
ДОКЛАДЫ БГУИР

Выходит два раза в квартал

Научный журнал издается с января 2003 года

Главный редактор М.П. Батура

Редакционная коллегия:

А.П. Кузнецов (зам. главного редактора),
Л.М. Лыньков (зам. главного редактора),

Т.В. Борботько (ответственный секретарь),
В.Е. Борисенко, С.Е. Карпович, Н.Т. Квасов, В.К. Конопелько,
А.А. Кураев, В.А. Лабунов, В.В. Муравьев, Р.Х. Садыхов

Редакционный совет:

И.И. Абрамов, В.Е. Агабеков, В.В. Баранов, А.И. Белоус, И.В. Боднар, С.В. Бордусов, С.В. Гапоненко, В.В. Голенков, В.Ф. Голиков, А.Л. Гурский, Л.И. Гурский, А.П. Достанко, В.А. Емельянов, И.Е. Зуйков, В.М. Колешко, Ф.Ф. Комаров, Ф.П. Коршунов, С.П. Кундас, В.А. Куренев, В.И. Курмашев, Н.И. Листопад, С.В. Лукьянец, В.Е. Матюшков, Л.И. Минченко, Ф.И. Пантелеенко, А.А. Петровский, В.А. Пилипенко, С.Л. Прищепа, А.М. Русецкий, В.А. Сокол, Н.К. Толчко, А.А. Хмыль, В.В. Цегельник, В.А. Чердынцев, Г.П. Яблонский, В.Н. Ярмолик

АДРЕС РЕДАКЦИИ:

220013, Минск, ул. П. Бровки, 6, к. 326
293-88-41

doklady@bsuir.by

www.doklady.bsuir.by

СОДЕРЖАНИЕ

ЭЛЕКТРОНИКА, РАДИОФИЗИКА, РАДИОТЕХНИКА, ИНФОРМАТИКА

Гирель К.В., Бондаренко А.В. Формирование наноструктур серебра методом иммерсионного осаждения на пористый кремний и исследование их оптических свойств	5
Чан Туан Чунг, Боровик А.М., Стемпицкий В.Р. Оптимизация параметров диффузионно-дрейфовой модели	11
Зеленина М.С., Козлова О.А. Моделирование электронных и магнитных свойств низкоразмерных материалов с использованием первопринципных методов	18
Лещёв А.Е., Минченко Л.И. Условия оптимальности второго порядка	24
Белоусова Е.С., Лыньков Л.М., Сеньют В.Т., Криштопова Е.А. Влияние термообработки порошка шунгита в вакууме на экранирование электромагнитного излучения	31
Саванович С.Э., Борботько Т.В., Соколов В.Б., Русакович М.В. Влияние размеров фракций влажосодержащего керамзита на его коэффициенты отражения и передачи электромагнитного излучения	36
Цветков В.Ю. Геометрические модели многокурсных изображений и проективная компенсация движения камеры	41

Овсянников А.В., Козел В.М. Информационная прогнозируемость стохастических процессов в дискретном времени.....	48
Рыбак В.А., Сулайман Х.М. Применение и оптимизация нейронных сетей для поддержки принятия решений на валютном рынке.....	55
Кузьмич А.И. Оценка риска выполнения логистических проектов на начальной стадии	60
Кун А.А., Шабан С.А., Еромин А.М., Кочеров А.Л. Синтез квазиоптимального контура телеуправляемого беспилотного летательного аппарата при инерционном измерителе	65
Липницкий В.А., Олексюк А.О. Теория норм синдромов и плюс-декодирование.....	72
Рабченко Д.И., Кругликов С.В. Способ реализации информационной модели обстановки в автоматизированной системе управления.....	79
Хохлов Е.А., Мысливец А.С., Смирнов А.Г. Формирование непрозрачного диэлектрического декоративного покрытия для сенсорных дисплеев	86
<i>КРАТКИЕ СООБЩЕНИЯ</i>	
Зеневич А.О. Повышение живучести волоконно-оптической системы связи в чрезвычайных ситуациях	90
Басов В.Г. Оценка эффективности направленного излучения встречно-штыревого преобразователя на неоднородно поляризованном звукопроводе.....	94

Учредитель: учреждение образования
"Белорусский государственный университет информатики и радиоэлектроники"

Редактор Т. В. МИРОНЕНКО
Компьютерный дизайн и верстка О. В. БОЙПРАВ

Подписано в печать 01.12.2014. Формат 60×84 1/8. Гарнитура "Таймс". Бумага офсетная.
Отпечатано на ризографе. Усл. печ. л. 11,86. Уч.-изд. л. 9,9. Тираж 100 экз. Заказ 466.

Индекс для индивидуальной подписки 00787.
Индекс для ведомственной подписки 007872.

Издатель: учреждение образования «Белорусский государственный университет информатики и радиоэлектроники».
Свидетельство о государственной регистрации средства массовой информации № 1087 от 23.01.2010.

Отпечатано в БГУИР. ЛП № 02330/264 от 14.04.2014. 220013, г. Минск, ул. П. Бровки, 6.

© УО «Белорусский государственный университет информатики и радиоэлектроники», 2014

DOKLADY BGUIR

Published twice quarterly

The journal has been published since January, 2003

Editor-In-Chief M.P. Batura

ADDRESS OF EDITORIAL OFFICE

220013, Minsk, P. Brovka Str., 6, Room 326

293-88-41

doklady@bsuir.by

www.doklady.bsuir.by

CONTENTS

ELECTRONICS, RADIOPHYSICS, RADIOENGINEERING, INFORMATICS

Girel K.V., Bandarenka H.V. Formation of silver nanostructures by immersion deposition method onto porous silicon and study of their optical properties.....	5
Trung Tran Tuan, Borovik A.M., Stempitsky V.R. Drift-diffusion model parameters optimization....	11
Zelenina M.S., Kozlova O.A. Simulation of electronic and magnetic properties of low-dimensional materials with use of first principles methods.....	18
Leschov A.E., Minchenko L.I. Second order optimality conditions.....	24
Belousova E.S., Lynkou L.M., Senyut' V.T., Krishtopova E.A. Influence of heat treatment in vacuum on shungite shielding properties.....	31
Savanovich S.E., Borbotko T.V., Sokolov V.B., Rusakovich M.V. Influence of water-containing claydite fraction sizes on the electromagnetic radiation reflection and transmission.....	36
Tsviatkou V.Yu. Geometric models of multi-angle images and projective compensation of camera motion.....	41
Ausiannikau A.V., Kozel V.M. Information predictability of stochastic processes in discrete time.....	48
Rybak V.A., Sulaiman H.M. Application and optimization of neural networks for decision support on exchange market.....	55
Kuzmich A.I. Risk assessment of logistics project implementation in the initial stage.....	60
Kun A.A., Shaban S.A., Eromin A.M., Kocherov A.L. The synthesis of quasioptimal contour of the long-distance drone's command and control with the inertia measuring device.....	65
Lipnitski V.A., Aliaksiuk A.O. Theory of normal syndrome and plus-decoding.....	72
Rabchenok D.I., Kruglikov S.V. Approach to realization of information model of the situation in management information system of the tactical link.....	79

Khokhlov E.A., Mislivets A.S., Smirnov A.G. Formation of a non transparent dielectric decorative coating for sensor displays	86
---	----

SHORT NOTES

Zenevich A.O. Improvement of survivability of fiber optic communication system used for data transmission in emergencies.....	90
--	----

Basov V.G. Evaluating the effectiveness of directional radiation interdigital transducer on inhomogeneously polarized acoustic line.....	94
---	----

Библиотека БГУИР

УДК 535.375.5; 543.424.2; 544.653.23

ФОРМИРОВАНИЕ НАНОСТРУКТУР СЕРЕБРА МЕТОДОМ ИММЕРСИОННОГО ОСАЖДЕНИЯ НА ПОРИСТЫЙ КРЕМНИЙ И ИССЛЕДОВАНИЕ ИХ ОПТИЧЕСКИХ СВОЙСТВ

К.В. ГИРЕЛЬ, А.В. БОНДАРЕНКО

*Белорусский государственный университет информатики и радиоэлектроники
П. Бровка, 6, Минск, 220013, Беларусь*

Поступила в редакцию 17 июля 2014

Представлены результаты исследований наноструктур, изготовленных при различных режимах иммерсионного осаждения серебра на пористый кремний (ПК), для их применения в качестве активных субстратов в спектроскопии гигантского комбинационного рассеяния (ГКР). ПК был сформирован методом анодирования монокристаллического кремния в водно-спиртовом растворе фтористоводородной кислоты. Исследованы спектры отражения полученных наноструктур серебра на ПК. Выявлено, что для создания оптимальных условий ГКР-спектроскопии с использованием наноструктур серебра на ПК необходимо использовать возбуждающий лазер с длиной волны 400–450 нм.

Ключевые слова: наноструктуры, серебро, иммерсионное осаждение, пористый кремний.

Введение

Наноструктуры серебра на ПК обладают уникальными оптическими свойствами, связанными с сильными локальными электромагнитными полями, возникающими благодаря возбуждению плазмонов на поверхности серебра [1, 2]. При этом на определенных частотах падающего излучения наблюдается эффект локализованного поверхностного плазмонного резонанса (ЛППР). Зная область проявления ЛППР можно выбрать оптимальные режимы снятия спектров ГКР, при которых интенсивность ГКР-сигнала будет максимальной. Спектроскопия ГКР используется для детектирования, идентификации и изучения структуры следовых количеств веществ в материаловедении [3, 4]. Кроме того, в области применения ГКР входят биомедицина, экология, пищевая промышленность, криминалистика и многие другие [5–11]. В настоящей работе проведено исследование закономерностей формирования наноструктур серебра на ПК и их спектров отражения.

Методика эксперимента

Нитрат серебра AgNO_3 (99,9999 %), 45 % водный раствор фтористоводородной кислоты HF (45 %), изопропиловый ($\text{C}_3\text{H}_7\text{OH}$) и этиловый ($\text{C}_2\text{H}_5\text{OH}$) спирты применялись без дополнительной очистки. Для приготовления растворов использовалась дистиллированная вода.

В качестве исходных подложек были использованы пластины монокристаллического кремния диаметром 100 мм, легированного сурьмой, с удельным сопротивлением ($\rho_{\text{уд}}$) 0,01 Ом·см и кристаллографической ориентацией поверхности (100). Поверхность кремниевых пластин была предварительно очищена от органических загрязнений в горячем (75 °С) перекисно-аммиачном растворе (ПАР) и от естественного слоя оксида в 4,5% водном растворе фтористоводородной кислоты. После чего пластины были высушены методом центрифугирования. Формирование ПК осуществлялось методом анодирования

монокристаллического кремния в водно-спиртовых растворах на основе фтористоводородной кислоты. Для формирования слоев ПК был использован электролит, состоявший из HF (45 %), H₂O и C₃H₇OH, смешанных в объемном соотношении 1:3:1. Процесс анодирования проводился при плотности тока 100 мА/см² в течение 85 с. Такие режимы позволили получить слои ПК толщиной 5 мкм, пористостью 72 % и средним диаметром пор около 100 нм.

Для получения наноструктур серебра на ПК образцы ПК помещались в водный раствор AgNO₃ с добавлением этилового спирта. Концентрация AgNO₃ варьировалась в пределах 1 – 10 мМ, время осаждения 5 – 180 мин, температура раствора составляла 20 или 40 °С. После образования на поверхности ПК пленки серебра образцы тщательно промывались в этиловом спирте, а затем высушивались. Процесс анодирования проводили при помощи потенцио/гальваностата AUTOLAB PGSTAT302N.

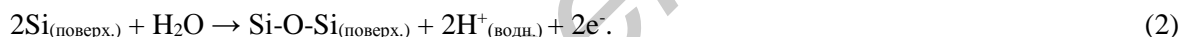
Морфологические и структурные параметры наноструктур серебра на ПК были изучены методом сканирующей электронной микроскопии на оборудовании Hitachi S4800. Измерения спектров отражения проводились на спектрофотометре MC 122 в диапазоне от 200 до 1100 нм.

Результаты и их обсуждение

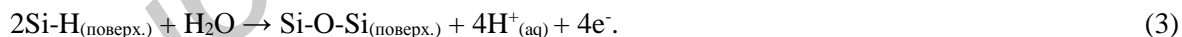
Механизм иммерсионного осаждения. Известно, что восстановление катионов серебра до атомарной формы происходит за счет присоединения электронов:



Источниками электронов восстановления при использовании подложек на основе кремния являются непосредственно атомы кремния. Благодаря положительному редокс-потенциалу серебра, катионы этого металла способны окислять поверхность кремния, забирая у него электроны. Поэтому при погружении кремния в водные растворы солей серебра происходит одновременное восстановление атомов металла (1) и образование под ними диоксида кремния:



Очевидно, длительная выдержка монокристаллического кремния в водных растворах солей серебра обуславливает формирование сплошного слоя оксида кремния, который препятствует контакту реагентов из раствора и атомов кремния, что приводит к прекращению восстановления металла. Это ограничивает количество и равномерность распределения осадка на подложке даже при высокой концентрации катионов серебра в исходном растворе. В случае применения ПК значительный вклад в процесс восстановления атомов серебра вносит присутствие на его развитой поверхности Si-H_x групп, которые возникают в результате гидратирования связей атомов кремния, оборванных в процессе электрохимического вытравливания пор. Связи Si-H_x характеризуются повышенной реакционной активностью и легко окисляются, также поставляя электроны для восстановления серебра:



Таким образом, ПК играет роль не только формообразующей подложки, придающей поверхности серебряного осадка шероховатостей наноразмерного диапазона, но и является источником гораздо большего числа центров зарождения и электронов для восстановления атомов серебра по сравнению с монокристаллическим кремнием.

Рентгеноструктурный анализ, проведенный для наноструктур серебра на ПК, изготовленных методом иммерсионного осаждения серебра на ПК, показал, что рефлексы, наблюдаемые на дифрактограммах, характерны для кристаллов серебра с ориентацией (111), (200), (220), (311) (рис. 1). То есть в процессе иммерсии на ПК сформировался поликристаллический серебряный осадок. Также наблюдается сильный рефлекс от подложки монокристаллического кремния.

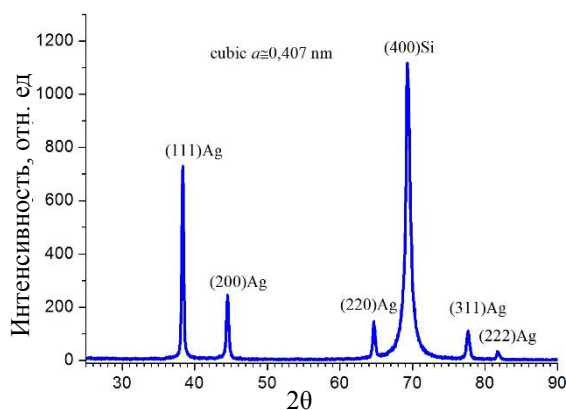


Рис. 1. Рентгеноструктурный спектр наноструктуры серебра на ПК

Закономерности формирования наноструктур серебра на ПК. На рис. 2 представлены СЭМ-фотографии поверхности образцов ПК, выдержанных в водном растворе 1 мМ AgNO_3 и 1 М $\text{C}_2\text{H}_5\text{OH}$ в течение (а) 15, (б) 120 и (в) 180 мин при температуре 20 °С. Видно, что для начальной стадии осаждения серебра характерно формирование на поверхности ПК отдельных частиц металла преимущественно сферической формы, зарождение которых в основном происходит на краях пор, что подтверждает описанный ранее факт повышенной реакционной активности областей нарушенной структуры кремния (рис. 2, а).

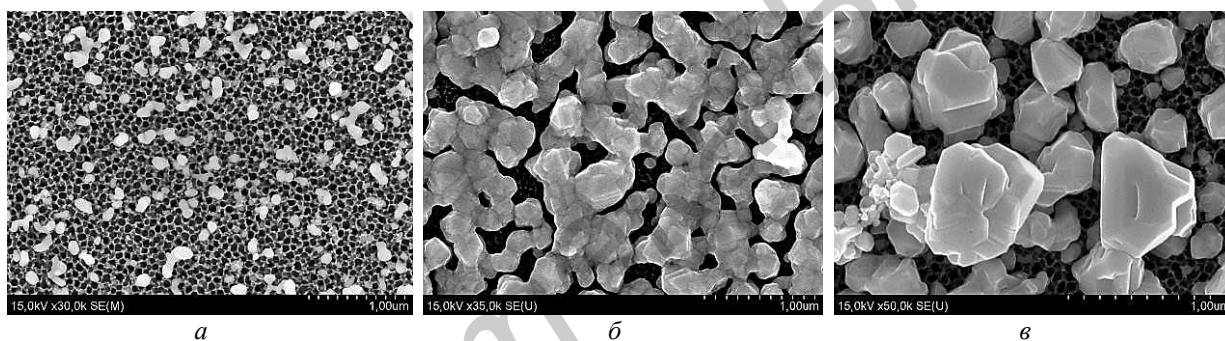


Рис. 2. СЭМ-фотографии поверхности образцов ПК, выдержанных в водном растворе 1 мМ AgNO_3 и 1 М $\text{C}_2\text{H}_5\text{OH}$ в течение 15 (а), 120 (б) и 180 (в) мин при температуре 20 °С

Диаметр частиц серебра колеблется от единиц до нескольких десятков нанометров. Расстояние между частицами составляет от 150 до 300 нм. Некоторые из них соединились между собой в цепочки, которые при дальнейшей выдержке образца в растворе до 120 мин превратились в агломераты неправильной формы, частично слившиеся между собой. Визуально, структура, представленная на рис. 2, б напоминает пористую серебряную пленку, в которой практически не наблюдается отдельно расположенных частиц серебра. Расстояние между элементами агломератов металла, которые не контактируют между собой, колеблется от 30 до 100 нм, что на порядок ниже, чем в случае 15 мин осаждения. Более длительный иммерсионный процесс обусловил срастание агломератов серебра в крупные кристаллы с явно выраженной огранкой (рис. 2, в).

Наблюдаемое явление позволяет сделать вывод о том, что иммерсионное осаждение серебра на ПК протекает в соответствии с известным механизмом Вольмера-Вебера, согласно которому формирование тонких пленок происходит в результате роста островков вещества, силы межатомного взаимодействия внутри которых выше, чем с атомами материала подложки.

При увеличении концентрации AgNO_3 до 3 мМ наблюдается аналогичная закономерность в формировании пленки серебра на поверхности ПК: переход от стадии отдельных частиц (рис. 3, а) к стадии почти сплошной пленки (рис. 3, б), а затем появление крупных частиц серебра и формирование вторичных островков (рис. 3, в). Это подтверждает сделанный ранее вывод о механизме осаждения.

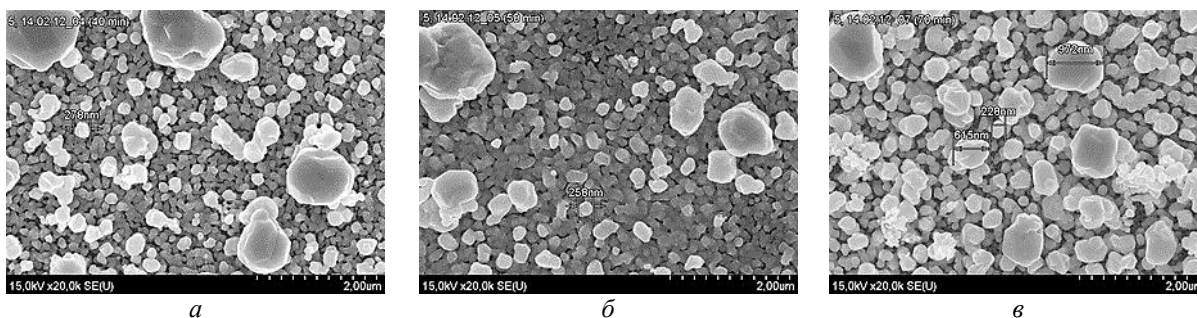


Рис. 3. СЭМ-фотографии поверхности образцов ПК, выдержанных в водном растворе 3 мМ AgNO_3 и 1 М $\text{C}_2\text{H}_5\text{OH}$ в течение 40 (а), 50 (б) и 70 (в) мин при температуре 20 °С

Также на рис. 4 приведены СЭМ-фотографии образцов наноструктур серебра на ПК, приготовленных в водном растворе 1 мМ AgNO_3 и 1 М $\text{C}_2\text{H}_5\text{OH}$ в течении 15 мин при температуре 20 °С или 40 °С. Как видно из рис. 4 механизм осаждения серебра такой же, как и при 20 °С, однако скорость реакции восстановления серебра увеличивается, что приводит к более быстрому росту частиц серебра на поверхности ПК. Таким образом, за то же время осаждения можно получить большее количество крупных частиц серебра.

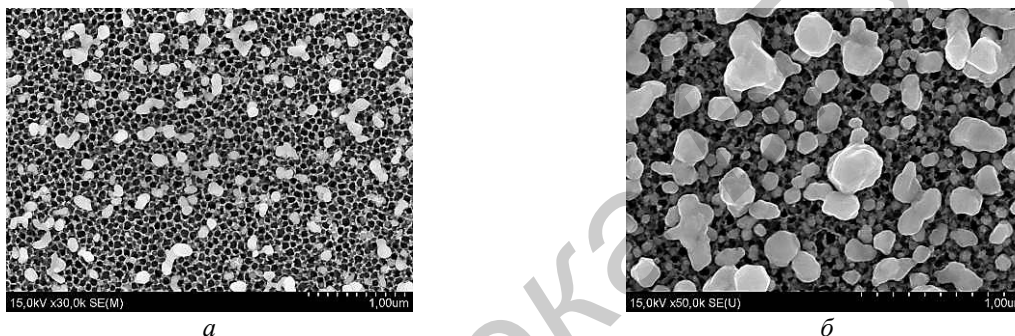


Рис. 4. СЭМ-фотографии поверхности образцов ПК, выдержанных в водном растворе 1 мМ AgNO_3 и 1 М $\text{C}_2\text{H}_5\text{OH}$ в течение 15 мин при температуре 20 °С (а), 40 °С (б)

Осаждение серебра при более высоких концентрациях AgNO_3 и увеличенной температуре раствора протекает по такому же механизму.

Исследование спектров отражения. Известно, что значительное увеличение ГКР-сигнала имеет место благодаря сильным локальным электромагнитным полям, которые возникают вблизи металлических наноструктур при возбуждении локализованных поверхностных плазмонов. При этом на определенных частотах наблюдается ЛППР, приводящий к существенному (резонансному) усилению поглощения/рассеяния падающего излучения.

Поскольку подложки ПК непрозрачны в видимой области, то для того, чтобы судить о положении ЛППР были получены спектры зеркального отражения наноструктур серебра на ПК. На рис. 5 показаны спектры отражения образцов, сформированных в течение различного времени осаждения серебра при концентрации 1 мМ AgNO_3 . В длинноволновой области спектров наноструктур серебра на ПК отчетливо наблюдаются полосы поглощения, обусловленные интерференцией света на ПК. Кроме того, в каждом из спектров отражения присутствуют две характерные полосы поглощения/рассеяния в интервалах 310–330 нм и 400–550 нм, связанные с поглощением/рассеянием излучения, вызванным ЛППР в наночастицах серебра. При более длительном времени осаждения серебра увеличивается диапазон разброса частиц металла по размерам, а также увеличивается их средний диаметр, что приводит к уширению полос поглощения на всех спектрах отражения и сдвиг их минимума в длинноволновую область.

При увеличении концентрации AgNO_3 до 3 и 10 мМ полосы поглощения/рассеяния находятся в тех же областях, что и при концентрации 1 мМ. Так же наблюдается уширение полос при более длительном осаждении и сдвиг их минимума в длинноволновую область (рис. 6, а, б).

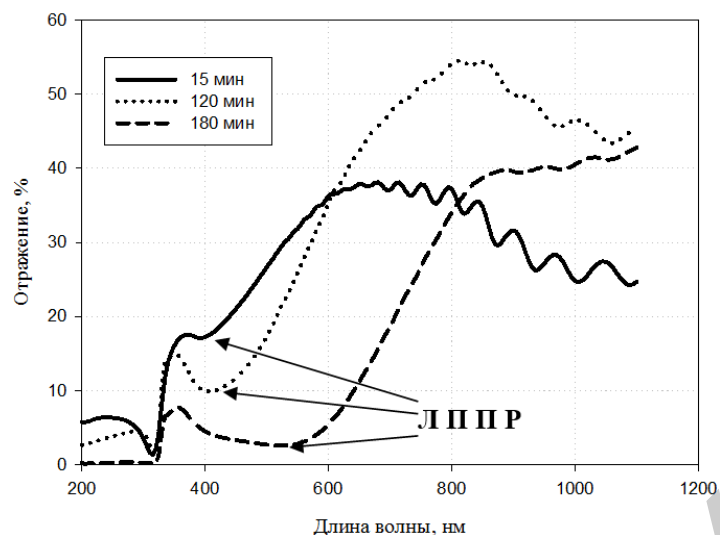
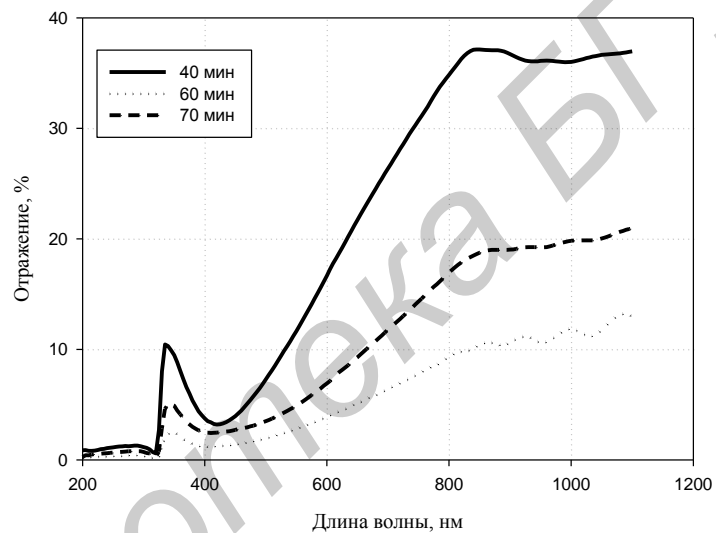
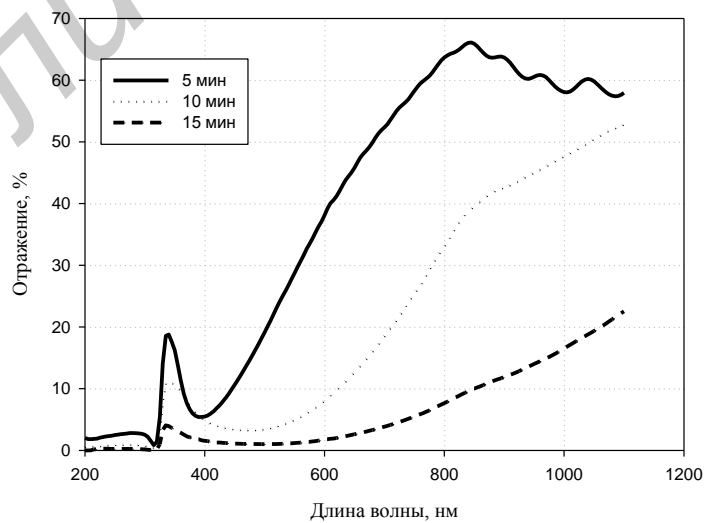


Рис. 5. Спектры отражения наноструктур серебра на ПК, сформированных путем осаждения частиц Ag на ПК из водного раствора 1 mM AgNO_3 в течение различных периодов времени



a



б

Рис. 6. Спектры отражения наноструктур серебра на ПК, сформированных путем осаждения частиц Ag на ПК из водного раствора 3 mM (а) и 10 mM (б) AgNO_3 в течение различных периодов времени

Оптимальные условия для проведения ГКР-спектроскопии от молекул, адсорбированных на поверхности наноструктур серебра на ПК, будут созданы при использовании возбуждающего излучения с длиной волны, попадающей в область ЛППР этих структур. Поэтому для обеспечения максимальной чувствительности ГКР-спектроскопии с применением полученных в данной работе наноструктур серебра на ПК целесообразно использовать лазер с длиной волны 400 – 450 нм.

Заключение

Путем иммерсионного осаждения серебра на ПК из водного раствора азотнокислого серебра и этилового спирта можно формировать наноструктуры серебра на ПК, демонстрирующие ЛППР в диапазоне длин волн возбуждающего излучения от 400 до 450 нм. Полагается, что полученные структуры могут быть использованы для проведения количественного и качественного анализа жидкостей методом ГКР-спектроскопии. Причем оптимальные условия ГКР-спектроскопии должно обеспечить использование возбуждающего лазера с длиной волны 400–450 нм.

Авторы выражают благодарность Бондаренко В.П. и Терехову С.Н. за плодотворные дискуссии при выполнении работы. Работа выполнялась в рамках задания 2.3.06 ГПНИ РБ «Электроника и фотоника».

FORMATION OF SILVER NANOSTRUCTURES BY IMMERSION DEPOSITION METHOD ONTO POROUS SILICON AND STUDY OF THEIR OPTICAL PROPERTIES

K.V. GIREL, H.V. BANDARENKA

Abstract

Nanostructures formed by immersion deposition of silver on porous silicon at different regimes for their application in surface enhanced Raman spectroscopy (SERS) has been presented. Porous silicon has been fabricated by anodization of monocrystalline silicon in water solution of isopropanol and hydrofluoric acid. Reflectance spectra of the obtained structures of silver/porous silicon have been studied. It has been revealed that the optimal regimes of SERS require using laser with wave length of 400–450 nm.

Список литературы

1. Natan J.M. // Faraday Discussions. 2006. № 132. P. 321–328.
2. Shalaev V.M. // Phys. Rep.: Rev. Sect. Phys. Lett. 1996. № 272. P. 61–137.
3. Lefrant S., Baltog I., Baibarac B. // Raman Spectrosc. 2005. Vol. 36. P. 676–698.
4. Li L., Zhou D.S., Xue G.J. // Raman Spectrosc. 2005. Vol. 36. P. 699–703.
5. Pinzaru S.C. // J. Raman Spectrosc. 2004. Vol. 35. P. 338–346.
6. Reyes-Goddard J.M., Barr H., Stone N. // Photodiag. and Photodyn. Therapy. 2005. Vol. 2. P. 223–233.
7. Farquharson S. // Molecules. 2008. Vol. 13. P. 2608–2627.
8. Sha M.Y. // J. Am. Chem. Soc. 2008. Vol. 130. P. 17214–17215.
9. Vo-Dinh T., Allain L.R., Stokes D.L. // J. Raman Spectrosc. 2002. Vol. 33. P. 511–516.
10. Yakes B.J. // Clin. Vaccine Immunol. 2008. Vol. 15. P. 227–234.
11. McLaughlin C. // Anal. Chem. 2002. Vol. 74. P. 3160–3167.

УДК 621.382

ОПТИМИЗАЦИЯ ПАРАМЕТРОВ ДИФFUЗИОННО-ДРЕЙФОВОЙ МОДЕЛИ

ЧАН ТУАН ЧУНГ, А.М. БОРОВИК, В.Р. СТЕМПИЦКИЙ

Белорусский государственный университет информатики и радиоэлектроники
П. Бровка, 6, Минск, 220013, Беларусь

Поступила в редакцию 21 июля 2014

Предложен новый подход расчета электрических характеристик наноразмерных МОП-транзисторов, сущность которого заключается в использовании корректирующих коэффициентов, а также таких значений параметров классических диффузионно-дрейфовых моделей, которые бы эффективно учитывали квантово-механические механизмы транспорта. Разработан и использован при оптимизации диффузионно-дрейфовой модели модифицированный метод прямого поиска для МОП-транзистора с длиной канала 90 нм.

Ключевые слова: наноразмерный МОП-транзистор, компьютерное моделирование, модель подвижности Дарвиша, оптимизация, метод прямого поиска.

Введение

Физико-математические модели полупроводниковых приборов описываются системой фундаментальных уравнений, которые связывают электростатический потенциал и плотность носителей заряда. Эта система уравнений выводится из законов Максвелла и состоит из уравнения Пуассона, уравнений непрерывности, транспортных уравнений и уравнений энергетического баланса. Для точного учета квантовых эффектов эти уравнения необходимо дополнить уравнением Шредингера [1–3]. Стандартной моделью транспорта носителей заряда является диффузионно-дрейфовая модель, которая дает физически адекватные результаты для всех приборов, изготовленных по микронным и субмикронным технологическим нормам. Однако уменьшение размеров элементов сверхбольших интегральных схем (СБИС) приводит к появлению новых физических эффектов, в том числе и квантовых, которые следует учитывать в процессе разработки моделей [3].

Особенности моделирования наноразмерных приборов

Рассмотрим некоторые физические эффекты, проявляющиеся в наноразмерных приборах, которые не учитываются в рамках стандартной диффузионно-дрейфовой модели, используемой в большинстве программных средств приборно-технологического моделирования.

Уменьшение подвижности вследствие увеличения вертикальной компоненты электрического поля.

Малая толщина подзатворного диэлектрика наноразмерных МОП-транзисторов приводит к увеличению напряженности вертикальной компоненты электрического поля под затвором до величины, когда происходит уменьшение подвижности вследствие рассеяния носителей на шероховатостях поверхности, на фонах и на неподвижных зарядах вблизи поверхности [3]. Зависимость эффективной подвижности от эффективного электрического поля может быть разделена на области влияния различных механизмов рассеяния носителей. При малых значениях напряженности (соответствующей пороговому напряжению) наблюдается наличие зависимости подвижности носителей заряда от концентрации примеси, которая увеличивается при уменьшении размеров приборов, что указывает на преобладание в этих условиях механизма рассеяния носителей на кулоновских центрах. При высоких поперечных

электрических полях рассеяние на шероховатостях поверхности имеет преобладающее влияние на подвижность в инверсном слое [4].

Модель подвижности Ломбарди [5], описывающая поведение носителей в инверсионных слоях и учитывающая снижение подвижности носителей заряда вследствие высокой степени их рассеяния на поверхности вблизи границы раздела полупроводник-диэлектрик, успешно используется при моделировании многих приборов микроэлектроники. Дарвишем была предложена улучшенная версия этой модели, в которой введены две модификации. Первая касается описания подвижности в объеме с использованием модели Клаасена, в которой учитываются эффекты кулоновского экранирования. Вторая модификация связана с использованием нового выражения для описания эффекта шероховатости поверхности [5, 6].

Вклад, обусловленный шероховатостью поверхности, определяется в модели Дарвиша следующими выражениями:

$$\frac{1}{\mu_{sr}} = \frac{\left(\frac{E_{\perp}}{E_2}\right)^{\gamma}}{\delta}, \quad \gamma = A + \frac{\alpha(n+p)}{\left(\frac{N}{N_2}\right)^{\eta}}, \quad (1)$$

где n и p – концентрации электронов и дырок соответственно, N – полная концентрация примесей ($N_D + N_A$), $N_2 = 1 \text{ см}^{-3}$, $E_2 = 1 \text{ В/см}$.

Вклад, обусловленный рассеянием на акустических фононах, определяется следующими соотношениями:

$$\mu_{ac} = \left(\frac{B}{\left(\frac{E_{\perp}}{E_2}\right)^E} + \frac{C \left(\frac{N}{N_2}\right)^{\tau}}{\left(\frac{E_{\perp}}{E_2}\right)^D} \right) \frac{1}{T}, \quad T = (T_L / 300)^k. \quad (2)$$

Данная модель успешно используется при моделировании субмикронных приборов, однако должна быть адаптирована для моделирования наноразмерных приборов.

Уменьшение продольной подвижности.

При малой напряженности поля дрейфовая скорость носителей линейно зависит от напряженности электрического поля. Однако, начиная с некоторой напряженности, называемой напряженностью насыщения, скорость носителей становится постоянной и, соответственно, происходит снижение подвижности носителей. Этот эффект является существенным именно для короткоканальных транзисторов, поскольку уменьшение длины канала происходит быстрее, чем рабочих напряжений транзистора. Эффект насыщения дрейфовой скорости носителей приводит к тому, что насыщение тока стока у короткоканальных транзисторов происходит раньше, чем у длинноканальных [3].

Описана методика независимого определения зависимости подвижности как от продольного, так и от поперечного поля для субмикронных транзисторов [4]. Методика основана на экстракции параметров β и v_{sat} в выражении для зависимости скорости носителей от напряженности продольного E_{\parallel} и поперечного эффективного E_{eff} полей:

$$\mu(E_{\parallel}, E_{\perp}) = \frac{\mu_0(E_{\perp})E_{\parallel}}{\left[1 + \left(\frac{\mu_0(E_{\perp})E_{\parallel}}{v_{sat}} \right)^{\beta} \right]^{1/\beta}}, \quad (3)$$

где μ_0 – низкополевая подвижность электронов при значении напряженности эффективного поля E_{\perp} ; v_{sat} – скорость насыщения;

$$v_{sat} = \frac{\alpha}{1 + \theta \exp\left(\frac{T_L}{T_{nom}}\right)}. \quad (4)$$

Эффект обеднения в поликремниевом затворе.

В субмикронных транзисторах толщина подзатворного диэлектрика настолько мала, что становится сравнима с толщиной области обеднения в поликремниевом затворе на границе с окислом. Слой обеднения увеличивает эффективное расстояние между затвором и полупроводником, что приводит к увеличению порогового напряжения. Необходимо также учитывать квантовомеханическую природу инверсионного слоя и вводить соответствующие поправки в классическую толщину области обеднения [3].

Эффект квантования инверсионного слоя.

Для эффективного управления каналом МОП-транзистора толщина окисла наноразмерных структур должна составлять несколько нанометров, а концентрация примеси в подложке должна быть увеличена. Но это уменьшает толщину области пространственного заряда и одновременно увеличивает напряженность электрического поля у поверхности. Поэтому потенциальная яма для электронов становится настолько узкой, что пренебречь квантовомеханической природой электронного газа уже нельзя [3]. Инверсионный слой действует как квантовый колодец для электронов. Потенциальные стенки этого колодца с одной стороны образованы границей диэлектрик/полупроводник, а с другой стороны – электрическим полем, прижимающим электроны к этой границе. Ввиду малой ширины колодца в нем формируется двумерный электронный газ. В результате разрешенные энергетические состояния электрона в яме также оказываются дискретными. Следствием этого является ненулевая минимальная энергия электрона в потенциальной яме. Конечное ненулевое минимальное значение энергии электронов и дискретность разрешенных энергетических состояний в низкоразмерной структуре, возникающие как следствие квантово-волнового поведения электрона в замкнутом пространстве, составляют суть эффекта квантового ограничения [4, 7].

Плотность состояний представляет собой параметр, определяющий количество энергетических уровней в определенном интервале энергий. Данная величина имеет важный физический смысл, так как она определяет концентрацию электронов в конкретной области любого материала или прибора, а также интенсивность рассеяния электронов в этой области. Выражение для плотности состояний при отсутствии квантового ограничения имеет вид:

$$n_{3D}(E) = \frac{\sqrt{2m^3}}{\hbar^3 \pi^2} \sqrt{E}. \quad (5)$$

Выражение для плотности состояний в случае квантового ограничения в одном направлении имеет вид:

$$n_{2D}(E) = \frac{m}{\pi \hbar^2 L_z} \sum_n \theta(E - E_n), \quad \theta(E - E_n) = \begin{cases} 1, E \geq E_n, \\ 0, E < E_n. \end{cases} \quad (6)$$

Плотность электронных состояний в квантовой пленке в зависимости от энергии представляет собой ступенчатую функцию, которая заменяет типичную параболическую зависимость для свободных электронов в трехмерных структурах [7, 8]. Следствием квантового ограничения является уменьшение концентрации электронов у поверхности и рост порогового напряжения. Соответственно, уменьшается также величина удельной емкости подзатворного диэлектрика и увеличивается его эффективная толщина. На рис. 1 представлено распределение электронов в канале при использовании классического и квантовомеханического приближения для структуры с толщиной подзатворного диэлектрика 4 нм, концентрацией акцепторной примеси в области канала $7 \cdot 10^{17} \text{ см}^{-3}$ при напряжении на затворе 2,5 В.

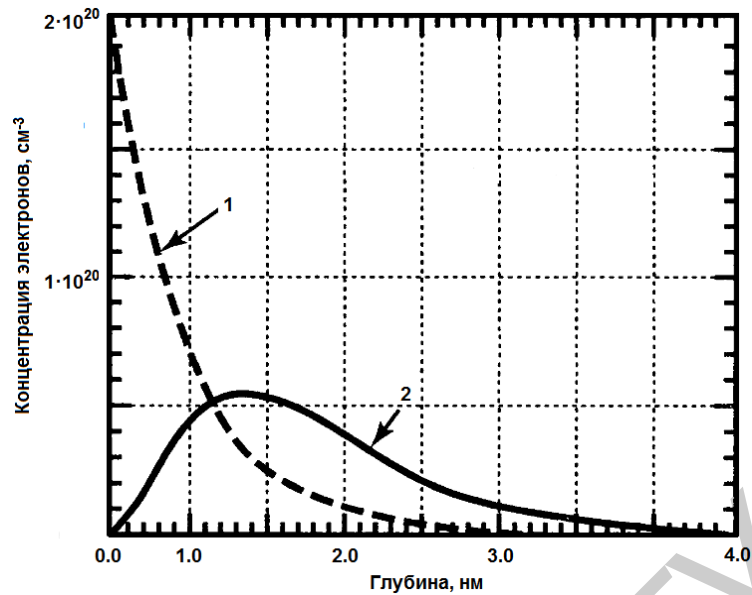


Рис. 1. Распределение электронов в канале: 1 – классическое приближение, 2 – квантовомеханическое приближение [3]

Метод прямого поиска

Оптимизация как эффективный алгоритм, позволяющий получить область экстремума целевой функции с заданной точностью, является основной областью использования методов и результатов проведения статистического анализа данных. Особую актуальность имеют методы и алгоритмы, позволяющие получать оценки вектора управляемых переменных, которому соответствует минимальное значение функции $f(x)$. Задача нахождения минимума функции $f(x)$ решается посредством процедуры систематического получения последовательности точек x_0, x_1, \dots, x_k таких, что $f(x_0) > f(x_1) > \dots > f(x_k) > \dots$. После выбора начальной точки x_0 необходимо выбрать направление, вдоль которого предполагается расположить следующую точку, и величину шага. Механизм образования последовательности точек и его эффективность в локализации точки минимума в сильной мере зависят от минимизируемой функции, а также информации, которая может быть использована для определения следующей точки [9].

Методы, ориентированные на решение задач безусловной оптимизации, можно разделить на три широких класса в соответствии с типом используемой при реализации того или иного метода информации, а именно:

- методы прямого поиска, основанные на вычислении только значений целевой функции $f(x)$;
- градиентные методы, в которых используются значения первых производных $f'(x)$;
- методы второго порядка, в которых наряду с первыми производными используются также вторые производные функции $f''(x)$.

