

Редакционный совет журнала
«Вестник Российского университета дружбы народов»:

В. М. Филиппов — председатель
Д. П. Билибин — заместитель председателя

Члены редакционного совета:

*Т. М. Балыхина, Н. У. Венсковский, Н. П. Гусаков,
В. В. Давыдов, В. Н. Денисенко, А. П. Ефремов,
А. Я. Капустин, Н. С. Кирабаев, Е. Б. Ланеев,
В. Г. Плющиков, Н. К. Пономарев, Н. А. Черных,
В. А. Фролов, В. А. Цвык, В. Н. Шаронов, В. В. Якушев*

Т. О. Сергеева — ответственный секретарь

Редакционная коллегия серии:

Ю. П. Рыбаков — главный редактор
Л. А. Севастьянов — заместитель главного редактора
Д. С. Кулябов — ответственный секретарь

*А. В. Арутюнов, Г. П. Башарин, А. Н. Гордеев,
В. В. Давыдов, В. А. Комоцкий, Н. Н. Лобанов,
В. П. Милантьев, В. И. Михеев, Р. Г. Мухарлямов,
Н. Э. Николаев, К. Е. Самуйлов, А. Л. Скубачевский,
В. Д. Степанов, М. Ф. Сухинин*

ISSN 0869-8732

Адрес редакции:
117923, ГСП-1, Москва, ул. Орджоникидзе, 3

©Российский университет дружбы народов, Издательство, 2007

Вестник Российского университета дружбы народов

НАУЧНЫЙ ЖУРНАЛ

Основан в 1993 году

Серия

Математика. Информатика. Физика
2007, № 1–2

Российский университет дружбы народов

СОДЕРЖАНИЕ

МАТЕМАТИКА

- Башарин Г. П., Клапоуцак С. Н.** Управление ресурсами соты в
обычных и чрезвычайных ситуациях 5
- Бочаров П. П., Зарядов И. С.** Стационарное распределение ве-
роятностей в системах массового обслуживания с обновлением . . . 14

ИНФОРМАТИКА

- Кулябов Д. С., Королькова А. В., Хохлов А. А.** Подсистема
маршрутизации Click 24
- Королькова А. В.** Метод расчета вероятности сброса пакетов в ал-
горитме RED 32
- Кулябов Д. С., Кокотчикова М. Г.** Аналитический обзор систем
символьных вычислений 38
- Асламов Е. Б., Висков А. В., Фомин М. Б.** Свойства аналитиче-
ских измерений 46

ФИЗИКА

- Бештоев Х. М.** Осцилляции нейтрино в схеме смешивания зарядов
(констант связей) 53
- Саха Б.** Спинорные поля в космологии Бианки VI 62
- Саха Б., Шикин Г. Н.** Спинорные поля в плоско-симметричном
пространстве-времени 66
- Богданов Ю. П., Дидковская И. Ю.** Упругое рассеяние реактор-
ных антинейтрино на электронах и ограничение на массу дополни-
тельного нейтрального векторного бозона 70
- Черницкий А. А.** Об основных параметрах электромагнитных ча-
стиц и электромагнитном взаимодействии 75
- Рудой Ю. Г., Кейта И.** Динамическое давление и его флуктуации
для классического идеального газа релятивистских частиц 84
- Копысов Ю. С., Стожков Ю. И.** Электрогазодинамическая мо-
дель ускорения заряженных частиц во вспышках карликовых звезд
Главной последовательности 94

Магаршак Ю. Б. Резонансная модель атома и формирование валентных связей	100
Богомолов Ф., Магаршак Ю. Б. Группа Лоренца и ее роль в нерелятивистской модели атома	117
Амирханов И. В., Барбашов Б. М., Гусев А. А., Первухин В. Н., Шувалов С. А., Виницкий С. И., Захаров А. Ф., Зинчук В. А. Ньютоновский потенциал в Общей теории относительности в конечном пространстве	123
Рыбаков Ю. П. О возможном геометрическом происхождении симметрий Ю. Магаршака	136
Комоцкий В. А., Соколов Ю. М. Оптоэлектронный измеритель угловых колебаний конструкций	138
Сулейман С. Ш., Саакян А. А., Никогосян С., Юнусова С. А. Удельная электрическая проводимость натуральных цеолитов из Армении (клиноптилолит) и Сирии	147

Bulletin
of Peoples'
Friendship
University
of Russia

SCIENTIFIC JOURNAL

Founded on 1993

Series

Mathematics. Information Scinces. Physics.
2007, No 1–2

Peoples' Friendship University of Russia

CONTENTS

MATHEMATICS

Basharin G.P., Klapouschak S.N. Cell Resource Management in Ordinary and Emergency Situations	5
Bocharov P.P., Zaryadov I.S. Queueing Systems with Renovation. Stationary Probability Distribution	14

COMPUTER SCIENCES

Kulyabov D.S., Korolkova A.V., Khokhlov A.A. Routing Subsystem Click	24
Korolkova A. V. The Methods of Drop Probability Calculation for RED Algorithm	32
Kulyabov D. S., Kokotchikova M. G. Analytic Review of Computer Algebra System	38
Aslamov E. B., Viskov A. V., Fomin M. B. Analytical Dimensions Properties	46

PHYSICS

Beshtoev Kh.M. Neutrino Oscillations in the Scheme of Charge (Couple Constant)	53
Saha B. Spinor Fields in Bianchi Type-VI Cosmology	62
Saha B., Shikin G.N. Spinor Fields in Plane-Symmetric Space-Time	66
Bogdanov Yu. P., Didkovskaya I. Yu. Elastic Reactor Antineutrino Scattering by Electron and Bound on the Mass of Extra Neutral Gauge Boson	70
Chernitskii A. A. On General Parameters of Electromagnetic Particles and Electromagnetic Interaction	75
Rudoy Yu. G., Keita I. Dynamic Pressure and its Fluctuations for the Ideal Gas of Relativistic Particles	84

Kopysov Yu.S., Stozhkov Yu.I. Electrogasodynamic Model of Charged Particle Acceleration in Flares of the Main Sequence Dwarf Stars	94
Magarshak Yu.B. Resonant Atom Model and the Formation of Valence Bonds	100
Bogomolov F., Magarshak Yu.B. Lorentz Group and it's Role in the Non-Relativistic Atom Model	117
Amirkhanov I.V., Barbashov B.M., Gusev A.A., Per- vushin V.N., Shuvalov S.A., Vinitzky S.I., Zakharov A.F., Zinchuk V.A. Newton Potential in General Relativity in a Finite Volume	123
Rybakov Yu.P. On Possible Geometric Origin of Symmetries by Y. Magarshak	136
Komotskii V.A., Sokolov Yu.M. Optoelectronic Device for Measurements Angular Oscillations of Constructions	138
Soulayman S.Sh., Sahakyan A.A., Nikogosyan S., Yunusova S.A. Specific Electrical Conductivity of Natural Syrian and Armenian Zeolites	147

МАТЕМАТИКА

УДК 621.39

Управление ресурсами соты в обычных и чрезвычайных ситуациях

Г. П. Башарин, С. Н. Клапоущак

*Кафедра систем телекоммуникаций
Российский университет дружбы народов
Россия, 117198, Москва, ул. Миклуто-Маклая, 6*

В работе исследовано функционирование одной соты ССПС, в которую поступают три типа вызовов: экстренные, новые неэкстренные и хэндовер-вызовы. Построена математическая модель соты с буферами ожидания и резервными каналами для экстренных и хэндовер-вызовов. Предложен алгоритм доступа к ресурсам соты этих типов вызовов, позволяющий регулировать приоритет экстренных вызовов для обычных и чрезвычайных ситуаций. Получены формулы для основных ВВХ и разработан эффективный алгоритм решения СУГБ. Результаты работы могут быть использованы в практической деятельности операторов сотовой связи как в обычных ситуациях, так и в случае ЧС.

КЛЮЧЕВЫЕ СЛОВА: сети сотовой подвижной связи (ССПС), экстренные вызовы, хэндовер, чрезвычайные ситуации (ЧС), система уравнений глобального баланса (СУГБ), вероятностно-временные характеристики (ВВХ).

Введение

В конце XX и в начале XXI века сети сотовой подвижной связи (ССПС) развиваются очень быстро, становясь частью современной цивилизации и способствуя глобализации. ССПС развиваются не только количественно (число мобильных телефонов превысило число стационарных как в мире в целом, так и в отдельных странах, включая РФ), но и качественно — происходит достаточно быстрая смена технологий и повсеместно завершён переход от технологии 1G к технологии 2G, происходит переход от технологии 2G к технологиям 2.5G и 3G. Возникает много новых радиофизических, радиотехнических и математических проблем и задач, о чём свидетельствует огромный объём публикаций [1].

Значительная часть проблем требует для своего решения привлечения вероятностных методов, включая методы ТМО и ТТ [1, 2]. Одной из таких проблем является классическая проблема перегрузок в сетях связи, которая возникла в начале XX века и, собственно, привела к возникновению ТТ и способствовала развитию ТМО.

Перегрузки могут возникать в определённые периоды (например, в праздничные дни, во время крупных спортивных мероприятий), что позволяет к ним заранее подготовиться, увеличив ресурс сети к началу периода больших нагрузок. Но наибольшую трудность представляет собой проблема перегрузок, время которых невозможно предсказать — это стихийные бедствия, природные или технические катастрофы, а в последние годы и террористические акты. Все эти причины принято называть чрезвычайными ситуациями. В эти периоды нагрузка может во много раз превышать нормальную, а часть сети, например базовые станции (БС) мобильной сети, может выходить из строя полностью или частично.

Существуют различные подходы к исследованию поведения ССПС после ЧС. В работах [3, 4] представлен макроподход, когда рассматривается большой фрагмент сети (регион) и исследуется живучесть сотовой сети при разных типах ЧС.

Предполагается, что часть БС полностью теряет работоспособность и с помощью вероятностно-графовых методов исследуется вероятность установления соединений между абонентами посредством оставшихся работоспособных БС. Эта проблема близка к рассматривавшимся в последней четверти XX века задачам расчёта надёжности и живучести сетевых структур.

В работе [5] в рамках микроподхода рассматривается отдельная сота ССПС, которая не выходит из строя после ЧС, но на неё резко увеличивается поток экстренных вызовов. Этот поток рассматривается отдельно от обычных новых вызовов из рассматриваемой соты и хэндовер-вызовов из смежных сот, поскольку экстренные вызовы должны иметь наивысший приоритет.

В настоящей статье рассматривается одна сота ССПС, в которую поступают три потока вызовов: экстренные, хэндовер и новые неэкстренные вызовы. Предлагается ввести в соте два буфера ожидания: для экстренных и хэндовер-вызовов. При выборе вызова, находящегося в буфере, на обслуживание вводится алгоритм управления доступом, заключающийся в гибком регулировании выбора канала для обслуживания между экстренными и хэндовер-вызовами с помощью задаваемого параметра. Благодаря этому параметру возможно оперативное реагирование на изменение трафика при резком переходе от обычного режима функционирования соты к режиму работы в условиях ЧС. Также для улучшения качества обслуживания приоритетных типов вызовов предусматривается введение резервных каналов в соте. С использованием методики, изложенной в [2], строится математическая модель соты в виде трёхмерного ступенчатого марковского процесса (СтМП). Предлагается новый эффективный алгоритм решения СУГБ и определяется равновесное распределение, получены формулы для основных ВВХ.

1. Математическая модель

Всего на БС соты имеется C разговорных каналов. В неё поступают три потока вызовов: экстренные вызовы из этой же соты (1-вызовы), хэндовер-вызовы из соседних сот (2-вызовы) и новые неэкстренные вызовы (3-вызовы), возникающие в пределах рассматриваемой соты, с интенсивностями λ_1 , λ_2 и λ_3 соответственно. Время обслуживания вызова любого из трёх типов имеет экспоненциальное распределение с параметром μ . Из общего числа C каналов выделено g резервных каналов исключительно для обслуживания приоритетных типов вызовов: экстренных и хэндовер-вызовов.

В системе реализован принцип обслуживания с ожиданием как для 1-вызовов, так и для 2-вызовов. Если 1-вызов или 2-вызов застает при поступлении в соту все каналы занятыми, он помещается в буфер ожидания ёмкостью r_1 и r_2 соответственно. 1-вызов, находящийся в очереди, может покинуть её с интенсивностью ν_1 , не дождавшись обслуживания, вследствие специфики чрезвычайных ситуаций, когда жизнь абонента находится под угрозой и возможно возникновение непредвиденных помех при установлении связи с БС соты. 2-вызов, находящийся в очереди, также может покинуть её с интенсивностью ν_2 вследствие пересечения абонентом границы хэндовер-зоны. При этом происходит разрыв уже установленного соединения.

Произвольному вызову может быть отказано в обслуживании по следующим причинам:

- вызов застал все каналы и все места в буфере для ожидания (для экстренных и хэндовер-вызовов) занятыми;
- экстренный или хэндовер-вызов после попадания в буфер ушёл из системы, не дождавшись обслуживания.

На систему накладываются следующие ограничения:

- все потоки вызовов, поступающие в соту, пуассоновские и независимые;
- после окончания обслуживания в соте разговор прекращается или вызов покидает соту, не оказывая влияния на другие вызовы;
- экстренные вызовы не используют процедуру хэндовера.

На рис. 1 представлена схема рассматриваемой системы. На нём дополнительно введены обозначения n_i , $i = 1, 2$ — количество i -вызовов в соответствующей очереди.

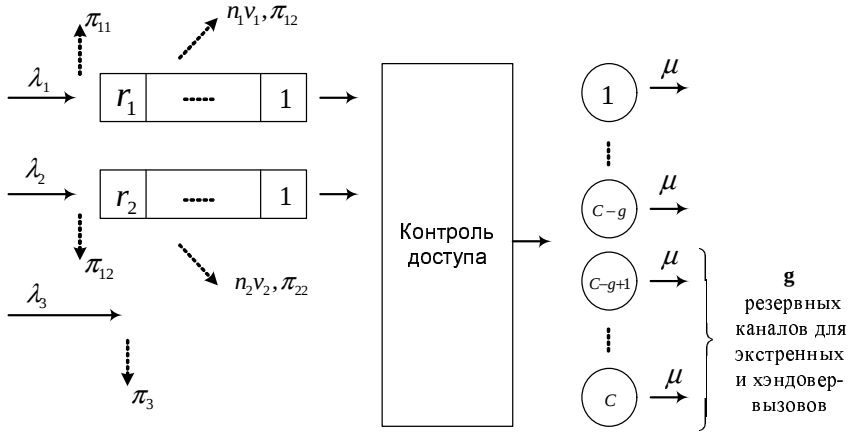


Рис. 1. Модель рассматриваемой соты.

Контроль доступа к ресурсам соты осуществляется следующим образом: зададимся некоторым коэффициентом b (т.н. коэффициент взвешивания). Тогда при наличии в обеих очередях хотя бы по одному вызову экстренный вызов будет принят на обслуживание с вероятностью

$$A = P\{bT_1 < T_2\}, \quad (1)$$

а хэндовер-вызов — с вероятностью

$$\bar{A} := 1 - A = P\{bT_1 > T_2\}, \quad (2)$$

где T_i , $i = 1, 2$ — время, оставшееся до ухода i -вызова из очереди.

Поскольку в рассматриваемом случае известно, что СВ T_1 и T_2 имеют экспоненциальное распределение, мы можем получить значения вероятностей (1) и (2) в явном виде. Как известно из курса теории вероятностей, если СВ T_1 распределена экспоненциально с параметром ν_1 , то СВ bT_1 также будет иметь экспоненциальное распределение, но с параметром $\frac{\nu_1}{b}$. Тогда для вероятностей (1) можно получить

$$\begin{aligned} P\{bT_1 < T_2\} &= \int_0^\infty P\{T_2 = c\}P\{bT_1 < c\}dc = \\ &= \int_0^\infty f_{T_2}(c)F_{bT_1}(c)dc = \int_0^\infty \nu_2 e^{-\nu_2 c} \left(1 - e^{-\frac{\nu_1}{b}c}\right) dc = \\ &= -e^{-\nu_2 c} \Big|_0^\infty + \frac{\nu_2 e^{-(\nu_2 + \frac{\nu_1}{b})c}}{\nu_2 + \frac{\nu_1}{b}} \Big|_0^\infty = 1 - \frac{\nu_2}{\nu_2 + \frac{\nu_1}{b}} = \frac{\frac{\nu_1}{b}}{\nu_2 + \frac{\nu_1}{b}}. \end{aligned}$$

Откуда, выполняя аналогичные преобразования для (2),

$$A = \frac{\frac{\nu_1}{b}}{\nu_2 + \frac{\nu_1}{b}} = \frac{\nu_1}{\nu_1 + b\nu_2} \rightarrow \begin{cases} 1, & b \rightarrow 0 \\ 0, & b \rightarrow \infty \end{cases}, \quad (3)$$

$$\bar{A} = \frac{\nu_2}{\nu_2 + \frac{\nu_1}{b}} = \frac{b\nu_2}{\nu_1 + b\nu_2} \rightarrow \begin{cases} 0, & b \rightarrow 0 \\ 0, & b \rightarrow \infty \end{cases}. \quad (4)$$

Предложенный алгоритм доступа включает в себя в качестве частных случаев алгоритмы выбора вызова из очереди на обслуживание с относительным приоритетом для любого из классов вызовов: при $b = 0$ относительный приоритет получают экстренные вызовы, а при $b = \infty$ — хэндовер-вызовы, поскольку при этом $A = 1$ и $A = 0$ соответственно.

Поведение рассматриваемой системы во времени описывается трёхмерным СтМП

$$\mathbf{X}(t) = (X_1(t), X_2(t), X_3(t)), \quad t \geq 0, \quad (5)$$

где

$X_1(t)$ — число вызовов, находящихся на обслуживании в момент времени t ,
 $X_2(t)$ — число экстренных вызовов в очереди в момент времени t ,
 $X_3(t)$ — число хэндовер-вызовов в очереди в момент времени t .

Пространство состояний этого процесса имеет вид

$$\mathbf{S} = \{(c, 0, 0) \cup (C, i, j) : c = \overline{0, C}; i = \overline{0, r_1}; j = \overline{0, r_2}\}.$$

Матрицу интенсивностей переходов процесса $\mathbf{X}(t)$ обозначим $A = [a(s, t)]_{s, t \in \mathbf{S}}$. Граф интенсивностей переходов процесса представлен на рис. 2.

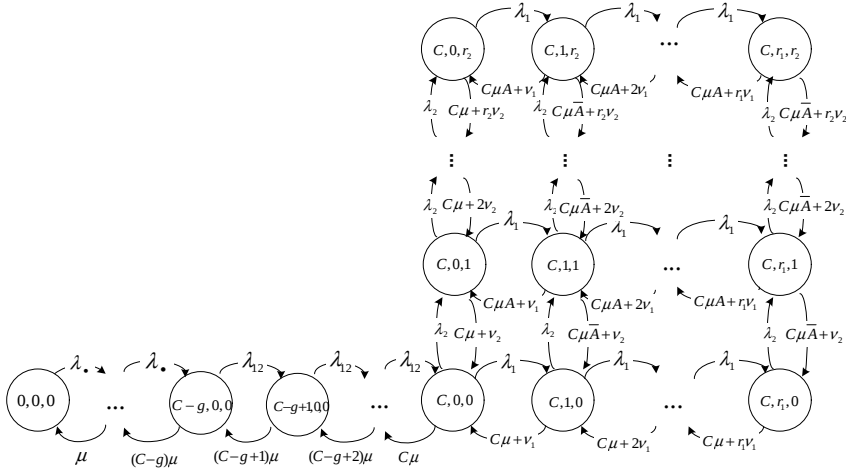


Рис. 2. Граф интенсивностей переходов случайного процесса $X(t)$.

Пусть выполняются все условия для существования равновесного распределения $P(c, i, j)$, $(c, i, j) \in \mathbf{S}$ МП (5) [2]. Для краткости обозначим равновесные вероятности процесса $\mathbf{X}(t)$, когда заняты все C разговорных каналов следующим образом

$$P(C, i, j) =: p_{i,j}; \quad i = \overline{1, r_1}, \quad j = \overline{1, r_2}.$$

По аналогии с [2] нетрудно показать, что СУГБ для процесса $X(t)$ имеет вид

$$\begin{aligned} [\lambda_1 + \lambda_2 + u(c - C + g + 1)\lambda_3 + u(c)c\mu]P(c, 0, 0) = \\ = u(c)[\lambda_1 + \lambda_2 + u(c - C + g)\lambda_3]P(c - 1, 0, 0) + \\ + (c + 1)\mu P(c + 1, 0, 0); \quad c = \overline{0, C - 1}; \end{aligned}$$

$$(\lambda_1 + \lambda_2 + C\mu)P(C, 0, 0) = (\lambda_1 + \lambda_2)P(C - 1, 0, 0) + (C\mu + \nu_1)p_{1,0} + (C\mu + \nu_2)p_{0,1};$$

$$\begin{aligned}
& \{u(r_1 - i)\lambda_1 + u(r_2 - j)\lambda_2 + u(i)C\mu[A - (A - 1)(1 - u(j))]\} + \\
& \quad + i\nu_1 + u(j)C\mu[\bar{A} - (\bar{A} - 1)(1 - u(i))] + j\nu_2\}p_{i,j} = \\
& = u(i)\lambda_1 p_{i-1,j} + u(j)\lambda_2 p_{i,j-1} + \\
& \quad + \{u(r_1 - i)C\mu[A - (A - 1)(1 - u(j))] + (i + 1)\nu_1\}p_{i+1,j} + \\
& \quad + \{u(r_2 - j)C\mu[\bar{A} - (\bar{A} - 1)(1 - u(i))] + (j + 1)\nu_2\}p_{i,j+1}; \\
& \quad i = \overline{0, r_1}, \quad j = \overline{0, r_2}, \quad i + j > 0.
\end{aligned} \tag{6}$$

2. Решение СУГБ

Решение полученной СУГБ стандартными матричными методами приводит к значительным вычислительным трудностям. В данной работе предложен алгоритм вычисления стационарных вероятностей, использующий специфический вид графа интенсивностей переходов процесса $X(t)$. Предложенный алгоритм при больших значениях структурных параметров системы показывает гораздо большую скорость вычислений, чем решение СУГБ напрямую. Суть алгоритма заключается в сведении СУГБ (6) размерности $C + (r_1 + 1)(r_2 + 1)$ к СЛУ размерности r_1 (или r_2 , в зависимости от значений этих параметров) с вероятностями $p_{0,k} =: x_k$; $k = \overline{1, r_2}$ (6) в качестве неизвестных. Для остальных вероятностей находится их выражение через линейную комбинацию значений (6).

2.1. Параметры производительности системы

Теперь вероятности блокировок (т. е. вероятности мгновенной потери на входе в систему вызовов соответствующего типа) для трёх типов вызовов могут быть записаны следующим образом

$$\pi_{11} = p_{r_1, \bullet}; \quad \pi_{21} = p_{\bullet, r_2}; \quad \pi_3 = p_{\bullet, \bullet} = 1 - \sum_{k=0}^{C-1} P\{k, 0, 0\} \tag{7}$$

для экстренных, хэндовер- и новых неэкстренных вызовов соответственно.

Вероятности ухода из очереди без обслуживания для экстренных и хэндовер-вызовов можно вычислить, исходя из следующих соображений.

Введём случайные величины ζ_1 и ζ_2 — количество вызовов, ожидающих обслуживания соответственно в очереди для 1-вызовов и 2-вызовов. Матожидания этих величин можно получить по формулам

$$\begin{aligned}
E\zeta_1 &= \sum_{i=1}^{r_1} i p_{i, \bullet}, \\
E\zeta_2 &= \sum_{j=1}^{r_2} j p_{\bullet, j}.
\end{aligned}$$

Если очередь 1-вызовов не пуста, то $\nu_1 E\zeta_1$ — интенсивность ухода вызовов из этой очереди без обслуживания, а $C\mu \sum_{i=1}^{r_1} p_{i,0} + C\mu A \sum_{i=1}^{r_1} \sum_{j=1}^{r_2} p_{i,j}$ — интенсивность ухода вызовов из неё же на обслуживание. Поэтому имеет место равенство интенсивностей потоков 1-вызовов, поступающих в очередь r_1 и покидающих её

$$\lambda_1 \sum_{i=0}^{r_1-1} p_{i, \bullet} = \nu_1 E\zeta_1 + C\mu \sum_{i=1}^{r_1} p_{i,0} + C\mu A \sum_{i=1}^{r_1} \sum_{j=1}^{r_2} p_{i,j}.$$

Из этого соотношения можно вычислить вероятность (по интенсивности) π_{12} ухода 1-вызова из очереди без обслуживания

$$\pi_{12} = \frac{\nu_1 E_{\zeta_1}}{\lambda_1 \sum_{i=0}^{r_1-1} p_{i,\bullet}}. \quad (8)$$

Аналогично, для очереди r_2 , $\nu_2 E_{\zeta_2}$ — интенсивность ухода вызовов из этой очереди без обслуживания, а $C\mu \sum_{j=1}^{r_2} p_{0,j} + C\mu \bar{A} \sum_{j=1}^{r_2} \sum_{i=1}^{r_1} p_{i,j}$ — интенсивность ухода вызовов из неё же на обслуживание. Таким образом, баланс интенсивностей потоков 2-вызовов, поступающих в очередь r_1 и покидающих её, имеет вид

$$\lambda_2 \sum_{j=0}^{r_2-1} p_{\bullet,j} = \nu_2 E_{\zeta_2} + C\mu \sum_{j=1}^{r_2} p_{0,j} + C\mu \bar{A} \sum_{j=1}^{r_2} \sum_{i=1}^{r_1} p_{i,j}.$$

Откуда, аналогично формуле (8), можно получить вероятность (по интенсивности) π_{22} разрыва соединения, ожидающего обслуживания в хэндовер-очереди r_2

$$\pi_{22} = \frac{\nu_2 E_{\zeta_2}}{\lambda_2 \sum_{j=0}^{r_2-1} p_{\bullet,j}}. \quad (9)$$

2.2. Численный анализ

Для примера возьмём систему с числом каналов 30 и двумя резервными каналами, буфером для экстренных вызовов из 3 мест для ожидания и буфером для хэндовер-вызовов из 3 мест для ожидания. Среднее время разговора примем равным 100 секундам, среднее время жизни в очереди 20 секунд и 10 секунд для экстренных и хэндовер-вызовов, соответственно. Будем считать, что наибольшую нагрузку на систему из трех типов вызовов обеспечивают новые неэкстренные вызовы, несколько меньшую — экстренные и хэндовер-вызовы. Все основные параметры системы перечислены в табл. 1.

Основные параметры системы для численного примера

Таблица 1

C	g	r_1	r_2	$\frac{\nu_1}{\mu}$	$\frac{\nu_2}{\mu}$	ρ_1	ρ_2	ρ_3
30	2	3	3	1.5	1.5	15 Эрл	15 Эрл	20 Эрл

На рис. 3 показана зависимость вероятностей блокировок вызовов всех трёх типов от значения коэффициента взвешивания. Как можно заметить, при значениях b , близких к нулю, вероятность π_{21} значительно превосходит вероятность π_{11} . Это объясняется тем фактом, что при этих значениях b на обслуживание выбираются, в основном, 1-вызовы из очереди. Если же b увеличивается, стремясь к бесконечности, значение π_{11} увеличивается и после определенного момента становится больше, чем значение π_{21} . При $b = 1$ значения двух вероятностей совпадают, поскольку 1-вызовы и 2-вызовы из очередей при этом выбираются на обслуживание с равной вероятностью. Надо отметить, что подобное совпадение имеет место именно в точке $b = 1$ вследствие того, что потоки 1-вызовов и 2-вызовов имеют одинаковые характеристики: создают одинаковую нагрузку, совпадают размеры буферов ожидания и интенсивности нетерпения. При различиях в перечисленных характеристиках двух потоков точка совпадения вероятностей будет смещена. Динамика изменений вероятностей π_{11} и π_{21} ярко демонстрирует возможность

гибкого разделения приоритета между 1-вызовами и 2-вызовами. Также на рис. 3 можно увидеть, что зависимости от коэффициента взвешивания и связь между собой вероятностей π_{12} и π_{22} ухода вызовов из очереди без обслуживания имеют такой же вид, как и для вероятностей π_{11} и π_{21} соответственно.

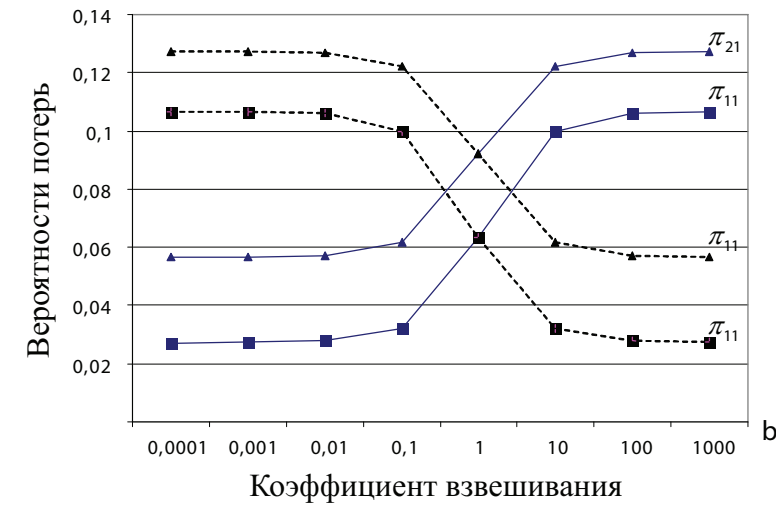


Рис. 3. График зависимости вероятностей блокировок вызовов от коэффициента взвешивания: $C = 30, g = 2, r_1 = 3, r_2 = 3, \rho_1 = 15, \rho_2 = 15, \rho_3 = 20, \frac{\nu_1}{\mu} = 1.5, \frac{\nu_2}{\mu} = 1.5$.

На рис. 4 представлен график зависимости вероятностей π_{11} и π_{21} блокировки 1-вызовов и 2-вызовов соответственно при входе в систему от значения коэффициента взвешивания при различных значениях ёмкости r_1 буфера накопителя для 1-вызовов.

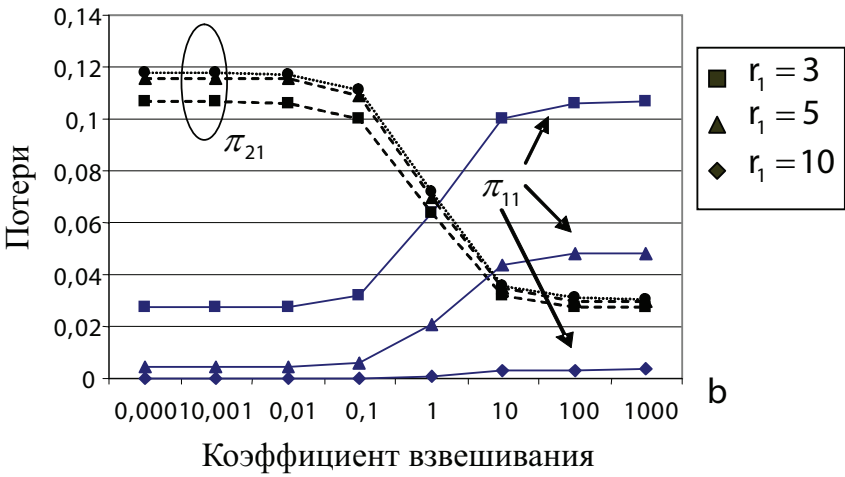


Рис. 4. График зависимости вероятностей потерь от коэффициента взвешивания при $r_1 = 3, 5, 10$: $C = 30, g = 2, r_2 = 3, \rho_1 = 15, \rho_2 = 15, \rho_3 = 20, \frac{\nu_1}{\mu} = 1.5, \frac{\nu_2}{\mu} = 1.5$.

На рис. 4 видно, что при увеличении ёмкости буфера для 1-вызовов вероятность π_{11} блокировки 1-вызовов сильно уменьшается. При $r_1 = 10$ её значение не превосходит 1% при любых значения коэффициента взвешивания b . Вероятность π_{21} блокировки 2-вызова при увеличении ёмкости буфера в диапазоне $r_1 = \overline{1,6}$

заметно увеличивается при изменении r_1 , однако при дальнейшем росте значений r_1 значения π_{21} практически стабилизируются.

Заметим также, что при $r_1 = 10$ $\pi_{11} < \pi_{21}$ для любых b . Таким образом увеличение ёмкости буфера может служить более радикальным методом уменьшения вероятности потерь экстренных вызовов, чем контроль доступа с помощью коэффициента взвешивания, однако организация буферов для ожидания в ССПС является сложной технической задачей.

Таким образом, после проведения численного анализа можно сделать вывод, что нельзя рассматривать каждый способ предоставления приоритета вызовам отдельно от других. Необходимо подбирать параметры системы так, чтобы при уменьшении вероятностей блокировок экстренных вызовов вероятности блокировок новых неэкстренных вызовов не принимали критических значений.

Заключение

В статье изучено функционирование одной соты ССПС с учётом поступающих в неё экстренных вызовов. Для обслуживания экстренных и хэндовер-вызовов вводятся резервные каналы и буферы для ожидания. Кроме того, вводится контроль доступа для гибкого регулирования приоритета между этими типами вызовов в обычных и чрезвычайных ситуациях. Для анализа параметров производительности разработана соответствующая математическая модель в виде трёхпоточной СМО с ожиданием. Предложен и реализован алгоритм решения СУГБ. Численный анализ показал, что комплекс предлагаемых методов позволяет обеспечить высокие показатели качества обслуживания для экстренных вызовов в период чрезвычайных ситуаций, а также улучшить обслуживание вызовов служб 01, 02, 03 в обычных ситуациях.

Литература

1. Handbook of wireless networks and mobile computing / Ed. by I. Stojmenovic. — NY: J. Wiley & sons, 2002. — 630 p.
2. Башарин Г. П. Лекции по математической теории телетрафика. — М.: Изд-во РУДН, 2004.
3. Величко В. В., Шахов В. В. Оценка живучести систем мобильной радиосвязи в условиях природных и техногенных катастроф // Мобильные системы. — № 10. — 2005. — С. 35–38.
4. Величко В. В., Юргенсон А. Н. Модели структурной надёжности в мобильных сетях передачи данных // Электросвязь. — № 3. — 2006. — С. 36–37.
5. Beard C., Zhou J. Weighted Earliest Deadline Scheduling and Its Analytical Solution for Admission Control in a Wireless Emergency Network // ITC19. — 2005. — Pp. 2229–2238.

UDC 621.39

Cell Resource Management in Ordinary and Emergency Situations

G. P. Basharin, S. N. Klapouschak

*Telecommunication Systems Department
Peoples' Friendship University of Russia
6, Miklukho-Maklaya str., Moscow, 117198, Russia*

A mobile network cell with emergency, handover and new arrival calls is investigated. A mathematical model with queuing and guard channels for handover and emergency call is developed. An admission control to cell resources for the above two types of calls is proposed that provides flexible priority to emergency calls in ordinary and emergency situations. The formulas for performance analysis are obtained and effective algorithm for solving system of balance equations is provided. This paper's results can be applied by cellular networks operators both in ordinary and emergency cases.

УДК 519.872

Стационарное распределение вероятностей в системах массового обслуживания с обновлением

П. П. Бочаров, **И. С. Зарядов**

*Кафедра теории вероятностей и математической статистики
Российский университет дружбы народов
Россия, 117198, Москва, ул. Миклухо-Маклая, 6*

Рассматриваются марковские системы массового обслуживания с обновлением. Для этих систем на основе доказанной в работе для обобщенного процесса размножения и гибели теоремы получены алгоритмы нахождения стационарных вероятностей состояний. В качестве примеров приведены алгоритмы расчетов для экспоненциальных системы с обновлением — системы $M/M/n/r$ без дообслуживанием и системы $M/M/n/r$ с дообслуживанием.

КЛЮЧЕВЫЕ СЛОВА: марковский процесс, матричное решение, стационарное распределение, обновление, дообслуживание.

Введение

Рассмотрим две системы массового обслуживания (СМО) с обновлением.

В СМО первого типа (без дообслуживания) заявка, окончившая обслуживание на приборе, с вероятностью q убивает все заявки в накопителе и покидает СМО, а с дополнительной вероятностью $p = 1 - q$ просто покидает систему. Вероятность q называется вероятностью обновления.

В СМО второго типа (с дообслуживанием) с вероятностью q обслужившаяся заявка остается в системе, убивая все заявки в накопителе, и с вероятностью $p = 1 - q$ покидает систему. Здесь вероятность q также называется вероятностью обновления.

Интерес к данному классу систем массового обслуживания связан с развитием компьютерных и телекоммуникационных систем. Впервые СМО с обновлением были рассмотрены А.Я. Крейниным [1, 2] и применялись для моделирования работы системы, в которую поступают потоки команд (заявок) нескольких типов (в том числе поток исполняемых команд) и возможен переход от исполняемой команды к иной команде с потерей промежуточных команд. В настоящее время этот подход начинают применять к моделированию систем, в которых обслуживание (а точнее говоря — момент окончания обслуживания) может вызвать сбой в работе системы, чреватый потерями заявок. В частности, в [3] были изучены вероятностные модели протоколов для систем, чувствительных к отказам. В этой же работе были построены системы массового обслуживания с приоритетами и обновлением. Отметим, что СМО с обновлением имеют также интересную интерпретацию с точки зрения математических моделей финансов и экономики, где с их помощью описывается стоимость рискованной ценной бумаги на финансовом рынке. При этом переход в нулевое состояние интерпретируется как банкротство (дефолт) компании, выпустившей данный рискованный актив.

В настоящей работе, в отличие от [1], в которой стационарное распределение вероятностей состояний найдено в терминах производящей функции, предлагается алгоритм поиска этого распределения, основанный на предложенном в [4] для обобщенного ПРГ методе.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (РФФИ грант №06-07-89056).

1. Вспомогательные результаты

Рассмотрим однородный неприводимый марковский процесс с непрерывным временем, конечным множеством состояний и инфинитезимальной матрицей Q следующего вида:

$$Q = \begin{pmatrix} N_0 & \Lambda & 0 & \dots & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ M_0 & N_1 & \Lambda & \dots & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ \tilde{M}_0 & M_1 & N_2 & \dots & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ \tilde{M}_0 & \tilde{M}_1 & M_2 & \dots & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \tilde{M}_0 & \tilde{M}_1 & \tilde{M}_2 & \dots & \Lambda & 0 & \dots & 0 & 0 & 0 & 0 \\ \tilde{M}_0 & \tilde{M}_1 & \tilde{M}_2 & \dots & N_l & \Lambda & \dots & 0 & 0 & 0 & 0 \\ \tilde{M}_0 & \tilde{M}_1 & \tilde{M}_2 & \dots & M_l & N & \dots & 0 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \tilde{M}_0 & \tilde{M}_1 & \tilde{M}_2 & \dots & \tilde{M}_l & 0 & \dots & 0 & M & N & \Lambda \\ \tilde{M}_0 & \tilde{M}_1 & \tilde{M}_2 & \dots & \tilde{M}_l & 0 & \dots & 0 & 0 & M & N_{r+1} \end{pmatrix}, \quad (1)$$

где $\Lambda, N, M, N_k, \tilde{M}_k, M_k$ — квадратные матрицы, причем $M_k = M + \tilde{M}_k, k = \overline{0, l}$ и $l < r - 2$.

Как видно из (1), начиная с $l + 1$ -го столбца матрица, становится трехдиагональной.

В частности, для многолинейной СМО с конечным накопителем и обновлением без дообслуживания матрица Q имеет вид

$$Q = Q_1 = \begin{pmatrix} N_0 & \Lambda & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ M_1 & N_1 & \Lambda & 0 & \dots & 0 & 0 & \dots & 0 & 0 \\ 0 & M_2 & N_2 & \Lambda & \dots & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & M_3 & N_3 & \dots & 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & \Lambda & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & N_{n-1} & \Lambda & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & \tilde{M}_n & N & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & \tilde{M} & M & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & \tilde{M} & 0 & \dots & \Lambda & 0 \\ 0 & 0 & 0 & 0 & \dots & \tilde{M} & 0 & \dots & N & \Lambda \\ 0 & 0 & 0 & 0 & \dots & \tilde{M} & 0 & \dots & M & N_{r+1} \end{pmatrix}, \quad (2)$$

где $\tilde{M}_n = M_n + \tilde{M}$, а для многолинейной СМО с конечным накопителем и обновлением с дообслуживанием вид

$$Q = Q_2 = \begin{pmatrix} N_0 & \Lambda & 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ M_1 & N_1 & \Lambda & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & M_2 & N_2 & \Lambda & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & M_3 & N_3 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & N_{n-1} & \Lambda & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & M_n & \tilde{N}_n & \Lambda & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & \tilde{M}_n & N & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & \tilde{M} & M & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 0 & \tilde{M} & 0 & \dots & \Lambda & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & \tilde{M} & 0 & \dots & N & \Lambda \\ 0 & 0 & 0 & 0 & \dots & 0 & \tilde{M} & 0 & \dots & M & N_{r+1} \end{pmatrix}, \quad (3)$$

где $\tilde{M}_n = M_n + \tilde{M}$, $\tilde{N}_n = N + \tilde{M}$.

Для марковского процесса с матрицей (1) стационарное распределение вероятностей состояний $\vec{p} = (\vec{p}_0, \vec{p}_1, \dots, \vec{p}_{r+1})^T$, где $\vec{p}_k$ — компоненты вектора $\vec{p}$, находятся из системы уравнений равновесия (СУР)

$$\vec{p}_0^T N_0 + \vec{p}_1^T M_0 + \sum_{j=2}^{r+1} \vec{p}_j^T \tilde{M} = \vec{0}^T, \quad (4)$$

$$\vec{p}_{k-1}^T \Lambda + \vec{p}_k^T N + \vec{p}_{k+1}^T M_k + \sum_{j=k+2}^{r+1} \vec{p}_j^T \tilde{M}_k = \vec{0}^T, \quad k = \overline{1, l}, \quad (5)$$

$$\vec{p}_{k-1}^T \Lambda + \vec{p}_k^T N + \vec{p}_{k+1}^T M = \vec{0}^T, \quad k = \overline{l+1, r}, \quad (6)$$

$$\vec{p}_r^T \Lambda + \vec{p}_{r+1}^T N_{r+1} = \vec{0}^T \quad (7)$$

с условием нормировки

$$\sum_{k=0}^{r+1} \vec{p}_k^T \vec{1} = 1. \quad (8)$$

Для марковских процессов с матрицами (2) и (3) имеем следующие СУР:

$$\vec{p}_0^T N_0 + \vec{p}_1^T M_1 = \vec{0}^T, \quad (9)$$

$$\vec{p}_{k-1}^T \Lambda + \vec{p}_k^T N_k + \vec{p}_{k+1}^T M_{k+1} = \vec{0}^T, \quad k = \overline{1, n-2}, \quad (10)$$

$$\vec{p}_{n-2}^T \Lambda + \vec{p}_{n-1}^T N_{n-1} + \vec{p}_n^T \tilde{M}_n + \sum_{n+1}^{n+r} \vec{p}_j^T \tilde{M} = \vec{0}^T, \quad (11)$$

$$\vec{p}_{k-1}^T \Lambda + \vec{p}_k^T N + \vec{p}_{k+1}^T M = \vec{0}^T, \quad k = \overline{n, n+r-1}, \quad (12)$$

$$\vec{p}_{n+r-1}^T \Lambda + \vec{p}_{n+r}^T N_{n+r} = \vec{0}^T, \quad (13)$$

и

$$\vec{p}_0^T N_0 + \vec{p}_1^T M_1 = \vec{0}^T, \quad (14)$$

$$\vec{p}_{k-1}^T \Lambda + \vec{p}_k^T N_k + \vec{p}_{k+1}^T M_{k+1} = \vec{0}^T, \quad k = \overline{1, n-1}, \quad (15)$$

$$\vec{p}_{n-1}^T \Lambda + \vec{p}_n^T \tilde{N}_n + \vec{p}_{n+1}^T \tilde{M}_{n+1} + \sum_{n+2}^{n+r} \vec{p}_j^T \tilde{M} = \vec{0}^T, \quad (16)$$

$$\vec{p}_{k-1}^T \Lambda + \vec{p}_k^T N + \vec{p}_{k+1}^T M = \vec{0}^T, \quad k = \overline{n+1, n+r-1}, \quad (17)$$

$$\vec{p}_{n+r-1}^T \Lambda + \vec{p}_{n+r}^T N_{n+r} = \vec{0}^T. \quad (18)$$

Рассмотрим теперь систему уравнений

$$\vec{z}_{k-1}^T A + \vec{z}_k^T B + \vec{z}_{k+1}^T C = \vec{0}^T, \quad k = \overline{1, l-1}, \quad (19)$$

где A , B и C — квадратные матрицы одинакового порядка, а $\vec{z}_0, \vec{z}_1, \dots, \vec{z}_l$ — векторы той же размерности.

Теорема 1. Пусть существуют матрицы R и S такие что $A + RB + R^2C = 0$, $S^2A + SB + C = 0$, и не вырождена матрица $SA + B + RC$. Тогда общее решение системы уравнений (19) представимо в виде

$$\vec{z}_k^T = \vec{f}^T R^k + \vec{g}^T S^{l-k}, \quad k = \overline{0, l}, \quad (20)$$

где $\vec{f}$ и $\vec{g}$ — произвольные векторы¹.

Отметим, что утверждение теоремы 1 применимо к системе уравнений (19) с любыми матрицами A , B и C (а не только с матрицами интенсивностей переходов).

Перейдем непосредственно к матрице интенсивностей переходов (1).

Теорема 2. Пусть существуют матрицы R и S такие что: $\Lambda + RN + R^2M = 0$ и $S^2\Lambda + SN + M = 0$, кроме того, не вырождена матрица $S\Lambda + N + RM$. В этом случае вектор стационарных вероятностей $\vec{p} = (\vec{p}_0, \vec{p}_1, \dots, \vec{p}_{r+1})^T$ удовлетворяет СУР $\vec{p}^T Q = \vec{0}^T$ с матрицей Q вида (1) тогда и только тогда, когда для некоторых векторов $\vec{f}$ и $\vec{g}$ справедливы равенства:

$$\vec{p}_k^T = \vec{f}^T R^{k-l} + \vec{g}^T S^{r+1-k}, \quad k = \overline{l, r+1}, \quad (21)$$

$$\begin{aligned} \vec{p}_0^T N_0 + \vec{p}_1^T M_0 + \sum_{j=2}^l \vec{p}_j^T \tilde{M}_0 + \vec{f}^T \left(\sum_{k=l+1}^{r+1} R^{k-l} \right) \tilde{M}_0 + \\ + \vec{g}^T \left(\sum_{k=l+1}^{r+1} S^{r+1-k} \right) \tilde{M}_0 = \vec{0}^T, \end{aligned} \quad (22)$$

$$\begin{aligned} \vec{p}_{k-1}^T \Lambda + \vec{p}_k^T N_k + \vec{p}_{k+1}^T M_k + \sum_{j=k+2}^l \vec{p}_j^T \tilde{M}_k + \vec{f}^T \left(\sum_{k=l+1}^{r+1} R^{k-l} \right) \tilde{M}_k + \\ + \vec{g}^T \left(\sum_{k=l+1}^{r+1} S^{r+1-k} \right) \tilde{M}_k = \vec{0}^T, \quad k = \overline{1, l-2}, \end{aligned} \quad (23)$$

$$\vec{p}_{l-2}^T \Lambda + \vec{p}_{l-1}^T N_{l-1} + (\vec{f}^T + \vec{g}^T S^{r+1-l}) M_{l-1} +$$

¹Доказательство теоремы см. в [4] или [5].

$$+ \vec{f}^T \left(\sum_{k=l+1}^{r+1} R^{k-l} \right) \tilde{M}_{l-1} + \vec{g}^T \left(\sum_{k=l+1}^{r+1} S^{r+1-k} \right) \tilde{M}_{l-1} = \vec{0}^T, \quad (24)$$

$$\begin{aligned} \vec{p}_{l-1}^T \Lambda + \vec{f}^T (N_l + R M_l) + \vec{g}^T S^{r-l} (M_l + S N_l) + \\ + \vec{f}^T \left(\sum_{k=l+2}^{r+1} R^{k-l} \right) \tilde{M}_l + \vec{g}^T \left(\sum_{k=l+2}^{r+1} S^{r+1-k} \right) \tilde{M}_l = \vec{0}^T, \end{aligned} \quad (25)$$

$$\vec{f}^T R^{r-l} (R N_{r+1} + \Lambda) + \vec{g}^T (N_{r+1} + S \Lambda) = \vec{0}^T. \quad (26)$$

Доказательство. Согласно теореме 1 уравнения (6) имеют место тогда и только тогда, когда для некоторых векторов $\vec{f}$ и $\vec{g}$ справедливы выражения (21)

$$\vec{p}_k^T = \vec{f}^T R^{k-l} + \vec{g}^T S^{r+1-k}, \quad k = \overline{l, r+1}.$$

Учитывая, что $M_k = M + \tilde{M}_k$, выразим из (21) векторы $\vec{p}_{l-1}$, $\vec{p}_l$:

$$\vec{p}_l = \vec{f}^T + \vec{g}^T S^{r+1-l}, \quad \vec{p}_{l-1} = \vec{f}^T R + \vec{g}^T S^{r-l}.$$

Подставляя полученные значения $\vec{p}_{l-1}$ и $\vec{p}_l$ в (6) при $k = l$ и $l-1$, приходим к уравнениям (24)–(25).

В выражении (4) представим сумму $\sum_{j=2}^{r+1} \vec{p}_j^T \tilde{M}$ в виде

$$\sum_{j=2}^{r+1} \vec{p}_j^T \tilde{M} = \sum_{j=2}^l \vec{p}_j^T \tilde{M} + \sum_{j=l+1}^{r+1} \vec{p}_j^T \tilde{M}.$$

Подставляя (21) в последнее слагаемое вместо $\vec{p}_j^T$ и производя аналогичную операцию с (5) при $k = \overline{1, l-2}$, приходим к (22)–(23).

Для векторов $\vec{p}_r$ и $\vec{p}_{r+1}$ из (21) имеем

$$\vec{p}_r = \vec{f}^T R^{r-l} + \vec{g}^T S, \quad \vec{p}_{r+1} = \vec{f}^T R^{r+1-l} + \vec{g}^T.$$

Подставляя полученные выражения в (7), получаем (26). $\square$

Из теоремы 2 непосредственно выводятся следующие следствия, позволяющие найти стационарное распределение числа заявок в СМО с матрицами интенсивностей (2) и (3).

Следствие 1. Пусть существуют матрицы R и S такие что: $\Lambda + RN + R^2M = 0$ и $S^2\Lambda + SN + M = 0$, кроме того, не вырождена матрица $S\Lambda + N + RM$. В этом случае вектор $\vec{p} = (\vec{p}_0, \vec{p}_1, \dots, \vec{p}_{n+r})^T$ удовлетворяет СУР $\vec{p}^T Q_1 = \vec{0}^T$ с матрицей Q_1 вида (2) тогда и только тогда, когда для некоторых векторов $\vec{f}$ и $\vec{g}$ справедливы равенства:

$$\vec{p}_k^T = \vec{f}^T R^{k-n+1} + \vec{g}^T S^{n+r-k}, \quad k = \overline{n-1, n+r},$$

$$\begin{aligned} \vec{p}_{n-2}^T \Lambda + \vec{f}^T (N_{n-1} + R \tilde{M}_n) + \vec{g}^T S^r (S N_{n-1} + \tilde{M}_n) + \\ + \vec{f}^T \left(\sum_{j=n+1}^{n+r} R^{j-n+1} \right) \tilde{M} + \vec{g}^T \left(\sum_{j=n+1}^{n+r} S^{n+r-j} \right) \tilde{M} = \vec{0}^T, \\ \vec{f}^T R^r (R N_{n+r} + \Lambda) + \vec{g}^T (N_{n+r} + S \Lambda) = \vec{0}^T. \end{aligned}$$

Следствие 2. Пусть для некоторых матриц R и S верны выражения $\Lambda + RN + R^2M = 0$ и $S^2\Lambda + SN + M = 0$, кроме того, не вырождена матрица $S\Lambda + N + RM$. В этом случае вектор $\vec{p} = (\vec{p}_0, \vec{p}_1, \dots, \vec{p}_{n+r})^T$ удовлетворяет СУР $\vec{p}^T Q_2 = \vec{0}^T$ с матрицей Q_2 вида (3) тогда и только тогда, когда для некоторых векторов $\vec{f}$ и $\vec{g}$ справедливы равенства:

$$\vec{p}_k^T = \vec{f}^T R^{k-n} + \vec{g}^T S^{n+r-k}, \quad k = \overline{n, n+r},$$

$$\vec{p}_{n-2}^T \Lambda + \vec{p}_{n-1}^T N_{n-1} + (\vec{f}^T + \vec{g}^T S^r) M_n = \vec{0}^T,$$

$$\begin{aligned} \vec{p}_{n-1}^T \Lambda + \vec{f}^T (N_n + RM_n) + \vec{g}^T S^{r-1} (SN_n + M_n) + \\ + \vec{f}^T \left(\sum_{j=n}^{n+r} R^{j-1} \right) \tilde{M} + \vec{g}^T \left(\sum_{j=n}^{n+r} S^{r+1-j} \right) \tilde{M} = \vec{0}^T, \end{aligned}$$

$$\vec{f}^T R^{r-1} (RN_{n+r} + \Lambda) + \vec{g}^T (N_{n+r} + S\Lambda) = \vec{0}^T.$$

В следующих разделах рассмотрим работу предложенного метода поиска решения на примере двух экспоненциальных многолинейных систем с обновлением: системы без дообслуживания и системы с дообслуживанием, которые будем кодировать, соответственно, как $M/M^R/n/r$ и $M/M^{RR}/n/r$. Для этих систем будем обозначать через λ интенсивность входящего потока, а через μ — интенсивность обслуживания.

2. Система $M/M^R/n/r$ с обновлением без дообслуживания

Система уравнений равновесия для СМО и условие нормировки имеют вид:

$$\begin{cases} -\lambda p_0 + \mu p_1 = 0, \\ \lambda p_{k-1} - (\lambda + k\mu)p_k + p(k+1)\mu p_{k+1} = 0, & k = \overline{1, n-2}, \\ \lambda p_{n-2} + (\lambda + (n-1)\mu)p_{n-1} + pn\mu p_n + qn\mu \sum_{j=n+1}^{n+r} p_j = 0, \\ \lambda p_{k-1} - (\lambda + n\mu)p_k + pn\mu p_{k+1} = 0, & k = \overline{n, n+r-1}, \\ \lambda p_{n+r-1} + \mu p_{n+r} = 0, \end{cases} \quad (27)$$

$$\sum_{k=0}^{n+r} p_k = 1.$$

Согласно следствию 1, решение такой СУР имеет вид

$$p_k = f R^{k-n+1} + g S^{n+r-k}, \quad k = \overline{n-1, n+r}, \quad (28)$$

где f и g — постоянные, определяемые из соотношений

$$\begin{aligned} \lambda p_{n-2} + f[-\lambda + (n-1)\mu + pn\mu R] + gS^r[-(\lambda + (n-1))S + pn\mu] + \\ + fnq\mu \left(\sum_{j=n}^{n+r} R^{j-n+1} \right) + gnq\mu \left(\sum_{j=n}^{n+r} S^{n+r-j} \right) = 0, \end{aligned} \quad (29)$$

$$fR^r(\lambda - n\mu R) + g(\lambda S - \mu) = 0, \quad (30)$$

а S и R — корни квадратных уравнений $\lambda - R(\lambda + \mu) + R^2p\mu = 0$ и $S^2\lambda - S(\lambda + \mu) + p\mu = 0$, удовлетворяющие условию $S\lambda - (\lambda + \mu) + R\mu \neq 0$.

$$R_{1,2} = \frac{\lambda + \mu \pm \sqrt{(\lambda + \mu)^2 - 4p\lambda\mu}}{2p\mu}, \quad S_{1,2} = \frac{\lambda + \mu \mp \sqrt{(\lambda + \mu)^2 - 4p\lambda\mu}}{2\lambda}.$$

Отметим, что с учетом условия $S\lambda - (\lambda + \mu) + R\mu \neq 0$ для дальнейших расчетов можно взять любую из двух пар корней: (S_1, R_1) или (S_2, R_2) . Естественно, использование любой из этих пар приводит к одному и тому же конечному результату.

