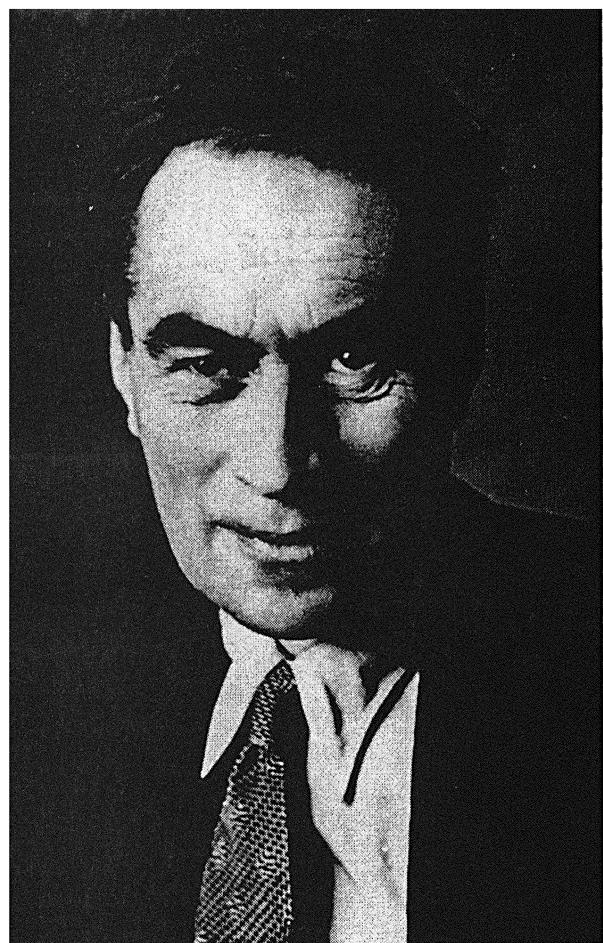


Содержание

Фейгин А. М., Мольков Я. И., Мухин Д. Н., Лоскутов Е. М. Прогноз качественного поведения динамической системы по хаотическому временному ряду .....	376
Зоннеманн Г., Семёнов В. Е. Автосолитоны в системах с глобальной нелинейностью.....	399
Литвак А. Г., Токман М. Д. Об эффекте электромагнитной индуцированной прозрачности в классических системах.....	407
Иудин Д. И., Трахтенгерц В. Ю. Фрактальная структура нелинейной динамики электрического заряда в грозовом облаке .....	419
Rabinovich M. I., Pinto R. D., Huerta R. Some problems of information neurodynamics .....	439
Корытин А. И., Крячко А. Ю., Сергеев А. М. Диссипативные солитоны в комплексном уравнении Гинзбурга—Ландау для фемтосекундных генераторов света .....	464
Kocharovskiy V. V. and Kocharovskiy VI. V. Cooperative parametric instability of natural modes and coherent mechanism of self-mode-locking in a ring class-B laser.....	480
Ostrovsky L. A. and Johnson P. A. Nonlinear dynamics of rock: hysteretic behaviour.....	487
Горшков К. А., Соустова И. А. Взаимодействие солитонов как составных структур в модели Гарднера .....	502
Троицкая Ю. И. Асимптотические модели диффузионных эффектов при нелинейном резонансном взаимодействии волн с потоками .....	515
Гинзбург Н. С., Песков Н. Ю., Сергеев А. С., Аржанников А. В., Синицкий С. Л. Генерация пространственно-когерентного излучения в лазерах на свободных электронах с двумерной распределённой обратной связью .....	533

# **СПЕЦИАЛЬНЫЙ ВЫПУСК**

**100-летию со дня рождения А. А. Андронова посвящается**



Уважаемые читатели,

Настоящий выпуск журнала «Радиофизика» посвящен 100-летнему юбилею академика Александра Александровича Андронова (1901–1952) — выдающегося русского ученого, одного из основоположников теории динамических систем и нелинейной науки в целом в бывшем Советском Союзе и, мы полагаем, в мире. Его основные научные достижения относятся к теории динамических систем, теории бифуркаций и их применению к автоматическому управлению. Несмотря на то, что он был физиком по образованию, его фундаментальная математическая подготовка в Московском университете, глубокая интуиция и вкус к постановке математически строгих задач позволили ему сформулировать и решить проблемы, определившие развитие этих теорий на многие годы. Благодаря его вкладу многие физические задачи получили адекватную математическую формулировку и интерпретацию. Отметим здесь автоколебания (генерация, эволюция и синхронизация), релаксационные колебания, конкуренцию и захват мод, шумовые эффекты в динамических системах, автоматические регуляторы и др. Он был талантливым учителем и научным руководителем для многих ученых и студентов, создателем известной в мире Горьковской школы нелинейных колебаний и соавтором известных книг, сыгравших важную роль в подготовке нескольких поколений ученых в области нелинейной науки: «Теория колебаний» Андронова, Витта и Хайкина, двухтомник «Качественная теория динамических систем второго порядка» и «Теория бифуркаций динамических систем на плоскости» Андронова, Леонтьевича, Гардона и Майера. Подчеркнем также, что А. А. Андронов был одним из основателей факультета радиофизики в Нижегородском (Горьковском) университете в 1945 г. и одним из его первых блестящих лекторов. Это был первый радиофизический факультет в бывшем Советском Союзе, и по сей день он остается одной из ведущих школ радиофизики и родственных областей (физики плазмы, СВЧ волн, квантовой электроники, акустики, радиотехники и др.). Без всякого преувеличения можно сказать, что Александр Александрович Андронов был одним из ученых, сформировавших современное лицо теории колебаний динамических систем и нелинейных колебаний. Его подход к науке и образованию, глубокое понимание будущего прогресса нелинейной науки и яркая индивидуальность останутся в памяти его многочисленных учеников, последователей и всех тех, кто ценит науку.

В данном выпуске журнала представлены работы, демонстрирующие развитие нелинейной науки в самых разных областях, от математики до биологии. Надеемся, что они интересны самой широкой аудитории и дадут возможность вспомнить человека, который был одним из основателей этой междисциплинарной области знаний.

УДК 538.56:517.33:621.373

## ПРОГНОЗ КАЧЕСТВЕННОГО ПОВЕДЕНИЯ ДИНАМИЧЕСКОЙ СИСТЕМЫ ПО ХАОТИЧЕСКОМУ ВРЕМЕННОМУ РЯДУ

*А. М. Фейгин, Я. И. Мольков, Д. Н. Мухин, Е. М. Лоскутов*

Предложен подход к построению прогноза качественного поведения динамической системы, основанный на нелинейно-динамическом анализе слабонестационарного хаотического временного ряда. Описан метод построения прогностических моделей по наблюдаемой эволюции одной динамической переменной, реализующий данный подход применительно к прогнозу бифуркаций низкоразмерных динамических систем. Метод применён для анализа временных рядов, сгенерированных системой Рёсслера и системой уравнений, моделирующей фотохимические процессы в мезосфере. Для анализа был использован временной ряд, рассчитанный при медленном изменении управляющего параметра системы. При этом в «наблюдаемой» реализации система демонстрировала только один — хаотический — тип поведения без каких-либо бифуркаций. Метод позволил на временах, существенно превышающих длину исходной реализации, сделать для обеих систем правильный прогноз последовательности бифуркаций, указать ожидаемые моменты бифуркационных переходов и точность определения этих моментов, а также рассчитать вероятность обнаружить в интересующий момент времени предсказываемые режимы поведения системы.

### ВВЕДЕНИЕ

1. Развитию методов анализа хаотических временных рядов (ВР) посвящено за последние двадцать лет несколько сот работ (см., например, [1] и цитируемую там литературу). Цель такого анализа — реконструкция фазового пространства, в котором происходит наблюдаемая эволюция системы, и отыскание характеристик (спектра показателей Ляпунова и размерности аттрактора) этой эволюции. Фундаментальной основой данных исследований являются теоремы Такенса [2], из которых следует, что по единственной бесконечной и стационарной реализации может быть точно восстановлена топологическая структура соответствующей области фазового пространства и перечисленные количественные характеристики. В реальности наблюдаемый ВР всегда ограничен во времени, что приводит к ошибке при реконструкции аттрактора и определении его характеристик. Тем не менее созданные к сегодняшнему дню алгоритмы позволяют достаточно аккуратно восстанавливать динамические свойства по конечному стационарному<sup>1</sup> временному ряду для широкого класса систем, в том числе делать количественный («локальный») прогноз: предсказывать будущую эволюцию системы в пределах наблюдаемого хаотического аттрактора [3].

2. Нередко (например, при исследовании природных динамических систем) наблюдаемый динамический процесс не является стационарным<sup>2</sup>. Было показано [8], что теоремы Такенса могут быть обобщены на случай нестационарных ВР, т. е. нестационарность не налагает ограничений с точки зрения принципиальной возможности реконструировать наблюдаемый аттрактор<sup>3</sup>. Цель данной работы — продемонстрировать, что нестационарный ВР даёт возможность получить информацию о качественных особенностях будущего поведения динамической системы (ДС).

<sup>1</sup> С точностью до ограниченности временного ряда.

<sup>2</sup> Предложены несколько способов выявления нестационарности ВР (см., например, [4–7]). Ещё один способ выявления нестационарности является частью излагаемого ниже алгоритма.

<sup>3</sup> Или аттракторы, если за время наблюдения имели место бифуркации. Один из способов оценки характеристик наблюдаемых аттракторов описан, например, в [7].

Утверждение о нестационарности наблюдаемого ВР означает наличие у соответствующего процесса, по меньшей мере, двух сильно различающихся временных масштабов. В данной работе мы рассматрим ситуацию, когда продолжительность наблюдаемого ВР существенно превышает меньший из этих двух масштабов, будучи при этом меньше большего из них. Такая ситуация характерна для многих долгоживущих природных систем, медленные тренды параметров которых обусловлены плавными изменениями внешних условий.

В описанной ситуации естественными с физической точки зрения являются следующие две гипотезы.

*Гипотеза 1.* Наблюдаемый нестационарный временной ряд сгенерирован слабонеавтономной «быстрой» динамической подсистемой, чьи параметры медленно меняются на характерных временных масштабах эволюции её динамических переменных.

*Гипотеза 2.* Характерное время эволюции «медленной» подсистемы, взаимодействие с которой является причиной неавтономности «быстрой» подсистемы, остаётся неизменным в не слишком дальнем будущем «медленной» подсистемы. Под недалёким будущим медленной подсистемы мы подразумеваем временной интервал порядка нескольких длин наблюдаемого ВР, на котором изменения параметров быстрой подсистемы относительно невелики.

Интерпретация нестационарности ВР как неавтономности наблюдаемой ДС (гипотеза 1) имеет смысл в случае, когда зависимость параметров системы от времени известна. При этом вторая гипотеза даёт принципиальную возможность восстановить тренды параметров быстрой подсистемы на достаточно протяжённом временном интервале.

3. Описанный подход позволяет поставить задачу о прогнозе качественного поведения динамической системы по наблюдаемому слабонестационарному ВР. В данной работе излагается алгоритм, реализующий данный подход применительно к прогнозу бифуркаций низкоразмерных ДС по наблюдаемой хаотической динамике.

Предлагаемый алгоритм включает три этапа. Первый из них заключается в создании модели, характеризуемой некоторым набором параметров и воспроизводящей поведение системы на временном интервале, являющемся частью наблюдаемой реализации. На этом этапе параметры модели полагаются фиксированными.<sup>4</sup> На втором этапе оставшаяся часть реализации используется для восстановления зависимости параметров модели от времени в предположении, что эта зависимость является существенно более медленной, чем характерные временные масштабы эволюции «мгновенной» модели, построенной на первом этапе. Созданную таким образом модель будем далее называть прогностической. Наконец, на третьем этапе строится прогноз бифуркаций: производится экстраполяция зависимостей параметров прогностической модели от времени и предсказывается последовательность бифуркаций и моменты бифуркационных переходов.

4. Очевидно, что любая построенная модель воспроизведёт наблюдаемую эволюцию системы с конечной точностью. Соответственно, с конечной точностью будет воспроизведён хаотический аттрактор и его характеристики. Нетрудно понять, что при предсказании будущего поведения ДС по нестационарному ВР такая погрешность чревата прогностической ошибкой. Действительно, погрешность в этом случае является в том числе и погрешностью восстановления неавтономности наблюдаемой ДС. Другими словами, возникает ошибка при реконструкции трендов параметров ДС, а следовательно, может оказаться ошибочным и прогноз бифуркаций. Сказанное означает, что оценка точности производимых операций является обязательным компонентом алгоритма такого прогноза. Ясно также, что сам прогноз может быть только вероятностным: речь должна идти о наиболее вероятных будущих бифуркациях, математических ожиданиях моментов бифуркаций, вероятности обнаружить в определённый момент времени определённый тип поведения системы и т. д.

<sup>4</sup> Модель, воспроизводящую с заданной точностью наблюдаемое поведение ДС, иногда называют глобальной моделью [1], а процесс построения такой модели — глобальной реконструкцией ДС по одномерной реализации [9].

5. В данной работе возможности предлагаемого метода демонстрируются с помощью компьютерных моделей: источником исходных данных является, во-первых, хорошо известная и изученная система Рёссlera [10] и, во-вторых, более сложная система пяти обыкновенных дифференциальных уравнений, моделирующая фотохимические процессы в мезосфере Земли [11, 12]. В качестве наблюдаемого ВР мы использовали зависимость от времени одной динамической переменной, рассчитанную при медленном изменении управляющего параметра системы. При этом продолжительность «наблюаемой» реализации ограничивалась таким образом, чтобы на всём её протяжении система демонстрировала только хаотический тип поведения. Несмотря на такую минимальную исходную информацию о возможных типах поведения «наблюаемой» системы, предлагаемый алгоритм позволил правильно предсказать последовательность бифуркаций и определить другие перечисленные выше в подразделе 4 характеристики прогноза на временах, существенно превышающих длину исходной реализации.

Первый раздел работы посвящён изложению общей методики построения прогностических моделей и основанного на них прогноза бифуркаций. Во втором разделе подробно рассмотрено применение предлагаемого алгоритма в случае, когда в качестве источника данных используется система Рёссlera. В третьем разделе продемонстрированы возможности метода в применении к мезосферной фотохимической системе (МФХС). В заключении работы сформулированы полученные результаты и обсуждены возможности развития изложенного подхода.

## 1. МЕТОДИКА МОДЕЛИРОВАНИЯ И ПОСТРОЕНИЯ ПРОГНОЗА

1.1. Предлагаемый алгоритм может быть представлен в виде следующей пошаговой процедуры:

1) производится реконструкция фазового пространства и выбирается секущая Пуанкаре;

2) по части исходной реализации строится «мгновенная» модель в виде записанной аналитически функции последования;

3) анализируется весь исходный временной ряд и восстанавливается неавтономность наблюдаемой системы (создаётся прогностическая модель);

4) отыскивается ковариационная матрица параметров прогностической модели и производится экстраполяция трендов параметров и их ковариационной матрицы в будущее;

5) строится прогноз бифуркаций: предсказывается последовательность бифуркаций, рассчитываются математические ожидания моментов бифуркационных переходов, определяются зависимости от времени вероятностей реализации предсказанных режимов поведения системы.

Операции, перечисленные в пунктах 1) и 2), обсуждались во многих работах (см., например, [1, 9] и цитируемую там литературу), поэтому ниже, в подразделе 1.2, мы ограничимся их кратким изложением с пояснением особенностей в рамках предлагаемого алгоритма. Содержание пункта 3), и в особенности пунктов 4) и 5), является ключевым при построении прогноза бифуркаций. Ранее данные операции, насколько нам известно, не обсуждались, в связи с чем они будут описаны более подробно в подразделах 1.3–1.5.

1.2. В данной работе модель наблюдаемой ДС строится в виде функции последования. В соответствии с выбором дискретной модели вектор состояния в восстанавливаемом фазовом пространстве в момент времени  $t_k$  задаётся координатами с задержками [2]:

$$\mathbf{X}(t_k) = \{x(t_k), x(t_k + \Delta t), \dots, x(t_k + (d_E - 1) \Delta t)\}, \quad (1)$$

где  $x(t)$  — исходный временной ряд,  $d_E$  — размерность фазового пространства (минимальная размерность вложения),  $\Delta t$  — время задержки, определяемое, например, из условия первого минимума функции взаимной информации [13]. Величина  $d_E$  рассчитывалась методом «фальшивых соседей» [1, 14].

Далее в восстановленном фазовом пространстве выбиралась секущая Пуанкаре. В результате из исходной временной реализации извлекался новый ряд данных  $x_j = x(t_j)$ , где  $t_j$  — моменты пересечения фазовой траекторией секущей Пуанкаре в определённую сторону. Модель строилась в виде некоторой функции  $f$ , аппроксимирующей функцию последования<sup>5</sup>:

$$x_j \approx f(x_{j-1}, \dots, x_{j-N}; \mu), \quad \mu \in \mathbb{R}^M. \quad (2)$$

Здесь  $N$  — порядок отображения,  $\mu$  — набор из  $M$  параметров модели, которые определялись из условия минимума среднеквадратичной невязки

$$\chi^2(\mu) = \left\langle (x_j - f(x_{j-1}, \dots, x_{j-N}; \mu))^2 \right\rangle_j, \quad (3)$$

где угловые скобки означают усреднение по соответствующему индексу. В результате создавалась модель наблюдаемой ДС в виде функции последования

$$f_{\text{in}} = f(x_{j-1}, \dots, x_{j-N}; \mu_0), \quad (4)$$

где  $\mu_0$  обозначает набор параметров, минимизирующих величину невязки:

$$\chi_0^2 = \chi^2(\mu_0) = \chi_{\min}^2.$$

Особенности операций, использованных при создании модели (4), состоят в следующем.

1) Для построения модели (4) используется начальная часть длительностью  $L$  исходной реализации  $T$ . По очевидным причинам продолжительность отрезка  $L$  существенно превышает характерное время изменения наблюдаемой динамической переменной. Одновременно  $L$  удовлетворяет неравенству  $L \ll T$ , что позволяет восстанавливать неавтономность системы с требуемой точностью (см. ниже в подразделе 1.4.2).

2) Предполагается, что на протяжении временного отрезка  $L$  наблюдаемая ДС является автономной, т. е. параметры модели (4) не зависят от времени. Построенную в этом приближении модель  $f_{\text{in}}$  будем называть мгновенной.

1.3. Следующий шаг состоит в восстановлении неавтономности наблюдаемого поведения ДС.

Мы исходили из достаточно очевидного соображения: изменение параметров ДС должно вызывать тренды параметров модели (4). Другими словами, неавтономность ДС должна приводить к зависимости параметров  $\mu_0$  мгновенной модели от аппроксимируемого участка исходной реализации.

Для восстановления неавтономности разобьём исходный временной ряд  $T$  на неперекрывающиеся отрезки длины  $L$ . В результате получим  $S = T/L$  отрезков ряда, на каждом из которых вычислим параметры мгновенной модели (4), руководствуясь описанной в предыдущем разделе процедурой. Будем предполагать, что полученные значения  $\mu_s = \mu_0(t_s)$  соответствуют параметрам в срединные моменты  $t_s$ , где  $s = 1, 2, \dots, S$ , отрезков исходного временного ряда, по которым они были вычислены. Такая техника позволяет построить зависимости каждого из параметров модели от «медленного» времени  $t_s$ , дискретизованного с шагом  $L$ , отражающие, в силу этого, изменение внешних условий с характерными временами больше или порядка  $L$ .<sup>6</sup>

Гипотеза 1, сформулированная в подразделе 2 введения, означает, что набор  $\{\mu_s, t_s\}$ , где  $s = 1, 2, \dots, S$ , может быть аппроксимирован полиномиальной зависимостью  $\mu_{0r}(t_s)$  невысокой степени. Модель (4) с полученной в результате аппроксимации зависимостью параметров от медленного

<sup>5</sup> Излагаемая в подразделах 1.2, 1.3 процедура построения прогностической модели наблюдаемой ДС не затрагивает вопроса о выборе вида функции  $f$ . Некоторые соображения по этому поводу приведены в разделе 4.

<sup>6</sup> Подобная техника использовалась в [15] для извлечения из временного ряда информации о зависимости от времени известных параметров известной ДС.

времени

$$\mu_{0r}^i(t_s) = \sum_{\gamma=0}^{\Gamma_i} \alpha_{\gamma}^i t_s^{\gamma} \quad (5)$$

будем называть прогностической. В (5)  $\Gamma_i$  — степень полинома, аппроксимирующего тренд  $i$ -го параметра,  $\alpha_{\gamma}^i$  — постоянные коэффициенты. Вторая гипотеза из подраздела 2 введения даёт возможность прогнозировать качественное поведение наблюдаемой ДС за пределами исходной реализации  $T$  путём экстраполяции зависимости  $\mu_{0r}(t_s)$  в будущее<sup>7</sup> и исследования поведения прогностической модели.

**1.4.** Как уже было сказано, важной составной частью алгоритма является оценка точности производимых операций. Выше говорилось о четырёх таких операциях:

1) о реконструкции фазового пространства и получении информации о пересечении фазовой траекторией выбранной секущей Пуанкаре;

2) о построении мгновенной модели (4), воспроизводящей эволюцию ДС на каждом из  $S$  отрезков длины  $L$  наблюдаемого ВР;

3) о полиномиальной аппроксимации  $\mu_{0r}(t_s)$  (см. (5)) наборов параметров мгновенной модели  $\mu_s$ , где  $s = 1, 2, \dots, S$ , в последовательные  $S$  моментов «медленного» времени;

4) об экстраполяции зависимости  $\mu_{0r}(t_s)$  в будущее.

Обсудим происхождение и оценим ошибки, возникающие при выполнении каждой из перечисленных операций.

**1.4.1.** Первая из ошибок определяется длиной  $T$  исходного ВР (пропорциональна  $1/T$ ) и его дискретизацией (увеличивается с ростом шага дискретизации).

Ошибка при построении мгновенной модели суть погрешность аппроксимации «наблюдаемого» точечного отображения аналитической функцией. Эта ошибка для каждого из  $S$  отрезков, составляющих ВР, характеризуется соответствующей минимальной невязкой  $\chi_0^2(t_s)$ , где  $s = 1, 2, \dots, S$ , и в рамках изложенной процедуры возникает по двум причинам. Первая из них — конечная точность «кусочно-автономной» аппроксимации данных, сгенерированных неавтономной ДС. Поскольку внешние условия успевают немного измениться за время порядка длительности  $L$  отрезка реализации, по которому строится мгновенная модель, ошибка аппроксимации всегда отлична от нуля. Вторая причина — неполное соответствие модели наблюдаемой динамической системе — приводит к конечной ошибке аппроксимации даже в отсутствие неавтономности.

Все перечисленные ошибки приводят к случайной погрешности  $\xi_j$  при построении мгновенной модели:

$$x_j = f_{in} + \xi_j. \quad (6)$$

Будем интерпретировать эту погрешность как следствие зашумлённости исходного ВР.

Предположим теперь, что мгновенная модель (4) абсолютно адекватна наблюдаемой ДС, т. е. для каждого отрезка  $L$  существует такой набор параметров  $\mu^*(t_s)$ , при котором модель идеально воспроизводит динамическую компоненту ВР. Точность вычисления параметров мгновенной модели в каждый момент медленного времени  $t_s$  определяется ковариационной матрицей параметров

$$C_{\mu}^{ij}(t_s) = \left\langle (\mu_0^i(t_s) - \mu^{*i}(t_s)) (\mu_0^j(t_s) - \mu^{*j}(t_s)) \right\rangle. \quad (7)$$

Здесь  $\mu_0(t_s)$  — набор параметров, при которых среднеквадратичная невязка минимальна, верхние индексыnumеруют компоненты вектора параметров, а угловые скобки означают усреднение по статистическому ансамблю. Последняя операция предполагает знание соответствующего распределения вероятностей, отыскание которого на основе имеющейся в нашем распоряжении единственной временной

<sup>7</sup> Очевидно, что «будущее» здесь является условным, т. к. предлагаемая процедура позволяет проводить экстраполяцию как по, так и против течения медленного времени.

реализации невозможно. Тем не менее достаточно общие представления о природе шумовой компоненты позволяют найти, опираясь на теоремы Байеса [16], соответствующее распределение вероятностей параметров модели.

В данной работе мы рассматриваем случай, когда стохастическая компонента  $\xi_j$  в (6) является нормально распределенным белым шумом с нулевым средним. Именно таким предположениям адекватен поиск параметров методом наименьших квадратов, использованный нами при построении модели (4). Если, кроме того, зависимость функции  $f$  от параметров линейна, то можно показать, что параметры  $\mu$  модели в каждый момент времени  $t_s$  являются нормально распределенными случайными величинами со средними  $\mu^*(t_s)$  и ковариациями  $C_\mu(t_s)$ , определяемыми следующим соотношением:

$$C_\mu = \frac{\langle \xi^2 \rangle}{N_L} U^{-1}, \quad U_{ij} = \left\langle \frac{\partial f(x_{k-1}, \dots, x_{k-N}; \mu)}{\partial \mu_i} \frac{\partial f(x_{k-1}, \dots, x_{k-N}; \mu)}{\partial \mu_j} \right\rangle \Big|_{\mu=\mu^*}. \quad (8)$$

В (8)  $\langle \xi^2 \rangle$  — дисперсия шума, усреднение в выражении для  $U_{ij}$  производится по отсчетам, принадлежащим отрезку длины  $L$ , центрированному относительно соответствующего момента медленного времени  $t_s$ ,  $N_L$  — число таких отсчетов.

В данной работе в качестве оценки  $\mu^*(t_s)$  мы будем использовать набор параметров  $\mu_0(t_s)$ , минимизирующий невязку (3).

Очевидно также, что в случае малых шумов сделанные предположения приводят к нормальному распределению параметров при произвольной зависимости мгновенной модели (4) от  $\mu_0$ .

**1.4.2.** Ошибка аппроксимации изменения  $i$ -го параметра мгновенной модели внутри наблюдаемого ВР (при  $0 < t_s < T$ ) определяется степенью  $\Gamma_i$  аппроксимирующего полинома  $\mu_{0r}^i(t_s)$  (см. (5)). В соответствии с предыдущим разделом ограничим эту степень наименьшим из значений, для которых минимальные невязки

$$[(M^i)^2]_{\min} = [\langle (\mu_{0r}^i(t_s) - \mu_0^i(t_s))^2 \rangle_s]_{\min},$$

где  $i = 1, 2, \dots, M$ , аппроксимации  $\mu_{0r}^i(t_s)$  меньше соответствующих элементов ковариационной матрицы  $C_\mu(t_s)$ , усредненных по отсчетам медленного времени внутри интервала  $T$ :

$$[(M^i)^2]_{\min} \leq \langle C_\mu^{ii} \rangle_s. \quad (9)$$

Здесь  $s = 1, 2, \dots, S$ , усреднение проводится по отсчетам медленного времени, верхние индексы обозначают компоненты вектора параметров. Условие (9) означает, что ошибка восстановления тренда параметров прогностической модели не превосходит по порядку величины ошибку определения самих величин  $\mu_s$ .

Мы можем теперь сказать, что прогностическая модель (4), (5) воспроизводит нестационарную эволюцию ДС внутри наблюдаемого ВР с погрешностью, возникающей вследствие случайной ошибки при отыскании параметров модели. При этом «быстрая» эволюция наблюдаемой ДС задается положением точки в пространстве параметров прогностической модели в фиксированный момент медленного времени  $t_s$ , а слабая неавтономность ДС отражается в «медленном» перемещении указанной точки, описываемом зависимостью (5) вектора параметров  $\mu_{0r}$  от  $t_s$ .

Погрешность в определении положения точки в пространстве параметров определяется, очевидно, распределением вероятностей значений параметров прогностической модели. В данной работе мы рассмотрим ситуацию, описанную в подразделе 1.4.1, и будем характеризовать случайный разброс параметров прогностической модели нормальным распределением, вектор средних  $\mu_{0r}(t_s)$  и ковариационная матрица  $C_\mu(t_s)$  которого изменяются с течением времени вследствие неавтономности системы.

**1.4.3.** Таким образом, для анализа будущего поведения наблюдаемой ДС необходимо экстраполировать за пределы наблюдаемого ВР восстановленные тренды параметров  $\mu_{0r}(t_s)$  и исследовать поведение прогностической модели при будущих значениях параметров. Поскольку в рамках излагаемого

алгоритма ошибки в реконструкции поведения наблюдаемой ДС при каждом фиксированном наборе параметров (т. е. в каждый момент медленного времени) целиком определяется погрешностью вычисления этих параметров, требуется оценить эту погрешность за пределами наблюдаемого ВР. Для этого следует экстраполировать в будущее, наряду со средним трендом  $\mu_{0r}(t_s)$ , ковариационную матрицу параметров  $C_\mu(t_s)$ . Продемонстрируем используемый метод экстраполяции на примере простейшего тренда параметров.

Пусть условию (9) удовлетворяет линейная регрессия набора  $\{\mu_s, t_s\}$ , где  $s = 1, 2, \dots, S$ :

$$\mu_{0r}(\tau) = \mathbf{P} + \mathbf{Q}\tau. \quad (10)$$

Параметры тренда (коэффициенты полинома (10)) ищем методом наименьших квадратов, что даёт

$$\mathbf{P} = \langle \mu \rangle_s = \frac{1}{S} \sum_{s=1}^S \mu_s, \quad (11)$$

$$\mathbf{Q} = \frac{\langle \mu \tau \rangle_s}{\langle \tau^2 \rangle_s}, \quad \langle \tau^2 \rangle_s = \frac{1}{S} \sum_{s=1}^S \tau_s^2 = \frac{12}{T(T+L)}, \quad \langle \mu \tau \rangle_s = \frac{1}{S} \sum_{s=1}^S \mu_s \tau_s. \quad (12)$$

Здесь введено новое медленное время  $\tau_s = t_s - T/2$ , симметризующее «будущее» и «прошлое», т. е. отсчитываемое от середины исходной реализации  $T$ ;  $\tau_s = -T/2 + L(2s-1)/2$  — срединные моменты отрезков  $L$ , отсчитываемые по симметризованной временной шкале.<sup>8</sup>

Чтобы оценить ошибку в определении параметров  $\mathbf{P}$  и  $\mathbf{Q}$  тренда (9) предположим, что реальный линейный тренд зашумлён стационарным дельта-коррелированным процессом  $\nu_s$  с нулевым средним:

$$\mu_s = \mathbf{p} + \mathbf{q}\tau_s + \nu_s, \quad (13)$$

так что

$$\langle \nu_s \rangle_\nu = 0, \quad C_\nu^{ij} = \langle \nu_s^i \nu_l^j \rangle_\nu = \langle C_\mu^{ij} \rangle_s \delta_{il}, \quad (14)$$

где  $\delta_{il}$  — символ Кронекера. Из (10)–(14) нетрудно получить выражения для ковариационных матриц  $C_Q^{ij}$  и  $C_P^{ij}$  параметров тренда (10), а также для их взаимной ковариации  $C_{PQ}^{ij}$ :

$$C_Q^{ij} = \langle (Q^i - q^i)(Q^j - q^j) \rangle_\nu = \frac{\langle C_\mu^{ij} \rangle_s}{S \langle \tau^2 \rangle_s}, \quad C_P^{ij} = \langle C_\mu^{ij} \rangle_s / S, \quad C_{PQ}^{ij} = 0. \quad (15)$$

Таким образом, ковариационная матрица  $C_{\mu_{0r}}^{ij}$  для экстраполированных параметров  $\mu_{0r}(\tau)$  будет иметь вид

$$C_{\mu_{0r}}^{ij} = \frac{\langle C_\mu^{ij} \rangle_s}{S} \left( 1 + \frac{\tau^2}{\langle \tau^2 \rangle_s} \right) = \frac{\langle C_\mu^{ij} \rangle_s}{S} \left( 1 + \frac{12\tau^2}{T(T+L)} \right) \approx \frac{\langle C_\mu^{ij} \rangle_s}{S} \left( 1 + \frac{12\tau^2}{T^2} \right). \quad (16)$$

В (16) мы ограничились (в силу очевидного условия  $T \gg L$ ) старшими членами разложения по параметру  $L/T$ .<sup>9</sup>

Очевидно, что изложенный подход позволяет экстраполировать зависимость  $C_\mu(t_s)$  за пределы исходной реализации  $T$  для любой степени полинома, аппроксимирующего зависимость  $\mu_0(t_s)$ .

1.5 Последнее, что необходимо сделать — построить прогноз бифуркаций и оценить его точность. В нашем распоряжении имеется прогностическая модель: функция последования (4) с восстановленной зависимостью её параметров  $\mu_0(t_s)$  от времени. Данная зависимость определяется, во-первых,

<sup>8</sup> Это обеспечивает, в частности, выполнение соотношения  $\langle \tau^n \rangle_s = S^{-1} \sum_{s=1}^S \tau_s^n = 0$  для всех нечётных  $n > 0$ .

<sup>9</sup> Отметим, что с такой точностью  $\langle \tau^n \rangle_s \approx T^{-1} \int_{-T/2}^{T/2} \tau^n d\tau$ , что удобно использовать при аппроксимации трендов полиномами более высокой степени.

средним трендом  $\mu_{0r}(t_s) = \{\mu_{0r}^i(t_s)\}$ , где  $i = 1, 2, \dots, M$ , в пространстве параметров модели (задаваемым, например, соотношениями (10)–(12)). Во-вторых, она характеризуется случайным разбросом величин этих параметров, описываемым соответствующим распределением вероятностей. Последнее (см. подраздел 1.4.2) определяется всё тем же средним трендом  $\mu_{0r}(t_s)$  и ковариационной матрицей  $C_\mu(t_s)$  (имеющей, например, вид (16)). Знание будущих зависимостей  $\mu_{0r}(t_s)$  и  $C_\mu(t_s)$  позволяет реконструировать бифуркационную диаграмму (БД), используя в качестве единственного управляющего параметра медленное время  $t_s$ . При этом, как пояснено в подразделах 4 введения и 1.4.3, случайный разброс параметров приводит к необходимости рассматривать бифуркации, последовательность которых прогнозируется в результате построения БД, как случайные события, а моменты бифуркационных переходов — как случайные величины. Поэтому прогноз бифуркаций должен дать ответы на следующие вопросы:

- 1) С какой вероятностью предсказываемые бифуркации произойдут до определённого момента?
- 2) Чему равно математическое ожидание момента времени конкретного бифуркационного перехода?
- 3) Какова точность определения среднего момента бифуркации?
- 4) Какова вероятность обнаружить в заданный момент времени интересующий тип поведения системы?

Чтобы ответить на поставленные вопросы, воспользуемся методом Монте-Карло. Рассмотрим набор из  $\Pi$  лучей (дополнительных трендов), построенных в пространстве параметров модели таким образом, чтобы задаваемые ими в произвольный момент времени  $t_s$  наборы параметров описывались распределением вероятностей, характеризуемым найденными зависимостями  $\mu_{0r}(t_s)$  и  $C_\mu(t_s)$ . Построив БД, отвечающую каждому дополнительному тренду, мы получим ансамбль моментов бифуркаций для каждого предсказанного бифуркационного перехода. Разобьём всю шкалу времени на достаточно малые интервалы  $\delta t_i$  и будем регистрировать количество попаданий  $\nu_b^i$  в каждый из них элементов каждого полученного ансамбля. Зависимость от времени плотности вероятности момента  $b$ -й бифуркации определим как

$$\varphi_b(t_s) = \frac{\nu_b^i}{\delta t_i \Pi}. \quad (17a)$$

Здесь  $b = 1, 2, \dots, B$ ,  $B$  — полное число прогнозируемых бифуркаций.

Знание  $\varphi_b(t_s)$  позволяет сразу ответить на три первых из четырёх сформулированных вопросов, определив стандартным образом зависимость интегральной вероятности  $\Phi_b$  момента соответствующей бифуркации от медленного времени:

$$\Phi_b(t_s^*) = \int_T^{t_s^*} \varphi_b(t_s) dt_s, \quad t_s^* > T, \quad (17b)$$

а также рассчитав математическое ожидание момента бифуркации и соответствующую дисперсию. Кроме того, анализ всего набора функций  $\varphi_b(t_s)$  даёт возможность определить вероятность  $P_k$  обнаружить в заданный момент времени  $t_s^*$   $k$ -й тип поведения системы:

$$P_k(t_s^*) = P_k^+(t_s^*) [1 - P_k^-(t_s^*)]. \quad (18)$$

Здесь

$$P_k^\pm(t_s^*) = \int_T^{t_s^*} \varphi_k^\pm(t_s) dt_s, \quad t_s^* > T, \quad (19)$$

а функции  $\varphi_k^+(t_s)$  и  $\varphi_k^-(t_s)$  есть суммы плотностей вероятностей всех возможных бифуркаций входа в  $k$ -й режим и выхода из него соответственно. В (17б) и (19)  $T$  обозначает границу исходного временного ряда.

## 2. ПОСТРОЕНИЕ ПРОГНОЗА БИФУРКАЦИЙ ДЛЯ СИСТЕМЫ РЁССЛЕРА

**2.1.** Для демонстрации предлагаемого метода мы использовали в качестве источника данных реализацию, сгенерированную хорошо известной системой Рёсслера [10], которая представляет собой систему трёх обыкновенных дифференциальных уравнений первого порядка:

$$\begin{aligned} \frac{dx}{dt} &= -y - z, \\ \frac{dy}{dt} &= x + ay, \\ \frac{dz}{dt} &= b - rz + xz. \end{aligned} \quad (20)$$

Параметры  $a$  и  $b$  фиксировались и полагались равными 0,2. Параметр  $r$  был выбран в качестве управляющего: он медленно (по сравнению с характерными временами эволюции системы) менялся со временем по линейному закону. На рис. 1 показана бифуркационная диаграмма (БД) системы (20). Ордината каждой точки на данной БД определяется локальным максимальным значением переменной  $z$ ,

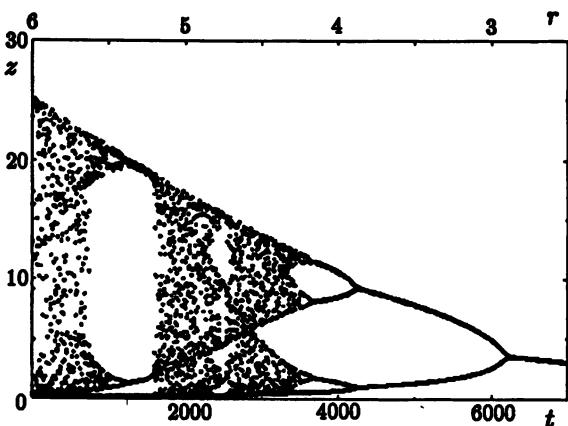


Рис. 1. Бифуркационная диаграмма системы Рёсслера. Переменная  $z$ , управляемый параметр  $r$  и время в относительных единицах

ещё один обратный каскад удвоений система переходит в однопериодический режим. БД, приведённая на рис. 1, будет использована в дальнейшем для контроля качества прогноза бифуркаций, построенного на основе сгенерированного системой Рёсслера временного ряда.

**2.2.** Для построения прогностической модели данной системы мы использовали участок реализации из правой широкой хаотической области (см. рис. 1)  $t \in [2000, 3500]$ , представленный на рис. 2а. Подчеркнём, что в соответствии с поставленной задачей вся дальнейшая процедура основана на ана-

а абсцисса — моментом времени  $t$ , в который данное значение достигнуто. При построении БД параметр  $r$  менялся от 6 при  $t = 0$  до 2,5 при  $t = 7000$ . Бифуркационные диаграммы такого вида мы будем использовать и в дальнейшем, поскольку они позволяют проследить не только изменения качественного поведения неавтономной системы, но и зафиксировать моменты времени бифуркационных переходов. На рис. 1, например, видно, что при заданной скорости тренда параметра  $r$  система в момент  $t \approx 1000$  выходит через обратный каскад удвоений периода из хаотического режима и реализуется окно регулярного поведения, соответствующее предельному циклу с периодом 3 в точечном отображении. Далее, при  $t \approx 1850$ , через перемежаемость происходит переход к хаотической динамике, которая, прерываемая узкими окнами регулярного поведения, существует до  $t \approx 3800$ . Далее через

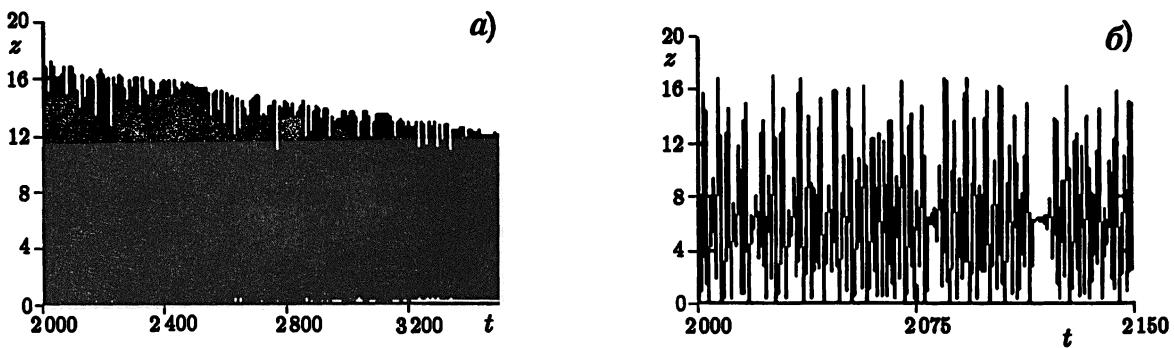


Рис. 2. Система Рёсслера: «наблюдаемый» временной ряд (а) и начальный отрезок этого ряда (б). Время  $t$  и переменная  $z$  в относительных единицах

лиze только данного участка реализации, без привлечения какой-либо дополнительной информации о свойствах сгенерировавшей его ДС.

Прежде всего по исходному участку реализации был реконструирован аттрактор в фазовом пространстве, в том числе определено, что минимальная размерность вложения равна 3. Далее исходный участок реализации был разбит на отрезки длины  $L = 150$ , и по первому из отрезков  $t \in [2000, 2150]$  (см. рис. 2б) была построена мгновенная модель. В качестве секущей Пуанкаре была выбрана поверхность  $dz/dt = 0$ . Регистрировались пересечения фазовой траекторией данной секущей со стороны  $dz/dt > 0$  (локальные максимумы переменной  $z$ ). После этого строилась функция последования первого порядка:

$$z_{k+1} = f(z_k, \mu).$$

Здесь индекс  $k$  нумерует максимумы переменной  $z$ , а  $\mu$  является вектором параметров модели. Мы аппроксимировали данное отображение, приведённое на рис. 3, функцией вида

$$f(z_k, \mu) = \mu_3 z_k e^{-\mu_1 z_k} + \mu_4 z_k e^{-\mu_2 z_k}. \quad (21)$$

Значения параметров  $\mu_{0i}$ , где  $i = 1; 2; 3; 4$ , найденные методом наименьших квадратов, равны 17,2; 0,46; 13,6 и 1,04 соответственно. При данных параметрах мгновенная модель (21) воспроизводит функцию последования, соответствующую эволюции на отрезке  $t \in [2000, 2150]$ , со среднеквадратичной невязкой  $\chi_0^2 = 0,03$ .

На следующем шаге была восстановлена неавтономность наблюдаемой системы. С этой целью «оптимальный» набор параметров функции последования (21) был найден для каждого отрезка с длительностью  $L$ . Нормированные значения параметров в зависимости от медленного времени приведены на рис. 4. Нормировочные коэффициенты выбирались так, чтобы параметры были порядка единицы. Найденные для  $T/L$  моментов медленного времени  $t_s$  наборы данных  $\mu_{0i}(t_s)$ , где  $i = 1; 2; 3; 4$ , аппроксимировались полиномиальной зависимостью  $\mu_{0r}(t_s)$ . Для всех параметров модели (21) адекватной (т. е. удовлетворяющей условию (9)) оказалась линейная аппроксимация

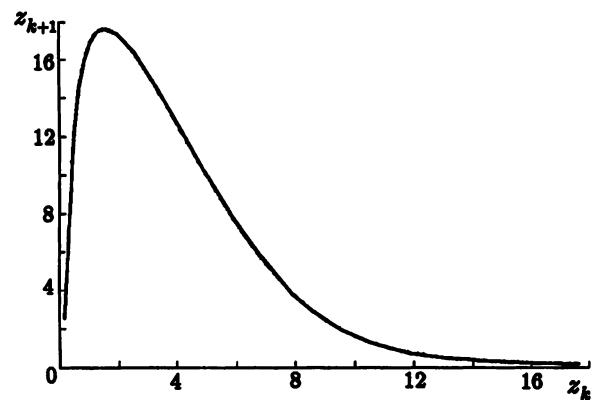


Рис. 3. Функция последования первого порядка  $z_{k+1}(z_k)$  для максимумов переменной  $z$ , построенная по начальному отрезку «наблюдаемого» временного ряда

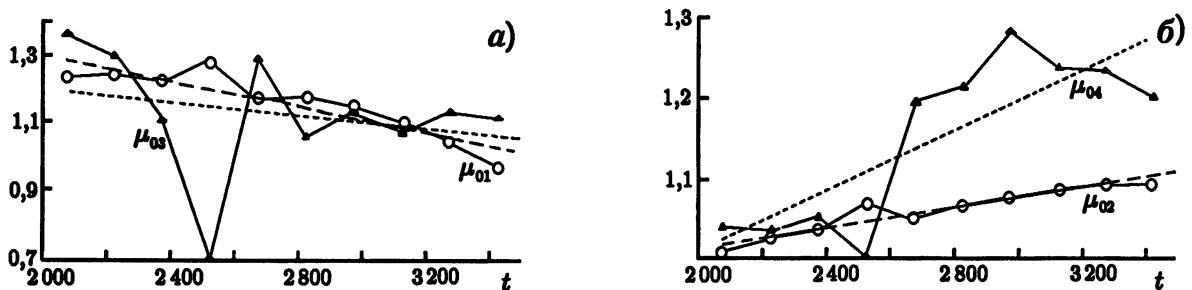


Рис. 4. Зависимость от времени параметров прогностической модели, реконструированная по «наблюдаемому» временному ряду (см. рис. 2а). Параметры  $\mu_{01}$ ,  $\mu_{02}$ ,  $\mu_{03}$  и  $\mu_{04}$  нормированы на коэффициенты 14; 0,455; 10 и 1 соответственно. Прямыми линиями показаны аппроксимации данных зависимостей. По горизонтальной оси отложено медленное время, по вертикальной — величины параметров

$$\mu_{0r}^i = \alpha_0^i + \alpha_1^i t_s, \quad (22)$$

где  $i = 1; 2; 3; 4$ . Коэффициенты  $\{\alpha_0^i; \alpha_1^i\}$ , найденные методом наименьших квадратов, равны  $\{1,68; -2 \cdot 10^{-4}\}$ ,  $\{1,4; -10^{-4}\}$ ,  $\{0,89; 6 \cdot 10^{-5}\}$ ,  $\{0,64; 1,8 \cdot 10^{-4}\}$  для  $i = 1; 2; 3; 4$  соответственно.

Функция последования (21) с восстановленной зависимостью (22) её параметров от времени является прогностической моделью динамической системы, сгенерировавшей исходную реализацию на временном интервале  $t \in [2000, 3500]$ .

Далее строился прогноз бифуркаций по среднему тренду  $\mu_{0r}(t_s)$ , для чего зависимости (22) экстраполировались на весь интервал прогноза  $t \in [0, 8000]$  исходной системы и строилась бифуркационная диаграмма (БД) прогностической модели с медленным временем в качестве управляющего параметра. На рис. 5а БД, отвечающая среднему тренду, приведена вместе с истинной БД с рис. 1. Видно, что на всём рассматриваемом временном интервале прогностическая модель верно воспроизводит последовательность бифуркаций «наблюдаемой» ДС. Отметим, что в данном случае правильно предсказаны два сценария выхода ДС из хаотического режима: через обратный каскад удвоения периода и через перемежаемость.

**2.3.** Перейдём к оценке точности прогноза. Как обсуждалось в подразделах 4 введения и 1.4.3, такая оценка подразумевает отыскание вероятностных характеристик предсказанных бифуркаций и режимов поведения системы. С этой целью для каждого отрезка длительностью  $L$  исходной реализации была рассчитана ковариационная матрица параметров модели  $C_\mu$  (см. (8)), которая затем экстраполировалась за пределы исходной временной реализации  $T$  способом, описанным в подразделе 1.4.3 (см. (16)). Далее в пространстве параметров модели был построен ансамбль из 10 000 параметрических траекторий (трендов) таким образом, чтобы в каждый момент времени компоненты вектора параметров модели являлись нормально распределёнными случайными величинами, чьи средние значения и ковариации представляют собой экстраполированные зависимости  $\mu_{0r}(t_s)$  и  $C_\mu(t_s)$ . Наконец, для каждого такого тренда была реконструирована БД и определена последовательность и моменты бифуркационных переходов на всём исследуемом временном интервале  $t \in [0, 8000]$ . Результаты статистической обработки полученных ансамблей моментов бифуркаций приведены на рис. 5б-г.

Прежде чем перейти к обсуждению полученных выше результатов, обратим внимание на следующее обстоятельство. Представление зависимостей средних значений параметров и компонент ковариационной матрицы в виде непрерывных функций медленного времени (см., например, (22)) позволила нам «забыть» о дискретности последнего при построении бифуркационных диаграмм, отвечающих как среднему, так и дополнительным трендам в пространстве параметров прогностической модели. Ясно, однако, что в рамках описываемого алгоритма имеет смысл проанализировать вероятностные характеристики бифуркаций, относящиеся к режимам, чья протяжённость во времени превышает шаг

$L = 150$  дискретизации медленного времени. На рис. 5а видно, что такими достаточно долгоживущими режимами являются трёхпериодический и хаотический режимы в прошлом и четырёх-, двух- и однопериодические режимы в будущем. В силу указанных обстоятельств на рис. 5б приведены плотности вероятности моментов бифуркаций  $\varphi_b(t_s)$ , рассчитанные по формуле (17), для «будущих» переходов: 1) из восьми- в четырёхпериодический режим ( $\varphi_{8 \rightarrow 4}(t_s)$ ); 2) из четырёх- в двухпериодический режим ( $\varphi_{4 \rightarrow 2}(t_s)$ ); 3) из двух- в однопериодический режим ( $\varphi_{2 \rightarrow 1}(t_s)$ ), а также 4) для бифуркации потери глобальной устойчивости однопериодического режима ( $\varphi_{1 \rightarrow \infty}(t_s)$ ), зарегистрированной при  $t \geq 6\,000$  для части дополнительных трендов. В «прошлом» показаны плотности вероятности моментов перехода из хаоса в трёхпериодический режим ( $\varphi_{ch \rightarrow 3}(t_s)$ ) и из трёхпериодического режима в шестипериодический ( $\varphi_{3 \rightarrow 6}(t_s)$ ).

Нетрудно понять, что рассчитанные с помощью данных плотностей вероятностей и показанные на рис. 5б математические ожидания и среднеквадратичные отклонения моментов бифуркаций определяются двумя факторами. Первый фактор — увеличение с течением времени характерной ширины функции распределения параметров, второй — форма конкретной бифуркационной поверхности в пространстве параметров модели. Например, видимое на рис. 5б монотонное убывание точности предсказания момента бифуркации при удалении события от границ исходного временного ряда является не только следствием монотонного расплывания с течением времени функции распределения параметров, но и свидетельствует о квазиподобной форме бифуркационных поверхностей, отвечающих различным бифуркациям.

Знание приведённого на рис. 5б набора плотностей вероятностей моментов бифуркаций позволяет рассчитать их интегральные вероятности, а также вероятности обнаружения в определённый момент времени определённого режима поведения системы. На рис. 5в приведены рассчитанные по формуле (17б) зависимости интегральной вероятности от времени для шести бифуркационных переходов, плотности вероятностей которых показаны на рис. 5б. Видно, что максимальные значения интегральной вероятности бифуркаций  $ch \rightarrow 3$  и  $3 \rightarrow 6$  ( $\Phi_{ch \rightarrow 3})_{max} = (\Phi_{3 \rightarrow 6})_{max} = 1$ , т. е. данные бифуркации зарегистрированы «в прошлом» (при  $t < 2\,000$ ) для всех дополнительных трендов. Максимальная интегральная вероятность «будущей» бифуркации  $8 \rightarrow 4$  немного меньше единицы:  $(\Phi_{8 \rightarrow 4})_{max} = 0,99$ . Это означает, что для незначительной части дополнительных трендов выход из восьмипериодического режима происходит иным путём (не в результате бифуркации  $8 \rightarrow 4$ ). Анализ последовательности бифуркаций для всех дополнительных трендов показал, что для небольшого их числа выход из восьмипериодического режима происходит вследствие глобальной неустойчивости модели. Аналогичная ситуация имеет место для четырёх- и двухпериодических режимов:  $(\Phi_{8 \rightarrow 4})_{max}$  немного больше  $(\Phi_{4 \rightarrow 2})_{max} \approx 0,97$ , которая в свою очередь превосходит  $(\Phi_{2 \rightarrow 1})_{max} \approx 0,86$  (достигается вне границ рис. 5в). Выход из однопериодического режима всегда связан с глобальной неустойчивостью модели:  $(\Phi_{1 \rightarrow \infty})_{max} = 1$  (также вне границ рис. 5в).

На рис. 5г построены зависимости от времени вероятностей обнаружить у системы различные режимы поведения, в том числе глобальную неустойчивость. Видно, во-первых, что отличная от нуля вероятность  $P_\infty$  глобальной неустойчивости действительно существует во времени со всеми будущими режимами поведения. Во-вторых, обратим внимание на следующее примечательное обстоятельство: соотношения между максимальными значениями вероятностей обнаружения конкретного режима поведения системы могут отличаться от соотношений между максимальными интегральными вероятностями моментов бифуркаций входа в соответствующие режимы. Действительно, несмотря на равную единице вероятность перехода  $ch \rightarrow 3$ , максимальная вероятность обнаружить у системы трёхпериодический режим ( $P_3)_{max} \approx 0,84$  меньше максимальной вероятности пребывания системы в двухпериодическом режиме ( $P_2)_{max} \approx 0,9$ , максимальная вероятность перехода в который меньше единицы. Объяснение данного «противоречия» почти тривиально: максимальная вероятность обнаружения некоторого режима поведения зависит не только от максимальной вероятности входа в этот режим, но

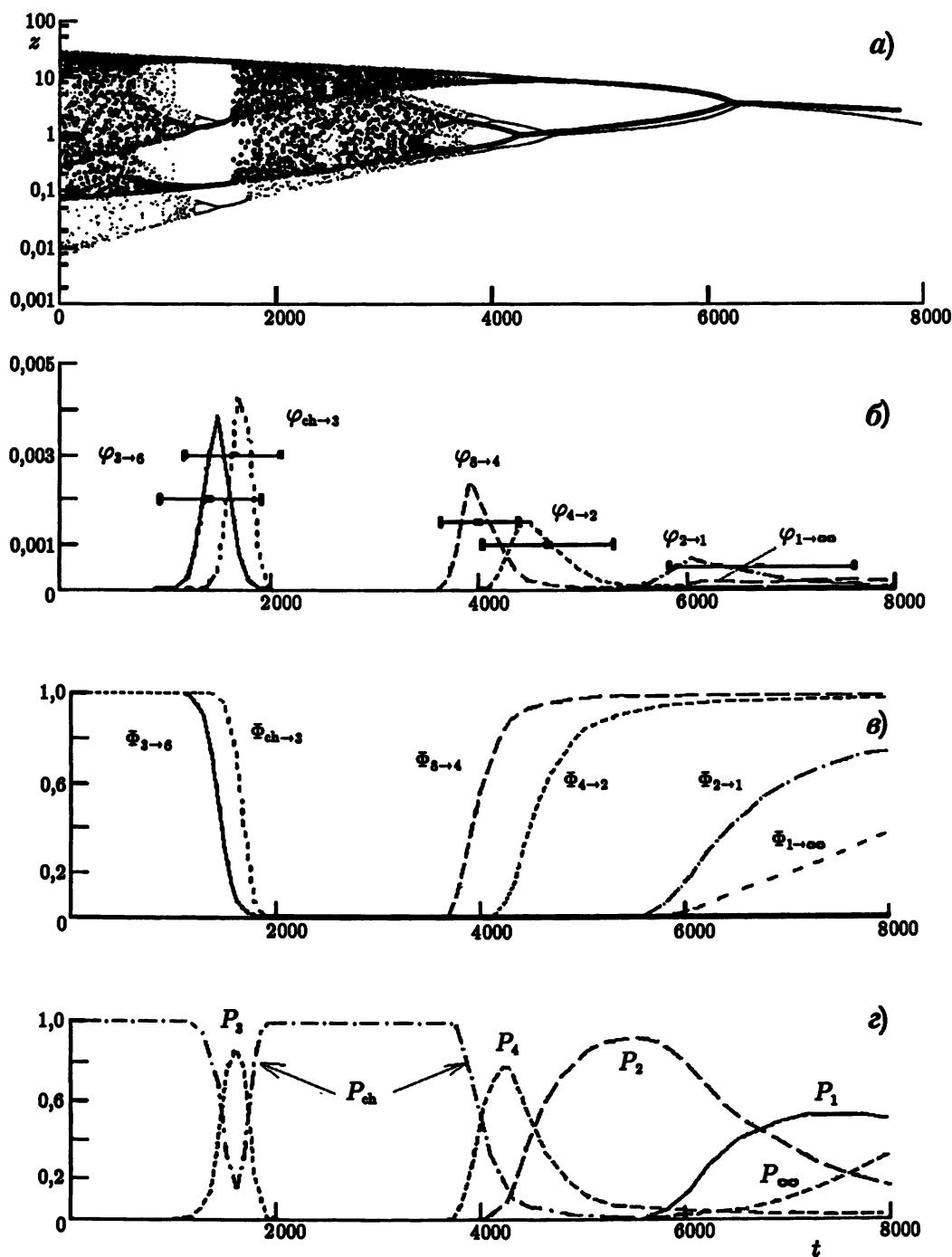


Рис. 5. Результаты прогноза: БД, соответствующая среднему тренду (серые точки) и исходная БД с рис. 1 (чёрные точки); переменная  $z$  приведена в логарифмическом масштабе (а); зависимости от времени плотностей вероятности моментов бифуркаций с математическими ожиданиями и среднеквадратичными отклонениями (б); зависимости от времени интегральных вероятностей моментов бифуркаций (в); зависимости от времени вероятностей обнаружения различных режимов (г)

и от степени перекрытия во времени плотностей вероятностей бифуркаций входа в режим и выхода из него. На рис. 5б видно, что степень перекрытия этих плотностей вероятностей для трёхпериодического режима велика, в то время как у двухпериодического режима плотности вероятностей бифуркаций вхо-

да в режим и выхода из него почти не перекрываются. По этой же причине максимальная вероятность обнаружить четырёхпериодический режим  $(P_4)_{\max} \approx 0,77 < (P_2)_{\max}$ .

**2.4.** В заключение данного раздела отметим, что приведённые на рис. 5б–г вероятностные характеристики могут быть использованы для оценки качества прогноза бифуркаций и определения его горизонта во времени. В зависимости от конкретной задачи в качестве условия достижения горизонта прогноза может быть выбрано определённое значение либо относительной точности прогноза момента бифуркации, либо максимальной интегральной вероятности, либо максимальной вероятности обнаружения некоторого режима поведения. Величины перечисленных характеристик предсказываемых бифуркаций и режимов поведения являются естественной основой и для оценки качества прогноза, в том числе для сравнения различных прогностических моделей<sup>10</sup> наблюдаемой ДС, построенных по одному и тому же ВР. Более подробно данные вопросы будут рассмотрены в отдельной работе.

### 3. ПРОГНОЗ БИФУРКАЦИЙ МЕЗОСФЕРНОЙ ФОТОХИМИЧЕСКОЙ СИСТЕМЫ

**3.1.** В данном разделе в качестве источника данных мы используем систему, описывающую поведение малых химических составляющих атмосферы Земли в районе мезопаузы (на высотах  $70 \div 90$  км) [11, 12, 17]. Эта система включает в себя пять нелинейных дифференциальных уравнений 1-го порядка с периодическим внешним воздействием:

$$\begin{aligned} dx_1/dt &= -(a_9 + 2a_{11}x_1 + a_{10}x_3 + a_4x_4 + a_5x_5)x_1 + a_1x_2x_5 + a_{15}x_4^2 + a_{16}s(t)x_3 + 2a_8s(t), \\ dx_2/dt &= -(a_6 + a_{12}x_3 + (a_1 + a_2 + a_{14})x_5)x_2 + a_4x_1x_4 + a_7s(t)r, \\ dx_3/dt &= -(a_{10}x_1 + a_{12}x_2 + a_{13}x_4 + a_{16}s(t))x_3 + a_9x_1, \\ dx_4/dt &= -(a_4x_1 + 2a_{15}x_4 + a_3x_5 + a_{13}x_3)x_4 + a_5x_1x_5 + a_{12}x_2x_3 + 2a_{14}x_2x_5 + a_7s(t)r, \\ dx_5/dt &= -(a_5x_1 + a_3x_4 + (a_1 + a_2 + a_{14})x_2)x_5 + a_6x_2 + a_{13}x_3x_4. \end{aligned} \quad (23)$$

Данные уравнения являются уравнениями химической кинетики и описывают 16 химических реакций (в том числе реакции фотолиза), протекающих в мезосфере с участием пяти химических компонент ( $\text{O}, \text{H}, \text{O}_3, \text{HO}, \text{HO}_2$ ), чьи концентрации являются динамическими переменными и обозначены  $x_1, \dots, x_5$  соответственно. Динамика этих переменных определяется коэффициентами  $a_1, \dots, a_{16}$  (константами реакций) и имеет в условиях мезосферы характерный временной масштаб одни сутки или менее. Концентрации остальных реагентов — участников реакций — существенно превышают концентрации перечисленных пяти компонент и считаются постоянными. Периодическое внешнее воздействие обусловлено суточными вариациями освещённости, связанными с восходами и заходами солнца и приводящими к периодической модуляции скорости фотодиссоциации молекул озона (концентрация которого является одной из динамических переменных системы (23)), кислорода, концентрация которого считается неизменной, и паров воды, чья относительная концентрация  $r$  является одним из управляющих параметров системы.<sup>11</sup> Такая модуляция учтена в системе (23) с помощью периодической функции  $s(t)$ , которая достигает максимума  $s_{\max} = 1$  в середине дня, а минимума  $s_{\min} = 0$  в середине ночи.<sup>12</sup> Вследствие указанной неавтономности размерность фазового пространства системы (23) равна шести.

<sup>10</sup> В рамках излагаемого алгоритма другие прогностические модели могут быть построены путём выбора иных видов функции последования (2).

<sup>11</sup> Всюду ниже  $r$  приводится в традиционных для фотохимии атмосферы единицах ppmv (particles per million in volume), определяемых как число молекул водяного пара на миллион молекул воздуха.

<sup>12</sup> Подробное обсуждение системы (23) (значения коэффициентов  $a_1, a_2, \dots, a_{16}$ , диапазон изменения контрольного параметра  $r$ , вид функции  $s(t)$ ) и её нелинейно-динамических свойств см. в работах [17, 18].

Динамической переменной, временной ряд которой «наблюдался», являлась переменная  $x_1$  (концентрация атомарного кислорода), хотя данный выбор принципиального значения не имеет. На рис. 6а показана бифуркационная диаграмма, по вертикальной оси которой отложена концентрация атомарного кислорода в момент восхода солнца, а по горизонтальной — относительная концентрация  $r$  водяного пара, которая, как отмечалось, является управляющим параметром системы. Как можно видеть, система обладает широким набором динамических режимов, реализация которых зависит от управляющего параметра. Соответствие какому-либо  $r$  нескольких значений  $x_1$  означает реализацию режима с периодом в соответствующее число суток. «Замазанные» области на бифуркационной диаграмме отвечают интервалам параметра, в которых система демонстрирует хаотическую динамику.

Видно, что в рассматриваемом диапазоне изменения управляющего параметра имеется три области хаотического поведения системы. Особый интерес представляют две из них, реализующиеся при относительно малых значениях  $r \in [1,5; 2,1]$ . На рис. 6б соответствующий фрагмент БД приведён в увеличенном по оси абсцисс масштабе. Отметим, что две указанные области хаотического поведения частично перекрываются. Эта особенность является следствием бистабильного поведения системы, сопровождающегося гистерезисом при изменении управляющегося параметра. При изменении параметра слева направо в пределах рис. 6б имеет место последовательность бифуркаций. Сначала движение с периодом в одни сутки теряет (при  $r \approx 1,65$ ) устойчивость и происходит удвоение периода (рождается устойчивый режим с двухсуточным периодом). Далее следует каскад таких удвоений, и при  $r \approx 1,8$  динамика хаотизуется. Будем называть получившееся в результате хаотическое множество внутренним аттрактором. При  $r \approx 1,87$  данное движение исчезает, и система скачком переходит в иной, совершенно отличный режим поведения.

При обратном изменении параметра мы стартуем при  $r = 2,1$  с устойчивого движения, имеющего период 3 суток, который при  $r \approx 2$  удваивается, при  $r \approx 1,9$  утверждается и через каскад удвоения периода переходит в хаотический режим. Полученное хаотическое множество будем называть внешним аттрактором. Данный аттрактор, как видно из рисунка, состоит из трёх областей, и с периодом внешнего воздействия (одни сутки) изображающая точка последовательно перескакивает из одной области в другую. При дальнейшем уменьшении управляющего параметра, при  $r \approx 1,7$ , начинается обратный каскад удвоений, при  $r \approx 1,63$  снова реализуется режим с трёхсуточным периодом, который при  $r \approx 1,58$  исчезает, и система переходит в устойчивый режим с периодом одни сутки. Таким образом, как при увеличении, так и при уменьшении параметра  $r$  в системе реализуются катастрофические бифуркации.

**3.2.** Мы использовали хаотическую эволюцию мезосферной фотохимической системы (МФХС) (23), «наблюдаемую» в пределах внешнего и внутреннего аттракторов, для построения соответствующих прогностических моделей данной динамической системы. Прогностическая модель, основанная на реализации, сгенерированной внешним аттрактором<sup>13</sup>, позволила с хорошей точностью предсказать последовательности всех «будущих» и «прошлых» бифуркаций и режимов, изображённых на рис. 6б. Ситуация с прогнозом бифуркаций внутреннего аттрактора оказалась более сложной. Поэтому ниже мы опишем построение прогностической модели и прогноза бифуркаций на основе реализации, порождённой внутренним аттрактором.

На основе системы (23) был сгенерирован временной ряд переменной  $x_1$  продолжительностью  $T = 2000$  суток (см. рис. 7а), в пределах которого параметр  $r$  убывал (линейно во времени) от 1,855 до 1,81. К реализации был применён алгоритм, изложенный в разделе 1.

<sup>13</sup> Как отмечено выше, внешний аттрактор является трёхсвязным (см. рис. 6б), поэтому на основе единственной 1500 суточной реализации были построены три прогностические модели. Каждая из них воспроизводила слабонеавтономное поведение МФХС в пределах соответствующей области внешнего аттрактора и представляла собой функцию последования, связывающую значения переменной  $x_1$  в конце ночи, разделённые интервалом в трое суток, и отличающуюся от двух других функций последования фазой переменной  $x_1$  внутри трёхсуточного интервала.

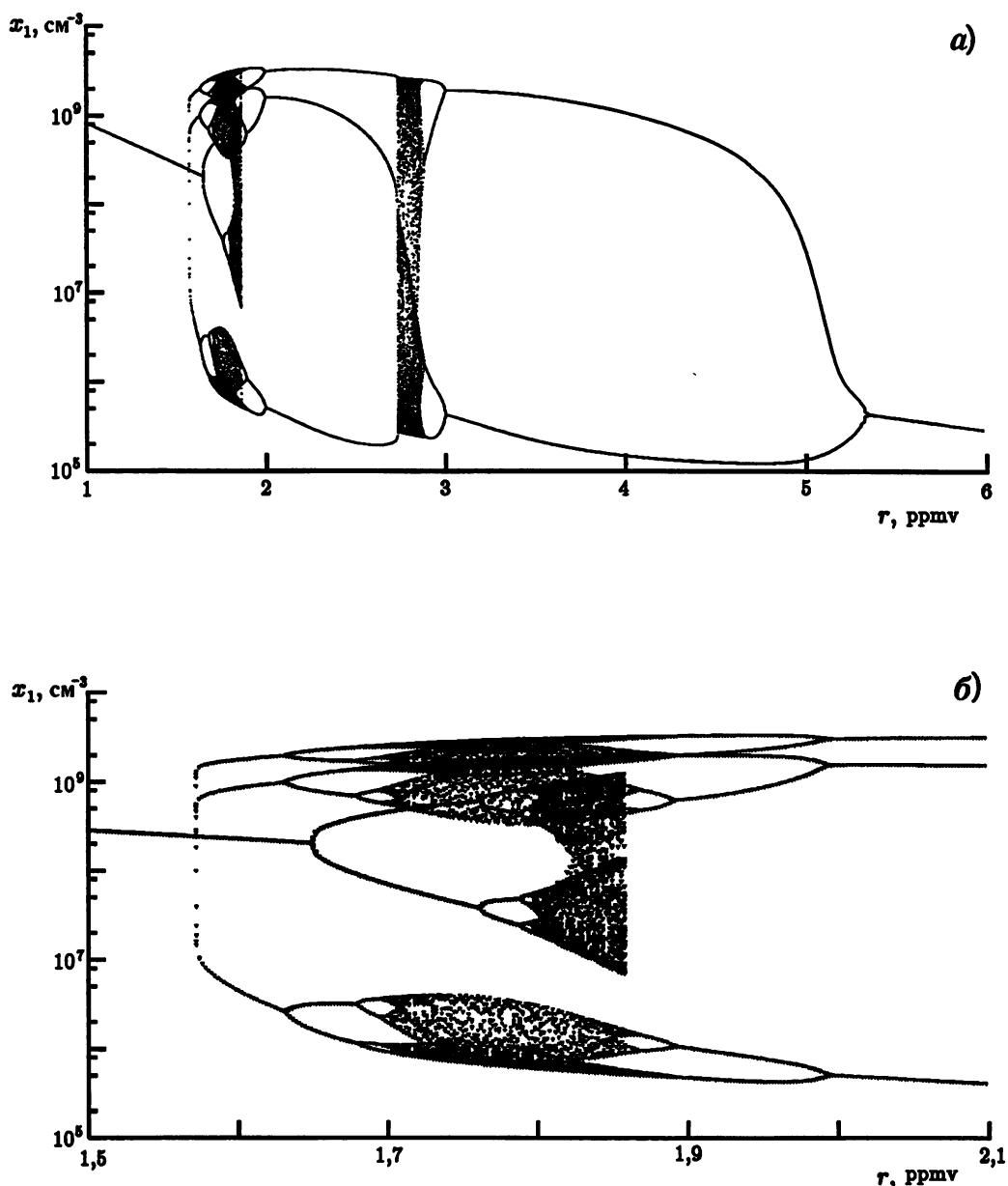


Рис. 6. БД мезосферной фотохимической системы (23) (а); фрагмент БД с увеличенным масштабом по оси абсцисс (б). Концентрация  $x_1$  атомарного кислорода в момент восхода солнца отложена по вертикальной оси, относительная концентрация  $r$  паров воды — по горизонтальной

Во-первых, был реконструирован аттрактор в фазовом пространстве и установлено, что минимальная размерность вложения системы (23) равна 3.<sup>14</sup> Во-вторых, исходный участок реализации был разбит на отрезки длиной  $L = 110$ , и по первому из отрезков  $t \in [0, 110]$  (см. рис. 7б) была построена мгновенная модель. В качестве секущей Пуанкаре была выбрана поверхность  $dx_1/dt = 0$ . Регистрировались пересечения фазовой траекторией данной секущей со стороны  $dx_1/dt < 0$  (локальные ми-

<sup>14</sup> Причина столь невысокой размерности системы (23) подробно обсуждается в работах [17, 18].

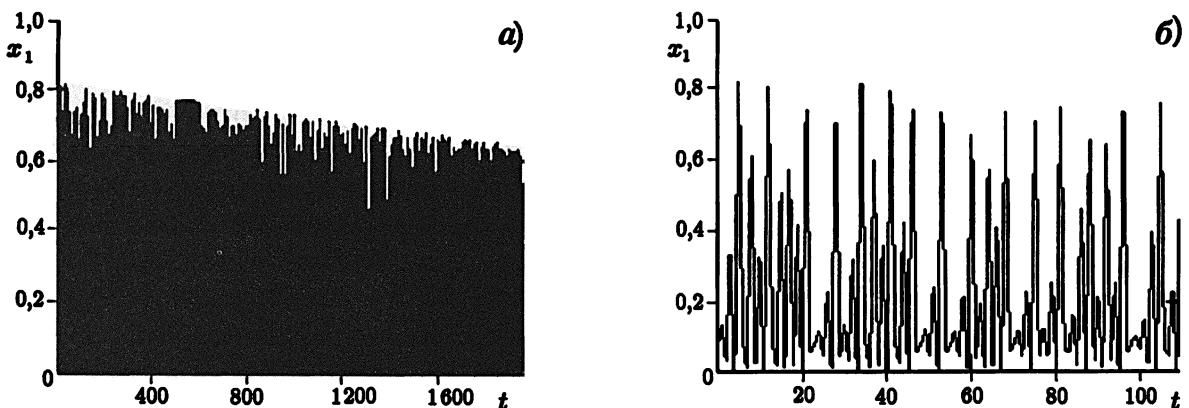


Рис. 7. Мезосферная фотохимическая система: «наблюдаемый» временной ряд, используемый для построения прогноза (а); начальный отрезок этого ряда (б). Время в сутках отложено по оси абсцисс, нормированная переменная  $x_1$  — по оси ординат; нормировочный коэффициент  $(x_1)_n = 1,51 \cdot 10^9 \text{ см}^{-3}$

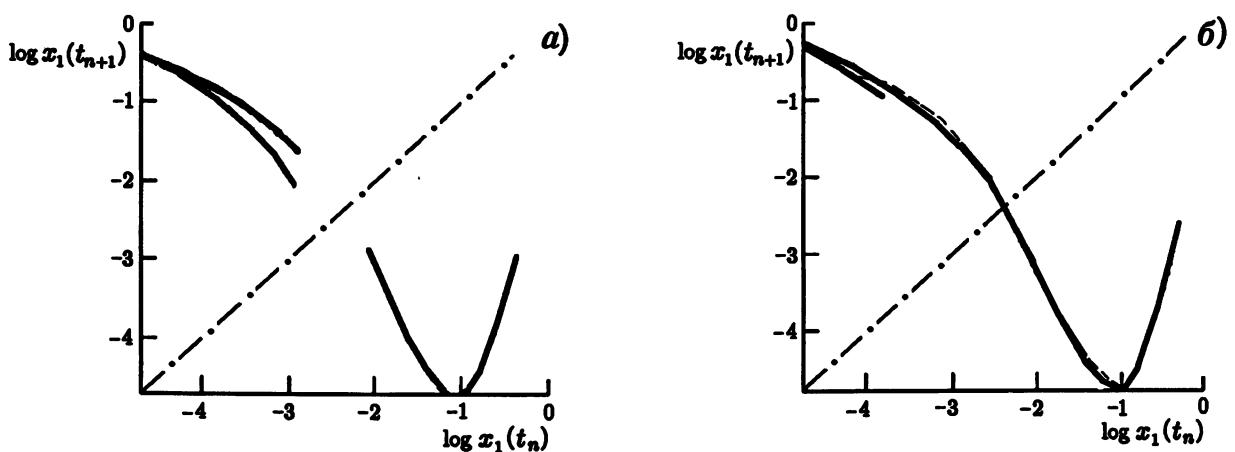


Рис. 8. Функция последовательности первого порядка, построенная по начальному отрезку «наблюдаемого» временного ряда (а); та же функция последовательности с отброшенной «дополнительной» ветвью и её аппроксимация полиномом (24) (б). По обеим осям отложены нормированные величины переменной  $x_1$  в логарифмическом масштабе

нимумы переменной  $x_1$ ). В расширенном фазовом пространстве такая секущая соответствует моменту восхода солнца.

Далее была построена функция последовательности первого порядка. Оказалось, что эта функция является, вообще говоря, неоднозначной (см. рис. 8а) из-за наличия небольшой «дополнительной» ветви в левой части точечного отображения. При построении мгновенной модели данная ветвь отбрасывалась (мы посчитали это допустимым, т. к. средняя плотность точек на этой ветви много меньше, чем на остальной части аттрактора). Оставшаяся часть функции последовательности, приведённая на рис. 8б, аппроксимировалась полиномом пятой степени:

$$(x_1)_{n+1} = \mu_1 + \mu_2 (x_1)_n + \mu_3 (x_1)_n^2 + \mu_4 (x_1)_n^3 + \mu_5 (x_1)_n^4 + \mu_6 (x_1)_n^5. \quad (24)$$

Как и в случае системы Рёссlera, параметры мгновенной модели (24) находились методом наименьших квадратов. Их значения  $\mu_{0i}$ , где  $i = 1, 2, \dots, 6$ , рассчитанные по отрезку  $t \in [0, 110]$  временного ряда,

равны  $-0,13; 11,78; 9,97; 3,44; 0,56$  и  $0,035$  соответственно. Зависимость (24) для данных значений параметров показана на рис. 8б пунктиром.

После этого по всем остальным отрезкам  $L \in T$  была восстановлена неавтономность «наблюдаемой» динамической системы в виде зависимости вектора параметров  $\mu_0$  от медленного времени  $t_s$ . В качестве примера на рис. 9а, б приведены такие зависимости для коэффициентов  $\mu_{0(5,6)}$  при двух старших степенях полинома. Для всех шести параметров модели (24) адекватной (удовлетворяющей условию (8)) оказалась линейная аппроксимация  $(\mu_{0r})_i$  зависимости  $\mu_{0i}(t_s)$ , показанная пунктиром на рис. 9а, б.<sup>15</sup> Аппроксимация коэффициентов при старших членах полинома (24) имеет вид

$$(\mu_{0r})_5 = 0,17 + 3,3 \cdot 10^{-4} t_s, \quad (\mu_{0r})_6 = 3,9 \cdot 10^{-3} + 2,6 \cdot 10^{-5} t_s. \quad (25)$$

В соответствии с общим алгоритмом построенная прогностическая модель (функция последования (24) с восстановленной зависимостью  $\mu_{0r}(t_s)$  её параметров от времени) использовалась далее для построения прогноза бифуркаций.

**3.3.** Во-первых, был построен прогноз бифуркаций по среднему тренду  $\mu_{0r}(t_s)$ , приведённый на рис. 10а. На этом же рисунке воспроизведён фрагмент истинной бифуркационной диаграммы с рис. 6б, включающий внутренний аттрактор. Видно, что верно предсказаны все бифуркации (в том числе катастрофическая, приводящая к исчезновению аттрактора с ростом  $r$ ), кроме одной: при уменьшении контрольного параметра не удалось воспроизвести выход системы на режим, соответствующий колебаниям с периодом в одни сутки (отметим, что именно этот бифуркационный переход наиболее удалён во времени от границ исходного ВР). Вместо этого прогноз по среднему тренду предсказывает при  $t \approx 9\,100$  переход из двухпериодического режима в отсутствующий у исходной системы четырёхпериодический и далее, через каскад удвоения, в хаотический режим, заканчивающийся при  $t \approx 10\,000$  катастрофой (глобальной неустойчивостью прогностической модели (24)).

Далее была рассчитана и экстраполирована за пределы наблюдаемого ВР ковариационная матрица  $C_\mu$  параметров прогностической модели (см. (8) и (16)). После этого были найдены зависимости от времени плотностей вероятности моментов бифуркационных переходов. Мы использовали набор из 10 000 трендов, распределённых, как это описано в подразделе 1.5, случайным образом в шестимерном пространстве параметров модели (24). Построению статистического ансамбля моментов бифуркационных переходов предшествовал анализ бифуркационных диаграмм, соответствующих различным трендам. Напомним, что для системы Рёссlera в результате такого анализа были выявлены бифуркации потери глобальной устойчивости модели, не имевшие место для среднего тренда. В пространстве параметров прогностической модели МФХС были обнаружены четыре бифуркационные поверхности, которые не пересекают средний тренд: 1) поверхность, соответствующая переходу  $2 \rightarrow 1$  от осцилляций с периодом двое суток к суточным колебаниям и 2)–4) поверхности, разделяющие области колебаний с периодом четверо, двое и одни сутки и область глобальной неустойчивости модели (переходы  $4 \rightarrow \infty$ ,  $2 \rightarrow \infty$  и  $1 \rightarrow \infty$ ). Отметим, что первая из данных бифуркаций действительно имеет место в системе (23) при уменьшении контрольного параметра  $r$ .

Как и в случае системы Рёссlera, визуально различные бифуркации, соответствующие переходам в режимы с временем жизни, меньшим шага дискретизации медленного времени, были исключены из статистического анализа. На рис. 10б приведены плотности вероятностей моментов «будущих» бифуркаций  $8 \rightarrow 4, 4 \rightarrow 2, 2 \rightarrow 1, 2 \rightarrow 4$  и  $2 \rightarrow \infty$ , а также бифуркации  $ch \rightarrow \infty$  в «прошлом». Видно, что в отличие от системы Рёссlera точность предсказания моментов бифуркаций не убывает монотонно

<sup>15</sup> Отметим, что зависимость параметров модели внешнего аттрактора от медленного времени оказалось необходимым аппроксимировать полиномом второй степени.

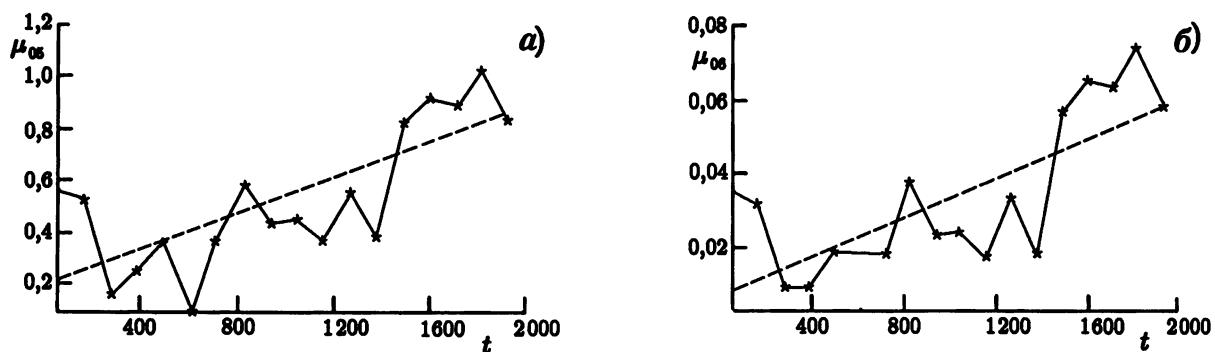


Рис. 9. Реконструированная зависимость параметров  $\mu_{05}$  и  $\mu_{06}$  от медленного времени

при удалении момента бифуркации от границ исходного временного ряда. Например, момент бифуркации  $2 \rightarrow 1$  предсказан почти вдвое точнее, чем менее удалённый «в будущее» момент бифуркации  $8 \rightarrow 4$ . Это означает, что соответствующие бифуркационные поверхности в пространстве параметров прогностической модели МФХС имеют существенно различную форму. На рис. 10 $\varepsilon$  приведены зависимости интегральной вероятности моментов бифуркаций от времени. Видно, что максимальное значение интегральной вероятности равно 1 для бифуркационных переходов  $ch \rightarrow \infty$ ,  $8 \rightarrow 4$  и  $4 \rightarrow 2$ , наиболее близких к границе исходной реализации. Максимальное значение интегральной вероятности для последней бифуркации  $2 \rightarrow 1$  обратного каскада удвоения периода существенно меньше единицы:  $(\Phi_{2 \rightarrow 1})_{\max} \approx 0,2$ , что объясняется, как и в случае системы Рёссlera, появлением альтернативных возможностей выхода из режима. Для двухпериодического типа поведения имеется три варианта выхода из режима: бифуркации  $2 \rightarrow 1$ ,  $2 \rightarrow 4$  и менее вероятный переход  $2 \rightarrow \infty$ . Такое размножение бифуркационных сценариев снижает, естественно, качество прогноза: максимальная интегральная вероятность всех будущих бифуркаций, начиная с перехода  $2 \rightarrow 1$ , существенно меньше 1.

На рис. 10 $\varepsilon$  построены вероятности обнаружения различных режимов поведения системы как функции времени. С вероятностью, равной единице, предсказана глобальная неустойчивость «в прошлом», действительно имеющая место у наблюдаемой ДС. Максимальные вероятности остальных режимов заметно меньше 1. Для четырёх- и двухпериодических режимов причиной этого является значительное перекрытие во времени плотностей вероятностей моментов бифуркаций входа в эти режимы и выхода из них (см. рис. 10 $\delta$ ). Для осцилляций с периодом в одни сутки дополнительным фактором, уменьшающим  $(P_1)_{\max}$ , является существенно меньшая единицы максимальное значение  $(\Phi_{2 \rightarrow 1})_{\max} \approx 0,2$  интегральной вероятности входа в этот режим.

В целом можно констатировать, что качество прогноза бифуркаций для МФХС хуже, чем для системы Рёссlera. Одна из причин этого — существенно более сложная структура пространства параметров прогностической модели. Вторым фактором, ухудшающим качество прогноза, является весьма грубая аппроксимация наблюдаемой функции последования при построении модели (24). Неоднозначность приведённой на рис. 8 $a$  функции последования первого порядка означает, что для улучшения качества прогноза следует строить прогностическую модель в виде функции последования более высокого порядка. Очевидно, что число параметров модели возрастёт при этом в несколько раз, что приведёт к ситуации, типичной для ДС более высокой размерности. Наши предварительные исследования показали, что в этом случае более эффективным оказывается построение модели в виде нейронной сети. Соответствующая модификация алгоритма построения прогноза будет предметом отдельной публикации.

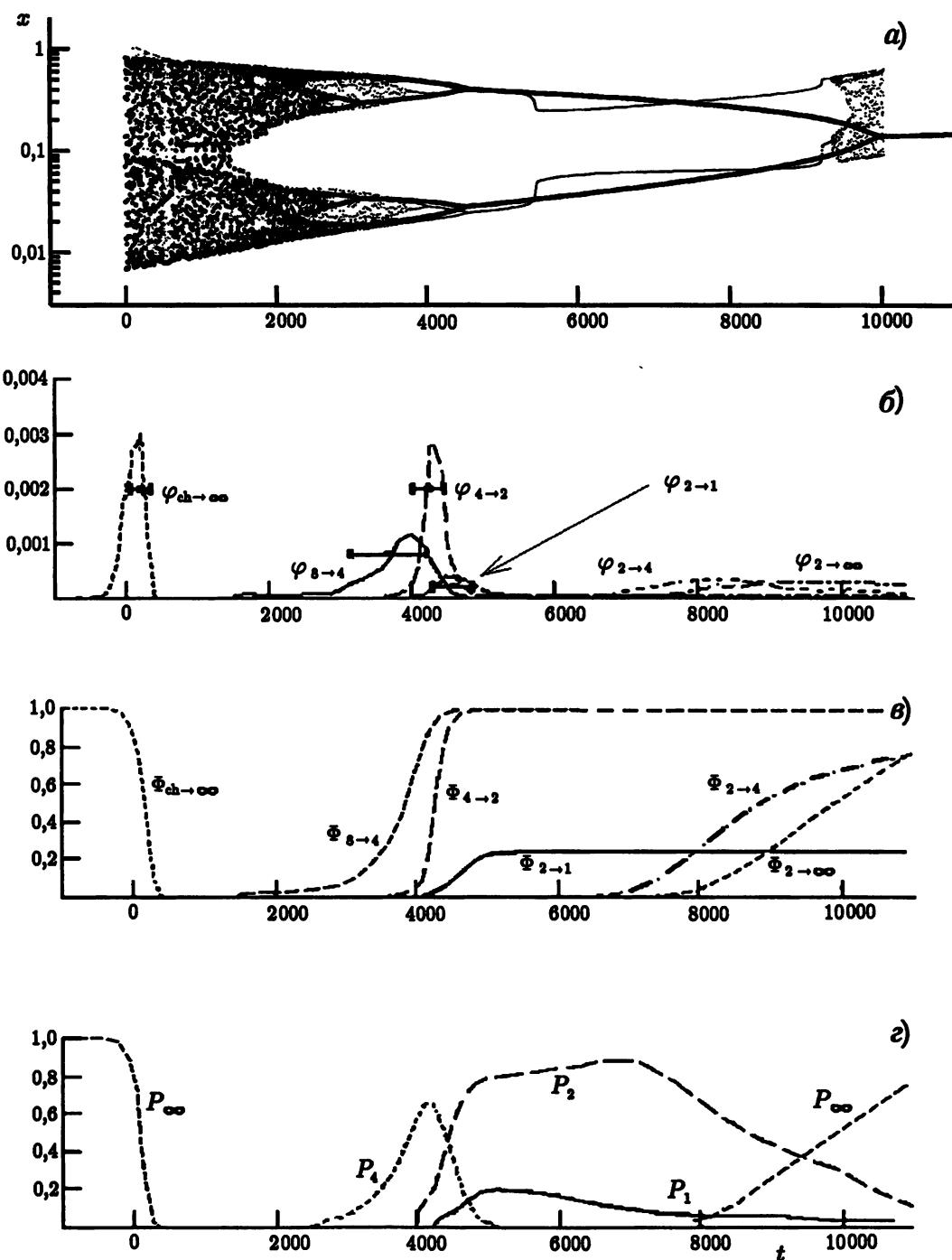


Рис. 10. Результаты прогноза: БД, соответствующая среднему тренду (серые точки) и исходная БД с рис. 6 (чёрные точки) (а); зависимости от времени плотностей вероятности моментов бифуркаций с математическими ожиданиями и среднеквадратичными отклонениями (б); зависимости от времени интегральных вероятностей моментов бифуркаций (в); зависимости от времени вероятностей обнаружения различных режимов (г)

#### 4. ЗАКЛЮЧИТЕЛЬНЫЕ ЗАМЕЧАНИЯ

**4.1.** В данной работе поставлена задача долгосрочного прогноза качественного поведения динамических систем на основе анализа наблюдаемых слабонестационарных хаотических временных рядов. Предложен метод построения прогностических моделей по наблюдаемой эволюции одной динамической переменной, решающий данную задачу применительно к прогнозу бифуркаций низкоразмерных ДС.

Метод применён для анализа временных рядов, сгенерированных системой Рёсслера и мезосферной фотохимической системой. Был использован временной ряд, рассчитанный при медленном изменении управляющего параметра системы. При этом продолжительность «наблюдаемой» реализации была ограничена таким образом, чтобы на всём её протяжении система демонстрировала только один — хаотический — тип поведения, без каких-либо бифуркаций.

Метод позволил на временах, существенно превышающих длительность исходной реализации, сделать для обеих систем правильный прогноз последовательности бифуркаций, указать ожидаемые моменты времени бифуркационных переходов и точность определения этих моментов, а также рассчитать вероятности обнаружить в интересующий момент времени предсказываемые режимы поведения системы.

**4.2.** Принципиальным для применения изложенного подхода является предсказание бифуркаций более сложного поведения в более простое: размерность аттракторов, возникающих в фазовом пространстве ДС в результате бифуркаций, не должна превышать размерности аттрактора, отвечающего наблюдаемому временному ряду. Подчеркнём, что предложенный алгоритм позволяет преодолевать окна регулярного поведения в хаотическом аттракторе, однако использование для построения прогностической модели наиболее сложного типа поведения системы имеет решающее значение. Достаточно очевидно, например, что описанная общая процедура не позволяет предсказать каскад бифуркаций удвоения периода в случае, когда исходный временной ряд содержит информацию о простейшем для такого каскада типе поведения. Отметим при этом, что круг реальных систем, демонстрирующих хаотическую динамику, даже на сегодняшний день весьма широк (от атмосферных и атмосферно-океанических систем [19–22] до различных систем нейронов в живых организмах [23, 24] и тектонической активности [25]) и имеет тенденцию к расширению по мере приложения современных методов нелинейно-динамического анализа к новым базам данных, относящимся к явлениям различной природы.

