МИНИСТЕРСТВО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Известия высших учебных заведений

РАДИОФИЗИКА

ежемесячный научно-технический журнал

Издаётся с апреля 1958 г.

Том XLVII № 2 **Нижний Новгород** **2004**

Содержание

УДК 524+520.27

МОДЕЛИРОВАНИЕ ПЕРЕНОСА ИЗЛУЧЕНИЯ В ЛИНИЯХ HCO^+ И $HC^{18}O^+$ ОБЛАКА L1544

Н. Р. Троицкий[1], А. В. Лапинов[1], С. Н. Замоздра[2]

Исследовано формирование линий излучения HCO^+ и $HC^{18}O^+$ в объекте L1544. Реализован численный расчёт переноса излучения, найдена его интенсивность при различных прицельных параметрах. Показано, что в случае оптически тонкой среды принципиальным для формирования наблюдаемого профиля линии является уменьшение содержания исследуемых молекул к центру объекта. Рассмотрены различные гидродинамические модели, учитывающие сжатие, турбулентность и возможные типы вращения облака, среди них выбрана обеспечивающая наилучшее соответствие наблюдательным данным. Численно показано, что в L1544 существует дифференциальное вращение. Проведены оценки плотности и скорости движения среды в рамках выбранной модели в предположении сферической симметрии объекта.

ВВЕДЕНИЕ

В спектрах межзвёздных облаков наблюдаются линии излучения большого числа молекул. Нередко на профиле этих линий имеется провал, который может указывать на некоторые особенности облака. В данной работе проверяется модель образования подобного провала, а также оцениваются распределение концентрации, поле скоростей и наличие вращения облака, а также относительное содержание молекул. В качестве объекта исследования выбрано облако L1544, наблюдавшееся в линиях J = 1–0 молекул HCO^+ [1] и изотопа $HC^{18}O^+$ [2]. Центральная область источника находится, по-видимому, на самой ранней стадии протозвёздного коллапса. На это указывает и ряд наблюдений в инфракрасном диапазоне [3], которые показали, что облако L1544 имеет плотное пылевое ядро. Детальная морфология L1544 приведена в работах [1, 2].

Как оказалось, наблюдаемые профили линий могут формироваться не только при сжатии облака, но и при наличии в нём дифференциального вращения. Образование описанной выше особенности в случае оптически толстой среды может быть объяснено поглощением в оболочке. Если оболочка облака разреженная и протяжённая, все молекулы находятся на нижних энергетических уровнях и, следовательно, поглощают излучение с переходом на более высокие уровни. В некоторых случаях провал наблюдается и в оптически тонких линиях. Это объясняется тем, что при больших концентрациях и низких температурах в ядре облака происходит вымораживание молекул на пыли, поэтому подавляющая часть излучения формируется не в ядре, а в оболочке облака. В этом случае на центральных лучевых скоростях, т. е. в области ядра, молекул очень мало, следовательно, их излучение менее интенсивное. Таким образом на профиле линии формируется провал.

1. РЕШЕНИЕ ЗАДАЧИ ПЕРЕНОСА ИЗЛУЧЕНИЯ

Для излучения молекул облака можно записать уравнение переноса, которое будет иметь вид

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}\tau_\nu} = -I_\nu + S_\nu,$$

где I_ν — интенсивность излучения на частоте ν, S_ν — функция источников, равная отношению коэффициента излучения к коэффициенту поглощения, τ_ν — оптическая толщина на частоте ν.

Аналитическое решение данного уравнения представлено, например, в [4]. В нашем случае аналитическое решение не представляется возможным, поэтому задача решалась численно методом Монте-Карло [5, 6]. Этот итерационный метод позволяет решать задачу переноса излучения в линиях в многоуровневой системе. Критерием найденного решения является получение устойчивых, мало меняющихся от итерации к итерации значений n_{km} — населённостей на уровне k в слое m. Основная идея метода состоит в том, что все фотоны, испущенные за единицу времени, рассматриваются как некоторое число модельных фотонов, каждый из которых представляет собой соответствующее количество реальных. В процессе прохождения сквозь облако статистический вес модельных фотонов на каждом шаге меняется из-за поглощения и стимулированного излучения. Все произошедшие акты поглощения приводят к изменению населённости уровней, расчёт которой является конечной целью. Процедура повторяется до получения устойчивых значений населённостей уровней, после чего из них находится интенсивность линий.

Фотоны излучаются благодаря спонтанным переходам с верхнего уровня u на нижний уровень l. Число N_{ph} таких излучений за единицу времени в единице объёма равно $N_{\mathrm{ph}} = n_{um}A_{ul}$, где A_{ul} — коэффициент Эйнштейна для спонтанного перехода с уровня u на уровень l. Исследуемая область (в нашем случае сферическая) подразделяется на некоторое число слоёв, в которых задаётся своя кинетическая температура и концентрация газа, относительное содержание исследуемых молекул и турбулентная скорость. Испущенный модельный фотон имеет случайную координату внутри слоя. Число таких фотонов внутри слоя m может быть найдено из локального значения N_{ph} и рассчитанного на предыдущем шаге значения n_{um}. Таким образом, мы получаем вес W_0 для фотонов, испускаемых за одну секунду в слое m. Каждый модельный фотон внутри слоя излучается в случайном направлении, задаваемым единичным вектором $\mathbf{n}$, и имеет случайную отстройку $\Delta\nu$ от центральной частоты перехода ν_{ul}. Излучение предполагается изотропным, и частотный профиль может быть представлен в доплеровском виде:

$$\phi(\nu) = (\delta\sqrt{\pi}\,)^{-1} \exp\left[-\left(\frac{\nu - \nu_{ul} - \mathbf{V}\mathbf{n}\,\nu_{ul}/c}{\delta}\right)^2\right],$$

где c — скорость света в вакууме, δ — доплеровская ширина, которая определяется кинетической температурой и микротурбулентной скоростью. Излучённый фотон проходит короткий путь s в заданном направлении, после чего координаты нового положения запоминаются. Оптическая толщина вдоль пути может быть представлена в виде

$$\tau_\nu = \frac{h\nu_{ul}}{4\pi}\,\phi(\nu)\,(n_{lm}B_{lu} - n_{um}B_{ul})\,s,$$

где n_{lm} и n_{um} — населённости нижнего и верхнего уровней в слое m, h — постоянная Планка, B_{ul} и B_{lu} — коэффициенты Эйнштейна для стимулированного излучения и поглощения. Предполагается, что на небольшом шаге s профили поглощения и излучения совпадают. Вес модельного фотона зависит от пройденной дистанции x внутри слоя следующим образом: $W(x) = W_0\exp(-\tau_\nu x/s)$. На пути от $x = 0$ до $x = s$ модельный фотон индуцирует излучение и поглощение. Полное число актов поглощения за время прохождения фотоном данного шага составляет

$$N_{lu} = \frac{h\nu_{ul}}{4\pi}\,\phi(\nu) n_{lm} B_{lu} \int\limits_0^s W(x)\,\mathrm{d}x.$$

Тогда число актов поглощения, приходящееся на одну молекулу, представляется в виде

$$S_{lu\,m} = \frac{N_{lu}}{n_{lm}V_m} = \frac{h\nu_{ul}}{4\pi}\,\phi(\nu) B_{lu}\,\frac{sW_0}{V_m\tau_\nu}\,[1 - \exp(-\tau_\nu)],$$

где V_m — объём слоя m. После этого делается новый шаг в том же направлении, и все вышеприведённые вычисления повторяются. Таким образом, после k-го шага можно написать

$$S_{lu\,m} = \frac{h\nu_{ul}s_k}{4\pi V_m \tau_{\nu k}}\,\phi(\nu)B_{lu}\,[1-\exp(-\tau_{\nu k})]\,W_0 \exp\left(-\sum_{i=1}^{k-1}\tau_{\nu i}\right). \tag{1}$$

Длина шага выбирается каждый раз таким образом, чтобы коэффициент поглощения из-за эффекта Доплера менялся не более чем на 6 % [7]. Расчёты ведутся до тех пор, пока модельный фотон не покинет облако или пока его вес не станет пренебрежимо мал. После этого излучается новый модельный фотон и т. д. После того, как все модельные фотоны испущены и их путь прослежен, можно определить населённость уровней (это делается в каждом слое) из уравнения статистического равновесия:

$$n_{lm}\left(\sum_{k<l}A_{lk} + \sum_{k\neq l}[S_{lk\,m} + C_{lk}]\right) = \sum_{k\neq l} n_{km}\left(\frac{g_l}{g_k}\,S_{lk\,m} + C_{kl}\right) + \sum_{k>l} n_{km}A_{kl}, \tag{2}$$

где $m = 1, 2, \ldots, N_{\rm shell}$, C_{lk} и C_{kl} — коэффициенты Эйнштейна для столкновительных переходов с l-го уровня на k-й и с k-го на l-й соответственно, $N_{\rm shell}$ — число слоёв в модели. В уравнении учтено соотношение $B_{kl}g_k = B_{lk}g_l$, где g_k и g_l — статистические веса k-го и l-го уровней. Эта система уравнений может быть решена с использованием дополнительного условия $\sum_k n_{km} = n_0$, где n_0 — концентрация молекул в слое m. После этого находятся $N_{\rm ph}$ через новые значения n_{km}. Теперь процесс может быть повторён, т. е. испущена новая серия модельных фотонов. Перед этим следует положить равным нулю $S_{lu\,m}$, но для уменьшения погрешности, связанной с модельным шумом, это значение копится несколько итераций. При этом в системе (2) нужно заменить $S_{lu\,m}$ на $\sum^{N_{\rm iter}} S_{lu\,m}/N_{\rm iter}$, где $N_{\rm iter}$ — число итераций до обнуления числа актов поглощения. После расчёта населённостей уровней можно посчитать интенсивность излучения:

$$I_{\rm em}(\nu) = I_{\rm bg}(\nu)\exp[-\tau_\nu(X)] + \int_0^X j(\nu,x)\exp[-\tau_\nu(x)]\;{\rm d}x,$$

где $j(\nu,x) = h\nu_{ul}/(4\pi)\,\phi(\nu)n_{um}A_{ul}$, $I_{\rm bg}(\nu)$ — интенсивность внешнего излучения, X — координата выхода луча из слоя.

За основу была взята оригинальная программа Бернеса [5], написанная для молекулы CO. На её основе была написана программа для HCO^+, в которой присутствуют такие компоненты, как свёртка с антенной функцией телескопа, что позволяет сравнить расчёты с наблюдениями на конкретных антеннах, и возможность задания переменных параметров объекта (скорости и плотности среды, а также содержания исследуемых молекул), зависящих от радиуса. Возможности данной программы не ограничиваются одной молекулой HCO^+ и заданной моделью облака.

2. МОДЕЛИ ЭВОЛЮЦИИ ОБЛАКА

Облако представляется в виде шара из слабо ионизованной плазмы с центрально-симметричным распределением концентрации. Шар пронизан магнитным полем с осевой симметрией. Самогравитация стремится сжать облако в точку, но магнитное натяжение, а также тепловое, магнитное и турбулентное давление замедляют этот процесс или даже приводят к вре́менной смене сжатия расширением. Кинетическая температура газа считается постоянной. Это оправдано тем, что

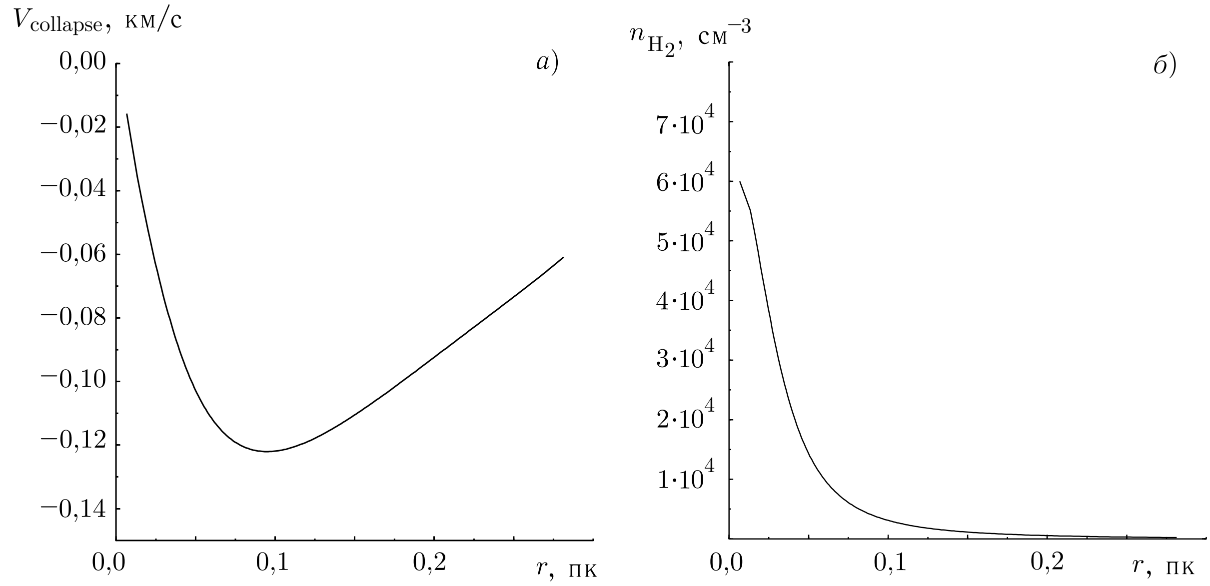


Рис. 1. Распределения регулярной радиальной скорости (*а*) и концентрации молекулярного водорода (*б*) в зависимости от расстояния r от центра объекта согласно модели Ли

области излучения исследуемых молекул малы по сравнению с полным размером облака. Рассмотрение в данных условиях более сложной модели, чем изотермическая, представляется нецелесообразным в силу незначительного изменения температуры на размерах излучающей области. Центробежной силой пренебрегаем вследствие медленности вращения облака, однако при расчёте профилей линий вращение учитывается. Для простоты считается, что при сжатии или расширении облака магнитное поле не нарушает центральной симметрии концентрации.

Эволюция облака описывается в рамках магнитогидродинамического (МГД) приближения с учётом магнитной амбиполярной диффузии. Используются две численные модели: модель Ли [8], а также авторская модель [9], основанная на подходе Дудорова и Сазонова [10, 11]. Модель Ли позволяет рассчитать зависимости концентрации водорода и крупномасштабной радиальной скорости от радиуса (см. рис. 1), но не даёт информацию о дисперсии турбулентной скорости $\sigma_{\rm t}$. В нашей модели этот недостаток устранён: $\sigma_{\rm t}$ расчитывается с помощью статистической модели МГД турбулентности [9]. Предполагается, что турбулентность обусловлена преимущественно альвеновскими волнами, поэтому плотность среды ρ не возмущается и, кроме того, плотности кинетической и магнитной энергии флуктуаций приблизительно равны. Следовательно, $\sigma_{\rm t} = \sqrt{\epsilon_{\rm t}/\rho}$, где $\epsilon_{\rm t}$ — плотность кинетической энергии флуктуаций. Уравнение эволюции $\epsilon_{\rm t}$ имеет вид

$$\frac{{\rm d}\epsilon_{\rm t}}{{\rm d}t} = -\frac{3}{2}\,\epsilon_{\rm t}\,\nabla\mathbf{U} - s\,\nabla\mathbf{v}_{\rm a}\,\epsilon_{\rm t} - \left(\frac{2\pi v_{\rm a}}{\lambda}\right)^2 t_{\rm ni}\epsilon_{\rm t}\,, \tag{3}$$

где ${\rm d}/{\rm d}t$ — субстанциональная производная, $\mathbf{U}$ — крупномасштабная скорость среды, $\mathbf{v}_a$ — альвеновская скорость, $t_{\rm ni}$ — время замедления нейтральных частиц при столкновениях с ионами, $s = 1$, если волны бегут в направлении среднего магнитного поля $\mathbf{H}$, и $s = -1$ в противоположном случае, λ — длина волн, дающих наибольший вклад в $\epsilon_{\rm t}$. Первый член в правой части (3) описывает изменение $\epsilon_{\rm t}$ вследствие крупномасштабного сжатия/расширения среды, второй член — перераспределение $\epsilon_{\rm t}$ из-за распространения волн относительно среды, третий член — диссипацию флуктуаций в ходе магнитной амбиполярной диффузии. Уравнение (3) отличается от аналогичного уравнения из работы [12] тем, что учитывает диссипацию не символично, а в конкретном

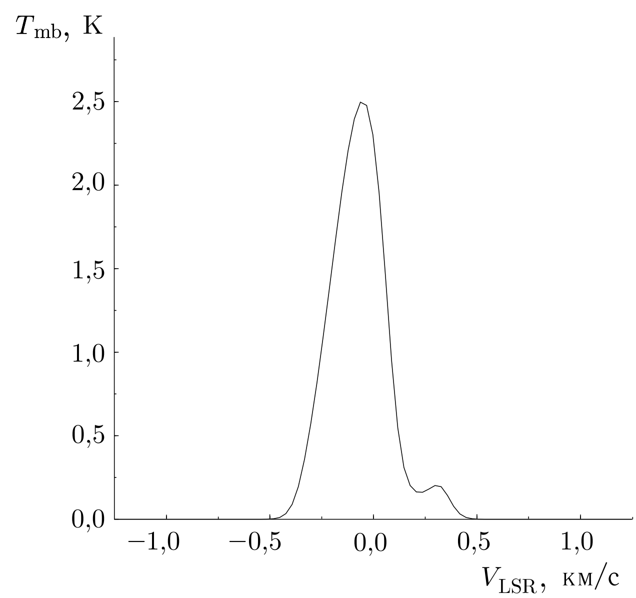


Рис. 2. Профиль линии J = 1–0 молекулы HCO^+ согласно основной модели. Здесь T_{mb} — результат свёртки яркостной температуры источника с диаграммой направленности телескопа, V_{LSR} — доплеровская скорость относительно местного стандарта отсчёта

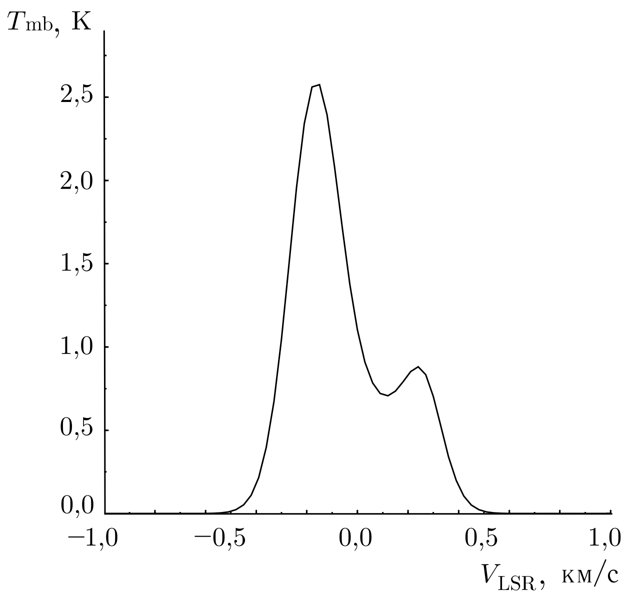


Рис. 3. Профиль линии J = 1–0 молекулы HCO^+ согласно модели Ли

виде, пригодном для расчётов. Для определения λ мы ввели дополнительное уравнение:

$$\frac{\partial\lambda}{\partial t} + v_{\mathrm{p}z}\,\frac{\partial\lambda}{\partial z} = \lambda\,\frac{\partial v_{\mathrm{p}z}}{\partial z}\,, \tag{4}$$

где $v_{\mathrm{p}z}$ — направленная вдоль магнитного поля компонента фазовой скорости (скорость течения плюс/минус альвеновская скорость; ось z сонаправлена с $\mathbf{H}$).

Уравнения (3), (4) решаются численно методом Лакса—Вендроффа совместно с уравнениями для крупномасштабных величин [10, 11]. Предполагается, что на границе облака $\epsilon_{\mathrm{t}} = \mathrm{const}$ и период альвеновских колебаний тоже постоянен. Считается, что турбулентность влияет на эволюцию облака через градиент турбулентного давления, которое в случае альвеновской турбулентности равно $\epsilon_{\mathrm{t}}/2$ [13].

Содержание молекул HCO^+ рассчитывается с помощью системы уравнений для псевдореакций [14].

3. РЕЗУЛЬТАТЫ

На основе метода Монте-Карло были проведены расчёты нескольких вариантов эволюции облака в моделях Ли [8, 15] и Замоздры (основная модель). На рис. 2 и 3 представлены расчёты излучения молекул HCO^+ в рамках указанных моделей. Для получения результатов, наиболее близких к наблюдаемым, в модель Ли была добавлена постоянная турбулентная скорость. Расчёты для модели Ли велись при микротурбулентной скорости 0,15 км/с, относительном содержании молекул HCO^+ $2\cdot 10^{-10}$ и концентрации водорода в центре облака 10^5 см$^{-3}$. Для основной модели концентрация водорода в центре облака составляла $6\cdot 10^5$ см$^{-3}$, относительное содержание молекул HCO^+ в центре облака — $5{,}2\cdot 10^{-10}$, микротурбулентная скорость

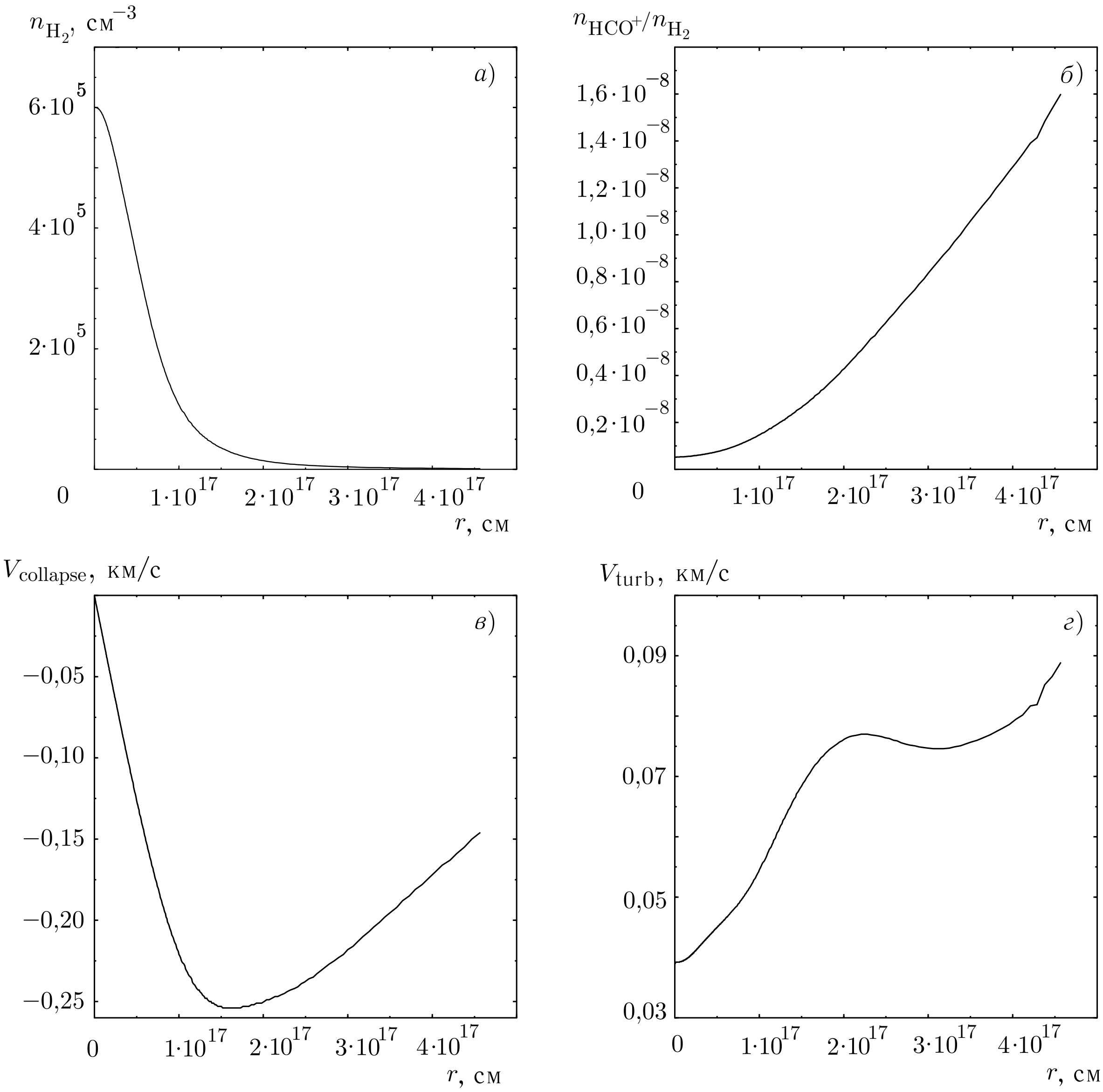


Рис. 4. На рисунке представлено распределение концентрации водорода n_{H_2} (*а*), относительного содержания молекул HCO^+ (*б*), регулярной радиальной $V_{\rm collapse}$ (*в*) и турбулентной $V_{\rm turb}$ (*г*) скоростей для основной модели Замоздры

рассчитывалась согласно модели (см. рис. 4). Для обеих моделей кинетическая температура принималась равной 10 K и учитывалось влияние реликтового фона на возбуждение молекул.

Как видно из сравнения рассчитанных профилей линий, характерный провал интенсивности можно получить и в более простой модели (в модели Ли с постоянным содержанием HCO^+). Однако невозможно получить наблюдаемый профиль для изотопа $HC^{18}O^+$, что налагает определённые требования на распределение его концентрации в облаке. Поскольку в нашей модели принималось, что концентрация изотопа $HC^{18}O^+$ убывает к центру облака (так же, как и концентрация основного элемента HCO^+), можно утверждать, что данный фактор является принципиальным для формирования провала на профиле линий $HC^{18}O^+$.

В предположении сферической симметрии облака результаты расчётов профилей линий некоторых переходов в молекуле HCO^+ для основной модели Замоздры представлены на рис. 5.

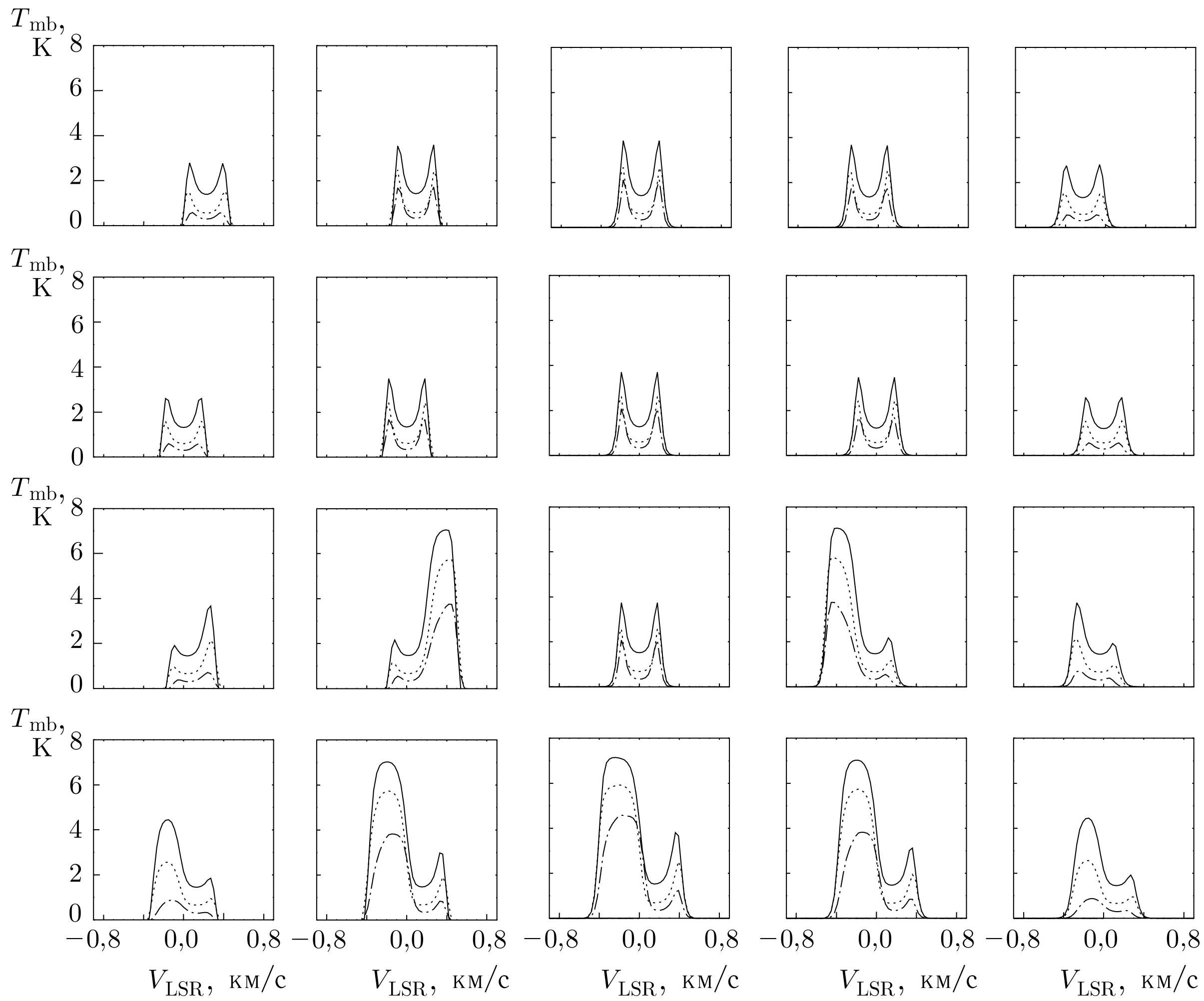


Рис. 5. Расчётные профили линий для переходов J = 3–2, J = 2–1, J = 1–0 молекулы HCO^{+} в облаке L1544 в рамках основной модели

Сплошной кривой показан профиль линии перехода J = 1–0, коротким пунктиром — перехода J = 2–1, длинным штрих-пунктиром — перехода J = 3–2. Профили в каждой горизонтальной строке рассчитаны при фиксированных гидродинамических параметрах облака, но при различных прицельных параметрах: слева направо −50 %, −20 %, 0 %, 20 %, 50 % радиуса облака соответственно, который составляет $4 \cdot 10^{17}$ см $\equiv$ 0,148 пк. Нижняя строка представляет собой профили, рассчитанные при наличии только коллапса, вторая строка снизу — при наличии вращения, третья — при отсутствии каких-либо регулярных (не турбулентных) движений, верхняя — при наличии твердотельного вращения (при постоянном значении угловой скорости). Как хорошо видно из рис. 5, твердотельное вращение приводит лишь к сдвигу профиля линии по скорости (частоте) без изменения его формы. Профиль линии излучения молекулы в коллапсирующем облаке очень похож на профиль при кеплеровском дифференциальном вращении. В то же время при дифференциальном вращении мы имеем различную ассиметрию линий при разных знаках прицельного параметра, что наилучшим образом согласуется с имеющимися наблюдениями. Данные профили рассчитывались без учёта свёртки с диаграммой направленности телескопа, что может сказаться на интенсивности линий, но никак не отражается на типе асимметрии. Таким образом, мы подтвердили предположение о вероятном вращении облака L1544. Параметры облака, при которых велись расчёты, следующие: концентрация водорода в центре облака 10^{6} см$^{-3}$, концентрация молекул HCO^+ в центре облака $5{,}22 \cdot 10^{-4}$ см$^{-3}$, скорость турбулентного движения изменяется от 0,03 до 0,09 км/с.

При расчётах линии $J = 1-0$ молекулы $HC^{18}O^+$ (см. рис. 6) использовалась основная модель с предположением, что содержание изотопа гораздо меньше, чем HCO^+ (примерно в 500 раз, как в земных условиях). Таким образом, логично предположить, что для данного изотопа оптическая толщина облака будет мала. Это подтверждается расчётами. Сравнение расчётных линий с наблюдаемыми [2] показало, что наиболее близкий к наблюдаемому профиль реализуется при относительном содержании $HC^{18}O^+$ не в 500, как в земных условиях, а в 300 раз меньшем, чем содержание HCO^+. Для случая оптически тонкой среды, можно сказать, что для формирования такого профиля необходима концентрация молекулярного водорода в центре облака не менее 10^6 см$^{-3}$ и молекул $HC^{18}O^+$ порядка $8 \cdot 10^{-6}$ см$^{-3}$. Для модели Ли при столь низких оптических толщинах не удалось получить характерного провала на профиле линии. Таким образом, возможные параметры облака L1544 найдены. Данный набор параметров не является единственным, но другие наборы параметров не являются характерными для данного вида объектов.

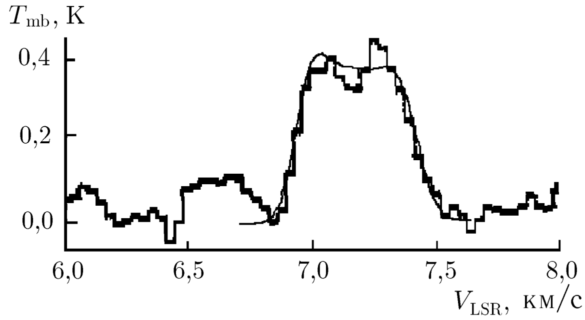


Рис. 6. Сравнение рассчитанного профиля линии $J = 1–0$ молекулы $HC^{18}O^+$ в модели со спаданием концентрации к центру (тонкая кривая) с наблюдениями [2]

4. ЗАКЛЮЧЕНИЕ

Проведено численное моделирование переноса излучения в линиях молекул HCO^+, а также изотопа $HC^{18}O^+$ на ранних стадиях формирования протозвёзд. Это позволило выполнить оценки физических параметров изучаемого объекта. Для облака L1544 концентрация водорода в центре объекта составляет $6 \cdot 10^5$ см$^{-3}$, относительное содержание молекул HCO^+ в центре облака равно $5{,}22 \cdot 10^{-10}$, скорость турбулентного движения изменяется в пределах от 0,03 до 0,09 км/с. Из сравнения профилей линий излучения, рассчитанных при различных гидродинамических моделях эволюции облака, для объекта L1544 следует, что основная модель Замоздры даёт наиболее близкий результат по форме и интенсивности линий. Показано, что важным фактором при формировании наблюдаемых профилей линий в оптически тонкой среде является уменьшение относительного содержания молекул $HC^{18}O^+$ к центру облака. Расчётами подтверждена возможность дифференциального вращения облака.

Авторы признательны рецензенту за полезные замечания и выражают благодарность А. В. Серберу за внимание к работе, а также И. И. Зинченко за полезные обсуждения.

Работа выполнена при поддержке РФФИ (проекты № 03–02–16307, 02–02–17642), программы по поддержке ведущих научных школ (грант № НШ–1483.2003.2) и INTAS (грант № 99.16.67).

СПИСОК ЛИТЕРАТУРЫ

1. Taffalla M., Mardones D., Myers P. C., et al. // Astrophys. J. 1998. V. 504. P. 900.
2. Caselli P., Walmsley C. M., Zucconi A., et al. // Astrophys. J. 2002. V. 565. P. 331.
3. Ward-Thomson D., Kirk J. M., Crutcher R. M., et al. // Astrophys. J. 2000. V. 537. P. L135.
4. Соболев В. В. Курс теоретической астрофизики. М.: Наука, 1979.

5. Bernes C. // Astron. Astrophys. 1979. V. 73. P. 67.
6. Hogergeijde M. R., van der Tak F. F. S. // Asron. Astrophys. 2000. V. 362. P. 697.
7. Соболев А. М. // Научные информации. 1982. вып. 50.
8. Li Z.-Y. // Astrophys. J. 1998. V. 493. P. 230.
9. Дудоров А. Е., Жилкин А. Г., Замоздра С. Н. // Аннотированный отчёт о НИР по гранту ГБ-73. Челябинск, 2002. 81 с.
10. Дудоров А. Е., Сазонов Ю. В. // Научные информации астрон. совета АН СССР. 1981. Т. 49. С. 114.
11. Дудоров А. Е., Сазонов Ю. В. // Научные информации астрон. совета АН СССР. 1987. Т. 63. С. 68.
12. McKee C. F., Zweibel E. G. // Astrophys. J. 1995. V. 440. P. 686.
13. Dewar R. L. // Physics of Fluids. 1970. V. 13. P. 2 710.
14. Ciolek G. E., Mouschovias T. Ch. // Astrophys. J. 1998. V. 504. P. 280.
15. Li Z.-Y. Private communication.
16. Monteiro T. S. // Mon. Not. R. Astron. Soc. 1985. V. 214. P. 419.
17. Flower D. R. // Mon. Not. R. Astron. Soc. 1999. V. 305. P. 651.

[1] Институт прикладной физики РАН, г. Нижний Новгород;
[2] Челябинский госуниверситет, г. Челябинск, Россия

Поступила в редакцию
14 апреля 2003 г.

RADIATIVE-TRANSFER MODELLING OF THE CLOUD L1544 IN THE EMISSION LINES HCO^+ AND $HC^{18}O^+$

N. R. Troitsky, A. V. Lapinov, and S. N. Zamozdra

We study the formation of the emission lines HCO^+ and $HC^{18}O^+$ in the cloud L1544. The radiative transfer is modelled numerically, and the line intensities for different impact parameters are calculated. It is shown that a decrease in the abundance of the studied molecules toward the object center Proved that in optically thing case decreasing of fractional abundance to the center of the object is very important. We consider various hydrodynamical models allowing for compression, turbulence, and different types of cloud rotation and find a model ensuring the best agreement to the observational data. Our numerical simulations are indicative of a nonuniform rotation in L1544. Assuming that the object is spherically symmetric, we estimate the density and velocity of the medium within the framework of the selected model.

УДК 52-77+621.391:53.08+520.86+004.93′1

ТОМОГРАФИЯ ПРИ ОГРАНИЧЕННОМ ЧИСЛЕ ПРОЕКЦИЙ. I. РАДИОАСТРОНОМИЧЕСКИЙ ПОДХОД К ПРОБЛЕМЕ И МЕТОД 2-CLEAN DSA

М. И. Агафонов

Развивается радиоастрономический подход к решению проблемы малоракурсной томографии. Показано, что эффективным способом решения задачи является метод 2-CLEAN DSA, предложенный для определения области допустимых решений. В его основе лежит решение задачи деконволюции с использованием синтезированного луча или синтезированной функции Грина. Устранение искажений томограммы, вызванных откликами от боковых лепестков синтезированной передаточной функции, выполняется на основе двух реализаций итерационного алгоритма с нелинейными ограничениями, известного в радиоастрономии под названием «чистка». Предложенный метод позволяет приблизительно в 10 раз уменьшить необходимое для двумерной реконструкции число проекций в сравнении с традиционным томографическим подходом при условии восстановления широкого, ограниченного лишь в верхней части спектра пространственных частот. Метод легко адаптируется к введению дополнительных ограничений. Приведены примеры реконструкции в астротомографии. Показана перспективность применения метода для широкого круга дистанционных исследований. Проводится сопоставление метода с другими известными способами реконструкции. Перечислены работы радиоастрономов, внёсшие существенный вклад в развитие компонентов метода.

ВВЕДЕНИЕ

Задача успешной реконструкции внутренней структуры объектов, включающих не только точечные компоненты, но и протяжённые области различных размеров, стоит во многих областях: при исследованиях плазмы, в медицине, при построении МГД двигателей, в токамаках, при реконструкции распределений яркости в астрономии при лунных покрытиях, при бесконтактном мониторинге различных объектов, при исследованиях антеннами с ножевыми диаграммами направленности и щелевыми оптическими системами. В 1978 году в докладе [1] были сформулированы базовые проблемы и принципы томографии, основными из которых являются: теорема индетерминантности [2], необходимость учёта физического характера решаемой задачи, требование приступать к конструированию прибора лишь после того, как алгоритм создан и проверен. Последнее, к сожалению, ранее почти никогда не выполнялось в медицине, и вследствие большого числа проекций доза облучения была значительной. В 1977 году математическим путём была доказана теорема индетерминантности — фундаментальное ограничение, вырожденный случай при восстановлении структуры объектов по проекциям, который, однако, указывает не на целесообразность увеличения их числа, а на необходимость использования различной априорной информации для ограничения области решений конкретной задачи реконструкции.