Из первого и второго уравнений СУР (27) имеем:

$$p_i = \frac{\rho^i}{i!} p_0, \quad i = \overline{1, n-1}, \quad (31)$$

где $\rho = \lambda/\mu$. Из (29)–(31) после преобразований получаем выражения:

$$g = fR^r \frac{\lambda - n\mu R}{n\mu - \lambda S}, \quad g = fR^{r+2} \frac{1 - S}{1 - R}.$$

Подставляя любое из них в уравнение (28) и используя условие нормировки, получим формулы для f и g :

$$f = \frac{\rho^{n-1}}{(n-1)!} \cdot \frac{1 - R}{1 - R + (RS)^{r+1}(R - RS)} p_0,$$

$$g = \frac{\rho^{n-1}}{(n-1)!} \cdot \frac{R^{r+2}(1 - S)}{1 - R + (RS)^{r+1}(R - RS)} p_0,$$

где

$$p_0 = \frac{1}{\sum_{i=0}^{n-1} \frac{\rho^i}{i!} + \frac{\rho^{n-1}}{(n-1)!} \left(\frac{R(1 - (RS)^{r+1})}{1 - R + (RS)^{r+1}(R - RS)} \right)}.$$

Таким образом, окончательное выражение для стационарных вероятностей имеет вид:

$$\begin{cases} p_0 = \frac{1}{\sum_{i=0}^{n-1} \frac{\rho^i}{i!} + \frac{\rho^{n-1}}{(n-1)!} \left(\frac{R(1 - (RS)^{r+1})}{1 - R + (RS)^{r+1}(R - RS)} \right)}, \\ p_i = \frac{\rho^i}{i!} p_0, \quad i = \overline{1, n-1}, \\ p_i = \frac{\rho^{n-1}}{(n-1)!} \cdot \frac{(1 - R)R^{i-n+1} + (1 - S)R^{r+2}S^{n+r-i}}{1 - R + (RS)^{r+1}(R - RS)} p_0, \quad i = \overline{n, n+r}. \end{cases} \quad (32)$$

Средние числа заявок в системе и очереди определяются формулами:

$$N = \frac{\lambda(1 - p_{n+r}) + n\mu(nq - 1) - \mu(1 - qn) \left(\sum_{i=0}^{n-1} i \frac{\rho^i}{i!} + n \sum_{i=0}^{n-1} \frac{\rho^i}{i!} \right) p_0}{nq\mu},$$

$$Q = N - n + \left(n \sum_{i=0}^{n-1} \frac{\rho^i}{i!} - \sum_{i=0}^{n-1} i \frac{\rho^i}{i!} \right) p_0.$$

Стационарная вероятность $P_{w=0}$ немедленного обслуживания имеет вид

$$P_{w=0} = p_0 \sum_{i=0}^{n-1} \frac{\rho^i}{i!}.$$

В частности, для однолинейной СМО $M/M^R/1/r$ с обновлением ($n=1$) получаем следующие формулы для стационарного распределения числа заявок в системе и средних чисел заявок в системе и в очереди:

$$p_k = \frac{1}{1 - (RS)^{r+2}} \left((1-R)R^k + R^{r+2}(1-S)S^{r+1-k} \right), \quad k = \overline{0, r+1},$$

$$N = \frac{\lambda(1-p_{r+1}) - \mu p(1-p_0)}{\mu q}, \quad Q = \frac{\lambda(1-p_{r+1}) - \mu(1-p_0)}{\mu q}.$$

3. Система $M/M^{RR}/n/r$ с обновлением и дообслуживанием

Система уравнений равновесия для СМО и условие нормировки имеют вид:

$$\begin{cases} 0 = -\lambda p_0 + \mu p_1, \\ 0 = \lambda p_{k-1} - (\lambda + k\mu)p_k + p(k+1)\mu p_{k+1}, \quad k = \overline{1, n-1}, \\ 0 = \lambda p_{n-1} + (\lambda + n\mu)p_n + qn\mu p_n + pn\mu p_{n+1} + qn\mu \sum_{j=n+1}^{n+r} p_j, \\ 0 = \lambda p_{k-1} - (\lambda + n\mu)p_k + pn\mu p_{k+1}, \quad k = \overline{n+1, n+r-1}, \\ 0 = \lambda p_{n+r-1} + \mu p_{n+r}, \end{cases} \quad (33)$$

$$\sum_{k=0}^{n+r} p_k = 1.$$

Согласно следствию 2 решение такой СУР имеет вид

$$p_k = fR^{k-n} + gS^{r+n-k}, \quad k = \overline{n, n+r}, \quad (34)$$

где $\rho = \lambda/\mu$, а f и g — постоянные, определяемые из соотношений

$$\begin{aligned} \lambda p_{n-1} - (\lambda + n\mu) (f + gS^r) + pn\mu (fR + gS^{r-1}) + \\ + fnq\mu \left(\sum_{j=n+1}^{n+r} R^{j-n} \right) + gnq\mu \left(\sum_{j=n+1}^{n+r} S^{n+r-j} \right) = 0, \end{aligned} \quad (35)$$

$$fR^{r-1}(\lambda - n\mu R) + g(\lambda S - n\mu) = 0, \quad (36)$$

а S и R — корни квадратных уравнений $\lambda - R(\lambda + \mu) + R^2\mu = 0$ и $S^2\lambda - S(\lambda + \mu) + \mu = 0$, удовлетворяющие условию $S\lambda - (\lambda + \mu) + R\mu \neq 0$.

$$R_{1,2} = \frac{\lambda + \mu \pm \sqrt{(\lambda + \mu)^2 - 4p\lambda\mu}}{2p\mu}, \quad S_{1,2} = \frac{\lambda + \mu \mp \sqrt{(\lambda + \mu)^2 - 4p\lambda\mu}}{2\lambda}.$$

Как и в предыдущем примере, отметим, что с учетом условия $S\lambda - (\lambda + \mu) + R\mu \neq 0$ для дальнейших расчетов можно взять любую из двух пар корней: (S_1, R_1) или (S_2, R_2) . Естественно, использование любой из этих

пар приводит к одному и тому же конечному результату. Из первого и второго уравнений СУР (33) имеем:

$$p_i = \frac{\rho^i}{p^i i!} p_0, \quad i = \overline{1, n}. \quad (37)$$

Из (37)–(36) после преобразований получаем выражения:

$$g = f R^{r-1} \frac{\lambda - n\mu R}{n\mu - \lambda S}, \quad g = f R^{r+1} \frac{1 - S}{1 - R}.$$

Подставляя любое из них в уравнение (34) и используя условие нормировки, получим формулы для f и g .

$$f = \frac{\rho^n}{p^n n!} \cdot \frac{1 - R}{1 - R + (RS)^r (R - RS)} p_0, \quad g = \frac{\rho^n}{p^n n!} \cdot \frac{R^{r+2} (1 - S)}{1 - R + (RS)^r (R - RS)} p_0,$$

где

$$p_0 = \frac{1}{\sum_{i=0}^n \frac{\rho^i}{p^i i!} + \frac{\rho^n}{p^n n!} \left(\frac{R(1 - (RS)^r)}{1 - R + (RS)^r (R - RS)} \right)}.$$

Таким образом, окончательное выражение для стационарных вероятностей имеет вид:

$$\begin{cases} p_0 = \frac{1}{\sum_{i=0}^n \frac{\rho^i}{p^i i!} + \frac{\rho^n}{p^n n!} \left(\frac{R(1 - (RS)^r)}{1 - R + (RS)^r (R - RS)} \right)}, \\ p_i = \frac{\rho^i}{p^i i!} p_0, \quad i = \overline{1, n}, \\ p_i = \frac{\rho^n}{p^n n!} \cdot \frac{(1 - R)R^{i-n} + (1 - S)R^{r+1}S^{n+r-i}}{1 - R + (RS)^r (R - RS)} p_0, \quad i = \overline{n+1, n+r}. \end{cases} \quad (38)$$

Средние числа заявок в системе и очереди определяются формулами:

$$N = \frac{\lambda(1 - p_{n+r}) - n\mu(p - qn) + \mu(qn - p) \left(\sum_{i=1}^n i \frac{\rho^i}{p^i i!} - n \sum_{i=0}^{n-1} \frac{\rho^i}{p^i i!} \right) p_0}{nq\mu},$$

$$Q = N - n + \left(n \sum_{i=0}^{n-1} \frac{\rho^i}{p^i i!} - \sum_{i=0}^{n-1} i \frac{\rho^i}{p^i i!} \right) p_0.$$

Стационарная вероятность $P_{w=0}$ немедленного обслуживания имеет вид:

$$P_{w=0} = p_0 \sum_{i=0}^{n-1} \frac{\rho^i}{p^i i!}.$$

В частности, для однолинейной СМО $M/M^{RR}/1/r$ с обновлением ($n = 1$) получаем следующие формулы для стационарного распределения числа заявок в системе и средних чисел заявок в системе и в очереди:

$$\begin{cases} p_0 = \frac{p\mu}{\lambda} \cdot \frac{(1 - R) + S^r R^{r+1} (1 - S)}{\frac{p\mu}{\lambda} (1 - R) + \frac{p\mu}{\lambda} S^r R^{r+1} (1 - S) + 1 - (RS)^{r+1}}, \\ p_k = \frac{R^{k-1} (1 - R) + R^{r+1} (1 - S) S^{r+1-k}}{\frac{p\mu}{\lambda} (1 - R) + \frac{p\mu}{\lambda} S^r R^{r+1} (1 - S) + 1 - (RS)^{r+1}}, \quad k = \overline{1, r+1}, \\ N = \frac{\lambda(1 - p_{r+1}) - \mu(p - q)(1 - p_0)}{\mu q}, \quad Q = \frac{\lambda(1 - p_{r+1}) - \mu p(1 - p_0)}{\mu q}. \end{cases} \quad (39)$$

Заключение

В работе рассмотрены марковские системы массового обслуживания с обновлением. Для этих систем на основе доказанной в работе для обобщенного процесса размножения и гибели теоремы получены алгоритмы нахождения стационарных вероятностей состояний. В качестве примеров приведены алгоритмы расчетов для экспоненциальных системы с обновлением — системы $M/M/n/r$ без дообслуживания и системы $M/M/n/r$ с дообслуживанием.

В дальнейшем планируется изучение систем массового обслуживания более общего вида с обновлением, а также рассмотрение случая, когда заявки, выбитые из очереди, формируют отдельный накопитель, из которого повторно могут пойти на обслуживание.

Литература

1. Kreinin A. Queueing Systems with Renovation // Journal of Applied Math. Stochast. Analysis. — Vol. 10, No 4. — 1997. — Pp. 431–443.
2. Kreinin A. Inhomogeneous Random Walks: Applications in Queueing and Finance // CanQueue / Fields Institute. — Toronto: 2003.
3. Towsley D., Tripathi S. K. A single server priority queue with server failure and queue flushing. — 1999.
4. Бочаров П. П., Печинкин А. В. Теория массового обслуживания. — М.: РУДН, 1995. — С. 529.
5. Naoumov V. A. Matrix-multiplicative approach to quasi- birth-and-death processes analysis // Proc. First Internat. Conf. on Matrix-Analytic Methods in Stochastic Models. — Detroit: 1995.

UDC 519.872

Queueing Systems with Renovation. Stationary Probability Distribution

P. P. Bocharov, I. S. Zaryadov

*Department of Probability Theory and Mathematical Statistics
Peoples' Friendship University of Russia
6, Miklukho-Maklaya str., Moscow, 117198, Russia*

The Marcovian queueing systems with renovation are revised. The matrix algorithm for computing the stationary distribution of the Markov process is described. As examples of the method the algorithms for exponential systems are given — queueing systems $M/M/n/r$ without reservice and $M/M/n/r$ with reservice.

ИНФОРМАТИКА

УДК 004.724:004.057.4

Подсистема маршрутизации Click

Д. С. Кулябов, А. В. Королькова, А. А. Хохлов

*Кафедра систем телекоммуникаций
Российский университет дружбы народов
Россия, 117198, Москва, ул. Миклухо-Маклая, 6*

В статье описывается программный комплекс Click Modular Router, разработанный в Массачусетском технологическом университете для построения быстрых, гибких и легко управляемых программных маршрутизаторов с поддержкой технологии QoS.

КЛЮЧЕВЫЕ СЛОВА: routing, QoS, router, Click Modular Router, Click, Linux, Device Polling.

Введение

С момента разработки стека протоколов TCP/IP неотъемлемой частью любой компьютерной сети стали маршрутизаторы, пересылающие пакеты между хостами. Наряду со специализированными компьютерами-маршрутизаторами, которые занимались только маршрутизацией, появилось много модулей для различных операционных систем, которые обеспечивали функциональность обычного компьютера в качестве маршрутизатора.

Разработчики каждой из сетевых операционных систем старались внедрять в свои продукты новые технологии, чтобы их ОС «шли в ногу со временем». Изначально, как в BSD UNIX, так и в Linux системах, вся статическая маршрутизация была представлена лишь простым перенаправлением пакетов между интерфейсами согласно таблице, а динамическая — сторонними программами gated и routed. Через небольшой промежуток времени требования к маршрутизаторам возросли, от них стало требоваться не только уметь пересылать трафик через свои интерфейсы, но и делать множество других вещей — исполнять функции брэндмауэров (firewall) практически на всех уровнях TCP/IP, кроме аппаратного, подсчитывать трафик, составлять статистику, выступать в роли прокси-серверов, ограничивать трафик от одних приложений либо адресов, и, наоборот, выделять больший приоритет трафику от других, осуществлять трансляцию сетевых адресов и их маскардинг. Так, маршрутизаторы стали не только перенаправлять трафик, а еще и изменять его, маркировать, классифицировать, сортировать, отбрасывать ненужный трафик — то есть управлять трафиком и обрабатывать его.

Естественно, что в развитых сетевых операционных системах — Linux и BSD Unix (например — FreeBSD) стала активно внедряться поддержка всех этих технологий.

Разработчики ядра Linux, начиная с ядер 2.2, ввели в систему программный пакет ipchains (цепочки IP), который впоследствии заменили на пакет iptables (таблицы IP) и систему IPRoute версий 1 и 2, работающие с модулем ядра netfilter и другими модулями. IPRoute 2 совместно с iptables (это оболочка к системе фильтрации трафика ядра) явили собой воистину выдающуюся связку, обладающую практически неограниченными возможностями в обработке трафика. Вот лишь некоторые возможности Linux в области обработки трафика [1]: управление полосой пропускания, разделение и приорегизация трафика, поддержка большого количества механизмов QoS, фильтрация, NAT, firewalling, туннелирование и много другое.

К сожалению, пакет IPRoute2, непрерывно развиваясь и разрастаясь, становится все более трудно управляемым.

Разработчики FreeBSD решили пойти другим путем [2,3]. В 1996 году программисты компании Whistle InterJet Арчи Коббс (Archie Cobbs) и Джулиан Элишер (Julian Elischer) встали перед задачей реализовать многопротокольный маршрутизатор на базе операционной системы FreeBSD. По некоторым причинам их не устроили уже существующие продукты, и они решили написать свою библиотеку. Ее назвали NetGraph. NetGraph не имеет недостатков, присущих продуктам — предшественникам, потому что ее изначально проектировали как систему, решающую множество задач, состоящую из модулей. Новые модули могут добавлять любые программисты, поэтому система масштабируема. Ее довольно простое устройство позволяет сторонним людям легко в ней разбираться и улучшать по необходимости.

NetGraph — это модульная сетевая подсистема, которая может использоваться для дополнения существующей сетевой инфраструктуры ядра. Многие существующие работающие модули поставляются с FreeBSD и включают поддержку практически всех сетевых стандартов.

NetGraph понятна для понимания, потому что конфигурация маршрутизатора описывается в виде ориентированного графа, по ребрам которого передается трафик. Это позволяет с легкостью изменять и масштабировать конфигурацию.

Недостатком NetGraph явилось опять же малое количество документации и отсутствие поддержки QoS. Опять же, в последнее время системы Linux получают гораздо большее распространение, чем BSD, и большее количество разработок ведется в этом направлении.

Одной из таких разработок стал программный комплекс Click Modular Router.

1. Описание Click

1.1. История возникновения

Click был разработан в Массачусетском технологическом университете США при поддержке национального агентства DARPA. Впервые авторы (Eddie Kohler, Robert Morris, Benjie Chen, John Jannotti, M. Frans Kaashoek) заявили о нем в декабре 1999 года на 17 симпозиуме об архитектуре операционных систем [4]. Click был представлен как гибкая, идеально масштабируемая, быстрая модульная система для построения любых, сколь угодно сложных маршрутизаторов. С тех пор система развивалась, дополнялась новыми модулями (на текущий момент она поддерживает работу с практически любыми сетевыми протоколами и стандартами) и поддерживалась. Обновление ее производится через CVS.

1.2. Идеология Click

В основу идеологии Click положены гибкость, масштабируемость, высокая скорость работы, легкость в управлении и возможность самостоятельно изменять систему.

Гибкость, масштабируемость и легкость в понимании достигаются тем, что конфигурации Click представлены в виде ориентированного графа. Возможность любому программисту изменять и дополнять систему предоставляется в силу того, что Click написан на C++, и в документации есть шаблоны классов, на основе которых можно писать новые компоненты системы.

Высокая скорость обработки трафика достигается за счет механизмов Device Polling и внутреннего механизма аннотаций.

Device Polling [5] — это новая технология работы ядра Linux с устройствами, которая позволяет ускорять обработку пакетов в несколько раз.

Click может заменить всю стандартную маршрутизацию Linux. Эмпирически доказано [4], что скорость его работы больше, хоть и немного, скорости работы стандартного ядра.

Например, на компьютере с процессором Intel Pentium III 700 MHz с двумя сетевыми картами DEC Tulip 100 Mbit/s, операционная система — Linux, ядро —

2.2.4, производительность маршрутизатора, построенного на базе Click Modular Router, по заявлению разработчиков, составила 144000 64-байтных IP пакетов в секунду [4]. Для сравнения, производительность маршрутизатора Cisco 3750 в режиме Fast Switching может достигать лишь до 120000 64-байтных IP пакетов в секунду.

1.3. Архитектура Click

Компоненты системы, которые обрабатывают пакеты, называются **обработчиками**. Основной код обработчика довольно мал, он включает в себя, в основном, только функции инициализации и передачи пакета, однако может быть дополнен дополнительными функциями, например, передачей размера очереди пакетов.

Для построения маршрутизатора пользователь должен выбрать набор обработчиков и соединить их в ориентированный граф. Ребра графа называются **соединениями**. Они представляют собой возможные пути следования пакета. Граф является ориентированным по причине того, что пакеты следуют через маршрутизатор в определенном направлении.

Для повышения функциональности пользователи могут создавать свои собственные обработчики, а также модифицировать должным образом имеющиеся. Обработчики можно комбинировать между собой любым образом при помощи каналов (pipes) операционной системы Unix.

Передача пакета по соединению может быть инициирована как источником (push processing), так и пунктом назначения (pull processing). Это повторяет большинство алгоритмов передачи пакетов маршрутизаторами и позволяет создавать полноценные планировщики передачи пакетов.

Обработчик Click представляет собой некий элемент, в котором происходит обработка пакета.

Каждое действие, произведенное программным комплексом Click Router, инкапсулировано в обработчик, начиная от аппаратной обработки и заканчивая подсчетом количества пакетов.

Внутри рабочей конфигурации каждый обработчик представлен объектом C++, соединения реализованы при помощи указателей на объекты, а прохождение пакета представлено вызовом виртуальных функций C++.

Ниже представлены основные свойства обработчика:

- **класс обработчика**: принадлежность к определенному классу определяет код, который работает, когда на обработчик попадает пакет;
- **порты**: каждый элемент может иметь определенное количество входных и выходных портов. Соединения между обработчиками создаются «между их портами» - между исходящим портом одного обработчика и входящим другого. Два соединения между одними и теми же обработчиками, но разными их портами могут иметь абсолютно разную семантику;
- **строка инициализации**: определяет состояние обработчика при первоначальном запуске программного комплекса;
- **интерфейсы**: каждый обработчик должен иметь как минимум интерфейс для приема и передачи пакетов, однако помимо у обработчика может существовать некоторое количество дополнительных интерфейсов, при помощи которых обработчики обмениваются информацией, полученной в результате исполнения их методов.

На рис. 1 представлен простейший обработчик **Tee(2)**, принимающий пакеты на единственный входящий порт и копирующий их на два исходящих. Tee — это класс обработчика. Здесь и далее треугольником обозначается входящий порт, а прямоугольником — исходящий. Простейший обработчик представлен на рис. 1.

Цифра 2 в строке инициализации означает запрос на два исходящих порта. В данном примере интерфейсов методов не имеется.

На рис. 2 показана конфигурация, в которой несколько элементов объединены в простейший маршрутизатор, который считает входящие пакеты и отбрасывает их.

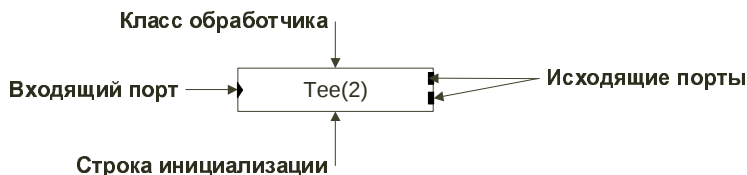


Рис. 1. Простейший обработчик Click.



Рис. 2. Простейший маршрутизатор.

1.4. Push и pull соединения

Теперь рассмотрим два механизма, которые поддерживает Click — передачу пакета по требованию источника (push processing) и по требованию адресата (pull processing). Push-соединения, при которых пакет посылается по инициативе источника, похожи на методы, которые используют все стандартные маршрутизаторы. При Pull-соединениях, наоборот, обработчик, которому пакет предназначен, посылает запрос на источник с тем, чтобы тот вернул ему пакет или нулевой указатель в том случае, если пакета нет. Оба этих способа реализованы при помощи виртуальных функций C++.

Следует заметить, что любой порт обработчика может быть либо push, либо pull типа. Типы портов определяются заранее, при инициализации маршрутизатора. Между портами разных типов соединение не может быть установлено. Существует специальный тип портов, которые, в зависимости от типа порта, с которым планируется взаимодействие, могут принимать разные типы.

Push-соединения используются, например, при поступлении пакетов на сетевую карту. После поступления каждого пакета он передается далее. Pull-соединения можно использовать, например, в том случае, когда для передачи пакета в сеть надо подождать освобождения передающего устройства, которое, как только освободится, инициирует Pull-соединение и передаст пакет.

1.5. Очереди пакетов

Обработчики Click не имеют встроенных буферов для построения очередей необработанных пакетов, однако такие очереди присутствуют в системе в явном виде. Они представлены специальными Queue-обработчиками, которые имеют push — входящий порт и pull — исходящий. Работают по алгоритму FIFO. Строкой инициализации является вместимость обработчика-очереди.

1.6. Механизм поиска обработчиков на основе потоков пакетов

Если обработчику А потребуется использовать обработчик В, он должен сначала найти В. Эта проблема не решается соединениями, так как между А и В может быть не установлено никаких соединений. Вопрос решается следующим образом: либо А ищет В по имени (например, ссылка на В имеется в его строке инициализации), либо использует автоматический механизм, названный механизмом поиска обработчиков на основе потоков пакетов, который на основе аннотаций пакетов определяет местоположение обработчиков в конфигурации.

Как раз этот механизм является одной из причин, по которой Click работает быстрее, чем IPRoute 2.

1.7. Язык конфигурации

Конфигурации системы Click Modular Router пишутся на простом языке, основами которого являются **декларации** (объявления) элементов, которые создают элементы, и **соединения**, которые описывают, каким образом обработчики должны быть соединены. Разработчики называют его C – подобным.

Этот язык является декларативным языком — то есть он описывает, какие обработчики надо создать и как их соединить, но не описывает, каким образом они работают. Это делает конфигурации Click понятными и легко изменяемыми. Простой пример:

```
frometh0 :: FromDevice(eth0); toeth1 :: ToDevice(eth1); frometh0 -> toeth1;
```

Как видно, обработчики объявляются при помощи оператора «::», а соединяются — при помощи «->». Этот язык очень легкий для усвоения, достаточно часа, чтобы научиться с ним работать в полной мере.

1.8. Способы взаимодействия с операционной системой

Click может работать либо на уровне ядра, для чего ядро придется перекомпилировать, наложив перед этим на его исходные коды специальный патч, либо на пользовательском уровне и взаимодействовать с сетью либо при помощи какого-либо пакетного фильтра, либо при помощи какого-нибудь другого механизма. Работа на пользовательском уровне больше подходит для настройки, написания и отладки конфигураций, в то время как работа на уровне ядра больше подходит для реальных задач, в силу того, что на уровне ядра скорость работы гораздо больше.

Для работы на уровне ядра существуют механизмы для «горячей» замены конфигураций «на лету».

1.9. Реализация обработчиков

Каждый класс обработчика Click соответствует подклассу класса *Element* C++, который имеет примерно 20 виртуальных функций. *Element* включает в себя базовые реализации большинства из них, так что подклассам требуется перепределить 6 или менее виртуальных функций. Всего три функции используются в процессе работы маршрутизатора — *push*, *pull* и *run_sheduled*. Все остальные нужны для идентификации, инициализации, статистики и др.

2. IP маршрутизатор

В этом разделе представлена схематическая модель простейшего маршрутизатора, который может перенаправлять пакеты между двумя сетевыми интерфейсами. Работает он следующим образом.

Обработчик *FromDevice* со строкой инициализации «*eth0*» принимает пакет от интерфейса *eth0* и помещает его в обработчик *Classifier*, который разделяет трафик на ARP-запросы, ARP-ответы и IP-трафик.

ARP-запросы будут передаваться обработчику *ARPResponder*, который будет отвечать на них. Строка инициализации у этого обработчика будет следующая: *ARPResponder(IP_адрес_интерфейса_eth0, MAC_адрес_интерфейса_eth0)*. ARP-ответы, которые приходят на интерфейс, очевидно, были получены в результате неких запросов. Их создавал обработчик *ARPQuerier*, ему они и направляются.

IP-пакеты, отфильтрованные обработчиком *Classifier*, направляются обработчику *Paint(1)*, который «метит» пакет параметром, заданным в строке инициализации. Это может быть целое число от 1 до 255. Оно означает сетевой интерфейс, откуда пришел пакет (для второго интерфейса будет вызываться обработчик *Paint(2)*). Теперь, при помощи механизма аннотаций, любой обработчик будет знать, с какого интерфейса пришел пакет. Далее пакет поступает обработчику *Strip*, который «отрезает» 14 байт от начала пакета (это указано в его строке

инициализации «14»), эти 14 байт являются ненужным для IP маршрутизации Ethernet-заголовком.

Далее IP-пакет поступает на обработчик *CheckIPHeader*, который проверяет пакет на правильность IP-адреса получателя, источника, вычисляет контрольную сумму пакета и сравнивает ее с той, что имеется в заголовке. Если все проверки проходят успешно, пакет передается на выходной интерфейс неизменным, если же что-то не так, пакет отбрасывается.

Далее пакет попадает на обработчик *GetIPAddress*, который копирует IP-адрес назначения из заголовка пакета в его аннотацию. Перенаправляет пакет неизменным. Потом пакет попадает в обработчик *LookupIPRoute*, который, имеет в качестве строки инициализации маршрутную таблицу. Этот обработчик начинает собой ветвление дерева, посылая пакет через ряд других обработчиков на первый или второй интерфейс, либо в TCP/IP стек Linux.

Рассмотрим одну из ветвей. После обработчика *LookupIPRoute* пакет попадает в обработчик *DropBroadcasts*, который на основании аннотации-флага, указывающего на широковещательные сообщения для физического соединения, с которого пришел пакет, установленного обработчиком *FromDevice*, отбрасывает эти широковещательные пакеты. Далее обработчиком *CheckPaint* проверяется аннотация, в которой указано, с какого интерфейса пришел пакет. Если оказывается так, что пакет идет в тот же интерфейс, откуда он пришел, то это не совсем корректная ситуация, генерируется ICMP ошибка.

Далее пакет поступает на обработчик *IPGWOptions*, который обрабатывает параметры Record Route и Timestamp. Если blackRecordblack blackRouteblack включен, маршрутизатор «вписывает» себя в список пройденных маршрутизаторов. Затем меняется параметр blackTimestampblack, который нужен для измерения времени передачи пакета от одного маршрутизатора к другому. В случае каких-либо ошибок генерируются ICMP-сообщения. Далее пакет передается обработчику *blackFixIpSrcblack* со строкой инициализации, которая представляет собой blackIPblack-адрес интерфейса, через который отправится пакет.

Этот обработчик, при условии, что аннотация-флаг Fix Ip Source включена, исправит IP-адрес источника на IP-адрес этого интерфейса. Этот флаг устанавливает обработчик *ICMPError*. Дело в том, что ICMP-ошибки должны приходить именно от этого маршрутизатора, а обработчик *ICMPError* не может предугадать, через какой интерфейс уйдет пакет в сеть, поэтому он просто включает аннотацию-флаг Fix Ip Source, а уже после того, как обработчик *LookupIpRoute* определит интерфейс, через который этот пакет уйдет, FixIpSrc исправит IP-адрес источника на IP-адрес этого интерфейса. Далее пакет передается обработчику *DecIPTTL*, который убавляет на 1 поле TTL заголовка пакета, и, если полученное число не стало равным 0, отправляет пакет дальше, а если TTL истекло, передает пакет обработчику *ICMPError* со строкой инициализации «11», что означает ICMP-ошибку «TTL Expired», который генерирует аналогичную ошибку. Пакеты, которые успешно проходят уменьшение TTL, попадают на обработчик *IPFragmenter* со строкой инициализации «1500». 1500 — это стандартный MTU для IP-сетей. Пакеты, меньшие, чем 1500 байт, обработчик пропускает беспрепятственно сразу. Те, которые больше — он делит на фрагменты, меньшие 1500 байт, формирует из этих фрагментов новые IP-пакеты и передает их на выходной интерфейс. Однако, если входной пакет больше, чем 1500 байт, а в заголовке установлен флаг don't-fragment, обработчик через второй выходной интерфейс посылает пакет обработчику *ICMPError* со строкой инициализации «3,4», что означает ICMP-сообщение «необходима фрагментация», который, в свою очередь, генерирует ICMP-пакет с аналогичной ошибкой.

После обработчика *IPFragmenter* мы получаем готовые к отправке по канальному уровню пакеты. Они поступают в обработчик *ARPQuerier*, строкой инициализации которого является IP-адрес интерфейса. Этот обработчик, во-первых, принимает ARP-ответы из обработчика *Classifier*, рассмотренного выше. Во-вторых, он инкапсулирует полученные IP-пакеты в Ethernet-кадры и отправляет их в очередь на отправку на сетевой интерфейс. Так же этот обработчик посылает в сеть ARP-запросы для распознавания MAC — IP необходимых адресов. Как было сказано выше, обработчик *ARPQuerier* при помощи push-соединения

«кладет» пакеты в очередь *Queue* со строкой инициализации, являющейся целым числом, которое является просто размером этой очереди. В данном примере очередь представляет собой структуру FCFS статического размера. Из очереди при помощи pull-соединения кадры «вытаскивает» обработчик ToDevice со строкой инициализации, являющейся именем сетевого интерфейса в системе, и отправляет их в сеть по мере ее освобождения. Этот обработчик работает на уровне протоколов MAC/LLC. Следует заметить, выше было рассмотрено 4 случая, когда пакет уходил в *ICMPError* обработчики, которые, в зависимости от ситуации, генерировали разные пакеты с ICMP-ошибками. Все эти пакеты передаются на вход обработчику *LookupIPRoute*, и он их направляет в нужные интерфейсы. Ниже представлен рисунок, который наглядно демонстрирует маршрутизатор, представляя его в виде связного графа, узлами которого являются обработчики, а ребрами — соединения. Схематическое изображение этого маршрутизатора представлено на рис. 3.

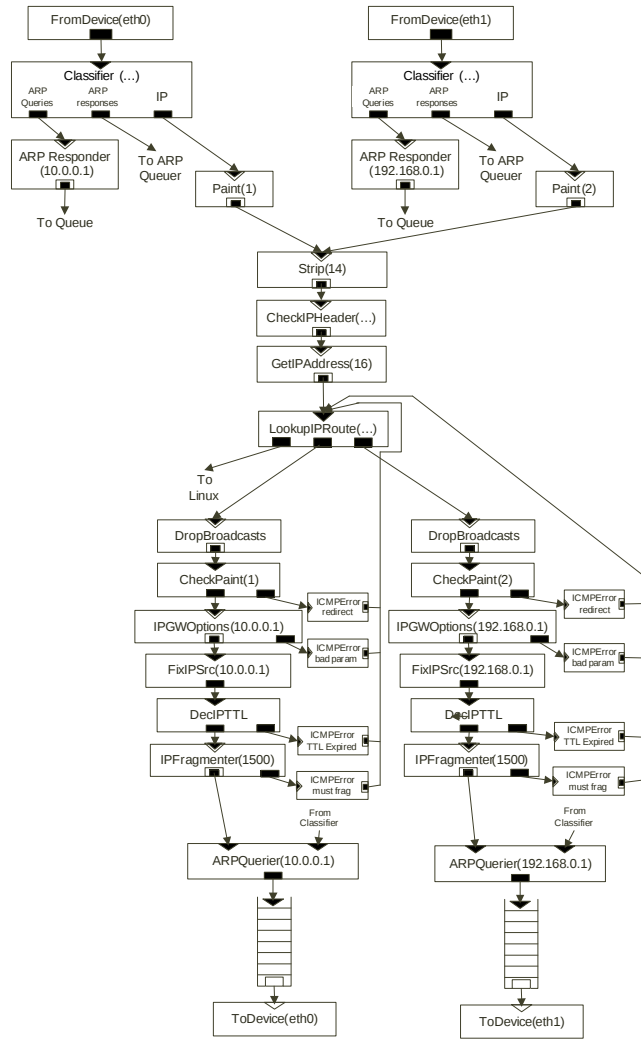


Рис. 3. Конфигурация простого IP — маршрутизатора.

Заключение

В статье рассмотрен программный комплекс, позволяющий реализовывать программные маршрутизаторы на базе Linux систем. Click уже с первого взгляда привлекает своей простотой, гибкостью и логичностью по сравнению с другими программными маршрутизаторами.

Заявленная разработчиками и измеренная эмпирически производительность высока и ограничивается пропускной способностью сетевых интерфейсов на аппаратном уровне. Однако первые эксперименты с сетевыми адаптерами на шине PCI Express показали, что на базе Click можно будет строить сверхбыстрые маршрутизаторы.

Глубокое изучение Click выявило, что он может поддерживать так или иначе практически все алгоритмы управления трафиком, которые присущи современному QoS и модели DiffServ в частности.

Все это позволяет сделать вывод, что у системы Click большое будущее в нише программных маршрутизаторов и что уже сейчас ее можно успешно применять в реальных ситуациях.

Также представляется перспективным применение данного пакета для имитационного моделирования, в рамках пакета ns-click. Данный пакет позволяет построить имитационную модель на ns2, используя при этом реальные модули системы Click.

Литература

1. *Горев А.* Advanced Routing и QoS. — http://www.opennet.ru/docs/RUS/adv_route_qos/.
2. *Коббс А.* Все о NetGraph. — 2003. — <http://citrin.ru/netgraph/>.
3. *Беляков С.* 12000 слов о Netgraph и FreeSBD. — 2006. — <http://www.unix.lviv.ua/content/view/33/27/>.
4. *Kohler E., Morris R., Chen B. et al.* The Click Modular Router. — 2000.
5. *Deri L.* Improving Passive Packet Capture: Beyond Device Polling // Proceedings of SANE 2004. — 2004. — <http://luca.ntop.org/>.

UDC 004.724:004.057.4

Routing Subsystem Click

D. S. Kulyabov, A. V. Korolkova, A. A. Khokhlov

*Telecommunication Systems Department
Peoples' Friendship University of Russia
6, Miklukho-Maklaya str., Moscow, 117198, Russia*

The Click Modular Router, a software for building powerful and flexible routers, is described in this article. It was developed in MIT in 1999. A QoS support is also included in this software.

УДК 519.217.2

Метод расчета вероятности сброса пакетов в алгоритме RED

А. В. Королькова

*Кафедра систем телекоммуникаций
Российский университет дружбы народов
Россия, 117198, Москва, ул. Миклухо-Маклая, 6*

В статье исследована модель функционирования алгоритма RED (Random Early Detection). Построена аналитическая модель в виде СМО $GI_{\lambda(\bar{q})}|M|1|R$ с пороговыми значениями для средней длины очереди и интенсивностью входящего потока $\lambda(\bar{q})$, зависящей от средней длины очереди $\bar{q}$.

Ключевые слова: RED, вложенная цепь Маркова, вероятность сброса, средняя длина очереди.

Постановка задачи

Алгоритм RED [1] был создан для сетей пакетной коммутации и хорошо себя зарекомендовал при работе с протоколами, ориентированными на установление соединения, в частности, с протоколом TCP.

Алгоритм включает две процедуры — оценку средней длины очереди и принятие решения о сбросе пакета. Пакеты сбрасываются перед поступлением в очередь, не дожидаясь ее переполнения, что позволяет избежать перегрузок и обрабатывать пульсирующий трафик. Алгоритм инициирует снижение скорости передачи для протокола TCP, поскольку при потере хотя бы одного пакета протокол заново проводит синхронизацию соединения и корректирует размер окна, снижая интенсивность трафика.

На рис. 1 изображена диаграмма функционирования RED. Здесь $\bar{q}$ — средняя длина очереди пакетов на передачу, r_1 — минимальное и r_2 — максимальное пороговые значения средней длины очереди, $\pi_{\max}$ — максимальное значение вероятности $\pi(\bar{q})$ сброса пакета. Заметим, что вероятность сброса пакета в алгоритме RED зависит от значения средней длины очереди. Если значение средней длины очереди меньше минимального порога, то все приходящие пакеты принимаются. Если значение средней длины очереди находится между минимальным и максимальным порогами, то вероятность сброса входящего пакета вычисляется по формуле (1) и, следовательно, линейно возрастает от 0 до значения $\pi_{\max}$. Когда значение средней длины очереди становится больше максимального порога, все поступающие пакеты сбрасываются. На практике в действующем на сети оборудовании значение средней длины очереди $\bar{q}$ определяется статистически в режиме реального времени, и по формуле (1), реализованной в алгоритме RED, рассчитывается вероятность $\pi(\bar{q})$.

Известен ряд работ (см., например, [2–4]), где исследованы характеристики алгоритма RED. Результаты этих исследований получены путём измерений либо с помощью имитационного моделирования. В данной статье для анализа и расчёта вероятностных характеристик алгоритма RED использована модель в виде системы массового обслуживания (СМО) типа $GI|M|1|R$ [5]. Предварительные результаты исследования для СМО $M|M|1|R$ и опубликованы в [6].

Суть проблемы состоит в том, что интенсивность поступающего на СМО потока зависит от значения $\bar{q}$ средней длины очереди, которую необходимо определить



Рис. 1. Диаграмма функционирования RED.

в моменты времени, непосредственно предшествующие поступлению заявок. Кроме того, величины $\lambda(\bar{q})$ и $\pi(\bar{q})$ связаны соотношениями (1)–(2), которые представляют собой систему трансцендентных уравнений «неподвижной точки», решение которой предлагается искать методом последовательных приближений.

$$\pi(\bar{q}) = \begin{cases} 0, & 0 \leq \bar{q} < r_1, \\ \frac{\bar{q} - r_1}{r_2 - r_1} \pi_{\max}, & r_1 \leq \bar{q} \leq r_2, \\ 1, & \bar{q} > r_2, \end{cases} \tag{1}$$

$$\lambda(\bar{q}) = \begin{cases} \lambda, & \bar{q} < r_1, \\ (1 - \pi(\bar{q}))\lambda, & r_1 \leq \bar{q} \leq r_2, \\ 0, & \bar{q} > r_2. \end{cases} \tag{2}$$

1. Метод расчета вероятности сброса

Обозначим через $\nu(t)$ число заявок в СМО в момент времени $t \geq 0$. Пусть τ_n , — момент поступления n -й заявки в СМО, $\nu_n^- = \nu(\tau_n - 0)$ — число заявок в СМО непосредственно перед поступлением n -й заявки. Последовательность $\{\nu_n^-, n \geq 0\}$ является цепью Маркова (ЦМ), для которой справедливо [5] следующее соотношение

$$\nu_{n+1}^- = \begin{cases} \min(\nu_n^- + 1 - \alpha_{n+1}, r_2 + 1), & \nu_n^- + 1 - \alpha_{n+1} > 0, \\ 0, & \nu_n^- + 1 - \alpha_{n+1} \leq 0. \end{cases} \tag{3}$$

Система уравнений равновесия для стационарных вероятностей q_i ЦМ $\{\nu_n^-, n \geq 0\}$ имеет вид:

$$\begin{cases} q_0 = \sum_{j=0}^{r_2+1} q_j \left(\sum_{m=j+1}^{\infty} \alpha_m \right), \\ q_i = \sum_{j=i-1}^{r_2} q_j \alpha_{j-i+1} + q_{r_2+1} \alpha_{r_2-i+1}, \quad i = \overline{1, r_2+1}, \end{cases} \quad (4)$$

где

$$\alpha_k = \int_0^{\infty} \frac{(\mu x)^k}{k!} e^{-\mu x} dA(x), \quad k = \overline{0, r_2}. \quad (5)$$

Пусть $g_i = \frac{q_i}{q_{i-1}}$, $i = \overline{1, r_2+1}$, $g_0 \equiv 1$. Тогда

$$q_i = g_i q_{i-1} = g_i g_{i-1} q_{i-2} = \dots = \prod_{m=1}^i g_m q_0, \quad i = \overline{1, r_2+1}. \quad (6)$$

Из (4) и (6) следует, что

$$\begin{cases} g_{r_2+1} = \frac{\alpha_0}{1 - \alpha_0}, \\ g_{r_2} = \frac{\alpha_0}{1 - \alpha_1 - g_{r_2+1} \alpha_1}, \\ g_{r_2-i} = \frac{\alpha_0}{1 - \alpha_1 - \sum_{k=2}^{i+1} \alpha_k \prod_{m=1}^{k-1} g_{r_2-i+m} - \alpha_{i+1} \prod_{m=1}^{i+1} g_{r_2-i+m}}, \quad i = \overline{1, r_2-1}, \\ g_0 \equiv 1, \end{cases} \quad (7)$$

а из условия нормировки получаем

$$q_0 = \frac{1}{1 + \sum_{i=1}^{r_2+1} \prod_{m=1}^i g_m}. \quad (8)$$

Формулы (6)–(8) определяют метод расчёта распределения q_i , $i = \overline{0, r_2+1}$, зная которые можно вычислить значение средней длины очереди $\bar{q}$ в моменты поступления в СМО заявок:

$$\bar{q} = \sum_{i=0}^{r_2+1} i q_i. \quad (9)$$

Ниже предложен итерационный алгоритм расчёта вероятности $\pi(\bar{q})$ по формулам (1)–(2), где значение $\bar{q}$ вычислено по формуле (9). Исходными данными для вычисления являются значения величин α_k , r_1 , r_2 и $\pi_{\max}$.

Алгоритм расчёта вероятности $\pi(\bar{q})$

-
- 1: $\ell \leftarrow 1, \bar{q}^\ell \leftarrow 0, \lambda(\bar{q}^\ell) \leftarrow \lambda$
 - 2: для $j = 0, 1, \dots, r_2 + 1$ вычислить q_j^ℓ по формулам (6)–(8)
 - 3: вычислить $\bar{q}^\ell, \pi(\bar{q}^\ell), \lambda(\bar{q}^\ell)$ по формулам (9), (1)–(2)
 - 4: если $|\pi(\bar{q}^\ell) - \pi(\bar{q}^{\ell-1})| \leq \varepsilon$, то
 - 5: $\ell \leftarrow \ell + 1$ и перейти на шаг 2
 - 6: иначе вернуть $\ell, \bar{q}^\ell, \pi(\bar{q}^\ell), \lambda(\bar{q}^\ell)$
-

Заметим, что на шаге 2 алгоритма требуются значения величин $\alpha_k, k = \overline{0, r_2}$, которые зависят от вида функции распределения $A(x)$ интервалов между поступлениями заявок в СМО $GI|M|1|R$. В приведённом далее примере численного анализа рассматривается случай системы $M|M|1|R$, в котором нетрудно убедиться, что величины α_k вычисляются рекуррентно по формулам

$$\begin{cases} \alpha_0 = \frac{\rho}{\rho + 1}, \\ \alpha_k = \frac{1}{\rho + 1} \alpha_{k-1}, \quad k = \overline{1, r_2}, \quad \rho = \lambda/\mu. \end{cases}$$

Численный анализ

Проиллюстрируем применение предложенного метода для случая пуассоновского входящего потока (СМО $M|M|1|r_2$). Исходные параметры имели следующие значения: $r_1 = 5, r_2 = 20, \pi_{\max} = 0,4, \varepsilon = 10^{-4}$.

На рис. 2 показано изменение значений вероятности сброса $\pi(\bar{q})$ относительно числа итераций. За конечное число итераций вероятность сброса сходится к единственному значению. Это говорит о том, что при различных параметрах можно прогнозировать оптимальное значение вероятности сброса для реальных систем.

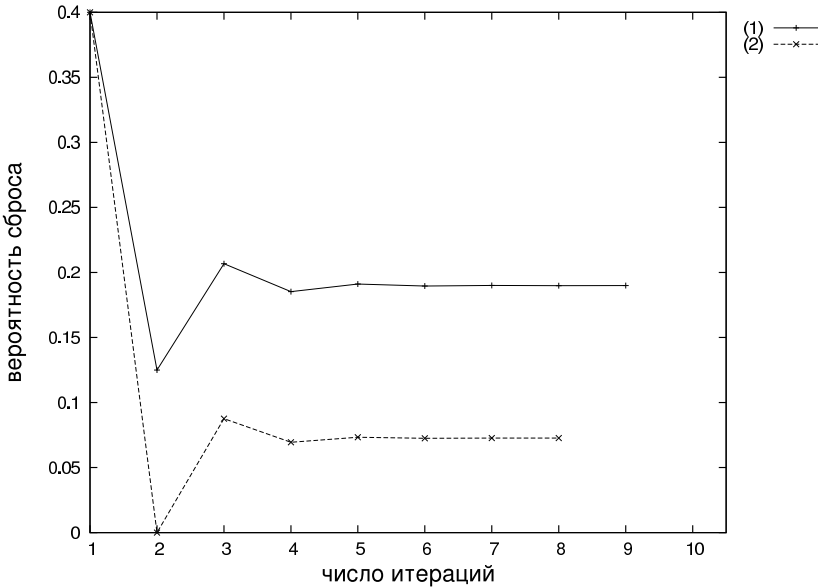


Рис. 2. Изменение значений вероятности сброса $\pi(\bar{q})$ относительно числа итераций: (1) $\lambda = 16, \mu = 4$; (2) $\lambda = 4, \mu = 2$.

На рис. 3 показано изменение оптимальных значений средней длины очереди $\bar{q}$ относительно нагрузки $\rho(\bar{q})$. Значение средней длины очереди $\bar{q}$ даже при высокой нагрузке не превышает максимально возможного значения.

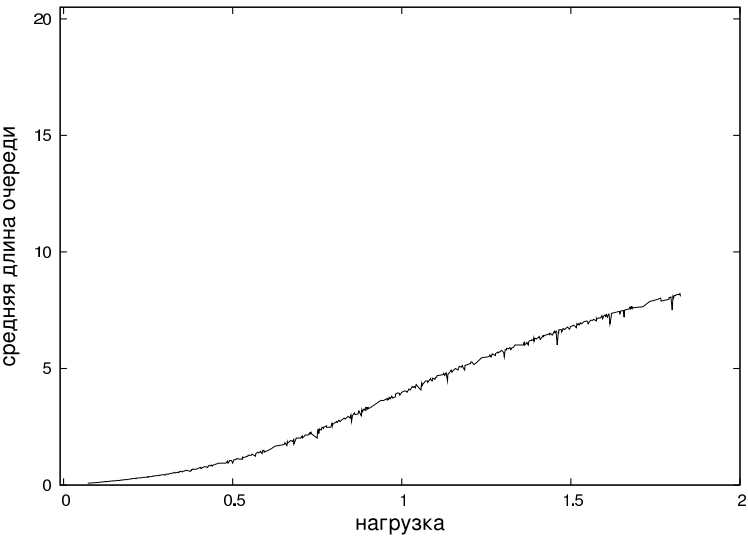


Рис. 3. Изменение оптимальных значений средней длины очереди $\bar{q}$ относительно нагрузки $\rho(\bar{q})$.

На рис. 4 показано изменение оптимальных значений вероятности сброса $\pi(\bar{q})$ относительно средней длины очереди $\bar{q}$. Значения вероятности сброса лежат на теоретической кривой и не превышают заданного максимального значения.

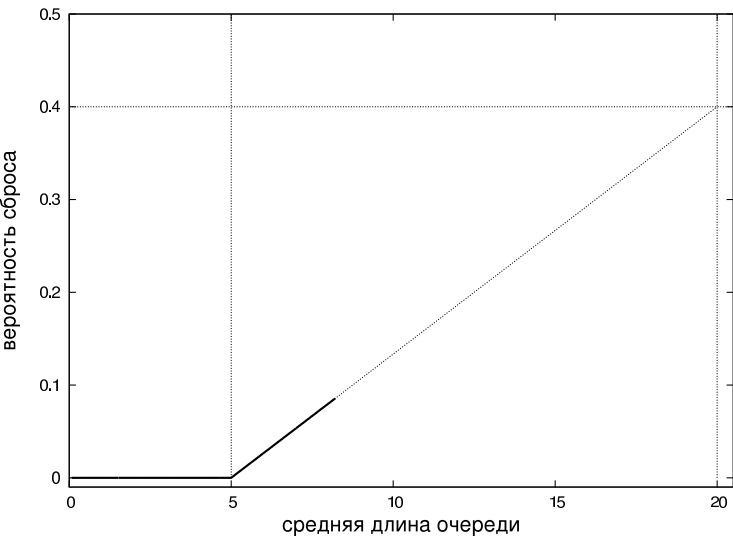


Рис. 4. Изменение оптимальных значений вероятности сброса $\pi(\bar{q})$ относительно средней длины очереди $\bar{q}$.

Заключение

Построена аналитическая модель для расчета вероятности сброса пакетов в алгоритме RED в виде системы $GI_{\lambda(\bar{q})}|M|1|R$ с порогами для значений средней длины очереди и интенсивностью входящего потока $\lambda(\bar{q})$, зависящей от средней

длины очереди $\bar{q}$. На основе аналитической модели разработан итерационный алгоритм, позволяющий за конечное число итераций рассчитать оптимальные значения вероятности сброса пакета при различных параметрах системы. Показано, что при высоком значении нагрузки оптимальное значение средней длины очереди никогда не превышает максимального порога и, следовательно, оптимальное значение вероятности сброса остаётся на низком уровне.

Литература

1. *Floyd S., Jacobson V.* Random Early Detection Gateways for Congestion Avoidance // IEEE/ACM Transactions on Networking. — No 1(4). — 1993. — Pp. 397–413.
2. *Floyd S.* RED: Discussions of Setting Parameters. — 1997. — <http://www.aciri.org/floyd/REDparameters.txt>.
3. Reasons not to deploy RED / M. May, J. Bolot, C. Diot, B. Lyles // Proceedings of IWQoS'99. — 1999. — Pp. 260–262.
4. *May M., Bonald T., Bolot J.-C.* Analytical evaluation of RED performance // Proceedings of IEEE INFOCOM 2000. — 2000.
5. *Бочаров П., Печинкин А. В.* Теория массового обслуживания. — М.: Изд-во РУДН, 1995. — С. 529.
6. *Королькова А. В., Кулябов Д. С.* Аналитическая модель для расчета вероятности сброса пакетов в алгоритме RED // Труды конференции «Телекоммуникационные и вычислительные системы» / МТУСИ. — М.: 2006. — С. 24.

UDC 519.217.2

The Methods of Drop Probability Calculation for RED Algorithm

A. V. Korolkova

*Telecommunication Systems Department
Peoples' Friendship University of Russia
6, Miklukho-Maklaya str., Moscow, 117198, Russia*

The analytic model of the functioning of RED algorithm is examined in the article. The model is based on the queueing system $GI_{\lambda(\bar{q})}|M|1|R$ with thresholds for the queue length average and the income flow rate $\lambda(\bar{q})$, dependent on queue length average, was developed.

УДК 519.673:519.682.6

Аналитический обзор систем символьных вычислений

Д. С. Кулябов, М. Г. Кокотчикова

Системы символьных вычислений способны существенно облегчить вычисления исследователя, освободить его от рутинной работы: вычисления интегралов, производных высоких порядков, решений систем дифференциальных уравнений, построения графиков сложных функций. Статья посвящена современным средствам проведения эксперимента, для облегчения расчетов которого используется компьютер. На основании проведенного аналитического обзора даются рекомендации по использованию программных продуктов. Предпочтение отдается свободно-распространяемому ПО.

Ключевые слова: CAS, Axiom, Maxima, системы компьютерной алгебры, системы символьных вычислений.

Введение

Система символьных вычислений (ССВ), или система компьютерной алгебры (СКА), — мощный программный продукт способный решать широкий круг задач с различным уровнем вычислительной сложности, начиная от простых преобразований выражений: полиномов, рядов, рациональных функций и формул, и вплоть до решения различных систем уравнений. У таких систем есть ряд преимуществ перед системами, которые производят вычисления численно.

Большинство математических систем, используемых в работе с компьютером, являются численными системами. Они превращают компьютер в мощный программируемый калькулятор, позволяющий выполнять арифметические вычисления с огромной скоростью. Хотя в наше время дисковые пространства, оперативная память и скорость обработки операции процессоров стремительно растут, редко такие результаты вычислений бывают абсолютно точными — как правило, при операциях с вещественными числами происходит их округление. Реализация большинства численных методов базируется на заведомо приближенных численных методах. Часто из-за накопления погрешности эти методы расходятся. За пределами возможностей численных математических систем оказались обширные области математики, связанные с проведением аналитических расчётов — от простых подстановок в выражениях до обучения компьютера новым математическим закономерностям и отношениям. Символьные операции — это то, что кардинально отличает системы символьной математики (или компьютерной алгебры) от систем для выполнения численных расчётов.

1. Историческая справка

В научных исследованиях и технических расчетах специалистам приходится гораздо больше заниматься преобразованиями формул, чем собственно численным счетом, однако с появлением компьютеров основное внимание уделялось автоматизации последнего, хотя компьютеры стали применяться для решения таких задач символьных вычислений, как, например, символьное дифференцирование, ещё в 50-е годы прошлого века. Активная разработка систем компьютерной алгебры началась в конце 60-х [1]. С тех пор создано значительное количество различных систем, получивших в разной степени распространение; некоторые системы продолжают развиваться, другие отмирают, появляются новые. Динамика систем непрерывна.

Эру создания компьютерной символьной математики принято отсчитывать с начала 60-х годов. Именно тогда в вычислительной технике возникла новая ветвь компьютерной математики, не совсем точно, но зато броско названная компьютерной алгеброй. Речь шла о возможности создания компьютерных систем, способных осуществлять типовые алгебраические преобразования: подстановки в выражениях, упрощение выражений, операции со степенными многочленами (полиномами), решение линейных и нелинейных уравнений и их систем, вычисление их корней. При этом предполагалась возможность получения аналитических (символьных) результатов везде, где это только возможно.

Большинство пользователей было заинтересовано в том, чтобы правильно выполнить конкретные аналитические преобразования: вычислить в символьном виде производную или первообразную заданной функции, разложить ее в ряд Тейлора или Фурье, а вовсе не в детальном и сложном математическом и логическом описании того, как это делается компьютером (или, точнее, его программистом).

Поняв это, многие западные фирмы приступили к созданию компьютерных систем символьной математики, ориентированных на широкие круги пользователей, не являющихся профессионалами в компьютерной алгебре. Учитывая невероятную большую сложность автоматизации решения задач в аналитическом виде (число математических преобразований и соотношений весьма велико), первые подобные системы удалось создать лишь для больших ЭВМ. Но затем появились и системы, доступные для мини-ЭВМ. Заметное развитие получили языки программирования для символьных вычислений Reduce, система muMath для малых ЭВМ, а в дальнейшем — интегрированные системы символьной математики для персональных компьютеров: Derive, MathCAD, Mathematica, Maple V.

В табл. 1 [1] приведена краткая хронология развития СКА, выделены ее основные этапы.

2. ССВ и системы численных вычислений. Области применимости каждого подхода

Системы численных вычислений применяются на первых этапах исследования, они позволяют оценить результаты. Это может быть предварительным оценочным этапом исследований. Компьютер позволяет облегчить анализ полученных данных. Для этого и предназначены пакеты ROOT и PAW (Physics Analysis Workstation). Сначала задача рассчитывается на наборе точек и при помощи пакета ROOT или PAW возможно приблизить результат какой-нибудь функцией.

PAW — это интерактивная программа, используемая для анализа и графического представления результатов, с возможностью автоматизации посредством скриптов.

PAW позволяет работать с большими объемами данных. Данные в основном представлены в виде списка не связанных между собой событий. Каждое событие представляет из себя набор именованных полей. Данные можно представить в виде одномерных или двумерных гистограмм. В PAW реализована фильтрация событий по условию. Для сравнения экспериментальных данных с теорией PAW предоставляет интерфейс к чрезвычайно мощной и гибкой процедуре подгонки и оценки ошибок.

Пакет PAW был создан для задач физики элементарных частиц (ФЭЧ). История PAW берет свое начало в 1986 году в Европейском Центре Ядерных Исследований (CERN). В процессе разработки основной упор делался на работу с очень большим объемом данных.

Для целей физического анализа PAW распространялась свободно с самого начала, но только в 2000 году она стала доступна по GPL. По сути PAW является интерфейсом к набору библиотек CERNLIB. Все, что можно сделать в PAW интерактивно, также можно реализовать в компилируемом программном коде. В PAW встроено интерпретатор языка программирования Fortran.

Чем же интересна PAW сейчас, когда есть ROOT? Во-первых, нет необходимости менять PAW в тех проектах, где она уже прижилась. Для целей интерактивного анализа PAW эффективнее, чем ROOT, из-за излишней многословности

Таблица 1

Хронология ССВ

Дата	Событие
Середина XIX века	Первые идеи автоматизации формульных выкладок — леди Ада Августа Лавлейс.
50-е годы XX века	Отдельные программы для решения конкретных задач на машинном языке. — Первые программы дифференцирования функций, США — Работы Л.В.Канторовича
60-е годы XX века	Исследования по созданию общих алгоритмов и программ в различных классах выражений. — Появление алгоритмических языков высокого уровня — Реализация полиномиальной алгебры — Алгоритм Риша (1967, система SIN) — Появление программных систем общего назначения (универсальных СКА): REDUCE, MACSYMA, MATHLAB, SCRATCHPAD, FORMAC, СИРИУС, АВТОАНАЛИТИК, АНАЛИТИК
70-е годы XX века	Первые в СССР систематические применения КА для решения физических задач, Дубна, ОИЯИ
80-е годы XX века	— Развитие вычислительной техники: суперкомпьютеры, рабочие станции, РС — Появление систем с развитым пользовательским интерфейсом: MAPLE (1981) и MATHEMATICA (1988) — СКА «нового поколения» (SCRATCHPAD-II, SMALLTALK) — учет абстрактных типов данных, использование средств ООП — Аппаратная реализация: АВТОАНАЛИТИК, LISP-машины
90-е годы XX века	— Интеграция символьных вычислений и численных расчетов — MATHEMATICA и MAPLE: все аспекты научного и технического компьютеринга — в одной системе — СКА становятся серийными коммерческими продуктами
Начало XXI века	— Система Axiom становится свободной (2001) — Появление системы Cadabra (2001) — Умер создатель системы Maxima Вильям Шеттлер (2001), проект стал медленнее разрабатываться

последнего (наследие C++). Многие базовые идеи, которые были заложены в фундамент PAW, достаточно интересны в плане изучения и копирования. Программы для того и делаются свободными, чтобы лежащие в их основе идеи не пропадали.

PAW оперирует векторами, гистограммами, списками событий и выборками подмножеств этих событий. Эти объекты являются для PAW базовыми. Сначала возник PAW, а 10 лет спустя появился ROOT. С 2006 года ROOT (<http://root.cern.ch/>) стал выпускаться под лицензией GNU, и, возможно, скоро попадёт во все основные дистрибутивы GNU/Linux. ROOT активно поддерживается и развивается. У ROOT есть довольно мощное сообщество. На сайте <http://root.cern.ch> можно найти ответ почти на все вопросы, касающиеся пакета.

При построении графиков, в отличие от своего предка PAW, ROOT позволяет интерактивно менять параметры картинки с помощью выпадающих меню. Тип меню зависит от того, на какой объект направлен указатель мыши. Также с помощью левой кнопки можно интерактивно изменять масштаб графика. Не стоит сильно изменять график, так как большое количество операций увеличивает время, потраченное на создание картинок. В отличие от набранных команд, осмысленные движения и клики мыши сохранить для повторного использования невозможно. Гистограмма является одним из основных элементов анализа. Интерпретатор для ROOT написан на языке C, поэтому возможно интегрирование в программу модулей, написанных на языке C.

После предварительной численной оценки можно переходить к аналитическим вычислениям. Таким образом, можно перейти к более детальному изучению построенной численно моделью. СКА снабжаются специальным процессором для выполнения аналитических (символьных) вычислений. Его основой является ядро, хранящее всю совокупность формул и формульных преобразований, с помощью которых производятся аналитические вычисления. Чем больше этих формул в ядре, тем надежней работа символьного процессора и тем вероятнее, что поставленная задача будет решена, если такое решение существует.

Визуализация является важнейшим инструментом исследователя. И тут проявляется различие между ССВ и СЧВ. Для численных расчётов результатом является набор значений, поэтому обычно стоит задача представить их в виде графика, сравнить с экспериментальными результатами. Поэтому важнейшим элементом пакетов численных вычислений являются графические библиотеки. При аналитических расчётах результатом является формула. И на первый план выходит привычное графическое представление формулы, то есть возрастает роль графического интерфейса.

3. Универсальные и специализированные ССВ

СКА традиционно делятся на специализированные и универсальные. Специализированные системы спроектированы и предназначены для решения конкретных задач, например, из небесной механики или квантовой теории поля. Универсальные системы ориентированы на широкий круг математических задач, хотя они имеют в своих библиотеках частные пакеты.

Результатом развития языков программирования, в частности Си и Lisp, стали универсальные математические пакеты, например, Maple, Mathematica, Maxima и Axiom. Эти системы имеют дружественный интерфейс, реализуют множество стандартных и специальных математических операций, снабжены графическими средствами и обладают собственными языками программирования. Все это предоставляет широкие возможности для эффективной работы специалистов разных профилей, о чем говорит активное применение математических пакетов в научных исследованиях и в преподавании. С помощью этих пакетов можно гораздо проще и быстрее решать инженерные задачи, готовить и выполнять домашние задания.

Однако за универсальность приходится расплачиваться. При решении некоторых задач эти системы слишком многословны, а зачастую требуют установки (или даже дописывания) дополнительных модулей. В качестве примера таких задач можно привести теорию поля, где возникают тензорные величины, индексы

(обладающие определёнными свойствами), симметрии. Описание этих сущностей требует усложнённого входного языка. Списочная внутренняя структура, применяемая универсальными системами, становится громоздкой для их представления. И тогда происходит смена парадигм: от универсальных к специализированным системам. Как яркий пример специализированной системы можно привести пакет Cadabra.