**4.3.** Другим условием применимости рассмотренного подхода является требование к медленности изменения внешних условий по сравнению с характерными временами эволюции быстрой динамической подсистемы. Требуемая степень медленности имеет естественные ограничения сверху и снизу. Ограничение сверху связано с необходимостью выделить тренды параметров в условиях конечной точности воспроизведения прогностической моделью наблюдаемой эволюции. Нижняя граница медленности определяется влиянием нестационарности наблюдаемого ВР на точность отыскания параметров мгновенной модели: погрешность кусочно-автономного приближения тем больше, чем больше скорость тренда.

**4.4.** Изложенный алгоритм включает в себя операции, минимизирующие погрешность воспроизведения наблюдаемого поведения ДС конкретной динамической моделью. Задача построения модели по временному ряду является обратной и в силу этого допускает множество решений. Выбор конкретной модели должен определяться целью моделирования. Естественным критерием отбора при построении прогностических моделей является качество прогноза, оцениваемое, например, по степени близости к единице максимальных интегральных вероятностей предсказываемых бифуркаций. Наши численные эксперименты показали, что для каждого вида модели (2) (полином, дробно-полиномиальная или экспоненциально-полиномиальная зависимости и т. д.) существует оптимальное с точки зрения каче-

ства прогноза число параметров прогностической модели. Здесь имеется в виду полное число параметров, равное

$$M_T = \sum_{i=1}^M (\Gamma_i + 1),$$

где  $M$  — число параметров мгновенной модели (2),  $\Gamma_i$  — степень полинома (5), аппроксимирующего тренд  $i$ -го параметра. Отметим, что на сегодняшний день нам неизвестен общий рецепт априорного выбора оптимального вида модели, и такой выбор производился путём сравнения качества прогнозов бифуркаций, построенных с помощью оптимальных моделей различных видов.

**4.5.** Достаточно очевидным является развитие алгоритма, позволяющее преодолеть указанные в разделе 4.3 ограничения, налагаемые кусочно-автономным приближением. Действительно, после того, как построена мгновенная модель и восстановлена зависимость её параметров от времени, процедура отыскания оптимальных значений всех параметров прогностической модели может быть повторена уже без разбиения наблюдаемого ВР на автономные отрезки. Погрешность воспроизведения наблюдаемого поведения ДС должна при этом уменьшиться не только вследствие «гладкой» неавтономности модели, но и из-за эффективного увеличения протяжённости ВР, используемого для отыскания оптимальных значений параметров. Модернизированная версия алгоритма будет изложена в отдельной публикации.

Авторы благодарны А. В. Гапонову-Грехову и М. И. Рабиновичу за плодотворные дискуссии.

Работа выполнена при поддержке РФФИ (проект № 99–02–16162) и Комплексной программы научных исследований Президиума РАН «Нелинейная динамика и солитоны» (проект 3.3).

#### СПИСОК ЛИТЕРАТУРЫ

1. Abarbanel H. D. I., Analysis of Observed Chaotic Data. New York: Springer-Verlag, 1997.
2. Takens F. // Lecture notes in mathematics. Berlin: Springer, 1981. V. 898. Dynamical systems and turbulence, Warwick, 1980 / Ed. by D. A. Rand, L.-S. Young. P. 366.
3. Farmer J. D., Sidorowich J. J. // Phys. Rev. Lett. 1987. V. 59. P. 845.
4. Manuka R., Savit R. // Physica D. 1996. V. 99. P. 134.
5. Schreiber T. // Phys. Rev. Lett. 1997. V. 78. P. 843.
6. Witt A., Kurths J., Pikovsky A. // Phys. Rev. E. 1998. V. 58. P. 1800.
7. Янсон Н. Б. и др. // Письма в ЖТФ. 1999. Т. 25. С. 74.
8. Stark J. et al. // Nonlinear Analysis, Theory, Methods and Applications. 1997. V. 30. P. 5 303.
9. Анищенко В. С., Вадивасова Т. Е., Астахов В. В. Нелинейная динамика хаотических и стохастических систем. Саратов: Изд-во СГУ, 1999. 368 с.
10. Rössler O. E. // Phys. Lett. A. 1976. V. 57. P. 397.
11. Fichtelmann B., Sonnemann G. // Annals of Geophysics. 1992. V. 10. P. 719.
12. Sonnemann G., Fichtelmann B. // J. Geophys. Res. D. 1997. V. 102. P. 1 193.
13. Fraser A. M., Swinney H. L. // Phys. Rev. A. 1986. V. 33. P. 1 134.
14. Kennel M. B., Brown R., Abarbanel H. D. I. // Phys. Rev. A. 1992. V. 45. P. 3 403.
15. Anishenko V. S., Pavlov A. N. // Phys. Rev. E. 1998. V. 57. P. 2 455.
16. Корн Г., Корн Т. Справочник по математике. М.: Наука, 1979.
17. Feigin A. M., Konovalov I. B., Molkov Y. I. // J. Geophys. Res. D. 1998. V. 103. P. 25 447.
18. Konovalov I. B., Feigin A. M. // Nonlinear Processes in Geophysics. 2000. V. 7. P. 87.
19. Yang P., Brasseur G. P., Gille J. C. et al. // Physica D. 1994. V. 76. P. 331.
20. Li I.-F., Biswas P., Islam S. // Atmospheric Environment. 1994. V. 28. P. 1 707.

21. Neelin J. D., Latif M. // Physics Today. 1998. No. 12. P. 32.
22. Wang B., Barcilon A., Fang Z. // J. Atmos. Sci. 1999. V. 56. P. 5.
23. Abarbanel H. D. I., Huerta R., Rabinovich M. I. et al. // Neural Comput. 1996. V. 8, No. 8. P. 1567.
24. Frank G. W., Lookman T., Nerenberg M. A. H. et al. // Physica D. 1990. V. 46. P. 427.
25. Srivastava H. N., Bhattacharya S. N., Sinha Ray K. C. // Geophys. Res. Lett. 1996. V. 23. P. 3519.

Институт прикладной физики РАН,  
г. Нижний Новгород, Россия

Поступила в редакцию  
11 апреля 2001 г.

## **PROGNOSIS OF QUALITATIVE BEHAVIOUR OF A DYNAMIC SYSTEM BY THE OBSERVED CHAOTIC TIME SERIES**

*A. M. Feigin, Ya. I. Molkov, D. N. Mukhin, and E. M. Loskutov*

An approach to the long-term prognosis of qualitative behavior of a dynamic system (DS) is proposed, which is based on the nonlinear-dynamical analysis of a weakly nonstationary chaotic time series (TS). A method for constructing prognostic models using the observed evolution of a single dynamical variable is described, which employs the proposed approach for prediction of bifurcations of low-dimensional DSs. The method is applied to analyze the TSs generated by the Rossler system and the system of equations modeling photochemical processes in the mesosphere. The analysis is performed for a TS calculated in the case of a slow variation in the control parameter of the system. The duration of the «observed» TS is limited such that the system demonstrates only one, chaotic, type of behavior without any bifurcations during the observed TS. The proposed algorithm allows us to predict correctly the bifurcation sequences for both systems at times much longer than the duration of the observed TS, to point out the expected instants of specific bifurcation transitions and accuracy of determining these instants, as well as to calculate the probabilities to observe the predicted regimes of system's behavior at the time of interest.

УДК 530.182

## АВТОСОЛИТОНЫ В СИСТЕМАХ С ГЛОБАЛЬНОЙ НЕЛИНЕЙНОСТЬЮ

Г. Зоннеманн<sup>1</sup>, В. Е. Семёнов<sup>2</sup>

Рассматривается простейшая математическая модель локализованных диссипативных структур, описываемых одним диффузионным уравнением с источником, содержащим как локальную, так и глобальную нелинейность, т. е. зависящим от интеграла по всему объёму от неизвестной функции. На основании качественного анализа делается вывод о возможности существования стабильных автосолитонов в рассматриваемой системе. Данный вывод подтверждается результатами численного моделирования в рамках трёхмерной задачи. Определена зависимость мощности автосолитона от параметров задачи. Обнаружены метастабильные многосолитонные состояния и исследованы их свойства. Обсуждается возможность стабилизации трёхмерных автосолитонов в системе без локальных потерь.

## ВВЕДЕНИЕ

В своё время изучение нелинейных диссипативных систем привело к открытию таких процессов, как автоколебания [1] и автоволны [2], которые в отличие от нелинейных колебаний и волн в консервативных системах слабо зависят от начальных условий. Непрерывно растущий интерес к этим явлениям обусловлен тем обстоятельством, что они встречаются в самых различных областях науки, от физики и химии до биологии и медицины [3]. По аналогии с автоколебаниями под автоволнами понимали первоначально самоподдерживающийся волновой процесс (в виде бегущих импульсов или ступенек) в неравновесной среде. Затем это понятие было обобщено и на стационарные диссипативные структуры, из которых со временем были выделены в отдельный класс (по аналогии с солитонами) автосолитоны [4], представляющие собой локализованные в пространстве структуры. Чаще всего автоволновые процессы ассоциируются с математическими моделями, опирающимися на уравнения диффузионного типа с нелинейными источниками [2, 3]. В рамках таких же моделей описываются и автосолитоны [4]. Однако в последнее время всё больший интерес вызывают системы с так называемой глобальной (или интегральной) нелинейностью [2, 3, 5–8], в которых источники включают в себя зависимость от интегральных характеристик неизвестных пространственных распределений параметров, т. е. от интегралов по всему пространству от искомых функций. Нелинейности такого типа описывают, в частности, истощение ресурсов источника, обусловленное коллективными эффектами. Как было показано в одной из последних работ [8], в отличие от случая локальной нелинейности источника, когда устойчивые автосолитоны существуют лишь при использовании моделей, содержащих не менее двух диффузионных уравнений [4], в условиях глобальной нелинейности устойчивые автосолитоны реализуются в рамках очень простых математических моделей, описываемых только одним уравнением диффузионного типа. Анализ одной из таких моделей и результаты соответствующих численных исследований и представлены в настоящей работе.

## 1. ОБЩИЙ АНАЛИЗ БАЗОВОЙ МАТЕМАТИЧЕСКОЙ МОДЕЛИ

Исходным уравнением рассматриваемой ниже базовой математической модели является параболическое уравнение с нелинейным источником:

$$\frac{\partial u}{\partial t} = D \nabla^2 u + A(J)f(u) - \nu u, \quad (1)$$

где неизвестная функция координат и времени  $u(x, y, z, t)$ , характеризующая искомое распределение плотности некоторого параметра в пространстве, предполагается неотрицательно определённой ( $u \geq 0$ ) и интегрируемой, функция  $f(u)$ , описывающая локальную нелинейность источника, предполагается монотонно возрастающей и обращающейся в нуль при  $u = 0$ , коэффициент диффузии  $D$  и эффективная частота потерь  $\nu$  полагаются постоянными, а интенсивность источника  $A$  монотонно убывает с ростом полной мощности параметра, т. е. интеграла  $J = \int u dx dy dz$  по всему пространству. Нелокальная нелинейность такого типа описывает, например, интенсивность электромагнитного поля в объёме высокодобротного резонатора, заполненного неоднородной разреженной плазмой. Соответственно, уравнение (1) может быть использовано для описания процесса контракции неравновесного микроволнового разряда в квазиоптическом резонаторе [8, 9].

Ниже основное внимание будет уделено частному случаю локальной нелинейности с насыщением:  $f(u) = u^\alpha / (1 + u^\alpha)$ , где постоянный показатель  $\alpha > 1$ , но основные результаты, по всей видимости, могут быть обобщены на достаточно широкий класс функций  $f(u)$ , которые при определённых фиксированных значениях  $A$  обеспечивают существование у уравнения (1) в одномерном случае локализованного в пространстве стационарного решения типа солитона с положительно определённой плотностью  $u$ .

При фиксированном значении интенсивности  $A$  анализ пространственно однородных решений уравнения (1) представляет собой тривиальную задачу. Его результаты для рассматриваемой нелинейности  $f(u) = u^\alpha / (1 + u^\alpha)$  можно сформулировать следующим образом. В зависимости от отношения  $\mu \equiv A/\nu$  имеется один (при  $\mu < \mu_1(\alpha) = \alpha(\alpha - 1)^{-1+1/\alpha}$ ) или три (при  $\mu > \mu_1(\alpha)$ ) корня уравнения  $Af(u) - \nu u = 0$ , из которых минимальный  $u = 0$  и максимальный  $u = u_m$  отвечают асимптотически устойчивым стационарным решениям.

В общем случае интегрируемого пространственного распределения  $u$  мощным инструментом анализа решения уравнения (1) при фиксированной интенсивности  $A$  является функционал Ляпунова

$$G[u] = \int \left\{ \frac{D(\nabla u)^2}{2} - AF(u) + \frac{\nu u^2}{2} \right\} dx dy dz, \quad (2)$$

где  $F(u) = \int_0^u f(u') du'$ . Вариационная производная данного функционала, как нетрудно убедиться, обращается в нуль на стационарных решениях уравнения (1). В то же время  $G[u]$  является невозрастающей функцией времени:

$$\frac{dG[u]}{dt} = - \int \left( \frac{\partial u}{\partial t} \right)^2 dx dy dz \leq 0. \quad (3)$$

Поэтому стационарные решения, обеспечивающие минимум  $G[u]$ , являются асимптотически устойчивыми. Анализ функционала Ляпунова позволяет, в частности, сделать вывод о том, что имеется ещё одно характерное значение параметра  $\mu = \mu_c(\alpha) > \mu_1$ , при котором  $\mu F(u_m) = u_m^2/2$ . Если  $\mu < \mu_c$ , то функционал Ляпунова на классе интегрируемых функций ограничен снизу, достигая минимального значения на тривиальном решении  $u \equiv 0$ , к которому и стремится с течением времени любое интегрируемое начальное распределение  $u(x, y, z)$ . Если же  $\mu > \mu_c$ , то  $\mu F(u_m) > u_m^2/2$ , и функционал Ляпунова снизу не ограничен. В результате у решения начальной задачи с интегрируемым начальным распределением  $u(x, y, z)$  появляется ещё одна возможность — приближение с течением времени к нетривиальному устойчивому пространственно однородному состоянию с  $u = u_m$ .<sup>1</sup>

Выполненный ранее анализ одномерного аналога задачи (1) с  $A = \text{const}$  [2, 3] показывает, что при  $\mu > \mu_1(\alpha)$  типичными устойчивыми решениями являются стационарные автоволны  $u = U(x - Vt)$  в виде перепада, бегущие с постоянной скоростью  $V$ . Значения  $u$  по разные стороны от бегущего фронта

<sup>1</sup> Как известно, в случае  $A = \text{const}$  возможные солитоноподобные решения уравнения (1) являются неустойчивыми [10, 11].

перепада совпадают с устойчивыми стационарными пространственно однородными состояниями  $u = 0$  и  $u = u_m$ , ширина фронта определяется диффузионной длиной  $L_d = \sqrt{D/\nu}$ , а его скорость  $V$  зависит от интенсивности  $A$  и локальной нелинейности источника  $f(u)$ . Если  $\mu < \mu_c$ , то стационарная автоволна описывает расширение области с  $u = 0$ . В этом случае релаксация широкого (по сравнению с  $L_d$ ) слоя с  $u \approx u_m$  на начальном этапе обусловлена уменьшением его толщины без заметного изменения максимального значения  $u$ , что может быть приближенно описано посредством двух встречных автоволн. В противоположном случае ( $\mu > \mu_c$ ) стационарная автоволна отвечает расширению области с  $u = u_m$ . Соответственно, две разбегающиеся автоволны будут описывать расширение широкого слоя с  $u \approx u_m$ , а релаксирующими будет лишь начальный пик  $u$  с недостаточно большой шириной или высотой. Стационарный солитон, который существует в данном случае, является, как уже отмечалось, неустойчивым образованием и представляет собой своеобразную границу между неограниченно расширяющимся и релаксирующими слоями. Другими словами, слой с плотностью  $u \approx u_m$  и толщиной, превышающей ширину солитона, будет неограниченно расширяться, а слой с такой же плотностью и толщиной меньше ширины солитона будет релаксировать к тривиальному однородному распределению  $u \equiv 0$ . Следует отметить, что с уменьшением параметра  $\mu$  ширина одномерного солитона монотонно возрастает и стремится к бесконечности при  $\mu \rightarrow \mu_c$ . Соответственно, при  $\mu = \mu_c$  появляется стационарное решение типа перепада, отвечающее автоволне с нулевой скоростью.

Полная мощность стационарного солитона  $J_s = L_d \varphi(\mu)$  (здесь функция  $\varphi(\mu)$  определяется видом локальной нелинейности  $f(u)$ ) оказывается также монотонно убывающей функцией параметра  $\mu$  (при фиксированной диффузионной длине  $L_d$ ). Неустойчивость же солитона позволяет с определённой степенью достоверности утверждать, что для квазисолитонных распределений плотности  $u$  и мощность  $J$  неограниченно возрастает с течением времени, если  $J > J_s$ , и убывает до нуля в противоположном случае.

Результаты данного качественного анализа одномерной задачи дают основания предполагать, что при контролируемом изменении интенсивности источника  $A$  с течением времени возможна, в принципе, стабилизация солитона. Необходимое для этого управление должно обеспечивать уменьшение интенсивности источника при расширении слоя с повышенной плотностью  $u$  и увеличение  $A$  в противоположном случае. Данный принцип управления может быть реализован в автоматическом режиме, если интенсивность  $A$  является монотонно убывающей функцией полной мощности  $J$ . Именно такой режим реализуется в процессе контракции как тлеющего [12], так и микроволнового [9, 13–15] разрядов, где увеличение полного числа электронов приводит к снижению напряжённости электрического поля в системе и экспериментально наблюдаются локализованные солитоноподобные структуры.

Иллюстрация необходимой для стабилизации солитона глобальной нелинейности  $A(J)$  приведена на рис. 1 (кривая 3). Кривые 1 и 2 изображают здесь зависимость мощности стационарного солитона от интенсивности источника  $J = J_s(A, D, \nu) = \sqrt{D/\nu} \varphi(A/\nu)$  для двух различных значений коэффи-

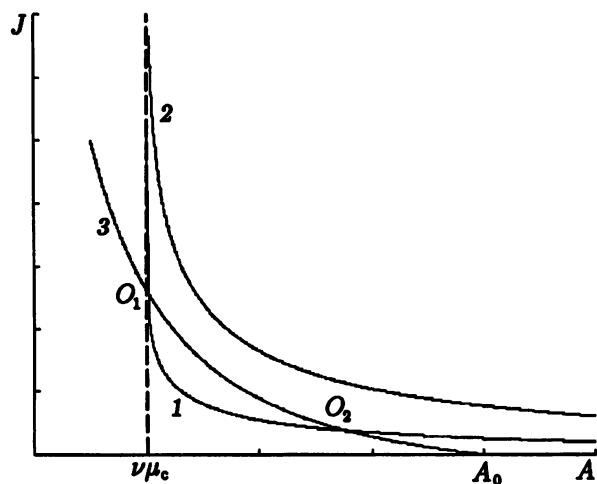


Рис. 1. Графический анализ возможных параметров автосолитона. Устойчивому автосолитону соответствует точка  $O_1$  пересечения кривых 1 и 3. Кривая 3 отражает обусловленную глобальной нелинейностью зависимость интенсивности источника от полной мощности  $J$  параметра  $u$ . Параметр  $A_0$  характеризует невозмущённую интенсивность источника, соответствующую  $J = 0$ . В случае нелинейности (4)  $A_0 = 1$

циента диффузии ( $D_1 < D_2$ ). Возможным параметрам автосолитона при рассматриваемой глобальной нелинейности отвечают точки  $O_1, O_2$  пересечения кривой 3 с кривыми 1 или 2. Заметим, что такие точки существуют только при достаточно малых значениях коэффициента диффузии  $D$  и эффективной частоты потерь  $\nu$ . Как уже отмечалось выше, при фиксированной интенсивности источника  $A$  полная мощность  $J$  квазисолитонного пика параметра  $u$  имеет тенденцию к возрастанию, если  $J > J_s$ , или убыванию, если  $J < J_s$ . Поэтому точка  $O_1$ , отвечающая автосолитону с большим значением интеграла  $J$ , может соответствовать устойчивому состоянию, а точка  $O_2$  соответствует неустойчивому состоянию. Данные качественные выводы были впервые подтверждены в работе [8], где методами численного моделирования исследовалось формирование автосолитона при степенном характере глобальной нелинейности:

$$A(J) = (1 + J)^{-\beta}. \quad (4)$$

## 2. РЕЗУЛЬТАТЫ ЧИСЛЕННОГО МОДЕЛИРОВАНИЯ

Численное решение уравнения (1) с локальной нелинейностью  $f(u) = u^\alpha/(1 + u^\alpha)$  и глобальной нелинейностью (4) было проведено в широком диапазоне основных параметров системы  $D, \nu, \beta$  и  $\alpha$ . В соответствии с выводами качественного анализа было обнаружено, что при  $\alpha > 1$  существуют устойчивые автосолитоны, если остальные параметры ограничены сверху определёнными пороговыми значениями. Автосолитоны характеризуются сферически-симметричным распределением плотности  $u$  в пространстве и формируются из всех начальных распределений с достаточно большими плотностью  $u$  и мощностью  $J$ .

Максимальное значение плотности в автосолитоне определяется в основном локальной нелинейностью и при фиксированном значении  $\alpha$  слабо зависит от параметров  $D, \nu, \beta$ , которые, однако, существенно влияют на ширину автосолитона и, следовательно, на его полную мощность  $J$ . В точном соответствии с результатами качественного анализа (см. рис. 1) мощность автосолитона монотонно возрастает с удалением параметров  $D, \nu, \beta$  от своих пороговых значений. Ширина автосолитона при этом увеличивается так, что его вершина уплощается (рис. 2). Если же параметры  $D, \nu, \beta$  приближаются к своим пороговым значениям, то мощность автосолитона уменьшается, а ширина его приближается к диффузионной длине  $L_d$ , т. е. пик плотности в нём становится игольчатым (рис. 3).

Характерный масштаб пограничного слоя (в котором плотность  $u$  падает от своего максимального значения  $u \approx u_m$  до нуля) широкого автосолитона мал по сравнению с радиусом кривизны этого слоя, совпадающим с радиусом  $R$  солитона. Поэтому структура пограничного слоя может быть рассмотрена в рамках одномерной модели, где этому слою соответствует неподвижная автоволна типа перепада. Следовательно, широкому автосолитону отвечает интенсивность источника  $A = (1 + J)^{-\beta} \approx \nu \mu_c$ , т. е. с удалением от пороговых значений параметров  $D, \nu, \beta$  мощность автосолитона приближается к

$$J_\infty = (\nu \mu_c)^{-1/\beta} - 1. \quad (5)$$

Используя методы качественного анализа и результаты численного моделирования, для широкого автосолитона можно определить и поправки к (5). Действительно, из уравнения (1) следует, что для любой заданной локальной нелинейности  $f(u)$  зависимость мощности автосолитона от параметров системы может быть представлена в виде

$$J_s(A, D, \nu) = (D/\nu)^{3/2} \psi(A/\nu), \quad (6)$$

где вид функции  $\psi(\nu)$  зависит от локальной нелинейности, т. е. от показателя  $\alpha$  в рассматриваемом случае  $f(u) = u^\alpha/(1 + u^\alpha)$ . Результаты численного моделирования показывают, что в предельном

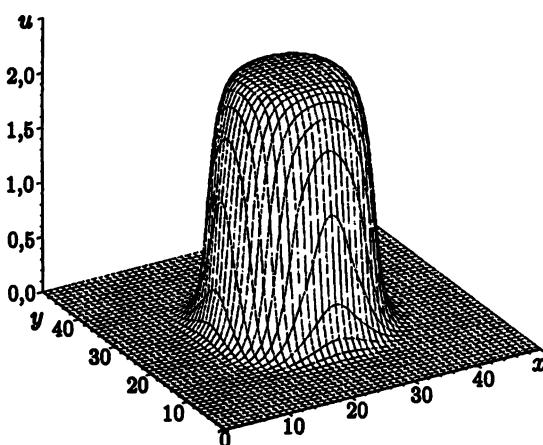


Рис. 2. Широкий автосолитон с плоской вершиной (распределение плотности  $u$  в плоскости симметрии автосолитона). Результаты численного моделирования при следующих значениях основных параметров задачи:  $\alpha = 5$ ;  $\beta = 0,4$ ;  $D = 0,01$ ;  $\nu = 0,01$

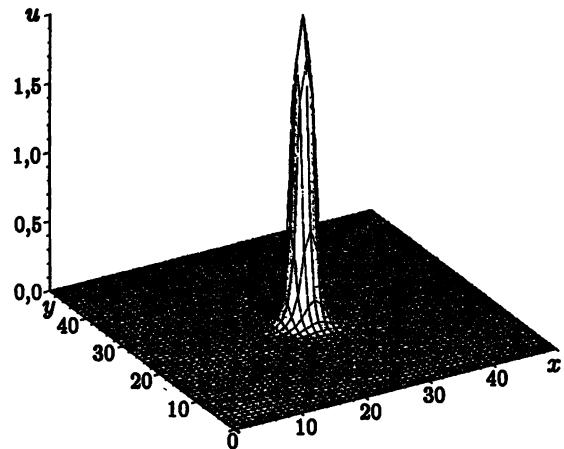


Рис. 3. Игольчатая структура автосолитона (распределение плотности  $u$  в плоскости симметрии автосолитона) при околовороговых значениях параметров задачи ( $\alpha = 5$ ;  $\beta = 0,7$ ;  $D = 0,01$ ;  $\nu = 0,01$ )

случае широкого автосолитона, когда  $A/\nu - \mu_c \ll \mu_c$ , функция  $\psi(A/\nu)$  имеет степенную асимптотику:

$$\psi(A/\nu) \approx a^3 (A/\nu - \mu_c)^{-3/\delta}, \quad (7)$$

где постоянные  $a$  и  $\delta$  зависят от показателя  $\alpha$ . Разрешая уравнения (4)–(6) относительно  $J$ , можно найти, что мощность широкого автосолитона равна

$$J \approx (\nu\mu_c)^{-1/\beta} - 1 - \frac{(\nu\mu_c)^{-1/\beta} (aL_d)^\delta}{\beta\mu_c [(\nu\mu_c)^{-1/\beta} - 1]^{\delta/3}}. \quad (8)$$

Необходимыми условиями корректности этого приближения являются положительность правой части (8) и малость последнего члена в правой части по сравнению с первым. При  $\alpha = 5$  результаты численного моделирования показывают, что величина  $\delta$  с большой степенью точности равна 2, и, следовательно, мощность широкого автосолитона линейно зависит от коэффициента диффузии:

$$J \approx (\nu\mu_c)^{-1/\beta} - 1 - \frac{(\nu\mu_c)^{-1/\beta} a^2 D}{\beta\mu_c \nu [(\nu\mu_c)^{-1/\beta} - 1]^{2/3}}, \quad (9)$$

где постоянные  $a$  и  $\mu_c$  не зависят от параметров задачи  $D$ ,  $\nu$  и  $\beta$  (в частности, при  $\alpha = 5$  имеем  $a \approx 5,7$ ;  $\mu_c \approx 2,1$ ).

Благодаря глобальной нелинейности в рассматриваемой системе оказывается возможным эффективное взаимодействие даже удаленных друг от друга солитоноподобных образований. Как уже отмечалось в рамках качественного анализа одномерной задачи, каждому фиксированному значению интенсивности источника  $A > \nu\mu_c$  соответствует неустойчивый стационарный солитон, мощность которого  $J_s = L_d\varphi(A/\nu)$  монотонно убывает с ростом  $A$ . Солитоноподобное образование с  $u \approx u_m$  расширяется с течением времени, если его мощность превышает  $J_s$ , и релаксирует в противоположном случае. Поэтому в рассматриваемой системе с глобальной нелинейностью следует ожидать, что состояние с несколькими удаленными друг от друга солитонами неустойчиво. Тот из них, который вследствие флуктуаций приобретет большую мощность, будет расширяться, обеспечивая снижение интенсивнос-

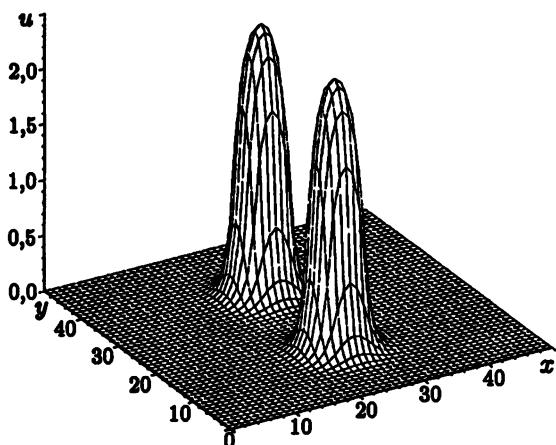


Рис. 4. Метастабильная двухсолитонная структура  
( $\alpha = 5$ ;  $\beta = 0,5$ ;  $D = 0,01$ ;  $\nu = 0,01$ )

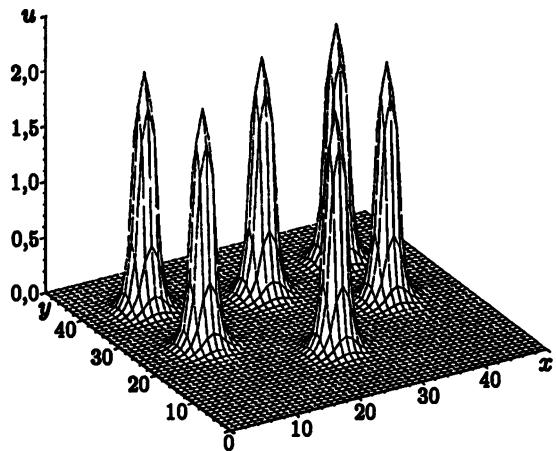


Рис. 5. Метастабильная многосолитонная структура (параметры  $\alpha$ ,  $\beta$ ,  $D$ , и  $\nu$  те же, что и на рис. 4)

ти источника  $A$ , что повлечёт за собой релаксацию всех остальных солитонов. Очевидно, что скорость развития данного процесса не зависит от расстояния между солитонами.

Проведённые численные исследования показали, что описанная выше конкуренция солитонов, приводящая к «выживанию» самого мощного, достаточно типична и в трёхмерной задаче. Однако для не очень широких и не очень удалённых друг от друга солитонов была зафиксирована возможность формирования метастабильных многосолитонных состояний (см. рис. 4, 5). Все солитоны в таких состояниях обладают примерно одинаковой мощностью. Их суммарная мощность оказывается меньше мощности односолитонного состояния, формирующегося в системе при тех же параметрах (см. рис. 6), т. к. интенсивность источника  $A$ , необходимая для поддержания каждого из солитонов в отдельности, возрастает с уменьшением его мощности. Динамика метастабильных многосолитонных состояний характеризуется слабым притяжением соседних солитонов, приводящим в конечном итоге к их слиянию (рис. 7). Запас устойчивости у многосолитонных состояний невелик, поэтому после слияния одной из пар исходных солитонов, приводящего к формированию одного солитона с примерно удвоенной мощностью, начинается описанный выше процесс расширения этого солитона за счёт релаксации всех остальных.

### 3. АВТОСОЛИТОНЫ В СИСТЕМАХ БЕЗ ЛОКАЛЬНЫХ ПОТЕРЬ

В рамках рассмотренной модели локальной нелинейности нет принципиальных отличий трёхмерных систем от систем с меньшей размерностью (двумерных или одномерных). Стационарные локализованные решения существуют в пространстве с любой размерностью. Поэтому следует ожидать, что предложенная глобальная нелинейность позволяет стабилизировать не только исследованный трёхмерный автосолитон, но и двумерный и одномерный автосолитоны. Следует отметить, однако, что при отсутствии локальных потерь ( $\nu = 0$ ) появляется качественное различие между системами в пространствах различной размерности. В этом случае стационарные локализованные структуры в одномерной и двумерной системах невозможны.<sup>2</sup> В трёхмерном же случае такие структуры при некоторых видах локальной нелинейности возможны, хотя и неустойчивы. В частности, для степенной нелинейности

<sup>2</sup> В неограниченной области двумерного или одномерного пространства вообще не существует стационарного решения линейного уравнения диффузии с заданным положительным источником.

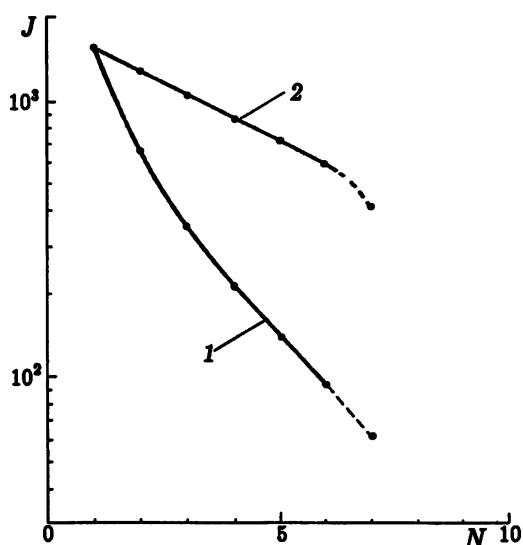


Рис. 6. Зависимость мощности отдельного автосолитона (кривая 1) и суммарной мощности всех автосолитонов (кривая 2) от числа  $N$  автосолитонов в метастабильных многосолитонных состояниях (параметры  $\alpha$ ,  $\beta$ ,  $D$  и  $\nu$  те же, что и на рис. 4). Значение  $J$  при  $N = 1$  соответствует мощности односолитонного состояния

$f(u) = u^\alpha$  локализованные стационарные решения уравнения (1) при фиксированном  $A$  и  $\nu = 0$  существуют, если  $\alpha \geq 5$  [16].<sup>3</sup> Вопрос о возможности стабилизации таких структур в условиях глобальной нелинейности остаётся пока открытым. Плотность  $u$  в этих структурах убывает с расстоянием  $r$  от их центра не быстрее, чем  $1/r$ . Поэтому их полная мощность  $J = \int u d^3r$  бесконечно велика и, следовательно, для их стабилизации не может быть использовано управление интенсивностью источника по этой мощности. Но для стабилизации можно было бы воспользоваться другой интегральной характеристикой структуры.

При фиксированном значении  $A$  и степенной локальной нелинейности характерным элементом динамики локализованных структур является взрывная неустойчивость с обострением пика плотности [16]. Анализ соответствующих автомодельных решений показывает, что при  $\alpha > 5$  такую неустойчивость очень трудно подавить, используя управление по интегральным характеристикам, т. к. её развитие практически не сопровождается увеличением интегралов  $J = \int u d^3r$ ,  $J = \int u^\alpha d^3r$ . Однако для рассмотренной выше локальной нелинейности с насыщением  $f(u) = u^\alpha / (1 + u^\alpha)$  взрывное решение с обострением невозможно. Поэтому при  $\alpha > 5$  и надлежащем выборе нелокальной нелинейности можно надеяться на стабилизацию соответствующего трёхмерного автосолитона. Данная гипотеза ждёт своей проверки.

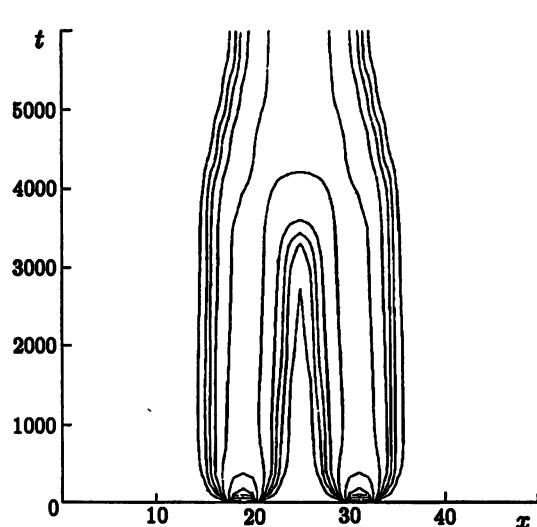


Рис. 7. Линии уровня плотности  $u$  на плоскости  $(x, t)$ , где  $x$  — координата вдоль линии, проходящей через центры двух солитонов, демонстрирующие процесс слияния двух близких автосолитонов (параметры  $\alpha$ ,  $\beta$ ,  $D$  и  $\nu$  те же, что и на рис. 4)

#### СПИСОК ЛИТЕРАТУРЫ

1. Андронов А. А., Витт А. А., Хайкин С. Э. Теория колебаний. М.: Физматгиз, 1959.
2. Васильев В. А., Романовский Ю. М., Яхно В. Г. Автоволновые процессы / Под ред. В. С. Чернавского. М.: Наука, 1987.

<sup>3</sup> При  $\alpha = 5$  данные решения могут быть представлены в явном виде  $u = (3D/A)^{1/4} \sqrt{c/(c^2 + r^2)}$ , где  $c$  — произвольная константа.

3. Vasiliev V. A., Romanovskii Yu. M., Chernavskii D. S., Yakhno V. G. Autowave processes in kinetics systems, spatial and temporal self-organization in physics, chemistry, biology, and medicine Ed. by W. Ebeling, Ch. Weißmantel. Berlin: VEB Deutscher Verlag der Wissenschaften, 1987.
4. Kerner B. S., Osipov V. V. Autosolitons / Ed. by A. van der Merwe. Dordrecht: Kluwer Academic Publishers, 1994.
5. Falke M. Pattern formation in reaction-diffusion systems under global constraints. Berlin: Wissenschaft und Technik Verlag, 1995 (in German).
6. Schenk C. P., Or-Guil M., Bode M., Purwins H.-G. // Phys. Rev. Lett. 1997. V. 78. P. 3781.
7. Sonnemann G. // Prog. Th. Phys. 1998. V. 99, No. 6. P. 931.
8. Sonnemann G., Semenov V. E. // Eur. Phys. J. D. 2000. V. 11. P. 481.
9. Вихарев А. Л., Иванов О. А., Иванова Л. С., Кузнецов О. Ю., Степанов А. Н. // ЖТФ. 1989. Т. 59, № 1. С. 40.
10. Баренблatt Г. И., Истратов А. Г., Зельдович Я. Б. // Ж. прикл. мех. и техн. физ. 1962. № 4. С. 21.
11. Каннель Я. И. // Матем. сб. 1964. Т. 65. С. 398.
12. Велихов Е. П., Ковалёв А. С., Рахимов А. Т. Физические явления в плазме газового разряда. М.: Наука, 1987.
13. Капица П. Л. // ЖЭТФ. 1969. Т. 5, № 6. С. 1801.
14. Вихарев А. Л., Иванов О. А., Кузнецов О. Ю., Степанов А. Н. // ДАН СССР. 1987. Т. 295, № 2. С. 358.
15. Вихарев А. Л., Иванов О. А., Кузнецов О. Ю., Степанов А. Н. // Физика плазмы. 1987. Т. 13, № 9. С. 1124.
16. Самарский А. А., Галактионов В. А., Курдюмов С. П., Михайлов А. П. Режимы с обострением в задачах для квазилинейных параболических уравнений. М.: Наука, 1987.

<sup>1</sup> Leibniz-Institute of Atmospheric Physics  
at the University of Rostock,  
Kühlungsborn, Germany,

<sup>2</sup> Институт прикладной физики РАН,  
г. Нижний Новгород, Россия

Поступила в редакцию  
6 апреля 2001 г.

## AUTOSOLITONS IN SYSTEMS WITH GLOBAL NONLINEARITY

*G. Sonnemann and V. E. Semenov*

We consider the simplest mathematical model of localized dissipative structures described by a single diffusive equation with the source containing both local and global nonlinearities, i.e., depending on an integral of the unknown function over the entire volume. On the basis of qualitative analysis, we find out that stable autosolitons exist in the medium considered. This conclusion is confirmed by the results of numerical simulation of a three-dimensional problem. We determine the dependence of autosoliton power on the parameters of the problem. Metastable multisoliton states are found and their properties are studied. The possibility of stabilization of three-dimensional autosolitons in a system without local losses is considered.

УДК 621.039

# ОБ ЭФФЕКТЕ ЭЛЕКТРОМАГНИТНОЙ ИНДУЦИРОВАННОЙ ПРОЗРАЧНОСТИ В КЛАССИЧЕСКИХ СИСТЕМАХ

*А. Г. Литвак, М. Д. Токман*

Построена классическая модель весьма популярного в последнее время параметрического эффекта электромагнитной индуцированной прозрачности (EIT) — формирования «окна прозрачности» внутри линии резонансного поглощения квантовой трёхуровневой системы, одновременно сопровождаемого рекордно сильным замедлением сигнальной световой волны. На основе этой модели рассмотрен эффект EIT для электромагнитных волн на частотах электронно-циклотронного резонанса в холодной плазме. Исследовано параметрическое (трёхвольновое) взаимодействие двух электромагнитных мод (частота одной из волн совпадает с гирочастотой электронов) с электростатической модой. Показано, что резонансная раскачка колебаний электронов на гирочастоте может демпфироваться за счёт параметрической связи с коллективными электростатическими колебаниями. Как и в аналогичной квантовой системе, в окне прозрачности в данном случае имеет место групповое замедление сигнальной электронно-циклотронной волны.

## ВВЕДЕНИЕ

Одним из перспективных направлений развития современной физики лазеров и нелинейной оптики является исследование процессов параметрического взаимодействия электромагнитных волн со средой, в которой возбуждена т. н. квантовая когерентность.<sup>1</sup> В рамках этого направления были предсказаны и реализованы такие интересные эффекты, как генерация стимулированного излучения в неинвертированной среде и электромагнитная индуцированная прозрачность<sup>2</sup> [1–3].

Первоначально данные эффекты исследовались в ансамблях квантовых осцилляторов: их реализация связывалась с интерференцией волновых функций «квантованных» электронов. Что касается эффекта EIT, то в настоящее время общепризнана перспективность его использования в задачах спектроскопии, формирования и коррекции пространственно-временной структуры лазерного излучения [3]; с точки зрения задач нелинейной оптики дисперсия электромагнитных волн в индуцированном окне прозрачности обладает чрезвычайно важными и интересными особенностями: групповая скорость волн может уменьшаться до нескольких метров в секунду при фазовой скорости, близкой к скорости света в вакууме, в режиме EIT существенно понижены мощностные пороги большинства нелинейнооптических эффектов [3, 4].

В последнее время проявляется заметный интерес к поиску классических аналогов квантового «фазониума». Этот интерес обусловлен как стремлением к более глубокому пониманию физики когерентных радиационных процессов, так и надеждами перенести новые идеи физики лазеров в классическую электронику СВЧ и физику плазмы. Если возможности реализации безынверсной генерации в классических системах уже теоретически исследовались в ряде работ (см. [5–8]), то теория электромагнитной индуцированной прозрачности в ансамблях классических электронов в настоящее время находится в самом начале своего становления. Здесь можно отметить работы [9–11], где рассматривается распространение в изотропной плазме волны с частотой меньше плазменной за счёт параметрического взаимодействия с мощной волной накачки. Этот пример, однако, лишь с известной натяжкой

<sup>1</sup> Для обозначения среды в таком состоянии часто используется термин «фазониум».

<sup>2</sup> Имея в виду формирование окон прозрачности именно в режиме параметрического взаимодействия волн, мы будем использовать общепринятую для этого эффекта аббревиатуру EIT (см. [3]).

может считаться классическим аналогом соответствующего квантового эффекта, т. к. обычно с ЕІТ связывают появление окна прозрачности внутри линии резонансного поглощения [3].

В данной работе рассматриваются прямые классические аналоги квантового эффекта электромагнитной индуцированной прозрачности. План статьи следующий. В разделе 1 кратко обсуждается реализация электромагнитной индуцированной прозрачности в квантовой трёхуровневой системе<sup>3</sup> и строится её простейший классический аналог — система связанных осцилляторов с параметрической связью. Показано, что классическим аналогом эффекта интерференции квантовых состояний трёхуровневой системы в режиме ЕІТ является один из вариантов хорошо известного в теории колебаний эффекта динамического демпфирования [12]. В разделе 2 развита теория электромагнитной индуцированной прозрачности в области электронно-циклotronного резонанса для электромагнитных волн в плазме. Раздел 3 содержит обсуждение полученных результатов и заключение.

## 1. «ПАРАМЕТРИЧЕСКОЕ» ДИНАМИЧЕСКОЕ ДЕМПФИРОВАНИЕ — КЛАССИЧЕСКИЙ АНАЛОГ ЭФФЕКТА ЭЛЕКТРОМАГНИТНОЙ ИНДУЦИРОВАННОЙ ПРОЗРАЧНОСТИ В АНСАМБЛЯХ КВАНТОВЫХ ОСЦИЛЛЯТОРОВ

Рассмотрим квантовую трёхуровневую систему (см. рис. 1) с собственными частотами переходов  $\Omega_1$ ,  $\Omega_2$  и  $\Omega_L = \Omega_1 - \Omega_2$ . Будем описывать её матрицей плотности [13]

$$\rho_{ij} = \begin{pmatrix} N_1 & \rho_L^* & \rho_1^* \\ \rho_L & N_2 & \rho_2^* \\ \rho_1 & \rho_2 & N_3 \end{pmatrix},$$

где  $N_1 = \rho_{11}$ ,  $N_2 = \rho_{22}$  и  $N_3 = \rho_{33}$  — населённости уровней, а недиагональные элементы  $\rho_1 = \rho_{31}$ ,  $\rho_2 = \rho_{32}$  и  $\rho_L = \rho_{21}$  — т. н. квантовые когерентности — характеризуют амплитуду и фазу колебаний системы на переходах  $3 \rightarrow 1$ ,  $3 \rightarrow 2$  и  $2 \rightarrow 1$  соответственно и определяют, в частности, электродипольный момент  $\mathbf{P}$  [13]:

$$\mathbf{P} = 2 \operatorname{Re}(\mathbf{d}_1^* \rho_1 + \mathbf{d}_2^* \rho_2 + \mathbf{d}_L^* \rho_L). \quad (1)$$

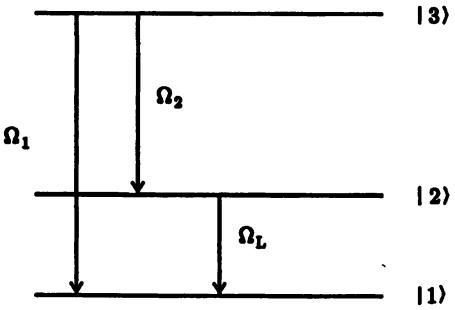


Рис. 1. Квантовая трёхуровневая система

Здесь  $\mathbf{d}_1 = \mathbf{d}_{31}$ ,  $\mathbf{d}_2 = \mathbf{d}_{32}$  и  $\mathbf{d}_L = \mathbf{d}_{21}$  — матричные элементы оператора дипольного момента для переходов  $3 \rightarrow 1$ ,  $3 \rightarrow 2$  и  $2 \rightarrow 1$  соответственно.

Исследуем взаимодействие данной системы с бихроматическим полем

$$\mathbf{E}(t) = \mathbf{x}_0 \operatorname{Re}(E_1 e^{-i\omega_1 t} + E_2 e^{-i\omega_2 t}), \quad (2)$$

где  $\mathbf{x}_0$  — единичный вектор поляризации, частоты  $\omega_1$  и  $\omega_2$  близки к частотам переходов  $\Omega_1$  и  $\Omega_2$ , а частота биений  $\omega_L = \omega_1 - \omega_2$ , следовательно, близка к частоте комбинационного перехода  $\Omega_L$ :

$$\left| \frac{\omega_{1,2,L} - \Omega_{1,2,L}}{\Omega_{1,2,L}} \right| \ll 1. \quad (3)$$

Определим линейный отклик рассматриваемой системы на частоте  $\omega_1$  в условиях воздействия накачки на частоте  $\omega_2$ . В этой ситуации распределение населённостей  $N_{1,2,3}$  определяется, вообще говоря, мощностью накачки и релаксационными процессами<sup>4</sup> на основе соответствующих уравнений

<sup>3</sup> Теория ЕІТ в квантовых системах рассматривалась в большом числе работ (см. обзор [3] и цитируемую там литературу); мы остановимся на ней лишь в той мере, в какой это полезно для перехода к классическому аналогу.

<sup>4</sup> Речь идёт о т. н. «продольной» релаксации [13].

баланса [13]. Мы, однако, будем для простоты полагать величины  $N_{1,2,3}$  параметрами задачи — это приближение вполне достаточно для анализа принципиальных (и наиболее интересных) особенностей рассматриваемых процессов.

Для определения недиагональных элементов матрицы плотности (когерентностей) мы воспользуемся известными уравнениями [13], в которых, учитывая условия синхронизма (3), сохраним только «резонансные» члены, т. е. используем метод укороченных уравнений<sup>5</sup> [15]. В результате имеем следующую систему уравнений:

$$\begin{aligned}\dot{\sigma}_1 + i(\Omega_1 - \omega_1 - i\gamma_1)\sigma_1 &= i(\alpha_1(N_1 - N_3) + \alpha_2\sigma_L), \\ \dot{\sigma}_L + i(\Omega_L - \omega_L - i\gamma_L)\sigma_L &= i(\alpha_2^*\sigma_1 - \alpha_1\sigma_2^*), \\ \dot{\sigma}_2 + i(\Omega_2 - \omega_2 - i\gamma_2)\sigma_2 &= i(\alpha_2(N_1 - N_2) + \alpha_1\sigma_L^*),\end{aligned}\quad (4)$$

где  $\sigma_{1,2,L}$  — комплексные амплитуды квантовых когерентностей:  $\rho_{1,2,L} = \sigma_{1,2,L} \exp(-i\omega_{1,2,L}t)$ ,  $\gamma_{1,2,L}$  — феноменологические релаксационные константы соответствующих переходов (обратные времена т. н. «поперечной» релаксации [13]),  $\alpha_{1,2} = d_{1,2}E_{1,2}/(2\hbar)$  — нормированные амплитуды ВЧ полей<sup>6</sup>,  $d_{1,2}$  — матричные элементы проекции оператора дипольного момента на орт  $\mathbf{x}_0$  для переходов  $3 \rightarrow 1$  и  $3 \rightarrow 2$ ,  $\hbar$  — постоянная Планка.

В линейном по  $\alpha_1$  приближении получаем следующее стационарное решение для когерентности на переходе  $3 \rightarrow 1$ :

$$\sigma_1 = \alpha_1(N_1 - N_3)\Phi(\omega_1, \omega_L, |\alpha_2|^2), \quad (5a)$$

где

$$\Phi = \frac{(\Omega_L - \omega_L - i\gamma_L) - \frac{(N_2 - N_3)|\alpha_2|^2}{(N_1 - N_3)(\Omega_2 - \omega_2 + i\gamma_2)}}{(\Omega_1 - \omega_1 - i\gamma_1)(\Omega_L - \omega_L - i\gamma_L) - |\alpha_2|^2}. \quad (5b)$$

Рассматривая среду из идентичных (для простоты — невзаимодействующих) трёхуровневых атомов, при помощи соотношений (5) и (1) можно перейти к выражению для макроскопической поляризуемости или диэлектрической проницаемости среды (см. [13]). Диэлектрическая проницаемость на частоте  $\omega_1$  (нелинейная по полю накачки на частоте  $\omega_2$ ) при этом равна

$$\epsilon = 1 + \omega_c \Phi(\omega_1, \omega_L, |\alpha_2|^2). \quad (6)$$

Здесь  $\omega_c = 4\pi|d_1|^2(N_1 - N_3)/\hbar$ , населённости  $N_{1,2,3}$  удовлетворяют нормировке  $N_1 + N_2 + N_3 = N$ , где  $N$  — концентрация атомов в среде.

Модификация линейной восприимчивости среды под действием накачки (т. е. поля  $E_2$ ) определяется величиной безразмерного параметра нелинейности  $\xi = |\alpha_2|^2/(\gamma_1\gamma_L)$ . В пределе  $\xi \rightarrow 0$  из (5б) следует стандартное выражение, определяющее линейную диэлектрическую восприимчивость в окрестности резонанса [13]:

$$\Phi \rightarrow \Phi_1 = \frac{1}{\Omega_1 - \omega_1 - i\gamma_1}. \quad (7)$$

Эффект электромагнитной индуцированной прозрачности наиболее наглядно проявляется в пределе  $N_{2,3}/N_1 \rightarrow 0$  (при этом дробью в числителе выражения (5б) можно пренебречь). В этом случае для частотной зависимости диэлектрической восприимчивости среды получаем выражение

$$\Phi \rightarrow \Phi_{\text{ЕИТ}} = \frac{\Omega_L - \omega_L - i\gamma_L}{(\Omega_1 - \omega_1 - i\gamma_1)(\Omega_L - \omega_L - i\gamma_L) - |\alpha_2|^2}. \quad (8)$$

<sup>5</sup> В квантовой электронике в этом смысле используют также термин приближение врачающейся волны [14].

<sup>6</sup> Величины  $\alpha_{1,2}$  определяют т. н. частоты Раби [13]:  $4|\alpha_{1,2}|^2 = \Omega_{R1,2}^2$ . Помимо условий синхронизма (3), условием применимости системы (4) является требование малости частот Раби по сравнению с частотами переходов между уровнями.

Сопоставляя выражения (7) и (8), можно убедиться, что в пределе сильной нелинейности, когда  $\xi \gg 1$ , линейное поглощение сигнальной волны в центре резонансной линии может быть подавлено в полосе частот порядка частоты Раби<sup>7</sup>:

$$\text{Im}(\Phi_{\text{EIT}}) \approx \text{Im}(\Phi_1)/\xi \quad (9)$$

при  $|\omega_{1,L} - \Omega_{1,L}| < \Omega_{R,2}/2 = |\alpha_2|$ .

Подробное обсуждение формы линии в режиме EIT и особенностей дисперсии волн внутри соответствующего нелинейного окна прозрачности (об этих особенностях мы упомянули во введении) можно найти, например, в обзоре [3] и цитируемой там литературе. Нашей же целью является переход к классическому аналогу рассмотренной выше трёхуровневой квантовой системы. В этой связи отметим, что система (4) соответствует, вообще говоря, уравнениям, описывающим возбуждение системы классических связанных осцилляторов внешним резонансным воздействием. Действительно, хотя в левые части стандартных уравнений классических осцилляторов входят вторые производные обобщённых координат, в рамках применимости укороченных уравнений (т. е. когда собственные частоты системы близки к частотам не слишком сильных внешних воздействий) уравнения осцилляторов обычно сводятся к уравнениям первого порядка для комплексных амплитуд колебаний соответствующих величин (см., например, [16]). С точки зрения реализации режима EIT важнейшей особенностью системы (4) является то обстоятельство, что в ней внешние источники не только непосредственно возбуждают «осцилляторы», но и обеспечивают параметрические связи между ними: действительно, в (4) переменные  $\sigma(j = 1, 2, L)$  входят в эволюционные уравнения для  $\sigma_{i \neq j}$  только в билинейных комбинациях с  $\alpha_{1,2}$ .

Выражению для формы линии в режиме EIT (8) соответствует приближение, в котором в системе (4) можно оставить лишь два первых уравнения (для когерентностей  $\sigma_{1,L}$ ), положив  $\sigma_2 = 0$  в уравнении для  $\sigma_L$ .<sup>8</sup> При этом в правой части уравнения для когерентности  $\sigma_1$  (в первом уравнении системы (4)) «параметрическое» вынуждающее слагаемое  $i\sigma_L\alpha_2$  компенсирует «линейный источник»  $i\alpha_1(N_1 - N_3)$ , подавляя возбуждение квантовой когерентности на переходе  $3 \rightarrow 1$ .

Аналогичная классическая колебательная система приведена на рис. 2. Она представляет собой два колебательных контура с собственными частотами  $\Omega_{1,L} = (L_{1,L}C_{1,L})^{-1/2}$  и сопротивлениями  $R_{1,L}$ , связанных через параметрическую взаимную индуктивность  $M = \text{Re}[\mu_2 \exp(-i\omega_2 t)]$ . При наличии действующей на первый контур сторонней ЭДС  $U = \text{Re}[i_1 \exp(-i\omega_1 t)]$ , при выполнении резонансных условий  $|\omega_{1,L} - \Omega_{1,L}| \ll \Omega_{1,L}$  и  $\gamma_{1,L} \ll \Omega_{1,L}$  (здесь  $\omega_L = \omega_1 - \omega_2$  — комбинационная частота,  $\gamma_{1,L} = R_{1,L}/(2L_{1,L})$  — релаксационные константы контуров), можно получить следующие укороченные уравнения для зарядов  $q_{1,L}$  на ёмкостях:

$$\begin{aligned} \dot{\sigma}_{1,L}^q + i(\Omega_1 - \omega_1 - i\gamma_1)\sigma_{1,L}^q &= i(\nu_1/(2L_1\omega_1) + \sigma_L^q\omega_L\mu_2/(4L_1)), \\ \dot{\sigma}_L^q + i(\Omega_L - \omega_L - i\gamma_L)\sigma_L^q &= i\sigma_1^q\omega_1\mu_2^*/(4L_L), \end{aligned} \quad (10)$$

где  $\sigma_{1,L}^q$  — комплексные амплитуды колебаний зарядов  $q_{1,L} = \text{Re}[\sigma_{1,L}^q \exp(-i\omega_{1,L}t)]$ . Очевидно, что система (10) в точности соответствует системе уравнений для элементов матрицы плотности (в пределе  $\sigma_2 = 0$ ). При этом гармонически изменяющаяся связь между контурами играет роль накачки на частоте  $\omega_2$  в квантовой системе. Из (10) следует выражение для комплексной амплитуды заряда, совершенно

<sup>7</sup> В часто рассматриваемом модельном приближении  $\gamma_L \rightarrow 0$  поглощение в центре линии вообще отсутствует (см. [3]); в этом случае, однако, система может быть неустойчивой даже при исчезающе малых насыщенностях верхних уровней. Действительно, в области параметров  $\xi > (N_1 - N_3)\gamma_2/[(N_2 - N_3)\gamma_1]$  знак мнимой части диэлектрической проницаемости  $\epsilon$  при  $\omega_1 - \Omega_1 = \omega_L - \Omega_L = 0$  соответствует неустойчивости пробной волны на частоте  $\omega_1$  даже в отсутствие инверсии, когда  $N_1 > N_2 > N_3$ . На этом эффекте основана идея т. н.  $P$ -схемы «безынверсного» лазера [2].

<sup>8</sup> При  $N_{2,3}/N_1 \rightarrow 0$  учёт конечной величины  $\sigma_2 \neq 0$  приводит к появлению лишь квадратичных по  $|\alpha_1|$  членов в итоговом выражении для  $\Phi_{\text{EIT}}$ .

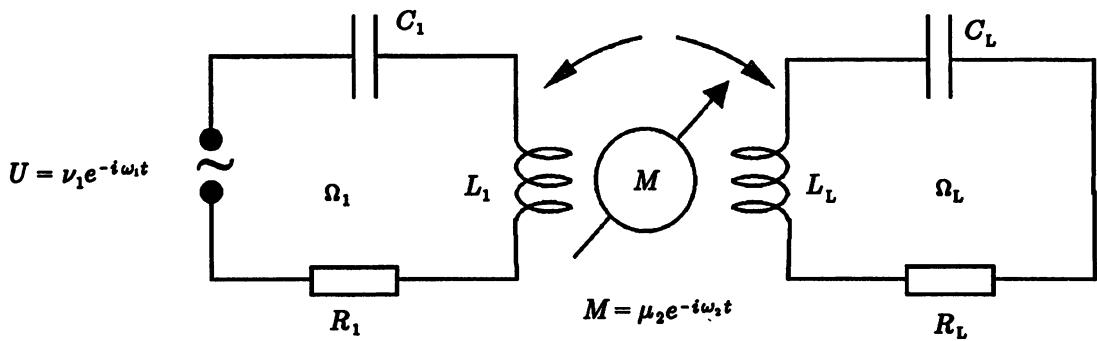


Рис. 2. Эквивалентная колебательная система с сосредоточенными параметрами

аналогичное выражению для квантовой когерентности ВЧ перехода  $3 \rightarrow 1$  в режиме EIT (ср. с (5а) и (8)):

$$\sigma_1^q = \frac{v_1}{2L_1\omega_1} \Phi_{\text{EIT}}(\omega_1, \omega_L, |\alpha_2|_{\text{eff}}^2), \quad (11)$$

где  $|\alpha_2|_{\text{eff}}^2 = \omega_1 \omega_L |\mu_2|^2 / (16 L_1 L_L)$ ; определяющая форму линии функция  $\Phi_{\text{EIT}}$  определена в (8). Как и в квантовой системе, резонансная раскачка колебаний первого колебательного контура подавлена в полосе частот  $|\omega_{1,L} - \Omega_{1,L}| < |\alpha_2|_{\text{eff}}$  при условии  $\xi_{\text{eff}} = |\alpha_2|_{\text{eff}}^2 / (\gamma_1 \gamma_L) \gg 1$ .

Отметим, что сам по себе эффект подавления резонансной раскачки осциллятора за счёт связи с другой колебательной системой хорошо известен: это т. н. динамическое демпфирование (см., например, [12]). Однако в своём стандартном варианте этот эффект обеспечивается обычной линейной связью двух осцилляторов, что приводит к возбуждению демпфирующей системы на частоте вынуждающей обобщённой силы. В рассматриваемом нами случае «параметрическая» связь осцилляторов приводит к раскачке демпфирующей системы на комбинационной частоте. В этом основная особенность рассматриваемого эффекта.

Таким образом, распространение волны с частотой  $\omega_1$  в режиме электромагнитной индуцированной прозрачности с точки зрения теории колебаний обусловлено описанным выше «параметрическим» динамическим демпфированием резонансного возбуждения за счёт раскачки колебаний демпфирующих осцилляторов на комбинационной частоте  $\omega_L = \omega_1 - \omega_2$ .

В заключение этого раздела отметим, что в квантовой системе возбуждение когерентности на «демпфирующем» переходе  $2 \rightarrow 1$  (рис. 1) вовсе не обязательно должно приводить к раскачке колебаний поляризации среды на комбинационной частоте: с целью минимизации диссипативного параметра  $\gamma_L$  режим EIT обычно реализуют в системе уровней с запрещённым электродипольным переходом ( $d_L = 0$ ) [3]. Можно, однако, убедиться, что несмотря на это обстоятельство при режиме EIT в трёхуровневой системе на переходе  $2 \rightarrow 1$  запасается энергия за счёт роста населённости состояния  $|2\rangle$ . Мы не приводим здесь соответствующий расчёт (он требует анализа системы (4) с учётом квадратичных по  $\alpha_1$  поправок<sup>9</sup>), ограничиваясь более простыми соображениями. Действительно, пусть в начальный момент времени отлична от нуля только населённость нижнего уровня, т. е.  $N_1 = N, N_{2,3} = 0$ . При установлении режима EIT на переходе  $2 \rightarrow 1$  возбуждается когерентность  $\rho_L \sim \alpha_1$ . Из свойств матрицы плотности (см. [13]) имеем  $N_2 N_1 \geq |\rho_L|^2$ , откуда следует, что населённость второго уровня должна стать отличной от нуля.

<sup>9</sup> Населённость второго уровня (состояния  $|2\rangle$ ) возрастает как  $|\alpha_1|^2$ , а верхнего уровня (состояния  $|3\rangle$ ) — только как  $\xi^{-2} |\alpha_1|^2$  (напомним, что здесь  $\xi \gg 1$ ).

## 2. ЭЛЕКТРОМАГНИТНАЯ ИНДУЦИРОВАННАЯ ПРОЗРАЧНОСТЬ В ЗОНЕ ЦИКЛОТРОННОГО РЕЗОНАНСА В МАГНИТОАКТИВНОЙ ПЛАЗМЕ

Эффект EIT в ансамбле квантовых осцилляторов, как уже отмечалось, обусловлен параметрическим взаимодействием колебаний на резонансных и комбинационных переходах. В плазме существует возможность подавления резонансного взаимодействия волн с частицами за счёт параметрического возбуждения коллективных степеней свободы. В качестве простого примера мы рассмотрим гидродинамическую теорию EIT для электронно-циклотронного резонанса в холодной магнитоактивной плазме. Пусть в плазме вдоль постоянного магнитного поля  $\mathbf{H} = z_0 H$  распространяются две циркулярно поляризованные волны с электрическим полем, вращающимся по направлению циклотронного вращения электронов (необыкновенные волны):

$$\mathbf{E}_\perp(z, t) = \operatorname{Re}[\mathbf{e}_+ (E_1 e^{-i\omega_1 t + ik_1 z} + E_2 e^{-i\omega_2 t + ik_2 z})], \quad (12)$$

где  $\mathbf{e}_+ = (\mathbf{x}_0 + i\mathbf{y}_0)/\sqrt{2}$  — вектор поляризации волн,  $\mathbf{z}_0$ ,  $\mathbf{x}_0$ ,  $\mathbf{y}_0$  — орты декартовых осей. Колебания поперечной и продольной (относительно постоянного магнитного поля) компонент скорости электронов определяются уравнениями Эйлера, где мы учтём действие силы Лоренца со стороны волновых полей:

$$\frac{\partial \mathbf{v}_\perp}{\partial t} + \Omega_H [\mathbf{v}_\perp, \mathbf{z}_0] + \gamma \mathbf{v}_\perp + v_\parallel \frac{\partial \mathbf{v}_\perp}{\partial z} = -\frac{e}{m} \left( \mathbf{E}_\perp(z, t) + v_\parallel \frac{\partial}{\partial z} \int_{-\infty}^t \mathbf{E}_\perp(z, t') dt' \right), \quad (13)$$

$$\frac{\partial v_\parallel}{\partial t} + \gamma v_\parallel + v_\parallel \frac{\partial v_\parallel}{\partial z} = \frac{e}{m} \left( \left( \mathbf{v}_\perp, \frac{\partial}{\partial z} \int_{-\infty}^t \mathbf{E}_\perp(z, t') dt' \right) + \frac{\partial \varphi}{\partial z} \right). \quad (14)$$

Здесь  $\Omega_H = eH/(mc)$  — гирочастота,  $e$  и  $m$  — заряд и масса электрона соответственно,  $\gamma$  — эффективная частота соударений,  $\varphi$  — электростатический потенциал, определяемый уравнением Пуассона

$$\frac{\partial^2 \varphi}{\partial z^2} = 4\pi e (N_e - N_i), \quad (15)$$

где  $N_{i,e}$  — концентрации ионов и электронов соответственно. Учитывая уравнение непрерывности

$$\frac{\partial N_e}{\partial t} + \frac{\partial}{\partial z} (v_\parallel N_e) = 0, \quad (16)$$

в рамках предположений о постоянстве концентрации ионов и квазинейтральности:

$$N_i = \operatorname{const} = N, \quad n = N_e - N_i \ll N_i,$$

из (14)–(16) получаем уравнения

$$\frac{\partial^2 n}{\partial t^2} + \Omega_p^2 n + \gamma \frac{\partial n}{\partial t} = -\frac{e N_e}{m} \frac{\partial}{\partial z} \left( \mathbf{v}_\perp, \frac{\partial}{\partial z} \int_{-\infty}^t \mathbf{E}_\perp(z, t') dt' \right), \quad (17)$$

$$v_\parallel = -\frac{1}{N} \int_{-\infty}^z \frac{\partial n}{\partial t} dz', \quad (18)$$

описывающие возбуждение продольных коллективных колебаний пондеромоторной силой, обусловленной ВЧ полем. Здесь  $\Omega_p = \sqrt{4\pi e^2 N/m}$  — электронная плазменная частота.

Будем интересоваться случаем, когда для волны на частоте  $\omega_1$  выполнено условие циклотронного синхронизма с электронами:

$$\{|\omega_1 - \Omega_H|, \gamma\} \ll \Omega_H. \quad (19)$$

Биения электромагнитных волн наиболее эффективно возбуждают плазменные колебания на комбинированной частоте  $\omega_L = \omega_1 - \omega_2$  при соответствующих условиях синхронизма:

$$\{|\omega_L - \Omega_p|, \gamma\} \ll \Omega_p. \quad (20)$$

Вводя комплексные амплитуды колебаний соответствующих величин

$$v_{\perp} = \operatorname{Re} \left( e_+ \sum_{j=1,2} \hat{v}_j e^{ik_j z - i\omega_j t} \right), \quad v_{||} = \operatorname{Re} (\hat{v}_{||} e^{i(k_1 - k_2)z - i\omega_L t}), \quad n = \operatorname{Re} (\hat{n} e^{i(k_1 - k_2)z - i\omega_L t}),$$

в линейном по полю  $E_1$  приближении получим из (12), (13), (17), (18) систему укороченных уравнений, которая имеет следующее стационарное решение:

$$\hat{v}_2 = \frac{eE_2}{imz_2}, \quad (21)$$

$$iz_1 \hat{v}_1 = \frac{eE_1}{m} + \frac{i\hat{v}_{||} k_2 U_2}{2}, \quad (22)$$

$$U_2 = \hat{v}_2 + \frac{ieE_2}{m\omega_2} = \frac{eE_2 (\Omega_H - i\gamma)}{imz_2 \omega_2}, \quad (23)$$

$$\hat{\eta} = \frac{k_1 - k_2}{\omega_L} \hat{v}_{||} N_e, \quad (24)$$

$$\hat{v}_{||} = -\frac{ie\omega_L}{2mz_p} \left( \frac{\hat{v}_1 E_2^* k_2}{\omega_2} + \frac{ieE_2^* E_1 k_1}{m\omega_1 z_2^*} \right), \quad (25)$$

где  $z_{1,2} = \omega_{1,2} + i\gamma - \Omega_H$ ,  $z_p = \omega_L^2 + i\omega_L \gamma - \Omega_p^2$ .

Из (21)–(25) получаем выражение для амплитуды осцилляций скорости на циклотронной частоте:

$$\hat{v}_1 = \frac{eE_1}{m} \frac{z_p + \frac{2k_1 \omega_2}{\omega_1 k_2} (\Omega_H - i\gamma) \omega_L \xi_{EC}}{i(z_1 z_2 + 2z_2^* (\Omega_H - i\gamma) \omega_L \xi_{EC})}, \quad (26)$$

где  $\xi_{EC} = |eE_2/(2mz_2)|^2 (k_2/\omega_2)^2 = |\hat{v}_2|^2/(2v_{ph})^2$  — параметр нелинейности, определяемый отношением квадратов осцилляторной и фазовой скоростей для поля накачки  $E_2 \exp(-i\omega_2 t)$ .

Перейдём теперь к выражению для электрического тока на частоте  $\omega_1$ . Учитывая колебания электронной плотности, получаем

$$\mathbf{j}_{\omega_1} = \operatorname{Re} (e_+ \hat{j}_1 e^{-i\omega_1 t + ik_1 z}),$$

где

$$\hat{j}_1 = -e(N\hat{v}_1 + \hat{\eta}\hat{v}_2/2). \quad (27)$$

Выражение (27) для комплексной амплитуды тока определяет показатель преломления для распространяющейся вдоль магнитного поля волны на частоте  $\omega_1$ :

$$N_{EIT}^2 = 1 - \frac{4\pi\hat{j}_1}{i\omega_1 E_1}. \quad (28)$$

Пренебрегая (там, где это корректно) малыми величинами порядка  $\gamma\xi_{\text{EC}}/\Omega_H$  и  $z_1\xi_{\text{EC}}/\Omega_H$ , получаем

$$N_{\text{EIT}}^2 = 1 - v \frac{(\Delta^2 - v + is\Delta) + 2\xi_{\text{EC}}\Delta\sqrt{u} \left( \frac{2ck_1}{\omega_1 N_2} - 1 + \Delta \right)}{(1 - \sqrt{u} + is)(\Delta^2 - v + is\Delta) - 2\xi_{\text{EC}}\Delta^2\sqrt{u}}, \quad (29)$$

где  $v = (\Omega_p/\omega_1)^2$ ,  $u = (\Omega_H/\omega_1)^2$ ,  $s = \gamma/\omega_1$ ,  $\Delta = \omega_L/\omega_1$ ,  $N_2 = ck_2/\omega_2$ .

В пределе исчезающей слабой накачки на частоте  $\omega_2$ , когда

$$\xi_{\text{EC}} \ll s^2/\Delta,$$

из (29) следует стандартное выражение для показателя преломления необыкновенной волны при продольном распространении относительно магнитного поля [17]:

$$N_1^2 = 1 - \frac{v}{1 + is - \sqrt{u}}. \quad (30)$$

В противоположном предельном случае

$$\xi_{\text{EC}} \gg s^2/\Delta \quad (31)$$

циклотронное поглощение существенно уменьшается по сравнению с линейным режимом в частотной полосе, соответствующей условию

$$|(1 - \sqrt{u} + is)(\Delta^2 - v + is\Delta)| \leq \xi_{\text{EC}}\Delta^2 \quad (32a)$$

(в этом случае получаем  $\text{Im}(N_{\text{EIT}}^2) \approx \text{Im}(N_1^2)[s^2/(\xi_{\text{EC}}\Delta)] \ll \text{Im}(N_1^2)$ ).

Что касается показателя преломления, то он порядка единицы, если вместе с условием (32a) выполнено неравенство

$$v|(\Delta^2 - v + is\Delta)| \ll \xi_{\text{EC}}\Delta^2. \quad (32b)$$

Как видно, имеет место полная аналогия с режимом EIT в ансамбле квантовых осцилляторов. Более того, в окрестности плазменного резонанса на биениях электромагнитных волн, когда в силу неравенства (20) можно использовать приближённое соотношение  $\omega_L^2 + i\omega_L\gamma - \Omega_p^2 \approx \omega_L(\omega_L + i\gamma/2 - \Omega_p)$ , в области параметров

$$s \gg \xi_{\text{EC}} \geq s^2/\Delta$$

из (29) следует выражение, точно соответствующее частотной зависимости комплексного показателя преломления для квантовой среды в режиме EIT:

$$N_{\text{EIT}}^2 \approx 1 + \frac{v(\sqrt{v} - \Delta - is/2)}{(\sqrt{v} - \Delta - is/2)(\sqrt{u} - 1 - is) - \xi_{\text{EC}}\Delta\sqrt{u}}. \quad (33)$$

### 3. ОСОБЕННОСТИ РАСПРОСТРАНЕНИЯ ЦИКЛОТРОННЫХ ВОЛН В УСЛОВИЯХ ЭЛЕКТРОМАГНИТНОЙ ИНДУЦИРОВАННОЙ ПРОЗРАЧНОСТИ

Полученное в предыдущем разделе выражение (29) для эффективного показателя преломления сигнальной волны позволяет исследовать особенности распространения излучения в окне индуцированной прозрачности.

В пределе слабой диссипации, когда  $|\text{Re}(N_{\text{EIT}}^2)| \gg |\text{Im}(N_{\text{EIT}}^2)|$ , действительное волновое число и групповая скорость в окне прозрачности могут быть определены из стандартных соотношений

$$c^2 k_1^2 \approx \omega_1^2 \text{Re}(N_{\text{EIT}}^2) \approx \omega_1^2 N_0^2, \quad (34)$$

$$V_{\text{gr}} = \frac{\partial \omega_1}{\partial k_1} = \frac{2cN_0 - \omega_1 (\partial N_0^2 / \partial k_1)}{2N_0^2 + \omega_1 (\partial N_0^2 / \partial \omega_1)}, \quad (35)$$

где  $N_0 = N_{\text{EIT}}|_{s \rightarrow 0}$ .

Строго говоря, в дисперсионном уравнении (34) вместо бесстолкновительного показателя преломления  $N_0^2$  можно использовать величину  $\text{Re}(N_{\text{EIT}}^2)$ , при этом учитывается влияние конечной величины  $s$  на дисперсию волн. Однако при определении групповой скорости в рамках стандартного соотношения  $\partial \omega / \partial k$  учёт соответствующих малых поправок в законе дисперсии не имеет смысла, т. к. не существует общего феноменологического выражения для групповой скорости волновых полей в диссипативных средах [17]. В этом смысле использование бесстолкновительного дисперсионного уравнения вполне оправдано.

Определяющие поглощение волн мнимые компоненты волнового вектора и частоты <sup>10</sup>  $\kappa_1$  и  $\Gamma_1$  для граничной (когда  $\Gamma_1 = 0$ ) и начальной (когда  $\kappa_1 = 0$ ) задач определяются выражениями [18]

$$\kappa_1 [2cN_0 - \omega_1 (\partial N_0^2 / \partial k_1)] \approx \omega_1 \text{Im}(N_{\text{EIT}}^2), \quad (36a)$$

$$\Gamma_1 [2N_0^2 + \omega_1 (\partial N_0^2 / \partial \omega_1)] \approx -\omega_1 \text{Im}(N_{\text{EIT}}^2). \quad (36b)$$

В области EIT (т. е. при условиях (32)) для умеренных значений  $N_{\text{EIT}}^2$  из (29) и (35), (36) получаем следующие приближённые соотношения:

$$V_{\text{gr}} \approx \frac{c\sqrt{u}\xi_{\text{EC}}}{\Delta} \left( N_{\text{EIT}} \frac{\Delta}{v} - \frac{1}{N_2} \right), \quad \Gamma_1 \approx -\frac{s\omega_1}{2}, \quad \kappa_1 \approx \frac{|\Gamma_1|}{V_{\text{gr}}} \quad (37)$$

(знак  $\kappa_1$  совпадает со знаком  $V_{\text{gr}}$ , т. е. волна всегда затухает в направлении распространения). Интересно отметить, что дисперсия циклотронных волн в режиме EIT имеет ряд особенностей, отсутствующих в квантовой трёхуровневой системе. Эти особенности связаны с тем фактом, что показатель преломления  $N_{\text{EIT}}$  зависит от волнового числа  $k_1$  (см. формулу (29)), т. е. несмотря на приближение холодной плазмы в этом режиме появляется пространственная дисперсия, индуцированная мощной накачкой. На рис. 3 показана структура дисперсионных ветвей при дополнительном упрощающем предположении  $\Omega_p + \omega_2 = \Omega_H$ , когда  $1 - \sqrt{u} = \Delta - \sqrt{v}$ , а на рис. 4 показан качественный вид зависимости  $N_{\text{EIT}}^2(v)|_{s \rightarrow 0} = N_0^2(v)$ .

Видно, что дисперсионные ветви оказываются асимметричными (приведённые на рис. 3, 4 кривые соответствуют случаю  $k_2 > 0$ ). Дисперсионная зависимость на рис. 3 имеет участки, соответствующие «обратным» волнам (когда направления фазовой и групповой скорости противоположны). Наконец, в отличие от квантового случая точка с нулевой групповой скоростью смешена с оси  $k_1 = 0$ , при этом данной точке соответствует ненулевой показатель преломления. В то же время при нулевом значении

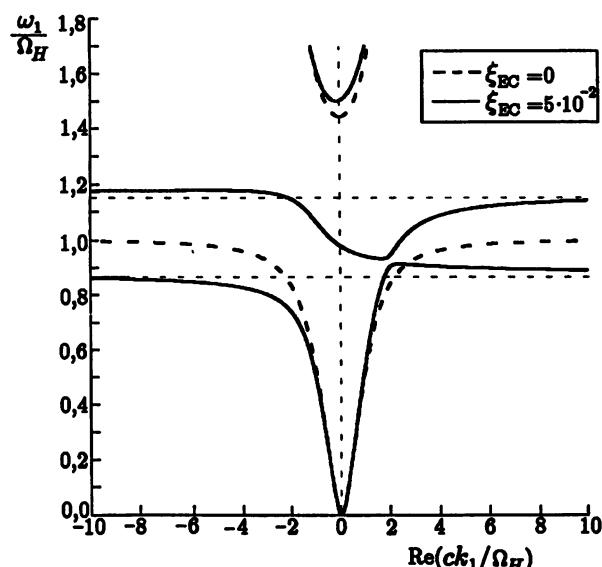


Рис. 3. Дисперсия электронно-циклотронной волны в условиях электромагнитной индуцированной прозрачности ( $v = 0,64$ ;  $\omega_2/\Omega_H = 0,2$ );  $s = 0$

<sup>10</sup> Напомним, что для принятой нами гармонической зависимости  $\exp(i k z - i \omega t)$  неравенство  $\kappa_1 > 0$  соответствует затуханию волны вдоль оси  $z$ , а временное затухание соответствует соотношению  $\Gamma_1 < 0$ .

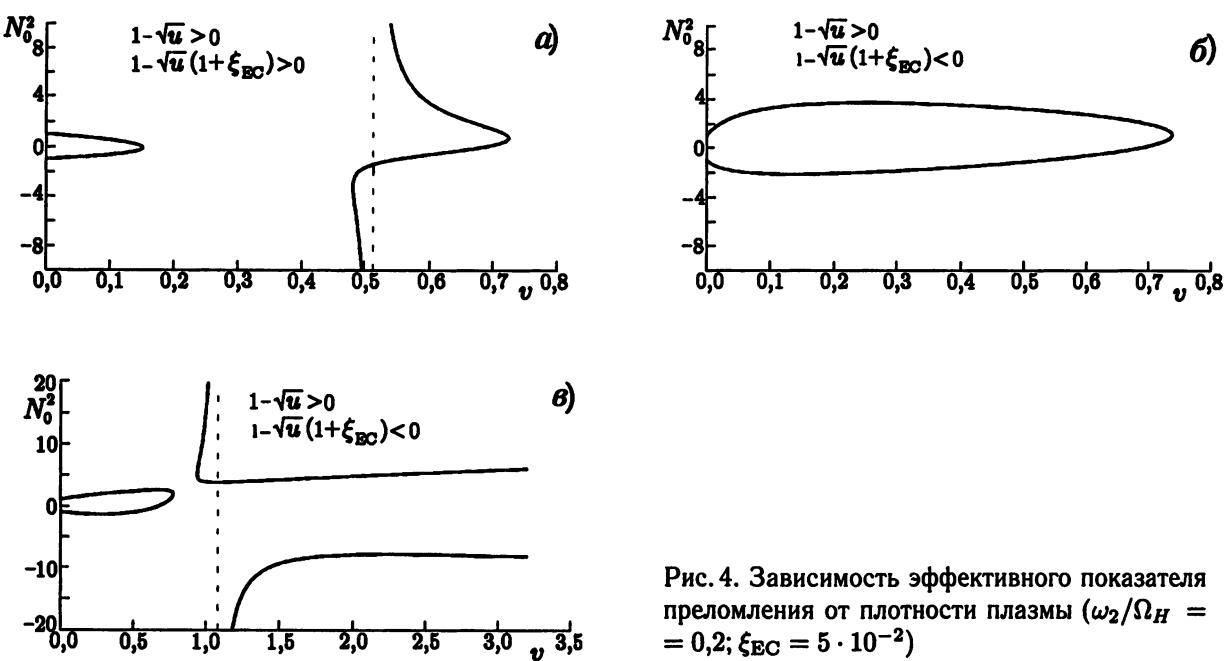


Рис. 4. Зависимость эффективного показателя преломления от плотности плазмы ( $\omega_2/\Omega_H = 0,2$ ;  $\xi_{EC} = 5 \cdot 10^{-2}$ )

показателя преломления имеется конечная групповая скорость.<sup>11</sup> Из рис. 3 и выражения (35) видно, что важнейшая особенность дисперсии волн в окне прозрачности EIT — значительное групповое замедление при фазовых скоростях порядка скорости света в вакууме (см. [3]) — имеет место и для циклотронных волн в плазме.

Обеспечение сильного группового замедления на фоне малого фазового замедления и слабого поглощения требует ограничений на релаксационную частоту  $\omega_1$ , а также предъявляет достаточно высокие требования к стабильности частот волн, концентрации плазмы и магнитного поля. Из (37) следует, что, хотя для относительного уменьшения линейного циклотронного поглощения интенсивность накачки должна удовлетворять условию (31), для достижения абсолютной малости поглощения на «вакуумной» длине волны необходимо выполнение более сильного, вообще говоря, неравенства

$$s \ll \xi_{EC}/v^{1/2}. \quad (38)$$

Что касается относительных частотных вариаций  $\eta = \delta/\omega_1$ , то для всех частот  $\omega_{1,2}$  и  $\Omega_{H,p}$  они не должны существенно превышать  $\xi_{EC}/v^{1/2}$ .

В заключение этого раздела рассмотрим следующий оценочный пример. При мощности излучения в 500 кВт, диаметре фокусировки пучка накачки около пяти вакуумных длин волн для  $(\Omega_H - \omega_2)/\omega_1 \approx \Delta \approx 0,1$  получаем (учитывая, что в зоне EIT  $\Delta \approx v^{1/2}$ ) допустимую относительную частотную расстройку  $\xi_{EC}/v^{1/2} \approx 10^{-3}$ . Для частоты сигнальной волны  $\omega_1/2\pi = 94$  ГГц величина  $v^{1/2} \approx 0,1$  соответствует плотности плазмы около  $1,5 \cdot 10^{12}$  см<sup>-3</sup>, при этом частота кулоновских соударений даже для низкой температуры электронов в 1 эВ соответствует параметру  $s \approx 10^{-4}$ , что обеспечивает выполнение неравенства (38). На рис. 5 проиллюстрировано формирование окна прозрачности внутри линии циклотронного поглощения для близких параметров. Отметим, что все использованные здесь параметры соответствуют техническим возможностям современных мазеров на циклотронном резонансе (см. [20]).

<sup>11</sup> Такой закон дисперсии оптимален для т. н. гиро-ЛБВ, т. к. при равном нулю показателе преломления циклотронное взаимодействие распространяющейся волны с частицами нечувствительно к питч-угловому разбросу электронов [19].

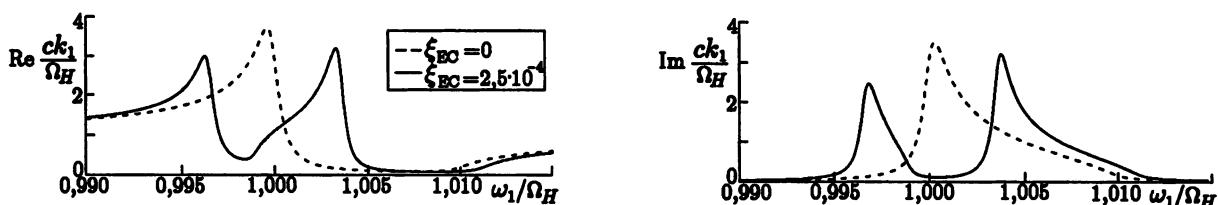


Рис. 5. Формирование окна прозрачности внутри линии циклотронного поглощения ( $v = 10^{-2}$ ;  $\omega_2/\Omega_H = 0,9$ ;  $s = 5 \cdot 10^{-4}$ )

#### 4. ОБСУЖДЕНИЕ И ЗАКЛЮЧЕНИЕ

Основным результатом данной работы является демонстрация прямых классических аналогов квантового эффекта EIT, связанного с интерференцией состояний квантовых осцилляторов, и основанное на исследовании этих аналогов объяснение механизма EIT с общефизической точки зрения, единой для квантовых и классических систем. Показано, в частности, что подавление резонансного поглощения в режиме EIT возможно, в принципе, при параметрическом взаимодействии бихроматического поля не только с индивидуальными степенями свободы электронов (осцилляторов), но и с их коллективными колебаниями.

Что касается возможных приложений EIT в ансамблях классических электронов, то этот эффект, в принципе, может быть использован для нелинейной транспортировки электронно-циклотронных волн в плазме, однако для разработки конкретных предложений необходимо обобщить развитую здесь теорию на случай непродольного распространения волн с учётом бесстолкновительного резонансного взаимодействия<sup>12</sup> волн и частиц.

Другая возможная область приложений связана с тем обстоятельством, что эффект подавления резонансного поглощения волн низкоэнергичными электронами (за счёт параметрического воздействия мощной накачки) может, по-видимому, обеспечить генерацию циклотронного излучения «горячими» электронами в той области параметров, где стандартный линейный инкремент является отрицательным. По существу, здесь идёт речь о переходе от режима EIT к безынверсной генерации в  $P$ -схеме.<sup>13</sup> Интересные перспективы продвижения в этом направлении могут быть связаны и с отмеченной выше особенностью дисперсии циклотронных волн в условиях EIT, позволяющей компенсировать доплеровское уширение линии циклотронного резонанса. Весьма интересным представляется также исследование этого эффекта на высших циклотронных гармониках, когда биения электромагнитных волн резонансы не плазменной частоте, а одной из циклотронных гармоник.

Авторы благодарны А. В. Гапонову-Грехову и В. А. Миронову за плодотворные дискуссии, А. Ю. Крячко за выполнение численных расчётов и М. А. Балакиной за помощь в оформлении работы.

Работа поддержана РФФИ (грант № 99–02–16230).

#### СПИСОК ЛИТЕРАТУРЫ

1. Kocharovskaya O. // Phys. Rep. 1992. V. 219. P. 175.

<sup>12</sup> Т. е. на случай неоднородно уширенной полосы резонансного взаимодействия.

<sup>13</sup> Как мы отмечали в разделе 1, EIT и генерация без инверсии в  $P$ -схеме являются родственными режимами. В отличие от т. н.  $\lambda$ -схемы лазера без инверсии, в  $P$ -схеме автоматически образуется необходимая для безынверсного лазера модуляция активной среды на комбинационной частоте (см. [1, 2]). Что касается возможности реализации классического лазера в рамках  $\lambda$ -схемы (т. е. с отдельной системой модуляции электронного потока), то этот вопрос подробно исследован в [7, 8].

2. Kocharovskaya O., Mandel P., Radeonychev Y. V. // Phys. Rep. 1992. V. 45. P. 1997.
3. Harris S. E. // Phys. Today. 1997. July. P. 36.
4. Harris S. E. // Phys. Rev. Lett. 1999. V. 82. P. 4611.
5. Nikonov D. E., Sherman B., Kurizki G., Scully M. O. // Opt. Commun. 1996. V. 123. P. 363.
6. Artemiev A. I., Fedorov M. V., Rostovtsev Y. V. et al. // Phys. Rev. Lett. 2000. V. 85. P. 4510.
7. Гапонов-Грехов А. В., Токман М. Д. // ЖЭТФ. 1997. Т. 112. С. 1176.
8. Ерухимова М. А., Токман М. Д. // ЖЭТФ. 2000. Т. 118. С. 291.
9. Harris S. E. // Phys. Rev. Lett. 1996. V. 77. P. 5357.
10. Gordon D. F., Mori W. B., Joshi C. // Phys. Plasmas. 2000. V. 7. P. 3145.
11. Gordon D. F., Mori W. B., Joshi C. // Phys. Plasmas. 2000. V. 7. P. 3156.
12. Мандельштам Л. И. Лекции по теории колебаний. М.: Наука, 1972.
13. Файн В. М. Фотоны и нелинейные среды. М.: Советское радио, 1972.
14. Аллен Л., Эберли Дж. Оптический резонанс и двухуровневые атомы. М.: Мир, 1978.
15. Андронов А. А., Витт А. А., Хайкин С. Э. Теория колебаний. М.: Наука, 1981.
16. Рабинович М. И., Трубецков Д. И. Введение в теорию колебаний. М.: Наука, 1984.
17. Гинзбург В. Л. Распространение электромагнитных волн в плазме. М.: Наука, 1967.
18. Гинзбург В. Л., Рухадзе А. А. Волны в магнитоактивной плазме. М.: Наука, 1970.
19. Bratman V. L., Gross A. W., Denisov G. G. et al. // Phys. Rev. Lett. 2000. V. 84. P. 2746.
20. Maysnikov V. E., Agapova M. V., Litvak A. G. et al. // Proc. of the International workshop "Strong Microwaves in plasmas", Nizhny Novgorod, 2–9 August, 1999. V. 2. P. 615.

Институт прикладной физики РАН,  
г. Нижний Новгород, Россия

Поступила в редакцию  
16 февраля 2001 г.