Для того, чтобы избежать значительных временных затрат на вычисление частных производных при решении задач с числом независимых переменных пять и более предпочтительно использовать прямые методы оптимизации (нулевого порядка). К числу общих особенностей этих методов следует отнести относительную простоту соответствующих вычислительных процедур, которые легко реализуются и быстро корректируются.

Идея используемой методики заключается в выборе базовой точки и оценивании значений целевой функции в точках, окружающих ее. При решении задачи с двумя переменными можно воспользоваться квадратным образцом, представляющим собой базовую точку в центре и пробные точки в вершинах. Затем «наилучшая» из пяти исследуемых точек выбирается в качестве следующей базовой точки. В задачах большой размерности вычисление значений целевой функции проводится во всех вершинах, а также в центре тяжести гиперкуба. Если количество переменных (размерность пространства, в котором ведется поиск) равно N , то поиск по кубическому образцу требует $2^N + 1$ вычислений значения функций для одного

образца. При увеличении размерности задачи необходимое количество вычислений значения целевой функции чрезвычайно быстро возрастает [10].

Адаптация диффузионно-дрейфовой модели для наноразмерных приборов

Решение выявленных проблем и противоречий заключается в оптимизации диффузионно-дрейфовой модели с целью достижения адекватности моделирования наноразмерных МОП-структур, используя экспериментальные данные или результаты более точного моделирования. С этой целью будем варьировать значения плотности состояний в области канала. Также введем корректирующие коэффициенты в выражения моделей подвижности в поперечном и продольном электрическом поле, основываясь на анализе уравнений модели и результатах проведенных отсеивающих экспериментов [11]. В качестве модели подвижности в поперечном электрическом поле рассматривается модель Дарвиша.

Для демонстрации возможностей предложенного подхода адаптации диффузионно-дрейфовой модели использовалась n -канальная МОП-структура с длиной канала 90 нм, результаты экспериментального измерения вольтамперных характеристик (ВАХ) которой получены Microsystems Technology Laboratory [12]. После введения корректирующих коэффициентов выражение для описания вклада, обусловленного рассеянием на акустических фонах, приняло вид:

$$\mu_{ac,n} = \left(\frac{B_n}{\left(\frac{E_{\perp}}{E_2}\right)^{0,97E_n}} + \frac{C_n \left(\frac{N}{N_2}\right)^{\tau_n}}{\left(\frac{E_{\perp}}{E_2}\right)^{0,97D_n}} \right) \frac{1}{T_n}. \quad (7)$$

Выражение для вычисления коэффициента γ :

$$\gamma_n = 1,0455A_n + \frac{\alpha_n(n+p)}{\left(\frac{N}{N_2}\right)^{\eta_n}}. \quad (8)$$

Выражения для скорости насыщения и подвижности в продольном электрическом поле соответственно имеют вид:

$$v_{sat} = \frac{0,97\alpha}{1 + \theta \exp\left(\frac{T_L}{T_{nom}}\right)}, \quad (9)$$

$$\mu(E_{\parallel}, E_{\perp}) = \frac{\mu_0(E_{\perp})E_{\parallel}}{\left[1 + \left(\frac{\mu_0(E_{\perp})E_{\parallel}}{v_{sat}} \right)^{0,56\beta} \right]^{1/0,56\beta}}. \quad (10)$$

Новое значение эффективной плотности состояний для электронов равно $2.5e18 \text{ см}^{-3}$.

На рис. 2 и 3 представлены ВАХ МОП-транзистора, полученные экспериментально и в результате моделирования с использованием диффузионно-дрейфовой модели. При этом использовались значения плотности состояний и выражения модели подвижности как исходные, так и оптимизированные.

Очевидно, что результаты, полученные с использованием оптимизированных уравнений модели, соответствуют экспериментальным.

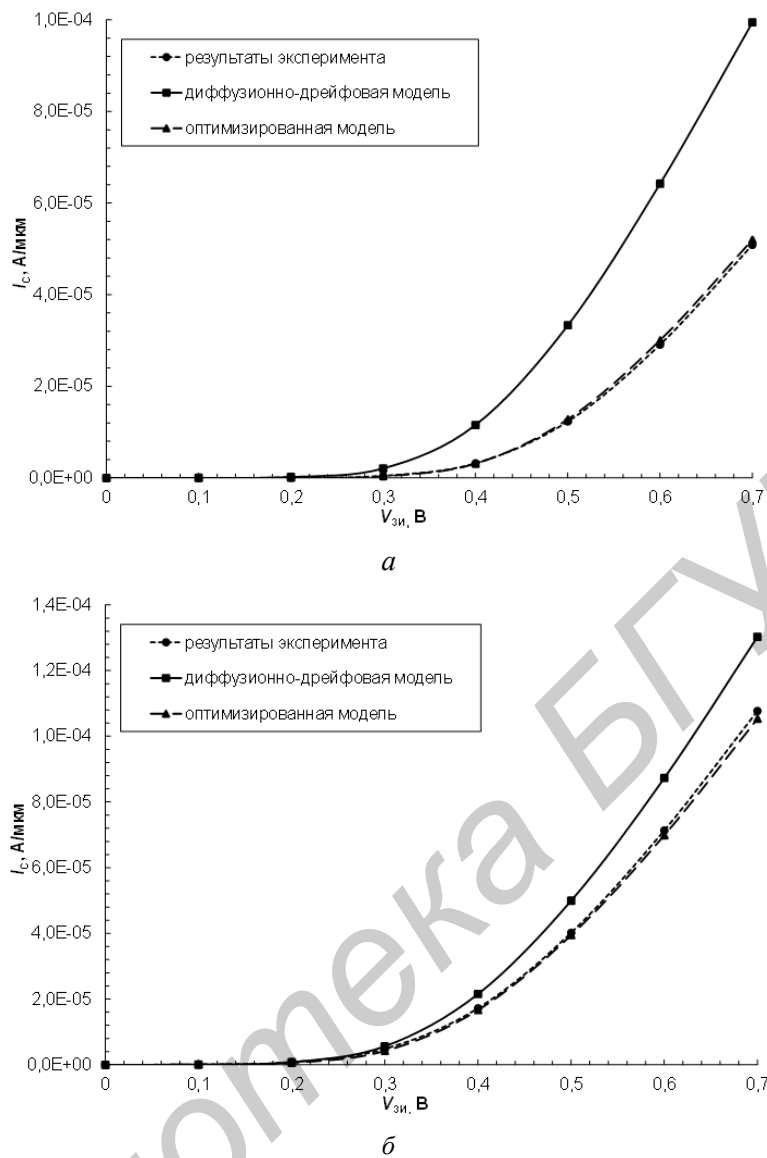


Рис. 2. ВАХ зависимости тока стока от напряжения на затворе МОП-транзистора при $U_{CИ} = 0,21$ В (а), $U_{CИ} = 2,01$ В (б)

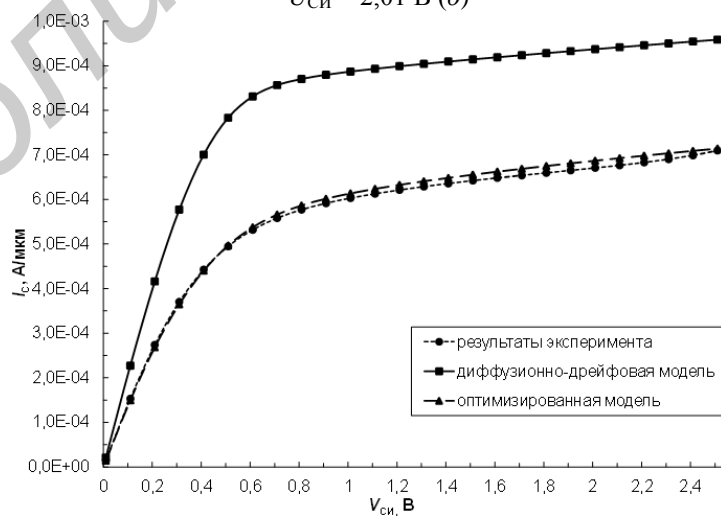


Рис. 3. ВАХ зависимости тока стока от напряжения на стоке МОП-транзистора при $U_{ZH} = 2$ В

Заключение

Рассмотрен новый подход расчета электрических характеристик наноразмерных МОП-транзисторов, сущность которого заключается в использовании корректирующих коэффициентов, а также таких значений параметров классических диффузионно-дрейфовых моделей, применяемых в стандартных комплексах приборно-технологического моделирования субмикронных приборов при описании физических механизмов транспорта носителей заряда, которые бы эффективно учитывали квантово-механические механизмы транспорта, характерные для наноразмерных приборов. С этой целью варьировались значения плотности состояний в области канала, а также были введены корректирующие коэффициенты в выражения моделей подвижности в поперечном и продольном электрических полях.

Разработан и использован при оптимизации параметров диффузионно-дрейфовой модели модифицированный метод прямого поиска для МОП-транзистора с длиной канала 90 нм. Анализ результатов свидетельствует о перспективности использования метода прямого поиска в задачах оптимизации параметров физико-математических моделей. Преимуществом представленного алгоритма является относительная простота реализации и отсутствие необходимости анализа большого количества комбинаций значимых параметров.

Результаты, полученные с использованием оптимизированной диффузионно-дрейфовой модели, соответствуют экспериментальным, что свидетельствует об эффективности применения предложенного подхода при адаптации физико-математических моделей.

DRIFT-DIFFUSION MODEL PARAMETERS OPTIMIZATION

TRUNG TRAN TUAN, A.M. BOROVİK, V.R. STEMPIŦSKY

Abstract

A new approach of nanoscale MOSFETs electrical characteristics calculating, the essence of which is the use of correction factors, as well as such values of classic drift-diffusion models, which would effectively take into account the quantum-mechanical transport mechanisms is proposed. Modified direct search method of drift-diffusion model optimization for MOSFET with a 90 nm channel length is developed and used.

Список литературы

1. *Vasileska D., Khan H.R., Ahmed S.S.* // Nano-Electronic Devices. 2005. Vol. 4. P. 305–361.
2. *Нелаев В.В., Стелпицкий В.Р.* Основы САПР в микроэлектронике. Моделирование технологии и прибора. Минск, 2008.
3. *Денисенко В.В.* Компактные модели МОП-транзисторов для SPICE в микро- и нанoeлектронике. М., 2010.
4. *Красников Г.Я.* Конструктивно-технологические особенности субмикронных МОП-транзисторов. М., 2011.
5. ATLAS User's Manual. Device simulation software. [Электронный ресурс]. – Режим доступа: atlas_users.pdf. – Дата доступа: 21.07.2014.
6. *Darwish M., Lentz J.L., Pinto M.R. et al* // IEEE Trans. Electron Devices. 1997. Vol. 44 (9). P. 1529–1538.
7. *Борисенко В.Е., Воробьева А.И., Данилюк А.Л. и др.* Нанoeлектроника: Теория и практика. М., 2013.
8. *Драгунов В.П., Неизвестный И.Г., Гридчин В.А.* Основы нанoeлектроники: Учебное пособие. Новосибирск, 2000.
9. *Аоки М.* Введение в методы оптимизации. М., 1977.
10. *Реклейтис Г., Рейвиндран А., Регсдел К.* Оптимизация в технике. Книга 1. М., 1986.
11. *Stempitsky V., Trung Tran Tuan, Borovik A.* // Proc. of Int. Workshop on New Approaches to High-Tech: Nano-Design, Technology, Computer Simulations (NDTCS'2013). Minsk, 11–15 June 2013. P. 85–87.
12. «Well-Tempered» Bulk-Si NMOSFET Device Home Page. [Электронный ресурс]. – Режим доступа: <http://www-mtl.mit.edu/researchgroups/Well/>. – Дата доступа: 21.07.2014.

УДК 538.911: 537.6/8

МОДЕЛИРОВАНИЕ ЭЛЕКТРОННЫХ И МАГНИТНЫХ СВОЙСТВ НИЗКОРАЗМЕРНЫХ МАТЕРИАЛОВ С ИСПОЛЬЗОВАНИЕМ ПЕРВОПРИНЦИПНЫХ МЕТОДОВ

М.С. ЗЕЛЕНИНА, О.А. КОЗЛОВА

Белорусский государственный университет информатики и радиоэлектроники
П. Бровка, 6, Минск, 220013, Беларусь

Поступила в редакцию 22 июля 2014

В рамках квантово-механического подхода исследованы электронные и магнитные свойства зернистых пленок ZnO и квазидвухмерных структур TiS₂ и TiSe₂. В качестве среды моделирования использован программный пакет VASP, в частности метод присоединенных плоских волн (PAW-метод). Рассчитаны значения плотности электронных состояний соединений, приведено распределение магнитной плотности зернистых пленок ZnO.

Ключевые слова: ab initio моделирование, теория плотности функционала, низкоразмерные структуры, электронные свойства.

Введение

Фундаментальные и прикладные исследования последних лет продемонстрировали широкие возможности применения наноструктурированных материалов в различных областях науки и техники. Эффективным инструментом в дальнейшем прогрессе понимания и объяснения физических явлений, происходящих в наноразмерных объектах, принадлежит использованию *ab initio*, первопринципных методов моделирования.

Особое внимание с настоящее время привлекает оксид цинка (ZnO). В частности, экспериментальные исследования показали, что наноструктурированный ZnO при определенных условиях приобретает ферромагнитные свойства, обладая при этом высокой температурой фазового перехода второго рода [1]. За последние десять лет было проведено огромное число экспериментальных исследований с целью выявления ферромагнетизма у ZnO и уже не раз удавалось получить соединения на основе оксида цинка с температурой Кюри выше комнатной [2]. Это открывает возможности использования оксида цинка в качестве материала структурных элементов запоминающих устройств, функционирующих на спиновых эффектах. Однако результаты, полученные исследователями, не сводятся к единому выводу, а именно, если одна группа коллективов экспериментаторов получила данные, свидетельствующие о наличии ферромагнетизма в ZnO, то другая группа не смогла выявить его [2–5].

Еще одной группой перспективных наноструктурированных материалов являются ультратонкие квази-двухмерные структуры слоистых дихалькогенидов переходных металлов (ХПМ). Материалы данного типа технологически просты в производстве, эффективны в использовании и обладают более широким диапазоном электро-химических свойств [6, 7]. В частности, структуры, состоящие из одного или нескольких слоев ХПМ, полученные посредством отслаивания объемных материалов или CVD-методом, являются прямозонными полупроводниками, значение запрещенной зоны и тип носителя заряда которых колеблется в зависимости от кристаллографической конфигурации, химического состава и размерности соединений. Таким образом, возможность модификации электронной структуры ХПМ делает их привлекательными для различных приложений микро- и нанoeлектроники.

В дополнение к составу и кристаллографической конфигурации атомов в соединениях ХПМ, размерность играет решающую роль в определении их основных электронных свойств. Наиболее ярким примером проявления аналогичного поведения является графен, уникальные свойства которого отсутствуют в объемном графите [8–10]. Высокие исследования электронных свойств графена и разработанных методологий подготовки и получения сверхтонких слоев привело к освоению других двухмерных материалов [6]. В частности, монослои ХПМ структур, подобные структуре графита, получили значительное внимание, так как некоторые из них являются полупроводниками со значительной шириной запрещенной зоны и широко распространены в природе.

В слоистых структурах ХПМ IV группы каждый слой имеет толщину $6 \sim 7 \text{ \AA}$, который состоит из гексагонально упакованного слоя атомов металла между двумя слоями атомов халькогена. Связи М - Х внутри слоя преимущественно ковалентные, слои соединены слабыми ван-дер-ваальсовыми силами, таким образом, возможно легко расщепить кристалл вдоль поверхности слоя. Соединения ХПМ IV группы являются узкозонными полупроводниками или полуметаллами, характеризуются низким значением удельного электрического сопротивления ($\sim 10 - 6 \text{ Ом}\cdot\text{см}$) и периодическим перераспределением в пространстве электронного, ионного и суммарного зарядов (волна зарядовой плотности) [7]. Длина связи М – М в структурах ХПМ IV группы изменяется в диапазоне от 3,15 до 4,03 \AA , в зависимости от размера ионов металла и халькогена. Эти величины на 15–25 % больше, чем длины связей, в соединениях переходных металлов твердых веществ, что указывает на ограниченное энергетическое и пространственное перекрытие орбиталей в соединениях ХПМ. Кристаллографическая конфигурация металлов М в структуре ХПМ может быть как тригонально призматической, так и октаэдрической [11].

Методология

Для анализа магнитного состояния наноструктуры традиционно используются современные методы зонных расчетов, основанные на теории функционала электронной плотности (density-functional theory – DFT) [12]. Наиболее популярными среди них являются полнопотенциальные методы, такие как линейаризированный метод присоединенных плоских волн (linearized augmented plane wave – LAPW), метод линейных muffin-tin-орбиталей (linear muffin-tin orbital – LMTO), первопринципный (*ab initio*) метод псевдопотенциалов, реализованный в программном пакете VASP (Vienna Ab initio Simulation Package), методы не эмпирического псевдопотенциала, Корринги – Кона – Ростокера (KKR) в приближении когерентного потенциала (coherent potential approximation) (KKR-CPA) и ряд других [13–15]. Получаемые результаты касаются распределения энергетических зон, плотностей состояний и характера их спинового расщепления, величин магнитных моментов атомов-компонентов в дефектных системах. Оценка энергий различных магнитных состояний системы используется для обоснования типа спинового упорядочения. Еще одним важным расчетным параметром, который привлекаемыми при изучении типа коллективного магнетизма, являются параметры обменных взаимодействий.

Особое внимание при расчетах в рамках DFT метода следует уделить обменно-корреляционному потенциалу, поскольку при описании сильно коррелированных систем, при оценках величины запрещенной зоны и в других случаях могут возникнуть несоответствия модели с реальной системой. Преодолеть эти затруднения возможно различными способами учета электронных корреляций, в частности приближения локальной спиновой плотности (local spin-density approximation) с учетом кулоновских корреляций (LSDA+U) и ряда сходных схем. Другим важным моментом, определяющим степень достоверности результатов, является учет эффектов структурной релаксации системы с дефектом. Дело в том, что энергетическая неустойчивость обсуждаемых далее примесных или нестехиометрических систем связана с высокой плотностью состояний на уровне Ферми и переход в устойчивое состояние может протекать как вследствие спинового расщепления прифермиевских зон (т.е. при переходе системы в магнитное состояние), так и за счет структурного фактора – отклонения атомов, соседствующих с дефектом, от их равновесных позиций в «идеальном» кристалле.

Наконец, при рассмотрении излагаемых результатов теоретических прогнозов следует учитывать тот факт, что получаемые результаты часто носят сугубо модельный характер, т.е.

относятся к гипотетическим системам с составами, реальный синтез которых в обычных условиях вряд ли может быть осуществлен.

Подобная ситуация отчетливо проявилась при попытках получения разбавленных ферромагнитных полупроводников с содержанием магнитных *d*-примесей, которые согласно прогнозам должны иметь высокие (превосходящие комнатную температуру) точки Кюри. При этом требуемые (достаточно большие) концентрации примесей приводили к изменению фазового состава и структуры исходного материала. Эти трудности в полной мере будут присутствовать при реализации теоретических прогнозов для рассматриваемого материала. Поэтому сам по себе факт обнаружения в рамках зонной теории спиновой поляризации примесных (или вакансионных) состояний в немагнитных *sp*-матрицах отнюдь не гарантирует возможности экспериментального наблюдения этого явления и тем более создания магнитного материала с высокой T_c , а может служить лишь предварительным ориентиром для продвижения в этом направлении [2].

Моделирование проводилось в программном комплексе VASP, который реализует первопринципные методы расчетов фундаментальных электронных, оптических и магнитных свойств материалов. Данная программа для квантово-механических расчетов применяет метод псевдопотенциала и разложения волновых функций в базисе плоских волн.

Расчеты проводились с учетом периодических условий, поэтому для того, чтобы исключить взаимное влияние границ зерен ZnO друг на друга использовалась супер-ячейка, размером 4x8x1 элементарных ячеек ZnO. Для того, чтобы исключить взаимодействие между слоями в структурах TiS₂ и TiSe₂, между ними закладывался вакуумный промежуток, равный 15 Å. Энергия обрезания плоских волн в расчетах составляла 500 эВ. Разбиение обратного пространства выполнялось методом Monkhorst Pack размером 4×4×1 для ZnO и 8×8×6 для TiS₂ и TiSe₂. Критерий энергии сходимости геометрии структур был равен 1·10⁻⁶ эВ. Расчеты структур TiS₂ и TiSe₂ выполнялись с использованием гибридного функционала PBE0, который является комбинацией обменных вкладов, определяемых методом Хартри-Фока и при помощи функционалов DFT. Гибридный функционал PBE0 содержит 25 % хартри-фоковской (HF) обменной энергии (E_x^{HF}), $E_{xc} = E_{xc}^{LDA} + \frac{1}{4}(E_x^{HF} - E_{xc}^{PBE})$, где в верхнем индексе (LDA, HF, и PBE) указан метод расчета обменно-корреляционной энергии. Вклад E_x^{HF} – эмпирически подобранный параметр [16].

Кристалло-химические параметры для TiS₂ и TiSe₂ представлены в табл. 1 [17].

Таблица 1. Кристалло-химические параметры для TiS₂ и TiSe₂

Материал	Параметры элементарной ячейки, a, c, соотв. (Å)	Уайкофф (Wyckoff) позиции	Радиус Вигнера-Зейтца R_{WS} (Å)	Координаты	
ZnO	3,350; 5,220;	Zn	2b	1,270	0,(3) 0,(6) 0,0
		O	2b	0,820	0,(3) 0,(6) 0,375
TiS ₂	3,407; 5,695	Ti	1a	1,2067	0,0 0,0 0,0
		S	2d	1,7209	0,(3) 0,(6) 0,2501
TiSe ₂	3,540; 6,008	Ti	1a	1,2067	0,0 0,0 0,0
		Se	2d	1,7209	0,(3) 0,(6) 0,2501

Результаты и их обсуждение

Межзеренную границу пленок оксида цинка можно рассматривать как область, в которой осуществляется контакт двух кристаллов, различающихся только ориентацией. Граница зерен – атомный слой толщиной 1–2 атомных диаметра, по обе стороны от которого кристаллические решетки различаются только пространственной разориентацией.

По такому принципу была воспроизведена структура межзеренной границы с углом разориентации 10 градусов (рис. 1). Расчеты проводили с учетом периодических условий, поэтому для того, чтобы исключить влияние дефектов границ друг на друга, использовалась супер-ячейка, которая имела размерность 16×4×1 элементарных ячейки ZnO, что соответствует 5,024×1,256 нм.

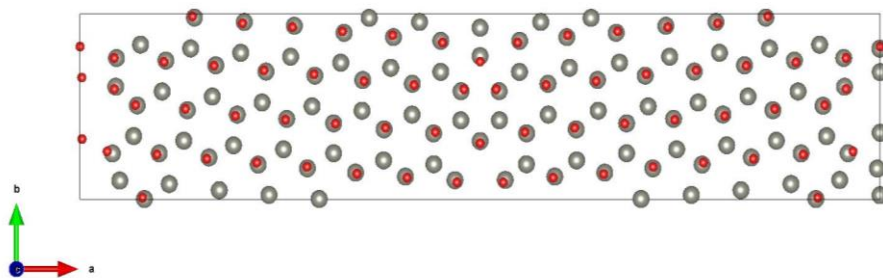


Рис. 1. Модель границы зерна ZnO после процесса релаксации

Проведено исследование ферромагнитного состояния системы. Установлено, что структура имеет стабильную ферромагнитную фазу. Магнитный момент равен $2,51 \mu_B$ (магнетон Бора). Получено распределение эффективной спиновой магнитной плотности в структуре (рис. 2).



Рис. 2. Распределение магнитной плотности в дефектной структуре ZnO

Из рис. 2 видно, что распределение магнитной плотности в дефектной структуре ZnO носит анизотропный характер: значение магнитной плотности достигает максимального значения в области межзеренной границы и равно нулю в областях, отдаленных от нее. Эта особенность позволяет сделать предположение о наличии зависимости ферромагнетизма ZnO от размера и разориентации зерен в его структуре. Такое предположение позволяет объяснить различия в результатах измерений ферромагнетизма в образцах ZnO и дает возможность использования данной структуры в приборах сенсорики.

Рассчитано распределение электронной плотности и зонная диаграмма для объемной и квазидвухмерной структуры TiS_2 и $TiSe_2$. На рис. 3 представлено распределение электронной плотности для объемной и квазидвухмерной структуры структуры TiS_2 .

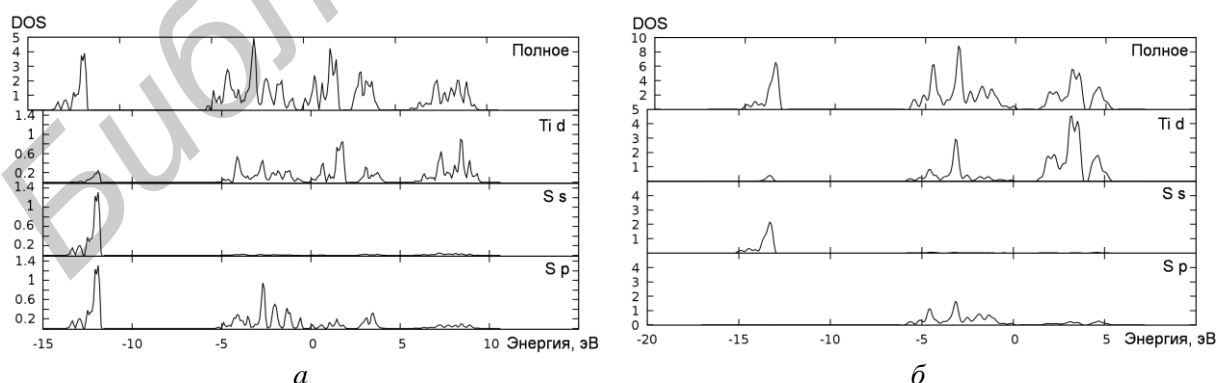


Рис. 3. Полное и поорбитальное распределение плотности электронных состояний (DOS):
a – для объемной структуры TiS_2 ; *б* – для квазидвухмерной структуры TiS_2

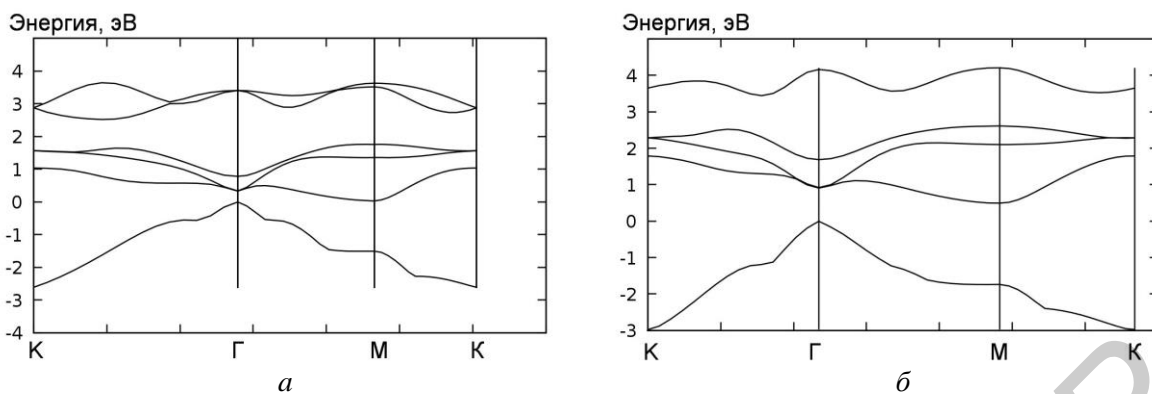


Рис. 4. Зонная диаграмма TiS_2 : *a* – для объемной структуры TiS_2 ;
б – для квазидвухмерной структуры TiS_2

В диапазоне от -15 до -10 эВ TiS_2 состоит в основном из энергетических состояний S $3s$ -орбиталей. В диапазоне от -5 до 5 эВ зоны заполнены состояниями S $3p$ -орбиталей, находящимися в гибридизации с Ti $3d$ орбиталями. (Максимум и минимум этого участка находится в Γ -точке). Зоны, расположенные над уровнем Ферми заполнены преимущественно состояниями Ti $3d$ – орбиталей.

Распределение электронной плотности и зонная диаграмма объемного TiSe_2 незначительно отличается от структуры TiS_2 . Энергетические зоны TiSe_2 , расположенные в диапазоне энергий от -14 эВ и до $-12,5$ эВ, образованы из состояний Se $3s$ -орбиталей. В диапазоне от -5 эВ до $4,5$ эВ зоны заполнены состояниями Se $3p$ -орбиталей, находящимися в гибридизации с Ti $3d$ орбиталями. Зоны, расположенные над уровнем Ферми заполнены преимущественно состояниями Ti $3d$ – орбиталей.

Распределение плотности электронных состояний квазидвухмерной структуры TiS_2 аналогично результатам, полученным для объемной структуры. Это связано с тем, что соединение имеет сильные ковалентные внутрислоевые взаимосвязи. Однако необходимо отметить ключевое отличие в свойствах объемной и квазидвухмерной структуры. Объемный TiS_2 является полуметаллом, а его квазидвухмерный аналог TiS_2 является полупроводником с шириной запрещенной зоны $1,1$ эВ (рис. 4). Полученные результаты согласуются с теоретическими и прикладными результатами, описанными в работе [18]. Обнаружено, что при переходе от объемной к двумерной структуре TiSe_2 электронные свойства меняются. Квазидвухмерная структура TiSe_2 является полупроводником с шириной запрещенной зоны $0,7$ эВ.

Таким образом, анализ результатов моделирования свидетельствует о многообразии физико-химических свойств указанных объемных и квазидвухмерных структур на основе ХПМ, что, в свою очередь, предоставляет множество возможностей для использования электронных и магнитных свойств указанных соединений для широкого круга приложений микро- и наноэлектроники, в частности для устройств хранения энергии, каталитически активных элементов в оптоэлектронике, и в структурных компонентах микроэлектронных приборов. Однако остаются нерешенными ключевые проблемы, связанные с увеличением электрической проводимости и повышением стабильности квазидвухмерных структур, решение которых требует интеграции экспериментальных и теоретических исследований.

Заключение

Таким образом, в рамках квантово-механического подхода исследованы электронные и магнитные свойства зернистых пленок ZnO и квазидвухмерных структур TiS_2 и TiSe_2 . Установлено, что зернистая структура ZnO имеет стабильную ферромагнитную фазу. Магнитный момент равен $2,51 \mu\text{B}$ (магнетон Бора). Получено распределение эффективной спиновой магнитной плотности в структуре. Обнаружено, что объемный TiS_2 является полуметаллом, а его квазидвухмерный аналог TiS_2 является полупроводником с шириной запрещенной зоны $1,1$ эВ. Электронные свойства TiSe_2 также изменяются при уменьшении размерности структуры. Квазидвухмерная структура TiSe_2 является полупроводником с

шириной запрещенной зоны 0,7 эВ. Магнитные свойства в исследованных структурах TiS₂ и TiSe₂ не обнаружены.

SIMULATION OF ELECTRONIC AND MAGNETIC PROPERTIES OF LOW-DIMENSIONAL MATERIALS WITH USE OF FIRST PRINCIPLES METHODS

M.S. ZELENINA, O.A. KOZLOVA

Abstract

The result of electronic and magnetic properties investigation of ZnO granular films and quasi-two-dimensional structures of TiS₂ and TiSe₂ were presented. Quantum-mechanical approach in software package VASP was used during the calculation procedure, in particular the method of augmented plane wave (PAW-method). The values of the electronic states' density of the compounds were presented, the distribution of the magnetic density of ZnO granular films were shown.

Список литературы

1. Страумал Б. Б., Протасова С. Г. // Письма в ЖЭТФ. 2013. Т. 97. С. 415–426.
2. Ивановский А.Л. // УФН. 2007. Т. 177, № 10, С. 1083–1104.
3. Byung-Sub Kang, Kwang-Pyo Chae. // J. Mag. 2012. № 17 (3). P. 163–167.
4. Bin Shao, Hong Liu, Jian Wu et al. // J. Appl. Phys. 2012. № 111. 07C301-07C301-3.
5. Bin Shao, Min Feng, Hong Liu et al. // J. Appl. Phys. 2013. 17C728-17C725-3.
6. Li H. // Small. 2012. Vol. 8. P. 63–67.
7. Castro Neto A.H. // Phys. Rev. Lett. 2001. Vol. 86. P. 4382–4385.
8. Novoselov K.S. // Science. 2011. Vol. 306. P. 666–669.
9. Castro Neto A. H., Guinea F., Peres N.M.R. et. al. // Rev. Mod. Phys. 2009. Vol. 81. P.109–162.
10. Geim A.K. // Science Vol. 2009. Vol. 324. P. 1530–1534.
11. Van de Walle C.G. First-principles calculations of hyperfine parameters // Physical Review. 1993. B 47. P. 4244 – 4248.
12. Chhowalla M., Shin H.S., Eda G. et. al. // Nature chemistry. 2013. Vol. 5. P. 263–275.
13. Kresse G. // Comput. Mat. Sci. 1996. № 6. P. 15–19.
14. Kresse G., Marsman M., Furthmüller J. VASP the guide: tutorial. Vienna, 2003.
15. Skriver H.L., Rosengaard N.M. // Phys. Rev. 1991. B 43, 12. 9538.
16. Perdew J., Ernzerhof M. Burke K. // J. Chem. Phys. 1996. № 105. 9982.
17. Bilbao. – [Электронный ресурс]. – Режим доступа: <http://cryst.ehu.es/>. – Дата доступа: 22.07.2014.
18. Fang C.M., de Groot R.A., Haas C. // Phys. Rev. 1997. B 56, 4455.

УДК 517.977

УСЛОВИЯ ОПТИМАЛЬНОСТИ ВТОРОГО ПОРЯДКА

А.Е. ЛЕЩЁВ, Л.И. МИНЧЕНКО

Белорусский государственный университет информатики и радиоэлектроники
П. Бровка, 6, Минск, 220013, Беларусь

Поступила в редакцию 9 октября 2014

Необходимые условия второго порядка играют важную роль в теории оптимизации. Это объясняется тем, что большинство используемых на практике численных алгоритмов сводится к нахождению стационарных точек, удовлетворяющих условиям оптимальности первого порядка. В то же время многие задачи оптимизации, особенно задачи высокой размерности, имеют достаточно большое число стационарных точек. В связи с этим возникает проблема усиления необходимых условий за счет привлечения необходимых условий второго порядка для удаления неоптимальных стационарных точек. В данной статье рассматриваются так называемые слабые необходимые условия оптимальности второго порядка и доказывается их справедливость при менее жестких требованиях по сравнению с известными ранее результатами.

Ключевые слова: нелинейное программирование, условия оптимальности второго порядка, условия регулярности.

Введение

Пусть $h_i(y)$, $i=1,2,\dots,p$ – дважды непрерывно дифференцируемые функции из R^m в R . Введем непустое множество допустимых точек $C = \{y \in R^m \mid h_i(y) \leq 0, i \in I, h_i(y) = 0, i \in I_0\}$, где $y \in R^m$, $I = \{1, \dots, s\}$, $I_0 = \{s+1, \dots, p\}$ или $I_0 = \emptyset$, и рассмотрим задачу (NLP) нелинейного программирования $f(y) \rightarrow \min, y \in C$ с дважды непрерывно дифференцируемой целевой функцией f .

Для задачи (NLP) введем функцию Лагранжа $L(y, \lambda) = f(y) + \langle \lambda, h(y) \rangle$, где $\lambda = (\lambda_1, \dots, \lambda_p)$, $h = (h_1, \dots, h_p)$, и множество множителей Лагранжа в точке y : $\Lambda(y) = \{ \lambda \in R^p \mid \nabla_y L(y, \lambda) = 0, \lambda_i \geq 0 \text{ и } \lambda_i h_i(y) = 0, i \in I \}$.

Обозначим через $I(y) = \{i \in I \mid h_i(y) = 0\}$ множество индексов активных в точке $y \in C$ ограничений типа неравенства.

Необходимые условия оптимальности в задачах математического программирования делятся на условия оптимальности первого порядка, когда для оптимальности в точке $y \in C$ требуется выполнение условий Куна-Таккера, т.е. существования множителей Лагранжа $\lambda \in \Lambda(y)$ (такая точка называется стационарной), и условия оптимальности второго порядка, когда дополнительно к существованию множителей Лагранжа требуется, чтобы в данной точке матрица вторых производных функции Лагранжа была неотрицательно определенной на некотором конусе касательных направлений к множеству C .

Наряду с необходимыми условиями оптимальности важную роль в теории оптимизации играют условиями регулярности, гарантирующие справедливость необходимых условий оптимальности в исследуемой точке $y^0 \in C$.

Введем касательный (контингентный) конус к множеству C в точке $y^0 \in C$:

$\widehat{T}_C(y) = \{\bar{y} \in R^m \mid \exists t_k \downarrow 0 \text{ и } \bar{y}_k \rightarrow \bar{y} \text{ такие, что } y + t_k \bar{y}_k \in C, \quad k=1,2,\dots\}$,

а также линеаризованный касательный конус

$\Gamma_C(y^0) = \{\bar{y} \in R^m \mid \langle \nabla h_i(y^0), \bar{y} \rangle \leq 0, \quad i \in I(y^0), \quad \langle \nabla h_i(y^0), \bar{y} \rangle = 0, \quad i \in I_0\}$.

Нетрудно заметить, что $\widehat{T}_C(y) \subset \Gamma_C(y^0)$. К условиям регулярности, гарантирующим справедливость необходимых условий Куна-Таккера, в частности, относится требование выполнения равенства $\widehat{T}_C(y) = \Gamma_C(y^0)$ (условие регулярности Абади). Отметим, что хотя данное условие регулярности носит весьма общий характер, оно не является конструктивным в смысле возможности практической проверки.

Одним из наиболее известных условий регулярности является условие линейной независимости градиентов $\nabla h_i(y^0), \quad i \in I(y^0) \cup I_0$. Более общий характер носит широко применяемое условие регулярности Мангасаряна-Фромовица, требующее чтобы в точке $y^0 \in C$ система векторов $\nabla h_i(y^0), \quad i \in I_0$ была линейно независимой и существовал вектор \bar{y}^0 такой, что $\langle \nabla h_i(y^0), \bar{y}^0 \rangle = 0, \quad i \in I_0, \quad \langle \nabla h_i(y^0), \bar{y}^0 \rangle < 0, \quad i \in I(y^0)$. Данное условие равносильно требованию $\Lambda_0(y^0) = \{\lambda \in R^p \mid \sum_{i=1}^p \lambda_i \nabla h_i(y^0) = 0, \quad \lambda_i \geq 0, \quad i \in I(y^0), \quad \lambda_i = 0, \quad i \in I \setminus I(y^0)\} = \{0\}$.

Введем конус критических направлений $D_C(y^0) = \{\bar{y} \in \Gamma_C(y^0) \mid \langle \nabla f(y_0), \bar{y} \rangle = 0\}$ и конус $S_C(y) = \{\bar{y} \in R^m \mid \langle \nabla h_i(y), \bar{y} \rangle = 0, \quad i \in I_0 \cup I(y^0)\}$.

Будем говорить, что в стационарной точке $y^0 \in C$ выполняется необходимое условие оптимальности второго порядка (SONC), если найдется множитель $\lambda \in \Lambda(y^0)$, для которого выполняется неравенство $\langle \bar{y}, \nabla_{yy}^2 L(y^0, \lambda) \bar{y} \rangle \geq 0$ для всех $\bar{y} \in D_C(y^0)$.

Необходимые условия SONC можно найти в работах [1–3], где они получены при различных условиях регулярности ограничений задачи в исследуемых точках. В частности, условия SONC справедливы при выполнении классического условия регулярности Мангасаряна-Фромовица.

Будем говорить, что в точке $y^0 \in C$ выполняется слабое необходимое условие оптимальности второго порядка (WSONC) [4], если существует множитель $\lambda \in \Lambda(y^0)$, для которого выполняется неравенство $\langle \bar{y}, \nabla_{yy}^2 L(y^0, \lambda) \bar{y} \rangle \geq 0$ для всех $\bar{y} \in S_C(y^0)$.

Известно, что хотя классические необходимые условия оптимальности второго порядка SONC более эффективны для проверки стационарных точек на оптимальность, большинство практических алгоритмов, использующих необходимые условия оптимальности второго порядка, имеют дело со слабыми условиями оптимальности второго порядка (WSONC) [4–7]. В частности, это относится к методам штрафных функций и методам с использованием расширенных функций Лагранжа [8–10]. Таким образом, слабые необходимые условия оптимальности второго порядка обладают значительной ценностью.

В то же время известно [4], что справедливость слабых необходимых условий оптимальности второго порядка не гарантируется традиционными условиями регулярности Мангасаряна-Фромовица без дополнительных предположений, а значит и более слабыми условиями регулярности. В работе [4] предложено дополнительное условие, наличие которого вместе с условиями регулярности способно обеспечить справедливость слабых необходимых условий оптимальности второго порядка в стационарных точках.

Следуя [4], будем говорить, что в точке $y^0 \in C$ выполняется слабое условие постоянного ранга (WCR), если $\text{rank}\{\nabla h_i(y), \quad i \in I_0 \cup I(y^0)\} = \text{const}$ в некоторой окрестности точки y^0 .

Из результатов [4] следует, что при совместном выполнении в стационарной точке условия WCR и условия регулярности Мангасаряна-Фромовица в данной точке выполняется и слабое необходимое условие оптимальности второго порядка WSONC.

Одной из целей данной статьи является доказательство слабых необходимых условий оптимальности второго порядка при более слабых условиях регулярности по сравнению с [4]. В данной статье также рассматриваются необходимые условия оптимальности второго порядка в усиленной форме SSONC (Strong Second Order Necessary Conditions) [11–13]. Говорят, что точка $y^0 \in C$ удовлетворяет условию SSONC, если при любом векторе $\lambda \in \Lambda(y^0)$ выполняется неравенство $\langle \bar{y}, \nabla_{yy}^2 L(y^0, \lambda) \bar{y} \rangle \geq 0$ для всех $\bar{y} \in K_C^\lambda(y^0)$,

где $K_C^\lambda(y^0)$ – конус критических направлений множества C в точке y^0 , связанный с множителем Лагранжа $\lambda \in \Lambda(y^0)$ и определяемый условием

$$K_C^\lambda(y^0) = \{ \bar{y} \in R^m \mid \langle \nabla h_i(y^0), \bar{y} \rangle = 0, i \in I_0, \\ \langle \nabla h_i(y^0), \bar{y} \rangle = 0 \text{ } i \in I^\oplus(y^0), \langle \nabla h_i(y^0), \bar{y} \rangle \leq 0, i \in I^\ominus(y^0) \},$$

где $I_\lambda^\oplus(y^0) = \{i \in I(y^0) \mid \lambda_i > 0\}$, $I_\lambda^\ominus(y^0) = \{i \in I(y^0) \mid \lambda_i = 0\}$.

Отметим, что конус критических направлений $K_C^\lambda(y^0)$ зависит от множителя Лагранжа λ и, следовательно, и от целевой функции f . Нетрудно показать, что, если $\Lambda(y^0) \neq \emptyset$, то $K_C^\lambda(y^0) = D_C(y^0)$ при всех $\lambda \in \Lambda(y^0)$.