На данный момент существует более 26 СКА [2]. Среди существующих математических пакетов можно выделить следующие системы: Maple и Mathematica (коммерческие), Maxima и Axiom (свободно-распространяемые). Одной из отличительных черт у этих систем является операционная система, системы Maple и Mathematica работают под ОС Windows, а системы Maxima и Axiom работают как под ОС Windows, так и под ОС Linux (следует отметить, что работа данных систем под ОС Windows осуществляется с существенными ограничениями, система Axiom под ОС Windows не поддерживает графических модулей).

В табл. 2 приведён список из 15 систем, в также указано применения конкретной системы. 4 системы — Axiom, Maple, Mathematica и Maxima — являются универсальными. Следует обратить внимание и на условия распространения. Предпочтение отдается свободно-распространяемым системам, однако не стоит отодвигать на второй план такую мощную универсальную и принятую за рубежом систему, как Maple.

Сравнение ССВ

Таблица 2

Название	Начало разработки	Лицензия	Применение
1. Axiom	1971	BSD	Универсальная ССВ
2. Cadabra	2001	GPL	Тензорная алгебра, теория поля
3. CoCoA	1987	GPL	Для операций с полиномами
4. Fermat	1986	shareware	Полиномы и матрицы
5. GAP	1986	GPL	Группы, теория характеров, дискретная математика
6. GiNaC	1997	GPL	Библиотека языка C++
7. GINV	2005	GPL	Метод базисов Гребнера для систем уравнений
8. Maple	1979	Коммерч.	Универсальная ССВ
9. MathCad	1985	Коммерч.	WYSIWYG интерфейс и типографское качество вывода информации
10. Mathematica	1986	Коммерч.	Универсальная ССВ
11. Mathomatic	1986	LGPL	Общая алгебра
12. Maxima	1967	GPL	Универсальная ССВ
13. Meditor	2000	GPL	Библиотека символьных вычислений языка Java и математический редактор
14. PARI/GP	1985	GPL	Теория чисел и заданная точность вычислений
15. Singular	1984	GPL	Полиномы, кольца, идеалы

4. Универсальные системы Axiom и Maxima

Перечислим общие особенности систем Maxima и Axiom перед системами, приведенными в табл. 2. Во-первых, эти две системы являются свободно-распространяемыми системами (лицензия GPL). Во-вторых, среда программирования язык Lisp, это, однако, не означает, что пользователь пишет команды на языке Lisp. Для каждой системы определен заранее список команд, которые предварительно писались создателем на языке Lisp, одной из целей создания которого являлась возможность вычислений в символьном виде. В-третьих, эти ССВ являются кроссплатформенными, то есть работают под ОС Linux и под ОС Windows. Для того, чтобы работать в этих ССВ, пользователю не обязательно переучиваться с Windows на Linux и наоборот.

Перечислим особенности системы Maxima:

- 1) Чувствительность к регистру.
- 2) Консольная программа, но возможны интеграции в графические оболочки. Следует подробно остановиться на этих оболочках: `TeXmacs` и `wxMaxima`.

В `wxMaxima` вводятся формулы в текстовом виде, а вывод отображается графически, привязанными математическими символами. Кроме того, большой упор здесь сделан на удобство ввода: командная строка отделена от окна ввода-вывода, а дополнительные кнопки и система меню позволяют вводить команды не только в текстовом, но и в диалоговом режиме. Так называемое «автодополнение» в командной строке на самом деле с таковым имеет лишь то сходство, что вызывается клавишей «Tab». Реализована как вызов истории команд, он вызывает ту команду из уже введенных в этой сессии, которая начинается с заданных в командной строке символов, но не дополняет до имен команд и их параметров. Таким образом, этот интерфейс наиболее удобен в том случае, когда нужно много вычислять и видеть результаты на экране; и еще для быстроты ввода команд. Кроме того, `wxMaxima` предоставляет удобный интерфейс к документации по системе; хотя, так как документация поставляется в формате `html`, вместо этого можно использовать обычный браузер.

Второй достаточно интересный интерфейс к Maxima — это дополнительный режим в редакторе `TeXmacs`, который подгружается в виде сессии. Maxima-сессия вызывается из меню: «Вставить → Сессия → Maxima», при этом появляется дополнительное меню с командами Максими. После запуска сессии можно уже внутри нее перейти в математический режим ввода и при вводе также использовать элементы математической нотации. Этот редактор разрабатывается для визуального редактирования текстов научной тематики, при котором вы видите на экране редактируемый текст практически в том же виде, в котором он будет распечатан. В частности, он имеет математический режим ввода, очень удобный для работы с самыми разнообразными формулами, и умеет импортировать/экспортировать текст в `LaTeX` и `XML/HTML`. Именно возможностями по работе с формулами пользуется Maxima, вызванная из `TeXmacs`. Фактически, формулы отображаются в привычной математической нотации, но при этом их можно редактировать и копировать в другие документы наподобие обыкновенного текста. Этот интерфейс будет наиболее удобен тем, кто хочет использовать результаты вычислений в своих текстах и любит редактировать их в визуальном режиме.

- 3) Не требовательна к ресурсам компьютера. Работает на маломощных компьютерах.
- 4) Каждый оператор в Maxima — функция.
- 5) Комплексные числа записываются в алгебраической форме: $a+b*i$, где i — мнимая единица, a и b — соответственно, действительная и мнимая части.
- 6) Символьное вычисление: пределы, упрощение выражения, производные, интегралы, упрощение тригонометрических функций, решение дифференциальных уравнений.
- 7) Внешняя графическая библиотека `GNUPlot`.

Следующей особенностью этих систем является возможность построение графиков. В системе Maxima построение осуществляется при помощи пакета GNUPlot.

Кроме перечисленных особенностей, для системы Axiom присущи и другие особенности:

- 1) Строгая типизация. Переменные и функции должны быть изначально четко определены, либо не определены вообще, в этом случае система сама подбирает тип для соответствующей переменной.
- 2) Возможность добавления модулей. Модуль представляет собой отдельный файл с расширением `*.input`, который при надобности подгружается потом.
- 3) Документированность. Эта система хорошо документирована, можно скачать подробную книгу около 1100 страниц.
- 4) Построение графиков. Особенностью является то, что в ОС Windows построение графиков пока невозможно.
- 5) Среда программирования: Lisp, Aldor, Fortran. Lisp — внутренний язык системы, существует возможность компиляции и на других языках: Aldor и Fortran.
- 6) Собственная встроенная графическая библиотека.
- 7) Символьные вычисления. Система с легкостью производит в символьном виде следующие операции:
 - преобразование;
 - вычисление производных, пределов, интегралов;
 - решение различных систем уравнений;
 - работа с матрицами.

Для тех, кто привык работать не в консоли, а графической оболочке, система Axiom предоставляет возможность работы в `TeXmacs`. Принципы работы сходны с описанными ранее принципами работы с системой Maxima.

Выводы

Дадим рекомендации по областям применения основных СКА. При исследовании сложной системы на начальном этапе следует применить численные системы: ROOT и PAW. Это позволит приблизить, сгладить численные результаты, аппроксимировать известными функциями. Далее стоит применять либо узкоспециализированные системы (см. табл. 2), либо универсальные СКА, Axiom и Maxima, речь о которых шла выше. При визуализации данных используются в зависимости от СКА либо встроенные библиотеки, либо внешние (например, GNUPlot). Этот самостоятельный пакет хорош тем, что позволяет визуализировать данные из разных программ, файлов. Если перед Вами стоит задача из

Таблица 3

Ссылки на программные продукты

Название	Официальный сайт
PAW	http://paw.web.cern.ch/paw/
ROOT	http://root.cern.ch/
Axiom	http://wiki.axiom-developer.org/
Maxima	http://maxima.sourceforge.net/
GNUPlot	http://www.gnuplot.info/
TeXmacs	http://www.texmacs.org/
wxMaxima	http://wxmaxima.sourceforge.net/

области дискретной математики или теории поля, обратите внимание на специализированные СКА. Ссылки на программные продукты, упомянутые в статье, представлены в табл. 3.

Основное направление, кроме научных разработок, где такие программы востребованы — это высшее образование; а использование для учебных нужд именно свободного ПО — это реальная возможность и для вуза, и для студентов и преподавателей иметь в своем распоряжении легальные копии такого ПО без больших и даже сколько-нибудь существенных денежных затрат.

Литература

1. *Панкратьев Е. В.* Элементы компьютерной алгебры. — California: Интернет-университет информационных технологий — ИНТУИТ.ру, БИНОМ. Лаборатория знаний Серия: Основы информатики и математики, 2007. — 248 с.
2. Comparison of computer algebra systems. — http://en.wikipedia.org/wiki/Comparison_of_computer_algebra_systems.

UDC 519.673:519.682.6

Analytic Review of Computer Algebra System

D. S. Kulyabov, M. G. Kokotchikova

A computer algebra system (CAS) is a software program that facilitates symbolic mathematics. The core functionality of a CAS is manipulation of mathematical expressions in symbolic form. This article describe some of CAS, opensource primary, and make his comparison.

УДК 681.3

Свойства аналитических измерений

Е. Б. Асламов, А. В. Висков, М. Б. Фомин

*Кафедра информационных технологий
Российский университет дружбы народов
Россия, 117198, Москва, ул. Миклухо-Маклая, 6*

В данной статье рассматривается формальное определение аналитического измерения, исходя из анализа которого сформулированы и доказаны некоторые свойства аналитических измерений

КЛЮЧЕВЫЕ СЛОВА: многомерные модели данных, аналитическое измерение, OLAP.

1. Введение

Анализ многомерных данных, как правило, производится в привязке к конкретному информационному контексту, определяемому предметной областью. Точка зрения на анализируемые данные задается аналитическими измерениями – сущностями предметной области, от которых зависит предмет анализа. Аналитическое измерение (далее просто измерение) состоит из множества объектов (называемых членами измерений) одного или нескольких типов, обеспечивающих информационный контекст анализируемых данных.

2. Определение измерения

Измерение включает объекты, представляющие различную степень детализации анализируемых данных. Совокупность элементов, обозначающая одну и ту же степень детализации, образует уровень измерения. Уровни показывают, как элементы измерения группируются для применения агрегационных функций. Связи, определенные между элементами на разных уровнях агрегации, являются композицией.

Таким образом, измерение можно определить как связный ориентированный граф. Каждая вершина в этом графе соответствует уровню агрегации, а дуга показывает, что каждый элемент уровня, определяющего конец дуги, декомпозируется на множество элементов нижележащего уровня, соответствующего началу дуги.

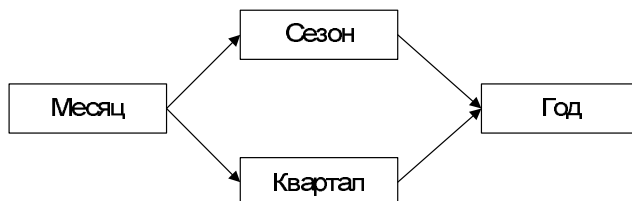


Рис. 1. Пример измерения

3. Свойства измерений

Как показано на рисунке 2, можно выделить три различных, независящих от области применения типа отношений «часть-целое», определяемых составной структурой целого. Выделяется три типа структуры целого:

- 1) масса – в случае отсутствия определенной структуры целого, считающегося гомогенным (например, кучка риса);
- 2) коллекция – в случае, когда рассматриваются различные элементы целого, имеющего однородную составную структуру (например, колонна грузовиков);
- 3) комплекс – в случае, когда различные части целого выполняют различные роли, целое является комплексом с гетерогенной составной структурой (например, части двигатель).

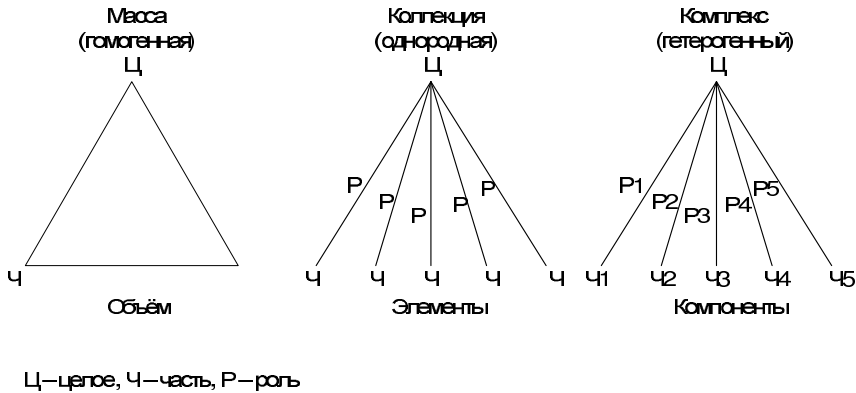


Рис. 2. Типы целого

Главная цель определения связей между различными сущностями измерения состоит в том, чтобы показать, каким образом могут применяться агрегационные функции (такие как сумма, минимум, максимум, среднее, и т.д.). Так как эти функции рассматривают элементы измерения как играющие в агрегации одну и ту же роль, то эти связи должны рассматриваться как коллекция. С этого момента, отношения «часть-целое» между уровнями агрегации в измерении должны пониматься как образующие коллекцию.

В случае рассмотрения коллекции, для описания и обоснования свойств целого, а также связей между составляющими элементами можно использовать аксиоматику классической экстенциональной мереологии [1, 2].

- 1) **Существование.** Если A является частью B , то и A и B существуют.
- 2) **Асимметричность.** Если A является частью B , то B не является частью A .
- 3) **Транзитивность.** Если A является частью B и B является частью C , то A является частью C .
- 4) **Дополнение.** Если A является некоторой частью B , тогда существует другой субъект C , который является оставшейся частью B .
- 5) **Экстенциональность.** A и B состоят из одних и тех же частей тогда и только тогда когда A и B совпадают.
- 6) **Сумма.** Для любых двух субъектов всегда найдется составленный из них третий субъект.

Исходя из определения и аксиом мереологии, можно вывести некоторые свойства измерений.

Свойство 1: Измерение не содержит циклов.

Доказательство. Предположим, что в графе измерения есть цикл. В цикле участвуют хотя бы два уровня. По аксиоме 3 (транзитивности) элемент A одного уровня есть часть элемента B другого уровня и наоборот. Получаем противоречие с аксиомой 2. Значит предположение не верно. □

Свойство 2: Для каждого измерения существует единственный атомарный уровень агрегации, содержащий элементарные сущности, которые не могут быть разделены на более мелкие части.

Доказательство. По свойству 1, существует как минимум один уровень, элементы которого не содержат частей.

Если существует более одного атомарного уровня, то, так как граф измерения связный, и, учитывая аксиому 3, будет существовать элемент E , который является композицией элементарных сущностей каждого из атомарных уровней. По аксиоме 5, все коллекции элементарных сущностей, составляющих E , должны являться одной и той же коллекцией элементов. Таким образом, существует только один атомарный уровень. $\square$

Свойство 3: Для любого измерения, может существовать уровень «Все», содержащий элементы, составленные всеми элементарными сущностями измерения. Если такой уровень существует то,

- 1) его элементы не являются частями никакого другого уровня агрегации;
- 2) этот уровень агрегации содержит в точности один элемент;
- 3) этот уровень единственный в измерении.

Доказательство. Последовательно применяя аксиому 6, можно построить элемент E , состоящий из всех элементарных сущностей измерения.

- 1) Если E будет являться частью некоторого элемента E' то, по аксиоме 4 будет существовать элементарная сущность, не входящая в E , что противоречит предположению. Поэтому, E принадлежит уровню, элементы которого не являющегося частью никакого другого уровня в измерении.
- 2) Если этот уровень будет содержать два элемента, оба содержащие все элементарные сущности, то, по аксиоме 5, они будут являться одним и тем же элементом.
- 3) Этот уровень единственный, потому, что если бы существовал другой уровень, элементы которого включают все элементарные сущности, то они бы совпадали с элементом, имеющимся на уровне «Все» (по аксиоме 5). $\square$

Свойство 4: Те уровни, элементы которых не являются коллекцией элементов любого другого уровня (т.е. они не являются началом дуги в графе) могут быть связаны некоторым ребром с уровнем «Все».

Доказательство. Элемент уровня «Все» может быть разложен на сущности любого уровня, покрывающие атомарный уровень. Если существует уровень, не покрывающий атомарный уровень, то, по аксиоме 6, к его элементам могут быть добавлены новые элементы, составленные из коллекций недостающих элементов атомарного уровня. $\square$

Свойство 5: Каждый элемент уровня не являющегося атомарным содержит как минимум одну часть.

Доказательство. Сущности, не включающие частей, являются элементарными, а все элементарные сущности, по свойству 2, располагаются на атомарном уровне. $\square$

Свойство 6: Каждый элемент уровня, не являющегося атомарным может иметь более чем одну часть.

Доказательство. По аксиоме 4, если элемент включает часть, то найдется другой элемент, являющийся оставшейся частью исходного элемента. $\square$

Свойство 7: Элемент может являться частью нескольких коллекций в одно и тоже время.

Доказательство. Не существует мереологической аксиомы запрещающей, чтобы элемент являлся частью нескольких коллекций. В качестве доказательства можно рассмотреть пример. Как показано на рисунке 3, заданный товар (на уровне «товаров») может принадлежать двум различным видам товаров на уровне «видов товаров». Такое допущение не создает каких-либо концептуальных проблем, однако может являться причиной возникновения вычислительных сложностей. Так, например, некоторые вычисляемые атрибуты сущностей на уровне «семейство» (которые состоят из элементов уровня «виды товаров») должны рассчитываться на основании элементов на уровне «товар» (например, размерность множества разложения элемента подарки (уровень «семейство товаров») не равна сумме размерностей множеств разложений элементов сладости и игрушки (уровень «типы товаров»): $card(\text{подарки}) \neq card(\text{сладости}) + card(\text{игрушки})$).

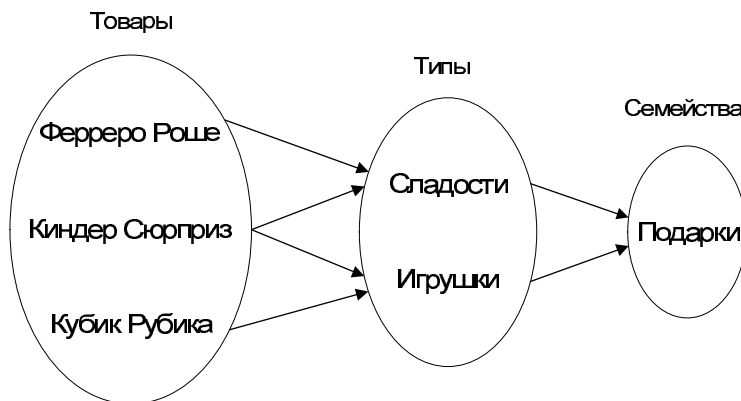


Рис. 3. Пример образования целого с пересечением

Исходя из определения измерения, и, по свойствам 1 и 2, можно сделать вывод, что в общем случае граф уровней агрегации измерения образует сеть. Кроме того, свойства 3 и 4 показывают, что в графе измерений, для того, чтобы получить сеть, всегда можно определить уровень «Все». Свойства 5 и 6 показывают, что отношение между двумя уровнями будет включать $1..N$ частей для каждого целого. То есть для каждого элемента на вышележащем уровне всегда можно определить его разложение на нижележащем уровне в соответствующей иерархии. Свойство 7 поясняет, что часть может входить в состав как одного целого, так и нескольких. Следует отметить, что если элемент нижележащего уровня может являться частью только одного элемента вышележащего уровня, то в таком случае он обязательно входит в состав некоторого элемента вышележащего уровня, если же множества разложений элементов могут пересекаться, то, вообще говоря, элемент нижележащего уровня может и не входить в элемент вышележащего уровня. Однако, учитывая аксиому 6, которая свидетельствует о том, что для любой части всегда можно определить целое, получаем, что в случае необходимости всегда можно доопределить элемент вышележащего уровня, который будет включать рассматриваемый элемент нижележащего уровня. Это означает, что элемент нижележащего уровня всегда включается как минимум в один элемент вышележащего уровня в соответствующей иерархии.

На рисунке 4 показаны допустимые кардинальные отношения между элементами уровней в агрегационных иерархиях. Можно выделить два вида отношений, таких что, каждому целому будет принадлежать хотя бы одна часть. В наиболее распространенном случае оказывается, что существует в точности одно целое для каждой части. Однако, также возможно, что определенная часть принадлежит нескольким целым. В таком случае, если найдется часть, которая не принадлежит ни одному целому, всегда могут быть

построены целые, так что каждая часть будет принадлежать хотя бы одному целому.

Кроме отношений «часть-целое» между элементами измерения можно выделить также связи, являющиеся ассоциациями. Такой тип связи используется для того, чтобы показать, что одна сущность описывает некоторое свойство другой сущности. Все сущности в измерении будут связаны с некоторыми другими сущностями, так как они являются их частями, либо для того чтобы показать некоторые их свойства.

Отношение «часть-целое» показывает, как различные элементы группируются вместе в измерении, тогда как ассоциация показывает, какие характеристики доступны для построения выборок сущностей. Таким образом, операции «подъём» («roll-up») и «спуск» («drill-down») будут выполняться вдоль связей «часть-целое» (по иерархии), тогда как выбор элементов («slice-dice») будет выполняться в соответствии с отношениями ассоциации.

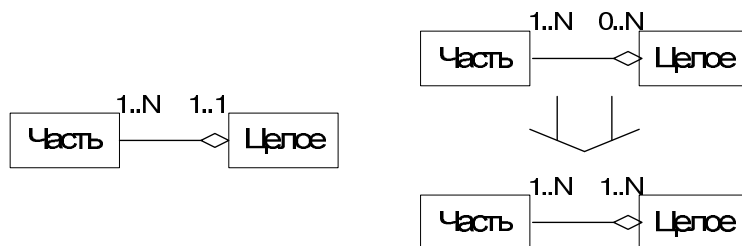


Рис. 4. Возможные кардинальные отношения между элементами уровней в иерархии измерения

Специализация – это одно из наиболее важных отношений, которое должно быть показано на многомерной схеме [3, 4]. Изолированные уровни агрегации не могут быть специализированы для отображения более специфичных аспектов, они должны рассматриваться в рамках измерения. □

Свойство 8: В общем случае, уровень и его специализация могут не принадлежать одному и тому же измерению.

Доказательство. Предположим, что и уровень L и его специализация L_s находятся в одном измерении. Для того чтобы определить полусеть с уровнем «Все», может потребоваться добавить на уровне L_s некоторые элементы, так как в таком случае L_s должен покрывать атомарный уровень. Эти элементы, могут не удовлетворять критерию специализации. Поэтому, не всегда возможно обеспечить принадлежность обоих уровней одному измерению.

На рисунке 5 показан пример, когда измерение «Люди» специализировано в виде измерения «Продавцы» на уровне «Профессия» (сплошная непрерывная стрелка). Эта специализированное измерение содержит уровень, элементами которого являются все люди, являющиеся продавцами, и еще один уровень с единственным элементом (который является сущностью «Профессии»), представляющим собой множество всех продавцов. Разрывная линия показывает, что уровень является специализацией другого (элемент уровня «Все» измерения «Продавцы» является сущностью «Профессии» в случае, когда выполняется критерий специализации «Профессия = Продавец»). Уровень «Возрастная Группа» для измерения «Продавцы» не рассматривается. Следует отметить, что если бы такой уровень в измерении «Продавцы» рассматривался, то он бы не являлся специализацией соответствующего уровня в измерении «Люди», так как его элементы отличались бы от элементов на уровне «Возрастная Группа» в измерении «Люди» (они включали бы меньше людей).

Обобщая пример, если D_s это специализация измерения D на уровне L , то в этом случае D_s включает, по меньшей мере, уровень L_s (специализацию уровня L), и специализации каждого из уровней в D , содержащих часть элементов L_s . Эти специализированные уровни содержат в точности те элементы

соответствующих уровней из D , которые являются любой частью элемента L_s . Кроме этих обязательных уровней, D_s также возможно включаются и другие уровни (которые не являются специализацией каких либо уровней в D), элементы которых не содержатся в D .

Специализации измерения предоставляет возможность показать специфику отношения «часть-целое» для некоторого подмножества элементов измерения, а также выделить общие для этого подмножества атрибуты. Все сущности некоторого уровня имеют общие свойства, так как уровень определяет заданный класс объектов, выполняющих одну и ту же роль в коллекции. Отношение ассоциации, также как и агрегации наследует в рамках специализации. Таким образом, отношение специализации имеет большое значение в вопросе повторного использования измерений.

Рассматриваемые семантические особенности отношения специализации важны не только для конечных пользователей – аналитиков, они позволяют также оптимизировать запросы и повышать их производительность, позволяя, например, осуществлять параллельные вычисления по разным специализациям измерения.

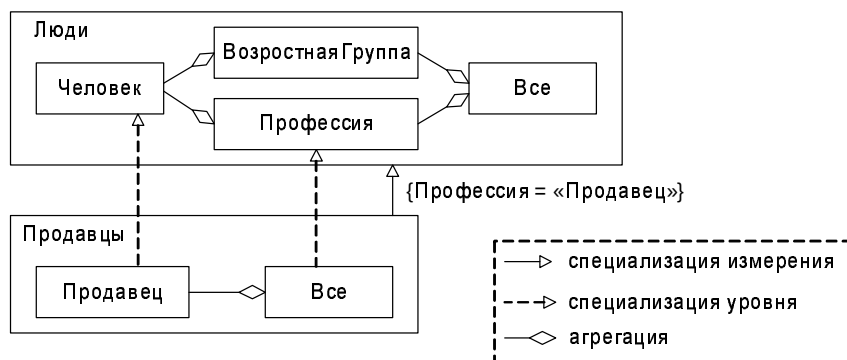


Рис. 5. Пример определения специализации

□

4. Заключение

В работе были рассмотрены и доказаны свойства аналитических измерений. Рассуждения основывались на мереологических аксиомах. Кроме описания связей между элементами одного измерения, также были разобраны возможные виды отношений между различными измерениями, как одной схемы данных, так и разных. Рассмотренный способ описания измерений и вытекающие из него свойства могут быть полезны для построения модели предметной области и позволяют разработчикам информационных систем более чётко проводить классификацию и группировку данных.

Литература

1. PartWhole relations in Object-centered systems: an overview / A. Artale, E. Franconi, N. Guarino, L. Pazzi // Data and Knowledge Engineering (DKE). — Vol. 20, No 3. — 1996. — Pp. 347–383.
2. Gerstl P., Pribbenow S. Midwinters, end games, and body parts: A classification of partwhole relations // International Journal of HumanComputer Studies. — Vol. 43(5,6). — 1995. — Pp. 865–889.
3. Trujillo J., Palomar M., Gomez J. Applying Object-Oriented Conceptual Modeling Techniques to the Design of Multidimensional Databases and OLAP applications //

In Proceedings of the 1st International Conference on WebAge Information Management – WAIM'2000, volume 1846 of LNCS. — Springer, 2000. — Pp. 83–94.

4. Tryfona N., Busborg F., Christiansen J. starER: A conceptual model for data warehouse design // In Proceedings of the 2nd International Workshop on Data Warehousing and OLAP – DOLAP 99. — ACM, 1999. — Pp. 3–8.

UDC 681.3

Analytical Dimensions Properties

E. B. Aslamov, A. V. Viskov, M. B. Fomin

In this paper, we propose a formal definition for an analytical dimension. We formulate and prove some properties of an analytical dimension within the proposed formal definition.

ФИЗИКА

УДК 539.123, 539.123.06

Осцилляции нейтрино в схеме смешивания зарядов (констант связей)

Х. М. Бештоев*Объединенный институт ядерных исследований
Россия, 141980, Моск. область, Дубна, ул. Жоллио Кюри, 6*

В стандартной теории осцилляций нейтрино используется схема массовых смешиваний, т. е. параметры осцилляций выражаются через элементы массовой матрицы. В данной работе рассматриваются осцилляции нейтрино, генерированные смешиванием зарядов (констант связей слабых взаимодействий). Получены выражения для углов смешивания и длин осцилляций. Также вычислены выражения для вероятностей переходов трех нейтринных осцилляций. Осцилляции нейтрино в этой схеме (механизме) являются реальными, если массы нейтрино равны, и виртуальными, если массы нейтрино не равны.

Введение

Предположение, что, по аналогии с K^0 , $\bar{K}^0$ осцилляциями, могли бы быть нейтрино–антинейтринные осцилляции ($\nu \rightarrow \bar{\nu}$), рассматривалось Понтекорво [1] в 1957. Впоследствии Маки и др. [2] и Pontecorvo [3] предполагали, что могут быть смешивания (и осцилляции) нейтрино различных ароматов (т. е. могут существовать $\nu_e \rightarrow \nu_\mu$ переходы).

В общем случае возможны две схемы (типы) смешивания (осцилляции) нейтрино: схема массовых смешиваний и схема зарядовых смешиваний (как это имеет место в модели векторной доминантности или при смешивании векторных бозонов в стандартной модели электрослабых взаимодействий) [4–6].

В стандартной теории осцилляции нейтрино [7–10] предполагается, что физически наблюдаемые нейтринные состояния ν_e , ν_μ , ν_τ не имеют определенные массы (в этом случае, мы не можем сформулировать закон сохранения энергии-импульса в строгой форме в реакциях с участием нейтрино [4–6]) и что они прямо рождаются как суперпозиции ν_1 , ν_2 , ν_3 нейтрино.

Вопрос об осцилляции нейтрино можно рассмотреть в двух подходах.

1. Не интересуясь причиной происхождения смешивания нейтрино, можно просто сформулировать, что ν_e , ν_μ , ν_τ нейтрино сразу рождаются в суперпозиционных состояниях ν_1 , ν_2 , ν_3 нейтрино, как это формулируется в стандартном подходе [7–10]. Тогда принцип причинности будет нарушен, так как мы предполагаем, что ν_e , ν_μ , ν_τ нейтрино в момент рождения уже знают, что они превратятся в суперпозицию ν_1 , ν_2 , ν_3 нейтрино. Нарушение принципа причинности означает, что в этом процессе закон сохранения энергии-импульса не будет выполняться. Действительно, в этом подходе считается, что ν_e , ν_μ , ν_τ нейтрино не имеют определенной массы. Это означает, что мы не можем сформулировать закон сохранения энергии-импульса в точной форме.
2. Очевидно, что так как ν_e , ν_μ , ν_τ нейтрино рождаются в слабых взаимодействиях и являются собственными состояниями этих взаимодействий, то они должны иметь определенные массы, и лептонные числа должны сохраняться. Если присутствует взаимодействие, нарушающее лептонные числа, которые

имеют собственные состояния ν_1, ν_2, ν_3 , то тогда ν_e, ν_μ, ν_τ нейтрино должны превратиться в суперпозиции ν_1, ν_2, ν_3 нейтрино. В этом случае массовая матрица становится недиагональной, и эти недиагональные члены ответственны за нарушение лептонных чисел. Было показано [4–6, 11–13], что эти недиагональные члены можно интерпретировать как ширины переходов между ν_e, ν_μ, ν_τ нейтрино и что они являются ответственными за возникновение осцилляций нейтрино. При таком подходе мы можем сформулировать закон сохранения энергии-импульса в строгом виде.

В случае, если осцилляции нейтрино генерируются массовой матрицей, тогда параметры смешивания нейтрино выражаются через элементы массовой матрицы нейтрино. Массовый лагранжиан двух нейтрино (ν_e, ν_μ) имеет следующий вид (для упрощения рассматривается случай двух нейтрино)

$$\begin{aligned} \mathcal{L}_M &= -\frac{1}{2} [m_{\nu_e} \bar{\nu}_e \nu_e + m_{\nu_\mu} \bar{\nu}_\mu \nu_\mu + m_{\nu_e \nu_\mu} (\bar{\nu}_e \nu_\mu + \bar{\nu}_\mu \nu_e)] \equiv \\ &\equiv -\frac{1}{2} (\bar{\nu}_e, \bar{\nu}_\mu) \begin{pmatrix} m_{\nu_e} & m_{\nu_e \nu_\mu} \\ m_{\nu_\mu \nu_e} & m_{\nu_\mu} \end{pmatrix} \begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix}, \end{aligned} \quad (1)$$

который диагонализуется вращением на угол θ , и тогда этот лагранжиан (1) приобретает следующий вид (см. [7–10]):

$$\mathcal{L}_M = -\frac{1}{2} [m_1 \bar{\nu}_1 \nu_1 + m_2 \bar{\nu}_2 \nu_2], \quad (2)$$

где

$$m_{1,2} = \frac{1}{2} \left[(m_{\nu_e} + m_{\nu_\mu}) \pm \left((m_{\nu_e} - m_{\nu_\mu})^2 + 4m_{\nu_e \nu_\mu}^2 \right)^{1/2} \right],$$

и угол θ определяется следующим выражением

$$\operatorname{tg}(2\theta) = \frac{2m_{\nu_e \nu_\mu}}{(m_{\nu_\mu} - m_{\nu_e})}, \quad (3)$$

$$\begin{aligned} \nu_e &= \cos \theta \nu_1 + \sin \theta \nu_2, \\ \nu_\mu &= -\sin \theta \nu_1 + \cos \theta \nu_2. \end{aligned} \quad (4)$$

Тогда массы ν_e, ν_μ есть

$$\begin{aligned} m_{\nu_e} &= m_1 \cos^2 \theta + m_2 \sin^2 \theta, \\ m_{\nu_\mu} &= m_1 \sin^2 \theta + m_2 \cos^2 \theta, \end{aligned} \quad (5)$$

в отличие от первоначального предположения, что ν_e, ν_μ, μ_τ нейтрино не имеют никакой определенной массы.

Вероятность $\nu_e \rightarrow \nu_e$ перехода дается следующим выражением:

$$P(\nu_e \rightarrow \nu_e) = 1 - \sin^2(2\theta) \sin^2((m_2^2 - m_1^2)/2p)t, \quad (6)$$

где

$$\sin \theta = \frac{1}{\sqrt{2}} \left[1 - \frac{|m_{\nu_\mu} - m_{\nu_e}|}{\sqrt{(m_{\nu_\mu} - m_{\nu_e})^2 + (2m_{\nu_e \nu_\mu})^2}} \right], \quad (7)$$

или

$$\sin^2(2\theta) = \frac{(2m_{\nu_e \nu_\mu})^2}{(m_{\nu_e} - m_{\nu_\mu})^2 + (2m_{\nu_e \nu_\mu})^2}. \quad (8)$$

Тогда недиагональный член массовой матрицы $m_{\nu_e \nu_\mu}$ в (1) может интерпретироваться как ширина $\nu_e \leftrightarrow \nu_\mu$ переходов [4–6].

В стандартной теории осцилляций нейтрино осцилляции нейтрино являются реальными, даже если массы нейтрино различаются, поэтому закон сохранения энергии и импульса нарушен. В скорректированной теории осцилляций нейтрино [11–13] законы сохранения энергии и импульса выполняются, и нейтринные осцилляции являются виртуальными, если массы нейтрино различаются, и реальными, если массы нейтрино равны.

Необходимо отметить, что в физике все процессы реализуются через динамику. К сожалению, в выше рассмотренной схеме массовых смешиваний динамика отсутствует. Вероятно, есть указание на то, что эта схема является неполной, т. е. эта схема требует физического обоснования. Ниже мы рассмотрим осцилляции нейтрино, которые появляются в схеме зарядовых (констант связей) смешиваний, т. е. при использовании динамики [4–6].

1. Осцилляции нейтрино в рамках схемы зарядовых смешиваний

Сначала мы рассматриваем случай смешивания (осцилляции) двух нейтрино, а затем перейдем к рассмотрению случая осцилляций трех нейтрино.

1.1. Случай смешивания (осцилляции) двух нейтрино в схеме зарядовых смешиваний

В этой схеме (или механизме) смешивания переходы между нейтрино может быть реализовано путем смешивания нейтринных полей по аналогии с моделью векторной доминантности ($\gamma - \rho^0$ или $Z^0 - \gamma$ смешивания), как это имеет место в физике элементарных частиц. Тогда, в случае двух нейтрино мы имеем

$$\begin{aligned}\nu_1 &= \cos \theta \nu_e - \sin \theta \nu_\mu, \\ \nu_2 &= \sin \theta \nu_e + \cos \theta \nu_\mu.\end{aligned}\tag{9}$$

Заряженный ток в стандартной модели слабых взаимодействий для двух семейств лептонов имеет следующий вид:

$$\begin{aligned}j^\alpha &= (\bar{e}\bar{\mu})_L \gamma^\alpha V \begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix}_L, \\ V &= \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix},\end{aligned}\tag{10}$$

и тогда лагранжиан взаимодействия есть

$$\mathcal{L} = \frac{\Gamma}{\sqrt{2}} j^\alpha W_\alpha^+ + \text{h.c.}\tag{11}$$

и

$$\begin{aligned}\nu_e &= \cos \theta \nu_1 + \sin \theta \nu_2, \\ \nu_\mu &= -\sin \theta \nu_1 + \cos \theta \nu_2.\end{aligned}\tag{12}$$

Лагранжиан (10)–(11) можно переписать в следующей форме

$$\mathcal{L} = \frac{\Gamma}{\sqrt{2}} j^\alpha W_\alpha^+ + \text{h.c.},\tag{13}$$

где j^α есть

$$j^\alpha = (\bar{e}\bar{\mu})_L \gamma^\alpha \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}_L,$$

А массовая матрица есть

$$\begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}.$$

В этом случае осцилляции нейтрино не могут иметь место, и даже если нейтринные осцилляции имеют место, то должны быть осцилляции между $\nu_1 \leftrightarrow \nu_2$ нейтрино, но не между $\nu_e \leftrightarrow \nu_\mu$ нейтрино.

В этом месте возникают некоторые вопросы. Откуда мы взяли ν_e, ν_μ нейтрино, если в слабых взаимодействиях, даваемых выражением (13), присутствуют только ν_1, ν_2 нейтрино? Из всех ускорительных экспериментальных данных хорошо известно, что в слабых взаимодействиях рождаются ν_e, ν_μ нейтрино, и лептонные числа l_{ν_e}, l_{ν_μ} хорошо сохраняются. Очевидно, мы должны решить эту проблему. ν_1, ν_2 нейтрино есть собственные состояния слабых взаимодействий, нарушающие лептонные числа, а ν_e, ν_μ нейтрино есть собственные состояния слабых взаимодействий с W, Z^0 бозон обменом. И мы должны переписать лагранжиан слабых взаимодействий в правильной форме, чтобы корректно описать рождения и осцилляции нейтрино. Тогда

$$\mathcal{L} = \frac{\Gamma}{\sqrt{2}} j^\alpha W_\alpha^+ + \text{h.c.}, \quad (14)$$

где

$$j^\alpha = (\bar{e}\bar{\mu})_L \gamma^\alpha \begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix}_L.$$

Каким образом нарушаются лептонные числа? Видимо, необходимо предположить, что после рождения ν_e, ν_μ нейтрино имеет место нарушение лептонных чисел, т. е.

$$\begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix}_L = V \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix}_L, \quad V = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \quad (15)$$

и тогда ν_e, ν_μ нейтрино становятся суперпозициями ν_1, ν_2 нейтрино:

$$\begin{aligned} \nu_e &= \cos \theta \nu_1 + \sin \theta \nu_2, \\ \nu_\mu &= -\sin \theta \nu_1 + \cos \theta \nu_2. \end{aligned} \quad (16)$$

Учитывая, что заряды ν_1, ν_2 нейтрино есть g_1, g_2 , мы получаем

$$g \cos \theta = g_1, \quad g \sin \theta = g_2, \quad (17)$$

т. е.

$$\cos \theta = \frac{g_1}{g}, \quad \sin \theta = \frac{g_2}{g}. \quad (18)$$

Так как $\sin^2 \theta + \cos^2 \theta = 1$, тогда

$$g = \sqrt{g_1^2 + g_2^2}$$

и

$$\cos \theta = \frac{g_1}{\sqrt{g_1^2 + g_2^2}}, \quad \sin \theta = \frac{g_2}{\sqrt{g_1^2 + g_2^2}}. \quad (19)$$

Если предположить, что $g_1 \cong g_2 \cong \frac{g}{\sqrt{2}}$, тогда

$$\cos \theta \cong \sin \theta \cong \frac{1}{\sqrt{2}}. \quad (20)$$

В общем случае константы связи g_1, g_2 и g могут не иметь никакой связи и тогда получим только выражение (19).

Что происходит с массовой матрицей нейтрино в этом случае? Первичная массовая матрица нейтрино имеет диагональный вид

$$\begin{pmatrix} m_{\nu_e} & 0 \\ 0 & m_{\nu_\mu} \end{pmatrix}, \quad (21)$$

так как в слабых взаимодействиях (с обменом W, Z^0 бозонами) лептонные числа сохраняются, тогда ν_e, ν_μ — собственные состояния этих взаимодействий.

Интересно заметить, что та же самая ситуация имеет место в кварковом секторе, когда мы рассматриваем $K^0, \bar{K}^0$ осцилляций. В сильных взаимодействиях рождаются только d, s, b кварки, и ароматические числа в этих взаимодействиях хорошо сохраняются, т. е. эти состояния являются собственными состояниями сильных взаимодействий. Тогда осцилляции появляются при нарушении ароматических чисел в слабых взаимодействиях при введении матриц Кабиббо–Кобаяши–Маскавы.

Тогда из-за присутствия нарушения лептонных чисел, в этой матрице появляются недиагональные члены, и тогда эта масса матрицы преобразовывается в следующую недиагональную матрицу (CP сохраняется):

$$\begin{pmatrix} m_{\nu_e} & m_{\nu_e \nu_\mu} \\ m_{\nu_\mu \nu_e} & m_{\nu_\mu} \end{pmatrix} \quad (22)$$

тогда массовый лагранжиан нейтрино приобретает следующий вид ($\nu \equiv \nu_L$)

$$\begin{aligned} \mathcal{L}_M &= -\frac{1}{2} [m_{\nu_e} \bar{\nu}_e \nu_e + m_{\nu_\mu} \bar{\nu}_\mu \nu_\mu + m_{\nu_e \nu_\mu} (\bar{\nu}_e \nu_\mu + \bar{\nu}_\mu \nu_e)] \equiv \\ &\equiv -\frac{1}{2} (\bar{\nu}_e, \bar{\nu}_\mu) \begin{pmatrix} m_{\nu_e} & m_{\nu_e \nu_\mu} \\ m_{\nu_\mu \nu_e} & m_{\nu_\mu} \end{pmatrix} \begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix}. \end{aligned} \quad (23)$$

Массовый лагранжиан новых состояний, полученный диагонализацией этой матрицы при вращении на угле θ , имеет следующий вид (эти состояния являются именно теми же самыми состояниями слабых взаимодействий, рассмотренными выше):

$$\begin{aligned} \mathcal{L}_M &= -\frac{1}{2} (\bar{\nu}_e, \bar{\nu}_\mu) V^{-1} \begin{pmatrix} m_{\nu_1} & 0 \\ 0 & m_2 \end{pmatrix} V \begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix} = \\ &= \frac{1}{2} (\bar{\nu}_e, \bar{\nu}_\mu) \begin{pmatrix} m_1 \cos^2 \theta + m_2 \sin^2 \theta & (m_2 - m_1) \cos \theta \sin \theta \\ (m_2 - m_1) \cos \theta \sin \theta & m_1 \sin^2 \theta + m_2 \cos^2 \theta \end{pmatrix} \begin{pmatrix} \nu_e \\ \nu_\mu \end{pmatrix} = \\ &= \frac{1}{2} [m_1 \bar{\nu}_1 \nu_1 + m_2 \bar{\nu}_2 \nu_2], \end{aligned} \quad (24)$$

где ν_1, ν_2 — собственные состояния, а m_1, m_2 — собственные массы. Из выражении (23), (24) мы получим

$$\begin{aligned} \nu_1 &= \cos \theta \nu_e - \sin \theta \nu_\mu, \\ \nu_2 &= \sin \theta \nu_e + \cos \theta \nu_\mu, \end{aligned} \quad (25)$$

$$\begin{aligned} m_{\nu_e} &= m_1 \cos^2 \theta + m_2 \sin^2 \theta, \\ m_{\nu_\mu} &= m_1 \sin^2 \theta + m_2 \cos^2 \theta, \\ m_{\nu_e \nu_\mu} &= (m_2 - m_1) \cos \theta \sin \theta \end{aligned} \quad (26)$$

или

$$\begin{aligned} m_1 &= \frac{(m_{\nu_e} \cos^2 \theta - m_{\nu_\mu} \sin^2 \theta)}{(\cos^2 \theta - \sin^2 \theta)}, \\ m_2 &= \frac{(m_{\nu_e} \sin^2 \theta - m_{\nu_\mu} \cos^2 \theta)}{(\cos^2 \theta - \sin^2 \theta)}, \end{aligned} \quad (27)$$

где $\sin \theta$, $\cos \theta$ определяется выражением (19).

Тогда $\Delta m^2 = m_2^2 - m_1^2$ есть

$$\Delta m^2 = \frac{(m_{\nu_\mu}^2 \cos^4 \theta - m_{\nu_e} \sin^4 \theta)}{(\cos^2 \theta - \sin^2 \theta)^2}. \quad (28)$$

Выражения для временной эволюции ν_1 , ν_2 нейтрино (см. (25)–(27)) с массами m_1 и m_2 есть

$$\nu_1(t) = e^{-iE_1 t} \nu_1(0), \quad \nu_2(t) = e^{-iE_2 t} \nu_2(0), \quad (29)$$

где

$$E_k^2 = (p^2 + m_k^2), \quad k = 1, 2.$$

Если нейтрино распространяются без взаимодействий, тогда

$$\begin{aligned} \nu_e(t) &= \cos \theta e^{-iE_1 t} \nu_1(0) + \sin \theta e^{-iE_2 t} \nu_2(0), \\ \nu_\mu(t) &= -\sin \theta e^{-iE_1 t} \nu_1(0) + \cos \theta e^{-iE_2 t} \nu_2(0). \end{aligned} \quad (30)$$

Используя выражения (25) для ν_1 и ν_2 и подставляя это в (20), получаем следующее выражение:

$$\begin{aligned} \nu_e(t) &= \left[e^{-iE_1 t} \cos^2 \theta + e^{-iE_2 t} \sin^2 \theta \right] \nu_e(0) + \\ &\quad + \left[e^{-iE_1 t} - e^{-iE_2 t} \right] \sin \theta \cos \theta \nu_\mu(0), \\ \nu_\mu(t) &= \left[e^{-iE_1 t} \sin^2 \theta + e^{-iE_2 t} \cos^2 \theta \right] \nu_\mu(0) + \\ &\quad + \left[e^{-iE_1 t} - e^{-iE_2 t} \right] \sin \theta \cos \theta \nu_e(0). \end{aligned} \quad (31)$$

Вероятность того, что нейтрино ν_e , произведенный во время $t = 0$ перейдет в ν_μ за время t , дается квадратом абсолютной величины следующего скалярного произведения:

$$P(\nu_e \rightarrow \nu_\mu) = \left| (\nu_\mu(0) \cdot \nu_e(t)) \right|^2 = \frac{1}{2} \sin^2(2\theta) \left[1 - \cos((m_2^2 - m_1^2)/2p)t \right], \quad (32)$$

где предполагается, что $p \gg m_1, m_2$ и $E_k \simeq p + m_k^2/2p$.

Выражение (32) определяет вероятность осцилляции ароматов нейтрино. Угол θ (угол смешивания) характеризует величину смешивания. Вероятность $P(\nu_e \rightarrow \nu_\mu)$ является периодической функцией расстояний, где период определяется следующим выражением:

$$L_o = 2\pi \frac{2p}{|m_2^2 - m_1^2|}. \quad (33)$$

И вероятность $P(\nu_e \rightarrow \nu_e)$, что нейтрино ν_e , произведенный во время $t = 0$, сохраняется как ν_e нейтрино во время t , дается квадратом абсолютной величины амплитуды $\nu_e(0)$ в (31). Так как состояния в (31) являются нормализованными состояниями, то

$$P(\nu_e \rightarrow \nu_e) + P(\nu_e \rightarrow \nu_\mu) = 1. \quad (34)$$

Итак, мы видим, что осцилляции аромата, вызванные недиагональностью массовой массы нейтрино, нарушает закон сохранения лептонных чисел $-\ell_e$ и ℓ_μ . Однако в этом случае, как можно видеть из (34), полное число лептонов $\ell = \ell_e + \ell_\mu$ сохраняется.

Необходимо подчеркнуть, что нейтринные осцилляции в этой схеме (механизме) являются виртуальными, если массы нейтрино различаются, и реальными, если массы нейтрино равны и если эти осцилляции происходят в рамках соотношения неопределенностей.

1.2. Случай смешивания (осцилляции) трех нейтрино в схеме зарядовых смешиваний

Для случая трех нейтрино мы можем выбрать параметризацию матрицы смешивания V в форме, предложенной Майани [14]:

$$V = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_\gamma & s_\gamma \\ 0 & -s_\gamma & c_\gamma \end{pmatrix} \begin{pmatrix} c_\beta & 0 & s_\beta \\ 0 & 1 & 0 \\ -s_\beta & 0 & c_\beta \end{pmatrix} \begin{pmatrix} c_\theta & s_\theta & 0 \\ -s_\theta & c_\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (35)$$

$$\begin{aligned} c_{e\mu} &= \cos \theta, & s_{e\mu} &= \sin \theta, & c_{e\mu}^2 + s_{e\mu}^2 &= 1; \\ c_{e\tau} &= \cos \beta, & s_{e\tau} &= \sin \beta, & c_{e\tau}^2 + s_{e\tau}^2 &= 1; \\ c_{\mu\tau} &= \cos \gamma, & s_{\mu\tau} &= \sin \gamma, & c_{\mu\tau}^2 + s_{\mu\tau}^2 &= 1. \end{aligned} \quad (36)$$

Перейдем к рассмотрению случая трех типов нейтрино ν_e, ν_μ, ν_τ .

Для первого и второго семейств (при осцилляции ν_e, ν_μ нейтрино) мы получаем

$$\cos \theta = \cos \theta_{\nu_e \nu_\mu} = \frac{g_1}{\sqrt{g_1^2 + g_2^2}}, \quad (37)$$

$$\sin(2\theta) = \frac{2g_1 g_2}{g_1^2 + g_2^2}.$$

Тогда вероятность $\nu_e \rightarrow \nu_e$ перехода определяется следующим выражением

$$P(\nu_e \rightarrow \nu_e) = 1 - \sin^2(2\theta) \sin^2(\pi t(m_2^2 - m_1^2)/2p_{\nu_e}). \quad (38)$$

В случае $g_1 \cong g_2$

$$\sin \theta_{\nu_e \nu_\mu} \cong \cos \theta_{\nu_e \nu_\mu} \cong \frac{1}{\sqrt{2}}. \quad (39)$$

Для первого и третьего семейств (при осцилляции ν_e, ν_τ нейтрино) мы получаем

$$\cos \beta = \cos \beta_{\nu_e \nu_\tau} = \frac{g_1}{\sqrt{g_1^2 + g_3^2}}, \quad (40)$$

$$\sin(2\beta) = \frac{2g_1 g_3}{g_1^2 + g_3^2}.$$

Тогда вероятность перехода $\nu_e \rightarrow \nu_e$ дается следующим выражением

$$P(\nu_e \rightarrow \nu_e) = 1 - \sin^2(2\beta) \sin^2(\pi t(m_3^2 - m_1^2)/2p_{\nu_e}). \quad (41)$$

В случае, когда $g_1 \cong g_3$

$$\cos \beta_{\nu_e \nu_\tau} \cong \sin \beta_{\nu_e \nu_\tau} \cong \frac{1}{\sqrt{2}}. \quad (42)$$

Для второго и третьего семейств (при осцилляции ν_μ, ν_τ нейтрино) мы получаем

$$\cos \gamma = \cos \gamma_{\nu_\mu \nu_\tau} = \frac{g_2}{\sqrt{g_2^2 + g_3^2}}, \quad (43)$$

$$\sin(2\gamma) = \frac{2g_2 g_3}{g_2^2 + g_3^2}.$$

Тогда вероятность $\nu_\mu \rightarrow \nu_\mu$ перехода дается следующим выражением

$$P(\nu_\mu \rightarrow \nu_\mu) = 1 - \sin^2(2\gamma) \sin^2(\pi t(m_3^2 - m_2^2)/2p_{\nu_\mu}). \quad (44)$$

В случае, когда $g_2 \cong g_3$

$$\cos \gamma_{\nu_\mu \nu_\tau} \cong \sin \gamma_{\nu_\mu \nu_\tau} \cong \frac{1}{\sqrt{2}}. \quad (45)$$

Итак, смешивания (осцилляции) нейтрино появляются вследствие того, что рождаются нейтрино, являющиеся собственными состояниями слабых взаимодействий (ν_e, ν_μ, ν_τ нейтрино), а не ν_1, ν_2, ν_3 нейтрино, являющиеся собственными состояниями слабых взаимодействий, нарушающих лептонные числа. И тогда, при прохождении нейтрино через вакуум, они преобразовываются в суперпозиции ν_1, ν_2, ν_3 нейтрино, и через эти промежуточные состояния происходят осцилляции (переходы) между ν_e, ν_μ, ν_τ нейтрино.

Заключение

Необходимо отметить, что в физике все процессы идут через динамику. К сожалению, в стандартной теории осцилляций нейтрино, основанной на схеме (механизме) массовых смешиваний, динамика отсутствует. Вероятно, это признак того факта, что эта схема является неполным, т. е. эта схема требует физического обоснования.

В данной работе рассматриваются осцилляции нейтрино, генерированные смешиванием констант связей слабых взаимодействий (зарядов) [4–6], как это имеет место в модели векторной доминантности [15, 16] или в модели электрослабых взаимодействий [17–19] при смешивании векторных бозонов. Получены выражения для углов смешивания и длин осцилляций при трех нейтринных осцилляциях. Осцилляции нейтрино в этой схеме (механизме) являются реальными, если массы нейтрино равны, и виртуальными, если массы нейтрино не равны.

Литература

1. Pontecorvo B. M. // JETP. — Vol. 33. — 1957. — P. 549.
2. Maki Z. et al // Prog. Theor. Phys. — Vol. 28. — 1962. — P. 870.
3. Pontecorvo B. M. // JETP. — Vol. 53. — 1967. — P. 1717.
4. Beshtoev K. M. — JINR Commun. E2-2004-58. — Dubna, 2004.
5. Beshtoev K. M. — hep-ph/0406124. — 2004.
6. Beshtoev K. M. — hep-ph/0503202. — 2005.
7. Bilenky S. M., Pontecorvo B. M. // Phys. Rep. — Vol. 41. — 1978. — P. 225.
8. Boehm F., Vogel P. Physics of Massive Neutrinos. — Cambridge Univ. Press, 1978. — Pp. 27, 121.
9. Bilenky S. M., Petcov S. T. // Rev. Mod. Phys. — Vol. 59. — 1977. — P. 631.
10. Gribov V., Pontecorvo B. M. // Phys. Lett. B. — Vol. 28. — 1969. — P. 493.
11. Beshtoev K. M. — JINR Commun. E2-92-318. — Dubna, 1992.
12. Beshtoev K. M. // Proc. of 27th Intern. Cosmic Ray Conf. — Vol. 4. — Hamburg, Germany: 2001. — P. 1186.
13. Beshtoev K. M. // Proc. of 28th Intern. Cosmic Ray Conf. — Vol. 1. — Tsukuba, Japan: 2003. — P. 1503.
14. Maiani L. // Proc. Intern. Symp. on Lepton–Photon Interaction, DESY. — Hamburg, Germany: 1977.
15. Sakurai J. J. Currents and Mesons. — Univ. of Chicago Press, 1967.
16. Beshtoev K. M. — INR Preprint P-217. — 1981.
17. Glashow S. L. // Nucl. Phys. — Vol. 22. — 1961. — P. 579.
18. Weinberg S. // Phys. Rev. Lett. — Vol. 19. — 1967. — P. 1264.
19. Salam A. // Proc. of the 8th Nobel Symp. — Stockholm: Almqvist and Wiksell, 1968. — P. 367.

UDC 539.123, 539.123.06

Neutrino Oscillations in the Scheme of Charge (Couple Constant)

Kh. M. Beshtoev

*Laboratory of Particle Physics
Joint Institute for Nuclear Research*

In the standard theory of neutrino oscillations, a scheme of mass mixings is used, i.e. oscillation parameters are expressed in terms of mass matrix. In this work, neutrino oscillations generated by charge (the weak interaction couple constant) mixings are considered. Expressions for angle mixings and lengths of oscillations are obtained. The expressions of the probability for three-neutrino oscillations are given. Neutrino oscillations in this scheme (mechanism) are virtual if neutrino masses are not equal and real if neutrino masses are equal.

Спинорные поля в космологии Бианки VI

Б. Саха

*Лаборатория информационных технологий
Объединенный институт ядерных исследований
Россия, 141980, Дубна, Московская обл.*

Получены точные решения самосогласованной системы уравнений нелинейных спинорных и скалярных полей в рамках космологической модели типа Бианки VI, заполненной идеальной жидкостью с учетом Λ -члена. Изучена роль пространственной неоднородности. Показано, что введение положительного Λ (наиболее широко распространенный вид темной энергии) приводит к быстрому росту вселенной, в то время как отрицательный, соответствующий дополнительной гравитационной энергии, генерирует колебательные или непериодические моды расширения.

Роль спинорного поля в развитии Вселенной была изучена множеством авторов [1–7]. Основная цель этих работ заключалась в нахождении регулярных решений соответствующих полевых уравнений. В качестве гравитационного поля в этих случаях рассматривалась анизотропная космологическая модель типа Бианки-I (BI). Было получено, что введение нелинейного спинорного поля в некоторых специальных случаях приводит к быстрому росту Вселенной. В свете этих исследований многие авторы [8–10] верят, что спинорное поле является одним из возможных кандидатов для объяснения ускоренного расширения Вселенной, которое наблюдается в последнее время.

Хотя в качестве анизотропной космологической модели космологи в основном рассматривают Бианки типа-I пространство-время, есть все еще несколько других моделей, которые описывают анизотропное пространство-время и производят специфический интерес среди физиков [11–18]. В недавних работах [19–21] мы изучили самосогласованную систему нелинейного спинорного и анизотропного неоднородного гравитационного полей, чтобы разъяснить роль нелинейности спинорного поля и неоднородности пространства-времени в формировании вселенной без сингулярностей. В качестве анизотропного пространства-времени мы выбрали Бианки типа-VI (BVI) модель, так как подходящий выбор его параметров приводит к появлению некоторых других Бианки моделей, включая вселенные BI и FRW. Отметим, это в отличие от BI, пространство-время BVI неоднородно. Включение неоднородность в поле тяготения значительно усложняет поиск точного решения системы. Была также исследована самосогласованная система спинорного, скалярного и гравитационного полей типа Бианки-VI в присутствии идеальной жидкости и космологической постоянной и изучена роль соответствующих материальных полей в эволюции вселенной. Цель статьи в подведении итогов сделанных работ.

Как было указано, эти работы проводились в течение нескольких лет и основные результаты были представлены на научных семинарах и конференциях и опубликованы в разных журналах. Поэтому мы опустим детали и уделим наше внимание на основному уравнению (уравнение для нахождения объем-фактора τ), так как все остальные, включая спинорное и скалярное поля, а также метрические функции, выражаются через него. Более того, другие физически измеряемые величины и инварианты пространства-времени также записываются в терминах τ . Отметим, что все эти величины обратно пропорциональны к τ (точнее $1/\tau^r$), где r — некоторые положительные числа.

Пространство-время Бианки-VI задается в виде

$$ds^2 = dt^2 - a^2 e^{-2mz} dx^2 - b^2 e^{2nz} dy^2 - c^2 dz^2, \quad (1)$$

где a , b , c зависят от времени t и m , n — некоторые постоянные, описывающие пространственную неоднородность.