#### THE EFFECT OF ELECTROMAGNETIC INDUCED TRANSPARENCY IN CLASSICAL SYSTEMS

*A. G. Litvak and M. D. Tokman*

We develop a classical model of the recently popular parametric effect of electromagnetic induced transparency (EIT), i.e., the formation of a «transparency window» inside a resonance absorption line of a three-level quantum system, which is accompanied by a record strong slowing of the signal wave. Basing on this model, we consider the EIT effect for electromagnetic waves at frequencies of the electron-cyclotron resonance in a cold plasma. The parametric (three-wave) interaction of two electromagnetic modes (the frequency of one of these modes is equal to the electron gyrofrequency) with the electrostatic mode is considered. It is shown that the resonance growth of the electron oscillations at the gyrofrequency can be damped due to the parametric coupling with the collective electrostatic oscillations. Similar to the analogous quantum system, the group slowing of the signal electron-cyclotron wave takes place in the case considered.

УДК 551.594

## ФРАКТАЛЬНАЯ СТРУКТУРА НЕЛИНЕЙНОЙ ДИНАМИКИ ЭЛЕКТРИЧЕСКОГО ЗАРЯДА В ГРОЗОВОМ ОБЛАЦЕ

*Д. И. Иудин, В. Ю. Трахтенгерц*

Исследуется фрактальная динамика внутриоблачного пространственного электрического заряда на предварительной стадии грозового разряда, в процессе которой формируется так называемая дренажная система сбора заряда со всего объёма облака, служащая основой для зарождения лидерного канала молнии. Формирование дренажной системы обусловлено появлением электрических микроразрядов, связанных с нелинейной стадией развития внутриоблачной неустойчивости пучково-плазменного типа. Последняя моделируется на основе двумерной решётки конечных автоматов. Результаты численного моделирования свидетельствуют о принадлежности возникающей дренажной системы к семейству переколяционных кластеров. Указана согласующаяся с предложенной моделью область параметров, в которой грозовое облако обнаруживает поведение, соответствующее режиму самоорганизованной критичности. Предложен метод расчёта радиоизлучения от внутриоблачных микроразрядов, позволяющий проводить непосредственное сопоставление теоретической модели с экспериментальными данными.

### ВВЕДЕНИЕ

Важным объектом исследований в области нелинейной динамики являются процессы в сильно неравновесных открытых системах. Потоки энергии и вещества, проходящие через эти системы, обеспечивают возникновение в них эффектов самоорганизации — образования макроскопических диссипативных структур. Наличие источников и стоков энергии делает невозможной релаксацию этих систем к состоянию теплового равновесия. Даже в установившемся состоянии сохраняется поток энергии, проходящий через систему и диссилирующий в окружающей среде. Именно благодаря этому потоку система становится активной, т. е. приобретает способность к автономному образованию структур. С ростом интенсивности источников система удаляется от теплового равновесия и проходит последовательность неустойчивостей, ведущих к качественным перестройкам установившегося кинетического режима. Каждая такая перестройка представляет собой кинетический переход, сопровождающийся изменением внутренней структуры открытой сильно неравновесной системы.

За последние десятилетия накопилось много примеров физических и химических систем, демонстрирующих эффекты самоорганизации [1–3]. Классические примеры кинетических превращений даёт гидродинамика: конвективные течения в задаче Бенара, вихри Тейлора и т. п. Как и живая материя, такие системы могут функционировать лишь за счёт подвода к ним потока энергии и вещества. Аналогия с живыми системами отразилась и на соответствующей терминологии. Примером может служить кинетический переход, называемый переходом типа «заселения среды». Изначально относящийся к задачам анализа поведения биологических сообществ, этот тип кинетического перехода был впоследствии обнаружен в явлениях автокатализа, при ядерных реакциях расщепления, при цепных химических реакциях, электрических разрядах в различных средах, процессах переноса в плазменных системах и т. д. — практически во всех неравновесных системах, характеризующихся размножением и распадом [4]. Сценарий этого кинетического превращения выглядит следующим образом. Сначала в системе происходит развитие неустойчивости, приводящей к неограниченному во времени росту численности некой популяции. В случае физических систем речь идёт о неограниченном росте соответствующих физических полей: концентрации носителей заряда, плотности тока, объёмной плотности

*Д. И. Иудин, В. Ю. Трахтенгерц*

419

продуктов ядерных и химических реакций и т. п. Далее неустойчивость может быть подавлена нелинейным механизмом запорогового ограничения. В этом случае выше порога неустойчивости в среде устанавливается отличная от нуля плотность популяции, т. е. происходит заселение среды.

Традиционный подход к описанию кинетического перехода предполагает пространственно-временную локальность нелинейного механизма подавления роста населённости и основывается на аналитическом или численном решении нелинейных эволюционных уравнений, причём основными компонентами решения являются нелинейные периодические либо уединённые волны — автоловновые фронты, спиральные структуры и т. д. [5].

Менее изученной является ситуация, когда нелинейный механизм запорогового ограничения является существенно нелокальным в пространстве и (или) во времени. При этом поле диссипации в системе имеет чрезвычайно сложную стохастическую структуру, и ограничение роста охватывает пространственно-временные области самых различных конфигураций и размеров. Природа устроена таким образом, что очень часто эти сложные диссипативные структуры являются физическими фракталами, т. е. демонстрируют скейлинговые свойства в широком диапазоне пространственных и временных масштабов. Пространственно-временной скейлинг характеризуется сильными спадающими по степенно-му закону корреляциями, которые типичны для критических явлений. Поэтому качественно динамика подобных систем производит впечатление самоорганизованного и самонастраивающегося критического режима. Явления такого рода недавно были объединены общим наименованием — системы с самоорганизованной критичностью (self-organized criticality) [6, 7]. Название не только удачно соединяет в себе два перспективных направления современной физики — самоорганизацию и критические явления, но и отражает суть происходящего в системе.

В данной работе мы обратимся к конкретному примеру потоковой системы — грозовому облаку. Сильные конвективные движения являются той внешней силой, которая обеспечивает всё богатство внутриоблачной динамики от разнообразнейших механизмов электризации облака до феерической картины атмосферного искрового разряда. В недавней работе [8] авторы построили качественную модель развития молнии, основанную на эвристических соображениях, навеянных идеями фрактальной геометрии. В основе этой модели лежит двухмасштабная структура электрического поля в грозовом облаке, где наряду с крупномасштабным полем существуют мелкомасштабные электрические ячейки с размером  $l \sim 10$  м, много меньшим размера облака. Такие ячейки возникают вследствие внутриоблачной неустойчивости, обусловленной многопотоковым течением в восходящем конвективном потоке. Как мы полагаем, электрические разряды внутри этих ячеек и создают своеобразную дренажную динамическую систему, обеспечивающую сбор заряда со всего объёма облака и предопределяющую появление лидера канала молнии, по которому заряд стекает под действием крупномасштабного поля в основание облака. Предположение о том, что возникающая сеть микроразрядов является в каждый момент перколяционным проводящим кластером, позволило нам определить основные электрические макропараметры грозового облака и записать закон Ома для фрактальной электрической цепи, характеризующий рост тока с увеличением размера перколяционного кластера. С этих позиций формирование лидера канала молнии можно рассматривать как переход к направленной перколяции электрического тока во внешнем электрическом поле. Направленная перколяция приводит к накоплению электрического заряда в локальном объёме на нижней кромке облака и последующему пробою промежутка облако—земля.

Сейчас мы хотим более последовательно изложить суть фрактального подхода к анализу нелинейной внутриоблачной динамики и, опираясь на компьютерное моделирование, получить некоторые количественные характеристики внутриоблачных электрических разрядов. Мы сосредоточимся на анализе предварительной стадии молниевой вспышки. Эта стадия предшествует формированию лидера канала молнии и развитию так называемого обратного удара, который является наиболее мощным проявлением молниевого разряда. Согласно экспериментальным данным предварительная стадия длится

от десятков до сотен миллисекунд и характеризуется аномально высокой внутриоблачной активностью. Она состоит из многочисленных коротких (порядка 1 мкс) импульсов, объединённых в цепочки различной длительности. Восстановленные по этим импульсам радиоизображения демонстрируют сильно разветвлённую внутриоблачную сеть микроразрядов, напоминающую структуру фрактальных кластеров.

Предварительная стадия молниевого разряда может быть осмыслена как кинетический переход типа заселения среды. Полидисперсный ансамбль заряженных внутриоблачных частиц представляет собой специфическую плазменную среду, в которой конвективные потоки вызывают развитие неустойчивости, аналогичной пучковой неустойчивости в плазме. Краткое знакомство с особенностями развития пучковой неустойчивости составляет содержание первого раздела работы. Неустойчивость приводит к росту мелкомасштабного электрического поля, а механизм, ограничивающий этот рост, обусловлен микроразрядами между максимумами и минимумами возникающей электрической волны. В первом разделе показано, что при самых общих предположениях относительно характера мелкомасштабного разряда механизм подавления роста мелкомасштабной электрической структуры оказывается существенно нелокальным. Мы ограничим рассмотрение анализом модельной ситуации, в которой отдельный мелкомасштабный пробой способен спровоцировать «эпидемию» внутриоблачных разрядов. При этом разыгрывается стохастический процесс фрактальной «металлизации» внутриоблачной среды, отражающий, с одной стороны, экспериментально наблюдаемые особенности предварительной стадии и, с другой стороны, обладающий универсальными скейлинговыми свойствами, типичными для критической кинетики сильно неравновесных систем. Фрактальный аспект динамики модельной системы подробно рассматривается во втором разделе с позиций теории протекания. Третий раздел посвящён анализу модели в приближении среднего поля. Высокочастотное радиоизлучение микроразрядов на предварительной стадии молниевого разряда обсуждается в четвёртом разделе. Представленные в пятом разделе количественные следствия развитой модели находятся в разумном согласии с экспериментом. В заключение авторы суммируют полученные результаты.

## 1. ПУЧКОВО-ПЛАЗМЕННЫЙ РАЗРЯД В ГРОЗОВОМ ОБЛАКЕ

В работах [9–11] была выдвинута гипотеза о физической связи предварительной стадии молниевого разряда с мелкомасштабной электрической стратификацией грозового облака — наличием тонкой структуры в распределении электрических зарядов и электрического поля, которая проявляется практически во всех натурных экспериментах [12–15].

Для существования такой мелкомасштабной электрической структуры в грозовом облаке есть веские физические основания. Во введении уже было отмечено, что грозовое облако на зрелой стадии развития представляет собой плазмоподобную многопотоковую систему. Взаимопроникающие потоки заряженных тяжёлой (крупные капли и град) и лёгкой (кристаллы льда и мелкие капли) компонент возникают в облаке благодаря совместному действию силы тяжести и восходящего конвективного потока воздуха. При наблюдаемых скоростях восходящего потока  $u = 5\text{--}20 \text{ м/с}$  крупные частицы (с массой  $M \sim 10^{-5} \text{ кг}$ ) оказываются взвешенными в потоке, заполняя в основном нижнюю половину облака, в то время как лёгкая фракция уносится вместе с потоком в верхнюю часть облака. Ситуация во многом напоминает пучково-плазменный разряд. Как было показано одним из авторов [9], эта система при достаточно большом заряде на крупных частицах становится неустойчивой по отношению к возбуждению электрических волн пространственного заряда. На линейной стадии развития неустойчивости поле волны можно представить в виде  $E \sim \exp(-i\omega t + i(\mathbf{k}, \mathbf{r}))$ , где  $\mathbf{r}$  и  $t$  — соответственно пространственная координата и время. Рассматривая в простейшем случае монодисперсный ансамбль тяжёлых частиц в потоке ионизированного воздуха, можно записать дисперсионное соотношение, связывающее

волновой вектор  $\mathbf{k}$  электрической волны с её частотой  $\omega$  [9]:

$$1 - \frac{\Omega^2}{\omega(\omega + i\nu_{st})} + \frac{4\pi i\sigma}{\omega - (\mathbf{k}, \mathbf{u})} = 0, \quad (1)$$

где  $\Omega^2 = 4\pi q^2 N/M$  и  $\nu_{st} \sim g/u$  — соответственно квадрат плазменной частоты и эффективная частота соударений для газа тяжёлых частиц,  $q$ ,  $N$  и  $M$  — соответственно заряд, концентрация и масса тяжёлых частиц,  $g$  — ускорение свободного падения,  $u$  и  $\sigma$  — скорость и проводимость потока ионизированного воздуха. Порог неустойчивости определяется условием  $(\Omega/\nu_{st})^2 \geq 1$ , темп роста неустойчивости  $\gamma \approx 2\pi\sigma\Omega/\nu_{st}$  для оптимального пространственного периода  $a \sim \pi/k \approx \pi u/\Omega$ . Для частиц льда радиуса примерно 0,5 см и  $N \sim 10^3 \text{ м}^{-3}$  порог неустойчивости достигается при  $q \sim 10^{-10} \text{ Кл}$ . Полагая  $u \sim 10 \text{ м/с}$  и  $\Omega \sim 2\nu_{st}$ , находим  $\nu_{st} \sim 1 \text{ с}^{-1}$ ,  $a \sim 10 \text{ м}$  и  $\gamma \sim 4\pi\sigma \sim 1 \text{ с}^{-1}$  ( $4\pi\sigma \sim \nu_{st}$ ). Дальнейшие исследования [11] с учётом обмена зарядом между взаимодействующими частицами показали значительную универсальность такой неустойчивости в широком интервале условий внутри облака. На нелинейной стадии амплитуда мелкомасштабного электрического поля по порядку величины определяется равновесием между электрической силой и силой тяжести, действующими на тяжёлую частицу:  $E \sim Mg/q \sim 10^6 \text{ В/м}$ . Следовательно, амплитуда мелкомасштабного электрического поля может значительно превышать среднюю напряжённость крупномасштабного внутриоблачного поля. Мы полагаем, что рост амплитуды мелкомасштабного поля ограничен некоторой критической величиной, по достижении которой между максимумами и минимумами электрической волны развивается искровой разряд. Именно возникновение мелкомасштабных пробоев ограничивает развитие неустойчивости и лежит в основе формирования лидера молниевого разряда.

## 2. МОДЕЛИРОВАНИЕ НА ОСНОВЕ СЕТИ КОНЕЧНЫХ АВТОМАТОВ

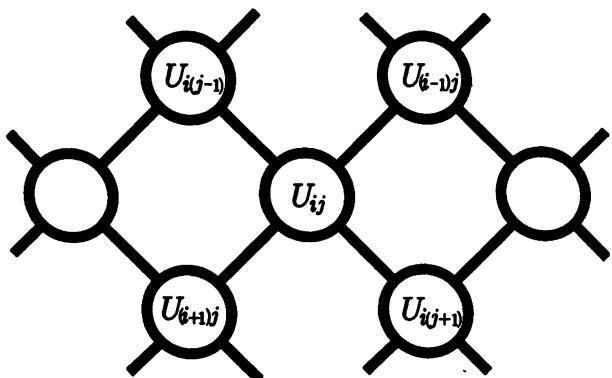


Рис. 1. Модельная решётка потенциалов

Эволюция мелкомасштабной электрической структуры на предварительной стадии молниевого разряда определяется взаимодействием процессов роста, обусловленных развитием пучковой неустойчивости, и процессов нейтрализации, обусловленных возникновением мелкомасштабных пробоев внутриоблачной среды. Мы проанализируем детали этого взаимодействия, используя моделирование на основе сети конечных автоматов. Для наглядности рассмотрим двумерный случай. Будем моделировать мелкомасштабную структуру облака элементами квадратной решётки (см. рис. 1). Шаг решётки соответствует размеру ячейки мелкомасштабной

электрической структуры и составляет порядка десяти метров. Если учесть, что максимальный пространственный масштаб системы определяется размером активной части грозового облака и составляет несколько километров, то линейный размер модельной решётки должен содержать несколько сотен пространственных периодов.

Каждому узлу решётки поставлена в соответствие зависящая от времени скалярная величина  $U_{ij}$ , характеризующая электрический потенциал. Эволюция потенциального рельефа на решётке определялась в рамках модели случайного роста: на каждом шаге модельного времени к электрическим потенциалам  $U_{ij}$  в узлах решётки прибавлялись нормально распределённые случайные добавки. При этом каждый узел совершил независимое от соседей броуновское движение в пространстве значений

электрического потенциала. Разности потенциалов между соседними узлами собраны в матрице  $\mathbf{A}$ , элементы которой определяются соотношениями

$$\begin{aligned} A_{(2i-1)2j} &= U_{i(j+1)} - U_{ij}, \\ A_{2i(2j-1)} &= U_{(i+1)j} - U_{ij}, \\ A_{(2i-1)(2j-1)} &= A_{2i2j} = 0. \end{aligned} \quad (2)$$

Ненулевые элементы матрицы  $\mathbf{A}$  характеризуют связи между узлами исходной решётки. Каждую связь можно рассматривать как конечный автомат, правила поведения которого определяются его собственным состоянием — перепадом потенциала на нём, и состоянием его ближайших соседей. Причём выбор ближайших соседей является дополнительным «правилом игры». Мы использовали две различные конфигурации ближайших соседей, изображённые схематически на рис. 2а и б. В первом случае ближайшими соседями являются четыре смежных связи, симметричных относительно исходной (а); ближайшими соседями являются все шесть связей, имеющих общий узел с исходной связью (б). Заметим, что сеть клеточных автоматов образует покрывающую решётку по отношению к исходной, представленной на рис. 1

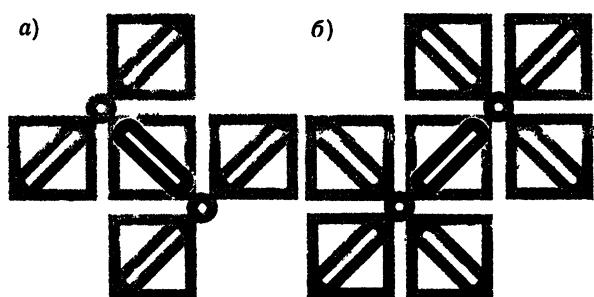


Рис. 2. Два варианта выбора ближайших соседей для фиксированной (выделена чёрным) связи: ближайшими соседями являются четыре связи, симметричных относительно исходной (а); ближайшими соседями являются все шесть связей, имеющих общий узел с исходной связью (б). Заметим, что сеть клеточных автоматов образует покрывающую решётку по отношению к исходной, представленной на рис. 1

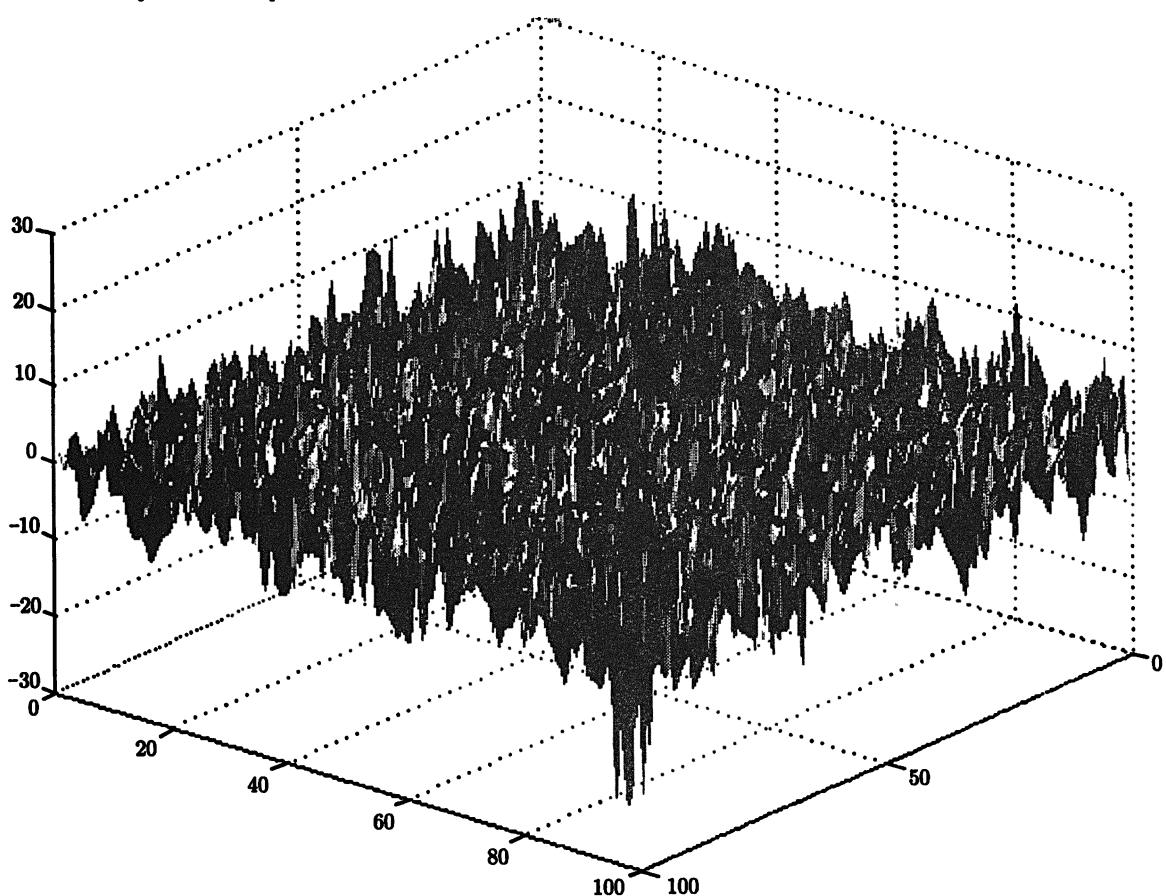


Рис. 3. Мгновенная фотография случайного потенциального рельефа, эволюционирующего на квадратной решётке  $100 \times 100$



Рис. 4. Эволюция разности потенциалов на отдельном элементе сети. Стрелками указаны моменты пробоя (металлизации) активированной связи, когда происходит выравнивание потенциалов в соседних ячейках мелкомасштабной электрической структуры. Абсолютная величина разности потенциалов на концах связи меньше критической ( $U_c = 50$ ), следовательно, указанные пробой стимулированы ближайшими соседями

Вернёмся к правилам игры. Рост потенциально-го рельефа ограничен некоторой критической разностью потенциалов  $U_c$  между соседними узлами, по достижении которой происходит пробой — «металлизация» связи. Предполагается, что на следующем шаге модельного времени возникший пробой может инициировать пробой соседних связей («инфицировать» соседей), если приложенная к ним разность потенциалов превышает некоторый фиксированный уровень активации  $U_a$ , который меньше критического. Здесь уместно провести аналогию с экспериментами по инициации искрового разряда лазерным пучком, где поле пробоя уменьшалось в несколько раз под действием излучения [16]. Пробитая связь на следующем шаге модельного времени возвращается в деактивированное состояние с нулевой разностью потенциалов.

На рис. 3. представлен характерный вид потенциального рельефа, реализованного в численном эксперименте. Поведение разности потенциалов на отдельном элементе сети в зависимости от модельного времени показано на рис. 4. Электризация облака и рост мелкомасштабного потенциального рельефа определяют медленную эволюцию внутриоблачной среды, на фоне которой происходят быст-

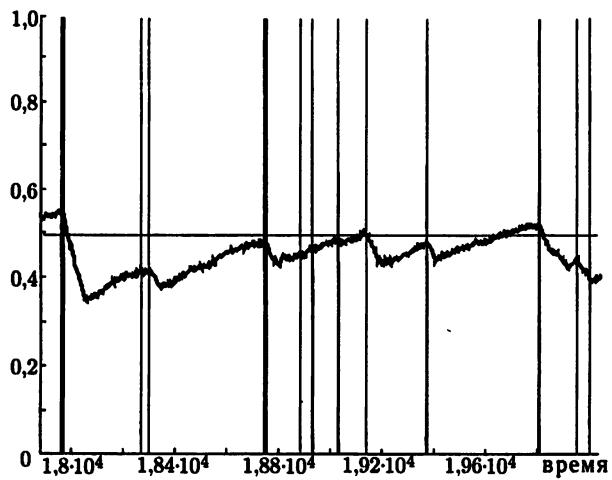


Рис. 5. Эволюция удельного числа активированных связей (жирная линия) и зависимость от модельного времени числа критических связей (дельтаобразные всплески единичной длины)

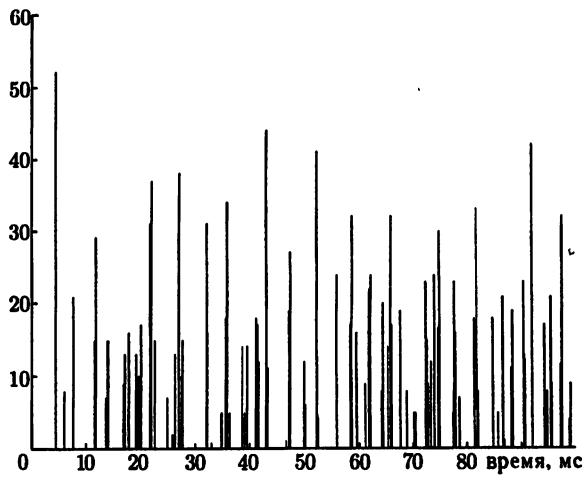


Рис. 6. Эволюция числа металлизированных связей. Грубое временное разрешение: цена деления по оси абсцисс соответствует десяти миллисекундам

рые релаксационные процессы, обусловленные мелкомасштабными разрядами. Параметры модели подобраны так, чтобы обеспечить сравнительно быструю диссипацию потенциала при медленном его росте: отдельный шаг в жизни автомата соответствует характерному времени передачи возбуждения

его ближайшим соседям, при этом изменение разности потенциалов между соседними узлами за один шаг модельного времени на несколько порядков меньше  $U_c$ . Зная размер ячейки (порядка 10 м) и полагая, что возбуждение передаётся со скоростью стримера (порядка  $10^7$  м/с), можно оценить шаг модельного времени величиной порядка микросекунды. Длительность предварительной стадии молниевого разряда, длящаяся от десятков до сотен миллисекунд, определяет полный интервал модельного счёта, который, следовательно, должен составлять от десятков до сотен тысяч шагов.

Обратимся к результатам численного эксперимента. На рис. 5 показана эволюция удельного числа активированных связей; там же вертикальными линиями отмечены моменты появления критических связей, инициирующих металлизацию. Легко заметить, что появление критических связей является чрезвычайно редким событием на фоне общего числа мелкомасштабных пробоев. Видно, кроме того, что удельное число активированных связей флюкутирует вблизи уровня  $p = p_c \sim 0.5$ . Этот критический уровень определяется равновесием между процессами роста мелкомасштабного потенциального рельефа и процессами выравнивания потенциалов в соседних ячейках при возникновении пробоя между ними. Появление новых активированных связей компенсируется их исчезновением в процессе металлизации. Можно сказать, что происходит заселение среды активированными элементами.

Соответствующая временная эволюция числа внутриоблачных разрядов представлена на рис. 6 и 7. На больших временах при грубом временном разрешении события представляют собой дельтаобразные вспышки (см. рис. 6), частота появления которых обратно пропорциональна их амплитуде. Напротив, при более высоком разрешении на коротких временах проявляется тонкая структура отдельного внутриоблачного разряда, определяемая процессом металлизации отдельного кластера активированных связей (см. рис. 7a–в). Под кластером активированных связей мы подразумеваем такую совокупность активированных элементов решётки, любые два из которых связаны непрерывной цепочкой активированных связей, причём каждая из этих связей является членом совокупности. Принципиально важным оказывается тот факт, что кластеры активированных связей оказываются фрактальными структурами. Мгновенная фотография процесса металлизации кластера активированных связей приведена на рис. 8. Чёрные точки на рисунке соответствуют фронту ионизации — элементам структуры, пробитым на данном шаге модельного времени. Вычисленная по результатам моделирования фрактальная размерность металлизированных кластеров составила  $d_f \approx 1.9$ , что практически совпадает, как мы увидим в разделе 4, с размерностью двумерного перколяционного кластера.

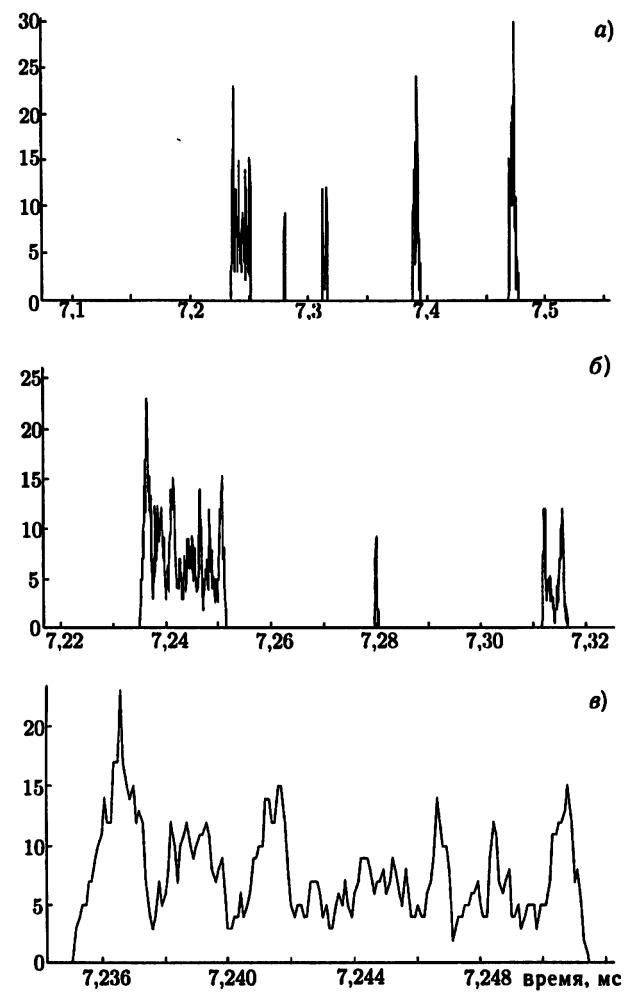


Рис. 7. Эволюция числа metallизированных связей. Высокое разрешение по времени: цена деления по оси абсцисс меняется от 500 мкс на верхнем графике до 20 мкс на нижнем

На первый взгляд удивительно, что однородная во времени и в пространстве накачка — случайно-равномерный рост потенциального рельефа приводит к появлению неоднородных фрактальных диссипативных структур. В облаке такая кластерная диссипация обусловлена тем, что величина, обратная инкременту роста мелкомасштабного потенциального рельефа, на несколько порядков превышает время передачи возбуждения от пробитого элемента его ближайшим активированным соседям. При этом даже самый крупный кластер активированных связей «выгорает» быстрее, чем среди его ближайших соседей появляются новые активированные элементы сети. Мы хотим подчеркнуть принципиальную важность такого резкого разделения характерных временных масштабов в рассматриваемой системе, которое наряду с пороговым характером быстрых диссипативных процессов вообще типично для сильно неравновесных систем, демонстрирующих критическое поведение. Ниже, в разделе 3, мы увидим, что этот резкий контраст типичных времён, характеризующих накачку и диссипацию, является необходимым условием пространственно-временного самоподобия диссипативных структур. Снятие этого требования существеннейшим образом изменяет характер динамики системы. Действительно, если кластер активированных связей не успевает «выгореть» прежде, чем среди его ближайших соседей подрастут новые активированные элементы сети, то процесс металлизации уже никогда не остановится даже в отсутствие новых спонтанных вспышек. По системе при этом будут распространяться сменяющие друг друга фронты горения, имеющие квазирегулярную спиральную структуру (см. рис. 9). Заметим, что характерный пространственный масштаб спиралей и типичное расстояние между ними обратно пропорциональны скорости роста мелкомасштабного рельефа.

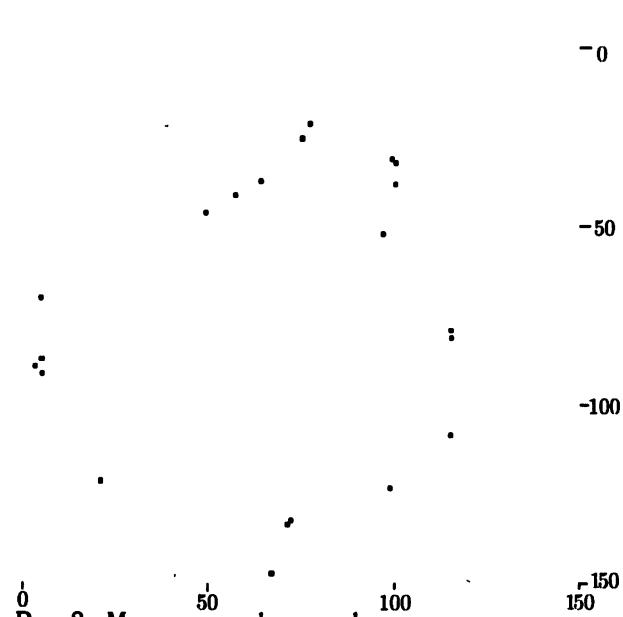


Рис. 8. Мгновенная фотография процесса металлизации кластера активированных связей на квадратной решётке размера  $150 \times 150$ . Чёрные точки соответствуют фронту ионизации — элементам структуры, пробитым на данном шаге модельного времени

-0

-50

-100

-150

При получении реализаций, показанных на рис. 6–8, мы использовали отношение уровня активации  $U_a$  к критической разности потенциалов  $U_c$ , близкое к 20 % ( $U_a/U_c \sim 0,2$  соответствует результатам экспериментов по лазерной инициации искры). При плавном увеличении параметра  $U_a/U_c$  происходит быстрое сокращение размеров металлизированных областей, хотя сам характер динамики не претерпевает серьёзных качественных изменений. Однако уменьшение отношения  $U_a$  к  $U_c$  при фиксированной скорости роста потенциального рельефа приводит к срыву критической динамики и установлению режима релаксационных автоколебаний, когда вся сеть ведёт себя как один сосредоточенный элемент. Такая синхронизация автоматов обусловлена тем, что все они к моменту появления первого спонтанного пробоя успевают преодолеть уровень активации. Восстановление критической динамики при фиксированном  $U_a/U_c$  требует ещё большего уменьшения скорости роста потенциального рельефа. Ниже, в разделе 3, рассматривая приближение среднего поля, мы увидим, что устойчивость критического режима наблюдается в пределе бесконечно малых скоростей роста.

Если отношение уровня активации  $U_a$  к критической разности потенциалов  $U_c$  и соотношение характерных времён роста и диссипации оказывают принципиальное влияние на динамику, то детали локальных взаимоотношений автоматов не оказывают заметного влияния на качественный характер поведения системы. Так, для обеих конфигураций ближайших соседей, схематически изображённых

на рис. 2а и б мы получили качественно совпадающие результаты. Количественные отличия проявились в росте характерных размеров металлизированных кластеров для случая на рис. 2б, что обусловлено увеличением числа ближайших соседей, следовательно, и увеличением вероятности передачи возбуждения. Кроме того, увеличение числа ближайших соседей существенно понижает уровень критической заселённости: мы получили  $p_c \approx 0,27$  в случае рис. 2б против  $p_c \approx 0,5$  в случае рис. 2а. Не внесла качественных изменений в динамику и смена последовательности выполнения процедур активации и выравнивания потенциалов соседних ячеек.

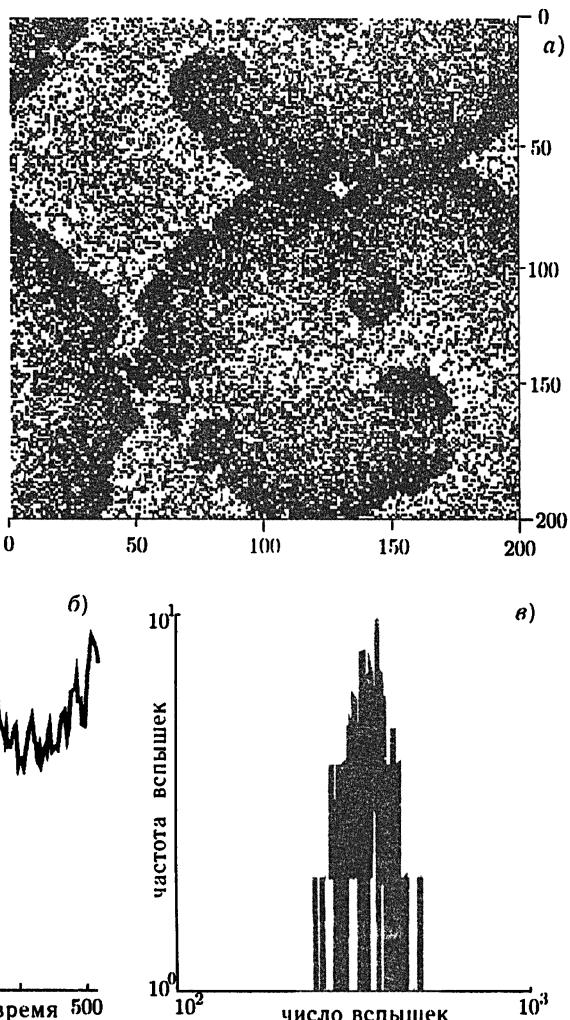


Рис. 9. Появление спиральных волн в сети конечных автоматов при нарушении условия существенного отличия временных масштабов накачки и диссипации: мгновенная фотография системы (а); временная эволюция числа вспышек (б) и распределение частоты вспышек в билогарифмическом масштабе (в)

Вернёмся к последовательности вспышек металлизации, представленной на рис. 6, и исследуем статистику событий. На рис. 10 частота появления металлизированных кластеров, содержащих  $N$  элементов, представлена как функция числа элементов  $N$  в билогарифмическом масштабе. Очевиден степенной характер зависимости, указывающий на самоподобие пространственно-временной динамики: в широком диапазоне значений нельзя выделить характерные размеры кластеров и характерные частоты их появления. Скейлинговый режим ограничен, с одной стороны, размерами системы и временем счёта, с другой — размером кластера, состоящего из единственного элемента, и обратной частотой появления этих единичных кластеров. Характерный интервал между частыми мелкими событиями (появление кластеров с размером не более десяти элементов) составляет сотни шагов модельного времени.

Примечательно, что и при более высоком разрешении на коротких интервалах (от одного до сотен шагов модельного времени) были обнаружены скейлинговые свойства процесса металлизации отдельного активированного кластера. Вычисляя, например, среднюю длину графика  $N(t)$  — числа металлизированных на данном шаге связей как функцию интервала усреднения мы обнаружили, что она растёт

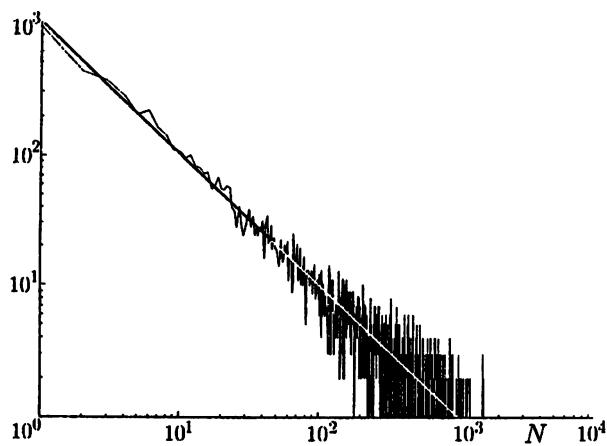


Рис. 10. Частота появления metallизированных кластеров как функция их размера в билогарифмическом масштабе

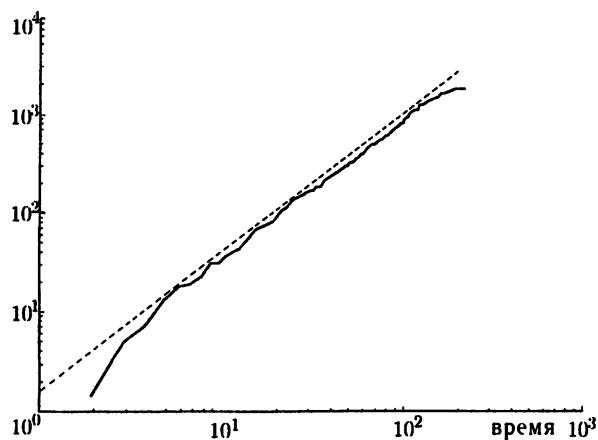


Рис. 11. Полученная в численном эксперименте зависимость длины графика  $N(t)$  от интервала времени. Наклон полученной зависимости в билогарифмическом масштабе определяет фрактальную размерность графика  $D_f \approx 1,28$

быстрее, чем линейная функция, т. е. является фракталом с размерностью  $D_f = 1,28$  (см. рис. 11). Используя известное определение  $D_f = 2 - Hu$ , где  $Hu$  — показатель Хёрста обобщённого броуновского процесса [17], мы получили следующее асимптотическое соотношение, подтверждённое в численном эксперименте:

$$N(t) \propto t^{0.72}. \quad (3)$$

Напомним для сравнения, что для графика смещений  $x(t)$  классической броуновской частицы фрактальная размерность  $D_f = 3/2$  в соответствии с известной асимптотикой  $x(t) \propto t^{0.5}$ . В разделе 4, используя представления теории перколяции, мы детально исследуем пространственную структуру активированных кластеров, найдём их размерность и с помощью средств фрактальной геометрии выведем асимптотическую формулу (3).

### 3. ПРИБЛИЖЕНИЕ СРЕДНЕГО ПОЛЯ

В этом разделе мы хотим выяснить общие критерии, обеспечивающие появление критической динамики. Численный эксперимент, например, показал существование критического уровня заселённости, флуктуации вблизи которого обладают скейлинговыми свойствами. Кроме того, в эксперименте была выявлена зависимость этого уровня от числа ближайших соседей для заданной решётки автоматов, т. е. от её координационного числа  $z$ . Чем большее число элементов находится в контакте с «заражённым», тем выше вероятность размножения «инфекции» и тем, следовательно, ниже критический уровень заселённости. Координационное число возрастает с увеличением размерности решётки. В пределе высоких координационных чисел для расчёта критического уровня заселённости может быть использовано приближение среднего поля.

Введём функцию распределения связей по абсолютной величине перепада потенциала на них  $f(|A_{ij}|)$ . Величина  $f(|A_{ij}|) d|A_{ij}|$  определяет долю связей с абсолютной величиной разности потенциалов в интервале  $[|A_{ij}|, |A_{ij}| + d|A_{ij}|]$ . Очевидно, что  $\int_0^{l_c} f(|A_{ij}|) d|A_{ij}| = 1$ . Удельная доля активи-

рованных связей определяется величиной

$$p = \int_{l_a}^{l_c} f(|A_{ij}|) d|A_{ij}| \quad (4)$$

и является своеобразным критическим параметром системы.

Развитая нами модель распространения «инфекции» микроразрядов в облаке находится в одном классе универсальности с хорошо известной моделью распространения лесных пожаров (forest-fire model [18]). По аналогии с этим классом моделей можно сформулировать достаточно наглядный критерий универсального критического поведения системы. Для этого мы используем среднеполевые переменные, характеризующие динамику активированных ( $p$ ), металлизированных ( $m$ ) и инертных ( $q$ ) связей. Зная среднюю скорость роста мелкомасштабного рельефа (инкремент неустойчивости)  $\Omega = \Delta U / \Delta t$ , введём вероятность появления в единицу модельного времени активированной связи  $\varepsilon = \Omega f(U_a)$  и вероятность спонтанной металлизации связи в единицу времени  $\xi = \Omega f(U_c)$ . В модели лесных пожаров  $\varepsilon$  характеризует скорость появления новых деревьев в лесу, а  $\xi$  — вероятность попадания молнии в дерево (возникновения пожара) в единицу времени. Следуя [18], записываем разностные уравнения для  $p$ ,  $m$  и  $q$ :

$$\begin{aligned} p(t+1) - p(t) &= -[\xi + zm(t)] p(t) + \varepsilon q(t), \\ m(t+1) - m(t) &= -m(t) + [\xi + zm(t)] p(t), \\ q(t+1) - q(t) &= -\varepsilon q(t) + m(t), \end{aligned} \quad (5)$$

совместно с законом сохранения

$$p(t) + m(t) + q(t) = 1, \quad (6)$$

где  $t$  — текущее время,  $z$  — координационное число (число соседей), а интервал  $\Delta t = 1$  соответствует одному шагу по времени в компьютерном эксперименте. Используя (6), можно исключить одну из переменных в уравнениях (5). Кроме того, будем рассматривать время как непрерывную переменную и, заменяя конечные разности производными, перепишем динамическую систему (5) в дифференциальной форме:

$$\dot{p} = -(\xi + zm) p + \varepsilon (1 - p - m) \equiv F(p, m), \quad \dot{m} = -m + (\xi + zm) p \equiv G(p, m). \quad (7)$$

Найдём сначала точки покоя системы (7), полагая  $\dot{p} = 0$  и  $\dot{m} = 0$ . Здесь существует единственное решение, совместимое с условием  $p \leq 1$ :

$$p^* = \frac{1 + z + k - [k^2 + 2(1+z)k + (z-1)^2]^{1/2}}{2z}, \quad m^* = \frac{\varepsilon}{1+\varepsilon} (1 - p^*), \quad (8)$$

где  $k = \xi(1+\varepsilon)/\varepsilon$ . Поведение системы вблизи неподвижной точки  $(p^*, m^*)$  определяется корнями характеристического уравнения

$$\begin{vmatrix} \partial F / \partial p - \lambda & \partial F / \partial m \\ \partial G / \partial p & \partial G / \partial m - \lambda \end{vmatrix} = 0, \quad (9)$$

которое можно переписать в виде  $\lambda^2 + B\lambda + C = 0$ , где

$$B = 1 + \xi + \varepsilon + z(p^* - m^*), \quad C = \xi(1 + \varepsilon) + \varepsilon[1 + z(p^* - m^*)] + zm^*. \quad (10)$$

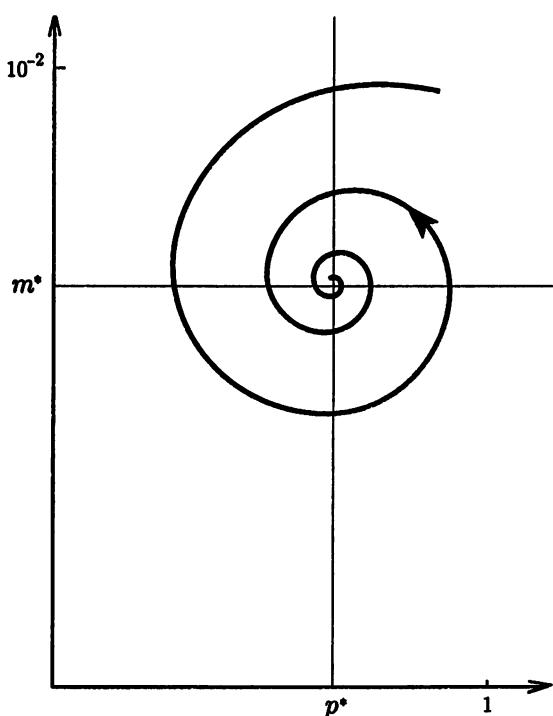


Рис. 12. Фазовая плоскость модели в приближении среднего поля

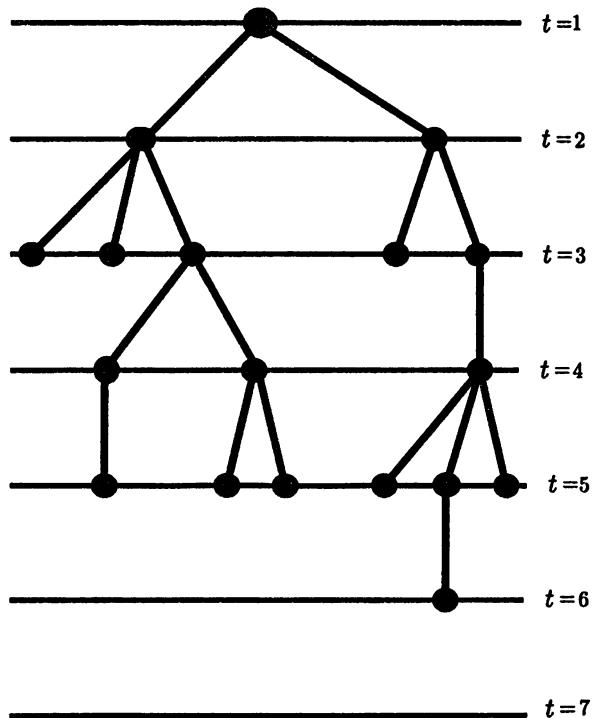


Рис. 13. Ветвящийся процесс металлизации активированных связей

В то время как характерное время развития тонкой электрической структуры облака составляет единицы—десятки секунд, её быстрая релаксация за счёт микроразрядов осуществляется за микросекунды. Это означает, что в условиях реального облака  $\xi \ll \varepsilon$ . В двойном пределе  $\xi/\varepsilon \rightarrow 0$  и  $\varepsilon \rightarrow 0$  оба корня характеристического уравнения оказываются комплексными числами с отрицательной реальной частью. Следовательно, точка покоя системы (7) является фокусом. Фазовый портрет системы изображён схематически на рис. 12.

Условие  $\xi/\varepsilon \ll 1$  обеспечивается экспоненциальной малостью функции распределения  $f(U)$  вблизи критического значения  $U_c$  и физически обусловлено чрезвычайной редкостью спонтанных микроразрядов. В пределе  $\varepsilon \rightarrow 0$  асимптотическое поведение удельного числа активированных связей определяется соотношением

$$p \approx z^{-1} \approx p_c, \quad (11)$$

которое не зависит от конкретных значений  $\varepsilon$  и  $\xi$ .

Металлизация связей активированного кластера может быть рассмотрена как ветвящийся процесс. Рис. 13 иллюстрирует древовидную структуру, созданную ветвящимся процессом. Процесс стартует с верхнего уровня ( $t = 1$ ), соответствующего моменту пробоя критической связи. Исходный узел на первом уровне может иметь  $n$  ветвей ( $0 \leq n \leq z$ ), соответствующим возможному числу ближайших активированных соседей. В следующий момент времени ( $t = 2$ ) каждая из вновь металлизированных связей возбуждает уже своих ближайших активированных соседей и т. д. Среднее число новых ветвей — пробитых связей, возникающих в процессе металлизации, называется коэффициентом ветвления  $\sigma = \langle n \rangle$ . По определению

$$\sigma = \sum_{n=0}^z n C_{z,n} p^n (1-p)^{z-n} = 1 - \frac{k}{z-1} + O(k^2), \quad (12)$$

где  $C_{z,n}$  — биномиальные коэффициенты,  $k = \xi(1 + \varepsilon)/\varepsilon$ . Видно, что ветвящийся процесс становится критическим, когда  $\xi \rightarrow 0$ .

Приведённое выше значение критической концентрации  $p$  активированных связей может быть обнаружено в рамках модели «направленного протекания» [19]. Мы видели, что если в данный момент времени связь находится в возбуждённом состоянии, то в следующий (дискретный) момент времени она возвращается в исходное состояние покоя, но может при этом с вероятностью  $p$  возбудить своих ближайших соседей по решётке. В результате разыгрывается стохастический процесс, напоминающий рассмотренное нами распространение «инфекции». Ясно, что когда вероятность передачи возбуждения очень мала, число возбуждённых элементов будет убывать со временем. В обратном пределе, когда вероятность передачи возбуждения близка к единице, число возбуждённых элементов решётки неограниченно возрастает. Критической является вероятность передачи возбуждения  $p_c$ , при которой появляется конечная вероятность роста «эпидемии» при единственном «инфицированном» в начальный момент.

В пределе высоких координационных чисел для расчёта порога протекания может быть использовано приближение среднего поля. Пусть в фиксированный момент времени  $t$  для элементов среды, находящихся в возбуждённом состоянии, составляет  $q_t$ . Если число ближайших соседей  $z$  у каждого элемента среды велико, можно считать, что средних в момент времени  $t$  имелось  $zq_t$  возбуждённых элементов. Выбранный нами элемент может быть «инфицирован» в следующий  $(t+1)$ -й момент времени с вероятностью

$$1 - (1 - p)^{zq_t}, \quad (13)$$

поскольку мы предполагаем независимость передачи «инфекции» через ближайших соседей. Выражение (13) даёт долю элементов среды, находящихся в возбуждённом состоянии в следующий  $(t+1)$ -й момент времени, т. е.

$$q_{t+1} = 1 - (1 - p)^{zq_t}. \quad (14)$$

Выше порога перколяции величина  $q_t$  должна стремиться к отличному от нуля конечному пределу:

$$\lim_{t \rightarrow \infty} q_t = \tilde{q}, \quad (15)$$

который даёт долю возбуждённых элементов в установившемся режиме, тогда как ниже порога перколяции  $\lim_{t \rightarrow \infty} q_t = 0$ .

Исследуя точечное отображение (14), нетрудно показать, что его аттракторами могут быть только стационарные точки, причём потеря устойчивости стационарной точки  $q = 0$  и рождение новой устойчивой стационарной точки  $q = \tilde{q} > 0$  происходит при прохождении вероятности  $p$  через значение  $p_c$ , удовлетворяющее условию

$$1 - p_c = \lim_{\tilde{q} \rightarrow 0} (1 - \tilde{q})^{1/(z\tilde{q})} = \exp(-1/z). \quad (16)$$

Таким образом, порог перколяции определяется соотношением

$$p_c = 1 - \exp(-1/z) = 1/z - 1/(2z^2) + \dots \quad (17)$$

При уменьшении координационного числа  $z$  приближение среднего поля становится неприменимым и возрастает роль флуктуационных эффектов. Так, в нашей компьютерной модели, вычисляя среднее число активированных связей  $p = \int_{I_a}^{I_c} f(|A_{ij}|) d|A_{ij}|$ , мы обнаружили  $p_c = 0,5$  при  $z = 4$  и  $p_c = 0,27$  при  $z = 6$ , что плохо согласуется с предсказаниями среднеполового рассмотрения. Для преодоления этих трудностей и адекватного учёта флуктуаций мы в следующем разделе воспользуемся средствами теории перколяции.

#### 4. ПЕРКОЛЯЦИОННАЯ СТРУКТУРА МЕТАЛЛИЗИРОВАННЫХ КЛАСТЕРОВ

Интуитивно ясно, что характерные масштабы металлизированных кластеров существенным образом зависят от удельного числа  $p$  активированных элементов решётки. Действительно, при малых концентрациях активированные связи либо изолированы друг от друга, либо формируют небольшие кластеры ближайших соседей. Соответственно и распространение микропробоев ограничено в этой ситуации отдельными связями и их небольшими группами — вероятность развития «эпидемии» ничтожна. Наоборот, при удельном числе активированных связей, близком к единице, возникновение «эпидемии» микроразрядов становится достоверным событием. Существует некоторое критическое значение  $p = p_c$ , при котором происходит переход от конечных очагов «инфекции» (при  $p < p_c$ ) к «эпидемии» — неограниченному распространению «инфекции» (при  $p > p_c$ ). Критическое значение  $p = p_c$  называют обычно порогом протекания (перколяции) [20].

Перколяционный переход является геометрическим фазовым превращением второго рода и характеризуется геометрическими свойствами активированных кластеров вблизи  $p = p_c$ . При  $p < p_c$  существуют только кластеры небольших размеров. По мере роста концентрации  $p$  средний размер кластеров увеличивается. При концентрации, близкой к критической, появляется кластер, связывающий противоположные стороны решётки. Такой кластер называют перколяционным. С дальнейшим ростом концентрации доля принадлежащих перколяционному кластеру связей  $P_\infty$  (плотность перколяционного кластера) возрастает, соответственно средний размер конечных кластеров, не принадлежащих перколяционному кластеру, уменьшается, и при  $p \sim 1$  все связи принадлежат, очевидно, только одному кластеру.

Величина  $P_\infty$  является параметром порядка перколяционного перехода и может быть определена как вероятность того, что связь принадлежит бесконечному кластеру. Когда  $p < p_c$ , существуют только конечные кластеры, и  $P_\infty = 0$ . При  $p > p_c$  параметр порядка  $P_\infty$  увеличивается по степенному закону:

$$P_\infty \propto (p - p_c)^\beta. \quad (18)$$

Радиус корреляции  $\xi$  перколяционного перехода имеет простой геометрический смысл: ниже  $p_c$  он определяет характерные линейные размеры конечных кластеров, а выше  $p_c$  — характерный размер «дыр» в бесконечном перколяционном кластере. С приближением к порогу  $p_c$  радиус корреляции расходится:

$$\xi \propto |p - p_c|^{-\nu}, \quad (19)$$

с одинаковым показателем  $\nu$  ниже и выше порога. Среднее число узлов (масса) конечного кластера тоже расходится вблизи порога:

$$S \propto |p - p_c|^{-\gamma}, \quad (20)$$

также с одинаковым показателем  $\gamma$  ниже и выше  $p_c$ . Для получения  $\xi$  и  $S$  требуется вычисление средних по всем конечным кластерам на решётке.

Показатели  $\beta$ ,  $\nu$  и  $\gamma$  описывают критическое поведение типичных величин, связанных с перколяционным переходом, и называются критическими индексами. Индексы эти универсальны и не зависят ни от структурных особенностей решётки (квадратная или треугольная и т. д.), ни от типа протекания (по узлам, по связям или континуальное), но только от размерности пространства.

Заметим, что все введённые выше величины определены в пределе бесконечно больших систем. В ограниченной системе  $P_\infty$ , например, не обязательно равно нулю ниже порога (при  $p < p_c$ ).

Перколяционный кластер является физическим фракталом. Он самоподобен на всех линейных масштабах, больших периода решётки и меньших размера системы. Фрактальная размерность кластера описывает, как в среднем меняется его масса (число принадлежащих ему элементов)  $M$  внутри евклидовой сферы с ростом её радиуса  $r$ :

$$M(r) \propto r^{d_f}. \quad (21)$$

Здесь  $M(r)$  представляет результат усреднения по многим различным конфигурациям переколяционного кластера.

Связем фрактальную размерность  $d_f$  переколяционного кластера с критическими индексами  $\beta$  и  $\nu$ . Вероятность того, что некоторый выбранный наугад узел внутри евклидовой сферы радиуса  $r < \xi$  принадлежит переколяционному кластеру, определяется отношением числа узлов, принадлежащих переколяционному кластеру, к полному числу узлов в сфере:

$$P_\infty \propto r^{d_f}/r^d, \quad r < \xi. \quad (22)$$

Здесь  $d$  — размерность пространства вложения. Последнее соотношение корректно для  $r = a\xi$ , где  $a$  — некоторая константа меньше 1. Подставляя  $r = a\xi$  в (22), получаем

$$P_\infty \propto \xi^{d_f}/\xi^d. \quad (23)$$

Обе части последнего соотношения являются степенными функциями выражения  $p - p_c$ . Подставляя (18) и (19) в (23), находим

$$d_f = d - \beta/\nu. \quad (24)$$

Таким образом, фрактальная размерность  $d_f$  переколяционного кластера на пороге протекания не является новым независимым параметром, но выражается через показатели  $\beta$  и  $\nu$ . Поскольку  $\beta$  и  $\nu$  являются универсальными индексами,  $d_f$  также является универсальной величиной, зависящей только от размерности пространства. Для двумерного случая  $d_f \approx 1,89$ . Ниже порога и на пороге (при  $p \leq p_c$ ) выражение (24) определяет и фрактальную размерность конечных кластеров с линейными размерами, не превышающими  $\xi$ .

Для получения новой существенной информации об устройстве переколяционного кластера, возникающего в процессе металлизации, рассмотрим наикратчайший путь между двумя его точками  $A$  и  $B$ . Элементы кластера, принадлежащие этому пути, формируют самоподобную структуру с фрактальной размерностью  $d_{\min}$ . Таким образом, длина  $l$  этого наикратчайшего пути, часто называемая химическим расстоянием, меняется с увеличением евклидовой дистанции между  $A$  и  $B$  по закону

$$l \sim r^{d_{\min}}. \quad (25)$$

Обратное соотношение

$$r \sim l^{1/d_{\min}} \equiv l^\nu, \quad (26)$$

показывает, как  $r$  меняется при изменении  $l$ .

С размерностями  $d_{\min}$  и  $d_f$  тесно связана так называемая химическая размерность  $d_l$ , которая описывает изменение числа  $M$  элементов кластера, удаленных от некоторого фиксированного элемента на химическое расстояние, не превышающее дистанцию  $l$ , с увеличением последней:

$$M(l) \propto l^{d_l}. \quad (27)$$

Если фрактальная размерность  $d_f$  характеризует рост массы кластера с увеличением его размера в евклидовом пространстве, то химическая размерность  $d_l$  описывает рост массы кластера в пространстве химических расстояний  $l$ . Комбинируя (21), (25) и (27), находим связь

$$d_l = d_f/d_{\min}. \quad (28)$$

В двумерном случае  $d_{\min} = 1,10 \pm 0,02$ ;  $d_l = 1,72 \pm 0,02$ .

Процесс металлизации инициируется отдельной пробитой связью активированного кластера, и каждый последующий шаг модельного времени соответствует увеличению химического расстояния

между исходным элементом и фронтом металлизации на пространственный период решётки. Следовательно, в нашем случае  $t \propto l$ , и для числа металлизированных на данном шаге связей  $N(t)$  мы можем записать следующее асимптотическое соотношение:

$$N(t) \propto \frac{\partial M(t)}{\partial t} \propto t^{d_l-1} \approx t^{0,72}, \quad (29)$$

или

$$N(t) \propto t^{Hu}, \quad (30)$$

где  $Hu \approx 0,72$  — показатель Хёрста для данного процесса. Соотношение (29) отражает геометрический смысл процесса металлизации:

$$M(t) \propto \int_0^t N(t') dt'. \quad (31)$$

Фрактальная размерность графика  $N(t)$  определяется соотношением  $D_f = 2 - Hu \approx 1,28$ . Близкое значение было получено нами в компьютерном эксперименте (см. рис. 11).

## 5. РАДИОИЗЛУЧЕНИЕ ОБЛАКА НА ПРЕДВАРИТЕЛЬНОЙ СТАДИИ МОЛНИЕВОГО РАЗРЯДА

В рамках развивающейся модели можно определить характеристики радиоизлучения грозового облака на предварительной стадии молниевого разряда. Поле излучения формируется совокупностью микроразрядов, возникающих в процессе роста описанных выше кластеров металлизации. В рассматриваемой модели каждая металлизированная связь представляет собой линейный ток с размером  $a$ , текущий от большего потенциала к меньшему. Поскольку в исследуемой модели знак разности потенциалов фиксируется, то, следовательно, известно и направление возникающих токов. Амплитуда тока  $j_{kl}$ , ассоциированного с  $kl$ -м элементом сети, пропорциональна разности потенциалов  $\Delta U_{kl} = A_{kl}$ , приложенной к элементу в момент пробоя, и зависит от электрической ёмкости ячейки  $C_{kl}$  и продолжительности разряда  $\tau$ . По порядку величины

$$j_{kl} \sim \frac{C_{kl} \Delta U_{kl}}{\tau}. \quad (32)$$

Для определения спектра излучения необходимо задать пространственно-временную зависимость тока, которую мы выберем в следующем виде:

$$j(t') = j_{kl} (\exp(-\alpha t') - \exp(-\delta t') - \theta(t - \tau) \{ \exp[-\alpha(t - \tau)] - \exp[-\delta(t - \tau)] \}), \quad t' = t - s/V, \quad (33)$$

где  $s$  — координата вдоль элемента тока,  $\theta(t)$  — единичная функция Хевисайда, а время запаздывания  $\tau$  определяется скоростью  $V$  распространения фронта ионизации в элементарном разряде:  $\tau \sim a/V$ . Таким образом, если при рассмотрении металлизации кластера активированных связей внутренняя структура отдельного пробоя была несущественной, то при расчёте радиоизлучения мы должны взять соответственно меньший, чем длительность тока, временной шаг, чтобы зависимость (33) была корректно описана. Заметим, что коэффициенты  $\alpha$  и  $\delta$  в (33), которые определяются внутренней динамикой отдельного микроразряда, могут быть выбраны из экспериментальных данных по радиоизлучению на предварительной фазе развития молнии [21, 22].

Для расчёта волновых форм радиосигналов от совокупности линейных токов с пространственно-временными характеристиками (33) мы воспользуемся результатами работы [23], согласно которой амплитуда электрического поля излучения  $E(r, t)$  совокупности линейных токов в дальней зоне

( $|r| \gg \lambda$ ) определяется выражением

$$E_\vartheta(r, t) = \frac{Z_0}{4\pi r} \sum_{k,l=1}^N \Psi_{\vartheta kl} \left[ j(t - t_{kl}^{(1)}) - j(t - t_{kl}^{(2)}) \right], \quad (34)$$

где  $r_{kl}$  — радиус-вектор от  $kl$ -го элемента тока в точку приёма (см. рис. 14),  $\lambda$  — характерная длина волн излучения,  $Z_0 = \mu_0/\epsilon_0$  — импеданс вакуума, ток  $j$  определяется формулой (33), амплитуда

$$\Psi_{\vartheta kl} = \frac{K_{\vartheta kl}}{c/V - \zeta_{kl}}, \quad K_{\vartheta kl} = (\vartheta_0, s_{0kl}), \quad \zeta_{kl} = (r_0, s_{0kl}), \quad (35)$$

$V$  — скорость фронта тока,  $c$  — скорость света,  $s_{0kl}$ ,  $r_0$  и  $\vartheta_0$  — единичные векторы в сферической системе координат, показанной на рис. 14; вектор  $s_{0kl}$  направлен вдоль  $kl$ -го тока, начало радиус-вектора  $r_{kl}$  лежит в средней точке  $kl$ -го линейного тока. Запаздывающие аргументы

$$t_{kl}^{(1)} = \frac{r_{kl}}{c} + t - \frac{r_{kl}}{2}, \quad t_{kl}^{(2)} = t_{kl}^{(1)} + \tau_{kl}, \quad (36)$$

где  $\tau_{kl} = (1 - \zeta_{kl}V/c)a/V$  — время запаздывания по элементу тока,  $t$  — текущее время. Суммирование в (34) ведётся по всем отличным от нуля элементарным токам в момент времени  $t$ .

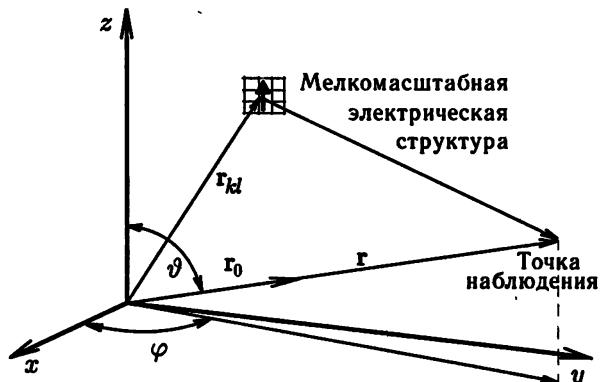


Рис. 14

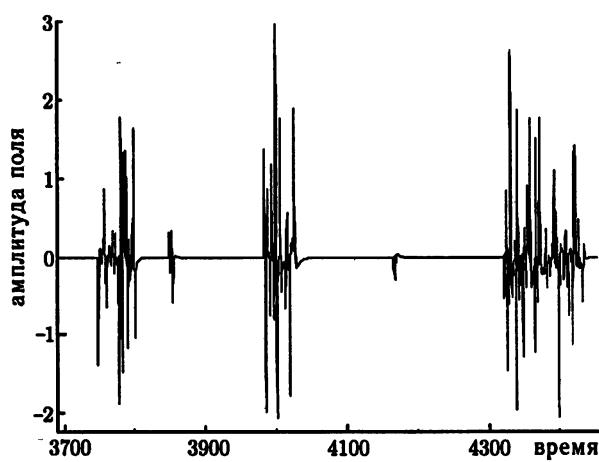


Рис. 15. Волновая форма излучения модельной системы

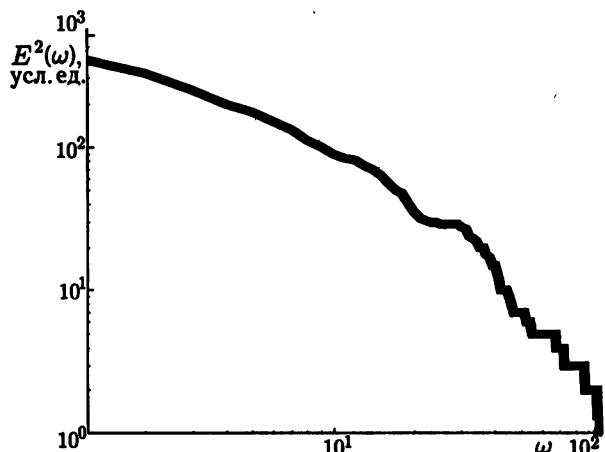


Рис. 16. Энергетический спектр поля излучения

Предполагается, что модельная решётка ориентирована в вертикальной плоскости и расстояние между её нижней кромкой и Землёй составляет один километр, а точка приёма излучения располагается невысоко (около 10 м) над поверхностью Земли на расстоянии до центра облака, много большем размера самого облака. В этом случае с учётом отражения поверхностью Земли полное поле излучения  $E_z(r, t) \approx 2E_\vartheta$ . Используя формулу (34) и пространственно-временные характеристики процесса металлизации, нетрудно найти поле излучения  $E_z(r, t)$  в каждый момент времени  $t = n\Delta t$ . На рис. 15 показаны волновые формы радиосигналов. В конкретных расчётах были выбраны следующие параметры:  $a \sim 10$  м,  $V \sim 10^7$  м/с и соответственно  $\tau$  порядка микросекунды.

По временным характеристикам поля  $E(t)$  находим спектральную плотность излучения, которая удовлетворяет асимптотике  $E^2(\omega) \propto N(\omega) \propto \omega^{-2H_u-1} \approx \omega^{-2,44}$  (см. рис. 16).

## 6. ЗАКЛЮЧЕНИЕ

Выше на примере динамики электрических разрядов на предварительной стадии формирования молнии мы продемонстрировали универсальный подход к описанию неравновесных систем, размеры которых существенно превышают внутренние масштабы характерной для данной системы неустойчивости. Такой подход широко используется в настоящее время при анализе неравновесных систем, состоящих из множества взаимодействующих между собой активных ячеек. Глобальное поведение таких систем слабо зависит от процессов внутри отдельной ячейки и определяется в основном характером взаимодействия ячеек между собой. В случае, когда корреляционные связи между ячейками простираются на конечное расстояние, много меньшее характерного масштаба системы, для описания макропараметров задачи применяются хорошо известные методы традиционной статистической физики. В нашем случае корреляционные длины сравнимы с масштабами рассматриваемой системы. Здесь универсальность подхода обусловлена тем, что многие физические процессы удается моделировать сетью конечных автоматов, причём системы обнаруживают фрактальные свойства (т. е. свойства самоподобия в широком интервале пространственно-временных масштабов).

В соответствии с указанным подходом мы смоделировали динамику электрических разрядов на основе сети конечных автоматов с учётом специфики неравновесности облачной среды. На основании этой модели нам удалось доказать принадлежность рассматриваемого процесса к классу переколяционных систем и определить область параметров задачи, где система обнаруживает критическое поведение. Мы демонстрируем, что параметры грозового облака лежат именно в этой области.

Важную информацию о физических процессах внутри облака содержит пространственно-временная динамика разрядов, которая следует непосредственно из компьютерного моделирования и может быть напрямую сопоставлена с экспериментальными данными по радиоизлучению грозового облака. Такая возможность открывается благодаря предложенному в работе способу определения характеристик радиоизлучения в прямой связи с процессом «размножения» электрических разрядов.

Целью настоящей работы не является детальное сопоставление модели с экспериментальными данными, поскольку для этого необходимо моделирование, по крайней мере, в пространстве трёх измерений. Однако уже первые сопоставления результатов моделирования с экспериментальными данными показывают хорошее качественное совпадение. Из фрактального характера микроразрядов следует и универсальность спектров радиоизлучения, которая обнаруживается в экспериментах.

В качестве иллюстрации количественного сопоставления результатов теории и эксперимента приведём оценку частоты появления микроразрядов по измерениям радиоимпульсов наземной приёмной антенной. Эта частота определяется соотношением

$$F \approx \gamma \frac{w}{a^3} p,$$

где  $\gamma$  характеризует темп роста электрического поля в ячейке,  $w$  — объём активной части грозового облака, где развивается неустойчивость,  $p$  — доля активированных ячеек. На рассматриваемой стадии, когда длина активированного кластера сопоставима с размерами облака,  $p \sim p_c$ , где  $p_c \approx 0,25$  соответствует порогу переколяции в трёхмерном случае. Полагая  $w \sim 10^{10} \text{ м}^3$ ,  $\gamma \sim 1,1 \text{ с}$  и  $a \sim 10 \text{ м}$ , находим  $F \sim 2,5 \cdot 10^5 \text{ с}^{-1}$ , что хорошо согласуется с экспериментом [24].

Работа выполнена при поддержке РФФИ (грант № 01–02–17403).

### СПИСОК ЛИТЕРАТУРЫ

1. Нелинейные волны: Сб. статей / Под ред. А. В. Гапонова-Грехова, М. И. Рабиновича. М.: Наука, 1983.
2. Нелинейные волны: Сб. статей / Под ред. А. В. Гапонова-Грехова, М. И. Рабиновича. М.: Наука, 1985.
3. Полак Л. С., Михайлов А. С. Самоорганизация в неравновесных физико-химических системах. М.: Наука, 1983.
4. Михайлов А. С., Упоров И. В. // УФН. 1984. Т. 144. С. 79.
5. Лоскутов А. Ю., Михайлов А. С. Введение в синергетику. М.: Наука, 1990. 272 с.
6. Bak P., Tang C., Wiesenfeld K. // Phys. Rev. Lett. 1987. V. 59. P. 381.
7. Bak P., Tang C., Wiesenfeld K. // Phys. Rev. A. 1988. V. 38. P. 364.
8. Иудин Д. И., Трахтенгерц В. Ю. // Физика атмосферы. 2000. Т. 36, № 5. С. 650.
9. Трахтенгерц В. Ю. // ДАН СССР. 1989. Т. 308. С. 584.
10. Trakhtengerts V. Y. // Proc. of ICAE-92. St.-Petersburg, Russia. 1992. V. 2. P. 416.
11. Мареев Е. А., Сорокин А. Е., Трахтенгерц В. Ю. // Физика плазмы. 1999. Т. 37, № 3. С. 123.
12. Именитов И. М., Чубарина Е. В., Шварц Я. М. Электричество облаков. Л.: Гидрометеоиздат, 1971.
13. Uman M. A. // Intern. Geophys. Series. 1987. V. 39. P. 377.
14. MacGorman D. R., Rust W. D. The electrical nature of storms. Oxford Univ. Press, 1998. P. 304.
15. Marshall T. C., Rust W. D. // J. Geophys. Res. 1991. V. 96. P. 22 297.
16. Wang D., Shimada Y., Uchida S., Fujiwara E., Kawasaki Z-I., Matsuura K., Izaka Y., Yamanaka C. // Noise and Chaos in Space and in the Laboratory / Ed. by H. Kikuchi. N. Y.: Plenum Press, 1994. P. 313.
17. Федер Е. Фракталы. М.: Мир, 1991.
18. Clar S., Drossel B., Schwabl F. // Phys. Rev. E. 1994. V. 50. P. 1 009.
19. Obukhov S. P. // Physica Ser. A. 1980. V. 101. P. 145.
20. Соколов И. М. // УФН. 1986. Т. 150, № 2. С. 221.
21. Mazur M., Krehbiel P. R., Shao X. M. // J. Geophys. Res. 1995. V. 100. P. 25 731.
22. Stiozenburg M., Rust W. D., Smull B. F., Marshall T. C. // J. Geophys. Res. D. 1998. V. 103, No. 12. P. 14 059.
23. Vecchi G., Labate D., Canavero F. // Radio Science. 1994. V. 29, No. 4. P. 691.
24. Labanne G., Richard P., Bondion A. // Lightning Electromagnetics / Ed. by R. L. Gardner.

Институт прикладной физики РАН,  
г. Нижний Новгород, Россия

Поступила в редакцию  
20 апреля 2001 г.

**FRACTAL STRUCTURE OF THE NONLINEAR DYNAMICS OF ELECTRIC CHARGE IN A THUNDERCLOUD***D. I. Iudin and V. Yu. Trakhtengert*

We study the fractal dynamics of the intracloud space charge during the preliminary stage of a lightning discharge, during which the so-called drainage system of charge collection from the entire cloud volume is formed. The drainage system is a basis of the initiation of the leader channel of a lightning. This system is formed due to the appearance of electric microdischarges related to the nonlinear stage of the development of an instability of the beam-plasma type. The latter is modeled using the two-dimensional lattice of finite-state automata. The results of numerical simulations show that the developed drainage system belongs to the percolation-cluster family. We point out the parameter region relevant to the proposed model, in which the thundercloud exhibits the behavior corresponding to the regime of self-organized criticality. A method of calculating the radio emission from the intracloud microdischarges is proposed, which allows one to compare directly the theoretical model with experimental data.

УДК 523.03

## SOME PROBLEMS OF INFORMATION NEURODYNAMICS

*M. I. Rabinovich,<sup>1,2</sup> R. D. Pinto,<sup>1</sup> and R. Huerta<sup>1,3</sup>*

The goal of neural science is to understand the brain, how we perceive, move, think and remember. All of these things are dynamical processes which are taking place in a complex, non-stationary and noisy environment. This means that these dynamical processes at all levels from small neural networks to behavior should be stable against perturbations but flexible and adaptive. The goal of neurodynamics is to formulate the main dynamical principles which can be a basis of such behavior and to predict the possible activities of neurons and neural ensembles using the tools of nonlinear dynamics. In this paper we discuss our last results related to the mostly challenging part of neurodynamics: information processing by dynamical neural ensembles.

Will it ever happen that mathematicians  
will know enough about the physiology of the brain,  
and neurophysiologists enough of mathematical discovery  
for efficient cooperation to be possible?

*Jacques Hadamard*

### INTRODUCTION

**1. Neuroscience and nonlinear dynamics.** The last decades have seen increasing efforts of neurophysiologists and neuroscientists to use dynamical models as a key element to understand how different neural systems function [1–6]. It is clear now that detailed physiological data alone are not sufficient to understand how even simple neural systems work. Experimentalists need a qualitative approach to the dynamical theory. Dynamical modeling can be important for prediction of the fundamental consequences of neural behavior. In fact, the new branch of Science that has been started to be built in the last few years is called Neurodynamics.

Nowadays, after 100 years of intensive studies of the nervous system, both experimentalists and theoretician agree that nerve cells and synapses are functional elements that process the information about the environment and control the behavior of the animals [7]. There is a lot of experimental evidence that neurons and neural ensembles behave as dynamical systems [8–11]. The major challenge of the dynamical approach arises from the diversity of neurons, synapses, network topologies and functions. A single neuron stands in the midst of a controversy of modelers deciding the level of detail to address in their specific problem. Many scientists think that details of a single neuron are superfluous when neurons operate in a large network. So, very simple input-output functions can be used to represent the neuron. On the other extreme, you can find researchers who model very carefully the morphology and electro-chemical details of the neurons to reveal how the various kind of dendrites, axons, membrane channels, synapses fit together to create a realistic functional element. The question is then: Which one of the two extreme approaches is right? We have no universal answer. From our point of view what really matters is to model functions and phenomena like synchronization, competition, rhythmic activity responsible for swimming, running, breathing and so on. Despite the variety of neurons and network topologies the phenomena and the functions are universal.

What is the nature of this universality? From the point of view of dynamics we can answer this question using the experience and the knowledge accumulated in the classical theory of oscillations, in particular the ideas put forth by Andronov in 1931 concerning the structural stability of dynamical systems and the investigation of their bifurcations. We briefly review the essential points of these ideas, the roots of which are traced back to Poincaré. In his book "The Validity of Science" (1905) in the chapter entitled "Analysis and Physics," Poincaré wrote that "The main thing for us to do with the equations of mathematical physics is to investigate what may and should be changed in them." Indeed, any description of real life is a model, and in modeling the dynamics of nonlinear systems one is usually confronted with ordinary or partial differential equations containing different nonlinear dependences. Ideally it would be great to obtain their general solutions and thus predict the behavior of a given model subject to specific initial conditions. However, it is impossible and even useless. Andronov's remarkable approach toward understanding such systems contained three key points:

1) Only those models demonstrating motions which do not vary with small changes of the parameters can be regarded as really interesting ones (Andronov referred to these as models or dynamical systems that are structurally stable);

2) To gain insight into the dynamics of the system means clarifying all principal types of its behavior under all possible initial conditions, i.e., one should analyze the behavior of the model as a whole rather than to find particular solution under specific initial conditions (this is why Andronov was fond of the methods of phase space analysis);

3) Consideration of the behavior of the system as a whole allows one to introduce the concept of topological equivalence of dynamical systems, and requires an understanding of local and global bifurcations as control parameters are varied.

Conservation of the topology of phase space partitioning into trajectories corresponds to a qualitatively stable motion of the system with small variation of the governing parameters. Partitioning of the space of parameters into regions with different behavior then furnishes a complete picture of the potentialities of a dynamical model [12, 13]. Thus the reasonable dynamical model of the neural unit has to capture the mostly essential structurally stable features that are required to reproduce the global behavior of the unit, instead of replicating the details.

Neurodynamics is distinguished from classical nonlinear dynamics by an explicit focus on how the neural systems process the information. Dynamical description, representation, processing of the information, learning and memory is a extremely promising way for an interpretation of the experimental data and the formulation of the direction of future experiments.

What kind of dynamical system are the neurons and neural networks? We proceed to answer this question using the following lessons learn from the experiments. First of all, as a rule, the response of neurons to incoming signal is reproducible; second, the behavior of neural groups is robust against fluctuations; and third, the action of the neuromodulators or sensory inputs is able to change qualitatively the neurons or neural networks dynamics.

Based on these messages, we can formulate the general features of the neural ensembles as a dynamical system (DS) in the following way:

(i) Strongly dissipative DS that has a short time memory about the initial conditions and, therefore, the responses of the system to the same input of different experiments need to be the equivalent (reproducibility).

(ii) The trajectories corresponding to the response of the system to the action of the stimuli need to have large basins of attraction (robustness).

(iii) DS must be nonlinear, nonequilibrium, and sufficiently complex to be able to separate many different inputs.

**2. Neural dynamics and information.** Now we can state the key questions related to the neural dynamics and information transduction and encoding as follows:

(i) How does the changing behavior of the DS influence the information that every neuron exchanges with each other?

(ii) How does the cooperative dynamics of interacting neurons encode the sensory information?

(iii) What is the role of the dynamical complexity and time as a variable in the representation of the information by neural ensembles?

The complete answers to these questions are not clear yet. Nevertheless some existing results are very important and promising.

Before we discuss these results let us to emphasize the main differences between living neural systems and communication channels traditionally considered in classical Shannon's information theory [14]. In classical theory the set of possible symbols (coding space) is a priori defined and the "state" of the recipients is not supposed to change. The information content of a message  $x$  with prior probability  $p(x)$  is  $I(x) = -\log p(x)$ . This quantification of information assumes some prior knowledge on the part of the recipient characterized by the probability distribution  $P$  for all the messages  $x$ . That means that the receiver knows in the statistical sense the composition of the character set to be used. The classical theory always refers to an ensemble of possible events and analyses the uncertainty with which the occurrence of these events is associated. The best way to describe the ability of the communication channel in the framework of such theory is the input-output statistical description [15, 16] which does not take into account the intrinsic "dynamical life" of the channel.

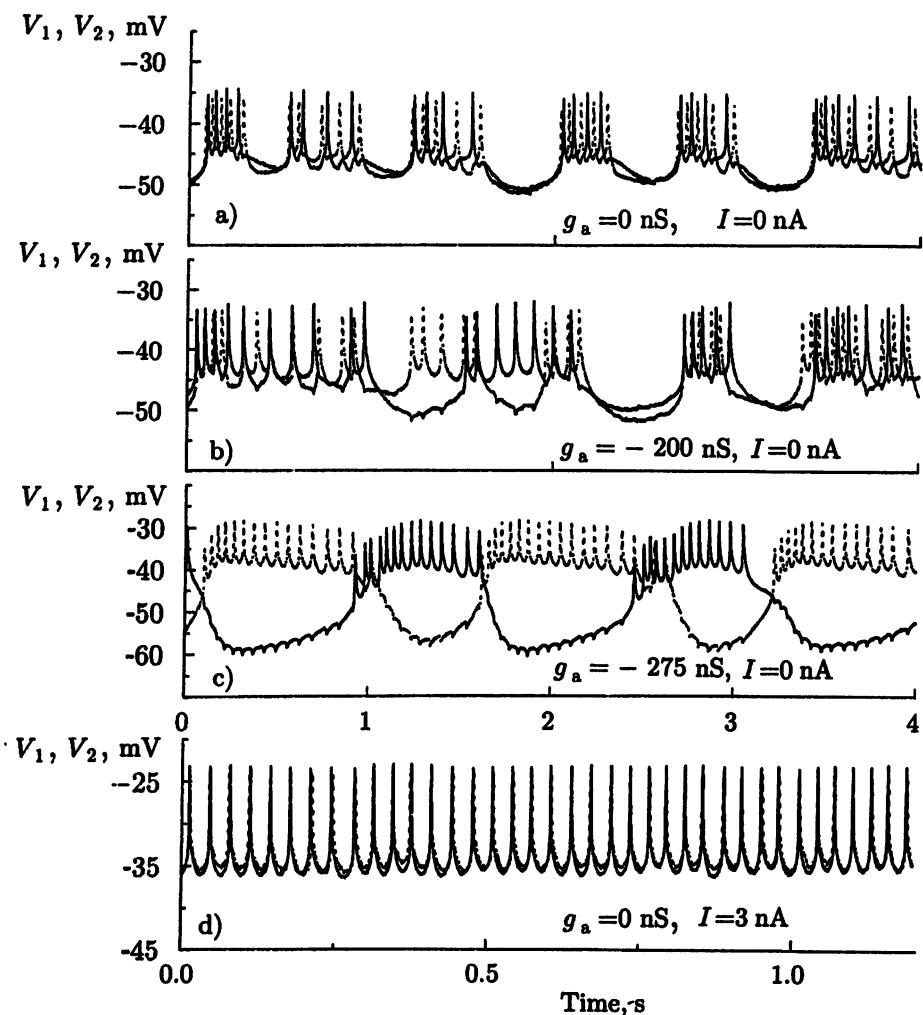
The input-output description of the nervous system can be useful in many ways [15, 16], but it can be improved for the following reasons: (i) In neural systems the set of possible symbols and coding space are not fixed a priori and the system may form the coding space itself; (ii) The neural code is changed at different stages of the information transduction, or, in other words, the system itself is able to transform one coding space to another one. Such transformation is the result of nonlinear dynamical activity of neurons and neural network as a whole under the action of the stimulus; (iii) In contrast with the classical communication channels a neural network is able, as a nonequilibrium system, to create new information which contains the information about the autonomous dynamics of the network. The simplest example is the chaotic behavior of neurons under the action of periodic stimulus; (iv) The nervous system is not fixed: its dynamical features and informational abilities are changing in the learning process. After learning, the way the information is encoded can change.

## 1. GENERATION OF THE SPATIO-TEMPORAL INFORMATION BY CPGs: SYNCHRONIZATION AND COMPETITION IN SMALL NEURAL NETWORKS

In the early 70s it was shown that the highly repetitive patterns of rhythmic motor activity of invertebrates could be sustained without any sensory stimulus, without any external influence or without neural signals from higher functions. These patterns are programmed by central pattern generators (CPGs). If a CPG is removed from an animal's body and placed in a salt solution that keeps the cells alive, this CPG may still generate essentially normal motor-output patterns for as long as many hours. There are several basic "minimal circuits" of neurons that are known to generate characteristic oscillatory behaviors. These circuits are responsible for the fast onset of synchronous behavior, rhythmic activity and regularization of neural signals. To understand the origin of the synchronization and competition phenomena in neural ensembles we have to begin from the simplest circuits: two neurons coupled together.

### 1.1. Minimal CPG circuit

Experimental studies of synchronization and competition in a pair of neurons that interact through naturally occurring electrical and inhibitory coupling have been reported in [9, 10]. Some of these experiments



**Fig. 1.** Regimes of oscillations in two coupled PD neurons from the stomatogastric ganglion of the California spiny lobster for different coupling conductances  $g_a$ . The first three rows show the bursting behavior of the two neurons with different levels of synchrony. The last row shows synchronous spiking behavior [9, 10],  $V_1$  and  $V_2$  are the membrane potentials of different PD neurons

were carried out on two pyrolic dilator (PD) neurons from the pyloric CPG of the California spiny lobster. Let us discuss in more detail the case of electrical coupling.

The strength of the natural electrical coupling can be altered during the observations by use of a feedback device. Individually, these neurons can generate highly irregular spiking–bursting activity (Fig. 1b). Varying the control parameters (injected DC current and interneuron conductance) we found the following regimes of behavior [9, 10].

Natural coupling produces state-dependent synchronization as shown in Fig. 1. With little or no applied current, the neurons fire spikes in irregular bursts in which the slow oscillations are well synchronized while the spikes are not; this is shown in Fig. 1a. Changing the magnitude and sign of electrical coupling restructures the cooperative dynamics. Increasing the strength of coupling produces complete synchronization of both irregular slow oscillations and fast spikes. Compensating the natural coupling of about 200 nS leads to the onset of independent irregular pulsations as in Fig. 1b. With net negative coupling, the neurons burst in antiphase, but now in a regularized pattern as in Fig. 1c. When depolarized by a positive DC current, both neurons fire a continuous pattern of synchronized spikes as we show in Fig. 1d. In this figure,  $g_a$  is the

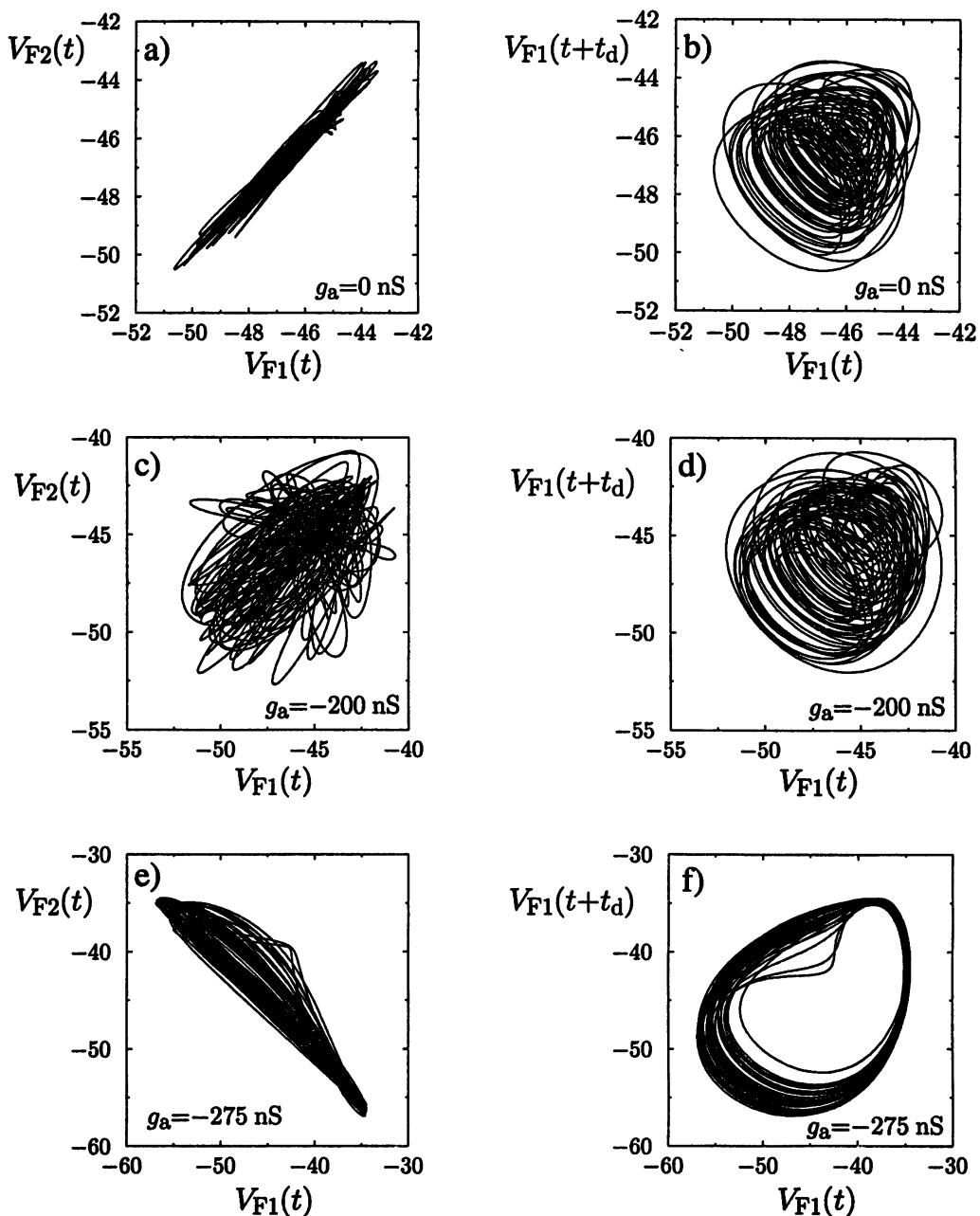


Fig. 2. Phase portraits of the slow components of oscillations in two coupled PD neurons as a function of the external conductance. The coordinates  $V_{F1}(t)$  and  $V_{F2}(t)$  are the filtered membrane potentials of the two neurons [9] and  $t_d = 0.3$  s. There is no externally injected DC current here

externally controlled conductance level.

