МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Известия высших учебных заведений

РАДИОФИЗИКА

ежемесячный научно-технический журнал

Издаётся с апреля 1958 г.

Том XLVIII № 7 **Нижний Новгород** **2005**

Содержание

УДК 524+520.27

ОБЗОР ОБЛАСТЕЙ ЗВЕЗДООБРАЗОВАНИЯ, СВЯЗАННЫХ С ИНФРАКРАСНЫМИ ИСТОЧНИКАМИ, В ЛИНИИ $J = 1$—0 МОЛЕКУЛЫ CO И ЕЁ ИЗОТОПОВ. РЕЗУЛЬТАТЫ НАБЛЮДЕНИЙ В ЛИНИИ $C^{18}O$ ($J = 1$—0)

Н. Р. Троицкий[1], *Л. Е. Пирогов*[1], *И. И. Зинченко*[1], *Дж. Янг*[2]

[1] Институт прикладной физики РАН, г. Нижний Новгород, Россия;
[2] Purple Mountain Observatory, Nanjing, China

Представлены результаты наблюдений 23 плотных ядер молекулярных облаков, связанных с «холодными» инфракрасными источниками IRAS, в линии $J = 1$—0 молекулы $C^{18}O$. Наблюдения проводились на 13,7-метровом радиотелескопе обсерватории Пурпурная Гора (Purple Mountain), Китай. Линия $C^{18}O$ ($J = 1$—0) зарегистрирована в излучении 21 объекта. В приближении локального термодинамического равновесия проведены оценки лучевых концентраций молекул $C^{18}O$ и H_2 в направлении максимума интегральной интенсивности излучения $C^{18}O$, которые составили $(2{,}5\div10{,}4)\cdot10^{15}$ см$^{-2}$ и $(1{,}5\div6{,}1)\cdot10^{22}$ см$^{-2}$ соответственно. Кинетические температуры, определённые по линиям CO для этих максимумов, варьируются от 14 до 45 K. Для шести объектов, для которых карты можно считать практически законченными, оценены размеры областей излучения $C^{18}O$, лежащие в диапазоне от 0,5 до 1,2 пк. Массы этих объектов лежат в диапазоне $(390\div1\,750)\,M_\odot$ и близки к оценкам, следующим из теоремы вириала. Диапазон средних плотностей объектов составил $(0{,}3\div1{,}4)\cdot10^4$ см$^{-3}$.

ВВЕДЕНИЕ

Спектральные исследования плотных молекулярных облаков, где по современным данным происходит процесс звездообразования, в последние годы всё чаще включают в себя большие массивы из десятков и даже сотен объектов, отобранных в соответствии с определёнными критериями, что должно способствовать выявлению общих закономерностей процесса образования звёзд и особенностей его различных стадий. Одной из важных задач таких систематических исследований, в частности, является изучение вариаций физических параметров звездообразующих плотных ядер в зависимости от их положения в Галактике.

Большая часть систематических спектральных исследований областей звездообразования опирается на данные обзоров инфракрасных (ИК) источников и водяных мазеров, являющихся индикаторами ранних стадий эволюции звёзд. Так, недавно были опубликованы данные обзора в линии $J = 1$—0 молекулы CO обширной выборки, основанной на каталоге ИК источников IRAS [1]. Данный обзор включал в себя 1 912 «холодных» ИК источников, удовлетворяющих критериям $\log(F_{12}/F_{25}) \le -0{,}4$ и $\log(F_{25}/F_{60}) \le -0{,}4$ или $\log(F_{12}/F_{60}) \le -0{,}4$, где F_{12}, F_{25} и F_{60} — потоки излучения на длине волны 12, 25 и 60 мкм соответственно. Данный критерий позволяет связать эти объекты с областями звездообразования на ранних стадиях развития. Для своего обзора авторы отбирали источники, расположенные в северной полусфере недалеко от галактической плоскости, где сосредоточена основная масса областей звездообразования, за исключением внутренней области Галактики, где идентификация может быть неоднозначной: $\alpha(1950) \le 8^{\mathrm{h}}$, $\alpha(1950) \ge 16^{\mathrm{h}}$, $\delta(1950) \ge -35^\circ$, $|b| \le 25^\circ$ и $|b| \ge 1^\circ$ при $l \le 60^\circ$ или $l \ge 300^\circ$; здесь $\alpha(1950)$ и $\delta(1950)$ — прямое восхождение и склонение источника для эпохи 1950, b и l — галактические широта и долгота соответственно. Кроме этого, отбирались источники, потоки излучения которых детектировались, как минимум, в трёх диапазонах IRAS, не связанные со звёздами поздних спектральных классов, планетарными туманностями и внегалактическими источниками. Большинство источников, вошедших в обзор, ранее не наблюдались в молекулярных линиях, также отсутствуют сведения об ассоциации их с какими-либо оптическими или радиообъектами. В результате данного обзора в 1 331 источнике было обнаружено излучение в линии $J = 1$—0 молекулы CO, также были проведены исследования особенностей профилей этой линии с целью обнаружения высокоскоростных потоков, являющихся индикаторами ранних стадий эволюции звезды. Картирование объектов не проводилось.

Немалый интерес, однако, представляет поиск областей с повышенной плотностью (плотных ядер) среди объектов, обнаруженных в обзоре [1], с целью оценки их физических параметров и последующего статистического анализа распределений этих параметров в Галактике. Для этого были отобраны источники, у которых антенная температура в линии CO превышает 10 K, что может указывать на существование областей тёплого газа с высокой плотностью (более 10^3 см$^{-3}$). Количество таких объектов составило 57 (считалось, что источники 05375+3536 и 05375+3540, а также 06055+2034 и 06055+2039 принадлежат единым молекулярным комплексам).

В данной работе приведены результаты наблюдений 22 объектов из 57 выбранных нами, а также источника 00338+6312 в линии $J = 1$—0 молекулы $C^{18}O$. Наличие излучения в указанной линии может служить индикатором плотного газа. Это позволило в нескольких случаях определить размеры ядер, их массу, лучевую концентрацию и среднюю плотность. Одновременно проводились наблюдения тех же объектов в линии $J = 1$—0 молекул ^{13}CO и ^{12}CO. Наблюдения линий молекулы ^{12}CO необходимы, в частности, для оценки кинетической температуры, а данные наблюдений излучения молекулы ^{13}CO могут быть использованы для исследования внутренней кинематики объектов, включая высокоскоростные выбросы газа, а также для моделирования источников (совместно с наблюдениями в линии $J = 1$—0 молекул ^{12}CO и $C^{18}O$). Одновременные наблюдения всех трёх линий позволяют существенно сократить время измерений и исключить погрешности, связанные с ошибками наведения телескопа, при сравнении параметров линий, относящихся к одной и той же точке облака. В настоящей статье приведены пиковые температуры излучения в линии $J = 1$—0 молекулы ^{12}CO в направлении максимума излучения молекулы $C^{18}O$. Более подробные результаты обработки данных наблюдений излучения молекул ^{13}CO и ^{12}CO будут представлены в последующих публикациях. Среди объектов выборки есть ранее исследовавшиеся в различных молекулярных линиях, однако для получения однородного массива данных эти источники не были исключены.

Поскольку наблюдения были осложнены рядом технических проблем, проявившихся в существенных погрешностях наведения телескопа (см. раздел 1), результаты, представленные в данной работе, следует рассматривать как предварительные.

1. ПРОЦЕДУРА НАБЛЮДЕНИЙ

Наблюдения проводились в апреле–мае 2004 года на 13,7-метровом радиотелескопе обсерватории Пурпурная Гора (Purple Mountain), Китай. Наблюдения велись одновременно в линии $J = 1$—0 молекул $C^{18}O$, ^{13}CO и ^{12}CO на частотах 109,8; 110,2 и 115,3 ГГц. При наблюдениях применялся охлаждаемый СИС-приёмник. Шумовая температура системы в двухполосном режиме изменялась в пределах от 250 до 470 K в зависимости от высоты источника и погодных условий. Для спектрального анализа принимаемого излучения использовались 3 акустооптических анализатора спектра (AOC), каждый из которых имел по 1 024 канала. Для наблюдений в линии CO использовалась верхняя полоса приёма и AOC с полосой 146 МГц и спектральным разрешением 0,37 км/с. Для наблюдений в линиях изотопов $C^{18}O$ и ^{13}CO использовалась нижняя полоса и 2 раздельных AOC, каждый из которых имел полосу 43 МГц и спектральное разрешение 0,12 км/с.

Предварительные ошибки наведения телескопа, оценённые из наблюдений Венеры и Юпитера в континууме, составили около $20''$. Однако поскольку многие источники расположены вне плоскости эклиптики, для контроля наведения каждые два часа наблюдался один из близких к исследуемому калибровочных источников (L134N, Orion A, S140, W51D, NGC2264, W3(OH) и IRC+10216). Как было выяснено, измеренные интенсивности линий в этих источниках подвержены значительным вариациям, что могло быть связано с проблемами в системах наведения и облучения телескопа, а также с погрешностями калибровки. По этой же причине ширина основного лепестка диаграммы направленности, определённая по наблюдениям Юпитера, представляла собой эллипс с главными осями около $60''$ и $110''$ на половинном уровне мощности на частоте 109,8 ГГц.

По результатам наблюдений калибровочных источников было установлено, что измеренные интенсивности линий систематически зависят от высоты, причём наиболее явной эта зависимость была для компактного и симметричного источника IRC+10216 [2]. На рис. 1 показана зависимость интегральной интенсивности линии молекулы CO от высоты для источника IRC+10216, а также аппроксимация вида $a\sin(bH+c)$, где H — высота источника над горизонтом в градусах, а коэффициенты a, b и c, рассчитанные методом наименьших квадратов, равны 52,8 К·км/с; 0,03 и 0,33 соответственно. Данная зависимость была использована для корректировки интенсивностей линии молекулы $C^{18}O$, поскольку области излучения в этой линии достаточно компактны. Для более протяжённых областей излучения в линиях CO и ^{13}CO зависимость интенсивности от высоты не учитывалась.

Как видно на рис. 1, наблюдается существенный разброс интегральной интенсивности для одной высоты, который и определяет погрешности измерений. Среднеквадратичное отклонение измеренной интегральной интенсивности от аппроксимирующей кривой составляет 9,7 К·км/с.

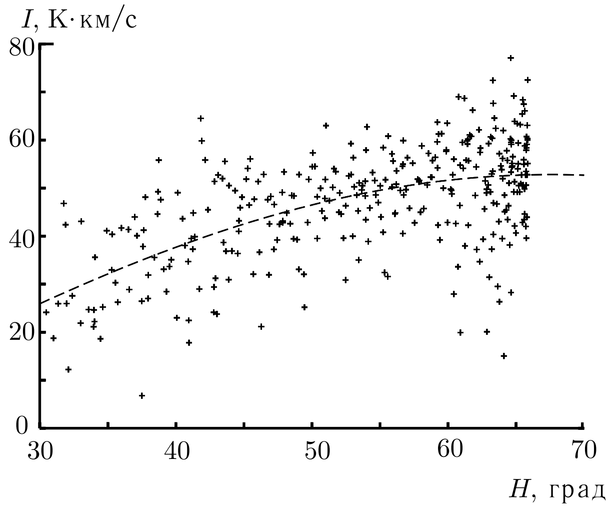


Рис. 1. Зависимость интегральной интенсивности линии $J=1$—0 молекулы CO от высоты над горизонтом для источника IRC+10216 и аппроксимирующая кривая (пунктир)

Из сравнения пиковых антенных температур излучения в линиях для источников IRC+10216 и Orion A с температурами тех же источников, полученными на 12-метровом телескопе NRAO, был определён коэффициент использования диаграммы направленности телескопа η_{mb}. Для линии молекулы $C^{18}O$ объекта Orion A максимальный коэффициент использования диаграммы направленности (соответствующий экстраполяции данных до высот, близких к 65°) $\eta_{\mathrm{mb}} \approx 0{,}5$. Оценки, сделанные по линии излучения ^{12}CO объектов Orion A и IRC+10216, оказались несколько выше. В работе [1] приводится значение $\eta_{\mathrm{mb}} = 0{,}44$ (авторы вели расчёт по Луне). Мы приняли значение $\eta_{\mathrm{mb}} = 0{,}5 \pm 0{,}1$ для всех трёх линий. Погрешность определялась исходя из разброса интегральной интенсивности калибровочных источников. В дальнейшем все температурные значения пересчитывались согласно формуле $T_{\mathrm{mb}} = T_{\mathrm{a}}/\eta_{\mathrm{mb}}$, где T_{a} — антенная температура источника.

Наблюдения шли по следующей методике: в течение 30 с наблюдалась область вне источника, затем в течение одной минуты — источник, затем в течение 30 с снова наблюдался фоновый сигнал. Разность между сигналом от источника и фоном давала величину полезного сигнала. Опрос АОС происходил раз в секунду, и значение усреднялось за время одного цикла. В связи с вышеописанными ошибками наведения было решено не увеличивать время накопления, а получить по возможности полные карты. Уточнение параметров ядер планируется провести в дальнейшем после технической модернизации телескопа.

Список источников, включающий их координаты и кинематические расстояния, взятые из работ [1, 3], приведён в табл. 1. Для каждого из источников проводилось картирование с шагом, равным одной угловой минуте.

Таблица 1. Список источников

Источник IRAS	α(1950), ч:мин:с	δ(1950), град:угл. мин:угл. с	$D_{\rm k}$, кпк
00338+6312	00:33:53,3	+63:12:32	1,14 [2]
01202+6133	01:20:15,2	+61:33:10	0,91 [1]
02531+6032	02:53:08,1	+60:32:19	4,22 [1]
03236+5836	03:23:39,0	+58:36:33	0,69 [1]
05327–0457	05:32:46,4	–04:57:38	0,85 [1]
05327–0529	05:32:42,6	–05:29:47	0,60 [1]
05379+3550	05:37:58,7	+35:50:38	1,6 [2]
05391–0152	05:39:07,1	–01:52:45	0,73 [1]
05391–0217	05:39:06,8	–02:17:18	0,73 [1]
05435–0011	05:43:34,2	–00:11:08	0,84 [1]
05450+0019	05:45:00,5	+00:19:08	0,83 [1]
06055+2039	06:05:33,9	+30:39:47	1,52 [1]
06068+2030	06:06:53,0	+20:30:41	1,39 [1]
06099+1800	06:09:57,9	+18:00:12	0,85 [1]
07028–1100	07:02:51,9	–11:00:05	1,0 [1]
18134–1942	18:13:24,6	–19:42:25	1,8 [2]
19442+2427	19:44:13,5	+24:28:00	2,3 [2]
19446+2505	19:44:41,4	+25:05:17	2,0 [2]
21418+6552	21:41:53,2	+65:52:42	1,02 [1]
22543+6145	22:54:20,2	+61:45:55	0,98 [1]
23030+5958	23:03:04,9	+59:58:28	4,38 [1]
23133+6050	23:13:21,5	+60:50:47	5,90 [1]
23140+6121	23:14:01,9	+61:21:22	4,71 [1]

2. РЕЗУЛЬТАТЫ НАБЛЮДЕНИЙ

Излучение в линии молекулы $\rm C^{18}O$ зарегистрировано в 21 объекте из 23. Линии молекул CO и $\rm ^{13}CO$ обнаружены во всех источниках. В табл. 2 приведены результаты обработки наблюдений в линии молекулы $\rm C^{18}O$ для точек с максимальной интегральной интенсивностью (пиков плотности). В колонке 2 табл. 2 указаны относительные координаты этих пиков.

Обработка полученных спектров осуществлялась с помощью пакета GILDAS (Grenoble Image and Line Data Analysis System) [4] и включала в себя расчёт и вычитание базовой линии (полиномы не выше 3-й степени) и аппроксимацию функцией Гаусса, в результате чего определялись интегральная интенсивность $\int T_{\rm mb}\,{\rm d}V$ (колонка 3 табл. 2), максимальная температура $T_{\rm mb}$ (колонка 4 табл. 2), ширина линии на уровне половины максимума ΔV и скорость в локальной системе

Таблица 2. Результаты обработки наблюдений в линии $J = 1$—0 молекулы $C^{18}O$ и лучевые концентрации молекул

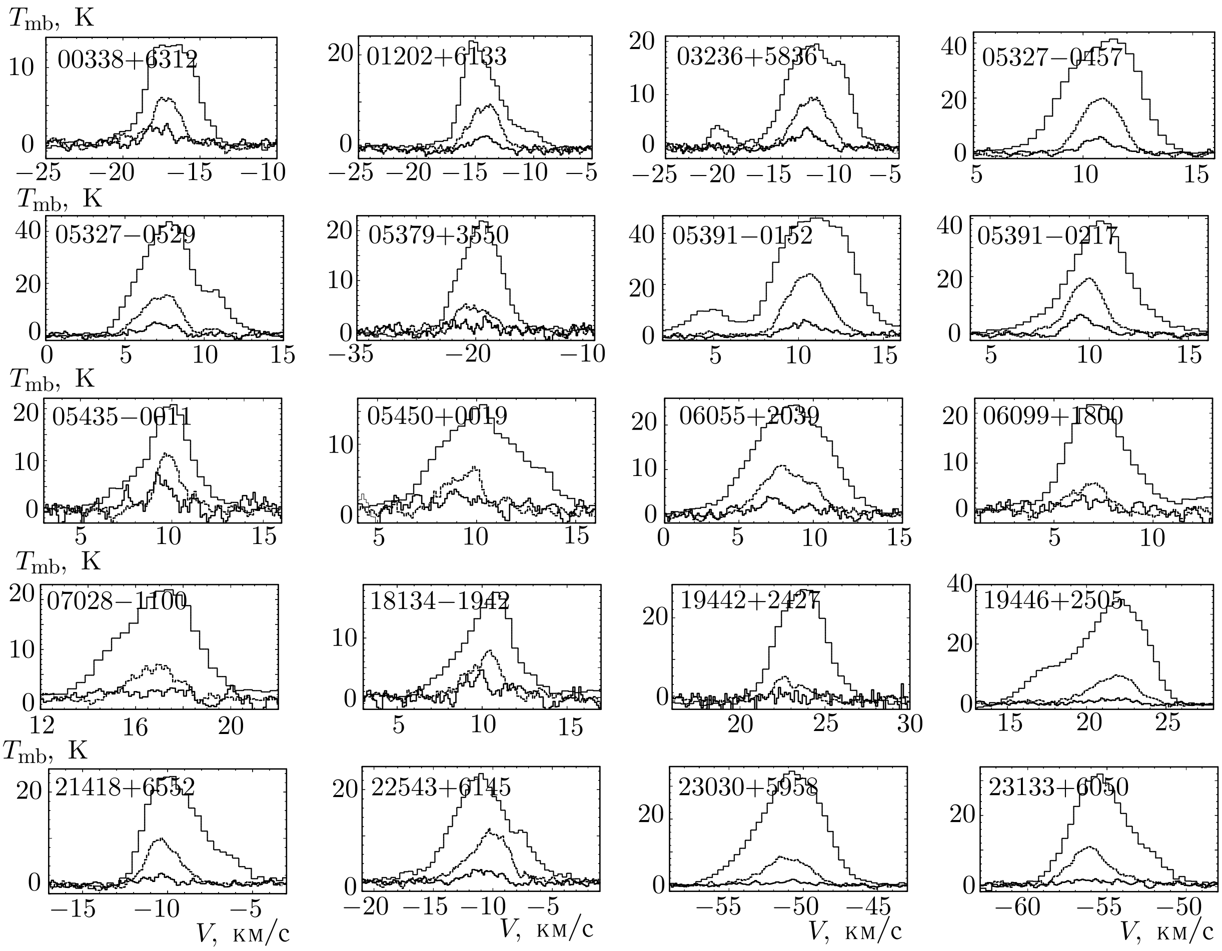
Источник IRAS	$(\Delta\alpha, \Delta\delta)$, (угл. мин, угл. мин)	$\int T_{\rm mb}\,{\rm d}V$, К · км/с	$T_{\rm mb}$, К	ΔV, км/с	$V_{\rm lsr}$, км/с	$T_{\rm mb}^{\rm CO}$, К	$N_{\rm L}(C^{18}O)$, 10^{15} см$^{-2}$	$N_{\rm L}(H_2)$, 10^{22} см$^{-2}$
00338+6312	(–1, –1)	4,5(0,3)	2,1(0,1)	2,0(0,2)	–17,6(0,1)	14	3,4	2,0
01202+6133	(0, –3)	5,9(0,4)	2,7(0,1)	2,1(0,1)	–14,3(0,1)	22	4,4	2,6
02531+6032	(–4, –3)	<1,4	—	—	—	28	—	—
03236+5836	(–4, –4)	8,3(0,4)	3,1(0,1)	2,5(0,1)	–13,0(0,1)	20	6,2	3,7
05327–0457	(0, –6)	9,4(0,5)	5,2(0,1)	1,7(0,1)	10,70(0,03)	42	7,1	4,1
05327–0529	(–1, –7)	11,2(0,7)	4,5(0,2)	2,3(0,2)	7,2(0,1)	44	8,4	4,9
05379+3550	(3, –1)	4,9(0,7)	1,2(0,2)	3,7(0,4)	–20,3(0,3)	22	3,7	2,2
05391–0152	(0, –6)	12,4(0,7)	4,8(0,1)	2,4(0,2)	10,4(0,1)	45	9,3	5,5
05391–0217	(–1, –3)	10,6(0,4)	6,3(0,2)	1,6(0,1)	9,66(0,03)	39	8,0	4,7
05435–0011	(–3, 4)	13,9(1,6)	4,9(0,3)	2,6(0,4)	9,3(0,1)	21	10,4	6,1
05450+0019	(3, 1)	8,0(1,1)	1,8(0,1)	4,1(0,7)	8,7(0,3)	15	6,0	3,5
06055+2039	(1, –4)	10,9(0,8)	2,5(0,1)	4,1(0,4)	7,5(0,2)	25	8,2	4,8
06068+2030	(1, –2)	<1,4	—	—	—	22	—	—
06099+1800	(–3, –5)	11,7(1,6)	1,9(0,1)	5,7(0,9)	6,2(0,4)	21	8,8	5,2
07028–1100	(–4, –1)	7,1(0,9)	1,7(0,1)	3,9(0,5)	16,0(0,3)	21	5,3	3,1
18134–1942	(–2, 2)	8,5(1,0)	2,6(0,2)	3,0(0,4)	10,2(0,2)	17	6,4	3,8
19442+2427	(2, 2)	6,5(1,6)	1,2(0,2)	5,1(1,5)	23,2(0,6)	26	4,9	2,9
19446+2505	(1, 1)	3,9(0,4)	1,4(0,1)	2,6(0,2)	21,6(0,1)	35	2,9	1,7
21418+6552	(1, –5)	4,4(0,4)	1,6(0,1)	2,5(0,3)	–10,6(0,1)	23	3,3	1,9
22543+6145	(–2, 0)	7,7(0,4)	2,6(0,1)	2,8(0,2)	–10,9(0,1)	23	5,8	3,4
23030+5958	(–2, –1)	3,3(0,4)	1,0(0,1)	3,2(0,5)	–51,1(0,2)	36	2,5	1,5
23133+6050	(–2, –3)	5,2(0,5)	1,3(0,1)	3,8(0,4)	–55,8(0,2)	33	3,9	2,3
23140+6121	(–1, –2)	3,8(0,6)	0,7(0,1)	5,1(0,7)	–52,1(0,4)	23	2,9	1,7

координат $V_{\rm lsr}$ (колонки 5 и 6 табл. 2 соответственно). В скобках приведены ошибки аппроксимации, которые определялись с учётом дисперсии шумов в каналах вне диапазона линии. Однако следует учесть, что для пиковых и интегральных интенсивностей линий реальные погрешности определяются в основном ошибками наведения (около 20 %, см. раздел 1). Для удобства сравнения линий, измеренных на различных частотах, частоты каналов пересчитывались в значения скорости с учётом того, что скорость, соответствующая центральному каналу полосы анализа, равна лучевой скорости объекта. Интегральные интенсивности рассчитывались путём суммирования температур в каналах по всему профилю линии. В колонке 7 табл. 2 приводятся пиковые температуры в линии молекулы ^{12}CO для указанных позиций на карте (14÷45 K), которые могут быть приняты в качестве оценки кинетической температуры.

На рис. 2 приведены примеры полученных спектров для 21 объекта, где было зарегистрировано излучение в линии молекулы $C^{18}O$. Спектры соответствуют точкам с максимальной интегральной интенсивностью. Для двух объектов, где линии не были зарегистрированы, приведены верхние пределы интегральной интенсивности (3σ).

Предполагая, что оптическая толщина в линии излучения молекулы $C^{18}O$ мала, в предположении локального термодинамического равновесия (ЛТР) можно сделать оценки концентрации исследуемых молекул на луче зрения по следующей формуле (см., например, [5]):

$$N_{\rm L}[\text{см}^{-2}] = \frac{3k\sum g_i \exp[-E_i/(kT_{\rm EX})]}{8\pi^3\nu_{\rm ul}\mu^2 J_u \exp[-E_u/(kT_{\rm EX})]\,[1 - I(T_{\rm BG})/I(T_{\rm EX})]}\int T_{\rm mb}\,{\rm d}V,$$



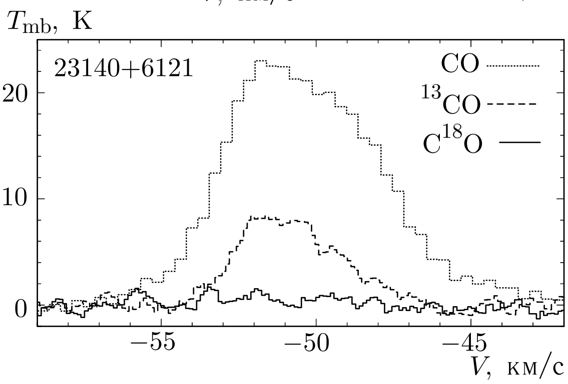


Рис. 2. Спектры линий молекул CO, ^{13}CO и C^{18}O для 21 объекта, в излучении которых зарегистрирована линия $J = 1$—0 молекулы C^{18}O. Спектры соответствуют точкам с максимальной интегральной интенсивностью в линии молекулы C^{18}O. При наличии нескольких локальных пиков на картах приведённые спектры соответствуют тому из пиков, у которого интегральная интенсивность излучения в линии молекулы C^{18}O выше. Стандартное отклонение для шумов вне линии составляет 0,6÷1,8 K для различных источников

где k — постоянная Больцмана, μ — дипольный момент молекулы, $I(T_{\rm BG})$ и $I(T_{\rm EX})$ — значения функции Планка для температуры фона, принятой равной 2,8 K, и температуры возбуждения уровня соответственно, $\nu_{\rm ul}$ — частота перехода с верхнего уровня u на нижний уровень l, E_i и g_i — энергия и статистический вес вращательного уровня с квантовым числом полного углового момента J_i.

Использование предположения о ЛТР в данной работе обосновано тем, что молекула ^{12}CO термализуется уже при плотностях порядка 10^3 см$^{-3}$ благодаря малому дипольному моменту. Хотя в реальных объектах условие ЛТР для линии C^{18}O ($J = 1$—0) может не выполняться, основные погрешности при использовании формулы расчёта лучевых концентраций связаны, в первую очередь, с неопределённостью температуры возбуждения для переходов между наиболее заселёнными нижними уровнями. При расчётах $N_{\rm L}$ было принято, что температура возбуждения для линий молекулы $C^{18}O$ равна 10 К для всех источников. При этой температуре возбуждения рассчитанные лучевые концентрации $C^{18}O$ близки к минимуму при фиксированной интегральной интенсивности. Для $T_{\rm EX} = 30$ К лучевые концентрации возрастают примерно в два раза. Зная относительное содержание молекул $C^{18}O$: $X(C^{18}{\rm O}) = 1{,}7 \cdot 10^{-7}$ [6], можно рассчитать лучевые концентрации водорода $N_{\rm L}({\rm H_2}) = N_{\rm L}({\rm C^{18}O})/X({\rm C^{18}O})$, откуда, зная размер объекта, можно получить оценку его массы. Кроме того, лучевую концентрацию водорода можно использовать в дальнейшем для оценки относительных концентраций других молекул в данных объектах.

В двух последних колонках табл. 2 приведены лучевые концентрации молекул $C^{18}O$ и H_2, которые лежат в диапазонах $(2{,}5{\div}10{,}4) \cdot 10^{15}$ см$^{-2}$ и $(1{,}5{\div}6{,}1) \cdot 10^{22}$ см$^{-2}$ соответственно.

3. ФИЗИЧЕСКИЕ ПАРАМЕТРЫ ИСТОЧНИКОВ И ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Из 21 объекта, в излучении которых обнаружена линия $J = 1$—0 молекулы $C^{18}O$, в шести случаях зарегистрировано спадание интенсивности по всем направлениям, что позволило оценить размеры областей излучения. На рис. 3 приведены их карты.

Для оценки размеров областей излучения молекулы $C^{18}O$ рассчитывалась суммарная площадь всех пикселей на карте (S), в которых интегральная интенсивность излучения в линии молекулы $C^{18}O$ не меньше половины максимальной интенсивности по источнику. Радиусы областей излучения ($R = \sqrt{S/\pi}$) варьируются в пределах 0,5÷1,2 пк (колонка 3 табл. 3). Соответствующие им радиусы в угловых единицах составили $3{,}2'{\div}8{,}2'$ (колонка 2 табл. 3).

Массы объектов рассчитывались по формуле

$$M_{\rm LTE} = 1{,}4 m_{\rm H_2} \sum N_{\rm L}^i({\rm H_2})\, \delta S,$$

где $m_{\rm H_2}$ — масса молекулы H_2, $N_{\rm L}^i({\rm H_2})$ — концентрация молекул H_2 на луче зрения для точек, в которых интегральная интенсивность излучения в линии молекулы $C^{18}O$ больше или равна половине максимальной, δS — площадь одного пикселя при картировании. Числовой коэффициент в данной формуле необходим для учёта как водорода, так и гелия, который тоже даёт весомый вклад в массу облака. Массой остальных элементов можно пренебречь в силу её малости по сравнению с массой водорода и гелия. Рассчитанные значения масс лежат в пределах 390÷1 750 масс Солнца (колонка 4 табл. 3).

В колонке 5 табл. 3 приведены массы, следующие из стандартной формулировки теоремы вириала в предположении, что сумма магнитного давления и газокинетического давления внешней среды пренебрежимо мала, и рассчитанные по формуле $M_{\rm VIR} = 5R\,\overline{\Delta V}^2\,(8G\ln 2)^{-1}$ (см., например, [7]), где G — гравитационная постоянная, $\overline{\Delta V}$ — средняя по источнику ширина линии $C^{18}O$. Последняя рассчитывалась как взвешенное среднее с весами, обратно пропорциональными квадрату ошибки гауссовой аппроксимации, по области, где интегральная интенсивность линий $C^{18}O$ больше или равна половине максимальной. Значения $\overline{\Delta V}$ для шести рассмотренных источников варьируются в пределах от 1,7 до 2,7 км/с.

Диапазон изменения вириальных масс составил 390÷1 100 масс Солнца. Отношение $M_{\rm LTE}/M_{\rm VIR}$ изменяется в пределах 0,5÷3,0. Поскольку погрешности определения масс, связанные в основном с ошибками калибровки и с погрешностями определения расстояний, могут достигать 50÷100 %, а погрешности определения $M_{\rm LTE}$ превышают погрешности $M_{\rm VIR}$ (массы по-разному зависят от расстояния: квадратично в случае $M_{\rm LTE}$ и линейно для $M_{\rm VIR}$, к тому же $M_{\rm VIR}$ значительно меньше зависит от ошибок калибровки), найденные значения в общем не противоречат

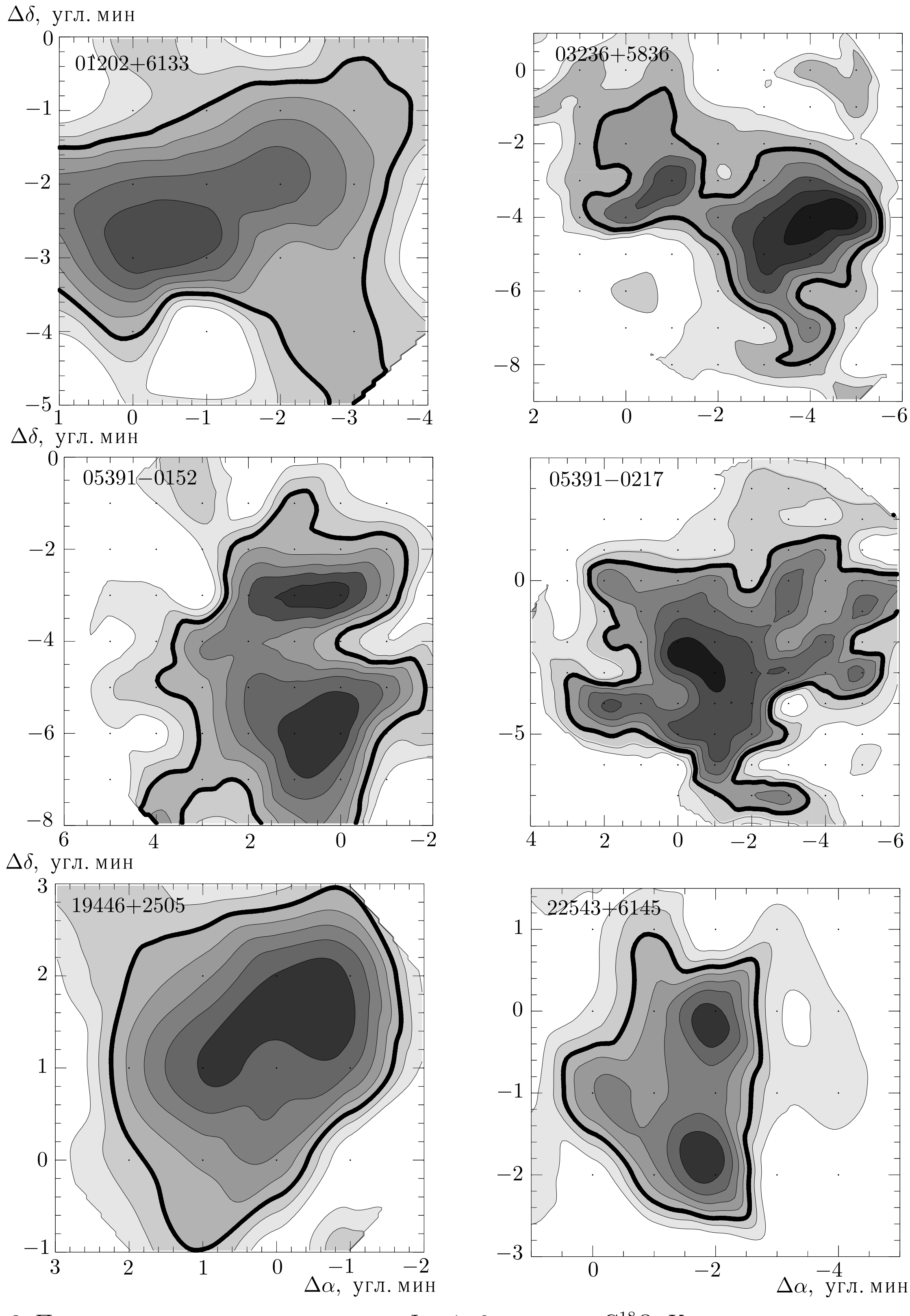


Рис. 3. Примеры полученных карт в линии $J = 1$—0 молекулы $C^{18}O$. Контуры соответствуют интегральной интенсивности от 90 % до 30 % с шагом 10 % от максимального значения. Уровень, равный половине интегральной интенсивности, выделен жирной линией

Таблица 3. Физические параметры ядер

Источник IRAS	Θ, угл. мин	R, пк	$M_{\rm LTE}$, $M_\odot$	$M_{\rm VIR}$, $M_\odot$	$\overline{n}({\rm H_2})$, 10^3 см$^{-3}$
01202+6133	4,9	0,7	540	394	6,8
03236+5836	5,3	0,5	493	442	11,3
05391−0152	6,6	0,7	1 332	676	13,5
05391−0217	8,2	0,9	1 749	546	9,1
19446+2505	4,2	1,2	1 346	1 133	2,5
22543+6145	3,2	0,5	388	696	14,2

условию близости величин $M_{\rm LTE}$ и $M_{\rm VIR}$, что характерно для ядер, связанных с областями образования массивных звёзд (см., например, [8]). В последней колонке табл. 3 приведены средние плотности ядер

$$\overline{n}({\rm H_2}) = (4\pi R^3/3)^{-1}\ \frac{M_{\rm LTE}}{1{,}4 m_{\rm H_2}}\,,$$

которые составили $(0{,}3\div 1{,}4)\cdot 10^4$ см$^{-3}$.

Несмотря на то, что ряд объектов, вошедших в данную выборку, ранее исследовался в различных молекулярных линиях, сравнение оценок физических параметров для конкретных источников представляется затруднительным из-за различных методов оценки и приближений, а также существенных погрешностей интенсивности линий для наших данных. Если рассмотреть в целом диапазоны физических параметров, найденных для больших выборок плотных ядер, связанных с областями образования массивных звёзд (например, [8–10]), нетрудно заметить, что оценки масс и размеров, найденные нами для шести объектов, лежат в этих диапазонах.

Как упоминалось ранее, представленные нами данные являются предварительными и будут в дальнейшем уточнены.

4. ЗАКЛЮЧЕНИЕ

С целью поиска плотных ядер и определения их физических параметров проведены наблюдения 23 молекулярных облаков, связанных с «холодными» ИК источниками IRAS, в линиях излучения $J = 1$—0 молекул $\rm C^{18}O$, $\rm ^{13}CO$ и $\rm ^{12}CO$. Линия молекулы $\rm C^{18}O$ зарегистрирована в 21 случае, линии молекул $\rm ^{13}CO$ и $\rm ^{12}CO$ зарегистрированы во всех случаях. В предположении локального термодинамического равновесия оценены лучевые концентрации молекул $\rm C^{18}O$ и $\rm H_2$ в направлении максимума интегральной интенсивности излучения в линии молекулы $\rm C^{18}O$, которые составили $(2{,}5\div 10{,}4)\cdot 10^{15}$ см$^{-2}$ и $(1{,}5\div 6{,}1)\cdot 10^{22}$ см$^{-2}$ соответственно. Кинетические температуры, определённые по линиям молекулы CO для этих максимумов, варьируются от 14 до 45 K. Для шести объектов получены полные карты интенсивности излучения в линии $\rm C^{18}O$. Для них оценены угловые и линейные размеры областей излучения, которые лежат в диапазоне $3{,}2'\div 8{,}2'$ и $0{,}5\div 1{,}2$ пк соответственно, и рассчитаны физические параметры. Массы этих объектов лежат в диапазоне $(390\div 1\,750)\ M_\odot$, диапазон средних плотностей $(0{,}3\div 1{,}4)\cdot 10^4$ см$^{-3}$, вириальные массы лежат в диапазоне $(390\div 1\,100)\ M_\odot$.

Авторы выражают благодарность коллективу обсерватории Пурпурная Гора (Purple Mountain Observatory) за помощь в проведении наблюдений. Работа проведена при поддержке РФФИ и ГФЕН Китая (совместный грант № 03–02–39016-ГФЕН) и Совета при Президенте РФ по поддержке ведущих научных школ (грант № 1483).

СПИСОК ЛИТЕРАТУРЫ

1. Yang J., Jiang Z., Wang M., Ju B., Wang H. // Astrophys. J. Suppl. Ser. 2002. V. 141. P. 157.
2. Truong-Bach, Morris D., Nguyen-Q-Reiu // Astron. Astrophys. 1991. V. 249. P. 435.
3. Blitz L., Fich M., Stark A. A. // Astrophys. J. Suppl. Ser. 1982. V. 49. P. 183.
4. http://www.iram.fr/IRAMFR/GILDAS.
5. Зинченко И. И., Лапинов А. В., Пирогов Л. Е. // Астрон. журн. 1989. Т. 66. С. 1142.
6. Frerking M. A., Langer W. D., Wilson R. W. // Astrophys. J. 1982. V. 262. P. 590.
7. Зинченко И. И. // Письма в Астрон. журн. 2000. Т. 26. С. 933.
8. Zinchenko I., Pirogov L., Toriseva M. // Astron. Astrophys. 1998. V. 133. P. 337.
9. Fontani F., Cesaroni R., Caselli P., Olmi L. // Astron. Astrophys. 2002. V. 389. P. 603.
10. Mueller K. E., Shirley Y. L., Evans II N. J., Jacobson H. R. // Astrophys. J. Suppl. Ser. 2002. V. 143. P. 469.

Поступила в редакцию 18 декабря 2004 г.; принята в печать 23 июня 2005 г.

THE CO ($J = 1$—0) AND ISOTOPE MOLECULAR LINE SURVEY OF STAR-FORMATION REGIONS ASSOCIATED WITH INFRARED SOURCES. THE RESULTS OF $C^{18}O$ ($J = 1$—0) OBSERVATIONS

N. R. Troitsky, L. E. Pirogov, I. I. Zinchenko, and J. Yang

We present observation results for 23 dense molecular cloud cores associated with cold IRAS sources in the $C^{18}O$ $J = 1$—0 molecular line. The observations took place at the 13.7-m radiotelescope of the Purple Mountain Observatory, Qinghai Station, China. The $C^{18}O$ (1—0) line was detected in 21 sources. The $C^{18}O$ and H_2 column densities toward the positions of the maximum $C^{18}O$ integrated intensity are estimated in the approximation of local thermodynamic equilibrium. They lie in the ranges $(2.5$–$10.4) \cdot 10^{15}$ cm^{-2} and $(1.5$–$6.1) \cdot 10^{22}$ cm^{-2}, respectively. Kinetic temperatures derived from CO lines for these positions vary from 14 K to 45 K. For 6 objects for which mapping is almost completed, the sizes of the $C^{18}O$ emission regions are calculated. They vary from 0.5 to 1.2 pc. The masses of these objects lie in the range $\sim (390$–$1\,750)$ $M_\odot$ and are close to the virial mass estimates. The range of mean number densities is $\sim (0.3$–$1.4) \cdot 10^4$ cm^{-3}.

УДК 533.951+537.868+621.371

О ФАЗОВЫХ ФЛУКТУАЦИЯХ РАДИОВОЛН ЗА ТУРБУЛЕНТНЫМ ФАЗОВЫМ ЭКРАНОМ

В. А. Алимов, А. В. Рахлин

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

Рассмотрен вопрос о фазовых флуктуациях радиоволн за турбулентным оптически тонким неоднородным слоем (фазовым экраном). Получены выражения для основных статистических характеристик флуктуаций фазы при современных прямых методах определения этих характеристик по результатам измерений комплексного поля принимаемого излучения. Проанализированы режимы слабых и насыщенных мерцаний сигналов. Показано, что в режиме слабых мерцаний статистические характеристики фазовых флуктуаций принимаемого излучения (структурная функция, спектры флуктуаций при одиночном и интерференционном приёме) практически совпадают с соответствующими статистическими характеристиками флуктуаций фазы на выходе экрана. В режиме насыщенных мерцаний получена информация о структурной функции фазовых флуктуаций принимаемого излучения в случаях сравнительно большого и малого пространственного разнесения точек наблюдения. Полученные результаты позволяют корректно анализировать данные современных экспериментов по дистанционному зондированию околоземной и космической плазмы с использованием прямых фазовых методов диагностики окружающей среды.

ВВЕДЕНИЕ

Модель оптически тонкого неоднородного слоя (фазового экрана) широко используется в исследованиях распространения радиоволн в случайно-неоднородных средах. С помощью этой модели решены многие задачи дифракции радиоизлучения в околоземной и космической плазме (см., например, [1, 2] и цитированную там литературу). Необходимо заметить, что подавляющее число работ, в которых авторы использовали модель турбулентного фазового экрана, было посвящено проблеме флуктуаций интенсивности при рассеянии волновых полей (см. обзорные работы [3, 4]). Вместе с тем, даже спустя несколько десятилетий после выхода основополагающей работы по дифракции радиоволн на фазовом экране [5], проблема флуктуаций фазы радиоволн за турбулентным фазовым экраном остаётся недостаточно изученной.

В настоящее время вопрос о флуктуациях фазы радиоволн после дифракции их на оптически тонком слое с турбулентными неоднородностями стал актуальным не только в теоретическом аспекте, но приобрёл и сугубо практическое значение. Дело в том, что в последние годы в космической радиофизике успешно развиваются фазовые методы диагностики околоземной и космической плазмы [6]. При этом используются прямые методы определения статистических характеристик флуктуаций фазы радиоволн (частотного спектра, структурной функции фазовых флуктуаций и т. п.) после дифракции их на оптически тонком неоднородном слое, например, околосолнечной плазмы с последующей регистрацией принимаемого радиоизлучения от удалённых космических источников с помощью радиоинтерферометров со сверхдлинными базами на Земле [7–9]. Во всех этих работах применяют прямые методы измерений фаз комплексных полей принимаемых сигналов. В результате непосредственно получают сведения о структурной функции или частотном спектре фазовых флуктуаций.

