МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Известия высших учебных заведений

РАДИОФИЗИКА

ежемесячный научно-технический журнал

Издаётся с апреля 1958 г.

Том XLVIII № 1 **Нижний Новгород** **2005**

Содержание

УДК 001.891.573+681.3.06

ТЕХНОЛОГИЯ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ПРОГНОЗИРОВАНИЯ КРИТИЧЕСКОЙ ЧАСТОТЫ ИОНОСФЕРНОГО СЛОЯ F_2

Н. А. Бархатов [1], *С. Е. Ревунов* [1], *В. П. Урядов* [2]

[1] Нижегородский государственный педагогический университет, г. Нижний Новгород, Россия
[2] Научно-исследовательский радиофизический институт, г. Нижний Новгород, Россия

В данной работе на основе технологии искусственных нейронных сетей разработан алгоритм прогноза критической частоты ионосферного слоя F_2 на 1, 2, 3, 12 и 24 часа. Проведён экспериментальный поиск подходящего обучающего множества и архитектуры нейронной сети. Дополнительное использование параметров солнечного ветра и межпланетного магнитного поля, а также индексов геомагнитной возмущённости позволило не только улучшить эффективность прогноза, но и установить ряд закономерностей поведения рассматриваемой критической частоты. Практическая ценность выполненной работы заключается в применении её результатов для оперативной коррекции модели ионосферы с целью улучшения ионосферной коротковолновой радиосвязи.

ВВЕДЕНИЕ

Прогноз ионосферных параметров на интервалы от 30 минут до нескольких часов используется для повышения надёжности коротковолновой радиосвязи. При решении этой задачи выясняется, что совершенствование используемых физических моделей часто не приводит к желаемому результату. Это происходит из-за того, что полная физическая картина ионосферных процессов крайне разнообразна, имеет много факторов и параметров. В этих условиях усложнение аналитической модели не приводит к улучшению прогноза. Кроме того, существенно ионосферно-магнитосферное взаимодействие, обусловленное особенностями солнечно-земных связей. Практически это означает, что учёт такого большого числа факторов при разработке аналитического описания физических моделей становится невозможным. Существует, однако, альтернативный подход для решения подобных задач — метод искусственных нейронных сетей (ИНС), основанный на использовании математического моделирования искусственного интеллекта [1]. Этот метод сочетает корреляционную обработку с нелинейным преобразованием изучаемой многофакторной последовательности данных.

Нейронные сети обучаются по экспериментальным данным и в процессе обучения автоматически подстраивают весовые коэффициенты между своими элементами (нейронами), после чего могут успешно прогнозировать изучаемый процесс. Достоинством метода ИНС является возможность получения неявной модели системы без построения её конкретной физической модели, используя только достаточно большие массивы экспериментальных данных. Вместе с тем предпосылки использования экспериментальных данных в технологии ИНС предполагают физические представления об их причинно-следственных связях.

Анализ работ, связанных с использованием нейросетей для решения физико-математических задач, показывает, что нейронно-сетевой подход имеет преимущества перед традиционными методами в трёх случаях. Во-первых, когда рассматриваемая задача в силу конкретных особенностей не поддаётся адекватной формализации традиционными математическими методами, поскольку содержит элементы неопределённости. Во-вторых, когда рассматриваемая задача формализуема, но аппарат для её решения в настоящее время отсутствует. В-третьих, когда для рассматриваемой хорошо формализуемой задачи существует соответствующий математический аппарат, но реализация вычислений с его помощью на базе имеющихся вычислительных систем требует слишком большого времени и т. д. В такой ситуации приходится или упрощать алгоритмы, что снижает качество решений, или применять соответствующий нейронно-сетевой подход при условии, что он обеспечит нужное качество решения задачи [2].

Необычайно высокий интерес к нейронным сетям, проявляемый специалистами из разных областей деятельности, объясняется, прежде всего, очень широким диапазоном решаемых с их помощью задач, а также рядом преимуществ по сравнению с другими методами. Нейронные сети интенсивно используются при обработке изображений и нелинейном управлении, распознавании образов и адаптивной фильтрации, идентификации и в финансовом прогнозировании.

В настоящее время метод ИНС активно применяется в задачах прогноза для различных геофизических приложений. Вот неполный круг задач, которые были успешно решены с привлечением этого метода: долгосрочный прогноз индексов солнечной активности [3], предсказание индекса геомагнитной активности $D_{\rm st}$ [4], восстановление пробелов в записях отдельных магнитных обсерваторий по данным других станций [5], определение воздействия суббурь на геомагнитные бури [6], моделирование воздействия солнечного ветра на магнитосферу [7], предсказание геомагнитных бурь [8, 9], идентификация суббурь с использованием пульсаций Pi2 [10], моделирование развития геомагнитной суточной вариации [11], прогноз солнечной активности [12, 13] и процессов в солнечном ветре [14].

Выполнение ряда работ осуществлялось с помощью разработанной компьютерной программы, реализующей ИНС обратного распространения ошибки с петлёй обратной связи Элмана [3, 4]. Подобная нейросеть обладает нелинейной внутренней памятью, заключённой в петлях обратной связи, что позволяет накапливать и использовать информацию о предыстории процесса. Поэтому этот вид ИНС является эффективным инструментом для решения проблемы восстановления и прогноза временны́х рядов.

Метод ИНС применялся и для прогноза параметров ионосферы [15]. Достоинство метода проявляется здесь в том, что обученная сеть нуждается только в непрерывном потоке значений критической частоты ионосферного слоя F_2. Результаты выполненного с помощью ИНС Элмана прогноза критической частоты использованы для коррекции ионосферной модели и синтеза ионограмм наклонного зондирования на трассе Инскип (Англия)—Нижний Новгород. Сопоставление экспериментальных и расчётных ионограмм продемонстрировало повышение качества прогноза по сравнению со старыми моделями. Однако в выполненной работе [15] не учитывались солнечно-земные связи. В связи с этим представляется важным совершенствование (аналогично [16]) физической модели с учётом воздействия изменяющихся параметров солнечного ветра и межпланетного магнитного поля на процессы в ионосфере. Кроме того, применяемая сеть Элмана вследствие использования простого алгоритма градиентного спуска обладает ограниченными интеллектуальными возможностями, что приводит к простому запоминанию анализируемого процесса.

Перспективы начатого исследования следует связывать с сочетанием различных методов прогноза и коррекции ионосферных данных с использованием как технических средств зондирования, так и разработанных и программно реализованных алгоритмов расчёта. Это существенно расширяет возможности обеспечения высокого качества прогноза ионосферного коротковолнового радиоканала.

Данная работа посвящена исследованию возможности прогноза критической частоты слоя F_2 при учёте не только предыстории процесса, но и при введении в рассмотрение параметров солнечного ветра и межпланетного магнитного поля, их физически обоснованных комбинаций, а также индексов локальной и глобальной геомагнитной возмущённости. При этом физические предпосылки исследования определяют и выбор более разветвлённой ИНС, построенной на более совершенном алгоритме, чем сеть Элмана.

Работоспособность реализованного метода продемонстрирована на массиве значений критической частоты $f_{\text{кр}}$ слоя F_2 в феврале—марте 2002 года (данные получены с сайта [17]) на станции вертикального зондирования Чилтон (Англия). Параметры солнечного ветра и межпланетного магнитного поля, а также значения индексов магнитной активности получены с сайта [18]. Для реализации метода ИНС использовался специализированный пакет «Neural Network Toolbox», являющийся прикладным математическим расширением системы MATLAB 6 [2, 19].

Объективная оценка качества прогноза делалась по интервалам измерений критической частоты слоя F_2 на основе вычисления так называемой эффективности прогнозирования PE [4]. Точность прогноза оценивалась также путём вычисления классического коэффициента корреляции R между реальными и спрогнозированными значениями критической частоты. Кроме этого, оценка точности прогноза определялась вычислением классического коэффициента корреляции R между реальными и спрогнозированными значениями критической частоты. Во всех численных экспериментах коэффициент корреляции подсчитывался по прогнозируемому временному интервалу каждые полчаса. Его динамика позволяла судить об изменении достоверности прогноза в течение суток. Решение поставленной задачи представляло собой поиск подходящего обучающего множества параметров задачи и их физически обоснованных комбинаций. Основной целью исследования было подтвердить правильность физических предпосылок выбора алгоритма и архитектуры ИНС, а также создать удовлетворительную для наших целей сеть с как можно меньшим количеством связей. Последнее позволяет исключить эффект запоминания и сделать процесс прогнозирования более интеллектуальным.

1. ВЫБОР АЛГОРИТМА И АРХИТЕКТУРЫ НЕЙРОСЕТИ

Архитектура ИНС определялась из физических представлений о причинно-следственных связях и развитии глобальных процессов в магнитосфере и в ионосфере. В связи с этим в архитектуре ИНС входные векторы подразделены на определяющие предысторию (описывающие комплекс параметров поведения критической частоты) и определяющие параметры околоземного пространства.

Поскольку поиск обучающей последовательности предполагал наличие во входном массиве данных параметров межпланетного магнитного поля и солнечного ветра, то исключительно для них был создан отдельный вход с блоком динамической задержки сигнала. Это объясняется тем, что эти параметры регистрируются на космическом аппарате далеко за пределами Земли, т. е. раньше остальных на несколько часов. Чтобы учесть разницу во времени и определить, какая задержка даёт наилучший результат, данный блок позволял регулировать время задержки с шагом в 30 минут. Из общих соображений это время должно лежать в пределах от одного до трёх часов. Второй вход не имеет задержки и предназначен для всех остальных параметров, регистрируемых непосредственно на Земле. Поскольку предполагается в результате получать единственную последовательность — временной ряд спрогнозированных значений критической частоты, то выход нейросети представляет собой один нейрон.

Как известно, алгоритм обучения нейронных сетей аналогичен алгоритмам поиска глобального экстремума функции многих переменных. В ИНС обратного распространения ошибки для этого рассчитывается возникающая в выходном слое ошибка и вычисляется градиент как функция весов и смещений. Этот вектор указывает направление кратчайшего спуска по поверхности пространства весов для данной точки, поэтому при продвижении в этом направлении ошибка уменьшится. Последовательность таких шагов в конце концов приведёт к минимуму того или иного типа. Значительную трудность здесь вызывает выбор шага. Применяемый ранее алгоритм градиентного спуска [4], использующий постоянный шаг итерации, оказывается чрезвычайно неэффективным в случае, когда производные по различным весам сильно отличаются и рельеф функции ошибки сложен. В этом случае для плавного уменьшения ошибки необходимо выбирать очень маленький темп обучения, диктуемый максимальной производной, тогда как расстояние до минимума по порядку величины определяется минимальной производной. В итоге обучение становится неприемлемо медленным. Кроме того, вблизи минимума неизбежно возникают осцилляции, и обучение теряет привлекательное свойство монотонности убывания ошибки. Среди алгоритмов, выполняющих подобную операцию и использующих метод обратного распространения ошибки, есть более совершенные, лишённые этого недостатка. Это, прежде всего, алгоритм сопряжённых градиентов, основанный на непрерывной коррекции шага итерации и поиска оптимального спуска в минимум. Ещё более совершенным и альтернативным предыдущему является алгоритм Левенберга—Марквардта (Levenberg—Marquardt) [20], основанный на дополнительном вычислении вторых частных производных по весам и смещениям от функционала ошибки. Этот алгоритм входит в группу квазиньютоновых методов.

Предварительные численные эксперименты подтвердили обоснованность теоретических предпосылок выбора подходящей архитектуры нейросети и алгоритма. В результате была рассмотрена двухслойная сеть, каждый скрытый слой которой содержал по два нейрона, имеющих петли обратной связи, что реализовалось внедрением в схему контекстных нейронов. Такое физически обусловленное разделение входных векторов (параметров) и их значительное количество удовлетворяют выбранному алгоритму Левенберга—Марквардта. Этот алгоритм требует достаточную разветвлённую сеть и многоступенчатость внутренней нелинейной памяти, что в данных условиях можно обеспечить только увеличением количества скрытых слоёв. Архитектура используемой сети показана на рис. 1.

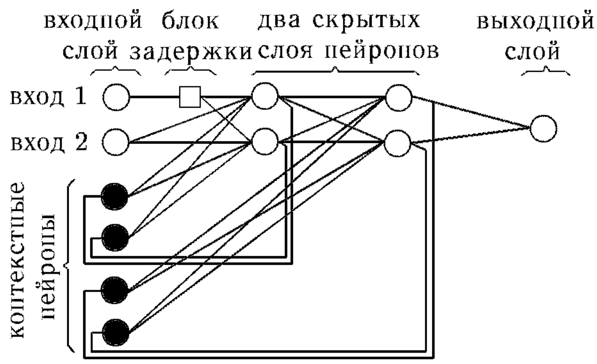

Рис. 1. Архитектура сети Элмана с добавленным блоком динамической задержки входного сигнала на одном из входов сети

Для ускоренного обучения выбранной ИНС используется метод Ньютона. Основной шаг этого метода определяется соотношением

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{H}_k^{-1}\mathbf{g}_k, \tag{1}$$

где $\mathbf{x}_k$ — вектор настраиваемых параметров, $\mathbf{H}_k^{-1}$ — обратная матрица Гессе вторых частных производных функционала ошибки по настраиваемым параметрам, $\mathbf{g}_k$ — вектор градиента функционала ошибки, k — шаг итерации. Процедуры минимизации на основе метода Ньютона, как правило, сходятся быстрее, чем те же процедуры на основе метода сопряжённых градиентов. Однако вычисление матрицы Гессе — весьма сложная и дорогостоящая в вычислительном отношении процедура. Поэтому был разработан класс алгоритмов, основанных на методе Ньютона, но не требующих вычисления вторых производных. Это класс квазиньютоновых алгоритмов, которые используют на каждой итерации некоторую приближённую оценку матрицы Гессе.

Алгоритм Левенберга—Марквардта реализует следующую стратегию для оценки матрицы Гессе. В предположении, что функционал определяется как сумма квадратов ошибок, гессиан может быть приближённо вычислен как

$$\mathbf{H} \approx \mathbf{J}^{\mathrm{T}}\mathbf{J}, \tag{2}$$

а градиент рассчитан по формуле

$$\mathbf{g} = \mathbf{J}^{\mathrm{T}}\mathbf{E}, \tag{3}$$

где $\mathbf{J}$ — матрица Якоби производных функционала ошибки по настраиваемым параметрам, $\mathbf{E}$ — полный вектор ошибок сети, индекс T обозначает транспонирование. Матрица Якоби может быть вычислена на основе стандартного метода обратного распространения ошибки, что существенно проще вычисления матрицы Гессе.

Алгоритм Левенберга—Марквардта использует указанную выше аппроксимацию гессиана (2):

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (\mathbf{H} + \mu\mathbf{I})^{-1}\mathbf{J}\mathbf{e}_k, \tag{4}$$

где $\mathbf{I}$ — единичная матрица; индекс $\mathbf{e}_k$ — вектор ошибок сети на шаге итерации k, μ — коэффициент настройки.

Когда коэффициент μ равен нулю, мы получаем метод Ньютона в форме (2); когда значение μ велико, получаем метод градиентного спуска с маленьким шагом. Поскольку метод Ньютона имеет бо́льшую точность и скорость сходимости вблизи минимума, задача состоит в том, чтобы в процессе минимизации как можно быстрее перейти к методу Ньютона. С этой целью параметр μ уменьшают после каждой успешной итерации и увеличивают только тогда, когда пробный шаг показывает, что функционал ошибки возрастает. Такая стратегия обеспечивает уменьшение ошибки после каждой итерации алгоритма.

Алгоритмы, основанные на приближённом методе Ньютона, требуют большого количества вычислений на каждой итерации и бо́льшего объёма памяти, чем градиентные алгоритмы обучения. При их использовании требуется на каждой итерации хранить оценку матрицы Гессе, размер которой определяется числом настраиваемых параметров. Очень трудно определить, какой обучающий алгоритм будет самым быстрым при решении той или иной практической задачи. Это зависит от многих факторов, включая сложность задачи, число элементов обучающего множества, число настраиваемых параметров сети и конечную ошибку. Для сетей, чьи матрицы весов и смещений имеют сотни элементов, число которых, в свою очередь, определяется размерами входных векторов, алгоритм Левенберга—Марквардта имеет самую быструю сходимость. Это преимущество особенно значимо, если требуется высокая точность обучения.

Главный недостаток алгоритма Левенберга—Марквардта состоит в том, что он требует значительного объёма оперативной памяти для хранения матриц больших размеров. Это означает, что при оценке (2) матрицы потребуются значительные ресурсы для её вычисления и хранения. Работа алгоритма обучения Левенберга—Марквардта может быть запрограммирована таким образом, что матрица оценки гессиана может быть разбита на несколько подматриц. Причём в процессе формирования матрицы Гессе использованные подматрицы могут быть удалены из оперативной памяти. Такая стратегия позволяет экономить объём оперативной памяти, но снижает быстродействие. Часто для стабильного численного счёта этого бывает недостаточно, т. к. сложные вычислительные алгоритмы системы MATLAB 6 требовательны к общим ресурсам компьютера. Перезапуск обучения ИНС по алгоритму Левенберга—Марквардта каждый раз показывает разную точность настройки, в том числе и самую неудовлетворительную. Это обычно связано с остаточным заполнением/переполнением оперативной памяти либо с занятостью ресурсов компьютера фоновыми программами.

Для обучения выбранной ИНС использовались данные за февраль—март 2002 года, представляющие собой последовательности значений критической частоты $f_{\text{кр}}(t)$, $\langle f_{\text{кр}}(t)\rangle$ (с устранённым сезонным эффектом), $\partial f_{\text{кр}}(t)/\partial t$, DI (разности между наблюдаемыми и средними значениями критической частоты), интенсивности рентгеновского излучения, скорости солнечного ветра V и концентрации частиц N в нём, модуля межпланетного магнитного поля и нормальной к плоскости эклиптики компоненты поля в солнечно-земной системе координат, а также индексов геомагнитной активности D_{st}, планетарной геомагнитной активности K_{p} и магнитной активности в полярных каспах PC. Всего было использовано свыше двух тысяч значений каждого параметра с дискретизацией 30 минут.

Кроме того, для поиска подходящего обучающего множества были реализованы физически обоснованные комбинации из входных последовательностей. Необходимо было выяснить, как скажется на общей эффективности прогноза добавление тех или иных параметров, а также какое время задержки для параметров солнечного ветра и межпланетного магнитного поля наиболее приемлемо. В каждом случае для прогноза на 1, 2, 3, 12 и 24 часа были проведены серии численных опытов с экспериментальным поиском наилучшей комбинации из имеющихся временны́х последовательностей. Очевидно, что привлечение сразу всех параметров не является рациональным, т. к. излишняя загрузка нейросети приводит к возрастанию времени адаптации к новым данным и продолжительному обучению. Поэтому была поставлена конкретная цель — определить ограниченный круг входных массивов, привлечение которых обеспечит высокое качество прогноза на каждый из исследуемых промежутков времени.

Обучение сетей проводилось по 1 661 значению (для каждого входного параметра). Прогнозирование предполагало получение $f_{\text{кр}}(t + \Delta t)$ на выходе ИНС для последующих 830 значений с уменьшением этого числа каждый раз на 2 при увеличении времени прогноза Δt на 1 час. Прогноз проводился на час, на два часа и т. д.

2. ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ ДЛЯ ПРОГНОЗА НА 1 ЧАС

Первая серия численных экспериментов проводилась с целью получения прогноза на 1 час. Первый численный эксперимент был направлен на выполнение прогноза по предыстории только самого процесса. Обучающая последовательность содержала значения двух параметров — критической частоты $f_{\text{кр}}(t)$ и её первой производной $\partial f_{\text{кр}}(t)/\partial t$. Выбор производной обусловлен желанием акцентировать «внимание» сети на нерегулярных изменениях критической частоты. В данном случае отсутствует физический смысл задержки на входе сети, и блок задержки был отключён. Результаты прогноза оказались удовлетворительными: PE $= 92\,\%$, $R = 0{,}96$. Для примера на рис. 2 приведены участки прогнозируемой и реальной последовательностей значений критической частоты за четверо суток из всего тестового интервала.

Как видно из рис. 2, прогноз достаточно успешен, но анализ поведения коэффициента корреляции в течение суток показал, что достоверность прогноза в вечернее и ночное время низка. Вариации суточного хода точности прогноза, скорее всего, связаны с различными физическими процессами, определяющими уровень электронной концентрации N_{e} в дневные и ночные часы в верхней ионосфере. Если днём уровень N_{e} в F-области определяется интенсивностью солнечного ионизирующего излучения, то в ночные часы он обеспечивается потоками плазмы из протоносферы и термосферными ветрами [21]. Определённую роль в снижении точности прогноза в вечерние и ночные часы может играть появление диффузных отражений от F-слоя. В переходное время суток (утро, вечер), когда происходит перестройка ионосферы, сочетание различных факторов, относительная роль которых подвержена заметным вариациям, находит своё отражение в снижении точности прогноза критической частоты ионосферы. На рис. 3 приведена суточная динамика коэффициента R.

Для повышения точности прогноза в вечерние и ночные часы при прогнозе на 1 час было

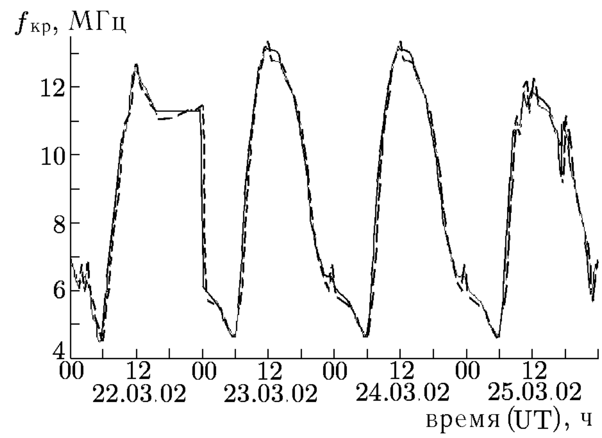

Рис. 2. Результат прогноза на 1 час для интервала с 00:00 UT 22 марта по 23:59 UT 25 марта 2002 года. Входной массив — последовательность значений критической частоты и её производной. Сплошной линией показана реальная последовательность, штриховой — прогноз

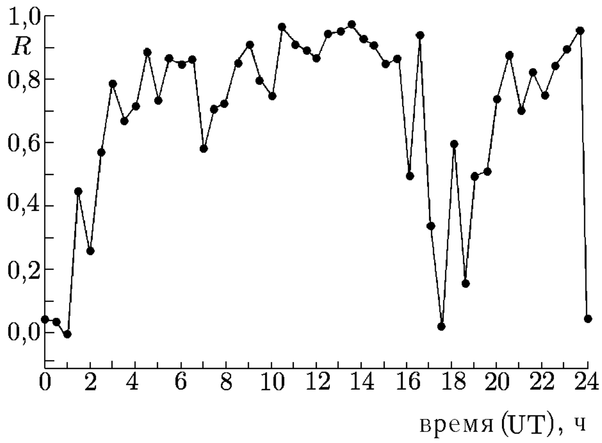

Рис. 3. Суточная динамика коэффициента корреляции R между реальными и спрогнозированными значениями критической частоты при прогнозе на 1 час. Входной массив — последовательность значений критической частоты и её производной

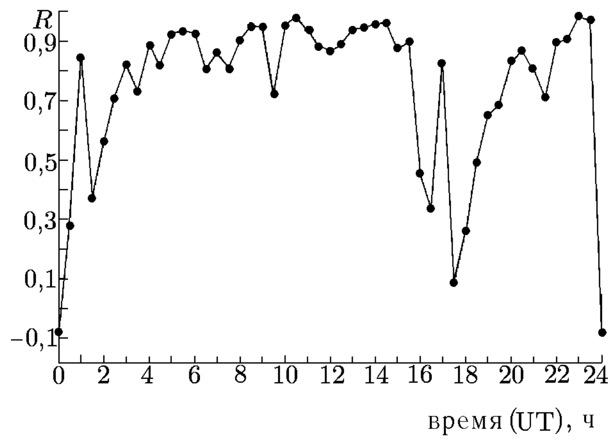

Рис. 4. Суточная динамика коэффициента корреляции между реальными и спрогнозированными значениями критической частоты при прогнозе на 1 час. Входной массив — последовательность значений критической частоты, её производной и индекса D_{st}

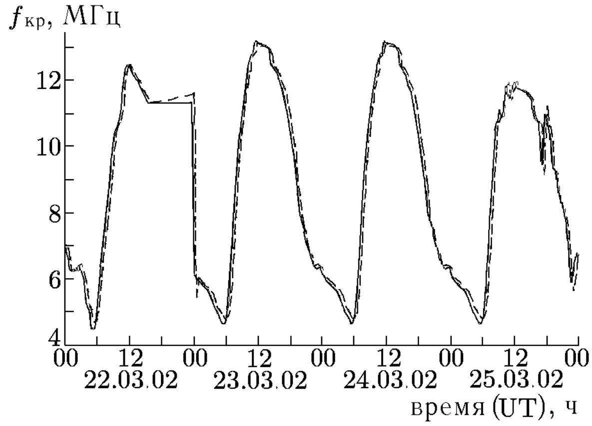

Рис. 5. Результат прогноза на 1 час для интервала с 00:00 UT 22 марта по 23:59 UT 25 марта 2002 года. Входной массив — последовательность значений критической частоты, её производной, модуля межпланетного магнитного поля и его компоненты, нормальной к плоскости эклиптики, с задержкой 2 часа 30 минут. Сплошной линией показана реальная последовательность, штриховой — прогноз

решено привлечь дополнительный параметр, в качестве которого в последующих численных экспериментах серии выступали индекс глобальной геомагнитной активности D_{st}, параметр DI и гидродинамическое давление солнечного ветра, полученное перемножением концентрации межпланетной плазмы N и квадрата скорости солнечного ветра V^2. В последнем случае параметры солнечного ветра подавались с разной задержкой (1, 2 или 3 часа), и исследовалась эффективность этой задержки. В итоге были получены следующие результаты. Учёт индекса D_{st} повышает PE с 92 % до 92,8 % при незначительном повышении общего коэффициента корреляции с 0,96 до 0,97. Учёт последовательности значений DI повышает PE до 93,1 % с повышением общего коэффициента корреляции до 0,97. Введение гидродинамического давления солнечного ветра с задержкой в 1 час не оказало существенного влияния на качество прогноза, но при двухчасовой задержке возросли PE до 93,1 % и R до 0,97. При задержке рассматриваемого массива на 3 часа нейросеть перестала обучаться, что косвенно свидетельствует о излишне большой для данной задачи задержке этого параметра.

На рис. 4 показана динамика коэффициента корреляции при дополнительной подаче вышеописанных параметров по очереди (отличия графика при разных дополнительных данных несущественны). Сопоставление с результатами, представленными на рис. 3, показало, что в вечернее время в районе 17–18 часов прогноз заметно улучшился, а в дневные часы он стал более стабилен.

Дальнейший поиск приемлемого обучающего множества привёл к нахождению ещё более удачной комбинации входных данных. Высокую точность прогноза на один час обеспечили последовательности значений модуля мажпланетного магнитного поля и его компоненты, нормальной к плоскости эклиптики (B_z), критической частоты и её первой производной. При этом наиболее удачной оказалась задержка для параметров межпланетного магнитного поля, равная 2,5 часа. Точность прогноза в этом случае составила PE = 93,9 %, $R = 0{,}97$. На рис. 5 показан результат прогноза. Как и в предыдущем опыте, трёхчасовая задержка для параметров межпланетного магнитного поля приводит к прекращению обучения сети.

Реализация других комбинаций из входных массивов не обеспечила улучшения результатов. Вместе с тем было подтверждено оптимальное время задержки для параметров солнечного ветра и межпланетного магнитного поля, равное 2,5 часа. Увеличение этого времени до 3 часов и более обычно ведёт к незначительному снижению качества прогнозирования, что подтверждает правильность выбора оптимальной задержки.

3. ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ ДЛЯ ПРОГНОЗА НА 2 И 3 ЧАСА

Данная серия численных экспериментов была проведена с целью изучения поведения нейронной сети при прогнозе на больший промежуток времени — на 2 и 3 часа. Как и в первой серии, анализ работы сети начат с эксперимента, направленного на её обучение по двум входным последовательностям: значениям критической частоты и её первой производной. Результаты оказались сильно отличными друг от друга: при прогнозе на 2 часа PE = 78,2 % и $R = 0{,}89$, при прогнозе на 3 часа PE = 58,7 % и $R = 0{,}78$. Следует заметить, что такая большая разница в точности прогноза при увеличении времени всего на 1 час с самого начала поставила под сомнение возможность дальнейшего увеличения времени прогноза до 12 и 24 часов. Как будет показано ниже, эти опасения не оправдались. На рис. 6 показан результат прогноза на 2 часа, а на рис. 7 — на 3 часа.

Динамика коэффициента корреляции для обоих вариантов прогноза показала низкое качество прогноза в некоторые интервалы суток. На рис. 8, 9 показаны полученные вариации коэффициента R.

Последующие численные эксперименты серии были посвящены поиску добавочного входного массива. При расчёте прогноза на 2 часа добавление значений индекса глобальной геомагнитной активности D_{st} повысило PE с 78,2 % до 80,6 %, и общий коэффициент корреляции R увеличился с 0,89 до 0,91. Введение последовательности значений DI не оказало существенного влияния, учёт значений интенсивности рентгеновского излучения или индексов K_{p} (планетарной геомагнитной активности) и PC (магнитной активности в полярных каспах) также не сказывается на повышении качества и точности прогноза. Комбинация из последовательности значений критической частоты, её первой производной по этой последовательности, модуля межпланетного магнитного поля и его компоненты B_z (без задержки) обеспечила качество прогноза с PE = 79,7 и $R = 0{,}91$. Однако параметры межпланетного магнитного поля и солнечного ветра способны оказать своё влияние на общую ситуацию в ионосфере лишь спустя некоторое время. Из физических соображений эта временна́я задержка должна находиться в пределах от одного до трёх часов.

Сконструированная модель нейросети позволяла симулировать задержку для параметров околоземного пространства по отношению к наблюдаемым критическим частотам с шагом 30 минут.

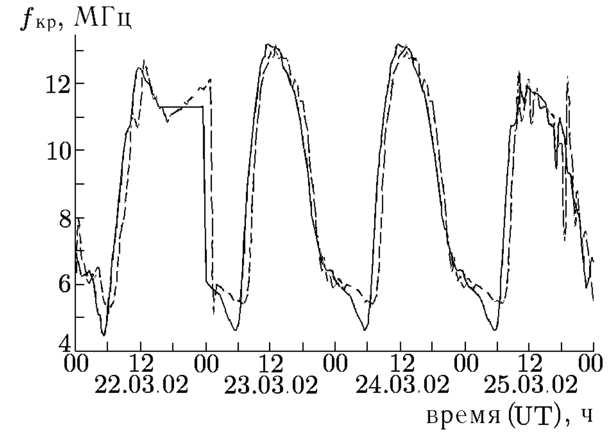

Рис. 6. Результат прогноза на 2 часа для интервала с 00:00 UT 22 марта по 23:59 UT 25 марта 2002 года. Входной массив — последовательность значений критической частоты и её производной. Сплошной линией показана реальная последовательность, штриховой — прогноз

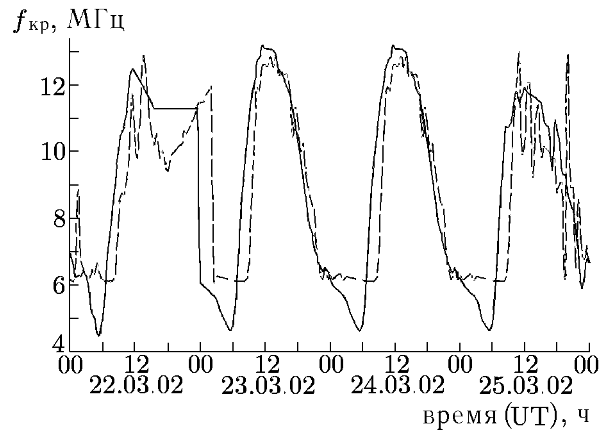

Рис. 7. Результат прогноза на 3 часа для интервала с 00:00 UT 22 марта по 23:59 UT 25 марта 2002 года. Входной массив — последовательность значений критической частоты и её производной. Сплошной линией показана реальная последовательность, штриховой — прогноз

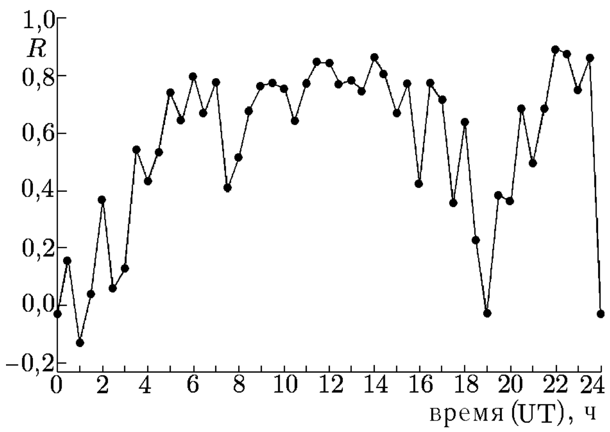

Рис. 8. Суточная динамика коэффициента корреляции между реальными и спрогнозированными значениями критической частоты при прогнозе на 2 часа. Входной массив — последовательность значений критической частоты и её производной

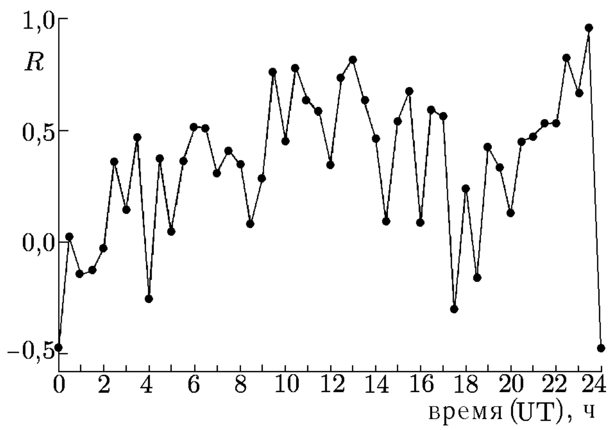

Рис. 9. Суточная динамика коэффициента корреляции при прогнозе на 3 часа. Входной массив — последовательность значений критической частоты и её производной

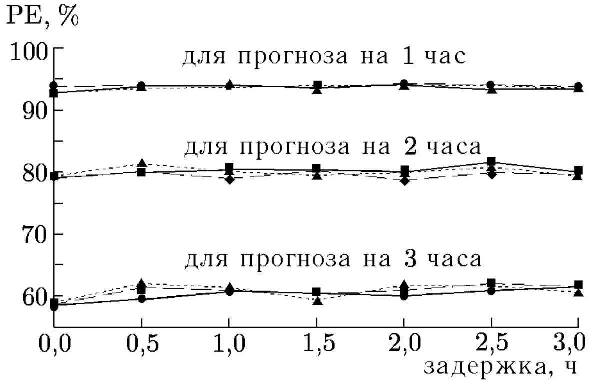

Рис. 10. Изменение максимальных достигнутых значений PE при разных задержках для параметров межпланетного магнитного поля и солнечного ветра. Первая серия экспериментов — сплошная линия, вторая — штриховая, третья — пунктирная

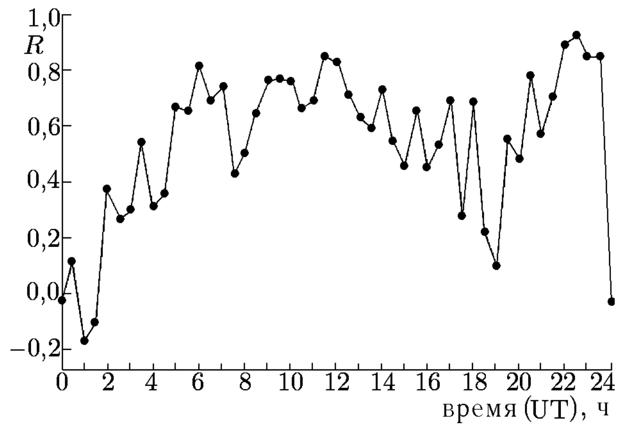

Рис. 11. Суточная динамика коэффициента корреляции при прогнозе на 2 часа. Входной массив — последовательности значений критической частоты, её производной, модуля магнитного поля и компоненты B_z с задержкой 1 час

Далее анализировалось введение временно́й задержки для параметров межпланетного магнитного поля и солнечного ветра. Результаты этого исследования приведены на рис. 10, где представлены максимальные коэффициенты PE, полученные в ходе серий численных экспериментов.

Выяснение оптимального времени задержки для каждого из вариантов прогноза проводилось в несколько этапов. На первом этапе входной порт нейросети регулировался таким образом, чтобы симулировать соответствующую задержку. На втором этапе при каждой фиксированной задержке нейронная сеть заново обучалась и тестировалась, т. е. прогнозировала тестовый интервал, 10 раз подряд, и подсчитывался коэффициент PE. Далее процедура повторялась для следующего времени задержки параметров межпланетного магнитного поля и солнечного ветра. На третьем этапе из каждого полученного 10-элементного массива значений PE выбиралось максимальное, которое представлено на рис. 10. Указанный алгоритм применялся для каждого варианта прогноза трижды, поэтому для каждого времени задержки получены три кривые. Как следует из рис. 10, для прогноза на 1 час наиболее приемлемой является задержка 1,5÷2,5 часа, для прогноза на 2 и 3 часа достаточна задержка 0,5 часа. Это говорит о соответствующей физической задержке в ионосферных событиях по отношению к параметрам межпланетного магнитного поля. Отметим, что по данным [22] задержка в 2 ÷ 4 часа наблюдается между моментами максимального отклонения критической частоты от среднего значения и максимального значения геомагнитного индекса D_{st}. Наличие экстремумов для бо́льших времён задержки характеризует сложные релаксационные процессы в ионосфере.

Комбинация из последовательности значений критической частоты, её первой производной по последовательности, модуля межпланетного магнитного поля и его компоненты B_z (с задержками 1 час) оказалась самой эффективной и повысила PE до 80,8 % и общий коэффициент корреляции до 0,91.

График, приведённый на рис. 11, иллюстрирует влияние общего повышения качества прогноза на суточную динамику коэффициента корреляции для двухчасового прогноза с задержкой 1 час для параметров межпланетного магнитного поля. Сравнение рис. 8 и 11 показывает улучшение прогноза в утренние и дневные часы и в вечернее время вблизи 19 часов.

При решении задачи повышения точности трёхчасового прогноза поиск оптимального обучающего множества привёл к следующим результатам. Учёт значений D_{st} повышает PE с 58,7 % до 60 % и R с 0,78 до 0,79; учёт значений DI и гидродинамического давления не оказывает влияния на прогноз. Ранее найденная удачная комбинация для двухчасового прогноза (последовательности значений критической частоты, производной по последовательности критических частот, модуля межпланетного магнитного поля и его компоненты B_z (с задержками 1 час)) не повышает качество трёхчасового прогноза.

Анализ суточной динамики коэффициента корреляции для трёхчасового прогноза показал незначительное изменение по сравнению с рис. 9. Наблюдается только общее повышение точности прогноза при добавлении новых входных данных.

4. ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ ДЛЯ ПРОГНОЗА НА 12 И 24 ЧАСА

Одним из достоинств метода ИНС является возможность его использования для автоматизированного прогноза ионосферных параметров с постоянной подстройкой под изменяющиеся внешние условия и учётом предыстории изучаемого процесса. В связи с этим важной частью данного исследования является постановка численных экспериментов по долговременному прогнозу — на 12 и 24 часа. Ниже представлены результаты прогнозирования по предыстории процесса.

При прогнозе на 12 часов при подаче на вход последовательности значений критической частоты и её производной точность прогноза составила PE $=57\,\%$ и $R=0{,}8$, а при прогнозе на сутки при тех же входных данных точность оказалась выше и составила PE $=80{,}3\,\%$ и $R = 0{,}9$ (см. рис. 12 и 13). В том, что прогноз на 24 часа оказался более достоверным, нет ничего удивительного. Дело в том, что обучение нейросети и дальнейшая её работа по прогнозу на 24 часа заключалась в пересчёте каждого значения критической частоты из последовательности на близкое значение из-за простого сдвига на сутки. Такой результат близок к выводам «наивного» прогноза, который выполняется заменой прогнозируемого значения на предыдущее.

Рис. 14 иллюстрирует зависимость качества предсказания критической частоты от вариантов прогнозирования: только по предыстории процесса, по предыстории, дополненной параметрами межпланетного магнитного поля и солнечного ветра, и методом «наивного» прогноза.

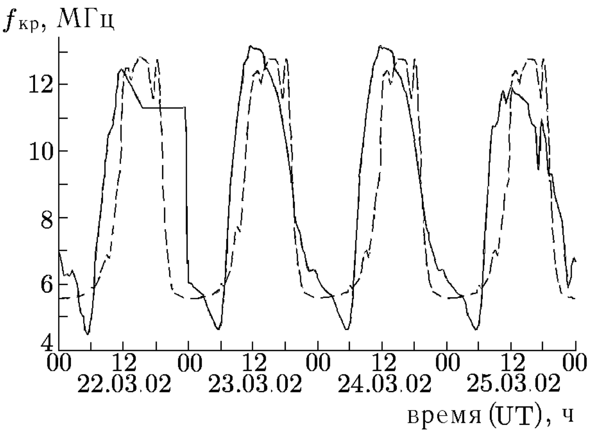

Рис. 12. Результат прогноза на 12 часов для интервала с 00:00 UT 22 марта по 23:59 UT 25 марта 2002 года. Входной массив — последовательность значений критической частоты и её производной. Сплошной линией показана реальная последовательность, штриховой — прогноз

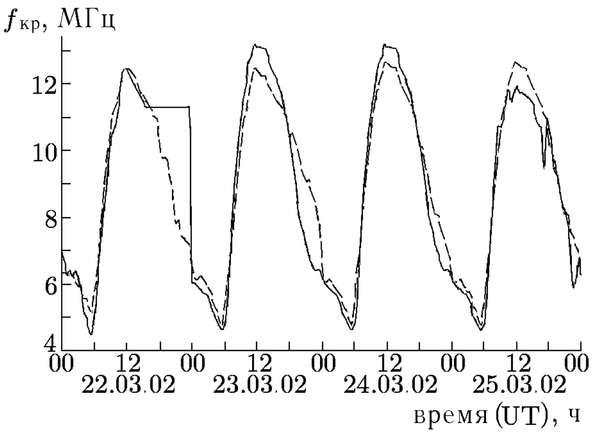

Рис. 13. Результат прогноза на сутки для интервала с 00:00 UT 22 марта по 23:59 UT 25 марта 2002 года. Входной массив — последовательность значений критической частоты и её производной. Сплошной линией показана реальная последовательность, штриховой — прогноз

Высокие отрицательные значения коэффициента корреляции при прогнозировании методом «наивного» прогноза на 12 и 32 часа соответствуют прямой уравнения регрессии, почти перпендикулярной к прямой уравнения регрессии для максимальной корреляции. В этом случае можно говорить о том, что корреляции нет, и «наивный» прогноз на этот период недопустим.

Дальнейший поиск приемлемого обучающего множества привёл к следующим результатам. Для 12-часового прогноза удачными оказались комбинации из входных последовательностей. Так, учёт значений критической частоты, гидродинамического давления (задержка 2,5 часа), производной критической частоты по последовательности и индекса D_{st} способствовал повышению PE с 57 % до 57,1 % и R с 0,8 до 0,9. Учёт значений критической частоты, её производной, модуля межпланетного магнитного поля и его компоненты B_z с задержкой 2,5 часа и индекса D_{st} повысили PE до 58 % и R до 0,9.

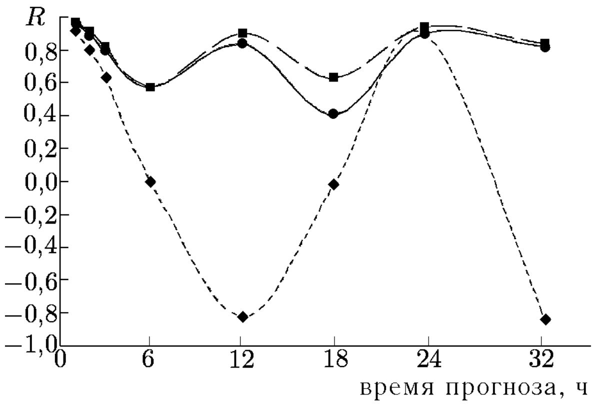

Рис. 14. Зависимость качества предсказания от варианта прогноза. Пунктиром показан «наивный» прогноз, сплошной линией — прогноз нейросети при обучении только по предыстории, штриховой линией — максимальное достигнутое качество прогноза с помощью нейросети

Качество 24-часового прогноза удалось повысить лишь за счёт единственной комбинации из обучающих данных. Таким набором оказался массив предыстории процесса — последовательность значений критической частоты, её первой производной и индекса DI (разности между наблюдаемым и средним значениями критической частоты). Эта комбинация значительно повысила достоверность прогноза, подняв значение PE с 80 % до 89 % и R с 0,9 до 0,94. Все остальные комбинации и включение других параметров не привнесли полезной информации в процесс обучения нейросети и не смогли повлиять на суточный прогноз.

Как и в других численных экспериментах, снова было подтверждено, что при задержке для параметров межпланетного магнитного поля и солнечного ветра относительно остальных 3 часа и более невозможно качественно обучить нейросеть, а оптимальным временем задержки является 2,5 часа.

5. ВЫВОДЫ И ПЕРСПЕКТИВЫ

В работе на основе технологии искусственных нейронных сетей разработаны алгоритмы прогнозирования критической частоты ионосферного слоя F_2 на 1, 2, 3 и более часов. Для этого проведён поиск соответствующего обучающего множества и архитектуры ИНС. Основным результатом проведённого комплексного исследования работы созданной ИНС явилось увеличение эффективности прогноза при использовании параметров солнечного ветра, межпланетного магнитного поля и индексов геомагнитной возмущённости. Это, в свою очередь, позволило установить особенности физического процесса, определяющего поведение критической частоты и, прежде всего, выяснить его характерное время. Практическая ценность выполненной работы заключается в возможности применения её результатов для оперативного уточнения ионосферной модели.

В результате был установлен ряд положений, касающихся разработки ИНС, и сделаны практические выводы, имеющие прикладное значение:

1) Применение двухслойной нейронной сети с обратными связями Элмана благодаря нелиней-

ной внутренней памяти является наиболее подходящим инструментом для прогноза временны́х последовательностей. Алгоритм обучения ИНС Левенберга—Марквардта на базе метода обратного распространения ошибки зарекомендовал себя как наиболее действенный при работе с большим числом входных массивов (3 и более). Именно его применение позволило настроить параметры нейросети таким образом, что при подаче новых, «незнакомых» для сети тестовых данных она могла быстро адаптироваться в новых условиях. Успешность работы ИНС при прогнозировании ионосферных параметров в режиме реального времени во многом будет зависеть от способности сети адаптироваться к динамически изменяющимся внешним факторам.

2) Прогноз критической частоты на 1 час может быть успешно проведён после обучения ИНС по двум потокам входных данных — последовательностям значений критической частоты и её первой производной. Точность прогноза в этом случае составляет PE = 92 %, $R = 0{,}96$. Понижение точности прогноза в ночные и вечерние часы можно объяснить особенностью поведения ионосферного слоя F_2, обусловленной перестройкой ионосферы.