The dynamics of slow oscillations changed as the external coupling conductance  $g_a$  was altered. With natural coupling  $g_a = 0$  nS the slow oscillations stayed synchronized as seen in Fig. 2a even though each neuron displays very complex dynamics as shown in Fig. 2b or Fig. 1a. Additional dissipative coupling ( $g_a < 0$  nS) led to desynchronization. The desynchronized slow oscillations remained complex and aperiodic as we see in Figs. 2c and 2d (see also Fig. 1b).

Adding further negative coupling conductance which could represent an inhibitory synaptic connection

caused the neurons to compete with each other and behave in an antiphase manner as seen in Fig. 2e. This regime of antiphase behavior was characterized by the onset of more regular, “almost periodic” bursts as we see in Fig. 2f.

We now consider the competition between neurons in more detail. Spatio-temporal patterns of neural activity can be generated by competition mechanisms among the cells. Competition means that several units are active at the same time and through inhibition between the component neurons, even with simultaneous excitation, their states alternate as in the antiphase bursting of our two PD neurons coupled electrically with negative coupling as shown in Fig. 2e. The results from another experiment in our laboratory with two different neurons (lateral pyrolic (LP) and PD from the pyloric CPG with an inhibitory connection from PD to LP) show that the neurons burst in a nearly periodic alternating temporal pattern and their individual chaotic activity is regularized [9]. When the polarity of one of the mutual connections is changed to excitation, regularization of the bursting behavior is lost.

Inhibitory synaptic connections between neurons appear to have a distinctive, even critical role to play in neuron assemblies. This type of connection between nonlinear oscillators is not typically found in physical systems, and this lesson from biology in itself represents an important new direction for the dynamical systems study of collections of nonlinear oscillators.

The experiments we have just described indicate that the slow bursting oscillations and the fast spiking oscillations of these two neurons have different thresholds for the onset of synchronization. This can be understood in terms of the different spatial sites of origin of the two types of voltage signal, the different mechanisms of synchronization, and the different conduction pathways and attenuation factors involved. The slow voltage oscillations that underlie bursting activity arise as a result of voltage-dependent ion channel activity in the membrane of neuropilar processes. The summed voltage signal will suffer some attenuation as it spreads by local current flow in the leaky cable array of the neuropil. However, two factors favor its effective transmission between the neurons: (a) the location of electrical coupling sites close to the site of slow wave generation, and (b) the slow time course of the voltage signal itself. In combination, these should allow a relatively strong and continuous interaction between the irregular slow oscillators. This mechanism resembles the synchronization seen in dissipatively coupled chaotic electrical circuits [17, 18]. In contrast, fast spike signals suffer strong attenuation as they spread between the spike initiation zone at the origin of the axon and the coupling sites in the neuropil. These factors argue for weak current flow between spike generators. If the spike generator of a neuron is close enough to its threshold, the transient current from the coupling pathway may drive it to phase-locked firing. In electrical circuits, this type of chaotic pulse synchronization is known as threshold synchronization [19]. With natural coupling, these threshold mechanisms can synchronize spike activity in tonic firing but not in the bursting regime. When the neurons generate slow voltage oscillations, ion channels open in neuropilar processes, decreasing the membrane resistance. This shunts the spike-evoked currents as they flow in their coupling pathway, causing a failure in threshold synchronization.

As the strength of net coupling is decreased, the slow oscillations remain irregular with little change in waveform, but make a sharp transition from synchronous to asynchronous behavior (see Fig. 2). When the net coupling reaches an expected, negative conductance, the slow oscillations resynchronize in antiphase and become regular. These bifurcations argue for a dynamical origin of the irregular neural activity.

## 1.2. Information as a global measure of the CPG activity

The inner properties of every neuron in a CPG together with the connection topology between them determines the phase relationship in the electrical activity that commands the muscles responsible for activities like chewing, walking and swimming [20].

A group of neurons can generate many different spatio-temporal patterns of activity, but only a small

subset of the possible solutions of activity of the CPG will work for a given mechanical device. Since the number of different spatio-temporal patterns that can be generated by a set of neurons connected using different topologies bears infinite, it should be very useful to find a way to reduce the large number of possible solutions without knowing a priori the specific function that the CPG is supposed to carry out [21, 22].

We discuss here the use of Shannon and Weaver information [14] as a global measure of the efficiency in the CPG activity as a method able to point out in a large ensemble those specific solutions found for some particular CPGs.

Since we do not intend to insight into information encoding and transmission between arriving spikes and the neural outputs we consider the input to the CPG as a set of synaptic conductance configurations  $g_{ij}$  [2]. In most of the CPGs, the rhythm changes due to neuromodulatory input that modifies the synaptic connections between neurons and the ionic conductances of the cell [3, 4, 23], one input generates a specific synaptic pattern and a set of conductance values. For simplicity we consider only variations in the synaptic connections. Therefore the input to the system is given by

$$\text{In} = \{g_{ij}; 0 \leq g_{ij} < g_{\max}, i \neq j; i, j = 1, \dots, N\}, \quad (1)$$

where  $g_{\max}$  is the maximum synaptic conductance and  $N$  is the number of neurons in the CPG.

Since the spikes of a neuron are carried through the axon to the muscles that integrate them to produce an action we disregarded in our calculations the frequency of the spike firing and consider that the information carried to the muscles is mainly contained in when the burst starts and ends.

Our CPG elements are chaotic elements based on the Hindmarsh–Rose model [24]. Random sets of input configurations implemented both in computer simulations and in electronic analog model neurons [25] were used to integrate the CPG so the conditional distributions and entropies could be calculated. The estimation of the maximum of the mutual information lead us to a subset of configurations in In, and we found out that a 15 % of the input configuration set were sufficient to obtain the maximum of the mutual information, that all of these configurations lead to a regular rhythmic activity (as observed in real CPGs), and all of them are closed topology configurations. This last property is widely observed in CPGs where there is not a neuron that do not receive any feedback from any other neuron in the CPG (see, e.g., [26–28]).

To estimate the information we acquired long time series of the membrane potential of all the neurons in the CPG for a given  $G_j \in \text{In}$ , where  $j = 1, \dots, M$  and  $M$  is the cardinality of the set In. We attributed to all neurons two counter arrays,  $C_{G_j}(R, t)$ , of length  $T$  sufficiently large (one for beginning of bursting and one for end of the bursting activity) and  $R$  represent the specific neuron and the event considered (beginning or end of bursting). We choose one of the neurons as a time reference, every time the reference neuron starts bursting the reference time was set to zero and we started looking for the beginning (end) of a burst in all neurons increasing their beginning (end) of burst counter array in the corresponding time. When the integration is finished in all of the time series we normalized the counters to 1 to obtain the conditional probabilities  $p_{G_j}(R, t)$ . We determined the values of the conditional entropies as follows:

$$H_{G_j} = - \sum_{t=1}^T \sum_{R=0}^{2N} p_{G_j}(R, t) \log_2 p_{G_j}(R, t). \quad (2)$$

By Bayes we know that the joint probability

$$p(R, t, G_j) = p_{G_j}(R, t)p(G_j), \quad (3)$$

where  $p(G_j)$  is the probability of the configuration  $G_j$  in the set In. The marginal probability

$$p(R, t) = \sum_{j=1}^M p_{G_j}(R, t)p(G_j), \quad (4)$$

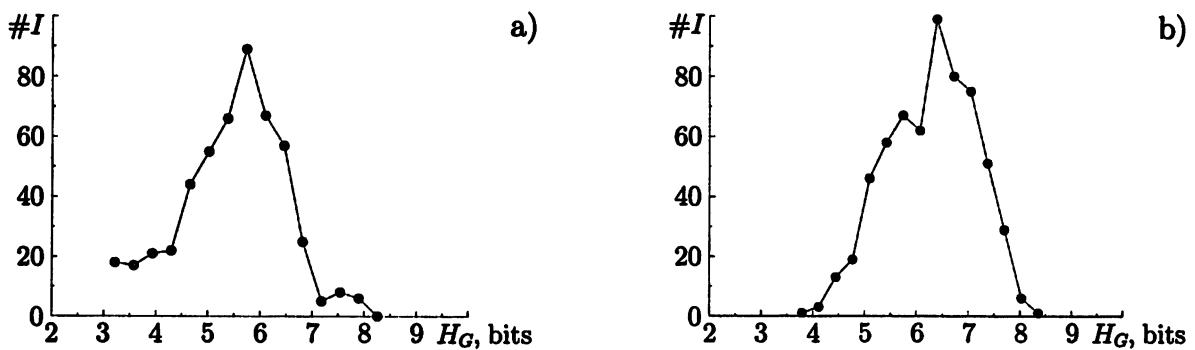


Fig. 3. Distribution of inputs with specific values of the conditional entropy  $\#I$  in  $[H_G - \epsilon, H_G + \epsilon]$  for  $\epsilon = 0.225$  in a 3-neuron CPG (a). Input distribution in a 4-neuron CPG for  $\epsilon = 0.58$  (b)

which is used to calculate the output entropy  $H(O)$  as

$$H(O) = - \sum_{t=1}^T \sum_{R=0}^{2N} p(R, t) \log_2 p(R, t) \quad (5)$$

and the conditional entropy

$$H(O/\text{In}) = \sum_{j=1}^M p(G_j) H_{G_j}. \quad (6)$$

In order to maximize the mutual information as a function of  $p(G_j)$  (we rename  $p(G_j) \equiv x_j$  and  $H_{G_j} \equiv h_j$  for convenience) we calculated the gradient of  $MI(\text{In}/O)$  on the  $x_j$  space with restrictions  $\sum_j x_j = 1$  and  $x_j \geq 0$ .

The single neuron model used in our simulations and analog implementations is a modified version of the Hindmarsh–Rose (HR) that is known to generate chaotic behavior [24]. The model is made of three dynamical variables comprising a fast subset,  $x$  and  $y$ , and a slower  $z$ :

$$\frac{dx}{dt} = 4y + 1.5x^2 - 0.25x^3 - 2z + 2e + I_{\text{syn}}, \quad (7)$$

$$\frac{dy}{dt} = 0.5 - 0.625x^2 - y, \quad (8)$$

$$\frac{dz}{dt} = \mu [-z + 2(x + 3.2)], \quad (9)$$

where  $e$  is a constant injected (DC) current, and  $\mu$  is the parameter that controls the time constant of the slow variable. The parameters were chosen to set the isolated neurons in the chaotic spiking-bursting regime ( $e = 3.281$ ,  $\mu = 0.0021$ ).  $I_{\text{syn}}$  represents the current injected in the neuron after the onset of a chemical graded synapse. We considered only inhibitory synapses, which is the main kind of interconnection present in most of the CPGs. The synaptic current has been simulated according to a first order kinetic model of the release of neurotransmitter [5, 29].

In the analog implementation of the CPGs the electronic neurons were connected using a program developed at the INLS for simulating synapses using dynamic clamp protocol [29]. Both in analog and computer simulations for every trial of synaptic configurations 20 % of the synapses were randomly chosen to have their  $g_{\text{max}} = 0$  and the other 80 % of the synapses were chosen to have their  $g_{\text{max}}$  ranging uniformly from 0 to a maximum conductance (500 nS for the analog implementations and 1 for the computer simulations).

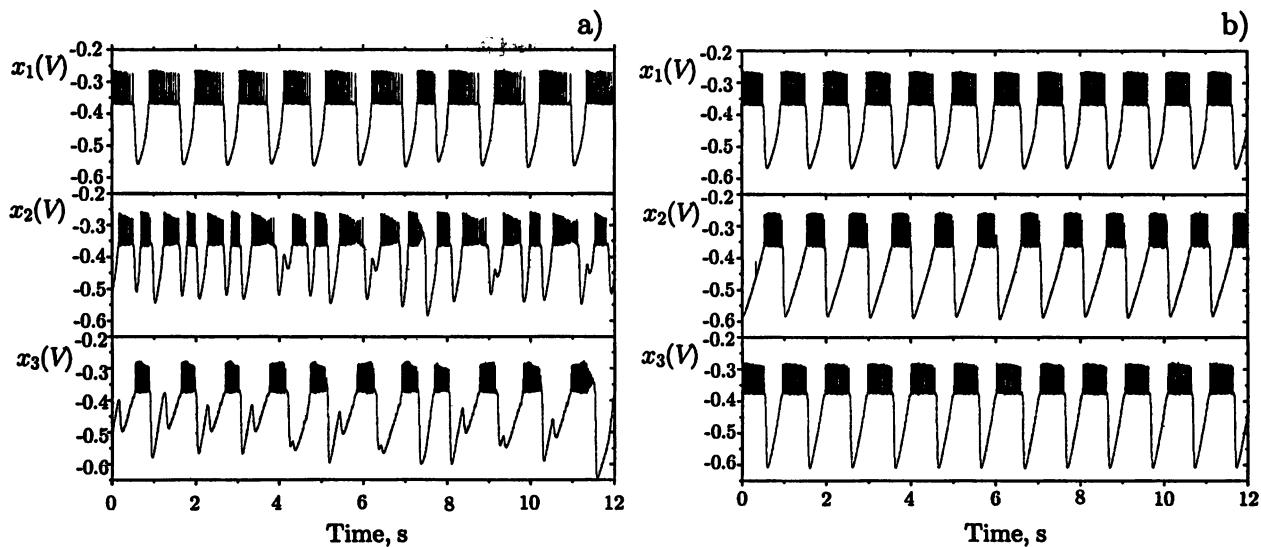


Fig. 4. Examples of time series for a 3-neuron analog CPG. Time series for a configuration  $g_{12} = 239$  nS,  $g_{13} = 374$  nS,  $g_{21} = 0$  nS,  $g_{23} = 277$  nS,  $g_{31} = 217$  nS, and  $g_{32} = 0$  nS, which generates a conditional entropy with a value of 6.44 bits (a). Time series for a configuration  $g_{12} = 277$  nS,  $g_{13} = 0$  nS,  $g_{21} = 260$  nS,  $g_{23} = 257$  nS,  $g_{31} = 0$  nS, and  $g_{32} = 443$  nS that generates a conditional entropy with a value of 3.37 bits (b).

Here we describe only the main results obtained for the analog computations but they are qualitatively similar to those obtained for the computer simulations.

We first performed our simulations in a CPG composed by three chaotic neurons where the connections between them were chosen at random as described. A pool of 500 random synaptic configurations were tested ( $\#I = 500$ ), their conditional probabilities and entropies calculated and the maximization of the mutual entropy  $MI(\text{In}; O)$  as a functions of  $p(G_j)$  was carried out. The same procedure was repeated for CPGs composed by four analog electronic neurons where we tested a pool of 630 synaptic configurations ( $\#I = 630$ ). In Fig. 3, we show the distribution of the conditional entropy values for a random set of In where all the elements were chosen from an uniform distribution for both three and four neurons CPGs.

We have most of the input configurations situated in a small range of entropies (between 5.0 and 6.5 bits). These values of entropy represent irregular activity (example time series in Fig. 4) and the values close to 3 bits represent regular rhythms.

The main results are summarized as follows. A small subset of configurations of In accounts for 99 % of the configurations that maximize  $MI(\text{In}; O)$ . The best preferred connectivity patterns are the ones with entropy values between 2 and 3 bits. These ones produced the most regular CPG activity than any others as can be seen in Fig. 4. Finally, all of these configurations are closed topologies. Most of the known CPGs have non open topology connections. Since the model neurons are chaotic, a neuron that do not receive any inhibitory feedback from any other neuron remains chaotic and the open configuration is not selected because it can not produce a regular spatio-temporal pattern.

## 2. INFORMATION PATTERNS IN LARGE NEURAL ENSEMBLES

### 2.1. Coarse grain patterns

We will discuss here time-periodic and spatially-ordered patterns that appears as a result of partial synchronization in randomly inhomogeneous media with chaotic local dynamics that is a very interesting

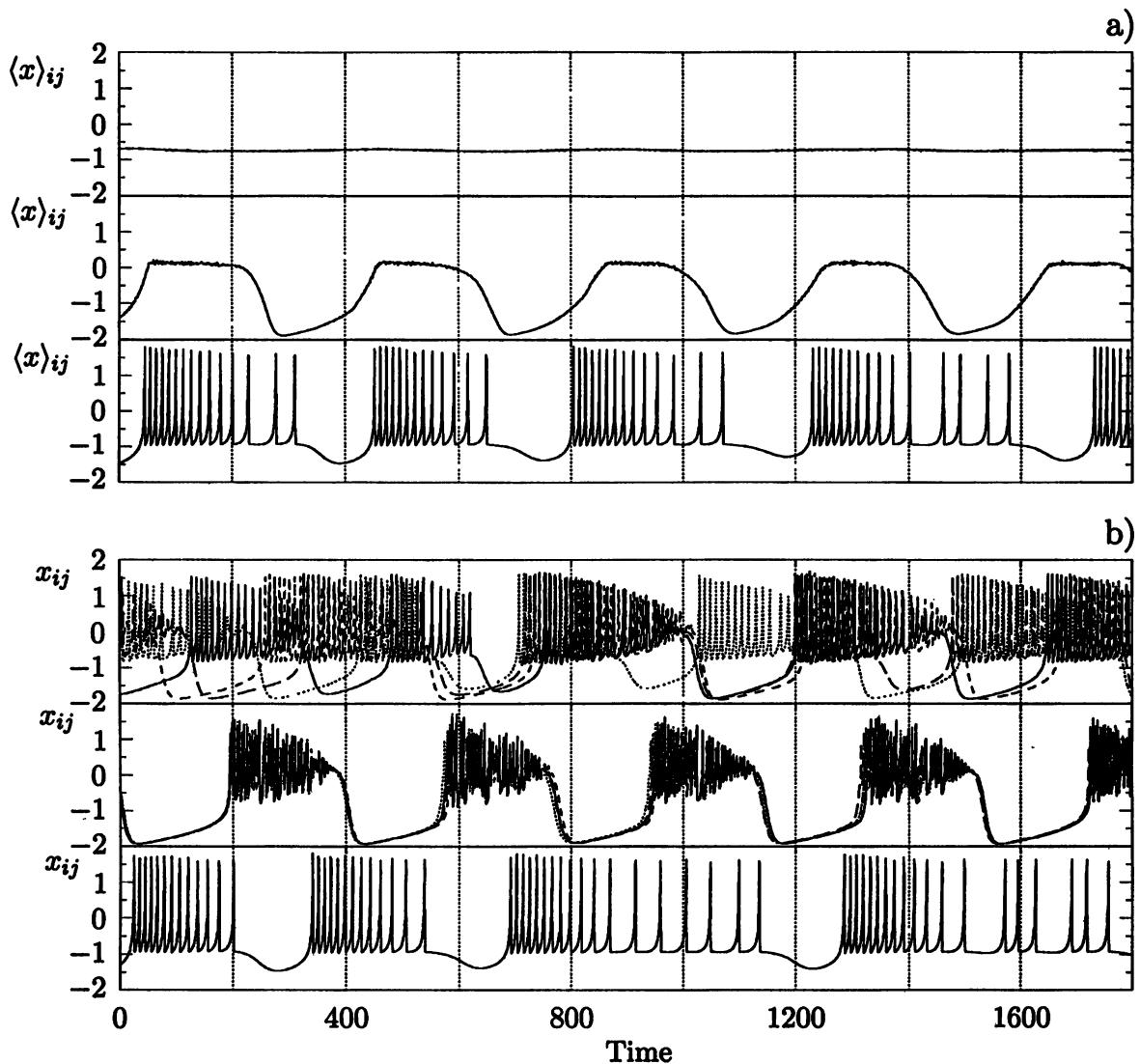
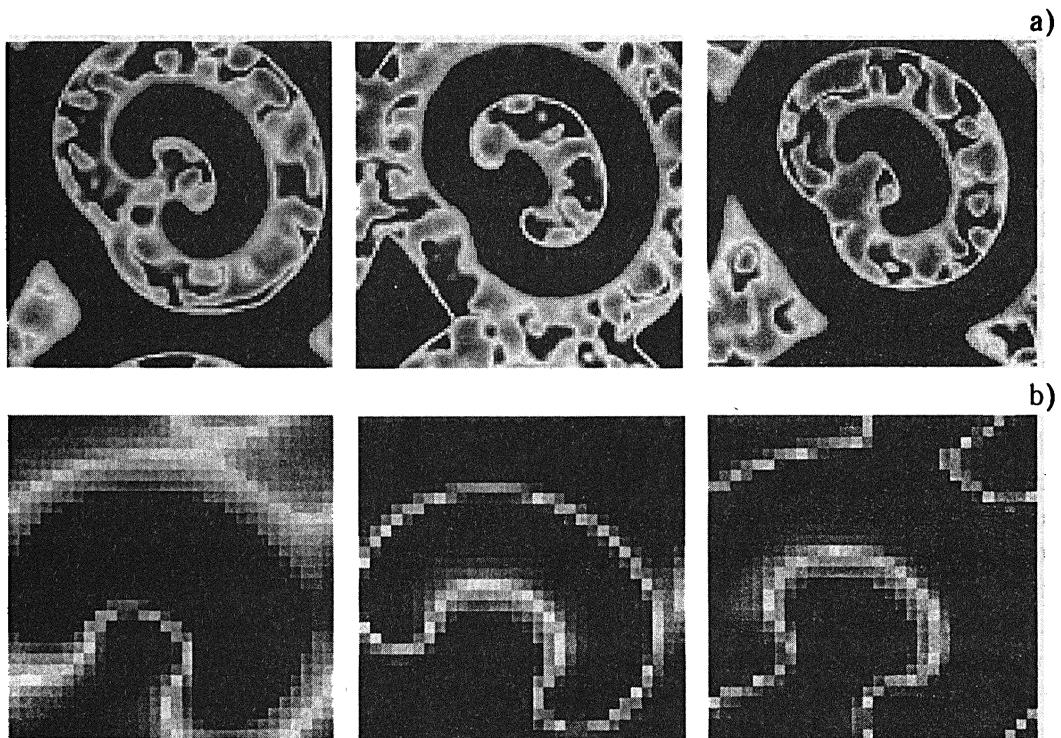


Fig. 5. Numerical simulations of the mean values  $\langle x \rangle_{ij}$  of all activities  $x_{ij}$  (a) and the time series of four different elements  $x_{ij}$  (b) described by Eqs. (10) for a lattice of  $100 \times 100$  Hindmarsh–Rose oscillators with different strengths of diffusive coupling: first rows  $g = 0.04$ , middle rows  $g = 0.4$ , and bottom rows  $g = 230$

and important problem, for example, in cellular dynamics. In particular, the continuous macromolecular oscillator functioning as a cell cycle attractor was observed in eucaryotic cells [30, 31]. Another important example is an ensemble of chaotic neurons [32].

The birth of coherent structures in such random media is paradoxical: every individual element of an ensemble of discrete elements pulsates chaotically and all of the elements are different. What can we expect when such oscillators are coupled to form a lattice medium? What kind of dynamics should an ensemble of chaotic elements with different parameters produce? The intuitive answer seems clear: it should be hyperchaos or developed turbulence. The mathematical image of the dynamical regime should be a chaotic set whose dimension increases with an increase in the number of elements in the ensemble. We consider a dissipative medium, therefore the mathematical image is a strange attractor with dimension of order of  $Nm$ , where  $N$  is the number of elements in the system and  $m$  is the number of positive Lyapunov exponents characterizing the chaotic dynamics of an individual element. Note that in the case of an unbounded



**Fig. 6.** Evolution of a periodic spatio-temporal pattern observed in a network of  $100 \times 100$  HR elements with  $g = 1.5$  (a). Periodic spatio-temporal patterns observed in a network of  $30 \times 30$  coarse-grain elements computed for  $R = 0.23$  and  $G = 0.5$  (the rest of the parameters have the same values used in the HR lattice) (b). The value of  $R$  is close to the bifurcation point and the individual coarse grain dynamics is periodic. Taken from [35]

medium we have to consider not dimension, but dimension density (see [33] for further discussion of this topic).

To check this prediction, we exhibit some numerical results obtained for a two-dimensional lattice of Hindmarsh–Rose (HR) chaotic neural oscillators, diffusively coupled to each other through coupling parameter  $g$  (see Eq. (10) below). We exhibit results for four different elements in a network of  $100 \times 100$  elements subjected to weak, moderate, and strong coupling in Fig. 5. Fig. 5a shows the average value  $\langle x \rangle_{ij}$  of the activity of all  $100 \times 100$  elements and Fig. 5b displays the individual activity  $x_{ij}$  responses of the four elements. Our intuition about the existence of hyperchaos is vindicated in the first rows of Figs. 5a, b for weak coupling  $g = 0.04$ . For  $g = 230$  in third rows of Figs. 5a, b the coupling is so strong that all elements in the system are chaotically synchronized. For the moderate coupling value  $g = 0.4$  in the middle rows of Figs. 5a, b, it is observed that the average cooperative dynamics of the chaotic elements produces regular pulsating patterns.

It turns out that many features of individual element dynamics are key features in the birth of ordered patterns in chaotic media. Nevertheless, universality of the topology of coherent structures (spirals, targets, etc.) in different chaotic media needs explanation in every particular case. It is important to understand that universality itself is a nontrivial phenomena.

Let us first determine what surely has to be common for different chaotic media. As we know, usually generators of periodic oscillations with sufficiently close frequencies will synchronize if the coupling is not too small. This is possible for chaotic generators as well, but their synchronization is a chaotic synchronization. We recall the essence of this phenomenon. The main difference between chaotic synchronization and the synchronization of periodic motions is that the former is the coordination of unstable motions. The

images of these motions in phase space are saddle trajectories. Since almost all trajectories that form a strange attractor are saddle trajectories, synchronization of even identical oscillators is nontrivial: owing to arbitrarily small perturbations that are different in different oscillators, even identical systems will select different trajectories among those present in the phase space. The result of the interaction of identical elements will depend on the magnitude of the dissipative coupling. If the coupling  $g$  is large, then the difference of signals  $\mathbf{u} = (\mathbf{u}_i - \mathbf{u}_j)$  must tend to zero with increasing  $g$  [17, 34]. This occurs because when  $g$  is large, the difference signal obeys  $\mathbf{u} \sim \mathbf{u}_0 e^{-gt}$  and two oscillators will evolve identically, although still chaotically. This is complete chaotic synchronization.

The situation is much more complicated when the individual dynamical elements are different. They do not have identical trajectories, the motion along which could be coordinated by strong dissipative coupling trying to zero the difference signals, independent of initial conditions. New, though similar motions are born in nonidentical subsystems under the action of coupling.

Consider a lattice composed of many different chaotic generators that can interact locally, e.g., each generator is electrically coupled to its nearest neighbors. Suppose that the individual chaotic dynamics of an element is characterized by slow and fast pulsations. This behavior is typical of many kinds of neurons in small and large neural systems. At present the most popular model of such a neural oscillator is the HR model [24].

It is particularly interesting that in the phase space of this system there exists a strange attractor that emerges after a sequence of period-doubling bifurcations. This attractor exists in a finite region of the control parameter space  $(s, e, d, \mu)$ . We will describe here an investigation of a lattice made of disparate chaotic HR oscillators placed randomly inside the chaotic regime found through an appropriate choice of parameters. The dynamics of the square lattice system is described by the following set of coupled ordinary differential equations [32]

$$\begin{aligned}\frac{dx_i}{dt} &= y_i + ax_i^2 - x_i^3 - z_i + e_i - g \sum_j (x_i - x_j), \\ \frac{dy_i}{dt} &= b - cx_i^2 - y_i, \\ \frac{1}{\mu} \frac{dz_i}{dt} &= -z_i + s(x_i + d),\end{aligned}\tag{10}$$

where  $i = 1, \dots, N$ , the index  $j$  runs over the four nearest neighbors of unit  $i$  in the lattice,  $a = 3$ ,  $b = 1$ ,  $c = 5$ ,  $d = 1.6$ ,  $\mu = 0.0021$ ,  $e_i = 3.281 \pm 0.005$ , and  $s = 4$ . The modifiable parameter is  $g$ , the strength of the electrical coupling. We suppose that the boundary conditions are periodic.

If the coupling is strong, the chaotic behavior of the whole lattice is identical to the chaotic pulsation of an individual element as one can see in bottom rows of Figs. 5a,b. This example illustrates the role of chaotic synchronization in the birth of trivial patterns in inhomogeneous chaotic media: spatially homogeneous irregular oscillations. One can easily imagine that, for not too strong coupling, and when the spatial scale of synchronization is smaller than the size of the system yet larger than the distance between just two or three elements, the system might exhibit nontrivial spatial patterns — clusters of synchronization. Before we discuss this phenomenon let us analyze in detail the cooperative behavior of two coupled chaotic generators with slow and fast dynamics.

The nature of the patterns observed in a lattice of chaotic neurons (cf. Fig. 6) is connected with the presence of two distinct time scales in the neural oscillations: fast chaotic spikes rising out of the background of slow yet chaotic pulsations. The coherent patterns observed in numerical experiments, like those displayed in Fig. 6, vary periodically in time. Therefore, one of the principal questions that needs to be answered in order to clarify the nature of these patterns is: Are fast chaotic oscillations able to change the dynamics of large-scale collective motion by making it regular? By large-scale patterns we understand structures whose characteristic size is much larger than a cell of the lattice. These together with the pattern topology, which

appears like waves of switching between states in regular excitable media, leads to a possible mechanism for the formation of patterns at moderate values of neural coupling. This may be analyzed by introducing the concept of neuron clusters, as discussed below.

The cluster with the average time periodic behavior will be called a coarse grain (CG) [35]. We suppose that the regular spatio-temporal patterns observed in the computer simulations are strongly related to the existence of the coarse grain for a moderate value of the coupling. The cooperative behavior of diffusively coupled coarse grains (periodic oscillators in our case) produce many different regular spatio-temporal patterns similar to those obtained with the discrete analog of the complex Ginzburg–Landau or the FitzHugh–Nagumo models. To analyze this behavior we need an equation that describes the average dynamics of the coarse grain.

The coarse grain dynamics are described using the cluster variables

$$X(t) = \frac{1}{M} \sum_{i=1}^M x_i(t) = \langle x_i \rangle_{\text{cg}}, \quad Y(t) = \langle y_i \rangle_{\text{cg}}, \quad Z(t) = \langle z_i \rangle_{\text{cg}}, \quad (11)$$

where  $M$  is the number of elements in the cluster. An approximate system of equations for  $X, Y, Z$  is obtained by substituting

$$x_i = X(t) + \xi_i(t), \quad y_i(t) = Y(t) + \eta_i(t), \quad z_i(t) = Z(t) + \zeta_i(t) \quad (12)$$

into Eq. (10). Ignoring terms of order higher than  $\xi_i^2$  gives, after averaging over  $M$  elements, the governing equations

$$\begin{aligned} \frac{dX}{dt} &= Y + aX^2 + ar(t) - X^3 - 3Xr(t) - Z + \epsilon, \\ \frac{dY}{dt} &= -cX^2 - Y - cr(t) - b, \\ \frac{1}{\mu} \frac{dZ}{dt} &= -Z + s(X + d), \end{aligned} \quad (13)$$

where  $\epsilon = \langle e_i \rangle_{\text{cg}}$ . We have taken into account from the definition of  $X, Y$ , and  $Z$  that  $\langle \xi_i(t) \rangle_{\text{cg}} = \langle \eta_i(t) \rangle_{\text{cg}} = \langle \zeta_i(t) \rangle_{\text{cg}} = 0$ , and consequently the only function left to be determined is  $r(t) = \langle \xi_i^2 \rangle_{\text{cg}}$ .

In order to describe the slow dynamics, we need to make a reasonable assumption about the nonautonomous terms on the right-hand side of Eqs. (13). Since  $r(t)$  varies much more rapidly than the slow coarse-grain oscillation, we suppose that the dynamics of an individual coarse grain will depend on the time-averaged value of  $r(t)$  defined as

$$R(t) = \frac{1}{\tau} \int_t^{t+\tau} r(t') dt' \quad (14)$$

with  $t_r \ll \tau < T$ , where  $t_r$  is the characteristic time scale of the fast oscillation  $r(t)$  and  $T$  is the characteristic time scale of  $X(t)$ . In Eq. (13), we now replace  $r(t)$  by the slow function of time  $R(t)$  given by Eq. (14) which also depends on the strength of the diffusive coupling  $g$  between elements and on the size  $M$  of a coarse grain. If our hypothesis is correct,  $R$  is very nearly a nonzero constant for small values of  $g$ , and  $R$  is almost zero for large values of the coupling; for moderate values of the coupling, prediction of the behavior of  $R$  is not intuitively clear. Computer simulations at  $g = 0.1$ , however, indicate that for moderate values of  $g$  the behavior of  $R$  becomes periodic. This  $g$ -dependent behavior of  $R$  infers that the averaged dynamics  $X(t)$  will vary as the coupling parameter is varied.

For sufficiently small values of  $g$ ,  $R$  is nearly constant taking on values in the range 0.4–0.5, and only a single stable fixed point appears corresponding to steady-state behavior of the cluster. For  $R < R_c$  ( $g > g_c$ ),

this fixed point becomes unstable and the limit cycle in the three-dimensional phase space of the average coarse-grain system undergoes a supercritical and sharp Andronov–Hopf bifurcation to a stable fixed point. Strictly speaking, at the moment of this bifurcation  $R$  becomes a periodic function of time. However, as the numerical results confirm, slightly above the threshold for bifurcation, the influence of this periodicity on the existence of the limit cycle is not important.

The dynamical mechanism giving rise to the ordering behavior of the coarse grain relies on the synchronization and regularization of the activity of the  $M$  elements inside the grain.

Using the above observations, we are now in a position to explain the existence of large ( $N \gg 1$ ) regular spatio-temporal patterns in a discrete diffusive medium. First, the existence of regular structures is impossible in weakly diffusive media because local oscillations of neighboring elements are not correlated for small couplings  $g$ , and the mean field of the coarse grains becomes homogeneous and stable. Direct computation of the Kolmogorov–Sinai entropy confirms that the level of spatially homogeneous chaos increases as  $g \rightarrow g_0 \ll 1$ .

For moderate coupling, the coarse-grain assembly should exhibit regular spatio-temporal patterns. As confirmation of this conjecture we have checked the behavior of a lattice medium consisting of coarse-grain elements with slow periodic behavior. The description of this medium is analogous to that given by the network of HR elements wherein  $(x_i, y_i, z_i)$  are replaced by  $(X_i, Y_i, Z_i)$ . We are looking for patterns in the coarse-grain system that have the same space scale, relative to the size of the lattice, as the pattern in the original HR lattice. Thus, the pattern in the coarse-grain lattice should be of identical structure but with a smaller absolute size. Since the patterns on the two lattices have the same time scale, the speed of front propagation in the HR lattice must be larger than in the coarse-grain lattice. The propagation speed of the front increases with increasing values of the diffusion. We surmise, based on this scaling argument, that a coarse-grain pattern with the same relative size as the original may be found only in the case when the coarse-grain lattice coupling  $G$  is smaller than the diffusive coupling  $g$  of the original HR network. Verification of this conjecture is given by the sequence of patterns obtained for the  $30 \times 30$  network of coarse-grain units shown in Fig. 6. These patterns, plotted for  $G = 0.5$  in Fig. 6b, are clearly similar to those produced by the original heterogeneous lattice of chaotic HR elements for  $g = 1.5$  in Fig. 6a. Thus, identical periodic boundary conditions applied to both the HR and the square coarse-grain networks give the same topology of the patterns observed. Furthermore, additional computations have revealed the same correspondence of pattern topology when hexagonal lattices were used. In the latter case, the strength of the coupling was reduced to take into account the larger number (six) of nearest neighbors.

We conclude that the formation of large-scale coherent structures in nonequilibrium media consisting of discrete and chaotic HR elements with fast and slow oscillations exhibits two key features. The first is the regularization phenomena in small assemblies of chaotic elements, i.e., coarse grains. This regularization of behavior is the result of the action of the averaged activity of fast oscillations in the slow coarse grain dynamics. The second feature is the instability of homogeneously oscillating modes in a media considered to be a coarse-grain lattice. It is important to keep in mind that the coarse grains are a temporal assembly of neurons whose relaxation time is smaller than the relaxation time of the coherent structures.

## 2.2 Coherent patterns in chaotic networks with lateral inhibition

Another example of regular spatio-temporal behavior in a nonregular lattice described by Eqs. (13) and (14) is shown in Fig. 7. The origin of this regular behavior is absolutely different from that of the regular behavior of the coarse-grain patterns discussed in the previous section.

The behavior we shall discuss now is typical of negative electrical couplings that model inhibitory connections in neural networks [8]. This regularization phenomena is intimately related to the behavior of two

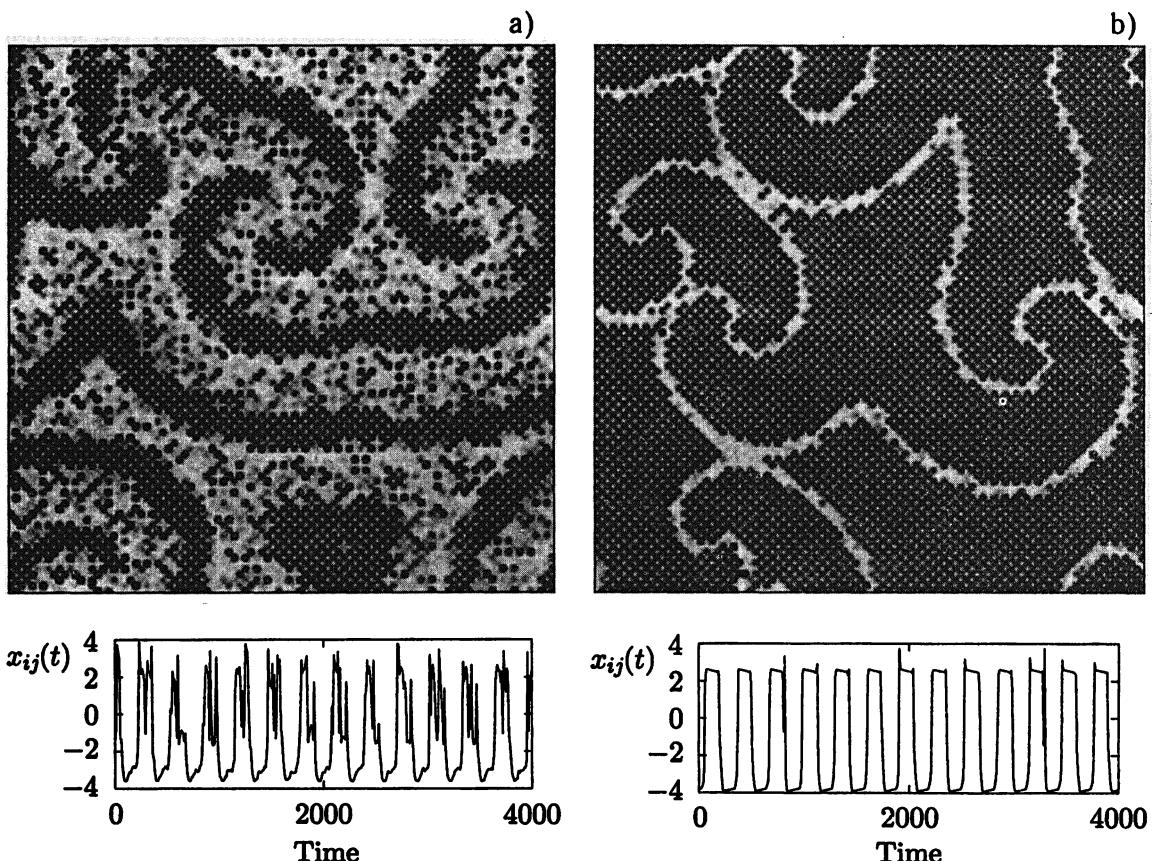


Fig. 7. Checkerboard patterns (top row) in a network of  $100 \times 100$  chaotic HR elements with negative electrical coupling between nearest neighbor units and the regular slow oscillations of a single neuron's activity (bottom row) for (a)  $g = -0.95$  and for (b) a stronger absolute value of the coupling,  $g = -1.12$ . The characteristic oscillation period is about 1 s. Taken from [35]

negatively coupled chaotic HR neurons. When the two oscillators are coupled with negative conductance ( $g \simeq -1$ ), the antiphase regularization is observed. The two neurons regulate their slow oscillations in the sense that the lengths and shapes of the bursts are kept uniform. This happens because the origin of the chaoticity of the model is related to the interaction of the fast subsystem  $(x, y)$  with the slow variable  $z$ : the homoclinic nature of the fast oscillations are regulated by the slow oscillations. In the absence of inhibitory action from other neurons to limit the rise of this slow variable, the system will be driven to a near homoclinic orbit which is unstable. Negative electrical coupling, however, will not permit individual neurons to reach a fast oscillation which is unstable.

When we have a lattice of such chaotic generators with negative coupling, they will form antiphase behavior with their nearest neighbors to form stable, regular checkerboard patterns as in Fig. 7. These types of regular spatio-temporal patterns are reminiscent of the regular envelope patterns of antiphase oscillation in the discrete variant of the complex Ginzburg–Landau model [36].

### 2.3. The robustness of patterns against noise

In our previous studies we dealt with systems free of noise (see [32]). However, neural systems in fact are influenced by large amounts of noise. It is interesting to investigate the effect of the noise on the spatio-temporal patterns. We investigated a lattice made of non-identical Hindmarsh–Rose neurons placed randomly inside the chaotic regime. Each element is electrically coupled to its nearest neighbors. The system

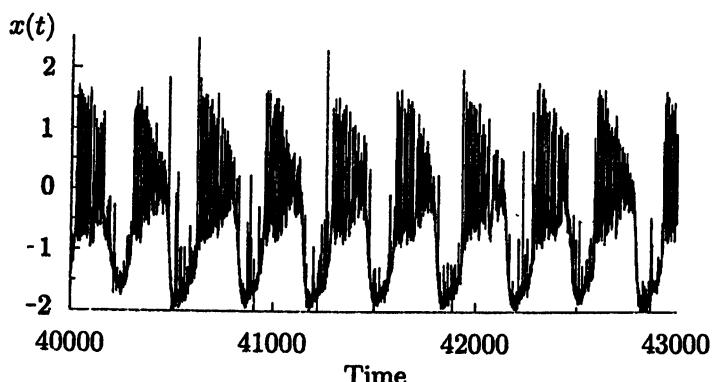


Fig. 8. Time series of the variable  $x$  of one randomly selected neurons in the lattice. This illustrates the high level of noise present in the network

is described by the set of coupled ordinary differential equations (13). We include additive noise  $\rho g_w(t)$  to the first  $x$  variable of these equations. Here  $g_w(t)$  is a Gaussian white noise with the following properties:  $\langle g_w(t) \rangle = 0$  and  $\langle g_w(t)g_w(t') \rangle = \delta(t - t')$ . First, we show in Fig. 8 the time series of one of the neurons of the lattice in the presence of value of noise  $\rho = 0.5$ .

In Fig. 9, we can see that the spatio-temporal patterns in the lattice survive even though the level of noise is extremely strong. We can conclude that the strong coupling prevents the patterns of synchronization from breaking.

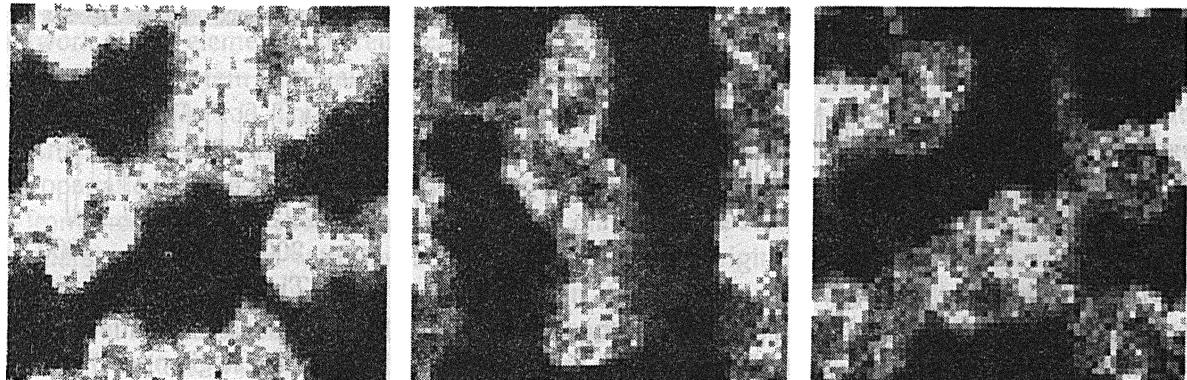


Fig. 9. Spatio-temporal snapshots of the HR lattice in the presence of value of noise  $\rho = 0.5$ . The snapshots are ordered by increasing time from left to right through one cycle

### 3. SPATIO-TEMPORAL REPRESENTATION OF THE SENSORY INFORMATION

#### 3.1. Computing with separatrices

There is a growing body of evidence [11, 37–39] that the information in neural systems is often recoded into a spatio-temporal format, where “space” is the identity of the neuron, then delivered to other nervous system functions. The use of time in encoding has a broad scientific interest ranging from understanding how real neural networks compute to insights on how nonlinear dynamical systems represent and transform information. Over the past decade there has been a developing interest in these spatio-temporal codes [6, 15], and our work will build on these developments.

One can imagine a variety of ways to transform a spatial pattern to a spatio-temporal code including a simple scanner which transforms a pattern into a pulse sequence. Experimental evidence suggests that nervous systems work as dynamical systems [6, 15]. We shall explore a class of dynamical systems called competitive networks or winner-less competition (WLC) networks. They produce spatio-temporal coding of temporal signals using deterministic trajectories moving along heteroclinic orbits connecting saddle fixed points or saddle limit cycles in the system state space. Instabilities of the nonlinear system are utilized to reliably encode and sensitively discriminate among sensory inputs.

We use observed features of biological sensory networks as a guide to computation using competitive networks. The experiments on which our ideas are based suggest the following features of neural encoding: the representation of input (sensory) information (i) employs both space and time; (ii) is nonperiodic in time; (iii) is deterministic and reproducible; (iv) is sensitively and uniquely dependent on the input stimulus; and (v) is robust against noise.

These observations suggest (a) that a dynamical system which possesses these characteristics should be strongly dissipative so that orbits rapidly “forget” the state of the system when the stimulus arrives, and (b) that the system should represent information by transient trajectories, not by attractors (regular or strange), of the unstimulated system. Indeed, a sensory system cannot have multistability because then noise could drive the system irreversibly to a new “representation.” On the other hand, a class of systems which we have been studying, exhibits transient near-heteroclinic trajectories which are robust against noise and always respond to a change in environmental stimuli in the same way. We will investigate the design of competitive information processing systems possessing the described features.

Let us illustrate the idea of “spatio-temporal coding with separatrices” on the following simple model,

$$\frac{dy_i}{dt} = y_i \left[ 1 - \left( y_i + \sum_{j \neq i}^N \rho_{ij}(S) y_j \right) \right] + S_i(t), \quad (15)$$

where  $y_i(t)$  is the intensity of the oscillations of one of the competitive elements, and  $\rho_{ij}(S)$  is the stimulus-dependent competitive matrix. This Lotka–Volterra model has been analyzed in detail for  $N = 3$  and  $S = 0$  in [40–42].

To begin, we ignore the additive sensory input in (15). This tells us how the network operates and allows an estimation of its encoding capacity. Information about the input resides solely in the couplings  $\rho_{ij}(S)$ , as in Hopfield models for which  $\rho_{ij} = \rho_{ji}$ . If the inhibitory connections are symmetric ( $\rho_{ij} = \rho_{ji}$ ) and identical ( $\rho_{ij} = \rho$ ,  $\rho_{ii} = 1$ ), the dynamics is very simple. For weak coupling,  $\rho \ll 1$ , the system has a global attractor, i.e., the stable fixed point  $y_i = [1 + \rho(N + 1)]^{-1}$  corresponding to simultaneous activity of all neurons. If  $\rho > 1$  the system is multistable: depending on the initial conditions, one neuron becomes active and the others are quiet. The nonsymmetrical case is more interesting and more realistic from the biological point of view. When the inhibitory connections are not symmetric, the system has different closed heteroclinic orbits that consist of saddle points and one dimensional separatrices connecting them. Such heteroclinic orbits are global attractors in phase space and are found in various regimes of the  $\rho_{ij}(S)$ . This implies that if the stimulus is changed, different heteroclinic orbits become global attractors.

To have a closed heteroclinic orbit, we need  $N \geq 3$ . Each saddle fixed point on the closed heteroclinic orbit must have one positive eigenvalue. The other  $N - 1$  are negative so that all other directions are attracted to the saddle. Movement from saddle point to saddle point along this sequence of unstable directions results. For this to occur when  $S_i = 0$ , the sequence from fixed point  $i$  to fixed point  $j$  must occur when  $\rho_{ii} = 1$ ,  $\rho_{ij} > 1$ , and  $\rho_{ji} < 1$  [43]. The only positive eigenvalue of the Jacobian at the fixed point is  $1 - \rho_{ji}$ .

One of the most important characteristics of the system is the number of different heteroclinic orbits that we can store in this system, that is, the capacity  $C$  of the system, which can be estimated as follows. If we satisfy the conditions for one heteroclinic orbit to exist, we can build another from it by permuting the indices of the  $y_i$  and of the matrix  $\rho_{ij}$ . There are  $N!$  permutations of the indices. Some of these generate the same heteroclinic orbit: firing as  $(1, 2, 3, 4, 5)$  or  $(2, 3, 4, 5, 1)$  is equivalent. For a given permutation there are  $N - 1$  permutations that are cyclically equivalent. The number of heteroclinic orbits involving all  $N$  neurons is  $(N - 1)!$ .

There are still more heteroclinic orbits. These are associated with the  $N - 1$ ,  $N - 2$ , ... dimensional subspaces which can be selected by eliminating one saddle point at a time from allowed orbits. The total

number of these is the capacity  $C$ :

$$C = \sum_{k=3}^N \binom{N}{k} (k-1)! = N! \sum_{k=3}^N \frac{1}{k(N-k)!}, \quad (16)$$

so  $C > \frac{N!}{N} \sum_{k=0}^{N-3} \frac{1}{k!}$  and  $C < \frac{N!}{3} \sum_{k=0}^{N-3} \frac{1}{k!}$ . For large  $N$ ,

$$1 - \frac{1}{e(N-2)!} < \frac{C}{e(N-1)!} < \frac{N}{3} \left( 1 - \frac{1}{e(N-2)!} \right). \quad (17)$$

Our networks are also robust against noise in the network present when the stimulus begins. This is achieved since each region is strongly unstable and a particular stimulus launches the trajectory in a particular direction much more rapidly than noise can deflect it to an alternative orbit. We have shown (not presented here) that quite high noise levels are required to overcome this important feature. The properties of these simplified models can also emerge from more realistic networks having the same dominance by asymmetrical inhibitory connections. We conjecture that a large network with sparse, random connections will effectively exhibit the same stimulus dependent sequential activation and deactivation of subgroups of neurons. Our central idea does not depend on the nature of the stimulus: It may thus apply to brain circuits other than olfactory processing system. It may perhaps underly interesting experimental observations such as the flipping between quasistationary states of activity seen in monkey cortex [39]. Beyond the biological observations which suggested these investigations WLC networks provide an attractive model for computation because of their large capacity as well as their robustness to noise contamination.

### 3.2. The role of synchronization on learning

There is a natural generalization of the previous model to understand the role of synchronization in learning. The generalized model of the network of oscillators has connections both in competitive (“inhibitory”) and supportive (“excitatory”) manners. We will use generalized Lotka–Volterra model discussed above (15) in the following form:

$$\frac{dx_i}{dt} = x_i \left[ 1 - \left( |x_i|^2 + \sum_{j \neq i}^N \rho_{ij}(\mathbf{S}) |x_j|^2 \right) \right] + \sum_{j \neq i}^N q_{ij} x_j. \quad (18)$$

Here  $x_i$  is the complex amplitude of the  $i$ -th oscillator,  $\rho_{ij}(\mathbf{S})$  is as before the stimulus-dependent matrix of inhibitory connections, and  $q_{ij}$  is the matrix of the excitatory connections.

Recent experimental results with sensory and cortical systems [44, 45] showed that neural network spatio-temporal dynamics are modified in a slow time scale when the stimulus is present. We argue that this kind of slow dynamics represent the mechanism for unsupervised learning in neural systems. The repeated application of a stimulus will allow an increasingly refined characterization of the input. The fundamental mechanism for such dynamical self-organization is claimed to be oscillatory synchronization. It is believed that the dynamical synapse increases its strength when the coupled neurons oscillate in coherent way. In the olfactory system of the locust the slow unsupervised learning enhances the discrimination between different spatio-temporal patterns [44]. To enhance the discrimination ability based on synchronization we propose the following slow evolution of excitatory connection matrix:

$$\frac{1}{\epsilon} \frac{dq_{ij}}{dt} = -q_{ij} + f(\mathbf{S}) + \gamma x_i x_j^*. \quad (19)$$

The last term in this equation is significant only if  $i$ th and  $j$ th neurons are synchronized. Thus, as an incoming stimulus via functions  $\rho_{ij}(S)$  and  $f(S)$  forms groups of synchronized neurons, the excitatory connections among them will strengthen according to Eq. (19), and these connections will remain even after the stimulus disappears. When a similar stimulus arrives again, the familiar pattern of synchronization will emerge much faster than before the original learning stage. We are planning to study a class of models of the form of Eqs. (18) and (19), because we believe they may bring new insights in understanding the role of synchronization for the unsupervised learning in complex systems.

#### 4. INFORMATION TRANSMISSION AND RECOVERY

##### 4.1. Stimulus-dependent propagation of information

The information propagation in nonequilibrium media has been addressed by a number of authors beginning from [46] (see also [47]). These authors used the information-theoretic concepts to characterize the information flow in spatio-temporal systems. The key remaining question is, how the information transmission depends on the statistical and dynamic features of the incoming signal. Let us illustrate this problem on the simplest chain of diffusively coupled neurons with complex dynamics which we model by the HR equations [24] (see also previous sections). The parameters of the elements are chosen such that, without an input signal, each neuron is in the periodic regime.

When the first neuron in a chain is excited by a stimulus, its dynamics changes to a specific temporal pattern of information. These changes propagate along the chain and thus transmit the input information. In our preliminary study we showed that the maximum distance at which the information can be recovered (synchronization cluster size) depends on the statistics or the dynamics of the incoming signal. We used as an input signal spike trains with varying inter-spike intervals  $t_i$ , with different entropies. An example of the stimulus depending clustering is presented in Fig. 10. As seen from the figure, the cluster size strongly depends on the entropy of the incoming signal. The higher the entropy, the shorter is the distance at which the signal can be detected, as determined by the mutual information analysis.

In order to see what happens when the stimulus event times are nonregular, we used Gaussian-distributed spike trains with different variances. Even with a small variance of 10, the message is less efficiently transmitted over the chain than for the periodic case (see Fig. 10). With increasing variance it gets even less efficient because the likelihood that the stimulus has an event while the first neuron is hyperpolarizing

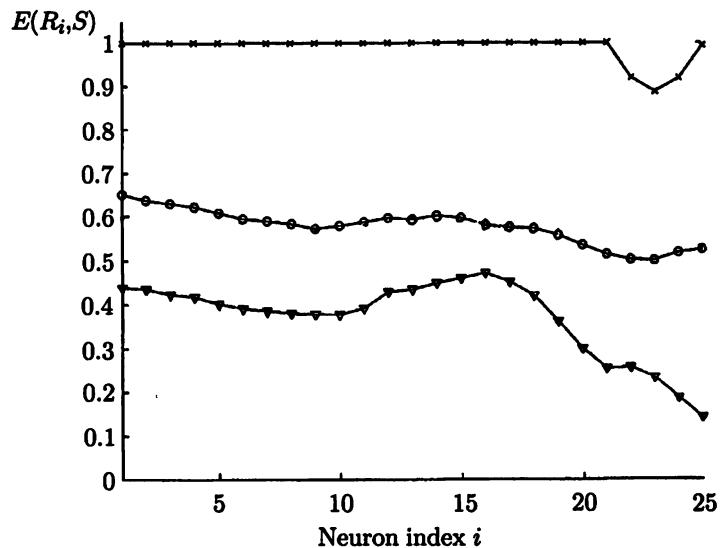
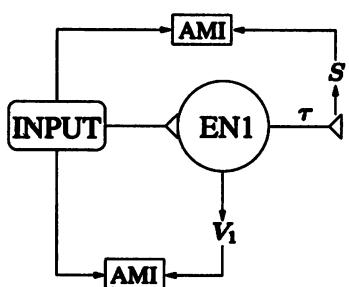


Fig. 10. Normalized mutual information between stimulus ( $S$ ) and  $i$ th neuron ( $R_i$ ) in an unidirectional coupled chain for three different statistics of the input signal, periodic spikes with inter-spike interval 400 (crosses), spikes with Gaussian distributed time intervals with mean value of 400 and variance of 10 (circle) and 200 (triangle). The entropy for each case is 0, 2.6, and 4 bits, respectively. The parameters of the neurons correspond to the regime of continuous spiking

increases, which means this event is overlooked by it. Since the excited neuron is kicked more arbitrarily with increasing variance, the deviations in the dynamics of neighboring neurons become larger, and the message is dissipated more rapidly along the chain. As a result, the normalized mutual information  $E(R_i, S)$  decreases stronger along the chain for larger variance, as can be seen in Fig. 10.

#### 4.2. Recovery of the hidden information by the dynamical synapse



**Fig. 11. Experimental layout.** The input signal consists of unimodal or bimodal Gaussian distributions of interspike intervals. The input induced by the spike sequence is generated using a digital to analog converter controlled by the PC. The input signal inhibits the neuron EN1 through an analog electronic model chemical synapse. The time series of the current synaptic activation  $S$  is used for the results described later

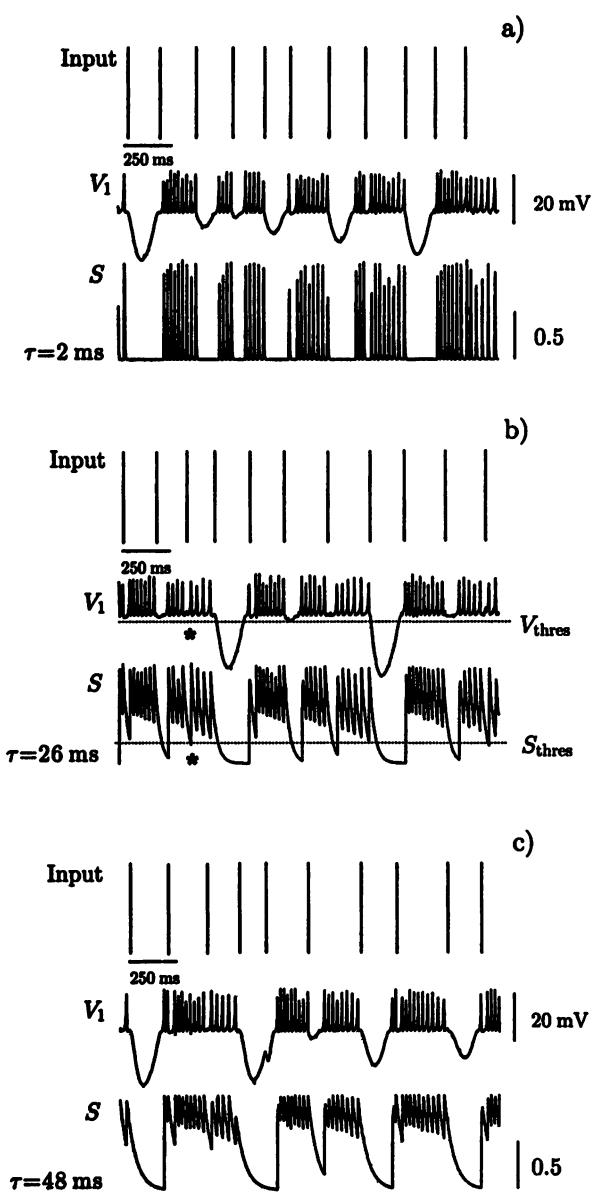
When information is transmitted through a neural medium, the modulation (information) can be carried from one variable to another one. It can be transformed from one coding space to another which can render it unreadable using membrane potential coding spaces. This is specially important if the information channel has a chaotic neuron. Nevertheless, the information that is not lost in this transformation can be recovered using specific nonlinear dynamical elements as decoders [48]. The main dynamical element that can recover the information hidden in the chaotic bursting model neuron is the synapse. We use the average mutual information (AMI) to show that the AMI between an input signal and a synaptic output is larger than the AMI between the input and the output of the presynaptic bursting neuron. This appears to violate the data processing inequality. However, it is not the case. It seems evident that a neuron with chaotic dynamics does not improve the transmission of the information but vice versa mask it making hidden. It is because the neuron generates the information about itself that has nothing related to the information carried by incoming signal. However, the dynamical synapse with appropriate parameters is able to recover the signal that has been scrambled by the chaotic neuron.

To better illustrate the phenomena we present in Fig. 11 the experimental setup we use to investigate how the hidden information can be recovered by a dynamical synapse. The experimental setup is composed of two electronic neurons (EN1 and EN2) that are connected through a dynamical synapse [49]. The ENs are electronic circuits that integrate in real time four ordinary differential equations of an enhanced Hindmarsh–Rose type model neuron [24, 49]. The synapse is modeled by the dynamic

clamp software developed in the INLS [29]. This software implements standard chemical synapses using a first order kinetic model. The most relevant parameter of the dynamical synapse is the integration time constant  $\tau$ . As we will show the parameter  $\tau$  is able to introduce a wide range of variation of average mutual information (AMI) between the input and the output signals. In the experiments the input signal consists of unimodal or bimodal Gaussian distributions of interspike intervals. The input induced by the spike sequence is generated using a digital to analog converter controlled by the PC. The input signal is numerically generated and consists of long sequences of spikes with a unimodal or bimodal distributions of the interspike intervals (ISIs). We used three different interspike distributions: L1 consists of a bimodal distribution composed of one “fast” Gaussian with a mean of 200 ms and a standard deviation of 50 ms along with a “slow” Gaussian centered at 500 ms and also with a standard deviation of 50 ms; L2 consists only of the “fast” Gaussian; L3, of just the “slow” Gaussian.

The input signal governed by the input distribution introduces changes in the dynamics of the neuron EN1 that can not be easily detected by using a membrane potential bursting code as shown in Fig. 12.

The bursting code detects an event every time the membrane potential drops under some given thresh-



**Fig. 12.** Examples of the time series obtained using the L2 (fast) distribution of ISI as input and different synaptic characteristic time constants at the dynamic clamp synapse: (a)  $\tau = 2$  ms, (b)  $\tau = 26$  ms, and (c)  $\tau = 48$  ms. In (b), we mark with the asterisk an event that can not be detected using a threshold level  $V_{\text{thres}}$  to encode the information in the membrane potential  $V_1$  for any value chosen for  $V_{\text{thres}}$  but that can easily detected using a threshold level  $S_{\text{thres}}$  in the synaptic activation  $S$ .

The synaptic activation variable assumes values between 0 and 1 and it is a measure of the modulation of the release of neurotransmitter in the synaptic cleft by the presynaptic cell. One can consider that the con-

old. Figure 12b shows an example (marked with the asterisk) of an event that is not possible to be detected using bursting code, in this sense the information about this event was lost. This happens because of the separation between the effect of the events and the effect of the baseline of the spikes in the membrane potential. The information is not actually lost but it found itself encoded in some other dynamical variables of the neuron not accessible to our threshold level coding. When a dynamical synapse is used we need to tune it up in order to recover the information that is not easily readable. If the synaptic characteristic time  $\tau$  is too small, as in Fig. 12a, the information can not be recovered using threshold levels because the effect of the spikes of the presynaptic cell have large amplitudes in the  $S$  trace and  $S$  decays so fast that mixes events and the baseline of the spikes at the bottom of the trace. If  $\tau$  is too large as in Fig. 12c, both spikes and events have small amplitudes, and is also more difficult to separate them. This suggests there might be an optimum value of  $\tau$  that, as shown in Fig. 12b, allows the bursting code applied to the synaptic activation variable  $S$  to be able to recover the information about the occurrence of the lost event.

The synapse activation variable  $S$  is a dynamical function of the membrane potential of the presynaptic cell,  $V_{\text{pre}}$ :

$$(1 - S_{\infty}) \tau_s \frac{dS}{dt} = S_{\infty} - S,$$

where

$$S_{\infty} = \begin{cases} \tanh\left(\frac{V_{\text{pre}} - V_{\text{thres}}}{V_{\text{slope}}}\right) & \text{if } V_{\text{pre}} > V_{\text{thres}}, \\ 0 & \text{otherwise,} \end{cases}$$

$\tau_s$  is the time constant for the synaptic decay,  $V_{\text{thres}}$  is the synaptic threshold voltage and  $V_{\text{slope}}$  is the synaptic slope voltage.

Although we did not need to make any assumptions on the properties of the postsynaptic cell we still need to rely on using a threshold level  $S_{\text{thres}}$  to codify the data using the  $S(t)$  time series. It is worthy noting that this threshold level has an important physiological meaning, closely related to the postsynaptic cell.

centration of neurotransmitter released is roughly proportional to  $S(t)$  times the synaptic strength. So, for our hypothetically designed inputs we can consider  $S_{\text{thres}}$  as proportional to a level of input neurotransmitter that separates the behavior of the postsynaptic cell in two different states of activity, and so it is directly related to the sensitivity of the postsynaptic cell to the particular neurotransmitter.

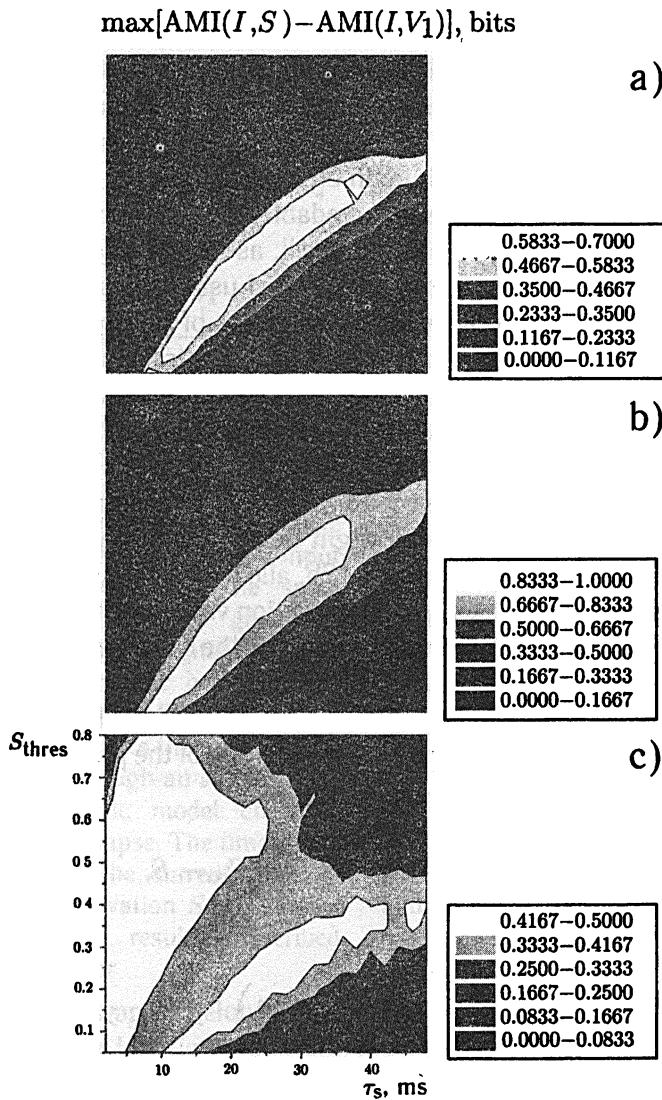


Fig. 13. Amount of information recovered by the dynamical synapse: L1 bimodal distribution of input ISI (a), L2 fast unimodal distribution of input ISI (b), and L3 unimodal slow distribution of input ISI (c)

threshold for  $S$  ( $S_{\text{thres}}$  related to the sensitivity of the postsynaptic cell), our experiments shown that a synapse can be tuned to specific values of the time constant to allow the system to transfer information in the most reliable way. If one need to optimize the dynamical synapse tuning the time constant  $\tau$  in order to transmit one specific information it is always still possible to fit the postsynaptic cell sensitivity to obtain the maximum information transmission by changing the strength of the synapse in the classical way during the learning process. These results may shed light on the origin of the heterogeneity of synaptic dynamics

Figure 13 shows bidimensional plots of the amount of information recovered as a function of  $\tau_s$  as well as of the different values  $S_{\text{thres}}$  chosen for codifying the data in the  $S$  variable and for the different distributions of ISI chosen as input.

For different distribution of input ISI and each value of  $\tau$  we acquired time series of the input, membrane potential of the cell and  $S$  variable one hour long. The analog signals were digitized using 500 Hz sample rate and the data was codified using fixed optimal threshold levels for the input and the membrane potential and variable  $S_{\text{thres}}$  as indicated in the  $y$  axis scales. We built words of 8 bits with the data and calculated the average mutual information between input and the membrane potential ( $\text{AMI}(I, V_1)$ ) and the average mutual information between input and  $S$  ( $\text{AMI}(I, S)$ ). We defined the recovered information as a positive difference  $\text{AMI}(I, S) - \text{AMI}(I, V_1)$ . The maximum values of  $\text{AMI}(I, S)$  are in all cases around 2.4 bits and roughly coincides in  $(\tau, S_{\text{thres}})$  parameter space with the maxima of the recovered information, which means that approximately 30 % of the total information transmitted from input to  $S$  came from the recovery phenomenon.

As it can be seen in Fig. 13, the dynamical synapse is able to recover input information that was hidden in other degrees of freedom in the membrane potential of the cell for a wide range of time constants and the gain of information is also dependent on the type of input (L1, L2, or L3) that is introduced in this neural information transport channel. Moreover, the dynamical synapse can be adjusted to maximize the gain of information. For a given codifying

in the brain [50, 51].

It may also be reasonable to hypothesize that during learning, alterations in synaptic dynamics leads to changes in the efficiency of information transmission in different ways depending on the specific features of the incoming temporal patterns. There is a need to find a reader or decoder that is able to extract the information from the channel, and this is provided by the dynamical synapse. The synaptic properties we have discussed will separate those states that seem to be mixed in the dynamics of the neuron so that they can be read at the next higher level. We hypothesize that this tuning of dynamical synaptic parameters may significantly contribute to the learning process for neural communication.

## 5. DISCUSSION

In fact, the subject of the information neurodynamics is to describe how to process the incoming information and generate new one by coupled neural networks with complex dynamics. In order to understand the information transformation and interaction in any dynamical system we need to know how the information about stimuli is to be fed into it and how the result of the dynamically activity is to be read out. These questions are nontrivial even for artificial systems and became much more complex for biological neural ensembles. In biological systems sometimes we even don't know who are the "readers." There are many publications about the right coding and decoding spaces, the role of time and so on but the final criteria of the validity of these ideas is the behavior only. From this point of view, all the results that we discussed in this paper do not give us a clear answer to the question: how the intrinsic dynamics of the nervous system is related to information processing? However, the discussed results build a good basis for future investigations. We think that in the light of it the experiments with simple animals (like mollusks) are very important because in such experiments we can find a direct connections between behavior and changing environment. In fact we are trying to solve the inverse problem: to reconstruct the dynamics of the nervous system using the knowledge about the incoming signal and behavior. The solution of this problem is not hopeless because the nervous system is not a "black box." We know a lot about the topology of the network and the dynamics of the individual neurons from the anatomical and physiological experiments. To show how the changes of the sensory environment mark events to which an organism is react let us describe the hunting behavior of the mollusk *Clione*. This behavior is really amazing. In the presence of the smell of the prey in the water the *Clione* behavior becomes a chaotic scanner of the surrounding space in search of the target. The way this scanning is performed is by loops in the water caused by a strong flexion of the tail and as a result the planes where this loops are carried out are changed chaotically. The dynamics can be represented by symbolic dynamics of the random sequence of the four symbols : L, R, A, B (Left, Right, Ahead, Back)[27]. To understand the origin of such chaotic activity of the mollusk we have to make a choice between two conceptions: (i) the *Clione* nervous system itself generates chaotic spatio-temporal patterns and the mechanical part of the system just follows the "commands from above," and (ii) random hunting is the result of the dynamically interaction of the position sensory system and "brain." Given the importance of robustness and stability in animal behavior we find the second concept as a more suitable mechanism. The anatomical and physiological data support our hypothesis [52, 53]. As our preliminary results showed, the dynamical model describing the interaction of the body motion with the nervous system activity in the presence of the prey smells has a strange attractor. This strange attractor is the mathematical image of the chaotic *Clione* hunting. From this point of view we can conclude that a simple mollusk brain generates new information in the form of a chaotic attractor in order to achieve survival.

This paper is the result of numerous discussions with our colleagues at INLS (Henry Abarbanel, Yury Arshavsky, Al Selverston, Rob Elson, Attila Szucs, Pablo Varona, Nikolai Rulkov, Alexander Volkovskii, and Lev Tsimring), at Caltech (Gilles Laurent and A. Bracker), and at Salk Institute (Terry Sejnowski

and Maxim Bazhenov). We acknowledge support from U. S. Department of Energy (grant DE-FG03-96ER14592), the Program “Leading Scientific Schools of the Russian Federation” (project No. 00-15-96582), and the Russian Foundation for Basic Research (project No. 01-02-17638). R. Huerta also thanks Ministerio de Ciencia y Tecnología for grant BFI2000-0157. R. D. Pinto was supported by the Brazilian agency Fundação de Amparo à Pesquisa do Estado de São Paulo — FAPESP under proc. 98/15124-5.

#### REFERENCES

1. Wilson H. R. Spikes, Decision and Actions: Dynamical foundations of neuroscience. Oxford, 1999.
2. Deco G., Schurmann B. // Phys. Rev. Lett. 1997. V. 79. P. 4697.
3. Strong S. P., Koberle R., De Ruyter Van Steveninck R. R., Bialek W. // Phys. Rev. Lett. 1998. V. 80. P. 197.
4. Stemmler M., Koch C. // Nat. Neurosci. 1999. V. 2. P. 521.
5. Destexhe A., Mainen Z. F., Sejnowski T. J. // Neural Computation. 1994. V. 6. P. 14.
6. Abeles M., Bergman H., Gat I., Seidelman E., Tishby N., Vaadia E. // Proc. National Acad. Sci. USA. 1995. V. 92. P. 8616.
7. Villa A. E., Abeles M. // Brain Research. 1990. V. 509. P. 325.
8. Abarbanel H. D. I., Huerta R., Rabinovich M. I., Rulkov N. F., Rowat P. F., Selverston A. I. // Neural Computation. 1996. V. 8. P. 1567.
9. Elson R., Selverston A. I., Huerta R., Rabinovich M. I., Abarbanel H. D. I. // Phys. Rev. Lett. 1998. V. 81. P. 5692.
10. Elson R., Huerta R., Abarbanel H. D. I., Rabinovich M. I., Selverston A. I. // J. Neurophysiol. 1999. V. 82. P. 115.
11. Laurent G., Stopfer M., Freidrich R. W., Rabinovich M., Volkovskii A., Abarbanel H. D. I. // Ann. Rev. Neurosci. 2001. V. 24. P. 263.
12. Andronov A. A., Leontovitch E. A., Gordon I. I., Maier A. G. Theory of Bifurcations of Dynamical Systems on a Plane. New York: Wiley, 1973.
13. Andronov A. A., Leontovitch E. A., Gordon I. I., Maier A. G. Qualitative Theory of Dynamical Systems of Second Order. New York: Wiley, 1973.
14. Shannon C. E. Weaver W. The Mathematical Theory of Communication. Urbana: Univ. Illinois Press, 1949.
15. Dayan P., Abbott L. F. Theoretical Neuroscience. <http://play.ccs.brandeis/abbott/book/TOC.html>
16. Rieke F., Warland D., de Ruyter van Steveninck R., Bialek W. Spikes. MIT Press, 1997.
17. Afraimovich V. S., Verichev N. N., Rabinovich M. I. // Izv. Vyssh. Uchebn. Zaved., Radiofiz. 1986. V. 29. P. 1050.
18. Heagy J. F., Pecora L. M., Carrol T. L. // Phys. Rev. E. 1994. V. 50. P. 1874.
19. Rulkov N. F., Volkovskii A. R. // Phys. Lett. A. 1993. V. 179. P. 332.
20. Selverston A. // Progr. Brain Res. 1999. V. 123. P. 247.
21. Huerta R., Varona P., Rabinovich M. I., Abarbanel H. D. I. // Biol. Cybern. 2001. V. 84. P. L1.
22. Huerta R., Pinto R. D., Varona P., Stiesberg G. R., Rabinovich M. I., Abarbanel H. D. I., Selverston A. // Neural Networks. (to be submitted).
23. Harris-Warrick R. M., Johnson B. R., Peck J. H., Kloppenburg P., Ayali A., and Skarbinski J. // Ann. N.Y. Acad. Sci. 1998. V. 860. P. 155.
24. Hindmarsh J. L., Rose R. M. // Proc. R. Soc. London. 1984. V. B221. P. 87.
25. Pinto R. D., Varona P., Volkovskii A. R., Szücs A., Abarbanel H. D. I., Rabinovich M. I. // Phys. Rev. E. 2000. V. 62. P. 2644.
26. Selverston A. I., Moulins M. The Crustacean Stomatogastric System. Berlin: Springer, 1987.

27. Arshavsky Y. I., Beloozerova I. N., Orlovsky G. N., Panchin Y. V., Pavlova G. A. // *Exp. Brain Res.* 1985. V. 58. P. 255.
28. Getting P. A. // *Ann. Rev. Neurosci.* 1989. V. 12. P. 185.
29. Pinto R. D., Elson R. C., Szűcs A., Rabinovich M. I., Abarbanel H. D. I., Selverston A. I. // *J. Neurosci. Methods.* 2001. (to be submitted). Versions of the program including source code as well as more information about system requirements and schematics can be download from <http://inls.ucsd.edu/~rproto>.
30. Klevecz R. R., Ruddle F. H. // *Science.* 1968. V. 159. P. 634.
31. Novak B., Mitchison J. M. // *J. Cell Sci.* 1986. V. 86. P. 191.
32. Huerta R., Bazhenov M., Rabinovich M. I. // *Europhys. Lett.* 1998. V. 43. P. 719.
33. Abarbanel H. D. I., Brown R., Sidorowich J. J., Tsimring L. S. // *Rev. Mod. Phys.* 1993. V. 64. P. 1331.
34. Pecora L. M., Carroll T. L. // *Phys. Rev. Lett.* 1990. V. 64. P. 821.
35. Rabinovich M. I., Torres J. J., Varona P., Huerta R., Weidman P. // *Phys. Rev. E.* 1999. V. 60. P. R1130.
36. Gaponov-Grekhov A. V., Rabinovich M. I. // *Chaos.* 1996. V. 6. P. 259.
37. Hudspeth A. J., Legothetis N. K. // *Curr. Opin. Neurobiol.* 2000. V. 10. P. 631.
38. Buzsaki G., Chrobak J. J. // *Curr. Opin. Neurobiol.* 1995. V. 5. P. 504.
39. Singer W., Gray C. M. // *Ann. Rev. Neurosci.* 1995. V. 18. P. 555.
40. May R. M., Leonard W. I. // *SIAM J. Applied Math.* 1975. V. 29. P. 243.
41. Grossberg S. // *J. Theor. Biol.* 1978. V. 73. P. 101.
42. Desmaisons D., Vincent J.-D., Lledo J.-M. // *J. Neurosc.* 1999. V. 19. P. 10727.
43. Afraimovich A., Rabinovich M. I. (in preparation).
44. Stopfer M., Laurent G. // *Nature.* 1999. V. 402. P. 664.
45. Vaadia E., Haalman I., Abeles M., Bergman H., Prut Y., Slovin H., Aertsen A., // *Nature.* 1995. V. 373. P. 515.
46. Vastano J. A., Swinney H. L. // *Phys. Rev. Lett.* 1998. V. 60. P. 1773.
47. Schreiber T. // *Phys. Rev. Lett.* 2000. V. 85. P. 461.
48. Eguia M. C., Rabinovich M. I., Abarbanel H. D. I. // *Phys. Rev. E.* 2000. V. 62. P. 7111.
49. Pinto R. D., Varona P., Volkovskii A. R., Szucs A., Abarbanel H. D. I., Rabinovich M. I. // *Phys. Rev. E.* 2000. V. 62. P. 2644.
50. Gupta A., Wang Y., Markram H. // *Science.* 2000. V. 287. P. 273.
51. Natschlaeger T., Wolfgang M. Neural Information Processing Systems 2000 (NIPS '2000), V. 13. Cambridge: MIT Press, 2001.
52. Deliagina T. G., Arshavsky Y. I., Orlovsky G. N. // *Nature.* 1998. V. 393. P. 172.
53. Orlovsky G. N., Deliagina T. G., Grillner S. *Neuronal Control of Locomotion: From Mollusc to Man.* Oxford, 1999.

<sup>1</sup> Institute for Nonlinear Science,  
University of California, San Diego,  
La Jolla, California, USA;

<sup>2</sup> Институт прикладной физики РАН,

Нижний Новгород, Россия;

<sup>3</sup> E.T.S.Informática,  
Universidad Autónoma de Madrid,  
Madrid, Spain

Поступила в редакцию  
26 февраля 2001 г.

УДК 517.9

## ДИССИПАТИВНЫЕ СОЛИТОНЫ В КОМПЛЕКСНОМ УРАВНЕНИИ ГИНЗБУРГА—ЛАНДАУ ДЛЯ ФЕМТОСЕКУНДНЫХ ГЕНЕРАТОРОВ СВЕТА

*А. И. Корытин, А. Ю. Крячко, А. М. Сергеев*

Рассмотрены модели генерации фемтосекундного излучения в твердотельных лазерах с пассивной синхронизацией мод. Развит под подход, в котором ультракороткие лазерные импульсы ассоциируются с диссипативными солитонами обобщённого комплексного уравнения Гинзбурга—Ландау. В ряде предельных случаев аналитически определены структура солитонов, области параметров их существования, сформулированы критерии устойчивости солитонов. Исследованы сценарии циклической спектральной и временной динамики диссипативных солитонов, связанной с неустойчивостью фонового поля.

### ВВЕДЕНИЕ

В последние годы одной из наиболее интересных проблем нелинейной оптики является теоретическое и экспериментальное исследование процессов генерации ультракоротких импульсов (УКИ) фемтосекундной длительности в твердотельных широкополосных лазерах. Область научно-технических применений таких генераторов света достаточно широка и включает в себя, например, диагностику быстро протекающих процессов, микромодификацию вещества, разработку телекоммуникационных технологий, создание источников сверхсильных лазерных полей и т. д.

Для описания физических процессов лазерной генерации коротких импульсов принято использовать два различных подхода: модовый и пространственно-временной. В модовом подходе, возникшем на основе идей параметрической теории нелинейных колебаний, в рассмотрение включается ограниченное (как правило, малое) число мод колебаний резонатора и исследуется условие их фазовой синхронизации (см., например, [1]). Это позволяет в явном виде сформулировать условие неустойчивости непрерывной генерации и детально исследовать энергообмен в небольшом семействе соседних мод резонатора. Пространственно-временной подход постепенно формировался по мере создания всё более короткоимпульсных квантовых генераторов [2–6]. В случае, если в процесс генерации вовлечено очень большое число продольных мод (в современных фемтосекундных источниках до  $10^6$ ), то для математического описания и физического понимания процессов формирования лазерных импульсов естественно наблюдать за эволюцией начального пробного импульса в системе. При этом установление лазерной генерации означает асимптотическую эволюцию пробного волнового пакета к универсальному локализованному временному распределению — автосолитону, или диссипативному солитону независимо от исходной формы. Математически это означает стремление траекторий к устойчивому состоянию равновесия в фазовом пространстве параметров импульса. С точки зрения традиционной нелинейной динамики процесс установления фемтосекундной лазерной генерации является частным случаем реализации автоколебательного режима [7] в лазерном резонаторе.

Целью данного обзора является изложение достаточно широкого круга вопросов, посвящённых поиску и исследованию устойчивости автосолитонных решений (соответствующих УКИ) в задачах генерации современных твердотельных лазеров. В работе рассмотрены следующие основные вопросы. В разделе 1 приведён вывод системы уравнений, описывающих процесс генерации УКИ в общем случае. В разделах 2 и 3 рассмотрены два важных случая режима лазерной генерации: при быстром и медленном насыщающемся поглотителе. При этом обсуждаются общие свойства решения, а также рассматриваются вопросы устойчивости лазерной генерации.

## 1. ОСНОВНЫЕ УРАВНЕНИЯ ДЛЯ ОПИСАНИЯ ПРОЦЕССОВ ГЕНЕРАЦИИ УКИ В СОВРЕМЕННЫХ ТВЕРДОТЕЛЬНЫХ ЛАЗЕРАХ

Наиболее просто модель лазерной генерации УКИ строится при следующих предположениях: 1) суммарное поле синхронизированных продольных мод лазера представляет собой компактный бегущий волновой пакет, временной масштаб которого  $\tau$  много меньше времени обхода резонатора  $T_R$ ; 2) изменения, происходящие с полем при прохождении отдельных элементов лазера, незначительны, а существенные изменения волнового пакета возможны лишь за время, много большее  $T_R$ ; 3) поле импульса остаётся в основной поперечной пространственной mode резонатора (приближение одномерных волновых пакетов). Последнее предположение сделано на основании того, что в рассматриваемом классе лазеров пиковая мощность импульса не превышает критическую мощность самофокусировки, и излучение лазера в режиме синхронизации мод имеет одномодовую поперечную структуру, что наблюдается в большинстве экспериментов. Тогда каждому обходу резонатора, состоящего из дискретных элементов, ставится в соответствие шаг распространения импульса в эквивалентной распределённой нелинейной среде, приводящий к аналогичному малому изменению параметров импульса. Ниже приведён вывод уравнения, описывающего изменение комплексной огибающей амплитуды импульса в процессе установления лазерной генерации.

В общем случае в лазерном резонаторе можно выделить следующие основные элементы (рис. 1) и соответствующие им механизмы трансформации импульса:

1) Нелинейный лазерный кристалл, обеспечивающий усиление проходящего импульса за счёт инверсии населённостей, фазовую самомодуляцию за счёт нелинейности показателя преломления вещества кристалла, расплывание импульса за счёт материальной дисперсии (дисперсии групповых скоростей) и фильтрацию боковых спектральных компонент импульса за счёт наличия конечной полосы усиления.

2) Компенсатор материальной дисперсии кристалла — система, обладающая отрицательной дисперсией групповых скоростей. Наличие данного элемента необходимо для компенсации дисперсионного расплывания импульса после прохождения через кристалл, что, в свою очередь, необходимо для достижения как можно меньшей длительности импульса.

3) Насыщающийся поглотитель, необходимый для эффективной синхронизации мод в резонаторе для достижения минимальной длительности импульса. Как правило, данный элемент представляет собой двухуровневую среду, насыщающуюся при прохождении импульса и, следовательно, обеспечивающую специфические нелинейные потери импульса.

4) Полупрозрачное зеркало, используемое для вывода излучения из резонатора. Данный элемент отвечает за линейные потери импульса.

Пусть после  $N$  обходов импульс находится в точке 1 резонатора ( $A_{(N)} \equiv A_1$ ). Рассмотрим эволюцию импульса за один обход по резонатору.

Наличие нелинейного лазерного кристалла обеспечивает сразу несколько механизмов трансформации импульса. Но в силу предположения о малости изменений импульса влияние данных факторов, во-первых, аддитивно:

$$A_2(t) = \exp \left[ g(t) - i\delta |A_{(N)}|^2 + \hat{D}(t) + \hat{G}(t) \right] A_{(N)}(t), \quad (1)$$

а во-вторых, экспонента в (1) может быть разложена в ряд с пренебрежением членами второго порядка и выше:

$$A_2(t) = A_1(t) + [g(t) - i\delta |A_1|^2 + \hat{D}(t) + \hat{G}(t)] A_1(t). \quad (2)$$

Здесь  $A_j(t)$  — комплексная огибающая импульса в  $j$ -й точке резонатора,  $t$  — время в локальной, т. е. движущейся с групповой скоростью  $V_{\text{gr}} = \partial\omega/\partial k|_{\omega=\omega_0}$  системе координат,  $\omega_0$  — центральная частота генерации лазера,  $g(t)$  — коэффициент усиления (гейн) активной среды,  $\delta$  — коэффициент фазовой самомодуляции, определяющий нелинейный набег фазы поля за один обход по резонатору (предполагается кубичная нелинейность показателя преломления),  $\hat{G}(t)$  и  $\hat{D}(t)$  — операторы, описывающие соответственно конечную полосу усиления активной среды и эффект дисперсии групповых скоростей. Наиболее просто данные операторы выглядят в предположении квадратичной дисперсии групповых скоростей и усиления:  $\hat{D}(t) = iD_{\text{mat}} \partial^2/\partial t^2$ ,  $\hat{G}(t) = D_g \partial^2/\partial t^2$ . Здесь  $D_{\text{mat}} = \partial^2 k/\partial\omega^2|_{\omega=\omega_0} L/2 > 0$ ,  $D_g = g(\omega_0)L/\Omega^2$ , где  $g(\omega_0)$  — коэффициент усиления на центральной частоте,  $\Omega$  — ширина полосы усиления,  $L$  — длина кристалла.

Влияние компенсатора дисперсии аналогично влиянию эффекта материальной дисперсии кристалла:

$$A_3(t) = \exp[\hat{D}_{\text{neg}}(t)] A_2(t) \approx A_2(t) + \hat{D}_{\text{neg}}(t) A_2(t), \quad \hat{D}_{\text{neg}}(t) = iD_{\text{neg}} \frac{\partial^2}{\partial t^2}, \quad (3)$$

но коэффициент дисперсии  $D_{\text{neg}}$  в компенсаторе отрицателен. Далее мы будем описывать влияние дисперсии групповых скоростей единым коэффициентом дисперсии  $D$ , характеризующим влияние дисперсии за один обход по резонатору:  $D = D_{\text{mat}} + D_{\text{neg}}$ . При этом  $D$  может иметь любой знак.

Трансформация импульса при прохождении насыщающегося поглотителя может быть описана соотношением

$$A_4(t) = \exp[-q(t)] A_3(t) \approx A_3(t) - q(t) A_3(t), \quad (4)$$

где  $q(t)$  — нормированная разность населённостей в поглотителе.

И, наконец, влияние линейных потерь может быть записано в виде

$$A_5(t) \equiv A_{(N+1)}(t) = \exp(-l_0) A_4(t) \approx A_4(t) - l_0 A_4(t), \quad (5)$$

где  $l_0$  — коэффициент линейных потерь за один обход по резонатору.