Мы рассматриваем случай, когда полевые функции зависят только от t . В качестве нелинейности спинорного поля рассматривалась как степенная функция, так и тригонометрическая функция от инвариантов билинейных спинорных форм. Так как решения, полученные во всех этих случаях, качественно не отличаются, мы рассмотрим случай со степенной нелинейностью. Кроме того, мы также изучим случай, когда нелинейный член в лагранжиане является функцией от $S = \bar{\psi}\psi$. Подробно эти и другие результаты можно найти в работах [19].

Из уравнения Эйнштейна находим выражения для метрических функций

$$a = [\mathcal{N}\tau^n c^{m-2n}]^{1/(m+n)}, \quad b = [\tau^m c^{n-2m}/\mathcal{N}]^{1/(m+n)}, \quad (2)$$

где мы определили

$$\tau = abc. \quad (3)$$

Уравнение для нахождения τ в данном случае имеет вид

$$\frac{\ddot{\tau}}{\tau} = 2\frac{m^2 - mn + n^2}{c^2} + \frac{3\kappa}{2} \left[\frac{M}{\tau} + F_1 + \frac{1 - \zeta}{\tau^{1+\zeta}} \right] + 3\Lambda. \quad (4)$$

Здесь M — масса спинорного поля, ζ — параметр, описывающий идеальную жидкость. F_1 связан с нелинейностью спинорного поля. Если нелинейность спинорного поля задается в виде $F = \lambda S^\eta = \lambda/\tau^\eta$, то в случае самодействия имеем $F_1 = \lambda(\eta-2)/\tau^\eta$, а в случае индуцированной нелинейности имеем $F_1 = \lambda\eta\tau^{\eta-2}/(\lambda + \tau^\eta)^2$. Как видно, уравнение (4) содержит функцию c в явном виде. Для разрешения этой проблемы мы предположим, что c является функцией от τ . Тогда уравнения (4) можно будет записать в виде

$$\ddot{\tau} = \mathcal{F}(\tau, q). \quad (5)$$

Уравнение (5) допускает следующий первый интеграл

$$\dot{\tau} = \sqrt{2[E - U(\tau, q)]}, \quad (6)$$

где $U(\tau, q) = -\int \mathcal{F}(\tau, q)d\tau$. С механической точки зрения уравнение (5) можно интерпретировать как уравнение движения одной частицы с единичной массой под действием силы $\mathcal{F}(\tau, q)$. В (6) постоянная интегрирования E может быть рассмотрена как уровень энергии, а $U(\tau, q)$ потенциал. В дальнейшем мы решим уравнение (5) при (i) $c = \tau$ и (ii) $c = \sqrt{\tau}$.

Рассмотрим случай при $c = \tau$. Для потенциала имеем

$$\mathcal{U}(\tau, q) = -\left\{ 4(m^2 - mn + n^2)\ln \tau + 3\kappa[M\tau - \lambda/\tau^{\eta-2} + \tau^{1-\zeta}] + 3\Lambda\tau^2 \right\} \quad (7)$$

в случае самодействия и

$$\mathcal{U}(\tau, q) = -\left\{ 4(m^2 - mn + n^2)\ln \tau + 3\kappa[M\tau - \lambda/(\tau^\eta + \lambda) + \tau^{1-\zeta}] + 3\Lambda\tau^2 \right\} \quad (8)$$

в случае индуцированной нелинейности. Как видно из рис. 1, а также из выражения (7), при $\eta < 2$ возникает потенциальный барьер при больших значениях τ . С другой стороны, $\ln \tau$ порождает бесконечно высокий барьер при $\tau = 0$. Это приводит к появлению осциллирующих решений независимо от значения Λ . В случае $\eta > 2$, также как при индуцированной нелинейности имеем осциллирующие решения только при $\Lambda < 0$, а при $\Lambda \geq 0$ имеем модель расширяющейся Вселенной. Соответствующий потенциал представлен на рис. 2. Заметим, что в случае индуцированной нелинейности потенциал имеет примерно тот же вид что показан на рис. 2. Во всех этих случаях решение всюду регулярно, т.е. нигде и никогда τ не может быть тривиальным.

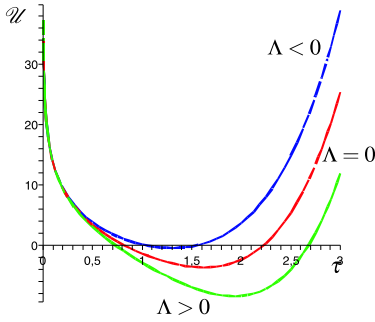


Рис. 1. Вид потенциала $\mathcal{U}$, заданного выражением (7) при $\eta = -3$

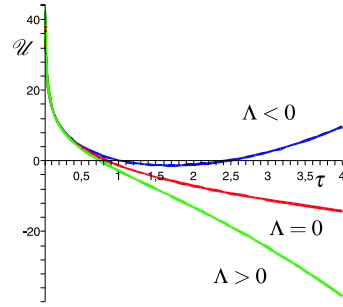


Рис. 2. Вид потенциала $\mathcal{U}$, заданного выражением (7) при $\eta = 3$

Теперь рассмотрим случай при $c = \sqrt{\tau}$. Для потенциала имеем

$$\mathcal{U}(\tau, q) = -\left\{4(m^2 - mn + n^2)\tau + 3\kappa[M\tau - \lambda/\tau^{\eta-2} + \tau^{1-\zeta}] + 3\Lambda\tau^2\right\} \quad (9)$$

в случае самодействия и

$$\mathcal{U}(\tau, q) = -\left\{4(m^2 - mn + n^2)\tau + 3\kappa[M\tau - \lambda/(\tau^\eta + \lambda) + \tau^{1-\zeta}] + 3\Lambda\tau^2\right\}. \quad (10)$$

При $\tau = 0$ потенциальный барьер в этом случае имеется только при $\eta > 2$. А при $\eta < 2$ возникает потенциальный барьер при больших значениях τ . При $\eta < 2$ имеется колебательное или непериодическое решение в зависимости от выбора уровня энергии E . В случае $\eta > 2$ решения всегда регулярные. При $\Lambda < 0$ оно колебательное, а при $\Lambda \geq 0$ имеем модель расширяющейся Вселенной, при этом положительная космологическая постоянная приводит к быстрому росту вселенной. При индуцированной нелинейности поведение τ совпадает с теми, что имеем при самодействии при $\eta > 2$, за исключением того, что решения в этом случае могут быть также сингулярными.

На рис. 3 и 4 представлены поведение τ при разных значениях Λ . На рис. 3 показан случай с ускоренными расширениями, а на рис. 4 представлены колебательное и непериодическое расширение Вселенной.

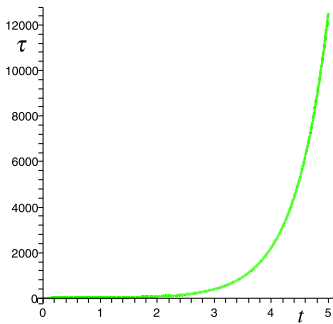


Рис. 3. Поведение τ при положительном Λ в случае индуцированной нелинейности

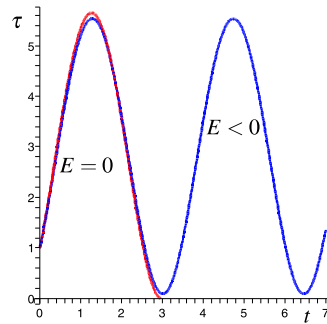


Рис. 4. Поведения τ при отрицательном Λ с разными значениями E

Литература

1. *Rybakov Y. P., Saha B., Shikin G. N.* // Commun. Theor. Phys. — Vol. 3. — 1994. — P. 199.
2. *Saha B., Shikin G. N.* // J. Math. Phys. — Vol. 38. — 1997. — P. 5305.
3. *Saha B., Shikin G. N.* // Gen. Relativ. Gravit. — Vol. 29. — 1997. — P. 1099.
4. *Saha B.* // Phys. Rev. D. — Vol. 64. — 2001.
5. *Saha B.* // Mod. Phys. Lett. A. — Vol. 16. — 2001. — P. 1287.
6. *Saha B., Boyadjiev T.* // Phys. Rev. D. — Vol. 69. — 2004. — http://www.jinr.ru/~bijan/my_papers/PRD24010.pdf.
7. *Armendáriz-Picón C., Greene P. B.* // Gen. Relativ. Gravit. — Vol. 35. — 2003. — P. 1637.
8. *Ribas M. O., Devecchi F. P., Kremer G. M.* // Phys. Rev. D. — Vol. 72. — 2005. — P. 123502.
9. *Saha B.* // Phys. Rev. D. — Vol. 74. — 2006. — P. 124030.
10. *Saha B.* // Gravitation & Cosmology. — Vol. 12, No 2–3 (46–47). — 2006. — P. 215.
11. *Singh T., Agrawal A. K.* // Int. J. Theor. Phys. — Vol. 32. — 1993. — P. 1041.
12. *Weaver M.* // Class. Quantum Grav. — Vol. 17. — 2000. — P. 421.
13. *Ibáñez J., van der Hoogen R., Coley A. A.* // Phys. Rev. D. — Vol. 51. — 1995. — P. 928.
14. *Socorro J., Medina E. R.* // Phys. Rev. D. — Vol. 61. — 2000. — P. 087702.
15. *Berger B. K.* // Class. Quantum Grav. — Vol. 13. — 1996. — P. 1273.
16. *Apostolopoulos P. A.* — e-print gr-qc/0310033.
17. *Christodoulakis T., Kofinas G., Papadopoulos G. O.* // Phys. Lett. B. — Vol. 514. — 2001. — P. 149.
18. *Christodoulakis T., Papadopoulos G. O.* — e-print gr-qc/0109058.
19. *Saha B.* // Phys. Rev. D. — Vol. 69. — 2004. — P. 124006.
20. *Saha B.* Nonlinear Spinor Field in Anisotropic Universes. — e-print gr-qc/0207025. — <http://xxx.lanl.gov/abs/gr-qc/0207025>.
21. *Saha B.* Interacting Spinor and Scalar Fields in Bianchi Cosmology. — e-print gr-qc/0701059. — talk given at the Conf. in memory of Prof. Ya. P. Terletski. Talk given at the Conf. in memory of Prof. Ya. P. Terletski.

Spinor Fields in Bianchi Type-VI Cosmology

B. Saha

*Laboratory of Information Technologies
Joint Institute for Nuclear Research
Dubna, Moscow reg., 141980, Russia*

Exact solutions to the self-consistent system of equations of interacting spinor and scalar fields are considered within the scope of Bianchi type VI cosmological model filled with a perfect fluid in account of a Λ -term have been obtained. The role of the spatial inhomogeneity has been studied. It has been shown that the introduction of a positive Λ , the most widespread kind of dark energy, leads to the rapid growth of the universe, while the negative one, corresponding to an additional gravitational energy gives rise to an oscillatory or non-periodic mode of expansion.

Спинорные поля в плоско-симметричном пространстве-времени

Б. Саха*, Г. Н. Шикин†

* Лаборатория информационных технологий
Объединенный институт ядерных исследований
Россия, 141980, Дубна, Московская обл.

† Кафедра теоретической физики
Российский университет дружбы народов
Россия, 117198, Москва, ул. Миклухо-Маклая, 6

В рамках плоско-симметричной космологической модели рассматривается система нелинейных спинорных и нелинейных скалярных полей с минимальной связью. Показано, что гравитационное поле играет решающую роль в формировании солитоноподобных решений, т. е. решений с ограниченными полной энергией, спином и зарядом.

В последние годы космологические модели, проявляющие плоскую симметрию, притягивают внимание многих специалистов. В настоящее время Вселенная сферически симметрична, и распределение материи в ней изотропно и однородно. Но в ранней стадии эволюции у нее не могло быть такой гладкой картины. Ближе к сингулярности большого взрыва, ни предположение о сферической симметрии, ни предположение об изотропности не могут быть строго выполнены. Именно поэтому многие специалисты рассматривают плоскую симметрию, так как она менее строга чем сферическая симметрия и обеспечивает поле для изучения неоднородности ранней Вселенной. Заметим, что неоднородные космологические модели играют важную роль для понимания некоторых существенных черт, таких как формирование галактики в ранней стадии эволюции и процесса гомогенизации Вселенной.

С другой стороны, нелинейность феномена остается одной из самых популярных тем исследования в течение последних лет. Хотя нелинейные классические поля не получили общего одобрения, нелинейное обобщение классического поля является одним из возможных путей для преодоления тех трудностей, которые испытывает теория, рассматривающая элементарные частицы как математические точки. Нахождение и исследование свойств регулярных локализованных решений нелинейных классических полевых уравнений (солитон или частицеподобных решений) связано с надеждой создать свободную от расходимостей теорию элементарных частиц, которая могла бы описывать сложную пространственную структуру частиц, наблюдаемую экспериментально [1]. Уравнение гравитационного поля является нелинейным по природе, и само поле является универсальным и незранированным. Эти свойства приводят к появлению определенного физического интереса к собственному гравитационному полю. Этот проект был реализован во множестве работ, где авторы подробно изучали скалярное и/или электромагнитное поля в рамках сферического и цилиндрического пространства-времени [2–8]. Однако в работах, связанных с солитоноподобным решением нелинейных уравнений полей, часто игнорируется собственное гравитационное поле. Под собственным гравитационным полем мы подразумеваем следующее: материальное поле ведет себя как источник гравитационного поля. А гравитационное поле со своей стороны эффективно влияет на конфигурацию материального поля. В этом смысле *собственное* гравитационное поле является решением самосогласованной системы материального и гравитационного полей.

В последние годы изучением космологических моделей, проявляющих плоскую симметрию, занимались множество авторов [9–18]. В это же время нами была изучена система нелинейных спинорного и скалярного полей с минимальной связью в плоско-симметричном фоне [19–22]. Целью настоящей заметки является обобщение всех этих результатов.

Лагранжиан спинорного и скалярного полей с минимальной связью выбран в виде

$$\mathcal{L} = \frac{i}{2} [\bar{\psi} \gamma^\mu \nabla_\mu \psi - \nabla_\mu \bar{\psi} \gamma^\mu \psi] - m \bar{\psi} \psi + F + \Psi(\Omega), \quad \Omega = \varphi_{,\alpha} \varphi^{,\alpha}. \quad (1)$$

Здесь F некоторые произвольные функции от инвариантов билинейных спинорных форм, а именно $F = F(I, J)$, где $I = S^2 = (\bar{\psi} \psi)^2$ и $J = P^2 = (i \bar{\psi} \gamma^5 \psi)^2$. Гравитационное поле мы выберем в виде

$$ds^2 = e^{2\rho} dt^2 - e^{2\alpha} dx^2 - e^{2\beta} (dy^2 + dz^2), \quad (2)$$

где метрические функции ρ, α, β зависят только от пространственной координаты x и подчиняются условию

$$\alpha = 2\beta + \rho. \quad (3)$$

Рассматриваемые спинорное и скалярное поля тоже зависят только от x . В силу этого из уравнения Эйнштейна получим

$$\beta(x) = \frac{1}{3}(\alpha(x) + BX), \quad \gamma(x) = \frac{1}{3}(\alpha(x) - 2Bx), \quad (4)$$

где некоторое B постоянное. Для α из уравнения Эйнштейна находим

$$\alpha'^2 - B^2 = -3\kappa e^{2\alpha} \left[mS - F(I, J) + 2\Upsilon \frac{d\Psi}{d\Upsilon} - \Psi(\Upsilon) \right]. \quad (5)$$

Правая часть уравнения (5) зависит только от α [20–22], следовательно его решение можно получить в квадратурах.

Уравнение (5) было подробно изучено при различных выборах правой части: линейное спинорное и линейное скалярное поля; нелинейное спинорное и линейное скалярное поля; нелинейное скалярное поле в отсутствии спинорного поля и, наконец, нелинейное спинорное и нелинейное скалярное поля. Мы подробно изучили это уравнение, когда нелинейный член в лагранжиане задавался в виде степенной функции от I, J .

Показано, что в случае линейного спинорного и линейного скалярного полей с минимальной связью заряд и спин спинорного поля являются ограниченной величиной, а плотность энергии системы определяется только за счет скалярного поля.

В случае нелинейного спинорного и линейного скалярного полей для S , которая связана с α как

$$S = C \exp -\alpha(x),$$

при некоторых специальных случаях допускает кинк-образное решение (рис. 1). На самом деле это возможно только нулевой массе спинорного поля, так как в противном случае ни при каких значениях x , $S(x)$ не может быть тривиальной. Заметим, что в рассматриваемом случае скалярное поле является линейным и безмассовым. Поскольку в отсутствии спинорного поля плотность энергии скалярного поля не локализована и полная энергия не ограничена, то в этом случае свойства полевых конфигураций определяются свойствами нелинейного спинорного поля.

В отсутствии спинорного поля в качестве нелинейного скалярного поля мы рассмотрели теорию Борна–Инфельда с лагранжианом

$$\Psi(\Upsilon) = -\frac{1}{\sigma} (1 - \sqrt{1 + \sigma \Upsilon}), \quad (6)$$

где σ есть параметр нелинейности. Показано, что в этом случае плотность энергии скалярного поля не локализована, однако в некоторых условиях в единицу инвариантного объема она локализована, и полная энергия скалярного поля ограничена. В этом случае Υ имеет кинк-образный вид.

Таким образом мы рассмотрели систему нелинейного спинорного и нелинейного скалярного полей с минимальной связью в рамках плоско-симметричной

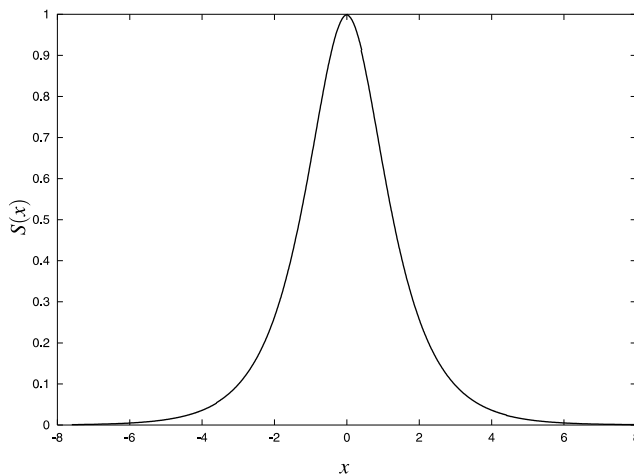


Рис. 1. Вид $S(x)$ в случае нелинейного спинорного поля с $n = 4$. Как видно, в рассматриваемом случае функция $S(x)$ имеет максимум в точке $x = 0$

космологической модели. Показано, что спинорное поле более чувствительно к гравитационному полю чем скалярное поле. Как ожидалось, плотность энергии и полная энергия системы линейного спинорного и линейного скалярного полей не ограничены, и система не имеет реальную физическую бесконечность. Введение нелинейного спинорного поля устраняет этот недостаток, и в этом случае система допускает конфигурацию с конечной плотностью и ограниченной полной энергией. В этом случае система становится наблюдаемой. Мы также установили, что свойства системы нелинейного спинорного и скалярного полей определяется той частью гравитационного поля, которая порождается нелинейным спинорным полем. Было показано, что вместе с нелинейным спинорным полем гравитационное поле тоже играет важную роль в формировании полевых конфигураций с ограниченной полной энергией, спином и зарядом. В общем было показано, что выбор нелинейности спинорного поля может устранять вклад скалярного поля в метрические функции, но при этом его вклад в полную энергию остается нетронутым.

Литература

1. *Thirring J. K., Skyrme T. H. R.* // Nuclear Physics. — Vol. 31. — 1962. — P. 550.
2. *Bronnikov K. A., Shikin G. N.* // Gravitation & Cosmology. — Vol. 7, No 3(27). — 2001. — P. 231.
3. *Бронников К. А., Шикин Г. Н.* Итоги науки и техники, Т. 2. Гравитация и космология. — М.: ВИНТИ, 1991. — С. 4.
4. *Рыбаков Ю. П., Саха Б., Шикин Г. Н.* // Известия ВУЗов. Физика. — Т. 35, № 10. — 1992. — С. 112.
5. *Rybakov Y. P., Saha B., Shikin G. N.* // Commun. Theor. Phys. — Vol. 3, No 1. — 1994. — P. 67.
6. *Rybakov Y. P., Saha B., Shikin G. N.* // Int. J. Theor. Phys. — Vol. 36. — 1997. — P. 1475.
7. *Rybakov Y. P., Saha B., Shikin G. N.* // Gravitation & Cosmology. — Vol. 4, No 2(14). — 1998. — P. 114.
8. *Saha B.* // Int. J. Mod. Phys. A. — Vol. 15. — 2000. — P. 1481.
9. *Rendall A. D.* // General Relat. Grav. — Vol. 27. — 1995. — P. 213.
10. *Taruya A., Nambu Y.* // Progress Theor. Phys. — Vol. 95. — 1996. — P. 295.
11. *Zhuravlev V. M., Chervon S. V., Shabalkin D. Y.* // Gravitation & Cosmology. — Vol. 3, No 4(14). — 1997. — P. 312.
12. *Silva M. F. A., Wang A.* // Phys. Lett. A. — Vol. 244. — 1998. — P. 462.
13. *Ori A.* // Phys. Rev. D. — Vol. 57. — 1998. — P. 4745.

14. *Chervon S. V., Shabalkin D. Y.* // Gravitation & Cosmology. — Vol. 6, No 1(21). — 2000. — P. 41.
15. *Anguige K.* // Class. Quant. Grav. — Vol. 17. — 2000. — P. 2117.
16. *Nouri-Zonoz M., Tavanfar A. R.* // Class. Quant. Grav. — Vol. 18. — 2001. — P. 4293.
17. *Pradhan A., Pandey H. R.* // Int. J. Mod. Phys. D. — Vol. 12. — 2003. — P. 941.
18. *Yazadjiev S. S.* // Class. Quant. Grav. — Vol. 20. — 2003. — P. 3365.
19. *Adomou A., Shikin G. N.* // Gravitation & Cosmology. — Vol. 4, No 2(14). — 1998. — P. 107.
20. *Saha B., Shikin G. N.* // J. Theor. Math. Comp. Phys. — Vol. 5. — 2002. — P. 54.
21. *Saha B., Shikin G. N.* // Czechoslovak J. Phys. — Vol. 54. — 2004. — P. 597.
22. *Saha B., Shikin G. N.* // Int. J. Theor. Phys. — Vol. 44. — 2005. — P. 1459.

Spinor Fields in Plane-Symmetric Space-Time

B. Saha^{*}, G. N. Shikin[†]

^{*} *Laboratory of Information Technologies
Joint Institute for Nuclear Research
Dubna, Moscow reg., 141980, Russia*

[†] *Department of Theoretical Physics
Peoples' Friendship University of Russia
6, Miklukho-Maklaya str., Moscow, 117198, Russia*

Within the framework of a plane-symmetric cosmological model system of a minimally coupled nonlinear spinor and scalar fields has been considered. It has been shown that the gravitational field plays crucial role in the formation of soliton-like solutions, i.e., solutions with limited total energy, spin and charge.

УДК 539.17.01

Упругое рассеяние реакторных антинейтрино на электронах и ограничение на массу дополнительного нейтрального векторного бозона

Ю. П. Богданов, И. Ю. Дидковская

*Кафедра физики**Российский университет дружбы народов**Россия, 117198, Москва, ул. Миклухо-Маклая, 6*

В статье изучено упругое рассеяние антинейтрино на электроны в рамках стандартной модели электрослабого взаимодействия. Получены ограничения на параметры и углы смешивания модели, а также ограничения на массу дополнительного нейтрального векторного бозона, существование которого допускается самой моделью.

Введение

Как известно, стандартная модель электрослабого взаимодействия Глешоу–Вайнберга–Салама (ГВС), основанная на калибровочной группе $SU(2) \otimes U(1)$, позволяет описывать имеющиеся экспериментальные данные с помощью следующих значений параметра смешивания $S_W \equiv \sin^2 \theta_w$ и масс промежуточных $W^\pm$ и Z^0 бозонов [1]

$$S_W = 0,226 \pm 0,004; \quad M_W = (81,8 \pm 1,5) \text{ ГэВ}; \quad M_Z = (92,6 \pm 1,7) \text{ ГэВ}. \quad (1)$$

Несмотря на высокую точность среднемирового значения S_W , в (1) существует заметное расхождение в величине этого параметра, измеренной в различных экспериментах. В частности, измерение электрослабых эффектов в процессе $e^+e^- \rightarrow e^+e^-$ на ускорителе PEP [2] дало $S_W = 0,24^{+0,18}_{-0,17}$. Путем измерения сечений $\nu_\mu e^-$ - и $\bar{\nu}_\mu e^-$ -рассеяния группой CHARM [3] было получено $S_W = 0,215 \pm 0,032$, тогда как измерение сечения рассеяния реакторных антинейтрино $\bar{\nu}_\mu e \rightarrow \bar{\nu}_\mu e$ [4] дало $S_W = 0,28 \pm 0,08$.

Поиски возможного отклонения от стандартной модели, а также попытки построения моделей с более широкой электрослабой группой предпринимались давно. Теоретической предпосылкой тому является необходимость объединения наряду со слабым и электромагнитным также сильного и гравитационного взаимодействий (модели большого объединения и супергравитация). Обнадеживающие результаты в этом направлении получены в последнее время в рамках так называемых суперструнных теорий [5], основанных на группе $E_8 \otimes E_8$. В качестве эффективной низкоэнергетической группы в этом подходе может выступать группа E_6 [6]. В зависимости от способа понимания ранга группы E_6 электрослабый сектор включает один или два дополнительных нейтральных промежуточных векторных бозона, которые могут иметь массы, не сильно отличающиеся от массы уже открытого Z^0 -бозона.

Анализу возможных проявлений дополнительного Z' -бозона при высоких энергиях посвящено значительное количество работ (см., например, [7, 8]).

В настоящей работе влияние дополнительного промежуточного векторного бозона на процесс рассеяния

$$\bar{\nu}_e + e^- \rightarrow \bar{\nu}_e + e^- \quad (2)$$

исследуется в области энергий реакторных антинейтрино $E_{\bar{\nu}} < 10$ МэВ.

1. Амплитуда и сечение процесса

Мы рассмотрим здесь влияние одного дополнительного нейтрального векторного бозона на характеристики процесса (2).

Лагранжиан взаимодействия электронного нейтрино (антинейтрино) с электроном дается выражением

$$\mathcal{L} = \mathcal{L}_{\text{ЗТ}} + \mathcal{L}_{\text{НТ}}, \quad (3)$$

где лагранжиан взаимодействия заряженных токов

$$\mathcal{L}_{\text{ЗТ}} = \frac{g_1}{2\sqrt{2}} \left[\bar{\nu}_e \widehat{W}^- (1 + \gamma_5) e + \bar{e} \widehat{W}^+ (1 + \gamma_5) \nu_e \right], \quad (4)$$

а лагранжиан взаимодействия нейтральных токов

$$\mathcal{L}_{\text{НТ}} = \sum_{\alpha=0,1} \frac{g}{4} \left\{ v_\alpha^\nu \bar{\nu}_e \widehat{Z}_\alpha (1 + \gamma_5) \nu_e + 2 \bar{e} \widehat{Z}_\alpha (v_\alpha^e + a_\alpha^e \gamma^5) e \right\}. \quad (5)$$

Здесь

$$g_1^2 = g_{SU(2)}^2 = \frac{G_F}{\sqrt{2}} \cdot 8M_W^2; \quad g = \frac{e}{S_W C_W}; \quad C_W \equiv \cos^2 \theta_W,$$

при $\alpha = 0$ соответствует известному Z^0 -бозону, а $\alpha = 1$ — дополнительному Z' -бозону. Константы v_α^ν , v_α^e и a_α^e зависят от углов

$$\theta_k, \beta_k \quad (k = 1, 2, 3; C_k \equiv \cos \theta_k; S_k \equiv \sin \theta_k),$$

которыми определяется характер смешивания генераторов и фермионных мультиплетов низкоэнергетической калибровочной группы

$$SU(3)_C \otimes SU(2)_L \otimes U(1) \otimes U'(1) \otimes U''(1).$$

Угол θ_1 фиксирует смешивание гиперзаряда Y и комбинации $\widehat{Y} = C_2 Y' - S_2 Y''$. В случае одного дополнительного бозона предполагается, что ортогональная комбинация $S_2 Y' + C_2 Y''$ связана с генератором, нарушаемым при более высоких энергиях, так что соответствующий ток устраняется из низкоэнергетической теории. Угол θ_3 входит в матрицу, диагонализующую массовую матрицу калибровочных бозонов.

В приближении малых передаваемых импульсов $\|q^2\| \ll M_W^2 < M_Z^2$ из (4) и (5) получим следующее выражение для амплитуды реакции (2)

$$M = M_{\text{ЗТ}} + M_{\text{НТ}} = -\frac{G_F}{\sqrt{2}} \bar{\nu}_e \gamma^\mu (1 + \gamma^5) \nu_e \bar{e} \gamma_\mu (C_V + C_A \gamma^5) e. \quad (6)$$

Эффективные векторная и аксиальные константы C_V и C_A связаны с соответствующими константами g_V и g_A для процессов с участием только нейтральных токов ($\nu_\mu e^- \rightarrow \nu_\mu e^-$; $\bar{\nu}_\mu e^- \rightarrow \bar{\nu}_\mu e^-$) соотношениями

$$C_V = 1 + g_V; \quad C_A = 1 + g_A. \quad (7)$$

для g_V и g_A в нашем случае можно записать

$$g_V = g_V^0 + \eta^{-2} v_1^\nu v_1^e; \quad g_A = g_A^0 + \eta^{-2} v_1^\nu a_1^e, \quad (8)$$

где

$$g_V^0 = -\frac{1}{2} + 2S_W; \quad g_A^0 = -\frac{1}{2}$$

— значения векторной и аксиальной констант нейтрального электронного тока в стандартной модели ГВС, а $\eta = M_{Z'}/M_Z$.

Из (6) следует выражение для дифференциального сечения процесса (2)

$$\frac{d\sigma}{dE_k}(\bar{\nu}_e e^-) = \frac{G_F^2 m_e}{2\pi} \left[g_-^2 + g_+^2 \left(1 - \frac{E_k}{E_\nu} \right)^2 - g_- g_+ \frac{m_e E_k}{E_\nu^2} \right], \quad (9)$$

в котором $g_\pm = C_v \pm C_A$, а E_k — кинетическая энергия электронов отдачи [9].

Ниже мы ограничимся анализом вклада дополнительного Z' -бозона в (9) в достаточно простом случае: $\theta_3 = 0$ (это эквивалентно справедливости массового соотношения стандартной модели $M_W = C_W M_Z$), а также $\theta_2 = 0$ (Y' и Y'' не смешиваются). При сделанных предположениях имеют место следующие соотношения

$$g_+ = 1 + 2S_W + S_W^2 \delta; \quad g_- = 2S_W + S_W^2 \delta, \quad (10)$$

из которых следует, что отклонение от величин $g_\pm$ стандартной модели определяется только одним эффективным параметром

$$\delta = \frac{1}{9\chi^2 \eta^2} [C_1^2 + \lambda^2 S_1^2 - 3C_1 S_1 (\lambda^2 - 1)]^2, \quad (11)$$

здесь $\lambda = g_Y/g_Y'$ и всегда может быть выбрано $\lambda \geq 1$.

Заметим, что в рассматриваемом случае, когда $\theta_2 = \theta_3 = 0$, в силу (11) вклад дополнительного бозона в параметры (10) неотрицателен.

2. Сравнение с экспериментом

Сечение упругого рассеяния реакторных антинейтрино было измерено [4]

$$\begin{aligned} \langle \sigma \rangle_1 &= (0,87 \pm 0,25) \langle \sigma^{V-A} \rangle_1; \\ \langle \sigma \rangle_2 &= (1,7 \pm 0,44) \langle \sigma^{V-A} \rangle_2, \end{aligned}$$

где угловые скобки означают усреднение по спектру реакторных антинейтрино и интегрирование по соответствующему интервалу E_k — кинетической энергии электронов отдачи

$$\begin{aligned} \text{1-й интервал: } & 1,5 \text{ МэВ} \leq E_k < 3,0 \text{ МэВ}; \\ \text{2-й интервал: } & 3,0 \text{ МэВ} \leq E_k < 4,5 \text{ МэВ}. \end{aligned}$$

После усреднения и интегрирования теоретического сечения (9) в каждом из данных интервалов возникает уравнение вида ($k = 1, 2$)

$$a_{11}^k g_-^2 + a_{22}^k g_+^2 + 2a_{12}^k g_- g_+ = \langle \sigma \rangle_k.$$

Относительно параметров S_W и δ это алгебраическое уравнение порядка 4 и 2 соответственно. Условие насыщения экспериментальных ошибок [10] дает в плоскости (S_W, δ) две кривые, полосы между которыми представляют области допустимых значений S_W и δ для каждого из интервалов [10]. Пересечение этих полос с учетом естественного условия $S_W, \delta \geq 0$ дает область значений S_W и δ , которые совместимы с результатом данным в [10].

Ниже нами использованы два несколько отличающихся друг от друга энергетических спектра для потока реакторных антинейтрино, рассчитанных в [10, 11].

В рамках стандартной модели ($\delta = 0$) расчет со спектром [11] дает следующие значения параметра Вайнберга

$$S_W^{(1)} = 0,26 \pm 0,06; \quad S_W^{(2)} = 0,30 \pm 0,05; \quad \bar{S}_W = 0,28 \pm 0,08,$$

а со спектром [10]:

$$S_W^{(1)} = 0,25 \pm 0,06; \quad S_W^{(2)} = 0,29 \pm 0,05; \quad \bar{S}_W = 0,27 \pm 0,08,$$

где через $\bar{S}_W$ обозначено среднее двух значений $S_W^{(1)}$ и $S_W^{(2)}$.

С помощью соотношения (11) эти ограничения могут быть трансформированы в ограничения на величину массы дополнительного Z' -бозона. В общем случае ($\lambda > 1$) ограничение на η зависит от угла η_1 , однако при $\lambda = 1$ эта зависимость исчезает. Очевидно, что ограничения сверху $M_{Z'} < 0,52M_Z$ или $M_{Z'} < 0,46M_Z$ могут свидетельствовать о том, что результаты эксперимента [4] плохо согласуются со среднемировым значением S_W^0 .

Отметим, что нижние границы для массы дополнительного Z' -бозона, полученные в области высоких энергий, из данных по pp -взаимодействиям составляют $M_{Z'} \gtrsim 140$ Гэв [7] и $M_{Z'} > 100$ Гэв [8], а из данных по $\nu_\mu e^-$ -рассеянию — $M_{Z'} \gtrsim 200$ Гэв.

Заключение

Полученные здесь результаты свидетельствуют о необходимости повышения точности экспериментальных данных по упругому рассеянию реакторных антинейтрино, что может позволить в принципе получить не только ограничения на массу дополнительного промежуточного векторного Z' -бозона, но исследовать их зависимость от углов смешивания и других параметров модели E_6 .

Литература

1. Review of Particle Properties // Phys. Lett. — Vol. 170B. — 1986.
2. Fernandez E. Search for single electron from supersymmetric-particle production // Phys. Lett. — Vol. 35D, No 1. — 1987. — P. 10.
3. Bergsma F. A new measurement of the ratio of the cross sections on Muon Neutrino and Antineutrino scattering on Electron // Phys. Lett. — Vol. 147B, No 6. — 1984. — P. 481.
4. Reines F., Gurr. H., Sobel H. Calculated spectrum of reactor antineutrino // Phys. Rev. Lett. — Vol. 37. — 1976. — P. 315.
5. Schwarz J. Superstring Theory // Phys. Rep. — Vol. 89. — 1982. — P. 223.
6. Witten E. Symmetry breaking patterns in superstring models // Nucl. Phys. — Vol. B258. — 1985. — P. 75.
7. Barger V., Deshpande. N., Whisnant K. Phenomenological Mass Limits on Extra Z of E_6 Superstrings // Phys. Rev. Lett. — Vol. 56. — 1986. — P. 30.
8. Del Aguila F. et al. Analysis of neutral currents in superstring inspired models // Nucl. Phys. — Vol. B283. — 1987. — P. 50.
9. Керимов Б. К. Спировые корреляции в нейтриноэлектронном рассеянии и нейтральные лептонные токи // Изв. АН СССР, сер. физ. — Т. 37. — 1973. — С. 1937.
10. Schreckenbach K. Determination of the antineutrino spectrum from ^{235}U thermal neutron fission products up to 9.5 Mev // Phys. Lett. — Vol. 160B. — 1985. — P. 325.
11. Avignon F. T., Greenwood. Z. D. Calculated spectra of antineutrino from the fission products of ^{235}U , ^{238}U and ^{239}Pu , and antineutrino-induced reactions // Phys. Rev. — Vol. C22. — 1980. — P. 594.

UDC 539.17.01

Elastic Reactor Antineutrino Scattering by Electron and Bound on the Mass of Extra Neutral Gauge Boson

Yu. P. Bogdanov, I. Yu. Didkovskaya

*Department of Physics
Peoples' Friendship University of Russia
6, Miklukho-Maklaya str., Moscow, 117198, Russia*

Elastic antineutrino-electron scattering is studied in the frame of standard electroweak model. Restrictions on the parameters and angles of model mixing as well as bounds for the mass of assumed extra gauge neutral boson are obtained.

УДК 539.12.01+537.8

Об основных параметрах электромагнитных частиц и электромагнитном взаимодействии

А. А. Черницкий

*Лаборатория теоретической физики им. А. А. Фридмана,
Институт гравитации и космологии
Российского университета дружбы народов;
Санкт-Петербургский Инженерно-экономический университет,
Россия, 191002, Санкт-Петербург, ул. Марата, 27*

Рассматривается модель нелинейной электродинамики типа Борна–Инфельда. Вводятся определения массы, спина, заряда и дипольного момента электромагнитной частицы. Классические уравнения движения массивной заряженной электромагнитной частицы со спином во внешнем электромагнитном поле получены из интегральных законов сохранения полевой модели. Получение этих уравнений доказывает обоснованность введенных определений массы, спина, заряда и дипольного момента частицы. В частности, также становится обоснованным известное Эйнштейновское соотношение эквивалентности массы и энергии.

Введение

Простейшая электромагнитная модель элементарной частицы ассоциируется со сферически симметричным решением уравнений линейной электродинамики с точечной заряженной сингулярностью. Это хорошо всем известный точечный электрон. Однако такая модель по существу обладает только одним параметром наблюдаемой элементарной частицы: электрическим зарядом. Остальные основные характеристики частицы: масса, спин и магнитный момент — должны в этом случае вводиться дополнительными математическими конструкциями, выходящими за рамки электродинамики как полевой модели. Идея же создания полноценной полевой модели частицы представляется весьма привлекательной.

Само поле должно подчиняться нелинейным уравнениям для возможности естественного описания взаимодействия между частицами. Соответствующая модель нелинейной электродинамики впервые была предложена Густавом Ми. Впоследствии Макс Борн и Леопольд Инфельд исследовали другую электродинамическую модель, которая в настоящее время рассматривается как весьма многообещающая [1].

Электромагнитная частица, выглядящая с большого расстояния как точечный объект, при достаточно близком рассмотрении должна иметь определённую структуру, которая и обеспечивает все её наблюдаемые характеристики. Такая интерпретация точечности элементарных частиц обсуждалась в литературе [2] и представляется обоснованной.

В настоящей статье рассматривается нелинейная электродинамическая модель типа Борна–Инфельда с произвольной функцией нелинейности. Показывается, как в рамках задачи о взаимодействии электромагнитной частицы с удалёнными частицами естественно проявляются основные её характеристики: масса, спин, заряд и дипольный момент.

1. Нелинейная вакуумная электродинамика

Общая система электродинамических уравнений в инерциальной системе координат плоского пространства-времени и для областей пространства вне сингулярностей поля имеет вид

$$\begin{aligned} \frac{\partial \mathbf{B}}{\partial x^0} + \text{curl} \mathbf{E} &= 0, & \text{div} \mathbf{B} &= 0, \\ \frac{\partial \mathbf{D}}{\partial x^0} - \text{curl} \mathbf{H} &= 0, & \text{div} \mathbf{D} &= 0, \end{aligned} \quad (1)$$

где $\mathbf{D}$ и $\mathbf{B}$ — электрическая и магнитная индукции, $\mathbf{E}$ и $\mathbf{H}$ — электрическая и магнитная напряжённости.

Система (1) дополняется алгебраическими соотношениями, связывающими векторы $\mathbf{D}$, $\mathbf{B}$, $\mathbf{E}$, $\mathbf{H}$. Эти соотношения будем называть материальными, которые в общем случае нелинейны. Пары векторов $(\mathbf{E}, \mathbf{B})$ и $(\mathbf{D}, \mathbf{H})$, как известно, представляют собой пространственно-временные антисимметричные электромагнитные тензоры и материальные соотношения могут быть записаны как связь между этими тензорами. В этом случае в качестве неизвестных функций задачи может быть взята пара $(\mathbf{E}, \mathbf{B})$ либо пара $(\mathbf{D}, \mathbf{H})$. Однако материальные соотношения могут быть преобразованы так, что они будут связывать пару векторов напряжённостей $(\mathbf{E}, \mathbf{H})$ с парой векторов индукций $(\mathbf{D}, \mathbf{B})$ [3]. Причем в этом случае

$$E_i = 4\pi \frac{\partial \mathcal{E}}{\partial D^i}, \quad H_i = 4\pi \frac{\partial \mathcal{E}}{\partial B^i}, \quad (2)$$

где $\mathcal{E} = \mathcal{E}(\mathbf{D}, \mathbf{B})$ — плотность электромагнитной энергии, ассоциированная с системой уравнений (1); латинские индексы принимают значения 1, 2, 3.

Общие нелинейные материальные соотношения тогда имеют вид

$$\mathbf{E} = \mathbf{E}(\mathbf{D}, \mathbf{B}), \quad \mathbf{H} = \mathbf{H}(\mathbf{D}, \mathbf{B}). \quad (3)$$

Для случая слабого поля эти соотношения должны быть асимптотически близки к линейным

$$\mathbf{E} = \mathbf{D}, \quad \mathbf{H} = \mathbf{B}. \quad (4)$$

Оказалось удобным в качестве неизвестных функций рассматривать именно электромагнитные индукции $(\mathbf{D}, \mathbf{B})$. Это, в частности, позволяет записывать формулы в виде, явно инвариантном относительно дуальных преобразований

$$\begin{pmatrix} \mathbf{D}' \\ \mathbf{B}' \end{pmatrix} = \begin{pmatrix} \cos \zeta & -\sin \zeta \\ \sin \zeta & \cos \zeta \end{pmatrix} \begin{pmatrix} \mathbf{D} \\ \mathbf{B} \end{pmatrix}, \quad (5a)$$

$$\begin{pmatrix} \mathbf{E}' \\ \mathbf{H}' \end{pmatrix} = \begin{pmatrix} \cos \zeta & -\sin \zeta \\ \sin \zeta & \cos \zeta \end{pmatrix} \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix}, \quad (5b)$$

где ζ — угол дуального поворота.

Наконец, весьма удобно использовать гиперкомплексное представление для электродинамики [4, 5]. В этом представлении электродинамическое уравнение (одно, вместо четырёх (1)) может быть записано для двух квазивекторов

$$\mathbf{Y} \doteq \mathbf{D} + \imath \mathbf{B}, \quad \mathbf{Z} \doteq \mathbf{E} + \imath \mathbf{H}, \quad (6)$$

где $\imath$ — гипермнимая единица или единичный псевдоскаляр, $\imath^2 = -1$. В общем случае умножение гиперкомплексных чисел некоммутативно. Однако $\imath$ коммутирует с векторами поля и с оператором дифференцирования, входящим в уравнение представления, в котором $\mathbf{Y}$ рассматривается как неизвестная функция. Таким образом, $\imath$ ведёт себя и используется как обычная мнимая единица.

2. Законы сохранения энергии-импульса и момента импульса

Запишем дифференциальные законы сохранения плотностей энергии-импульса и момента энергии-импульса (в декартовых координатах) для системы уравнений (1)

$$\frac{\partial T^{\mu\nu}}{\partial x^\nu} = 0, \quad \frac{\partial M^{\mu\nu\rho}}{\partial x^\rho} = 0. \quad (7)$$

Греческие индексы принимают значения 0, 1, 2, 3 и по повторяющимся индексам подразумевается суммирование.

Компоненты симметричного тензора плотности энергии-импульса выражаются через компоненты поля следующим образом

$$\begin{aligned} T^{00} &\doteq \mathcal{E}, & T^{0i} &= T^{i0} \doteq \mathcal{P}^i, & \mathcal{P} &= \frac{1}{4\pi} (\mathbf{D} \times \mathbf{B}), \\ T^{ij} &= \frac{1}{4\pi} \left[\mathfrak{m}^{ij} (D^l E_l + B^l H_l) - (D^i E^j + B^i H^j) \right] - \mathfrak{m}^{ij} \mathcal{E}, \end{aligned} \quad (8)$$

где $\mathcal{E}$ — плотность энергии, $\mathcal{P}$ — плотность импульса, $\mathfrak{m}^{\mu\nu}$ — компоненты метрического тензора инерциальной системы координат ($\mathfrak{m}^{0i} = 0$, $|\mathfrak{m}^{00}| = 1$). Функция плотности энергии $\mathcal{E}(\mathbf{D}, \mathbf{B})$ задаёт, согласно (2), конкретную электродинамическую модель. Для линейных соотношений (3) плотность энергии имеет вид

$$\mathcal{E} = \mathcal{E} \doteq \frac{1}{8\pi} (\mathbf{D}^2 + \mathbf{B}^2) = \frac{1}{8\pi} (\mathbf{Y} \cdot {}^*\mathbf{Y}), \quad (9)$$

где ${}^*\mathbf{Y}$ означает комплексное сопряжение к $\mathbf{Y}$ (6).

Компоненты квазитензора¹ плотности момента энергии-импульса связаны с тензором плотности энергии-импульса следующим образом:

$$M^{\mu\nu\rho} \doteq x^\mu T^{\nu\rho} - x^\nu T^{\mu\rho}. \quad (10)$$

Рассмотрим трёхмерный объём V , ограниченный замкнутой двумерной поверхностью $\mathcal{S}$ с направленным векторным элементом $d\mathcal{S}$. Считаем объём и поверхность фиксированными (не зависящими от времени) в выбранной системе координат. Определим энергию и импульс поля в объёме V как интегралы от соответствующих плотностей

$$\mathbb{E}_V \doteq \mathbb{P}_V^0 \doteq \int_V \mathcal{E} dV, \quad \mathbb{P}_V^i \doteq \int_V \mathcal{P}^i dV. \quad (11)$$

Аналогично определяется антисимметричный тензор момента импульса

$$\mathbb{M}_V^{\mu\nu} \doteq \int_V M^{\mu\nu 0} dV. \quad (12)$$

Для компонент последнего тензора, содержащих нулевой индекс, согласно определениям (10) и (11), имеем следующее тождество

$$\mathbb{M}_V^{0i} \equiv x^0 \mathbb{P}_V^i - x_V^i \mathbb{E}_V. \quad (13)$$

¹Плотность момента энергии-импульса является тензором только относительно линейных преобразований координат: сдвигов, пространственно-временных поворотов, масштабных преобразований. Это связано с тем, что соответствующее определение содержит координаты. В дальнейшем, однако, приставку «квази» для аналогичных величин будем опускать.

Здесь введено обозначение для координат энергетического центра поля в объёме V :

$$\mathbf{x}_V^i \doteq \frac{1}{\mathbb{E}_V} \int_V x^i \mathcal{E} dV. \quad (14)$$

Очевидно, что в общем случае интегральные величины $\mathbb{P}_V^\mu$, $\mathbb{M}_V^{\mu\nu}$, $\mathbf{x}_V$ могут являться функциями времени.

Три независимые компоненты тензора момента импульса, не содержащие нулевого индекса, можно представить компонентами трёхмерного вектора. Используем для этого вектора ту же букву, что и для антисимметричного тензора, но компоненты будут нумероваться одним индексом:

$$\mathbb{M}_{Vi} \doteq \frac{1}{2} \epsilon_{ijk} \mathbb{M}_V^{ik}, \quad \mathbb{M}_V \doteq |\mathbf{M}_V|, \quad \mathbf{M}_V = \int_V \mathbf{r} \times \mathcal{P} dV, \quad (15)$$

где ϵ_{ijk} — компоненты полностью антисимметричного единичного тензора третьего ранга.

Интегральные законы сохранения для поля в объёме V получаются из дифференциальных обычным образом и имеют вид

$$\frac{d\mathbb{E}_V}{dx^0} = - \int_S T^{0i} d\mathcal{S}_i, \quad \frac{d\mathbb{P}_V^i}{dx^0} = - \int_S T^{ij} d\mathcal{S}_j. \quad (16)$$

$$\frac{d\mathbb{M}_V^{\mu\nu}}{dx^0} = - \int_S M^{\mu\nu p} d\mathcal{S}_p. \quad (17)$$

3. Масса и спин

Считаем, что частице соответствует некоторое пространственно локализованное решение нелинейной электродинамической модели, которое назовём корпускулярным. Будем обозначать корпускулярное решение кружком сверху. Так, поле корпускулярного решения обозначаем $(\overset{\circ}{\mathbf{D}}, \overset{\circ}{\mathbf{B}})$.

Пространственная локализация поля $(\overset{\circ}{\mathbf{D}}, \overset{\circ}{\mathbf{B}})$ предполагает то, что его плотность энергии-импульса имеет сравнительно большие значения в некоторой ограниченной области трёхмерного пространства, в то время как в остальном пространстве плотность энергии-импульса достаточно мала.

Естественно определить энергию-импульс электромагнитной частицы как полные энергию и импульс соответствующего корпускулярного решения. Однако для того, чтобы такое определение было возможно, необходимо ввести более узкое понятие области локализации корпускулярного решения: область локализации (полной) энергии корпускулярного решения. Обозначим эту область $\overset{\circ}{V}$, а ограничивающую поверхность $\overset{\circ}{S}$.

Положим

$$\overset{\circ}{\mathbb{P}}^\mu \doteq \overset{\circ}{\mathbb{P}}_{\overset{\circ}{V}}^\mu \approx \lim_{V \rightarrow \infty} \mathbb{P}_V^\mu. \quad (18)$$

Предполагается, что для любой заданной точности $\varepsilon > 0$ существует такая область $\overset{\circ}{V}$, что

$$|\overset{\circ}{\mathbb{P}}_{\overset{\circ}{V}}^\mu - \lim_{V \rightarrow \infty} \mathbb{P}_V^\mu| < \varepsilon. \quad (19)$$

В этом смысле понимается приближённое равенство в (18). Поскольку объём интегрирования здесь считается постоянным, определение области локализации энергии в общем случае относится к определённому малому интервалу времени.

Фактически, существование области локализации энергии равносильно существованию конечного предела $|\lim_{V \rightarrow \infty} \mathbb{P}_V^\mu| < \infty$.

Здесь будем рассматривать корпускулярные решения, для которых выполняется соотношение

$$|\mathbf{m}_{ij} \mathring{\mathbb{P}}^i \mathring{\mathbb{P}}^j| < \mathring{\mathbb{E}}^2. \quad (20)$$

Для таких решений существует инерциальная система координат, в которой все компоненты импульса обращаются в ноль $\mathring{\mathbb{P}}^i = 0$. Переход к ней осуществляется обычным пространственно-временным поворотом. Корпускулярные решения, для которых выполняется соотношение (20), соответствуют массивным частицам, скорость которых меньше скорости света.

Определённое корпускулярное решение соответствует определённой частице. Многочастичным конфигурациям с взаимодействием физических частиц соответствуют многочастичные решения полевой модели: решения, в которых можно выделить несколько пространственных областей со сравнительно большой плотностью энергии. Такие решения часто могут исследоваться при помощи метода возмущений с начальным приближением в виде суммы соответствующих корпускулярных решений [6]. Взаимодействие будет проявляться, в частности, как зависимость энергии-импульса частиц от времени. Для корпускулярного же решения энергия и импульс от времени не зависят.

Естественно определить собственную систему координат частицы как такую, в которой равны нулю компоненты её импульса $\mathring{\mathbb{P}}^i = 0$ и энергетического центра (14) $\mathring{x}^i \doteq \mathring{x}_{\mathring{V}}^i = 0$. Однако в случае взаимодействия и, следовательно, изменения траектории частицы её собственная система не будет инерциальной. Поэтому удобно также определить такую инерциальную систему координат, для которой в заданный момент времени выполняются соотношения $\mathring{\mathbb{P}}^i = 0$, $\mathring{x}^i = 0$. Будем называть такую систему собственной инерциальной системой частицы. Для одиночной частицы (корпускулярного решения) собственная система является и инерциальной. Будем обозначать величины, относящиеся к собственной инерциальной системе координат частицы точкой снизу:

$$\mathring{\mathbb{P}}^i \doteq 0, \quad \mathring{x}^i \doteq 0. \quad (21)$$

Из тождества (13) и определения (21) следует, что в собственной инерциальной системе координат $\mathring{\mathbb{M}}^{0i} \equiv 0$. Произведём временной сдвиг и временной поворот из собственной инерциальной системы координат

$$\mathring{x}^i = \frac{v^i (x^0 - q^0)}{\sqrt{1 - v^2}}, \quad x^0 = \frac{(x^0 - q^0)}{\sqrt{1 - v^2}}, \quad \mathring{\mathbb{P}}^i = \frac{v^i \mathring{\mathbb{E}}}{\sqrt{1 - v^2}}, \quad \mathring{\mathbb{E}} = \frac{\mathring{\mathbb{E}}}{\sqrt{1 - v^2}}. \quad (22)$$

Подставляя (22) в (13), получаем тождество

$$\mathring{\mathbb{M}}^{0i} \equiv 0, \quad (23)$$

справедливое в инерциальных системах координат, полученных временным сдвигом и временным поворотом из собственной без пространственного сдвига в ней.

Корпускулярное решение в собственной системе координат в общем случае может содержать часть, зависящую от времени. Запишем разложение поля корпускулярного решения $\mathring{\mathbf{Y}}$ в виде временного ряда Фурье

$$\mathring{\mathbf{Y}} = \sum_{n=-\infty}^{\infty} \mathring{\mathbf{Y}}_{|n} e^{-i n \omega x^0}, \quad (24)$$

где $\mathring{\mathbf{Y}}_{|n} = \mathring{\mathbf{Y}}_{|n}(x^i)$. Здесь подразумевается периодическая зависимость от времени с периодом $2\pi/\omega$, которая может быть переведена в непериодическую путём устремления $\omega \rightarrow 0$ и замены суммы интегралом.

Также можно разложить и плотность энергии

$$\overset{\circ}{\xi} = \sum_{n=-\infty}^{\infty} \overset{\circ}{\xi}_{|n} e^{-i n \omega x^0}. \quad (25)$$

Если мы говорим о массе как о некоторой постоянной характеристике частицы, то, очевидно, для её определения можно использовать только часть плотности энергии, независимую от времени $\overset{\circ}{\xi}_{|0}$. Однако эта часть плотности энергии определяется также и переменной частью поля (24), поскольку содержит все функции $\overset{\circ}{Y}_{|n}(x^i)$. Для слабого поля, которое должно таким стать при удалении от возможной области локализации, согласно (9) имеем

$$\overset{\circ}{\xi}_{|0} = \frac{1}{8\pi} \sum_{n=-\infty}^{\infty} \overset{\circ}{Y}_{|n} \cdot {}^* \overset{\circ}{Y}_{|n}. \quad (26)$$

Можно показать, используя, например, математический аппарат, изложенный в статье [5], что для плотности энергии (26), содержащей слагаемые с $n \neq 0$, соответствующий интеграл энергии расходится на бесконечности, т.е. при удалении от возможной области локализации. Следовательно, для поля с не исчезающей переменной по времени составляющей (только в представлении решения на бесконечности) нельзя ввести область локализации энергии, определённую в соответствии с формулами (18), (19). Значит, для того, чтобы можно было говорить о такой области локализации энергии, надо рассматривать решения, для которых переменная составляющая не содержится в представлении решения на бесконечности: $\overset{\circ}{Y} = \overset{\circ}{Y}_{|0}$. При этом, конечно, вблизи центра частицы может иметь место общий случай: $\overset{\circ}{Y} \neq \overset{\circ}{Y}_{|0}$.

Также может рассматриваться случай, когда с корпускулярным решением связана периодическая по времени часть, не исчезающая из представления решения на бесконечности. Полная энергия такой конфигурации поля бесконечна.

Часть, периодическая по времени в собственной системе координат, даст бегущую волну в системе координат, относительно которой частица движется. Эта волна может быть ответственна за волновое поведение частицы. Здесь же, однако, рассматривается случай, когда влиянием волны на движение частицы можно пренебречь и частица проявляет только корпускулярные свойства.

Таким образом, рассматриваем корпускулярное решение, для которого $\overset{\circ}{Y} = \overset{\circ}{Y}_{|0}$ на бесконечности. Кроме того, считаем, что полная энергия и момент импульса не зависят от времени: $\overset{\circ}{E} = \overset{\circ}{E}_{|0}$, $\overset{\circ}{M} = \overset{\circ}{M}_{|0}$. При этих предположениях определим массу и вектор спина электромагнитной частицы как полную энергию и вектор момента импульса соответствующего корпускулярного решения в собственной инерциальной системе координат:

$$m \doteq \overset{\circ}{E}, \quad s \doteq \overset{\circ}{M}. \quad (27)$$

С учётом данного определения массы имеем следующее известное выражение для энергии-импульса частицы в любой инерциальной системе координат:

$$\overset{\circ}{E} = \frac{m}{\sqrt{1-v^2}}, \quad \overset{\circ}{P}^i = \frac{m v^i}{\sqrt{1-v^2}}, \quad (28)$$

где v^i — компоненты скорости частицы.

Как известно, интеграл энергии поля точечного электрона линейной электродинамики расходится вблизи сингулярности, что, согласно определению (27), даёт бесконечную массу. В электродинамике же Борна-Инфельда масса соответствующего корпускулярного решения с точечной сингулярностью конечна.

Существуют статические конфигурации электромагнитного поля со спином. Это, например, конфигурация с двумя точечными дионами [6]. Также с этой точки зрения заслуживает внимания конфигурация с кольцевой сингулярностью [7].

4. Заряд и дипольный момент

На краю области локализации поле должно быть достаточно мало так, что с некоторой точностью можно считать справедливыми материальные соотношения, характерные для линейной модели (4). Соответствующее решение известно и может быть представлено в виде мультипольного разложения (см., например, [5]). Так, для статической части имеем

$$\hat{\mathbf{Y}}_{|0} = \mathcal{C} \frac{\mathbf{r}}{r^3} + 3 (\mathbf{D} \cdot \mathbf{r}) \frac{\mathbf{r}}{r^5} - \mathbf{D} \frac{1}{r^3} + \dots, \quad (29)$$

где $\mathcal{C} \doteq \mathcal{C} + i\mathcal{Q}$ — монопольный момент или электромагнитный заряд, $\mathbf{D} \doteq \mathbf{D} + i\mathbf{Q}$ — вектор дипольного электромагнитного момента, соответственно $\mathcal{C}$, $\mathcal{Q}$, $\mathbf{D}$, $\mathbf{Q}$ — электрические и магнитные заряды и векторы дипольных моментов.

Здесь надо подчеркнуть, что вид разложения (29) не зависит от конфигурации и характера сингулярностей поля вблизи центра энергии частицы. Сингулярности могут быть точечные, с магнитными зарядами или без, кольцевые или поверхностные. В любом случае вдали от области локализации поле будет иметь вид (29). Различаться могут только мультипольные константы $\mathcal{C}$, $\mathbf{D}$, ...

5. Частица во внешнем поле

Рассмотрим ситуацию взаимодействия частицы с другими частицами, удалёнными от неё на достаточно большое расстояние так, что области локализации частиц не перекрываются. Можно искать соответствующее многочастичное решение методом возмущений, взяв в качестве начального приближения сумму корпускулярных решений [6]. В области локализации рассматриваемой частицы начальное приближение запишем в виде

$$\mathbf{D} = \mathring{\mathbf{D}} + \tilde{\mathbf{D}}, \quad \mathbf{B} = \mathring{\mathbf{B}} + \tilde{\mathbf{B}}, \quad (30)$$

где $\mathring{\mathbf{D}}$, $\mathring{\mathbf{B}}$ — малые поля удалённых частиц в области локализации рассматриваемой.