3) При расчёте прогноза на 1 час учёт значений параметров $D_{\rm st}$ либо DI способствует повышению PE приблизительно на 1 %. Такому же приросту PE соответствует введение с задержкой 1 час в обучающую последовательность значений гидродинамического давления или модуля и компоненты B_z магнитного поля. Добавленные параметры повышают точность прогноза в ночные и вечерние часы.

4) Наиболее эффективная задержка для параметров межпланетного магнитного поля и солнечного ветра относительно остальных данных составляет 2 часа. Задержка 3 и более часов затрудняет обучение нейронной сети, что проявляется в низком качестве прогноза. Этот результат согласуется с физической обоснованностью задержки.

5) При расчёте прогноза на 2 часа учёт индекса глобальной геомагнитной активности $D_{\rm st}$ к последовательности значений критической частоты и её производной повышает PE с 78 % до 80 %. Учёт значений модуля межпланетного магнитного поля и его компоненты B_z (с задержками 1 час) оказалось наиболее эффективным и повысило PE до 81 %.

6) При решении задачи повышения точности трёхчасового прогноза учёт значений $D_{\rm st}$ в дополнение к последовательности значений критической частоты и её производной повышает PE с 59 % до 60 % и R с 0,78 до 0,8. При этом учёт DI или гидродинамического давления не оказывает существенного влияния.

7) Возможен более долговременный прогноз критической частоты, в некоторых случаях с эффективностью до 89 %. В целом на качество и точность прогноза на изучаемые промежутки времени положительно влияет учёт параметров межпланетного магнитного поля (с задержкой 2,5 часа) и солнечного ветра в форме гидродинамического давления, а также индекса глобальной геомагнитной активности $D_{\rm st}$, что приводит к повышению PE приблизительно на 1 %. Точность прогноза на сутки возрастает на 8 %, если в обучающую последовательность включены значения индекса DI или усреднённая последовательность критических частот.

8) Число обучающих параметров должно отвечать физическим представлениям о процессах, происходящих в солнечно-ионосферных связях, и быть достаточно ограниченным, чтобы не происходила перегрузка сети — адаптация к большому объёму данных и рост времени обучения.

Таким образом применение искусственных нейронных сетей может обеспечить универсальный автоматизированный прогноз параметров ионосферных процессов по непрерывным спутниковым данным, магнитным измерениям на земной поверхности и результатам вертикального и наклонного зондирования ионосферы. В настоящее время необходимые данные можно получать в режиме реального времени. В перспективе вполне возможно создание вычислительного комплекса, включающего в себя ИНС и сервис, позволяющий с высокой надёжностью в автоматическом режиме прогнозировать параметры ионосферного коротковолнового канала по данным наклонного зондирования на несколько часов вперёд и устанавливать связь характеристик радиоканалов с изменяющейся солнечно-геомагнитной обстановкой, учитывающей явления космической погоды. Сочетание различных методов прогнозирования и коррекции ионосферных данных с использованием как технических средств зондирования, так и разработанных и программно реализованных алгоритмов расчёта, существенно расширяет возможности обеспечения высокого качества прогноза параметров ионосферного коротковолнового радиоканала.

Прогноз параметров ионосферы на основе метода ИНС позволяет получать достоверные результаты без привлечения аналитического описания физического процесса. Для этого обученной сети необходимо всего лишь обеспечить непрерывный поток необходимых данных. Выполненный прогноз уже использован для коррекции ионосферной модели и синтеза ионограмм наклонного зондирования на трассе Инскип (Англия)—Нижний Новгород [15]. Сопоставление экспериментальных и расчётных ионограмм продемонстрировало увеличение качества прогнозируемых ионограмм по сравнению с долгосрочным прогнозом.

В заключение заметим, что развитый применительно к прогнозированию ионосферных параметров метод ИНС предполагается использовать для прогноза основных параметров ионосферного коротковолнового радиоканала — максимальной применимой частоты и оптимальной рабочей частоты по данным наклонного ЛЧМ-зондирования, полученным на трассе Инскип (Англия)—Ростов-на-Дону. Наблюдения на данной трассе ведутся круглосуточно уже более полугода, получен большой массив данных (около 50 000 ионограмм), необходимых для обучения нейронных сетей в различных геофизических условиях. На основе прогноза с использованием большого массива данных наклонного зондирования ионосферы будут сделаны оценки диапазона изменения ключевых параметров солнечного ветра и межпланетного магнитного поля, определяющих последовательность развития магнитных бурь, ионосферным откликом которых являются отрицательные возмущения электронной концентрации, приводящие к снижению точности прогнозирования и надёжности коротковолновой радиосвязи. Будут выработаны рекомендации для изменения стратегии работы сети ЛЧМ-зондов во время возмущений (изменение скважности зондирования, получение данных зондирования в реальном времени) для адаптации радиоэлектронных систем к текущему состоянию ионосферы. На наш взгляд, такая задача представляется сейчас весьма актуальной в связи с развёртыванием и наращиванием российской сети трасс наклонного ЛЧМ-зондирования для обеспечения эффективной работы радиоэлектронных систем различного назначения. В таком приложении метод искусственных нейронных сетей позволит оптимизировать работу сети ЛЧМ-зондов для обеспечения высокого качества прогноза на различные временны́е интервалы в различных геофизических условиях.

Работа выполнена при финансовой поддержке РФФИ (гранты № 02–05–64383 и 03–05–65137).

СПИСОК ЛИТЕРАТУРЫ

1. Круглов В. В., Борисов В. В. Искусственные нейронные сети. Теория и практика. М.: Горячая линия—Телеком, 2000.
2. Дьяконов В. П., Круглов В. В. Математические пакеты расширения MATLAB. Специальный справочник. СПб.: Питер, 2001.
3. Бархатов Н. А., Королёв А. В., Пономарёв С. М., Сахаров С. Ю. // Изв. вузов. Радиофизика. 2001. Т. 44, № 9. С. 806.
4. Бархатов Н. А., Беллюстин Н. С., Левитин А. Е., Сахаров С. Ю. // Изв. вузов. Радиофизика. 2000. Т. 43, № 5. С. 385.
5. Бархатов Н. А., Левитин А. Е., Сахаров С. Ю. // Геомагнетизм и аэрономия. 2002. Т. 42, № 2. С. 195.

6. Munsami V. // J. Geophys. Res. A. 2000. V. 105, No. 12. P. 27 833.
7. Gleisner H., Lundstedt H. // J. Geophys. Res. A. 2001. V. 106, No. 5. P. 8 425.
8. Lundstedt H., Wintoft P. // Ann. Geophys. 1994. V. 12, No. 1. P. 19.
9. Gleisner H., Lundstedt H., Wintoft P. // Ann Geophys. 1996. V. 14, No. 7. P. 679.
10. Sutcliffe P. R. // Ann. Geophys. 1997. V. 15, No. 10. P. 1 257.
11. Sutcliffe P. R. // Ann. Geophys. 2000. V. 18, No. 1. P. 120.
12. Conway A. J., Macpherson K. P., Blacklaw G., Brown J. C. // J. Geophys. Res. A. 1998. V. 103, No. 12. P. 29 733.
13. Fessant F., Bengio S., Collobert D. // Ann Geophys. 1996. V. 14, No. 1. P. 20.
14. Wintoft P., Lundstedt H. // J. Geophys. Res. A. 1999. V. 104, No. 4. P. 6 729.
15. Бархатов Н. А., Валов В. А., Макаров А. В. и др. // Труды IX Междунар. научно-техн. конф. «Радиолокация, навигация и связь», 22–24 апреля 2003 г., Воронеж. Т. 3. С. 1 853.
16. Бархатов Н. А., Левитин А. Е., Рябкова Г. А. // Солнечно-земная физика. Иркутск, 2002. Вып. 2 (115). С. 104.
17. http://www.wdc.rl.ac.uk.
18. http://www.nssdc.gsfc.nasa.gov.
19. Дьяконов В. П. MATLAB 6. Учебный курс. СПб.: Питер, 2001.
20. Медведев В. С., Потёмкин В. Г. Нейронные сети. MATLAB 6. М.: Диалог-МИФИ, 2002.
21. Брюнелли Б. Е., Намгаладзе А. А. Физика ионосферы. М.: Наука, 1988. 527 с.
22. Зевакина Р. А., Киселёва М. В. // Прогнозирование ионосферы и условий распространения радиоволн. М.: Наука, 1985. С. 39.

Поступила в редакцию 29 октября 2003 г.; принята в печать 11 февраля 2004 г.

FORECASTING THE CRITICAL FREQUENCY OF THE IONOSPHERIC F_2 LAYER BASED ON ARTIFICIAL NEURAL NETWORKS

N. A. Barkhatov, S. E. Revunov, and V. P. Uryadov

We employ artificial neural networks (ANNs) to develop an algorithm of forecasting the critical frequency of the ionospheric F_2-layer for 1, 2, 3, 12, and 24 h. A search for suitable training set and ANN architecture is performed. The use of auxiliary input data, such as the solar-wind and the interplanetary magnetic-field parameters, as well as the geomagnetic-activity indices, made it possible not only to improve the prediction efficiency but also to find some regularities in the evolution of the critical frequency. The results of this work can be applied to the prompt correction the ionosphere model, aimed at improving the ionospheric HF radio communication.

УДК 551.510.536

ЗАХОДНО-ВОСХОДНЫЕ ХАРАКТЕРИСТИКИ СПОРАДИЧЕСКИХ СЛОЁВ ИОНИЗАЦИИ В НИЖНЕЙ ИОНОСФЕРЕ, НАБЛЮДАЕМЫЕ МЕТОДОМ РЕЗОНАНСНОГО РАССЕЯНИЯ РАДИОВОЛН НА ИСКУССТВЕННЫХ ПЕРИОДИЧЕСКИХ НЕОДНОРОДНОСТЯХ ИОНОСФЕРНОЙ ПЛАЗМЫ

Н. В. Бахметьева, В. В. Беликович, Л. М. Каган, А. А. Понятов

Научно-исследовательский радиофизический институт, г. Нижний Новгород, Россия

Приведены результаты исследований спорадических слоёв ионизации в нижней ионосфере по наблюдениям в заходно-восходные часы в летние месяцы 2000 и 2001 годов. Измерения выполнены на нагревном комплексе «Сура» (НИРФИ, г. Нижний Новгород). На основе анализа высотно-временно́й зависимости амплитуды, времени релаксации и фазы рассеянного сигнала получены новые сведения о формировании и динамике среднеширотного спорадического слоя E в вечерние и ночные часы, приведены результаты измерения скорости вертикального движения плазмы, сделаны оценки относительной молекулярной массы и общей концентрации металлических ионов, сгонка которых за счёт ветрового сдвига могла быть причиной образования наблюдавшихся спорадических слоёв.

ВВЕДЕНИЕ

В работе [1] приведены результаты исследования неоднородной структуры нижней ионосферы методом искусственных периодических неоднородностей (ИПН), полученные по данным измерений в дневные часы в августе 1999 года. В частности, получены и проанализированы яркостные записи (высотно-временны́е зависимости) рассеянного сигнала, высотные профили и временны́е зависимости его амплитуды и времени релаксации, исследованы высотно-временны́е вариации температуры и плотности атмосферы, скорости вертикального движения плазмы и вертикальной компоненты турбулентной скорости. Приведены новые сведения о рассеянии радиоволн на ИПН, идентифицированы различные ионосферные образования, включая спорадические слои и крупномасштабные ионосферные неоднородности, проанализирована динамика этих неоднородностей, высо́ты их появления и временны́е характеристики. Было показано, в частности, что амплитуда рассеянного сигнала во многих случаях определяется интерференцией волн от ИПН и естественных ионосферных образований, что вызывает быстрые вариации сигнала и изменения высотной зависимости его амплитуды и времени релаксации.

В 2000 и 2001 годах были проведены новые эксперименты в вечерние и ночные часы, включавшие периоды захода и восхода Солнца, а именно с 19:00 до 05:00 LT (местного времени) 16, 17 августа 2000 года (с перерывом с 22:00 до 02:00 LT) и 15, 16 июня 2001 года с 18:00 до 05:30 LT. Заметим, что согласно многолетним ионосферным наблюдениям, проводимым различными методами, на июнь в северном полушарии приходится главный максимум вероятности появления спорадического слоя E ($E_{\rm s}$) [2], а в целом 80 % отражений от слоя $E_{\rm s}$ регистрируется в течение года с мая по август. Задачей экспериментов 2000 года было исследование методом ИПН ночной ионизации D- и E-областей, включая спорадические слои. Одним из полученных результатов было наблюдение сигналов от E-области, рассеянных ИПН, с максимальной амплитудой до 50 дБ после захода Солнца на этих высотах и появление их в довосходные часы. Можно предположить, что сигналы, рассеянные ИПН, а значит, и ионизация E-области существовали в течение всей ночи (об этом сообщалось в [3]). Другим результатом, полученным на основе анализа яркостных записей амплитуды и фазы рассеянного сигнала, был вывод о том, что в восходно-заходные часы спорадические образования различных пространственно-временны́х масштабов наблюдаются в верхней части D-области и в нижней части E-области практически постоянно. По данным наблюдений 2000 года отражения от этих слоёв могли быть относительно «гладкими», а могли иметь и выраженную «облачную» структуру. Результаты оказались интересными, и в июне 2001 года эксперимент был повторён, причём наблюдения продолжались практически всю ночь. Совместные результаты этих двух экспериментов представлены в настоящей работе.

1. ПОСТАНОВКА ЭКСПЕРИМЕНТА

Метод исследования основан на создании в ионосфере искусственных периодических неоднородностей электронной концентрации в поле мощной стоячей волны от наземного радиопередатчика, локации ИПН пробными (т. е. слабыми, не возмущающими среду распространения) радиоволнами и измерении амплитуды и фазы обратно рассеянного сигнала. Суть метода заключается в том, что в поле мощной стоячей радиоволны, образующейся в результате интерференции падающей на ионосферу и отражённой от неё волн, возникают искусственные периодические неоднородности ионосферной плазмы. В пучностях стоячей волны происходит нагрев электронного газа, в результате чего образуется периодическая структура с пространственным периодом, равным половине длины мощной радиоволны λ (рис. 1*а*). Температурные неоднородности, в свою очередь, формируют неоднородности электронной концентрации. Процессы образования и релаксации ИПН после выключения нагрева в каждой области ионосферы определяются различными физическими процессами: изменениями состава, плотности, температуры атмосферы, степени диссоциации и ионизации с высотой, что позволяет применять этот метод для диагностики ионосферы [4]. Метод резонансного рассеяния на ИПН позволяет регистрировать сигналы от сравнительно слабых неоднородностей (уже при относительной концентрации больше 10^2).

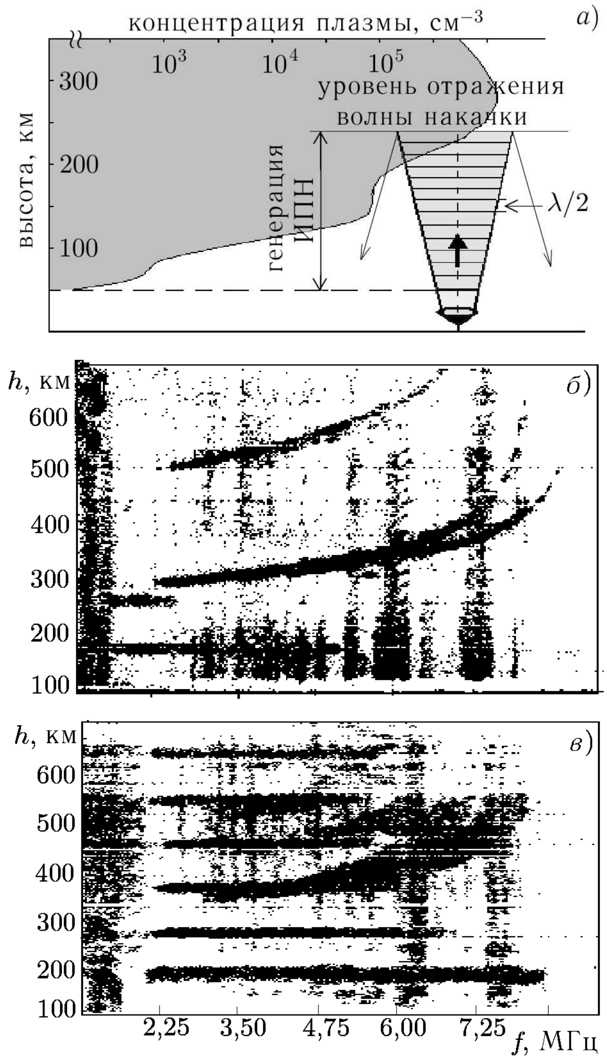

Рис. 1. Схема создания искусственных периодических неоднородностей ионосферной плазмы (*а*) и примеры ионограмм: 20:45:00 LT 16.08.2000 (*б*) и 23:15:00 LT 15.06.2001 (*в*)

Для создания ИПН использовался стенд «Сура» с координатами 56,1° с. ш., 46,1° в. д., эффективная мощность излучения в экспериментах 2000 и 2001 годов составила приблизительно 70 и 170 МВт соответственно. Каждый сеанс продолжался 20 с, из них в течение первых 3 с осуществлялся нагрев ионосферы радиоволной X-поляризации на частоте $f = 5{,}67$ МГц. За трёхсекундным нагревом следовала пауза с длительностью 17 с, в первые 4 секунды которой излучались пробные импульсы той же частоты с длительностью 25 мкс и частотой повторения 50 Гц. Антенна, состоящая из 12 синфазных диполей, принимала рассеянные сигналы X-поляризации. Усиление сигналов осуществлялось приёмником с полосой пропускания 40 кГц. Квадратурные компоненты рассеянного сигнала регистрировались в цифровом виде с шагом по высоте 1,4 км и кодировались 12-разрядным аналого-цифровым преобразователем. Далее на каждой высоте рассчитывались амплитуда и фаза сигнала, временны́е зависимости которых затем аппроксимировались линейными функциями вида $\ln A(t) = \ln A_0 - t/\tau$ и $\varphi(t) = \varphi_0 + 4\pi V t/\lambda$. Здесь λ — длина пробной (зондирующей) волны в плазме, τ — время релаксации сигнала, характеризующее время жизни ИПН после прекращения нагрева, V — скорость их вертикального движения (более подробно см. [4]). Уровень естественных шумов в разное время наблюдений составлял 10÷15 дБ, хотя были периоды с интенсивными импульсными помехами. Сканирование проводилось от высоты 50 км до высоты отражения мощной радиоволны. В настоящей работе все результаты приведены для нижней ионосферы, т. е. для интервала высот 50÷120 км.

Для контроля общего состояния ионосферы каждые 15 минут снимались ионограммы вертикального зондирования. Примеры двух ионограмм приведены на рис. 1*б*, *в*.

2. АМПЛИТУДА И ВРЕМЯ РЕЛАКСАЦИИ РАССЕЯННОГО СИГНАЛА

2.1. Результаты наблюдений 16 и 17 августа 2000 года

На рис. 2 представлены яркостные записи (высотно-временны́е зависимости) рассеянного сигнала для измерений 16 и 17 августа 2000 года. Здесь же тёмной линией показана зависимость от высоты времени захода (левая панель) и восхода (правая панель) Солнца на высотах ионосферы. Данные по амплитуде A и скорости вертикального движения V представлены в первоначальном виде, без предварительной обработки, а значения времени релаксации τ сглажено по высоте в интервале 2,8 км и пятиминутному интервалу времени. Отрицательные значения V соответствуют движению вверх. На рис. 2 на высотах 56÷80 км и 84÷115 км видны сигналы от ИПН в D- и E-областях соответственно.

Вначале несколько слов о сигналах от ИПН в D-области ионосферы. Достаточно слабые сигналы наблюдались ещё около часа после захода Солнца на этих высотах, затем исчезали и появились приблизительно через 30 минут после восхода. Обращает на себя внимание явление, определённое в [3] как заходно-восходная асимметрия высотно-временно́й зависимости рассеянного сигнала. Она выражается, в частности, в том, что на заходе Солнца сигналы от ИПН имеют бо́льшую амплитуду и занимают бо́льший интервал высот, чем на восходе, что хорошо видно на рис. 2*а*. Интервал высот между D- и E-областями, в котором рассеянные сигналы отсутствовали, достигал 7 км в дозаходные часы и 15 км после восхода Солнца. В [3] на основе качественных рассуждений и численных расчётов времени релаксации и амплитуды рассеянного сигнала по модели аэрономических процессов с одним отрицательным ионом показано, что эти особенности обусловлены увеличением скорости отлипания электронов вследствие повышенной концентрации атомарного кислорода. Отметим, что в D-области главную роль в процессе образования ИПН играет температурная зависимость коэффициента прилипания электронов к нейтральным молекулам [4].

По данным ионозонда критические частоты E-области составили 2,0÷2,5 МГц. На ионограммах виден полупрозрачный слой $E_{\rm s}$ с частотой экранирования $f_{\rm b} = 1{,}5\div 3{,}0$ МГц и предельной частотой отражения $f_0 = 1{,}5\div 5{,}0$ МГц. Сигналы от ИПН в E-области наблюдались в течение всего периода измерений. На яркостных записях каждого отдельного сеанса видны также спорадические сигналы характерного вида. Поскольку на ионограммах в это время наблюдался слой $E_{\rm s}$ (рис. 1*б*), эти сигналы можно было идентифицировать как рассеяние на неоднородностях по-

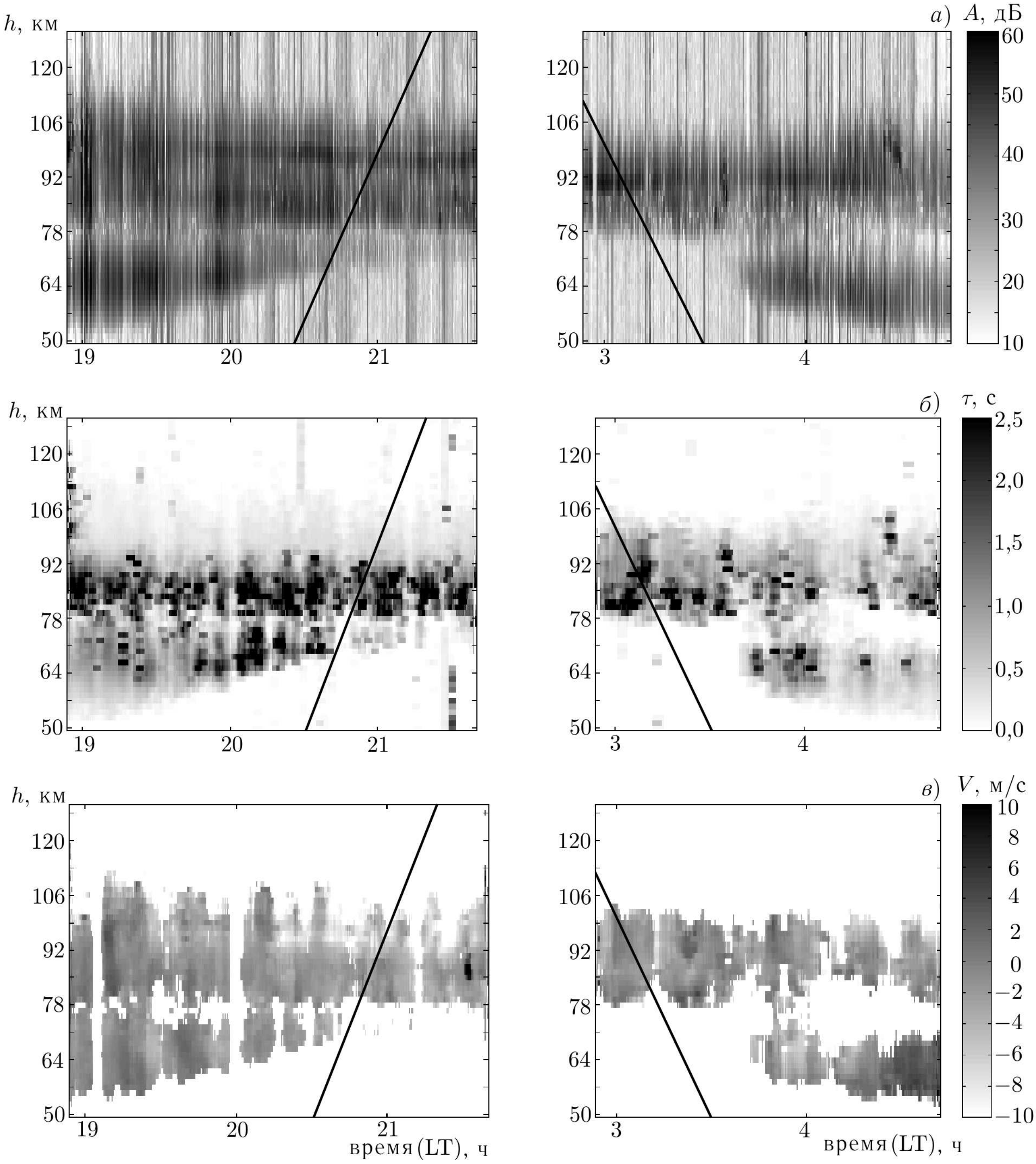

Рис. 2. Яркостные записи (высотно-временны́е зависимости) рассеянного сигнала, полученные 16 и 17 августа 2000 года: амплитуда (*а*), время релаксации (*б*) и скорость вертикального движения плазмы (*в*)

лупрозрачного спорадического слоя E. На левой панели рис. 2*а* в E-области хорошо видны два выделенных высотных интервала максимальной амплитуды рассеянного сигнала — вблизи 92 и 106 км. Амплитуда этих сигналов достигала 55 дБ, а время релаксации значительно превышало (до 5 раз) величины, характерные для ИПН в E-области (0,5÷2,5 с). Вблизи первой высоты практически всё время существовал ярко выраженный тонкий слой ионизации, который часто имел «облачную» структуру. Такие спорадические облака ионизации хорошо видны на 3-секундных

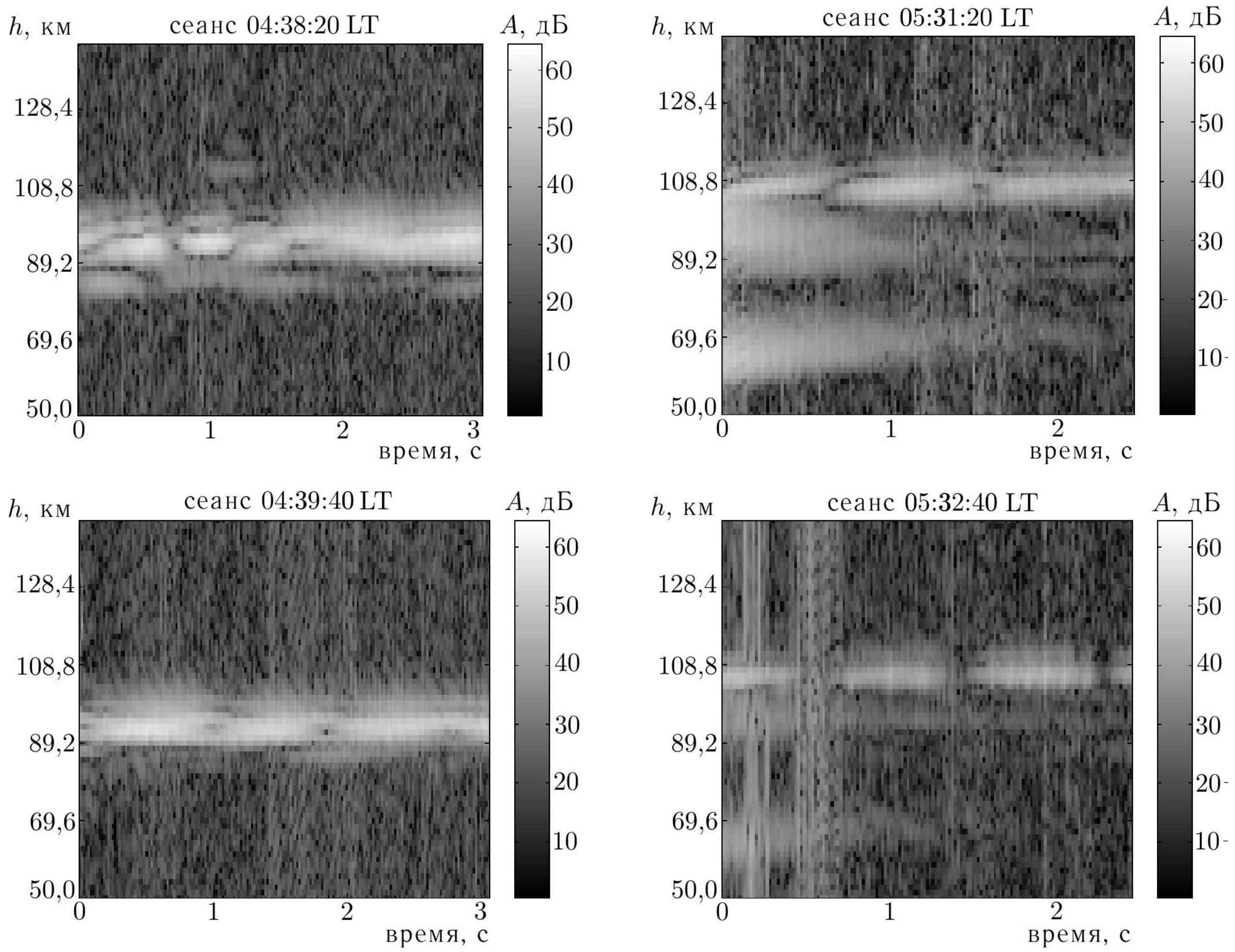

Рис. 3. Пример яркостных записей амплитуды сигнала, рассеянного на спорадическом слое E с выраженной «облачной» структурой — подслойного (левые панели) и возникшего выше максимума E-области (правые панели), полученных 17.08.2000

«снимках» резонансного рассеяния, приведённых на рис. 3. Амплитуда сигнала в центре «облака» могла быть на $10\div40$ дБ выше, чем по краям. Временна́я протяжённость «облаков» ξ составляла, как правило, порядка $0{,}5\div2{,}0$ с.

Выше 96 км отражения от нижнего слоя плавно переходили в отражения от E-области, а ниже, на высотах $80\div86$ км, наблюдались, по-видимому, сигналы от естественных крупномасштабных неоднородностей плазмы, так называемые сигналы частичных отражений. Кроме того, почти всегда наблюдались сигналы от слоя $E_{\rm s}$ с высот $105\div108$ км, т. е. максимум слоя $E_{\rm s}$ находился на высоте максимума слоя E. Эпизодически появлялся и верхний спорадический слой на высоте $h\sim116$ км, часто также с выраженной «облачной» структурой (рис. 3). С течением времени отражения от нижней границы E-области переместились на более низкие высоты, что соответствует наблюдениям ночных профилей электронной концентрации в E-области. Ночью регулярная E- и, тем более, D-области обычно не детектируются радарами обратного рассеяния, поскольку фоновая ионизация падает на два порядка за исключением спорадической ионизации, плотность которой может достигать и превышать максимум ионизации в F-области. Из рис. 2 видно, что за ночные часы слой E снизился приблизительно на 5 км. С восходом Солнца в E-области высота слоя начала постепенно возрастать.

2.2. Результаты наблюдений 15 и 16 июня 2001 года

В 2001 году измерения характеристик сигналов ИПН проводились в течение ночи с 15 на 16 июня. Яркостные записи амплитуды, времени релаксации сигналов и скорости вертикального движения плазмы для этого периода наблюдений представлены на рис. 4*а*–*в*. Ниже, на рис. 4*г*, показан временно́й ход частоты экранирования f_b и предельной частоты отражения f_0 для слоя E_s, а также ход критических частот областей E и $F2$.

Эти наблюдения подтвердили замеченную в предыдущем эксперименте заходно-восходную асимметрию высотно-временно́й зависимости рассеянного сигнала. Слабые сигналы от ИПН в D-области наблюдались на этих высотах почти до захода Солнца, а появились спустя час после восхода. Их амплитуда была в целом выше в восходные часы. Высотный интервал, который занимали сигналы от ИПН в D-области, и интервал высот между D- и E-областями, в котором сигналы отсутствовали, соответствует наблюдениям 2000 года.

По данным ионозонда с 19:45 LT 15 июня до 03:15 LT 16 июня наблюдался полупрозрачный слой E_s с частотой экранирования $f_\mathrm{b} = 3{,}0\div5{,}5$ МГц и предельной частотой отражения $f_0 = 5\div\div10$ МГц, в результате чего регистрировалось до 5 кратных отражений как на ионограммах, так и по регистрации ИПН (рис. 1*в*).

Интересной особенностью эксперимента 2001 года явилось наблюдение нескольких типов спорадических образований на высотах E-области. Первый тип можно условно определить как «классический» долгоживущий слой E_s, сформировавшийся на высотах $95\div105$ км (это высота максимальной амплитуды сигнала от слоя), с «толщиной» (по уменьшению амплитуды рассеянного сигнала в 2 раза) до 5 км. Амплитуда рассеянного ИПН сигнала от такого слоя составляла до 70 дБ, а время релаксации достигало 10 с.

Второй тип спорадического слоя можно назвать подслойным (underlying). Этот, как правило, тонкий слой, обычно наблюдался на высоте $85\div90$ км и часто имел ярко выраженную «облачную» структуру (масштабы «облаков» приведены ниже). Амплитуда рассеянного ИПН сигнала от такого слоя достигала 75 дБ, а время его релаксации — $7\div10$ с.

И, наконец, третий тип спорадического слоя — промежуточный (sequential), или спускающийся (descending). Подобные слои регулярно наблюдаются различными методами (см., например, [2, 5, 6]). В ходе эксперимента наблюдался достаточно толстый слой с «толщиной» до 10 км, амплитудой рассеянного сигнала более 75 дБ и временем релаксации более 10 с. Он появился на ионограммах 15 июня 2001 года в 19:45 LT и в это же время был зарегистрирован методом ИПН. Слой появился в области высот $113\div123$ км и постепенно снижался до высот $100\div113$ км с эффективной скоростью порядка 1 м/с. Волнообразная модуляция спускающегося слоя хорошо видна на рис. 4*а* (в интервале 21:00–22:00 пропущено 8 минут по техническим причинам, поэтому после 21:00 шкала времени несколько искажена).

На графике зависимости критических частот слоя E_s от времени на рис. 4*г* видно, что после 21:00 LT частоты f_b и f_0 начинают возрастать и за 45 минут увеличиваются почти в 2 раза: f_0 растёт с 4,8 до 10 МГц, а f_b с 3,4 до 5,3 МГц, достигая максимума в 21:45 LT, после чего частоты слоя E_s начинают быстро уменьшаться, и в 22:00 LT наблюдается их глубокий минимум. На рис. 4*а* этот интервал времени соответствует резкому уменьшению амплитуды рассеянного сигнала и «расслоению» верхнего спускающегося слоя E_s на два более тонких слоя. Аналогичная ситуация имела место и с 20:20 до 20:40 LT, когда амплитуда отражений слоя упала с 70 до $40\div50$ дБ, а его «толщина» уменьшилась до $2\div3$ км. Интересно, что в это же время наблюдался небольшой рост f_b и небольшой локальный максимум f_0E_s. Что касается слоя $F2$, то амплитуда сигнала, рассеянного ИПН, была стабильно высокой (всегда более 75 дБ, т. е. намного превышала динамический диапазон приёмника), критическая частота слоя слабо менялась в окрестности

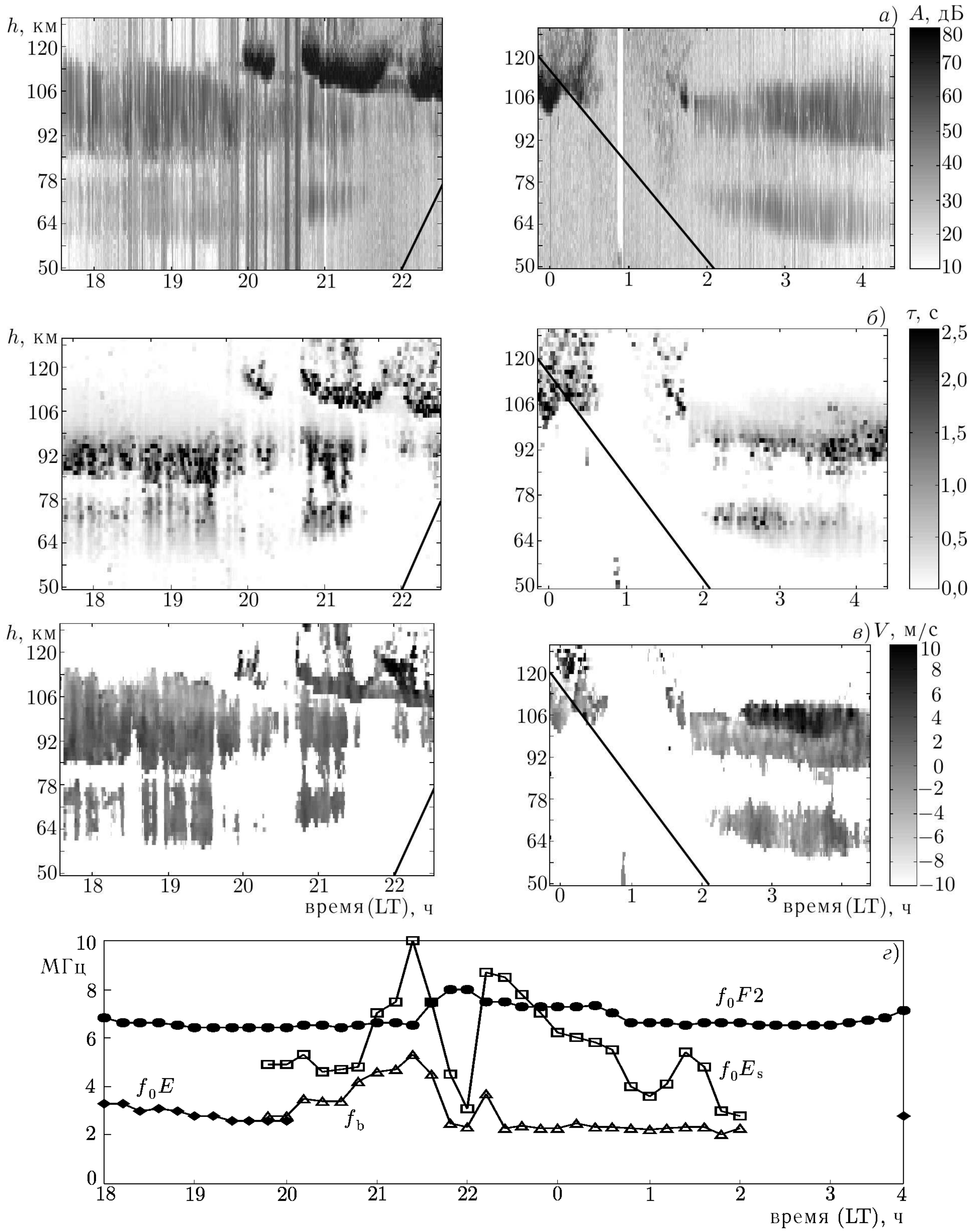

Рис. 4. Высотно-временны́е зависимости рассеянного сигнала, полученные 15 и 16 июня 2001 года: амплитуда (*а*), время релаксации (*б*), скорость вертикального движения (*в*), а также временны́е зависимости критических частот ионосферы и спорадического слоя E

7 МГц, наблюдалась лишь её небольшая вариация (до 1 МГц) с 21:45 до 01:00 LT. Любопытно, что в течение почти часового интервала времени с 00:45 до 01:30 LT на записях зарегистрирован только сигнал от ИПН в области $F2$ и «следы» от его многочисленных кратных отражений (широкая вертикальная полоса на рис. 4*а*). Несмотря на то, что на ионограммах в это время виден слой $E_{\rm s}$ с предельной частотой отражения $f_0 \sim 3{\div}5$ МГц и частотой экранирования $f_{\rm b} \sim 2{,}5$ МГц, методом ИПН он регистрировался лишь эпизодически. Это, скорее всего, связано с тем, что диаграмма направленности ионозонда существенно шире, чем диаграмма направленности стенда «Сура», излучение которого формирует ИПН, поэтому ионозонд принимает рассеянный сигнал со значительно большей площади. Поскольку в эксперименте наблюдался полупрозрачный слой $E_{\rm s}$ (рис. 1*в*), а такие слои состоят из отдельных крупных «облаков» ионизации (подобный пример приведён на рис. 1 в [5]), то возможны интервалы времени, когда эти «облака» не попадали в область формирования ИПН.

Приблизительно в 01:30 LT появились рассеянные сигналы от слоя $E_{\rm s}$ вблизи высоты $113\div$ $\div 116$ км, и практически одновременно появились сигналы от ИПН в E-области. Они регистрировались до конца наблюдений, причём одновременно с ними на яркостных записях видны отражения от слоя $E_{\rm s}$ с высоты чуть ниже высоты максимума E-области.

На восходе Солнца реже наблюдался подслойный тип $E_{\rm s}$, а слой «классического» типа регистрировался практически до конца наблюдений. Возможно, наблюдавшийся подслойный тип $E_{\rm s}$ является результатом расслоения слоя «классического» типа. Наиболее долгоживущие ионы вносят вклад в перераспределение остаточной ионизации в E-области после захода Солнца. Вероятно также, что в тех случаях, когда подслойный тип $E_{\rm s}$ наблюдался на восходе, оба слоя, «классический» и подслойный, существовали всю ночь, но были настолько слабы, что какое-то время не детектировались.

3. ФАЗА РАССЕЯННОГО СИГНАЛА И СКОРОСТЬ ВЕРТИКАЛЬНОГО ДВИЖЕНИЯ ПЛАЗМЫ. ДИНАМИКА СПОРАДИЧЕСКИХ ОБРАЗОВАНИЙ

В монографии [4] подробно обсуждается тот факт, что в процессе релаксации ИПН увлекаются движением нейтрального газа, поэтому регистрация доплеровского смещения частоты рассеянного сигнала может быть использована для определения скорости V вертикального движения плазмы. Мы, однако, не можем непосредственно измерять доплеровскую частоту сигнала, рассеянного ИПН, т. к. характерное время жизни неоднородностей меньше периода доплеровских колебаний. Преодолеть эту трудность позволяет измерение фазы рассеянного сигнала, по которой затем рассчитывается V. При этом погрешность определения V составляет в среднем $0{,}05 \div 0{,}08$ м/с [4], и лишь в условиях сильного нагрева, когда частота мощной волны близка к критической частоте слоя, погрешность может быть значительной. В условиях эксперимента фаза рассеянного ИПН сигнала в некоторые интервалы времени очень быстро менялась, что вызывало глубокие вариации скорости V. Такие результаты из анализа скорости исключались. По наблюдениям в августе 2000 года основная масса измерений относится к интервалу значений скорости от -10 до 8 м/с со средним значением $\langle V\rangle \approx -1$ м/с. Из рис. 2*в* видно, что распределение вертикальной скорости по высоте значительно отличалось в вечерние часы и на восходе Солнца. Так, например, в вечерние и ночные часы как в D-, так и в E-области скорость была преимущественно направлена вверх со сдвигом $\mathrm{d}V/\mathrm{d}h \sim 10^{-5} \div 10^{-3}$ с$^{-1}$. После восхода Солнца с периодичностью $15 \div 20$ минут происходила смена преимущественного направления скорости в E-области, что делало возможным сгонку металлических ионов в спорадический слой, а на высотах области D преимущественно наблюдались движения вверх.

По измерениям 2001 года большинство значений скорости лежат в интервале от -10 до 10 м/с.

Средняя за весь период наблюдений вертикальная скорость на высотах E-области составила на разных высотах от 0,5 до $-3{,}8$ м/с.

На рис. 4*в* хорошо видно, что в D-области в предзаходные часы скорости направлены из области ионизации как вверх, так и вниз. Вблизи 90 км чаще наблюдались движения вниз с отрицательным градиентом по высоте, который должен вызывать сгонку ионизации в «подслойный» слой $E_{\rm s}$. В то же время на высотах спускающегося спорадического слоя скорости были направлены преимущественно внутрь области ионизации.

Изменение средней величины и направления скорости в E-области можно проследить по гистограммам распределения V для высот $85 \div 124{,}2$ км, которые для вечерних часов 15 июня 2001 года показаны на рис. 5. Для каждой гистограммы приведены высота и среднее значение вертикальной скорости $\langle V \rangle$ на этой высоте. Хорошо видно, что средние величина и направление скорости зависели от высоты. На высотах примерно 92 и 114 км имела место смена знака средней скорости, т. е. могло происходить образование спорадического слоя E за счёт механизма ветрового сдвига путём сгонки положительных ионов. Средние за весь период скорости невелики по модулю, но в отдельных конкретных сеансах они были значительными и имели нужное для сгонки направление. Здесь мы приводим лишь средние значения, т. к. в этой статье подробно не анализируются вариации вертикальной скорости, но даже по средним значениям видно, что на высотах 92 и 114 км могло происходить образование слоя $E_{\rm s}$ за счёт сгонки металлических ионов. Отметим, что спорадические слои мы наблюдали именно на этих высотах.

Определённые по высотным профилям $V(h)$ сдвиги вертикальной скорости на высотах $100 \div 113$ км составили по модулю $|\mathrm{d}V/\mathrm{d}h| \approx 5 \cdot 10^{-5} \div 10^{-4}$ с$^{-1}$. Такие сдвиги являются достаточными для формирования слоя $E_{\rm s}$ путём сгонки долгоживущих металлических ионов. Во многих случаях максимум амплитуды сигнала от слоя $E_{\rm s}$ наблюдался на высоте, где $V \sim 0$, а также на высотах, где имелся достаточный сдвиг скорости для сгонки (по оценкам, приведённым в [1], сгонка ионов может происходить уже при сдвигах порядка $5 \cdot 10^{-4}$ с$^{-1}$).

Особенности заходно-восходной перестройки динамического режима ионосферы нашли отражение в изменении вертикальной скорости с высотой, которые иллюстрирует рис. 6, где показано несколько взятых подряд профилей $V(h)$ за интервалы, в которые включены времена восхода и захода Солнца в E-области. Видно, что высотные вариации $V(h)$ имеют явно волнообразный характер с периодом по высоте порядка 15 км.

4. ВЛИЯНИЕ СЛОЯ $E_{\rm s}$ НА АМПЛИТУДУ И ВРЕМЯ РЕЛАКСАЦИИ РАССЕЯННОГО СИГНАЛА. МЕТАЛЛИЧЕСКИЕ ИОНЫ

Влияние слоя $E_{\rm s}$ на амплитуду A рассеянного сигнала заключается в том, что на высотной зависимости $A(h)$ наблюдается значительное увеличение амплитуды (иногда на $30 \div 40$ дБ) вследствие роста электронной концентрации в слое и коэффициента отражения. Понятно, что время релаксации также должно возрастать.