Целью настоящей работы является получение теоретических соотношений, связывающих характеристики флуктуаций свойств среды с соответствующими характеристиками флуктуаций фазы радиоволн в случае произвольных фазовых флуктуаций принимаемого излучения. При этом в качестве модели среды используется известное приближение неоднородного слоя в форме хаотического фазового экрана, а фазовые характеристики принимаемого излучения определяются после дифракции излучения в свободном пространстве за этим фазовым экраном. Режимы слабых ($F_I \ll 1$) и насыщенных ($F_I \approx 1$) мерцаний анализируются отдельно (здесь F_I — индекс флуктуаций интенсивности принимаемого излучения [2]).[1]

1. ПРОСТРАНСТВЕННЫЕ СПЕКТРЫ ФЛУКТУАЦИЙ КОМПЛЕКСНОГО ПОЛЯ РАДИОВОЛН

Рассмотрим следующую задачу. Пусть плоская волна проходит неоднородный слой и принимается в плоскости, удалённой на расстояние z от него. Слой будем считать оптически тонким, т. е. вызывающим только фазовые флуктуации проходящего излучения (фазовым экраном [1, 2]). Флуктуации излучения на выходе такого турбулентного экрана будем характеризовать структурной функцией изотропных фазовых флуктуаций [1, 2]:

$$D_S(\rho) = 4\pi \int\limits_0^\infty \kappa_\perp \Phi_S(\kappa_\perp, 0)\,[1 - J_0(\kappa_\perp \rho)]\,\mathrm{d}\kappa_\perp, \tag{1}$$

где $\rho \equiv |\boldsymbol{\rho}|$ — расстояние между точками наблюдения, $\kappa_\perp$ — пространственное волновое число, $\kappa_\perp = \sqrt{\kappa_x^2 + \kappa_y^2}$, x и y — декартовы координаты в плоскости экрана, $J_0(\kappa_\perp \rho)$ — функция Бесселя нулевого порядка, $\Phi_S(\kappa_\perp, 0)$ — трёхмерный спектр фазовых флуктуаций, который в дальнейших вычислениях будем считать равным [1, 2]

$$\Phi_S(\kappa_\perp, 0) = \frac{C_S^2}{|\kappa_\perp|^p}\,\exp(-\kappa_\perp^2/\kappa_\mathrm{m}^2). \tag{2}$$

Здесь C_S^2 — структурная постоянная, κ_m — пространственное волновое число, соответствующее внутреннему масштабу турбулентности l_m [1, 2], p — показатель трёхмерного спектра неоднородностей фазового экрана ($2 < p < 4$; для случая колмогоровской турбулентности $p = 11/3$ [1, 2]).

Представим структурную функцию $D_S(\rho)$ в виде суммы двух слагаемых:

$$D_S(\rho) = D_{S_\text{м}}(\rho) + D_{S_\text{кр}}(\rho). \tag{3}$$

При этом сами слагаемые могут быть представлены в двух эквивалентных видах (см. (1)):

$$D_{S_\text{м}}(\rho) = 4\pi \int\limits_0^\infty \kappa_\perp \Phi_S(\kappa_\perp, 0)\,[1 - J_0(\kappa_\perp \rho)]\,[1 - \exp(-\kappa_\perp^2/\kappa_\text{эф}^2)]\,\mathrm{d}\kappa_\perp,$$

$$D_{S_\text{кр}}(\rho) = 4\pi \int\limits_0^\infty \kappa_\perp \Phi_S(\kappa_\perp, 0)\,[1 - J_0(\kappa_\perp \rho)]\,\exp(-\kappa_\perp^2/\kappa_\text{эф}^2)\,\mathrm{d}\kappa_\perp, \tag{4}$$

или

$$D_{S_\text{м}}(\rho) = 4\pi \int\limits_{\kappa_\text{эф}}^\infty \kappa_\perp \Phi_S(\kappa_\perp, 0)\,[1 - J_0(\kappa_\perp \rho)]\,\mathrm{d}\kappa_\perp,$$

[1] Для модели фазового экрана ранее была решена соответствующая задача только в приближении слабых фазовых флуктуаций волны на экране, а в случае сильных флуктуаций фазы получено решение, справедливое в зоне Фраунгофера за экраном с одномасштабными неоднородностями лишь для интервала изменения фазы принимаемого излучения в пределах $[0, 2\pi]$ (см. [1] и цитированную там литературу).

$$D_{S_{\text{кр}}}(\rho) = 4\pi \int\limits_0^{\kappa_{\text{эф}}} \kappa_\perp \Phi_S(\kappa_\perp, 0)\,[1 - J_0(\kappa_\perp \rho)]\,\mathrm{d}\kappa_\perp. \tag{5}$$

В соотношениях (4), (5) введено понятие эффективного пространственного волнового числа $\kappa_{\text{эф}}$, которое разделяет спектр фазовых флуктуаций на экране на две компоненты (два слагаемых), соответствующих мелкомасштабным и крупномасштабным неоднородностям флуктуаций фазы волны на выходе экрана (ср. [10]).

В режиме слабых мерцаний, когда структурная функция фазовых флуктуаций на масштабе первой зоны Френеля мала ($D_{\rm S}(\sqrt{z/k}) \ll 1$), см. [11]), в качестве параметра $\kappa_{\text{эф}}$ целесообразно выбрать френелевское волновое число $\kappa_{\text{эф}} = \kappa_{\text{фр}} = \sqrt{k/z}$, поскольку флуктуации интенсивности в этом случае формируются в основном за счёт неоднородностей с размерами, меньшими первой зоны Френеля; здесь $k = 2\pi/\lambda$, λ — длина волны излучения. В условиях насыщенных мерцаний ($D_S(\sqrt{z/k}) \gg 1$, см. [11]), когда основной вклад во флуктуации интенсивности вносят фазовые неоднородности с волновыми числами, бо́льшими $\kappa_{\perp\text{н}} = \kappa_{\text{фр}}\,[D_S(\sqrt{z/k})]^{-1/(p-2)}$ (ср. [11]), в качестве параметра $\kappa_{\text{эф}}$ целесообразно выбрать это волновое число: $\kappa_{\text{эф}} = \kappa_{\text{фр}}\,[D_S(\sqrt{z/k})]^{-1/(p-2)}$.

Функция пространственной корреляции флуктуаций комплексного поля, которая для крупных (в масштабе длины волны) неоднородностей фазы на экране сохраняется при распространении флуктуирующего излучения в свободном пространстве за неоднородным слоем [1, 2], может быть записана в следующем виде (ср. [1, 2]):

$$\Gamma_{EE^*}(\rho) = \exp\bigl[-[D_{S_{\text{м}}}(\rho) + D_{S_{\text{кр}}}(\rho)]/2\bigr] \equiv \Gamma_{E_{\rm g}E_{\rm g}^*}(\rho)\Gamma_{E_{\rm p}E_{\rm p}^*}(\rho). \tag{6}$$

Здесь

$$\Gamma_{E_{\rm g}E_{\rm g}^*}(\rho) = \exp[-D_{S_{\text{м}}}(\rho)/2]\,, \qquad \Gamma_{E_{\rm p}E_{\rm p}^*}(\rho) = \exp\bigl[-D_{S_{\text{кр}}}(\rho)/2\bigr] \tag{7}$$

— корреляционные функции флуктуаций комплексных полей радиоволн, соответствующие дифракционной и рефракционной компонентам рассеянного поля на фазовом экране (см., например, [12] и цитированную там литературу). [2]

С учётом соотношений (3)–(7) двумерный пространственный спектр флуктуаций комплексного поля радиоволн может быть представлен в виде свёртки двух спектров (ср. [5]):

$$\overline{F^2}(\boldsymbol{\kappa}_\perp) = \frac{1}{(2\pi)^2} \iint\limits_{-\infty}^{+\infty} \Gamma_{EE^*}(\boldsymbol{\rho}) \exp(-i\boldsymbol{\kappa}_\perp\boldsymbol{\rho})\,\mathrm{d}\boldsymbol{\rho} = \iint\limits_{-\infty}^{+\infty} \overline{F^2_{\rm g}}(\boldsymbol{\kappa}'_\perp)\overline{F^2_{\rm p}}(\boldsymbol{\kappa}_\perp - \boldsymbol{\kappa}'_\perp)\,\mathrm{d}\boldsymbol{\kappa}'_\perp, \tag{8}$$

где

$$\overline{F^2_{\rm g}}(\boldsymbol{\kappa}_\perp) = \frac{1}{(2\pi)^2} \iint\limits_{-\infty}^{+\infty} \Gamma_{E_{\rm g}E_{\rm g}^*}(\boldsymbol{\rho}) \exp(-i\boldsymbol{\kappa}_\perp\boldsymbol{\rho})\,\mathrm{d}\boldsymbol{\rho} = \frac{1}{2\pi}\int\limits_0^\infty \exp[-D_{S_{\text{м}}}(\rho)/2] J_0(\kappa_\perp\rho)\rho\,\mathrm{d}\rho,$$

$$\overline{F^2_{\rm p}}(\boldsymbol{\kappa}_\perp) = \frac{1}{(2\pi)^2} \iint\limits_{-\infty}^{+\infty} \Gamma_{E_{\rm p}E_{\rm p}^*}(\boldsymbol{\rho}) \exp(-i\boldsymbol{\kappa}_\perp\boldsymbol{\rho})\,\mathrm{d}\boldsymbol{\rho} = \frac{1}{2\pi}\int\limits_0^\infty \exp[-D_{S_{\text{кр}}}(\rho)/2] J_0(\kappa_\perp\rho)\rho\,\mathrm{d}\rho \tag{9}$$

— двумерные спектры флуктуаций дифракционной и рефракционной компонент поля принимаемого излучения соответственно.

[2] Понятие дифракционного и рефракционного рассеяния впервые было введено Буккером в 1981 г. (подробнее см. [13]).

Дальнейшее рассмотрение мы будем проводить отдельно для случая слабых мерцаний излучения после его дифракции на фазовом экране ($D_S(\sqrt{z/k}) \ll 1$) и случая насыщенных мерцаний ($D_S(\sqrt{z/k}) \gg 1$).

В случае слабых мерцаний для структурной функции фазы $D_{S_{\text{м}}}(\rho)$ удобно воспользоваться соотношением (4) в эквивалентном комплексном виде (ср. [2]):

$$D_{S_{\text{м}}}(\boldsymbol{\rho}) = 2 \iint\limits_{-\infty}^{+\infty} [1 - \exp(i\boldsymbol{\kappa}_\perp \boldsymbol{\rho})]\, \Phi_S(\boldsymbol{\kappa}_\perp, 0)\, [1 - \exp(-\kappa_\perp^2/\kappa_{\text{фр}}^2)]\, \mathrm{d}\boldsymbol{\kappa}_\perp. \tag{10}$$

Подставляя это выражение в соотношение (9) и учитывая, что в данном случае $D_{S_{\text{м}}}(\sqrt{z/k}) \ll 1$, в результате несложных преобразований для спектра флуктуаций дифракционной компоненты комплексного поля получаем следующее приближённое соотношение:

$$\overline{F_{\text{g}}^2}(\kappa_\perp) \approx \delta(\kappa_\perp) + \Phi_S(\boldsymbol{\kappa}_\perp, 0)\, [1 - \exp(-\kappa_\perp^2/\kappa_{\text{фр}}^2)], \tag{11}$$

где $\delta(\kappa_\perp)$ — дельта-функция Дирака.

Выражение для пространственного спектра флуктуаций рефракционной компоненты комплексного поля можно записать в следующем приближённом виде:

$$\overline{F_{\text{p}}^2}(\kappa_\perp) \approx \frac{1}{2\pi} \int\limits_0^{\rho_{\text{кор}}} \rho J_0(\kappa_\perp \rho)\, \mathrm{d}\rho = \frac{\rho_{\text{кор}}^2}{2\pi}\, \frac{J_1(\kappa_\perp \rho_{\text{кор}})}{\kappa_\perp \rho_{\text{кор}}}\,, \tag{12}$$

где $\rho_{\text{кор}} = \sqrt{z/k}\ [D_S(\sqrt{z/k})]^{-1/(p-2)}$, $J_1(\kappa_\perp \rho_{\text{кор}})$ — функция Бесселя первого порядка.

При выводе соотношения (12) было учтено, что структурная функция фазовых флуктуаций для спектра (2) имеет вид (ср. [1])

$$D_S(\rho) \approx \begin{cases} D_S(\sqrt{z/k})\, (\rho/\sqrt{z/k})^{p-2}, & \rho \ge l_{\text{m}}; \\ D_S(\sqrt{z/k}) \rho^2/(\sqrt{z/k})^{p-2}\, l_{\text{m}}^{p-4}, & \rho < l_{\text{m}}. \end{cases} \tag{13}$$

Кроме того, предполагалось, что выполнено неравенство $\kappa_{\text{фр}} \ll \kappa_{\text{m}}$ ($l_{\text{m}} \ll \sqrt{z/k}$), так что можно было опустить экспоненциальный множитель в формуле (2) для $\Phi_S(\kappa_\perp, 0)$ и из соотношения (4) для структурной функции $D_{S_{\text{кр}}}(\rho)$ получить выражение (ср. (13))

$$D_{S_{\text{кр}}}(\rho) \approx \begin{cases} D_S(\sqrt{z/k})\, (\rho/\sqrt{z/k})^{p-2}, & \rho \ge l_{\text{m}}; \\ D_S(\sqrt{z/k})\, (\rho/\sqrt{z/k})^2, & \rho < l_{\text{m}}. \end{cases} \tag{14}$$

После подстановки (14) в (9) и некоторых преобразований для спектра $\overline{F_{\text{p}}^2}(\kappa_\perp)$ получаем соотношение (12).

При выводе выражений для спектров флуктуаций дифракционной и рефракционной компонент комплексного поля в случае насыщенных мерцаний ($D_S(\sqrt{z/k}) \gg 1$) следует воспользоваться соотношениями (9). При этом надо учесть, что соответствующие структурные функции $D_{S_{\text{м}}}(\rho)$ и $D_{S_{\text{кр}}}(\rho)$, входящие в эти соотношения, могут быть определены непосредственно из соотношения (5), где эффективное пространственное волновое число $\kappa_{\text{эф}} = \kappa_{\text{фр}}\, [D_S(\sqrt{z/k})]^{-1/(p-2)}$ (см. выше). Фактически это означает, что масштаб $l_{\text{эф}} = \kappa_{\text{фр}}^{-1} = \sqrt{z/k}\ [D_S(\sqrt{z/k})]^{1/(p-2)}$ играет роль внутреннего масштаба турбулентности при вычислениях $D_{S_{\text{кр}}}(\rho)$ и роль внешнего масштаба турбулентности при вычислениях $D_{S_{\text{м}}}(\rho)$ [1,2]. В результате для $D_{S_{\text{кр}}}(\rho)$ получаем следующее выражение (ср. (14)):

$$D_{S_{\text{кр}}}(\rho) \approx \begin{cases} D_S(\sqrt{z/k})\,(\rho/\sqrt{z/k})^{p-2}, & \rho \ge l_{\text{эф}};\\ \rho^2/l_{\text{эф}0}^2, & \rho < l_{\text{эф}0}, \end{cases} \tag{15}$$

где $l_{\text{эф}0} = \sqrt{(z/k)\,D_S(\sqrt{z/k})}$.

Соответственно, для $D_{S_{\text{м}}}(\rho)$ можно записать следующее приближённое соотношение (ср. [1] и (13)):

$$D_{S_{\text{м}}}(\rho) \approx \begin{cases} D_S(\sqrt{z/k})\,(l_{\text{эф}}/\sqrt{z/k})^{p-2}, & \rho \ge l_{\text{эф}};\\ D_S(\sqrt{z/k})\,(\rho\big/\sqrt{z/k})^{p-2}, & l_{\text{m}} < \rho < l_{\text{эф}};\\ D_S(\sqrt{z/k})\rho^2\big/(\sqrt{z/k})^{p-2}\,l_{\text{m}}^{p-4}, & \rho \le l_{\text{m}}. \end{cases} \tag{16}$$

Подставляя соотношения (15), (16) в (9) и проводя необходимые преобразования, получаем следующие приближённые выражения:

$$\overline{F_{\text{g}}^2}(\kappa_\perp) \approx \frac{\rho_{\text{кор}}^2}{2}\,\frac{J_1(\kappa_\perp\rho_{\text{кор}})}{\kappa_\perp\rho_{\text{кор}}}\,, \tag{17}$$

$$\overline{F_{\text{p}}^2}(\kappa_\perp) \approx \frac{l_{\text{эф}0}^2}{2}\,\frac{J_1(\kappa_\perp l_{\text{эф}0})}{\kappa_\perp l_{\text{эф}0}}\,. \tag{18}$$

Напомним, что здесь $\rho_{\text{кор}} = \sqrt{z/k}\,[D_S(\sqrt{z/k})]^{-1/(p-2)}$, $l_{\text{эф}0} = \sqrt{(z/k)\,D_S(\sqrt{z/k})}$; также в режиме насыщенных мерцаний выполнено неравенство $D_S(\sqrt{z/k}) \gg 1$.

Соотношения (11), (12) и (17), (18) определяют пространственные спектры флуктуаций дифракционной и рефракционной компонент комплексного поля радиоволн при их дифракции на турбулентном фазовом экране в режиме слабых ($D_S(\sqrt{z/k}) \ll 1$) и насыщенных ($D_S(\sqrt{z/k}) \gg 1$) мерцаний соответственно. Анализ этих соотношений позволяет сделать следующие выводы.

В режиме слабых мерцаний пространственный спектр флуктуаций дифракционной составляющей комплексного поля (11) имеет две выделенные компоненты: $\kappa_\perp = 0$ и $\kappa_{\perp\text{м}} \approx \kappa_{\text{фр}} = \sqrt{k/z}$. Первая отвечает регулярной компоненте в дифракционном поле принимаемого сигнала, вторая — слабому дифракционному рассеянию радиоволн на неоднородностях с размерами порядка первой зоны Френеля ($l_{\text{м}} \approx \sqrt{z/k}$). Характерный масштаб спектра флуктуаций рефракционной компоненты комплексного поля (12) соответствует $\kappa_{\perp\text{кр}} \approx l_{\text{кр}}^{-1}$, где $l_{\text{кр}} \approx \rho_{\text{кор}} = \sqrt{z/k}\,[D_S(\sqrt{z/k})]^{-1/(p-2)}$.

В режиме насыщенных мерцаний дифракционная и рефракционная составляющие пространственного спектра флуктуаций комплексного поля (17), (18) характеризуются спектральными компонентами $\kappa_\perp \le \kappa_{\perp\text{g}} = \rho_{\text{кор}}^{-1}$ и $\kappa_\perp \le \kappa_{\perp\text{p}} \approx l_{\text{эф}0}^{-1}$ соответственно. При этом характерный пространственный масштаб флуктуаций дифракционной составляющей комплексного поля $\rho_{\text{кор}} \approx \sqrt{z/k}\,[D_S(\sqrt{z/k})]^{-1/(p-2)}$ много меньше размера первой зоны Френеля $\sqrt{z/k}$, а характерный пространственный масштаб рефракционной составляющей спектра комплексного поля $l_{\text{эф}0} = \sqrt{(z/k)\,D_S(\sqrt{z/k})}$ много больше $\sqrt{z/k}$.

2. КОМПЛЕКСНОЕ ПОЛЕ ПРИНИМАЕМОГО СИГНАЛА

Следуя [5], комплексное поле принимаемого сигнала запишем в виде спектрального преобразования Фурье:

$$E(\boldsymbol{\rho}, z) \approx \exp(-ikz) \iint\limits_{-\infty}^{+\infty} F(\boldsymbol{\kappa}_\perp) \exp(-i\kappa_\perp^2 z/2k) \exp(i\boldsymbol{\kappa}_\perp \boldsymbol{\rho}) \, \mathrm{d}\boldsymbol{\kappa}_\perp. \quad (19)$$

Здесь $F(\boldsymbol{\kappa}_\perp)$ — текущий пространственный спектр комплексного поля на выходе фазового экрана.

Поле на экране мы представим в виде произведения двух комплексных полей:

$$E(\boldsymbol{\rho}, 0) = E_{\mathrm{g}}(\boldsymbol{\rho}, 0) E_{\mathrm{p}}(\boldsymbol{\rho}, 0) = \exp\{i\,[S_{\text{м}}(\boldsymbol{\rho}) + S_{\text{кр}}(\boldsymbol{\rho})]\}, \quad (20)$$

формируемых мелкомасштабными ($S_{\text{м}}(\boldsymbol{\rho})$) и крупномасштабными ($S_{\text{кр}}(\boldsymbol{\rho})$) неоднородностями флуктуаций фазы на экране. Тогда для текущего спектра комплексного поля на экране справедливо соотношение, аналогичное (8):

$$F(\boldsymbol{\kappa}_\perp) = \iint\limits_{-\infty}^{+\infty} F_{\mathrm{g}}(\boldsymbol{\kappa}'_\perp) F_{\mathrm{p}}(\boldsymbol{\kappa}_\perp - \boldsymbol{\kappa}'_\perp) \, \mathrm{d}\boldsymbol{\kappa}'_\perp. \quad (21)$$

Здесь $F_{\mathrm{g}}(\boldsymbol{\kappa}_\perp)$, $F_{\mathrm{p}}(\boldsymbol{\kappa}_\perp)$ — текущие фурье-спектры мелкомасштабной и крупномасштабной компонент комплексных полей $E_{\mathrm{g}}(\boldsymbol{\rho}, 0)$ и $E_{\mathrm{p}}(\boldsymbol{\rho}, 0)$ на экране соответственно.

Подставляя (21) в (19), получаем следующее выражение для комплексного поля принимаемого сигнала:

$$E(\boldsymbol{\rho}, z) \approx \exp(-ikz) \iint\limits_{-\infty}^{+\infty} \iint\limits_{-\infty}^{+\infty} F_{\mathrm{g}}(\boldsymbol{\kappa}_{\perp\mathrm{g}}) F_{\mathrm{p}}(\boldsymbol{\kappa}_{\perp\mathrm{p}}) \exp\left[-i\,\frac{(\boldsymbol{\kappa}_{\perp\mathrm{g}} + \boldsymbol{\kappa}_{\perp\mathrm{p}})^2\, z}{2k}\right] \times$$
$$\times \exp[i\,(\boldsymbol{\kappa}_{\perp\mathrm{g}} + \boldsymbol{\kappa}_{\perp\mathrm{p}})\,\boldsymbol{\rho}] \, \mathrm{d}\boldsymbol{\kappa}_{\perp\mathrm{g}} \, \mathrm{d}\boldsymbol{\kappa}_{\perp\mathrm{p}}. \quad (22)$$

Из соотношения (22) следует, что при выполнении неравенств

$$\frac{\kappa_{\perp\mathrm{g}} \kappa_{\perp\mathrm{p}}}{k}\, z \ll 2\pi, \qquad \frac{\kappa_{\perp\mathrm{p}}^2 z}{k} \ll 2\pi \quad (23)$$

комплексное поле принимаемого сигнала может быть записано в виде произведения двух сомножителей (ср. (6)):

$$E(\boldsymbol{\rho}, z) \approx \left[\exp(-ikz) \iint\limits_{-\infty}^{+\infty} F_{\mathrm{g}}(\boldsymbol{\kappa}_{\perp\mathrm{g}}) \exp[-i\kappa_{\perp\mathrm{g}}^2 z/(2k)] \exp(i\boldsymbol{\kappa}_{\perp\mathrm{g}} \boldsymbol{\rho}) \, \mathrm{d}\boldsymbol{\kappa}_{\perp\mathrm{g}}\right] \times$$
$$\times \left[\iint\limits_{-\infty}^{+\infty} F_{\mathrm{p}}(\boldsymbol{\kappa}_{\perp\mathrm{p}}) \exp(i\boldsymbol{\kappa}_{\perp\mathrm{p}} \boldsymbol{\rho}) \, \mathrm{d}\boldsymbol{\kappa}_{\perp\mathrm{p}}\right] \equiv E_{\mathrm{g}}(\boldsymbol{\rho}, z) E_{\mathrm{p}}(\boldsymbol{\rho}, 0). \quad (24)$$

Здесь $E_{\mathrm{g}}(\boldsymbol{\rho}, z)$ описывает комплексное поле дифракционной (мелкомасштабной) компоненты принимаемого излучения, а $E_{\mathrm{p}}(\boldsymbol{\rho}, 0)$ — комплексное поле рефракционной (крупномасштабной) компоненты, причём поле $E_{\mathrm{p}}(\boldsymbol{\rho}, z)$ совпадает с крупномасштабной компонентой поля волны на выходе экрана.

Как было показано в разделе 1, характерный масштаб спектра дифракционной компоненты комплексного поля принимаемого излучения в режиме слабых мерцаний ($D_S(\sqrt{z/k}) \ll 1$) равен $\kappa_{\perp \mathrm{g}} = \kappa_{\text{фр}} = \sqrt{k/z}$, а масштаб рефракционной компоненты в этих же условиях составляет $\kappa_{\perp \mathrm{p}} \approx \sqrt{k/z}\,[D_S(\sqrt{z/k})]^{1/(p-2)}$. В режиме насыщенных мерцаний ($D_S(\sqrt{z/k}) \gg 1$) спектры дифракционной и рефракционной компонент сосредоточены в областях $\kappa_\perp \le \kappa_{\perp \mathrm{g}} \approx \sqrt{k/z} \times [D_S(\sqrt{z/k})]^{1/(p-2)}$ и $\kappa_\perp \le \kappa_{\perp \mathrm{p}} \approx \sqrt{k/z}\,[D_S(\sqrt{z/k})]^{-1/2}$ соответственно. Таким образом, необходимые условия (23) выполнены, и в первом приближении можно считать, что комплексное поле принимаемого сигнала за турбулентным фазовым экраном может быть представлено в виде произведения двух комплексных полей (24). При этом статистические свойства флуктуаций дифракционного $E_{\mathrm{g}}(\boldsymbol{\rho}, z)$ и рефракционного $E_{\mathrm{p}}(\boldsymbol{\rho}, 0)$ комплексных полей в принимаемом излучении характеризуются соответствующими пространственными спектрами флуктуаций комплексных полей (11), (12) и (17), (18) для режимов слабых и насыщенных мерцаний.

Следует ещё раз подчеркнуть, что рефракционное комплексное поле $E_{\mathrm{p}}(\boldsymbol{\rho}, z)$ принимаемого сигнала совпадает с рефракционной компонентой $E_{\mathrm{p}}(\boldsymbol{\rho}, 0)$ на выходе оптически тонкого неоднородного слоя (фазового экрана). С учётом этого обстоятельства для комплексного поля принимаемого излучения можно записать следующее приближённое выражение:

$$E(\boldsymbol{\rho}, z) \approx A_{\mathrm{g}}(\boldsymbol{\rho}, z) \exp[i\varphi_{\mathrm{g}}(\boldsymbol{\rho}, z)] \exp[iS_{\text{кр}}(\boldsymbol{\rho})], \tag{25}$$

где $A_{\mathrm{g}}(\boldsymbol{\rho}, z)$ и $\varphi_{\mathrm{g}}(\boldsymbol{\rho}, z)$ — амплитуда и фаза дифракционной компоненты поля $E(\boldsymbol{\rho}, z)$, $S_{\text{кр}}(\boldsymbol{\rho})$ — флуктуационный набег фазы на крупномасштабных рефракционных неоднородностях оптически тонкого неоднородного слоя. Для случая слабых мерцаний это неоднородности фазы с масштабами $l_{\text{кр}} > \sqrt{z/k}$, а в случае насыщенных мерцаний — неоднородности с размерами $l_{\text{кр}} > \sqrt{z/k}\,[D_S(\sqrt{z/k})]^{1/(p-2)}$.

3. СТАТИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ ФЛУКТУАЦИЙ ФАЗЫ ПРИНИМАЕМОГО ИЗЛУЧЕНИЯ

Представим поле принимаемого сигнала в общепринятом комплексном виде:

$$E(\boldsymbol{\rho}, z) = A(\boldsymbol{\rho}, z) \exp[i\varphi(\boldsymbol{\rho}, z)], \tag{26}$$

где $A(\boldsymbol{\rho}, z)$ и $\varphi(\boldsymbol{\rho}, z)$ — амплитуда и фаза принимаемого флуктуирующего сигнала соответственно. Используя это соотношение, путём несложных преобразований получаем следующее выражение для структурной функции фазовых флуктуаций принимаемого излучения:

$$D_\varphi(\boldsymbol{\rho}, z) = \overline{[\varphi(\boldsymbol{\rho}_1, z) - \varphi(\boldsymbol{\rho}_2, z)]^2} = -\frac{1}{4}\, \overline{\{\mathrm{Ln}[E(\boldsymbol{\rho}_1, z) E^*(\boldsymbol{\rho}_2, z)] - \mathrm{Ln}[E^*(\boldsymbol{\rho}_1, z) E(\boldsymbol{\rho}_2, z]\}^2}, \tag{27}$$

где $\boldsymbol{\rho} = \boldsymbol{\rho}_1 - \boldsymbol{\rho}_2$ — пространственное разнесение точек наблюдения в плоскости приёма, удалённой на расстояние z от рассеивающего слоя. Подставляя в соотношение (27) выражение для комплексного поля (25), получаем

$$D_\varphi(\rho, z) = D_{\varphi_{\mathrm{g}}}(\rho, z) + D_{S_{\text{кр}}}(\rho), \tag{28}$$

где $D_{\varphi_{\mathrm{g}}}(\rho, z) = \overline{[\varphi_{\mathrm{g}}(\boldsymbol{\rho}_1, z) - \varphi_{\mathrm{g}}(\boldsymbol{\rho}_2, z)]^2}$ — структурная функция фазовых флуктуаций дифракционной компоненты рассеянного поля, $D_{S_{\text{кр}}}(\rho)$ — структурная функция крупномасштабных флуктуаций фазы волны на экране, рассчитанная в геометрооптическом приближении по формулам (4) или (5).

В режиме слабых мерцаний выражение для функции $D_{\varphi_{\rm g}}(\rho,z)$ несложно получить, следуя вычислениям корреляционных характеристик флуктуаций фазы за хаотическим фазовым экраном в приближении слабых флуктуаций [1, 5]. В частности, в зоне Фраунгофера искомая структурная функция равна

$$D_{\varphi_{\rm g}}(\rho,z) \approx D_{S_{\rm м}}(\rho)/2, \tag{29}$$

где $D_{S_{\rm м}}(\rho)$ — структурная функция мелкомасштабных флуктуаций фазы волны на экране (см. (4), (5)).

Из соотношений (28), (29) с учётом (3) имеем

$$D_{\varphi}(\rho,z) = D_S(\rho) - \frac{D_{S_{\rm м}}(\rho)}{2} \approx D_S(\rho). \tag{30}$$

В соотношении (30) учтено, что в условиях слабых мерцаний $D_{S_{\rm м}}(\rho) \ll 1$. Таким образом, в режиме слабых мерцаний ($D_S(\sqrt{z/k}) \ll 1$) структурная функция флуктуаций фазы принимаемого излучения практически совпадает со структурной функцией флуктуаций фазы на выходе оптически тонкого неоднородного слоя (фазового экрана). Отличие будет заметным лишь при малых расстояниях между точками наблюдения (при $\rho < \sqrt{z/k}$), когда второе слагаемое в формуле (30) существенно.

В режиме насыщенных мерцаний ($D_S(\sqrt{z/k}) \gg 1$) ситуация более сложная. В этих условиях структурная функция флуктуаций фазы принимаемого излучения $D_\varphi(\rho,z)$ практически совпадает со структурной функцией флуктуаций фазы на выходе фазового экрана лишь при больших расстояниях между точками наблюдения ($\rho > \sqrt{z/k}\ [D_S(\sqrt{z/k})]^{1/(p-2)}$), когда она фактически равна структурной функции $D_{S_{\rm кр}}(\rho)$, формируемой крупномасштабными (с размерами $l > l_{\rm эф} = \sqrt{z/k}\ [D_S(\sqrt{z/k})]^{1/(p-2)}$) неоднородностями фазы волны на экране (см. (28) и разделы 1, 2). При меньших ρ основной вклад в функцию $D_{\varphi_{\rm g}}(\rho,z)$ дают более мелкомасштабные неоднородности (с размерами $l < l_{\rm эф}$). Проблема заключается именно в расчёте этой функции, входящей в соотношение (28).

Для случая относительно малых ρ функция $D_{\varphi_{\rm g}}(\rho,z)$ может быть рассчитана с привлечением известной формулы для плотности распределения $W_h(\vartheta)$ флуктуаций разности фаз ϑ нормального случайного процесса на интервале $\vartheta \in [-\pi,\pi]$ (см. [14]).

В нашем случае в условиях насыщенных мерцаний нормальное распределение с нулевым средним значением имеет дифракционная компонента комплексного поля $E_{\rm g}$ принимаемого излучения (см. выше). Поэтому соответствующее выражение для функции $W_h(\vartheta)$ из монографии [14] можно записать в виде

$$W_h(\vartheta) = \frac{1-R_0}{2\pi}\left[\frac{1}{1-R_0^2\cos^2\vartheta} + R_0\cos\vartheta\,\frac{\pi/2+\arcsin(R_0\cos\vartheta)}{[1-R_0^2\cos^2\vartheta]^{3/2}}\right], \tag{31}$$

где $R_0 = \Gamma_{EE^*}(\rho) = \exp[-D_{S_{\rm м}}(\rho)/2]$, $|\vartheta| \le \pi$.

По определению искомая структурная функция равна

$$D_{\varphi_{\rm g}}(\rho,z) = \int\limits_{-\infty}^{+\infty} \vartheta^2 W(\vartheta)\,{\rm d}\vartheta, \tag{32}$$

где $W(\vartheta)$ — плотность распределения флуктуаций разности фаз нормального случайного процесса во всей области определения ϑ.

Из [14] следует, что при выполнении неравенства $R_0 \ge 0{,}95$ (соответственно, из соотношения (31) имеем $D_{S_{\rm м}}(\rho) \le 10^{-1}$) функция $W_h(\vartheta)$ практически полностью сосредоточена в области

$|\vartheta| \le \pi/2$, т. е., другими словами, при выполнении неравенства $D_{S_\text{м}}(\rho) \le 10^{-1}$ функция $W(\vartheta)$ может быть аппроксимирована функцией $W_h(\vartheta)$ на интервале $[-\pi/2, \pi/2]$. Учитывая это обстоятельство и проводя необходимые интегральные преобразования, из соотношений (31), (32) получаем следующее приближённое соотношение (при $D_{S_\text{м}}(\rho) \le 10^{-1}$):

$$D_{\varphi_\text{g}}(\rho) \approx \int\limits_{-\pi/2}^{+\pi/2} \vartheta^2 W_h(\vartheta)\,\mathrm{d}\vartheta \approx -\frac{1}{2}\, D_{S_\text{м}}(\rho) \ln[D_{S_\text{м}}(\rho)]. \qquad (33)$$

Из соотношения (33) следует, что в области относительно малых ρ, когда $D_{S_\text{м}}(\rho) \le 10^{-1}$, величина структурной функции фазовых флуктуаций принимаемого излучения в режиме насыщенных мерцаний больше не только, чем в условиях слабых мерцаний (29), но и чем величина структурной функции мелкомасштабных фазовых флуктуаций $D_{S_\text{м}}(\rho)$ на экране.[3] Фактически, речь идёт о расстояниях $\rho < \rho_\text{кор} = \sqrt{z/k}\ [D_S(\sqrt{z/k})]^{-1/(p-2)}$, поскольку в режиме насыщенных мерцаний $D_{S_\text{м}}(\rho) \approx 1$ (см. (16)). Кроме того, следует иметь в виду, что формула (33) получена с помощью соотношения (31), поэтому область её применимости ограничена значениями $D_{\varphi_\text{g}}(\rho) \le 1$.

Полученные выше результаты имеют простое физическое толкование. Дело в том, что в зоне насыщенных мерцаний за хаотическим фазовым экраном дисперсия (средний квадрат) сильных фазовых флуктуаций практически совпадает с её величиной на экране [15], а коэффициент пространственной корреляции этих флуктуаций при малом разнесении точек наблюдения, хотя и близок к единице, но всё же немного меньше, чем на экране. В конечном счёте это и приводит к тому, что структурная функция флуктуаций фазы при малом пространственном разнесении точек наблюдения в зоне насыщенных мерцаний за сильным фазовым экраном становится заметно большей, чем на экране (см. (33)). В случае слабых флуктуаций в зоне Фраунгофера дисперсия фазовых флуктуаций равна дисперсии амплитудных флуктуаций, которые возникают при распространении волн в свободном пространстве за слабым фазовым экраном, а сумма их равна дисперсии флуктуаций фазы на экране [1, 5].[4] За счёт уменьшения вдвое дисперсии фазовых флуктуаций в плоскости наблюдения примерно в два раза уменьшается структурная функция фазовых флуктуаций принимаемого излучения по сравнению с её значением на экране (см. (29)).

Наряду со структурной функцией фазовых флуктуаций, в современных экспериментальных исследованиях широко используются измерения спектральных характеристик флуктуаций фазы [6]. Фактически, они являются фурье-преобразованиями от соответствующих кросс- и автокорреляционных характеристик флуктуаций фазы принимаемого излучения. В частности, при радиоинтерферометрических измерениях это функция пространственно-временно́й корреляции флуктуаций разности фаз излучения, принимаемого отдельными элементами интерферометра с базой $\boldsymbol{\rho}$ [10, 16]:

$$\Gamma_{\Delta\varphi}(\tau) = \frac{1}{2}\,[D_\varphi(\boldsymbol{\rho} - \mathbf{V}_\perp\tau) + D_\varphi(\boldsymbol{\rho} + \mathbf{V}_\perp\tau) - 2D_\varphi(\mathbf{V}_\perp\tau)], \qquad (34)$$

где $\mathbf{V}_\perp$ — скорость дрейфа дифракционной картины в плоскости наблюдения.

[3] Это утверждение хорошо соответствует результатам прямого численного моделирования структурной функции фазовых флуктуаций радиоволн за турбулентным фазовым экраном в зоне насыщенных мерцаний [9].

[4] В зоне насыщенных мерцаний за сильным фазовым экраном дисперсия относительных амплитудных флуктуаций порядка единицы, но она мала по сравнению с дисперсией сильных фазовых флуктуаций принимаемого излучения в плоскости наблюдения [15].

С учётом соотношения (30) можно утверждать, что в условиях слабых мерцаний ($D_S(\sqrt{z/k}) \ll \ll 1$) функция $\Gamma_{\Delta\varphi}(\tau)$ (а значит, и соответствующий частотный спектр $S_{\Delta\varphi}(\Omega)$ фазовых флуктуаций) будет практически совпадать с функцией пространственно-временно́й корреляции флуктуаций разности фаз $\Gamma_{\Delta S}(\tau)$ (или с её спектром $S_{\Delta S}(\Omega)$) на выходе оптически тонкого слоя (фазового экрана):

$$S_{\Delta\varphi}(\Omega) = \frac{1}{\pi} \int\limits_0^\infty \Gamma_{\Delta\varphi}(\tau) \cos(\Omega\tau)\,\mathrm{d}\tau \approx \frac{1}{\pi} \int\limits_0^\infty \Gamma_{\Delta S}(\tau) \cos(\Omega\tau)\,\mathrm{d}\tau = S_{\Delta S}(\Omega). \tag{35}$$

Нетрудно убедиться, что для режима слабых мерцаний подобный результат оказывается справедливым и в случае, когда спектр фазовых флуктуаций излучения измеряется с помощью одиночного радиотелескопа ($S_\varphi(\Omega) \approx S_S(\Omega)$).

4. ЗАКЛЮЧЕНИЕ

Проведённый анализ произвольных (в том числе сильных) фазовых флуктуаций радиоволн за турбулентным оптически тонким неоднородным слоем (фазовым экраном) позволяет утверждать, что в режиме слабых мерцаний статистические характеристики флуктуаций фазы принимаемого излучения (структурная функция, спектры флуктуаций при одиночном и интерферометрическом приёме) практически совпадают с соответствующими статистическими характеристиками флуктуаций фазы на выходе фазового экрана. Это обстоятельство позволяет корректно интерпретировать результаты современных прямых измерений флуктуаций фазы радиоволн в экспериментах по дистанционному зондированию околоземной и космической плазмы, в том числе и с применением радиоинтерферометров со сверхдлинными базами [6–9].

В режиме насыщенных мерцаний структурная функция фазовых флуктуаций принимаемого излучения практически совпадает со структурной функцией флуктуаций фазы на выходе фазового экрана лишь для больших расстояний между точками наблюдения. При сравнительно малых расстояниях между точками наблюдения величина структурной функции фазовых флуктуаций принимаемого излучения больше, чем величина структурной функции мелкомасштабных фазовых флуктуаций на экране. Вопрос о фазовых флуктуациях дифракционной компоненты рассеянного поля за турбулентным фазовым экраном в режиме насыщенных мерцаний требовал специального рассмотрения, которое было проведено в [15].

Работа выполнена в рамках проекта РФФИ (грант № 03–02–17303).

СПИСОК ЛИТЕРАТУРЫ

1. Рытов С. М., Кравцов Ю. А., Татарский В. И. Введение в статистическую радиофизику. Т. 2. М.: Наука, 1978.
2. Гершман Б. Н., Ерухимов Л. М., Яшин Ю. Я. Волновые явления в ионосфере и космической плазме. М.: Наука, 1984.
3. Прохоров А. М., Бункин Ф. В., Гочелашвили К. С., Шишов В. И. // УФН. 1974. Т. 114, вып. 3. С. 415.
4. Якушкин И. Г. // Изв. вузов. Радиофизика. 1985. Т. 28, № 5. С. 535.
5. Денисов Н. Г. // Изв. вузов. Радиофизика. 1961. Т. 4, № 4. С. 630.
6. Яковлев О. И. Космическая радиофизика. М.: Научная книга, 1998.
7. Grall R. R., Coles W. A., Spangler S. R., Sakurai T. // J. Geophys. Res. 1997. V. 102, No. 1. P. 263.
8. Spangler S. R., Sakurai T. // Astrophys. J. 1995. V. 445. P. 999.

9. Coles W. A., Liu W., Harmon J. K., Martin C. L. // J. Geophys. Res. 1991. V. 96, № 2. P. 1745.
10. Алимов В. А., Гавриленко В. Г., Липатов Б. Н., Нечаева М. Б. // Изв. вузов. Радиофизика. 2004. Т. 47, № 3. С. 167.
11. Gochelashvily K. S., Shishov V. I. // Optica Acta. 1991. V. 18, No. 4. P. 313.
12. Алимов В. А., Рахлин А. В., Выборнов Ф. И. // Изв. вузов. Радиофизика. 1997. V. 40, № 11. С. 1323.
13. Booker H. G., Majidiahi G. // J. Atm. Terr. Phys. 1981. V. 43, No. 11. P. 1199.
14. Левин Б. Р. Теоретические основы статистической радиотехники. Т. 1. М.: Сов. радио, 1974.
15. Алимов В. А., Рахлин А. В. // Изв. вузов. Радиофизика. 2005. Т. 48, № 4 С. 275.
16. Татарский В. И. Распространение радиоволн в турбулентной атмосфере. М.: Наука, 1967.

Поступила в редакцию 12 декабря 2003 г.; принята в печать 28 апреля 2005 г.

ON PHASE FLUCTUATIONS OF RADIO WAVES AFTER A TURBULENT PHASE SCREEN

V. A. Alimov and A. V. Rakhlin

We consider the problem on phase fluctuations of radio waves having passed a turbulent optically thin inhomogeneous layer (phase screen). Expressions are derived for main statistical characteristics of the phase fluctuations obtained using modern direct methods based on measurements of complex field of the received radiation. Regimes of weak and saturated scintillations of a signal are analyzed. It is shown that in the case of weak scintillations, the statistical characteristics of phase fluctuations of the received radiation, such as the structure function and the fluctuation spectra upon single-point and interference reception, practically coincide with those at the screen output. In the regime of saturated scintillations, we obtain an information on the structure function of phase fluctuations of the received radiation in the cases of relatively large and small spatial separations of the reception points. The results obtained allow one to correctly analyze data of modern experiments on remote sensing of the near-Earth and space plasmas with help of direct phase methods for diagnosing the environment.

УДК 533.951

О ГРАДИЕНТНО-ТОКОВОМ МЕХАНИЗМЕ ОБРАЗОВАНИЯ НЕОДНОРОДНОЙ СТРУКТУРЫ ВЫСОКОШИРОТНОЙ ВЕРХНЕЙ ИОНОСФЕРЫ

Е. Н. Мясников

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

В приближении двухжидкостной магнитной гидродинамики рассмотрена градиентно-токовая неустойчивость неоднородной магнитоактивной плазмы. В отличие от известных градиентно-дрейфовой и токово-конвективной неустойчивостей, градиентно-токовая неустойчивость связана с генерацией непотенциальных квазистатических электрических полей, поляризованных ортогонально к внешнему магнитному полю $\mathbf{B}_0$ и возбуждаемых вихревыми токами, вектор плотности которых лежит в плоскости, проходящей через векторы магнитного поля $\mathbf{B}_0$ и крупномасштабного градиента электронной концентрации. Показано, что в высокоширотной верхней ионосфере в областях, содержащих крупномасштабные токи, втекающие и вытекающие из ионосферы вдоль магнитного поля, градиентно-токовая неустойчивость может приводить к появлению плоско-слоистых (sheet-like) неоднородностей, вытянутых преимущественно в плоскости, проходящей через векторы геомагнитного поля и скорости регулярного дрейфа плазмы.

ВВЕДЕНИЕ

Впервые неустойчивость вертикально-неоднородного слоя плазмы, находящегося во взаимно ортогональных магнитном и гравитационном полях, была предложена Данжи для объяснения неоднородной структуры экваториальной ионосферы [1]. Аналогичная неустойчивость неоднородной плазмы в скрещенных электрическом $\mathbf{E}_0$ и магнитном $\mathbf{B}_0$ полях либо при наличии поперечного к $\mathbf{B}_0$ нейтрального ветра в дальнейшем была названа обобщённой градиентно-дрейфовой (ГД) [2]. Как известно, причиной развития ГД неустойчивости являются поляризационные электрические поля, вызывающие дрейф возмущений плотности плазмы в плоскости, ортогональной к $\mathbf{B}_0$, приводящий к их конвективному росту на фоне крупномасштабного градиента электронной концентрации. Неустойчивость неоднородной плазмы с продольным (вдоль магнитного поля) током впервые была расмотрена Кадомцевым и Недоспасовым и названа токово-конвективной (ТК) [3]. В [4] токово-конвективная неустойчивость была рассмотрена применительно к условиям высокоширотной ионосферы. Учёт дополнительного нагрева неоднородностей плазмы за счёт флуктуационных продольных токов может привести к понижению порога ТК неустойчивости и развитию термомагнитной неустойчивости [5].