В силу предположения о малости изменения импульса за один обход резонатора разность амплитуд на  $(N+1)$ -м и  $N$ -м обходах можно заменить на производную по некоторой эффективной координате  $z = T/T_R$ , где  $T$  — время в лабораторной системе отсчёта, или, в терминах эффективной распределённой среды, по координате вдоль данной среды:

$$A_{(N+1)}(t) - A_{(N)}(t) \approx \partial A(t, z)/\partial z. \quad (6)$$

Из соотношений (2)–(6) вытекает следующее уравнение для комплексной огибающей поля:

$$\frac{\partial A}{\partial z} = [g(t, z) - l_0] A - i\delta |A|^2 + (D_g + iD) \frac{\partial^2 A}{\partial t^2} - q(t) A. \quad (7)$$

Данное уравнение носит название обобщённого комплексного уравнения Гинзбурга—Ландау (УГЛ) и широко используется для теоретического исследования процессов генерации и устойчивости УКИ в современных твердотельных лазерах.

Выход уравнения Гинзбурга—Ландау (точнее, различных его частных случаев) и поиск его решения в аналитическом виде проводились в большом количестве работ, начиная с 60-х годов. В связи с этим

интересно проследить хронологию основных этапов в исследовании данного уравнения. УГЛ в рамках исследования волновых процессов в плазме было получено в [8]. Уравнение, описывающее лазерную генерацию (имеющее вид уравнения переноса) было получено в [3, 4] с учётом наличия усиления с бесконечно широкой полосой, линейных и нелинейных потерь. В частности, в данных работах была высказана идея о выводе данного уравнения путём рассмотрения малых изменений импульса после прохождения очередного элемента резонатора. В [5] уравнение лазерной генерации было исследовано при дополнительном учёте конечной полосы усиления. Наконец, в [9] было проведено наиболее полное исследование уравнения лазерной генерации в виде УГЛ: учтены все факторы, присутствующие в (7), найдено в аналитическом виде и исследовано решение УГЛ.

Уравнение (7) должно быть дополнено уравнениями, описывающими динамику коэффициента усиления  $g(t, z)$  и насыщающегося поглотителя  $q(t)$ . Насыщение поглотителя при некогерентном взаимодействии поля с двухуровневой средой может быть описано уравнением

$$\frac{dq}{dt} = -\frac{q - q_0}{\tau_A} - \frac{|A|^2 q}{E_A}. \quad (8)$$

Здесь  $q_0$  — разность населённостей при слабом сигнале,  $\tau_A$  и  $E_A$  — время релаксации и энергия насыщения поглотителя. Поскольку время релаксации поглотителя в твердотельных лазерах, как правило, много меньше времени обхода резонатора, разность населённостей зависит лишь от локального (в системе отсчёта, связанной с импульсом) времени  $t$ .

Коэффициент усиления  $g$  испытывает двухмасштабную временную эволюцию — быстрое уменьшение на небольшую величину за счёт насыщения усиления при прохождении импульса и медленное восстановление своих свойств под действием постоянной накачки  $P$ . Такое разделение характерных времён изменения гейна возможно, т. к. современные твердотельные лазеры характеризуются большой энергией насыщения активной среды:  $W \ll W_s$ . Здесь  $W = \int |A|^2 d\tau$  — энергия импульса (на единицу площади поперечного сечения),  $W_s$  — энергия насыщения коэффициента усиления. Вследствие этого гейн можно представить в виде произведения:

$$g = g_m(z) \left[ 1 - \int_{-\infty}^t \frac{|A|^2 d\tau}{W_s} \right], \quad (9)$$

а для медленной огибающей использовать уравнение баланса:

$$V_{gr} \frac{dg_m}{dz} = -\frac{g_m - g_0}{T_1} - \frac{g_m W(z)}{T_R W_s}. \quad (10)$$

Здесь  $T_1$  — время релаксации разности населённостей активной среды,  $g_0 = PT_1$  — ненасыщенный гейн.

Система уравнений (7)–(10) описывает процесс генерации УКИ в общем случае. В силу своей сложности исследование данной системы возможно лишь в рамках численного моделирования, однако существует ряд частных случаев, в которых данные уравнения существенно упрощаются и в то же время позволяют получить ряд важных результатов в аналитическом виде.

Для дальнейшего изложения удобно проводить рассмотрение для двух противоположных предельных случаев: быстрого насыщающегося поглотителя ( $\tau_A \ll \tau$ ) [4–6, 9–14] и медленного насыщающегося поглотителя ( $\tau_A \gg \tau$ ) [9, 15–21].

## 2. СЛУЧАЙ БЫСТРОГО НАСЫЩАЮЩЕГОСЯ ПОГЛОТИТЕЛЯ

В данном случае разность населённостей  $q$  является мгновенной функцией интенсивности поля. Следовательно, можно пренебречь производной в левой части (8) по сравнению с  $(q - q_0)/\tau_A$ . При этом выражение для  $q(t)$  будет иметь вид

$$q(t) = \frac{q_0}{1 + |A|^2 \tau_A / E_A}. \quad (11)$$

Предположим далее, что энергия насыщения поглотителя велика, так что  $|A|^2 \tau_A / E_A \ll 1$ . Тогда (11) принимает вид

$$q(t) = q_0 - \frac{q_0 \tau_A |A|^2}{E_A}. \quad (12)$$

Рассмотрим также наиболее простое предположение относительно динамики гейна — случай бесконечно большой энергии насыщения активной среды:  $W_S \rightarrow \infty$ . Тогда соотношение для  $g(t)$  принимает вид

$$g(t) = g_0. \quad (13)$$

С учётом (12) и (13) УГЛ принимает вид

$$\frac{\partial A}{\partial z} = (g_0 - l) A + D_g (1 + id) \frac{\partial^2 A}{\partial t^2} + (\gamma - i\delta) A |A|^2, \quad (14)$$

где  $l = l_0 + q_0$ ,  $\gamma = q_0 \tau_A / E_A$ ,  $d = D/D_g$  — безразмерный коэффициент дисперсии.

### 2.1. Общие свойства решения

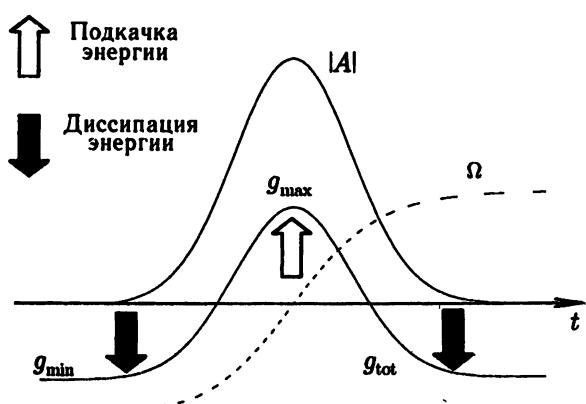


Рис. 2. Распределение амплитуды поля  $|A|$ , отстройки частоты  $\Omega$  в импульсе и полного гейна  $g_{tot} = g_0 - l + \gamma |A|^2$ , соответствующих (14), (15). Здесь  $g_{min} = g_0 - l$ ,  $g_{max} = g_0 - l + \gamma |A|_0^2$ . Для (15)  $\beta < 0$ , т. е. частота нарастает от переднего фронта импульса к заднему

ем энергии, амплитуды и обратной длительности импульса в бесконечность на конечном расстоянии распространения вдоль  $z$  по закону [22, 23]

$$W, A_0 \propto (z_0 - z)^{-1/2}, \quad \tau \propto (z_0 - z)^{1/2}.$$

Уравнение Гинзбурга—Ландау в форме (14) имеет аналитическое решение в виде локализованного импульса [9, 14]:

$$A = \frac{A_0 e^{ihz}}{\text{ch}^{1+i\beta}(t/\tau)}, \quad (15)$$

характеризующееся четырьмя параметрами: амплитудой  $A_0$ , длительностью  $\tau$ , модуляцией ( chirpом) частоты  $\beta$  и постоянной распространения  $h$ . Эти структуры интегрально балансируют накачку энергии и потери, так что в центральной части импульса  $g_0 + \gamma |A|^2 > l$ , а на крыльях солитона  $g_0 < l$ , чем достигается устойчивость среды относительно нарастания шумов вне импульса (рис. 2). Формально такие решения существуют в полосе параметров  $\delta < \delta_* = \gamma [3(1+d^2)^{1/2} - d]$ , а в области  $\delta > \delta_*$  система характеризуется сложным динамическим поведением. При этом эволюция импульса в резонаторе происходит в режиме коллапса с обращением

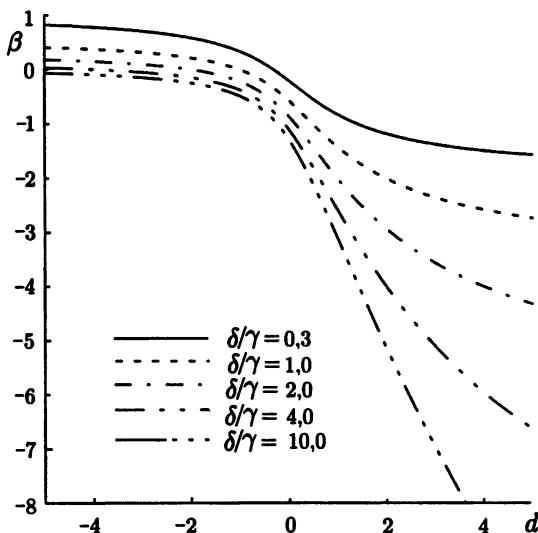


Рис. 3. Зависимость модуляции  $\beta$  от коэффициента дисперсии  $d$  для разных значений параметра  $\delta/\gamma$

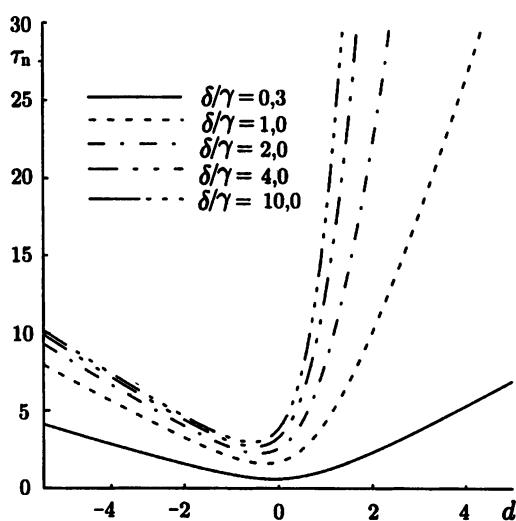


Рис. 4. Зависимость длительности импульса  $\tau_n$  от коэффициента дисперсии  $d$  для разных значений параметра  $\delta/\gamma$

Подставляя решение в виде (15) в уравнение (14), можно получить выражения для параметров импульса:

$$\beta = \frac{-3[(\delta/\gamma)d - 1] - \sqrt{9[(\delta/\gamma)d - 1]^2 + 8(\delta/\gamma + d)^2}}{2(\delta/\gamma + d)},$$

$$\tau_n = W\tau\delta/D_g = 2(\delta/\gamma)(2 - \beta^2 - 3\beta d), \quad h = (D_g/\tau^2)[d(1 - \beta^2) + 2\beta], \quad A_0^2 = W/(2\tau). \quad (16)$$

Здесь введена безразмерная длительность импульса  $\tau_n$ .

На рис. 3 и 4 изображена зависимость  $\beta$  и  $\tau_n$  от коэффициента дисперсии  $d$ . В зависимости от знака  $d$  существуют два существенно различных режима генерации [6, 10]. В случае положительных  $d$  абсолютное значение  $\beta$  достаточно велико, и длительность импульса определяется действием двух факторов: уменьшением диапазона изменения частотной отстройки (а следовательно, и укорочением импульса) за счёт фильтрации крайних спектральных компонент полосой усиления и дисперсионным расплыванием импульса в следствие положительной дисперсии групповых скоростей. Для отрицательных  $d$  модуляция частоты импульса близка к нулю, а длительность импульса существенно меньше, чем в случае положительных  $d$ . В данном режиме основное влияние на формирование импульса оказывают дисперсия групповых скоростей и фазовая самодуляция, обеспечивая существование локализованных решений, которые описываются уравнением типа нелинейного уравнения Шредингера (так называемые шредингеровские солитоны) [6].

## 2.2. Исследование устойчивости солитонного решения

В фемтосекундной оптике широко распространено мнение, что именно солитоны вида (15) описывают излучение генераторов сверхкоротких импульсов света. Однако любые стационарные локализованные решения (15) УГЛ глобально неустойчивы относительно взрывного нарастания амплитуды в центральной части солитона. Стабилизация этой неустойчивости возможна при учёте насыщения гейна генерируемыми импульсами на медленном временном масштабе его изменения (т. е. в среднем за обход) в уравнении (10) [11]. Это приводит к модификации (14) путём замены

$$g_0 \rightarrow g_m = g_0/(1 + W/W_0), \quad (17)$$

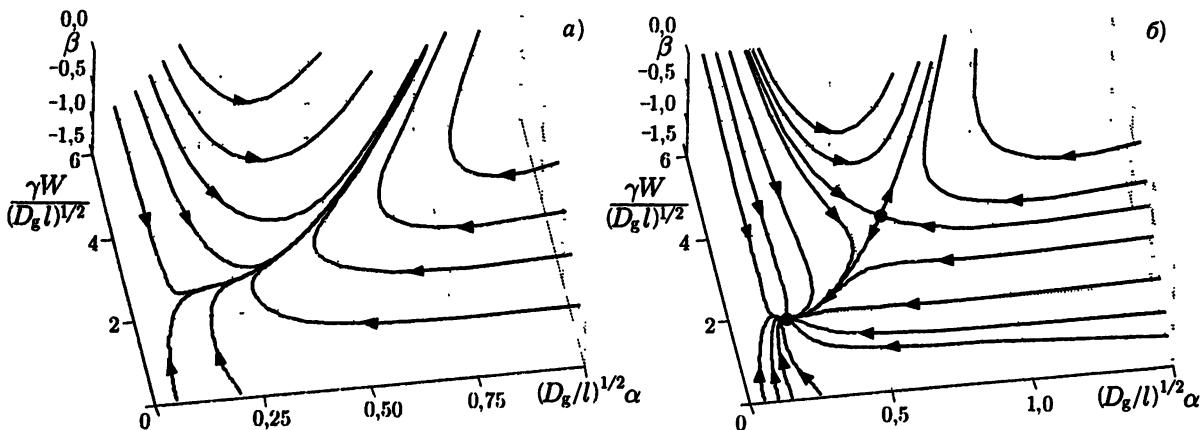


Рис. 5. Фазовое пространство  $(W, \alpha, \beta)$  в случае слабого насыщения поглотителя и  $W_0 > W_{0*}$ . (а),  $W_0 < W_{0*}$ . (б). Вычисления проводились для  $g_0/l = 1,2$ ;  $d = 0,5$ ;  $\delta/\gamma = 0,5$  и  $\gamma W_0 / (D_g l)^{1/2} = 5$  (а),  $\gamma W_0 / (D_g l)^{1/2} = 3$  (б)

где  $W_0 = W_S T_R / T_1 \ll W_S$ . При этом (14) принимает вид

$$\frac{\partial A}{\partial z} = [g_0/(1 + W/W_0) - l] A + D_g (1 + id) \frac{\partial^2 A}{\partial t^2} + (\gamma - i\delta) A |A|^2. \quad (18)$$

Данную задачу удобно исследовать на фазовой плоскости параметров импульса  $(W, \alpha, \beta)$ , где  $\alpha = 1/\tau$  — обратная длительность импульса. Как показано в [12, 13], в новом модельном уравнении существует бифуркационное значение параметра  $W_0$ :

$$W_{0*} = \frac{2(1 - \xi_*) \sqrt{\mu \xi_*}}{\sqrt{\gamma(g_0/l - 1 + \xi_*)[1 - 2\beta_*(\beta_* + d)/(\mu\gamma)]^{1/2}}}, \quad (19)$$

где  $\xi_* = 1 - 3g_0[1 - \sqrt{1 - 8l/(9g_0)}]/(2l)$ ,  $\mu = 2(2 - \beta_*^2 - 3d\beta_*)/\gamma$ ,  $\beta_* = \lambda \pm \sqrt{\lambda^2 + 2}$ ,  $\lambda = 3(\gamma - \delta d)/[2(d\gamma + \delta)]$  (значение  $\beta_*$  выбирается таким, чтобы  $\mu > 0$ ). При уменьшении  $W_0$  и прохождении через это значение в фазовом пространстве  $(W, \alpha, \beta)$  системы, описывающем динамику импульсов солитонной формы (15), появляется пара состояний равновесия (рис. 5). Одно из них представляет собой седло и соответствует импульсу с большей энергией, а другое — устойчивый узел с меньшей энергией. При этом узел соответствует режиму устойчивой импульсной генерации с мягким возникновением из непрерывной. Бифуркационному значению  $W_0$  отвечает максимальная энергия солитона

$$W_{sol*}^{\max} = \frac{2\sqrt{\xi_*\mu}}{\sqrt{\gamma[1 - 2\beta_*(\beta_* + d)/(\mu\gamma)]^{1/2}}}, \quad (20)$$

при этом энергия импульса для случая  $g_0 - l \ll g_0$  ровно вдвое превышает энергию  $W_{cw} = W_0(g_0/l - 1)$ , соответствующую запасённой в резонаторе при непрерывной генерации.

Несмотря на существование режима устойчивой импульсной генерации в (18), фазовое пространство  $(W, \alpha, \beta)$  системы при достаточно большой энергии волновых пакетов по-прежнему, как и в случае  $W_S \rightarrow \infty$ , заполнено неустойчивыми траекториями, уходящими на бесконечность, а эволюция волнового пакета вдоль них происходит в режиме коллапса. Это означает, что учёт насыщения усреднённого коэффициента усиления стабилизирует динамику не любого начального распределения. Более того, практически самая интересная область параметров с большими энергиями импульса и малыми длительностями генерации описывается в (18) некорректно.

Решить эту проблему можно при учёте насыщения нелинейного поглотителя, рассмотрев, например, следующий член в разложении (11) по степеням  $|A|^2$  [24, 25]. Однако при достаточно больших

интенсивностях импульса, когда все члены в разложении (11) оказываются одного порядка, подход, описанный в [24, 25], неприменим. В данном случае можно учесть насыщение нелинейного поглотителя, взяв в исходном уравнении (7)  $q = q_0/(1 + |A|^2 \tau_A/E_A)$  и сохраняя для  $g$ ,  $\hat{D}$  и  $\hat{G}$  аналогичные (18) выражения. Тогда даже при  $W_0 \gg W_{0*}$  в фазовом пространстве появляется новый атTRACTор в виде устойчивого узла при больших значениях энергии  $W_{\text{sat}}$  и меньших длительностях импульсов  $\tau_{\text{sat}}$  (рис. 6). Так, в случае  $\delta = 0$ ,  $d = 0$  имеем  $W_{\text{sat}} \approx \gamma W_0(g_0/l_0 - 1)$ ,  $\tau_{\text{sat}} \approx [D_g W_0(g_0/l_0 - 1)]^{1/2}/E_A$ . Если  $\Delta g = g_0 - (l_0 + q_0) \ll g_0$  и  $l_0 \approx q_0$ , генерация УКИ может иметь в  $g_0/\Delta g$  раз большую мощность, чем непрерывная.

При изменении  $W_0$  и  $E_A$  фазовое пространство системы  $(W, \alpha, \beta)$  имеет различную топологическую структуру. Для достаточно больших полей насыщения  $E_A$  и  $\Delta g \ll g_0$  имеются два устойчивых состояния равновесия, соответствующие слабому (аналогично (18)) и сильному насыщению поглотителя, при этом относительно длинноимпульсная генерация возможна при мягком режиме возбуждения, а короткоимпульсная — только при жёстком. При уменьшении  $W_0$  седловая точка и короткоимпульсный атTRACTор могут аннигилировать, оставляя возможность существования только длинноимпульсной генерации при слабом насыщении поглотителя. Если же нарушено условие  $\Delta g \ll g_0$  или  $l_0 \approx q_0$ , то бифуркации рождения длинноимпульсной генерации не существует, и устойчивое солитонное решение всегда соответствует режиму сильно насыщенного поглотителя при мягком режиме возбуждения.

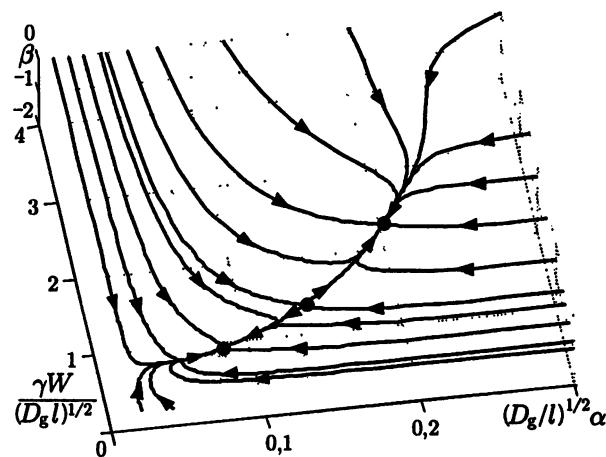


Рис. 6. Фазовое пространство  $(W, \alpha, \beta)$  в случае сильного насыщения поглотителя. Вычисления проводились для  $g_0/l = 1,007$ ;  $d = 0,5$ ;  $(l_0 + q_0)/l_0 = 2$ ;  $\delta/\gamma = 0,5$  и  $\gamma W_0/(D_g l)^{1/2} = 27$

### 2.3. Солитонная генерация в режиме полного насыщения нелинейного поглотителя

Режим генерации УКИ с достаточно большой интенсивностью ( $|A|^2 \tau_A/E_A \gg 1$ ) можно описать в приближении полностью насыщенного нелинейного поглотителя, когда практически во всей энергосодержащей области волнового пакета  $q = 0$ , и нелинейный поглотитель служит только для стабилизации среды от возбуждения низкоинтенсивных шумов вне импульса (рис. 7). При этом гайн на обходе превышает линейные потери, т. е.  $g_m > l_0$ , а энергетический баланс достигается либо за счёт конечной полосы усиления ( $D_g > 0$ ), либо вследствие локального насыщения гейна (9). Как оказывается, учёт локального насыщения гейна практически полностью аналогичен случаю медленного насыщающегося поглотителя и поэтому будет рассмотрен в соответствующем разделе ниже. Здесь же мы остановимся на случае конечной полосы усиления.

При этом генерация УКИ описывается уравнением (7), в котором следует положить  $q = 0$

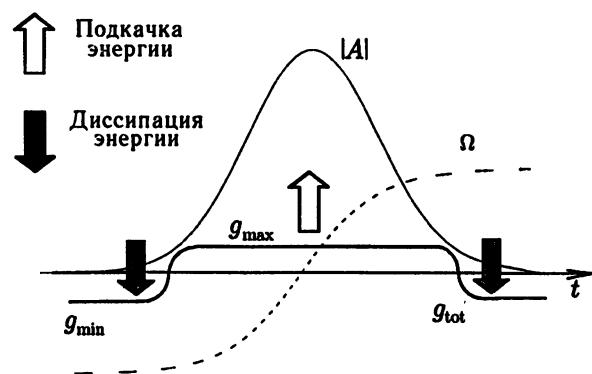


Рис. 7. Распределение  $|A|$ ,  $\Omega$  и  $g_{\text{tot}} = g_0 - l_0 - q_0/(1 + |A|^2 \tau_A/E_A)$  для случая полного насыщения поглотителя. Здесь  $g_{\text{min}} = g_m - l$ ,  $g_{\text{max}} = g_m - l$ ,  $g_{\text{tot}} = g_m - l_0$

и  $g = g_m = g_0/(1 + W/W_0)$ . Заметим, что в рамках данного приближения можно отвлечься от конкретного механизма насыщения поглотителя. Можно показать, что в полученном уравнении с формально неустойчивым нулевым фоном ( $g_m > l_0$ ) солитонные решения вида (15) существуют при любых физических параметрах задачи, более того, они всегда глобально устойчивы (аттрактор в фазовом пространстве  $(W, \alpha, \beta)$  системы). Эти импульсы характеризуются нарастанием частоты от переднего фронта импульса к заднему с коэффициентом модуляции  $\beta_S = -[(9d^2 + 8b^2)^{1/2} + 3d]/(2b)$ . Энергия солитона однозначно определяется из решения уравнения  $(W_{sol}/W_0 + 1)W_{sol}^2/W_*^2 + W_{sol}/W_0 + 1 - g_0/l = 0$ , где  $W_* = 36D_g l_0 (1 + d^2)(2 - \beta_S d)/\delta^2$ , а ширина и амплитуда находятся из следующих соотношений:  $\tau_S = -6D_g \beta_S (1 + d^2)/(W_{sol} \delta)$  и  $A_{0S} = [W_{sol}/(2\tau_S)]^{1/2}$ .

Заметим, что в случае  $d \gg 1$  модуляция частоты импульса значительно превосходит его обратную ширину, т. е.  $|\beta_S| \gg 1$ . Это означает, что для получения спектрально ограниченного импульса (длительность которого определяется обратной шириной спектра) генерируемый импульс должен быть эффективно сжат (в идеальном случае — приблизительно в  $|\beta_S|$  раз) на выходе из активной среды (лазерного резонатора) в оптической системе с отрицательной дисперсией групповой скорости. В зависимости от конкретной реализации лазера это может быть как линейная, так и нелинейная оптическая система [26].

### 3. СЛУЧАЙ МЕДЛЕННОГО НАСЫЩАЮЩЕГОСЯ ПОГЛОТИТЕЛЯ

В данном предельном случае можно, напротив, пренебречь слагаемым  $(q - q_0)/\tau_A$  по сравнению с производной  $dq/dt$  в (8). Предположим также слабость насыщения поглотителя:  $W/E_A \ll 1$ . В этом случае выражение для разности населённостей  $q$  имеет вид

$$q(t) = q_0 - \frac{q_0}{E_A} \int_{-\infty}^t |A|^2 dt, \quad (21)$$

а для определения коэффициента усиления мы по-прежнему воспользуемся приближением (17). УГЛ при этом принимает вид

$$\frac{\partial A}{\partial z} = [g/(1 + W/W_0) - l] A + D_g (1 + id) \frac{\partial^2 A}{\partial t^2} - i\delta A |A|^2 + \frac{q_0}{E_A} \int_{-\infty}^t |A|^2 dt. \quad (22)$$

#### 3.1. Общие свойства решения

Уравнение (22) также имеет решение в виде локализованного импульса [9, 14]:

$$A = \frac{A_0 e^{i(\Omega t - hz)}}{\text{ch}^{1+i\beta} [(t - Vz)/\tau]}, \quad (23)$$

а параметры импульса имеют вид

$$\begin{aligned} \beta &= -\frac{3d}{2} - \sqrt{\frac{9d^2}{4} + 2}, \quad \tau_n = W\tau\delta/D_g = -6\beta(1 + d^2), \\ h &= ld(1 - D_g\Omega^2) + l\beta(2 - 3d^2), \quad \Omega = -\frac{\beta}{2(1 + \beta^2)} \frac{q_0}{E_A D_g} A_0^2 \tau^2, \quad A_0^2 = W/(2\tau), \\ g_m &- l + \frac{q_0}{E_A} A_0^2 \tau - l D_g \Omega^2 - \frac{1 + \beta^2}{3} \frac{D_g}{\tau^2} = 0, \quad V = \frac{q_0}{3E_A} A_0^2 \tau^2. \end{aligned} \quad (24)$$

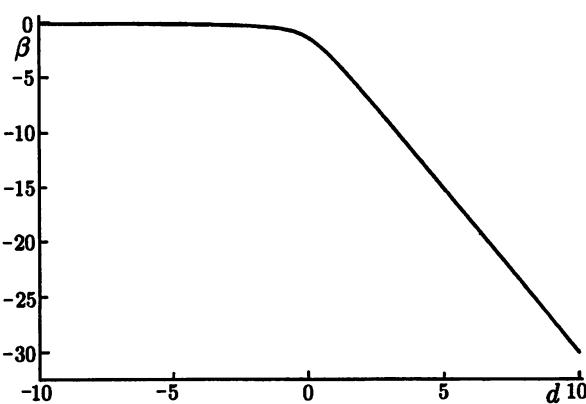


Рис. 8. Зависимость коэффициента частотной модуляции автосолитона  $\beta$  от нормированной дисперсии

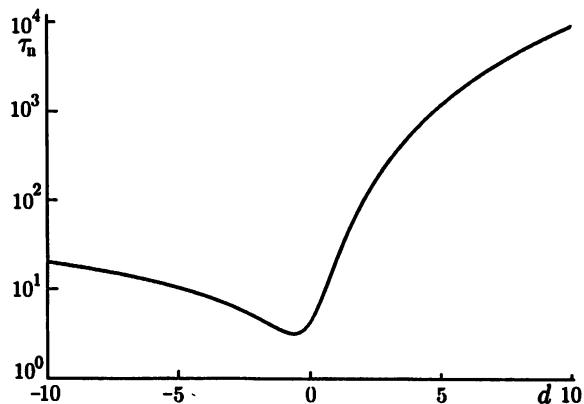


Рис. 9. Зависимость длительности солитона  $\tau_n$  от нормированной дисперсии

Как и в случае быстрого насыщающегося поглотителя, зависимость параметров  $\beta$  и  $\tau_n$  импульса от коэффициента дисперсии  $d$  обладают ярко выраженной асимметрией в зависимости от знака  $d$  (рис. 8, 9): модуляция  $\beta$  близка к нулю при отрицательных  $d$  и стремится к бесконечности при больших положительных  $d$ , а длительность импульса  $\tau_n$  существенно меньше для отрицательных значений  $d$ . Длительность автосолитона и его энергия определяются последним из уравнений (24). Это уравнение баланса энергообмена между автосолитоном и средой. Источниками пополнения энергии автосолитона являются его усиление и просветление нелинейного поглотителя. Убыль энергии определяется линейным поглощением и спектральными потерями, которые связаны с отстройкой частоты автосолитона от минимума спектра поглощения среды и фильтрацией его крайних спектральных компонент.

Баланс влияния физических факторов различной природы на структуру автосолитона реализуется за счёт специфического энергообмена между полем и средой и внутри автосолитона. Ключевую роль здесь играет модуляция импульса, источниками которой служат фазовая самомодуляция и дисперсия групповых скоростей. Генерация новых спектральных компонент поля стабилизируется на высокочастотном краю спектрально-селективным поглощением среды, а на низкочастотном — неоднородными во времени потерями, связанными с насыщением поглотителя. На переднем фронте автосолитона преобладают потери, в то время как на заднем фронте усиление больше потерь (рис. 10). Баланс структуры импульса поля поддерживается за счёт перетока энергии от его заднего фронта к переднему. Энергообмен внутри автосолитона связан с влиянием дисперсий групповых скоростей и поглощения на модулированный импульс, обладающий отстройкой частоты. Направление потока энергии, вызванного дисперсией групповых скоростей, зависит от знака дисперсии, поэтому зависимости параметров автосолитона от  $d$  обладают ярко выраженной асимметрией.

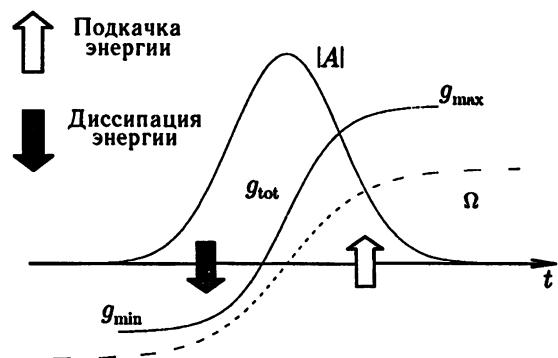


Рис. 10. Распределение амплитуды поля  $|A|$ , отстройки частоты  $\Omega$  в импульсе и полного гейна  $g_{\text{tot}} = g_m - l + q_0 \int (|A|^2 / E_A) dt$ , соответствующих (22), (23). Здесь  $g_{\min} = g_m - l$ ,  $g_{\max} = g_m - l + q_0 W / E_A$

### 3.2. Формирование шрёдингеровского солитона в случае пренебрежения диссипативными факторами

В работах [15, 16] рассмотрен важный частный случай формирования УКИ в случае, когда основной вклад в формирование импульса вносят консервативные факторы — дисперсия групповых скоростей (причём она предполагается отрицательной) и фазовая самомодуляция. Действие остальных факторов при этом учитывается по теории возмущений.

В случае учёта только дисперсии и фазовой самомодуляции уравнение (22) сводится к нелинейному уравнению Шрёдингера, решение которого (шрёдингеровский солитон) хорошо известно:

$$A = \frac{A_0}{\operatorname{ch}[(t + 2D\nu_0 z - t_0)/\tau]} \exp[-i\nu_0 t - i\tau^{-2} D\nu_0^2 z/\tau^2 + i\theta_0], \quad (25)$$

Здесь  $\nu_0$ ,  $t_0$  и  $\theta_0$  — произвольные константы. Для шрёдингеровского солитона справедливо соотношение

$$\tau = 4|D|/(\delta W), \quad (26)$$

определенное длительность импульса в условиях стационарной генерации.

В работах [15, 16] проведено экспериментальное исследование и численное моделирование лазерной генерации в данном режиме и показано, что в определённой области параметров имеет место устойчивая генерация лазерных импульсов с длительностью, соответствующей (26). При этом длительность импульса много меньше времени релаксации поглотителя. Несмотря на существование окна усиления на заднем фронте импульса за счёт насыщения поглотителя, нарастание шумов в резонаторе не происходит из-за разницы групповых скоростей импульса и шумов, и решение (25) является устойчивым.

### 3.3. Диссипативные оптические солитоны

Учтём в (22) влияние диссипативных факторов, но рассмотрим для простоты случай бесконечно широкой полосы усиления ( $D_g = 0$ ). Тогда у (22) появляется новый класс солитонных решений — диссипативные оптические солитоны [17].

В данном случае среда до прихода импульса по-прежнему считается устойчивой за счёт нелинейного поглотителя с малой энергией насыщения, при насыщении которого полный гейн  $g_m - l + + q_0 \int (|A|^2/E_A) dt$  становится положительным. Полученное уравнение обладает важным интегральным свойством: энергия любого локализованного волнового распределения достаточно быстро выходит на постоянное значение

$$W_{\text{const}} = \left( \frac{l}{q_0} E_A - \frac{W_0}{2} \right) + \sqrt{\left( \frac{l}{q_0} E_A - \frac{W_0}{2} \right)^2 - 4E_A W_0 \frac{g_0 - l}{q_0}} \quad (27)$$

(при этом пакет может и не прекращать своей эволюции), которое изменяет полный коэффициент усиления с отрицательного  $-q_0 W_{\text{const}}/(2E_A)$  на переднем фронте до такого же по величине, но положительного на заднем фронте (рис. 11). Спектральные компоненты шума, имеющие меньшую по сравнению с самим импульсом групповую скорость, оказываются на его заднем фронте, где коэффициент усиления положительный. В то же время спектральные компоненты на переднем фронте импульса уменьшают свою энергию за счёт наличия поглотителя. Очевидно, в таких условиях в среде с дисперсией групповой скорости должны появляться волновые структуры со спектром, постоянно смещающимся по частотной полосе в область, соответствующую меньшим групповым скоростям. В координатном представлении им соответствуют самозамедляющиеся волновые пакеты — диссипативные солитоны: в системе отсчёта локального времени импульс начинает сдвигаться в сторону больших значений  $t$ .

В классе стационарных импульсов формально имеются солитонные решения типа (23) со сдвигом, но без модуляции частоты и движущиеся с постоянной скоростью (шредингеровские солитоны) [18]. Однако они оказываются неустойчивыми за счёт нарастания шумов на заднем фронте импульса, где гейн положительный.

В [17] показано, что атTRACTорами в эволюции произвольного начального распределения в этой задаче являются солитоны со стационарным профилем амплитуды:

$$A(z, t) = \left[ \frac{E_A (l - g_m)^{3/2}}{q_0 |D|^{1/2}} \right]^{1/2} \tilde{A} \left[ \sqrt{l - g_m} t / \sqrt{|D|} - a (l - g_m)^2 z^2 \right]. \quad (28)$$

В системе отсчёта локального времени они смещаются с постоянным ускорением

$$\frac{d\omega}{dz} = \frac{|D|^{1/2} (l - g_m)^{3/2} a}{D}, \quad (29)$$

имеющим для заданных параметров задачи единственное безразмерное собственное значение  $a = a(\chi)$ , где  $\chi = \delta(E_A/q_0)[(l - g_m)/|D|]^{1/2}$ . Солитоны существуют при любой величине консервативной нелинейности (в том числе и при  $\delta = 0$ ) и любом знаке дисперсии групповой скорости волны, хотя модуляция частоты и направление смещения «центра масс» спектра  $\omega$  меняют знак при изменении знака  $D$ . Вследствие ускорения волна имеет асимметричный профиль интенсивности, который, например, для случая  $\delta = 0$  представляется в виде

$$\tilde{A}(\xi) = \frac{1}{(4a\xi)^{1/4} \operatorname{ch}[(\xi/a)^{1/2} + C]},$$

где  $a = a(0) = 0,478$  и  $C = 2,96$ . Собственные значения ускорения и собственные функции  $\tilde{A}(\xi)$  для произвольных  $g$  найдены в работе [17]. Таким образом, в широкополосных активных средах, в том числе моделирующих лазерную динамику, существуют диссиPATивные солитоны с постоянным дрейфом частотного спектра, энергетический баланс которых осуществляется за счёт перекачки энергии из активной области на заднем фронте импульса в область действия поглотителя на переднем фронте, а временной профиль определяется балансом нелинейности гейна и дисперсии групповой скорости.

Заметим, что структуры, аналогичные рассмотренным диссиPATивным солитонам, могут существовать и в лазерах с быстрым насыщающимся поглотителем при учёте локального насыщения гейна. Действительно, при этом уравнение (18) аналогично (22) с точностью до замены  $q_0/E_A$  на  $-g_m/W_S$ , т. е. при этом коэффициент перед  $\int |A|^2 dt$  имеет противоположный знак. Поэтому у таких импульсов усиление на переднем фронте больше, чем на заднем, т. е. знак ускорения противоположен, и солитоны ускоряются в лабораторной системе отсчёта. В остальном к ним применимы математические соотношения и выводы, указанные выше.

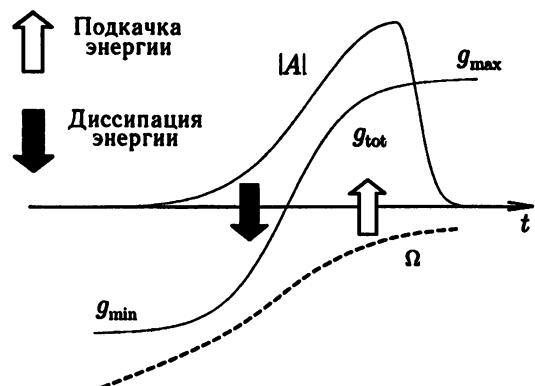


Рис. 11. Распределение  $|A|$ ,  $\Omega$  и  $g_{tot}$  для диссиPATивного оптического солитона. Здесь  $g_{min} = -q_0 W_{const}/(2E_A)$ ,  $g_{max} = |g_{min}|$ . Асимметрия распределения  $|A|$  и  $\Omega$  обусловлена дрейфом импульса в спектральном пространстве

### 3.4. Режим лазерной генерации в области неустойчивости солитонного решения

В области параметров, соответствующих неустойчивости решения (23), имеет место специфическая циклическая динамика лазерного поля. Как оказывается, решение (23) может обладать двумя типами неустойчивости в зависимости от параметров.

Первый тип неустойчивости связан с нарастанием мелкомасштабных флюктуаций поля в резонаторе. Дело в том, что при насыщении инерционного нелинейного поглотителя ультракоротким лазерным импульсом в течение определённого промежутка времени усиление преобладает над потерями. Это временное окно усиления жёстко привязано к самому импульсу, а его длительность определяется глубиной модуляции нелинейного поглощения и временем  $\tau_A$ . Таким образом, флюктуация поля с масштабом меньше  $\tau_A$ , возникшая вслед за импульсом, будет обладать меньшими нелинейными потерями. Однако существуют стабилизирующие факторы, связанные с дисперсионным расплыванием мелкомасштабной флюктуации поля и отличием групповых скоростей флюктуации поля и автосолитона. Баланс этих факторов определяет границы области неустойчивости мелкомасштабных флюктуаций лазерного поля. Можно показать, что граница зон устойчивости солитона определяется соотношением

$$l_{\text{sol}}^{(1)} + l_{\text{sol}}^{(2)} = l_{\text{noise}}^{\text{group}} + l_{\text{noise}}^{\text{disp}}, \quad (30)$$

где

$$l_{\text{sol}}^{(1)} = \beta^2 (1 + d^2)/4, \quad l_{\text{sol}}^{(2)} = \tilde{W}^2/[12\beta^2 (1 + \beta^2) (1 + d^2)],$$

$$l_{\text{noise}}^{\text{group}} = 9\beta^4 (1 + d^2)^2/4 [1 + \beta^2]^2, \quad l_{\text{noise}}^{\text{disp}} = 2,338 (\tilde{W}\xi)^{2/3} \operatorname{Re} \sqrt[3]{(1 + id)/4}, \quad \xi = \delta E_A / (\tau_A q_0),$$

а критерий устойчивости решения имеет вид

$$\xi < \xi^* = \frac{1}{8} \left( \frac{3}{2,338} \right)^{3/2} \frac{1}{[\operatorname{Re} \sqrt[3]{1 + id}]^3} \frac{\beta}{1 + \beta^2} \left[ \frac{9\beta^2 (1 + d^2)}{(1 + \beta^2)^2} - 1 \right].$$

Здесь введена безразмерная энергия  $\tilde{W} = W\delta^2 E_A/(D_g q_0)$ . В левой части (30) сгруппированы спектрально зависимые потери автосолитона, а в правой — потери шумовой моды. При этом  $l_{\text{noise}}^{\text{group}}$  — потери шума за счёт различия в групповых скоростях шума и солитона, а  $l_{\text{noise}}^{\text{disp}}$  — потери за счёт дисперсионного расплывания мелкомасштабных флюктуаций шума. Как оказывается, для солитона (23) всегда справедливо соотношение  $l_{\text{sol}}^{(1)} > l_{\text{noise}}^{\text{group}}$ , т. е. основным фактором, стабилизирующим автосолитон, в отличие от [16] являются потери шума за счёт дисперсионного расплывания мелкомасштабных флюктуаций, а отличие в групповых скоростях автосолитона и континуума не способно скомпенсировать потери автосолитона, связанные с его частотной отстройкой.

Зависимость границ зон устойчивости автосолитона от дисперсии для  $\xi = 0,1; 1; 10$  приведены на рис. 12. Область устойчивости автосолитона имеет перетяжку вблизи  $d = 0$ , которая сжимается с уменьшением  $\xi$ . При  $\xi < \xi^* = 0,16$  область устойчивости расщепляется на две зоны (см. рис. 12), что позволяет сформулировать достаточный критерий устойчивости автосолитона относительно мелкомасштабных возмущений континуума:  $\xi > 0,16$ .

Спектральные потери автосолитона служат источником ещё одного типа неустойчивости континуума. Этот механизм неустойчивости присущ крупномасштабным (с длительностью порядка  $T_R$ ) флюктуациям поля, поскольку в данном случае влияние дисперсии, отстройки групповой скорости автосолитона и окна усиления на баланс энергии шума несущественны. Как оказывается, границы области устойчивости автосолитона относительно роста крупномасштабных возмущений шума определяются соотношением

$$\tilde{W}_{1,2} = \frac{27\beta^2 (1 + d^2)^2}{1 + \beta^2} \left[ 1 \pm \sqrt{1 - \frac{\beta^2}{3(1 + \beta^2)}} \right]. \quad (31)$$

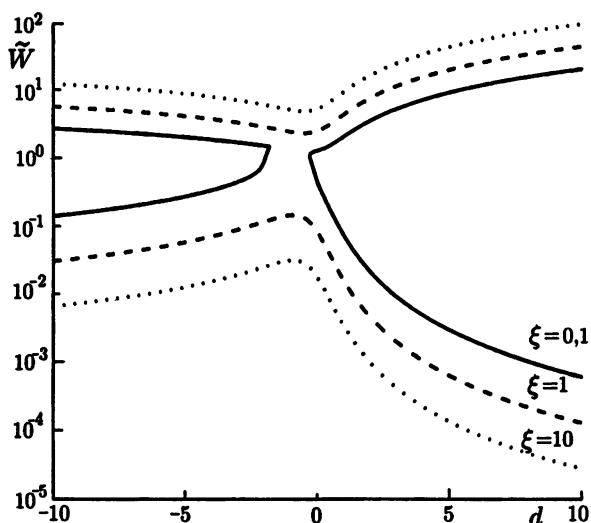


Рис. 12. Зоны устойчивости мелкомасштабных возмущений континуума для  $\xi = 10; 1; 0,1$  на плоскости  $(\tilde{W}, d)$

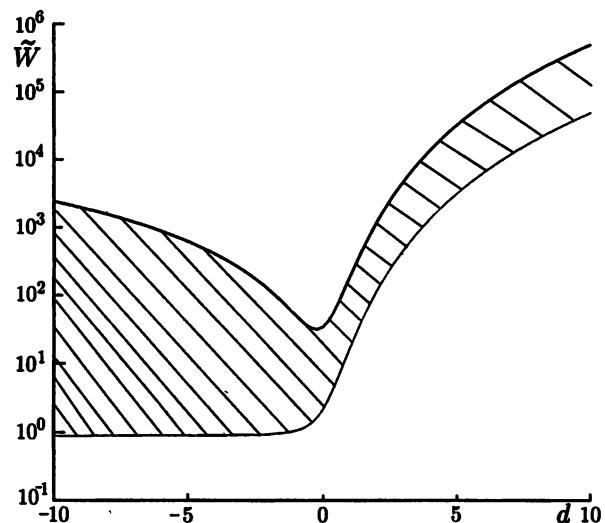


Рис. 13. Область устойчивости крупномасштабных флюктуаций шума на плоскости  $(\tilde{W}, d)$

Нижняя граница ( $W = W_2$ ) связана в основном с частотной отстройкой автосолитона от центра полосы усиления, а верхняя ( $W = W_1$ ) — с эффективной фильтрацией краёв спектра автосолитона полосой усиления. Зависимость границ зон устойчивости от дисперсии приведены на рис. 13.

Рассмотренные выше два типа неустойчивости континуума приводят к специфической циклической динамике. Особенность динамики поля обусловлена обратной связью по усилию  $g_m = g_0/(1 + W/W_0)$  и избыточными спектральными потерями автосолитона. Уровень усиления в среде зависит от полной энергии поля, поэтому рост энергии континуума сопровождается уменьшением энергии автосолитона, и наоборот. Нелинейный поглотитель создаёт условия для формирования автосолитонной структуры и перекачки энергии из континуума в автосолитон, а избыточные спектральные потери установленного автосолитона делают неустойчивым континуум, что сопровождается перекачкой энергии в обратном направлении. С помощью численных расчётов можно показать, что поведение системы можно условно разбить на три этапа: 1) формирование автосолитона; 2) неустойчивость континуума, уменьшение энергии автосолитона; 3) гибель автосолитона и зарождение в континууме новой автосолитонной структуры. В зависимости от типа неустойчивости циклическая динамика может происходить либо во временном, либо в спектральном пространстве. На рис. 14 изображён фрагмент эволюции огибающей поля при неустойчивости мелкомасштабных возмущений континуума. Начальное распределение соответствовало стационарному автосолитону (23). Здесь циклическая динамика автосолитона носит характер временных скачков. Пример эволюции спектра поля при неустойчивости крупномасштабных возмущений континуума для  $d \ll -1$  и  $W < W_2$  приведён на рис. 15. В этом случае стационарное решение (23) неустойчиво и не участвует в циклической динамике поля, которая связана с генезисом другого типа нелинейной волновой структуры — диссипативного оптического солитона. В данном случае зависимость линейных потерь от частоты проявляется в адиабатическом изменении параметров ускоряющегося автосолитона, в том числе и темпа смещения его спектра. Зарождаясь в области малых частотных отстроек, ускоряющийся автосолитон смещается в сторону низких частот, увеличивая свои спектральные потери. Это приводит к неустойчивости крупномасштабных возмущений континуума и последующей диссипации автосолитона. В целом, динамика поля носит характер циклических скачков в спектральном пространстве.

В общем случае динамика поля имеет более сложный характер, она может содержать циклические движения как во временном, так и в спектральном пространствах, а также периодическую деформацию

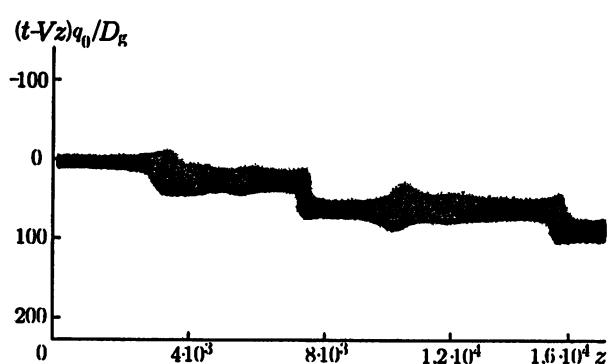


Рис. 14. Эволюция нормированного распределения интенсивности поля при неустойчивости мелкомасштабных флуктуаций континуума ( $\xi < \xi^*$ ). Вычисления проводились для  $d = -2$ ;  $\delta = \delta E_A \sqrt{l/D_g}/q_0 = 12$ ;  $\tilde{W}_0 = 38,88$ ;  $\tilde{W} = 14,4$ ;  $g_0/l = 1,32$ ;  $\xi = 0,044$ ;  $\xi^* = 0,106$

спектра и временного распределения волнового пакета, вызванную фазомодуляционной неустойчивостью [19].

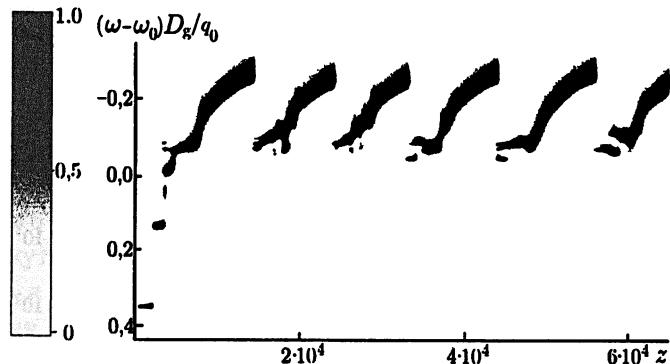


Рис. 15. Эволюция спектра волнового пакета при неустойчивости крупномасштабных флуктуаций шумового поля ( $W < W_2$ ). Вычисления проводились для  $d = -10$ ;  $\delta = \delta E_A \sqrt{l/D_g}/q_0 = 2$ ;  $\tilde{W}_0 = 1,08$ ;  $\tilde{W} = 0,4$ ;  $g_0/l = 1,45$ ;  $\xi = 0,016$ ;  $\xi^* = 0,0137$

#### 4. ЗАКЛЮЧЕНИЕ

В данном обзоре рассмотрен ряд вопросов, посвящённых исследованию различных режимов генерации ультракоротких импульсов в современных твердотельных фемтосекундных лазерах. На основе пространственно-временного подхода к описанию лазерной динамики с учётом самосогласованного поведения коэффициента усиления среды сформулирована и исследована модель генерации фемтосекундных импульсов в форме диссипативных оптических солитонов. Получен аналитический критерий существования устойчивой солитонной генерации со слабым насыщением нелинейного поглотителя. Показано, что наряду с данным режимом возможна солитонная генерация с сильным (или полным) насыщением нелинейного поглотителя, характеризующаяся минимальной длительностью и максимально возможной амплитудой импульсов. Аналитически определены структура солитонов с сильным насыщением поглотителя и область параметров их существования. Для широкополосных усиливающих сред найден новый класс диссипативных оптических солитонов, балансирующих дисперсию групповой скорости и неконсервативную нелинейность насыщения поглотителя, в форме волновых пакетов со стационарным профилем амплитуды во времени и равномерно смещающимся внутри полосы усиления спектральным профилем. Исследован сценарий циклической спектральной динамики диссипативных оптических солитонов внутри неоднородной частотной полосы усиления.

Заметим, что приведённые уравнения, а следовательно, и все результаты, полученные при их исследовании, справедливы лишь в предположении малых изменений импульса при прохождении каждого элемента резонатора. Однако по мере достижения всё более коротких импульсов (менее 10 фс, что составляет небольшое число периодов оптического поля) относительные изменения импульса в резонаторе могут составить сотни процентов, и данное приближение будет заведомо несправедливо. В настоящее время теория такой лазерной генерации развита ещё недостаточно, и данный режим изучается, как правило, в численных расчётах. Построение теории генерации предельно коротких лазерных импульсов является предметом дальнейших исследований авторов данной работы.

## СПИСОК ЛИТЕРАТУРЫ

1. Ханин Я. И. Динамика квантовых генераторов. М.: Советское радио, 1975.
2. Garmire E. M., Yariv A. // IEEE J. Quant. Electr. 1967. V. 3, No. 6. P. 222.
3. Беспалов В. И. // Изв. вузов. Радиофизика. 1968. Т. 11, № 2. С. 320.
4. Летохов В. С. // ЖЭТФ. 1968. Т. 55, № 3. С. 1077.
5. Haus H. A. // J. Appl. Phys. 1975. V. 46, No. 7. P. 3049.
6. Haus H. A., Fujimoto J. G., Ippen E. P. // JOSA B. 1991. V. 8, No. 10. P. 2068.
7. Андронов А. А. Собрание трудов. М.: Изд-во АН СССР, 1955.
8. Pereira N. R., Stenflo L. // Phys. Fluids. 1977. V. 20, No. 10. P. 1733.
9. Комаров К. П. // Оптика и спектроскопия. 1986. Т. 60, № 2. С. 379.
10. Belanger P. A., Gagnon L., Pare C. // Opt. Lett. 1989. V. 14, No. 17. P. 943.
11. Chen C.-J., Wai P. K. A., Menyuk C. R. // Opt. Lett. 1994. V. 19, No. 3. P. 198.
12. Sergeev A. M., Vanin E. V., Wise F. W. // Opt. Comm. 1997. V. 140. P. 61.
13. Sergeev A. M., Vanin E. V. // Proc. SPIE. 1995. V. 2377. P. 32.
14. Martinez O. E., Fork R. L., Gordon J. P. // JOSA B. 1985. V. 2. P. 753.
15. Kartner K. X., Kopf D., Keller U. // JOSA B. 1995. V. 12, No. 3. P. 486.
16. Kartner K. X., Jung I. D., Keller U. // IEEE J. Selected Topics in Quantum Electronics. 1996. V. 2. P. 540.
17. Vanin E. V., Korytin A. I., Sergeev A. M., Anderson D., Lisak M., Vazquez L. // Phys. Rev. A. 1994. V. 49, No. 4. P. 2806.
18. Григорьян В. С., Маймистов А. И., Скляров Ю. М. // ЖЭТФ. 1988. Т. 67. С. 530.
19. Комаров А. К., Комаров К. П., Кучьянин А. С. // Кvantовая электроника. 1998. Т. 25, № 7. С. 629.
20. Калашников В. Л., Калоша В. П., Полойко И. Г., Михайлов В. П. // Кvantовая электроника. 1996. Т. 23, № 3. С. 243.
21. Akhmediev N. N., Ankiewicz A., Lederer M. J., Luther-Davies B. // Opt. Lett. 1998. V. 23, No. 4. P. 280.
22. Chernykh A. I., Turitsyn S. K. // Opt. Lett. 1995. V. 20, No. 4. P. 398.
23. Kramer L., Kuznetsov E. A., Popp S., Turitsyn S. K. // Письма в ЖЭТФ. 1995. Т. 61, № 11. С. 887.
24. Akhmediev N. N., Afanasjev V. V. // Phys. Rev. Lett. 1995. V. 75, No. 12. P. 2320.
25. Soto-Crespo J. M., Akhmediev N. N., Afanasjev V. V. // JOSA B. 1996. V. 13, No. 7. P. 1439.
26. Jaskorzynska B., Nilsson J., Sergeev A., Vanin E. // Opt. Lett. 1995. V. 20, No. 20. P. 2123.

Институт прикладной физики РАН,  
г. Нижний Новгород, Россия

Поступила в редакцию  
6 апреля 2001 г.

## DISSIPATIVE SOLITONS IN THE COMPLEX GINZBURG–LANDAU EQUATION FOR FEMTOSECOND LASERS

*A. I. Korytin, A. Yu. Kryachko, and A. M. Sergeev*

We consider the models of generation of femtosecond radiation in solid-state lasers with passive mode locking. An approach in which ultrashort laser pulses are associated with the complex Ginzburg–Landau equation is developed. The structure of solitons and the range of parameters of their existence are analytically determined and the stability criteria of those solitons are formulated in a number of limiting cases. The scenarios of the cyclic spectral and temporal dynamics of dissipative solitons due to the instability of the background field are studied.

УДК 621.373:535

## COOPERATIVE PARAMETRIC INSTABILITY OF NATURAL MODES AND COHERENT MECHANISM OF SELF-MODE-LOCKING IN A RING CLASS-B LASER

*V. V. Kocharovskiy<sup>1,2</sup> and Vl. V. Kocharovskiy<sup>1</sup>*

Basing on recent experimental observations and numerical simulations [1–4], we develop an analytical theory of spontaneous low-threshold mode-locking in a long homogeneously-broadened fiber laser with slow relaxation of inversion. We show that the phenomenon originates from the self-consistent parametric resonance and beat-frequency-locking of paired “hot” modes taking place due to coherent nonlinear effects in the system of active two-level centres. An important role of the Risken–Nummedal–Graham–Haken [5–7] multimode instability is discussed.

### I. STATUS OF THE PROBLEM

In the present paper, we give a qualitative physical explanation and present a quantitative theory of the regular self-pulsation behavior of ring-cavity class-B lasers in the absence of any external modulators and saturable absorbers. A class-B laser [8–11] is characterized by extremely low relaxation rate of inversion,  $T_1^{-1}$ , and very high relaxation rate of polarization,  $T_2^{-1}$ , as compared to a field relaxation rate,  $T_E^{-1}$ :

$$T_1^{-1} \ll T_E^{-1} \lesssim \omega_c \ll T_2^{-1} \ll \omega_0. \quad (1)$$

We assume a homogeneous broadening of a spectral line in an active two-level medium, though a weak inhomogeneous broadening,  $1/T_2^* \lesssim T_2^{-1}$ , does not change our interpretation of the origin of self-pulsing emission under continuous pumping.<sup>1</sup> An important parameter is the so-called cooperative frequency of inverted active centers with a transition optical frequency  $\omega_0$ ,

$$\omega_c(\Delta N) = \sqrt{2\pi\omega_0 d^2 \Delta N / (3\hbar)}, \quad (2)$$

where  $d$  and  $\Delta N/N = (N_2 - N_1)/N$ , respectively, is the dipole moment and the population difference for a two-level lasing transition of an active center,  $N$  the density of active centers. For simplicity, we assume random orientations of dipole moments, which can hardly influence the consideration unless a laser is strongly saturated; in what follows the latter is not the case.

Our investigation was initiated by pioneering experimental observation of spontaneous cw-mode-locking in an erbium-doped traveling-wave fiber laser [1, 2], when the laser emits a regular train of pulses with a period  $\Delta t = L/c_0$  equal to the round-trip time in the ring cavity of length  $L$ . (For definiteness, in our model we assume the homogeneous distribution of an active medium in the cavity and neglect the dielectric permittivity of a matrix except for the expression for the velocity of light,  $c_0 = c/\sqrt{\epsilon_0} \simeq c/1.5$ .) The corresponding multimode instability of steady-state lasing under a continuous pumping, predicted in [5–7], is known as Risken–Nummedal–Graham–Haken (RNGH) one and has been studied theoretically in many works (see, e.g., [8–15] and references therein). There were estimated the instability threshold and the necessary length of a ring cavity, which have the following forms in a standard model of resonance two-level laser:

$$\eta \equiv T_2 T_E \omega_c^2 (\Delta N_p) / 2 > 1 + \left( \frac{9}{8} - \frac{1+R}{1-R} \ln R^{-1/2} \right)^{-1} > 1, \quad (3)$$

<sup>1</sup> In the case of strong inhomogeneous broadening,  $1/T_2^* \gg T_2^{-1}$ , say, in femtosecond solid-state lasers, the theory will be modified, but the effect is still present and the threshold of parametric self-mode-locking may be even lower.

$$L \gtrsim L_{\text{cr}} = 4\pi c_0 \sqrt{T_1 T_2 / [3(3 \mp 1)(\eta - 1)]}. \quad (4)$$

Here  $T_E = L/(c_0 \ln R^{-1/4})$ , where  $R$  is an intensity feedback factor determined by an output coupler,  $\Delta N_p$  is an inversion per unit volume created by pumping in the absence of lasing, and we consider longitudinal modes with given transverse structure paying no attention to possible transverse-mode (low-frequency) instabilities. According to [11, 13, 16, 17], the minus sign in Eq. (4) describes the two-level and four-level laser systems, and the plus sign should be taken for the three-level laser systems. Steady lasing of a longitudinal mode with wave number  $k_j \equiv \omega_0/c_0 + \pi(2j - 1)/L$ , where  $j$  is an integer, is unstable if

$$\left[ \frac{4\pi^2 (2j-1)^2 T_1 T_2 c_0^2}{(3 \mp 1) L^2} - 3(\eta - 1) \right]^2 < 8(\eta - 1) \left[ (\eta - 1) \left( \frac{9}{8} - \frac{1+R}{1-R} \ln R^{-1/2} \right) - 1 \right]. \quad (5)$$

Basing on the preliminary reports [18, 19], we will discuss the collective parametric effects in the coherent multimode dynamics of laser beyond this instability threshold, focusing on the self-consistent oscillations of each pair of “hot” neighboring modes. The paper is organized as follows. We begin with recent experimental results. Then, we introduce “hot” laser modes and their pairing off. After that, we discuss the physics of parametric mode-locking and present the results of the quantitative analysis. Namely, the growth rate of cooperative parametric instability is calculated and the number of self-locked modes is found. Finally, we draw the main conclusions.

## 2. EXPERIMENTAL DATA

First of all, note that the RNGH instability exists only if [13]  $R > R_{\text{RNGH}} \approx 0.29$ , when the right-hand-side inequality in Eq. (3) is fulfilled. In the experiments [1, 2], they observed self-mode-locking for two values of feedback,  $R = 0.95$  and  $R = 0.5$ , and the main features of laser dynamics were quite similar in both cases. For definiteness, we will consider the simplest case  $R = 1$ , when the second laser threshold (described by the left-hand-side inequality in Eq. (3)) is minimal,  $\eta_{\text{cr}} = 9$ ; the first (lasing) threshold is  $\eta_l = 1$  for any  $R > 0$ , so that  $\eta_{\text{cr}}/\eta_l = 9$ .

The erbium-doped single-polarization fiber laser at 1550 nm investigated in the works [1, 2] is an inverted system of three-level active centers, with a dopant density about 300 ppm, pumped by a laser diode at 980 nm with a power  $P \sim 70$  mW. In fact, due to the well-known nonlinear relation between the dimensionless pump parameter  $\eta$  of the equivalent two-level model and the physical pump parameter  $P$  [3, 11], for three-level lasers the instability-to-lasing threshold ratio,  $P_{\text{cr}}/P_l$ , can be more close to unity than the limiting value for actual two-level lasers ( $\eta_{\text{cr}}/\eta_l = 9$ ) and the corresponding value for four-level lasers (e.g., with neodimium-doped fibers) which always exceeds 9. However, this threshold scaling is not important for our consideration, which is based on the standard two-level model.

The main temporal and frequency parameters of the ring laser demonstrating self-mode-locking were the following [1, 2]:

$$T_1 \sim 5 \text{ ms}, \quad T_E \sim 0.1-5 \mu\text{s}, \quad \Delta t \sim 40-200 \text{ ns}, \quad T_2 \sim 0.15 \text{ ps}; \quad (6)$$

$$\delta = 2\pi c_0/L \sim (0.3-1.5) \cdot 10^8 \text{ s}^{-1}, \quad \omega_c(\Delta N_p) \sim 10^{10} \text{ s}^{-1}, \quad \omega_0 \simeq 1.2 \cdot 10^{15} \text{ s}^{-1}. \quad (7)$$

The above values of the round-trip time,  $\Delta t \equiv 2\pi/\delta$ , correspond to the cavity fiber length  $L \sim 10-50$  m (up to 60% of the whole fiber was erbium-activated). The average inversion of activated centers was about 1 % ( $\omega_c(N) \sim 10^{11} \text{ s}^{-1}$ ) and free lasing of many thousands of modes was possible inside the gain line. The single pulse width beyond the multimode instability threshold was about several nanoseconds. It means that from tens to hundreds of modes were locked without any external modulation and installation of additional

nonlinear elements.<sup>2</sup> The average output power of lasing was in the microwatt range, so that the resonance two-level saturation nonlinearity seems to be the only one which can be responsible for the strong mode-mode coupling.

For the above-mentioned parameters we have  $L_{\text{cr}} \sim 5-10 \text{ m}$  so that at typical values  $L \sim 20 \text{ m}$  and  $\eta \sim 10-100$  which are not very much above the thresholds (3) and (4), the RNGH instability is possible only for few modes with indices  $|j|$  close to  $L/L_{\text{cr}} \sim \sqrt{\eta - 1}$ ; see Eq. (5). Nevertheless, numerical simulations, based on the optical Maxwell–Bloch equations and carried out for this class of ring fiber lasers [1–4], show that the dynamics of the emitted intensity evolves from the sinusoidal oscillations of few initially unstable modes satisfying Eq. (5) to an instability of a large number of modes outside the interval (5) and then come to a regular regime of mode locking. These simulations are in qualitative agreement with the experimental data (where, of course, extra noise is present), but a clear physical interpretation and analytical theory are desirable.

### 3. "HOT" MODES AND THEIR PAIRING OFF

Let us begin with noting the fact that all usual necessary conditions of using the standard rate equations to describe laser dynamics [10–12] are fulfilled in the case under consideration:

$$T_2^{-1} \gg \{T_1^{-1}, T_E^{-1}, \omega_c, \omega_R\}, \quad (8)$$

where  $\omega_R = d|E_S|/\hbar$  is a typical Rabi frequency in the multimode laser field  $E_S$ . At the same time, spontaneous mode locking is impossible in the rate regime of class-B lasers and has to be attributed to weak coherent phenomena. We will show that the latter are amplified due to parametric resonance in the collective process of self-modulation of the intermode frequency spacing and can involve a large number of modes lying far outside the domain of the RNGH instability. Moreover, we expect (it will be verified elsewhere) that such an effect is possible even in the absence of the multimode RNGH instability if a sufficient number of long-living modes is excited somehow, for instance, via external initiation by a short coherent resonance pulse (hard excitation of mode locking) or due to multimode laser generation in the presence of essential inhomogeneous broadening.

Complex eigenfrequencies of natural (so-called "hot") electromagnetic modes in a ring-cavity class-B laser are well-known and can be easily found from the standard dispersion relation [23]:

$$\omega_e(k_j) \simeq \pi(2j-1)c_0/L + \omega_0 - iT_E^{-1} - \omega_c^2/[\pi(2j-1)c_0/L + iT_2^{-1}]. \quad (9)$$

For simplicity, we assume that the value  $\omega_0L/(2\pi c_0)$  is an integer and  $j - 1/2 = (k_j - \omega_0/c_0)L/(2\pi)$ . We will consider the case of quasi-stationary multimode generation, no matter what is its origin:

$$\text{Im}[\omega_e(k_j)] \simeq 0. \quad (10)$$

This makes our theory quite general, so that it may be applied to different lasers, even without knowing specific factors of energy balance in lasing modes (these factors may be attributed to various nonlinearities which differ from the resonance saturation nonlinearity of a two-level lasing medium).

Hot modes (9) are almost equidistant, and the intermode frequency spacing is as follows:

$$2\Omega \simeq \delta(1 - \omega_c^2 T_2^2) + 2\delta_j, \quad \delta_j = 3\delta^3 \omega_c^2 T_2^4 (j - 1/2)^2. \quad (11)$$

<sup>2</sup> Note that passive Q-switching of such erbium-doped fiber laser needs insertion a fast saturable absorber, e.g., a segment of samarium-doped fiber [20]. Apart from mode-locking, class-B lasers may exhibit a chaotic dynamical behavior, e.g., due to the cross-relaxation effect owing to ion clustering [21] or due to the nonlinear Kerr effect [22]; it is not what is discussed here.

Here  $\delta = 2\pi c_0/L$  is the intermode frequency spacing in a ring “cold” cavity without active two-level medium; a non-equidistant term  $\delta_j$  can include additional contribution from the cold-cavity dispersion if present. If it would be  $\delta \sim T_2^{-1}$ , then in the absence of exact medium-cavity resonance the instability of mode lasing may be brought about by a destabilization of the field phase [15]. In our case where  $\delta \ll T_2^{-1}$ , the multimode (RNGH) instability results from a destabilization of the field amplitude, and the medium-cavity frequency detuning is not important.

Let us peak out  $2j_0$  self-locking modes, unknown for a time. From a formal point of view, their field at any point in a cavity can be expressed as a superposition of fields,  $E_n$ , produced by pairs  $\{j, j+1\}$  of neighboring modes ( $n$  is an integer):

$$E_S = 0.5 E_S \exp(-i\omega_0 t) + \text{c.c.}, \quad E_S = \sum_{|n| \leq (j_0-1)/2} E_n \exp(-4in\Omega_0 t). \quad (12)$$

Here the complex amplitudes  $E_n(t)$  can be time-dependent, but we expect that all  $2j_0$  modes become equidistant,  $\Omega_j = \Omega_{n\pm} = \Omega_0$  (see Eqs. (16) and (17) below,  $\Omega_0 = \Omega_{j=1/2}$  by definition), due to parametric beat-frequency locking. In the following sections we will indicate the necessary condition of this locking and find the number  $2j_0$ .<sup>3</sup>

In the quasi-stationary regime, the field of each couple of hot modes obeys the same pendulum equation with almost zero decay rate,  $\gamma \rightarrow 0$  (cf., e.g., [23, 24]):

$$\frac{d^2 E_n}{dt^2} + 2\gamma \frac{dE_n}{dt} + \Omega^2 E_n = 0; \quad T_1^{-1} \ll \Omega \ll \omega_c. \quad (13)$$

Here the intermode frequency (11) is time-dependent due to weak coherent oscillations of inversion,  $\Delta N(t)$ , which is present in the cooperative frequency (2), according to the optical Bloch equations:

$$\frac{d(\Delta N)}{dt} + T_1^{-1} (\Delta N - \Delta N_p) = \text{Im}(E_S^* P_S)/\hbar, \quad (14)$$

$$\frac{dP_S}{dt} + T_2^{-1} P_S = -i E_S \Delta N d^2 / (3\hbar) \quad (15)$$

( $P_S$  is a complex amplitude of polarization). The system of equations (13)–(15) is quite sufficient to develop the theory of self-mode-locking. In general, using a perturbation method, we can find the solution in the following form:

$$E_n = A_{n+} \exp(-i\Omega_{n+}t - i\varphi_{n+}) + A_{n-} \exp(i\Omega_{n-}t + i\varphi_{n-}), \quad (16)$$

$$P_n = \frac{d^2 \Delta N}{3\hbar} \left( \frac{A_{n+}}{\Omega_{n+} + iT_2^{-1}} e^{-i\Omega_{n+}t - i\varphi_{n+}} - \frac{A_{n-}}{\Omega_{n-} - iT_2^{-1}} e^{i\Omega_{n-}t + i\varphi_{n-}} \right), \quad (17)$$

where  $P_S$  and  $P_n$  are defined similarly to  $E_S$  and  $E_n$  in Eq. (12). Small values  $\Omega_{n\pm} \simeq \Omega_0$  in the denominators of Eq. (17), which are much less than  $T_2^{-1}$ , play the principle part because they are responsible for the effects of coherence in the self-consistent dynamics of modes paired off.

#### 4. COOPERATIVE PARAMETRIC INSTABILITY

To simplify the calculations as much as possible we assume the inequality (see Eqs. (6) and (7))

$$T_1 \delta \gg 1 + 2j_0 A^2 / E_0^2, \quad (18)$$

<sup>3</sup> In the general case, the above-mentioned modes can be shifted from the center of spectral line even for a large frequency  $\Delta\omega \gg 2j_0\Omega_0$ ; it will introduce a minor factor  $(1 - T_2^2 \Delta\omega^2)/(1 + T_2^2 \Delta\omega^2)^2$  in the value  $\omega_c^2$  in Eq. (11) and below, and also makes the values  $\gamma_e$  and  $\delta_{0j}$  in Eqs. (19) and (20) smaller in  $(1 + T_2^2 \Delta\omega^2)$  times.

where  $E_0 = \sqrt{3} \hbar / (d \sqrt{T_1 T_2})$  is the value of saturation field, and consider the limiting case of equal mode amplitudes,  $A_{n\pm} = A$ , and zero phases,  $\varphi_{n\pm} = 0$ . A general case of “phased mode locking” with a linear frequency dependence of mode phases may be analyzed in a similar way (cf., e.g., [11]). Note that the universal form of Eq. (13) for beating of two neighboring modes makes the final result independent on the shift  $j \rightarrow j + 1$  in mode pairing if  $2j_0 \gg 1$ .

Thus, from Eqs. (14)–(17) we find the following self-consistent frequency of mode-mode beating:

$$\Omega(t) = \Omega_0 + \delta_{0j} + 2\gamma_e \sin(2\Omega_0 t); \quad \gamma_e = 2j_0 (T_2 dA/2\hbar)^2 \omega_{c0}^2 T_2 / 3, \quad (19)$$

$$2\Omega_0 = \delta (1 - \omega_{c0}^2 T_2^2), \quad \delta_{0j} = 3\delta^3 \omega_{c0}^2 T_2^4 (j - 1/2)^2, \quad (20)$$

where the time-independent cooperative frequency  $\omega_{c0} = \omega_c(\Delta N_0)$  is defined via an average inversion,

$$\Delta N_0 = \Delta N_p / \left( 1 + E_0^{-2} \sum_j |A_j|^2 \right), \quad (21)$$

saturated by all modes, including even not locked if present ( $|j| \leq j_0$ ). If the inequality (18) is broken, the frequency  $\Omega_0$  in the sine argument in Eq. (19) and in the exponent in Eq. (12) is replaced by  $\Omega_0 - \gamma_e (1 + 2j_0 A^2/E_0^2)/(T_1 \delta)$ , the phases  $\varphi_{n\pm}$  in Eq. (17) become finite, and the growth rate  $\gamma_e$  in Eq. (19) is multiplied by the factor  $[1 + (1 + 2j_0 A^2/E_0^2)^2/(T_1^2 \delta^2)]^{1/2}$ . In fact, the result (19) follows also from standard calculations of the modulation of inversion by coherent mode-mode beating, even in the rate approximation corresponding to an algebraic relation between the amplitudes of polarization and field in each mode,  $P_j/E_j \simeq -iT_2 \Delta N_0 d^2/(3\hbar)$ .

From Eqs. (13) and (19) we clearly see the effect of parametric resonance which guarantees synchronous mode beating in all couples having small enough intermode non-equidistancy,  $2|\delta_{0j}| < 2\gamma_e$ . (Again, for an actual laser we can add to the value  $\delta_{0j}$  in Eq. (20) a non-equidistant term caused by the cold-cavity dispersion.) The maximum growth rate of parametric instability,  $\gamma_e$ , is equal to the laser amplification factor  $\omega_{c0}^2 T_2 / c_0$  multiplied by the speed of light and the square of ratio between the Rabi frequency of all self-locked modes,  $d(2j_0)^{1/2} A / (\sqrt{3} \hbar)$ , and the line width  $2/T_2$  of amplifying medium. In fact, under steady-state lasing the parametric growth rates,  $(\gamma_e^2 - \delta_{0j}^2)^{1/2}$ , including non-equidistant terms will be compensated due to small nonlinear shifts of intermode frequency spacing, which are much less than  $\gamma_e$  and different for different pairs and depend on an actual inhomogeneous spectrum of mode amplitudes, as well as due to decay rates of modes,  $\gamma$ , which are influenced by standard nonlinear frequency-dependent effects of cavity losses, pump depression, etc. We won't discuss these subtle effects as well as analyze the stability of the mode-locking state. It will be done elsewhere (together with the space-time analysis of the transformation of initial field fluctuations into a final soliton-like solution).<sup>4</sup>

The well-known results of the self-mode-locking is the formation of a narrow pulse,

$$\mathcal{E}_S \simeq A \cos(\omega_0 \tilde{t}) \sin(2j_0 \Omega_0 \tilde{t}) / \sin(\Omega_0 \tilde{t}) \quad (22)$$

(in an ideal case of equal mode amplitudes with zero phases), where  $\tilde{t} = t - z / [c_0 (1 - \omega_{c0}^2 T_2^2)]$  and  $z$  is the length along a fiber. It is this pulse traveling along the cavity which produces periodical laser emission under a continuous pumping. One can expect also that in the case of a broad-band active medium with rather weak coherent modulation of intermode beating frequency the analogous parametric self-mode-locking can be obtained due to a suitable resonance absorber which provides higher value of the parametric growth rate and does not influence much the lasing threshold.

<sup>4</sup> Previous analysis of this kind [25] did not include the notion of “hot” modes and their pairing off and coherent interaction.

## 5. THE NUMBER OF LOCKED MODES

According to the parametric instability condition,  $\delta_{0j} < \gamma_e$ , in the ideal case of a non-dispersive cavity the number of self-locked modes,

$$2j_0 \simeq [dA/(\hbar\delta)]^2/(9T_2\delta), \quad (23)$$

is much larger than the number of RNGH-unstable modes and does not depend on the saturation level unless the inequality (18) is broken. However, when the value  $j_0$  becomes large, the depression of pumping efficiency per one mode may become principle because the number (23) is less than the total number of modes within the spectral line,  $2/(T_2\delta)$ , by a factor of the order of one-mode Rabi frequency over intermode spacing squared. The limiting values,

$$2j_0^* \sim [3T_1 T_2^2 \delta^3]^{-1/2}, \quad (24)$$

$$\gamma_e \sim \omega_c^2 (\Delta N_p) T_2^2 / (8T_1), \quad (25)$$

correspond to the saturation threshold,  $2j_0^* A^2 \sim E_0^2$ , in the absence of free generation of modes which are not locked (see Eq. (21)). In the experiments [1, 2] the laser radiation was weak, far below the above-mentioned saturation threshold, and the number of self-locked modes was much smaller than the characteristic value (24),  $2j_0^* \sim 10^4$ . At the same time, the estimate of a mode amplitude  $A$  from Eq. (23) with the known number  $2j_0 \sim 100$  corresponds approximately the experimental data.

In the case when  $2j_0 A^2 \gg E_0^2$  and the inequality (18) is still valid, the number of locked modes can be greater than  $2j_0^*$  in many times although the growth rate of parametric instability remains at the same level (25) because the cooperative frequency squared decreases as  $(2j_0 A^2)^{-1}$  (see Eqs. (2) and (19)).

The latter is especially important if the mode non-equidistancy,  $\delta_j$ , is defined by the linear dispersion of a cold cavity instead of the nonlinear dispersion of an active medium (cf. Eq. (11) where  $\delta_j \propto \omega_c^2 (\Delta N) \propto (2j_0 A^2)^{-1}$ ). Then the self-mode-locking is more difficult and the condition  $|\delta_j| < \gamma_e$  limits a possible growth of the number of locked modes beyond the saturation threshold,  $2j_0 A^2 \gtrsim E_0^2$ , where the parametric growth rate is limited by the value

$$\max \gamma_e = \omega_c^2 (\Delta N_p) T_2^2 / (4T_1). \quad (26)$$

This value may be increased by increasing the relaxation rate of inversion,  $T_1^{-1}$ , though it is also limited by the inequality (18) as well as by an additional parasite non-equidistancy of mode spacing which can appear due to the nonlinear dispersive effects in laser fields exceeding the saturation field  $E_0 = \sqrt{3} \hbar / (d \sqrt{T_1 T_2})$ .

Anyway, a decrease in the intermode spacing in longer cavities will favor the self-mode-locking. This fact as well as other features of the phenomenon show that the mode locking of parametric origin described above is essentially different from usual passive mode locking in class-B lasers with saturable absorber [11].

## 6. CONCLUDING REMARKS

In this paper, we introduce the pairing decomposition of a multimode laser field, calculate the parametric growth rate for each pair of neighboring modes, find the necessary conditions of the corresponding collective parametric instability, and estimate the number of pairs which can be self-locked. As a result we clarify the physical interpretation of the original experimental and numerical data [1–3] describing the spontaneous pulsed generation in long ring fiber lasers.

The importance of self-parametric effects in laser physics was noted many times since, to our knowledge, the work [8]. In the present paper we demonstrate that the coherent effect of intermode parametric resonance is extremely efficient in class-B lasers under continuous pumping. When for some reasons, e.g., due to the multimode RNGH instability, a quasi-stationary generation at least of two neighboring longitudinal modes becomes possible, the cooperative parametric resonance of mode-mode beating can favor

the generation of many other pairs of modes, even if their individual generation (or instability) threshold is not satisfied. This effect, originating from weak coherent self-modulation of frequency spacing between hot modes, gives natural explanation of the multimode instability in class-B lasers and open new way of spontaneous mode locking, in particular, in femtosecond solid-state lasers.

#### REFERENCES

1. Fontana F., Begotti M., Pessina E. M., Lugiato L. A. // Opt. Comm. 1995. V. 114. P. 89.
2. Pessina E. M., Bonfrate G., Fontana F., Lugiato L. A. // Phys. Rev. A. 1997. V. 56. P. 4086.
3. Pessina E. M., Prati F., Redondo J., Roldan E., de Valcarcel G. J. // Phys. Rev. A. 1999. V. 60. P. 2517.
4. Voigt T. M., de Valcarcel G., Roldan E., Lenz M. O., Mitschke F. // Digest of Int. Quantum Electronics Conference (September 10–15, 2000, Nice, France). P. 65.
5. Risken H., Nummedal K. // Phys. Lett. A. 1968. V. 26 P. 275.
6. Risken H., Nummedal K. // J. Appl. Phys. 1968. V. 39 P. 4662.
7. Graham R., Haken H. // Z. Phys. 1968. V. 213. P. 420.
8. Dement'ev V. A., Zubazev T. N., Oraevsky A. N. // Trudy FIAN. 1977. V. 91. P. 3.
9. Instabilities and Chaos in Quantum Optics / Ed. by F. T. Arecchi, R. G. Harrison. Berlin: Springer-Verlag, 1987.
10. Weiss C. O., Vilaseca R. Dynamics of Lasers. Weinheim: VCH, 1991.
11. Khanin Ya. I. Fundamentals of Laser Dynamics. Moscow: Nauka-Fizmatlit, 1999.
12. Samson A. M., Kotomtseva L. A., Loiko N. A. Self-oscillations in Lasers. Minsk: Navuka i Tekhnika, 1990.
13. de Valcarcel G. J., Roldan E., Prati F. // Opt. Comm. 1999. V. 163. P. 5.
14. Gerber P. R., Bütticker M. // Z. Phys. 1979. V. 53. P. 219.
15. Narducci L. M., Tredicce J. R., Lugato L. A., Abraham N. B., Bandy D. K. // Phys. Rev. A. 1986. V. 33. P. 1842.
16. Roldan E. // Opt. Comm. 1997. V. 143. P. 235.
17. Roldan E., de Valcarcel G. J. // Europhys. Lett. 1998. V. 43. P. 255.
18. Kocharovskiy V. V., Kocharovskiy Vl. V. // Bull. Russian Acad. Sci. 2000. V. 64. Proc. VIII Intern. Readings on Quantum Optics, Kazan, Russia, October 27–29, 1999. P. 1722.
19. Kocharovskiy V. V., Kocharovskiy Vl. V. // Digest of Int. Quantum Electronics Conference (September 10–15, 2000, Nice, France). P. 143.
20. Luo L. G., Chu P. L. // Opt. Comm. 1999. V. 161. P. 257.
21. Sanchez F., Stephan G. // Phys. Rev. E. 1996. V. 53. P. 2110.
22. Abarbanel H. D. I., Kennel M. B., Buhl M., Lewis C. T. // Phys. Rev. A. 1999. V. 60. P. 2360.
23. Zheleznyakov V. V., Kocharovskiy V. V., Kocharovskiy Vl. V. // Sov. Phys. Uspekhi. 1989. V. 32. P. 835.
24. Kocharovskiy V. V., Kocharovskiy Vl. V. // Radiophys. Quantum Electron. 1985. V. 28. P. 1099.
25. Haus H. A. // J. Appl. Phys. 1975. V. 46. P. 3049.

<sup>1</sup> Институт прикладной физики РАН,  
г. Нижний Новгород, Россия

Поступила в редакцию  
17 мая 2001 г.

<sup>2</sup> Department of Physics and Institute for  
Quantum Studies, Texas A&M University,  
College Station, Texas, USA

УДК 530.182

## NONLINEAR DYNAMICS OF ROCK: HYSTERETIC BEHAVIOUR

*L. A. Ostrovsky<sup>1,2</sup> and P. A. Johnson<sup>3</sup>*

We discuss the dynamics of media with hysteretic stress-strain properties. First, experimental evidences of hysteretic behaviour of rock are presented. This evidence leads to a pattern of unifying behaviour whose origin is within the “bond system” of the material, that includes small cracks, intergrain contacts and the cement that holds the grains in place. Nonlinear response is evident over a large frequency interval (dc to several MHz at least), and it is significant from strains that plasticity down to the smallest measurable strains (order  $10^{-8}$ – $10^{-9}$ ). Second, some models related to such a behaviour are outlined. Finally, oscillations and waves in hysteretic media are discussed.

## INTRODUCTION

The famous book “Oscillation theory” by Andronov, Witt, and Khaikin [1] provides a clear idea of phase plane, limit cycle, and, finally, hysteresis as a multivalued response of a system to a forcing dependent on the history of the process. Topologically speaking, these notions are associated with some closed figures, loops, on the plane of dynamic parameters (phase plane) or that of material parameters (hysteretic equation of state, EOS). Although hysteretic behaviour of ferromagnetics and ferroelectrics and their mechanical analogues has been studied by physicists for decades, little has been addressed regarding the theory of waves in such media except for some specific problems such as electromagnetic shock waves (e.g., [2]). It has long been known that metals can have a hysteretic mechanical EOS which in the simplest form is a dependence between the stress  $\sigma$  and strain  $\epsilon$ .

Beginning in the 1940's, Birch's group at Harvard University began the study of the static nonlinear properties of rock. These studies, conducted in large mechanical presses, were designed to interrogate the EOS in rock samples at strong, low frequency forcing (near dc). Static tests were aimed at understanding the nature of the earth's interior in terms of pressure and temperature response and to predict mineral assemblages and their phases in the earth's lower crust and mantle. These studies have been invaluable in understanding and modeling physical properties and inferring the composition of the earth's interior. In the early 1970's, Stacy's group at the University of Queensland wrote a series of papers on mechanical hysteresis in rocks at low strains. Their work showed that nonlinear response persists in quasistatic tests at strains as low as  $10^{-6}$ , and perhaps lower. In the early 1980's, Bakulin and Protosenya and a group at the Radiophysical Institute and the Institute of Applied Physics in Nizhny Novgorod began studies of nonlinear response in earth materials.

As early as 1986, there was an international symposium on nonlinear seismology organised by the Moscow Earth Physics Institute, held in Suzdal, USSR, and soon thereafter, a special issue of *Physics of the Earth and Planetary Interiors* (V. 50, No. 1, 1987) devoted to nonlinear seismology appeared.

Studies of mechanical hysteresis in the EOS began to build up when our groups at the Institute of Applied Physics and the Los Alamos National Laboratory began to actively investigate dynamical (vibrational, acoustical) properties in earth materials. Most attention in our recent research has been paid to rock. The cumulative results of these studies confirmed that rock nonlinearity is very strong compared to that of “normal” media like fluids and ideal crystals, and the nonlinearity reveals itself even at extremely small, “acoustical” strain amplitudes. Relevant to that, in 1996–1999, the first four International Workshops on Nonlinear Mesoscopic Elasticity were held at the Institute of Geophysics and Planetary Physics

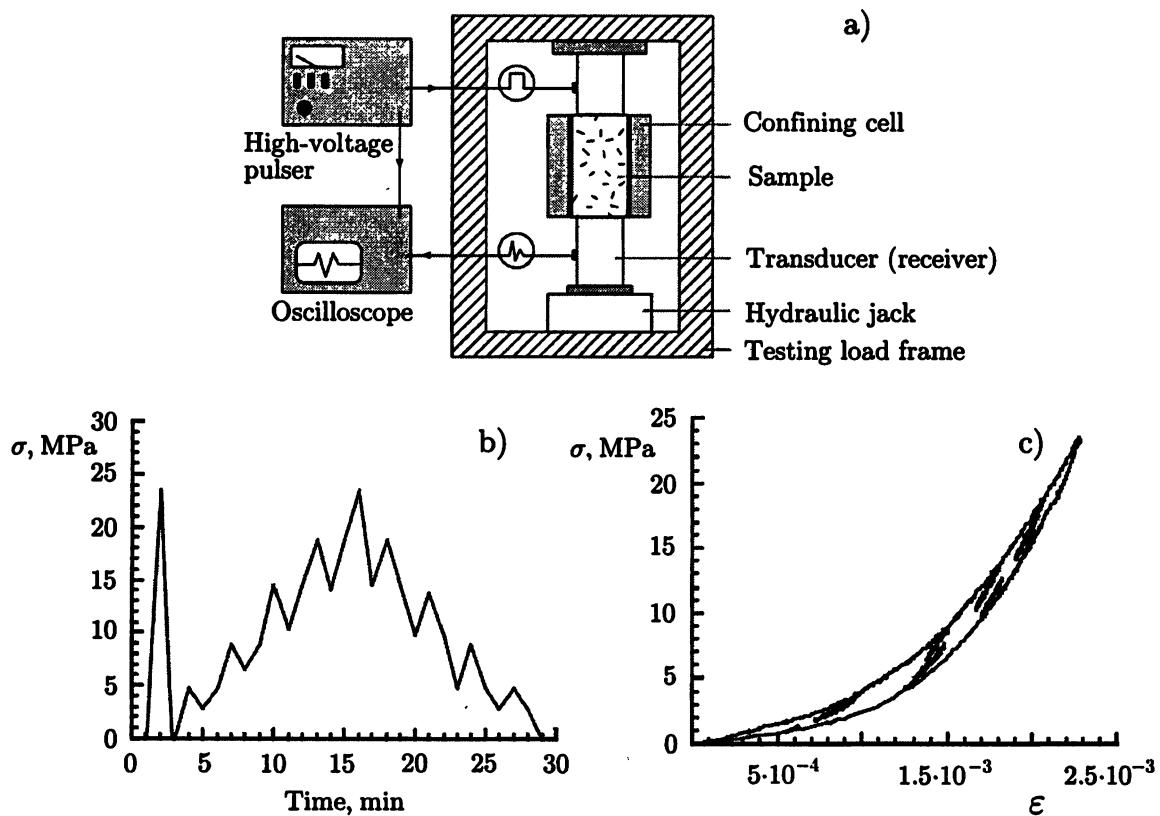


Fig. 1. Static stress-strain experiment: Typical experimental configuration for a uniaxial stress (a). Stress history or "protocol" (b). Resulting stress-strain dependence for sandstone (c). The plot illustrates a nonlinear stress-strain relation, hysteresis, and end-point (discrete) memory

at Los Alamos National Laboratory. Here, we outline some facts and ideas; an interested reader can find an additional material and detailed references in the illustrative paper [3] and a detailed review [4].

## 1. SOME EXPERIMENTAL INDICATORS OF HYSTERETIC NONLINEARITY IN ROCKS

Dynamic nonlinear response may manifest itself in a variety of manners. Let us outline some key indicators of nonlinear hysteretic behaviour of rock from static and, mostly, dynamic data of laboratory experiments.