Со времени доклада [1] существенный вклад в решение задачи томографии при ограниченном числе проекций внесла радиоастрономия. В настоящей работе изложен радиоастрономический подход к указанной проблеме. Показано, что эффективным способом решения задачи является метод 2-CLEAN DSA, предложенный для определения области допустимых решений. В хронологическом порядке перечислены работы различных исследователей, посвящённые решению проблемы, показана роль метода 2-CLEAN DSA в приложении к экспериментальным задачам и его место в сравнении с другими известными методами. Кроме того, указано на ряд принципиальных моментов, без учёта которых реконструкция при ограниченном числе проекций не может быть успешной. В работе кратко изложены принципы решения проблемы и приведены примеры практической реализации метода.

1. ПОДХОД К ПРОБЛЕМЕ РЕКОНСТРУКЦИИ

Рассматриваемая задача соответствует реконструкции распределения яркости двумерного (2D) изображения или сечения оптически тонкого эмиссионного объекта по одномерным (1D) проекциям, полученным на основе линейного перемещения сканирующего луча для каждого из используемых позиционных углов. Набор сечений с необходимым шагом позволяет восстановить трёхмерное строение объекта. Исследованиям вопросов, возникающих для оптически толстых объектов, искажениям лучевого хода в окружающих их средах, в объектах с дифракцией, а также проблемам, связанным с переходом к трансмиссионной схеме, посвящено большое число работ, и здесь эти вопросы не обсуждаются. По сходной причине мы не касаемся в настоящей работе фазовой проблемы и вопросов когерентного излучения.

Используем следующую терминологию: реконструкция распределения яркости подразумевает восстановление пространственных частот в непрерывном двумерном спектре с полосой от нуля до верхней граничной частоты ω_{b}; реконструкция изображений допускает введение в этом спектре различных спектральных ограничений $\Delta\omega_{\mathrm{i}}$ в указанной полосе.

Классическое решение получено в 1967 году Брейсуэллом и Риддлом [3]. Ими показано, что количество проекций, необходимое для реконструкции распределения яркости на всех пространственных частотах в круге радиуса ω_{b} на плоскости uv составляет $N \geq \pi D/\varphi$, где $\varphi = 1/\omega_{\mathrm{b}}$ — желаемое угловое разрешение, D — диаметр объекта. Число проекций $N_{\mathrm{BR}} = \pi D/\varphi$ назовём числом Брейсуэлла—Риддла. Неполное заполнение области пространственных частот в пределах окружности радиуса ω_{b} требует экстраполяции решения, т. е. нелинейных методов обработки.

Томографический подход к задаче восстановления распределения яркости двумерного изображения, как правило, сводится к нескольким основным вычислительным схемам (см., например, [4, 5]): фурье-синтезу, суммированию фильтрованных обратных проекций и фильтрации суммарного изображения. На методе интегральных преобразований Радона основывается большинство разнообразных алгоритмов томографии, которые достаточно подробно описаны в [6]. Однако все эти алгоритмы, как правило, требуют большого числа проекций, слабо противодействуют искажениям томограммы при уменьшении числа проекций и достаточно сложны в вычислительном отношении. Более привлекательными являются итерационные алгоритмы с нелинейными ограничениями [7, 8], введение в схему которых различной априорной информации не представляет существенных затруднений.

Перечислим основные отличительные признаки радиоастрономического подхода, основанного на решении уравнения свёртки. Первый признак — это введение синтезированной функции Грина, аналога синтезированного луча в радиоастрономии, или суммарной функции отклика на точечный источник и соответствующей передаточной функции. Синтезированная функция имеет размерность выше, чем реальная приёмная система, с помощью которой регистрируются проекции. Благодаря этому задачу можно свести к задаче обращения свёртки, или деконволюции реального изображения с синтезированной функцией Грина. Если проекции равномерно распределены в пространстве и число их так велико, что обеспечивает требуемое разрешение, то синтезированная функция Грина отлична от нуля только в небольшой области, характеризующей разрешение. При ограниченном числе проекций она отлична от нуля также на расходящихся от центра направлениях, т. е. имеет так называемые боковики (боковые лепестки синтезированного луча). Отклики от боковых лепестков вносят искажения (артефакты) в суммарное изображение (томограмму). Вторым отличительным признаком радиоастрономического подхода является процедура «чистки» (CLEAN) для исключения искажений, вызванных откликами от боковиков синтезированного луча. Алгоритмы чистки — это реализации итерационного алгоритма с нелинейными ограничениями, разработанные для решения радиоастрономических задач. Третьим признаком является разработанный нами метод определения области допустимых решений [9–12] для сложных случаев, реализуемый при помощи двух алгоритмов чистки, 2-CLEAN DSA (от английского Determination of the permissible Solutions Area).

В процессе развития радиоастрономического подхода было выполнено исследование сходимости [9, 13] алгоритмов чистки ST-CLEAN (Standard CLEAN) [14] и TC-CLEAN (Trim Contour CLEAN) [15] для томографической задачи с ограниченным набором проекций. Кроме того, существовала необходимость предварительного исследования поведения алгоритма TC-CLEAN в различных практических ситуациях, на что было указано в [16]. Имелась также некоторая неопределённость с выбором усиления λ даже для алгоритма ST-CLEAN. Основные положения томографической задачи и геометрия синтезированного луча существенно отличались от соответствующих параметров радиоастрономических задач апертурного синтеза. Наличие расходящихся лучей у синтезированной передаточной функции в томографической задаче означает, что внутренняя структура двумерного объекта исследуется не игольчатым лучом, а лучом, в котором уровень боковых лепестков тем больше, чем меньше имеется проекций, а следовательно, и ножевых диаграмм. Вследствие этого возникают искажения томограммы, которые при числе проекций $N \geq N_{\mathrm{BR}}$ становятся незначительными, однако с уменьшением количества проекций существенно возрастают. Из геометрии синтезированного луча (см. рис. 1) на плоскости xy следует, что уровень боковых лепестков для четырёх проекций составляет 0,25 от амплитуды пика в центральной части. Это объясняется тем, что при суммировании четырёх ножевых диаграмм их интенсивность в месте пересечения возрастает в 4 раза. Заметим, например, что при диаметре объекта $D \sim 360''$ и требуемом угловом разрешении примерно $28''$ количество проекций $N_{\mathrm{BR}} = 40$. При этом максимальная интенсивность боковых лепестков синтезированного луча спадает до уровня 0,025, и искажения на суммарном изображении незначительны. При компьютерном моделировании [13] была показана возможность восстановления распределения яркости объектов в случае, когда количество проекций $N \sim 0{,}1 N_{\mathrm{BR}}$. Реконструкция выполнялась при осложняющих обстоятельствах: углы наблюдения располагались лишь в секторе 77° при незаполненном секторе 103° из возможных 180°. Принципиальным является то обстоятельство, что радиоастрономический подход к реконструкции позволяет получить решение в виде полного двумерного набора пространственных частот без введения априорных спектральных ограничений. Наличие информации о какой-либо части спектра решения способствует либо улучшению качества реконструкции, либо смягчает требования к необходимому числу проекций.

Радиоастрономический подход рассматривает проблему как решение уравнения свёртки:

$$g = h * f + n, \tag{1}$$

где $f(x,y)$ — распределение яркости объекта, $h(x,y)$ — синтезированный луч из ножевых диаграмм, $g(x,y)$ — суммарное изображение объекта, $n(x,y)$ — интенсивность шума.

Поскольку уравнение (1) является уравнением Фредгольма 1-го рода типа свёртки, то требуется найти решение обратной некорректно поставленной задачи. Возникают вопросы существования, единственности и устойчивости решения [17]. Весьма привлекательными для решения задачи реконструкции изображений являются итерационные алгоритмы с нелинейными ограничениями, математическое обоснование применения которых для восстановления изображений, наряду с изложением физической сути проблем наиболее полно и корректно описано в [18]. В указанной монографии, пожалуй, впервые в доступной форме была показана общая идеология принципа восстановления изображения с введением как можно большего числа ограничений. Формализованное математическое описание численных методов, используемых для реконструкции, изложено в [19].

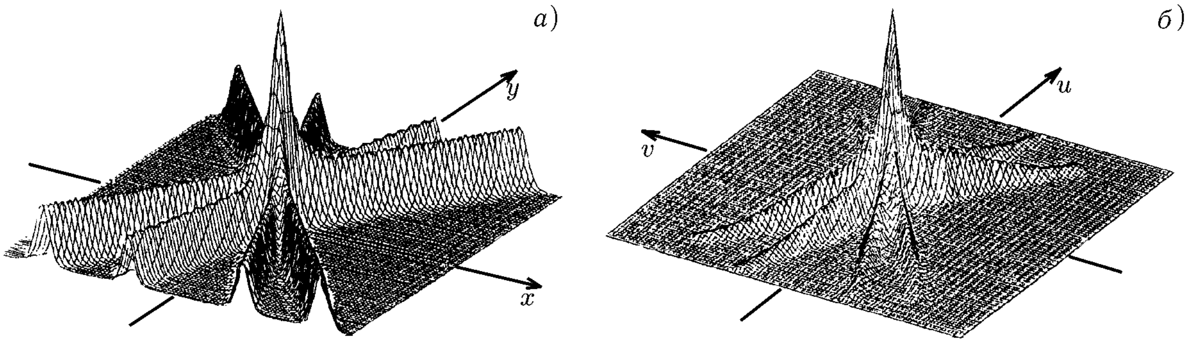


Рис. 1. a — интенсивность синтезированного луча ($N = 4$, позиционные углы $\theta_1 = 105°$, $\theta_2 = 127°$, $\theta_3 = 230°$, $\theta_4 = 249°$); $б$ — изображение синтезированного луча на плоскости uv показывает обширные незаполненные области; углы наблюдения расположены лишь в секторе 77°, сектор 103° — пустой

Общая схема итерационного алгоритма с нелинейными ограничениями определяется выражением

$$f^{k+1} = r_\alpha C f^k + \lambda\,(g - hCf^k), \tag{2}$$

где λ — параметр усиления в итерации (параметр сжатия) [7, 8, 18], интервал изменения которого в общем случае $0 < \lambda < 2/\max[h(u,v)]$, $C = C_1 C_2\,\ldots\,C_n$ — ограничения, r_α — стабилизатор. Решение уравнения (1) с использованием итерационных алгоритмов с нелинейными ограничениями можно упрощённо представить как разложение в ряд обратного оператора, собственные значения которого определяются параметром λ. Заметим, однако, что структура реализаций алгоритмов значительно сложнее. К примеру, в алгоритме TC-CLEAN, кроме параметра λ и установленного уровня контура (trim contour level), используется масштабирующий коэффициент, который индивидуален для каждой итерации. В итоге итерационный процесс включает не одну, а две пары быстрых преобразований Фурье в каждой итерации, что обеспечивает, однако, быструю сходимость: при одинаковой модели распределения яркости объекта (см. рис. 2) ST-CLEAN требуется более 300 итераций, тогда как алгоритму TC-CLEAN необходимо лишь 17 итераций.

Проблема реконструкции итерационными алгоритмами с ограничениями относится к задачам, которые в строгом математическом плане могут рассматриваться методом проекций на выпуклые множества (см., например, [18]), и в некоторых случаях в отсутствие шума вопрос о единственности решения может быть выяснен теоретически. Известно, что при введении ряда ограничений (на пространственную протяжённость источника при знании низкочастотной составляющей спектра и в отсутствие шумов) в задачах восстановления изображений и сигналов применяется итерационный алгоритм Гершберга—Папулиса [20], который позволяет восстановить «отрезанные» системой формирования спектральные составляющие. Однако задача томографической реконструкции при ограниченном числе проекций требует исследования возможности восстановления двумерной структуры, которое можно выполнить лишь при использовании компьютерного моделирования. Кроме того, реконструкция распределения яркости предполагает отсутствие каких-либо ограничений на спектр пространственных частот в широком интервале $0 < \omega < \omega_{\rm b}$. Заметим, что и условия затмения Крабовидной туманности Луной, которое можно представить как постепенное покрытие объекта краем плоского экрана, позволяют регистрировать весь спектр пространственных частот, в случае лунных покрытий лишь сверху ограниченный зоной Френеля. Наконец, поскольку для исключения откликов от боковых лепестков синтезированного луча целесообразно применить специально разработанные для этой цели радиоастрономические алгоритмы, необходимо было провести исследование возможностей применения указанных алгорит-

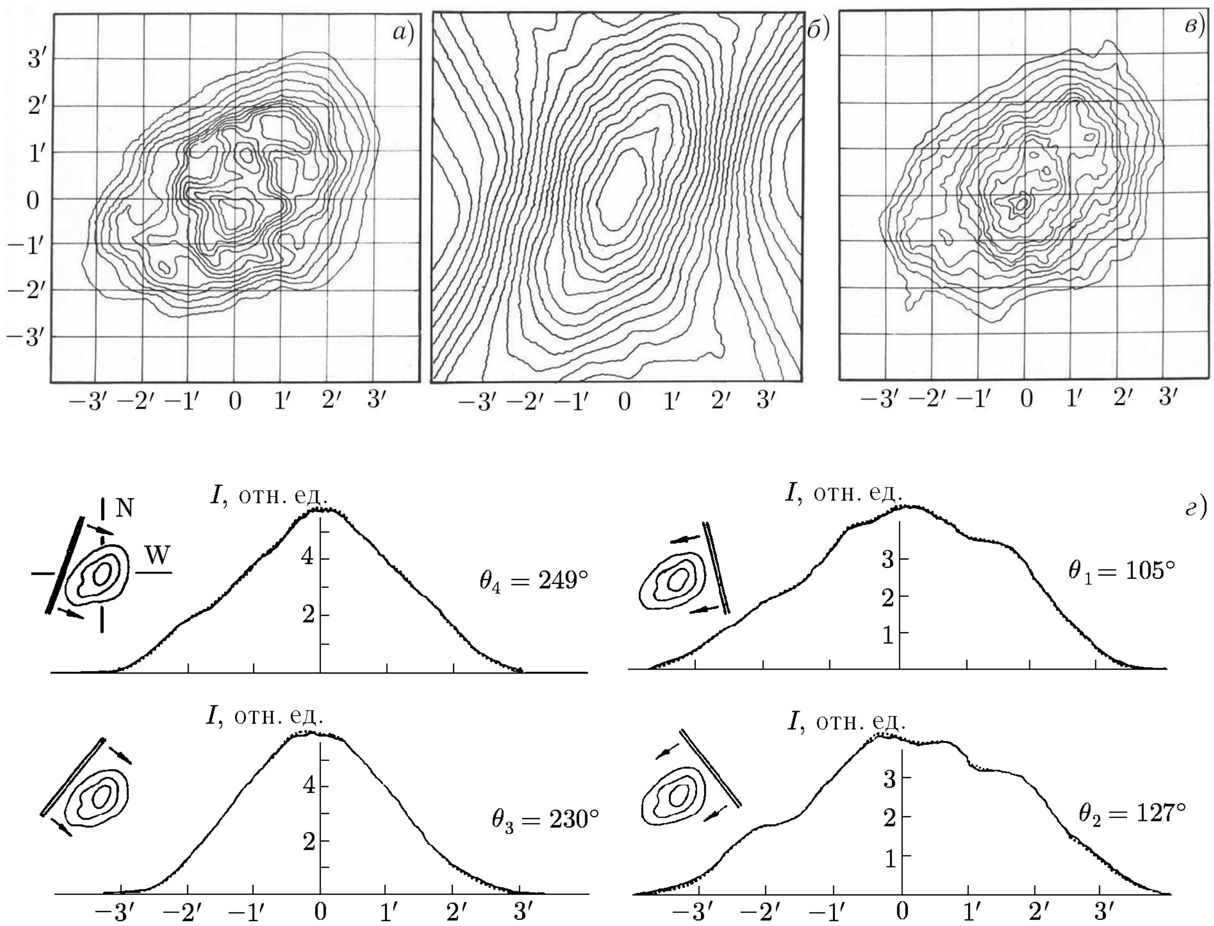


Рис. 2. *а* — исходное модельное распределение яркости $f(x,y)$; *б* — суммарное изображение $g(x,y)$; *в* — результат реконструкции распределения $f(x,y)$ с помощью алгоритма TC-CLEAN ($N = 4$; позиционные углы $\theta_1 = 105°$, $\theta_2 = 127°$, $\theta_3 = 230°$, $\theta_4 = 249°$; $D \sim 360''$, $\varphi_{x,y} = 20{\times}35''$, $N \approx 0{,}1 N_{\rm BR}$); *г* — исходные, полученные из модели $f(x,y)$, и контрольные (пунктир), полученные из решения $f'(x,y)$, одномерные профили интенсивности

мов и сходимости решения для томографической задачи. В частности, выбор коэффициента λ был не совсем понятен, а алгоритм TC-CLEAN [15], который давал надежду на улучшение реконструкции протяжённых областей, нуждался в предварительной апробации в различных практических ситуациях [16].

Следуя обобщённому подходу [18], томографическая задача восстановления изображения эквивалентна минимизации функционала

$$\Phi[f] = \Phi_0[f] + C[f,\lambda] + \alpha\Omega[f], \tag{3}$$

где $\Phi_0[f]$ — функционал-критерий (мера качества, которой в нашей задаче является норма ошибки σ исходных и контрольных профилей; в общем случае в этой роли может выступать, например, энтропия или другие информационные критерии), $C[f,\lambda]$ — функционал ограничений (например, уравнение формирования изображения $g = h * f + n$; неотрицательность решения $f(x,y)$, если известно, что в каком-то интервале $f(x,y) \neq 0$, или если задан спектр $f(\omega)$ на плоскости uv при $\omega \le \omega_{\rm b}$ и $f(\omega) = 0$ при $\omega > \omega_{\rm b}$; априорная информация о спектре решения $f(\omega)$ при $\omega \in \Delta\omega_{\rm i}$, когда $f(\omega) = 0$ при $\omega \notin \Delta\omega_{\rm i}$ в общем интервале пространственных частот $[0, \omega_{\rm b}])$, $\Omega[f]$ — стабилизирующий функционал для устойчивости к шумам, α — стабилизирующий множитель. Проблеме стабилизации, непосредственно касающейся слагаемого $\alpha\Omega[f]$ в правой части (3), уделялось большое внимание, например, в [18], поэтому этот вопрос здесь почти не рассматривается. Отметим, что решение задачи может быть в общем случае сформировано на основе трёх методов: метода Тихонова, метода невязки и метода квазирешений [19]. В нашей задаче невязкой является норма ошибки σ исходных и контрольных проекций. Естественно, что исходные профили должны проходить процедуру фильтрации (начального сглаживания), исходя из размеров ножевых диаграмм, что может быть выполнено и по методу Тихонова. Выбранная нами мера качества соответствует решению по методу невязки. В случае возникновения неустойчивости, что возможно в ряде случаев для алгоритма ST-CLEAN, в [13] было предложено использовать комбинацию метода невязки и метода квазирешений, включающего статистический подход.

2. 2-CLEAN DSA — МЕТОД ДЛЯ ТОМОГРАФИЧЕСКОЙ РЕКОНСТРУКЦИИ ПРИ ОГРАНИЧЕННОМ ЧИСЛЕ ПРОЕКЦИЙ

2.1. Исследование сходимости решений при использовании двух алгоритмов. Моделирование

Сходимость решений при минимизации σ (ошибки исходных и контрольных одномерных профилей) с изменением параметров λ и TC исследовалась нами в работе [13] для каждого из алгоритмов с использованием блок-схемы, показанной на рис. 3.

В работе [13] описаны также последствия необоснованного выбора параметров λ и уровня контура. На рис. 4 приведены графики, которые иллюстрируют достижение минимальной погрешности σ при коррекции λ (ST-CLEAN) или λ и уровня контура (TC-CLEAN) для каждого алгоритма.

Процесс моделирования для объекта сложной структуры, включающего как протяжённые, так и более компактные компоненты различных пространственных масштабов, иллюстрирует рис. 2. Как известно, алгоритм ST-CLEAN формирует решение из суммы пиков, следовательно, результат является экстремальным, наиболее высокочастотным (так называемым «острым»), допустимым в рамках установленных ограничений. С другой стороны, TC-CLEAN аккумулирует результат, начиная с наиболее протяжённых компонентов, следствием чего является также экстремальное, наиболее низкочастотное, сглаженное решение, допустимое в рамках тех же установленных ограничений [9].

2.2. Определение области допустимых решений в сложных случаях

Как было показано в [9], объект, состоящий из компактных источников (δ-функций), может быть успешно восстановлен методом ST-CLEAN. Метод TC-CLEAN даёт схожие результаты для простого объекта, состоящего из отдельных компонентов, однако TC-CLEAN эффективнее в вычислительном отношении. Для профилей с гладкой структурой, лишь с небольшими возвышениями, решение может быть получено [9] как на основе сглаженных компонентов (TC-CLEAN), так и из обособленных компонентов (ST-CLEAN). Алгоритм ST-CLEAN увеличивает контраст мелких деталей, однако более протяжённый основной фон понижается вследствие так называемых «оврагов». Если минимальная невязка при реконструкции двумя алгоритмами приблизительно совпадает: $\min\sigma(\text{ST-CLEAN}) \approx \min\sigma(\text{TC-CLEAN})$, что соответствует примеру на рис. 4 (графики приведены для моделирования, показанного на рис. 2), то решения, полученные с помощью

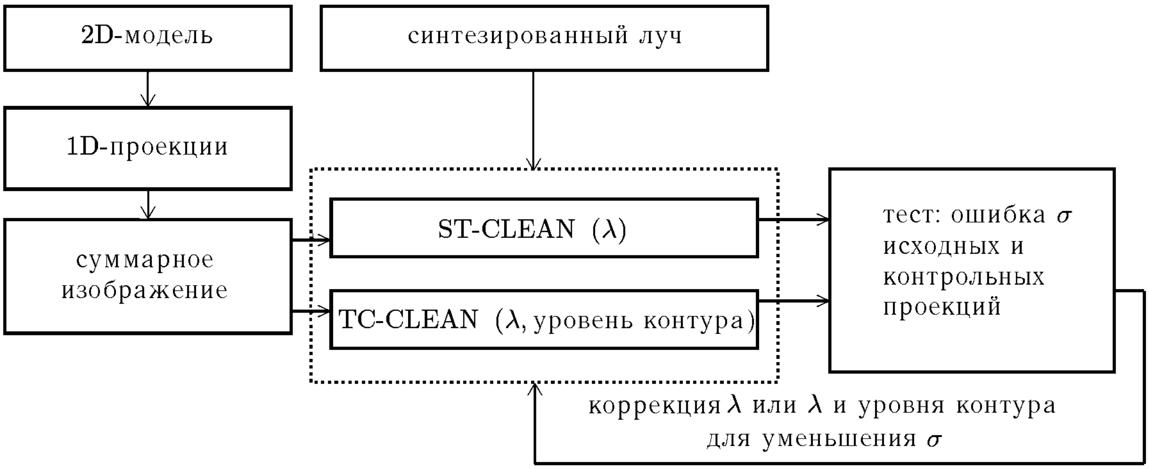


Рис. 3. Блок-схема для моделирования и нахождения решения методом 2-CLEAN DSA

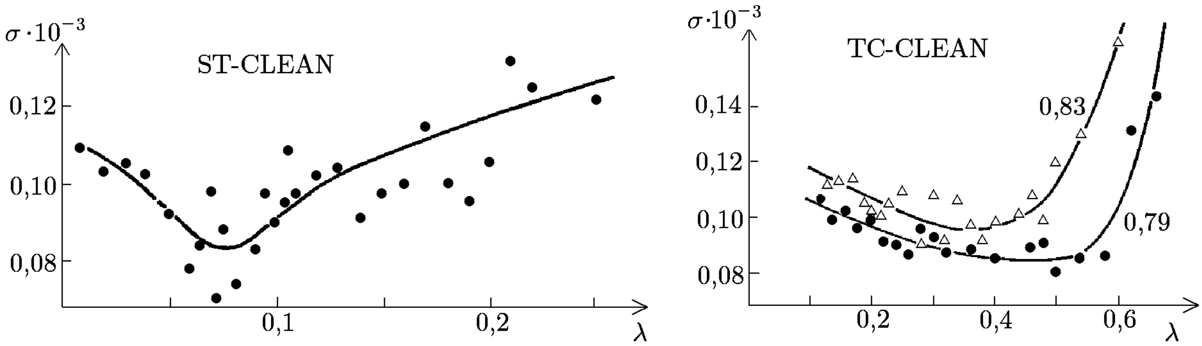


Рис. 4. Среднеквадратичная ошибка исходных и контрольных профилей $\sigma(\lambda)$ для каждого из алгоритмов. Для алгоритма TC-CLEAN цифры у кривых соответствуют выбранному уровню контура

двух алгоритмов, формально эквивалентны. В итоге мы должны: (*а*) предположить существование целого класса допустимых решений в интервале между сглаженным и высокочастотным распределениями либо (*б*) отдать предпочтение результату, соответствующему физическим свойствам объекта на основании имеющейся априорной информации.

Отметим, что если при сложной структуре объекта различия двух экстремальных вариантов решения при одинаковой погрешности исходных и контрольных проекций становятся столь велики, что между ними наблюдается лишь слабая корреляция, это означает, что в эксперименте необходимо либо увеличить число используемых проекций N, либо понизить желаемое угловое разрешение, т. е. верхнюю граничную частоту ω_{b}. В [9] было предложено проводить исследование области допустимых решений при сложной структуре объектов в задачах радиоастрономии и томографии в случае ограниченного числа проекций с помощью двух алгоритмов. Несколько позднее в [21] этот метод, включающий также процесс предварительного моделирования, получил название 2-CLEAN DSA, которое закрепилось за ним в дальнейших работах [11, 12, 22].

На рис. 5 представлены особенности накопления решения алгоритмами ST-CLEAN и TC-CLEAN с увеличением номера итерации, а также вид модулей спектров решений $F(u, v)$. В случае простых изображений спектры решений, полученные двумя алгоритмами, практически полностью совпадают; в сложных случаях происходит перераспределение энергии: некоторое смещение спектра, полученного с помощью алгоритма ST-CLEAN, в область более высоких простран-

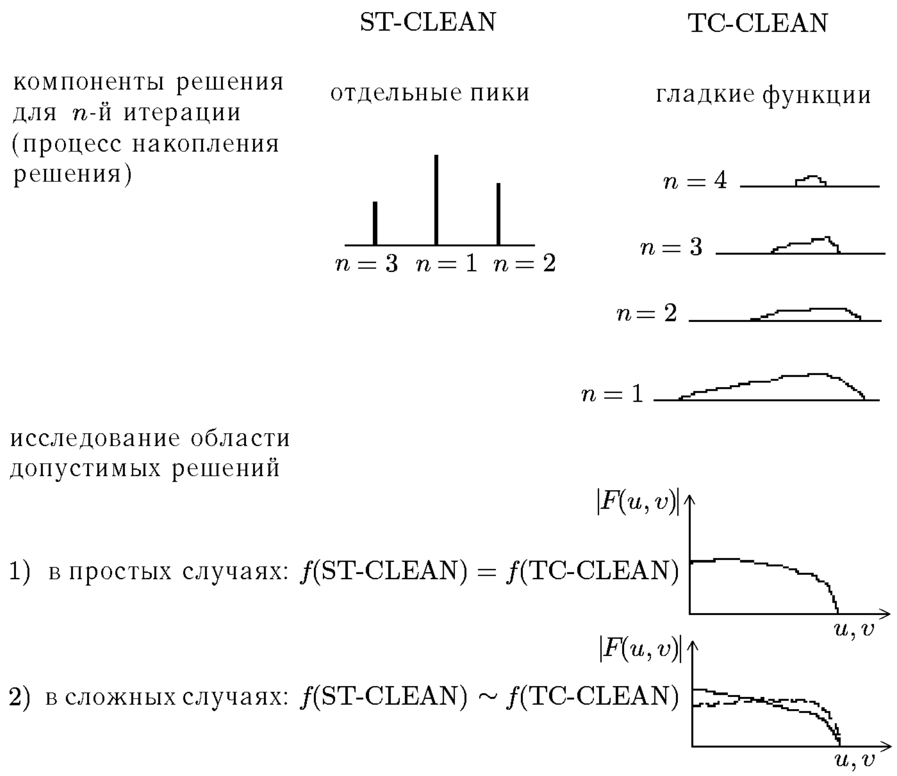


Рис. 5. Особенности итерационных процессов накопления решения для алгоритмов ST-CLEAN и TC-CLEAN

ственных частот. Следует подчеркнуть, что возможно лишь небольшое перераспределение, т. к. итерационные процессы проходят в рамках одних и тех же установленных ограничений. Кроме того, говорить об одинаковом угловом разрешении изображений, полученных с помощью двух алгоритмов, правомерно лишь при неизменной граничной частоте ω_{b}.

На рис. 6 можно проследить условную схему суммирования компонентов решения алгоритмом TC-CLEAN в области пространственных частот. Процесс начинается с наиболее низкочастотных, протяжённых компонентов; далее область, соответствующую установленным ограничениям, заполняют более высокочастотные составляющие. Первым приближением $f^1(x,y)$ решения является суммарное распределение яркости в пределах области, ограниченной приспособленным контуром метода TC-CLEAN. Уровень контура следует выбирать наименьшим, чтобы он охватывал наиболее низкие пространственные частоты, однако не выходил за пределы зоны существования решения. Рис. 6 иллюстрирует также схему возможного перераспределения энергии в сложных случаях, когда для определения области допустимых решений используется метод 2-CLEAN DSA, включающий два алгоритма реконструкции. Кроме того, на рис. 6 условно изображено решение в узком интервале пространственных частот $\Delta\omega_{\text{i}}$, которое получается при подходе к задаче [4, 5] методом Гершберга—Папулиса [20, 23, 24]. Рис. 6 наглядно иллюстрирует отличие возможностей представленного в известных монографиях [4, 5] метода решения, предполагающего априорное знание спектра, от радиоастрономического подхода, который позволяет получить решение в широкой полосе пространственных частот.

При сопоставлении методов решения следует также обратить внимание на то, что изложенный в упомянутых монографиях [4, 5] подход не использует синтезированный луч, а также никакие кардинальные способы устранения искажений, которые происходят из-за откликов от его боковых

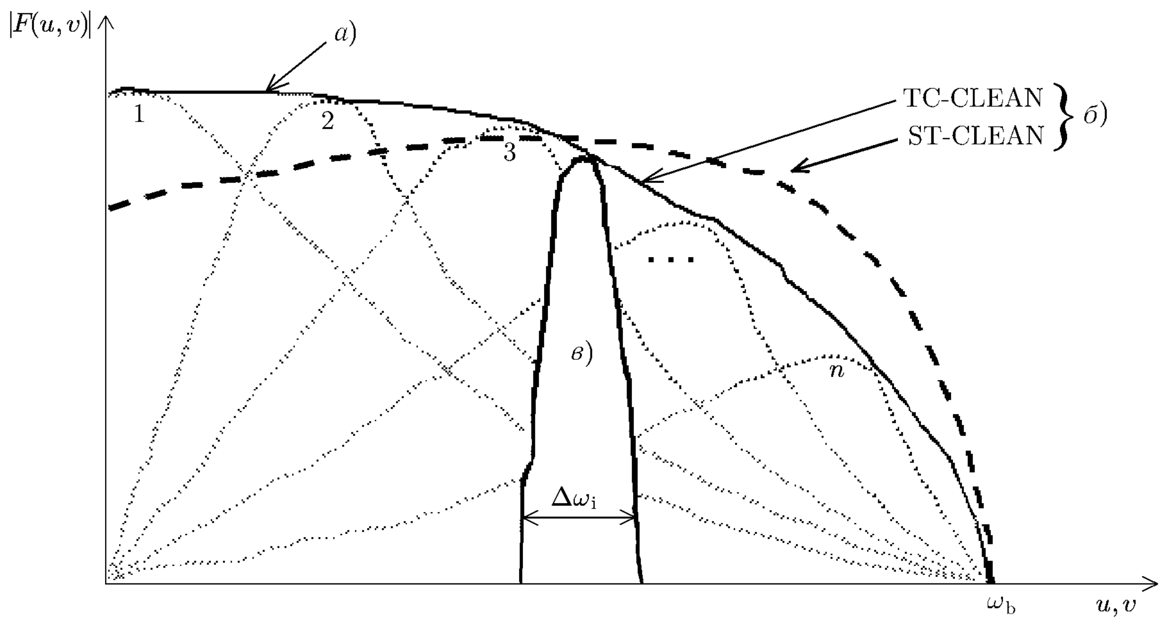


Рис. 6. Схема процесса получения решений: *а*) суммирование компонентов решения алгоритмом TC-CLEAN ($1, 2, 3, \ldots, n$ — компоненты итераций); *б*) перераспределение энергии по спектру в небольших пределах при определении области допустимых решений методом 2-CLEAN DSA с помощью двух алгоритмов: ST-CLEAN и TC-CLEAN; *в*) решение, полученное в узком интервале $\Delta\omega_{\mathrm{i}}$, которое иллюстрирует описанный в [4, 5] подход к задаче с использованием метода Гершберга—Папулиса

лепестков. Процесс включает итерационную процедуру с введением нелинейных ограничений (неотрицательность и финитность решения, спектральные ограничения и т. д.) с целью последовательной подгонки контрольных проекций к исходным, однако в целом решение сводится к фильтрации. Эффективное удаление искажений, обусловленных влиянием боковых лепестков, может быть выполнено лишь с помощью проведения такой процедуры в пространстве изображений с использованием нескольких необходимых суммарных функций, к которым, в первую очередь, относится синтезированный луч и суммарное изображение. Тем не менее следует заметить, что описанный в [4, 5] подход для целого ряда задач может быть оправдан и обеспечивает достаточно точное решение, но при априорном знании характерных размеров восстанавливаемых пространственных компонентов и при условии их подобия.

3. РЕКОНСТРУКЦИЯ ПО РЕАЛЬНЫМ ДАННЫМ. АСТРОТОМОГРАФИЯ

Реконструкция распределения яркости Крабовидной туманности по одномерным профилям интенсивности, полученным при её лунных затмениях, является примером двумерной малоракурсной астротомографии. Наблюдения покрытий Крабовидной туманности Луной детально описаны в [25, 26]. Описание реконструкции распределения яркости объекта с помощью алгоритма ST-CLEAN, выполненной на раннем этапе работы, представлено в [27].

Численное моделирование и исследование возможностей восстановления [9, 13] позволили выполнить более качественную реконструкцию распределения яркости Крабовидной туманности [28] с исследованием границ области допустимых решений при использовании метода 2-CLEAN DSA. Обсуждение восстановленной структуры представлено в работах [28–30]. На рис. 7*а* показана условная схема покрытия Крабовидной туманности Луной, а также вид интегральных и дифференциальных кривых изменений интенсивности. На рис. 7*б* показано располо-

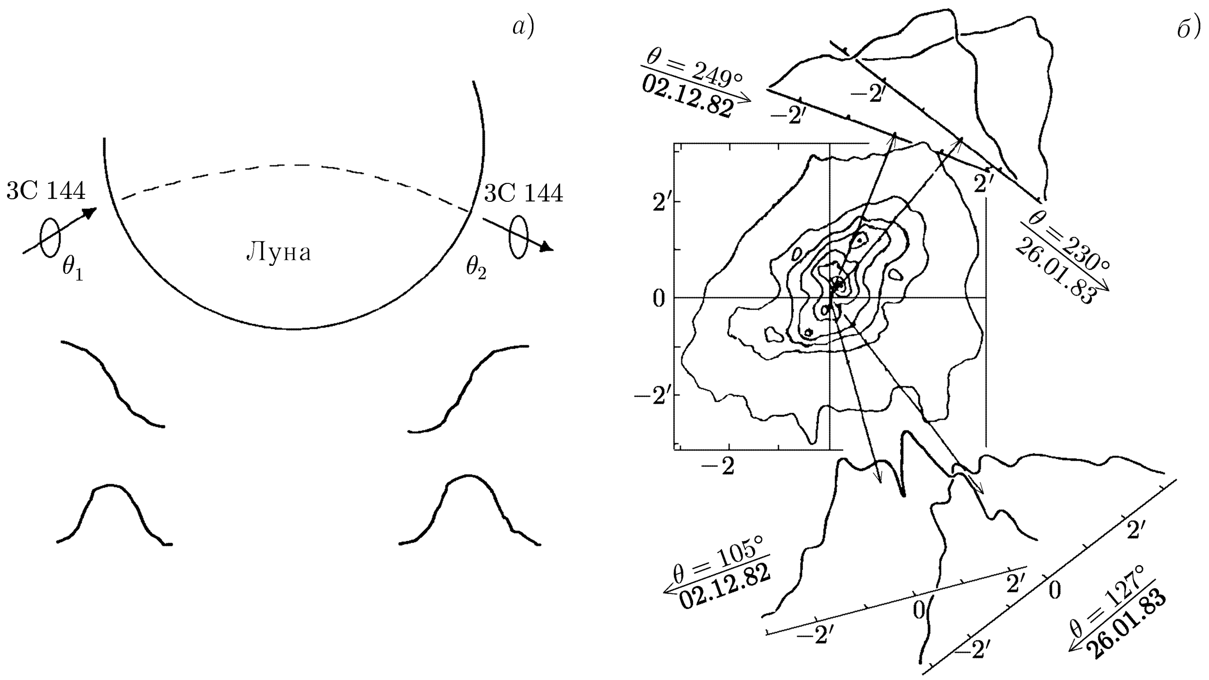


Рис. 7. Схема покрытия Крабовидной туманности лунным диском (*а*); ниже качественно приведены интегральные и дифференциальные кривые изменения интенсивности. Расположение проекций при двух лунных покрытиях (*б*)

жение проекций в соответствии с условиями покрытий 1982 и 1983 годов. Отметим, что проекции, полученные на частоте 750 МГц, имеют очень высокое отношение сигнал/шум (последний составляет около 2 % от максимальной интенсивности профилей), поскольку для наблюдений использовался расположенный в Евпатории радиотелескоп РТ-70, имеющий зеркало с диаметром 70 м. Проекции, суммарное изображение, а также результаты реконструкции [28], полученные методом 2-CLEAN DSA при использовании двух алгоритмов чистки ST-CLEAN и TC-CLEAN, иллюстрирует рис. 8.

Определены пределы области возможных решений от наиболее изрезанного (ST-CLEAN) до наиболее сглаженного варианта (TC-CLEAN), допустимые в рамках заданных ограничений. В то время как алгоритм ST-CLEAN повышает контрастность компактных компонентов, и полученное с его помощью распределение яркости имеет значение для выявления мелких деталей, карта, полученная с помощью TC-CLEAN, находится в большем согласии с известной априорной информацией о структуре туманности. В целом, два распределения яркости имеют сильное сходство, подобны в основных деталях, обладают большой корреляцией. Угловое разрешение карты $20'' \times 35''$ (размер ячейки сетки на карте $60'' \times 60''$).

На рис. 9 показан пример реконструкции распределения яркости Крабовидной туманности на частоте 178 МГц по пяти профилям, полученным при покрытиях лунным диском [28]. Два первых профиля относятся к покрытию 1982 года, третий и четвертый — к 1983 году, пятый профиль получен при центральном покрытии 1974 года. Поскольку отношение сигнал/шум в этом примере значительно хуже, угловое разрешение карты составило лишь $45'' \times 65''$ дуги. Вследствие высокой погрешности исходных экспериментальных одномерных профилей исследование области допустимых решений двумя алгоритмами не представляло интерес. Реконструкция выполнена методом TC-CLEAN.