В статье получены необходимые условия оптимальности второго порядка в усиленной форме, обобщающие известные ранее результаты [11–13].

1. Слабые условия оптимальности второго порядка

Следующая теорема дает слабые необходимые условия оптимальности второго порядка в задаче (NLP).

Теорема 1. Пусть точка $y^0 \in C$ является локальным решением задачи (NLP). Тогда существуют числа λ_i , $i = 0, 1, \dots, p$, такие, что выполнено условие $\lambda_0 \nabla f(y^0) + \sum_{i=1}^p \lambda_i \nabla h_i(y^0) = 0$, $\lambda_0^2 + \lambda_1^2 + \dots + \lambda_p^2 = 1$, где $\lambda_0 \geq 0$, $\lambda_i \geq 0$, $i \in I(y^0)$, $\lambda_i = 0$, $i \in I \setminus I(y^0)$.

Если дополнительно в точке $y^0 \in C$ выполняется условие WCR, то

$$\langle \bar{y}, [\lambda_0 \nabla^2 f(y^0) + \sum_{i=1}^p \lambda_i \nabla^2 h_i(y^0)] \bar{y} \rangle \geq 0 \text{ для всех } \bar{y} \in S_C(y^0).$$

Доказательство. Обозначим $h_i^+(y) = \max\{0, h_i(y)\}$ и для каждого целого положительного k введем вспомогательную задачу минимизации функции

$$G_k(y) = f(y) + \frac{k}{3} \sum_{i \in I(y^0)} (h_i^+(y))^3 + \frac{k}{2} \sum_{i \in I_0} (h_i(y))^2 + \frac{1}{4} |y - y^0|^4$$

на множестве $S = \{y \mid |y - y^0| \leq \varepsilon\}$, где $\varepsilon > 0$ такое, что $f(y^0) \leq f(y)$ и $h_i(y) < 0$ $i \in I \setminus I(y^0)$ для всех точек $y \in S \cap C$.

Пусть y^k – решение данной вспомогательной задачи. Очевидно, $G_k(y^k) \leq G_k(y^0)$ для любого k , т.е.

$$f(y^k) + \frac{k}{3} \sum_{i \in I(y^0)} (h_i^+(y^k))^3 + \frac{k}{2} \sum_{i \in I_0} (h_i(y^k))^2 + \frac{1}{4} |y^k - y^0|^4 \leq f(y^0).$$

Не убавив общности, можно считать, что $y^k \rightarrow y^* \in S$. Поскольку $f(y^k)$ ограничена на S , то $h_i^+(y^k) \rightarrow 0$ при $i \in I(y^0)$ и $h_i(y^k) \rightarrow 0$ при $i \in I_0$. Следовательно, $y^* \in C$, и тогда $f(y^0) \leq f(y^*)$.

С другой стороны, $f(y^k) + \frac{1}{4} |y^k - y^0|^4 \leq f(y^0)$ и, значит, $f(y^*) + \frac{1}{4} |y^* - y^0|^4 \leq f(y^0)$. Отсюда следует $y^* = y^0$. Таким образом, $y^k \rightarrow y^0$, оставаясь при достаточно больших

k внутренней точкой множества S . Тогда для больших k можно записать необходимые условия оптимальности для функции $G_k(y)$ в точке y^k :

$$\nabla G_k(y^k) = 0 \text{ и } \langle \bar{y}, \nabla^2 G_k(y^k) \bar{y} \rangle \geq 0 \text{ для всех } \bar{y} \in R^m.$$

Данные условия можно переписать в развернутом виде:

$$\nabla f(y^k) + \sum_{i \in I(y^0)} \xi_i^k \nabla h_i(y^k) + \sum_{i \in I_0} \xi_i^k \nabla h_i(y^k) + |y^k - y^0|^2 (y^k - y^0) = 0, \quad (1)$$

$$\text{где } \xi_i^k = k(h_i^+(y^k))^2, \quad i \in I(y^0), \quad \xi_i^k = k(h_i(y^k)), \quad i \in I_0,$$

и

$$\begin{aligned} & \langle \bar{y}, \nabla^2 f(y^k) \bar{y} \rangle + \langle \bar{y}, \sum_{i \in I_0 \cup I(y^0)} \xi_i^k \nabla^2 h_i(y^k) \bar{y} \rangle + \\ & + k \langle \bar{y}, \sum_{i \in I(y^0)} 2h_i^+(y^k) \nabla h_i(y^k)^T \nabla h_i(y^k) \bar{y} + \sum_{i \in I_0} \nabla h_i(y^k)^T \nabla h_i(y^k) \bar{y} \rangle + \\ & + 3|y^k - y^0|^2 \langle \bar{y}, E \bar{y} \rangle \geq 0. \end{aligned} \quad (2)$$

Положим

$$\delta_k = (1 + \sum_{i \in I_0 \cup I(y^0)} (\xi_i^k)^2)^{1/2}, \quad \lambda_0^k = 1/\delta_k, \quad \lambda_i^k = \xi_i^k / \delta_k, \quad i \in I_0 \cup I(y^0), \quad \lambda_i^k = 0, \quad i \in I \setminus I(y^0).$$

Разделив (1) на δ_k , получим

$$\lambda_0^k \nabla f(y^k) + \sum_{i=1}^p \lambda_i^k \nabla h_i(y^k) + \frac{1}{\delta_k} |y^k - y^0|^2 (y^k - y^0) = 0, \quad (3)$$

Поскольку $(\lambda_0^k)^2 + \sum_{i=1}^p (\lambda_i^k)^2 = 1$ и последовательность $\{\lambda_0^k, \lambda_1^k, \dots, \lambda_p^k\}$ ограничена, можно, не ограничив общности, считать ее сходящейся: $\lambda_0^k \rightarrow \lambda_0, \lambda_i^k \rightarrow \lambda_i$.

Тогда из (3) следует

$$\lambda_0 \nabla f(y^0) + \sum_{i=1}^p \lambda_i \nabla h_i(y^0) = 0, \quad (\lambda_0)^2 + \sum_{i=1}^p (\lambda_i)^2 = 1, \quad (4)$$

$$\lambda_0 \geq 0, \quad \lambda_i \geq 0, \quad i \in I(y^0), \quad \lambda_i = 0, \quad i \in I \setminus I(y^0).$$

Разделив (2) на δ_k , получим

$$\begin{aligned} & \langle \bar{y}, [\lambda_0^k \nabla^2 f(y^k) + \sum_{i=1}^p \lambda_i^k \nabla^2 h_i(y^k)] \bar{y} \rangle + \\ & + \frac{k}{\delta_k} \langle \bar{y}, \sum_{i \in I(y^0)} 2h_i^+(y^k) \nabla h_i(y^k)^T \nabla h_i(y^k) \bar{y} + \sum_{i \in I_0} \nabla h_i(y^k)^T \nabla h_i(y^k) \bar{y} \rangle + \\ & + \frac{3}{\delta_k} |y^k - y^0|^2 \langle \bar{y}, E \bar{y} \rangle \geq 0 \end{aligned} \quad (5)$$

для всех $\bar{y} \in R^m$.

Оценим третье слагаемое в (5):

$$\begin{aligned} & \left| \frac{k}{\delta_k} \langle \bar{y}, \sum_{i \in I(y^0)} 2h_i^+(y^k) \nabla h_i(y^k)^T \nabla h_i(y^k) \bar{y} + \sum_{i \in I_0} \nabla h_i(y^k)^T \nabla h_i(y^k) \bar{y} \rangle \right| \leq \\ & \leq \frac{k}{\delta_k} \sum_{i \in I_0 \cup I(y^0)} M(y^k) |\langle \nabla h_i(y^k), \bar{y} \rangle|, \end{aligned}$$

где $M(y^k)$ – ограниченная величина.

Для любого $\bar{y} \in S_c(y^k)$ из (5) следует

$$\langle \bar{y}, [\lambda_0^k \nabla^2 f(y^k) + \sum_{i=1}^p \lambda_i^k \nabla^2 h_i(y^k)] \bar{y} \rangle + \frac{3}{\delta_k} |y^k - y^0|^2 \langle \bar{y}, E \bar{y} \rangle \geq 0. \quad (6)$$

В силу условия *WCR* в точке y^0 справедливо равенство $\text{rank}\{\nabla h_i(y), i \in I_0 \cup I(y^0)\} = \text{rank}\{\nabla h_i(y^0), i \in I_0 \cup I(y^0)\} = l$ для всех y достаточно близких к точке y^0 .

Не ограничивая общности будем считать, что в системе $\langle \nabla h_i(y^0), \bar{y} \rangle = 0, i \in I_0 \cup I(y^0)$ (7)

ранг достигается для первых l уравнений и первых l переменных $\bar{y}_1, \dots, \bar{y}_l$. Тогда система (7) равносильна системе

$$B(y^0) \bar{y}^1 + D(y^0) \bar{y}^2 = 0 \text{ или } \bar{y}^1 = -B^{-1}(y^0) D(y^0) \bar{y}^2,$$

где $\bar{y}^1 = (\bar{y}_1, \dots, \bar{y}_l)^T, \bar{y}^2 = (\bar{y}_{l+1}, \dots, \bar{y}_m)^T,$

$$B(y) = \left[\frac{\partial h_i(y)}{\partial y_j} \quad i, j = 1, \dots, l \right], \quad D(y) = \left[\begin{array}{c} \frac{\partial h_i(y)}{\partial y_j} \quad i = 1, \dots, l \\ j = l+1, \dots, m \end{array} \right].$$

Тогда для любого вектора $\bar{y} = (\bar{y}^1, \bar{y}^2)^T \in S_C(y^0)$ можно построить вектор $\bar{y}^k = (\bar{y}^{1k}, \bar{y}^{2k})^T \in S_C(y^k)$, такой, что $\bar{y}^{1k} = -B^{-1}(y^k) D(y^k) \bar{y}^2, \bar{y}^{2k} = \bar{y}^2$.

Тогда $\bar{y}^k \rightarrow \bar{y}$ и, следовательно, подставив в (6) $\bar{y} = \bar{y}^k$ и переходя к пределу, получим

$$\langle \bar{y}, [\lambda_0 \nabla^2 f(y^0) + \sum_{i=1}^p \lambda_i \nabla^2 h_i(y^0)] \bar{y} \rangle \geq 0 \text{ для всех } \bar{y} \in S_C(y^0).$$

Отметим, что теорема 1 обобщает результат [4].

Следствие 1. Если точка $y^0 \in C$ является локальным решением задачи (NLP) и в ней выполнены условие регулярности Мангасаряна-Фромовица и условие *WCR*, то в этой точке необходимо выполняется условие *WSOFC*.

Теорема 1 является более общим утверждением по сравнению со следствием 1, которое повторяет результат [4].

2. Условие критической регулярности и необходимые условия оптимальности второго порядка в усиленной форме

Необходимые условия оптимальности второго порядка в усиленной форме имеют определенные преимущества по сравнению с классическими необходимыми условиями оптимальности второго порядка (SONC) (см., например, [3, 14]), которые требуют чтобы для любого вектора $\bar{y} \in D_C(y^0)$ существовал хотя бы один множитель $\lambda \in \Lambda(y^0)$ такой, что

$$\langle \bar{y}, \nabla_{yy}^2 L(y^0, \lambda) \bar{y} \rangle \geq 0.$$

Вообще говоря, условие Куна-Таккера является необходимым условием для точки локального минимума задачи (NLP) только при выполнении некоторых дополнительных условий регулярности в этой точке. Одним из наиболее известных условий регулярности является условие линейной независимости градиентов $\nabla h_i(y^0), i \in I(y^0) \cup I_0$, где $y^0 \in C$. Более общий характер носит широко применяемое условие регулярности Мангасаряна-Фромовица, требующее чтобы в точке $y^0 \in C$ система векторов $\nabla h_i(y^0), i \in I_0$ была линейно независимой и существовал вектор \bar{y}^0 такой, что

$$\langle \nabla h_i(y^0), \bar{y}^0 \rangle = 0, i \in I_0, \quad \langle \nabla h_i(y^0), \bar{y}^0 \rangle < 0, i \in I(y^0).$$

В [15, 16] предложена ослабленная версия условия Мангасаряна-Фромовица, названная в [15] ослабленным (обобщенным) условием Мангасаряна-Фромовица (RMFCQ), а в [16] названная CRSC.

Представим множество индексов $I(y^0)$ активных ограничений в точке $y^0 \in C$ в виде $I(y^0) = I_C^0(y^0) \cup I_C^+(y^0)$, где $I_C^0(y^0) = \{i \in I(y^0) \mid \langle \nabla h_i(y^0), \bar{y} \rangle = 0, \forall \bar{y} \in \Gamma_C(y^0)\}$, $I_C^+(y^0) = I(y^0) \setminus I_C^0(y^0)$.

Будем говорить, что в точке $y^0 \in C$ выполнено ослабленное условие Мангасаряна-Фромовица (RMFCQ), если в некоторой окрестности точки y^0 система векторов $\{\nabla h_i(y), i \in I_0 \cup I_C^0(y^0)\}$ имеет постоянный ранг. Ослабленное условие Мангасаряна-Фромовица имеет достаточно общий характер и выполняется, если имеет место условие регулярности Мангасаряна-Фромовица или какое-либо из условий регулярности, предложенных в работах [15–18].

Известно [13], что необходимое условие SSONC выполняется в точке локального минимума задачи (NLP), если в этой точке выполнено условие постоянного ранга [18] или ослабленное условие постоянного ранга RCRCQ [13].

Целью данной статьи является получение более общих по сравнению с [13] условий, которые позволяют обеспечить справедливость необходимых условий оптимальности второго порядка в усиленной форме.

Положим $I_D(y^0) = \{i \in I(y^0) \mid \langle \nabla h_i(y^0), \bar{y} \rangle = 0, \forall \bar{y} \in D_C(y^0)\}$, $I_{\#}(y^0) = I(y^0) \setminus I_D(y^0)$.

Определение 1. Будем говорить, что в точке $y^0 \in C$ выполнено условие критической регулярности, если $\text{rank}\{\nabla h_i(y), i \in I_0 \cup I_D(y^0)\} = \text{const}$ для всех y из некоторой окрестности точки y^0 .

Отметим, что введенное условие критической регулярности всегда выполняется, если выполнено условие постоянного ранга [18] или ослабленное условие постоянного ранга [13].

Также можно видеть структурную схожесть условия критической регулярности с ослабленным условием Мангасаряна-Фромовица (RMFCQ). Однако данное условие критической регулярности является значительно более жестким по сравнению с RMFCQ.

Следующая теорема обобщает аналогичные результаты [12, 13]. Ее доказательство следует схеме, предложенной в [13] с учетом специфики условия критической регулярности.

Теорема 2. Пусть в точке $y^0 \in C$, являющейся решением задачи (NLP), выполнено условие критической регулярности и $\Lambda(y^0) \neq \emptyset$. Тогда в данной точке выполняется условие SSONC.

Заключение

В статье доказываются необходимые условия оптимальности второго порядка для задачи математического программирования. Результаты обобщают необходимые условия оптимальности, полученные другими авторами.

SECOND ORDER OPTIMALITY CONDITIONS

A.E. LESCHOV, L.I. MINCHENKO

Abstract

Nonlinear programming problems are studied. Necessary second order optimality conditions are proved under minimal assumptions about constraints.

Список литературы

1. Габасов Р.Ф., Кириллова Ф.М. Методы оптимизации. Минск, 1981.
2. Гороховик В.В. Конечномерные задачи оптимизации. Минск, 2007.

3. *Luderer B., Minchenko L., Satsura T.* Multivalued analysis and nonlinear programming problems with perturbations. Dordrecht. 2002.
4. *Andreani R., Martinez J.M., Schuverdt M.L.* // Optimization. 2007. Vol. 56. P. 529–542.
5. *Bertsekas D.P.* Nonlinear Programming. Massachusetts. 1999.
6. *Fletcher R.* Practical Methods of Optimization. London. 1987.
7. *Nocedal J., Wright S.J.* Numerical Optimization. New York, 1999.
8. *Guo L., Lin G.H., Ye J.J.* // SIAM J. Optim. 2012. Vol. 22. P. 1151–1176.
9. *Hu X.M., Ralph D.* // J. Optim. Theory and Appl. 2004. Vol. 123. P. 365–390.
10. *Izmailov A.F., Solodov M.V.* // SIAM J. Optim. 2008. Vol. 19. P. 1003–1027.
11. *Vaccari A., Trad A.* // SIAM J. Optimization. 2004. № 15. P. 394–408.
12. *Andreani R., Eshague C.E., Schverdt M.L.* // J. Optimization Theory and Appl. 2010. № 146. P. 255–266.
13. *Minchenko L., Stakhovski S.* // SIAM Journal on Optimization. 2011. Vol. 21, № 1. P. 314–332.
14. *Andreani R., Martinez J.M., Schuverdt M.L.* // Optimization. 2007. № 56. P. 529–542.
15. *Минченко Л.И., Стаховский С.М.* // Докл. БГУИР. 2010. № 8. С. 104–109.
16. *Andreani R., Haeser G., Schuverdt M.L. et. al.* // SIAM Journal on Optimization. 2012. Vol. 22, № 3. P. 1109–1125.
17. *Mangasarian O.L., Fromovitz S.* // J. Mathematical Analysis and Appl. 1967. № 17. P. 37–47.
18. *Janin R.* // Mathematical Programming Study 1984. № 21. P. 110–126.

УДК 621.785.061:628.518

ВЛИЯНИЕ ТЕРМООБРАБОТКИ ПОРОШКА ШУНГИТА В ВАКУУМЕ НА ЭКРАНИРОВАНИЕ ЭЛЕКТРОМАГНИТНОГО ИЗЛУЧЕНИЯ

Е.С. БЕЛОУСОВА, Л.М. ЛЫНЬКОВ, В.Т. СЕНЮТЬ, Е.А. КРИШТОПОВА

*Белорусский государственный университет информатики и радиоэлектроники
П. Бровки, 6, Минск, 220013, Беларусь*

Поступила в редакцию 22 октября 2014

Представлены результаты исследования изменения структуры, химического состава и коэффициентов отражения и передачи электромагнитной энергии порошкообразного шунгита после термической обработки в вакууме при температуре 900 °С в течение 2 ч. Рентгенодифракционное исследование порошка шунгита после термической обработки показало образование ортоклаза и сохранение модификаций углерода. Выявлено увеличение коэффициента передачи электромагнитного излучения на 12 дБ в диапазоне 8... 12 ГГц.

Ключевые слова: вакуум, коэффициент передачи электромагнитного излучения, коэффициент отражения электромагнитного излучения, термическая обработка, шунгит.

Введение

В настоящее время ввиду активного использования радиосвязи и различных систем передачи и дистанционной обработки информации, телевидения, мобильной связи, радиолокации и радионавигации, становится актуальной разработка конструкций экранов и поглотителей, обладающих способностью отражать или поглощать электромагнитное излучение (ЭМИ) в определенном диапазоне частот. Одним из перспективных веществ для синтеза материалов с управляемыми электромагнитными свойствами является шунгит.

Шунгитовую структуру определяют как некристаллическую метастабильную неграфитируемую, глобулярную, фуллереноподобную. Главным элементом этой структуры является глобула с размерами 100–300 Å. Весьма актуальным представляется направление изучения изменений свойств шунгита при различных режимах термической обработки. Углерод в составе шунгитовых пород при разных температурах переходит в другие модификации (кристаллическая, аморфная и др.).

В [1] показано, что термообработка шунгита в инертной атмосфере не приводит к изменению структурных параметров шунгитового углерода и не влияет на удельную поверхность до температуры 1000 °С. Исследования, представленные в [2–4], показывают, что при высокотемпературной обработке в восстановительной (NH₃) и окислительной (CO₂) средах с последующей термобарической обработкой при 1000 °С и 1,5 ГПа наблюдается слияние глобул с образованием более крупных, увеличение размеров кристаллитов шунгитового углерода (с 0,1...0,3 до 5...10 мкм) и частичное разрушение шунгитовых глобул с графитизацией. Результаты рентгенодифракционного исследования порошка шунгита, подверженного термической обработке при температуре 900 °С в течение 2 ч в среде с добавлением хлорида аммония представлены в [5]. После термической обработки в порошке шунгита углерода обнаружено не было, образовался ортоклаз (21,5 %).

Цель работы – исследование химического состава и экранирующих ЭМИ свойств шунгита до и после термической обработки при температуре 900 °С в течение 2 ч в вакууме.

Методика эксперимента

Для исследования влияния высокой температуры на свойства порошкообразного шунгита предложена методика проведения эксперимента, которая включает следующие этапы:

- 1) приготовление порошка шунгита:
 - помол шунгита с размером фракций не более 20 мкм;
 - промывка порошка шунгита в дистиллированной воде;
 - сушка порошкообразного шунгита при температуре 50 °С в сушильном шкафу;
- 2) термообработка:
 - помещение порошка шунгита в тигель из инертного материала;
 - помещение тигля с порошком шунгита в печь;
 - термическая обработка в вакууме в течение 2 ч при температуре 900 °С;
 - после окончания термообработки печь должна остыть, затем образец изымается из печи и охлаждается в течение суток;
- 3) исследование свойств порошка шунгита после термической обработки:
 - электронно-микроскопический анализ порошка шунгита, подверженного термической обработке;
 - рентгенодифракционное исследование порошка шунгита на установке ДРОН-3М с использованием $\text{CuK}\alpha$ -излучения (длина волны $\lambda=1,5417737 \text{ \AA}$) и графитового фильтра в диапазоне углов от 15° до 100°;
 - исследование экранирующих электромагнитное излучение характеристик порошков шунгита, подверженных термической обработке (измерение коэффициента отражения и передачи) с помощью панорамного измерителя ослабления и КСВН Я2Р-67 с ГКЧ-61.

Результаты и их обсуждение

В результате термической обработки вес образцов уменьшился на 32 %, что свидетельствует об удалении органических компонентов, кристаллической воды и адсорбированных газов, содержащихся в порах материала.

Результаты рентгенодифракционного исследования порошка шунгита после термообработки представлены на дифрактограмме (рис. 1). Определение химического состава полученного вещества производилось с помощью программы для идентификации рентгенодифракционных максимумов веществ – МАТСН!, которая позволила определить фазы в образце путем сравнения его дифракции с эталонными дифрактограммами веществ, которые находятся в базе данных "Crystallography Open Database" (COD). Также МАТСН! позволяет провести качественный анализ содержания выбранных веществ в порошке шунгита. Были определены фазы веществ в порошке шунгита до и после термообработки, в таблице представлены сравнения примерной массовой доли веществ в порошке исходного материала и материала, подверженного термообработке.

В результате рентгенодифракционного анализа можно сделать вывод об увеличении в химическом составе вещества углерода (С) и фуллерена (C_{60}) на 9,8 и 6,4 % соответственно, что говорит о сохранении углеродом кристаллической формы. При воздействии температуры 900 °С α -кварц (SiO_2) взаимодействует с различными оксидами, что приводит к образованию силикатов, вследствие этого значительно уменьшается содержание α -кварца (SiO_2) на 36 %, и увеличивается процентное содержание таких слюд, как пеллиит (6,6 %) и кальсилит (2,2 %), а также образовывается ортоклаз с процентным содержанием 13,8 %, что свидетельствует о взаимодействии данных веществ с диоксидом кремния (SiO_2) и мусковитом ($\text{KAl}_3\text{Si}_3\text{O}_{11}\cdot\text{H}_2\text{O}$), который не был обнаружен в шунгите после термообработки. При температуре выше 850 °С в мусковите ($\text{KAl}_3\text{Si}_3\text{O}_{11}\cdot\text{H}_2\text{O}$) удаляется вода и он переходит в ортоклаз ($\text{K}_2\text{O}\cdot\text{Al}_2\text{O}_3\cdot 6\text{SiO}_2$).

Как видно из таблицы, значительно уменьшилось процентное содержание рутила и анатаза (TiO_2). Анатаз, подверженный температуре более 800 °С, гидролизуется по месту остаточных гидроксильных групп на поверхности по схеме $\text{Ti-OH}\rightarrow\text{Ti}^++\text{OH}^-$ и заряжается положительно, а рутил – по схеме $\text{Ti-OH}\rightarrow\text{Ti}^--\text{OH}^-$ и заряжается отрицательно [6]. Также уменьшается содержание магнетита на 3,7 %, что объясняется выделением оксидов железа при температуре более 500–600 °С, в соответствии с чем увеличивается процентное содержание

карбоната железа (FeCO_3) (на 10,7 %). Температура плавления оксида циркония (ZrO_2) составляет 2900 °С, поэтому увеличение его процентного содержания, как и углерода (С) и фуллерена (C_{60}), можно объяснить перераспределением содержания других элементов в структуре шунгита.

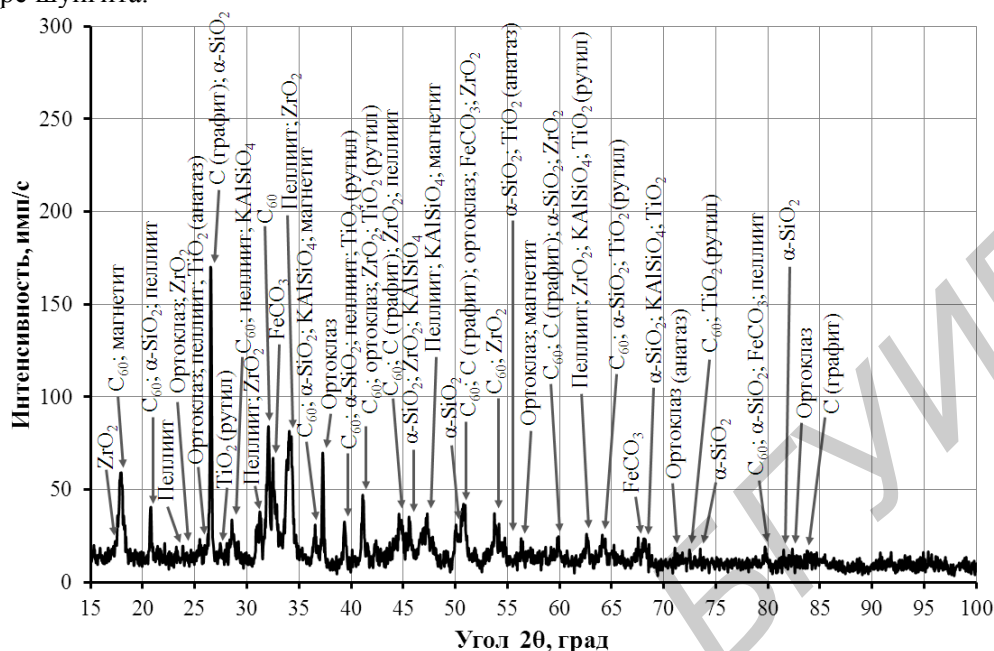


Рис. 1. Дифрактограмма шунгита, подверженного термообработке в вакууме при температуре 900 °С в течение 2 ч

Результаты рентгенофазового анализа шунгита, подверженного термообработке в вакууме при температуре 900 °С в течение 2 ч

Название фазы	Химическая формула	Номер записи в базе данных COD	Содержание в шунгите, масс. %	Содержание в шунгите после термообработки, масс. %
Ортоклаз	$\text{K}_2\text{O} \cdot \text{Al}_2\text{O}_3 \cdot 6\text{SiO}_2$	96-900-0312	–	13,8
Диоксид кремния (α-кварц)	SiO_2	96-101-1173	50,1	14,1
Углерод (графит)	C	96-901-2231	11,3	21,1
Фуллерен	C_{60}	96-901-1581	18,8	25,2
Мусковит (слюда)	$\text{KA}_3\text{Si}_3\text{O}_{11} \cdot \text{H}_2\text{O}$	96-101-1050	7,7	–
Оксид циркония	ZrO_2	96-900-7449	1,1	3,4
Пеллиит	$\text{CaMg}_{0,22}\text{Ba}_2\text{Al}_{0,2}\text{Fe}_{1,48}\text{Mn}_{0,18}\text{Zn}_{0,12}\text{Si}_{5,8}\text{O}_{17}$	96-900-0483	3,0	6,6
Карбонат железа	FeCO_3	96-500-0037	0,5	11,2
Магнетит	$\text{Ca}_{0,01}\text{Mg}_{0,06}\text{Al}_{0,05}\text{Fe}_{2,1}\text{Mn}_{0,02}\text{V}_{0,01}\text{Ti}_{0,74}\text{Si}_{0,01}\text{O}_4$	96-900-4157	4,4	0,8
Кальсилит	KAlSiO_4	96-900-9435	0,2	2,2
Рутил	TiO_2	96-900-7433	1,2	0,8
Анализ	TiO_2	96-900-8214	1,7	0,8

На электронно-микроскопическом снимке шунгита (рис. 2, а) видны затемненные участки поверхности, представляющие углерод, и светлые вкрапленности мелких рассеянных зерен белого и серого оттенков (кварц с примесью слюды и карбонатов). На рис. 2, б светлых зерен значительно больше, чем черной массы, что говорит об увеличении в составе шунгита в результате термообработки в вакууме содержания слюды и карбонатов за счет перераспределения α-кварца (SiO_2), размер частиц которых увеличивается до 20–30 мкм, за счет этого структура шунгита приобретает более шероховатую и разрыхленную поверхность.

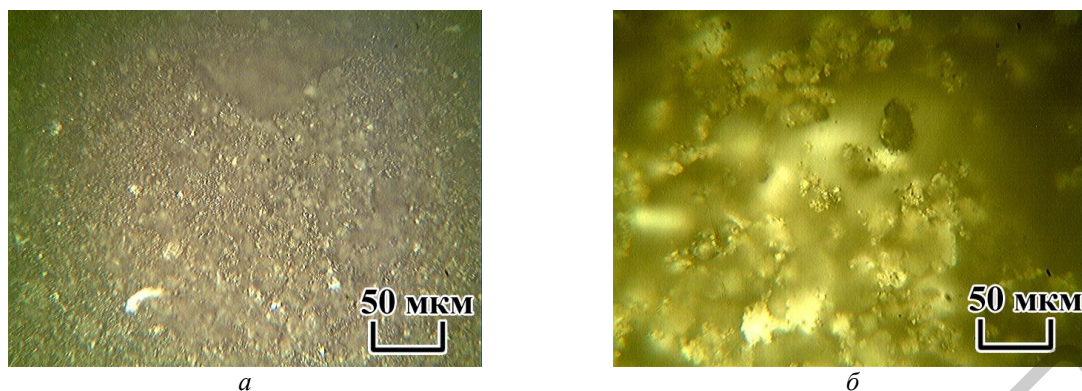


Рис. 2. Электронно-микроскопический снимок поверхности порошкообразного шунгита: *a* – исходного; *б* – шунгита, подверженного термообработке в вакууме при температуре 900 °С в течение 2 ч

Исследование экранирующих ЭМИ свойств порошка шунгита основывалось на изучении изменений коэффициентов отражения и передачи ЭМИ в диапазон частот 8...12 ГГц, для чего использовался панорамный измеритель ослабления и коэффициента стоячей волны по напряжению (КСВН) Я2Р-67 с ГКЧ-61 и волноводным трактом, который обеспечивает выделение и детектирование уровней падающей и отраженной волн электромагнитного излучения, прошедших и отраженных от образца. Калибровка оборудования перед началом измерений экранирующих характеристик производилась по стандартной методике [7]. Коэффициент отражения определяется через КСВН измерительного тракта с образцом в режимах согласованной нагрузки и короткого замыкания (с установленным металлическим отражателем), связь коэффициента отражения и КСВН, а также ослабления ЭМИ и коэффициента передачи представлена в [8]. Частотные зависимости коэффициента отражения и передачи ЭМИ в диапазоне 8...12 ГГц представлены на рис. 3 (1 – для исходного порошка шунгита; 2 – для порошка шунгита, подверженного термообработке в вакууме при температуре 900 °С в течение 2 ч).

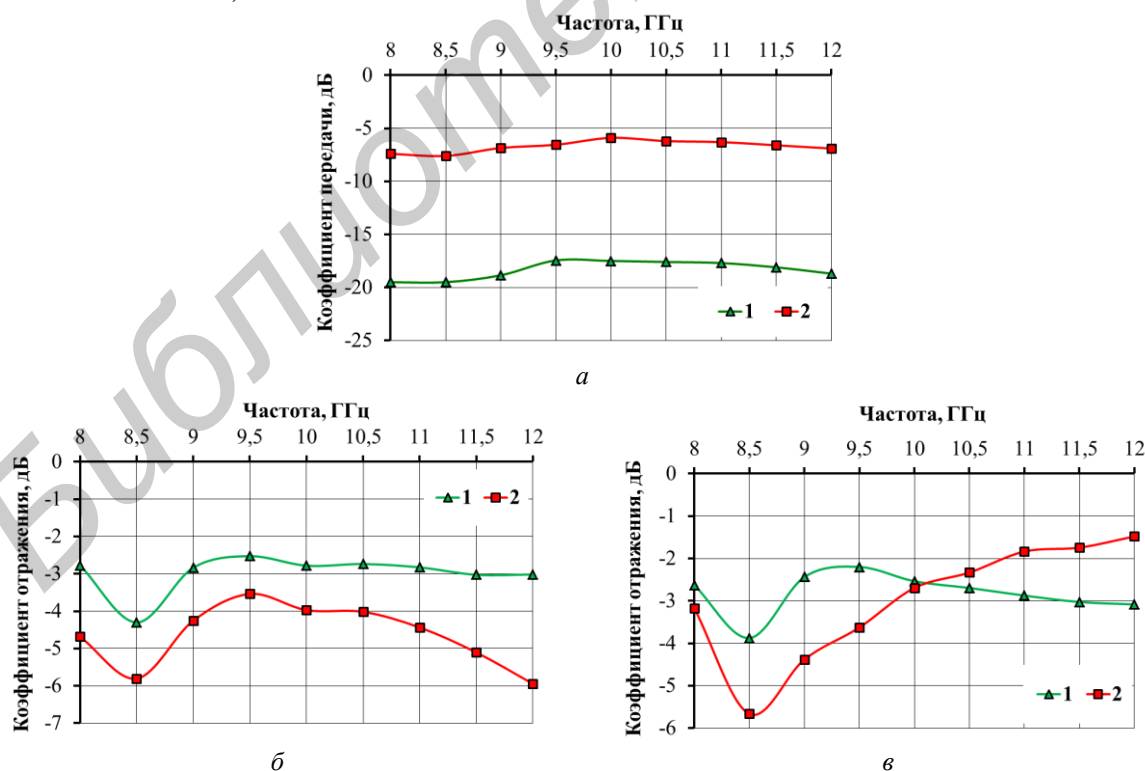


Рис. 3. Графики частотной зависимости коэффициента передачи (*a*), коэффициента отражения, измеренного в режиме согласованной нагрузки (*б*), коэффициента отражения, измеренного в режиме короткого замыкания (*в*) в диапазоне частот 8...12 ГГц

В результате проведенных измерений было обнаружено, что коэффициент передачи при термообработке порошка шунгита увеличивается на 11... 12 дБ (рис. 3, а), что объясняется изменением структуры шунгита, возникновением крупных частиц слюд. Коэффициент отражения снижается на 2 дБ в режиме согласованной нагрузки (рис. 3, б) и короткого замыкания (рис. 3, в), за счет взаимодействия α -кварца (SiO_2) с различными оксидами и образования новых силикатов (ортоклаз). Незначительное изменение коэффициента отражения ЭМИ происходит за счет сохранения кристаллической формы углерода в составе порошка шунгита, подверженного термообработке в вакууме. Для частотных характеристик коэффициентов отражения наблюдается резонансный эффект на частоте 8,5 ГГц, который может быть объяснен содержанием металлов в структуре шунгита и их влияния на отражение ЭМИ.

Заключение

Установлено, что шунгит, подверженный термообработке в вакууме при температуре 900 °С в течение 2 ч, может быть использован для создания композиционных материалов на основе полярных и неполярных полимеров. Так как за счет реакций, протекающих в шунгите при воздействии высокой температуры, увеличивается концентрация слюды (пеллит, кальсилит), образуются новые силикаты (ортоклаз), а также сохраняется кристаллическая форма углерода и его процентное содержание составляет 46,3 %, что обеспечивает уменьшение коэффициента отражения ЭМИ на 1,5... 2 дБ на частотах 8–12 ГГц. Структура шунгита становится разрыхленной и шероховатой после термообработки в вакууме, по сравнению с плотной кристаллической структурой исходного порошка шунгита, что влияет на уменьшение коэффициента передачи ЭМИ на 11...12 дБ.

INFLUENCE OF HEAT TREATMENT IN VACUUM ON SHUNGITE SHIELDING PROPERTIES

E.S. BELOUSOVA, L.M. LYNKOU, V.T. SENYUT', E.A. KRISHTOPOVA

Abstract

The research results of structure, chemical composition, reflection and transmission coefficients of powder shungite after heat treatment at 900 °C for 2 h are presented. According X-ray diffraction study, orthoclase was formed, modifications of carbon saved. Transmission coefficient increased in 12 dB in the range 8–12 GHz.

Список литературы

1. Касаточкин В.И., Элизен В.М., Мельниченко В.М. и др. // Химия твердого топлива. 1978. № 3. С. 17–21.
2. Хейфец М.Л., Железняк В.К., Кухта С.В. и др. // Вестник ПГУ. 2012. № 4. С. 113–120.
3. Колмаков А.Г., Витязь П.А., Сенють В.Т. и др. // Сб. тез. докл. 8-й междунар. конф. «Углерод: фундаментальные проблемы науки, материаловедение, технология». Троицк, 25–28 сентября 2012 г. С. 254–256.
4. Витязь П.А., Сенють В.Т., Хейфец М.Л. и др. // Вестник ФФИ. 2011. № 3 (57). С. 80–88.
5. Белоусова Е.С., Сенють В.Т., Лыньков Л.М. // Матер. 23-ей междунар. Крымской конф. «СВЧ-техника и телекоммуникационные технологии». Севастополь, 8–13 сентября 2013 г. С. 794–795.
6. Гольдберг М.М. Сырье и полупродукты для лакокрасочных материалов. М., 1978.
7. Данилин А.А. Измерения в технике СВЧ. М., 2008.
8. Белоусова Е.С., Лыньков Л.М. // Приборы. 2014. № 7 (169). С. 28–34.

УДК 004.353.25-026.66:666.9

ВЛИЯНИЕ РАЗМЕРОВ ФРАКЦИЙ ВЛАГОСОДЕРЖАЩЕГО КЕРАМЗИТА НА ЕГО КОЭФФИЦИЕНТЫ ОТРАЖЕНИЯ И ПЕРЕДАЧИ ЭЛЕКТРОМАГНИТНОГО ИЗЛУЧЕНИЯ

С.Э. САВАНОВИЧ, Т.В. БОРБОТЬКО, В.Б. СОКОЛОВ, М.В. РУСАКОВИЧ

*Белорусский государственный университет информатики и радиоэлектроники
П. Бровка, 6, Минск, 220013, Беларусь**Поступила в редакцию 29 октября 2014.*

Приведены результаты исследований коэффициентов отражения и передачи влагосодержащего керамзита с размерами фракций до 4 мм и 10...20 мм в диапазоне частот 0,7...17 ГГц.

Ключевые слова: экранирование, электромагнитное излучение, керамзит, размер фракций, коэффициент отражения, коэффициент передачи.

Введение

В настоящее время для обработки информации, в том числе ограниченного распространения, используют средства вычислительной техники. Такие устройства являются источником побочных электромагнитных излучений (ПЭМИ) [1]. Одним из способов защиты информации от утечки по электромагнитным каналам является экранирование, позволяющее снизить уровень ПЭМИ за счет применения радиоэкранирующих материалов. Для экранирования средств вычислительной техники (СВТ) и помещений, где они размещаются, широко используются металлические материалы, характеризующиеся высокой эффективностью, технологичностью и вместе с тем большими стоимостью и массой. Альтернативой металлическим материалам являются композиты, однако их стоимость также значительная, при сравнимой эффективности. Таким образом, широкое использование радиоэкранирующих материалов для решения задач снижения уровня ПЭМИ сдерживается в первую очередь стоимостью таких материалов.

Перспективным направлением в разработке экранов электромагнитного излучения (ЭМИ) представляется их создание на основе влагосодержащих материалов [2]. Радиоэкранирующие свойства конструкций экранов определяются их влагосодержанием, примесями и их концентрациями в водном растворе, используемом для пропитки, структурой и свойствами матрицы, заполняемой таким раствором. В работах [3–6] были предложены конструкции экранов ЭМИ на основе влагосодержащих капиллярно-пористых (волокнистых, тканых и нетканых, порошкообразных и гранулированных) материалов с различными растворными наполнителями. Известны конструкции влагосодержащих экранов ЭМИ из силикагелевых и бентонитовых материалов, пропитанных различными водными растворами [7]. В объеме каждого влагосодержащего элемента конструкции, независимо от типа используемого материала подложки, находился растворный наполнитель.

Существенными недостатками подобных конструкций являются большой вес, обусловленный в основном высоким влагосодержанием растворного наполнителя до 70 %, а также необходимость стабилизации влагосодержания для сохранения радиоэкранирующих характеристик. Для герметизации конструкций экранов ЭМИ в некоторых случаях применялись полимерные плёнки и связующие вещества, позволяющие значительно снизить испарение влаги [8].

Одним из решений по минимизации массы влагосодержащих конструкций экранов ЭМИ является применение пористых материалов с малым удельным весом, к которым относится керамзит – лёгкий пористый строительный материал, получаемый путем обжига глины. Он характеризуется малой плотностью, относительно малым весом, высокой удельной пористостью, невысокой стоимостью по сравнению с рассмотренными в [3–7] материалами ЭМИ. Ячеистая структура керамзита позволяет инкорпорировать в поры такого материала влагосодержащий наполнитель, и изменять коэффициенты отражения и передачи конструкций экранов ЭМИ, выполненных на его основе, за счет варьирования размеров его фракций, влагосодержания и концентрации растворного наполнителя.

Таким образом, целью данной работы являлось установление закономерностей влияния размеров фракций влагосодержащего керамзита на его коэффициенты отражения и передачи в диапазоне частот 0,7...17 ГГц.

Методика эксперимента

Для проведения исследования были изготовлены конструкции экранов ЭМИ, выполненные в виде модулей с плоской формой поверхности, внутренний объем которых заполнялся керамзитом с размерами фракций до 4 мм и 10...20 мм (толщина слоя керамзита в обоих случаях составляла 30 мм). Перед заполнением модулей керамзит пропитывался 10 % водным раствором хлорида натрия (NaCl) при температуре 18...20 °С.

Выбор хлорида натрия обусловлен тем, что его водный раствор является электролитом. Растворимость в воде при 20 °С для NaCl составляет 36,0 г на 100 г воды, диэлектрическая проницаемость 23. Содержание водного раствора NaCl в исследуемых конструкциях экрана ЭМИ оценивалось гравиметрически, относительная погрешность измерений составляла ± 2 %. Влагосодержание керамзита с размерами фракций до 4 мм и 10...20 мм, пропитанного 10 % водным раствором NaCl, варьировалось в пределах 22...23 % и 23...24 % соответственно [9].