Для упрощения будем решать задачу в собственной инерциальной системе координат частицы. Частицу же будем считать статической, т.е. не содержащей временной части. Это соответствует обсуждавшемуся выше случаю возможности пренебрежения периодической по времени частью. Добавление заданного поля удалённых частиц $\hat{\mathbf{Y}}$ должно изменять свободные параметры корпускулярного решения $\hat{\mathbf{Y}}$. Свободные параметры здесь — это параметры пространственно-временного поворота и параметры пространственного сдвига.

Параметр масштабного преобразования, которым обладает корпускулярное решение в силу соответствующей инвариантности исходной полевой модели, предполагается фиксированным. Его фиксация соответствует сохранению заряда. Заряд же в модели рассматриваемого типа связан с сингулярностью поля. Сингулярности могут быть точечные, линейные (струнные) и поверхностные. Для точечной электрически заряженной сингулярности в модели нелинейной электродинамики доказательство сохранения заряда имеется в статье [8]. Принимая во внимание электромагнитную дуальность электродинамики [6], аналогично может быть доказано сохранение магнитного точечного заряда. Подобным же образом может быть доказано сохранение полного заряда струны или поверхности. В этой связи надо также упомянуть об обсуждавшейся проблеме устойчивости солитонов [9] и способе её обеспечения путём фиксации заряда.

Представим малое поле удалённых частиц двумя членами степенного ряда вблизи центра рассматриваемой частицы:

$$\tilde{D}_i = \tilde{D}_{i|} + \tilde{D}_{i|j} x_j, \quad \tilde{B}_i = \tilde{B}_{i|} + \tilde{B}_{i|j} x_j, \quad (31)$$

где $\tilde{D}_i$, $\tilde{B}_i$ — значения i -й компоненты поля в центре рассматриваемой частицы, $\tilde{D}_{i|j}$, $\tilde{B}_{i|j}$ — значения производных от i -й компоненты поля по j -й координате там же.

Рассмотрим интегральные законы сохранения (16) и (17) для поля (30) в области локализации рассматриваемой частицы. При этом считаем, что область выбрана достаточно большой так, что на её границе поле мало и модель можно считать линейной. Это значит, что на границе выполняются соотношения (4) и поле подчиняется линейным уравнениям Максвелла. Кроме того, слабое поле $(\tilde{\mathbf{D}}, \tilde{\mathbf{B}})$ подчиняется линейным уравнениям везде в области локализации рассматриваемого корпускулярного решения. Подставим сумму (30) в интегральные законы сохранения (16) и (17). При подстановке в правые части берём вид корпускулярного решения на бесконечности (29) и разложение заданного поля (31). После выполнения поверхностного интегрирования и учёта линейных уравнений Максвелла для поля $(\tilde{\mathbf{D}}, \tilde{\mathbf{B}})$ получаем следующие уравнения

$$\frac{d\mathbf{P}}{dx^0} = \mathcal{E} \tilde{\mathbf{D}} + \mathcal{G} \tilde{\mathbf{B}} + (\mathbf{P} \cdot \nabla) \tilde{\mathbf{D}} + (\mathbf{P} \cdot \nabla) \tilde{\mathbf{B}}, \quad (32)$$

$$\frac{d\mathbf{M}}{dx^0} = \mathbf{P} \times \tilde{\mathbf{D}} + \mathbf{P} \times \tilde{\mathbf{B}}. \quad (33)$$

Эти уравнения совпадают с классическими уравнениями движения заряженной частицы со спином и дипольным моментом во внешнем поле, записанными в собственной инерциальной системе координат частицы.

Заключение

Таким образом, выведены классические уравнения движения массивной заряженной частицы со спином и дипольным (в частности, магнитным) моментом во внешнем поле. Эти уравнения получились естественно как следствие интегральных законов сохранения для поля. Но именно динамические законы движения частицы в электромагнитном поле, появившиеся сначала феноменологически, требуют задания её массы, спина, заряда и дипольного момента. Значит, здесь можно говорить об обосновании введения этих характеристик. В частности, становится обоснованным известное Эйнштейновское соотношение эквивалентности массы и энергии.

Литература

1. Chernitskii A. A. 2004. — Born–Infeld Equations. — // Encyclopedia of Nonlinear Science, Ed. by A. Scott. New York and London: Routledge, Pp. 67–69.
2. Kuryshkin V. V., Entralgo E. E. Classical Theory of Point Spin-Particle and Structural Point Objects // Nuovo Cim. A. — Vol. 103. — 1990. — Pp. 561–575.
3. Bialynicki-Birula I. 1983. — Nonlinear Electrodynamics: Variations on a Theme by Born and Infeld. — // Quantum Theory of Particles and Fields, Ed. by B. Janciewicz, J. Lukierski. World Scientific, Pp. 31–47.
4. Chernitskii A. A. Born–Infeld Electrodynamics: Clifford Number and Spinor Representations // Int. J. Math. & Math. Sci. — Vol. 31. — 2002. — Pp. 77–84.
5. Chernitskii A. A. Basic Systems of Orthogonal Functions for Space-Time Multivectors // Advances in applied Clifford algebras. — Vol. 15. — 2005. — Pp. 27–53.
6. Chernitskii A. A. Dyons and Interactions in Nonlinear (Born–Infeld) Electrodynamics // J. High Energy Phys. — Vol. 12. — 1999. — P. 10.
7. Буринский А. Я. Микрогеон со спином // ЖЭТФ. — Т. 66. — 1974. — С. 406–411.
8. Chernitskii A. A. Nonlinear Electrodynamics with Singularities (Modernized Born–Infeld Electrodynamics) // Helv. Phys. Acta. — Vol. 71. — 1998. — Pp. 274–287.

9. Рыбаков Ю. П. Устойчивость солитонов. — В сб. Проблемы теории гравитации и элементарных частиц / Под ред. К. П. Станюковича, Вып. 16. — М.: Энергоатомиздат. — 1985.

UDC 539.12.01+537.8

On General Parameters of Electromagnetic Particles and Electromagnetic Interaction

A. A. Chernitskii

*A. Friedmann Laboratory for Theoretical Physics,
Institute of Gravitation and Cosmology at Peoples Friendship University of Russia,
St.-Petersburg University of Engineering and Economics,
27, Marata str., St.-Petersburg, 191002, Russia*

Nonlinear electrodynamics model of Born–Infeld type is considered. Definitions of mass, spin, charge, and dipole moment for electromagnetic particle are introduced. The classical equations of motion for massive charged electromagnetic particles with spin and dipole moment in external electromagnetic field are obtained from integral conservation laws for the field model. The obtaining of these equations demonstrates the validity of introduced definitions for mass, spin, charge, and dipole moment for electromagnetic particle. Also, in particular, the known Einstein’s relation for equivalence of mass and energy becomes valid.

УДК 536.7

Динамическое давление и его флуктуации для классического идеального газа релятивистских частиц

Ю. Г. Рудой *, И. Кейта †

* Центр естественно-научного образования РУДН

† Кафедра теоретической физики

Российский университет дружбы народов

Россия, 117198, Москва, ул. Миклухо-Маклая, 6

Дан анализ динамических величин — давления и сжимаемости идеального газа классических частиц с релятивистским законом дисперсии. Анализ основан на обобщенной теореме Боголюбова–Зубарева для квазидинамических величин. Показано, что учет релятивистских поправок как в области малых, так и больших импульсов может быть реализован посредством эффективного показателя однородности.

Введение

Важной проблемой равновесной статистической механики Гиббса [1] является вычисление равновесных термодинамических флуктуаций одной из обобщенных сил — динамического¹ давления P макроскопической системы. Согласно известной лемме Гиббса [1] (см. также, например, [2, 3], для этого помимо знания динамического давления P необходимо также знание динамической сжимаемости Ψ (соответствующие определения даны ниже).

Выражение для давления и его термодинамических флуктуаций требует нахождения гиббсовских статистических средних $\langle P \rangle$ и $\langle \Psi \rangle$ от динамических величин P и Ψ . При этом задача усреднения для P сводится к вычислению статистической суммы и ее производных, так что $\langle P \rangle$ относится к классу так называемых «термодинамических» средних, тогда как для Ψ она может иногда потребовать независимых дополнительных вычислений так называемых «нетермодинамических» средних.

Длительное время проблема вычисления флуктуаций давления — прежде всего нахождение Ψ и $\langle \Psi \rangle$ — не имела последовательного решения в рамках подхода Гиббса или нуждалась в привлечении каких-либо дополнительных предположений (см., например, обсуждение в [2, 4, 5], а также в [6, 7]). Лишь недавно в работе Рудого и Суханова [3] на основе идей метода квазисредних Боголюбова [8] было получено полное и последовательное решение проблемы флуктуаций давления в макроскопической системе, находящейся в классическом и слабовырожденном квантовом режиме.

В качестве конкретного примера в [3] был рассмотрен идеальный f -мерный газ частиц с f трансляционными степенями свободы в нерелятивистском и ультрарелятивистском предельных случаях. В этих случаях функция Гамильтона, или кинетическая энергия $H(p)$ свободной частицы с импульсом p , является однородной в смысле Эйлера. Представляет, однако, интерес и более общий случай релятивистской частицы, когда полная энергия $H(p)$ частицы складывается из

Авторы благодарны проф. Ю. П. Рыбакову за внимание к работе и полезные обсуждения.

¹ Динамическими здесь и ниже называются величины, заданные на фазовом пространстве макроскопической системы. Если эта система находится в тепловом контакте с термостатом (т. е. является термодинамической), то все динамические величины становятся стохастическими, или случайными, и для них имеют смысл понятия среднего значения и флуктуаций (подробнее см. ниже).

энергии покоя $E_0 \equiv H(0)$ и ее кинетической энергии $H_k(p)$ (для которой, по определению, $H_k(0) = 0$), так что $H(p)$ заведомо не является однородной функцией.

Согласно релятивистской механике, для свободной частицы энергия $H(p)$ (и, следовательно, $H_k(p)$) изотропна в импульсном пространстве, т. е. зависит только от величины $p = |\mathbf{p}|$ вектора $\mathbf{p}$:

$$\begin{aligned} H(p) &\equiv E_0 + H_k(p) = [E_0^2 + (cp)^2]^{1/2}, \\ H(p) &= E_0 h(p), \quad h(p) = 1 + h_k(p), \end{aligned} \quad (1)$$

где c — скорость света в вакууме, $h(p)$ и $h_k(p)$ — безразмерные энергии.

Как показано в [3], динамические величины P и Ψ в случае идеального газа определяются только кинетической энергией $H_k(p)$, для которой обычно используются следующие приближенные выражения:

$$H_k^{\text{нр}}(p) \approx \frac{(cp)^2}{2E_0} \quad \text{при} \quad \frac{cp}{E_0} \ll 1, \quad \text{нерелятивистский («нр») предел}; \quad (2)$$

$$H_k^{\text{ур}}(p) \approx cp \quad \text{при} \quad \frac{cp}{E_0} \gg 1, \quad \text{ультрарелятивистский («ур») предел}. \quad (3)$$

Очевидно, в обоих предельных случаях (2) и (3) кинетическая энергия

$$H_k(p) = \alpha_k p^k \quad \left(k_{\text{ур}} = 1, \quad \alpha_1 \equiv \alpha_{\text{ур}} = c; \quad k_{\text{нр}} = 2, \quad \alpha_2 \equiv \alpha_{\text{нр}} = \frac{\alpha_1^2}{2E_0} \right) \quad (4)$$

является степенной и потому однородной (в смысле Эйлера) функцией импульса p с показателем однородности k , равным 2 и 1 соответственно³.

Настоящая работа посвящена применению подхода, развитого в [3], к общему релятивистскому случаю (1), когда $H_k(p)$ не является однородной функцией импульса p . Можно показать, однако, что и в этом случае полученные в [3] соотношения для давления P и сжимаемости Ψ с предельными (однородными) выражениями для функции $H_k(p)$ вида (4) сохраняют свой вид, если формально ввести в этих выражениях вместо одного постоянного показателя k два переменных (зависящих от импульса p) «показателя однородности»

$$k(p) = 1 + [h(p)]^{-1}, \quad m(p) = 1 + [h(p)]^{-2} \quad (1 \ll k(p), m(p) \ll 2). \quad (5)$$

Очевидно, что при всех p показатели (5) строго ограничены — как сверху «нерелятивистским» значением $k_{\text{нр}} = 2$ при $cp/E_0 \rightarrow 0$, когда $h_k(p) \rightarrow 0$, $h(p) \rightarrow 1$, так и снизу — «ультрарелятивистским» значением $k_{\text{ур}} = 1$ при $cp/E_0 \rightarrow \infty$, когда $h_k(p)$ и $h(p) \rightarrow \infty$, причем $k = m \equiv 1$ в случае $E_0 = 0$ (более подробно см. раздел 2).

1. Функция Гамильтона и ее связь с квазидинамическими давлением и сжимаемостью

Рассмотрим макроскопическую систему с фиксированным числом N одинаковых f -мерных частиц, находящуюся в классическом режиме (т. е. при достаточно

²В частном (но важном) случае безмассовых частиц (например, фотонов) с $E_0 = 0$ выражение (3) для $H_k^{\text{ур}}(p)$ становится точным.

³Заметим, что развиваемый нами подход сохраняет применимость и для более общих случаев, когда показатель k в (4) принимает не только «предельные» значения 1 и 2, но и любые натуральные, а также дробные и даже отрицательные значения.

высоких температурах и/или низких плотностях, см. ниже сноску 23) и описываемую функцией Гамильтона $H_N(q, p; a) \equiv H_N(\Gamma; a)$. Здесь $\Gamma = \Gamma_q + \Gamma_p - 2fN$ -мерное фазовое пространство координат q и импульсов p всех частиц системы; a — набор механических внешних параметров, из которых мы будем рассматривать только объем V .

Это означает, что координатная часть Γ_q пространства Γ (для одной частицы) ограничена фиксированным объемом V , т. е. система находится внутри сосуда произвольной формы (для простоты — f -мерного куба) с непроницаемыми для частиц гладкими стенками. Предположим, что система находится в тепловом равновесии с термостатом при фиксированной абсолютной температуре $T = 1/k_B\beta$ (k_B — постоянная Больцмана).

Последнее условие делает динамическую систему термодинамической, т. е. способной к случайному обмену энергией с термостатом в виде теплоты; при этом V и T (или β) являются для системы внешними параметрами (механическим и термическим соответственно).

В статистической механике Гиббса подобная система описывается стационарной функцией плотности распределения по состояниям в фазовом пространстве Γ . Эта функция предполагается нормируемой и имеет экспоненциальный, или канонический вид

$$\rho_N(\Gamma; \beta, V) = Z_N^{-1}(\beta, V) \exp[-\beta H_N(\Gamma; V)]. \quad (6)$$

Здесь

$$H_N(\Gamma; V) = H_N(\Gamma) + H_N + \varepsilon U(\Gamma; V), \quad (7)$$

где $U(\Gamma; V)$ — внешний по отношению к системе сингулярный потенциал стенок сосуда, равный нулю всюду внутри объема V сосуда, но принимающий сколь угодно большие значения вне сосуда и на поверхности его стенок.

Формальный параметр $\varepsilon \geq 0$ может быть произвольным, и лишь предел $\varepsilon \rightarrow 0$ соответствует пределу $V \rightarrow \infty$, т. е. переходу к неограниченной («свободной») системе, описываемой функцией Гамильтона $H_N(\Gamma)$.

Нормировочный множитель для распределения (6)

$$Z_N(\Gamma; V) = \int d\Gamma \exp[-\beta H_N(\Gamma; V)] \quad (8)$$

является статистической суммой для термодинамической системы, причем интегрирование по $d\Gamma = d\Gamma_q d\Gamma_p$ предполагается по всему фазовому пространству $d\Gamma$ системы N частиц с f измерениями по q и p у каждой⁴.

Ограничение координатной части Γ_q объемом V для каждой из N частиц в (8) происходит автоматически за счет наличия в (7) сингулярного слагаемого $U(\Gamma_q; V)$: при этом входящий в (8) «режущий» множитель $\exp[-\beta U(\Gamma_q; V)]$ играет роль проекционного оператора на указанную часть Γ_q . Что касается импульсной части Γ_p , то, как правило, функция $H(q, p)$ является растущей⁵ функцией импульса p (во всяком случае, при больших p). Поэтому интегрирование по p можно (с точностью до экспоненциально малых поправок) распространить до сколь угодно больших значений.

Любая функция $X(\Gamma; \beta, V)$, заданная на фазовом пространстве системы Γ , является случайной при тепловом контакте. Среднее значение такой функции здесь и всюду ниже определяется как гиббсовское, или каноническое среднее с функцией распределения (8)

$$X(\beta, V) \equiv \langle X(\Gamma; \beta, V) \rangle = \int d\Gamma X(\Gamma; \beta, V) \rho_N(\Gamma; \beta, V). \quad (9)$$

⁴В дальнейшем мы будем опускать индекс N у величин H_N , ρ_N и Z_N , считая их аддитивными; в силу одинаковости частиц $Z_N = Z^N$ для идеальной системы, а все средние динамические величины (энергия, давление, сжимаемость и т. п.) пропорциональны N .

⁵Мы отвлекаемся здесь от случаев ограниченных сверху функций $H(q, p)$ — например, для периодической в Γ_q динамической системы на решетке, когда импульс p становится квазиимпульсом и его изменение ограничено пределами зоны Бриллюэна.

Согласно Гиббсу [1] (см. также [2]), обобщенная сила $A(\Gamma; a)$, сопряженная внешнему механическому параметру a (в частности, объему), определяется как производная $[-\partial H(\Gamma; a)/\partial a]$. Обычно параметр a входит в функцию Гамильтона линейно, т. е. $H(\Gamma; a) = H(\Gamma) - A(\Gamma; a)$, причем $A(\Gamma; a)$ фактически не зависит от величин⁶ a , так что $\partial A(\Gamma)/\partial a = -\partial^2 H(\Gamma; a)/\partial^2 a = 0$. Это означает, что обобщенная сила $A(\Gamma)$ может быть определена независимо от наличия или отсутствия сопряженного поля⁷ a .

Однако в интересующем нас случае это не так: как видно из (7), внешний параметр $a = V$ входит в $H(\Gamma; V)$ посредством потенциала $U(\Gamma; V)$ существенно нелинейным (и, более того, даже сингулярным) образом. Поэтому сопряженная объему обобщенная сила — динамическое давление $P(\Gamma; V)$ — будет зависеть от значения параметра V , причем очевидно, что $P(\Gamma; V) \rightarrow 0$ при $V \rightarrow \infty$ («нет стенок — нет давления»).

Из общих принципов статистической механики Гиббса следует, что вид функции $P(\Gamma; V)$ не должен зависеть от конкретного вида потенциала стенки $U(\Gamma; V)$, но только от вида «свободной» функции Гамильтона $H(\Gamma)$ из (7) при $\varepsilon = 0$ (то же справедливо и для $\Psi(\Gamma; V)$). С учетом этих соображений корректными определениями давления и сжимаемости следует считать предложенные в [3] определения этих величин как *квазидинамических* в смысле Боголюбова

$$P(\Gamma; V) = \lim_{(\varepsilon \rightarrow 0)} \left[-\frac{\partial H_N(\Gamma; V)}{\partial V} \right], \quad (10)$$

$$\Psi(\Gamma; V) = \lim_{(\varepsilon \rightarrow 0)} \left[-\frac{\partial^2 H_N(\Gamma; V)}{\partial V^2} \right], \quad \Psi(\Gamma; V) \neq \frac{\partial P(\Gamma; V)}{\partial V}. \quad (11)$$

Эти определения следуют основной идее метода квазисредних Боголюбова [3], согласно которой для снятия имеющегося вырождения «свободной» системы (в данном случае — пространственного) необходимо «виртуально» ввести в функцию Гамильтона соответствующий физический «источник» (в данном случае — потенциал стенок $U(\Gamma; V)$), снимающий это вырождение. После проведения требуемого вычисления (в данном случае — производных по V) этот источник следует аннулировать; отличный от нуля результат подобного вычисления означает неаналитичность этого результата по параметру ε (или, что то же, непостоянность операций указанного вычисления и предельного перехода по ε).

Определения (10) и (11) представляются целесообразными в связи с тем, что вычисление производных по V непосредственно от полной функции Гамильтона вида (7) из-за наличия сингулярного слагаемого $U(\Gamma; V)$ не является математически корректной операцией и должно быть каким-либо образом «регуляризовано». Математически это соответствует тому, что вычисление производных от сингулярных, или обобщенных, функций сводится к вычислению производных от регулярных «носителей» (см., например, [9]), для чего необходимо перейти от самих функций к функционалам для них (в данном случае — к статистической сумме).

По-видимому, именно недооценкой указанных обстоятельств были обусловлены трудности в работах [5–7] (см. также [4, 5]). Авторы этих работ пытались провести необходимую регуляризацию, сперва моделируя «реальный» потенциал стенки, а затем переходя к сингулярному пределу «бесконечно высокой» стенки; в результате, однако, для $\Psi(\Gamma; V)$ расходящиеся выражения.

Как показано впервые в работе Рудого и Суханова [3], избежать подобных трудностей позволяет использование в этой проблеме идей метода квазисредних Боголюбова [8]. Именно этот метод представляет собой физически естественную,

⁶Среднее значение $A(\beta, V, a) \equiv \langle A(\Gamma) \rangle$ приобретает зависимость от β , V и a благодаря наличию этих переменных в функции распределения (5) (зависимость от внешнего поля a появляется за счет дополнительного слагаемого $[-\beta A(\Gamma)a]$ в показателе экспоненты).

⁷Например, это имеет место для таких пар a и A как внешнее электрическое поле и поляризация, внешнее магнитное поле и намагнитченность в линейных средах (т. е. исключая сегнетоэлектрические и ферромагнитные среды).

математически корректную (и при том универсальную) процедуру подобной регуляризации для широкого класса физических задач с различным характером «вырождения».

2. Динамические уравнения состояния

2.1. Общие соотношения

Детали вычислений, дающих конструктивную реализацию определений (10) и (11), можно найти в работе [3]; их основной результат состоит в следующем. Пусть макроскопическая динамическая система ограничена конечным объемом V и описывается функцией Гамильтона вида (7). Тогда явные выражения для $P(q, p; V)$ и $\Psi(q, p; V)$ определяются только «свободной» частью функции Гамильтона $H(q, p)$ и не зависят от конкретного вида «потенциала стенок» $U(\Gamma; V)$

$$P_N(q, p; V) = -\frac{1}{fV} \left. \frac{\partial H_N(\lambda q, p/\lambda)}{\partial \lambda} \right|_{\lambda=1}, \quad (12)$$

$$\begin{aligned} \Psi_N(q, p; V) = \frac{1}{V} P_N(q, p; V) + \\ + \left(\frac{1}{fV} \right)^2 \left[\frac{\partial H_N(\lambda q, p/\lambda)}{\partial \lambda} + \frac{\partial^2 H_N(\lambda q, p/\lambda)}{\partial \lambda^2} \right] \Big|_{\lambda=1}. \end{aligned} \quad (13)$$

Выражения (12) и (13) определены практически для любых функций Гамильтона $H(q, p)$ — эти функции должны быть лишь достаточно «гладкими», а именно дважды дифференцируемы по своим аргументам p и q .

Величина λ является параметром канонического (т. е. сохраняющего элемент фазового объема $d\Gamma$) масштабного преобразования. Это преобразование устанавливает связь между изменением объема с эквивалентным ему изменением координат (условие каноничности требует также соответствующего изменения импульсов); после вычисления производных следует всюду положить $\lambda = 1$.

Заметим, что выражение (12) известно как *теорема Боголюбова–Зубарева* [10], а выражение (13) было впервые получено в работе Рудого и Суханова [3]. Выражения (12) и (13) естественно называть *динамическими уравнениями состояния*, поскольку они связывают между собой (квази)динамические величины — давление P и сжимаемость Ψ с основной характеристикой динамической системы — функцией Гамильтона H .

Существенно, что в динамические уравнения состояния (12) и (13) не входит внешний тепловой параметр — температура T , однако эти уравнения содержат явную зависимость от внешнего механического параметра — объема V ; подчеркнем, что все динамические функции H , P и Ψ заданы на фазовом пространстве Γ системы.

Все функции, входящие в (12) и (13), являются аддитивными относительно составляющих систему частиц (и, следовательно, их средние значения пропорциональны числу частиц N). С другой стороны, эти функции имеют различный характер поведения относительно объема V

$$H(q, p) = O(V^0), \quad P(q, p; V) = O(V^{-1}), \quad \Psi(q, p; V) = O(V^{-2}).$$

Действительно, внешний параметр V входит в правые части (12) и (13) только в виде множителя V^{-1} и V^{-2} , поэтому в пределе $V \rightarrow \infty$ (т. е. для «свободной» системы) величины P и Ψ действительно стремятся к нулю, тогда как H остается в этом пределе неизменной⁸.

В условиях теплового контакта динамической системы с термостатом все функции H , P и Ψ становятся *случайными*, и при наличии теплового равновесия

⁸Ясно, что дифференцирование по q и/или p не может изменить характер поведения по V правых частей уравнений (12) и (13).

в соответствии с общим определением (9) можно усреднить обе части динамических уравнений состояния (12) и (13).

Эта процедура приводит к *термодинамическим уравнениям состояния*, причем среднее значение функции P из (12) всегда является термодинамическим средним (определение см. ниже разделы 3 и 4), тогда как для среднего значения функции Ψ из (13) это не всегда так. Заметим, что усреднение уравнения (12) приводит к так называемой *теореме вириала*, или термодинамическому уравнению состояния, связывающему термическое $P(T, V)$ и калорическое $H(T, V)$ уравнения состояния.

Для неидеальной макроскопической системы функция Гамильтона⁹ $H(q, p)$ может быть представлена в виде суммы трех слагаемых: энергии покоя E_0 , кинетической $H_k(p)$ и потенциальной $H_{\Pi}(q)$ энергий. В свою очередь, эти энергии обычно также являются аддитивными по всем частицам (E_0 и $H_k(p)$) и их парам ($H_{\Pi}(q)$). Заметим, что постоянная энергия E_0 не дает вклада в динамические уравнения состояния системы (12) и (13) для давления и сжимаемости, что физически вполне обосновано.

2.2. Однородный случай

В [3] в качестве примера применения формул (10) и (11) был рассмотрен частный случай, когда обе энергии $H_k(p)$ и $H_{\Pi}(q)$ являются *однородными* (в смысле Эйлера) функциями своих аргументов с показателями k и l соответственно. Это означает, что

$$H_{\Pi}(\lambda q) = \lambda^l H_{\Pi}(q), \quad H_k(\lambda^{-l} p) = \lambda^{-k} H_k(p), \quad (14)$$

так, что выражения (12) и (13) принимают вид соответственно

$$P(q, p; V) = \frac{1}{fV} [k H_k(p) - l H_{\Pi}(q)], \quad (15)$$

$$\Delta \Psi(q, p; V) = \left(\frac{1}{fV} \right)^2 [k^2 H_k(p) + l^2 H_{\Pi}(q)], \quad (16)$$

$$\Delta \Psi(q, p; V) \equiv \Psi(q, p; V) - \frac{1}{V} P(q, p; V).$$

«Однородные» выражения (15) и (16) обладают следующими полезными свойствами:

- 1) При любых k и l обе энергии $H_k(p)$ и $H_{\Pi}(q)$ входят в правые части (15) и (16) *линейно*, причем каждое дифференцирование по λ увеличивает на единицу степень показателей k и l , входящих в виде множителей перед $H_k(p)$ и $H_{\Pi}(q)$.
- 2) Физическая размерность выражения (15) для давления соответствует объемной плотности энергии, а выражения (16) для динамической сжимаемости — объемной плотности давления (каждое дифференцирование по λ приводит к увеличению на единицу степени множителя $1/V$).
- 3) Существуют условия, при которых давление $P(q, p; V)$, а также сжимаемость $\Psi(q, p; V)$ пропорциональны полной энергии $H(q, p) = H_k(p) + H_{\Pi}(q)$. При этих условиях согласно (15) и (16) среднее значение $\langle \Psi \rangle$ пропорционально $\langle P \rangle$ и/или $\langle H \rangle$, т. е. является термодинамическим, и его вычисление не представляет никакой дополнительной проблемы. Очевидно, указанные условия выполняются только в двух случаях: при $k = -l$ или при $l = 0$ (само значение k может быть при этом произвольным).

В данной работе мы ограничиваемся случаем *идеальной* динамической системы с $l = 0$, когда отсутствует зависящая от координат потенциальная энергия $H_{\Pi}(q)$ взаимодействия между частицами. Полная энергия $H(q, p)$ такой системы

⁹В дальнейшем мы, как правило, будем опускать явное указание индекса N у динамических величин, тогда как зависимость от N термодинамических величин будет указываться явно.

помимо постоянного слагаемого E_0 содержит зависящую только от импульсов частиц кинетическую энергию $H_k(p)$

$$H_\Pi(q) = 0, \quad H(\Gamma) \equiv H(q, p) = E_0 + H_k(p). \quad (17)$$

В случае, если функция $H_k(p)$ однородна (в смысле Эйлера) с показателем k , выражения (16) и (17) принимают следующий простой вид

$$P(p; V) = \kappa \mathcal{H}_k(p; V), \quad \kappa = \frac{k}{f}, \quad \mathcal{H}_k(p; V) \equiv \frac{H_k(p)}{V}, \quad (18)$$

$$\Delta\Psi(q, p; V) = \frac{1}{V} \kappa P(p; V) = \frac{1}{V} \kappa^2 \mathcal{H}_k(p; V). \quad (19)$$

Отметим, что в выражения (18) и (19) входит постоянная $\kappa = k/f$, равная отношению показателя однородности k к числу степеней свободы f . Величина κ характеризует данную динамическую систему при описании ее динамических и термодинамических свойств как в классическом, так и в квантовом режимах; κ является своего рода «показателем подобия» системы. При заданном значении f величина κ может изменяться в пределах от $\kappa_{\text{нр}} = k_{\text{нр}}/f = 2/f$ до $\kappa_{\text{ур}} = k_{\text{ур}}/f = 1/f$ (см. соотношения (2)–(4)); таким образом, для обычных случаев $1 \leq k \leq 2$, $1 \leq f \leq 3$ для κ имеют место ограничения $1/3 \leq \kappa \leq 2$. Однако в некоторых специальных моделях «идеального газа» интервалы изменения параметров k , f и $\kappa = k/f$ могут быть иными как по величине, так и по знаку (см. в связи с этим сноску³); выражения (18) и (19) сохраняют применимость и для этих случаев¹⁰.

2.3. Неоднородный случай

В данной работе рассматривается идеальный газ (17) с *неоднородной* функцией Гамильтона вида (1), описывающей общий релятивистский закон связи кинетической энергии и импульса. В этом случае выражения (14)–(16) (и, соответственно, (18) и (19)) уже неприменимы, и для получения явного вида динамических уравнений состояния следует вернуться к исходным выражениям (12) и (13).

Введем безразмерную переменную ξ , для которой в соответствии с (1) имеем¹¹

$$h(\xi) = 1 + h_k(\xi) = (1 + \xi^2)^{1/2}, \quad \xi = \frac{cp}{E_0} \quad (E_0 \neq 0), \quad H(\xi) = E_0 h(\xi). \quad (20)$$

Тогда вместо (18) и (19) получаем динамические уравнения состояния следующего вида

$$P(\xi; V) = \frac{E_0}{fV} \frac{h^2(\xi) - 1}{h(\xi)} = \mathcal{H}(\xi; V) \mu^{(-)}(\xi), \quad (21)$$

$$\begin{aligned} \Delta\Psi(q, p; V) &= E_0 \left(\frac{1}{fV} \right)^2 \frac{h^4(\xi) - 1}{h^3(\xi)} = \frac{1}{V} P(\xi; V) \mu^{(+)}(\xi) = \\ &= \frac{1}{V} \mathcal{H}(\xi; V) \mu^{(-)}(\xi) \mu^{(+)}(\xi). \end{aligned} \quad (22)$$

Здесь с учетом определения (18) для $\mathcal{H}_k(p; V)$ введена объемная плотность полной энергии $\mathcal{H}(\xi; V) = (E_0/V) + \mathcal{H}_k(p; V)$. Иногда удобнее вместо (21) и (22) использовать представления, обобщающие (18) и (19)

$$P(\xi; V) = \mathcal{H}_k(\xi; V) \kappa^{(+)}(\xi), \quad \mathcal{H}_k(\xi; V) = \mathcal{H}(\xi; V) f \kappa^{(-)}(\xi), \quad (23)$$

¹⁰Заметим, что если входящая в (18) и (19) плотность энергии положительна, то флуктуации давления также положительны (независимо от знака κ); таким образом, система механически устойчива, даже если давление отрицательно (в случае $\kappa < 0$).

¹¹Случай $E_0 = 0$ точно соответствует ультрарелятивистскому пределу и, следовательно, является однородным с $k = 1$.

$$\Delta\Psi(\xi; V) = \frac{1}{V}P(\xi; V)\mu^{(+)}(\xi) = \frac{1}{V}\mathcal{H}_k(\xi; V)\kappa^{(+)}(\xi)\mu^{(+)}(\xi). \quad (24)$$

Входящие в (21)–(24) безразмерные множители $\mu^{(\pm)}(\xi)$ и $\kappa^{(\pm)}(\xi)$ имеют смысл обобщенных «показателей однородности» и зависят (хотя и весьма слабо) от переменной ξ посредством функции $h(\xi)$

$$\begin{aligned} f\kappa^{(\pm)}(\xi) &= 1 \pm [h(\xi)]^{-1}, \quad f\mu^{(\pm)}(\xi) = 1 \pm [h(\xi)]^{-2}, \quad f\nu^{(\pm)}(\xi) = 1 \pm [h(\xi)]^{-4}; \\ f\kappa^{(+)}(\xi)\kappa^{(-)}(\xi) &= \mu^{(-)}(\xi), \quad f\mu^{(-)}(\xi)\mu^{(+)}(\xi) = f\nu^{(-)}(\xi), \\ h_k(\xi) &= h(\xi)f\kappa^{(-)}(\xi). \end{aligned} \quad (25)$$

Очевидно, что основную роль в динамическом описании системы играет функция $h_k(\xi) = 1 - h(\xi)$, которая обладает следующими свойствами: $h_k(\xi)$ и ее производные $h'_k(\xi) = \xi[h(\xi)]^{-1}$ и $h''_k(\xi) = [h(\xi)]^{-3}$ при всех $0 \leq \xi < \infty$ знакопостоянны (все три — неотрицательны) и монотонны: $h_k(\xi)$ и $h'_k(\xi)$ возрастают от нуля до ∞ и до 1 соответственно, а $h''_k(\xi)$ убывает от 1 до нуля.

Функция $h'''_k(\xi)$ также знакопостоянна (неположительна), но уже не монотонна: она имеет минимум при $\xi = 1/2$ и достигает максимума, равного нулю, в обоих пределах: при $\xi = 0$ и $\xi \rightarrow \infty$. Наконец, функция $h''''_k(\xi)$ не является уже ни знакопостоянной, ни монотонной: она возрастает от минимального значения, равного (-3) при $\xi = 0$, меняет знак при $\xi = 1/2$ и, пройдя через положительный максимум, убывает до нуля при $\xi \rightarrow \infty$.

Представляют интерес разложения для функций $h_k(\xi)$ и $1/h_k(\xi)$ в предельных случаях: в нерелятивистском (^{нр}) пределе ($\xi = 0$) и в ультрарелятивистском (^{ур}) пределе ($\xi \rightarrow \infty$). Низшие поправки к «однородным» формулам (2) и (3) имеют вид

$$\begin{aligned} h_k(\xi) &\approx h_k^{\text{нр}}(\xi) \left[1 - \frac{1}{4}\xi^2 \right] = h_k^{\text{нр}}(\xi) \left[1 - \frac{1}{2}h_k^{\text{нр}}(\xi) \right], \\ h_k^{\text{нр}}(\xi) &= \frac{1}{2}\xi^2 \quad (\xi \rightarrow 0), \\ 1/h_k(\xi) &= [h_k^{\text{ур}}(\xi)]^{-1} \left\{ 1 - \frac{1}{2}\xi^2 \right\} = [h_k^{\text{ур}}(\xi)]^{-1} \left\{ 1 - \frac{1}{2}[h_k^{\text{ур}}(\xi)] \right\}, \\ [h_k^{\text{ур}}(\xi)]^{-1} &= \xi^{-1} \quad (\xi \rightarrow \infty). \end{aligned} \quad (26)$$

Заметим: $h_k^{\text{нр}}(0) = 1/h_k^{\text{ур}}(\infty) = 0$. Это и позволяет рассматривать эти величины как малые в соответствующих областях изменения переменной ξ .

Соответственно, поведение функций $\kappa^{(\pm)}(\xi)$ и $\mu^{(\pm)}(\xi)$ таково: все четыре функции при $\xi \rightarrow \infty$, $h(\xi) \rightarrow 0$ стремятся к одному и тому же предельному значению $\kappa_{\text{ур}} = 1/f$, однако при $\xi = 0$, $h(\xi) = 1$ пределы для них совпадают уже только для каждой из пар: $\kappa^{(\pm)}(0) = \mu^{(\pm)}(0) = \kappa_{\text{ур}} \pm (1/f)$, так, что $\kappa^{(+)}(0) = \mu^{(+)}(0) = \kappa_{\text{нр}} = 2/f$, тогда как $\kappa^{(-)}(0) = \mu^{(-)}(0) = 0$.

Кроме того, пары функций $\kappa^{(+)}(\xi), \mu^{(+)}(\xi)$ и $(\kappa^{(-)}(\xi) = \mu^{(-)}(\xi))$ обладают схожими общими свойствами: обе функции $\kappa^{(+)}(\xi)$ и $\mu^{(+)}(\xi)$ положительны и монотонно убывают с ростом ξ (от $\kappa_{\text{нр}}$ при $\xi = 0$ до $\kappa_{\text{ур}}$ при $\xi \rightarrow \infty$)¹², тогда как обе функции $\kappa^{(-)}(\xi)$ и $\mu^{(-)}(\xi)$ неотрицательны и монотонно возрастают (от нуля при $\xi = 0$ до $\kappa_{\text{ур}}$ при $\xi \rightarrow \infty$).

В ряде физических задач может представлять интерес получение поправок к предельным «однородным» уравнениям (18) и (19), обусловленных переменным характером функций $\kappa^{(\pm)}(\xi)$ и $\mu^{(\pm)}(\xi)$, входящих в динамические уравнения состояния (23) и (24) для $P(\xi; V)$ и $\Psi(\xi; V)$.

Для проведения подобной процедуры в духе теории возмущений при малых значениях $\xi \ll 1$ в нерелятивистском (^{нр}) пределе и больших значениях $\xi \gg 1$ в

¹²Очевидно, в указанных предельных случаях уравнения (23) и (24) переходят в свои «однородные» аналоги (18) и (19).

ультрарелятивистском (yp) пределе удобно использовать в уравнениях (23)–(25) в качестве малого параметра не ξ и $1/\xi$, а величины $h_k(\xi)$ и $1/h_k(\xi)$ соответственно.

Очевидно, даже простейший учет поправок на неоднородность (в смысле Эйлера) функции Гамильтона для общего релятивистского случая приводит к отклонению от присущей однородному случаю линейной зависимости между $P(\xi; V)$ и $\mathcal{H}_k(\xi; V)$, а также между $\Delta\Psi(\xi; V)$ и $P(\xi; V)$.

Сопоставляя формулы (23)–(24) со схожими по структуре «однородными» аналогами (18)–(19), легко видеть, что эти отклонения удобно описывать с помощью пары обобщенных показателей «однородности» ($\kappa^{(+)}(\xi)$, $\mu^{(+)}(\xi)$). Используя разложения (26), нетрудно получить в низшем приближении для нерелятивистского (np) предела

$$\begin{aligned}\kappa^{(+)}(\xi) &= \frac{1}{f} \{1 + [1 + h_k(\xi)]^{-1}\} \approx \kappa_{np} \left[1 - \frac{1}{2} h_k(\xi)\right], \\ \mu^{(+)}(\xi) &= \frac{1}{f} \{1 + [1 + h_k(\xi)]^{-2}\} \approx \kappa_{np} [1 - h_k^{np}(\xi)] \quad (\xi \rightarrow 0); \end{aligned} \quad (27)$$

и для ультрарелятивистского (yp) предела

$$\begin{aligned}\kappa^{(+)}(\xi) &= \frac{1}{f} \{1 + h_k^{-1}(\xi) [1 + h_k^{-1}(\xi)]^{-1}\} \approx \kappa_{yp} \{1 + [h_k(\xi)]^{-1}\}, \\ \mu^{(+)}(\xi) &= \frac{1}{f} \{1 + h_k^{-2}(\xi) [1 + h_k^{-2}(\xi)]^{-2}\} \approx \kappa_{yp} \{1 + [h_k(\xi)]^{-2}\} \quad (\xi \rightarrow 0). \end{aligned} \quad (28)$$

С учетом (27) и (28) соотношения (23) и (24) принимают приближенный вид

$$\begin{aligned}P(\xi; V) &\approx \kappa_{np} \kappa_k \left[1 - \frac{1}{2} h_k(\xi)\right], \\ \Delta\Psi(\xi; V) &\approx \kappa_{np}^2 \mathcal{H}_k(\xi; V) \left[1 - \frac{3}{2} h_k(\xi)\right] \quad (\xi \rightarrow 0); \end{aligned} \quad (29)$$

$$\begin{aligned}P(\xi; V) &\approx \kappa_{yp} \mathcal{H}_k(\xi; V) \{1 + [h_k(\xi)]^{-1}\}, \\ \Delta\Psi(\xi; V) &\approx \kappa_{yp}^2 \mathcal{H}_k(\xi; V) \quad (\xi \rightarrow 0). \end{aligned} \quad (30)$$

Выражения (26)–(30) обнаруживают общую тенденцию к «выполаживанию» зависимостей от ξ как кинематических ($\kappa^{(+)}$, $\mu^{(+)}$), так и динамических (h_k , P , $\Delta\Psi$) величин: при малых (но конечных) ξ все эти величины становятся *меньше* своих «однородных» пределов при $\xi = 0$, тогда как при больших (но конечных) ξ — наоборот, *больше*, чем при $1/\xi = 0$.

Динамические уравнения состояния (21) и (22) представляют интерес с точки зрения получения термического уравнения состояния для давления $P = P(V, T)$, а также для его флуктуаций $\overline{(\Delta P)^2} = +\langle\Psi\rangle$. Для этого необходимо вычисление термодинамических средних значений согласно формуле (9), а также статической суммы (8) с функцией Гамильтона (17). Соответствующие расчёты в настоящее время проводятся и будут нами опубликованы.

Литература

1. Гиббс Д. В. Основные принципы статистической механики. — М.: Наука, 1982.
2. Терлецкий Я. П. Статистическая физика. — 3-е издание. — М.: Высшая школа, 1994. — 350 с.
3. Рудой Ю. Г., Суханов Ю. Г. Термодинамические флуктуации в подходах Гиббса и Эйнштейна // Успехи физических наук. — Т. 170, № 12. — 2000. — С. 1265–1296.

4. Мюнстер А. Термодинамика необратимых процессов. — М.: ИЛ, 1962.
5. Munster A. // Physica. — Vol. 26. — 1960. — P. 1117.
6. Fowler R. H. Statistical Mechanics. — 2nd edition. — Cambridge: Cambridge University Press, 1936. — 350 p.
7. Klein M. J. // Physica. — Vol. 26. — 1960. — P. 1073.
8. Боголюбов Н. Н. Избранные труды. — Киев: Наукова думка, 1971. — Т. 3.
9. Владимиров В. С. Обобщенные функции в математической физике. — М.: Наука, 1976.
10. Zubarev D. N. Неравновесная статистическая термодинамика. — М.: Наука, 1971.

UDC 536.7

Dynamic Pressure and its Fluctuations for the Ideal Gas of Relativistic Particles

Yu. G. Rudoy *, I. Keita †

* Center of Scientific Education of PFUR

† Department of Theoretical Physics
Peoples' Friendship University of Russia
6, Miklukho-Maklaya str., Moscow, 117198, Russia

The analysis of the dynamical quantities, i.e. the pressure and compressibility for the ideal gas of classical particles with the relativistic dispersion law is given. The analysis is based on the generalized Bogoliubov–Zubarev theorem for the quasi-dynamical quantities. It is shown that the account of the relativistic corrections in the both regions of small and large moments may be realized by means of the effective uniformity index

UDC 539.12

Electrogasodynamic Model of Charged Particle Acceleration in Flares of the Main Sequence Dwarf Stars

Yu. S. Kopysov ^{*}, Yu. I. Stozhkov [†]

^{*} *Institute of Nuclear Research of RAS
7a, Prospect of the 60th October Anniversary, Moscow, 117312, Russia*

[†] *P. N. Lebedev Physical Institute of RAS
53, Leninsky Prospect, Moscow, 119991, Russia*

Dwarf stars of the main sequence generate charged particles in stellar flares. In our Galaxy the total number of such flash stars is about 10^{11} . The power and frequency of stellar flares are enough to populate all cosmic ray spectrum up to particle energy 10^{15} eV. Electrogasodynamic model of charged particle acceleration in flares of dwarf stars is considered. The physical mechanism includes the acceleration of particles by electric field behind plasma shock wave originating by stellar flare.

Introduction

Supernova explosions are considered as the main sources of galactic cosmic rays (GCR) [1]. However, our Sun and the stars in the low part of the Main Sequence also accelerate particles in flare processes. The bulk of star population in our Galaxy makes up dwarf stars. The most of them with a later spectral class than the Sun (spectral class G2) belongs to flare-stars. These stars with frequent and powerful flares are much more active in comparison with the Sun [2]. Below it is shown that these flare-stars can provide the necessary energy of GCRs. Maximum energy of particles, which can be obtained in explosion processes, is defined by acceleration mechanism of charged particles. Mechanism of particle acceleration up to $\sim 10^{15}$ eV in the coronal electric fields of flare-stars is considered.

1. Dwarf Stars as GCR Sources

The stellar population of our Galaxy contains $\sim 2 \times 10^{11}$ stars. More than 90% of them belong to dwarf stars. These stars are in the low part of the Main Sequence. The most part of them is much more active than our Sun. In other spiral galaxies many dwarf stars are observed too. For example, the galaxy NGC4565 has halo of numerous red dwarf stars (it was observed in the infrared radiation) [3]. It is probable that our Galaxy has such halo too. The number of dwarf stars in Galactic disc is more than 10^{11} . As these stars have low luminosity, they can be observed at the distances less than $(20-30)pc$. The red dwarfs have spots like sunspots. The ratio of total area of spots on red dwarf to star surface can get tens of percents (sometimes till 90%). For comparison, this ratio for the Sun is $\leq 0.05\%$. Weak global magnetic fields and strong fields of active regions define the solar and dwarf star activity. The strength of the solar global magnetic field is $\leq 1G$ and in active regions it is $\sim 3 kG$. These values for red dwarfs are $(10-20)G$ and several tens kG , respectively. Numerous starspots, strong magnetic fields and their complex structure create favorable situations for frequent and powerful stellar flares. Hydrogen emission (e.g., H_α line) observed during flares is the direct proof of the existence of flare activity on these objects [2].

This work was supported in part by the Neutrino Physics Program of the Russian Academy of Sciences (RAS).

2. Flares on Red Dwarfs and Energy of Cosmic Rays in Galaxy

Our Galaxy has the following characteristics: radius $r_{\text{Gal}} \approx 10 \text{ kpc}$, thickness of optical disc $d \approx 0.3 \text{ kpc}$. The Galactic disc volume $V_{\text{disc}} \approx 2.8 \times 10^{66} \text{ cm}^3$. The GCR energy in Galactic disc plus halo is $W_{\text{CRs}} \approx 0.6 \times 10^{68} \text{ eV}$ (the density of GCR energy was taken $w_{\text{CR}} \approx 0.5 \text{ (eV/cm)}^3$). If GCR contain in Galactic disc only, their energy is $W_{\text{GCRdisc}} \approx 1.4 \times 10^{66} \text{ eV}$. Let us consider stellar flares. We take that the number of active red dwarfs in our Galaxy as $N_{\text{dwarf}} \approx 5 \times 10^{10}$. The energy of one powerful stellar flare is $W_{\text{dwarf}} \approx 3 \times 10^{36} \text{ erg} = 1.9 \times 10^{48} \text{ eV}$ [2]. Let the energy of each stellar flare is $W_{\text{dwarf}} \approx 2 \times 10^{47} \text{ eV}$ (in 10 times less than maximum value) and such flares occur each (3–4) days. About 10% of flare energy is transferred to cosmic rays [2]. During one year the energy released to GCRs from all active red dwarfs is about 10^{60} eV . In our Galaxy cosmic rays spend $\sim (10^7\text{--}10^8)$ years. Taking into account the values given above the total energy released to GCRs from stellar flares equals to $W_{\text{GCRs}} \approx (10^{67}\text{--}10^{68}) \text{ eV}$. As the values discussed are the evaluations, we think that active dwarf stars can provide the total energy of cosmic rays in our Galaxy and halo, $W_{\text{GCRs}} \approx 10^{68} \text{ eV}$. If GCRs are in Galactic disc only, the energy released to GCRs from dwarf stellar flares is all the more enough to provide its total energy in Galaxy.

3. Electric Fields in Solar and Dwarf Star Coronas

As it is known from observations, our Sun accelerates particles up to energies $E \leq 10^{11} \text{ eV}$. The solar magnetic fields of active regions could keep and accelerate particles up to $E \leq 10^{15} \text{ eV}$ [4]. One of the possible ways of particle acceleration is Syrovatsky's mechanism in which the reconnection of magnetic field lines in the flare regions takes place and neutral current sheet is formed [5]. Below we present another mechanism, in which the process of stellar heat energy transformation into electric energy plays the main role. Experiments with laser beams are known. When powerful laser beam strikes crystal, the hot electrons are evaporated from crystal and dense electron cloud is formed. These clouds produce strong electric fields, which, in turn, accelerate protons and nuclei up to energies of several tens of MeV and more [6–8]. In this mechanism up to 50% of laser beam energy can be transformed into electric field accelerating particles [8]. One can expect that such kind of mechanisms can work in solar and stellar flares where the electron temperature is much higher than the temperature of environmental medium. Such electric fields can arise in a quiet stellar atmosphere when electrons have got energy, for example, from acoustic waves or ascending magnetic fields.

Let us start the investigation of solar atmospheric electricity from the simple events when electric charges have spherically symmetric distribution. In this case the Poisson's equation has the form

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d\phi}{dr} \right) = -4\pi q(r), \quad (1)$$

where $q(r) = e \sum Z_a n_a(r)$ is the charge volume density, $n_a(r)$ is the concentration of particles with electric charge Z in the units of electron charge ($Z_e = 1$). We will look for the solution of the equation (1) as

$$\phi_0 = \text{const} \quad \text{at} \quad r \leq r_0, \quad (2)$$

$$\phi(r) = eZ[U_1(r)/r]; \quad U_1(r) = C \sin[k(r - r_0) + \delta] \quad \text{at} \quad r_0 < r \leq r_1, \quad (3)$$

$$\phi(r) = eZ[U_2(r)/r]; \quad U_2(r) = D \exp[-\chi(r - r_1)] \quad \text{at} \quad r_1 \leq r < \infty, \quad (4)$$

$$q(r) = -\frac{eZ}{4\pi r} \frac{d^2 U_1}{dr^2} = \frac{eZ}{4\pi r} = \begin{cases} k^2 U_1(r) & \text{at} \quad r_0 < r \leq r_1, \\ -\chi^2 U_2(r) & \text{at} \quad r_1 \leq r < \infty, \end{cases} \quad (5)$$

where Z is positive electric charge of the solar sphere concentrated at the photosphere level with radius $r_0 \approx R_0$ (R_0 — radius of the Sun), $r_1 \approx 2.3R_0$ is the radius of spherical surface at the boundary between K - and F -corona. Let us take the boundary conditions as

1) Condition of logarithm derivatives U_1 and U_2 equality (at $r = r_1$)

$$\left. \frac{1}{U_1} \frac{dU_1(r)}{dr} \right|_{r=r_1} = \left. \frac{1}{U_2} \frac{dU_2(r)}{dr} \right|_{r=r_1}.$$

It gives the following condition of interfacing two solutions

$$k \operatorname{ctg}[k(r_1 - r_0) + \delta] = -\chi, \quad \delta = n\pi - k(r_1 - r_0) - \operatorname{arctg}\left(\frac{k}{\chi}\right), \quad n = 0, \pm 1, \pm 2, \dots$$

2) Continuity condition for potential ϕ at $r = r_1$

$$U_1(r_1) = U_2(r_2).$$

It gives the relation between constants C and D

$$D = \pm Ck(k^2 + \chi^2)^{1/2}.$$

3) Homogeneity condition for electric field at $r \geq r_0$

$$\left. \frac{d^2\phi(r)}{dr^2} \right|_{r=r_0} = 0.$$

This additional condition gives the relation between k and χ

$$\frac{2kr_0}{2 - k^2r_0^2} + \frac{k + \chi \operatorname{tg}[k(r_1 - r_0)]}{\chi - k \operatorname{tg}[k(r_1 - r_0)]} = 0. \quad (6)$$

From these equations we can get the discrete values of k_m , taking the scale of height in the external F -corona with the use of the parameter chi . These discrete values of k_m correspond to different configurations of electric charge distribution in the inner part of the solar atmosphere. In this case the energy field decrease with the increase of k_m at the same values of χ and Z and the conservation of electro-neutrality condition of the Sun as a star.

4) Electro-neutrality condition.

The density of volume charge q defined by formulas (4) depends on the constant C , which can be found from the electro-neutrality condition. Let us write this condition as

$$eZ + \int_{r_0}^{\infty} dr 4\pi q r^2 = 0. \quad (7)$$

After integration one can find

$$C = (\sin \delta - kr_0 \cos \delta)^{-1}.$$

The solutions do not depend on the value of Z and can be used to describe the electric fields of solar and red dwarf coronas, and of the Earth's atmosphere too. As an example we calculated electric field configuration similar to that of electric field of the Earth. The important parameter of this configuration is the decrement of the field χ at $r \rightarrow \infty$ because χ can be used to define main sizes of the field configuration: $(r_1 - r_0)$ and k_m . For $r_0 = R_0$ and $r_1 = 2.3R_0$ we took $\chi = 0.3543$. Then the first 3 values of k_m equal to $k_0 = 0.8276$, $k_1 = 2.1301$, $k_2 = 4.0823$. The state with $m = 0$ corresponds to monotonic decrease of the field with the distance increase. At $m = 1$ the electric field strength has minimum at $r \approx 2R_0$ and also decreases with the distance increase. At this point the maximum values of potential barrier and electron volume density take place.

4. Solar Atmosphere as Thermoemission Converter of Heat Energy into Electric One and into Energy of Accelerated Particles

In the chromosphere at the altitude $h \approx 2000$ km above the photosphere the sharp increase of temperature takes place and in the corona at $h \approx 10^4$ km the temperature reaches 10^6 K. The heating is made by the dissipation of energy of mechanical and magnetic fluxes arising from the photosphere and by the scattering light quanta on free electrons. Electrons are heated and the high-energy electrons cannot be kept by gravitation field. So, one can expect the flow of such electrons as it takes place in the laboratory conditions. The low energy electrons are kept by gravitational field and produce large-scale electric fields.

Let us write the system of equations for electron gas flow. The continuity equation of “hot” electron flow has the form

$$4\pi e r^2 n_{eh} u_{eh} = I_{eh}, \quad (8)$$

where n_{eh} , u_{eh} and I_{eh} denote volume concentration, velocity and current of “hot” electrons, respectively. Equation of momentum has the form

$$m_{eh} n_{eh} u_{eh} \frac{du_{eh}}{dr} = -\frac{dp_{eh}}{dr} - n_{eh} Z_e e \frac{d\phi}{dr} - \frac{GM_0 m_e n_{eh}}{r^2} + \sum_k P_{eh,k}, \quad (9)$$

where p_{eh} is the pressure of electron gas flow, M_0 — mass of the Sun, G — gravitational constant, ϕ — electric potential (see formulae (2)), $\sum P_{eh,k}$ — total momentum transferring to ions and “cold” electrons in the unit volume. Equation of energy conservation is written as

$$n_{eh} u_{eh} \left(\frac{3}{2} k \frac{dT_{eh}}{dr} - \frac{k T_{eh}}{n_{eh}} \frac{dn_{eh}}{dr} \right) = \frac{1}{r^2} \frac{d}{dr} \left(r^2 \lambda_{eh} \frac{dT_{eh}}{dr} \right) - \frac{3}{2} \nu n_{eh} k (T_{eh} - T_{ec}),$$

where T_{eh} and T_{ec} are the temperatures of “hot” and “cold” electrons, λ_{eh} — coefficient of thermal conductivity of fast electrons, ν — constant characterizing the velocity of heat transfer from “hot” electrons to “cold” ones. There are two types of Parker’s solutions corresponding to electron wind and electron breeze. The large-scale electric fields considered arise from the interaction of several Maxwell’s groups of electrons — running electrons of electron wind and returning electrons of electron breeze. Besides these flows, the fluxes of electrons and ions accelerated by electric field are produced. These fluxes compensate negative electric charge which solar wind carries away. Thus, the current circuit of coronal thermoemission converter is closed and such converter is working at the background of neutral solar wind. The source of energy is the thermal energy of “hot” electrons. The existence of be-Maxwell’s electron spectrum in the solar wind at $r = 1$ a. u. ($T_{eh} = 6.9 \times 10^5$ K and $T_{ec} = 1.25 \times 10^5$ K) is the confirmation of this model [9]. One can consider two stages of the acceleration process. At the first stage the hot electrons’ cloud leads positive ions (process is like the ambipolar diffusion) and electric field is produced between clouds of the hot electrons and ions. Maximum energy which can get a positive ion is $E_m \leq Mc^2 \gamma$, where Mc^2 is the rest energy of ion and γ is the Lorentz-factor of electron cloud. If flare temperature can get $\sim 10^8$ K then the value of E_m will be several tens of MeV. Second stage of the positive ions acceleration is due to the large-scale electric field arising between two clouds. This field will accelerate ions up to energies $E \sim \epsilon L$, where ϵ is the electric field strength and L is the linear scale of electric field. If we take $\epsilon \approx 100$ V/cm and $L \approx 10^{13}$ cm then one has got $E_m \approx 10^{15}$ eV. We suppose that the overturning shockwave front and/or ion-sound turbulence could be responsible for ion injection.

5. Red Dwarfs and Problem of Cold Dark Matter

Hypothesis on red dwarfs as CGR sources involves the problem of cold dark matter (CDM) in Galaxy. The signature of CDM can be high-energy γ -quanta. Neutralino,

as weakly interacting particle, is the widely discussed candidate on the role of CDM particle. The annihilation of neutralino pairs produces high-energy γ -quanta and antiprotons as by-product of this reaction. The new results were obtained by processing data of the detector EGRET [10]. The excess of galactic high-energy γ -quanta (0.1–10 GeV) over background expected in γ -ray spectrum from GCR was found. The authors of the paper consider this result as the indicator of CDM presence in Galaxy (neutralino annihilation, if the neutralino mass is (50–100) GeV). In this case one can expect the excess flux of antiprotons in the energy range of (10–30) GeV. If this excess of antiprotons exists, it will be soon detected by experiment PAMELA [11]. In the case of negative result of PAMELA experiment, the excess of high-energy γ -quanta will testify that flare red dwarfs can be sources of these γ -quanta, and that averaged γ -spectrum has an universal form as it was observed by EGRET γ -telescope.

Conclusion

- The energy released into cosmic rays during the flares on dwarf stars is enough to explain the density of cosmic ray energy observed in our Galaxy, both in Galactic disc and in halo.
- When upper chromosphere and low corona are heated the detachment of electrons from ions takes place and the large-scale electric fields are produced. These fields accelerate charged particles up to high energies. The strongest electric fields arise from solar flares when temperature reaches 10^7 – 10^8 K.