### 1.1. Quasi-static experiments

The most direct observation of elastic nonlinearity in solids comes from quasi-static tests of stress versus strain. Figure 1 shows experimental results illustrating such a dependence [5]. Primary characteristics illustrated by such an experiment are: (1) extreme nonlinearity in the stress-strain dependence, (2) hysteresis (i.e., the behaviour depending on stress history) and, (3) "discrete memory" (also called "end point memory"). Discrete memory can be described as follows. If a partial stress cycle is conducted during the quasistatic cycle (e.g., small loops inside the big loop in Fig. 1c), the outer (low frequency) loop is maintained; discrete memory is a memory of the previous maximum strain state. Similar results have been obtained by numerous other researchers since the 1970s (e.g., [6, 7]).

These manifestations have important consequences for the material elastic modulus defined as the derivative of the stress over the strain,  $\partial\sigma/\partial\epsilon$ . In short, static tests indicate that the value of the modulus depends on the stress history and the current EOS amplitudes, and changes discontinuously at the stress-strain cusps.

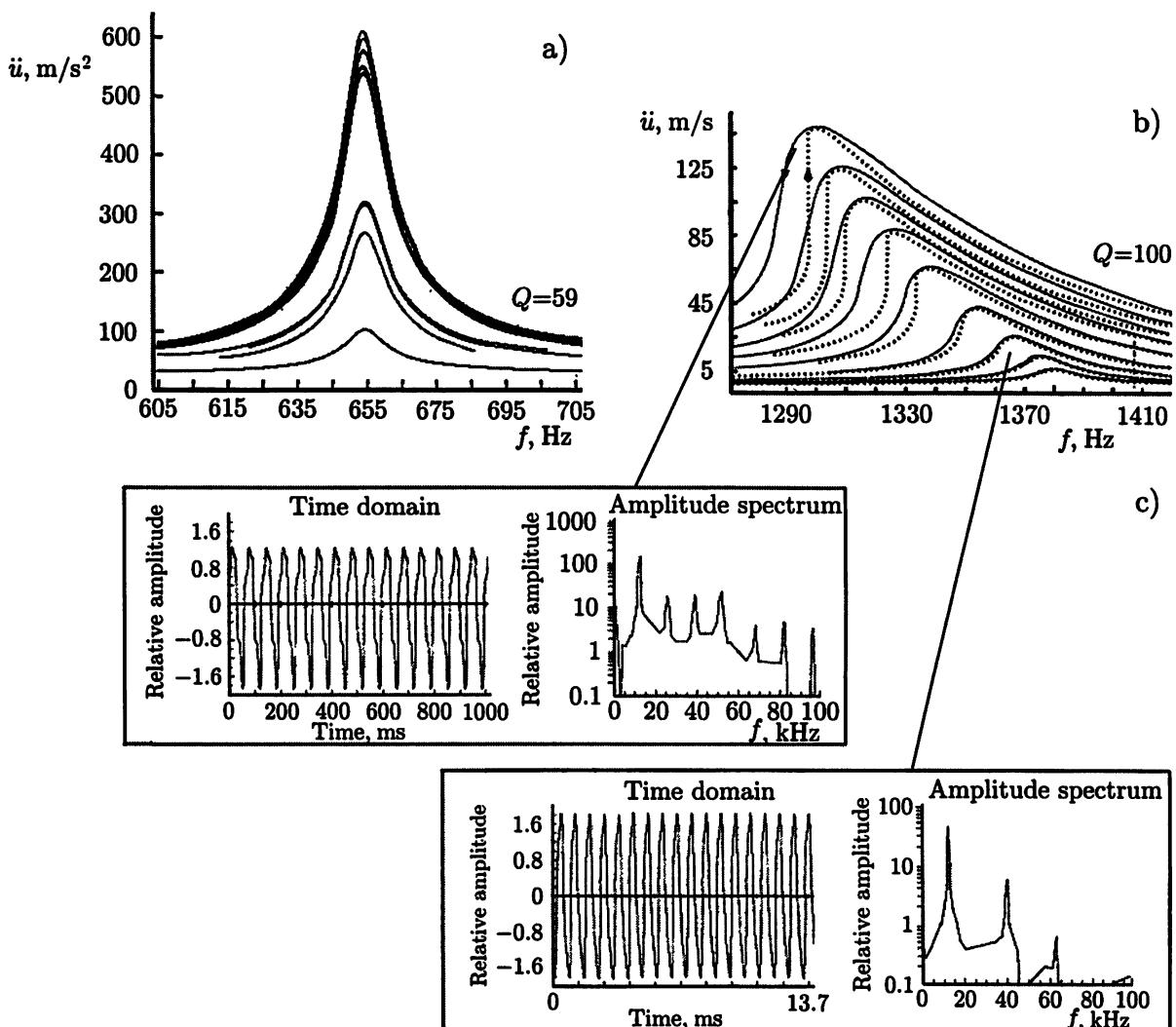


Fig. 2. Resonance acceleration response of polyvinylchloride (PVC) for several drive levels (a). Resonance acceleration response of a Fontainebleau sandstone bar, for increasing drive (b). Time and frequency domain signals from relatively low-amplitude (albeit already nonlinear) drive levels (bottom), and at large drive levels (top) (c). The time signals are obtained at peak resonances in each case,  $Q$  is the quality factor of a sample, and  $f = \omega / 2\pi$

### 1.2. Dynamic experiments

There exist numerous methods by which to observe dynamic nonlinear effects. In the ultrasonic range typically used in the acoustics of liquids and gases, the harmonics of a periodic travelling wave can be monitored out to the distance of shock wave formation. However, for relatively low-frequency and low-amplitude sound in solids, it is often difficult to obtain quantitative results from travelling wave experiments. The majority of quantitative measurements for rocks have been performed with resonant bars. Due to the amplification that resonance provides, it is perhaps the most sensitive manner by which to observe nonlinear behaviour, even at extremely small exciting strains, even at  $\varepsilon = 10^{-9}$  in a simple one-dimensional configuration  $\varepsilon = \partial u / \partial x$ , where  $u$  is the displacement.

The corresponding dynamic experiments in solids are based on the relation between the detected strain amplitude at the drive frequency and, typically, the following: (1) harmonic amplitudes, (2) wave cross-

modulation amplitudes, (3) resonance frequency shift, and (4) amplitude dependent losses. It is the observation of these effects that tell us about the nature of the nonlinearity, for instance, whether or not nonclassical behaviour such as hysteresis is present in dynamic processes.

### 1.3. Nonlinear resonance frequency shift

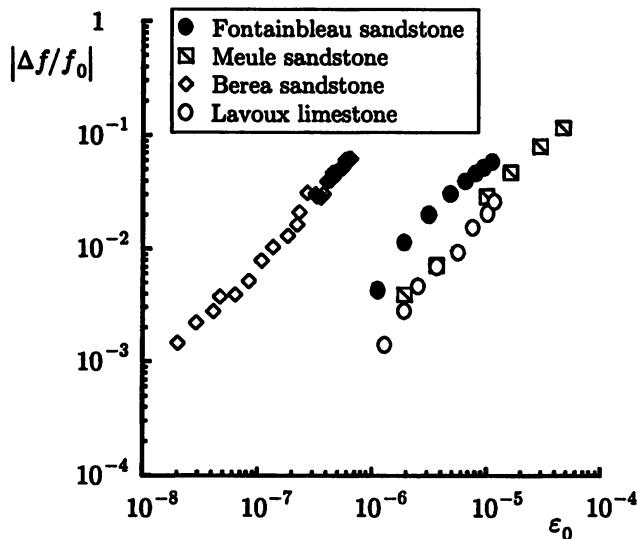


Fig. 3. Normalized frequency shift  $|f - f_0|/f_0$ , where  $f_0$  is the linear resonance peak, versus strain amplitude  $\epsilon_0$  for various rocks under various experimental conditions. A slope of approximately 1 indicates that nonclassical nonlinearity is responsible for the peak shift

representative for nonlinear resonant behaviour in a rock observed in Los Alamos [8]. The material is Fontainebleau sandstone under ambient temperature and pressure conditions. The solid and dotted lines in Fig. 2b indicate that the resonance response is dependent on the direction of the frequency shift (up or down the frequency axis). Clearly, the intensity of the distortion increases significantly with wave amplitude. Some relevant wave forms and signal spectra are shown in Fig. 2c.

Figure 3 illustrates the dependence between the detected frequency shift and the strain for different rocks. In all cases, this dependence is close to linear rather than to quadratic. This is unexpected behaviour for classical nonlinear elasticity, where the dependence is quadratic, and implies that the equation of state contains singularities. More specifically, hysteresis is present. Note also that the nonlinearity is well measurable at strains as small as  $10^{-8}$ – $10^{-7}$ , and that strain is actually equivalent to the acoustic Mach number!

A complementary series of experiments was performed at the Institute of Applied Physics [9]. The measurements were performed in samples of granite and for a cylinder filled with a wet sand; the latter was pre-consolidated by exposing it to intensive sound over 4 hours to obtain stable results. Fig. 4 shows the resonance frequency shift for these materials as a function of strain amplitude  $\epsilon_0$ . As with the data shown in Fig. 3, the shift is linearly proportional to  $\epsilon_0$  for the granite and the sand. Of significance is that these are some of the very few existing measurements in unconsolidated material, the sand, and they show the same dependence with strain as the rocks. However, a different dependence was found in marble.

It is notable that rock samples demonstrate slow dynamical (relaxation) response. Slow dynamics in this context means that the average material modulus is temporarily altered (lowered) during wave excitation.

The dependence of a mode resonance frequency in a sample on the oscillation amplitude is a sensitive measure that can be used for calculation of the average modulus and wave speed in rock.

Nonlinear resonance in a classical system is due to cubic nonlinearity (or its equivalent) and can be attributed to the Duffing-type equation

$$\ddot{u} + \omega_0^2 (u + pu^3) + g\dot{u} = F_0 \cos(\omega t), \quad (1)$$

where  $p$  defines nonlinearity and  $g$ , dissipation. It is well known that the resonance response of such an oscillator to the force amplitude  $F_0$  is such that (for small  $g$ ) the amplitude maximum is shifted from the linear resonance frequency  $\omega_0$  at a value of  $\Delta\omega$  proportional to  $F_0^2$ . This result is rather universal for all systems when the equation of state can be represented as a polynomial.

Let us look now at the experimental results. Figure 2a shows resonances in a "classical" medium (polyvinylchloride) at different amplitudes; it is typical of the linear behavior. Figure 2b is repre-

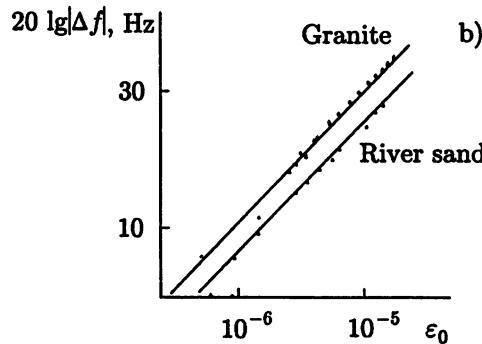
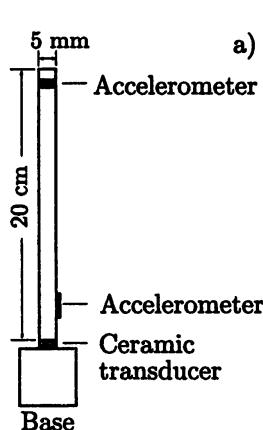


Fig. 4. Resonance experiments in wet sand (grain diameters of about 0.3 mm) and granite: Experimental setup (a). Resonance peak shift versus strain amplitude (b). The excitation frequency was 3.6 kHz (first-mode resonance) and  $\Delta f = f - f_0$

After wave excitation, it takes some time (of the order of  $10^3$  s) for the material modulus to recover to its original state. One manner in which to observe this behaviour is to monitor the resonance frequency before and after large excitation (e.g., [10, 11]). That is, after measuring the linear resonant peak, the sample is driven at a large amplitude for several minutes. The low-amplitude resonance is then monitored until the resonant peak has returned to its original frequency. An example of slow dynamics is illustrated in Fig. 5 for several different rock types and for concrete. It is interesting that the resonance frequency recovery is universally logarithmic in time. All that is more demonstration of the non-classical behaviour of materials.

#### 1.4. Harmonic generation

A large series of resonance experiments dealt with measurement of harmonic amplitudes via Fourier analysis. For instance, a typical result for Berea sandstone taken from [8] is shown in Fig. 6a. The fact that the second and third harmonic amplitude slopes are both proportional to the square of the driving force amplitude also indicates that classical nonlinearity is not sufficient to explain such behaviour. Indeed, a classical stress-strain relation, like Eq. (2) below, always provides a cubic dependence of the third harmonic amplitude on the main one.

The corresponding dependencies obtained in other experiments [9] are shown in Figs. 6b and 6c for granite and wet sand. These results further indicate that the EOS does not correspond to classic nonlinearity and should be singular.

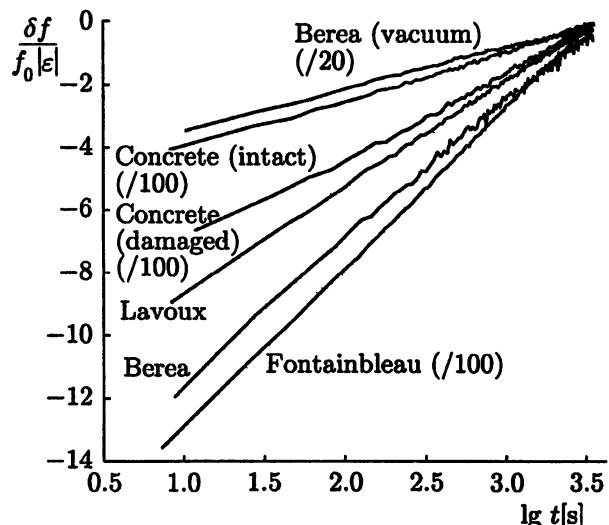


Fig. 5. Slow dynamical response in several rock types and in concrete. The time-dependent shift  $\delta f$  of the recovering resonant frequency, normalized to the asymptotic value  $f_0$ , per unit driving strain  $|\epsilon|$ . Sample names are indicated in the figure, and some data were divided by the indicated factors for plotting convenience. Lavoux is a limestone, Berea and Fontainebleau are sandstones, and one sample of concrete was damaged

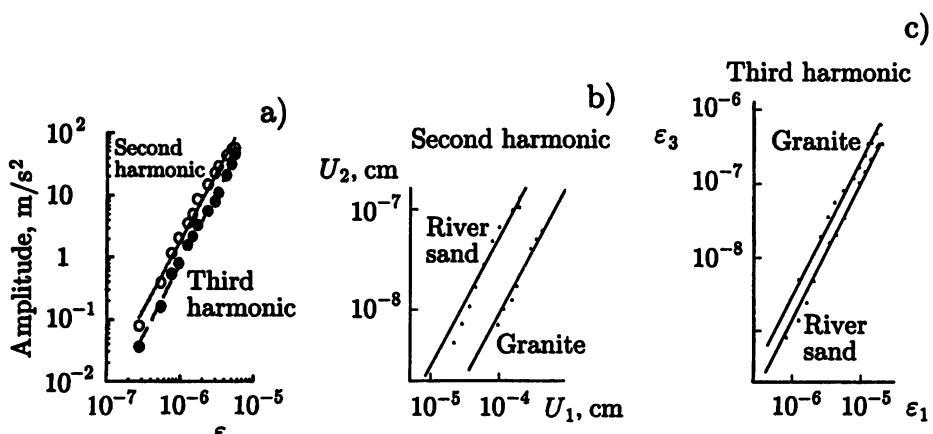


Fig. 6. Second- and third-harmonic amplitudes at resonance peaks as functions of the measured strain amplitude in Berea sandstone (a). Displacement amplitude of the second harmonic  $U_2$  versus the fundamental displacement amplitude  $U_1$  in river sand and granite (b). Amplitude  $\epsilon_3$  of the third harmonic of strain versus fundamental strain amplitude  $\epsilon_1$  in these materials (c). The fits correspond to a power law of 2, indicating nonclassical nonlinearity

### 1.5. Nonlinear dissipation

Amplitude-dependent loss is a well known phenomenon in metals where it is attributed to hysteresis due to dislocations. For rocks, for soils, and in earthquake studies (e.g., [12]) nonlinear dissipation is nearly always observed and is an additional indicator of hysteresis in the EOS. Below, we give two examples.

Amplitude-dependent attenuation in earth materials was observed in experiments on nonlinear interaction between low- and high-frequency signals [13, 14]. During the low-frequency, high amplitude resonant excitation ("pump wave"), a longitudinal ultrasound pulse (frequency 200 kHz, pulse duration 70 ms) was generated to propagate for some distance along the bar, after which its amplitude was measured, and the spatial damping rate was calculated. In the presence of the low-frequency mode, the ultrasound damping rate increased in proportion to the low-frequency strain amplitude  $\epsilon_1$  in granite and sand, and to  $\epsilon_1^2$  in marble (Fig. 7). Similar experiments for control samples (glass) did not reveal any significant nonlinear effects.

Figures 7b and 7c show results from a resonance experiment in Berea sandstone [15] under vacuum conditions at very small acceleration/strain levels. In the resonance data shown in Fig. 7b, a very small frequency shift and peak broadening can be observed. Figure 7c shows the actual change in the sample quality factor extracted from Fig. 7b. It is remarkable that nonlinear attenuation can occur at extremely small strain levels when the nonlinear frequency shift is not well noticeable yet.

### 2. MODELS OF STRUCTURAL NONLINEARITY

An adequate physical model of rocks must be associated with their complex structure. Mechanical properties of rock appear to be a part of a broader class of materials, one we call the "Structural Nonlinear Elasticity" class (also "Mesoscopic/Nanoscale Elasticity"). These terms are in contrast to materials that display classical, "atomic" elasticity, such as most fluids and monocrystalline solids. The nonlinearity of atomic elastic materials is due to the atomic/molecular lattice anharmonicity. The latter is relatively small because the intermolecular forces are extremely strong. In contrast, the materials considered below contain small soft features that we term the "bond system" (cracks, grain contacts, dislocations, etc.) occupying a small total volume (of up to nanoscale sizes) within a hard matrix (grains, crystals) but subject to strong deformation and is the origin of strong nonlinearity, whereas the hard phase is relatively insensitive to deformation (e.g., [16, 17]).

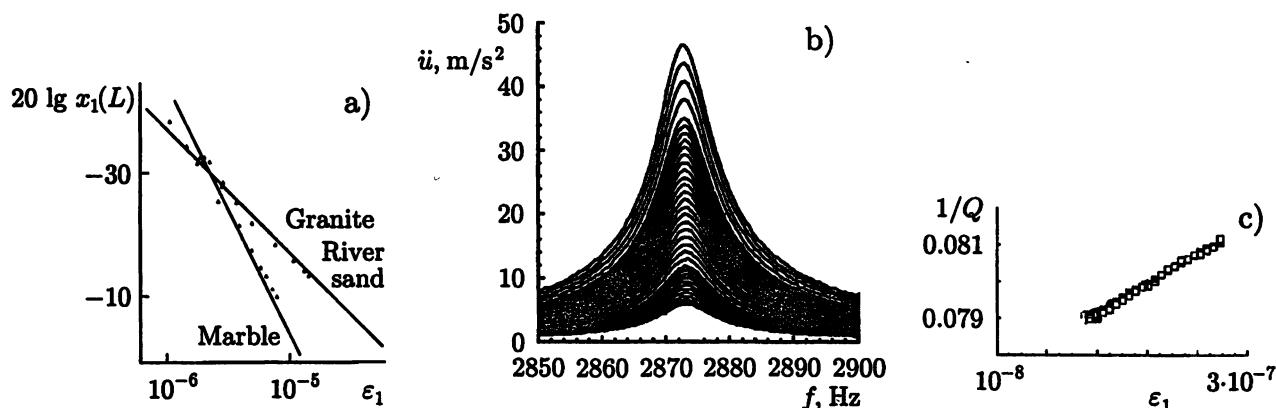


Fig. 7. Dependence of the damping rate of a 200 kHz ultrasound pulse ( $X_1(L) = \ln[A_0/A(L)]$ ), where  $A_0$  is the input amplitude of the pulse and  $A(L)$  is output amplitude in the rod), on the amplitude  $\epsilon_1$  (logarithmic scale) of the low-frequency resonance “pumping” wave for three materials, obtained from the experimental configuration shown in Fig. 4. For sand and granite, the dependence is linear and for the marble, quadratic (a). Resonance acceleration amplitude versus frequency at very small acceleration levels (b). From the data in (b): normalized attenuation rate  $1/Q$  versus strain (logarithmic scale) (c)

An effect of the soft/hard system is well known for liquids with bubbles where the maximum nonlinearity is observed for a gas volume fraction of less than  $10^{-3}$  (e.g., [18]). A similar behaviour is demonstrated by waterlike porous media in which the shear modulus is small compared to the bulk modulus (i.e., shear wave velocity is much smaller than that of longitudinal waves). In such cases, the parameter of nonlinearity  $\beta$  may reach values of  $10^4$ – $10^5$  compared to 1 to 10 for “classical” gases, liquids and solids. However, no hysteresis exists in these systems.

From the above measurements one can calculate a set of fundamental nonlinear parameters of the material and attempt to use them in order to infer the nature of the nonlinear response and to create models.

The classical nonlinear theory for atomic elasticity is thoroughly described in literature (see [19]). It begins with the expansion of the elastic strain energy,  $E$ , in powers of the components of strain tensor,  $\epsilon_{ij}$ . The expansion coefficients designate the components of the second-order elastic tensor and the third-order elastic tensor, respectively. These tensors are characterized, respectively, by 21 and 56 independent components for an arbitrary anisotropic medium (in the lowest-order, triclinic material symmetry) and only by 2 and, respectively, 3 components in the highest-order symmetry (isotropic material). The result is an equation of state relating the stress tensor  $\sigma_{ik}$  to the strain tensor. To gain insight, one can consider the one-dimensional case with only nonzero components  $\sigma_{xx} = \sigma$  and  $u_x = u$  or  $\epsilon_{xx} = \epsilon = \partial u / \partial x$ . Then,

$$\sigma = M(\epsilon + \beta\epsilon^2 + \delta\epsilon^3 + \dots), \quad (2)$$

where  $M$  is the elastic modulus, and  $\beta$  and  $\delta$  are nonlinear coefficients that can be expressed in terms of combinations of the elastic moduli. A typical order of linear moduli for atomic elastic solids is  $10^{11}$ – $10^{12}$  Pa. From here, it easily follows that the amplitude of the second harmonic of a signal,  $A_2$ , is proportional to the square of the amplitude of the applied force,  $A_0$ , and the third harmonic is proportional to  $A_0^2$ .

Nonlinear response in hysteretic materials should contain an additional, singular term in EOS that is sensitive to the history of the process. The latter can be characterized by the sign of  $\dot{\epsilon} = \partial\epsilon/\partial t$ :

$$\sigma = M(\epsilon + \beta\epsilon^2 + \delta\epsilon^3 + \dots) + \hat{A}[\epsilon, \text{sign}(\dot{\epsilon})], \quad (3)$$

where  $\hat{A}$  is a functional describing “nonclassical” effects. A specific form of  $\hat{A}$  should follow from the material physics. In an early work by Asano [20], two basic types of nonclassical (hysteretic in stress-strain)

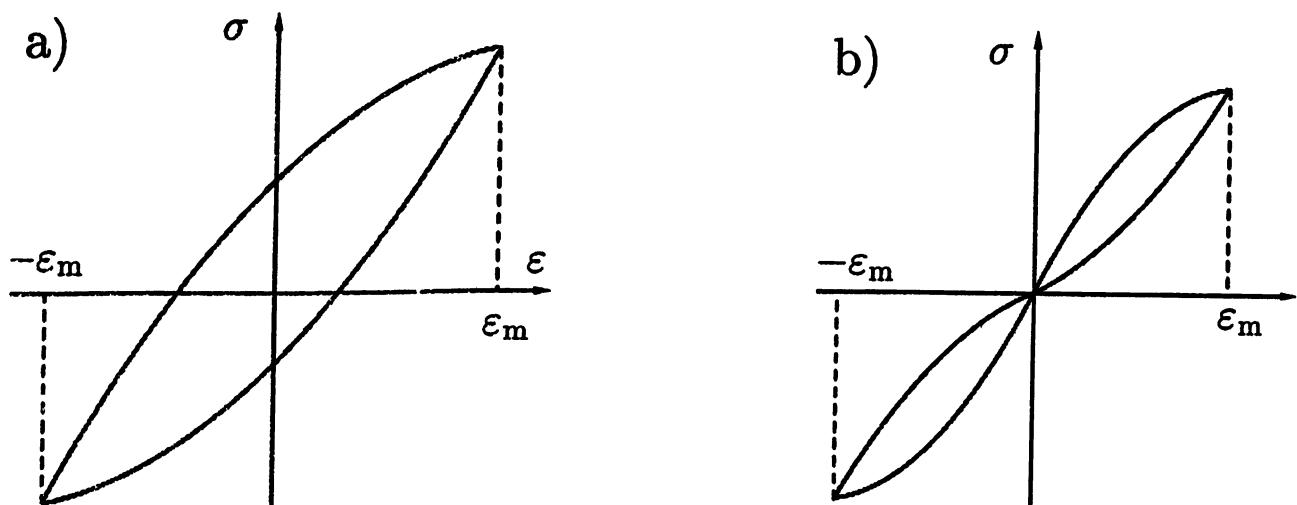


Fig. 8. Simplified forms of two hysteretic models of the EOS: irreversible (type 1) hysteresis (a) and reversible (type 2) hysteresis (b)

behaviour were considered. One of them (type 1) surrounds zero point on the  $(\sigma, \varepsilon)$  plane for a periodic motion (Fig. 8a). The second (type 2) includes zero point and has a "butterfly" shape (Fig. 8b); it is related to the Granato–Lücke model discussed below. Note that Asano associates them with two simple mechanical models, "slider" for type 1 and "ratchet" for type 2 which could help to understand their physical nature. In many cases, these forms were introduced as a best fit for experimental values and dependences. For example, using a specific phenomenological EOS enabled the authors of [14] to consistently describe the results of their experiments.

A semi-phenomenological model called the Priesach–Mayergoyz space (P–M space) model, that successfully describes hysteretic nonlinear behaviour of rock elasticity with discrete memory, was developed in a series of papers by our colleagues (see [3, 21] and references therein). The model assumes that the elastic properties of a macroscopic sample of material results from the integral response of a large number of individual, elastic elements (of order  $10^{12}$ , a rough estimate of the number of grain-to-grain contacts, microcracks, etc., in one cubic centimetre of rock). Each elastic element may or may not demonstrate hysteretic behaviour. The individual elements are combined for analysis in what is known as P–M space (also referred to as Priesach space). To obtain an equation of state, it is assumed that most of the elements are non-hysteretic (with weak nonlinearity). As a result, the stress–strain relation of type 2 hysteresis follows. The P–M space model gives some idea of composition of a hysteretic media (note that dislocations considered below can serve as its elements). However, it still remains a phenomenological description that does not contain the physical mechanisms of nonlinear response. In the following discussion, several relatively simple physical models will be tested here to see if they may provide some insight into the mechanism of nonlinear response.

### 2.1. Hertzian contacts

A starting point model of nonlinearity in rock can be based on representing the rock as a system of dry, contacting grains as shown in Fig. 9. These contacts are much softer than the matrix material, the grains themselves, and therefore play the primary role in the nonlinear elastic response of the medium. In this model, the distance change  $\Delta$  between the grain centres is related to the compressing force,  $F$ , by the Hertzian contact law (e.g., [19])

$$\Delta = \left[ \frac{3(1-\nu^2)F}{4ER^{1/2}} \right]^{2/3}, \quad (4)$$

where  $E$  is the Young modulus of the material,  $\nu$  is the Poisson ratio, and  $R$  is the grain radius.

For a dry medium composed of spheres, this readily yields the following one-dimensional EOS relation [16] in which the effective stress,  $\sigma_{\text{eff}}$ , is proportional to  $\varepsilon^{3/2}$ . As a result, the contact contribution to the sound speed,  $c(\varepsilon) = (\rho^{-1} d\sigma_{\text{eff}}/d\varepsilon)^{1/2} \propto \sqrt{\varepsilon}$ , where  $\rho$  is density, tends to zero at small positive strains. (Negative strain means that grains separate, and there are no contact forces present.) However,  $dc/d\varepsilon$ , which is a measure of nonlinearity, tends toward infinity! In real experiments, the aggregate is subject to a static pressure creating a constant pre-strain,  $\varepsilon_0$ , and for small one-dimensional perturbations, we can expand  $\sigma$  into the series (2), where the modulus  $M$  is proportional to  $\varepsilon_0^{1/2}$ , and the quadratic and cubic nonlinearity coefficients are

$$\beta = 1/(2\varepsilon_0), \quad \delta = 1/(6\varepsilon_0^2). \quad (5)$$

In rock, the role of pre-strain can be played by a hard, consolidated fraction of contacts and/or by the pressure from upper layers of earth.

Some interesting properties of granular materials follow from these simple results. For instance, the nonlinearity parameters do not depend on grain size or on their composition, but on pre-strain  $\varepsilon_0$ , i.e., on static pressure. These results were confirmed in experiments with lead shot and tuff excited at frequencies of a few kilohertz by a vibrating table [22]. The values of  $\varepsilon_0$  were controlled by loading masses. From measurements of second and third harmonics, the formulae (5) both gave a good agreement with the experiment, whereas the parameter  $\beta$  exceeded  $10^3$  and  $\delta$  exceeded the values of  $5 \cdot 10^6$ .

Hence, this model predicts a very strong nonlinearity as compared with the solid matrix, and it admits many effective extensions. One of them deals with small-scale multicontact interfaces between larger grains, yielding a significant local amplification of nonlinearity from local stress concentrations. Estimates [17] show that for larger grains (of radius  $R$ ) in contact with smaller hemispheres of radius  $r$ , the nonlinear portion of the EOS and, correspondingly, nonlinearity parameters (5) acquire an additional term of  $s(R/r)^{3/2}$ , where  $s$  is the relative contact area occupied by small-sphere contacts (the remaining area is assumed to be cemented). This approach can be extended to more complicated fractal structures.

An interesting physical problem is associated with the effect of intergrain fluid. A 100% saturation decreases nonlinearity due to decreasing linear compressibility. However, a small amount of liquid can increase nonlinearity thanks to the effect of capillary forces or, for even thinner contacts of grains, to the Van der Waals force. The increase of nonlinearity at small and moderate fluid saturation has been confirmed experimentally [23]. This area is, however, beyond the scope of this paper.

The above effects are sufficient to provide extremely strong nonlinearity but still fail to include hysteresis. For the latter, transverse deformation can be important. In a series of papers by D. Johnson and colleagues (see [24] and references therein), a detailed analysis of nonlinear properties of granular media under the influence of static pressure was investigated. These authors took into account both the Hertz theory and the Mindlin relation (and its variations) stating that upon normal compression, a tangential displacement,  $\tau$ , of contacting may arise that, in general, creates an additional transverse force:

$$\delta T = \frac{4\mu a(\Delta)\tau}{1 - \nu}. \quad (6)$$

Here again,  $\Delta$  is the relative displacement of spheres,  $\mu$  is shear modulus of the material, and  $a$  is a characteristic length depending on the nature of surface contacts. For a "pure" Hertzian contact with reversible

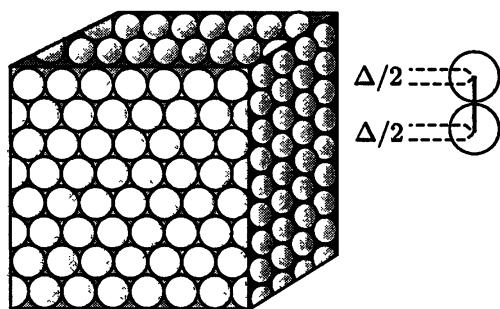


Fig. 9. An aggregate of contacting grains

slip,  $a = 0$ , but for a rough, nonslip contact, it is equal to  $\sqrt{R\Delta}$  or, in the case of pre-compression with an initial contact radius  $b$ ,  $a = [(R^2\Delta^2 + b^4/4)^{1/2} + b^2/2]^{1/2}$ . This results in new features such as the dependence of forces and energy on the path of deformation. Indeed, in general, transposing of normal and transverse displacements changes the work of external force. This is actually a hysteretic phenomenon that causes, for example, attenuation of an elliptically polarized acoustic wave. However, a role of this mechanism in rock hysteresis is still unclear, and below, we shall incorporate the description suggested several decades earlier for hysteretic behaviour of metals.

## 2.2. Granato–Lücke model

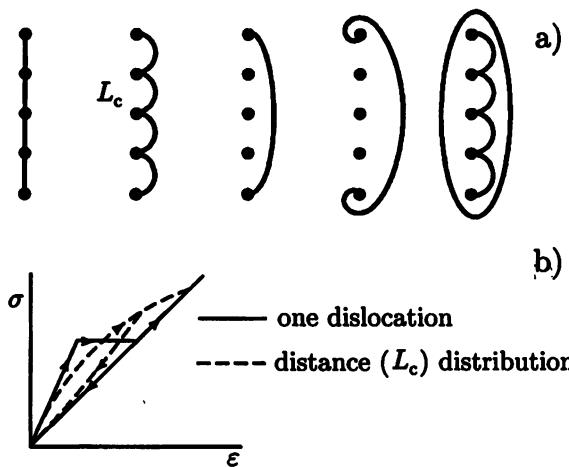


Fig. 10. Granato–Lücke model of dislocations (a). Resulting stress–strain curve for a single dislocation (solid line) and a distribution of dislocations (dashed line) (b)

state may be the same before and after inducing the dislocations to react, so that we have a type I hysteresis. The model also incorporates slow dynamics because the equilibrium state takes some time to restore. In reality, the distances between the sticking points are statistically distributed, which smooths the hysteretic loop (Fig. 10b, dashed line). Hybrids of the G-L model include other aspects including frequency dependence. In spite of some disadvantages, this model was truly a pioneering micromodel for hysteretic dynamic behaviour of structurally inhomogeneous materials.

## 3. NONLINEAR OSCILLATIONS AND WAVES IN ROCKS

In order to explain experimental data, it is necessary, along with construction of material models, to understand peculiarities of waves in these materials. From the above considerations and experimental data, it can be expected that these processes are more complex than those in nonlinear acoustics describing most fluids and intact solids.

### 3.1. Classical theory

In the classical theory of elasticity, the equation of motion in Lagrangian coordinates is

$$\rho \ddot{u}_i = \frac{\partial \sigma_{ij}}{\partial x_j}, \quad (7)$$

where  $u_i$  are the components of the displacement vector,  $\mathbf{u}$ ; and  $\rho$ ,  $\sigma_{ij}$ , and  $\ddot{\mathbf{u}}$  designate the density, the stress tensor, and the particle acceleration, respectively.

For a longitudinal wave ( $P$ -wave) propagating in an isotropic medium, the resulting equation of motion can be written in the one-dimensional form

$$\rho_0 \frac{\partial^2 \varepsilon}{\partial t^2} = \frac{\partial^2 \sigma(\varepsilon)}{\partial x^2}. \quad (8)$$

From the energy expansion, the EOS can be written as Eq. (2). Correspondingly, the local sound velocity is

$$c = \sqrt{\rho^{-1} d\sigma/d\varepsilon} \approx c_0(1 + 2\beta\varepsilon + \delta\varepsilon^2 + \dots), \quad (9)$$

with  $c_0 = \sqrt{M/\rho}$ .

Note that even if the nonlinearity is anomalously large in rocks in comparison with that of atomic elastic media, the nonlinear terms in the EOS are generally much smaller than the linear term. This is because  $\varepsilon$  is order  $10^{-9}\text{--}10^{-5}$  in dynamic experiments. Therefore, one can use the relation (8) and its generalizations for media with strong nonlinearity.

This nonlinearity can be caused by two mechanisms. The first is of a “geometrical” or “kinematic” type, associated with the difference between the Lagrangian and Eulerian descriptions of motion (such as  $(\mathbf{u}, \nabla \mathbf{u})$  in the Eulerian equation of motion). The other type is “physical” elastic nonlinearity. Physical nonlinearity is described by third-order (and higher) terms in the expansion of the elastic energy in  $\varepsilon$ , and accounts for the fact that stress is not a linear function of strain. The effects of geometrical nonlinearity are typically comparable in order to physical nonlinearity in atomic elastic materials, such as fluids and intact solids. In rock and other highly nonlinear media, physical nonlinear response is typically orders of magnitude larger than geometric nonlinear response, and therefore, the latter (and hence the difference between Eulerian and Lagrangian descriptions) can be ignored, which somewhat simplifies the mathematics.

### 3.2. Traveling waves and hysteresis

For hysteretic media, we begin by considering traveling (progressive) waves in an unbounded material. For a one-dimensional traveling wave, the strain  $\varepsilon \approx -v/c$ , where  $v = u_t$  is the particle velocity, and  $c$  is sound velocity. In this case,  $\varepsilon$  plays a role of the acoustic Mach number. When the Taylor expansion (9) for  $c$  is valid, a well-known solution in the form of a simple (Riemann) wave follows from Eq. (8):

$$\varepsilon = \Psi [x - c(\varepsilon)t], \quad (10)$$

where  $\Psi$  is an arbitrary function defined by the initial condition, and  $c(\varepsilon)$  is the local wave speed. Propagation of such a wave in acoustics is known to result in the appearance of shocks and then the formation of a sawtooth wave (e.g., [18]), which dissipates asymptotically as  $t^{-1}$  at a sinusoidal initial condition.

Looking back at the experiments, we see nonclassical, hysteretic behaviour of rock and turn to wave solutions that can describe this behavior. The evolution of a nonlinear wave in hysteretic media described by Eq. (3) is considerably different from that of a wave in classical nonlinear media. Several problems of this kind have already been addressed in publications (e.g., [26–29]). Here we give only a few simple illustrations. Available phenomenological models of hysteretic stress-strain dependence can include rather many parameters, and the choice should be based on experimental results. We mention here both basic hysteretic models outlined above, of types 1 and 2 (Figs. 8a and 8b), for which the wave distortion processes are significantly different from each other and from the classical case. Here, we restrict ourselves by symmetric hysteretic loops. Thus, for type 1

$$\sigma = M \begin{cases} (\varepsilon - \alpha\varepsilon_m)\varepsilon + (\gamma/2)(\varepsilon_m^2 - \varepsilon^2), & \dot{\varepsilon} > 0, \\ (\varepsilon - \alpha\varepsilon_m)\varepsilon - (\gamma/2)(\varepsilon_m^2 - \varepsilon^2), & \dot{\varepsilon} < 0, \end{cases} \quad (11)$$

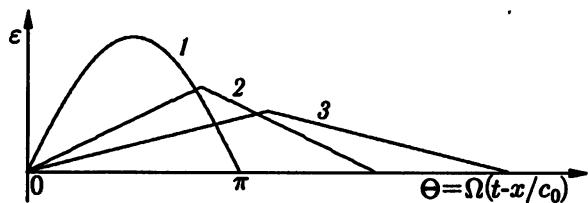
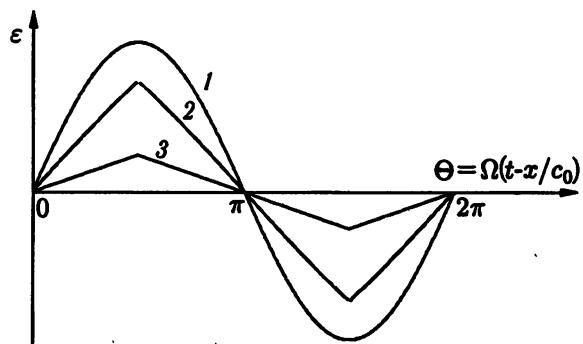
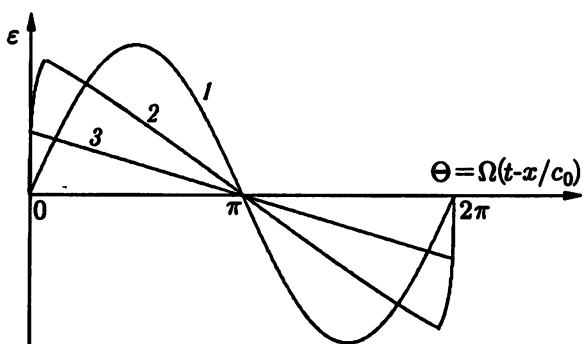


Fig. 11. Qualitative evolution of the wave profile as a function of the “traveling phase”  $\Theta = \Omega(t - x/c_0)$  in a nonhysteretic medium with a quadratic nonlinearity (a), in a type 1 hysteretic model (b), and in a type 2 hysteretic model (c). Curves 1, 2, and 3 correspond to three successive points along the propagation path. For (a) and (b) the input wave is sinusoidal with frequency  $\Omega$ , for (c) it is a unipolar pulse in the form of a half-period of a sinusoid with the frequency  $\Omega$ .

where  $E$  is the linear elasticity modulus,  $\varepsilon_m$  is maximal strain corresponding to the singular points of EOS, and  $\alpha$  and  $\gamma$  are the nonlinearity parameters; for weak nonlinearity,  $\alpha\varepsilon_m$  and  $\gamma\varepsilon_m$  are small. This relationship corresponds to an irreversible, hysteresis in the sense that  $\varepsilon \neq 0$  for  $\sigma = 0$  and vice versa. If the process starts from zero, it first goes along some path up to the singular point  $\varepsilon = \varepsilon_m$ , and then continues periodically along the loop (11).

The type 2 model is described by the equation

$$\sigma(\varepsilon, \dot{\varepsilon}) = M \begin{cases} \varepsilon - \gamma\varepsilon^2/2, & \text{if } \varepsilon > 0 \text{ and } \dot{\varepsilon} > 0; \\ \varepsilon + \gamma\varepsilon^2/2 - \gamma\varepsilon_m\varepsilon, & \text{if } \varepsilon > 0 \text{ and } \dot{\varepsilon} < 0; \\ \varepsilon + \gamma\varepsilon^2/2, & \text{if } \varepsilon < 0 \text{ and } \dot{\varepsilon} < 0; \\ \varepsilon - \gamma\varepsilon^2/2 + \gamma\varepsilon_m\varepsilon, & \text{if } \varepsilon < 0 \text{ and } \dot{\varepsilon} > 0. \end{cases} \quad (12)$$

In this case, the loops beginning from zero are reversible.

Let us now describe traveling waves within the framework of these two EOS. As long as nonlinearity is small, each piece of a wave corresponding to a given branch of hysteretic EOS propagates as a simple wave (10) without reflections. They should be matched at  $\varepsilon = \varepsilon_m$ . Some analytical expressions for the wave (10) can be found, for example, in [29]. They are rather cumbersome due to the dependence of wave amplitude  $\varepsilon_m$  on time. Qualitative pictures of nonlinear wave evolution for different cases are shown in Fig. 11. Figure 11a demonstrates a classical case: the evolution of an initially sinusoidal wave in a non-hysteretic medium with quadratic nonlinearity. In this case, a sawtooth wave is eventually formed. Due to hysteresis, an additional kind of singularity arises. In general, due to singularity at the wave maximum, the portions on either side of the wave peak move with different velocities, thus “consuming” each other and resulting in the formation of a cusp. Figure 11b shows such a process for the type 1 EOS and Fig. 11c, for the type 2 EOS. In the latter case, a unipolar pulse is shown; otherwise, additional singularities in zero point may occur.

The corresponding asymptotic behaviour of wave amplitude at large distances is also different. For example, for the type 1 hysteresis (11), the wave amplitude  $\varepsilon_m$  changes as  $1/t$ , as in a classical periodic sawtooth wave. Needless to say, the wave profiles are radically different from their classical “prototypes.” Correspondingly, whereas all harmonics are represented in the classical sawtooth wave, only odd harmonics exist in the symmetric hysteretic case.

In short, the presence of hysteresis qualitatively changes the wave profile evolution in comparison with the classical case.

### 3.3. Resonance oscillations

The theory for nonlinear standing waves is more complicated due to nonlinear interactions of oppositely propagating waves. In the general case, one should solve a nonlinear wave equation (8) with an external forcing and boundary conditions at the end of the bar. In terms of displacement  $u$ , this equation is

$$\rho \left( \ddot{u} + \frac{\omega}{Q} \dot{u} \right) = \frac{\partial \sigma}{\partial x} + \frac{F_0(x)}{2} \left[ e^{i(\omega+\Delta)t} + \text{c.c.} \right], \quad (13)$$

where  $F_0(x)$  is the external force amplitude,  $Q$  is the linear quality factor,  $\Delta$  is the frequency shift (detuning) from a linear resonance mode, and c.c. denotes the complex-conjugated term.

For a small nonlinearity, the structure of resonance modes is close to that of the linear system, although their amplitudes are strongly affected by nonlinear interaction. Thus, it is adequate to use an expansion of the solution for  $u$  into the modal series:

$$u = \sum_n A_n(t) \Phi_n(x), \quad (14)$$

where  $\Phi$  is the eigenfunction of the linear wave to be found with the use of corresponding boundary conditions at the bar ends (for a bar of length  $L$  with free ends,  $\Phi_n = \cos(k_n x)$ , where  $k = n\pi/L$  and  $n$  is an integer mode number). As a result, we obtain a system of ordinary differential equations for  $A_n$  that can usually be reduced to a system of a few interacting resonant modes (e.g., [30]).

Substituting a solution in the form (14) for the  $n$ th mode, e.g.,  $u = U \cos(k_n x) \exp[-i(\omega_n + \Delta)t/2]$ , into Eq. (13), multiplying by  $\cos(k_n x)$  and integrating over the length  $L$  of the bar, we obtain the following equation for the amplitude  $U$  in equilibrium:

$$-\rho (\omega_n U \Delta + i \omega_n^2 / (2Q)) = k_n \langle \sigma_1^N \rangle + \langle F_{0n} \rangle. \quad (15)$$

The stress is  $\sigma = E(\epsilon) + \sigma^N(\epsilon)$  with  $N$  denoting the small nonlinear portion,  $\sigma_1^N$  is the amplitude of the first temporal harmonic of nonlinear part of the stress, and  $E$  is the linear modulus, so that  $c_0 = \omega_n/k_n = (E/\rho)^{1/2}$ . Here  $Q$  is the linear quality factor, and  $\langle \sigma_1^N \rangle = L^{-1} \int_0^L \sigma_1^N(x) \cos(k_n x) dx$  (the same for  $F$ ) denotes spatial averaging. Hence, we have an equation for the amplitude-frequency resonance dependence. In terms of strain  $\epsilon = \partial u / \partial x$ , the result is

$$\epsilon_m = \frac{\epsilon_0}{L \sqrt{[(\Delta + \Delta^N)/c_0]^2 + [k_n/(2Q_N)]^2}}, \quad (16)$$

where  $Q_N < Q$  is the quality factor taking into account both linear and nonlinear losses and  $\Delta^N$  is the nonlinear frequency shift. Nonlinear losses and frequency shift depend on the imaginary and real parts of  $\langle \sigma_1^N \rangle$ . Here, it is supposed that the excitation occurs in a fixed point  $x = 0$ , i.e.,  $u(0, t) = u_0 \sin(\omega t)$ , and  $\epsilon_0 = k_1 u_0$ .

The result clearly depends on the EOS of the medium. In a classical “cubic” medium, where  $\sigma^N \propto \epsilon^3$ , this relation defines a well-known nonlinear resonance curve corresponding to the known Duffing oscillator, with the nonlinear frequency shift proportional to  $\epsilon_0^2$ . In hysteretic media considered above, the wave profile is distorted according to the above consideration of traveling waves. Note that, unlike in classical media, the Taylor expansion of  $\langle \sigma_1^N \rangle$  contains not only odd, but also even powers of  $\epsilon$ . As a result the third harmonic in hysteretic media can be proportional to  $\epsilon^2$  as observed.

For strong wave distortion, a relatively simple way to treat the problem is to consider a ring resonator [17] (similar to that used in experiments described in [30]) where a resonance mode can be a wave traveling

around the ring and undergoing nonlinear distortion until it is balanced with a harmonic source applied at some point. Hysteretic media described by the two aforementioned types of phenomenological stress-strain loops yield different results.

Let us denote by  $\alpha(\varepsilon_1)$  the relative change of the amplitude of the fundamental harmonic  $\varepsilon_1$  of the wave upon propagation at the distance of the bar length, and suppose that  $g \ll 1$ . The wave excited at  $x = 0$  has the form  $u(0, t) = u_0 \sin(\omega t)$ , so that the first harmonic arriving at  $x = 0$  after passing the ring length  $L$  will have the delay of  $L/c$  and decay by a factor of  $g$ , which in the steady-state regime is to be compensated for by the source. Thus, the balance equation for  $\varepsilon_1$  takes the form

$$\varepsilon_1 - \varepsilon_1 [1 - \alpha(\varepsilon_1)] e^{-ik_1 L} = k_1 u_0. \quad (17)$$

The solution of this equation can be represented as Eq. (16) plus the expression for  $\langle \sigma_1^N(\varepsilon_m) \rangle$  depending on the EOS model. For illustration, we suppose that the excitation amplitude is large enough and, consequently, a steady-state travelling nonlinear wave excited in the ring has an asymptotic form with cusps corresponding to Figs. 11a and 11b. Briefly, the results are the following.

In both cases, the nonlinear  $Q$ -factor for the  $n$ th mode is given by

$$Q_N^{-1} = \frac{\omega_n}{2\pi c_0} g \varepsilon_1, \quad (18)$$

where  $g = 2\pi^2\gamma$  for the type 1 model (11) and  $g = 4\pi^2\gamma$  for the type 2 model (12).

Nonlinear resonance frequency shift is

$$\delta\omega_n = -\frac{\pi\omega_n\varepsilon_1\beta}{2c_0}, \quad (19)$$

with  $\beta = \gamma$  for type 1 hysteresis and  $\beta = \alpha$  for type 2 one. Hence, both the losses and the frequency shift are proportional to the amplitude of excited oscillations, and the resonance curves look as in Fig. 2b. However, the relation between  $\varepsilon_1$  and the input strain  $\varepsilon_0$  in resonance depends itself on the nonlinear losses and, as a result, on the EOS.

Similarly, a higher  $m$ th harmonic of oscillations at a given mode can be considered by using the perturbation method (small harmonic amplitudes) and the corresponding stress harmonic amplitude,  $\langle \sigma_m^N \rangle$ . For hysteretic models with singularities, we obtain different dependencies which are often observed in experiments with rocks.

## CONCLUSIONS

Oscillations and waves in hysteretic media may appear an arcane subject. At the same time, as illustrated above, in dynamics of rocks, some ceramics, some metals, and damaged materials, hysteresis is rather a rule than an exception. The cause of that is seen from the experimental results outlined above: the micro-to-nanoscale "bond system" (microcracks, dislocations, etc.) responsible for strong hysteretic nonlinearity. Such "violations" are so ubiquitous that a regular crystalline structure can be considered almost as an exception for macroscopic solids found in nature. This is an extremely rich area of material physics which can be considered as a bridge between the macroscopic mechanical properties, mesoscopic granular structure, and nanoscale features that appear responsible for the hysteretic behaviour. A joint mechanical, thermodynamical, and possibly even quantum-physics approach may be necessary for understanding the physics of nonlinearity in rocks. In addition, practical significance of nonlinear methods of non-destructive evaluation of materials (e.g., [31]) has been already demonstrated. At the same time, it opens an interesting chapter in nonlinear wave theory which deserves to be more actively addressed.

## REFERENCES

1. Andronov A., Vitt A., Khaikin S. Theory of Oscillations. New York: Dover, 1987.
2. Gaponov A., Ostrovsky L., Freidman G. // Radiophys. Quantum Electron. 1967.
3. Guyer R., Johnson P. // Physics Today. 1999. V. 52. P. 30.
4. Ostrovsky L., Johnson P. // Rivista del Nuovo Chimento. 2001. (in press).
5. Hilbert L., Hwong T., Cook N., Nihei K., Myer L. // Rock Mechanics Models and Measurements Challenges from Industry / Ed. P. P. Nelson, S. E. Laubach, A. A. Balkema. Rotterdam, 1994. P. 497.
6. Jaeger J., Cook N. // Fundamentals of Rock Mechanics, 3rd edition. London: Chapman and Hall, 1979. P. 78.
7. Guyer R., McCall K., Boitnott G., Hilbert L., Plona T. // J. Geophys. Res. 1997. V. 102. P. 5281.
8. Johnson P., Zinszner B., Rasolofosaon P. // J. Geophys. Res. B. 1996. V. 101, No. 5.
9. Nazarov V., Ostrovsky L., Soustova I., Sutin A., // Phys. Earth Planet. Interiors. 1988. V. 50. P. 65.
10. Tencate J., Shankland T. // Geophys. Res. Lett. 1996. V. 23. P. 3019.
11. Tencate J., Smith J., Guyer R. // Phys. Rev. Lett. 2000. V. 85. P. 1020.
12. Ishihara K. Soil Behaviour in Earthquake Geotechnics. Oxford: Clarendon Press, 1996.
13. Nazarov V., Ostrovsky L., Soustova I., Sutin A. // Acoust. Phys. 1994. V. 30, No. 5. P. 827.
14. Zimenkov S. V., Nazarov V. // Izv. Fiz. Zemli. 1994. V. 29. P. 12.
15. Guyer R., Johnson P., Tencate J. // Phys. Rev. Lett. 1999. V. 82. P. 3280.
16. Belyaeva I., Ostrovsky L., Zaitsev V. // Nonlinear Processes in Geophysics. 1997. V. 4. P. 1.
17. Ostrovsky L., Johnson P., Shankland T. // Nonlinear Acoustics at the Turn of the Millennium, ISNA 15, Göttingen, 1999 / Ed. W. Lauterborn, T. Kurz. Melville, N.Y.: Amer. Inst. Phys, 2000. P. 75.
18. Naugolnykh K. A., Ostrovsky L. A. // Nonlinear Wave Processes in Acoustics. Cambridge—New York: Cambridge Univ. Press, 1998.
19. Landau L., Lifshits E. // Theory of Elasticity, 3rd edition. Oxford: Pergamon Press, 1986.
20. Asano S. // J. Phys. Soc. Japan. 1970. V. 29. P. 952.
21. Guyer R., McCall K., Boitnott G., Hilbert L., Plona T. // J. Geophys. Res. 1997. V. 102. P. 5281.
22. Belyaeva I., Ostrovsky L., Timanin E. // Acoust. Lett. 1992. V. 15. P. 221.
23. Van den Abeele K., Carmeliet J., Johnson P., Zinszner B. // J. Geophys. Res. 2000. (submitted).
24. Johnson D., Makse H., Gland N., Schwartz L. // Physica B. 2000. V. 279. P. 134.
25. Granato A., Lücke K. // J. Appl. Phys. 1956. V. 27. P. 583.
26. Nazarov V. // Acoust. Phys. 1997. V. 43. P. 192.
27. Van den Abeele K., Johnson P., Guyer R., McCall K. // J. Acoust. Soc. Am. 1997. V. 101. P. 1885.
28. Gusev V., Glorieux C., Lauriks W., Thoen J. // Phys. Lett. A. 1997. V. 232. P. 77.
29. Gusev V. // J. Acoust. Soc. Am. 2000. V. 107. P. 3047.
30. Ostrovsky L., Soustova I., Sutin A. // Acustica. 1978. V. 5. P. 298.
31. Van den Abeele K., Johnson P., Sutin A. // Research on Nondestructive Evaluation. 2000. V. 12/1. P. 17.

<sup>1</sup> Zel Technology / NOAA Environmental Technology Laboratory, Boulder, Colorado, USA;

Поступила в редакцию  
22 марта 2001 г.

<sup>2</sup> Институт прикладной физики РАН,  
Нижний Новгород, Россия;

<sup>3</sup> Los Alamos National Laboratory,  
Los Alamos, New Mexico, USA

УДК 530.16

## ВЗАИМОДЕЙСТВИЕ СОЛИТОНОВ КАК СОСТАВНЫХ СТРУКТУР В МОДЕЛИ ГАРДНЕРА

*К. А. Горшков, И. А. Соустова*

Анализ модернизированного приближённого описания взаимодействия уединённых волн как частиц позволяет найти новое представление известных точных  $N$ -солитонных решений, имеющих вид суперпозиции солитонов с относительно медленно меняющимися параметрами; приближённые решения имеют одинаковую с точными структуру и отличаются лишь точностью представления пространственно-временной зависимости параметров уединённых волн. Приближённое описание эффективно обнаруживает детали столкновения, существенно дополняющие общую картину взаимодействия солитонов в конкретных моделях. Эти положения обсуждаются на примере взаимодействия солитонов в модели Гарднера. Специфика взаимодействия в этом случае обусловлена составной структурой солитонов, позволяющей рассматривать их как суперпозицию кинков (перепадов поля) разной полярности. Построена наглядная и строгая картина взаимодействия солитонов, позволяющая найти не отмеченные ранее решения, в частности, в виде кинков огибающей, осуществляющих при распространении вдоль бесконечных последовательностей (цепочек) солитонов перестройку структуры исходного волнового поля.

### ВВЕДЕНИЕ

Взаимодействие солитонов — задача, традиционно решаемая в рамках различного рода точных подходов (метода обратной задачи рассеяния, преобразования Бэклунда, метода Хироты и др.) [1, 2]. Возможность применения какого-либо из указанных подходов к конкретной модели, в принципе решает задачу о столкновении солитонов. В типичных ситуациях после взаимодействия остаются те же солитоны, что были до взаимодействия, а единственным следствием столкновения является появление сдвигов траекторий центров солитонов. Извлечение более детальных сведений о процессе взаимодействия затруднено из-за громоздкости и многопараметричности формул для  $N$ -солитонных решений (особенно при больших  $N$ ). Вместе с тем именно детали взаимодействия даже для известных интегрируемых моделей оказываются иногда весьма неожиданными и существенно дополняют общую картину столкновения. Так, в рамках модели, описываемой уравнением Кадомцева—Петвиашвили (К-П), были обнаружены стационарные связанные состояния и аномально медленные процессы рассеяния двумерных солитонов с бесконечными фазовыми сдвигами [3], а для системы, описываемой нелинейным уравнением Шредингера (НУШ), установлено существование режимов распространения последовательностей из  $N$  солитонов с квазиэквидистантным расположением их друг относительно друга [4]. Заметим, что наличие подобных режимов имеет принципиальное значение для длинных оптоволоконных линий связи.

Оказывается, что особенности взаимодействия солитонов удобнее выявить и исследовать в рамках приближённого описания столкновения уединённых волн [5], а не путём непосредственного анализа точных  $N$ -солитонных решений.

Как показано в настоящей работе, приближённое описание может быть улучшено таким образом, чтобы обнаруживать и весьма общие свойства взаимодействия солитонов в интегрируемых системах, которые не очевидны в рамках точных подходов. Модернизация описания, с одной стороны, связана с привлечением метода сращиваемых асимптотических разложений, вскрывающего составной (аддитивный по полям солитонов и поправок к ним) характер общего решения. С другой стороны, замена

сосредоточенных характеристик солитонов, таких как координаты их центров  $S_i(t)$ , на распределённые фазовые переменные  $S_i(x, t)$  позволяет, сохраняя представление о солитонах как о целостных образований, в значительной мере учесть их волновую природу (конечность скорости распространения возмущений и связанные с этим деформации формы солитона, различного рода задержки и т. п.). Решения, получаемые таким путём, демонстрируют уникальное свойство: их структура с учётом каждого последующего порядка теории возмущений остаётся соответствующей суперпозиции квазистационарных уединённых волн, а каждое приближение лишь уточняет пространственно-временную зависимость  $S_i(x, t)$ . Это обстоятельство служит ключевым моментом при нахождении нового представления известных  $N$ -солитонных решений в виде строгой суперпозиции солитонов с относительно медленно меняющимися параметрами. Ряды теории возмущений в приближённом описании являются, по существу, такими разложениями точных  $N$ -солитонных решений, что главные члены этих разложений превосходят оставшиеся суммы при любых соотношениях параметров сталкивающихся уединённых волн. В результате решения задачи о взаимодействии солитонов, полученные в первом приближении, оказываются определяющими. Важно, что эти решения могут быть найдены из той же системы уравнений движения для классических частиц, что и в немодифицированном варианте, но теперь эти решения определяют не только временную, но и пространственную эволюцию параметров солитонов  $S_i(x, t) = S_i(x - vt)$ , где  $v$  — «групповая» скорость солитонов, зависящая от конкретной модели. Таким образом, новое представление  $N$ -солитонных решений позволяет трактовать взаимодействие солитонов как столкновение частиц, вообще говоря, протяжённых и деформируемых.

Эти общие положения можно проследить на примере любой точно решаемой модели; в данной работе они обсуждаются в рамках расширенного уравнения Кортевега-де-Вриза (КдВ)

$$\frac{\partial \Phi}{\partial t} + 6\Phi(1 - \Phi)\frac{\partial \Phi}{\partial x} + \frac{\partial^3 \Phi}{\partial x^3} = 0, \quad (1)$$

называемого также уравнением Гарднера. Это уравнение простой заменой переменных  $\Phi = 1/2 + \psi$  сводится к более известному модифицированному уравнению Кортевега-де-Вриза (МКдВ).

## 1. ПОСТРОЕНИЕ ПРИБЛИЖЁННОГО РЕШЕНИЯ

Солитонное решение уравнения (1) может быть представлено в виде

$$\Phi_s(x, t) = \frac{k}{2} \left\{ \operatorname{th} \left[ \frac{k}{2} (x - k^2 t + \Delta) \right] - \operatorname{th} \left[ \frac{k}{2} (x - k^2 t - \Delta) \right] \right\}, \quad (2)$$

где  $\Delta = k^{-1} \operatorname{Arth} k$ ,  $0 < k < 1$ . Уединённые волны (2) имеют ограниченный диапазон изменения амплитуд  $A = \max \Phi_s = 1 - \sqrt{1 - k^2}$  и скоростей  $k^2$ :  $0 < \{A, k^2\} < 1$ . В пределе  $k \rightarrow 0$  решение (2) представляет собой солитон уравнения КдВ, при этом  $\Delta \rightarrow 1$ . Принципиальной для задачи о взаимодействии оказывается возможность интерпретации солитона (2) как составной структуры, сформированной из более простых, по сравнению с уединёнными волнами, образований — кинков. Такие решения, описывающие распространение стационарных перепадов поля, существуют в рамках уравнения (1) лишь при единственном значении скорости  $k^2 = 1$ :

$$\Phi_{\pm}(x, t) = \frac{1}{2} \left[ 1 \pm \operatorname{th} \left( \frac{x - t}{2} \right) \right]. \quad (3)$$

Сравнивая (3) и (2) вблизи  $k^2 = 1$ , когда солитон (2) имеет вид плато большой протяжённости ( $\Delta \gg \gg 1$ ), ограниченного относительно резкими перепадами поля, решение (2) можно приближённо записать следующим образом:

$$\Phi_s \simeq \Phi_{+k} + \Phi_{-k} - 1 = \frac{1}{2} \left[ \operatorname{th} \left( \frac{x - t + \Delta}{2} \right) - \operatorname{th} \left( \frac{x - t - \Delta}{2} \right) \right]. \quad (4)$$

Структура выражения (4) соответствует составному решению в методе сращиваемых асимптотических разложений [6]: глобальное решение состоит из суммы решений во «внутренней» области (в данном случае это области с быстрыми перепадами поля) за вычетом их общей асимптотики, являющейся, в свою очередь, решением во «внешней» области (с медленным изменением поля; в данном случае это единица — решение для вершины плато). Процедура сращивания и представление (4) не определяют параметры солитона (2) (в частности, ширину солитона  $\Delta$ ), однако правильно передают структуру решения, которая и закладывается в основу искомого более общего решения, описывающего взаимодействие  $N$  солитонов. В нулевом приближении, когда взаимодействие ещё не учитывается, глобальное решение будем искать в виде суммы  $2N$  чередующихся кинков (3) за вычетом суммы их общих асимптотик, равной  $N$ :

$$\Phi^{(0)}(x, t) = \frac{1}{2} \sum_{i=1}^{2N} (-1)^{i+1} \operatorname{th}\left(\frac{\xi - S_i}{2}\right). \quad (5)$$

Здесь  $\xi = x - t$ ,  $S_i(\tau, \rho)$  — медленно меняющиеся по сравнению с  $\xi$  фазы кинков,  $\tau = \varepsilon t$ ,  $\rho = \varepsilon x$ ;  $\varepsilon$  — малый параметр, по порядку величины равный  $1 - k$ . Таким образом, задачу о взаимодействии  $N$  солитонов вида (2) предполагается решать как задачу о взаимодействии  $2N$  кинков (3).

Алгоритм построения решений в следующих приближениях состоит в нахождении локальных решений и их последующем сращивании. Локальные решения  $\Phi_i(x, t)$  отыскиваются независимо вблизи каждого из кинков в виде разложения по малому параметру  $\varepsilon$ :

$$\Phi_i(x, t) = \Phi_i^{(0)}(\xi - S_i) + \sum_n \varepsilon^n \Phi_i^{(n)}(\xi - S_i, \tau, \rho), \quad (6)$$

где  $\Phi_i^{(0)}(\xi - S_i)$  — решение в виде кинка соответствующей полярности, а каждая из поправок  $\Phi_i^{(n)}$  как функция  $\xi - S_i$  находится из решения линейной задачи, получающейся после подстановки (6) в (1) и последующего разложения по степеням  $\varepsilon$ :

$$\hat{L}_i \Phi_i^{(n)}(\xi - S_i) = H_i^{(n)}(\Phi_i^{(n-1)}, \dots, \Phi_i^{(0)}). \quad (7)$$

Здесь

$$\hat{L}_i = \frac{d}{d\xi} \left( -1 + \frac{3}{2} \operatorname{ch}^{-2}\left(\frac{\xi - S_i}{2}\right) + \frac{d^2}{d\xi^2} \right)$$

— оператор, получающийся при линеаризации уравнения (1) вблизи  $i$ -го кинка, а правые части  $H_i^{(n)}$  зависят от локальных поправок  $\Phi_i^{(n)}$  предыдущих приближений. Сращивание полей, описываемых локальными решениями  $\Phi_i^{(n)}$  и  $\Phi_{i+1}^{(n)}$ , проводится в областях между кинками, где решения ищутся в виде разложения типа (6) со слагаемым  $\Phi_i^{(0)}$ , равным либо 1 (область вершин солитонов), либо 0 (область между солитонами). Поправки  $\Phi_{i+1}^{(n)}$  в этих областях удовлетворяют уравнениям  $\hat{L}_0 \Phi_{i+1}^{(n)} = H_{i+1}^{(n)}$ , где  $\hat{L}_0 = -d/d\xi + d^3/d\xi^3$  — оператор, имеющий одинаковый вид в любых промежутках между кинками.

При построении локальных решений будем иметь в виду процедуру их последующего сращивания. Поля соседних кинков с точностью до нуля (либо единицы) уже сращены, результатом чего является представление решения в нулевом приближении в виде (5). В первом приближении необходимо осуществить сращивание первых экспоненциально убывающих от центра каждого кинка асимптотик

$$\begin{pmatrix} \Phi_+^{(0)} \\ \Phi_-^{(0)} \end{pmatrix}_{\xi \rightarrow +\infty} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \end{pmatrix} e^{-\xi} + O(e^{-2\xi}), \quad \begin{pmatrix} \Phi_{+k}^{(0)} \\ \Phi_{-k}^{(0)} \end{pmatrix}_{\xi \rightarrow -\infty} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \end{pmatrix} e^{\xi} + O(e^{2\xi}) \quad (8)$$

с полями поправок  $\Phi_i^{(1)}$  вблизи соседних кинков. Поле асимптотик (8) в местах расположения соседних кинков будем считать по порядку величины равным  $\varepsilon$ . Эти асимптотики удобно явным образом включить в искомые решения первого приближения:

$$\Phi_i^{(1)} = \tilde{\Phi}_i^{(1)} + (-1)^i \left[ e^{\xi - S_{i+1}} - e^{-(\xi - S_{i-1})} \right]. \quad (9)$$

При этом процедура сращивания экспоненциально убывающих асимптотик полей соседних  $i \pm 1$  кинков (8) с экспоненциально растущими асимптотиками  $\Phi_i^{(1)}$  (9) осуществляется тождественно, если величины  $\tilde{\Phi}_i^{(1)}$  ограничены при  $|\xi - S_i| \rightarrow \infty$ . Уравнения для  $\tilde{\Phi}_i^{(1)}$  имеют вид (7), где

$$\tilde{H}_i^{(1)} = \frac{\partial S_i}{\partial t} \frac{\partial \Phi_i^{(0)}}{\partial \xi} + \frac{\partial S_i}{\partial x} \left[ 6\Phi_i^{(0)} (1 - \Phi_i^{(0)}) \frac{\partial \Phi_i^{(0)}}{\partial \xi} + 3 \frac{\partial^3 \Phi_i^{(0)}}{\partial \xi^3} \right] - (-1)^i \hat{L}_i \left[ e^{\xi - S_{i+1}} - e^{-(\xi - S_{i-1})} \right]. \quad (10)$$

Несмотря на наличие растущих асимптотик в представлении (9), правая часть  $\tilde{H}_i$  ограничена при  $|\xi - S_i| \rightarrow \infty$ , поскольку  $\hat{L}_i e^{\pm \xi} \sim d[e^{\pm \xi} \operatorname{ch}^{-2}[(\xi - S_i)/2]]/d\xi$  в силу  $\hat{L}_0 e^{\pm \xi} \equiv 0$ , и для разрешимости задачи (7), (10) в классе ограниченных при  $|\xi - S_i| \rightarrow \infty$  функций необходимо и достаточно лишь выполнения условий ортогональности  $\iint_{-\infty}^{+\infty} \partial \Phi_i^{(0)} / \partial \xi \tilde{H}_i^{(1)} d\xi' d\xi = 0$ , которые представляют собой уравнения для фаз кинков  $S_i(x, t)$ :

$$\frac{dS_i}{dt} = -2D_i + 4(-1)^i \left[ e^{-(S_{i+1}-S_i)} - e^{-(S_i-S_{i-1})} \right], \quad \frac{d}{dt} = \frac{\partial}{\partial t} + \frac{\partial}{\partial x}. \quad (11)$$

Здесь  $D_i$  — константы интегрирования, не зависящие от  $\xi$ , но являющиеся медленно меняющимися функциями  $x$  и  $t$ . В результате функции  $\tilde{\Phi}_i^{(1)}$  ограничены при любых  $\xi$  и стремятся к ненулевым пределам при  $(\xi - S_i) \rightarrow \pm\infty$ :

$$\begin{aligned} \tilde{\Phi}_{i+}^{(1)} &= \tilde{\Phi}_i^{(1)}[(\xi - S_i) \rightarrow +\infty] = (-1)^i \left[ \frac{1}{2} \frac{dS_i}{dt} + 4e^{-(S_{i+1}-S_i)} + 2e^{-(S_i-S_{i-1})} \right], \\ \tilde{\Phi}_{i-}^{(1)} &= \tilde{\Phi}_i^{(1)}[(\xi - S_i) \rightarrow -\infty] = -(-1)^i \left[ \frac{1}{2} \frac{dS_i}{dt} + 4e^{-(S_i-S_{i-1})} + 2e^{-(S_{i+1}-S_i)} \right]. \end{aligned} \quad (12)$$

В целом же асимптотики локальных решений  $\Phi_i^{(1)}$  (см. (9), (12)) имеют точно такую же структуру, что и решения в первом приближении для возмущений поля в областях между кинками:

$$\hat{L}_0 \Phi_{i:i+1}^{(1)} = 0, \quad \Phi_{i:i+1}^{(1)} = c_{1(i:i+1)} e^\xi + c_{2(i:i+1)} e^{-\xi} + c_{3(i:i+1)}, \quad (13)$$

что сводит процедуру сращивания решения в любой области между  $i$ -м и  $(i+1)$ -м кинками к выбору коэффициентов:

$$c_{1(i:i+1)} = (-1)^i e^{-S_{i+1}}, \quad c_{2(i:i+1)} = (-1)^i e^{+S_{i-1}}, \quad c_{3(i:i+1)} = \tilde{\Phi}_{i+}^{(1)} = \tilde{\Phi}_{(i+1)-}^{(1)}. \quad (14)$$

Первые два из условий (14) фиксируют положения экспоненциальных решений в области между кинками, последнее же условие приводит к ещё одному уравнению, связывающему неизвестные величины  $S_i$  и  $D_i$ :

$$[1 + (-1)^{i+1}] \frac{d}{dt} (S_{i+1} - S_i) + 2(D_{i+1} - D_i) = 0. \quad (15)$$

В соответствии с общим алгоритмом построения решения запишем глобальное решение в виде суммы локальных решений  $\Phi_i^{(1)}$  (9) за вычетом суммы их общих асимптотик  $\Phi_{i:i+1}^{(1)}$  (13):

$$\Phi^{(1)}(x, t) = \sum_i \Phi_i^{(1)} - \sum_i \Phi_{i:i+1}^{(1)} = \sum_i (\tilde{\Phi}_i^{(1)} - \tilde{\Phi}_{i+}^{(1)}). \quad (16)$$

Выражение (16), очевидно, уже не содержит неограниченно растущих слагаемых. Отметим, что алгоритм построения таких ограниченных решений, включающий представление (9) для поля локальных поправок и последующую процедуру их сращивания (14) с учётом условий ортогональности, достаточно универсален и может быть использован в других, в том числе неодномерных и неинтегрируемых системах. С неформальной стороны важно, что выявляемый при этом составной характер решения позволяет складывать поля локальных поправок  $\Phi_i^{(1)}$ , что является необходимым условием суперпозиции квазисолитонов (см. раздел 5).

## 2. УРАВНЕНИЯ ДВИЖЕНИЯ ЦЕПОЧКИ КИНКОВ

Процедура сращивания, завершившая построение общего решения в первом приближении (16), одновременно замкнула и систему уравнений (11), (15) для неизвестных величин  $S_i$  и  $D_i$ . Последовательно исключая  $D_2, D_3, \dots$  из уравнений (11), с помощью (15) можно получить цепочку уравнений непосредственно для фаз кинков:

$$\frac{dS_i}{dt} = -4 \left[ e^{-(S_{i+1}-S_i)} + e^{-(S_i-S_{i-1})} \right] + 2D_1. \quad (17)$$

Уравнения (17) примечательны в следующем отношении. Продифференцируем их по  $t$  при  $D_1 = \text{const}$  и исключим возникающие при этом слагаемые, пропорциональные  $dS_i/dt$ :

$$\frac{d^2S_i}{dt^2} = 16 \left[ e^{-(S_{i+2}-S_i)} - e^{-(S_i-S_{i-2})} \right]. \quad (18)$$

Система (18) представляет собой известную точно интегрируемую модель решётки Тоды [7]. Очевидно, что уравнения (18) распадаются на две независимые цепочки уравнений, каждая из которых описывает эволюцию последовательностей кинков с чётными и нечётными номерами соответственно. При этом уравнения (17), играющие роль условий связи, представляют собой преобразование Бэкунда для указанных подсистем [7]. Таким образом, решения цепочки уравнений (17) имеют композиционную структуру и составляются непосредственно из решений цепочек уравнений Тоды (18), отвечающих соответственно подсистемам фронтов и спадов уединённых волн, при условии, что эти решения связаны между собой преобразованием Бэкунда. В общем случае части, составляющие такие композиционные решения, имеют различную структуру и отличаются друг от друга, образно говоря, на один солитон цепочки Тоды. В то же время представляют интерес и вырожденные решения ( $D_1 = 0$ ), когда преобразования Бэкунда, не меняя структуры частных решений цепочек уравнений Тоды, приводят лишь к условиям, связывающим их параметры (см. разделы 3–5).

В случае  $D_i \equiv 0$  общее решение (16) с помощью уравнений (17) приводится к виду

$$\Phi^{(1)} = \frac{1}{4} \sum_i (-1)^{i+1} \left[ \frac{dS_i}{dt} \operatorname{th} \left( \frac{S - S_i}{2} \right) + \left( \frac{dS_i}{dt} + 2 \frac{\partial S_i}{\partial x} \right) \left( \frac{\xi - S_i}{2} \right) \operatorname{ch}^{-2} \left( \frac{\xi - S_i}{2} \right) \right]. \quad (19)$$

Наконец, отметим, что все уравнения движения цепочки кинков являются уравнениями в частных производных, и обсуждаемые ниже решения этих уравнений учитывают это обстоятельство. При этом удобно перейти от переменных  $x, t$  к переменным  $\xi = x - t, t$ . Поскольку  $\xi$  является характеристической координатой оператора  $d/dt = \partial/\partial t + \partial/\partial x$ , в уравнениях (17), (18) производная  $d/dt$  заменяется на частную производную  $\partial/\partial t$ .

### 3. ДВУХСОЛИТОННОЕ РЕШЕНИЕ

Паре взаимодействующих солитонов (2) отвечает система из четырёх кинков, описываемая в рамках приближённого подхода уравнениями для двух независимых цепочек кинков:

$$\frac{\partial^2 S_{1,3}}{\partial t^2} = \mp 16e^{-(S_3 - S_1)}, \quad \frac{\partial^2 S_{2,4}}{\partial t^2} = \mp 16e^{-(S_4 - S_2)}. \quad (20)$$

Решая эти уравнения как систему обыкновенных дифференциальных уравнений, получаем

$$S_{1,3} = S_{10} + V_1 t \mp \ln \left\{ \frac{4}{v_1} \operatorname{ch}[v_1(t - t_1)] \right\}, \quad S_{2,4} = S_{20} + V_2 t \mp \ln \left\{ \frac{4}{v_2} \operatorname{ch}[v_2(t - t_2)] \right\}. \quad (21)$$

Параметры  $v_{1,2}$ ,  $V_{1,2}$ ,  $S_{(1,2)0}$ ,  $t_{1,2}$ , возникшие как постоянные интегрирования, являются, вообще говоря, функциями характеристической координаты  $\xi$ . Зависимость этих параметров от  $\xi$  проще всего определить при  $t \rightarrow -\infty$ , когда решение имеет вид суперпозиции бесконечно разнесённых стационарных солитонов (2) с заданными параметрами  $k_{1,2} = 1 - \varepsilon_{1,2}$ ;  $k_1 > k_2$ . В этом пределе асимптотики решений (21)

$$S_{1,2}(\xi, t \rightarrow -\infty) = (V_{1,2} + v_{1,2})t + S_{(1,2)0} - v_{1,2}t_{1,2}; \\ S_{3,4}(\xi, t \rightarrow -\infty) = (V_{1,2} - v_{1,2})t + S_{(1,2)0} + v_{1,2}t_{1,2}$$

должны быть линейными функциями  $\xi$  и  $t$  и совпадать с разложениями фазовых переменных соответствующих стационарных солитонов  $k_{1,2} - k_{1,2}^3 t = \xi - \varepsilon_{1,2} (\xi - 2t) + O(\varepsilon_{1,2}^2)$ . Отсюда находим

$$v_1 = v_2 = \varepsilon_2 - \varepsilon_1, \quad V_1 = V_2 = \varepsilon_1 + \varepsilon_2, \\ 2S_{(1,2)0} = (\varepsilon_1 + \varepsilon_2)\xi \pm \tilde{\Delta}_+, \quad 2v_{1,2}t_{1,2} = (\varepsilon_2 - \varepsilon_1)\xi \pm \tilde{\Delta}_-, \quad (22)$$

где  $\tilde{\Delta}_{\pm}$  — пока неизвестные, но не зависящие от  $\xi$  и  $t$  постоянные. Подстановка (22) в (21) даёт явную зависимость величин  $S_i$  от  $\xi$ ,  $t$ :

$$S_{1,2} = \frac{1}{2}(\varepsilon_1 + \varepsilon_2)(\xi - 2t) \mp \tilde{\Delta}_+ - \ln \left\{ \frac{4}{\varepsilon_2 - \varepsilon_1} \operatorname{ch} \left[ \left( \frac{\varepsilon_2 - \varepsilon_1}{2} \right) (\xi - 2t) \pm \tilde{\Delta}_- \right] \right\}, \\ S_{3,4} = \frac{1}{2}(\varepsilon_1 + \varepsilon_2)(\xi - 2t) \mp \tilde{\Delta}_+ + \ln \left\{ \frac{4}{\varepsilon_2 - \varepsilon_1} \operatorname{ch} \left[ \left( \frac{\varepsilon_2 - \varepsilon_1}{2} \right) (\xi - 2t) \pm \tilde{\Delta}_- \right] \right\}. \quad (23)$$

Неизвестные  $\tilde{\Delta}_{\pm}$ , через которые определяются расстояния между кинками в солитонах, находятся из условий связи (17) при  $D_1 = 0$  (вырожденное решение):  $\tilde{\Delta}_{\pm} = \tilde{\Delta}_1 \pm \tilde{\Delta}_2$ ,  $2\tilde{\Delta}_{1,2} = \ln(2/\varepsilon_{1,2})$ .

Поскольку  $S_i$  зависят от  $\xi$ ,  $t$  только в комбинации  $\xi - 2t$ , производные  $\partial/\partial t + 3\partial/\partial x$  от таких решений тождественно равны нулю, и общее решение в первом приближении (19), сложенное с нулевым приближением (5), приобретает простой структурный вид:

$$\Phi(x, t) = \frac{k_0}{2} \sum_{i=1}^4 (-1)^{i+1} \left( 1 - \frac{\partial S_i}{\partial x} \right) \operatorname{th} \left\{ \frac{1}{2} [x - t - S_i(x, t)] \right\}, \\ S_{1,2} = \mp \tilde{\Delta}_+ - \ln \left\{ \frac{4}{\varepsilon_2 - \varepsilon_1} \operatorname{ch} [\varepsilon(x - 3t) \pm \tilde{\Delta}_-] \right\}, \\ S_{3,4} = \mp \tilde{\Delta}_+ + \ln \left\{ \frac{4}{\varepsilon_2 - \varepsilon_1} \operatorname{ch} [\varepsilon(x - 3t) \pm \tilde{\Delta}_-] \right\}. \quad (24)$$

Здесь  $k_0 = (k_1 + k_2)/2 = 1 - (\varepsilon_1 + \varepsilon_2)/2$ ,  $\varepsilon = (k_1 - k_2)/(k_1 + k_2) = (\varepsilon_2 - \varepsilon_1)/(2 - \varepsilon_1 - \varepsilon_2)$ ,  $x$  и  $t$  нормированы соответственно на  $k_0$  и  $1 - 3(\varepsilon_1 + \varepsilon_2)/2$ , которая приближённо равна  $k_0^3$ . Важно, что с учётом первого приближения общее решение сохраняет вид, отвечающий суперпозиции отдельных кинков. Оказывается, что и точное двухсолитонное решение может быть представлено в аналогичной форме:

$$\begin{aligned}\Phi(x, t) &= \frac{k_0}{2} \sum_{i=1}^4 (-1)^{i+1} \left( 1 - \frac{\partial \Theta_i}{\partial x} \right) \operatorname{th} \left\{ \frac{1}{2} [x - (1 + 3\varepsilon^2)t - \Theta_i(x, t)] \right\}, \\ \Theta_{1,2}(x, t) &= \mp \Delta_+ - \operatorname{Arch} \left\{ \frac{1}{\varepsilon} \operatorname{ch} [\varepsilon (x - (3 + \varepsilon^2)t) \pm \Delta_-] \right\}, \\ \Theta_{3,4}(x, t) &= \mp \Delta_+ + \operatorname{Arch} \left\{ \frac{1}{\varepsilon} \operatorname{ch} [\varepsilon (x - (3 + \varepsilon^2)t) \pm \Delta_-] \right\}. \end{aligned} \quad (25)$$

Здесь для двузначной функции  $\operatorname{Arch} y$  выбрана одна (положительная) ветвь,  $\Delta_\pm = \Delta_1 + \Delta_2$ ,  $2\Delta_{1,2} = \ln[(1 + k_{1,2})/(1 - k_{1,2})] = \ln[(2 - \varepsilon_{1,2})/\varepsilon_{1,2}]$ ,  $x$  и  $t$  нормированы на  $k_0$  и  $k_0^3$  соответственно.

Совпадение общей структуры приближённого и точного решений (ср. выражения для  $\Phi(x, t)$  из (24) и (25)) позволяет свести аналитическую оценку построенного решения (24) к сравнению фазовых переменных  $S_i(x, t)$  и  $\Theta_i(x, t)$ . Решающим обстоятельством при этом является возможность разложения функций  $\operatorname{Arch} y$  в функциональные ряды, равномерно сходящиеся по  $x, t$  при всех допустимых значениях  $\varepsilon \in (0, 1)$ :

$$\operatorname{Arch} u_\pm = \ln(2u_\pm) - \sum_{n=1}^{\infty} \frac{(2n-1)!!}{2n(2n)!!} u_\pm^{-2n}, \quad u_\pm = \frac{1}{\varepsilon} \operatorname{ch} [\varepsilon (x - (3 + \varepsilon^2)t) \pm \Delta_-]. \quad (26)$$

Нетрудно видеть, что фазовые переменные  $S_i$  близки к главным членам  $\ln(2u_\pm)$  разложений (26): тождественный функциональный вид сравниваемых выражений дополняется правильным определением как нормировочных масштабов и коэффициентов при переменных  $x, t$  (с точностью до  $O(\varepsilon_{1,2}^2)$ ), так и постоянных  $\Delta_\pm, \varepsilon^{-1}$  (с точностью до  $O(\varepsilon_{1,2})$ ).

Важной особенностью этих разложений является определяющая роль главных членов  $\ln(2u_\pm)$ , которые по абсолютной величине превосходят оставшиеся суммы в (26) при всех  $x, t$  и  $\varepsilon \in (0, 1)$ . Это позволяет в рамках приближённых решений адекватно описывать взаимодействие солитонов как с близкими, так и со значительно (в несколько раз) отличающимися параметрами  $k_{1,2}$ . Ввиду равномерной пригодности решения по  $x, t$  приближённое описание справедливо на всех этапах столкновения, включая и моменты сильного перекрытия полей отдельных уединённых волн, когда теряется их визуальная идентификация (см. раздел 4).

#### 4. ВЗАИМОДЕЙСТВИЕ СОЛИТОНОВ КАК ЧАСТИЦ В ТОЧНОМ И ПРИБЛИЖЁННОМ ОПИСАНИИ

Составной характер общего решения, позволяющий складывать поля солитонов и поправки к ним, реализует представление этого решения в виде суперпозиции квазисолитонов лишь при определённой зависимости фазовых переменных  $S_i$  от  $x$  и  $t$ :  $S_i(x, t) = S_i(\tilde{\eta} = x - 3t)$ . При таком виде решений уравнения в частных производных (20) становятся обыкновенными дифференциальными уравнениями, имеющими вид уравнений движения для классических частиц:

$$\frac{d^2 S_{i,i+2}}{d\tilde{\eta}^2} = - \frac{\partial}{\partial S_{i,i+2}} \tilde{W}(S_{i+2} - S_i), \quad \tilde{W}(\Delta S) = 16 \exp(-\Delta S), \quad (27)$$

где  $i = 1, 2$ . Фазы  $\Theta_i(x, t)$  точных солитонных решений также зависят от одной бегущей переменной  $\eta = x - (3 + \varepsilon^2)t$  и являются решениями следующей системы:

$$\frac{d^2\Theta_{i,i+2}}{d\eta^2} = -\frac{\partial}{\partial\Theta_{i,i+2}}W(\Theta_{i+2} - \Theta_i), \quad W(\Delta\Theta) = 4(1 - \varepsilon^2)\operatorname{sh}^{-2}(\Delta\Theta/2), \quad (28)$$

где  $i = 1, 2$ . Таким образом, точное двухсолитонное решение может быть представлено в виде суперпозиции квазисолитонов с пространственно-временной зависимостью их фазовых переменных (в виде бегущей стационарной волны  $\Theta_i(\eta = x - v(\varepsilon)t)$ , определяемой из решения уравнений движения для классических частиц.

Обсуждаемые представления двухсолитонных решений примечательны тем, что в наглядной форме выявляют свойства солитонов и как частиц, и как волновых объектов. Динамика квазисолитонов как целостных образований сходна с описываемой уравнениями (28) динамикой взаимно отталкивающихся частиц, не проникающих друг сквозь друга. Непосредственно из решений (25) этих уравнений следует, что в асимптотиках при  $t \rightarrow \pm\infty$  кинки одинаковой полярности, принадлежащие разным солитонам, обмениваются фазами:

$$\begin{aligned}\Theta_1(x, t \rightarrow \pm\infty) &= -(\Delta_+ \mp \Delta_-) \mp \varepsilon[x - (3 + \varepsilon^2)t] - \ln\varepsilon = \Theta_3(x, t \rightarrow \mp\infty) + \ln\varepsilon, \\ \Theta_2(x, t \rightarrow \pm\infty) &= (\Delta_+ \mp \Delta_-) \mp \varepsilon[x - (3 + \varepsilon^2)t] - \ln\varepsilon = \Theta_4(x, t \rightarrow \mp\infty) + \ln\varepsilon.\end{aligned} \quad (29)$$

Это означает, что и квазисолитоны, отождествлённые с соответствующими парами кинков в (25), также обмениваются своими параметрами, не проникая друг сквозь друга. Этот вывод иллюстрируется поведением траекторий центров кинков  $X_i(t)$  на рис. 1, определяемых нулями аргументов гиперболических тангенсов в (25):

$$X_i(t) = (1 + 3\varepsilon^2)t + \Theta_i[X_i(t) - (3 + \varepsilon^2)t], \quad (30)$$

где  $i = 1, \dots, 4$ . Из соотношений (29) следует типичный для интегрируемых систем результат: после столкновения солитоны приобретают сдвиг фаз, равный  $\pm\ln(\varepsilon^{-1})$  соответственно для быстрой (+) и медленной (−) уединённых волн. Таким образом, известная аналогия между взаимодействием солитонов и классических частиц [5] находит в обсуждаемых представлениях строгое подтверждение и распространяется на столкновения уединённых волн с любым отношением их параметров  $k_{1,2}$ .

Волновые свойства солитонов связаны с пространственно-временной зависимостью их параметров и фаз в виде стационарной бегущей волны и проявляются в деформации формы уединённых волн (см. рис. 2) и различного рода задержках, обусловленных конечной скоростью распространения стационарной волны  $\Theta_i[x - (3 + \varepsilon^2)t]$ . Важным примером таких задержек могут служить сдвинутые по времени экстремумы траекторий центров кинков одинаковой полярности, принадлежащих разным солитонам. Из рис. 1 видно, что сначала начинает тормозиться фронт первоначально быстрого солитона и лишь затем (с не малой, вообще говоря, задержкой, равной по порядку величины  $\ln(\varepsilon^{-1})$ ) начинает ускоряться фронт первоначально медленного солитона. Это вполне согласуется с причинно-следственными представлениями о конечной скорости передачи энергии и волнового импульса от быстрого солитона медленному.

