

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Известия высших учебных заведений

РАДИОФИЗИКА

ежемесячный научно-технический журнал

Издаётся с апреля 1958 г.

Том XLV №5

Нижний Новгород

2002

Содержание

Данилкин Н. П., Котонаева Н. Г. Расчёт высотных профилей электронной концентрации в ионосфере по ионограммам космической станции «Мир»	367
Бенедиктов Е. А. , Беликович В. В., Бахметьева Н. В., Толмачёва А. В. Искусственные периодические неоднородности ионосферной плазмы — перспективное направление в исследовании ионосферы	375
Чернов Ю. А. О пространственной корреляции поля коротких волн при наклонном отражении от ионосферы	392
Рапопорт В. О., Митяков Н. А. Об использовании метеорадаров миллиметрового диапазона длин волн для измерения температурных профилей в атмосфере	403
Фисанов В. В. О спектре волн плоскопараллельного кироволновода	406
Белашов В. Ю. О структуре решений обобщённых уравнений Кадомцева—Петвиашвили	411
Божков В. Г. О природе «низкотемпературной аномалии» в контактах металл—полупроводник с барьером Шоттки	416
Заярный В. П. Исследование зарядовых и частотных свойств твердотельных гетерогенных структур с зарядовой связью	427
Шкелёв Е. И., Кисляков А. Г., Лупов С. Ю. Методы ослабления эффектов интермодуляции в распределении Вигнера—Виля	433
Павленко А. А. О точности определения параметров реверберации при восстановлении сигнала с реверберационными искажениями в случае дискретных задержек	443
Костылев В. И. Характеристики энергетического обнаружения квазидетерминированных радиосигналов с нерелеевским распределением флуктуаций амплитуды	450

УДК 550.3

РАСЧЁТ ВЫСОТНЫХ ПРОФИЛЕЙ ЭЛЕКТРОННОЙ КОНЦЕНТРАЦИИ В ИОНОСФЕРЕ ПО ИОНОГРАММАМ КОСМИЧЕСКОЙ СТАНЦИИ «МИР»

Н. П. Данилкин, Н. Г. Котонаева

Институт прикладной геофизики РАН, г. Москва, Россия

В статье приведены результаты расчётов профилей концентрации электронов во внутренней и внешней ионосфере (ниже и выше высоты максимума ионосферы соответственно) по данным ионозонда космической станции «Мир», летавшей на высотах около 350 км. Отмечено, что по данным наземных ионосферных станций высота максимума ионосферы ночью определяется недостаточно корректно.

ВВЕДЕНИЕ

Установка ионозонда на борту обитаемой орбитальной космической станции (ОС) «Мир», летавшей на высотах около 350 км, означала новый качественный этап зондирования ионосферы с ИСЗ. Основанием для этого были результаты экспериментов по трансionoсферному зондированию, выполненных на ИСЗ «Интеркосмос-19» и «Космос-1809» [1], которые показали, что основные параметры максимума слоя F ионосферы ($h_{\max F_2}$ — высота максимума слоя F_2 , f_{oF_2} и f_{xF_2} — критические частоты обыкновенной и необыкновенной волн в этом слое, q — полутолщина области F в модельном приближении) могут быть определены при любой высоте расположения ИСЗ [2].

При этом такое существенное изменение расположения ионозонда (от 1 000 до 350 км) оказалось весьма благоприятным для определения параметров ионосферы. В частности, это относится к методам расчёта высотных профилей электронной концентрации по спутниковым ионограммам. Целью настоящей статьи является описание метода и первых результатов расчёта указанных профилей в ионосфере по ионограммам, полученным при радиозондировании ионосферы с борта ОС «Мир».

Расчёт высотных профилей электронной концентрации $N_e(h)$ в ионосфере является завершающим этапом вертикального радиозондирования сверху (из космоса) или снизу (с поверхности Земли) и во многом определяет реальную мощь этого классического метода. Формально такой расчёт представляет собой решение интегрального уравнения Вольтерра для каждой в отдельности из магнитоионных компонент отражённого от ионосферы сигнала или для той или иной их комбинации. На практике расчёт сводится к обращению треугольных матриц. В классе монотонных функций решение является единственным, если используется только одна магнитоионная компонента отражённого от ионосферы сигнала. Одновременное использование обеих компонент при наличии на ионограмме достаточной информации позволяет получать решения и в классе немонотонных функций. Однако в области немонотонности (как правило, это «долина» между слоями E и F) найти однозначное решение не удаётся. Для получения большей информации используются поляризационные ионозонды, на ионограммах которых отсутствуют области, где невозможно прочесть отсчёты действующих высот отражённых от ионосферы сигналов из-за поляризационного наложения сигналов один на другой. Появление подобных приборов позволяет уменьшить диапазон неопределённости решения при нахождении профилей электронной концентрации в областях «ненаблюдаемой ионизации» [3].

Разработка метода, создание аппаратуры и проведение космических экспериментов по трансionoсферному радиозондированию привело к созданию методов расчёта высотных профилей электронной концентрации в ионосфере на основе решения интегральных уравнений Фредгольма, которое формально сводилось к обращению прямоугольных матриц. Единственность решения в этом случае достигается в классе «монотонных эквивалентов» [4]. Разработка метода нахождения высотного профиля

электронной концентрации на основе решения уравнений Фредгольма продвинула вперёд и возможности обычного метода определения зависимости $N_e(h)$ за счёт использования в расчётах отражения от Земли, которое в этой терминологии является двойной трансionoграммой.

Высота орбиты станции «Мир» в среднем соответствовала высоте расположения основного максимума электронной концентрации в ионосфере. Однако ионосфера весьма изменчива во времени и в пространстве, поэтому во время экспериментов станция «Мир» находилась то выше, то ниже главного максимума концентрации электронов.

Появление ионограмм спутникового радиозондирования при расположении ионозонда ниже максимума области F предоставило новые возможности в методах расчёта профилей $N_e(h)$ во внутренней ионосфере. Эти возможности появляются вследствие комбинации решений на основе методов, использующих отражение радиоволн от ионосферы (т. е. на основе решения уравнения Вольтерра), и решений на основе радиопросвечивания (т. е. на основе решения уравнения Фредгольма). Разработанные методы опираются на возможность использования в расчётах, по крайней мере, одной точки на профиле $N_e(h)$, положение которой определяется из навигационных спутниковых данных с беспрецедентной для такого рода задач точностью. Последнее означает, что и весь профиль $N_e(h)$ рассчитывается значительно точнее.

Заметим, что все прежние спутниковые ионозонды в основном располагались на космических аппаратах с орбитами, лежащими на высотах в окрестности 1 000 км, и ионограммы внешнего зондирования, полученные ими, давно и хорошо известны [5]. Когда в нашем эксперименте станция «Мир» находилась выше основного максимума ионосферы, регистрируемые ионограммы качественно не отличались от тех, которые были получены на других спутниках ранее. На рис. 1 приведена подобная достаточно типичная ионограмма с ионозонда на ОС «Мир». На рис. 1 цифрами обозначены следующие детали ионограммы: 1 — след отражения от ионосферы обыкновенной компоненты излучения, 2 — след отражения от ионосферы необыкновенной компоненты, 3 — частота верхнего электронно-ионного гибридного резонанса, 4 — след отражения от Земли необыкновенной компоненты, 5 — след отражения от Земли обыкновенной компоненты, 6 — критическая частота обыкновенной компоненты $f_{oF_2} = 7,6$ МГц, 7 — критическая частота необыкновенной компоненты $f_{xF_2} = 8,3$ МГц.

Хотя след отражения от ионосферы достаточно короткий, тот факт, что высота станции известна с очень высокой точностью (до десятых или сотых долей километра), уменьшает ошибку в определении высоты максимума электронной концентрации. Указанное обстоятельство представляется весьма перспективным для развития метода определения профиля $N_e(h)$ с повышенной точностью. Это связано с тем, что накопленная ошибка при вычислении высоты каждого из нижележащих использованных слоёв, расположенных между спутником и максимумом концентрации электронов, будет мала.

Следует также иметь в виду, что на ионограммах внешнего зондирования необходимо различать частоту отсечки необыкновенной компоненты и частоту верхнего электронно-ионного гибридного резонанса.

1. ХАРАКТЕРНЫЕ ОСОБЕННОСТИ ИОНОГРАММ

В зависимости от положения спутника относительно максимума области F на ионограммах возникают существенно различные детали. Примеры таких ионограмм от 21.04.99 представлены на рис. 2.

Ионограммы на рис. 2а, б получены, когда станция находилась выше максимума электронной концентрации, и напоминают обычные ионограммы внешнего зондирования с отчётливыми следами отражения обыкновенной и необыкновенной компонент от слоя F_2 .

Ионограмма на рис. 2в соответствует редкому случаю, когда станция пересекает максимум слоя. При этом след отражения от максимума стал почти вертикальным (плазменная частота обыкновен-

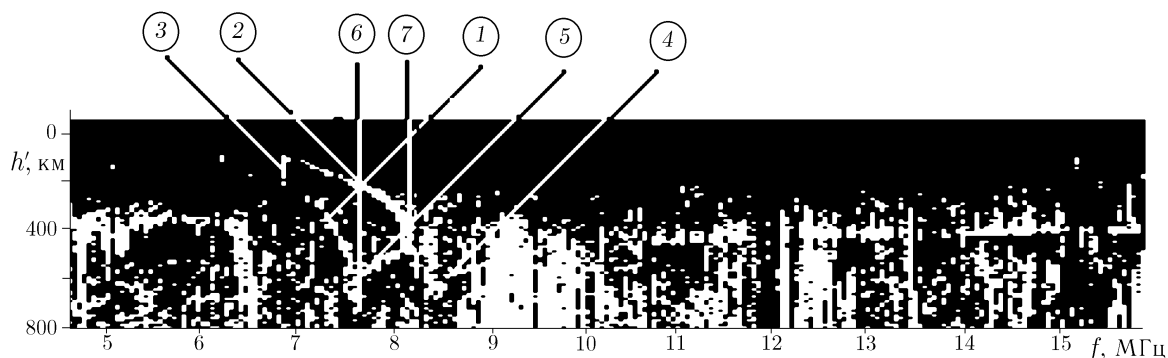


Рис. 1. Ионограмма радиозондирования ионосферы, полученная на ОС «Мир» 23.04.99 в 10:45:11 MSK. Координаты станции в этот момент времени 46,13° с. ш., 62,72° в. д., высота 346,56 км

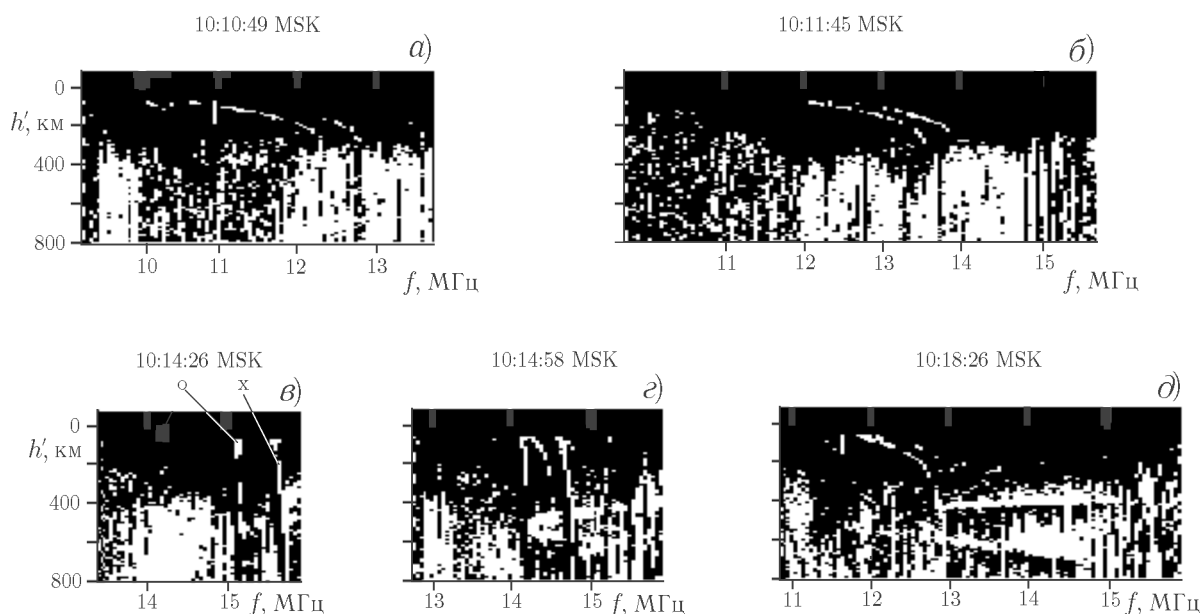


Рис. 2

ной компоненты меняется от 15,05 до 15,1 МГц (последнее значение равно f_{oF_2}), а плазменная частота необыкновенной компоненты — от 15,65 до 15,7 МГц, $f_{xF_2} = 15,7$ МГц). Соответствующие следы отмечены на рисунке значками о и х.

На рис. 2г, д представлены характерные примеры ионограмм в случае, когда ОС «Мир» находилась ниже максимума слоя F . Основной принципиальной особенностью таких ионограмм является отличие низшей частоты отражения зондирующих радиоволн от Земли от критической частоты f_{oF_2} области F . Совпадение же указанных параметров является характерной чертой внешнего зондирования (т. е. радиозондирования, при котором ионозонд расположен выше (как правило, значительно) максимума концентрации электронов в ионосфере) при вертикальном распространении зондирующих радиоволн. В случае, если спутник расположен ниже максимума электронной концентрации области F , то низшая частота отражения радиоволн от Земли определяет плазменную частоту ионосферы в месте расположения станции.

2. МЕТОД

Для практического использования расчётов профилей $N_e(h)$ особое значение имеет точное определение высоты максимума электронной концентрации. Поэтому в данной публикации этому вопросу уделено особое внимание.

Расчёт зависимости электронной концентрации от высоты в настоящей работе основан на идеях и алгоритмах, описанных в [6]. В кратком изложении метод состоит в следующем.

Действующая высота (глубина) отражения h' для радиоволны с частотой f , падающей на ионосферу вертикально, вычисляется по формуле

$$h' = \int_{h_s}^h \mu'(f, h) dh, \quad (1)$$

где μ' — групповой показатель преломления на частоте f , h — истинная высота (глубина) отражения, h_s — высота ИСЗ. Если разбить отрезок $[h, h_s]$ на N слоёв и ввести

$$\bar{\mu}'_{rn} = \frac{1}{h_r - h_{r-1}} \int_{h_{r-1}}^{h_r} \mu'(f_n, f_r, h) dh \quad (2)$$

— среднее значение группового показателя преломления в r -м слое (при $h_{r-1} < h < h_{r+1}$) на частоте f_n , где $r = 1, 2, \dots, N$, f_n — электронная плазменная частота в n -м слое, то вместо (1) получается система из N уравнений, каждое из которых выражает действующую высоту отражения на частоте f_n :

$$h'_n = \sum_{r=1}^n \bar{\mu}'_{rn} (h_r - h_{r-1}). \quad (3)$$

Из этой системы уравнений выражаются $\Delta h_r = h_r - h_{r-1}$, причём при вычислении Δh_r используются уже найденные величины Δh_i при $i < r$, т. е. (3) преобразуется в систему из N рекурсивных формул для вычисления приращений истинных высот отражения:

$$\Delta h_n = \frac{1}{\beta_{nn}} (\Delta h'_n + \sum_{r=1}^{n-1} \beta_{rn} \Delta h_r), \quad (4)$$

где $\Delta h'_r = h'_r - h'_{r-1}$ — приращения действующих высот отражения,

$$\beta_{rn} = \begin{cases} \bar{\mu}'_{r(n-1)} - \bar{\mu}'_{rn}, & r \neq n; \\ \bar{\mu}'_{nn}, & r = n. \end{cases}$$

Полученные рекурсивные формулы позволяют последовательно вычислять приращения истинной высоты отражения, зная средние значения $\bar{\mu}'_{rn}$ соответствующих групповых показателей преломления.

Для расчёта коэффициентов $\bar{\mu}'_{rn}$ используется предположение, что в каждом из элементарных слоёв истинная высота (глубина) отражения h является линейной функцией отношения $w = f/f_n$, где f — частота волны. В этом случае из соотношения (2) получаем

$$\bar{\mu}'_{rn} = \frac{\Delta(\mu_{rn} w_{rn})}{\Delta w_{rn}} - \frac{1}{\Delta w_{rn}} \int_{w_{r-1}}^{w_r} u \frac{\partial \mu}{\partial u} dw \approx 0,98 \frac{\Delta(\mu_{rn} w_{rn})}{\Delta w_{rn}}, \quad (5)$$

где μ — фазовый показатель преломления, $u = f_h/f_n$, f_h — гирочастота, $\Delta(\mu_{rn}w_{rn}) = \mu_{rn}w_{rn} - \mu_{(r-1)n}w_{(r-1)n}$, $\Delta w_{rn} = w_{rn} - w_{(r-1)n}$, $w_{rn} = f_r/f_n$.

Последнее приближённое равенство справедливо из-за того, что второй член в средней части (5) для обыкновенной компоненты излучения мал, постоянен и составляет не более 2 % от первого слагаемого. Особенности расчётов для необыкновенной компоненты обсуждаются в [6].

Коэффициенты μ'_{rn} зависят также от гирочастоты f_h и магнитного наклона θ , которые определяются из навигационных данных ИСЗ и могут уточняться (f_h) по ионограммам.

Используя соотношения (5) для вычисления коэффициентов системы (4), можно найти приращение истинной высоты. Суммируя их для каждой частоты f_n , получаем истинную высоту отражения и тем самым восстанавливаем искомый профиль $N_e(h)$ в ионосфере.

3. РЕЗУЛЬТАТЫ

Эксперименты по радиозондированию с борта ОС «Мир» проводились с 12.08.98 по 25.06.99. Было получено свыше 10 000 ионограмм в различных районах планеты в разных условиях и в разное время года. Расчёт профилей $N_e(h)$ применялся во многих случаях для решения тех или иных конкретных задач. Описание результатов этих расчётов приведено в соответствующих тематических публикациях [7–9].

В настоящей работе рассматривается расчёт высотных профилей электронной концентрации в ионосфере в случае, когда станция «Мир» пересекает максимум концентрации электронов в слое F_2 при движении станции сверху вниз. Этот пример, с одной стороны, как нам кажется, хорошо иллюстрирует возможности метода, а с другой — интересен в геофизическом смысле.

Ионограммы, представленные на рис. 2, получены при радиозондировании со станции «Мир» в течение небольшого (менее 8 минут) интервала времени, когда ОС «Мир» пролетала над о. Тайвань. В это время в местном университете С. А. Пулинцом было организовано наблюдение ионосферы методом вертикального зондирования с помощью наземного ионозонда с хорошим разделением сигналов различной поляризации, а также приём ионограмм с ОС «Мир» в аналоговом режиме [7]. Высотные профили электронной концентрации, при зондировании сверху и снизу, рассчитанные С. А. Пулинцом, хорошо согласуются, и высота максимума ионосферы в обоих случаях оказывается одинаковой. Кроме того, было отмечено, что, когда ОС «Мир» заходила за горизонт для наблюдателя на о. Тайвань, она опустилась ниже максимума слоя. Мы рассмотрели этот случай пролёта ОС «Мир» над о. Тайвань до её пересечения с экватором, используя не аналоговые ионограммы, а цифровые, полученные на станции «Мир» в режиме с запоминанием информации (см. рис. 2).

В течение всего периода наблюдений станция «Мир» меняла своё положение относительно максимума слоя F_2 . Ионограммы на рис. 2а, б, получены с интервалом приблизительно 1 минута, когда станция находилась выше максимума слоя. Используя метод расчёта истинной глубины отражения, изложенный выше, по следам отражения от ионосферы мы рассчитали профили $N_e(h)$ выше максимума слоя F_2 (кривые 1 и 2 на рис. 3). Для сравнения на рис. 3 помещён высотный профиль электронной концентрации во внутренней и внешней ионосфере (отмечен звёздочкой), полученный С. А. Пулинцом и опубликованный в [5, 7]. Это результат расчёта по двум ионограммам, одна из которых получена в аналоговом режиме наземной приёмной станцией по каналу телеметрии на частоте 137 МГц, а вторая — при одновременном зондировании ионосферы наземным ионозондом в момент, когда ОС «Мир» находилась наиболее близко к точке над наземной станцией зондирования.

На полученной на станции «Мир» ионограмме, показанной на рис. 2б, след отражения сигнала от Земли отсутствует, поэтому нельзя ничего сказать о профиле концентрации электронов ниже максимума слоя F_2 , а выше максимума обе методики расчёта дали одинаковые результаты. Аналогично и

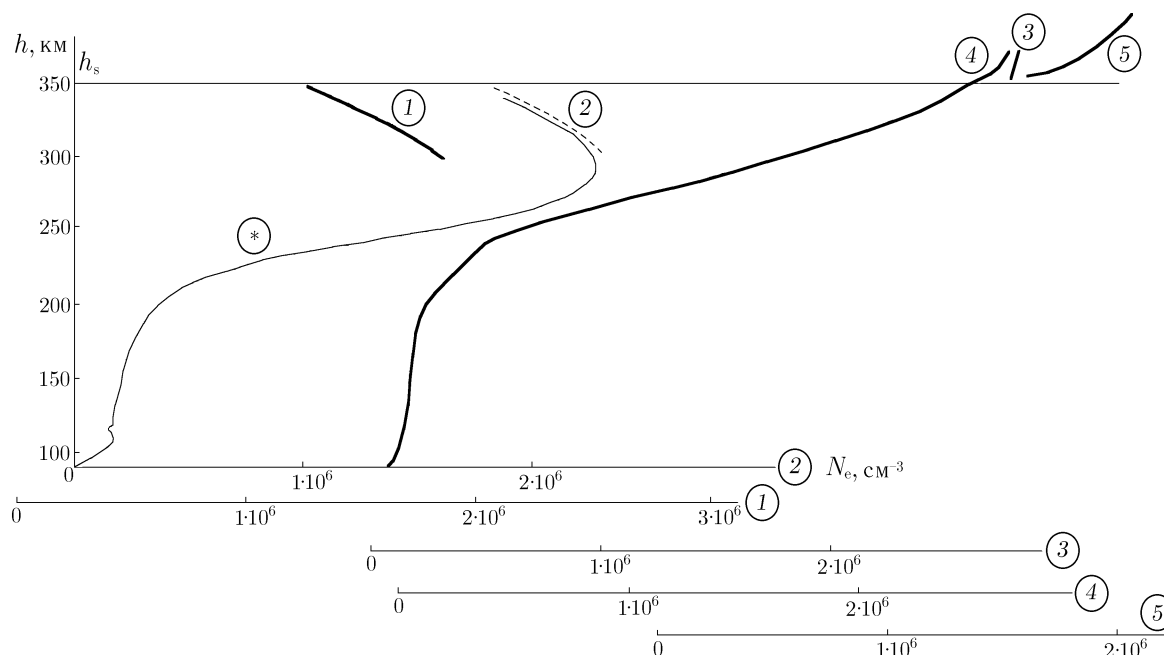


Рис. 3

на ионограмме на рис. 2а след отражения сигнала от Земли отсутствует, поэтому также ничего нельзя сказать о профиле $N_e(h)$ ниже максимума.

В момент регистрации ионограммы, приведённой на рис. 2в, станция пересекала максимум слоя F_2 . В этот момент времени для наблюдателя на о. Тайвань станция находилась за горизонтом. След отражения от максимума слоя стал почти вертикальным, частотный диапазон кривой $h'(f)$ очень мал, однако, т. к. высота спутника известна, ошибка в определении высоты максимума электронной концентрации для кривой 3 на рис. 3 совершенно незначительна и намного меньше ошибки определения $h_{\max F_2}$ с ИСЗ при обычном внешнем ($h_s \sim 1000$ км) зондировании. Временной интервал регистрации ионограмм, показанных на рис. 2б и в, менее четырёх минут.

Через 30 с после регистрации ионограммы, показанной на рис. 2в, станция находилась уже значительно ниже максимума слоя F_2 , что и отражает рис. 2г. Профиль $N_e(h)$ выше станции «Мир» до максимума слоя F_2 рассчитывался так, как описано выше. Поскольку на ионограмме на рис. 2г отчётливо виден след отражённых от Земли сигналов, по нему с помощью метода трансионосферного зондирования рассчитывался профиль концентрации ниже станции [1] (кривая 4 на рис. 3).

Следует отметить, что на рис. 3 сдвиг осей концентрации пропорционален временным интервалам между соответствующими экспериментами.

Ионограмма на рис. 2д, полученная через 3,5 минуты после ионограммы на рис. 2г, существенно не отличается от предыдущей. Спутник находился значительно ниже максимума слоя F_2 . Результаты расчёта профиля $N_e(h)$ во внутренней ионосфере от $h_{\max F_2}$ до высоты станции «Мир» рассчитаны и приведены на рис. 3 (кривая 5). Расчёт профиля электронной концентрации ниже станции «Мир» совпадает с кривой 4 и на рис. 3 не отмечен.

Утверждение о том, что ОС «Мир» находится выше или ниже максимума слоя F_2 , относительны. Высота орбиты станции существенно не менялась, следовательно, изменение её положения относительно максимума слоя F_2 свидетельствует в данном случае об увеличении высоты максимума. Менее чем за 8 минут наблюдения высота максимума слоя F_2 вдоль траектории станции изменилась примерно

на 100 км.

Таким образом, приведённый расчёт показывает, что, комбинируя различные описанные методы расчёта профилей $N_e(h)$, можно всегда определить профиль концентрации электронов ниже ИСЗ, на котором получены соответствующие ионограммы. При этом, по крайней мере, часть параметров этого профиля будет вычислена более точно по сравнению со стандартными методами расчёта профилей $N_e(h)$ при зондировании снизу или сверху, с высот порядка 1 000 км.

Дальнейшее сравнение профилей $N_e(h)$, рассчитанных по данным станции «Мир» и наземного вертикального радиозондирования, проводилось с помощью результатов наиболее аккуратной (с точки зрения авторов настоящей работы) станции вертикального зондирования на планете, расположенной под Лондоном, в Чилтоне (после переезда из Слау). Ионограммы вертикального зондирования и соответствующие профили $N_e(h)$, полученные на этой станции, публикуются в сети Интернет.

Были отобраны сеансы, когда станция «Мир» пролетала почти точно над наземным ионозондом и отсутствовали резкие изменения в ионосфере. Сравнения проводились в разное время суток при различных условиях в ионосфере и различной удалённости станции от максимума слоя F_2 . Для иллюстрации были выбраны случаи, когда станция находилась строго в максимуме слоя F_2 . Напомним, что в этом случае (см. рис. 2в) на ионограммах регистрировались только плазменные резонансы и/или частоты отсечки этого максимума, а также отражения от Земли. Высота появления плазменного резонанса, т. е. высота максимума области F_2 , определяется в этом случае высотой расположения станции «Мир», которая известна с беспрецедентной для расчёта профилей $N_e(h)$ точностью. При этом авторы учитывали, что по этой причине происходит не исследование точности расчёта профиля $N_e(h)$ по данным станции «Мир», а наоборот, исследование точности расчёта профиля электронной концентрации в ионосфере по данным одной из лучших наземных станций по сравнению с заведомо более точными данными станции «Мир».

Сравнение основных параметров профилей электронной концентрации, определяемых по данным станции «Мир» и по данным станции вертикального зондирования в Чилтоне, показало их полное совпадение для тех случаев, когда в ионосфере присутствует слой E . Однако для ночных ионограмм сравнение показало, что, если критическая частота слоя F_2 ночью определяется так же точно, как и днём, то высота максимума слоя определяется значительно менее точно. В табл. 1 приведены данные, иллюстрирующие эти различия.

Таблица 1

Местоположение станции и дата радиозондирования	Время, UT	Критическая частота f_{xF_2} необыкновенной компоненты, МГц	Высота $h_{\max F_2}$ максимума слоя F_2 , км
Чилтон (Лондон), 07.05.99	04:00:00	5,35	326
«Мир» (координаты 51,83° с. ш., 0,26° з. д.), 07.05.99	03:55:32	5,34	358,30
Чилтон (Лондон), 07.05.99	02:00:00 02:30:00	5,9 5,7	368 373
«Мир» (координаты 49,15° с. ш., 0,04° в. д.), 07.05.99	02:19:59	5,84	357,55

ВЫВОДЫ

Экспериментально показано, что расположение ионозонда на борту орбитальной станции, находящейся вблизи максимума слоя F_2 , позволяет определять основные параметры ионосферы.

Комбинация методов трансionoсферного зондирования (расчёт профиля $N_e(h)$ по следу отражения от Земли) и стандартных методов расчёта высотных профилей электронной концентрации позволяет строить распределение концентрации электронов $N_e(h)$ во внутренней ионосфере с точностью, сравнимой с точностью наземных ионозондов, а в некоторых диапазонах высотного профиля значительно превосходящей её.

Показано, что высота максимума ионосферы ночью определяется по данным наземных ионосферных станций недостаточно корректно. Можно ожидать, что использование спутникового ионозонда на обитаемой космической станции в системе постоянного оперативного контроля ионосферы позволит значительно повысить точность расчётов высоты максимума слоя F ночью на всех станциях вертикального зондирования на планете.

СПИСОК ЛИТЕРАТУРЫ

1. Danilkin N. P. // J. Atmos. Terr. Phys. 1994. V. 56, No. 11. P. 1 423.
2. Данилкин Н. П., Вайсман Г. М. // Геомагнетизм и аэрoномия. 1997. Т. 37, № 1. С. 191.
3. Данилкин Н. П. // Ионосферно-магнитная служба / Под ред. С. И. Авдюшина, А. Д. Данилова. Л.: Гидрометеoиздат, 1987. С. 46.
4. Данилкин Н. П. // Ионосферно-магнитная служба / Под ред. С. И. Авдюшина, А. Д. Данилова. Л.: Гидрометеoиздат, 1987. С. 79.
5. Pulinets S. A., Benson R. F. // Rev. Radio Sci. 1996–1999. Oxford Univ. Press, 1999. P. 711.
6. Данилкин Н. П., Мальцева О. А. Ионосферные радиоволны. Теория, алгоритмы, программы. Изд-во РГУ, 1976.
7. Pulinets S. A., Jann-Yeng Liu, Chuo Y. J., Danilkin N. P., Depuev V. Kh., Kotonaeva N. G. // Terr. Atmos. Ocean. Sci. J. 2001. V. 12, No. 3. P. 525.
8. Danilkin N. P. // Int. J. Geomagn. Aeron. 2001. V. 2, No. 3. P. 173.
9. Danilkin N. P., Kotonaeva N. G. // Int. J. Geomagn. Aeron. 2001. <http://ijga.wdcb.ru/gpo/2001/gai01372/gai01372.htm>.

Поступила в редакцию Институт прикладной геофизики РАН, г. Москва, Россия; принята в печать 4 июля 2001 г.

CALCULATION OF ALTITUDE PROFILES OF THE IONOSPHERIC ELECTRON DENSITY BASED ON THE IONOGRAMS OBTAINED ONBOARD THE “MIR” SPACE STATION

N. P. Danilkin and N. G. Kotonaeva

We present the results of calculating the altitude profiles of the electron density below and above the ionospheric maximum using the ionosonde data obtained by the “Mir” space station flying at altitudes about 350 km. It is pointed out that the data by ground-based ionosondes impede correct determination of the altitude of the nighttime ionospheric maximum.

УДК 550.388.2:551.510.536

ИСКУССТВЕННЫЕ ПЕРИОДИЧЕСКИЕ НЕОДНОРОДНОСТИ ИОНОСФЕРНОЙ ПЛАЗМЫ ПЕРСПЕКТИВНОЕ НАПРАВЛЕНИЕ В ИССЛЕДОВАНИИ ИОНОСФЕРЫ *

Е. А. Бенедиктов, В. В. Беликович, Н. В. Бахметьева, А. В. Толмачёва

Статья посвящена новому методу исследования ионосферы, разработанному в НИРФИ и основанному на создании искусственных периодических неоднородностей ионосферной плазмы (ИПН). Дан обзор способов и приведены результаты определения основных ионосферных и атмосферных параметров.

ВВЕДЕНИЕ

Во второй половине 20-го столетия был достигнут значительный прогресс в исследованиях высоких слоёв атмосферы и ионосферы — областей, где газ, окружающий Землю, частично ионизован. В этот период появились новые методы исследований: метод частичных отражений, МСТ-радары, метод некогерентного рассеяния, лазерное зондирование. Были развиты радиотехнические и оптические методы, а использование ракет и космических аппаратов позволило проводить измерения непосредственно в высоких слоях атмосферы.

Несмотря на большие успехи в исследованиях атмосферы и ионосферы, проблема их изучения и мониторинга на высотах $50 \div 200$ км остаётся актуальной до сих пор. Такая ситуация обусловлена тем, что на этих высотах не летают искусственные спутники Земли, а запуск ракет возможен только с оборудованных полигонов и стоит довольно дорого. МСТ-радары дают сведения только о движениях неоднородностей на высотах ниже 100 км, а метод некогерентного рассеяния имеет, как правило, не слишком хорошее высотное и временное разрешение в силу низкого уровня сигнала.

В последние десятилетия уделяется большое внимание исследованию средней атмосферы (высоты $10 \div 100$ км) в целом, а также входящих в этот высотный интервал мезосферы ($30 \div 90$ км) и области D ($50 \div 90$ км). Эта область атмосферы играет важную роль в работе радиосредств в традиционных диапазонах длинных, средних и коротких волн, а также во взаимодействии термосферы, подверженной влиянию всплесков солнечной активности, и тропосферы, прилегающей непосредственно к Земле. В этой связи разработка новых высокоинформативных методов исследования нижней ионосферы является актуальной и важной задачей. Статья посвящена новому методу исследования ионосферы, разработанному в НИРФИ и основанному на создании искусственных периодических неоднородностей ионосферной плазмы (ИПН).

1. ИСТОРИЯ ОТКРЫТИЯ ИСКУССТВЕННЫХ ПЕРИОДИЧЕСКИХ НЕОДНОРОДНОСТЕЙ

Впервые мысль об образовании периодических неоднородностей в нижней ионосфере под действием мощной стоячей радиоволны была высказана И. М. Виленским [1] в 1970 г. Он предполагал, что эти неоднородности являются неоднородностями температуры электронов. Затем в 1973 г. в работе И. М. Виленского и В. В. Плоткина [2] была рассмотрена возможность образования неоднородностей электронной концентрации вследствие зависимости эффективного коэффициента рекомбинации

* Статья подготовлена к юбилею НИРФИ.

от температуры электронов. В 1972 г. Т. Н. Селига в работе [3] указал на возможность создания искусственных периодических неоднородностей плазмы при отражении мощной радиоволны от F -области ионосферы, однако конкретные механизмы образования ИПН рассмотрены не были. Эти теоретические работы предшествовали экспериментальному открытию искусственных периодических неоднородностей ионосферной плазмы.

Впервые ИПН были обнаружены в 1975 г. в экспериментах по модификации ионосферы мощным высокочастотным радиоизлучением [4]. Это направление исследований в НИРФИ было инициировано Г. Г. Гетманцевым. Под его руководством с целью изучения воздействия мощного радиоизлучения на ионосферную плазму в 1975 г. был создан нагревный стенд «Ястреб». В первых экспериментах, проводимых под руководством Г. Г. Гетманцева, было обнаружено дискретное рассеяние пробных радиоволн необыкновенной поляризации на частоте 5,6 МГц при воздействии на ионосферу радиоизлучением обыкновенных радиоволн с частотой 4,6 МГц. Авторы интерпретировали это явление как рассеяние от искусственных периодических неоднородностей [5]. В дальнейшем было установлено, что ИПН формируются в поле стоячей волны, возникающей вследствие интерференции падающей на ионосферу и отражённой от неё волн, в высотном интервале от 60 км до точки отражения мощной радиоволны от ионосферы.

Благодаря теоретическим и экспериментальным исследованиям [6–8] удалось понять основные физические процессы, приводящие к образованию ИПН. Эти исследования позволили разработать новые методы диагностики ионосферной плазмы и атмосферы. С 1986 г. в НИРФИ ведутся целенаправленные исследования по диагностике ионосферной плазмы с помощью ИПН.

Долгое время НИРФИ был практически единственной организацией, ведущей исследования по данной тематике. Наконец в 1984 г. в печати появилась работа [9], в которой сообщалось, что на нагревном стенде в Аресибо (Пуэрто-Рико, США) наблюдались ИПН в области F ионосферы. В 90-х годах сотрудники НИРФИ развернули наблюдения ИПН на стенде «Гиссар» Института астрофизики Таджикской академии наук и получили интересные результаты [10, 11].

В последние десять лет интерес к исследованию ИПН за рубежом заметно усилился. В 1993 и 1996 гг. при участии сотрудника НИРФИ Н. П. Гончарова были проведены первые исследования ИПН на стенде EISCAT в Тромсё (Норвегия) [12, 13]. В 1995 г. наблюдения ИПН в авроральной ионосфере осуществлены на нагревном стенде HIPAS, расположенном на Аляске [14].

2. СУТЬ МЕТОДА И СПОСОБЫ ДИАГНОСТИКИ ПАРАМЕТРОВ ИОНОСФЕРЫ И АТМОСФЕРЫ

ИПН формируются в поле мощной стоячей радиоволны, образующейся в результате интерференции падающей на ионосферу и отражённой от неё волн. В пучностях стоячей волны происходит нагрев электронного газа, в результате чего возникает периодическая структура с пространственным периодом Λ , равным половине длины λ_1 мощной радиоволны. Температурные неоднородности, в свою очередь, формируют неоднородности электронной концентрации.

Процессы образования и релаксации ИПН после выключения нагрева определяются физическими условиями, характерными для данной области ионосферы: изменениями состава, плотности и температуры атмосферы, а также степени диссоциации и ионизации с высотой, что позволяет применять этот метод для диагностики ионосферы.

Благодаря теоретическим и экспериментальным исследованиям, проведённым в НИРФИ, удалось понять основные физические процессы, приводящие к образованию ИПН, и изучить механизмы их формирования в разных слоях ионосферы. Показано, что на высотах области D образование ИПН обусловлено температурной зависимостью коэффициента прилипания электронов к молекулам кислорода, в области E (высоты 90–130 км) неоднородности создаются вследствие диффузионного перераспределения плазмы под действием избыточного давления электронной компоненты, нагретой в пуч-

ностях стоячей радиоволны. Наконец, в области F ионосферы ИПН образуются под действием ponderomotorной стрикционной силы с возбуждением ионного звука [15].

Понимание процессов образования ИПН на разных высотах привело к разработке способов определения целого ряда основных параметров ионосферной плазмы и атмосферы. Все они основаны на создании ИПН при воздействии на ионосферу мощной радиоволной, локации их пробными радиоволнами, приеме сигналов, рассеянных неоднородностями, измерении их характеристик (амплитуды, фазы и их изменений с высотой и во времени) и определении по ним параметров ионосферной плазмы.

Рассеяние пробных радиоволн на ИПН обладает резонансными свойствами, т. е. интенсивный рассеянный сигнал появляется только в случае, когда волны, рассеянные отдельными неоднородностями, складываются синфазно.¹ Это условие интенсивного рассеяния (условие Вульфа—Брэгга) имеет вид

$$\lambda_2 = 2\Lambda \cos \theta, \quad (1)$$

где Λ — пространственный период неоднородностей, λ_2 — длина волны излучения радиолокатора в плазме, θ — угол падения волны на неоднородности. В случае обратного рассеяния $\theta = 0$, и $\lambda_2 = 2\Lambda$. Т. к. $\Lambda = \lambda_1/2$, то приходим к условию пространственного синхронизма длин волн нагревного и пробного передатчиков:

$$\lambda_1 = \lambda_2, \quad (2)$$

которое можно выполнить в двух случаях, определяющих два способа регистрации ИПН.

Один из них состоит в том, что для создания ИПН и их локации используют волны одной и той же частоты и поляризации, т. е. $f_1 = f_2$. Эксперимент в этом случае ставится следующим образом. Нагревный стенд излучает вверх мощную радиоволну на частоте f_1 , которая ниже критической частоты ионосферы. Отражаясь от ионосферы, мощная волна интерферирует с падающей и образует стоячую волну, в поле которой создаются неоднородности электронной концентрации. После этого мощный передатчик нагревного стенда выключается, и на той же частоте начинает работать импульсный радиолокатор. Излученные радиоимпульсы рассеиваются ИПН и принимаются приемником. Амплитуда и фаза рассеянного сигнала регистрируются в зависимости от высоты и времени в течение всего периода релаксации ИПН.

Таким образом, рассеянные сигналы приходят со всех высот, где создаются неоднородности, т. к. равенство (2) выполняется независимо от величины электронной концентрации. В этом случае нагревный передатчик можно использовать и в качестве локационного, переводя его в импульсный режим.

Другой способ выполнения равенства (2) — применение для создания и локации ИПН радиоволн разных частот и поляризаций. В этом случае условие (2) приводит к уравнению:

$$f_1 n_1^{o,x} = f_2 n_2^{x,o}, \quad (3)$$

где n^o и n^x — показатели преломления обыкновенной и необыкновенной волн соответственно. Так как n^o и n^x зависят от электронной концентрации и магнитного поля, то соотношение (3) следует рассматривать как уравнение, связывающее f_1 , f_2 , электронную плазменную частоту $f_0 = \sqrt{e^2 N / (\pi m)}$ и гирочастоту электронов f_H . Выбрав рабочие частоты соответствующим образом, можно наблюдать рассеяние только с высот, где электронная концентрация удовлетворяет уравнению (3). На этом основан так называемый частотный способ определения электронной концентрации [15]. Принципиальной особенностью этого способа, связанной с использованием эффекта рассеяния, является возможность измерения высотного профиля электронной концентрации $N(h)$ в долине между слоями E и F . Разумеется, этот метод позволяет определить профили $N(h)$ в D - и F -областях. При этом верхний предел

¹ Это обстоятельство обусловило другое часто употребляемое название метода — метод резонансного рассеяния радиоволн на ИПН.

интервала измерений определяется высотой отражения мощной волны, а нижний — шириной спектра Δf_2 зондирующего импульса.

При $|f_1 - f_2| \leq \Delta f_2$ условие пространственного синхронизма выполняется почти на всех высотах области D , и определить высоту отражения рассеянного сигнала становится трудно. Преодоление этой трудности возможно путём измерения фазы рассеянного сигнала, т. к. высотный профиль $N(h)$ пропорционален изменению фазы сигнала по высоте. На этом основан фазовый способ измерения электронной концентрации.

Таким образом, на основе экспериментальных и теоретических представлений о формировании ИПН были разработаны и реализованы способы определения параметров ионосферной плазмы и атмосферы, основными среди которых являются:

- профиль электронной концентрации в интервале от нижней части области D до высоты максимума F -слоя;
- скорость вертикального движения атмосферы в D - и E -областях;
- температура и плотность атмосферы на высотах E -области;
- относительные концентрации атомарного и возбуждённого кислорода в D -области;
- электронная и ионная температуры в F -области;
- параметры турбулентности на высотах $90 \div 105$ км: вертикальная компонента турбулентной скорости, коэффициент турбулентной диффузии и высота турбопаузы;
- параметры внутренних гравитационных волн и их спектральные характеристики;
- характеристики спорадического слоя E .

3. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ ИОНОСФЕРЫ И АТМОСФЕРЫ С ПОМОЩЬЮ ИСКУССТВЕННЫХ ПЕРИОДИЧЕСКИХ НЕОДНОРОДНОСТЕЙ

3.1. Экспериментальная база

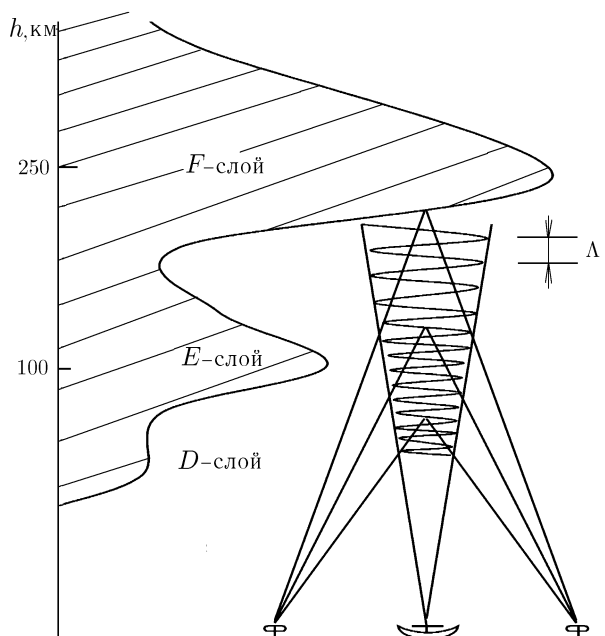


Рис. 1. Схема создания искусственных периодических неоднородностей

Для диагностики ионосферы с помощью ИПН необходим мощный передатчик — нагревный стенд, создающий в ионосфере электромагнитное поле, соизмеримое с плазменным [15]. Экспериментальные результаты, представленные ниже, получены с помощью нагревных стендов «Ястреб» (эффективная излучаемая мощность $P = 20$ МВт) и «Сура» ($P = 150 \div 280$ МВт). Для диагностики ИПН используются импульсные радиолокаторы, работающие в том же частотном диапазоне, что и нагревный стенд. Наиболее удобными для этой цели являются установки для исследования ионосферы методом частичных отражений и ионозонды с повышенным потенциалом. На рис. 1 схематично показано формирование ИПН при нагреве ионосферы.

Целенаправленные исследования ионосферы методом ИПН были начаты в 1986 г. Наиболее продолжительный цикл измерений, во время

которого был получен наибольший объём экспериментальных данных, проведён в начале 90-х годов, с сентября 1990 по январь 1992 года. Наблюдения проводились в дневное время в течение 5–10 дней каждый месяц за исключением летнего сезона.

В этих экспериментах ИПН создавались нагревным стендом «Ястреб», который работал циклами длительностью 10–15 с и излучал волны необыкновенной поляризации на частоте $f = 5,828$ или $5,545$ МГц. Мощные радиоволны излучались в течение $4 \div 5$ с, во время паузы в работе стенда осуществлялась диагностика ИПН. Для этого применялась установка для исследования ионосферы методом частичных отражений, расположенная вблизи нагревного стенда, со следующими параметрами: мощность передатчика при излучении импульса приблизительно 500 кВт, коэффициент усиления приёмно-передающей антенны $G \approx 200$. При измерении профилей $N(h)$ частота пробной волны изменялась от 5,455 до 5,75 МГц с шагом 20 кГц; принимались обратно рассеянные сигналы обыкновенной поляризации и измерялись высоты, на которых выполнялось условие (2) пространственного синхронизма нагревной и пробной радиоволн. Длительность зондирующего импульса составляла 25 мкс. В высотном интервале $90 \div 140$ км амплитуда рассеянного сигнала в $10 \div 100$ раз превышала уровень шумов.

Измерения высотных профилей электронной концентрации проводились, как правило, несколько раз в течение дня, а в остальное время измерялись амплитуды и фазы сигналов, рассеянных ИПН. Для этого использовались нагревные и пробные волны с одинаковой частотой и поляризацией. Для регистрации амплитуды и фазы обратно рассеянных сигналов применялась специально разработанная приёмно-регистрирующая аппаратура. Запись велась одновременно с шести высот с шагом $3 \div 5$ км. Полученные данные обрабатывались по оригинальной методике [15].

Летом 1996 года была проведена модернизация приёмно-регистрирующего комплекса и введена цифровая регистрация рассеянного сигнала, что позволило уменьшить шаг регистрации по высоте и производить компьютерную обработку данных в реальном времени. На рис. 2 показана временная диаграмма работы комплекса в двух режимах: при измерении высотного профиля электронной концентрации (а) и характеристик рассеянного сигнала (б), а также блок-схема аппаратного комплекса (в).

Методика измерений построена следующим образом. Для более точного определения атмосферных параметров: вертикальной V и турбулентной V_t скоростей, температуры T и плотности ρ требуется знание профиля электронной концентрации, поэтому первым шагом является измерение $N(h)$, которое является более сложным в техническом отношении, т. к. нагревная и диагностическая (пробная) волны должны иметь разные частоты и поляризации. Остальные параметры можно получить одновременно из одного эксперимента, т. к. для их определения используется только высотно-временная зависимость амплитуды $A(h, t)$ и фазы $\phi(h, t)$ рассеянного сигнала. Исходя из этого, алгоритм изме-

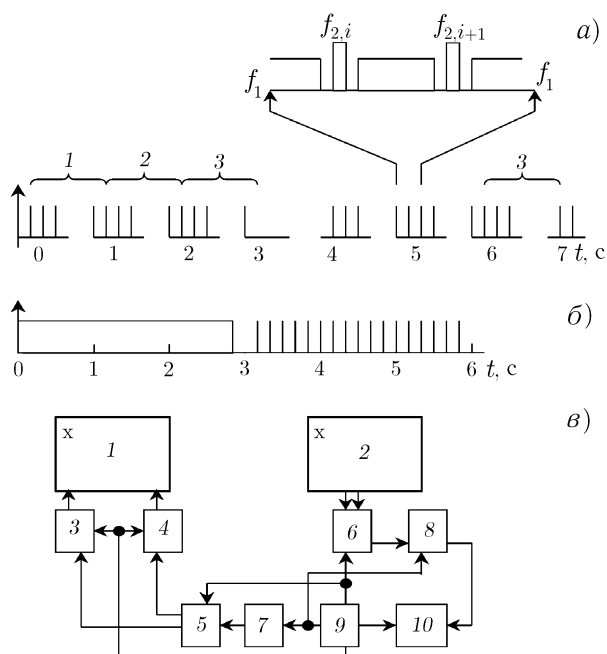


Рис. 2. Временная диаграмма работы комплекса в двух режимах (а, б) и блок-схема аппаратного комплекса (в): 1, 2 — передающая и приёмная антенны; 3, 4 — усилители мощности; 5, 6 — поляризаторы; 7 — задающий генератор; 8 — приёмник; 9 — управляющая ЭВМ; 10 — ЭВМ для регистрации, обработки и хранения первичной информации

рения должен состоять из двух частей: измерения $N(h)$ и измерения амплитуды и фазы рассеянного сигнала в процессе релаксации ИПН.

Измерение $N(h)$ начинается с излучения пробных радиоимпульсов обыкновенной поляризации на частоте $f_{2,i}$ и регистрации амплитуды помеховых сигналов $A_i(h)$. Поскольку для определения $N(h)$ требуется изменение частоты в диапазоне $(f_1; f_1 + 1 \text{ МГц})$, частота $f_{2,i}$ меняется от импульса к импульсу с шагом $\delta f \simeq 20 \text{ кГц}$. При частоте повторения $F \simeq 50 \text{ Гц}$ для пробега частотного интервала в 1 МГц потребуется приблизительно 50 импульсов, при этом длительность цикла составляет около 1 с. Для получения надёжных данных цикл зондирования повторяется несколько раз (например трижды, как показано на рис. 2а, где i — номер импульса в цикле). Далее излучается нагревная радиоволна необыкновенной поляризации на частоте f_1 , и в течение $t \geq 1 \text{ с}$ создаются ИПН. После этого циклы зондирования на частоте $f_{2,i}$ обыкновенной компоненты повторяются, и регистрируется амплитуда рассеянного сигнала $A_i^*(h)$. По окончании циклов зондирования нагревная волна выключается. Для определения высоты рассеянного сигнала определяются величины $y_i(h) = \Sigma A_i^*(h) - \Sigma A_i(h)$ и для каждой частоты $f_{2,i}$ определяется высота, на которой $y_i(h)$ достигает максимума, уверенно превышающего уровень шумов. Здесь знаком Σ обозначено суммирование амплитуд сигналов по всем импульсам цикла. В результате измеряется действующая высота $h_d(f_i)$, которая используется для расчёта $N(h)$ по формулам или по соответствующей таблице. Таким образом, профиль $N(h)$ можно определить за время порядка 10 с.

Временной алгоритм работы комплекса при измерении амплитуды и фазы рассеянного ИПН сигнала представлен на рис. 2б. Он начинается с излучения нагревной волны необыкновенной поляризации на частоте f_1 в течение приблизительно 3 с. Затем нагревная волна выключается, излучаются короткие диагностические импульсы той же частоты и поляризации. Длительность импульсов составляет $20 \div 50 \text{ мкс}$, частота повторения $F \simeq 50 \text{ Гц}$. Рассеянные сигналы от каждого зондирующего импульса принимаются и подаются на амплитудный и фазовый детекторы, после которых регистрируются зависимости $A(h, i)$ и $\phi(h, i)$ (или синусная и косинусная составляющие сигнала). Зондирование продолжается около 3 с. Весь цикл повторяется $3 \div 5$ раз. Длительность циклов нагрева и измерения не превышает 10 с, поэтому суммарное время работы по второй части алгоритма составляет примерно 50 с, а общее время измерений по алгоритму в целом составит около 1 мин.

Подобный алгоритм работы реализован с помощью аппаратного комплекса, блок-схема которого изображена на рис. 2в. В данном комплексе совмещены функции нагревного стенда и диагностической установки.

На рис. 3а приведён пример высотно-временной зависимости амплитуды сигнала, рассеянного ИПН. Хорошо видно уменьшение амплитуды со временем в интервале высот $55 \div 140 \text{ км}$ и быстро затухающие периодические колебания в области F .

В августе 1996 года на стенде «Сура» в загородной лаборатории НИРФИ «Васильсурск» был проведён эксперимент с использованием модернизированной аппаратуры. Воздействие на ионосферу осуществлялось радиоволной необыкновенной поляризации на частоте $4,7 \text{ МГц}$ при эффективной излучаемой мощности $60 \div 100 \text{ МВт}$ в течение 3 с с паузой продолжительностью 12 с. После каждого цикла нагрева в течение двух секунд излучались зондирующие импульсы длительностью 20 мкс с частотой повторения 40 Гц. Рассеянные сигналы принимались антенной, имеющей 12 синфазных диполей на каждой линейной поляризации и выделяющей необыкновенную компоненту излучения. Усиление сигналов осуществлялось приёмником с полосой пропускания 50 кГц. Сигналы, рассеянные ИПН, регистрировались с высотным шагом 1,2 км в виде двух квадратурных компонент, кодировались десятирядным аналого-цифровым преобразователем и подвергались дальнейшей компьютерной обработке.

Следующий этап измерений на стенде «Сура» был проведён в августе 1999 г., включая день солнечного затмения 11 августа. Каждый сеанс измерений продолжался 20 с. В течение первых 3 с осуществлялся нагрев ионосферы необыкновенной радиоволной на частоте $f = 5,67 \text{ МГц}$ при эффективной

излучаемой мощности $P \approx 130$ МВт 10 августа и $P \approx 60$ МВт 11 и 12 августа. За трёхсекундным нагревом следовала пауза длительностью 17 с, в первые 4 с которой излучались пробные импульсы той же частоты длительностью $25 \div 50$ мкс с частотой повторения 100 Гц. Усиление сигналов осуществлялось приёмником с полосой пропускания 40 кГц.

В августе 2000 г. на стенде «Сура» был проведён эксперимент по исследованию нижней ионосферы в восходно-заходные часы. Для создания неоднородностей использовались две секции стенда, которые излучали радиоволны необыкновенной поляризации на частоте 5,6 МГц. Сигналы регистрировались в течение первых 4 с после прекращения нагрева с высотным шагом 1,4 км в виде двух квадратурных компонент и кодировались 12-ти разрядным аналого-цифровым преобразователем. Работа проводилась в период с 20:00 до 23:00 MSK 16 августа и с 03:50 до 06:30 MSK 17 августа. На каждой высоте рассчитывались амплитуда и фаза сигнала, временные зависимости которых затем аппроксимировались линейными функциями вида $\ln A(t) = \ln A_0 - t/\tau$, $\phi(t) = \phi_0 + 4\pi Vt/\lambda$. В этих выражениях λ — длина волны зондирующего сигнала в плазме, τ — время релаксации ИПН, которое характеризует время жизни искусственных неоднородностей после прекращения нагрева, V — скорость их вертикального движения [15]. Параметры аппроксимации определялись методом простой линейной регрессии с весовой функцией

$$F = \exp(-2t/\tau_0),$$

которая согласно [16] позволяет минимизировать дисперсию рассчитываемых параметров (здесь τ_0 — значение параметра в первом приближении). Уровень шума в разное время составлял $10 \div 15$ дБ. При дальнейшей обработке по зависимости $\tau(h)$ согласно [17, 18] определялись температура T и плотность ρ атмосферы на высотах $95 \div 110$ км, вертикальная компонента скорости турбулентных движений вблизи турбопаузы [19, 20], а по зависимости $\phi(t)$ — вертикальная скорость V плазмы на каждой высоте [21, 22]. Пример комплексной диагностики параметров ионосферы приведён на рис. 3б.

Для контроля общего состояния ионосферы каждые 5 минут с помощью АИС снимались ионограммы вертикального зондирования.

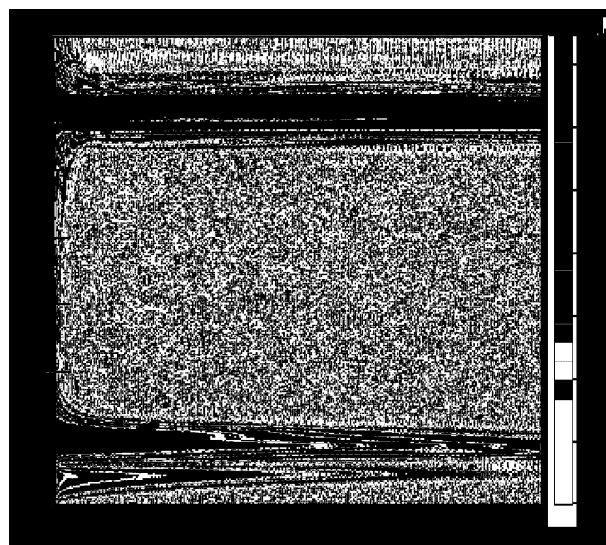


Рис. 3а

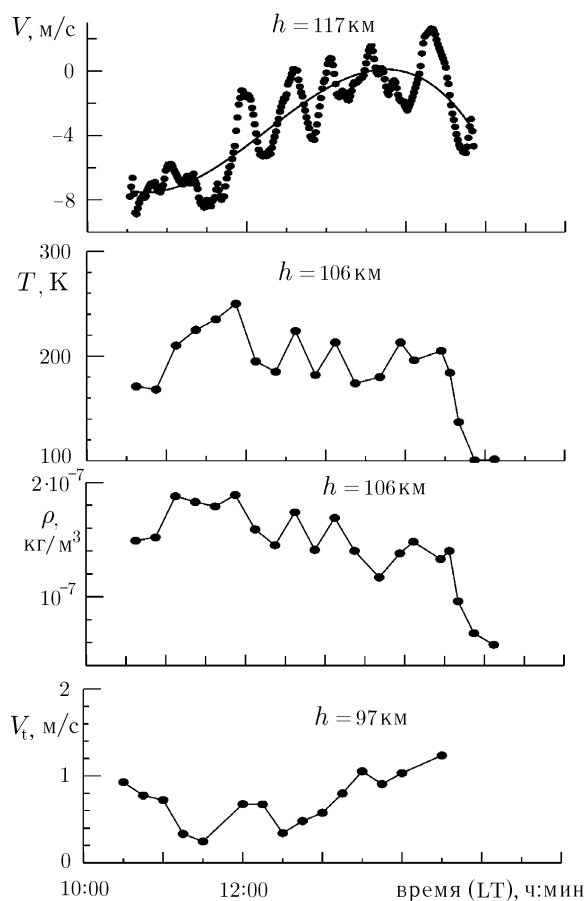


Рис. 3б. Пример комплексной диагностики параметров ионосферы 27.02.91

3.2. Результаты измерений высотных профилей электронной концентрации

3.2.1. Регулярный E -слой и спорадические слои

За время измерений был получен большой массив данных — более чем 350 высотных профилей электронной концентрации. На рис. 4 приведены изменения во времени профиля электронной концентрации $N(h)$ на интервале с 09:00 до 12:35 MSK 17.12.91. В 09:00 наблюдалось плавное возрастание концентрации с высотой, в 09:35 появилась долина (межслоевая впадина) на высотах между 120 и 140 км, а затем возникла более тонкая структура в области долины: появился спорадический слой на высоте 115 км. На рис. 4 хорошо видны два спорадических слоя: нижний с максимумом на высоте 115 км и верхний с максимумом на высоте приблизительно 140 км, в межслоевой впадине (высоты действующие).

Проанализированы временные изменения профиля $N(h)$, обусловленные следующими факторами: изменением зенитного угла χ Солнца, определяющим обычные суточные и сезонные вариации, магнитными возмущениями и нерегулярными изменениями состава атмосферы и ионосферы. Последнее обстоятельство вызывает такие явления, как зимняя аномалия и спорадические слои, и тесно связано с динамикой нижней термосферы. Отметим, что заметные изменения характера высотной зависимости электронной концентрации появлялись в сильно возмущённые в планетарном масштабе дни (при планетарном трёхчасовом индексе $K_p \geq 5$). При этом в ионосфере, как правило, появлялось несколько дополнительных слоёв.

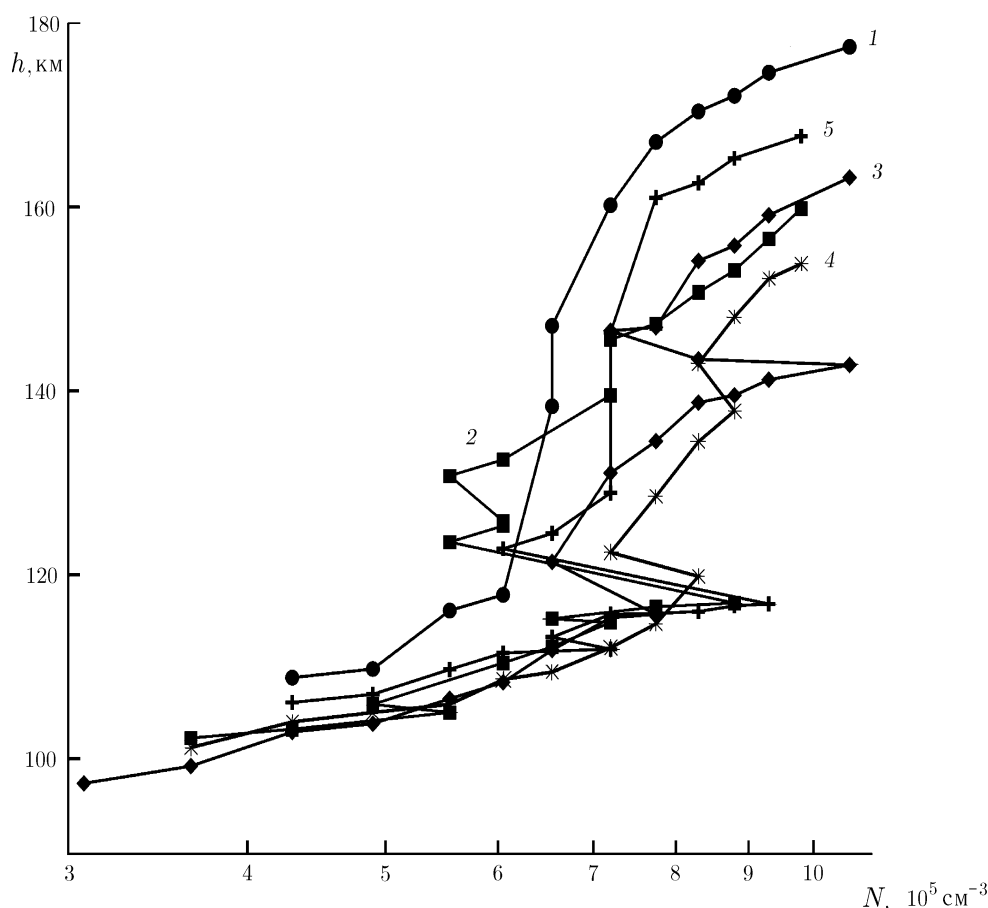


Рис. 4. Высотные профили электронной концентрации, полученные 17.12.91 в различные моменты местного времени: кривая 1 — в 09:00; 2 — в 09:35; 3 — в 10:00; 4 — в 10:45; 5 — в 12:35

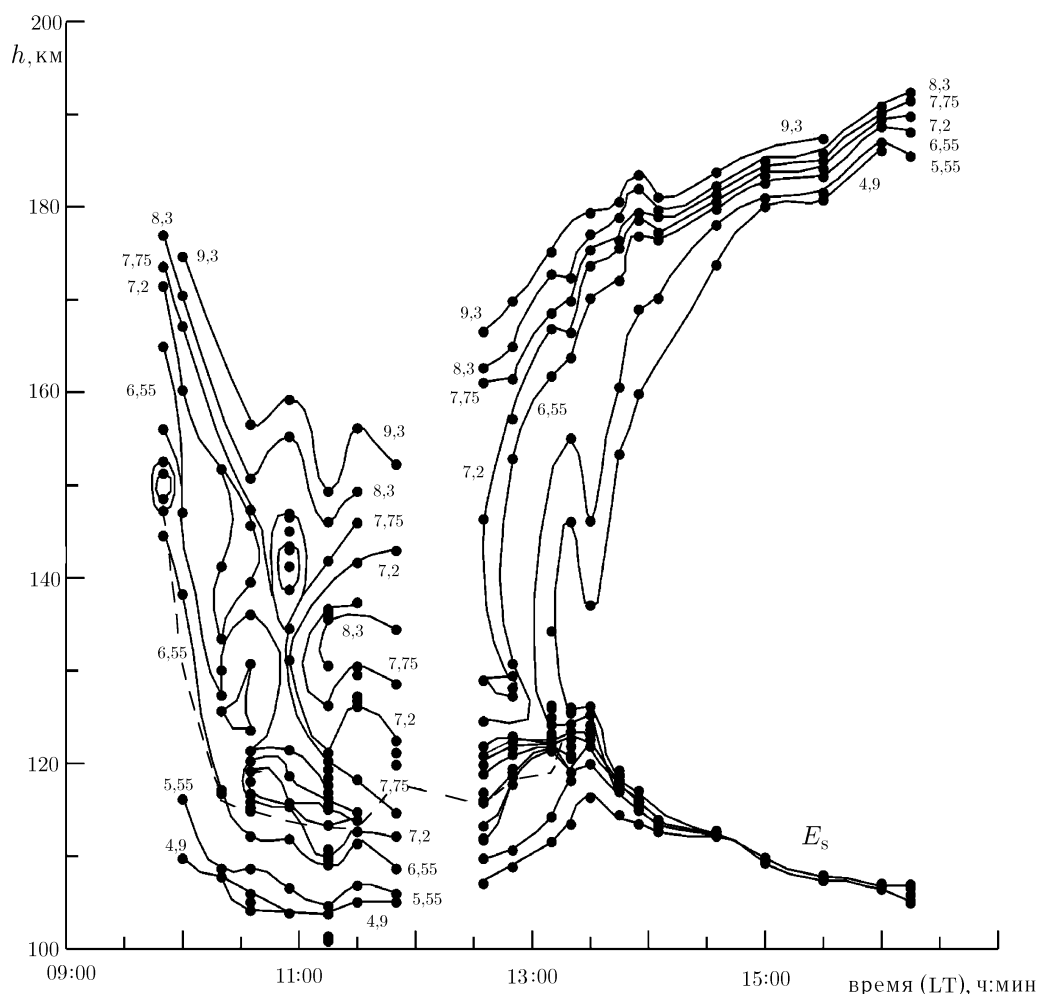


Рис. 5. Изолинии электронной концентрации $N [10^4 \text{ см}^{-3}]$ 17.12.91 при $K_p = 5,4$. Пунктирная кривая даёт изменение высоты максимума E -слоя

Измерения высотных профилей электронной концентрации позволяют получить сведения о спорадических слоях, расположенных ниже максимума E -слоя [15, 23]. По результатам измерений, выполненных в 1990—1992 гг., почти в половине всех полученных профилей отмечались тонкие спорадические слои E_s , расположенные ниже или вблизи максимума регулярного E -слоя. Большинство из них наблюдались на высотах $90 \div 115$ км. Последовательные измерения профилей $N(h)$ дают возможность наблюдать динамику спорадических слоёв и оценивать скорость их смещения по высоте. На рис. 5, например, виден ночной слой E_s , сформировавшийся перед заходом Солнца и наблюдавшийся более двух часов. Толщина слоя уменьшалась с 3,2 км в 15:00 LT до менее чем 1 км между 16:00 и 17:15 LT, а максимальная электронная концентрация была равна $N_{\max} = (8,8 \div 9,3) \cdot 10^4 \text{ см}^{-3}$. При этом высота максимума уменьшалась от 112 до 105 км. Таким образом, ИПН представляет собой удобный способ диагностики спорадических слоёв.

3.2.2. Исследование долины между слоями E и F

Межслоевая впадина (долина) между слоями E и F по-прежнему является наименее изученной областью ионосферы. Метод вертикального зондирования неприменим здесь из-за экранирования до-

лины E -слоем, ракетные измерения проводятся эпизодически, а интенсивность некогерентного рассеяния от этого высотного интервала существенно ниже, чем от F -области ионосферы из-за значительно меньшей концентрации электронов, что существенно снижает высотное и временное разрешение метода. Привлекательность использования ИПН в этом случае трудно переоценить.

Методом ИПН в осенне-зимние периоды 1990–1992 гг. получено более 80 профилей $N(h)$ [23, 24], охватывающих высотный интервал с долиной, которая наблюдалась в половине случаев. По этим данным экспериментально изучена тонкая структура долины и обнаружены дополнительные слои электронной концентрации внутри неё. Оказалось, что параметры слоёв, лежащих ниже 135 км, определяются в основном высотой Солнца и связаны с динамикой регулярной E -области, а их нижняя граница испытывает нерегулярные вариации из-за влияния естественных неоднородностей и расслоений E -области ионосферы.

3.3. Скорость вертикального движения плазмы в нижней ионосфере

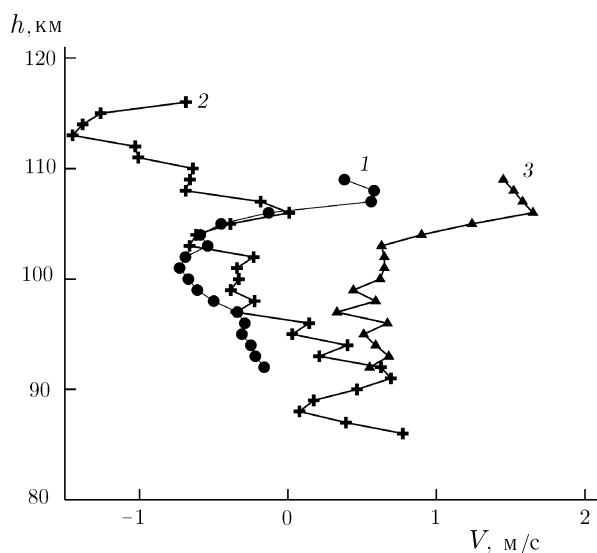


Рис. 6. Высотные профили вертикальной скорости ионосферной плазмы для осенних месяцев 1991 г.: 1 — сентябрь, 2 — октябрь, 3 — ноябрь

Вертикальные движения являются одной из составляющих общей циркуляции атмосферы. До настоящего времени этот тип движений остаётся наименее изученным.

В процессе релаксации искусственные периодические неоднородности увлекаются движением нейтрального газа, поэтому наблюдения доплеровского смещения частоты рассеянного сигнала могут быть использованы для измерения скорости вертикального движения. Однако непосредственно измерять доплеровский сдвиг F_D частоты сигнала, рассеянного ИПН, невозможно, т. к. характерное время жизни неоднородностей меньше периода $1/F_D$. Преодолеть эту трудность позволяет измерение фазы рассеянного сигнала [21]. В этом случае

$$V = \frac{\lambda}{4\pi} \frac{d\Phi}{dt}. \quad (4)$$

По этой формуле определяется скорость вертикального движения плазмы, которая вплоть до высоты 120 км совпадает со скоростью движения нейтральной компоненты.

3.3.1. Сезонно-суточные вариации и волновые движения

Отметим некоторые общие закономерности вариаций скорости V вертикального движения ионосферной плазмы по результатам измерений 1990–1992 гг. [22]. Было установлено, что среднемесячные значения V на высотах ниже 90 км составили менее 1 м/с, возрастая до 5 м/с с увеличением высоты до 120 км. На высотах, больших 90 км, преобладали движения вверх (более 70 % всех данных). На высотах ниже 65 км все полученные значения V соответствовали движению вниз.

На рис. 6 показаны высотные профили вертикальной скорости ионосферной плазмы для трёх осенних месяцев 1991 г., дающие представление о среднем характере изменений величины и направления вертикальной скорости с высотой. Положительные значения V соответствуют движению вниз.

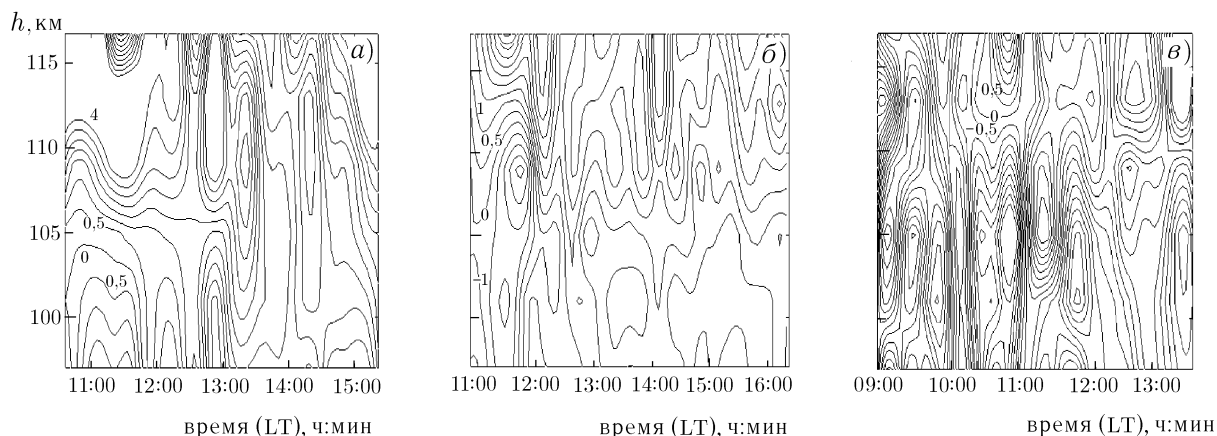


Рис. 7. Изолинии скорости вертикального движения V по измерениям 1991 г.: а) — 27.02.91, б) — 15.03.91, в) — 06.12.91

Временные вариации вертикальной скорости обусловлены рядом причин, одной из которых является постоянное наличие в нижней ионосфере волновых процессов различной природы, в т. ч. внутренних гравитационных волн (ВГВ) и приливов. Измерения V с помощью ИПН показывают, что мгновенные значения скорости на высотах больше 90 км могут составлять до десяти и более метров в секунду, на меньших высотах — до нескольких метров в секунду. Сравнительно большие значения вертикальной скорости обусловлены волновыми движениями в атмосфере [25]. На рис. 7 хорошо видна волновая структура вариаций вертикальной скорости на всём интервале.

Пример комплексной диагностики ионосферных параметров по результатам измерений 27.02.91 на разных высотах, приведённый на рис. 3б, показывает волнообразные изменения временных вариаций вертикальной V и турбулентной V_t скоростей, температуры T и плотности ρ атмосферы [15, 26]. То же самое можно видеть на рис. 5, где показаны изолинии электронной концентрации N [23].

3.3.2. Вертикальные движения плазмы и спорадический слой E

Результаты измерений скорости вертикального движения плазмы методом ИПН использованы для расчётов характеристик среднеширотного спорадического слоя E : эффективного коэффициента рекомбинации и относительного содержания метеорных и атмосферных ионов в слое [27]. В рамках представлений о формировании среднеширотного спорадического слоя E из долгоживущих металлических ионов предложен способ определения их концентрации и эффективного коэффициента рекомбинации в максимуме слоя E_s . Указанный способ основан на предположении о преобладающей роли атмосферных ионов в фотохимических процессах и использует частотные характеристики слоя E_s и области E , полученные из ионограмм вертикального зондирования ионосферы, и скорости вертикальных движений плазмы, измеренные с помощью ИПН.

3.4. Исследование атмосферной турбулентности

Анализ высотной зависимости времени релаксации сигнала, рассеянного ИПН, позволяет с помощью простых соотношений определить одну из важнейших характеристик атмосферы — вертикальную компоненту турбулентной скорости [19], которая на высотах D - и E -областей оказывает существенное влияние на характеристики и свойства ионосферы.

Мелкомасштабные движения, обусловленные турбулентностью, влияют на характеристики рассеянного сигнала, т. к. нарушают упорядоченную структуру ИПН. Следствием этого является нарушение

экспоненциального хода времени релаксации с высотой, который обусловлен амбиполярной диффузией. Измеряя в эксперименте время релаксации τ сигнала и зная характерное время τ_d амбиполярной диффузии, легко найти «турбулентное» время [19]

$$\tau_t = (\tau^{-1} - \tau_d^{-1})^{-1} \quad (5)$$

и рассчитать по нему характерную турбулентную скорость:

$$V_t/\sqrt{2} = 1/(\sqrt{2} K \tau_t), \quad (6)$$

где $K = 4\pi/\lambda$ — волновое число стоячей волны. По измерениям 1991 г. получено [20], что средние за минуту значения V_t изменяются от нуля до 5 м/с. Часто наименьшие значения V_t наблюдались вблизи местного полдня. Наблюдались также большие вариации V_t в объёме рассеяния, что указывает на нестационарность турбулентного процесса в относительно малом объёме (несколько десятков кубических метров). В среднем за шести-семичасовой сеанс наблюдений значения V_t на высотах, меньших высоты h_t турбопаузы, составили около $1 \div 2$ м/с, уменьшаясь с приближением к ней до нескольких десятков сантиметров в секунду (см. также рис. 3б).

3.5. Температура и плотность атмосферы на высотах области E

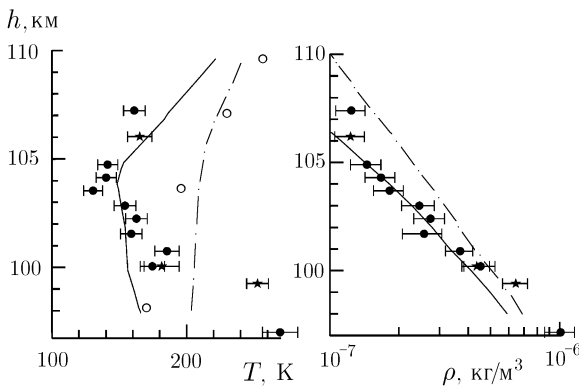


Рис. 8

Установлено, что релаксация ИПН на высотах E-области ионосферы происходит вследствие амбиполярной диффузии. В этом случае характерное время уменьшения амплитуды неоднородностей в e раз, равное $\tau = (K^2 D)^{-1}$, определяется пространственным периодом неоднородностей и коэффициентом амбиполярной диффузии $D = \kappa(T_e + T_i)/(M\nu_{im})$, где T_e и T_i — электронная и ионная температуры соответственно, κ — постоянная Больцмана, M — масса иона, ν_{im} — частота столкновений ионов с нейтральными частицами. В свою очередь, D зависит от температуры ионизованных компонент и частоты ν_{im} . Последняя пропорциональна плотности атмосферы: $\nu_{im} = \beta\rho/M$, где β — численный коэффициент. Эти рассуждения показывают,

что характерное время релаксации ИПН связано с температурой и плотностью атмосферы, что и позволяет определять эти параметры. Подробное изложение метода можно найти в работах [15, 21].

В 1990–1992 гг. был проведён цикл измерений температуры и плотности атмосферы на высотах E-слоя данным методом. Отмечены вариации зависимостей $T(t)$ и $\rho(t)$ с периодами от $15 \div 30$ минут до $2 \div 3$ часов, связанные с распространением атмосферных волн (см. рис. 3б).

На рис. 8 показаны результаты измерений, выполненных 21.09.90 в 09:09–09:49 MSK (звёздочки) и в 10:12–10:30 MSK (точки). Для сравнения показаны данные, полученные методом некогерентного рассеяния [28] (светлые кружки), а также на ракетах 15.10.86 [29] (сплошные кривые). Штрихпунктиром показаны модельные профили CIRA-72.

3.6. Исследование неоднородной структуры нижней ионосферы

Наблюдения 10–12 августа 1999 г. с высотным шагом 0,7 км и цифровой регистрацией квадратурных компонент сигнала показали, что амплитуда рассеянного сигнала во многих случаях определяется интерференцией волн от ИПН и естественных ионосферных образований, включая спорадические слои и крупномасштабные неоднородности. Это позволяет изучать неоднородную структуру

нижней ионосферы на основе анализа высотно-временной зависимости амплитуды и фазы рассеянного сигнала. На рис. 9 приведены примеры яркостных записей амплитуды рассеянного сигнала для таких случаев. На рис. 9а видны отражения от слоя E_s с действующей высоты 140 км, от ИПН в E - и D -областях, эхо от естественных неоднородностей с высот $75 \div 90$ км и отражение от слоя E_s на высоте 85 км. Рис. 9б иллюстрирует «расслоение» E -области, когда рассеяние на ИПН существенно искажается рассеянием сигнала на естественных неоднородностях ионосферы. Это приводит к искажению высотного профиля амплитуды и времени релаксации рассеянного сигнала, что позволяет оценивать горизонтальные масштабы и скорость перемещения естественных неоднородностей.

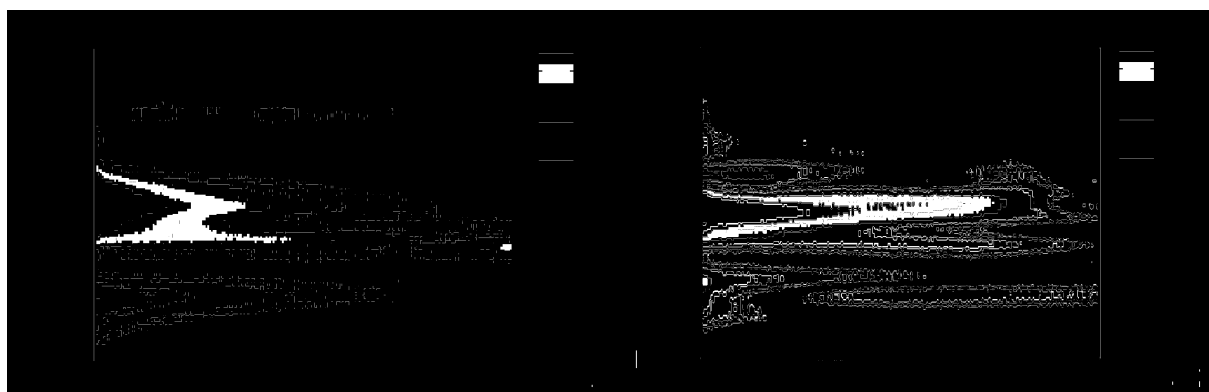


Рис. 9. Примеры яркостных записей рассеянных сигналов: часто встречающийся случай с рассеянием на ИПН и естественных ионосферных образованиях, включая E_s (а) и «расслоение» E -области (б)

3.7. Исследование D -области. Определение концентрации атомного и возбуждённого кислорода.

Восходно-заходные явления

На высотах области D ИПН образуются в результате увеличения скорости прилипания электронов к молекулам кислорода, что приводит к снижению электронной концентрации и росту концентрации отрицательных ионов кислорода [24, 30, 31]. Экспериментальные высотные профили амплитуды и времени релаксации сигналов, рассеянных ИПН, характеризуются плавным возрастанием до высоты $h \approx 77$ км. Выше этого уровня амплитуда и время релаксации рассеянного сигнала резко спадают, что объясняется увеличением концентрации атомарного кислорода. Высокая концентрация атомарного кислорода в атмосфере приводит к увеличению суммарной скорости отлипания электронов от отрицательных ионов, что фактически прекращает действие механизма образования ИПН, характерного для области D . В работах [30, 31] приведены экспериментальные высотные профили амплитуды и времени релаксации сигналов, рассеянных ИПН в дневной области D , и дано количественное объяснение полученных профилей на основе модели фотохимических процессов с одним типом отрицательных ионов O_2^- .

Известно, что D -область претерпевает сильные изменения на восходе и заходе Солнца. Это подтвердили измерения, проведённые в августе 2000 г. Яркостные записи амплитуды рассеянного сигнала во время захода и восхода Солнца представлены на рис. 10. Из них видно, что в области D уменьшение амплитуды сигнала происходит при изменении зенитного угла Солнца от 90° до 105° (см. рис. 10а; шкала зенитных углов помещена на верхней границе рисунка, сплошные линии показывают время прохождения соответствующих высот тенью от Земли и озоносферы; высота последней принята равной 30 км). Увеличение рассеянного сигнала происходит при уменьшении зенитного угла Солнца от 97° до 90° (см. рис. 10б). При этом хорошо заметна восходно-заходная асимметрия, которая выражается в

том, что на заходе Солнца сигналы имеют бóльшую амплитуду и занимают бóльший высотный интервал, чем на восходе. На заходе Солнца слабые сигналы в верхней части области D видны ещё около часа после того, как область полностью погружается в тень (после захода Солнца на этих высотах). Обращает на себя внимание также более широкий и глубокий минимум, формирующийся между областями D и E на восходе Солнца.

Эти процессы хорошо описываются моделью с одним типом отрицательных ионов O_2^- (рис. 11). Такая модель позволяет определять изменение концентрации атомарного кислорода и возбуждённых молекул кислорода в состоянии $^1\Delta g$ и показывает значительное увеличение концентрации атомарного кислорода во время восхода Солнца [31].

Зависимость температуры и плотности атмосферы на высоте 99 км на заходе Солнца приведена на рис. 12. Видны колебания с периодом 15÷60 минут, связанные, возможно, с возбуждением ВГВ на линии терминатора.