References

1. *Yodh G. B.* Cosmic Rays, Particle Physics and High Energy Frontier // Proceedings of the 29th ICRC. — Vol. 10. — Pune, India: August 3–10, 2005. — Pp. 11–36.
2. *Hershberg P. E.* Solar Type Activity of Main Sequence Stars. — Odessa: Astroprint, 2002. — 688 p. — In Russian.
3. Physics of Space. — Small Encyclopedia. — 1986. — in Russian.
4. *Kopysov Y. S., Stozhkov Y. I.* About Cosmic Ray Sources with Energies up to $(10^{13}$ – $10^{14})$ eV // Proceedings of the 29th ICRC. — Vol. 3. — Pune, India: August 3–10, 2005. — Pp. 141–144.
5. *Podgornyy I. M., Podgornyy A. I.* Acceleration of Particles in Current Sheet // Izvestiya Russian Academy of Sciences, ser. Phys. — Vol. 61, No 6. — 1997. — Pp. 1067–1069. — In Russian.
6. Electron, Photon, and Ion Beams from Relativistic Interaction of Petawatt Laser Pulses with Solid Target / S. P. Hatchet, C. G. Brown, T. E. Cowan et al // Phys. Plasmas. — Vol. 7, No 5. — 2000. — Pp. 2076–2082.
7. Intense High-Energy Proton Beams from Petawatt Laser Irradiation of Solids / R. A. Snavely, M. H. Key, S. P. Hatchet et al // Phys. Rev. Lett. — Vol. 85, No 14. — 2000. — Pp. 2945–2948.
8. *Bochkarev S. G., Bychenkov V. Y., Tikhonchuk V. T.* Investigation of Ion Acceleration at Flying Apart of Laser Plasma Using Boltzman–Vlasov–Poisson Model // Plasma Physics. — Vol. 32, No 3. — 2006. — Pp. 230–247. — In Russian.
9. *Feldman Y., Asbridge J., Beim S., Gosling J.* Solar Plasma and Magnetic Fields. — In the book “Solar Output and its Variation”/ Editor O. R. White. — Colorado Associated University Press. — Boulder, Colorado. — 1977.
10. *de Boer W., Sander C., Zhukov V. et al.* EGRET Excess of Diffusive Galactic Gamma Rays as a Tracer of Dark Matter. — In arXiv: astro-ph/0508617. — V. 1, 29 Aug. 2005.
11. *Picozza P.* on behalf of PAMELA collaboration, Experiment PAMELA // Proceedings of the 20th ECRS. — Lisbon, Portugal: 2006. — in press.

УДК 539.12

Электрогазодинамическая модель ускорения заряженных частиц во вспышках карликовых звёзд Главной последовательности

Ю. С. Копысов ^{*}, Ю. И. Стожков [†]

^{*} *Институт Ядерных Исследований РАН
Россия, 117312, Москва, Проспект 60-летия Октября, 7а*

[†] *Физический Институт РАН им. П.Н. Лебедева
Россия, 119991, Москва, Ленинский проспект, 53*

Во вспышках карликовых звёзд Главной последовательности происходит генерация заряженных частиц. В нашей Галактике полное число таких звездных вспышек составляет около 10^{11} . Мощности и частоты звездных вспышек вполне достаточно для обеспечения спектра космических лучей до энергии частиц 10^{15} eV. Рассматривается электрогазодинамическая модель ускорения заряженных частиц во вспышках карликовых звёзд. Физический механизм включает в себя ускорение частиц электрическим полем, создаваемым плазменной ударной волной, порождённой звездной вспышкой.

UDC 530.145

Resonant Atom Model and the Formation of Valence Bonds

Yu. B. Magarshak

MathTech Inc., New York, USA

The nature of the valence bond is traditionally attributed to the displacement of electronic orbitals in space between interacting atoms. At the same time, a number of experimental facts, in particular, the existence of four-dimensional structural symmetry between the filling of electron shells and periodical properties of elements require a change in the standard valence concept. A model in which valence bonds are caused by the resonant interaction between the shell electrons of one atom and the protons of another atom seems to be more adequate.

Introduction

In 1939, Linus Pauling in his classical book *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry* [1] proposed valence theory according to which (disregarding details) the valence bond is formed due either to the collectivization of the wave function by atoms (covalence bond) or to the transition of the electronic cloud from one atom to another. If such transition is complete, the bond is called ionic; otherwise, it is called polar. In this case, a number of additional effects and phenomena such as the hybridization of orbitals, the electronegativity scale of elements, resonant valence bonds (as, for instance, in the benzene molecule) may appear. However, according to Pauling, the nature of the chemical bond causing these and many other processes is in the aforementioned displacement of the functions. Pauling's valence model is now standard in spite of some obvious difficulties. One of them is presented by coordination numbers [2]. It was found that one electron in some compounds can interact with several atoms (e.g. magnesium and iron atoms interact with four nitrogen atoms in chlorophyll and hemoglobin molecules, respectively), and the number of atoms with which one electron in so-called coordination compounds interacts can reach 12. The explanation that the wave function of an electron localized in one of the atoms is redistributed between a large number of nearby atoms is an obviously strained interpretation and directly contradicts the third of six Pauling's valence rules [1, 3]: "*An electron involved in the formation of a bond cannot be involved in other bonds.*" (which actually was later revised). The explanation of the hydrogen bond in the Pauling model also seems artificial. It is found that the single electron of a hydrogen atom in many molecules (e.g. in protein molecules) can form not only the basic bond, but also the second weaker bond in the direction opposite to the former bond. The location of the stationary wave function of the electron on "crest" between two atoms, which is of key importance, in particular for the structure of proteins, seems to require a more natural explanation. Thus, although the Pauling model (which constitutes the core of the modern valence concept) well explains the structure of many molecules and provides evident representation of the corresponding processes, the further development of valence theory by analyzing data collected for more than half-century since its appearance seems to be natural and necessary, particularly because the nature of the valence bond is of fundamental importance not only in chemistry,

I am grateful to Profs. F. Bogomolov, I. Vodyanoi, L. Gribov and D. Chernavskii for stimulating discussions.

but also in molecular biology, solid state physics, and many other fields of natural sciences.

Another difficulty follows from the analysis of the periodic properties of chemical elements and the structural symmetry of the electronic configurations of atoms. In 1869, Mendeleev [4, 5] showed that the properties of elements periodically repeat. He managed to place elements in a rectangular table consisting of eight groups (in the current interpretation, with valences increasing from +1 in the first group to +7 (or –1) in the seventh group; these groups end with noble gases, which have zero valence in most cases and are not involved in compounds). In addition, Mendeleev predicted the properties of several elements opened later. The main features of the Periodic Table remain unchanged to date. However, some questions remain unanswered. Let us mention two of them. First, why the number of elements on the k -th shell is equal to $2k^2$, i.e. increases parabolically, while the basic part of the Periodic Table (where elements are divided into periods and groups) is rectangular? Second, why the Periodic Table is two-dimensional, while the number of quantum numbers is known to be equal to four? Maybe the structure of the periodicity of the properties of elements simply is flat, because Mendeleev searched it in such form from the very beginning in order to ensure optimal visualization? Whether substantial elements of the structure of periodically repeating properties are lost in such visually clear form? Answers to these questions were given in [6, 7]. The key to the answer is the presence of two coexisting reference systems in the plane of the principal n and orbital l quantum numbers in the set of atoms and elements (as chemists conventionally call atoms involved in valence interaction with each other) [7]. To visualize these structures, it is necessary to specially order the electronic configurations of all known elements at four hierarchical levels (see Fig. 1):

$$\textit{Supercycle} \Rightarrow \textit{Cycle} \Rightarrow \textit{Subshell} \Rightarrow \textit{Chemical element} \tag{1}$$

in contrast to the traditional ordering at three hierarchical levels (see Fig. 2):

$$\textit{Shell} \Rightarrow \textit{Subshell} \Rightarrow \textit{Chemical element} \tag{2}$$

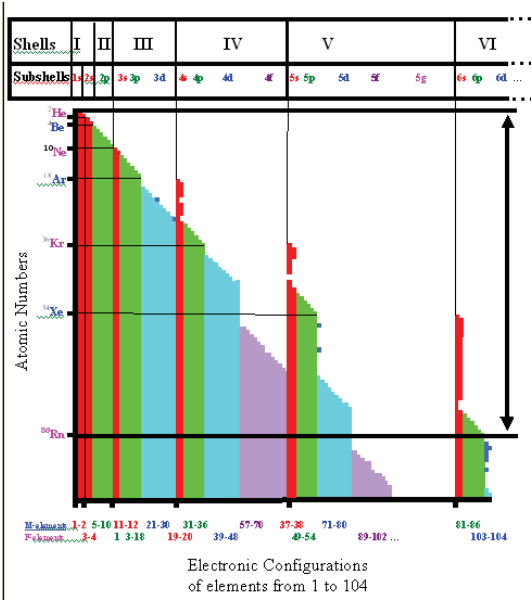


Figure 1. Electron configurations of the first 104 elements of the Periodic Table that are ordered in the following sequence: *Supercycle* $\Rightarrow$ *Cycle* $\Rightarrow$ *Subshell* $\Rightarrow$ *Chemical element*. For the definition of the cycle, supercycle, and the detailed discussion of the structures associated with them, see [1].

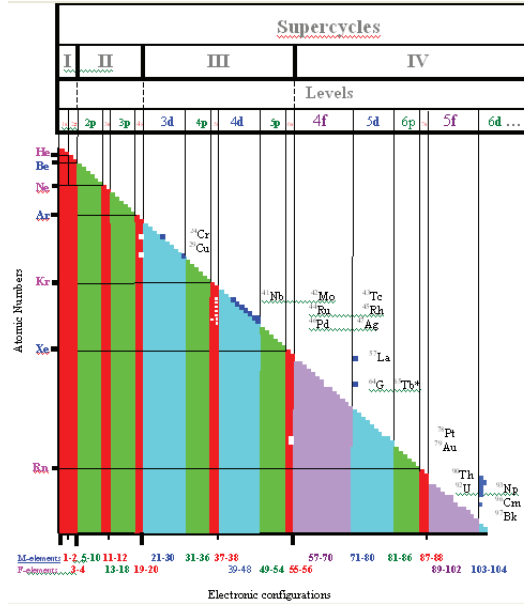


Figure 2. Electron configurations of the first 104 elements of the Periodic Table that are ordered in the following sequence: *Shell* $\Rightarrow$ *Subshell* $\Rightarrow$ *Chemical element* forming a subshell.

No structure is observed in Fig. 2, but it is obviously seen in Fig. 1. It is worth noting that shells are absent among levels forming the structure in Fig. 1 and in relations (1). For this reason, the term *subshell* (i.e. a part of the shell) traditionally used in physics (and almost not used in chemistry) seems to be inappropriate. According to Fig. 1, the subshell is the structure element of the classification of the electronic states of the set of atoms and is the basis for the classification of chemical elements. In [7], subshells were called *cells* in order to emphasize that, first, they are independent of shells and, second, they are structural “blocks” of the “building” of the Periodic Table of chemical elements so as atoms are the structural elements of molecules. Dividing the structure in the upper part of Fig. 1 into several rows and appropriately displacing these rows, one can represent the arrangement of subshells in Fig. 1 in the form [7]

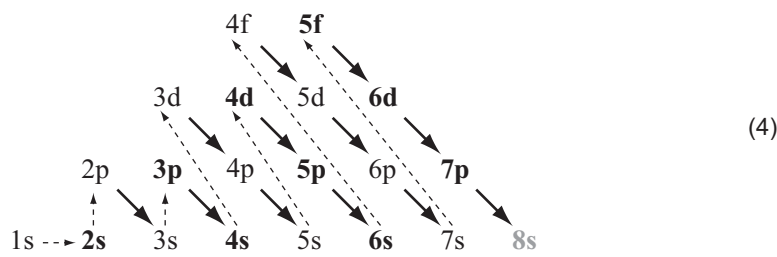
$$\begin{aligned}
 &5f \rightarrow 6d \rightarrow 7p \rightarrow 8s \\
 &4f \rightarrow 5d \rightarrow 6p \rightarrow 7s \rightarrow \\
 &\quad 4d \rightarrow 5p \rightarrow 6s \rightarrow \\
 &\quad 3d \rightarrow 4p \rightarrow 5s \rightarrow \\
 &\quad 3p \rightarrow 4s \rightarrow \\
 &\quad 2p \rightarrow 3s \rightarrow \\
 &\quad 2s \rightarrow \\
 &\quad 1s \rightarrow
 \end{aligned} \tag{3}$$

In this representation, vertical, horizontal, and diagonal structures of indices are clearly seen. It is also seen that every odd row and following even row (from bottom to top) contain the same number of chemical elements and are identically structured in subshells (but the orbital quantum number l in even rows is one unit larger than that in odd rows). The elements entering into a pair of rows with the numbers $2k - 1$ and $2k$, where $k = 1, 2, 3$, and 4 are called **supercycles** [7]. The first (lowest) and second rows of a supercycle are referred to as ***M*** and ***F*** ***cycles***, respectively. The rows of relations (1) are called ***levels***. Thus, all currently known 118 elements form eight levels dividing into four supercycles (rather than seven periods as in the traditional classification of elements by according to Mendeleev), consisting of four ***M*** cycles and

four F cycles. Thus, a pyramid of atomic electron configurations appears with four hierarchically ordered structural levels:

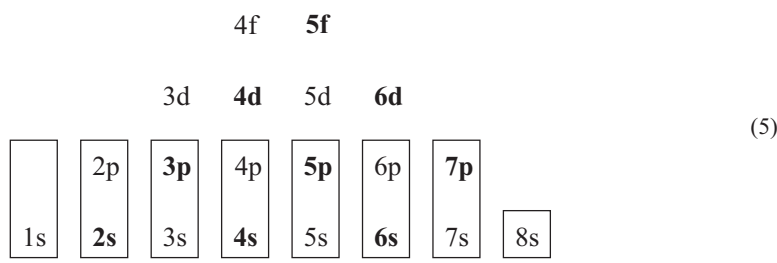
$$\textit{Supercycle} \Rightarrow \textit{Cycle} \Rightarrow \textit{Subshell} \Rightarrow \textit{Chemical element}.$$

In order to pass to the standard quantum-mechanical representation in the (n, l) coordinates, it is necessary “to rotate” relations (1) by an angle of $\pi/4$: where boldface



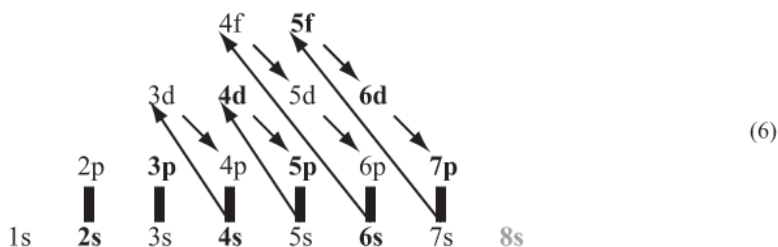
indicates the subshells of F cycles (ending supercycles), solid arrows specify the order of transitions between subshells (cells) in a cycle (also called a level), and dashed arrows indicate the order of transitions between cycles as a function of the charge of a nucleus. According to relations (1) and (2), the set of atomic electron configurations can be structured in two coordinate systems (n, l) and $(n + l, n - l)$.

Let us now analyze the connection between the electronic configurations of atoms and the chemical properties of elements (as traditionally called atoms involved in the valence interaction with each other). Imposing the Mendeleev’s Periodic Table of elements on relations (4) for the electronic configurations of atoms, one obtains [8]



The elements of the Periodic-Table groups from the first to eight that form valence bonds are given in frames. The filling of subshells (cells) inside a cycle occurs in parallel to the main diagonal from top to bottom and from left to right. Cycles are filled along the secondary diagonal of the pyramid from bottom to top and from left to right, beginning with the 1s subshell consisting of hydrogen and helium. Thus, in addition to the four structural levels including supercycles, cycles, subshells, and chemical elements in the $(n + l, n - l)$ coordinates, the set of atoms and chemical elements also has the structure of the valence properties of the chemical elements in the (n, l) coordinates. These structures, which can be represented both in the (n, l) coordinates [relations (5)] and in the $(n + l, n - l)$ coordinates [relations (3)], coexist and supplement each other. If chemical elements forming subshells are represented in relations (5) by introducing an additional dimension, a pyramidal structure appears. Then, separating M cycles from F cycles by displacing the latter, for example, deeper into the figure by half the edge of cubes symbolizing chemical elements, we arrive at the four-dimensional pyramid of atoms and elements (Figs. 3, 4, and 5). However, not all elements are on the surface of the pyramid in these figures. If, for better visualization, the chemical elements belonging to one subshell are represented vertically rather than perpendicularly to the figure, the pyramid of atoms and elements that is similar to skyscraper is obtained (see Fig. 6) [7]. The groups of the Periodic Table are located vertically on the first eight “floors” of the pyramid of atoms and elements. In this “3.5-dimensional” representation, alkali metals and noble gases are located on the first and eighth floors of the step pyramid, respectively.

The pyramid of elements (in which a structure appears both in the (n, l) coordinates and in the $(n + l, n - l)$ coordinates) differs from the Periodic Law not only in the dimension of the structure (four rather than two hierarchical levels), but also in the way of dividing elements into structural blocks. In particular, the second period according to Mendeleev consists of eight elements, whereas the second level of relations (3) (the F cycle of the first supercycle) contains two elements (lithium and beryllium) and the first supercycle includes four elements H, He, Li, and Be [7]. Which of the pyramid-forming coordinate systems, (n, l) or $(n + l, n - l)$, is dominant in chemistry? By the tradition going back to Mendeleev, the periods begin with s terms and end with p terms, whereas the last elements in division into levels are s terms ending a cycle. The division of elements into the periods in the (n, l) coordinates is represented as



The group subshells entering into a period (according to Mendeleev's classification) are connected by thick line segments. They specify the subshells that begin and end periods, whereas the order of the filling of the subshells in the periods is specified by a sequence of arrows that begin on the first floor and end on the second floor. The comparison of the composition of solid arrows in relations (4) (the *levels*, which are also called *cycles*, which are disposed parallel to the main diagonal) with the disposition of arrows in the relations (6), shows that Mendeleev's choice of the beginning and end of the cyclically repeating process seems to be geometrically and logically artificial. It seems more natural to begin the periods with the subshell from which the filling of the cycle starts. However, when quantum mechanics and, later, quantum chemistry appeared, the hybridization of orbitals was found to occur in the (n, l) coordinates on the first two floors of relations (5) in vertical and from bottom to top [see frames in relations (5)] [8]. In particular, this is the way of the hybridization of two s and two p orbitals in the carbon atom. Owing to this hybridization, organic molecules are formed and life becomes possible [9]. Thus, we have to value Mendeleev's intuition who had constructed the Periodic Table in the (n, l) coordinates half-century before the development of quantum mechanics. However, the chemical properties of elements and electron shells can be structured in both coordinate systems, which are called *fundamental*. To date, it is not clear whether the monotonic increase in valences in the Periodic Table from 1 to 7 and the end of the periods with noble gases having eight electrons on the outer shell are due to any physical mechanism or this is simply one of the possible choices of the beginning and end of the cyclically repeating process at which the filling of subshells from the first floor (alkali and alkali earth metals, $l = 0$) in the periods can pass to the *fourth floor* (lanthanides and actinides, $l = 3$), then/or to the *third floor* (transition metals $l = 2$), and only then comes back to the *second floor* ($l = 1$), in groups of which valence begins with 3. In the latter case, **the monotonic increase in valence with the ordinal number of the groups of the Periodic Table is nothing else than an artifact of this choice.**

1. Pyramids of Elements and the Nature of the Valence Bond

In the light of the analysis performed in the preceding section, an additional disadvantage of standard valence theory is obvious. None of relations (3)—(6) follows from Pauling's theory. Moreover, the presence of the rotation by $\pi/4$, the existence of two

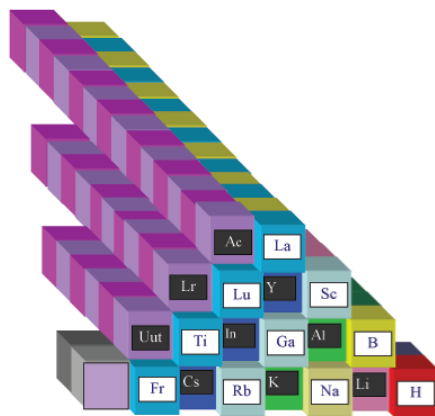


Figure 3. Four-dimensional pyramid of atoms and elements in the level (cycle) representation (cubes representing the elements of the F levels are displaced deeper into the figure by half the cube edge).

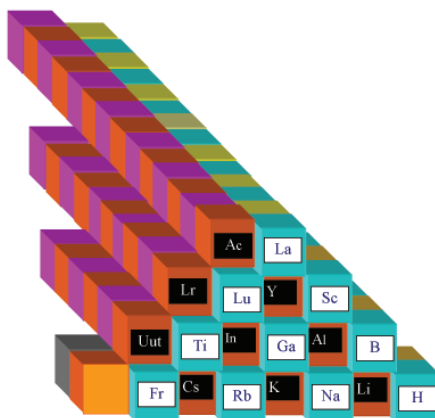
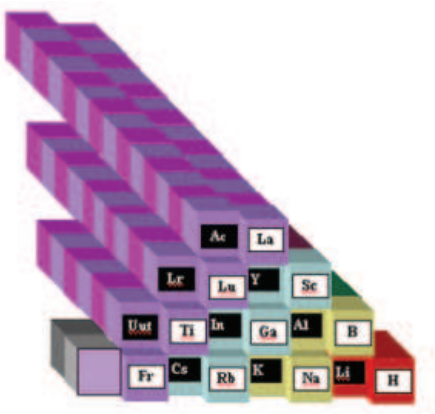


Figure 4. Four-dimensional pyramid of atoms and elements in the cluster representation.



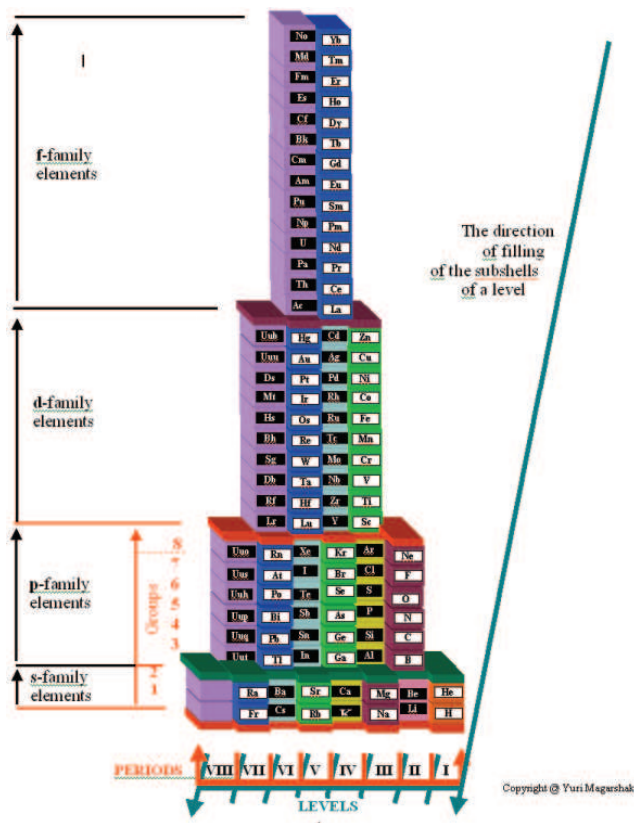


Figure 6. Pyramid of elements in the “3.5-dimensional representation”, in which (in contrast to four-dimensional representation on Figs. 3–5) all elements of the Periodic Table are shown. Colors designate the division of the elements into the periods (according to the tradition going back to Mendeleev). Steps correspond to the division of the elements into levels (cycles) and supersteps (pairs of odd and next even steps) correspond to the division in supercycles. The elements of the M cluster in the 3.5-dimensional representation, as well as in the four-dimensional pyramid, are displaced forward (perpendicularly to the figure plane) by half the cube edge with respect to the corresponding elements of F clusters.

reference systems (n, l) and $(n + l, n - l)$, the presence of the triangular structure in the arrangement at the four hierarchical levels shown in Fig. 1, and the existence of four-dimensional pyramids of elements (Figs 3–5 and Fig. 6) also do not follow from this theory. From the nature of chemical bonds as it is understood now, it is fundamentally unclear why the filling of electron shells occurs in one system of quantum numbers, whereas the valence properties of elements, as well as atomic spectra, are determined in the other system of quantum numbers that is related to the former system by rotation (by the angle $\pi/4$). The model of the valence bond that is due to the hybridization of orbitals and to the displacement of wave functions does not answer the question: “Why does the set of elements in Figs. 1 and 3–5, as well as in relations (3), (5), and (6) (multielectron problems), behave similarly to one electron in the formation of atomic spectra, i.e. so as if all elements are the states of a certain hyperparticle?” To explain all listed phenomena and femtochemical experiments, which will be briefly reviewed below, a physicochemical valence model that differs from the standard model is proposed in this work.

2. Presence of Several Coexisting Symmetries and the Commutation of Operators

As follows from the preceding sections, the set of atomic electron configurations can be structured in two systems of coordinates (n, l) and $(n + l, n - l)$, which are rotated by an angle of $\pi/4$ in the plane of the principal and orbital quantum numbers. This property of the system is not trivial. A question of under which conditions a physical system (and the corresponding differential equations and Hamiltonians describing it) has two coexisting symmetries that do not asymptotically displace one another in the $t \rightarrow \infty$ limit seems to be very important. When searching for an answer to this question, it was proved that, **for two symmetries to coexist in the system described by differential equations (and/or by Hamiltonians), operators generating these symmetries must commute with one another** [8].

A more precise formulation of this theorem is as follows. Let the structures and symmetries that are described above and shown in Figs. 1, 3–5 and 6 be generated by a pair of Hamiltonian-type operators, which naturally appear in the presence of two commutative Lie symmetry algebras in the quantum system. From the existence of simple commutative Lie symmetry algebras in a physical problem, it follows the existence of a single commuting pair of energy-type operators for localized quantum states.

Let us consider a finite-dimensional complex representation of the product of semisimple Lie algebras. Any such representation V of $g_1 \times g_2$ decomposes into a direct sum of tensor products $V_k \times W_j$, where V_k are irreducible representations of g_1 and W_j are representations of g_2 .

Theorem 1. *Assume that g_1, g_2 are simple Lie algebras. Then there is a canonically defined pair of commuting operators Δ_1, Δ_2 such that both $\Delta_i, i = 1, 2$ commute also with the action of g_1, g_2 and have integer nonnegative eigenvalues on V .*

This theorem was proved in [8]. Thus, to explain symmetry between the properties of elements and filling of electron shells within the rotation on the plane of the principal and orbital quantum numbers, **two commuting operators are required**. One of these operators is specified by the Schrödinger equation. Since an atom consists of a shell and a nucleus, which contains protons and neutrons, the only possible generator of the second operator can be the interaction of the electrons of the shell with the protons of the nucleus. Owing to the negligible smallness of the probability of the random appearance of structures (1), (3), (4) and (6), as well as of their appearance due to any “empirical rule” (the term used, in particular, for Hund’s “rule”) rather than a process fundamental for the atomic structure, the hypothesis of the existence of physical mechanisms that generate the second operator and ensure the appearance of these structures seems to have no alternative to explain the above phenomena.

3. Resonant Atom Model

Postulate 1. In atoms there are no stationary electronic states.

Postulate 2. All electrons, which constitute the shell of the atom, are involved in the process with the following basic steps (collectively called **resonant path**):

- (i) The $\psi\beta$ -capture of a shell electron by atom nuclei.
- (ii) Proton–Electron $\rightarrow$ Neutron resonance interaction (**pen-resonance**) of the shell electron with the proton of the nucleus: $e + p \rightarrow n$
- (iii) $\beta\psi$ -release of the electron from the nuclei and its return to nonlocalized ψ -state in the shell.

Postulate 3. The physical nature of $\psi\beta$ -shell electron capture by nuclei, and $\beta\psi$ -electron release from nuclei, is not the same. $\psi\beta$ -capture takes place due to the overlap of the electronic ψ -function with the volume occupied by atom nuclei. $\beta\psi$ -release of the electron from nuclei is possible only at energies determined by Schrödinger equation for the space around the nuclei, in which electronic shell is disposed.

Comments.

- 1) According to the resonance atom model, the stationary Schrödinger equation is valid only because electron-nuclei cyclic processes is very quick, much quicker than the measurement time.
- 2) At present, unidirectional processes of the field interaction between electrons, protons, and neutrons are well known. These processes are, first, K-capture at which an electron from the shell nearest to the nucleus interacts with a proton of the nucleus and, as a result, a neutron appears and the atomic number of the element decreases by one, and, second, β -decay at which the nucleus emits an electron and the atomic number of the element increases by one [2]. According to the resonance atom hypothesis, K-capture and β -decay are comparatively rare irreversible processes that are easily detectable and constitute the “top of the iceberg” of frequent reversible interactions between the proton, neutron, and electron.
- 3) A consequence of quantum mechanics is wave-particle dualism due to which the electron at the time of measurement exhibits corpuscular properties and is located in a volume that is approximately 10^{15} times smaller than the volume occupied by its orbital. On the other hand, the overlap of the electronic ψ -state with the volume occupied by the nuclei can provide the localization of the electron inside the nuclei. Such localization is a necessary condition for the beginning of the *pen*-resonance process. For these reasons, localization of a shell electron in the nucleus, necessary for the (presumably resonant) interaction between proton in the nuclei with shell electron, is undoubtedly possible.
- 4) Other particles (in particular, neutrinos) can be involved in the *pen*-resonance process.

4. Superresonant Nature of the Valence Bond

Postulate 1. In the valence interaction between the atoms in a molecule, an electron on the shell of one of the atoms is involved in the *pen*-resonant interaction with a proton in the nucleus of another atom.

Postulate 2. A valence bond is formed due to jumps of the electron of the shell between the nuclei connected by the valence bond, and each of jumps to a certain nucleus is of *pen*-resonance nature.

Postulate 3. In molecules consisting of more than two atoms, resonances between resonances can occur. The number of the hierarchical levels of resonances in nature is unlimited.

Comments.

- 1) According to the superresonant valence model, the physical nature of the valence bond is the same as the nature of the atom: it is superresonance that is resonance between *pen*-resonances with nuclei participating in the bond formation.
- 2) According to the postulates of the (super)resonant nature of valence, covalent, ionic, and polar bonds, as well as the presence of hybrid orbitals and resonances in molecules, are explained not as in the traditional model. These explanations are as follows.
 - Covalence bonds, in which a common electron pair is formed between atoms, arise when the electron of the outer shell of each of two atoms resonantly interacts with the nucleus of the atom with which the valence bond is formed with the same, or approximately the same, energy and frequency as those of the interaction with the nucleus of the “parent atom”.
 - The ionic bond between two atoms in a molecule is formed when the outer electron of one atom resonantly interacts with the nucleus of another atom and completely leaves resonance with the nucleus of the parent atom.
 - The polar bond appears when the ratio of the probabilities of the resonance localization of the electron in the nucleus of the “parent” atom and in the nucleus of the atom with which the valence bond is formed is between zero and one and differs from $1/2$.

- The hybridization of orbitals is determined by correlation in superresonant jumps of several electrons between the nuclei of the atoms connected by the valence bond.
 - Resonant valence bonds (e.g. in the benzene molecule in which resonance occurs between the single and double bonds) are attributed to the resonance of superresonant jumps of electrons between the nuclei of the atoms which take part in the multi-atomic resonance (for example, between the atoms of the six-membered benzene ring in the case of the benzene molecule).
- 3) Time in which atoms diverge at a distance where the formation of the valence bond is impossible is of the same order of magnitude as the period of atomic vibrations in molecules (10^{-15} – 10^{-12} s). If the **superresonant cycle** (which includes at least two *pen*-resonances with nuclei which form the bond) is several orders of magnitude faster than the upper limit of this range, the electron has more than enough time to form the valence bond that holds the atoms connected by this bond. In a similar way, i.e. by means of the superresonant interactions between the shell electrons of all atoms of a molecule with the nuclei of the atoms connected by valence bonds, the formation of molecules of any size occurs.
 - 4) According to the superresonance model, an electron may not only vibrate between the nuclei of two atoms, but also “travel” between the nuclei of the atoms along a certain trajectory. In particular, the existence of chemical compounds with coordination numbers in which an electron of one atom is connected by valence bonds with several neighboring atoms is explained by the fact that the valence bonding is a fast nonstationary process. If the time of the superresonant bonding of an electron on an atomic shell with each of the atoms coordinationally connected with this coordinating atom is several orders of magnitude shorter than the period of the oscillations of the atoms, the electron coordinating the configuration of the molecule has more than enough time to sequentially couple with all atoms with nuclei of which it superresonantly interacts.
 - 5) The resonance model of the atom suggests the possibility of the existence of the hierarchy of resonances. The first-level *pen*-resonance provides the existence of atoms. The second-level resonance ensure the formation of valence bonds between atoms. The third-level resonances provide the synchronization of the formation of bonds between atomic nuclei (forming, e.g. a ring as in benzene). This process can be continued. There is likely no restriction imposed by nature on the number of hierarchical levels of resonances. For this reason, the application of resonance atom concept, which generate higher levels resonances, to explain the functioning of biological systems, in particular, electronic transport in macromolecules and enzymatic activity, as well as the levels of the organization of matter in living organisms that are higher than biological macromolecules seems to deserve a detailed investigation.
- According to the atomic resonance concept, resonances underlie not only the structures of atoms and valence bond, but also the general structure of matter at all levels beginning with the atomic structure.
- 6) The resonance atom model and superresonant bond model provide an explanation of structural symmetries in relations (4) and (6) between (i) the valence properties of chemical elements, (ii) the properties of electron shells of atoms, and (iii) atomic spectra. This explanation is that the existence of **two** fundamental processes different in nature that underlie the structure of atoms and valence bonds gives rise to the existence two operators (Hamiltonian) that can have properties required for the satisfaction of relations (3), (4) and (6), in particular, can commute with one another [9].
 - 7) Superresonant bonds can provide the basis for the formation of not only molecules, but also solids. However, the discussion of this problem is beyond the scope of this paper.

5. Estimation of the Parameters of the Superresonance Valence Model

A small difference between the rest masses of the neutron and proton, which is less than 1/700 of the mass of these particles, is explained in the framework of the resonance atom model of the structure of matter. Such ratio of the masses is a necessary condition for the possibility of the occurrence of the *pen*-resonance and is determined by the time–energy uncertainty relation [10, 11]

$$\Delta E \Delta t \approx \hbar, \quad (4)$$

where Δt is interval between the beginning and the end of the process and ΔE is related to the rest masses of the electron, neutron, and proton as $\Delta E = c^2 (m_n - m_p - m_e)$. The substitution of the neutron mass value $m_n = 939.5731$ MeV, proton mass value $m_p = 938.2796$ MeV, and electron mass value $m_e = 0.5110034$ MeV, the Planck's constant, and electron charge yields the value $\Delta t \cong 8 \cdot 10^{-22}$ s, which is close to the frequency range in which resonances between elementary particles occur (10^{22} – 10^{24} s $^{-1}$)¹. Thus, **the closeness of the masses of the neutron and proton (the cause of which remains mysterious in many respects) is a necessary condition for the existence of atoms according to the resonance atom model. Moreover, according to the hypothesis of the superresonance nature of the valence bond, this closeness is a necessary condition for the existence of not only atoms, but also molecules and solids.**

What is the frequency of the occurrence of the *pen*-resonance? Light covers the distance from the nucleus to the outer electron orbit in time $\tau = 5.3 \cdot 10^{-11}$ m/($3 \cdot 10^8$ m/s) = $1.8 \cdot 10^{-19}$ s. It is natural to assume that the frequency with which resonance occurs between the electron localized in the nucleus and the electron delocalized on the shell is lower than this value. Thus, the electron is in the state described by the Schrödinger equation for a time that are several orders of magnitude longer than the time for which the electron is in the state localized in the nucleus. This relation completely corresponds to the fact that, in experiments where measurement is performed for a time that are several orders of magnitude longer than the time of fundamental resonances, the electron states are described by the time-independent Schrödinger equation and the number of protons in the nucleus is equal to the number of electrons on the shell. At the same time, for valence bonds to be formed due to superresonances, the superresonance cycle frequency must be larger than the maximum atomic vibration frequency, i.e. in the range 10^{15} – 10^{18} s $^{-1}$.

6. History of the Development of Representations of the Nature of the Valence Bond

Walter Heitler and Fritz London [12] were the first physicists who applied the principles of quantum mechanics to determine the nature of the chemical bond in the hydrogen molecule. According to their model, when atoms approach each other, a negatively charged electron of one atom is attracted to the positive nucleus of another atom due to the Heisenberg exchange energy. As a result, beginning with a certain distance between the atoms, electrons oscillate between the nuclei of the atoms. Thus, according to the model, electrons in the hydrogen molecule belong to both atoms, forming the valence bond, whose length and energy are calculated from the Schrödinger equation.

In 1928, London [13] applied an approach based on the Heisenberg exchange energy to H_3 triatomic molecule. He showed that electrons in this case also oscillate between the nuclei. In the next several years, Polanyi and Wigner [14] and Henry Eyring and Michael Polanyi [15] formulated transition-state theory. In addition, using the London

¹More accurate estimations, in particular, with the inclusion of other particles (e.g. neutrinos), which involvement in the superresonance process might be discovered experimentally in the future, can only refine this estimate, but cannot change it by several orders of magnitude.

equation, Eyring and Polanyi [15] for the first time introduced the concepts of reaction dynamics and potential energy surface. The Arrhenius-like equations, which take into consideration both statistical mechanics and quantum theory, have been derived. The chemical reaction rate k in the van't Hoff equation [16] was expressed in terms of the Boltzmann constant k_B , Planck's constant h , absolute temperature T and Gibbs free energy of activation ΔG^{**} . It was shown that the maximum rate $k_{\max}$ of a chemical reaction (which was defined in transition-state theory as the frequency of penetration through the potential barrier)

$$k = k_{\max} \cdot e^{-\Delta G^{**}/RT} = k_B T/h \cdot e^{-\Delta G^{**}/RT} \quad (5)$$

limited by the pre-exponential factor in Eq. (5), is equal to about $k_B T/h$ [17]. Its value $6 \times 10^{12} \text{ s}^{-1}$ at room temperature is in the range of atomic vibration frequencies in molecules (from 0.01 ps to 0.1 ps). The hopping frequency of the electron forming the valence bond between the nuclei was estimated in these works as $> 10^{12} \text{ s}^{-1}$.

In 1929, Pauling published a work in which the concept of the hybridization of orbitals was introduced for the first time in application to the carbon atom. In 1931, He formulated six rules valid for valence bonds [1]:

- (i) A valence bond is formed as a result of the interaction of uncoupled electrons, one of each atom.
- (ii) The spins of the electrons forming the valence bond should be oppositely directed.
- (iii) The electron that forms a bond cannot be involved in other bonds.
- (iv) One and only one wave function from each atom is involved in the formation of a bond.
- (v) The electron having the minimum energy forms the strongest valence bond.
- (vi) Among two orbitals of an atom, an orbital that has the largest overlapping with the atomic orbital of another atom forms the stronger bond, which is directed toward the overlapping of the orbitals.

The comparison of rules (i)–(iii) (based on works by Heitler and London) with rules (iv)–(vi) (which opened by Pauling) obviously shows that Pauling oriented to the needs of practical chemistry, where it was difficult to apply the conceptual results obtained by Heitler and London. Simultaneously with the work published by Pauling in 1931, Slater published paper [18] in which he tried to justify a model of the valence bond that was based on the fundamental physical principles and was similar to Pauling's model. In the 1930s, Pauling published seven papers under the general title *The Nature of the Chemical Bond*, which have become classical and, considered together, created the basis of the present canonical chemical bond model. Those papers are as follows.

- *The Nature of the Chemical Bond. Application of Results Obtained from the Quantum Mechanics and from a Theory of Paramagnetic Susceptibility to the Structure of Molecules.* April 1931.
- *The Nature of the Chemical Bond. II. The One-Electron Bond and the Three-Electron Bond.* September 1931.
- *The Nature of the Chemical Bond. III. The transition from One Extreme Bond Type to Another.* March 1932.
- *The Nature of the Chemical Bond. IV. The energy of Single Bonds and the Relative Electronegativity of Atoms.* September 1932.
- *The Nature of the Chemical Bond. V. The Quantum-Mechanical Calculation of the Resonance Energy of Benzene and Naphthalene and the Hydrocarbon Free Radicals.* June 1933.
- *The Nature of the Chemical Bond. VI. The Calculation from Thermochemical Data of the Energy of Resonance of Molecules Among Several Electronic Structures.* August 1933.
- *The Nature of the Chemical Bond. VII. The Calculation of Resonance Energy in Conjugated Systems.* October 1933.

In each of these works, practically useful rules were proposed and they were not all be derived from the fundamental principles of quantum mechanics or from the Schrödinger equation. The most important of them are as follows.

- a) The introduction of the concept of hybrid bond.

- b) The introduction of the concept of electronegativity, which for the first time allowed chemists to predict the energy of a chemical bond arising between elements by using the tables composed by Pauling instead of using the Schrödinger equation and wave functions.
- c) The introduction of the concept of resonant structures in which the resonance occurs between single and double valence bonds. By the example of the benzene molecule, it has been shown how resonance occurs between five canonical conformations. According to Pauling, the properties of a resonant structure are the average properties of states between which resonance occurs. (*The Nature of the Chemical Bond. V. The Quantum-Mechanical Calculation of the Resonance Energy of Benzene and Naphthalene and the Hydrocarbon Free Radicals*).
- d) The introduction of the concept of resonance between the ionic and covalence bonds. Owing to this resonance, valence bonds can cease to be integer (single, double, or triple) and can have intermediate values (*The Nature of the Chemical Bond. VI. The Calculation from Thermochemical Data of the Energy of Resonance of Molecules Among Several Electronic Structures*).
- e) The concept of stabilizing resonance in the conjugate systems, in which single and double bonds alternate (*The Nature of the Chemical Bond. VII. The Calculation of Resonance Energy in Conjugated Systems*) and a number of practical rules for calculating the parameters of such systems have been proposed.

These results were summarized in classical book [1] published in 1939. Many of rules proposed by Pauling, which appeared to be highly efficient in applied chemistry, were based on semiempirical reasons and their justification continues to be criticized. Nevertheless, Pauling's theory of chemical bonding is considered now as standard to a much greater degree than Pauling might think, developing it step by step.

An alternative approach to the nature of the chemical bond was developed by Friedrich Hund, Born's pupil, who studied atomic spectra in 1926 and opened the order of the filling of electron shells with changing in the charge of a nucleus (Hund's rule [19]). Studying molecular spectra, Hund found that molecular emission and absorption spectra are in many respects similar to the spectra of free atoms. As a result, Hund and Mulliken [20–26] proposed a model of molecular orbitals according to which electrons are located on the surface of a molecule. Hund and Mulliken represented a molecule as an atom with several nuclei that are separated in space and surrounded by an electron shell common for all nuclei. Thus, in the Hund & Mulliken model of the molecular structure, the difference between atoms and molecules is not radical. This concept was significantly different from the Heitler–London–Pauling model; however, it was confirmed in molecular spectroscopy experiments.

Despite significant differences between the Heitler–London–Pauling valence bond model and the Hund–Mulliken representation of molecules as atoms with the separated nuclei surrounded by molecular orbitals, both results can be derived from the Schrödinger equation as it was shown in the next decade. At present, the concepts of molecular orbitals (proposed by Hund and Mulliken) and valence based on works by Pauling, coexist.

Experimental data collected in the last decade due to the creation of femtosecond lasers and to the development of femtosecond transition-state spectroscopy force us to return to the problem of the nature of bonding between atoms in molecules. In particular, the presence of resonance transitions between covalent and ionic configurations in the atoms of the NaI and NaBr molecules was shown in the real-time experiment [27, 28]. Thus, the concept that the chemical bond is stationary, follows from the time-independent Schrödinger equation, and is caused by the stationary states obtained by solving this equation appears to be in doubt, although it is presented as standard in all school and high school textbooks and most monographs. Investigation of the rupture of the valence bond in real time is another area of fundamental interest. Using femtosecond spectroscopy, A. Zewail [28] analyzed the rupture of the I-CN bond in the ICN molecule. This analysis made it possible to experimentally follow the trajectory of the reaction, including the dynamics of the transition through a potential barrier. The irradiation of molecules by femtosecond pulses with a length of about 0.05 Å (the characteristic sizes of analyzed molecules are Angstroms) in order

to generate and analyzes Single-Molecule-Type Dynamics is the third research area that seems to be most promising for the analysis of the nature of the chemical bond. The action of such short pulses whose length is less than one hundredth of the atomic size almost does not perturb the ground state except for a small part of a molecule. This circumstance allows the real-time analysis of the propagation of perturbations in the molecule [28, 29]. Zewail and his coworkers believe that the creation of attosecond (10^{-18} s) lasers will make it possible to follow the trajectory of the electron during chemical reactions in real time [28]. Thus, direct experimental proofs of the validity or invalidity of the valence-bond models both reviewed above and proposed in this work can be obtained in the next decade.

7. Comparison of the Superresonance Valence Model with the Heitler–London, Hund–Mulliken, and Pauling Models

In the first decade after the development of quantum mechanics, Heitler, London, Polanyi, Hund, Mulliken, Slater, Wigner, and others seriously tried to explain chemistry using the fundamental physical principles. Nevertheless, the conceptual gap between physics and chemistry exists to date. The superresonance model includes a number of elements of the previous models.

- 1) As well as in the Heitler–London model, the valence bond is based on the motion of an electron between nuclei with very high frequency. However, in contrast to the Heitler–London model, the superresonance model of the formation of molecules attributes this motion not to the Heisenberg exchange energy, but to the *pen*-resonance between elementary particles, with a frequency characteristic of resonances between particles (10^{22} – 10^{24} s $^{-1}$) which is a number of orders of magnitude higher than Heitler–London evaluation.
- 2) As well as in the work by Hund and Mulliken, the structure of molecules in the superresonant model is determined by nuclei. Therefore, the consideration of molecules “*as atoms with more than one nucleus*” in both models has both physical and chemical senses. However, the physics of the formation of the electron shell in the superresonant valence model — the resonant interaction of electrons with protons in a nucleus — differs from the physical representations underlying the Hund–Mulliken model.
- 3) The superresonant bond model, as well as Pauling’s rules, provides the possibility of the occurrence of resonance structures. However, in contrast to the standard model, the superresonant valence model considers resonances not only as the cause of the formation of resonant bonds in some molecules (such as the benzene molecule), but also as the physical basis of the chemical bond itself. Moreover, according to the model proposed in this work, resonances constitute the physical mechanism that is responsible for the appearance and existence of not only molecules, but also atoms.

Discussion

- 1) The efficiency of the Occam’s razor principle (according to which new concepts should be introduced only if they are absolutely necessary) has been manifested during centuries in various sciences. The introduction of new models and concepts is not required if the existing representations successfully explain the basic experiments. However, too many questions concerning mechanisms of the formation of chemical compounds are collected. The consideration of alternative models becomes necessary. The superresonance hypothesis of the nature of valence seems to be most adequate and “soft” among all mechanisms listed in the review. On the one hand, it can hold the quantum-mechanical results, which are described by the Schrödinger equation whenever one is applicable (in particular, in the adiabatic approximation). On the other hand, the efficiency of the application of

Butlerov's "hyphens" becomes understandable, because electrons (according to the superresonant valence model) are involved one by one in *pen*-resonance with the nucleus, forming an inter-nucleic "bridge". Thus, the efficiency of the widespread chemical representation of atoms as spheres and points, which proved to be highly effective in numerous applications, is directly explained in the physics of resonant processes underlying (according to the superresonance hypothesis) the formations of molecules. The superresonant model most naturally explains the hybridization of orbitals, the hydrogen bond, and coordination numbers. Femto-chemical results, which directly demonstrate the inadequacy of the representation of the chemical bond as a stationary process, are completely consistent with the existence of the fundamental resonances underlying the nature of the valence bond. Moreover, the existence of structural symmetries in relations (3)–(6), as well as the representation of the set of electronic configurations and the periodical properties of chemical elements in the form of four-dimensional pyramids, can be strictly explained in the framework of the superresonance model.

- 2) According to the superresonant valence model, the interaction between atoms in molecules is caused either by the resonant interaction of valence electrons with the nuclei of both valently connected atoms (covalent and polar bonds) or by the complete transition of an electron of one atom to interaction with the nucleus of the other atom (ionic bond). In the latter case, the atoms become oppositely charged, which is equivalent to the appearance of the Coulomb interaction between the centers of these atoms in the classical approach. This fully corresponds to the empirical representation of the atomic structure that is accepted by chemists whose practical experience forces them skeptically estimate the utility of the use of quantum-mechanical calculations in chemistry. However, in the light of the proposed superresonant valence model, the skepticism by practical chemists concerning the adequacy of the current way of the application of atomic theory in their science seems to be completely justified.
- 3) According to the resonance hypothesis, the structure of matter is based on resonances, resonances between resonances, interaction between resonances, and the hierarchy of resonances. At the most general understanding level, such representation is similar, on the one hand, to the ideas underlying string theory and, on the other hand, to philosophical representations of the harmonic structure of the universe beginning with Plato and Pythagoras. However, string theory is applied to a scale of about 10^{-35} m, whereas the characteristic sizes of atoms and molecules are 10^{25} – 10^{27} times larger.
- 4) Resonance model of the structure of atoms and molecules establishes a straight connection between chemistry and the theory of elementary particles, molecular biology, and nuclear resonance states. The role of relativistic effects in the superresonant process may also be essential [9].
- 5) The increase in entropy is now explained by the fact that nearby trajectories at multiple atomic collisions diverge exponentially such as it is shown both analytically and in computer simulation for the Sinai billiard [30] and Bunimovich stadium [31]. However, the problem of the consistency between the time reversibility of the Newtonian equations of motion and the irreversibility of the entropy increase, strictly speaking, remains. If at least some colliding atoms interact not only as elastic spheres (traditional model in the kinetic theory of gases), but also *pen*-resonantly, these collisions cease to be reversible under time reversal in the Newtonian equations of motion, already because the duration of the *pen*-resonance is a random variable with dispersion. Moreover, since time of the *pen*-resonance is small but nonzero, this duration is enough for the interacting nuclei to move in space. Therefore, the model of collisions between atoms as elastic spheres is inapplicable at least for collisions at which *pen*-resonance occurs (whose energy, in contrast to valence bonds in molecules, is insufficient to hold atoms). Thus, if *pen*-resonances occur at least in some atomic collisions, the entropy of the system increases.
- 6) The superresonance model of valence bonds forces us to revise both understanding and modeling of the enzymatic activity and functioning of biological molecules in

vivo. The traditional bond model compels specialists in molecular-biology simulations to reduce molecular dynamics, including the enzymatic activity, to the Coulomb law and Newton's law of gravity with the selection of the corresponding coefficients. This approach, although numerically effective, seems to be inadequate and too simplified. Representation that bonds between atoms are caused by resonances, as well as the possibility of the existence of hierarchical resonances in molecules, provides a new sight and new opportunities both for explaining existing processes and for designing molecules with given properties. Possible connection of enzymatic activity with the theory of elementary particles (on which proton-electron-neutron-resonances are based according to the proposed model) is equally intriguing and stimulating.

References

1. *Pauling L.* The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry. — Ithaca: Cornell Univ. Press, 1939.
2. *Levine I. N.* Quantum Chemistry. — 5th ed. edition. — Englewood Cliffs: Prentice Hall, 2000.
3. *Pauling L.* The Nature of the Chemical Bond // J. Am. Chem. Soc. — Vol. 53. — 1931. — Pp. 1367–1400.
4. *Mendeleev D.* On the Relationship of the Properties of the Elements to Their Atomic Weights // Zh. Russ. Fiz.-Khim. O-va. — Vol. 1. — 1869. — Pp. 60–77.
5. *Mendeleev D.* On the Relationship of the Properties of the Elements to Their Atomic Weights // Z. Chem. — Vol. 12. — 1869. — Pp. 405–406.
6. *Magarshak Y., Malinsky J.* A Three-Dimensional Periodic Table // Nature. — Vol. 360. — 1992. — Pp. 114–115.
7. *Magarshak Y.* Four-Dimensional Pyramidal Structure of the Periodic Properties of Atoms and Chemical Elements // Sci. Isr. Technol. Adv. — Vol. 7, No 1, 2. — 2006. — Pp. 134–150.
8. *Bogomolov F., Magarshak Y.* On Commuting Operators Related to Asymptotic Symmetries in the Atomic Theory // Sci. Isr. Technol. Adv. — Vol. 7, No 3. — 2006.
9. *Pauling L.* The Shared-Electron Chemical Bond // Proc. Natl. Acad. Sci. USA. — Vol. 14. — 1928. — Pp. 359–362.
10. *Mandelshtam L. I., Tamm I. E.* The Uncertainty Relation between Energy and Time in Nonrelativistic Quantum Mechanics // Izv. Akad. Nauk SSSR, Ser. Fiz. — Vol. 9. — 1945. — Pp. 122–128.
11. *Landau L. D., Lifshitz E. M.* Nonrelativistic Quantum Mechanics. — Moscow: Nauka, 1974.
12. *Heitler W., London F.* Wechselwirkung neutraler Atome und homopolare Bindung nach der Quantenmechanik (Interaction Between Neutral Atoms and Homopolar Binding According to Quantum Mechanics) // Z. Phys. — Vol. 44. — 1927. — Pp. 455–472.
13. *London F.* Probleme der Modernen Physik / Ed. by S. Herzl. — Leipzig: Sommerfeld Festschrift, 1928. — Pp. 104–113.
14. *Polanyi M., Wigner E.* // Z. Phys. Chem. A. — Vol. 139. — 1928. — P. 439.
15. *Eyring H., Polanyi M.* On Simple Gas Reactions // Z. Phys. Chem. B. — Vol. 12. — 1931. — P. 279.
16. *van't Hoff J. H.* Etudes de Dynamiques Chimiques. — Amsterdam: F. Muller and Co., 1884. — P. 114.
17. *Arrhenius S.* On Simple Gas Reactions // Z. Phys. Chem. — Vol. 4. — 1889. — P. 226.
18. *Slater J. C.* Directed Valence in Polyatomic Molecules // Phys. Rev. — Vol. 37. — 1931. — P. 481.
19. *Hund F.* Linienspektren und Periodisches System der Elemente. — Berlin: Springer, 1927.
20. *Hund F.* // Z. Phys. — Vol. 52. — 1929. — P. 606.

21. *Mulliken R. S.* // *Rev. Mod. Phys.* — Vol. 2. — 1930. — Pp. 99–100.
22. *Mulliken R. S.* // *Phys. Rev.* — Vol. 33. — 1929. — P. 507.
23. *Mulliken R. S.* // *Phys. Rev.* — Vol. 30. — 1927. — P. 138.
24. *Mulliken R. S.* // *Phys. Rev.* — Vol. 33. — 1929. — P. 738.
25. *Mulliken R. S.* // *Phys. Rev.* — Vol. 36. — 1930. — P. 1449.
26. *Hund F.* // *Z. Phys.* — Vol. 63. — 1930. — Pp. 719–751.
27. *The Chemical Bond: Structure and Dynamics* / Ed. by A. H. Zewail. — Boston: Academic, 1992.
28. *Zewail A. H.* *Femtochemistry: Atomic-Scale Dynamics of the Chemical Bond* // *J. Phys. Chem. A.* — Vol. 104. — 2000. — Pp. 5660–5694.
29. *Femtochemistry and Femtobiology: Ultrafast Events in Molecular Science* / Ed. by M. M. Martin, J. T. Hynes. — New York: Elsevier, 2002.
30. *Sinai Y. G.* *Dynamical Systems with Elastic Reflections. Ergodic Properties of Dispersing Billiards* // *Russ. Math. Surveys.* — Vol. 25, No 2. — 1970. — Pp. 137–189.
31. *Bunimovich L. A., Sinai Y. G.* *On a Fundamental Theorem in the Theory of Dispersing Billiards* // *Math. USSR, Sb.*19. — Vol. 3. — 1973. — Pp. 407–423.

УДК 530.145

Резонансная модель атома и формирование валентных связей

Ю. Б. Магаршак

МатТек Инк., Нью-Йорк, США

Природу валентных связей традиционно объясняют смещением электронных орбиталей в пространстве между взаимодействующими атомами. В то же время ряд экспериментальных фактов, например, существование четырехмерной симметрии между заполнением электронных оболочек и периодическими свойствами химических элементов, требует пересмотра стандартной концепции валентности. Модель, в которой валентные связи вызываются резонансным взаимодействием между электронной оболочкой одного атома и протонами другого атома, представляется более адекватной.

UDC 531

Lorentz Group and it's Role in the Non-Relativistic Atom Model

F. Bogomolov^{*}, Yu. B. Magarshak[†]

^{*} Courant Institute of Mathematical Sciences of New York University

[†] MathTech Inc., New York, USA

In this article we discuss new symmetry in the atomic structure which was discovered by the second author. We also consider mathematical and physical arguments, which can potentially explain these symmetries.

Introduction

Let us first remind the description of the symmetry. As it was shown in [1], there is a symmetry between the order of completion of the electronic levels of different atoms [2] and the spectra of the hydrogen atom [3–7]. Let us consider the multielectron problem of the electronic shells and subshells structure as a function of atomic number. If one orders the electronic configuration in the coordinates (n, l) , e.g. in the same coordinates in which were ordered the energy levels of the hydrogen atom (relations (1)), the picture obtained does not reveal any structure or regularity. However, if the subshells are ordered differently, one gets clear-cut triangular structure. This ordering as a function of principal and orbital quantum numbers, might be restructured in the form [1].

$$\begin{aligned}
 &5f \rightarrow 6d \rightarrow 7p \rightarrow 8s \\
 &4f \rightarrow 5d \rightarrow 6p \rightarrow 7s \rightarrow \\
 &\quad 4d \rightarrow 5p \rightarrow 6s \rightarrow \\
 &\quad 3d \rightarrow 4p \rightarrow 5s \rightarrow \\
 &\quad \quad 3p \rightarrow 4s \rightarrow \\
 &\quad \quad 2p \rightarrow 3s \rightarrow \\
 &\quad \quad \quad 2s \rightarrow \\
 &\quad \quad \quad 1s \rightarrow
 \end{aligned} \tag{1}$$

The shell and subshell structure in the relations (1) possesses a clear cut symmetry along both main and secondary diagonals, as well as along the horizontal line and the vertical one. This means that relations (1) are structured both in the coordinates (n, l) and in the coordinates $(n+l, n-l)$. Note that odd line number $2k-1$ and even line $2k$, $k = 1, 2, 3, 4$ in (1) (counting bottom-up) have the same structure. The atoms (chemical elements), which constitute the pair of lines $2k-1$ and $2k$ in relations (1) have been called supercycle [1]. Hence there are four supercycles, consisting of 4, 16, 36 and 64 elements of the periodic table [8]. One should notice also the formal tie between relations (1) (which appear from the Schrödinger equation solution for the hydrogen atom [3–7]) and relations (1), which do represent the ordering of the electronic configuration of all atoms [9, 10]. In other words, one electron problem is structured in the coordinates (n, l) , while the multi electron problem (the electronic configurations of all atoms) becomes structured in the coordinates $(n+l, n-l)$. Two pairs of axes (n, l) and $(n+l, n-l)$ are turned on the plane of the orbital and the principal quantum numbers by $\pi/4$ relative to each other [1]. Let's now compare

2. Coexistence of Symmetries and Lorentz Group

Note that any element of the center of $U(g_i)$ has similar properties. If the algebra g_i has rank r (dimension of the Cartan subalgebra) then the central subalgebra is a polynomial algebra with r generators. If the rank of g_i is > 1 then we have other similarly defined canonical operators coming from g_i . However Casimir operator is the only operator in the center of $U(g_i)$, g_i — simple Lie algebra, which is equivalent to second order differential operator (Laplace operator) in natural representations of g_i . These representations appear on the spaces of special functions (or sections of vector bundles) over smooth manifolds with geometric action of the algebra g_i .