Для измерения коэффициентов отражения и передачи исследуемых конструкций экранов ЭМИ в диапазоне частот 0,7...17 ГГц использовался панорамный измеритель коэффициентов передачи и отражения SNA 0,01–18, работающий по принципу раздельного выделения и непосредственного детектирования уровней падающей и отраженной волн, и антенны П6 - 23М. Измерения проводились по методике, приведенной в [10].

Результаты и их обсуждение

На основе анализа полученных результатов установлено, что в диапазоне частот 0,7...17 ГГц для конструкций экранов ЭМИ, выполненных на основе влагосодержащего керамзита, значения коэффициента отражения варьируются в пределах – 0,1...–18 дБ (для размера фракций до 4 мм), в пределах – 0,1...–17 дБ (для размера фракций 10...20 мм) (рис. 1, 2).

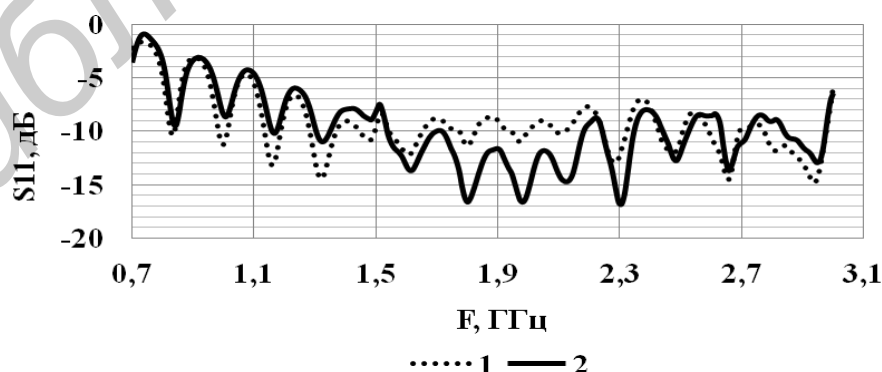


Рис. 1. Частотные зависимости (диапазон частот 0,7 ... 3 ГГц) коэффициента отражения конструкций экранов ЭМИ, выполненных на основе влагосодержащего керамзита, с размерами фракций: 1 – до 4 мм, 2 – 10...20 мм

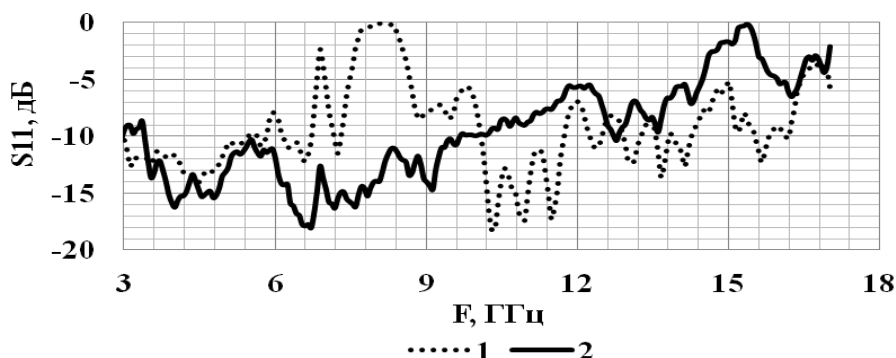


Рис. 2. Частотные зависимости (диапазон частот 3 ... 17 ГГц) коэффициента отражения конструкций экранов ЭМИ, выполненных на основе влагосодержащего керамзита, с размерами фракций: 1 – до 4 мм, 2 – 10 ... 20 мм

Показано, что уменьшение размера фракций влагосодержащего керамзита в конструкциях экранов ЭМИ позволяет снизить значения коэффициента отражения на частотах 0,7...1,5 ГГц, 2,4...3,4 ГГц и 10...17 ГГц до 10 дБ. Увеличение размера фракций влагосодержащего керамзита в исследуемых конструкциях экранов ЭМИ позволяет снизить значения коэффициента отражения на частотах 1,5...2,4 ГГц, 3,4...10 ГГц до 10 дБ.

Показано, что в диапазоне частот 0,7...17 ГГц значения коэффициента передачи для конструкций экранов ЭМИ, выполненных на основе влагосодержащего керамзита варьируются в пределах – 3...–29 дБ (для размера фракций до 4 мм), и в пределах –0,1...–29 дБ (для размера фракций 10...20 мм) (рис. 3, 4).

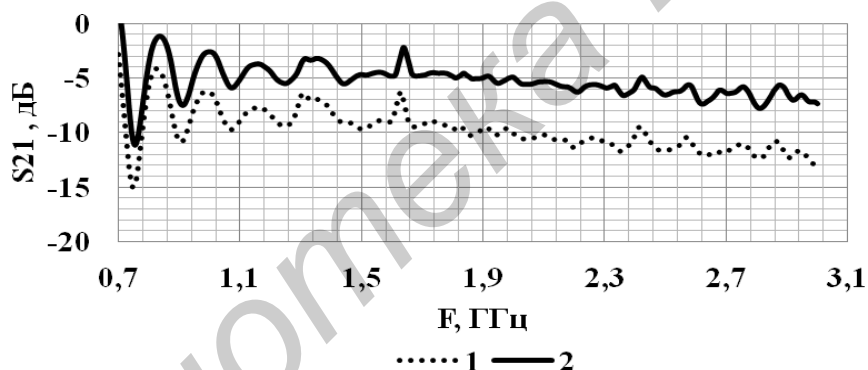


Рис. 3. Частотные зависимости (диапазон частот 0,7 ... 3 ГГц) коэффициента передачи конструкций экранов ЭМИ, выполненных на основе влагосодержащего керамзита, с размерами фракций: 1 – до 4 мм, 2 – 10...20 мм

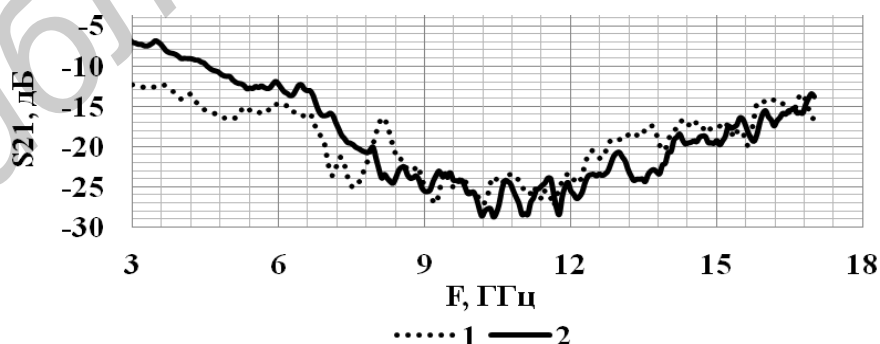


Рис. 4. Частотные зависимости (диапазон частот 3 ... 17 ГГц) коэффициента передачи конструкций экранов ЭМИ, выполненных на основе влагосодержащего керамзита, с размерами фракций: 1 – до 4 мм, 2 – 10...20 мм

Установлено, что уменьшение размера фракций влагосодержащего керамзита в конструкциях экранов ЭМИ позволяет снизить значения коэффициента передачи на частотах 0,7 ... 8 ГГц до 7 дБ. Увеличение размера фракций влагосодержащего керамзита позволяет

снизить значения коэффициента передачи на частотах 8 ... 8,4 ГГц, 10...11 ГГц и 12...15 ГГц до 8 дБ.

При закреплении исследуемых конструкций экранов ЭМИ на металлической подложке, значения коэффициента отражения варьируются в пределах – 0,1...–19 дБ (для размера фракций до 4 мм), и в пределах – 0,1...–23 дБ (для размера фракций 10...20 мм) в диапазоне частот 0,7...17 ГГц.

Введение в воду примесей влияет на структуру воды и позволяет изменять ее характеристики: электропроводность, диэлектрическую проницаемость, вязкость. Еще в большей степени структура воды изменяется вблизи твердой пористой поверхности керамзита при образовании связанной воды. Свойства связанной воды, находящейся в пористой структуре керамзита в виде тонких граничных слоев вблизи твердой поверхности, существенно отличаются от свойств свободной воды. Эти отличия объясняются изменениями структуры связанной воды вблизи поверхности, сетки межмолекулярных водородных связей в ее структуре под влиянием поля поверхностных сил [11]. Введение в воду 10 % хлорида натрия повышает проводимость водного раствора, одновременно с этим его динамическая вязкость изменяется от $1,002 \cdot 10^3 \text{ Па} \cdot \text{с}$ до $1,19 \cdot 10^3 \text{ Па} \cdot \text{с}$ [12].

Инкорпорирование в поры керамзита 10 % водного раствора электролита приводит к уменьшению толщины граничного слоя вблизи твердой поверхности керамзита и, как следствие этому, снижению значений коэффициента отражения и передачи в исследуемом диапазоне частот. Влагосодержащий керамзит размера фракций до 4 мм и 10...20 мм отличается неоднородностью структуры и размерами пор, этим обусловлено отличие его радиоэкранирующих характеристик для исследуемых конструкций экранов ЭМИ в диапазоне частот 0,7...17 ГГц.

Заключение

По результатам исследований установлено, что для снижения значения коэффициента отражения электромагнитного излучения предпочтительно использование в конструкциях экранов ЭМИ влагосодержащего керамзита с размером фракций до 4 мм для диапазонов частот 0,7...1,5 ГГц, 2,4...3,4 ГГц и 10...17 ГГц, для диапазонов частот 1,5...2,4 ГГц и 3,4...10 ГГц – влагосодержащего керамзита с размером фракций 10...20 мм.

Для ослабления электромагнитных излучений предпочтительно использование в конструкциях экранов влагосодержащего керамзита с размером фракций до 4 мм в диапазоне частот 0,7...8 ГГц, в диапазоне частот 8...8,4 ГГц и 12...15 ГГц влагосодержащий керамзит с размером фракций 10...20 мм.

Установлено, что конструкции экранов ЭМИ с плоской формой поверхности, выполненные на основе влагосодержащего керамзита, с размером фракций до 4 мм и 10...20 мм, содержащие 10 % водный раствор хлорида натрия, имеют более высокую эффективность экранирования по сравнению с аналогичными по толщине экранами из мелкодисперсных порошковых сорбентов на основе диоксида кремния и углеродосодержащих материалов и экранами с различным рельефом поверхности на основе порошкообразных торфов [13, 14].

Показано, что уменьшение размеров фракций влагосодержащего керамзита в конструкциях экранов ЭМИ позволяет обеспечить более высокую эффективность ослабления электромагнитного излучения в диапазоне частот 0,7...17 ГГц, что показывает перспективность применения влагосодержащего керамзита в конструкциях экранов ЭМИ для решения задач информационной безопасности.

INFLUENCE OF WATER-CONTAINING CLAYDITE FRACTION SIZES ON THE ELECTROMAGNETIC RADIATION REFLECTION AND TRANSMISSION

S.E. SAVANOVICH, T.V. BORBOTKO, V.B. SOKOLOV, M.V. RUSAKOVICH

Abstract

Results of investigations of the reflection coefficients and transmission moisture-expanded clay fractions with dimensions of up to 4 mm and 10 ... 20 mm in the frequency range 0.7 ... 17 GHz.

Список литературы

1. Михалев Л. А., Новичков И. С., Сергеев и др. // Системы безопасности связи и телекоммуникаций. 2001. № 39. С. 50–53.
2. Лыньков Л.М., Богуш В.А., Борботько Т.В. и др. // Докл. БГУИР. 2004. №3 (7). С. 152–157.
3. Богуш В.А., Зубаревич О.И., Колбун Н.В. и др. // Матер. 15-й междунар. Крымской конференции «СВЧ-техника и коммуникационные технологии». Севастополь. Украина, сентябрь 2005 г. С. 637–639.
4. Колбун Н.В. // Вестн. ПГУ Серия В. Прикладные науки. 2004. № 12. С. 30–35.
5. Казанцева Н.Е., Рывкина Н.Г., Чмутин И.А. // Радиотехника и электроника. 2003. Т. 48. № 2. С. 196–209.
6. Лыньков Л.М., Богуш В.А., Глыбин В.П. и др. Гибкие конструкции экранов электромагнитного излучения. Минск, 2000.
7. Занг Фан Ньят Формирование водосодержащих силикагелевых и бентонитовых материалов для высокостабильных экранов электромагнитного излучения: Автореф. дис. ... канд. техн. наук. Минск, 2007.
8. Колбун Н.В. Технология изготовления экранов электромагнитного излучения на основе влагосодержащих капиллярно-пористых материалов: Автореф. дис. ... канд. техн. наук. Минск, 2005.
9. Лыньков Л.М., Борботько Т.В., Колбун Н.В. и др. // Матер. II Белор.-росс. науч.-техн. конф. «Технические средства защиты информации». Нарочь, 17–21 мая 2004 г. С. 48–50.
10. Неамах М. Р., Бойправ О.В., Борботько Т.В. и др. Радиоэкранирующие модульные конструкции на основе порошкообразных материалов. Минск, 2013.
11. Королёв В.А. // Соросовский образовательный журнал. 1996. № 9. С. 79–85.
12. Краткий справочник физико-химических величин / Под ред. А.А. Равделя. СПб., 2003.
13. Пухир Г.А., Махмуд М.Ш., Лыньков Л.М. // Докл. БГУИР. 2011. № 8 (62). С. 99–102.
14. Лыньков Л.М., Столер Д.В. // Докл. БГУИР. 2013. №7 (77). С. 32–36.

УДК 621.391

ГЕОМЕТРИЧЕСКИЕ МОДЕЛИ МНОГОРАКУРСНЫХ ИЗОБРАЖЕНИЙ И ПРОЕКТИВНАЯ КОМПЕНСАЦИЯ ДВИЖЕНИЯ КАМЕРЫ

В.Ю. ЦВЕТКОВ

Белорусский государственный университет информатики и радиоэлектроники
П. Бровки, 6, Минск, 220013, Беларусь

Поступила в редакцию 24 сентября 2014

Разработаны геометрические модели формирования и предсказания многоракурсных изображений. Для предсказания предложено использовать методы проективной компенсации движения камеры. Показано, что проективная компенсация движения позволяет реализовать эффективное сжатие многоракурсных изображений.

Ключевые слова: геометрическая модель изображения, многоракурсное изображение, проективная компенсация движения камеры.

Введение

Для систем распределенного видеомониторинга, использующих множество камер (видеокамер и тепловизоров), актуальна проблема сокращения объема передаваемой видеоинформации. Наиболее эффективные методы сжатия видеоданных от одной камеры основаны на предсказании и компенсации движения [1–4]. В случае нескольких камер с перекрывающимися секторами обзора формируются многоракурсные изображения, предсказание которых возможно за счет использования некоторого одноракурсного фрагмента (опорного изображения) для формирования приближений (прогнозных изображений) других одноракурсных фрагментов (прогнозируемых изображений). Для предсказания многоракурсных изображений необходим учет проективных искажений при переходе от одного ракурса к другому. В связи с этим актуальна разработка геометрических моделей формирования и предсказания многоракурсных изображений.

Постановка задачи

Формируемое R камерами многоракурсное изображение $\{I(r)\}_{(r=0, \overline{R-1})}$ представляет собой множество перспективных проекций $I(r)$ трехмерного пространства P^3 регистрации, полученных с ракурсов r , с помощью перспективного проективного преобразования T_{pp}^3 (без учета дисторсии оптической системы камеры) в соответствии с выражением

$$I(r) = T_{pp}^3(P^3, r). \quad (1)$$

Если известна реконструкция P_R^3 пространства P^3 , то прогнозное изображение $I_p(r)$ для произвольного ракурса r может быть получено с помощью выражения

$$I_p(r) = T_{pp}^3(P_R^3, r). \quad (2)$$

Поэтому для предсказания изображения $I(r)$ по реконструкции P_R^3 достаточно иметь значение ракурса r . Однако, получение реконструкции P_R^3 связано с существенными вычислительными и временными затратами. Поэтому для реальных условий актуальна

проблема формирования прогнозного изображения $I_p(r)$ на основе некоторого опорного изображения $I_B(k)$, полученного с другого ракурса $k = r - \Delta r$, в соответствии с выражением

$$I_p(r) = T_x^2(I_B(k), \Delta r), \text{ где } \Delta r - \text{ перемещение камеры.} \quad (3)$$

При известных k и Δr данная проблема сводится к отысканию преобразования T_x^2 . Если T_x^2 известно, то для предсказания изображения $I(r)$ достаточно иметь $I_B(k)$ и Δr . Формирование прогнозного изображения $I_p(r)$ в этом случае по известному перемещению Δr камеры является процессом компенсации движения камеры.

При неизвестных ракурсах k и Δr опорное I_B и прогнозные I_p изображения связаны с помощью следующего выражения:

$$I_p = T_x^2(I_B). \quad (4)$$

Следовательно, для предсказания изображения I необходимо иметь I_B и T_x^2 . Преобразование T_x^2 может быть получено в результате сопоставления опорного I_B и прогнозируемого I изображений.

В методах сжатия MPEG-4, H.264, H.265 [3, 4] предполагается, что преобразование T_x^2 является переносом T_T^2 и его параметры определяются с помощью блочной компенсации движения (рис. 1, а). Перенос T_T^2 не учитывает изменение масштаба, поворот и перспективные искажения базового изображения (рис. 1, б). Фактически, предполагается, что пространство регистрации является плоскостью в двумерном пространстве P^2 , а смещение камеры незначительно. Это снижает эффективность использования блочной компенсации движения для предсказания многокакурсных изображений.

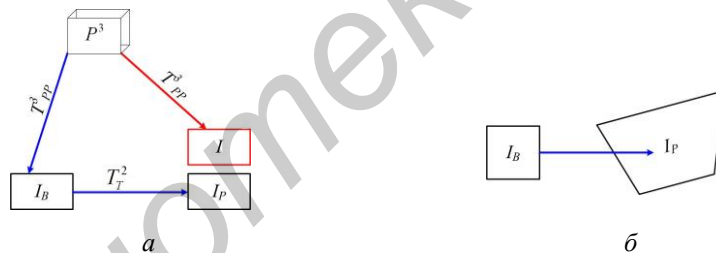


Рис. 1. Блочная компенсация движения: а – геометрическая модель; б – проективные искажения опорного изображения

Целью работы является разработка геометрических моделей формирования многокакурсных изображений и определение наиболее эффективных методов компенсации движения для их предсказания.

Геометрические модели пространства регистрации многокакурсных изображений

В зависимости от величины смещения камеры от одного ракурса к другому (значительное/незначительное) относительно расстояния до объектов трехмерное пространство P^3 регистрации может быть представлено приближенно одной или несколькими плоскостями в двумерном или трехмерном пространстве (рис. 2).

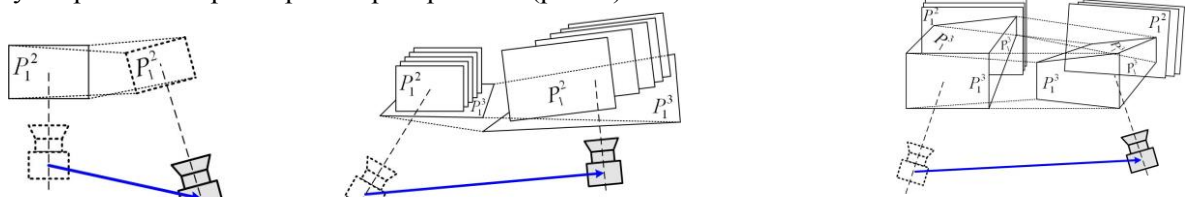


Рис. 2. Геометрические модели пространства регистрации: а – P_1^2 ; б – $P_{N/1}^{2/3}$ в – $P_{N/M}^{2/3}$

Наибольший практический интерес представляют следующие геометрические модели пространства регистрации:

- P_1^2 – плоскость в двумерном пространстве P^2 , в которой лежат все объекты (рис. 2, а);
- P_N^2 – N параллельных плоскостей P_1^2 , каждая из которых расположена в двумерном пространстве P^2 и включает равноудаленные от камеры объекты;
- P_1^3 – плоскость в трехмерном пространстве P^3 , в которой лежат все объекты;
- P_M^3 – M произвольно ориентированных плоскостей P_1^3 в трехмерном пространстве P^3 ;
- $P_{N/1}^{2/3}$ ($P_{1/1}^{2/3}$ в частном случае при $N = 1$) – N плоскостей P_1^2 и плоскость P_1^3 , перпендикулярная плоскостям P_1^2 (рис. 2, б);
- $P_{N/M}^{2/3}$ ($P_{1/M}^{2/3}$ в частном случае при $N = 1$) – N плоскостей P_1^2 и M плоскостей P_1^3 (рис. 2, в).

Модель P_1^2 представляет плоскую сцену при любом перемещении камеры, например, при аэрокосмической съемке, когда оптическая ось камеры перпендикулярна земной поверхности. Модели P_N^2 , $P_{N/1}^{2/3}$ и $P_{1/1}^{2/3}$ представляют неглубокую трехмерную сцену при относительно небольшом смещении камеры, например, при наземной съемке, когда земная или водная поверхность не попадает (P_N^2) или попадает ($P_{N/1}^{2/3}$ и $P_{1/1}^{2/3}$) в кадр. Модель P_1^3 представляет плоскую поверхность в трехмерной сцене при любом смещении камеры, например, при наземной съемке, когда земная или водная поверхность попадает в кадр, или при аэросъемке, когда положение оптической оси камеры не перпендикулярно земной поверхности. Модели P_M^3 , $P_{N/M}^{2/3}$ и $P_{1/M}^{2/3}$ представляют глубокую трехмерную сцену при значительном смещении камеры в условиях отсутствия (P_M^3) и наличия ($P_{N/M}^{2/3}$ и $P_{1/M}^{2/3}$) удаленных объектов, например, при наземной съемке в городских условиях.

Геометрические модели формирования и предсказания многокурсных изображений

Общая геометрическая модель формирования многокурсного изображения $\{I(r)\}_{(r=0, R-1)}$ показывает получение нескольких изображений с помощью перспективного проективного преобразования T_{pp}^3 согласно выражению (1). Выделяя среди этих изображений два: опорное I_B и прогнозируемое I , можно построить модель предсказания прогнозируемого изображения I . Данная модель показывает преобразование T_{pp}^3 пространства P^3 регистрации в опорное изображение I_B и, далее, преобразование T_x^2 опорного изображения I_B в прогнозное изображение I_P : $P^3 \xrightarrow{T_{pp}^3} I_B \xrightarrow{T_x^2} I_P$. Тип преобразования T_x^2 зависит от модели пространства P^3 регистрации. В случае представления пространства P^3 регистрацией моделью P_1^2 ($P^3 \rightarrow P_1^2$) или моделью P_N^2 ($P^3 \rightarrow P_N^2$) преобразование T_x^2 является преобразованием T_s^2 подобия, сохраняющим параллельность линий. При представлении пространства P^3 моделью P_1^3 или моделью P_M^3 преобразование T_x^2 является проективным преобразованием T_p^2 , сохраняющим коллинеарность точек.

Рассмотренным геометрическим моделям формирования многокурсных изображений соответствуют геометрические модели предсказания: $P_1^2 \xrightarrow{T_{pp}^3} I_B \xrightarrow{T_s^2} I_P$ (рис. 3, а); $P_N^2 \xrightarrow{T_{pp}^3} I_B \xrightarrow{\{T_s^2\}_N} I_P$ (рис. 3, б); $P_{N/1}^{2/3} \xrightarrow{T_{pp}^3} I_B \xrightarrow{\{T_s^2\}_N, T_p^2} I_P$ ($P_{1/1}^{2/3} \xrightarrow{T_{pp}^3} I_B \xrightarrow{T_s^2, T_p^2} I_P$) (рис. 3, в); $P_{N/M}^{2/3} \xrightarrow{T_{pp}^3} I_B \xrightarrow{\{T_s^2\}_N, \{T_p^2\}_M} I_P$ ($P_{1/M}^{2/3} \xrightarrow{T_{pp}^3} I_B \xrightarrow{T_s^2, \{T_p^2\}_M} I_P$) (рис. 3, г); $P_M^3 \xrightarrow{T_{pp}^3} I_B \xrightarrow{\{T_p^2\}_M} I_P$ (рис. 3, д); $P_1^3 \xrightarrow{T_{pp}^3} I_B \xrightarrow{T_p^2} I_P$ (рис. 3, е).

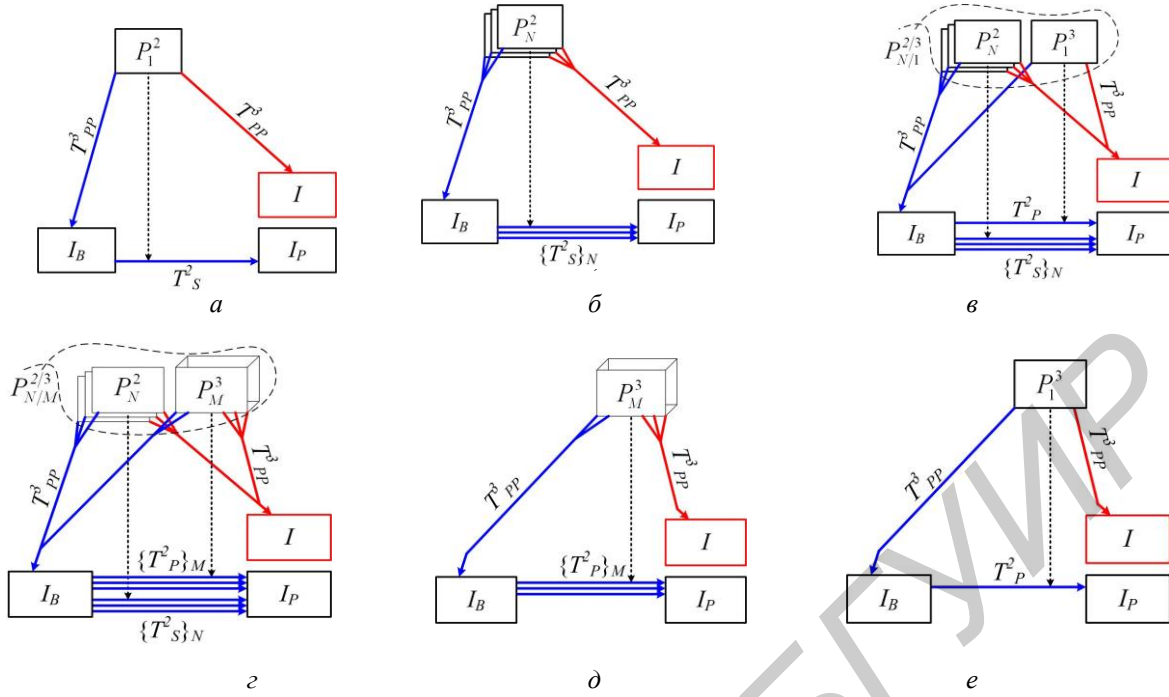


Рис. 3. Геометрические модели предсказания многокурсных изображений

Данные модели уточняются с помощью выражений

$$\begin{cases}
 T_S^2(I_B) = T_S^2(T_{PP}^3(P_1^2)) \text{ при } P^3 \rightarrow P_1^2, \\
 \bigcup_{n=0}^{N-1} T_S^2(I_B, n) = \bigcup_{n=0}^{N-1} T_S^2(T_{PP}^3(P_N^2), n) \text{ при } P^3 \rightarrow P_N^2, \\
 \bigcup_{n=0}^{N-1} T_S^2(I_B, n) \cup T_P^2(I_B) = \bigcup_{n=0}^{N-1} T_S^2(T_{PP}^3(P_{N/1}^{2/3}), n) \cup T_P^2(T_{PP}^3(P_{N/1}^{2/3})) \text{ при } P^3 \rightarrow P_{N/1}^{2/3}, \\
 T_S^2(I_B) \cup T_P^2(I_B) = T_S^2(T_{PP}^3(P_{1/1}^{2/3})) \cup T_P^2(T_{PP}^3(P_{1/1}^{2/3})) \text{ при } P^3 \rightarrow P_{1/1}^{2/3}, \\
 I_P = \begin{cases}
 \bigcup_{n=0}^{N-1} T_S^2(I_B, n) \cup \bigcup_{m=0}^{M-1} T_P^2(I_B, m) = \bigcup_{n=0}^{N-1} T_S^2(T_{PP}^3(P_{N/M}^{2/3}), n) \cup \bigcup_{m=0}^{M-1} T_P^2(T_{PP}^3(P_{N/M}^{2/3}), m) \text{ при } P^3 \rightarrow P_{N/M}^{2/3}, \\
 T_S^2(I_B) \cup \bigcup_{m=0}^{M-1} T_P^2(I_B, m) = T_S^2(T_{PP}^3(P_{1/M}^{2/3})) \cup \bigcup_{m=0}^{M-1} T_P^2(T_{PP}^3(P_{1/M}^{2/3}), m) \text{ при } P^3 \rightarrow P_{1/M}^{2/3}, \\
 \bigcup_{m=0}^{M-1} T_P^2(I_B, m) = \bigcup_{m=0}^{M-1} T_P^2(T_{PP}^3(P_M^3), m) \text{ при } P^3 \rightarrow P_M^3, \\
 T_P^2(I_B) = T_P^2(T_{PP}^3(P_1^3)) \text{ при } P^3 \rightarrow P_1^3,
 \end{cases}
 \end{cases} \quad (5)$$

где \cup – операция объединения фрагментов изображения.

Методы проективной компенсации движения камеры

Для предсказания многокурсных изображений $\{I(r)\}_{(r=0, R-1)}$ предлагается использовать методы проективной компенсации движения камеры, основанные на использовании в общем случае проективного преобразования T_x^2 опорного изображения I_B , полученного в некотором исходном ракурсе, для формирования прогнозного изображения I_P , соответствующего другому ракурсу, согласно выражениям (3) или (4). В зависимости от используемой геометрической модели пространства P^3 регистрации методы проективной компенсации движения камеры можно разделить на плоскостные (для пространств P_1^2 и P_1^3) и объектные

(для пространств P_N^2 и P_M^3), двумерные (для пространств P_1^2 и P_N^2) и трехмерные (для пространств P_1^3 и P_M^3). Для реализации методов проективной компенсации движения камеры необходимо разделить пиксели опорного изображения I_B по плоскостям (объектная декомпозиция), установить соответствие между плоскостями опорного и прогнозируемого изображений, определить тип и параметры преобразования для каждой плоскости.

При плоскостной компенсации движения камеры пиксели опорного изображения I_B рассматриваются в одной плоскости, параллельной [5] (двухмерная плоскостная компенсация движения) или произвольно ориентированной [6] (трехмерная плоскостная компенсация движения) относительно плоскости регистрирующей системы камеры. Двухмерная плоскостная компенсация движения камеры основана на преобразовании T_S^2 подобия, трехмерная плоскостная компенсация движения – на проективном преобразовании T_P^2 . Для примера на рис. 4, а представлен фрагмент видеопоследовательности из 15 кадров, сформированной с борта беспилотного летательного аппарата, для сжатия которой использованы кодеки H.264, MJPEG-2000, MJPEG и кодек FMC на основе двухмерной плоскостной компенсации движения камеры. В качестве опорного изображения для двухмерной плоскостной компенсации движения камеры использован фрагмент фотоплана (рис. 4, б).

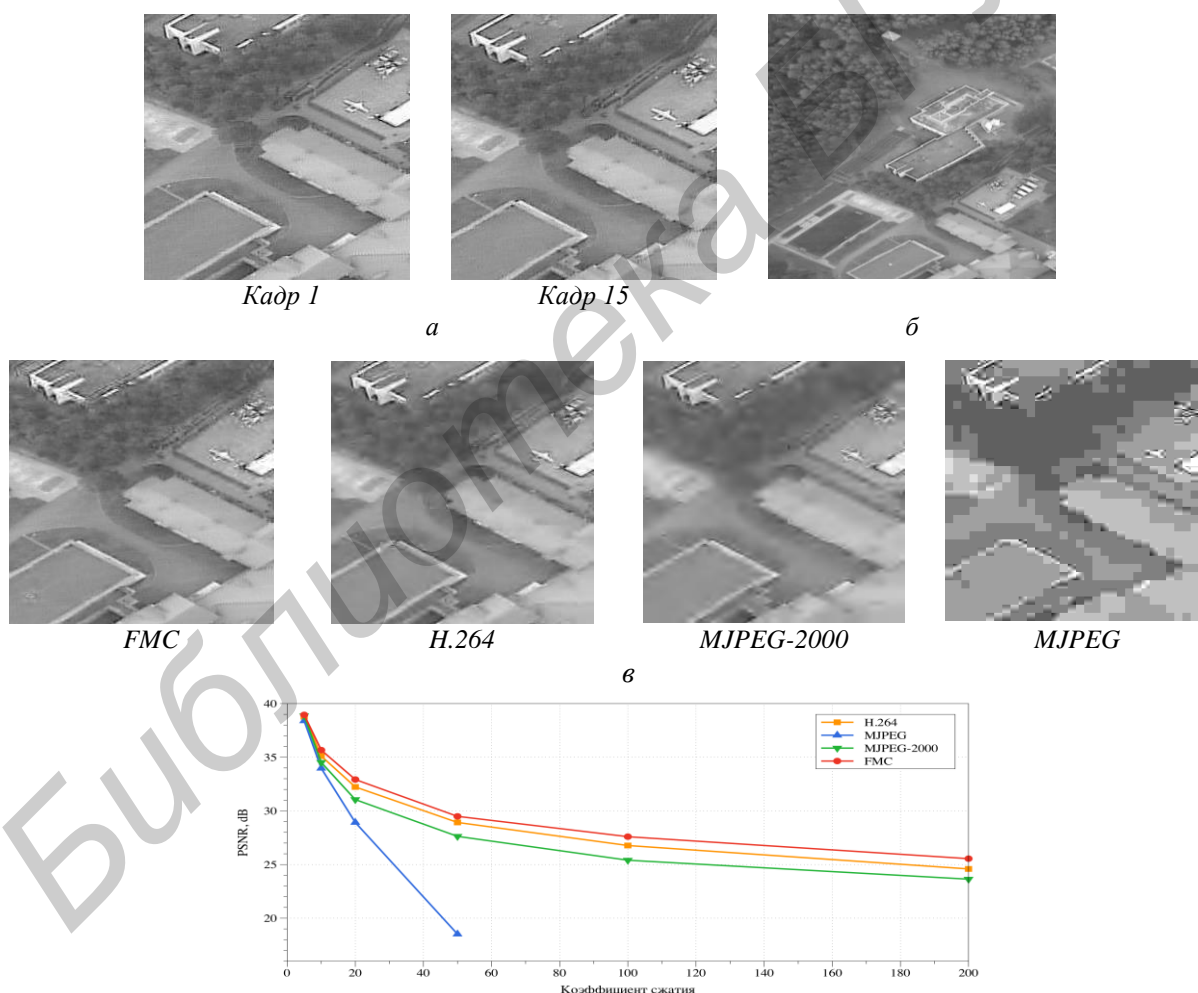


Рис. 4. Оценка эффективности двухмерной плоскостной компенсации движения камеры: а – видеопоследовательность (15 кадров); б – фотоплан; в – восстановленный 10-й кадр; з – зависимость PSNR от коэффициента сжатия

Анализ кадров, восстановленных после сжатия в 50 раз различными кодеками (рис. 4, з), показывает, что двухмерная плоскостная компенсация движения камеры обеспечивает лучшее качество. Экспериментально установлено, что при коэффициентах сжатия

больше 10 раз кодек на основе двумерной плоскостной компенсации движения камеры обеспечивает увеличение пикового отношения сигнал-шум (PSNR) на 1дБ по сравнению с кодеком H.264 и на 2 дБ по сравнению с кодеком MJPEG-2000 (рис. 4, з).

При объектной компенсации движения камеры пиксели опорного изображения I_B рассматриваются в нескольких плоскостях, параллельных (двухмерная объектная компенсация движения) или произвольно ориентированных (трехмерная объектная компенсация движения) относительно плоскости регистрирующей системы камеры [6]. Двухмерная объектная компенсация движения камеры основана на преобразованиях T_S^2 подобия, трехмерная объектная компенсация движения – на проективных преобразованиях T_P^2 , применяемых к каждой плоскости независимо. Для примера на рис. 5 представлены опорный (рис. 5, а) и прогнозируемый (рис. 5, б) кадры видеопоследовательности, сформированные с помощью видеокамеры, установленной на движущемся автомобиле. Для этих кадров с помощью методов двумерной объектной и трехмерной плоскостной компенсации движения камеры сформированы прогнозный (рис. 5, в) и разностный (рис. 5, з) кадры. Для сравнения на рис. 5, д приведен разностный кадр, сформированный с помощью кодека H.264. Экспериментально установлено, что выигрыш в пиковом отношении сигнал-шум предсказания от использования проективной компенсации движения достигает 1,5 дБ по сравнению с блочной компенсацией движения, используемой в кодеке H.264.



Рис. 5. Оценка эффективности объектной и плоскостной компенсации движения камеры: а – опорный кадр; б – прогнозируемый кадр; в – прогнозный кадр, сформированный с помощью объектной и плоскостной компенсации движения; з – разностный кадр, сформированный с помощью объектной и плоскостной компенсации движения; д – разностный кадр, сформированный с помощью блочной компенсации движения кодеком H.264

Заключение

Предложены геометрические модели, представляющие пространство регистрации камеры одной или несколькими плоскостями в двумерном или трехмерном пространстве. На их основе разработаны геометрические модели формирования и предсказания многокурсных изображений. Показано, что эффективным средством предсказания является проективная компенсация движения камеры, основанная на использовании в общем случае проективного преобразования опорного изображения, полученного в некотором исходном ракурсе, для формирования прогнозного изображения, соответствующего другому ракурсу. Предсказание на основе проективной компенсации движения камеры позволяет реализовать эффективные методы сжатия многокурсных изображений. Установлено, что выигрыш в пиковом отношении сигнал-шум предсказания от использования проективной компенсации движения достигает 1,5 дБ по отношению к блочной компенсацией движения, используемой в кодеке

H.264. При сжатии многокадровых изображений с коэффициентами больше 10 раз кодек на основе проективной компенсации движения камеры обеспечивает увеличение пикового отношения сигнал-шум на 1дБ по сравнению с кодеком H.264 и на 2 дБ по сравнению с кодеком MJPEG-2000.

GEOMETRIC MODELS OF MULTI-ANGLE IMAGES AND PROJECTIVE COMPENSATION OF CAMERA MOTION

V.Yu. TSVIATKOU

Abstract

Geometric models for the formation and prediction of multi-angle images are developed. As a means of prediction is proposed to use projective techniques compensate of camera motion. It is shown that the projective motion compensation allows to create an efficient compression techniques of multi-angle images.

Список литературы

1. *Wallace G.K.* // Communications of the ACM. 1991. Vol. 34. № 4. P. 30–44.
2. *Yu W., Qiu R., Fritts J.* // Proc. SPIE, Visual Communications and Image Processing. 2002. Vol. 4671. P. 635–645.
3. *Iain R.* H.264 and MPEG-4 Video Compression and Video Coding for Next-generation Multimedia. UK, 2003.
4. *Sullivan G.J., Ohm J.-R., Han W.-J. et al.* // IEEE Transactions on Circuits and Systems for Video Technology. 2012. Vol. 22, № 12. P. 1649–1668.
5. *Журавлёв А.А., Цветков В.Ю.* // Докл. БГУИР. № 1 (79). 2014. С. 5–10.
6. *Аль-Джубури Т.М., Конопелько В.К., Цветков В.Ю.* // Инж. вест. 2010. № 2 (30). С. 20–26.

УДК 519.2:005

ИНФОРМАЦИОННАЯ ПРОГНОЗИРУЕМОСТЬ СТОХАСТИЧЕСКИХ ПРОЦЕССОВ В ДИСКРЕТНОМ ВРЕМЕНИ

А.В. ОВСЯННИКОВ, В.М. КОЗЕЛ

*Белорусский государственный университет информатики и радиоэлектроники
П. Бровка, 6, Минск, 220013, Беларусь*

*Белорусский государственный университет
пр. Независимости, 4, Минск, 220030, Беларусь*

Поступила в редакцию 4 июня 2014

Приведено определение одношаговой информационной прогнозируемости стохастического процесса в дискретном времени. Исследовано влияние накопления данных о процессе на его информационную прогнозируемость. Получены соотношения, связывающие прогнозируемость процесса в целом с прогнозируемостью его отдельных параметров. Приведены примеры определения информационной прогнозируемости для процессов, описываемых линейными разностными схемами.

Ключевые слова: прогнозируемость, стохастический процесс, оптимальная разностная схема, одношаговая плотность перехода, одношаговая информационная прогнозируемость.

Введение

Решение теоретических и прикладных задач прогнозирования, множество методов, методик и подходов к их реализации зачастую оставляет без внимания вопрос о самом понятии прогнозируемости исследуемого процесса или системы. В данном случае под прогнозируемостью понимается функция, количественно характеризующая возможные знания об исследуемом явлении (процессе или системе) в любые будущие моменты времени. Такая функция, очевидно, должна являться функцией времени. В то же время, в обширной литературе по прогнозированию отсутствует единое понимание того, что следует рассматривать под прогнозируемостью. Во многом такое положение связано с различием областей применения и разнообразием методов прогноза. Так, во-первых, исторически прогнозируемость связывалась со скоростью потери информации о состоянии исследуемой системы с течением времени [1,2] (энтропия Колмогорова); во-вторых, с позиций теории случайных процессов, в стационарном случае, возможности прогнозирования связывают с исследованием корреляционно-спектральных характеристик процессов, в частности с интервалом корреляции; в-третьих, теория оценивания, фильтрации, управления предполагает построение алгоритмов обработки, включающих элементы прогноза и коррекции. В связи с этим, предложенное в работе [3] определение информационной прогнозируемости в непрерывном времени, базирующееся на обобщении информации Фишера на случай нестационарных плотностей вероятности, может оказаться интересным с теоретической и практической точки зрения, позволяющим с единых позиций, однозначно трактовать это понятие.

В данной работе, предложенный в статье [3], подход к определению информационной прогнозируемости распространен на стохастические процессы в дискретном времени, представленные соответствующими оптимальными разностными схемами или одношаговыми плотностями перехода. Зависимость одношаговой информационной прогнозируемости от шага дискретизации и текущего момента времени позволяет расширить классификацию стохастических процессов, учитывая «хорошо» и «плохо» прогнозируемые. Количественная

мера информационной прогнозируемости позволяет также оценить адекватную поставленной задаче величину горизонта прогноза применяемой прогностической модели. Особый интерес, в этом случае, представляет исследование информационной прогнозируемости с учетом эффекта накопления данных о процессе.

Цель работы – дать определение одношаговой информационной прогнозируемости стохастического процесса и его параметров в дискретном времени, установить основные соотношения, связывающие прогнозируемость параметров с прогнозируемостью стохастического процесса в целом, а также на примере процессов, представленных оптимальными линейными разностными схемами, получить выражения информационной прогнозируемости, учитывающие эффект накопления данных.