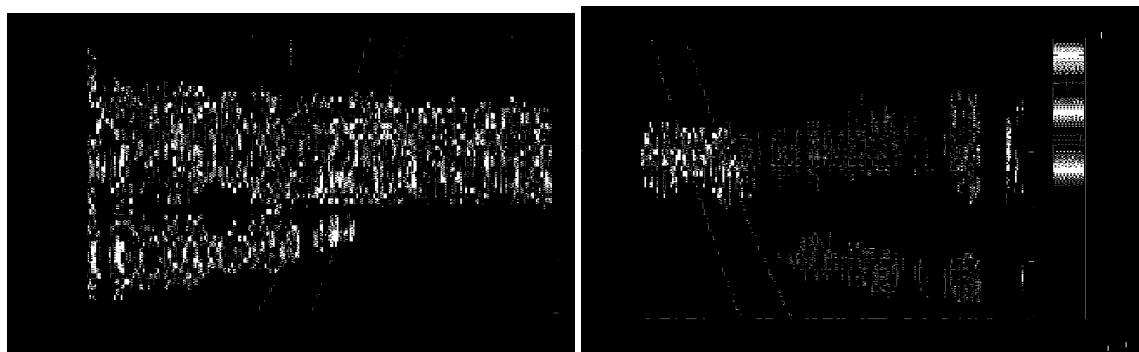


Рис. 10. Яркостные записи амплитуды рассеянного ИПН сигнала во время захода и восхода Солнца 16–17 августа 2000 г. Верхняя шкала соответствует зенитному углу Солнца

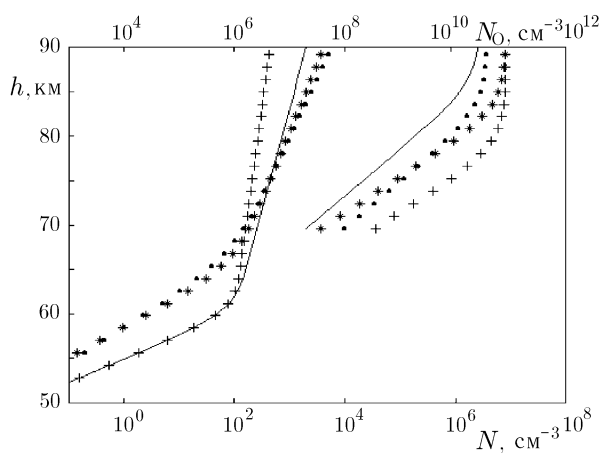


Рис. 11. Модельные профили $N(h)$ (левые кривые, нижняя шкала) и концентрация атомарного кислорода N_O (правые кривые, верхняя шкала): сплошные линии и точки — на заходе Солнца, звёздочки и крестики — на восходе. Сплошные линии и крестики соответствуют зенитному углу $\chi = 89^\circ$, точки и звёздочки — $\chi = 94^\circ$

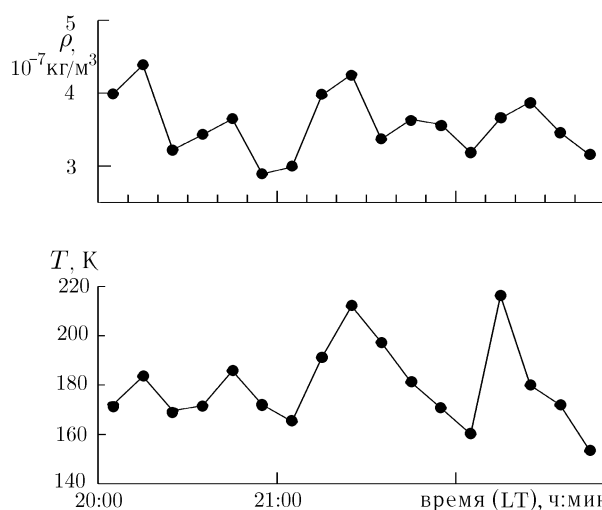


Рис. 12. Изменение плотности и температуры атмосферы на заходе Солнца на высоте 99 км по измерениям 16 августа 2000 г.

3.8. Определение параметров ионосферной плазмы на высотах области F . Температуры электронов и ионов и частота ионно-молекулярных соударений

Метод определения электронной и ионной температур и частоты ионно-молекулярных соударений в F -области ионосферы основан на ударном возбуждении ионно-звуковых волн в процессе формирования искусственных периодических неоднородностей плазмы. Первые измерения T_e , T_i и ν_{im} проведены с помощью стенда «Сура» летом и зимой 1983–1986 гг. Сопоставив теоретически рассчитанные значения частоты Ω и декремента затухания Γ ионно-звуковых колебаний с экспериментальными данными, можно определить два параметра из трёх: T_e , T_i и ν_{im} . В [31] предложен способ определения всех трёх указанных параметров путём возбуждения в исследуемом объёме ионно-звуковых колебаний различных масштабов, а в работе [32] приведены результаты их определения.

4. ПЕРСПЕКТИВЫ ИССЛЕДОВАНИЙ

Применение ИПН позволяет проводить систематические исследования плазмы в нижней ионосфере. Способы определения параметров плазмы с помощью ИПН имеют высокое временное разрешение, чем выгодно отличаются от, например, метода некогерентного рассеяния. По информативности они конкурируют с ракетными исследованиями, являясь при этом значительно более экономичными и экологически чистыми. Все способы диагностики ионосферной плазмы с помощью ИПН легко объединить в одну комплексную методику, которая позволит проводить исследования и вести мониторинг ионосферы в широком высотном интервале от области D до области F .

Установки для ионосферного мониторинга с помощью ИПН можно относительно просто создать на базе нагревных стендов НИРФИ и исследовательского комплекса EISCAT в Тромсё. Для этого требуется лишь некоторое дооборудование действующих установок, сводящееся к созданию приёмного комплекса, оснащению его вычислительной техникой и разработке соответствующих алгоритмов и программ. В ближайшей перспективе планируются совместные эксперименты с использованием нагревных стендов «Сура» и EISCAT.

Применение ИПН является перспективным как для дальнейшего изучения структуры ионосферы, так и для исследования динамических явлений. Наиболее интересные результаты могут быть получены в следующих направлениях:

- сезонные и суточные вариации различных параметров ионосферы, в т. ч. в восходно-заходные периоды и во время солнечных затмений;
- неоднородная структура нижней ионосферы: расслоение регулярного E -слоя, спорадические слои, дополнительные слои в межслоевой E — F -впадине (долине);
- высотное распределение скорости вертикального движения в D - и E -областях ионосферы;
- высотные-временные параметры внутренних гравитационных волн и их спектральные характеристики;
- параметры турбулентности вблизи мезопаузы и их связь с волновой активностью в атмосфере;
- химия отрицательных ионов и аэрономическая модель области D .

Все перечисленные направления являются чрезвычайно актуальными в области исследования ионосферы и верхней атмосферы. Комплексные измерения временных вариаций ионосферных параметров позволяют расширить наше понимание динамики верхней атмосферы и энергетической взаимосвязи между различными слоями ионосферы.

Работа выполнена при поддержке Российского фонда фундаментальных исследований (гранты № 99–06–64464, 00–05–64695 и 01–05–64504).

СПИСОК ЛИТЕРАТУРЫ

1. Виленский И. М. // Доклады АН СССР. 1970. Т. 191, № 5. С. 1 041.
2. Виленский И. М., Плоткин В. В. // Изв. вузов. Радиофизика. 1973. Т. 16, № 6. С. 886.
3. Seliga T. N. // J. Atmos. Terr. Phys. 1972. V. 34, No. 10. P. 1 827.
4. Беликович В. В., Бенедиктов Е. А., Гетманцев Г. Г., Игнатьев Ю. А., Комраков Г. П. // Письма в ЖЭТФ. 1975. Т. 22, вып. 10. С. 497.
5. Беликович В. В., Бенедиктов Е. А., Иткина М. А., Митяков Н. А., Терина Г. И., Толмачёва А. В., Шавин П. Б. // Изв. вузов. Радиофизика. 1977. Т. 20, № 12. С. 1 821.
6. Беликович В. В., Бенедиктов Е. А., Терина Г. И. // Изв. вузов. Радиофизика. 1978. Т. 21, № 10. С. 1 418.
7. Беликович В. В., Бенедиктов Е. А., Дмитриев С. А., Терина Г. И. // Изв. вузов. Радиофизика. 1981. Т. 24, № 4. С. 504.
8. Беликович В. В., Бенедиктов Е. А., Дмитриев С. А., Терина Г. И. // Изв. вузов. Радиофизика. 1981. Т. 24, № 7. С. 905.
9. Fejer J. A., Djuth F. T., Gonzales C. A. // J. Geophys. Res. 1984. V. 89. P. 9 145.
10. Беликович В. В., Вяхирев В. Д., Гончаров Н. П., Ежов А. И., Кунтенкова Н. В., Ковалёв В. Я., Марченко С. Ф., Рубцов Л. Н. // Геомагнетизм и аэрономия. 1996. Т. 36, № 4. С. 135.
11. Бахметьева Н. В., Беликович В. В., Бенедиктов Е. А., Бубукина В. Н., Гончаров Н. П., Игнатьев Ю. А., Рубцов Л. Н. // Геомагнетизм и аэрономия. 1996. Т. 36, № 6. С. 43.
12. Rietveld M. T., Turunen E., Matveinen H., Goncharov N. P., Pollari P. // Ann. Geophys. 1996. No. 14. P. 1 437.
13. Rietveld M. T., Goncharov N. P. // Adv. Space Res. 1998. V. 21, No. 5. P. 693.
14. Djuth F. T., Groves K. M., Elder J. H., Shinn E. R., Quinn J. M., Villasenor J., Wong A. Y. // J. Geophys. Res. A. 1997. V. 102, No. 11. P. 24 023.
15. Беликович В. В., Бенедиктов Е. А., Толмачёва А. В., Бахметьева Н. В. Исследование ионосферы с помощью искусственных периодических неоднородностей. Н. Новгород: ИПФ РАН, 1999. 156 с.
16. Беликович В. В. // Изв. вузов. Радиофизика. 1993. Т. 36, № 12. С. 1 105.
17. Bakhmet'eva N. V., Belikovich V. V., Benediktov E. A., Vyakhirev V. D., Goncharov N. P., Tolmacheva A. V., Korotina G. S. // Radio Science. 1998. V. 33, No. 3. P. 583.
18. Беликович В. В., Бенедиктов Е. А., Толмачева А. В. // Геомагнетизм и аэрономия. 1994. Т. 34, № 1. С. 141.
19. Беликович В. В., Бенедиктов Е. А. // Геомагнетизм и аэрономия. 1995. Т. 35, № 2. С. 91.
20. Бахметьева Н. В., Беликович В. В., Коротина Г. С. // Геомагнетизм и аэрономия. 1996. Т. 36, № 5. С. 180.
21. Беликович В. В., Бенедиктов Е. А., Гончаров Н. П. // Геомагнетизм и аэрономия. 1991. Т. 31, № 2. С. 381.
22. Бахметьева Н. В., Беликович В. В., Бенедиктов Е. А., Бубукина В. Н., Гончаров Н. П., Игнатьев Ю. А. // Геомагнетизм и аэрономия. 1996. Т. 36, № 5. С. 120.
23. Бенедиктов Е. А., Вяхирев В. Д., Толмачёва А. В. // Геомагнетизм и аэрономия. 1997. Т. 37, № 5. С. 79.
24. Беликович В. В., Бенедиктов Е. А. // Изв. вузов. Радиофизика. 1986. Т. 29, № 11. С. 1 283.
25. Бахметьева Н. В., Беликович В. В., Бенедиктов Е. А., Бубукина В. Н., Игнатьев Ю. А. // Изв. вузов. Радиофизика. 1997. Т. 40, № 3. С. 308.
26. Бахметьева Н. В., Беликович В. В., Григорьев Г. И., Толмачёва А. В. // Изв. вузов. Радиофизика. Т. 45, № 3. С. 233.

27. Бахметьева Н. В., Беликович В. В., Игнатъев Ю. А., Понятов А. А. // Геомагнетизм и аэрономия. 1996. Т. 36, № 6. С. 36.
28. Kirkwood S. // J. Atmos. Terr. Phys. 1986. V. 48. P. 817.
29. Lubken F. J., von Zahn U., Manson A. et. al. // J. Atmos. Terr. Phys. 1990. V. 52, No. 10–11. P. 955.
30. Беликович В. В., Бенедиктов Е. А., Бубукина В. Н., Вяхирев В. Д. // Изв. вузов. Радиофизика. 1999. Т. 42, № 5. С. 431.
31. Беликович В. В., Бенедиктов Е. А., Трунов Д. В. // Геомагнетизм и аэрономия. 2000. Т. 40, № 6. С. 55.
32. Бенедиктов Е. А., Беликович В. В., Толмачёва А. В. // Изв. вузов. Радиофизика. 1997. Т. 40, № 3. С. 308.

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

Поступила в редакцию
16 октября 2001 г.

**ARTIFICIAL PERIODIC INHOMOGENEITIES OF THE IONOSPHERIC PLASMA:
A PROMISING LINE IN THE IONOSPHERIC RESEARCH**

E. A. Benediktov, *V. V. Belikovich, N. V. Bakhmet'eva, and A. V. Tolmacheva*

This paper is dedicated to a new technique for studying the ionosphere, which was developed at the Radiophysical Research Institute (NIRFI). The technique is based on creating the artificial periodic inhomogeneities (APIs) in the ionospheric plasma. The paper reviews the methods for determination of the basic ionospheric and atmospheric parameters and present the results of their determination.

УДК 621.371.38

О ПРОСТРАНСТВЕННОЙ КОРРЕЛЯЦИИ ПОЛЯ КОРОТКИХ ВОЛН ПРИ НАКЛОННОМ ОТРАЖЕНИИ ОТ ИОНОСФЕРЫ

Ю. А. Чернов

Рассмотрены пространственные статистические характеристики коротковолнового сигнала и взаимная связь корреляционных функций амплитуды и фазы поля при наклонном отражении от ионосферы. Характеристики амплитуды и фазы выражены через наиболее просто определяемый параметр мутности ионосферы β или только через коэффициент корреляции амплитуд в разнесённых точках. Получены выражения для корреляционных функций фаз и амплитуд, удовлетворительно описывающие экспериментальные данные.

1. ОБЩИЕ ПОЛОЖЕНИЯ

Для решения многих практических задач коротковолновой радиосвязи, а также для исследования распространения радиоволн в ионосфере необходимо знание пространственных корреляционных связей случайных изменений амплитуды и фазы отражённых от ионосферы радиоволн. В частности, это необходимо для совершенствования систем приёма с пространственным разнесением, для исследования работы антенн с большой апертурой, выбора приёмных и передающих антенн для эффективной работы радиолиний и дальнейшего изучения свойств ионосферы. Непосредственные измерения пространственной функции корреляции фаз в различных условиях — практически невыполнимая задача, поэтому желательно воспользоваться другими, более исследованными параметрами ионосферы или принимаемого сигнала, с которыми фазовые характеристики поля связаны вполне определённым образом.

Наиболее тесно быстрые замирания и корреляционные свойства коротковолновых (КВ) сигналов связаны с параметром мутности ионосферы β , равным отношению мощностей зеркальной и рассеянной составляющих (в радиотехнической литературе, когда речь идёт об электрических сигналах, этот параметр часто обозначается как α). Наибольшее число работ по изучению статистических характеристик замираний, коэффициентов корреляции и степени их взаимной связи было выполнено при вертикальном зондировании ионосферы, причём фазовым характеристикам уделялось сравнительно мало внимания. Лишь в нескольких работах рассматривались корреляции амплитуд и фаз (или косинусов фаз) в разнесённых точках при наклонном падении (см., например, [1, 2] и библиографию в них). Вместе с тем интерес к более глубокому пониманию физической природы статистических свойств КВ сигнала сохраняется. В частности, в последнее время предпринимаются попытки оценить в общем параметре β для многоскачкового сигнала две составляющие, одна из которых обусловлена статистическим характером ионосферы, а другая — неоднородностью отражающей поверхности Земли [3]. Продвижение в этом вопросе позволит более обоснованно подойти к статистическому описанию поведения амплитуды и фазы КВ сигнала.

Несмотря на значительное число публикаций о замираниях на коротких волнах, имеющих в литературе сведений о корреляционных связях фаз и амплитуд в разнесённых точках недостаточно для решения многих практических задач.

В настоящей работе с учётом имеющихся экспериментальных данных рассмотрены пространственные статистические характеристики КВ сигнала и взаимная связь корреляционных функций амплитуды и фазы поля при наклонном распространении, выраженные через наиболее просто определяемый

параметр мутности ионосферы β или только через коэффициент корреляции амплитуд в разнесённых точках. Получены выражения для корреляционных функций фаз и амплитуд, удовлетворительно описывающие экспериментальные данные. Изучаемый процесс предполагается узкополосным.

2. ТЕОРЕТИЧЕСКАЯ СВЯЗЬ МЕЖДУ КОЭФФИЦИЕНТАМИ КОРРЕЛЯЦИИ АМПЛИТУДЫ И ФАЗЫ

Будем считать, что на рассматриваемом ограниченном участке поверхности Земли поле локально однородно и имеет место эргодичность. Поэтому все результаты, которые будут получены для случайных процессов во времени, в равной степени относятся и к пространственным случайным функциям. Будем также считать, что амплитуда быстрых замираний в большинстве случаев (особенно на одноканальных трассах) подчиняется распределению Райса (β произвольно), а на протяжённых трассах с отражением от земной поверхности распределение амплитуды замираний близко к распределению Релея ($\beta \ll 1$) [4]. Однако, судя по опубликованным экспериментальным данным, в ряде случаев распределение Релея наблюдается и на коротких трассах, вплоть до вертикального падения.

В общем случае произвольного отношения мощностей зеркальной и флуктуационной составляющих для двумерного распределения Райса корреляционная функция фазы имеет достаточно громоздкий вид [5, ф. 8.79] и не может быть использована в практических расчётах или аналитических исследованиях. Для установления связи между корреляционной функцией $B_\varphi(\tau)$ фазы и коэффициентом корреляции $R_u(\tau)$ огибающей суммарного (гармонической и случайной составляющих) сигнала можно учесть, что обе эти функции выражаются через коэффициент корреляции $R_0(\tau)$ текущих мгновенных значений исходного нормального процесса и его стандартного отклонения σ . Используя это обстоятельство, можно получить приближённые, но достаточно точные формулы, связывающие непосредственно $B_\varphi(\tau)$ и $R_u(\tau)$.

Функция корреляции $B_u(\tau)$ суммарного процесса для малых и больших a приведена в [6, ф. 7.8.11, 7.8.13] и имеет вид

$$\begin{aligned} B_{u1}(\tau) &= \pi\sigma^2 R_0(\tau) [R_0(\tau) + (4/\pi)^2 a^2]/8, & a \ll 1; \\ B_{u2}(\tau) &= \sigma^2 R_0(\tau) [1 + R_0(\tau)/(2a^2)], & a \gg 1, \end{aligned} \quad (1)$$

где σ — стандартное отклонение исходного нормального процесса. Если теперь привести эти функции к единице, чтобы получить коэффициент корреляции, можно записать зависимость $R_u(\tau)$ в двух вариантах:

$$R_{u1}(\tau) = R_0(\tau) [R_0(\tau) + (4/\pi)^2 a^2] / [1 + (4/\pi)^2 a^2], \quad a \ll 1; \quad (2)$$

$$R_{u2}(\tau) = R_0(\tau) [1 + R_0(\tau)/(2a^2)] / [1 + 1/(2a^2)], \quad a \gg 1. \quad (3)$$

Сопоставляя последние два выражения, можно видеть, что они схожи и справедливы во всём диапазоне значений a , от 0 до ∞ . Максимальное расхождение между зависимостями (2) и (3) наблюдается при $a \approx 0,8$ и $R_0(\tau) = 0,5$ и составляет 0,012. Если провести сравнение (3) с точной зависимостью $B_u(\tau)$ для $a = 0$ [5, ф. 8.31'], то получим аппроксимацию $R_u(\tau) = R_0^{2,1}(\tau)$, достаточно близкую к приближённому выражению $R_{u1}(\tau) = R_0^2(\tau)$. Различие между этими зависимостями составляет не более 2,5 %.

Отсюда следует, что для нахождения обратной зависимости можно использовать, в частности, коэффициент корреляции $R_{u1}(\tau)$ из (2), что даёт

$$R_0(\tau) = (4/\pi)^2 \left\{ -a^2 + [a^4 + 4R_u(\tau) (1 + (4/\pi)^2 a^2)/(4/\pi)^4]^{1/2} \right\} / 2. \quad (4)$$

Выразить так же просто $B_\varphi(\tau)$ как функцию переменных $R_0(\tau)$ и a не удаётся. Однако, используя аппроксимации громоздких выражений, полученных в [5] и [6], можно построить сравнительно простой набор формул, обеспечивающий удобство и высокую точность расчётов. Кратко схема получения указанных формул изложена в Приложении, сами же формулы имеют вид

$$B_\varphi(R_0|a) = g_0 + g_1 (1 - R_0)^{1/2} + g_2 R_0,$$

$$g_0 = -0,04022 + 1,248 \exp(-1,05a) + 0,000767a + 0,5088/(0,2 + a);$$

$$g_1 = 0,042 - 1,247 \exp(-1,05a) - 0,00083a - 0,50915/(0,2 + a);$$

$$g_2 = 0,07375 - 0,339 \exp(-a) - 0,0023a - 0,0397/(0,2 + a). \quad (5)$$

Подставляя в (5) $R_0(\tau)$ из (4), получим функцию корреляции фазы, непосредственно выраженную через коэффициент корреляции $R_u(\tau)$ огибающей (для узкополосных процессов вместо распределения огибающей часто используется распределение амплитуды, что в нашем случае также может быть вполне оправдано).

3. ЭКСПЕРИМЕНТАЛЬНЫЕ ХАРАКТЕРИСТИКИ РЕАЛЬНЫХ ТРАСС

Для определения автокорреляционных функций фазовых отклонений в реальных условиях можно воспользоваться выражениями (4), (5) и экспериментальным коэффициентом корреляции $R_u(l)$ амплитуд в точках, разнесённых на расстояние l . Сравнительно полные данные измерений $R_u(l)$ содержатся в [4, 7, 8]. Из этих данных следует, что каждая кривая коэффициента корреляции, характеризующаяся формой и радиусом корреляции, достаточно хорошо соответствует некоторому определённому интервалу протяжённостей трасс и среднему периоду замираний τ_z на них. Эта связь наиболее полно может быть прослежена в работах [4, 7] для совершенно различных условий экспериментов и способов определения скорости (периодов) замираний. Приведённые в этих работах периоды замираний и соответствующие протяжённости радиотрасс оказались возможным объединить в девять групп, удобных для использования. Каждой из этих групп, характеристики которых приведены в табл. 1, соответствует своя кривая коэффициента корреляции. Полученные зависимости $R_u(l)$ достаточно точно отражены на рис. 1; в дальнейшем будем называть эти аппроксимирующие кривые экспериментальными (подробно их построение рассмотрено ниже).

Таблица 1

Номер кривой на рис. 1	Период замираний τ_z , с	Тип трассы	Протяжённость D трассы, км
1	45	односкачковая	меньше 3 500
2	20		
3	10		
4	8		
5	6	многоскачковая	3 500
6	5		4 000
7	4		5 000
8	3,5		6 000
9	2 ÷ 3		больше 8 000

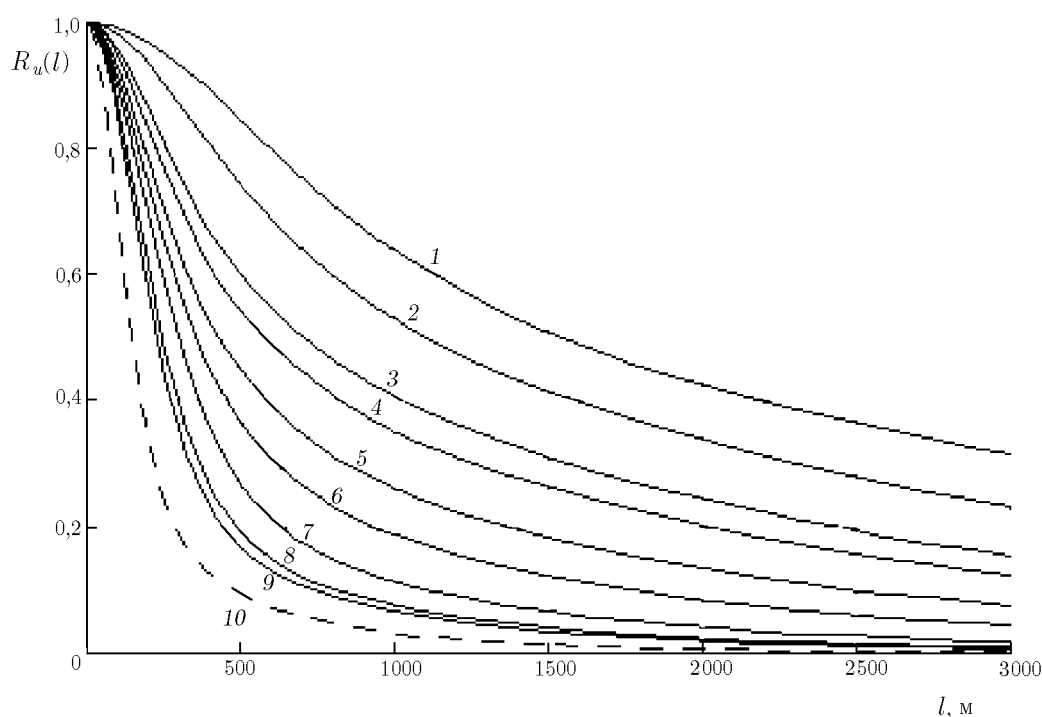


Рис. 1

Необходимо отметить, что, несмотря на конкретные цифры в табл. 1 и однозначность кривых на рис. 1, в действительности исходные данные имеют заметный разброс, и приведённые на рис. 1 зависимости получены после усреднения и округления, не превышающего половины от разницы значений коэффициента корреляции на соседних экспериментальных кривых. Практически среди экспериментальных точек каждой группы проводилась плавная кривая (абсолютная полуширина полосы коэффициента корреляции не превышает 0,05).

В приведённом на рис. 1 семействе кривых для указанной в табл. 1 шкалы периодов замираний некоторые кривые построены путём интерполяции. Кривая 10, которая будет отражена в одной из последующих таблиц, определяет нижнюю границу коэффициента корреляции и относится к трассам с протяжённостью свыше 10 000 км [4] или к трассам с полным рассеянием в ионосфере [9].

Сравнивая кривые на рис. 1 с традиционно применяющимися для описания зависимости $R_u(l)$ гауссовой и экспоненциальной функциями, можно убедиться, что обе они не являются удовлетворительными аппроксимациями.

Отличительной особенностью приведённых реальных функций является то, что их начальная часть имеет отрицательную кривизну и равную нулю первую производную при $l = 0$, а основная часть после перегиба спадает медленнее, чем экспонента. Наличие значительных расхождений следовало ожидать, т. к. экспоненциальная и гауссова функции большинством исследователей выбирались почти исключительно из соображений удобства вычислений.

4. ПРОСТРАНСТВЕННАЯ КОРРЕЛЯЦИЯ АМПЛИТУДЫ И ФАЗЫ

Для определения вида функции, описывающей коэффициент пространственной корреляции амплитуд на поверхности Земли и удовлетворительно аппроксимирующей семейство кривых на рис. 1, можно использовать следующий подход. Он заключается в том, что исходя из характеристик фазовых

искажений в ионосфере (коэффициента автокорреляции ρ_φ фазовых отклонений при прохождении ионосферы) строится автокорреляционная функция B_u амплитуды сигнала на поверхности Земли. Этот путь, вообще говоря, известен, но неизвестны сами характеристики вносимых фазовых искажений. После их определения может быть найдена и автокорреляционная функция B_φ фазы на поверхности Земли по формулам (4) и (5).

Следуя описанной схеме, будем рассматривать ионосферу как фазовый экран со своей пространственной автокорреляционной функцией $\sigma_{\varphi 0}^2 \rho_\varphi$ фазовых искажений. Тогда расчётный коэффициент корреляции амплитуд $R_{up}(l)$ в месте приёма на поверхности Земли с небольшими приближениями может быть записан в виде [10; 11, §10]

$$R_{up}(l) = \frac{\exp\{-\sigma_{\varphi 0}^2 [1 - \rho_\varphi(l)]\} - \exp(-\sigma_{\varphi 0}^2)}{1 - \exp(-\sigma_{\varphi 0}^2)}, \quad (6)$$

где $\sigma_{\varphi 0}^2$ — дисперсия фазовых искажений при выходе волны из ионосферы.

Если считать формулу (6) справедливой, необходимо определить вид функции $\rho_\varphi(l)$ и дисперсию $\sigma_{\varphi 0}^2$. Рассмотрим в роли пробных наиболее простые зависимости.

Подставив в (6) в качестве функции $\rho_\varphi(l)$ часто используемые выражения

$$\rho_{\varphi 1}(l) = \exp(-|l[m]|/200), \quad \rho_{\varphi 2}(l) = \exp[-(l[m]/200)^2] \quad (7)$$

и сравнивая результаты расчёта $R_{up}(l)$ с экспериментальными кривыми на рис. 1, можно видеть их существенное расхождение. На рис. 2 приведён пример таких расчётов. Полученные кривые $R_{up}(l)$ для экспоненциальной и гауссовой функций имеют номера 1 и 2 соответственно, кривая 3 приведена для сравнения и совпадает с кривой 9 на рис. 1. Радиус фазовой корреляции l_φ для двух выбранных функций подобран так, чтобы при $l = l_\varphi$ коэффициент корреляции амплитуд $R_{up}(l)$ был равен 0,37. Кривая 9 (см. рис. 1) выбрана из тех соображений, что для неё с наибольшей вероятностью выполняется условие максимальной дисперсии фаз $\sigma_{\varphi 0}^2 = \sigma_{\varphi m}^2$, что даёт право подставить в (6) $\sigma_{\varphi 0}^2 = \pi^2/3$.

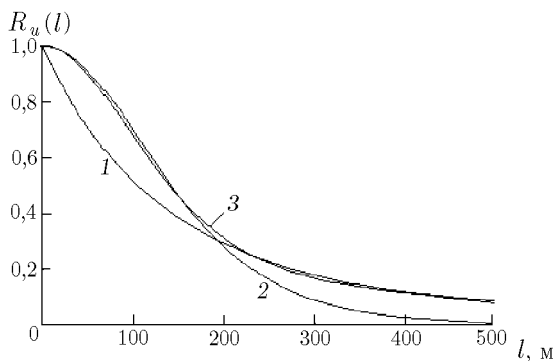


Рис. 2

Из рис. 2 видно, что экспериментальная кривая 3 при малых l удовлетворительно совпадает с зависимостью $R_{up}(l)$, построенной для гауссового коэффициента автокорреляции фазы, а при больших l — с расчётной зависимостью для экспоненциального коэффициента автокорреляции $\rho_\varphi(l)$.

Приведённые расчёты показывают также, что кривые $R_{up}(l)$ при разных $\rho_\varphi(l)$, каждая в своей части, близки к функции $R_u(l)$, соответствующей кривой 9 на рис. 1. Поэтому функцию, аппроксимирующую коэффициент корреляции $\rho_\varphi(l)$ фазовых искажений от ионосферных неоднородностей, целесообразно искать в виде подходящего сочетания двух

зависимостей: гауссовой (при малых l) и экспоненциальной (спадающая часть $\rho_\varphi(l)$ при больших l). Физически это в некоторой мере оправдано, т. к. турбулентные неоднородности ионизации в ионосфере могут быть достаточно плавными и обуславливать равенство нулю первой производной автокорреляционной функции при $l = 0$. Вместе с тем эти же неоднородности в большом объёме могут в совокупности обладать статистическими характеристиками, близкими к марковским, и в области больших l формировать близкую к экспоненте автокорреляционную функцию. В такой весьма упрощённой схеме имеют место оба указанных свойства, и соответствующий ей комбинированный аппроксимирующий

коэффициент автокорреляции после подбора подходящих числовых коэффициентов может иметь, в частности, следующий вид:

$$\rho_{\varphi}(l) = 0,23 \exp(-dl^2/l_{\varphi}^2) + 0,77q + 0,77(1 - q) \exp(-bl/l_{\varphi}), \quad (8)$$

где $q = \exp[l^2/(2l_{\varphi}^2)]$, l_{φ} — радиус корреляции, b и d — числовые коэффициенты.

Вспомогательная функция q выполняет роль ключа, связывающего гауссову и экспоненциальную области в (8). Благодаря этой функции выражение (8) при малых l повторяет гауссову кривую, при больших — спадает как экспонента, при $l = 0$ имеет производную, равную нулю. При подстановке (8) в (6) получим функции $R_{up}(l)$, которые при соответствующем подборе параметров хорошо аппроксимируют экспериментальные данные.

Для построения всего семейства кривых, приведённого на рис. 1, необходимо оценить дисперсию фазы $\sigma_{\varphi 0}^2$ ионосферного фазового экрана в случае распределения Райса для амплитуд сигнала, т. е. для средних и малых протяжённостей трасс.

Случайные фазовые набегии при прохождении радиоволнами эквивалентного фазового экрана формируют случайное поле, аналогичное полю на поверхности Земли, т. е. состоящее из когерентной и хаотической составляющих. Поэтому для оценки дисперсии фазы $\sigma_{\varphi 0}^2$ можно использовать развитый выше подход и выразить её через параметр $\beta = a$ с помощью формул (5) с заменой $B_{\varphi}(a)$ на $B_{\varphi 0}(\beta)$. С учётом равенства $\sigma_{\varphi 0}^2 = B_{\varphi 0}(\beta)|_{R_0=1}$ это даёт

$$\sigma_{\varphi 0}^2 = B_{\varphi 0}(\beta)|_{R_0=1} = 0,0279 + 0,8668 \exp(-\beta) - 0,001251\beta + 0,479/(0,2 + \beta). \quad (9)$$

Для оценки параметра β можно воспользоваться связью между β и периодом замираний τ_z . В [12] приведены экспериментальные зависимости $\beta(\tau_z)$. Диапазон изменений периода замираний τ_z , приведённый в [12] для случая вертикального падения ($2 \div 32$ с), близок к полученному при наклонном падении [9] ($2,8 \div 47$ с), поэтому можно приближённо оценить $\sigma_{\varphi 0}^2$ из (9) с помощью зависимости $\beta(\tau_z)$ и значений τ_z , полученных при наклонном падении. Результаты измерений, приведённые в [12] и показанные на рис. 3, в отличие от предложенной в [12] линейной аппроксимации могут быть лучше описаны нелинейной регрессионной зависимостью периода замираний от коэффициента мутности (τ_z выражено в секундах):

$$\tau_z = 3,296 + 5,774\beta^{1/2} + 2,22\beta,$$

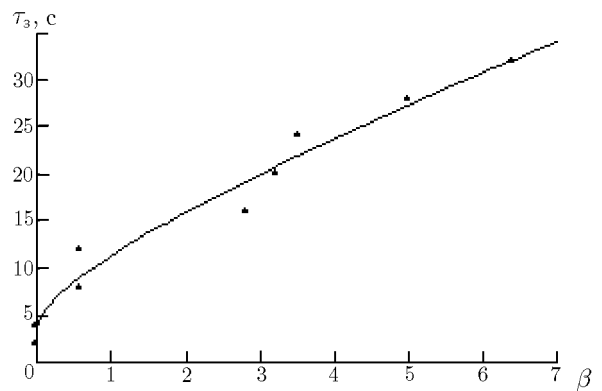


Таблица 2

Номер кривой на рис. 1	Период замираний τ_z , с	β	$\sigma_{\varphi 0}^2$, рад ²
1	45	10,4	0,06
2	20	3,0	0,216
3	10	0,75	0,936
4	8	0,42	1,36
5	6	0,16	2,08
6	5	0,07	2,6
7	4	0,013	3,13
8	3,5	0,0011	3,276
9	2 ÷ 3	0	3,29

$$\beta = \begin{cases} [-1,3 + (0,202 + 0,45\tau_3)^{1/2}]^2, & \tau_3 \geq 3,3, \\ 0, & 0 < \tau_3 < 3,3. \end{cases} \quad (10)$$

Используя (9), (10) и табл. 1, получим дисперсии фазы на фазовом экране (при выходе из ионосферы) для отдельных кривых на рис. 1 (см. табл. 2).

Из проведённого анализа следует, что для построения функции корреляции фазы достаточно знать β или τ_3 (параметр β находится по τ_3 с помощью (10)). Далее в (6) подставляются выражения (8), (9) с учётом приведённых в табл. 3 значений l_φ , b , d , и строится зависимость $R_{up}(l)$, соответствующая заданному β . Используя другие значения β и, соответственно, другие значения l_φ , b и d , получим семейство аналитических кривых $R_u(l)$, представляющих коэффициенты корреляции амплитуд (семейство аппроксимирующих кривых $R_u(l)$ на рис. 1 рассчитывалось именно таким образом). Затем, подставляя нужную зависимость $R_{up}(l)$ в (4), а (4) в (5), получим искомую функцию корреляции фазы $B_\varphi(l|\beta)$. Ниже индекс p у обозначения коэффициента корреляции будем опускать.

Коэффициент корреляции $R_u(l)$ задаётся кривой (8) и параметрами β или τ_3 , которые приведены в табл. 2. При необходимости расчёта функции корреляции фазы для значений β , отсутствующих в табл. 2, необходимо провести графическую интерполяцию или интерполяцию коэффициентов в (8). В приведённой схеме расчётов не только зависимости $R_u(l)$ и $B_\varphi(l|\beta)$, но и входящие в них величины $\sigma_{\varphi 0}^2$, $\rho_\varphi(l)$ должны рассчитываться при одном и том же значении β . Коэффициенты в выражении (8) при различных значениях l приведены в табл. 3. В этой же таблице представлены параметры для построения предельной кривой 10 на рис. 1. При интерполяции удобно воспользоваться следующими выражениями:

$$l_\varphi = 473,98 - 3,138 \cdot 10^{-6} \exp(6x) + 12,238/x - 112,99x^{1/2} + 6,9x^2, \quad (11)$$

$$b = 0,145 + 0,05 \exp(-x) + 0,00268 \exp(x) - 0,0114 [2,722 - (1,65 - x)^2]^{1/2}, \quad (12)$$

$$d = 0,7 - 0,0000157 \exp(3x) + 1,252x^{1/4} - 0,3348x, \quad (13)$$

где $x = \sigma_{\varphi 0}^2$ находится из β с помощью (9).

Т а б л и ц а 3

Номер кривой на рис. 1	$\sigma_{\varphi 0}^2$, рад ²	l_φ , м	b	d
1	0,06	650	0,19	1,3
2	0,216	480	0,18	1,48
3	0,936	380	0,154	1,61
4	1,36	365	0,15	1,6
5	2,08	350	0,155	1,5
6	2,6	330	0,17	1,38
7	3,13	300	0,2	1,12
8	3,276	250	0,215	1,0
9	3,29	230	0,217	0,98
10	3,29	150	0,22	0,95

В качестве примера на рис. 4, 5 показаны построенные по полученным выше формулам зависимости коэффициента корреляции $R_u(l)$ амплитуд и функции корреляции $B_\varphi(l|\beta)$ фаз для $\beta = 10,4$ и $0,013$, что охватывает практически весь диапазон реальных значений этого параметра. Там же для сравнения точками показаны исходные экспериментальные значения коэффициентов корреляции [7], соответствующие этим кривым, которые наряду с другими данными были использованы для построения семейства кривых на рис. 1.

Семейство функций корреляции фазы, соответствующее рис. 1, построено с помощью (4), (5) и показано на рис. 6, 7. Характерно, что функции корреляции фазы в отличие от

функций корреляции амплитуды имеют ненулевую первую производную при $l = 0$.

В практической работе любая из исходных величин, β или τ_3 , может быть оценена экспериментально по соответствующим методикам.

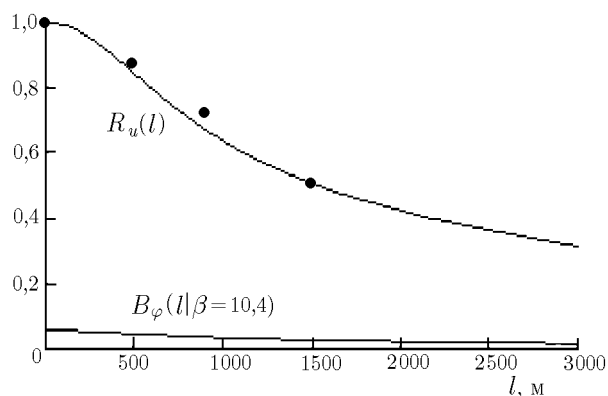


Рис. 4

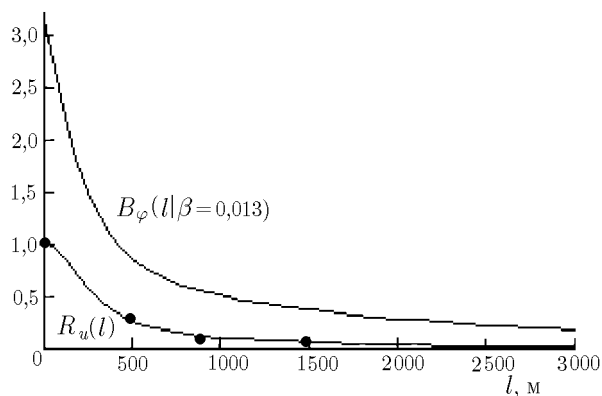


Рис. 5

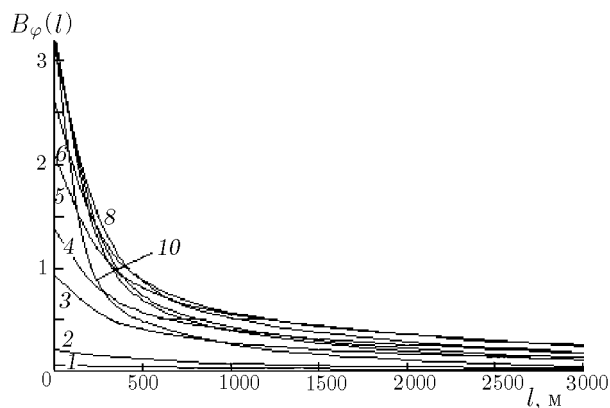


Рис. 6

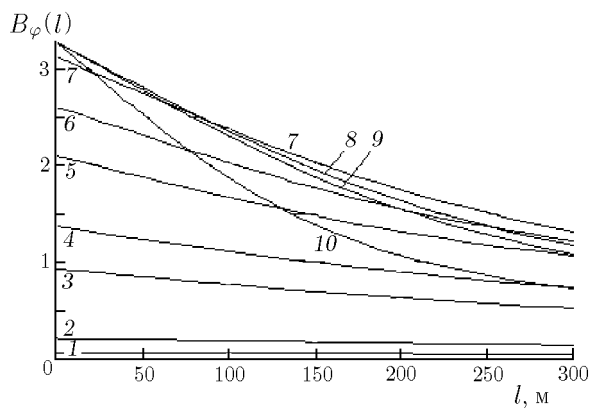


Рис. 7

5. ЗАКЛЮЧЕНИЕ

Проведённый анализ показал, что реальные функции корреляции амплитуд и фаз отражённого от ионосферы сигнала не совпадают с гауссовой кривой или экспонентой, которые чаще всего используются в расчётах. Реальные зависимости более сложны и соответствуют указанным модельным кривым лишь в определённых интервалах аргумента.

В работе изложен один из возможных путей оценки корреляционных функций с использованием промежуточных аппроксимаций с некоторой потерей точности, что однако вполне компенсируется простотой решений и широкими возможностями их применения. Полученные семейства кривых можно рассматривать скорее как иллюстрацию предложенного метода, а не как универсальное решение. Используемые при их построении экспериментальные данные получены при определённых условиях, и остаётся широкое поле приложения предложенного подхода для других ионосферных условий.