Без учёта влияния атмосферной турбулентности, что справедливо на высотах выше турбопаузы, можно полагать, что релаксация неоднородностей в E-области обусловлена амбиполярной диффузией [4]. В этом случае в отсутствие слоя $E_{\rm s}$ высотная зависимость времени релаксации ИПН $\tau(h)$ является экспоненциальной. Влияние турбулентности выражается в уменьшении времени релаксации относительно времени диффузионного расплывания ИПН [4]. Если же на какой-либо высоте существует слой $E_{\rm s}$, то на графике зависимости $\tau(h)$ на этой высоте наблюдаются локальные максимумы, в которых τ возрастает иногда в несколько раз. Это объясняется тем, что увеличение электронной концентрации в слое $E_{\rm s}$ по сравнению с фоновой концентрацией в E-области приводит к уменьшению показателя преломления, увеличению длины волны в слое и

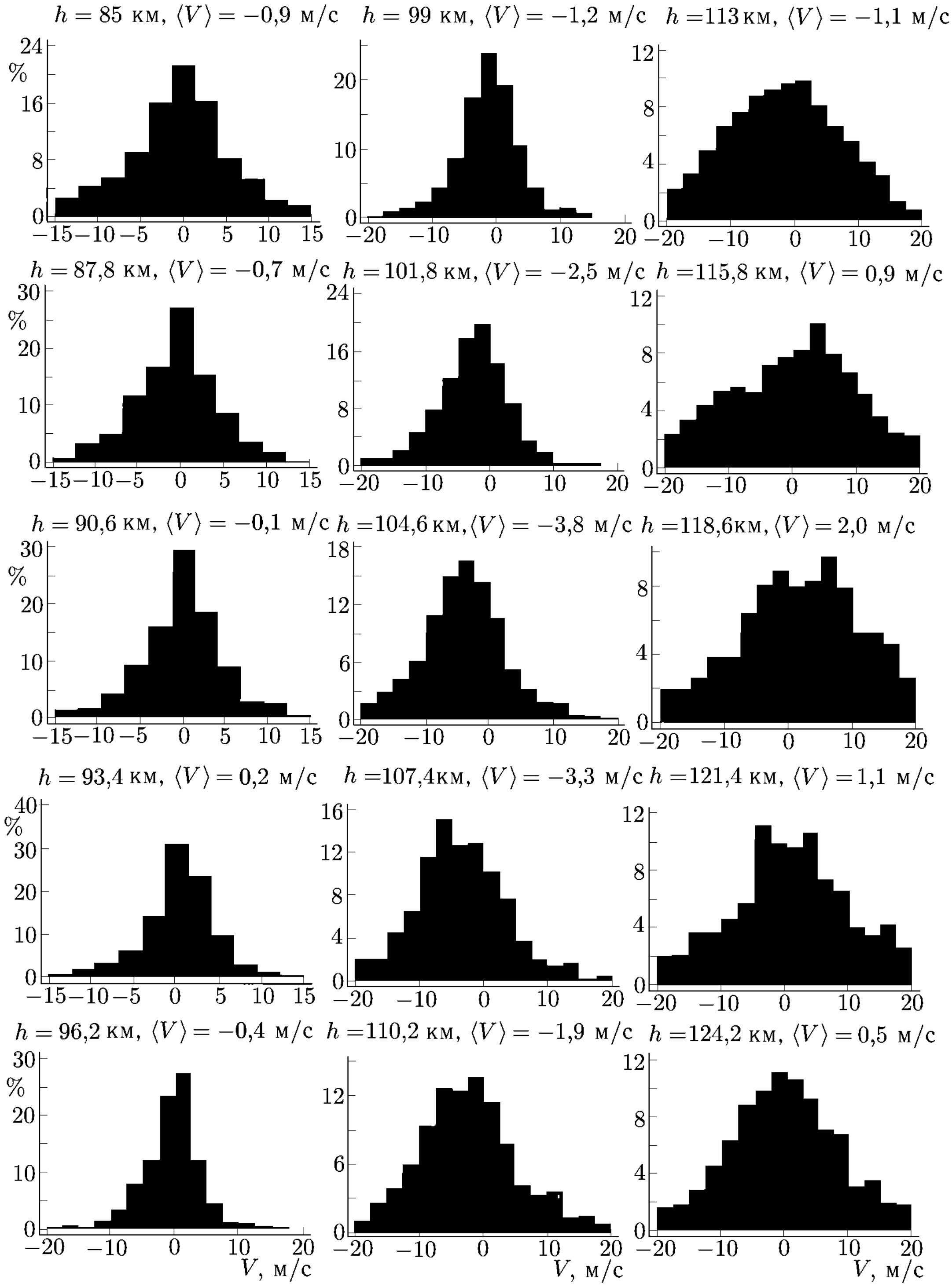

Рис. 5. Гистограммы распределения скорости вертикального движения на разных высотах в вечерние и ночные часы 15 июня 2001 года. На каждой гистограмме указаны высота и среднее значение скорости на данной высоте

увеличению времени диффузионного расплывания неоднородностей. При этом относительное изменение τ выражается соотношением

$$\delta\tau = \frac{\Delta\tau}{\tau} = \frac{1}{2}\,\frac{\Delta n}{n}\,, \qquad (1)$$

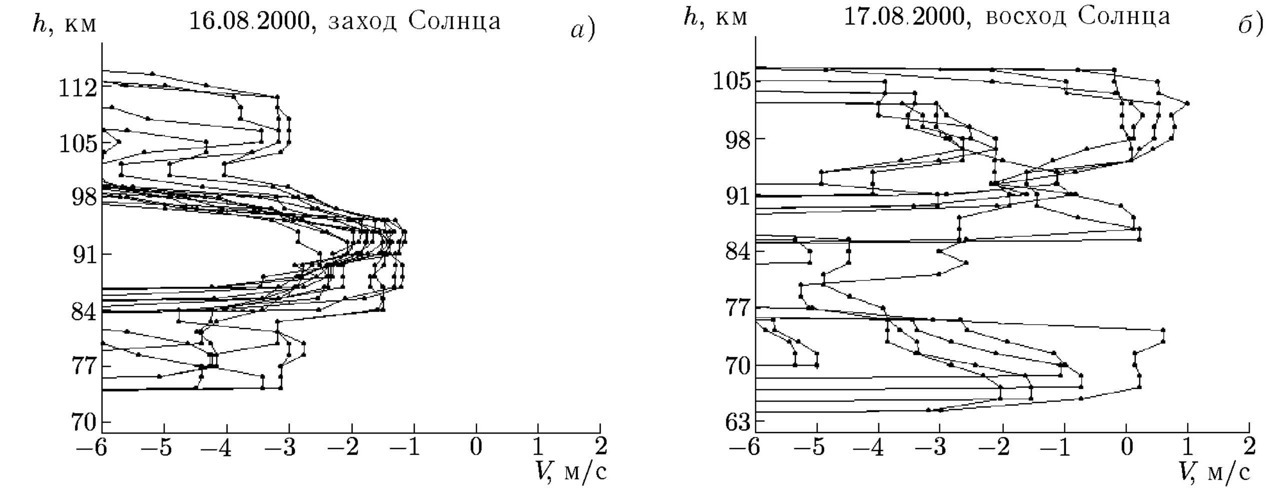

Рис. 6. Высотные профили скорости вертикального движения при заходе Солнца 16.08.2000 (*а*) и восходе Солнца 17.08.2000 (*б*)

а изменение плазменной частоты слоя $E_{\rm s}$ относительно фоновой в E-области записывается выражением [4]

$$\delta f_0 = \frac{f_0 E_{\rm s}}{f_0 E} = \frac{1}{2}\,\delta\tau \left[\frac{f\,(f - f_{\rm L})}{(f_0 E)^2} - 1\right], \tag{2}$$

где $f_0 E_{\rm s}$ и $f_0 E$ — плазменные частоты слоя $E_{\rm s}$ и E-области соответственно, $f_{\rm L} = f_H \cos\theta$, f_H — гирочастота электронов, θ — угол между направлением геомагнитного поля и волновым вектором радиоволны, $\Delta\tau$ — увеличение времени релаксации на высоте слоя $E_{\rm s}$ по сравнению с диффузионным, τ — диффузионное время релаксации в E-области в отсутствие спорадического слоя, n — показатель преломления на данной высоте, Δn — его изменение из-за влияния слоя $E_{\rm s}$. Подставив в (2) значения $f = 5{,}67$ МГц, $f_0 E = 3{,}5\div 4{,}0$ МГц, $f_{\rm L} = 1{,}3$ МГц и $\delta\tau = 0{,}1\div 0{,}5$, получим $\delta f_0 = 0{,}03\div 0{,}25$. Таким образом, даже не слишком интенсивные слои заметно увеличивают время релаксации ИПН.

Другим фактором, вызывающим увеличение времени релаксации ИПН, является наличие на этих высотах долгоживущих металлических ионов, сгонка которых в узкие слои согласно теории ветрового сдвига и приводит на средних широтах к образованию слоя $E_{\rm s}$ [2]. Между тем диффузионное время релаксации ИПН τ пропорционально относительной молекулярной массе преобладающих ионов M и выражается соотношением

$$\tau = \frac{1}{K^2 D_{\rm a}} = \frac{M \nu_{\rm in}}{k\,(T_{\rm e0} + T_{\rm i0})\,K^2}\,, \tag{3}$$

где k — постоянная Больцмана, $K = 4\pi/\lambda$ — волновое число стоячей волны, $\lambda = \lambda_0/n$ — длина волны в среде, $D_{\rm a}$ — коэффициент амбиполярной диффузии, M — относительная молекулярная масса ионов, $T_{\rm e0}$ и $T_{\rm i0}$ — невозмущённые электронная и ионная температуры, $\nu_{\rm in}$ — частота столкновений ионов с нейтральными молекулами.

По результатам спектрометрических измерений (см., например, [7, 8]) в спорадических слоях были обнаружены ионы металлов $\rm Fe^+$ ($M = 56$), $\rm Mg^+$ ($M = 24$), $\rm Ca^+$ ($M = 40$), $\rm Al^+$ ($M = 27$), $\rm Na^+$ ($M = 22$), $\rm Si^+$ ($M = 28$). Из (3) с очевидностью следует, что наибольшее влияние на τ должны оказывать ионы железа с атомной массой $M = 56$, которая почти в 2 раза больше средней молекулярной массы преобладающих на высотах E-области атмосферных ионов $\rm NO^+$ ($M = 30$), $\rm O_2^+$ ($M = 32$), или кальция ($M = 40$). Измерения показывают, что тяжёлые ионы могут составлять до 80 % от общего количества металлических ионов в слое $E_{\rm s}$ [7–12], и в этом случае можно ожидать заметного увеличения τ на высотах спорадического слоя. Если критические частоты слоя $E_{\rm s}$ и E-области известны, то используя (1) и (3) можно одновременно оценить влияние как показателя преломления n, так и относительной молекулярной массы металлических ионов M на время релаксации рассеянного сигнала:

$$\frac{\tau_{E_{\rm s}}}{\tau} = \left(\frac{n_E}{n_{E_{\rm s}}}\right)^2 \left(\frac{M_{\rm m}}{M_{\rm a}}\right), \tag{4}$$

где $M_{\rm m}$ и $M_{\rm a}$ — относительная молекулярная масса металлических и атмосферных ионов.

Таким образом, измеряя в эксперименте $\Delta\tau/\tau$ или $\tau_{E_{\rm s}}/\tau$, можно оценить относительную молекулярную массу металлических ионов. [1]

При измерениях с 19:45 до 20:45 LT 15.06.2001 было получено $\Delta\tau/\tau \approx 1{,}0 \div 1{,}8$. Взяв $f_0E = 2{,}5$ МГц, $f_{\rm b} = 3 \div 3{,}5$ МГц, $f = 5{,}67$ МГц, получим для относительной молекулярной массы ионов значения 39 и 57, которые близки к массам ионов $\rm Ca^+$ и $\rm Fe^+$, что выглядит достаточно правдоподобно. В то же время для «классического» слоя $E_{\rm s}$, наблюдавшегося в предзаходные часы на высотах 95÷105 км, $f_0E = 2{,}0$ МГц, $f_{\rm b} = 2{,}2$ МГц и $\Delta\tau/\tau \approx 0{,}3$, откуда для массы ионов получим значение $M_{\rm m} = 37$.

По ионосферным данным во время наблюдений 16, 17 августа 2000 года критическая частота E-слоя и частота экранирования слоя $E_{\rm s}$ практически совпадали. В этом случае увеличение времени релаксации ИПН должно было определяться лишь массой металлических ионов. Характерно, что в это время изменения τ в подавляющем большинстве сеансов составили 30÷50 %, что даёт для $M_{\rm m}$ значения 37÷42.

Можно оценить и общую концентрацию всех металлических ионов на высоте максимума слоя $E_{\rm s}$, используя для этой цели метод расчёта, предложенный в [13, 14]. Метод основан на измерении скорости вертикального движения плазмы с помощью искусственных периодических неоднородностей и использовании данных вертикального зондирования. Исходным являлось уравнение непрерывности для электронов в стационарном состоянии в приближении плоскослоистой ионосферы. В качестве эффективного коэффициента рекомбинации α принималось средневзвешенное значение коэффициентов рекомбинации различных сортов положительных ионов, имеющихся на высотах области E. Считалось также, что слой $E_{\rm s}$ не влияет на функцию образования электронов. Таким образом, было получено, что концентрации атмосферных ($\rm NO^+$ и $\rm O_2^+$) и металлических ионов зависят от высотного сдвига вертикальной скорости $\mathrm{d}V/\mathrm{d}h$ и $N_{\rm s}$ — электронной концентрации в максимуме слоя $E_{\rm s}$ — и определяются выражениями [15]

$$\alpha = \alpha_0 \left(\frac{N_0}{N_{\rm s}}\right)^2 - \frac{1}{N}\,\frac{\mathrm{d}V}{\mathrm{d}h}\,, \quad N_{\rm a} = N_{\rm s}\left(\frac{\alpha}{\alpha_0}\right) = \frac{N_0^2}{N_{\rm s}} - \frac{1}{\alpha_0}\,\frac{\mathrm{d}V}{\mathrm{d}h}\,, \quad N_{\rm m} = N_{\rm s}\left[1 - \left(\frac{N_0}{N_{\rm s}}\right)^2\right] + \frac{1}{\alpha_0}\,\frac{\mathrm{d}V}{\mathrm{d}h}\,, \tag{5}$$

где α — эффективный коэффициент рекомбинации, $\alpha_0 = 4{,}4\cdot 10^{-7}\,(300/T_{\rm e}[\mathrm{K}])$ см^3 с$^{-1}$, $T_{\rm e}$ — электронная температура на высоте максимума слоя $E_{\rm s}$, также определяемая методом ИПН [4], $N_{\rm a}$ и $N_{\rm m}$ — концентрации атмосферных и металлических ионов, N_0 и $N_{\rm s}$ — электронная концентрация в E-области и в слое $E_{\rm s}$ на высоте его максимума. Таким образом, используя измеренные методом

[1]Строго говоря, метод позволяет оценить эффективную массу ионов, содержащихся в ионосфере на высоте максимума слоя $E_{\rm s}$, которая выражается формулой $M_{\rm eff} = N^{-1}(\Sigma_{i=1}^{n} M_{{\rm m}i} N_{{\rm m}i} + \Sigma_{j=1}^{k} M_{{\rm a}j} N_{{\rm a}j})$, где $N_{{\rm m}i}$ — концентрация разных сортов металлических ионов, $N_{{\rm a}j}$ — атмосферных. Концентрация всех ионов N в приближении квазинейтральности плазмы равна электронной концентрации $N_{\rm e}$. Известно, что в слое $E_{\rm s}$ наблюдается резкое уменьшение содержания атмосферных ионов по сравнению с фоновыми значениями, в результате величина $M_{\rm eff}$ будет отличаться от массы преобладающих ионов не более чем на 3÷5 %.

ИПН значения скорости вертикального движения плазмы V и частотные характеристики слоя E_{s} и области E по данным вертикального зондирования, из выражений (5) можно определить эффективный коэффициент рекомбинации и концентрации атмосферных и металлических ионов.

Оценки, сделанные для двух рассмотренных выше случаев, дали значения $N_{\mathrm{m}} \sim 10^4$ см$^{-3}$ и $N_{\mathrm{a}} \sim 8{\cdot}10^4$ см$^{-3}$ для слоя, спускавшегося с высоты 113 км (в начале спуска), и $N_{\mathrm{m}} \sim 7{,}3{\cdot}10^3$ см$^{-3}$ и $N_{\mathrm{a}} \sim 5{,}3\cdot10^4$ см$^{-3}$ для слоя, максимум которого находится на высоте около 100 км. К сожалению, не удалось сделать таких оценок на протяжении всего времени наблюдения снижавшегося слоя E_{s}, т. к. в течение ночи ионозонд не регистрировал низкую электронную концентрацию в E-области (рис. 4*г*), в результате чего данные по f_0E отсутствуют.

Сопоставим полученные результаты с приведёнными в литературе данными измерений высотного распределения металлических ионов на высотах спорадического слоя E.

За последние тридцать лет спектрометрические измерения на ракетах, а позднее лидарные измерения на разных широтах дали достаточно полную картину распределения металлических ионов и основных атмосферных ионов NO^+ и O_2^+ в диапазоне высот $80\div120$ км [7–12], включая периоды наблюдений метеорных потоков. Из приведённых в [11] данных для $30^\circ\div55^\circ$ с. ш. можно сделать вывод о том, что в высотном распределении металлических ионов существуют три основных локальных максимума, определяющих высоты слоя E_{s}. Один максимум в интервале $90\div95$ км с общей концентрацией всех сортов металлических ионов до $N_{\mathrm{m}} \sim 2\cdot10^4$ дают слои, состоящие в основном из ионов Fe^+ (до $6\cdot10^3$ см$^{-3}$), Mg^+ (до $4\cdot10^3$ см$^{-3}$), Na^+ (до $7\cdot10^2$ см$^{-3}$), Ca^+ (до $2\cdot10^2$ см$^{-3}$). Другой максимум вблизи высоты 105 км с общей концентрацией всех металлических ионов до $N_{\mathrm{m}} \sim 10^5$ см$^{-3}$ дают слои тех же металлов, иногда регистрируются также ионы кремния Si^+ ($M = 28$), концентрация которых может быть сопоставима с концентрацией ионов железа и даже превышать её. Наконец, третий максимум концентрации ионов металлов с $N_{\mathrm{m}} \sim 2\cdot10^4$ см$^{-3}$ во многих случаях наблюдается на высотах $113\div115$ км. В ряде измерений зарегистрированы ионные слои с высоким содержанием Na^+ (до $4\cdot10^3$ см$^{-3}$). Высоты максимальных концентраций разных сортов ионов отличаются, как правило, на $1\div2$ км.

Сопоставляя наши оценки с приведёнными выше результатами прямых измерений, можно заключить, что общая концентрация металлических ионов, найденная согласно (5), хорошо согласуется с результатами лидарных и ракетных измерений.

Располагая данными непосредственных измерений ионного состава, можно оценить концентрацию определённого сорта ионов. Так, по данным, приведённым в [12, 15], на высоте $113\div115$ км доля ионов Fe^+ составляла до 55 % от общей концентрации ионов металлов, в то время как доля кальция была существенно меньше — от 0,5 до 5 %. На высотах $95\div100$ км концентрация Ca^+ по отношению к концентрации Fe^+ составляет обычно не более $5\div10$ %. Если принять долю ионов Fe^+ порядка 50 %, то при общей концентрации всех сортов ионов $N_{\mathrm{m}} \sim 10^4$ см$^{-3}$ концентрация ионов железа составит $N_{\mathrm{m}}(\mathrm{Fe}^+) \sim 5\cdot10^3$ см$^{-3}$, что не противоречит результатам прямых измерений.

Проведённое сопоставление показывает перспективность предложенного метода исследования ионного состава на высотах спорадического слоя E.

Как отмечалось выше, наиболее точно метод, использующий ИПН, позволяет определять тип ионов металлов с массой, превышающей массу основных атмосферных ионов. В этой связи возникает необходимость усовершенствования метода с возможностью идентификации более лёгких, но не менее существенных ионных фракций в слое E_{s}, например Na^+ и Mg^+, концентрации которых в спорадических слоях могут быть значительно выше концентраций Fe^+ [11].

Таким образом, тип относительно тяжёлых метеорных ионов, участвующих в образовании слоя E_{s}, и общую концентрацию всех сортов металлических ионов можно определять по высотной зависимости времени релаксации ИПН, сдвигу скорости вертикального движения ионосферной плазмы и данным вертикального зондирования.

5. ВНУТРЕННИЕ ГРАВИТАЦИОННЫЕ ВОЛНЫ И ОБРАЗОВАНИЕ СЛОЯ E_s. ПАРАМЕТРЫ НЕОДНОРОДНОСТЕЙ В СЛОЕ E_s

В рамках теории ветрового сдвига узкие интенсивные спорадические слои ионизации образуются в среднеширотной ионосфере под действием вертикального сдвига горизонтальной скорости плазмы (или горизонтального ветра, т. к. в нижней ионосфере плазма является примесью и перемещается со скоростью нейтрального ветра), что вызывает вертикальный перенос плазмы. Необходимые для сгонки плазмы ветровые сдвиги могут создаваться, в частности, прохождением интенсивных внутренних гравитационных волн [2, 5, 6]. Не вызывает сомнения факт возбуждения внутренних гравитационных волн в периоды прохождения через пункт наблюдения солнечного терминатора, т. е. в периоды захода и восхода Солнца [14, 16, 17].

В литературе (см., например, обзоры [2, 6]) приводится достаточно много данных наблюдений слоя E_s и их связи с внутренними гравитационными волнами, приливами, перемещающимися ионосферными возмущениями. Теоретически рассмотрен вопрос о возникновении необходимых для сгонки слоя E_s сдвигов горизонтального ветра, инициированных внутренними гравитационными волнами. Согласно нашим измерениям методом ИПН скорости вертикального движения плазмы и профилей электронной концентрации в нижней ионосфере в 1990–1991 гг., результаты которых подробно представлены в [4], внутренние гравитационные волны во многом определяют характер изменений вертикальной скорости с высотой.

Спектральный анализ амплитуды и времени релаксации рассеянного сигнала, а также вертикальной скорости на высотах E-области по результатам измерений в июне 2001 года выявил волнообразные вариации этих параметров с наиболее характерными периодами 15, 30, 120 и 240 минут. Период, соответствующий частоте Бранта—Вяйсяля, на которой происходит резкое уменьшение интенсивности спектра, составил от 6,5 до 8,5 минут. По одновременным измерениям характеристик рассеянного сигнала в большом диапазоне высот можно было проследить за вариациями фазовой скорости волны, которая была направлена преимущественно вниз. На высотных профилях вертикальной скорости выделен период порядка 15 км по высоте. Можно предположить, что необходимые для образования снижающегося спорадического слоя E ветровые сдвиги вертикальной скорости по наблюдениям 15, 16 июня 2001 года могли быть обусловлены прохождением интенсивных внутренних гравитационных волн. Дополнительным аргументом в пользу такой интерпретации служит наличие глубоких вариаций критических частот слоёв (рис. 4*г*) в период наблюдения снижавшегося слоя E_s.

Интересную особенность процесса формирования верхнего слоя E_s в предзаходные часы 15 июня 2001 года иллюстрирует рис. 7. Чуть ниже высоты 115 км, на которой потом появляется слой E_s, виден слой, амплитуда рассеянного сигнала от которого близка к амплитуде сигнала от ИПН E-области, однако его время релаксации существенно меньше. На последовательных кадрах видно, как развивается слой E_s, интенсивность которого с течением времени значительно увеличивается, и он сливается с этим маленьким слоем. Приблизительно через 30 минут слой E_s ослабевает, а потом процесс начинается вновь. При этом маленький слой никуда не исчезает. Происхождение таких долгоживущих слоёв не удаётся однозначно интерпретировать. Можно лишь предположить, что эти слои появляются в результате расслоения регулярного E-слоя [4].

Как отмечалось выше, спорадические образования, особенно «подслойные», имели ярко выраженную «облачную» структуру, и амплитуда сигнала в центре «облака» могла быть на 10÷40 дБ выше, чем по краям. Согласно подробно рассмотренной в [4] теории образования и релаксации ИПН в области E, амплитуда рассеянного сигнала пропорциональна концентрации в неоднородностях, т. е. указанное изменение A соответствует относительной концентрации до $\delta N \sim 10^2$.

Размеры «облаков» грубо можно оценить как $l = V_0\xi$, где V_0 — скорость горизонтального движения «облаков», а ξ — их «протяжённость во времени». Полагая $V_0 \sim 100$ м/с, получим масштабы $l \sim 50 \div 200$ м и $l \sim 150 \div 300$ м по результатам измерений в 2000 и 2001 году соответственно. Таким образом, можно полагать, что наблюдавшиеся облака ионизации представляли собой до-

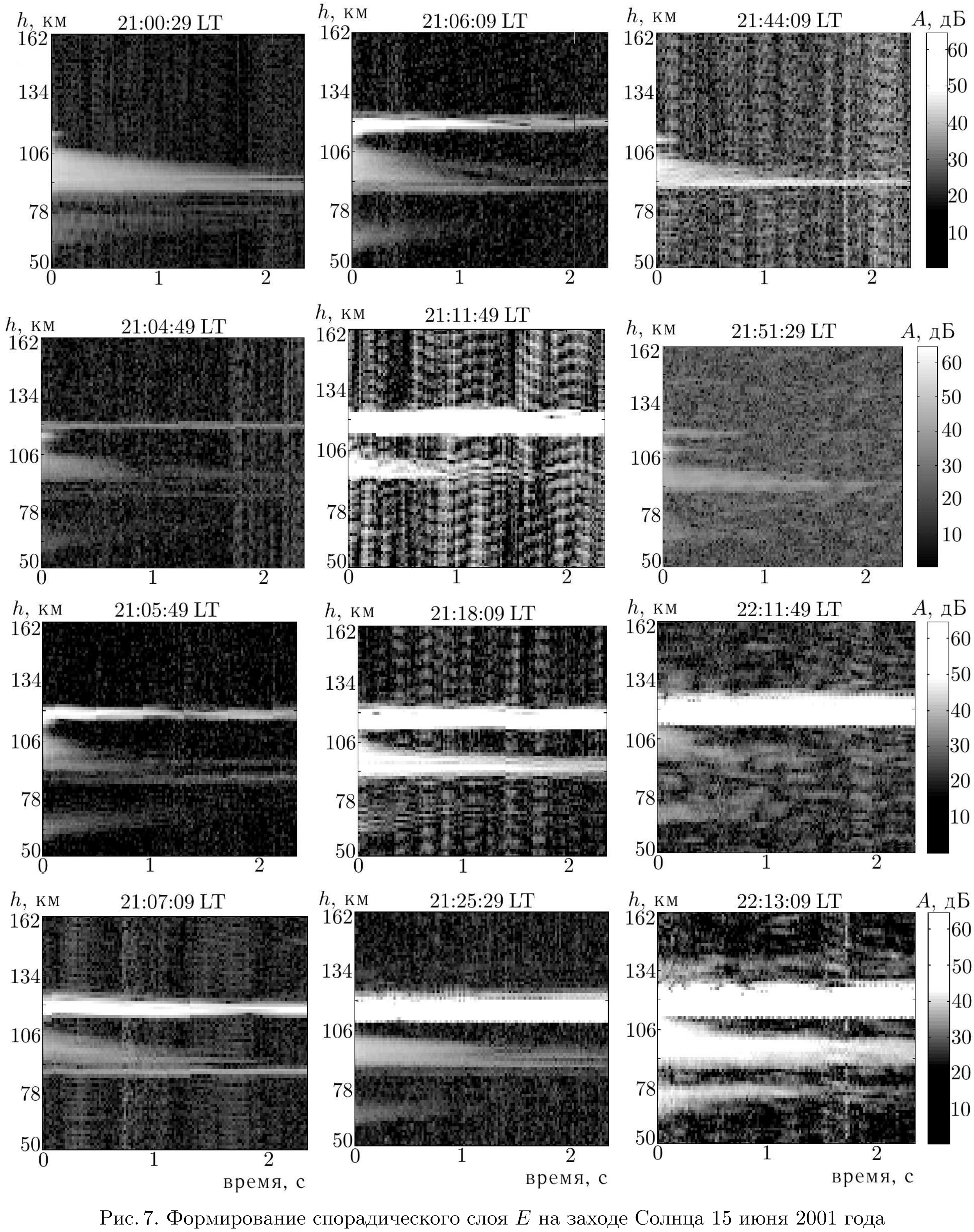

Рис. 7. Формирование спорадического слоя E на заходе Солнца 15 июня 2001 года

статочно интенсивные среднемасштабные неоднородности. Подобные результаты были получены в [1] по результатам дневных измерений в августе 1999 года.

ВЫВОДЫ

В работе приведены результаты исследований спорадических слоёв ионизации в нижней ионосфере методом обратного рассеяния радиоволн на искусственных периодических неоднородностях ионосферной плазмы по наблюдениям в заходно-восходные часы в летние месяцы 2000 и 2001 года. Измерения выполнены на нагревном комплексе «Сура» (НИРФИ, г. Нижний Новгород). На основе анализа высотно-временно́й зависимости характеристик рассеянного сигнала получены новые сведения о формировании и динамике среднеширотного спорадического слоя E в вечерние и ночные часы. На высотах E-области наблюдались различные типы спорадических образований. Одновременные измерения амплитуды, времени релаксации и скорости вертикального движения ионосферной плазмы позволили исследовать их высотные профили и временны́е зависимости, оценить параметры волновых движений на этих высотах.

Обсуждается роль внутренних гравитационных волн в формировании снижающегося слоя $E_{\rm s}$. Выяснено, что наиболее стабильно наблюдались «классические» слои $E_{\rm s}$, высота максимума которых была близка к высоте максимума E-области. Исследованы особенности заходно-восходных вариаций характеристик рассеянного сигнала, сделаны оценки относительной молекулярной массы и общей концентрации металлических ионов в слое. Получено, в частности, что наблюдавшиеся вариации времени релаксации ИПН могли вызывать спорадические слои, имевшие в составе ионы кальция и железа с общей концентрацией всех сортов металлических ионов на высотах $100\div115$ км до $7{,}3\cdot10^3\div10^4$ см$^{-3}$.

Работа выполнена при поддержке Российского фонда фундаментальных исследований (гранты № 01–05–65025 и 02–05–65281). Работа Л. М. Каган была частично поддержана Канадским советом по естественным наукам и инженерным исследованиям (Canadian Natural Sciences and Engineering Research Council).

СПИСОК ЛИТЕРАТУРЫ

1. Бахметьева Н. В., Беликович В. В., Бенедиктов Е. А., Толмачёва А. В. // Изв. вузов. Радиофизика. 2001. Т. 44, № 12. С. 1 003.
2. Whitehead J. D. // J. Atmos. Terr. Phys. 1989. V. 51, No. 5. P. 401.
3. Беликович В. В., Бенедиктов Е. А. // Изв. вузов. Радиофизика. 2002. Т. 45, № 6. С. 502.
4. Бенедиктов Е. А., Беликович В. В., Толмачёва А. В., Бахметьева Н. В. Исследование ионосферы с помощью искусственных периодических неоднородностей. Н. Новгород: ИПФ РАН, 1999. 156 с.
5. Kagan L. M., Bakhmet'eva N. V., Belikovich V. V., et al. // Radio Science. 2002. V. 37, No. 6. P. 1 106.
6. Гершман Б. Н., Игнатьев Ю. А., Каменецкая Г. Х. Механизмы образования ионосферного спорадического слоя на различных широтах. М.: Наука, 1976. 108 с.
7. Narcisi R. S., Bailey A. D. // J. Geophys. Res. 1965. V. 70. P. 3 687.
8. Narcisi R. S., Bailey A. D., Wlodyka L. E., Philbrick C. R. // J. Atmos. Terr. Phys. 1972. V. 34. P. 647.
9. Mathews J.D. // J. Atmos. Terr. Phys. 1998. V. 60, No. 4. P. 413.
10. Kumar S., Hanson W. B. // J. Geophys. Res. 1980. V. 85. P. 6 783.

11. Huuskonen A., Nygren T., Jalonen L., et al. // J. Geophys. Res. 1988. V. 93. P. 14 603.
12. Grebovsky J. M., Goldberg R. A., Pesnall W. D. // J. Atmos. Terr. Phys. 1998. V. 60. P. 607.
13. Бахметьева Н. В., Беликович В. В., Игнатьев Ю. А., Понятов А. А. // Геомагнетизм и аэрономия. 1996. Т. 36, № 6. С. 36.
14. Бахметьева Н. В., Беликович В. В., Игнатьев Ю. А., Понятов А. А. // Изв. вузов. Радиофизика. 1999. Т. 42, № 1. С. 26.
15. Gerding M., Alpers M., Hoffner J., Von Zahn U. // Ann. Geophisicae. 2001. V. 19, No. 1. P. 47.
16. Сомсиков В. М. Солнечный терминатор и динамика атмосферы. Алма-Ата: Наука, 1983. 192 с.
17. Hocke K., Schlegel K. // Ann. Geophysicae. 1996. V. 14, No. 9. P. 917.

Поступила в редакцию 16 декабря 2003 г.; принята в печать 18 октября 2004 г.

SUNSET–SUNRISE CHARACTERISTICS OF SPORADIC LAYERS OF IONIZATION IN THE LOWER IONOSPHERE AS OBSERVED BY THE METHOD OF RESONANCE SCATTERING OF RADIO WAVES FROM ARTIFICIAL PERIODIC IRREGULARITIES OF THE IONOSPHERIC PLASMA

N. V. Bakhmet'eva, V. V. Belikovich, L. M. Kagan, and A. A. Ponyatov

We present the results of studying the sporadic layers of ionization in the lower ionosphere as observed in the sunset–sunrise time during the summer months of 2000 and 2001. The measurements were performed at the "Sura" heating facility of the Radiophysical Research Institute, Nizhny Novgorod, Russia. Based on the analysis of the altitude–time dependences of the amplitude, relaxation time, and phase of the scattered signal, we obtain new data on the formation and dynamics of the evening and nighttime mid-latitude sporadic E layer, present the results of measuring the vertical-motion velocity of the plasma, and estimate the relative molecular mass and total concentration of metal ions whose pile-up induced by the wind shear can cause the formation of the observed sporadic layers.

УДК 537.874+534.87

ЭЛЕКТРОАКУСТИЧЕСКОЕ ЗОНДИРОВАНИЕ АТМОСФЕРЫ

В. О. Рапопорт, Н. А. Митяков, В. А. Зиничев, Г. П. Комраков, Н. А. Рыжов, Ю. А. Сазонов

Научно-исследовательский радиофизический институт, г. Нижний Новгород, Россия

В статье приведены результаты измерения электрического поля, возбуждаемого в атмосфере мощным звуком (электроакустическое зондирование). Излучаемая акустическая мощность составляла около 1 кВт, частота сигнала изменялась по линейному закону от 17 до 21 Гц и обратно с периодом 8 мин. Для регистрации электрического поля использовалась Г-образная электрическая антенна. Результаты обработки записей электрического поля показали, что в ряде сеансов зарегистрирован электрический отклик на частоте, близкой к частоте зондирующего акустического сигнала. Амплитуда сигнала на входе приёмного устройства составляла порядка 0,1÷1 мкВ.

ВВЕДЕНИЕ

В работе [1] был предложен новый метод дистанционного исследования пространственного распределения электрического заряда в турбулентной атмосфере — метод электроакустического зондирования. Идея этого метода основана на генерации переменного электрического поля при воздействии мощных звуковых волн на области турбулентной атмосферы, содержащие локальные неоднородности распределения электрического заряда. При прохождении звукового импульса возникают вынужденные осцилляции заряженных частиц в таких неоднородностях (электрических ячейках) на частоте звука. Если в атмосфере есть ячейки с размерами порядка длины звуковой волны, то такие области являются источниками электрического поля на частоте, близкой к частоте звука. Это индуцированное электрическое поле может быть зарегистрировано наземным датчиком. Направление луча диаграммы направленности акустической антенны и запаздывание приходящего на электрический датчик сигнала по отношению к моменту посылки акустического импульса определяют положение зондируемой области.

Впервые оценки электрического поля, которое возбуждается при электроакустическом зондировании атмосферы в предгрозовых и грозовых условиях, были выполнены в работе [1]. Было показано, что поля на частотах, близких к частоте акустического сигнала, достаточно велики и могут быть обнаружены в эксперименте.

Первые попытки проведения таких измерений были предприняты летом 1993 года на полигоне НИРФИ «Зимёнки» с использованием акустического излучателя установки радиоакустического зондирования атмосферы, описание которой приведено в [2]. Акустический излучатель работал на частоте 40÷42 Гц, излучаемая мощность составляла около 1 кВт, коэффициент усиления акустической антенны был порядка 20. Электроакустические измерения проводились одновременно с радиоакустическими, поэтому режимы работы акустического излучателя и аналого-цифрового преобразователя в этих экспериментах определялись требованиями, предъявляемыми к радиоакустическим измерениям.

В предгрозовых условиях (в предполуденное время, тогда как гроза началась ближе к вечеру) наблюдался электроакустический сигнал на высотах до 1 км с максимальным превышением уровня сигнала над шумом около 10 дБ. К сожалению, сигнал не удалось записать по техническим причинам.

1. УСТАНОВКА ДЛЯ ЭЛЕКТРОАКУСТИЧЕСКОГО ЗОНДИРОВАНИЯ АТМОСФЕРЫ И ПЕРВЫЕ РЕЗУЛЬТАТЫ

Цикл экспериментов по электроакустическому зондированию атмосферы был проведён в августе–сентябре 2003 года на полигоне НИРФИ «Васильсурск», расположенном в 120 км к востоку от Нижнего Новгорода. В этих экспериментах в качестве источника звука использовалась низкочастотная сирена с рупорным излучателем.

Сирена подключалась к резервуару объёмом около 10 м^3, в котором поддерживалось избыточное давление около 5 кПа при расходе воздуха 1,5 м^3/с. В горле рупора установлен прерыватель воздушной струи, изготовленный в виде круглого цилиндра с двумя щелевыми отверстиями на поверхности. При вращении цилиндра воздушная струя прерывается дважды за один оборот цилиндра. Рупор длиной 12 м ориентирован горизонтально. Ширина раскрыва рупора 7 м, высота 3,5 м. Параметры акустического излучателя следующие: акустическая мощность 1 кВт, диапазон рабочих частот $17\div24$ Гц, ширина диаграммы направленности по азимуту составляла $\pm60^\circ$, по углу места — $0^\circ\div90^\circ$.

Для регистрации электрического поля использовалась Г-образная антенна с высотой подвеса 3 м и длиной горизонтального плеча 30 м, установленная на расстоянии около 600 м от рупора в зоне минимума его диаграммы направленности. Сигнал с антенны подавался на вход усилителя с полосой пропускания от 5 Гц до 3 кГц.

Сеансы наблюдений проводились как в грозовых условиях (23 августа 2003 года), так и в спокойных условиях (17–19 сентября 2003 года). Наблюдения 17–19 сентября проводились отдельными получасовыми сеансами с 11 до 20 часов (около 5 сеансов в день). В эти дни стояла почти безветренная малооблачная погода. Частота звука во время сеансов изменялась по линейному закону (ЛЧМ сигнал) в пределах $17\div21$ Гц, как правило, с периодом 8 мин, при этом первую половину периода частота росла, вторую — уменьшалась. Для контроля излучаемого акустического сигнала использовался микрофон, установленный вблизи раскрыва рупора. Во время сеансов проводилась регистрация сигналов как с выхода усилителя электрического канала, так и с выхода микрофонного усилителя. Частота оцифровки данных составляла 512 Гц. [1]

Ожидаемый электрический отклик на зондирующий акустический сигнал должен иметь характерный вид зондирующего ЛЧМ сигнала (или отдельных его фрагментов), однако он может быть сдвинут по времени относительно излучаемого акустического сигнала. Величина этого сдвига определяется временем распространения акустического сигнала до активных областей атмосферы, являющихся источниками переменного электрического поля. Кроме того, электрический отклик может быть также несколько размыт и/или сдвинут по частоте из-за эффекта Доплера, связанного с ветровым движением этих областей.

Спектральная обработка записей электрического поля показала, что 18 сентября в двух вечерних сеансах (начало сеансов в 18:18 и 19:24 LT) имелись отдельные, хотя и слабо выраженные, фрагменты ЛЧМ сигнала. На рис. 1 приведены результаты обработки сигналов для сеанса с началом в 18:18 LT (рис. 1*а*) и для сеанса в 19:24 LT (рис. 1*б*). Для обоих сеансов на нижней панели показан динамический спектр зондирующего акустического сигнала, на средней панели — динамический спектр сигнала с выхода электрического канала. Стрелками на динамических спектрах электрического сигнала отмечены те места, где наблюдались фрагменты, соответствующие динамике акустического сигнала. На верхней панели приведены спектральные амплитуды норми-

[1] Для уменьшения влияния импульсных помех на чувствительность датчика электрического поля, что, несомненно, необходимо во время грозы, в последующих экспериментах предполагается расширить полосу частот датчика электрических сигналов до $50\div100$ кГц, а частоту оцифровки соответственно до $150\div300$ тысяч отсчётов в секунду.

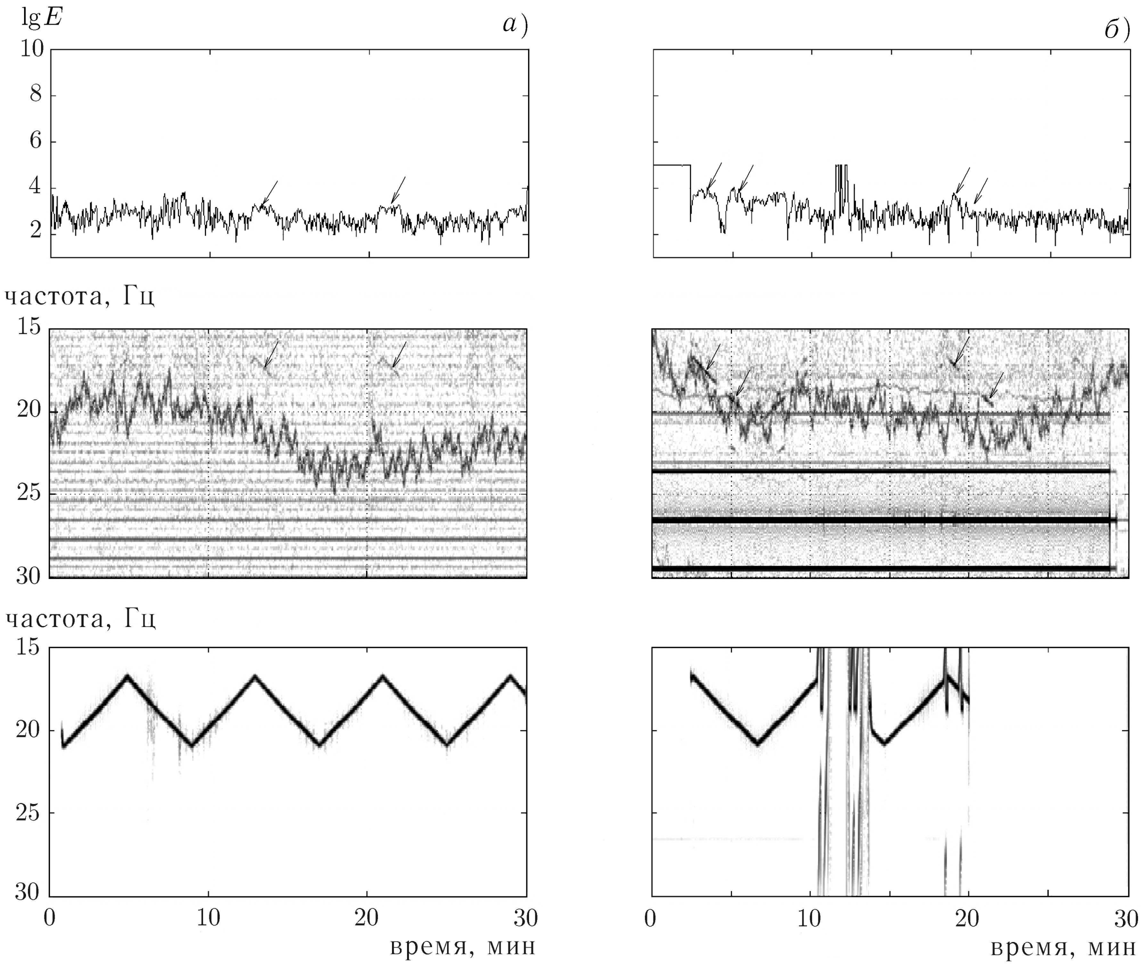

Рис. 1

рованного электрического сигнала (поля E) на частоте, соответствующей частоте акустического сигнала в данный момент времени (сигнала с микрофона). Можно видеть, что на верхних рисунках указанным фрагментам на динамических спектрах электрического канала соответствуют временны́е интервалы с повышенной спектральной амплитудой.

К сожалению, во втором сеансе (с началом в 19:24 LT) с 11 по 13 минуты и около 18 и 19 минуты имел место сбой в работе сирены, что видно на динамическом спектре акустического канала. Для проверки возможного проникновения сигнала с канала на канал микрофон во втором сеансе после двадцатой минуты и до конца сеанса был отключён от регистрирующей аппаратуры (при этом сирена продолжала работать до конца сеанса). Как видно из рисунка, это не привело к изменению характера спектра в электрическом канале. В частности, уже после отключения микрофона (примерно на 21 мин) наблюдался фрагмент ЛЧМ сигнала. Это свидетельствует о том, что наблюдаемые фрагменты ЛЧМ сигнала на динамическом спектре не являются наводками с акустического канала. Следует отметить, что в сеансе с началом в 19:24 LT на верхнем рисунке вблизи 21 мин (указано стрелкой) не наблюдается увеличение спектральной амплитуды электрического сигнала. Это связано с тем, что в отсутствие управляющего акустического сигнала нарушается синхронизация выборки спектральных амплитуд, поэтому сигнал теряется.

Для дальнейшего анализа была использована стандартная методика обработки ЛЧМ сигналов. Излучаемый акустический ЛЧМ сигнал представим в виде

$$U = A\exp(-i\omega t - i\gamma t^2/2), \tag{1}$$

где параметр γ характеризует скорость перестройки частоты.

Пусть область, где расположены заряды, находится на расстоянии r от акустического излучателя и движется со средней скоростью $v \ll c_{\rm s}$, где $c_{\rm s}$ — скорость звука в воздухе. Тогда индуцированное электрическое поле можно представить в виде

$$E \propto B\exp\left[-i\omega\,(t - r/c_{\rm s}) - i\gamma\,(t - r/c_{\rm s})^2/2 + i\mathbf{kv}t\right]. \tag{2}$$

В последнем выражении учтено запаздывание акустического воздействия на время $t_0 = r/c_{\rm s}$, а в показателе экспоненты появилось дополнительное слагаемое $\mathbf{kv}t$, учитывающее доплеровское смещение частоты звуковой волны, воздействующей на движущийся объект — электрические ячейки (здесь $\mathbf{k}$ — волновой вектор звуковой волны).