Основная трудность, с которой приходится сталкиваться при рассмотрении ГД неустойчивости в верхней ионосфере, связана с эффектом «короткого замыкания», который приводит к резкому уменьшению амплитуды поляризационного электрического поля (а следовательно, и инкремента) в случае, когда степень вытянутости возмущений в направлении $\mathbf{B}_0$ не является экстремально большой. В F-слое максимальный инкремент ГД неустойчивости имеет место при $l_\parallel/l_\perp \gtrsim 10^3\div10^4$, откуда, в частности, следует, что на высоких и средних широтах при поперечном масштабе неоднородностей $l_\perp \gtrsim 100$ м их продольный масштаб должен превышать вертикальный размер ионосферного слоя. Последнее противоречит экспериментальным данным, полученным методом радиопросвечивания ионосферы сигналами искусственных спутников, согласно которым степень вытянутости километровых неоднородностей в F-слое составляет $l_\parallel/l_\perp \lesssim 10\div30$ [6, 7].

Чтобы избежать указанного противоречия, в большинстве теоретических работ при рассмотрении ГД неустойчивости используется предположение об эквипотенциальности силовых линий поля $\mathbf{B}_0$. Последнее позволяет проинтегрировать исходные уравнения вдоль силовых линий магнитного поля и, положив продольные токи на границе возмущения равными нулю, исключить их влияние на амплитуду поляризационного электрического поля (см. [8, 9]).

Вместе с тем измерения, проведённые с космических аппаратов, свидетельствуют о том, что в высокоширотной верхней ионосфере и магнитосфере мелкомасштабные продольные токи оказывают непосредственное влияние на генерацию низкочастотных электрических и магнитных полей [10]. Эксперимент показывает, что отношение поперечных к $\mathbf{B}_0$ пространственных спектральных компонент магнитного и электрического полей обратно пропорционально поперечному волновому числу. Это отличает их от полей, возбуждаемых низкочастотными магнитогидродинамическими волнами, для которых в приближении одножидкостной магнитной гидродинамики отношение спектральных компонент магнитного и электрического полей остаётся постоянным в широком диапазоне поперечных волновых чисел [11].

Одна из возможностей согласовать теорию с экспериментальными данными состоит в рассмотрении механизмов образования мелкомасштабных квазистатических электрических полей в приближении двухжидкостной магнитной гидродинамики. Таким механизмом может явиться эффект Холла, возникающий при протекании в плоскости, ортогональной к $\mathbf{B}_0$, тока, заданного внешним источником. В отличие от потенциального поляризационного поля, к возникновению которого приводит частичное запирание флуктуационного тока, индукционное электрическое поле может возбуждаться вследствие изменяющегося во времени магнитного потока через контуры проводящей жидкости, связанные с движущимися со скоростями порядка дрейфовой возмущениями плотности плазмы. Аналогичная задача была рассмотрена в [12, 13], где в качестве источника дополнительного электрического поля рассматривался диамагнитный ток, определяющий состояние квазистатического равновесия возмущений плотности плазмы в магнитном поле. Было показано, что необходимыми условиями генерации индукционного электрического поля являются дифференциальное вращение возмущения плотности плазмы в плоскости, ортогональной к $\mathbf{B}_0$, и нарушение его отражательной симметрии относительно внешнего магнитного поля. В теории магнитного динамо указанные условия рассматриваются в качестве необходимых для генерации магнитного поля в движущейся проводящей среде [14].

В данной работе рассмотрен механизм генерации электрических полей за счёт вихревых токов, протекающих в плоскости, проходящей через направление внешнего магнитного поля. В случаях слабо и сильно ионизованной плазмы найдены выражения для инкрементов ГТ неустойчивости, возникающей при наличии градиента электронной концентрации, лежащего в плоскости замыкания тока. Показано, что ГТ неустойчивость может приводить к образованию наблюдаемых в верхней высокоширотной ионосфере плоско-слоистых (sheet-like) неоднородностей электронной концентрации, вытянутых в плоскости, проходящей через векторы геомагнитного поля и скорости регулярного дрейфа плазмы [15].

1. КВАЗИГИДРОДИНАМИЧЕСКОЕ ПРИБЛИЖЕНИЕ, ГРАДИЕНТНО-ДРЕЙФОВАЯ И ТОКОВО-КОНВЕКТИВНАЯ НЕУСТОЙЧИВОСТИ

Будем описывать квазинейтральную плазму с помощью системы уравнений двухжидкостной квазигидродинамики (см. [2, 16]):

$$\partial n/\partial t + \operatorname{div}(n\mathbf{v}_\alpha) = 0, \tag{1}$$

$$m_\alpha n \frac{\mathrm{d}\mathbf{v}_\alpha}{\mathrm{d}t} = e_\alpha n\mathbf{E} + e_\alpha n \frac{[\mathbf{v}_\alpha \times \mathbf{B}]}{c} - \nabla p_\alpha - \frac{\nu_{\mathrm{ei}} m_{\mathrm{e}}}{e_\alpha}\mathbf{j} - m_\alpha n \nu_{\alpha\mathrm{n}} \mathbf{v}_\alpha, \tag{2}$$

где e_α — заряд частицы сорта α, положительный для иона ($\alpha = \mathrm{i}$) и отрицательный для электрона ($\alpha = \mathrm{e}$), n — концентрация, $p_\alpha = nT_\alpha$ — давление частиц сорта α, T_α — их температура, выраженная в энергетических единицах, $\mathbf{v}_\alpha$ — скорость заряженных частиц сорта α, $\mathbf{E}$ — напряжённость электрического поля, $\mathbf{B}_0$ — индукция магнитного поля, $\mathbf{j} = en\,(\mathbf{v}_{\mathrm{i}} - \mathbf{v}_{\mathrm{e}})$ — плотность тока, $\nu_{\alpha\mathrm{n}}$ — частота соударений заряженных частиц с нейтральными, ν_{ei} — частота электрон-ионных соударений, c — скорость света. Рассмотрим случай, когда плазма находится во внешнем однородном магнитном поле, направленном вдоль оси z, давление которого существенно больше газокинетического: $\beta = 8\pi n\,(T_{\mathrm{e}} + T_{\mathrm{i}})/B_0^2 \ll 1$, и между частотами соударений заряженных частиц $\nu_{\mathrm{e}} = \nu_{\mathrm{ei}} + \nu_{\mathrm{en}}$, ν_{in} и их гирочастотами $\omega_{B\alpha} = eB_0/(m_\alpha c)$ выполнены соотношения

$$\frac{\nu_{\mathrm{e}}}{\omega_{B\mathrm{e}}} \ll \frac{\nu_{\mathrm{in}}}{\omega_{B\mathrm{i}}} \ll 1. \tag{3}$$

Здесь $e > 0$ — элементарный заряд. Если в плазме задано внешнее однородное электрическое поле $\mathbf{E}_0$, вектор которого лежит в плоскости xz, то в стационарном состоянии $\mathrm{d}\mathbf{v}_\alpha/\mathrm{d}t = 0$ из уравнений движения (2) следуют выражения для скорости дрейфа плазмы в скрещенных электрическом $\mathbf{e}_x E_{0x}$ и магнитном $\mathbf{B}_0$ полях:

$$v_{0y} = -cE_{0x}/B_0, \tag{4}$$

и плотности тока (закон Ома):

$$\mathbf{j}_0 = en\,(\mathbf{v}_{\mathrm{i}0} - \mathbf{v}_{\mathrm{e}0}) = \sigma_{\mathrm{P}} E_{0x}\mathbf{e}_x + \sigma_{\mathrm{H}} E_{0x}\mathbf{e}_y + \sigma_\parallel E_{0z}\mathbf{e}_z. \tag{5}$$

Здесь σ_{P} — педерсеновская, σ_{H} — холловская, $\sigma_\parallel$ — продольная проводимости соответственно, $\mathbf{e}_x$, $\mathbf{e}_y$ и $\mathbf{e}_z$ — орты осей декартовой системы координат. При выполнении условий (3) имеют место следующие выражения для компонент тензора проводимости плазмы:

$$\sigma_\parallel = \frac{e^2 n}{m_{\mathrm{e}}\nu_{\mathrm{e}}}, \qquad \sigma_{\mathrm{P}} \approx \frac{\nu_{\mathrm{e}}\nu_{\mathrm{in}}}{\omega_{B\mathrm{e}}\omega_{B\mathrm{i}}}\,\sigma_\parallel, \qquad \sigma_{\mathrm{H}} \approx \frac{\nu_{\mathrm{in}}}{\omega_{B\mathrm{i}}}\,\sigma_{\mathrm{P}}. \tag{6}$$

В силу соотношений $\sigma_{\mathrm{H}} \ll \sigma_{\mathrm{P}} \ll \sigma_\parallel$ вектор плотности протекающего в плазме тока (5) лежит приблизительно в той же плоскости xz, что и поле $\mathbf{E}_0$.

Выражение для потенциального поляризационного электрического поля $\mathbf{E} = -\nabla\varphi$ находится из уравнения непрерывности плотности полного тока, протекающего в плазме:

$$\operatorname{div}\mathbf{j} = 0. \tag{7}$$

Для возмущений в виде плоских волн $\exp(i\mathbf{kr})$ из линеаризованного уравнения (7) получим

$$\mathbf{E}_{\mathbf{k}} = -i\mathbf{k}\varphi_{\mathbf{k}} = -\mathbf{k}\ \frac{\mu k_x E_{0x} + k_z E_{0z}}{\mu k_\perp{}^2 + k_z^2}\ \delta n_{\mathbf{k}}, \tag{8}$$

где $\delta n_{\mathbf{k}} = n_{\mathbf{k}}/n$ — спектральная компонента относительных возмущений концентрации плазмы, $\mu = \sigma_{\mathrm{P}}/\sigma_\parallel \ll 1$ — малый параметр, $\mathbf{E}_{\mathbf{k}}$ и $\varphi_{\mathbf{k}}$ — спектральные компоненты электрического поля и потенциала соответственно, $\mathbf{k} = \mathbf{e}_x k_x + \mathbf{e}_y k_y + \mathbf{e}_z k_z$ — волновой вектор, $k_\perp^2 = k_x^2 + k_y^2$. Если в плазме задано поперечное электрическое поле E_{0x}, из (8) следует, что максимальное поляризационное поле $E_{\mathbf{k}x} \approx -E_{0x}\,\delta n_{\mathbf{k}}$ соответствует возмущениям, волновой вектор которых составляет малый угол $\theta \approx k_z/k_x \lesssim \mu^{1/2} \ll 1$ с нормалью к полю $\mathbf{B}_0$ и направлен вдоль регулярного электрического поля: $k_y \ll k_x$. Из (8) следует, что результирующее электрическое поле усиливается в отрицательных возмущениях концентрации ($\delta n_\mathbf{k} < 0$) и ослабляется в положительных ($\delta n_\mathbf{k} > 0$). При наличии продольного электрического поля E_{0z} максимальное поле $E_{\mathbf{k}x} \approx -E_{0z}\,\delta n_\mathbf{k}/(2\mu^{1/2})$ имеет место при условии $\theta = \theta_\mathrm{m} = \mu^{1/2}$.

Рассмотрим неоднородный вдоль оси y слой плазмы, характерный масштаб которого $L_y = n\,|\partial n/\partial y|^{-1}$ удовлетворяет условию $kL_y \ll 1$. Подставив выражения для спектральных компонент флуктуационной скорости дрейфа возмущений плотности плазмы

$$v_{\mathbf{k}x} = cE_{\mathbf{k}y}/B_0\,, \qquad v_{\mathbf{k}y} = -cE_{\mathbf{k}x}/B_0 \tag{9}$$

в линеаризованное уравнение непрерывности (1) для возмущений вида $\exp(-i\omega t)$:

$$-i\omega\,\delta n_\mathbf{k} + ik_x v_{\mathbf{k}x} + ik_y v_{\mathbf{k}y} + v_{\mathbf{k}y}/L_y = 0, \tag{10}$$

приходим к выражениям для частоты и инкрементов ГД и ТК неустойчивостей:

$$\omega = 0, \qquad \gamma = \gamma_\text{ГД} + \gamma_\text{ТК} = \frac{c}{B_0 L_y}\,\frac{k_x}{(\mu k_\perp^2 + k_z^2)}\,(\mu k_x E_{0x} + k_z E_{0z}). \tag{11}$$

Максимальный инкремент ГД неустойчивости равен $\gamma_\text{ГД} = v_{0y}/L_y$. Для неустойчивости необходимо, чтобы направления градиента концентрации и скорости дрейфа плазмы совпадали. В этом случае отрицательное возмущение дрейфует в область с большей, а положительное — с меньшей концентрацией фоновой плазмы, что отвечает неустойчивости конвективного типа. Максимальный инкремент ГД неустойчивости имеет место для возмущений, волновой вектор которых составляет с нормалью к $\mathbf{B}_0$ угол $\theta \approx k_z/k_x \lesssim \mu^{1/2}$ и ортогонален к направлению регулярного градиента плазмы: $\mathbf{k} \perp \nabla_y n$, где $\nabla_y n = \mathbf{e}_y\,\partial n/\partial y$.

Максимальный инкремент ТК неустойчивости $\gamma_\text{ТК} = [m_\mathrm{e}\nu_\mathrm{e}/(m_\mathrm{i}\nu_i)]^{1/2} v_{0z}/(2L_y)$ реализуется при $\theta = \mu^{1/2}$ и $k_y \to 0$, где $v_{0z} = eE_{0z}/(m_\mathrm{e}\nu_\mathrm{e})$ — скорость электрона в поле E_{0z}. Инкремент ТК неустойчивости пропорционален произведению $k_x k_z$, поэтому она может развиваться при любом направлении регулярного градиента концентрации.

Оценим возмущение магнитного поля, возникающее при развитии ГД и ТК неустойчивостей. Для этого подставим спектральную плотность тока $\mathbf{j}_\mathbf{k}$, найденную из линеаризованного уравнения (5) с учётом электрического поля (8), в уравнение

$$i\,[\mathbf{k} \times \mathbf{B}_\mathbf{k}] = 4\pi \mathbf{j}_\mathbf{k}/c. \tag{12}$$

Возмущение магнитного поля содержит две взаимно ортогональные компоненты, одна из которых, $B_{\mathbf{k}\theta} \approx B_{\mathbf{k}z}$, лежит в плоскости, проходящей через вектор $\mathbf{B}_0$, другая, $B_{\mathbf{k}\psi}$, ортогональна $\mathbf{B}_0$. В случае, когда задано поле E_{0x}, компоненты магнитного поля описываются выражениями

$$B_{\mathbf{k}\theta} = -i\,\frac{k_y}{k_\perp^2}\,\frac{4\pi\sigma_\perp}{c}\,E_{0x}\,\delta n_\mathbf{k}, \qquad \mathbf{B}_{\mathbf{k}\psi} = i\,\frac{k_x k_z}{k_\perp^4}\,\frac{4\pi\sigma_\parallel}{c}\,E_{0x}\,(k_y \mathbf{e}_x - k_x \mathbf{e}_y)\,\delta n_\mathbf{k}. \tag{13}$$

При оптимальном для возбуждения ГД и ТК неустойчивостей условии $\theta \approx \mu^{1/2}$ имеет место соотношение $B_{\mathbf{k}\psi} \gg B_{\mathbf{k}z}$. В случае $k_z = 0$ продольные токи отсутствуют, и возмущение магнитного поля содержит только компоненту $B_{\mathbf{k}z}$.

2. ДВУХЖИДКОСТНАЯ МАГНИТНАЯ ГИДРОДИНАМИКА, ГРАДИЕНТНО-ТОКОВАЯ НЕУСТОЙЧИВОСТЬ ПЛАЗМЫ

2.1. Система квазистатических крупномасштабных возмущений

Рассмотрим систему крупномасштабных возмущений плазмы, показанную на рис. 1, которая является характерной для авроральной верхней ионосферы в северном полушарии. Ось z системы координат направлена вдоль магнитного поля $\mathbf{B}_0$ и близка к вертикали, ось x направлена с юга на север и лежит в плоскости геомагнитного меридиана, ось y направлена на восток. Будем считать, что через плазму проходит крупномасштабная система токов, причём токи, втекающие в ионосферу, находятся на южной стороне аврорального овала, а вытекающие — на северной, что характерно для дневных и вечерних условий [2, 17]. Данной системе токов отвечает крупномасштабное возмущение магнитного поля B_{0y}, которое может быть определено из уравнений

$$j_{0x} = -\frac{c}{4\pi}\,\frac{\partial B_{0y}}{\partial z}\,, \qquad j_{0z} = \frac{c}{4\pi}\,\frac{\partial B_{0y}}{\partial x}\,. \tag{14}$$

Будем считать что в плазме существует неоднородное по x электрическое поле $E_{0x}(x)$, направленное с юга на север, которому соответствует скорость дрейфа

$$v_{0y}(x) = -cE_{0x}/B_0\,, \tag{15}$$

направленная с востока на запад. Как показывают измерения, в зоне существования продольных токов фоновая плазма неоднородна в направлении север—юг, в частности, здесь существует область сильного обеднения плазмы, называемая главным ионосферным провалом электронной

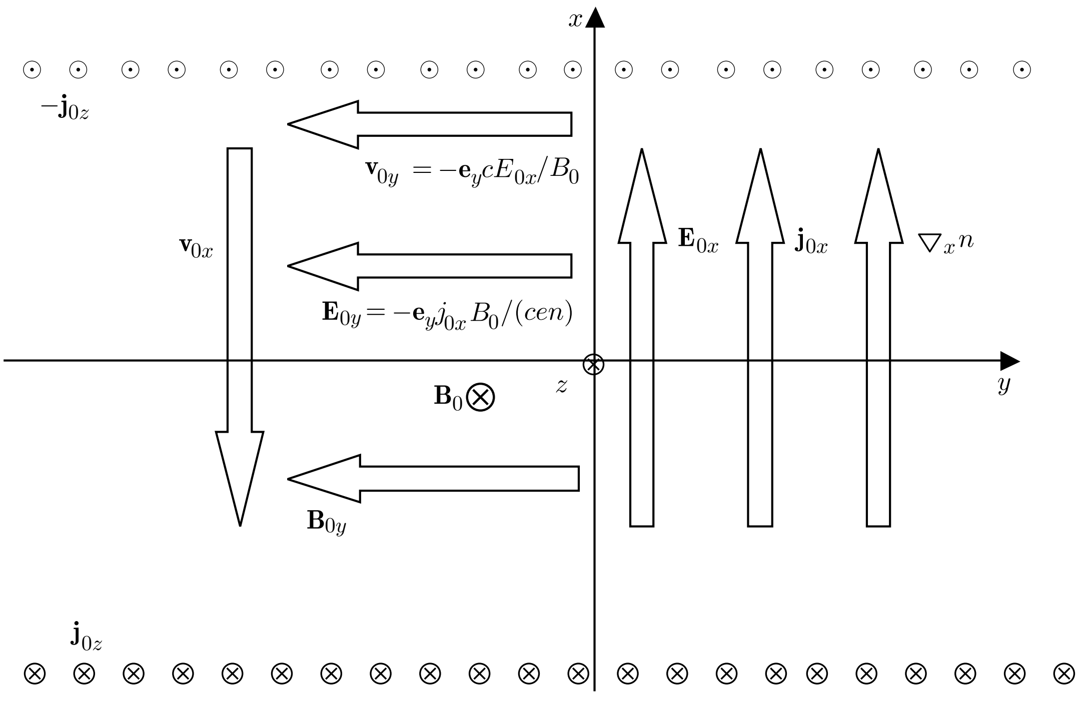


Рис. 1. Схема квазистатических крупномасштабных полей и токов в авроральной верхней ионосфере для вечернего времени суток в северном полушарии

концентрации. При этом как крупномасштабный градиент $\nabla_x n$, так и волновые векторы $\mathbf{k}$ образующихся мелкомасштабных неоднородностей ориентированы преимущественно в направлении электрического поля $\mathbf{e}_x E_{0x}$ и ортогональны к скорости дрейфа плазмы. Будем считать концентрацию $n(x,z)$ фоновой плазмы неоднородной по координатам x и z. Тогда условие непрерывности тока

$$\frac{\partial j_{0x}}{\partial x} = -\frac{\partial j_{0z}}{\partial z} \tag{16}$$

может быть выполнено только за счёт существования продольных токов, которые обеспечивают условие квазинейтральности плазмы. Примем приведённую на рис. 1 систему крупномасштабных возмущений в качестве исходного состояния квазистатического равновесия для фоновой плазмы.

2.2. Квазистатическое индукционное электрическое поле

В соответствии с приближением двухжидкостной магнитной гидродинамики определим поперечное к $\mathbf{B}_0$ электрическое поле из стационарного уравнения движения для электронной компоненты плазмы (см. [16], с. 32):

$$\nabla p_{\rm e} = -en\mathbf{E} - \frac{en}{c}\,[\mathbf{v}\times\mathbf{B}_0] + \frac{1}{c}\,[\mathbf{j}_\perp\times\mathbf{B}_0] - \frac{en\mathbf{j}}{\sigma_\parallel}\,, \tag{17}$$

где $\mathbf{v}$ — гидродинамическая (дрейфовая) скорость возмущения плотности плазмы, $\mathbf{j}_\perp$ — плотность поперечного тока. Слагаемое $[\mathbf{j}_\perp\times\mathbf{B}_0]/c$ в правой части (17) описывает эффект Холла — возникновение э. д. с. в проводнике с током, помещённом во внешнее магнитное поле. Пренебрегая в (17) градиентом давления и столкновительным слагаемым и переходя в локальную систему отсчёта, где $\mathbf{v}_\perp = 0$, получим

$$\mathbf{E}_\perp = -[\mathbf{v}_{{\rm e}\perp}\times\mathbf{B}_0]/c = [\mathbf{j}_\perp\times\mathbf{B}_0]/(cen). \tag{18}$$

Заметим, что в отличие от электрического поля, рассматриваемого в одножидкостной магнитной гидродинамики, поле (18) не обращается в нуль в локальной системе отсчёта, движущейся вместе с возмущением плотности проводящей жидкости. Будем считать, что в лабораторной системе отсчёта оно приводит к движению возмущения плотности плазмы с дрейфовой скоростью, которая, как и в приближении одножидкостной магнитной гидродинамики, определяется уравнением

$$\mathbf{v}_\perp = c\,[\mathbf{E}_\perp\times\mathbf{B}_0]/B_0^2 = -\mathbf{j}_\perp/(en). \tag{19}$$

Согласно (19) для рассматриваемого типа возмущений дрейфовая скорость всегда направлена противоположно плотности поперечного тока. Поскольку скорость возмущения плотности плазмы совпадает со скоростью более тяжёлой ионной компоненты: $\mathbf{v}_\perp \approx \mathbf{v}_{{\rm i}\perp}$, в силу (19) скорость электронной компоненты относительно ионной в лабораторной системе отсчёта должна иметь то же направление, что и дрейфовая скорость. Отметим, что это не соответствует решению, описывающему движение плазмы под действием потенциального электрического поля. Согласно (5) гидродинамическая (дрейфовая) скорость возмущения и плотность тока определяются ионной компонентой, при этом электронная компонента в силу условия $\nu_{\rm e}/\omega_{B{\rm e}} \ll \nu_{\rm in}/\omega_{B{\rm i}}$ тормозится магнитным полем и не участвует в движении. Заметим, что электрическое поле (18) удовлетворяет определению $\mathbf{E}_\perp = \mathbf{F}_{\rm A}/e$, согласно которому оно равно отношению силы Ампера $\mathbf{F}_A = [\mathbf{j}\times\mathbf{B}_0]/(nc)$, действующей на элемент тока, к заряду e. В отличие от силы Лоренца, которая не совершает работу по перемещению заряженных частиц, сила Ампера может совершать работу и ускорять возмущение плотности плазмы до скорости порядка дрейфовой.

Далее мы покажем, что необходимыми условиями для генерации поля $\mathbf{E}_\perp$ являются распространение в плазме с внешним током дрейфовой МГД волны, фазовая скорость которой пропорциональна разности регулярных скоростей ионной и электронной компонент, определяющих плотность поперечного тока $j_{0\perp} = en\,\Delta v_{0\perp}$, где $\Delta v_{0\perp} = v_{0\mathrm{i}\perp} - v_{0\mathrm{e}\perp}$, и наличие шира дрейфовой скорости, приводящего к нарушению зеркальной симметрии исходного возмущения в плоскости замыкания вихревого квазистатического тока.

2.3. Градиентно-токовая неустойчивость слабо ионизованной плазмы

Рассмотрим неоднородную плазму с концентрацией $n(x,z)$ и внешним током $\mathbf{j}_0(x,z)$, протекающим в плоскости xz. Будем искать решение системы линеаризованных уравнений (14), (16) в виде плоских волн $\exp(ik_x x + ik_z z)$. При выполнении условия $\theta_\mathrm{m} \ll \theta \ll 1$ пренебрежём возмущением тока, вызванным флуктуационным поляризационным полем (8), по сравнению с изменением тока $j_{\mathbf{k}x} = j_{0x}\,\delta n_\mathbf{k}$, связанным с возмущением проводимости. Тогда выражения для плотности поперечного и продольного к $\mathbf{B}_0$ флуктуационных токов принимают вид

$$j_{\mathbf{k}x} = j_{0x}\,\delta n_\mathbf{k} = en\,\Delta v_{0x}\,\delta n_\mathbf{k}, \qquad j_{\mathbf{k}z} = -\frac{k_x}{k_z}\,j_{\mathbf{k}x} = -\frac{k_x}{k_z}\,j_{0x}\,\delta n_\mathbf{k}, \tag{20}$$

где $\Delta v_{0x} = v_{\mathrm{i}0x} - v_{\mathrm{e}0x}$. Заметим, что условие квазинейтральности флуктуаций концентрации может быть выполнено только за счёт протекания продольных токов, благодаря которым заряды на границах возмущений не накапливаются. Флуктуационное магнитное поле, отвечающее плотности тока (20), направлено вдоль оси y, его величина равна

$$B_{\mathbf{k}y} = i\,\frac{4\pi}{ck_z}\,j_{0x}\,\delta n_\mathbf{k}. \tag{21}$$

Используя (18) и (19), получим выражения для электрического поля:

$$E_{\mathbf{k}y} = -\frac{B_0}{cen_0}\,j_{\mathbf{k}x}, \tag{22}$$

и скорости дрейфа возмущений плотности плазмы:

$$v_{\mathbf{k}x} = \frac{cE_{\mathbf{k}y}}{B_0} = -\frac{j_{0x}}{en_0}\,\delta n_\mathbf{k} = -\Delta v_{0x}\,\delta n_\mathbf{k}. \tag{23}$$

Подставляя (23) в линеаризованное уравнение непрерывности (1)

$$-i\omega\,\delta n_\mathbf{k} + ik_x v_{\mathbf{k}x} + v_{\mathbf{k}x}/L_{0x} = 0, \tag{24}$$

приходим к выражениям для частоты и инкремента градиентно-токовой неустойчивости:

$$\omega_{\text{ГТ}} = -k_x\,\frac{j_{0x}}{en_0} = -k_x\,\Delta v_{0x}, \qquad \gamma_{\text{ГТ}} = \frac{j_{0x}}{en_0}\,\frac{\partial n/\partial x}{n_0} = \frac{\Delta v_{0x}}{L_x}\;. \tag{25}$$

Из полученных соотношений следует, что для развития ГТ неустойчивости необходимо, чтобы направления регулярного градиента концентрации $\nabla_x n$ и плотности тока $\mathbf{j}_{0x} = \mathbf{e}_x j_{0x}$ совпадали. Действительно, согласно (23) знак скорости дрейфа возмущения, например, $\delta n_\mathbf{k} < 0$ совпадает со знаком компоненты j_{0x}, при этом оно будет смещаться в направлении возрастания фоновой концентрации. Отметим, что в отличие от ГД неустойчивости, максимальный инкремент которой имеет место для возмущений с волновым вектором, ортогональным к исходному градиенту концентрации, инкремент ГТ неустойчивости максимален для возмущений, волновой вектор которых направлен вдоль крупномасштабного градиента.

Вычислив шир дрейфовой скорости $\boldsymbol{\omega} = \operatorname{rot} \mathbf{v}$, получим

$$\omega_{\mathbf{k}y} = \partial\theta_{\mathbf{k}}/\partial t = ik_z v_{\mathbf{k}x} = -ik_z\, \Delta v_{0x}\, \delta n_{\mathbf{k}}. \tag{26}$$

Выражение (26) можно рассматривать как мнимую поправку к дрейфовой частоте $\omega_{\text{ГТ}}$, пропорциональную относительному возмущению концентрации $\delta n_{\mathbf{k}}$.

2.4. Неустойчивость сильно ионизованной плазмы

Рассмотрим условия равновесия сильно ионизованной магнитоактивной плазмы с внешним током. Предположим, что крупномасштабный вихревой ток $\mathbf{j}_0(x,z)$ однороден в направлении y, причём его линии замкнуты в плоскости xz. Подставим плотность тока $\mathbf{j}$ и величину создаваемого им возмущения магнитного поля B_{0y} в уравнение движения заряженных частиц (2) и сравним слагаемые, отвечающие плотностям нелинейной силы Ампера $\mathbf{f}_{\rm A} = [\mathbf{j}_0 \times \mathbf{e}_y]\, B_{0y}/c$ и силы трения $\mathbf{f}_{\text{тр}\ \alpha} = -n m_\alpha \nu_\alpha \mathbf{v}_\alpha$. Введём выражения для электронного и ионного чисел Рейнольдса как отношения соответствующих инерционного и столкновительного слагаемых в уравнениях движения заряженных частиц:

$$\mathrm{Re}_{\rm e} = \frac{B_{0y}}{B_{0z}}\,\frac{\omega_{B\rm e}}{\nu_{\rm e}}\,, \qquad \mathrm{Re}_{\rm i} = \frac{B_{0y}}{B_{0z}}\,\frac{\omega_{B\rm i}}{\nu_{\rm in} + \nu_{\rm ei} m_{\rm e}/m_{\rm i}}\,. \tag{27}$$

Для сильно ионизованной плазмы ($m_{\rm e}\nu_{\rm ei} \gtrsim m_{\rm i}\nu_{\rm in}$) значения $\mathrm{Re}_{\rm e}$, $\mathrm{Re}_{\rm i}$ совпадают с магнитным числом Рейнольдса:

$$\mathrm{Re}_{\rm e} \approx \mathrm{Re}_{\rm i} \approx \mathrm{Re}_{\rm m} = \Delta v_z\, L_x/D_{\rm m} = \Delta v_x\, L_z/D_{\rm m}. \tag{28}$$

Здесь $D_{\rm m} = c^2/(4\pi\sigma_\|)$ — коэффициент диффузии магнитного поля, $L_x = B_{0y}\,|\partial B_{0y}/\partial x|^{-1}$ и $L_z = B_{0y}\,|\partial B_{0y}/\partial z|^{-1}$ — характерные пространственные масштабы возмущения магнитного поля, создаваемого током $\mathbf{j}_0$.

Рассмотрим случай $\mathrm{Re}_{\rm m} \gg 1$ и предположим, что крупномасштабный вихревой ток возникает вследствие перезамыкания в плоскости xz линий противоположно направленных продольных токов. Будем считать выполненным условие равновесия фоновой изотермической плазмы в магнитном поле, которое определяется уравнением

$$(T_{\rm e} + T_{\rm i}) \left(\mathbf{e}_x\, \frac{\partial}{\partial x} + \mathbf{e}_z\, \frac{\partial}{\partial z} \right) n = \frac{B_{0y}}{c}\ [\mathbf{j}_0 \times \mathbf{e}_y]. \tag{29}$$

Исследуем устойчивость стационарного состояния (29) по отношению к мелкомасштабным неоднородностям плотности плазмы, волновой вектор которых находится в плоскости xz. Линеаризуя (29), получим выражения для плотности мелкомасштабного тока:

$$j_{\mathbf{k}x} = ik_z\, \frac{cn\,(T_{\rm e} + T_{\rm i})}{B_{0y}}\ \delta n_{\mathbf{k}}, \qquad j_{\mathbf{k}z} = -ik_x\, \frac{cn\,(T_{\rm e} + T_{\rm i})}{B_{0y}}\ \delta n_{\mathbf{k}}, \tag{30}$$

и соответствующего ему возмущения магнитного поля:

$$B_{\mathbf{k}y} = -\,\frac{4\pi n\,(T_{\rm e} + T_{\rm i})}{B_{0y}}\ \delta n_{\mathbf{k}}. \tag{31}$$

Далее запишем выражения для флуктуационного электрического поля:

$$E_{\mathbf{k}y} = -ik_z \, \frac{(T_\mathrm{e} + T_\mathrm{i})}{e} \, \frac{B_{0z}}{B_{0y}} \, \delta n_\mathbf{k}, \tag{32}$$

и скорости дрейфа возмущений плотности плазмы:

$$v_{\mathbf{k}x} = -ik_z \, \frac{c\,(T_\mathrm{e} + T_\mathrm{i})}{eB_{0y}} \, \delta n_\mathbf{k}. \tag{33}$$

Подставляя (33) в уравнение непрерывности (24), приходим к выражению для инкремента неустойчивости сильно ионизованной плазмы:

$$\gamma_{\text{ГТ}} = -k_x k_z \, \frac{c\,(T_\mathrm{e} + T_\mathrm{i})}{eB_{0y}} \,. \tag{34}$$

Видно, что инкремент (34), как и инкремент ТК неустойчивости, пропорционален произведению $k_x k_z$, поэтому данная неустойчивость может развиваться при любом направлении поля B_{0y}. Заметим, что компоненту магнитного поля B_{0y} в (34) нельзя устремлять к нулю, т. к. считалось выполненным условие равновесия (29), определяемое равенством изменения газокинетического давления вследствие возмущения плотности плазмы и давления возмущения магнитного поля B_{0y}. Выражение для шира скорости имеет вид

$$\omega_{\mathbf{k}y} = k_z^2 \, \frac{c\,(T_\mathrm{e} + T_\mathrm{i})}{eB_{0y}} \, \delta n_\mathbf{k}. \tag{35}$$

В отличие от (26) частота (35) является действительной и характеризует дифференциальное смещение возмущения плотности плазмы в плоскости xz, которое должно приводить к нарушению его зеркальной симметрии относительно направления внешнего магнитного поля $\mathbf{B}_0$.

3. ОБСУЖДЕНИЕ, СОПОСТАВЛЕНИЕ С ЭКСПЕРИМЕНТАЛЬНЫМИ ДАННЫМИ

Исследуем условия, необходимые для генерации квазистатического электрического поля (18). Последнее, являясь индукционным, должно удовлетворять уравнению

$$-\frac{1}{c} \, \frac{\partial \mathbf{B}'_\mathbf{k}}{\partial t} = i\,[\mathbf{k} \times \mathbf{E}_\mathbf{k}], \tag{36}$$

где $\mathbf{B}'_\mathbf{k}$ — эффективное магнитное поле, возникающее за счёт изменения потока поля $\mathbf{B}_0$ через контуры, связанные с возмущением плотности плазмы. При этом поле $\mathbf{B}'_\mathbf{k}$ должно быть пропорциональным индукции поля $\mathbf{B}_0$ и относительному возмущению электронной концентрации и может возникать за счёт локальных разрежений и сжатий возмущения плотности плазмы $\partial\, \delta n_\mathbf{k}/\partial t$, которые, в частности, в уравнении непрерывности (24) определяются слагаемым $i\mathbf{k}\mathbf{v}_\mathbf{k}$, либо за счёт дифференциального поворота возмущения в плоскости xz, вызывающего шир дрейфовой скорости $\boldsymbol{\omega} = i\,[\mathbf{k} \times \mathbf{v}_\mathbf{k}]$. Действительно, подставив в (36) $\partial \mathbf{B}'_\mathbf{k}/\partial t = \mathbf{B}_0\, \partial\, \delta n_\mathbf{k}/\partial t$ и записав проекцию этого уравнения на ось z, приходим к уравнению непрерывности (24), решение которого определяет дрейфовую частоту и инкремент ГТ неустойчивости (25), (34). Подставляя в (36) выражение $\partial B'_{\mathbf{k}x}/\partial t = B_0\, \partial \theta_\mathbf{k}/\partial t$, где $\theta_\mathbf{k}$ — угол между волновым вектором $\mathbf{k}$ и нормалью к полю $\mathbf{B}_0$, приходим к выражениям (26), (35). Таким образом, мы показали, что для рассмотренного

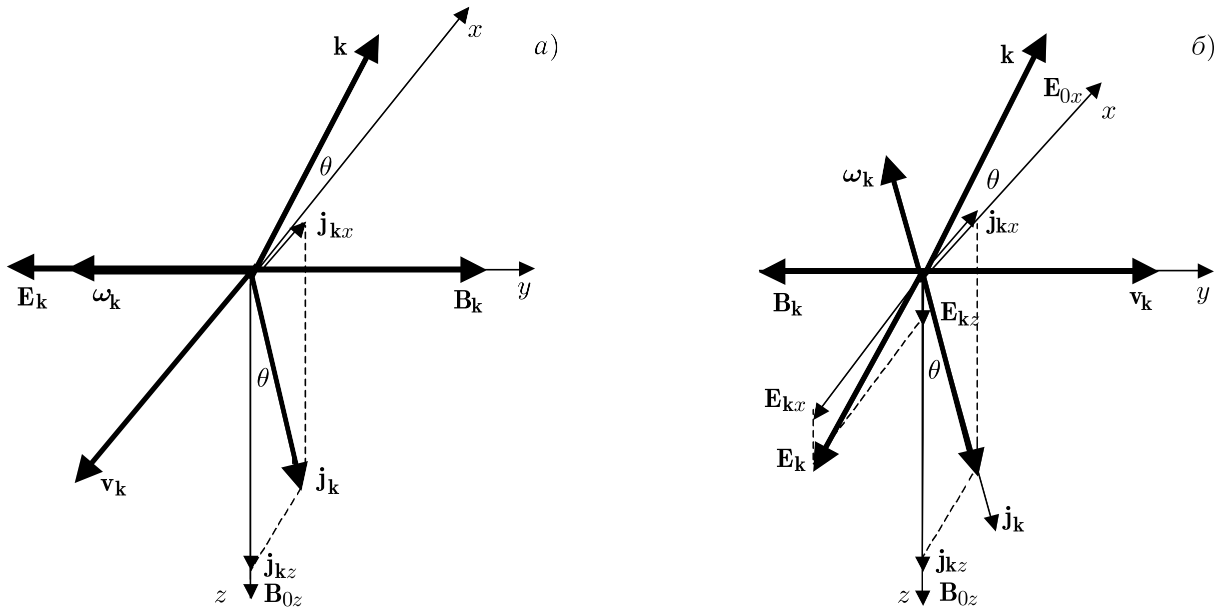


Рис. 2. Поляризация электромагнитных и гидродинамических возмущений в случае ГТ неустойчивости (*а*) и ГД и ТК неустойчивостей при условии $k_z/k_x \sim \sqrt{\nu_\mathrm{e}\nu_\mathrm{in}/(\omega_{B\mathrm{e}}\omega_{B\mathrm{i}})}$ (*б*)

класса возмущений уравнение электромагнитной индукции (36) эквивалентно гидродинамическим уравнениям, описывающим движение возмущений проводящей жидкости в магнитном поле. Отметим, что дифференциальный поворот приводит к нарушению отражательной симметрии исходного возмущения относительно направления $\mathbf{B}_0$. Аналогичная ситуация имеет место в теории магнитного динамо, где дифференциальное вращение и нарушение отражательной симметрии системы, вызываемое движением проводящей среды, рассматриваются в качестве необходимых условий генерации квазистатического магнитного поля (см. [14]).

Рассмотрим поляризацию возмущений для градиентно-токовой, градиентно-дрейфовой и токово-конвекционной неустойчивостей. На рис. 2*а* показаны направления электрического $\mathbf{E}_\mathbf{k}$ и магнитного $\mathbf{B}_\mathbf{k}$ полей, плотности тока $\mathbf{j}_\mathbf{k}$, дрейфовой скорости $\mathbf{v}_\mathbf{k}$ и её шира $\boldsymbol{\omega}_\mathbf{k}$ для ГТ неустойчивости, на рис. 2*б* — для ГД и ТК неустойчивостей в случае, когда отношение поперечных и продольных волновых чисел удовлетворяет условию $k_\parallel/k_\perp \sim \sqrt{\nu_\mathrm{e}\nu_\mathrm{in}/(\omega_{B\mathrm{e}}\omega_{B\mathrm{i}})}$, т. е. токи короткого замыкания являются определяющими. Как видно из рис. 2, при развитии ГТ неустойчивости выполняются условия $\mathbf{E}_\mathbf{k} \parallel \mathbf{B}_\mathbf{k} \parallel \boldsymbol{\omega}_\mathbf{k}$, причём каждое из этих полей ортогонально к $\mathbf{k}$, т. е. является вихревым, а также ортогонально к $\mathbf{j}_{\mathbf{k}\perp}$ и $\mathbf{v}_\mathbf{k}$. Отношение величин магнитного и электрического полей составляет

$$\frac{B_\mathbf{k}}{E_\mathbf{k}} = -\frac{i4\pi en}{k_z B_0}\ . \tag{37}$$

Максимальный инкремент ГТ неустойчивости реализуется для возмущений, у которых проекция волнового вектора на плоскость, ортогональную к $\mathbf{B}_0$, совпадает с направлением регулярного градиента концентрации.

Для ГД и ТК неустойчивостей электрическое поле является потенциальным: $\mathbf{E}_\mathbf{k} \parallel \mathbf{k}$, поэтому при $k_z \neq 0$ оно имеет компоненту, параллельную $\mathbf{B}_0$, вызывающую ток короткого замыкания. Отношение спектральных компонент магнитного и электрического полей в случае, когда ток короткого замыкания является определяющим, составляет

$$\frac{B_{\mathbf{k}}}{E_{\mathbf{k}}} = -\frac{i4\pi\sigma_{\|}k_z}{ck_x^2}\ . \tag{38}$$

Вектор дрейфовой скорости $\mathbf{v_k} \perp \mathbf{k}$, поэтому при развитии ГД и ТК неустойчивостей максимальными инкрементами обладают возмущения, волновой вектор которых ортогонален к регулярному градиенту концентрации. Отметим, что возмущения электромагнитных и гидродинамических полей, определяющих развитие описанных выше неустойчивостей, пропорциональны относительным возмущениям электронной концентрации $\delta n_{\mathbf{k}}$, поэтому соотношения между любыми из них являются линейными и не зависят от характера нелинейных процессов, определяющих форму пространственного спектра.

Рассмотрим зависимости пространственных спектров флуктуационных магнитного и электрического полей от поперечного волнового числа. Как следует из (38), при оптимальном для развития ГД и ТК неустойчивостей условии $k_z/k_x \sim \sqrt{\nu_{\mathrm{e}}\nu_{\mathrm{in}}/(\omega_{B\mathrm{e}}\omega_{B\mathrm{i}})}$ отношение поперечных к $\mathbf{B}_0$ компонент пространственных спектров мощности составляет $B_{\mathbf{k}}^2/E_{\mathbf{k}}^2 \propto k_x^{-2}$, в то время в соответствии с (37) для ГТ неустойчивости $B_{\mathbf{k}}^2/E_{\mathbf{k}}^2 \propto k_z^{-2}$. Поэтому, если продольный масштаб неоднородностей не зависит от поперечного масштаба, как это имеет место при гауссовом обрезании спектра в направлении $\mathbf{B}_0$ [7], либо степень вытянутости неоднородностей уменьшается с ростом их поперечного масштаба, как, например, для искусственных неоднородностей, возбуждаемых мощным радиоизлучением [12], рассматриваемое отношение при развитии ГТ неустойчивости будет иметь более слабую зависимость от поперечного волнового числа $k_\perp$, чем в случаях ГД и ТК неустойчивостей. Согласно локальным измерениям, проведённым в высокоширотной ионосфере с помощью космических аппаратов, в диапазоне масштабов $l_\perp \sim 0{,}1\div10$ км наблюдается зависимость $\langle B_k^2\rangle/\langle E_k^2\rangle \propto {k_\perp}^{-1}$ [11].

Обсудим пространственно-временну́ю картину образования неоднородной структуры высокоширотной верхней ионосферы в районе главного ионосферного провала, представляющего собой крупномасштабную область уменьшения концентрации плазмы с характерным размером L_x от ста и более километров в направлении север—юг, сильно вытянутую в направлении восток—запад. В качестве примера рассмотрим данные, полученные при радиопросвечивании ионосферы когерентными сигналами искусственных спутников земли (ИСЗ), имеющих полярные круговые орбиты, на частотах 150 и 400 МГц. Результаты измерений позволяют сопоставить местоположение локальных сцинтилляционных пэтчей, представляющих собой области с повышенным по сравнению с фоновым уровнем флуктуаций амплитуды сигналов, и крупномасштабных градиентов полного электронного содержания (ПЭС) — интеграла электронной концентрации вдоль луча ИСЗ. Мониторинг состояния ионосферы с использованием ряда последовательных (в течение суток) пролётов ИСЗ показывает, что при переходе от дневных к ночным часам наблюдений имеет место смещение главного ионосферного провала в сторону южных широт, которое часто сопровождается увеличением градиента ПЭС на северной стороне главного ионосферного провала, где концентрация плазмы возрастает в сторону высоких широт. Именно в области резкого северного градиента ПЭС в вечерних и ночных условиях возникают интенсивные сцинтилляционные пэтчи [18, 19].

Описанная морфологическая картина образования мелкомасштабных неоднородностей плазмы в районе главного ионосферного провала согласуется с предложенным ниже сценарием развития градиентно-токовой неустойчивости. Как известно, пространственное положение главного ионосферного провала совпадает с положением системы крупномасштабных биркеландовских токов, находящейся на авроральных широтах и представляющей собой зоны втекающего в ионосферу ($\mathbf{j}_{0z}$) и вытекающего из ионосферы ($-\mathbf{j}_{0z}$) крупномасштабных продольных токов [17]. Измерения показывают, что в магнитоспокойных условиях при переходе от дневных к ночным часам втекающий ток находится на южной стороне авроральной зоны, а вытекающий — на северной. В ночные и утренние часы направления токов меняются на противоположные. На не освещённой солнцем стороне ионосферы, когда E-слой отсутствует либо выражен слабо, замыкание противоположно направленных продольных токов должно происходить за счёт поперечного тока $\mathbf{j}_{0x}$, протекающего преимущественно в верхней ионосфере и направленного с юга на север (см. рис. 1). В этом случае направление тока $\mathbf{j}_{0x}$ совпадает с направлением северного градиента концентрации плазмы $\nabla_x n$, что, как отмечалось, является необходимым условием для развития ГТ неустойчивости. Эксперимент также показывает [15], что в авроральной ионосфере развиваются в основном плоско-слоистые (sheet-like) неоднородности, обладающие максимальной степенью вытянутости в направлении скорости $\mathbf{v}_{0y}$ регулярного дрейфа плазмы. Последняя ориентирована преимущественно в направлении восток–запад и ортогональна к направлению регулярного градиента плазмы. Описанная выше картина хорошо укладывается в рамки ГТ механизма развития неоднородной структуры. Здесь также важно, что наблюдаемое в вечерние часы увеличение северного градиента главного ионосферного провала может быть обусловлено нелинейным перераспределением концентрации за счёт неоднородного по x и z крупномасштабного индукционного электрического поля $E_{0y} = -j_{0x} B_0/(enc)$, вызывающего дрейф фоновой плазмы на юг со скоростью $v_{0x} = -cE_{0y}/B_0$.