Одношаговая информационная прогнозируемость марковского процесса

Рассмотрим задачу прогнозируемости стохастического процесса в дискретном времени в условиях регулярного статистического эксперимента. Пусть имеется наблюдаемая последовательность $\Xi = [\xi_1, \dots, \xi_k]$ и соответствующая многомерная плотность вероятности $P_{\Xi} = P(\Xi | X)$, где $X = \{X_q\}$, $q = \overline{1, l}$ – набор параметров плотности.

В непрерывном времени модель процесса представляется стохастическим дифференциальным уравнением (СДУ) $\dot{\xi}(t) + a(t, \xi) = g(t)\zeta(t)$, где $a(t, \xi)$, $g(t)$ – известные детерминированные функции удовлетворяющие условию Липшица, $\zeta(t)$ – нормальный белый шум с нулевым средним $\langle \zeta(t) \rangle = 0$ и дельтаобразной корреляционной функцией $\langle \zeta(t)\zeta(t') \rangle = N\delta(t'-t)/2$, N – односторонняя спектральная плотность. Оптимальная разностная схема, при понимании СДУ в форме Ито, имеет вид [4]

$$\xi_{i+1} = \xi_i - \Delta a_i + \Delta g_i \zeta_i, \quad (1)$$

где $\Delta = t_{i+1} - t_i$ малый интервал времени, на котором слева и справа интегрируется СДУ, $a_i = a(t_i, \xi_i)$, $\zeta_i = (1/\Delta) \int_{t_i}^{t_i+\Delta} \zeta(t) dt$, $b(t_i) = \lim_{\Delta \rightarrow 0} (N/2\Delta) \int_{t_i}^{t_i+\Delta} g^2(x) dx$ – коэффициент диффузии (в случае $g = \text{const}$ этот коэффициент равен $b = Ng^2/2$), $M_{\xi_i} = \xi_i - \Delta a_i$ – математическое ожидание, $D_{\xi_i} = b(t_i)\Delta$ – дисперсия. Глобальная среднеквадратическая погрешность разностной схемы (1) определяется величиной [4]: $\sigma \leq \Delta \left(\int_0^T M \left[(a'_\xi g(t))^2 \right] dt \right)^{1/2}$, $T = k\Delta$. В дальнейшем, там где это не вызывает сомнений, зависимость функций от аргументов опускается.

Одношаговая плотность перехода (ОПП) марковского процесса (1) имеет вид

$$\pi_{i+1,i} = \exp \left\{ -(2D_{\xi_i})^{-1} \left[\xi_{i+1} - M_{\xi_i} \right]^2 \right\} / \sqrt{2\pi D_{\xi_i}}. \quad (2)$$

Введем следующее определение.

Определение 1. Одношаговая информационная прогнозируемость по Фишеру вектора параметров X марковского процесса определяется матрицей

$$\begin{aligned} IP_X(t_i) &= \left\langle \left(\frac{\partial \ln \pi_{i+1,i}}{\partial X} \right) \left(\frac{\partial \ln \pi_{i+1,i}}{\partial X} \right)^T \right\rangle_{\xi_i, \xi_{i+1}} = \\ &= \int_{\xi_i} \int_{\xi_{i+1}} \left(\frac{\partial \ln \pi_{i+1,i}}{\partial X} \right) \left(\frac{\partial \ln \pi_{i+1,i}}{\partial X} \right)^T P(\xi_i, \xi_{i+1}) d\xi_i d\xi_{i+1}, \quad t_i > 0, \end{aligned} \quad (3)$$

представляющей собой информационную матрицу Фишера [5], зависящую от момента времени t_i , при выполнении условий регулярности для $\pi_{i+1,i}$.

В частном случае одномерного параметрического множества, информационная прогнозируемость параметра X представляется функцией информационного количества

Фишера, зависящего от параметра t_i : $IP_X(t_i) = \left\langle \left(\partial \ln \pi_{i+1,i} / \partial X \right)^2 \right\rangle_{\xi_i, \xi_{i+1}}$. Очевидно также, что если $\left\langle \left(\partial \ln \pi_{i+1,i} / \partial X_p \right) \left(\partial \ln \pi_{i+1,i} / \partial X_q \right) \right\rangle_{\xi_i, \xi_{i+1}} = 0$, где $p \neq q$ и $p, q = \overline{1, l}$, то матрица $IP_X(t_i)$ имеет диагональный вид.

Определение 2. Одношаговая информационная прогнозируемость марковского процесса ξ_i по Фишеру в целом есть неотрицательная функция

$$IP_\xi(t_i) = \left\langle \left(\partial \ln \pi_{i+1,i} / \partial t_i \right)^2 \right\rangle_{\xi_i, \xi_{i+1}}, t_i > 0. \quad (4)$$

Установим связь между одношаговой информационной прогнозируемостью по параметрам $IP_X(t_i)$ и информационной прогнозируемостью марковского процесса в целом $IP_\xi(t_i)$. Эта связь определяется следующими выражениями:

$$\begin{aligned} IP_\xi(t_i) &= \left\langle \left(\partial \ln \pi_{i+1,i} / \partial t_i \right) \left(\partial \ln \pi_{i+1,i} / \partial t_i \right) \right\rangle_{\xi_i, \xi_{i+1}} = \left\langle \left[\dot{X}^T \left(\partial \ln \pi_{i+1,i} / \partial X \right) \right] \left[\left(\partial \ln \pi_{i+1,i} / \partial X \right)^T \dot{X} \right] \right\rangle_{\xi_i, \xi_{i+1}} = \\ &= \dot{X}^T \left\langle \left(\partial \ln \pi_{i+1,i} / \partial X \right) \left(\partial \ln \pi_{i+1,i} / \partial X \right)^T \right\rangle_{\xi_i, \xi_{i+1}} \dot{X} = \dot{X}^T IP_X(t_i) \dot{X}. \end{aligned}$$

Таким образом, с учетом условия $\left\langle \left(\partial \ln \pi_{i+1,i} / \partial X_p \right) \left(\partial \ln \pi_{i+1,i} / \partial X_q \right) \right\rangle_{\xi_i, \xi_{i+1}} = 0$, $p \neq q$, $p, q = \overline{1, l}$ получаем уравнение взаимосвязи

$$IP_\xi(t_i) = \sum_q IP_{X_q}(t_i) \dot{X}_q^2. \quad (5)$$

Вклад (степень влияния) отдельных параметров ОПП в информационную прогнозируемость процесса в целом $IP_\xi(t_i)$ можно оценить коэффициентом влияния $\gamma_{X_q}(t_i) = IP_{X_q}(t_i) \dot{X}_q^2 / \sum_q IP_{X_q}(t_i) \dot{X}_q^2$, $q = \overline{1, l}$.

В характерном частном случае, когда имеем двухпараметрическую ОПП $X = [M_{\xi_j}, D_{\xi_j}]$ получаем

$$IP_X(t_i) = \begin{bmatrix} IP_M(t_i) & 0 \\ 0 & IP_D(t_i) \end{bmatrix}.$$

В матрице одношаговая информационная прогнозируемость математического ожидания и дисперсии соответственно равны

$$IP_M(t_i) = D_{\xi_j}^{-1}, \quad IP_D(t_i) = 0,5D_{\xi_j}^{-2}, \quad IP_D(t_i) = 0,5IP_M(t_i)^2. \quad (6)$$

С учетом формулы (5) одношаговая информационная прогнозируемость марковского процесса (2) в целом имеет вид

$$IP_\xi(t_i) = \dot{M}_{\xi_j}^2 / D_{\xi_j} + 0,5 \left[\dot{D}_{\xi_j} / D_{\xi_j} \right]^2 = \Delta \dot{a}_i^2 / b_i + 0,5 \left[\dot{b}_i / b_i \right]^2, \quad a_i = a(t_i, \xi_i), \quad b_i = b(t_i). \quad (7)$$

Из выражения (7) видно, что одношаговая информационная прогнозируемость отдельных параметров ОПП и процесса в целом полностью определяется параметрами оптимальной разностной схемы (1). При достаточно малых величинах $\Delta \rightarrow 0$ получаем $IP_\xi(t_i) \approx 0,5 \left[\dot{b}_i / b_i \right]^2$.

Для формирования ОПП может быть предложен более общий подход – конструктивное формирование ОПП с формальной заменой в одномерной плотности $P(\xi | X)$ переменных $\xi \rightarrow \xi_{i+1} - M_{\xi_j}$ и $D_{\xi_j} \rightarrow D_{\xi_j}$, требуя при этом сохранения условий нормировки. Строго говоря, полученная таким образом конструкция ОПП, как и в случае априорного выбора функции ядра в методах непараметрической оценки плотности, в общем случае, не является физически

обоснованной или математически доказанной. Однако сконструированные таким образом ОПП являются удобными аналитическими моделями с физически понятными при их использовании результатами. В качестве такой конструкции удобно использовать следующую:

$$\pi_{i+1,i} = c_i^{-1} \exp(-B_{i+1,i}), \quad \pi_0 = P(\xi_0), \quad (8)$$

где $B_{i+1,i} = B[\xi_{i+1} - M_{\xi_i}, D_{\xi_i}, X_i]$ – семейство параметрических функций, $X_i = X(t_i)$ – значение параметра ОПП в i -тый момент времени, c_i – параметры нормировки (в общем случае $c_i = c(X_i)$). Так, в частности, при решении практических задач нашли применение модели, совпадающие по форме с обобщенно-нормальным распределением следующего вида:

$$\pi_{i+1,i} = B(m) \exp\left(-A(m) \left(\left| \xi_{i+1} - M_{\xi_i} \right| / \sqrt{D_{\xi_i}} \right)^m\right), \quad m \geq 0,5, \quad (9)$$

где $A(m) = (\Gamma(3/m)\Gamma^{-1}(1/m))^{m/2}$, $B(m) = A(m)^{1/m} / (2\sqrt{D_{\xi_i}}\Gamma(1+1/m))$, Γ – гамма-функция. Наиболее часто такие модели используются с параметром $m \in \{0,5; 1; 2\}$. Например, для $m = 1$ прогнозируемость по параметрам ОПП следующая:

$$IP_M(t_i) = 2 / D_{\xi_i}, \quad IP_D(t_i) = 0,25D_{\xi_i}^{-2}, \quad IP_D(t_i) = IP_M(t_i)^2 / 16, \quad (10)$$

а для $m = 2\nu$, $\nu = 1, 2, \dots$, соответственно, получаем

$$IP_M(t_i) = \frac{m^2 \Gamma(2 - 1/m) A(m)^{2/m}}{D_{\xi_i} \Gamma(1/m)}, \quad IP_D(t_i) = 0,25m D_{\xi_i}^{-2}. \quad (11)$$

Информационная прогнозируемость стохастических процессов представленных оптимальными линейными разностными схемами

В теоретическом аспекте представляет интерес исследование и анализ различных моделей оптимальных разностных схем с точки зрения возможности прогнозирования процессов, которые могут быть описаны этими моделями.

Рассмотрим примеры, поясняющие смысл определенных величин одношаговой информационной прогнозируемости. Для процесса, описываемого линейной разностной схемой с переменными параметрами

$$\xi_{i+1} = \xi_i - \Delta\mu_i \xi_i + \Delta g_i \zeta_i, \quad \mu_i = \mu(t_i), \quad g_i = g(t_i), \quad (12)$$

математическое ожидание и дисперсия определяются выражениями: $M_{\xi_i} = \xi_i [1 - \Delta\mu_i]$, $D_{\xi_i} = b(t_i)\Delta$.

В табл. 1 приведены параметры разностных схем, параметры ОПП и одношаговая информационная прогнозируемость двух процессов – чисто диффузионного (столбец №1) и гауссовского (столбец №2). В табл. 1, в условиях малых $\Delta \ll 1$, введены обозначения $r = 1 - \Delta\mu \approx e^{-\mu\Delta}$, $\sigma^2 = Ng^2 / 4\mu$. Результаты, представленные в табл. 1, позволяют заметить, что для диффузионного и гауссовского процессов прогнозируемость в целом (5), (7) равна нулю, в то время как одношаговая информационная прогнозируемость отдельных параметров есть фиксированная величина, зависящая только от шага дискретизации Δ : $IP_M(t_i) = \sigma^{-2}(1 - r^2)^{-1}$, $IP_D(t_i) = 0,5\sigma^{-4}(1 - r^2)^{-2}$. При этом с увеличением интервала дискретизации $\lim_{\Delta \rightarrow \infty} IP_M(t_i) = \lim_{\Delta \rightarrow \infty} 2(Ng^2\Delta)^{-1} = 0$, $\lim_{\Delta \rightarrow \infty} IP_D(t_i) = 0$.

Полученный результат одношаговой информационной прогнозируемости для моделей с разностными схемами № 1, № 2 (табл. 1) полностью согласуется с теоретическими представлениями.

Таблица 1. Информационная прогнозируемость простейших марковских процессов

Разностная схема		№ 1	№ 2
Параметры разностной схемы		$\mu_i = 0, g_i = g$	$\mu_i = \mu, g_i = g$
Параметры ОПП $\pi_{i+1,i}$	M_{ξ_i}	ξ_i	$\xi_i [1 - \Delta\mu],$ $r\xi_i$ при $\Delta \ll 1$
	D_{ξ_i}	$Ng^2\Delta / 2$	$Ng^2\Delta / 2,$ $\sigma^2(1 - r^2)$ при $\Delta \ll 1$
$IP_M(t_i)$		$2(Ng^2\Delta)^{-1}$	$2(Ng^2\Delta)^{-1},$ $\sigma^{-2}(1 - r^2)^{-1}$ при $\Delta \ll 1$
$IP_D(t_i)$		$2(Ng^2\Delta)^{-2}$	$2(Ng^2\Delta)^{-2},$ $0,5\sigma^{-4}(1 - r^2)^{-2}$ при $\Delta \ll 1$

Рассмотрим прогнозируемость марковских процессов, описываемых оптимальными разностными схемами с более сложными функциями $\mu_i = \mu(t_i)$ и $g_i = g(t_i)$ (табл. 2). Такого рода разностные схемы могут быть использованы для описания в дискретном времени процессов с медленными нестационарными изменениями, процессов установления, процессов ухода из контрольной зоны, процессов, описывающих метрологические характеристики аппаратно-технических средств в теории надежности.

В табл. 2 приведены параметры трех оптимальных разностных схем и соответствующие им параметры ОПП, позволяющие определить одношаговую информационную прогнозируемость как по параметрам (3), так и по процессу в целом (4), (5).

Таблица 2. Информационная прогнозируемость марковских процессов, описываемых разностными схемами со сложными функциями $\mu(t_i)$ и $g(t_i)$

№	Параметры разностной схемы	Параметры ОПП
1	$\mu(t_i) = \mu, g(t_i) = \sqrt{\sum_{k=0}^K g_k t_i^k}$	$D_{\xi_i} = 2\mu\Delta \sum_{k=0}^K \sigma_k^2 t_i^k$ $M_{\xi_i} = \xi_i [1 - \Delta\mu], \sigma_k^2 = Ng_k / 4\mu$
2	$\mu(t_i) = \sum_{k=0}^K \mu_k t_i^k, g(t_i) = g_i$	$D_{\xi_i} = Ng^2\Delta / 2, D_{\xi_i} = \sigma^2(1 - r^2)$ при $\Delta \ll 1$ $M_{\xi_i} = \xi_i [1 - \Delta \sum_{k=0}^K \mu_k t_i^k], \sigma^2 = Ng^2 / 4\mu$
3	$\mu(t_i) = \sum_{k=0}^{K1} \mu_k t_i^k,$ $g(t_i) = \sqrt{\sum_{k=0}^{K2} g_k t_i^k}$	$D_{\xi_i} = 2\mu\Delta \sum_{k=0}^K \sigma_k^2 t_i^k$ $M_{\xi_i} = \xi_i [1 - \Delta \sum_{k=0}^K \mu_k t_i^k], \sigma_k^2 = Ng_k / 4\mu$
4	$a(t_i, \xi_i) = \sum_{k=0}^{K1} \mu_k t_i^k \xi_i^k,$ $g(t_i) = \sqrt{\sum_{k=0}^{K2} g_k t_i^k}$	$D_{\xi_i} = 2\mu\Delta \sum_{k=0}^K \sigma_k^2 t_i^k, \sigma_k^2 = Ng_k / 4\mu$ $M_{\xi_i} = \xi_i - \Delta \sum_{k=0}^K \mu_k t_i^k \xi_i^k$

Поскольку в реальных задачах прогнозирования информация о процессе последовательно накапливается, представляет интерес получение зависимости величины информационной прогнозируемости от объема накопленных данных. Пусть накопленная информация о процессе на интервале времени $[t_1; t_k]$ содержится в векторе данных $\Xi = [\xi_1, \dots, \xi_k]$. Тогда в скалярном случае на основании (3) можно записать

$$IPk_X(t_k) = \int_{\Xi} (\partial \ln P_{\Xi} / \partial X)^2 P_{\Xi} d\Xi.$$

Представим входящие в формулу компоненты в виде

$$\ln P_{\Xi} = \sum_{i=0}^{k-1} \ln \pi_{i+1,i},$$

$$\left(\frac{\partial \ln P_{\Xi}}{\partial X} \right)^2 = \left(\sum_{i=0}^{k-1} \frac{\partial \ln \pi_{i+1,i}}{\partial X_i} \right)^2 = \sum_{i=0}^{k-1} \left(\frac{\partial \ln \pi_{i+1,i}}{\partial X_i} \right)^2 + 2 \sum_{i=0}^{k-2} \sum_{j=i+1}^{k-1} \frac{\partial \ln \pi_{i+1,i}}{\partial X_i} \frac{\partial \ln \pi_{j+1,j}}{\partial X_j}.$$

Для иллюстрации эффекта накопления ограничимся случаем информационной прогнозируемости математического ожидания и параметра масштаба, т.е. $X = [M_{\xi_j}, D_{\xi_j}]$. На основании общей формулы (8) получаем

$$IPk_X(t_k) = \sum_{i=0}^{k-1} \int_{\xi_{i+1}} \int_{\xi_i} \left(\frac{\partial \ln \pi_{i+1,i}}{\partial X_i} \right)^2 P(\xi_i, \xi_{i+1}) d\xi_i d\xi_{i+1}.$$

Например, для ОПП (2), на основании последней формулы, информационная прогнозируемость на момент времени t_{k+r} с учетом накопления определяется выражениями:

$$IPk_M(t_{k+r}) = \sum_{i=1}^k D_{\xi_{i+r}}^{-1},$$

$$IPk_D(t_{k+r}) = 0,5 \sum_{i=1}^k D_{\xi_{i+r}}^{-2}, \quad r = 0, 1, \dots$$

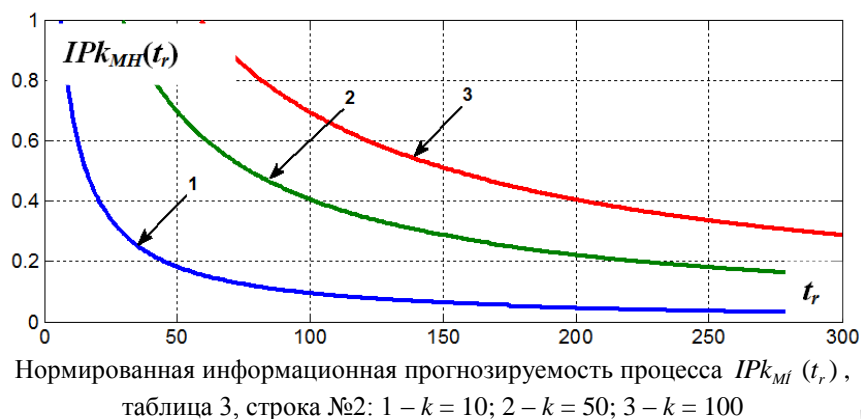
В табл. 3 представлены зависимости информационной прогнозируемости от величины k и r , определенные для разностного уравнения $\xi_{i+1} = (1 - \Delta\mu_i)\xi_i + \Delta\sqrt{g_1 + g_2 t_i} \zeta_i$. Здесь величина дисперсии $D_{\xi_j} = b(t_j)\Delta = d_1 + d_2 t_j$, $d_v = N\Delta g_v / 2$, $t_i = i\Delta$, $v = 1, 2$. В табл. 3 обозначено $\Psi_n(y) = \Psi^{(n)}(y)$, $\Psi(y) = \Gamma'(y) / \Gamma(y)$, $\delta = d_1 / d_2 \Delta$.

На рисунке представлена зависимость нормированной величины информационной прогнозируемости математического ожидания $IPk_{MH}(t_r) = IPk_M(t_r) d_1 \delta^{-1}$ процесса, заданного разностной схемой (табл. 3, строка № 2) от величины t_r при $\Delta = 1$. Зависимости, приведенные на рисунке, наглядно демонстрируют влияние накопления данных о процессе на его информационную прогнозируемость. С увеличением количества накопленных данных растет и прогнозируемость параметров и самого процесса в целом.

В заключении отметим еще одно важное практическое приложение – генерация на основе оптимальных разностных схем последовательностей с заданными свойствами информационной прогнозируемости. Так, на примере формул (1), (6), (7), (10)–(12), табл. 1 и 2 видно, что определяющими информационную прогнозируемость величинами являются параметры масштаба D_{ξ_j} и сдвига M_{ξ_j} , которые, в свою очередь, связаны с диффузией $b(t_i)$ и сносом $a(t_i, \xi_i)$. Поэтому, задаваясь желаемыми функциями $IP_M(t_i)$, $IP_D(t_i)$, $IP_{\xi}(t_i)$ можно решать обратную задачу по нахождению $g(t_i)$, $a(t_i, \xi_i)$, входящих в уравнения (1), (12).

Таблица 3. Информационная прогнозируемость процесса, описываемого линейной разностной схемой с $D_{\xi_j} = d_1 + d_2 t_j$

№	Параметры разностной схемы	Параметры ОПП
1	$d_1 > 0,$ $d_2 = 0$	$IPk_M(t_{k+r}) = kd_1^{-1}$ $IPk_D(t_{k+r}) = 0,5kd_1^{-2}$
2	$d_1 = 0,$ $d_2 > 0$	$IPk_M(t_{k+r}) = d_1^{-1} \delta [\Psi_0(k+r) - \Psi_0(r)]$ $IPk_D(t_{k+r}) = 0,5d_1^{-2} \delta^2 [\Psi_1(k+r) - \Psi_1(r)]$
3	$d_1, d_2 \neq 0,$ $d_1, d_2 > 0$	$IPk_M(t_{k+r}) = d_1^{-1} \delta [\Psi_0(k+r+\delta) - \Psi_0(r+\delta)]$ $IPk_D(t_{k+r}) = 0,5d_1^{-2} \delta^2 [\Psi_1(k+r+\delta) - \Psi_1(r+\delta)]$



Заключение

Предложенное в статье определение прогнозируемости стохастического процесса и его параметров в дискретном времени на основе информационного подхода позволяет теоретически обоснованно дать количественную оценку этой характеристики. В отличие от энтропийных мер, предложенная в работе мера основана на информационном количестве Фишера и непосредственно связана с асимптотическими алгоритмами теории оценивания. В дискретном времени прогнозируемость процесса определяется одношаговой информационной прогнозируемостью определяемой по одношаговой плотности перехода. Учет эффекта накопления данных о процессе, как и следовало ожидать, численно отражается в количественном увеличении прогнозируемости.

Практическая значимость введенной величины одношаговой и (с учетом накопления) информационной прогнозируемости состоит в том, что, во-первых, эта величина может способствовать обоснованному выбору наиболее адекватного метода прогнозирования, во-вторых, появляется возможность генерации стохастических процессов на основе разностных схем с заданными свойствами относительно их прогнозируемости. В частности, это может оказаться полезным при тестировании и сравнительной оценке алгоритмов прогнозирования.

INFORMATION PREDICTABILITY OF STOCHASTIC PROCESSES IN DISCRETE TIME

A.V. AUSIANNIKAU, V.M. KOZEL

Abstract

The definition of one-step information predictability stochastic process in discrete time is given. The influence of the accumulation process data on its informational predictability is investigated. Relations connecting the predictability of the whole process with predictability of the individual parameters are obtained. The examples of the definition of information predictability for processes described by linear difference schemes are shown.

Список литературы

1. Заславский Г.М. Стохастичность динамических систем. М., 1984.
2. Шустер Г. Детерминированный хаос. М., 1988.
3. Овсянников А.В. // Докл. БГУИР. 2014. № 6 (84). №. С. 48–54.
4. Никитин Н.Н., Разевиг В.Д. // Журнал вычислительной математики и математической физики. 1978. Т 18, № 1. С.106–117.
5. Ибрагимов И.А., Хасьминский Р.З. Асимптотическая теория оценивания. М., 1979.

УДК 519.711: 004.85

ПРИМЕНЕНИЕ И ОПТИМИЗАЦИЯ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ НА ВАЛЮТНОМ РЫНКЕ

В.А. РЫБАК, Х.М. СУЛАЙМАН

Белорусский государственный университет информатики и радиоэлектроники
П. Бровка, 6, Минск, 220013, Беларусь

Поступила в редакцию 16 октября 2014

Представлены результаты исследования по разработке нейро-нечеткой системы, которая может использоваться как инструмент поддержки принятия решений на валютном рынке. В рамках данной работы создан алгоритм формирования торговых сигналов, выделены наиболее значимые технические индикаторы, сформирована обучающая выборка, спроектирована и протестирована нейронная сеть. При этом точность полученной модели превосходит линейную и квадратичную регрессию.

Ключевые слова: валютный рынок, торговые системы, нейронные сети.

Введение

Развитие современных информационных технологий позволило существенно расширить круг лиц, которые могут являться участниками валютного рынка. Сегодня каждый из нас при наличии доступа к сети Интернет может выступать в роли трейдера на глобальных мировых торговых площадках, включая Forex. Успешность осуществления сделок при этом будет зависеть от наших умений и навыков оперативно обрабатывать существенный объем различной информации, включая экономические и политические новости, котировки и технические индикаторы.

Для целей автоматизации процесса торговли часто разрабатывают и используют системы с применением нейронных сетей. При этом научной задачей является как подбор оптимальной структуры, так и формирование обучающей выборки. Решению последней проблемы и посвящена данная статья [1].

Методика эксперимента

Идея формирования обучающей выборки состоит в том, чтобы на основании ретроспективных данных попытаться предположить, каким будет курс валюты в будущем. При этом нам в любой момент доступны десятки технических индикаторов, изменяющих свое значение в зависимости от ситуации на рынке.

Одним из наиболее часто используемых индикаторов при построении механических торговых систем (МТС) является скользящая средняя (Moving Average – MA), которую используют как сигнал на покупку, если MA с меньшим периодом пересекает снизу вверх MA с большим периодом, и наоборот.

Указанный индикатор многократно доказал свою полезность в качестве инструмента поддержки принятия решений, однако для формирования обучающей выборки нам необходим выходной числовой ряд, связанный с входными значениями индикаторов, который будет представлять собой торговые сигналы.

Для формирования указанного ряда был разработан соответствующий алгоритм, блок-схема которого представлена ниже на рис. 1.

Идея данного алгоритма заключается в следующем. Имея массив значений цен за истекший период времени (C) сформируем новый массив A , элементами которого будут значения максимального отклонения от текущей цены за период Tn . При этом положительное значение будет означать возможность открытия в данной точке сделки на покупку, отрицательное – на продажу. Таким образом, задача построения МТС сводится к нахождению таких индикаторов, значения которых бы статистически значимо были связаны с торговыми сигналами. Или другими словами – если на хронологических данных ряд индикаторов был достаточно связан с ценой, можно предположить, что данная связь сохранится и в дальнейшем.

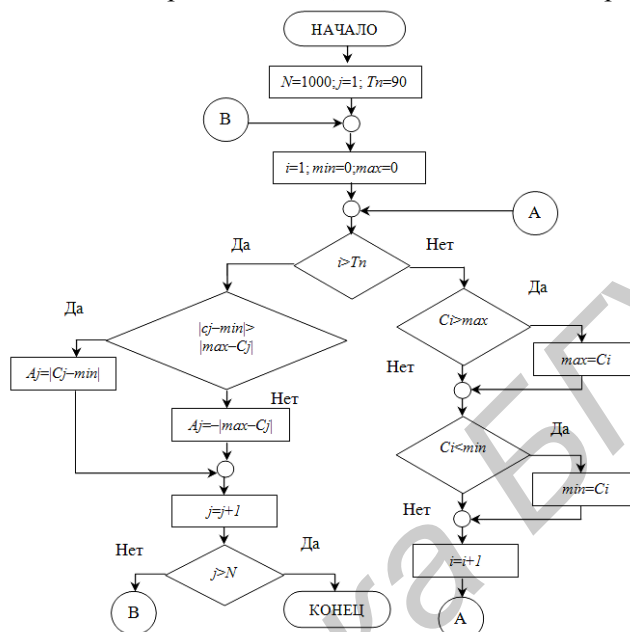


Рис. 1. Блок-схема алгоритма формирования торговых сигналов

Располагая историческими данными о котировках валют (взятых из программы MetaTrader), проведем корреляционный анализ полученного ряда торговых сигналов и разностью между скользящими средними различного периода. Результаты представлены в табл. 1, из которой видно, что для выбранных валютной пары EURUSD и четырехчасового интервала наибольший коэффициент корреляции достигается, если рассматривается разность МА с периодами 120 и 100–110. Из полученных результатов можно предположить, что дальнейшее увеличение периода вызовет еще большее увеличение связи, однако проведенные исследования показали, что при увеличении периода более 140 корреляция начинает убывать.

Исходя из сказанного выше, индикатор МА будет использоваться как один из входных параметров для нейронной сети. Однако для повышения устойчивости МТС необходимо выделить и другие индикаторы, тесно связанные с изменением цены.

Таблица 1. Коэффициенты корреляции между рядом торговых сигналов и разностью значений МА различных периодов

Периоды	5	10	15	20	25	50	60	70	80	90	100	110	120
120	0,36	0,38	0,41	0,43	0,45	0,54	0,56	0,58	0,60	0,61	0,62	0,62	–
110	0,34	0,36	0,39	0,41	0,44	0,53	0,55	0,57	0,59	0,60	0,61	–	–0,62
100	0,32	0,34	0,37	0,39	0,42	0,51	0,54	0,56	0,58	0,59	–	–0,61	–0,62
90	0,29	0,32	0,34	0,37	0,39	0,49	0,52	0,54	0,56	–	–0,59	–0,60	–0,61
80	0,26	0,29	0,31	0,34	0,36	0,47	0,50	0,53	–	–0,56	–0,58	–0,59	–0,60
70	0,23	0,25	0,28	0,31	0,33	0,44	0,48	–	–0,53	–0,54	–0,56	–0,57	–0,58
60	0,19	0,22	0,24	0,27	0,30	0,41	–	–0,48	–0,50	–0,52	–0,54	–0,55	–0,56
50	0,15	0,18	0,20	0,23	0,26	–	–0,41	–0,44	–0,47	–0,49	–0,51	–0,53	–0,54
25	0,05	0,07	0,09	0,11	–	–0,26	–0,30	–0,33	–0,36	–0,39	–0,42	–0,44	–0,45
20	0,03	0,05	0,07	–	–0,11	–0,23	–0,27	–0,31	–0,34	–0,37	–0,39	–0,41	–0,43
15	0,02	0,03	–	–0,07	–0,09	–0,20	–0,24	–0,28	–0,31	–0,34	–0,37	–0,39	–0,41
10	0,01	–	–0,03	–0,05	–0,07	–0,18	–0,22	–0,25	–0,29	–0,32	–0,34	–0,36	–0,38
5	–	–0,01	–0,02	–0,03	–0,05	–0,15	–0,19	–0,23	–0,26	–0,29	–0,32	–0,34	–0,36

С этой целью был проведен корреляционный анализ между наиболее распространенными техническими индикаторами, используемыми трейдерами, и сформированным рядом торговых сигналов. Результат представлен в табл. 2.

Таблица 2. Коэффициенты корреляции между значениями ряда торговых сигналов и индикаторами

Обозначение индикатора	Название индикатора	Коэффициент корреляции
MACD	Схождение/расхождение скользящих средних (Moving Average Convergence/Divergence)	-0,0795
MACD signal	Сигнальная линия MACD	-0,1065
MACD - signal	Разность между MACD и сигнальной линией	0,0734
AC	Технический индикатор ускорения/замедления (Acceleration/Deceleration, AC)	
AD	Индикатор накопления/распределения (accumulation/distribution, A/D)	-0,2610
Alligator	Аллигатор (Bill Williams' Alligator)	-0,3116
Gator	Гатор осциллятор (Gator Oscillator)	0,0788
ADX	Индекс среднего направления движения (Average Directional Movement Index, ADX)	0,1072
ATR	Средний истинный диапазон (Average True Range, ATR)	0,0055
AO	Замечательный осциллятор Билла Вильямса (Awesome Oscillator, AO)	-0,0316
BearsPower	Индикатор медвежьей силы	-0,0164
Bands	Полосы боллинджера (Bollinger Bands®, BB)	-0,2950
BullsPower	Индикатор бычьей силы	-0,0057
CCI	Индекс товарного канала (Commodity Channel Index, CCI)	-0,0312
DeMarker	Индикатор Демарка (DeMarker, DeM)	-0,0600
Envelopes	Индикатор огибающие линии (конверты, envelopes)	-0,3114
Force	Индекс силы (Force Index, FRC)	0,0087
Ichimoku	Индикатор ишимоку кинко хайо (Ichimoku Kinko Hyo)	-0,2783
MFI	Индекс денежных потоков (Money Flow Index, MFI)	0,0076
OsMA	Индикатор скользящей средней осциллятора (Moving Average of Oscillator)	0,0734
SAR	Параболическая система SAR (Parabolic SAR, stop & revers)	-0,2963
RSI	Индекс относительной силы (Relative Strength Index, RSI)	-0,0845
RVI	Индекс относительной бодрости (Relative Vigor Index, RVI)	-0,0636
StdDev	Стандартное отклонение (Standard Deviation, StdDev)	-0,0112
Stochastic	Стохастический осциллятор (Stochastic Oscillator)	-0,0179
WPR	Технический индикатор процентный диапазон Вильямса (Williams' Percent Range, %R)	-0,0371

Из табл. 2 видно, что наибольший по модулю коэффициент корреляции имеют следующие индикаторы: AD, Alligator, Bands, Envelopes, Ichimoku и SAR. Изменяя их параметры и временные периоды попытаемся улучшить связь со статистическим рядом торговых сигналов.

Перечень параметров индикатора AD не предполагает возможность его оптимизации. Подбор периодов и режимов работы индикатора Alligator изменили коэффициент корреляции только в третьем знаке после запятой. Оптимизация параметров индикатора Bands позволила увеличить модуль коэффициента корреляции до 0,3535 за счет подбора режима и периода (45). Не удалось существенно оптимизировать степень связи у индикатора Envelopes – подбор параметров позволил выйти на значение минус 0,31224 за счет изменения режима сигнальной линии. Несущественно – до минус 0,3136 удалось увеличить модуль коэффициента корреляции между индикатором Ichimoku и рядом торговых сигналов. Улучшить связь индикатора SAR не удалось. Для проверки на устойчивость выбранных на основании табл. 2 индикаторов в табл. 3 представлены коэффициенты корреляции их значений на графиках различных временных периодов.

Таблица 3. Коэффициенты корреляции для выбранных индикаторов на различных таймфреймах

Период	AD	Alligator	Bands	Envelopes	Ichimoku	SAR	MA140-MA120
один час	-0,1887	-0,4013	-0,3804	-0,4119	-0,4214	-0,4307	0,2721
четыре часа	-0,2611	-0,4638	-0,4694	-0,4884	-0,4715	-0,4439	0,5867
один день	0,1122	-0,3219	-0,1819	-0,5647	-0,4128	-0,3114	0,5799

Анализируя представленные в табл. 3 данные, необходимо отметить изменение знака у индикатора AD на дневном графике. Этот факт требует исключения его из перечня

индикаторов, значения которых будут в дальнейшем использованы для формирования обучающей выборки и поддержки принятия решений. Для остальных индикаторов сильная статистическая связь их значений с торговыми сигналами сохранилась. При этом с учетом размера выборки ($N=1000$) и критических значений коэффициента корреляции (0,1460) величина ошибки p составляет 0,001. То есть с вероятностью 99,9 % выявленные связи являются статистически значимыми.

Таким образом, обучающая выборка для нейронной сети будет состоять из семи рядов, первые шесть из которых будут являться входными параметрами, а седьмой – рядом торговых сигналов.

Построение нейро-нечеткой системы

Для построения нейро-нечеткой системы воспользуемся пакетом MatLab, предоставляющим возможность проектировать нейронные сети и работать с нечеткими множествами [2]. Для примера сформируем текстовый файл с тремя рядами, первые два из которых будут содержать значения технических индикаторов (Envelops и MA140-120), имеющих наибольший коэффициент корреляции, третий – торговые сигналы. Структура сгенерированной нейронной сети представлена на рис. 2.

Проведем обучение нечеткой системы на протяжении 25 итераций. Динамика обучения выводится в основном окне ANFIS-редактора: ось абсцисс соответствует итерациям алгоритма обучения, а ось ординат – значениям средней квадратичной ошибки. После настройки ошибка обучения составляет 123,6. Если увеличить число вершин второго уровня сети (inputmf, рис. 2) с трех до пяти, а количество циклов обучения установить равным 50, то ошибка обучения снизится до 104,5. Ошибка на тестовой выборке составляет 104,5.

Сравним результаты нечеткой идентификации с линейной и квадратичной регрессиями. В результате получаем модель, ошибки которой на обучающей и тестовой выборках составляют 180,5 и 156,5. Для нелинейной модели ошибки на обучающей и тестовой выборках составляют 165,5 и 143,3 соответственно.

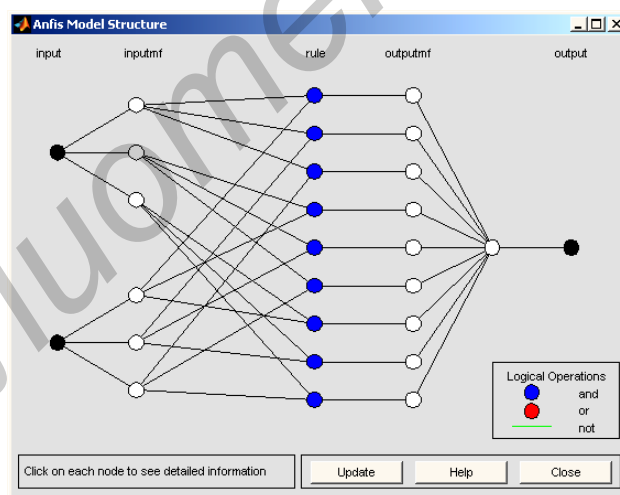


Рис. 2. Структура сгенерированной нейро-нечеткой системы

Исследуя вопрос оптимизации структуры созданной модели с большим количеством входными сигналами следует отметить, что при прочих равных условиях искусственная нейронная сеть (ИНС) с большим количеством правил показывает лучшую точность. Это подтверждает и ИНС с шестью нейронами и 216 правилами, ошибка – 1506,9, с семью нейронами и 343 правилами – 1107,5. Однако наращивание количества правил также как и количества входных сигналов упирается в вычислительные мощности и время, требуемое для обучения системы.

Кроме количества входных параметров и структуры ИНС на точность системы существенное влияние оказывают и другие параметры, в частности, *Input MF Type*. Так, для предлагаемого по умолчанию типа «gbellmf» ошибка обучения (3139,8) получается выше, чем для типа «gaussmf» (2896,1), который из всего множества признан нами наиболее

оптимальным. При этом при изменении типа выходного узла с «constant» (предлагается по умолчанию) на «linear» ошибку обучения удастся снизить еще больше – до 1944,6.

Из предложенных методов обучения оптимальным является «hybrid», который по сравнению "backproga" позволяет получать более точные системы.

Заключение

Таким образом, нечеткая модель точнее линейной и квадратичной регрессий описывает зависимость торговых сигналов от значений технических индикаторов. Важно также заметить, что использование вместо простого метода решетчатого разбиения (Grid partition) метода сложной кластеризации (Sub. clustering) позволяет еще больше снизить уровень ошибки синтезированной нейро-нечеткой сети и тем самым повысить ее адекватность [3]. Проведенные нами исследования позволили выявить оптимальные параметры ИНС в рамках модели ANFIS, которая в полной мере позволяет нам создавать и эффективно использовать СППР на рынке Forex.

APPLICATION AND OPTIMIZATION OF NEURAL NETWORKS FOR DECISION SUPPORT ON EXCHANGE MARKET

V.A. RYBAK, H.M. SULAIMAN

Abstract

The results of investigation of developing neuron-fuzzy system that can be used as a decision support tool on the currency market are presented. The algorithm of trading signals forming is created, the most important technical indicators are chosen, the training selection is formed, and the neuron net is created and tested. The precision of developed model exceeds linear and quadratic regressions.

Список литературы

1. *Яремчук А.В.* Информационно-аналитические методы и алгоритмы поддержки принятия решений при управлении портфелем ценных бумаг на основе сценарного подхода к прогнозированию: Автореф. дис. ... канд. экон. наук. СПб, 2011.
2. *Штовба С.Д.* Проектирование нечетких систем средствами MATLAB. М., 2007.
3. *Рыбак В.А.* Методологические основы принятия решений для управления природоохранной деятельностью. Минск, 2009.

УДК 658.511

ОЦЕНКА РИСКА ВЫПОЛНЕНИЯ ЛОГИСТИЧЕСКИХ ПРОЕКТОВ НА НАЧАЛЬНОЙ СТАДИИ

А.И. КУЗЬМИЧ

*Белорусский государственный университет информатики и радиоэлектроники
П. Бровка, 6, Минск, 220013, Беларусь*

Поступила в редакцию 24 сентября 2014

Рассмотрена задача оценки риска при выполнении логистических проектов. Представлено решение для BOL-стадии в рамках методологии CL2M. В результате построена теоретическая основа для автоматизации оценки риска проекта на начальной стадии проектирования, что снижает вероятность ошибок на других стадиях.

Ключевые слова: сложные технические объекты, оценка риска.

Введение

Жизнеспособность экономики любого государства во многом зависит от эффективности логистических проектов (ЛП), обеспечивающих перемещение грузов от производителей к потребителям с использованием мобильных гетерогенных объектов (МГО) (железнодорожных составов, автопоездов, танкеров и т.д.) [1]. Данная группа проектов отличается сложностью, большим объемом затрат и высокими требованиями к персоналу. Вопросы повышения эффективности и безопасности проектов активно исследуются ведущими научными учреждениями. В частности, 22 организации из Евросоюза, Швейцарии, Японии, Австралии и США разработали методологию Closed Loop Lifecycle Management (CL2M), соответствующую современным требованиям бизнеса и экологической безопасности с использованием преимуществ информационной среды [2]. Согласно CL2M жизненный цикл проекта должен включать три стадии: начальную (Beginning of life, BOL) – проектирование и создание объектов; среднюю (Middle of life, MOL) – применение, техническое обслуживание; конечную (End of life, EOL) – утилизация или восстановление и повторное применение компонентов объекта.

