

ISSN 1816-0301

ИНФОРМАТИКА

3 (43)

ИЮЛЬ-СЕНТЯБРЬ
2014

Редакционная коллегия:

Главный редактор

А.В. Тузиков

Заместитель главного редактора

М.Я. Ковалев

Члены редколлегии

С.В. Абламейко, В.В. Анищенко, П.Н. Бибило, М.Н. Бобов,
А.Н. Дудин, А.Д. Закревский, С.Я. Килин, В.В. Краснопрошин,
С.П. Кундас, Н.А. Лиходед, П.П. Матус, С.В. Медведев, А.А. Петровский,
Ю.Н. Сотсков, Ю.С. Харин, А.Ф. Чернявский, В.Н. Ярмолик
Н.А. Рудая (*заведующая редакцией*)

Адрес редакции:

220012, Минск,
ул. Сурганова, 6, к. 305
тел. (017) 284-26-22
e-mail: rio@newman.bas-net.by
<http://uiip.bas-net.by>

ИНФОРМАТИКА

ЕЖЕКВАРТАЛЬНЫЙ НАУЧНЫЙ ЖУРНАЛ

Издается с января 2004 г.

№ 3(43) • июль-сентябрь 2014

СОДЕРЖАНИЕ

ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ И РЕЧИ

- Снежко Э.В., Карлье П., Аззабу Н., Ковалев В.А., Дмитрук А.А., Шукелович А.В.** Количественная оценка нервно-мышечных нарушений на основе текстурного анализа изображений магнитно-резонансной томографии..... 5
- Залесский Б.А.** Метод нечеткой кластеризации k -средних со сглаживающей штрафной функцией..... 14
- Ткачяня А.В.** Адаптация скрытых марковских моделей к распознаванию эмоционально окрашенной речи..... 21

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

- Ерофеев В.Т., Бондаренко В.Ф.** Экранирование электромагнитных полей экранами из матричных композитов, содержащих биизотропные частицы 28
- Левин Г.М., Розин Б.М., Долгий А.Б.** Линейная аппроксимация задачи оптимизации интенсивностей последовательно-параллельного выполнения пересекающихся множеств операций..... 44
- Зяц Г.М., Комаров Ф.Ф., Комаров А.Ф.** Численное исследование влияния радиационных излучений на параметры МДП-приборов 52
- Побегайло А.П.** Моделирование сплайн-поверхностей Бэзье на четырехугольных сетках кривых..... 62

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Казанцева О.Г., Барановский Е.А., Ландарский Е.А. Шейдерная метамодель как компонент архитектуры интерактивного приложения	72
Сычев В.А. Реализация дискретных нелинейных динамических систем с хаотическими режимами на арифметике с фиксированной запятой	80
Бородуля А.В., Соловьев А.Н., Напрасников В.В., Мирзаванд М.А. Выбор рациональных параметров конструкции каркаса морской платформы	89
Несенчук А.А. Исследование динамических систем третьего порядка с неопределенностью на основе корневых портретов	96

ЗАЩИТА ИНФОРМАЦИИ

Поляков А.С., Самсонов В.Е. Анализ возможностей алгоритмов международного стандарта «Облегченная криптография» – ISO/IEC 29192-2:2012	107
--	-----

Редактор Г.Б. Гончаренко
Корректор А.А. Михайлова
Компьютерная верстка Д.С. Гавинович

Сдано в набор 04.08.2014. Подписано в печать 08.09.2014.
Формат 60×84 1/8. Бумага офсетная. Гарнитура Таймс.
Усл. печ. л. 13,0. Уч.-изд. л. 12,8. Тираж 100 экз. Заказ 24.

Государственное научное учреждение «Объединенный институт проблем информатики Национальной академии наук Беларуси».
Свидетельство о государственной регистрации издателя, изготовителя, распространителя печатных изданий № 1/274 от 04.04.2014.
ЛП № 02330/444 от 18.12.13.
Ул. Сурганова, 6, 220012, Минск.