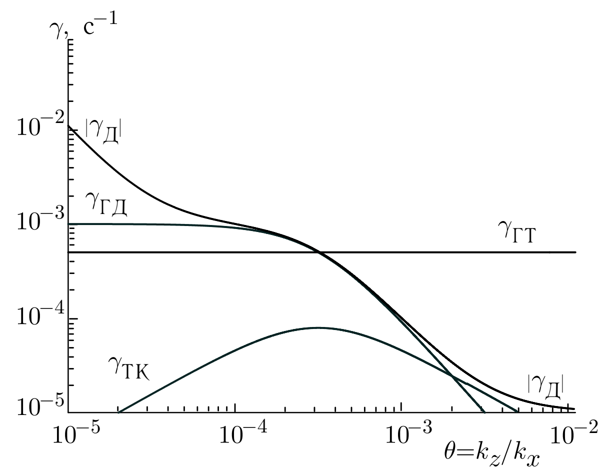


Рис. 3. Сопоставление инкрементов градиентно-токовой ($\gamma_{\text{ГТ}}$), градиентно-дрейфовой ($\gamma_{\text{ГД}}$), токово-конвективной ($\gamma_{\text{ТК}}$) неустойчивостей с декрементом униполярной диффузии ($|\gamma_{\text{Д}}|$)

В отличие от рассмотренной выше картина развития неоднородностей вследствие ГД неустойчивости должна выглядеть иначе. Если предположить, что причиной неустойчивости является дрейф плазмы со скоростью $v_{0y} = -cE_{0x}/B_0$, вызванный крупномасштабным электрическим полем, которое на авроральных широтах ориентировано преимущественно параллельно направлению север—юг, неустойчивым может стать только градиент $\nabla_y n$, который существенно меньше градиента $\nabla_x n$ вдоль направления север—юг. Другая особенность ГД и ТК неустойчивостей состоит в том, что максимальным инкрементом обладают возмущения, волновой вектор которых ортогонален к регулярному градиенту плазмы, откуда можно предположить, что их развитие приведёт к изотропизации крупномасштабного градиента в плоскости, ортогональной к $\mathbf{B}_0$.

Оценим инкремент ГТ неустойчивости (25) применительно к условиям авроральной верхней ионосферы и сопоставим его с инкрементами ГД и ТК неустойчивостей (11). Из уравнения непрерывности тока (16) следует $\Delta v_{0x}/L_x \approx -\Delta v_{0z}/L_z$, где $\Delta v_{0x} = j_{0x}/(en)$ и $\Delta v_{0z} = j_{0z}/(en)$ — разности скоростей заряженных частиц, отвечающих крупномасштабным поперечному и продольному токам. Определим Δv_{0z} из уравнения (14). Взяв величину градиента магнитного поля в области противоположно направленных продольных токов $\Delta B_{0y}/\Delta x \approx 1$ нТл/км (см. [2, 17]), получим $\Delta v_{0z} \approx 20\div 50$ м/с, что отвечает инкременту $\gamma_{\text{ГТ}} \approx (2\div 5)\cdot 10^{-4}$ с$^{-1}$. Для сильно ионизованной плазмы инкремент (34) при $B_{0y} \approx 5\cdot 10^{-3} B_0 \approx 500$ нТл и $l_z = 2\pi/k_z \approx 50$ км в диапазоне поперечных масштабов $l_x = 2\pi/k_x \sim 1\div 10$ км составляет $\gamma_{\text{ГТ}} \gtrsim 10^{-3}\div 10^{-2}$ с$^{-1}$. В случае ГД неустойчивости при $v_{0y} \approx 500$ м/с и $L_y \approx 500$ км максимальный инкремент составляет

$\gamma_{\text{ГД}} \approx v_{0y}/L_y \approx 10^{-3}$ с$^{-1}$. Максимальный инкремент ТК неустойчивости при использовавшихся выше значениях $\Delta v_{0z} \approx 20{\div}50$ м/с и $L_y \approx 100$ км составляет $\gamma_{\text{ТК}} = [m_{\text{e}}\nu_{\text{e}}/(m_{\text{i}}\nu_i)]^{1/2}\, v_{0z}/(2L_y) \approx$ $\approx (0{,}2{\div}1) \cdot 10^{-4}$ с$^{-1}$.

Зависимости инкрементов рассмотренных выше неустойчивостей от параметра $\theta = l_\perp/l_\parallel$ построены на рис. 3. На этом же рисунке приведён декремент униполярной диффузии [20]

$$|\gamma_{\text{Д}}| \approx \gamma_{\text{i}}\gamma_{\text{e}}/(|\gamma_{\text{i}}| + |\gamma_{\text{e}}|), \tag{39}$$

где $|\gamma_{\text{i}}| = D_{\text{i}\perp}k_x^2 + D_{\text{i}\parallel}k_z^2$, $|\gamma_{\text{e}}| = D_{\text{e}\perp}k_x^2 + D_{\text{e}\parallel}k_z^2$ — декременты ионной и электронной униполярной диффузии соответственно. Были взяты следующие характерные для верхней ионосферы значения коэффициентов униполярной ионной и электронной диффузии:

$$D_{\text{i}\perp} = (T_{\text{e}} + T_{\text{i}})\,\nu_{\text{in}}/(m_{\text{i}}\omega_{B\text{i}}^2) \approx 10^5 \text{ см}^2/\text{с}, \qquad D_{\text{i}\parallel} = (T_{\text{e}} + T_{\text{i}})/(m_{\text{i}}\nu_{\text{in}}) \approx 10^{10} \text{ см}^2/\text{с},$$

$$D_{\text{e}\perp} = (T_{\text{e}} + T_{\text{i}})\,\nu_{\text{e}}/(m_{\text{e}}\omega_{B\text{e}}^2) \approx 10^3 \text{ см}^2/\text{с}, \qquad D_{\text{e}\parallel} = (T_{\text{e}} + T_{\text{i}})/(m_{\text{e}}\nu_{\text{e}}) \approx 10^{12} \text{ см}^2/\text{с}.$$

Продольный масштаб неоднородностей считался равным $l_z \approx 2\pi/k_z \approx 300$ км. Из приведённых графиков видно, что даже при столь большом значении l_z существует довольно узкий интервал значений $10^{-4} \lesssim \theta \lesssim 10^{-3}$, где $\gamma_{\text{ГД}}$ совпадает с $|\gamma_{\text{Д}}|$. Инкремент $\gamma_{\text{ТК}}$ всегда остаётся меньше $|\gamma_{\text{Д}}|$, что не позволяет рассматривать ТК неустойчивость в качестве действующего механизма образования неоднородной структуры верхней высокоширотной ионосферы. В то же время $\gamma_{\text{ГТ}}$ значительно превосходит $|\gamma_{\text{Д}}|$ при $\theta \gtrsim 10^{-3}$, при этом ГТ неустойчивость может развиваться, когда имеет место ионная униполярная диффузия. Для сильно ионизованной плазмы определяющей становится электронная униполярная диффузия (см. [20]). Сопоставив инкремент (34) с декрементом электронной продольной диффузии $|\gamma_{\text{e}\parallel}| = k_z^2 D_{\text{e}\parallel}$, получим $k_z/k_x \lesssim \text{Re}_{\text{e}}^{-1} =$ $= B_{0z}\nu_{\text{e}}/(B_{0y}\omega_{B\text{e}})$. Таким образом, видно, что при переходе к сильно ионизованной плазме на высотах, превышающих положение максимума F-слоя, степень вытянутости неоднородностей в направлении геомагнитного поля определяется магнитным числом Рейнольдса. Отметим, что образование в результате развития ГТ неустойчивости замкнутых квазистатических токов и нарушение симметрии возмущения могут существенно ослабить его расплывание вдоль $\mathbf{B}_0$, как это, например, имеет место при диффузии мелкомасштабных искусственных неоднородностей [12] или при релаксации θ-пинча в перекошенном магнитном поле [16].

4. ЗАКЛЮЧЕНИЕ

В приближении двухжидкостной магнитной гидродинамики рассмотрен механизм генерации квазистатических электромагнитных полей в плазме с регулярным током, вектор плотности которого лежит в плоскости, проходящей через направление регулярного магнитного поля. В случаях слабо и сильно ионизованной плазмы получены выражения для инкрементов ГТ неустойчивости, показано, что в условиях высокоширотной верхней ионосферы она может приводить к генерации наблюдаемых в эксперименте плоско-слоистых неоднородностей электронной концентрации, вытянутых вдоль направлений геомагнитного поля и скорости регулярного дрейфа плазмы. В сильно ионизованной плазме ГТ неустойчивость может развиваться вследствие перезамыкания системы противоположно направленных крупномасштабных продольных токов и приводить к образованию локальных областей, содержащих интенсивные мелкомасштабные неоднородности плотности плазмы. Предложенный механизм генерации флуктуаций плотности плазмы и мелкомасштабных квазистатических электрических полей может быть использован при разработке теории низкочастотной турбулентности плазмы верхней ионосферы и магнитосферы.

Работа выполнена при поддержке РФФИ (грант № 03–05–64636) и INTAS (грант № 03–51–5583).

СПИСОК ЛИТЕРАТУРЫ

1. Dungey J. W. // J. Atm. Terr. Phys. 1956 V. 9. P. 304.
2. Kelley M. C. The Earth's Ionosphere: Plasma Physics and Electrodynamics. Int. Geophys. Ser. V. 43. New York: Academic Press, 1989.
3. Kadomtsev B. B., Nedospasov A. V. // J. Nucl. Energy. C. 1960. V. 1. P. 230.
4. Ossakow S. L., Chaturvedi P. K. // Geophys. Res. Lett. 1979. V. 6. P. 332.
5. Erukhimov L. M., Kagan L. M. // J. Atm. Terr. Phys. 1994. V. 56. P. 133.
6. Livingston R. C., Rino C. L., Owen J., Tsunoda R. T. // J. Gephys. Res. 1982. V. 87. P. 10 519.
7. Erukhimov L. M., Kosolapenko V. I., Myasnikov E. N., Lerner A. M. // Planet. Space Sci. 1981. V. 29. P. 931.
8. Huba J. D., Ossakow S. L., Satyanarayana P., Guzdar P. N. // J. Geophys. Res. 1983. V. 88. P. 425.
9. Keskinen M. J., Mitchell H. G., Fedder J. A., et al. // J. Geophys. Res. 1988. V. 93. P. 137.
10. Nagatsuma T., Fukunishi H., Hayakawa H., Mikai T., Matsuoka A. // J. Geophys. Res. 1996. V. 101. P. 21 715.
11. Kintner P. M., Seyler C. E. // Space Sci. Rev. 1985. V. 41. P. 91.
12. Ерухимов Л. М., Мясников Е. Н. // Изв. вузов. Радиофизика. 1998. Т. 41. С. 194.
13. Мясников Е. Н. // Изв. вузов. Радиофизика. 1999. Т. 42. С. 691.
14. Моффат Г. Возбуждение магнитного поля в проводящей среде. М.: Мир, 1980.
15. Tereshchenko E. D., Khudukon B. Z., Kozlova M. O., et al. // Ann. Geophysicae. 2000. V. 18. P. 918.
16. Кадомцев Б. Б. Коллективные явления в плазме. М.: Наука, 1988.
17. Kamide Y. // Space Sci. Rev. 1982. V. 31. P. 127.
18. Erukhimov L. M., Myasnikov E. N., Kosolapenko V. I., et al. // Radio Sci. 1994. V. 29. P. 311.
19. Kagan L. M., Myasnikov E. N., Kosolapenko V. I., et al. // J. Atm. Terr. Phys. 1995. V. 57. P. 917.
20. Гуревич А. В., Цедилина Е. Е. // УФН. 1967. Т. 91. С. 609.

Поступила в редакцию 30 июля 2004 г.; принята в печать 6 апреля 2005 г.

ON THE GRADIENT-CURRENT MECHANISM OF FORMATION OF THE INHOMOGENEOUS-STRUCTURE OF THE HIGH-LATITUDE UPPER IONOSPHERE

E. N. Myasnikov

We consider the gradient-current instability of an inhomogeneous magnetoactive plasma in the approximation of two-liquid magnetic hydrodynamics. Unlike the known gradient-drift and current-convective instabilities, the gradient-current instability is related to generation of nonpotential quasistatic electric fields polarized orthogonal to the external magnetic field $\mathbf{B}_0$ and excited by eddy currents whose density vector lies in the plane passing through the vectors of the magnetic field $\mathbf{B}_0$ and large-scale electron-density gradient. It is shown that in the high-latitude upper ionosphere, in the regions containing large-scale currents flowing in and out of the ionosphere along the magnetic field, the gradient-current instability can lead to the appearance of sheet-like inhomogeneities extended predominantly in the plane passing through the vectors of the geomagnetic field and regular plasma-drift velocity.

УДК 537.874.37

РАСПРОСТРАНЕНИЕ УЛЬТРАКОРОТКИХ ВОЛН НА МОРСКИХ ТРАССАХ В ЮЖНЫХ ПОЛЯРНЫХ ШИРОТАХ

В. К. Иванов, В. Н. Лановой, В. Н. Шаляпин, Л. А. Егорова, А. С. Васильев, А. А. Могила

Институт радиофизики и электроники им. А. Я. Усикова НАНУ, г. Харьков, Украина

Приведены результаты метеорологических и радиометеорологических наблюдений, а также измерений ослабления ультракоротких радиоволн во время 28-й Советской Антарктической экспедиции в январе–марте 1983 года. Показано, что радиометеорологические параметры над акваторией Южного океана в летний период практически не отличаются от соответствующих усреднённых параметров над сушей в зимний период. Основным механизмом распространения радиосигналов на расстояния 100 и более километров является однократное рассеяние на турбулентных флуктуациях коэффициента преломления атмосферы. Отмечены низкие по сравнению с измерениями в других климатических районах абсолютные значения погонного ослабления на высоких частотах.

ВВЕДЕНИЕ

Распространение ультракоротких радиоволн определяется, в первую очередь, распределением коэффициента преломления атмосферы. В соответствии с рекомендациями Международного консультативного комитета по радиосвязи [1] широкое применение получила так называемая усреднённая стандартная модель атмосферы, в которой коэффициент преломления монотонно спадает с высотой по линейному закону.

Стандартная модель позволяет аналитически рассчитать зависимость ослабления радиоволн от дистанции, и на небольших расстояниях её предсказания неплохо соответствуют реальным измерениям. Однако в результате экспериментов, проведённых над сушей, было показано, что уровни сигналов на расстояниях порядка 100 км и более значительно превышают расчётные значения, полученные по теории дифракции [2–6]. Такое явление дальнего тропосферного распространения обусловлено, прежде всего, рассеянием излучения на флуктуациях коэффициента преломления. В то же время теоретические и экспериментальные исследования распространения ультракоротких волн над различными акваториями Мирового океана показали, что в среднем на приводных трассах наблюдается более высокий над сушей уровень сигнала по сравнению с измерениями [7–9]. Связано это, прежде всего, с наличием волновода испарения и более высокой вероятностью образования приподнятых отражающих слоёв. Отметим также, что уровень сигналов существенно зависит от климатического района и сезона измерения [10], из чего следует необходимость проведения исследований в различных районах Земного шара.

В последние десятилетия отмечен активный интерес к исследованиям Антарктиды и охватывающих её морей трёх океанов: Атлантичнского, Индийского и Тихого, объединяемых сейчас единым понятием Южный океан. Однако, судя по количеству публикаций, исследования распространения радиоволн в этом районе очень немногочисленны [11]. Среди них отметим экспериментальное изучение условий распространения ультракоротких волн над морем с помощью метеорологического радиолокатора МРЛ-1М и встроенного в него СВЧ радиометра на длине волны 3,2 см [12]. Радиолокатор был установлен в точке с координатами $67°40'$ ю. ш. и $46°08'$ в. д. на берегу залива Алашеева на высоте 232,3 м над уровнем моря. Однако эти измерения проводились в прибрежной зоне, а физика образования приводных, приповерхностных и приподнятых инверсионных слоёв возле берега, как известно, в корне отличается от механизмов, характерных для акватории открытого океана. В прибрежной зоне инверсионные слои образуются в основном за счёт адвекции воздушных масс с континента.

В данной статье нами предпринята попытка частично восполнить пробел в изучении распространения ультракоротких волн в южных полярных широтах на открытых морских трассах и представить результаты метеорологических, радиометеорологических и радиофизических исследований 28-й Советской Антарктической экспедиции, относящиеся к изучению дальнего тропосферного распространения ультракоротких радиоволн.

1. МЕТЕОРОЛОГИЧЕСКИЕ И РАДИОМЕТЕОРОЛОГИЧЕСКИЕ ИЗМЕРЕНИЯ

Экспедиция проходила в период южного лета с января по март 1983 года. В измерениях участвовали два судна, находившихся в совместном плавании. Их маршруты приведены на рис. 1.

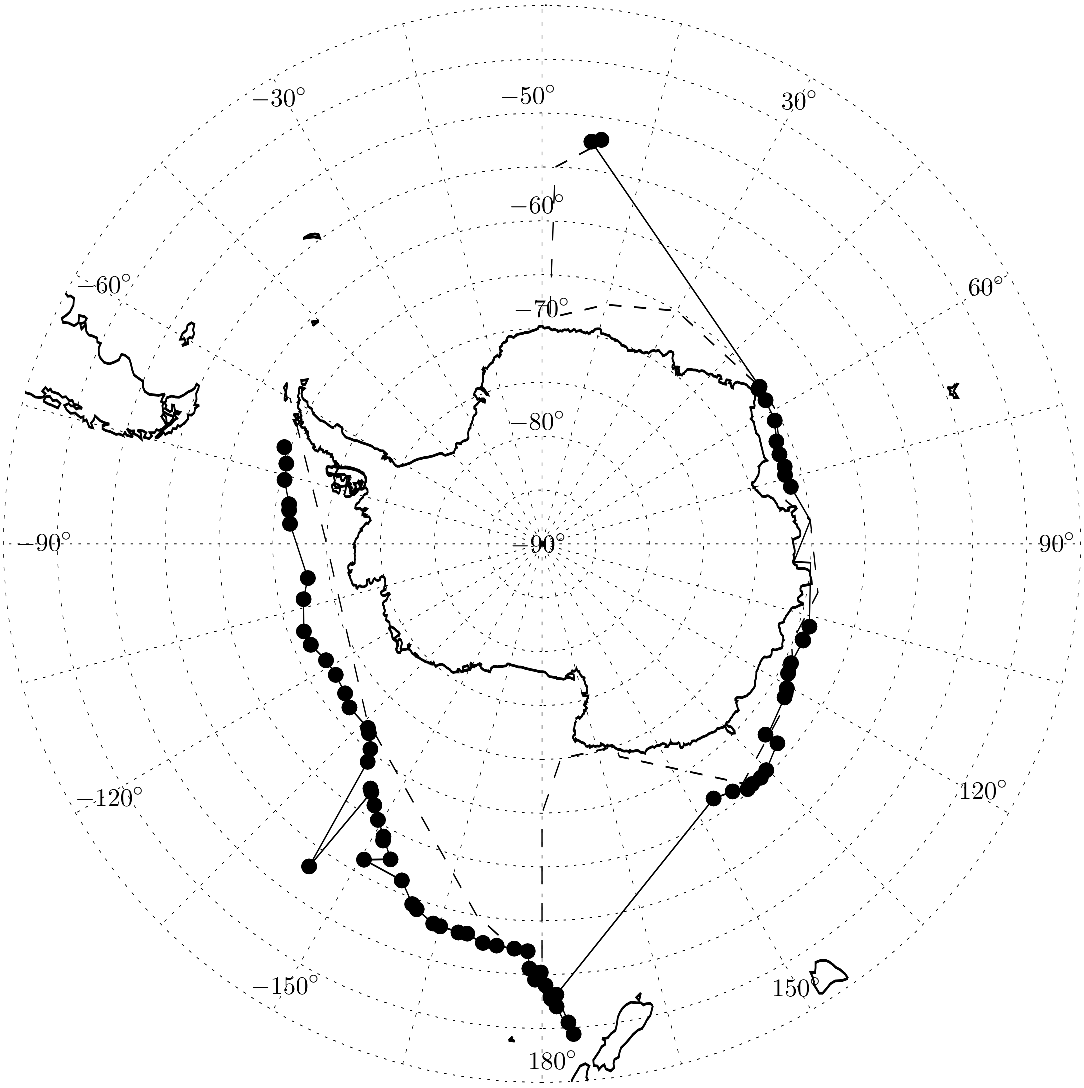
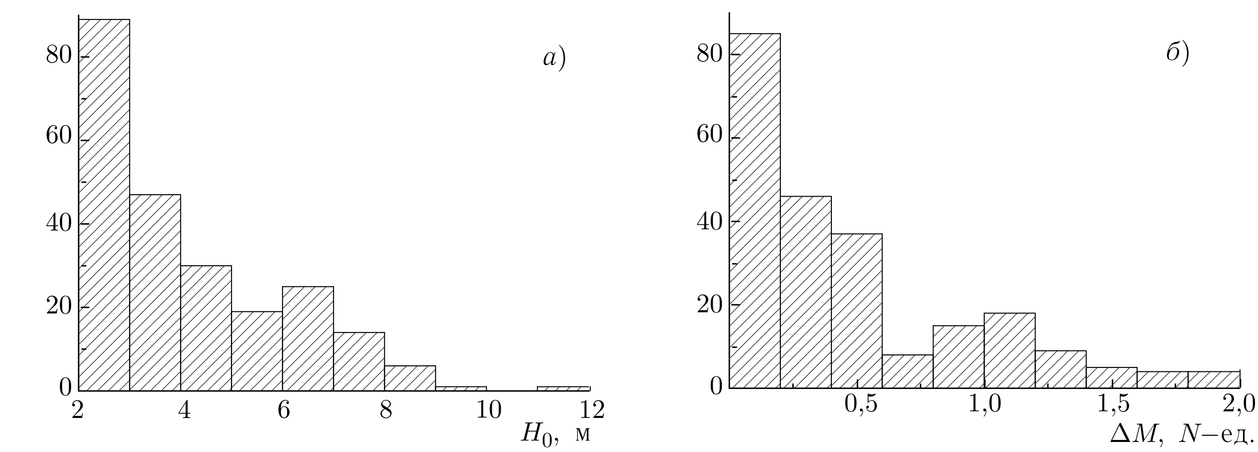


Рис. 1. Схема маршрута 28-й Советской Антарктической экспедиции в 1983 году. Точками обозначено положение корабля «Адмирал Владимирский» во время сеансов радиосвязи. Маршруты судов обозначены сплошной и штриховой линиями



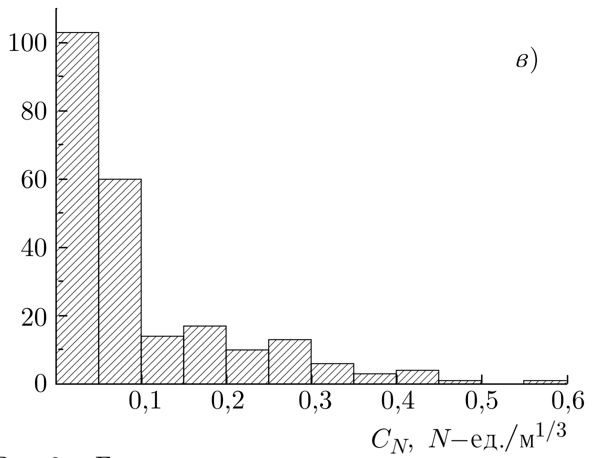


Рис. 2. Гистограммы распределений параметров волновода испарения: высоты H_0 (*а*), M-дефицита ΔM и структурной постоянной C_N

На обоих судах 8 раз в сутки проводились стандартные гидрометеорологические наблюдения (измерения температуры, давления, влажности и скорости ветра на уровне палубы и измерения температуры воды) и 4 раза в сутки осуществлялось аэрологическое зондирование атмосферы до высоты 5 000 м. Фиксировались особые метеорологические условия, степень морского волнения в баллах, наличие и тип облаков, тумана, осадков в виде дождя и снега.

Судя по данным, полученным на меридианном разрезе по 10° в. д., северная граница полярной фронтальной зоны располагалась между 39° и 40° ю. ш., причём субарктический фронт был выражен, как обычно, очень резко: перепад температур на трёх градусах широты превышал 7 °C. Южнее зоны антарктической конвергенции, расположенной примерно на 53° ю. ш., температура воды не превышала 2 °C. Таким образом, между параллелями 39° и 53° ю. ш. температура воды снизилась с 16 °C до 2 °C, что связано с выходом к поверхности главного термоклина [13]. Давление воздуха южнее 50° ю. ш. в течение января–февраля изменялось в пределах 900÷1 000 ГПа.

Сильное понижение температуры воды и воздуха, частые туманы, интенсивное ветровое перемешивание обуславливались характером атмосферных процессов вблизи побережья Антарктиды, окружённой линией атмосферных фронтов.

По данным стандартных гидрометеорологических измерений были рассчитаны распределения параметров приводного слоя атмосферы [14]: высоты H_0, разности значений коэффициента преломления на верхней и нижней границе приводного волновода (M-дефицита) ΔM, структурной постоянной C_N в колмогоровском спектре флуктуаций, а также коэффициент преломления $N_{\rm s}$ у поверхности раздела, необходимые как для учёта высвечивания энергии из волновода испарения за счёт рассеяния на анизотропных флуктуациях, так и для расчёта уровня сигналов при дальнем тропосферном распространении ультракоротких радиоволн [15]. Из полученных гистограмм (рис. 2) следует, что высота волновода в основном не превышает 6 м, а M-дефицит, измеряемый в N-единицах, меняется от десятых долей до 2 N-единиц. Структурная постоянная имеет низкие

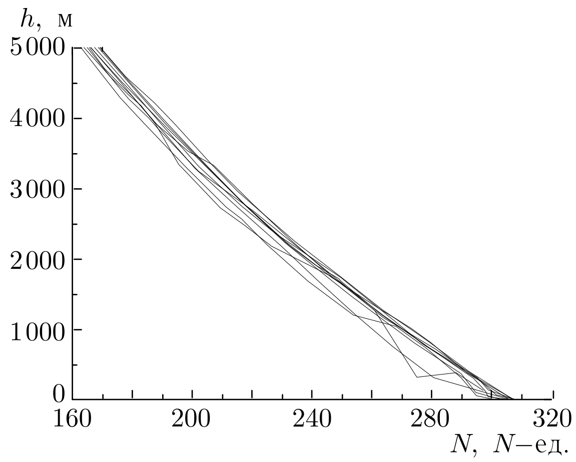


Рис. 3. Вертикальные профили коэффициента преломления $N(h)$

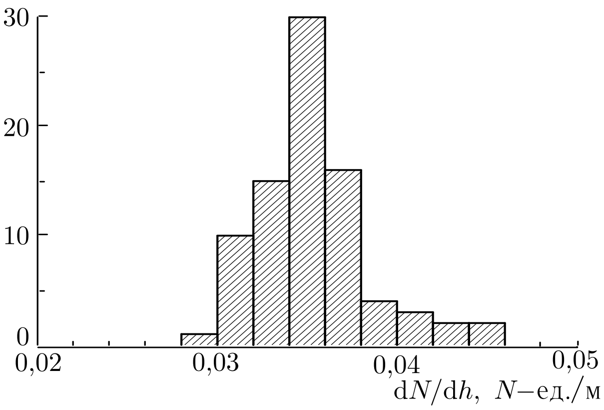


Рис. 4. Гистограмма распределения усреднённого градиента показателя преломления на высотах до 1 000 м

значения. Это связано с низкой температурой поверхности океана (от 0 до 2°C) и высокой из-за частых штормов, туманов и осадков влажностью. Лишь в редких случаях волноводы достигают высоты 7÷9 м, обусловленной адвективными процессами с материка.

Интенсивная циклоническая деятельность и связанные с ней восходящие потоки над акваторией океана не благоприятствовали образованию инверсионных слоёв. Нижняя граница «инверсионной кромки» морского пограничного слоя располагалась на высотах от 600 до 1 400 м. В среднем толщина пограничного слоя, т. е. слоя перемешивания, составляла около 1 100 м. Слои на этих широтах встречались очень редко. Типичное распределение турбулентных потоков эпизодически нарушалось адвекцией воздушных масс, в основном обусловленных выносом холодных масс с шельфовых ледников материка.

По результатам аэрологического зондирования определялись также градиенты коэффициента преломления на высотах до 5 000 м, а также параметры приподнятых инверсионных слоёв. На рис. 3 приведены типичные примеры вертикального профиля коэффициента преломления $N(h)$ для Южного океана в летний период, рассчитанные по данным аэрологического зондирования. Более полный анализ аэрологических данных указывает на отсутствие приподнятых инверсионных слоёв в данном районе. Градиент коэффициента преломления в нижнем километровом слое имеет значение $-0{,}035\pm0{,}003$ N-ед./м очень близкое к величине $-0{,}04$ N-ед./м, характерной для стандартной атмосферы над сушей (рис. 4). Таким образом, радиометеорологические параметры над акваторией Южного океана в летний период практически не отличаются от соответствующих параметров усреднённой стандартной атмосферы над сушей в зимний период.

2. ИЗМЕРЕНИЯ ОСЛАБЛЕНИЯ РАДИОВОЛН

Измерения ослабления радиосигналов проводились для различных дистанций R между судами, движущимися в попутном направлении. Максимальное расстояние достигало 513 км. На одном из судов размещался передающий измерительный комплекс, а на втором – приёмный. Характеристики и высоты расположения антенн приведены в табл. 1.

Взаимное наведение антенн между судами осуществлялось по данным спутниковых навигационных систем. Тщательные измерения поля в интерференционных максимумах, где множитель ослабления V при коэффициенте отражения от подстилающей поверхности, близком к единице,

Таблица 1. Характеристики и высоты расположения приёмных и передающих антенн

Частота, МГц	Высота приёмника, м	Высота передатчика, м	Усиление, дБ	Ширина диаграммы, град.	Поляризация
150	22	12	15	27	гориз.
600	17	10	14	16	гориз.
3 000	12	17	26	5	верт.
10 000	12	17	22	17	верт.

имеет нормированное значение, равное 6 дБ, позволили определить множитель ослабления для других дистанций, не прибегая к абсолютным измерениям параметров передающего и приёмного комплексов, включая антенны [9]. Из экспериментально определённой зависимости относительных значений принятой мощности от дистанции следует, что вариации уровня принимаемой мощности на дистанции первого максимума интерфереционной картины $R_{1\,\max}$ за ряд галсов не превышают $\pm(3\div4)$ дБ, а на дистанции $2R_{1\,\max}$ составляют $\pm(2\div3)$ дБ. Таким образом, проводившаяся на близких расстояниях привязка на мощности позволяла определять уровень поля на малых дистанциях с погрешностью, не превышающей ±3 дБ. При этом суммарная ошибка в определении экспериментальных значений множителя ослабления на других дистанциях не превышала $\pm(2\div3)$ дБ и определялась в основном нестабильностью параметров измерительного комплекса за время, прошедшее между двумя привязками.

Натурные измерения уровня сигналов на трассе проводились отдельными сеансами продолжительностью по $15\div30$ мин. Дистанция R между приёмным и передающим судном от сеанса к сеансу изменялась случайным образом. По результатам каждого сеанса рассчитывались медианные значения множителя ослабления $V_{\rm m}$ для данных расстояний между судами.

Измерения множителя ослабления проводились в основном на дистанциях в диапазоне $100\div300$ км. Лишь в некоторых случаях дистанция понижалась до 50 км или возрастала до 500 км. На рис. 5 показаны измеренные медианные значения ослабления на частотах 150 МГц, 600 МГц, 3 ГГц и 10 ГГц вместе с аппроксимацией дистанционной зависимости прямой линией (сплошные линии) с помощью соотношения

$$V_{\rm m} = V_{\rm m}(100) + b\,(R - 100), \tag{1}$$

где $V_{\rm m}(100)$ — ослабление на дистанции 100 км, b — погонное ослабление в дБ/км. Коэффициенты $V_{\rm m}(100)$ и b приведены в табл. 2.

Из анализа следует, что уровень сигналов на расстоянии 100 км на всех длинах волн довольно низкий по сравнению с морскими трассами в других районах Земного шара. Это объясняется отсутствием приводных волноводов и практически стандартным градиентом коэффициента преломления в нижнем слое атмосферы. Отсюда следует, что основным механизмом распространения ультракоротких волн на антарктических морских трассах на больших расстояниях, подобно распространению над сушей, является однократное рассеяние на турбулентных флуктуациях показателя преломления атмосферы. Для сравнения на рис. 5 штриховыми линиями приведены экспериментальные аппроксимации зависимости уровня сигналов от дистанции над сушей для зимнего периода при различных длинах волн [16]. Погонное ослабление было выбрано равным $b = -0{,}05$ дБ/км, а ослабление на дистанции 100 км получено с помощью эмпирической зависимости от длины волны λ:

$$V_{\rm m} = -74 + 10\lg\lambda\ [\text{см}]. \tag{2}$$

Погонное ослабление для $\lambda = 200$ см и $\lambda = 50$ см практически совпадает со значением $b = -0{,}05$ дБ/км, полученным над сушей. Для $\lambda = 10$ см и $\lambda = 3$ см погонное ослабление зна-

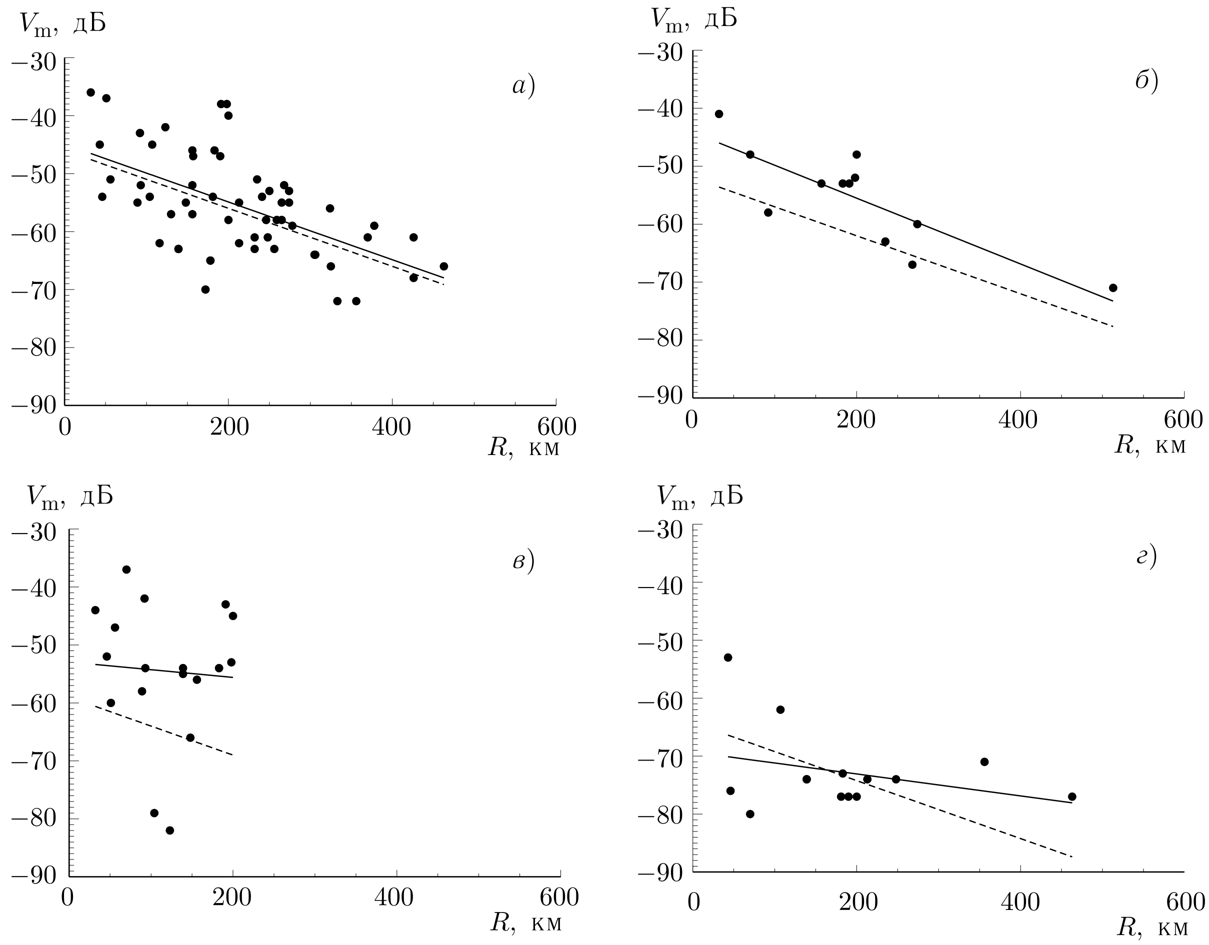


Рис. 5. Измеренные медианные значения множителя ослабления на частотах 150 МГц (*а*), 600 МГц (*б*), 3 ГГц (*в*) и 10 ГГц (*г*). Сплошными линиями показаны результаты аппроксимации прямыми линиями, штриховыми линиями — экспериментальные зависимости для суши в зимний период

Таблица 2. Параметры аппроксимации дистанционной зависимости множителя ослабления прямой линией и средние значения показателя преломления воздуха на уровне моря

Частота, МГц	$V_{\rm m}(100)$, дБ	b, дБ/км	$\overline{N_{\rm s}}$, N-ед.
150	$-49{,}9 \pm 2{,}3$	$-0{,}050 \pm 0{,}010$	$310{,}1 \pm 3{,}7$
600	$-49{,}8 \pm 2{,}9$	$-0{,}057 \pm 0{,}013$	$310{,}2 \pm 7{,}3$
3 000	$-54{,}3 \pm 6{,}9$	$-0{,}013 \pm -0{,}054$	$308{,}2 \pm 2{,}1$
10 000	$-71{,}2 \pm 3{,}6$	$-0{,}019 \pm 0{,}017$	$311{,}0 \pm 5{,}6$

чительно слабее (см. табл. 2). В то же время для большинства измерений в других климатических районах наблюдается прямо противоположная тенденция: абсолютное значение погонного ослабления возрастает с увеличением частоты. На наш взгляд, более медленное спадание уровня сигнала на высоких частотах может быть связано с меньшей высотой пограничного слоя. Как известно, на границе пограничного слоя, как правило, образуются инверсионные слои и значительно увеличивается интенсивность флуктуаций коэффициента преломления, что способствует повышению уровня сигнала на более высоких частотах.

ВЫВОДЫ

Радиометеорологические условия в летний период над акваторией Южного океана соответствуют усреднённым стандартным радиометеоусловиям над сушей в средних и высоких широтах в зимний период и характеризуются монотонным спаданием коэффициента преломления атмосферы с высотой. Температура поверхности океана в этот сезон колеблется от 0 до 2° C. Из-за низкой температуры высота и мощность волновода испарения очень малы, и он практически не влияет на распространение радиоволн на расстояния 100 и более километров. Частые штормы и ветры, а также осадки в виде дождя и снега способствуют перемешиванию атмосферы и препятствуют образованию как приповерхностных, так и приподнятых инверсий. Исключением являются редкие случаи адвекции холодных сухих масс с материка. Граница атмосферного пограничного слоя в этой части Мирового океана ниже, чем в остальных морских и сухопутных районах, что приводит к некоторому возрастанию сигналов на высоких частотах в зоне дальнего тропосферного распространения.

Радиометеорологические исследования позволяют сделать вывод, что основным механизмом распространения ультракоротких волн на антарктических морских трассах на больших расстояниях является рассеяние на турбулентных флуктуациях показателя преломления атмосферы. В зоне прямой видимости и дифракции также следует ожидать характеристики сигналов, аналогичные полученным над сушей при отсутствии атмосферных волноводов.

Авторы выражают благодарность рецензенту за полезные замечания.

СПИСОК ЛИТЕРАТУРЫ

1. Международный консультативный комитет по радиосвязи. / Сб. материалов VIII Пленарной Ассамблеи по вопросам радиовещания, радиосвязи и телевидения. М.: ЦБНТИ, 1957. 27 с.
2. Арманд Н. А., Введенский Б. А., Калинин А. И. и др. // Радиотехника и электроника. 1961. Т. 6, № 6. С. 867.
3. Gordon W. E. // Proc. IRE. 1955. V. 43. P. 23.
4. Троицкий В. Н. // Радиотехника. 1956. Т. 11, № 5. С. 3.
5. Шифрин Я. С., Чёрный Ф. Б., Тихомиров Ю. А. и др. // Труды радиотехнической академии им. Говорова. Харьков, 1964. С. 101.
6. Шур А. А. Характеристики сигналов на тропосферных радиолиниях. М.: Связь, 1972. 103 с.
7. Дудко Б. П., Корнеев И. Л., Тисленко В. И., Шарыгин Г. С. // Межведомственное совещание по распространению ультракоротких радиоволн и электромагнитной совместимости. Улан-Уде, 1983. С. 21.
8. Hitney H. V., Richter J. H., Pappert R. A., et al. // Proc. IEEE. 1985. V. 73. P. 265.
9. Braude S. Ya., Ivanov V. K., Ostrovsky I. E., Fuks I. M. // Telecommunication and Radioengineering. 1997. V. 51. P. 1.
10. Иванов В. К., Кивва Ф. В., Островский И. Е. // XVII конф. по распространению радиоволн. Ульяновск, 1993. С. 2.
11. Balaklitski I. M., Kivva F. V., Lanovoy V. N. // 6th Intern. School on Microwave Physics and Technique. Bulgaria: Varna, 1989. P. 184.
12. Михайлов Н. Ф., Рыжков А. В., Щукин Г. Г. Радиометеорологические исследования над морем. Л.: Гидрометеоиздат, 1990. 207 с.
13. Лагун В. Е., Васильев В. Ф., Романов В. Ф., Арискина Н. В. // Сб. докладов на II Всесоюзном симпозиуме «Метеорологические исследования в Антарктике» Ленинград, 19–22 октября 1981 г. Л.: Гидрометеоиздат, 1986. С. 217.

14. Гаврилов А. С., Петров Ю. С. // Рассеяние и дифракция радиолокационных сигналов и их информативность. Л.: СЗПИ, 1984. С. 31.
15. Белоброва М. В., Иванов В. К., Кукушкин А. В. и др. // Изв. вузов. Радиофизика. 1990. Т. 33, № 12. С. 1315.
16. Дальнее тропосферное распространение ультракоротких радиоволн. / Под ред. Б. А. Введенского, М. А. Колосова, А. И. Калинина, Я. С. Шифрина. М.: Сов. радио, 1965. 415 с.

Поступила в редакцию 5 июля 2004 г.; принята в печать 18 марта 2005 г.

PROPAGATION OF VHF RADIO WAVES ON SEA ROUTES IN THE SOUTH POLAR LATITUDES

V. K. Ivanov, V. N. Lanovoy, V. N. Shalyapin, L. A. Egorova, A. S. Vasilyev, and A. A. Mogila

We present the results of meteorological and radio-meteorological observations as well as the results of measurements of the attenuation factor of very-high-frequency (VHF) radio waves during the 28th Soviet Antarctic expedition in January–March 1983. It is shown that radio-meteorological parameters over the South Ocean area in summer almost coincide with the corresponding averaged parameters over the land in winter. The main mechanism of radio-wave propagation at distances over 100 km is single scattering by turbulent fluctuations of the refractive index of the atmosphere. Absolute values of the running attenuation at high frequencies are low compared to those in other climatic areas.

УДК 621.372+621.378.1

ОТКРЫТЫЕ РЕЗОНАТОРЫ С ПРОВОДЯЩИМИ ЦИЛИНДРИЧЕСКИМИ ВСТАВКАМИ. 1. ДВУМЕРНАЯ МОДЕЛЬ

П. Н. Мележик, В. С. Мирошниченко, Е. Б. Сенкевич

Институт радиофизики и электроники им А. Я. Усикова НАНУ, г. Харьков, Украина

Представлены результаты вычислительного эксперимента по исследованию свойств колебаний в двумерном открытом резонаторе с проводящей цилиндрической вставкой. Обнаружен и исследован эффект повышения на 2÷3 порядка дифракционной добротности основного типа колебаний за счёт проводящей цилиндрической вставки. Обсуждаются перспективы использования данного эффекта для управления спектром, частотой и добротностью возбуждаемых колебаний. Продемонстрирована возможность осуществления перестройки открытого резонатора с проводящей вставкой на одном (основном) типе колебаний в широком диапазоне частот.

ВВЕДЕНИЕ

Открытые резонаторы (ОР) благодаря разреженности спектра их собственных колебаний являются практически безальтернативными резонансными системами в миллиметровом и субмиллиметровом диапазонах длин волн [1, 2]. При размещении в объёме ОР активных элементов (лавинно-пролётные диоды, диоды Ганна и т. п.), подводе питающих напряжений и обеспечении отвода тепла возникает необходимость ввода в резонатор проводящих неоднородностей, не ухудшающих спектральный состав и не приводящих к существенному снижению добротности возбуждаемых колебаний [3]. Так, для размещения активных элементов в объёме полусферического ОР ранее использовались как полуволновые решётки из прямоугольных проводящих брусьев, так и одиночные металлические брусья [4]. При этом было обнаружено явление аномально слабого рассеяния поля ОР на одиночном проводящем бруске прямоугольного сечения, размещённом в объёме резонатора [5].