Перемножив (1) и (2), получим

$$S(t) = UE^* \propto \exp\left[-i\Omega t - i\left(\omega\,\frac{r}{c_{\rm s}} - \frac{\gamma}{2}\,\frac{r^2}{c_{\rm s}^2}\right)\right], \tag{3}$$

где Ω — частота преобразованного электрического сигнала:

$$\Omega = \gamma r/c_{\rm s} + \mathbf{kv}. \tag{4}$$

Как видно из (4), частота Ω определяется как скоростью перестройки частоты γ и расстоянием r от источника звука до области генерации, так и доплеровским сдвигом $\mathbf{kv}$. В нашем случае $\gamma \approx \pm 0{,}1$ рад/с^2, и для расстояния r порядка одного километра величина $\gamma r/(2\pi c_{\rm s}) \approx$ $\approx \pm 0{,}05$ Гц. Величина и знак доплеровского сдвига определяется произведением $\mathbf{kv}$. Для продольной (вдоль $\mathbf{k}$) составляющей скорости $v_\parallel \sim 1$ м/с величина $|\mathbf{kv}|/(2\pi)$ приблизительно равна 0,06 Гц. Поэтому для определения расстояния до области генерации необходимо проводить фурье-анализ сигнала $S(t)$ с разрешением 0,01÷0,02 Гц.

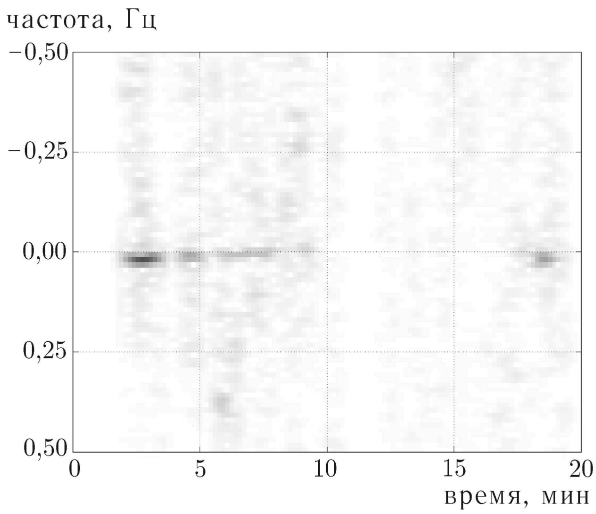

Рис. 2

Дальнейшая обработка акустического и электрического сигналов проводилась для сеанса с началом в 19:24 LT 18.09.2003, в котором уровень электрических помех был меньше. Проводилась фильтрация сигналов в полосе 15÷25 Гц. Затем с помощью преобразования Гильберта [3] исходные сигналы были представлены в экспоненциальном виде с мнимыми аргументами. Затем оба сигнала перемножались, после чего проводился фурье-анализ произведения с использованием окна Хэмминга. При длительности выборки 80 секунд частотное разрешение составляло около 0,012 Гц. Фурье-анализ проводился по выборкам с большим взаимным перекрытием по времени.

Результаты обработки приведены на рис. 2, из которого видно, что сдвиг частоты сигнала $F \leq 0{,}02$ Гц. Отсюда, в частности, следует, что наблюдаемый эффект не может быть связан с непосредственным воздействием акустического сигнала на электрическую антенну, т. к. в этом случае частота сигнала должна изменять знак при смене знака параметра γ и составлять не менее

0,03 Гц (4), поскольку в условиях эксперимента ветер отсутствовал, а расстояние между акустическим источником и антенной составляло около 600 м.

Одним из возможных объяснений наблюдаемого малого сдвига по частоте принимаемого сигнала может быть то, что электрическое поле индуцируется вблизи акустического излучателя. Проведём оценки возможного электрического сигнала в этом случае. Ток хорошей погоды $J \sim 2 \cdot 10^{-12}$ А/м^2, подвижность частиц 0,01 м/с. Из этого следует, что объёмная плотность зарядов одного знака $qN \sim 2 \cdot 10^{-10}$ Кл/м^3. При мощности излучателя 1 кВт осцилляторная скорость частиц воздуха в поле акустической волны вблизи раскрыва рупора будет порядка 0,1 м/с, а амплитуда осцилляций порядка $\triangle x \sim 10^{-3}$ м. Объём ближней зоны $V \sim \lambda^3 \sim 10^4$ м^3. Таким образом, дипольный момент $P \sim qNV \triangle x$ может составить величину порядка 10^{-9} Кл$\cdot$м. В этом случае напряжённость электрического поля на расстоянии $L \sim 600$ м составит $E = P/(4\pi\varepsilon_0 r^3) \sim 10^{-7} \div 10^{-6}$ В/м, что по порядку величины соответствует наблюдаемому значению.

2. ЗАКЛЮЧЕНИЕ

Проведённые пробные эксперименты по электроакустическому зондированию атмосферы показали принципиальную возможность получения электрического отклика при акустическом воздействии на заряженные области атмосферы. Дальнейшее развитие методики и техники эксперимента, в частности построение более мощного низкочастотного акустического излучателя, позволит успешно реализовать этот новый метод дистанционного зондирования для исследования электрического состояния нижней атмосферы.

Работа выполнена в рамках программы ОФН РАН «Физика атмосферы: электрические процессы, радиофизические методы исследования», при поддержке РФФИ (грант № 01–02–16680) и ИНТАС (грант № 01–0456).

СПИСОК ЛИТЕРАТУРЫ

1. Поляков С. В., Рапопорт В. О., Трахтенгерц В. Ю. // Изв. вузов. Радиофизика. 1992. Т. 35, № 1. С. 15.
2. Зиничев В. А., Рапопорт В. О., Трахтенгерц В. Ю. и др. // Изв. вузов. Радиофизика. 1991. Т. 34, № 3. С. 234.
3. Marple S. L. // IEEE Trans. Signal Processing. 1999. V. 47, No. 9. P. 2600.

Поступила в редакцию 30 декабря 2003 г.; принята в печать 5 июля 2004 г.

ELECTRO-ACOUSTIC SOUNDING OF THE ATMOSPHERE

V. O. Rapoport, N. A. Mityakov, V. A. Zinichev, G. P. Komrakov, N. A. Ryzhov, and Yu. A. Sazonov

This paper presents the results of measurement of the electrical field generated by a high-power acoustic signal in the atmosphere (electro-acoustic sounding). The radiated acoustic power was about 1 kW. The acoustic-signal frequency varied linearly within the interval 17−21 Hz with a period of 8 min. The electric field was registered by a Г-shaped antenna. The results of processing of the electric-field records for a part of sessions indicate the presence of an electric response with frequency close to the acoustic-signal frequency. The signal amplitude at the input of the receiving device was 0.1−1 μV.

УДК 535.361

НАБЛЮДЕНИЕ ПОДВОДНЫХ ОБЪЕКТОВ ЧЕРЕЗ БЛИКОВЫЕ УЧАСТКИ МОРСКОЙ ПОВЕРХНОСТИ

В. Л. Вебер

Институт прикладной физики РАН, г. Нижний Новгород, Россия

Рассматривается возможность уменьшения негативного влияния волнения на видимость подводных объектов, основанная на использовании информации о бликах на морской поверхности, необходимой для определения не искажённых волнением участков в изображении подводного объекта. Исследовано влияние углового разрешения системы видения на качество изображения подводных объектов.

ВВЕДЕНИЕ

Наблюдение подводного объекта через взволнованную морскую поверхность, как правило, сопровождается геометрическими искажениями изображения, вызванными отклонениями световых лучей от направления прямолинейного распространения при преломлении их на неровной границе раздела воздух—вода. Эти искажения проявляются в случайном смещении и дроблении (размножении) точек изображения объекта, зачастую настолько значительным, что распознавание объекта становится невозможным. Устранение такого рода геометрических искажений представляется привлекательной, хотя и весьма сложной задачей.

Одной из плодотворных идей по устранению влияния волнения на видимость подводных объектов является использование информации о бликах на взволнованной поверхности для определения тех участков в изображении подводного объекта, которые образованы световыми лучами, прошедшими через поверхность воды без изменения направления распространения. Суть этой идеи, принадлежащей И. М. Левину и высказанной им в частной беседе в 1980 году, заключается в следующем.

Допустим, что система наблюдения подводного объекта содержит импульсный источник подсветки и два стробирумых по времени приёмника, один из которых открывается в момент прихода импульса излучения от поверхности воды, а другой — в момент прихода импульса от поверхности объекта, расположенного на глубине h. Световое излучение, отражённое объектом, формирует изображение, распределение яркости которого описывается выражением [1]

$$B_1(\mathbf{r}) = B_0[\mathbf{r} + a\mathbf{q}(\mathbf{r})], \tag{1}$$

где $B_0(\mathbf{r})$ — распределение яркости на объекте, $\mathbf{q}(\mathbf{r})$ — вектор-градиент взволнованной морской поверхности, $a = h\,(m-1)/m$, m — показатель преломления воды, $\mathbf{r} = x\mathbf{x}_0 + y\mathbf{y}_0$ — радиус-вектор в горизонтальной плоскости, $\mathbf{x}_0$ и $\mathbf{y}_0$ — орты декартовых осей x и y. Выражение (1) получено для условий надирного наблюдения подводного объекта в пренебрежении эффектами рассеяния светового излучения в водной толще и в приближении малости углов наклона границы раздела воздух—вода.

Излучение, отражённое водной поверхностью, формирует изображение бликов, нормированная яркость которого описывается выражением (см. следствие формулы (15))

$$T(\mathbf{r}) = \exp[-4\pi q^2(\mathbf{r})/\Delta_{\mathrm{s}}], \tag{2}$$

где $\Delta_{\rm s}$ — телесный угол, под которым с взволнованной морской поверхности виден источник света. Заметим, что выражение (2) может быть получено из простых эвристических соображений при выполнении следующих условий:

а) поверхность освещается изотропным источником с гауссовым распределением яркости по апертуре;

б) горизонтальные координаты центра излучающей апертуры и приёмника совпадают;

в) наблюдение ведётся в надир.

Эти условия используются и при выводе (15). Они не носят принципиального характера, но существенно упрощают выкладки.

Используя бликовое изображение (2) в качестве маски (транспаранта), можно выделить в искажённом изображении объекта (1) лишь те области, которые соответствуют горизонтальным участкам морской поверхности. В итоге получается хотя и фрагментарное, но неискажённое изображение подводного объекта в точках, соответствующих положению бликов от поверхности:

$$B_2(\mathbf{r}) = B_1(\mathbf{r}) T(\mathbf{r}). \tag{3}$$

Поскольку состояние водной поверхности меняется во времени, описанную процедуру можно многократно повторять через промежуток времени, определяемый временем корреляции изображения бликов. Складывая (накапливая) множество различных фрагментарных изображений (3), можно получить изображение, достаточно близкое к неискажённому изображению объекта. Покажем это. Поскольку процедура накопления с формальной точки зрения идентична операции статистического усреднения по случайным реализациям уклонов взволнованной морской поверхности, то распределение яркости в изображении (3), накопленном за достаточно большой промежуток времени, может быть описано следующим образом:

$$\overline{B}_2(\mathbf{r}) = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} B_0(\mathbf{r} + a\mathbf{q}) \exp(-4\pi q^2/\Delta_{\rm s}) w_1(\mathbf{q})\, \mathrm{d}\mathbf{q}, \tag{4}$$

где $w_1(\mathbf{q}) = (2\pi\sigma_q^2)^{-1} \exp[-q^2/(2\sigma_q^2)]$ — одноточечная функция распределения вероятностей уклонов морской поверхности, σ_q^2 — дисперсия уклонов. Здесь для простоты взята модель статистически изотропной взволнованной поверхности. Ниже, в основной части работы, рассматривается одномерная модель морского ветрового волнения.

Выберем в качестве тестового объект с синусоидальным распределением яркости:

$$B_0(\mathbf{r}) = B_0\, [1 + K_0 \cos(\boldsymbol{\omega}_0 \mathbf{r})], \tag{5}$$

где $\boldsymbol{\omega}_0$ — пространственная частота распределения яркости. Подставляя выражение (5) в формулу (4), в результате несложных преобразований получим

$$\overline{B}_2(\mathbf{r}) = B_0\, \frac{\Delta'_{\rm s}}{\Delta'_{\rm s} + \Delta_w}\, [1 + K_0 K_{\rm i} \cos(\boldsymbol{\omega}_0 \mathbf{r})], \tag{6}$$

где $K_{\rm i} = \exp[-\Delta_{\rm s}^0 a^2 \omega_0^2/(4\pi)]$, $\Delta_{\rm s}^0 = \Delta'_{\rm s}\Delta_w/(\Delta'_{\rm s} + \Delta_w)$, $\Delta'_{\rm s} = \Delta_{\rm s}/4$, $\Delta_w = 2\pi\sigma_q^2$. Параметр $K_{\rm i}$ в полученном выражении определяет контраст усреднённого (накопленного) изображения тестового объекта. В том случае, когда информация о бликах не используется ($T(\mathbf{r}) \equiv 1$, или $\Delta_{\rm s} \to \infty$, что соответствует обычной системе наблюдения), из (6) следует известный результат, впервые полученный в [2]:

$$K_{\rm i} = \exp[-\Delta_w a^2 \omega_0^2/(4\pi)]. \tag{7}$$

При использовании информации о бликах (при обычно выполняющемся условии $\Delta_{\rm s} \ll \Delta_w$) контраст изображения оказывается существенно более высоким:

$$K_{\rm i} = \exp[-\Delta'_{\rm s} a^2 \omega_0^2/(4\pi)]. \tag{8}$$

Вышеизложенные соображения свидетельствуют о принципиальных преимуществах бликовой системы наблюдения подводных объектов перед обычными системами. Однако изложенное выше представляет собой идеальную схему. В ней используется предположение о системе наблюдения (как бликов, так и объектов) с бесконечной разрешающей способностью (т. е. приёмник представляет собой яркомер). Между тем на практике приходится иметь дело с приёмниками светового излучения, имеющими конечную разрешающую способность. Нетрудно, например, представить себе ситуацию, в которой размер пятна разрешения приёмника на поверхности воды настолько велик, что на нём укладываются несколько элементов водной поверхности, имеющие нулевой наклон. При этом будет принято решение о наличии в данном элементе изображения не нескольких, а одного блика. Это решение, разумеется, является ошибочным. Анализу влияния такого рода ошибок при реставрации изображения подводного объекта посвящены следующие разделы статьи.

1. МОДЕЛЬ ПЕРЕНОСА ИЗОБРАЖЕНИЯ ПОДВОДНОГО ОБЪЕКТА ЧЕРЕЗ МОРСКУЮ ПОВЕРХНОСТЬ И ТОЛЩУ ВОДЫ

Рассмотрим схему формирования изображения подводного объекта, приведённую на рис. 1. Над водой на высоте H в плоскости z_1 расположена система наблюдения подводного объекта с приёмной диаграммой, описываемой апертурной функцией $D_1(\mathbf{r}_1 - \mathbf{r}, \mathbf{\Omega}'_1 - \mathbf{\Omega}')$. Здесь $\mathbf{r}$ — координата центра апертуры приёмника, $\mathbf{\Omega}^0$ — единичный вектор, определяющий направление визирования, $\mathbf{\Omega}'_i$ — проекция вектора $\mathbf{\Omega}_i^0$ на плоскость $z = \mathrm{const}$. Система наблюдения формирует изображение равномерно освещённого объекта, находящегося на глубине h в плоскости z_3 и имеющего распределение коэффициента отражения $R_0(\mathbf{r}_3)$, через случайно-неровную поверхность, заданную функцией уклонов $\mathbf{q}(\mathbf{r}_2)$ в плоскости z_2.

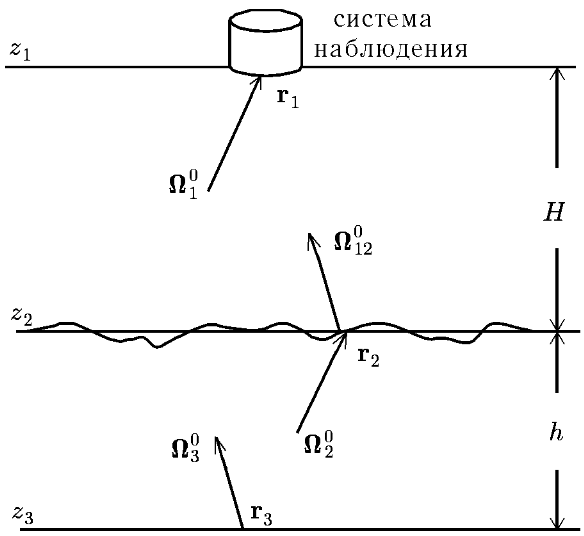

Рис. 1. Схема наблюдения подводного объекта

Выведем соотношение для мощности принимаемого оптического сигнала, опираясь на схему рассуждений, приведённую в [3–5]. Элементарный световой поток $\mathrm{d}F_0 = B_0(\mathbf{r}_3)\,\mathrm{d}\mathbf{r}_3\,\mathrm{d}\mathbf{\Omega}'_3$, где $B_0(\mathbf{r}_3) = E_0 R_0(\mathbf{r}_3)/\pi$, E_0 — освещённость в плоскости z_3, идущий от объекта в направлении $\mathbf{\Omega}_3^0$, создаёт в плоскости z_2 поле с яркостью

$$B(\mathbf{r}_2, \mathbf{\Omega}'_2) = \int \ldots \int \mathrm{d}F_0\, G_{\rm c}(\mathbf{r}_3 \to \mathbf{r}_2, \mathbf{\Omega}'_3 \to \mathbf{\Omega}'_2),$$

где $G_{\rm c}(\mathbf{r}_3 \to \mathbf{r}_2, \mathbf{\Omega}'_3 \to \mathbf{\Omega}'_2)$ — функция Грина уравнения переноса излучения в водной среде, определяющая яркость светового поля в точке $\mathbf{r}_2$ и в направлении $\mathbf{\Omega}_2^0$ при облучении среды точечным мононаправленным источником света с единичной мощностью, расположенным в точке $\mathbf{r}_3$ и излучающим в направлении $\mathbf{\Omega}_3^0$. Яркость светового излучения непосредственно над границей воздух—вода определяется соотношением

$$B(\mathbf{r}_2, \mathbf{\Omega}'_{12}) = \iint B(\mathbf{r}_2, \mathbf{\Omega}'_2) G_{\mathrm{w}}(\mathbf{\Omega}'_2 \to \mathbf{\Omega}'_{12}, \mathbf{r}_2)\, \mathrm{d}\mathbf{\Omega}'_2,$$

где $G_{\mathrm{w}}(\mathbf{\Omega}'_2 \to \mathbf{\Omega}'_{12}, \mathbf{r}_2)$ — функция Грина поверхности раздела, определяющая изменение направления распространения световых лучей при прохождении границы раздела сред с различным показателем преломления.

Яркость светового поля в плоскости z_1 (на входе оптического приёмника) определяется соотношением

$$B(\mathbf{r}_1, \mathbf{\Omega}'_1) = \int \ldots \int B(\mathbf{r}_2, \mathbf{\Omega}'_{12}) G_{\mathrm{a}}(\mathbf{r}_2 \to \mathbf{r}_1, \mathbf{\Omega}'_{12} \to \mathbf{\Omega}'_1)\, \mathrm{d}\mathbf{r}_2\, \mathrm{d}\mathbf{\Omega}'_{12},$$

где $G_{\mathrm{a}}(\mathbf{r}_2 \to \mathbf{r}_1, \mathbf{\Omega}'_{12} \to \mathbf{\Omega}'_1)$ — функция Грина уравнения переноса излучения в атмосфере.

Мощность светового излучения, принятого фотоприёмником системы наблюдения, составляет

$$P_1(\mathbf{r}, \mathbf{\Omega}') = \int \ldots \int D_1(\mathbf{r}_1 - \mathbf{r}, \mathbf{\Omega}'_1 - \mathbf{\Omega}') B(\mathbf{r}_1, \mathbf{\Omega}'_1)\, \mathrm{d}\mathbf{r}_1\, \mathrm{d}\mathbf{\Omega}'_1.$$

Данное выражение после использования оптической теоремы взаимности может быть записано в следующей обобщённой форме:

$$P_1(\mathbf{r}, \mathbf{\Omega}) = \iint B_0(\mathbf{r}_3) E_1(\mathbf{r}_3, \mathbf{r}, \mathbf{\Omega})\, \mathrm{d}\mathbf{r}_3, \tag{9}$$

где

$$E_1(\mathbf{r}_3, \mathbf{r}, \mathbf{\Omega}) = \int \ldots \int D_1(\mathbf{r}_1 - \mathbf{r}, \mathbf{\Omega}_1 - \mathbf{\Omega}) e_1(\mathbf{r}_1 \to \mathbf{r}_3, \mathbf{\Omega}_1)\, \mathrm{d}\mathbf{r}_1\, \mathrm{d}\mathbf{\Omega}_1,$$

$$e_1(\mathbf{r}_1 \to \mathbf{r}_3, \mathbf{\Omega}_1) = \int \ldots \int G_{\mathrm{a}}(\mathbf{r}_1 \to \mathbf{r}_2, \mathbf{\Omega}_1 \to \mathbf{\Omega}_{12}) G_{\mathrm{w}}(\mathbf{\Omega}_{12} \to \mathbf{\Omega}_2, \mathbf{r}_2) e_{\mathrm{c}}(\mathbf{r}_2 \to \mathbf{r}_3, \mathbf{\Omega}_2)\, \mathrm{d}\mathbf{r}_2\, \mathrm{d}\mathbf{\Omega}_{12}\, \mathrm{d}\mathbf{\Omega}_2,$$

$$e_{\mathrm{c}}(\mathbf{r}_2 \to \mathbf{r}_3, \mathbf{\Omega}_2) = \iint G_{\mathrm{c}}(\mathbf{r}_2 \to \mathbf{r}_3, \mathbf{\Omega}_2 \to \mathbf{\Omega}_3)\, \mathrm{d}\mathbf{\Omega}_3;$$

здесь осуществлена замена переменных $\mathbf{\Omega}_i = -\mathbf{\Omega}'_i$.

Входящие в данное выражение функции имеют простой физический смысл. Функция $e_1(\mathbf{r}_1 \to \mathbf{r}_3, \mathbf{\Omega}_1)$ определяет распределение освещённости в точке $\mathbf{r}_3$ плоскости z_3 при облучении «составной» среды (атмосфера, граница раздела, водная толща) точечным мононаправленным источником единичной мощности, расположенным в точке $\mathbf{r}_1$ и излучающим в направлении $-\mathbf{\Omega}_1$. Функция $E_1(\mathbf{r}_3, \mathbf{r}, \mathbf{\Omega})$ описывает распределение освещённости в плоскости z_3, создаваемое источником единичной мощности с диаграммой направленности $D_1(\mathbf{r}_1 - \mathbf{r}, \mathbf{\Omega}_1 - \mathbf{\Omega})$.

В условиях надирного наблюдения (под небольшими углами к вертикали) и в предположении малости уклонов взволнованной морской поверхности для всех функций Грина можно использовать малоугловое приближение, которое формулируется следующим образом:

$$G_{\mathrm{w}}(\mathbf{\Omega}_{12} \to \mathbf{\Omega}_2, \mathbf{r}_2) = \delta[\mathbf{\Omega}_{12} - m\mathbf{\Omega}_2 + (m-1)\,\mathbf{q}(\mathbf{r}_2)]/m^2, \tag{10}$$

$$e_{\mathrm{c}}(\mathbf{r}_2 \to \mathbf{r}_3, \mathbf{\Omega}_2) = e_{\mathrm{c}}(\mathbf{r}_3 - \mathbf{r}_2 - \mathbf{\Omega}_2 h), \tag{11}$$

где $\delta(\mathbf{r})$ — дельта-функция Дирака.

Поскольку наблюдение подводных объектов обычно ведётся со сравнительно небольших высот, эффектами рассеяния света в атмосфере можно пренебречь. Функция Грина слоя атмосферы при этом имеет вид

$$G_{\mathrm{a}}(\mathbf{r}_1 \to \mathbf{r}_2, \mathbf{\Omega}_1 \to \mathbf{\Omega}_{12}) = \delta(\mathbf{r}_2 - \mathbf{r}_1 - \mathbf{\Omega}_1 H)\delta(\mathbf{\Omega}_{12} - \mathbf{\Omega}_1). \tag{12}$$

Подставляя соотношения (10)–(12) в выражение (9), после несложных преобразований получим

$$P_1(\mathbf{r}, \mathbf{\Omega}) = \frac{1}{(2\pi)^4\, m^2} \int\limits_{-\infty}^{+\infty} \ldots \int\limits_{-\infty}^{+\infty} F_0(\mathbf{k}) F_{\mathrm{c}}(\mathbf{k}) F_1(\boldsymbol{\omega}, \boldsymbol{\omega} H - \mathbf{k} h_0) \times \\ \times \exp\bigl\{i\,[a\mathbf{k}\mathbf{q}(\mathbf{r}_2) + (\boldsymbol{\omega} + \mathbf{k})\,\mathbf{r}_2 - \boldsymbol{\omega}\mathbf{r} - (\boldsymbol{\omega} H - \mathbf{k} h_0)\,\mathbf{\Omega}]\bigr\}\, \mathrm{d}\mathbf{k}\, \mathrm{d}\boldsymbol{\omega}\, \mathrm{d}\mathbf{r}_2, \tag{13}$$

где $F_0(\mathbf{k})$, $F_{\mathrm{c}}(\mathbf{k})$, $F_1(\mathbf{k}, \mathbf{p})$ — фурье-образы функций $B_0(\mathbf{r})$, $e_{\mathrm{c}}(\mathbf{r})$, $D_1(\mathbf{r}, \mathbf{\Omega})$ (функции F_1 и F_{c} называются оптическими передаточными функциями приёмника и среды), $a = (m-1)\, h_0$, $h_0 = h/m$.

Для упрощения дальнейшего анализа сделаем ряд допущений.

Как видно из выражения (13), изображение подводного объекта может быть сформировано посредством сканирования — углового (центр апертуры приёмника $\mathbf{r}$ неподвижен, направление визирования $\mathbf{\Omega}$ изменяется) либо пространственного (направление $\mathbf{\Omega}$ фиксировано, $\mathbf{r}$ изменяется). В дальнейшем мы будем рассматривать вариант пространственного сканирования объекта наблюдения с направлением визирования в надир ($|\mathbf{\Omega}| = 0$). Заметим, что этот выбор сделан лишь из соображений упрощения математической стороны задачи. Кроме того, без ущерба для общности рассуждений мы не будем учитывать конечность размеров апертуры приёмника системы наблюдения, т. е. положим $F_1(\mathbf{k}, \mathbf{p}) = \Sigma_1 F_1(\mathbf{p})$, где Σ_1 — площадь апертуры приёмника. И, наконец, входящее в выражение (13) произведение спектров объекта и среды $F_0(\mathbf{k}) F_{\mathrm{c}}(\mathbf{k})$ мы в дальнейшем для краткости будем называть спектром объекта и обозначать его $F_0(\mathbf{k})$.

С учётом этих допущений выражение для мощности оптического сигнала в элементе изображения подводного объекта принимает вид

$$P_1(\mathbf{r}) = \frac{\Sigma_1}{(2\pi)^4\, m^2} \int\limits_{-\infty}^{+\infty} \ldots \int\limits_{-\infty}^{+\infty} F_0(\mathbf{k}) F_1(\boldsymbol{\omega} H - \mathbf{k} h_0) \exp\bigl\{i\,[a\mathbf{k}\mathbf{q}(\mathbf{r}_2) + (\boldsymbol{\omega} + \mathbf{k})\mathbf{r}_2 - \boldsymbol{\omega}\mathbf{r}]\bigr\}\, \mathrm{d}\mathbf{k}\, \mathrm{d}\boldsymbol{\omega}\, \mathrm{d}\mathbf{r}_2. \tag{14}$$

Заметим, что формула (1), описывающая распределение яркости изображения подводного объекта, является частным случаем общих формул (13) и (14). Нетрудно видеть, что она следует из выражения (14) при условии $F_1(\mathbf{p}) = \mathrm{const}$; это условие определяет приёмник системы наблюдения с бесконечным угловым разрешением (яркомер).

2. МОДЕЛЬ ФОРМИРОВАНИЯ ИЗОБРАЖЕНИЯ БЛИКОВ НА ВЗВОЛНОВАННОЙ МОРСКОЙ ПОВЕРХНОСТИ

Рассмотрим схему наблюдения морской поверхности (рис. 2). На высоте H_{s} над водой расположен изотропный источник излучения с распределением яркости $B_{\mathrm{s}}(\mathbf{r}_3 - \mathbf{r}) = B_{\mathrm{s}} D_{\mathrm{s}}(\mathbf{r}_3 - \mathbf{r})$, где $D_{\mathrm{s}}(\mathbf{r}_3 - \mathbf{r})$ — апертурная функция источника, $\mathbf{r}$ — координата центра апертуры. На высоте H расположен фотоприёмник системы наблюдения бликов, имеющий приёмную диаграмму $D_2(\mathbf{r}_1 - \mathbf{r}, \mathbf{\Omega}_1)$. Система наблюдения, ориентированная в надир, формирует изображение источника излучения после отражения света от взволнованной поверхности, заданной функцией уклонов

$\mathbf{q}(\mathbf{r}_2)$. Как и в предыдущем разделе, мы полагаем, что изображение бликов формируется посредством пространственного сканирования морской поверхности.

Соотношение для мощности принимаемого оптического сигнала может быть получено на основе рассуждений, приведённых в предыдущем разделе. Аналогично (9) оно представляется в виде

$$P_2(\mathbf{r}) = B_{\mathrm{s}} \iint D_{\mathrm{s}}(\mathbf{r}_3 - \mathbf{r}) E_2(\mathbf{r}_3, \mathbf{r})\, \mathrm{d}\mathbf{r}_3,$$

где

$$E_2(\mathbf{r}_3, \mathbf{r}) = \\ = \int \ldots \int D_2(\mathbf{r}_1 - \mathbf{r}, \boldsymbol{\Omega}_1) e_2(\mathbf{r}_1 \to \mathbf{r}_3, \boldsymbol{\Omega}_1)\, \mathrm{d}\mathbf{r}_1\, \mathrm{d}\boldsymbol{\Omega}_1,$$

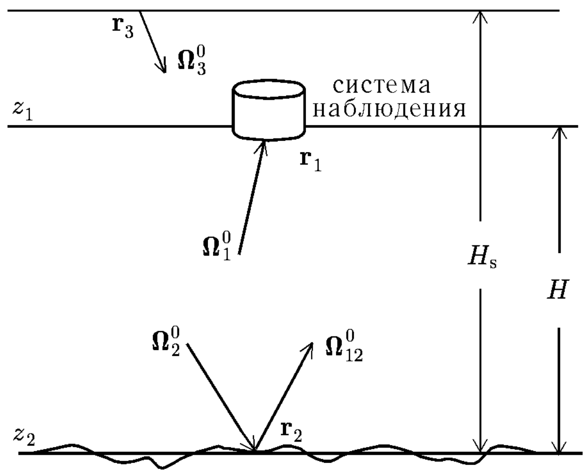

Рис. 2. Схема наблюдения бликов на морской поверхности

$$e_2(\mathbf{r}_1 \to \mathbf{r}_3, \boldsymbol{\Omega}_1) = \int \ldots \int G'_{\mathrm{a}}(\mathbf{r}_1 \to \mathbf{r}_2, \boldsymbol{\Omega}_1 \to \boldsymbol{\Omega}_{12}) \times \\ \times G'_{\mathrm{w}}(\boldsymbol{\Omega}_{12} \to \boldsymbol{\Omega}_2, \mathbf{r}_2) G''_{\mathrm{a}}(\mathbf{r}_2 \to \mathbf{r}_3, \boldsymbol{\Omega}_2 \to \boldsymbol{\Omega}_3)\, \mathrm{d}\mathbf{r}_2\, \mathrm{d}\boldsymbol{\Omega}_{12}\, \mathrm{d}\boldsymbol{\Omega}_2\, \mathrm{d}\boldsymbol{\Omega}_3.$$

В условиях надирного наблюдения и в приближении малости углов наклона взволнованной морской поверхности входящие в это выражение функции Грина имеют вид

$$G'_{\mathrm{a}}(\mathbf{r}_1 \to \mathbf{r}_2, \boldsymbol{\Omega}_1 \to \boldsymbol{\Omega}_{12}) = \delta(\mathbf{r}_2 - \mathbf{r}_1 - H\boldsymbol{\Omega}_1)\delta(\boldsymbol{\Omega}_{12} - \boldsymbol{\Omega}_1),$$

$$G'_{\mathrm{w}}(\boldsymbol{\Omega}_{12} \to \boldsymbol{\Omega}_2, \mathbf{r}_2) = \delta[\boldsymbol{\Omega}_{12} - \boldsymbol{\Omega}_2 - 2\mathbf{q}(\mathbf{r}_2)],$$

$$G''_{\mathrm{a}}(\mathbf{r}_2 \to \mathbf{r}_3, \boldsymbol{\Omega}_2 \to \boldsymbol{\Omega}_3) = \delta(\mathbf{r}_3 - \mathbf{r}_2 - H_{\mathrm{s}}\boldsymbol{\Omega}_2)\delta(\boldsymbol{\Omega}_3 - \boldsymbol{\Omega}_2).$$

Подставляя полученные соотношения в выражение для $P_2(\mathbf{r})$, после преобразований получим выражение для яркости изображения бликующей поверхности моря в следующей «частотной» форме:

$$P_2(\mathbf{r}) = \frac{B_{\mathrm{s}}\Sigma_2}{(2\pi)^4} \int\limits_{-\infty}^{+\infty} \ldots \int\limits_{-\infty}^{+\infty} F_{\mathrm{s}}^0(-\mathbf{p}) F_2(\mathbf{p} + \mathbf{k}H) \exp\bigl\{i\,[(\mathbf{k} - \mathbf{p}/H_{\mathrm{s}})\,(\mathbf{r}_2 - \mathbf{r}) + 2\mathbf{p}\mathbf{q}(\mathbf{r}_2)]\bigr\}\, \mathrm{d}\mathbf{k}\, \mathrm{d}\mathbf{p}\, \mathrm{d}\mathbf{r}_2,$$

где $F_{\mathrm{s}}^0(\mathbf{p}) = H_{\mathrm{s}}^{-2} F_{\mathrm{s}}(\mathbf{p}/H_{\mathrm{s}})$, функции $F_{\mathrm{s}}(\mathbf{p})$ и $F_2(\mathbf{p})$ — фурье-образы функций $D_{\mathrm{s}}(\boldsymbol{\Omega})$ и $D_2(\boldsymbol{\Omega})$ (как и в предыдущем разделе, мы пренебрегли здесь конечностью размеров апертуры приёмника).

Для упрощения дальнейших рассуждений допустим, что источник света расположен достаточно высоко над поверхностью воды ($H_{\mathrm{s}} \to \infty$). При этом выражение для $P_2(\mathbf{r})$, фактически, описывает яркость изображения солнечных бликов на взволнованной поверхности моря (см. также [6]):

$$P_2(\mathbf{r}) = \frac{B_{\mathrm{s}}\Sigma_2}{(2\pi)^4} \int\limits_{-\infty}^{+\infty} \ldots \int\limits_{-\infty}^{+\infty} F_{\mathrm{s}}(-\mathbf{p}) F_2(\mathbf{p} + \mathbf{k}H) \exp\bigl\{i\,[\mathbf{k}(\mathbf{r}_2 - \mathbf{r}) + 2\mathbf{p}\mathbf{q}(\mathbf{r}_2)]\bigr\}\, \mathrm{d}\mathbf{k}\, \mathrm{d}\mathbf{p}\, \mathrm{d}\mathbf{r}_2, \qquad (15)$$

где $F_{\mathrm{s}}(\mathbf{p})$ — фурье-спектр углового распределения яркости Солнца.

Допустим, что приёмник системы наблюдения идеальный, т. е. $F_2(\mathbf{p}+\mathbf{k}H) \equiv \Delta_2$, где Δ_2 — телесный угол приёма излучения, а угловой спектр излучения описывается гауссовой функцией $F_\mathrm{s}(\mathbf{p}) = \Delta_\mathrm{s}\exp[-\Delta_\mathrm{s}p^2/(4\pi)]$. При этом из выражения (15) следует простое соотношение для яркости изображения солнечных бликов, тождественное формуле (2):

$$P_2(\mathbf{r}) = P_0 \exp\left[-q^2(\mathbf{r})\,\frac{4\pi}{\Delta_\mathrm{s}}\right],$$

где $P_0 = B_\mathrm{s}\Sigma_2\Delta_2$.

3. ХАРАКТЕРИСТИКИ ФРАГМЕНТАРНОГО ИЗОБРАЖЕНИЯ

Выражение (14) описывает распределение яркости искажённого изображения подводного объекта, а (15) — распределение яркости изображения бликующей поверхности моря. Произведение

$$P_\mathrm{i}(\mathbf{r}) = P_1(\mathbf{r})P_2(\mathbf{r}) \tag{16}$$

описывает распределение яркости во «фрагментарно неискажённом» изображении объекта, наблюдаемого через взволнованную морскую поверхность. Накопление реализаций (16) (при достаточно большом их числе) позволяет устранить фрагментарность изображения. Статистически усреднённое по реализациям уклонов взволнованной поверхности выражение (16) определяет распределение яркости в накопленном изображении, которое, как мы надеемся, будет достаточно близким к идеальному изображению, полученному при наблюдении подводного объекта через плоскую границу раздела воздух—вода. Так ли это — покажет дальнейший анализ.

Усредняя произведение функций (14), (15) по реализациям уклонов поверхности $\mathbf{q}$, получим следующую формулу:

$$\bar{P}_\mathrm{i}(\mathbf{r}) = \frac{B_\mathrm{s}\Sigma_1\Sigma_2}{m^2\,(2\pi)^6}\int\limits_{-\infty}^{+\infty}\dots\int\limits_{-\infty}^{+\infty} F_0(\mathbf{k})F_\mathrm{s}(-\mathbf{p})F_1(-\boldsymbol{\omega}H-\mathbf{k}L)\,\times$$
$$\times\, F_2(\mathbf{p}-\boldsymbol{\omega}H)\theta_2(-a\mathbf{k},2\mathbf{p},\boldsymbol{\rho})\exp[i\,(\boldsymbol{\omega}\boldsymbol{\rho}+\mathbf{k}\mathbf{r})]\,\mathrm{d}\mathbf{k}\,\mathrm{d}\boldsymbol{\omega}\,\mathrm{d}\mathbf{p}\,\mathrm{d}\boldsymbol{\rho}, \tag{17}$$

где $\theta_2(-a\mathbf{k},2\mathbf{p},\boldsymbol{\rho})$ — двухточечная характеристическая функция распределения вероятностей уклонов взволнованной морской поверхности, $L = H + h/m$.

Для характеристики качества системы формирования изображения, как и ранее, воспользуемся тестовым объектом с синусоидальным распределением яркости:

$$B_0(\mathbf{r}) = B_0\,[1 + K_0\cos(\boldsymbol{\omega}_0\mathbf{r})],$$

где $\boldsymbol{\omega}_0 = \mathbf{x}_0\omega_0$, K_0 — контраст синусоидальной структуры яркости на объекте. Пространственный спектр распределения яркости на объекте имеет вид

$$F_0(\mathbf{k}) = (2\pi)^2\,B_0\left\{\delta(\mathbf{k}) + \frac{K_0}{2}\,[\delta(\mathbf{k}-\boldsymbol{\omega}_0)+\delta(\mathbf{k}+\boldsymbol{\omega}_0)]\right\}. \tag{18}$$

В целях дальнейшего упрощения анализа в данной работе будем рассматривать случай одномерного волнения: $\mathbf{q}(\mathbf{r}) = q(r)$; здесь подразумевается, что $q \equiv q_x$ и $r \equiv x$. В этом случае характеристическая функция уклонов имеет вид

$$\theta_2(-a\mathbf{k},2\mathbf{p},\boldsymbol{\rho}) = \theta_2(-ak_x,2p_x,\rho_x). \tag{19}$$

Подставляя соотношения (18) и (19) в формулу (17), получим

$$\overline{P}_{\rm i}(r) = \overline{P}_{\rm i}^{(0)}(r) + \frac{K_0}{2}\,\overline{P}_{\rm i}^{(1)}(r) + \frac{K_0}{2}\,\overline{P}_{\rm i}^{(2)}(r), \tag{20}$$

где

$$\overline{P}_{\rm i}^{(1)}(r) = \frac{B_0 B_{\rm s}\Sigma_1\Sigma_2}{m^2\,(2\pi)^3} \int\limits_{-\infty}^{+\infty} \ldots \int\limits_{-\infty}^{+\infty} F_{\rm s}(-\mathbf{p}) F_1(-\omega H - \omega_0 L) \times$$
$$\times F_2(p_x - \omega H, p_y)\theta_2(-a\omega_0, 2p_x, \rho) \exp[i\,(\omega\rho + \omega_0 r)]\,{\rm d}\omega\,{\rm d}\mathbf{p}\,{\rm d}\rho; \tag{21}$$

слагаемое $\overline{P}_{\rm i}^{(0)}$ следует из (21) при $\omega_0 = 0$, $\overline{P}_{\rm i}^{(2)}$ следует из (21) при $\omega_0 = -\omega_0$; переменные интегрирования: $\omega \equiv \omega_x$, $\rho \equiv \rho_x$.

Дальнейший анализ требует задания конкретного вида входящих в выражение (21) функций. Зададим оптические передаточные функции излучателя и приёмников в виде гауссовых функций:

$$F_{\rm s}(\mathbf{p}) = \Delta_{\rm s} \exp\!\left(-\frac{\Delta_{\rm s}}{4\pi}\,p^2\right), \qquad F_1(\mathbf{p}) = \Delta_1 \exp\!\left(-\frac{\Delta_1}{4\pi}\,p^2\right), \qquad F_2(\mathbf{p}) = \Delta_2 \exp\!\left(-\frac{\Delta_2}{4\pi}\,p^2\right), \tag{22}$$

где $\Delta_{\rm s}$ и Δ_1, Δ_2 — телесные углы излучения и приёма.

Двухточечная характеристическая функция распределения вероятностей уклонов одномерного морского волнения имеет следующий вид [1]:

$$\theta_2(p_1, p_2, \rho) = \exp\!\left\{-\frac{\sigma_q^2}{2}\left[p_1^2 + p_2^2 + 2p_1p_2R_q(\rho)\right]\right\}, \tag{23}$$

где $R_q(\rho)$ — нормированная корреляционная функция уклонов взволнованной морской поверхности.

Подставим соотношения (22), (23) в выражение (21), после чего, проведя последовательное интегрирование по переменным p_y, p_x, получим

$$\bar{P}_{\rm i}^{(1)}(r) = \frac{C_1}{2\pi} \exp(i\omega_0 r) \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} I_p F_1(\omega H + \omega_0 L) \exp(i\omega\rho)\,{\rm d}\omega\,{\rm d}\rho,$$

где

$$I_p = \frac{1}{\sqrt{\Delta_0}} \exp\!\left\{-\beta\omega_0^2 - \frac{S_2}{4\pi}\,\omega^2 + \frac{[\omega\Delta_2 H + f(\rho)\omega_0]^2}{4\pi\Delta_0}\right\}, \qquad C_1 = B_0 B_{\rm s}\Sigma_1\Sigma_2\Delta_2\Delta_{\rm s}/(m^2\,\sqrt{\Delta_{\rm s2}}),$$
$$f(\rho) = \Delta_q a R_q(\rho)/2, \qquad \beta = a^2\sigma_q^2/2, \qquad S_2 = \Delta_2 H^2,$$
$$\Delta_q = 8\pi\sigma_q^2, \qquad \Delta_0 = \Delta_{\rm s2} + \Delta_q, \qquad \Delta_{\rm s2} = \Delta_{\rm s} + \Delta_2.$$

Интегрируя это выражение по переменной ω, получим

$$\overline{P}_{\rm i}^{(1)}(r) = C_2 \exp(i\omega_0 r) \exp[-\Delta_1 L^2\omega_0^2/(4\pi)] \exp(-\beta\omega_0^2)\Phi(\omega_0), \tag{24}$$

где

$$\Phi(\omega_0) = \frac{1}{\sqrt{S_0}} \int\limits_{-\infty}^{+\infty} \exp\!\left(-\frac{\pi\rho^2}{S_0}\right) \exp\!\left[-i\,\omega_0\rho\,\frac{S_1}{S_0}\left(1 - f(\rho)\,\frac{\Delta_2 H}{\Delta_0 S_1}\right)\right] \times$$
$$\times \exp\!\left\{\frac{\omega_0^2}{4\pi}\left[\frac{f^2(\rho)}{\Delta_0}\left(1 + \frac{S_2\Delta_2}{S_0\Delta_0}\right) + \frac{S_1^2}{S_0}\left(1 - 2f(\rho)\,\frac{\Delta_2 H}{\Delta_0 S_1}\right)\right]\right\} {\rm d}\rho,$$

$$C_2 = C_1\Delta_1/\sqrt{\Delta_0}\,, \qquad S_1 = \Delta_1 HL, \qquad S_0 = (\Delta_1 + \Delta_2')\,H^2, \qquad \Delta_2' = \Delta_2\,(1 - \Delta_2/\Delta_0).$$

Дальнейшее упрощение формулы (24) связано с некоторыми ограничениями, налагаемыми на геометрию освещения и наблюдения объекта и морской поверхности. Положим, что выполняются условия

$$\Delta_1 \approx \Delta_2, \qquad \Delta_q \gg \Delta_2 + \Delta_{\mathrm{s}}, \qquad H \gg h/4.$$

Заметим, что эти условия не представляют собой ничего из ряда вон выходящего: они, как правило, выполняются в большинстве реальных ситуаций при мониторинге морских акваторий.

Опираясь на эти условия, нетрудно показать, что

$$\Delta_2 S_2/(\Delta_0 S_0) \ll 1, \qquad 2f(\rho)\Delta_2 H/(\Delta_0 S_1) \ll 1.$$

С учётом последних неравенств выражение (24) принимает более простой вид:

$$\overline{P}_{\mathrm{i}}^{(1)}(r) = P_0^2 \exp(i\omega_0 r) \exp[-\Delta_1 L^2 \omega_0^2/(4\pi)]\Psi(\omega_0), \tag{25}$$

где

$$\Psi(\omega_0) = \exp\!\left(\frac{\omega_0^2}{4\pi}\,\frac{S_1^2}{S_0}\right) \frac{1}{\sqrt{S_0}} \int\limits_{-\infty}^{+\infty} \exp\!\left(-\,\frac{\pi\rho^2}{S_0}\right) \exp\!\left(-i\omega_0\rho\,\frac{S_1}{S_0}\right) \exp\!\left\{-\beta\omega_0^2\,[1 - \gamma R_q^2(\rho)]\right\} \mathrm{d}\rho, \tag{26}$$

$P_0^2 = B_0 B_{\mathrm{s}} \Sigma_1 \Sigma_2 \Delta_1 \Delta_2 \Delta_{\mathrm{s}}/(m^2\,\sqrt{\Delta_0 \Delta_{\mathrm{s}2}})$, $\gamma = \Delta_q/\Delta_0 \approx 1$; размерность P_0^2 — квадрат мощности.

Подставляя соотношения (25), (26) в (20), получим окончательное выражение для яркости усреднённого (накопленного) изображения подводного объекта, сформированного посредством бликовой системы наблюдения:

$$\overline{P}_{\mathrm{i}}(r) = P_0\,[1 + K_0 K_{\mathrm{i}} \cos(\omega_0 r)], \tag{27}$$

где $K_{\mathrm{i}} = \exp[-\Delta_1 L^2 \omega_0^2/(4\pi)]\Psi(\omega_0)$.