INFORMATICS

PUBLISHED QUATERLY

Issued since 2004

№ 3(43) • July-September 2014

CONTENTS

SIGNAL, IMAGE AND SPEECH PROCESSING

- Snezhko E.V., Carlier P., Kovalev V.A., Azzabou N., Dmitruk A.A., Shukelovich A.V.** Application of texture analysis techniques to NMR images for quantitative assessment of muscle disorders 5
- Zalesky B.A.** Method fuzzy clustering k -means with smoothing penalty function..... 14
- Tkachenia A.V.** Adaptive learning of hidden markov models for emotional speech 21

MATHEMATICAL MODELING

- Erofeenko V.T., Bondarenko V.F.** Screening electromagnetic fields by means of screens from matrix composites, containing biisotropic fractions 28
- Levin G.M., Rozin B.M., Dolgui A.B.** Linear approximation for intensities optimization problem of sequential-parallel execution of intersecting operation sets 44
- Zayats G.M., Komarov F.F., Komarov A.F.** Numerical simulation of the influence of radiation on the MOS devices parameters..... 52
- Pobegailo A.P.** Design of Bezier spline surfaces over bivariate networks of curves 62

APPLIED INFORMATION TECHNOLOGIES

Kazantsava V., Baranouski Y., Landarski Y. Shader metamodel as a component of an interactive application architecture.....	72
Sychou U.A. Realization of discrete nonlinear dynamical systems with chaotic regimes based on fixed-point arithmetic.....	80
Borodulia A.V., Soloveev A.N., Naprasnikov V.V., Mirzavand M.A. Choice of rational parameters of the frame construction of offshore platform	89
Nesenchuk A.A. Research of uncertain dynamic systems of the third order based on the root locus portraits	96

INFORMATION SECURITY

Poljakov A.S., Samsonov V.E. Analysis of the characteristics of international standard algorithms «Lightweight cryptography» – ISO/IEC 29192-3:2012.....	107
---	-----

ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ И РЕЧИ

УДК 004.9

Э.В. Снежко¹, П. Карлье², Н. Аззабу², В.А. Ковалев¹, А.А. Дмитрук¹,
А.В. Шукелович^{1,2}

**КОЛИЧЕСТВЕННАЯ ОЦЕНКА НЕРВНО-МЫШЕЧНЫХ НАРУШЕНИЙ
НА ОСНОВЕ ТЕКСТУРНОГО АНАЛИЗА ИЗОБРАЖЕНИЙ
МАГНИТНО-РЕЗОНАНСНОЙ ТОМОГРАФИИ**

Предлагаются результаты исследований по использованию базовой концепции методики анализа медицинских изображений посредством адаптации универсальных дескрипторов изображений и их использования для количественной интерпретации областей мышц на изображениях магнитно-резонансной томографии.

Введение

В настоящей работе рассматриваются характеристические признаки текстур с целью количественной оценки уровня дегенерации мышц по изображениям магнитно-резонансной томографии (МРТ). Эксперименты по оценке эффективности характеристических признаков проводились на изображениях мышц животных и человека.

Всего были протестированы следующие группы признаков: интегральные признаки текстур, матрицы совместной встречаемости Харалика, групповые матрицы, обобщенные матрицы совместной встречаемости различных типов с использованием значений яркостей элементов изображений, обобщенные матрицы совместной встречаемости различных типов с использованием значений производных первого и второго порядков, признаки на основе анизотропии текстур изображений и анизотропии матриц совместной встречаемости текстур.

По результатам полученных оценок выбраны объединенные количественные признаки, экспериментально установлена их эффективность по параметрам специфичности, чувствительности и точности.

1. Исходные данные

В качестве исходных данных, по которым вычислялись количественные признаки, использовались две группы выборок МРТ-изображений. Изображения были предоставлены лабораторией ядерного магнитного резонанса Института миологии (Париж, Франция).

Первая выборка содержала изображения здоровых (контрольная группа) и больных мышечной дистрофией собак породы золотистый ретривер [1, 2]. Течение заболевания мышц (дистрофии Дюшенна) у данной породы собак схоже с течением заболевания у людей [3], поэтому они часто используются в качестве тестовых объектов при изучении упомянутой патологии мышц. Всего выборка содержала МРТ-изображения (типа T2 и T2FS) 39 животных. Внутри каждой группы были выделены пять возрастных групп: 2, 4, 6, 8 и 12 месяцев. Разделение по возрастным группам было сделано по той причине, что степень поражения мышц увеличивается с прогрессированием болезни. Ожидалось, что на ранней стадии заболевания текстуры мышц будут отличаться незначительно либо не отличаться вообще, в то время как значительные различия проявятся у более взрослых собак. На всех изображениях специалистами-радиологами были сегментированы восемь групп мышц: EDL-G, EDL-D, GASLAT-G, GASLAT-D, GASMED-G, GASMED-D, TC-G, TC-D (рис. 1). Все изображения были преобразованы из формата данных DICOM в формат Analyze.