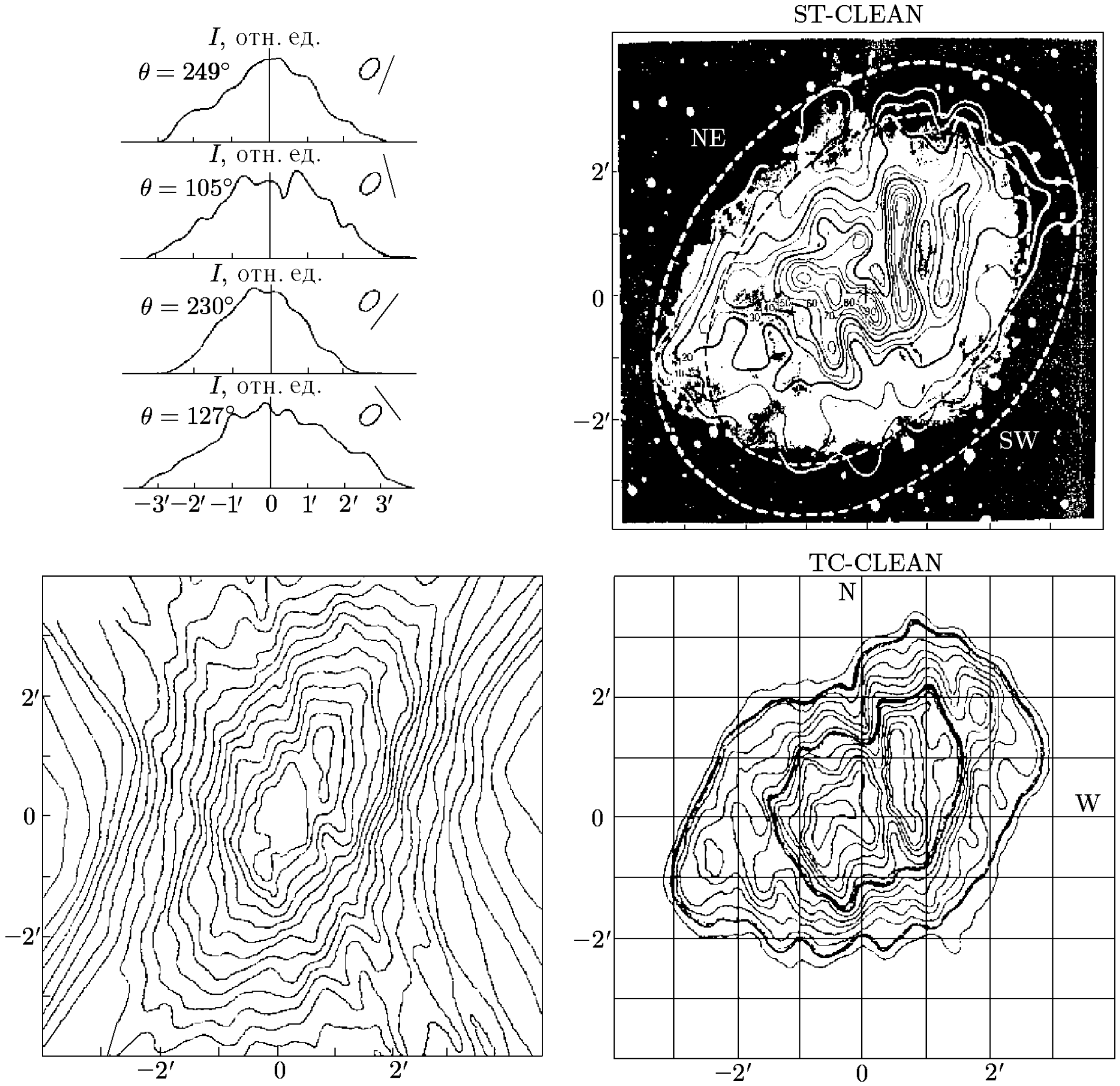


Рис. 8. Реконструкция распределения яркости Крабовидной туманности методом 2-CLEAN DSA с использованием алгоритмов ST-CLEAN и TC-CLEAN на частоте 750 МГц ($D \sim 360''$, $\varphi_{x,y} = 20'' \times 35''$); слева показаны одномерные проекции и суммарное изображение

Следует отметить, что низкое отношение сигнал/шум является серьёзным ограничением для реконструкции при ограниченном числе проекций. Использование большого числа проекций может быть в этом случае эффективным и даже необходимым вследствие достижения позитивного статистического эффекта от увеличения числа независимых измерений.

4. ОБСУЖДЕНИЕ

Сопоставление радиоастрономического подхода к восстановлению распределения радиояркости с известной по монографиям [4, 5] реконструкцией для ограниченного числа проекций методом Гершберга—Папулиса [20, 23] показывает, что последний является, на первый взгляд, частным случаем изложенного здесь подхода, и стремится обеспечить итерационную подгонку решения исходя из соответствия исходных и контрольных профилей за счёт введения весьма су-

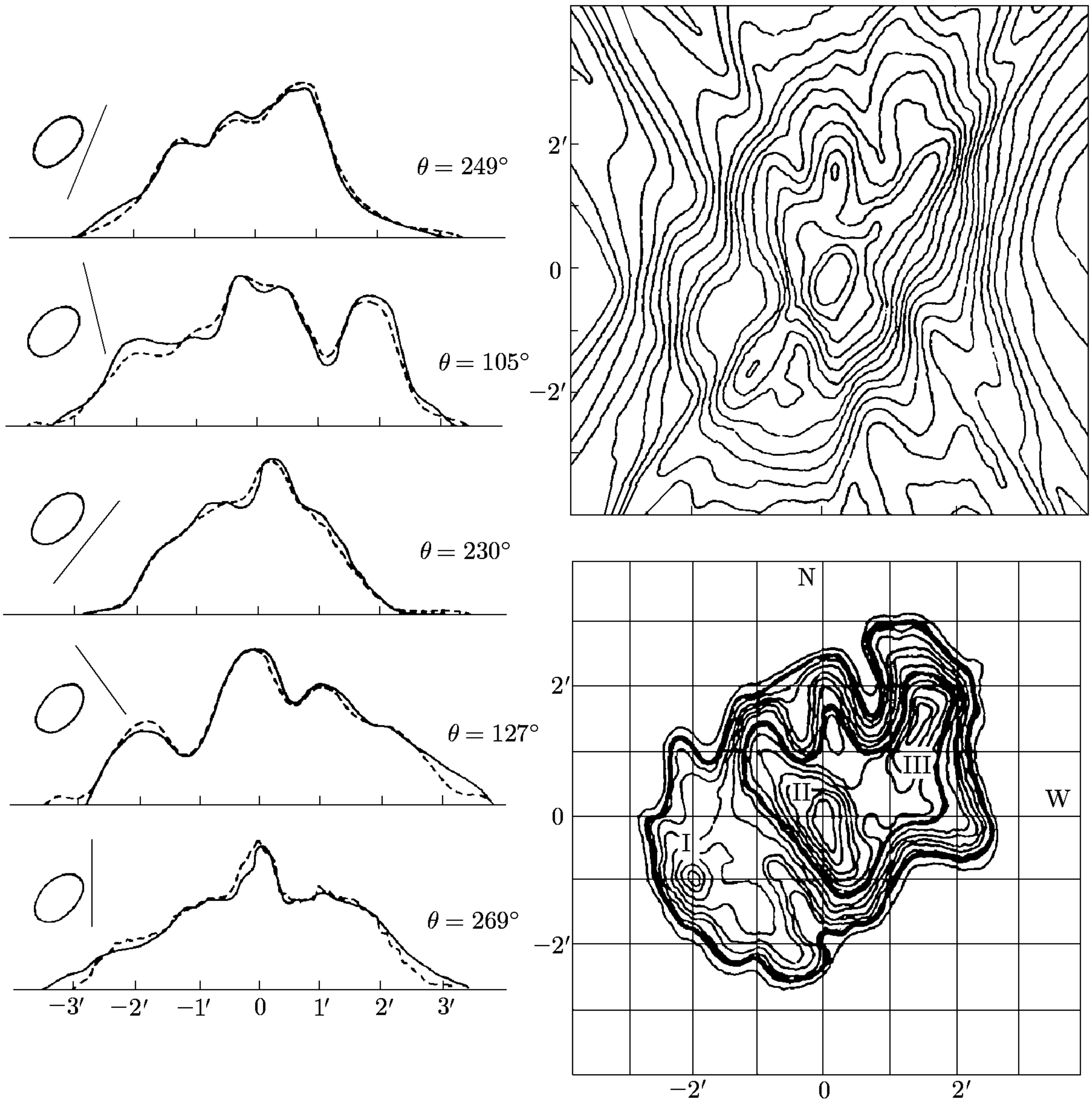


Рис. 9. Иллюстрации к реконструкции распределения яркости Крабовидной туманности на частоте 178 МГц по пяти проекциям (слева) трёх лунных покрытий (справа показаны суммарное изображение и реконструкция по методу TC-CLEAN)

щественного спектрального ограничения — узкой полосы пространственных частот $\Delta\omega_{\rm i}$. Использование априорной информации об ограничении спектра пространственных частот является, фактически, восстановлением при априорном знании размеров и единообразия компонентов решения. Это означает, что энергия спектра одномерных проекций должна распределиться определённым образом на плоскости пространственных частот, удовлетворяя как спектральным ($\Delta\omega_{\rm i}$), так и всем прочим установленным ограничениям. Можно рассматривать такой подход как частный случай, однако было бы правильнее применить спектральное ограничение $\Delta\omega_{\rm i}$ лишь после или вместе с процедурой устранения искажений на томограмме («грязной» карте), связанных с вкладом боковых лепестков синтезированной передаточной функции.

Введение узкого спектрального интервала $\Delta\omega_{\rm i}$ перераспределяет энергию спектра одномерных проекций в априорно заданные области на плоскости пространственных частот. Заметим, что ошибочная априорная информация об ограничении спектра пространственных частот полосой $\Delta\omega_{\rm i}$ будет приводить к реконструкции ложных компонентов. В таком случае иногда может быть получено ложное двумерное решение, однако наверняка с существенно большей ошибкой контрольных и исходных проекций.

Если в предложенном методе реконструкции полного спектра в полосе пространственных частот от 0 до $\omega_{\rm b}$ необходимое количество проекций N оценено как $0{,}1N_{\rm BR}$, то введение жёсткого спектрального ограничения может существенно снизить эту оценку. Конкретные значения можно получить лишь с помощью компьютерного моделирования.

Статистический подход к томографическим задачам, направленный на исследование влияния шумов, изложен в работах [31, 32]. Поскольку из теории статистических решений следует, что никакая обработка первичных данных не может увеличить содержащуюся в них информацию, напрашивается естественный вывод о том, что поиск оптимальных решений должен ориентироваться непосредственно на радоновские образы, т. е. на проекции или их спектры. Однако при изучении сложных объектов, как правило, осуществляется восстановление изображения и анализируется именно томографическое изображение, а не образ в пространстве Радона. Это связано с тем, что при решении задачи обычно используется априорная информация об образах. Следует ли восстанавливать изображения или целесообразнее осуществить требуемый анализ по радоновскому образу? Статистический подход в классическом варианте решения томографической задачи даёт ответ, что целесообразнее работать с радоновскими образами [31]. Доказательством является то положение, что если ориентироваться на восстановление изображения, то для обеспечения той вероятности распознавания, которая достигается при использовании радоновских образов, необходимо увеличить отношение сигнал/шум в 1,21 раза.

Изложенный радиоастрономический подход с использованием априорной информации, т. е. ограничений, приводит к кардинальному изменению ситуации. Использование этого подхода эквивалентно использованию упомянутой выше априорной информации. Кроме того, он эффективен для исключения искажений, т. к. использует пространство изображений. В итоге становится совершенно очевидно, почему в задаче малоракурсной томографии при введении ограничений целесообразно анализировать именно качественно восстановленное двумерное изображение или его двумерный спектр, которые удаётся получить при радиоастрономическом подходе.

Статистический подход к решению проблемы с использованием метода максимальной энтропии и других информационных критериев (максимума информации и т. д.) в высшей степени эффективен при наличии импульсных компонентов, особенно в условии слабых сигналов. На практике это широко используется в ряде медицинских задач, а также в астрофизике и других областях.

Всё вышесказанное — развитие общей идеологии использования и введения в процесс решения задачи наибольшего количества априорной информации в контексте развития тезиса Гордона [1], с которого мы начали изложение. Общие положения такого подхода наиболее полно сформулированы в монографии [18]. Кратко их можно выразить утверждением о том, что количественная оценка влияния различных ограничений на сужение класса допустимых решений весьма сложна из-за нелинейного характера ограничений, математические трудности вынуждают отказаться от глобального решения проблемы. Введение ограничений приводит к сужению области допустимых решений, а количественную оценку этого сужения можно получить лишь на основе практических результатов. Предложенный подход находится в полном согласии с приведённым утверждением.

ВЫВОДЫ

Представлен радиоастрономический подход к проблеме малоракурсной томографии, в основе которого лежит предложенный метод определения области допустимых решений 2-CLEAN DSA. Решение задачи деконволюции происходит с введением синтезированной передаточной функции, а также двух реализаций итерационного алгоритма с нелинейными ограничениями, известного в радиоастрономии под названием «чистка». Метод 2-CLEAN DSA можно рассматривать как универсальный алгоритм для данной задачи, использующий широкий спектр пространственных частот и легко адаптирующийся к введению дополнительных ограничений.

Метод 2-CLEAN DSA прошёл апробацию в эксперименте, обсуждался в процессе своего развития на отечественных и международных конференциях [10–12, 29, 33–35]. В настоящее время он является перспективным способом решения томографической задачи при ограниченном наборе проекций. Метод прост в вычислительном отношении, его структура легко адаптируется к введению дополнительных ограничений. Указанные качества позволяют использовать метод 2-CLEAN DSA для широкого круга задач, а также для моделирования перспективных томографических систем и обработки данных реальных экспериментов. Разработанный подход позволяет предложить информационно-вычислительную технологию реконструкции, распознавания и анализа изображений малоракурсной томографии в широкой полосе пространственных частот.

Сформулируем основные выводы проведённого исследования.

1. Реконструкция распределения яркости при ограниченном числе проекций ($N < N_{\rm BR}$) в широкой полосе пространственных частот $[0, \omega_{\rm b}]$ не может быть по-настоящему успешной без удаления откликов от боковых лепестков синтезированной диаграммы, а также без предварительной оценки возможностей реконструкции с использованием компьютерного моделирования при установленных параметрах (числе проекций N, распределении позиционных углов $\theta_{\rm i}$, других возможных ограничениях) для вероятных вариантов структуры объекта.

2. Показано, что реконструкция распределения яркости в широкой полосе пространственных частот методом 2-CLEAN DSA возможна при числе проекций N, равном 0,1 от необходимого количества проекций $N_{\rm BR}$ при классическом томографическом подходе согласно Брейсуэллу—Риддлу. Метод позволяет определить границы области допустимых решений в сложных случаях.

3. На основе метода 2-CLEAN DSA предложена информационно-вычислительная технология реконструкции, распознавания и анализа изображений малоракурсной томографии в широкой полосе пространственных частот. Технология включает процедуру удаления откликов от боковых лепестков синтезированного луча, соответствующего условиям задачи; возможность приложения двух алгоритмов чистки: либо одного из алгоритмов, либо обоих алгоритмов для проведения предварительного компьютерного моделирования и исследования области допустимых решений; реконструкцию изображения по экспериментальным данным.

Информационно-вычислительная технология реконструкции на основе метода 2-CLEAN DSA может иметь приложение в самых различных областях, некоторые из которых перечислены во введении. В следующих работах мы предполагаем рассмотреть некоторые потенциальные возможности применения предложенного метода.

Разработанный подход к решению томографической задачи при ограниченном числе проекций базируется на работах радиоастрономов и может называться радиоастрономическим. Существенный вклад в решение томографической задачи при ограниченном числе проекций внесли следующие работы радиоастрономов.

1. Работа [3] — классическая в томографии и радиоастрономии, в которой показано, что при числе проекций $N \geq \pi D/\varphi$, где D — диаметр исследуемой области, а φ — желаемое угловое разрешение, восстановление распределения яркости можно считать полным (что эквивалентно

заполнению суммарной передаточной функцией плоскости пространственных частот до радиуса ω_b).

2. Использование синтезированного луча и удаление откликов от его боковых лепестков, расходящихся радиально от центра [13, 27, 28, 36]. Следует отметить, что работа [36] с использованием синтезированной диаграммы из ножевых лучей является пионерской, хотя выполнена без необходимого обоснования решения.

3. Алгоритмы чистки ST-CLEAN [14] и TC-CLEAN [15], являющиеся радиоастрономическими реализациями итерационных алгоритмов с нелинейными ограничениями.

4. Исследование сходимости решений [13] при использовании алгоритмов чистки ST-CLEAN и TC-CLEAN для выяснения возможностей их применения в томографической задаче с ограниченным числом проекций в широкой полосе пространственных частот $[0, \omega_b]$.

5. Определение предельно допустимых вариантов решения при помощи двух вариантов алгоритма чистки (ST-CLEAN и TC-CLEAN) в задачах радиоастрономии и томографии [9].

6. Метод 2-CLEAN DSA [11, 12] решения томографической задачи при ограниченном числе проекций в широкой полосе пространственных частот [10, 21, 22, 37].

7. Реконструкция по данным эксперимента [28] распределения яркости Крабовидной туманности на основе четырёх профилей двух лунных покрытий с определением предельно допустимых вариантов решений, возможных в рамках заданных ограничений.

Заметим, что ранние пионерские работы, связанные с получением информации о двумерной структуре Крабовидной туманности по профилям лунных покрытий, относятся к концу шестидесятых годов и выполнены Л. И. Матвеенко [38, 39].

Автор выражает благодарность К. С. Станкевичу за основополагающую инициативу проведения наблюдений лунных покрытий Крабовидной туманности в 1982 и 1983 годах, а также В. П. Иванову, И. Т. Бубукину, армянским коллегам А. М. Асланяну, А. Г. Гуляну, Р. М. Мартиросяну, без вклада которых в успешное проведение наблюдений на 70-метровом радиотелескопе РТ-70 в Евпатории настоящая работа не получила бы необходимый исходный импульс. Особой благодарности достойна О. А. Подвойская за проявленное терпение и неоценимую помощь при проведении серии вычислений, которая легла в основу работы 1989 года по исследованию сходимости алгоритмов. Автор благодарен О. И. Шаровой, полезный обмен мнениями с которой стимулировал изложение проблемы в настоящей работе, Е. Н. Виняйкину и П. А. Фридману за информацию об интересных публикациях, а также Г. И. Василенко и А. М. Тараторину за плодотворные дискуссии.

Работа выполнена при частичной поддержке Минобразования России в рамках научно-технической программы «Научные исследования высшей школы по приоритетным направлениям науки и техники» (проект 209.01.01.003), а также программы по поддержке ведущих научных школ Российской Федерации (проект НШ-1483.2003.2).

СПИСОК ЛИТЕРАТУРЫ

1. Гордон Р. // Построение изображений в астрономии по функциям когерентности. М.: Мир, 1982. С. 306.
2. Smith K. T., Solmon D. C., Wagner S. L. // Bull. Amer. Math. Soc. 1977. V. 83. P. 1 227.
3. Bracewell R. N., Riddle A. C. // Astrophys. J. 1967. V. 150. P. 427.
4. Левин Г. Г., Вишняков Г. Н. Оптическая томография. М.: Радио и связь, 1989.
5. Низкотемпературная плазма. Т. 13. Томография плазмы / В. В. Пикалов, Т. С. Мельникова. Новосибирск: Наука, 1995.
6. ТИИЭР. 1983. Т. 71, № 3. Реконструктивная вычислительная томография.

7. Шафер Р., Мерсеро Р., Ричардс М. // ТИИЭР. 1981. Т. 69, № 4. С. 432.
8. Даджион Д., Мерсеро Р. Цифровая обработка многомерных сигналов. М.: Мир, 1988.
9. Агафонов М. И., Подвойская О. А. // Изв. вузов. Радиофизика. 1990. Т. 33, № 10. С. 1 185.
10. Agafonov M. I. // Abstract of 24th General Assembly of the International Union of Radio Science, Kyoto, 1993. P. 454.
11. Agafonov M. I. // ASP Conference Series. 1997. V. 125. Astronomical Data Analysis Software and Systems. VI. P. 202. (http://www.cv.nrao.edu/adass/adassVI/agafonovm.html).
12. Agafonov M. I. // ASP Conference Series. 1998. V. 145. Astronomical Data Analysis Software and Systems. VII. P. 58. (http://www.stsci.edu/stsci/meetings/adassVII/agafonovm.html).
13. Агафонов М. И., Подвойская О. А. // Изв. вузов. Радиофизика. 1989. Т. 32, № 6. С. 742.
14. Hogbom J. A. // Astron. Astrophys. Suppl. Ser. 1974. V. 15, No. 3. P. 417.
15. Steer D. G., Dewdney P. E., Ito M. R. // Astron. Astrophys. 1984. V. 137, No. 2. P. 159.
16. Cornwell T. J. // Astron. Astrophys. 1988. V. 202. P. 316.
17. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач. М.: Наука, 1986.
18. Василенко Г. И., Тараторин А. М. Восстановление изображений. М.: Радио и связь, 1986.
19. Иванов В. В. Методы вычислений на ЭВМ: Справочное пособие. Киев: Наук. думка, 1986.
20. Papoulis A. // IEEE Trans. CAS. 1975. V. 22. P. 735.
21. Agafonov M. I. // Abstracts of the Astronomical Data Analysis Software and Systems (ADASS'96) Conference, Charlottesville, Virginia, USA, 1996. P. 50.
22. Agafonov M. I. // Absracts of the 9-th European and 5-th Euro-Asian Astronomical Society Conference (Joint European and National Astronomical Meeting — JENAM) Moscow, Russia, 2000. P. 170.
23. Gerchberg R. W. // Optica Acta. 1974. V. 21, No. 9. P. 709.
24. Defrise M., De Mol C. A. // Optica Acta. 1983. V. 30, No. 4. P. 403.
25. Агафонов М. И., Асланян А. М., Барабанов А. П. и др. // Письма в Астрон. журн. 1984. Т. 10, № 10. С. 730.
26. Агафонов М. И., Асланян А. М., Барабанов А. П. и др. // Астрофизика. АН Арм. ССР, 1984. Т. 21. С. 283.
27. Агафонов М. И., Асланян А. М., Гулян А. Г. и др. // Письма в Астрон. журн. 1986. Т. 12, № 4. С. 275.
28. Агафонов М. И., Иванов В. П., Подвойская О. А. // Астрон. журн. 1990. Т. 67, № 3. С. 549.
29. Agafonov M. I., Stankevich K. S. // Abstracts of International Symposium on Computerized Tomography, Novosibirsk, 1993. P. 11.
30. Agafonov M. I., Stankevich K. S. // ASP Conference Series. 1996. V. 93. Radioemission from the Sstars and the Sun. P. 147.
31. Троицкий И. Н. // Автометрия. 1987. № 2. С. 94.
32. Троицкий И. Н. Статистическая теория томографии. М.: Радио и связь, 1989.
33. Агафонов М. И., Подвойская О. А. // Тез. докладов 13 Всесоюзной научно-технической конф. «Высокоскоростная фотография, фотоника и метрология быстропротекающих процессов», Москва, 1987. С. 114.
34. Агафонов М. И., Подвойская О. А. // Тез. докл. 18 Всесоюзной конф. «Радиотелескопы и интерферометры», Иркутск, 1986. С. 154.
35. Агафонов М. И., Подвойская О. А., Станкевич К. С. // Тез. докл. 4 Всесоюзного симпозиума по вычислительной томографии, Ташкент, 1989. С. 60.
36. Maloney F. P., Gottesman S. T. // Astrophys. J. 1979. V. 234. P. 485.
37. Agafonov M. I. // Abstracts of the Astronomical Data Analysis Software and Systems (ADASS'97) Conference, Sonthofen, Germany, 1997. P. P1:01.

38. Матвеенко Л. И. // Астрон. журн. 1968. Т. 45, № 1. С. 160.
39. Матвеенко Л. И. // Астрон. журн. 1969. Т. 46, № 2. С. 250.

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

Поступила в редакцию
4 февраля 2003 г.

FEW-PROJECTION TOMOGRAPHY. I. RADIOASTRONOMICAL APPROACH TO THE PROBLEM AND THE 2-CLEAN DSA METHOD

M. I. Agafonov

The radioastronomical approach to the few-projection tomography is developed. It is shown that the 2-CLEAN DSA method proposed for determination of the permissible solution area is an efficient technique for solving this problem. The method is based on solving the deconvolution problem with allowance for the synthesized beam or Green's function. The distortions due to the sidelobe responses are eliminated using two realizations of the well-known radioastronomical CLEAN algorithm with nonlinear constraints. In comparison with the traditional tomographic approach, the 2-CLEAN DSA method permits about tenfold decrease in the number of projections required for two-dimensional reconstruction in the case of a wide spatial-frequency spectrum limited from above. The method can be easily adapted to additional constraints. Examples of astrotomographic reconstruction are presented. We show that the proposed method is prospective for a number of applications in remote sensing and compare it with other well-known reconstruction techniques. The papers by radioastronomers contributed significantly to the development of the method components are pointed out.

УДК 534.26

МОДЕЛИРОВАНИЕ ВЫСОКОЧАСТОТНЫХ АКУСТИЧЕСКИХ ПОЛЕЙ, РАССЕЯННЫХ НА ТЕЛАХ В РЕФРАКЦИОННЫХ ВОЛНОВОДАХ

И. П. Смирнов, А. А. Хилько, А. И. Хилько

В работе исследуются структура высокочастотного поля, рассеянного объектом наблюдения — телом, помещённым в плоскослоистый волновод океанического типа с неровным дном. Решение задачи позволяет проанализировать возможности построения акустических систем томографического типа, включающих в себя набор акустических источников и приёмников, располагающихся в пределах ограниченного по размерам района наблюдений.

ВВЕДЕНИЕ

При решении широкого круга практических задач, связанных с контролем скоплений рыб, морских животных и других малоразмерных подводных объектов в шельфовых районах океана с помощью акустической томографии мелкого моря, необходимо осуществлять наблюдение на относительно небольших дистанциях при повышенных требованиях к пространственному разрешению [1–5]. В этих случаях целесообразно использовать высокочастотные (ВЧ) звуковые поля [6–9].

Поскольку в океане одновременно присутствуют большое число различных неоднородностей, возникает задача выделения неоднородностей одних типов на фоне других. Примером может служить наблюдение локализованных неоднородностей на фоне случайно распределённых неоднородностей поверхности океана, возникающих вследствие ветрового волнения. При этом случайно распределённые неоднородности образуют помехи. Другим источником помех являются аддитивные шумы океана [1, 2, 10].

Для оптимального построения подобного рода акустических систем необходимо разработать физико-математические, а на их основе — компьютерные модели конкретных районов наблюдения. Это позволит проводить компьютерные эксперименты, варьируя параметры зондирующих сигналов, конструкцию источников и приёмников, а также условия наблюдения, с целью подбора оптимальных элементов и алгоритмов томографических систем подводного акустического наблюдения.

В настоящей работе исследуются задача о структуре ВЧ поля, рассеянного объектом наблюдения — телом, помещённым в рефракционный плоскослоистый волновод океанического типа. Решение этой задачи позволит проанализировать возможности развития высокочастотных акустических систем томографического типа, включающих в себя набор акустических источников и приёмников, располагающихся в пределах ограниченного по размерам района наблюдений. Построив в дальнейшем модели шумов и помех, можно будет определить зону уверенного наблюдения объектов (поле зрения системы).

1. РАССЕЯНИЕ ПОЛЯ НА ТЕЛАХ В ЛУЧЕВОМ ПРИБЛИЖЕНИИ

1.1. Приближение Кирхгофа

Для описания океанического волновода воспользуемся моделью плавно-неоднородной слоистой по глубине среды. В этом случае для распространяющейся моды с частотой ω имеем соотношение $\omega^2/c^2 = k_0^2 n_{\rm c}^2(\mathbf{r})$, где $k_0 = \omega/c_0$, $n_{\rm c}(\mathbf{r}) = c_0/c(z)$, $c(z)$ — локальное значение скорости звука,

$c_0 = 1{,}46$ км/с — характерное значение скорости, $\mathbf{r} = (x, y, z)$ — радиус-вектор произвольной точки пространства. Обозначая потенциал скорости смещения частиц U, можно вывести для него уравнение Гельмгольца [2, 11], в правой части которого располагается функция $f_0(\mathbf{r})$, определяющая пространственное распределение спектральных компонент сторонних источников:

$$\Delta U + k_0^2 n_{\mathrm{c}}^2(\mathbf{r}) U = f_0(\mathbf{r}).$$

Кроме рефракционной водной среды в океаническом волноводе присутствуют также границы $\partial\Omega_0$: верхняя (поверхность) и нижняя (дно). В общем случае можно считать границы импедансными, так что поле на них должно удовлетворять условиям

$$\left.\left(\alpha_0 U + \beta_0 \frac{\partial U}{\partial n}\right)\right|_{\partial\Omega_0} = 0, \tag{1}$$

где $\alpha_0(r)$, $\beta_0(r)$ определяют условия в различных точках, $\mathbf{n}$ — нормаль к границе. В случае океанического волновода обычно считается, что верхняя граница является свободной (плотностью воздуха пренебрегают) и для неё приближённо выполняется условие абсолютно мягкой границы. Дно океанического волновода в реальных условиях имеет сложную слоистую структуру. Для его описания часто используют упрощённые модели, в частности, считают дно импедансной границей. Ещё более простое приближение — модель абсолютно жёсткого дна [2, 10].

По определению функция Грина G_0 является откликом среды в точке наблюдения $\mathbf{r}$ на воздействие точечного источника с координатой $\mathbf{r}_S$:

$$\Delta G_0 + k^2 n_0^2(\mathbf{r}) G_0 = \delta(\mathbf{r} - \mathbf{r}_S), \quad \mathbf{r} \in \Omega_0; \qquad \left.\left(\alpha_0 G_0 + \beta_0 \frac{\partial G_0}{\partial n}\right)\right|_{\partial\Omega_0} = 0, \tag{2}$$

В рассматриваемой нами задаче в океаническом волноводе присутствуют также малые тела либо неровности стенок волновода, имеющие криволинейные границы. Такие неоднородности вызывают дополнительное рассеяние акустических волн [7, 12]. Существуют также слабые и плавные неоднородности, не имеющие резких границ. К такому типу неоднородностей можно отнести случайно распределённые объёмные вариации акустических параметров среды, возникающие в результате турбулентных пульсаций, воздействия ветрового волнения и других факторов, связанных с движением морской среды [2, 10]. Как уже указывалось, различные типы неоднородностей присутствуют в океанических волноводах одновременно.

Рассмотрим подробнее структуру высокочастотного гидроакустического поля в океаническом волноводе в случае, когда в волновод помещено тело. Возмущённое поле при наличии тел удовлетворяет неоднородному волновому уравнению Гельмгольца и граничным условиям на свободной поверхности и дне. Кроме того, возмущённое поле должно удовлетворять также и граничным условиям на поверхности тела $\partial\Omega_1$, которые записываются аналогично условиям на границах волновода. При этом внутренняя краевая задача для неоднородных волноводов может быть сформулирована следующим образом: возмущённое поле, представляющее собой сумму невозмущённой составляющей и дифрагированного поля, должно удовлетворять уравнению Гельмгольца и граничным условиям на границах волновода и поверхности возмущающего тела:

$$\begin{gathered} \Delta U + k^2 n_0^2(\mathbf{r}) U = f_0(\mathbf{r}), \qquad \mathbf{r} \in \Omega_0 \setminus \Omega_1; \\ \left.\left(\alpha_0 U + \beta_0 \frac{\partial U}{\partial n}\right)\right|_{\partial\Omega_0} = 0; \qquad \left.\left(\alpha_1 U + \beta_1 \frac{\partial U}{\partial n}\right)\right|_{\partial\Omega_1} = 0. \end{gathered} \tag{3}$$

Система уравнений (3) является формальной постановкой задачи о рассеянии акустического поля на теле в волноводе океанического типа. Для дальнейшего анализа рассмотрим указанную задачу в эквивалентной интегральной форме. Для этого применим вторую интегральную теорему Грина [11] к функциям U и G_0 с учётом того, что границы не имеют общих точек, и в любой точке границы $\partial\Omega_0$ выражение в соответствующем граничном интеграле тождественно равно нулю:

$$U(\mathbf{R}) - \iiint G_0 f_0 \,\mathrm{d}\xi = \iint\limits_{\partial\Omega_0 \cup \partial\Omega_1} \left(G_0 \frac{\partial U}{\partial n} - U \frac{\partial G_0}{\partial n} \right) \mathrm{d}S_\xi = \iint\limits_{\partial\Omega_1} \left(G_0 \frac{\partial U}{\partial n} - U \frac{\partial G_0}{\partial n} \right) \mathrm{d}S_\xi . \tag{4}$$

Отсюда получаем следующее выражение для возмущённого поля в точке наблюдения через функцию Грина невозмущённого волновода и значений возмущённого поля и его нормальной производной на поверхности тела (уравнение Липпмана—Швингера) [13]:

$$U(\mathbf{R}) = U_0(\mathbf{R}) + \iint\limits_{\partial\Omega_1} \left(G_0 \frac{\partial U}{\partial n} - U \frac{\partial G_0}{\partial n} \right) \mathrm{d}S_\xi \equiv U_0(\mathbf{R}) + \iiint G_0 f_\mathrm{s} \,\mathrm{d}\xi, \tag{5}$$

где

$$f_\mathrm{s} \equiv \frac{\partial U}{\partial n}\, \delta_\mathrm{s} + \frac{\partial}{\partial n}\,(U \delta_\mathrm{s})$$

— обобщённая поверхностная плотность вторичных источников поля на поверхности тела, δ_s — поверхностная δ-функция [3], $\mathbf{n}$ — нормаль к поверхности тела, $U_0(\mathbf{R})$ — невозмущённое поле в волноводе от источника зондирующего поля (подсветки) f.

При рассмотрении импедансных тел приближённо можно считать, что рассеяние происходит на локальных не зависящих друг от друга элементах поверхности тела. Суммарное рассеянное телом поле можно приближённо представить как совокупность всех полей, отражённых от подобных локальных площадок. Более сложная картина формирования рассеянного поля имеет место в случае упругого тела, когда под воздействием стороннего поля тело колеблется сообразно своей форме и внутреннему строению [12].

Во многих случаях в качестве достаточного приближения можно использовать модель касательной плоскости — приближение Кирхгофа [11]. Согласно этому приближению отражение и преломление квазиплоской волны на локальной площадке поверхности тела происходит так же, как на плоскости, касательной к поверхности тела в центральной точке рассматриваемой площадки. Используя для отражённой волны данное приближение, можно заменить возмущённое поле U в подынтегральном выражении (5) на сумму падающего невозмущённого поля U_0 и отражённого поля. Это позволяет свести задачу о нахождении рассеянного телом поля к вычислению интегралов от известных функций. Разумеется, указанное приближение налагает свои ограничения на форму и размеры тела: необходимо, чтобы в масштабах нескольких длин волн падающего поля подсветки поверхность тела можно было считать плоской. В рамках приближения Кирхгофа для возмущённого поля на поверхности получаем выражения

$$U \approx (1 + V)\, U_0, \qquad \frac{\partial U}{\partial n} \approx (1 - V)\, \frac{\partial U_0}{\partial n}\,, \tag{6}$$

где V — локальный коэффициент отражения в точке поверхности. В простейших случаях тело можно считать идеальным: абсолютно неподатливым ($V = 1$) или абсолютно мягким ($V = -1$).

Для обобщённой плотности вторичных источников в приближении Кирхгофа с учётом (6) получим

$$f_\mathrm{s} \approx (1 - V)\, \frac{\partial U_0}{\partial n}\, \delta_\mathrm{s} + (1 + V)\, \frac{\partial}{\partial n}\,(U_0 \delta_\mathrm{s}). \tag{7}$$

1.2. Лучевое представление прямого и рассеянного сигнала

В дальнейших рассуждениях для описания высокочастотного поля используем модель лучевого поля, или приближение геометрической акустики [2, 6]. При волновом толковании мы будем придерживаться, лучи образуют только геометрическую структуру, на которую «нашивается» поле. Рассеяние поля на неоднородностях проявляется, прежде всего, в отклонении лучей от невозмущённых траекторий. Для неоднородностей с резкими изменениями акустических параметров, т. е. имеющих чётко выраженные границы, рассеяние лучевого поля включает в себя отражение и преломление лучей (в случае полупрозрачных импедансных тел), а также соскальзывание с кромок тел и другие дифракционные эффекты [6–8].

В приближении геометрической акустики поле в рефракционном волноводе представляется в виде суммы лучевых компонент. Амплитуды компонент определяются мощностью источника, формой спектра излучаемого сигнала, а также особенностями распространения лучей в рефракционной среде, т. е. фактором фокусировки и расстоянием между соответствующими точками. Существенную роль играет также отражение от границ волновода и затухание в среде, определяющие интегральный фактор ослабления поля.

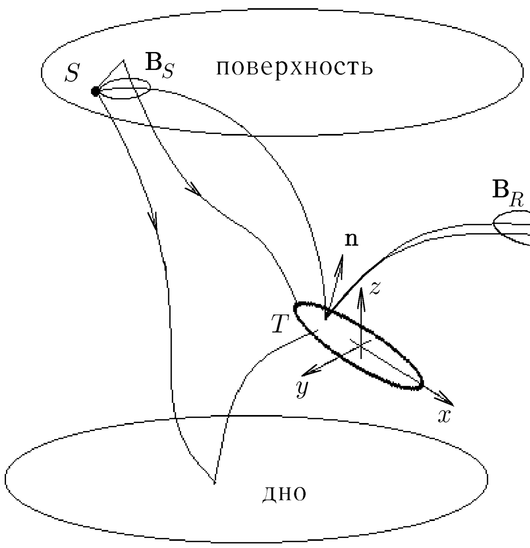


Рис. 1. Геометрия задачи о рассеянии высокочастотного поля телом в волноводе

Рассмотрим особенности формирования ВЧ лучевого поля, рассеянного криволинейными поверхностями и телами в плоскослоистой рефракционной среде. Пусть поле подсветки возбуждается источником, расположенным в точке S, а наблюдение осуществляется приёмником, расположенным в точке R (рис. 1).

В лучевом приближении решение уравнения Гельмгольца сводится к решению уравнений лучевой акустики: уравнения эйконала, которое определяет геометрию лучей, и уравнения переноса энергии поля вдоль лучевых трубок. Вместе с траекторией луча уравнение эйконала определяет также и время пробега сигнала по лучу. В неоднородной среде часть лучей многократно отражается от границ волновода. Суммарное поле можно представить в виде суммы лучей двух типов. К первому типу относятся водные лучи, которые не выходят на границы. Ко второму типу относятся лучи, отражающиеся от поверхности и/или дна океанического волновода. Обычно при отражении от дна поле существенно ослабляется. В этой связи в поле выделяют группу из водных лучей и лучей, отражающихся только от поверхности, несущих основную энергию поля.

Для вычисления поля, рассеянного криволинейной поверхностью, в точке наблюдения необходимо найти блики — точки отражения лучей. Аналогично при исследовании поля преломлённых лучей необходимо найти точки преломления. В рамках геометрической теории дифракции (ГТД) могут быть рассмотрены дополнительные компоненты рассеянного поля, связанные с особенностями строения поверхности отражающего тела [7, 8] (кромки, рёбра, острия и пр.).

При падении луча на криволинейную поверхность соответствующая ему лучевая трубка изменяет свою расходимость. Например, при отражении от выпуклой поверхности трубка расширяется, и поэтому поле ослабляется. При увеличении кривизны поверхности поле ослабляется больше. Задача вычисления амплитуды лучевого поля, отражённого от криволинейной поверхности в неоднородной среде, изучалась в достаточно общей постановке (см. [6–8]). Полученные аналитические выражения довольно сложны (за исключением некоторых практически важных асимптотических случаев), и на практике часто приходится использовать приближённые численные методы. Случай слабонеоднородной среды рассмотрен в работе [14]. Более простые выражения приведены в [6] для случая рассеяния лучевых полей криволинейными поверхностями в однородной среде.

В лучевом приближении функция Грина (2) записывается в виде

$$G_0(\mathbf{r}_S;\mathbf{r};\omega)=\sum_{\mu=1}^{M}\hat{A}_\mu(\mathbf{r}_S;\mathbf{r})\,\exp[-i\omega t_\mu(\mathbf{r}_S;\mathbf{r})]\,, \tag{8}$$

где M — число лучей, соединяющих источник и приёмник, $\hat{A}_\mu(\mathbf{r}_S;\mathbf{r})$ — комплексные амплитуды волн, приходящих по этим лучам, $t_\mu(\mathbf{r}_S;\mathbf{r})$ — время распространения акустической волны по лучу от точки источника $\mathbf{r}_S$ до текущей точки $\mathbf{r}$. Амплитуда волны, распространяющейся вдоль μ-го луча, определяется следующим выражением:

$$\hat{A}_\mu(\mathbf{r}_S;\mathbf{r})=\frac{\sqrt{f_\mu(\mathbf{r}_S;\mathbf{r})}}{R(\mathbf{r}_S;\mathbf{r})}\,\hat{V}_\mu, \tag{9}$$

где f_μ — геометрический фактор фокусировки соответствующей лучевой трубки, $R(\mathbf{r}_S;\mathbf{r})$ — расстояние между соответствующими точками в горизонтальной плоскости, $\hat{V}_\mu$ — интегральный фактор ослабления поля в океаническом волноводе с учётом затухания в толще жидкости (с декрементом $\beta_{\rm at}$) и потерь при отражении от поверхности ($\hat{V}_i^{(\rm surf)}$) и дна ($\hat{V}_j^{(\rm bot)}$):

$$\hat{V}_\mu=\prod_{i,\,j}\hat{V}_i^{(\rm surf)}\hat{V}_j^{(\rm bot)}\,10^{-\beta_{\rm at}l_\mu(\mathbf{r}_S;\mathbf{r})}\,, \tag{10}$$

где произведение берётся по всем точкам отражения луча от дна и свободной поверхности волновода, $l_\mu(\mathbf{r}_S;\mathbf{r})$ — длина дуги луча.