The above decomposition may exist in the complexification of the symmetry algebra and not in the algebra itself. Since complex finite-dimensional representations of the complexified Lie algebra are the same as of its real form, the above decomposition holds anyway. In our application the complexification of both real algebras are isomorphic to $so(3, C) = sl(2, C)$ after a complexification and hence have rank 1. Thus there are two natural options:

- 1) both algebras are coming from the independent rotation invariance of both the nucleus and the electron envelope;
- 2) the local symmetry group is the Lorentz group $SO(3, 1)$ and though its Lie algebra $so(3, 1)$ has no such a decomposition but the complexification $so(3, 1) \times C = sl(2, C) + sl(2, C)$. The actual Hamiltonian operator of the problem splits into a sum of two commuting energy time operators. In our opinion second case is physically more plausible since it opens a possibility to connect the problem with a Lorentz group and its nonramified double cover $Spin(3, 1)$ which are natural infinitesimal symmetry groups of general relativity.

It is a well established fact that one-electron problem is governed by the representations of the natural rotation group $Spin(2)$ and the corresponding level structure can be completely deduced from the representations theory of this group. In the multielectron problem we can see the presence of a similar but more complex structure. One of its demonstrations is the natural decomposition of all possible types of atoms into groups which are naturally organized into supercycles [1]. The numerical structure of the latter reveals a connection with representation theory of the product of groups $Spin(3) \times Spin(3)/Z_2 = SO(4)$ with an additional symmetry between both $Spin(3)$ -components and nontrivial representation on the center $Z_2 \subset Spin(3)$. Indeed, any supercycle has length $(2n)^2$ for $n = 1, 2, 3, 4$ and complex irreducible representation of $Spin(3)$, which are nontrivial on the center have even dimensions $2n$.

Thus the size of the supercycles corresponds to dimensions of first irreducible representations of $Spin(3) \times Spin(3)/Z_2$, which are symmetric with respect to both $Spin(3)$ components of the above group. This is hardly a simple coincidence and we would like to discuss possible physical and mathematical reasons for such a structure.

Note that symmetries (1) and (2) are observed within the set of atomic structures which are quantum objects. The appearance of complex valued Ψ -functions and complex representations of the symmetry groups identifies the spaces of complex (though not unitary!) representations of the groups $SO(4)$ and $SO(3, 1)$ where the latter is the Lorentz group of the relativity theory.

The structure of irreducible finite-dimensional representations of the group of real matrices $Spin(3, 1)$ and its Lie algebra is the same as for the Lie algebras $sl(2, C) + sl(2, C)$ (or $so(3) + so(3)$) only if we consider finite-dimensional representations. Physically this is the case of the interactions for ensembles localized within small spatial domains like atom. However the class of infinite-dimensional irreducible unitary representations of the group $Spin(3, 1)$ is substantially bigger than the class of representations of $so(3) + so(3)$. Thus the additional symmetry appears only in the description of *a priori* localized systems of particles where the whole interaction process is restricted to such a domain and space where quantum effects are dominating. The representations of $Spin(3, 1)$ which appear on the corresponding system are of finite dimension. However this effect does not appear when we consider nonlocal interaction

problem. Corresponding irreducible representations of $Spin(3, 1)$ are infinite dimensional. In particular, this symmetry would not show up on the quasi-classical level. The important detail is that the action of the Lorentz group is mixing the electric and magnetic components. Namely both components constitute a tensor represented by skewsymmetric form on four-dimensional space with a standard representation of $SO(3, 1)$. The action of both commuting Lie algebras $so(3, R)$ of pseudosymmetry on the complexification of the above representations does the same. It exactly corresponds to the coordinate change (n, l) to $(n + l, n - l)$ on the set of quantum states so that electric n and magnetic l components are mixing via the symmetry transform. The group $so(3)$ acts on the space of the tensors of the electromagnetic field: in $\Lambda^2 R^4$ and $so(3)_\Delta$ acts splits $\Lambda^2 R^4 = \Lambda^2 R^3$ (magnetic) $+ R^3$ electric charge. The group $so(3)_L$ acts on complexified space $\Lambda^2 C^4 = \Lambda^2 S^+ \times S^-$. The algebra $so(3)_L$ acts as $su(2)$ on $S^{\{+\}}$ and identically on $S^{\{-\}}$. Thus $so(3)_L$ acts as $S^{\{+\}} \times S^{\{+\}} + C^2 = C_S^3 + C_L^3$. The representation $S \times S^+ = C_S^3 + C$ of $so(3)_L$ coincides with C^4 -complexification of the space-time R^4 and rotates “complexified” tensors of magnetic and electric fields. Thus the group $Spin(3) \times Spin(3)Z_2 = SO(4)$ is not the real symmetry group of the problem but a shadow of Lorentz group $SO(3, 1)$ with a given spatial subgroup $SO(3)$.

3. *I*-Particle

How could such a symmetry appear in reality? Currently the standard answer to such question is symmetry break. Namely the existence of symmetry [15] within an ensemble of different particles is usually explained by considering them as different states of an “ideal object” with the above symmetry so that the particles are degenerate states of the above object where the symmetry is broken due to some process of natural degeneration.

If we try to apply similar explanation we are brought to the idea to view the atom as one of the possible states of an ideal particle which we will be denoting as *I*-particle. This particle has Lorentz symmetry algebra within a bigger algebra of local symmetries and supersymmetries corresponding to other potential fields. The *I*-particle exists in this ideal state only 0-time which in practice means a very short time depending on the total energy of the *I*-particle according to the “uncertainty principle”. While degenerating the *I*-particle creates an avalanche of intermediate particles which retain only pieces of the initial symmetry but previous symmetry is manifested in the set of possible degenerate states.

Note that due to the short time existence of *I*-particle it is more relevant to speak about the Lie algebra of symmetry and not of the group in this case. The process of the symmetry breaking can be put into phases. Namely one of the first symmetries to be broken is the symmetry between all possible space-type three-dimensional subspaces.

Thus the first break occurs on the level of space definition. It corresponds to the appearance of the selected space subgroup — subalgebra $so(3)_\Delta \subset sl(2, C)$ which is contained within a product of two commuting algebras: $so(3) \times so(3)$ in the complexification $sl(2, C) \times sl(2, C)$ of the Lorentz algebra $so(3, 1)$.

Indeed $so(3)_\Delta \subset sl(2, C)$ is contained in the unique product $so(3)_L \times so(3)_R \subset sl(2, C) \times sl(2, C)$. Note that $so(3)_L \times so(3)_R$ appear symmetrically which explains why the atomic structures are distributed via representations belonging to the tensor product of equivalent representations of $so(3)_L$ and $so(3)_R$.

However this process proceeds with a different speed in different regions and atomic nuclei [16] can be considered as a domain where this process stalls at some moment. To visualize the picture we can view the *I*-particle as a cluster of vortices (or oscillators) rotating with different speeds comparable with the speed of light which makes impossible to fix a spatial direction. Thus it opens up the possibility for the existence of the Lorentz symmetry if we assume an equisitrubition of the angles of the rotating vortices. The absence of a preferred space and time directions within the *I*-particle brings naturally Lorentz group in the picture. It is certainly a very simple model which only hints the possibility of such an effect.

The process of degeneration of an I -particle is inhomogeneous and involves the creation of radiation and some elementary particles. At the same time it goes much slower in some domains and stalls in the intermediate state in some minuscule domains insulated by surrounding clouds of electrons. This is the process which leads to the creation of big atoms.

Thus the nuclei (at least sufficiently big ones) can be considered as cooled down states of the I -particle and the electronic envelope consists of the cloud of emanated electrons which compensate the electric charge of the nuclei. There is a supporting evidence for this picture in recent experimental data which shows the existence of sublight velocities in some processes within the nuclei [16].

This general picture provides with a potential explanation of the structure and symmetries within the set of existing atoms. It also indicates the role of the electronic envelope of the atom. Being a relatively stable degenerate state of an I -particle the atom is electrically neutral and its stability is sustained by the interaction between the nucleus and the electronic cloud. Some recent experimental data which show that sublight velocities of processes within the nucleus points in this direction [16].

This general picture provides with a potential explanation of the structure and symmetries within the set of existing atoms. It also indicates the role of the electronic cloud in the atom.

Being a relatively stable degenerate state of an I -particle the atom is electrically neutral and its stability is sustained by the interaction between the nucleus and the electronic cloud.

References

1. *Magarshak Y.* Four-Dimensional Pyramidal Structure of the Periodic Properties of Atoms and Chemical Elements // *Sci. Isr. Technol. Adv.* — Vol. 7, No 1, 2. — 2006. — Pp. 134–150.
2. *Magarshak Y., Malinsky J.* A Three-Dimensional Periodic Table // *Nature.* — Vol. 360. — 1992. — Pp. 114–115.
3. *Schrödinger E.* // *Annalen der Physik.* Leipzig. — Vol. 79. — 1926. — P. 361.
4. *Schrödinger E.* // *Annalen der Physik.* Leipzig. — Vol. 79. — 1926. — P. 489.
5. *Schrödinger E.* // *Annalen der Physik.* Leipzig. — Vol. 79. — 1926. — P. 734.
6. *Schrödinger E.* // *Annalen der Physik.* Leipzig. — Vol. 80. — 1926. — P. 437.
7. *Schrödinger E.* // *Annalen der Physik.* Leipzig. — Vol. 81. — 1926. — P. 109.
8. *Mendelejeff D.* On the Relationship of the Properties of the Elements to their Atomic Weights // *Zeitschrift für Chemie.* — Vol. 12. — 1869. — Pp. 405–406.
9. *Cowan R. D.* The Theory of Atomic Structure and Spectra. — Science, 1981.
10. *Messiah A.* Quantum Mechanics. — Courier Dover Publications, 2000.
11. *Pauling L.* The Nature of the Chemical Bond. — Cornell Univ. Press, 1939.
12. *Pauling L.* Chemistry. — San Francisco: W.H. Freeman Publishing House, 1978.
13. *Levine I.* Quantum Chemistry. — 5th edition edition. — Prentice Hall, 1999.
14. *Fulton W. J.* Harris Representation Theory. — Springer, 1991.
15. *Murray Gell-Mann.* Symmetry Properties of Fields // *Proceedings of Solvay Congress.* — 1961.
16. *Niyazov R. A., Weinstein L. B. et al.* Two-Nucleon Momentum Distributions Measured in ${}^3\text{He}(e, e'pp)n$ // *Phys. Rev. Lett.* — Vol. 92. — 2004. — P. 052303.

УДК 531

Группа Лоренца и ее роль в нерелятивистской модели атома

Ф. Богомолов^{*}, Ю. Б. Магаршак[†]

^{} Институт Куранта*

[†] МатТех Инк., Нью-Йорк, США

В работе мы обсуждаем новые симметрии атомных структур, недавно открытых вторым соавтором. Мы также обсуждаем математические и физические аспекты возможного объяснения этих симметрий.

УДК 519.633.2, 517.958: 530.145.6

Ньютоновский потенциал в Общей теории относительности в конечном пространстве

И. В. Амирханов, Б. М. Барбашов, А. А. Гусев,
В. Н. Первушин, С. А. Шувалов, С. И. Веницкий,
А. Ф. Захаров, В. А. Зинчук

*Объединенный институт ядерных исследований
Россия, 141980, Дубна, Московской обл., ул. Жолио-Кюри, 6*

Ньютоновский потенциал вычисляется в гамильтоновом подходе к Общей теории относительности (ОТО) в конечном пространстве, где координатное время калибровочно-инвариантно и не может рассматриваться как измеряемая величина теории. Наблюдаемый (калибровочно-инвариантный) параметр эволюции отождествляется с однородным космологическим масштабным фактором $a(x^0)$, определенным с помощью усреднения логарифма детерминанта пространственной метрики по масштабнo инвариантному пространству Лихнеровича, а соответствующая калибровочно-инвариантная энергия в ОТО определяется как решение уравнения энергетической связи относительно канонического импульса масштабного фактора. В этом случае дается космологическое обобщение ньютоновского потенциала, заданного в неоднородном классе функций.

Введение

Потенциалами в релятивистской теории поля называют полевые переменные, которые удовлетворяют уравнениям типа Лапласа $\Delta A_0 = J_0$. В отличие от динамических переменных, которые удовлетворяют уравнениям типа Даламбера $\square A_0 = J_0$ и которые будем далее называть степенями свободы, так как они имеют свободные начальные данные.

Определение потенциала в релятивистской теории поля связано с так называемой сопутствующей системой отсчета, выделяемой единичным временноподобным вектором $l_\mu^0 = (1, 0, 0, 0)$. Этот вектор помогает разделить все координаты и компоненты полей на пространственные и временноподобные части, в соответствии с теорией неприводимых унитарных представлений группы Пуанкаре.

Начальные данные в теории калибровочных полей, включая общую теорию относительности (ОТО) [1, 2], должны быть определены как инварианты относительно группы калибровочных и общековариантных преобразований, которые радикально отличаются от преобразований систем отсчета. Это различие было показано двумя теоремами Нетер [3]. Параметры преобразований систем отсчета непосредственно измеряемы как начальные данные, в то время как параметры калибровочных преобразований не наблюдаемы; калибровочная инвариантность ведет лишь к связям начальных данных, выявленных симметриями теории относительно преобразований систем отсчета.

Для того, чтобы отделить группу калибровочных преобразований от преобразований систем отсчета в общей теории относительности (ОТО), Фок ввел описание ОТО в терминах лоренц-ковариантного и калибровочно-инвариантного ортогонального репера [4].

Главной трудностью в ОТО при вычислении потенциала является инвариантность этой теории относительно параметризации координатного времени $x^0 \rightarrow \tilde{x}^0 = \tilde{x}^0(x^0)$.

В случае, когда мы имеем дело с конечным пространством-временем, которое активно рассматривается в современной космологии, роль репараметризационно-инвариантного параметра эволюции играет космологический масштабный фактор [5–8]. Этот фактор отделяется масштабным преобразованием от метрики. Отделение космологического масштабного фактора от метрики в ОТО известно как космологическая теория возмущений, предложенная Лившицем [9] и применяемая для анализа современных наблюдательных данных в астрофизике и космологии [10]. Обычно предполагают, что этот фактор является дополнительной переменной, заданной в классе однородных функций. Это предположение ведет к удвоению числа однородных переменных, что не позволяет использовать гамильтонов подход к ОТО [11, 12].

В данной статье мы вычисляем ньютоновский потенциал в последовательном гамильтоновом подходе к ОТО [11, 12], где этот потенциал задан в классе неоднородных локальных функций.

1. Инвариантный параметр эволюции в ОТО

1.1. Действие, интервал и симметрии ОТО

ОТО [1, 2] дается двумя фундаментальными величинами: «динамическим» действием

$$S = \int d^4x \sqrt{-g} \left[-\frac{\varphi_0^2}{6} R(g) + \mathcal{L}_{(M)} \right], \quad (1)$$

где $\varphi_0^2 = \frac{3}{8\pi} M_{\text{Planck}}^2$ — постоянная Ньютона, а $\mathcal{L}_{(M)}$ — лагранжиан поля материи, и «геометрическим интервалом»

$$g_{\mu\nu} dx^\mu dx^\nu \equiv \omega_{(\alpha)} \omega_{(\alpha)} = \omega_{(0)} \omega_{(0)} - \omega_{(1)} \omega_{(1)} - \omega_{(2)} \omega_{(2)} - \omega_{(3)} \omega_{(3)}, \quad (2)$$

где $\omega_{(\alpha)}$ — линейные дифференциалы, определяемые как компоненты ортогонального репера [4]. ОТО в терминах компонент фоковского репера содержит два принципа относительности: «геометрический» — общее преобразование координат

$$x^\mu \rightarrow \tilde{x}^\mu = \tilde{x}^\mu(x^0, x^1, x^2, x^3), \quad \omega_{(\alpha)}(x^\mu) \rightarrow \omega_{(\alpha)}(\tilde{x}^\mu) = \omega_{(\alpha)}(x^\mu) \quad (3)$$

и «динамический» принцип, сформулированный как преобразования Лоренца систем отсчета заданных ортогональным репером

$$\omega_{(\alpha)} \rightarrow \bar{\omega}_{(\alpha)} = L_{(\alpha)(\beta)} \omega_{(\beta)}. \quad (4)$$

1.2. Система отсчета гамильтонова похода к ОТО

Гамильтонов подход к ОТО сформулирован в работе [5, 6] в определенной системе координат в терминах переменных

$$\omega_{(0)} = \psi^6 N_d dx^0 \equiv \psi^2 \omega_{(0)}^{(L)}, \quad \omega_{(b)} = \psi^2 \mathbf{e}_{(b)i} (dx^i + N^i dx^0) \equiv \psi^2 \omega_{(b)}^{(L)}; \quad (5)$$

где $\omega_{(\mu)}^{(L)}$ [13] — преобразование Лихнеровича к скалярно-инвариантным величинам, а триады $\mathbf{e}_{(a)i}$ формируют пространственную метрику с $\det |\mathbf{e}| = 1$, N_d — функция смещения Дирака, N^i — вектора сдвига начала координат, а ψ — детерминант пространственной метрики.

Зельманов установил [7, 8], что система координат, определенная в виде (5), является инвариантной относительно преобразований

$$x^0 \rightarrow \tilde{x}^0 = \tilde{x}^0(x^0), \quad x_i \rightarrow \tilde{x}_i = \tilde{x}_i(x^0, x_1, x_2, x_3), \quad (6)$$

$$\tilde{N}_d = N_d \frac{dx^0}{d\tilde{x}^0}, \quad \tilde{N}^k = N^i \frac{\partial \tilde{x}^k}{\partial x_i} \frac{dx^0}{d\tilde{x}^0} - \frac{\partial \tilde{x}^k}{\partial x_i} \frac{\partial x^i}{\partial \tilde{x}^0}. \quad (7)$$

Группа преобразований сохраняет семейство гиперповерхностей $x^0 = \text{const}$, которое называют «кинеметрической» подгруппой группы общих преобразований координат. «Кинеметрическая» подгруппа содержит репараметризацию координатного параметра эволюции (x^0). Это означает, что координатный параметр эволюции (x^0) не измеряем, роль наблюдаемого параметра эволюции в этом случае играет одна из переменных теории. В космологических приложениях ОТО в качестве такой репараметризационно-инвариантной времени-подобной переменной выбирают однородный масштабный фактор [14, 15].

1.3. Однородный масштабный фактор, как нулевая Фурье-гармоника

Космология Фридмана и космологическая теория возмущения [9, 10], применяемые для анализа современных наблюдательных данных в астрофизике и космологии, отождествляют инвариантный «параметр эволюции» с космологическим масштабным фактором $a(x_0)$. Последний вводится масштабным преобразованием метрики $g_{\mu\nu} = a^2(x^0)\tilde{g}_{\mu\nu}$ и любого поля $F^{(n)}$ с конформным весом (n): $F^{(n)} = a^n(x_0)\tilde{F}^{(n)}$. В частности кривизна

$$\sqrt{-g}R(g) = a^2\sqrt{-\tilde{g}}R(\tilde{g}) - 6a\partial_0\left[\partial_0 a\sqrt{-\tilde{g}}\tilde{g}^{00}\right] \quad (8)$$

может быть выражена в терминах новой функции смещения $\tilde{N}_d$ и пространственного детерминанта $\tilde{\psi}$ в уравнении (5)

$$\tilde{N}_d = [\sqrt{-\tilde{g}}\tilde{g}^{00}]^{-1} = a^2 N_d, \quad \tilde{\psi} = (\sqrt{a})^{-1}\psi. \quad (9)$$

Чтобы сохранить число переменных, мы определяем $\log \sqrt{a}$ следующим образом

$$\log \sqrt{a} = \langle \log \psi \rangle \equiv \frac{1}{V_0} \int d^3x \log \psi, \quad (10)$$

где $V_0 = \int d^3x < \infty$ — конечный объем. В этом случае новый детерминант переменной $\tilde{\psi}$ удовлетворяет тождеству

$$\int d^3x \log \tilde{\psi} = \int d^3x [\log \psi - \langle \log \psi \rangle] \equiv 0. \quad (11)$$

Можно назвать эти функции «отклонениями» от среднего значения. Можно заметить, что операции «усреднения» $\Pi_{av} \cdot \mu = \langle \mu \rangle$ и «отклонения» $\Pi_{de} \cdot \mu = \bar{\mu} = \mu - \langle \mu \rangle$ имеют свойства операторов проектирования: $\Pi_{de} \cdot \mu + \Pi_{av} \cdot \mu = I\mu$, $\Pi_{de}^2 = \Pi_{de}$ и $\Pi_{av}^2 = \Pi_{av}$.

После масштабного преобразования (8), (9) действие (1) принимает вид

$$S[\varphi_0] = \tilde{S}[\varphi] - \int dx^0 (\partial_0 \varphi)^2 \int \frac{d^3x}{\tilde{N}_d}; \quad (12)$$

здесь $\tilde{S}[\varphi]$ — действие (1) в терминах метрики $\tilde{g}$. Энергетическая связь $\delta S[\varphi_0]/\delta \tilde{N}_d = 0$ принимает вид [14]

$$-\frac{\delta \tilde{S}[\varphi]}{\delta \tilde{N}_d} \equiv \tilde{T}_0^0 = \frac{(\partial_0 \varphi)^2}{\tilde{N}_d^2}, \quad (13)$$

где $\tilde{T}_0^0$ — локальная плотность энергии. Это уравнение имеет точное решение

$$\left[\frac{d\varphi}{d\zeta} \right]^2 \equiv \varphi'^2 = \left\langle \sqrt{\tilde{T}_0^0} \right\rangle^2, \quad \langle (\tilde{N}_d)^{-1} \rangle \tilde{N}_d = \frac{\left\langle \sqrt{\tilde{T}_0^0} \right\rangle}{\sqrt{\tilde{T}_0^0}}, \quad (14)$$

где

$$\zeta(\varphi_0|\varphi) \equiv \int dx^0 \langle (\tilde{N}_d)^{-1} \rangle^{-1} = \pm \int_{\varphi}^{\varphi_0} \frac{d\tilde{\varphi}}{\langle \sqrt{\tilde{T}_0^0(\tilde{\varphi})} \rangle} \quad (15)$$

— инвариантное время. Данная эволюция в ОТО является аналогией закона Хаббла в космологии. Эти уравнения появляются, если $\tilde{T}_0^0$ не зависит от однородной скорости φ' .

2. Модель массивной электродинамики

2.1. Действие

В качестве модели материи рассмотрим массивную электродинамику в ОТО

$$S = S_{\text{GR}} + S_{\text{m}} = \int d^4x \sqrt{-g} [\mathcal{L}_{\text{GR}} + \mathcal{L}_{\text{m}}], \quad (16)$$

где $\mathcal{L}_{\text{GR}}$ — лагранжиан ОТО

$$\mathcal{L}_{\text{GR}} = -\frac{\varphi_0^2}{6} R(g) \quad (17)$$

и $\mathcal{L}_{\text{m}}$ — лагранжиан массивных векторных и спинорных полей

$$\mathcal{L}_{\text{m}} = -\frac{1}{4} F_{\mu\nu} F_{\alpha\beta} g^{\mu\alpha} g^{\nu\beta} - M_0^2 A_\mu A_\nu g^{\mu\nu} - \bar{\Psi} i \gamma^\sigma (D_\sigma - ie A_\sigma) \Psi - m_0 \bar{\Psi} \hat{\Psi}. \quad (18)$$

Здесь

$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ — тензор напряженности,
 $D_\delta = \partial_\delta - i\frac{1}{2}[\gamma_{(\alpha)}\gamma_{(\beta)}]\sigma_{\delta(\alpha)(\beta)}$ — ковариантная производная Фока [4],
 $\gamma_{(\beta)} = \gamma^\mu e_{(\beta)\mu}$ — γ -матрицы Дирака, свернутые с тетрадами $e_{(\beta)\nu}$, и
 $\sigma_{\sigma(\alpha)(\beta)} = e_{(\beta)}^\nu (\nabla_\mu e_{(\alpha)\nu})$ — коэффициенты спиновой связности [4].

2.2. Лагранжиан ОТО

Для вычисления канонических импульсов можно записать гильбертово действие (1) в терминах новых переменных Дирака (9)

$$S_{\text{GR}}[g = a^2 \tilde{g}] = \int d^4x (\mathbf{K}[\varphi|\tilde{g}] - \mathbf{P}[\varphi|\tilde{g}]) - \int dx^0 \frac{(\partial_0 \varphi)^2}{\tilde{N}_d} = \int dx^0 L_{\text{GR}}, \quad (19)$$

где

$$\mathbf{K}[\varphi|e] = \tilde{N}_d \varphi^2 \left[-4\tilde{v}^2 + \frac{v_{(ab)}^2}{6} \right], \quad (20)$$

$$\mathbf{P}[\varphi|e] = \frac{\tilde{N}_d \varphi^2 \tilde{\psi}^7}{6} \left[{}^{(3)}R(\mathbf{e}) \tilde{\psi} + 8\Delta \tilde{\psi} \right] \quad (21)$$

являются соответственно кинетическими и потенциальными условиями

$$\tilde{v} = \frac{1}{\tilde{N}_d} \left[(\partial_0 - N^l \partial_l) \log \tilde{\psi} - \frac{1}{6} \partial_l N^l \right], \quad (22)$$

$$v_{(ab)} = \frac{1}{2} (\mathbf{e}_{(a)i} v_{(b)}^i + \mathbf{e}_{(b)i} v_{(a)}^i), \quad (23)$$

$$v_{(a)i} = \frac{1}{\tilde{N}_d} \left[(\partial_0 - N^l \partial_l) \mathbf{e}_{(a)i} + \frac{1}{3} \mathbf{e}_{(a)i} \partial_l N^l - \mathbf{e}_{(a)l} \partial_i N^l \right] \quad (24)$$

— скорости метрических компонент, $\Delta\psi = \partial_i (\mathbf{e}_{(a)}^i \mathbf{e}_{(a)}^j \partial_j \psi)$ — ковариантный оператор Лапласа, ${}^{(3)}R(\mathbf{e})$ — трехмерная кривизна, в терминах триад $\mathbf{e}_{(a)i}$

$${}^{(3)}R(\mathbf{e}) = -2\partial_i [\mathbf{e}_{(b)}^i \sigma_{(c)|(b)(c)}] - \sigma_{(c)|(b)(c)} \sigma_{(a)|(b)(a)} + \sigma_{(c)|(d)(f)} \sigma_{(f)|(d)(c)}. \quad (25)$$

Здесь

$$\sigma_{(a)|(b)(c)} = \mathbf{e}_{(c)}^j \nabla_i \mathbf{e}_{(a)k} \mathbf{e}_{(b)}^k = \frac{1}{2} \mathbf{e}_{(a)j} [\partial_{(b)} \mathbf{e}_{(c)}^j - \partial_{(c)} \mathbf{e}_{(b)}^j] \quad (26)$$

— коэффициенты спиновой-связности, $\nabla_i \mathbf{e}_{(a)j} = \partial_i \mathbf{e}_{(a)j} - \Gamma_{ij}^k \mathbf{e}_{(a)k}$ — ковариантные производные, а $\Gamma_{ij}^k = \frac{1}{2} \mathbf{e}_{(b)}^k (\partial_i \mathbf{e}_{(b)j} + \partial_j \mathbf{e}_{(b)i})$.

2.3. Лагранжиан массивных полей

Лагранжиан массивных полей (18) можно переписать в терминах переменных Лихнеровича [13]

$$A^L_\mu = A_\mu, \quad \Psi^L = a^{3/2} \psi^3 \Psi, \quad (27)$$

что приводит к полям с массами, зависящими от масштабного фактора $a\psi^2$

$$m_{(L)} = m_0 a \psi^2 = m \psi^2, \quad M_{(L)} = M_0 a \psi^2 = M \psi^2, \quad (28)$$

находящимся в пространстве, заданном компонентами репера (5)

$$\omega_{(0)}^{(L)} = \tilde{\psi}^4 \tilde{N}_d dx^0, \quad (29)$$

$$\omega_{(a)}^{(L)} = \mathbf{e}_{(a)i} (dx^i + N^i dx^0) \quad (30)$$

с единичным детерминантом метрики $|\mathbf{e}| = 1$.

В результате лагранжиан массивных полей (18) принимает вид

$$\begin{aligned} \sqrt{-g} \mathcal{L}_m(A, \bar{\Psi}, \Psi) = & \frac{1}{i} \bar{\Psi}^L \gamma^0 \left(\partial_0 - N^k \partial_k + \frac{1}{2} \partial_l N^l - ie A_0 \right) \Psi^L - \\ & - \tilde{N}_d \mathcal{H}_\Psi + \tilde{N}_d \left[-J_{5(c)} v_{[ab]} \varepsilon_{(c)(a)(b)} + \right. \\ & + \frac{\tilde{\psi}^4}{2} \left(v_{i(A)} v_{(A)}^i - \frac{1}{2} F_{ij} F^{ij} \right) - \tilde{\psi}^8 M^2 A_{(b)}^2 - \frac{\pi_0^2}{M^2} \left. \right] - \\ & - \tilde{N}_d \tilde{\psi}^6 m \bar{\Psi}^L \Psi^L - \pi_0 [N^i A_i - A_0], \quad (31) \end{aligned}$$

где для линеаризации массового члена использовано преобразование Лежандра $A_0^2/(2\tilde{N}_d) = \pi_0 A_0 - \tilde{N}_d \pi_0^2/2$ с вспомогательным полем π_0 ;

$$\mathcal{H}_\Psi = -\tilde{\psi}^4 [i \bar{\Psi}^L \gamma_{(b)} D_{(b)} \Psi^L + J_5^0 \sigma - \partial_k J^k] \quad (32)$$

— гамильтонова плотность фермионов,

$$v_{[ab]} = \frac{1}{2} (\mathbf{e}_{(a)i} v_{(b)}^i - \mathbf{e}_{(b)i} v_{(a)}^i), \quad (33)$$

$$D_{(b)} \Psi^L = \left[\partial_{(b)} - \frac{1}{2} \partial_k \mathbf{e}_{(b)}^k - ie A_{(b)} \right] \Psi^L, \quad (34)$$

$$v_{i(A)} = \frac{1}{\psi^4 \tilde{N}_d} [\partial_0 A_i - \partial_i A_0 + F_{ij} N^j] \quad (35)$$

— скорости полевых переменных, а величины

$$J_{5(c)} = \frac{i}{2} (\bar{\Psi}^L \gamma_5 \gamma_{(c)} \Psi^L), \quad J_5^0 = \frac{i}{2} (\bar{\Psi}^L \gamma_5 \gamma^0 \Psi^L), \quad J_k = \frac{i}{2} \bar{\Psi}^L \gamma_k \Psi^L \quad (36)$$

являются потоками, $\sigma = \sigma_{(a)(b)|(c)} \varepsilon_{(a)(b)(c)}$, где $\varepsilon_{(a)(b)(c)}$ обозначает тензор Леви–Чивита.

2.4. Гамильтонов подход

Перейдем от скоростей к канонически сопряженным импульсам:

$$P_\varphi = -2V_0 \frac{\partial_0 \varphi}{N_0} = -2V_0 \frac{d\varphi}{d\zeta} \equiv -2V_0 \varphi', \quad (37)$$

$$\tilde{p}_\psi = \frac{\partial \mathbf{K}[\varphi|e]}{\partial(\partial_0 \ln \psi)} = -8\varphi^2 \tilde{v}, \quad (38)$$

$$p_{(b)}^i = \frac{\partial [\mathbf{K}[\varphi|e] + \sqrt{-g} \mathcal{L}_m]}{\partial(\partial_0 \mathbf{e}_{(a)i})} = \mathbf{e}_{(a)}^i \left[\frac{\varphi^2}{3} v_{(ab)} - J_{5(c)} \varepsilon_{(c)(a)(b)} \right], \quad (39)$$

$$P_{(A)}^i = \frac{\partial [\sqrt{-g} \mathcal{L}_m]}{\partial(\partial_0 A_i)} = \tilde{\psi}^4 v_{(A)}^i, \quad (40)$$

$$P_{(\Psi)} = \frac{\partial [\sqrt{-g} \mathcal{L}_m]}{\partial(\partial_0 \Psi^L)} = \frac{1}{i} \bar{\Psi}^L \gamma^0. \quad (41)$$

Тогда действие (16) можно записать в гамильтоновой форме

$$S = \int dx^0 \left[-P_\varphi \partial_0 \varphi + N_0 \frac{P_\varphi^2}{4V_0} + \int d^3x \left(\sum_F P_F \partial_0 F + \mathcal{C} - \tilde{N}_d T_{0t}^0 \right) \right], \quad (42)$$

где P_F — набор полевых импульсов (38)–(41), а

$$\tilde{T}_0^0 = \tilde{\psi}^7 \hat{\Delta} \tilde{\psi} + \sum_{I=0,4,6,8} \tilde{\psi}^I \tau_I \quad (43)$$

— сумма гамильтоновых плотностей, включая плотность гравитации

$$\tilde{\psi}^7 \hat{\Delta} \tilde{\psi} \equiv \tilde{\psi}^7 \frac{4\varphi^2}{3} \partial_{(b)} \partial_{(b)} \tilde{\psi}, \quad (44)$$

$$\tau_{I=0} = \frac{6\tilde{p}_{(ab)} \tilde{p}_{(ab)}}{\varphi^2} - \frac{16}{\varphi^2} \tilde{p}_\psi^2 + \frac{\pi_0^2}{2a^2 M^2}, \quad (45)$$

$$\tau_{I=4} = \frac{P_{i(A)} P_{(A)}^i + F_{ij} F^{ij}}{2} - \left[i \bar{\Psi}^L \gamma_{(b)} D_{(b)} \Psi^L + J_5^0 \sigma - \partial_k J^k \right], \quad (46)$$

$$\tau_{I=6} = m \bar{\Psi}^L \Psi^L \quad (47)$$

$$\tau_{I=8} = \frac{\varphi^2}{6} R^{(3)}(\mathbf{e}) + \frac{M^2 A_{(b)}^2}{2}, \quad (48)$$

где $\tilde{p}_{(ab)} = \frac{1}{2} (\mathbf{e}_{(a)}^i \tilde{p}_{(b)i} + \mathbf{e}_{(b)}^i \tilde{p}_{(a)i})$, $\tilde{p}_{(b)i} = p_{(b)i} + \mathbf{e}_{(a)}^i \varepsilon_{(c)(a)(b)} J_{(c)}$,

$$\mathcal{C} = A_0 [\partial_i P_{(A)}^i + e J_0 - \pi_0] + N_{(b)} T_{(b)t}^0 + \lambda_0 \tilde{p}_\psi + \lambda_{(a)} \partial_k \mathbf{e}_{(a)}^k \quad (49)$$

($J_0 = \bar{\Psi}^L \gamma_0 \Psi^L$ — нулевая компонента тока; $A_0, N_d, N^i, \lambda_0, \lambda_{(a)}$ являются множителями Лагранжа) обозначает сумму связей, включающую условие Дирака минимальной трехмерной гиперповерхности [5, 6]

$$p_{\tilde{\psi}} = \tilde{v} = 0 \rightarrow (\partial_0 - N^l \partial_l) \log \tilde{\psi} = \frac{1}{6} \partial_l N^l, \quad (50)$$

означающей положительную определенность плотности гамильтониана (45). Если обозначить компоненты полного тензора энергии-импульса $T_{(a)}^0 = T_i^0 \mathbf{e}_{(a)}^i$, то

$$T_{(a)}^0 = -p_{\psi} \partial_{(a)} \psi + \frac{1}{6} \partial_{(a)} (p_{\psi} \psi) + 2p_{(b)(c)} \gamma_{(b)|(a)(c)} - \partial_{(b)} p_{(b)(a)} - \\ - \frac{1}{i} \bar{\Psi}^I \gamma^0 \partial_{(a)} \Psi^I - \frac{1}{2i} \partial_{(a)} (\bar{\Psi}^I \gamma^0 \Psi^I) - P_{(A)}^j F_{jk} \mathbf{e}_{(a)}^k - \pi_0 A_{(a)}. \quad (51)$$

2.5. Гамильтонова редукция

Энергетическая связь $\delta S[\varphi|\tilde{F}]/\delta \tilde{N}_d = 0$ в виде уравнения $N_0^2 P_{\varphi}^2 / (\tilde{N}_d)^2 = \tilde{T}_0^0(\varphi|\tilde{F})$ имеет точное решение

$$P_{\varphi}^2 = E_{\varphi}^2, \quad \frac{N_0}{\tilde{N}_d} = \frac{\sqrt{\tilde{T}_0^0}}{\langle \sqrt{\tilde{T}_0^0} \rangle}, \quad (52)$$

где

$$E_{\varphi} = 2V_0 \langle \sqrt{\tilde{T}_0^0} \rangle, \quad (53)$$

и подстановка этого решения $P_{\varphi(\pm)} = \pm E_{\varphi}$ в действие

$$S = \int dx^0 \left[-P_{\varphi} \partial_0 \varphi + N_0 \frac{P_{\varphi}^2}{4V_0} + \int d^3x \left(\sum_F P_F \partial_0 F + \mathcal{C} - \tilde{N}_d \tilde{T}_0^0 \right) \right], \quad (54)$$

дает редуцированное действие

$$S_{\pm}[\varphi_I|\varphi_0]|_{\text{energy constraint}} = \int_{\varphi_I}^{\varphi_0} d\varphi \left\{ \left[\int d^3x \left(\sum_{\tilde{F}} \tilde{P}_F \partial_{\varphi} \tilde{F} + \bar{\mathcal{C}} \right) \pm E_{\varphi} \right] \right\} \quad (55)$$

в полевом пространстве событий $[\varphi|\tilde{F}]$. Здесь $\bar{\mathcal{C}} = \mathcal{C}/\partial_0 \varphi$, φ_I есть начальные данные рождения (или уничтожения) Вселенной.

Окончательно для десяти компонент дираковской тетрады [4] имеем следующие уравнения

$$\frac{\delta S}{\delta e_{(b)i}^{(T)}} = -T_{(b)}^{i(T)} = 0, \\ \frac{\delta S}{\delta N^{j(T)}} = -T_0^{j(T)} = 0, \\ \partial_j e_{(b)}^{(T)} = 0, \\ \partial_j N^{jT} = 0.$$

Для N_{int} и $\partial_i N^0$ имеем соответственно

$$N_{int} = \langle (\tilde{N}_d)^{-1} \rangle \tilde{N}_d = \frac{\langle \sqrt{\tilde{T}_0^0} \rangle}{\sqrt{\tilde{T}_0^0}}, \quad \left[\frac{d\varphi}{d\zeta} \right]^2 \equiv \varphi'^2 = \langle \sqrt{\tilde{T}_0^0} \rangle^2,$$

$$p_{\tilde{\psi}} = \tilde{v} = 0 \rightarrow (\partial_0 - N^l \partial_l) \log \tilde{\psi} = \frac{1}{6} \partial_l N^l,$$

а для $\tilde{\psi}$ — уравнение

$$\frac{\delta S}{\delta \ln \tilde{\psi}} = \frac{\delta S}{\delta \ln \psi} - \left\langle \frac{\delta S}{\delta \ln \psi} \right\rangle = 0,$$

которое будет рассмотрено в следующем разделе.

3. Ньютоновский потенциал в конечном пространстве

3.1. Эффективное действие для скалярных потенциалов в конечном пространстве

В диффеоинвариантной формулировке в ОТО скалярные потенциальные возмущения в определенной системе отсчета могут быть определены как $N_{\text{int}}^{-1} = 1 + \bar{\nu}$ и $\tilde{\psi} = e^{\bar{\mu}}$, где $\bar{\mu}$, $\bar{\nu}$ даются в классе функций, определенных оператором проектирования (11) $\bar{F} = F - \langle F \rangle$, причем $\langle \bar{F} \rangle \equiv 0$.

Явную зависимость метрики и тензора энергии $\tilde{T}_0^0$ от $\tilde{\psi}$ можно получить в терминах масштабно-инвариантных переменных Лихнеровича [13]

$$\omega_{(0)}^{(L)} = \tilde{\psi}^4 N_{\text{int}} d\zeta, \quad \omega_{(b)}^{(L)} = \mathbf{e}_{(b)k} [dx^k + \mathcal{N}^k d\zeta], \quad (56)$$

$$\tilde{T}_0^0 = \tilde{\psi}^7 \hat{\Delta} \tilde{\psi} + \sum_I \tilde{\psi}^I a^{\frac{I}{2}-2} \tau_I, \quad \tau_I \equiv \langle \tau_I \rangle + \bar{\tau}_I, \quad (57)$$

где $\hat{\Delta} \tilde{\psi} \equiv \frac{4\varphi^2}{3} \partial_{(b)} \partial_{(b)} \tilde{\psi}$ — оператор Лапласа, а τ_I — плотности энергии, определенные уравнениями (44)–(48). Отрицательный вклад $-(16/\varphi^2) \bar{p}_{\psi}^2$ пространственного детерминанта импульса $\tau_{I=0}$ убирается из нулевой скорости пространственного элемента объема условием положительности локальной плотности энергии, введенном Дираком в [5, 6]

$$p_{\tilde{\psi}} = -8\varphi^2 \frac{\partial_{\zeta} \tilde{\psi}^6 - \partial_l [\tilde{\psi}^6 \mathcal{N}^l]}{\tilde{\psi}^6 N_{\text{int}}} = 0. \quad (58)$$

Для упрощения уравнений скалярных потенциалов $N_{\text{int}}, \tilde{\psi}$, можно ввести новые обозначения

$$N_s = \psi^7 N_{\text{int}}, \quad T(\tilde{\psi}) = \sum_I \tilde{\psi}^{(I-7)} a^{\frac{I}{2}-2} \tau_I, \quad \rho_{(0)} = \left\langle \sqrt{\tilde{T}_0^0} \right\rangle^2 = \varphi'^2. \quad (59)$$

В терминах этих обозначений действие (55) можно представить, используя инвариантное время ζ , в виде

$$S[\varphi_0] = \int d\zeta \int d^3x \left[\sum_F P_F \partial_{\zeta} F + C_{\zeta} - N_s (\hat{\Delta} \tilde{\psi} + T(\tilde{\psi})) - \frac{\tilde{\psi}^7 \rho_{(0)}}{N_s} \right], \quad (60)$$

где $C_{\zeta} = C(dx^0/d\zeta)$.

3.2. Потенциальные уравнения

Вариация этого действия относительно $N_s, \tilde{\psi}$ ведет к уравнениям

$$\hat{\Delta}\tilde{\psi} + T = \frac{\tilde{\psi}^7 \rho_{(0)}}{N_s^2}, \quad (61)$$

$$\tilde{\psi}\hat{\Delta}N_s + N_s\tilde{\psi}\partial_{\tilde{\psi}}T + 7\frac{\tilde{\psi}^7 \rho_{(0)}}{N_s} = \rho_{(1)}, \quad (62)$$

соответственно, где

$$\rho_{(0)} = \langle N_s[\hat{\Delta}\tilde{\psi} + T] \rangle, \quad (63)$$

$$\rho_{(1)} = \left\langle \tilde{\psi}\hat{\Delta}N_s + N_s\tilde{\psi}\partial_{\tilde{\psi}}T + 7\frac{\tilde{\psi}^7 \rho_{(0)}}{N_s} \right\rangle. \quad (64)$$

в пределе бесконечного объема $\rho_{(n)} = \langle \tau_I \rangle = 0$ уравнения (61) и (62) сводятся к уравнениям $\hat{\Delta}\tilde{\psi} = 0$, $\hat{\Delta}N_s = 0$ с решениями Шварцшильда $\tilde{\psi} = 1 + \frac{r_g}{4r}$; $N_s = 1 - \frac{r_g}{4r}$.

3.3. Космологические возмущения

Для $N_s = 1 - \nu_1$, $\tilde{\psi} = 1 + \mu_1$ и малых отклонений $\mu_1, \nu_1 \ll 1$ уравнения (61) и (62) принимают вид

$$[-\hat{\Delta} + 14\rho_{(0)} - \rho_{(1)}]\mu_1 + 2\rho_{(0)}\nu_1 = \bar{\tau}_{(0)}, \quad (65)$$

$$[7 \cdot 14\rho_{(0)} - 14\rho_{(1)} + \rho_{(2)}]\mu_1 + [-\hat{\Delta} + 14\rho_{(0)} - \rho_{(1)}]\nu_1 = 7\bar{\tau}_{(0)} - \bar{\tau}_{(1)}, \quad (66)$$

где $\rho_{(n)} = \langle \tau_{(n)} \rangle \equiv \sum_I I^n a^{\frac{I}{2}-2} \langle \tau_I \rangle$. Этот выбор переменных позволяет определить $\tilde{\psi} = 1 + \mu_1$ и $N_s = 1 - \nu_1$ в виде суммы

$$\tilde{\psi} = 1 + \frac{1}{2} \int d^3y [D_{(+)}(x, y)\bar{T}_{(+)}^{(\mu)}(y) + D_{(-)}(x, y)\bar{T}_{(-)}^{(\mu)}(y)], \quad (67)$$

$$N_s = 1 - \frac{1}{2} \int d^3y [D_{(+)}(x, y)\bar{T}_{(+)}^{(\nu)}(y) + D_{(-)}(x, y)\bar{T}_{(-)}^{(\nu)}(y)], \quad (68)$$

где β имеет вид

$$\beta = \sqrt{1 + [\langle \tau_{(2)} \rangle - 14\langle \tau_{(1)} \rangle] / (98\langle \tau_{(0)} \rangle)}, \quad (69)$$

$$\bar{T}_{(\pm)}^{(\mu)} = \bar{\tau}_{(0)} \mp 7\beta[7\bar{\tau}_{(0)} - \bar{\tau}_{(1)}], \quad \bar{T}_{(\pm)}^{(\nu)} = [7\bar{\tau}_{(0)} - \bar{\tau}_{(1)}] \pm (14\beta)^{-1}\bar{\tau}_{(0)} \quad (70)$$

— локальные токи, $D_{(\pm)}(x, y)$ — функции Грина, удовлетворяющие уравнениям

$$[\pm\hat{m}_{(\pm)}^2 - \hat{\Delta}]D_{(\pm)}(x, y) = \delta^3(x - y). \quad (71)$$

Здесь $\hat{m}_{(\pm)}^2 = 14(\beta \pm 1)\langle \tau_{(0)} \rangle \mp \langle \tau_{(1)} \rangle$.

В случае распределения массы в конечном объеме V_0 с нулевыми давлением и плотностью $\bar{\tau}_{(0)}(x) = \bar{\tau}_{(1)}(x)/6 \equiv M[\delta^3(x - y) - 1/V_0]$, решения (67), (68) принимают вид

$$\tilde{\psi} = 1 + \mu_1 = 1 + \frac{r_g}{4r} [\gamma_1 e^{-m_{(+)}(z)r} + (1 - \gamma_1) \cos m_{(-)}(z)r], \quad (72)$$

$$N_s = 1 - \nu_1 = 1 - \frac{r_g}{4r} [(1 - \gamma_2) e^{-m_{(+)}(z)r} + \gamma_2 \cos m_{(-)}(z)r], \quad (73)$$

где $\gamma_1 = \frac{1+7\beta}{2}$, $\gamma_2 = \frac{14\beta-1}{28\beta}$, $r_g = \frac{3M}{4\pi\varphi^2}$, $r = |x - y|$, $m_{(\pm)}^2 = \frac{3\widehat{m}_{(\pm)}^2}{4\varphi^2}$.

Минимальная поверхность (55) $\partial_i[\bar{\psi}^6 \mathcal{N}^i] - (\bar{\psi}^6)' = 0$ дает вектор сдвига начала координат в процессе эволюции

$$\mathcal{N}^i = \left(\frac{x^i}{r}\right) \left(\frac{\partial_\zeta V}{\partial_r V}\right), \quad V(\zeta, r) = \int^r d\tilde{r} \tilde{r}^2 \tilde{\psi}^6(\zeta, \tilde{r}). \quad (74)$$

Решения (72), (73) имеют пространственные осцилляции и ненулевой вектор сдвига начала координат (74). В бесконечном объеме $\langle \tau_{(n)} \rangle = 0$, $a = 1$ решения (72) и (73) совпадают с изотропными решениями Шварцшильда $\tilde{\psi} = 1 + \frac{r_g}{4r}$, $N_s = 1 - \frac{r_g}{4r}$, $N^k = 0$.

3.4. Пример

Рассмотрим случай распределения массы в конечном объеме $V_0 = 4\pi R_0^3/3$ с плотностью энергии (57), содержащей только вклад частице-подобных сингулярностей и массивной пыли $I = 6$, так что $\tau_{I=6} = \Omega_M \rho_{cr} + M\delta^3(x)$, где $\rho_{cr} = \varphi^2 H^2$. В этом случае плотность энергии (59) может быть записана в виде

$$T = \frac{4\varphi^2}{3} \tilde{\psi}^{-1} \tilde{T}, \quad \tilde{T} = \left[\frac{3}{4} H^2 \Omega_M + \pi r_g \delta^3(x) \right], \quad (75)$$

здесь H — параметр Хаббла, и $r_g = M3/(4\pi\varphi^2)$ — радиус гравитации.

Уравнения (61) и (62) сведем к двум интегро-дифференциальным уравнениям соответственно

$$N_s [\Delta \tilde{\psi} + \tilde{\psi}^{-1} \tilde{T}] = \frac{\tilde{\psi}^7}{N_s} \rho_{(0)}, \quad (76)$$

$$\tilde{\psi} \Delta N_s - N_s \tilde{\psi}^{-1} \tilde{T} = \rho_{(1)} - 7 \left[\frac{\tilde{\psi}^7}{N_s} - 1 \right] \rho_{(0)}, \quad (77)$$

где $\Delta f = \frac{d^2 f}{dr^2} + \frac{2}{r} \frac{df}{dr}$ в интервале $r_b \leq r \leq R_0$, а r_b — естественная разница «частице-подобных» объектов (предполагаем что $r_b \sim r_g \ll R_0 \sim H^{-1}$) и

$$\rho_{(0)} = \frac{3}{R_0^3 - r_b^3} \int_{r_b}^{R_0} dr r^2 N_s [\Delta \tilde{\psi} + \tilde{\psi}^{-1} \tilde{T}], \quad (78)$$

$$\rho_{(1)} = \frac{3}{R_0^3 - r_b^3} \int_{r_b}^{R_0} dr r^2 [\tilde{\psi} \Delta N_s - N_s \tilde{\psi}^{-1} \tilde{T}], \quad (79)$$

с нормировкой в интервале $r_b \leq r = \sqrt{x_1^2 + x_2^2 + x_3^2} \leq R_0$

$$\frac{3}{R_0^3 - r_b^3} \int_{r_b}^{R_0} dr r^2 \frac{\tilde{\psi}^7}{N_s} = 1, \quad (80)$$

$$\frac{3}{R_0^3 - r_b^3} \int_{r_b}^{R_0} dr r^2 \log \tilde{\psi} = 0, \quad (81)$$

и граничные условия $N_s(r = R_0) = 1$, $\tilde{\psi}(r = R_0) = 1$. Для $N_s = 1 - \nu_1$, $\tilde{\psi} = 1 + \mu_1$ и малых отклонений $\mu_1, \nu_1 \ll 1$ решение этих уравнений имеет вид (72) и (73),

$$\tilde{\psi} = 1 + \mu_1 = 1 + \frac{rg}{4r} \left[\gamma_1 e^{-m_{(+)}(z)r} + (1 - \gamma_1) \cos m_{(-)}(z)r \right], \quad (82)$$

$$N_s = 1 - \nu_1 = 1 - \frac{rg}{4r} \left[(1 - \gamma_2) e^{-m_{(+)}(z)r} + \gamma_2 \cos m_{(-)}(z)r \right], \quad (83)$$

где $\beta = \sqrt{1 + [36 - 14 \cdot 6]/(98)} = \sqrt{1 - 48/98} \simeq 1/1.4$, $m_{(-)} = H\sqrt{\frac{3}{2}\Omega_M}$, $m_{(+)} = 3m_{(-)}$, $\gamma_1 = 3$, $\gamma_2 = 9/20$.

Можно видеть, что это изменение координаты $r \rightarrow x = rH\sqrt{(3/4\Omega_M)}$, а $\tilde{T}$ в уравнении (75) содержит $\tilde{T} \rightarrow 1$.

Уравнения (61) и (62) сводим к двум интегро-дифференциальным уравнениям соответственно в интервале $x_b \leq x \leq 1$, $\rho_M \ll x_b \sim x_g \ll 1$, $R_0 \sim H\sqrt{(3/4\Omega_M)} = 1$

$$\Delta\tilde{\psi} + \tilde{\psi}^{-1} = \frac{\tilde{\psi}^7}{N_s^2} \rho_{(0)}, \quad (84)$$

$$\tilde{\psi}\Delta N_s - N_s\tilde{\psi}^{-1} = \rho_{(1)} - 7\left[\frac{\tilde{\psi}^7}{N_s} - 1\right]\rho_{(0)}, \quad (85)$$

где $\Delta f = \frac{d^2 f}{dx^2} + \frac{2}{x} \frac{df}{dx}$ и интегральные термины имеют вид

$$\rho_{(0)} = 3 \int_{x_b}^1 dx x^2 N_s \left[\Delta\tilde{\psi} + \tilde{\psi}^{-1} \right], \quad x_b \ll 1, \quad (86)$$

$$\rho_{(1)} = 3 \int_{x_b}^1 dx x^2 \left[\tilde{\psi}\Delta N_s - N_s\tilde{\psi}^{-1} \right], \quad x_b \ll 1, \quad (87)$$

с пределами интегрирования

$$3 \int_{x_b}^1 dx x^2 \frac{\tilde{\psi}^7}{N_s} = 1, \quad (88)$$

$$3 \int_{x_b}^1 dx x^2 \log \tilde{\psi} = 0. \quad (89)$$

В предельном случае $x \rightarrow 1 \gg x_g$ имеем

$$\psi_0(x) = 1 - \varepsilon_1 + \frac{x_g}{4x} \frac{1}{20} \left[11e^{-3\sqrt{6}x} + 9 \cos \sqrt{6}x \right], \quad (90)$$

$$N_{0s}(x) = 1 + \varepsilon_2 - \frac{x_g}{4x} \left[3e^{-3\sqrt{6}x} - 2 \cos \sqrt{6}x \right]. \quad (91)$$

Следовательно, в этом случае мы можем наложить граничные условия

$$N_s(x = 1) \simeq 1, \quad (92)$$

$$\tilde{\psi}(x = 1) \simeq 1. \quad (93)$$

Проблема состоит в нахождении значений функций $\tilde{\psi}(x)$, $N_s(x)$ в области $x \sim x_g$ и сравнении этих функций с (90) и (91).

Прежде всего можно видеть, что уравнение (76) — алгебраическое относительно N_s и имеет решение

$$N_s^{-1} = \tilde{\psi}^{-7} \frac{[\tilde{\psi}^7 \Delta \tilde{\psi} + \tilde{\psi}^6]^{1/2}}{\rho_{(0)}^{1/2}} \gg 1, \quad (94)$$

из которого видно, что N_s никогда в ноль не обращается. Следовательно черная дыра в космологии не существует.

Заключение

Мы вычислили потенциал Ньютона в ОТО в конечном пространстве-времени, отделяя преобразования систем отсчета от общих преобразований координат и различая координатное время x^0 (как объект репараметризации $x^0 \rightarrow \tilde{x}^0 = \tilde{x}^0(x^0)$) и репараметризационно-инвариантное конформное время η (как результат решения уравнений движения и энергетической связи). В космологической теории возмущения [9, 10] в качестве инвариантного параметра эволюции выбирают космологический масштабный фактор. Гамильтонов подход [11, 12] обнаружил двойной учет космологического масштабного фактора в стандартной теории возмущения Лившица [9]. Это означает, что стандартная теория возмущения не совпадает с теорией Эйнштейна [1, 2]. Чтобы избежать двойного учета и вернуться к ОТО, мы отделили масштабный фактор с помощью усреднения детерминанта пространственной метрики. В этом случае показано, что сохранение числа переменных в ОТО позволяет исключить кинетические возмущения и получить новый тип потенциальных возмущений, как космологическое обобщение решения Шварцшильда и потенциала Ньютона.

Литература

1. *Einstein A.* Die Gründlange der allgemeinen Relativitätstheorie // Ann. d. Phys. — Vol. 49. — 1916. — Pp. 769–826.
2. *Hilbert D.* Die Grundlangen der Physik // Nachrichten von der Kön. Ges. der Wissenschaften zu Göttingen, Math.-Phys. Kl. — Vol. 3. — 1915. — Pp. 395–407.
3. *Noether E.* Invariante Variationsprobleme // Göttinger Nachrichten, Math.-Phys. Kl. — Vol. 2. — 1918. — Pp. 235–251.
4. *Fock V. A.* Geometrisierung der Diracschen Theorie des Electrons // Zs. Phys. — Vol. 57. — 1929. — Pp. 261–277.
5. *Dirac P. A. M.* Generalized Hamiltonian Dynamics // Proc. Roy. Soc. — Vol. A 246. — London: 1958. — Pp. 326–332.
6. *Arnowitt R., Deser S., Misner C. W.* Canonical Variables for General Relativity // Phys. Rev. — Vol. 117. — 1960. — Pp. 1595–1602.
7. *Zelmanov A. L.* Kinemetric Invariants and Their Relation to Chronometric Ones in Einstein's Gravitation Theory // Dokl. AN USSR. — Vol. 209. — 1973. — Pp. 822–825.
8. *Vladimirov Y. S.* Frame of References in Theory of Gravitation. — Moscow: Energoizdat, 1982. — Vol. 209, Pp. 822–825. — In Russian.
9. *Lifshits E. M.* On the Gravitational Stability of the Expanding Universe // ZhETF. — Vol. 16. — 1946. — Pp. 587–602.
10. *Mukhanov V. F., Feldman H. A., Brandenberger R. H.* Theory of Cosmological Perturbations // Phys. Rep. — Vol. 215. — 1992. — Pp. 203–333.
11. *Barbashov B. M. et al.* Hamiltonian Cosmological Perturbation Theory // Phys. Lett. — Vol. B 633. — 2006. — in press.
12. *Zakharov A. F., Zinchuk V. A., Pervushin V. N.* Tetrad Formalism and Frames of References in General Relativity // Physics of Particles and Nuclei. — Vol. 37. — 2006. — to be published.

13. York J. W. Gravitational Degrees of Freedom and the Initial-Value Problem // *Phys. Rev. Lett.* — Vol. 26. — 1971. — Pp. 1656–1658.
14. Pawlowski M., Pervushin V. N. Reparametrization-Invariant Path Integral in GR and “Big Bang” of Quantum Universe // *Int. J. Mod. Phys.* — Vol. A 16. — 2001. — Pp. 1715–1742.
15. Pervushin V., Smirichinski V. Bogoliubov Quasiparticles in Constrained Systems // *J. Phys. A: Math. Gen.* — Vol. 32. — 1999. — Pp. 6191–6201.

UDC 519.633.2, 517.958: 530.145.6

Newton Potential in General Relativity in a Finite Volume

I. V. Amirkhanov, B. M. Barbashov, A. A. Gusev, V. N. Pervushin,
S. A. Shuvalov, S. I. Vinitzky, A. F. Zakharov, V. A. Zinchuk

*Joint Institute for Nuclear Research
6, Joliot-Curie str., Dubna, Moscow Region, 141980, Russia*

The Newton potential is calculated in the Hamiltonian approach to general relativity (GR) in finite volume, where coordinate “time” is gauge-invariant and therefore can’t be considered as a measurable variable of the theory. The evolution (gauge-invariant) parameter under study is identified with a homogenous cosmological scale factor $a(x^0)$, determined by means of averaging logarithm of spatial metric determinant over the scale-invariant Licnerowicz space, whereas respective gauge-invariant energy in GR is determined as a solution of the energy constraint equation in relation to the canonical momentum of the scale factor. In this case, cosmological generalization of the Newton potential, given in a non-homogenous class of functions, is specified.