Важно отметить, что за рамками обсуждения осталась частотная зависимость корреляционных функций, которая практически не изучалась, хотя некоторые измерения рассеянных сигналов и показывают, что корреляция сигналов в разнесённых точках зависит не от числа длин волн между антеннами, а только от расстояния между ними [13]. Возможно, вид корреляционных кривых связан с отношением рабочей частоты к максимальной применимой частоте (МПЧ).

Требуются также дополнительные исследования зависимости периода замираний от параметра мутности ионосферы в других ионосферных условиях. Приведённая на рис. 3 кривая является лишь частным случаем.

Несмотря на отмеченные ограничения, построенные семейства корреляционных функций могут служить хорошей основой для ориентировочных решений в большом числе задач.

ПРИЛОЖЕНИЕ

КРАТКИЙ ВЫВОД КОРРЕЛЯЦИОННОЙ ФУНКЦИИ ФАЗЫ СЛУЧАЙНОГО НОРМАЛЬНОГО ПРОЦЕССА

Для вывода приближённого выражения для корреляционной функции фазы случайного нормального процесса достаточно использовать несколько формул.

1) Корреляционная функция $B_\varphi(\tau)$ фазы нормального шума без постоянного сигнала, непосредственным аргументом которой является коэффициент автокорреляции $R_0(\tau)$ исходного нормального процесса, имеет вид [5, ф. 8.81]

$$B_\varphi(\tau) = \frac{1}{2} \sum_{r=1}^{\infty} \sum_{n=0}^{\infty} \Gamma^2(n+r/2) (n! (n+r)!)^{-1} R_0^{r+2n}(\tau). \quad (\text{П1})$$

2) Функция корреляции шумового сигнала без когерентной составляющей может быть представлена как $B_\varphi(\tau) = \sigma_{\varphi m}^2 R(\tau)$, где $\sigma_{\varphi m}^2$ — максимальная дисперсия фазы случайного процесса ($\sigma_{\varphi m}^2 = \pi^2/3$), $R(\tau)$ — нормированный коэффициент автокорреляции фазы этого процесса. Зависимость дисперсии фазы от отношения a мощностей когерентной и шумовой составляющих сигнала имеет вид [5, ф. 8.65]

$$\sigma_{\varphi m}^2 = \pi^2/3 + 4\pi \sum_{n=1}^{\infty} (-1)^n c_n/n^2, \quad (\text{П2})$$

Таблица П1

$R_0(\tau)$	$B_\varphi(R_0)/\sigma_{\varphi m}^2$
0,00	0,000
0,90	0,6420
0,05	0,024
0,95	0,7380
0,10	0,0486
0,98	0,8277
0,20	0,099
0,99	0,8750
0,40	0,209
0,995	0,9090
0,60	0,340
1,000	1,0000
0,80	0,5130
-	-

Таблица П2

a	$B_\varphi(\tau=0) = \sigma_{\varphi m}^2$	$B_\varphi(\tau=0) = 0,5a^{-1}$
0,0	3,29	-
0,2	1,866	-
0,4	1,408	-
0,6	1,117	-
0,8	0,9119	-
1,0	0,759	-
1,5	0,511	-
2,0	0,368	0,25
3,0	0,221	0,1667
5,0	0,1158	0,100
7,0	0,0782	0,0714
10	0,0529	0,050
20	0,0257	0,0250
30	0,0169	0,0167

где $c_n = \Gamma(1+n/2) (\pi n! 2^{n/2})^{-1} (2a)^{n/2} {}_1F_1(n/2, n+1, -a)$, $a = Um^2/(2\sigma_u^2)$, U_m и $U_m^2/2$ — амплитуда и мощность когерентной составляющей, σ_u^2 — мощность хаотической составляющей шумового сигнала,

${}_1F_1(x, y, z)$ — гипергеометрическая функция. При $a = 0$ дисперсия фазы $\sigma_\varphi^2(0)$ максимальна и равна $\sigma_{\varphi m}^2 = \pi^2/3$.

3) Функция корреляции при большом отношении сигнал/шум ($a \gg 1$) имеет вид [6, ф. 7.9.5]

$$B_\varphi(\tau) = (2a)^{-1} R_0(\tau). \quad (\text{ПЗ})$$

Зависимости (П1)–(П3) позволяют воссоздать автокорреляционную функцию фазы при любом отношении сигнал/шум и произвольном коэффициенте корреляции $R_0(\tau)$.

Рассчитанные по формуле (П1) значения $B_\varphi(R_0)$ при $a = 0$ приведены в табл. П1.

Построенная по данным табл. П1 аппроксимирующая формула имеет вид

$$B_\varphi(\tau)/\sigma_{\varphi m}^2 = 1,1338 - 1,128 (1 - R_0(\tau))^{1/2} - 0,1448 R_0(\tau). \quad (\text{П4})$$

В табл. П2 приведены результаты численных расчётов дисперсии по формуле (П2), которые также относятся и к функции корреляции, поскольку $\sigma_\varphi^2(a) = B_\varphi(a)|_{R_0=1}$. Для сравнения в табл. П2 приведены результаты расчётов по формуле (П3) для больших a . Видно, что при $R_0 = 1$ удовлетворительное совпадение достигается при $a > 4$.

Аппроксимирующая формула, соответствующая табл. П2, имеет вид

$$B_\varphi(a, \tau = 0) = \sigma_{\varphi m}^2(a) = 0,0261 + 0,957 \exp(-a) - 0,001407a + 0,4615/(0,2 + a). \quad (\text{П5})$$

С помощью формул (П3)–(П5) можно построить семейство кривых $B_\varphi(a, R_0)$, по которым конкретное значение коэффициента корреляции при любом сочетании параметров может быть получено путём несложной интерполяции.

Начала всех кривых указанного семейства при $R_0(\tau \rightarrow \infty) = 0$ имеют одно значение $B_\varphi(a)|_{R_0=0} = 0$. Коэффициенты корреляции при $R_0(\tau) = 1$ могут быть получены из (П5) для ряда значений a (табл. П2) и по формуле (П3) для больших a . Промежуточные значения $B_\varphi(a, R_0)$ для $0 < R_0 < 1$ получим следующим образом.

При малых значениях a вид функции $B_\varphi(a, R_0)$ будет аналогичен зависимости $B_\varphi(a = 0, R_0)$ с той лишь разницей, что значение функции при $R_0 = 1$ при заданной величине a будет равно рассчитанному по (П5) или приведённому в табл. П2. Промежуточные значения функции $B_\varphi(a, R_0)$ при различных R_0 и малых a будут изменяться в той же пропорции, что и на исходной шкале $B_\varphi(a, R_0)$ при $a = 0$ в соответствии с (П4) или табл. П1. Для больших значений a функция $B_\varphi(a, R_0)$ для различных R_0 вычисляется из (П3). Соединение этих двух частей, рассчитанных при одном значении R_0 , в одну функцию от a может быть произведено по методу наименьших квадратов. Например, для $R_0 = 0,6$ соединение происходит при $a = 2$, и в этой точке расхождение между кривыми равно примерно 1 %. Подобным образом построены и другие кривые, составляющие семейство $B_\varphi(a|R_0)$.

Путём перестроения получим семейство кривых $B_\varphi(R_0|a)$, у которых аргументом вместо a является R_0 . Ниже, на рис. П1, приведён пример семейства кривых $B_\varphi(R_0|a)$ для параметра a , изменяющегося от 0 до 5 с шагом 0,5 (значение $B_\varphi(R_0|a)$ в общем случае не ограничивается единицей, как на рис. П1).

Формулы, полученные описанным выше путём, имеют вид

$$B_\varphi(R_0|a) = g_0 + g_1 (1 - R_0)^{1/2} + g_2 R_0,$$

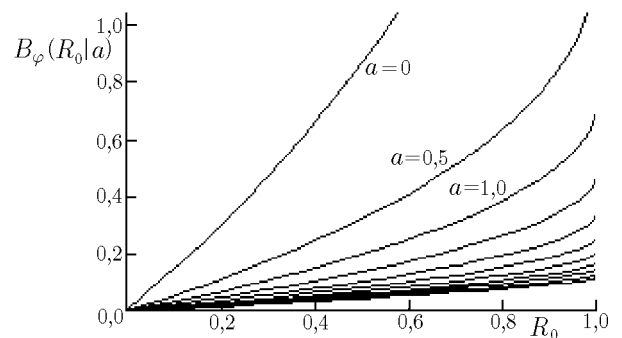


Рис. П1

$$\begin{aligned}
 g_0 &= -0,04022 + 1,248 \exp(-1,05a) + 0,000767a + 0,5088/(0,2 + a), \\
 g_1 &= 0,042 - 1,247 \exp(-1,05a) - 0,00083a - 0,50915/(0,2 + a), \\
 g_2 &= 0,07375 - 0,339 \exp(-a) - 0,0023a - 0,0397/(0,2 + a).
 \end{aligned}
 \tag{П6}$$

Если известен исходный коэффициент корреляции $R_0(\tau)$, то, подставляя его в (П6), получим функцию автокорреляции фазы.

СПИСОК ЛИТЕРАТУРЫ

1. Цветнов В. В. // Исследование точности и помехоустойчивости фазовых радиопеленгаторов: Тр. МАИ. Ленинград: Судпромгиз, 1959. С. 26.
2. Всехсвятская И. С. Статистические свойства сигналов, отражённых от ионосферы. М.: Наука, 1973. 136 с.
3. Миркотан С. Ф., Белов С. Ю., Захаров В. И. // Радиотехника и электроника. 1999. № 10. С. 1 190.
4. Грисдейл Г. Л., Моррис Д. Г., Палмер Д. С. // Краткая информация по материалам зарубежных журналов. М.: Связьиздат, 1958. № 9(38). С. 3.
5. Левин Б. Р. Теоретические основы статистической радиотехники. М.: Сов. радио, 1966. 728 с.
6. Тихонов В. И. Статистическая радиотехника. М.: Сов. радио, 1966. 678 с.
7. Богданов С. Е. Исследование замираний: Сб. трудов НИИ. Вып. 2 (6). М.: Связьиздат, 1957. 38 с.
8. Коноплёва Е. Н., Хмельницкий Е. А. // Электросвязь. 1970. № 1. С. 20.
9. Sugar G. // Proc. IRE. 1955. V. 43. P. 1 432.
10. Ratcliffe J. // Rep. Progr. Phys. 1956. V. 19. P. 188.
11. Рытов С. М., Кравцов Ю. А., Татарский В. И. Введение в статистическую радиофизику. Ч. 2, Случайные поля. М.: Наука, 1978. 464 с.
12. Чеча В. А., Зеленков В. Е. // Дрейфы и неоднородности в ионосфере: Сб. статей. М.: АН СССР, 1959. № 1. С. 50.
13. Ветшев Ж. Н., Бочаров В. И., Афонина Л. Я. // Труды Сибирского физико-технического института. Томск: Изд-во Томского университета, 1962. Вып. 41. С. 109.

Научно-исследовательский институт радио,
г. Москва, Россия

Поступила в редакцию
18 июля 2001 г.

SPATIAL CORRELATION OF AN HF WAVE FIELD IN THE CASE OF OBLIQUE REFLECTION FROM THE IONOSPHERE

Yu. A. Chernov

We consider the spatial statistical characteristics of an HF signal and the interdependence of the correlation functions of the field amplitude and phase in the case of oblique reflection from the ionosphere. The amplitude and phase parameters are expressed either in terms of the parameter β of ionospheric turbidity, which can be determined most easily, or only in terms of the correlation coefficient of the amplitudes at different points. We obtain the expressions for the amplitude and phase correlation functions, which describe well the experimental data.

УДК 537.874

ОБ ИСПОЛЬЗОВАНИИ МЕТЕОРАДАРОВ МИЛЛИМЕТРОВОГО ДИАПАЗОНА ДЛИН ВОЛН ДЛЯ ИЗМЕРЕНИЯ ТЕМПЕРАТУРНЫХ ПРОФИЛЕЙ В АТМОСФЕРЕ

В. О. Рапопорт, Н. А. Митяков

Рассмотрен метод радиоакустического зондирования атмосферы с использованием миллиметровых метеорадаров на основе мощных мазеров (гиротронов), работающих в режиме внутриимпульсной линейной частотной модуляции. Предложенный метод позволяет проводить измерения температурных профилей в атмосфере без какой-либо модернизации приёмно-передающей системы метеорадара, и существенно расширяет возможности дистанционного мониторинга атмосферы.

Предложение А. Л. Фабриканта об использовании метода радиоакустического зондирования атмосферы при проведении исследований с помощью метеорадаров, работающих в режиме амплитудной модуляции радиоимпульса в миллиметровом диапазоне длин волн [1], открывает новые возможности в исследованиях атмосферы. Однако выбранные А. Л. Фабрикантом схема и методика измерений температуры далеки от оптимальных. Даже использование акустических излучателей с рекордной мощностью даёт возможность проводить измерения лишь до высот $1,5 \div 2$ км. В [2] рассматривались способы оптимизации системы измерений. Предложенные технические решения позволяют существенно (на 4 порядка) увеличить отношение сигнал/шум по сравнению с оценками А. Л. Фабриканта, что даёт возможность измерить температурные профили до высот порядка 13 км. Однако использование модулированных по амплитуде импульсов требует существенной модернизации как передающего, так и приёмного трактов системы. В настоящее время разрабатываются метеорадары, использующие достижения мощной микроволновой электроники [3–5] и обладающие уникальными возможностями для проведения исследований пространственно-временной структуры атмосферы. Использование внутриимпульсной линейной частотной модуляции (ЛЧМ) в этих установках позволяет проводить измерения температурных профилей без какой-либо модернизации приёмно-передающей системы радара. Рассмотрим эту возможность более подробно.

При работе радаров с ЛЧМ импульсами сдвиг частот между зондирующим и рассеянным сигналами определяется временной задержкой между излучаемым и принимаемым импульсами, т. е. высотой отражающего слоя. Поэтому энергетический спектр произведения зондирующего импульса на принимаемый сигнал характеризует распределение рассеивателей по высоте. Такая обработка возможна в предположении, что во время прохождения импульса среда неподвижна и отдельные рассеиватели (в случае метеорадара это микрочастицы) дельта-коррелированы. При этом пространственное разрешение определяется частотным интервалом ЛЧМ импульса, а объём рассеяния — длительностью импульса.

Пусть концентрация аэрозолей модулирована проходящей через рассеивающий объём акустической волной с частотой Ω_a и волновым числом K_a . В применении к радиоакустике задача сводится к тому, чтобы в принимаемом сигнале выделить две спектральные компоненты с волновыми числами k_1 и k_2 , для которых выполняется условие, аналогичное условию Брэгга: $k_1 - k_2 = K_a/2$.

Излучаемый сигнал с частотой ω_0

$$E_{tr} \propto \exp[-i\omega_0 t - i\gamma t^2/2], \quad (1)$$

принимаемый на уровне Земли рассеянный сигнал

$$E_s \propto \int p(z) \exp\left[-i\omega_0 \left(t - \frac{2z}{c}\right) - i\frac{\gamma}{2} \left(t - \frac{2z}{c}\right)^2\right] dz. \quad (2)$$

В. О. Рапопорт, Н. А. Митяков

403

Здесь $p(z)$ характеризует распределение дипольных моментов аэрозольных частиц, на которых происходит рассеяние радиоволн, вдоль направления зондирования, c — скорость света. Спектр произведения излучаемого и рассеянного сигналов имеет вид

$$\Phi(\Omega) = \int p(z) \exp \left[i\omega_0 \frac{2z}{c} + i \frac{2\gamma tz}{c} - i \frac{2\gamma z^2}{c^2} - i\Omega t \right] dz dt = \frac{c}{2\gamma} p \left(\frac{c\Omega}{2\gamma} \right) \exp [i\omega_0 \Omega / \gamma - i\Omega^2 / (2\gamma)]. \quad (3)$$

На этом этапе обработки используются узкополосные фильтры с шириной полосы порядка τ^{-1} , где τ — длительность импульса. Энергетический спектр

$$\Phi(\Omega)\Phi^*(\Omega) = \frac{c^2}{4\gamma^2} p \left(\frac{c\Omega}{2\gamma} \right) p^* \left(\frac{c\Omega}{2\gamma} \right) \propto \sigma \left(\frac{c\Omega}{2\gamma} \right),$$

где звёздочка означает комплексное сопряжение, определяет зависимость поперечного сечения рассеяния σ от расстояния, если учесть, что $z = c\Omega / (2\gamma)$.

Обычно на этом заканчивается стандартная процедура обработки ЛЧМ сигналов. Рассмотрим теперь процедуру обработки радиоакустического сигнала. Выполним двумерное преобразование Фурье для функции $\Phi(\Omega_1)\Phi^*(\Omega_2)$. Учтём, что $p(z_1)p^*(z_2) = \sigma(z_1)\delta(z_1 - z_2)$ и $\sigma = \sigma_0 + \sigma_1 \exp(-i\Omega_a t + iK_a z)$. В результате получим

$$\begin{aligned} \Psi(\Delta\chi) &= \int \Phi(\Omega_1)\Phi^*(\Omega_2) \exp(i\Omega_1\chi_1 - i\Omega_2\chi_2) d\Omega_1 d\Omega_2 = \int \sigma \left(\frac{c\Omega}{2\gamma} \right) \exp(i\Omega\chi_1 - i\Omega\chi_2) d\Omega = \\ &= \sigma_1 \delta \left(\frac{K_a c}{2\gamma} - \Delta\chi \right), \quad (4) \end{aligned}$$

где $\Delta\chi = \chi_2 - \chi_1$.

Наличие в результате дельта-функции связано с тем, что интегрирование проводилось в бесконечных пределах. При учёте конечных размеров окна интегрирования (вопрос о выборе окна интегрирования в рамках этой статьи детально рассматриваться не будет) получим

$$\Psi(\Delta\chi) = \sigma_1 \exp \left[-i\Omega_a t - \frac{\zeta^2}{4} \left(K_a - \frac{2\gamma \Delta\chi}{c} \right)^2 \right]. \quad (5)$$

Здесь $\zeta = c\tau$ характеризует длину посылаемого импульса. Волновое число K_a акустической волны должно быть, по крайней мере, в 2–3 раза меньше $\Delta f / c$, где Δf — частотный интервал ЛЧМ импульса. Этот критерий определяет максимальную частоту акустического сигнала.

Изложенная выше процедура позволяет определить волновое число акустической волны K_a . При этом мы предполагаем, что рассеивающие частицы неподвижны в течение времени прохождения отдельного радиоимпульса (т. е. за это время смещение частиц существенно меньше длины акустической волны). Частота акустического сигнала определяется с помощью фурье-анализа последовательности принимаемых посылок.

Перейдём теперь к энергетическим оценкам. Локационная формула для радиоакустики имеет вид

$$P_s = \frac{P_{tr} L_0 k^4 \langle d \rangle^6 N l A}{4\pi z^3}. \quad (6)$$

Здесь P_{tr} — импульсная мощность передатчика, $L_0 = \sqrt{W_a G_a / (2\pi \rho_a c_s^3)}$ — параметр, характеризующий акустический излучатель, k — волновое число радиоволны, $\langle d \rangle$ — средний размер капель, N — концентрация капель, $l = c\tau$ — длина импульса, A — площадь приёмной антенны, z — дальность,

W_a — мощность акустического излучателя, G_a — коэффициент усиления антенны акустического излучателя, ρ_a — плотность воздуха, $c_s = \Omega_a/K_a$ — скорость звука. Для приведённых в [4] параметров установки $P_{tr} = 10^5$ Вт; $W_a = 100$ Вт; $G_a = 10$; $\rho_a = 1,3$ кг/м³; $L_0 = 2 \cdot 10^{-3}$ м; $\lambda = 2\pi/k = 3 \cdot 10^{-3}$ м; $\langle d \rangle^6 N = 10^{-18}$ м³; $l = 3 \cdot 10^3$ м; $A = 1$ м²; $z = 10^3$ м получим мощность сигнала $P_s = 10^{-10}$ Вт.

При температуре входа приёмника $T_r = 10^4$ К и полосе принимаемых частот (после свёртки с излучаемым импульсом) $\delta f = 10^6$ Гц мощность шумов приёмника составляет $P_r = 10^{-13}$ Вт. Таким образом, при приёме единичного импульса отношение сигнал/шум $P_s/P_r \approx 10^3$.

Фурье-анализ серии принимаемых импульсов обеспечивает дополнительное (некогерентное) накопление и, следовательно, увеличение отношения сигнал/шум в $\sqrt{N_{sc}}$ раз. Например, при частоте посылок импульсов $N_{sc} = 10^4$ с⁻¹ и частотном разрешении 1 Гц (время накопления 1 с) получим дополнительный выигрыш в отношении сигнал/шум в 100 раз.

Таким образом, согласно вышеприведённым грубым оценкам при высоте зондирования 1 км можно получить отношение сигнал/шум порядка 10^5 .

Возможности анализа ЛЧМ сигнала позволяют рассчитывать на то, что вместо второго фурье-преобразования можно использовать более оптимальную фильтрацию. Преимуществом радиоакустической системы на базе радара, работающего в миллиметровом диапазоне длин волн, является возможность использовать несколько акустических излучателей на различных частотах. Это может оказаться существенным, т. к. распространение звука в облаке, как и деформация волнового фронта, связанная с сильной турбулентностью, существенным образом зависит от частоты звука. Если полоса перестройки ЛЧМ сигнала составляет 100 МГц, для радиоакустического зондирования может быть использован диапазон акустических частот примерно до 100 Гц.

Работа выполнена при финансовой поддержке РФФИ (проекты № 00-02-17372, 01-02-16680).

СПИСОК ЛИТЕРАТУРЫ

1. Фабрикант А. Л. // Изв. вузов. Радиофизика. 1991. Т. 34, № 1. С. 12.
2. Рапопорт В. О., Митяков Н. А. // Изв. вузов. Радиофизика. 2000. Т. 43, № 12. С. 1 059.
3. Manheimer W. M. // Int. J. Electronics. 1992. V. 72, No. 5–6. P. 1 165.
4. Manheimer W. M. // Plasma Science and Environment. Woodbury, New York: AIP Press, 1997. P. 77.
5. Моченёва О. С., Токман М. Д. // Изв. АН. Сер. физическая. 1999. Т. 63, № 12. С. 2 384.

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

Поступила в редакцию
14 февраля 2001 г.

USING A MILLIMETER-WAVE METEORADAR FOR MEASUREMENT OF THE ATMOSPHERIC TEMPERATURE PROFILE

V. O. Rapoport and N. A. Mityakov

We consider a method of radio-acoustic sounding of the atmosphere using millimeter-wave radars based on powerful masers (gyrotrons) operated in the regime of intrapulse chirp modulation. The proposed method makes it possible to measure the atmospheric temperature profile without any modernization of the meteoradar transceiving system and to improve significantly the capabilities of remote monitoring of the atmosphere.

В. О. Рапопорт, Н. А. Митяков

УДК 621.372.821:538.561

О СПЕКТРЕ ВОЛН ПЛОСКОПАРАЛЛЕЛЬНОГО КИРОВОЛНОВОДА

В. В. Фисанов

Исследуется сложное поведение дисперсионных кривых собственных волн плоскопараллельных кироволноводов с идентичными (электрическими) и смешанными проводящими стенками, которое обусловлено многократным пересечением кривых, соответствующих двум семействам мод и испытывающих бифуркацию на частоте отсечки. Найдено полное число точек пересечения модовой пары, включающее точки пересечения как для распространяющихся, так и для нераспространяющихся мод. Определено асимптотическое поведение дисперсионных кривых в области высоких частот.

1. ВВЕДЕНИЕ. ДИСПЕРСИОННЫЕ УРАВНЕНИЯ ДЛЯ ВОЛНОВОДА С ИДЕНТИЧНЫМИ СТЕНКАМИ

Термин «кироволновод» был предложен около 15 лет назад, когда наметилась возможность практического осуществления эффектов оптической активности на СВЧ [1]. Целиком заполненный изотропной киральной средой плоскопараллельный кироволновод с идеально проводящими стенками, как наиболее простая волноведущая структура, был рассмотрен в первую очередь [1–3], а затем неоднократно подвергался дополнительным исследованиям. Такое внимание объясняется новизной и необычностью изучаемого объекта, а также отчасти отсутствием единого подхода к феноменологическому описанию киральной среды. Хотя впервые дисперсионное уравнение для собственных волн в таком волноводе было получено ранее в [4], в указанной работе оно не было факторизовано и детально не изучалось. Пропущенная в [1–3] основная мода спектра волн была впоследствии описана в [5–8], хотя возможность её существования была отмечена уже в [4].

Независимо от вида используемой системы материальных уравнений трансцендентные дисперсионные уравнения для собственных волн плоскопараллельного кироволновода высотой d могут быть представлены единообразно. Для чётных волн (поперечные к направлению распространения компоненты электромагнитного поля являются чётными функциями относительно поперечного смещения от оси волновода) уравнение имеет вид

$$\frac{q_+}{k_+} \operatorname{tg} \frac{q_+ d}{2} + \frac{q_-}{k_-} \operatorname{tg} \frac{q_- d}{2} = 0, \quad (1)$$

тогда как для нечётных волн (поперечные компоненты поля волны являются нечётными функциями смещения) происходит замена в (1) тангенсов на котангенсы:

$$\frac{q_+}{k_+} \operatorname{ctg} \frac{q_+ d}{2} + \frac{q_-}{k_-} \operatorname{ctg} \frac{q_- d}{2} = 0, \quad (2)$$

где $q_+ = (k_+^2 - K^2)^{1/2}$ и $q_- = (k_-^2 - K^2)^{1/2}$ — поперечные волновые числа, K — постоянная распространения, k_+ и k_- — волновые числа собственных волн безграничной изотропной киральной среды [2, 5, 7]. Для частот отсечки мод, когда длина волны в волноводе становится бесконечной, т. е. $K = 0$, из обоих уравнений выводится единое условие

$$\operatorname{tg} \frac{k_+ d}{2} + \operatorname{tg} \frac{k_- d}{2} = 0, \quad \text{или} \quad \sin \frac{(k_+ + k_-) d}{2} = 0. \quad (3)$$

Из (3) следует, что точки отсечки на дисперсионной диаграмме одновременно являются точками бифуркации мод, которая является характерным свойством кироволноводов.

Из совместного рассмотрения уравнений (1) и (2) выясняется, что существуют и другие точки пересечения дисперсионных кривых чётных и нечётных мод, на что впервые ясно было указано в работе [8]. Однако в ней, а также в [7], не обсуждается важный для структуры дисперсионных кривых вопрос о числе таких точек. Цель данной работы состоит в том, чтобы найти полное число точек пересечения в каждой модовой паре спектра волн, рассматривая всеволновую эволюцию дисперсионных кривых, определить их поведение в высокочастотной области и на этой основе уточнить влияние параметра киральности на формирование структуры дисперсионных кривых.

2. АНАЛИЗ ДИСПЕРСИОННЫХ ЗАВИСИМОСТЕЙ

Для анализа дисперсионных зависимостей следует определиться с выбором системы материальных уравнений киральной среды. Применяемые в настоящее время системы Поста, Теллегена, Кондона и Друде—Борна—Фёдорова в сочетании с однородными уравнениями Максвелла для гармонической временной зависимости обладают взаимно однозначным соответствием по материальным параметрам. Однако связанные с ними волновые параметры среды проявляют качественно различные свойства. В частности, отсутствует зависимость от параметра киральности у среднего арифметического волновых чисел k_+ и k_- в системах Теллегена и Кондона, у их среднего геометрического в системе Поста и у среднего гармонического в системе Друде—Борна—Фёдорова. Из (3) следует, что только в системах Теллегена и Кондона положение частот отсечки не зависит от параметра киральности. Поэтому ниже применяется система уравнений связи типа Кондона—Теллегена, которую для зависимости от времени вида $\exp(-i\omega t)$ можно представить так:

$$\mathbf{D} = \varepsilon \mathbf{E} + i \sqrt{\varepsilon \mu} G \mathbf{H}, \quad \mathbf{B} = \mu \mathbf{H} - i \sqrt{\varepsilon \mu} G \mathbf{E}, \quad (4)$$

где G — безразмерный параметр киральности, ε и μ — абсолютные значения диэлектрической и магнитной проницаемостей; обозначения электрических и магнитных полей общепринятые. В отсутствие потерь в среде псевдоскаляр G является вещественной величиной, $|G| < 1$. Пусть $G > 0$, тогда $k_+ > k_-$, причём

$$k_+ = k(1 + G), \quad k_- = k(1 - G), \quad (5)$$

где $k = \omega \sqrt{\varepsilon \mu}$ является их средним арифметическим и имеет смысл волнового числа в отсутствие киральности [7]. Условие отсечки (3) преобразуется к виду

$$kd = m\pi, \quad (6)$$

где $m = 0, 1, 2, \dots$, т. е. оказывается идентичным условию отсечки для некирального волновода с теми же значениями диэлектрической и магнитной проницаемостей. Каждому числу $m \geq 1$ соответствует пара мод с общей частотой отсечки, тогда как основная гибридная мода ($m = 0$) принадлежит к семейству (1) и при $G \rightarrow 0$ приобретает структуру ТЕМ-волны.

Дополнительные точки пересечения дисперсионных кривых находятся из пары условий

$$\sin[(q_+ + q_-)d/2] = 0, \quad \sin[(q_+ - q_-)d/2] = 0, \quad (7)$$

откуда следует, что

$$q_+ d = (l + n)\pi, \quad q_- d = (l - n)\pi, \quad (8)$$

где l, n — натуральные числа. Будем изучать дисперсионные зависимости относительно квадратичных параметров $x = (kd/\pi)^2$ и $y = (Kd/\pi)^2$, но не в квадранте $x > 0, y > 0$, как в [7], а на полуплоскости $x > 0$, что позволит рассмотреть эволюцию мод как в области распространения ($y > 0$), так и в запределной области ($y < 0$). Исходя из (8) и используя выражение $k_+^2 - k_-^2 = 4k^2 G$, найдём абсциссы точек пересечения:

$$x_{ln} = ln/G. \quad (9)$$

Тогда их ординаты находятся из двух равносильных выражений, следующих из (5):

$$y_{ln} = (1 + G)^2 x_{ln} - (l + n)^2, \quad y_{ln} = (1 - G)^2 x_{ln} - (l - n)^2. \quad (10)$$

Соотнесём число l с порядковым номером пары мод, полагая $l = m \neq 0$, и исключим из (9) и (10) число n . В результате получим уравнение линии расположения точек пересечения на плоскости xy , относящихся к m -й модовой паре:

$$y = [(x/m) - m] [m - (G^2 x/m)]. \quad (11)$$

Из (11) становится понятно, что на нулевой частоте дисперсионные кривые начинаются в запредельной области из общей точки $y = -m^2$.

Для того чтобы определить полное число точек пересечения каждой из мод, следует нанести на плоскость xy два семейства прямых:

$$y = (1 + G)^2 x - (m + n)^2, \quad (12)$$

$$y = (1 - G)^2 x - (m - n)^2. \quad (13)$$

Линии с $m \pm n = 0$ проходят в области распространяющихся мод; они разграничивают её на три части: сектор медленных волн ($K > k_+$, где дисперсионных кривых нет), промежуточный сектор ($k_- < K < k_+$, где q_- является мнимой величиной) и сектор быстрых волн ($K < k_-$, где оба аргумента тригонометрических функций являются вещественными величинами). Это полезное разграничение впервые было применено в [3]. Остальные линии семейств (12), (13) образуют на плоскости xy сетку с ячейками в форме параллелограмма, двумя вершинами которого, как очевидно, являются точки с абсциссами $x = mn/G$ и $x = m(n+1)/G$, где пересекаются дисперсионные кривые с модовым числом m . Две другие вершины имеют абсциссы $x = (m \pm 1/2)(n+1/2)/G$. Примечательно, что пересечение прямых (13) с линией (11) происходит в точках, принадлежащих семейству (9). Но поскольку при $n = m$ прямая (13) пересекает в точке с абсциссой $x = m^2/G$ только одну дисперсионную кривую, принадлежащую семейству чётных волн (1), значение $n = m$ следует исключить из подсчёта точек пересечения, производимого путём перебора значений n . Точка пересечения дисперсионной кривой семейства нечётных волн и прямой $y = (1 - G)^2 x$ находится из решения трансцендентного уравнения

$$\frac{\operatorname{tg}(\pi \sqrt{Gx})}{\pi \sqrt{Gx}} = -\frac{1 - G}{1 + G},$$

которое следует непосредственно из выражения (2), если положить в нём $q_- = 0$ и устранить неопределённость во втором слагаемом. Эта прямая является также касательной для линии точек пересечения (11), так что все точки пересечения сосредоточены ниже границы промежуточного сектора, при $y < (1 - G)^2 x$.

Итак, дисперсионные кривые каждой пары мод с номером m имеют m точек пересечения с координатами

$$x_{mn} = mn/G, \quad y_{mn} = [n/G - m] [m - Gn], \quad (14)$$

где $m = 1, 2, \dots$; $n = 0, 1, \dots, m - 1$, причём точки отсечки $x_m = m^2$, $y_m = 0$ в это семейство не входят. Как видно из (14), наложение точек пересечения на точки отсечки возможно, но происходит только при параметре киральности G , равном отношению целых чисел. Пользуясь (14), нетрудно также определить число точек пересечения, расположенных в области распространяющихся мод. Пусть $n = n^*$ — минимальное число, при котором $y_{mn} > 0$. Тогда, исходя из неравенства $n^*/G - m > 0$, находим, что на частотах выше частоты отсечки моды с номером m пересекаются $m - 1 - [mG]$ раз, где квадратные скобки обозначают антье (целую часть числа).

В промежуточном секторе обе дисперсионные кривые растут с увеличением x , асимптотически приближаясь к ближайшим прямым семейства (12): асимптотой нечётных волн является прямая $y =$

$= (1 + G)^2 x - (2m)^2$, асимптотой чётных волн — прямая $y = (1 + G)^2 x - (2m + 1)^2$. Таким образом, подтверждается наблюдение [5], что на высоких частотах определяющее влияние на спектр волн в волноводе оказывает только одна собственная волна среды, характеризующаяся бóльшим волновым числом (k_+ при $G > 0$). На частотах, которые находятся выше частоты последней точки пересечения, т. е. при $x > x_{m(m-1)}$, чётные волны m -й модовой пары распространяются с большей фазовой скоростью.

3. ВОЛНОВОД СО СМЕШАННЫМИ СТЕНКАМИ

Применим изложенный выше подход к исследованию спектра волн плоскопараллельного кироволновода со смешанными стенками, т. е. при задании условия идеальной электрической границы $\mathbf{n} \times \mathbf{E} = 0$ на одной из них и условия идеальной магнитной границы $\mathbf{n} \times \mathbf{H} = 0$ на другой, где $\mathbf{n}$ — нормаль к границе стенки. Возможной технической реализацией магнитной стенки является поверхность плазмы с критической плотностью [9]. Структура со смешанными стенками находит применение в теории волноводов для эквивалентного описания поля в волноводе удвоенной высоты, которое обладает определённым типом симметрии относительно серединной плоскости [10].

Представляя электромагнитное поле в волноводе в виде суперпозиции «волновых полей» [11] с общей постоянной распространения K и учитывая граничные условия, получим пару дисперсионных уравнений:

$$(q_+/k_+) \operatorname{tg}[(q_+d)/2] - (q_-/k_-) \operatorname{ctg}[(q_-d)/2] = 0, \quad (15)$$

$$(q_+/k_+) \operatorname{ctg}[(q_+d)/2] - (q_-/k_-) \operatorname{tg}[(q_-d)/2] = 0. \quad (16)$$

Моды обоих семейств, характеризующихся уравнениями (15) и (16), имеют общее условие отсечки $\cos[(k_+ + k_-)d/2] = 0$, следовательно, частоты отсечки, определяемые выражением

$$kd = \pi(M + 1/2), \quad (17)$$

где $M = 0, 1, 2, \dots$, соответствуют точкам бифуркации дисперсионных кривых на дисперсионной диаграмме. Из совместного рассмотрения уравнений (15) и (16) следуют условия пересечения дисперсионных кривых:

$$\cos \frac{(q_+ + q_-)d}{2} = 0, \quad \cos \frac{(q_+ - q_-)d}{2} = 0, \quad (18)$$

откуда находим

$$q_+d = [(L + 1/2) + (N + 1/2)]\pi, \quad q_-d = [(L + 1/2) - (N + 1/2)]\pi, \quad (19)$$

где $1 \leq N \leq L$, $L = 1, 2, \dots$. Выражения (17) и (19) формально переводятся в соответствующие выражения (6) и (8), если положить $m = M + 1/2$, $l = L + 1/2$, $n = N + 1/2$, так что положение точек пересечения и другие свойства дисперсионных кривых волновода со смешанными стенками можно найти, пользуясь результатами анализа структуры мод кироволновода с идентичными стенками. При этом каждой модовой паре с номером $M > 0$ соответствуют согласно (14) M точек пересечения с координатами

$$x_{MN} = \frac{(M + 1/2)(N + 1/2)}{G}, \quad y_{MN} = \left[\frac{N + 1/2}{G} - \left(M + \frac{1}{2} \right) \right] \left[M + \frac{1}{2} - G \left(N + \frac{1}{2} \right) \right], \quad (20)$$

где $M = 1, 2, \dots$; $N = 0, 1, \dots, M - 1$. В отличие от волновода с идентичными стенками, на нулевой частоте дисперсионные кривые начинаются из разных точек:

$$y_M(0) = - \left[\frac{1}{2} + M \pm (-1)^M \frac{2}{\pi} \operatorname{arctg} \frac{G}{1 + \sqrt{1 - G^2}} \right]^2, \quad (21)$$

где $M = 0, 1, 2, \dots$. Здесь для кривых семейства (15) берётся верхний знак, а для кривых семейства (16) — нижний знак. В промежуточном секторе, где $(1-G)^2 x < y < (1+G)^2 x$, дисперсионные кривые семейства (15) имеют асимптоты $y = (1+G)^2 x - (2M+1)^2$, а кривые семейства (16) — асимптоты $y = (1+G)^2 x - (2M+2)^2$.

4. ЗАКЛЮЧЕНИЕ

У каждой пары мод плоскопараллельного кироволновода, испытывающей бифуркацию на общей частоте отсечки, происходит многократное наложение дисперсионных кривых. Полное число общих точек (не считая точку бифуркации) равно порядковому номеру модовой пары, если принимать во внимание эволюцию кривых как в области распространения, так и на частотах ниже частоты отсечки. Число точек пересечения дисперсионных кривых распространяющихся мод изменяется в зависимости от их номера и от параметра киральности, что следует принимать во внимание при использовании кироволноводов в качестве направленных ответвителей [1] и для иных применений. Устойчивое различие фазовых скоростей в высокочастотной области может быть положено в основу идентификации двух семейств мод плоскопараллельных волноводов.

СПИСОК ЛИТЕРАТУРЫ

1. Engheta N., Pelet P. // Opt. Lett. 1989. V. 14, No. 11. P. 593.
2. Varadan V. K., Lakhtakia A., Varadan V. V. // J. Wave—Mater. Interact. 1988. V. 3, No. 3. P. 267; 1988. V. 3, No. 4. P. 351. (Errata)
3. Pelet P., Engheta N. // IEEE Trans. Antennas Propagat. 1990. V. 38, No. 1. P. 90.
4. Альтшулер Е. Л., Москалев А. Н., Рыжов В. А., Рындин Р. М., Сильвестров П. Г., Фомичев В. Н., Хриплович И. Б. // ЖТФ. 1984. Т. 54, № 10. С. 1956.
5. Mahmoud S. F. // J. Electromagn. Waves Appl. 1992. V. 6, No. 10. P. 1381.
6. Zhang L.-Y., Jiao Y.-C., Liang C.-H. // IEEE Trans. Microwave Theory Tech. 1994. V. 42, No. 10. P. 2009.
7. Lekner J. // Pure Appl. Opt. 1997. V. 6, No. 3. P. 373.
8. Cory H., Waxman S. // IEE Proc. Microw. Antennas Propagat. 1994. V. 141, No. 4. P. 299.
9. Генчев Ж. Д. // Изв. инст. электрон. Бълг. АН. 1969. Т. 4. С. 17.
10. Миттра Р., Ли С. Аналитические методы теории волноводов. М.: Мир, 1974. 328 с.
11. Третьяков С. А. // Радиотехника и электроника. 1994. Т. 39, № 10. С. 1457.

Сибирский физико-технический институт Томского
государственного университета, г. Томск, Россия

Поступила в редакцию
18 июля 2001 г.

MODE SPECTRUM OF A PARALLEL-PLATE CHIROWAVEGUIDE

V. V. Fisanov

We study the complicated behavior of the dispersion curves of eigenmodes of parallel-plate chirowaveguides with identical (electric) and mixed (nonidentical) conducting walls, which stems from multiple crossovers of the curves corresponding to two sets of modes and undergoing a bifurcation at the cutoff frequency. We find the total number of the crossover points in each modal pair, including both propagating and evanescent waves. The asymptotic behavior of the dispersion curves in a high-frequency region is determined.

УДК 501+530.1

О СТРУКТУРЕ РЕШЕНИЙ ОБОБЩЁННЫХ УРАВНЕНИЙ КАДОМЦЕВА ПЕТВИАШВИЛИ

В. Ю. Белашов

Рассматриваются обобщения полученных ранее для одномерного случая результатов на двумерные и трёхмерные уравнения КП-класса с произвольным показателем нелинейности, учётом дисперсионной поправки высшего порядка и членов, описывающих диссипацию и неустойчивость. Построены асимптотики солитонных и несолитонных решений. Обсуждается конструирование фазовых портретов в 8-мерном фазовом пространстве на основе результатов качественного анализа обобщённых уравнений Кортевега-де-Вриза.

В работе на основе результатов, полученных ранее для одномерного случая [1], рассматривается подход к изучению структуры неодномерных решений уравнения Кадомцева—Петвиашвили (КП), обобщённого введением дисперсионной поправки высшего порядка и членов, описывающих диссипацию «вязкостного» типа и неустойчивость [2, 3]:

$$\partial_{\eta} (\partial_t u + \alpha u \partial_{\eta} u - \mu \partial_{\eta}^2 u + \beta \partial_{\eta}^3 u + \delta \partial_{\eta}^4 u + \gamma \partial_{\eta}^5 u) = \kappa \Delta_{\perp} u, \quad (1)$$

где $\Delta_{\perp} = \partial_{\zeta_1}^2 + \partial_{\zeta_2}^2$. Данная модель описывает распространение нелинейных волн в средах со слабыми нелинейностью и дисперсией и имеет широкое приложение в физике ионосферной и магнитосферной плазмы, гидродинамике и аэродинамике.

При $\mu = \delta = \gamma = 0$ уравнение (1) переходит в известное уравнение КП [4], представляющее собой полностью интегрируемую гамильтонову систему и имеющее в случае $\Delta_{\perp} = \partial_{\zeta_1}^2$ решения в виде одномерных (при $\beta\kappa < 0$) или двумерных (при $\beta\kappa > 0$) солитонов [4, 5]. Структура и динамика решений не интегрируемой аналитически модели (1) в случае $\delta = 0$ была детально исследована в работах [6, 7], где было показано, что при $\mu = 0$ в зависимости от знаков коэффициентов β , γ и κ могут иметь место двумерные или трёхмерные решения солитонного типа с монотонными либо осциллирующими асимптотиками, которые при наличии в среде диссипации «вязкостного» типа ($\mu > 0$) в ходе эволюции теряют симметрию своей структуры и затухают. В работе [1] методами асимптотического и качественного анализа были детально изучены асимптотики одномерного аналога уравнения (1) и построена достаточно полная классификация его решений, включая решения солитонного и несолитонного типов.