Обычно для формирования акустических сигналов, оптимизированных для решения конкретных практических задач, применяют протяжённые излучающие и приёмные антенны и сложные (чаще всего широкополосные импульсные) сигналы. Используя лучевое представление функции Грина (8), запишем формулу для принимаемого поля. С этой целью введём диаграммы направленности излучающей и приёмной антенн по амплитуде: $\mathbf{B}_S(\omega,\mathbf{e}_\mu^{(S)})$ и $\mathbf{B}_R(\omega,\mathbf{e}_\mu^{(R)})$, где $\mathbf{B}_S^2$, $\mathbf{B}_R^2$ — коэффициенты концентрации антенн, $\mathbf{e}_\mu^{(S)}$, $\mathbf{e}_\mu^{(R)}$ — единичные касательные векторы лучей в центрах антенн. С учётом указанных определений, для спектральной амплитуды давления прямого сигнала в точке $\mathbf{r}_l$ размещения приёмного гидрофона антенны в отсутствие рассеивающего тела можно записать

$$\hat{P}_S(\mathbf{r}_l,\omega)=\sqrt{W\rho_0c_0}\,\hat{F}_0(\omega)\sum_{\mu=1}^{M}A_\mu(\mathbf{r}_S;\mathbf{r}_R)\mathbf{B}_S\big(\omega,\mathbf{e}_\mu^{(S)}\big)\exp[-i\omega t_\mu(\mathbf{r}_S;\mathbf{r}_l)]. \tag{11}$$

Здесь W — суммарная мощность излучающей антенны, ρ_0c_0 — импеданс воды, $\hat{F}_0(\omega)$ — спектральная амплитуда излучаемого сигнала, l — индекс приёмного гидрофона. Время распространения сигнала до данного гидрофона можно приближённо записать в виде

$$t_\mu(\mathbf{r}_S;\mathbf{r}_l)\approx t_\mu(\mathbf{r}_S;\mathbf{r}_R)-\big(\mathbf{e}_\mu^{(R)},\mathbf{r}_l-\mathbf{r}_R\big)/c_0. \tag{12}$$

Для широкополосных импульсов каждая спектральная компонента испытывает при распространении трансформацию, описываемую выражением (11). При использовании согласованной фильтрации прямой сигнал на выходе фильтра может быть записан в виде

$$P_S(\mathbf{r}_l, t) = \frac{1}{2\pi} \int\limits_{-\infty}^{+\infty} \hat{P}_S(\mathbf{r}_l, \omega) \hat{F}_0^*(\omega) \exp(-i\omega t)\, \mathrm{d}\omega, \tag{13}$$

где $\hat{F}_0^*(\omega)$ является комплексно-сопряжённым спектром излучаемого сигнала. В частности, при излучении сложных импульсов, после их свёртки с репликой (комплексно-сопряжённым излучаемым импульсом), на выходе получится сжатый импульс для каждого из лучей. Суммарный сигнал будет набором лучевых импульсов, часть из которых будет перекрываться.

Рассмотрим структуру рассеянного телом эхо-сигнала. Учитывая вид функции Грина невозмущённого волновода (8), для спектральной амплитуды суммарного рассеянного сигнала в точке $\mathbf{r}_l$ получим

$$\hat{P}_T(\mathbf{r}_l, \omega) = \sqrt{W\rho_0 c_0}\, \hat{F}_0(\omega) \sum_{\mu=1}^{M} \sum_{\mu'=1}^{M'} A_\mu(\mathbf{r}_S; \mathbf{r}_T) A_{\mu'}(\mathbf{r}_T; \mathbf{r}_R) \mathbf{B}_S\bigl(\omega, \mathbf{e}_{\mu'}^{(S)}\bigr) \times$$
$$\times R_T\bigl(\mathbf{e}_\mu^{(T)}; \mathbf{e}_{\mu'}^{(T)}\bigr) \exp\bigl[-i\omega\, (t_\mu(\mathbf{r}_S; \mathbf{r}_T) + t_{\mu'}(\mathbf{r}_T; \mathbf{r}_l))\bigr], \tag{14}$$

где M и M' — числа лучей, соединяющих фазовые центры источника и приёмника с точкой отражения T на поверхности тела (речь идёт о лучах, впервые пересекающих поверхность, т. е. падающих на поверхность в данной точке), R_T — эквивалентный радиус цели (радиус абсолютно отражающей сферы, создающей в зоне Фраунгофера сигнал с такой же амплитудой), $\mathbf{e}_\mu^{(T)}$ и $\mathbf{e}_{\mu'}^{(T)}$ — направления лучей от источника и приёмника в точке T (см. рис. 1). Для эквивалентного радиуса цели будем использовать следующее асимптотическое (геометроакустическое) выражение [6, 8]:

$$R_T\bigl(\mathbf{e}_\mu^{(T)}; \mathbf{e}_{\mu'}^{(T)}\bigr) = \bigl|\Gamma\bigr| \bigl|K(\mathbf{n}_{\mu\mu'})\bigr|^{-1/2}, \qquad \mathbf{n}_{\mu\mu'} = -\bigl(\mathbf{e}_\mu^{(T)} + \mathbf{e}_{\mu'}^{(T)}\bigr) \Big/ \bigl\|\mathbf{e}_\mu^{(T)} + \mathbf{e}_{\mu'}^{(T)}\bigr\|, \tag{15}$$

где $K(\mathbf{n})$ — полная (гауссова) кривизна тела в той точке (блика), где внешняя нормаль к поверхности $\mathbf{n}_{\mu\mu'}$ обеспечивает отражение приходящей от источника волны с направлением $\mathbf{e}_\mu^{(T)}$ в направлении приёмника $-\mathbf{e}_{\mu'}^{(T)}$, Γ — коэффициент отражения от элемента поверхности тела в точке T, зависящий от угла падения луча на поверхность, $\|\mathbf{e}\|$ — длина вектора $\mathbf{e}$. Условия применимости данной модели определяются неравенствами

$$|\lambda| \ll L \approx \rho_1 \approx \rho_2,$$

где λ — длина акустической волны, L — линейный размер тела, ρ_1 и ρ_2 — главные радиусы кривизны поверхности в точке отражения.

Из выражений (14), (15) видно, что для вычисления эхо-сигнала необходимо найти точки бликов, от которых отражаются лучи, пришедшие из источника и попадающие в приёмник, вычислить в них радиусы кривизны поверхности тела и коэффициенты отражения, вычислить амплитуды лучевых компонент, соответствующих рассеянным (только отражённым в рассматриваемом нами приближении) лучам, и сложить все такие компоненты.

2. АНАЛИЗ РАССЕЯННОГО ПОЛЯ

Задача о нахождении рассеянных на неоднородностях высокочастотных полей, прежде всего, приводит к проблеме построения лучей, отражённых от тела и пришедших в точку наблюдения, т. е. к задаче нацеливания [1, 8, 15, 16]. Для этого необходимо найти все точки на поверхности тела (точки бликов), для которых отражённый луч попадает в приёмник. В рефракционной среде такого рода задача не является тривиальной. Она требует перебора большого числа лучей и характеризуется большой чувствительностью к изменениям их параметров. Сложность задачи заключается также в её принципиальной трёхмерности, что не позволяет напрямую применить для её решения стандартные методы уточнения (например, дихотомии).

В случае малых в сравнении с характерным масштабом лучевой картины размерах тела, представляющем основной практический интерес, более эффективными в сравнении с прямым перебором являются специальные методы. Задачу нацеливания можно ставить как задачу отбора пар лучей, соединяющих в неоднородной среде точки источника и приёмника с поверхностью отражателя (либо рефрактора), которые в общей точке соприкосновения с поверхностью удовлетворяют соответствующим условиям сопряжения. Такого рода условия определяются типом луча (отражённый, преломлённый, соскальзывающий и т. д.). Использование дополнительной информации о задаче позволяет сузить множества возможных точек поверхности, в которых могут реализовываться условия сопряжения.

Для произвольных неоднородных сред и локальных неоднородностей не существует универсального алгоритма априорного определения направлений лучей на источник (приёмник) в точке соприкосновения. Однако при малых размерах неоднородности в качестве таковых могут быть выбраны направления лучей, соединяющих некоторую заранее выбранную («центральную») точку тела с источником и приёмником. Точка поверхности, в которой выполнены условия сопряжения для указанных направлений, принимается за начальную при итерационном поиске искомой точки. Далее процесс повторяется до тех пор, пока в найденной точке поверхности условия сопряжения не будут выполнены с заданной точностью. В достаточно общем случае описанный итерационный процесс имеет геометрическую скорость сходимости со знаменателем, равным отношению среднего радиуса кривизны тела к среднему расстоянию до источника и приёмника.

Величина эхо-сигнала будет определяться параметрами гидроакустического волновода (в частности, зависимостью скорости звука от глубины) и формой рассеивателя. В дальнейшем анализе в качестве тела будем рассматривать абсолютно жёсткий вытянутый эллипсоид вращения. При этом величина эхо-сигнала может существенно варьироваться как при изменении положения эллипсоида относительно источника и приёмника, так и при изменении его размеров. Такого рода вариации оказывают существенное влияние на возможности измерения или наблюдения сигнала в океанических волноводах, поскольку эхо-сигнал в реальных условиях всегда будет измеряться на фоне реверберационных помех и аддитивных шумов моря. При анализе возможностей зондирования шельфовых морских зон высокочастотными акустическими сигналами важной является оценка эффективности измерений или наблюдений, т. е. оценка области параметров, где эхо-сигналы от эллипсоида могут быть измерены с заданной достоверностью. По аналогии с оптическим видением такую область можно назвать полем зрения системы наблюдения. В рамках настоящей работы в качестве системы наблюдения будем рассматривать лишь одну из томографических проекций, состоящую из одной излучающей и одной приёмной антенн (бистатическая схема). Проанализируем структуру рассеянного эллипсоидом поля и попытаемся исследовать поле зрения рассматриваемой нами схемы наблюдения в предположении малости реверберационных помех.

Для эллипсоида вращения точка блика находится из выражения

$$\mathbf{r}(\mathbf{n}) = \left(a^2 n_x \mathbf{i} + b^2 n_y \mathbf{j} + c^2 n_z \mathbf{k}\right) \left(a^2 n_x^2 + b^2 n_y^2 + c^2 n_z^2\right)^{-1/2}, \tag{16}$$

где $\mathbf{i}$, $\mathbf{j}$ и $\mathbf{k}$ — орты декартовой системы координат с началом в центральной точке эллипсоида, направленные вдоль его главных осей, $\mathbf{n} = (n_x, n_y, n_z)$ — нормаль в данной точке, a, b и c — длины соответствующих полуосей. Для гауссовой кривизны эллипсоида вращения ($b = c$) в точке (x, y, z) получаем

$$K = a^2 \left[\frac{c^2}{a^2}\, x^2 + \frac{a^2}{c^2}\left(y^2 + z^2\right)\right]^{-2}. \tag{17}$$

Затухание звука в морской среде определяется тепловыми потерями, вязкостью жидкости, рассеянием на мелкоструктурных неоднородностях и др. Существует целый ряд эмпирических формул, удовлетворительно описывающих частотные зависимости затухания звука в морской среде. На низких частотах (менее 3 кГц) обычно используется известная формула Торпа; на более высоких частотах — формула Марша—Шулькина [2]. Мы в своих расчётах применяем упрощённую эмпирическую формулу [9]

$$\beta_{\text{at}} = \begin{cases} 0{,}0279\, f^{3/2}, & f \le 6; \\ 0{,}0114\, f^2, & f > 6, \end{cases} \tag{18}$$

где частота f измеряется в кГц, а коэффициент затухания β_{at} — в дБ/км.

Вследствие влияния ветра поверхность океана является случайно неровной [2, 11]. Рассеяние сигналов на ней обычно рассматривается в рамках общей теории рассеяния на шероховатых поверхностях [11]. Для определения статистических характеристик сигналов, рассеянных на шероховатой поверхности, применяют эмпирические модели спектра ветрового волнения, в частности известную модель Пирсона—Московитца [2]. Не вдаваясь в детали, будем использовать упрощённую модель Релея [2, 11] для среднего коэффициента отражения от взволнованной поверхности океана:

$$\hat{V}^{(\text{surf})} = -\exp(-P^2/2), \tag{19}$$

где $P = 2k_0 \sigma_\xi \sin\chi$ — параметр Релея, χ — угол скольжения луча в точке падения на поверхность, σ_ξ^2 — дисперсия отклонений взволнованной поверхности, определяемая скоростью ветра [2, 11].

Для коэффициента отражения от дна в зеркальном направлении (когерентная компонента поля) используем формулу Френеля [11]:

$$\hat{V}^{(\text{bot})} = \frac{\hat{W} - 1}{\hat{W} + 1}, \qquad \hat{W} \equiv \frac{m \sin\chi}{n \sin\chi_1}, \qquad \cos\chi = n_{\text{t}} \cos\chi_1,$$

где m — отношение плотностей воды и грунта, n_{t} — отношение скоростей звука в воде и в грунте (модель дна в виде жидкого слоя осадков).

В качестве источника и приёмника рассмотрим вертикальные решётки с диаграммами направленности

$$\mathbf{B}^2(\omega, \mathbf{e}) = C \sin^2\left(N\pi\, \frac{d}{\lambda}\, \sin\theta\right) \Big/ \left[N^2 \sin^2\left(\pi\, \frac{d}{\lambda}\, \sin\theta\right)\right], \tag{20}$$

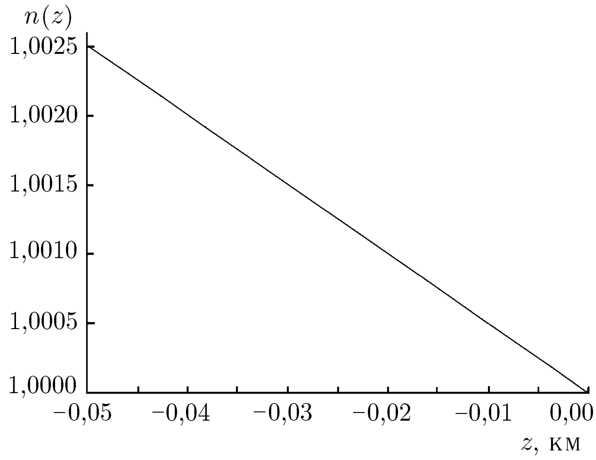


Рис. 2. Структура гидрологии рефракционного волновода

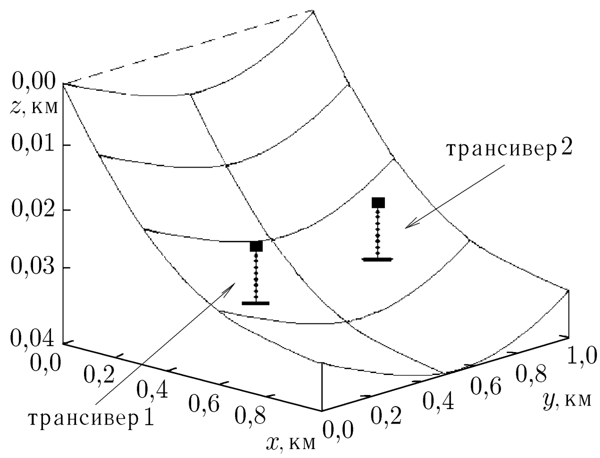


Рис. 3. Структура дна волновода

где N — число гидрофонов, d —расстояние между ними, $\sin\theta$ — вертикальная компонента вектора направления $\mathbf{e}$, C — безразмерный нормирующий коэффициент. Фактически, комплексные амплитуды лучей, выходящих из источника или приходящих к приёмнику под различными углами, будут промодулированы по амплитуде и фазе функцией $\mathbf{B}(\omega,\mathbf{e})$.

Рассмотрим зону наблюдения в горизонтальной плоскости в виде квадрата с ориентированными вдоль осей x и y сторонами в 1 км. В дальнейшем координаты в пределах указанного района будем обозначать в безразмерных величинах, нормированных на 1 км.

В качестве модели океанического волновода рассматривался типичный волновод придонного типа, который обычно реализуется в мелководных районах в летних условиях. Профиль скорости звука в волноводе аппроксимировался линейной зависимостью $n^2(z)$, скорости на границах слоя задавались из таблиц (рис. 2). Использовалась следующая модель криволинейного дна в виде параболоида вращения:

$$z = z_0 - \left[(x-x_0)^2 + (y-y_0)^2\right]/h^2,$$

где $z_0 = 0{,}04$; $x_0 = 1$; $y_0 = 0{,}5$; $h^2 = (x_0^2 + y_0^2)/z_0$ (см. рис. 3). Проводились также расчёты для более сложных моделей профиля скорости и дна, при этом использовалась кубическая аппроксимация таблично заданных функций одной и двух переменных. Взаимодействующие с дном лучи не рассматривались, поскольку они существенно ослабляются при отражении. Влияние неровного дна при этом проявлялось как ограничение глубины волновода в пределах рассматриваемой зоны наблюдения.

Для излучения и приёма сигналов использовались модели вертикальных излучающей и приёмной решёток, состоящих из двенадцати излучателей и приёмников, с расстоянием между элементами $\lambda_0/2$, где λ_0 — длина волны излучения. Источник располагался в точке S с координатами $(0{,}5; 0{,}25; 0{,}025)$, приёмник — в точке R с координатами $(0{,}5; 0{,}75; 0{,}025)$. Для согласования подсветки и приёма с условиями распространения для диаграмм направленности (20) на основании численных экспериментов были подобраны оптимальные углы компенсации: $(5^\circ, -90^\circ)$ для излучающей решётки и $(5^\circ, 90^\circ)$ для приёмной решётки, что обеспечивало эффективную засветку глубин $5 \div 10$ м.

В качестве зондирующего импульса $F(t) = \sqrt{W}\,F_0(t)$ рассматривался сигнал с линейной частотной модуляцией:

$$f_0(t) = \sqrt[4]{\frac{2}{\pi T^2}}\,\exp\!\left[-\frac{t^2}{T^2} + i\,(\omega_0 + \alpha t)\,t\right], \tag{21}$$

со спектральной плотностью

$$|F_0(\omega)|^2 \approx \frac{1}{\sqrt{2\pi s}} \exp\left[-\frac{(\omega-\omega_0)^2}{2s}\right],$$

где $s \equiv T^{-2} + \alpha^2 T^2/4$. При $\alpha T \gg T^{-1}$ полоса данного сложного импульсного сигнала $\Delta\omega = 2\pi\,\Delta f \approx \alpha T$; $\omega_0 = 2\pi f_0$. В расчётах выбирались следующие параметры импульса: центральная частота импульса $f_0 = 20\div30$ кГц ($\lambda_0 = 5\div7$ см), длительность сигнала $T = 1\div10$ с, полоса $\Delta f \approx 4\div6$ кГц.

В качестве модели рассеивателя рассматривался абсолютно жёсткий эллипсоид вращения с осями $2a = 2$ м, $2b = 2c = 1$ м. Следует отметить, что используемая нами модель эквивалентного радиуса цели (15) является в рамках использованных приближений предельно допустимой, учитывая, что в соответствии с (17) для указанного эллипсоида значения возможных эквивалентных радиусов цели лежат в интервале от 25 до 200 см, что соответствует диапазону от 5 до 40 длин волн излучения λ_0. Для тел ещё меньших (относительно длины волны) размеров следует использовать другие возможные модели эквивалентного радиуса цели, рассчитываемые в рамках более строгих волновых приближений [9, 10].

Для поиска точек бликов применялся описанный выше итерационный алгоритм. В качестве начального шага итерационного процесса поиска использована нормаль в точке пересечения поверхности тела биссектрисой, построенной для лучей, пришедших в центральную точку эллипсоида от источника и приёмника (см. рис. 1).

Уровень рассеянного сигнала определяется мощностью излучателей, затуханием в морской среде и ослаблением при распространении и рассеянии. При этом для приёма сигналов на фоне аддитивных шумов должна быть обеспечена заданная достоверность (вероятность обнаружения сигнала на фоне шумов) при заданной (например, гауссовой) статистике шумов при необходимом превышении уровня сигнала над уровнем шума. В численных экспериментах полагалось, что мощность излучающей решётки $W = 10$ Вт, уровень аддитивных шумов моря составлял 75 дБ [2, 10]. Как известно [11], для обнаружения сигналов на фоне шумов с вероятностью 0,95 при вероятности ложной тревоги 10^{-7} уровень сигнала должен превышать уровень шума на величину порядка 24 дБ. В дальнейших рассуждениях будем называть область параметров задачи, при которых сигнал превышает уровень 100 дБ, полем зрения системы наблюдения.

Как указывалось в предыдущих разделах, коэффициент ослабления сигнала при отражении от взволнованной поверхности океана зависит от ветрового волнения. В расчётах коэффициент отражения когерентной компоненты поля определялся приближённым выражением (19), скорость ветра полагалась равной 10 м/с.

Основная задача, которая решалась в ходе численных экспериментов, — исследование зависимости рассеянных телом акустических сигналов от положения и ориентации эллипсоида.

При расположении эллипсоида в различных точках относительно источника S и приёмника R лучи будут подходить к рассеивающей поверхности с различных сторон. При этом точки бликов будут перемещаться по поверхности эллипсоида с изменяющейся кривизной. При удалении от линии SR рассеянное поле будет уменьшаться, что объясняется как уменьшением угловых размеров рассеивателя, так и перемещением бликовых точек. Для исследования рассеянного поля будем помещать эллипсоид в различные точки рефракционной среды так, чтобы ориентация эллипсоида не изменялась. При этом рассмотрим сечения трёхмерной зоны чувствительности (поля зрения) системы различными координатными плоскостями (рис. 4).

Для интерпретации конфигурации зон чувствительности рассмотрим лучи, формирующие поле. На рис. 5 показана картина лучей от источника и приёмника, из которой видно, что в рассматриваемой модели волновода в каждую точку приходит набор лучей. Рассеивающий эллипсоид на рис. 5 показан сильно увеличенным (в реальном масштабе он был бы неразличим). Всю область можно разбить на подобласти, в которых результирующее поле формируется различным числом лучей. В областях, расположенных над источником и приёмником, поле мало́ за счёт влияния диаграммы направленности излучающей и приёмной решёток. На рис. 5 это условно показано как области, в которых лучи отсутствуют. В ближней зоне источника и приёмника поле формируется только лучами, не касающимися дна и поверхности (водными). Начиная с некоторых расстояний поле формируется двумя типами лучей: водными и отражёнными от поверхности (поверхностными). В придонном волноводе существенная часть лучей будет отражаться от дна (донные лучи). Как указывалось выше, в настоящем исследовании донные лучи, как и многократно отражённые лучи ввиду существенного ослабления их амплитуд рассматриваться не будут. Отметим, что при возбуждении поля в ограниченном интервале углов при наблюдении на фиксированной глубине начиная с некоторых расстояний поле в результате влияния рефракции существенно ослабляется. Такие области геометрической тени формируются главным образом вблизи поверхности. При помещении эллипсоида в различные точки среды лучи будут падать на поверхность под различными углами и отражаться от различных участков поверхности эллипсоида. Число лучей также будет меняться. Поскольку рассматриваемый в расчётах эллипсоид вытянут вдоль одной из координат, эквивалентный радиус цели будет существенно меняться при перемещении точки блика по его поверхности. При выбранных параметрах эллипсоида эквивалентный радиус цели может меняться приблизительно от 0,25 м при попадании луча на острые концы эллипсоида до 2 м при отражении от более плоских боковых участков эллипсоида. При отражении совокупности лучей результирующий эквивалентный радиус цели приближённо можно характеризовать его средним значением.

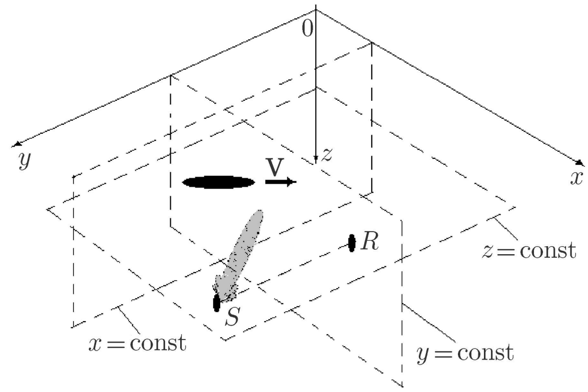


Рис. 4. Расположение плоскостей, секущих область чувствительности при наблюдении

При падении на эллипсоид двух лучей (водного и поверхностного), от источника и от приёмника, что возможно в области, расположенной между источником и приёмником, могут реализоваться четыре канала связи источника и приёмника. Степень связанности каждой из возможных пар лучей будет зависеть от того, существует ли соответствующая точка блика на поверхности эллипсоида, а также от кривизны поверхности в соответствующей точке блика.

Для удобства анализа полезно выделить потенциальные характеристики связанности точек волновода — трансляционные характеристики волновода [8]. Одной из таких характеристик служит произведение интенсивностей полей, возбуждаемых точечными излучателями с единичной амплитудой, помещёнными в точку расположения источника и приёмника:

$$V\,(\mathbf{r}_S \,|\, \mathbf{r} \,|\, \mathbf{r}_R) = |G_0(\mathbf{r}_S, \mathbf{r}) G_0(\mathbf{r}, \mathbf{r}_R)|^2 \,. \tag{22}$$

При суммировании лучей следует учесть особенности распространения ВЧ поля в океанических волноводах. Поскольку при отражении от случайно неоднородной границы волновода фаза волн флуктуирует, то интерференционная структура, вызываемая суперпозицией лучей,

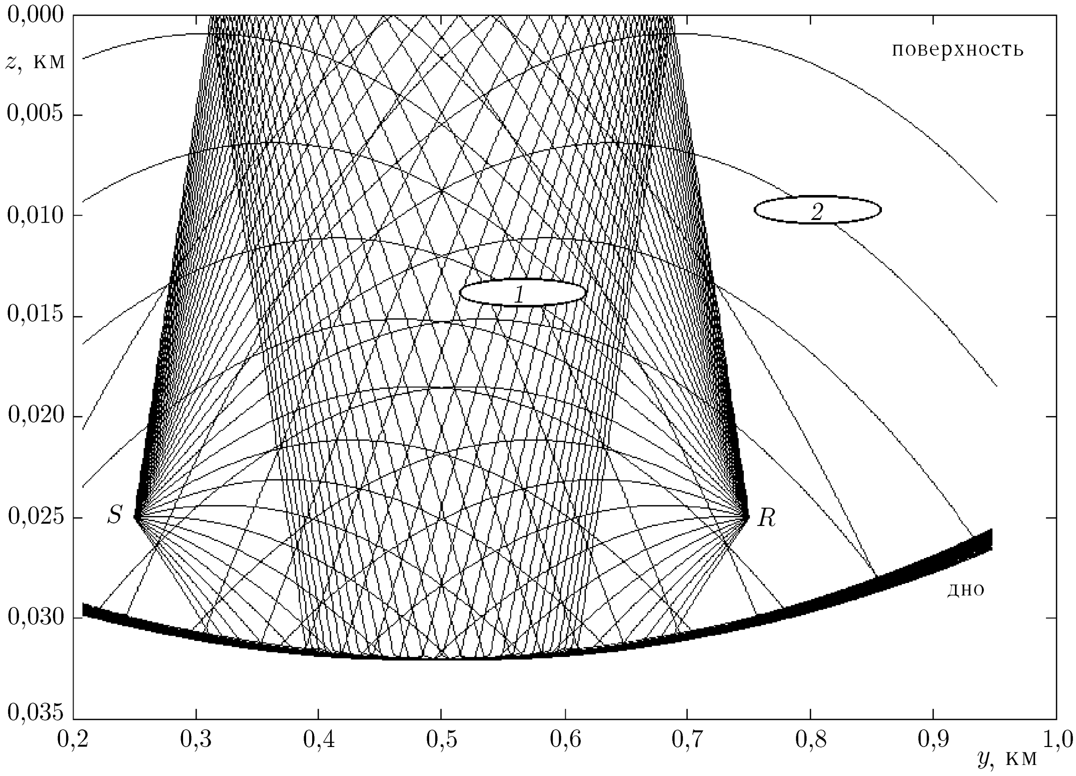


Рис. 5. Структура лучей при численных расчётах; *1* и *2* соответствуют различным положениям тела

будет случайным образом варьироваться во времени. При оценке поля в таких условиях целесообразно воспользоваться некогерентным сложением лучей. В наших вычислениях рассматривалось рассеянное поле как при когерентном, так и при некогерентном суммировании.

Для того, чтобы более наглядно выявить влияние анизотропии формы эллипсоида и разделить факторы, существенные при формировании рассеянного поля, рассмотрим прежде всего рассеяние от сферы единичного радиуса, помещая её в различные точки волновода. При этом рассчитываем трансляционные характеристики, эквивалентные радиусы цели и другие характеристики рассеянного поля.

Как видно из результатов вычислений, структура трансляционных характеристик при когерентном и некогерентном сложении лучей для рассматриваемой геометрии задачи в предположении малости вклада донных лучей отличается несущественно. Это можно объяснить малым отличием задержек водных и поверхностных лучей, отражённых от тела, при распространении по трассе от источника к телу и далее к приёмнику.

Из вычислений видно, что существует верхняя квазисферическая граница поля зрения, за пределами которой трансляционные характеристики равны нулю. Подобного рода зоны возникают вследствие того, что лучи (за исключением донных) не попадают в эти области из-за рефракции. Конические области над источником и приёмником и под ними также остаются незасвеченными из-за влияния диаграммы направленности излучающей антенны (в соответствии с (20) источник и приёмник имеют направленность лишь в вертикальной плоскости).

Ослабление поля подсветки объясняется расширением лучевых трубок по мере увеличения расстояния. В пространственном распределении трансляционных характеристик проявляется также и интерференция водных и поверхностных лучей. В частности, при когерентном сложении проявляется ослабление поля вблизи поверхности, возникающее за счёт интерференции водных и по-

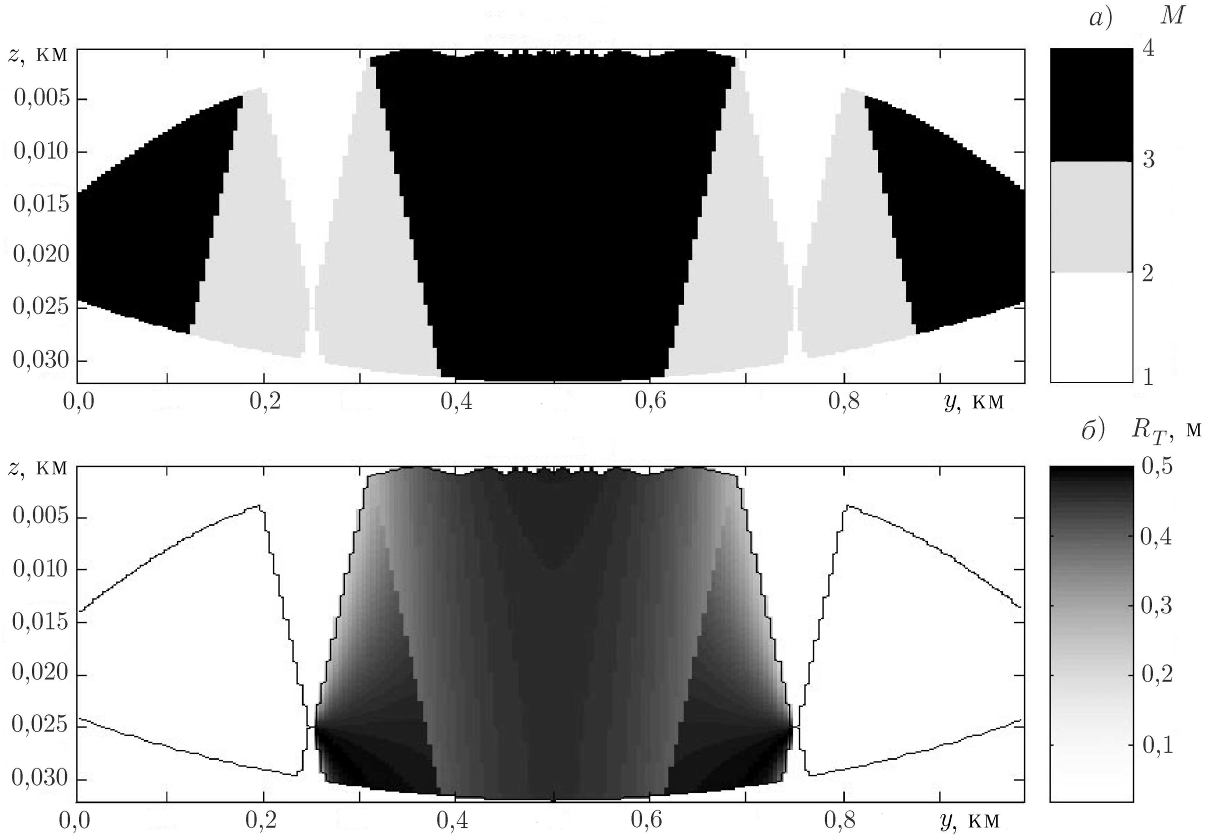


Рис. 6. Пространственные распределения числа лучей M (*а*) и эквивалентного радиуса цели R_T (*б*) при перемещении эллипсоида в плоскости yz при $x = 0{,}5$ км

верхностных лучевых компонент. Поскольку эквивалентный радиус цели для рассеивающего тела в виде шара постоянен, структура эхо-сигнала совпадает со структурой трансляционной функции. Таким образом, ВЧ поле, рассеянное сферическим телом в рефракционном волноводе, формируется только за счёт фактора распространения (перемножения факторов распространения от источника и приёмника).

При формировании рассеянного поля от отражателя в виде вытянутого эллипсоида существенную роль играет также изменение положения точек отражения на поверхности тела при изменении ракурса подсветки из точек расположения источника и приёмника. Это приводит к изменению кривизны поверхности тела в точке отражения, определяющей амплитуду отражённого поля каждой из лучевых компонент. В этом случае при формировании рассеянного поля важными являются все факторы: характеристики распространения в рефракционном волноводе; положение и ориентация тела относительно источника и приёмника в рефракционном волноводе; анизотропия и строение тела [17].

При проведении численных расчётов были исследованы пространственные распределения числа лучей, эквивалентных радиусов, трансляционных характеристик и эхо-сигналов в сечениях зоны чувствительности координатными плоскостями x, y, z (рис. 4) при различных положении и ориентации эллипсоида.

Как следует из представленных расчётов, число лучей и трансляционные характеристики остаются практически такими же, как и при расчётах с шаром, поскольку перечисленные функции определяются только характеристиками волновода и положением рассеивателя. Однако для вытянутого эллипсоида распределения эквивалентного радиуса и связанные с ними распределения эхо-сигналов существенно изменяются. В горизонтальных сечениях при ориентации эллипсоида вдоль оси x эквивалентный радиус цели изменяется в интервале от 0,25 до 2 м. Максимальные значения эквивалентный радиус принимает в области между источником и приёмником, а также при отражении назад под малыми углами от боковых стенок эллипсоида. По мере удаления от источника и приёмника эквивалентный радиус цели уменьшается, поскольку в большинстве случаев точки бликов смещаются к острым концам эллипсоида (см. рис. 6).

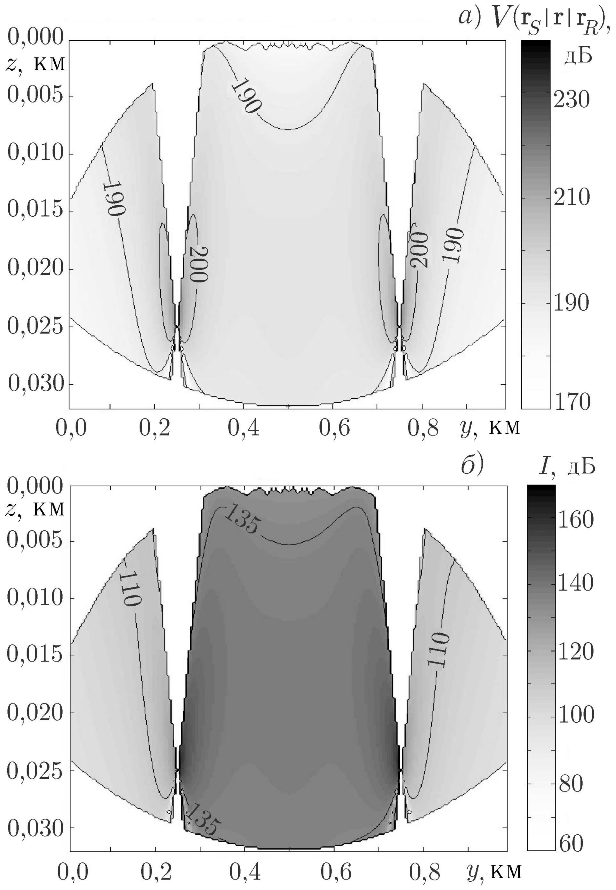


Рис. 7. Пространственное распределение трансляционной характеристики (*а*) и интенсивности эхо-сигнала I (*б*) при перемещении эллипсоида в плоскости yz при $x = 0{,}5$ км

Эхо-сигнал формируется с учётом как распространения в волноводе, так и рассеяния на эллипсоиде. Из расчётов спектральной амплитуды эхо-сигнала на центральной частоте импульса (рис. 7) видно, что в горизонтальной плоскости сигнал максимален для областей, прилегающих к вертикальной плоскости SR, проходящей через источник и приёмник. При изменении ориентации эллипсоида область бóльших значений эхо-сигналов перемещается вследствие изменения пространственного распределения эквивалентного радиуса цели. В частности, в случае, когда эллипсоид ориентирован вдоль плоскости SR, эквивалентный радиус цели принимает бóльшие значения, когда эллипсоид находится между источником и приёмником. Лучи отражаются от боковых стенок эллипсоида и при смещении эллипсоида от плоскости SR. По мере приближения тела к источнику или приёмнику эквивалентный радиус цели быстро уменьшается вследствие смещения точек бликов к острым концам эллипсоида. Интересен случай, когда эллипсоид, источник и приёмник находятся в одной горизонтальной плоскости (рис. 5). В этом случае при расположении тела на линии SR около источника или приёмника бóльшая часть водных лучей отражается от боковой поверхности. Однако начиная с некоторых расстояний от источника или приёмника водные лучи на горизонте расположения эллипсоида становятся пологими. При этом они начинают отражаться от точек, расположенных в районе острой носовой части эллипсоида, где эквивалентный радиус цели меньше. При дальнейшем удалении тела от источника или приёмника начинают играть роль и поверхностные лучи.