Для анализа явлений, происходящих в ОР с металлическими и диэлектрическими включениями, были развиты методы строгой теории дифракции волн на идеально проводящих незамкнутых экранах и построена спектральная теория электромагнитных колебаний в ОР с металлическими и металлодиэлектрическими включениями [6, 7]. Было отмечено, что для некоторых типов колебаний в ОР с цилиндрической диэлектрической вставкой наблюдался резонансный характер поведения добротности в зависимости от диаметра диэлектрической вставки и её расположения между зеркалами резонатора [1, 6, 7].

В данной работе проведён целенаправленный поиск способов управления спектром и добротностью колебаний, возбуждаемых в ОР с проводящей цилиндрической вставкой. Обнаружен и исследован эффект повышения добротности основного типа колебаний в ОР за счёт проводящей цилиндрической вставки, обсуждаются перспективы использования данного эффекта для разрежения спектра колебаний ОР и размещения в объёме резонатора активных элементов.

1. ПОСТАНОВКА ЗАДАЧИ И ОБЪЕКТ ИССЛЕДОВАНИЯ

Как известно, разрежение спектра собственных колебаний в ОР с цилиндрическими зеркалами возможно только за счёт увеличения дифракционных потерь для высших поперечных типов колебаний [2]. Этот способ селекции колебаний в ОР реализуется при уменьшении апертуры зеркал резонатора либо при увеличении межзеркального расстояния. При этом неизбежно происходит снижение дифракционной добротности основного типа колебаний TEM_{00q}. Задача данного исследования состоит в том, чтобы за счёт размещения в ОР проводящего кругового цилиндра достичь эффективной селекции высших поперечных типов колебаний при сохранении или повышении дифракционной добротности для основного типа колебаний.

В качестве объекта исследования выбран двумерный ОР, образованный одинаковыми цилиндрическими зеркалами с радиусом кривизны R_c, размещённых на расстоянии $2l$ вдоль оси z (см. рис. 1). Между зеркалами резонатора помещён проводящий цилиндр с диаметром $2a$, причём образующая цилиндра параллельна оси y. Апертуру зеркал будем характеризовать углом раскрыва $2\varphi_0$ для соответствующего сегмента поверхности кругового цилиндра с диаметром $2R_c$. Будем рассматривать случай H-поляризации возбуждаемых колебаний (вектор магнитного поля $\mathbf{H}$ параллелен оси y), а для численного анализа поставленной задачи применим алгоритм расчёта спектральных характеристик ОР с металлодиэлектрическим цилиндром, приведённый в [6]. Следует отметить, что выбранная электродинамическая модель и алгоритм её решения допускают изменение в широких пределах как геометрических параметров ОР, так и длины волны возбуждаемых колебаний.

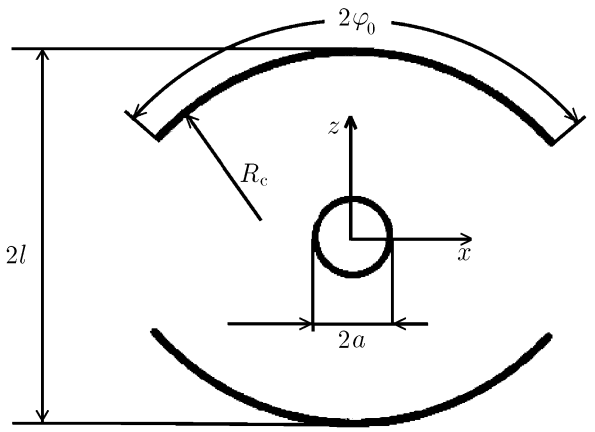


Рис. 1. Открытый резонатор с проводящей цилиндрической вставкой

Типы колебаний, возбуждаемые в рассматриваемом ОР, обозначим как квази-TEM_{mnq}, где индексы m, n и q описывают вариации поля вдоль осей x, y и z соответственно, как и в пустом резонаторе (для случая двумерного ОР $n \equiv 0$). В качестве ориентира для резонансных частот колебаний в ОР с проводящей вставкой возьмём спектр пустого ОР. В квазиоптическом приближении спектр TEM_{mnq}-колебаний в двумерном ОР без проводящей вставки описывается дисперсионным уравнением [2]

$$2kl = \pi q + \frac{1}{2}\,(1 + 2m)\arccos(1 - 2l/R_c) + 2\pi p_x, \qquad (1)$$

где $k = 2\pi/\lambda$ — волновое число в свободном пространстве, p_x — поправка за счёт дифракции на краях зеркал вдоль оси x, q — продольный индекс колебаний, равный числу полуволн вдоль оси z, m — число вариаций поля вдоль оси x.

Для практических приложений представляют интерес такие значения геометрических параметров исследуемого ОР, которые позволили бы создать одномодовый перестраиваемый резонатор с активными элементами, размещёнными в его объёме. Поэтому для численного анализа свойств двумерного ОР с проводящей цилиндрической вставкой были выбраны такие параметры $2l$, R_c, $2\varphi_0$, при которых происходит эффективное высвечивание высших поперечных типов колебаний пустого резонатора (TEM_{10q}, TEM_{20q} и т. д.), а уровень дифракционной добротности основного типа колебаний TEM_{00q} остаётся достаточно высоким ($Q = 10^3 \div 10^4$).

2. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ ВЫЧИСЛИТЕЛЬНЫХ ЭКСПЕРИМЕНТОВ

2.1. Поиск резонансных значений диаметра проводящей вставки

Поиск резонансного увеличения дифракционной добротности TEM_{00q}-колебаний проводился как при фиксированном расстоянии между зеркалами ($2l = \mathrm{const}$) и изменении диаметра проводящего цилиндра $2a$, так и при перестройке ОР за счёт изменения межзеркального расстояния $2l$ при фиксированном диаметре проводящего цилиндра ($2a = \mathrm{const}$). Следует отметить, что независимо от способа перестройки ОР ось проводящего цилиндра всегда совпадала с пересечением плоскостей симметрии xy и yz исследуемого резонатора.

В дальнейшем для сравнения полученных численных результатов с экспериментальными данными геометрические параметры ОР выбирались для диапазона длин волн $\lambda \approx 10$ мм. Так, для пустого ОР при $R_{\mathrm{c}} = 23$ мм, $l/R_{\mathrm{c}} = 0{,}57$ и $2\varphi_0 = 80^\circ$ были получены следующие значения дифракционной добротности основного типа колебаний TEM_{005} и близлежащих типов колебаний TEM_{104}, TEM_{204}: $Q_{005} = 7\,000$, $Q_{104} = 570$, $Q_{204} = 390$. Таким образом, для выбранных геометрических параметров ОР наблюдается эффективное высвечивание высших типов колебаний и приемлемый уровень дифракционной добротности основного типа колебаний TEM_{005}.

В ходе вычислительного эксперимента при симметричном размещении проводящей вставки между зеркалами ОР было обнаружено резонансное повышение дифракционной добротности нечётных типов колебаний TEM_{00q}, где $q = 3, 5, 7$, для двух значений диаметра проводящей вставки, причём больший из диаметров проводящей вставки составляет $2a \sim \lambda_{\mathrm{r}}/2$, где λ_{r} — длина волны в ОР.

Рассмотрим более подробно поведение дифракционной добротности квази-TEM_{005}-колебания в зависимости от диаметра проводящей вставки, симметрично размещённой между зеркалами ОР. Геометрические параметры зеркал ОР были следующими: $R_{\mathrm{c}} = 23$ мм, $l/R_{\mathrm{c}} = 0{,}57$ и $2\varphi_0 = 80^\circ$. Результаты расчётов дифракционной добротности и резонансной частоты квази-TEM_{005}-колебания в зависимости от диаметра проводящей вставки приведены на рис. 2. Как видно, размещение в ОР вдоль оси y проводящей вставки малого диаметра ($0 < 2a < 0{,}2$ мм) практически не приводит к изменению частоты и добротности H-поляризованного квази-TEM_{005}-колебания в ОР.

При дальнейшем увеличении диаметра проводящего цилиндра происходит изменение как дифракционной добротности, так и резонансной частоты возбуждаемого типа колебаний. Для двух диаметров проводящего цилиндра $2a = 1{,}15$ мм ($a/R_{\mathrm{c}} = 0{,}025$) и $2a = 5{,}14$ мм ($a/R_{\mathrm{c}} = 0{,}112$) наблюдается резонансное увеличение дифракционной добротности квази-TEM_{005}-колебания более чем на два порядка по сравнению с добротностью колебания TEM_{005} в пустом резонаторе. В интервалах значений диаметра проводящего цилиндра $1{,}86$ мм $< 2a < 4{,}58$ мм и $2a > 5{,}50$ мм, наоборот, происходит снижение дифракционной добротности квази-TEM_{005}-колебания по сравнению с добротностью колебания TEM_{005} в пустом резонаторе.

Рассмотрим характер поведения частоты и добротности квази-TEM_{007}-колебания в зависимости от диаметра проводящей вставки. Геометрические параметры ОР были следующими: $R_{\mathrm{c}} = 19{,}15$ мм, $l/R_{\mathrm{c}} = 0{,}94$, $2\varphi_0 = 100^\circ$. Как видно на рис. 3, при увеличении межзеркального расстояния и продольного индекса ($q = 7$) резонансы дифракционной добротности квази-TEM_{00q}-колебаний теряют свою остроту, а малый и большой резонансные диаметры проводящей вставки сближаются. Так, первый максимум добротности квази-TEM_{007}-колебания наблюдается при диаметре проводящей вставки $2a = 2{,}30$ мм ($a/R_{\mathrm{c}} = 0{,}060$), а второй максимум — при $2a = 4{,}02$ мм ($a/R_{\mathrm{c}} = 0{,}105$).

Следует отметить, что для квази-TEM_{007}-колебания оба резонансных диаметра проводящего

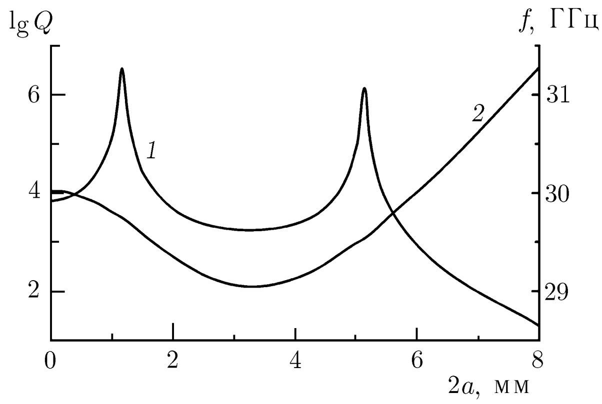


Рис. 2. Изменение добротности (кривая *1*) и резонансной частоты (кривая *2*) квази-TEM_{005}-колебания при увеличении диаметра проводящего цилиндра

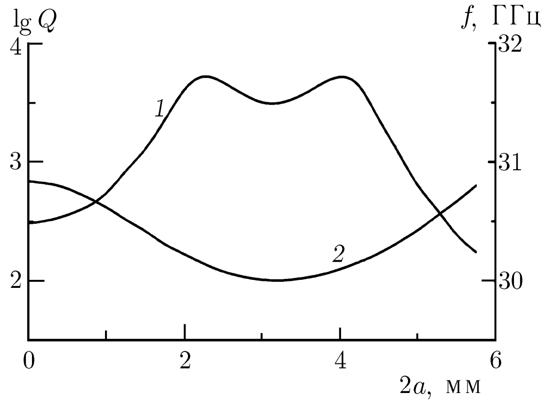


Рис. 3. Изменение добротности (кривая *1*) и резонансной частоты (кривая *2*) квази-TEM_{007}-колебания в зависимости от диаметра проводящей вставки

цилиндра уже существенно отличаются от $\lambda_{\text{г}}/2$. Более того, при изменении диаметра проводящего цилиндра в интервале 2,0 мм $< 2a <$ 4,5 мм добротность квази-TEM_{007}-колебания остаётся на порядок выше, чем в пустом резонаторе.

2.2. Физическая природа эффекта повышения добротности колебаний

Для выяснения физической причины повышения дифракционной добротности колебаний за счёт вставки в виде проводящего цилиндра были проведены расчёты структуры магнитной компоненты поля в объёме ОР и распределения амплитуды тока на зеркалах резонатора для TEM_{005}-колебания в конфокальном ОР с геометрическими параметрами $R_{\text{c}} = 24{,}0$ мм, $l/R_{\text{c}} = 0{,}5$, $2\varphi_0 = 80^\circ$. Как оказалось, размещение в объёме ОР проводящего цилиндра с первым (меньшим) резонансным значением диаметра ($a/R_{\text{c}} = 0{,}0163$) приводит к снижению амплитуды тока на краях зеркал резонатора и приближению распределения амплитуды тока на зеркалах к гауссовому виду (рис. 4*а*). Анализ распределения магнитной составляющей поля в объёме ОР показал, что размещение проводящего цилиндра между зеркалами резонатора приводит к устранению боковых лепестков, возникающих вследствие затекания токов на наружную поверхность зеркал. Это хорошо видно при отображении магнитной компоненты поля с низкой амплитудой $H_y = (0{,}01{\div}0{,}10)\, H_{y_{\max}}$ (см. рис. 4*б, в*). При этом структура поля в объёме ОР тоже приближается к гауссовому виду, а диаметр пятна поля колебания квази-TEM_{005} в плоскости xy остаётся таким же, как и в пустом резонаторе. Следовательно, при меньшем диаметре проводящей вставки происходит противофазное сложение за пределами резонатора собственного поля квази-TEM_{005}-колебания и поля, рассеянного на проводящей вставке.

Для второго (большего) резонансного диаметра вставки ($a/R_{\text{c}} = 0{,}101$) наблюдается другой механизм повышения дифракционной добротности колебания TEM_{005}. В этом случае происходит как снижение амплитуды тока на краях зеркал, так и стягивание распределения амплитуды тока к центру зеркал (см. рис. 4*а*), что также приводит к снижению дифракционных потерь. Следует отметить, что для второго резонанса ($a/R_{\text{c}} = 0{,}101$) происходит также стягивание поля во всём объёме резонатора к его оси (рис. 4*г*), т. е. диаметр пятна поля для колебания квази-TEM_{005} в плоскости xy уменьшается по сравнению с диаметром пятна поля в пустом резонаторе. Причём оптимальным является проводящий цилиндр с диаметром $2a \approx \lambda_{\text{г}}/2$, расположенный в объёме

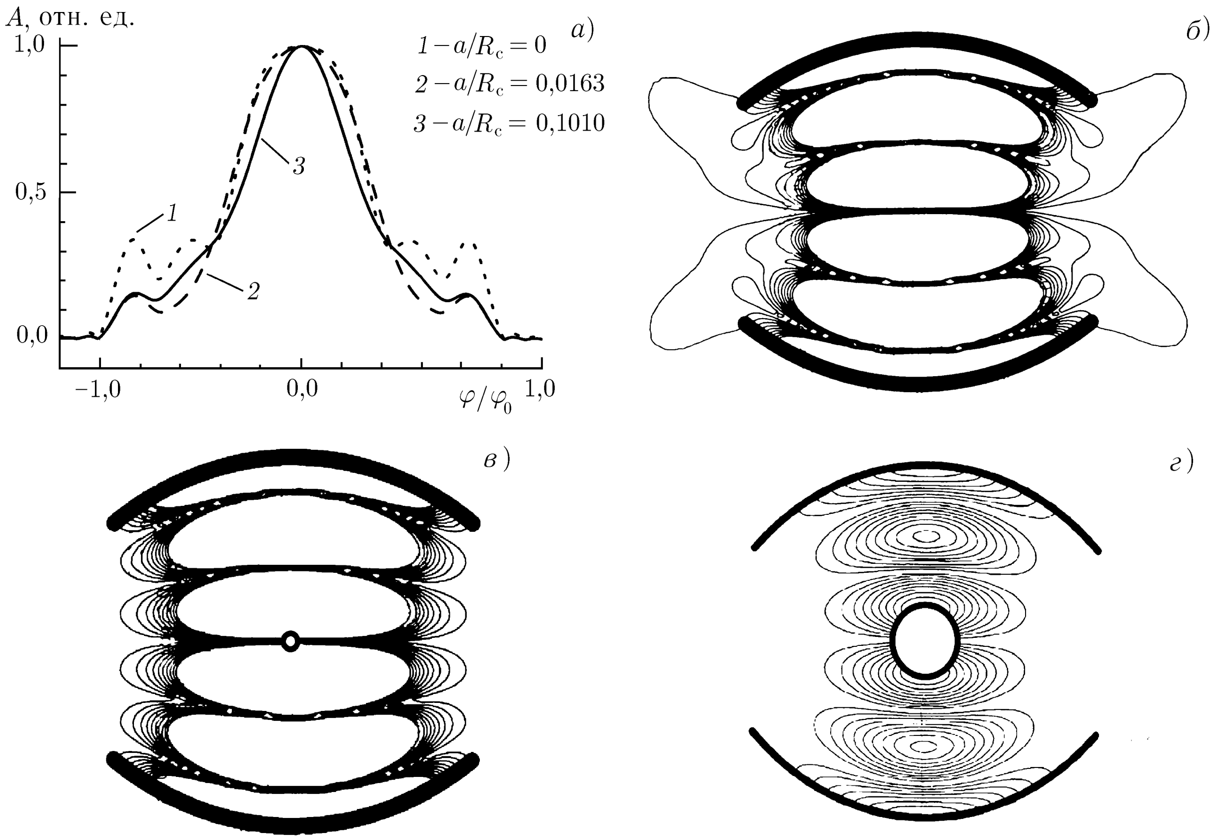


Рис. 4. Распределение амплитуды тока на зеркале с апертурой $2\varphi_0 = 80°$ (*а*) и магнитной составляющей поля в объёме ОР (*б*) при $a/R_{\rm c} = 0$ (шаг изолиний 0,01), (*в*) при $a/R_{\rm c} = 0{,}0163$ (шаг изолиний 0,01) и (*г*) при $a/R_{\rm c} = 0{,}101$ (шаг изолиний 0,1)

ОР таким образом, что его поверхность пересекает пучности магнитного поля стоячей волны в ОР. Это отчётливо видно на рис. 4*г*, где приведены линии равных амплитуд для магнитной составляющей поля $H_y = (0{,}1{\div}1{,}0)\,H_{y_{\max}}$.

Таким образом, эффект повышения добротности для колебания квази-ТЕМ$_{005}$ при большем диаметре проводящей вставки связан как с противофазным сложением полей за пределами резонатора, так и с перестройкой структуры поля в самом резонаторе. Здесь прослеживается некоторая аналогия с эффектом аномального рассеяния поля собственного колебания ОР на проводящем бруске квадратного сечения со стороной квадрата порядка $\lambda_{\rm г}/2$, размещённом в объёме резонатора [5].

Для подтверждения высказанных выше предположений о природе резонансного повышения добротности колебаний в ОР с цилиндрической проводящей вставкой мы провели сравнительный анализ поведения добротности квази-ТЕМ$_{005}$-колебания в ОР с идеально проводящим металлическим цилиндром и в ОР с диэлектрическими цилиндрами, имеющими разную диэлектрическую проницаемость ε ($\operatorname{Im}\varepsilon = \varepsilon'' = 0$). Геометрические параметры зеркал резонатора были следующие: $R_{\rm c} = 23$ мм, $l/R_{\rm c} = 0{,}534$, $2\varphi_0 = 80°$. Результаты расчётов добротности колебания квази-ТЕМ$_{005}$ в зависимости от диаметра цилиндрической вставки приведены на рис. 5. Как видно на рис. 5, для первого (меньшего) резонансного диаметра цилиндрической вставки природа эффекта повышения добротности квази-TEM_{005}-колебания связана с коэффициентом отражения электромагнитной волны от материала цилиндрической вставки. Чем выше диэлектрическая проницаемость материала диэлектрической вставки, тем меньше должен быть диаметр диэлектрического стрежня для обеспечения необходимой амплитуды рассеянной волны, компенсирующей поле за пределами резонатора. Для больши́х значений ε резонансный диаметр диэлектрической вставки приближается к диаметру идеально проводящей вставки.

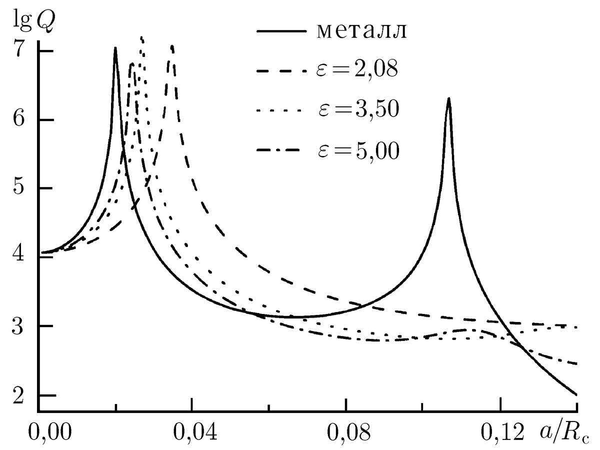


Рис. 5. Изменение добротности колебания квази-TEM_{005} в зависимости от диаметра и материала цилиндрической вставки

Для второго (большего) резонансного диаметра цилиндрической вставки природа эффекта повышения добротности колебания квази-TEM_{005} для ОР с проводящей вставкой и ОР с диэлектрический вставкой различна. Так, для ОР с диэлектрической вставкой наблюдающееся резонансное увеличение добротности при $a/R_c = 0{,}112$ ($\varepsilon = 5{,}0$) слабо выражено, а добротность колебания квази-TEM_{005} меньше, чем в пустом ОР.

2.3. Селекция колебаний и диапазон одномодовой перестройки в ОР с проводящей цилиндрической вставкой

При симметричном расположении проводящего цилиндра между зеркалами ОР эффект повышения добротности TEM_{00q}-колебаний существует только для H-поляризованных колебаний с нечётным продольным индексом (TEM_{003}, TEM_{005}, TEM_{007}), т. е. наблюдается дополнительное разрежение спектра ОР как по поляризации возбуждаемых колебаний, так и по продольному индексу за счёт высвечивания близлежащих продольных мод $TEM_{00\,q-1}$ и $TEM_{00\,q+1}$.

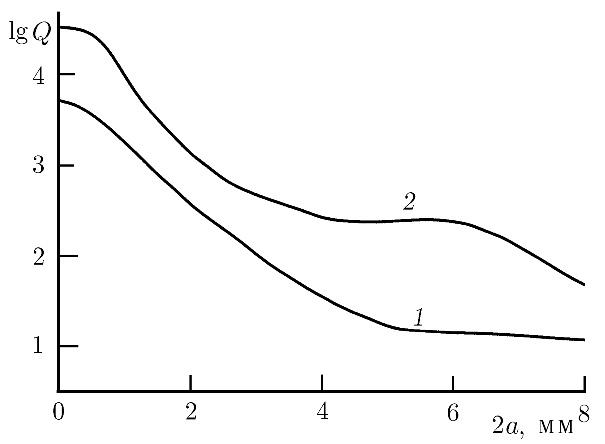


Рис. 6. Добротность колебаний квази-TEM_{004} (кривая *1*) и квази-TEM_{006} (кривая *2*) в зависимости от диаметра проводящей вставки

Высвечивание колебаний с чётным продольным индексом при симметричном размещении в ОР проводящей вставки проиллюстрирован на рис. 6. Здесь представлены результаты расчёта добротности колебаний квази-TEM_{004} и квази-TEM_{006} в зависимости от диаметра проводящего цилиндра ($R_c = 23$ мм, $l/R_c = 0{,}57$ и $2\varphi_0 = 80°$). Как видно на рис. 6, с увеличением диаметра проводящей вставки происходит монотонное снижение добротности колебаний с чётным продольным индексом, а при диаметре проводящей вставки $2a > 4$ мм добротность колебаний квази-TEM_{004} и квази-TEM_{006} снижается более чем на 2 порядка. Это позволяет реализовать ОР с широким диапазоном перестройки практически на одном типе колебаний. Так, для пустого ОР резонансная частота колебания TEM_{004} составляла $f_{004} = 25{,}226$ ГГц, а колебания TEM_{006} — $f_{006} = 37{,}142$ ГГц. Следовательно, при размещении в объёме ОР проводящего цилиндра с диаметром $2a = 5{,}14$ мм (см. второй резонанс по добротности на рис. 2) имеется возможность осуществить одномодовую перестройку резонатора на колебании TEM_{005} в полосе частот 25÷37 ГГц.

При перестройке ОР за счёт изменения расстояния между цилиндрическими зеркалами диаметр пятна поля в окрестности проводящего цилиндра меняется незначительно, поэтому эффект повышения дифракционной добротности квази-TEM_{005}-колебания наблюдается в широкой полосе частот (см. рис. 7, кривые *1* и *2*). Расчёты выполнены для ОР с радиусом кривизны зеркал $R_\mathrm{c} = 23$ мм, апертурой зеркал $2\varphi_0 = 80^\circ$ и цилиндрической вставкой с диаметром $2a = 1{,}15$ мм в диапазоне длин волн $\lambda \approx 10$ мм. Омические потери в зеркалах резонатора ещё более снижают резонансный характер повышения добротности колебаний в ОР с проводящим цилиндром за счёт снижения дифракционных потерь. Так, с учётом омических потерь в зеркалах суммарная добротность Q_Σ для TEM_{00q}-колебаний составляет

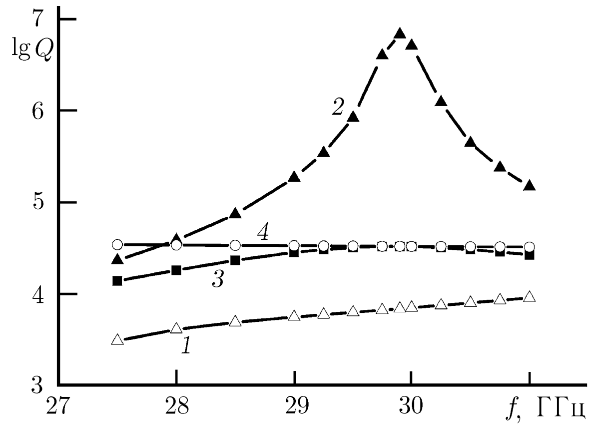


Рис. 7. Сравнение добротности колебаний TEM_{005} в пустом ОР и в резонаторе с проводящей вставкой: кривая *1* — дифракционная добротность колебания TEM_{005} в пустом ОР, *2* — дифракционная добротность квази-TEM_{005}-колебаний при $2a = 1{,}15$ мм, *3* — суммарная добротность квази-TEM_{005}-колебания с учётом омических потерь, *4* — омическая добротность TEM_{005} колебания в пустом ОР

$$\frac{1}{Q_\Sigma} = \frac{1}{Q_\mathrm{D}} + \frac{1}{Q_\Omega} = \frac{1}{Q_\mathrm{D}} + \frac{1}{\pi q}\,\frac{4R_\mathrm{S}}{W_0}\,, \qquad (2)$$

где Q_D — дифракционная добротность квази-TEM_{00q}-колебаний в резонаторе с проводящей вставкой, Q_Ω — омическая добротность TEM_{00q}-колебаний в пустом ОР, R_S — действительная часть поверхностного импеданса зеркал, W_0 — волновое сопротивление свободного пространства. Здесь омические потери в самой цилиндрической вставке не учитываются, т. к. при резонансном диаметре $2a = 1{,}15$ мм площадь поверхности проводящей вставки намного меньше, чем площадь пятна поля на зеркалах ОР.

Таким образом, для ОР с зеркалами из меди размещение в объёме резонатора проводящего цилиндра позволяет повысить добротность колебания квази-TEM_{005} в 3÷6 раз в широкой полосе частот. На частоте резонанса $f = 30{,}0$ ГГц удаётся практически достичь омической добротности TEM_{005}-колебания в пустом ОР (см. рис. 7, кривые *3* и *4*).

2.4. Несимметричное размещение проводящей цилиндрической вставки

Резонансное повышение добротности для TEM_{00q}-колебаний можно ожидать и при несимметричном размещении проводящей вставки в открытом резонаторе, например в соседних пучностях электрического поля. Нами были проведены исследования свойств квази-TEM_{005}-колебания при перемещении проводящей вставки вдоль оси z между зеркалами ОР. Исследования проведены в 10-миллиметровом диапазоне длин волн для конфокального ОР со следующими геометрическими параметрами: $R_\mathrm{c} = 2l = 24{,}15$ мм, $2\varphi_0 = 80^\circ$. Для симметрично расположенной цилиндрической вставки перый резонанс по добротности TEM_{005}-колебания наблюдался при $a/R_\mathrm{c} = 0{,}016$. Изменение резонансной частоты исследуемого колебания при смещении проводящей вставки (при $a/R_\mathrm{c} = 0{,}016$) вдоль оси z носит периодический характер (см. рис. 8*а*), а максимумы добротности колебания квази-TEM_{005} наблюдаются при размещении вставки в пучностях электрического поля стоячей волны в ОР (рис. 8*б*).

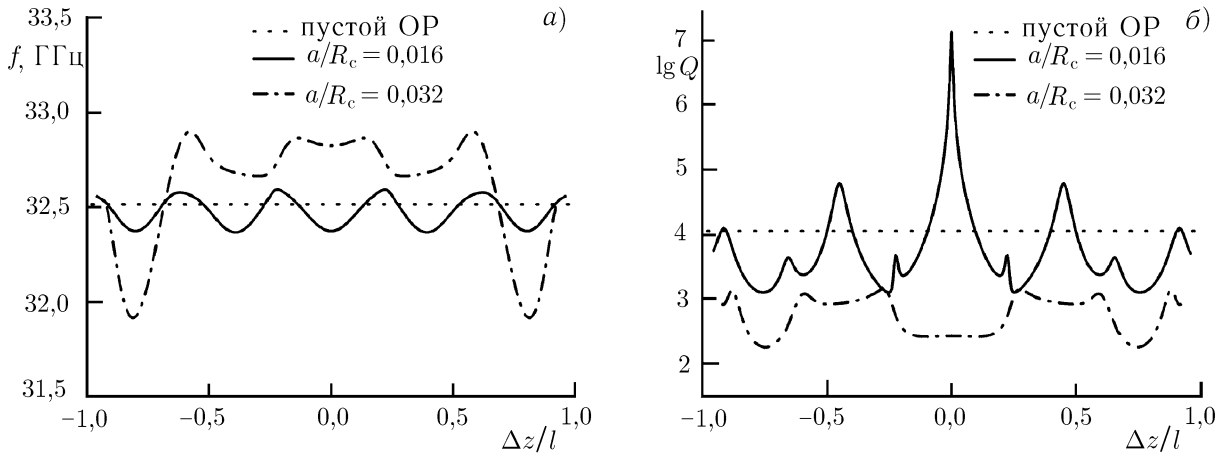


Рис. 8. Зависимость резонансной частоты (*а*) и добротности (*б*) для квази-TEM_{005}-колебания от смещения проводящей цилиндрической вставки вдоль оси z

Отметим, что при размещении проводящей вставки в соседних с центральной пучностях стоячей волны (при $\Delta z/l = \pm 0{,}45$) добротность квази-TEM_{005}-колебания выше, чем добротность колебания TEM_{005} в пустом резонаторе. Правда, повышение добротности колебаний при несимметричном размещении проводящей вставки в ОР уже не столь значительно, как при $\Delta z/l = 0$. При смещении вдоль оси z проводящей вставки с нерезонансным диаметром ($a/R_{\text{c}} = 0{,}032$) изменение резонансной частоты и добротности квази-TEM_{005}-колебания носит сложный характер (рис. 8), а добротность колебания снижается на 1,5−2 порядка по сравнению с добротностью TEM_{005}-колебания в пустом ОР.

Для чётных типов колебаний TEM_{004}, TEM_{006} также наблюдается резонансное повышение добротности при размещении проводящей вставки в пучностях электрического поля с нарушением симметрии ОР. Так, для TEM_{006}-колебания в конфокальном ОР ($R_{\text{c}} = 23$ мм, $l/R_{\text{c}} = 0{,}5$ и $2\varphi_0 = 80^\circ$) наибольший эффект резонансного повышения добротности наблюдается при размещении проводящей вставки в третьей или четвёртой пучностях электрического поля, расположенных вблизи плоскости симметрии резонатора (xy). При внесении проводящей вставки с меньшим резонансным диаметром ($a/R_{\text{c}} = 0{,}0118$) добротность TEM_{006}-колебания повышается с $\lg Q = 4{,}08$ в пустом ОР до $\lg Q = 6{,}06$. Для большего резонансного диаметра проводящей вставки ($a/R_{\text{c}} = 0{,}0785$) повышение добротности TEM_{006}-колебания уже не столь значительно ($\lg Q = 4{,}26$).

ВЫВОДЫ

Обнаружен и исследован эффект резонансного повышения добротности колебаний в ОР со вставкой в виде проводящего кругового цилиндра. При помощи вычислительного эксперимента обнаружены два резонансных диаметра проводящей вставки и установлена физическая природа повышения добротности квази-TEM_{00q}-колебаний для меньшего и большего резонансных диаметров проводящей вставки.

Рассмотрены возможности управления спектром, частотой и добротностью колебаний, возбуждаемых в ОР с проводящими вставками. Показано, что с использованием проводящих вставок имеется практическая возможность реализовать ОР с перестройкой на одном (основном) типе колебаний с добротностью, приближающейся к омическому пределу.

Второй (больший) резонансный диаметр проводящего цилиндра составляет $2a \approx \lambda_{\text{г}}/2$, что вполне достаточно для обеспечения подвода питающих напряжений и эффективного отвода тепла от активных элементов, размещаемых в объёме ОР.

Работа выполнена при частичной финансовой поддержке фонда УНТЦ (проект № 1770).

СПИСОК ЛИТЕРАТУРЫ

1. Шестопалов В. П. Физические основы миллиметровой и субмиллиметровой техники. Т. 1. Открытые структуры. Киев: Наук. думка, 1985. 216 с.
2. Вайнштейн Л. А. Открытые резонаторы и открытые волноводы. М.: Сов. радио, 1966. 474 с.
3. Коцержинский Б. А., Мачусский Е. А., Першин Н. А. и др. // Изв. вузов. Радиоэлектроника. 1987. Т. 30, № 10. С. 13.
4. Бородкин А. И., Селезнёв Д. Г., Смородин В. В. // Изв. вузов. Радиоэлектроника. 1981. Т. 24, № 10. С. 27.
5. Булгаков Б. М., Велиев Э. И., Веремей В. В. и др. // ЖТФ. 1990. Т. 60, № 6. С. 182.
6. Бровенко А. В., Мележик П. Н., Поединчук А. Е. и др. // Радиотехника и электроника. 1997. Т. 42, № 10. С. 1 174.
7. Мележик П. Н., Бровенко А. В., Поединчук А. Е. // Изв. вузов. Радиофизика. 1998. Т. 41, № 10. С. 1 336.

Поступила в редакцию 8 октября 2004 г.; принята в печать 18 марта 2005 г.

OPEN RESONATORS WITH CONDUCTIVE CYLINDRICAL INSERTS. 1. TWO DIMENSIONAL MODEL

P. N. Melezhik, V. S. Miroshnichenko, and Ye. B. Senkevich

We present the results of a computing experiment on the study of properties of oscillations in a two dimensional open resonator with conductive cylindrical insert. An increase in the diffractive Q-factor of the fundamental oscillation mode by 2–3 orders due to the conductive cylindrical insert is found and studied. The prospects of using this phenomenon to ensure control over the spectrum, frequency, and Q-factor of the excited oscillations are discussed. The possibility of retuning the open resonator with conductive cylindrical insert for one (fundamental) oscillation mode in the wide frequency range is shown.

УДК 621.371+621.372.833.1

ДИФРАКЦИЯ ЭЛЕКТРОМАГНИТНЫХ ВОЛН НА АНИЗОТРОПНОЙ ЦИЛИНДРИЧЕСКОЙ НЕОДНОРОДНОСТИ В ПЛОСКОМ ВОЛНОВОДЕ

Е. А. Шорохова [1], *В. А. Яшнов* [2]

[1] ФГУП «ФНПЦ НИИ измерительных систем им Ю. Е. Седакова»;
[2] Нижегородский госуниверситет им. Н. И. Лобачевского, г. Нижний Новгород, Россия

Рассмотрена дифракция электромагнитных волн на анизотропной цилиндрической неоднородности, расположенной в плоском волноводе с идеально проводящими стенками. Анизотропия учтена с использованием приближения одноосного кристалла. Представлено строгое аналитическое решение в виде двойных сумм по собственным функциям плоского волновода с идеально проводящими стенками и азимутальным собственным функциям цилиндра. Проведены численные расчёты компонент напряжённости электрического поля, рассеянного на анизотропной неоднородности. Проанализировано влияние анизотропии и размеров неоднородности на рассеянное поле.

ВВЕДЕНИЕ

В последнее время большое внимание уделяется исследованию особенностей распространения электромагнитных волн внутри плоских волноводов, содержащих регулярные неоднородности. Это связано с широким использованием модели плоских волноводов в гидроакустике, сейсмологии, в теории распространения радиоволн ОНЧ диапазона [1–6]. Хорошо известны и широко применяются модели плоского волновода Земля—ионосфера [2–5] и литосферного волновода [7–9]. Важно отметить, что существенное влияние на распределение электромагнитного поля внутри волноводов оказывают как электрические, так и геометрические свойства неоднородностей [10, 11].

Интерес к проблемам распространения радиоволн в земных породах обусловлен, прежде всего, необходимостью радиосвязи с подземными объектами, расположенными в рудниках, шахтах, горных выработках и т. д. [12]. Особенно актуальной задача радиосвязи через массивы земной породы становится в аварийных ситуациях: при взрывах, обвалах и т. п. Для радиосвязи через однородные горные массивы существуют оптимальные частоты, что обусловлено, с одной стороны, уменьшением затухания радиоволн в проводящих средах с уменьшением частоты, а с другой — уменьшением эффективности антенн с уменьшением их электрических размеров.

Неоднородности земной породы, в качестве которых можно рассматривать скважины, шахты, вулканические каналы и другие цилиндрические структуры как естественного, так и искусственного происхождения [12], при определённых условиях могут способствовать увеличению дальности подземной радиосвязи. Например, в угольных шахтах, где проводимость угля достаточно низка по сравнению с проводимостью окружающих пород, наблюдается волноводный механизм распространения радиоволн.

В скважинных геофизических исследованиях при поисках полезных ископаемых или обнаружении карстовых пустот и провалов возникает аналогичная ситуация: использование в скважинах бурового раствора, проводимость которого на несколько порядков превышает проводимость окружающего пространства, способствует волноводному механизму распространения радиоволн [13]. Другим примером такого механизма может служить распространение радиоволн ОНЧ и СНЧ диапазонов в литосфере Земли. Литосферный волновод представляет собой слой плохо проводящих базальтовых пород толщиной $20\div30$ км, расположенный на глубине от 2 до

20 км под земной поверхностью или дном океана. Снизу такой волновод ограничен хорошо проводящей мантией, а сверху — слоем осадочных пород, проводимость которых также велика по сравнению с проводимостью базальтовых пород.

Детальное исследование особенностей распространения электромагнитных волн в литосферном волноводе проведено в работе [7], где использована простая модель волновода в виде плоского плохо проводящего слоя, ограниченного с двух сторон проводящими полупространствами. В [8] проанализировано распространение радиоволн ОНЧ и СНЧ диапазонов в литосферном волноводе с использованием модели, включающей четыре среды: океан, слой осадочных пород, слой базальтовых пород и мантию Земли. В рамках такой модели исследованы характеристики нормальных волн и полей, возбуждаемых вертикальным электрическим диполем, расположенным внутри волновода. В этой работе рассмотрено несколько способов возбуждения литосферного волновода и показано, что наиболее предпочтительным является вариант, когда источник находится внутри волновода. Если же излучатель невозможно поместить в волновод, то горизонтальный диполь, расположенный в верхнем полупространстве, более эффективно возбуждает волновод по сравнению с размещённым там же вертикальным диполем. Аналогичная ситуация имеет место при возбуждении волновода Земля—ионосфера источниками, расположенными в ионосфере [14].

Заметим, что задачам дифракции нормальных волн на цилиндре внутри плоского волновода уже уделялось значительное внимание [9, 15, 16]. В частности, в [9] рассмотрена дифракция нормальных волн на цилиндрической неоднородности с заданным поверхностным импедансом, расположенной в плоском волноводе с импедансными границами. При этом определены собственные функции такого волновода, а методом последовательных приближений — и его собственные значения. Получены выражения для продольной компоненты напряжённости электрического поля, с помощью которых проведены численные расчёты дифракции низкочастотных радиоволн на цилиндрической неоднородности внутри литосферного волновода. Показано, что при расчётах электромагнитных полей в плоских слоях, ограниченных породами с высокой удельной проводимостью, можно использовать модель волновода с идеально проводящими стенками.

Работы [15] и [16] посвящены дифракции электромагнитных волн на идеально проводящем и магнитодиэлектрическом цилиндрах соответственно, расположенных внутри плоского волновода с идеально проводящими стенками. Получены точные аналитические решения этих задач в виде двойных рядов и проанализированы дифракционные поля в волноводе. В [15] даны алгоритмы расчёта электромагнитных полей в приближении Рэлея и в коротковолновом приближении и оценено влияние идеально проводящего цилиндра на мощность излучения диполя в волноводе. Вопросы взаимодействия и трансформации двух типов нормальных волн (TE и TM) на границе изотропной цилиндрической неоднородности подробно исследованы в [16]. Особое внимание уделено анализу влияния такой неоднородности на энергетические характеристики излучения вертикального электрического диполя, расположенного в волноводе.

Здесь на основе результатов работы [17] проведено исследование рассеяния электромагнитных волн на анизотропной цилиндрической неоднородности, расположенной внутри плоского волновода с идеально проводящими стенками.

1. ПОСТАНОВКА ЗАДАЧИ. ОСНОВНЫЕ УРАВНЕНИЯ И ВЫРАЖЕНИЯ ДЛЯ ПОЛЕЙ

Рассмотрим плоский волновод с толщиной L, между идеально проводящими стенками которого расположена анизотропная неоднородность в виде круглого цилиндра радиуса b, как показано на рис. 1. Источником излучения электромагнитных волн является элементарный вертикальный электрический диполь, расположенный в плоском волноводе в точке с координатами (a, φ_0, z_0) в цилиндрической системе координат $(r,\ \varphi,\ z)$. Электромагнитные свойства неоднородности описываются диагональным тензором относительной диэлектрической проницаемости

$$\hat{\boldsymbol{\varepsilon}} = \begin{pmatrix} \varepsilon_{1\perp} & 0 & 0 \\ 0 & \varepsilon_{1\perp} & 0 \\ 0 & 0 & \varepsilon_{1\|} \end{pmatrix} \qquad (1)$$

и скалярной относительной магнитной проницаемостью μ_1. Предполагается, что волновод заполнен однородной изотропной средой, характеризуемой относительной диэлектрической проницаемостью ε_2 и магнитной проницаемостью μ_2.

Уравнения Максвелла для гармонических процессов $\exp(-i\omega t)$ и материальные уравнения для сред внутри и вне неоднородности можно записать в следующем виде:

1) внутри цилиндра (область 1)

$$\mathrm{rot}\,\mathbf{H}_1 = -i\omega\mathbf{D}_1, \qquad \mathbf{D}_1 = \hat{\boldsymbol{\varepsilon}}\varepsilon_0\mathbf{E}_1, \qquad (2)$$

$$\mathrm{rot}\,\mathbf{E}_1 = i\omega\mathbf{B}_1, \qquad \mathbf{B}_1 = \mu_1\mu_0\mathbf{H}_1; \qquad (3)$$

2) вне цилиндра (область 2)

$$\mathrm{rot}\,\mathbf{H}_2 = -i\omega\mathbf{D}_2 + \mathbf{j}, \qquad \mathbf{D}_2 = \varepsilon_2\varepsilon_0\mathbf{E}_2, \qquad (4)$$

$$\mathrm{rot}\, bfE_2 = i\omega\mathbf{B}_2, \qquad \mathbf{B}_2 = \mu_2\mu_0\mathbf{H}_2. \qquad (5)$$

Здесь ε_0 и μ_0 — электрическая и магнитная постоянные вакуума, ω — круговая частота, $\mathbf{j}$ — плотность стороннего тока, которая в цилиндрической системе координат определяется формулой

$$\mathbf{j} = \mathbf{e}_z\,\frac{I_0 l}{\sqrt{ra}}\,\delta(r-a)\delta(\varphi-\varphi_0)\delta(z-z_0) = \mathbf{e}_z j_z, \qquad (6)$$

где $\mathbf{e}_z$ — единичный вектор вдоль оси z, I_0 — амплитуда тока на частоте ω, l — эффективная длина диполя, $\delta(x)$ — дельта-функция Дирака.

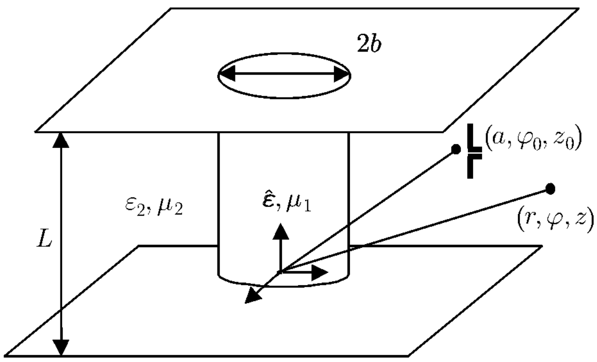


Рис. 1

Для решения поставленной задачи удобно ввести электрический ($\mathbf{A}$) и магнитный ($\mathbf{F}$) векторные потенциалы следующим образом:

$$\mathbf{B}_1 = \mathrm{rot}\,\mathbf{A}_1, \qquad \mathbf{B}_2 = \mathrm{rot}\,\mathbf{A}_2, \qquad (7)$$

$$\mathbf{D}_1 = -\mathrm{rot}\,\mathbf{F}_1, \qquad \mathbf{D}_2 = -\mathrm{rot}\,\mathbf{F}_2. \qquad (8)$$

Отметим, что в случае излучения вертикального диполя электрический и магнитный векторные потенциалы будут иметь лишь продольные составляющие (A_1, A_2 и F_1, F_2 соответственно).