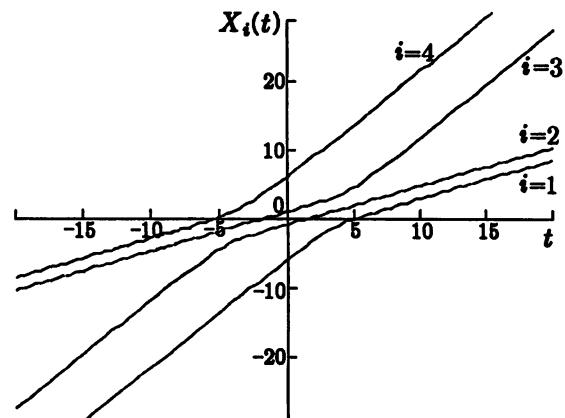


Рис. 1. Траектории центров кинков. Координаты и время в относительных единицах

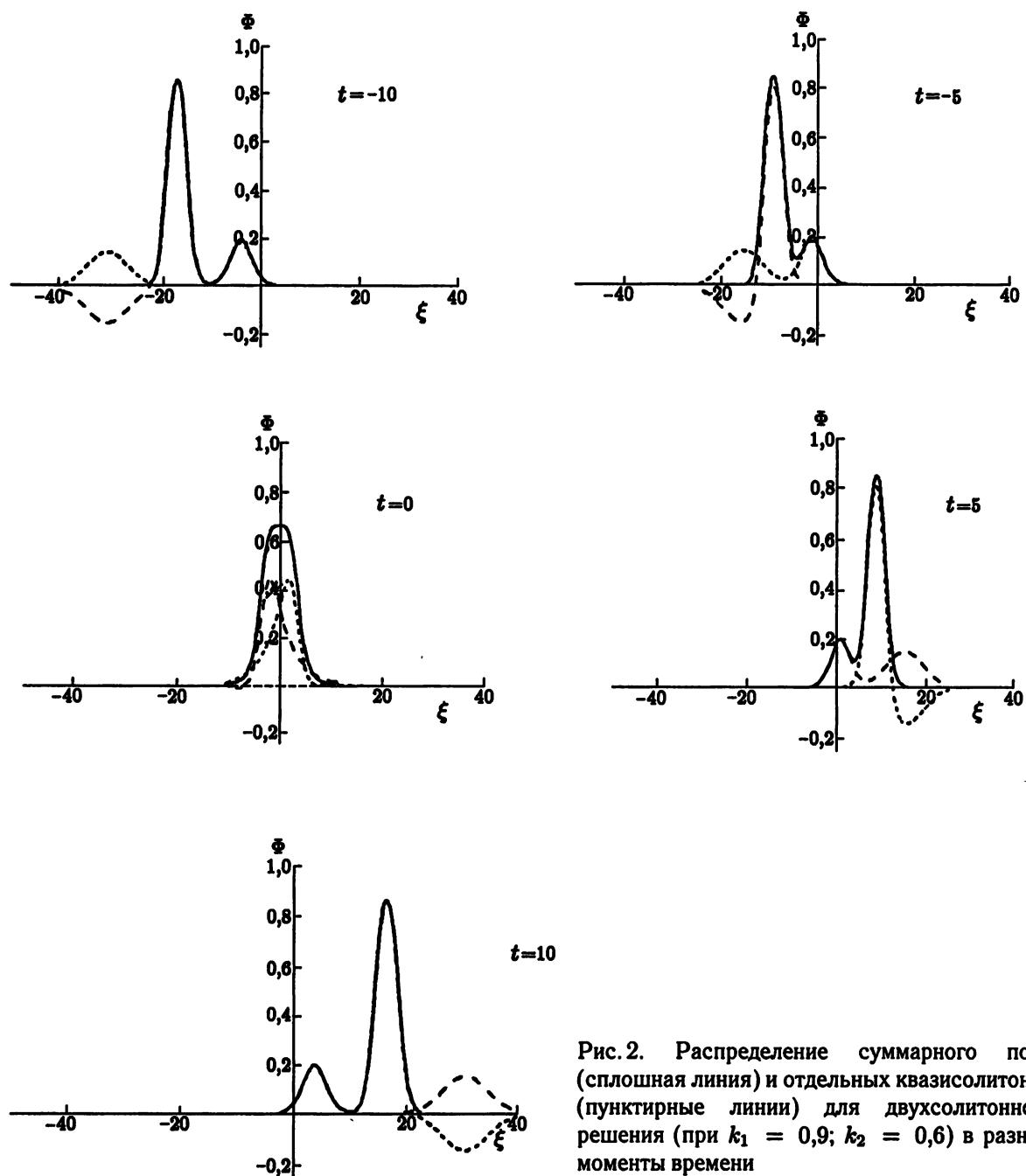


Рис. 2. Распределение суммарного поля (сплошная линия) и отдельных квазисолитонов (пунктирные линии) для двухсолитонного решения (при  $k_1 = 0,9$ ;  $k_2 = 0,6$ ) в разные моменты времени

Отличительной особенностью солитонов в модели Гарднера по сравнению с уединёнными волнами в других интегрируемых моделях является наличие у них внутренней структуры. В столкновении уединённых волн эта особенность проявляется в том, что по мере сближения солитонов сначала начинают взаимодействовать фронты солитонов (а не фронт быстрого и спад медленного солитонов, как можно было бы интуитивно ожидать) и лишь затем, с некоторым запозданием, во взаимодействие вступают спады уединённых волн (см. изгибы траекторий на рис. 1).

## 5. ЦЕПОЧКИ СОЛИТОНОВ

Особенности взаимодействия солитонов в модели Гарднера, описанные в предыдущем разделе, локальны во времени, проявляясь на этапе наибольшего сближения уединённых волн, и никак себя не обнаруживают вдали от области столкновения, при  $t \rightarrow \pm\infty$ . Ситуация кардинально меняется, если во взаимодействии участвует большое число солитонов.

Обсудим здесь решения, типичные для бесконечных последовательностей (решёток) солитонов. Как уже отмечалось выше (см. раздел 2), эти решения имеют композиционную структуру и могут составляться непосредственно из решений цепочек уравнений Тоды (18), отвечающих фронтам и спадам уединённых волн, при условии, что эти решения связаны преобразованием Бэкунда. Так,  $N$ -солитонному решению одной из подсистем будет соответствовать  $(N+1)$ -солитонное решение другой подсистемы, и решение уравнений цепочки (18) в целом имеет вид

$$S_i(t) = \begin{cases} S_i^{(N)}(t), & i \text{ — чётное (нечётное);} \\ S_i^{(N+1)}(t), & i \text{ — нечётное (чётное),} \end{cases} \quad (31)$$

где  $S_i^{(N)}(t), S_i^{(N+1)}(t)$  —  $N$ - и  $(N+1)$ -солитонные решения цепочки уравнений Тоды. В каждой из подсистем (31) асимптотики решений  $S_i^{(N)}(t), S_i^{(N+1)}(t)$  при  $t \rightarrow \pm\infty$  имеют вид соответствующим образом упорядоченных по скоростям последовательностей солитонов цепочки Тоды. Поскольку  $S_i^{(N+1)}(t)$  получено путём однократного применения преобразования Бэкунда к решению  $S_i^{(N)}(t)$ ,  $N$  солитонов в каждой из подсистем имеют попарно одинаковые параметры, так что решение уравнения (18) в общем случае описывает столкновение  $N$  пар связанных солитонов Тоды из разных подсистем и одного дополнительного солитона из одной из этих подсистем. Такие столкновения обладают, очевидно, всеми атрибутами взаимодействий солитонов в интегрируемых системах: отсутствует излучение, число и параметры участвующих в столкновении локализованных образований не меняются, единственным результатом столкновения являются сдвиги фаз. В то же время в отличие от отдельно взятой цепочки Тоды система (18) допускает существование двух типов локализованных образований, соответствующих возбуждению солитонов Тоды либо только в одной, либо одновременно в обеих подсистемах (вырожденное решение). Волна модуляции первого типа приводит к сдвигу кинков, принадлежащих одной из подсистем (фронтов или спадов исходных уединённых волн), причём это происходит таким образом, что приводит к своеобразному «обмену»: размеры солитонов и расстояния между ними в конечном состоянии равны соответственно расстояниям между солитонами и их размерам в начальном состоянии. Поскольку волны модуляции этого типа осуществляют перестройку структуры поля, их естественно назвать кинками огибающих.

Волны модуляции второго типа осуществляют сдвиг кинков обеих подсистем на одинаковое расстояние. Поэтому профили волнового поля в начальном и конечном состоянии одинаковы. Такие волны модуляции естественно назвать солитонами огибающих.

Солитоны огибающей можно трактовать как составные образования, сформированные парой кинков огибающей из разных подсистем, причём эта интерпретация прослеживается не только в рамках стационарных волн модуляции, но и в динамике их взаимодействия.

Не прибегая к процедуре построения приближённого решения (как это было сделано в разделе 2), укажем, что приближённые уравнения движения для координат кинков огибающей с точностью до масштабных коэффициентов имеют вид (18). Это означает, что в бесконечной последовательности кинков огибающей также возможно распространение возмущений в виде солитонов и кинков огибающей второго порядка со всеми свойствами, описанными выше. Повторяя эти рассуждения, можно построить семейство решений в виде иерархии волн огибающих разных порядков, описываемых моногородническими функциями с некратными, вообще говоря, периодами. Подчеркнём, что на каждом

иерархическом уровне воспроизводится составной характер солитонов огибающих, заложенный исходной моделью (1). В связи с этим отметим, что подобные решения в виде иерархии волн огибающих, построенные для моделей с простыми (не составными) солитонами (например КdВ), демонстрируют наличие только таких же (простых) солитонов огибающей на любом иерархическом уровне [8].

## 6. ЗАКЛЮЧЕНИЕ

Представление точных  $N$ -солитонных решений в виде суперпозиции квазисолитонов оказывается возможным при любом  $N$ . Рассмотрим  $N$ -солитонное решение уравнения (1) в виде, полученном методом Хироты [2]:

$$\begin{aligned} \Phi_N(x, t) &= \frac{\partial}{\partial x} \ln \frac{F_+}{F_-}, \quad F_{\pm} = 1 + f_{\pm}^{(1)} + f_{\pm}^{(2)} + \dots + f_{\pm}^{(N)}, \\ f_{\pm}^{(1)} &= \sum_{i=1}^N e^{\eta_i \pm k_i \Delta_i}, \quad f_{\pm}^{(2)} = \sum_{1 \leq i < j}^N \sum_{j=1}^N e^{\eta_i \pm k_i \Delta_i + \eta_j \pm k_j \Delta_j + A_{ij}}, \quad \dots, \\ f_{\pm}^{(N)} &= \exp \left[ \sum_{i=1}^N \eta_i \pm k_i \Delta_i + \sum_{1 \leq i < j}^N \sum_{j=1}^N A_{ij} \right], \\ \eta_i &= k_i x - k_i^3 t + \eta_{0i}, \quad \Delta_i = k_i^{-1} \operatorname{Arth} k_i, \quad A_{ij} = \left( \frac{k_i - k_j}{k_i + k_j} \right)^2, \end{aligned} \quad (32)$$

где  $k_i$  и  $\eta_{0i}$  — параметры решения. Если представить функции  $F_{\pm}$  в виде произведений:

$$F_{\pm} = \prod_{i=1}^N \left( 1 + e^{\eta - \Theta_i^{(\pm)}} \right), \quad \eta = \frac{1}{N} \sum_{i=1}^N \eta_i, \quad (33)$$

то неизвестные фазовые переменные  $\Theta_i^{(\pm)}(x, t)$  находятся как корни ( $Z_i^{(\pm)} = e^{-\Theta_i^{(\pm)}}$ ) алгебраических уравнений  $N$ -й степени:

$$\begin{aligned} \sum_{n=0}^N a_n^{(\pm)} \left( Z_n^{(\pm)} \right)^n &= 0, \quad a_N^{(\pm)} = 1, \quad -a_{N-1}^{(\pm)} = \sum_{i=1}^N e^{\eta_i - \eta_0 \pm k_i \Delta_i}, \\ a_{N-2}^{(\pm)} &= \sum_{1 \leq i < j}^N \sum_{j=1}^N e^{\eta_i - \eta \pm k_i \Delta_i + \eta_j - \eta \pm k_j \Delta_j + A_{ij}}, \quad (-1)^N a_0^{(\pm)} = \exp \left[ \sum_{1 \leq i < j}^N \sum_{j=1}^N A_{ij} \right]. \end{aligned} \quad (34)$$

Из (32) и (33) получаем искомое представление  $N$ -солитонного решения в виде суперпозиции пар кинк-ков чередующейся полярности, отвечающих квазисолитонам:

$$\Phi_N(x, t) = \frac{k}{2} \sum_{i=1}^N \left\{ \left( 1 - \frac{\partial \Theta_i^{(+)}}{\partial x} \right) \operatorname{th} \left( \frac{\eta - \Theta_i^{(+)}}{2} \right) - \left( 1 - \frac{\partial \Theta_i^{(-)}}{\partial x} \right) \operatorname{th} \left( \frac{\eta - \Theta_i^{(-)}}{2} \right) \right\}, \quad (35)$$

где  $k = N^{-1} \sum_{i=1}^N k_i$ . Выражения для фаз  $\Theta_i$  двухсолитонных решений, приведённые в разделе 3, следуют из решений квадратных уравнений (см. (34) при  $N = 2$ ).

Помимо аддитивности важным свойством представления (35) является относительно медленное (по сравнению с  $\eta$ ) пространственно-временное изменение фаз  $\Theta_i^{(\pm)}(x, t)$ . Корни уравнения (34) как функции коэффициентов  $a_n^{(\pm)}$  зависят лишь от разностей:

$$\eta - \eta_i = \varepsilon_i x - \left[ 3\varepsilon_i - 3 \left( \varepsilon_i^2 - \frac{1}{N} \sum_{j=1}^N \varepsilon_j^2 \right) - \left( \varepsilon_i^3 - \frac{1}{N} \sum_{j=1}^N \varepsilon_j^3 \right) \right] t, \quad \varepsilon_i = \frac{k - k_i}{k}, \quad (36)$$

так что характерные масштабы изменения  $\Theta_i(x, t)$  определяются параметрами  $\varepsilon_i$ , которые меньше единицы при всех допустимых значениях  $k_i$ .

Таким образом, общая структура и характер пространственно-временной зависимости точных решений, а также наличие параметров  $\varepsilon_i$ , идентичных используемым в разделе 1, позволяют заключить, что задача о взаимодействии солитонов может быть сформулирована как квазистационарная задача теории возмущений, по существу, без ограничений на параметры сталкивающихся уединённых волн; при этом точные  $N$ -солитонные решения в представлении (34), (35) демонстрируют принципиальную возможность решения этой задачи в любом порядке теории возмущений.

В первом порядке теории возмущений, т. е. в рамках обыкновенных дифференциальных уравнений, описание взаимодействия солитонов как частиц возможно благодаря одинаковой (в этом приближении) зависимости от  $x, t$  всех фаз  $\Theta_i(x, t) \simeq S_i(x - 3t)$ : определяемые посредством разностей  $\eta - \eta_i$  фазы  $\Theta_i^{(\pm)}$  с точностью до  $O(\varepsilon_i^2)$  зависят лишь от одной бегущей координаты  $x - 3t$  (см. (36)). В высших приближениях, при  $N \geq 3$ , зависимости  $\Theta_i^{(\pm)}$  от  $x, t$  становятся индивидуальными и не могут быть решениями обыкновенных дифференциальных уравнений (исключение представляет случай  $N = 2$ , см. раздел 3).

Тем не менее, как показывает проведённое в работе сравнение, уже первое приближение теории возмущений даёт адекватное описание взаимодействия уединённых волн, обладающих не только близкими, но и значительно отличающимися параметрами. Это позволяет в рамках относительно простой системы обыкновенных дифференциальных уравнений моделировать волновые процессы с достаточно произвольной пространственно-временной структурой, а не только близкой к последовательности солитонов с почти одинаковыми параметрами [5]. При этом в рамках первого приближения естественно перейти к ансамблям с бесконечным числом солитонов. В связи с этим укажем, что полученные в разделе 5 решения в виде иерархии волн огибающих, по всей видимости, являются приближениями конечно-зонных решений, описывающих эволюцию обширного класса периодических и условно периодических возмущений в интегрируемых системах [1, 2]. Можно предположить, что и для  $N$ -зонных решений могут быть найдены представления, аналогичные приведённым выше для  $N$ -солитонных решений, в виде суперпозиции  $N$  бесконечных последовательностей квазисолитонов.

Доказательство справедливости подобного представления позволило бы не только оценить приближенные решения, приведённые в разделе 5, но и дать наглядную и строгую интерпретацию конечно-зонных решений, что значительно расширило бы представление о солитонах как об элементарных образованиях, играющих конструктивную роль в самых разнообразных волновых процессах.

Проведённое обсуждение затронуло лишь наиболее очевидные стороны представления (34), (35). Невыясненными остаются вид точных уравнений движения для фаз  $\Theta_i(x, t)$ , распределение корней алгебраических уравнений и связанные с этим общие свойства поведения траекторий центров кинков и др. Вместе с тем укажем, что представление  $N$ -солитонных решений в виде суперпозиции квазисолитонов, фазовые переменные которых определяются корнями алгебраических уравнений вида (34), типичны для интегрируемых систем и могут быть получены аналогичным образом для других точно решаемых моделей (КdВ, цепочка Тоды, К-П и др.).

Работа выполнена при поддержке Российского фонда фундаментальных исследований (проекты № 00-02-16596, 01-02-17568) и ИНТАС (проект № 00-242).

## СПИСОК ЛИТЕРАТУРЫ

1. Захаров В. Е., Манаков С. В., Новиков С. П., Питаевский Л. П. Теория солитонов. М.: Наука, 1980.
2. Абловиц М., Сигур Х. Солитоны и метод обратной задачи. М.: Мир, 1987.
3. Горшков К. А., Пелиновский Д. Е., Степанянц Ю. А. // ЖЭТФ. 1993. Т. 104. С. 2704.
4. Gerdjikov V. S., Kaup D. J., Uzunov I. M., Evstatiev E. G. // Phys. Rev. Lett. 1996. V. 77. P. 3943.
5. Gorshkov K. A., Ostrovsky L. A. // Physica D. 1981. V. 3. P. 428.
6. Коул Дж. Методы возмущений в прикладной математике. М.: Мир, 1972.
7. Тода М. Теория нелинейных решёток. М.: Мир, 1984.
8. Горшков К. А., Папко В. В. // ЖЭТФ. 1977. Т. 72. С. 178.

Институт прикладной физики РАН,  
г. Нижний Новгород, Россия

Поступила в редакцию  
6 апреля 2001 г.

## INTERACTION OF SOLITONS AS COMPOUND STRUCTURES IN THE GARDNER MODEL

*K. A. Gorshkov and I. A. Soustova*

Analysis of the modernized approximate description of interaction of solitary waves as particles allows one to find a new representation of the known exact  $N$ -soliton solutions which have the form of a superposition of solitons whose parameters vary relatively slowly. The approximate solutions have the same form as the exact ones, differing from the latter by only the accuracy of representation of the spatio-temporal dependence of the solitary-wave parameters. The approximate description exhibits efficiently the details of the collision, which significantly complements the general picture of soliton interaction within the framework of specific models. These statements are discussed using the example of soliton interaction within the framework of the Gardner model. In this case, the feature of the interaction is caused by the compound soliton structure, which allows for analyzing them as a superposition of kinks (field steps) of various polarity. We develop a clear and rigorous picture of soliton interaction, which allows one to find previously unknown solutions. In particular, such solutions can have the form of envelope kinks, whose propagation over infinite sequences (chains) of solitons changes the structure of the initial wave field.

УДК 533.9

## АСИМПТОТИЧЕСКИЕ МОДЕЛИ ДИФФУЗИОННЫХ ЭФФЕКТОВ ПРИ НЕЛИНЕЙНОМ РЕЗОНАНСНОМ ВЗАИМОДЕЙСТВИИ ВОЛН С ПОТОКАМИ

*Ю. И. Троицкая*

Построена асимптотическая теория, описывающая нелокальные эффекты, обусловленные процессом слабой диффузии, при резонансном взаимодействии квазигармонических волн малой, но конечной амплитуды с потоками разной физической природы, в случае произвольного соотношения между нелинейностью и диффузией.

Изучено взаимодействие внутренних гравитационных волн с плоскопараллельными стратифицированными сдвиговыми потоками в нелинейно-диссипативном критическом слое (КС), формирующемся в окрестности резонансного уровня, в котором скорость потока совпадает с фазовой скоростью волны. Показано, что совместное действие радиационной силы во внутренней области КС и диффузии завихренности во внешнюю область приводит к установлению течения, в котором асимптотические значения средней завихренности по разные стороны от КС постоянны, но различны по величине. При выполнении условия линейной динамической устойчивости (число Ричардсона  $Ri > 1/4$ ) возникающие перепады завихренности оказываются сравнимыми по порядку величины с её невозмущённым значением. Возникает волна, отражённая от неоднородности завихренности в КС. С ростом амплитуды падающей волны средняя завихренность со стороны падения стремится к пороговому значению линейной устойчивости (число Ричардсона  $Ri > 1/4$ ), а коэффициент отражения — к минус единице.

В режиме нелинейного диссипативного КС исследовано квазистационарное асимптотическое поведение течения, формирующегося при падении внутренней гравитационной волны на динамически устойчивый стратифицированный по скорости и плотности поток, скорость которого на некотором уровне совпадает с фазовой скоростью волны. Показано, что диффузия завихренности приводит к формированию нелокальной переходной области от КС к невозмущённому течению, названной диффузионным пограничным слоем (ДПС). При этом происходит смещение КС навстречу падающей волне. Для средних полей найдено автомодельное решение, справедливое в случае постоянного перепада завихренности на КС, определены его параметры в зависимости от внутреннего числа Рейнольдса в КС, которое определяет соотношение между нелинейными и диффузионными эффектами для волнового поля в резонансной области. Определена структура и временная динамика ДПС, формирующегося при обтекании неровной поверхности потоком стратифицированной жидкости, меняющим направление на некотором уровне.

Показано, что при нелинейном резонансном взаимодействии электронов плазмы с волной Ленгмиора диффузия электронов в пространстве скоростей приводит к существенной нелокальной деформации их функции распределения за пределами области захвата. Определён вид деформированной функции распределения и вычислен декремент нелинейного затухания Ландау волны конечной амплитуды при произвольном соотношении частоты столкновений электронов и периода колебаний захваченных электронов.

### ВВЕДЕНИЕ

Исследование взаимодействия волн с потоками занимает значительное место в различных областях физики: механике жидкости и газа, включая геофизическую гидродинамику, физике плазмы, астрофизике. Наиболее интенсивным является резонансное взаимодействие, при котором фазовые скорости волн совпадают со скоростями определённой части частиц потока. Примером такого взаимодействия является линейное и нелинейное затухание Ландау продольных волн пространственного заряда на электронах разреженной плазмы [1]. Аналогом этого эффекта в общей теории нелинейных волн является адиабатическое взаимодействие длинных и коротких волн в условиях группового синхронизма [2]. Необходимость рассмотрения резонансного взаимодействия волн с потоками возникает и в гидродинамике при исследовании устойчивости плоскопараллельных потоков [3, 4]. При этом в задачах геофизической гидродинамики становятся существенными эффекты плавучести, обусловленные

*Ю. И. Троицкая*

плотностной стратификацией жидкости [3, 5, 6]. К рассмотрению аналогичного резонансного взаимодействия волн и частиц сводится и задача о нелинейном взаимодействии спиральной волны со звёздами вблизи радиуса коротации в дисках галактик [7].

Наиболее просто волны малой амплитуды в потоках можно описать в линейном приближении, пренебрегая диссипативными эффектами (столкновениями частиц, вязкостью и т. п.). В этом случае уравнение, описывающее гармонические волновые возмущения (линеаризованное уравнение Власова, уравнение Релея, уравнение Тейлора—Гольдштейна и т. п.), имеет особенность в резонансной точке. Это означает, что в некоторой её окрестности линейное бездиссипативное приближение неприменимо, и необходимо учитывать ряд дополнительных факторов: конечность амплитуды волновых полей (нелинейность), диссипацию, конечность ширины спектра волновых полей или комбинацию этих факторов. При этом область, в которой существенны указанные факторы, в гидродинамике называется критическим слоем (нелинейным, диссипативным, нестационарным или диссипативно-нелинейным и т. п. в зависимости от учтённых факторов). В физике плазмы критическому слою (КС) можно поставить в соответствие область захваченных частиц.

Если перечисленные выше факторы достаточно слабы, то их влияние существенно только внутри резонансной области. Во внешней области для волновых полей применимо линейное бездиссипативное стационарное приближение. При этом для описания взаимодействия волны с потоком можно использовать метод сращиваемых асимптотических разложений, разработанный применительно к задачам гидродинамики [8].

К настоящему времени наиболее изучено резонансное взаимодействие волн с потоками в бездиссипативном приближении. При этом энергообмен волны с потоком имеет место в узкой резонансной области (КС, область захвата и т. п.). В этой же области происходит деформация распределения частиц (образование плато на функции распределения частиц плазмы, выравнивание профиля завихренности сдвигового потока и т. п.).

Бездиссипативное приближение применимо для достаточно быстро протекающих процессов. Для изучения асимптотики поведения системы волна—поток на больших временах необходимо учитывать диффузию частиц потока в пространстве, в котором задано их распределение. В случае частиц плазмы это пространство скоростей, в случае адиабатического взаимодействия длинных и коротких волн — пространство волновых чисел коротких волн, в случае КС в сдвиговом потоке — физическое пространство. При этом возникает новый нелинейный эффект. Нелинейное взаимодействие волн с частицами или потоком приводит к существенной деформации функции распределения в резонансной области, т. е. распределение становится неравновесным. Диффузия частиц в пространстве скоростей, стремящаяся восстановить равновесие, приводит к глобальной деформации всего распределения частиц. Причём сколь угодно слабая диффузия на достаточно больших временах приводит к принципиальному изменению картины деформации среднего распределения частиц по сравнению с бездиссипативным случаем. Это сильный нелинейный эффект, который определяет главным образом нелинейные свойства рассматриваемой системы.

Совместное действие нелинейности и диффузии было впервые рассмотрено в гидродинамической задаче о квазистационарном КС в плоскопараллельном потоке однородной несжимаемой жидкости, скорость которого меняется в поперечном направлении [9]. Впоследствии этот эффект был рассмотрен в случае слабостратифицированной жидкости [10] и в случае сильной устойчивой стратификации [11].

В настоящей работе будут рассмотрены особенности эффекта нелокальной деформации среднего распределения частиц на примерах из различных областей физики. В разделе 1 будет рассмотрен пример из геофизической гидродинамики — диссипативный нелинейный критический слой в динамически устойчивом стратифицированном сдвиговом потоке. В разделе 2 будет рассмотрена нелокальная деформация распределения электронов горячей плазмы при резонансном взаимодействии с продольной волной пространственного заряда.

## 1. НЕЛИНЕЙНЫЙ ДИССИПАТИВНЫЙ КРИТИЧЕСКИЙ СЛОЙ В СТРАТИФИЦИРОВАННОМ СДВИГОВОМ ПОТОКЕ

### 1.1. Стационарный критический слой

**1.1.1. Постановка задачи. Внешняя задача, скейлинг.** Рассмотрим двумерное течение стратифицированной несжимаемой жидкости с учётом вязкости и теплопроводности. В приближении Буссинеска оно описывается системой уравнений для завихренности и плотности, которую можно записать в безразмерном виде:

$$\frac{\partial}{\partial t} \Delta \psi + J(\Delta \psi, \psi) - \frac{\partial b}{\partial x} = \frac{\Delta \Delta \psi}{Re_0}, \quad \frac{\partial b}{\partial t} + J(b, \psi) = \frac{\Delta b}{Re_0 Pr}. \quad (1)$$

Здесь  $\Delta = \partial^2 / \partial x^2 + \partial^2 / \partial z^2$  — оператор Лапласа,  $J(a, b) = \partial(a, b) / \partial(x, z)$  — оператор Якоби,  $x = x_{\text{dim}} / L_0$ ,  $z = z_{\text{dim}} / L_0$  — безразмерные горизонтальная и вертикальная координаты,  $L_0$  — характерный масштаб течения,  $t = t_{\text{dim}} U_0 / L_0$  — безразмерное время,  $U_0$  — характерная скорость течения,  $\psi = \psi_{\text{dim}} / (U_0 L_0)$  — безразмерная функция тока,  $b = (\rho - \rho_0) g / (N_0^2 L_0 \rho_0)$  — безразмерная плотность,  $\rho_0$  — невозмущённая плотность,  $g$  — ускорение свободного падения,  $N_0$  — характерная частота плавучести,  $Re_0 = U_0 L_0 / \nu$  — число Рейнольдса, определённое по параметрам среднего течения,  $Pr = \nu / \kappa$  — число Прандтля,  $\nu$  и  $\kappa$  — коэффициенты кинематической вязкости и диффузии.

Рассмотрим возмущение в виде гармонической плоской волны малой амплитуды  $\epsilon$  на фоне среднего течения с профилем скорости  $U(z)$ , стратифицированного по плотности по закону  $b_0(z)$ . В этом случае решение системы (1) можно искать в виде разложения в ряд по степеням  $\epsilon$ . При этом

$$\begin{aligned} \psi &= \int U(z) dz + \epsilon \operatorname{Re} \left\{ \Psi^{(1)}(z) \exp[i(\omega t - kx)] \right\} + \epsilon^2 \Psi^{(2)} + \dots, \\ b &= \int b_0(z) dz + \epsilon \operatorname{Re} \left\{ b^{(1)}(z) \exp[i(\omega t - kx)] \right\} + \epsilon^2 b^{(2)} + \dots. \end{aligned} \quad (2)$$

Здесь  $\omega$  и  $k$  — частота и волновое число возмущения.

В первом порядке по  $\epsilon$  и в нулевом порядке по  $Re_0^{-1}$  (в линейном бездиссипативном приближении) функция  $\Psi^{(1)}(z)$  удовлетворяет уравнению Тейлора—Гольдштейна

$$\frac{d^2 \Psi^{(1)}}{dz^2} + \frac{d^2 U / dz^2}{c - U} \Psi^{(1)} + \left[ \frac{N^2}{(c - U)^2} - k^2 \right] \Psi^{(1)} = 0. \quad (3)$$

Здесь  $N^2 = db_0 / dz$  — квадрат нормированной частоты плавучести,  $c = \omega / k$  — фазовая скорость волны.

Уравнение (3) имеет особенности в точке  $U(z_c) = c$ , а его решение — точку ветвления при  $z = z_c$ : при числе Ричардсона  $Ri > 1/4$  функция  $\Psi^{(1)} \sim (z - z_c)^{1/2 \pm i\mu}$ , где  $\mu = \sqrt{Ri - 1/4}$ . Это означает, что в окрестности КС уравнение (3) и его решения, вообще говоря, несправедливы. В малой окрестности КС необходимо учесть дополнительные факторы: диссипацию [12–16], нелинейность [17–19] (см. также обзоры [6, 20] и ссылки в них) или нестационарность [21–23]. Области, в которых существенны диссипация, нелинейность или нестационарность, называют соответственно вязким, нелинейным и нестационарным критическими слоями; их вертикальные масштабы составляют [19]

$$\delta_{\text{vis}} = (Re_0 k dU / dz)^{-1/3}; \quad \delta_{\text{nl}} = \epsilon^{2/3}; \quad \delta_t = (|\partial \psi / \partial t| / |\psi|) (k dU / dz)^{-1}. \quad (4)$$

В настоящей работе рассматривается совместное действие нелинейности и вязкости в стационарном КС, т. е.  $\delta_{\text{vis}} \sim \delta_{\text{nl}} \gg \delta_t$ . Если амплитуда волны мала, а число Рейнольдса велико, так что  $\delta_{\text{vis}}$  и

$\delta_{nl}$  малы, то задачу можно решать методом сращиваемых асимптотических разложений. При этом внутреннее разложение внешнего решения представляет собой сумму средних полей и решения уравнения Тейлора—Гольдштейна (3), представленного в виде разложения Фробениуса:

$$\psi_{\pm} = \varepsilon^{2/3} c_{\pm} z + \Gamma_{\pm} z^2/2 + \dots + \varepsilon \sum_{n=1}^{\infty} \operatorname{Re} \left\{ \left[ A_{\pm}^{(n)} |z|^{1/2+i\mu_{\pm}} \left( 1 + \alpha_1^{(n)} z + \dots \right) + B_{\pm}^{(n)} |z|^{1/2-i\mu_{\pm}} \left( 1 + \beta_1^{(n)} z + \dots \right) e^{i\xi n} \right] \right\}. \quad (5)$$

Здесь индексы + и – относятся к решениям выше и ниже КС,  $\xi = k(x - ct)$  — нормированная горизонтальная «бегущая» координата. Слагаемые  $\Gamma_{\pm} z^2/2$  соответствуют значениям завихренности по разные стороны от КС, т. е. имеется скачок завихренности на КС, аналогичный полученному в [9, 10], возникающий, как это будет показано ниже, в нулевом порядке по  $\varepsilon$ . Кроме того, в решении (5) учитываются высшие гармоники, появление которых связано с нелинейностью КС. Причём подобно тому, как скачок завихренности возникает в нулевом порядке по  $\varepsilon$ , высшие гармоники имеют тот же порядок, что и основная. Это свойство сингулярных возмущений.

Задача состоит в нахождении связи между величинами с индексами + и –, т. е. правил обхода. Для этого необходимо найти внутреннее решение. Из (5) следует, что вертикальная внутренняя (внутри КС) координата должна быть определена следующим образом:  $\eta = z/\varepsilon^{2/3}$  (см. [9]), а внутреннее решение надо искать в виде рядов по степеням  $\varepsilon^{2/3}$ :

$$\psi = \varepsilon^{4/3} (\Phi + \varepsilon^{2/3} \varphi_1 + \dots); \quad b = \varepsilon^{2/3} (B + \varepsilon^{2/3} b_1 + \dots).$$

В нулевом порядке по  $\varepsilon$  уравнения для  $\Phi$  и  $B$  имеют вид

$$\frac{\partial^3 \Phi}{\partial \eta^2} \frac{\partial \Phi}{\partial \eta} - \frac{\partial^3 \Phi}{\partial \eta^3} \frac{\partial \Phi}{\partial \xi} - \operatorname{Ri} \frac{\partial b}{\partial \xi} = \lambda \frac{\partial^4 \Phi}{\partial \eta^4}, \quad (6a)$$

$$\frac{\partial B}{\partial \xi} \frac{\partial \Phi}{\partial \eta} - \frac{\partial b}{\partial \eta} \frac{\partial \Phi}{\partial \xi} = \frac{\lambda}{\operatorname{Pr}} \frac{\partial^2 B}{\partial \eta^2}. \quad (6b)$$

Здесь  $\operatorname{Ri} = N_-^2/\Gamma_-^2$  — число Ричардсона при  $\eta \rightarrow -\infty$ ,  $\lambda = (\operatorname{Re}_0 \Gamma_- \varepsilon^2)^{-1}$  — параметр, характеризующий отношение вязкости и нелинейности (более точно  $\lambda = (\delta_{vis}/\delta_{nl})^3$ ).

Из вида уравнений (6) ясно, что  $\lambda$  имеет смысл обратного внутреннего числа Рейнольдса ( $\operatorname{Re}_i = \lambda^{-1}$ ), определённого по амплитуде волнового возмущения в окрестности КС. Предел  $\lambda \rightarrow \infty$  ( $\operatorname{Re}_i \ll \ll 1$ ) соответствует вязкому линейному течению в окрестности КС, а  $\lambda \rightarrow 0$  ( $\operatorname{Re}_i \gg 1$ ) — невязкому нелинейному течению.

Границные условия для  $\Phi$  и  $B$  определяются условиями сращивания с внешним решением. Внутреннее решение очевидно зависит от амплитуд гармоник волнового поля вне КС ( $A_{\pm}^{(n)}$ ,  $B_{\pm}^{(n)}$ ), т. е. от постановки задачи. В [11] рассмотрена задача о падении волны на КС, её отражения с коэффициентом  $R$  и прохождения с коэффициентом  $T$ . Такой постановке задачи соответствуют следующие амплитуды гармоник при  $\eta \rightarrow \pm\infty$ :

$$\begin{aligned} |B_+^{(1)}| &= 1; \quad A_+^{(1)} = R; \quad B_-^{(1)} = T; \quad A_-^{(1)} = 0; \\ A_+^{(n)} &= R_n; \quad B_-^{(n)} = T_n; \quad B_+^{(n)} = A_-^{(n)} = 0, \end{aligned} \quad (7)$$

где  $n = 2, 3, \dots$ ,  $R_n$  и  $T_n$  — амплитуды  $n$ -х гармоник, излучаемых соответственно вверх и вниз от КС. В [11] использована нормировка на амплитуду падающей волны. Учитывалось также, что КС излучает волны высших гармоник.

**1.1.2. Особенности асимптотического поведения средних полей при переходе через стационарный критический слой.** Ряд интегральных соотношений может быть получен без решения системы (6). Так, чтобы найти выражение для скачка завихренности, проинтегрируем уравнение (6а) по  $\eta$  от  $-\infty$  до  $\eta$  и два раза и усредним по  $\xi$ . С учётом периодичности решения по  $\xi$  и поведения полей на бесконечности (см. (5) и (7)) после несложных преобразований получим

$$2\lambda(\Gamma_+ - 1) = \mu_+(1 - |R|^2) - \mu_-|T|^2 + \mu_+\sum_{n=1}^{\infty}|R_n|^2 - \mu_-\sum_{n=1}^{\infty}|T_n|^2. \quad (8)$$

Уравнение (8) показывает, что радиационная сила, равная разности волновых потоков импульса выше и ниже КС (левая часть (8)), в стационарном течении компенсируется вязкой силой, равной разности вязких напряжений выше и ниже КС (правая часть (8)). В результате формируется течение, в котором завихренность при удалении вверх и вниз от КС стремится к некоторым постоянным, которые, вообще говоря, различны. Аналогично из уравнения (6б) можно получить  $N_+^2 = N_+^2, N_-^2 = 1$ . Отсутствие скачка градиента плотности при переходе через КС связано с тем, что волновые потоки массы выше и ниже КС одинаковы (равны нулю), поэтому в стационарном состоянии должны быть равны и диффузионные потоки, а значит, и градиенты плотности.

Итак, резонансное взаимодействие волны с потоком в резонансной области при учёте диффузии завихренности, обусловленной вязкостью, приводит к появлению перепада завихренности при переходе через КС. Это означает, что возникают возмущения среднего профиля скорости, растущие при удалении от КС, т. е. возмущения средних полей становятся нелокальными. Как отмечалось в [3, 6], возникает вопрос о практической реализуемости такого течения. Рост возмущения с удалением от КС означает, что возмущение среднего течения становится нелокальным, и вопрос о его реализуемости должен решаться с учётом граничных условий на бесконечности (этот вопрос рассматривается в разделе 2).

**1.1.3. Характеристики волновых полей в нелинейном диссипативном критическом слое.** По результатам численного решения системы (6) были построены зависимости от  $\lambda$  всех величин, входящих в формулу (8):  $\Gamma_+$  (рис. 1), нормированного коэффициента прохождения  $|T| e^{\pi\mu_-}$  (рис. 2) и  $|R|$  (рис. 3), а также фазы коэффициента отражения  $f_R$  (рис. 4) при четырёх значениях числа Ричардсона:  $Ri = 0,5; 1; 2; 3$  (подробности численной процедуры см. в [11]).

Зависимость  $\Gamma_+(\lambda)$  в широком диапазоне значений  $\lambda$  можно описать простой формулой. Действительно, из рис. 2 и 3 легко видеть, что для всех  $Ri \geq 1$  при  $\lambda \geq 0,15$  справедливы неравенства  $|R|^2 \ll 1$ ;  $|T|^2 \ll 1$ . Но в этом случае из (8) следует

$$\lambda(\Gamma_+) = \frac{\sqrt{Ri - 1/4}}{2(\Gamma_+ - 1)}. \quad (9)$$

На рис. 1 эта зависимость отмечена пунктиром. Видно, что почти при всех значениях  $\lambda$  она совпадает с зависимостью, полученной в результате численного расчёта. Отличия имеются лишь в узкой области вблизи  $\lambda = 0$ , т. е. в сильно нелинейном режиме.

Как из численного расчёта, так и из формулы (9) следует, что при стремлении  $\lambda$  к нулю  $\Gamma_+$  стремится к значению  $2Ri^{1/2}$ , при котором число Ричардсона выше КС  $Ri_+ = 1/4$ , а  $\mu_+ = 0$ . Однако характер зависимости, следующий из (9), отличается от полученного численно. В последнем случае, как видно из рис. 2–4, коэффициент отражения  $R$  стремится к минус единице, а коэффициент прохождения  $T$  — к нулю. Этот численный результат можно пояснить с помощью следующих качественных рассуждений. При стремлении  $Ri_+$  к  $1/4$  поле волны, пропорциональное  $z^{i\sqrt{Ri_+-1/4}}$ , становится всё менее осциллирующим. Это означает, что эффективная вертикальная длина волны выше КС стремится к бесконечности. В то же время ниже КС число Ричардсона не близко к  $1/4$ , и длина волны конечна.

Таким образом, возникает ситуация, эквивалентная отражению волн на границе раздела сред с сильно отличающимися показателями преломления при переходе волны из оптически менее плотной среды в более плотную. Из вида асимптотики волнового поля (5) ясно, что роль вертикальных компонент волновых векторов вблизи КС играют параметры  $\mu_{\pm}$ . Но тогда коэффициенты отражения и прохождения имеют вид

$$R = -1 + O(\mu_+/\mu_-), \quad T = O(\mu_+/\mu_-). \quad (10)$$

С учётом (10) из формулы (8) следует, что  $\lambda = O\left\{\mu_+^2/[\mu_- (2Ri^{1/2} - 1)]\right\}$ . При этом комплексная амплитуда волнового поля выше КС есть малая величина порядка  $\mu_+$ :

$$\varphi_{1+} = \eta^{1/2-i\mu_+} + [-1 + O(\mu_+/\mu_-)] \eta^{1/2+i\mu_+} \approx \eta^{1/2} O(\mu_+/\mu_-) - 2i\mu_+ \eta^{1/2} \ln \eta.$$

В этом случае естественными внутренними переменными, в которых амплитуды волновых полей порядка единицы, являются следующие функции:  $\eta_{\text{new}} = \eta/\mu_+^{2/3}$ ;  $\varphi_{\text{new}} = \varphi/\mu_+^{4/3}$ ;  $b_{\text{new}} = b/\mu_+^{2/3}$ . В новых переменных скачок завихренности остаётся тем же, а поля  $\varphi_{\text{new}}$  и  $b_{\text{new}}$  удовлетворяют системе (6), в которой  $\lambda_{\text{new}}$  порядка  $\lambda\mu_+^{-2}$ , или  $\lambda_{\text{new}} = O\left[(Ri - 1/4)^{-1/2}/(2Ri^{1/2} - 1)\right]$ , т. е. с ростом  $Ri$  параметр  $\lambda_{\text{new}}$  убывает (КС становится всё более нелинейным).

Итак, с ростом числа Рейнольдса, определяемого по амплитуде падающей волны ( $\lambda \rightarrow 0$ ),  $\Gamma_+$  стремится к  $2Ri^{1/2}$ , при котором  $Ri_+ = 1/4$ . При этом возникает отражённая волна, коэффициент отражения которой близок к единице, находящаяся в противофазе с падающей. В результате амплитуда эффективного волнового поля существенно меньше, чем у падающей волны, причём при увеличении амплитуды падающей волны суммарное поле остаётся неизменным ( $\lambda_{\text{new}}$  не зависит от  $\lambda$ ). В результате при  $\lambda \rightarrow 0$  течение внутри КС не меняется, а число Рейнольдса этого течения  $Re_i$ , определяемое параметром  $\lambda_{\text{new}}$ , а не  $\lambda$ , остаётся неизменным и небольшим. Оценка внутреннего числа Рейнольдса  $Re_i$  по  $\lambda$  для  $Ri = 0,5$  даёт  $Re_i \approx 10$ , для  $Ri = 1 - Re_i \approx 30$ , для  $Ri = 2 - Re_i \approx 60$ , для  $Ri = 3 - Re_i \approx 100$ .

## 1.2. Нелинейный квазистационарный диссипативный критический слой в стратифицированном сдвиговом потоке

**1.2.1. Постановка задачи.** Итак, при переходе волны конечной амплитуды через КС в вязком потоке формируется профиль средней скорости со скачком завихренности

$$\langle \Omega \rangle / \Omega_0 = 1 + \Gamma + \Delta\Gamma \operatorname{sign}(z - z_c)/2, \quad (11)$$

где  $\tau = t\varepsilon^{2/3}$ ,  $\Omega_0$  — невозмущённая завихренность,  $z$  — вертикальная координата,  $z_c$  — координата КС,  $\Delta\Gamma$  — нормированный скачок завихренности,  $\Gamma$  — константа. Оказывается, что  $\Delta\Gamma$  больше по порядку величины, чем возмущение амплитуды  $\varepsilon$ , вызвавшее его. Возмущение скорости, соответствующее возмущению завихренности (11), неограниченно растёт при удалении от критического слоя. Чтобы ограничить этот рост, необходимо выйти из рамок стационарного приближения и учсть временнюю зависимость, как это предложено в [3, 6] и реализовано в [24].

Следует также принять во внимание, что скачок завихренности  $\Delta\Gamma$  может быть получен в рамках стационарного приближения, но  $\Gamma$  — произвольная постоянная стационарной задачи. Это значит, что завихренность по разные стороны от КС ( $\Gamma_+ = 1 + \Gamma + \Delta\Gamma/2$  и  $\Gamma_- = 1 + \Gamma - \Delta\Gamma/2$ ) может быть вычислена только в рамках нестационарной задачи, где  $\Gamma$  может быть определено, причём результат

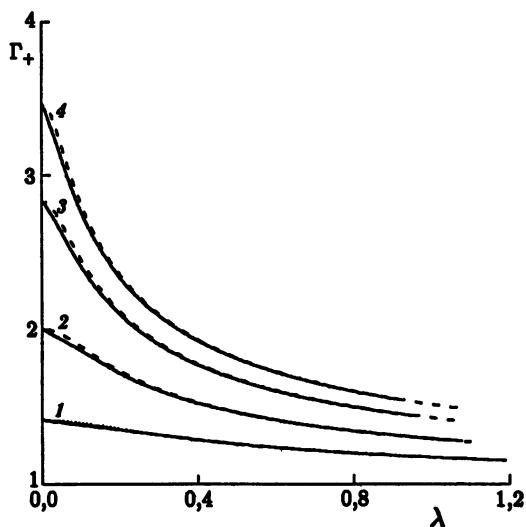


Рис. 1. Зависимость завихренности выше критического слоя от  $\lambda$  для различных чисел Ричардсона: кривая 1 соответствует  $Ri = 0,5$ ; 2 —  $Ri = 1$ ; 3 —  $Ri = 2$ ; 4 —  $Ri = 3$ . Число Прандтля  $Pr = 0,71$ . Пунктирные линии — функции  $\Gamma_+(\lambda)$ , рассчитанные по формуле (9)

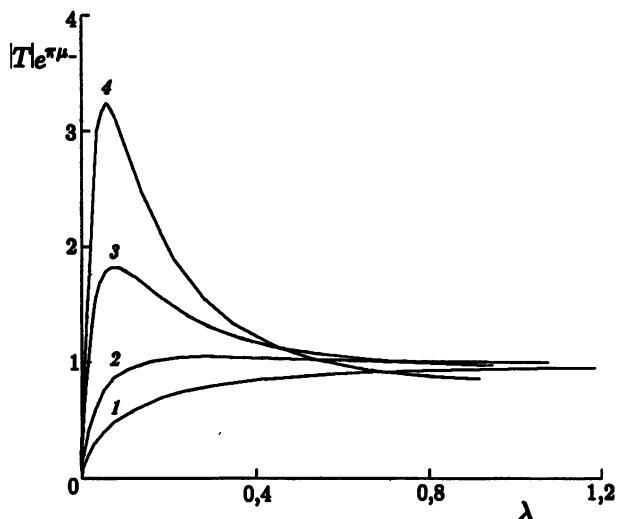


Рис. 2. Зависимость нормированного коэффициента прохождения от  $\lambda$  для различных чисел Ричардсона: кривая 1 соответствует  $Ri = 0,5$ ; 2 —  $Ri = 1$ ; 3 —  $Ri = 2$ ; 4 —  $Ri = 3$ . Число Прандтля  $Pr = 0,71$

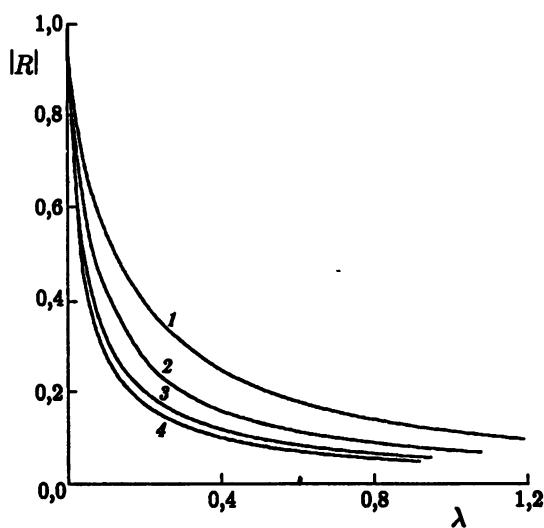


Рис. 3. Зависимость модуля коэффициента отражения  $|R|$  от  $\lambda$  для различных чисел Ричардсона: кривая 1 соответствует  $Ri = 0,5$ ; 2 —  $Ri = 1$ ; 3 —  $Ri = 2$ ; 4 —  $Ri = 3$ . Число Прандтля  $Pr = 0,71$

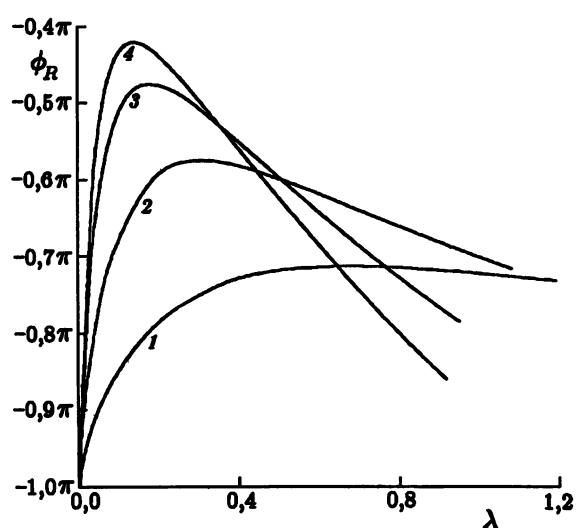


Рис. 4. Зависимость фазы коэффициента прохождения  $\phi_R$  от  $\lambda$  для различных чисел Ричардсона: кривая 1 соответствует  $Ri = 0,5$ ; 2 —  $Ri = 1$ ; 3 —  $Ri = 2$ ; 4 —  $Ri = 3$ . Число Прандтля  $Pr = 0,71$

зависит от граничных условий во внешней области КС.

Рассмотрим, как происходит установление скачка завихренности в двумерном плоскопараллельном потоке стратифицированной жидкости с  $Ri > 1/4$ . В системе уравнений гидродинамики в этом случае надо принять во внимание зависимость полей от времени. Тогда система уравнений, которой удовлетворяет внутреннее решение, принимает вид (ср. с (6а, б))

$$\frac{\partial \chi}{\partial \tau} + \frac{\partial \chi}{\partial \xi} \frac{\partial \varphi}{\partial \eta} - \frac{\partial \chi}{\partial \eta} \frac{\partial \varphi}{\partial \xi} - Ri \frac{\partial b}{\partial \xi} = \lambda \frac{\partial^2 \chi}{\partial \eta^2},$$

$$\frac{\partial b}{\partial \tau} + \frac{\partial b}{\partial \xi} \frac{\partial \varphi}{\partial \eta} - \frac{\partial b}{\partial \eta} \frac{\partial \varphi}{\partial \xi} = \frac{\lambda}{Pr} \frac{\partial^2 b}{\partial \eta^2}, \quad \chi = \frac{\partial^2 \varphi}{\partial \eta^2}. \quad (12)$$

Как и раньше, волновое поле предполагается периодической функцией горизонтальной координаты  $\xi$ , поэтому удобно представить гидродинамические поля в виде суммы двух компонент — средней по периоду возмущения и переменной (пространственно-периодической), т. е.

$$\begin{bmatrix} \varphi \\ b \\ \chi \end{bmatrix} = \begin{bmatrix} \varphi_0(\eta, \tau) \\ b_0(\eta, \tau) \\ \chi_0(\eta, \tau) \end{bmatrix} + \begin{bmatrix} \varphi'(\eta, \xi, \tau) \\ b'(\eta, \xi, \tau) \\ \chi'(\eta, \xi, \tau) \end{bmatrix}.$$

Уравнения для средних полей легко могут быть получены усреднением уравнений системы (12) по периоду возмущений. При этом уравнение для средней горизонтальной компоненты скорости имеет вид:

$$\frac{\partial u_0}{\partial \tau} - \lambda \frac{\partial^2 u_0}{\partial \eta^2} = F_{\text{rad}}, \quad (13)$$

где  $F_{\text{rad}} = \partial T_p / \partial \eta$ , а

$$T_p = -\frac{1}{2\pi} \int_0^{2\pi} \frac{\partial \varphi}{\partial \eta} \frac{\partial \varphi}{\partial \xi} d\xi \quad (14)$$

— вертикальный поток горизонтальной компоненты импульса (или радиационное напряжение). Уравнение (13) представляет собой закон движения жидкой частицы, усреднённый по периоду возмущения. При этом среднее ускорение определяется балансом радиационной и вязкой сил.

Будем полагать, что в начальный момент времени профиль скорости линеен, т. е. с учётом нормировки начальное условие имеет вид

$$u_0 = \eta.$$

Из принципа причинности следует, что при  $\eta \rightarrow \pm\infty$  выполняются граничные условия

$$u_0(\eta \rightarrow \pm\infty, \tau) = \eta. \quad (15)$$

Уравнения для средней плотности аналогичны (13)–(15).

Рассмотрим качественно эволюцию среднего потока на больших временах. Будем полагать, что временная изменчивость полей в окрестности КС определяется только диффузионными процессами. Это предположение правомочно, поскольку при  $Ri > 1/4$  течение устойчиво. Очевидно, что КС становится квазистационарным на временах, когда характерный масштаб диффузионного расплывания  $\sqrt{\lambda\tau}$  (диффузионная длина) велик по сравнению с масштабом  $\delta$  КС, т. е.

$$\tau \gg \delta^2 / \lambda. \quad (16)$$

Для квазистационарных внутренних волн хорошо известно [25], что поток импульса  $T_p$  в них постоянен по  $\eta$  вне КС, т. е. производная  $\partial T_p / \partial \eta$  отлична от нуля лишь в окрестности КС толщиной  $\delta$ . Качественно картина деформации среднего течения выглядит следующим образом. В КС на течение действует радиационная сила. Под действием вязкости происходит диффузия завихренности от КС, причём на достаточно больших временах диффузионный масштаб много больше  $\delta$ . Допустим, следуя работе [11], что волна распространяется из области  $\eta > 0$ . В этом случае (см. [11]) направление радиационной силы в КС таково, что при деформации профиля скорости завихренность в области  $\eta > 0$  больше, а при  $\eta < 0$  — меньше её невозмущённого значения, а достаточно далеко от КС (при  $\eta \rightarrow \pm\infty$ ) возмущение завихренности отсутствует. При этом формируется переходная область от КС к невозмущённому течению, которую мы назвали диффузионным пограничным слоем (ДПС), характерный масштаб которого — диффузионная длина  $\sqrt{\lambda\tau}$ . КС представляет собой точку на профиле скорости, в которой по определению скорость постоянна и равна фазовой скорости волны. Такое течение может быть реализовано только в случае, если КС смещается в сторону  $\eta > 0$ , т. е. навстречу падающей волне. Такое смещение КС отмечалось в численных экспериментах [26–28] и наблюдалось в лабораторных условиях [29, 30]. В общем виде закон смещения КС можно записать следующим образом:  $\eta = s(\tau)\sqrt{\lambda\tau}$ .

При описании гидродинамических полей в окрестности смещающегося КС удобно перейти к «бегущей» вертикальной координате

$$h = \eta - s(\tau)\sqrt{\lambda\tau}. \quad (17)$$

Кроме того, рассматривая диффузионные процессы, удобно вместо временной переменной  $\tau$  пользоваться величиной  $\nu = 1/(2\sqrt{\lambda\tau})$ , обратной диффузионной длине.

Оценки показывают, что на больших временах (при  $\nu \ll 1$ ) уравнения, описывающие течение жидкости внутри КС, совпадают с (6а, б) с точностью до замены  $\eta \rightarrow h$ .

При рассмотрении течений в ДПС естественно перейти к новым «диффузионным» переменным:  $\beta = \nu b^{(0)}$ ,  $V = \nu u^{(0)} = \nu \partial \varphi^{(0)} / \partial h$ ,  $X = \chi^{(0)}$ . Здесь индекс 0 обозначает нулевое приближение по параметру  $\nu$ ,  $H = h/\nu$ . В [24] показано, что в нулевом порядке по  $\nu$  средняя завихренность удовлетворяет уравнению

$$\frac{\partial^2 X_0}{\partial H^2} + 2\nu \frac{\partial X_0}{\partial \nu} + 2 \left( H + \frac{s}{2} - \frac{\nu}{2} \frac{ds}{d\nu} \right) \frac{\partial X_0}{\partial H} = 0. \quad (18)$$

Из условий сращивания внешнего и внутреннего решений можно получить три граничных условия на критическом слое для внешнего решения. Первое из них означает, что перепад средней завихренности на квазистационарном КС равен перепаду среднего нормированного волнового потока импульса:

$$X_0(+0, \nu) - X_0(-0, \nu) = \Gamma_+ - \Gamma_- = \Xi(\nu), \quad (19)$$

где  $\Xi = [T_p(+\infty) - T_p(-\infty)]/\lambda$ . Второе и третье граничные условия  $V_0(\pm 0, \nu) = 0$  имеют простой физический смысл: в системе отсчёта, движущейся вместе с волной, средняя скорость потока в КС по определению равна нулю.

С учётом (15) и выражения для  $H$  граничные условия для скорости и завихренности на бесконечности имеют вид

$$V_0(\pm\infty, \nu) = H + s/2, \quad X_0(\pm\infty, \nu) = 1. \quad (20)$$

Рассмотрим примеры точных и приближённых решений полученных уравнений.

**1.2.2. Автомодельная деформация среднего течения в диффузионном пограничном слое.** Пример точного решения. В простейшем случае постоянного скачка потока импульса при переходе

через КС можно найти автомодельное решение уравнения (18). В этом случае  $X_0$  и  $V_0$  не зависят от  $\nu$ , и для  $X_0$  получим обыкновенное дифференциальное уравнение второго порядка:

$$\frac{d^2 X_0}{dH^2} + 2 \left( H + \frac{s}{2} \right) \frac{dX_0}{dH} = 0, \quad (21)$$

решение которого, удовлетворяющее граничным условиям (20), имеет вид

$$X_0(H) = 1 + C_{\pm} \int_{\pm\infty}^{H+s/2} e^{-H_1^2} dH_1. \quad (22)$$

Интегрируя (22) по  $H$  с учётом граничных условий (20) на бесконечности имеем

$$V_0(H) = H + \frac{s}{2} + C_{\pm} \left[ \left( H + \frac{s}{2} \right) \int_{\pm\infty}^{H+s/2} e^{-H_1^2} dH_1 + \frac{e^{-(H+s/2)^2}}{2} \right]. \quad (23)$$

Используя граничные условия на критическом слое, можно найти завихренности  $\Gamma_+$  и  $\Gamma_-$  по разные стороны от него:

$$X_0 = \begin{cases} \Gamma_+ = \frac{e^{-s^2/4}}{e^{-s^2/4} - sJ}, & H = +0; \\ \Gamma_- = \frac{e^{-s^2/4}}{e^{-s^2/4} + s(\sqrt{\pi} - J)}, & H = -0. \end{cases} \quad (24)$$

Уравнение, связывающее скачок завихренности  $\Xi = \Gamma_+ - \Gamma_-$  и  $s$ , имеет вид

$$\sqrt{\pi} e^{-s^2/4} s = \Xi \left( e^{-s^2/4} - sJ \right) \left[ e^{-s^2/4} + s(\sqrt{\pi} - J) \right]. \quad (25)$$

Зависимость  $s(\Xi)$  изображена на рис. 5а, а зависимости  $\Gamma_{\pm}(\Xi)$  — на рис. 5б.

При малых и больших  $\Xi$  можно найти достаточно простые выражения для  $s(\Xi)$  и  $\Gamma_{\pm}(\Xi)$ . Так, при  $\Xi \ll 1$  указанные зависимости имеют вид  $s = \Xi/\sqrt{\pi}$ ,  $\Gamma_{\pm} = 1 \pm \Xi/2$ , а при  $\Xi \gg 1$  —  $s = \sqrt{2\Xi}$ ,  $\Gamma_+ = \Xi$ ,  $\Gamma_- = e^{-\Xi/2}/\sqrt{2\pi\Xi}$ .

Зависимость  $V_0(H)$  при нескольких значениях  $\Xi$  изображена на рис. 6, а на рис. 7 показана временная эволюция профилей  $u_0(\eta, \tau)$ , соответствующая зависимостям (23) при одном значении  $\Xi$ . Видно, что происходит диффузия завихренности из КС и его смещение навстречу падающей волне.

Таким образом, в случае постоянного перепада потока импульса на КС мы полностью ответили на вопросы, сформулированные в начале этого параграфа: 1) как может реализоваться нелокальная деформация профиля скорости, в которой возмущения растут при удалении от КС (ДПС и КС, смещающийся навстречу падающей волне); 2) чему равны значения завихренности по разные стороны от КС (формула (25)).

**1.2.3. Особенности нелокальной деформации среднего профиля скорости при излучении волн конечной амплитуды в потоках с резонансными слоями.** Рассмотрим теперь более реалистичный источник волнового возмущения, например волнистую поверхность во внешней области течения. Тогда легко видеть, что потоки импульса по обе стороны от КС, а также их перепад на КС будут зависеть от времени. Действительно, волна, излучаемая источником во внешней области течения, распространяется по направлению к ДПС, частично отражается от неоднородного профиля завихренности и частично проходит. Затем прошедшая волна взаимодействует с КС: частично отражается, частично проходит, частично поглощается. После этого волна, прошедшая через КС, взаимодействует

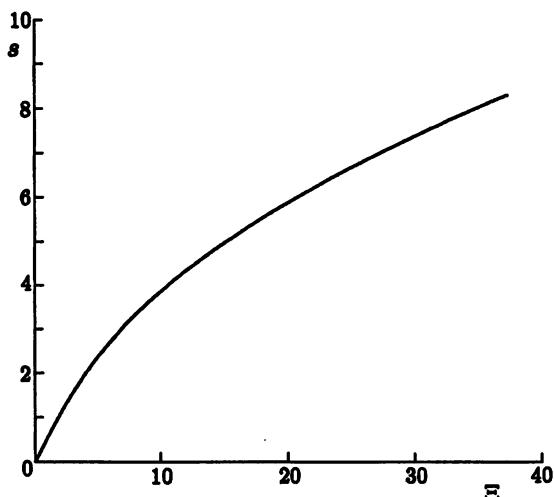


Рис. 5а. Зависимость константы в законе движения КС от скачка завихренности

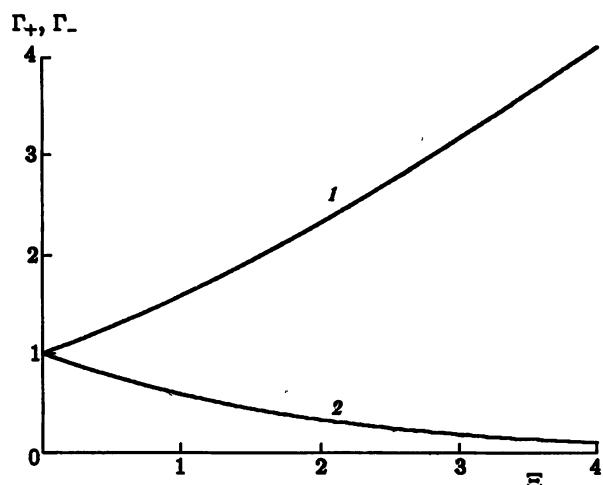


Рис. 5б. Зависимость средней завихренности по разные стороны от КС от  $\Xi$ . Кривая 1 — средняя завихренность  $\Gamma_+$  со стороны падения волны; 2 — средняя завихренность  $\Gamma_-$  со стороны прошедшей волны

с ДПС ниже КС. Поскольку профиль скорости в ДПС зависит от времени, то коэффициент отражения также зависит от времени. Это означает, что волновой поток импульса, который равен алгебраической сумме потоков в падающей и отражённой волнах, зависит от времени. В этом случае не удается получить простого автомодельного профиля скорости в ДПС, и необходимо решать уравнение (18) вместе с уравнением для волновых возмущений в ДПС. При этом невозможно получить решение задачи в общем виде, но в некотором достаточно широком диапазоне параметров потока может быть найдено приближённое решение.

Во-первых, если число Ричардсона не близко к  $1/4$ , то амплитуда волны, прошедшей через КС, очень мала, поэтому потоком импульса в прошедшей волне ( $T_p(-\infty)$ ) можно пренебречь и считать, что перепад волнового потока импульса приблизительно равен  $T_p(+\infty)$ , т. е.  $\Xi = T_p(+\infty)/\lambda$ . Во-вторых, поскольку зависимость волнового потока импульса от времени возникает из-за отражения волны от зависящего от времени профиля скорости в ДПС, то можно ожидать, что зависящая от времени компонента волнового потока импульса мала, когда мал коэффициент отражения. Как показано в [11] (см. также рис. 3), коэффициент отражения волны от КС и ДПС  $R < 0,1$  для умеренных чисел Ричардсона (не близких к  $1/4$ ) и  $\lambda > 0,2$ . Тогда решение можно искать в виде ряда по степеням  $R$ . Такая процедура подробно описана в [24]. В нулевом порядке по  $R$  перепад завихренности на КС не зависит от времени. При этом поля скорости в ДПС описываются автомодельными зависимостями (23). Для нахождения  $R$  необходимо вычислить коэффициент отражения волны от неоднородного профиля завихренности в ДПС  $r$  и коэффициент отражения от КС  $R$ . Комплексная величина  $r$  определялась в [24] на основании рассмотрения волновых полей в ДПС, а процедура нахождения коэффициента  $R$  описана выше.

Полученный в результате коэффициент отражения имеет вид суммы двух слагаемых:

$$R = R \cos[2\mu \ln \tau + \phi_1] + r \cos[2(\mu - \mu_+) \ln \tau + \phi_2].$$

Аналогичный вид имеют в первом порядке по  $R$  все средние характеристики течения в окрестности КС: перепад завихренности, её значения по разные стороны от КС, закон смещения КС и т. п.

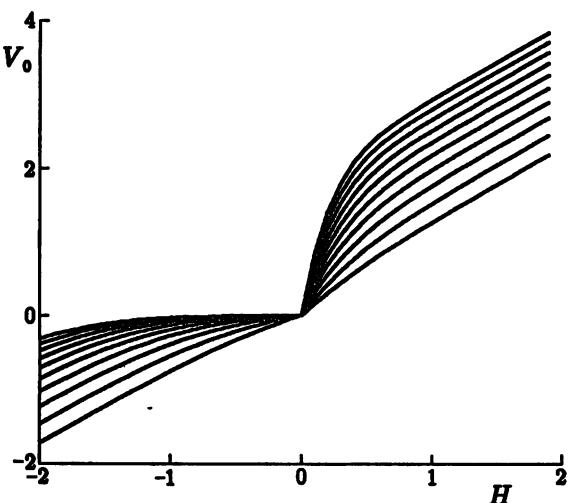


Рис. 6. Профиль средней скорости в ДПС. Нормированный скачок завихренности принимает значения от 1 (нижняя кривая) до 10 (верхняя кривая)

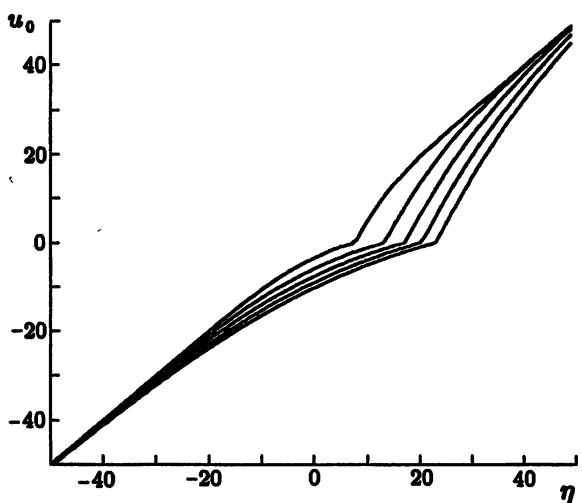


Рис. 7. Средние профили скорости в ДПС для моментов времени  $\tau = 50, 100, 150, 200$  и  $250$  (соответствующие кривые отчитываются слева направо). Скачок завихренности  $\Xi = 2$

## 2. О ДЕФОРМАЦИИ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ ЭЛЕКТРОНОВ ПРИ НЕЛИНЕЙНОМ ЗАТУХАНИИ ЛАНДАУ ЛЕНГМЮРОВСКОЙ ВОЛНЫ В СЛАБОСТОЛКНОВИТЕЛЬНОЙ ПЛАЗМЕ

Хорошо известно, что резонансное взаимодействие волн с частицами в плазме приводит к деформации функции распределения  $f$ . Рассмотрим, как в этом случае проявляются нелокальные эффекты, обусловленные диффузией в пространстве скоростей, вызванной столкновениями частиц. При этом будем использовать интеграл столкновений в форме Фоккера—Планка. Подобная модель была использована в [31] и обоснована в [32].

Рассмотрим в простейшей постановке задачу о распространении продольной волны пространственного заряда в плазме. Пусть потенциал электрического поля

$$\tilde{\Phi} = \tilde{\Phi}_0 \cos[k(x - ct)]. \quad (26)$$

Тогда кинетическое уравнение имеет вид [32]

$$\frac{\partial f}{\partial t} + v \frac{\partial f}{\partial x} - \frac{e}{m} \frac{\partial \tilde{\Phi}}{\partial x} \frac{\partial f}{\partial v} = \frac{3}{2\tau_D} \frac{\partial}{\partial v} \left( \frac{\partial f}{\partial v} v_T^2 + v f \right), \quad (27)$$

где  $\tau_D = m^2 c^3 / (8\pi e^4 n L)$ ,  $v_T$  — тепловая скорость электронов,  $e$  и  $m$  — заряд и масса электрона,  $L$  — кулоновский логарифм,  $n$  — концентрация плазмы,  $c$  — фазовая скорость ленгмюровских волн с дисперсионным соотношением  $\omega^2 = \omega_0^2 + 3v_T^2 k^2$ ,  $\omega_0^2$  — электронная плазменная частота.

Как показано в [32], интеграл столкновений, входящий в (27), справедлив в достаточно узкой области в пространстве скоростей вдали от тепловой скорости. В общем случае, строго говоря, необходимо использовать интеграл столкновений Ландау [1]. Однако здесь для простоты будет принята модель, в которой интеграл Фоккера—Планка используется во всём пространстве скоростей (аналогично [31]) как наиболее простое выражение, удовлетворяющее требованиям размерности и сохраняющее основные черты интеграла столкновений Ландау [1]: дивергентную форму и свойства диффузии в пространстве скоростей.

Основной задачей настоящего параграфа является анализ деформации функции распределения при резонансном взаимодействии волна—частица. При этом мы будем использовать квазистатическое приближение, справедливое, если декремент линейного затухания Ландау  $\gamma_L \ll 1/\tau_D$ . Тогда в системе отсчёта, сопровождающей волну, в безразмерных переменных

$$\tilde{V} = (v - c)/v_T, \quad \xi = kx$$

уравнение для функции распределения с учётом выражения (26) для потенциала электрического поля примет вид

$$\tilde{V} \frac{\partial f}{\partial \xi} + \epsilon \sin \xi \frac{\partial f}{\partial \tilde{V}} = \mu \frac{\partial}{\partial \tilde{V}} \left[ (C + \tilde{V}) f + \frac{\partial f}{\partial \tilde{V}} \right], \quad (28)$$

где  $\epsilon = e\tilde{\Phi}_0/(mv_T^2)$ ,  $\mu = 3/(2\tau_D k v_T^2)$ ,  $C = c/v_T$ .

Будем считать, что  $\{\epsilon, \mu\} \ll 1$ . В этом случае задача допускает решение методом сращиваемых асимптотических разложений. При этом существует глубокая аналогия между динамикой захваченных электронов и движением частиц в КС — плазменно-гидродинамическая аналогия [33, 34]. Причём в [35] была показана не просто качественная применимость плазменно-гидродинамической аналогии, но и доказана тождественность эволюционных уравнений, описывающих взаимодействие волна—поток в бесстолкновительной плазме и плоскопараллельном потоке невязкой жидкости.

Из этой аналогии следует, что при решении задачи можно воспользоваться схемой, предложенной в [9]. При этом надо искать решение уравнения для функции распределения в виде разложения по степеням  $\epsilon^{1/2}$ . В [9] было показано, что при учёте диффузии завихренности на КС формируются перепады её среднего значения порядка  $\epsilon^{1/2}$ . Аналогичные скачки функции распределения при переходе через область захваченных частиц возникают при учёте диффузии распределения электронов в пространстве скоростей, т. е. деформация является нелокальной. Вид функции распределения следует из решения внешней задачи. Используя условие нормировки, можно показать, что при  $C \gg 1$  с точностью до  $\epsilon^{1/2}$

$$f(\tilde{V}) = f_0 \left\{ \left[ 1 - \epsilon^{1/2} (F_+ - F_-) \frac{e^{-C^2/2}}{C} \right] + \epsilon^{1/2} \left[ F_- + \theta(\tilde{V}) e^{-(\tilde{V}+C)^2/2} (F_+ - F_-) \right] \right\}, \quad (29)$$

где  $\theta(\tilde{V})$  — единичная функция,

$$f_0 = \frac{n e^{-(\tilde{V}+C)^2/2}}{\sqrt{2\pi} v_T}$$

— равновесная функция распределения электронов (см. рис. 8). Видно, что деформация функции распределения имеет место не только в резонансной области, но и за её пределами, т. е. является нелокальной. Заметим, что распределение (29) характеризуется ненулевой средней скоростью в покоящейся системе отсчёта, которая пропорциональна скачку функции распределения. Действительно, средняя скорость

$$\langle v \rangle = \frac{v_T}{n} \int_{-\infty}^{+\infty} (\tilde{V} + C) f(\tilde{V}) d\tilde{V} = c (F_+ - F_-) \sqrt{\epsilon}. \quad (30)$$

Скачок функции распределения  $F_+ - F_-$  зависит от параметра  $\lambda = \mu/\epsilon^{3/2}$ , определяющего отношение эффективного периода столкновений электронов и периода колебаний захваченных электронов. Параметр  $\lambda$  полностью эквивалентен параметру нелинейности, имеющему смысл числа Рейнольдса внутри КС и определяющему соотношение нелинейных и диффузионных эффектов. Скачок функции распределения определяется параметром  $\lambda$ . Зависимость  $F_+ - F_-$  от  $\lambda$  представлена на рис. 9.

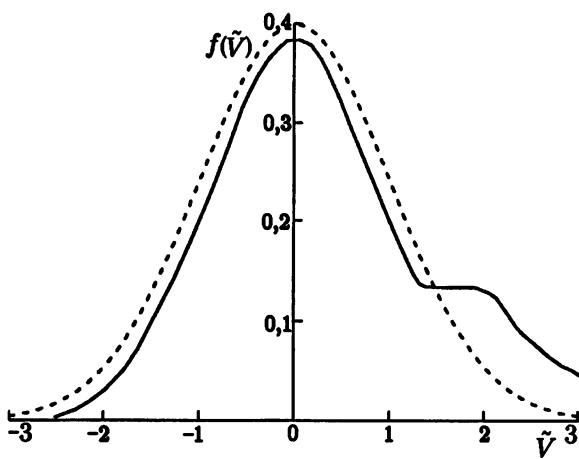


Рис. 8. Схема нелокальной деформации функции распределения (29) при учёте диффузии частиц в пространстве скоростей. Пунктиром показана равновесная функция распределения электронов  $f_0$

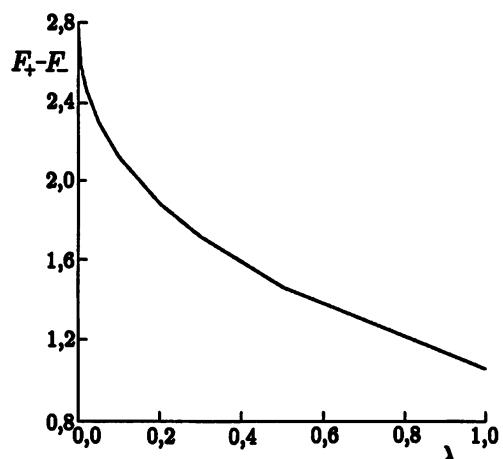


Рис. 9. Зависимость скачка функции распределения от  $\lambda$

Подобная нелокальная деформация функции распределения была найдена в работе [36] при численном решении двумерного кинетического уравнения, используемого для моделирования нагрева электронной компоненты плазмы под действием СВЧ излучения в условиях циклотронного резонанса. Ранее рассмотрение влияния столкновений на резонансное затухание плазменной волны было проведено в [32] и повторено в [37] в режиме сильно нелинейной волны, который в обозначениях этой статьи соответствует малым значениям параметра нелинейности  $\lambda$ . Здесь рассмотрен более общий случай произвольного  $\lambda$ , соответствующий произвольному соотношению эффектов конечной амплитуды волны и частоты столкновений. При этом случай малых  $\lambda$  здесь получен как предельный переход в задаче с конечным  $\lambda$ . Полученное решение принципиально отличается от решения [32] прежде всего наличием скачка средней функции распределения при переходе через критический уровень.

Скачок функции распределения определяет декремент нелинейного затухания Ландау  $\gamma$ , что с точностью до обозначений аналогично результатам [9]. Действительно, вычислим  $\gamma$  аналогично [32]. Очевидно, что при достаточно слабом затухании  $\gamma = \dot{W}/(2W)$ . Здесь  $\dot{W}$  — среднее на периоде волны затухание её энергии. В системе отсчёта, где волна покоятся,  $\dot{W} = c \langle \partial \tilde{\Phi} / \partial x \rho \rangle$ , где  $\rho = e \int_{-\infty}^{\infty} f dv$  — плотность заряда,  $W = \tilde{\Phi}_0^2 k^2 / (16\pi)$  — средняя плотность энергии волны. Отсюда сразу видно, что декремент  $\gamma$  пропорционален  $\int_{-\infty}^{\infty} \langle f \partial \tilde{\Phi} / \partial x \rangle dv$ . Усредним теперь по периоду волны кинетическое уравнение (27):

$$\left\langle \varepsilon \sin \xi \frac{\partial f}{\partial \tilde{V}} \right\rangle = \mu \frac{d}{d \tilde{V}} \left( (C + \tilde{V}) \langle f \rangle + \frac{d \langle f \rangle}{d \tilde{V}} \right).$$

Аналогично процедуре вывода соотношения (8) проинтегрируем это уравнение сначала по  $\tilde{V}$  от  $-\infty$  до текущего значения  $\tilde{V}$ , что приводит к уравнению

$$\varepsilon \langle f \sin \xi \rangle = \mu \left( (C + \tilde{V}) \langle f \rangle + \frac{d \langle f \rangle}{d \tilde{V}} \right).$$

Теперь проинтегрируем это уравнение по  $\tilde{V}$  от  $-\infty$  до  $+\infty$ , тогда с учётом граничных условий на бес-

конечности получим

$$\varepsilon \int_{-\infty}^{+\infty} \langle f \sin \xi \rangle d\tilde{V} = \mu \int_{-\infty}^{+\infty} (C + \tilde{V}) \langle f \rangle d\tilde{V}. \quad (31)$$

В левой части (31) стоит величина, пропорциональная  $\gamma$ . Очевидно, она отлична от нуля только в случае асимметричной функции распределения. Сравнивая (31) и (30), легко видеть, что денкремент  $\gamma$  пропорционален скачку функции распределения  $F_+ - F_-$ .

После преобразований, аналогичных [32], можно получить выражение для  $\gamma$ :

$$\gamma = \frac{4}{\pi} \lambda (F_+ - F_-) \gamma_L, \quad \gamma_L = \sqrt{\frac{\pi}{8}} \omega_0 \frac{c^2}{v_T^2} e^{-c^2/(2v_T^2)}. \quad (32)$$

Будем называть функцию  $\Phi_L(\lambda) = 4\lambda(F_+ - F_-)$  скачком фазы. Происхождение этого термина связано с тем, что при получении формулы (32) можно формально положить

$$\frac{1}{\omega - kv} = P \frac{1}{\omega - kv} + i\Phi_L(\lambda)\delta(\omega - kv),$$

где  $\delta(\omega - kv)$  — дельта-функция.

Зависимость  $\Phi_L(\lambda)$  показана на рис. 10. Видно, что при  $\lambda \rightarrow \infty$  (линейный режим)  $\Phi_L \rightarrow \pi$ , а при  $\lambda \rightarrow 0$  (нелинейный режим)  $\Phi_L \sim \lambda$ , поскольку при этом, как следует из рис. 9,  $(F_+ - F_-) \rightarrow \text{const}$ . Заметим, что  $\lambda \sim \varepsilon^{-3/2}$ , т. е. в слабодиссипативном пределе

$$\gamma \sim \varepsilon^{-3/2}.$$

Этот результат согласуется с полученным в работе [32], хотя численный коэффициент отличается по величине. При  $\lambda \gg 1$  можно показать [38], что  $\gamma$  стремится к постоянному значению, равному декременту линейного затухания Ландау.

Итак, главный результат этого раздела состоит в демонстрации нелокального эффекта деформации функции распределения за счёт диффузии в пространстве скоростей, обусловленной столкновениями частиц. Кроме того, получен закон затухания волны конечной амплитуды при произвольном соотношении частоты столкновений и периода колебаний резонансных электронов, захваченных электрическим полем продольной волны пространственного заряда. Подобное взаимодействие полностью эквивалентно взаимодействию волн с однородным сдвиговым потоком в окрестности КС, в котором соотношение нелинейности и диссипации произвольно.

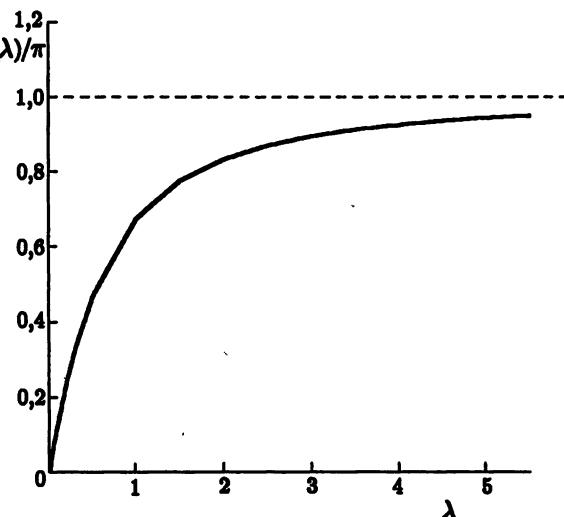


Рис. 10. Зависимость «скакка фазы» от параметра нелинейности

## ВЫВОДЫ

В настоящей работе рассмотрено взаимодействие волн с потоками в случаях, когда имеется узкая резонансная область, в которой скорости частиц потока близки к фазовой скорости волны. В такой области происходит существенное возрастание амплитуд и градиентов волновых полей, в результате чего становятся существенными нелинейные и диффузионные (диссипативные) эффекты. Здесь рассмотрены случаи, когда соотношение нелинейных и диффузионных (диссипативных) эффектов произвольно, т. е. волновые поля внутри резонансной области являются, вообще говоря, сильно нелинейными. Напротив, вне резонансной области нелинейность и диссипация пренебрежимо малы. В этом

случае для описания волновых полей может быть использован метод сращиваемых асимптотических разложений. Появление резонансных областей означает возникновение неравновесности распределения частиц потока, усреднённого по волновым возмущениям, а диффузионные процессы приводят к нелокальному эффекту деформации всего распределения частиц (а не только его узкой резонансной области).

В качестве примера, демонстрирующего особенности таких процессов, подробно исследовано нелинейно-диссипативное резонансное взаимодействие внутренних гравитационных волн малой амплитуды  $\epsilon$  с потоками жидкости или газа, в которых скорость и плотность неоднородны по вертикали, в окрестности критического слоя (КС), где фазовая скорость волны совпадает со скоростью потока. Рассмотрен случай динамически устойчивого потока, в котором градиентное число Ричардсона  $Ri > 1/4$ .

Показано, что за счёт нелинейного взаимодействия внутренней гравитационной волны с течением при переходе через КС в нулевом порядке по амплитуде волны возникает скачок средней завихренности потока. Происхождение этого скачка можно пояснить следующим образом. За счёт поглощения волн в КС средний волновой поток импульса испытывает скачок, т. е. в КС действует горизонтальная радиационная сила, которая в стационарном случае должна быть уравновешена вязкой силой. При этом устанавливается течение, в котором асимптотические значения завихренности при удалении вверх и вниз от КС постоянны, но различны. Скачок градиента плотности при переходе через КС отсутствует, т. к. отсутствует вертикальный поток массы во внутренних волнах. Таким образом, числа Ричардсона по разные стороны от КС различны. Кроме того, на КС существуют скачки средних скорости и плотности порядка  $\epsilon^{2/3}$ . Скачок завихренности зависит от  $\lambda$  (обратного значения вертикального числа Рейнольдса внутри КС) и растёт при стремлении  $\lambda$  к нулю, т. е. с ростом амплитуды падающей волны. Волна, падающая на КС, испытывает отражение и частично проходит. При  $\lambda \rightarrow 0$  средняя завихренность со стороны падения волны стремится к значению, при котором  $Ri = 1/4$ , коэффициент отражения стремится к минус единице, а коэффициент прохождения — к нулю.

Возникновение скачка завихренности (или излома профиля скорости) в окрестности КС означает, что возмущение средней скорости растёт при удалении от КС, но в силу принципа причинности на большом расстоянии от КС эти возмущения должны стремиться к нулю. Течение, удовлетворяющее этим условиям, реализуется при учёте нестационарного процесса диффузии завихренности из критического слоя за счёт вязкости. При этом формируется переходная область от КС к невозмущенному течению — диффузионный пограничный слой (ДПС), а область резонансного взаимодействия волны с потоком (КС) смещается по направлению к падающей волне.

Рассмотрена нелокальная деформация функции распределения электронов в плазме с учётом столкновений. Показано, что учёт диффузии распределения частиц в пространстве скоростей приводит к деформации функции распределения не только в резонансной области, но и во всем пространстве скоростей. Найдена форма деформированной функции распределения и вычислен декремент нелинейного затухания Ландау продольной волны пространственного заряда конечной амплитуды при произвольном соотношении частот столкновений и колебаний захваченных электронов.

Работа выполнена при финансовой поддержке РФФИ (проекты № 00–15–99399, 00–15–96772).

#### СПИСОК ЛИТЕРАТУРЫ

- Лифшиц Е. М., Питаевский Л. П. Физическая кинетика. М.: Наука, 1981.
- Басович А. Я., Таланов В. И. // Нелинейные волны. Самоорганизация. 1985. С. 147.
- Маслоу С. А. // Гидродинамические неустойчивости и переход к турбулентности / Под ред. Х. Сунни, Дж. Голлаба. М.: Мир, 1981. С. 218.
- Craik A. D. D. Wave interactions and fluid flows. Cambridge University Press, 1985.

5. Триттон Д. Дж., Девис П. А. // Гидродинамические неустойчивости и переход к турбулентности / Под ред. Х. Суинни, Дж. Голлаба. М.: Мир, 1981. С. 271.
6. Maslowe S. A. // Ann. Rev. Fluid Mech. 1986. V. 18. P. 405.
7. Шухман И. Г. Вопросы теории нелинейных и диссипативных процессов в динамике гравитирующих систем, жидкости и плазмы: Дисс. ... д. ф.-м. н. Иркутск, 1986.
8. Ван Дайк М. Теория возмущений в механике жидкости. М.: Мир, 1967.
9. Haberman R. // Stud. Appl. Math. 1972. V. 51, No. 2. P. 139.
10. Haberman R. // J. Fluid Mech. 1973. V. 58. P. 727.
11. Troitskaya Yu. I. // J. Fluid Mech. 1991. V. 233. P. 25.
12. Koppel D. // J. Math. Phys. 1964. V. 5. P. 963.
13. Hazel P. // J. Fluid Mech. 1967. V. 30. P. 775.
14. Baldwin P., Roberts P. H. // Mathematics. 1970. V. 17. P. 102.
15. Bowman M. R., Thomas L., Thomas R. H. // Planet. Space Sci. 1980. V. 28. P. 119.
16. Van Duin C. A., Kelder H. // J. Fluid Mech. 1986. V. 169. P. 293.
17. Kelly R. E., Maslowe S. A. // Stud. Appl. Math. 1970. V. 49. P. 302.
18. Maslowe S. A. // Stud. Appl. Math. 1972. V. 51, No. 1. P. 1.
19. Maslowe S. A. // Boundary Layer Meteorol. 1973. V. 5. P. 45.
20. Stewartson K. // J. Appl. Math. 1981. V. 27. P. 133.
21. Booker J. R., Bretherton F. P. // J. Fluid Mech. 1967. V. 27. P. 513.
22. Miles J. W. // J. Fluid Mech. 1961. V. 10, pt. 4. P. 496.
23. Hovard L. N. // J. Fluid Mech. 1961. V. 10, pt. 4. P. 509.
24. Troitskaya Yu. I., Reznik S. N. // Phys. Fluids. 1996. V. 8, No. 12. P. 3313.
25. Ле Блон П., Майсек Л. Волны в океане. Т. 1, 2. М.: Мир, 1981.
26. Geller M., Tanaka, Fritts D. // J. Atm. Sci. 1976. V. 33. P. 2276.
27. Fritts D. // J. Atm. Sci. 1978. V. 35. P. 397.
28. Fritts D. // J. Atm. Sci. 1979. V. 36. P. 12.
29. Koop G. // J. Fluid Mech. 1981. V. 113. P. 347.
30. Koop G., McGee B. // J. Fluid Mech. 1986. V. 172. P. 453.
31. Lennard A., Bernstein I. // Phys. Rev. 1957. V. 108. P. 546.
32. Захаров В. Е., Карпман В. И. // ЖЭТФ. 1962. Т. 43, № 2. С. 490.
33. Михайловский А. Б. Теория плазменных неустойчивостей. Т. 1, 2. М.: Атомиздат, 1975.
34. Андронов А. А., Фабрикант А. Л. // Нелинейные волны / Под ред. А. В. Гапонова-Грекова). М.: Наука, 1979. С. 68.
35. Реутов В. П. // Изв. АН СССР. ФАО. 1980. Т. 16, № 12. С. 1266.
36. Суворов Е. В., Шалашов А. Г. Численное решение двумерного кинетического уравнения для моделирования ЭЦР нагрева: Препринт ИПФ РАН № 462. Н. Новгород, 1998. 26 с.
37. Галеев А. А., Сагдеев Р. З. // Вопросы теории плазмы. Вып. 7 М.: Энергоатомиздат, 1973, с. 3.
38. Троицкая Ю. И. Диффузионные и радиационные эффекты при нелинейном резонансном взаимодействии волн с потоком: Дисс. ... д. ф.-м. н., Н. Новгород, 1998.

Институт прикладной физики РАН,  
г. Нижний Новгород, Россия

Поступила в редакцию  
6 апреля 2001 г.

**ASYMPTOTIC MODELS OF DIFFUSION EFFECTS DUE TO NONLINEAR RESONANCE INTERACTION OF WAVES  
WITH FLOWS**

*Yu. I. Troitskaya*

We develop an asymptotic theory describing nonlocal effects, caused by weak-diffusion processes, in the case of a resonance interaction of quasi-harmonic waves of small but finite amplitudes with flows of various physical nature under the conditions of an arbitrary relation between the nonlinearity and diffusion.

We analyze the interaction of inner gravity waves with plane-parallel stratified shear flows in the nonlinearly-dissipative critical layer (KL) formed in the vicinity of the resonance level where the flow velocity is equal to the phase velocity of the wave. It is shown that the combined effect of the radiation force in the inner region of the KL and vorticity diffusion to the outer region results in the formation of a flow in which the asymptotic values of average vorticity at different sides of the KL are constant but different. If the criterion of the linear dynamic stability is satisfied (the Richardson number  $Ri > 1/4$ ), the resulting vorticity steps are comparable to the unperturbed vorticity. As a result, a wave reflected from the inhomogeneity of the vorticity in the KL is formed. The average vorticity from the incidence side approaches the linear-stability threshold (Richardson number  $Ri > 1/4$ ) and the reflection coefficient, to  $-1$  with increasing amplitude of the incident wave.