При ориентации эллипсоида под углом к линии SR также возникают области, где эквивалентный радиус цели велик. Вид и положение этих зон обусловливаются тем, что в этих областях лучи отражаются от боковой поверхности эллипсоида. При увеличении глубины расположения эллипсоида одна из таких зон исчезает, что связано с влиянием дна. На рис. 8 показано изменение формы зон с бóльшим уровнем эхо-сигнала в горизонтальной плоскости в зависимости от

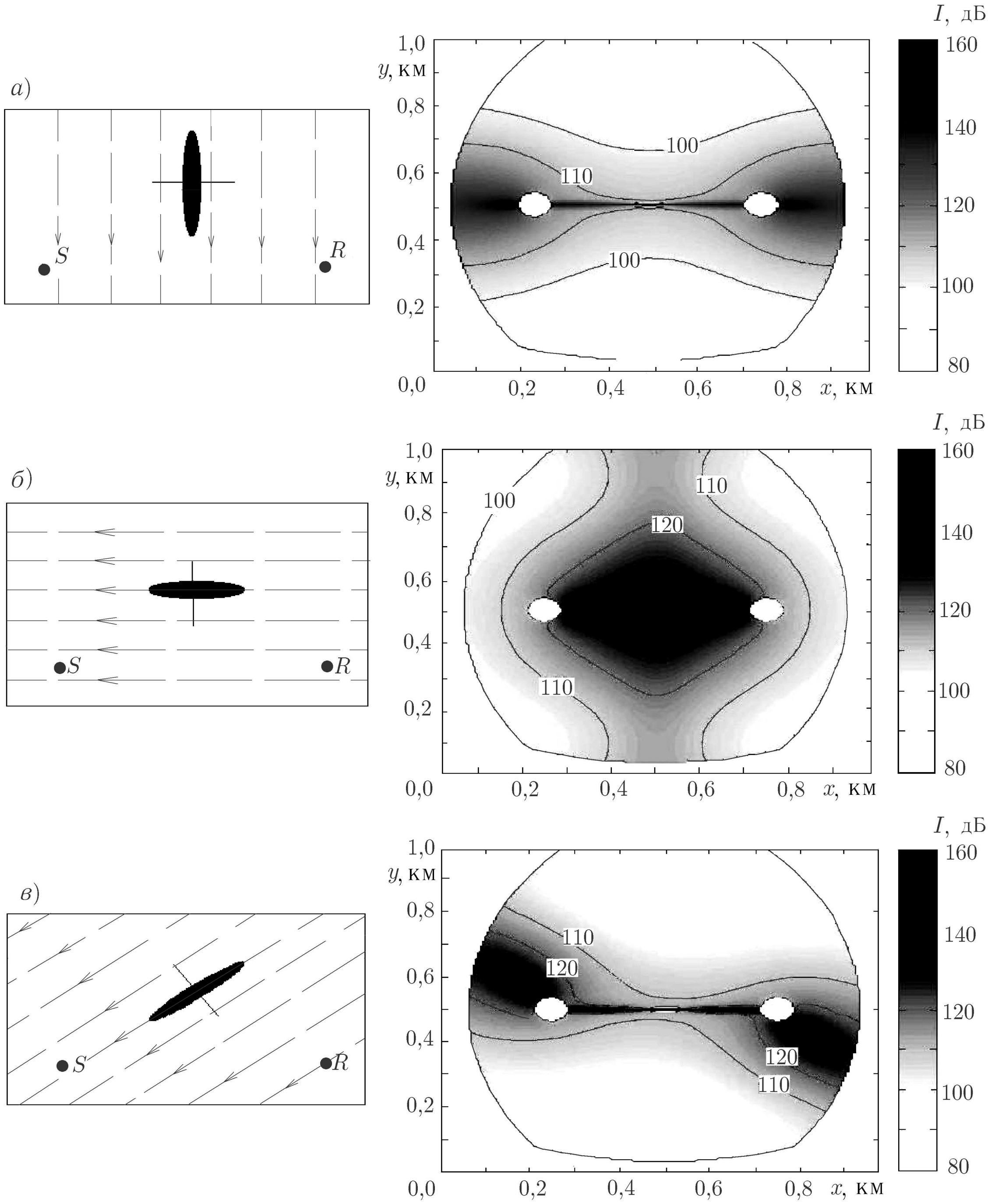


Рис. 8. Пространственное распределение областей повышенного уровня эхо-сигнала I при перемещении эллипсоида вдоль различных траекторий относительно источника и приёмника в горизонтальной плоскости $z = 10$ м. Ориентация эллипсоида и совокупность траекторий его движения (штриховые линии) показаны в левых частях панелей *а*–*в*

ориентации эллипсоида. Видно, что при изменении угла поворота эллипсоида относительно вертикальной плоскости, проходящей через линию SR, области повышенных значений эхо-сигнала перемещаются, располагаясь всегда по нормали к оси, вдоль которой вытянут эллипсоид. Аналогичные эффекты наблюдаются и в сечениях по осям x и y.

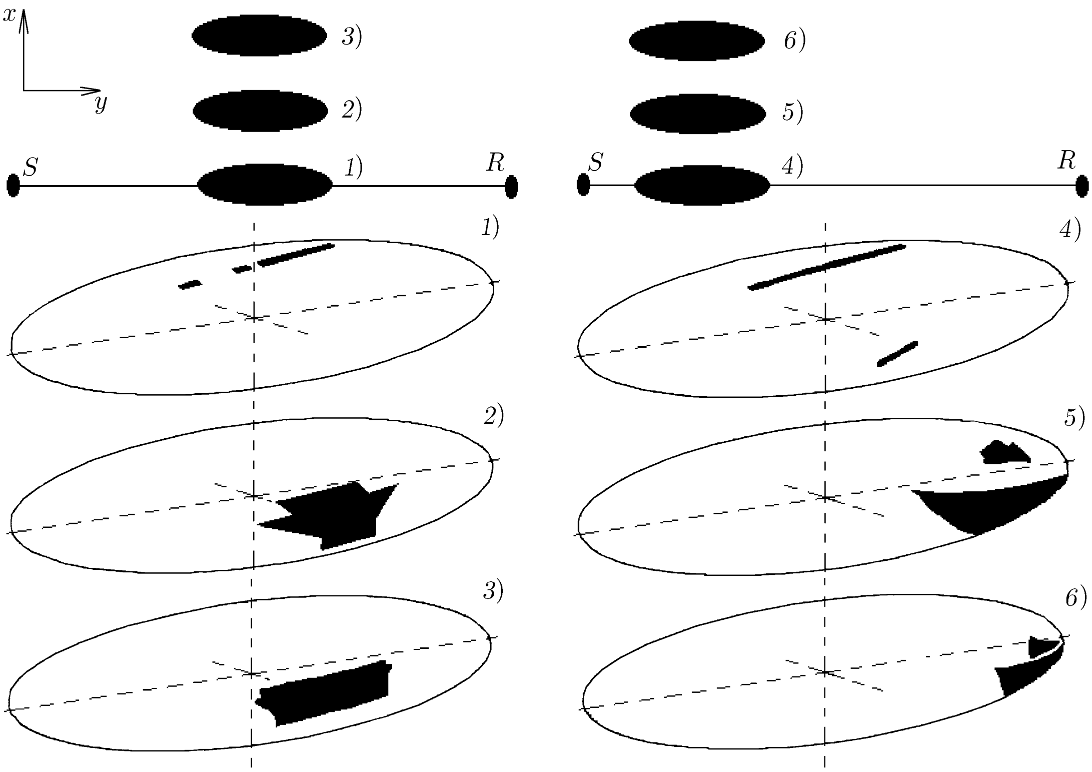


Рис. 9. Полученные численно области локализации бликов (тёмные области) на поверхности эллипсоида при различном его расположении в плоскости xy относительно источника S и приёмника R

Дополнительную информацию о структуре рассеянного поля несут аналогичные распределения, построенные для координатных плоскостей y. В такого рода сечениях можно отметить влияние диаграммы направленности источника и приёмника. При приближении к источнику и приёмнику поле концентрируется по вертикали вблизи углов компенсации. По мере удаления от источника и приёмника (максимальное удаление соответствует точке, лежащей посередине между источником и приёмником) максимум поля поднимается к поверхности волновода. Этот эффект проявляется и в структуре эхо-сигнала.

На рис. 9 показано смещение бликов на поверхности эллипсоида при перемещении эллипсоида. Видно, что при увеличении смещения от середины трассы точки бликов смещаются к острым носовым областям эллипсоида, что приводит к уменьшению амплитуды отражённого сигнала.

3. ЗАКЛЮЧЕНИЕ

В работе в рамках лучевого приближения исследована структура акустических сигналов, рассеянных телами в рефракционных волноводах океанического типа с неровным дном. В численных расчётах рассматривалось тело в виде абсолютно жёсткого эллипсоида вращения, для которого были рассмотрены особенности формирования рассеянных сигналов в зависимости от характеристик плоскослоистого рефракционного волновода с кривым дном, а также от параметров эллипсоида.

Аналитически и на основе численных расчётов установлено, что при расположении эллипсоида под ненулевым углом к линии источник—приёмник формируется несимметричное пространственное распределение областей с бóльшим значением эхо-сигнала. Показано, что важным параметром при измерении эхо-сигналов являются углы компенсации диаграмм направленности источника и приёмника в вертикальной плоскости. Их изменение приводит к различной эффективности засветки области расположения эллипсоида в фиксированных интервалах глубин. Установлено, что оптимальные углы компенсации зависят от гидрологии и геометрии задачи. При рассмотренных в работе параметрах задачи соответствующее значение углов компенсации оказалось равным 9°.

Работа осуществлена при поддержке РФФИ (проект № 00–02–17157).

СПИСОК ЛИТЕРАТУРЫ

1. Гончаров В. В., Зайцев В. Ю., Куртепов В. Н. и др. Акустическая томография океана. Н. Новгород: ИПФ РАН, 1997. 254 с.
2. Бреховских Л. М., Лысанов Ю. П. Теоретические основы акустики океана. Л.: Гидрометеоиздат, 1982. 264 с.
3. Смирнов И. П., Карузерс Дж. В., Хилько А. И. Томографическое наблюдение локализованных неоднородностей в плоскослоистых волноводах: Препринт № 505 ИПФ РАН. Н. Новгород, 1999. 25 с.
4. Смирнов И. П., Смирнов А. И., Карузерс Дж. В., Хилько А. И. Томографическая реконструкция локализованных неоднородностей в океанических волноводах: Препринт № 538 ИПФ РАН. Н. Новгород, 2000. 26 с.
5. Бурдуковская В. Г., Лучинин А. Г., Хилько А. И. Томографическая реконструкция неоднородностей океанических волноводов с помощью маломодовых акустических импульсов: Препринт № 535 ИПФ РАН. Н. Новгород, 2000. 27 с.
6. Кравцов Ю. А., Орлов Ю. И. Геометрическая оптика неоднородных сред. М.: Наука, 1980. 304 с.
7. Боровиков Л. Д., Кинбер А. С. Геометрооптическая теория дифракции. М.: Наука, 1981.
8. Зорин А. Ю., Смирнов И. П., Хилько А. И. // Формирование акустических полей в океанических волноводах. Реконструкция неоднородностей. Н. Новгород: ИПФ РАН, 1994. С. 214.
9. Borodin V. V., Galaktionov M. Yu. // The Formation of Acoustical Fields in Oceanic Waveguides. Reconstructions of Inhomogeneities in Shallow Water. Nizhny Novgorod: Inst. Appl. Phys. RAS, 1998. V. 2. P. 259.
10. Урик Р. Д. Основы гидроакустики. Л.: Судостроение, 1978. 444 с.
11. Рытов С. М., Кравцов Ю. А., Татарский В. И. Введение в статистическую радиофизику. Часть 2. Случайные поля. М.: Наука, 1978.
12. Горский С. М., Зверев В. А., Хилько А. И. // Формирование акустических полей в океанических волноводах. Н. Новгород: ИПФ РАН, 1991. С. 82.
13. Горюнов А. А., Сасковец А. В. Обратные задачи рассеяния в акустике. М.: Изд-во МГУ, 1989. 152 с.
14. Жидко Ю. М. // Изв. вузов. Радиофизика. 1969. Т. 12, № 8. С. 1 205.
15. Смирнов А. И., Хилько А. А. // Труды (пятой) научной конференции по радиофизике, посвященной 100-летию со дня рождения А. А. Андронова. 7 мая 2001 г. Н. Новгород: ТАЛАМ, 2001. С. 258.
16. Пермитин А. В., Шарова А. Л. Нацеливание лучей в геометрической оптике: Препринт № 287 ИПФ РАН. Н. Новгород, 1999. 30 с.
17. Смирнов И. П., Хилько А. А., Смирнова И. Р. // Труды Нижегородской акустической научной сессии. Н. Новгород: ТАЛАМ, 2002. С. 88.

Институт прикладной физики РАН, г. Нижний Новгород, Россия

Поступила в редакцию 7 марта 2003 г.

MODELLING OF HIGH-FREQUENCY ACOUSTIC FIELDS SCATTERED BY BODIES IN REFRACTIVE WAVEGUIDES

I. P. Smirnov, A. A. Khil'ko, and A. I. Khil'ko

We study the structure of the sound field scattered by a body (observed object) in a refractive plane-layered oceanic waveguide with a curved bottom. The solution of the problem allows one to analyze the possibilities of development of a acoustic tomography system with finite numbers of sources and receivers based at a given bounded region of the ocean.

УДК 621.385.6

ИССЛЕДОВАНИЕ НЕЛИНЕЙНОЙ ДИНАМИКИ ЛБВ-ГЕНЕРАТОРА С ЗАПАЗДЫВАЮЩЕЙ ОБРАТНОЙ СВЯЗЬЮ

Н. М. Рыскин

Представлены результаты численного моделирования нелинейной динамики генератора на лампе бегущей волны с запаздывающей обратной связью. Рассмотрены свойства стационарных режимов одночастотной генерации, подробно изучены процессы возникновения автомодуляции. Проведено моделирование сценариев перехода к хаосу по мере увеличения параметра неравновесности, пропорционального току электронного пучка. Показано, что основным сценарием является разрушение квазипериодического движения. В то же время торможение пучка в сильно нелинейных режимах вызывает переходы через перемежаемость к режимам на базе мод с более высокими частотами. Конкуренция этих сценариев приводит к сложной картине регулярных и хаотических автомодуляционных режимов в пространстве параметров, что является типичным для электронно-волновых распределённых автогенераторов с запаздывающей обратной связью.

ВВЕДЕНИЕ

Исследование нелинейной динамики приборов вакуумной СВЧ электроники в последние годы привлекает большое внимание в связи с возможными применениями в современных системах радиолокации и связи, основанных на использовании динамического хаоса, в установках нагрева плазмы и т. д. Среди приборов, демонстрирующих сложное, в том числе хаотическое, поведение особое место занимают генераторы на лампе бегущей волны (ЛБВ) с внешней запаздывающей обратной связью, которые широко изучались как теоретически, так и экспериментально [1–9]. Тем не менее многие особенности нелинейной динамики таких приборов остаются неисследованными. В частности, в экспериментах наблюдаются все известные сценарии перехода к хаосу, характерные для конечномерных динамических систем: удвоение периода, разрушение квазипериодического движения, перемежаемость [5–7]. Однако неясно, какие факторы ответственны за тот или иной тип динамики. Что касается теоретических исследований, то в большинстве работ доминируют различные приближённые подходы, основанные, например, на описании при помощи точечных [1, 2] или функциональных [8, 9] отображений. Они служат, скорее, для качественного объяснения механизмов возникновения автомодуляции и хаоса. В то же время для указанных выше приложений представляет интерес построение достаточно полной картины нелинейной динамики, которая на сегодняшний день в литературе отсутствует. В частности, это актуально в связи с последними исследованиями, направленными на создание ЛБВ-генераторов миллиметрового и субмиллиметрового диапазонов длин волн с замедляющей системой типа «петляющий волновод» (folded waveguide) [10, 11], причём в [10] были экспериментально обнаружены разнообразные автомодуляционные и хаотические режимы генерации.

В настоящей работе излагаются подробные результаты численного моделирования уравнений нестационарной нелинейной теории ЛБВ-генератора, восполняющие отмеченные выше пробелы. Проводится сопоставление с полученными ранее теоретическими и экспериментальными результатами, а также с динамикой ряда других приборов: клистронов с запаздывающей обратной связью, распределённых параметрических генераторов, лампы обратной волны (ЛОВ) с отражениями.

1. ОСНОВНЫЕ УРАВНЕНИЯ

Заметим, что под ЛБВ-генераторами с запаздывающей обратной связью можно понимать, вообще говоря, различные приборы. Так, в [1, 3] рассматривается замкнутая в кольцо цепочка из двух ЛБВ, одна из которых работает в линейном, а другая — в нелинейном режиме. В работах [4–7] в цепь обратной связи включён узкополосный перестраиваемый фильтр, который ограничивает число возбуждаемых мод и существенным образом влияет на динамику системы. В данной работе рассматривается наиболее простая ситуация, когда внешняя обратная связь осуществляется посредством широкополосной линии передачи без дисперсии. Используем хорошо известные уравнения нестационарной нелинейной теории взаимодействия электронного потока с электромагнитной волной в случае узкополосного сигнала, которые для ЛБВ-генератора с запаздывающей обратной связью имеют вид

$$\frac{\partial^2\theta}{\partial\xi^2} = -L^2\,\mathrm{Re}[F\exp(i\theta)], \tag{1}$$

$$\frac{\partial F}{\partial\tau} + \frac{\partial F}{\partial\xi} = -\frac{L}{\pi}\int\limits_0^{2\pi}\exp(-i\theta)\,\mathrm{d}\theta_0. \tag{2}$$

В (1), (2) $\theta = \omega_0\,(t - x/v_0)$ — фаза электрона относительно волны, ω_0 — частота синхронизма, на которой фазовая скорость волны равна скорости пучка v_0, θ_0 — начальная фаза, F — нормированная медленно меняющаяся амплитуда волны с несущей частотой, равной ω_0, $L = 2\pi CN$ — безразмерный управляющий параметр, зависящий от тока пучка и длины пространства взаимодействия (здесь введены обычные в теории взаимодействия О-типа величины: C — параметр усиления Пирса, N — электрическая длина [12]). Безразмерные координата ξ и время τ введены следующим образом:

$$\xi = x/l, \qquad \tau = \frac{t - x/v_0}{l/v_\mathrm{g} - l/v_0}\,, \tag{3}$$

где l — длина системы, v_g — групповая скорость волны (предполагается, что $v_\mathrm{g} < v_0$). Более подробно нормировка переменных описана, например, в [13].

Уравнение (1) описывает движение электронов в поле электромагнитной волны, которое записано в приближении малого изменения скорости электронов в процессе взаимодействия, когда

$$\frac{1}{2\pi N}\,\frac{\partial\theta}{\partial\xi} \ll 1. \tag{4}$$

При этом (1) принимает универсальную форму, единую как для нерелятивистского, так и для релятивистского пучка, а в уравнениях (1), (2) остаётся единственный управляющий параметр L. Уравнение (2) описывает нестационарное возбуждение волны током медленно меняющейся амплитуды и справедливо в случае сигналов с достаточно узким спектром, когда можно пренебречь дисперсией групповой скорости и аппроксимировать дисперсионную характеристику в окрестности частоты синхронизма прямой линией. Эти приближения заведомо нарушаются, например, вблизи границы полосы пропускания замедляющей системы, где следует пользоваться другим вариантом теории [14, 15], в котором дисперсионная характеристика аппроксимируется параболой. С осторожностью можно говорить о применимости указанных приближений в случае широкополосной спиральной ЛБВ, где наиболее строгий подход заключается в отказе от приближения медленно меняющихся амплитуд [16]. Наиболее адекватно сформулированная система уравнений описывает приборы с умеренной полосой (например, релятивистские ЛБВ, ЛБВ с цепочкой связанных резонаторов или с замедляющей системой типа «петляющий волновод») вдали от границы полосы пропускания.

Систему (1), (2) следует дополнить граничными условиями для пучка, которые отражают тот факт, что пучок на входе системы не имеет модуляции ни по скорости, ни по плотности:

$$\theta\big|_{\xi=0} = \theta_0, \qquad \frac{\partial\theta}{\partial\xi}\bigg|_{\xi=0} = 0, \tag{5}$$

и граничным условием для поля:

$$F(\tau;\xi=0) = RF(\tau-\delta;\xi=1). \tag{6}$$

Здесь $R \equiv \rho\exp(i\psi)$ — комплексный параметр обратной связи, δ — нормированное время запаздывания. Предполагаем, что обратная связь осуществляется с помощью внешней линии передачи, через которую сигнал проходит за время Δt, причём считается, что Δt от частоты не зависит. В используемых здесь безразмерных переменных

$$\delta = \left(\Delta t + \frac{l}{v_0}\right)\frac{v_0 v_{\rm g}}{l\,(v_0 - v_{\rm g})}\,. \tag{7}$$

В принципе, эта величина может принимать любые положительные значения.

Приведённая выше система уравнений описывает не только генератор с внешней обратной связью, но и резонансную ЛБВ (ЛБВ с отражениями), если пренебречь взаимодействием пучка с несинхронной отражённой волной. В этом случае параметр R пропорционален произведению коэффициентов отражения от границ замедляющей структуры, а $\Delta t = l/v_{\rm g}$, так что

$$\delta = \frac{v_0 + v_{\rm g}}{v_0 - v_{\rm g}} > 1. \tag{8}$$

Для резонансной ЛОВ, напротив, $0 < \delta < 1$ [13].

Наконец, отметим, что уравнения нестационарной нелинейной теории лазера на свободных электронах (ЛСЭ) с нефиксированной структурой поля при определённых приближениях также могут быть приведены к виду (1)–(6) (см., например, [17–19]). Однако о применении данной модели к теории ЛСЭ следует говорить с определённой осторожностью. Для ЛСЭ безразмерное время запаздывания может быть очень велико, точнее, полоса усиления оказывается очень широкой, т. к. и скорость пучка, и групповая скорость волны близки к скорости света. Хотя учитывать частотную дисперсию групповой скорости теперь нет необходимости, существенную роль начинает играть дисперсия обратной связи [18, 19]. Для ЛБВ с запаздывающей обратной связью аналогичная ситуация реализуется, когда в цепь обратной связи включён узкополосный фильтр (резонатор) [4–7]; здесь эти эффекты не рассматриваются. Существенно также, что уравнения (1)–(6) справедливы только в случае $v_0 > v_{\rm g}$, тогда как для ЛСЭ это условие может не выполняться, если реализуется взаимодействие с незамедленными волнами.

2. УСЛОВИЯ САМОВОЗБУЖДЕНИЯ И СТАЦИОНАРНЫЕ РЕЖИМЫ ГЕНЕРАЦИИ

Прежде всего, кратко обсудим условия самовозбуждения генератора. Отметим, что колебательная система ЛБВ-генератора с запаздывающей обратной связью представляет собой кольцевой резонатор, частоты «холодных» собственных мод которого в принятых здесь обозначениях определяются выражением (см. [13])

$$\Omega_n = \frac{2\pi n + \psi}{1+\delta}\,, \tag{9}$$

где $n = 0, \pm 1, \pm 2, \ldots$. Динамика нестационарных процессов в значительной степени определяется конкуренцией этих мод. В принципе, строгое построение линейной нестационарной теории представляет собой достаточно сложную задачу, которая полностью может быть решена только численно (см., например, [20], где обсуждается постановка аналогичной задачи для ЛОВ-генератора). Однако можно достаточно просто получить приближённое решение, если условию самовозбуждения придать вид

$$\rho G(\Omega) \exp[i\,(\psi - \delta\Omega)] = 1, \tag{10}$$

где $G(\Omega)$ — коэффициент усиления ЛБВ-усилителя, Ω — нормированная отстройка от частоты синхронизма. Физический смысл соотношения (10) очевиден: сигнал, прошедший по кольцу обратной связи, должен поступать на вход усилителя в соответствующей фазе, а усиление должно превышать потери. Для $G(\Omega)$ в используемых здесь обозначениях имеем выражение

$$G(\Omega) = [1 - L^3\,(f_{\rm a} + i f_{\rm r})] \exp(-i\Omega), \tag{11}$$

где

$$f_{\rm a} = -\frac{2\,(1-\cos\Omega) - \Omega\sin\Omega}{\Omega^3}\,, \qquad f_{\rm r} = \frac{2\sin\Omega - \Omega\,(1+\cos\Omega)}{\Omega^3} \tag{12}$$

— известные выражения для активной и реактивной составляющих мощности взаимодействия электромагнитной волны и электронного потока, получаемые в рамках метода последовательных приближений [12]. Из этих соотношений можно найти пороговые значения параметра L (т. е. стартовый ток пучка) и частоту генерации. Простые качественные соображения позволяют заключить, что оптимальными для самовозбуждения являются такие значения фазы ψ параметра обратной связи, при которых одна из резонансных частот (9) близка к частоте, на которой активная мощность, отдаваемая пучком волне, максимальна (эта частота примерно равна $0{,}8\pi$). Наоборот, самовозбуждение затрудняется, когда эта частота лежит посередине между двумя собственными частотами. Таким образом, на плоскости параметров (L, ψ) будет наблюдаться система периодически расположенных зон генерации: граница зоны самовозбуждения при $\psi \approx 2\pi n + 0{,}8\pi\,(1+\delta)$ будет иметь минимумы, а при $\psi \approx 2\pi n + \pi + 0{,}8\pi\,(1+\delta)$ — максимумы. Такая картина типична для автоколебательных систем с запаздыванием: клистронов с запаздывающей обратной связью, ЛОВ с отражениями, параметрических генераторов и др. [13, 21–24].

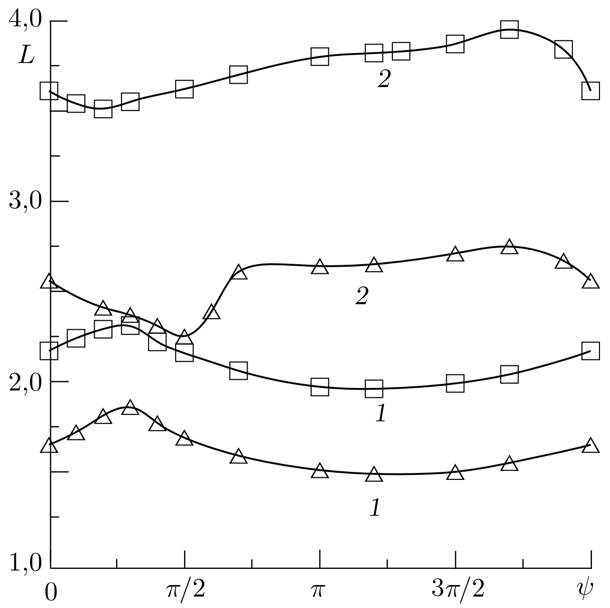


Рис. 1. Границы зон самовозбуждения (*1*) и автомодуляции (*2*) на плоскости параметров (L, ψ) при $\delta = 0{,}5$ и различных значениях ρ: $\square$ — $\rho = 0{,}5$; $\triangle$ — $\rho = 0{,}7$

Следует подчеркнуть, что эти рассуждения носят качественный характер, поскольку основываются на выражении для коэффициента усиления (11), полученном на первом шаге метода последовательных приближений. Оно достаточно хорошо согласуется с точным решением при $CN < 0{,}6$ (или $L < 3{,}77$), когда усиление не превышает 20 дБ [12]. Поскольку в данном случае

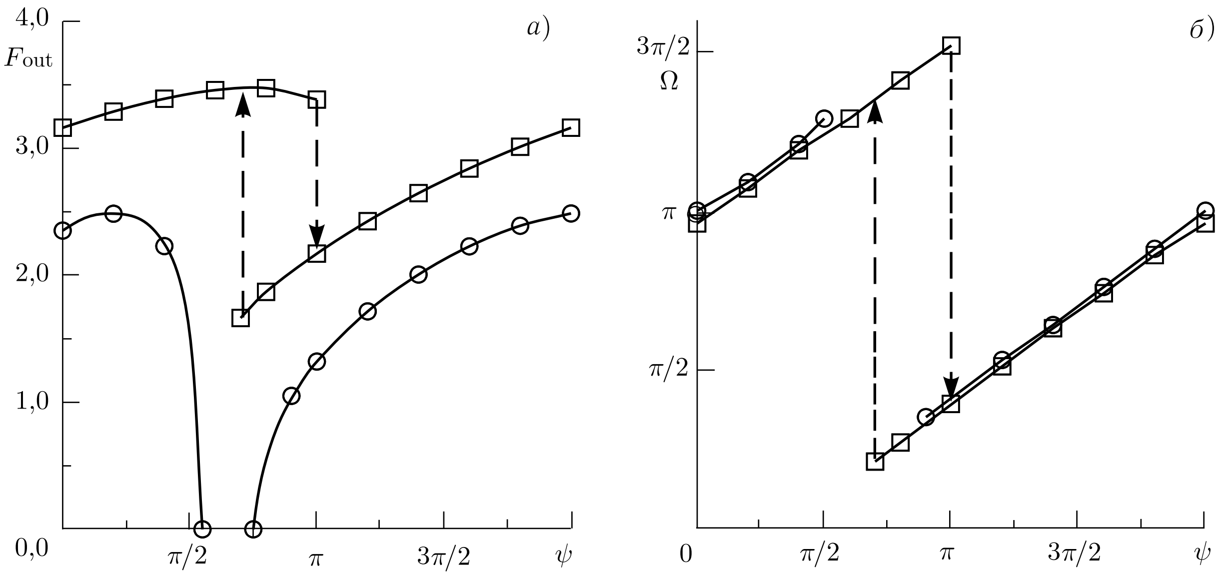
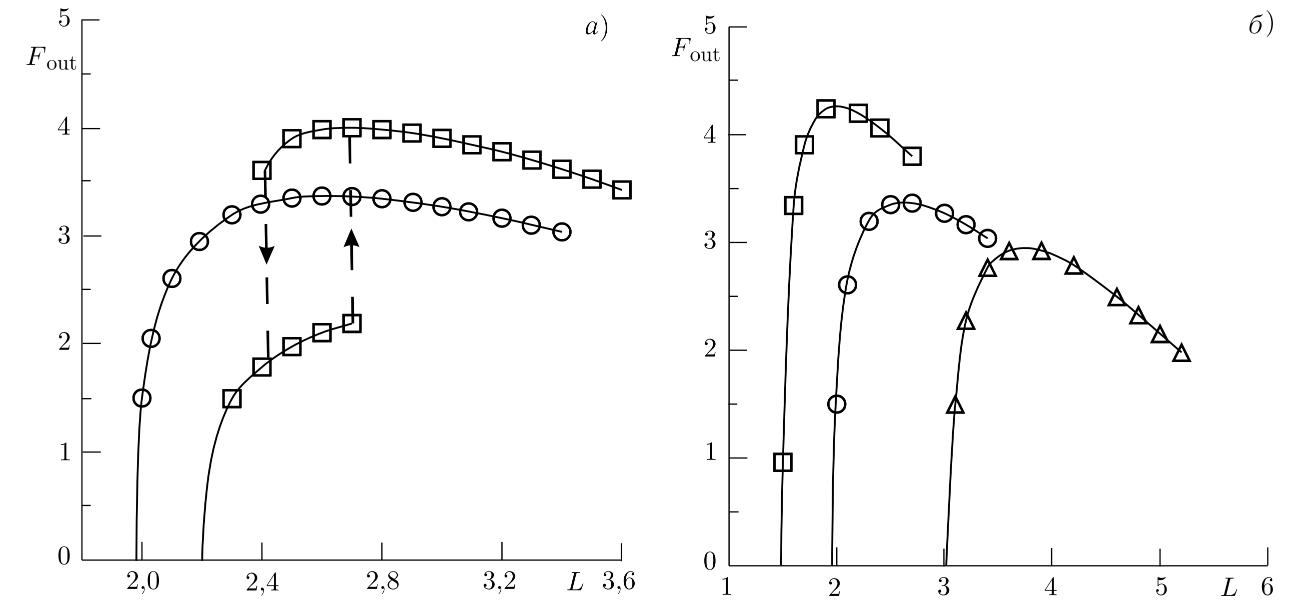


Рис. 2. Зависимости амплитуды (*а*) и частоты стационарной генерации (*б*) от ψ при $\delta = 1$, $\rho = 0{,}5$ и различных значениях L: $\bigcirc$ — $L = 2{,}1$; $\square$ — $L = 2{,}3$

речь идёт о пусковых условиях генератора, это одновременно означает, что глубина обратной связи должна быть не слишком малой. Если это условие выполнено, теоретические оценки хорошо соответствуют результатам численного моделирования. В качестве примера на рис. 1 приведены полученные численно границы зоны самовозбуждения на плоскости параметров (L, ψ) при $\delta = 0{,}5$ и различных значениях ρ. В силу того, что зависимость от ψ является периодической, достаточно ограничиться интервалом $0 < \psi < 2\pi$. На рис. 1 также построены границы зоны автомодуляции: линии, выше которых на плоскости параметров (L, ψ) стационарные режимы одночастотной генерации становятся неустойчивыми и сменяются многочастотными колебаниями. Центр зоны генерации расположен при $\psi \approx 1{,}3\pi{\div}1{,}4\pi$, а граница между двумя соседними зонами — при $\psi \approx 0{,}3\pi$. С увеличением ρ границы зон генерации становятся более резко выраженными и опускаются вниз, что обусловлено усилением резонансных свойств колебательной системы. Однако разность между максимальным и минимальным значениями L относительно невелика, и в целом зависимость от фазы выражена слабее, чем у резонансной ЛОВ при аналогичных значениях ρ [13]. Физически такое поведение обусловлено гораздо более плавной частотной зависимостью коэффициента усиления у ЛБВ-усилителя. Границы самовозбуждения на плоскости (L, ρ) при различных значениях ψ также имеют вид, типичный для усилителей с запаздывающей обратной связью: порог самовозбуждения монотонно убывает с ростом глубины обратной связи и стремится к нулю при $\rho \to 1$, а при малых ρ обращается в бесконечность. Соответствующие графики мы здесь не приводим, т. к. они вполне аналогичны представленным, например, в [17, 23].

Отметим, что для построения полной картины динамики ЛБВ-генератора нельзя ограничиваться каким-либо одним значением фазы параметра обратной связи, как это делается в ряде работ, например считать R чисто вещественным [8, 9, 17]. Фиксированные значения ψ могут соответствовать самым различным положениям относительно центра зоны генерации в зависимости от времени запаздывания δ. Поскольку динамика генератора в центре и на краю зоны может существенно различаться, необходимо рассматривать весь диапазон изменения фазы.

При превышении порога самовозбуждения переходный процесс заканчивается установлением стационарного одночастотного режима, в котором $F(\xi, \tau) = F(\xi) \exp(i\Omega\tau)$. Возбуждается мо-



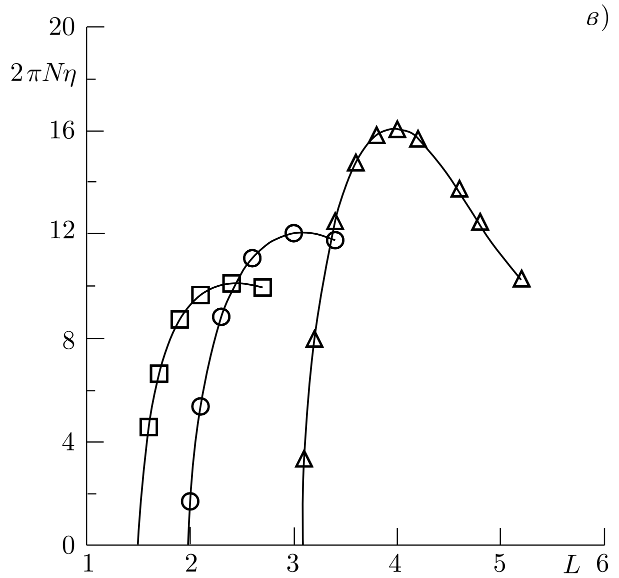


Рис. 3. Зависимости амплитуды выходного сигнала в стационарном режиме от L: ($а$) — $\delta = 0{,}5$; $\rho = 0{,}5$; $\psi = 0{,}5\pi$ (□) и $\psi = 1{,}5\pi$ (○); ($б$) — $\delta = 0{,}5$; $\psi = 1{,}5\pi$; $\rho = 0{,}2$ (△), $\rho = 0{,}5$ (○) и $\rho = 0{,}7$ (□); ($в$) — зависимости «приведённого» КПД, соответствующие рис. 3*б*

да, частота которой наиболее близка к $\Omega_{\max} \approx \approx 0{,}8\pi$, т. е. к частоте, на которой активная мощность взаимодействия электромагнитной волны и электронного потока максимальна. На рис. 2 приведены примеры зависимостей амплитуды выходного сигнала $F_{\text{out}} = |F(\xi = 1)|$ и частоты генерации от параметра ψ при различных значениях L; период зависимостей по ψ равен 2π. При достаточно малых L наблюдаются две отдельные зоны генерации. С ростом L зоны расширяются и начинают перекрываться. В области перекрытия наблюдается бистабильность: в некотором диапазоне параметров сосуществуют два устойчивых режима генерации. То, какой именно из этих режимов реализуется, определяется начальными условиями. При плавном изменении ψ происходят жёсткие переходы от одной моды к другой (они показаны штриховыми линиями), которые сопровождаются гистерезисом. Это типичная картина для автоколебательных систем с запаздыванием [13, 22–24], которая также хорошо согласуется с результатами экспериментальных исследований ЛБВ-генераторов (см., например, [4, 7]). Частота генерации, как видно из рис. 2*б*, зависит от L достаточно слабо. Отметим, что более высокочастотная мода имеет бóльшую амплитуду. Это связано с торможением пучка в нелинейном режиме, в результате чего он начинает более эффективно взаимодействовать с волнами, имеющими меньшую фазовую скорость (а следовательно, более высокую частоту, поскольку замедляющая система имеет нормальную дисперсию).

На рис. 3*а* приведены характерные зависимости амплитуды выходного сигнала от параметра L при различных значениях ψ: в центре и на краю зоны генерации. Вначале с увеличением L наблюдается быстрый рост амплитуды сигнала, затем происходит насыщение, после чего амплитуда начинает медленно уменьшаться (в отличие от ЛОВ, где автомодуляция возникает сразу же, как только амплитуда стационарной генерации достигает максимума). Одна из кривых демонстрирует наличие бистабильности: верхняя ветвь кривой соответствует моде с номером $n = 1$ (см. формулу (9)), нижняя — моде с номером $n = 0$. Наибольшее значение амплитуды выходного сигнала, а значит, бо́льшая выходная мощность и КПД соответствуют именно моде с номером $n = 1$. Аналогичный вид имеют и зависимости амплитуды стационарной генерации от параметра L при различных значениях ρ (рис. 3*б*). Для простоты выбрано такое значение ψ, при котором бистабильность отсутствует. Чем больше ρ, тем больше максимальная амплитуда выходного сигнала и тем меньше значения L, соответствующие порогам самовозбуждения и автомодуляции. Очевидно, что это обусловлено увеличением добротности колебательной системы с ростом ρ.

Практический интерес, однако, представляет не безразмерная амплитуда колебаний, а выходная мощность и КПД генерации. В данном случае можно показать, что КПД в стационарном режиме определяется соотношением

$$\eta = \frac{(1-\rho^2)\, C F_{\rm out}^2}{2}\,, \tag{13}$$

что позволяет найти мощность как $P = \eta P_0$, где P_0 — постоянная мощность пучка. Однако, строго говоря, нам известны не C и N по отдельности, а их произведение. Будем считать, что в эксперименте увеличивается ток пучка, так что N остаётся постоянным. Тогда можно построить графики «приведённого» КПД $2\pi N\eta = (1-\rho^2)\, L F_{\rm out}^2/2$ (рис. 3*в*). Видно, что наибольшие значения КПД соответствуют как раз меньшим значениям глубины обратной связи. Задав конкретное значение N, по этим графикам можно определить КПД. Однако необходимо учитывать, что мы используем уравнение движения (1), справедливое в условиях малого изменения энергии электронов (4), т. е. при низком КПД.