Из выражения (27) следует, что усреднённое изображение подводного объекта имеет то же гармоническое распределение яркости, что и распределение светимости объекта, однако контраст этого изображения отличается от контраста структуры яркости на объекте. Изменение контраста определяется параметром K_{i}, который довольно сложно зависит от характеристик волнения и параметров системы наблюдения. Заметим, что первый сомножитель в выражении для K_{i} определяет контраст изображения, полученного через плоскую границу раздела воздух—вода. Второй сомножитель — $\Psi(\boldsymbol{\omega}_0)$ — можно определить как частотно-контрастную характеристику бликовой системы наблюдения. Если в (27) положить $\Delta_{\mathrm{s}} \approx \Delta_1 \approx \Delta_2 \approx 0$, то, используя фильтрующие свойства функции $\exp(-\pi\rho^2/S_0)/\sqrt{S_0} \sim \delta(\rho)$ в формуле для $\Psi(\omega_0)$, можно получить выражение для контраста изображения подводного объекта, полученного с помощью «идеальной» бликовой системы наблюдения: $K_{\mathrm{i}} = \exp[-\Delta_1 L^2 \omega_0^2/(4\pi)]$. При этих условиях видение подводного объекта осуществляется как бы через плоскую границу раздела.

4. АНАЛИТИЧЕСКОЕ ИССЛЕДОВАНИЕ ЧАСТОТНО-КОНТРАСТНОЙ ХАРАКТЕРИСТИКИ

Для пояснения физической сути полученных формул рассмотрим две предельные ситуации. Во-первых, обратимся к случаю мелкошероховатой (в масштабе элементов разрешения приёмников) границы раздела. При этом в формуле (26) можно положить $R_q(\rho) \equiv 0$. В результате контраст изображения, полученного с помощью бликовой системы наблюдения, оказывается тождественным контрасту обычной системы видения (ср. с (7)):

$$K_{\rm i} = \exp[-\Delta_1 L^2 \omega_0^2/(4\pi)] \exp(-\beta\omega_0^2). \tag{28}$$

Здесь первый сомножитель описывает оптическую передаточную функцию приёмника изображения, второй сомножитель определяет частотно-контрастную характеристику взволнованной морской поверхности [2]. Нетрудно видеть, что ожидаемого улучшения качества изображения в данном случае нет: бликовая система наблюдения не даёт никакого выигрыша по сравнению с обычными системами.

Обратимся теперь к случаю крупномасштабных (в масштабе элементов разрешения приёмников) неровностей на границе раздела. При этом в формуле для $\Psi(\omega_0)$ можно положить $R_q(\rho) \equiv 1$. В результате контраст оказывается весьма близким к величине, определяемой оптической передаточной функцией приёмника изображения (при этом наблюдение объекта осуществляется как бы через плоскую границу раздела; ср. с (8)):

$$K_{\rm i} = \exp[-\Delta_1 L^2 \omega_0^2/(4\pi)] \exp[-\beta\omega_0^2\,(1-\gamma)]. \tag{29}$$

Нетрудно убедиться в том, что в данной ситуации использование «опорной» информации о бликах приводит к весьма ощутимому повышению качества изображения подводного объекта.

Из анализа этих предельных ситуаций следует, что в общем случае при использовании информации о бликах, если и не удастся достигнуть «идеального» результата (29), то всё же можно ожидать некоторого улучшения качества изображения по сравнению с «плохим» результатом (28).

Рассмотрим поведение контраста накопленного изображения на низких и высоких пространственных частотах.

На низких частотах, условием которых является соотношение $\beta\omega_0^2 \ll 1$, выражение для контраста имеет следующий вид:

$$K_{\rm i} = \exp[-(\Delta_1 L^2 - S_1^2/S_0)\,\omega_0^2/(4\pi)] \exp(-\beta\omega_0^2)\Psi(\omega_0),$$

где

$$\Psi(\omega_0) = \frac{1}{\sqrt{S_0}} \int\limits_{-\infty}^{+\infty} \exp\left(-\frac{\pi\rho^2}{S_0}\right) \left[1 + \beta\gamma\omega_0^2 R_q^2(\rho)\right] \exp\left(-i\omega_0\rho\,\frac{S_1}{S_0}\right) {\rm d}\rho.$$

Аппроксимируя корреляционную функцию уклонов гауссовой функцией: $R_q(\rho) = \exp(-\pi\rho^2/S_q)$, где S_q — характерная площадь корреляции уклонов взволнованной морской поверхности, получим выражение для контраста в виде

$$K_{\rm i} = \exp[-\Delta_1 L^2 \omega_0^2/(4\pi)] \exp(-\beta\omega_0^2) \left[1 + \beta\gamma\omega_0^2 \sqrt{\frac{S_q}{S_q + 2S_0}} \exp\left(\frac{\omega_0^2}{4\pi}\,\frac{2S_1^2}{S_q + 2S_0}\right)\right]. \tag{30}$$

Из выражения (30), в частности, следует, что в случае высокого разрешения системы наблюдения, что соответствует условию $2S_0 \ll S_q$, контраст изображения оказывается близким к «идеальному» значению:

$$K_{\rm i} \approx \exp[-\Delta_1 L^2 \omega_0^2/(4\pi)].$$

В случае низкого разрешения ($2S_0 \gg S_q$) контраст изображения, полученного бликовой системой наблюдения, оказывается равным контрасту изображения, полученного посредством обычной системы видения:

$$K_{\rm i} \approx \exp[-\Delta_1 L^2 \omega_0^2/(4\pi)] \exp(-\beta\omega_0^2).$$

На высоких пространственных частотах, условием которых является соотношение $\beta\omega_0^2 \gg 1$, корреляционную функцию уклонов взволнованной морской поверхности можно представить в виде укороченного ряда Тейлора: $R_q^2(\rho) \approx 1 - 2\pi\rho^2/S_q$. При этом интеграл в выражении (26) берётся аналитически, и выражение для контраста записывается следующим образом:

$$K_{\rm i} = \exp[-\Delta_1 L^2\omega_0^2/(4\pi)]\exp[-\beta\omega_0^2\,(1-\gamma)]\,\sqrt{\frac{\tilde S_q}{S_0+\tilde S_q}}\,\exp\left(\frac{\omega_0^2}{4\pi}\,\frac{S_1^2}{S_0+\tilde S_q}\right),$$

где $\tilde S_q = S_q/(2\beta\gamma\omega_0^2)$.

Поскольку обычно выполняется условие $S_0 \gg \tilde S_q$, это выражение может быть записано в более простом виде:

$$K_{\rm i} = \frac{C}{\omega_0}\exp[-\Delta_{\min} L^2\omega_0^2/(4\pi)]\exp[-\beta\omega_0^2\,(1-\gamma)], \tag{31}$$

где

$$C = \sqrt{\frac{S_q}{2\beta\gamma S_0}}\,, \qquad \Delta_{\min} = \frac{\Delta_1\Delta_2'}{\Delta_1+\Delta_2'}\,.$$

Рассмотрим отношение контрастов изображений, полученных с помощью бликовой и обычной систем наблюдения, в качестве оценки выигрыша от применения последней. Разделив выражение (31) на (28), получаем оценку:

$$Q = \frac{C}{\omega_0}\exp(\beta\gamma\omega_0^2)\exp\left(\Delta_1 L^2\,\frac{\Delta_1}{\Delta_1+\Delta_2'}\,\frac{\omega_0^2}{4\pi}\right).$$

Отсюда следует, что контраст изображения, полученного с помощью бликовой системы наблюдения, на высоких частотах может значительно превышать контраст изображения, полученного посредством обычной системы наблюдения.

5. РЕЗУЛЬТАТЫ ЧИСЛЕННОГО АНАЛИЗА ЧАСТОТНО-КОНТРАСТНОЙ ХАРАКТЕРИСТИКИ

Выведенные в предыдущих разделах соотношения позволяют провести анализ зависимости контраста изображения подводного объекта, полученного через морскую поверхность с помощью бликовой системы наблюдения, от физических факторов задачи. Ниже излагаются результаты численных расчётов контраста изображения для следующих параметров: высота размещения источника и приёмников $H_{\rm s} = H = 100$ м, угловая ширина тела яркости освещающего поля над поверхностью воды $\varphi_{\rm s} = 0{,}1^\circ$ ($\Delta_{\rm s} = \pi\varphi_{\rm s}^2/4$), скорость надводного ветра $v = 4$ м/с; эти параметры в расчётах являются фиксированными, остальные варьируются в зависимости от условий задачи. При проведении расчётов использовалась гауссова модель корреляционной функции уклонов взволнованной морской поверхности, имеющая характерный радиус корреляции

$$\rho_q = \sqrt{2\sigma_q^2/\sigma_p^2}\,,$$

где $\sigma_q^2 = 2\cdot 10^{-3}\,(3+1{,}92v[\text{м/с}])$ [7], $\sigma_p^2[\text{м}^{-2}] = 8\,(-4{,}13+1{,}23v[\text{м/с}])^2/3$ [8], σ_p^2 — дисперсия кривизны взволнованной поверхности моря.

Перейдём к обсуждению полученных результатов. На рис. 3 и 4 приведены зависимости контраста изображения подводного тестового объекта от частоты синусоидальной структуры (частотно-контрастная характеристика) для двух глубин расположения объекта. Эти зависимости рассчитаны для случая одинаковых диаграмм направленности приёмников обычной и бликовой систем наблюдения с угловой шириной, соответствующей разрешению человеческого глаза: $\varphi_1 = \varphi_2 = 1/30^\circ$ ($\Delta_1 = \pi\varphi_1^2/4$, $\Delta_2 = \pi\varphi_2^2/4$).

Из приведённых на рис. 3 и 4 зависимостей видно, что частотно-контрастная характеристика бликовой системы наблюдения занимает промежуточное положение между частотно-контрастной характеристикой обычной системы наблюдения и «идеальной» частотно-контрастной характеристикой, соответствующей ситуации наблюдения через плоскую границу раздела. С увеличением глубины расположения объекта частотно-контрастная характеристика бликовой системы наблюдения становится уже. Однако то, что контраст изображения, полученного с помощью бликовой системы наблюдения, всегда (т. е. на всех частотах и на любой глубине расположения объекта) больше контраста изображения, полученного посредством обычной системы наблюдения, с очевидностью свидетельствует в поддержку идеи использования информации о бликах для коррекции геометрических искажений в изображении подводного объекта.

На рис. 5 приведены аналогичные частотно-контрастные характеристики, рассчитанные для глубины расположения объекта $h = 20$ м и скорости надводного ветра $v = 6$ м/с (остальные параметры соответствуют рис. 4).

Из сопоставления зависимостей рис. 4 и 5 следует, что увеличение скорости надводного ветра приводит к снижению эффективности работы бликовой системы наблюдения. Это можно понять, если сравнить характерные размеры пятна разрешения системы наблюдения на поверхности воды (5,8 см) и радиуса корреляции уклонов (16 см при скорости ветра 4 м/с и 4,5 см при скорости ветра 6 м/с). С увеличением скорости ветра в пятне разрешения укладывается всё больше неоднородностей границы раздела. Это означает, что отдельные блики на поверхности бликовой системы наблюдения воспринимаются как один блик, и коррекция изображения осуществляется с ошибкой. Тем не менее и в этих случаях качество изображения, полученного с помощью бликовой системы наблюдения, оказывается значительно лучше качества изображения, полученного посредством обычной системы наблюдения.

На рис. 6 приведены зависимости контраста изображения подводного объекта, полученные для узкой диаграммы направленности приёмника бликового изображения: $\varphi_2 = 1/150^\circ$ (остальные

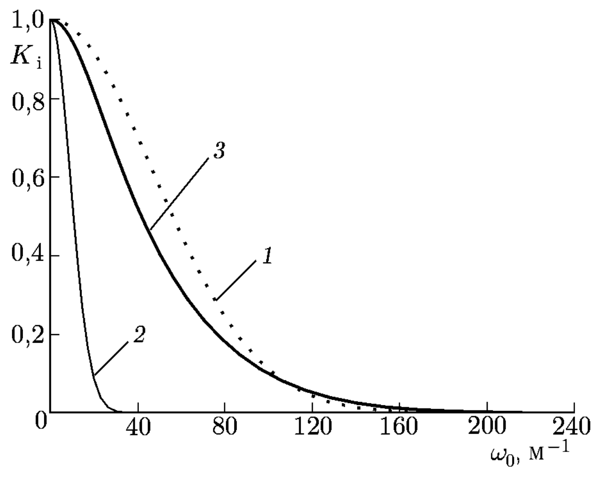

Рис. 3. Частотно-контрастная характеристика для глубины расположения объекта $h = 3$ м: *1* — видение через плоскую границу раздела, *2* и *3* — видение через взволнованную морскую поверхность с помощью обычной и бликовой систем наблюдения соответственно

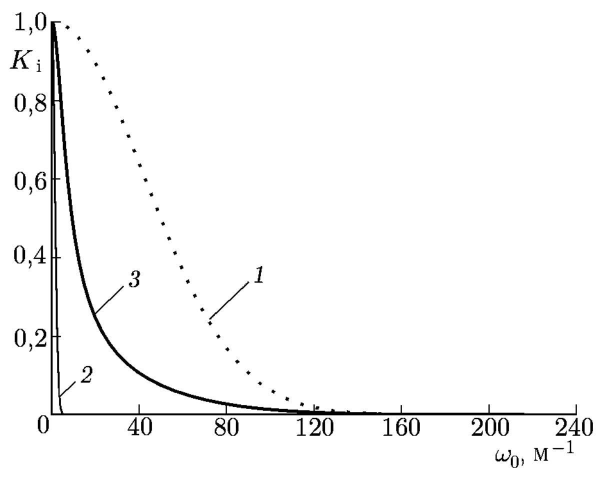

Рис. 4. Частотно-контрастная характеристика для глубины расположения объекта $h = 20$ м; Обозначения кривых те же, что на рис. 3

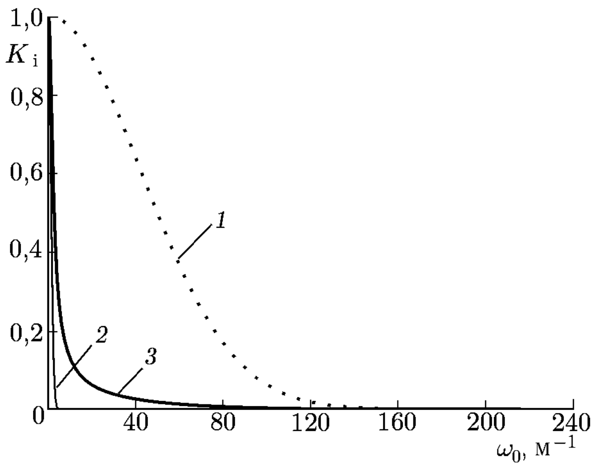

Рис. 5. Частотно-контрастная характеристика для скорости ветра $v = 6$ м/с: *1* — видение через плоскую границу раздела, *2* и *3* — видение через взволнованную морскую поверхность с помощью обычной и бликовой систем наблюдения соответственно

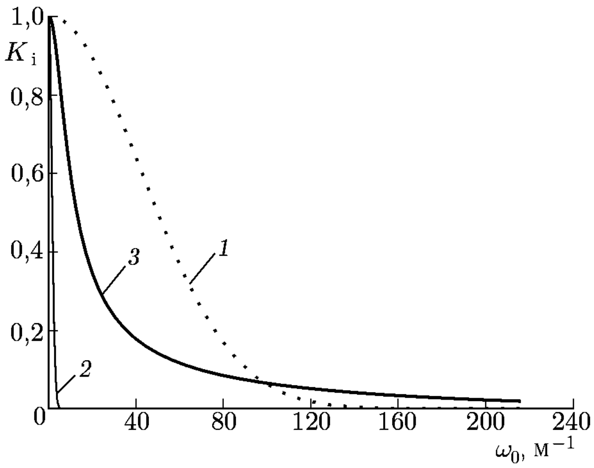

Рис. 6. Частотно-контрастная характеристика в системе наблюдения с узкой диаграммой направленности приёмника излучения, отражённого от морской поверхности

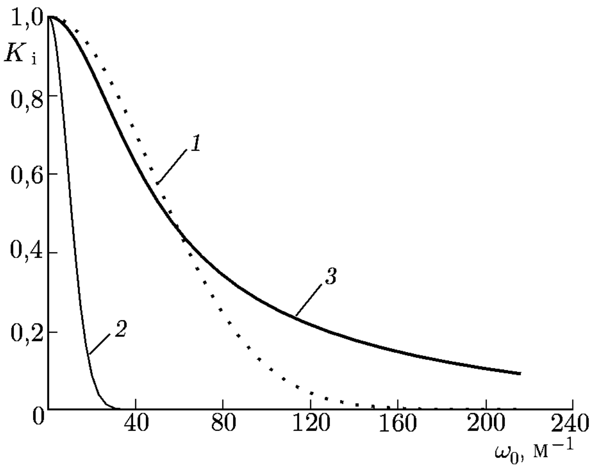

Рис. 7. Частотно-контрастная характеристика для глубины расположения объекта $h = 3$ м

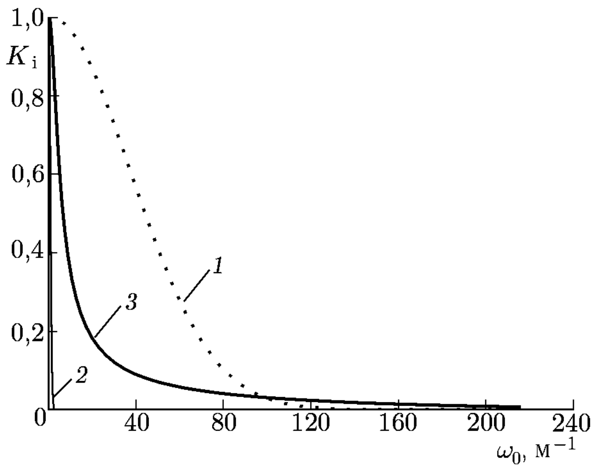

Рис. 8. Частотно-контрастная характеристика для глубины расположения объекта $h = 40$ м

параметры: $\varphi_1 = 1/30°$, $v = 4$ м/с, $h = 20$ м). Из сравнения зависимостей рис. 4 и 6 следует, что увеличение разрешающей способности системы, формирующей изображение бликов, приводит к заметному повышению контраста, особенно на высоких пространственных частотах.

Отметим, что на высоких пространственных частотах (начиная с некоторой частоты) контраст изображения в бликовой системе наблюдения может превосходить контраст «идеального» изображения, полученного через плоскую границу раздела. Этот не совсем тривиальный эффект достаточно отчетливо иллюстрируют зависимости, приведённые на рис. 7 и 8. При проведении расчётов использовались следующие параметры: $v = 4$ м/с, $\varphi_1 = 1/30°$, $\varphi_2 = 1/150°$. Зависимости на рис. 7 соответствуют глубине расположения объекта $h = 3$ м, зависимости на рис. 8 — $h = 40$ м.

Следует подчеркнуть, что эффект улучшения качества изображения на высоких пространственных частотах в бликовой системе наблюдения по сравнению со случаем наблюдения через плоскую границу раздела хорошо заметен, как правило, лишь в системах наблюдения с достаточно большой шириной диаграммы направленности приёмника φ_1.

6. ПРОСТАЯ АНАЛИТИЧЕСКАЯ ФОРМУЛА ДЛЯ КОНТРАСТА ИЗОБРАЖЕНИЯ

Формула (26) для определения контраста накопленного фрагментарного изображения, полученного с помощью бликовой системы наблюдения, содержит однократный интеграл, который в общем случае должен браться численно. Упростим нашу задачу с целью получения аналитической формулы для контраста. Представим корреляционную функцию уклонов в виде отрезка параболической функции:

$$R_q^2(\rho) \approx \begin{cases} 1 - 2\rho^2/\rho_q^2, & |\rho| \le \rho_q; \\ 0, & |\rho| > \rho_q. \end{cases} \tag{32}$$

Подставим функцию корреляции (32) в выражение (26) и представим $\Psi(\omega_0)$ в следующем виде:

$$\Psi(\omega_0) \propto \int\limits_{-\infty}^{+\infty} B \exp(-\pi\rho^2/S_0) \exp(-i\omega_0\rho S_1/S_0)\,\mathrm{d}\rho + \int\limits_{-\rho_q}^{+\rho_q} (A - B) \exp(-\pi\rho^2/S_0) \exp(-i\omega_0\rho S_1/S_0)\,\mathrm{d}\rho,$$

где $A = \exp[-\beta\omega_0^2\,(1-\gamma)] \exp(-2\pi\beta\gamma\omega_0^2\rho^2/S_q)$, $B = \exp(-\beta\omega_0^2)$. Заменим в этом выражении прямоугольное «окно» интегрирования на гауссовое и, проведя несложные преобразования, получим следующую аналитическую формулу, определяющую контраст изображения:

$$K_\mathrm{i} = \exp[-\Delta_1 L^2\omega_0^2/(4\pi)] \exp(-\beta\omega_0^2)\,[1 + M_1(\omega_0) - M_2(\omega_0)], \tag{33}$$

где

$$M_1(\omega_0) = \exp(\beta\gamma\omega_0^2) \sqrt{\frac{\breve{S}_q}{\breve{S}_q + S_0}} \exp\left(\frac{\omega_0^2}{4\pi}\,\frac{S_1^2}{\breve{S}_q + S_0}\right),$$

$$M_2(\omega_0) = \sqrt{\frac{S_q}{S_q + S_0}} \exp\left(\frac{\omega_0^2}{4\pi}\,\frac{S_1^2}{S_q + S_0}\right), \qquad S_q = \pi\rho_q^2, \qquad \breve{S}_q = S_q/(1 + 2\beta\gamma\omega_0^2).$$

На низких частотах ($\beta\omega_0^2 \ll 1$) выражение (33), как нетрудно показать, преобразуется к виду, совпадающему с (30). На высоких частотах ($\beta\omega_0^2 \gg 1$) оно имеет тот же вид, что и выражение (31). Это позволяет считать предложенную модель «локально параболической» границы раздела пригодной для проведения оценок контраста изображений подводных объектов, полученных с помощью бликовой системы наблюдения.

Численные расчёты частотно-контрастных характеристик бликовой системы наблюдения, проведённые по формуле (33), дали результаты, близкие к полученным на основе строгих соотношений (26), (27). Это говорит об адекватности предложенной аналитической формулы, несмотря на то, что при её выводе использованы достаточно грубые приближения.

7. ЗАКЛЮЧЕНИЕ

Разработанная в данной работе модель формирования изображения подводного объекта, использующая информацию о бликовых участках взволнованной морской поверхности, и полученные на её основе результаты позволяют рекомендовать этот способ наблюдения для использования на практике. Часть материала, изложенного здесь, впервые докладывалась на III Межреспубликанском симпозиуме «Оптика атмосферы и океана» [9].

Заметим, что при практической реализации бликовой системы наблюдения можно применять различные методы разделения оптических сигналов от объекта и поверхности моря. Например, кроме временнóго разделения сигналов, реализуемого в системах с импульсной подсветкой и временны́м стробированием приёмника изображения, можно применить разделение по оптической длине волны, при котором сигнал от объекта принимается в видимом участке спектра излучения источника, а сигнал от поверхности — в инфракрасной области спектра.

Рассмотренные в данной работе вопросы относятся к одномерной модели поверхностного морского волнения. Обобщение полученных результатов на случай двумерной морской поверхности является задачей дальнейших исследований.

Важными вопросами, не исследованными в данной работе, являются также практически важные вопросы, связанные с процедурой накопления фрагментированных изображений. К ним, прежде всего, можно отнести вопрос о необходимом количестве реализаций изображения, а также вопрос о времени, необходимом для накопления этого количества реализаций. Решение этих вопросов требует изучения временны́х интервалов корреляции бликового изображения в зависимости от балльности волнения на морской поверхности.

Автор выражает признательность Л. С. Долину и А. Г. Лучинину за плодотворные дискуссии по затронутой проблеме.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект № 02–05–64975).

СПИСОК ЛИТЕРАТУРЫ

1. Вебер В. Л. // Изв. вузов. Радиофизика. 1979. Т. 22, № 8. С. 989.
2. Мулламаа Ю.-А. Р. // Изв. АН. ФАО. 1975. Т. 11, № 2. С. 199.
3. Браво-Животовский Д. М., Долин Л. С., Левин И. М. и др. // Вопросы радиоэлектроники. Серия ТТ. 1972. № 3. С. 35.
4. Вебер В. Л. // Оптика моря и атмосферы. Л.: ГОИ, 1984. С. 199.
5. Вебер В. Л. // Изв. АН. ФАО. 1988. Т. 24, № 6. С. 647.
6. Вебер В. Л. // Изв. АН. ФАО. 1997. Т. 33, № 2. С. 274.
7. Cox C. S., Munk W. H. // J. Mar. Res. 1954. V. 13. P. 198.
8. Бурцев Е. Г., Пелевин В. Н. // Световые поля в океане. М.: ИО АН СССР, 1979. С. 231.
9. Вебер В. Л. // Оптика атмосферы и океана. Томск: ИОА СО РАН, 1996. С. 53.

Поступила в редакцию 6 декабря 2003 г.; принята в печать 2 августа 2004 г.

OBSERVATION OF UNDERWATER OBJECTS THROUGH GLITTER PARTS OF THE SEA SURFACE

V. L. Weber

We consider the possibility of decreasing the negative influence of roughness on the visibility of underwater objects. The technique is based on the use of information about glitters on the sea surface (this information is needed for determination of undistorted parts of the underwater object image). The effect of the angular resolution of the imaging system on the underwater object image quality is studied.

УДК 539.21

ЭЛЕКТРОДИНАМИЧЕСКИЕ СВОЙСТВА ВОЛНОВОДА СО СЛОИСТО-ПЕРИОДИЧЕСКИМИ СТЕНКАМИ

А. А. Булгаков[1], *О. В. Костылёва*[2], *А. В. Мериуц*[2]

[1] Институт радиофизики и электроники им. А. Я. Усикова НАН Украины
[2] Национальный технический университет «Харьковский политехнический институт»
г. Харьков, Украина

В данной статье исследуется волновод со слоисто-периодическими стенками при различных соотношениях между диэлектрической проницаемостью среднего (волноведущего) слоя и слоёв сверхрешёток. Рассмотрено волноводное распространение в таком волноводе, предсказано появление поверхностных волн в волноведущем слое, а также рассмотрена ситуация, когда вся энергия переносится в основном в периодических стенках.

ВВЕДЕНИЕ

В последнее время устройствам с периодическими структурами (сверхрешётками) уделяется много внимания благодаря особенностям свойств сверхрешёток. По аналогии с разрешёнными и запрещёнными зонами для электрона, движущегося в периодическом потенциале, в случае периодических структур образуются зоны прозрачности и полного внутреннего отражения для электромагнитных волн. Это связано с тем, что волновые процессы в каждом из слоёв сверхрешётки определяются параметрами слоя и условиями непрерывности тангенциальных составляющих полей на границах между слоями. В результате интерференции волн отдельных слоёв и формируется зонная структура спектра. В сверхрешётках могут распространяться лишь волны, лежащие в пределах жёстко ограниченных полос пропускания.

Одним из возможных применений сверхрешёток являются волноводы. Волноводы, в которых обычно используемая подложка заменена периодической структурой, можно разделить на два типа. Для первого типа показатель преломления среднего (волноведущего) слоя больше показателя преломления стенок. В этом случае возникает стандартная волноводная мода, которая является результатом отличия показателей преломления и подобна моде в стандартном волноводе. Для второго типа показатель преломления среднего слоя меньше, чем показатель преломления стенок, поэтому волна не может распространяться в среднем слое. Однако, если распространяющаяся волна находится в запрещённой зоне, то волноводная мода может существовать, т. к. на границах среднего слоя волна отражается согласно брэгговскому закону. Этот случай (когда волновод образован воздушным зазором между двумя сверхрешётками) рассматривался в [1], где на примере такого волновода с одномерной периодической структурой описаны основные свойства брэгговских волн. Кроме того, авторы описали распространение электромагнитного поля в цилиндрическом волноводе, стенки которого имеют коаксиальную периодическую структуру.

В работе [2] получено дисперсионное соотношение для брэгговского волновода (волновода со слоисто-периодическими стенками) и исследовано распространение блоховских волн в таком волноводе. На основе этого исследования было показано, что, в отличие от стандартного диэлектрического волновода, ограниченное распространение со сколь угодно малыми потерями возможно, даже когда показатель преломления волноведущего слоя меньше показателя преломления периодических слоёв. Такое свойство волновода со слоисто-периодическими стенками позволяет применять его в рентгеновском диапазоне. Трудность создания волновода с металлическими стенками для этого диапазона состоит в том, что диэлектрическая проницаемость металлов в области рентгеновского излучения близка к единице, поэтому металлические стенки волновода становятся прозрачными для электромагнитных волн. Использование периодических структур позволяет обойти эту трудность, поскольку в сверхрешётках отражения связаны с фазовыми соотношениями между полями в слоях структуры. В связи с этим коэффициент отражения может быть близок к единице, а свойства волновода близки к свойствам металлических систем на низких частотах.

В последнее время, кроме теоретических работ по исследованию волновода со слоисто-периодическими стенками, появились и экспериментальные работы. Так, в статье [3] представлено экспериментальное изучение генерации второй гармоники в коротковолновой части миллиметрового диапазона. Структура, исследуемая в [3], состояла из двух идентичных сверхрешёток, разделённых воздушным зазором; на фронтальную поверхность одной из сверхрешёток был нанесён нелинейный слой.

В статьях [1, 2] авторы рассматривали только волноводное распространение. В данной статье будет проведено более полное исследование брэгговского волновода (предсказано появление поверхностных волн, также рассмотрена ситуация, когда вся энергия переносится в основном в периодических стенках). Кроме того, будет рассмотрен случай, когда показатель преломления среднего слоя больше показателя преломления стенок.

1. ПОСТАНОВКА ЗАДАЧИ. ДИСПЕРСИОННОЕ СООТНОШЕНИЕ

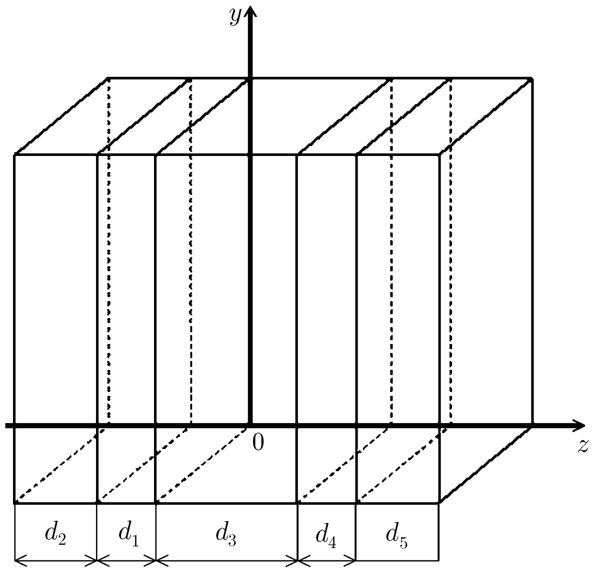

Рис. 1. Схема волновода со слоисто-периодическими стенками

Геометрия волновода со слоисто-периодическими стенками представлена на рис. 1. Левая стенка волновода представляет собой плоскослоистую периодическую структуру, период которой составлен из двух слоёв с различной толщиной (d_1 и d_2) и диэлектрической проницаемостью (ε_1 и ε_2). Правая стенка волновода — плоскослоистая периодическая структура, которая характеризуется толщинами d_4 и d_5 и диэлектрическими проницаемостями ε_4 и ε_5. Волноведущий слой имеет толщину d_3 и диэлектрическую проницаемость ε_3. Ось z выбираем перпендикулярно границам слоёв, а начало координат — на левой стенке волноведущего слоя. Вдоль оси y слои предполагаются однородными, поэтому зависимостью от координаты y можно пренебречь, положив $\partial/\partial y = 0$.

Распространение электромагнитных волн в каждом из слоёв описывается уравнениями Максвелла:

$$\operatorname{rot}\mathbf{E} = -\frac{1}{c}\,\frac{\partial \mathbf{H}}{\partial t}\,, \qquad \operatorname{rot}\mathbf{H} = \frac{\varepsilon}{c}\,\frac{\partial \mathbf{E}}{\partial t}\,. \tag{1}$$

Рассматривается плоская волна вида $\exp[-i\,(\omega t - k_x x - k_z z)]$. В результате уравнения Максвелла распадаются на две системы для так называемых TE- и TM-волн. Для определённости будем рассматривать TM-волны, у которых магнитное поле имеет только компоненту, нормальную к волновому вектору и параллельную границам слоёв. Для другой поляризации методика рассмотрения аналогична. Для ТМ-волны отличными от нуля оказываются следующие комплексные амплитуды электромагнитного поля: E_x, H_y, E_z. Для получения явного вида этих компонент воспользуемся методом матрицы преобразования [4, 5], которая связывает комплексные амплитуды в начале координат с их значениями в конце периода структуры:

$$\begin{pmatrix} H_y(0) \\ E_x(0) \end{pmatrix} = \mathrm{m}(d) \begin{pmatrix} H_y(d) \\ E_x(d) \end{pmatrix}.$$

Решение уравнений Максвелла будем искать в виде

$$\begin{gathered} H_y(z) = A_1 \exp(ik_z z) + A_2 \exp(-ik_z z), \\ E_x(z) = -\frac{ic}{\omega}\,\frac{1}{\varepsilon}\,\frac{\partial H_y(z)}{\partial z} = \frac{c}{\omega}\,\frac{k_z}{\varepsilon}\,[A_1 \exp(ik_z z) - A_2 \exp(-ik_z z)], \end{gathered} \tag{2}$$

где

$$k_z = \pm\sqrt{\frac{\omega^2}{c^2}\,\varepsilon - k_x^2}\,, \tag{3}$$

k_z — поперечное волновое число.

На плоскостях границ раздела должны выполняться граничные условия, состоящие в непрерывности тангенциальных компонент комплексных амплитуд электромагнитного поля. Кроме того, решение должно удовлетворять условию периодичности. В случае безграничной периодической структуры поля́ на границах периода могут отличаться только на фазовый множитель (теорема Флоке):

$$H_y(0) = H_y(d)\exp(i\bar{k}d), \qquad E_x(0) = E_x(d)\exp(i\bar{k}d), \tag{4}$$

где d — период структуры, $\bar{k}$ — блоховское волновое число.

Для ограниченной структуры соотношение между компонентами поля H_y и E_x должно быть таким же, как для безграничной структуры [6]:

$$H_y(z) = E_x(z)\,\frac{m_{12}(z)\exp(i\bar{k}d)}{1 - m_{11}(z)\exp(i\bar{k}d)}\,. \tag{5}$$

Здесь $m_{ik}(z)$ — компоненты матрицы преобразования.

Записав матрицы преобразования для двух сверхрешёток и волноведущего слоя, перемножив их в соответствующем порядке с учётом граничных условий и связи H_y и E_x (соотношение (5)), получим систему уравнений, имеющую определитель второго порядка. Приравняв определитель системы нулю, а затем раскрыв его, после некоторых преобразований получим дисперсионное соотношение для волновода со слоисто-периодическими стенками:

$$g_{11}\,\frac{b_{12}\exp(i\bar{k}_b d_b)}{1 - b_{11}\exp(i\bar{k}_b d_b)} + g_{12} - \frac{a_{12}\exp(i\bar{k}_a d_a)}{1 - a_{11}\exp(i\bar{k}_a d_a)}\left[g_{21}\,\frac{b_{12}\exp(i\bar{k}_b d_b)}{1 - b_{11}\exp(i\bar{k}_b d_b)} + g_{22}\right] = 0. \tag{6}$$

Здесь $\hat{\mathbf{a}}$ — матрица преобразования для одного периода структуры левой стенки волновода. Элементы этой матрицы, входящие в (6), равны

$$a_{11} = \cos(k_{z1}d_1)\cos(k_{z2}d_2) - \frac{\varepsilon_1}{\varepsilon_2}\,\frac{k_{z2}}{k_{z1}}\,\sin(k_{z1}d_1)\sin(k_{z2}d_2),$$

$$a_{12} = \frac{i\omega}{c}\,\frac{\varepsilon_1}{k_{z1}}\,\sin(k_{z1}d_1)\cos(k_{z2}d_2) + \frac{i\omega}{c}\,\frac{\varepsilon_2}{k_{z2}}\,\cos(k_{z1}d_1)\sin(k_{z2}d_2),$$

$$a_{22} = \cos(k_{z1}d_1)\cos(k_{z2}d_2) - \frac{\varepsilon_2}{\varepsilon_1}\,\frac{k_{z1}}{k_{z2}}\,\sin(k_{z1}d_1)\sin(k_{z2}d_2); \tag{7}$$

матрица $\hat{\mathbf{b}}$ отличается от матрицы $\hat{\mathbf{a}}$ заменой индексов 1 и 2 в правой части выражений (7) на 4 и 5 соответственно; $\hat{\mathbf{g}}$ — матрица преобразования для волноведущего слоя:

$$\hat{\mathbf{g}} = \begin{pmatrix} \cos(k_{z3}d_3) & -\dfrac{i\omega}{c}\,\dfrac{\varepsilon_3}{k_{z3}}\,\sin(k_{z3}d_3) \\ -\dfrac{ic}{\omega}\,\dfrac{k_{z3}}{\varepsilon_3}\,\sin(k_{z3}d_3) & \cos(k_{z3}d_3) \end{pmatrix}. \tag{8}$$

Множители $\exp(i\bar{k}_a d_a)$ и $\exp(i\bar{k}_b d_b)$ в формуле (6) описывают периодичность полей в слоистых структурах вдоль оси z и могут быть найдены из дисперсионного соотношения для безграничной структуры:

$$\begin{aligned} \cos(\bar{k}_a d_a) &= F_a, \qquad F_a = (a_{11} + a_{22})/2\,, \\ \cos(\bar{k}_b d_b) &= F_b, \qquad F_b = (b_{11} + b_{22})/2\,, \end{aligned} \tag{9}$$

где d_a и d_b — периоды левой и правой решёток соответственно. В случае ограниченной среды (9) представляет собой характеристическое уравнение.

2. ЧИСЛЕННОЕ ИССЛЕДОВАНИЕ ДИСПЕРСИОННОГО СООТНОШЕНИЯ

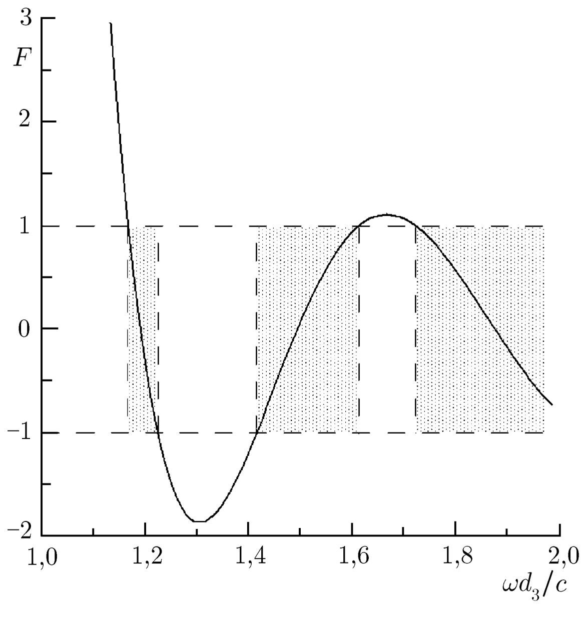

Рис. 2. Графическое определение разрешённых и запрещённых зон

Дисперсионное соотношение (6) анализировалось численно. Параметры сверхрешёток подбирались таким образом, чтобы зоны непропускания оказались достаточно широкими. Для этого была построена зависимость правой части характеристического уравнения (9) (в расчётах было принято, что периодические решётки, ограничивающие волноведущий слой, одинаковы: $F = F_a = F_b$) в зависимости от частоты при фиксированном значении продольного волнового числа (рис. 2). Области, в которых кривая проходит между значениями 1 и -1, являются зонами пропускания сверхрешётки. Эти области на рис. 2 заштрихованы. Если правая часть (9) становится по модулю больше единицы, то имеем область непропускания. Из рис. 2 следует, что с ростом частоты ширина разрешённых зон увеличивается, а запрещённых — уменьшается. Чем больше значение $|F|$, тем больше блоховское волновое число и тем быстрее волна затухает вглубь сверхрешётки.

Чтобы энергия переносилась вдоль волноведущего слоя, стенки волновода должны быть непрозрачны для электромагнитных волн. Поэтому решение необходимо искать в области пересечения запрещённых зон обеих сверхрешёток. В этом случае распределение амплитуды полей вдоль оси z должно убывать по экспоненте от границ волноведущего слоя, а блоховские волновые числа должны быть комплексными. Для этого необходимо выбирать знаки $\mathrm{Im}\,\bar{k}_a$ и $\mathrm{Im}\,\bar{k}_b$ таким образом, чтобы экспонента $\exp(i\bar{k}_a z)$ убывала при $z \to -\infty$, а $\exp(i\bar{k}_b z)$ убывала при $z \to +\infty$. Математически выбор делается по знаку логарифма:

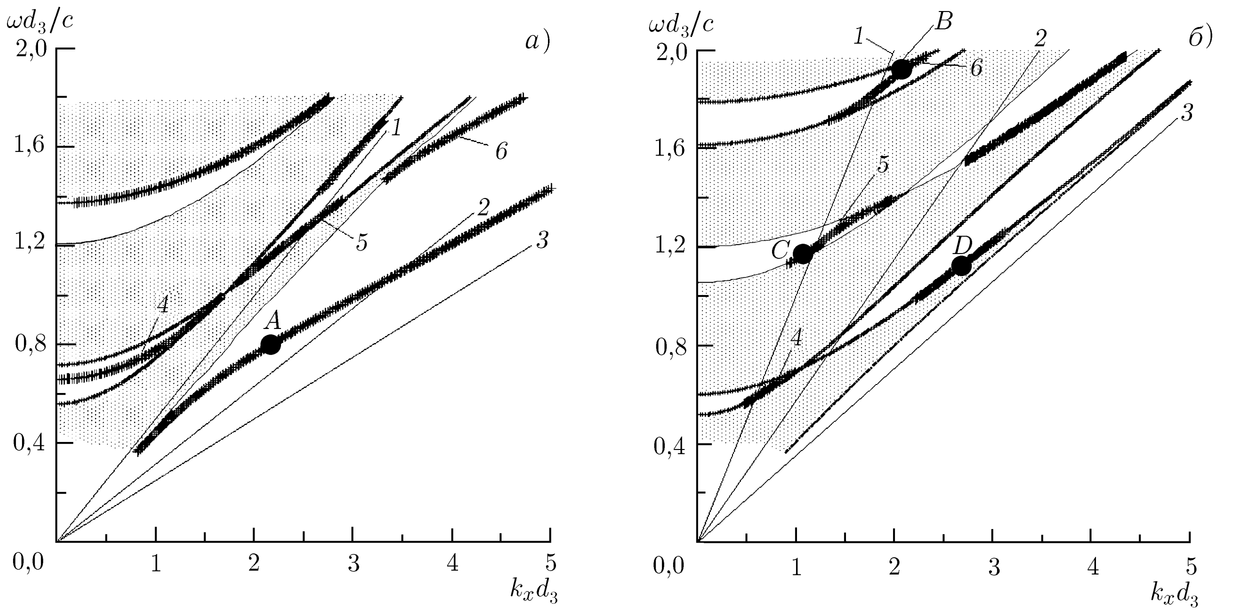

Рис. 3. Дисперсионные зависимости для волноводов с максимальным (*а*) и минимальным (*б*) значениями диэлектрической проницаемости волноведущего слоя

$$\bar{k}_a d = \ln\left[\frac{a_{11}+a_{22}}{2} \pm \sqrt{\left(\frac{a_{11}+a_{22}}{2}\right)^2 - 1}\right], \qquad \bar{k}_b d = \ln\left[\frac{b_{11}+b_{22}}{2} \pm \sqrt{\left(\frac{b_{11}+b_{22}}{2}\right)^2 - 1}\right]. \quad (10)$$

В работе исследовались два типа волноводов. Для первого типа волноведущий слой представлял собой среду с диэлектрической проницаемостью $\varepsilon_3 = 16$, а проницаемости слоёв сверхрешёток имели значения $\varepsilon_1 = \varepsilon_4 = 4$; $\varepsilon_2 = \varepsilon_5 = 10$. Волновод второго типа был образован воздушным зазором ($\varepsilon_3 = 1$) между двумя периодическими стенками с $\varepsilon_1 = \varepsilon_4 = 8$ и $\varepsilon_2 = \varepsilon_5 = 3$. В обоих случаях волноведущий слой располагался между слоями с одинаковой диэлектрической проницаемостью. Дисперсионные зависимости для волноводов первого и второго типов представлены на рис. 3*а* и *б* соответственно.

Для безграничной периодической структуры имеются два типа запрещённых зон, различающихся распределением полей: 0-зона, находящаяся между дисперсионными кривыми с $\mathrm{Re}(\bar{k}d) = 2\pi n$, где $n = 0, 1, 2, \ldots$, для которой период повторения поля в сверхрешётке равен периоду структуры; и π-зона, находящаяся между дисперсионными кривыми, которым соответствует $\mathrm{Re}(\bar{k}d) = \pi\,(2n+1)$. Для этой зоны период повторения составляет два периода структуры. На рис. 3*а* и 3*б* 0-зоны и π-зоны обозначены соответственно самыми тонкими линиями и линиями средней толщины. Решению дисперсионного соотношения для волновода со слоистопериодическими стенками соответствуют наиболее толстые линии.

На рис. 3*а*, *б* прямыми обозначены световые линии, для которых поперечные волновые числа равны нулю. Левее световой линии соответствующее k_z — действительное число. При переходе через световую линию k_z становится мнимым. В зависимости от комплексности поперечного волнового числа распределение полей в слоях может выражаться тригонометрическими или гиперболическими функциями. Во втором случае поле волны убывает от границ по экспоненте, т. е. волны «просачиваются» (туннелируют) через слой. Линии *1*, *2* и *3* на рис. 3*а* соответствуют слоям с диэлектрической проницаемостью 4, 10 и 16 соответственно, а на рис. 3*б* — слоям с $\varepsilon = 1$, 3, 8.

Отметим, что дисперсионные кривые должны быть непрерывны [7]. Дисперсионные кривые на рис. 3 из зон непропускания переходят в зоны пропускания (заштрихованы) непрерывным образом. Это означает, что энергия волны вытекает из волноведущего слоя, следовательно, решение отсутствует, поэтому эта часть дисперсионной кривой на рисунках не показана.