Применение методологии CL2M на практике вызывает ряд проблем, основная из которых заключается в отсутствии механизмов анализа риска на начальной стадии проекта [3]. В результате руководитель проекта не может объективно оценить возможность выполнения проекта, что неизбежно приводит к росту неопределенности и снижает эффективность стадий MOL и EOL.

В данной работе рассматривается метод оценки риска ЛП на этапе BOL в рамках парадигмы CL2M на основе экспертных знаний и теории нечетких множеств.

Основные понятия и определения

Для обеспечения однозначного понимания тематики статьи приведем определения основных терминов.

Проект – совокупность задач и мероприятий, обеспечивающих достижение запланированной цели. В данном случае рассматриваются проекты, относящиеся к логистике. Жизненный цикл проекта – комплекс процессов реализации проекта с момента его инициализации до завершения. В данном случае – BOL, MOL, EOL. Акторы – одушевленные и искусственные участники проекта. В данном случае к актерам относятся менеджер (С),

диспетчер (D), исполнитель (P) и МГО. МГО – сложный мобильный гетерогенный технический объект с изменяемой структурой, используемый для перемещения грузов. Неопределенность – отсутствие или недостаток информации для реализации проекта. Риск – оценка опасности снижения эффективности или невыполнения проекта. Внешняя среда (среда) – комплекс внешних факторов, влияющие на возможность выполнения проекта. Приведенные выше термины используются в литературе для проектировщиков и программистов, поэтому их в известной степени можно считать универсальными.

Постановка задачи

Стадия BOL является важной частью проекта, т.к. от нее зависит возможность и эффективность реализации стадий MOL и EOL. Спектр возможных рисков при ее выполнении непредсказуем и полностью их избежать в принципе невозможно. Однако учесть наиболее значимые риски, приводящие к тяжелым последствиям, можно, что значительно повысит вероятность успеха.

Постановку задачи оценки риска на стадии BOL сформулируем следующим образом. Дано: проект $proj$ реализуется в среде с различным уровнем неопределенности Z . В реализации участвуют факторы C, D, P и МГО M , состоящий из N компонентов, принадлежащих к m типам. Требуется разработать механизм оценки риска выполнения проекта на стадии BOL в рамках парадигмы CL2M. Основное требование к решению – унификация представления разнородных исходных данных и алгоритмов их обработки.

Факторы риска

На основе анализа теоретических работ [1, 3–5], можно сделать вывод, что среди прочих на эффективность ЛП влияют три основных фактора риска: уровень неопределенности проекта в целом, уровень компетентности акторов и уровень сложности МГО. Совокупность этих факторов опишем кортежем: $rBOL = (rZ, rC, rD, rP, rM)$

Фактор неопределенности проекта. Первым фактором является неопределенность проекта. Поэтому основным условием выполнимости проекта является возможность уменьшения неопределенности на основе знаний акторов. Соответственно неопределенность можно подразделить на три типа: Zp, Zd, Zc . Если неопределенность выходит за уровень акторов, проект не следует начинать. Неопределенность выражается словами, текстом, цифрами, сигналами. Несмотря на столь большое разнообразие, все эти способы относятся к компетенции акторов. Соответственно, в рамках каждой компетенции неопределенность можно ранжировать и отобразить на континуум $0..1$. Например: $rZC \rightarrow 0,00 - 0,50 - 1,00$. Этот очень удобный способ, предложенный Л.Заде [6], будем также применять для оцифровки других факторов риска. В данном случае используется три уровня градации, но в зависимости от специфики проекта может быть пять и более градаций, например: $rZC \rightarrow 0,00 - 0,25 - 0,50 - 0,75 - 1,00$.

В результате появляется возможность формального описания неопределенности для каждого актора и обработки формальными методами.

Фактор компетентности акторов. Второй фактор – компетентность акторов – является краеугольным фактором успеха любого проекта. Риск включения в проект некомпетентного актора велик, поэтому задаче выбора персонала уделяется столь много внимания в литературе и на практике. Для этого, в частности, используются многочисленные профессиональные тесты, дорогостоящие тренажеры и т.д. Независимо от способа проверки, результаты сводятся к некоторым численным оценкам. В нашем случае с учетом требования унификации, классифицируем уровень компетентности каждого актора на три класса и отобразим на континуум $[0..1]$. Например, «высший» – $0,00$, «средний» – $0,05$, «низкий» – $1,00$.

Фактор сложности МГО. МГО, применяемые для перемещения грузов, являются сложными техническими устройствами, потенциально опасными для человека и природной среды. В [4] показано, что оценка риска тесно связана со сложностью объекта. Для определения сложности необходимо построить модель, формирующую цельный взгляд на объект на основе анализа наиболее значимых сущностей объекта. В качестве модели МГО используем кортеж:

$$M = (proj, id, FL, N, \langle i. T. q. \rangle, com), \quad (1)$$

где id – идентификатор; FL – отношения компонентов (F – жесткие; L – гибкие); N – общее количество компонентов; i – порядковый номер компонента; q – количество компонентов данного типа; com – коммуникации.

В кортеже имеются четыре параметра, определяющих сложность МГО. В данном случае сложность является синонимом риска. Первый фактор сложности – это изменение структуры МГО, например: F-L-F. Такие операции, включая разъединение объекта на компоненты с последующим их объединением в единое целое, относятся к высшему уровню сложности. Соответственно, минимальная сложность будет у объекта, не изменяющего свою первоначальную структуру. Сложность существенно растет при одном изменении и резко увеличивается при двух и более изменениях. Формализуем эти факты и отобразим результат на шкале 0..1: {количество изменений : сложность} $\rightarrow 0 - 0,00, 1 - 0,50, \geq 2 - 1,00$.

Количество изменений оказывает столь значимое влияние на сложность, что в некоторых случаях значение параметра умножается на некоторый коэффициент, который изменяется в зависимости от типа МГО и значимости проекта.

Второй фактор сложности – количество типов компонентов. Чем больше типов, тем сложнее организация управления, мониторинга, ремонта и, соответственно, выше затраты. Соответственно, минимальная сложность будет при использовании одного типа и будет расти при увеличении их количества. Усредняя множество мнений в литературе, получим следующую картину: {количество типов : сложность} $\rightarrow 1 - 0,00, 2-3 - 0,50, \geq 4 - 1,00$.

Третий фактор – количество компонентов. Чем их больше, тем выше вероятность возникновения аварийных ситуаций. Поэтому важно выбрать то количество, которое соответствует условиям траектории движения и возможностям ремонта в пути. В результате усреднения данных, представленных в литературе, получим следующую шкалу: {количество компонентов : сложность} $\rightarrow 1-20 - 0,00, 20-40 - 0,50, \geq 40 - 1,00$.

Четвертый фактор – связь. При перемещении грузов на большие расстояния с изменяющимся уровнем неопределенности важнейшую роль играют средства и каналы связи. Неопределенность Z_p исполнитель снижает сам, знания для уменьшения неопределенности Z_D и Z_C он должен получать дистанционно по каналам связи от акторов D и C . Поэтому коммуникационная сложность прямо зависит от количества типов коммуникаций. Соответствующую шкалу можно представить строкой: {количество видов связи : сложность} $\rightarrow 1 - 0,00, 2 - 0,50, \geq 3 - 1,00$.

Таким образом, факторы риска с различным генезисом удалось привести к одному типу данных и детерминировать рамками шкалы [0..1] по принципу увеличения риска. Результат представлен в табл. 1.

Таблица 1. Факторы риска

	Фактор риска	Класс	Нормализованное значение риска
X1	Неопределенность	Низкая	0,00
		Средняя	0,50
		Высокая	1,00
X2	Компетентность	Высокая	0,00
		Средняя	0,50
		Низкая	1,00
X3	Изменение структуры	0	0,00
		1	0,50
		≥ 2	1,00
X4	Типы компонентов	1	0,00
		2-5	0,50
		> 5	1,00
X5	Количество компонентов	1-9	0,00
		10-30	0,50
		> 30	1,00
X6	Тип коммуникации	1	0,00
		2	0,50
		≥ 3	1,00

Очевидно, что при построении таблицы должны использоваться экспертные знания, статистика отказов, аварий и т.д. В данном случае риск классифицируется на три уровня. При необходимости список факторов, их количество можно детализировать до пяти, семи или более уровней.

Алгоритм оценки сложности

На практике оценка риска на начальной стадии проекта является более искусством, чем объективным процессом. Найти истинного эксперта не всегда удастся, поэтому многие компании пытаются формализовать этот процесс. К сожалению, их результаты описаны в литературе в самых общих чертах, что затрудняет их применение. Для решения в некоторой степени этой проблемы предлагается интуитивно понятный и легко проверяемый алгоритм оценки риска на основе данных таблицы, написанный на языке C#:

```
{
    double porog = 0.40; // порог
    double [] X = { 1.00, 0.00, 0.00*7, 1.00, 0.50, 0.00 }; // исходный вектор
    double sr = Math.Round (X.Sum() / X.Length, 3);
    if (sr > porog) Console.WriteLine("Риск = " + sr + " Высокий уровень.");
    else Console.WriteLine("Риск = " + sr + " Допустимый уровень.");
    Console.ReadKey();
}
```

На вход алгоритма подается вектор параметров риска X , который строится на основе данных таблицы, требований проекта и ресурсов компании, и порог ($porog$), разделяющий допустимый риск от высокого, который может отличаться в зависимости от оперативности, стоимости и других особенностей проекта. В рассматриваемом случае используется коэффициент 7 для параметра изменчивости структуры, порог равен 0,40. В результате выполнения алгоритма будет выведено сообщение:

Риск = 0,417 Высокий уровень.

Таким образом, любой специалист среднего уровня, знающий требования проекта, может формально оценить риск его выполнения.

Для практического применения алгоритм должен быть интегрирован в графический интерфейс, обеспечивающий ввод элементов вектора X простым указанием курсора в списке вариантов, представленных в табл. 1.

Для проверки правильности выбранного подхода и алгоритма необходимо сравнить формальный и экспертный результаты. В табл. 2 приведен результат эксперимента для семи векторов с различным значением порога.

Таблица 2. Результаты эксперимента

Вектор	Порог	Формальная оценка	Экспертная оценка
0,50, 0,00, 0,00*7, 1,00, 0,00, 0,50	0,50	0,333	Допустимый уровень риска
0,00, 1,00, 0,00*7, 0,50, 0,00, 0,00	0,50	0,250	Допустимый уровень риска
1,00, 0,00, 0,00*7, 0,50, 1,00, 0,00	0,60	0,417	Допустимый уровень риска
0,50, 1,00, 0,00*7, 0,50, 0,50, 1,00	0,40	0,583	Высокий уровень риска
1,00, 0,00, 1,00*7, 0,50, 1,00, 1,00	0,80	1,750	Высокий уровень риска
0,00, 0,00, 1,00*7, 0,50, 0,00, 0,50	0,50	1,333	Высокий уровень
0,00, 0,00, 0,00*7, 1,00, 1,00, 1,00	0,30	0,500	Высокий уровень риска

Сравнив третий и четвертый столбцы, несложно заметить, что оценка эксперта совпадает с формальной оценкой. Если выполнить типизацию проектов, можно получить унифицированный тандем: <тип проекта – значение порога>. Целесообразно так же установить коэффициент для X_3 в зависимости от типа компонентов и свойств среды (воздушной, наземной, водной), где они будут использоваться. В результате уровень достоверности и стандартизации процесса оценки риска существенно повысится.

Заключение

Разработан метод оценки риска на стадии BOL в рамках концепции CL2M. Применение метода позволяет существенно сократить время оценки риска и повысить достоверность результата. Разработан интуитивно понятный алгоритм, инвариантный количеству входных параметров и позволяющий адаптировать параметры с помощью коэффициентов и порогового значения риска к специфическим требованиям проекта.

RISK ASSESSMENT OF LOGISTICS PROJECT IMPLEMENTATION IN THE INITIAL STAGE

A.I. KUZMICH

Abstract

The problem of risk assessment in the logistics projects implementation is studied. The solution for BOL-step in the methodology CL2M is presented. As a result, a theoretical basis for automating the risk assessment of the project in the initial stage, which reduces the likelihood of errors in other stages, is offered.

Список литературы

1. *Рушкевич А.В.* // Беспроводные технологии. 2010. № 3. С. 56–60.
2. *Stark J.* Product Lifecycle Management. New York, 2004.
3. *Kuzmich A.I., Shakah G., Valvachev A.N.* // Proceedings of The 10-th International Conference on Pattern Recognition (PRIP'2011). Minsk, May, 18–20, 2011. P. 272–275.
4. *Rausand M.* Risk Assessment: Theory, Methods, and Applications. San Francisco, 2011.
5. *Casti J.* Connectivity, Complexity and Catastrophe in Large-scale Systems. San Francisco, 1979.
6. *Zadeh L., Klir G., Yuan B.* Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems. Singapore, 1996.

УДК 629.7.05

СИНТЕЗ КВАЗИОПТИМАЛЬНОГО КОНТУРА ТЕЛЕУПРАВЛЯЕМОГО БЕСПИЛОТНОГО ЛЕТАТЕЛЬНОГО АППАРАТА ПРИ ИНЕРЦИОННОМ ИЗМЕРИТЕЛЕ

А.А. КУН, С.А. ШАБАН, А.М. ЕРОМИН, А.Л. КОЧЕРОВ

*Военная академия Республики Беларусь
Минск-57, 220057, Беларусь**Поступила в редакцию 16 июня 2014*

Решена задача синтеза квазиоптимального контура телеуправляемого беспилотного летательного аппарата с учетом инерционности координатного блока цели. Предложен способ компенсации инерционности измерительного устройства. Проведен анализ зависимости параметров контура телеуправления при различных дальностях до точки встречи беспилотного летательного аппарата с целью.

Ключевые слова: синтез контура телеуправляемого беспилотного летательного аппарата, измерительное устройство, устройство выработки команд.

Введение

Одной из задач синтеза контура телеуправляемого беспилотного летательного аппарата (БЛА) является определение оптимальной структуры и параметров контура. Наличие в контуре управления измерительного устройства приводит к увеличению инерционности системы в целом.

Рассматривается контур телеуправляемого БЛА малой дальности. При синтезе принимаются следующие допущения: динамические свойства координатного блока БЛА, устройства передачи команд и системы стабилизации БЛА считаются безинерционными. В качестве управления принимается нормальное ускорение БЛА.

Постановка задачи

Представим кинематику плоского телеуправления в виде линеаризованного уравнения [1, 2]:

$$\ddot{h}(t) = W_k(t) - W_p(t), \quad (1)$$

где $h(t)$ – ошибка наведения БЛА на цель, $W_k(t)$ и $W_p(t)$ – требуемое (кинематическое) и нормальное ускорения БЛА соответственно.

Будем полагать, что $W_k(t)$ является случайной функцией времени и описывается стохастическим дифференциальным уравнением [1, 3]:

$$\dot{W}_k(t) = -\frac{1}{T} W_k(t) + \frac{1}{T} \xi(t), \quad W_k(t_0) = W_{k0},$$

где $\xi(t)$ – гауссовый процесс типа белого шума с интенсивностью Q .

Представим уравнение (1) в форме Коши [4]. Положив $h(t) = x_1(t)$, $W_k(t) = x_3(t)$ при $t_0 < t < t_k$, получим систему уравнений для фазовых переменных с учетом модели задающего воздействия:

$$\begin{cases} \dot{x}_1(t) = x_2(t), & x_1(t_0) = x_{10}; \\ \dot{x}_2(t) = x_3(t) - W_p(t), & x_2(t_0) = x_{20}; \\ \dot{x}_3(t) = -\frac{1}{T}x_3(t) + \frac{1}{T}\xi(t), & x_3(t_0) = x_{30}. \end{cases} \quad (2)$$

В векторно-матричной форме (2) примет вид:

$$\dot{\mathbf{X}}(t) = \mathbf{F}\mathbf{X}(t) + \mathbf{G}\mathbf{U}(t) + \mathbf{J}\zeta(t), \quad \mathbf{X}(t_0) = \mathbf{X}_0, \quad (3)$$

где $\mathbf{F} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -\frac{1}{T} \end{bmatrix}$; $\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$; $\mathbf{G} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$; $\mathbf{U} = \begin{bmatrix} 0 \\ W_p \\ 0 \end{bmatrix}$; $\mathbf{J} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{1}{T} & 0 & 0 \end{bmatrix}$.

Вектор состояния в начальный момент времени $\mathbf{X}(t_0)$ – гауссовый случайный вектор с характеристиками: $M[\mathbf{X}(t_0)] = 0$, $M[\mathbf{X}(t_0)\mathbf{X}^T(t_0)] = \mathbf{P}_0$, где \mathbf{P}_0 – матрица априорных дисперсий фазовых координат наблюдаемой системы.

Ошибка наведения в контуре телеуправления формируется в виде [1, 2]:

$$h = r_p(t) \cdot (\hat{\varepsilon}'_{ц} - \hat{\varepsilon}'_p), \quad (4)$$

где $r_p(t)$ – дальность до БЛА; $\hat{\varepsilon}'_{ц}$, $\hat{\varepsilon}'_p$ – относительные угловые координаты цели и БЛА, измеряемые координатным блоком цели и координатным блоком БЛА соответственно.

Передаточная функция координатного блока цели определяется выражением [1]:

$$K_{ц0}(p) = \frac{1 + pT_{ц}}{\frac{1}{K_{ц}}p^2 + pT_{ц} + 1} = \frac{z_1(p)}{x_1(p)}, \quad (5)$$

где $K_{ц}$, $T_{ц}$ – параметры передаточной функции; $z_1 = \hat{\varepsilon}'_{ц}$ – измеренная координата; $x_1 = \varepsilon'_{ц} = \varepsilon_{ц} - \varepsilon_a$ – ошибка между направлением на цель и положением антенны.

Для того, чтобы найти систему дифференциальных уравнений в форме Коши, соответствующую передаточной функции (5), составим структурную схему координатного блока цели как показано на рис. 1. Такой подход исключает дифференцирование белого шума η .

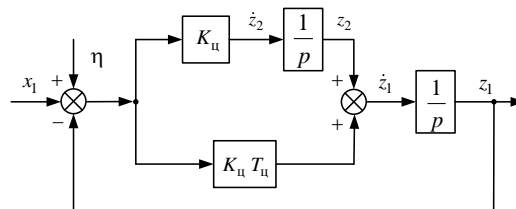


Рис. 1. Структурная схема координатного блока цели

Представим уравнения динамики координатного блока цели в форме Коши

$$\begin{cases} \dot{z}_1(t) = (x_1(t) - z_1(t)) \cdot K_{ц}T_{ц} + z_2(t) + \eta(t) \cdot K_{ц}T_{ц} = (x_1(t) - z_1(t)) \cdot k_1 + z_2(t) + \eta(t) \cdot k_1; \\ \dot{z}_2(t) = (x_1(t) - z_1(t)) \cdot K_{ц} + \eta(t) \cdot K_{ц} = (x_1(t) - z_1(t)) \cdot k_2 + \eta(t) \cdot k_2; \end{cases} \quad (6)$$

где $k_1 = K_{ц}T_{ц}$; $k_2 = K_{ц}$; $\eta(t)$ – гауссовый процесс типа белого шума с интенсивностью R .

Полученную систему уравнений можно записать в векторно-матричной форме:

$$\dot{\mathbf{Z}}(t) = \mathbf{L}\mathbf{Z}(t) + \mathbf{C}\mathbf{X}(t) + \mathbf{N}\eta(t), \quad (7)$$

где $\mathbf{Z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$; $\mathbf{L} = \begin{bmatrix} -k_1 & 1 \\ -k_2 & 0 \end{bmatrix}$; $\mathbf{C} = \begin{bmatrix} k_1 & 0 & 0 \\ k_2 & 0 & 0 \end{bmatrix}$; $\mathbf{N} = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix}$.

Процессы $\xi(t)$ и $\eta(t)$ – некоррелированные белые шумы с нулевыми математическими ожиданиями и корреляционными функциями:

$$M[\xi(t)\xi^T(t')] = Q\delta(t-t'); \quad M[\eta(t)\eta^T(t')] = R\delta(t-t'),$$

где M – знак математического ожидания.

Требуется найти такое управление $W_p(t)$, которое минимизирует функционал

$$J(W_p) = \frac{1}{2} M \left\{ \int_{t_0}^{t_k} [h^2(t) + kW_p^2(t)] dt \right\}, \quad (8)$$

где $k = \frac{h_{\text{доп}}^2}{W_{\text{расп}}^2}$ – коэффициент штрафа на управление; $h_{\text{доп}}$ и $W_{\text{расп}}$ – допустимое значение ошибки наведения, и располагаемое нормальное ускорения БЛА.

Компенсация инерционности измерительного устройства

Для построения оптимального линейного фильтра при инерционном измерителе применим предварительное преобразование выражения (7), введя в рассмотрение преобразованный вектор оценивания [5]:

$$\mathbf{Z}^*(t) = \dot{\mathbf{Z}}(t) - \mathbf{L}\mathbf{Z}(t) = \mathbf{C}\mathbf{X}(t) + \mathbf{N}\eta(t), \quad (9)$$

который в скалярной форме запишется в виде:

$$\begin{cases} \dot{z}_1^*(t) = \dot{z}_1(t) + z_1(t) \cdot k_1 - z_2(t); \\ \dot{z}_2^*(t) = \dot{z}_2(t) + z_1(t) \cdot k_2. \end{cases} \quad (10)$$

Теперь задача аналитического построения фильтра свелась к рассмотренной в [5], но для преобразованного вектора оценивания $\mathbf{Z}^*(t)$.

С учетом (9) уравнения фильтра для рассматриваемой задачи запишутся в виде [5]:

$$\dot{\hat{\mathbf{X}}}(t) = \mathbf{F}\hat{\mathbf{X}}(t) + \mathbf{G}\mathbf{U}(t) + \mathbf{B}(t)(\mathbf{Z}^*(t) - \mathbf{C}\hat{\mathbf{X}}(t)), \quad \hat{\mathbf{X}}(t_0) = \langle X_0 \rangle, \quad (11)$$

или в скалярной форме

$$\begin{cases} \dot{\hat{x}}_1(t) = \hat{x}_2(t) + b_{11}(z_1^*(t) - k_1\hat{x}_1(t)) + b_{12}(z_2^*(t) - k_2\hat{x}_1(t)), & \hat{x}_1(t_0) = \hat{x}_{10}; \\ \dot{\hat{x}}_2(t) = \hat{x}_3(t) - W_p(t) + b_{21}(z_1^*(t) - k_1\hat{x}_1(t)) + b_{22}(z_2^*(t) - k_2\hat{x}_1(t)), & \hat{x}_2(t_0) = \hat{x}_{20}; \\ \dot{\hat{x}}_3(t) = -\frac{1}{T}\hat{x}_3(t) + b_{31}(z_1^*(t) - k_1\hat{x}_1(t)) + b_{32}(z_2^*(t) - k_2\hat{x}_1(t)), & \hat{x}_3(t_0) = \hat{x}_{30}. \end{cases} \quad (12)$$

Матрица коэффициентов усиления $\mathbf{B}(t)$ определяется из условия минимума среднего квадрата ошибки фильтрации [1, 2, 5]:

$$\mathbf{B}(t) = \mathbf{P}(t)\mathbf{C}^T(\mathbf{N}\mathbf{R}\mathbf{N}^T)^{-1}. \quad (13)$$

Входящая в (13) ковариационная матрица ошибок фильтрации $\mathbf{P}(t)$ симметрична, положительно определена и удовлетворяет уравнению Риккати [1, 2, 5]:

$$\dot{\mathbf{P}}(t) = \mathbf{F}\mathbf{P}(t) + \mathbf{P}(t)\mathbf{F}^T - \mathbf{P}(t)\mathbf{C}^T(\mathbf{N}\mathbf{R}\mathbf{N}^T)^{-1}\mathbf{C}\mathbf{P}(t) + \mathbf{J}\mathbf{Q}\mathbf{J}^T, \quad \mathbf{P}(t_0) = \mathbf{P}_0. \quad (14)$$

Анализируя выражения (13), (14) с целью выявления условий компенсации инерционности поступающих на вход фильтра (11) измерений (7), выразим коэффициенты усиления b_{ij} (элементы матрицы $\mathbf{B}(t)$) через параметры системы и характеристики протекающих процессов. Так как матрица

$$\mathbf{N}\mathbf{N}^T = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} \begin{bmatrix} k_1 & k_2 \end{bmatrix} = \begin{bmatrix} k_1^2 & k_1 k_2 \\ k_1 k_2 & k_2^2 \end{bmatrix} \quad (15)$$

вырождена [3, 5], то, умножив левую и правую части (13) на (15), можно найти систему скалярных уравнений:

$$\begin{cases} b_{11} k_1 + b_{12} k_2 = \frac{1}{R} p_{11}; \\ b_{21} k_1 + b_{22} k_2 = \frac{1}{R} p_{12}; \\ b_{31} k_1 + b_{32} k_2 = \frac{1}{R} p_{13}. \end{cases} \quad (16)$$

Покажем, что условия оптимальной фильтрации выполняются независимо от выбора соотношения между b_{i1} и b_{i2} при следующих условиях:

$$b_{i1} k_1 = d p_{1i}; \quad b_{i2} k_2 = (1-d) p_{1i}; \quad d = \text{const}; \quad i = \overline{1,3}. \quad (17)$$

Найдем передаточные функции $K_{z_1 x_1}(p) = \frac{z_1(p)}{x_1(p)}$ и $K_{z_2 x_1}(p) = \frac{z_2(p)}{x_1(p)}$.

Используя преобразование Лапласа, систему уравнений (6) при нулевых начальных условиях преобразуем к виду:

$$\begin{cases} (k_1 + p) z_1(p) - z_2(p) = k_1 x_1(p); \\ k_2 z_1(p) + p z_2(p) = k_2 x_1(p). \end{cases} \quad (18)$$

Решая систему (18) методом Жордана [4], получим

$$\begin{cases} K_{z_1 x_1}(p) = \frac{z_1(p)}{x_1(p)} = \frac{k_1 p + k_2}{p^2 + k_1 p + k_2} = \frac{k_1 p + k_2}{A(p)}; \\ K_{z_2 x_1}(p) = \frac{z_2(p)}{x_1(p)} = \frac{k_2 p}{p^2 + k_1 p + k_2} = \frac{k_2 p}{A(p)}, \end{cases} \quad (19)$$

где $A(p) = p^2 + k_1 p + k_2$.

Линейную комбинацию $b_{i1} z_1^*(p) + b_{i2} z_2^*(p)$ ($i = \overline{1,3}$) в уравнении фильтра (12) с учетом (10), (17) и (19) можно представить в виде:

$$\begin{aligned} b_{i1} z_1^*(p) + b_{i2} z_2^*(p) &= b_{i1} \left(\frac{d z_1(p)}{dt} + k_1 z_1(p) - z_2(p) \right) + b_{i2} \left(\frac{d z_2(p)}{dt} + k_2 z_1(p) \right) = \\ &= \frac{p_{1i}}{2R} \left(\frac{p(k_1 p + k_2)}{k_1 A(p)} + \frac{k_1 p + k_2}{A(p)} - \frac{k_2 p}{k_1 A(p)} + \frac{k_2 p^2}{k_2 A(p)} + \frac{k_1 p + k_2}{A(p)} \right) x_1(p) = \\ &= \frac{p_{1i}}{2RA(p)} (2p^2 + 2k_1 p + 2k_2) x_1(p) = \frac{1}{R} p_{1i} x_1(p). \end{aligned} \quad (20)$$

С учетом (17) и (20) систему уравнений (12) запишем в виде:

$$\begin{cases} \dot{\hat{x}}_1(t) = \hat{x}_2(t) + \frac{p_{11}(t)}{R}(x_1(t) + \eta(t) - \hat{x}_1(t)) = \hat{x}_2(t) + \frac{p_{11}(t)}{R}(z(t) - \hat{x}_1(t)); \\ \dot{\hat{x}}_2(t) = \hat{x}_3(t) - W_p(t) + \frac{p_{12}(t)}{R}(x_1(t) + \eta(t) - \hat{x}_1(t)) = \hat{x}_3(t) - W_p(t) + \frac{p_{12}(t)}{R}(z(t) - \hat{x}_1(t)); \\ \dot{\hat{x}}_3(t) = -\frac{\hat{x}_3(t)}{T} + \frac{p_{13}(t)}{R}(x_1(t) + \eta(t) - \hat{x}_1(t)) = -\frac{\hat{x}_3(t)}{T} + \frac{p_{13}(t)}{R}(z(t) - \hat{x}_1(t)), \end{cases} \quad (21)$$

где $z(t) = x_1(t) + \eta(t)$.

Сравнивая уравнения (21) с аналогичными уравнениями из [6], можно сделать вывод о том, что условия оптимальной фильтрации выполняются независимо от выбора соотношения между b_{i1} и b_{i2} , но при этом должны быть выполнены условия (16).

Другими словами, при выполнении условия (17) и формировании сигнала в соответствии с (21) состояние системы (3) оценивается фильтром [5]:

$$\dot{\hat{\mathbf{X}}}(t) = \mathbf{F}\hat{\mathbf{X}}(t) + \mathbf{G}\mathbf{U}(t) + \mathbf{B}(t)(\mathbf{Z}(t) - \mathbf{C}\hat{\mathbf{X}}(t)), \quad \hat{\mathbf{X}}(t_0) = \mathbf{X}_0, \quad (22)$$

где $\mathbf{Z}(t)$ – вектор измерения (в данном случае это измеренное значение ошибки наведения $z(t) = \hat{h}(t)$, формируемое из состояний $z_1(t), z_2(t)$ инерционного измерителя (7) в соответствии с условиями (17) и (21)).

Таким образом, компенсация инерционности измерительного устройства осуществляется за счет добавления ошибки сопровождения к оценке угловой координаты, что позволяет уменьшить динамическую ошибку разности угловых координат цели и БЛА.

Синтез квазиоптимального контура управления

Для модели (1), (3), (21) и критерия качества (8) оптимальным является управление [6]:

$$W_p(t) = b_1 \hat{x}_1(t) + b_2 \hat{x}_2(t) + b_3 \hat{x}_3(t), \quad t \in [0; t_k]; \quad (23)$$

$$\text{где } b_1 = \frac{1}{\sqrt{k}}; \quad b_2 = 4\sqrt[4]{\frac{4}{k}}; \quad b_3 = \frac{(T + 4\sqrt[4]{4k})T}{T^4\sqrt[4]{4k} + \sqrt{k} + T^2}; \quad k = \frac{h_{\text{доп}}^2}{W_{\text{расп}}^2}.$$

Объединив (21), (23) и перейдя к изображениям по Лапласу при нулевых начальных условиях, получим систему уравнений, описывающих контур управления БЛА:

$$\begin{cases} W_p(p) - b_1 \hat{x}_1(p) - b_2 \hat{x}_2(p) - b_3 \hat{x}_3(p) = 0; \\ (p + M_1) \hat{x}_1(p) - \hat{x}_2(p) = M_1 x_1(p); \\ W_p(p) + M_2 \hat{x}_1(p) + p \hat{x}_2(p) - \hat{x}_3(p) = M_2 x_1(p); \\ M_3 \hat{x}_1(p) + \left(p + \frac{1}{T}\right) \hat{x}_3(p) = M_3 x_1(p). \end{cases} \quad (24)$$

где $M_i = \frac{1}{R} p_{1i}; \quad i = \overline{1,3}$.

Решая систему (24) методом Жордана [4], получим:

$$\begin{aligned} W_p(p) &= x_1(p) \times \\ &\times \frac{(M_1 b_1 + M_2 b_2 + M_3 b_3) p^2 + \left(\frac{M_1 b_1 + M_2 b_2}{T} + M_2 b_1 + M_3 b_2\right) p + M_3 b_1 + \frac{M_2 b_1}{T}}{p^3 + \frac{1 + M_1 T + b_2 T}{T} p^2 + \frac{M_1 + b_2 + b_1 T + M_1 b_2 T + M_2 T}{T} p + \frac{M_2 + M_3 T + b_1 + M_1 b_2 - M_3 b_3 T}{T}} = \\ &= \frac{K_w (T_1^2 p^2 + 2\xi_1 T_1 p + 1)}{a_3 p^3 + a_2 p^2 + a_1 p + 1} x_1(p), \end{aligned} \quad (25)$$

$$\text{где } K_w = \frac{M_3 b_1 T + M_2 b_1}{M_2 + M_3 T + b_1 + M_1 b_2 - M_3 b_3 T} ; a_1 = \frac{M_1 + b_2 + b_1 T + M_1 b_2 T + M_2 T}{M_2 + M_3 T + b_1 + b_2 M_1 - M_3 b_3 T} ;$$

$$a_3 = \frac{T}{M_2 + M_3 T + b_1 + b_2 M_1 - M_3 b_3 T} ; a_2 = \frac{1 + M_1 T + b_2 T}{M_2 + M_3 T + b_1 + b_2 M_1 - M_3 b_3 T} ;$$

$$T_1 = \sqrt{\frac{T(M_1 b_1 + M_2 b_2 + M_3 b_3)}{M_3 b_1 T + M_2 b_1}} ; \xi_1 = \frac{M_1 b_1 + M_2 b_2 + (M_2 b_1 + M_3 b_2) T}{2\sqrt{T(M_1 b_1 + M_2 b_2 + M_3 b_3)(M_3 b_1 T + M_2 b_1)}} .$$

Знаменатель выражения (25) представляет собой полином третьей степени. Поэтому хотя бы один из его корней является действительным числом. В результате передаточная функция квазиоптимального контура телеуправления, обеспечивающая выполнение критерия качества (8), примет вид

$$W_p(p) = \frac{K_w (T_1^2 p^2 + 2\xi_1 T_1 p + 1)}{p^2 (1 + p T_3) (T_2^2 p^2 + 2\xi_2 T_2 p + 1)} x_1(p). \quad (26)$$

Пример. Дано: 1) дисперсия нормального ускорения цели при маневре $\sigma_m^2 = 400 \text{ м}^2/\text{с}^4$, среднее время переключения знака ускорения $T = 5 \text{ с}$; 2) СКО ошибки угловых измерений $\sigma_\varepsilon = 1 \text{ угл. мин}$, время корреляции шума измерений $T_{ш} = 0,06 \text{ с}$; 3) величина допустимого значение ошибки наведения $h_{\text{доп}} = 25 \text{ м}$, располагаемое нормальное ускорения БЛА $W_{\text{расп}} = 30 \text{ м}/\text{с}^2$.

Необходимо определить параметры передаточной функции (26) при различных дальностях до точки встречи r_b . Результаты вычислений приведены в таблице.

Параметры контура управления

Дальность r_b , км	Значения параметров контура управления					
	K_w , с^{-2}	T_1 , с	ξ_1	T_3 , с	T_2 , с	ξ_2
3	6,4	1,26	0,741	4,356	0,171	0,707
5	5,227	1,326	0,745	4,482	0,197	0,707
7	4,541	1,376	0,747	4,555	0,215	0,707
10	3,886	1,435	0,75	4,624	0,236	0,707
15	3,228	1,511	0,754	4,693	0,262	0,707

Как видно из примера, постоянная времени аperiodического звена T_3 приблизительно определяется временем корреляции нормального ускорения цели ($T = 5 \text{ с}$). Дифференцирующее звено второго порядка с постоянной времени T_1 обеспечивает нормальные запасы устойчивости по фазе. Опорная частота колебательного звена с постоянной времени T_2 находится за частотой среза. Логарифмические амплитудно-частотная и фазо-частотная характеристики (ЛАЧХ и ЛФЧХ) контура управления представлены на рис. 2.

Анализ зависимости параметров контура телеуправления при различных дальностях до точки встречи r_b показал, что с увеличением r_b коэффициент преобразования разомкнутого контура управления K_w уменьшается, постоянные времени T_2 увеличиваются, а остальные параметры контура существенно не изменяются.

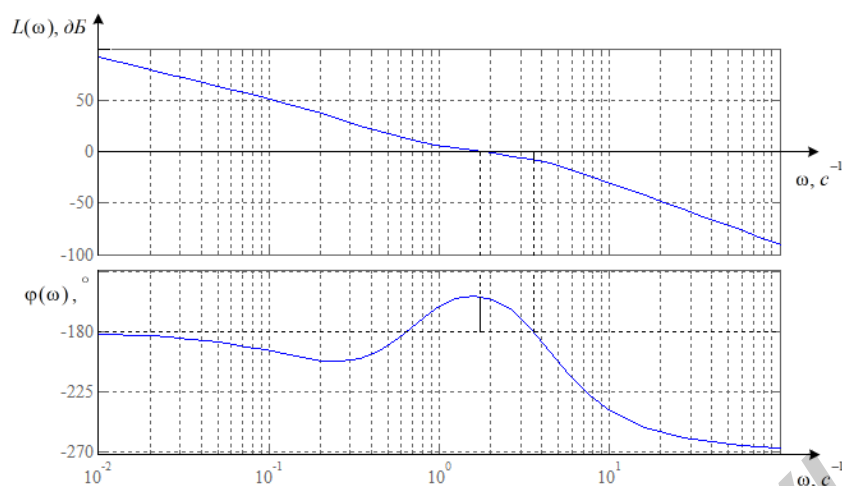


Рис. 2. ЛАЧХ и ЛФЧХ контура управления при $r_B = 10$ км

Заключение

В статье проведен синтез квазиоптимального контура телеуправляемого БЛА с учетом инерционности координатного блока цели. Анализ ЛАЧХ и ЛФЧХ синтезированного контура показал необходимость компенсации инерционности измерительного устройства за счет добавления ошибки сопровождения к оценке угловой координаты. Такой подход позволил избавиться от динамической ошибки разности угловых координат цели и БЛА.

Анализ параметров синтезированного квазиоптимального контура показал, что при полете БЛА на большие дальности необходимо сужать полосу пропускания системы за счет уменьшения коэффициента преобразования контура и увеличения постоянной времени фильтра устройства выработки команд управления.

THE SYNTHESIS OF QUASIOPTIMAL CONTOUR OF THE LONG-DISTANCE DRONE'S COMMAND AND CONTROL WITH THE INERTIA MEASURING DEVICE

A.A. KUN, S.A. SHABAN, A.M. EROMIN, A.L. KOCHEROV

Abstract

The problem of synthesis of a quasioptimal contour of the long-distance drone's command and control considering the inertia of the target coordinate unit is solved. The mean of the compensation of the inertia measuring device is offered. The analysis of the distance influence on the parameters of the long-distance drone's command and control contour before the contact of the drone with the target is carried out.

Список литературы

1. Кун А.А., Лукьянов В.Ф., Шабан С.А. Основы построения систем управления ракетами. Минск, 2001.
2. Орлов Е.В. Проектирование систем телеуправления. Ижевск, 2000.
3. Батков А.М., Тарханов И.Б. Системы телеуправления. М., 1971.
4. Корн Г., Корн Т. Справочник по математике для научных работников и инженеров. М., 1973.
5. Казаков И.Е., Гладков Д.И. Методы оптимизации стохастических систем. М., 1987.
6. Кун А.А., Шабан С.А., Еромин А.М. и др. // Сб. науч. статей Воен. акад. РБ. 2014. № 26. С. 95–102.

УДК 004.056.55

ТЕОРИЯ НОРМ СИНДРОМОВ И ПЛЮС-ДЕКОДИРОВАНИЕ

В.А. ЛИПНИЦКИЙ, А.О. ОЛЕКСЮК

*Военная академия Республики Беларусь
Минск, 220057, Беларусь**Поступила в редакцию 8 октября 2014*

Представлены результаты исследования не примитивных БЧХ-кодов, имеющих декодирующий потенциал, многократно превышающий конструктивные возможности. Показана эффективность автоморфизмов кодов, теории норм синдромов при коррекции всех допустимых минимальным расстоянием ошибок в названных кодах.

Ключевые слова: помехоустойчивое кодирование, БЧХ-коды, конструктивное и минимальное расстояния кода, циклические подстановки, теория норм синдромов.

Введение

Помехоустойчивое кодирование призвано вести синхронную борьбу с ошибками при передаче информации в цифровых системах связи, вынужденных функционировать в реальных зашумленных каналах [1, 2]. Реалии информационной эпохи XXI века, всеобщая компьютеризация, стремительный рост широчайших потоков информации предъявляют различные и порой противоречащие друг другу требования к применяемым кодам, создавая ряд проблем в теории и практике помехоустойчивого кодирования. Наиболее острая из них – проблема «селектора» - проблема быстрого и надежного выбора конкретной ошибки среди обширного корректируемого многообразия этих ошибок [3, 4].

Корректирующий потенциал кода обеспечивают спектры ошибок с попарно различными синдромами. Однако прямые развязки типа «синдром - ошибка» эффективны лишь при исправлении одиночных ошибок.

Применяемые на практике коды нередко обладают корректирующим потенциалом, выходящим за рамки конструктивных возможностей. Сказанное наиболее характерно для популярнейшего на практике класса кодов Боуза-Чоудхури-Хоквингема (БЧХ-кодов), особенно не примитивных БЧХ-кодов [2, 5]. Для реализации названного потенциала стандартные средства неприменимы. Здесь требуется разработка принципиально новых подходов, применение новых идей. Круг вопросов, связанных с коррекцией ошибок, выходящих за рамки конструктивных возможностей кодов, предлагаем назвать «плюс-декодированием».

Теория норм синдромов (ТНС) [4, 5], на порядок снизившая влияние проблемы «селектора», предоставляет также и конструктивные подходы к решению проблемы «плюс-декодирования». Их обсуждению, особенно в применении к не примитивным БЧХ-кодам, рассчитанным на коррекцию двукратных ошибок, и посвящена данная работа.

О БЧХ-кодах с конструктивным расстоянием 5

Общее определение и основные свойства БЧХ-кодов приведены в монографии [2]. Среди БЧХ-кодов с конструктивным расстоянием $2t+1$, рассчитанных на исправление t – кратных случайных ошибок, наибольшую размерность и скорость передачи информации, а, следовательно, и наибольший практический интерес имеют коды C_{2t+1} с проверочной матрицей

$$H = (\beta^i, \beta^{3i}, \dots, \beta^{2t-1})^T. \quad (1)$$

Здесь β – элемент мультипликативной группы $GF(2^m)^*$ поля Галуа $GF(2^m)$, имеющий порядок $n = (2^m - 1) / \tau$ для некоторого делителя τ числа $|GF(2^m)^*| = 2^m - 1$, $0 \leq i \leq n - 1$. Длина кода C_{2t+1} равна n и всегда является нечетной величиной. Группа $GF(2^m)^*$ циклическа. Если α – образующая этой группы – примитивный элемент поля $GF(2^m)$, то в качестве β можно взять $\beta = \alpha^\tau$. При $\tau = 1$ длина $n = 2^m - 1$, элемент $\beta = \alpha$ и потому код C_{2t+1} называется примитивным; если же $\tau > 1$, $\beta \neq \alpha$ и потому код C_{2t+1} называют не примитивным [5, 6].