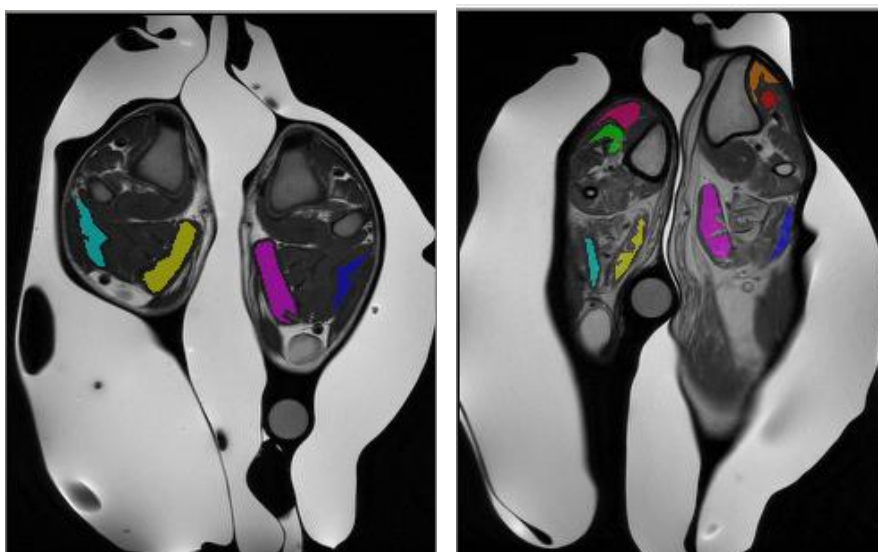


Рис. 1. Примеры МРТ-изображений мышц животных, выделенных разными цветами (серые области вокруг лап относятся к приспособлению для фиксации животных)

Вторая выборка состояла из МРТ-изображений мышц типа T2FS 20 здоровых людей (добровольцев). Выборка изображений была разбита на две подвыборки AG_1 и AG_2 по возрасту: на условно молодых и пожилых людей соответственно (рис. 2). Таким образом в ходе работ с данной выборкой была осуществлена оценка эффективности признаков текстур для разделения изображений на два соответствующих класса. На всех изображениях специалистами-радиологами были сегментированы 16 групп мышц: AL_G, AL_D, AM_G, AM_D, BF_G, BF_D, RF_G, RF_D, SM_G, SM_D, ST_G, ST_D, VI_G, VI_D, VL_G, VL_D. Все изображения были преобразованы из формата данных DICOM в формат Analyze (рис. 3).

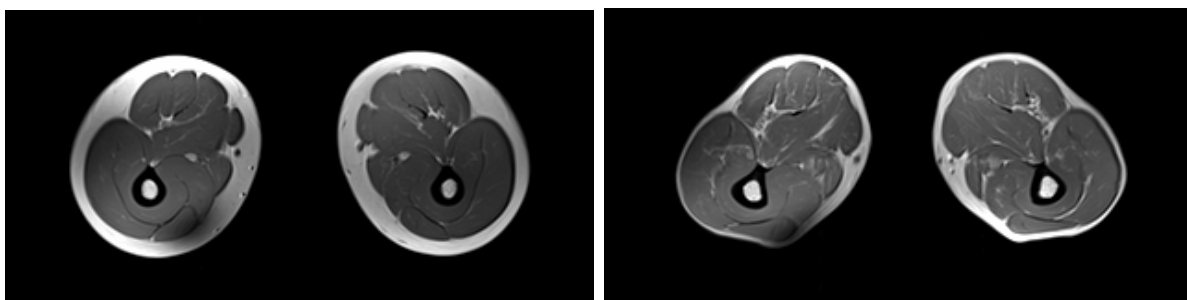


Рис. 2. Примеры срезов изображений мышц людей из подвыборок AG_1 и AG_2 соответственно



Рис. 3. Примеры срезов изображений сегментированных мышц людей из подвыборок AG_1 и AG_2 соответственно

2. Интегральные количественные признаки текстур мышц на МРТ-изображениях

С целью предварительного анализа возможности количественной оценки уровня деградации мышц было использовано специализированное программное обеспечение (ПО) MaZda, разработанное в рамках европейского проекта COST B11 «Количественный анализ текстур изображений магнитно-резонансной томографии» (1998–2002) [4, 5]. Данное ПО позволяет вычислять более 300 различных текстурных признаков как нулевого (признаков гистограмм), так и более высокого порядка (матриц совместной встречаемости и групповых матриц).