3. ПОЛЯ И ПОТОКИ ЭНЕРГИИ

Рассмотрим распределение компонент полей в исследуемой структуре. Для расчёта решение уравнений Максвелла удобно представить в следующем виде:

$$\begin{gathered} H_y(z) = \tilde{A}_1 \cos(k_z z) + i\tilde{A}_2 \sin(k_z z), \\ E_x(z) = -\frac{ic}{\omega}\,\frac{1}{\varepsilon}\,\frac{\partial H_y}{\partial z} = \frac{ic}{\omega}\,\frac{k_z}{\varepsilon}\,\big[\tilde{A}_1 \sin(k_z z) - i\tilde{A}_2 \cos(k_z z)\big]. \end{gathered} \tag{11}$$

Уравнения типа (11) записываются для каждого слоя. На границах слоёв комплексные амплитуды тангенциальных компонент электромагнитного поля связаны граничными условиями. В периодических структурах комплексные амплитуды H_y и E_x связаны с помощью теоремы Флоке (см. (4)). Таким образом, для границ волноведущего слоя и границ одного периода в каждой из периодических стенок получаем систему из десяти однородных линейных уравнений с десятью неизвестными. Её определитель — дисперсионное уравнение, полученное ранее с помощью метода матрицы преобразования. Решим эту систему, выразив все коэффициенты через один, например через A_1 (A_1 и A_2 — коэффициенты первого слоя левой сверхрешётки). Тогда для коэффициентов волноведущего слоя имеем

$$\tilde{A}_1 = A_1, \qquad \tilde{A}_2 = \frac{k_{z1}}{\varepsilon_1}\,\frac{\varepsilon_3}{k_{z3}}\,A_2, \tag{12}$$

а для прилегающего к волноведущему слою слоя правой сверхрешётки можно записать

$$\begin{aligned} \tilde{A}_1 = {} & A_1\Big[\cos(k_{z3}d_3)\cos(k_{z4}d_3) + \frac{k_{z3}}{\varepsilon_3}\,\frac{\varepsilon_4}{k_{z4}}\,\sin(k_{z3}d_3)\sin(k_{z4}d_3)\Big] + \\ & + i\,\frac{k_{z1}}{\varepsilon_1}\,\frac{\varepsilon_3}{k_{z3}}\,A_2\Big[\sin(k_{z3}d_3)\cos(k_{z4}d_3) - \frac{k_{z3}}{\varepsilon_3}\,\frac{\varepsilon_4}{k_{z4}}\,\cos(k_{z3}d_3)\sin(k_{z4}d_3)\Big], \\ \tilde{A}_2 = {} & -iA_1\Big[\cos(k_{z3}d_3)\sin(k_{z4}d_3) - \frac{k_{z3}}{\varepsilon_3}\,\frac{\varepsilon_4}{k_{z4}}\,\sin(k_{z3}d_3)\cos(k_{z4}d_3)\Big] + \\ & + \frac{k_{z1}}{\varepsilon_1}\,\frac{\varepsilon_3}{k_{z3}}\,A_2\Big[\sin(k_{z3}d_3)\sin(k_{z4}d_3) + \frac{k_{z3}}{\varepsilon_3}\,\frac{\varepsilon_4}{k_{z4}}\,\cos(k_{z3}d_3)\cos(k_{z4}d_3)\Big]. \end{aligned} \tag{13}$$

Коэффициенты A_1 и A_2 связаны между собой следующим образом:

$$\begin{aligned} A_2 = {} & -iA_1 \times \\ & \times \frac{\sin(k_{z2}d_a) - \Big[\cos(k_{z1}d_1)\sin(k_{z2}d_1) - \dfrac{k_{z1}}{\varepsilon_1}\,\dfrac{\varepsilon_2}{k_{z2}}\,\sin(k_{z1}d_1)\cos(k_{z2}d_1)\Big]\exp(i\bar{k}_a d_a)}{\dfrac{k_{z1}}{\varepsilon_1}\,\dfrac{\varepsilon_2}{k_{z2}}\,\cos(k_{z2}d_a) - \Big[\sin(k_{z1}d_1)\sin(k_{z2}d_1) + \dfrac{k_{z1}}{\varepsilon_1}\,\dfrac{\varepsilon_2}{k_{z2}}\,\cos(k_{z1}d_1)\cos(k_{z2}d_1)\Big]\exp(i\bar{k}_a d_a)}\,. \end{aligned} \tag{14}$$

Для других слоёв формулы имеют вид, аналогичный (13).

Были также рассчитаны потоки энергии в соответствии с выражением

$$\mathbf{P} = \frac{c}{8\pi}\,\mathrm{Re}\,[\mathbf{E}, \mathbf{H}^*]. \tag{15}$$

Рассматривалась только проекция потока энергии на ось x:

$$P_x = -\frac{c^2}{8\pi}\,\frac{k_x}{\omega\varepsilon}\,H_y^2, \tag{16}$$

где H_y — действительная величина.

Исследование полей и потоков энергии проводилось численно. В ходе исследования был рассмотрен ряд точек, отмеченных на рис. 3*а, б*. Наиболее чётко волноводные свойства центрального слоя выражены для случая, когда его диэлектрическая проницаемость равна единице. Тогда основная часть энергии переносится внутри этого слоя. Это видно на эпюрах рис. 4, 5, которые соответствуют точкам B, C на рис. 3*б*. Рис. 4 соответствует случаю, когда по границам волноведущего слоя распространяются поверхностные волны, а поля внутри этого слоя описываются гиперболическими функциями. Для случая, изображённого на рис. 5, компонента H_y в центральном слое пропорциональна $\sin(k_{z3}d_3)$, а $E_x \propto \cos(k_{z3}d_3)$, т. е. распределение полей подобно распространению в волноводе с металлическими стенками. Именно этот частный случай рассматривался в работе [2].

Для центрального слоя с максимальным значением диэлектрической проницаемости основная энергия также может переноситься в волноведущем слое. При этом поля в стенках волновода убывают по экспоненте вглубь сверхрешёток, т. к. диэлектрические проницаемости их слоёв меньше, чем ε_3 (рис. 6*а, б*). Этот случай аналогичен диэлектрическому волноводу.

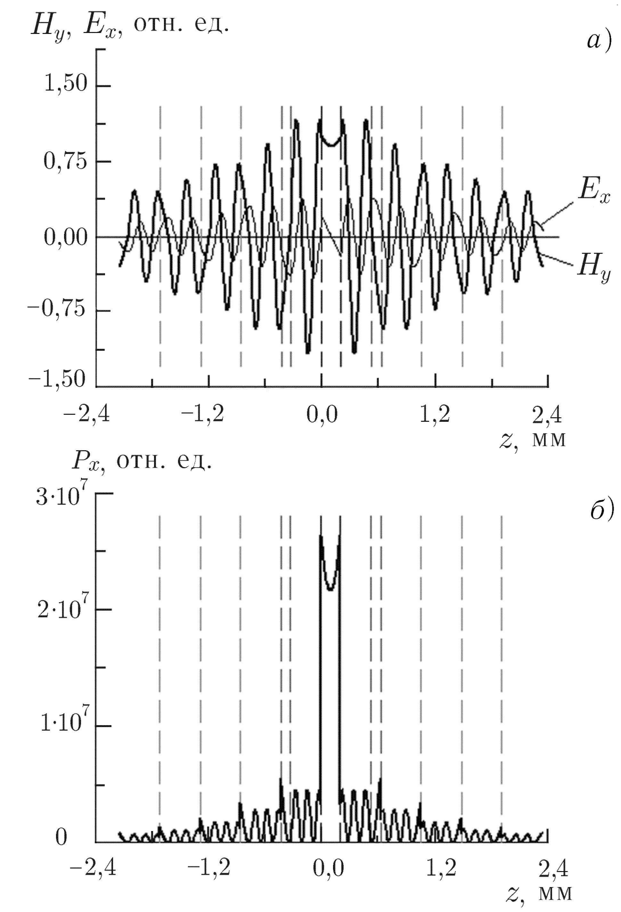

Рис. 4. Распределение компонент H_y и E_x электромагнитного поля (*а*) и потока энергии (*б*) для точки B рис. 3*б*. Штриховыми линиями обозначены границы слоёв

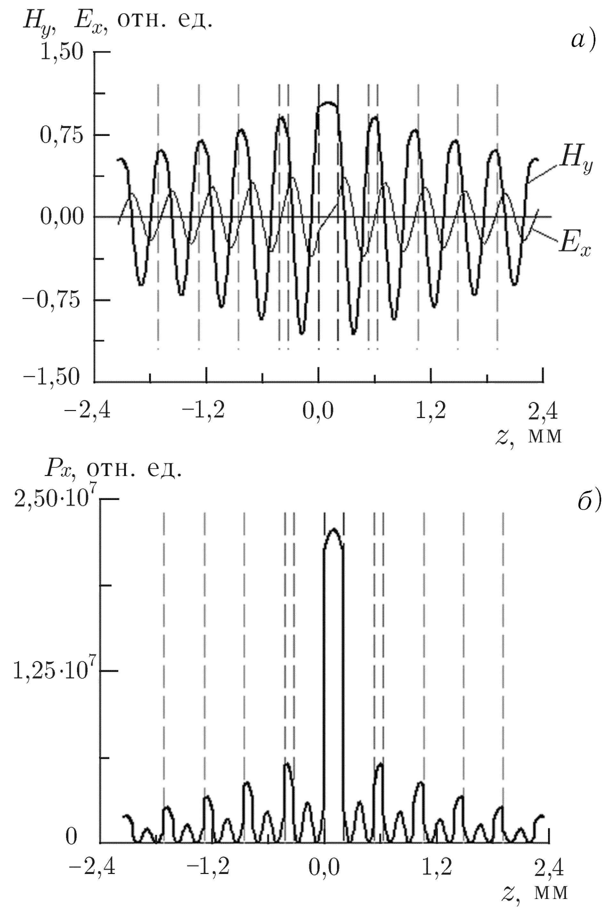

Рис. 5. Распределение компонент H_y и E_x электромагнитного поля (*а*) и потока энергии (*б*) для точки C рис. 3*б*. Штриховыми линиями обозначены границы слоёв

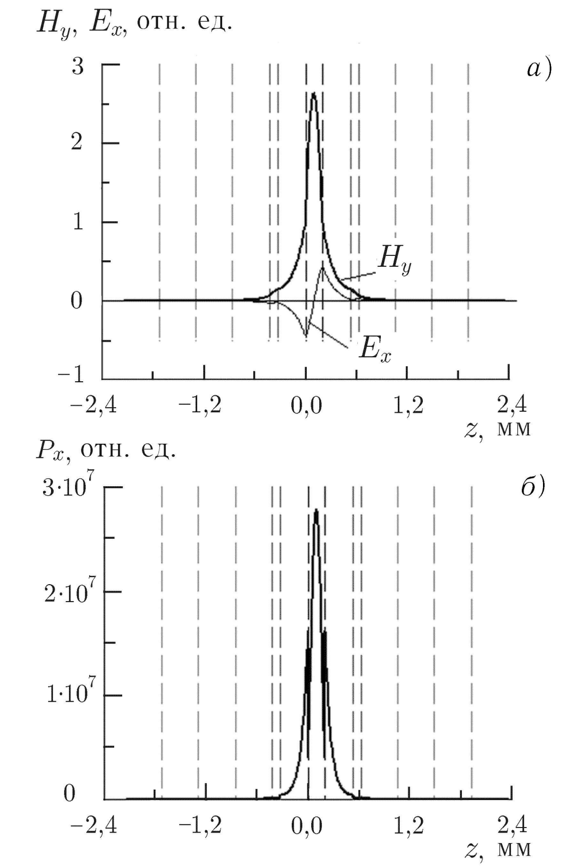

Рис. 6. Распределение компонент H_y и E_x электромагнитного поля (*а*) и потока энергии (*б*) для точки А рис. 3*а*. Штриховыми линиями обозначены границы слоёв

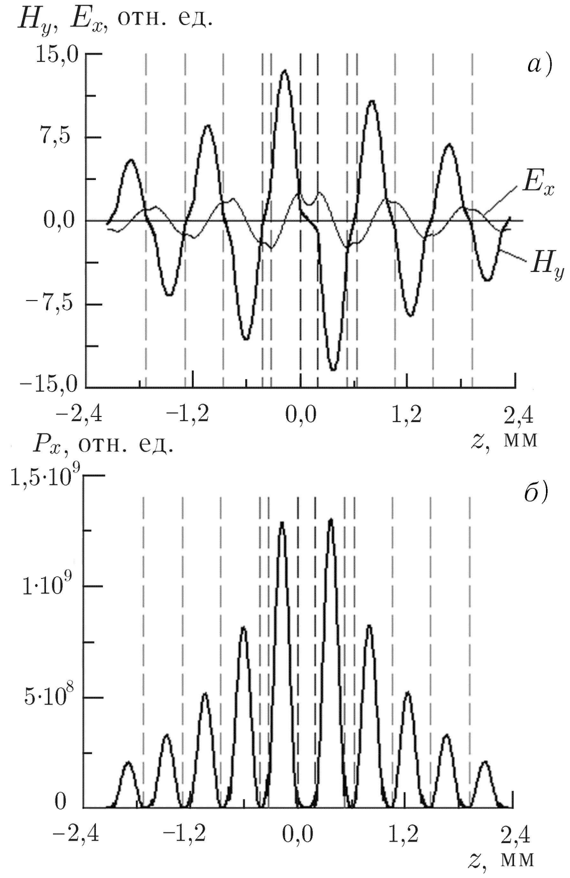

Рис. 7. Распределение компонент H_y и E_x электромагнитного поля (*а*) и потока энергии (*б*) для точки D рис. 3*б*. Штриховыми линиями обозначены границы слоёв

В волноводах обоих типов некоторая часть энергии переносится в слоях сверхрешёток. В случае, изображённом на рис. 6, эта доля энергии меньше, чем в случае, изображённом на рис. 5. Интересным представляется случай, когда точка дисперсионной кривой расположена правее световых линий волноведущего слоя и одного из слоёв решётки (рис. 3*б*, линии *1* и *2*). Тогда энергия, переносимая в центральном слое, может оказаться почти равной нулю, а весь поток энергии оказывается сосредоточенным в стенках волновода (рис. 7). Такая ситуация может быть использована для разделения сигнала, например, для получения двух когерентных, разделённых в пространстве пучков из одного лазера. Заметим, что в работе [2] данный случай не был рассмотрен.

На рис. 8*а, б* представлена относительная глубина проникновения поля в стенки волновода

$$\delta = |\bar{k}d|^{-1} \tag{17}$$

для дисперсионных кривых *4, 5, 6* на рис. 3*а, б*. Из сравнения рис. 8 и 3 видно, что при подходе дисперсионных кривых к краю запрещённой зоны глубина проникновения стремится к бесконечности, а дисперсионные кривые непрерывным образом переходят в зоны пропускания сверхрешёток, и волноводный эффект не имеет места. Отметим, что наименьшей глубиной проникновения обладает волновод поверхностных волн (кривая *6* (точка B) рис. 3*б* и кривая *6* рис. 8*б*).

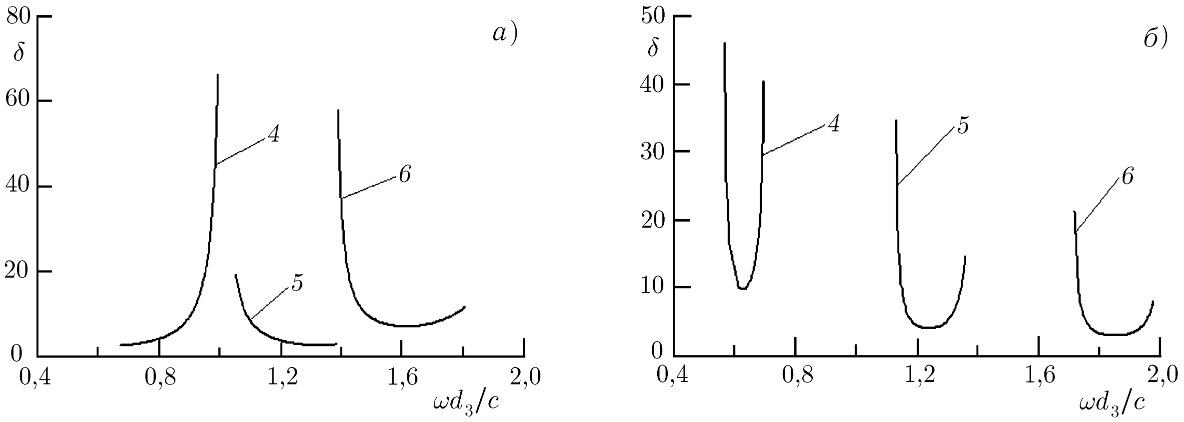

Рис. 8. Глубина проникновения поля в стенки волноводов с максимальным (*а*) и минимальным (*б*) значениями диэлектрической проницаемости волноведущего слоя

4. ЗАКЛЮЧЕНИЕ

В итоге можно сказать, что в рассматриваемой структуре возможно волноводное распространение в случаях наименьшей и наибольшей диэлектрической проницаемости (по сравнению с проницаемостями слоёв сверхрешёток) волноведущего слоя. Показано, что эффективным может оказаться волновод поверхностных волн, в котором поток энергии сосредоточен вблизи стенок волноведущего слоя. В этом случае глубина проникновения в сверхрешётки наименьшая. Также показана возможность разделения сигнала, когда поток в волноведущем слое отсутствует, а вся энергия переносится в периодических стенках.

Рассмотренные свойства волновода с периодическими стенками показывают, что перспективными областями применения таких структур является диапазон от миллиметровых до субмикронных длин волн. Такие волноводы обладают избирательными свойствами, существенным образом зависящими от ширины запрещённых зон, управлять которыми можно, меняя параметры слоёв сверхрешёток. Кроме того, ввод энергии в такие волноводы возможен непосредственно через периодические стенки, если параметры сверхрешётки в области ввода и вывода информации близки к краю запрещённой зоны. Эти системы также могут представлять интерес для вывода энергии из твердотельных лазеров, чтобы избежать специфических для твёрдого тела потерь (например, на возбуждение оптических фононов [8]).

СПИСОК ЛИТЕРАТУРЫ

1. Kawanishi T., Izutsu M. // Opt. Express. 2000. V. 7, No. 1. P. 10.
2. Cho A. Y., Yariv A., Yeh P. // Appl. Phys. Lett. 1977. V. 30, No. 9. P. 471.
3. Trull J., Vilaseca R., Martorell J., Corbalan R. // Opt. Letters. 1995. V. 7, No. 17. P. 1746.
4. Басс Ф. Г., Булгаков А. А., Тетервов А. П. Высокочастотные свойства полупроводников со сверхрешёткой. М.: Наука, 1989. 288 с.
5. Борн М., Вольф Э. Основы оптики. М.: Наука, 1973. 720 с.
6. Лившиц И. М., Розенцвейг Л. Н. // ЖЭТФ. 1948. Т. 18, № 11. С. 1012.
7. Физическая акустика / Под ред. У. Мэзона. Т. 1. М.: Мир, 1966.
8. Köhler R., Tredicucci A., Beltram F., et al. // Nature. 2002. V. 417, No. 9. P. 156.

Поступила в редакцию 13 октября 2003 г.; принята в печать 18 июля 2004 г.

ELECTRODYNAMIC PROPERTIES OF A WAVEGUIDE WITH LAYERED-PERIODIC WALLS

A. A. Bulgakov, O. V. Kostylyova, and A. V. Meriuts

We study a waveguide with layered-periodic walls for different relations between the dielectric permittivities of the central (guiding) layer and the superlattice layers. We consider guided propagation in such a waveguide, predict the appearance of surface waves in the guiding layer, and discuss the case where almost all of energy is transferred in periodic walls.

УДК 517.5+519.64+535.8+538.56+621.3

ТЕОРИЯ ВОЛНОВОДНОГО РАССЕЯНИЯ СВЕТА В ИНТЕГРАЛЬНО-ОПТИЧЕСКОМ ВОЛНОВОДЕ ПРИ НАЛИЧИИ ШУМА

А. А. Егоров

Институт общей физики им. А. М. Прохорова РАН,
Российский университет дружбы народов,
г. Москва, Россия

Представлена векторная теория рассеяния монохроматического света в интегрально-оптическом волноводе с произвольными трёхмерными нерегулярностями при наличии шума. Решение трёхмерной электродинамической задачи о рассеянии лазерного излучения в нерегулярном волноводе получено методом связанных мод в первом приближении теории возмущений. Приближённое решение неоднородного трёхмерного волнового уравнения найдено с помощью метода мод и метода функций Грина. Анализ полученной функции Грина выполнен для случаев распространяющихся и затухающих мод нерегулярного несимметричного оптического волновода. Приведены выражения для полей излучения в ближней и дальней зонах, дан их предварительный анализ.

ВВЕДЕНИЕ

В наших предыдущих работах [1–5] продемонстрирована возможность использования волноводного рассеяния лазерного излучения для получения информации о статистических свойствах нерегулярностей из зашумлённых данных измерения в дальней зоне. Измеренные в дальней зоне диаграммы рассеяния использовались для нахождения приближённого корректного решения обратной задачи рассеяния на трёхмерных шероховатостях подложки планарного оптического волновода. Понятно, что такой двумерный анализ применим далеко не всегда, а в ряде случаев может использоваться только приближённо, например в случае рассеяния монохроматического света в устройствах интегральной оптоэлектроники, созданных на основе различных волноводов с трёхмерной геометрией элементов. В связи с этим развитие векторной теории волноводного рассеяния света в нерегулярном интегральном волноводе при наличии шума является актуальной задачей современной интегральной оптики и волноводной оптоэлектроники. В опубликованных ранее работах по приближённому трёхмерному рассмотрению рассеяния в оптических волноводах [6–9], например, не учитывается рассеяние в ближней зоне, а электродинамическая задача решается, как правило, без учёта шумов. Оба этих фактора являются принципиально важными в свете дальнейшей миниатюризации и интеграции элементов и устройств интегральной оптики и волноводной оптоэлектроники. Шумы и рассеяние лазерного излучения в интегрально-оптических анализаторах спектра или мультиплексорах/демультиплексорах являются критически важными факторами, ограничивающими их работоспособность, особенно при субмикронных размерах основных элементов устройств.

1. ТРЁХМЕРНАЯ ЗАДАЧА ВОЛНОВОДНОГО РАССЕЯНИЯ. МЕТОД ФУНКЦИЙ ГРИНА

Рассмотрим рассеяние монохроматического лазерного излучения волноводной моды в интегрально-оптическом волноводе, содержащем произвольные нерегулярности. Нерегулярный трёхслойный планарный волновод и схема регистрации рассеянного излучения показаны на рис. 1, где

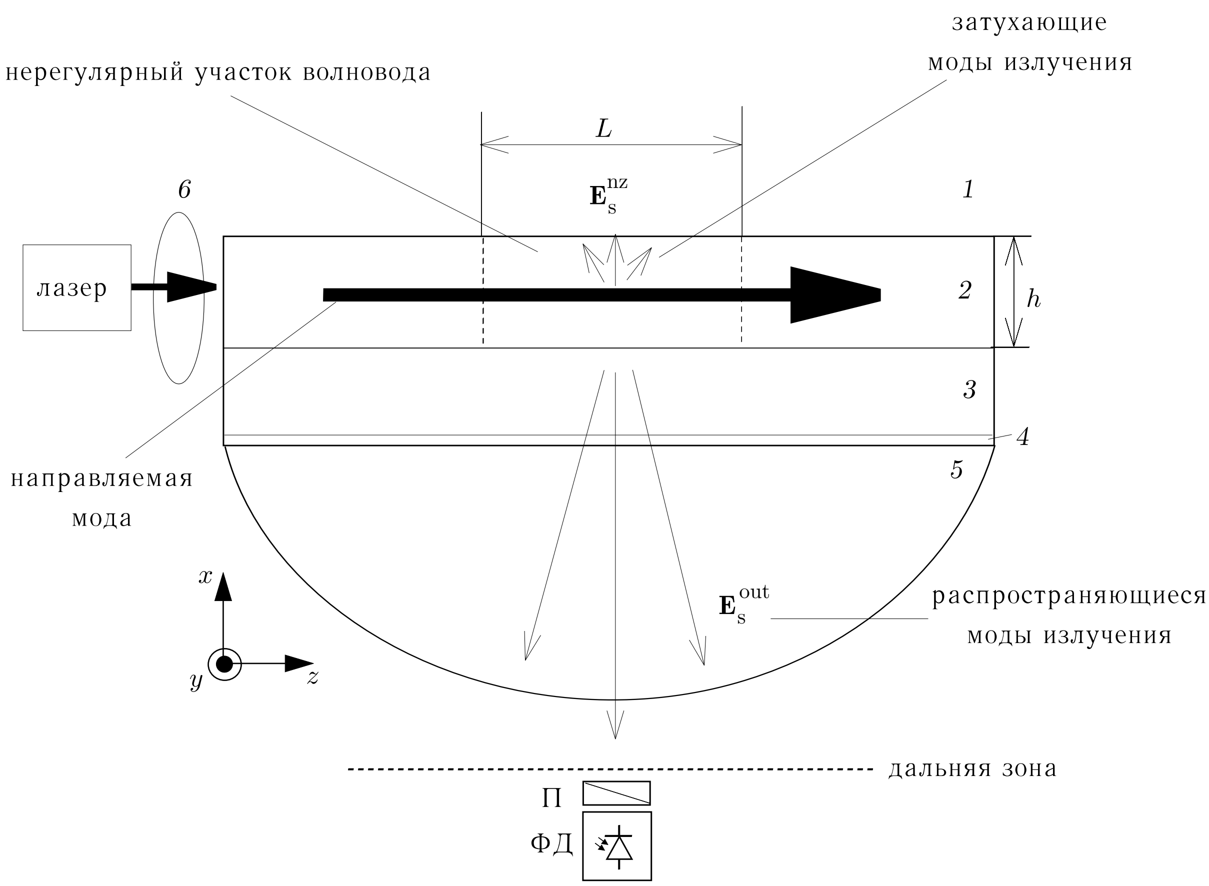

Рис. 1

1 — обрамляющая среда (воздух), *2* — волноводный слой, *3* — подложка, *4* — тонкий слой иммерсии, *5* — кварцевый полукруг (или полусфера), *6* — линза, L — длина нерегулярной области, h — толщина волновода, П — поляризатор, ФД — фотодетектор. Для повышения отношения сигнал/шум может применяться механический или электрооптический модулятор вводимого в волновод лазерного излучения, а также синхронное детектирование. Сигнал с фотоприёмника поступает на резонансный усилитель и регистрируется в аналоговом виде с помощью самописца или в цифровом виде (после аналого-цифрового преобразователя) на компьютере для последующей обработки данных измерений. Ряд устройств на рис. 1 не показан.

Как правило, планарный волновод состоит из следующих слоёв: обрамляющей среды, волноводного слоя и подложки с показателями преломления n_1, n_2 и n_3 соответственно. Нерегулярности структуры планарного волновода могут быть обусловлены неровностями (шероховатостями) границ раздела сред, образующих волновод, подповерхностными дефектами (так называемый нарушенный или трещиноватый слой) и неоднородностями показателя преломления волноводного слоя. Неоднородности волноводного слоя и подповерхностные дефекты при рассмотрении задачи рассеяния могут быть описаны однотипно — как неоднородности показателя преломления соответствующей среды волновода. Для упрощения анализа пренебрежём этими неоднородностями, а также кросскорреляционными связями между неровностями границ раздела.

Электродинамическая задача о рассеянии направляемой волноводной моды в оптическом волноводе, содержащем нерегулярности, решается методом связанных мод с помощью теории возмущений [1, 5, 6]. Для описания электромагнитного поля в нерегулярном волноводе используется уравнение, которое в трёхмерных декартовых координатах имеет вид [1, 7]

$$\nabla^2 \mathbf{E} + \nabla \left(\mathbf{E} \, \frac{\nabla \varepsilon_i}{\varepsilon_i} \right) + \omega^2 \mu_i \varepsilon_i \mathbf{E} = 0, \tag{1}$$

где ε_i и μ_i — диэлектрическая и магнитная проницаемости в i-м слое планарного волновода ($i = 1, 2, 3$), $\omega = 2\pi f$, f — частота поля, $\omega \sqrt{\mu_i \varepsilon_i} = n_i k$, n_i — показатель преломления i-го слоя, $k = 2\pi/\lambda$, λ — длина волны света в вакууме, $\nabla^2 \mathbf{E} = \Delta \mathbf{E}$ — лапласиан электрической компоненты.

Рассматривается случай распространения в волноводе вдоль оси z фундаментальной ТЕ-моды с компонентами E_{0y}, H_x, H_z (для ТМ-моды анализ проводится аналогично). Полное поле в нерегулярном планарном оптическом волноводе можно записать в виде суммы электрических полей падающей волноводной моды $\mathbf{E}_{0y}$, пóля рассеянной волны $\mathbf{E}_{\mathrm{s}}$ и пóля шумовой компоненты $\mathbf{E}_{\mathrm{w}}(x, y, z)$: $\mathbf{E} = [\mathbf{E}_{0y}(x, y, z) + \mathbf{E}_{\mathrm{s}}(x, y, z)] + \mathbf{E}_{\mathrm{w}}(x, y, z)$. В случае мультипликативной компоненты шума второй знак плюс надо заменить знаком умножения. При такой записи полагается, что все источники шума (независимо от их природы) дают вклад в полное поле $\mathbf{E}$ как один эффективный источник шума $\mathbf{E}_{\mathrm{w}}$, приведённый к плоскости волновода. При решении задачи для конкретного интегрального волновода и заданной экспериментальной установки роль каждого источника шума может быть оценена в рамках данного подхода. Например, может быть использован метод решения, описанный в нашей работе [3], где использовалась модель аддитивного действительного «белого» шума, близкого по уровню к экспериментальному шуму. Более того, вполне допустимо считать, что эффективная сумма шумов, дающая вклад в измеряемую диаграмму рассеяния, по центральной предельной теореме, как и большинство реальных случайных процессов в оптике, подчиняется гауссовому распределению [3]. В пределах ограниченного диапазона регистрируемых мод рассеяния эта гауссова функция вполне может быть заменена равномерным спектром. Отметим также, что основными источниками шума в установке, показанной на рис. 1, являются лазер, модулятор, фотоприёмник, усилитель и источники внешней засветки. Как известно [3], шумы, возникающие при работе этих устройств, описываются в основном гауссовыми функциями. Поэтому сделанное выше предположение о представлении этих шумов некоторым эффективным суммарным «белым» шумом, дающим вклад в измеряемую диаграмму рассеяния, является вполне оправданным. Спектр вибрационных шумов (механический модулятор и т. д.) обычно ограничен сверху частотой 5÷10 кГц и эффективно компенсируется, например, синхронным детектированием. Шум, связанный с вычислениями на компьютере, во многом аналогичен шуму квантования в аналого-цифровых преобразователях, и его действие также подобно «белому» шуму. Погрешность, вносимая им в расчёты, не превышает 1÷2 %, и в большинстве случаев этот шум можно не учитывать. Учитывая сложность достоверного определения спектра экспериментального шума в широком диапазоне постоянных распространения мод рассеяния, можно считать на данном этапе такой подход вполне удовлетворительным. При исследовании статистического ансамбля однотипных микрообъектов (шероховатость поверхности или объёмные неоднородности волноводного слоя) следует учитывать, что фазовая информация содержится в измеряемой диаграмме рассеяния (и, соответственно, в энергетическом спектре) в усреднённом виде. Этот факт необходимо принимать во внимание при анализе фазовой проблемы, т. к. из-за сложного характера формирования диаграммы при волноводном рассеянии лазерного излучения на статистическом ансамбле нерегулярностей (особенно при наличии шума) трудно воспользоваться комплексной, т. е. амплитудно-фазовой диаграммой рассеяния, получаемой, например, методами интерферометрии. В общем случае для решения этой проблемы используется пространственная ограниченность объекта и свойства аналитических сигналов, позволяющие найти фазу с помощью преобразования Гильберта.

Как и ранее, будем полагать, что диэлектрическая проницаемость может быть представлена в виде $\varepsilon_i(x, y, z) = \varepsilon_i(\mathbf{r}) = \varepsilon_{0i}(x, z) + \Delta\varepsilon_i(x, y, z)$. Здесь $\varepsilon_{0i}(x, z)$ описывает регулярные свойства соответствующего слоя планарного волновода, а добавка $\Delta\varepsilon_i(x, y, z)$ описывает произвольные (в общем случае — случайные с произвольной статистикой нерегулярностей) трёхмерные нерегулярности структуры волновода (как неровности границ раздела сред планарного волновода, так и неоднородности показателя преломления в каждом i-м слое). При решении поставленной электродинамической задачи для волновода с определённым типом нерегулярностей вид $\Delta\varepsilon_i$ конкретизируется, позволяя в результате получить явную зависимость амплитудных коэффициентов направляемых мод и мод рассеяния (а следовательно, и полей рассеяния) как от функции неровностей границ раздела сред планарного волновода, так и от функции неоднородностей показателя преломления. Изложение этого метода выходит за рамки данной статьи. Отметим только, что в первом приближении теории возмущений рассеянное поле $\mathbf{E}_{\mathrm{s}}$ (см. ниже формулу (4)) зависит от коэффициента разложения q, который пропорционален фурье-преобразованию от $\Delta\varepsilon_i$, т. е. зависит от функции неровности границы волновода. Для применения теории возмущений добавка $\Delta\varepsilon_i$ не обязательно должна быть величиной высокого порядка малости. Вполне достаточно, чтобы область, в пределах которой эта добавка отличается от нуля, была очень узкой [1, 7]. Например, для планарного оптического волновода с толщиной волноводного слоя h данное приближение заведомо выполняется при среднеквадратичной высоте неровностей (шероховатостей) границ волновода $\sigma \ll h$. Более точная оценка, полученная нами из анализа соотношения неопределённостей Гейзенберга для случая рассеяния лазерного излучения в интегрально-оптическом волноводе, даёт следующий возможный диапазон изменения среднеквадратичной высоты шероховатостей границ волновода: 10^{-2} Å $< \sigma < 10^{3}$ Å, и соответствующий интервал корреляции шероховатостей: 10^{-3} мкм $< r < 10^{2}$ мкм.[1] В работах [3–5] показано, что нашим методом решения обратной задачи волноводного рассеяния света можно восстановить статистические характеристики второго порядка и вполне надёжно определить соответствующие параметры шероховатости в широком диапазоне значений интервала корреляции шероховатостей, включая и размеры порядка λ, при изменении σ в ограниченном интервале: примерно от 1 до 500 Å, и отношении сигнал/шум около 10. При этом субволновые радиусы корреляции из интервала $\lambda/35 \div \lambda/15$ могут быть определены с относительной ошибкой примерно $5 \div 15$ %, а радиусы корреляции из интервала $\lambda/2 \div 15\lambda$ — с ошибкой не более $25 \div 30$ % [4]. Использование информации о локальных полях вблизи источника, несущих важнейшую информацию о субволновой структуре нерегулярностей (см. раздел 1.2), позволит улучшить разрешение метода волноводного рассеяния света.

С учётом приведённых выше выражений для полного поля и диэлектрической проницаемости можно получить из (1) приближённое трёхмерное уравнение. Оставляя в нём только члены первого порядка малости относительно $\mathbf{E}_{\mathrm{s}}$, $\Delta\varepsilon_i\,\mathbf{E}_{0y}$ и $\Delta\varepsilon_i\,\mathbf{E}_{\mathrm{w}}$, найдём приближённое неоднородное трёхмерное волновое уравнение, которое можно рассматривать как однородное волновое уравнение с возмущением в виде «источника» в правой части:

[1] Данная оценка означает, что при последовательном изменении параметров нерегулярностей в указанном диапазоне наблюдаются изменения в диаграмме рассеяния, которые могут быть зарегистрированы. По ним, в принципе, может быть определена величина нерегулярностей, обусловившая эти изменения в диаграмме рассеяния. Параметры реальных нерегулярностей вполне укладываются в указанный диапазон [3, 6–9]. В рамках теоретической работы можно только ещё отметить, что в прикладных исследованиях полезно делать как априорные оценки параметров нерегулярностей, так и учитывать особенности изготовления конкретного волновода, позволяющие определить, какие из нерегулярностей — неровности границ раздела (шероховатости) и/или (объёмные) неоднородности показателя преломления — будут доминировать в процессе рассеяния. Например, в волноводах с волноводным слоем из оптически прозрачной чистой жидкости можно пренебречь рассеянием на объёмных неоднородностях показателя преломления волноводного слоя и учитывать только рассеяние на неровностях границ раздела. В волноводных плёнках, полученных методом термического испарения, ионного распыления или химического осаждения пренебречь рассеянием на неоднородностях показателя преломления волноводного слоя можно только в рамках некоторых модельных предположений, учтённых при проведении эксперимента, т. к. радиус корреляции неоднородностей показателя преломления плёнок в этом случае может изменяться от приблизительно 0,05 до 10 мкм.

$$\nabla^2 \mathbf{E}_\mathrm{s}(x,y,z) + \omega^2\mu\varepsilon_{0i}\mathbf{E}_\mathrm{s}(x,y,z) \approx -\omega^2\mu\varepsilon_{0i}\,\Delta\varepsilon_i(x,y,z)\,[\mathbf{E}_{0y}(x,y,z) + \mathbf{E}_\mathrm{w}(x,y,z)], \qquad (2)$$

где E_{0y} — решение однородного невозмущённого уравнения, описывающего распространение TE_0-моды в волноводе. С энергетической точки зрения «источником» в правой части уравнения (2) является интенсивность моды, падающей на нерегулярный трёхмерный участок волновода при наличии шума $\mathbf{E}_\mathrm{w}$ и рассеиваемой в сферу во всём окружающем пространстве. Решение данного неоднородного волнового уравнения методом функций Грина может быть получено в виде свёртки некоторой трёхмерной функции Грина $G(x,y,z;x',y',z')$ с выражением для источника [1, 2, 6–9]:

$$\mathbf{E}_\mathrm{s}(x,y,z) = -\omega^2\mu\varepsilon_{0i}\iiint \Delta\varepsilon_i(x',y',z')G(x,y,z;x',y',z')\times \\ \times[\mathbf{E}_{0y}(x',y',z') + \mathbf{E}_\mathrm{w}(x',y',z')]\,\mathrm{d}x'\,\mathrm{d}y'\,\mathrm{d}z'. \qquad (3)$$

Этот подход отражает возникновение в волноводе с трёхмерными нерегулярностями гибридных мод, имеющих шесть компонент поля, а не три, как TE- и TM-моды [2, 7–9]. Для гибридных мод не выполняется условие $\partial\mathbf{E}/\partial y = 0$, т. е. существуют вариации полей в направлении оси y. Следовательно, в случае трёхмерных нерегулярностей любое произвольное распределение поля планарного оптического волновода необходимо в общем случае представлять в виде разложения по всем возможным модам плоского волновода, включая суммирование и интегрирование по модам, соответствующим вариациям поля по второй поперечной координате — вдоль оси y.

1.1. Поиск выражения для трёхмерной волноводной функции Грина методом мод

Найдём аналитическое выражение для функции Грина $G(\mathbf{r};\mathbf{r}')$ с помощью метода мод [1, 2, 7–9]. Для этого представим решение неоднородного трёхмерного уравнения (2) в следующем виде:

$$\mathbf{E}_{\mathrm{s}y}(x,y,z) = \sum_\nu \int\limits_{-\infty}^{+\infty} C(\nu,\beta_y)\mathbf{E}_{\nu y}(x,z)\exp(-i\beta_y y)\,\mathrm{d}\beta_y + \\ + \int\limits_{-\infty}^{+\infty}\mathrm{d}\beta_y \int\limits_{-\beta_3}^{+\beta_3} q(\beta,\beta_y)\mathbf{E}_{\beta y}(x,z)\exp(-i\beta_y y)\,\mathrm{d}\beta, \qquad (4)$$

где выполнено суммирование по распространяющимся чётным и нечётным TE-модам (переменная ν принадлежит множеству натуральных чисел I), а комбинация из двух интегралов учитывает все моды излучения (для TM-мод анализ проводится аналогично). В (4) β и β_y — постоянные распространения мод рассеяния, формирующих диаграмму рассеяния, вдоль осей z и y соответственно, $\beta_3 = kn_3$; интегрирование по постоянным распространения β_y учитывает вариации поля вдоль оси y, $C(\nu,\beta_y)$ и $q(\beta,\beta_y)$ — коэффициенты разложения, которые находятся с помощью соотношений ортогональности [7–9], $\mathbf{E}$ — электрические составляющие полей направляемых (дискретных) мод и (непрерывных) мод излучения несимметричного интегрального оптического волновода. Нижний индекс y у составляющей поля $\mathbf{E}$ означает, что в уравнении (4) пока предполагается сохранение исходной TE-поляризации. При решении неоднородного трёхмерного уравнения (2) в виде (4) мы используем теоретические методы, изложенные как в работах независимых авторов [7–11], так и в наших публикациях (см., например, [1–5]).

Подставив выражение (4) в исходное неоднородное уравнение (2) и выполнив дифференцирование, получим

$$\sum_\nu \int\limits_{-\infty}^{+\infty} C(\nu,\beta_y)\mathbf{E}_{\nu y}(x,z)\exp(-i\beta_y y)\,(\beta_\nu^2-\beta_y^2)\,\mathrm{d}\beta_y + \int\limits_{-\infty}^{+\infty}\mathrm{d}\beta_y \int\limits_{-\beta_3}^{+\beta_3} q(\beta,\beta_y)\mathbf{E}_{\beta y}(x,z)\times$$
$$\times\exp(-i\beta_y y)\,(\beta^2-\beta_y^2)\,\mathrm{d}\beta = -\omega^2\mu\varepsilon_{0i}\,\Delta\varepsilon_i(x,y,z)\,[\mathbf{E}_{0y}(x',y',z')+\mathbf{E}_{\mathrm{w}}(x',y',z')]. \quad (5)$$

Для нахождения коэффициентов разложения $C(\nu,\beta_y)$ и $q(\beta,\beta_y)$ умножим обе части уравнения (5) на $(1/2\pi)\exp(i\beta_y' y)$ и проинтегрируем его по y от $-\infty$ до $+\infty$, используя свойства дельта-функции Дирака [10]. В результате получим следующее уравнение:

$$\sum_\nu \int \mathbf{E}_{\nu y}(x,z)\,\delta(\beta_y'-\beta_y)\,(\beta_\nu^2-\beta_y^2)\,C(\nu,\beta_y)\,\mathrm{d}\beta_y + \iint q(\beta,\beta_y)\mathbf{E}_{\beta y}(x,z)\,\delta(\beta_y'-\beta_y)\,(\beta^2-\beta_y^2)\times$$
$$\times\,\mathrm{d}\beta\,\mathrm{d}\beta_y = -\frac{1}{2\pi}\int \omega^2\mu\varepsilon_{0i}\,\Delta\varepsilon_i(x,y,z)\,[\mathbf{E}_{0y}(x',y',z')+\mathbf{E}_{\mathrm{w}}(x',y',z')]\,\exp(i\beta_y' y)\,\mathrm{d}y, \quad (6)$$

где $\delta(\beta_y'-\beta_y)$ — одномерная дельта-функция Дирака.

Выполнив интегрирование по β_y и опустив штрихи, получим уравнение

$$\sum_\nu C(\nu,\beta_y)\,(\beta_\nu^2-\beta_y^2)\,\mathbf{E}_{\nu y}(x,z) + \int\limits_{-\beta_3}^{+\beta_3} q(\beta,\beta_y)\,(\beta^2-\beta_y^2)\,\mathbf{E}_{\beta y}(x,z)\,\mathrm{d}\beta =$$
$$= -\frac{1}{2\pi}\int\limits_{-\infty}^{+\infty}\omega^2\mu\varepsilon_{0i}\,\Delta\varepsilon_i(x,y,z)\,[\mathbf{E}_{0y}(x',y',z')+\mathbf{E}_{\mathrm{w}}(x',y',z')]\exp(i\beta_y' y)\,\mathrm{d}y. \quad (7)$$

Для получения системы связанных уравнений необходимо умножить полученное уравнение (7) на комплексно-сопряжённое поле $\mathbf{E}^*_{\nu' y}(x,z)$ и проинтегрировать результат по поперечному сечению волновода, используя свойства ортонормированности волноводных полей [5, 7–9]:

$$\frac{\beta_\nu}{2\omega\mu_0}\iint \mathbf{E}^*_{\nu' y}(x,z)\mathbf{E}_{\nu y}(x,z)\,\mathrm{d}x\,\mathrm{d}z = P\delta_{\nu'\nu}, \quad (8)$$

где $\mathbf{E}_{\nu y}(x,z)$ и $\mathbf{E}_{\nu' y}(x,z)$ — поля соответствующих мод волновода, P — мощность (например, единичная), переносимая направляемой модой планарного волновода, $\delta_{\nu'\nu}$ — символ Кронекера, $\nu\in I$. В результате получаем

$$\sum_\nu C(\nu,\beta_y)\,(\beta_\nu^2-\beta_y^2)\,\delta_{\nu'\nu} =$$
$$= -\frac{1}{2\pi}\iiint \omega^2\mu\varepsilon_{0i}\,\Delta\varepsilon_i(x,y,z)\,\mathbf{E}^*_{\nu y}(x,z)\,[\mathbf{E}_{0y}(x,y,z)+\mathbf{E}_{\mathrm{w}}(x,y,z)]\exp(i\beta_y y)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z. \quad (9)$$

Из формулы (9) можно найти выражение для коэффициента разложения направляемых волноводных мод:

$$C(\nu,\beta_y) = -\frac{1}{2\pi\,(\beta_\nu^2-\beta_y^2)}\times$$
$$\times\iiint \omega^2\mu\varepsilon_{0i}\,\Delta\varepsilon_i(x,y,z)\,\mathbf{E}^*_{\nu y}(x,z)\,[\mathbf{E}_{0y}(x,y,z)+\mathbf{E}_{\mathrm{w}}(x,y,z)]\exp(i\beta_y y)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z. \quad (10)$$

Для излучательных мод аналогично (8) имеем

$$\frac{\beta}{2\omega\mu_0}\iint \mathbf{E}^*_{\beta' y}(x,z)\mathbf{E}_{\beta y}(x,z)\,\mathrm{d}x\,\mathrm{d}z = P\,\delta(\beta'-\beta). \tag{11}$$

Тогда коэффициент разложения для мод рассеяния имеет вид

$$q(\beta,\beta_y) = -\frac{1}{2\pi\,(\beta^2-\beta_y^2)}\times \\ \times\iiint \omega^2\mu\varepsilon_{0i}\,\Delta\varepsilon_i(x,y,z)\,\mathbf{E}^*_{\beta y}(x,z)\,[\mathbf{E}_{0y}(x,y,z)+\mathbf{E}_{\mathrm{w}}(x,y,z)]\exp(i\beta_y y)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z. \tag{12}$$

Как отмечено выше, решение исходного неоднородного волнового уравнения ищется в виде свёртки некоторой трёхмерной функции Грина $G(x,y,z;x',y',z')$ с выражением для источника $\omega^2\mu\varepsilon_{0i}\,\Delta\varepsilon_i(x,y,z)\,[\mathbf{E}_{0y}(x,y,z)+\mathbf{E}_{\mathrm{w}}(x,y,z)]$ (см. уравнение (3)). Для того, чтобы найти явное выражение для функции Грина уравнения (3), необходимо подставить полученные выражения для коэффициентов разложения $C(\nu,\beta_y)$ и $q(\beta,\beta_y)$ в уравнение (4). После подстановки получаем трёхмерное векторное выражение для рассеянного поля (штрихи у функций и пределы интегрирования восстановлены):

$$\mathbf{E}_{\mathrm{s}}(x,y,z) = -\frac{1}{2\pi}\Bigg\{\sum_\nu\int\limits_{-\infty}^{+\infty}\mathrm{d}\beta_y\iiint\omega^2\mu\varepsilon_{0i}\,\Delta\varepsilon_i(x',y',z')\,\mathbf{E}^*_{\nu y}(x',z')\,[\mathbf{E}_{0y}(x',y',z')+\mathbf{E}_{\mathrm{w}}(x',y',z')]\times \\ \times\exp(i\beta_y y)\mathbf{E}_{\nu y}(x,z)\exp(-i\beta_y y')\,\frac{\mathrm{d}x'\,\mathrm{d}y'\,\mathrm{d}z'}{(\beta_\nu^2-\beta_y^2)} + \int\limits_{-\infty}^{+\infty}\mathrm{d}\beta_y\int\limits_{-\beta_3}^{+\beta_3}\mathrm{d}\beta\iiint\omega^2\mu\varepsilon_{0i}\,\Delta\varepsilon_i(x',y',z')\,\mathbf{E}^*_{\beta y}(x',z')\times \\ \times[\mathbf{E}_{0y}(x',y',z')+\mathbf{E}_{\mathrm{w}}(x',y',z')]\exp(i\beta_y y)\mathbf{E}_{\beta y}(x,z)\exp(-i\beta_y y')\,\frac{\mathrm{d}x'\,\mathrm{d}y'\,\mathrm{d}z'}{\beta^2-\beta_y^2}\Bigg\}. \tag{13}$$

Пределы интегрирования по поперечному сечению волновода (по x и y) и вдоль направления распространения волноводной моды (по z) не указаны, т. к. зависят от размеров конкретного волновода и трёхмерной области нерегулярности.