БЧХ-коды C_5 – частный случай кодов с проверочной матрицей (1) – задаются проверочными двоичными $(2m \times n)$ -матрицами, $2m < n = (2^m - 1) / \tau$,

$$H = (\beta^i, \beta^{3i})^T. \quad (2)$$

В [2] доказано, что у примитивных кодов C_5 при $m \geq 4$ ранг матрицы (2) равен $2m$, циклотомические классы $C^1 \neq C^3$, их размерность $k = n - 2m$, а минимальное расстояние $d = \delta = 5$. У не примитивных кодов C_5 любое из названных соотношений может нарушиться.

Ранг подматрицы β^i равен m [5]. Ранг подматрицы $(\beta^{3i}) = (1, \beta^3, \beta^6, \dots, \beta^{3(n-1)})$ матрицы (2) и подматрицы (1) не всегда равен m . Действительно, из каждых трех последовательных нечетных значений n одно делится на 3, а два – не делятся на три. Если $\text{НОД}(3, n) = 1$, то отображение φ_3 циклической группы $\{1, \beta, \beta^2, \dots, \beta^{n-1}\}$ порядка n в себя, действующее по правилу $\varphi_3(x) = x^3$ является автоморфизмом этой группы. Отсюда следует, в частности, что β^3 имеет тот же порядок в мультипликативной группе $GF(2^m)^*$, что и его прообраз β при отображении φ_3 . Это означает, что неприводимый полином над $GF(2)$ с корнем β^3 должен иметь степень m , а следовательно, элементы $1, \beta^3, \beta^6, \dots, \beta^{3(m-1)}$ образуют линейно независимую над $GF(2)$ систему. Следовательно, $\text{rank}(\beta^{3i}) = m$.

Пусть n делится на три. Здесь возможны два варианта: 1) элемент β^3 поля $GF(2^m)$ не принадлежит никакому подполю этого поля; 2) существует подполе $GF(2^\mu)$ поля $GF(2^m)$, содержащее β^3 . В первом случае неприводимый над полем $GF(2)$ полином элемента β^3 должен иметь степень m , а тогда $\text{rank}(\beta^{3i}) = m$. Во втором случае по тем же причинам $\text{rank}(\beta^{3i}) = \mu$ для некоторого делителя μ числа m . Второй случай хорошо иллюстрирует следующий пример. БЧХ-код C_5 длиной 219 определен над полем $GF(2^{18})$. $2^{18} - 1 = 7 \cdot 9 \cdot 57 \cdot 73$. В этом случае матрица (2) задается элементом $\beta = \alpha^{57 \cdot 21} = \alpha^{1197}$, а элемент $\beta^3 = \alpha^{57 \cdot 21 \cdot 3} = \alpha^{57 \cdot 7 \cdot 9}$ имеет порядок 73. Но такой элемент должен принадлежать мультипликативной группе $GF(2^9)^*$, которая имеет порядок $2^9 - 1 = 7 \cdot 73$. В таком случае $\text{rank}(\beta^{3i}) \leq 9$.

Элементы β и β^3 в коде C_5 не должны быть сопряженными в поле $GF(2^m)$, что, очевидно, эквивалентно неравенству $C^1 \neq C^3$ циклотомических классов. И это свойство не всегда выполняется. К примеру, у БЧХ-кода C_5 длиной 95 имеет место совпадение циклотомических классов: $C^1 = C^3$; данный код реально относится к классу кодов Хемминга.

Во всех конкретных случаях проверка каждого параметра не примитивного кода C_5 требует внимания, дополнительных вычислений, а порой и серьезных компьютерных ресурсов.

Потенциал плюс-декодирования для БЧХ-кодов C_5

Ранее проблема декодирования ошибок, выходящих за конструктивные рамки, имела частный характер, поскольку количество таких ошибок было незначительным (см. [7, 8]). В

случае не примитивных БЧХ-кодов картина резко меняется. БЧХ-код C_5 конструктивно рассчитан на исправление одиночных и двойных ошибок в количестве $K_{\text{констр}} = C_n^1 + C_n^2 = n(n+1)/2$. Если у данного кода минимальное расстояние $d=7$ (в табл. 1 отмечено пять таких кодов), то код должен исправлять и тройные ошибки в количестве $C_n^3 = n(n-1)(n-2)/6$. Данное количество ошибок превосходит почти в $n/3$ раз $K_{\text{констр}}$ и составляет потенциал K^+ плюс-декодирования для рассматриваемых кодов. Если же у кода C_5 реальное значение $d=9$ (в табл. 1 отмечено пять таких кодов), то $K^+ = C_n^3 + C_n^4 = (n+1)n(n-1)(n-2)/24$, что почти в $n^2/12$ раз превосходит $K_{\text{констр}}$ и т.д. Точные данные приведены ниже в табл. 1. Из нее следует, что на плюс-декодирование приходится в десятки тысяч раз больше векторов-ошибок, чем на конструктивное.

ТНС как средство реализации возможностей плюс-декодирования

Конструктивные подходы к реализации в БЧХ-кодах C_5 потенциала плюс-декодирования предоставляет теория норм синдромов [4, 5]. Данная теория опирается на цикличность БЧХ-кодов с проверочными матрицами (1) и (2). Группам автоморфизмов $Aut(C)$ названных кодов C принадлежит подгруппа Γ порядка n (длина кода) циклических сдвигов, которая состоит из степеней линейного преобразования σ двоичного векторного пространства V_n , действующего на каждый вектор $\bar{x} = (x_1, x_2, \dots, x_n) \in V_n$ по правилу

$$\sigma(x_1, x_2, \dots, x_n) = (x_n, x_1, x_2, \dots, x_{n-1}). \quad (3)$$

Под действием группы Γ пространство V_n , а с ним и совокупность K_C декодируемых конкретным БЧХ-кодом C векторов-ошибок, разбивается на попарно непересекающиеся классы – Γ -орбиты. Каждая Γ -орбита состоит из переходящих друг в друга под действием степеней σ векторов. Поэтому каждая Γ -орбита J имеет следующую стандартную структуру:

$$J = \{\bar{e}, \sigma(\bar{e}), \dots, \sigma^{\mu-1}(\bar{e})\} \quad (4)$$

для произвольного фиксированного вектора $\bar{e} \in J$. Здесь μ – наименьшее натуральное число с условием: $\sigma^{\mu}(\bar{e}) = \bar{e}$, делитель длины кода n . Чаще всего $\mu = n$. В последнем случае Γ -орбита содержит максимально возможное количество векторов и потому называется полной. Равенство (4) служит причиной для следующего, более точного обозначения Γ -орбит: $J = \langle \bar{e} \rangle$.

Отображение Φ_H из двоичного пространства V_n в двоичное пространство V_{2m} , действующее по правилу $\bar{y} = \bar{x} \cdot H^T$, есть линейный оператор между названными пространствами. Согласно базовым результатам линейной алгебры, полный образ $\Phi_H(V_n)$ есть подпространство пространства V_{2m} размерностью $n - \dim Ker H = n - k = 2m$. Это означает, что $\Phi_H(V_n) = V_{2m}$.

Каждая вектор-ошибка \bar{e} в БЧХ-коде C_5 над полем $GF(2^m)$ имеет в силу формулы (2) синдром $S(\bar{e}) = H \cdot \bar{e}^T = (s_1, s_2)^T$ для некоторых $s_1, s_2 \in GF(2^m)$. Из равенства $\Phi_H(V_n) = V_{2m}$ следует, что для любых $s_1^*, s_2^* \in GF(2^m)$ найдется вектор $\bar{e} \in V_n$, такой, что $S(\bar{e}) = H \cdot \bar{e}^T = (s_1^*, s_2^*)^T$.

Действие оператора σ на каждый вектор \bar{e} однозначно отражается в коде C_5 на синдромах по формуле [4, 5]:

$$S(\sigma(\bar{e})) = (\beta \cdot s_1, \beta^3 \cdot s_2)^T. \quad (5)$$

Формула (5) определяет структуру спектра синдромов $S(\langle \bar{e} \rangle)$ каждой Γ -орбиты $\langle \bar{e} \rangle$:

$$S(\langle \bar{e} \rangle) = \{\beta^i \cdot s_1, \beta^{3i} \cdot s_2\}, \quad 0 \leq i \leq n-1. \quad (6)$$

Из формул (5), (6) следует, что, если у двух Γ -орбит нашлись векторы с одинаковыми синдромами, то спектры синдромов этих орбит совпадают полностью.

Формула (5) послужила отправной точкой для следующего определения.

Нормой синдрома $S(\bar{e}) = (s_1, s_2)^T$ в БЧХ-коде C_5 называется величина

$$N = N(S(\bar{e})) = \begin{cases} s_2 / s_1^3; & s_1 \neq 0; \\ +\infty; & s_1 = 0, s_2 \neq 0. \end{cases} \quad (7)$$

Норма синдрома не определена для единственного нулевого значения синдрома: $S(\bar{e}) = (0, 0)^T$. Нормы синдромов обладают рядом замечательных свойств. Очевидно, норма может быть любым элементом поля $GF(2^m)$, а также имеет одно особое значение: $+\infty$. Всего, таким образом, норма принимает $n+2=2^m+1$ значений. Если две Γ -орбиты J_1 и J_2 имеют различные нормы, то спектры синдромов этих орбит не пересекаются, иными словами, данные орбиты не могут содержать векторы с одинаковыми синдромами.

К наиглавнейшим свойствам норм синдромов относится следующее: у всех векторов, принадлежащих каждой отдельно взятой Γ -орбите J норма синдрома одинакова. Это единственное значение естественно назвать нормой $N(J)$ данной Γ -орбиты J . Таким образом, $N(J) = N(S(\bar{e}))$ для произвольного вектора $\bar{e} \in J$.

Приведем еще одно достаточно важное свойство норм синдромов – синдромы равномерно распределены по значениям норм синдромов: для каждого из $n+2=2^m+1$ значений N норм синдромов найдется в точности $n=2^m-1$ различных синдромов, норма которых равна N . Действительно, если у синдрома $S(\bar{e}) = (s_1, s_2)^T$ компонента $s_1 \neq 0$ и $N(S(\bar{e})) = N$, то для примитивного элемента α поля $GF(2^m)$ различные n синдромов $(\alpha^i \cdot s_1, \alpha^{3i} \cdot s_2)^T$, $0 \leq i \leq n-1$, принимают то же самое значение нормы. Для всех $n=2^m-1$ синдромов вида $(0, s_2)^T$, где $s_2 \neq 0$, $s_2 \in GF(2^m)$ норма равна $+\infty$. В целом, таким образом мы учли уже $n(n+2) = (2^m-1) \cdot (2^m+1) = 2^{2m}-1$ синдромов. Добавив к ним нулевой синдром, мы получим весь спектр возможных синдромов в БЧХ-коде C_5 , что и завершает доказательство свойства равномерного распределения синдромов по значениям норм синдромов.

У примитивных БЧХ-кодов из равномерного свойства вытекает еще одно важное свойство норм синдромов – из равенства $N(J_1) = N(J_2)$ норм синдромов двух полных Γ -орбит J_1 и J_2 с полными спектрами синдромов следует и равенство самих синдромов: для каждого вектора $\bar{f} \in J_1$ найдется вектор $\bar{g} \in J_2$, такой, что их синдромы равны $S(\bar{f}) = S(\bar{g})$.

Для не примитивных БЧХ-кодов ситуация сложнее. Полная Γ -орбита J с полным спектром синдромов $S(J)$ содержит $n = (2^m-1)/\tau$ векторов с $n = (2^m-1)/\tau$ синдромами в $S(J)$. Отсюда и из сформулированного выше равномерного свойства следует, что существует τ различных полных Γ -орбит с попарно непересекающимися полными спектрами синдромов и с одинаковой нормой.

Теория норм синдромов дает новый взгляд на коррекцию ошибок. Она предлагает иметь дело не с отдельными векторами ошибок, а с их Γ -орбитами. Любую декодируемую кодом C совокупность K_C векторов-ошибок, обязательно имеющих попарно различные синдромы, можно разбить на множество K_C/Γ Γ -орбит этих ошибок. Для задания каждой конкретной Γ -орбиты J достаточно зафиксировать один из ее представителей \bar{e}_j . Остальные векторы орбиты

легко строятся действием группы Γ – циклическими сдвигами координат вектора \bar{e}_j .

Все Γ -орбиты J декодируемой данным кодом C совокупности K_C векторов-ошибок имеют попарно непересекающиеся спектры синдромов $S(J)$. Как уже отмечалось, спектр $S(J)$ однозначно восстанавливается по формуле (5) из синдрома $S(\bar{e}_j)$.

Имея список 1 образующих \bar{e}_j совокупности K_C / Γ , список 2 синдромов $S(\bar{e}_j)$, а также список 3 норм $N(S(\bar{e}_j))$, можно легко определять «ряд и место» подлежащей определению вектор-ошибки \bar{e} в каждом очередном сообщении \bar{x} , принятом ТКС на основе кода C .

Алгоритм 1 – математический алгоритм норменного декодирования с помощью Γ -орбит

ТКС, получив очередное сообщение \bar{x} , в обязательном порядке вычисляет синдром $S(\bar{x}) = S(\bar{e})$. Если $S(\bar{x}) = S(\bar{e}) \neq \bar{0}$, что однозначно свидетельствует о наличии неизвестной и подлежащей определению вектор-ошибки \bar{e} , то вычисляем $N^* = N(S(\bar{x}))$. Совпадение N^* с $N(S(\bar{e}_j))$ из списка 3 резко сужает круг Γ -орбит декодируемой совокупности, которые могут содержать искомую вектор-ошибку \bar{e} из сообщения \bar{x} , до небольшой группы орбит $J_1, J_2, \dots, J_\theta$, $1 \leq \theta \leq \tau$, с одинаковой нормой N^* .

Синдром $S(\bar{x}) = S(\bar{e}) = (s_1, s_2)$ принадлежит спектру синдромов только одной Γ -орбиты J_l , $1 \leq l \leq \theta$. Пусть N^* является элементом поля Галуа $GF(2^m)$. Тогда у всех рассматриваемых синдромов первая компонента отлична от нуля. Пусть синдром образующей \bar{e}_{j_l} Γ -орбиты J_l имеет первую компоненту $s_1^{j_l} = \alpha^v$ для некоторого целого v , $0 \leq v < n$. Пусть у синдрома $S(\bar{x})$ первая компонента $s_1 = \alpha^\lambda$, $0 \leq \lambda < n$. Тогда для подходящего целого i , $0 \leq i < n$, в силу формулы (6), $\alpha^\lambda = \beta^i \cdot \alpha^v = \alpha^{\tau i + v}$. Полученное равенство означает, что либо $\lambda - v$ (если $\lambda > v$), либо $2^m - 1 + \lambda - v$ (если $\lambda < v$) делится на τ и частное i однозначно определяет искомую вектор-ошибку: $\bar{e} = \sigma^i(\bar{e}_{j_l})$. Величина же l и есть то единственное число из множества целых $\{1, 2, \dots, \theta\}$, для которого одна из величин $\lambda - v$ или $2^m - 1 + \lambda - v$ делится нацело на τ .

Пусть $N^* = +\infty$. Тогда у всех рассматриваемых синдромов первая компонента равна нулю. Пусть синдром образующей \bar{e}_{j_l} Γ -орбиты J_l имеет вторую компоненту $s_2^{j_l} = \alpha^v$ для некоторого целого v , $0 \leq v < n$. Пусть у синдрома $S(\bar{x})$ вторая компонента $s_2 = \alpha^\lambda$, $0 \leq \lambda < n$. Тогда для подходящего целого i , $0 \leq i < n$, в силу формулы (6), $\alpha^\lambda = \beta^{3i} \cdot \alpha^v = \alpha^{\tau 3i + v}$. Полученное равенство означает, что либо $\lambda - v$ (если $\lambda > v$), либо $2^m - 1 + \lambda - v$ (если $\lambda < v$) делится на τ . Величина l и есть то единственное число из множества целых $\{1, 2, \dots, \theta\}$, для которого одна из величин $\lambda - v$ или $2^m - 1 + \lambda - v$ делится нацело на τ . Очевидно, этим целым частным будет число $3i$.

Пусть n не делится на 3. Для взаимно простых чисел n и 3 выполняется соотношение Безу, то есть существуют такие целые числа u и v , что $3u + nv = 1$. Искомую вектор-ошибку \bar{e} определяем формулой: $\bar{e} = \sigma^{3ui}(\bar{e}_{j_l})$.

Действительно, $\sigma^{3ui}(\bar{e}_{j_l}) = \sigma^{3iu + nvi}(\bar{e}_{j_l}) = \sigma^{(3u + nv)i}(\bar{e}_{j_l}) = \sigma^i(\bar{e}_{j_l}) = \bar{e}$. Ситуация: $N^* = +\infty$ и n делится на 3 – крайне редкая, присуща неполным Γ -орбитам и рассматривается в индивидуальном порядке.

Итак, работа норменного декодера достаточно легко реализуется, если созданы списки 1 – 3, характеризующие Γ -орбиты декодируемой совокупности векторов-ошибок.

Эффективность работы норменных декодеров особенно ярко наблюдается на примитивных БЧХ-кодах [5, 9]. Отличительным и несколько усложняющим фактором непримитивного случая является возможное наличие отдельных значений $\theta > 1$.

Циклотомические подстановки для норменного декодирования

Следует признать, что при $d > 7$ списки 1 – 3 становятся достаточно обширными, а работа с ними – затруднительной – проблема «селектора» начинает проявлять себя на новом уровне. Эффективным в преодолении названных затруднений мог бы быть метод «сжатия» – преобразования исправляемых векторов-ошибок в ошибки с узким спектром значений норм синдромов [10]. Однако имеющиеся подходы рассчитаны на примитивные коды и ошибки конкретного веса, автоматически эти подходы не переносятся на ошибки других весов. Для не примитивных же кодов они попросту не применимы.

Богатство группы автоморфизмов кода остается наиболее реальным и конструктивным средством сжатия обрабатываемой декодерами информации. Группа автоморфизмов любого из кодов C_5 достаточно богата, содержит, к примеру, циклотомические подстановки. Их действие, свойства и применение частично уже рассматривались в [5].

Циклотомические подстановки составляют циклическую группу Φ порядка m с образующей ϕ , такой, что для каждого вектора-ошибки \bar{e} с синдромом $S(\bar{e}) = (s_1, s_2)$ синдром $S(\phi(\bar{e})) = (s_1^2, s_2^2)$. Тогда при условии $N(S(\bar{e})) = N \in GF(2^m)$ норма $N(S(\phi(\bar{e}))) = N^2$. Вектор $\phi(\bar{e})$ получается из вектора \bar{e} по правилу: для каждого целого $i, 1 \leq i \leq n$, i -я координата вектора \bar{e} становится $(2i-1)$ -й координатой вектора $\phi(\bar{e})$, если $2i-1 \leq n$, и $(2i-1-n)$ -й координатой вектора $\phi(\bar{e})$, если $2i-1 > n$.

Циклическая подстановка σ и циклотомическая подстановка ϕ связаны соотношением: $\phi\sigma = \sigma^3\phi$ [2, 5]. Вместе они образуют некоммутативную группу G порядка mn , подгруппу группы $Aut(C_5)$. Если J – некоторая Γ -орбита векторов-ошибок, то $\phi(J)$ – новая Γ -орбита векторов-ошибок [5]. Таким образом, группа Φ действует на множестве Γ -орбит K_C / Γ декодируемой кодом $C = C_5$ совокупности K_C векторов-ошибок, разбивая его на Φ -орбиты. Соответственно, множество K_C разбивается на G -орбиты, содержащие, как правило, по mn векторов-ошибок. Зафиксировав одну вектор-ошибку \bar{e} , можно восстановить все вектор-ошибки G -орбиты $\langle \bar{e} \rangle_G$. Таким образом, списки 1–3 можно сократить примерно в m раз, оставив по одной образующей каждой G -орбиты декодируемой совокупности (табл. 1).

Таблица 1. Оценка количества Γ -орбит и G -орбит корректируемой совокупности для БЧХ-кодов C_5

№п/п	1	2	3	4	5	6
N	33	39	43	49	57	69
M	10	12	14	21	18	22
D	9	10	13	7	9	7
$K_{\text{констр}}$	561	780	946	1225	1653	2415
K^+	46376	91390	7194803	18424	424770	52394
$\Gamma_{\text{констр}}$	17	20	22	25	29	35
Γ^+	1405	2344	167321	376	7453	760
G^+	141	196	11952	18	415	35
№п/п	7	8	9	10	11	12
N	73	77	87	89	91	99
M	9	30	28	11	12	30
D	7	7	9	9	7	9
$K_{\text{констр}}$	2701	3003	3828	4005	4186	4950
K^+	62196	73150	2331890	2555190	121485	3921225
$\Gamma_{\text{констр}}$	37	39	44	45	46	50
Γ^+	852	950	2680	28710	1335	39609
G^+	95	32	958	2610	112	1321

Алгоритм 2 – математический алгоритм коррекции ошибок с помощью G -орбит

Этот алгоритм является некоторой модификацией алгоритма 1. Если для принятого сообщения \bar{x} с синдромом $S(\bar{x}) = S(\bar{e}) = (s_1, s_2)$ норма $N^* = N(S(\bar{x}))$ является элементом поля Галуа $GF(2^m)$, но не принадлежит третьему списку, то находим такое наименьшее целое i , $1 \leq i \leq m$, что для $\lambda = 2^i$ величина $(N^*)^\lambda = N(S(\bar{e}_j))$ для одного или нескольких векторов \bar{e}_j из списка 1 образующих G -орбит декодируемой совокупности. Полученное равенство норм означает, что для искомого вектора-ошибки \bar{e} в сообщении \bar{x} вектор $\phi^i(\bar{e})$ имеет синдром $S(\phi^i(\bar{e})) = (s_1^\lambda, s_2^\lambda)$, принадлежащий спектру синдромов $S(\langle \bar{e}_j \rangle)$ одной Γ -орбиты $\langle \bar{e}_j \rangle$, порожденной конкретным вектором-ошибкой \bar{e}_j из списка 1. Алгоритм 1 однозначно определяет вектор \bar{e}_j и находит выражение вектора $\phi^i(\bar{e})$ через него: $\phi^i(\bar{e}) = \sigma^s(\bar{e}_j)$ для подходящего целого s , $0 \leq s \leq n-1$. Далее, вектор \bar{e} однозначно восстанавливается в соответствии с формулой: $\bar{e} = \phi^{m-i}(\sigma^s(\bar{e}_j))$.

Заключение

Примерно треть не примитивных БЧХ-кодов имеют декодирующий потенциал, многократно превышающий конструктивные возможности. Применение автоморфизмов кодов, теория норм синдромов обеспечивают эффективные конструктивные норменные методы коррекции всех допустимых минимальным расстоянием ошибок в названных кодах. Перечисленные факторы обеспечивают перспективность для приложений многих представителей класса не примитивных БЧХ-кодов.

THEORY OF NORMAL SYNDROME AND PLUS-DECODING

V.A. LIPNITSKI, A.O. ALIAKSIUK

Abstract

The results of the study are not primitive BCH codes with decoding the guide, the potential is much greater than the design possibilities. The efficiency of automorphisms of codes, norms theory syndromes in the correction of all admissible-Mykh minimum distances errors in the above code is shown.

Список литературы

- 1 Шеннон, К. Работа по теории информации и кибернетике. М., 1963.
- 2 Мак-Вильямс Ф.Дж., Слоэн Н. Дж.А. Теория кодов, исправляющих ошибки. М., 1979.
- 3 Колесник В.Д., Мирончиков Е.Т. Декодирование циклических кодов. М., 1968.
- 4 Конопелько В.К., Липницкий В.А. Теория норм синдромов и перестановочное декодирование помехоустойчивых кодов. М., 2004.
- 5 Липницкий В.А., Конопелько В.К. Норменное декодирование помехоустойчивых кодов и алгебраические уравнения. Минск, 2007.
- 6 Курилович А.В., Липницкий В.А., Михайловская Л.В. // Технологии информатизации и управления. 2011. Вып. 2. С. 43–49.
- 7 Сагалович Ю.Л. // Тез. докл. IX Симпозиума по проблеме избыточности в информационных системах. Ленинград, 3–8 июня 1986 г. С. 135–138.
- 8 Давыдов А.А., Дрожжина-Лабинская А.Ю., Калинин В.В. // Вопросы кибернетики. Проблемы программного обеспечения супер-ЭВМ. 1990. С. 151–174.
- 9 Конопелько В.К. Устройство декодирования для коррекции двоичных ошибок / Патент СССР SU1833968 A1.
- 10 Липницкий В.А., Аль-Хайдар Е.К. // Докл. БГУИР. 2009. № 5 (43). С. 12–16.

УДК 681.396.96

СПОСОБ РЕАЛИЗАЦИИ ИНФОРМАЦИОННОЙ МОДЕЛИ ОБСТАНОВКИ В АВТОМАТИЗИРОВАННОЙ СИСТЕМЕ УПРАВЛЕНИЯ

Д.И. РАБЧЕНОК, С.В. КРУГЛИКОВ

Военная академия Республики Беларусь
Независимости, 220, Минск, 220057, Беларусь

Поступила в редакцию 2 декабря 2013

Предложен способ нахождения количественной оценки числа иерархических уровней управления информационной моделью автоматизированной системы управления (АСУ) с целью оптимизации загруженности лица, принимающего решение.

Ключевые слова: автоматизированная система управления, информационная модель, интерфейс, иерархическая структура, переборный алгоритм, функционал стоимости, оптимальное управление.

Введение

Одной из задач автоматизированного управления, направленной на повышение эффективности боевых действий и сокращение цикла управления, является формирование информационной модели, отражающей своим содержанием и объемом реальную действительность в требуемом районе.

При этом информационная модель должна соответствовать задачам и возможностям оператора, а также обеспечивать оптимальный информационный баланс предоставляемой ему информации.

Проанализировав работы по проблемам математического моделирования [1–5], можно отметить, что имеющиеся в настоящее время информационные модели автоматизированных систем управления касаются, как правило, отдельных аспектов их функционирования. Отсутствует унифицированный методологический подход к построению информационных моделей АСУ тактического звена как многоуровневых систем со сложной иерархической структурой.

В [4] предложены общие методы синтеза оптимальных иерархических структур. Опираясь на данную монографию и предположив, что описанные методы будут справедливы и для построения информационных моделей, описывающих такие иерархические структуры, необходимо отметить следующее.

Структурный синтез в данном контексте сводится к построению и оптимизации иерархии управления информационной моделью с целью максимизации эффективности либо минимизации затрат на функционирование. Такую задачу решают в режиме статической оптимизации. Но, так как внешние условия изменяются, то очевидна необходимость оптимизации иерархической структуры информационной модели в динамическом режиме.

Таким образом, необходимо рассмотреть частную задачу оптимизации иерархической структуры информационной модели АСУ, представленной лицу, принимающему решение (ЛПР) на автоматизированном рабочем месте, как выбор оптимального числа уровней иерархии управления информационной моделью в зависимости от динамики изменений внешних условий.

Выбор исходных данных

Полагаем, что полученные результаты не будут зависеть от содержательной интерпретации и будут касаться только иерархической структуры информационной модели АСУ, как таковой.

Информационную модель АСУ тактического уровня можно представить в виде множества элементов подсистемы отображения обстановки или как их иерархическую структуру в виде ориентированного ациклического аддитивного графа G организации над множеством объединенных элементов отображения-управления (ЭОУ) N с определенным функционалом стоимости такой организации (рис. 1).

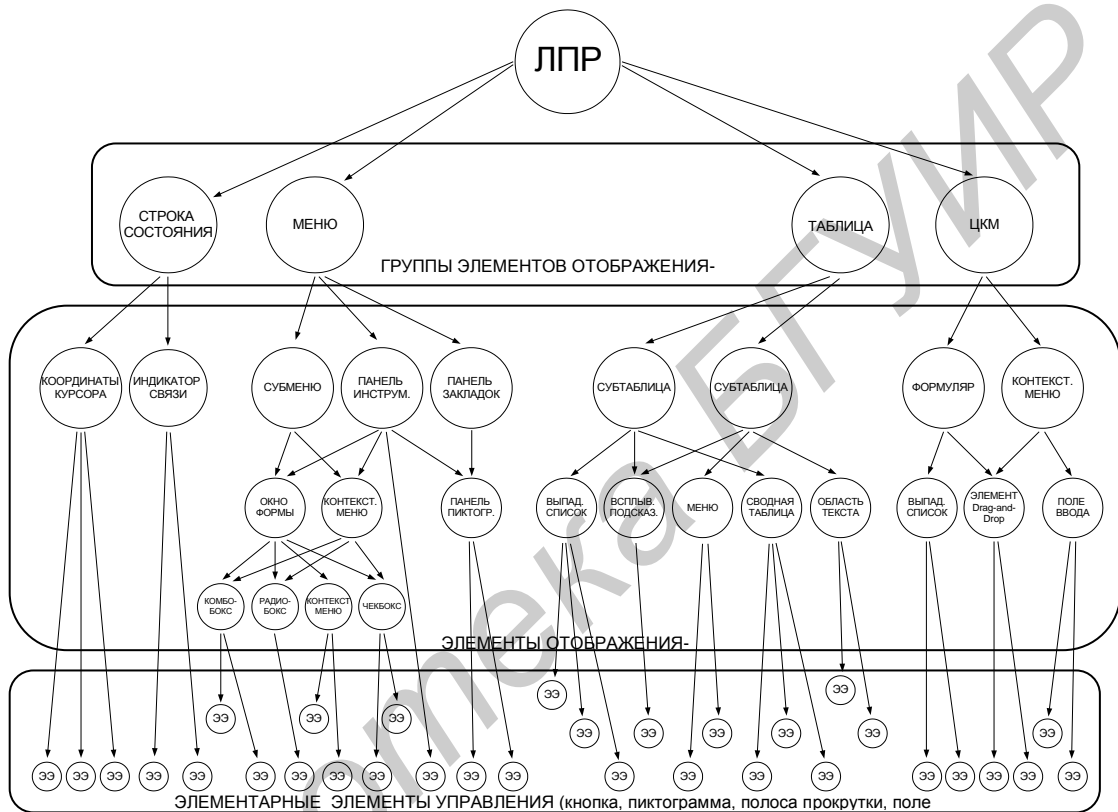


Рис. 1. Представление информационной модели АСУ тактического уровня в виде графа над множеством элементов подсистемы отображения обстановки

Под ребрами такого графа можно понимать управленческие связи, наделенные некоторыми весовыми коэффициентами: в вершинах будут находиться группы ЭОУ, объединенные ЭОУ непосредственно и элементарные элементы управления (ЭЭУ), обладающие своими весовыми коэффициентами.

Функционал стоимости $P(G)$ такого графа можно представить в виде суммы значений стоимости управления отдельными объединенными ЭОУ (1):

$$P(G) = \sum_{m=1}^{N_{\max}} P(N_m), \quad (1)$$

где $P(N_m)$ – стоимость управления объединенными ЭОУ N_m , определяется экспериментально, N_m – количество объединенных ЭОУ, N_{\max} – максимальное, для данных условий, количество объединенных ЭОУ, выбирается эмпирически.

Необходимо отметить, что в таком функционале организация новых подгрупп из элементов N_m , либо удаление существующих подгрупп с одновременной организацией на их основе новых подгрупп, не приведет к уменьшению значения функционала (1) [4]. Это утверждение вытекает из свойства аддитивности графа.

Понятие ЭОУ целесообразно ввести в связи с тем, что на элементе подсистемы отображения в большинстве случаев присутствуют ЭЭУ, в том числе управления и им самим. Появление в поле зрения оператора такого элемента подразумевает не только его визуальный анализ, но и управление им, используя содержащиеся на нем же самые элементы управления.

Под множеством конечных исполнителей подсистемы отображения автоматизированного рабочего места комплекса средств автоматизации пункта управления будем понимать множество объединенных ЭОУ:

$$N_m = \{a_1, a_2, K, a_4, \dots, a_n\}, \quad (2)$$

где K – элементарный управляющий элемент ЭОУ N_m , а $a_1, a_2, a_4, \dots, a_n$ – ЭЭУ, принадлежащие множеству $A = \{a_1, a_2, a_4, \dots, a_n\}$, их количество также выбирается эмпирически.

Множеством групп объединенных ЭОУ назовем множество всевозможных непустых конечных подмножеств f , составленных из N_m . Через

$$f^t = \{f_1^t, f_2^t, f_3^t, K, f_5^t, f_6^t, \dots, f_i^t\} \quad (3)$$

обозначим набор групп объединенных ЭОУ (их количество выбирается эмпирически), которые должны быть организованы системой на протяжении интервала времени t для выполнения подсистемой отображения своих функций. Набор организуемых групп элементов или выполняемых функций может изменяться при изменении t , что влечет необходимость изменения структуры системы отображения.

Иерархией над множеством элементарных элементов A назовем дерево организации набора $D(f^t)$, в узлах которого располагаются множества f_i^t или группы объединенных ЭОУ N_m , а корнем является некоторый набор групп f^t .

На рис. 2 изображен фрагмент иерархии информационной модели АСУ и фрагмент предлагаемой иерархии в терминологии излагаемого материала.

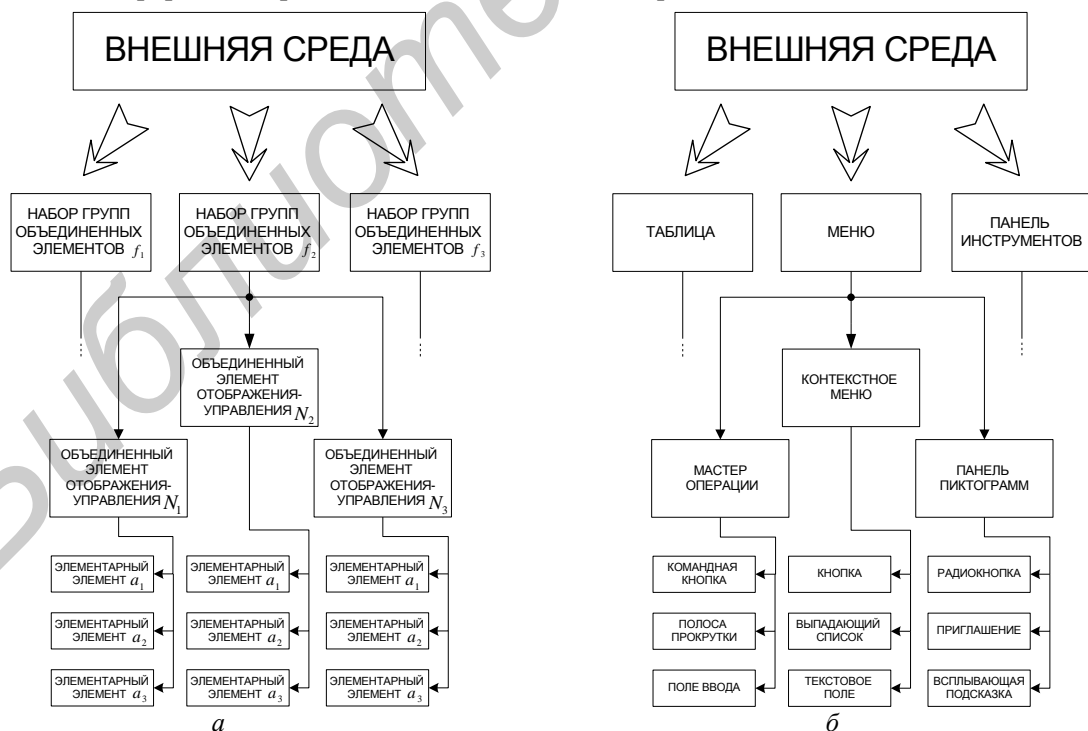


Рис. 2. Сравнение фрагмента предлагаемой информационной модели (а) и фрагмента информационной модели АСУ (б)

Под внешней средой будем понимать условия, которые в каждый момент времени вынуждают ЛПР организовывать соответствующие наборы групп ЭОУ. Следует отметить, что

предмет исследования – это структура иерархии системы отображения, поэтому и решение ЛПР также следует относить к внешним условиям или к внешней среде, считая, что иерархия осуществляется только лишь по заданному набору групп элементов. Таким образом, задача оптимизации структуры системы отображения решается отдельно от остальных задач управления. Такой подход позволит исследовать структурные явления «сами по себе» и будет способствовать построению общей модели, оптимальной с точки зрения восприятия управления, системой отображения.

Информацией о внешней среде, известной к началу определенного момента времени t , считаем наборы групп, определенные выражением (3).

Динамикой внешней среды назовем новые наборы групп ЭОУ, определенные выражением: $f^{(t+1)} = \{f_1^{(t+1)}, f_2^{(t+1)}, f_3^{(t+1)}, K, f_5^{(t+1)}, f_6^{(t+1)}, \dots, f_i^{(t+1)}\}$.

Под структурой будем понимать граф G^l организации заданного внешней средой набора групп. Под l будем понимать количество уровней управления.

Под управлением структурой $\Psi^l = [G^{(t,l)}, (f_1^{(t+1)}, f_2^{(t+1)}, f_3^{(t+1)}, K, f_5^{(t+1)}, f_6^{(t+1)}, \dots, f_i^{(t+1)})]$ будем понимать произвольное отображение, из всего множества графов организации над N_m , текущей структуры и известной информации об изменении внешней среды в структуру на следующем шаге.

Под стоимостью функционирования системы будем понимать величину $P(G^l)$ (трансформируем $P(G)$ из выражения (1)).

Стоимость реорганизации системы рассмотрим на примере реорганизации группы f (не учитывая временной интервал). Под стоимостью реорганизации группы f_1 в группу f_2 будем понимать величину $\rho(f_1, f_2) = \rho_1(N_m) + \rho_2(N_m)$, где $\rho_1(N_m)$ – заданная стоимость исключения объединенного элемента N_m из группы f_1 , а $\rho_2(N_m)$ – заданная стоимость включения объединенного элемента N_m в группу f_2 .

Введем понятие скорости изменения внешней среды. Под ней будем понимать величину S , равную количеству новых групп, появляющихся в наборе f^t в течение единицы времени, и количеству старых групп, которые удаляются из набора в течение единицы времени.

Результат управления Ψ^l информационной моделью АСУ тактического уровня в каждый момент времени t складывается из затрат на функционирование информационной модели, определяемых функционалом $P(G^l)$ и затрат на реструктуризацию $\rho(f_1, f_i)$ и зависит от динамики внешней среды и от управления, определяющего структуру информационной модели АСУ:

$$R(\Psi^l) = P(G^l) + \rho(f_1, f_i). \quad (4)$$

Определение условий перехода от статических моделей к динамическим

Говоря об информационных моделях АСУ в статике, достаточно сказать о поиске оптимальной структуры с наименьшим функционалом стоимости среди некоторого множества альтернатив, т.е. об оптимальной иерархии. Ключевым моментом при определении статической информационной модели является выбор функционала. Для различных АСУ накоплен огромный эмпирический материал, позволяющий определить некоторые агрегированные параметры структуры информационной модели. Данные разработки позволяют в самом общем виде сравнивать некоторые «типичные» структуры информационных моделей АСУ и выбирать из них наиболее подходящие для конкретной задачи.

При переходе от статических моделей АСУ к динамическим будем считать, что искомым функционал состоит из суммы стоимостей функционирования и реорганизации системы, т.е. затратам в течение некоторого интервала времени. Можно сказать, что если ситуация в системе остается неизменной, то достаточно минимизировать затраты на функционирование для получения наиболее эффективной системы и информационная модель

сводится к статической. Если же ситуация изменяется, то старая структура информационной модели АСУ может не отвечать новым требованиям и появится необходимость в ее реструктуризации, которая неизбежно повлечет за собой затраты на перестроение. И уже в этом случае информационная модель системы носит динамический характер и оптимальной будет модель, требующая меньших затрат на перестроение.

Под условием перехода от статической информационной модели АСУ к динамической, будем понимать условия, создаваемые внешней средой и характеризующиеся высокой скоростью изменения. Под высокой скоростью будем понимать такую скорость s , при которой временные затраты на реструктуризацию информационной модели, обусловленные появлением группы f_1 будут превышать временной интервал появления новой группы f_2 .

Нахождение оптимальной иерархической структуры информационной модели АСУ

Оценивая сложность нахождения оптимальной иерархической структуры информационной модели АСУ с помощью переборного алгоритма, достаточно сказать, что переборный алгоритм становится крайне трудоемким, с точки зрения вычислительного процесса для нахождения значения функционала, уже начиная с количества элементов N_m в упорядочиваемой группе равного одиннадцати.

Поэтому очевидный интерес представляют вопросы построения эвристических алгоритмов меньшей вычислительной сложности. Данные алгоритмы, возможно, будут давать сколь угодно большую погрешность, однако их все же можно использовать для конкретного функционала после предварительного тестирования.

Для поиска оптимальной иерархической структуры подсистемы отображения АСУ достаточно найти ее оптимальную организацию на множестве N_m . Иными словами, задача об оптимальной иерархии трансформируется в задачу об оптимальной организации.

Предположим, что к началу единицы времени t известно, какой набор групп объединенных элементов f^t нужно организовать, и какие наборы групп нужно было организовать на протяжении предыдущих единиц времени. То есть известна «история» изменений.

Оптимальное управление, как говорилось выше, минимизирует средние затраты на функционирование и на реструктуризацию системы на протяжении конечного числа единиц времени t . Оптимальное управление зависит от динамики внешней среды, в отношении которой, как правило, имеется некоторый прогноз или вероятностное распределение, позволяющее достаточно приблизительно, в среднем, наметить верное направление оптимизации. Оптимальным в среднем управлением назовем управление Ψ^l , минимизирующее средний результат управления (4), т.е.

$$\min R(f_1^{(t+1)}, f_2^{(t+1)}, f_3^{(t+1)}, K, f_5^{(t+1)}, f_6^{(t+1)}, \dots, f_i^{(t+1)}, \Psi^l),$$

$$\text{где } R(f_1^{(t+1)}, f_2^{(t+1)}, f_3^{(t+1)}, K, f_5^{(t+1)}, f_6^{(t+1)}, \dots, f_i^{(t+1)}, \Psi^l) = R(\Psi^l) = P(\Psi^l) + \rho(\Psi^l) \quad (5)$$

средний результат управления, где $P(\Psi^l)$ – затраты на функционирование информационной модели, а $\rho(\Psi^l)$ – затраты на реорганизацию информационной модели. Математическое ожидание динамики внешней среды предполагается известным на уровне его закона распределения.

Величина $P(\Psi^l)$ из выражения (5) очевидно не зависит от скорости изменения внешней среды, так как для всех временных интервалов, все возможные организуемые наборы объединенных групп ЭОУ будут иметь заранее определенную мощность и структуру пересечения друг с другом.

На рис. 3 изображены эмпирические зависимости $P(\Psi^l)$ и $\rho(\Psi^l)$ от l при различных s для структуры с набором из тридцати объединенных ЭОУ с одиннадцатью возможными уровнями иерархии управления, и мощностью одного набора ЭЭУ равную двенадцати, рассмотрен временной интервал равный тридцати периодам возможного появления новых

наборов групп (аналог периодов обновления информации на экране автоматизированного рабочего места ЛПР). Рассмотрены случайные наборы из десяти случайных групп. Вероятность появления любой из них одинакова [4]. При минимальной скорости изменения внешней среды, набор организуемых групп меняется один раз за t , при максимальной скорости тридцать раз.