УДК 539.12

О возможном геометрическом происхождении симметрий Ю. Магаршака

Ю. П. Рыбаков

*Кафедра теоретической физики
Российский университет дружбы народов
Россия, 117198, Москва, ул. Миклухо-Маклая, 6*

Предлагается возможная интерпретация симметрий многомерной таблицы элементов Ю. Магаршака.

В очень интересных работах [1, 2] были предложены принципы построения многомерных аналогов периодической таблицы химических элементов Д. И. Менделеева. В дальнейшем было выяснено [3], что в предложенной схеме должна существовать асимптотическая симметрия, порожденная двумя коммутирующими генераторами. В данной заметке предлагается наглядная геометрическая интерпретация отмеченной симметрии.

Рассмотрим сначала бесконечную (открытую) струну, или вихрь, ориентированную вдоль оси Z . Тогда одна из симметрий может порождаться сдвигами вдоль струны, а другая — вращениями вокруг неё. Соответствующие векторы Киллинга будут иметь вид:

$$\partial_z, \quad \partial_\varphi,$$

где φ — азимутальный угол.

Если взять отрезок такой струны конечной длины, то его можно замкнуть в тор. Полученная замкнутая тороидальная структура по-прежнему будет сохранять указанные выше симметрии, порождённые соответственно поворотами вдоль и поперёк тора. В подходящих тороидальных координатах это должно отвечать разделению переменных в уравнении Шрёдингера. Последнее возможно, если эффективный самосогласованный потенциал атома оказывается асимптотически зависящим только от радиальной тороидальной координаты.

Литература

1. *Magarshak Y., Malinsky J.* A Three-Dimensional Periodic Table // *Nature*. — Vol. 360. — 1992. — Pp. 114–115.
2. *Magarshak Y.* Four-Dimensional Pyramidal Structure of the Periodic Properties of Atoms and Chemical Elements // *Scientific Israel — Technological Advances*. — Vol. 7, No 1, 2. — 2006. — Pp. 134–150.
3. *Bogomolov F., Magarshak Y.* On Commuting Operators Related to Asymptotic symmetries in the Atomic Theory // *Scientific Israel — Technological Advances*. — Vol. 7, No 3. — 2006. — Pp. 134–150.

UDC 539.12

On Possible Geometric Origin of Symmetries by Y. Magarshak

Yu. P. Rybakov

*Department of Theoretical Physics
Peoples' Friendship University of Russia
6, Miklukho-Maklaya str., Moscow, 117198, Russia*

A possible interpretation of symmetries in the many-dimensional table of elements by Y. Magarshak is suggested.

УДК 681.7

Оптоэлектронный измеритель угловых колебаний конструкций

В. А. Комоцкий, Ю. М. Соколов

*Кафедра радиофизики
Российский университет дружбы народов
Россия, 117198, Москва, ул. Миклухо-Маклая, 6*

Исследовано устройство для измерения угловых колебаний конструкций на основе фазовой дифракционной решетки с профилем в виде меандра. Выполнен теоретический анализ аппаратных характеристик и расчет оптимальных параметров датчика. Изготовлено несколько образцов датчиков с различными параметрами и проведено экспериментальное исследование угловых колебаний пробной конструкции.

Введение

Измерение малых угловых колебаний различных деталей и конструкций представляет практический интерес во многих областях науки и техники. Для этой цели могут использоваться различные варианты датчиков, например, пьезоэлектрические, волоконно-оптические, датчики с зеркальными отражателями. В настоящей работе мы рассмотрим оригинальное устройство измерителя малых угловых колебаний [1], которое дает возможность исследовать их форму, проводить измерение зависимости амплитуды колебаний от координаты расположения блока-сенсора на конструкции и от частоты, выявлять резонансные частоты.

Датчики для измерения угловых колебаний на основе фазовой дифракционной решетки рассматривались ранее в [2], где была исследована схема измерения угловых колебаний с применением дифракционной решетки, отделенной от исследуемой поверхности. К недостаткам этой схемы следует отнести возможность искажений результатов измерений за счет неучтенных вибраций держателя решетки. Новый датчик является более удобным в работе и не требует сложной настройки. По сравнению с предыдущим вариантом датчика в нем уменьшено количество отражающих поверхностей, поэтому снижается уровень помех из-за переотражений оптического пучка. Существенно, что данный датчик чувствителен только к угловым колебаниям поверхности и не реагирует на ее поступательное движение.

1. Устройство и физический принцип работы измерителя

Основным элементом измерителя, схема которого приведена на рис. 1, является блок-сенсор, представляющий собой прозрачную пластину, на одной стороне которой находится фазовая дифракционная решетка с профилем в виде меандра, а на другой стороне — зеркальная отражающая пленка. Блок-сенсор закрепляем на исследуемой конструкции и просвечиваем лазерным пучком. Луч лазера, направленный на блок, дифрагирует на фазовой дифракционной решетке, распространяется в прозрачном материале пластины, отражается от зеркальной пленки и повторно дифрагирует на фазовой решетке. Таким образом, в основе работы устройства лежит двойная дифракция оптического пучка на фазовой дифракционной решетке, имеющей профиль в виде меандра.

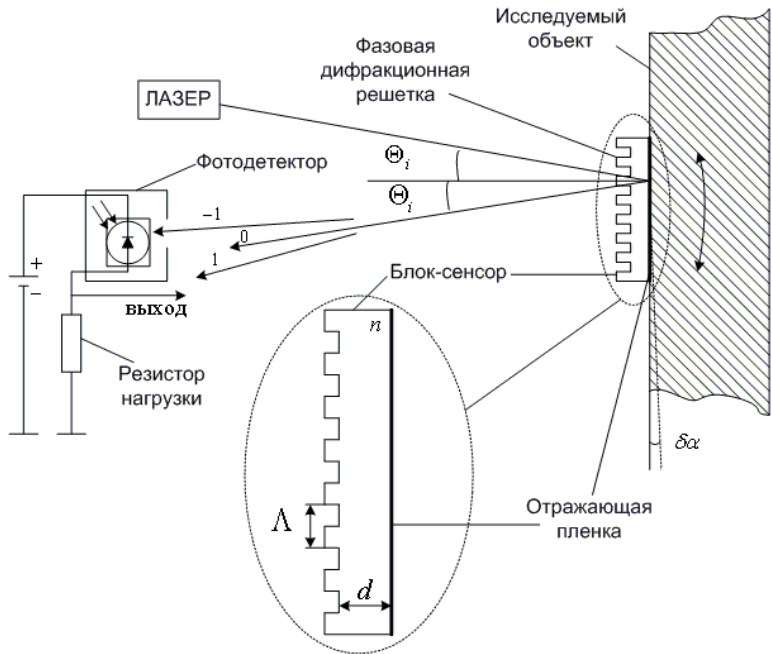


Рис. 1. Схема оптоэлектронного устройства для измерения угловых колебаний конструкций.

При наклоне блока-сенсора, закрепленного на конструкции, на небольшой угол $\delta\alpha$ происходит смещение следов падающего и отраженного пучков по поверхности решетки. Если заменить отраженный луч зеркально симметричным, то анализ взаимодействия оптического пучка с блоком-сенсором можно заменить анализом эквивалентной схемы из двух решеток, находящихся на расстоянии, равном удвоенной толщине блока-сенсора $2d$. Ранее в работе [3] был проведен теоретический анализ системы из двух фазовых решеток при просвечивании ее лазерным пучком, поэтому все теоретические результаты, полученные в [3], могут быть использованы также при анализе данного датчика.

При прохождении плоского волнового фронта с длиной волны λ через фазовую дифракционную решетку образуется пространственная фазовая модуляция (ПФМ) с амплитудой Φ_M , которая равна половине разности фаз между лучами, проходящими через выступ и впадину решетки (рис. 2). При падении светового пучка по нормали к поверхности решетки выражение для амплитуды Φ_M имеет вид

$$\Phi_M = \frac{\Delta\varphi}{2} = \frac{1}{2} \frac{2\pi}{\lambda} h(n-1). \tag{1}$$

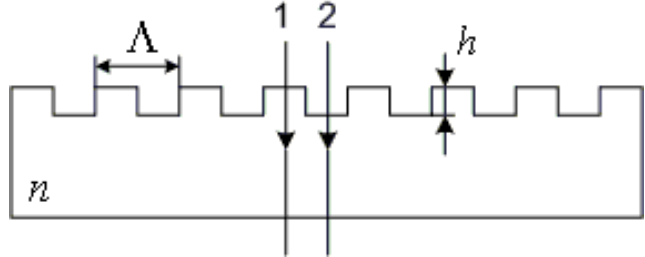


Рис. 2. Прохождение оптического излучения через выступ и впадину дифракционной решетки.

В [3] было показано, что для решеток с профилем в виде идеального меандра зависимости интенсивностей первых дифракционных порядков от смещения следа оптического пучка вдоль решетки описываются гармоническими функциями вида

$$I_{\pm 1}(\Delta x) = 2a_0^2 a_1^2 + 2a_0^2 a_1^2 \cos\left(\frac{2\pi}{\Lambda} \Delta x \mp L\right). \quad (2)$$

Выражение (2) записано для нормального падения луча. В нем $a_0 = \cos(\Phi_M)$ и $a_1 = \frac{2}{\pi} \cos(\Phi_M)$ — коэффициенты разложения в ряд Фурье функции пропускания фазовой решетки с формой меандра, величина $L = 2d\pi \frac{\lambda}{n\Lambda^2}$ — безразмерный параметр расстояния, соответствующий датчику на рис. 1. Для упрощения анализа в (2) коэффициент отражающей пленки R и интенсивность падающей волны $I_{\text{пад}}$ приняты равными единицам.

Второе слагаемое выражения (2) представляет собой переменную составляющую зависимости $I_{\pm 1}(\Delta x)$, амплитуда которой равна $I_{\pm 1}^{\approx} = 2a_0^2 a_1^2 = \frac{2}{\pi^2} \sin^2(2\Phi_M)$. Максимальное значение этого слагаемого достигается при $\Phi_M = \frac{\pi}{4}$. Это значение Φ_M и соответствующее ему значение физической глубины рельефа h являются оптимальными, так как при этом достигается максимальное значение крутизны линейного участка этой зависимости. Выражение (2) при $\Phi_M^{\text{opt}} = \frac{\pi}{4}$ принимает вид

$$I_{\pm 1}(\Delta x) = \frac{2}{\pi^2} + \frac{2}{\pi^2} \cos\left(\frac{2\pi}{\Lambda} \Delta x \mp L\right). \quad (3)$$

Приведем численную оценку оптимальной глубины рельефа. При значениях $\Phi_M^{\text{opt}} = \frac{\pi}{4}$, $\lambda = 0,65$ мкм, $n = 1,51$ (стекло) оптимальная глубина рельефа равна $h = 0,319$ мкм.

В формуле (3) смещение Δx следа оптического пучка вдоль решетки пропорционально углу наклона $\delta\alpha$, поэтому зависимости интенсивностей первых порядков от угла наклона блока-сенсора также будут описываться гармоническими функциями. Связь между Δx и $\delta\alpha$ найдем с помощью рис. 3, на котором показан вид блока-сенсора при наклоне на небольшой угол (*а*) и эквивалентная схема для анализа (*б*).

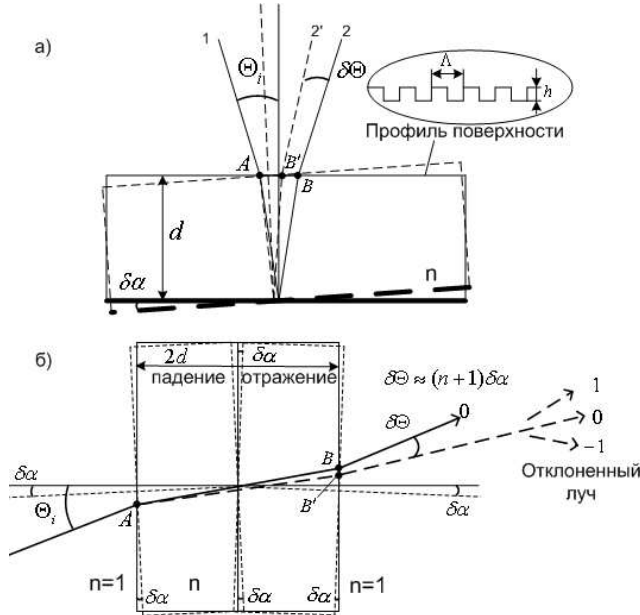


Рис. 3. Схема блока-сенсора при наклонном падении пучка (*а*). Эквивалентная схема из двух решеток в среде с показателем преломления n (*б*).

Линейное смещение Δx определится как разница расстояний между следами падающего и отраженного пучков до и после наклона датчика на угол $\delta\alpha$ с учетом преломления на границах разделов двух сред:

$$\Delta x = AB - AB' = 2d \left\{ \tan\left(\frac{\Theta_i}{n}\right) - \tan\left(\frac{\Theta_i - \delta\alpha}{n}\right) \right\} \approx 2d \frac{\delta\alpha}{n}. \tag{4}$$

Соотношение (4) справедливо для малых углов Θ_i и $\delta\alpha$.

Используя последнее выражение (4), связывающее смещение следа оптического пучка Δx с наклоном блока-сенсора на малый угол $\delta\alpha$, получим выражения для зависимостей интенсивностей первых дифракционных порядков от угла наклона блока

$$I_{\pm 1}(\delta\alpha) = \frac{2}{\pi^2} + \frac{2}{\pi^2} \cos\left(\frac{4\pi d}{n\Lambda} \Delta x \mp L\right). \tag{5}$$

Дифференцируя выражение (5), определим максимальное значение крутизны преобразования изменения интенсивности при изменении угла наклона блока-сенсора

$$S_{I\alpha \max} = \left| \frac{dI_{\pm 1}}{d(\delta\alpha)} \right|_{\max} = \frac{8d}{\pi n\Lambda}. \tag{6}$$

Для регистрации и измерения малых угловых колебаний используется один из множества линейных участков гармонической зависимости интенсивности одного из первых дифракционных порядков от угла наклона блока-сенсора $I_{\pm 1}(\delta\alpha)$ или в других терминах от эквивалентного смещения следа пучка $I_{\pm 1}(\Delta x)$. Зависимость от эквивалентного смещения следа пучка показана на рис. 4. Для работы устройства требуется обеспечить расположение рабочей точки на середине линейного участка. Это достигается выбором начального угла падения.



Рис. 4. Зависимость интенсивности первого дифракционного порядка от относительного смещения решеток. Рабочая точка находится на середине линейного участка.

Расчет начальных углов падения оптического пучка. Обозначим через Θ_{ik} углы падения, соответствующие положению рабочей точки на середине линейного участка зависимости $I_{\pm 1}(\Delta x)$. Расчет Θ_{ik} можно провести с помощью формулы (5), заменив угол $\delta\alpha$ углом Θ_{ik} и приравняв аргумент косинуса величине $\frac{\pi}{2} + \pi k$. В результате получим выражение для расчета Θ_{ik}

$$\Theta_{ik} = \frac{n\Lambda}{8d}(1 + 2k), \quad k = 0, 1, 2, \dots \tag{7}$$

На практике угол падения определялся в ходе калибровки, методика которой описана ниже.

2. Результаты экспериментальных исследований

Для проведения экспериментов были изготовлены образцы блоков-сенсоров с различными параметрами, которые приведены в табл. 1.

Таблица 1
Параметры используемых в экспериментах блоков-сенсоров

датчик 1	датчик 2	датчик 3	датчик 4
$\Lambda = 100 \text{ мкм}$	$\Lambda = 200 \text{ мкм}$	$\Lambda = 100 \text{ мкм}$	$\Lambda = 200 \text{ мкм}$
$d = 7,5 \text{ мм}$	$d = 11,2 \text{ мм}$	$d = 1,5 \text{ мм}$	$d = 7,2 \text{ мм}$
$\Phi_M = 22^\circ$	$\Phi_M = 40^\circ$	$\Phi_M = 38^\circ$	$\Phi_M = 40^\circ$

Измерение статических характеристик и калибровка датчиков. После изготовления блоков-сенсоров проводилось измерение статических характеристик и калибровка датчиков. Схема измерений приведена на рис. 5.

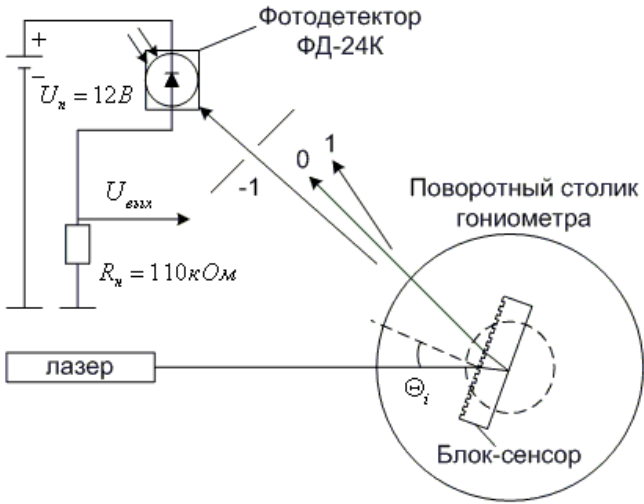


Рис. 5. Схема измерения статических характеристик и калибровки датчика.

При измерении статических характеристик блок-сенсор помещали на поворотный столик гониометра так, чтобы штрихи решетки были параллельны оси вращения столика. После этого выставляли некоторый небольшой начальный угол падения пучка и, поворачивая столик на угол $\delta\alpha$, проводили измерение величин выходных напряжений на нагрузках фотодетекторов первых порядках дифракции. Результаты измерений для датчика 1 приведены на рис. 6а. Измерения проводились с шагом $\delta\alpha = 1'$, начальный угол падения был равен $\Theta_i = 4^\circ$. На рис. 6б показаны расчетные зависимости интенсивностей первых дифракционных порядков от угла поворота для датчика 1 (толщина $d = 7.5 \text{ мм}$). Из сравнения рис. 6а и 6б видно, что экспериментальные и теоретические зависимости датчика 1 хорошо согласуются между собой.

Одновременно с измерением статических характеристик проводилась калибровка датчиков. Калибровка датчика включала следующие операции:

а) определение угла падения Θ_i , соответствующего положению рабочей точки на середине одного из множества линейных участков гармонической зависимости мощности первого дифракционного порядка от угла поворота датчика,

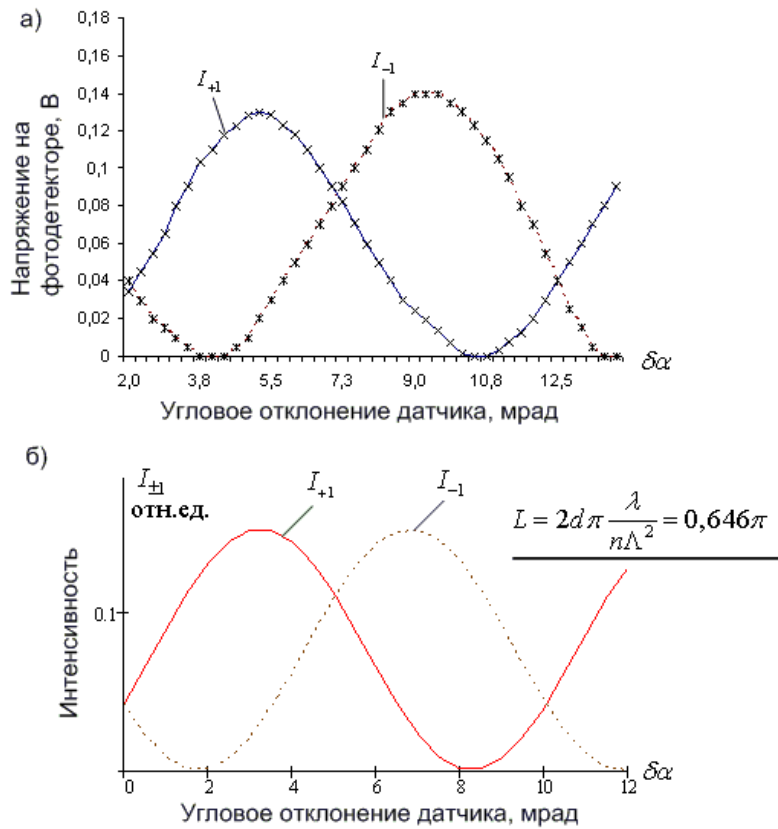


Рис. 6. Экспериментальные (а) и расчетные (б) статические характеристики первых дифракционных порядков для датчика 1.

б) определение крутизны преобразования $S_{U\alpha}$ изменения выходного сигнала ΔU при изменении угла поворота датчика.

Начальный угол падения пучка Θ_i выбирался таким, чтобы мощность дифрагированного пучка первого порядка была равна средней величине между ее максимальным и минимальным значениями. Поскольку ток и напряжение на нагрузке фотодетектора пропорциональны мощности падающего излучения, то на практике установка рабочей точки осуществлялась по уровню постоянного напряжения на выходе схемы, которое должно быть равно

$$U_{cp} = \frac{U_{max} + U_{min}}{2}, \tag{8}$$

где U_{max} и U_{min} — максимальное и минимальное значение напряжения выходного сигнала.

После установки рабочей точки на середину линейного участка производили калибровку крутизны преобразования. Для этого, давая малые приращения поворота блока-сенсора $\pm \delta\alpha$, измеряли приращения $\pm \Delta U$, а затем рассчитывали крутизну преобразования для данного блока-сенсора $S_{U\alpha} = \frac{\Delta U}{\delta\alpha}$. Для датчика 1 измеренное значение крутизны составило $S_{U\alpha} = 0,675$ В/мрад.

Результаты измерений колебаний конструкции. В качестве наглядного примера использования датчика на практике были измерены угловые колебания эталонной конструкции, которая представляла собой металлическую планку, закрепленную на двух вертикальных стойках. На одной из стоек устанавливали электромагнитный возбудитель, который возбуждал в пластине колебания с частотой от $F = 20$ Гц до $F = 1$ кГц. В качестве внешнего источника колебаний использовали генератор низких частот ГЗ-2. Схема экспериментальной установки показана на рис. 7.

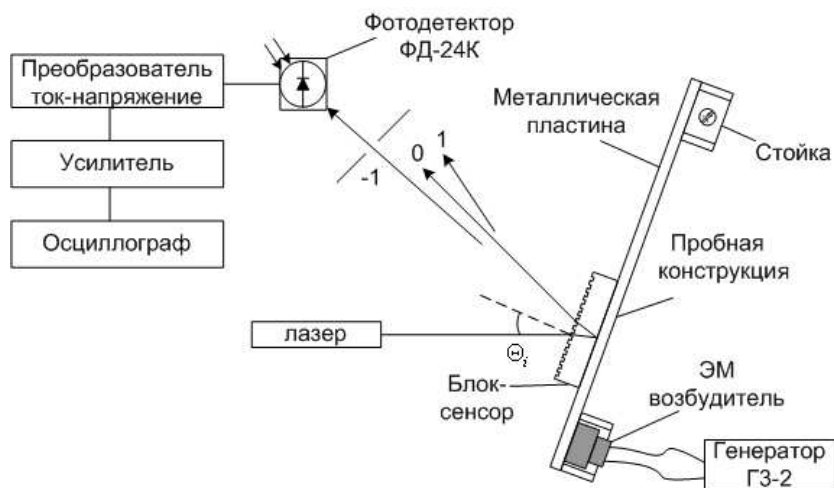


Рис. 7. Схема экспериментальной установки для измерения угловых колебаний эталонной конструкции.

При проведении эксперимента блок-сенсор приклеивали к металлической планке в различных точках. Угол падения, соответствующий середине линейного участка, определяли в процессе калибровки. Под этим углом на датчик направляли лазерный пучок. В образованной дифракционной картине выделяли один из первых порядков и направляли его на фотодетектор. После усиления сигнал с выхода фотодетектора поступал на осциллограф. Зная крутизну преобразования $S_{U\alpha}$, определенную при калибровке датчика и измерив амплитуду выходного сигнала, рассчитывали величину амплитуды угловых колебаний металлической конструкции $\Delta\alpha = \frac{\Delta U}{S_{U\alpha}}$. Изменяя частоту возбуждения и положение блока-сенсора на конструкции, измеряли частотные зависимости амплитуды угловых колебаний планки в различных ее точках. Результаты экспериментов, проведенных с применением датчика 1, показаны на рис. 8а, 8б, 8в. Из графиков на рис. 8 можно легко определить резонансные частоты колебаний. Так, на всех графиках хорошо проявляется резонансная область колебаний в районе $F = 95\text{--}120$ Гц.

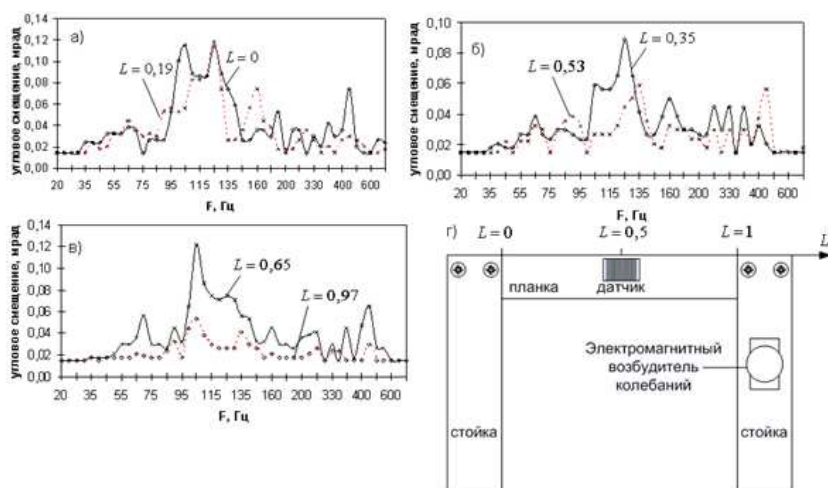


Рис. 8. Зависимости амплитуд угловых колебаний от частоты внешнего возбуждения при разных положениях датчика на конструкции: (а, б, в). Схема расположения датчика в различных точках на металлической планке показана на рисунке (г).

Так как датчик имеет небольшие размеры, то его легко перемещать вдоль конструкции. Таким образом было измерено распределение колебаний в различных точках конструкции на фиксированных частотах. Результаты измерений показаны на рис. 9 для четырех значений частот: $F = 105$ Гц, $F = 125$ Гц, $F = 160$ Гц и $F = 300$ Гц.

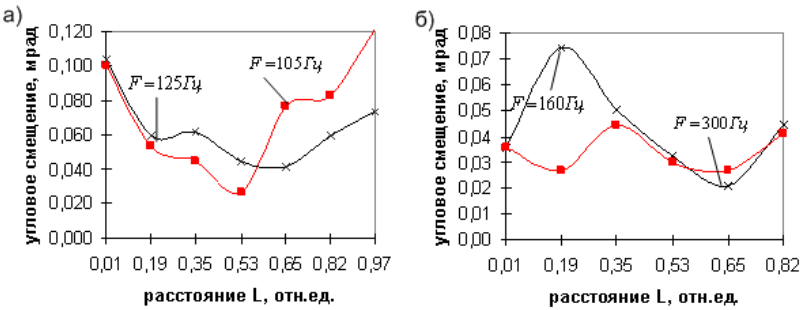


Рис. 9. Зависимости амплитуд угловых колебаний при разных положениях датчика на конструкции на фиксированных частотах (а, б).

В дополнение к проведенным исследованиям был выполнен специальный эксперимент по измерению колебаний конструкции с увеличенной массой. В этом эксперименте на конструкцию прикрепляли дополнительный груз, как это показано на рис. 10б, и проводили измерения частотных зависимостей амплитуды колебаний конструкции с грузом. Результаты этих измерений показаны на рис. 10а. Из сравнения двух экспериментальных кривых, соответствующих различным массам колеблющихся конструкций, видно, что при увеличении массы конструкции происходит уменьшение резонансных частот. Этот факт хорошо согласуется с теорией.

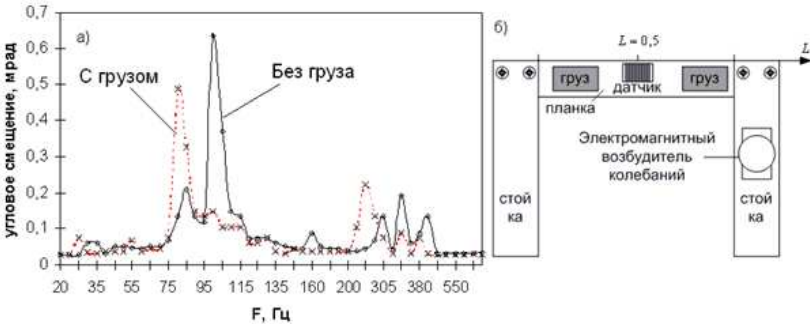


Рис. 10. Зависимости амплитуд угловых колебаний от частоты внешнего возбуждения при разных массах конструкций.

Заключение

Предложенный в настоящей работе датчик дает возможность проводить измерения зависимостей амплитуды колебаний от частоты, выявлять резонансные частоты, исследовать зависимость амплитуды колебаний от координаты расположения датчика на конструкции, а также исследовать форму колебаний. Он может быть использован на практике для исследования низкочастотных колебаний крупногабаритных конструкций. Отличительной чертой измерителя является то, что он способен регистрировать колебания с очень низкими частотами, вплоть до нулевых частот.

Технология изготовления датчика и методика проведения эксперимента достаточно просты. Экспериментально измеренная пороговая чувствительность датчика составила величину порядка сотых долей миллирадиана при полосе канала регистрации от 0 Гц до 1000 Гц.

Датчик нечувствителен к поступательным колебаниям конструкций и случайным вибрациям фотодетектора и позволяет получать результаты, выраженные в абсолютных единицах углов, а не в относительных единицах.

Литература

1. *Комоцкий В. А., Соколов Ю. М.* Оптоэлектронное устройство для измерения угловых колебаний конструкций. — Патент на полезную модель (приоритет от 23.05.2006 по заявке № 2006117556).
2. *Комоцкий В. А., Никулин В. Ф.* Бесконтактный дифракционный метод измерения угловых смещений и вибраций отражающих поверхностей // *Оптика и спектроскопия*. — Т. 72, № 2. — 1992. — С. 479–486.
3. *Ниибизи А., Комоцкий В. А.* Теоретический анализ взаимодействия оптической волны с системой пространственно разделенных периодических решеток // *ВИНИТИ* № 661-В86. — 1985.

UDC 681.7

Optoelectronic Device for Measurements Angular Oscillations of Constructions

V. A. Komotskii, Yu. M. Sokolov

*Department of Radiophysics
Peoples' Friendship University of Russia
6, Miklukho-Maklaya str., Moscow, 117198, Russia*

New optoelectronic device for measurements angular displacements of constructions based on phase diffraction grating with square waveform is investigated. Theoretical calculation and optimization of characteristics are executed. Some samples of transducers are manufactured, and experimental investigation of angular oscillations of standard construction is executed.

UDC 537.3113

Specific Electrical Conductivity of Natural Syrian and Armenian Zeolites

S. Sh. Soulayman ^{*}, A. A. Sahakyan [†], S. Nikogosyan [†], S. A. Yunusova [‡]

^{*} Department of Applied Physics
Higher Institute of Applied Sciences and Technology
Damascus, P. O. Box 31983, Syria

[†] Yerevan Physics Institute
Alikhaniyan Bros. Str., Yerevan, 375036, Armenia

[‡] Department of Theoretical Physics
Peoples' Friendship University of Russia
6, Miklukho-Maklaya str., Moscow, 117198, Russia

The direct current (dc) conductivity of several natural Syrian and Armenian zeolitic samples is measured in order to understand the mechanism of electrical conductivity in these materials. The influence of the sample's water content on its electrical conductivity is studied in details. We find that with the increase of hydration time or, equivalently, the increase of water content in the sample, the electrical conductivity of the samples increases up to a definite moment at which the increase stops. This moment is characteristic of each sample and it corresponds to the saturation state which means the state where the hydration process is finished and that the sample has reached its equilibrium state corresponding to an air-dried sample. We compare our results with available experimental data of relatively other materials and find that they agree in general.

Introduction

The measurement of the specific electrical conductivity σ of zeolites is one of the most attractive procedures to determine their properties [1–10]. Usually, the measurement of σ of zeolitic samples takes place at the alternative current (ac) condition (variable electrical field) in order to avoid various effects such as polarization, relaxation processes, electrolyze, etc [1–8]. In this case, however, σ is highly dependent on the applied frequency of the electrical field. Therefore, it is recommended that one measures σ within a wide range of frequencies in order to have an idea, sufficiently correct, about the mechanism of electrical conductivity in the zeolites [2]. Moreover, the measurement of σ at constant electrical field conditions (direct current (dc) conditions) somehow gives an overestimate. The results of the measurements, obtained using these two methods, could be integrated in order to give a real picture of the electrical conductivity [3–7].

In addition to the above mentioned properties, it is well known that zeolites have a large forbidden energy gap ($\sim 7\text{eV}$) [8], and that their electrical conductivity, in value, is near to that of bad dielectrics. On the other hand, the characteristic behavior of the electrical conductivity of dielectrics is such that when applying a constant electrical field on a dielectric sample, the intensity of the current, passed through the sample, decreases with time [11–13]. In fact, at the start, bias current with quickly decreasing intensity passes through the circuit, then, this current disappears at a time equal to the order of the RC system (generator-sample) constant time, which is very small (less than 1 s). However, the current can continue its decrease for some minutes, voire some hours in cases of samples with big dimensions. The time required for current stabilization could be decreased by adjusting the dimensions of the sample.

The component of slow changing current is due to the redistribution of free charges in the volume of the dielectric. This component is known as the absorption current,

and it is caused by the absorption of free current carriers in the volume of the dielectric. Part of these carriers, during their movement, will meet defects-traps of crystalline lattice causing them to be trapped and stopped. With time, the traps could be filled by current carriers, then, the absorption current disappears. The rest of the current does not depend on time and it is due to the charge carriers which pass from one side of the sample to the other (i.e. from one electrode to another). Thus, the absorption current leads to the accumulation of current carriers in separate parts of the dielectric, for example in the destroyed parts of the crystalline lattice, or at the grain boundaries, or/and in other types of inhomogeneities. As a result of the space charge appearance, the distribution of the electrical field becomes inhomogeneous. The accumulation of space charge in the dielectric leads to an undesired phenomenon in that the charged capacitor does not discharge totally after wiring its plates (short circuit). This effect is characterized by the coefficient of absorption which is determined as the ratio of the residual to the initial voltage.

When measuring the specific electrical conductivity, the absorption current could be eliminated by applying temporarily a constant voltage on the sample electrodes. All what we mentioned of phenomena do happen in the zeolitic samples and the arguments above do apply. Thus, the measurement of the specific electrical conductivity of zeolites is completed, in fact, by measuring the current which passes through the samples, i.e. the steady leakage current [11].

Our work deals with the measurement of the specific electrical conductivity of natural Syrian and Armenian zeolites.

1. Experimental Set-Up

In order to measure the dc conductivity σ , a universal voltmeter V 7-16 is used with an input resistance equal to $10\text{ M}\Omega$ (for dc), in addition to a nano-voltmeter P-341 having an input resistance of $100\text{ }\Omega$ approximately and minimal scaling of 0.005 nA (for dc). Since the input resistance of the voltmeter is less than the resistance of the sample itself (or near to it), the voltmeter is not connected to the sample directly. As usually done, in the case of samples with low resistance, the voltmeter should be connected in parallel to the sample and to the Amber meter simultaneously in order to avoid current pass through the voltmeter. The mentioned scheme is well described in the literature. In fact, the voltmeter is connected to one side of the sample while the other side is connected to the Amber meter.

2. Sample Preparation

In this work, the investigated samples have been prepared from grounded Syrian (S and M) and Armenian (A) natural zeolites. The Armenian zeolite, denoted by A11, contains up to 85% of clinoptilolite with a ratio of Si/Al equal to 9.8 [14] and amorphous SiO_2 less than 3%. Around 50% of the Syrian zeolitic substance contains, mainly, the three types of zeolite (analcime, phillipsite and chabazite). Other contaminations in raw materials for Armenian natural zeolite are: feldspar — 5%, quartz — 5%, mica — 2%, clays — 3%, whereas these values are quite different in the natural Syrian zeolite where the availability of calcite, quartz, clays and amorphous phase is proven in most of the samples. The quantity of each of the mentioned minerals depends on the location of the site from which the sample was taken along the borehole. These results were obtained by chemical and thermo-gravimetric analysis. According to this analysis and measurements, the endothermic effect of dehydration is observed at the temperatures between 70° and 300°C with simultaneous mass loss of up to 8.2% [14].

The chemical composition of the studied samples is given in Table 1. Zeolite powder was pressed into pellets at pressure of 200 kg/cm^2 . The pellet was mechanically cut out into rectangular samples with area of approximately 30 mm^2 and thickness of 2–3 mm. On both sides of the sample surfaces, silver contacts were made. Then, the samples were annealed at temperature 120°C for two hours, followed by thermal treatment at

220°C for two extra hours in the surrounding environment. This processing makes it possible to obtain reliable silver contacts. Measurement was carried out in the air and at temperature 24°C at relative humidity of 50–55%. The sample was stored at normal atmospheric pressure and it contained, of its weight, about 11.2% water. The sample temperature was measured by germanium diode with accuracy of ± 0.2 K.

The chemical composition of the studied samples Table 1

L.I.	P ₂ O ₅	Cr ₂ O ₃	Mn ₂ O ₃	TiO ₂	MgO	Fe ₂ O ₃	CaO	K ₂ O	Na ₂ O	Al ₂ O ₃	SiO ₂	Sample
14.19	—	—	—	0.20	1.28	2.23	4.9	2.22	0.79	11.69	67.11	A11
12.3	0.44	0.03	0.2	1.60	12.21	11.68	8.77	1.01	1.22	10.48	39.51	2S14
14.10	0.46	0.03	0.17	1.48	11.09	10.41	11.01	0.89	0.93	9.98	38.69	4S14
15.16	0.52	0.03	0.17	1.60	9.74	10.65	11.92	1.08	0.99	10.15	37.64	6S14
16.52	0.52	0.03	0.15	1.37	8.55	8.97	14.95	0.84	0.75	9.80	37.14	8S14
10.98	0.41	0.04	0.22	1.64	14.17	13.20	6.73	1.02	1.13	10.42	39.95	10S14
10.46	0.49	0.03	0.20	1.84	9.99	12.02	8.24	1.13	1.50	11.65	40.76	12S14
10.71	0.47	0.03	0.18	1.75	10.65	11.75	8.95	1.13	1.66	10.97	40.17	M5
9.28	0.52	0.03	0.20	2.10	9.01	12.60	8.26	1.24	1.45	13.04	41.36	16S14
9.31	0.49	0.03	0.19	1.92	10.45	11.99	8.18	1.19	1.43	12.00	41.08	18S14
9.24	0.47	0.03	0.19	1.87	10.62	12.30	7.96	1.18	1.65	11.63	41.23	20s14

Measurements of electro physical parameters were carried out on pressed samples. The samples are made of Syrian and Armenian natural zeolite powder (with grain size of 50 Micron). It should be noted that, after switching the constant electric field, the current passing through the sample is established at a time interval which gets beyond several hours sometimes (especially for large samples). This means that, in these cases, relaxation processes are taking place. The main reason for such relaxations is the polarization of the sample which can be taken into account by a well known method [11–13].

We should mention here that during the period of storage, relative changes in the weight of the samples due to weather changes did not surpass 0.5 l. The absolute accuracy in the weight measurement was 0.05 mg; hence the accuracy of measurement for σ is about 5%.

3. Results and Discussion

After having carried out the thermal treatment according to the above mentioned methodology, the dc conductivity of dried air samples was measured at a temperature equal to 24°C using a constant electrical field of $1 \div 3$ V/cm. The weight of the sample was measured also with an absolute accuracy of ± 0.05 mg.

In order to study the influence of the sample’s water content on its electrical conductivity, all the studied samples were put in a closed container for 4–5 days at a saturated vapor pressure corresponding to a temperature equal to 24°C. This procedure allows the samples to reach their state of total saturation. After that, each one of the studied samples was put in a special cell and its dc electrical conductivity was measured dynamically as a function of time at open air conditions. Thus, we could acquire the dependence of the dc conductivity on the dehydration degree. The dehydration degree is determined as $d_h = (m_0 - m) \cdot 100\% / m_0$, where m_0 is the mass of the sample with maximal water content and m is the current mass of the same sample. Some of the obtained results are given in Table 2. In addition, we present in Figure 1 the dependence of dc conductivity of Armenian sample (A_{11}) on the dehydration period length. After dehydration, Syrian samples became unsuitable for measuring the dc conductivity since they conserved weakly their shape after the pressing. This is due, perhaps, to the inclusion of different phases in these samples

(especially quartz and different zeolitic types). These samples could be destroyed when they reached total saturation. Therefore, the electrical conductivity of the Syrian samples was measured at the hydration mode, starting from partially dehydrated state up to totally dehydrated one. Therefore, the results corresponding to the state of total saturation are not given in Table 2 (see 2nd column). We think that these data can be completec

technology.

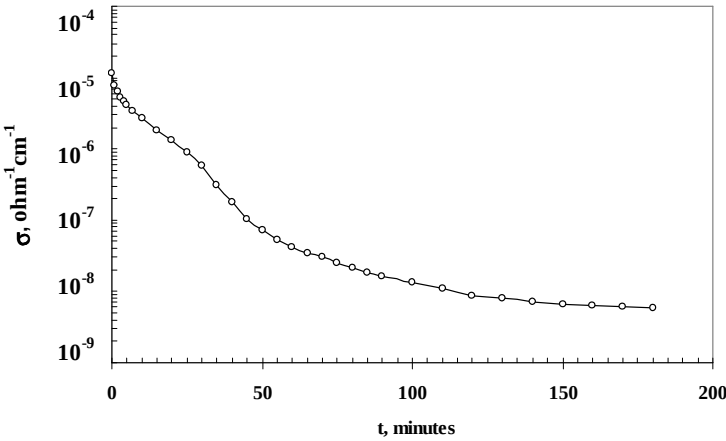


Figure 1. The hydration time variation of the specific conductivity of natural Armenian zeolite (clinoptilolite) at temperature 24C

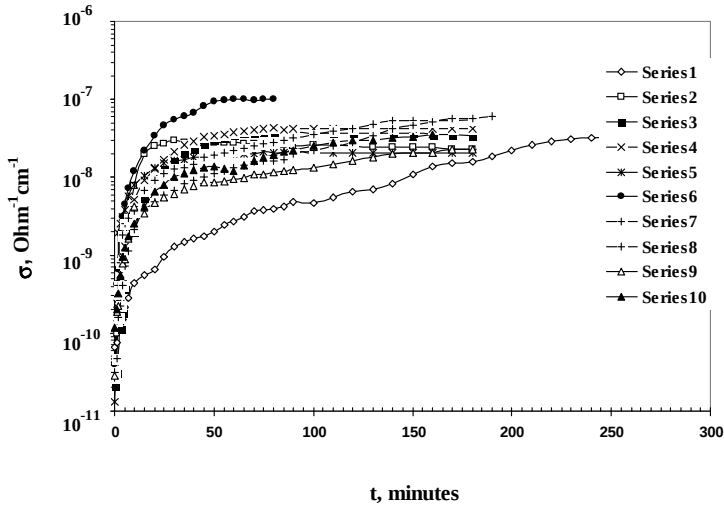


Figure 2. The hydration time variation of the specific conductivity of natural zeolitic samples at temperature 24°C. 1-A11; 2-M5; 3-2S14; 4-4S14; 5-6S14; 6-8S14; 7-10S14; 8-12S14; 9-16S14; 10-20S14

It should be noted also that the dehydration degree (c.f. the 1st row in Table 2) is evaluated using for m_0 the value of the mass when the sample was at the state of total saturation. In contrast, the evaluation, in the other rows, was made using for m_0 the value at the partial dehydration state which corresponded to when the sample was thermally treated for two hours in open air conditions at temperature 220°C. It is seen from Figure 1 that the sample A_{11} reaches its equilibrium state

Table 2
Specific electrical conductivity at T=24°C for natural Armenian (A₁₁) and Syrian (M and S) zeolites corresponding to different degrees of dehydration d_h

Samples	σ (in $10^{-11} \Omega^{-1}\text{cm}^{-1}$) for corresponding degree of dehydration (in %)		
A ₁₁	1200000 for d _h = 0	677 for d _h = 9.22	1.5 for d _h = 14.8
A ₁₁	—	1220 at ambient air	6.58 for d _h = 5.73
M5	—	618 at ambient air	3.52 for d _h = 6.65
2S14	—	1190 at ambient air	2.07 for d _h = 5.32
4S14	—	1930 at ambient air	4.25 for d _h = 4.8
6S14	—	2210 at ambient air	1.31 for d _h = 6.4
8S14	—	4030 at ambient air	4.04 for d _h = 4.3
10S14	—	4290 at ambient air	2.68 for d _h = 6.16
12S14	—	2860 at ambient air	3.14 for d _h = 7.4
16S14	—	1480 at ambient air	2.84 for d _h = 5.52
20S14	—	1730 at ambient air	11.9 for d _h = 5.3

within three hours, and here the dc conductivity becomes independent of time. In this case, the sample passes from total saturation state to dried-air state as a result of water emigration from the sample to the surrounding air. In other words, we have a dehydration sample at that time. This situation is accompanied by σ , decrease from $1.2 \cdot 10^{-5} \Omega^{-1}\text{cm}^{-1}$ to $6.77 \cdot 10^{-9} \Omega^{-1}\text{cm}^{-1}$ in the case of Armenian samples. The measured values of dc electrical conductivity for all samples at some values of dehydration degree are presented in Table 2. It is characteristic of sample A₁₁ that the dependence of its conductivity on the water content in the mode of hydration differs from that of the hydration mode. This can be seen clearly by dividing the values of the dc conductivity, given in the 3rd column of Table 2, by the corresponding values in the last (4th) column in the same table to give the dc value of σ , the partially dehydrated state. In such a way, for Armenian samples, we find ($677/1.5 = 451$) and ($1220/6.58 = 185$) respectively. This behavior of conductivity at the mentioned two modes agrees well with the isothermal curves of sorption and desorption. For the Syrian samples, the following values were obtained (see Table 2, from top to bottom) 175; 575; 454; 1687; 998; 1600; 911; 521 and 145 respectively. The dependence of $\sigma(t)$ on the hydration time is illustrated in Figures 2–5. Figure 2 contains the data for all samples, whereas, for clarity and to make the comparison easier, the samples' data have been divided into groups and are presented in Figures 3–5.

It is seen from Figures 2–5 that with the increase of hydration time (i.e. with the increase of water content in the sample) the electrical conductivity of the samples increases until a definite moment when the increase stops. This moment is characteristic of each sample and it corresponds to the saturation state indicating that the hydration process has been finished and that the sample has reached its equilibrium state (air-dried sample). For some samples, this process goes relatively quickly (from 0.5 hour to 1 hour) while for others it goes more slowly. One can also notice additional stages on the curve of time dependence of electrical conductivity $\sigma(t)$. This is due, probably, to the inhomogeneous phase content of the corresponding samples.

Moreover, when comparing σ for natural Armenian and Syrian zeolitic samples one finds that during the process of establishing total hydration, the value of σ increases 2.5 times in magnitude for the Armenian samples, while for the Syrian samples, the same value triples its magnitude, and sometimes increases even more.

On the other hand, one can compute from Table 1, where the oxide contents of Armenian and Syrian samples is given, the molar ratio $\text{SiO}_2/\text{Al}_2\text{O}_3$ for these samples. When doing this, we find that this ratio is in the range ($3.17 \div 3.88$) for the Syrian

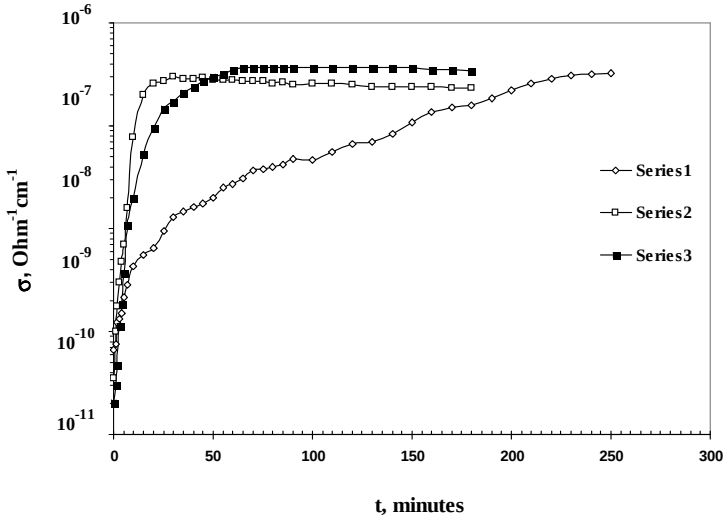


Figure 3. The hydration time variation of the Specific conductivity of natural zeolitic samples (1-A11; 2-M5; 3-2S14) at temperature 24°C.

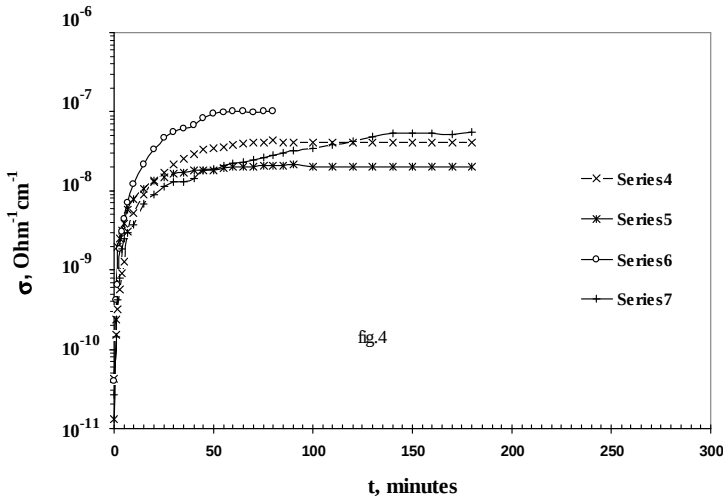


Figure 4. The hydration time variation of the Specific conductivity of natural Syrian zeolitic samples (4-4S14; 5-6S14; 6-8S14; 7-10S14) at temperature 24°C

samples, while its value, for Armenian zeolite, could change in the range (5.74 ÷ 5.91) [14]. It is known [2,9,10] that samples with smaller values of this ratio are more conductive and less stable. Whence, the higher values of electrical conductivity of Syrian samples could be explained. However, it should be mentioned that, at relatively low temperatures, the measured values of σ and the stability of zeolitic samples are determined by their cation contents [1,9].

The mechanism of conductivity in zeolitic samples has a jumpy character and the ions or protons play, in general, the role of current carriers [1,2,5–7]. The content and the structure of the zeolitic sample determine its sorbitizing properties, in relation to water vapors [2,9,10]. In this case, with the increase of water content, σ increases while the tilt of the curve of $\lg \sigma$ dependence on water concentration decreases. This

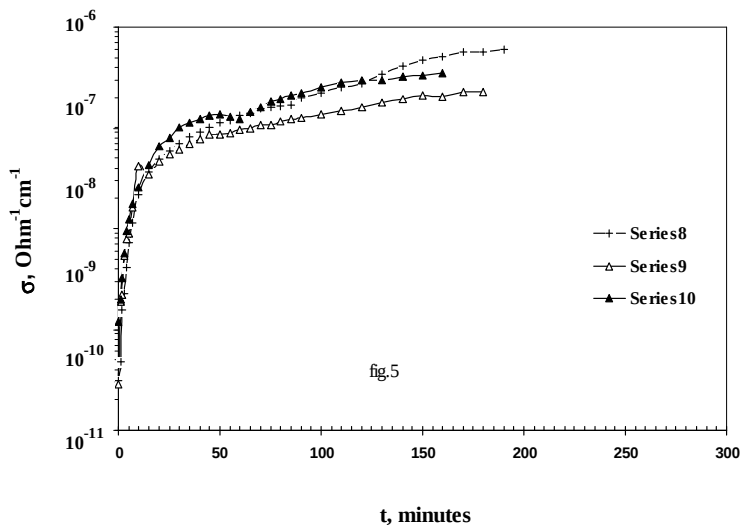


Figure 5. The hydration time variation of the specific conductivity of natural Syrian zeolitic samples (8-12S14; 9-16S14; 10-20S14) at temperature 24°C

means that the activation energy of conduction decreases also [2, 9, 10]. In our case, the availability of different tilts on the curve $\lg \sigma - t$ confirms this result.

Conclusions

One can conclude that the complete explanation of the observed phenomena requires measuring σ in wider ranges of temperature, hydration and studied samples. This can be achieved using different chemical and radiation methods [5–7, 14, 15]. These results should be accompanied as well by the results of other related physical measurements. We have to mention also that the measurement of surface conductivity in addition to the volume conductivity could help in understanding many phenomena especially the process of sorption which passes through the surface. At present, we are doing this kind of measurement and the preliminary results show that it is highly sensitive to the humidity and temperature in comparison with the volume conductivity.

References

1. Freeman D. C. J., Stamires D. N. // The Journal of Chemical Physics. — Vol. 35. — 1961. — P. 799.
2. Simon U., Franke M. E. // Micropor. Mesopor. Mater. — Vol. 41. — 2000. — P. 1.
3. Sayed M. B. AC-Conductivity Analysis of the Partial Dealumination and Gamma-Irradiation on the Electric Features of HZSM-5 // Zeolites. — Vol. 16. — 1996. — P. 157.
4. Sayed M. B. // Microporous Materials. — Vol. 6. — 1996. — P. 181.
5. Sayed M. B. // Micropor. Mesopor. Mater. — Vol. 37. — 2000. — P. 107.
6. Higazy A. A., Kassen M. E., Sayed M. B. // J. Phys. Chem. Solids. — Vol. 53. — 1992. — P. 549.
7. Sayed M. B. // J. Phys. Chem. Solids. — Vol. 53. — 1992. — P. 1041.
8. Franke M. E., Simon U. // Phys. Stat. Sol. (b). — Vol. 218. — 2000. — P. 287.
9. Breck D. N. Zeolite Molecular Sieves. — New York: Willey, 1974. — P. 571.
10. Stamires D. N. // The Journal of Chemical Physics. — Vol. 36. — 1962. — P. 3174.
11. Tareev B. M. et al. Electrорadiomateriali. — Vysshaya Shkola, 1978. — 336 p.

12. *Borisova M. E., Koikov S. N.* Physics of Dielectrics. — Leningrad: Izdatel'stvo Leningradskogo Universiteta, 1979. — 240 p.
13. *Volkenshtein F. F.* Electrons and Crystals. — Moscow: Nauka, 1983. — P. 128.
14. *Gevorkyan R. et al.* Zeolite Modification and Application Study for Decontamination Nuclear Liquid Waste Part1. // 13th Inter. Zeolite Conf. Montpellier. — 2001. — 31-R-09. R.G.Gevorkyan, H.H.Sargsyan, G.G.Karamyan, Y.M.Keheyanyan, H.N.Yeritsyan, A.S. Hovhannisyan, and A.A. Sahakyan, Study of "Absorption Properties of Modified Zeolites", Chem.Erde **62** (2002) 237.
15. *Ermakov S. E. et al* // J. Phys. Chem. — Vol. 54 (10). — 1980. — P. 2524.

УДК 537.3113

Удельная электрическая проводимость натуральных цеолитов из Армении (клиноптилолит) и Сирии

С. Ш. Сулейман *, А. А. Саакян[†], С. Никогосян[†], С. А. Юнусова[‡]

* Кафедра прикладной физики
Институт прикладных технологий
Дамаск, П.Я. 31983, Сирия

[†] Ереванский физический институт
Ереван, 375036, Армения

[‡] Кафедра теоретической физики
Российский университет дружбы народов
Россия, 117198, Москва, ул. Миклухо-Маклая, 6

В этой работе мы проводили измерения электрической проводимости (σ) натуральных цеолита Сирии и Армении в постоянном электрическом поле. Изучены влияния влажности на проводимость образцов. Для всех образцов проводились временные зависимости проводимости (в режиме гидратации) от частично обезвоженного состояния до воздушно-сухого состояния. Можно отметить следующие основные результаты: с ростом времени гидратации (т. е. с ростом содержания воды) наблюдается увеличение проводимости образцов, и через некоторое время она достигает насыщения, что означает конец процесса гидратации, образец переходит в равновесное состояние (воздушно-сухой образец).

ПРАВИЛА ОФОРМЛЕНИЯ СТАТЕЙ

В журнале печатаются оригинальные и обзорные статьи российских и зарубежных авторов по следующей тематике: 1) Математика; 2) Математическая теория телетрафика; 3) Математическое моделирование; 4) Теоретические основы информатики; 5) Физика; 6) Квантовая и статистическая физика; 7) Радиофизика.

Редколлегия журнала «Вестник Российского университета дружбы народов», серия «Математика. Информатика. Физика» просит авторов придерживаться следующих правил при представлении статьи в журнал.

1) Статьи представляются на русском или английском языке.
2) Объём статьи не должен превышать 1 печ. л.
3) Автор представляет в редакцию электронную версию рукописи, набранную в системе $\text{\LaTeX}$ (используется версия $\text{\LaTeX} 2_{\epsilon}$, для набора формул используется макропакет $\mathcal{A}\mathcal{M}\mathcal{S}\text{-}\text{\LaTeX}$). К электронному варианту прилагается отпечатанный на бумаге экземпляр или файл в формате Postscript или PDF.

4) Для каждой статьи указываются коды УДК и MSC (PACS), а также ключевые слова (на языке статьи).

5) Название, аннотация, фамилии и инициалы авторов, название организации, где работают авторы, предоставляются на русском и английском языках.

6) Текст статьи должен включать аннотацию на русском и английском языках (в аннотации не допускаются ссылки на цитированную литературу и громоздкие формулы), введение, заключение. Глубина разбивки текста не должна превышать трёх уровней (разделы, пункты и подпункты).

7) Рисунки принимаются в электронном виде. Каждый рисунок должен быть помещён в отдельный файл. Принимаемые форматы файлов: 1) векторные: EPS, PDF, $\text{\TeX}$; 2) растровые: TIFF, GIF, PNG (возможна инкапсуляция в EPS).

8) Размер рисунка вместе с подписью не должен превышать 14x19 см. Разрешение растрового рисунка должно находиться в пределах 300–600 dpi.

9) Рисунки должны быть чёрно-белые. Возможность использования полутонных и фотографических изображений может быть рассмотрена отдельно. Фоны должны быть только штрихованные. Сеточные фоны и полутона не допускаются.

10) Список литературы подготавливается в системе $\text{\BibTeX}$ и должен соответствовать требованиям ГОСТ 7.1-84. Ссылки на неопубликованные работы не допускаются.

11) Рукопись должна быть тщательно выверена. Необходимо указать точный почтовый и электронный адрес места работы авторов и телефоны. После подготовки редакцией к набору размеченный и исправленный автором текст статьи и исправленная электронная версия возвращаются в редакцию. Корректра для просмотра высылается по e-mail.

12) Возвращение статьи автору на доработку не означает, что она принята к опубликованию. Доработанный вариант статьи редколлегия рассматривает вновь. В случае отклонения статьи редколлегия оставляет за собой право не возвращать автору один ее экземпляр.

ВЕСТНИК
Российского университета дружбы народов
Научный журнал. Основан в 1993 г.

Серия
МАТЕМАТИКА. ИНФОРМАТИКА. ФИЗИКА
2007, № 1–2

Редактор *Т. О. Сергеева*
Технический редактор *Н. А. Лощенова*
Компьютерный набор *Н. Э. Николаев, А. В. Королькова*

Издательство Российского университета дружбы народов
117923, ГСП-1, г. Москва, ул. Орджоникидзе, 3

Типография ИПК РУДН
117923, ГСП-1, г. Москва, ул. Орджоникидзе, 3, тел. 952-04-41