После группировки подобных членов в правой части уравнения (13) получаем следующее выражение для электрической составляющей рассеянного поля:

$$\mathbf{E}_{\mathrm{s}}(x,y,z) = -\frac{1}{2\pi}\Bigg\{\iiint\omega^2\mu\varepsilon_{0i}\,\Delta\varepsilon_i(x',y',z')\,[\mathbf{E}_{0y}(x',y',z')+\mathbf{E}_{\mathrm{w}}(x',y',z')]\exp(i\beta_y y)\,\mathrm{d}x'\,\mathrm{d}y'\,\mathrm{d}z'\times \\ \times\Bigg[\sum_\nu\int\limits_{-\infty}^{+\infty}\frac{\mathbf{E}^*_{\nu y}(x',z')\mathbf{E}_{\nu y}(x,z)}{\beta_\nu^2-\beta_y^2}\exp[-i\beta_y\,(y-y')]\,\mathrm{d}\beta_y + \\ + \int\limits_{-\infty}^{+\infty}\mathrm{d}\beta_y\int\limits_{-\beta_3}^{+\beta_3}\frac{\mathbf{E}^*_{\beta y}(x',z')\mathbf{E}_{\beta y}(x,z)}{\beta^2-\beta_y^2}\exp[-i\beta_y\,(y-y')]\,\mathrm{d}\beta\Bigg]\Bigg\}. \tag{14}$$

Сравнивая (14) с уравнением (3), можно записать явное выражение для искомой функции Грина $G(x,y,z;x',y',z')$ исходного неоднородного трёхмерного волнового уравнения (2):

$$G(x,y,z;x',y',z') = \frac{1}{2\pi}\left[\sum_{\nu}\int\limits_{-\infty}^{+\infty}\exp[-i\beta_y\,(y-y')]\,\frac{\mathbf{E}^{*}_{\nu y}(x',z')\mathbf{E}_{\nu y}(x,z)}{\beta_\nu^2-\beta_y^2}\,\mathrm{d}\beta_y + {}\right.$$
$$\left. + \int\limits_{-\infty}^{+\infty}\mathrm{d}\beta_y\int\limits_{-\beta_3}^{+\beta_3}\exp[-i\beta_y\,(y-y')]\,\frac{\mathbf{E}^{*}_{\beta y}(x',z')\mathbf{E}_{\beta y}(x,z)}{\beta^2-\beta_y^2}\,\mathrm{d}\beta\right]. \quad (15)$$

1.2. Анализ полученного выражения для функции Грина. Случаи распространяющихся и затухающих мод

Анализ полученного выражения для функции Грина (15) проведём с помощью диаграммы со спектрами постоянных распространения β направляемых мод и мод излучения нерегулярного несимметричного интегрального оптического волновода (рис. 2).

Диапазон возможных значений β_ν постоянных распространения направляемых мод дискретного спектра определяется неравенством $\beta_3 < |\beta_\nu| < \beta_2$, где $\beta_2 = kn_2$, $\beta_3 = kn_3$. Диапазон возможных значений β постоянных распространения мод излучения непрерывного спектра определяется неравенством $-\beta_3 < \beta < \beta_3$. Этот диапазон включает диапазон покровных (воздушных) мод излучения: $-\beta_1 < \beta < \beta_1$, и два диапазона подложко-покровных мод излучения: $-\beta_3 < \beta < -\beta_1$ и $\beta_1 < \beta < \beta_3$, где $\beta_1 = kn_1$. Отметим, что диапазон возможных значений β_ν для направляемых мод существует при действительном коэффициенте замедления волновода γ (эффективном показателе преломления волновода): $n_3 < \gamma < n_2$. Диапазон возможных значений β распространяющихся мод излучения существует при действительных поперечных волновых числах ρ. При действительных ρ существует также ещё один диапазон возможных постоянных распространения β, которые задаются мнимыми значениями $\beta_{\mathrm{im}} = -i\,|\beta|$. Этот диапазон β определяет затухающие моды нерегулярного волновода, которые описывают локальные поля вблизи источников излучения (в области нерегулярности). Используя их, можно описать поле излучения в ближней зоне.

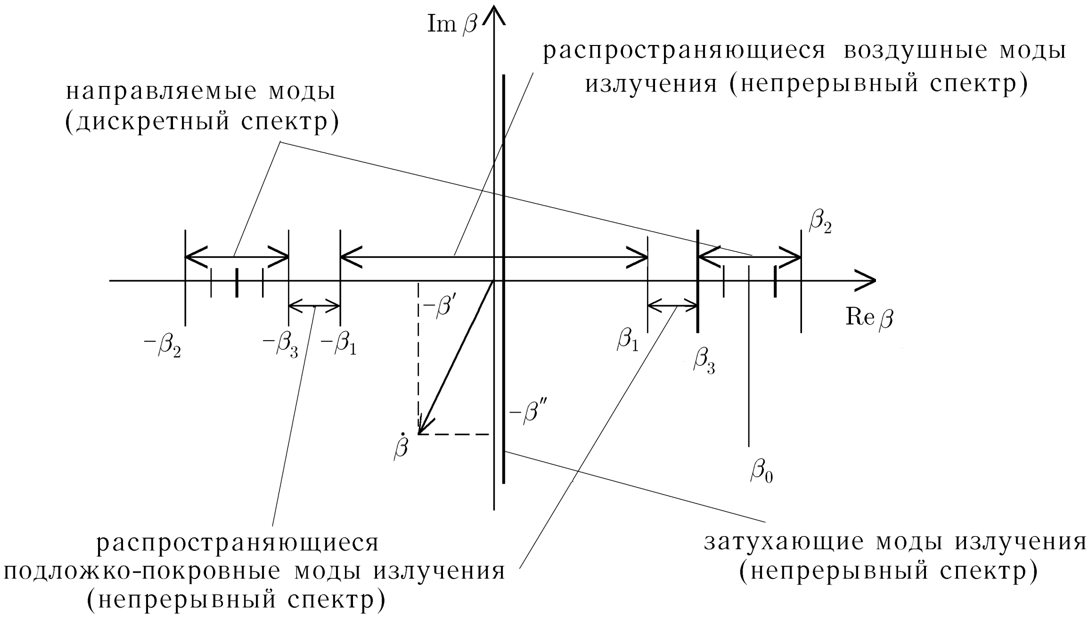

Рис. 2

В общем случае представим постоянную распространения мод излучения как некоторое комплексное число $\dot{\beta} = \beta' - i\beta''$, где β' — вещественная часть комплексного $\dot{\beta}$, т. е. то β, которое мы обычно рассматриваем, а $-\beta''$ — мнимая часть $\dot{\beta}$. При таком рассмотрении полная рассеянная мощность состоит из суммы мощностей, переносимых затухающими и распространяющимися модами излучения [9]. Обычно первую составляющую называют реактивной (мнимой) мощностью, а вторую — активной. Реактивная мощность убывает экспоненциально в направлении распространения затухающих мод (неоднородные волны) и слабо участвует в формировании диаграммы рассеяния в дальней зоне (в зоне дифракции Фраунгофера). Необходимо отметить, что при большом отношении уровней сигнала и шума и/или специальной регистрации и обработке данных рассеяния в дальней зоне можно извлечь некоторые данные о локальных полях вблизи источника, которые несут важнейшую информацию о субволновой структуре нерегулярностей [11].

С учётом вышесказанного приведём окончательные выражения для функции Грина (15) для двух теоретически и практически очень важных случаев: (1) распространяющихся и (2) затухающих мод излучения.

Для вычисления в (15) интегралов вида $\int_{-\infty}^{+\infty} \exp[-i\beta_y\,(y-y')]/(\beta^2-\beta_y^2)\,\mathrm{d}\beta_y$ в случае распространяющихся мод воспользуемся интегрированием по методу вычетов [12]:

$$2\pi i \sum_{m=1}^{2} \mathrm{Res}_m(\beta_y = \pm\beta) = -\pi i\,\frac{\exp(-i\beta\,|y-y'|)}{\beta}\,, \tag{16}$$

где для простоты дальнейших выкладок вторая экспонента опущена. Полученное значение интеграла (16) одинаково как при $y' > y$, так и $y > y'$, т. к. в показателе экспоненты разность $y - y'$ берётся по модулю (свойство симметрии функции Грина).

Теперь для случая распространяющихся мод излучения можно написать окончательное выражение для функции Грина $G(\mathbf{r};\mathbf{r}')$ исходного неоднородного волнового уравнения (2):

$$\begin{aligned} G(x,y,z;x',y',z') = i\Biggl\{\sum_{\nu} \exp[-i\beta_{\nu y}\,(y-y')]\,\frac{\mathbf{E}^*_{\nu y}(x',z')\mathbf{E}_{\nu y}(x,z)}{\beta_\nu} + {}\\ + \int\limits_{-\infty}^{+\infty} \exp[-i\beta_y\,(y-y')]\,\frac{\mathbf{E}^*_{\beta y}(x',z')\mathbf{E}_{\beta y}(x,z)}{\beta}\,\mathrm{d}\beta\Biggr\}, \end{aligned} \tag{17}$$

где сумма учитывает вклад в потери направляемой моды на межмодовое преобразование, а интеграл — вклад в потери направляемой моды на рассеяние в окружающее пространство (3D-scattering). Оба вида потерь связаны с нерегулярностью структуры рассматриваемого интегрального оптического волновода, а именно с неровностью его границ и неоднородностью показателя преломления образующих его сред. Важно отметить, что выражение для функции Грина (17) учитывает эффекты многократного рассеяния, а следовательно, учитывает нелокальную связь между падающей и рассеянными волнами.

С учётом полученного выражения для функции Грина электрическая составляющая поля излучения, обусловленного потерями направляемой моды на межмодовое преобразование и на рассеяние в окружающее трёхмерное пространство, вне слоя волновода принимает вид

$$\mathbf{E}_{\mathrm{s}}^{\mathrm{out}}(x,y,z) = -\frac{ik^2\bar{n}_i^2}{2}\int\limits_{-L_y/2}^{L_y/2}\mathrm{d}y'\int\limits_{-\infty}^{+\infty}\mathrm{d}x'\int\limits_{-\infty}^{+\infty}\mathrm{d}z'\times$$
$$\times\left\{\sum_{\nu}\exp[-i\beta_{\nu y}\,(y-y')]\,\frac{\mathbf{E}_{\nu y}^{*}(x',z')\mathbf{E}_{\nu y}(x,z)}{\beta_\nu}+\int\limits_{-\infty}^{+\infty}\exp[-i\beta_y\,(y-y')]\,\frac{\mathbf{E}_{\beta y}^{*}(x',z')\mathbf{E}_{\beta y}(x,z)}{\beta}\,\mathrm{d}\beta\right\}\times$$
$$\times\,\Delta n_i^2(x',y',z')\,[\mathbf{E}_{0y}(x',y',z')+\mathbf{E}_{\mathrm{w}}(x',y',z')]\exp(i\beta_y y),\quad(18)$$

где $\omega^2\mu_i\varepsilon_i = k^2\bar{n}_i^2$, $\Delta n_i^2 = \omega^2\mu_i\,\Delta\varepsilon_i$, $\bar{n}_i$ — среднее значение показателя преломления i-го слоя, $k=2\pi/\lambda$, L_y — протяжённость нерегулярного трёхмерного участка вдоль оси y.

В случае затухающих мод для анализа интегралов вида $\int_{-\infty}^{+\infty}\exp[-i\beta_y\,(y-y')]/(\beta^2-\beta_y^2)\,\mathrm{d}\beta_y$ в полученном выражении для функции Грина (15) необходимо рассмотреть диапазон постоянных распространения β, которые задаются соотношением $\beta_{\mathrm{im}}=-i\,|\beta|$. Как указано выше, этот диапазон значений β определяет затухающие моды нерегулярного волновода. В этом случае подынтегральные выражения также имеют два простых полюса [12], но при $\beta_y=\pm i\beta_\nu$ и $\beta_y=\pm i\beta$. Для вычисления интегралов в качестве контура интегрирования взята совокупность двух отдельных замкнутых контуров, каждый из которых содержит один из рассматриваемых простых полюсов $\beta_y=\pm i\beta''$:

$$2\pi i\sum_{m=1}^{2}\mathrm{Res}_m(\beta_y=\pm i\beta)=\frac{\pi}{\beta}\left[\exp(\beta\,|y-y'|)-\exp(-\beta\,|y-y'|)\right].\quad(19)$$

В результате для случая затухающих мод излучения можно записать окончательное выражение для функции Грина $G(\mathbf{r};\mathbf{r}')$ исходного неоднородного волнового уравнения (2):

$$G(x,y,z;x',y',z')=\frac{\pi}{\beta}\left\{\sum_{\nu}\Big[\exp(\beta_{\nu y}\,|y-y'|)-\exp(-\beta_{\nu y}\,|y-y'|)\Big]\,\frac{\mathbf{E}_{\nu y}^{*}(x',z')\mathbf{E}_{\nu y}(x,z)}{\beta_\nu}+\right.$$
$$\left.+\int\limits_{-\infty}^{+\infty}\Big[\exp(\beta_y\,|y-y'|)-\exp(-\beta_y\,|y-y'|)\Big]\,\frac{\mathbf{E}_{\beta y}^{*}(x',z')\mathbf{E}_{\beta y}(x,z)}{\beta}\,\mathrm{d}\beta\right\},\quad(20)$$

где, как и в (17), сумма учитывает вклад в потери направляемой моды на межмодовое преобразование, а интеграл — вклад в потери направляемой моды на рассеяние в окружающее пространство (трёхмерное рассеяние), но в ближней зоне. С учётом полученного выражения (20) для функции Грина поле излучения, обусловленное потерями направляемой моды на межмодовое преобразование и на рассеяние в окружающее пространство в ближней зоне, соответствующее экспоненциально затухающим в направлении распространения волнам, принимает вид

$$\mathbf{E}_{\mathrm{s}}^{\mathrm{nz}}(x,y,z)=-\frac{\pi i}{2\beta}\int\limits_{-L_y/2}^{L_y/2}\mathrm{d}y'\int\limits_{-L_x/2}^{L_x/2}\mathrm{d}x'\int\limits_{-L_z/2}^{L_z/2}\mathrm{d}z'\times$$
$$\times\left\{\sum_{\nu}\big[-\exp(-\beta_{\nu y}\,|y-y'|)\big]\,\frac{\mathbf{E}_{\nu y}^{*}(x',z')\mathbf{E}_{\nu y}(x,z)}{\beta_\nu}-\int\limits_{-\infty}^{+\infty}\exp(-\beta_y\,|y-y'|)\,\frac{\mathbf{E}_{\beta y}^{*}(x',z')\mathbf{E}_{\beta y}(x,z)}{\beta}\,\mathrm{d}\beta\right\}\times$$
$$\times\,k^2\bar{n}_i^2\,\Delta n_i^2(x',y',z')\big[\mathbf{E}_{0y}(x',y',z')+\mathbf{E}_{\mathrm{w}}(x',y',z')\big]\exp(i\beta_y y),\quad(21)$$

где L_x, L_y и L_z — протяжённость локальной трёхмерной нерегулярности в направлении осей x, y и z соответственно.

При рассмотрении поля излучения в ближней, промежуточной и дальней зонах в (18) и (21) необходимо взять только второе слагаемое под знаком тройного интеграла:

$$\mathbf{E}_{\mathrm{s}}^{\mathrm{out}}(x,y,z) = -\frac{ik^2\bar{n}_i^2}{2}\int\limits_{-L_y/2}^{L_y/2}\mathrm{d}y'\int\limits_{-\infty}^{+\infty}\mathrm{d}x'\int\limits_{-\infty}^{+\infty}\mathrm{d}z'\int\limits_{-\infty}^{+\infty}\exp[-i\,(\beta_{0y}-\beta)\,y']\;\frac{\mathbf{E}^*_{\beta y}(x',z')\mathbf{E}_{\beta y}(x,z)}{\beta}\times$$
$$\times\,\Delta n_i^2(x',y',z')\big[\mathbf{E}_0(x',z')+\mathbf{E}_{0\mathrm{w}}(x',z')\big]\exp(-i\beta_y y)\,\mathrm{d}\beta, \quad (22)$$

где $\beta_{0y} = k_y\gamma = kn_2\sin\varphi_{0y}$ — модуль продольной составляющей $\boldsymbol{\beta}_{0y}$ вектора распространения $\mathbf{k}_y n_2$ волноводной моды вдоль оси y. В (22) учтено, что $\mathbf{E}_{0y}(x',y',z') = \mathbf{E}_0(x',z')\exp(-i\beta_{0y}y')$, $\mathbf{E}_{\mathrm{w}}(x',y',z') = \mathbf{E}_{0\mathrm{w}}(x',z')\exp(-i\beta_{0y}y')$. Заметим, что учёт опущенной в (16) второй экспоненты даёт в (22) вместо $\exp(-i\beta_y y)$ множитель вида $\sin(\beta_y y)/(\beta_y y)$.

Аналогично (22) получаем для поля излучения в ближней зоне

$$\mathbf{E}_{\mathrm{s}}^{\mathrm{nz}}(x,y,z) = -\frac{\pi ik^2\bar{n}_i^2}{2\beta}\int\limits_{-L_y/2}^{L_y/2}\mathrm{d}y'\int\limits_{-L_x/2}^{+L_x/2}\mathrm{d}x'\int\limits_{-L_z/2}^{+L_z/2}\mathrm{d}z'\Bigg\{-\int\limits_{-\infty}^{+\infty}\exp(-\beta_y\,|y-y'|)\times$$
$$\times\,\frac{\mathbf{E}^*_{\beta y}(x',z')\mathbf{E}_{\beta y}(x,z)}{\beta}\,\mathrm{d}\beta\Bigg\}\,\Delta n_i^2(x',y',z')\big[\mathbf{E}_{0y}(x',y',z')+\mathbf{E}_{\mathrm{w}}(x',y',z')\big]\exp(i\beta_y y). \quad (23)$$

Найденные выражения для полей излучения позволяют решить на данном этапе, по крайней мере, прямую задачу теории векторного рассеяния света в нерегулярном интегрально-оптическом волноводе в ближней, промежуточной и дальней зонах. Отметим, что амплитудно-фазовые, т. е. комплексные, выражения для полей рассеянного в волноводе излучения позволяют оценить фазовое распределение, характеризующее весь ансамбль нерегулярностей в целом. Для измерения комплексного сигнала могут быть использованы стандартные методы интерферометрии (использование опорного лазерного сигнала), позволяющие определить действительную и мнимую части вектора $\mathbf{E}_{\mathrm{s}}$, а затем найти его фазу. Для решения обратной задачи волноводного рассеяния света [1–5] может быть использован, например, метод сравнения, который наиболее часто применяется экспериментаторами для определения физических параметров исследуемого явления по результатам измерения его проявлений [10]. Действительно, в этом случае обратная задача теории волноводного рассеяния решается наиболее просто: измеренная диаграмма рассеяния сравнивается с теоретически рассчитанными диаграммами рассеяния для некоторой заданной нерегулярности $\Delta n_i^2(x',y',z')$, и из них подбирается диаграмма, наиболее близкая (например, в среднеквадратичной метрике) к измеренной. В результате определяются параметры нерегулярностей. В этом случае $\Delta n_i^2(x',y',z')$ предварительно оценивается из априорных данных независимых исследований. В общем случае нерегулярности могут представлять собой как неровности границ раздела сред волновода, так и неоднородности волноводного слоя [1–9]. Строгое решение обратной задачи теории волноводного рассеяния в нерегулярном интегрально-оптическом волноводе выходит за рамки данной работы и будет дано в одной из наших следующих работ (см., например, [13, 14]). Действительно, обратная задача волноводного рассеяния света относится к классу некорректно [2]

[2] Напомним, что задача называется корректно поставленной, если выполняются следующие требования: 1) её решение существует, 2) это решение определяется однозначно, 3) решение непрерывно зависит от входных данных, т. е. малым изменениям входных данных соответствуют малые изменения в решении задачи (свойство устойчивости решения). Задачи, не удовлетворяющие данным требованиям, называются некорректно поставленными.

поставленных задач [1–5, 10, 11], причём корректность решения обратной задачи теории волноводного рассеяния света до сих пор подробно не исследовалась. Наше решение предполагает использование комплексного алгоритма восстановления автокорреляционной функции нерегулярностей, основанного на комбинации классической регуляризации и квазиоптимальной фильтрации. Приближённое корректное решение обратной задачи волноводного рассеяния света удаётся получить в классе аналитических функций [13].

В заключение данного раздела заметим, что для анализа полученных в работе выражений, описывающих теорию векторного волноводного рассеяния света, необходимо использовать компьютерные методы трёхмерного векторного моделирования, требующие разработки сложных программ. Несомненно, в перспективе это позволит проводить исследование рассеяния в нерегулярных интегральных оптических волноводах и устройствах на их основе с произвольной трёхмерной геометрией. Очевидно, что решение этой задачи имеет фундаментальное значение для развития субмикронных технологий в интегральной оптике и волноводной оптоэлектронике.

2. ЗАКЛЮЧЕНИЕ

В настоящей работе развита векторная теория волноводного рассеяния монохроматического света на трёхмерных нерегулярностях при наличии шума. Получены выражения для электрических составляющих полей рассеянного излучения при наличии шума. Их вывод основан на последовательном решении неоднородного волнового трёхмерного уравнения. Проведён анализ полученных формул. Развитая в настоящей работе векторная теория волноводного рассеяния лазерного излучения на трёхмерных нерегулярностях при наличии шума может стать основой для построения полной векторной теории волноводного рассеяния света в интегрально-оптическом волноводе с произвольной трёхмерной топологией.

СПИСОК ЛИТЕРАТУРЫ

1. Егоров А. А. // Квантовая электроника. 2002. Т. 32, № 4. С. 357.
2. Егоров А. А. // Известия вузов. Радиофизика. 2002. Т. 45, № 7. С. 577.
3. Егоров А. А. // Квантовая электроника. 2003. Т. 33, № 4. С. 335.
4. Егоров А. А. // Оптика и спектроскопия. 2003. Т. 95, № 2. С. 294.
5. Yegorov A. A. // Laser Physics. 2003. V. 13, No. 9. P. 1143.
6. Hall D. G. // Optics Letters. 1981. V. 6, No. 12. P. 601.
7. Маркузе Д. Оптические волноводы. М.: Мир, 1974.
8. Содха М. С., Гхатак А. К. Неоднородные оптические волноводы. М.: Связь, 1980.
9. Снайдер А., Лав Дж. Теория оптических волноводов. М.: Радио и связь, 1987.
10. Арсенин В. Я. Методы математической физики и специальные функции. М.: Наука, 1974.
11. Обратные задачи в оптике / Под ред. Г. П. Болтса. М.: Машиностроение, 1984.
12. Маркушевич А. И. Краткий курс теории аналитических функций. М.: Наука, 1978.
13. Egorov A. A. // Laser Physics. 2004. V. 14, No. 8. P. 1299.
14. Egorov A. A. // Laser Physics Letters. 2005. V. 2, No. 2. P. 77.

Поступила в редакцию 8 декабря 2003 г.; принята в печать 16 сентября 2004 г.

THEORY OF WAVEGUIDE LIGHT SCATTERING IN AN INTEGRATED OPTICAL WAVEGUIDE IN THE PRESENCE OF NOISE

A. A. Egorov

We develop a vector theory of waveguide monochromatic light scattering in an integrated optical waveguide with arbitrary three-dimensional irregularities in the presence of noise. The electrodynamical problem of laser radiation scattering in an irregular waveguide is solved by the method of coupled modes with the help of the theory of perturbations. The approximate solution of inhomogeneous three-dimensional wave equation is derived by the method of modes and the method of Green's functions. The obtained Green's function is analyzed for the cases of propagating and evanescent modes of an irregular asymmetric optical waveguide. Analytical formulas for the near- and far-zone radiation fields are presented. Preliminary analysis of these formulas is given.

УДК 519.23

ОЦЕНКИ ДИСПЕРСИИ ГАУССОВОГО ПРОЦЕССА, УСТОЙЧИВЫЕ К ИМПУЛЬСНОЙ ПОМЕХЕ

А. А. Родионов, В. И. Турчин

Институт прикладной физики РАН, г. Нижний Новгород, Россия

Применительно к пассивной локации предложен метод оценки средней мощности гауссового сигнала, устойчивый к импульсной помехе. В основе метода лежит следующее соображение: при определённых условиях гистограмма отсчётов сигнала, «загрязнённого» импульсной помехой, близка к нормальному распределению в определённом интервале значений сигнала, т. е. гистограмма является оценкой, мало чувствительной к наличию импульсных загрязнений. Соответственно, для оценки средней мощности «полезной» гауссовой компоненты предложена модификация известного метода минимума χ^2 (в статье она названа методом устойчивой аппроксимации). Показано, что предложенная оценка даёт более высокую точность, например, по сравнению с медианными оценками и близка к границе Крамера—Рао. Приведены результаты экспериментальной апробации метода на гидроакустических данных.

ВВЕДЕНИЕ

В пассивной локации (см., например, [1]) существует большой класс задач, заключающихся в измерении средней мощности случайного сигнала. Такие задачи характерны, например, для гидроакустики или радиоастрономии. Часто считается, что принимаемый сигнал представляет собой белый гауссовый процесс с нулевым средним. Хотя, как правило, его средняя мощность во время измерений меняется [1], обычно можно разбить процесс наблюдения на временны́е интервалы длительности T, внутри которых изменением дисперсии (средней мощности) сигнала можно пренебречь. Оценка $\hat{\sigma}^2$ дисперсии σ^2 внутри такого интервала строится общеизвестным способом:

$$\hat{\sigma}^2 = J^{-1} \sum_{j=1}^{J} x_j^2, \tag{1}$$

где x_j — отсчёт вещественного сигнала в j-й момент времени, $J/F_\mathrm{s} = T$, F_s — частота дискретизации. Дисперсия этой оценки, т. е. точность измерения, определяется известным выражением

$$\mathrm{disp}[\hat{\sigma}^2] = 2\sigma^4/J. \tag{2}$$

Известно, что оценка (1) является неулучшаемой, т. е. (2) совпадает с границей Крамера—Рао [2].

На практике, однако, наряду с гауссовой составляющей (такой, например, как смесь шума источника и шумов моря в гидроакустике или смесь шума внеземного радиоисточника и шумов аппаратуры, атмосферы и т. д. в радиоастрономии), в отсчётах сигнала может присутствовать аддитивная импульсная помеха. Происхождение импульсных выбросов может быть различным: от сбоев в аппаратуре до шумов биологического происхождения в гидроакустике, импульсные помехи от радиолокационных станций, искровых разрядов и т. д. в радиоастрономии. В силу разнообразия природы импульсной помехи, как правило, её статистические характеристики весьма неопределённы, и построение детальной априорной статистической модели импульсной помехи невозможно. Очевидно также, что использование (1) при наличии импульсных помех со средней

[1] Например, при сканировании внеземного радиоисточника диаграммой направленности радиотелескопа или при перемещении источника шума относительно приёмника в акустике.

мощностью, сопоставимой с мощностью случайного сигнала, приведёт к существенному смещению оценки, поэтому необходима разработка специальных процедур оценки дисперсии гауссовой компоненты при наличии импульсной помехи. Этот вопрос и является предметом исследования данной работы.

Традиционным способом борьбы с импульсными помехами является клиппирование — «вырезание» участков записи, на которых присутствовала импульсная помеха. Обнаружение таких участков возможно за счёт использования порогового устройства, реагирующего на большие выбросы. Для выбора порога, однако, необходимо априори знать (хотя бы приближённо) дисперсию гауссовой компоненты, что противоречит изначальной постановке задачи. Естественно, анализируя отдельные участки записи внутри интервала T, можно с помощью многошаговой процедуры постепенно найти приближённое значение порога, однако такая процедура является достаточно громоздкой (см., например, [3]); кроме того, её статистическое обоснование весьма трудоёмко.

Другим распространённым методом подавления импульсной помехи является медианная фильтрация. Один из способов её практического использования при оценке дисперсии заключается в следующем: J отсчётов из интервала T разбиваются на K групп по M отсчётов ($KM = J$), после чего путём некогерентного усреднения отсчётов в каждой группе получаются новые K отсчётов $y_k = M^{-1}\sum_{i=1}^{M} x^2_{(k-1)M+i}$. За оценку дисперсии принимается выборочная медиана последовательности y_k. Величины y_k, как известно, в отсутствие импульсной помехи имеют χ^2-распределение с M степенями свободы, которое уже при небольших M близко к нормальному распределению со средним значением σ^2. Известно, что в отсутствие импульсной помехи оценка такого вида является несмещённой и имеет дисперсию, превышающую (2) приблизительно в $\pi/2$ раз [4], что может оказаться разумной платой за избавление от импульсной компоненты. Другим известным способом является оценка с помощью так называемого абсолютного медианного отклонения [5, 6]. Очевидно, что в отсутствие импульсной помехи дисперсия σ^2 величин x_j и медиана m величин $y_j = |x_j|$ связаны однозначно: $\sigma^2 = \alpha m^2$, где численный коэффициент $\alpha \approx 2{,}2$ находится из уравнения $\mathrm{erf}(1/\sqrt{2\alpha}) = 1/2$, в котором $\mathrm{erf}(x)$ — интеграл вероятности. Оба способа имеют определённые недостатки. В первом случае выбор длины группы зависит от средней частоты следования импульсной помехи (очевидно, что среди элементов последовательности должно найтись достаточно много результатов текущего усреднения, не затронутых импульсной помехой). Во втором случае, строго говоря, при наличии импульсной помехи распределение отсчётов x_j уже не будет нормальным, и коэффициент α будет отличаться от приведённого выше значения.

Одной из принципиальных трудностей решения поставленной задачи является негауссовость распределения отсчётов сигнала, причём трудно рассчитывать на априорное знание плотности вероятности импульсной помехи. Поэтому искомая оценка дисперсии гауссовой компоненты должна быть робастной по отношению к этой неизвестной плотности вероятности. Известно достаточно много подходов к построению таких оценок. Например, для оценки параметра сдвига некоторого симметричного распределения с целью повышения устойчивости получаемых оценок можно вместо выборочного среднего применять различные линейные комбинации порядковых статистик, так называемые L-оценки (см. [7]; к этому классу относится, в частности, оценка среднего, использованная в первом варианте медианной оценки дисперсии). Наряду с L-оценками используются так называемые R- и M-оценки. Примером последней является как тривиальная оценка (1), так и второй вариант оценки, использующей выборочную медиану. Более общим видом M-оценок является оценка Хьюбера параметра масштаба σ^2 некоторого распределения:

$$\hat{\sigma}^2 = \frac{1}{e(\alpha)} \arg\min_{\sigma^2} \left| \sum_{j=1}^{J} \psi(x_j/\sigma) \right|, \tag{3}$$

где

$$\psi(x) = \begin{cases} x^2 - 1, & |x| \le \alpha; \\ \alpha^2 - 1, & |x| > \alpha, \end{cases}$$

$e(\alpha)$ устраняет смещение оценки. Для различных α значения $e(\alpha)$ приведены в [5]. Отметим также классы A-, D-, P-, S- и W-оценок, рассмотренные в [6].

Цель нашей работы заключается в построении достаточно простой и устойчивой по отношению к неизвестным характеристикам импульсной помехи оценки дисперсии гауссовой компоненты и сопоставлении характеристик такой оценки с оценками, приведёнными выше.

1. МОДЕЛЬ ИМПУЛЬСНОЙ ПОМЕХИ И ОЦЕНКА ДИСПЕРСИИ ГАУССОВОЙ КОМПОНЕНТЫ ТИПА D-КЛАССА

Как отмечалось выше, форма, длительность и другие характерные параметры импульсов помехи априори неизвестны, поэтому в данной работе мы ограничимся статистическим описанием наблюдаемого процесса одномерной плотностью распределения вероятности. Обычно для этих целей используется модель Тьюки (см., например, [7]). Суть её заключается в том, что плотность распределения вероятности $w_1(x)$ «полезного» (в данном случае — гауссового) процесса в каждый момент времени с вероятностью p замещается некоторой «загрязняющей» плотностью распределения вероятности $w_2(x)$. При этом результирующая плотность распределения вероятности записывается как

$$w(x) = w_1(x)\,(1-p) + w_2(x)p. \tag{4}$$

Свойства помехи в модели (4) определяются вероятностью её появления p и видом (параметрами) одномерной плотности распределения вероятности $w_2(x)$. Типичным примером конкретного вида (4) является модель Лихтера (см., например, [8]) для огибающей комплекснозначного сигнала, в которой в качестве $w_1(x)$ берётся распределение Рэлея, а в качестве $w_2(x)$ — распределение Райса. В данной работе в качестве отправной точки удобно рассматривать нормальную плотность распределения вероятности помехи $w_2(x)$ с нулевым средним и дисперсией σ_p^2, при этом

$$w(x,\sigma^2,\sigma_p^2,p) = \frac{1}{\sqrt{2\pi}}\left[\frac{1-p}{\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right) + \frac{p}{\sigma_p}\exp\left(-\frac{x^2}{2\sigma_p^2}\right)\right]. \tag{5}$$

Отметим, что в дальнейшем вид $w_2(x)$ не будет играть особой роли, так что (5) следует рассматривать как некоторый конкретный пример.

Поскольку в соответствии с (4) мы ограничились описанием смеси гауссовой компоненты и импульсной помехи одноточечным распределением, оценку дисперсии гауссовой компоненты можно искать, ориентируясь на приближение эмпирической плотности распределения вероятности (гистограммы) к (4); заметим, что такого рода оценки принадлежат к классу D-оценок. Известным примером из этого класса является метод минимума χ^2 [2, 9, 10], в котором оценка $\hat{\boldsymbol{\theta}}$ вектора неизвестных параметров плотности распределения вероятности $\boldsymbol{\theta}$ находится путём минимизации взвешенного среднеквадратичного отклонения (СКО):

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \Phi(\boldsymbol{\theta}), \qquad \Phi(\boldsymbol{\theta}) = \sum_{k=1}^{K} g_k [N_k - E_k(\boldsymbol{\theta})]^2. \tag{6}$$

Для получения (6) область изменения значений сигнала разбивается на K неперекрывающихся интервалов (a_k, b_k); N_k представляет число отсчётов сигнала, попавших в k-й интервал, $E_k(\boldsymbol{\theta}) = J\int_{a_k}^{b_k} w(x,\boldsymbol{\theta})\,\mathrm{d}x$ — ожидаемое число попаданий в k-й интервал. Весовые коэффициенты g_k

берутся в виде $g_k \sim 1/E_k(\boldsymbol{\theta})$ или $g_k \sim 1/N_k$. Как известно, оценки метода минимума χ^2 являются асимптотически эффективными [9], т. е. при достаточно большом числе отсчётов J обеспечивают наилучшую точность. Недостатком процедуры (6) является использование явного вида плотности распределения вероятности импульсной помехи w_2. Суть предлагаемого метода заключается в устранении этой трудности.

Разумным для модели (5) является предположение о том, что σ_p^2, по крайней мере, в несколько раз больше, чем σ^2, а p невелико. В дальнейшем будем строить оценку в рамках этого условия, которое нередко выполняется на практике. Нетрудно заметить, что при сделанном предположении для весьма широкого класса симметричных «колоколообразных» распределений $w_2(x)$ с параметром масштаба σ_p внутри интервала $[-x_0, x_0]$, где $x_0 = C\sigma$, C — численная константа порядка нескольких единиц, $pw_2 \sim p/\sigma_p$, в то время как максимальное значение $(1-p)\, w_1$ при $x = 0$ будет равно $(2\pi)^{-1/2}\,(1-p)/\sigma$. Тем самым внутри указанного интервала $[-x_0, x_0]$ гауссовая компонента будет существенно превалировать в случае малости параметра γ:

$$\gamma = \frac{\sigma p}{\sigma_p\,(1-p)} \ll 1. \tag{7}$$

В качестве иллюстрации этого факта на рис. 1 сплошной линией показана плотность распределения вероятности (5) при $\sigma = 1$; $\sigma_p = 5$; $p = 0{,}1$ ($\gamma \approx 0{,}02$), пунктиром — при $p = 0$. Вертикальные прямые обозначают границы интервала $[-x_0, x_0]$. Рис. 1*а* дан в логарифмическом масштабе, рис. 1*б* — в линейном. В логарифмическом масштабе хорошо заметно отличие w от гауссовой компоненты, в то время как в линейном масштабе это отличие несущественно.

Таким образом, при малом γ внутри интервала $[-x_0, x_0]$ результирующая плотность распределения вероятности w почти не отличается от гауссовой, т. е., по сути, внутри интервала $[-x_0, x_0]$ гистограмма при $\gamma \ll 1$ представляет собой статистику, мало чувствительную к импульсным выбросам. Соответственно, в (6) можно положить

$$E_k(\boldsymbol{\theta}) = E_k^{(n)}(\sigma^2), \tag{8}$$

где $E_k^{(n)}$ — среднее число попаданий отсчётов нормального процесса с дисперсией σ^2, играющей роль неизвестного параметра $\boldsymbol{\theta}$, в интервал (a_k, b_k), либо, считая плотность распределения ве-

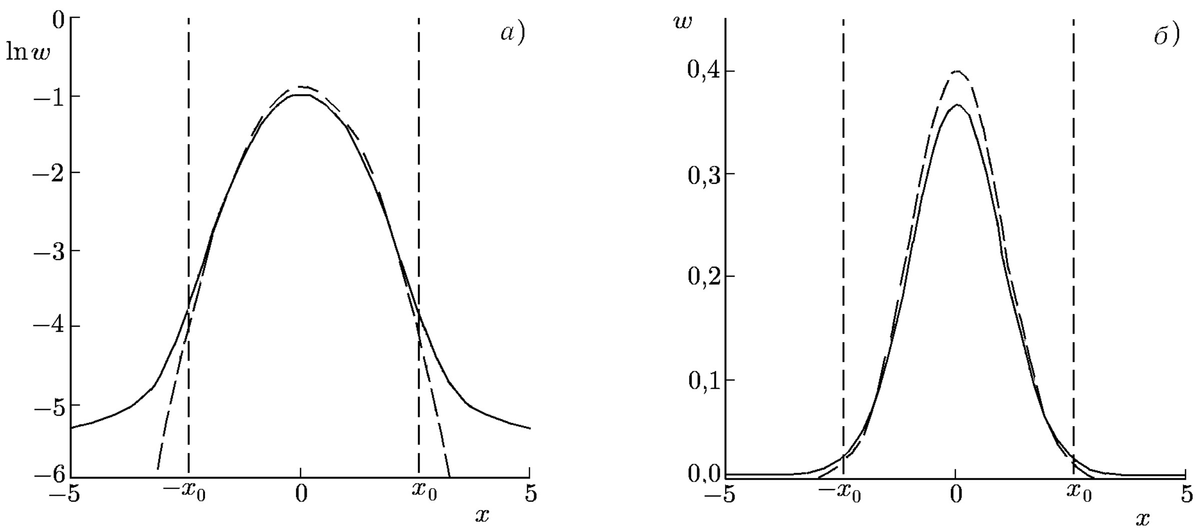

Рис. 1. Сплошная линия — плотность распределения вероятности (5) при $\gamma \approx 0{,}02$, штриховая линия — при $p = 0$

роятности w_2 приблизительно постоянной внутри $[-x_0, x_0]$, использовать приближение

$$E_k(\boldsymbol{\theta}) = \theta_0 E_k^{(n)}(\sigma^2) + \theta_1, \tag{9}$$

где θ_0, θ_1 — дополнительные неизвестные параметры, характеризующие p и σ_p. Минимум (6) по этим параметрам находится аналитически, после чего оценка дисперсии приобретает вид

$$\hat{\sigma}^2 = \arg\max_{\sigma^2} \frac{A_{11}B_2^2 - 2A_{12}B_1B_2 + A_{22}B_1^2}{A_{11}A_{22} - A_{12}^2}\ , \tag{10}$$

где $A_{11} = \mathbf{a}^{\mathrm{T}}\mathbf{G}\mathbf{a}$, $A_{12} = \mathbf{a}^{\mathrm{T}}\mathbf{G}\mathbf{e}$, $A_{22} = \mathbf{e}^{\mathrm{T}}\mathbf{G}\mathbf{e}$, $B_1 = \mathbf{a}^{\mathrm{T}}\mathbf{G}\mathbf{n}$, $B_2 = \mathbf{e}^{\mathrm{T}}\mathbf{G}\mathbf{n}$, $\mathbf{a} = (1, \ldots, 1)^{\mathrm{T}}$, $\mathbf{e} = (E_1^{(n)}, \ldots, E_K^{(n)})^{\mathrm{T}}$, $\mathbf{n} = (N_1, \ldots, N_K)^{\mathrm{T}}$; $\mathbf{G} = \mathrm{diag}(g_1, \ldots, g_K)$ — диагональная матрица, составленная из весов g_k, входящих в (6); индекс T означает транспонирование.

Как следует из вышеизложенного, существенную роль в предлагаемом методе играет интервал $[-x_0, x_0]$, границы которого определяются как $x_0 = C\sigma$. Численная константа C должна удовлетворять двум противоречивым требованиям. С одной стороны, с увеличением C повышается точность оценки дисперсии гауссовой компоненты. С другой стороны, при расширении интервала $[-x_0, x_0]$ вклад «загрязняющего» распределения увеличивается, что приводит к снижению точности оценки. Численное моделирование показало, что дисперсия оценки минимальна и слабо меняется, если константа находится в достаточно широких пределах: $2 \lesssim C \lesssim 4$. Таким образом, для первоначального определения x_0 достаточно выполнить грубую оценку σ^2. Для этого может быть использован (*а*) метод абсолютного медианного отклонения или (*б*) та же самая оценка в рамках метода минимума χ^2 для интервала $[\min\{x_j\}, \max\{x_j\}]$, но с постоянными весами g_k, что увеличивает робастность оценки по отношению к априорно неизвестному распределению $w_2(x)$. [2]

Таким образом, суть предложенного метода определения дисперсии гауссового «полезного» сигнала на фоне импульсной помехи сводится к использованию модифицированного метода минимума χ^2. Модификация сводится к тому, что

1) в качестве модельной плотности распределения вероятности берётся нормальное распределение (8) либо распределение (9) с двумя дополнительными неизвестными параметрами; в последнем случае оценка строится в соответствии с (10);

2) приближение распределения с неизвестным параметром к гистограмме проводится на интервале $[-x_0, x_0]$, границы которого находятся по первоначальной грубой оценке дисперсии.

Поскольку, как было показано выше, гистограмма в определённой области весьма устойчива к наличию импульсных помех, предложенный способ оценки будем в дальнейшем именовать методом устойчивой аппроксимации. Как отмечалось выше, метод минимума χ^2 обладает свойством асимптотической эффективности, поэтому можно ожидать, что точность (дисперсия) оценки будет достаточно высокой; свойства предложенных оценок анализируются в следующем разделе.

В заключение данного раздела остановимся на выборе интервалов разбиения области значений сигнала — длине интервалов (a_k, b_k). Будем считать, что их длина одинакова, а область значений ограничена экстремальными значениями в выборке. Как можно предположить, слишком большая длина интервалов сильно ухудшит свойства оценки (6) или (10), да и любой оценки, использующей гистограмму. Интуитивно понятно, что интервалы должны быть достаточно малы, чтобы гистограмма при увеличении объёма выборки всё больше «напоминала» по виду теоретическую плотность распределения вероятности, адекватно отображая все её особенности. Для выяснения зависимости СКО (6) от количества интервалов разбиения было проведено численное

[2] Нетрудно показать, что при использовании постоянных весов дисперсия оценки (6) в случае гауссового процесса увеличивается в 2,74 раза, однако при этом мы избегаем сложностей, связанных с большими весами в (6), появляющимися при больших значениях сигнала.

моделирование. На рис. 2 приведён график зависимости СКО оценки дисперсии по методу минимума χ^2 от числа интервалов K для нормальной плотности распределения вероятности. СКО на рис. 2 представлена в децибелах по отношению к (2).

Как видно из рис. 2, начиная с некоторого $K > K_0$ можно пренебречь отличием СКО оценки (6) от (2). Можно предположить, что K_0 связано с величиной дисперсии. Эмпирическим путём было получено, что с определённым запасом можно принять $K_0 \approx 4\Delta/\sigma$, где Δ — разница между максимальным и минимальным значениями выборки, т. е. длина интервала равна $\sigma/4$. Поскольку высокая точность при выборе K не требуется, для оценки дисперсии можно воспользоваться абсолютным медианным отклонением, соответствующая оценка является в определённом смысле наиболее устойчивой оценкой дисперсии [5, 6].

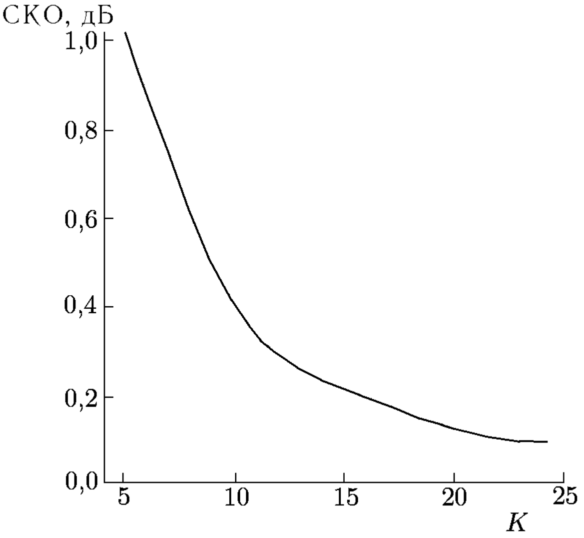

Рис. 2. Зависимость СКО (6) от числа интервалов K

2. СРАВНИТЕЛЬНЫЙ АНАЛИЗ ТОЧНОСТИ ОЦЕНОК

В данном разделе мы ограничимся исследованием свойств оценок в зависимости от дисперсии помехи σ_p^2 и вероятности p для модели Тьюки (5) с нормальной плотностью распределения вероятности $w_2(x)$. Очевидно, что полученные результаты будут, по крайней мере, качественно справедливы и для других «похожих» распределений $w_2(x)$ с параметром масштаба σ_p.