Цель настоящей работы состоит в обобщении результатов, полученных в [1], на неодномерные случаи с учётом результатов [3]. Во избежание громоздкости получающихся выражений будем рассматривать уравнение (1) в двумерном виде, положив $\Delta_{\perp} = \partial_{\zeta_1}^2$. Обобщение используемой техники и получаемых при этом результатов на случай $\Delta_{\perp} = \partial_{\zeta_1}^2 + \partial_{\zeta_2}^2$, как мы увидим ниже, достаточно тривиально. Положим $\zeta_1 \equiv \zeta$ и примем для определённости $\alpha = 6$ (что легко получить путём масштабного преобразования $u \rightarrow 6u/\alpha$ в уравнении (1)).

Введём новые переменные $\bar{\eta} = \eta + \zeta$, $\bar{\zeta} = \eta - \zeta$. В результате, выражая переменные η и ζ в уравнении (1) сначала через $\bar{\eta}$, а затем через $\bar{\zeta}$, получим пару одномерных уравнений

$$\begin{aligned} \partial_{\bar{\eta}} (\partial_t u + 6u \partial_{\bar{\eta}} u - \mu \partial_{\bar{\eta}}^2 u + \beta \partial_{\bar{\eta}}^3 u + \delta \partial_{\bar{\eta}}^4 u + \gamma \partial_{\bar{\eta}}^5 u) &= \kappa \partial_{\bar{\eta}}^2 u, \\ \partial_{\bar{\zeta}} (\partial_t u + 6u \partial_{\bar{\zeta}} u - \mu \partial_{\bar{\zeta}}^2 u + \beta \partial_{\bar{\zeta}}^3 u + \delta \partial_{\bar{\zeta}}^4 u + \gamma \partial_{\bar{\zeta}}^5 u) &= \kappa \partial_{\bar{\zeta}}^2 u, \end{aligned} \quad (2)$$

записанных в системе координат, оси $\bar{\eta}$ и $-\bar{\zeta}$ которой повернуты относительно осей η и ζ на 45° . Представление (2) фактически означает, что исходное уравнение (1) допускает два типа одномерных решений вида $u(\bar{\eta}, t)$ и $u(\bar{\zeta}, t)$, удовлетворяющих первому и второму уравнениям системы (2) соответственно.

В. Ю. Белашов

411

При этом, однако, необходимо иметь в виду, что «одномерность» этих решений всё же неявно предполагает линейную зависимость каждой из новых переменных $\bar{\eta}$ и $\bar{\zeta}$ от обеих координат η и ζ .

Интегрируя уравнения (2) по $\bar{\eta}$ и $-\bar{\zeta}$ соответственно, получим одинаковые по форме обобщённые уравнения Кортевега-де-Вриза (КдВ):

$$\begin{aligned}\partial_t u + (-\kappa + 6u) \partial_{\bar{\eta}} u - \mu \partial_{\bar{\eta}}^2 u + \beta \partial_{\bar{\eta}}^3 u + \delta \partial_{\bar{\eta}}^4 u + \gamma \partial_{\bar{\eta}}^5 u &= 0, \\ \partial_t u + (-\kappa + 6u) \partial_{\bar{\zeta}} u - \mu \partial_{\bar{\zeta}}^2 u + \beta \partial_{\bar{\zeta}}^3 u + \delta \partial_{\bar{\zeta}}^4 u + \gamma \partial_{\bar{\zeta}}^5 u &= 0,\end{aligned}\quad (3)$$

связанные друг с другом через произведённую выше замену переменных. Переходя теперь в систему координат, движущуюся вдоль соответствующей оси со скоростью $-\kappa$, т. е. выполняя в уравнениях (3) замену $\eta' = \bar{\eta} + \kappa t$, $\zeta' = \bar{\zeta} + \kappa t$ и опуская штрихи, запишем (3) в стандартном виде:

$$\begin{aligned}\partial_t u + 6u \partial_{\eta} u - \mu \partial_{\eta}^2 u + \beta \partial_{\eta}^3 u + \delta \partial_{\eta}^4 u + \gamma \partial_{\eta}^5 u &= 0, \\ \partial_t u + 6u \partial_{\zeta} u - \mu \partial_{\zeta}^2 u + \beta \partial_{\zeta}^3 u + \delta \partial_{\zeta}^4 u + \gamma \partial_{\zeta}^5 u &= 0.\end{aligned}\quad (4)$$

Таким образом, теперь можно провести анализ одного обобщённого уравнения системы (3), а затем, выполнив обратную замену переменных, распространить результаты на двумерные решения $u(\eta, \zeta, t)$ уравнения (1) с $\Delta_{\perp} = \partial_{\zeta}^2$.

В работе [1] рассмотрен наиболее общий случай, когда класс уравнений (4) расширен введением произвольного положительного показателя степени нелинейности p , при этом, например, первое уравнение системы (4) приобретает вид

$$\partial_t u + 6u^p \partial_{\eta} u - \mu \partial_{\eta}^2 u + \beta \partial_{\eta}^3 u + \delta \partial_{\eta}^4 u + \gamma \partial_{\eta}^5 u = 0 \quad (5)$$

и в случае $\mu = \delta = \gamma = 0$ представляет собой известное уравнение КдВ, если $p = 1$, и модифицированное уравнение КдВ (МКдВ), если $p = 2$. Отметим, что случаи, когда в уравнении (5) и, соответственно, в (1) нелинейный член имеет вид $6u^p \partial_{\eta} u$, где $p = 1, 2$, наиболее интересны с физической точки зрения, приложения же уравнения (5) с $p > 2$ в настоящее время неизвестны. Однако в связи с тем, что уравнения (5) с произвольным целым $p > 0$ обнаруживают в значительной степени сходные математические свойства, в [1] был осуществлён общий подход к выяснению зависимости параметров решений от показателя нелинейности.

С учётом того, что $\mu > 0$ и $\delta > 0$ (в соответствии с физическим смыслом соответствующих членов, описывающих диссипацию и неустойчивость), положив без потери общности аналогично работе [1] $\gamma > 0$, $\beta = \pm 1$ и сделав замену $u = Vw$, для случаев различных знаков V и β результаты [1] могут быть обобщены на уравнения (2) и, соответственно, (1) с $p \geq 1$ следующим образом.

1) Значение показателя нелинейности p определяет характер зависимости $V = f(u)$: при $p > 1$ такая зависимость для уравнения (1), как и в одномерном случае (см. [1]), становится нелинейной в отличие от известной линейной зависимости при $p = 1$ (например, в случае уравнения КП). Кроме того, для чётных p решения уравнения (1) могут иметь как положительную ($u > 0$), так и отрицательную ($u < 0$) поляризованность для любого знака V .

2) В случае консервативных уравнений класса (1) (случай $\mu = \delta = \gamma = 0$) асимптотики решений определяются следующими соотношениями:

а) для случаев $V > 0$; $\beta = -1$ и $V < 0$; $\beta = 1$ (верхний и нижний знаки соответственно)

$$w = A_1 \exp \left[(2\gamma)^{-1/2} \left(C^2 + \sqrt{C^4 \pm 4\gamma} \right)^{1/2} \chi \right]; \quad (6)$$

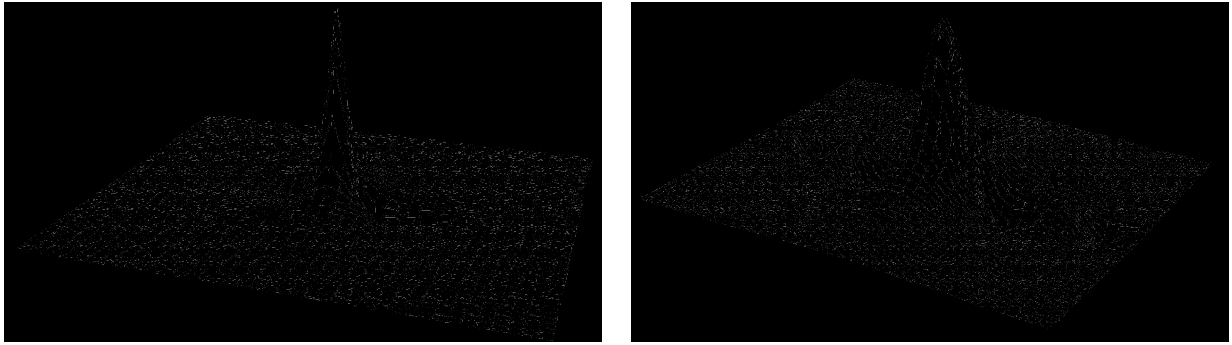


Рис. 1. Общий вид двумерного солитона уравнения (1) с $\Delta_{\perp} = \partial_{\zeta}^2$ при $\mu = \delta = 0, p = 1$: (а) $\gamma = 1; \beta = -0,6$ ($t = 0,2$); (б) $\gamma = 1; \beta = -3,16$ ($t = 0,5$)

б) для случая $V < 0, \beta = 1$

$$w = A_2 \exp \left[(2C^{-1}\gamma^{-1/2})^{-1} (2C^{-2}\gamma^{1/2} - 1)^{1/2} \chi \right] \cos \left[(2C^{-1}\gamma^{-1/2})^{-1} (2C^{-2}\gamma^{1/2} + 1) \chi + \Theta \right], \quad (7)$$

где A_1, A_2 и Θ — произвольные постоянные, $C = |V|^{-1/4}$, $\chi = [\eta \pm \zeta + (\kappa - V)t]$; здесь знаки плюс и минус относятся соответственно к первому и второму уравнениям системы (2). Как можно видеть из выражений (6), (7),¹ в решениях $u(\eta, \zeta, t)$ уравнения (1) при $\mu = \delta = 0$ в зависимости от знаков V и β могут иметь место солитоны как с монотонными, так и с осциллирующими асимптотиками. Отметим, что при $\beta = 0$ и любом значении $\gamma > 0$ решения уравнений (2) с $\mu = \delta = 0$ имеют вид $w = (A_1 + A_2 C^{-1} \chi) \exp(\gamma^{-1/4} C^{-1} \chi)$ и, следовательно, также описывают солитон с монотонными асимптотиками [7]. На рис. 1 показаны результаты численного интегрирования уравнения (1) при $\mu = \delta = 0$ при начальном условии $u = u_0 \exp(-\zeta^2/l_1^2 - \eta^2/l_2^2)$, где l_1, l_2 — некоторые постоянные, которые подтверждают результаты асимптотического анализа.

3) В случае диссипативных уравнений класса (1) с неустойчивостью (случай $\beta = \gamma = 0$) асимптотики решений определяются следующими соотношениями:

а) для $\delta > (4/27) \mu^3 C^8$

$$w = A_1 \exp \left[(2\delta C)^{-1/3} Q_1^+ \chi \right] + \exp \left[-(16\delta C)^{-1/3} Q_1^+ \chi \right] \left\{ A_2 \cos \left[\sqrt{3} (16\delta C)^{-1/3} Q_1^- \chi + \Theta_1 \right] + A_3 \sin \left[\sqrt{3} (16\delta C)^{-1/3} Q_1^- \chi + \Theta_2 \right] \right\}; \quad (8)$$

б) для $\delta = (4/27) \mu^3 C^8$

$$w = A_1 \exp \left[(2\delta C/4)^{-1/3} \chi \right] + A_2 (1 + A_3 \chi) \exp \left[-(2\delta C)^{-1/3} \chi \right]; \quad (9)$$

в) для $\delta < (4/27) \mu^3 C^8$

$$w = A_1 \exp \left[(2\delta C/4)^{-1/3} \operatorname{Re}(Q^+) \chi \right] + A_2 \exp \left\{ -(2\delta C)^{-1/3} \chi \left[\operatorname{Re} Q^+ - \sqrt{3} |\operatorname{Im} Q^+| \right] \right\} + A_3 \exp \left\{ -(2\delta C)^{-1/3} \chi \left[\operatorname{Re} Q^+ + \sqrt{3} |\operatorname{Im} Q^+| \right] \right\}, \quad (10)$$

где $A_1, A_2, A_3, \Theta_1, \Theta_2$ — произвольные постоянные, $Q_1^+ = Q^+ + Q^-$, $Q_1^- = Q^+ - Q^-$,

$$Q^+ = \left[1 + \sqrt{1 - 4\mu^3 C^8 / (27\delta)} \right]^{1/3}, \quad Q^- = \left[1 - \sqrt{1 - 4\mu^3 C^8 / (27\delta)} \right]^{1/3};$$

¹ Случаи других соотношений знаков V и β не реализуются (см. [1]).

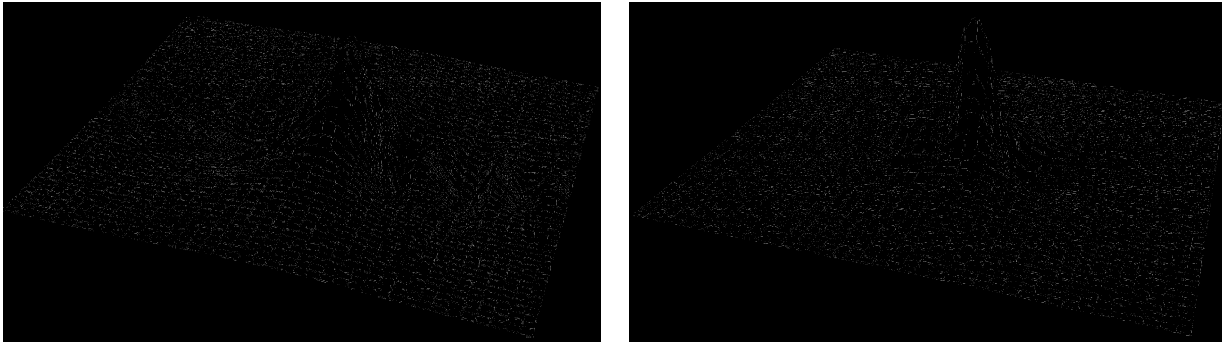


Рис. 2. Общий вид двумерного солитона, описываемого уравнением (1) с $\Delta_{\perp} = \partial_{\zeta}^2$ при $\beta = \gamma = 0$, $V > 0$:
(а) $\mu = 1$; $\delta = 1$ ($\delta > (4/27)\mu^3 C^8$) при $t = 3$; (б) $\mu = 1$; $\delta = 1 \cdot 10^{-6}$ ($\delta < (4/27)\mu^3 C^8$) при $t = 0,2$

параметры Q^+ и Q^- действительные в случаях (а) и (б) и комплексные в случае (в).

Из формул (8)–(10) легко видеть, что решения $u(\eta, \zeta, t)$ уравнения (1) имеют осциллирующие асимптотики в случае (а) и экспоненциальные — в случаях (б) и (в). На рис. 2 показаны численные решения уравнения (1), отвечающие случаям (а) и (в) при начальном условии $u = u_0 \exp(-\zeta^2/l_1^2 - \eta^2/l_2^2)$.

4) Что касается соответствующей трансформации фазовых портретов и «связывания» их в случае двумерного уравнения, то здесь в связи с тем, что при $\mu = \delta = 0$ фазовое пространство является 8-мерным, а при $\beta = \gamma = 0$ — 6-мерным, мы вынуждены ограничиться полученными в [1] результатами, связывая характеристики каждой особой точки каждого из уравнений системы (2) соответственно в 8-мерном и 6-мерном пространствах. При этом тип особых точек в каждом из 4-мерных или 3-мерных подпространств (см. [1]) при обратном преобразовании координат $\eta = (\bar{\eta} + \bar{\zeta})/2$, $\zeta = (\bar{\eta} - \bar{\zeta})/2$ не изменится; изменяются лишь те параметры фазовых портретов, которые для решений одного и того же класса соответствуют изменению таких параметров, как амплитуда, крутизна фронтов, частота осцилляций и т. п.

В заключение отметим, что нами рассмотрены специальные случаи, когда в уравнении (1) $\mu = \delta = 0$ и $\beta = \gamma = 0$. Для других значений коэффициентов могут наблюдаться более сложные волновые структуры, обусловленные наличием всех рассмотренных эффектов в их совокупности. Так, результаты, полученные численно в [8, 9], показывают, что для ненулевых параметров β , μ , δ при наличии гауссовских случайных флуктуаций волнового поля для гармонических начальных условий и начальных условий в форме уединённого импульса в процессе временной эволюции также могут формироваться устойчивые волновые структуры солитонного типа. Более того, устойчивые солитонные структуры могут формироваться и при $\gamma \neq 0$. Однако аналитическое изучение таких случаев весьма сложно, хотя при этом также может использоваться рассмотренный выше подход. Отметим также, что полученные в настоящей работе результаты могут быть весьма полезны при изучении решений и интерпретации многомерных фазовых портретов более сложных неоднородных модельных уравнений (см., например, работу [3]).

Работа выполнена при финансовой поддержке РФФИ (грант № 01–02–16116)

СПИСОК ЛИТЕРАТУРЫ

1. Белашов В. Ю., Тюнина С. Г. // Изв. вузов. Радиофизика. 1997. Т. 40, № 3. С. 328.
2. Карпман В. И., Белашов В. Ю. О структуре двумерных осциллирующих солитонов в слабо диспергирующих средах: Препринт № 25(972) ИЗМИРАН. М., 1991. 19 с.
3. Белашов В. Ю. Уравнение КП и его обобщения. Теория, приложения. Магадан: СВКНИИ ДВО РАН, 1997.
4. Кадомцев Б. Б., Петвиашвили В. И. // ДАН СССР. 1970. Т. 192, № 4. С. 753.

5. Петвиашвили В. И. // Физика плазмы. 1976. Т. 2, вып. 3. С. 496.
6. Karpman V. I., Belashov V. Yu. // Phys. Lett. A. 1991. V. 154, No. 3–4. P. 131.
7. Karpman V. I., Belashov V. Yu. // Phys. Lett. A. 1991. V. 154, No. 3–4. P. 140.
8. Belashov V. Yu. // Phys. Lett. A. 1995. V. 197. P. 282.
9. Kawahara T. // Phys. Rev. Lett. 1983. V. 51, No. 5. P. 381.

Северо-Восточный комплексный НИИ ДВО РАН,
г. Магадан, Россия

Поступила в редакцию
2 февраля 1999 г.

STRUCTURE OF THE SOLUTIONS OF GENERALIZED KADOMTSEV–PETVIASHVILI EQUATIONS

V. Yu. Belashov

We consider generalizations of the earlier results concerning the one-dimensional case to two- and three-dimensional equations of Kadomtsev–Petviashvili (KP) class with an arbitrary nonlinearity index, taking into account the higher-order dispersion correction and the terms describing dissipation and instability. The asymptotics of soliton and nonsoliton solutions are derived. Constructing phase portraits in the 8-dimensional space based on the results of a qualitative analysis of generalized Korteweg–de Vries equations is discussed.

УДК 621.315.592

О ПРИРОДЕ «НИЗКОТЕМПЕРАТУРНОЙ АНОМАЛИИ» В КОНТАКТАХ МЕТАЛЛ—ПОЛУПРОВОДНИК С БАРЬЕРОМ ШОТКИ

В. Г. Божков

Показано, что «низкотемпературная аномалия» вольт-амперной характеристики однородных контактов металл—полупроводник с барьером Шоттки (рост показателя идеальности n и уменьшение измеряемой по току насыщения высоты барьера φ_{bm} с понижением температуры, а также близость их произведения к реальной высоте барьера) является следствием влияния двух факторов: 1) нелинейной зависимости реальной высоты барьера от смещения, причём такой, что показатель идеальности вольт-амперной характеристики растёт с увеличением смещения, и 2) измерения параметров n и φ_{bm} в диапазоне температур при одном и том же токе. Существенно уточнено выражение для высоты барьера φ_{bf} , соответствующей плоским зонам в контакте.

ВВЕДЕНИЕ

Как известно, реальные вольт-амперные характеристики (ВАХ) контактов металл—полупроводник (КМП) с барьером Шоттки (БШ) могут значительно отличаться от идеальных [1, 2]. Существенное отклонение в поведении ВАХ, получившее название «низкотемпературная аномалия», было обнаружено уже в одной из наиболее ранних работ, посвящённых КМП с БШ Au—GaAs [3]. В дальнейшем эти особенности подтвердились в целом ряде работ для контактов не только на основе GaAs [4–8], но и на основе Si [8–10], InP [8, 11] и других полупроводников, в т. ч. в работах самого последнего времени. Суть «низкотемпературной аномалии» заключается в следующем:

1) с понижением температуры в КМП с БШ наблюдаются рост показателя идеальности n ВАХ и уменьшение измеряемой по току насыщения высоты барьера φ_{bm} , при этом практически неизвестны случаи, когда величины n и φ_{bm} изменялись бы независимо друг от друга; температурная зависимость показателя идеальности в целом близка к виду $n = 1 + T_0/T$ (так называемый T_0 -эффект), где T_0 — константа или слабо зависящая от температуры величина;

2) величина φ_{bn} , равная произведению показателя идеальности на измеряемую высоту барьера ($\varphi_{bn} \equiv n\varphi_{bm}$), слабо меняется с температурой и близка к реальной высоте барьера в КМП, определённой другими методами, например из вольт-фарадных характеристик.

Заметим, что речь идёт о контактах со слаболегированными полупроводниками. ВАХ таких контактов при комнатной температуре близка к идеальной ($n < 1,1$), и объяснение указанных эффектов туннелированием носителей заряда через барьер исключено.

Анализ причин «низкотемпературной аномалии» ВАХ стал предметом многочисленных исследований и дискуссий, продолжающихся до настоящего времени. Однако необходимой ясности в этом вопросе по-прежнему нет (см., например, [1, с. 139]). Между тем эта проблема имеет и очевидное практическое значение, поскольку контакты металл—полупроводник с барьером Шоттки являются основой широкого круга диодов и транзисторов и в значительной степени определяют их свойства. Например, параметр n непосредственно определяет шумовую температуру диодов с БШ (ДБШ) (см., например, [12]), а значит, и шумовую температуру охлаждаемых приёмных устройств миллиметрового (прежде всего) диапазона длин волн, в смесителях которых широко используются ДБШ на основе GaAs.

Заслуживает специального рассмотрения и вопрос о высоте барьера при плоских зонах (ВБПЗ). Особое его значение связано с тем, что использование этого понятия, в т. ч. и при низких температурах, стало довольно распространённым среди исследователей КМП с барьером Шоттки. По существу,

ВБПЗ получила статус фундаментальной характеристики контактов с БШ. В подтверждение её «фундаментальности» авторы указывают на слабую зависимость ВБПЗ от параметров обработки контакта [13], от температуры [14], от концентрации носителей заряда [15], на её соответствие высоте барьера, определённой из вольт-фарадных характеристик, и на другие свойства контактов [16, 17].

Вместе с тем это понятие введено в предположении независимости показателя идеальности ВАХ от смещения [13], что равносильно требованию линейной зависимости высоты барьера от смещения [1, 2]. Поскольку это допущение, на наш взгляд, неоправдано в широком диапазоне смещений и в случае однородных контактов несовместимо с наличием в них «низкотемпературной аномалии» (что будет показано ниже), возникает необходимость в дополнительном анализе этого вопроса.

В связи с вышесказанным целью предлагаемой работы является попытка более строгого, систематического подхода к анализу измеряемых параметров n и φ_{bm} ВАХ, их связи с реальной высотой барьера (при заданных условиях измерения), с высотой барьера при плоских зонах и с проявлением T_0 -эффекта в широком диапазоне температур. Для достижения этой цели принципиально важными в нашем анализе оказываются два обстоятельства: учёт нелинейной зависимости высоты барьера от смещения, приводящей к росту показателя идеальности ВАХ с увеличением смещения, и условие постоянства тока при определении высоты барьера и показателя идеальности ВАХ в диапазоне температур.

ТЕОРЕТИЧЕСКИЙ АНАЛИЗ

Мы полагаем, что экспериментальная ВАХ реальных контактов металл—полупроводник с БШ может быть представлена в виде

$$I = AR^*T^2 \exp\left(-\frac{q\varphi_{\text{b}}(V)}{kT}\right) \exp\left(\frac{qV}{kT}\right). \quad (1)$$

Здесь A — площадь контакта, R^* — эффективная константа Ричардсона, k — постоянная Больцмана, q — заряд электрона, T — температура, $\varphi_{\text{b}}(V)$ — зависящая от смещения V высота барьера, которую в дальнейшем будем называть реальной или истинной высотой барьера. Наше предположение не противоречит известным экспериментальным данным [1, 2] и соответствует всем наиболее известным моделям КМП с БШ: модели, учитывающей эффект сил зеркального изображения [1], модели Бардина [1], модели Левина [18] и другим [8, 19]. В то же время нужно учесть, что при наличии на контакте промежуточного слоя, прозрачность которого зависит от смещения, при учёте туннелирования через барьер, при участии поверхностных электронных состояний в прохождении тока и в некоторых других случаях высота барьера может иметь смысл эффективной. Это обстоятельство может быть существенным при исследовании механизма прохождения тока в КМП, но при корректном проведении и интерпретации измерений (чему дальше уделено особое внимание) оно не играет значительной роли.

В таком же виде ВАХ реальных КМП с БШ принята и в работе [13], где было получено выражение, связывающее высоту барьера при нулевом смещении на барьере ($\varphi_{\text{b}0} \equiv \varphi_{\text{b}}(0)$) и при плоских зонах ($\varphi_{\text{bf}} \equiv \varphi_{\text{b}}(V_{\text{bf}})$) с показателем идеальности ВАХ:

$$\varphi_{\text{bf}} = n\varphi_{\text{b}0} - (n - 1)\varphi_{\text{s}}. \quad (2)$$

Здесь $\varphi_{\text{bf}} = V_{\text{bf}} + \varphi_{\text{s}}$ — высота барьера при плоских зонах (по определению), V_{bf} — смещение, соответствующее состоянию плоских зон в БШ, $\varphi_{\text{s}} = (kT/q) \ln(N_{\text{C}}/N_{\text{D}})$ — потенциал уровня Ферми, который отсчитывается от дна зоны проводимости, N_{C} — эффективная плотность состояний в зоне проводимости, N_{D} — концентрация донорной примеси в полупроводнике.

Принципиальным при выводе (2) является предположение о линейной зависимости высоты барьера от смещения, т.е.

$$\varphi_b(V) = \varphi_{b0} + \left. \frac{d\varphi_b(V)}{dV} \right|_{V=0} V. \quad (3)$$

В этом случае в соответствии с известным определением показателя идеальности ВАХ (см. ниже) он не зависит от смещения, поскольку связан с высотой барьера соотношением

$$n(V) = \left[1 - \left. \frac{d\varphi_b(V)}{dV} \right|_{V=0} \right]^{-1}. \quad (4)$$

Это приближение, как показывают известные результаты [13–16, 20], вполне адекватно описывает свойства реальных КМП с БШ при достаточно высоких температурах. Однако, как уже было сказано, оно не позволяет объяснить наличие «низкотемпературной аномалии» в различных однородных контактах с БШ.

Положим, что зависимость $\varphi_b(V)$ в выражении (1) нелинейна. Формально она может быть представлена в виде ряда, ограниченного квадратичным членом разложения:

$$\varphi_b(V) = \varphi_{b0} + \left. \frac{d\varphi_b(V)}{dV} \right|_{V=0} V + \frac{1}{2} \left. \frac{d^2\varphi_b(V)}{dV^2} \right|_{V=0} V^2. \quad (5)$$

Из (4) и (5) следует, что показатель идеальности ВАХ становится при этом функцией смещения. Мы полагаем далее, что показатель идеальности растёт с увеличением смещения (вторая производная в (5) положительна), что также обычно выполняется для реальных контактов, особенно при достаточно больших токах (см., например, [12]). В такой ситуации ВАХ (1) в полулогарифмическом масштабе становится нелинейной и может быть представлена условной кривой 1 на рис. 1.

Как известно, для определения параметров n и φ_{bm} из экспериментальной (реальной) ВАХ к последней (в полулогарифмическом масштабе) в некоторой точке $(\bar{V}, \bar{I})$ строится касательная (прямая 2 на рис. 1). Высоту барьера φ_{bm} определяют по точке пересечения касательной с осью ординат (по току насыщения I_s , который на рис. 1 не показан), показатель идеальности n — по её наклону:

$$\varphi_{bm}(\bar{V}, \bar{I}) = \frac{kT}{q} \ln \left(\frac{AR^*T^2}{I_s(\bar{V}, \bar{I})} \right), \quad n(\bar{V}, \bar{I}) = \left[\frac{q}{kT} \frac{dV}{d \ln I} \right]_{V=\bar{V}, I=\bar{I}}. \quad (6)$$

(Заметим, что на увеличение показателя идеальности с ростом смещения указывает сравнение наклона касательных 2 и 3 в точках $(\bar{V}, \bar{I})$ и (V', I') на рис. 1: как видно, $n(I') > n(\bar{I})$ при $I' > \bar{I}$). Сказанное означает, что, по существу, мы анализируем не реальную ВАХ, а касательную к ней, или, иначе говоря, идеальную ВАХ, которая может быть представлена в виде

$$I = AR^*T^2 \exp \left(-\frac{q\varphi_{bm}(\bar{V})}{kT} \right) \exp \left(\frac{qV}{n(\bar{V})kT} \right). \quad (7)$$

Её параметры $\varphi_{bm}(\bar{V})$ и $n(\bar{V})$ и являются измеряемыми параметрами реальной ВАХ контакта металл—полупроводник с барьером Шоттки.

Для достаточно совершенных контактов, когда $n < 1,1$, зависимость от точки касания не вносит существенной неопределённости в указанные параметры. Но при низкой температуре, когда n может существенно зависеть от смещения (прежде всего в силу более широкого диапазона прикладываемых смещений), выбор точки касания может быть принципиальным с точки зрения сравнения с теорией. Поэтому, а также исходя из соображений удобства определения и необходимой строгости анализа результатов измерения, значения n и φ_{bm} следует определять при заранее оговоренных условиях, например при одном и том же токе (а значит, при различных смещениях) во всём температурном интервале

измерений. Условно это показано на рис. 2 для двух температур: высокой (комнатной T_k) и низкой (например, для температуры жидкого азота T_a). ВАХ 1 и 2 (условные) соответствуют указанным температурам, а V_k и V_a — смещения при этих температурах, соответствующие одному и тому же току $\bar{I}$, при котором проводятся измерения. Очевидно, что выбрать приемлемый для всех температур ток для проведения измерений не составляет труда (проблемы могут возникнуть только для низких барьеров Шоттки). В то же время выбор одинакового смещения в достаточно широком диапазоне температур практически невозможен из-за неприемлемо (неизмеримо) малых токов при низких температурах.

Следует подчеркнуть, что отсутствие определённости относительно того, измеряются ли параметры n и φ_{bm} в диапазоне температур при одинаковом токе, одинаковом напряжении или произвольных значениях тока и напряжения, может привести к совершенно различным выводам относительно характера теоретических зависимостей $n(T)$ и $\varphi_{bm}(T)$. В некоторых случаях при условии $V = \text{const}$ заметная зависимость параметров n и φ_{bm} от температуры отсутствует и появляется при использовании условия $I = \text{const}$, отвечающего реальным экспериментам. Впервые на это обратил внимание Левин [18], однако, возможно, из-за недостаточной обоснованности использованной им модели КМП подход Левина не привлёк должного внимания.

Сравнение идеальной аппроксимации (7) и реальной ВАХ (1) в точке касания ($\bar{V}$, $\bar{I}$) (см. рис. 1) позволяет установить связь между реальной высотой барьера $\varphi_b(V)$ (или $\varphi_b(I)$), соответствующей заданному смещению (току), и измеряемыми при этом смещении значениями n и φ_{bm} [21]. Учитывая произвольный выбор точки касания, получим

$$\varphi_{bV} = \varphi_{bm}(V) + \frac{n(V) - 1}{n(V)} \quad \text{или} \quad \varphi_{bI} = n(I)\varphi_{bm}(I) - [n(I) - 1] \frac{kT}{q} \ln \frac{AR^*T^2}{I}. \quad (8)$$

Новые обозначения φ_{bV} и φ_{bI} использованы, чтобы различать высоту барьера как функцию смещения (тока), $\varphi_b(V)$, и реальную высоту барьера, определяемую при заданном смещении (φ_{bV}) или заданном токе (φ_{bI}). По сути, это одна и та же величина. По соображениям, изложенным выше, далее мы будем использовать выражение для φ_{bI} , в котором легко учесть условие постоянства тока при измерениях n и φ_{bm} в диапазоне температур.

Введём в последнее выражение ток плоских зон $I_{bf} = AR^*T^2 \exp[-q\varphi_s/(kT)]$, соответствующий согласно (1) их смещению. Полученное в результате соотношение

$$\varphi_{bI} = n(I)\varphi_{bm}(I) - [n(I) - 1] \varphi_s - [n(I) - 1] \frac{kT}{q} \ln \frac{I_{bf}}{I} \quad (9)$$

оказывается весьма удобным для сравнения с известным выражением (2). Если $n = \text{const}$, что предполагалось при выводе (2), то при $I = I_{bf}$ (т. е. в условиях плоских зон) (9) переходит в (2). При этом в силу постоянства n измеряемая высота барьера φ_{bm} также не зависит от тока и оказывается равной высоте барьера φ_{b0} при нулевом смещении. Если же $n = n(I)$, то при $I = I_{bf}$ (9) переходит в точное выражение для ВБПЗ, учитывающее зависимость показателя идеальности ВАХ от смещения:

$$\varphi_{bf} = n(I_{bf})\varphi_{bm}(I_{bf}) - [n(I_{bf}) - 1] \varphi_s. \quad (10)$$

Заметим, что, строго говоря, ВАХ барьера Шоттки в условиях, близких к плоским зонам, не определена [2]. Это вносит некоторую неопределённость в понятие «ток плоских зон». Однако поскольку мы не уточняем здесь физическую природу процессов прохождения токов в БШ (см. комментарии к (1)), то для нас достаточно, чтобы при плоских зонах сохранилось выражение ВАХ в форме (1). Очевидно, что это можно достичь за счёт подгонки зависимости $\varphi_b(V)$. Тем не менее следует признать, что из-за трудностей реализации условий плоских зон для определения ВБПЗ желательно использовать другие условия. К этой проблеме мы вернёмся ниже.

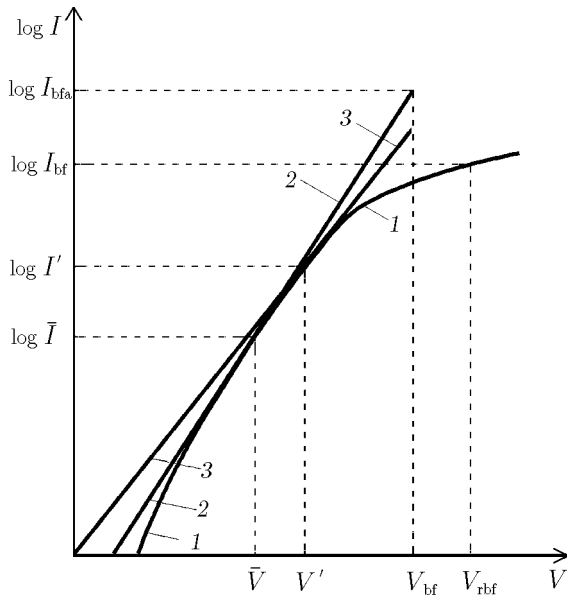


Рис. 1

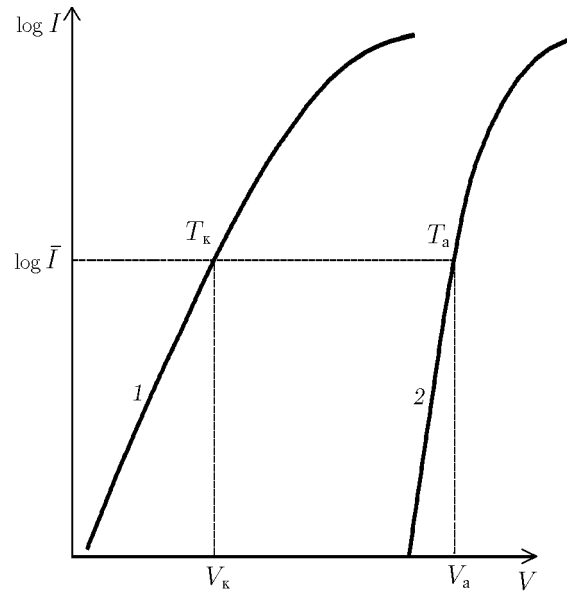


Рис. 2

Анализ выражения (9) для φ_{bI} показывает, что в достаточно широком диапазоне параметров и условий измерений ($I_{bf}/I < 10^2$; $N_C/N_D < 10^2$; $n(T_k) < 1,1$; $n(T_a) < 1,5$) вклад второго и третьего слагаемых в правой части (9) незначителен (около 3 %). При высоких температурах вклад указанных слагаемых уменьшается за счёт приближения $n(I)$ к единице, а при низких — за счёт непосредственного уменьшения температуры. В результате для реальной высоты барьера получим

$$\varphi_{bI} \approx \varphi_{bn} \equiv n(I)\varphi_{bm}(I). \quad (11)$$

Естественно, что ещё лучшим приближением для φ_{bI} является соотношение

$$\varphi_{bI} \approx n(I)\varphi_{bm}(I) - [n(I) - 1]\varphi_s. \quad (12)$$

Таким образом, соотношения (9) и (11) объясняют (см. Введение) близость величины $\varphi_{bn} \equiv n\varphi_{bm}$ к реальной высоте барьера и её слабую зависимость от температуры (поскольку реальная высота барьера не может существенно зависеть от температуры [1, 2]). Одновременно последнее означает, что для сохранения постоянства произведения φ_{bn} величины n и φ_{bm} в диапазоне температур должны изменяться разнонаправленно.

Объяснение изменений n и φ_{bm} с понижением температуры напрямую связано с ростом показателя идеальности с увеличением смещения. Поскольку для поддержания заданного тока при измерениях при более низкой температуре необходимо приложить большее смещение (см. рис. 2), это одновременно приводит к росту показателя идеальности ВАХ. При этом в соответствии с рис. 1 уменьшается измеряемая высота барьера, определяемая согласно (6) по точке пересечения касательных 2 и 3 с осью ординат (по току насыщения). Фактически это означает, что если показатель идеальности ВАХ некоторого контакта металл—полупроводник с барьером Шоттки растёт с увеличением смещения, то для него неизбежно проявление в той или иной степени «низкотемпературной аномалии»: роста n и уменьшения φ_{bm} при понижении температуры. Нетрудно убедиться, что эта ситуация характерна для самых различных моделей КМП с БШ, о которых говорилось выше: модели, учитывающей эффект сил изображения, модели Бардина [21], модели Левина [18], контактов со сферической и цилиндрической (с малым радиусом кривизны) поверхностью [8, 19] и других.

В конечном счёте сказанное выше означает, что слабая температурная зависимость величины $\varphi_{bn} \equiv n\varphi_{bm}$ при одновременном разнонаправленном изменении показателя идеальности и измеряемой высоты барьера с понижением температуры является следствием (или отражением) подобного поведения указанных величин в зависимости от смещения (тока). Действительно, численные оценки для конкретных моделей [8, 19, 21] указывают на относительно слабую зависимость величины φ_{bn} , выполняющей роль реальной высоты барьера, от смещения (тока), хотя параметры n и φ_{bm} могут варьироваться при изменении смещения довольно существенно. Это обстоятельство, в свою очередь, указывает на относительно слабое влияние выбора тока, при котором проводятся измерения параметров ВАХ, на результат измерений: в любом случае произведение $n\varphi_{bm}$ даст величину, близкую к реальной высоте барьера. Это заключение представляется важным для практических исследований. Нельзя, однако, упускать из вида, что оно справедливо при не слишком малых токах, когда выполняется приближение (11).

Приведённое объяснение поведения ВАХ при низких температурах неприменимо, однако, если зависимость высоты барьера от смещения линейна, как это предполагается в [13] (см. выше) и ряде последующих работ [14–16, 20, 22]. Действительно, поскольку при этом показатель идеальности ВАХ (4) не зависит от смещения, то исключается и его рост с понижением температуры, и одновременное уменьшение измеряемой высоты барьера. В принципе, возможной причиной «низкотемпературной аномалии» ВАХ может быть неоднородность КМП с БШ [22], никак не связанная с характером зависимости $\varphi_b(V)$. Однако в нашем случае, как подчёркивалось выше, речь идёт об однородных контактах. О возможной роли неоднородности КМП будет сказано ниже.

Очевидным недостатком реальной высоты барьера φ_{bI} (8) как характеристики КМП является зависимость (хотя и достаточно слабая) от тока. Безусловно, более желательно использовать для характеристики БШ высоту барьера, которая в принципе не зависит от смещения. Таковой согласно [1, 13–16] может быть высота барьера при плоских зонах φ_{bf} . Однако ВБПЗ в виде (2) основывается на некорректном предположении относительно зависимости $\varphi_b(V)$, а точным выражением (10) трудно воспользоваться из-за сложности реализации условий плоских зон.

Можно получить удобное и корректное выражение для φ_{bf} , если воспользоваться понятием эффективного (кажущегося) тока плоских зон I_{bfa} , который находится из (7) подстановкой $V = V_{bf}$. Смысл этого тока вполне понятен из рис. 1. Из рис. 1 следует также, что величина I_{bfa} (подобно n и φ_{bm}) является функцией тока, т. е. зависит от координат точки, в которой проводится касательная к реальной ВАХ. В результате преобразований с использованием равенства $\varphi_{bf} = V_{bf} + \varphi_s$ получим

$$I_{bfa}(I) = I_{bf} \exp \left\{ \frac{q}{nkT} [\varphi_{bf} - (n(I)\varphi_{bm}(I) - (n(I) - 1)\varphi_s)] \right\}, \quad (13)$$

откуда следует, что

$$\varphi_{bf} = n\varphi_{bm} - (n - 1)\varphi_s + \frac{nkT}{q} \ln \left(\frac{I_{bfa}}{I_{bf}} \right). \quad (14)$$

В выражении (14) опущена зависимость от тока, т. к. выбор конкретной точки на ВАХ не оказывает влияния на величину φ_{bf} . Очевидно, что выражение (14) при $I_{bfa} = I_{bf}$ (что возможно только при $n = \text{const}$) переходит в известное выражение (2).

Все входящие в правую часть (14) величины (а значит, и высота барьера при плоских зонах) могут быть определены непосредственно из реальной ВАХ (из эксперимента). Для повышения точности определения тока плоских зон I_{bf} постоянную Ричардсона можно найти из температурной зависимости $\ln(I_s/T^2) = f[1/(nT)]$ (см., например, [9]), поскольку при высоких температурах «низкотемпературная аномалия» практически не оказывает влияния на точность измерения R^* . Для оценки тока I_{bfa} необходимо учесть наличие последовательного сопротивления r_s в КМП и определить его величину из ВАХ при достаточно больших токах (например, в области пересечения ВАХ с линией тока

плоских зон $I = I_{bf}$, см. рис. 1), где сопротивление барьера несущественно. Очевидно, что смещение V_{rbf} , соответствующее указанной точке пересечения, определяет состояние плоских зон в КМП. Оно равно сумме падений напряжения на последовательном сопротивлении $r_s I_{bf}$ и в барьере V_{bf} , так что $V_{bf} = V_{rbf} - r_s I_{bf}$. Пересечение линии $V = V_{bf}$ с аппроксимацией ВАХ (7), которой соответствует кривая 2, и даст необходимое значение тока I_{bfa} .