3. ВОЗНИКНОВЕНИЕ АВТОМОДУЛЯЦИИ

Изучение режимов автомодуляции начнём со случая небольшого времени запаздывания $\delta = 0{,}5$, который в основном рассматривался в предыдущем разделе. Во всех случаях автомодуляция вызывается жёстким возбуждением одной из собственных мод кольцевой резонансной колебательной системы. Такой механизм автомодуляции получил название частотного, т. к. он связан с наличием вогнутого участка на амплитудно-частотной характеристике усилителя [8, 9]. Частотный механизм характерен также для ЛСЭ, параметрических генераторов, ЛОВ при больших отражениях, линейных ускорителей с запаздывающей обратной связью [13, 21, 23, 25]. Важно, что переход к автомодуляции происходит жёстко, т. е. сразу же за порогом возникают глубокие пульсации амплитуды выходного сигнала. При этом в спектре имеется большое число гармоник частоты автомодуляции, которые примерно совпадают с частотами «холодных» собственных мод (9); такой режим можно интерпретировать как режим самосинхронизации мод [13, 21, 23]. Другой механизм автомодуляции, называемый амплитудным, связан с наличием крутого падающего участка на амплитудной характеристике усилителя; в этом случае автомодуляция возникает мягко [8, 9, 13, 22, 24]. Отметим, что в ЛОВ при больших значениях ρ доминирует частотный механизм, а при малых — амплитудный [13]. Напротив, в ЛБВ-генераторе частотный механизм имеет место при любых значениях ρ, однако можно реализовать амплитудный механизм, если включить в цепь обратной связи узкополосный фильтр, ограничивающий число возбуждаемых

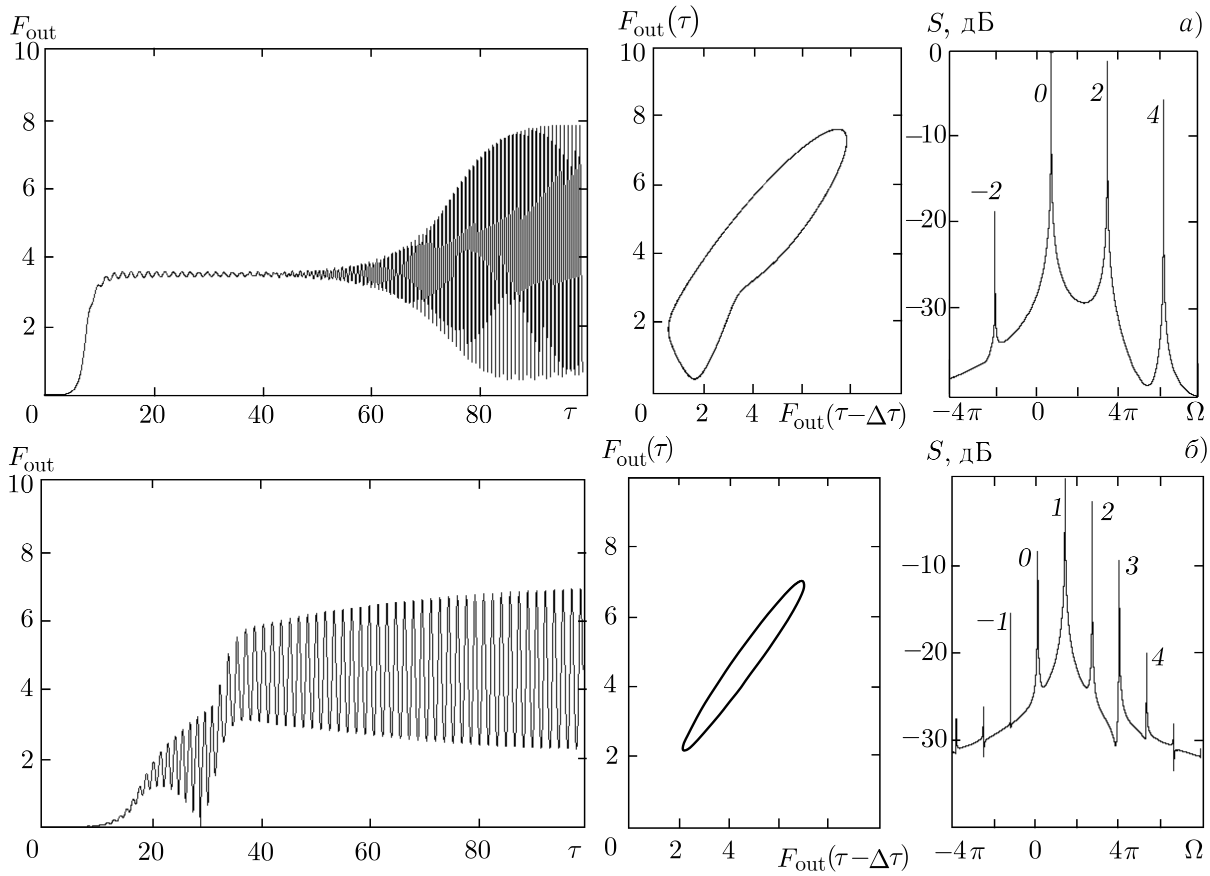


Рис. 4. Переходные процессы, фазовые портреты и спектры выходного сигнала в режимах периодической автомодуляции при $\rho = 0{,}7$ и $\delta = 0{,}5$: (*а*) — $\psi = 1{,}5\pi$; $L = 3{,}0$; (*б*) — $\psi = 0{,}2\pi$; $L = 2{,}5$

мод. Именно этим объясняется мягкий характер возбуждения автомодуляции, наблюдавшийся в экспериментах [4–7].

Типичный пример переходного процесса, фазового портрета и спектра выходного сигнала приведён на рис. 4*а*. Фазовый портрет строится при помощи так называемого метода задержек (метода Паккарда—Такенса), когда в качестве координат выбираются значения амплитуды выходного сигнала $F_{\rm out}$ в моменты времени, разделённые некоторым интервалом $\Delta\tau$. Это эквивалентно проекции аттрактора на некоторую плоскость в фазовом пространстве. Интервал $\Delta\tau$ выбирается из соображений удобства, чтобы восстановленный аттрактор имел наиболее наглядный вид. Цифрами на спектрограмме обозначены номера резонансных мод. На начальном этапе возбуждаются моды с номерами $n = 0$ и $n = 1$; затем нулевая мода подавляет первую ($\tau > 30$), и начинает медленно нарастать мода с номером $n = 2$. В результате устанавливается режим периодической автомодуляции, при котором возбуждаются моды с чётными номерами. Такое поведение характерно практически для всех параметров при данном значении δ, однако если ранее произошел переход к режиму на базе моды с $n = 1$, описанный в разделе 2, в спектре присутствуют только нечётные моды. Отметим, что спектр является существенно асимметричным: амплитуда высокочастотных сателлитов существенно больше. Как уже отмечалось выше, физически это объясняется тем, что электронный поток более эффективно отдаёт свою энергию волнам, фазовая скорость которых меньше скорости электронов. В ЛОВ-генераторе [13] реализуется противоположный случай, и доминирующими являются низкочастотные сателлиты.

Несколько иной характер носит динамика генератора при достаточно больших ρ вблизи гра-

Таблица 1. Порог автомодуляции, основная частота и частота автомодуляции при $\rho = 0{,}5$

δ	$\psi_{\rm c}/\pi$	$L_{\rm sm}$	Ω_0/π	$\Omega_{\rm sm}/\pi$	Ω_n/π (теория)
0,10	0,88	3,10	0,55; $n=0$	2,58; $n=1$	2,61
0,20	0,96	3,47	0,55; $n=0$	2,32; $n=1$	2,46
0,25	1,00	3,80	0,55; $n=0$	4,06; $n=2$	4,00
0,50	1,20	3,40	0,61; $n=0$	3,52; $n=2$	3,46
0,75	1,40	3,22	0,66; $n=0$	3,11; $n=2$	3,09
1,00	1,60	3,17	0,68; $n=0$	2,79; $n=2$	2,80
1,25	1,80	3,35	0,68; $n=0$	3,52; $n=3$	3,46
1,50	0,00	3,25	0,70; $n=1$	3,22; $n=4$	3,20
1,75	0,20	3,20	0,72; $n=1$	2,99; $n=4$	2,98
2,00	0,40	3,20	0,74; $n=1$	2,79; $n=4$	2,80
2,50	0,80	3,20	0,74; $n=1$	3,11; $n=5$	3,09
2,75	1,00	3,20	0,74; $n=1$	2,95; $n=5$	2,93
3,00	1,20	3,20	0,74; $n=1$	2,79; $n=5$	2,80

ницы зон генерации, где на границе зон автомодуляции имеется характерный «клюв» (см. рис. 1). В этом случае (рис. 4*б*) на начальном этапе конкурируют моды с номерами $n = 0$ и $n = 1$. За время $\tau \approx 30$ мода с номером $n = 1$ становится доминирующей, и начинается медленное нарастание моды с номером $n = 2$. В результате частота автомодуляции оказывается примерно равной межмодовому расстоянию, т. е. вдвое меньше, чем в предыдущем случае. Однако этот режим существует только в небольшом диапазоне параметров внутри «клюва»; с ростом L он теряет устойчивость и сменяется другим, в спектре которого присутствуют только моды с нечётными номерами. Частота автомодуляции, соответственно, увеличивается.

Видно, что возникновение автомодуляции в большинстве случаев обусловлено возбуждением моды с $n = 2$. При других значениях δ это может быть связано с другими модами, поэтому было проведено подробное исследование возникновения автомодуляции в зависимости от δ в широком диапазоне параметров. В табл. 1 представлены полученные результаты при умеренной глубине обратной связи ($\rho = 0{,}5$). Чтобы остальные условия оставались неизменными, для определённости рассматривалась динамика системы в центре зоны генерации, что достигалось подстройкой фазы параметра обратной связи. В табл. 1 приведены значения $\psi_{\rm c} = 0{,}8\pi\,(1+\delta) \bmod 2\pi$, соответствующие центру зоны, пороговые значения $L_{\rm sm}$, часто́ты основной Ω_0 и автомодуляционной $\Omega_{\rm sm}$ мод, а также их номера. Кроме того, приведены часто́ты «холодных» собственных мод, рассчитанные по формуле (9). Отметим, что Ω_0 отличается от частоты соответствующей «холодной» моды, которая, напомним, в центре зоны совпадает с $\Omega_{\max} \approx 0{,}8\pi$. В то же время $\Omega_{\rm sm}$ всегда близка к частоте одной из «холодных» мод. Номер автомодуляционной моды увеличивается с ростом δ. Это связано с тем, что имеется определённый диапазон наиболее «опасных» частот, в котором инкремент неустойчивости максимален, т. е. максимально усиление паразитной моды на фоне полезного сигнала (см., например, [18]). С ростом δ часто́ты мод уменьшаются, и в область максимального инкремента попадают моды со всё более высокими номерами. При переходе от одной автомодуляционной моды к другой их конкуренция приводит к увеличению порога автомодуляции $L_{\rm sm}$, которое при небольших δ может быть весьма значительным. В целом данная ситуация аналогична описанной в [23] для распределённого параметрического генератора. Поскольку с увеличением δ спектр мод сгущается, то динамика системы усложняется, а длительность переходных процессов увеличивается.

Необходимо заметить, что моделировалось возбуждение генератора малыми шумовыми флуктуациями, поэтому начальное распределение поля задавалось в виде случайной функции и проводилась серия расчётов при различных начальных условиях. Длительность и детали переходного процесса, вообще говоря, зависят от начальных условий. Особенно сильно это проявляется при больших δ, когда на начальной стадии происходит конкуренция большого числа близко расположенных мод. Кроме того, поскольку автомодуляция возникает жёстко, в окрестности порога автомодуляции имеет место бистабильность, т. е. сосуществуют автомодуляционный и одночастотный режимы. Если, например, проводить моделирование при адиабатически медленном увеличении параметра L, мы получим несколько бóльшие значения L_{sm}.

4. СЦЕНАРИЙ ПЕРЕХОДА К ХАОСУ

Дальнейшее увеличение тока пучка (параметра L) приводит к разрушению режима самосинхронизации мод. Подчеркнём, что детали последовательности бифуркаций, наблюдающиеся при увеличении L, отличаются чрезвычайным разнообразием, что характерно для распределённых систем с большим числом управляющих параметров [7, 13, 24, 26, 27]. Однако можно выделить некоторые типичные черты, которые проиллюстрируем для случая $\rho = 0{,}5$; $\delta = 0{,}5$; $\psi = 1{,}5\pi$ (рис. 5). Разрушение режима самосинхронизации связано с возбуждением мод с промежуточными номерами, которые прежде подавлялись в переходном процессе.[1] Частóты этих мод близки к частотам «холодных» мод (9), однако не равны им. Поэтому модуляция амплитуды выходного сигнала становится квазипериодической (рис. 5*а*), хотя частóты достаточно близки к резонансным. В данном случае это моды с нечётными номерами, расположенные примерно посередине между частотами, присутствовавшими в спектре периодического автомодуляционного режима. Аналогичное поведение для резонансной ЛОВ было отмечено в [13]. Далее наблюдается переход к хаосу через разрушение квазипериодического движения (сценарий Рюэля—Такенса, рис. 5*б, в*).

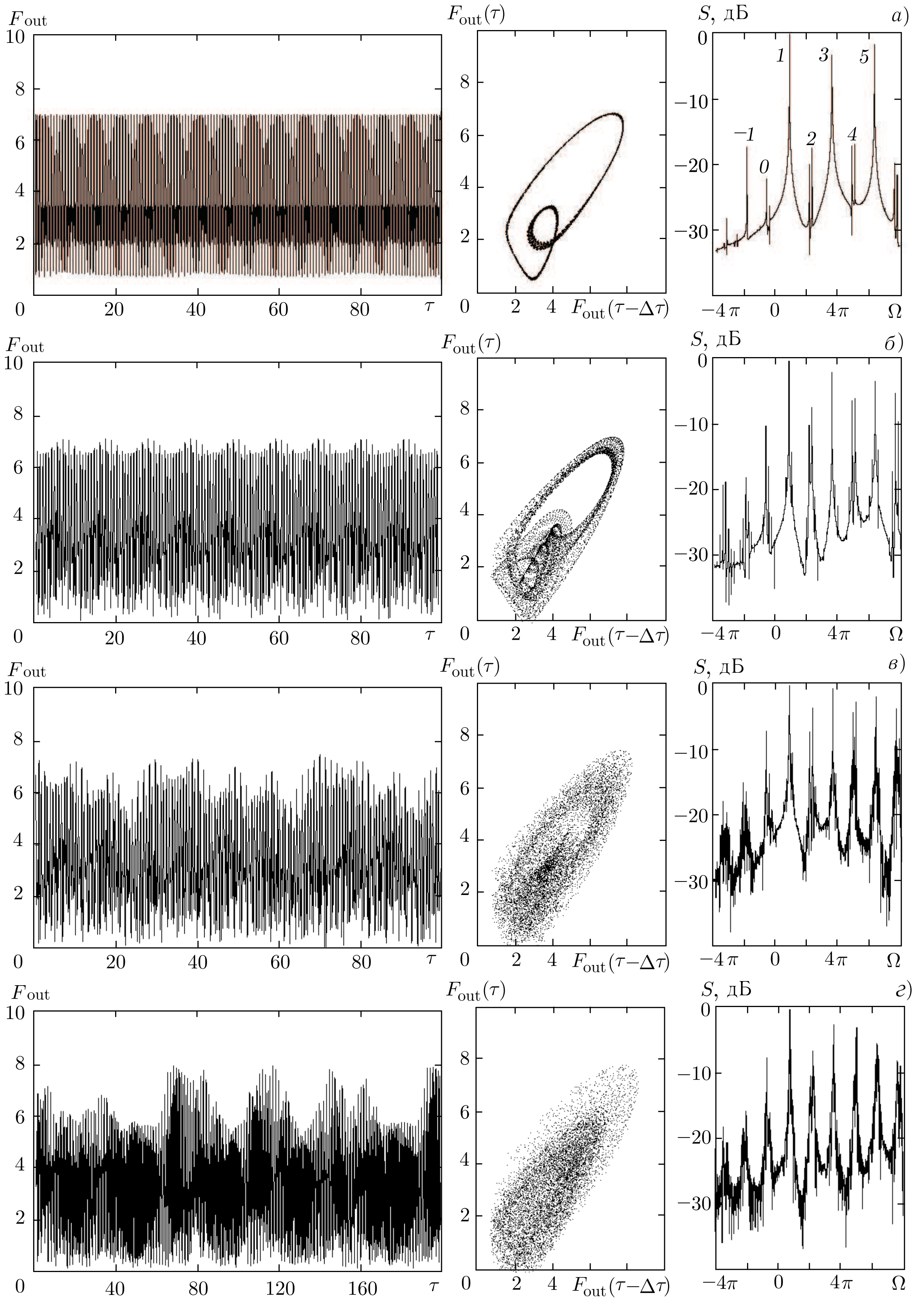
Однако одновременно с этим процессом проявляется тенденция перехода к режимам, в которых доминируют моды с более высокими номерами. Как уже отмечалось выше, это обусловлено торможением пучка в сильно нелинейном режиме, что приводит к преимущественному взаимодействию с более высокочастотными модами. Подобные переходы обычно происходят через перемежаемость. Конкуренция двух описанных выше тенденций приводит к тому, что наблюдать в чистом виде сценарий Рюэля—Такенса удаётся не всегда.[2] По-видимому, с этим связано утверждение о том, что в подобных системах доминирует переход через перемежаемость [8, 9]. Аналогичная картина для ЛОВ с отражениями была описана в [13].

В рассматриваемом случае происходит переход через перемежаемость (рис. 5*г*) к квазипериодическому режиму на базе мод с $n = 1$ и $n = 4$ (рис. 5*д*), причём их амплитуда примерно одинакова. На временнóй реализации на рис. 5*г* можно различить участки почти периодического движения, прерываемые сильно нерегулярными всплесками. Обратим внимание, что временной масштаб на рис. 5*г* уменьшен в два раза. Далее вновь наблюдается переход к хаосу через разрушение квазипериодического движения (рис. 5*е*). Таким образом, в области хаоса имеются «окна» регулярного поведения, т. е. фактически при увеличении L наблюдается сложная последовательность чередования регулярных и хаотических автомодуляционных режимов. Подобная картина типична для распределённых автоколебательных электронно-волновых систем (см., например, [13, 24, 26, 27]).

При большой надкритичности реализуются так называемые режимы «развитого» хаоса, т. е.

[1] За исключением случая малых δ, когда автомодуляционная мода соседствует с основной.

[2] В особенности это сложно сделать при больших δ.


$F_{\rm out}$
τ
$F_{\rm out}(\tau)$
$F_{\rm out}(\tau-\Delta\tau)$
S, дБ
Ω
а)
б)
в)
г)
1
3
5
−1
2
4
0

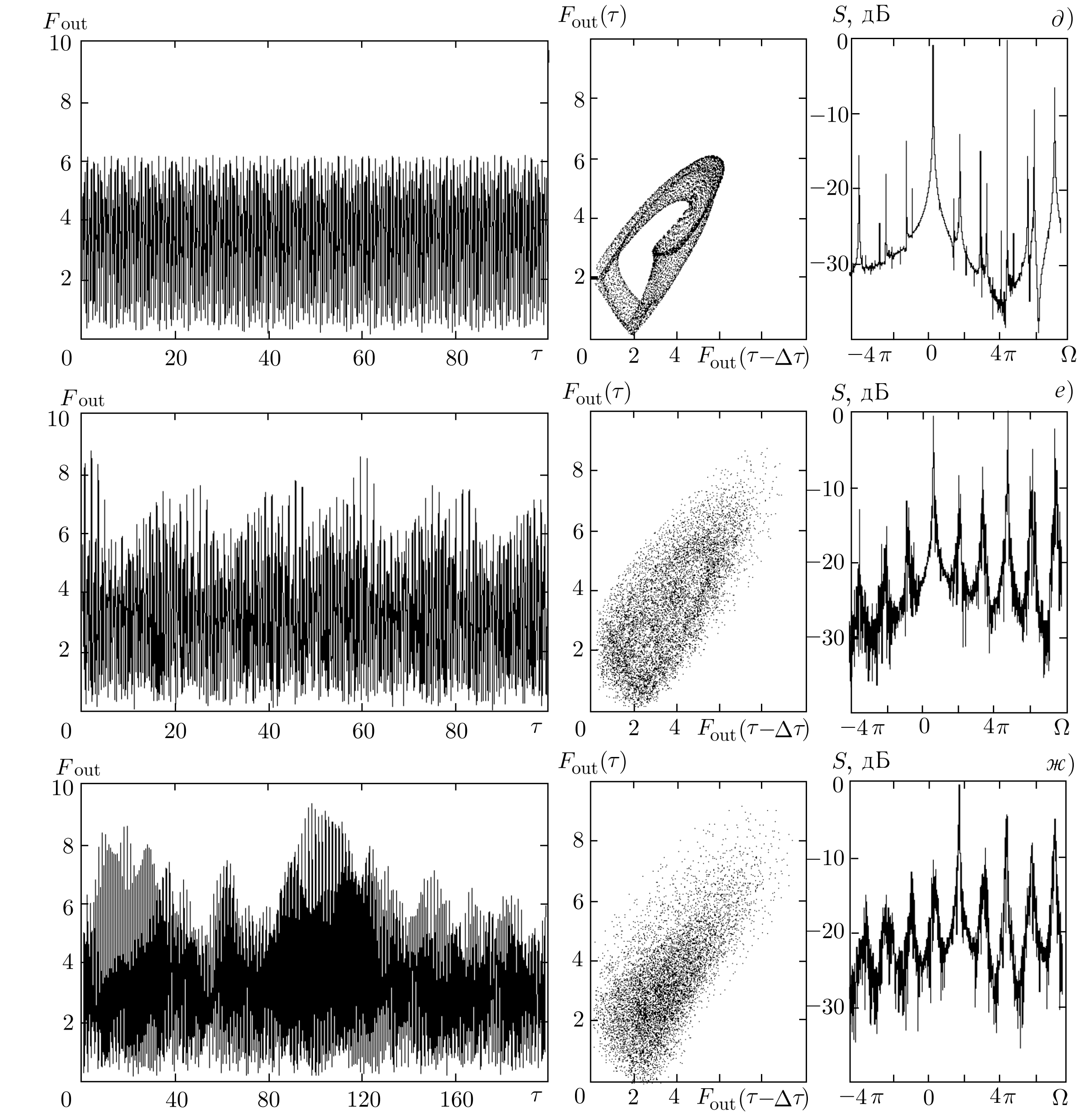


Рис. 5. Временны́е реализации, фазовые портреты и спектры выходного сигнала при $\rho = 0{,}5$; $\delta = 0{,}5$; $\psi = 1{,}5\pi$ и $L = 4{,}2$ (*а*), $L = 4{,}4$ (*б*), $L = 4{,}6$ (*в*), $L = 4{,}8$ (*г*), $L = 5{,}0$ (*д*), $L = 5{,}4$ (*е*), $L = 5{,}8$ (*ж*)

сильно нерегулярные колебания, когда на фазовом портрете не удаётся выделить какую-либо крупномасштабную структуру. При этом возбуждается очень большое число собственных мод (рис. 5*е, ж*). Однако даже в таких режимах в спектре сигнала отчётливо выделяются компоненты «холодных» мод на фоне относительно низкого шумового пьедестала (см. [13]), который медленно растёт с увеличением надкритичности. Вместе с тем происходят новые переходы к режимам на базе мод со всё более высокими частотами. Так, на рис. 5*ж* можно видеть, что основной является мода с $n = 2$.

Наконец, отметим, что в основном рассматривался случай, когда параметр δ не слишком велик ($\delta \le 3$). Это сделано, прежде всего, потому, что с ростом δ исследование системы затрудняется, поскольку в полосу усиления попадает много мод, находящихся примерно в одинаковых условиях. Эффекты конкуренции этих мод приводят к усложнению динамики ЛБВ-генератора в целом, увеличению длительности переходных процессов и т. д. Однако такие особенности, как частотный механизм автомодуляции, переход от периодической автомодуляции к квазипериодической за счёт возбуждения мод с промежуточными номерами, которые ранее были подавлены, конкуренция сценариев разрушения квазипериодического движения и перемежаемости, остаются в силе и при $\delta > 3$.

5. ЗАКЛЮЧЕНИЕ

В работе представлены результаты исследования режимов автомодуляционной и хаотической генерации в ЛБВ-генераторе с запаздывающей обратной связью. Численное моделирование показало, что автомодуляция возникает по так называемому частотному механизму, т. е. связана с жёстким возбуждением одной из резонансных мод кольцевой системы. Номер автомодуляционной моды определяется главным образом безразмерным временем запаздывания δ. Основным сценарием перехода к хаосу является разрушение квазипериодического движения. Однако по мере увеличения надкритичности наблюдается ряд переходов к режимам на базе всё более высокочастотных мод, которые, как правило, происходят через перемежаемость. Конкуренция между этими двумя тенденциями приводит к сложной картине чередующихся в пространстве параметров зон квазипериодических и хаотических автомодуляционных режимов. Даже в развитых хаотических режимах в спектре чётко видны компоненты на частотах «холодных» собственных мод с невысоким шумовым пьедесталом. В целом описанная картина нелинейной динамики обнаруживает много общего с такими системами, как резонансная ЛОВ, ЛСЭ, распределённый параметрический генератор и др., и является достаточно характерной для распределённых электронно-волновых автогенераторов с запаздывающей обратной связью.

Автор выражает искреннюю благодарность А. Г. Зайцевой за помощь в проведении расчётов. Работа выполнена при финансовой поддержке CRDF (грант № REC–006) и РФФИ (грант № 03–02–16192а).

СПИСОК ЛИТЕРАТУРЫ

1. Кислов В. Я., Залогин Н. Н., Мясин Е. А. // Радиотехника и электроника. 1979. Т. 24, № 6. С. 1118.
2. Кислов В. Я. // Радиотехника и электроника. 1980. Т. 25, № 8. С. 1683.
3. Кислов В. Я., Мясин Е. А., Залогин Н. Н. // Радиотехника и электроника. 1980. Т. 25, № 10. С. 2160.
4. Калинин В. И., Залогин Н. Н., Кислов В. Я. // Радиотехника и электроника. 1983. Т. 28, № 10. С. 2001.
5. Анисимова Ю. В., Дмитриев А. С., Залогин Н. Н. и др. // Письма в ЖЭТФ. 1983. Т. 37, № 8. С. 387.
6. Кац В. А., Трубецков Д. И. // Письма в ЖЭТФ. 1984. Т. 39, № 3. С. 116.
7. Кац В. А. // Изв. вузов. Радиофизика. 1985. Т. 28, № 2. С. 161.
8. Блиох Ю. П., Бородкин А. В., Любарский М. Г. и др. // Изв. вузов. Прикладная нелинейная динамика. 1993. Т. 1, № 1–2. С. 34.
9. Блиох Ю. П., Любарский М. Г., Подобинский В. О., Файнберг Я. Б. // Физика плазмы. 1994. Т. 20, № 7–8. С. 718.

10. Bhattacharjee S., Kory C. L., Lee W.-J., et al. // Third IEEE International Vacuum Electronics Conference, April 23–25, 2002. Monterey, California, USA. P. 26.
11. Han S.-T., Kim J.-I., Park G.-S. // Third IEEE International Vacuum Electronics Conference, April 23–25, 2002. Monterey, California, USA. P. 94.
12. Шевчик В. Н., Трубецков Д. И. Аналитические методы расчёта в электронике СВЧ. М.: Сов. радио, 1971.
13. Рыскин Н. М., Титов В. Н. // Изв. вузов. Радиофизика. 2001. Т. 44, № 10. С. 860.
14. Булгакова Л. В., Кузнецов С. П. // Изв. вузов. Радиофизика. 1988. Т. 31, № 2. С. 207.
15. Булгакова Л. В., Кузнецов С. П. // Изв. вузов. Радиофизика. 1988. Т. 31, № 5. С. 612.
16. Манькин И. А., Школьников В. Г. // Радиотехника и электроника. 1984. Т. 29, № 2. С. 307.
17. Гинзбург Н. С., Сергеев А. С. // ЖТФ. 1991. Т. 61, № 6. С. 133.
18. Братман В. Л., Савилов А. В. // Изв. вузов. Прикладная нелинейная динамика. 1994. Т. 2, № 6. С. 27.
19. Savilov A. V., Bratman V. L., Denisov G. G., et al. // Phys. Plasmas. 2001. V. 8, No. 2. P. 638.
20. Гинзбург Н. С., Кузнецов С. П., Федосеева Т. Н. // Изв. вузов. Радиофизика. 1978. Т. 21, № 7. С. 1 037.
21. Гинзбург Н. С., Петелин М. И. // Изв. вузов. Прикладная нелинейная динамика. 1994. Т. 2, № 6. С. 3.
22. Дмитриев Б. С., Жарков Ю. Д., Кижаева К. К. и др. // Изв. вузов. Прикладная нелинейная динамика. 2002. Т. 10, № 5. С. 37.
23. Дмитриева Т. В., Рыскин Н. М. // ЖЭТФ. 2001. Т. 120, № 6. С. 1 517.
24. Рыскин Н. М., Шигаев А. М. // ЖТФ. 2002. Т. 72, № 7. С. 1.
25. Айзацкий Н. И. // ЖТФ. 1987. Т. 57, № 8. С. 1 671.
26. Рыскин Н. М., Титов В. Н., Трубецков Д. И. // ДАН. 1998. Т. 358, № 5. С. 620.
27. Рыскин Н. М., Титов В. Н. // Изв. вузов. Прикладная нелинейная динамика. 1998. Т. 6, № 1. С. 75.

Саратовский госуниверситет им. Н. Г. Чернышевского,
г. Саратов, Россия

Поступила в редакцию
1 апреля 2003 г.

STUDY OF THE NONLINEAR DYNAMICS OF A TRAVELING-WAVE-TUBE OSCILLATOR WITH A DELAYED FEEDBACK

N. M. Ryskin

We present the results of numerical simulations of the nonlinear dynamics in a traveling-wave-tube (TWT) oscillator with a delayed feedback are presented. Basic properties of stationary single-frequency oscillation regimes are considered, the onset of self-modulation is studied in detail. Various route-to-chaos scenarios corresponding to successively increasing values of the beam current are simulated numerically. It is shown that the basic scenario is a quasi-periodic route to chaos, while the beam deceleration in strongly nonlinear regimes causes transitions via intermittency to the regimes based on modes with higher frequencies. Competition between these two scenarios leads to a complex picture of regular and chaotic self-modulation regimes in the parameter space. Such a behavior is typical of distributed electron–wave self-oscillators with delayed feedback.

УДК 621.391.1

ЭФФЕКТИВНОСТЬ ПРОСТРАНСТВЕННОГО РАЗДЕЛЕНИЯ ПОЛЬЗОВАТЕЛЕЙ В MIMO-СИСТЕМАХ СВЯЗИ С ПАРАЛЛЕЛЬНОЙ ПЕРЕДАЧЕЙ ИНФОРМАЦИИ

В. Т. Ермолаев [1], *А. Г. Флаксман* [1], *И. М. Аверин* [2], *Д. В. Грибов* [2]

Рассматриваются MIMO-системы (multiple-input multiple-output systems) связи с антенными решётками на обоих концах линии и передачей данных по параллельным собственным каналам, согласованным со случайным пространственным каналом. Исследуется эффективность метода пространственного разделения пользователей, не требующего оценки направлений прихода сигналов и основанного на ортогонализации всех параллельных каналов всех пользователей. В случае релеевских замираний сигналов получены приближённые аналитические выражения для среднего отношения мощности сигнала к мощности шума и пропускной способности MIMO-системы, которые являются значительно более простыми по сравнению с точными формулами, однако обеспечивают высокую точность при произвольном числе передающих и приёмных антенн и произвольной мощности передатчика. Приведённые результаты показывают высокую эффективность рассмотренного способа пространственного разделения пользователей.

ВВЕДЕНИЕ

В настоящее время широко используются временно́й, частотный и кодовый способы разделения пользователей в системах сотовой связи. Системы с кодовым разделением обладают возможностью одновременного обслуживания пользователей, работающих в одном частотном диапазоне, и поэтому имеют наибольшую пропускную способность. Известен также способ пространственного разделения пользователей, называющийся SDMA (Space Division Multiple Access), который может быть реализован в MIMO-системах, использующих антенные решётки на одном или на обоих концах линии связи [1–3]. Физический принцип SDMA основан на формировании системы ортогональных лучей, каждый из которых обеспечивает передачу информации только одному пользователю. В [1–3] для реализации SDMA предлагается предварительно оценивать направления прихода сигналов от пользователей, а затем формировать систему ортогональных лучей.

Оценка азимута направления прихода сигналов от пользователей в условиях случайной среды распространения является сложной проблемой, поскольку пользователь окружён отражателями, которые рассеивают его сигнал. Поэтому для базовой станции пользователь представляет собой распределённый источник, угловые размеры которого могут составлять несколько десятков градусов [4]. Более того, центр излучения такого источника может флуктуировать в достаточно больших пределах, т. к. число рассеивателей, их угловое положение и эффективная поверхность рассеяния являются случайными величинами.

В [5] рассмотрен способ пространственного разделения пользователей в MIMO-системах, не требующий оценки направлений прихода сигналов. Предполагается, что передача информации осуществляется по параллельным собственным каналам, которые формируются с помощью адаптивных диаграммообразующих схем (ДОС) на обоих концах линии связи [6]. Такие каналы называются собственными, т. к. ДОС построены на собственных векторах матрицы коэффициентов передачи рассеивающей среды. При этом обеспечивается согласование со случайным каналом связи не только приёмника, но и передатчика, вследствие чего значительно увеличивается пропускная способность MIMO-системы [7, 8]. SDMA обеспечивается за счёт дополнительной обработки сигналов, на основе ортогонализации пространственных каналов всех пользователей без оценки направлений прихода сигналов. Однако, хотя полученные в [5] результаты для отношения мощности сигнала к мощности шума (ОСШ) на выходе собственных каналов и пропускной способности MIMO-системы справедливы в случае произвольных замираний сигналов, они являются достаточно сложными и дают возможность получить значения ОСШ и пропускной способности только с помощью математического моделирования. Кроме того, в [5] считалось, что мощность передатчика увеличивается пропорционально числу пространственно разделяемых пользователей. Более естественным является предположение о том, что излучаемая мощность является ограниченной и не зависит от числа обслуживаемых пользователей.

В настоящей работе рассмотрен случай некоррелированных релеевских замираний сигналов в приёмных антеннах, который является наиболее актуальным для систем связи, работающих в городских условиях. Получены приближённые аналитические выражения для средних значений ОСШ и пропускной способности MIMO-системы, которые являются достаточно простыми и обеспечивают высокую точность при произвольном числе передающих и приёмных антенн. Мощность передатчика также является произвольной и не зависит от числа обслуживаемых пользователей. Показано, что существует оптимальное число пространственно разделяемых пользователей, обеспечивающее максимальную пропускную способность системы связи. Это объясняется тем, что, с одной стороны, пропускная способность возрастает с увеличением числа пользователей за счёт увеличения числа собственных каналов. Однако, с другой стороны, потери в ОСШ на выходе этих каналов при увеличении числа пользователей также возрастают. Оптимальное число пользователей увеличивается с ростом мощности передатчика и числа передающих антенн и уменьшается с ростом числа приёмных антенн.

1. ПРОСТРАНСТВЕННОЕ РАЗДЕЛЕНИЕ ПОЛЬЗОВАТЕЛЕЙ В MIMO-СИСТЕМАХ С СОБСТВЕННЫМИ КАНАЛАМИ

Предположим сначала, что MIMO-система состоит из M передающих и N приёмных антенн и обеспечивает передачу данных одному пользователю. Если пространственный канал связи является частотно-неселективным, он может быть описан матрицей $\mathbf{H}$ размерности $N \times M$ комплексных коэффициентов передачи h_{nm} сигналов из m-й передающей антенны в n-ую приёмную антенну. В случае распространения сигналов в случайной однородной рассеивающей среде коэффициенты h_{nm} имеют релеевское распределение амплитуд и равномерное распределение фаз в интервале $[0, 2\pi]$ [9]. Без ограничения общности можно считать, что коэффициенты передачи в среднем нормированы к единице, т. е. $\langle |h_{nm}|^2 \rangle = 1$, где угловые скобки обозначают статистическое усреднение.

Передаваемые сигналы разделяются на параллельные потоки, число K которых не может быть больше минимального числа передающих или приёмных антенн, т. е. $K \le \min(M, N)$. Обозначим $\mathbf{C}(t) = (c_1(t), c_2(t), \ldots, c_K(t))^{\mathrm{T}}$ K-мерный вектор-столбец параллельно передаваемых сигналов, где индекс T обозначает операцию транспонирования. Для формирования этих информационных потоков приёмник должен непрерывно оценивать канальную матрицу $\mathbf{H}$, затем эта информация должна сообщаться на передающий конец линии. Оценка матрицы $\mathbf{H}$ производится с помощью псевдошумовых обучающих последовательностей, состоящих из конечного числа известных символов. Обычно используются максимально правдоподобные оценки или оценки, основанные на поиске минимума среднеквадратической ошибки [10]. Матрица $\mathbf{H}$ всегда оценивается с некоторой ошибкой, обусловленной влиянием собственных шумов приёмных устройств и изменением состояния канала за время между двумя его последовательными оценками. Будем рассматривать потенциальные характеристики MIMO-системы, реализуемые при точно известной на обоих концах линии матрице $\mathbf{H}$.

M-мерный вектор-столбец сигналов $\mathbf{S}(t) = (s_1(t), s_2(t), \ldots, s_M(t))^{\mathrm{T}}$ на выходе передающей

ДОС можно представить в виде

$$\mathbf{S}(t) = \mathbf{V}\mathbf{P}^{1/2}\mathbf{C}(t), \tag{1}$$

где матрица $\mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_K)$ размерности $M \times K$ состоит из весовых векторов $\mathbf{V}_i$ передающей ДОС, $\mathbf{P} = \mathrm{diag}(p_1, p_2, \ldots, p_K)$ — диагональная матрица, составленная из чисел p_i, которые дают распределение мощности между параллельными каналами.

Вектор $\mathbf{X}(t) = (x_1(t), x_2(t), \ldots, x_N(t))^{\mathrm{T}}$ сигналов в элементах приёмной антенной решётки равен

$$\mathbf{X}(t) = \mathbf{H}\mathbf{S}(t) + \mathbf{Z}(t), \tag{2}$$

где $\mathbf{Z}(t) = (z_1(t), z_2(t), \ldots, z_N(t))^{\mathrm{T}}$ — вектор собственных шумов, которые будем считать гауссовскими некоррелированными во времени и в приёмных каналах случайными процессами с нулевым средним и дисперсией σ_0^2.

Опишем преобразование сигнала в приёмной ДОС с помощью матрицы $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_K)$ размерности $N \times K$. Тогда вектор сигнала на выходе ДОС равен $\mathbf{Y}(t) = \mathbf{U}^{\mathrm{H}}\mathbf{X}(t)$, где индекс H обозначает эрмитовое сопряжение. Учитывая (1) и (2), получим

$$\mathbf{Y}(t) = \mathbf{U}^{\mathrm{H}}\mathbf{H}\mathbf{V}\mathbf{P}^{1/2}\mathbf{C}(t) + \mathbf{U}^{\mathrm{H}}\mathbf{Z}(t). \tag{3}$$

Если в качестве весовых векторов $\mathbf{V}_i$ передающей ДОС использовать собственные векторы матрицы $\mathbf{H}^{\mathrm{H}}\mathbf{H}$, а весовые векторы приёмной ДОС являются собственными векторами матрицы $\mathbf{H}\mathbf{H}^{\mathrm{H}}$, то матрица $\mathbf{U}^{\mathrm{H}}\mathbf{H}\mathbf{V}$ в (3) будет иметь диагональный вид: $\mathbf{U}^{\mathrm{H}}\mathbf{H}\mathbf{V} = \mathbf{\Lambda}^{1/2}$, где $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_K)$ — диагональная матрица ненулевых собственных чисел λ_i матрицы $\mathbf{H}\mathbf{H}^{\mathrm{H}}$ или матрицы $\mathbf{H}^{\mathrm{H}}\mathbf{H}$. Сформированные таким образом каналы называются собственными.

Вектор сигнала на выходе собственных каналов равен

$$\mathbf{Y}(t) = \mathbf{\Lambda}^{1/2}\mathbf{P}^{1/2}\mathbf{C}(t) + \tilde{\mathbf{Z}}(t), \tag{4}$$

где $\tilde{\mathbf{Z}}(t) = \mathbf{U}^{\mathrm{H}}\mathbf{Z}(t)$ — вектор шумов на выходе ДОС. Отсюда следует, что на выходе приёмной ДОС в i-м собственном канале присутствует только i-й передаваемый символ. Кроме того, собственные шумы в этих собственных каналах являются некоррелированными между собой, т. к. их корреляционная матрица является единичной: $\langle \tilde{\mathbf{Z}}(t)\tilde{\mathbf{Z}}^{\mathrm{H}}(t)\rangle \equiv \mathbf{I}$. Следовательно, взаимная связь между собственными каналами отсутствует. Это означает, что с помощью линейного преобразования сигналов в передающей и приёмной диаграммообразующих схемах MIMO-система может быть представлена в виде $K \le \min(M, N)$ независимых параллельных каналов (подсистем). Общая схема MIMO-системы с собственными каналами показана на рис. 1.