Параметры, при которых проводился эксперимент, соответствуют функционированию подсистемы отображения АСУ в режиме боевой работы с высоким коэффициентом напряженности для ЛПР.

Составляющая $\rho(\Psi^l)$ из выражения (5) значительно трансформируется при увеличении скорости изменения внешней среды. Характерно, что максимальный рост затрат на реструктуризацию управления возникает при переходе от наиболее простой (веерной) организации, к организации, которая имеет два уровня управления. При дальнейшем увеличении l рост затрат на перестроение замедляется. Однако совершенно точно прослеживается тот факт, что с увеличением скорости изменения внешней среды, затраты на реорганизацию управления значительно увеличиваются, увеличение на порядок одного приводит к увеличению на порядок другого.

Однако характер изменения кривой $P(\Psi^l)$ свидетельствует об обратном, так как при переходе от веерной организации к двухуровневой приводит к максимальному уменьшению затрат на функционирование. В дальнейшем поведение кривой говорит об ослаблении данного эффекта. Таким образом, кривые $P(\Psi^l)$ и $\rho(\Psi^l)$ ведут себя в некотором смысле противоположным образом при увеличении либо уменьшении количества уровней управления.

Для поиска оптимального управления необходимо найти компромисс или точку пересечения кривых таким образом, чтобы либо минимизировать значение среднего результата управления (5) в конкретной ситуации, либо минимизировать нагрузку ЛПР, выбирая в качестве целевой функции минимум затрат на функционирование $\rho(\Psi^l)$.

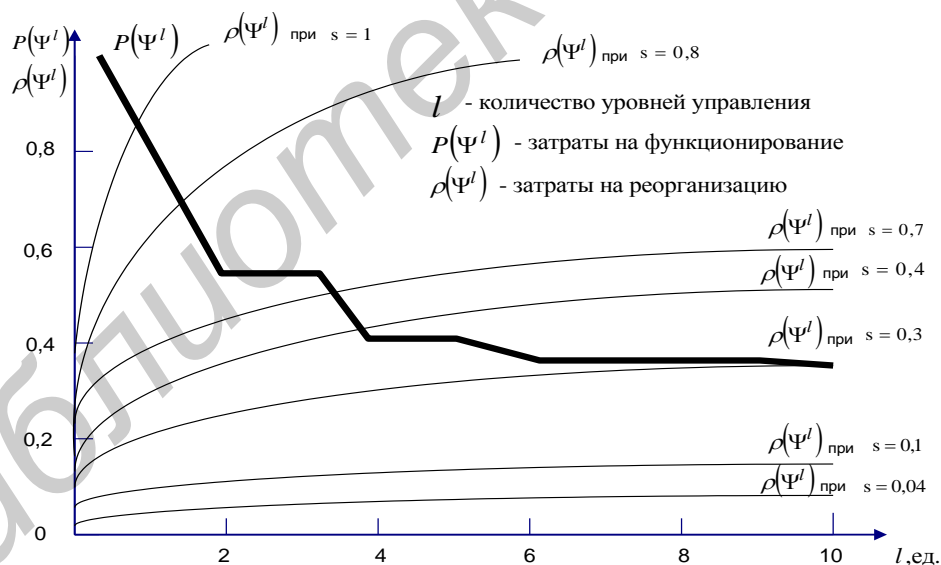


Рис. 3. Зависимости $P(\Psi^l)$ и $\rho(\Psi^l)$ от l при разных s

Из рис. 3 можно сделать следующий вывод: при жестких (интенсивных) изменениях внешней среды выгодно поддерживать простую (веерную) организацию системы, усложняя ее по мере смягчения внешних воздействий (увеличивая число уровней иерархии управления).

Качественно это утверждение соответствует тому, что в нестабильной внешней среде могут «выживать» лишь организационные системы с максимально простой структурой за счет быстрой приспособляемости, в стабильной же среде наоборот доминируют системы со сложной иерархической структурой за счет их высокой эффективности.

Заключение

Динамическая оптимизация информационной модели АСУ в первую очередь связана с проблемой выбора оптимального числа уровней иерархии управления элементами модели в зависимости от внешних условий. Данная проблема обсуждается в большинстве работ лишь на качественном уровне.

Использование предложенного способа нахождения количества иерархических уровней управления информационной моделью при ее построении характеризуется оптимальным информационным балансом, с точки зрения загруженности ЛПР, соответствием задачам и возможностям ЛПР по эффективной реализации им своих функций.

Использование способа повысит «живучесть» АСУ в ситуациях, связанных с высокой нагрузкой ЛПР, например, при больших плотностях налета и повысит качество управления за счет оптимизации иерархической структуры подсистемы отображения. Позволит построить информационную модель, отличающуюся оптимальной организацией и использованием интерфейса подсистемы отображения, что в условиях динамично изменяющейся боевой обстановки позволит оператору успешно ликвидировать возникающие конфликтные ситуации.

Результаты, полученные авторами, в дальнейшем, могут быть использованы как часть общей методики проектирования информационных моделей АСУ.

APPROACH TO REALIZATION OF INFORMATION MODEL OF THE SITUATION IN MANAGEMENT INFORMATION SYSTEM OF THE TACTICAL LINK

D.I. RABCHENOK, S.V. KRUGLIKOV

Abstract

The new way, allowing to define quantity of hierarchical levels in management, in relation to information model of an automated control system, and allowing to optimize load of the person making the decision is offered.

Список литературы

1. *Меньков А.В.* Теоретические основы автоматизированного управления: учебник для вузов. М., 2005.
2. *Бурков В.Н., Новиков Д.А.* // Материалы междунар. науч.-практ. конф. «Теория активных систем». Москва, 19–21 ноября 2001 г. С. 12–16.
3. *Месарович М., Мако Д., Такахара И.* Теория иерархических многоуровневых систем. М., 1973.
4. *Воронин А.А., Мишин С.П.* Оптимальные иерархические структуры. М., 2003.
5. *Кульба В.В., Микрин Б.В., Павлов Б.В.* Теоретические основы проектирования информационно-управляющих систем космических аппаратов. М., 2006.

УДК 621.37.46

ФОРМИРОВАНИЕ НЕПРОЗРАЧНОГО ДИЭЛЕКТРИЧЕСКОГО ДЕКОРАТИВНОГО ПОКРЫТИЯ ДЛЯ СЕНСОРНЫХ ДИСПЛЕЕВ

Е.А. ХОХЛОВ, А.С. МЫСЛИВЕЦ, А.Г. СМИРНОВ*

ООО «Изовак», ул. Богдановича 155, Минск, 220040, Беларусь

**Белорусский государственный университет информатики и радиоэлектроники
П. Бровка, 6, Минск, 220013, Беларусь**Поступила в редакцию 13 мая 2014*

Описаны процессы формирования непрозрачных диэлектрических декоративных покрытий для сенсорных дисплеев. Представлены их электрические и оптические характеристики, а также интерференционные структуры, позволяющие управлять цветом покрытия без изменения их диэлектрических свойств.

Ключевые слова: устройства отображения информации (дисплеи), сенсорная панель, интерференционные структуры.

Введение

Традиционно сенсорный дисплей представляет собой многослойную структуру, состоящую из защитного стекла, сенсора и самого дисплея. Для уменьшения габаритов и массы сенсорного дисплея в последнее время предпринимаются попытки создания сенсора непосредственно на защитном стекле, что вызывает технологические сложности из-за применения декоративных полимерных покрытий, которые скрывают разводку контрактных площадок сенсора [1, 2]. Нанесение прозрачного токопроводящего покрытия высокого качества, необходимого для создания сенсора, требует высокотемпературного нагрева, что невозможно из-за наличия полимерного покрытия [3]. В данной работе описана технология, позволяющая формировать такие покрытия методами вакуумного напыления, что решает данную проблему [4, 5].

Экспериментальная часть

Нанесение пленок нитрида производилось методом магнетронного и ионно-лучевого распыления мишени германия в среде смеси аргона и азота. Пленки оксидов наносились методом ионно-лучевого распыления. Магнетроны работали на постоянном токе при мощности 500 Вт. Для ионно-лучевого распыления использовался кольцевой источник типа «ускоритель анодного слоя». Работы проводились на лабораторном оборудовании Аспира-150. Мишени представляли собой диск диаметром 4 дюйма и толщиной 0,25 дюйма. В качестве подложек применялось стекло фирмы Corning Gorilla Glass. Напыление проводилось при комнатной температуре подложки. Оптические характеристики пленок были измерены на спектрофотометре SP-1024-USB. Показатель преломления и коэффициент поглощения определялись аналитическим методом по спектрам отражения и пропускания с помощью программы расчета оптических покрытий OptiLayer. Для формирования прозрачного окна использовалась фотолитография.

Схема поперечного сечения сенсорного дисплея представлена на рис. 1. Он состоит из самой подложки (в нашем случае это упрочненное стекло), декоративного непрозрачного покрытия (маски), прозрачного токопроводящего покрытия, металлических контактов.

Прозрачное токопроводящее покрытие является одной из обкладок емкостной сенсорной панели. Металлический контакт необходим для подключения к электрическому разъему ячеек сенсора. Декоративное непрозрачное покрытие скрывает от пользователя разводку, поэтому оно должно быть диэлектрическим, чтобы не влиять на работоспособность сенсора. Традиционно такое покрытие формируется печатью органических полимерных покрытий.

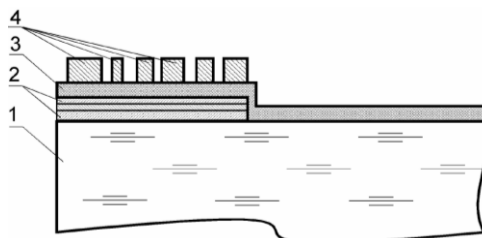


Рис. 1. Схема поперечного сечения сенсорного дисплея: 1 – стекло, 2 – декоративное непрозрачное покрытие, 3 – прозрачное токопроводящее покрытие, 4 – металлические контакты

После нанесения маски необходимо нанести прозрачное токопроводящее покрытие с максимально возможной степенью прозрачности. Известно, что для получения пленок высокого качества необходимо применять высокотемпературный процесс нанесения, однако органический полимер не может быть нагрет до высоких температур. В этом и заключается одна из основных технологических сложностей создания сенсорного дисплея непосредственно на поверхности защитного стекла.

Как правило, использование неорганических материалов позволяет решить обозначенную нами проблему нагрева, однако традиционно используемые декоративные непрозрачные покрытия, формируемые из металлов, например хрома, либо их оксидов, не обладают требуемыми диэлектрическими свойствами. Их поверхностное сопротивление меньше 2×10^{12} Ом/квадрат, что неприемлемо.

Применение полупроводниковых материалов, например, кремния, также не дает требуемых диэлектрических свойств, а его оксиды и нитриды, которые, с одной стороны, обладают приемлемыми диэлектрическими свойствами, с другой стороны прозрачны в видимом диапазоне спектра.

Нами предлагается использовать в качестве поглощающего слоя пленок на основе германия, позволяющих достигать поверхностного сопротивления на уровне 100 МОм/кв. и выше. Так, например, пленки нитрида германия обладают как требуемыми диэлектрическими свойствами, так и достаточным коэффициентом поглощения для возможности их применения в сенсорных дисплеях. Следует отметить, что метод нанесения существенно влияет на коэффициент их поглощения. Так, например, пленки, формируемые с помощью ионного источника, обладают большим коэффициентом поглощения, чем при магнетронном распылении, что объясняется более высокой плотностью пленки (см. таблицу).

Зависимость коэффициента пропускания пленки нитрида германия от толщины и метода нанесения

Метод нанесения	Толщина, нм	Пропускание на 550 нм, %	Поверхностное сопротивление, ГОм/кв
Магнетрон	100	1,5	>200
Магнетрон	200	<0,1	>200
Ионный источник	100	0,8	>200
Ионный источник	200	<0,1	>200

Экспериментально установлено, что однослойное покрытие нитрида германия имеет «серый» цвет с небольшим металлическим оттенком, поскольку коэффициент отражения пленки достаточно высокий, что обуславливается высоким коэффициентом преломления материала. Применение интерференционной структуры на основе пленок оксида ниобия и германия (рис. 2), позволяет уменьшить коэффициент отражения до единиц процентов, что обеспечивает «черный» цвет покрытия (рис. 3).

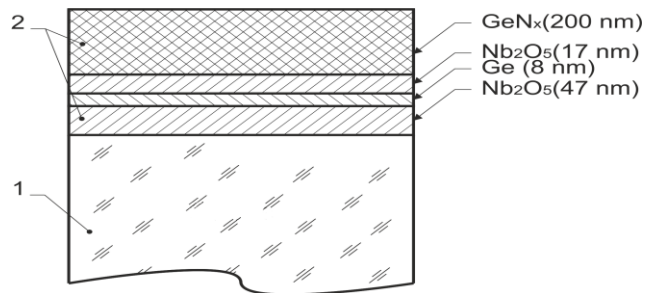


Рис. 2. Структура слоев «черного» цвета: 1 – стекло, 2 – покрытие.

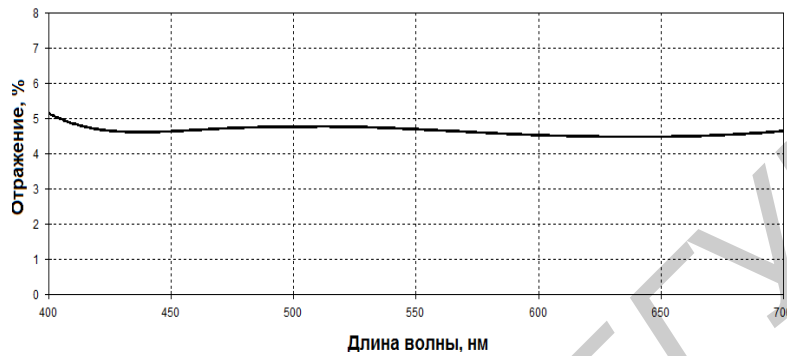


Рис. 3. Спектральная зависимость коэффициента отражения «черного» покрытия

Применение многослойных интерференционных покрытий на основе кремния и оксида кремния (рис. 4) позволяет получать покрытия других цветов.

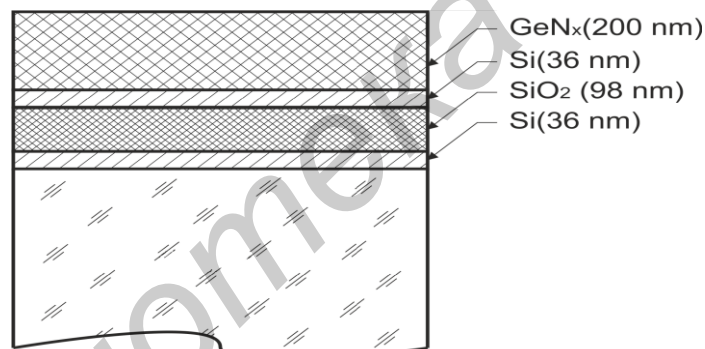


Рис. 4. Структура слоев «золотого» цвета

Увеличение коэффициента отражения в областях спектра, соответствующих «желтому» и «красному» диапазону, одновременно с уменьшением отражения «синей» области спектра, придает покрытию «золотистую» окраску (рис. 5).



Рис. 5. Спектральная зависимость коэффициента отражения «золотого» покрытия

Изменяя толщину слоев интерференционной структуры и материалы покрытий, можно добиваться широкой гаммы цветов декоративного покрытия.

Заключение

В работе были показаны зависимость оптических свойств пленок нитрида германия от толщины и метода нанесения. Использование неорганических декоративных покрытий вместо органических в современных сенсорных дисплеях дает возможность увеличения температуры последующих технологических операций, что существенно улучшает потребительские свойства дисплея, позволяет уменьшать общую толщину сенсора и, соответственно, устройств, в которых они используются. Возможность создания покрытий различного цвета также является существенным преимуществом перед органическими непрозрачными покрытиями. Применяемые методы нанесения давно используются в дисплейной промышленности, поэтому разработанная нами технология может быть успешно внедрена в существующие производства.

FORMATION OF A NON TRANSPARENT DIELECTRIC DECORATIVE COATING FOR SENSOR DISPLAYS

E.A. KHOKHLOV, A.S. MISLIVETS, A.G. SMIRNOV

Abstract

The processes of formation of opaque dielectric decorative coatings for touch screens is described. Their electrical and optical properties, as well as interference structures that control the color of the coating without changing of their dielectric properties are presented.

Список литературы

1. *Ching-Tai Chen, Frank C. Sulzbach*. Opaque low reflecting coating aperture on glass / Патент US6071616A.
2. *Robert Bellman, Ljerka Ukrainczyk*. Opaque chrome coating suitable for etching / Патент US7459095B2.
3. *Robert Bellman, Ljerka Ukrainczyk*. Opaque chrome coating having increased resistance to pinhole formation // Патент US7160628B2.
4. *Ширшов В.Я., Марышев С.П., Хохлов Е.А., Len Sen Lim*. Дисплейное сенсорное стекло, покрытие для светонепроницаемой диэлектрической маски для дисплейного сенсорного стекла и способ нанесения такого покрытия / Патент ЕА01200613.
5. *Ширшов В.Я., Марышев С.П., Хохлов Е.А.* Дисплейное сенсорное стекло, покрытие светонепроницаемой маски и способ нанесения такого покрытия / Патент ЕА201300333

КРАТКИЕ СООБЩЕНИЯ

УДК 621.383.92

ПОВЫШЕНИЕ ЖИВУЧЕСТИ ВОЛОКОННО-ОПТИЧЕСКОЙ СИСТЕМЫ СВЯЗИ В ЧРЕЗВЫЧАЙНЫХ СИТУАЦИЯХ

А.О. ЗЕНЕВИЧ

*Высший государственный колледж связи
Ф. Скорины, 8/2, Минск, 220114, Беларусь**Поступила в редакцию 3 ноября 2014*

Предложена волоконно-оптическая система связи для передачи данных в чрезвычайных ситуациях. Показано, что с целью повышения живучести таких систем необходимо использовать для передачи данных оптическое излучение с длиной волны 850 нм, счетчик фотонов и асинхронный способ передачи данных.

Ключевые слова: макроизгиб, волоконно-оптическая линия связи, чрезвычайная ситуация, счетчик фотонов.

Введение

В настоящее время в кабельных сетях связи используются волоконно-оптические линии (ВОЛС). Значительное увеличение числа волоконно-оптических линий связано с тем, что они позволяют осуществлять трансляцию данных с достаточно большой скоростью. Однако современные ВОЛС достаточно чувствительны к высоким механическим нагрузкам. Такие механические нагрузки могут появиться в результате чрезвычайной ситуации, возникающей в результате аварии, катастрофы, стихийного или экологического бедствия, а также при преднамеренном воздействии на ВОЛС. Последствия этих нагрузок могут привести к частичной или полной потере работоспособности системы связи. Потеря связи может привести к гибели людей или невозможности выполнения техникой или людьми поставленных перед ними задач. Поэтому требуется обеспечить надежность и живучесть ВОЛС на участках, длина которых не превышает расстояния между ретрансляционными участками. В связи с этим целью данной работы являлось повышение надежности и живучести, используемых волоконно-оптических линий к механическим воздействиям за счет использования счетчиков фотонов и асинхронного способа передачи данных.

Описание экспериментальной установки

В качестве объектов исследования использовались серийно выпускаемое оптическое одномодовое волокно КСО КСЦЗПБ-1×8Е-7,0, а также кремневые лавинные фотодиоды ФД-115Л и со структурой n^+p-p^+ .

Структурная схема экспериментальной установки представлена рис. 1. Принцип работы этой установки следующий. Источник оптического излучения И1 с длиной волны λ_1 используется для передачи данных. Рефлектометр тестирует оптическое волокно ОВ на наличие механических дефектов оптическими импульсами с длиной волны λ_2 . Оптический смеситель СМ смешивает оптические излучения с длинами волн λ_1 и λ_2 и направляет их в оптическое волокно ОВ. С выхода оптического волокна излучение подается на монохроматор М. Монохроматор используется для того, чтобы отделить оптическое излучение

с длиной волны λ_1 от излучения с длиной волны λ_2 . В результате чего только излучение с длиной волны λ_1 проходит через монохроматор и подается на его выход 1. С выхода 1 оптическое излучение поступает на фотоприемное устройство Φ . Устройство Φ используется для регистрации оптического излучения с длиной волны λ_1 .

При увеличении коэффициента затухания оптического излучения в волокне $ОВ$ до такого значения, при котором устройство Φ не может зарегистрировать излучение, от рефлектометра P поступает управляющий сигнал к источнику оптического излучения $И$. В результате чего источник $И$ переключается на передачу данных по асинхронному способу, описанному в работе [1]. В этом случае с управляющего выхода устройства Φ на монохроматор M поступает сигнал, который изменяет направление поступления оптического излучения с выхода 1 устройства Φ на выход 2. Оптическое излучение с выхода 2 подается на счетчик фотонов $СФ$. Счетчик фотонов используется в качестве приемника оптической информации, передаваемой от источника $И$ с использованием асинхронного способа передачи данных.

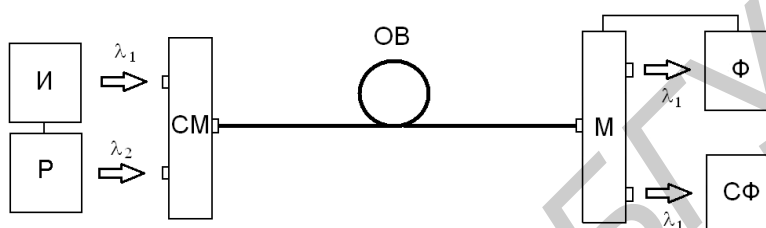


Рис. 1 Структурная схема экспериментальной установки: $И$ – источники оптического излучения (лазер), $СМ$ – оптический смеситель, $ОВ$ – оптическое волокно, $М$ – монохроматор, $СФ$ – счетчик фотонов, P – рефлектометр, Φ – фотоприемное устройство

При помощи цилиндров различных диаметров D формировались макроизгибы оптического волокна. Рефлектометр оценивал затухания волокна $ОВ$ с наличием макроизгиба и без него. Длины волн оптического излучения λ_1 и λ_2 изменялись в процессе проведения исследований и принимали значения 1625, 1550, 1450, 1350 и 850 нм. Мощность оптического излучения источника $И$ в процессе измерений регулировалась и выбиралась такой, при которой счетчик фотонов мог зарегистрировать поступающее на него оптическое излучение.

Общая длина волокна составляла $L = 398$ м, микроизгиб формировался на расстоянии $l = 198$ м от источника оптического излучения. В процессе проведения эксперимента расстояние l не изменялось.

Для поддержания постоянной рабочей температуры фотоприемника счетчика фотонов использовался термоэлектрический холодильник с максимальной температурой охлаждения 263 К. Исследования выполнены при температурах фотоприемника 300 и 263 К.

Экспериментальные результаты

Для оценки влияния макроизгиба оптического волокна на скорость передачи информации используется коэффициент $K = C/C_0$, где C_0 и C – скорости передачи информации по оптическому волокну в отсутствии и при наличии макроизгиба. На рис. 2 представлена зависимость коэффициента K от диаметра макроизгиба.

Отметим, что при проведении измерений мощность источника оптического излучения для каждой длины волны λ_1 подбиралась такой, чтобы скорость передачи данных для всех длин волн оставалась одинаковой. Из полученных зависимостей следует, что с уменьшением диаметра D уменьшается скорость передачи данных. Такая зависимость наиболее сильно проявляется для длины волны 1625 нм, а наиболее слабо – для длины волны 850 нм. Так для диаметра макроизгиба $D \leq 13$ мм скорость передачи информации для длины волны 1625 нм становилась равной нулю. Для длины волны 850 нм это наблюдалось для $D \leq 3$ мм. Поэтому в условиях механических воздействий на оптическое волокно, приводящих к появлению макроизгибов, целесообразно использовать длину волны 850 нм.

При диаметрах макроизгиба $D \leq 3$ мм затухание излучения в оптическом волокне для

длины волны 850 нм становится настолько большим, что для мощности источника оптического излучения 1 мВт фотоприемное устройство Φ нечувствительно к падающему на него излучению.

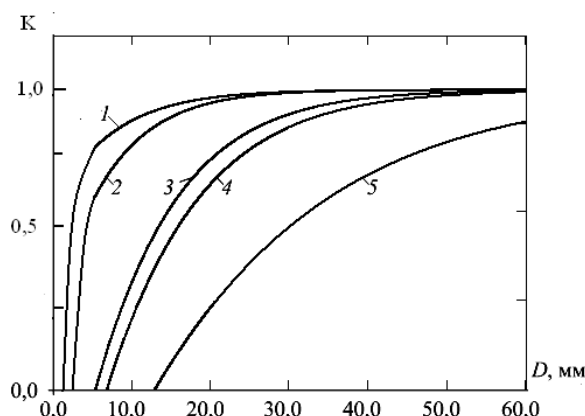


Рис. 2 Зависимость коэффициента K от диаметра макроизгиба: 1 – длина волны оптического излучения $\lambda_1=850$ нм; 2 – $\lambda_1=1310$ нм; 3 – $\lambda_1=1490$ нм; 4 – $\lambda_1=1550$ нм; 5 – $\lambda_1=1625$ нм

Если невозможно увеличить мощность источника оптического излучения, то необходимо повысить чувствительность фотоприемного устройства. Это можно обеспечить за счет использования счетчика фотонов СФ [2]. Применение режима счета фотонов позволит повысить пороговую чувствительность фотоприемного устройства в 100 и более раз.

При использовании счетчика фотонов в качестве фотоприемного устройства для приема информации передаваемой по оптическому волокну необходимо применять асинхронный способ передачи данных [1]. Асинхронный способ передачи данных не требует передачи оптических импульсов синхронизации работы источника и приемника информации. При появлении макроизгибов оптическое излучение синхроимпульсов подвержено сильному затуханию, также как и излучение, используемое для передачи данных. Это увеличивает вероятность ошибки регистрации синхроимпульсов и приводит к уменьшению скорости передачи информации по сравнению с асинхронным способом передачи данных.

В таблице представлены для сравнения характеристики счетчиков фотонов на основе лавинных фотодиодов, при регистрации оптического излучения с длиной волны 850 нм. Для определения скорости передачи информации при асинхронном способе передачи данных использовалась методика, описанная в работе [1]. Как следует из таблицы наибольшую скорость передачи данных позволяют обеспечить лавинные фотодиоды со структурой n^+p-p^+ при температуре 263 К.

Характеристики лавинных фотодиодов в режиме счета фотонов

Тип лавинного фотодиода	Лавинный фотодиод ФД – 115Л		Лавинный фотодиод со структурой n^+p-p^+	
	300	263	300	263
Рабочая температура, К	300	263	300	263
Квантовая эффективность регистрации, %	3	10	2	12
Скорость счета темновых импульсов, 10^4 с^{-1}	1,00	0,90	0,80	0,07
Длительность мертвого времени, мкс	1,0	1,0	1,0	1,0
Средняя длительность бита (символа), мкс	8	3	12	2
Скорость передачи информации, Кбит/с	50	130	30	190
Диаметр макроизгиба оптического волокна, мм	3			

Заключение

На основании проделанных исследований можно сделать вывод, что повысить надежность и живучесть волоконно-оптической линии к механическим воздействиям, формирующим макроизгибы, можно за счет использования для передачи данных оптического излучения с длиной волны 850 нм, в качестве приемника оптического излучения счетчика фотонов и асинхронного способа передачи данных.

При этом наибольшую скорость передачи данных для макроизгибов с диаметром $D \leq 3$ мм удастся обеспечить при использовании счетчика фотонов на основе лавинного фотоприемника со структурой $n^+p\text{-}\pi\text{-}p^+$ при температуре 263 К и длине волны оптического излучения 850 нм. Такая скорость передачи может быть использована для трансляции по волоконно-оптической линии связи речи и передачи данных.

Работа выполнена при поддержке Белорусского республиканского фонда фундаментальных исследований (договор № T13-018).

IMPROVEMENT OF SURVIVABILITY OF FIBER OPTIC COMMUNICATION SYSTEM USED FOR DATA TRANSMISSION IN EMERGENCIES

A.O. ZENEVICH

Abstract

Fiber optic communication system used for data transmission in emergencies has been suggested. It has been shown in this paper that for increasing survivability of such systems, optical radiation with a wavelength of 850 nm, photon counter and asynchronous data transfer mode are to be used.

Список литературы

1. Гулаков И.Р., Зеневич А.О., Тимофеев А.М. // Приборы и методы измерений. 2013. № 2 (7). С. 80–87.
2. Гулаков И.Р., Холондырев С.В. Метод счета фотонов в оптико-физических измерениях. Минск, 1989.

УДК 621.372.85

ОЦЕНКА ЭФФЕКТИВНОСТИ НАПРАВЛЕННОГО ИЗЛУЧЕНИЯ ВСТРЕЧНО-ШТЫРЕВОГО ПРЕОБРАЗОВАТЕЛЯ НА НЕОДНОРОДНО ПОЛЯРИЗОВАННОМ ЗВУКОПРОВОДЕ

В.Г. БАСОВ

Белорусский государственный университет информатики и радиоэлектроники
П. Бровка, 6, Минск, 220013, Беларусь

Поступила в редакцию 11 декабря 2012

С применением теории связанных волн получена оценка эффективности направленного излучения встречно-штыревого преобразователя, выполненного с неоднородной поляризацией пьезозвукопровода и возбуждающего акустические источники с фазовым сдвигом 90° .

Ключевые слова: возбуждение, поверхностные акустические волны, направленность излучения, пьезозвукопровод.

Введение

Широко применяемый для построения акустоэлектронных устройств встречно-штыревой преобразователь (ВШП), расположенный на звукопроводе с однородной поляризацией, характеризуется двунаправленным возбуждением поверхностной акустической волны (ПАВ). ВШП, кроме потерь на преобразование электрической энергии в энергию акустических волн, вносит дополнительные потери, связанные с двунаправленным возбуждением. Для устранения указанного свойства применяются различные способы получения однонаправленного возбуждения. Одним из таких способов является последовательное расположение двух ВШП, соединенных между собой фазосдвигающим звеном, который обеспечивает относительный сдвиг фазы на угол $\varphi = 90^\circ$. Продолжением этого метода является «вложение» одного ВШП в структуру другого, а также применение трехфазного преобразователя [1]. Однако все эти способы получения однонаправленного возбуждения приводят к увеличению протяженности структуры ВШП, что сопровождается уменьшением эффективности и полосы частот возбуждения ПАВ.

В случае неоднородной поляризации пьезокерамического звукопровода при возбуждении ВШП последним формируются периодические акустические источники, сдвинутые по фазе на 90° [2]. Наличие таких акустических источников позволяет создавать структуры ВШП с определенной направленностью возбуждения ПАВ. В данной работе проведена оценка эффективности направленного возбуждения ВШП, расположенного на пьезозвукопроводе с неоднородной поляризацией.

Теоретический анализ

Для оценки эффективности направленного излучения ВШП с неоднородной поляризацией пьезозвукопровода рассмотрим некоторую пару разнополярных электродов 1 и 2 ВШП, которые формируют по паре акустических источников A_1 и B_1 , A_2 и B_2 с фазовым сдвигом 90 градусов, возбуждающих парциальные акустические волны u_1 и u_2 , распространяющиеся в положительном (индекс «+») и отрицательном (индекс «-») направлениях (рис. 1).

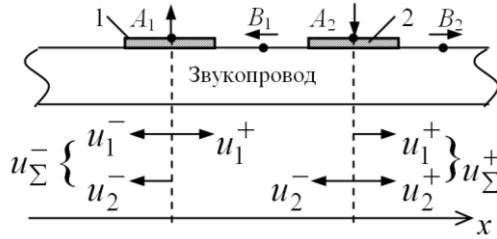


Рис. 1. Возбуждение парциальных волн ВШП в пьезозвукопроводе с неоднородной поляризацией

Тогда суммарные составляющие этих волн u_{Σ}^{-} и u_{Σ}^{+} можно определить, применив для решения такой задачи теорию связанных волн [3]. Для волн, распространяющихся соответственно в положительном и отрицательном направлениях, получим

$$\begin{cases} \frac{\partial u_1^+}{\partial x} - jk_1 u_1^+ = j\alpha u_2^+ \\ \frac{\partial u_2^+}{\partial x} - jk_2 u_2^+ = j\alpha u_1^+ \end{cases}, \quad (1)$$

$$\begin{cases} \frac{\partial u_1^-}{\partial x} - jk_1 u_1^- = j\alpha u_2^- \\ \frac{\partial u_2^-}{\partial x} - jk_2 u_2^- = j\alpha u_1^- \end{cases}, \quad (2)$$

где k_1 и k_2 – постоянные распространения указанных волн, создаваемых акустическими источниками A_1 и B_1 ; α – коэффициент связи между этими волнами.

Проводя решение системы уравнений (1) и (2) в предположении, что

$$u_{1,2}^{\pm} = a_{1,2} \exp(\gamma_{\pm} x), \quad (3)$$

где γ_{\pm} – постоянные распространения парциальных акустических волн с амплитудами $a_{1,2}$ – возбуждаемые акустическими источниками A_1 и B_1 , и распространяющимися в положительном и отрицательном направлениях вдоль оси x , получим суммарные значения u_{Σ}^{+} и u_{Σ}^{-} в следующем виде

$$u_{\Sigma}^{+} = a_1 \exp\left(-j\frac{\pi}{2}\right) \exp(\gamma_+ x) + a_2 \exp(-j\varphi_{\ell}) \exp(\gamma_+ x), \quad (4)$$

$$u_{\Sigma}^{-} = a_1 \exp\left(-j\frac{\pi}{2}\right) \exp(-j\varphi_{\ell}) \exp(\gamma_- x) + a_2 \exp(\gamma_- x), \quad (5)$$

где φ_{ℓ} – фазовый сдвиг между акустическими источниками A_1 и B_1 .

В анализируемом случае выполняется условие $\gamma_- = \gamma_+$ и коэффициент направленного излучения N_n [4] может быть определен как

$$N_n = \frac{u_{\Sigma}^{+} - u_{\Sigma}^{-}}{u_{\Sigma}^{+}} = 1 - \frac{1 + P \exp\left[-j\left(\frac{\pi}{2} + \varphi_{\ell}\right)\right]}{1 + P \exp\left[-j\left(\frac{\pi}{2} - \varphi_{\ell}\right)\right]}, \quad (6)$$

где $P = \frac{a_1}{a_2}$.

Амплитудно-частотные характеристики (АЧХ) анализируемого ВШП для отрицательного (индекс «-») и положительного (индекс «+») направлений распространения парциальных акустических волн определяются соотношениями

$$H^-(f) = \sum_{n=1}^M u_{zn}^- \exp(-j\gamma_{-x}), \quad (7)$$

$$H^+(f) = \sum_{n=1}^M u_{zn}^+ \exp(+j\gamma_{+x}), \quad (8)$$

где M – количество электродов ВШП.

Результаты расчетов

Используя выражение (6), были рассчитаны графики коэффициента направленности N_n в зависимости от отношения амплитуд P парциальных акустических волн a_1 и a_2 при различных фазовых сдвигах между ними φ_ℓ (рис. 2). Из графиков рис. 2 следует, что наибольший коэффициент направленного излучения ВШП будет в случае, когда $P = 1$, т.е. амплитуды парциальных акустических волн равны $a_1 = a_2$, и имеют фазовый сдвиг $\varphi_\ell = \pi/2$.

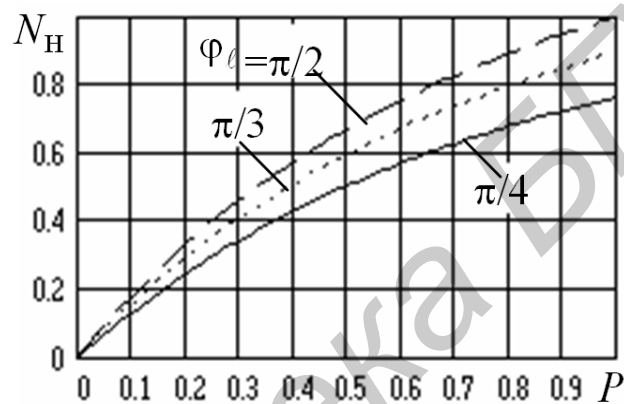


Рис. 2. Зависимость коэффициента направленности N_n от фазового угла φ между парциальными волнами

Рассчитанные по (7, 8) АЧХ для ВШП, расположенного на пьезозвукопроводе с неоднородной поляризацией, показаны на рис. 3, где сплошной линией изображена АЧХ в случае распространения ПАВ в положительном, а пунктирной в отрицательном направлениях.

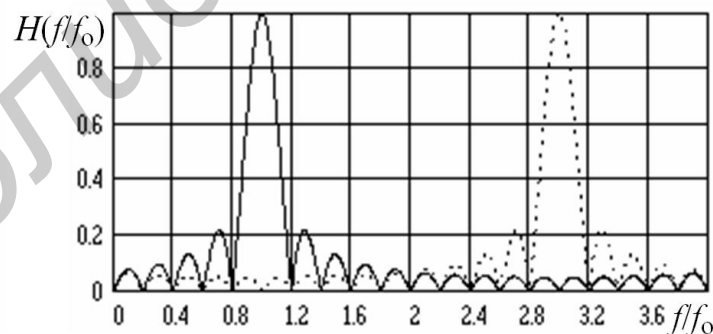


Рис. 3. Частотные зависимости коэффициента передачи $H(f/f_0)$ от направления распространения ПАВ

Анализ графиков рис. 3 показывает, что на основной частоте возбуждения ПАВ имеется АЧХ только для положительного направления, а для отрицательного направления на этой частоте частотная характеристика не формируется, что указывает на то, что излучение отсутствует. Из рассмотренных графиков следует, что применение пьезозвукопроводов на основе пьезокерамик с неоднородной поляризацией последних позволяет формировать многофазные структуры ВШП, которые обладают однонаправленным излучением ПАВ. Применение таких структур для построения устройств на ПАВ позволяет снизить вносимые ими потери до 6 дБ без увеличения протяженности структур ВШП.

Заключение

Применение на практике неоднородной поляризации пьезокерамических звук-проводов позволяет широко использовать их для построения фильтрующих устройств с меньшими вносимыми потерями по сравнению с обычно распространенными ВШП с однородной поляризацией пьезозвукопровода.

EVALUATING THE EFFECTIVENESS OF DIRECTIONAL RADIATION INTERDIGITAL TRANSDUCER ON INHOMOGENEOUSLY POLARIZED ACOUSTIC LINE

V.G. BASOV

Abstract

With application of connected waves theory the evaluation of directed radiation efficiency of interdigital executed by inhomogeneous polarization piezosoundlead transducer exciting acoustic sources with a phase shift 90 degrees has obtained.

Список литературы

1. *Речицкий В.И.* Акустоэлектронные радиокомпоненты. Схемы, топология, конструкции. М., 1987.
2. *Басов В.Г., Синица В.Н.* // Радиотехника и электроника: Сборник научных трудов. 2000. Вып. 25. С. 125–128.
3. *Басов В.Г., Чиркин Н.М.* // Известия АН БССР, сер. Физико-техническая. 1972. № 3. С. 114–120.
4. *Басов В.Г., Синица В.Н., Лайков Г.Д.* // Радиотехника и электроника. 1988. Вып. 17. С. 55–58.

ДЛЯ ЗАМЕТОК

Lined area for notes, featuring 20 horizontal lines. A large, faint watermark is oriented diagonally from the bottom-left to the top-right, reading "Библиотека БГУИР".

ПРАВИЛА ОФОРМЛЕНИЯ СТАТЕЙ, НАПРАВЛЯЕМЫХ В РЕДАКЦИЮ ЖУРНАЛА «ДОКЛАДЫ БГУИР»

1. К публикации принимаются статьи, отражающие результаты **оригинальных научно-технических исследований и разработок, не опубликованные и не предназначенные для публикации в других изданиях**. Тематика представляемой статьи должна соответствовать рубрикам журнала (электроника, радиофизика, радиотехника, информатика).

2. Статьи должны быть написаны в сжатой и ясной форме и содержать соответствующий индекс УДК; название на русском и английском языках; инициалы и фамилии авторов на русском и английском языках; полное название учреждений, в которых выполнялось исследование (с указанием адреса); ключевые слова (до 10 слов) на русском языке; аннотацию на русском и английском языках (3–5 предложений). Рекомендуется стандартизировать текст статьи, используя подзаголовки **введение, теоретический анализ, методика, экспериментальная часть, результаты и их обсуждение, заключение, список литературы**.

3. Статья представляется в двух экземплярах, распечатанных на листах формата А4, и в электронном варианте в виде файла формата MS Word (расширение – *.doc). Также необходимо представить *акт экспертизы* о возможности опубликования статьи в открытой печати.

Количество страниц и иллюстраций:

- для оригинальной статьи – не более 7 страниц и 8 иллюстраций;
- для краткого сообщения – не более 3 страниц и 2 иллюстраций.

4. При наборе текста используется гарнитура Times New Roman и Symbol, кегль – 11 pt. Устанавливаемый размер бумаги – А4 (210×297 мм). Поля слева, справа, сверху и снизу – по 25 мм. Межстрочный интервал – 1,0. Абзацный отступ – 1,25 см. *Запрещается* при наборе текста использовать гиперссылки, маркированные списки и ручные переносы слов.

5. **Формульные выражения** выполняются только в редакторе формул MathType, стиль – «Математика» («Math»), гарнитура – Times New Roman и Symbol, кегль – 11 pt. Как в тексте статьи, так и в формульных выражениях, математические функции, цифры, буквы греческого и русского алфавитов набираются прямым начертанием, латинские буквы – курсивом. Нумеровать рекомендуется те формульные выражения, на которые имеются ссылки в последующем тексте, заключая при этом номер в круглые скобки.

6. Рисунки выполняются в соответствии со следующими требованиями:

- размер рисунка – не более 15×20 см, разрешение – не менее 300 dpi;
- рисунок должен быть оформлен как графический объект;
- подрисовочные подписи обязательно должны включать в себя номер и название рисунка;
- словесных обозначений на рисунках рекомендуется избегать, используя при этом цифровые либо буквенные обозначения, пояснения к которым следует давать в тексте статьи либо подрисовочных подписях;
- цифровые и буквенные обозначения на рисунках по начертанию и размеру должны соответствовать обозначениям в тексте статьи либо подрисовочных подписях.

7. Наименования, обозначения, определения и правила применения единиц физических величин, употребляемых в статье, должны соответствовать ГОСТ 8.417–2002.

8. Ссылки на литературные источники нумеруются в порядке их цитирования в тексте. Номера ссылок заключаются в квадратные скобки.

9. Набор статьи рекомендуется производить в шаблоне, доступном для скачивания на сайте БГУИР (www.bsuir.by → «Журнал "Доклады БГУИР"» → раздел «Авторам»).

10. Адрес для пересылки статей в электронном варианте и переписки с редакцией: doklady@bsuir.by.

***Ответственность за оригинальность и достоверность
материалов, поданных в печать, несут авторы.***