Чтобы оценить погрешность, возникающую при определении ВБПЗ согласно (2) в сравнении с точными выражениями (10) или (14), необходимо учесть, что в силу условия $n = n(I)$ в действительности измерения проводят при некотором токе I , и вместо величины φ_{bf} (2) вычисляется некоторая величина

$$\varphi_{bfI} = n(I)\varphi_{bm}(I) - [n(I) - 1]\varphi_s, \quad (15)$$

практически равная истинной высоте барьера при указанном токе (см. (12)). Это значит, что погрешность $\Delta\varphi_{bf}$ при оценке φ_{bf} равна третьему слагаемому в правой части выражения (14): $\Delta\varphi_{bf} = (nkT/q) \ln(I_{bfa}/I_{bf})$. Более наглядную оценку $\Delta\varphi_{bf}$, как нам кажется, можно получить из сравнения выражений (10) и (15). Учитывая, что с ростом n измеряемая высота барьера уменьшается, а их произведение, как указывалось выше, может слабо зависеть от тока, для $\Delta\varphi_{bf}$ получим

$$\Delta\varphi_{bf} \approx [n(I_{bf}) - n(I)]\varphi_s. \quad (16)$$

Эта величина для низколегированных полупроводников и в КМП с заметно выраженной зависимостью $n(I)$ может быть существенной. Но для достаточно совершенных ДБШ, для высоколегированных полупроводников с малыми значениями φ_s , а также при низких температурах погрешность $\Delta\varphi_{bf}$, по видимому, мала. Этим, возможно, и объясняется довольно успешное использование высоты барьера φ_{bf} , определённой согласно (2), а практически согласно (15), в качестве фундаментальной характеристики КМП с БШ. Более того, во многих случаях из-за малости второго слагаемого в правой части (15) выражение для ВБПЗ принимает вид

$$\varphi_{bf} \approx n\varphi_{bm}. \quad (17)$$

Как видно, ВБПЗ в этом случае полностью соответствует величине φ_{bn} , играющей роль реальной высоты барьера в КМП с БШ (см. Введение). Совпадение (11) и (17) не должно смущать, т. к. это лишь свидетельствует, что при достаточной близости тока, при котором определяется φ_{bf} , к току плоских зон I_{bf} реальная высота барьера достаточно близка к ВБПЗ и обе эти величины (при выполнении указанных выше условий) близки к $\varphi_{bn} \equiv n\varphi_{bm}$. Это обстоятельство представляется важным, поскольку оно подчёркивает несостоятельность известной интерпретации экспериментальных результатов [20, 23], согласно которой реальная высота барьера φ_{b0} (в отсутствие смещения) уменьшается с понижением температуры, а ВБПЗ при этом слабо изменяется с температурой или не зависит от неё. Согласно изложенным выше представлениям реальная высота барьера φ_{bf} и ВБПЗ ведут себя (в т. ч. в диапазоне температур) одинаково, и их отличие (часто незначительное) связано только с тем, что они соответствуют разным токам: I и I_{bf} .

Полученные соотношения позволяют сделать ещё одно интересное заключение. Очевидно, что величина $\varphi_{bn} \equiv n\varphi_{bm}$ при определённых условиях может быть ближе к точному значению ВБПЗ (14), чем используемое в литературе выражение (2), а фактически (15). Это следует из сравнения (14) и (15). В первом из них в правой части присутствуют два относительно малых слагаемых с противоположными знаками, во втором — только одно, совпадающее с одним из слагаемых в правой части (14). Следовательно, можно сделать вывод, что (17) является лучшим приближением для (14), чем (15). Это значит, что использование (15) (т. е. (2)) для оценки ВБПЗ может привести к большей погрешности, чем выбор в качестве ВБПЗ величины φ_{bn} .

Экспериментальным критерием соответствия вычисленной величины φ_{bf} «истинной» высоте барьера при плоских зонах является, согласно практически всем цитированным работам, соответствие φ_{bf} величине φ_{bc} — высоте барьера, определённой из вольт-фарадной (или $C(V)$ -) характеристики и являющейся по своей сути ВБПЗ [1]. Согласно известному виду $C(V)$ -характеристики для ДБШ [1, 2] φ_{bc} определяется при стремлении $1/C^2 \propto \varphi_{bc} - \varphi_s - V_b - kT/q$ к нулю. Здесь V_b — часть внешнего смещения V , падающая на барьере. Точное определение φ_{bc} возможно, если $V_b \approx V$, т. е. в случае достаточно совершенных контактов с БШ. Однако в уже цитированных работах [13–16] и других речь идёт о КМП, которые не всегда можно отнести к совершенным. В этом случае соответствие (или, напротив, несоответствие) φ_{bf} величине φ_{bc} может быть недостаточным аргументом для вывода об «истинности» определённого значения φ_{bf} .

Как уже говорилось выше, явление «низкотемпературной аномалии» в ряде работ связывается с неоднородностью КМП с БШ [14, 22]. В то же время для анализа неоднородных контактов в этих работах используются те же представления, на основе которых в работе [13] получено выражение (2) для ВБПЗ однородного контакта. Попытки совместить оба подхода путём введения усреднённых характеристик неоднородного контакта [17, 22] хотя и интересны, но, на наш взгляд, противоречивы и требуют специального обсуждения.

Между тем последовательное рассмотрение однородных контактов позволяет сделать некоторые общие заключения и о T_0 -эффекте. Представим для этого выражение (14) для ВБПЗ в виде

$$n \left(1 - \frac{\ln(I_{bfa}/I_{bf})}{\ln(N_C/N_D)} \right) = 1 + \frac{n\varphi_{bm} - \varphi_{bf}}{\varphi_s}. \quad (18)$$

По крайней мере, для достаточно слаболегированных полупроводников, когда $\ln(N_C/N_D) \gg \gg \ln(I_{bfa}/I_{bf})$, выражение (18) преобразуется к известному виду

$$n \approx 1 + \frac{n\varphi_{bm} - \varphi_{bf}}{\varphi_s} = 1 + \frac{1}{T} \left[\frac{q(n\varphi_{bm} - \varphi_{bf})}{k \ln(N_C/N_D)} \right] = 1 + \frac{T_0}{T}. \quad (19)$$

Очевидно, что $T_0 \approx \text{const}$ только при слабой температурной зависимости величины $n\varphi_{bm}$, когда $n\varphi_{bm} - \varphi_{bf} \approx \text{const}$. Следовательно, не только рост показателя идеальности n и уменьшение высоты барьера φ_{bm} с понижением температуры и слабое изменение с температурой их произведения $\varphi_{bn} \equiv n\varphi_{bm}$, но и сам характер температурной зависимости $n(T)$ в виде $n \approx 1 + T_0/T$ (T_0 -эффект) связаны между собой и являются в конечном счёте следствием общего вида (1) ВАХ реальных КМП и определения параметров n и φ_{bm} вольт-амперной характеристики в виде (6) при некотором заданном токе в диапазоне температур. Очевидным представляется и другое: выражение (17) может быть хорошим приближением для φ_{bf} только в случае достаточно совершенных КМП, когда $T_0 \ll T$.

С другой стороны, учитывая различие физических закономерностей, допускающих представление ВАХ БШ в виде (1) и лежащих в основе зависимости $n(T)$, можно утверждать, что соотношение (19) с $T_0 \approx \text{const}$ не носит характер строгого закона и допускает существенные отклонения для конкретных механизмов, формирующих зависимость $\varphi_b(V)$. Это нетрудно показать, и об этом свидетельствует целый ряд работ, где условие $T_0 \approx \text{const}$ выполняется плохо или не выполняется совсем (см., например, [6, 9]).

Полученные результаты и выводы в основных чертах нашли подтверждение в численных расчётах для КМП с учётом сил изображения (оценки получены при выполнении настоящей работы), для КМП, подчиняющихся модели Бардина [21], и в однородных контактах с искривлённой поверхностью [19]. Было показано, что в широком диапазоне параметров КМП показатель идеальности ВАХ n растёт с понижением температуры. При этом измеряемая высота барьера φ_{bm} уменьшается, а произведение $\varphi_{bn} \equiv n\varphi_{bm}$ изменяется очень слабо и близко к истинной высоте барьера. В рамках данной работы

анализ различных моделей КМП с точки зрения более детального соответствия полученным выводам и соотношениям невозможен без подробной характеристики КМП. Предполагается, что такому анализу будет посвящена специальная публикация.

ЗАКЛЮЧЕНИЕ

Проведённый анализ даёт основания утверждать, что «низкотемпературная аномалия» ВАХ в однородных контактах с БШ (рост показателя идеальности ВАХ n (T_0 -эффект) и уменьшение высоты барьера φ_{bm} , измеряемой по току насыщения, с понижением температуры, а также слабая зависимость от температуры их произведения $n\varphi_{bm}$) является следствием нелинейной зависимости от смещения реальной высоты барьера $\varphi_b(V)$, причём эта зависимость такова, что показатель идеальности ВАХ растёт с увеличением смещения. Вторым условием проявления «низкотемпературной аномалии» является измерение показателя идеальности ВАХ n и высоты барьера φ_{bm} в диапазоне температур при одном и том же токе (что в экспериментах выполняется, как правило, автоматически).

Условия, необходимые для проявления «низкотемпературной аномалии», выполняются для всех наиболее известных моделей КМП с БШ: при учёте эффекта сил изображения, в моделях Бардина [21] и Левина [18], в однородных контактах с искривлённой границей раздела [8, 19]. Выполняются они и в реальных контактах с БШ, на что указывает наблюдаемый рост показателя идеальности ВАХ последних с увеличением смещения, хотя при низких температурах в силу ограниченной области смещений, в которой измеряется ВАХ, этот эффект не всегда заметно выражен. Сказанное означает, что «низкотемпературная аномалия» в КМП не связана с конкретным физическим механизмом, ответственным за наблюдаемый рост показателя идеальности n . Этим, как мы полагаем, и объясняется универсальный характер «низкотемпературной аномалии» в КМП с БШ.

В КМП, удовлетворяющих указанным выше требованиям и условиям измерений, высота барьера φ_{bf} , соответствующая плоским зонам в барьере, описывается полученным в настоящей работе выражением (14) и может быть определена непосредственно из экспериментальной ВАХ. Это выражение отличается от используемого в литературе [13] выражения (2), соответствующего контактам с независящим от смещения показателем идеальности ВАХ, и свидетельствует о систематической погрешности, допускаемой при определении φ_{bf} известным способом.

С методической и физической точки зрения важно различать используемые при анализе ВАХ высоты барьера. В каждой точке (в силу наличия зависимостей $\varphi_b(V)$ и $n(V)$) ВАХ определяется измеряемой высотой барьера φ_{bm} , которая находится согласно (6), и реальной высотой барьера φ_{bI} (или φ_{bV}), определяемой согласно (8). Учёт этого обстоятельства позволяет естественным образом разрешить противоречие между кажущимся уменьшением высоты барьера при понижении температуры (при этом уменьшается именно φ_{bm}) и отсутствием уменьшения реальной высоты барьера φ_{bI} в действительности.

Что касается объяснения «низкотемпературной аномалии» ВАХ неоднородностью контакта, то не вдаваясь в детали, отметим, что такой подход не встречает принципиальных препятствий. Формально он предполагает лишь доказательство того, что ВАХ неоднородного контакта имеет вид (1), а функции $\varphi_b(V)$ и $n(V)$ и условия измерений параметров ВАХ соответствуют сформулированным выше требованиям. Очевидно, что если однородные участки неоднородного контакта проявляют, как было показано выше, «низкотемпературную аномалию», то и весь контакт в целом будет неизбежно проявлять подобные свойства. Более того, в неоднородных контактах появляются дополнительные возможности для «усиления» зависимостей $\varphi_b(V)$ и $n(V)$ [22].

Наконец, необходимо сделать ещё одно важное замечание. Показанная теоретически неизбежность «низкотемпературной аномалии» в КМП с БШ (рост показателя идеальности ВАХ и, следовательно, рост высокочастотных и, как правило, низкочастотных шумов с понижением температуры)

не должна приводить к заключению о бесплодности усилий по совершенствованию ДБШ для работы в области низких температур (в охлаждаемых приёмных устройствах). Достаточно сказать, что эффект сил изображения в относительно слаболегированных полупроводниках проявляется в заметном увеличении показателя идеальности только при $T = 20$ К, тогда как высокая плотность поверхностных состояний при наличии тонкого промежуточного слоя в КМП или существование локальных областей термополевой эмиссии в контакте проявляются значительно сильнее уже при $T \leq 150$ К [19, 21].

Работа выполнена при поддержке Межотраслевой научно-технической программы «Физика микроволн» (проект 3.12).

СПИСОК ЛИТЕРАТУРЫ

1. Rhoderick E. H., Williams R. H. Metal-Semiconductor Contacts: 2 ed. Oxford: Clarendon, 1988.
2. Зи С. М. Физика полупроводниковых приборов: Пер. с англ. под ред. Р. А. Сурица. М.: Мир, 1984.
3. Padovani F. A., Sumner G. G. // J. Appl. Phys. 1965. V. 38, No. 12. P. 3744.
4. Hackam R., Harrop P. // IEEE Trans. ED. 1972. V. 19, No. 12. P. 1231.
5. Гольдберг Ю. А., Поссе Е. А., Царенков Б. В. // ФТП. 1971. Т. 5. С. 468.
6. Божков В. Г., Ковтуненко Г. Ф., Суроткина Г. М. и др. // Электронная техника. Сер. 2. Полупроводниковые приборы. 1978. Вып. 4 (122). С. 14.
7. Божков В. Г., Ковтуненко Г. Ф., Аврамчикова Т. А. // Электронная техника. Сер. 2. Полупроводниковые приборы. 1979. Вып. 9 (135). С. 35.
8. Божков В. Г., Куркан К. И. // Полупроводниковые приборы с барьером Шоттки. Киев: Наукова думка, 1979. С. 43.
9. Saxena A. N. // Surface Science. 1969. V. 13. P. 151.
10. Jager D., Kassing R. // J. Appl. Phys. 1977. V. 48. P. 4414.
11. Коротченков Г. С., Молодян И. П. // Микроэлектроника. 1978. Т. 7, вып. 2. С. 168.
12. Божков В. Г., Малаховский О. Ю., Леуский В. Е. и др. // Радиотехника и электроника. 1983. Т. 28, № 6. С. 1182.
13. Wagner L. F., Joung R. W., Sugerman A. // IEEE Electron. Dev. Lett. 1983. V. 4, No. 9. P. 320.
14. Werner J. H., Guttler H. H. // J. Appl. Phys. 1993. V. 73. P. 1315.
15. Broom R. F., Meier H. P., Walter W. // J. Appl. Phys. 1986. V. 60, No. 5. P. 1832.
16. Chin V. W. L., Green M. A., Storey J. W. V. // J. Appl. Phys. 1990. V. 68, No. 7. P. 3470.
17. Werner J. H., Guttler H. H. // Physica Scripta. T. 1991. V. 39. P. 258.
18. Levine J. D. // J. Appl. Phys. 1971. V. 42. P. 3991.
19. Божков В. Г., Кашкан А. А. // Электронная техника. Сер. 2. Полупроводниковые приборы. 1981. Вып. 7 (150). С. 12.
20. Unewisse M. H., Storey J. W. V. // J. Appl. Phys. 1993. V. 73. P. 3873.
21. Божков В. Г. // Изв. вузов. Физика. 1987. Т. 30, № 2. С. 29.
22. Werner J. H., Guttler H. H. // J. Appl. Phys. 1991. V. 69, No. 4. P. 1522.
23. Chand S., Kumar J. // Semicond. Sci. Technol. 1995. V. 10. P. 1680.

Федеральное государственное
унитарное предприятие «НИИПП»,
г. Томск, Россия

Поступила в редакцию
28 августа 2001 г.

ON THE NATURE OF THE "LOW-TEMPERATURE ANOMALY" IN METAL-SEMICONDUCTOR SCHOTTKY-BARRIER CONTACTS

В. Г. Божков

425

V. G. Bozhkov

The "low-temperature anomaly" of current–voltage characteristics of homogeneous Schottky-barrier contacts consists in an increase in the ideality factor n and a decrease in the barrier height φ_{bm} , determined using the saturation current, with decreasing temperature, as well as the proximity of their product $\varphi_{\text{bn}} \equiv n\varphi_{\text{bm}}$ to the actual barrier height $\varphi_{\text{b}}(V)$. We show that this well-known anomaly occurs due to the effect of two factors: (i) the nonlinear voltage dependence of $\varphi_{\text{b}}(V)$ resulting in a growth of the ideality factor with increasing bias voltage (current) and (ii) the fact that n and φ_{bm} are measured for the same current in some interval of the temperature. A new, much more accurate expression for the flat-band barrier height φ_{bf} is derived.

УДК 537.8+621.38

ИССЛЕДОВАНИЕ ЗАРЯДОВЫХ И ЧАСТОТНЫХ СВОЙСТВ ТВЕРДОТЕЛЬНЫХ ГЕТЕРОГЕННЫХ СТРУКТУР С ЗАРЯДОВОЙ СВЯЗЬЮ

В. П. Заярный

Выявлены зарядовые характеристики гетерогенных твердотельных структур (усреднённая плотность поверхностных состояний и эффективный частотный фактор), определяющие их частотные свойства в приборах с зарядовой связью. Получены характеристики, определяющие особенности взаимодействия носителей заряда в слоях гетероструктур с электрическими полями радиочастотного диапазона.

Множество существующих разновидностей гетерогенных твердотельных структур для новейших микроэлектронных приборов по-прежнему требуют решения проблем, обусловленных наличием примесей вблизи границ раздела слоёв указанных структур, наличием в их запрещённых зонах разрешённых состояний и моноэнергетических уровней, определяющих особенности происходящих в слоях зарядовых процессов при взаимодействии с внешними (в нашем случае электрическими) полями. Важность изучения этих проблем подчёркивается в работах [1, 2].

Приборы с зарядовой связью (ПЗС) на основе гетерогенных структур (в нашем случае типа металл (Al)—поликремний (Si*)—диэлектрик (SiO₂)—полупроводник (Si)) представляют собой их периодическую последовательность, на которую подаются сдвинутые по фазе импульсы переноса зарядовых пакетов [3]. Основной характеристикой работы таких приборов является неэффективность переноса зарядовых пакетов [4], а цель данного исследования — выявление влияния зарядовых характеристик гетероструктур, определяющих их высокочастотные свойства, на неэффективность переноса заряда приборами с зарядовой связью.

В [3] утверждается, что высокочастотные свойства ПЗС определяются следующими физическими механизмами. Для относительно малых зарядовых пакетов доминирующим механизмом переноса заряда является термодиффузия. При этом полный заряд уменьшается экспоненциально с постоянной времени $\tau_{\text{диф}} = 4L^2/(\pi^2 D_n)$, где L — длина электрода переноса, D_n — коэффициент диффузии неосновных носителей заряда.

Перемещение относительно больших зарядовых пакетов происходит за счёт самоиндуцированного дрейфа в продольном электрическом поле, пропорциональном градиенту поверхностной плотности сигнального заряда. Однако даже в отсутствие сигнального заряда под электродом переноса на границе с диэлектриком существует продольное (краевое) электрическое поле, способствующее ускорению перемещения заряда на последней стадии его переноса, когда самоиндуцированное поле уже меньше краевого.

Таким образом, улучшение частотных свойств ПЗС можно достичь, уменьшая длину L электрода переноса или используя полупроводники с наибольшей подвижностью заряда (наилучший вариант — арсенид галлия). В последнем случае ПЗС могут работать на частотах до 500 МГц при неэффективности переноса заряда $\varepsilon_{ss} \approx 10^{-4}$. В [4] показано, что усилением краевого электрического поля можно повысить рабочую частоту ПЗС до 1 ГГц, однако проблема дальнейшего повышения быстродействия ПЗС в литературе не исследуется. Если исходить из предположения, что на сверхвысоких частотах структуры ПЗС обладают свойствами микрополосковых (компланарных [5]) линий, то, используя их электродинамические свойства [6], можно существенно усилить влияние краевых полей на перемещение зарядовых пакетов вдоль структуры и, следовательно, увеличить быстродействие ПЗС.

В. П. Заярный

427

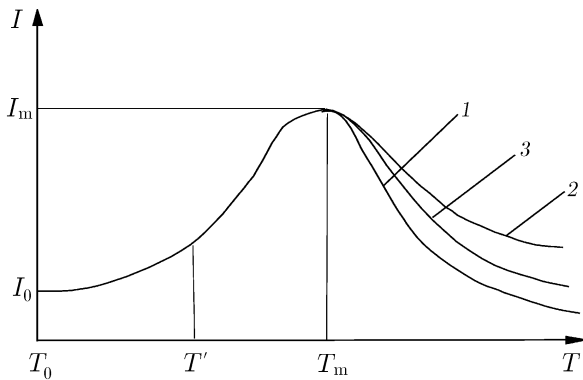


Рис. 1. Релаксация термостимулированных токов при различном порядке зарядовых процессов: кривая 1 соответствует случаю $m = 1$; 2 — $m = 2$; 3 — промежуточному случаю

Однако это возможно в случае минимизации захвата носителей заряда на разрешённых состояниях и моноэнергетических уровнях в запрещённой зоне полупроводника и повышения скорости протекания зарядовых процессов в гетероструктурах.

Кинетика зарядовых процессов в приповерхностной области полупроводника (Si) рассматриваемых гетероструктур, являющейся зоной перемещения сигнальных зарядовых пакетов в ПЗС, в значительной степени будет определяться наличием разрешённых состояний и моноэнергетических уровней в запрещённой зоне полупроводника. В [7] показано, что при исследовании твердотельных материалов методами термостимулированных токов поляризации/деполяризации, когда зависимость тока от температуры имеет вид пиков (см. рис. 1), последние можно описать одним из следующих выражений, несущих информацию о порядке зарядовых процессов:

$$I(T) = -q\omega_3 r^2 / (2L) (1 - r/L) n_{t0} \exp \left[-E_t / (kT) - \int_{T_0}^T (\omega_t / \beta) \exp[-E_t / (kT')] dT' \right] \quad (1)$$

в случае релаксации первого порядка ($m = 1$, см. рис. 1),

$$I(T) = -I_0 [\tau_t / \tau_f(T_0)] \exp[-E_t / (kT)] \left[1 + \int_{T_0}^T \omega_t \tau_t / [\beta \tau_f(T_0)] \exp[-E_t / (kT')] dT' \right]^{-2} \quad (2)$$

в случае релаксации второго порядка ($m = 2$),

$$I(T) \approx -q^2 \mu N_c M_t^2 r^2 / (2\epsilon \epsilon_0 L N_{ss}) (1 - r/L) \exp[-E_t / (kT)] A'(T) / [1 - A'(T)]^2, \quad (3)$$

где

$$A'(T) = n_{t0} / (n_{t0} + M_t) \exp \left[- \int_{T_0}^T q \mu N_c M_t / (\epsilon \epsilon_0 \beta N_{ss}) \exp[-E_t / (kT')] dT' \right]$$

в промежуточном случае. Здесь q — элементарный заряд, k — постоянная Больцмана, N_{ss} — плотность поверхностных состояний в запрещённой зоне полупроводника, r и L — толщины области локализации объёмного заряда и образца соответственно, n_{t0} — начальная концентрация носителей заряда, локализованного в ловушках, β — скорость нагрева образца, E_t — энергия активации моноэнергетического уровня, I_0 — начальная величина термостимулированного тока, τ_f и τ_t — время пролёта и захвата свободных носителей заряда в ловушки, T и T_0 — текущая и начальная температуры образца, ϵ и ϵ_0 — диэлектрическая проницаемость и электрическая постоянная полупроводника, N_c — эффективная плотность состояний в разрешённой зоне, M_t — концентрация глубоких ловушек, не опустошающихся при нагревании, ω_t и ω_3 — частотный фактор опустошения ловушек и эффективный частотный фактор соответственно.

В случае, когда зарядовые процессы в гетероструктурах описываются выражением (1), имеет место слабый перезхват носителей заряда, для которого справедливо соотношение $\omega_3 = \omega_t$. Если зарядовые процессы в гетероструктурах описываются выражением (2), то имеет место сильный перезхват носителей заряда, для которого $\omega_3 = \omega_t \tau_t / \tau_f$, где τ_f — время рекомбинации носителей заряда.

Используя температурную зависимость времени релаксации τ носителей заряда [7], можно определить энергию активации моноэнергетического уровня E_t и эффективный частотный фактор ω_3 . Эта зависимость описывается выражением

$$\tau(T) = \omega_3^{-1} \exp[E_t/(kT)],$$

где $\tau(T)$ определяется для нескольких точек пика термостимулированного тока $\tau(T')$ (см. рис. 1) по формуле

$$\tau(T') = \frac{1}{I(T')} \int_{T'}^{\infty} \frac{I(T)}{\beta} dT.$$

Зависимость времени релаксации τ носителей заряда от абсолютной температуры T в координатах $(1/(kT), \ln \tau)$ представляет собой прямую (см. рис. 2), по которой определяется энергия активации моноэнергетического уровня (по углу α наклона прямой: $E_t \propto \tan \alpha$) и эффективный частотный фактор ω_3 (по координате точки пересечения указанной прямой с осью ординат).

Порядок m зарядовых процессов определим путём сопоставления спадающей части измеренного пика термостимулированного тока с моделями (1)–(3). Для их адекватного использования при исследовании свойств гетероструктур необходимо также иметь достоверную информацию о плотности поверхностных состояний в запрещённой зоне полупроводника исследовавшихся гетероструктур, которая в нашем случае определялась методами, описанными в [5, 6]. Эксперименты проводились с применением термоочистки пиков термостимулированного тока от влияния фонового заряда.

Экспериментально установленная зависимость неэффективности переноса зарядовых пакетов от эффективного частотного фактора $\varepsilon_{ss}(\omega_3)$ при разном порядке m зарядовых процессов и постоянной частоте тактовых импульсов $f_T = 10$ МГц приведена на рис. 3. Видно, что с увеличением ω_3 неэффективность переноса сигнального заряда уменьшается. Из графиков также видно, что с увеличением m наклон прямой резко уменьшается, что говорит о существенном влиянии порядка процесса на неэффективность переноса заряда. При наличии слабого перезахвата носителей заряда (когда $m \approx 1$) прямая $\varepsilon_{ss}(\omega_3)$ имеет более крутой наклон. Следовательно, для уменьшения неэффективности переноса зарядовых пакетов приборами с зарядовой связью необходимо по возможности увеличивать эффективный частотный фактор ω_3 и уменьшать порядок m зарядовых процессов на имеющихся центрах захвата носителей заряда.

При наличии сильного перезахвата носителей заряда в гетероструктуре (в данном случае при $m \approx 1,5$) зависимость $\varepsilon_{ss}(\omega_3)$ проявляется значительно слабее. Увеличение перезахвата ведёт к росту неэффективности переноса заряда, и при больших значениях m (особенно когда порядок процесса приближается к другому предельному значению $m = 2$) неэффективность переноса может быть столь значительна, что ПЗС будет неработоспособным даже в режиме обработки логических сигналов.

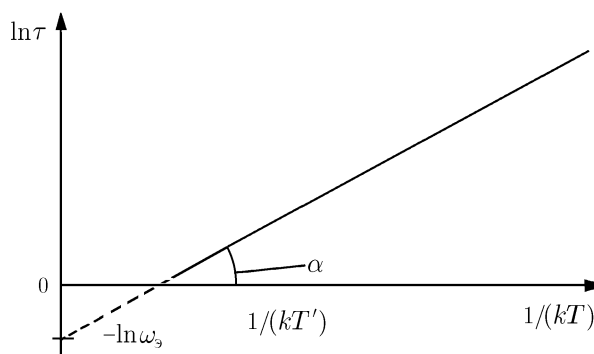


Рис. 2. Прямая для определения характеристик центров захвата носителей заряда на моноэнергетических уровнях

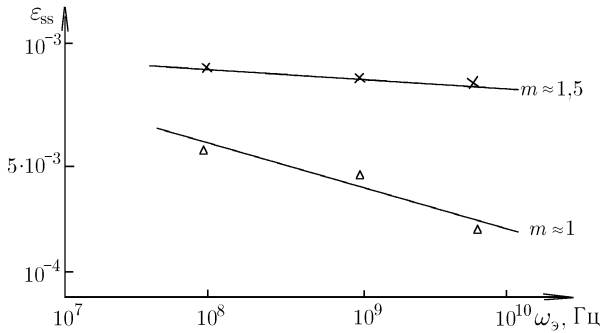


Рис. 3. Зависимость неэффективности переноса заряда приборами с зарядовой связью от эффективного частотного фактора при различном порядке зарядовых процессов

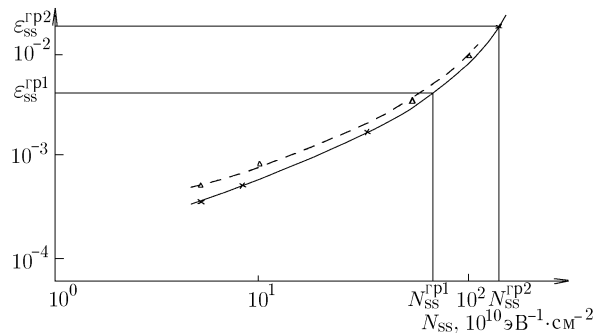


Рис. 4. Экспериментальная (сплошная кривая) и теоретическая (штриховая кривая) зависимости неэффективности переноса зарядовых пакетов ПЗС от усреднённой плотности поверхностных состояний

На рис. 4 приведена зависимость неэффективности переноса зарядовых пакетов ε_{ss} исследовавшихся образцов линейных ПЗС (определялась по методике, приведённой в [3]) от усреднённой плотности N_{ss} поверхностных состояний в запрещённой зоне полупроводника вблизи границы раздела диэлектрик—полупроводник. Для исследованных ПЗС наименьшие и наибольшие неэффективности переноса зарядовых пакетов достигались соответственно при наименьших и наибольших усреднённых значениях N_{ss} . Неэффективность переноса определялась при постоянной частоте тактовых импульсов $f_T = 10^5$ Гц и числе переносов зарядовых пакетов в линейном ПЗС $n_p = 16$. При этом было установлено, что в режиме формирования сигналов изображения ПЗС имеют удовлетворительную работоспособность до граничного значения неэффективности переноса зарядовых пакетов $\varepsilon_{ss}^{rp1} = 5 \cdot 10^{-3}$. На приведённом графике значению ε_{ss}^{rp1} соответствует усреднённая плотность поверхностных состояний гетероструктур, на основе которых выполнены исследовавшиеся ПЗС, $N_{ss}^{rp1} = 6,5 \cdot 10^{11}$ эВ $^{-1}$ ·см $^{-2}$. В случае логической обработки сигналов ПЗС сохраняют работоспособность вплоть до $\varepsilon_{ss}^{rp2} = 2 \cdot 10^{-2}$, чему соответствует значение $N_{ss}^{rp2} = 1,1 \cdot 10^{12}$ эВ $^{-1}$ ·см $^{-2}$.

Соответствующая теоретическая зависимость $\varepsilon_{ss}(N_{ss})$ имеет вид [3]

$$\varepsilon_{ss} = qkTN_{ss}/(C_d V_g) \ln(p + 1), \quad (4)$$

где $C_d V_g$ — поверхностная плотность сигнального заряда, p — число фаз у используемого ПЗС. В нашем случае при ёмкости диэлектрика гетероструктуры $C_d \approx 110$ пФ, приложенном к гетероструктуре напряжении $V_g = 10$ В и $p = 3$ теоретическая зависимость показана на рис. 4 штриховой линией. Имеющиеся расхождения с экспериментально полученной зависимостью, по всей видимости, объясняются методической погрешностью при использовании соотношения (4).

Неэффективность переноса зарядовых пакетов, обусловленная наличием поверхностных состояний в полупроводнике вблизи границы раздела диэлектрик—полупроводник, с увеличением тактовой частоты имеет тенденцию к уменьшению. На рис. 5 приведены экспериментально полученные зависимости $\varepsilon_{ss}(f_T)$ для разной плотности поверхностных состояний. Уменьшение неэффективности переноса зарядовых пакетов объясняется тем, что с ростом частоты f_T снижается количество активных ловушек относительно их общего числа. Характер приведённых зависимостей свидетельствует о том, что с увеличением частоты тактовых импульсов величина ε_{ss} уменьшается тем быстрее, чем больше плотность поверхностных состояний N_{ss} , и совсем слабо проявляется при $f_T > 10^7$ Гц.

На частотах $f_T > 10^7$ Гц сказываются только зарядовые процессы, обусловленные наличием мо-
ноэнергетических уровней в приповерхностной области полупроводника, для которых эффективный

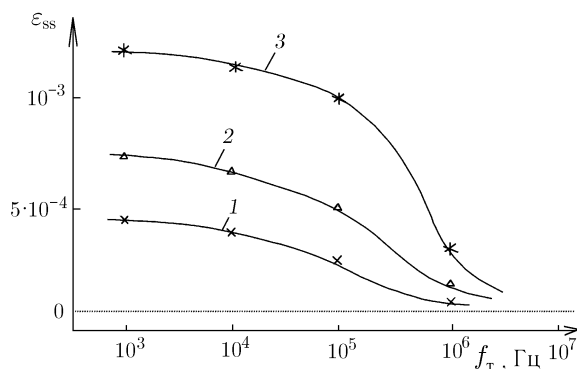


Рис. 5. Зависимость неэффективности переноса заряда в ПЗС, обусловленной влиянием поверхностных состояний гетероструктур вблизи границы раздела диэлектрик—полупроводник, от тактовой частоты: кривая 1 соответствует $N_{ss} = 5 \cdot 10^{10} \text{ см}^{-2} \cdot \text{эВ}^{-1}$, кривая 2 — $N_{ss} = 8 \cdot 10^{10} \text{ см}^{-2} \cdot \text{эВ}^{-1}$, кривая 3 — $N_{ss} = 3 \cdot 10^{11} \text{ см}^{-2} \cdot \text{эВ}^{-1}$

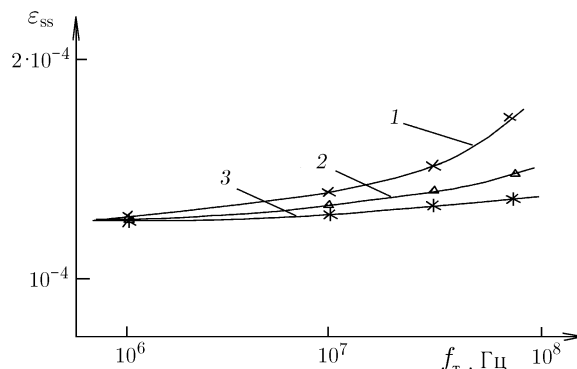


Рис. 6. Зависимость неэффективности переноса заряда в ПЗС, обусловленной влиянием моноэнергетических уровней вблизи границы раздела диэлектрик—полупроводник, от тактовой частоты: кривая 1 соответствует $\omega_э = 10^8 \text{ с}^{-1}$, кривая 2 — $\omega_э = 10^9 \text{ с}^{-1}$, кривая 3 — $\omega_э = 10^{10} \text{ с}^{-1}$

частотный фактор также может превышать 10^7 Гц (согласно [7] $\omega_э$ может достигать 10^{12} Гц). Экспериментально полученные зависимости $\varepsilon_{ss}(f_T)$ для различных значений эффективного частотного фактора $\omega_э$ приведены на рис. 6 (порядок m процесса во всех случаях был близок к единице). Из рис. 6 видно, что при увеличении частоты f_T в диапазоне СВЧ влияние моноэнергетических уровней, для которых $\omega_э > 10^7$ Гц, приводит к некоторому увеличению ε_{ss} , что ухудшает качество работы ПЗС. При меньших значениях $\omega_э$ увеличение неэффективности переноса заряда происходит на более низких частотах, и наоборот. Это объясняется тем, что за время переноса сигнального зарядового пакета часть заряда не успевает высвободиться из ловушек. Чем больше тактовая частота, тем меньше вероятность эмиссии захваченного заряда из ловушек, что приводит к увеличению ε_{ss} .

Характер данных зависимостей хорошо согласуется с эмпирическими выражениями, полученными из следующих рассуждений. Пусть неэффективность переноса зарядовых пакетов определяется отношением захваченного в ловушках остаточного заряда к полному заряду в пакете. Пик термостимулированного тока деполяризации $I(T)$ для данного моноэнергетического уровня при $m = 1$ описывается выражением (1). Тогда с учётом числа переносов зарядового пакета и тактовой частоты определим неэффективность переноса заряда следующим образом:

$$\varepsilon_{ss} \approx n_{\pi} \int_{T_0}^{T_k} \frac{kT_m I(T) dT}{\beta E_t C_d V_g \exp[1 - \ln(f_T/\omega_э)]},$$

где T_m и T_k — температуры, соответствующие максимуму и окончанию пика термостимулированного тока деполяризации, E_t — энергия активации ловушек моноэнергетического уровня.

Приведённые выше характеристики определялись для ПЗС на основе гетероструктур, имевших наибольший эффективный частотный фактор $\omega_э = 10^{10} \text{ с}^{-1}$, что не является его предельным значением. Поэтому при увеличении $\omega_э$ следует ожидать расширения частотного диапазона ПЗС. При увеличении тактовой частоты до значений, превышающих $\omega_э$, следует ожидать снижения активности ловушек на данном моноэнергетическом уровне.

Из установленных частотных зависимостей характеристик, определяющих зарядовое состояние исследовавшихся гетероструктур и ПЗС на их основе, видно, что влияние поверхностных состояний

сказывается при тактовой частоте меньше нескольких мегагерц. На более высоких частотах их влияние пренебрежимо мало. В диапазоне СВЧ на частотные свойства ПЗС влияет только наличие моноэнергетических уровней. Ухудшение качества работы ПЗС вызывают моноэнергетические уровни с меньшим эффективным частотным фактором ω_3 . Наоборот, наличие в полупроводнике гетероструктуры с большим значением ω_3 может практически не сказываться на качестве работы ПЗС. Следует отметить, что с увеличением тактовой частоты качество работы ПЗС оказывается тем выше, чем меньше порядок процесса, т. е. чем слабее перезахват носителей заряда.

Из приведённых результатов также следует, что на высоких и сверхвысоких частотах могут оказаться работоспособными ПЗС, которые неработоспособны на более низких частотах, что особенно важно при создании ПЗС для работы в составе объёмных интегральных схем СВЧ/КВЧ диапазона [5].

СПИСОК ЛИТЕРАТУРЫ

1. Алфёров Ж. И. // Физика сегодня и завтра. Прогнозы науки. М.: Наука, 1975.
2. Алфёров Ж. И., Конников С. Г., Корольков В. И. // Физика и техника полупроводников. 1973. Т. 7. С. 1 423.
3. Приборы с зарядовой связью: Пер. с англ. / Под ред. Д. Ф. Барба. М.: Мир, 1982. 240 с.
4. Зи С. Физика полупроводниковых приборов: В 2-х книгах. М.: Мир, 1984. 456 с.
5. Гвоздёв В. И., Нефёдов Е. И. Объёмные интегральные схемы СВЧ. М.: Наука. 1985. 256 с.
6. Неганов В. А., Раевский С. Б., Яровой Г. П. Линейная макроскопическая электродинамика. Т. 1. / Под ред. В. А. Неганова. М.: Радио и связь, 2000. 509 с.
7. Гороховатский Ю. А., Бордовский Г. А. Термоактивационная токовая спектроскопия высокоомных полупроводников и диэлектриков. М.: Наука, 1991. 248 с.

Волгоградский госуниверситет,
г. Волгоград, Россия

Поступила в редакцию
18 июля 2001 г.

ANALYSIS OF CHARGE AND FREQUENCY CHARACTERISTICS OF CHARGE-COUPLED SOLID-STATE HETEROSTRUCTURES

V. P. Zayarniy

We reveal the charge characteristics, such as the average density of surface states and the effective frequency factor, which determine the frequency properties of heterostructures in charge-coupled devices. The parameters determining the features of the interaction between charge carriers in heterogenic-structure layers with RF electric fields are found.

УДК 621.391.828

МЕТОДЫ ОСЛАБЛЕНИЯ ЭФФЕКТОВ ИНТЕРМОДУЛЯЦИИ В РАСПРЕДЕЛЕНИИ ВИГНЕРА—ВИЛЯ

Е. И. Шкелёв, А. Г. Кисляков, С. Ю. Лунов

Рассмотрены характеристики типичных нелинейных искажений в динамических спектрах сигналов, полученных по методу Вигнера—Виля. Обсуждаются существующие способы ослабления этих искажений. Предложен ряд методов «чистки» распределений Вигнера—Виля от продуктов нелинейного взаимодействия исследуемой компоненты сигнала с его более мощными гармоническими или импульсными составляющими. Методы основаны главным образом на предварительной линейной обработке (например, с помощью быстрого преобразования Фурье) и/или на частотной фильтрации сигнала. Создан пакет программ для расчёта распределений Вигнера—Виля на ПЭВМ, с помощью которого предложенные алгоритмы успешно апробированы на тестовых и реальных сигналах.

ВВЕДЕНИЕ

Спектральный анализ с высоким качеством невозможно выполнить без использования априорной информации, касающейся свойств исследуемого сигнала, или без постулирования его определённой модели. Так, например, априорно задаваемая модель сигнала используется в параметрических методах анализа. В частности, для метода максимума энтропии принимается авторегрессионная модель [1, 2]. В методе Писаренко использована модель с добавлением белого шума к «чистым» частотам. Конкретизация свойств сигнала, как и использование априорной информации о нём, позволяет добиться эффекта сверхразрешения. В противном случае было бы необходимо анализировать сигнал на промежутке времени от $t = -\infty$ до $t = +\infty$.

Анализ сигналов с изменяющимся спектром реализуется различными спектрально-временными методами и тесно связан с понятием «мгновенный спектр». Определение мгновенного спектра зависит от используемого метода анализа и имеет свои особенности в каждом конкретном случае, что отчётливо проявилось во множестве определений этого понятия [1]. Однако все разновидности спектрально-временного анализа совпадают или непосредственно связаны с функцией неопределённости Вудворда [3–5]

$$\chi(\theta, \tau) = \int_{-\infty}^{+\infty} s(t - \tau/2) s^*(t + \tau/2) \exp(j\theta t) dt, \quad (1)$$

где звёздочка означает комплексное сопряжение. Указанную функцию можно интерпретировать как отклик фильтра, согласованного с аналитическим сигналом $s(t)$, у которого одновременно измеряются задержка во времени и смещение по частоте. Поэтому с функцией неопределённости имеют дело всякий раз, когда желают оценить временной τ и частотный $\delta\omega$ сдвиги двух сигналов. Связанные с этим вопросы достаточно широко обсуждались [1] по той причине, что функция неопределённости позволяет судить о предельной достижимой точности одновременного измерения τ и $\delta\omega$ в условиях априорной неопределённости анализируемых сигналов. Однако в силу нелинейности, заложенной в алгоритм вычисления, функция неопределённости имеет «боковые» выбросы, которые могут породить неоднозначность в интерпретации результатов измерения. Связано это с перекрёстными искажениями, возникающими при взаимодействии спектральных и временных компонент сигнала. В активной радиолокации,

применительно к которой и была введена функция неопределённости, роль перекрёстных искажений может быть уменьшена, поскольку существует возможность подбора используемого для локации сигнала под требуемую функцию неопределённости.