Вертикальные компоненты потенциалов A_1, A_2 и F_1, F_2 удовлетворяют следующим уравнениям:

1) внутри цилиндра

$$\Delta_\perp A_1 + \frac{1}{\Lambda}\,\frac{\partial^2 A_1}{\partial z^2} + k_{1\|}^2 A_1 = 0, \tag{9}$$

$$\Delta F_1 + k_{1\perp}^2 F_1 = 0; \tag{10}$$

2) вне цилиндра

$$\Delta A_2 + k_2^2 A_2 = -\mu_0 \mu_2 j_z, \tag{11}$$

$$\Delta F_2 + k_2^2 F_2 = 0. \tag{12}$$

Здесь $k_{1\perp}$, $k_{1\|}$ и k_2 — волновые числа внутри цилиндра и вне его: $k_{1\perp} = k_0\sqrt{\varepsilon_{1\perp}\mu_1}$, $k_{1\|} = k_0\sqrt{\varepsilon_{1\|}\mu_1}$, $k_2 = k_0\sqrt{\varepsilon_2\mu_2}$; $k_0 = \omega\sqrt{\varepsilon_0\mu_0}$ — волновое число в вакууме, $\Lambda = \varepsilon_{1\perp}/\varepsilon_{1\|}$ — коэффициент анизотропии, Δ — оператор Лапласа,

$$\Delta_\perp = \frac{1}{r}\,\frac{\partial}{\partial r}\left(r\,\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\,\frac{\partial^2}{\partial \varphi^2}\ .$$

Следует заметить, что уравнения (9) и (10) описывают обыкновенную и необыкновенную волны, которые характерны для одноосного кристалла.

Сформулируем граничные условия на стенках волновода и на поверхности цилиндра. Во-первых, тангенциальные компоненты электрического поля на идеально проводящей поверхности равны нулю ($E_r = 0, E_\varphi = 0$), что приводит к следующим граничным условиям при $z = 0$ и $z = L$:

$$\partial A/\partial z = 0, \qquad F = 0. \tag{13}$$

Во-вторых, на поверхности круглого анизотропного диэлектрического цилиндра при $r = b$ должны выполняться условия непрерывности тангенциальных компонент напряжённостей электрического и магнитного полей:

$$E_{\varphi 1} = E_{\varphi 2}, \qquad H_{\varphi 1} = H_{\varphi 2}, \qquad E_{z1} = E_{z2}, \qquad H_{z1} = H_{z2}. \tag{14}$$

Из соотношений (14) легко получить граничные условия для потенциалов.

В соответствии с [17] запишем решение в виде двойных рядов:

$$A_1 = \sum_{n=-\infty}^{+\infty}\sum_{s=-\infty}^{+\infty} A_{1ns}(r)\cos\!\left(n\pi\,\frac{z_0}{L}\right)\cos\!\left(n\pi\,\frac{z}{L}\right)\cos[s\,(\varphi-\varphi_0)], \tag{15}$$

$$F_1 = \sum_{n=-\infty}^{+\infty}\sum_{s=-\infty}^{+\infty} F_{1ns}(r)\cos\!\left(n\pi\,\frac{z_0}{L}\right)\sin\!\left(n\pi\,\frac{z}{L}\right)\sin[s\,(\varphi-\varphi_0)], \tag{16}$$

$$A_2 = \sum_{n=-\infty}^{+\infty}\sum_{s=-\infty}^{+\infty} A_{2ns}(r)\cos\!\left(n\pi\,\frac{z_0}{L}\right)\cos\!\left(n\pi\,\frac{z}{L}\right)\cos[s\,(\varphi-\varphi_0)], \tag{17}$$

$$F_2 = \sum_{n=-\infty}^{+\infty}\sum_{s=-\infty}^{+\infty} F_{2ns}(r)\cos\!\left(n\pi\,\frac{z_0}{L}\right)\sin\!\left(n\pi\,\frac{z}{L}\right)\sin[s\,(\varphi-\varphi_0)], \tag{18}$$

где функции $A_{1ns}(r)$, $A_{2ns}(r)$, $F_{1ns}(r)$ и $F_{2ns}(r)$ удовлетворяют обыкновенным дифференциальным уравнениям

$$\frac{\mathrm{d}^2 A_{1ns}}{\mathrm{d}r^2} + \frac{1}{r}\,\frac{\mathrm{d}A_{1ns}}{\mathrm{d}r} + \left[k_{1\|}^2 - \frac{1}{\Lambda}\left(\frac{n\pi}{L}\right)^2 - \frac{s^2}{r^2}\right] A_{1ns} = 0, \tag{19}$$

$$\frac{\mathrm{d}^2F_{1ns}}{\mathrm{d}r^2}+\frac{1}{r}\,\frac{\mathrm{d}F_{1ns}}{\mathrm{d}r}+\left[k_{1\perp}^2-\left(\frac{n\pi}{L}\right)^2-\frac{s^2}{r^2}\right]F_{1ns}=0, \tag{20}$$

$$\frac{\mathrm{d}^2A_{2ns}}{\mathrm{d}r^2}+\frac{1}{r}\,\frac{\mathrm{d}A_{2ns}}{\mathrm{d}r}+\left[k_2^2-\left(\frac{n\pi}{L}\right)^2-\frac{s^2}{r^2}\right]A_{2ns}=-\frac{I_0l\mu_0\mu_2}{2\pi L}\,\frac{\delta(r-a)}{\sqrt{ra}}\,, \tag{21}$$

$$\frac{\mathrm{d}^2F_{2ns}}{\mathrm{d}r^2}+\frac{1}{r}\,\frac{\mathrm{d}F_{2ns}}{\mathrm{d}r}+\left[k_2^2-\left(\frac{n\pi}{L}\right)^2-\frac{s^2}{r^2}\right]F_{2ns}=0. \tag{22}$$

Общее решение уравнения (21) представим в виде

$$A_{2ns}=\frac{iI_0l\mu_0\mu_2}{4L}\times\begin{cases}J_s(\kappa_2r)H_s^{(1)}(\kappa_2a)+R_{ns}^{\mathrm{ee}}H_s^{(1)}(\kappa_2r), & r\le a;\\ J_s(\kappa_2a)H_s^{(1)}(\kappa_2r)+R_{ns}^{\mathrm{ee}}H_s^{(1)}(\kappa_2r), & r\ge a,\end{cases}$$

где κ_2 — радиальное волновое число в волноводе, определяемое выражением

$$\kappa_2=\sqrt{k_2^2-((n\pi)/L)^2}\,,$$

где $n=0,1,2,\ldots;$ J_s и $H_s^{(1)}$ — функции Бесселя и Ханкеля I-го рода порядка s. Тогда выражение для векторного потенциала A_2 в волноводе в соответствии с выражением (17) можно записать в виде

$$A_2=\frac{iI_0l\mu_0\mu_2}{4L}\times$$
$$\times\sum_{n,s=-\infty}^{+\infty}\begin{cases}\left[J_s(\kappa_2r)H_s^{(1)}(\kappa_2a)+R_{ns}^{\mathrm{ee}}H_s^{(1)}(\kappa_2r)\right]\cos(n\pi z_0/L)\cos(n\pi z/L\cos[s\,(\varphi-\varphi_0)], & r\le a;\\ \left[J_s(\kappa_2a)H_s^{(1)}(\kappa_2r)+R_{ns}^{\mathrm{ee}}H_s^{(1)}(\kappa_2r)\right]\cos(n\pi z_0/L)\cos(n\pi z/L)\cos[s\,(\varphi-\varphi_0)], & r\ge a.\end{cases} \tag{23}$$

Заметим, что первое слагаемое в квадратных скобках формулы (23) соответствует прямой волне, распространяющейся от источника до точки наблюдения, а второе — волне, рассеянной на анизотропной неоднородности.

Аналогичным образом можно записать частные решения однородных дифференциальных уравнений (19), (20) и (22), подстановка которых в (15), (16) и (18) даёт следующие выражения для потенциалов:

$$A_1=\frac{iI_0l\mu_0\mu_2}{4L}\sum_{n=-\infty}^{+\infty}\sum_{s=-\infty}^{+\infty}T_{ns}^{\mathrm{ee}}J_s(\kappa_{1\parallel}r)\cos\!\left(n\pi\,\frac{z_0}{L}\right)\cos\!\left(n\pi\,\frac{z}{L}\right)\cos[s\,(\varphi-\varphi_0)], \tag{24}$$

$$F_1=\frac{iI_0l\mu_0\mu_2\zeta_0}{4L}\sum_{n=-\infty}^{+\infty}\sum_{s=-\infty}^{+\infty}T_{ns}^{\mathrm{em}}J_s(\kappa_{1\perp}r)\cos\!\left(n\pi\,\frac{z_0}{L}\right)\sin\!\left(n\pi\,\frac{z}{L}\right)\sin[s\,(\varphi-\varphi_0)], \tag{25}$$

$$F_2=\frac{iI_0l\mu_0\mu_2\zeta_0}{4L}\sum_{n=-\infty}^{+\infty}\sum_{s=-\infty}^{+\infty}R_{ns}^{\mathrm{em}}H_s^{(1)}(\kappa_2r)\cos\!\left(n\pi\,\frac{z_0}{L}\right)\sin\!\left(n\pi\,\frac{z}{L}\right)\sin[s\,(\varphi-\varphi_0)]. \tag{26}$$

Здесь введены обозначения для импеданса свободного пространства $\zeta_0=\sqrt{\mu_0/\varepsilon_0}$ и радиальных волновых чисел $\kappa_{1\perp}$ и $\kappa_{1\parallel}$:

$$\kappa_{1\perp}=\sqrt{k_{1\perp}^2-(n\pi/L)^2}\,,\qquad \kappa_{1\parallel}=\sqrt{k_{1\parallel}^2-(n\pi/L)^2}\,.$$

Отметим, что решения (23) и (26) удовлетворяют условию излучения на бесконечности.

В соотношениях (24) и (25) коэффициент $T_{ns}^{\rm ee}$ описывает амплитуду электрических волн, прошедших внутрь цилиндра, $T_{ns}^{\rm em}$ — амплитуду магнитных волн внутри цилиндра, которые возникают в результате трансформации электрических волн, падающих на поверхность неоднородности; $R_{ns}^{\rm em}$ — коэффициент трансформации электрических волн в магнитные при отражении от цилиндра. Эти коэффициенты вместе с коэффициентом $R_{ns}^{\rm ee}$ определяются из граничных условий на цилиндрической поверхности неоднородности (14). Решение соответствующей системы уравнений приводит к выражениям

$$R_{ns}^{\rm ee} = -\,\frac{(D_n^\mu M_n^\varepsilon - s^2C_n^2)}{(D_n^\mu D_n^\varepsilon - s^2C_n^2)}\,\frac{J_s(\beta_2)}{H_s^{(1)}(\beta_2)}\,H_s^{(1)}(\alpha_2), \tag{27}$$

$$R_{ns}^{\rm em} = -\,\frac{2\varepsilon_2}{\pi\beta_2^2}\,\frac{sC_n}{(D_n^\mu D_n^\varepsilon - s^2C_n^2)}\,\frac{H_s^{(1)}(\alpha_2)}{[H_s^{(1)}(\beta_2)]^2}\ , \tag{28}$$

$$T_{ns}^{\rm ee} = -i\,\frac{2\varepsilon_{1\perp}}{\pi\beta_{1\perp}^2}\,\frac{\mu_1}{\mu_2}\,\frac{D_n^\mu}{(D_n^\mu D_n^\varepsilon - s^2C_n^2)}\,\frac{H_s^{(1)}(\alpha_2)}{J_s(\beta_{1\|})H_s^{(1)}(\beta_2)}\ , \tag{29}$$

$$T_{ns}^{\rm em} = \frac{2\varepsilon_{1\perp}}{\pi\beta_{1\perp}^2}\,\frac{\mu_1}{\mu_2}\,\frac{sC_n}{(D_n^\mu D_n^\varepsilon - s^2C_n^2)}\,\frac{H_s^{(1)}(\alpha_2)}{J_s(\beta_{1\perp})H_s^{(1)}(\beta_2)}\ . \tag{30}$$

В формулах (27)–(30) использованы следующие обозначения:

$$D_n^\mu = \frac{\mu_1}{\beta_{1\perp}}\,\frac{\dot J_s(\beta_{1\perp})}{J_s(\beta_{1\perp})} - \frac{\mu_2}{\beta_2}\,\frac{\dot H_s^{(1)}(\beta_2)}{H_s^{(1)}(\beta_2)}\ ,\qquad D_n^\varepsilon = \frac{\varepsilon_{1\perp}\beta_{1\|}}{\beta_{1\perp}^2}\,\frac{\dot J_s(\beta_{1\|})}{J_s(\beta_{1\|})} - \frac{\varepsilon_2}{\beta_2}\,\frac{\dot H_s^{(1)}(\beta_2)}{H_s^{(1)}(\beta_2)}$$

$$M_n^\varepsilon = \frac{\varepsilon_{1\perp}\beta_{1\|}}{\beta_{1\perp}^2}\,\frac{\dot J_s(\beta_{1\|})}{J_s(\beta_{1\|})} - \frac{\varepsilon_2}{\beta_2}\,\frac{\dot J_s(\beta_2)}{J_s(\beta_2)}\ ,\qquad C_n = \frac{n\pi}{k_0L}\,\frac{\beta_2^2 - \beta_{1\perp}^2}{\beta_{1\perp}^2\beta_2^2}\ ,$$

$$\beta_{1\perp} = \kappa_{1\perp}b,\qquad \beta_{1\|} = \kappa_{1\|}b,\qquad \beta_2 = \kappa_2 b,\qquad \alpha_2 = \kappa_2 a;$$

точка над функцией обозначает производную по аргументу.

Можно отметить, что дифракционные электромагнитные поля полностью описываются потенциалами A_2 и F_2, т. е. коэффициентами $R_{ns}^{\rm ee}$ и $R_{ns}^{\rm em}$. Чтобы определить электромагнитное поле внутри цилиндра, следует использовать потенциалы A_1 и F_1 (коэффициенты $T_{ns}^{\rm ee}$ и $T_{ns}^{\rm em}$).

Полученное решение (23)–(30) задачи для потенциалов позволяет вычислить компоненты электрического и магнитного полей как в волноводе, так и внутри неоднородности. В качестве примера приведём выражения для азимутальной $E_{2\varphi}$ и продольной (по отношению к оси неоднородности) E_{2z} компонент напряжённости электрического поля в волноводе:

$$E_{2\varphi} = \frac{P^{\rm e}\mu_2}{4}\sum_{n=-\infty}^{+\infty}\sum_{s=-\infty}^{+\infty}\left\{\frac{n\pi s}{k_2Lk_2r}\left[J_s(\tilde\kappa_2k_2r)H_s^{(1)}(\tilde\kappa_2k_2a) + R_{ns}^{\rm ee}H_s^{(1)}(\tilde\kappa_2k_2r)\right] + \right.$$
$$\left. + \frac{\mu_2}{\varepsilon_2}\,i\zeta_0^2\tilde\kappa_2R_{ns}^{\rm em}\dot H_s^{(1)}(\tilde\kappa_2k_2r)\right\}\cos\!\left(n\pi\,\frac{z_0}{L}\right)\sin\!\left(n\pi\,\frac{z}{L}\right)\sin[s\,(\varphi-\varphi_0)], \tag{31}$$

$$E_{2z} = \frac{P^{\rm e}\mu_2}{4}\sum_{n=-\infty}^{+\infty}\sum_{s=-\infty}^{+\infty}\tilde\kappa_2^2\left[J_s(\tilde\kappa_2k_2r)H_s^{(1)}(\tilde\kappa_2k_2a) + R_{ns}^{\rm ee}H_s^{(1)}(\tilde\kappa_2k_2r)\right]\times$$
$$\times\cos\!\left(n\pi\,\frac{z_0}{L}\right)\cos\!\left(n\pi\,\frac{z}{L}\right)\cos[s\,(\varphi-\varphi_0)], \tag{32}$$

где $P^{\rm e} = (I_0 l/L)\zeta_0 k_0$, $\tilde{\kappa}_2 = \sqrt{1 - n^2\pi^2/(k_2 L)^2}$ — безразмерное поперечное волновое число.

В работе [17] проанализированы два предельных случая: длинноволновый, когда длина волны в среде велика по сравнению с радиусом цилиндра (приближение Рэлея): $|k_j b| \ll 1$, и коротковолновый, когда выполняется обратное условие (приближение геометрической оптики) $|k_j b| \gg 1$. В обоих случаях получены простые соотношения для напряжённости электрического поля в волноводе, позволяющие исследовать влияние анизотропной неоднородности на структуру полей.

2. РЕЗУЛЬТАТЫ ЧИСЛЕННЫХ РАСЧЁТОВ

Полученные выражения для азимутальной $E_{2\varphi}$ и продольной E_{2z} компонент напряжённости электрического поля (31), (32) позволяют выполнить численное моделирование дифракции электромагнитных волн на анизотропной цилиндрической неоднородности в литосферном волноводе и в плоскослоистой толще земной породы. Здесь ограничимся анализом азимутальных зависимостей компонент $E_{2\varphi}$ и E_{2z} напряжённости электрического поля, рассеянного на анизотропной цилиндрической неоднородности в плоском слое с идеально проводящими стенками.

На рис. 2, 3 представлены индикатрисы рассеяния азимутальной $E_{2\varphi}$ и продольной E_{2z} компонент поля, нормированные на величины $P^{\rm e} f_{\varphi\,\max}$ и $P^{\rm e} f_{z\,\max}$ соответственно, где $f_{\varphi\,\max} = E_{2\varphi\,\max}/P^{\rm e}$ и $f_{z\,\max} = E_{2z\,\max}/P^{\rm e}$ — безразмерные максимальные значения соответствующих компонент поля. Заметим, что далее на рисунках будут приведены такие же зависимости. Расчёты выполнены для следующих параметров задачи: $k_0 b = 0{,}5$ (рис. 2); $k_0 b = 3$ (рис. 3); $k_0 L = 3$; $k_0 a = 4$; $k_0 r = 4$; $z/L = 0{,}5$; $z_0/L = 0{,}3$; $\varphi_0 = 0$; $\varepsilon_2 = 1$; $\mu_1 = \mu_2 = 1$. На рисунках приведены индикатрисы рассеяния для анизотропной неоднородности, показанные сплошной линией ($\varepsilon_{1\perp} = 3, \varepsilon_{1\parallel} = 9$), а также для изотропной неоднородности, показанные штриховыми ($\varepsilon_{1\perp} = \varepsilon_{1\parallel} = 3$) и штрих-пунктирными ($\varepsilon_{1\perp} = \varepsilon_{1\parallel} = 9$) линиями. При выбранных параметрах задачи нормированные коэффициенты составили $f_{\varphi\,\max} = 2{,}37 \cdot 10^{-7}$, $f_{z\,\max} = 0{,}13$.

Из анализа представленных зависимостей видно, что форма индикатрисы рассеяния азимутальной компоненты поля для анизотропной неоднородности определяется в основном поперечной составляющей тензора диэлектрической проницаемости $\varepsilon_{1\perp} = 3$ (см. сплошную и штриховую линии на рис. 2). В отличие от этого, форма индикатрисы рассеяния компоненты E_{2z} в случае анизотропной неоднородности зависит в основном от продольной компоненты тензора диэлек-

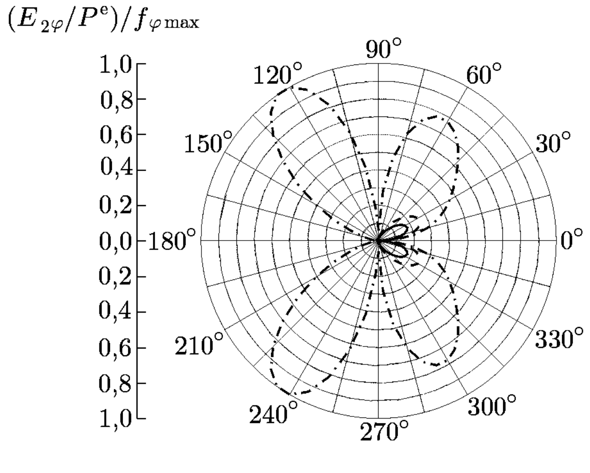


Рис. 2

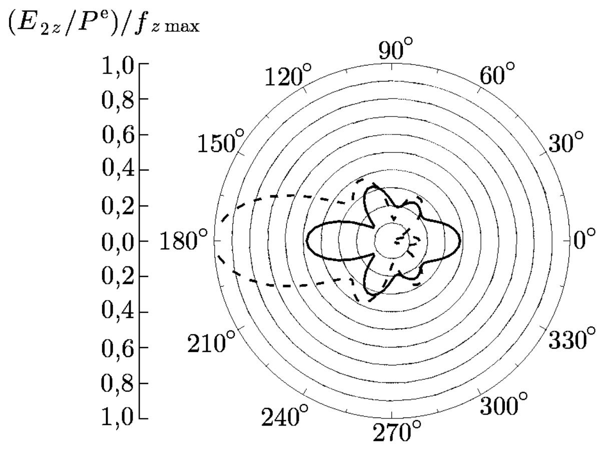


Рис. 3

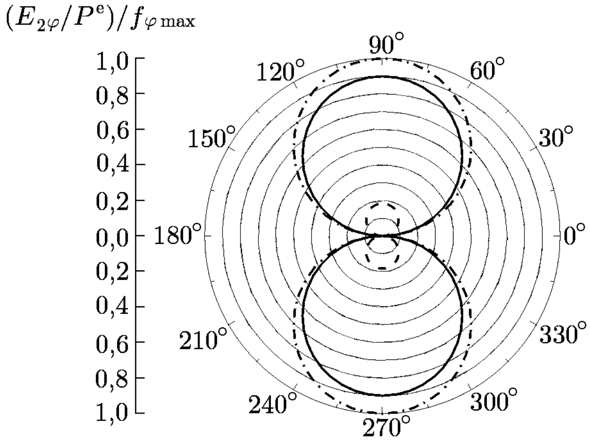


Рис. 4

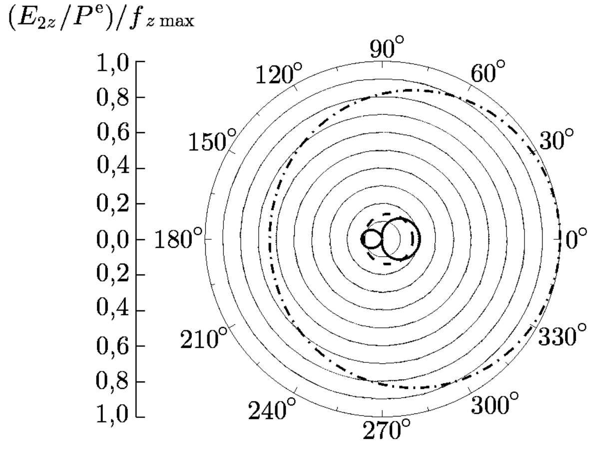


Рис. 5

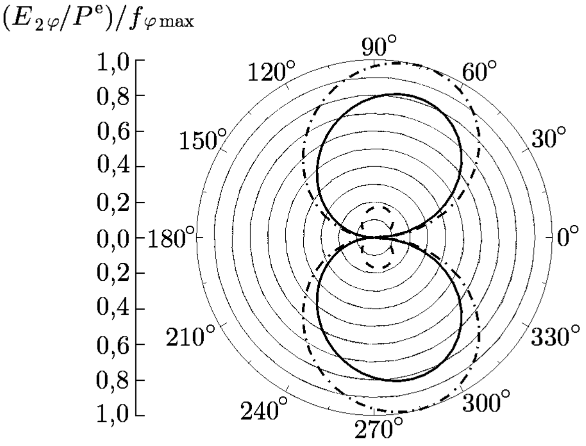


Рис. 6

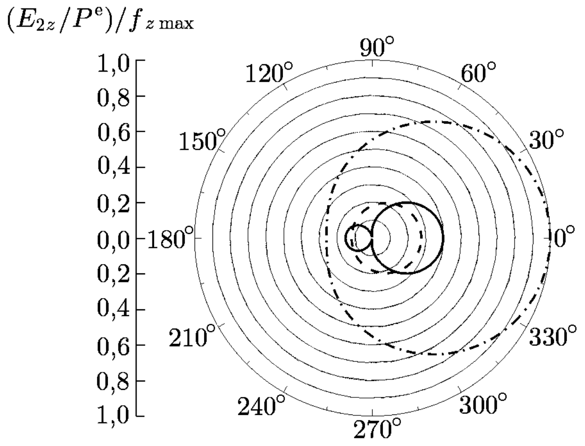


Рис. 7

трической проницаемости $\varepsilon_{1\|} = 9$: на рис. 3 зависимости для анизотропной неоднородности и изотропной неоднородности с диэлектрической проницаемостью, равной 9, полностью совпадают. Кроме того, амплитуда продольной компоненты рассеянного поля в волноводе значительно превышает амплитуду азимутальной составляющей. Указанные особенности поведения индикатрис рассеяния азимутальной и продольной компонент напряженности электрического поля определяются анизотропией неоднородности.

Известно, что для слоистых пород — осадочных и метаморфических — наблюдается анизотропия электрического сопротивления: сопротивление пород поперёк плоскости слоёв всегда больше сопротивления вдоль плоскости слоёв [18]. У горных пород существует анизотропия диэлектрической проницаемости, причём вдоль плоскости слоёв проницаемость больше, чем поперёк. Кроме того, для влагонасыщенных пород анизотропия выражена более резко, чем для сухих [18].

Учитывая эти особенности реальных осадочных и горных пород, на основе выражений (31), (32) выполнены численные расчёты азимутальных зависимостей компонент $E_{2\varphi}$ и E_{2z} напряжённости рассеянного электрического поля на частоте 10 кГц внутри плоского слоя толщиной 30 м, состоящего из слабопроводящих изотропных доломитов, с идеально проводящими границами.

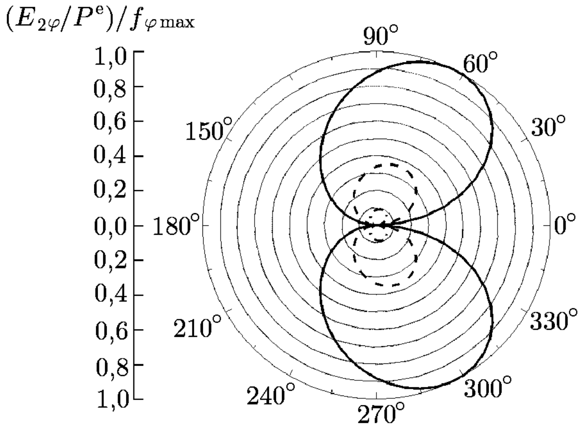


Рис. 8

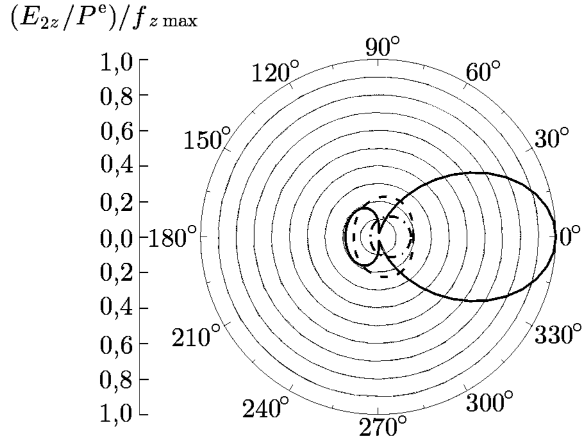


Рис. 9

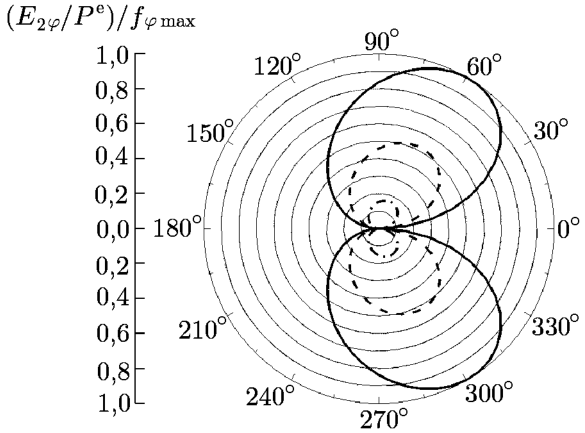


Рис. 10

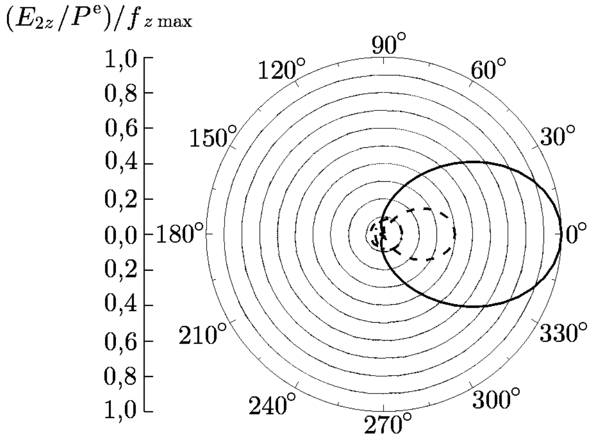


Рис. 11

Относительная комплексная диэлектрическая проницаемость доломитов, насыщенных пресными водами, на частоте 10 кГц равна $\varepsilon_2 = 11{,}9 + 36 \cdot 10^2 i$. Здесь моделируется ситуация, когда слой из слабопроводящих пород сверху и снизу ограничен осадочными породами с высокой удельной проводимостью. Напомним, что возможность проведения таких расчётов была обоснована в работе [9]. В качестве примера анизотропной неоднородности выбраны глинистые сланцы, относительная диэлектрическая проницаемость и удельная проводимость которых в зависимости от влагонасыщенности и солёности могут изменяться в широких пределах [18]: тензор относительной диэлектрической проницаемости сухих глинистых сланцев имеет компоненты — $\varepsilon_{1\perp} = 10{,}3 + 9{,}18 \cdot 10^2 i$ и $\varepsilon_{1\|} = 11{,}2 + 18 \cdot 10^2 i$, влажных глинистых сланцев — $\varepsilon_{1\perp} = 12{,}2 + 46 \cdot 10^2 i$ и $\varepsilon_{1\|} = 14{,}1 + 9 \cdot 10^3 i$, глинистых сланцев, насыщенных солёными водами и рассолами — $\varepsilon_{1\perp} = 18{,}4 + 18 \cdot 10^3 i$ и $\varepsilon_{1\|} = 20{,}6 + 69 \cdot 10^3 i$.

Влияние электрических размеров анизотропной цилиндрической неоднородности иллюстрируют рис. 4–7, где представлены азимутальные зависимости компонент $E_{2\varphi}$ (рис. 4 и 6) и E_{2z} (рис. 5 и 7) поля для трёх указанных выше моделей анизотропной неоднородности. Сплошной линией обозначены кривые для сухих, штриховой — для влажных, а штрих-пунктирной — для солёных глинистых сланцев. Графики зависимостей, представленные на рис. 4 и 5, построены для $b = 0{,}1$ м, а на рис. 6 и 7 — для $b = 3$ м. При этом считалось, что $a = 8$ м и $r = 20$ м. Нормировочные коэффициенты имеют следующие значения: $f_{\varphi\,\max} = 3{,}13 \cdot 10^{-2}$ (рис. 4), $f_{z\,\max} = 9{,}06 \cdot 10^{-5}$ (рис. 5), $f_{\varphi\,\max} = 32{,}25$ (рис. 6), $f_{z\,\max} = 5{,}16 \cdot 10^{-2}$ (рис. 7). Анализ полученных зависимостей показывает, что амплитуда поля максимальна при рассеянии на неоднородности с более высокой удельной проводимостью или диэлектрической проницаемостью (см. штрих-пунктирные линии). Такое поведение азимутальных диаграмм связано, по существу, с увеличением электрических размеров неоднородности. Из рисунков также видно, что для влажных сланцев (штриховые линии) максимальная амплитуда рассеянного поля мала по сравнению с другими рассматриваемыми случаями, хотя удельная проводимость влажных сланцев значительно превышает удельную проводимость сухих сланцев. Это обусловлено малым контрастом сред в волноводе (влажные доломиты) и внутри неоднородности (влажные сланцы).

На рис. 8–11 представлены аналогичные предыдущим азимутальные зависимости напряжённости электрического поля, рассчитанные для солёных сланцев при радиусе неоднородности $b = 3$ м. Рисунки 8, 9 построены для трёх расстояний до точки наблюдения (сплошная линия соответствует $r = 7$ м, штриховая — $r = 10$ м, штрих-пунктирная — $r = 20$ м) при $a = 8$ м. Зависимости, изображённые на рис. 10 и 11, рассчитаны для трёх положений источника (сплошная линия соответствует $a = 6$ м, штриховая — $a = 8$ м, штрих-пунктирная — $a = 12$ м) при $r = 7$ м. Нормировочные коэффициенты в этом случае оказались следующими: $f_{\varphi\,\max} = 1\,302{,}3$ (рис. 8), $f_{z\,\max} = 0{,}87$ (рис. 9), $f_{\varphi\,\max} = 2\,499{,}9$ (рис. 10), $f_{z\,\max} = 4{,}37$ (рис. 11). Анализируя представленные зависимости, можно сделать вывод о том, что форма азимутальных диаграмм компоненты $E_{2\varphi}$ очень слабо зависит от вариаций как электрических, так и геометрических параметров задачи (см. рис. 4, 6, 8, 10). Продольная составляющая E_{2z} поля более чувствительна к подобного рода изменениям. Как видно из рис. 9 и 11, форма азимутальных зависимостей компоненты E_{2z} поля зависит от положения источника и точки наблюдения. Такое поведение азимутальных диаграмм, скорее всего, связано с моделью анизотропии цилиндрической неоднородности.

3. ЗАКЛЮЧЕНИЕ

В работе представлено строгое аналитическое решение задачи дифракции электромагнитных волн на анизотропной цилиндрической неоднородности, расположенной в плоском волноводе с идеально проводящими стенками. Проведённый анализ численных расчётов азимутальных зависимостей компонент E_φ и E_z напряжённости рассеянного электрического поля показал: 1) характер азимутальной зависимости компоненты E_φ поля в основном определяется поперечной составляющей тензора диэлектрической проницаемости анизотропной среды внутри неоднородности, а компоненты E_z — продольной составляющей тензора; 2) в отличие от компоненты E_φ поля, форма азимутальных зависимостей продольной компоненты поля сильно зависит от положения точек передачи и приёма сигнала.

Результаты работы могут быть использованы при моделировании электромагнитных процессов в волноводе Земля—ионосфера, литосферном волноводе и других направляющих структурах при наличии анизотропной цилиндрической неоднородности.

Авторы признательны В. П. Докучаеву за постоянный интерес к работе и В. В. Кириллову за полезные обсуждения результатов.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 05–02–16493).

СПИСОК ЛИТЕРАТУРЫ

1. Wait J. R. Electromagnetic waves scattering media. Oxford: Pergamon Press, 1962. 372 p.
2. Альперт Я. Л., Гусева Э. Г., Флигель Д. С. Распространение электромагнитных низкочастотных волн в волноводе Земля—ионосфера. М.: Наука, 1967. 124 с.
3. Wait J. R. // IEEE Trans. Antennas Propagat. 1991. V. 39, No. 7. P. 1 051.
4. Заборонкова Т. М., Кондратьев И. Г. // Изв. вузов. Радиофизика. 1977. Т. 20, № 12. С. 1 895.
5. Коган Л. П. // Изв. вузов. Радиофизика. 1997. Т. 40, № 4. С. 457.
6. Макаров Г. И., Новиков В. В. // Проблемы дифракции и распространения волн. Л.: ЛГУ, 1968. Вып. 7. С. 19.
7. Wait J. R. // Radio Sci. 1966. V. 1, No. 8. P. 913.
8. Шорохова Е. А. // Новые промышленные технологии. 2000. Вып. 5 (298). С. 64.
9. Шорохова Е. А., Яшнов В. А. // Геофизика. 2001. № 5. С. 57.
10. Левин Л. Теория волноводов. М.: Радио и связь, 1981. 311 с.
11. Швингер Ю. Неоднородности в волноводе. М.: Сов. радио, 1970. 110 с.
12. Рязанцев А. М. Теоретические и экспериментальные исследования по проблемам радиосвязи в шахтах, туннелях и других подземных образованиях. М.: Экос, 1982.
13. Ванзин П. А., Шорохова Е. А. // Геофизика. 1999. № 1. С. 45.
14. Котик Д. С., Поляков С. В., Яшнов В. А. // Изв. вузов. Радиофизика. 1978. Т. 21, № 7. С. 938.
15. Докучаев В. П., Яшнов В. А. // Распространение и дифракция электромагнитных волн. М.: МФТИ, 1993. С. 58.
16. Докучаев В. П., Можжухин С. Б. // Изв. вузов. Радиофизика. 1994. Т. 37, № 8. С. 1 032.
17. Docuchaev V. P., Shorochova E. A., Yashnov V. A. // J. Appl. Electromag. 2000. V. 3, No. 2. P. 1.
18. Дортман Н. Б. Физические свойства горных пород и полезных ископаемых. Справочник геофизика. М.: Недра, 1984. 455 с.

Поступила в редакцию 26 августа 2003 г.; принята в печать 28 марта 2005 г.

DIFFRACTION OF ELECTROMAGNETIC WAVES BY AN ANISOTROPIC CYLINDRICAL INHOMOGENEITY IN A PLANAR WAVEGUIDE

E. A. Shorokhova and V. A. Yashnov

We consider diffraction of electromagnetic waves by an anisotropic cylindrical inhomogeneity located in a planar waveguide with perfectly conducting walls. The anisotropy is allowed for with the use of the uniaxial-crystal approximation. A rigorous analytical solution is presented in the form of double sums over eigenfunctions of a planar waveguide with perfectly conducting walls and azimuthal eigenfunctions of a cylinder. Different components of the intensity of the electric field scattered by an anisotropic inhomogeneity are numerically calculated. Effect of the anisotropy and inhomogeneity sizes on the scattered field is analyzed.

УДК 621.396.96.01

ОЦЕНИВАНИЕ ПЛОТНОСТЕЙ ВЕРОЯТНОСТИ МОДУЛЯ И АРГУМЕНТА КОМПЛЕКСНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

М. А. Островский, М. В. Уханов

Нижегородский государственный технический университет, г. Нижний Новгород, Россия

Критерии оценивания плотностей вероятности N-мерного случайного вектора, предложенные в работе одного из авторов, конкретизируются применительно к случаю чётного N. При этом для удобства изложения двумерные векторы трактуются как комплексные случайные величины, и оценке подвергаются плотности вероятности модулей и аргументов последних. На основе найденных критериев синтезируются алгоритмы оценки в условиях полной априорной информации о статистике выборки и при отсутствии такой информации. На ряде примеров доказывается оптимальность найденных алгоритмов в среднестатистическом смысле.

ВВЕДЕНИЕ

При проектировании радиолокационных и радионавигационных комплексов, а также систем радиотехнической разведки и радиосвязи возникают схожие задачи, состоящие в необходимости перестройки структуры и параметров приёмных устройств в зависимости от изменения статистических характеристик помех и полезных сигналов. В частности, подобная перестройка требуется при отличии указанных законов распределения от гауссового. Этой проблеме посвящены публикации [1–7], в том числе и работы одного из авторов [5–7]. В [5, 6] показано, что при воздействии асимптотически слабых сигналов структура нелинейного приёмника разбивается на линейный пространственно-временно́й обеляющий фильтр и безынерционный нелинейный преобразователь, характеристика которого целиком определяется видом одномерной плотности вероятности помехи. В случае воздействия сильного флуктуирующего сигнала характеристика преобразователя, как и характеристика детектора огибающей, зависит от законов распределения амплитудных и фазовых пульсаций сигнала. Поскольку, как показано в [5], на выходах многомерного обеляющего фильтра устраняется статистическая связь между отсчётами помехи и сигнала, совокупность последних может интерпретироваться как случайная величина.

Одним из возможных методов перестройки нелинейных приёмников является оценка плотностей вероятности принятой реализации либо плотностей вероятности её амплитуд и фаз и использование оценок для управления приёмником. В связи с этим большое практическое и научное значение приобретает решаемая ниже задача оптимального оценивания законов распределения модуля и аргумента комплексных случайных величин.

Аналогичная задача возникает также при создании систем картографирования и классификации типов действующих на радиолокаторы помех, а также при формировании радиолокационных портретов целей, поскольку оцениваемые законы распределения содержат достаточную информацию о пространственной конфигурации и интенсивности расположенных на поверхности радиолокационных целей доминирующих «блестящих точек».

1. ФОРМУЛИРОВКА ЗАДАЧИ

В [8] рассмотрен метод нелинейного оценивания плотности вероятности N-мерной выборки $\mathbf{x} = \|x_1, \ldots, x_N\|$, образованной последовательностью независимых испытаний комплексной случайной величины ξ. Полагается, что выборка характеризуется плотностью вероятности $w(x_\nu)$, где $\nu = 1, 2, \ldots, N$, неопределённость которой проявляется в зависимости отсчёта x_ν от априорно неизвестного n-мерного вектора параметров $\|\lambda_1, \ldots, \lambda_n\|^{\mathrm{T}} = \mathbf{\Lambda}$, где индекс T обозначает транспонирование. Сущность метода состоит в определении плотности вероятности по вектору состояния $\hat{\mathbf{\Lambda}} = \|\hat{\lambda}_1, \ldots, \hat{\lambda}_n\|^{\mathrm{T}}$ обучающегося измерителя с зависящей от $\mathbf{\Lambda}$ и $\hat{\mathbf{\Lambda}}$ функцией преобразования $y_\nu = \hat{F}_\nu[x_\nu(\mathbf{\Lambda}), \hat{\mathbf{\Lambda}}]$. Обучение достигается максимально возможным приближением оцениваемой функции $\hat{F}_\nu[x_\nu(\mathbf{\Lambda}), \hat{\mathbf{\Lambda}}]$ к заданной $F_\nu[x_\nu(\mathbf{\Lambda})]$. В [8] показано, что для минимизации информационных потерь при преобразовании желаемая функция не произвольно зависит от оцениваемой плотности вероятности. Она должна удовлетворять условию

$$F_\nu[x_\nu(\mathbf{\Lambda})] = -\frac{\partial}{\partial x_\nu(\mathbf{\Lambda})} \ln w[x_\nu(\mathbf{\Lambda})], \tag{1}$$

т. е. являться ν-м элементом достаточной статистики при неограниченном стремлении объёма выборки N к бесконечности. При этом оценка плотности вероятности комплексной случайной величины однозначно зависит от функции преобразования измерителя:

$$\hat{w}[x_\nu(\mathbf{\Lambda})] = w(0) \exp\Big\{-\int \hat{F}_\nu[x_\nu(\mathbf{\Lambda}), \hat{\mathbf{\Lambda}}]\, \mathrm{d}x_\nu(\mathbf{\Lambda})\Big\}. \tag{2}$$

Предложенный в [8] метод синтеза основан на сочетании информационного подхода с критерием «минимального расстояния» между оцениваемой и желаемой вектор-функциями. В этом состоит его основное отличие от известных методов [9]. Преимущества предложенного метода в точности и скорости сходимости исследованы и подтверждены экспериментально в работе [8].

Наряду с оцениванием плотности вероятности комплексных случайных величин $x_\nu = x_{\mathrm{c}\nu} + jx_{\mathrm{s}\nu} = \rho_\nu \exp(j\varphi_\nu)$, во многих случаях требуется определение плотностей вероятности их модуля $P(\rho_\nu)$ и аргумента $Q(\varphi_\nu)$. С принципиальной точки зрения возможно использование для этой цели результатов нелинейного оценивания плотности вероятности [8]. Для этого необходимо перейти в (2) от декартовых к полярным координатам ρ_ν, φ_ν, записать в явном виде оценку совместной плотности вероятности модуля и аргумента $\hat{W}_2(\rho_\nu, \varphi_\nu)$ и осуществить интегрирование по переменной φ_ν, если отыскивается функция $\hat{P}(\rho_\nu)$, или по переменной ρ_ν, если определяется функция $\hat{Q}(\varphi_\nu)$. Например, такой метод используется при гауссовом распределении комплексной величины [10]. Однако для произвольного закона $w(x_\nu)$ интегрирование совместной плотности вероятности неосуществимо. В связи с этим целесообразно, минуя этап вычисления плотности вероятности комплексного числа, использовать предложенный в [8] метод для непосредственного оценивания распределений модуля и аргумента.

Переходя как в самом комплексном отсчёте x_ν, так и в дифференциальном операторе к полярным координатам:

$$\frac{\partial}{\partial x_\nu} = \frac{\partial \rho_\nu}{\partial x_\nu}\, \frac{\partial}{\partial \rho_\nu} + \frac{\partial \varphi_\nu}{\partial x_\nu}\, \frac{\partial}{\partial \varphi_\nu}\,,$$

преобразуем зависимость (1) к виду

$$F_\nu(\rho_\nu, \varphi_\nu) = x_\nu^*[A(\rho_\nu) + jB(\varphi_\nu)], \tag{3}$$

где

$$A(\rho_\nu) = -\frac{1}{\rho_\nu}\, \frac{\partial}{\partial \rho_\nu} \ln\left[\frac{P(\rho_\nu)}{\rho_\nu}\right], \qquad B(\varphi_\nu) = \frac{1}{\rho_\nu^2}\, \frac{\partial}{\partial \varphi_\nu} \ln Q(\varphi_\nu)$$

— желаемые функции измерения, а индекс звёздочка обозначает комплексное сопряжение.

Пусть на вход оценивающего устройства действует $2N$-мерная выборка $\mathbf{x} = \|\rho_1, \varphi_1, \ldots, \rho_N, \varphi_N\|$. Выборка одновременно характеризуется плотностью вероятности модуля $P(\rho_\nu, \mathbf{C})$ и плотностью вероятности аргумента $Q(\varphi_\nu, \mathbf{G})$, где параметры $\mathbf{C} = \|c_1, \ldots, c_n\|^{\mathrm{T}}$ и $\mathbf{G} = \|g_1, \ldots, g_m\|^{\mathrm{T}}$ априорно неизвестны. Выборка является обучающей для устройства оценивания, которое характеризуется двумя функциями измерения: $\hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}})$ и $\hat{B}_\nu(\varphi_\nu, \hat{\mathbf{G}})$.