Эффективность передачи данных в системах связи определяется вероятностью битовой ошибки, которая зависит как от статистических свойств флуктуаций сигналов в пространственном канале связи, так и от выбранного способа кодирования передаваемой информации. Существует большое число методов кодирования [10], которые существенно влияют на битовую ошибку. Поэтому эффективность системы удобно характеризовать шенноновской пропускной способностью, не зависящей от способа кодирования информации и равной максимальному числу бит, которые можно передать без ошибки за одну секунду в полосе частот 1 Гц (бит $\cdot$ с$^{-1}$ $\cdot$ Гц$^{-1}$).

Из (4) следует, что ОСШ на выходе i-го собственного канала равно $\eta_i = \rho_i\lambda_i$, где $\rho_i = p_i/\sigma_0^2$. Поэтому для пропускной способности этого канала имеем $C_i = \log_2(1 + \rho_i\lambda_i)$ [10]. Поскольку собственные каналы являются независимыми, то пропускная способность всей MIMO-системы

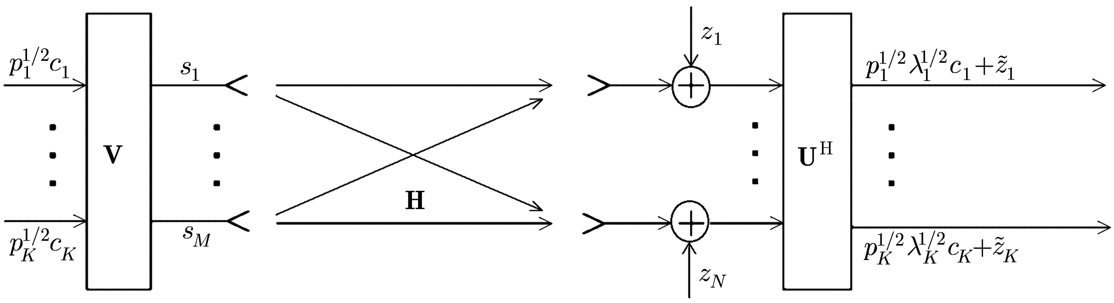


Рис. 1

определяется суммой пропускных способностей отдельных каналов, т. е.

$$C = \sum_{i=1}^{K} \log_2(1 + \rho_i \lambda_i). \tag{5}$$

Пропускная способность MIMO-системы зависит от полной мощности P_0 передатчика и от способа её распределения между собственными каналами. Оптимальным является распределение мощности в соответствии с так называемым правилом «наполнения водой» («water pouring») [10]. Чем больше коэффициент передачи λ_i соответствующего собственного канала, тем бóльшая доля мощности распределяется в этот канал. Мощность в i-м собственном канале равна $p_i = \mu - \sigma_0^2 \lambda_i^{-1}$, где константа μ находится из условия ограничения полной мощности: $p_1 + p_2 + \ldots + p_K = P_0$. Отметим, что если в некотором канале $\mu < \sigma_0^2 \lambda_i^{-1}$, то $p_i = 0$, т. е. мощность в данный канал не распределяется, и он не будет использоваться для передачи данных.

Рассмотрим теперь одновременную передачу данных произвольному числу Q пользователей, каждый из которых может иметь разное число приёмных антенн. Сделаем не принципиальное предположение, что все пользователи имеют одинаковое число N антенн. При этом для каждого из них может быть сформировано $K = N$ собственных каналов. Очевидно, что передающая антенна базовой станции должна состоять из большего числа M антенн ($M > N$), а максимальное число Q пространственно разделяемых пользователей должно удовлетворять условию $Q \le \mathrm{int}(M/N)$, где $\mathrm{int}(x)$ обозначает целую часть x. Будем также считать, что полная мощность P_0 передатчика является ограниченной и не зависит от числа пользователей, т. е. при обслуживании Q пользователей мощность, предназначенная каждому из них, уменьшается в Q раз и равна P_0/Q.

Для обозначения номера пользователя будем использовать верхний индекс. Тогда пространственный канал связи можно описать с помощью Q матриц $\mathbf{H}^{(1)}, \mathbf{H}^{(2)}, \ldots, \mathbf{H}^{(Q)}$, каждая из которых имеет одинаковую размерность $N \times M$ и состоит из коэффициентов передачи $h_{nm}^{(q)}$ сигналов из m-й передающей антенны в n-ую приёмную антенну q-го пользователя. В условиях случайной однородной рассеивающей среды все коэффициенты $h_{nm}^{(q)}$ имеют одинаковые статистические свойства. Передающую и приёмную диаграммообразующие схемы, формирующие собственные каналы для q-го пользователя, будем описывать матрицами $\mathbf{V}^{(q)}$ и $\mathbf{U}^{(q)}$, а распределение мощности P_0/Q по собственным каналам q-го пользователя зададим диагональной матрицей $\mathbf{P}^{(q)} = \mathrm{diag}(p_1^{(q)}, p_2^{(q)}, \ldots, p_K^{(q)})$.

Информация, предназначенная всем пользователям, смешивается и передаётся одновременно. Пусть $\mathbf{C}^{(q)}(t)$ — вектор сигналов, одновременно передаваемых q-му пользователю. Тогда для вектора сигналов на выходе передающей ДОС вместо (1) имеем

$$\mathbf{S}(t) = \sum_{q=1}^{Q} \mathbf{V}^{(q)} \left(\mathbf{P}^{(q)}\right)^{1/2} \mathbf{C}^{(q)}(t). \tag{6}$$

Вектор сигнала $\mathbf{X}^{(q)}(t)$, принимаемого q-м пользователем, определяется выражением (2). Представляя матрицу $\mathbf{H}^{(q)}$ канальных коэффициентов в виде сингулярного разложения $\mathbf{H}^{(q)} = \mathbf{U}^{(q)}[\mathbf{\Lambda}^{(q)}]^{1/2}\mathbf{V}^{(q)\mathrm{H}}$ и учитывая (6), получим

$$\mathbf{X}^{(q)}(t) = \sum_{p=1}^{Q} \mathbf{U}^{(q)} \left(\mathbf{\Lambda}^{(q)}\mathbf{P}^{(p)}\right)^{1/2} \left(\mathbf{V}^{(q)\mathrm{H}}\mathbf{V}^{(p)}\right) \mathbf{C}^{(p)}(t) + \mathbf{Z}^{(q)}(t). \tag{7}$$

После преобразования в приёмной ДОС вектор сигнала для q-го пользователя на выходе собственных каналов будет равен

$$\mathbf{Y}^{(q)}(t) = \left(\mathbf{\Lambda}^{(q)}\mathbf{P}^{(q)}\right)^{1/2} \mathbf{C}^{(q)}(t) + \sum_{\substack{p=1\\ p\neq q}}^{Q} \left(\mathbf{\Lambda}^{(q)}\mathbf{P}^{(p)}\right)^{1/2} \left(\mathbf{V}^{(q)\mathrm{H}}\mathbf{V}^{(p)}\right) \mathbf{C}^{(p)}(t) + \mathbf{U}^{(q)\mathrm{H}}\mathbf{Z}^{(q)}(t). \tag{8}$$

Полная система весовых векторов $\mathbf{V}_i^{(q)}$, где $i = 1, 2, \ldots, K$; $q = 1, 2, \ldots, Q$, передающих диаграммообразующих схем всех пользователей не является ортогональной, т. к. векторы $\mathbf{V}_i^{(q)}$ и $\mathbf{V}_j^{(p)}$ ортогональны друг другу только при $q = p$ и $j \neq i$, т. е. весовые векторы, формирующие собственные каналы для разных пользователей, не ортогональны между собой. Неортогональность системы векторов $\mathbf{V}_i^{(q)}$ приводит к тому, что q-й пользователь принимает информацию, предназначенную не только ему (первое слагаемое в (8)), но и другим пользователям (второе слагаемое в (8)). Это означает, что разные пользователи оказываются пространственно неразделёнными, и независимая параллельная передача всех KQ символов становится невозможной.

В [5] рассмотрена процедура пространственного разделения пользователей, основанная на ортогонализации всех KQ собственных каналов всех Q пользователей. Процедура ортогонализации представляет собой дополнительное преобразование сигналов на передающем конце линии связи и может быть выполнена с помощью матриц-проекторов $\mathbf{\Pi}_i^{(q)}$, где $i = 1, 2, \ldots, K$; $q = 1, 2, \ldots, Q$, на подпространство, ортогональное всем векторам, кроме вектора $\mathbf{V}_i^{(q)}$. Общая схема системы с SDMA показана на рис. 2.

Пусть $\mathbf{A} = (\mathbf{V}_1^{(1)}, \ldots, \mathbf{V}_K^{(1)}, \mathbf{V}_1^{(2)}, \ldots, \mathbf{V}_K^{(2)}, \ldots, \mathbf{V}_1^{(Q)}, \ldots, \mathbf{V}_K^{(Q)})$ — матрица размерности $M \times QK$, составленная из всех весовых векторов $\mathbf{V}_i^{(q)}$, а $\mathbf{A}_i^{(q)}$ — матрица, образованная из матрицы $\mathbf{A}$ исключением вектора $\mathbf{V}_i^{(q)}$. Проекция вектора $\mathbf{V}_i^{(q)}$ на подпространство, ортогональное всем другим векторам, определяется вектором $\mathbf{\Pi}_i^{(q)}\mathbf{V}_i^{(q)}$, где матрица-проектор равна [11] $\mathbf{\Pi}_i^{(q)} = \mathbf{I} - \mathbf{A}_i^{(q)} (\mathbf{A}_i^{(q)\mathrm{H}}\mathbf{A}_i^{(q)})^{-1}\mathbf{A}_i^{(q)\mathrm{H}}$. Соотвествующий нормированный вектор равен [5]

$$\tilde{\mathbf{V}}_i^{(q)} = \frac{\mathbf{\Pi}_i^{(q)}\mathbf{V}_i^{(q)}}{\sqrt{\mathbf{V}_i^{(q)\mathrm{H}}\mathbf{\Pi}_i^{(q)}\mathbf{V}_i^{(q)}}}\,. \tag{9}$$

Векторы $\tilde{\mathbf{V}}_i^{(q)}$, где $i = 1, 2, \ldots, K$; $q = 1, 2, \ldots, Q$, образуют ортогональную систему векторов и являются весовыми векторами результирующей передающей ДОС, которая обеспечивает формирование всех собственных каналов, а также пространственное разделение пользователей.

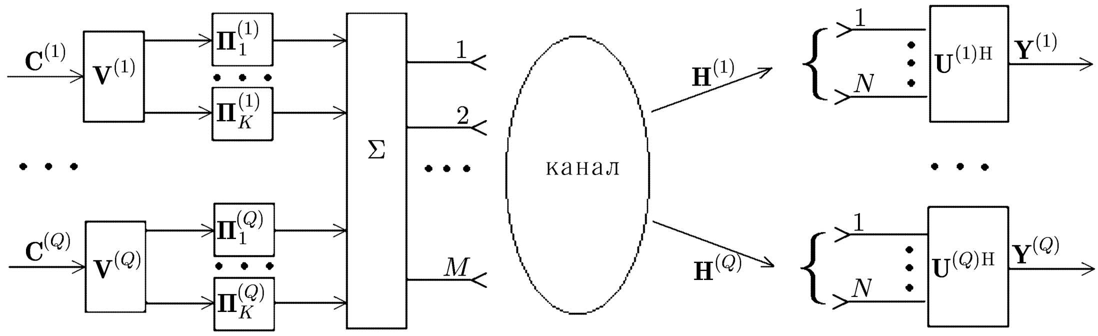


Рис. 2

Теперь вектор $\mathbf{S}(t)$ излучаемых сигналов и вектор сигналов $\mathbf{X}^{(q)}(t)$ на входе приёмной ДОС q-го пользователя будут определяться с помощью (6) и (7), если в этих формулах заменить вектор $\mathbf{V}_i^{(q)}$ вектором $\tilde{\mathbf{V}}_i^{(q)}$. Далее учтём, что матрица $\mathbf{V}^{(q)\mathrm{H}}\tilde{\mathbf{V}}^{(p)}$ является диагональной, т. к. её элемент равен $(\mathbf{V}^{(q)\mathrm{H}}\tilde{\mathbf{V}}^{(p)})_{lm} = \sqrt{\mathbf{V}_l^{(q)\mathrm{H}}\mathbf{\Pi}_m^{(p)}\mathbf{V}_m^{(p)}} \propto \delta_{pq}\delta_{lm}$, где δ_{lm} — символ Кронекера. В результате для вектора сигнала, принимаемого q-м пользователем, вместо (8) имеем

$$\mathbf{Y}^{(q)}(t) = \left(\mathbf{\Lambda}^{(q)}\right)^{1/2}\left(\mathbf{\Lambda}_{\Pi}^{(q)}\right)^{1/2}\left(\mathbf{P}^{(q)}\right)^{1/2}\mathbf{C}^{(q)}(t) + \tilde{\mathbf{Z}}^{(q)}(t), \tag{10}$$

где $\mathbf{\Lambda}_{\Pi}^{(q)}$ — диагональная матрица, составленная из чисел $\mathbf{V}_i^{(q)\mathrm{H}}\mathbf{\Pi}_i^{(q)}\mathbf{V}_i^{(q)}$ $(i = 1, 2, \ldots, K)$. Из (10) следует, что рассмотренная процедура ортогонализации обеспечивает пространственное разделение Q пользователей и, таким образом, даёт возможность параллельной передачи KQ символов.

Из (10) следует, что ОСШ на выходе i-го собственного канала q-го пользователя равно

$$\eta_i^{(q)} = \rho_i^{(q)}\lambda_i^{(q)}\left(\mathbf{V}_i^{(q)\mathrm{H}}\mathbf{\Pi}_i^{(q)}\mathbf{V}_i^{(q)}\right). \tag{11}$$

Для любой системы нормированных векторов $\tilde{\mathbf{V}}_i^{(q)}$ справедливо неравенство $\mathbf{V}_i^{(q)\mathrm{H}}\mathbf{\Pi}_i^{(q)}\mathbf{V}_i^{(q)} < 1$, поэтому ОСШ на выходе всех собственных каналов всех пользователей будет уменьшаться в $\mathbf{V}_i^{(q)\mathrm{H}}\mathbf{\Pi}_i^{(q)}\mathbf{V}_i^{(q)}$ раз за счёт пространственного разделения пользователей.

Полную пропускную способность MIMO-системы можно получить, суммируя удельные пропускные способности, приходящиеся на каждого из пространственно разделённых пользователей:

$$C = \sum_{q=1}^{Q}\sum_{i=1}^{K}\log_2\left[1 + \rho_i^{(q)}\lambda_i^{(q)}\left(\mathbf{V}_i^{(q)\mathrm{H}}\mathbf{\Pi}_i^{(q)}\mathbf{V}_i^{(q)}\right)\right]. \tag{12}$$

2. ВЫВОД АНАЛИТИЧЕСКИХ ВЫРАЖЕНИЙ ДЛЯ ОСШ И ПРОПУСКНОЙ СПОСОБНОСТИ MIMO-СИСТЕМЫ С ПРОСТРАНСТВЕННЫМ РАЗДЕЛЕНИЕМ ПОЛЬЗОВАТЕЛЕЙ

Выражения (11) и (12) для ОСШ на выходе собственных каналов и пропускной способности MIMO-системы справедливы в случае произвольных замираний сигналов. Однако они являются достаточно сложными и дают возможность получить значения ОСШ и пропускной способности только с помощью математического моделирования.

Покажем, что в практически наиболее важном случае однородной рассеивающей среды с некоррелированными релеевскими коэффициентами передачи между передающими и приёмными антеннами можно получить значительно более простые аналитические выражения для средних значений ОСШ и пропускной способности MIMO-системы. Будем действовать в следующей последовательности. Сначала предположим, что число пользователей равно двум ($Q = 2$) и каждый из них имеет по одной антенне ($N = 1$), а число M передающих антенн является произвольным. Далее сделаем обобщение на случай произвольного числа Q пользователей, а затем рассмотрим MIMO-систему с произвольным числом N приёмных антенн у каждого пользователя. Отметим, что в мобильных системах сотовой связи пользователи имеют по одной приёмной антенне, поэтому случай $N = 1$ имеет самостоятельный интерес.

2.1. Случай одной приёмной антенны

В случае двух пользователей, имеющих по одной приёмной антенне, матрица коэффициентов передачи $\mathbf{H}^{(q)}$ для q-го пользователя представляет собой вектор-строку, а матрица весовых векторов передающей ДОС состоит из одного M-мерного вектора $\mathbf{V}^{(q)}$, где $q = 1, 2$. Нормированные весовые векторы ДОС, обеспечивающие согласованную с пространственным каналом передачу данных первому и второму пользователям, равны

$$\mathbf{V}^{(1)} = \frac{1}{\sqrt{\mathbf{H}^{(1)}\mathbf{H}^{(1)\mathrm{H}}}}\,\mathbf{H}^{(1)\mathrm{H}}, \qquad \mathbf{V}^{(2)} = \frac{1}{\sqrt{\mathbf{H}^{(2)}\mathbf{H}^{(2)\mathrm{H}}}}\,\mathbf{H}^{(2)\mathrm{H}}. \tag{13}$$

Среднее значение пропускной способности MIMO-системы в случае двух пользователей, имеющих по одной антенне, в соответствии с (12) определяется выражением

$$\langle C\rangle = \left\langle \log_2\!\left[1 + 0{,}5\rho_0\lambda^{(1)}\left(\mathbf{V}^{(1)\mathrm{H}}\mathbf{\Pi}^{(1)}\mathbf{V}^{(1)}\right)\right] + \log_2\!\left[1 + 0{,}5\rho_0\lambda^{(2)}\left(\mathbf{V}^{(2)\mathrm{H}}\mathbf{\Pi}^{(2)}\mathbf{V}^{(2)}\right)\right]\right\rangle, \tag{14}$$

где $\rho_0 = P_0/\sigma_0^2$. Множитель 0,5 в (14) отражает факт разделения полной мощности между двумя пользователями. Матрица-проектор $\mathbf{\Pi}^{(1)}$ проектирует весовой вектор $\mathbf{V}^{(1)}$ первого пользователя на подпространство, ортогональное весовому вектору $\mathbf{V}^{(2)}$ второго пользователя. Аналогично матрица $\mathbf{\Pi}^{(2)}$ проектирует весовой вектор $\mathbf{V}^{(2)}$ второго пользователя на подпространство, ортогональное весовому вектору $\mathbf{V}^{(1)}$ первого пользователя. С учётом вышесказанного имеем [11]

$$\mathbf{\Pi}^{(1)} = \mathbf{I} - \mathbf{V}^{(2)}\mathbf{V}^{(2)\mathrm{H}}, \qquad \mathbf{\Pi}^{(2)} = \mathbf{I} - \mathbf{V}^{(1)}\mathbf{V}^{(1)\mathrm{H}}. \tag{15}$$

Собственные числа $\lambda^{(1)}$ и $\lambda^{(2)}$ в (14) определяются выражением $\lambda^{(q)} = \mathbf{H}^{(q)\mathrm{H}}\mathbf{H}^{(q)}$, т. е. представляют собой сумму M случайных комплексных величин с нулевым средним и единичной дисперсией и, следовательно, подчиняются χ^2-распределению с $2M$ степенями свободы. При достаточно большом числе передающих антенн ($M \gg 1$) их относительные флуктуации являются малыми, и $\lambda^{(1)} \approx \lambda^{(2)} \approx M$. Это даёт возможность заменить в (14) усреднение логарифмов усреднением их аргументов. Следовательно, нахождение средней пропускной способности сводится к вычислению средних значений ОСШ для обоих пользователей.

Подставляя (15) в (11), получим, что среднее значение ОСШ для первого пользователя равно

$$\left\langle \eta^{(1)}\right\rangle = \frac{\rho_0}{2}\left(\left\langle \mathbf{H}^{(1)}\mathbf{H}^{(1)\mathrm{H}}\right\rangle - \left\langle \frac{\left|\mathbf{H}^{(1)}\mathbf{H}^{(2)\mathrm{H}}\right|^2}{\mathbf{H}^{(2)}\mathbf{H}^{(2)\mathrm{H}}}\right\rangle\right). \tag{16}$$

Аналогичное выражение для ОСШ $\langle\eta^{(2)}\rangle$ для второго пользователя можно получить из (16), если сделать в нём замену верхних индексов ($1 \to 2$ и $2 \to 1$).

Выполнить усреднение дроби в (16) достаточно сложно. Покажем, что в случае большого числа ($M \gg 1$) передающих антенн флуктуации знаменателя являются малыми по сравнению со средним значением, и ими можно пренебречь. Обозначим случайную величину $\xi^{(2)} = \mathbf{H}^{(2)}\mathbf{H}^{(2)\mathrm{H}}$. Её среднее значение и дисперсия равны соответственно $\left\langle \xi^{(2)} \right\rangle = \left\langle \mathbf{H}^{(2)}\mathbf{H}^{(2)\mathrm{H}} \right\rangle = M$ и $D_\xi = \left\langle (\mathbf{H}^{(2)}\mathbf{H}^{(2)\mathrm{H}})^2 \right\rangle - M^2$.

Чтобы найти дисперсию D_ξ, необходимо вычислить момент четвёртого порядка некоррелированных комплексных гауссовских процессов с нулевым средним, который в соответствии с [12] равен

$$\left\langle \left(\mathbf{H}^{(2)}\mathbf{H}^{(2)\mathrm{H}} \right)^2 \right\rangle = \sum_{p=1}^{M}\sum_{n=1}^{M} \left\langle h_p^{(2)} h_p^{(2)*} h_n^{(2)*} h_n^{(2)} \right\rangle =$$

$$= \sum_{p=1}^{M}\sum_{n=1}^{M} \left(\left\langle h_p^{(2)} h_p^{(2)*} \right\rangle \left\langle h_n^{(2)*} h_n^{(2)} \right\rangle + \left\langle h_p^{(2)} h_n^{(2)*} \right\rangle \left\langle h_n^{(2)} h_p^{(2)*} \right\rangle \right) = M^2 + M, \quad (17)$$

где индекс звёздочка обозначает комплексное сопряжение. Из (17) следует, что дисперсия $D_\xi = M$, а относительные флуктуации величины $\xi^{(2)}$ равны $\sqrt{D_\xi}/\langle \xi^{(2)} \rangle = M^{-1/2}$ и являются малыми при $M \gg 1$.

Аналогичный результат можно получить для случайной величины $\xi^{(1)} = \mathbf{H}^{(1)}\mathbf{H}^{(1)\mathrm{H}}$. Следовательно, при $M \gg 1$ флуктуации величин $\xi^{(1)}$ и $\xi^{(2)}$ являются малыми по сравнению с их средними значениями и могут не учитываться.

Поступая аналогично (17), найдём, что

$$\left\langle \left| \mathbf{H}^{(1)}\mathbf{H}^{(2)\mathrm{H}} \right|^2 \right\rangle = \sum_{p=1}^{M}\sum_{n=1}^{M} \left(\left\langle h_p^{(1)} h_p^{(2)*} \right\rangle \left\langle h_n^{(1)*} h_n^{(2)} \right\rangle + \left\langle h_p^{(1)} h_n^{(1)*} \right\rangle \left\langle h_n^{(2)} h_p^{(2)*} \right\rangle \right) = M. \quad (18)$$

Учтём, что $\left\langle \mathbf{H}^{(1)}\mathbf{H}^{(1)\mathrm{H}} \right\rangle = \left\langle \mathbf{H}^{(2)}\mathbf{H}^{(2)\mathrm{H}} \right\rangle$, и подставим (17) и (18) в (16). В результате получим

$$\left\langle \eta^{(1)} \right\rangle = \left\langle \eta^{(2)} \right\rangle \approx 0{,}5\rho_0 M\,(1 - M^{-1}). \quad (19)$$

Оптимальная (когерентная) передача сигналов в MIMO-системе без пространственного разделения пользователей обеспечила бы увеличение ОСШ пропорционально числу передающих антенн M, т. е. в M раз. Таким образом, энергетические потери в ОСШ за счёт разделения двух пользователей в среднем составляют величину $\nu^{(1,2)} = 1 - M^{-1}$. Этот результат имеет ясное физическое толкование. Матрица $\mathbf{\Pi}^{(1)}$ проектирует M-мерный весовой вектор $\mathbf{V}^{(1)}$ на $(M-1)$-мерное подпространство. Релеевским замираниям соответствует однородная рассеивающая среда. Поэтому вектор $\mathbf{V}^{(1)}$ является случайным вектором, равновероятно распределённым в M-мерном пространстве. При его проектировании в подпространство меньшей размерности потери в ОСШ определяются степенью уменьшения размерности этого подпространства, которая в случае двух пользователей составляет величину M^{-1}.

Из (14) получаем, что суммарная средняя пропускная способность системы с двумя разделёнными пользователями равна

$$\langle C \rangle = \left\langle C^{(1)} \right\rangle + \left\langle C^{(2)} \right\rangle \approx 2\log_2\!\left[1 + 0{,}5\rho_0 M\,(1 - M^{-1})\right]. \quad (20)$$

Полученный результат можно обобщить на случай произвольного числа $Q > 2$ пользователей, имеющих по одной приёмной антенне ($N = 1$). В этом случае матрица $\mathbf{\Pi}^{(1)}$ проектирует M-мерный весовой вектор $\mathbf{V}^{(1)}$ на $(M-Q)$-мерное подпространство. Следовательно, энергетические потери в ОСШ будут составлять величину $\nu^{(1,2)} = \left[1 - (Q-1)\,M^{-1}\right]$. Учтём также, что полная мощность теперь должна делиться равномерно между Q пользователями. Тогда для средних ОСШ и пропускной способности получим следующие выражения:

$$\left\langle \eta^{(1,2)} \right\rangle \approx \frac{\rho_0}{Q}\,M\left(1 - \frac{Q-1}{M}\right), \qquad \langle C\rangle \approx Q\log_2\left[1 + \frac{\rho_0}{Q}\,M\left(1 - \frac{Q-1}{M}\right)\right]. \tag{21}$$

2.2. Случай произвольного числа приёмных антенн

При произвольном числе N приёмных антенн у каждого пользователя собственные числа $\lambda_i^{(q)}$ в (11) и (12) удовлетворяют условию [6] $(\sqrt{M} - N)^2 < \lambda_i^{(q)} < (\sqrt{M} + N)^2$ и испытывают бóльшие флуктуации, чем при $N = 1$. Тем не менее аналогично выводу (20) при вычислении средней пропускной способности усреднение логарифмов в (12) заменим усреднением их аргументов. В результате получим

$$\langle C\rangle \approx \sum_{q=1}^{Q}\sum_{i=1}^{K}\log_2\left[1 + \left\langle \rho_i^{(q)}\lambda_i^{(q)}\left(\mathbf{V}_i^{(q)\mathrm{H}}\mathbf{\Pi}_i^{(q)}\mathbf{V}_i^{(q)}\right)\right\rangle\right]. \tag{22}$$

Отметим, что при малом числе антенн у каждого пользователя ($N \ll M$) сделанное приближение является достаточно точным из-за малых флуктуаций собственных чисел $\lambda_i^{(q)}$. При увеличении N число разделяемых пользователей уменьшается, т. к. их максимальное число удовлетворяет условию $Q \le \mathrm{int}(M/N)$. Поэтому число слагаемых в (22) уменьшается и, следовательно, погрешность выражения (22) также будет уменьшаться.

Поскольку каждый пользователь имеет N собственных каналов, его пространственное разделение с другими пользователями предполагает использование N матриц-проекторов. Полная мощность передатчика теперь должна распределяться между N собственными каналами каждого из Q пользователей. Поэтому, обобщая (21), получим, что среднее ОСШ на выходе собственных каналов каждого пользователя равно

$$\langle \eta\rangle \approx \frac{\rho_0}{QN}\,M\left(1 - \frac{QN-1}{M}\right). \tag{23}$$

Отсюда следует, что выходное ОСШ уменьшается при увеличении числа пространственно разделяемых пользователей.

Среднюю пропускную способность MIMO-системы, обеспечивающей пространственное разделение Q пользователей, можно найти с помощью следующего выражения:

$$\langle C\rangle \approx QN\log_2\left[1 + \frac{\rho_0}{QN}\,M\left(1 - \frac{QN-1}{M}\right)\right]. \tag{24}$$

Таким образом, средняя пропускная способность MIMO-системы увеличивается с ростом числа Q пользователей за счёт увеличения общего числа собственных каналов. Однако среднее ОСШ на выходе собственных каналов уменьшается при увеличении Q. Поэтому существует оптимальное число Q_{opt} пользователей для их пространственного разделения, которое зависит от полной мощности P_0 передатчика. Аналитическое выражение для Q_{opt}, которое можно получить путём дифференцирования правой части (24) по параметру Q, в общем случае имеет достаточно сложный вид.

Если излучаемая мощность является малой ($\rho_0 M/N \ll 1$), то из (24) можно получить

$$\langle C \rangle \approx \rho_0 \frac{M}{\ln 2} \left(1 - \frac{QN-1}{M} \right). \tag{25}$$

Отсюда следует, что при малой мощности передатчика оптимальным является обслуживание только одного пользователя.

В другом предельном случае достаточно большой излучаемой мощности, дифференцируя правую часть (24) по параметру Q, можно показать, что оптимальным является одновременное обслуживание максимально возможного числа пользователей, т. е. $Q_{\rm opt} = {\rm int}(M/N)$.

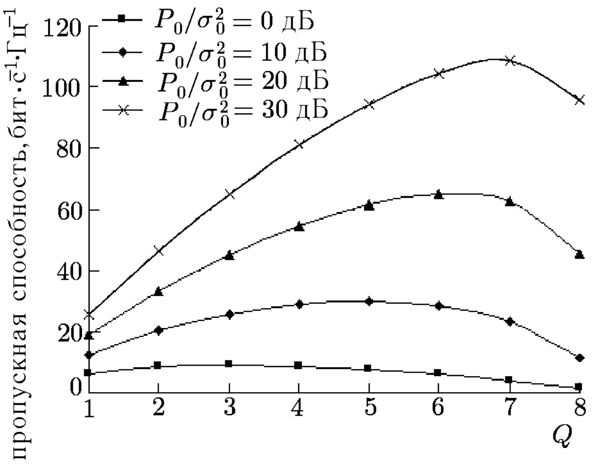


Рис. 3

На рис. 3 показана зависимость средней пропускной способности случайного релеевского канала от числа Q пользователей для разных значений относительной мощности передатчика $P_0/\sigma_0^2 = 0$; 10; 20 и 30 дБ, когда число передающих и приёмных антенн равно соответственно $M = 16$ и $N = 2$. Видно, что полная пропускная способность MIMO-системы сначала возрастает при увеличении числа пользователей, а затем начинает убывать из-за потерь в ОСШ на выходе собственных каналов. Оптимальное число пользователей также увеличивается с увеличением полной мощности и составляет $Q_{\rm opt} \approx 3$; 5; 6 и 7 для мощности $P_0/\sigma_0^2 = 0$; 10; 20 и 30 дБ соответственно, т. е. приближается к максимально возможному числу разделяемых пользователей $Q_{\max} = M/N = 8$.

3. РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Рассмотрим результаты моделирования при распространении сигналов в случайной однородной рассеивающей среде, когда коэффициенты $h_{nm}^{(q)}$ имеют релеевское распределение амплитуд и равномерное распределение фаз, а также являются некоррелированными между собой в передающих и приёмных антеннах. Будем сравнивать точные результаты для средней пропускной способности MIMO-системы, основанные на (12), с результатами, полученными с помощью приближённой аналитической формулы (24).

При использовании выражения (12) вначале формировалась одна реализация случайных канальных матриц $\mathbf{H}^{(q)}$ для всех Q пользователей. Для этой реализации матриц $\mathbf{H}^{(q)}$, где $q = 1, 2, \ldots, Q$, вычислялись векторы $\mathbf{V}^{(q)}$ и $\mathbf{U}^{(q)}$ передающей и приёмной ДОС q-го пользователя как векторы сингулярного разложения матрицы $\mathbf{H}^{(q)}$, а также собственные числа $\lambda_i^{(q)}$. Далее осуществлялось пространственное разделение пользователей. Для этого находились проекционные матрицы $\mathbf{\Pi}_i^{(q)}$ для всех пользователей и вычислялась пропускная способность MIMO-системы. Затем формировалась следующая выборка канальных матриц $\mathbf{H}^{(q)}$, и аналогичным образом вычислялось соответствующее значение пропускной способности MIMO-системы. Всего использовалось 3 000 реализаций каждой из матриц $\mathbf{H}^{(q)}$. Моделирование завершалось нахождением среднего значения $\langle C \rangle$ пропускной способности MIMO-системы.

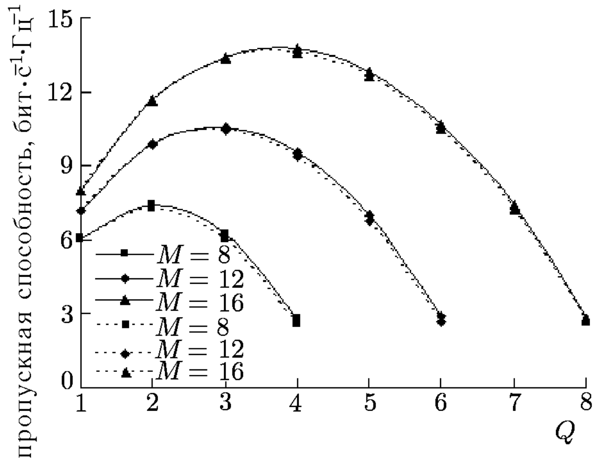


Рис. 4

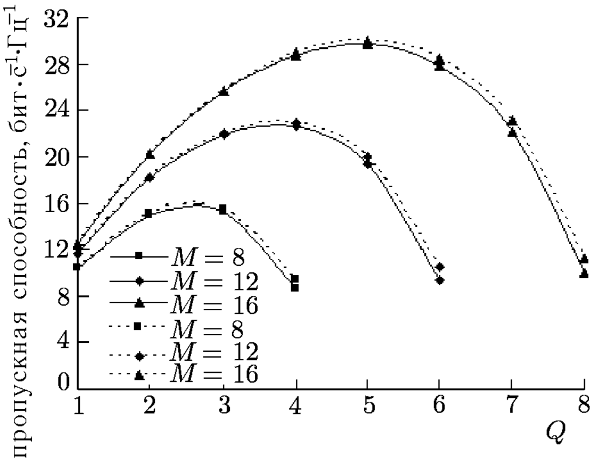


Рис. 5

На рис. 4 показана средняя пропускная способность MIMO-системы в зависимости от числа Q пользователей для разного числа передающих антенн $M = 8$; 12 и 16. Число приёмных антенн $N = 2$, а полная относительная мощность передатчика $P_0/\sigma_0^2 = 3$ дБ. Аналогичные результаты для $P_0/\sigma_0^2 = 10$ дБ приведены на рис. 5. Сплошные кривые соответствуют точной формуле (12), а пунктирные кривые получены с помощью приближённого аналитического выражения (24). Видно, что формула (24) обеспечивает высокую точность, достаточную для практического использования.

Приведённые результаты показывают высокую эффективность пространственного разделения пользователей в MIMO-системе, использующей параллельные собственные каналы для передачи данных. Так, пропускная способность увеличивается в 2,4 раза (с 12,5 до 30 бит $\cdot$ с$^{-1} \cdot$ Гц$^{-1}$) при 16-ти передающих и 2-х приёмных антеннах и относительной мощности $P_0/\sigma_0^2 = 10$ дБ. Из рис. 4 и 5 также следует, что пространственное разделение пользователей увеличивает пропускную способность тем значительнее, чем больше антенн используется на базовой станции. Например, если относительная мощность $P_0/\sigma_0^2 = 10$ дБ, то оптимальное число пользователей составляет $Q_{\mathrm{opt}} \approx 3$; 4 и 5, при $M = 8$; 12 и 16 соответственно. При этом пропускная способность MIMO-системы увеличивается в 1,5; 2,0 и 2,4 раза.

4. ЗАКЛЮЧЕНИЕ

В настоящей работе рассмотрен случай некоррелированных релеевских замираний сигналов в приёмных антеннах, который является наиболее актуальным для систем связи, работающих в городских условиях. Получены приближённые аналитические выражения для средних значений ОСШ и пропускной способности MIMO-системы, использующей параллельные собственные каналы для передачи информации. Эти выражения являются значительно более простыми по сравнению с точными формулами, однако обеспечивают высокую точность при произвольном числе передающих и приёмных антенн и произвольной мощности передатчика. Показано, что пространственное разделение пользователей приводит к уменьшению ОСШ на выходе собственных каналов, поэтому полная пропускная способность MIMO-системы сначала возрастает с ростом числа пользователей, а затем начинает убывать из-за энергетических потерь в ОСШ. Установлено, что оптимальное число пользователей, обеспечивающее максимальную пропускную способность, увеличивается при увеличении мощности передатчика и числа передающих антенн и уменьшается с ростом числа приёмных антенн. При увеличении мощности оптимальное число пользователей приближается к их максимальному возможному числу, равному отношению числа передающих антенн к числу приёмных антенн. Полученные результаты показывают высокую эффективность пространственного разделения пользователей в MIMO-системах связи, использующих параллельные собственные каналы для передачи информации.

Данная работа выполнена при поддержке Совета по поддержке ведущих научных школ (грант № НШ-1729.2003.2) и РФФИ (грант № 03–02–17141).

СПИСОК ЛИТЕРАТУРЫ

1. Kohno R. // IEEE Personal Communications. 1998. P. 28.
2. Vornefeld U., Walk C., Walk B. // IEEE Communications Magazine. 1999. V. 37, No. 11. P. 52.
3. Sheikh K., Gesbert D., Gore D., Paulraj A. // IEEE Communications Magazine. 1999. V. 37, No. 11. P. 100.
4. Pedersen K. I., Mogensen P. E., Fleury B. H. // IEEE Trans. Vehicular Technology. 2000. V. 49, No. 2. P. 437.
5. Флаксман А. Г. // Изв. вузов. Радиофизика. 2002. Т. 45, № 11. С. 986.
6. Andersen J. B. // IEEE Anten. Propag. Magazine. 2000. V. 42, No. 2. P. 12.
7. Ермолаев В. Т., Маврычев Е. А., Флаксман А. Г. // Зарубежная радиоэлектроника. Успехи современной радиоэлектроники. 2001. № 9. С. 50.
8. Maltsev A. A., Rubtsov A. E., Tiraspolsky S. A. // Proc. 5th Sci. Conf. Radio Physics Devoted to the 100th Anniversary of A. A. Andronov's Birth, May 7, 2001, Nizhny Novgorod, Russia.
9. Parsons J. D. The Mobile Radio Propagation Channel. London: Pentech Press Publisher, 1994.
10. Прокис Дж. Цифровая связь. М.: Радио и связь, 2000.
11. Воеводин В. В. Линейная алгебра. М.: Наука, 1980.
12. Евсиков Ю. А., Чапурский В. В. Преобразование случайных процессов в радиотехнических устройствах. М.: Высшая школа, 1977.

[1] Нижегородский госуниверситет им. Н. И. Лобачевского;
[2] Нижегородский государственный технический университет, г. Нижний Новгород, Россия

Поступила в редакцию 5 мая 2003 г.

EFFECTIVENESS OF SPACE DIVISION MULTIPLE ACCESS IN MIMO SYSTEMS WITH PARALLEL DATA TRANSMISSION

V. T. Ermolayev, A. G. Flaksman, I. M. Averin, and D. V. Gribov

We consider multiple-input multiple-output (MIMO) cellular communication systems with antenna arrays at both link ends and parallel channels for data transmission. These channels (the so-called eigenchannels) are matched with the random spatial propagation channel. We analyze the effectiveness of a space-division multiple-access (SDMA) method, which does not require estimation of signal-arrival directions and is based on orthogonalization of the parallel channels of all users. Approximate analytical formulas for the mean signal-to-noise ratio and the total capacity of a MIMO system are derived in the case of Rayleigh fading of signals. The obtained formulas are much simpler than the exact ones, but ensure high accuracy for arbitrary numbers of transmitting and receiving antennas and an arbitrary power of transmitter. Our results demonstrate the high effectiveness of the proposed SDMA method.

УДК 519.217:517.977.57

АЛГОРИТМ ОБНАРУЖЕНИЯ ПОСЛЕДОВАТЕЛЬНОСТИ ИМПУЛЬСНЫХ СИГНАЛОВ СО СЛУЧАЙНЫМИ МОМЕНТАМИ ПОЯВЛЕНИЯ

А. В. Королёв, А. М. Силаев

Методами марковской теории оптимальной нелинейной фильтрации получен алгоритм оптимального обнаружения последовательности импульсных сигналов со случайными моментами появления на фоне белых гауссовских шумов в дискретном времени. Проведено исследование статистических характеристик синтезированного алгоритма с помощью компьютерного моделирования. Получены зависимости вероятности правильного обнаружения последовательности сигналов от отношения сигнал/шум и интервалов между импульсами. Показано, что применение методов оптимальной нелинейной фильтрации позволяет улучшить точность обнаружения последовательности импульсных сигналов по сравнению с алгоритмами, использующими согласованную фильтрацию отдельных импульсов в пакете.