По результатам классификации с использованием модуля b11 среди интегральных гистограммных признаков текстур определены те, которые лучше всего разделяют классы текстур здоровых и больных дистрофией собак, а также значения ошибочной классификации данных признаков. Построена таблица ошибок классификации мышц здоровых и больных дистрофией собак по интегральным признакам, вычисленным при помощи специализированного ПО MaZda (табл. 1).

Таблица 1

Фрагмент таблицы ошибок классификации мышц здоровых и больных дистрофией собак по интегральным признакам для мышцы GASLAT-D

Месяц	Признак	Количество ошибок классификации, %	Признак	Количество ошибок классификации, %
2-й	Variance	17,65	Variance	5,88
4-й	S(1,0)SumVarnc	20,00	S(1,0)SumVarnc	15,00
6-й	Variance	12,50	S(1,-1)InvDfMom	12,50
8-й	S(1,0)Correlat	6,25	S(1,0)Correlat	6,25
12-й	S(1,0)Entropy	8,70	S(1,-1)Entropy	8,70

Среди интегральных гистограммных признаков текстур мышц собак чаще всего наиболее информативным признаком являлось значение среднеквадратического отклонения вокселей из областей интереса (16 %).

В то же время уровень ошибочной классификации с использованием среднеквадратического отклонения, а также тот факт, что во многих других случаях в качестве наиболее информативных выбирались и другие признаки, не позволяют выделить некий интегральный признак, с использованием которого можно было бы не только достоверно разделять классы изображений, относящихся к здоровым или больным дистрофией Дюшенна собакам, но и определять степень поражения мышц. Поэтому были продолжены работы по разработке количественных признаков текстур мышц на МРТ-изображениях.

3. Признаки текстур на основе матриц совместной встречаемости и групповых матриц с использованием значений яркостей элементов изображений

Кроме интегральных (например, моменты различных порядков) и геометрических признаков ПО MaZda позволяет вычислять матрицы совместной встречаемости Харалика, а также групповые матрицы (run-length matrices).

Для каждого из слоев МРТ-изображений мышц (EDL-G, EDL-D, GASLAT-G, GASLAT-D, GASMED-G, GASMED-D, TC-G, TC-D) с использованием ПО MaZda были вычислены 313 количественных признаков. Так, для каждой мышцы, признака и возрастной группы были получены два вектора чисел. Первый вектор соответствовал изображениям собак с мышечными нарушениями (GRMD), а второй – контрольной группе изображений (Healthy).

Пример 1

Возрастная группа «4-й месяц», мышца GASLAT-D:

– группа GRMD (собаки с мышечными нарушениями) (vector_g):

2470,6626 3767,0305 2357,6574 2409,0150 3372,9133 1445,1748
 1815,3884 2639,4784 1779,9983 2416,6319 2649,6683 1244,0339
 1433,8159 2634,6316;

– контрольная группа (vector_h):

1479,8324 817,5634 2773,8365 1143,4609.

С использованием скользящего порогового значения от минимального ($\min(\text{vector}_g, \text{vector}_h)$) до максимального ($\max(\text{vector}_g, \text{vector}_h)$) для каждого из пороговых значений были вычислены чувствительность (True positive rates, TPR) и показатель ложных срабатываний (False positive rates, FPR). На основе вычисленных данных были построены кривые оперативной характеристики надежности (Receiver operating characteristic curves, ROC-curves (рис. 4)), а также определены значения площади под кривой (Area under curve, AUC).

Пример 2

Возрастная группа, мышца и признак те же, что и в примере 1; $\min(\text{vector}_g, \text{vector}_h) = 817,5634$; $\max(\text{vector}_g, \text{vector}_h) = 3767,0305$; значение площади под кривой $AUC = 0,74362$.