Целью обучения является стремление $\hat{A}_\nu$ и $\hat{B}_\nu$ к желаемым функциям $A(\rho_\nu)$ и $B(\varphi_\nu)$ соответственно, зависящим от оцениваемых плотностей вероятности модуля и аргумента. В результате обучения на выходах измерителя в соответствии с (3) формируются оценки плотностей вероятности

$$\hat{P}(\rho_\nu) = K_\rho \rho_\nu \exp\left[-\int \rho_\nu \hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}})\,\mathrm{d}\rho_\nu\right], \qquad \hat{Q}(\varphi_\nu) = K_\varphi \exp\left[\int \rho_\nu^2 \hat{B}_\nu(\varphi_\nu, \hat{\mathbf{G}})\,\mathrm{d}\varphi_\nu\right], \tag{4}$$

где коэффициенты пропорциональности K_ρ и K_φ находятся из условия единичной нормировки интегралов от этих оценок.

Как следует из (4), вычисление плотностей вероятности модуля и аргумента предполагает оценку оптимальных функций измерения $\hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}})$ и $\hat{B}_\nu(\varphi_\nu, \hat{\mathbf{G}})$. Для подобной оценки необходимо сформулировать критерии качества и синтезировать оптимальные алгоритмы оценивания как при наличии, так и при отсутствии априорной информации о статистике выборки.

2. КРИТЕРИИ КАЧЕСТВА ОПТИМАЛЬНОГО ОЦЕНИВАНИЯ

Если законы $P(\rho_\nu)$ и $Q(\varphi_\nu)$ известны, оптимальное обучение измерителя сводится к наилучшей аппроксимации желаемых функций $A_\nu(\rho_\nu, \mathbf{C})$, $B_\nu(\varphi_\nu, \mathbf{G})$ их оценками $\hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}})$ и $\hat{B}_\nu(\varphi_\nu, \hat{\mathbf{G}})$. Так же, как и в [8], в качестве критерия аппроксимации используем усиленный критерий минимума среднего квадрата разности этих функций:

$$\int\limits_0^\infty \rho_\nu^{N_1}\,[\hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}}) - A_\nu(\rho_\nu, \mathbf{C})]^2\, P(\rho_\nu)\,\mathrm{d}\rho_\nu = \min,$$

$$\int\limits_{-\pi}^{\pi} \rho_\nu^{N_1}\,[\hat{B}_\nu(\varphi_\nu, \hat{\mathbf{G}}) - B_\nu(\varphi_\nu, \mathbf{G})]^2\, Q(\varphi_\nu)\,\mathrm{d}\varphi_\nu = \min. \tag{5}$$

Умножение подынтегральных функций на $\rho_\nu^{N_1}$ (усиление критерия) вызвано необходимостью сохранения их непрерывности при стремлении ρ_ν к нулю и к бесконечности, где N_1 — максимальный отрицательный показатель степени аппроксимирующего полинома [8]. Минимизация (5) достигается соответствующими вариациями вектор-параметров $\hat{\mathbf{C}}$ и $\hat{\mathbf{G}}$, при которых оценочные функции наилучшим (в среднеквадратическом смысле) образом соответствуют оптимальным.

Поскольку оптимальные функции $A_\nu(\rho_\nu, \mathbf{C})$ и $B_\nu(\varphi_\nu, \mathbf{G})$ не зависят от оценок параметров, систему (5) можно упростить как для амплитудного, так и для фазового функционалов:

$$\int\limits_0^\infty \rho_\nu^{N_1}\,[\hat{A}_\nu^2(\rho_\nu, \hat{\mathbf{C}}) - 2\hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}}) A_\nu(\rho_\nu, \mathbf{C}]\, P(\rho_\nu)\,\mathrm{d}\rho_\nu = \min, \tag{6}$$

$$\int\limits_{-\pi}^{\pi} \rho_\nu^{N_1}\,[\hat{B}_\nu^2(\varphi_\nu, \hat{\mathbf{G}}) - 2\hat{B}_\nu(\varphi_\nu, \hat{\mathbf{G}}) B_\nu(\varphi_\nu, \mathbf{G})]\, Q(\varphi_\nu)\,\mathrm{d}\varphi_\nu = \min. \tag{7}$$

Введём множество L однозначных, непрерывных и непрерывно дифференцируемых функций $\hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}})$, удовлетворяющих граничным условиям

$$\lim_{\rho \to 0} \rho_\nu^{N_1 - 1} \hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}}) P(\rho_\nu) = \lim_{\rho \to \infty} \rho_\nu^{N_1 - 1} \hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}}) P(\rho_\nu) = 0. \tag{8}$$

Несложно доказать выполнение условия (8) для большинства известных распределений модуля и аппроксимирующих функций $\hat{A}$, заданных в форме степенных полиномов [11] с максимальным отрицательным показателем $N_1 = 2$:

$$\hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}}) = \sum_{i=-2}^{n-3} \hat{c}_i(\nu)\rho_\nu^i. \tag{9}$$

Выполнение условия (8) для указанных функций очевидно, поскольку спадание начальных и периферийных областей плотности вероятности происходит быстрее, чем рост любой конечной степени модуля.

Интегрируя по частям второе слагаемое в подынтегральном выражении в (6) и используя граничные условия (8), получим окончательный критерий качества аппроксимации:

$$\int\limits_0^\infty \left\{ \rho_\nu^2 \hat{A}_\nu^2(\rho_\nu, \hat{\mathbf{C}}) - \frac{2}{\rho_\nu}\,\frac{\partial}{\partial \rho_\nu}\left[\rho_\nu^2 \hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}})\right] \right\} P(\rho_\nu)\,\mathrm{d}\rho_\nu = \min. \tag{10}$$

Аналогично можно было бы из (7) найти функционал качества оценивания плотности вероятности аргумента. Однако использованный в [11] аппроксимирующий ряд

$$\hat{B}_\nu(\varphi_\nu, \hat{\mathbf{G}}) = \sum_{\mu=1}^{m} \hat{g}_\mu(\nu)\varphi_\nu^\mu \tag{11}$$

не всегда удовлетворяет граничным условиям, аналогичным (8):

$$\lim_{\varphi\to\pi} \hat{B}_\nu(\varphi_\nu, \hat{\mathbf{G}})Q(\varphi_\nu) = \lim_{\varphi\to-\pi} \hat{B}_\nu(\varphi_\nu, \hat{\mathbf{G}})Q(\varphi_\nu) = 0. \tag{12}$$

Например, условие (12) не выполняется при равенстве нулю среднего значения выборочной величины $\langle x_\nu\rangle = \rho_0 \exp(j\varphi_0)$ и равномерном распределении аргумента: $Q(\varphi_\nu) = 1/(2\pi)$, $\varphi_\nu \in [-\pi, \pi]$. В этом случае функционал (7) принимает вид

$$\int\limits_{-\pi}^{\pi} \left[\frac{\rho_\nu^2}{2}\,\hat{B}_\nu^2(\varphi_\nu, \hat{\mathbf{G}}) + \frac{\partial}{\partial\varphi_\nu}\,\hat{B}_\nu(\varphi_\nu, \hat{\mathbf{G}})\right] Q(\varphi_\nu)\,\mathrm{d}\varphi_\nu - 2\hat{B}_\nu(\pi, \hat{\mathbf{G}})Q(\pi) = \min. \tag{13}$$

Во всех иных случаях, когда величина ρ_0 отлична от нуля и достаточно велика для выполнения граничных условий (12), функционал (7) имеет более компактный вид:

$$\int\limits_{-\pi}^{\pi} \left[\frac{\rho_\nu^2}{2}\,\hat{B}_\nu^2(\varphi_\nu, \hat{\mathbf{G}}) + \frac{\partial}{\partial\varphi_\nu}\,\hat{B}_\nu(\varphi_\nu, \hat{\mathbf{G}})\right] Q(\varphi_\nu)\,\mathrm{d}\varphi_\nu = \min. \tag{14}$$

3. СИНТЕЗ ФУНКЦИЙ ИЗМЕРЕНИЯ ПРИ АПРИОРНО ИЗВЕСТНОЙ СТАТИСТИКЕ ВЫБОРКИ

Начнём синтез с рассмотрения функционала (10). Поскольку ниже плотность вероятности модуля полагается известной, минимизация функционала приводит к зависимости параметров полинома $\hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}})$ от усреднённых по ансамблю моментов $\langle \rho_\nu^k \rangle = \langle \rho^k \rangle$. Подставляя в (10) выражение (9), вычисляя градиент функционала по каждой из оценок параметра c_k и приравнивая его к нулю, получим систему уравнений

$$\left\langle \rho_\nu^2 \hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}}) \, \frac{\partial \hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}})}{\partial \hat{c}_k} \right\rangle - \left\langle \frac{1}{\rho_\nu} \, \frac{\partial}{\partial \rho_\nu} \left[\rho_\nu^2 \, \frac{\partial \hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}})}{\partial \hat{c}_k} \right] \right\rangle = \sum_{i=-2}^{n-3} \hat{c}_i \, \langle \rho^{i+k+2} \rangle - (2+k) \, \langle \rho^k \rangle = 0, \quad (15)$$

где $k = -2, \ldots, n-3$.

Докажем оптимальность системы (15) и возможность её использования для синтеза функции измерения модуля $\hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}})$ на ряде конкретных примеров.

1. В случае релеевской плотности вероятности $P(\rho_\nu) = \rho_\nu \sigma_x^{-2} \exp[-\rho_\nu^2/(2\sigma_x^2)]$ с начальными моментами $\langle \rho^k \rangle = (2\sigma_x^2)^{k/2} \Gamma(k/2+1)$ принадлежность плотности вероятности множеству L обусловлена выполнением условия (8): $\sigma_x^{-2} \sum_{i=-1}^{n-3} \hat{c}_i \lim\limits_{\rho \to 0,\infty} \rho_\nu^{i+2} \exp(-\rho_\nu^2/2\sigma_x^2) = 0$ при $i > -2$. Отсюда функция измерения плотности вероятности модуля выражается полиномом $\hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}}) = \hat{c}_{-1} \rho_n^{-1} + \hat{c}_0 + \hat{c}_1 \rho_\nu$, а система оптимизационных уравнений (15), представленная в векторно-матричной форме, принимает вид

$$\left\| \begin{matrix} \hat{c}_{-1} \\ \hat{c}_0 \\ \hat{c}_1 \end{matrix} \right\| \left\| \begin{matrix} \langle \rho \rangle & \langle \rho^2 \rangle & \langle \rho^3 \rangle \\ \langle \rho^2 \rangle & \langle \rho^3 \rangle & \langle \rho^4 \rangle \\ 1 & \langle \rho \rangle & \langle \rho^2 \rangle \end{matrix} \right\| = \left\| \begin{matrix} 2 \\ 3 \, \langle \rho \rangle \\ \langle \rho^{-1} \rangle \end{matrix} \right\|. \quad (16)$$

Учитывая, что входящие в уравнение оптимизации (16) элементы моментной матрицы равны $\langle \rho^{-1} \rangle = 1{,}2533 \sigma_x^{-1}$, $\langle \rho \rangle = 1{,}2533 \sigma_x$, $\langle \rho^2 \rangle = 2\sigma_x^2$, $\langle \rho^3 \rangle = 3{,}7598 \sigma_x^3$, $\langle \rho^4 \rangle = 8\sigma_x^4$, найдём его решение $\hat{\mathbf{C}} = \|0, \sigma_x^{-2}, 0\|^{\mathrm{T}}$ и соответствующую функцию измерения плотности вероятности модуля

$$\hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}}) = \sigma_x^{-2}.$$

Подставляя эту функцию в (4) и находя нормировочный множитель $K_\rho = \sigma_x^{-2}$, получим выражение для оптимальной оценки релеевского закона распределения:

$$\hat{P}(\rho_\nu) = \frac{\rho_\nu}{\sigma_x^2} \exp\left(-\frac{1}{\sigma_x^2} \int \rho_\nu \, \mathrm{d}\rho_\nu \right) = \frac{\rho_\nu}{\sigma_x^2} \exp\left(-\frac{\rho_\nu^2}{2\sigma_x^2} \right), \quad (17)$$

не отличающейся от оцениваемого закона распределения выборки.

2. Рассмотрим случай экспоненциального распределения плотности вероятности модуля $P(\rho_\nu) = \lambda^2 \rho_\nu \exp(-\lambda \rho_\nu)$ с начальными моментами $\langle \rho^{-1} \rangle = \lambda$, $\langle \rho \rangle = 2\lambda^{-1}$, $\langle \rho^2 \rangle = 6\lambda^{-2}$, $\langle \rho^3 \rangle = 24\lambda^{-3}$, $\langle \rho^4 \rangle = 120\lambda^{-4}$. Граничные условия (8) также выполняются при $i > -2$, поэтому вид функции измерения и системы оптимизационных уравнений в этом случае остаётся таким же, как и для первого примера. Подставляя в (16) значения моментов, найдём решение этого векторного уравнения $\hat{\mathbf{C}} = \|\lambda, 0, 0\|^{\mathrm{T}}$ и соответствующую функцию измерения $\hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}}) = \lambda \rho_\nu^{-1}$. Находя далее коэффициент пропорциональности $K_\rho = \lambda^2$ и подставляя функцию измерения в (4), получаем выражение для оценки плотности вероятности:

$$\hat{P}(\rho_\nu) = \lambda^2 \rho_\nu \exp(-\lambda \rho_\nu). \quad (18)$$

3. Пусть плотность вероятности модуля подчиняется закону Накагами

$$P(\rho_\nu) = \frac{2\rho_\nu^{2\vartheta-1}}{\Gamma(\vartheta)} \left(\frac{\vartheta}{\langle \rho^2 \rangle} \right)^{\vartheta} \exp\left(-\frac{\vartheta \rho_\nu^2}{\langle \rho^2 \rangle} \right)$$

с начальными моментами

$$\langle \rho^k \rangle = \frac{\Gamma(\vartheta + k/2)}{\Gamma(\vartheta)} \left(\frac{\langle \rho^2 \rangle}{\vartheta} \right)^{k/2}$$

и единственным параметром $\vartheta \geq 1/2$, где $\Gamma(\vartheta)$ — гамма-функция. Граничные условия (8) для $i_{\min} = -N_1 = -2$ оказываются нулевыми:

$$\frac{2}{\Gamma(\vartheta)} \left(\frac{\vartheta}{\langle \rho^2 \rangle} \right)^{\vartheta} \sum_{i=-2}^{n-3} \hat{c}_i \lim_{\rho \to 0, \infty} \rho_\nu^{1+2\vartheta} \exp\!\left(-\frac{\vartheta \rho_\nu^2}{\langle \rho^2 \rangle} \right) = 0,$$

и выполняются при $\vartheta > 1$. Выбирая $\vartheta = 2$, $N_1 = 2$, $n = 4$, $\hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}}) = \hat{c}_{-2} \rho_\nu^{-2} + \hat{c}_{-1} \rho_\nu^{-1} + \hat{c}_0 + \hat{c}_1 \rho_\nu$, получим систему оптимизационных уравнений (15):

$$\left\| \begin{matrix} \hat{c}_{-2} \\ \hat{c}_{-1} \\ \hat{c}_0 \\ \hat{c}_1 \end{matrix} \right\| \left\| \begin{matrix} \langle \rho^{-2} \rangle & \langle \rho^{-1} \rangle & 1 & \langle \rho \rangle \\ \langle \rho^{-1} \rangle & 1 & \langle \rho \rangle & \langle \rho^2 \rangle \\ 1 & \langle \rho \rangle & \langle \rho^2 \rangle & \langle \rho^3 \rangle \\ \langle \rho \rangle & \langle \rho^2 \rangle & \langle \rho^3 \rangle & \langle \rho^4 \rangle \end{matrix} \right\| = \left\| \begin{matrix} 0 \\ \langle \rho^{-1} \rangle \\ 2 \\ 3 \langle \rho \rangle \end{matrix} \right\|. \tag{19}$$

Подставляя в неё значения начальных моментов модуля $\langle \rho^{-2} \rangle = 2/\langle \rho^2 \rangle$, $\langle \rho^{-1} \rangle = 1{,}2533/\sqrt{\langle \rho^2 \rangle}$, $\langle \rho \rangle = 0{,}94 \sqrt{\langle \rho^2 \rangle}$, $\langle \rho^3 \rangle = 1{,}1748 \langle \rho^2 \rangle^{3/2}$, $\langle \rho^4 \rangle = 1{,}5 \langle \rho^2 \rangle^2$ и находя решение (19) $\hat{\mathbf{C}} = \|-2, 0, 4 \langle \rho^2 \rangle^{-1}, 0\|^{\mathrm{T}}$, получаем искомую функцию измерения $\hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}}) = 4/\langle \rho^2 \rangle - 2/\rho_\nu^2$. Отсюда оценка плотности вероятности (4) для рассматриваемого примера принимает вид

$$\hat{P}(\rho_\nu) = K_\rho \rho_\nu^3 \exp\!\left(-2\rho_\nu^2 \langle \rho^2 \rangle^{-1} \right), \tag{20}$$

где коэффициент пропорциональности $K_\rho = 8 \langle \rho^2 \rangle^{-2}$. Как несложно заметить, при $\vartheta = 2$ оценка плотности вероятности в точности совпадает с измеряемым законом распределения.

В [10] указывается, что при $\vartheta > 1$ закон Накагами даёт хорошую аппроксимацию для обобщённой плотности вероятности Релея, или закона Райса:

$$P(\rho_\nu) = \frac{\rho_\nu}{\sigma_x^2} \exp\!\left(-\frac{\rho_\nu^2 + \rho_0^2}{2\sigma_x^2} \right) I_0\!\left(\frac{\rho_\nu \rho_0}{\sigma_x^2} \right), \tag{21}$$

где $\rho_\nu > 0$, когда условие центрированности принимаемой выборки нарушается, т. е. ν-й выборочный элемент имеет вид $x_\nu = \rho_0 + \rho_\nu \exp(j\varphi_\nu)$; здесь $I_0(x)$ — функция Бесселя нулевого порядка. При этом между параметрами ϑ, $\langle \rho^2 \rangle$ и σ_x^2, ρ_0 имеется следующая связь: $\langle \rho^2 \rangle = \sigma_x^2 + \rho_0^2$, $\vartheta = \langle \rho^2 \rangle^2 / (\langle \rho^2 \rangle^2 - \rho_0^4)$, $\rho_0^2 = \langle \rho^2 \rangle / \vartheta$.

Таким образом, в рассмотренных примерах установлено точное совпадение оценки плотности вероятности с заданным. Это доказывает оптимальность предложенного критерия оценки плотности вероятности модуля комплексной случайной величины (10) и пригодность его для синтеза функции измерения.

Перейдём к исследованию критериев качества оценки плотности вероятности аргумента (13) и (14). Целью исследования является оценка возможности их использования для синтеза оптимальных функций измерения $\hat{B}_\nu(\varphi_\nu, \hat{\mathbf{G}})$ и определение условий, ограничивающих применение каждого из этих критериев.

Подставляя (11) в (13), дифференцируя критерий по каждой из оценок $\hat{g}_k$ и приравнивая градиент к нулю, получаем систему оптимизационных уравнений:

$$\rho_\nu \left\langle \hat{B}_\nu(\varphi_\nu, \hat{\mathbf{G}}) \, \frac{\partial \hat{B}_\nu(\varphi_\nu, \hat{\mathbf{G}})}{\partial \hat{g}_k} \right\rangle + \left\langle \frac{\partial^2 \hat{B}_\nu(\varphi_\nu, \hat{\mathbf{G}})}{\partial \varphi_\nu \partial \hat{g}_k} \right\rangle - 2Q(\pi) \, \frac{\partial \hat{B}_\nu(\pi, \hat{\mathbf{G}})}{\partial \hat{g}_k} =$$
$$= \rho_\nu^2 \sum_{\mu=1}^{m} \hat{g}_\mu \left\langle \varphi^{\mu+k} \right\rangle + k \left\langle \varphi^{k-1} \right\rangle - 2\pi^k Q(\pi) = 0, \quad (22)$$

где $k = 1, \ldots, m$. Поскольку функция измерения плотности вероятности аргумента удовлетворяет условию нечётности и обращается в нуль в точке $\varphi_\nu = 0$, индексы μ и k в (22) принимают только нечётные целочисленные значения. Простейшим примером такой функции является двучленный полином $\hat{B}_\nu(\varphi_\nu, \hat{\mathbf{G}}) = \hat{g}_1 \varphi_\nu + \hat{g}_3 \varphi_\nu^3$, эффективность которого исследуем на конкретных примерах.

1. Пусть выборка $\mathbf{x}$ центрирована, а плотность вероятности аргумента характеризуется равномерным распределением $Q(\varphi_\nu) = 1/(2\pi)$, $|\varphi_\nu| \le \pi$. Все нечётные моменты распределения обращаются в нуль, а чётные равны $\langle \varphi^{2i} \rangle = \pi^{2i}/(2i+1)$, т. е. $\langle \varphi^2 \rangle = \pi^2/3$, $\langle \varphi^4 \rangle = \pi^4/5$, $\langle \varphi^6 \rangle = \pi^6/7$. Учитывая, что при равномерном распределении справедливы соотношения $2\pi Q(\pi) = 1$, $2\pi^3 Q(\pi) = \pi^2$, из (22) находим

$$\left\| \begin{matrix} \hat{g}_1 \\ \hat{g}_3 \end{matrix} \right\| \left\| \begin{matrix} \langle \varphi^2 \rangle & \langle \varphi^4 \rangle \\ \langle \varphi^4 \rangle & \langle \varphi^6 \rangle \end{matrix} \right\| = 0. \quad (23)$$

Поскольку уравнение оптимизации (23) однородно, а его определитель отличен от нуля, оба корня этого уравнения, как и функция измерения, тождественно обращаются в нуль. Отсюда оценка плотности вероятности аргумента (4) при условии её единичной интегральной нормировки $K_\varphi \int_{-\pi}^{\pi} \mathrm{d}\varphi_\nu = 1$ принимает вид $\hat{Q}(\varphi_\nu) = 1/(2\pi)$, $|\varphi_\nu| \in [-\pi, \pi]$.

2. Пусть на вход измерителя действует нецентрированная выборка $\mathbf{x}$ с независимыми нормальными квадратурными компонентами, среднее значение которой ρ_0 отлично от нуля. Плотность вероятности результирующего аргумента выборки приведена в [10]:

$$Q(\varphi_\nu) = \frac{1}{2\pi} \, \exp\!\left(-\frac{s^2}{2}\right) \left[1 + \sqrt{2\pi} \; s \cos(\varphi_\nu) \Phi(s \cos \varphi_\nu) \exp\!\left(\frac{s^2}{2} \cos^2 \varphi_\nu\right)\right], \quad (24)$$

и характеризуется чётными моментами

$$\langle \varphi^{2i} \rangle = \frac{\pi^{2i}}{2i+1} + \sum_{r=1}^{\infty} a_r \int\limits_{-\pi}^{\pi} t^{2i} \cos(rt) \, \mathrm{d}t, \quad (25)$$

где

$$a_r = \frac{\Gamma(1 + r/2) s^r}{\pi \, r! \sqrt{2^r}} \; {}_1F_1\!\left(\frac{r}{2}, r+1, -\frac{s^2}{2}\right),$$

$s = \rho_0/\sigma_x$, $\Phi(x)$ — интеграл вероятности, ${}_1F_1(x)$ — вырожденная гипергеометрическая функция.

По мере увеличения параметра s форма плотности вероятности меняется от равномерной зависимости при $s = 0$ до гауссовой при $s \gg 1$. Действительно, при $\rho_0 \gg \sigma_x$ аргумент φ_ν и модуль ρ_ν результирующего комплексного числа стремятся к аргументу $\varphi_0 = 0$ и модулю ρ_0 среднего значения выборки. В исследуемом предельном случае справедливы следующие допущения: $\exp(-s^2/2) \approx 0$, $\cos \varphi_\nu \approx 1$, $\sin \varphi_\nu \approx \varphi_\nu$, $\Phi(s \cos \varphi_\nu) \approx 1$, приводящие (24) к виду

$$Q(\varphi_\nu) = \frac{s}{\sqrt{2\pi}} \, \exp(-s^2 \varphi_\nu^2/2). \quad (26)$$

При равномерном распределении совпадение оптимальной оценки с измеряемой плотностью вероятности было доказано выше. Покажем, что такое же совпадение имеет место и при $s \gg 1$.

Из (26) следует, что плотность вероятности аргумента быстро спадает в периферийных областях, причём это убывание происходит быстрее, чем соответствующий рост любой положительной степени φ_ν. Поэтому при некоторых критических значениях $s = s_{\text{кр}}$, которые будут установлены ниже, следует ожидать выполнения граничных условий (12). При этом для синтеза измерителя можно использовать более простой критерий оптимальности (14), приводящий к уравнению оптимизации

$$\rho_\nu^2 \begin{Vmatrix} \hat{g}_1 \\ \hat{g}_3 \end{Vmatrix} \begin{Vmatrix} \langle\varphi^2\rangle & \langle\varphi^4\rangle \\ \langle\varphi^4\rangle & \langle\varphi^6\rangle \end{Vmatrix} = \begin{Vmatrix} -1 \\ -3\,\langle\varphi^2\rangle \end{Vmatrix}. \tag{27}$$

Элементы моментной матрицы представляют собой чётные моменты распределения (26):

$$\langle\varphi^{2i}\rangle = \frac{2^i}{\sqrt{\pi}\; s^{2i}}\;\Gamma\!\left(i+\frac{1}{2}\right),$$

и для $i = 1,2,3$ равны соответственно $\langle\varphi^2\rangle = s^{-2}$, $\langle\varphi^4\rangle = 3s^{-4}$, $\langle\varphi^6\rangle = 15s^{-6}$. Подставляя значения моментов в (27), находим оптимальное решение этого уравнения $\hat{\mathbf{C}} = \| -(\rho_0/\rho_\nu\sigma_x)^2, 0\|^{\mathrm{T}}$ и функцию измерения $\hat{B}_\nu(\varphi_\nu, \hat{\mathbf{G}}) = -[\rho_0/(\rho_\nu\sigma_x)]^2\,\varphi_\nu$. При этом оценка плотности вероятности аргумента принимает вид

$$\hat{Q}(\varphi_\nu) = K_\varphi \exp\!\left(-s^2 \int \varphi_\nu\,\mathrm{d}\varphi_\nu\right) = K_\varphi \exp(-s^2\varphi_\nu^2/2), \tag{28}$$

где коэффициент пропорциональности

$$K_\varphi = \left[\int\limits_{-\pi}^{\pi} \exp\!\left(-\frac{s^2\varphi_\nu^2}{2}\right)\mathrm{d}\varphi_\nu\right]^{-1} \approx \frac{s}{\sqrt{2\pi}}\,.$$

Таким образом, и в этом случае оценка плотности вероятности совпадает с измеряемым законом распределения.

Определим условия, при которых использование критерия (14) допустимо. Для этого сравним оптимизационное уравнение (27) с уравнением, полученным в результате применения более общего критерия (13):

$$\rho_\nu^2 \begin{Vmatrix} \hat{g}_1 \\ \hat{g}_3 \end{Vmatrix} \begin{Vmatrix} \langle\varphi^2\rangle & \langle\varphi^4\rangle \\ \langle\varphi^4\rangle & \langle\varphi^6\rangle \end{Vmatrix} = \begin{Vmatrix} -1 + 2\pi Q(\pi) \\ -3\,\langle\varphi^2\rangle + 2\pi^3 Q(\pi) \end{Vmatrix}. \tag{29}$$

Уравнение (29) справедливо для произвольных значений параметра s и отличается от (27) лишь свободным членом. В случае выполнения неравенств

$$Q(\pi,s) \le Q_{\text{кр1}} \approx \frac{0{,}1}{2\pi} = 1{,}591\cdot 10^{-2}, \qquad Q(\pi,s) \le Q_{\text{кр2}}(s) \approx \frac{0{,}3}{2\pi^3}\,\langle\varphi^2\rangle = 4{,}84\cdot 10^{-3}\,\langle\varphi^2\rangle \tag{30}$$

уравнение (29) переходит в (27), поэтому система (30) является условием допустимости использования критерия (14) для синтеза функции измерения плотности вероятности аргумента.

Находя из (24), (25) при $\varphi_\nu = \pi$ зависимости $Q(\pi,s) = (2\pi)^{-1}\exp(-s^2/2) - s\,[1-\Phi(s)]/\sqrt{2\pi}$ и

$$\langle\varphi^2\rangle = \frac{\pi}{3} + 4\pi \sum_{r=1}^{\infty} (-1)^r\,\frac{a_r(s)}{r^2}\,,$$

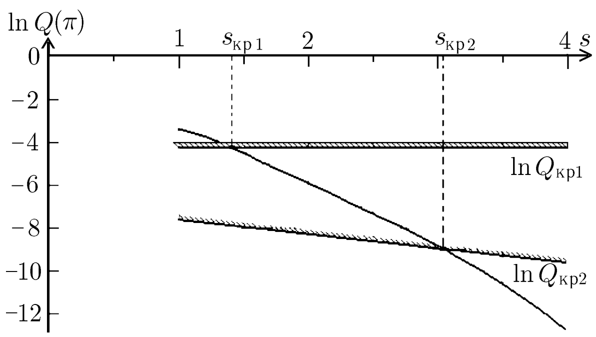


Рис. 1

вычислим величины $\ln Q(\pi, s)$, $\ln Q_{\text{кр}1}$ и $\ln Q_{\text{кр}2}(s)$ и изобразим их в виде графиков в системе координат $\ln Q(\pi), s$ (см. рис. 1).

Находя из графиков значения $s_{\text{кр}1} = 1{,}43$ и $s_{\text{кр}2} = 3{,}07$ как проекции на ось s точек пересечения кривой $\ln Q(\pi)$ с $\ln Q_{\text{кр}1}$ и с $\ln Q_{\text{кр}2}(s)$, определим условие справедливости критерия (14):

$$s \geq s_{\text{кр}2} = 3{,}07 > s_{\text{кр}1}. \tag{31}$$

4. СИНТЕЗ ФУНКЦИЙ ИЗМЕРЕНИЯ ПРИ АПРИОРНО НЕИЗВЕСТНОЙ СТАТИСТИКЕ ВЫБОРКИ

Если априорная информация о законах распределения $P(\rho_\nu)$ и $Q(\varphi_\nu)$ отсутствует, нахождение функционалов (10), (14) становится неосуществимым из-за невозможности реализации в них усреднения функций измерения $\hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}})$ и $\hat{B}_\nu(\varphi_\nu, \hat{\mathbf{G}})$ по ансамблю. Именно в этом случае возникает необходимость в обучении синтезируемого устройства неизвестным функциям измерения и, следовательно, неизвестным плотностям вероятности модуля и аргумента. Обучение должно быть организовано так, чтобы по наблюдаемой выборке $\mathbf{x}$ и измеряемому градиенту случайных функционалов

$$J_\rho(\nu) = \rho_\nu^2 \hat{A}_\nu^2(\rho_\nu, \hat{\mathbf{C}}) - \frac{2}{\rho_\nu}\,\frac{\partial}{\partial \rho_\nu}\,[\rho_\nu^2 \hat{A}_\nu(\rho_\nu, \hat{\mathbf{C}})], \qquad J_\varphi(\nu) = \frac{\rho_\nu}{2}\,\hat{B}_\nu^2(\varphi_\nu, \hat{\mathbf{G}}) + \frac{\partial}{\partial \varphi_\nu}\,\hat{B}_\nu(\varphi_\nu, \hat{\mathbf{G}}) \tag{32}$$

определить с течением времени оптимальные оценки $\hat{\mathbf{C}}$ и $\hat{\mathbf{G}}$.

В [9, 12] для реализации поставленной цели обучения рассматривается так называемый алгоритм стохастической аппроксимации, в котором, в отличие от ранее использовавшегося безынерционного вычисления градиентов ансамблевых средних $\langle J_\rho \rangle$, $\langle J_\varphi \rangle$, применяется итеративное, т. е. временнóе, усреднение градиентов входящих под знак интегралов (10) и (14) случайных функционалов $J_\rho(\nu)$ и $J_\varphi(\nu)$:

$$\hat{\mathbf{C}}(\nu+1) = \hat{\mathbf{C}}(\nu) - \mathbf{\Gamma}_\rho(\nu)\,\nabla_c J_\rho[\nu, \rho_\nu, \hat{\mathbf{C}}(\nu)], \qquad \hat{\mathbf{G}}(\nu+1) = \hat{\mathbf{G}}(\nu) - \mathbf{\Gamma}_\varphi(\nu)\,\nabla_g J_\varphi[\nu, \rho_\nu, \varphi_\nu, \hat{\mathbf{G}}(\nu)], \tag{33}$$

где $\mathbf{\Gamma}_\rho$ и $\mathbf{\Gamma}_\varphi$ — коэффициенты сходимости, представляющие собой диагональные матрицы:

$$\mathbf{\Gamma}_\rho = \begin{Vmatrix} \gamma_{\rho 1} & 0 & \cdots & 0 \\ 0 & \gamma_{\rho 2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \gamma_{\rho n} \end{Vmatrix}, \qquad \mathbf{\Gamma}_\varphi = \begin{Vmatrix} \gamma_{\varphi 1} & 0 & \cdots & 0 \\ 0 & \gamma_{\varphi 2} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \gamma_{\varphi m} \end{Vmatrix}.$$

Структурная схема устройства, реализующего эти алгоритмы, в самом общем виде изображена на рис. 2. Схема формирования функционалов $J_\rho(\nu)$ и $J_\varphi(\nu)$ приведена на рис. 3. Из рисунков видно, что на выходе устройства осуществляется накопление мгновенных оценок параметров функций измерения за время, соответствующее длительности выборки N. Устройство является асимптотически оптимальным, поскольку оптимальное значение оценок достигается в нём не сразу, а с течением времени. Только при бесконечном объёме выборки следует ожидать совпадения

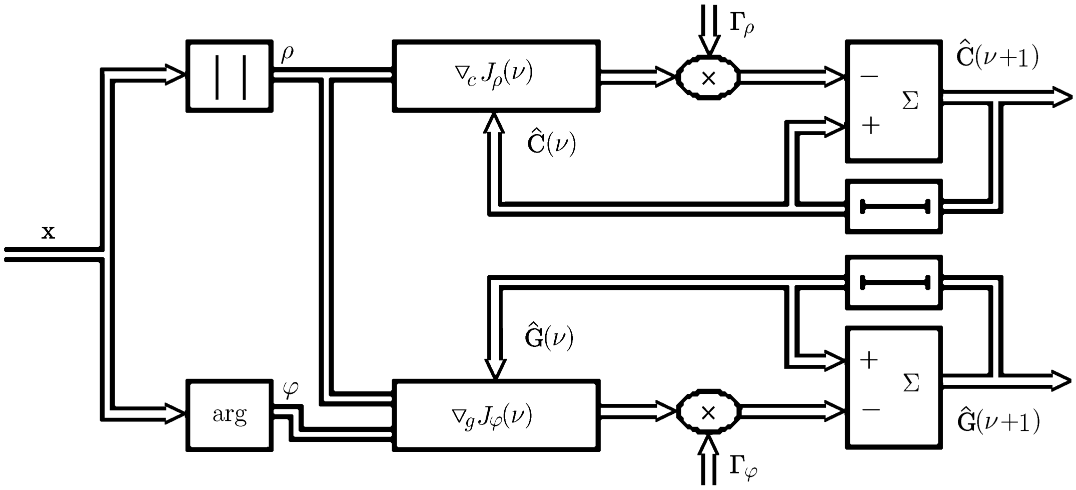


Рис. 2

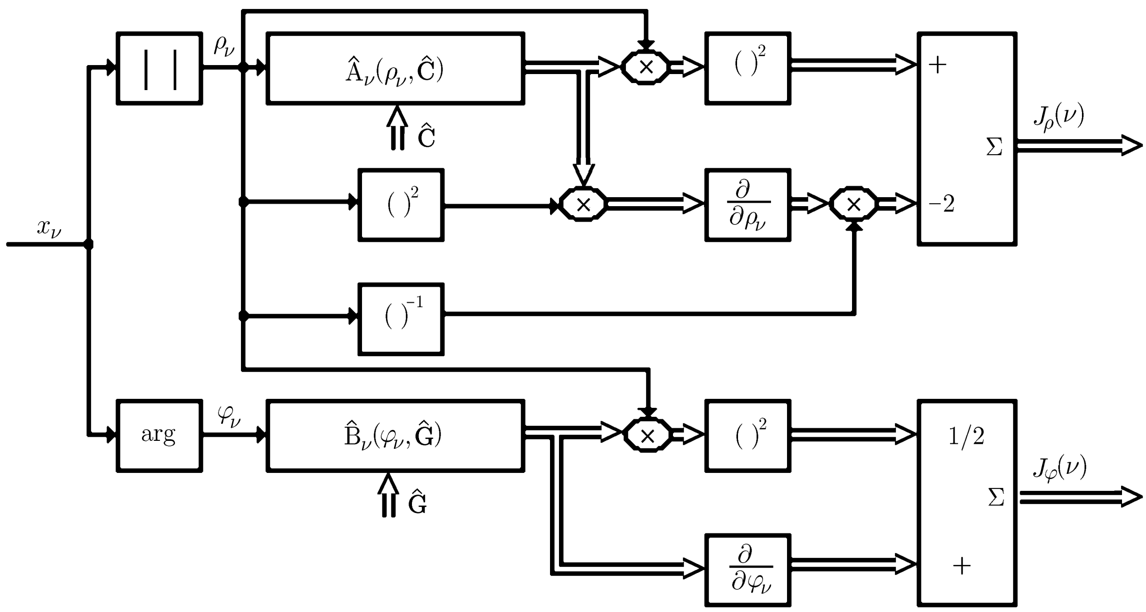


Рис. 3

оценок (33) с ранее рассмотренными оптимальными оценками. Подставляя в функционалы J_ρ и J_φ степенные полиномы (9) и (11), перепишем (33) в окончательном виде:

$$\hat{c}_k(\nu+1) = \hat{c}_k(\nu) - 2\gamma_{\rho k}(\nu)\rho_\nu^k \sum_{i=-2}^{n-3} \left[\hat{c}_i(\nu)\rho_\nu^{i+2} - 2 - k\right], \qquad (34а)$$

где $k = -2, \ldots, n-3$,

$$\hat{g}_r(\nu+1) = \hat{g}_r(\nu) - 2\gamma_{\varphi r}(\nu)\rho_\nu^2 \sum_{\mu=1}^{m} \left[\hat{g}_\mu \varphi_\nu^{\mu+r} + r\varphi_\nu^{r-1}\right], \qquad (34б)$$

где $r = 1, \ldots, m$.

При независимости от номера и безразмерного времени коэффициентов сходимости: $\gamma_{\rho k}(\nu) = \gamma_\rho$, $\gamma_{\varphi r}(\nu) = \gamma_\varphi$, и выполнении условий устойчивости синтезированного алгоритма последовательность итераций (34) в среднем сходится к оптимальным векторам (19) и (27). Покажем это, полагая, что при значительной инерционности устройства ($N \gg 1$) закон распределения компонент векторов $\hat{\mathbf{C}}(\nu)$ и $\hat{\mathbf{G}}(\nu)$ достаточно близок к нормальному, а сами векторы статистически не связаны с флуктуациями модуля и аргумента принятой выборки [13].

Действительно, в рамках принятого допущения первые моменты (34) равны

$$\langle \hat{c}_k(\nu+1)\rangle = \langle \hat{c}_k(\nu)\rangle - 2\gamma_\rho \sum_{i=-2}^{n-3} \left[\langle \hat{c}_i(\nu)\rangle \langle \rho^{k+i+2}\rangle - (2+k)\langle \rho^k\rangle\right],$$

$$\langle \hat{g}_r(\nu+1)\rangle = \langle \hat{g}_r(\nu)\rangle - 2\gamma_\varphi \rho_\nu^2 \sum_{\mu=1}^{m} \left[\langle \hat{g}_\mu(\nu)\rangle \langle \varphi^{\mu+r}\rangle + r\langle \varphi^{r-1}\rangle\right].$$

Переходя в последней системе уравнений от временнóй к операторной форме записи (дискретное преобразование Лапласа), получим

$$\langle \hat{C}_k(q)\rangle [\exp(q) - 1] = -2\gamma_\rho \sum_{i=-2}^{n-3} \left[\langle \hat{C}_i(q)\rangle \langle \rho^{k+i+2}\rangle - (2+k)\langle \rho^k\rangle\right],$$

$$\langle \hat{G}_r(q)\rangle [\exp(q) - 1] = -2\gamma_\varphi \rho_\nu^2 \sum_{\mu=1}^{m} \left[\langle \hat{G}_\mu(q)\rangle \langle \varphi^{\mu+r}\rangle + r\langle \varphi^{r-1}\rangle\right]. \qquad (35)$$

Согласно теореме об установившемся значении оригиналов $\lim_{\nu\to\infty} \hat{c}_k(\nu) = \lim_{q\to 0} \hat{C}_k(q)$, $\lim_{\nu\to\infty} \hat{g}_r(\nu) = \lim_{q\to 0} \hat{G}_r(q)$, из (35) несложно получить систему оптимизационных уравнений:

$$\sum_{i=-2}^{n-3} \langle \hat{C}_i(0)\rangle \langle \rho^{k+i+2}\rangle - (2+k)\langle \rho^k\rangle = 0, \qquad \rho_\nu^2 \sum_{\mu=1}^{m} \langle \hat{G}_\mu(0)\rangle \langle \varphi^{\mu+r}\rangle + r\langle \varphi^{r-1}\rangle = 0, \qquad (36)$$

которая по форме ничем не отличается от системы уравнений (15) и (22) при выполнении условия (31). Последнее означает, что итерационные алгоритмы (34) оптимальны в среднестатистическом смысле, т. к. обеспечивают сходимость первого момента векторов состояния измерительной системы и соответствующих функций измерения к оптимальным значениям.

Таким образом, выше были сформулированы критерии качества оценивания плотностей вероятности модуля и аргумента комплексного случайного числа, установлены границы справедливости каждого из критериев, синтезированы алгоритмы оптимальной оценки плотностей вероятности при априорно известной и неизвестной статистике выборки. Показано, что итеративные алгоритмы оценивания обеспечивают сходимость в среднестатистическом смысле функций измерения к оптимальным значениям.

СПИСОК ЛИТЕРАТУРЫ

1. Теория обнаружения сигналов / Под ред. П. А. Бакута. М.: Радио и связь, 1984.
2. Сосулин Ю. Г. Теория обнаружения и оценивания стохастических сигналов. М.: Сов. радио, 1978.
3. Левин Б. Р. Теоретические основы статистической радиотехники. Кн. 3. М.: Сов. радио, 1976.

4. Миддлтон Д. Введение в статистическую теорию радиосвязи. Т. 2. М.: Сов. радио, 1962.
5. Островский М. А., Пахомов Ю. И. // Изв. вузов. Радиофизика. 1990. Т. 33, № 6. С. 689.
6. Островский М. А. // Изв. вузов. Радиофизика. 1995. Т. 38, № 8. С. 870.
7. Ширман Я. Д., Островский М. А. // Радиотехника и электроника. 1990. Т. 35, № 8. С. 1 655.
8. Островский М. А. // Изв. вузов. Радиофизика. 1997. Т. 40, № 11. С. 1 416.
9. Левин Б. Р. Теоретические основы статистической радиотехники. Кн. 2. М.: Сов. радио, 1969.
10. Левин Б. Р. Теоретические основы статистической радиотехники. Кн. 1. М.: Сов. радио, 1968.
11. Островский М. А., Рябинин С. А. // Изв. вузов. Радиофизика. 1998. Т. 41, № 10. С. 1 314.
12. Цыпкин Я. З. Основы теории обучающихся систем. М.: Сов. радио, 1976.
13. Лейких М. А., Островский М. А. // Динамика систем. Устойчивость динамических систем и процессов управления. Горький, 1979. С. 86.

Поступила в редакцию 21 июня 2004 г.; принята в печать 31 марта 2005 г.

ESTIMATION OF PROBABILITY DENSITY OF THE ABSOLUTE VALUE AND ARGUMENT OF COMPLEX RANDOM VARIABLES

M. A. Ostrovsky and M. V. Ukhanov

Criteria for estimating the probability densities of the N-dimensional random vector proposed in the paper written by one of the authors are specified for the case of even N. In this case, the two-dimensional vectors are treated as complex random values and the probability densities of the absolute values and the arguments of the latter are estimated for the presentation convenience. The estimation algorithms are synthesized on the basis of the obtained criteria under the conditions of both complete *a priori* information on the sample statistic and its absence. Examples are shown to prove the optimality of the obtained algorithms in the average statistical sense.

УДК 519.217:517.977.57

ОЦЕНИВАНИЕ МАРКОВСКИХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ СО СКАЧКООБРАЗНЫМ ИЗМЕНЕНИЕМ ПАРАМЕТРОВ МЕТОДОМ ИНТЕРПОЛЯЦИИ

А. В. Королёв, А. М. Силаев

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

Методами теории оптимальной нелинейной фильтрации разработан алгоритм нахождения оптимальных оценок последовательности скрытых состояний дискретнозначных марковских процессов со скачкообразно изменяющимися параметрами в неизвестный момент времени. Оптимальные оценки состояний марковских процессов и момента появления скачка параметров получены в результате интерполяции путём обработки всей последовательности наблюдений. Приведены результаты моделирования работы алгоритма.

ВВЕДЕНИЕ

Проблеме оптимального оценивания случайных марковских процессов со скачкообразными изменениями параметров, а также задачам оптимального обнаружения скачков к настоящему времени посвящено большое число работ (см., например, [1–7]). Для решения данных задач используются различные методы. Например, в [4] рассматриваются процедуры оценивания случайных процессов и моментов скачков, осуществляемые после окончания наблюдений. При этом для оценивания моментов скачкообразного изменения свойств случайных процессов используется метод максимального правдоподобия или байесовский подход (если задана априорная плотность вероятности моментов скачков). Однако в случае сложных многокомпонентных процессов или при широком диапазоне возможных значений моментов скачков реализация подобных процедур анализа требует слишком большого объёма вычислений. Последовательные статистические методы анализа, когда обработка случайного процесса проводится в текущем времени по мере поступления новой информации, характеризуются более высоким быстродействием, но в определённых задачах не позволяют добиться высокой точности оценивания. В ряде случаев при решении задач интерполяции сигналов удаётся избежать сложных многоканальных схем обработки за счёт отсутствия последействия марковских процессов и синтеза рекуррентных алгоритмов обработки наблюдаемых процессов в прямом и обратном времени [8, 9].