Для целей спектрально-временного анализа более удобно использовать распределение Вигнера—Виля (ВВР) [6–8]

$$P(f, t) = \int_{-\infty}^{+\infty} s(t - \tau/2) s^*(t + \tau/2) \exp(j2\pi f\tau) d\tau, \quad (2)$$

которое даёт не относительное (как в случае с функцией неопределённости), а абсолютное распределение мощности сигнала по частоте и времени. Однако наличие перекрёстных искажений, как и в случае с функцией неопределённости, усложняет процедуру спектрально-временного анализа: последняя должна включать средства, которые позволяли бы отделять спектрально-временные компоненты сигнала от продукта их нелинейного взаимодействия [9]. Достаточно уверенное выделение компонент сигнала возможно только для относительно простых сигналов (например, для сигналов с гауссовой формой огибающей [9]).

В настоящей работе предполагается рассмотреть как характеристики типичных перекрёстных искажений, так и некоторые методы их фильтрации. Предложенные методы иллюстрируются практическими примерами на основе результатов спектрально-временного анализа пульсаций солнечного радиоизлучения.

1. ИНТЕРМОДУЛЯЦИОННЫЕ ИСКАЖЕНИЯ СИГНАЛОВ

Спектрально-временные компоненты сигнала и продуктов перекрёстных искажений имеют свои отличительные особенности. В частности, из (2) следует, что временные вариации соответствующих мгновенному спектру компонент распределения Вигнера—Виля обусловлены только свойствами сигнала $s(t)$, в то время как интермодуляционные компоненты, возникающие из-за взаимодействия между составляющими мгновенного спектра, имеют осциллирующие фазовые множители. Поясним это на примере сигнала с большой базой, представленного суперпозицией некоторого числа мод, мгновенные частоты $f_i(t)$ которых непрерывно изменяются во времени (здесь i — индекс моды). Мгновенный спектр такого сигнала в каждый момент времени локализован в относительно узких спектральных областях вблизи $f_i(t)$. Области локализации энергии интермодуляционных компонент на плоскости частота—время лежат в окрестности частот $f_{ij}(t)$, определяемых совокупностью полусумм мгновенных частот $f_i(t)$ и $f_j(t)$: $f_{ij}(t) = [f_i(t) + f_j(t)]/2$.

Отличия между компонентами активного спектра и продуктами перекрёстных искажений иллюстрирует рис. 1. На нём представлено дискретное распределение Вигнера—Виля для тестового цифрового сигнала с частотой дискретизации f_D . По оси абсцисс графика отложены номера отсчётов N времени t , а по оси ординат — отношение отсчётов частоты f к частоте дискретизации. Тестовый сигнал содержит три частотно-модулированные составляющие $s_i(t)$, где $i = 1, 2, 3$, амплитуды которых равны и постоянны. Сигналы $s_1(t)$ и $s_2(t)$ имеют линейную частотную модуляцию (ЛЧМ) в соответствии с выражениями

$$f_1(t) = f_{\max} - \lambda t, \quad f_2(t) = f_{\min} + \lambda t.$$

Модуляция сигнала $s_3(t)$ отличается от модуляции сигнала $s_2(t)$ наличием дополнительной фазовой модуляции по закону гиперболического тангенса:

$$f_3(t) = f_2(t) - \frac{d\phi(t)}{dt},$$

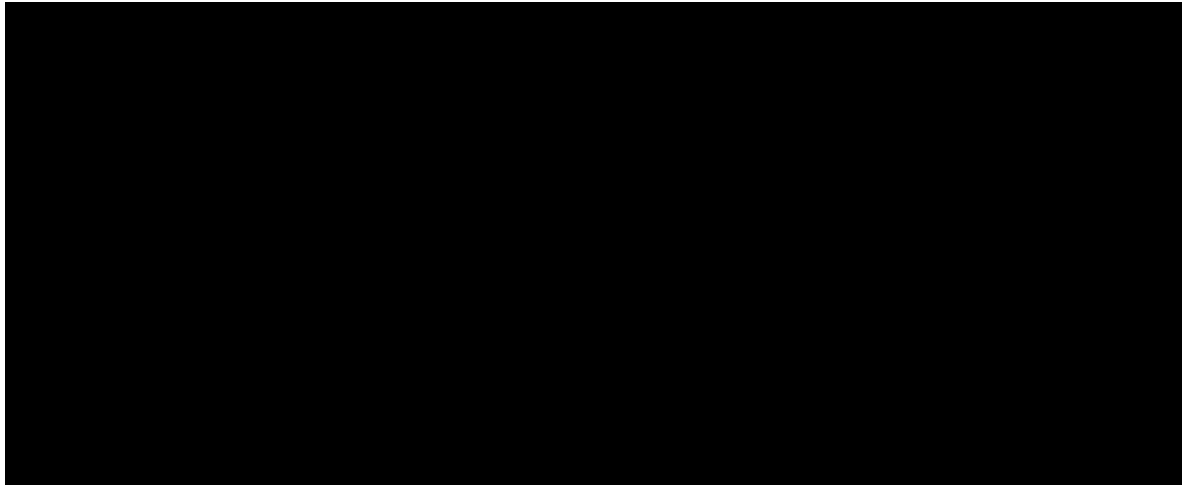


Рис. 1. Распределение Вигнера—Виля (ВВР), полученное с помощью 256-точечного быстрого преобразования Фурье для суммы 3-х сигналов с частотной модуляцией

где

$$\phi(t) = 0,04\pi t \operatorname{th}[k(t - t_s)].$$

В рассматриваемом примере время t безразмерно и отсчитывается в номерах отсчётов N , скорость λ изменения частот f_1 и f_2 равна 0,0001, коэффициент $k = 0,01$, а сдвиг t_s во временной зависимости фазы $\phi(t)$ равен $(f_{\max} - f_{\min})/(6\lambda)$. Соответствующие этим сигналам моды в распределении Вигнера—Виля представлены на рис. 1 двумя прямыми линиями (1 и 2) и кривой 3. Временная зависимость мгновенной частоты $f_3(t)$ имеет минимум на интервале между точками пересечения кривой 3 с прямыми 1 и 2.

Основные моды ВВР, представленные линиями 1–3, отражают распределение активного спектра мощности тестового сигнала на плоскости частота—время, поэтому с течением времени их амплитуды изменяются плавно и без смены знака. В отличие от основных мод продукты перекрёстных искажений знакопеременны в зависимости как от времени, так и от частоты. На рис. 1 это представлено чередованием светлых и тёмных участков в областях 4–6. «Пунктирные» области между линиями 2 и 3 примыкают к частоте $f_{23}(t) = [f_2(t) + f_3(t)]/2$ и представляют собой продукт взаимодействия сигналов $s_2(t)$ и $s_3(t)$ на тех интервалах времени, где частота сигнала $s_3(t)$ изменяется почти по линейному закону. Между этими интервалами тот же продукт взаимодействия выглядит как относительно широкая изрезанная полоса 5, но и она по-прежнему следует временной зависимости $f_{23}(t)$, к которой теперь добавляется обусловленная $f_3(t)$ нелинейность. Взаимодействие ЛЧМ мод $s_1(t)$ и $s_2(t)$ порождает интермодуляционные компоненты вблизи частоты $f_{12}(t) = [f_1(t) + f_2(t)]/2$ (области 4). Сюда же даёт вклад и продукт взаимодействия мод $s_1(t)$ и $s_3(t)$, локализованный вблизи частоты $f_{13}(t) = [f_1(t) + f_3(t)]/2$. Разделение указанных интермодуляционных компонент происходит в области 6, где мгновенная частота $f_3(t)$ значительно отличается от $f_2(t)$. Заметим, что резкость границ в графическом отображении интермодуляционных компонент зависит от выбора, с одной стороны, уровней градации изображения $P(f, t)$ по яркости, а с другой — шага изменения N и размерности использовавшегося при вычислении ВВР быстрого преобразования Фурье (БПФ). В данном случае применялось 256-точечное БПФ, а шаг сдвига по N временного окна анализа равнялся шести. При увеличении размерности БПФ и уменьшении шага по N картина вариаций интермодуляционных компонент становится более детальной, но не меняется по существу. Это иллюстрирует рис. 2, на котором приведён увеличенный фрагмент рис. 1 (область увеличения на рис. 1 выделена пунктирным прямоугольником).

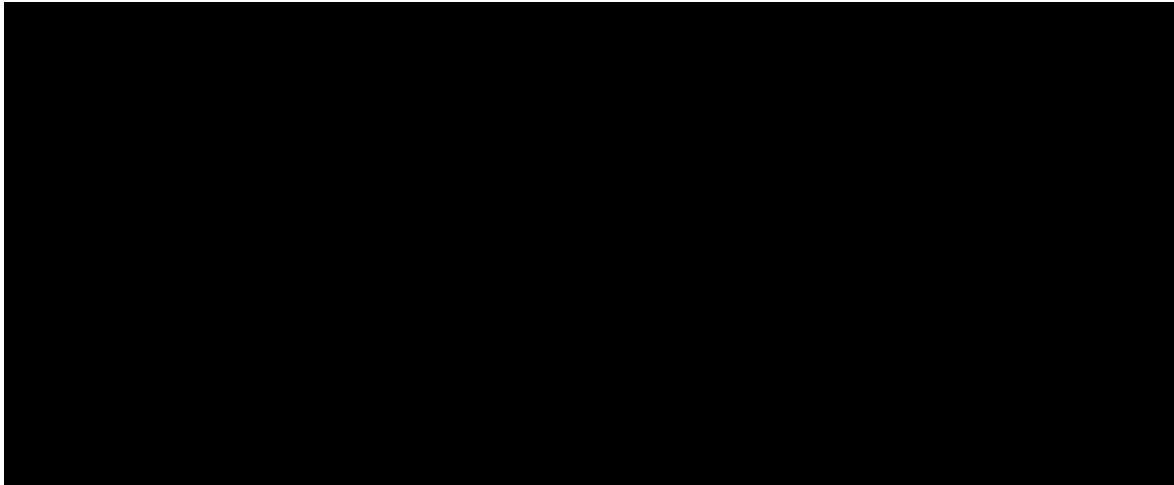


Рис. 2. Фрагмент представленного на рис. 1 распределения с бóльшим частотно-временным разрешением (1024-точечное БПФ)

2. МЕТОДЫ ФИЛЬТРАЦИИ ИНТЕРМОДУЛЯЦИОННЫХ КОМПОНЕНТ

Отмеченные выше отличительные особенности перекрёстных искажений, в принципе, дают возможность их отделения от соответствующих сигналу компонент ВВР. Это можно делать разными способами. Один из них — двумерная (по частоте и по времени) фильтрация соответствующего изображения ВВР. Однако этот способ эффективен лишь в случае, когда изображение достаточно качественное, и в нём обнаруживаются компоненты сигнала. Для реальных сигналов, которые целесообразно анализировать с помощью ВВР, получение таких изображений чаще всего затруднено, поскольку перекрёстные искажения по интенсивности превосходят компоненты мгновенного спектра сигнала. Кроме того, фильтрация уменьшает разрешающую способность, которую можно достичь с помощью ВВР.

В основе другого способа лежит использование в процессе вычисления ВВР элементов фильтрации и изменяемого окна анализа. При этом параметры фильтра и окна подбираются либо заранее по некоторому задаваемому правилу, либо в процессе анализа. В частности, так вычисляется распределение Чои—Вильямса (Choi—Williams distribution) [10]

$$C(f, t) = \int_{-\infty}^{+\infty} \exp(2\pi j f \tau) \int_{-\infty}^{+\infty} \sqrt{\frac{\sigma}{4\pi\tau^2}} \exp\left[-\frac{\sigma(\mu - t)^2}{4\tau^2}\right] s(\mu - \tau/2) s^*(\mu + \tau/2) d\mu d\tau. \quad (3)$$

Произведение $s(t - \tau/2)s^*(t + \tau/2)$ в (2) есть функция двух переменных — времени t и временного сдвига τ . Свёртка этого произведения с гауссовой функцией $\sqrt{\sigma/(4\pi\tau^2)} \exp[-\sigma t^2/(4\tau^2)]$ в (3) соответствует низкочастотной фильтрации по t . Полоса пропускания эквивалентного фильтра не постоянна и сужается по мере роста τ , поскольку τ является одним из параметров гауссовой функции. Кроме этого, гауссова функция формирует два симметрично расположенных относительно $\tau = 0$ окна, определяющих соответствующую им область интегрирования по τ . С помощью параметра σ задаётся приемлемое качество анализа за счёт подавления эффектов интермодуляции. Величина σ подбирается в процессе анализа путём адаптивного перебора.

Однако и низкочастотная фильтрация, и введение окна интегрирования по τ снижают точность спектрально-временного анализа. Это можно пояснить, исключив зависимость гауссовой функции от

τ и заменив распределение $C(f, t)$ на

$$\begin{aligned}\tilde{C}(f, t) &= \frac{1}{\sqrt{4\pi\delta^2}} \int_{-\infty}^{+\infty} \exp\left[-\frac{(\mu-t)^2}{4\delta^2}\right] d\mu \int_{-\infty}^{+\infty} s(\mu - \tau/2) s^*(\mu + \tau/2) \exp(2\pi j f \tau) d\tau = \\ &= \frac{1}{\sqrt{4\pi\delta^2}} \int_{-\infty}^{+\infty} \exp\left[-\frac{(\mu-t)^2}{4\delta^2}\right] P(f, \mu) d\mu. \quad (4)\end{aligned}$$

В алгоритме (4) параметр δ задаёт ширину гауссовой функции в подынтегральном выражении, а $P(f, \mu)$ является распределением Вигнера—Виля, определённым в соответствии с (2). Свёртка ВВР с функцией Гаусса эквивалентна «пропусканию» зависимости $P(f, t)$ как функции времени через гауссов фильтр нижних частот (ФНЧ). Таким приёмом можно подавить какую-то часть интермодуляционных компонент, но при этом разрешающая способность по времени t не может быть выше длительности импульсной реакции ФНЧ, определяемой величиной δ (или величиной $2\tau/\sqrt{\sigma}$, если иметь в виду распределение (3)). Введение в процедуру спектрально-временного анализа, наряду с фильтрацией, окна анализа по τ уменьшает также и разрешающую способность по частоте f , ограничивая разрешение снизу величиной $\Delta f \approx 1/T_a$, где T_a — протяжённость окна анализа, которая регулируется параметром σ .

В данной работе обсуждается применение преобразования Вигнера—Виля для анализа сигналов, обладающих компонентами с резко различающимися уровнями сосредоточенной в них энергии. Как правило, такие сигналы имеют достаточно сильно выраженные спектральные компоненты, выделить и измерить которые можно с помощью обычных методов спектрального анализа, например посредством измерения скользящего спектра. Эти значительные по мощности спектральные компоненты, взаимодействуя с остальными компонентами сигнала и друг с другом, создают значительные интермодуляционные искажения, на фоне которых практически не обнаруживаются слабые компоненты сигнала. Однако исследуемый сигнал можно подвергнуть предварительной обработке, подавив в нём ненужные с точки зрения анализа составляющие, и тем самым предотвратить появление вызываемых ими эффектов интермодуляции.

Часто анализируемые сигналы имеют характер нестационарного процесса

$$s(t) = \bar{s}(t) + \Delta s(t)$$

с незначительными отклонениями $\Delta s(t)$ от среднего значения $\bar{s}(t)$, которое относительно медленно меняется во времени. Нестационарность $\bar{s}(t)$ может быть вызвана, например, аддитивными добавками (как непрерывными, так и импульсными) с низкочастотными спектральными составляющими. На них приходится значительная доля энергии сигнала, поэтому низкочастотную область можно анализировать независимо. Анализ же значительно более высокочастотных, но менее энергичных колебаний $\Delta s(t)$ можно выполнять после предварительной низкочастотной фильтрации сигнала $s(t)$, что даёт возможность значительно снизить эффекты их взаимодействия с низкочастотными компонентами $s(t)$.

Проиллюстрируем сказанное на примере тестового сигнала, который образован двумя разными по частоте синусоидами и тремя разными по длительности гауссовскими импульсами:

$$\begin{aligned}s(N) &= A \sin(0,5\pi f_D N) + A \sin(0,45\pi f_D N) + 5 \exp\left[-\frac{(N-225)^2}{50a^2}\right] + \\ &+ 5A \exp\left[-\frac{(N-300)^2}{2a^2}\right] + 5A \exp\left[-\frac{(N-325)^2}{2a^2}\right]. \quad (5)\end{aligned}$$



Рис. 3. Распределение Вигнера—Виля для сигнала (а) и после пропускания сигнала через низкочастотный режекторный фильтр (б)

Здесь зависимость от времени t заменена на соответствующую зависимость от номера отсчёта N , параметр $a = 5$ задаёт длительность импульсов, A — амплитудный множитель, конкретное значение которого для дискретного анализа несущественно. Распределение Вигнера—Виля, вычисленное с помощью 256-точечного БПФ для сигнала (5), представлено на рис. 3а. Линии 1 и 2 соответствуют синусоидальным составляющим $s(t)$, область 3 — импульсу с длительностью $t_i = 10a$, а область 4 — двум импульсам с $t_i = 2a$. Штриховая прямая между линиями 1 и 2 — продукт взаимодействия гармонических составляющих, область 5 представляет собой результат их интерференции с импульсами. Поскольку это взаимодействие наиболее значительно для временных интервалов, в которых сосредоточена основная доля энергии импульсов (интервалов, по которым определяется их длительность), то ширина области 5 увеличивается в зоне расположения импульсных составляющих $s(t)$. Кроме того, заметно некоторое «возмущение» линий 1 и 2 продуктами интермодуляции. После предварительной обработки $s(t)$ низкочастотным режекторным фильтром с полосой запираания $\Pi = 0,08f_D$ принадлежащие области 5 интермодуляционные компоненты значительно ослабляются (см. рис. 3б), а влияние

импульсов на линии 1 и 2 в ВВР практически не обнаруживается. При этом в области, соответствующей продукту взаимодействия синусоидальных составляющих $s(t)$ (область между линиями 1 и 2), никаких изменений не наблюдается.

3. ОБРАБОТКА РЕЗУЛЬТАТОВ НАБЛЮДЕНИЙ ПУЛЬСАЦИЙ СОЛНЕЧНОГО РАДИОИЗЛУЧЕНИЯ

Столь подробное описание особенностей ВВР для тестовых сигналов имело своей целью пояснить ситуацию, с которой нередко приходится сталкиваться в процессе спектрально-временного анализа реальных наблюдательных данных. Обсудим это, опираясь на результаты исследования солнечной радиации по записям временных вариаций интенсивности радиоизлучения Солнца на частоте 37 ГГц, которые были получены в радиообсерватории Метсахови (Metsähovi Radio Observatory, Финляндия) и любезно предоставлены авторам сотрудником радиообсерватории С. Урпо (S. Urpo) при содействии В. В. Зайцева (ИПФ РАН) [11, 12].

На рис. 4 представлена запись сопровождавшего солнечную вспышку радиоизлучения, выполненная 24.03.91 на интервале 14:04:06–14:26:06 UT. Для наблюдавшегося сигнала $s(t)$ характерно значительное (более чем на четыре порядка) изменение уровня мощности. Поэтому приведённая на рис. 4 осциллограмма отражает в основном только изменение среднего значения $\bar{s}(t)$ без деталей, свойственных высокочастотной части $\Delta s(t)$. Частота дискретизации f_D равнялась 2 Гц. Анализируемый отрезок записи с длительностью T_i ограничен начальным N_1 и конечным N_2 отсчётами. При вычислении ВВР применялось 1024-точечное БПФ с окном Парзена. Шаг перемещения окна по времени (т. е. по N) составлял $\Delta N = 4$.

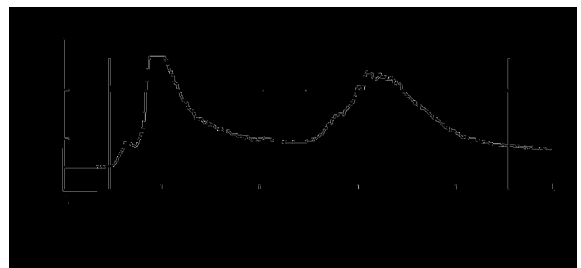


Рис. 4. Фрагмент записи радиоизлучения Солнца 24.03.91 в период 14:04:06–14:26:06 UT. По оси ординат отложена интенсивность излучения в относительных единицах

Распределение Вигнера—Вилля для этого сигнала приведено на рис. 5. Полученная спектрограмма соответствует частотному интервалу $(0,04 \div 0,26) f/f_D$. Нижняя граница по частоте увеличена до $0,04 f/f_D$ по той причине, что графическое отображение нижележащей части спектрограммы ухудшает её качество в целом. На спектрограмме видны только продукты перекрёстных искажений, что подтверждается сопоставлением её со спектрограммой, полученной путём вычисления скользящего спектра с помощью 1024-точечного БПФ с тем же перемещающимся окном анализа (см. рис. 6). Текущий спектр, представленный на рис. 6, показывает наличие в сигнале, по крайней мере, четырёх составляющих: 1 — с линейной частотной модуляцией, 2, 3 и 4 — с постоянными частотами. Ни одна из них не видна на спектрограмме на рис. 5, однако на ней присутствует интермодуляционный след 1 от взаимодействия ЛЧМ сигнала 1 с интенсивной низкочастотной составляющей исследуемого сигнала $\bar{s}(t)$, представленной двумя импульсами на рис. 4.

Предварительная фильтрация исходного сигнала с помощью низкочастотного режекторного фильтра с полосой подавления $\Pi = 0,16 f_D$ позволило в значительной степени уменьшить эффекты интермодуляции и выделить на полученном ВВР компоненты, соответствующие слабым по сравнению с $\bar{s}(t)$ вариациям $\Delta s(t)$. Последствия предварительной обработки (см. рис. 7) здесь во многом аналогичны тем, что представлены на рис. 3 для тестового сигнала (5). Во вновь полученном спектрально-временном распределении компоненты 1–4 соответствуют тем, которые обнаруживаются и на рис. 6, но со значительно большим разрешением по частоте и времени. Продукты интермодуляции достаточно

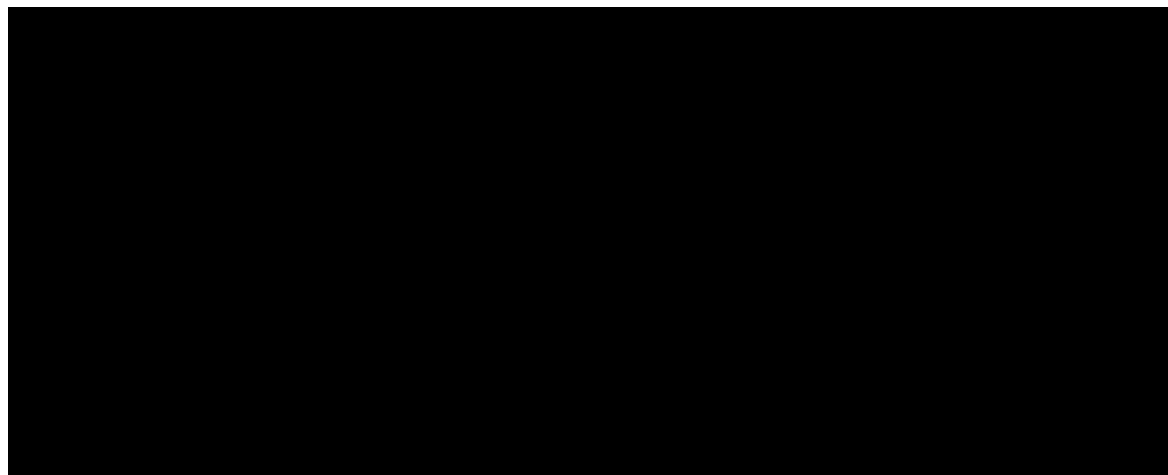


Рис. 5. ВВР, полученное с помощью 1024-точечного БПФ, для сигнала, представленного на рис. 4

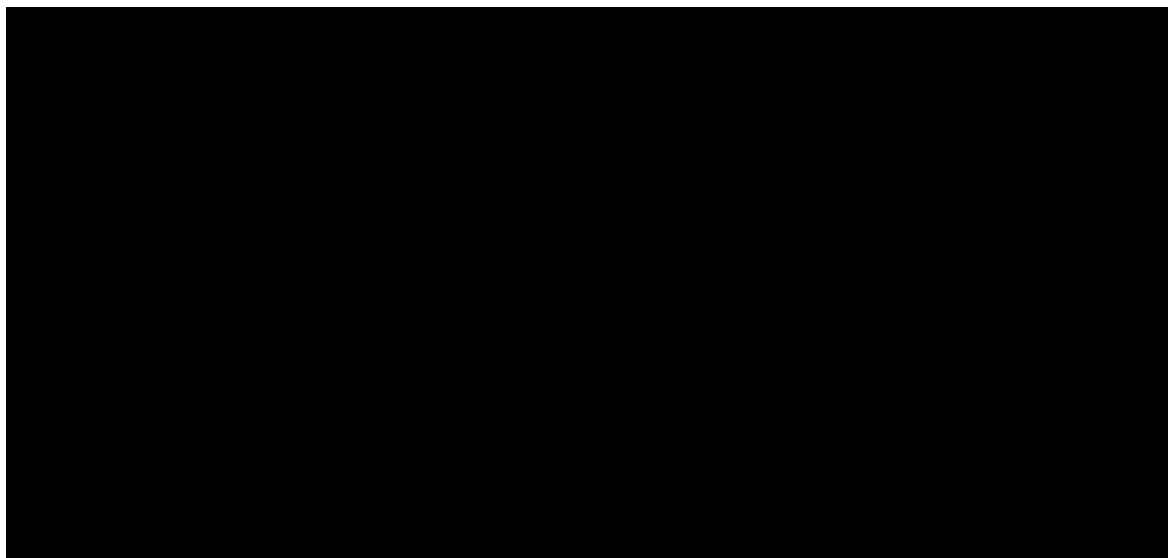


Рис. 6. Фурье-спектр сигнала, представленного на рис. 4

слабы, и на их фоне спектральные компоненты сигнала уверенно различаются. Среди оставшихся интермодуляционных компонент выделяется область 5 — результат взаимодействия ЛЧМ составляющей 1 с фиксированной по частоте составляющей 2 сигнала $s(t)$. Различимыми стали интермодуляционные компоненты, возникшие из-за взаимодействия составляющих 2, 3 (область 6) и 3, 4 (область 7).

Предварительная фильтрация сигнала может быть выполнена и другими способами. Возможна аппроксимация среднего значения сигнала методом наименьших средних квадратов с последующим вычитанием расчётной и экспериментальной кривых. Как показал опыт [13], при этом удаётся существенно увеличить контраст слабых ЛЧМ сигналов в динамических спектрах солнечного радиоизлучения. Очевидно, что вычитание среднего аналогично пропусканию сигнала через некоторый фильтр верхних частот. Более кардинальное решение этой проблемы заключается в нормировке исследуемого сигнала на его огибающую. Действительно, аналитический сигнал $s(t) = x(t) + jy(t)$ можно представить в виде

$$s(t) = E(t) \exp[j\Theta(t)],$$

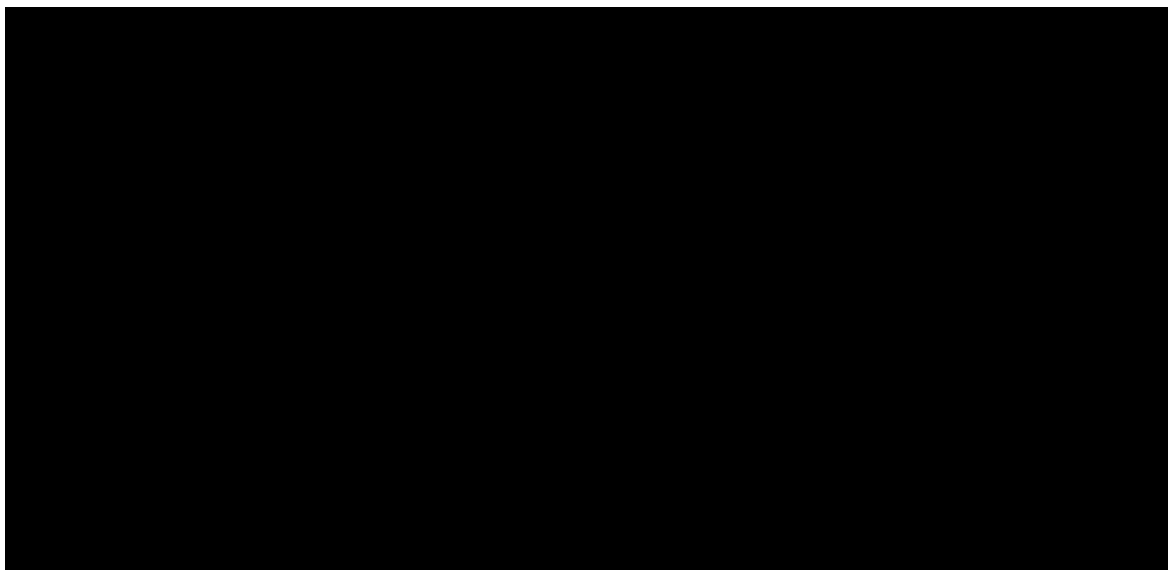


Рис. 7. ВБР для представленного на рис. 4 сигнала после обработки последнего низкочастотным режекторным фильтром с полосой запираания $\Pi = 0,16f_D$

где $E(t) = \sqrt{x^2(t) + y^2(t)}$ — огибающая, а $\Theta(t) = \arctg[y(t)/x(t)]$ — мгновенная фаза сигнала. Составив отношение $s(t)/E(t) = s_N(t)$, придём к нормированной сигнальной функции, в которой максимально подчёркнуты быстро осциллирующие составляющие сигнала. Численные эксперименты и применение указанной нормировки при обработке экспериментальных данных показали, что использование $s_N(t)$ приводит к положительным результатам, когда необходимы подавление кратковременных интенсивных импульсов и поиск слабых квазигармонических компонент. В то же время использование $s_N(t)$ не избавляет от необходимости фильтрации продуктов интерференции этих компонент.

4. ЗАКЛЮЧЕНИЕ

Таким образом, применение предварительной обработки исследуемых сигналов в определённых ситуациях позволяет повысить качество спектрально-временного анализа с применением ВБР. Вопрос о конкретном способе такой обработки можно решить, основываясь как на виде самого сигнала (его временной зависимости), так и на результатах применения к нему более простых линейных методов спектрального анализа, которые позволяют обнаружить интенсивные спектральные компоненты и в последующем исключить их влияние на процесс спектрально-временного анализа с высокой разрешающей способностью. Иногда положительного эффекта можно достичь, добавляя к исследуемому сигналу значительные по уровню компоненты с известными и заранее определёнными параметрами (амплитудой, частотой, видом и законом модуляции). Такие добавки, порождая новые продукты интермодуляции, позволяют по их следам обнаруживать полезные спектральные компоненты и область их локализации и дают информацию, по которой можно выбрать подходящий метод обработки сигнала, предшествующей его анализу.

На основе представленных в работе алгоритмов создан пакет программ для получения ВБР с ослабленными (или подавленными) эффектами интермодуляции. При этом также удаётся, как следует из приведённых выше результатов, отделить истинные сигналы от их продуктов взаимодействия с более мощными импульсными и гармоническими «помехами». Указанные программы успешно использовались при исследовании динамических спектров пульсаций радиоизлучения Солнца в миллиметровом диапазоне длин волн [13].

Работа поддержана РФФИ (грант № 01–02–16435а).

СПИСОК ЛИТЕРАТУРЫ

1. Гун С., Рао Д., Арун К. // Сверхбольшие интегральные схемы и современная обработка сигналов: Пер с англ. / Под ред. С. Гуна, Х. Уайтхауса, Т. Кайлата. М.: Радио и связь, 1980.
2. Макс Ж. Методы и техника обработки сигналов при физических измерениях: В 2-х томах. Пер. с франц. М.: Мир, 1983. Т. 2. 256 с.
3. Wigner E. P. // Phys. Rev. 1932. V. 40. P. 749.
4. O'Connell R. F. // Found. Phys. 1983. V. 13, No. 1. P. 83.
5. Woodworth F. M. Probability and information theory with applications to radar. Pergamon press, 1953.
6. Boualem Boashash. // IEEE Trans. on Acoustics, Speech, and Signal Processing. 1988. V. 36, No. 9. P. 1518.
7. Johnson J. A. // IEE Proc. F. 1989. V. 136, No. 2. P. 81.
8. Бабанов Ю. Н., Лебедев Ю. П., Воинов Б. С. Спектрально-временной анализ детерминированных сигналов: Учебное пособие. Горький: Изд-во ГГУ, 1980. 91 с.
9. Shubha Kadambe // IEEE Trans. on Signal Processing. 1992. V. 40, No. 10. P. 2498.
10. Коэн Л. // ТИИЭР. 1989. Т. 77, № 10. С. 72.
11. Zaitsev V. V., Kislyakov A. G., Urpo S., Stepanov A. V., Shkelev E. I. The Wigner—Ville analysis of solar microwave emission. Metsähoivi Publications on Radio Science, Espoo 2000, HUT-MET-32.
12. Зайцев В. В., Кисляков А. Г., Степанов А. В., Урпо С., Шкелёв Е. И. // Изв. вузов. Радиофизика. 2001. Т. 44, № 1—2. С. 38.
13. Зайцев В. В., Кисляков А. Г., Урпо С., Шкелёв Е. И. // Изв. вузов. Радиофизика. 2001. Т. 44, № 9. С. 756.

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

Поступила в редакцию
2 июля 2001 г.

METHODS FOR ATTENUATION OF THE CROSS-MODULATION EFFECTS IN THE WIGNER—VILLE DISTRIBUTION

E. I. Shkelev, A. G. Kislyakov, and S. Yu. Lupov

We consider the characteristics of typical nonlinear distortions in the dynamical spectra of signals, obtained using the Wigner—Ville method, and discuss existing techniques for attenuating these distortions. Certain methods are proposed to “clean” the Wigner—Ville distributions of the products of the nonlinear interaction between the signal component under study and its more powerful harmonic or pulsed components. The proposed techniques are mainly based on a preliminary linear processing (e.g., using the fast Fourier transform) and/or frequency filtering of the signal. A software package for calculating the Wigner—Ville distribution on a PC is developed. Using this package, we approved successfully the proposed algorithms for test and real signals.

УДК 534.222

О ТОЧНОСТИ ОПРЕДЕЛЕНИЯ ПАРАМЕТРОВ РЕВЕРБЕРАЦИИ ПРИ ВОССТАНОВЛЕНИИ СИГНАЛА С РЕВЕРБЕРАЦИОННЫМИ ИСКАЖЕНИЯМИ В СЛУЧАЕ ДИСКРЕТНЫХ ЗАДЕРЖЕК

А. А. Павленко

В работе рассмотрена точность определения параметров реверберации, необходимая для восстановления сигнала с реверберационными искажениями. Показано, что если сигнал с искажениями оцифрован с выполненными условиями теоремы отсчётов Котельникова, то недостаточно определять время задержки запаздывающего сигнала с точностью до интервала дискретизации. Приведён итерационный алгоритм нахождения параметров реверберации с достаточной точностью.

При проведении акустических измерений часто приходится иметь дело с интерференционными искажениями. К примеру, звуковая волна, излучённая точечным акустическим источником в море, претерпевает многочисленные отражения от дна и поверхности воды. В результате возникает эффект реверберации, который заключается в том, что к приёмнику приходят многочисленные сигналы, распространяющиеся различными путями и запаздывающие относительно прямого сигнала, приходящего непосредственно от источника к приёмнику. Интерферируя, эти сигналы могут существенно исказить результаты измерений. В частности, такие интерференционные искажения, связанные с реверберацией помещений, затрудняют распознавание речи. Поэтому интерференционные искажения желательно устранять.

В принципе, такая задача имеет решение, т. к. процесс реверберации представляет собой линейное преобразование, описываемое эквивалентным фильтром (ЭФ). Чтобы устранить реверберацию, достаточно пропустить сигнал с реверберационными искажениями через фильтр, частотная характеристика которого обратна частотной характеристике ЭФ. Заметим, что такое преобразование для физических сигналов, подвергшихся реверберации, всегда существует, т. к. существует реальный прямой сигнал, не подвергшийся действию реверберации.

Для исключения реверберации (дереверберации) необходимо знать характеристику ЭФ. Такая характеристика может быть определена на основе исследований среды, в которой распространяются волны, или сигнала, искажённого реверберацией. В последнем случае мы имеем дело с так называемой слепой дереверберацией (blind dereverberation). Известен метод слепой дереверберации [1–3], в котором частотная характеристика ЭФ определяется на основе принципа минимального фазового сдвига при наличии двух точек приёма сигнала с реверберацией. Другой метод [4], применимый в случае дискретных задержек реверберирующих сигналов, основан на извлечении необходимой информации о параметрах ЭФ непосредственно из принятого искажённого сигнала с помощью кепстрального анализа [5].

Несмотря на принципиальное отличие подходов к решению задачи, все методы дереверберации как слепой, так и основанный на исследованиях среды, базируются на определении характеристики ЭФ, вид которой не зависит от применяемого метода. От метода зависит лишь точность, с которой определяются параметры ЭФ. Целью настоящей работы является установление точности, с которой достаточно определять параметры ЭФ для эффективного подавления реверберации.

Пользуясь тем, что результат дереверберации сигнала не зависит от способа её осуществления, рассмотрим вопрос о точности определения параметров ЭФ на примере слепой дереверберации сигнала с дискретными задержками [4].

А. А. Павленко

443

В общем случае сигнал, содержащий реверберационные искажения, описывается формулой

$$R(t) = \int_0^{+\infty} X(\tau) S(t - \tau) d\tau, \quad (1)$$

где $S(t)$ — исходный сигнал без искажений, $X(\tau)$ — отклик некоторого линейного фильтра на импульсное воздействие в виде δ -функции. В спектральном представлении выражение (1) имеет вид

$$\Phi_t[R(t)] = \Phi_t[X(t)] \Phi_t[S(t)], \quad (2)$$

где $\Phi_t[\]$ — преобразование Фурье по переменной t для функции в квадратных скобках. Очевидный и единственный способ восстановления исходного сигнала из искажённого реверберацией состоит в том, чтобы поделить спектр искажённого сигнала на частотную характеристику фильтра $X(\omega)$ и взять от полученной функции обратное преобразование Фурье:

$$S(t) = \Phi_\omega^{-1}[\Phi_t[R(t)]/\Phi_t[X(t)]], \quad (3)$$

где $\Phi_\omega^{-1}[\]$ — обратное преобразование Фурье для выражения в квадратных скобках.

Для нахождения функции $X(\tau)$ можно воспользоваться одним из указанных выше методов. Однако здесь мы не будем останавливаться на этой процедуре, а попробуем выяснить, с какой точностью должны быть проведены вычисления, чтобы впоследствии успешно «очистить» сигнал от реверберационных искажений.

Примем следующую математическую модель принимаемого сигнала. Пусть имеется первоначальный сигнал $S(t)$, который повторяется через интервалы времени τ_n с изменёнными амплитудами a_n . Согласно этой модели принимаемый сигнал можно записать в следующем виде:

$$R(t) = S(t) + \sum_n a_n S(t - \tau_n).$$

Отклик фильтра на импульсное воздействие (см. формулу (1)) при такой модели сигнала представляет собой следующую сумму δ -функций:

$$X(t) = \delta(t) + \sum_n a_n \delta(t - \tau_n). \quad (4)$$

Исходя из (4) спектральное представление (2) запишется в виде

$$\Phi_t[R(t)] = \Phi_t[S(t)] \left[1 + \sum_n a_n \exp(-i\omega\tau_n) \right]. \quad (5)$$

Из (5) легко получается выражение для нахождения исходного сигнала:

$$S(t) = \Phi_\omega^{-1} \left[\frac{\Phi_t[R(t)]}{1 + \sum_n a_n \exp(-2\pi i f \tau_n)} \right]. \quad (6)$$

Из (6) следует, что основными параметрами ЭФ в данном случае являются амплитуды a_n запаздывающих сигналов и задержки τ_n . Точность восстановления амплитуд a_n определяется исходя из вида выражения (6). Относительная точность определения амплитуд a_n должна быть не ниже той, которую мы хотим достичь при подавлении реверберационных искажений. Вопрос о точности определения задержек τ_n требует дополнительного исследования.

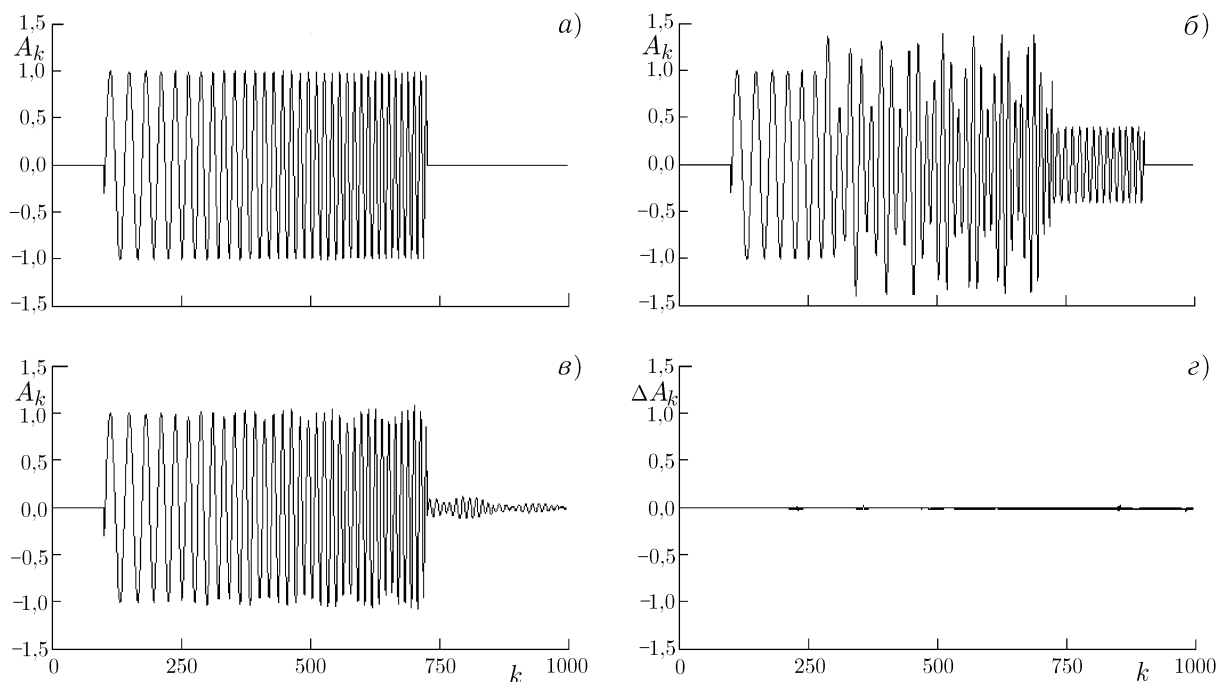


Рис. 1

Сигналы предполагаются непрерывными и обладающими ограниченным спектром. В соответствии с теоремой Котельникова такой сигнал может быть представлен в виде дискретных отсчётов по времени. Такое представление необходимо для ввода сигнала в ЭВМ. В то же время при таком представлении удобно ввести понятие точности ε_n определения параметра τ_n : $\varepsilon_n = |t_n - \tau_n|$, где τ_n — действительные времена запаздывания сигнала, а t_n — вычисленные. Наиболее точное значение τ_n получается в случае, когда оно равно истинному значению. Однако истинное значение этого параметра может быть некрatным шагом дискретизации. В этом случае, оставаясь в рамках дискретной сетки интервалов задержки τ_n , мы никогда не получим значение, в точности равное истинному. Качество восстановления сигнала при помощи фильтра (6) с найденными с точностью до шага дискретизации Δt параметрами t_n при этом резко ухудшается. Более того, даже если задержка определена абсолютно точно, восстановление сигнала при помощи фильтра (6) даёт неудовлетворительные результаты. Последнее обстоятельство можно объяснить следующим образом. Умножение спектра сигнала на $\exp(2\pi i f \tau_n)$ соответствует сдвигу сигнала во временной области на τ_n . Если $\tau_n = k \Delta t$, где k — целое, то сигнал просто сдвигается по оси времени на целое число отсчётов. Если же k не является целым числом, то результат операции сдвига не определён, поскольку изначально не определены значения сигнала в промежутках между отсчётами.