Для заданного вида $w(x)$ с неизвестными параметрами σ^2, σ_p^2, p может быть найдена граница Крамера—Рао: минимальная граница дисперсии для параметра σ^2, которая представляет собой предельную достижимую дисперсию оценки. С этой границей мы и будем сравнивать дисперсии оценок, полученных разными методами. Соответствующие графики отношений $D_{\text{мгд}}/D_0$ в зависимости от σ_p/σ и p приведены на рис. 3; здесь $D_{\text{мгд}}$ — минимальная граница дисперсии оценки σ^2, D_0 — дисперсия (2) тривиальной оценки (1) при отсутствии импульсной помехи.

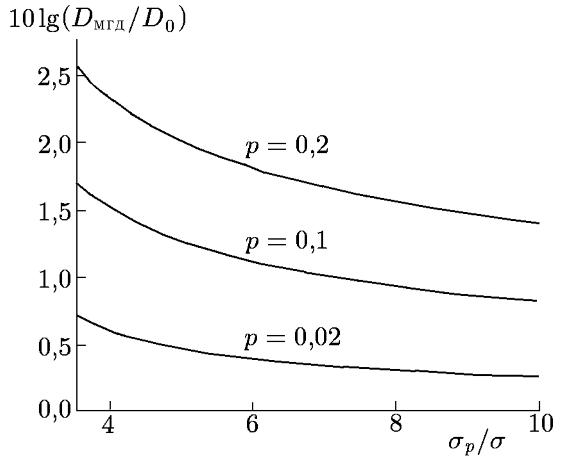

Рис. 3. Граница Крамера—Рао для параметра σ^2 модели (5) в зависимости от параметров σ_p и p

Как видно из рис. 3, зависимость $D_{\text{мгд}}$ монотонна по обоим параметрам. С увеличением вероятности появления выброса $D_{\text{мгд}}$ возрастает, а с увеличением амплитуды выбросов — уменьшается. Такое поведение соответствует интуитивным представлениям о том, что чем ближе помеха по амплитуде к сигналу и чем чаще она появляется, тем сложнее оценить параметры «полезного» сигнала на фоне помехи. Следует, однако, заметить, что введённая таким способом минимальная граница дисперсии теряет смысл при $\sigma_p \to \sigma$, т. к. при этом возникает неопределённость в оценке вероятности p появления отсутствующей помехи; математически это выражается в том, что минимальная граница дисперсии для σ^2 обращается в бесконечность. Поэтому корректно проводить

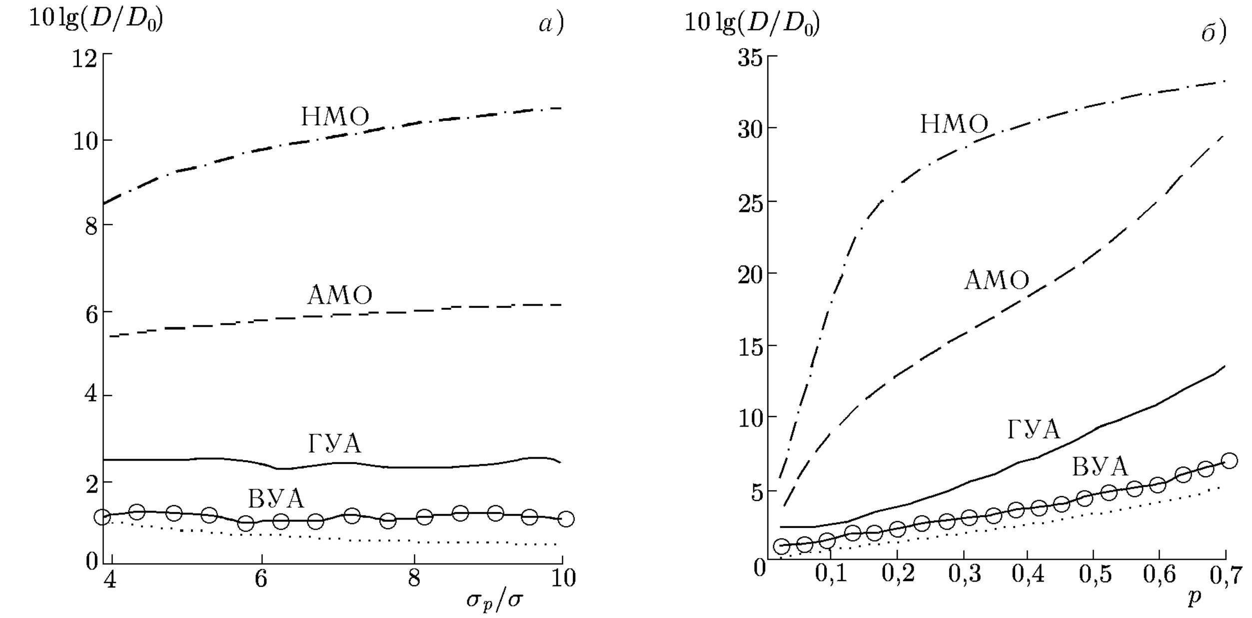

Рис. 4. СКО оценок дисперсии, полученные численным моделированием. Штриховая линия соответствует оценке с помощью абсолютного медианного отклонения (АМО); штрих-пунктирная — медианной оценке с предварительным некогерентным усреднением (НМО), сплошная линия — оценке (10) с постоянными весами без ограничения области аппроксимации (ГУА), сплошная линия с метками — методу устойчивой апроксимации (10) с весами $g_k = N_k^{-1}$, $C = 4$ (ВУА), пунктирная линия — границе Крамера—Рао

сравнение результатов с минимальной границей дисперсии, например, при $\sigma_p \geq 3\sigma$.

С помощью численного моделирования были получены СКО различных оценок дисперсии. На рис. 4 приведены СКО оценок дисперсии в зависимости от σ_p при $p = 0{,}05$ (рис. 4*а*) и в зависимости от p при $\sigma_p/\sigma = 10$ (рис. 4*б*). Из рис. 4 видно, что (10) обладает меньшим СКО по сравнению с обеими медианными оценками. Моделирование также показало, что СКО оценки Хьюбера (3) весьма близко к СКО оценки с использованием абсолютного медианного отклонения. Таким образом, предложенный метод устойчивой аппроксимации в варианте (10) с ограничением области аппроксимации и введением весов, фактически, реализует наилучшую из рассмотренных методов точность, а СКО этой оценки всего на 1÷1,5 дБ превышает (2) при достаточно сильных «загрязняющих» выбросах и незначительно отличается от границы Крамера—Рао.

3. АПРОБАЦИЯ МЕТОДА НА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Предложенный метод был апробирован на экспериментальных данных, представляющих собой запись сигнала от движущегося источника гидроакустического шума при наличии фоновой гауссовой помехи. При этом зависимость средней мощности принимаемого «полезного» сигнала от времени $P(t)$ в идеализированном случае может быть представлена как $P(t) = P_s r_0^2/r^2(t) + P_0$, где $r^2(t) = r_0^2 + (vt)^2$ — расстояние от источника до приёмной системы в момент времени t; r_0 и v — минимальное расстояние и скорость источника соответственно, $P_{\rm s}$ — мощность шума источника на расстоянии r_0, P_0 — мощность фоновой помехи. [3] Такого рода эксперименты являются типичными в гидроакустике при измерении мощности шума, излучаемого различными источниками

[3] На практике могут также наблюдаться зависимости, достаточно сильно отличающиеся от приведённого выше идеализированного случая из-за отражений от поверхности и дна, протяжённости источника и его собственной нестационарности.

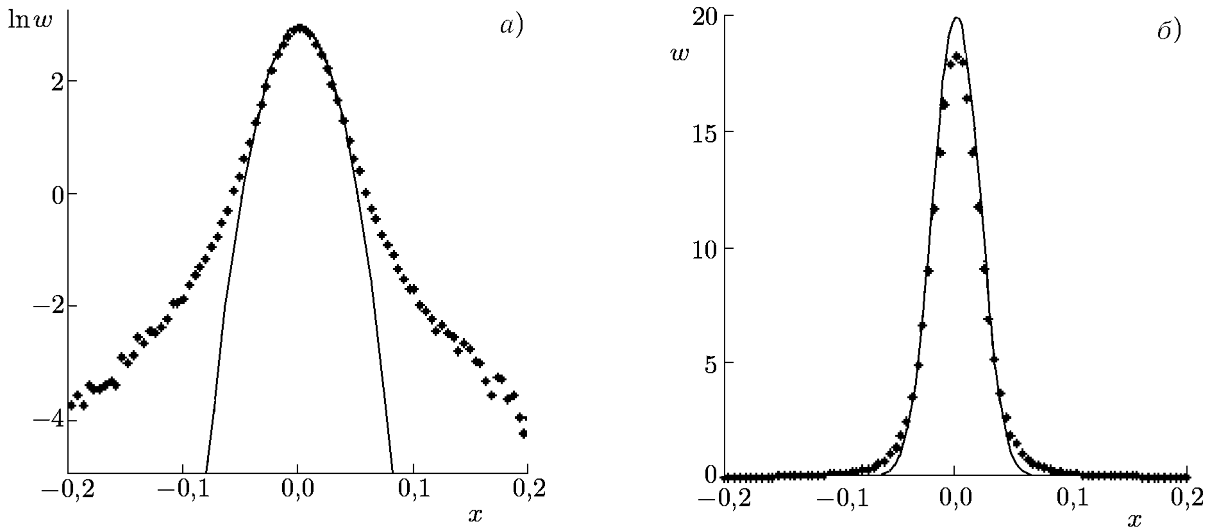

Рис. 5. Точками показана гистограмма экспериментального сигнала, полученная по выборке из приблизительно $2 \cdot 10^4$ отсчётов в логарифмическом масштабе (*а*) и в линейном масштабе (*б*). Сплошная линия — гауссовая плотность распределения вероятности, аппроксимирующая гистограмму в окрестности среднего

(см., например, [11]). Сигнал от источника наблюдался на фоне интенсивной импульсной помехи, которая возникала из-за механических колебаний приёмной системы, вызванных её случайными рывками. На рис. 5 приведена типичная гистограмма принятого сигнала, построенная по выборке из приблизительно $2 \cdot 10^4$ отсчётов. На рис. 5 в логарифмическом масштабе хорошо видны «приподнятые» по сравнению с гауссовой плотностью распределения вероятности (сплошная линия) хвосты гистограммы, определяемые импульсными помехами. Типичная осциллограмма «загрязняющего» импульсного выброса приведена на рис. 6.

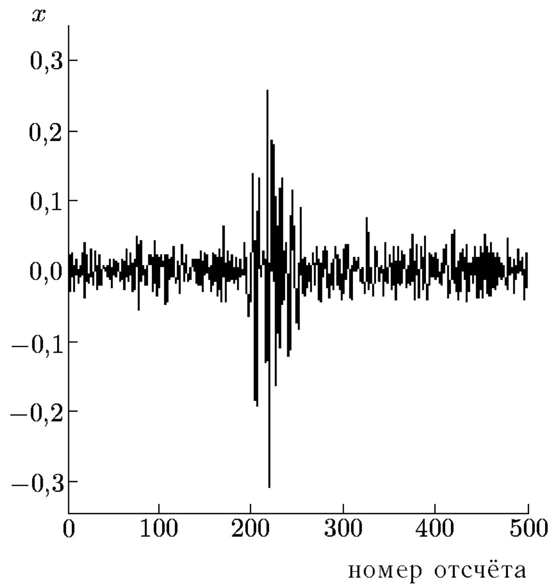

Рис. 6. Осциллограмма отдельного импульсного выброса

На рис. 7 приведены две зависимости оценки текущей дисперсии от времени (в относительных единицах). Сплошной линией показаны зависимости, полученные с использованием (10) для каждого интервала времени T. Пунктиром показаны зависимости, полученные с использованием (1) для временны́х интервалов $T_1 \approx T/10$. Выбор меньшего интервала усреднения для (1), чем для (10), позволяет качественно оценить моменты появления помехи.

Как видно из рис. 7, в отдельные моменты времени частота появления выбросов весьма высока, и тривиальная оценка (1) привела бы к существенному искажению зависимости средней мощности сигнала от положения источника; в то же время метод оценки (10) слабо реагирует на появление помехи, что говорит о его практической эффективности.

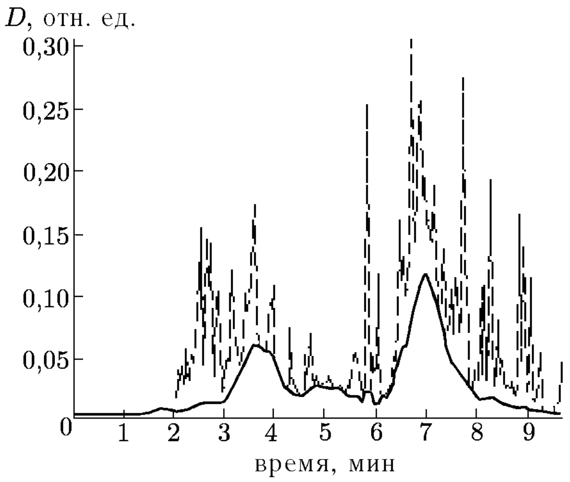

Рис. 7. Зависимости от времени оценки текущей дисперсии. Сплошная линия соответствует оценке (10) по интервалу T, штриховая линия — оценке (1) по интервалу $T_1 \approx T/10$

ВЫВОДЫ

Предложен новый метод оценки дисперсии гауссового процесса, «загрязнённого» импульсной помехой — метод устойчивой аппроксимации, основанный на определённой модификации метода минимума χ^2. Численное моделирование показало, что погрешность оценки с помощью метода устойчивой аппроксимации весьма близка к минимальной границе дисперсии и существенно меньше погрешности методов, использующих порядковые статистики. Метод прошёл успешную апробацию на экспериментальных данных.

Следует заметить, что в современных методах обработки сигналов метод минимума χ^2 практически не используется: лидирующее положение занимает метод максимума правдоподобия. Возможно, данная работа может стимулировать интерес к дальнейшему развитию статистик на основе метода минимума χ^2 и их использованию в различных приложениях.

Работа поддержана РФФИ (грант № 03–02–17035).

СПИСОК ЛИТЕРАТУРЫ

1. Караваев В. В., Сазонов В. В. Статистическая теория пассивной локации. М.: Радио и связь, 1987. 240 с.
2. Тихонов В. И. Статистическая радиотехника. М.: Радио и связь, 1982. 624 с.
3. Симкин А. В., Светников О. Г., Масленников Б. С., Гриць В. М. // Радиотехника. 1992. № 5–6. С. 35.
4. Романовский В. И. Математическая статистика. Кн. 2. Оперативные методы математической статистики. Ташкент: Изд-во АН Узбекской ССР, 1963. 794 с.
5. Хьюбер Дж. Робастность в статистике: Пер. с англ. М.: Мир, 1984. 304 с.
6. Хампель Ф., Рончетти Э., Рауссеу П., Штаэль В. Робастность в статистике. Подход на основе функций влияния: Пер. с англ. М.: Мир, 1989. 512 с.
7. Леман Э. Теория точечного оценивания: Пер. с англ. М.: Наука, 1991. 448 с.
8. Колданов А. П. // Радиотехника. 2001. № 9. С. 17.
9. Кендалл М., Стьюарт А. Статистические выводы и связи: Пер. с англ. М.: Наука, 1973. 900 с.
10. Бард Й. Нелинейное оценивание параметров: Пер. с англ. М.: Статистика, 1979. 349 с.

11. Урик Р. Основы гидроакустики: Пер. с англ. Л.: Судостроение, 1978. 448 с.

Поступила в редакцию 3 ноября 2003 г.; принята в печать 30 апреля 2004 г.

ROBUST ESTIMATION OF GAUSSIAN-NOISE VARIANCE IN THE PRESENCE OF PULSED INTERFERENCE

A. A. Rodionov and V. I. Turchin

A robust estimator of the mean power of a Gaussian noise signal in the presence of pulsed interference is proposed for remote sensing. The technique is based on the following rationale: under certain conditions, the histogram of samples of a signal contaminated by pulsed interference is close to a normal distribution in a certain range of snapshot values, i.e., the histogram is an estimate almost insensitive to the presence of pulsed interference. Correspondingly, we propose a modified minimum χ^2 method, which is called the method of stable approximation, for estimating the mean power of the “useful” Gaussian component. It is shown that the proposed estimate yields a higher accuracy compared with, e.g., the median estimates and is close to the Cramer–Rao lower bound. The results of experimental testing of the technique using hydroacoustic data are presented.

УДК 530.162

ВЕРОЯТНОСТНОЕ РАСПРЕДЕЛЕНИЕ СКОРОСТЕЙ БРОУНОВСКОЙ ЧАСТИЦЫ, РЕГИСТРИРУЕМОЙ ЗАМКНУТЫМ ДЕТЕКТОРОМ

Е. З. Грибова

Нижегородский госуниверситет им. Н. И. Лобачевского, г. Нижний Новгород, Россия

Для частиц пассивной примеси в потоке газа найдено вероятностное распределение скоростей на замкнутой границе заданной области пространства. Рассмотрены случаи различного соотношения между регулярным сносом и диффузией. Приведён пример вычисления плотности вероятностей скорости на границе области, окружающей источник частиц.

Естественное, а также постоянно усиливающееся антропогенное загрязнение окружающей среды делает всё более актуальными вопросы, связанные с изучением циркуляции пассивной примеси в атмосфере [1, 2]. В общем случае частицы примеси участвуют одновременно в регулярном сносе средой и в случайном (броуновском) перемешивании. Например, в [3] изучены закономерности диффузии частиц, движущихся в турбулентной среде под действием сил гравитации и Стокса. Для практических приложений часто бывает важно знать статистические характеристики частиц примеси в некоторой заранее заданной области пространства (в частности, при оседании дымового аэрозоля в окрестности промышленного узла это могут быть среднее и наиболее вероятное время достижения границ загрязняемого района либо вероятностные свойства скорости частиц вблизи этих границ). Назовём такие (заранее заданные) границы детектором. Для приложений интересно рассмотреть случай, когда поверхность детектора является замкнутой, т. е. задаётся уравнением

$$\varphi(\mathbf{r}) = C. \tag{1}$$

Пусть детектор регистрирует скорости пересекающих его частиц. При этом будем считать, что свойства среды (диффузия и снос) внутри и вне детектора одинаковы, а сам детектор никак не влияет на характер движения частиц (случай «прозрачного» детектора). Найдём $W(\mathbf{v};\varphi)$ — вероятностное распределение скоростей частиц, впервые пересекающих поверхность детектора. Заметим, что аналогичная задача решалась в [4, 5], однако детектор предполагался незамкнутым (его поверхность задавалась уравнением $z = \varphi(\mathbf{x})$, где z — продольная координата (вдоль вектора начальной скорости частицы), $\mathbf{x}$ — поперечная координата) и «поглощающим», т. е. частица могла только один раз пересечь поверхность. Здесь мы, во-первых, обсудим особенности, связанные с замкнутостью поверхности детектора, и, во-вторых, рассмотрим случай многократного пересечения границы.

1. ПОСТАНОВКА ПРОБЛЕМЫ

Совместим начало системы координат $\mathbf{r} = (r_1, r_2, r_3)$ с источником, который частица покидает в момент $t = 0$, тогда лагранжевы уравнения для координаты частицы $\mathbf{R}(t)$ и её скорости $\mathbf{V}(t)$ имеют вид [6, 7]

$$\frac{\mathrm{d}\mathbf{R}}{\mathrm{d}t} = \mathbf{V}, \qquad \frac{\mathrm{d}\mathbf{V}}{\mathrm{d}t} = -k\,(\mathbf{V} - \mathbf{v}_0') + \boldsymbol{\xi}(t),$$
$$\mathbf{R}(t=0) = 0, \qquad \mathbf{V}(t=0) = \mathbf{v}_0. \tag{2}$$

Здесь k — эффективный (т. е. приходящийся на единицу массы) коэффициент вязкого трения, $\mathbf{v}_0'$ — скорость окружающего потока, $\boldsymbol{\xi}(t)$ — случайная сила в расчёте на единицу массы, учитывающая броуновское взаимодействие. Будем считать, что $\boldsymbol{\xi}(t)$ — случайный гауссов процесс с нулевым средним и корреляционным тензором

$$\left\langle \xi_i(t)\xi_j(t+t')\right\rangle = 2D\,\delta_{ij}\,\delta(t'),$$

где D — коэффициент диффузии, δ_{ij} — символ Кронекера, $\delta(t')$ — дельта-функция Дирака; коэффициенты i, j принимают значения $1, 2, 3$. Это значит, что совместная плотность вероятностей координат и скорости частицы

$$f(\mathbf{r},\mathbf{v};t) = \left\langle \delta[\mathbf{r}-\mathbf{R}(t)]\,\delta[\mathbf{v}-\mathbf{V}(t)]\right\rangle,$$

где усреднение проводится по ансамблю случайных реализаций координат и скорости, удовлетворяет уравнению Фоккера—Планка

$$\frac{\partial f}{\partial t} + \mathbf{v}\,\frac{\partial f}{\partial \mathbf{r}} - k\,\frac{\partial}{\partial \mathbf{v}}\,[(\mathbf{v}-\mathbf{v}_0')\,f] = D\,\Delta_v f \tag{3}$$

с начальным условием

$$f(\mathbf{r},\mathbf{v};t=0) = \delta(\mathbf{r})\,\delta(\mathbf{v}-\mathbf{v}_0). \tag{4}$$

В правой части (3) лапласиан Δ_v вычисляется по компонентам скорости $\mathbf{v}$. Заметим, что рассматриваемая здесь постановка задачи позволяет вообще отказаться от граничного условия (в отличие, например, от [4], где оно, с одной стороны, было принципиально важным, а с другой — сильно затрудняло решение). Задача Коши (3), (4) является классической [6], и её точное решение — плотность вероятностей $f(\mathbf{r},\mathbf{v};t)$ — известно: это совместное гауссово распределение координаты и скорости. Наша цель — установить связь между плотностью вероятностей координат и скорости в фиксированный момент времени $f(\mathbf{r},\mathbf{v};t)$ и плотностью вероятностей скорости в фиксированной области пространства $W(\mathbf{v};\varphi)$. С точки зрения гидродинамики первая из упомянутых плотностей вероятностей описывает лагранжеву статистику броуновской частицы, вторая — эйлерову.

2. УСТАНОВЛЕНИЕ СВЯЗЕЙ МЕЖДУ ЭЙЛЕРОВОЙ И ЛАГРАНЖЕВОЙ СТАТИСТИКАМИ СКОРОСТИ ЧАСТИЦЫ

Рассмотрим интеграл

$$G(\mathbf{v};\varphi) \equiv \int\limits_0^\infty \mathrm{d}t \iiint\limits_{-\infty}^{+\infty} \mathrm{d}^3\mathbf{r}\, f(\mathbf{r},\mathbf{v};t)\,|\mathbf{v}_\perp|\,|\nabla\varphi(\mathbf{r}_C)|\,\delta[\varphi(\mathbf{r})-C] =$$
$$= \int\limits_0^\infty \mathrm{d}t \iiint\limits_{-\infty}^{+\infty} \mathrm{d}^3\mathbf{r}\,\langle\delta[\mathbf{r}-\mathbf{R}(t)]\,\delta[\mathbf{v}-\mathbf{V}(t)]\rangle\,|\mathbf{v}_\perp|\,|\nabla\varphi(\mathbf{r}_C)|\,\delta[\varphi(\mathbf{r})-C].$$

Здесь $\nabla\varphi(\mathbf{r}_C)$ означает градиент функции $\varphi(\mathbf{r}) = C$, взятый в точках поверхности детектора. Последняя дельта-функция в подынтегральном выражении «выкалывает» из всех значений $\mathbf{r}$ те, которые удовлетворяют уравнению (1), т. е. концы векторов $\mathbf{r}$ лежат на поверхности детектора. Введём в произвольной точке поверхности локальную систему координат $\{\mathbf{r}_\perp, \mathbf{r}_\parallel\}$, где вектор $\mathbf{r}_\perp$

направлен вдоль внешней нормали к поверхности в данной точке, а вектор $\mathbf{r}_\parallel$ лежит в касательной плоскости. Воспользуемся представлением дельта-функции [8, 9]

$$\delta[q(z)] = \sum_i \frac{\delta(z - z_i)}{|q'(z_i)|} \,, \tag{5}$$

где z_i — вещественные корни уравнения $q(z) = 0$, суммирование ведётся по всем корням. С помощью (5) преобразуем интеграл $G(\mathbf{v};\varphi)$ к виду

$$G(\mathbf{v};\varphi) = \int\limits_0^\infty \mathrm{d}t \iint\limits_{(\varphi)} \mathrm{d}^2\mathbf{r}_\parallel \left\langle \delta[\mathbf{r} - \mathbf{R}(t)]\, \delta[\mathbf{v} - \mathbf{V}(t)] \right\rangle |\mathbf{v}_\perp|,$$

где интегрирование ведётся по поверхности детектора.

В моменты времени t_i, где $i = 1, 2, \ldots$, частица пересекает поверхность детектора. Считая, что локальный базис $\{\mathbf{r}_\perp, \mathbf{r}_\parallel\}$ введён именно в точке пересечения, в той же точке представим уравнение траектории частицы в виде $\mathbf{R}(t) = \{\mathbf{R}_\perp(t), \mathbf{R}_\parallel(t)\}$. В моменты пересечения $\mathbf{r}_\perp = \mathbf{R}_\perp(t_i)$, а нормальная к поверхности компонента скорости равна $\mathbf{v}_{\perp i}$. Это позволяет снова воспользоваться формулой (5) и переписать дельта-функцию от координаты в виде

$$\delta[\mathbf{r} - \mathbf{R}(t)] = \delta[\mathbf{r}_\perp - \mathbf{R}_\perp(t)]\, \delta[\mathbf{r}_\parallel - \mathbf{R}_\parallel(t)] = \sum_i \frac{\delta[\mathbf{r}_\parallel - \mathbf{R}_\parallel(t)]\, \delta(t - t_i)}{|\mathbf{v}_{\perp i}|} \,,$$

тогда

$$G(\mathbf{v};\varphi) = \int\limits_0^\infty \mathrm{d}t \iint\limits_{(\varphi)} \mathrm{d}^2\mathbf{r}_\parallel \left\langle \sum_i \delta[\mathbf{r}_\parallel - \mathbf{R}_\parallel(t)]\, \delta[\mathbf{v} - \mathbf{V}(t)]\, \delta(t - t_i) \right\rangle .$$

Интегрируя сначала по времени t, находим

$$G(\mathbf{v};\varphi) = \iint\limits_{(\varphi)} \mathrm{d}^2\mathbf{r}_\parallel \left\langle \sum_i \delta[\mathbf{r}_\parallel - \mathbf{R}_\parallel(t_i)]\, \delta[\mathbf{v} - \mathbf{V}(t_i)] \right\rangle .$$

Теперь проинтегрируем по поверхности детектора в предположении, что частица пересекает её N раз. Это даёт

$$G(\mathbf{v};\varphi) = \left\langle \sum_{i=1}^N \delta[\mathbf{v} - \mathbf{V}(t_i)] \right\rangle . \tag{6}$$

Учитывая, что число пересечений N — случайное, и пользуясь формулой полной вероятности, окончательно получим

$$G(\mathbf{v};\varphi) = \sum_{N=1}^\infty P(N) \sum_{i=1}^N w_i(\mathbf{v};\varphi \mid N), \tag{7}$$

где $P(N)$ — вероятность того, что частица N раз пересекает поверхность детектора, $w_i(\mathbf{v};\varphi\,|N)$ — условная плотность вероятностей скорости в момент i-го пересечения (при условии, что всего пересечений N).

Учтём, что искомая плотность вероятностей $W(\mathbf{v};\varphi)$ — это вероятностное распределение скоростей частиц, впервые достигающих детектора. Тогда эта плотность вероятностей связана со слагаемыми суммы (7) следующим образом:

$$W(\mathbf{v};\varphi) = \sum_{N=1}^\infty P(N) w_1(\mathbf{v};\varphi \mid N). \tag{8}$$

Отметим, что полученное соотношение (8) является строгим, в то время как соответствующая «поглощающему» детектору граничная задача, рассмотренная в [4], не допускает точного аналитического решения.

Для дальнейшего анализа вычислим два вспомогательных интеграла: проинтегрируем функцию $G(\mathbf{v};\varphi)$ по скорости $\mathbf{v} = \{\mathbf{v}_\perp, \mathbf{v}_\parallel\}$, при этом отдельно найдём интегралы по $v_n > 0$ и $v_n < 0$, где v_n — проекция вектора $\mathbf{v}$ на внешнюю нормаль $\mathbf{n}$ к поверхности детектора. Для этого заметим, что каждое нечётное пересечение поверхности всегда происходит с $v_n < 0$ (частица влетает в детектор снаружи), а каждое чётное соответствует тому, что частица вылетает из детектора, т. е. $v_n > 0$. С учётом нормировки плотностей вероятности $w_i(\mathbf{v};\varphi \mid N)$ на единицу это даёт

$$J_+ \equiv \int\limits_{-\infty}^{+\infty} \mathrm{d}\mathbf{v}_\parallel \int\limits_0^{\infty} \mathrm{d}\mathbf{v}_\perp\, G(\mathbf{v};\varphi) = P(1)+P(2)+2P(3)+2P(4)+\ldots = \sum_{N=1}^{\infty} N\,[P(2N-1)+P(2N)], \tag{9а}$$

$$J_- \equiv \int\limits_{-\infty}^{+\infty} \mathrm{d}\mathbf{v}_\parallel \int\limits_0^{\infty} \mathrm{d}\mathbf{v}_\perp\, G(\mathbf{v};\varphi) = P(2)+P(3)+2P(4)+2P(5)+\ldots = \sum_{N=1}^{\infty} N\,[P(2N)+P(2N+1)]. \tag{9б}$$

3. ПРЕДЕЛЬНЫЕ ЧАСТНЫЕ СЛУЧАИ

Применим полученные выше общие соотношения для анализа некоторых характерных частных случаев. Для этого, пользуясь соображениями размерности, прежде всего введём параметр, определяющий конкуренцию диффузии и регулярного сноса:

$$\alpha = k v_0'^2/(2D).$$

Очевидно, что характер движения частицы существенно отличается при больших и малых α, поэтому рассмотрим эти случаи отдельно.

3.1. Сильный снос при слабой диффузии ($\alpha \gg 1$)

Будем считать, что скорость сноса направлена в сторону детектора и свойства среды вне и внутри детектора одинаковы. Тогда очевидно, что в этом случае частица либо вообще пролетит мимо детектора, либо дважды пересечёт его поверхность (влетит и вылетит). При этом с учётом условия нормировки вероятностей на единицу можно считать, что вероятность $P(2) = 1 - P(0)$, а вероятности любого другого числа пересечений (при $N \neq 0$ и $N \neq 2$) пренебрежимо малы. Тогда по формуле (7) получим

$$G(\mathbf{v};\varphi) = P(2)\,[w_1(\mathbf{v};\varphi \mid 2) + w_2(\mathbf{v};\varphi \mid 2)].$$

С другой стороны, формула (8) даёт

$$W(\mathbf{v};\varphi) = P(2) w_1(\mathbf{v};\varphi \mid 2).$$

Поскольку детектор никак не влияет на движение частицы, то при сильном сносе условное вероятностное распределение скорости не зависит от того, с какой стороны происходит пересечение, т. е.

$$w_1(\mathbf{v};\varphi \mid 2) = w_2(\mathbf{v};\varphi \mid 2),$$

откуда

$$W(\mathbf{v};\varphi) = G(\mathbf{v};\varphi)/2. \tag{10}$$

3.2. Слабый снос при сильной диффузии ($\alpha \ll 1$)

Качественные представления о характере движения частицы подсказывают, что в этом случае частица может вообще не попасть в детектор (не долетает, хаотически двигаясь вблизи источника). Пример такого движения можно видеть на рис. 1, где представлены траектории, полученные численным моделированием стохастических уравнений (2) в двумерном случае при $\alpha_x = \alpha_z = 0{,}01$, где z — координата вдоль вектора начальной скорости частицы, x — поперечная координата, $\alpha_x = kv_{0x}'^2/(2D)$, $\alpha_z = kv_{0z}'^2/(2D)$. На этом рисунке детектор Д выбран в виде отрезка, перпендикулярного оси z. С точки зрения нашей задачи этот случай не представляет интереса. Поэтому сначала рассмотрим такие параметры α и расстояния от источника до детектора, когда вероятность хотя бы однократного попадания частицы в детектор становится заметно отличной от нуля за конечный промежуток времени.

Такая возможность, прежде всего, реализуется, когда детектор расположен настолько близко к источнику, что даже при $\alpha \ll 1$ частица пересекает границу детектора. В пределе, когда сносом можно пренебречь, за бесконечное время число пересечений поверхности детектора стремится к бесконечности, а вероятности всех пересечений одинаковы. Обозначим их P_∞. С учётом нормировки вероятностей на единицу получим $P_\infty \to 0$, и

$$W(\mathbf{v};\varphi) = 0.$$

Физический смысл последнего равенства очевиден. Вследствие изотропии диффузии равновозможны любые направления единичных толчков. Это приводит к тому, что распределение скоростей становится гауссовым с дисперсией $\sigma_v^2 \approx D/k$, которая растёт с уменьшением параметра α (с ростом коэффициента диффузии). В результате при $\alpha \to 0$ скорость равновероятно принимает любые значения из интервала $(-\infty, +\infty)$.

Для приложений более интересен другой случай, когда значения α и расстояние от источника до детектора таковы, что выполнено условие однократного попадания частицы в детектор $P(1) \gg P(2)$, которое сводится тогда к неравенству $J_+ \gg J_-$. Оценим необходимые для этого значения α с помощью численного моделирования.

Предположим, что радиус R кривизны поверхности детектора всюду положителен, а в точке, где частица впервые пересекает детектор, удовлетворяет неравенству $R \geq k^{-1}v_0'$ («плоская» поверхность). Вычисляя интегралы J_+ и J_- для плоского детектора, получаем, что при $\alpha_x = \alpha_z = 0{,}001$ частица вообще не достигает детектора за конечное время, в то время как при $\alpha_x = 0{,}001$ и $\alpha_z = 0{,}01$ вероятность даже двукратного пересечения $P(2)$ составляет не более $0{,}1P(1)$ (вероятности $P(1;kt)$ и $P(2;kt)$ для этого случая приведены на рис. 2). На основании этого будем считать, что при данных значениях α вероятностями более чем одного пересечения детектора частицей можно пренебречь. Сравнение формул (7) и (8) в этом случае показывает, что интеграл $G(\mathbf{v};\varphi)$ как раз даёт нужную плотность вероятностей:

$$W(\mathbf{v};\varphi) = G(\mathbf{v};\varphi). \tag{11}$$

Анализ наиболее интересных предельных случаев (10), (11) позволяет считать, что при произвольном соотношении между диффузией и сносом выполнено соотношение

$$W(\mathbf{v};\varphi) = qG(\mathbf{v};\varphi), \tag{12}$$

где q — множитель, который найдём из условия нормировки плотности вероятностей $W(\mathbf{v};\varphi)$ на единицу. Интегрируя (12) по всем скоростям, с учётом формул (9а), (9б) получим

$$q = \left\{\sum_{N=1}^{\infty} N\left[P(2N-1) + 2P(2N) + P(2N+1)\right]\right\}^{-1}.$$

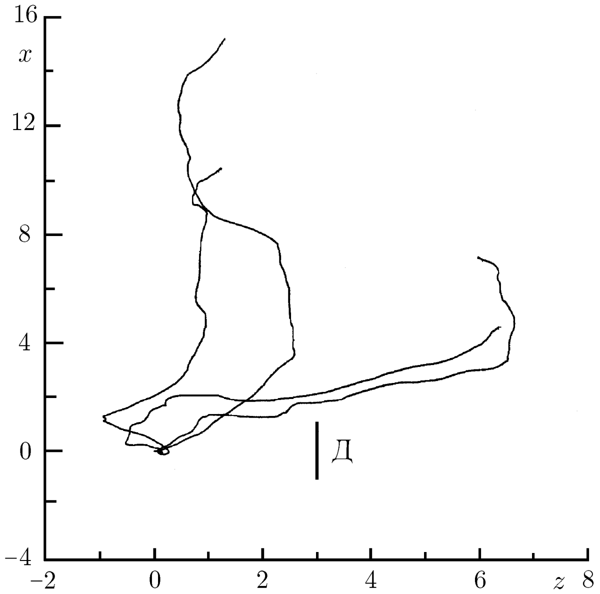

Рис. 1. Траектории движения частиц в двумерном случае

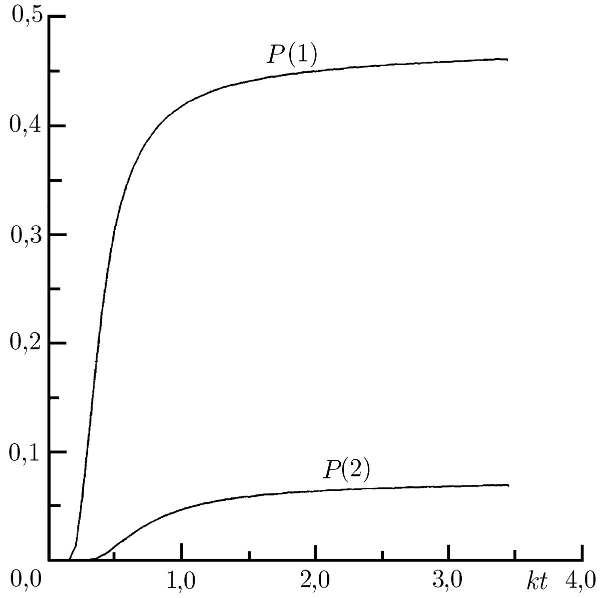

Рис. 2. Вероятности одно- и двукратного пересечения детектора

3.3. Источник внутри детектора

В качестве примера рассмотрим ситуацию, когда точечный источник находится в центре детектора-шара радиуса C. Начало координат по-прежнему свяжем с источником, ось $\mathbf{r}$ сферической системы координат направим вдоль вектора начальной скорости частицы $\mathbf{v}_0$. Вектор сноса $\mathbf{v}'_0$ имеет составляющие $v'_{0r}, v'_{0\theta}, v'_{0\phi}$.

Здесь, в отличие от рассмотренного в разделе 3.2 случая (когда источник находился вне детектора), при сильном сносе вероятность однократного достижения границы детектора $P(1)$ стремится к единице, а всеми другими вероятностями (с $N \neq 1$) можно пренебречь. Если же диффузия преобладает над сносом, то, вообще говоря, могут быть отличны от нуля все вероятности с $N \geq 1$. В то же время численное моделирование, проведённое при различных значениях $\alpha \ll 1$, показывает, что в любой момент времени выполнены соотношения $P(2) \ll P(1)$ и $P(3) \ll P(2)$, что позволяет и в этом случае пренебрегать всеми вероятностями, кроме $P(1)$. Тогда формула (11) даёт нужное вероятностное распределение независимо от параметра α.

Необходимое для вычисления интеграла $G(\mathbf{v};\varphi)$ решение задачи Коши (3), (4) известно: по любому из направлений совместное распределение координаты и скорости частицы $f_i(r, v_i; t)$, где $i = r, \theta, \phi$, является гауссовым. Вводя более удобные для численного анализа безразмерные координаты $\rho = kr/v'_{0r}$, скорости $u_i = v_i/v'_{0i}$ и время $\tau = kt$, безразмерную плотность вероятностей

$$\gamma_i(\rho, u_i; \tau) = (v'^2_{0i}/k)\, f_i(r, v_i; t)$$

запишем в виде

$$\gamma_i(\rho, u_i; \tau) = \frac{\alpha_i}{\pi\,\sqrt{c(\tau)}}\, \exp\biggl\{-\frac{\alpha_i}{c(\tau)} \Bigl[\bigl(1 - \exp(-2\tau)\bigr)\,[\rho - \tau - q_i\,(1 - \exp(-\tau))]^2 - \\ - 2\,\bigl(1 - \exp(-\tau)\bigr)^2\,[\rho - \tau - q_i\,(1 - \exp(-\tau))]\,\bigl(u_i - 1 - q_i \exp(-\tau)\bigr) + \\ + \bigl(2\tau + 4\exp(-\tau) - \exp(-2\tau) - 3\bigr)\bigl(u_i - 1 - q_i \exp(-\tau)\bigr)^2\Bigr]\biggr\}, \quad (13)$$

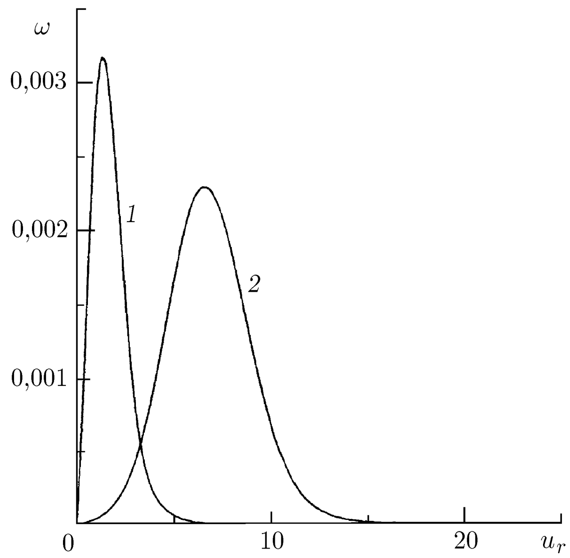

Рис. 3. Зависимость плотности вероятностей $\omega(u_r;\varphi)$ от параметра $\alpha=\alpha_r=\alpha_\theta=\alpha_\phi$: кривая *1* соответствует $\alpha=0{,}1$, *2* — $\alpha=0{,}01$

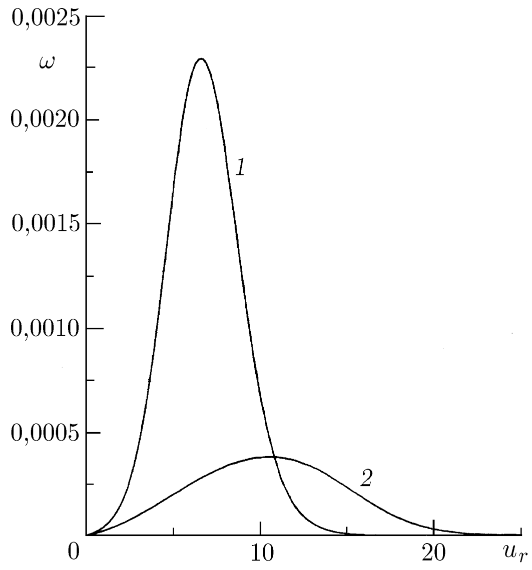

Рис. 4. Зависимость плотности вероятностей $\omega(u_r;\varphi)$ от размеров детектора: кривая *1* соответствует $\zeta=1$, *2* — $\zeta=5$

где

$$c(\tau)=2\tau\,[1-\exp(-2\tau)]-4\,[1-\exp(-\tau)]^2,\qquad q_i=(v_{0i}-v'_{0i})/v'_{0i},\qquad \alpha_i=kv'^2_{0i}/(2D).$$

Все направления независимы, поэтому

$$\gamma(\mathbf{r},\mathbf{v};t)=\prod_{i=1}^{3}\gamma_i(r,v_i;t). \tag{14}$$

Таким образом, вместо размерной плотности вероятностей $W(\mathbf{v};\varphi)$ мы ищем безразмерную величину

$$\omega(\mathbf{u};\varphi)=v'_{0r}v'_{0\theta}v'_{0\phi}W(\mathbf{v};\varphi).$$

В сферических координатах интеграл

$$g(\mathbf{u};\varphi)=v'_{0r}v'_{0\theta}v'_{0\phi}G(\mathbf{v};\varphi)$$

равен

$$g(\mathbf{u};\varphi)=\zeta^2\int\limits_0^\infty \mathrm{d}\tau\int\limits_0^{2\pi}\mathrm{d}\phi\int\limits_0^{\pi}\mathrm{d}\theta\,|u_r|\,\gamma(\zeta,\theta,\phi;\tau)\sin\theta, \tag{15}$$

где ζ — безразмерный радиус детектора.

На рис. 3 приведены результаты вычисления плотности вероятностей $\omega(u_r;\varphi)$ по формулам (6), (11), (13), (14), (15) при разных значениях α (диффузия и снос предполагаются изотропными, радиус детектора $\zeta=1$). Результат легко понять: при постоянном сносе уменьшение параметра α означает, что диффузия всё сильнее преобладает над сносом. Это приводит к росту дисперсии скорости $\sigma_u^2\approx D/k$.

Зависимость плотности вероятностей $\omega(u_r;\varphi)$ от размеров детектора ζ представлена на рис. 4 (диффузия и снос также изотропны: $\alpha_r=\alpha_\theta=\alpha_\phi=0{,}01$). С ростом расстояния от источника до детектора частица при движении к его поверхности испытывает всё большее число случайных толчков, в результате плотность вероятностей скорости из-за беспорядочного движения частиц «расплывается», и растёт вероятность попадания в детектор частиц с бо́льшими значениями u_r.

4. ЗАКЛЮЧЕНИЕ

Численно-аналитическими методами получено решение важной для приложений задачи: найдено вероятностное распределение скорости броуновской частицы, регистрируемой замкнутым детектором. Интересно отметить существенное отличие данной задачи от рассмотренной в [4]: в случае «поглощающего» детектора задача сводится к краевой для уравнения Фоккера—Планка и не допускает в общем случае аналитического решения. В то же время для детектора, который никак не влияет на движение частиц, удаётся найти достаточно точное выражение для вероятностного распределения скорости первого пересечения поверхности.

Автор выражает признательность А. И. Саичеву за постоянное внимание к работе. Работа выполнена при финансовой поддержке РФФИ (грант № 03–02–16680), Совета по грантам Президента РФ для поддержки ведущих научных школ (грант № 00–15–96619), а также Минобразования РФ (грант № Е02–3.5–232).

СПИСОК ЛИТЕРАТУРЫ

1. Csanady G. T. Turbulent diffusion in the environment. D. Reidel Publ. Comp., 1980. 250 p.
2. Григорьев А. И., Сидорова Т. И. // ЖТФ. 1998. Т. 68, вып. 3. С. 20.
3. Грибова Е. З., Жукова И. С., Лапинова С. А. и др. // ЖЭТФ. 2003. Т. 123, вып. 3. С. 543.
4. Грибова Е. З., Саичев А. И. // Изв. вузов. Радиофизика. 1998. Т. 41, № 10. С. 1301.
5. Грибова Е. З., Саичев А. И. // ЖТФ. 2000. Т. 70, вып. 9. С. 1.
6. Чандрасекар С. Стохастические проблемы в физике и астрономии. М.: Изд-во иностр. лит., 1947. 168 с.
7. Кляцкин В. И. Стохастическое описание динамических систем с флуктуирующими параметрами. М.: Наука, 1975. 239 с.
8. Gurbatov S. N., Malakhov A. N., Saichev A. I. Nonlinear random waves and turbulence in nondispersive media: waves, rays and particles. Manchester and New York: Manchester University Press, 1991. 308 p.
9. Saichev A. I., Woyczynski W. A. Distributions in the physical and engineering sciences. Boston: Birkhänser, 1997. 336 p.

Поступила в редакцию 15 октября 2003 г.; принята в печать 16 января 2004 г.

VELOCITY DISTRIBUTION OF BROWNIAN PARTICLES AT THE INPUT OF A CLOSED DETECTOR

E. Z. Gribova

The velocity distribution function of passive-tracer particles in a gas flow on a closed boundary of a given spatial region is found for various ratios between the regular drift and diffusion. An example of calculation of the velocity probability density on the boundary of a region comprising a source of particles is given.