Модели условных марковских процессов в настоящее время широко используются для решения задач обработки сигналов. В большинстве работ рассматриваются диффузионные марковские процессы, которые описывают сигналы на выходе динамических систем при возбуждении их белым гауссовским шумом. В другом классе процессов, привлекающем в последнее время внимание исследователей, для описания каналов связи используются скрытые марковские модели (hidden Markov model, HMM) с конечным числом состояний, между которыми происходят переходы в случайные моменты времени [9, 10].

В настоящей работе при выводе алгоритма оценивания марковских последовательностей со скачкообразным изменением параметров методом интерполяции применяется математический аппарат теории нелинейной фильтрации марковских случайных процессов [1, 11–14]. При этом используются также результаты работ [15, 16], в которых решались задачи оптимальной фильтрации марковских сигналов и получены алгоритмы оптимального оценивания моментов скачков параметров непосредственно в текущем времени.

1. ПОСТАНОВКА ЗАДАЧИ

Пусть в дискретном времени на интервале $k = 1, 2, \ldots, T$ последовательность наблюдений $\{y_k\}$ статистически связана с последовательностью скрытых состояний $\{x_k\}$ следующими условными вероятностями:

$$b_j(Y_m) = \begin{cases} b_j^0(Y_m) = P^0\bigl(y_k = Y_m \mid x_k = X_j\bigr), & k \le \tau; \\ b_j^1(Y_m) = P^1\bigl(y_k = Y_m \mid x_k = X_j\bigr), & \tau < k \le T. \end{cases} \tag{1}$$

Здесь $P^0\bigl(y_k = Y_m \mid x_k = X_j\bigr)$ и $P^1\bigl(y_k = Y_m \mid x_k = X_j\bigr)$ — известные условные вероятности наблюдений, Y_m — наблюдаемая дискретная величина из набора M значений, X_j — скрытое состояние из дискретного набора N значений, τ — неизвестный момент времени, при котором происходит скачкообразное изменение параметров модели, $m = 1, 2, \ldots, M$, $j = 1, 2, \ldots, N$.

Пусть марковская последовательность скрытых состояний описывается вероятностью переходов следующего вида при $k = 1, 2, \ldots, T$:

$$a_{ij} = \begin{cases} a_{ij}^0 = P^0\bigl(x_k = X_j \mid x_{k-1} = X_i\bigr), & k \le \tau; \\ a_{ij}^1 = P^1\bigl(x_k = X_j \mid x_{k-1} = X_i\bigr), & \tau < k \le T, \end{cases} \tag{2}$$

где $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, N$, и вероятностью начальных состояний при $k = 0$:

$$P(x_0 = X_j) = \begin{cases} P^0\bigl(x_0 = X_j\bigr), & \tau > 0; \\ P^1\bigl(x_0 = X_j\bigr), & \tau = 0. \end{cases} \tag{3}$$

Здесь $P^0(x_k = X_j \mid x_{k-1} = X_i)$, $P^1(x_k = X_j \mid x_{k-1} = X_i)$, $P^0(x_0 = X_j)$ и $P^1(x_0 = X_j)$ — известные величины.

Задача состоит в том, чтобы по данным реализации наблюдений $\{y_k\}$ найти оптимальные оценки состояний скрытого марковского процесса $\{\hat{x}_k\}$ и момента дискретного времени $\hat{\tau}$, при котором происходит скачок параметров. Предполагается, что ещё до начала оценивания известна вся последовательность наблюдений $\{y_k\}$, где $k = 1, 2, \ldots, T$. Также предполагаются известными начальные и переходные условные вероятности $P^0(x_0), a_{ij}^0, b_j^0(Y_m), P^1(x_0), a_{ij}^1, b_j^1(Y_m)$ и априорные вероятности случайного момента появления скачка $P_\tau(\tau)$ при $\tau \ge 0$.

2. АЛГОРИТМ ВЫЧИСЛЕНИЯ АПОСТЕРИОРНОЙ ВЕРОЯТНОСТИ СОСТОЯНИЙ МОДЕЛИ

Наиболее полную информацию об оцениваемом состоянии модели содержит апостериорная вероятность $P(x_k \mid y_1^T)$, найденная при условии, что на интервале времени $k = 1, 2, \ldots, T$ наблюдается процесс $\{y_k\}$.

Запишем выражение для $P(x_k \mid y_1^T)$, воспользовавшись формулой Байеса:

$$\begin{aligned} P(x_k \mid y_1^T) = \frac{P(x_k, y_{k+1}^T \mid y_1^k)}{P(y_{k+1}^T \mid y_1^k)} = \\ = \frac{P(x_k, y_{k+1}^T, \tau < k \mid y_1^k) + P(x_k, y_{k+1}^T, \tau = k \mid y_1^k) + P(x_k, y_{k+1}^T, \tau > k \mid y_1^k)}{P(y_{k+1}^T \mid y_1^k)}\ , \end{aligned} \tag{4}$$

где τ — момент времени, при котором происходит скачок параметров модели. Таким образом, $P(x_k, y_{k+1}^T \mid y_1^k)$ представляется в виде суммы вероятностей, найденных до скачка, в момент скачка и после скачка параметров. Получим рекуррентные уравнения для каждого из слагаемых суммы.

Сначала вычислим первое слагаемое в числителе (4):

$$P(x_k, y_{k+1}^T, \tau < k \mid y_1^k) = P(\tau < k \mid y_1^k) P(x_k \mid \tau < k, y_1^k) P(y_{k+1}^T \mid x_k, \tau < k, y_1^k) = \\ = P(\tau < k \mid y_1^k) P_1(x_k, k) P(y_{k+1}^T \mid x_k, \tau < k) = P(\tau < k \mid y_1^k) P_1(x_k, k) P_1(y_{k+1}^T \mid x_k). \quad (5)$$

Здесь $P_1(x_k, k) = P(x_k \mid \tau < k, y_1^k)$ — апостериорная вероятность нахождения процесса в состоянии x_k при условии, что скачок параметров к моменту времени k уже произошёл, $P(\tau < k \mid y_1^k)$ — вероятность появления скачка к моменту времени k. Последний сомножитель $P_1(y_{k+1}^T \mid x_k) \equiv \equiv P(y_{k+1}^T \mid x_k, \tau < k)$ в (5) не зависит от y_1^k, т. к. в силу предположений модели (1)–(3) реализация y_1^k при известном x_k не несёт дополнительной информации о других состояниях процесса.

Запишем рекуррентное уравнение для функции $P_1(y_{k+1}^T \mid x_k)$:

$$P(y_{k+1}^T \mid x_k, \tau < k)\big|_{x_k = X_i} \equiv P_1(y_{k+1}^T \mid x_k)\big|_{x_k = X_i} = \sum_{j=1}^{N} P(y_{k+1}, y_{k+2}^T, x_{k+1} \mid x_k, \tau < k)\big|_{x_k = X_i, x_{k+1} = X_j} = \\ = \sum_{j=1}^{N} P(y_{k+1} \mid x_{k+1}, x_k, \tau < k) P(x_{k+1} \mid x_k, \tau < k) P(y_{k+2}^T \mid y_{k+1}, x_k, x_{k+1}, \tau < k)\big|_{x_k = X_i, x_{k+1} = X_j} = \\ = \sum_{j=1}^{N} a_{ij}^1 b_j^1(y_{k+1}) P(y_{k+2}^T \mid x_{k+1}, \tau < k+1)\big|_{x_{k+1} = X_j} = \sum_{j=1}^{N} a_{ij}^1 b_j^1(y_{k+1}) P_1(y_{k+2}^T \mid x_{k+1})\big|_{x_{k+1} = X_j}, \quad (6)$$

где $k = T-2, T-3, \ldots, 0$, $i = 1, 2, \ldots, N$.

При $k = T-1$ для $P_1(y_T \mid x_{T-1})$ выполняется соотношение

$$P_1(y_T \mid x_{T-1})\big|_{x_{T-1} = X_i} = \sum_{j=1}^{N} a_{ij}^1 b_j^1(y_T). \quad (7)$$

Таким образом, при разных значениях x_k, начиная от конечного момента времени наблюдений, последовательность функций $P_1(y_{k+1}^T \mid x_k)$ можно рекуррентно вычислять в обратном времени при $k = T-1, T-2, \ldots, 1$ в соответствии с формулами (6), (7). В этом алгоритме не задействованы момент скачка τ и параметры модели до скачка параметров $P^0(x_0), a_{ij}^0, b_j^0(Y_m)$, поэтому функции правдоподобия $P_1(y_{k+1}^T \mid x_k)$ не зависят от $\tau, P^0(x_0), a_{ij}^0, b_j^0(Y_m)$.

Аналогично можно получить, что функция правдоподобия $P(y_{k+1}^T \mid x_k, \tau = k)$ находится из тех же уравнений (6), (7), что и $P(y_{k+1}^T \mid x_k, \tau < k)$. При этом выполняется равенство

$$P(y_{k+1}^T \mid x_k, \tau \le k) = P(y_{k+1}^T \mid x_k, \tau = k) = P(y_{k+1}^T \mid x_k, \tau < k) \equiv P_1(y_{k+1}^T \mid x_k). \quad (8)$$

Вычислим второе слагаемое в числителе правой части (4), разложив $P(x_k, y_{k+1}^T, \tau = k \mid y_1^k)$ на условные вероятности и используя равенство (8):

$$P(x_k, y_{k+1}^T, \tau = k \mid y_1^k) = P(\tau = k \mid y_1^k) P(x_k \mid \tau = k, y_1^k) P(y_{k+1}^T \mid x_k, \tau = k, y_1^k) = \\ = P(\tau = k \mid y_1^k) P_0(x_k, k) P_1(y_{k+1}^T \mid x_k). \quad (9)$$

Здесь $P(\tau = k \mid y_1^k)$ — апостериорная вероятность появления скачка параметров модели в момент времени k, $P_0(x_k, k) \equiv P(x_k \mid \tau \ge k, y_1^k) = P(x_k \mid \tau = k, y_1^k) = P(x_k \mid \tau > k, y_1^k)$ — апостериорная вероятность нахождения модели в состоянии x_k при условии, что к моменту времени k скачок параметров ещё не произошёл.

Вычислим третье слагаемое в числителе правой части (4):

$$P(x_k, y_{k+1}^T, \tau > k \mid y_1^k) = P(\tau > k \mid y_1^k) P(x_k \mid \tau > k, y_1^k) P(y_{k+1}^T \mid x_k, \tau > k, y_1^k) = \\ = P(\tau > k \mid y_1^k) P_0(x_k, k) P(y_{k+1}^T \mid x_k, \tau > k) = P(\tau > k \mid y_1^k) P_0(x_k, k) P_0(y_{k+1}^T \mid x_k). \quad (10)$$

Отметим, что в силу уравнений модели (1)–(3) функция правдоподобия $P_0(y_{k+1}^T \mid x_k) \equiv \equiv P(y_{k+1}^T \mid x_k, \tau > k, y_1^k)$ не зависит от y_1^k при заданном x_k. Найдём рекуррентный алгоритм для вычисления функции $P_0(y_{k+1}^T \mid x_k)$ в правой части формулы (10):

$$P_0(y_{k+1}^T \mid x_k) = \frac{P(y_{k+1}^T, x_k \mid \tau > k)}{P(x_k \mid \tau > k)} = \frac{P(y_{k+1}^T, x_k, \tau = k+1 \mid \tau > k) + P(y_{k+1}^T, x_k, \tau > k+1 \mid \tau > k)}{P(x_k \mid \tau > k)} = \\ = \frac{P(\tau = k+1 \mid \tau > k) P(y_{k+1}^T, x_k \mid \tau = k+1, \tau > k) + P(\tau > k+1 \mid \tau > k) P(y_{k+1}^T, x_k \mid \tau > k+1, \tau > k)}{P(x_k \mid \tau > k)} = \\ = \frac{P(\tau = k+1 \mid \tau > k) P(y_{k+1}^T, x_k \mid \tau = k+1) + P(\tau > k+1 \mid \tau > k) P(y_{k+1}^T, x_k \mid \tau > k+1)}{P(x_k \mid \tau > k)}. \quad (11)$$

Введём функцию $\nu_0(k) = P(\tau = k \mid \tau \geq k)$, которая имеет смысл априорной вероятности появления скачка параметров в момент времени k при условии, что до этого он не произошёл [14, 15]. Функция $\nu_0(k)$ выражается через априорную вероятность момента появления скачка следующим образом: $\nu_0(k) = P_\tau(k) / \sum_{\tau=k}^{\infty} P_\tau(\tau)$. Тогда из (11) получим

$$P_0(y_{k+1}^T \mid x_k) = \frac{\nu_0(k+1) P(y_{k+1}^T, x_k \mid \tau = k+1) + [1 - \nu_0(k+1)] P(y_{k+1}^T, x_k \mid \tau > k+1)}{P(x_k \mid \tau > k)} = \\ = \frac{\nu_0(k+1) P(y_{k+1}^T \mid x_k, \tau = k+1) P(x_k \mid \tau = k+1)}{P(x_k \mid \tau \geq k+1)} + \\ + \frac{[1 - \nu_0(k+1)]\, P(y_{k+1}^T \mid x_k, \tau > k+1) P(x_k \mid \tau > k+1)}{P(x_k \mid \tau \geq k+1)} = \\ = \nu_0(k+1) P(y_{k+1}^T \mid x_k, \tau = k+1) + [1 - \nu_0(k+1)]\, P(y_{k+1}^T \mid x_k, \tau > k+1). \quad (12)$$

Здесь учтено, что $P(x_k \mid \tau > k+1) = P(x_k \mid \tau = k+1) = P(x_k \mid \tau \geq k+1)$, т. к. в силу модели процесса (2), (3) данные функции определяются только переходными вероятностями a_{ij}^0 и начальной вероятностью $P^0(x_0)$.

Функцию правдоподобия в первом слагаемом правой части (12) при $k = T-2, T-3, \ldots, 0$ представим в виде

$$P(y_{k+1}^T \mid x_k, \tau = k+1)\big|_{x_k = X_i} = \sum_{j=1}^{N} P(y_{k+1}, y_{k+2}^T, x_{k+1} \big| x_k, \tau = k+1)\big|_{x_k = X_i, x_{k+1} = X_j} = \\ = \sum_{j=1}^{N} P(x_{k+1} \mid x_k, \tau = k+1) P(y_{k+1} \mid x_{k+1}, x_k, \tau = k+1) \times \\ \times P(y_{k+2}^T \mid y_{k+1}, x_{k+1}, x_k, \tau = k+1)\big|_{x_k = X_i, x_{k+1} = X_j} = \sum_{j=1}^{N} a_{ij}^0 b_j^0(y_{k+1}) P_1(y_{k+2}^T \mid x_{k+1})\big|_{x_{k+1} = X_j}. \quad (13)$$

Здесь учтено, что в силу (8) выполняется равенство $P(y_{k+2}^T \mid x_{k+1}, \tau = k+1) = P_1(y_{k+2}^T \mid x_{k+1})$.

Аналогично функция правдоподобия во втором слагаемом правой части (12) представляется в виде

$$P(y_{k+1}^T \mid x_k, \tau > k+1)\big|_{x_k = X_i} = \sum_{j=1}^{N} P(y_{k+1}, y_{k+2}^T, x_{k+1} \mid x_k, \tau > k+1)\big|_{x_k = X_i, x_{k+1} = X_j} =$$
$$= \sum_{j=1}^{N} P(x_{k+1} \mid x_k, \tau > k+1) P(y_{k+1} \mid x_{k+1}, x_k, \tau > k+1) \times$$
$$\times P(y_{k+2}^T \mid y_{k+1}, x_{k+1}, x_k, \tau > k+1)\big|_{x_k = X_i, x_{k+1} = X_j} = \sum_{j=1}^{N} a_{ij}^0 b_j^0(y_{k+1}) P_0(y_{k+2}^T \mid x_{k+1})\big|_{x_{k+1} = X_j}, \quad (14)$$

где $k = T-2, T-3, \ldots, 0$.

В результате имеем рекуррентную формулу

$$P_0(y_{k+1}^T \mid x_k)\big|_{x_k = X_i} = \nu_0(k+1) \sum_{j=1}^{N} a_{ij}^0 b_j^0(y_{k+1}) P_1(y_{k+2}^T \mid x_{k+1})\big|_{x_{k+1} = X_j} +$$
$$+ [1 - \nu_0(k+1)] \sum_{j=1}^{N} a_{ij}^0 b_j^0(y_{k+1}) P_0(y_{k+2}^T \mid x_{k+1})\big|_{x_{k+1} = X_j}, \quad (15)$$

где $k = T-2, T-3, \ldots, 0$.

При $k = T-1$ для функции $P_0(y_T \mid x_{T-1})$ выполняется соотношение

$$P_0(y_T \mid x_{T-1})|_{x_k = X_i} = \sum_{j=1}^{N} a_{ij}^0 b_j^0(y_T)\big|_{x_{k+1} = X_j}. \quad (16)$$

Таким образом, последовательность функций $P_0(y_{k+1}^T \mid x_k)$ также можно рекуррентно вычислять в обратном времени. При этом значения функций правдоподобия $P_0(y_{k+1}^T \mid x_k)$ не зависят от τ, $P^1(x_1)$, a_{ij}^1 и $b_j^1(Y_m)$.

С учётом выражений (5), (9) и (10) уравнение (4) для апостериорной вероятности процесса x_k принимает вид

$$P(x_k \mid y_1^T) = \frac{P(\tau < k \mid y_1^k) P_1(x_k, k) P_1(y_{k+1}^T \mid x_k)}{P(y_{k+1}^T \mid y_1^k)} + \frac{P(\tau = k \mid y_1^k) P_0(x_k, k) P_1(y_{k+1}^T \mid x_k)}{P(y_{k+1}^T \mid y_1^k)} +$$
$$+ \frac{P(\tau > k \mid y_1^k) P_0(x_k, k) P_0(y_{k+1}^T \mid x_k)}{P(y_{k+1}^T \mid y_1^k)}. \quad (17)$$

Введём обозначения для функций в правой части формулы (17):

$$p_1(k) \equiv P(\tau < k \mid y_1^k), \qquad p_0(k) \equiv P(\tau \geq k \mid y_1^k). \quad (18)$$

Эти функции имеют смысл апостериорных вероятностей соответственно наличия и отсутствия скачка к моменту времени k. Выполняются следующие соотношения [14]:

$$P(\tau > k \mid y_1^k) = P(\tau > k, \tau \geq k \mid y_1^k) = P(\tau > k \mid \tau \geq k) P(\tau \geq k \mid y_1^k) =$$
$$= P(\tau > k \mid \tau \geq k) p_0(k) = p_0(k) \, \frac{P(\tau > k)}{P(\tau \geq k)} = p_0(k) \, [1 - \nu_0(k)],$$

$$P(\tau = k \mid y_1^k) = \frac{P_\tau(k)}{\sum\limits_{\tau=k}^{T} P_\tau(\tau)}\, p_0(k) = \nu_0(k) p_0(k). \tag{19}$$

С учётом соотношений (18), (19) выражение (17) принимает вид

$$P(x_k \mid y_1^T) = \frac{p_1(k) P_1(x_k, k) + \nu_0(k) p_0(k) P_0(x_k, k)}{P(y_{k+1}^T \mid y_1^k)}\, P_1(y_{k+1}^T \mid x_k) + \\ + \frac{p_0(k)\,[1 - \nu_0(k)] P_0(x_k, k)}{P(y_{k+1}^T \mid y_1^k)}\, P_0(y_{k+1}^T \mid x_k), \tag{20}$$

где $k = 0, 1, \ldots, T-1$.

Таким образом, формула (20) позволяет вычислять апостериорные вероятности состояния модели $P(x_k \mid y_1^T)$ с помощью апостериорных вероятностей наличия ($p_1(k)$) и отсутствия ($p_0(k)$) скачка к моменту времени k, условных вероятностей состояний модели $P_1(x_k, k)$, $P_0(x_k, k)$ и функций правдоподобия $P_1(y_{k+1}^T \mid x_k)$, $P_0(y_{k+1}^T \mid x_k)$ и $P(y_{k+1}^T \mid y_1^k)$. При этом функции $P(y_{k+1}^T \mid y_1^k)$ в (20) имеют смысл нормировочных коэффициентов и могут быть найдены из условия нормировки $\sum_{j=1}^N P(x_k \mid y_1^T)\big|_{x_k = X_j} = 1$. Для функций $P_1(y_{k+1}^T \mid x_k)$ и $P_0(y_{k+1}^T \mid x_k)$ получены рекуррентные уравнения (6), (7), (15), (16) в обратном времени.

Вероятности наличия и отсутствия скачка и условные вероятности состояния модели можно найти, используя методы теории оптимальной нелинейной фильтрации марковских процессов [1–3]. Далее изложен алгоритм нахождения функций $p_0(k)$, $p_1(k)$ и $P_1(x_k, k)$, $P_0(x_k, k)$.

3. АЛГОРИТМ ВЫЧИСЛЕНИЯ СОСТОЯНИЙ СКРЫТЫХ МАРКОВСКИХ ПРОЦЕССОВ

Задачи фильтрации дискретно-непрерывных марковских процессов решались в [14, 15]. Применяя результаты указанных работ к рассматриваемому случаю, для функций $p_1(k)$, $p_0(k)$, $P_1(x_k, k)$ и $P_0(x_k, k)$ можно записать рекуррентные уравнения:

$$p_0(k+1) = \frac{\Phi_0(y_{k+1}, k)}{\Phi(y_{k+1}, k)}\, [1 - \nu_0(k)]\, p_0(k), \qquad p_1(k+1) = \frac{\Phi_1(y_{k+1}, k)}{\Phi(y_{k+1}, k)}\, p_1(k) + \frac{\Phi_{01}(y_{k+1}, k)}{\Phi(y_{k+1}, k)}\, \nu_0(k) p_0(k),$$

$$P_0(x_{k+1}, k+1)\big|_{x_{k+1} = X_j} = \frac{b_j^0(y_{k+1}) \sum\limits_{i=1}^N a_{ij}^0 P_0(x_k, k)\big|_{x_k = X_i, x_{k+1} = X_j}}{\Phi_0(y_{k+1}, k)}\ ,$$

$$P_1(x_{k+1}, k+1)\big|_{x_{k+1} = X_j} = \frac{b_j^1(y_{k+1}) p_1(k) \sum\limits_{i=1}^N a_{ij}^1 P_1(x_k, k)\big|_{x_k = X_i}}{\Phi_1(y_{k+1}, k) p_1(k) + \Phi_{01}(y_{k+1}, k) \nu_0(k) p_0(k)} + \\ + \frac{b_j^1(y_{k+1}) \nu_0(k) p_0(k) \sum\limits_{i=1}^N a_{ij}^1 P_0(x_k, k)\big|_{x_k = X_i}}{\Phi_1(y_{k+1}, k) p_1(k) + \Phi_{01}(y_{k+1}, k) \nu_0(k) p_0(k)}\, , \tag{21}$$

где $k = 0, 1, \ldots, T-1$, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, N$ с начальными условиями

$$p_0(0) = \sum_{\tau=1}^{\infty} P_\tau(\tau), \qquad p_1(0) = 1 - p_0(0), \qquad P_0(x_0, 0) = P^0(x_0), \qquad P_1(x_0, 0) = P^1(x_0). \tag{22}$$

В формулах (21) приняты следующие обозначения:

$$\Phi_0(y_{k+1},k)=\sum_{i=1}^{N}\sum_{j=1}^{N} b_j^0(y_{k+1})a_{ij}^0 P_0(x_k,k)\big|_{x_k=X_i}, \qquad \Phi_1(y_{k+1},k)=\sum_{i=1}^{N}\sum_{j=1}^{N} b_j^1(y_{k+1})a_{ij}^1 P_1(x_k,k)\big|_{x_k=X_i},$$

$$\Phi_{01}(y_{k+1},k)=\sum_{i=1}^{N}\sum_{i=j}^{N} b_j^1(y_{k+1})a_{ij}^1 P_0(x_k,k)\big|_{x_k=X_i},$$

$$\Phi(y_{k+1},k)=p_0(k)\Phi_0(y_{k+1},k)+p_1(k)\Phi_1(y_{k+1},k)+\nu_0(k)p_0(k)\left[\Phi_{01}(y_{k+1},k)-\Phi_0(y_{k+1},k)\right]. \qquad (23)$$

Введём вспомогательные вероятности

$$P_{11}(x_{k+1},k+1)\big|_{x_{k+1}=X_j}=\frac{b_j^1(y_{k+1})\sum\limits_{i=1}^{N} a_{ij}^1 P_1(x_k,k)\big|_{x_k=X_i}}{\Phi_1(y_{k+1},k)},$$

$$P_{01}(x_{k+1},k+1)\big|_{x_{k+1}=X_j}=\frac{b_j^1(y_{k+1})\sum\limits_{i=1}^{N} a_{ij}^1 P_0(x_k,k)\big|_{x_k=X_i}}{\Phi_{01}(y_{k+1},k)} \qquad (24)$$

и коэффициент

$$\beta_1(k+1)=\frac{\nu_0(k)p_0(k)\Phi_{01}(y_{k+1},k)}{\nu_0(k)p_0(k)\Phi_{01}(y_{k+1},k)+p_1(k)\Phi_1(y_{k+1},k)}\,. \qquad (25)$$

Тогда уравнение для условной вероятности $P_1(x_k,k)$ в (21) можно представить в более наглядной форме:

$$P_1(x_{k+1},k+1)=\beta_1(k+1)P_{01}(x_{k+1},k+1)+\left[1-\beta_1(k+1)\right]P_{11}(x_{k+1},k+1). \qquad (26)$$

Уравнения (21), (26) для функций $p_1(k)$, $p_0(k)$, $P_1(x_k,k)$ и $P_0(x_k,k)$ с начальными условиями (22) и обозначениями (23)–(25) составляют искомый рекуррентный алгоритм нахождения апостериорных вероятностей моментов появления скачков и условных вероятностей состояний модели. Для нахождения апостериорных функций $p_1(k)$, $p_0(k)$, $P_1(x_k,k)$ и $P_0(x_k,k)$ в момент времени $k+1$ необходимо знать их значения в предыдущий момент времени k и новые наблюдения y_{k+1}.

Таким образом, выражение (20) и рекуррентные уравнения (6), (7), (15), (16), (21), (22), (26) образуют искомый алгоритм нахождения оценки апостериорной вероятности состояния модели $P(x_k\mid y_1^T)$. Ещё раз опишем структуру алгоритма.

Для того, чтобы оценить $P(x_k\mid y_1^T)$, необходимо:

1) используя начальные условия (22) и рекуррентные уравнения (21) и (26), в прямом времени найти оценки апостериорных вероятностей наличия и отсутствия скачка $p_1(k)$, $p_0(k)$ и оценки условных вероятностей состояния модели $P_1(x_k,k)$, $P_0(x_k,k)$ при $k=1,2,\ldots,T$;

2) используя конечные условия (7), (16) и рекуррентные уравнения (6), (15), в обратном времени найти оценки функций $P_1(y_{k+1}^T\mid x_k$, $P_0(y_{k+1}^T\mid x_k)$ при $k=T-1,T-2,\ldots,0$;

3) подставить найденные значения функций $p_1(k)$, $p_0(k)$, $P_1(x_k,k)$, $P_0(x_k,k)$, $P_1(y_{k+1}^T\mid x_k)$, $P_0(y_{k+1}^T\mid x_k)$ в выражение (20) для $P(x_k\mid y_1^T)$ при $k=1,2,\ldots,T$ и учесть условия нормировки $\sum\limits_{j=1}^{N} P(x_k\mid y_1^T)\big|_{x_k=X_j}=1$ для вычисления нормировочных коэффициентов $P(y_{k+1}^T\mid y_1^k)$.

4. АЛГОРИТМ ВЫЧИСЛЕНИЯ АПОСТЕРИОРНЫХ ВЕРОЯТНОСТЕЙ МОМЕНТА ПОЯВЛЕНИЯ СКАЧКА ПАРАМЕТРОВ

Апостериорная вероятность момента появления скачка параметров в момент времени k выражается через ранее найденные функции следующим образом:

$$P_\tau(\tau \mid y_1^T)\big|\tau = k = P(\tau = k \mid y_1^k) = \frac{P(\tau = k, y_{k+1}^T \mid y_1^k)}{P(y_{k+1}^T \mid y_1^k)} = \frac{\sum\limits_{j=1}^{N} P(x_k, \tau = k, y_{k+1}^T \mid y_1^k)\big|x_k = X_j}{P(y_{k+1}^T \mid y_1^k)} . \quad (27)$$

Функции в числителе правой части формулы (27) согласно выражениям (9), (19) записываются в виде

$$P(x_k, \tau = k, y_{k+1}^T \mid y_1^k) = \nu_0(k) p_0(k) P_0(x_k, k) P_1(y_{k+1}^T \mid x_k). \quad (28)$$

Подставляя (28) в (27), получим выражение для апостериорной вероятности момента появления скачка:

$$P_\tau(\tau = k \mid y_1^T) = \frac{\nu_0(k) p_0(k) \sum\limits_{j=1}^{N} P_0(x_k, k) P_1(y_{k+1}^T \mid x_k)\big|x_k = X_j}{P(y_{k+1}^T \mid y_1^k)} , \qquad P(\tau = T \mid y_1^T) = \nu_0(T) p_0(T), \quad (29)$$

где $k = 0, 1, \ldots, T-1$.

Отметим, что функции $p_0(k)$, $P_0(x_k, k)$, $P_1(y_{k+1}^T \mid x_k)$ и $P(y_{k+1}^T \mid y_1^k)$ могут быть вычислены с помощью описанного выше алгоритма. Аналогично, следуя (5), (10), (18) и (19), находятся апостериорные вероятности наличия и отсутствия скачка к моменту времени k:

$$P_\tau(\tau < k \mid y_1^T) = \frac{P(\tau < k, y_{k+1}^T \mid y_1^k)}{P(y_{k+1}^T \mid y_1^k)} = \frac{\sum\limits_{j=1}^{N} P(x_k, k < \tau, y_{k+1}^T \mid y_1^k)\big|x_k = X_j}{P(y_{k+1}^T \mid y_1^k)} =$$
$$= \frac{p_1(k) \sum\limits_{j=1}^{N} P_1(x_k, k) P_1(y_{k+1}^T \mid x_k)\big|x_k = X_j}{P(y_{k+1}^T \mid y_1^k)} , \quad (30)$$

$$P_\tau(\tau > k \mid y_1^T) = \frac{P(\tau > k, y_{k+1}^T \mid y_1^k)}{P(y_{k+1}^T \mid y_1^k)} = \frac{\sum\limits_{j=1}^{N} P(x_k, k > \tau, y_{k+1}^T \mid y_1^k)\big|x_k = X_j}{P(y_{k+1}^T \mid y_1^k)} =$$
$$= \frac{[1 - \nu_0(k)]\, p_0(k) \sum\limits_{j=1}^{N} P_0(x_k, k) P_0(y_{k+1}^T \mid x_k)\big|x_k = X_j}{P(y_{k+1}^T \mid y_1^k)} , \quad (31)$$

где $k = 0, 1, \ldots, T-1$. В конечный момент времени $P_\tau(\tau < T \mid y_1^T) = p_1(T)$, $P_\tau(\tau > T \mid y_1^T) = [1 - \nu_0(T)]\, p_0(T)$.

Апостериорные распределения вероятностей состояния модели $P(x_k \mid y_1^T)$ и момента появления скачка $P_\tau(\tau \mid y_1^T)$ содержат полную информацию об исследуемом процессе с учётом наблюдений. В зависимости от выбираемого критерия оптимальности по этим вероятностям можно находить оценки как момента появления скачка параметров, так и состояний модели, а также исследовать точность формируемых оценок.

5. РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

С целью проверки работоспособности проводилось численное моделирование алгоритма. По реализации наблюдений $\{y_k\}$, где $k = 1, 2, \ldots, T$, оценивалась последовательность скрытых состояний марковской модели и проводилась оценка дискретного времени τ, при котором происходил скачок параметров. Предполагалось, что вся последовательность наблюдений $\{y_k\}$ известна ещё до начала оценивания. Параметры системы до и после скачка считались известными. Априорная вероятность момента появления скачка выбиралась равномерной на всём временно́м интервале наблюдений $T = 100$: $P_\tau(\tau) = 1/T$. При моделировании число N возможных скрытых состояний модели выбиралось равным 2, а число M различных дискретных значений наблюдений считалось равным 4. Для определённости можно считать, что скрытый процесс x_k принимает значения из множества $\{1, 2\}$, а наблюдаемый процесс y_k принимает значения из множества $\{1, 2, 3, 4\}$.

Условные вероятности наблюдений $b_j(Y_m)$, где $j = 1, 2, \ldots, N$, $m = 1, 2, \ldots, M$, до и после скачка выбирались одинаковыми и для всех значений дискретного времени $k = 1, 2, \ldots, T$ описывались элементами матрицы $b_{jm} = b_j(Y_m)$:

$$\mathbf{b} = \begin{pmatrix} 0{,}7 & 0{,}1 & 0{,}1 & 0{,}1 \\ 0{,}1 & 0{,}1 & 0{,}1 & 0{,}7 \end{pmatrix}.$$

Например, если модель находится в состоянии $x_k = 1$, то с вероятностью 0,7 значение наблюдений равно $y_k = 1$ и с равными вероятностями 0,1 наблюдаются другие значения из множества $\{2, 3, 4\}$. Если же модель находится в состоянии $x_k = 2$, то с вероятностью 0,7 наблюдается значение $y_k = 4$ и с равными вероятностями 0,1 реализуются другие значения наблюдений из множества $\{1, 2, 3\}$.

Матрица, составленная из условных вероятностей переходов a_{ij} между скрытыми состояниями, изменялась в момент скачка τ:

$$\mathbf{a} = \begin{cases} \mathbf{a}^0 = \begin{pmatrix} 0{,}95 & 0{,}05 \\ 0{,}05 & 0{,}95 \end{pmatrix}, & 1 \le k \le \tau; \\[2ex] \mathbf{a}^1 = \begin{pmatrix} 0{,}50 & 0{,}50 \\ 0{,}50 & 0{,}50 \end{pmatrix}, & \tau < k \le T. \end{cases}$$

Таким образом, при $1 \le k \le \tau$ вероятность остаться в состоянии $x_k = 1$ была равна 0,95, вероятность перехода из состояния $x_k = 1$ в состояние $x_{k+1} = 2$ выбиралась равной 0,05. В момент времени $k = \tau$ вероятность перехода между состояниями скачкообразно изменялась. Далее при $\tau < k \le T$ вероятности перехода между состояниями были равны 0,5. Момент времени τ, при котором происходил скачок параметров, был равен 50. Моделирование проводилось при условии, что алгоритму «не известны» ни момент скачка параметров, ни последовательность скрытых состояний.

На рис. 1 представлена реализация случайной последовательности скрытых состояний $\{x_k\}$ с вероятностями переходов, описываемыми матрицей $\mathbf{a}$. На рис. 2 представлена реализация наблюдений $\{y_k\}$, описываемая матрицей $\mathbf{b}$ с учётом реализации $\{x_k\}$, где $k = 1, 2, \ldots, T$.

По наблюдаемым реализациям $\{y_k\}$ в результате работы алгоритма формировались вероятности различных значений скрытого марковского процесса $P(x_k \mid y_1^T)$, где $k = 1, 2, \ldots, T$. Например, на рис. 3 приведён график апостериорной вероятности $P(x_k = 2 \mid y_1^T)$ пребывания процесса в состоянии $x_k = 2$, полученный по реализации наблюдений на рис. 2. Апостериорные вероятности $P(x_k \mid y_1^T)$ учитывают всю информацию о процессе x_k и могут быть использованы

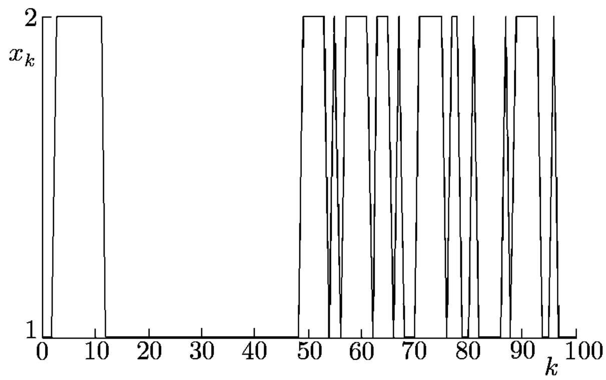


Рис. 1. Случайная последовательность скрытых состояний модели

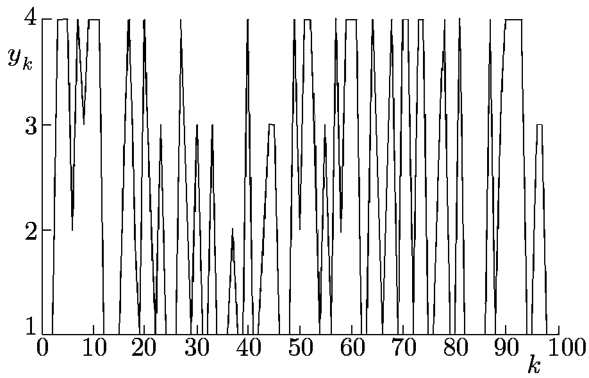


Рис. 2. Случайная последовательность наблюдений, статистически связанная с последовательностью скрытых состояний x_k матрицей $\mathbf{b}$

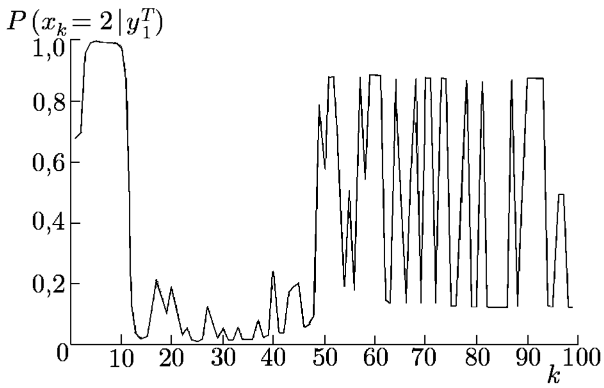


Рис. 3. График апостериорной вероятности состояния $x_k = 2$

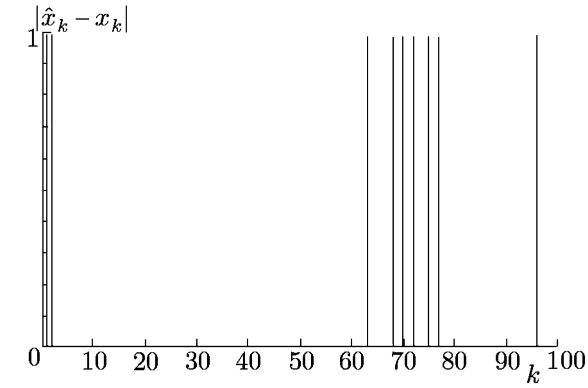


Рис. 4. График ошибок оценок состояний $|\hat{x}_k - x_k|$

для вычисления оптимальных оценок $\hat{x}_k$ в соответствии с различными критериями оптимальности. В рассматриваемом примере выбирался критерий максимума апостериорной вероятности. Поэтому оценка величины x_k считалась равной $\hat{x}_k = 1$, если $P(x_k = 1 \mid y_1^T) \geq 0{,}5$. Если же $P(x_k = 1 \mid y_1^T) < 0{,}5$, то оценка считалась равной $\hat{x}_k = 2$. Полученные таким образом оценки состояний процесса x_k затем сравнивались с истинной скрытой последовательностью. На рис. 4 представлен график ошибок оценок состояний. Каждое ненулевое значение на этом графике соответствует неправильной оценке. Очевидно, точность оценки зависит от параметров задачи и улучшается с ростом точности наблюдений.

Одновременно с оценкой состояний находились и апостериорные вероятности для случайного момента появления скачка. Отметим, что оценки вероятностей для момента появления скачка формируются при работе алгоритма как непосредственно в текущем времени (используя методы фильтрации), так и после анализа всей последовательности наблюдений (используя методы интерполяции).

На рис. 5 показан график вероятности момента появления скачка параметров $P(k = \tau \mid y_1^k)$, полученный непосредственно в текущем времени. На рис. 6 показан график вероятности момента появления скачка параметров $P(k = \tau \mid y_1^T)$, полученный в результате интерполяции по всей последовательности наблюдений. Из графиков видно, что при появлении скачка параметров в

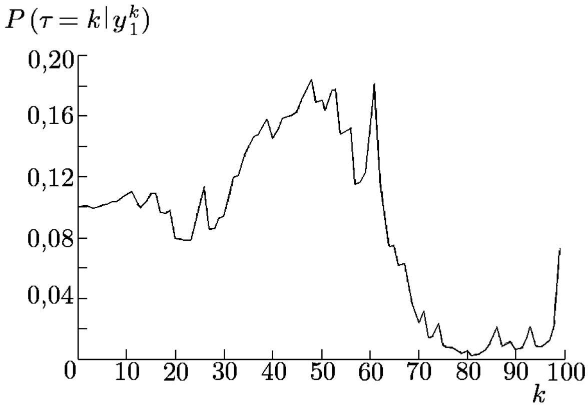


Рис. 5. График оценки вероятности момента появления скачка параметров $P(k = \tau \mid y_1^k)$, полученной непосредственно в текущем времени

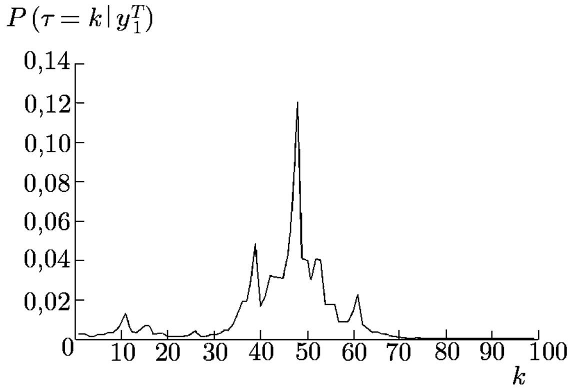


Рис. 6. График оценки вероятности момента появления скачка параметров $P(k = \tau \mid y_1^T)$, полученной после анализа всей реализации последовательности наблюдений

результате обработки наблюдений алгоритмом удаётся сформировать функцию распределения вероятностей $P(\tau \mid y_1^T)$ при $1 \leq \tau \leq 100$, по которой можно находить оптимальные оценки момента появления скачка, используя различные критерии оптимальности. Например, оптимальная по критерию максимума апостериорной вероятности оценка момента появления скачка для рассматриваемого примера равна $\hat{\tau} = 48$, при этом ширина пика апостериорной функции распределения вероятности $P(\tau \mid y_1^T)$ на рис. 6, равная приблизительно 10 шагам дискретного времени, характеризует точность оценивания величины τ.

В заключение отметим, что результаты моделирования продемонстрировали работоспособность и подтвердили высокую точностью оценивания и хорошее быстродействие полученного алгоритма.

Работа выполнена при поддержке Совета при Президенте Российской Федерации по поддержке ведущих научных школ (грант № НШ-1729.2003.2).

СПИСОК ЛИТЕРАТУРЫ

1. Ширяев А. Н. Статистический последовательный анализ. М.: Наука, 1976.
2. Сосулин Ю. Г., Фишман М. М. // Изв. АН СССР. Техническая кибернетика. 1977. № 3. С. 149.
3. Davis M. H. A. // IEEE Transactions on Automatic Control. 1975. V. 20, No. 2. P. 257.
4. Клигене Н., Телькснис Л. // Автоматика и телемеханика. 1983. № 10. С. 5.
5. Никифоров И. В. Последовательное обнаружение изменения свойств случайных процессов. М.: Наука, 1983.
6. Бассвиль М., Вилски А., Банвенист А. и др. Обнаружение изменения свойств сигналов и динамических систем. М.: Мир, 1989.
7. Бухалев В. А. Распознавание, оценивание и управление в системах со случайной скачкообразной структурой. М.: Физматлит, 1996.
8. Forney G. D., Jr. // Proc. IEEE. 1973. V. 61, No. 3. P. 268.
9. Rabiner L. R. // Proc. IEEE. 1989. V. 77, No. 2. P. 257.
10. MacDonald I. L., Zucchini W. Hidden Markov and Other Models for Discrete-Valued Time Series. CRC Press, 1997.

11. Стратонович Р. Л. Условные марковские процессы и их применение к теории оптимального управления. М.: Изд-во МГУ, 1966.
12. Липцер Р. Ш., Ширяев А. Н. Статистика случайных процессов. М.: Наука, 1974.
13. Тихонов В. И., Кульман Н. К. Нелинейная фильтрация и квазикогерентный приём сигналов. М.: Сов. радио, 1975.
14. Силаев А. М. // Автоматика и телемеханика. 1997. № 10. С. 58.
15. Ванжа А. В., Мальцев А. А., Силаев А. М. // Изв. вузов. Радиофизика. 1993. Т. 36, № 6. С. 498.
16. Королёв А. В., Силаев А. М. // Изв. вузов. Радиофизика. 2002. Т. 45, № 3. С. 62.

Поступила в редакцию 1 июля 2004 г.; принята в печать 18 июля 2005 г.

ESTIMATION OF MARKOV SEQUENCES WITH JUMP-LIKE VARIATION OF PARAMETERS BY THE INTERPOLATION METHOD

A. V. Korolev and A. M. Silaev

Using the methods of the theory of optimal nonlinear filtering, we develop an algorithm for obtaining optimal estimates of the sequence of hidden states of discrete-valued Markov processes with jump-like varying parameters at unknown time. Optimal estimates of the states of the Markov processes and the time of appearance of parameter jump are obtained as a result of interpolation by processing the entire observation sequence. The results of simulation of the algorithm work are presented.