In the regime of nonlinear dissipative KL, we study quasi-stationary asymptotic behavior of the flow formed by an inner gravity wave incident on a dynamically stable flow with velocity and density stratification, whose velocity at some level is equal to the phase velocity of the wave. It is shown that the vorticity diffusion results in the formation of a nonlocal transition region from the KL to the unperturbed flow, which is called the diffusive boundary layer (DBL). In this case, the KL is shifted toward the incident wave. We obtain a self-similar solution for the average fields, which is valid in the case of a constant vorticity step in the KL, and determine its parameters depending on the inner Reynolds number in the KL, which describes the relation between the nonlinear and diffusive effects for the wave field in the resonance region. We determine the structure and temporal dynamics of the DBL formed by a rough surface streamlined by a stratified fluid whose velocity changes direction at some level.

It is shown that in the case of the nonlinear resonance interaction of plasma electrons with a Langmuir wave, the electron diffusion in the velocity space leads to a significant nonlocal distortion of the electron distribution function outside the region of trapping. We determine the distorted distribution function and calculate the decay rate of the nonlinear Landau damping of a finite-amplitude wave for an arbitrary ratio of the electron collision rate and the oscillation period of the trapped electrons.

УДК 517.958:537.874

# ГЕНЕРАЦИЯ ПРОСТРАНСТВЕННО-КОГЕРЕНТНОГО ИЗЛУЧЕНИЯ В ЛАЗЕРАХ НА СВОБОДНЫХ ЭЛЕКТРОНАХ С ДВУМЕРНОЙ РАСПРЕДЕЛЁННОЙ ОБРАТНОЙ СВЯЗЬЮ

*Н. С. Гинзбург<sup>1</sup>, Н. Ю. Песков<sup>1</sup>, А. С. Сергеев<sup>1</sup>,  
А. В. Аржанников<sup>2</sup>, С. Л. Синицкий<sup>2</sup>*

Исследована возможность использования двумерной распределённой обратной связи для получения пространственно-когерентного излучения релятивистских электронных потоков ленточной и трубчатой геометрии, поперечные размеры которых на несколько порядков превышают длину волн. Такая обратная связь может быть реализована в планарных и коаксиальных двумерных брэгговских резонаторах, имеющих двояко-периодическую гофрировку боковых стенок. Возникающие на подобной гофрировке дополнительные поперечные (по отношению к поступательному движению электронов) потоки электромагнитной энергии синхронизуют излучение отдельных частей электронного пучка. Моделирование процесса установления автоколебаний в лазерах на свободных электронах (ЛСЭ) с двумерной распределённой обратной связью показывает возможность получения одномодовой одночастотной когерентной генерации от пучков с поперечными размерами, достигающими  $10^2 \div 10^3$  длин волн. Проведён также анализ использования гибридных резонаторов, составленных из входного «двумерного» и выходного «одномерного» брэгговских зеркал. В такой схеме «двумерное» зеркало обеспечивает синхронизацию излучения поперёк электронного потока, а отражений от выходного «одномерного» зеркала оказывается достаточно для самовозбуждения генератора. В случае закрытых в поперечном направлении систем подобная схема позволяет существенно снизить омические потери, обусловленные запертыми в поперечном направлении электромагнитными потоками. Показано, что двумерная распределённая обратная связь может быть использована также для синхронизации излучения в многопучковом генераторе, представляющем собой совокупность ЛСЭ-модулей планарной геометрии, каждый из которых запитывается ленточным релятивистским электронным пучком, а связь различных модулей осуществляется через возникающие в двумерных брэгговских структурах поперечные потоки электромагнитной энергии. Обсуждаются экспериментальные исследования, направленные на реализацию сверхмощных ЛСЭ с двумерной распределённой обратной связью.

## ВВЕДЕНИЕ

В настоящее время достаточно широкое распространение получили лазеры на свободных электронах (ЛСЭ) с брэгговскими резонаторами, выполненными в виде отрезков волноводов с однопериодической слабой гофрировкой стенок («одномерные» брэгговские резонаторы). Такие ЛСЭ были предложены в [1, 2] и, фактически, явились аналогом лазеров с распределённой обратной связью [3]. Достоинством брэгговских резонаторов в микроволновом диапазоне является, во-первых, их совместимость с системами транспортировки сильноточных релятивистских электронных пучков (РЭП), во-вторых, возможность обеспечения обратной связи для волн, распространяющихся под малым углом к электронному потоку (что является необходимым требованием для обеспечения в ЛСЭ большого доплеровского преобразования частоты) и, в-третьих, высокая селективность обратной связи, которая имеет место только в узком частотном интервале вблизи брэгговской частоты. Использование брэгговских резонаторов позволило экспериментально реализовать узкополосные ЛСЭ-генераторы миллиметрового диапазона длин волн [4–9].

Следует, однако, отметить, что во всех проведённых экспериментах поперечные размеры брэгговских резонаторов, реализующих одномерную распределённую обратную связь (РОС), составляли не более нескольких длин волн излучения, а генерируемая мощность не превышала 50 МВт. Дальнейшее увеличение поперечных размеров таких электродинамических систем сопряжено с потерей их

селективности. Вместе с тем увеличение мощности СВЧ генераторов требует перехода к существенно более сверхразмерным электродинамическим системам. В частности, для получения мощного излучения в миллиметровом диапазоне длин волн представляется привлекательным использовать существующие в настоящий момент РЭП трубчатой [10] и ленточной [11, 12] конфигураций с характерными поперечными размерами до  $10^2$  см, запас энергии в которых достигает  $10^2 \div 10^3$  кДж. Очевидно, что для указанного диапазона длин волн ширина электронного потока и, соответственно, электродинамической системы на несколько порядков превосходят длину волны, и на первый план выступает проблема синхронизации излучения от различных частей электронного потока.

Для решения проблемы пространственной синхронизации излучения в [13, 14] было предложено использование двумерной РОС, которая может быть реализована в «двумерных» брэгговских резонаторах планарной и коаксиальной геометрии с двоякопериодической гофрировкой стенок. На подобной гофрировке возникает взаимное рассеяние четырёх парциальных волн, две из которых распространяются в направлении электронного потока и навстречу ему (как в одномерных аналогах), а две других — в перпендикулярных ему направлениях. Благодаря последним оказывается возможной синхронизация излучения различных частей электронного потока.

Настоящая работа посвящена теоретическим исследованиям ЛСЭ с новым механизмом обратной связи. В разделе 1 описана базирующаяся на методе связанных волн модель двумерной брэгговской структуры. Анализ собственных мод двумерных брэгговских резонаторов в отсутствие электронных потоков демонстрирует их высокую селективность как по продольному, так и по поперечному (азимутальному) индексу. Процесс установления автоколебаний в планарном и коаксиальном ЛСЭ-генераторах с двумерной РОС описан в разделах 2 и 3 соответственно, где показано, что в идеальном случае рассматриваемая схема позволяет практически неограниченно увеличивать поперечные размеры системы, а следовательно, и интегральную мощность излучения. В разделе 4 рассмотрена комбинированная схема двухзеркального резонатора, составленного из «одномерного» и «двумерного» брэгговских зеркал. В этой схеме расположение на катодном конце «двумерное» зеркало обеспечивает синхронизацию излучения поперёк электронного потока, а выходное традиционное «одномерное» зеркало замыкает кольцо обратной связи. Подобная комбинированная схема представляется оптимальной для закрытых в поперечном направлении систем как планарной, так и коаксиальной геометрии. С помощью такой схемы удаётся снизить до оптимального уровня поперечные потоки энергии, уменьшив тем самым напряжённость поля внутри резонатора, а также омические потери. В разделе 5 показано, что двумерная РОС может быть также использована для получения пространственно-когерентного излучения в системах, развитых по трём пространственным координатам в виде многослойных структур. Практическая реализация такой схемы представляет собой многопучковый генератор, который состоит из нескольких планарных секций (модулей), синхронизуемых посредством двумерной РОС. Каждый модуль запитывается ленточным РЭП, связь различных модулей осуществляется через возникающие в двумерных брэгговских структурах поперечные потоки электромагнитной энергии. В заключительном разделе 6 обсуждаются экспериментальные исследования ЛСЭ с двумерной РОС планарной геометрии, проводимые в ИЯФ СО РАН (Новосибирск), и коаксиальной геометрии, ведущиеся в Стратклайдском университете (Глазго, Великобритания).

## 1. СОБСТВЕННЫЕ МОДЫ ПЛАНАРНОГО ДВУМЕРНОГО БРЭГГОВСКОГО РЕЗОНАТОРА

Рассмотрим планарный двумерный брэгговский резонатор, образованный двумя металлическими пластинами ширины  $l_x$  и длины  $l_z$ , расположенными на расстоянии  $a_0$  (рис. 1a), на которые нанесена двоякопериодическая гофрировка

$$a = a_1 [\cos(\bar{h}x - \bar{h}z) + \cos(\bar{h}x + \bar{h}z)], \quad (1)$$

где  $2a_1$  — глубина гофра,  $\bar{h} = \sqrt{2}\pi/d$  — проекции векторов  $\bar{h}$  элементарных решёток (рис. 1б),

$d$  — период гофра. На этой решётке возникает взаимное рассеяние четырёх парциальных волн, распространяющихся вдоль осей  $z$  (волны  $A_{\pm}$ ) и  $x$  (волны  $B_{\pm}$ ):

$$\mathbf{E} = \mathbf{E}^0 \operatorname{Re}\{[A_+ \exp(-ihz) + A_- \exp(ihz) + B_+ \exp(-ihx) + B_- \exp(ihx)] \exp(i\omega t)\}, \quad (2)$$

где  $\mathbf{E}^0$  — функции, описывающие поперечные (по оси  $y$ ) распределения волн и совпадающие со структурой соответствующих мод невозмущённого планарного волновода, если волновые числа парциальных волн  $h$  удовлетворяют условию брэгговского резонанса (см. рис. 1б):

$$h \approx \bar{h}. \quad (3)$$

Рассеяние этих волн может быть описано с помощью уравнения связанных волн для медленно меняющихся амплитуд  $A_{\pm}(x, z)$  и  $B_{\pm}(x, z)$  [14–16]:

$$\begin{aligned} \frac{\partial A_{\pm}}{\partial z} \mp i\delta A_{\pm} \pm i\alpha (B_+ + B_-) &= 0, \\ \frac{\partial B_{\pm}}{\partial x} \mp i\delta B_{\pm} \pm i\alpha (A_+ + A_-) &= 0, \end{aligned} \quad (4)$$

где  $A_{\pm} = A_{\pm} \exp(\mp i\delta z)$ ,  $B_{\pm} = B_{\pm} \exp(\mp i\delta x)$ ,  $\delta = \bar{h} - h$  — отстройка от точного брэгговского резонанса,  $\alpha$  — коэффициент связи волн на двумерной брэгговской структуре, который для случая рассеяния основных ТЕМ-волн планарного волновода равен [15, 16]

$$\alpha = a_1 \bar{h} / (2a_0). \quad (5)$$

Заметим, что при получении уравнений (4) размеры брэгговского резонатора  $l_{x,z}$  и соответствующие размеры волновых пучков предполагались существенно превосходящими длину волны ( $l_{x,z} \gg \lambda$ ), что позволяет пренебречь дифракцией и использовать геометрооптическое приближение. Кроме того, глубина гофрировки предполагалась малой ( $a_1 \bar{h} \ll 1$ ), что допускает использование метода связанных волн.

При определении частот и добротностей собственных мод резонатора будем считать, что потоки электромагнитной энергии извне отсутствуют и имеет место идеальное согласование для парциальных волн при излучении из резонатора. Границные условия к уравнениям (4) при этом можно представить в виде

$$A_+(x, 0) = 0, \quad A_-(x, l_z) = 0; \quad B_+(0, z) = 0, \quad B_-(l_x, z) = 0. \quad (6)$$

Анализ уравнений (4) с граничными условиями (6) показывает (см. подробнее [15, 16]), что в условиях сильной связи волн ( $\alpha l_{x,z} \gg 1$ ) рассматриваемая система обладает дискретным спектром высокодобротных мод

$$Q_{n,m} \approx \frac{\bar{h}}{2 |\operatorname{Im} \delta_{n,m}|} \gg 1. \quad (7)$$

Собственные частоты этих мод

$$\omega_{n,m} \approx c(\bar{h} + \operatorname{Re} \delta_{n,m}), \quad (8)$$

где  $c$  — скорость света в вакууме, расположены в окрестности точного брэгговского резонанса  $\delta \approx 0$ , а также вблизи  $\delta \approx \pm 2\alpha$  (рис. 2). Соответствующие выражения для отстройки имеют вид

$$\delta_{n,m} = -\frac{\pi^2 mn}{2\alpha l_z l_x} - i \frac{\pi^2}{2\alpha^2 l_z l_x} \left( \frac{n^2}{l_z^2} + \frac{m^2}{l_x^2} \right) \quad (9a)$$

при  $\delta \approx 0$  и

$$\delta_{n,m} = \pm \left[ 2\alpha + \frac{\pi^2}{4\alpha} \left( \frac{n^2}{l_z^2} + \frac{m^2}{l_x^2} \right) \right] - i \frac{\pi^2}{2\alpha^2} \left( \frac{n^2}{l_z^3} + \frac{m^2}{l_x^3} \right) \quad (9b)$$

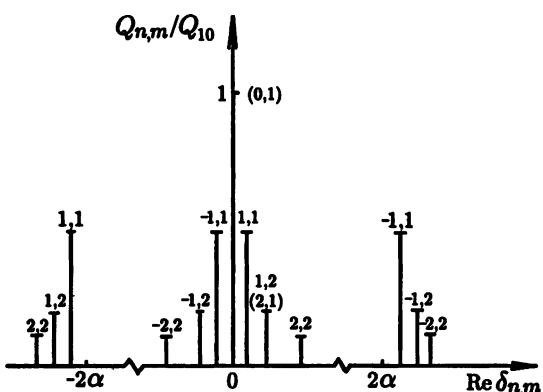


Рис. 2. Спектр частот и добротности собственных мод открытого брэгговского резонатора планарной геометрии ( $l_x = l_z$ )

при  $\delta \approx \pm 2\alpha$ . Здесь  $n$  и  $m$  — продольный и поперечный индексы мод соответственно. Максимальную добротность имеют моды с индексами  $n = 0$ ;  $m = 1$  и  $n = 1$ ;  $m = 0$  (рис. 2). Эти моды вырождены по собственной частоте ( $\operatorname{Re} \delta_{n,m} = 0$ ), а в случае  $l_x = l_z$  — и по добротности. На рис. 3 для моды  $n = 1$ ,  $m = 0$  приведена пространственная структура парциальных волн  $A_+$  и  $B_+$  (структура волны  $A_-$  идентична  $A_+$ , а  $B_-$  зеркально симметрична  $B_+$ ). Важно подчеркнуть, что для указанной моды амплитуда поля волны  $A_+$  (считаем, что волна  $A_+$  синхронна с электронным пучком, движущимся в направлении оси  $z$ , а остальные парциальные волны с пучком не взаимодействуют) не зависит от поперечной координаты  $x$ . Заметим также, что максимальные амплитуды волн  $A_{\pm}$  существенно превы-

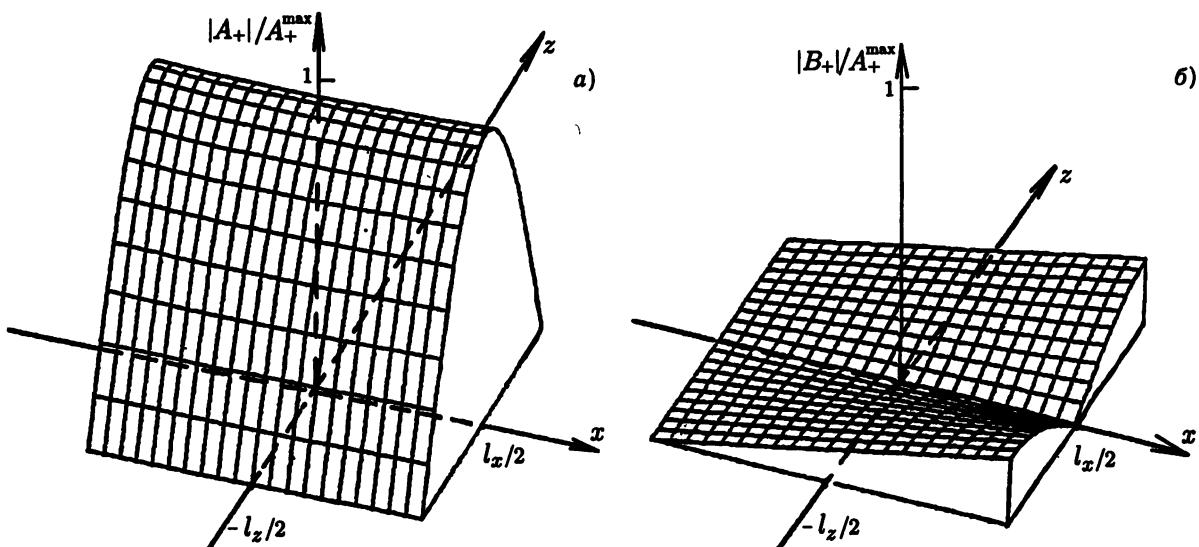


Рис. 3. Пространственные структуры парциальных волн  $A_+$  и  $B_+$  для наиболее высокодобротной моды резонатора  $n = 1$ ;  $m = 0$  ( $\alpha l_x = \alpha l_z = 5$ )

шают максимальные амплитуды волн  $B_{\pm}$ :  $A_{\pm}^{\max}/B_{\pm}^{\max} \approx \alpha l_z$  (см. рис. 3). Для моды  $n = 0; m = 1$  ситуация обратная, здесь пространственные структуры волн  $A_{\pm}$  и  $B_{\pm}$  совпадают соответственно со структурами волн  $B_{\pm}$  и  $A_{\pm}$  моды  $n = 1, m = 0$ , если в последних поменять местами координаты  $z$  и  $x$ . В результате эта мода имеет относительно малую амплитуду синхронной волны  $A_+$  ( $A_{\pm}^{\max}/B_{\pm}^{\max} \approx 1/(\alpha l_x)$ ) и существенно меньшую связь с электронным пучком, а поэтому должна менее эффективно им возбуждаться (что подтверждается численным моделированием в разделе 2).

Таким образом, двумерный брэгговский резонатор обладает достаточно высокой селективностью как по продольному ( $n$ ), так и по поперечному ( $m$ ) индексам, что обеспечивается за счёт вывода излучения не только в продольном направлении (параллельно оси  $z$ ), как в одномерных брэгговских резонаторах, но и дополнительно в поперечном направлении.

## 2. ПРОСТРАНСТВЕННАЯ СИНХРОНИЗАЦИЯ ИЗЛУЧЕНИЯ В ПЛАНАРНОМ ЛСЭ С ДВУМЕРНОЙ РАСПРЕДЕЛЁННОЙ ОБРАТНОЙ СВЯЗЬЮ

Рассмотрим процесс возбуждения двумерного брэгговского резонатора релятивистским ленточным электронным пучком. Будем считать, что только парциальная волна  $A_+$  синхронна с пучком электронов, движущихся в направлении оси  $z$  с поступательной скоростью  $v_{||}$  и осциллирующих в поле ондулятора с баунс-частотой  $\Omega_b = h_w v_{||}$ , где  $h_w = 2\pi/d_w$ ,  $d_w$  — период ондулятора. В условиях ондуляторного синхронизма

$$\omega - hv_{||} \approx \Omega_b \quad (9)$$

этая волна может усиливаться электронным потоком. На периодической решётке она рассеивается в волны  $B_{\pm}$ , распространяющиеся в поперечном направлении и синхронизующие (после обратной трансформации в волну  $A_+$ ) излучение различных частей электронного потока. Одновременно волны  $B_{\pm}$  рассеиваются в волну  $A_-$ , распространяющуюся навстречу поступательному движению частиц и замыкающую кольцо обратной связи.

Для исследования процесса возбуждения двумерного брэгговского резонатора широким ленточным электронным пучком будем полагать в (2) комплексные амплитуды волн  $A_{\pm}(x, z, t)$ ,  $B_{\pm}(x, z, t)$  медленно меняющимися во времени, а в качестве несущей частоты выберем брэгговскую частоту  $\bar{\omega} = \bar{h}c$ . В нестационарном приближении возбуждение ленточных РЭП парциальной волны  $A_+$  и последующее рассеяние этой волны на двумерной брэгговской структуре может быть описано с помощью следующей системы уравнений [16, 17]:

$$\left( \frac{\partial}{\partial Z} + \beta_{gr}^{-1} \frac{\partial}{\partial \tau} \right) \hat{A}_+ + \sigma \hat{A}_+ + i\hat{\alpha} (\hat{B}_+ + \hat{B}_-) = J, \quad J = \frac{1}{\pi} \int_0^{2\pi} \exp(-i\theta) d\theta_0, \quad (11a)$$

$$\left( -\frac{\partial}{\partial Z} + \beta_{gr}^{-1} \frac{\partial}{\partial \tau} \right) \hat{A}_- - \sigma \hat{A}_- + i\hat{\alpha} (\hat{B}_+ + \hat{B}_-) = 0, \quad (11b)$$

$$\left( \pm \frac{\partial}{\partial X} + \beta_{gr}^{-1} \frac{\partial}{\partial \tau} \right) \hat{B}_{\pm} \pm \sigma \hat{B}_{\pm} + i\hat{\alpha} (\hat{A}_+ + \hat{A}_-) = 0, \quad (11c)$$

$$\left( \frac{\partial}{\partial Z} + \beta_{||}^{-1} \frac{\partial}{\partial \tau} \right)^2 \theta = \operatorname{Re}(\hat{A}_+ \exp(i\theta)). \quad (11d)$$

Границные условия к уравнениям (11) для парциальных волн совпадают с (6), а для уравнений движения частиц первоначально стационарного моноэнергетического электронного потока имеют вид

$$\theta|_{Z=0} = \theta_0 \in [0, 2\pi), \quad \left( \frac{\partial}{\partial Z} + \beta_{||}^{-1} \frac{\partial}{\partial \tau} \right) \theta|_{Z=0} = -\Delta. \quad (12)$$

Здесь использованы следующие обозначения:  $Z = \bar{h}zC$ ,  $X = \bar{h}xC$ ,  $\tau = \bar{\omega}tC$ ,  $\hat{\alpha} = \alpha/(\bar{h}C)$ ,  $(\hat{A}_\pm, \hat{B}_\pm) = e\mu(A_\pm, B_\pm)/(mc\bar{\omega}\gamma_0 C^2)$ ,  $\Delta = (\bar{\omega} - \bar{h}v_{||} - h_w v_{||})/(\bar{\omega}C)$  — начальная расстройка синхронизма на несущей частоте,  $\theta = \bar{\omega}t - \bar{h}z - h_w z$  — фаза электрона относительно синхронной волны,

$$C = \left( \frac{eI_b \lambda^2 \kappa^2 \mu}{mc^3 8\pi \gamma_0 a_0} \right)^{1/3}$$

— параметр Пирса,  $\mu \approx \gamma^{-2}$  — параметр инерционной группировки [5],  $\gamma$  — релятивистский масс-фактор,  $\kappa = \beta_{\perp 0}/(2\beta_{||0})$  — коэффициент связи электронов с волной,  $\beta_{\perp,||} = v_{\perp,||}/c$ ,  $v_\perp$  — осцилляторная скорость частиц (индекс 0 соответствует начальным значениям величин при  $z = 0$ ),  $v_{gr} = \beta_{gr}c$  — групповая скорость парциальных волн,  $I_b$  — погонный ток пучка,  $\sigma$  — коэффициент омических потерь,  $e$  и  $m$  — элементарный заряд и масса электрона соответственно. Отметим, что учёт потерь является принципиально важным при описании закрытых в поперечном направлении систем, рассмотренных далее в разделах 3 и 4, в частности коаксиальных ЛСЭ с двумерной РОС. В то же время для открытых планарных систем омические потери не имеют существенного значения, и можно положить  $\sigma = 0$ .

Интегральный КПД определяется соотношениями

$$\eta = \frac{C\hat{\eta}}{\mu(1 - \gamma_0^{-1})}, \quad \hat{\eta} = \frac{1}{2\pi L_x} \int_0^{L_x} dX \int_0^{2\pi} \left( \frac{\partial \theta}{\partial Z} + \Delta \right) \Big|_{Z=L_x} d\theta_0,$$

где  $L_{x,z} = \bar{h}l_{x,z}C$ .

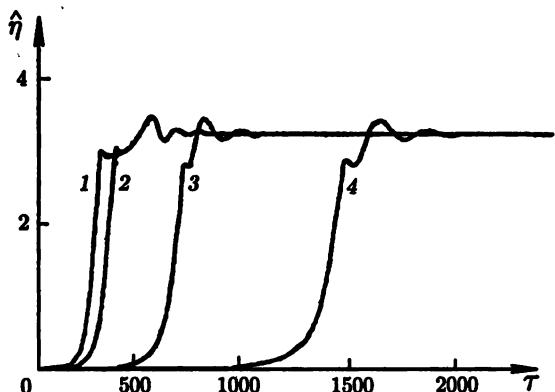


Рис. 4. Установление стационарного режима генерации в планарном ЛСЭ-генераторе с открытым двумерным брэгговским резонатором. Зависимость электронного КПД от времени при различной ширине системы ( $L_z = 4.0$ ;  $\Delta = -1.8$ ;  $\hat{\alpha}^2 L_z = 1.25$ ): кривая 1 —  $L_z = 0.8$ ;  $\hat{\alpha} = 1.25$ ; 2 —  $L_z = 3.2$ ;  $\hat{\alpha} = 0.63$ ; 3 —  $L_z = 12.8$ ;  $\hat{\alpha} = 0.32$ ; 4 —  $L_z = 28.8$ ;  $\hat{\alpha} = 0.21$

перечных размерах генератора, при которых может осуществляться режим синхронизации излучения ленточного РЭП. Согласно уравнениям (11) в стационарных условиях ( $\partial/\partial\tau = 0$ ) при возбуждении основной моды зависимость структуры парциальных волн от поперечной координаты может быть представлена в форме

$$\hat{B}_+ = \hat{\alpha}X(\hat{A}_+ + \hat{A}_-), \quad \hat{B}_- = \hat{\alpha}(L_x - X)(\hat{A}_+ + \hat{A}_-). \quad (13)$$

$$\hat{\alpha}^2 L_x L_z^4 \geq 250, \quad \Delta L_z \approx \pi.$$

Подчеркнём, что поперечное распределение амплитуды поля синхронной волны  $A_+$  не зависит от поперечной координаты  $i$ , следовательно, эффективность энергообмена с высокочастотным полем у всех частей электронного потока одинакова.

Принципиальным для рассматриваемой схемы ЛСЭ является вопрос о предельно допустимых по-

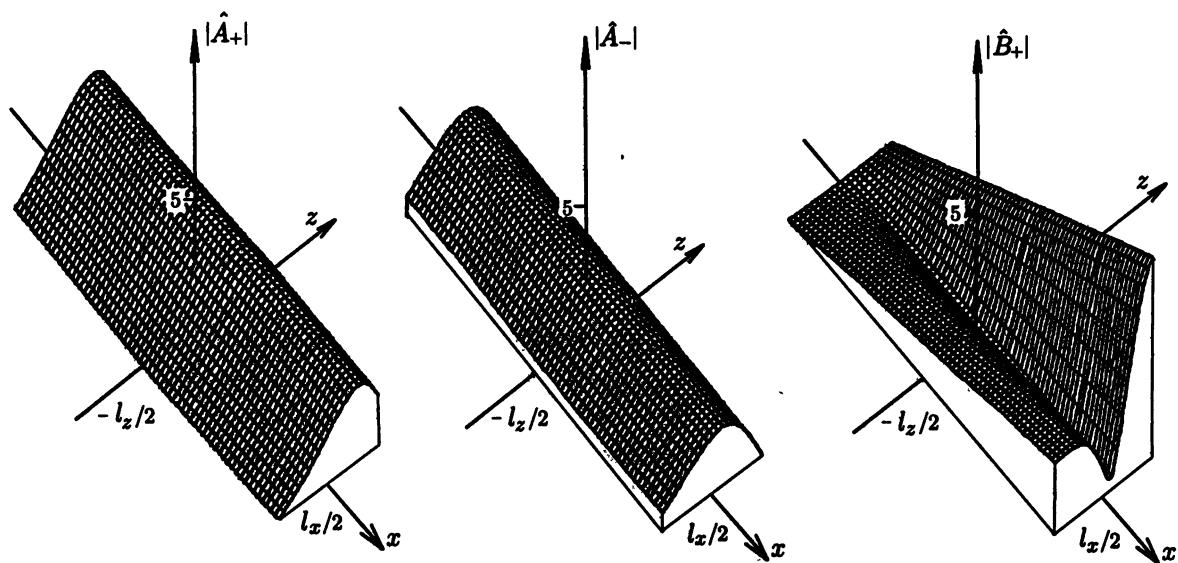


Рис. 5. Пространственное распределение амплитуд полей парциальных волн  $\hat{A}_{\pm}$  и  $\hat{B}_{\pm}$  в стационарном режиме генерации для параметров ЛСЭ, соответствующих кривой 3 на рис. 4

Это позволяет привести систему уравнений (11) к виду

$$\begin{aligned} \frac{d\hat{A}_+}{dZ} + i\hat{\alpha}^2 L_x (\hat{A}_+ + \hat{A}_-) &= \frac{1}{\pi} \int_0^{2\pi} \exp(-i\theta) d\theta_0, \\ \frac{d\hat{A}_-}{dZ} - i\hat{\alpha}^2 L_x (\hat{A}_+ + \hat{A}_-) &= 0, \\ \frac{d^2\theta}{dZ^2} &= \operatorname{Re}(\hat{A}_+ \exp(i\theta)). \end{aligned} \quad (14)$$

Следовательно, при фиксированном продольном размере системы распределение полей вдоль координаты  $z$  и электронный КПД остаются неизменными при выполнении условия  $\hat{\alpha}^2 L_x = \text{const}$ . Подобная автомодельность позволяет увеличивать ширину пространства взаимодействия за счёт снижения коэффициента связи волн  $\alpha$  (например, уменьшая глубину гофра  $a_1$ ). Численное моделирование нестационарных уравнений (11) показывает, что при  $L_z \leq 5$  найденный режим синхронизации оказывается устойчивым, по крайней мере, до  $L_x \leq 30$  (при параметре Пирса  $C \approx 10^{-2}$  это соответствует  $l_x/\lambda \approx 10^3$ ). Однако время установления увеличивается с ростом ширины системы (см. рис. 4).

Отметим, что при увеличении тока пучка над порогом генерации наблюдалось возбуждение высших мод двумерного брэгговского резонатора (с одной и более вариациями поля по обеим координатам). При большем превышении порога возникают автомодуляционные и стохастические режимы генерации (рис. 6).

Используем полученные выше результаты для оценки параметров сверхмощного ЛСЭ, который может быть реализован на базе сильноточного ускорителя У-2 (ИЯФ СО РАН), формирующего ленточный РЭП с длительностью до 10 мкс, энергией 1 МэВ, погонным током 1 кА/см и поперечным сечением  $0,5 \times 140$  см. Фокусировка пучка осуществляется продольным магнитным полем с напряжённостью около 1 Тл, для обеспечения поперечной скорости частиц  $\beta_{\perp} \approx 0,2$  предполагается использовать планарный токовый ондулятор с периодом 4 см и амплитудой поперечного поля на оси до 0,2 Тл. При указанных параметрах для излучения 4-мм диапазона длин волн параметр Пирса  $C \approx 0,006$ , и кривой 3 на рис. 4 ( $L_z = 4,0$ ;  $L_x = 12,8$ ;  $\hat{\alpha} = 0,32$ ) при глубине гофра 0,4 мм соответствует длина резонатора 50 см и ширина 150 см. Таким образом, при КПД 10 % мощность излучения будет достигать 15 ГВт. Время установления колебаний составит 150 нс.

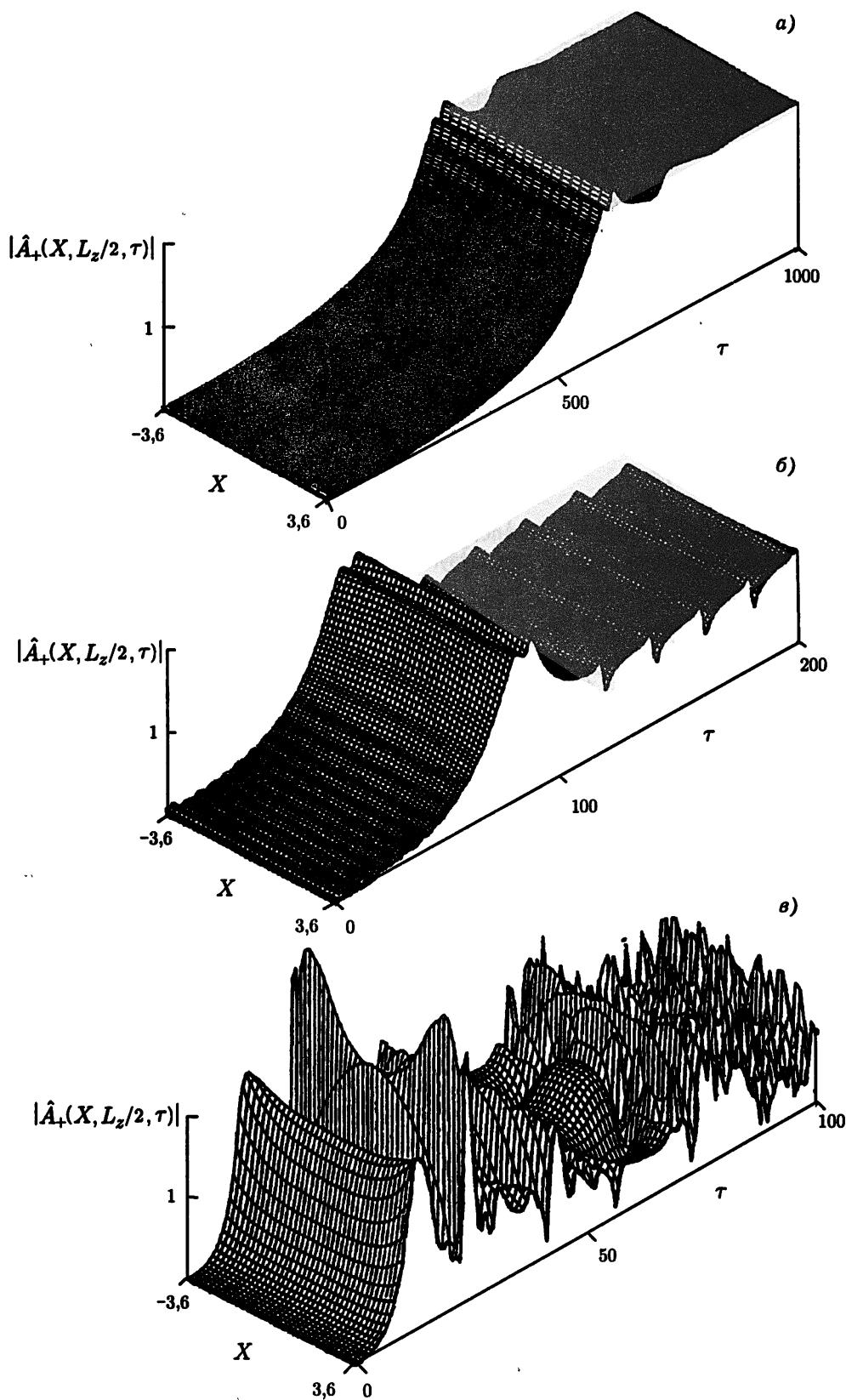


Рис. 6. Эволюция пространственного распределения выходного излучения ( $L_x = 7,2; \Delta = -1,8; \hat{\alpha} = 0,42$ ): при установлении стационарного режима генерации,  $L_z = 4,0$  (а); в режиме периодической автомодуляции,  $L_z = 5,0$  (б); в режиме стохастической автомодуляции,  $L_z = 12,0$  (в)

### 3. ДИНАМИКА ЛСЭ-ГЕНЕРАТОРОВ КОАКСИАЛЬНОЙ ГЕОМЕТРИИ С ДВУМЕРНОЙ РАСПРЕДЕЛЁННОЙ ОБРАТНОЙ СВЯЗЬЮ

Принципиальная схема коаксиального ЛСЭ-генератора с двумерной РОС и трубчатым электронным пучком изображена на рис. 7. Коаксиальный брэгговский резонатор [17], внутренняя поверхность которого гофрирована по закону<sup>1</sup>

$$a = a_1 [\cos(\bar{m}\varphi - \bar{h}z) + \cos(\bar{m}\varphi + \bar{h}z)], \quad (15)$$

обеспечивает связь четырёх парциальных волн, распространяющихся в продольном (волны  $A_{\pm}$ ) и азимутальном (волны  $B_{\pm}$ ) направлениях. Для сверхразмерного трубчатого РЭП можно считать, что средний радиус  $r_0$  резонатора существенно превосходит расстояние между проводниками ( $r_0 \gg a_0$ ), а также длину волны ( $r_0 \gg \lambda$ ). В этих условиях кривизной резонатора можно пренебречь и для описания эволюции полей в поперечном относительно движения РЭП направлении ввести координату  $x = r_0\varphi$ . Возбуждение электронным потоком коаксиального брэгговского резонатора будет по-прежнему описываться уравнениями (11) [17]. Основное отличие от планарной модели сведётся к тому, что в поперечном направлении парциальные волны должны удовлетворять условию цикличности

$$A_{\pm}(x + l_x, z, t) = A_{\pm}(x, z, t), \quad B_{\pm}(x + l_x, z, t) = B_{\pm}(x, z, t), \quad (16)$$

где  $l_x = 2\pi r_0$  — средний периметр резонатора. При этом граничные условия в продольном направлении остаются аналогичными (6) и (12). Условие цикличности (16) позволяет разложить поля в ряд Фурье:

$$A_{\pm}(x, z, t) = \sum_{m=-\infty}^{+\infty} A_{\pm}^m(z, t) \exp(imx/r_0), \quad B_{\pm}(x, z, t) = \sum_{m=-\infty}^{+\infty} B_{\pm}^m(z, t) \exp(imx/r_0), \quad (17)$$

и рассматривать каждую гармонику как моду резонатора с азимутальным индексом  $m$ .

Анализ спектра собственных мод «холодного» резонатора, аналогичный приведённому в разделе 1, показывает, что отличительной особенностью коаксиального двумерного брэгговского резонатора является наличие азимутально-симметричной моды ( $m = 0$ ), у которой отсутствуют дифракционные потери. Собственная частота этой моды совпадает с брэгговской частотой, а добротность ограничена только омическими потерями  $\sigma$ . В то же время остальные, азимутально-несимметричные моды имеют дифракционные потери и, следовательно, обладают существенно меньшей добротностью. Таким образом, при возбуждении такого резонатора электронным потоком происходит селективная дискриминация азимутально-несимметричных мод. Эволюция азимутального состава электромагнитного поля при установлении стационарного режима генерации иллюстрируется на рис. 8, где показана зависимость от времени различных азимутальных гармоник в спектре амплитуды волны  $B_+$  в сечении  $z = l_z$ . В рассматриваемом примере в момент времени  $t = 0$  было задано начальное возмущение для волны  $A_+$ , в спектре которого в равной степени были представлены гармоники с номерами  $m \in [-2, 2]$ . На начальном

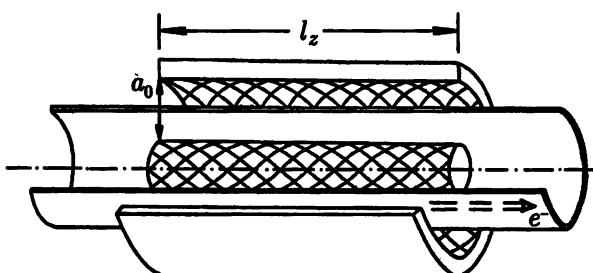


Рис. 7. Схема ЛСЭ-генератора с коаксиальным двумерным брэгговским резонатором и трубчатым релятивистским электронным пучком

<sup>1</sup> Отметим, что в случае коаксиального брэгговского резонатора с «одномерной» гофрировкой  $a = a_1 \cos(\bar{m}\varphi - \bar{h}z)$  система также будет обладать спектром высокодобротных мод [18, 19]. Однако при большой сверхразмерности селективность по азимутальному индексу в такой системе отсутствует.

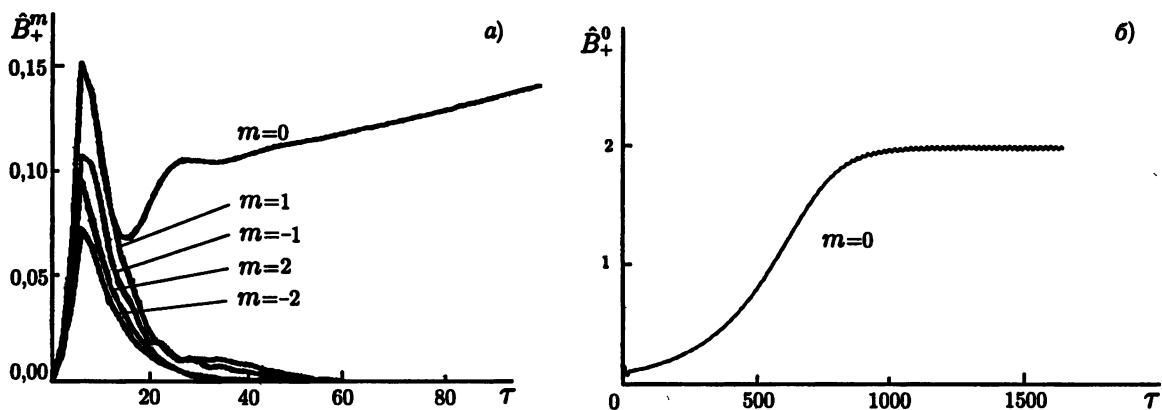


Рис. 8. Эволюция амплитуд азимутальных гармоник в спектре парциальной волны  $\mathcal{B}_+$  при установлении стационарного режима генерации в коаксиальном ЛСЭ с двумерной распределённой обратной связью ( $L_z = 6,8; L_x = 12,8; \Delta = -0,65; \hat{\alpha} = 0,1; \sigma = 0,1$ ): стадия установления азимутально-симметричной структуры излучения (а); стадия накопления энергии в резонаторе (б)

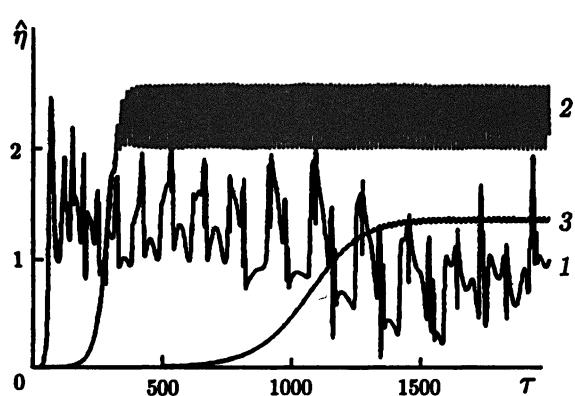


Рис. 9. Зависимость электронного КПД от времени в коаксиальном ЛСЭ с двумерной РОС при различных коэффициентах омических потерь  $\sigma$  ( $L_z = 6,8; L_x = 12,8; \Delta = -0,65; \hat{\alpha} = 0,1$ ): кривая 1 соответствует  $\sigma = 0$ ; 2 —  $\sigma = 0,08$ ; 3 —  $\sigma = 0,1$

вится определяющим на второй стадии переходного процесса — стадии накопления энергии в резонаторе. Поскольку эта стадия занимает основную часть переходного процесса (ср. рис. 8а и б), а распределение поля синхронной волны  $\mathcal{A}_+$  на этой стадии азимутально-симметрично, то общее время установления лишь в малой степени зависит от периметра системы  $l_x$ . Важно подчеркнуть, что установление стационарного режима колебаний наблюдалось, по крайней мере, при нормированном периметре  $L_x \leq 30$ , что при параметре усиления  $C \approx 10^{-2} \div 10^{-3}$  соответствует  $r_0/\lambda \approx 10^2 \div 10^3$ . Ограничение поперечного размера системы, как и в случае планарного ЛСЭ с двумерной РОС, было обусловлено не физическими факторами, а объёмом памяти компьютера, использованного для данного моделирования.

В то же время следует подчеркнуть, что для установления стационарного режима генерации в коаксиальном ЛСЭ с двумерной РОС принципиальное значение имеет наличие омических потерь. Как видно из рис. 9, на котором представлены зависимости интегрального КПД от времени при различных параметрах  $\sigma$ , стационарный режим генерации устанавливается, только если параметр потерь доста-

этапе происходит трансформация этой волны в волны  $\mathcal{B}_{\pm}$ , чем объясняется рост амплитуд всех гармоник в спектре  $\mathcal{B}_+$  (см. рис. 8а). Однако затем амплитуды всех гармоник, за исключением основной  $m = 0$ , падают практически до нуля. Соответственно, как видно из рис. 8, переходный процесс в коаксиальном двумерном брэгговском резонаторе может быть разделён на две стадии. На первой, относительно короткой линейной стадии в течение нескольких проходов парциальных волн по резонатору формируется азимутально-симметричная мода с частотой, близкой к брэгговской частоте. Наличие усиления, вносимого электронным пучком, на этой стадии переходного процесса существенной роли не играет, а дискриминация азимутально-несимметричных мод происходит благодаря наличию в резонаторе высокой электродинамической селективности. Взаимодействие с электронным пучком станов-

точно велик. Если же  $\sigma$  меньше, чем некоторое критическое значение (определенное, в частности, длиной системы), то возникает незатухающая периодическая или стохастическая автомодуляция. Кроме того, даже в стационарном режиме генерации, если параметр омических потерь  $\sigma$  меньше определенного оптимального значения, то с уменьшением  $\sigma$  имеет место увеличение амплитуды полей парциальных волн  $B_{\pm}$ , которые циркулируют в азимутальном направлении и, фактически, заперты в резонаторе. Таким образом, возникает достаточно парадоксальная ситуация, когда интегральные потери в генераторе увеличиваются по мере уменьшения параметра потеря  $\sigma$ .

#### 4. ЛСЭ С КОМБИНИРОВАННЫМИ РЕЗОНАТОРАМИ, СОСТАВЛЕННЫМИ ИЗ ОДНОМЕРНОГО И ДВУМЕРНОГО БРЭГГОВСКИХ ЗЕРКАЛ

Как было показано в разделах 1–3, использование двумерной РОС позволяет получить пространственно-когерентное излучение от пучков ленточной и трубчатой геометрии при практически неограниченных поперечных размерах. Отметим, однако, что при экспериментальной реализации указанного механизма возникают определенные затруднения, связанные, прежде всего, с рядом конструктивных ограничений. Так, например, для ЛСЭ планарной геометрии идеальной является ситуация, когда система открыта в поперечном направлении и возможен вывод во внешнее пространство поперечных (по отношению к поступательному движению электронного пучка) потоков электромагнитной энергии. В то же время для практической реализации ЛСЭ с двумерной РОС подобная схема является достаточно сложной из-за наличия непрозрачной для излучения электронно-оптической системы ЛСЭ, включающей в себя, в частности, соленоид. Таким образом, желательна разработка более простых, закрытых в поперечном направлении систем, обладающих односторонним выводом излучения. Как показано в [15], односторонний вывод излучения, в принципе, может быть осуществлен при использовании дополнительных одномерных брэгговских решеток, установленных на торцах резонатора и поворачивающих поперечные потоки энергии на  $90^\circ$  в направлении движения электронного потока. Однако для первых экспериментальных реализаций ЛСЭ с двумерной РОС подобная схема представляется достаточно сложной, и желательна разработка закрытых в поперечном направлении систем.

Для ЛСЭ коаксиальной геометрии вопросов, связанных с обеспечением одностороннего вывода энергии, не возникает, и большая часть излучаемой электронным потоком энергии здесь выводится в направлении поступательного движения электронного пучка [17]. Тем не менее, если двумерная брэгговская структура занимает все пространство взаимодействия, достаточно большими оказываются омические потери, связанные с наличием «запертых» (циркулирующих по азимутальной координате) поперечных потоков энергии. Так, для случая, приведенного на рис. 9, интегральные омические потери составляют около 25 %. Возможна реализация двухзеркальной схемы, когда две двоякопериодические брэгговские структуры (зеркала) разделены участком регулярного волновода. Однако для такой схемы характерно, что в зеркале, расположенном у коллекторного конца системы, оказываются достаточно большие азимутальные потоки энергии. Это в конечном счете может привести к развитию высокочастотных пробоев.

В качестве одного из возможных вариантов решения обозначенных выше проблем в данном разделе проанализирована схема ЛСЭ с комбинированным двухзеркальным резонатором, составленным из одномерного и двумерного брэгговских зеркал [20]. Планарный ЛСЭ с комбинированным резонатором изображен на рис. 10. В этой схеме расположение на катодном конце двумерное брэгговское зеркало обеспечивает синхронизацию излучения поперек ленточного электронного потока. При этом на противоположном, коллекторном конце оказывается достаточным использование традиционного одномерного брэгговского зеркала, отражающего часть выходного излучения и обеспечивающего замыкание кольца обратной связи. Важно, что для планарной геометрии оба зеркала являются закрытыми в поперечном направлении (с точностью до небольших омических потерь, которые следует внести

в двумерный брэгговский отражатель для стабилизации режима генерации). В результате, как будет показано ниже, при оптимально выбранных параметрах, когда коэффициент отражения от выходного зеркала не слишком высок, и, соответственно, процесс усиления излучения в основном происходит после входного зеркала, данные потери не оказывают существенного влияния на баланс энергии. Таким образом, преобладающая часть мощности, излучаемой электронным потоком, выдается в направлении поступательного движения частиц, и только небольшая её доля тратится на омические потери.

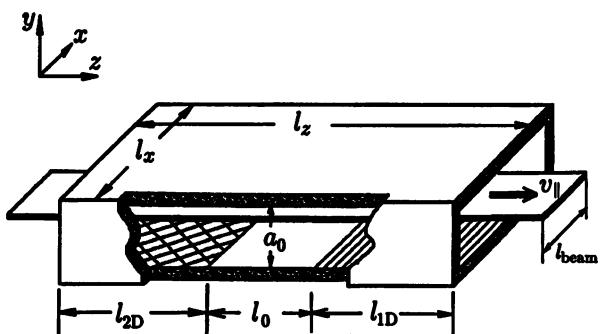


Рис. 10. Схема планарного ЛСЭ с ленточным релятивистским электронным пучком и комбинированным резонатором, составленным из одномерного и двумерного брэгговских зеркал

го резонанса (3), т. е.  $h \approx \bar{h}_{2D}$ . Взаимное рассеяние этих волн и связь с электронным потоком синхронной волны  $A_+$  по-прежнему описываются системой уравнений (11). С учётом отражений поперечных потоков энергии от боковых стенок резонатора планарной геометрии граничные условия для волн  $B_{\pm}$  приобретают вид

$$(\hat{B}_+ + R\hat{B}_-) \Big|_{x=0} = 0, \quad (\hat{B}_- + R\hat{B}_+) \Big|_{x=L_x} = 0, \quad (18)$$

где  $R \leq 1$  — соответствующий коэффициент отражения. Кроме того, будем считать, что ленточный электронный поток имеет ширину  $l_{beam}$  и заполняет, таким образом, только часть резонатора (см. рис. 10).

В выходном одномерном брэгговском зеркале, как и в регулярной части резонатора, присутствуют только две парциальные волны:

$$\mathbf{E} = \mathbf{E}_0 \operatorname{Re}\{[A_+ \exp(-ihz) + A_- \exp(ihz)] \exp(i\bar{\omega}t)\}, \quad (19)$$

распространяющиеся в попутном и встречном по отношению к поступательной скорости пучка направлениях. Одномерная брэгговская структура

$$a = a_{1D} \cos(\bar{h}_{1D} z) \quad (20)$$

при выполнении условия

$$h \approx \bar{h}_{1D}/2, \quad (21)$$

где  $\bar{h}_{1D} = 2\pi/d_{1D}$ ,  $2a_{1D}$  и  $d_{1D}$  — глубина и период одномерной гофрировки соответственно, обеспечивает рассеяние этих волн, которое описывается уравнениями

$$\left( \frac{\partial}{\partial Z} + \beta_{gr}^{-1} \frac{\partial}{\partial \tau} \right) \hat{A}_+ + i\alpha_{1D} \hat{A}_- = J, \quad \left( \frac{\partial}{\partial Z} - \beta_{gr}^{-1} \frac{\partial}{\partial \tau} \right) \hat{A}_- + i\alpha_{1D} \hat{A}_+ = 0. \quad (22)$$

Здесь  $\alpha_{1D}$  — коэффициент связи волн на одномерной брэгговской структуре [1–2]. В регулярной части резонатора процесс усиления волны  $A_+$  описывается уравнением, аналогичным (22), в котором

коэффициент связи волн следует положить равным нулю. Границные условия для парциальных волн в продольном направлении имеют вид, аналогичный (6), в которых  $l_z = l_{2D} + l_0 + l_{1D}$  — полная длина резонатора (см. рис. 10). Естественно, что амплитуды парциальных волн  $A_{\pm}$  следует считать непрерывными на границах различных секций.

Отметим, что предполагаются одновременно выполненные условия двумерного (3) и одномерного (21) брэгговских резонансах. Это, очевидно, возможно, когда  $\bar{h}_{2D} = \bar{h}_{1D}/2$ , и, соответственно, необходимо обеспечить следующее соотношение между периодами гофрировок:  $d_{2D} = \sqrt{2}d_{1D}$ .

На рис. 11, 12 представлены результаты моделирования ЛСЭ в 4-мм диапазоне длин волн с комбинированным резонатором при параметрах пучка, формируемого ускорителем У-2 (ИЯФ СО РАН). При моделировании длина входного (двумерного) брэгговского зеркала принималась равной 18 см, длина выходного (одномерного) — 9 см, регулярная часть резонатора составляла 25 см, поперечное сечение резонатора  $1 \times 140$  см. Параметр Пирса составлял  $C \approx 6 \cdot 10^{-3}$  (см. раздел 2), что соответствует безразмерным  $L_x = 12,8$ ;  $L_{2D} = 1,7$ ;  $L_0 = 2,1$ ;  $L_{1D} = 0,8$ .

На рис. 11 показана зависимость электронного КПД от времени при различной ширине пространства взаимодействия, иллюстрирующая возможность установления стационарного режима генерации. Заметим, что с увеличением ширины при сохранении прочих геометрических размеров варьировался (в сторону уменьшения) коэффициент связи волн. Для стабилизации режима генерации в модель вносились определенные потери для волн  $B_{\pm}$ , коэффициент отражения которых во входном двумерном брэгговском зеркале полагался  $R \approx 0,95$ . Очевидно, указанные потери могут быть реализованы путем нанесения поглотителя на боковые стенки волновода.

Для случая  $L_x = 12,8$ , когда ленточный РЭП занимает примерно 80 % ширины резонатора, пространственные распределения амплитуды полей парциальных волн в стационарном режиме генерации показаны на рис. 12. Видно, что в области, занятой электронным пучком, амплитуда поля синхронной волны  $A_+$  является практически постоянной по всей поперечной координате (аналогично случаю открытого в поперечном направлении резонатора). Это обеспечивает одинаковую энергоотдачу всех фракций электронного потока и достаточно высокий средний КПД. Заметим, что отличие от единицы коэффициентов отражения для волн  $B_{\pm}$  во входном зеркале не оказывает заметного влияния на баланс мощностей, и связанные с этим потери не превышают 1÷2 %.

Аналогичная комбинированная схема брэгговского резонатора представляется также перспективной и для ЛСЭ коаксиальной геометрии. Она позволяет оптимизировать максимальные амплитуды полей внутри резонатора, а также снизить до 5÷10 % омические потери. При этом моделирование показывает устойчивое возбуждение азимутально-симметричной моды при периметре системы, достигающем  $10^3$  длин волн.

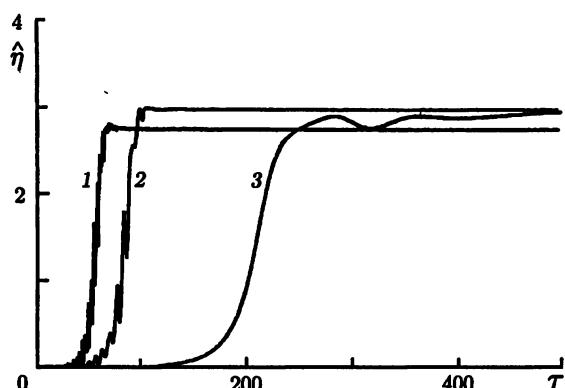


Рис. 11. Моделирование процесса установления автоколебаний в ЛСЭ-генераторе с комбинированным брэгговским резонатором. Зависимость электронного КПД от времени при различной ширине системы ( $L_{2D} = 1,7$ ;  $L_0 = 2,1$ ;  $L_{1D} = 0,8$ ;  $R = 0,95$ ;  $\Delta = -1,5$ ): кривая 1 соответствует  $L_x = 1,8$ ;  $\hat{\alpha}_{2D} = 0,9$ ;  $\hat{\alpha}_{1D} = 1,1$ ; кривая 2 —  $L_x = 6,5$ ;  $\hat{\alpha}_{2D} = 0,5$ ;  $\hat{\alpha}_{1D} = 0,35$ ; кривая 3 —  $L_x = 12,8$ ;  $\hat{\alpha}_{2D} = 0,3$ ;  $\hat{\alpha}_{1D} = 0,3$

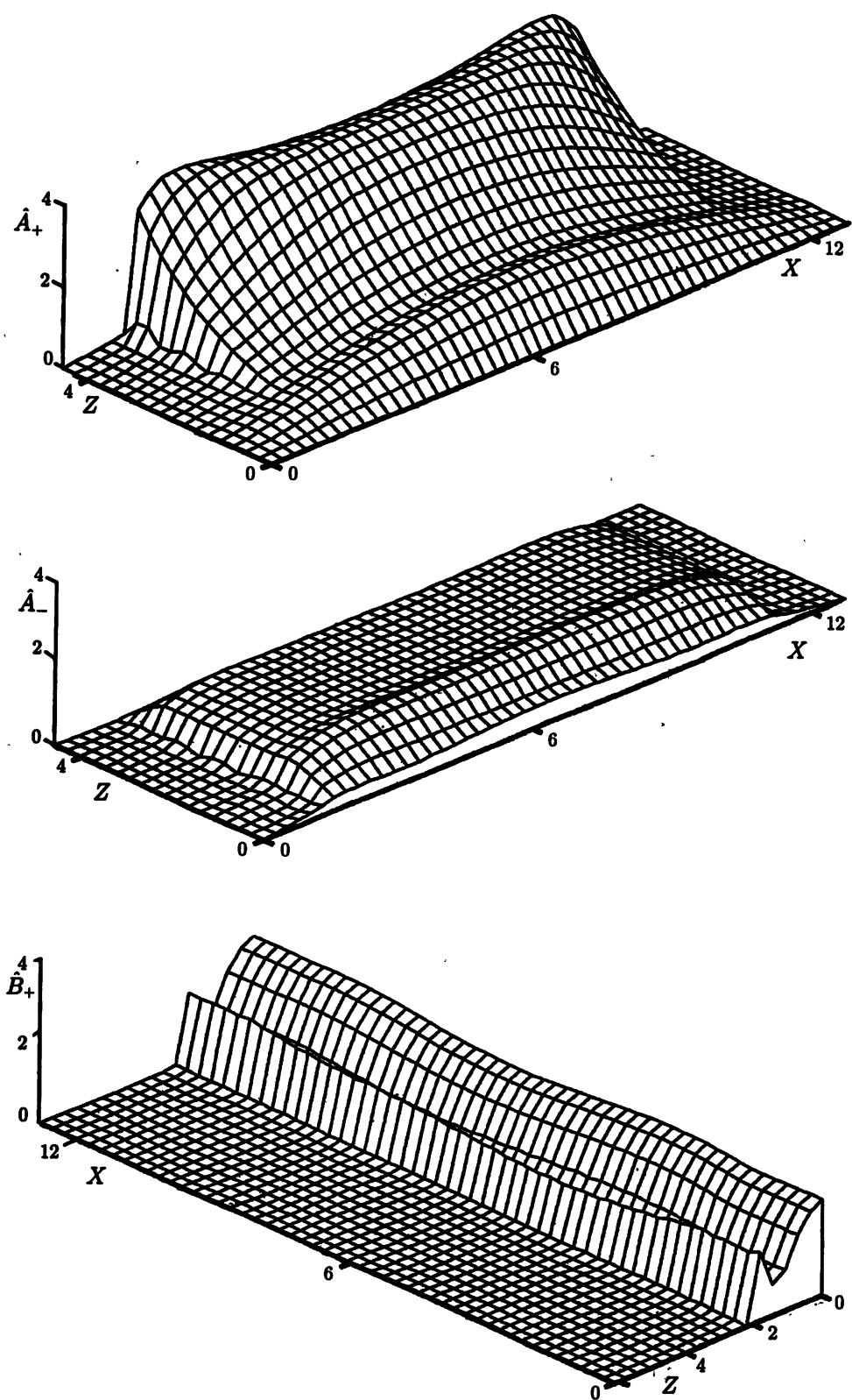


Рис. 12. Пространственное распределение амплитуд полей парциальных волн  $\hat{A}_{\pm}$  и  $\hat{B}_+$  в стационарном режиме генерации для параметров ЛСЭ, соответствующих кривой  $\mathcal{Z}$  на рис. 11

## 5. ИСПОЛЬЗОВАНИЕ ДВУМЕРНОЙ РАСПРЕДЕЛЁННОЙ ОБРАТНОЙ СВЯЗИ ДЛЯ СИНХРОНИЗАЦИИ МНОГОПУЧКОВОГО ЛСЭ

Естественным путём дальнейшего радикального повышения мощности излучения является развитие пространства взаимодействия вдоль второго поперечного размера системы. Практическая реализация такой схемы представляет собой многопучковый генератор, который состоит из нескольких планарных секций (модулей), синхронизуемых посредством двумерной РОС [21]. Каждый модуль запитывается ленточным РЭП, связь различных модулей осуществляется через возникающие в двумерных брэгговских структурах поперечные потоки электромагнитной энергии. При этом возможна реализация как генераторных, так и усилительных схем. В первом случае должно иметь место замыкание в кольцо потоков энергии с первого и последнего модуля. Во втором случае данное кольцо оказывается разомкнутым, и его концы используются для подачи входного сигнала.

Общая схема многомодульного ЛСЭ-генератора изображена на рис. 13. Каждый из модулей (рис. 13б) представляет собой генератор с двумерным планарным брэгговским резонатором, имеющим гофрировку вида (1). ЛСЭ-модуль запитывается ленточным РЭП, электроны которого, фокусируемые продольным магнитным полем, осциллируют в поле планарного ондулятора, примыкающего к каждому из модулей. Различные модули связаны последовательно через поперечные потоки энергии, возникающие за счёт рассеяния при возбуждении двумерного брэгговского резонатора ленточным РЭП. При этом для замыкания кольца обратной связи первый и последний модули должны быть также связаны друг с другом (рис. 13а).

Электромагнитное поле внутри модуля с индексом  $n$  будем по-прежнему представлять как совокупность 4-х парциальных волн (2), а процесс электронно-волнового взаимодействия в  $n$ -м модуле описывать с помощью системы уравнений (11). Границные условия для электронного пучка и парциальных волн, распространяющихся вдоль пучка, имеют вид, аналогичный (6), (12). Для парциальных волн, распространяющихся в поперечном направлении и осуществляющих связь различных модулей, граничные условия могут быть представлены в виде [21]

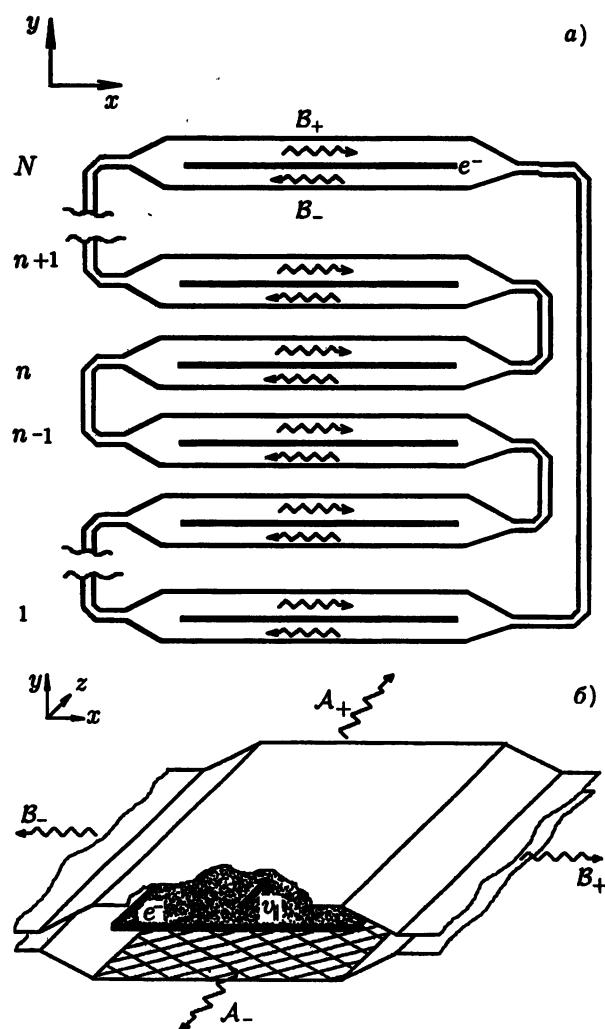


Рис. 13. Общая схема многопучкового ЛСЭ, состоящего из  $N$  планарных модулей: в каждом модуле показано поперечное сечение ленточного РЭП и направление поперечных потоков электромагнитной энергии (а); Схема отдельного ЛСЭ-модуля, использующего двумерный брэгговский резонатор и запитываемого ленточным РЭП (б)

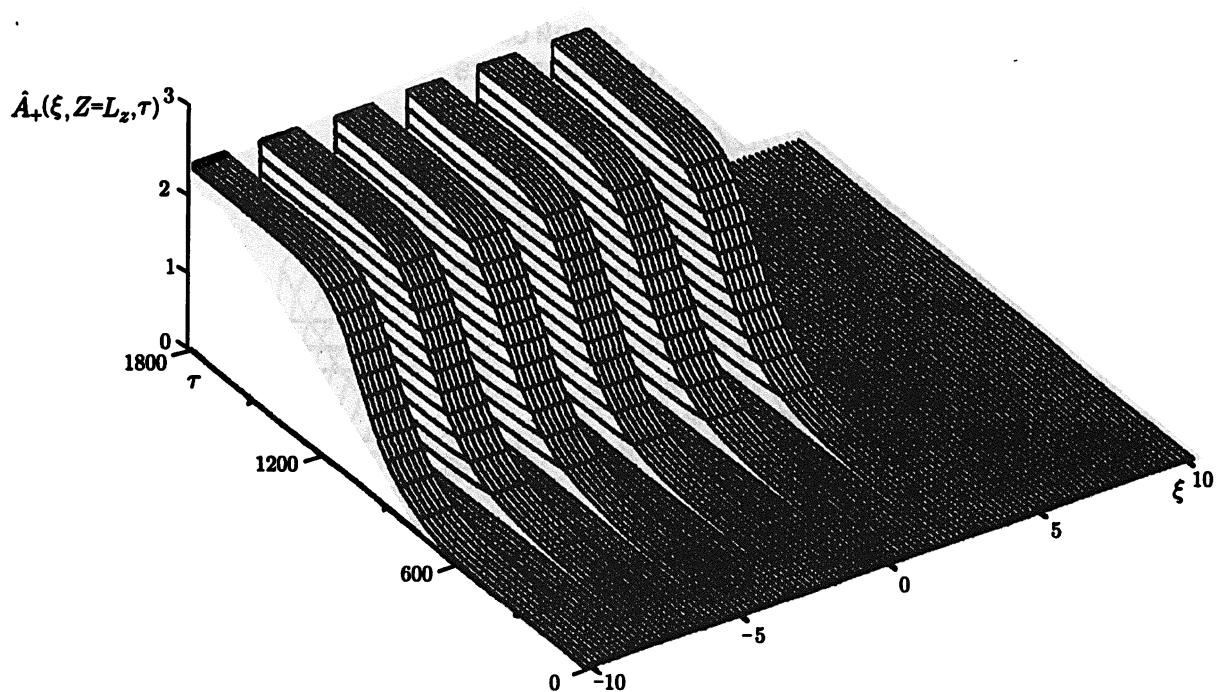


Рис. 14. Моделирование процесса установления автоколебаний в шестимодульном ( $N = 6$ ) ЛСЭ-генераторе. Зависимости от времени амплитуды поля на выходе системы, при  $Z = L_z$  ( $L_z = 6,8; L_x = 1,2; T = 1,2; \hat{\alpha} = 0,1; \Delta = -0,2; \Gamma = 0,95$ )

$$\begin{aligned} [B_{+,n}(\tau) = \Gamma B_{-,n-1}(\tau - T)]|_{X=0}, & \quad [B_{-,n}(\tau) = \Gamma B_{+,n+1}(\tau - T)]|_{X=L_z}, \\ [B_{-,N}(\tau) = \Gamma^N B_{+,1}(\tau - NT)]|_{X=L_z}, & \quad [B_{-,1}(\tau) = \Gamma^N B_{+,N}(\tau - NT)]|_{X=L_z}, \end{aligned} \quad (23)$$

где  $L_x$  — ширина модулей,  $T$  — время запаздывания при распространении сигнала между соседними модулями,  $\Gamma < 1$  — коэффициент омических потерь. При этом для связи между крайними модулями естественно предположить, что время запаздывания равно  $NT$ , где  $N$  — полное число модулей. Таким образом, различные модули оказываются связанными в кольцо (см. рис. 13а), и процесс синхронизации автоколебаний в системе  $N$  раздельных модулей в значительной степени оказывается аналогичным процессу синхронизации в ЛСЭ-генераторе с двумерным брэгговским резонатором коаксиальной геометрии. Следует также отметить, что подобно ЛСЭ коаксиальной геометрии наличие омических потерь для волн  $B_{\pm}$  является необходимым условием установления стационарного режима автоколебаний.

Моделирование процесса синхронизации ЛСЭ-модулей проводилось при параметрах, близких к условиям эксперимента, планируемого на ускорителе «ЭЛМИ» (1 МэВ/3 кА/5 мкс). Ширина отдельных пучков принималась равной 15 см, и, таким образом, при длине волны излучения 4 мм параметр Пирса был равен  $C \approx 4 \cdot 10^{-3}$ . На рис. 14, 15 представлены результаты моделирования процесса синхронизации шести генераторных модулей с приведённой нормализованной шириной пространства взаимодействия  $L_x = 1,2$  (что в условиях эксперимента на «ЭЛМИ» соответствует  $l_x = 20$  см), длине модулей  $L_z = 6,8$  ( $l_z = 110$  см), нормированным временем запаздывания  $T = 1,2$  (0,67 нс) и коэффициентом омических потерь  $\Gamma \approx 0,95$ . Рис. 14 иллюстрирует процесс установления стационарного режима генерации, на нём представлена временная зависимость распределения поля  $A_+$  в выходном сечении  $Z = L_z$ . Частота генерации в установившемся режиме близка к частоте точного брэгговского резонанса. Здесь введена координата  $\xi$ , описывающая распределение поля по общему контуру системы, включая все модули и волноводы связи. На рис. 15 показано пространственное распределение поля

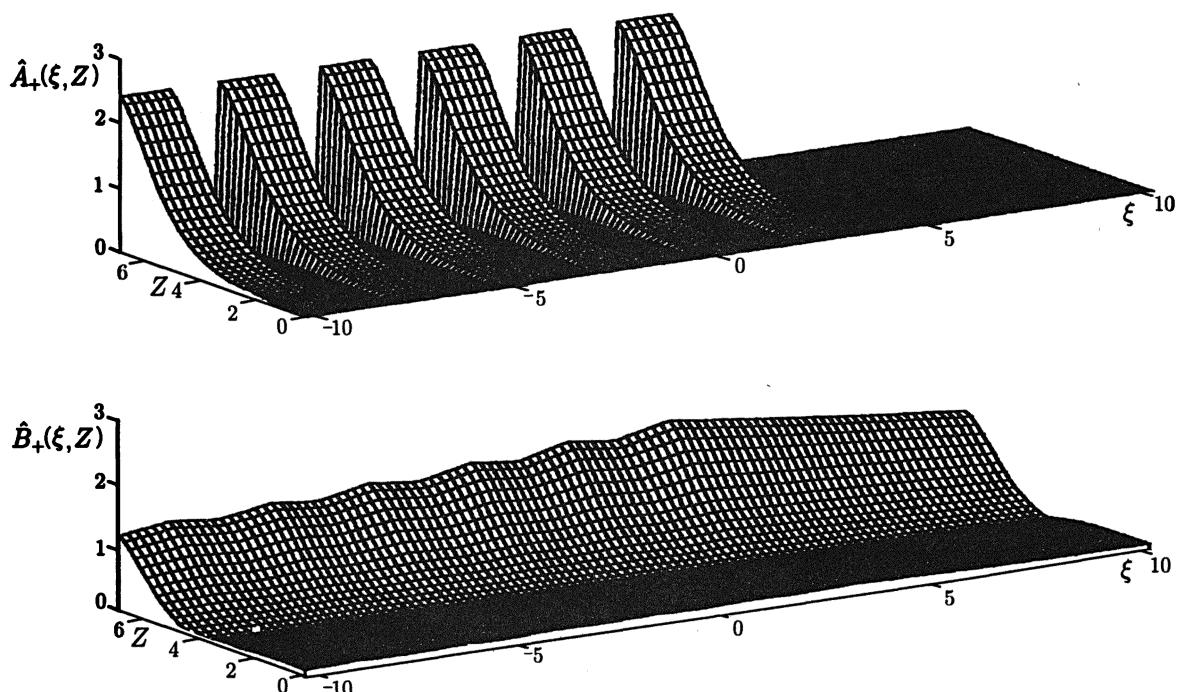


Рис. 15. Пространственное распределение амплитуд полей парциальных волн  $A_+$  и  $B_+$  в стационарном режиме генерации для параметров ЛСЭ, соответствующих рис. 14

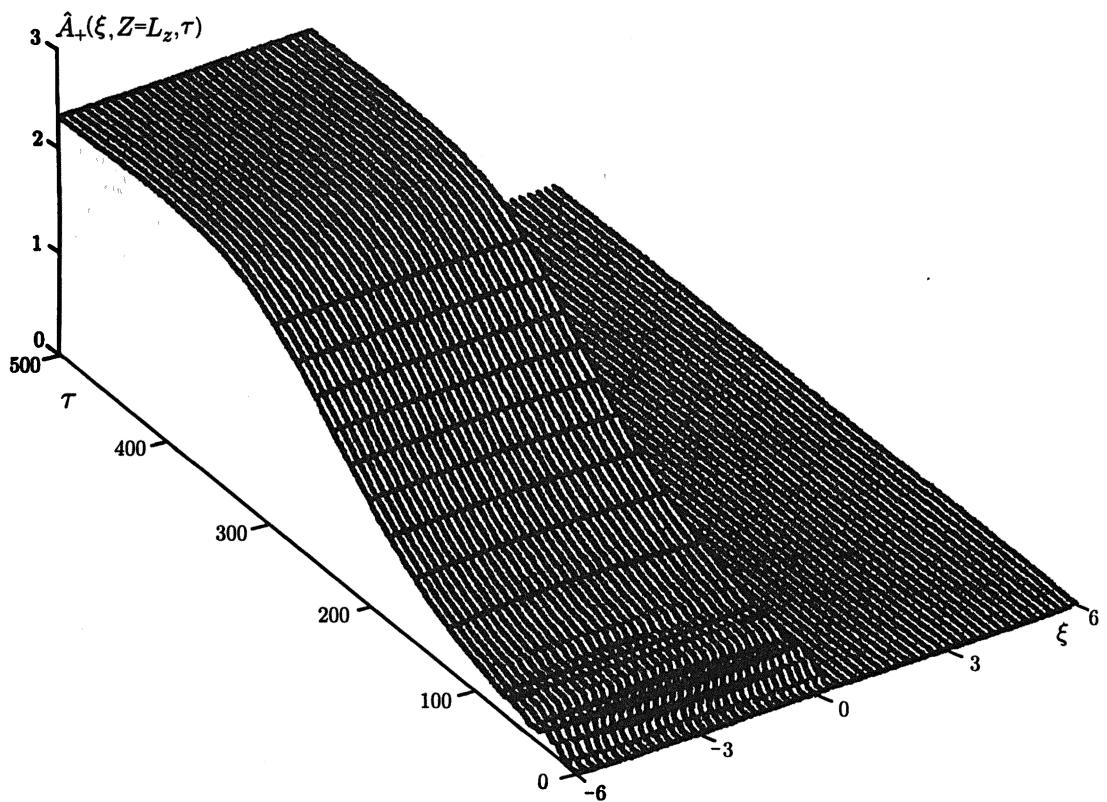


Рис. 16. Моделирование процесса установления автоколебаний в односекционном ( $N = 1$ ) планарном ЛСЭ при замыкании в кольцо поперечных электромагнитных потоков. Зависимость от времени амплитуды поля на выходе системы, при  $Z = L_z$  ( $L_z = 6,8; L_x = 6; T = 6; \hat{\alpha} = 0,1; \Delta = 0; \Gamma = 0,95$ )

синхронной волны  $A_+$  и одной из распространяющихся поперёк волн  $B_+$ , осуществляющих синхронизацию различных модулей.

Важно подчеркнуть, что синхронизация имеет место даже в условиях, когда различные модули запитываются электронными потоками с несколько отличающимися энергиями и, соответственно, имеющими разные расстройки синхронизма  $\Delta_n$ . При рассматриваемых параметрах самовозбуждение и синхронизация различных модулей имели место, пока относительный разброс расстроек составлял менее 50 % ширины полосы самовозбуждения. Далее автоколебания срывались. Ситуации, в которых режим синхронизации нарушался и отдельные модули генерировали на разных частотах, в проведенном моделировании не наблюдались.

Отметим в заключение, что поперечная синхронизация излучения в генераторе путём замыкания правого и левого конца пространства взаимодействия может быть эффективна и в случае одного модуля. В разделе 2 было показано, что использование планарных двумерных брэгговских резонаторов, открытых в поперечном направлении, позволяет практически неограниченно увеличивать ширину пространства взаимодействия, однако экспериментальная реализация таких устройств сталкивается с принципиальными трудностями, связанными с выводом поперечных потоков энергии. В разделе 4 обсуждались закрытые в поперечном направлении комбинированные системы резонаторов, обеспечивающие односторонний вывод излучения. В то же время использование дополнительного волновода, замыкающего в кольцо поперечные потоки энергии с правого и левого концов пространства взаимодействия, может рассматриваться в качестве одной из возможностей обеспечивать односторонний вывод излучения. На рис. 16 показан процесс формирования поперечного распределения амплитуды выходного излучения в такой схеме, при котором устанавливается стационарный режим генерации с частотой, совпадающей с брэгговской.

## 6. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ ЛСЭ С ДВУМЕРНОЙ РАСПРЕДЕЛЁННОЙ ОБРАТНОЙ СВЯЗЬЮ

Проведённый теоретический анализ подтверждает перспективность использования двумерной РОС для синхронизации излучения мощных РЭП ленточной и трубчатой геометрий. Экспериментальные исследования планарного ЛСЭ с частотой излучения 75 ГГц с новым механизмом обратной связи ведутся в сотрудничестве ИПФ РАН и ИЯФ СО РАН [22–25], коаксиальная геометрия исследуется на базе ЛСЭ с рабочей частотой 38 ГГц, разрабатываемого совместно ИПФ РАН и Стратклайдским университетом (Глазго, Великобритания) [26].

К настоящему времени проведены экспериментальные исследования электродинамических характеристик «холодных» двумерных брэгговских резонаторов обеих упомянутых геометрий в отсутствие электронного потока [25]. Измерялись частотные зависимости коэффициентов отражения, прохождения и рассеяния в поперечном направлении. При проведении этих исследований выяснилось, что указанные характеристики существенно зависят от формы поверхности, аппроксимирующей идеальную синусоидальную решётку, задаваемую соотношением (1). Так, в простейшем случае при изготовлении указанной решётки в виде двух последовательностей прямоугольных канавок, нарезанных под углом 90° друг к другу (рис. 17б), возникает дополнительная зона отражения, обусловленная примесью одномерной брэгговской решётки и непосредственным взаимным рассеянием парциальных волн  $A_+$  и  $A_-$  (см. подробнее [25]). В этой связи более оптимальной оказалась двумерная брэгговская «шахматная» решётка, представленная на рис. 17в, которая может быть изготовлена, в частности, фотохимическим методом. Заметим, наконец, что в университете Карлсруе (Германия) путём механической обработки поверхности удалось изготовить планарную двумерную синусоидальную брэгговскую решётку (рис. 17а). Для решёток с формой поверхности, показанной на рис. 17а и б, были измерены частотные зависимости коэффициента отражения от брэгговских структур как планарной, так и коаксиальной геометрии, которые находятся в хорошем соответствии с результатами вычислений. Частоты и доброт-

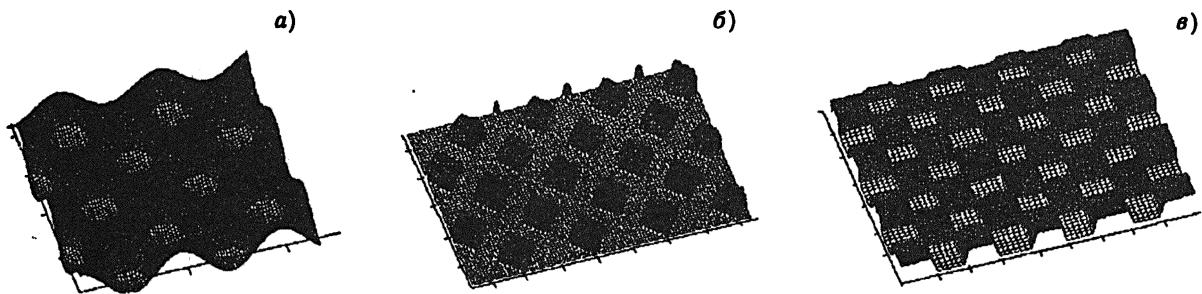


Рис. 17. Различные профили гофрировок двумерных брэгговских резонаторов: «идеальная» синусоидальная гофрировка (а), гофрировка в виде прямоугольных взаимно перпендикулярных канавок (б), гофрировка в виде «шахматных» ячеек (в)

ности собственных мод также соответствуют расчётным значениям, что подтверждает высокие селективные свойства резонаторов нового типа.

Работоспособность новой схемы обратной связи впервые продемонстрирована экспериментально в ЛСЭ, реализованном на базе ускорителя «ЭЛМИ» (ИЯФ СО РАН), формирующего ленточный РЭП с шириной до 15 см [23, 24]. В результате экспериментов в импульсе микросекундной длительности при КПД около 10 % получена мощность излучения на уровне 100 МВт. Частота излучения лежала внутри брэгговской полосы вблизи 75 ГГц. Дальнейшее увеличение мощности предполагается достигнуть при переходе на полномасштабный ленточный пучок. Первоначальную экспериментальную проверку предложенного механизма синхронизации для многопучковой схемы ЛСЭ предполагается осуществить на установке «ЭЛМИ» путём формирования двух параллельных пучков и объединения в единое устройство двух планарных мазеров.

Отметим в заключение, что в обсуждаемых экспериментах исследуются генераторы, основанные на вынужденном торможении излучении электронных потоков в периодическом магнитном (ондуляторном) поле. Рассматриваемый механизм обратной связи может быть применен и к другим типам генераторов, прежде всего к мазерам на циклотронном резонансе. Существует также возможность использования черенковского излучения прямолинейно движущихся электронных пучков. При этом двумерная брэгговская решётка может служить одновременно и резонатором, и замедляющей системой, в которой электроны будут синхронно взаимодействовать с медленной пространственной гармоникой одной из парциальных волн.

Работа выполнена при частичной поддержке РФФИ (гранты № 00–02–17232 и 01–02–16749) и программы Министерства науки «Физика микроволн».

#### СПИСОК ЛИТЕРАТУРЫ

1. Резонатор: А. с. 720592 СССР / Ковалёв Н. Ф., Петелин М. И., Резников М. Г. Опубл. 1980. Бюл. № 9.
2. Братман В. Л., Гинзбург Н. С., Денисов Г. Г. // Письма в ЖТФ. 1981. Т. 7, № 21. С. 1 320.
3. Ярив А. Квантовая электроника. М.: Сов. радио, 1980.
4. Ботвинник И. Е., Братман В. Л., Волков А. Б., Гинзбург Н. С., Денисов Г. Г., Офицеров М. М., Петелин М. И. // Письма в ЖЭТФ. 1982. Т. 35, № 10. С. 418.
5. Bratman V. L., Denisov G. G., Ginzburg N. S., Petelin M. I. // IEEE J. Quant. Electron. 1983. V. 19, No. 3. P. 282.
6. Mima K., Imasaki K., Kiguma S., Akiba T., Ohigashi N., Tsunawaki Y., Tanaka K., Yamanaka C., Nakai S. // Nucl. Instr. and Meth. in Phys. Res. A. 1991. V. 285. P. 47.

7. Chu T. S., Hartemann F. V., Danly B. G., Temkin R. J. // Phys. Rev. Lett. 1994. V. 72, No. 15. P. 2391.
8. Zambon P., Witteman W. J., Van der Slot P. J. M. // Nucl. Instr. and Meth. in Phys. Res. A. 1994. V. 341. P. 88.
9. Песков Н. Ю., Гинзбург Н. С., Каминский А. А., Каминский А. К., Седых С. Н., Сергеев А. П., Сергеев А. С. // Письма в ЖТФ. 1999. Т. 25, № 11. С. 19.
10. Бастриков А. Н., Бугаев С. П., Киселёв И. Н., Кошелев В. И., Сукушин К. Н. // ЖТФ. 1988. Т. 58, № 3. С. 483.
11. Arzhannikov A. V., Nikolaev V. S., Sinitsky S. L., Yushkov M. V. // J. Appl. Phys. 1992. V. 72, No. 4. P. 1657.
12. Arzhannikov A. V., Bobylev V. B., Nikolaev V. S., Sinitsky S. L., Tarasov A. V. // Proc. 10th Int. Conf. On High-Power Particle Beams, San Diego, USA, 1994. V. 1. P. 136.
13. Гинзбург Н. С., Песков Н. Ю., Сергеев А. С. // Письма в ЖТФ. 1992. Т. 18, № 9. С. 23.
14. Ginzburg N. S., Peskov N. Yu., Sergeev A. S., Arzhannikov A. V., Sinitsky S. L. // Nucl. Instr. and Meth. in Phys. Res. A. 1995. V. 358. P. 189.
15. Гинзбург Н. С., Песков Н. Ю., Сергеев А. С. // Радиотехника и электроника. 1995. Т. 40, № 3. С. 401.
16. Ginzburg N. S., Peskov N. Yu., Sergeev A. S., Phelps A. D. R., Konoplev I. V., Robb G. R. M., Cross A. W., Arzhannikov A. V., Sinitsky S. L. // Phys. Rev. E. 1999. V. 60, No. 1. P. 935.
17. Ginzburg N. S., Peskov N. Yu., Sergeev A. S. // Optics Comm. 1994. V. 112. P. 151.
18. Denisov G. G., Cooke S. J. // Digest of 21st Int. Conf. on Infrared and Millimeter Waves, Berlin, Germany, 1996. P. AT2.
19. Denisov G. G., Bratman V. L., Phelps A. D. R., Samsonov S. V. // IEEE Trans. on Plasma Science. 1998. V. 26, No. 3. P. 508.
20. Гинзбург Н. С., Песков Н. Ю., Сергеев А. С., Аржанников А. В., Синицкий С. Л. // Письма в ЖТФ. 2000. Т. 26, № 16. С. 8.
21. Гинзбург Н. С., Песков Н. Ю., Сергеев А. С., Аржанников А. В., Синицкий С. Л. // Письма в ЖТФ. 2001. Т. 27, № 6. С. 50.
22. Arzhannikov A. V., Agafonov M. A., Ginzburg N. S., Ivanenko V. G., Kalinin P. V., Kuznetsov S. A., Peskov N. Yu., Sinitsky S. L. // IEEE Trans. on Plasma Science. 1998. V. 26, No. 3. P. 531.
23. Arzhannikov A. V., Agarin N. V., Bobylev V. B., Ginzburg N. S., Ivanenko V. G., Kalinin P. V., Kuznetsov S. A., Peskov N. Yu., Sergeev A. S., Sinitsky S. L., Stepanov V. D. // Strong microwaves in plasmas / Ed. by A. G. Litvak. N. Novgorod: IAP RAS, 2000. V. 2. P. 802.
24. Arzhannikov A. V., Agarin N. V., Bobylev V. B., Ginzburg N. S., Ivanenko V. G., Kalinin P. V., Kuznetsov S. A., Peskov N. Yu., Sergeev A. S., Sinitsky S. L., Stepanov V. D. // Nucl. Instr. and Meth. in Phys. Res. A. 2000. V. 445. P. 222.
25. Песков Н. Ю., Гинзбург Н. С., Денисов Г. Г., Сергеев А. С., Аржанников А. В., Калинин П. В., Синицкий С. Л., Степанов В. Д., Петров П. В. // Письма в ЖТФ. 2000. Т. 26, № 8. P. 72.
26. Cross A. W., Ginzburg N. S., He W., Konoplev I. V., Peskov N. Yu., Phelps A. D. R., Robb G. R. M., Ronald K., Sergeev A. S., Whyte C. G. // The 22nd Int. FEL Conf., Durham, USA, 2000. P. 176.

<sup>1</sup> Институт прикладной физики РАН,  
г. Нижний Новгород,

Поступила в редакцию  
17 апреля 2001 г.

<sup>2</sup> Институт ядерной физики им. Будкера  
СО РАН, г. Новосибирск, Россия

**GENERATION OF SPATIALLY-COHERENT RADIATION IN  
FREE-ELECTRON LASERS WITH A TWO-DIMENSIONAL DISTRIBUTED FEEDBACK**

*N. S. Ginzburg, N. Yu. Peskov, A. S. Sergeev, A. V. Arzhannikov, and S. L. Sinitsky*

We study the possibility of using a two-dimensional distributed feedback to generate spatially-coherent radiation of relativistic ribbon and hollow electron beams whose transverse dimensions exceed the wavelength by several orders of magnitude. Such a feedback can be realized in planar and coaxial two-dimensional Bragg resonators with a two-period corrugation of their side walls. This corrugation gives rise to the additional transverse (with respect to the reciprocal motion of the electrons) fluxes of electromagnetic energy, which synchronize emission from different parts of the electron beam. Simulations of the process of the autooscillation onset in free-electron lasers (FELs) with the two-dimensional distributed feedback shows the possibility of obtaining a single-mode monochromatic coherent generation by beams with transverse dimensions up to  $10^2$ – $10^3$  wavelengths. We also analyze the use of hybrid resonators comprising the input «two-dimensional» and output «one-dimensional» Bragg mirrors. In such a scheme, the «two-dimensional» mirror provides synchronization of the emission perpendicular to the electron beam, while reflection from the output «one-dimensional» mirror is enough for the self-excitation of the generator. In the case of a system closed in the transverse direction, such a scheme allows one to reduce significantly the ohmic losses resulting from the electromagnetic fluxes locked in the transverse direction. It is shown that the two-dimensional distributed feedback can also be used to synchronize radiation in a multi-beam generator consisting of planar FEL modules fed by a ribbon electron beam and coupled via transverse electromagnetic-energy fluxes which are formed by two-dimensional Bragg structures. The experimental studies aimed at realization of super-powerful FELs with the two-dimensional distributed feedback are discussed.