1. ПОСТАНОВКА ЗАДАЧИ

Проблема оптимального обнаружения последовательности импульсных сигналов известной формы со случайными амплитудами и моментами появления возникает в радиосвязи при многолучевом распространении волн, в радиолокации при приёме отражённых сигналов от целей со сложной конфигурацией, а также в ряде других областей науки и техники. Для решения подобных задач обычно применяется согласованная фильтрация импульсных сигналов [1, 2]. Однако алгоритмы согласованной фильтрации могут оказаться неэффективными, если мало́ отношение сигнал/шум для каждого отдельного импульсного сигнала в наблюдаемой реализации, а интервалы между импульсами заранее неизвестны. В настоящей работе для обнаружения последовательности импульсных сигналов со случайными моментами появления используются методы нелинейной марковской фильтрации [3–6]. В работах [7, 8] рассматривались задачи оптимальной нелинейной фильтрации импульсных сигналов со случайными моментами появления и оптимального оценивания неизвестных моментов появления импульсов на фоне шумов. Были получены алгоритмы обработки сигналов в дискретном времени, которые, как показано в настоящей работе, можно применить и для задачи оптимального обнаружения последовательности случайных импульсных сигналов.

Пусть наблюдаемая векторная последовательность $\mathbf{y}^{(T)} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T\}$ описывается линейными уравнениями в дискретном времени:

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{z}_k + \boldsymbol{\eta}_k, \tag{1}$$

где $\mathbf{H}_k$ — заданная матрица, $\boldsymbol{\eta}_k$ — аддитивный шум наблюдений, T — число выборочных значений сигналов за время наблюдения, $k = 1, 2, 3, \ldots, T$. Последовательность шумов наблюдений $\{\boldsymbol{\eta}_k\}$ образована из взаимно независимых случайных векторных величин с плотностями вероятности $\rho_{\boldsymbol{\eta}}(\boldsymbol{\eta}_k, k)$.

Предположим, что входящий в наблюдения (1) информационный векторный сигнал $\{\mathbf{z}_k\}$ в случае статистической гипотезы H_0 представляет собой отклик некоторой линейной системы на входной шум $\boldsymbol{\xi}_k$. В случае же статистической гипотезы H_1 данный сигнал является реакцией линейной системы на сумму шумовых и импульсных воздействий:

$$\mathrm{H}_0\colon\ \mathbf{z}_{k+1} = \mathbf{F}_k\mathbf{z}_k + \mathbf{G}_k\boldsymbol{\xi}_k;$$

$$\mathrm{H}_1\colon\ \mathbf{z}_{k+1} = \mathbf{F}_k\mathbf{z}_k + \mathbf{G}_k\boldsymbol{\xi}_k + \sum_{j=1}^{M}\mathbf{A}_j\delta(k,\tau_j). \tag{2}$$

Здесь $\mathbf{F}_k$, $\mathbf{G}_k$ — матрицы, описывающие параметры линейной системы и шумовых возмущений, $\sum_{j=1}^{M}\mathbf{A}_j\delta(k,\tau_j)$ — последовательность M импульсных возмущений со случайными моментами появления τ_j и амплитудами $\mathbf{A}_j$, $\delta(k,\tau)$ — символ Кронекера, $k = 0,1,2,\ldots,T-1$. Предположим, что амплитуды импульсных возмущений $\mathbf{A}_j$ статистически взаимно независимы и имеют известные плотности вероятности $P_{\mathbf{A}_j}(\mathbf{A}_j)$. Будем считать, что последовательность $\{\boldsymbol{\xi}_k\}$ составлена из взаимно независимых в различные моменты времени случайных векторных величин, которые статистически не зависят от начального значения вектора состояния системы $\mathbf{z}_0$, от амплитуд и моментов появления импульсов и описываются плотностями вероятности $\rho_{\boldsymbol{\xi}}(\boldsymbol{\xi}_k,k)$. Предположим также, что при $k=0$ задана плотность вероятности $P_{\mathbf{z}_0}(\mathbf{z}_0)$ начального значения сигнала $\mathbf{z}_0$.

Задача состоит в том, чтобы в момент окончания наблюдений $k=T$ по известной реализации наблюдаемой последовательности $\mathbf{y}^{(T)}$ определить, какая из статистических гипотез H_0 и H_1 верна. Различение статистических гипотез H_0 и H_1 можно интерпретировать как задачу оптимального обнаружения последовательности случайных импульсных возмущений $\sum_{j=1}^{M}\mathbf{A}_j\delta(k,\tau_j)$ динамической системы, описываемой уравнениями (2).

2. АЛГОРИТМ ВЫЧИСЛЕНИЯ ОТНОШЕНИЯ ПРАВДОПОДОБИЯ В ДИСКРЕТНОМ ВРЕМЕНИ

Как известно [1], для решения задачи различения статистических гипотез H_0 и H_1 по наблюдаемой реализации $\mathbf{y}^{(T)}$ необходимо вычислить отношение правдоподобия

$$\Lambda(k) = P(\mathbf{y}^{(k)}\,|\,\mathrm{H}_1)/P(\mathbf{y}^{(k)}\,|\,\mathrm{H}_0) \tag{3}$$

и сравнить его при $k=T$ с порогом h, величина которого определяется заранее исходя из выбранного критерия оптимальности.

Для рассматриваемой задачи оптимального различения статистических гипотез H_0 и H_1 отношение правдоподобия $\Lambda(k)$ можно представить в виде разложения в ряд по числу импульсов [8]:

$$\Lambda(k) = P(\mathbf{y}^{(k)}\,|\,\mathrm{H}_1)/P(\mathbf{y}^{(k)}\,|\,\mathrm{H}_0) = \sum_{j=0}^{M} P(\mathbf{y}^{(k)}\,|\,\mathrm{H}_1,\Gamma_j)P(\Gamma_j\,|\,\mathrm{H}_1)/P(\mathbf{y}^{(k)}\,|\,\mathrm{H}_0) =$$
$$= \sum_{j=0}^{M} P(\mathbf{y}^{(k)}\,|\,\mathrm{H}_1,\Gamma_j)P(\Gamma_j\,|\,\mathrm{H}_1)/P(\mathbf{y}^{(k)}\,|\,\mathrm{H}_1,\Gamma_0) = \sum_{j=0}^{M} P(\Gamma_j\,|\,\mathrm{H}_1)\Lambda_j(k). \tag{4}$$

Здесь Γ_j — случайное событие, состоящее в том, что к моменту времени k в модели сигнала (2) появилось j импульсов. Функции $P(\Gamma_j\,|\,\mathrm{H}_1) = P(\Gamma_j) = p_{j\,\mathrm{pr}}(k) = P(\tau_j < k \le \tau_{j+1})$ имеют смысл априорных вероятностей появления ровно j импульсов к моменту времени k. С учётом упорядоченности моментов появления импульсов $0 < \tau_1 < \tau_2 < \ldots < \tau_M$ они выражаются через функции априорных вероятностей $P_{\tau_j}(\tau_j)$ следующим образом:

$$p_{0\mathrm{pr}}(k) = \sum_{\tau_1=k}^{\infty} P_{\tau_1}(\tau_1), \qquad p_{M\mathrm{pr}}(k) = \sum_{\tau_M=0}^{k-1} P_{\tau_M}(\tau_M),$$

$$p_{n\mathrm{pr}}(k) = \sum_{\tau_{n+1}=k}^{\infty} P_{\tau_{n+1}}(\tau_{n+1}) - \sum_{\tau_n=k}^{\infty} P_{\tau_n}(\tau_n), \tag{5}$$

где $n = 1, 2, \ldots, M-1$; $k \geq 0$.

В результате из (4) получаем соотношение

$$\Lambda(k) = \sum_{j=0}^{M} p_{j\mathrm{pr}}(k)\Lambda_j(k), \tag{6}$$

где введены частные отношения правдоподобия $\Lambda_j(k)$:

$$\Lambda_j(k) = \frac{P(\mathbf{y}^{(k)} \,|\, \mathrm{H}_1, \Gamma_j)}{P(\mathbf{y}^{(k)} \,|\, \mathrm{H}_1, \Gamma_0)} = \frac{P(\mathbf{y}^{(k)} \,|\, \Gamma_j)}{P(\mathbf{y}^{(k)} \,|\, \Gamma_0)} = \frac{P(\Gamma_j \,|\, \mathbf{y}^{(k)}) P(\Gamma_0)}{P(\Gamma_0 \,|\, \mathbf{y}^{(k)}) P(\Gamma_j)} = \frac{p_j(k) p_{0\mathrm{pr}}(k)}{p_0(k) p_{j\mathrm{pr}}(k)}\,. \tag{7}$$

Отсюда видно, что отношение правдоподобия $\Lambda(k)$ может быть вычислено с помощью уравнений (6), (7) через функции апостериорных и априорных вероятностей появления различного количества импульсов к моменту времени k. При этом априорные вероятности $p_{j\mathrm{pr}}(k)$ находятся заранее до проведения наблюдений по формулам (5). А для нахождения функций апостериорных вероятностей появления импульсов $p_j(k) = P(\Gamma_j \,|\, \mathbf{y}^{(k)}) = P(\tau_j < k \leq \tau_{j+1} \,|\, \mathbf{y}^{(k)})$ необходимо сделать определённые предположения относительно статистических характеристик моментов появления импульсов $\tau_1, \tau_2, \ldots, \tau_M$.

Будем считать, что условные вероятности случайных моментов появления импульсов τ_n удовлетворяют условиям

$$P(\tau_n \,|\, \tau_{n-1}, \tau_{n-2}, \ldots, \tau_1) = P(\tau_n \,|\, \tau_{n-1}),$$
$$P(\tau_n \,|\, \tau_{n-1} < k \leq \tau_n, \tau_{n-1}) = P(\tau_n \,|\, \tau_{n-1} < k \leq \tau_n). \tag{8}$$

Отметим, что для совокупности $\boldsymbol{\tau} = \{\tau_1, \tau_2, \ldots, \tau_M\}$ при этом выполняются «марковские» соотношения

$$P_{\boldsymbol{\tau}}(\boldsymbol{\tau}) = P(\tau_1) \prod_{n=2}^{M} P(\tau_n \,|\, \tau_{n-1}),$$

а для нахождения апостериорных вероятностей появления импульсов $p_j(k)$ можно использовать алгоритмы оптимальной нелинейной фильтрации марковских случайных импульсных сигналов. В частности, при таких предположениях в [8] получены уравнения для вычисления функций апостериорных вероятностей $p_j(k)$ по наблюдениям $\{\mathbf{y}^{(k)}\}$ непосредственно в текущем времени. В итоге уравнения (6), (7) настоящей работы вместе с уравнениями (16)–(22) статьи [8] позволяют найти отношение правдоподобия $\Lambda(k)$ и составляют искомый алгоритм оптимального обнаружения последовательности импульсных сигналов со случайными моментами появления в дискретном времени.

3. РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Компьютерное моделирование и исследование статистических характеристик полученного алгоритма проводились для частного случая скалярных моделей наблюдений и сигналов:

$$\mathrm{H}_0\colon\ y_k = \eta_k; \qquad \mathrm{H}_1\colon\ y_k = z_k + \eta_k; \tag{9}$$

$$z_k = 0{,}95 z_{k-1} - 0{,}65 z_{k-2} + \sum_{j=1}^{M} \delta(k, \tau_j), \tag{10}$$

где $k = 0, 1, \ldots, T$, а начальные условия выбирались нулевыми: $z_{-2} = z_{-1} = z_0 = 0$. Предполагалось, что шумы наблюдений $\{\eta_k\}$ составлены из последовательности независимых гауссовских случайных величин с нулевыми средними значениями и одинаковыми дисперсиями Q.

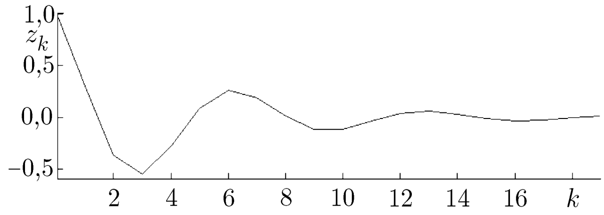


Рис. 1. Отклик динамической системы на отдельный импульс

Таким образом, выходной сигнал z_k динамической системы формировался под влиянием импульсных воздействий в случайные моменты времени τ_j с одинаковыми единичными амплитудами. При этом шумовые возмущения, действующие на динамическую систему, для простоты считались нулевыми. При таких параметрах отклик системы на отдельное импульсное возмущение представлял собой синусоиду с характерным временем затухания порядка 15 отсчётов дискретного времени (см. рис. 1).

Отметим, что уравнения (9), (10) могут быть представлены в виде (1), (2), если ввести двумерные векторы и матрицу

$$\mathbf{z}_k = \begin{pmatrix} z_k \\ z_{k-1} \end{pmatrix}, \qquad \mathbf{H} = (1 \quad 0), \qquad \mathbf{A}_j = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \mathbf{F} = \begin{pmatrix} 0{,}95 & -0{,}65 \\ 1 & 0 \end{pmatrix}.$$

Для описанной модели статистических гипотез исследовались показатели точности обнаружения при использовании алгоритма, работающего в соответствии с уравнениями (6), (7). Для сравнения исследовались также характеристики алгоритма обнаружения, основанного на согласованной фильтрации отдельных импульсов в пакете. При этом наблюдаемый сигнал пропускался через линейный фильтр, согласованный с откликом динамической системы на отдельный импульс (см. рис. 1), а затем выходной сигнал согласованного фильтра использовался для решения задачи обнаружения. Отметим, что в случае, когда заранее известно, что может появиться всего один импульс, т. е. $M = 1$, алгоритм обнаружения, основанный на согласованной фильтрации, при гауссовских шумах наблюдений является наилучшим для достаточно широкого класса различных критериев оптимального обнаружения [1]. Если же импульсных сигналов в принимаемой реализации может быть достаточно много и они имеют случайные моменты появления, то характеристики алгоритма обнаружения, основанного на согласованной фильтрации, можно улучшить, применяя методы нелинейной фильтрации сигналов.

При моделировании для каждого алгоритма анализировались случаи появления двух, четырёх и восьми импульсов в наблюдаемой реализации. При этом для предотвращения перекрытия импульсных сигналов в наблюдениях моменты появления соседних импульсов в пакете специально выбирались достаточно далеко друг от друга, а длительность интервала наблюдений менялась пропорционально числу приходящих импульсов. При $M = 2$ расчёт проводился для интервала времени длительностью $T = 100$, при $M = 4$ — для интервала $T = 200$, а при $M = 8$ — для $T = 400$. В алгоритме обнаружения моменты появления импульсов были заранее неизвестны, но априорные вероятности моментов появления импульсов предполагались распределёнными по геометрическому закону:

$$P(\tau_1) = \nu\,(1-\nu)^{\tau_1},$$

$$P(\tau_n \,|\, \tau_{n-1}) = \begin{cases} \nu\,(1-\nu)^{\tau_n - \tau_{n-1} - 1}, & \tau_n > \tau_{n-1}; \\ 0, & \tau_n \le \tau_{n-1}, \end{cases}$$

где $n = 2, 3, \ldots, M$; $\tau_1 > 0$; $\tau_n > \tau_{n-1}$, а величина $\nu < 1$ имеет смысл частоты появления импульсов. При моделировании эта величина выбиралась равной $\nu = 0{,}02$, что соответствовало появлению в среднем одного импульса за пятьдесят отсчётов времени.

При статистическом анализе показателей обнаружении пакета импульсов в рассматриваемых моделях использовался критерий Неймана—Пирсона. В процессе моделирования сначала вычислялось отношение правдоподобия для реализаций, когда в наблюдениях импульсов нет, а присутствует только шум наблюдений, т. е. верна статистическая гипотеза H_0. В результате этих вычислений были получены дифференциальные функции распределения $W_\Lambda(\Lambda)$ для отношения правдоподобия в момент окончания наблюдений $k = T$ при различных значениях дисперсии шума наблюдений Q. Например, на рис. 2 представлен график функции $W_\Lambda(\Lambda)$, построенный при $Q = 1$ по ансамблю из 10 000 независимых реализаций наблюдений. Затем, исходя из определённого значения вероятности ложной тревоги P_{FA}, выбирался порог обнаружения для каждого конкретного значения Q. На рис. 2 вертикальной чертой отмечены уровни порогов Λ при $P_{\mathrm{FA}} = 10^{-2}$ и $P_{\mathrm{FA}} = 10^{-3}$. Отметим, что площади под графиком выше пороговых значении соответственно равны 10^{-2} и 10^{-3}.

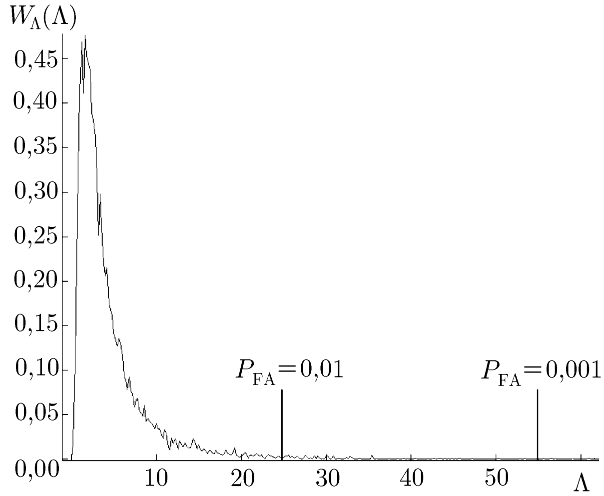


Рис. 2. График функции $W_\Lambda(\Lambda)$, построенный по ансамблю из 10 000 независимых реализаций наблюдений при отсутствии импульсного сигнала в наблюдениях

Затем для полученных таким образом порогов обнаружения вычислялись кривые обнаружения, т. е. вероятности правильного обнаружения P_{CD} последовательности импульсов при фиксированных уровнях вероятности ложной тревоги в зависимости от отношения сигнал/шум для отдельного импульса. Под величиной сигнал/шум понималось соотношение $A_j^2/Q = 1/Q$, где $A_j = 1$ — амплитуда импульсного воздействия, Q — дисперсия шума наблюдений. На рис. 3*а–в* приведены кривые обнаружения для оптимального алгоритма (6), (7) при вероятности ложной тревоги $P_{\mathrm{FA}} = 10^{-2}$ и 10^{-3} для двух, четырёх и восьми импульсов в наблюдаемой реализации соответственно.

Для сравнения на рис. 4 показаны кривые обнаружения, полученные при тех же параметрах сигналов и наблюдений с помощью алгоритма, основанного на согласованной фильтрации отдельных импульсных сигналов.

Как показали результаты моделирования (см. рис. 4), качество обнаружения при использовании согласованных фильтров практически не зависит от числа импульсов в наблюдаемой последовательности. В то же время характеристики алгоритма обнаружения с применением методов нелинейной фильтрации, как видно из рис. 3, заметно улучшаются с ростом числа учитываемых импульсных возмущений в модели сигнала. Сравнение кривых обнаружения на рис. 3 и 4 позволяет сделать вывод, что уже при $M = 8$ для обнаружения целесообразно использовать алгоритм, основанный на нелинейной фильтрации марковских процессов.

С помощью компьютерного моделирования проводился также статистический анализ характеристик обнаружения при малых интервалах времени между соседними импульсами, когда происходит наложение отдельных импульсных сигналов в наблюдаемой реализации y_k.

На рис. 5 показаны результаты моделирования — кривые обнаружения для реализаций наблюдений, когда интервалы между импульсами были равны соответственно одному, трём, пяти и де-

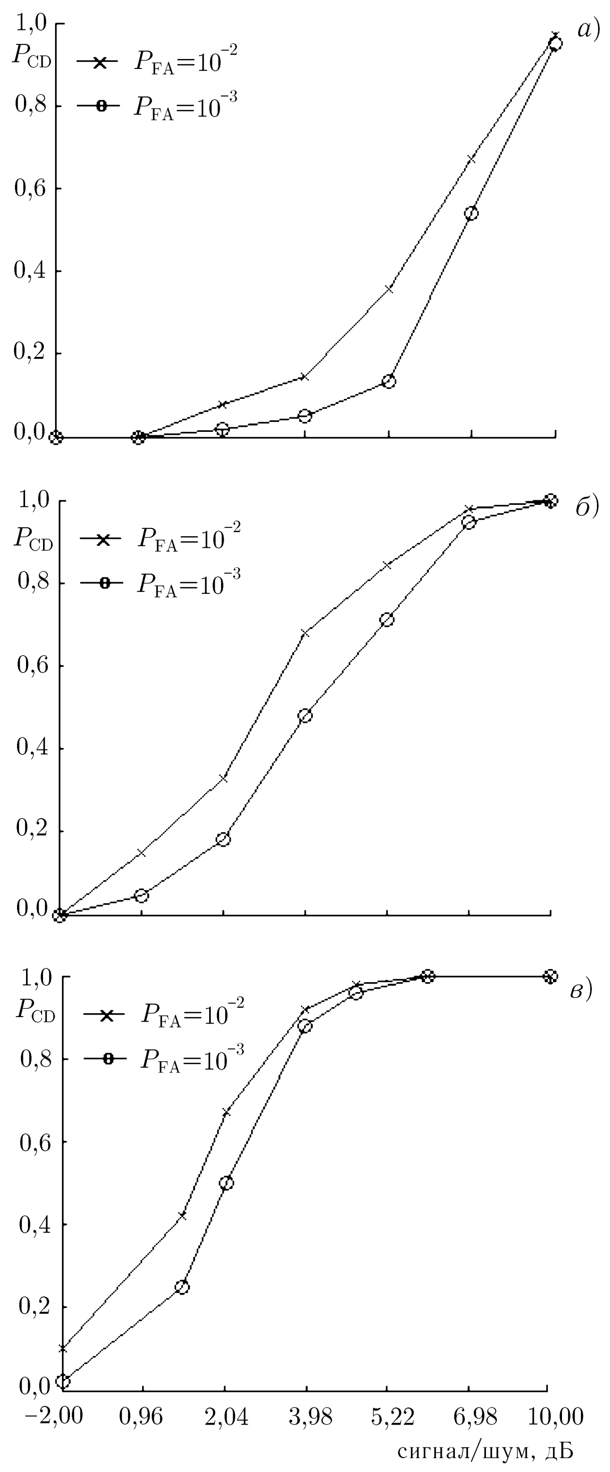


Рис. 3. Кривые обнаружения для полученного алгоритма при различных вероятностях ложной тревоги для двух (*а*), четырёх (*б*) и восьми (*в*) импульсов в наблюдаемой реализации

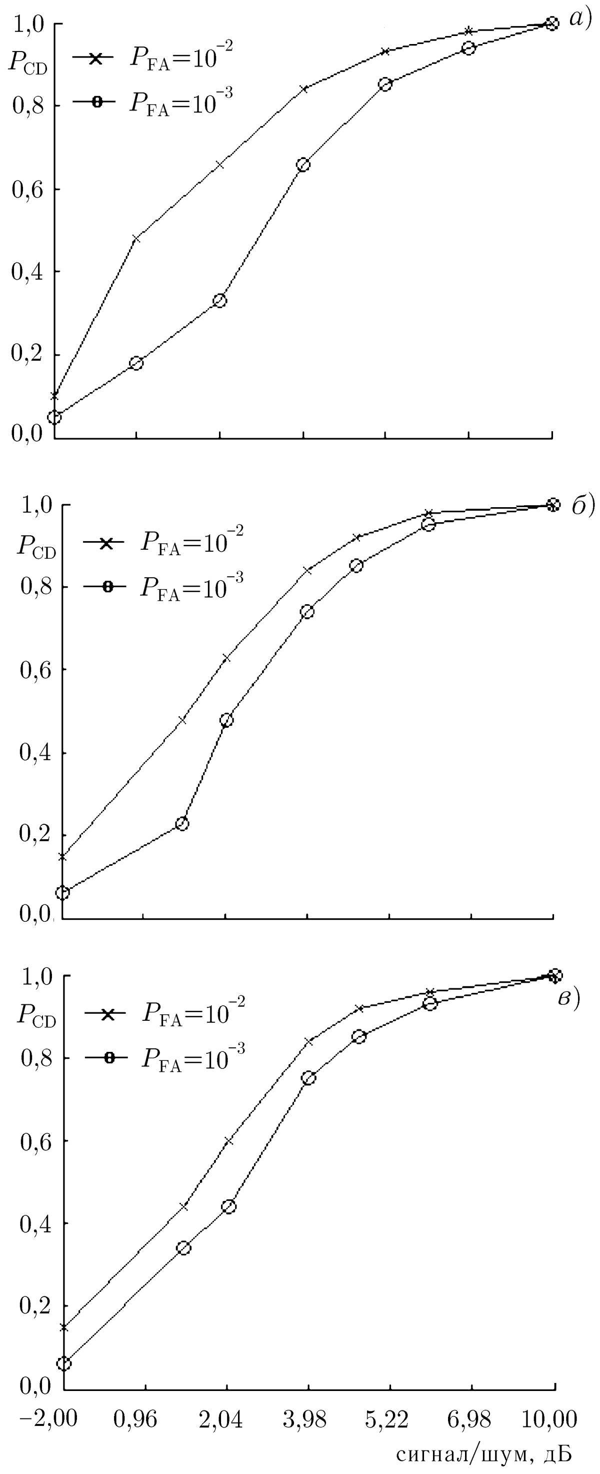


Рис. 4. Кривые обнаружения для алгоритма, основанного на согласованной фильтрации отдельных импульсных сигналов, при различных вероятностях ложной тревоги для двух (*а*), четырёх (*б*) и восьми (*в*) импульсов в наблюдаемой реализации

сяти отсчётам в дискретном времени. Данные кривые были получены для алгоритма, учитывающего два импульса ($M = 2$) при вероятности ложной тревоги $P_{\rm FA} = 10^{-2}$. Как видно из графиков, характеристики обнаружения улучшаются, если импульсы приходят так, что их отклики в наблюдениях складываются с одним знаком, и ухудшаются, если отклики импульсов складываются с разными знаками. Однако в целом показатели обнаружения остаются достаточно стабильными при различных временны́х задержках импульсов: оптимальный алгоритм позволяет обнаруживать последовательности импульсных сигналов даже в случае перекрытия отдельных импульсов в противофазе.

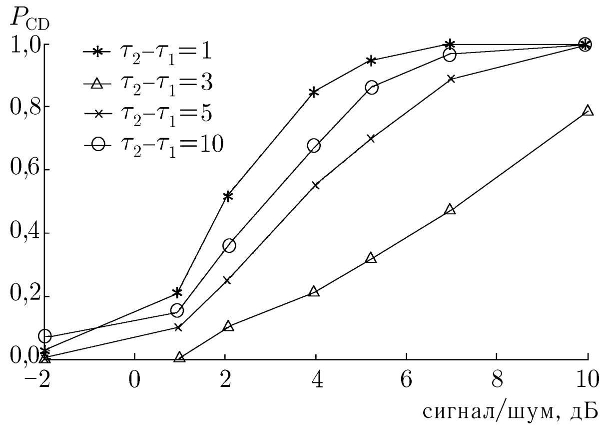


Рис. 5. Кривые обнаружения для реализации с различными интервалами между импульсами

Результаты математического моделирования показали, что исследуемый алгоритм оптимального обнаружения импульсных сигналов со случайными моментами появления вполне работоспособен, обладает хорошими характеристиками в случае большого числа импульсов и в то же время в силу особенности рекуррентного вычисления отношения правдоподобия характеризуется сравнительно высоким быстродействием.

Работа выполнена при поддержке Совета по грантам Президента РФ для поддержки ведущих научных школ (грант НШ-1729.2003.2).

СПИСОК ЛИТЕРАТУРЫ

1. Акимов П. С., Бакут П. А. Теория обнаружения сигналов. М.: Радио и связь, 1984.
2. Тихонов В. И. Оптимальный приём сигналов. М.: Радио и связь, 1983.
3. Стратонович Р. Л. Условные марковские процессы и их применение к теории оптимального управления. М.: Изд-во МГУ, 1966.
4. Липцер Р. Ш., Ширяев А. Н. Статистика случайных процессов. М.: Наука, 1974.
5. Тихонов В. И., Кульман Н. К. Нелинейная фильтрация и квазикогерентный приём сигналов. М.: Сов. радио, 1975.
6. Тихонов В. И., Харисов Н. К. Статистический анализ и синтез радиотехнических устройств и систем. М.: Радио и связь, 1991.
7. Ванжа А. В., Мальцев А. А., Силаев А. М. // Изв. вузов. Радиофизика. 1993. Т. 36, № 6. Т. 498.
8. Королёв А. В., Силаев А. М. // Изв. вузов. Радиофизика. 2002. Т. 45, № 3. С. 254.

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

Поступила в редакцию
27 марта 2003 г.

ALGORITHM FOR DETECTION OF A GROUP OF PULSED SIGNALS WITH RANDOM MOMENTS OF APPEARANCE

A. V. Korolev and A. M. Silaev

Using the methods of optimal nonlinear Markov filtering, we obtain an algorithm for optimal detection of a group of pulsed signals with random amplitudes and random moments of appearance against the background of white Gaussian noise. Statistical characteristics of the algorithm are studied numerically. Dependences of the probability of correct detection on the signal-to-noise ratio and the time interval between the pulses are obtained. It is shown that the use of optimal nonlinear filtering methods makes it possible to improve the quality of detection of a group of pulsed signals in comparison with methods based on the matched filtering of individual pulsed signals in a group.

УДК 519.217.4

К ВОПРОСУ ОБ ОБОБЩЁННОМ ПРОЦЕССЕ ОРНШТЕЙНА—УЛЕНБЕКА

А. И. Саичев, С. Г. Уткин

Основываясь на аномально диффузионной природе случайных блужданий, сконструирован модельный обобщённый процесс Орнштейна—Уленбека. Найдены асимптотики его основных статистических характеристик. Показано, что при выполнении центральной предельной теоремы асимптотики подчиняются классическому закону диффузии, а при её нарушении — аномальному.

ВВЕДЕНИЕ

В последнее время большое внимание уделяется аномально диффузионным явлениям. Они рассматриваются в различных областях современной науки: хаотической динамике гамильтоновых систем [1], физике плазмы [2], теории турбулентности [3], физике твёрдого тела [4], физической химии [5] (см. также [6] и обзор [7]).

Один из наиболее плодотворных методов описания диффузионных явлений основывается на детальном исследовании модельных процессов, чьи вероятностные распределения являются точными решениями тех или иных диффузионных уравнений. Одним из таких процессов является процесс Орнштейна—Уленбека, описывающий поведение совершающей броуновское движение частицы, при воздействии на неё упругой силы. Характерным его свойством является то, что он может быть легко преобразован в винеровский процесс, плотность вероятностей которого подчиняется классическому уравнению диффузии.

Многие авторы для удобства вводят так называемый «дробный процесс Орнштейна—Уленбека» [5–7], заменяя в уравнении Фоккера—Планка простые дифференциальные операторы их дробными аналогами (см., например, [5]). В этой статье сконструирован обобщённый процесс Орнштейна—Уленбека и исследованы некоторые его статистические характеристики.

ОБОБЩЁННЫЙ ПРОЦЕСС ОРНШТЕЙНА—УЛЕНБЕКА

Как известно, обычный процесс Орнштейна—Уленбека удовлетворяет следующему уравнению Ланжевена:

$$\mathrm{d}v/\mathrm{d}t = -\gamma v + \xi(t),$$

где $\xi(t)$ — гауссов белый шум. Дисперсия такого процесса подчиняется классическому линейному закону диффузии. Подставим вместо белого шума случайные блуждания, благодаря которым система приобретает аномально диффузионный характер [8].

Рассмотрим процесс $X(t)$, удовлетворяющий уравнению Ланжевена:

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = V(t), \qquad \frac{\mathrm{d}V(t)}{\mathrm{d}t} + hV(t) = \sum_k \eta_k\, \delta(t - t_k), \tag{1}$$

где $\{\eta_k\}$ — гауссовы независимые случайные величины с нулевыми средними и одинаковыми дисперсиями σ^2, в то время как $\{\ldots, t_{-1}, t_0, t_1, \ldots, t_k, \ldots\}$ — стационарный поток случайных моментов времени со статистически независимыми интервалами между ними $\tau_n = t_n - t_{n-1}$, $\delta(t)$ — дельта-функции Дирака. Будем также считать, что плотности вероятностей случайных длительностей интервалов не зависят от номера интервала и равны $f(\tau)$. Случайный процесс в правой части второго уравнения (1) представляет собой последовательные скачки случайной силы с амплитудой $\{\eta_k\}$, действующей на частицу, в моменты времени $\{t_k\}$, между которыми скорость частицы не изменяется.

Решение уравнения (1) можно записать в виде

$$X(t) = \sum_{k=1}^{N(t)} \frac{\eta_k}{h} \left[1 - \exp(-hT_k)\right].$$

Здесь $N(t)$ — количество скачков за время наблюдения t (начало наблюдения совпадает с моментом нулевого скачка t_0), а T_n — интервал от начала наблюдений до n-го следующего момента столкновения. Плотность вероятностей этого интервала равна

$$h_n(\tau) = f(\tau) \underbrace{* \ldots *}_{n-1} f(\tau), \tag{2}$$

где звёздочка обозначает операцию свёртки.

Видно, что среднее процесса $X(t)$ равно нулю. Вычислим его второй момент:

$$\langle X^2(t)\rangle = \left\langle \sum_{k=1}^{N(t)} \sum_{l=1}^{N(t)} \frac{\eta_k \eta_l}{h^2} \{1 - \exp(-hT_k) - \exp(-hT_l) + \exp[-h\,(T_k + T_l)]\} \right\rangle .$$

Воспользовавшись свойствами случайных величин $\{\eta_k\}$, перепишем полученное равенство в виде

$$\langle X^2(t)\rangle = \frac{\sigma^2}{h^2} \left\langle \sum_{k=1}^{N(t)} [1 - 2\exp(-hT_k) + \exp(-2hT_k)] \right\rangle .$$

Последнее выражение неудобно для исследования, поскольку в предел суммы входит случайная величина с неизвестным распределением. Поэтому введём вспомогательную функцию

$$\Pi_n[N(t)] = \chi(N(t) - n) - \chi(N(t) - n - 1),$$

где $\chi(t)$ — функция Хэвисайда. Тогда

$$\langle X^2(t)\rangle = \frac{\sigma^2}{h^2} \left\langle \sum_{n=1}^{\infty} \sum_{k=1}^{n} [1 - 2\exp(-hT_k) + \exp(-2hT_k)]\, \Pi_n[N(t)] \right\rangle .$$

Заметим, что для величин $N(t)$ и T_n неравенства $N(t) \geq n$ и $T_n < t$ эквивалентны, а значит

$$\langle X^2(t)\rangle = \frac{\sigma^2}{h^2} \left\langle \sum_{n=1}^{\infty} \sum_{k=1}^{n} \big[1 - 2\exp(-hT_k) + \exp(-2hT_k)\big] \big[\chi(t - T_n) - \chi(t - T_{n+1})\big] \right\rangle .$$

В дальнейшем удобнее будет рассматривать лаплас-образ второго момента

$$g(s) = \int_0^{\infty} \exp(-st)\, \langle X^2(t)\rangle \, \mathrm{d}t,$$

$$g(s) = \frac{\sigma^2}{sh^2} \sum_{n=1}^{\infty} \sum_{k=1}^{n} \Big\langle \big[1 - 2\exp(-hT_k) + \exp(-2hT_k)\big] \big[\exp(-sT_n) - \exp(-sT_{n+1})\big] \Big\rangle. \tag{3}$$

Вычислим входящие сюда средние с помощью плотности вероятностей $h_n(\tau)$ (2). С их помощью получим

$$\langle \exp(-sT_n) \rangle = \hat{f}^n(s), \qquad \langle \exp(-hT_k - sT_n) \rangle = \hat{f}^k(h+s) \hat{f}^{n-k}(s),$$

где $\hat{f}(s)$ — лаплас-образ распределения $f(\tau)$. Подставив эти выражения в (3) и просуммировав по k и n, запишем окончательно

$$g(s) = \frac{\sigma^2}{sh^2} \left\{ \frac{1}{1 - \hat{f}(s)} - \frac{2}{1 - \hat{f}(h+s)} + \frac{1}{1 - \hat{f}(2h+s)} \right\}. \tag{4}$$

Если средняя длительность интервала между моментами $\{t_n\}$ ограничена, то имеет место соотношение

$$\hat{f}(s) = 1 - s \langle \tau \rangle \qquad (s \to 0),$$

а асимптотика выражения (4) описывается формулой

$$g(s) \sim \frac{\sigma^2}{h^2 \langle \tau \rangle} \frac{1}{s^2}.$$

Соответствующая ей временна́я асимптотика среднеквадратичного отклонения

$$\langle X^2(t) \rangle \sim \frac{\sigma^2}{h^2 \langle \tau \rangle} t \qquad (t \to \infty)$$

растёт по закону линейной диффузии. Если же асимптотика соответствующей характеристической функции интервалов $\tau = t_m - t_{m-1}$ имеет вид [9]

$$\hat{f}(s) = 1 - \varepsilon s^{\beta} \qquad (s \to 0, \quad 0 < \beta < 1),$$

то для (4) имеем

$$g(s) \sim \frac{\sigma^2}{\varepsilon h^2} \frac{1}{s^{1+\beta}} \qquad (0 < \beta < 1).$$

Соответственно, в этом случае

$$\langle X^2(t) \rangle \sim \frac{\sigma^2}{\varepsilon h^2} \frac{1}{\Gamma(1+\beta)} t^{\beta} \qquad (t \to \infty), \tag{5}$$

что отвечает аномально диффузионному поведению системы.

ВЫВОДЫ

В правой части уравнения Ланжевена для обычного процесса Орнштейна—Уленбека стоит гауссов белый шум $\xi(t)$. В рассмотренном выше процессе вместо него присутствуют случайные блуждания (1), состоящие из совокупности скачков величины $\{\eta_k\}$, также имеющих гауссово распределение. Для полученного процесса выведена зависимость второго момента от коэффициента затухания, совпадающая с аналогичной зависимостью для процесса Орнштейна—Уленбека. Однако временна́я асимптотика среднеквадратичного отклонения уже не подчиняется закону линейной диффузии, хотя и переходит в него в пределе $\beta \to 1$ (5). Из всего вышеизложенного следует, что рассмотренный процесс является естественным обобщением процесса Орнштейна—Уленбека на случай нарушения центральной предельной теоремы.

СПИСОК ЛИТЕРАТУРЫ

1. Zaslavsky G. M., Edelman M., Niyazov B. // Chaos. 1997. V. 7, No. 1. P. 159.
2. Забурдаев В. Ю., Чукбар К. В. // ЖЭТФ. 2002. Т. 121, вып. 2. С. 299.
3. Klafter J., Shlesinger M. F., Zumofen G. // Phys. Today. 1996. V. 49, No. 2. P. 33.
4. Учайкин В. В. // ТМФ. 1998. Т. 115, № 1. С. 154.
5. Barkai E., Silbey R. // J. Phys. Chem. B. 2000. V. 104. P. 3 866.
6. Barkai E. // Phys. Rev. E. 2001. V. 63. P. 046118–1/17.
7. Metzler R., Klafter J. // Phys. Rep. 2000. V. 339. P. 1.
8. Саичев А. И., Уткин С. Г. // Актуальные проблемы статистической радиофизики (малаховский сборник). 2002. Т. 1, № 1. С. 5.
9. Saichev A. I., Woyczyński W. A. Distributions in the Physical and Engineering Sciences. V. I. Boston: Birkhäuser, 1997. 336 p.

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

Поступила в редакцию
10 апреля 2003 г.

ON THE GENERALIZED ORNSTEIN–ULENBECK PROCESS

A. I. Saichev and S. G. Utkin

Based on the anomalous-diffusion nature of random-walk processes, we construct a model generalized Ornstein–Ulenbeck process and find the asymptotics of its main statistical characteristics. It is shown that the asymptotics obey the classical diffusion law if the central limit theorem is valid and the anomalous diffusion law if this theorem is violated.