Чтобы убедиться в истинности последнего утверждения, рассмотрим следующий пример. В качестве исходного возьмём сигнал с линейной частотной модуляцией, который описывается выражением

$$A_k = \sin[2\pi (fk + ak^2)/L]. \quad (7)$$

Здесь индекс $k = 1, 2, \dots, L$, длину реализации L возьмём равной 1 024 точкам, $f = 15$; $a = 0,05$ (см. рис. 1а). Для наглядности сигнал A_k занулён на концах интервала при $k < 100$ и при $k > 730$. В сигнале смоделируем одну задержку с длительностью $178,4 \Delta t$ и амплитудой, равной 0,4 амплитуды исходного сигнала (см. рис. 1б). Теперь попробуем восстановить сигнал при помощи фильтра (6). Результаты восстановления сигнала приведены на рис. 1в. Даже без каких-либо количественных оценок видно, что восстановленный сигнал имеет ярко выраженную периодическую компоненту. Это свидетельствует о том, что задача восстановления не решена.

Из всего вышесказанного можно сделать простой вывод: модель (3) не учитывает дробность (по отношению к интервалу дискретизации) задержек запаздывающих сигналов. Это означает, что для решения задачи нужно либо рассматривать иную модель описания сигнала с реверберационными искажениями, либо каким-то образом модифицировать метод обработки сигнала, чтобы учитывалась дробность задержки.

На первый взгляд, проще всего увеличить частоту квантования исходных данных. В этом случае задержка может стать целой (кратной интервалу дискретизации), и будет работать рассмотренный выше алгоритм. Однако в большинстве случаев приходится иметь дело с уже оцифрованными сигналами. Имея только отсчёты и не располагая оригинальным аналоговым сигналом, частоту квантования можно лишь уменьшить. Кроме того, чтобы существенно (например, в 10 раз) увеличить точность, во столько же раз необходимо увеличить и частоту квантования. Поэтому необходим более универсальный и оптимальный (в смысле количества вычислительных операций) алгоритм. По сути дела, нам необходимо получить значения функции в промежутках между отсчётами. Для этой цели можно применить описанную в [6] процедуру, называемую «дополнением нулями». Операция дополнения нулями позволяет применить преобразование Фурье для интерполяции между исходными отсчётами. Рассмотрим некоторые свойства данной операции.

Пусть имеющиеся отсчёты данных $x_0, \dots, x_{L-1}$ дополнены нулевыми значениями $x_L, \dots, x_{2L-1}$. Дискретно-временной ряд Фурье этой дополненной нулями последовательности данных будет определяться выражением

$$X_k = T \sum_{n=0}^{2L-1} x_n \exp[-2\pi i n k / (2L)] = T \sum_{n=0}^{L-1} x_n \exp[-2\pi i n k / (2L)], \quad (8)$$

где верхний предел суммы в правой части изменён таким образом, чтобы учесть наличие введённых нулевых отсчётов. Пусть теперь $k = 2m$, так что

$$X_m = T \sum_{n=0}^{L-1} x_n \exp(-2\pi i n m / L), \quad (9)$$

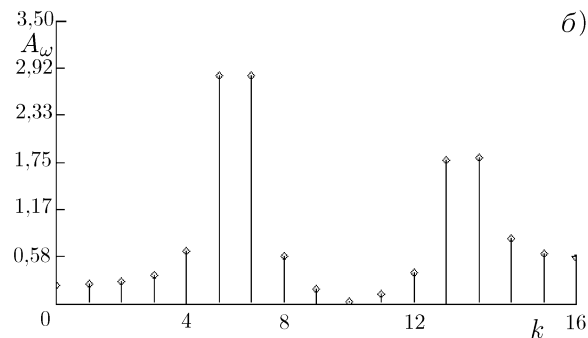
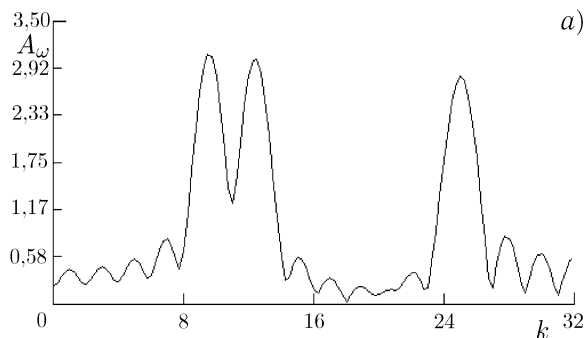
где $m = 0, 1, \dots, L-1$, определяет X_k при чётных k . Следовательно, при чётном индексе k $2L$ -точечный дискретно-временной ряд Фурье сводится к L -точечному дискретно-временному ряду. Нечётные индексы k соответствуют интерполированным значениям, расположенным между значениями исходного L -точечного ряда.

Влияние дополнительных нулей на интерполирование иллюстрирует рис. 2. На нём представлены спектры 32-точечной реализации сигнала, определяемого выражением

$$A_k = \sin(2\pi f_1 k) + \sin(2\pi f_2 k) + \sin(2\pi f_3 k), \quad (10)$$

где параметры f_1 , f_2 и f_3 равны соответственно 0,15; 0,194 и 0,389. На рис. 2а представлено непрерывное преобразование Фурье для последовательности (10), на рис. 2б — дискретно-временной ряд Фурье для последовательности A_k , а на рис. 2в представлен спектр той же последовательности, но с восьмикратным увеличением числа отсчётов за счёт дополнения исходного ряда нулями. Как видно из рис. 2, операция дополнения нулями позволяет получить значения сигнала в промежуточных точках между отсчётами.

Однако при непосредственном дополнении исходного сигнала нулями мы получим интерполированные значения в спектре сигнала. Чтобы интерполировать сам сигнал, очевидно, нужно дополнять нулями его спектр. В итоге процесс фильтрации (6) разбивается на три этапа. Сначала вычисляется спектр сигнала с искажениями $R(t)$. Затем полученный спектр дополняется нулями, и вычисляется



спектр исходного сигнала $S(t)$. Очистив таким образом спектр сигнала от реверберационных искажений, нам необходимо получить сам сигнал. Для этого нужно сделать преобразование, обратное дополнению нулями. Эту операцию можно выполнять, поскольку в результате такого преобразования мы ни коим образом не искажаем форму сигнала, а лишь убираем промежуточные интерполированные точки. Результат выполнения описанных действий иллюстрирует рис. 1г, где представлена разность восстановленного и исходного сигналов после вычисления уточнённых значений τ_n .

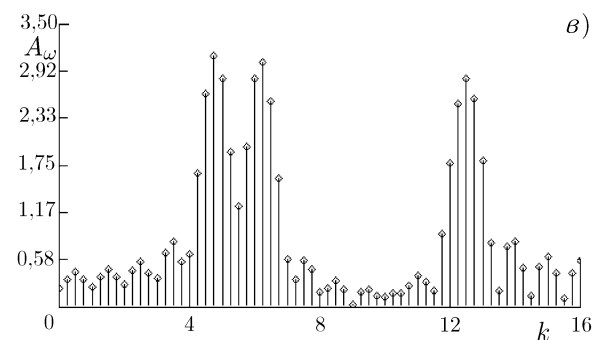


Рис. 2

Далее нам требуется установить, во сколько раз (обозначим это число M) необходимо увеличить количество отсчётов, чтобы получить удовлетворительный результат. Теоретически наилучшим вариантом является такое количество дополнительных нулей, при котором задержка становится целой (т.е. кратной новому интервалу дискретизации $\Delta q = \Delta t/M$). Таким образом, чем больше M , тем больше вероятность того, что задержка станет целой. Однако более актуальным является вопрос о выборе числа M при практических вычислениях. Решить эту задачу аналитически невозможно, поскольку нам заранее неизвестна форма принимаемого сигнала. Решение этой задачи можно найти только в численном эксперименте.

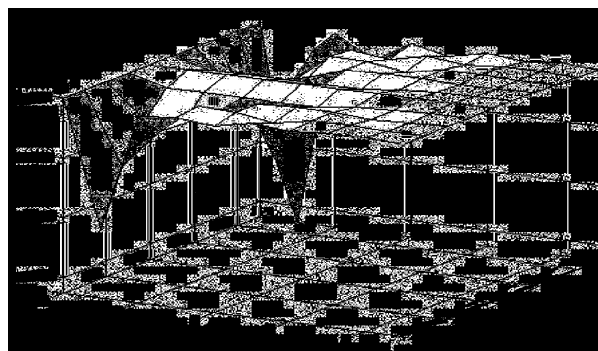


Рис. 3

Такой эксперимент был проведён и состоял в следующем. Модельный сигнал (7) подвергался фильтрации (6) с дополнением спектра нулями, причём число M варьировалось от 1 до 10. Как было указано выше, кроме количества дополнительных нулей на точность восстановления сигнала существенно влияет отклонение реальной задержки от её округлённого значения. В сущности, при $M = 1$ это как раз будет дробная часть d задержки. Чтобы получить количественную информацию о том, каким образом точность восстановления зависит от d , была проведена серия экспериментов, в которых величина d изменялась от $-0,5$ до $0,4$ с шагом $0,1$. Точность восстановления оценивалась с помощью коэффициента корреляции r восстановленного сигнала и сигнала, не содержащего искажений, который нам был известен заранее. Зависимость коэффициента корреляции от количества вводимых нулей и дробной части задержки показана на рис. 3. Из рис. 3 видно, что уже при $M = 8$ алгоритм даёт практически одинаковые результаты для любых значений d , включая $d = 0$.

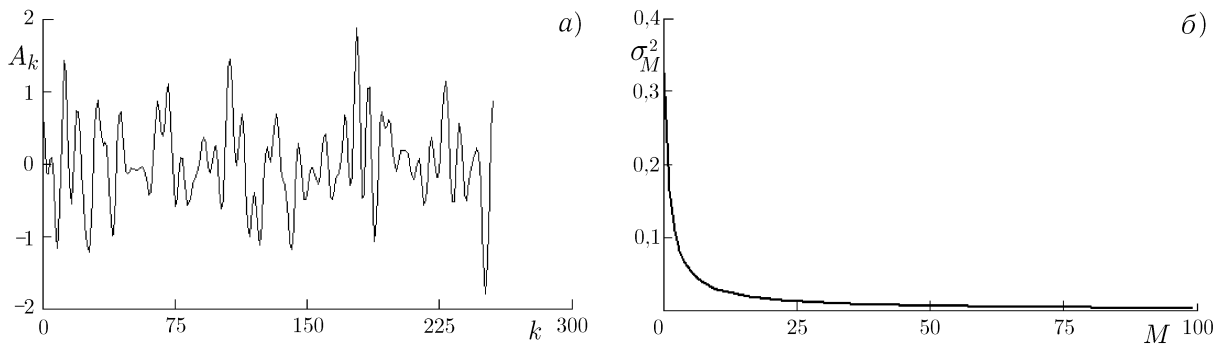


Рис. 4

В общем случае алгоритм выбора числа M можно описать следующим образом. Количество нулей подбирается итерационным методом. Прежде всего задаётся некоторая точность определения задержки $\varepsilon > 0$ и выбирается произвольное число M (можно взять $M = 8$, как показано в рассмотренном примере). Для этого числа M проводятся все вычисления, в результате которых получается реализация восстановленного сигнала S_M . Затем число M увеличивается и выполняется следующая итерация, в результате которой получаем реализацию S_{M+1} . Полученные реализации сравниваются путём вычисления коэффициента корреляции r или взаимного среднеквадратичного отклонения σ_M^2 . Полученная оценка сравнивается с заданным значением ε . Если $\sigma_M^2 > \varepsilon$, то итерация повторяется. Очевидно, что данный процесс сходится, т. к. при увеличении M искусственно уменьшается интервал дискретизации, а следовательно, найденное округлённое значение τ_n приближается к реальной задержке, и сам сигнал восстанавливается более точно.

Рис. 4 иллюстрирует это утверждение. На рис. 4а показан тестовый сигнал, который представляет собой последовательность из $L = 256$ дискретных отсчётов, которая получена методом, подробно описанным в [7, 8]. Эта последовательность является реализацией случайной нормально распределённой функции с заданным спектром, который определялся отсчётами, отстоящими один от другого на интервалы $2\pi/L$. Всего задавалось 50 частот (на каждую частоту приходится две точки спектра), примыкающих к нулю, из 128 возможных. Остальные отсчёты, не вошедшие в указанные 50, принимались равными нулю. Количество точек в спектре сигнала изменялось за счёт дополнения нулями. Нули добавлялись сериями длиной L , число M таких серий изменялось в пределах от 0 до 100.

Что касается амплитуд a_n запаздывающих сигналов, то ситуация с ними сравнительно простая, поскольку ни моделью, ни алгоритмом никаких ограничений на них не налагается. Амплитуды a_n могут быть найдены с любой наперёд заданной точностью. Для их вычисления применим тот же итерационный метод, что и для задержек τ_n .

Таким образом, получено, что независимо от способа определения параметров ЭФ задержки должны быть определены с точностью, намного превышающей интервал временного квантования сигнала. Путём численного моделирования на конкретном примере показано, что достаточным может быть определение задержек с точностью до одной восьмой или до одной шестнадцатой интервала квантования. Полученный на конкретном примере результат, по-видимому, имеет достаточно общий характер.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проекты № 99–02–16401 и 00–15–9674).

СПИСОК ЛИТЕРАТУРЫ

1. Hayes M. H., Lim J. S., Oppenheim A. V. // IEEE Trans. Acoustics, Speech and Processing. 1980. V. ASSP-28, No. 6. P. 672.
2. Subramaniam S., Petropulu A. P., Wendt C. // IEEE Trans. Speech and Audio Processing. 1996. V. 4, No. 5, P. 392.
3. Petropulu A. P., Nikias C. L. // IEEE Trans. Signal Processing. 1993. V. 41, No. 6. P. 2088.
4. Зверев В. А., Павленко А. А., Соколов А. Д., Шаронов Г. А. // Акустический журнал. 2001. Т. 47, № 1. С. 76.
5. Oppenheim A. V., Schaffer R. W., Stockham T. G. // IEEE Trans. Audio and Electroacoustics. 1968. V. AU-16, No. 3. P. 1264.
6. Марпл-мл. С. Л. Цифровой спектральный анализ и его приложения: Пер. с англ. М.: Мир, 1990.
7. Зверев В. А. Физические основы формирования изображений волновыми полями. Нижний Новгород: ИПФ РАН, 1998. 252 с.
8. Зверев В. А., Литвак Н. В. // Акустический журнал. 1999. Т. 45, № 6. С. 807.

Институт прикладной физики РАН,
г. Нижний Новгород, Россия

Поступила в редакцию
25 января 2001 г.

**ACCURACY OF DETERMINATION OF THE REVERBERATION PARAMETERS FOR RETRIEVAL OF A SIGNAL WITH
REVERBERATION DISTORTIONS IN THE CASE OF DISCRETE DELAYS**

A. A. Pavlenko

In this paper, we consider the accuracy of determination of the reverberation parameters, which is necessary for retrieval of a signal with reverberation distortions. We show that if the signal with distortions is digitized under the conditions of the Kotelnikov theorem, then it is insufficient to determine the temporal delay of the retarded signal with an accuracy of the sampling interval. We present an iterative algorithm allowing for determination of the reverberation parameters with the sufficient accuracy.

УДК 537.86:519.2

ХАРАКТЕРИСТИКИ ЭНЕРГЕТИЧЕСКОГО ОБНАРУЖЕНИЯ КВАЗИДЕТЕРМИНИРОВАННЫХ РАДИОСИГНАЛОВ С НЕРЭЛЕЕВСКИМ РАСПРЕДЕЛЕНИЕМ ФЛУКТУАЦИЙ АМПЛИТУДЫ

В. И. Костылев

Найдено распределение решающей статистики дискретного энергетического обнаружителя в случае приёма аддитивной смеси квазидетерминированного сигнала с нерэлеевским распределением флуктуаций амплитуды и белого шума. Получено выражение для вероятности правильного обнаружения.

В [1] проанализирована эффективность дискретного энергетического обнаружения квазидетерминированных радиосигналов с рэлеевским распределением флуктуаций амплитуды. Однако известно [2], что рэлеевские флуктуации амплитуды отражённого объектом сигнала имеют место лишь для некоторых типов объектов, участков частотного диапазона, секторов углов наблюдения и конкретных характеристик среды распространения радиоволн. Поэтому рэлеевская модель флуктуаций амплитуды квазидетерминированного радиосигнала зачастую слишком идеализирует реальный процесс формирования вторичного излучения [2]. Для описания статистических характеристик флуктуаций амплитуды реальных радиосигналов было разработано или привлечено из других областей знания большое число так называемых нерэлеевских моделей [2, 3]. В отличие от однопараметрического распределения Рэлея нерэлеевские модели могут иметь два, три и более параметров, имеющих различный физический смысл.

Независимо от закона распределения флуктуаций случайной амплитуды A выражение для условной характеристической функции $\Theta(j\eta|A)$ решающей статистики Ξ дискретного энергетического обнаружителя [1] можно представить в виде

$$\Theta(j\eta|A) = \exp\left[\frac{j\eta(AQ)^2}{1-2j\eta}\right] (1-2j\eta)^{-\mu/2}, \quad (1)$$

где $Q = z/\bar{A}$ — вспомогательный детерминированный параметр, z — параметр обнаружения, равный корню из отношения средней энергии обнаруживаемого сигнала к спектральной плотности мощности шума, $\mu = 2T\Delta f + 1$ — число степеней свободы, T — длительность обработки, Δf — ширина спектра обнаруживаемого сигнала. Чертой сверху обозначено статистическое усреднение по ансамблю реализаций.

Найти безусловную характеристическую функцию $\Theta(j\eta)$ статистики Ξ можно посредством усреднения характеристической функции (1) по случайной амплитуде:

$$\Theta(j\eta) = (1-2j\eta)^{-\mu/2} \int_0^\infty \exp\left[\frac{j\eta(AQ)^2}{1-2j\eta}\right] W(A) dA = (1-2j\eta)^{-B} \vartheta(j\eta), \quad (2)$$

где $W(A)$ — плотность вероятности амплитуды A , $B = T\Delta f$ — база обработки,

$$\vartheta(j\eta) = \int_0^\infty \exp\left[\frac{j\eta(AQ)^2}{1-2j\eta}\right] \frac{W(A) dA}{\sqrt{1-2j\eta}}. \quad (3)$$

Правую часть (2) можно рассматривать как произведение двух характеристических функций: характеристической функции $(1 - 2j\eta)^{-B}$ χ^2 -распределения с $2B$ степенями свободы и характеристической функции $\vartheta(j\eta)$ некоторого распределения, определяемого статистическими свойствами амплитуды A . Следовательно, в случае дискретного энергетического обнаружения квазидетерминированного сигнала со случайной амплитудой и произвольной начальной фазой на фоне белого гауссовского шума решающая статистика имеет вид

$$\Xi|H_1 = \chi_{2B}^2 + \zeta, \quad (4)$$

где H_1 — гипотеза о наличии обнаруживаемого сигнала в обрабатываемом колебании, χ_ν^2 — случайная величина (статистика), имеющая χ^2 -распределение с ν степенями свободы, ζ — случайная величина, имеющая характеристическую функцию (3), причём слагаемые в правой части (4) независимы. Как нетрудно убедиться, (4) обобщает формулу, полученную в [1] для случая рэлеевского распределения флуктуаций амплитуды.

В частности, если флуктуации амплитуды вторичного излучения распределены по закону Райса (П1) (см. Приложение), то можно показать, что

$$\vartheta(j\eta) = (1 - 2j\eta\varsigma)^{-1} \exp\left(\frac{j\eta\lambda\varsigma}{1 - 2j\eta\varsigma}\right) \quad (5)$$

и решающая статистика дискретного энергетического обнаружителя может быть представлена в виде взвешенной суммы двух независимых χ^2 -статистик, одна из которых нецентральная:

$$\Xi|H_1 = \chi_{2B}^2 + \varsigma\chi_{2,\lambda}^2. \quad (6)$$

Здесь $\varsigma = (2z^2 + a^2 + 2)/(a^2 + 2)$ — весовой коэффициент, $\lambda = 2a^2z^2/(2z^2 + a^2 + 2)$ — параметр нецентральности, a — параметр распределения Райса, $\chi_{\nu,\lambda}^2$ — случайная величина, имеющая нецентральное χ^2 -распределение с ν степенями свободы и параметром нецентральности λ .

Если амплитуда вторичного излучения имеет распределение Накагами (П2), то

$$\vartheta(j\eta) = (1 - 2j\eta)^{m-1} [1 - 2j\eta(1 + z^2/m)]^{-m}, \quad (7)$$

где m — параметр распределения Накагами, и

$$\Xi|H_1 = \chi_{2(B+1-m)}^2 + \gamma_{2(1+z^2/m),m}. \quad (8)$$

В (8) $\gamma_{\alpha,A}$ — гамма-статистика с плотностью вероятности $w_\gamma(x; a, A) = 1(x)a^{-A}x^{A-1} \exp(-x/a)/\Gamma(A)$,

$$1(x) = \begin{cases} 1, & x \geq 0; \\ 0, & x < 0 \end{cases}$$

— единичная ступенчатая функция, $\Gamma(x)$ — гамма-функция.

Плотность вероятности $W(x)$ решающей статистики Ξ дискретного энергетического обнаружителя также может быть найдена посредством усреднения условной плотности вероятности

$$W(x|A) = \exp\left(-\frac{(AQ)^2}{2}\right) \sum_{k=0}^{\infty} \frac{(Q^2/2)^k}{k!} A^{2k} w_{\chi^2}[x; 2(k+B+1)], \quad (9)$$

где $w_{\chi^2}(x; \nu)$ — плотность вероятности χ^2 -распределения с ν степенями свободы. В результате усреднения получаем

$$W(x) = \sum_{k=0}^{\infty} \wp_k w_{\chi^2}[x; 2(k+B+1)], \quad (10)$$

где

$$\wp_k = \frac{1}{k! 2^k} \int_0^\infty (AQ)^{2k} \exp[-(AQ)^2/2] W(A) dA. \quad (11)$$

Можно доказать, что коэффициенты $\wp_k$ удовлетворяют условию

$$\sum_{k=0}^{\infty} \wp_k = 1. \quad (12)$$

Нетрудно также убедиться, что все коэффициенты $\wp_k$ положительны вследствие неотрицательности подынтегральной функции в (11). Таким образом, распределение решающей статистики дискретного энергетического обнаружителя при приёме аддитивной смеси квазидетерминированного радиосигнала со случайной амплитудой и белого гауссовского шума в общем случае имеет бесконечномерное гипер- χ^2 -распределение.

Из (11) вытекает формула для функции распределения статистики Ξ :

$$\mathcal{F}(x) = 1(x) \sum_{k=0}^{\infty} \wp_k P(B+k+1, x/2), \quad (13)$$

где $P(x, y)$ — ро-функция (нормированная неполная гамма-функция) [4]. Из (13) можно получить выражение для вероятности правильного обнаружения:

$$P_{\text{ПО}} = \sum_{k=0}^{\infty} \frac{\wp_k}{\Gamma(B+k+1)} \Gamma(B+k+1, h/2), \quad (14)$$

где $\Gamma(x, y)$ — неполная гамма-функция [4], $h = \Lambda(B+1, P_{\text{ЛТ}})$ — пороговый уровень, определяемый в соответствии с критерием Неймана—Пирсона допустимым уровнем вероятности ложной тревоги (ошибки первого рода) $P_{\text{ЛТ}}$, $\Lambda(x, y)$ — функция, обратная по второму параметру к ро-функции $P(x, y)$: $P[x, \Lambda(x, y)] = y$.

Подставляя в (11) плотность вероятности распределения Райса (П1), можно получить

$$\wp_k = \frac{\exp(-\lambda/2)}{\varsigma} \left(\frac{\varsigma-1}{\varsigma} \right)^k L_k \left(-\frac{\lambda}{2(\varsigma-1)} \right), \quad (15)$$

где $L_k(x)$ — полиномы Лагерра, а параметры ς и λ имеют тот же смысл, что и в (5), (6).

Воспользовавшись формулой [5]

$$\sum_{k=0}^{\infty} t^k L_k(x) = (1-t)^{-1} \exp\left(\frac{tx}{t-1}\right), \quad (16)$$

нетрудно доказать, что коэффициенты $\wp_k$ вида (15) удовлетворяют условию (12).

Подставляя в (11) плотность вероятности распределения Накагами (П2), можно получить

$$\wp_k = \frac{\Gamma(k+m)}{\Gamma(m) k!} \left(\frac{m}{m+z^2} \right)^m \left(\frac{z^2}{m+z^2} \right)^k. \quad (17)$$

Воспользовавшись формулой [5]

$$\sum_{k=0}^{\infty} \frac{\Gamma(k+\nu)}{k!} x^k = \Gamma(\nu) (1-x)^{-\nu}, \quad (18)$$

нетрудно убедиться, что коэффициенты (17) также удовлетворяют необходимому условию нормировки (12).

После громоздких преобразований удалось получить формулы для коэффициентов $\wp_k$ и при других законах распределения случайной амплитуды обнаруживаемого сигнала, а именно:

$$\wp_k = \frac{2}{z^2 \sqrt{1 - \alpha^2}} \left[\frac{z^2 (1 - \alpha^2)}{z^2 (1 - \alpha^2) + 2} \right]^{k+1} {}_2F_1 \left(\frac{k+1}{2}, \frac{k}{2} + 1; 1; \frac{4\alpha^2}{[z^2 (1 - \alpha^2) + 2]^2} \right) \quad (19)$$

— для распределения Хойта 3-го типа (ПЗ),

$$\wp_k = \frac{Q^{2k} \beta^\alpha}{(Q^2 + \beta)^{k+\alpha}} \exp \left[-\frac{(Q\gamma)^2}{2\beta(Q^2 + \beta)} \right] L_k^{(\alpha-1)} \left[-\frac{\gamma^2}{2(Q^2 + \beta)} \right] \quad (20)$$

— для распределения Накагами—Райса (П8),

$$\wp_k = \frac{(2p)^{m+l+1} \Gamma(l+1/2) \Gamma(m+1/2) \Gamma(m+l+k+1) Q^{2k}}{2^{k+1} \pi (2l-1)!! (2m-1)!! (m+l)! k! (p+Q^2/2)^{k+m+l+1}} \quad (21)$$

— для бета-экспоненциального распределения (П10),

$$\wp_k = \frac{\beta^{2k} \sigma_y / \sigma_x}{(1 + \beta^2)^{k+1}} \exp \left(-\frac{m_y^2 \beta^2}{2\sigma_y^2 (1 + \beta^2)} - \frac{m_x^2}{2\sigma_x^2} \right) \sum_{l=0}^{\infty} \left(\frac{1 - \sigma_y^2 / \sigma_x^2}{1 + \beta^2} \right)^l \times \\ \times L_l^{(-1/2)} \left[-\frac{m_x^2}{2\sigma_x^2 (1 - \sigma_x^2 / \sigma_y^2)} \right] L_k^{(l)} \left[-\frac{m_y^2}{2\sigma_y^2 (1 + \beta^2)} \right] \quad (22)$$

— для четырёхпараметрического распределения (П12). В (20) и (22) $L_n^{(\lambda)}(z)$ — обобщённые полиномы Лагерра, в (19) ${}_2F_1(a, b; c; z)$ — гипергеометрическая функция Гаусса, в (22) используется обозначение $\beta^2 = z^2 \sigma_y^2 / (\sigma_x^2 + \sigma_y^2 + m_x^2 + m_y^2)$.

В общем случае коэффициенты $\wp_k$ можно выразить через начальные моменты случайной амплитуды A . В самом деле, раскладывая под знаком интеграла в (11) экспоненту в ряд и меняя местами интегрирование и суммирование, нетрудно получить

$$\wp_k = \frac{Q^{2k}}{2^k k!} \sum_{l=0}^{\infty} \frac{(-1)^l}{2^l l!} Q^{2l} \overline{A^{2(k+l)}}. \quad (23)$$

Из (23) можно найти, что

$$\wp_k = \frac{Q^{2k} \exp[2k(\alpha + k\sigma^2)]}{2^k k!} \sum_{l=0}^{\infty} \frac{(-1)^l}{2^l l!} Q^{2l} \exp[2l(\alpha + 2k\sigma^2) + 2\sigma^2 l^2] \quad (24)$$

для логарифмически нормального распределения (П4),

$$\wp_k = \frac{Q^{2k}}{2^k k! c^{2k/\alpha}} \sum_{l=0}^{\infty} \frac{(-1)^l Q^{2l}}{2^l l! c^{2l/\alpha}} \Gamma[1 + 2(k+l)/\alpha] \quad (25)$$

для распределения Вейбулла (П5),

$$\wp_k = \frac{1}{2^k k!} \sum_{l=0}^{\infty} \frac{(-1)^l z^{2(k+l)}}{2^l l! \alpha} \frac{\Gamma(\alpha + k + l) \Gamma(1 + k + l)}{\Gamma(\alpha)} \quad (26)$$

для K -распределения (П6),

$$\wp_k = \frac{1}{k!} \sum_{l=0}^{\infty} \frac{(-1)^l}{l!} (\sigma^2 Q)^{2(k+l)} (1 - \alpha^2)^{(k+l+1)/2} \exp\left(-\frac{a^2}{2} \sqrt{\frac{1-\alpha}{1+\alpha}}\right) \times \\ \times \sum_{n=0}^{\infty} \varepsilon_n \left(\alpha \sqrt{1-\alpha^2}\right)^n \left[\frac{a^2}{2} \left(\frac{1-\alpha}{1+\alpha}\right)\right]^{2n} \times \\ \times \sum_{m=0}^{\infty} \frac{\Gamma(3m+3n+1)}{m! \Gamma(m+n+1)} \alpha^{2m} {}_1F_1\left(k+l+2(n+m)+1; 2n+1; \frac{a^2(1-\alpha)^{3/2}}{2\sqrt{1+\alpha}}\right) \quad (27)$$

для распределения Бекмана (П7),

$$\wp_k = \frac{1}{2^{k/2} k!} \sum_{l=0}^{\infty} \frac{(-1)^l}{2^{l/2} l!} Q^{2(k+l)} \sigma^{k+l} \frac{\Gamma(k+l+\alpha/2)}{\Gamma(\alpha/2)} \frac{D_{-k-l-\alpha/2}[-p/(\sqrt{2}\sigma^2)]}{D_{-\alpha/2}[-p/\sqrt{2}\sigma^2]} \quad (28)$$

для распределения параболического цилиндра (П9),

$$\wp_k = \frac{1}{2^k k!} \sum_{l=0}^{\infty} \frac{(-1)^l}{2^l l!} Q^{2(k+l)} \left[\alpha^{2(k+l)} + \sum_{m=1}^{2(k+l)} \binom{2(k+l)}{m} \exp\left(\frac{m^2 - 2m\eta\gamma}{2\eta^2}\right) \right] \quad (29)$$

для S_L -распределения Джонсона (П11). Здесь ${}_1F_1(a; b; c)$ — вырожденная гипергеометрическая функция Куммера; $D_\nu(z)$ — функции параболического цилиндра;

$$\binom{n}{k} = \frac{n!}{k! (n-k)!}$$

— биномиальный коэффициент.

Таким образом, в статье получены выражения, позволяющие рассчитывать характеристики дискретного энергетического обнаружения квазидетерминированного радиосигнала с произвольной начальной фазой и нерэлеевским распределением флуктуаций амплитуды.

ПРИЛОЖЕНИЕ

Наиболее очевидным обобщением распределения Рэлея является двухпараметрическое распределение Райса, соответствующее физической ситуации, когда одновременно реализуются два механизма отражения зондирующей волны от объекта — зеркальный и диффузный. Плотность вероятности распределения Райса имеет вид

$$W(A) = 1(A) \frac{A}{\omega^2} \exp(-a^2/2) \exp[-A^2/(2\omega^2)] I_0(aA/\omega), \quad (П1)$$

где a и ω — параметры распределения, причём параметр a характеризует соотношение между зеркальной и диффузной компонентами вторичного излучения (если присутствует только диффузная компонента, то $a = 0$, а если только зеркальная, то $a = \infty$), а масштабный параметр ω пропорционален среднеквадратичному значению случайной амплитуды A ; $I_0(x)$ — модифицированная функция Бесселя нулевого порядка.

В том случае, когда вторичное излучение попадает на приёмную антенную систему вследствие многолучевого распространения (например, в ионосфере), в качестве модели флуктуаций амплитуды излучения обычно используют двухпараметрическое распределение Накагами [2, 3], называемое также

m -распределением. Плотность вероятности распределения Накагами имеет вид

$$W(A) = 1(A) \frac{2A^{2m-1}}{\Gamma(m)} \left(\frac{m}{\Upsilon}\right)^m \exp\left(-\frac{mA^2}{\Upsilon}\right), \quad (\text{П2})$$

где $\Upsilon = \overline{A^2}$ и $m = \Upsilon/(\overline{A^4} - \Upsilon^2)$ — параметры.

При $m = 1/2$ распределение Накагами переходит в одностороннее нормальное распределение, а при $m = 1$ — в распределение Рэлея. Формально (при переходе в выражении (П2) к новой переменной $B = A^2$ и замене m на $n/2$, где n — целое положительное число) распределение Накагами может быть сведено и к χ^2 -распределению.

Ещё одной широко распространённой двухпараметрической статистической моделью флуктуаций амплитуды радиосигнала является распределение Хойта 3-го типа, согласно которому

$$W(A) = 1(A) \frac{2A}{\Upsilon \sqrt{1-\alpha^2}} \exp\left[-\frac{A^2}{\Upsilon(1-\alpha^2)}\right] I_0\left[\frac{\alpha A^2}{\Upsilon(1-\alpha^2)}\right], \quad (\text{П3})$$

где α и $\Upsilon = \overline{A^2}$ — параметры. Распределение Хойта 3-го типа описывает флуктуации амплитуды огибающей периодически нестационарного гауссовского вторичного излучения (что может явиться, например, следствием его распространения в случайно-неоднородной среде, являющейся источником мультипликативных помех) [2]. Видно, что при $\alpha = 0$ распределение Хойта (П3) преобразуется в закон Рэлея.

В ряде случаев, в частности когда источниками вторичного излучения являются пространственно протяжённые объекты типа земной или водной поверхности, модели Райса, Хойта и Накагами не соответствуют экспериментальным результатам [2]. Для учёта возможной высокой вероятности больших значений амплитуды излучения используют логарифмически нормальное распределение

$$W(A) = \frac{1(A)}{\sqrt{2\pi} \sigma A} \exp\left[-\frac{(\ln A - \alpha)^2}{2\sigma^2}\right], \quad (\text{П4})$$

где $\sigma^2 = \overline{\ln^2 A} - \alpha^2$ и $\alpha = \overline{\ln A}$ — дисперсия и среднее значение логарифма амплитуды соответственно. Логарифмически нормальное распределение нашло широкое применение при аппроксимации результатов экспериментальных исследований вторичного излучения космических аппаратов и статистически неровных поверхностей [2].

Наряду с логарифмически нормальным распределением в качестве модели флуктуаций амплитуды волны, отражённой от статистически неровной поверхности (и, в частности, от морской поверхности) используют распределение Вейбулла

$$W(A) = 1(A) c \alpha A^{\alpha-1} \exp(-cA^\alpha), \quad (\text{П5})$$

где $\alpha > 0$ и $c > 0$ — параметры. Это распределение занимает промежуточное положение между рэлеевским и логарифмически нормальным распределениями. При $\alpha = 2$ распределение Вейбулла переходит в распределение Рэлея, при $\alpha > 2$ изменяется асимметрия распределения, что в ряде случаев соответствует реальным условиям отражения сигналов от умеренно взволнованной (до 3 баллов) морской поверхности [2].

В последнее время для описания флуктуаций амплитуды волн, отражённых статистически неровной поверхностью, всё чаще используют так называемое K -распределение [2]

$$W(A) = 1(A) \frac{4A^\alpha (\alpha/\Upsilon)^{(\alpha+1)/2}}{\Gamma(\alpha)} K_{\alpha-1}\left(2A \sqrt{\alpha/\Upsilon}\right), \quad (\text{П6})$$

где $K_\nu(z)$ — функция Макдональда (модифицированная функция Бесселя третьего рода).

Двухпараметрические модели, обладая всего одним параметром формы и одним параметром масштаба, не могут учесть всего разнообразия свойств амплитуды вторичного радиоизлучения [2]. Более гибкими являются трёхпараметрические модели, включающие в себя более простые двухпараметрические распределения как частные случаи.

Трёхпараметрическое распределение Бекмана (Райса—Хойта)

$$W(A) = 1(A) \frac{A}{\sigma^2} \exp\left(-\frac{a^2(1-\alpha) + A^2/\sigma^4}{2\sqrt{1-\alpha^2}}\right) \sum_{k=0}^{\infty} \varepsilon_k I_k\left(\frac{\alpha A^2/\sigma^4}{2\sqrt{1-\alpha^2}}\right) I_{2k}\left(\frac{Aa}{\sigma^2} \sqrt{\frac{1-\alpha}{1+\alpha}}\right), \quad (\text{П7})$$

где $\varepsilon_0 = 1$, $\varepsilon_n = 2$ при $n > 1$, было получено в качестве модели флуктуаций амплитуды волны, рассеянной статистически шероховатой поверхностью с высотой неровностей, превышающей длину волны [2]. Здесь a , α и σ^2 — параметры распределения.

Трёхпараметрическое распределение Накагами—Райса

$$W(A) = 1(A) \frac{\beta^\alpha A^\alpha}{\gamma^{\alpha-1}} \exp\left(-\frac{\gamma^2}{2\beta} - \frac{\beta A^2}{2}\right) I_{\alpha-1}(\gamma A) \quad (\text{П8})$$

обладает значительно большей гибкостью по сравнению с распределениями Райса и Накагами (П1), (П2) и включает в себя многие одно- и двухпараметрические модели флуктуаций амплитуды как частные случаи. Здесь α , β и γ — параметры распределения.

Практически такой же общностью [2], что и распределение Накагами—Райса, обладает трёхпараметрическое распределение параболического цилиндра

$$W(A) = 1(A) \frac{A^{\alpha-1} \exp\left(-\frac{A^4 - 2pA^2}{4\sigma^4}\right)}{D_{-\alpha/2}[-p/(\sqrt{2}\sigma^2)] 2^{\alpha/4-1} \sigma^\alpha \Gamma(\alpha/2) \exp[p^2/(8\sigma^4)]}, \quad (\text{П9})$$

где $D_\nu(z)$ — функции параболического цилиндра, α , p и σ^2 — параметры распределения.

Трёхпараметрическая модель бета-экспоненциального распределения

$$W(A) = 1(A) \frac{(2p)^{(b+1)} \Gamma(l+1/2) \Gamma(m+1/2)}{\pi (2l-1)!! (2m-1)!! b!} A^{2b+1} \exp(-pA^2), \quad (\text{П10})$$

где $b = m + l \geq 0$, достаточно близка к распределению Накагами, однако в отличие от последнего она позволяет описывать более глубокие флуктуации амплитуды вторичного излучения [2]. Здесь p , m и l — параметры распределения.

В миллиметровом диапазоне длин волн флуктуации амплитуды вторичного излучения земной или водной поверхности иногда описывают при помощи трёхпараметрической логарифмически нормальной модели

$$W(A) = 1(A) \frac{\eta}{(A-\alpha)\sqrt{2\pi}} \exp\left\{-\frac{\eta^2}{2} \left[\frac{\gamma}{\eta} + \ln(A-\alpha)\right]^2\right\}, \quad (\text{П11})$$

где $A \geq \alpha$, называемой S_L -распределением Джонсона [6]. Здесь η , α и γ — параметры распределения.

Из всего многообразия четырёхпараметрических моделей можно отметить распределение [3]

$$W(A) = 1(A) \frac{A}{\sigma_x \sigma_y} \exp\left(-\frac{m_y^2 + A^2}{2\sigma_y^2} - \frac{m_x^2}{2\sigma_x^2}\right) \sum_{k=0}^{\infty} \frac{H_{2k}(\alpha)}{(2k)!! 2^k} A^k \left(\frac{\sigma_y^2}{m_y}\right)^k \left(\frac{1}{\sigma_x^2} - \frac{1}{\sigma_y^2}\right)^k I_k\left(\frac{A m_y}{\sigma_y^2}\right), \quad (\text{П12})$$

где σ_x , σ_y , m_x и m_y — четыре параметра распределения, $\sigma_x^2 \leq \sigma_y^2$, $H_i(x)$ — полиномы Эрмита и используется обозначение [3] $\alpha = m_x / \left(\sqrt{2} \sigma_x^2 \sqrt{1/\sigma_x^2 - 1/\sigma_y^2} \right)$. Четырёхпараметрическое распределение (П12) переходит в трёхпараметрическое распределение Бекмана при $m_x = 0$, в распределение Райса — при $\sigma_x = \sigma_y = \sigma$, в распределение Хойта — при $m_x = m_y = 0$, в распределение Рэлея — при $\sigma_x = \sigma_y = \sigma$ и $m_x = m_y = 0$, в одностороннее нормальное распределение — при $m_x = m_y = 0$ и $\sigma_x \rightarrow \infty$ и т. д. [3].

Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект № 01-01-00356).

СПИСОК ЛИТЕРАТУРЫ

1. Костылев В. И. // Изв. вузов. Радиофизика. 2000. Т. 43, № 10. С. 926.
2. Шляхин В. М. // Радиотехника и электроника. 1987. Т. 32, № 9. С. 1793.
3. Кловский Д. Д. Передача дискретных сообщений по радиоканалам. М.: Радио и связь, 1982. 304 с.
4. Справочник по специальным функциям с формулами, графиками и математическими таблицами / Под ред. М. Абрамовица, И. Стиган. М.: Наука, 1979. 832 с.
5. Прудников А. П., Брычков Ю. А., Маричев О. И. Интегралы и ряды. Специальные функции. М.: Наука, 1983. 752 с.
6. Кендалл М., Стьюарт А. Теория распределений. М.: Наука, 1966. 587 с.

Воронежский госуниверситет,
г. Воронеж, Россия

Поступила в редакцию
18 июля 2001 г.

CHARACTERISTICS OF THE ENERGY DETECTION OF QUASI-DETERMINISTIC RF SIGNALS WITH NON-RAYLEIGH DISTRIBUTION OF THE AMPLITUDE FLUCTUATIONS

V. I. Kostylev

We determine the distribution of the decision statistic of a discrete energy detector receiving an additive mixture of a white noise and a quasideterministic signal with a non-Rayleigh distribution of amplitude fluctuations. An expression for the probability of the correct detection is derived.