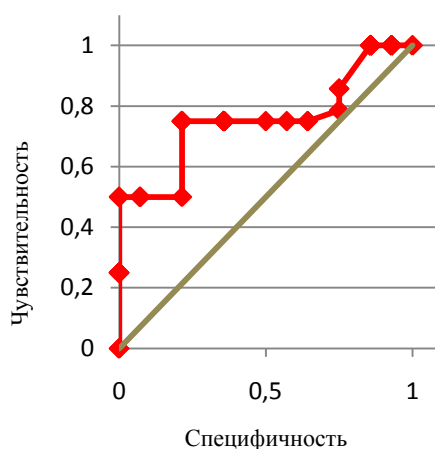


Рис. 4. Пример ROC-кривой надежности для возрастной группы «4-й месяц», мышцы GASLAT-D и одного из признаков

Для всех 313 количественных признаков вычислены средние значения по всем мышцам площадей под ROC-кривыми внутри каждой возрастной группы собак. Для фиксированной возрастной группы количественные признаки были отсортированы в убывающем порядке в соответствии с их средней эффективностью для разделения изображений, соответствующих собакам из групп Healthy и GRMD.

Пример 3

Для вычисленных значений площади под кривой (AUC) возрастной группы «4-й месяц» всех рассматриваемых мышц среднее значение площади под кривой $\overline{AUC} = 0,8677$. Аналогично вычислены средние значения для всех 313 количественных признаков MaZda каждой возрастной группы.

В дополнение к рассмотренным выше проведены эксперименты по оценке эффективности количественных признаков на основе многосортных матриц совместной встречаемости [6], которые представляют собой многомерные гистограммы. Каждый элемент матрицы содержит частоту встречаемости элементарных структур изображения. В качестве элементарных структур использовались пары и тройки пикселей изображения, расположенные в вершинах равносторонних треугольников с различной длиной стороны. Соответствующие матрицы обозначаются как IID и IID, где I – интенсивность пикселей изображения, D – расстояние между пикселями изображения. Название количественного признака кодируется следующим образом: <тип признака>–<количество уровней разбиения по яркости>–<расстояние между элементами изображения>. Например, IID-128-3 является признаком типа IID со стороной треугольников, равной трем элементам изображения, для которого диапазон яркостей элементов изображения был квантизован на 128 равных уровней. В ходе экспериментов были использованы количественные признаки

следующих типов: IID-8-1, IID-8-3, IID-16-1, IID-16-3, IID-32-1, IID-32-3, IID-64-1, IID-64-3, IID-128-1, IID-128-3, IID-8-1, IID-8-3, IID-16-1, IID-16-3, IID-32-1, IID-32-3, IID-64-1, IID-64-3, IID-128-1, IID-128-3.

Для всех количественных признаков на основе многосортных матриц совместной встречаемости (общее количество типов – 20) были вычислены оценки эффективности для разделения МРТ-изображений мышц в соответствии с принадлежностью изображений классам здоровых собак и собак с мышечными нарушениями. Количественные признаки вычислялись для всех используемых в экспериментах мышц: EDL-G, EDL-D, GASLAT-G, GASLAT-D, GASMED-G, GASMED-D, TC-G, TC-D. Возрастные группы собак были теми же, что и в описанных ранее экспериментах: 2, 4, 6, 8 и 12 месяцев. Тип релаксации при формировании МРТ-изображений – T2.

В результате для каждой возрастной группы было проведено 160 (20 типов признаков на 8 типов мышц) экспериментов. Данные для каждого эксперимента были подготовлены в виде специальной таблицы (табл. 2), в которой каждая строка соответствовала слою изображения класса GRMD либо Healthy, а столбцы – элементам некоторого данного количественного признака.

Таблица 2

Пример таблицы, подготовленной для оценки эффективности многосортных матриц совместной встречаемости

Класс	Признак 65	Признак 76	Признак 77	Признак 79
month2 GRMD T2 EDL-D	110	1932	3032	61
month2 GRMD T2 EDL-D	36	484	1426	126
month2 GRMD T2 EDL-D	24	94	717	47
month2 Healthy T2 EDL-D	326	1699	467	60
month2 Healthy T2 EDL-D	96	1154	196	20
month2 Healthy T2 EDL-D	169	1921	723	12

Признак 65 является 65-м элементом вектора количественного признака, остальные элементы кодируются аналогично. Столбцы с нулевой дисперсией значений были удалены.

Таблицы в виде, аналогичном табл. 2, являются входными данными для последующего анализа методом кривых оперативной характеристики надежности (ROC-кривых). Выходными данными являются значения площади под ROC-кривой. В экспериментах для вычисления характеристик ROC-кривых использовались два классификатора: «случайный лес» (Random Forest) с 50 деревьями и метрический классификатор « k -ближайших соседей» с евклидовой метрикой, для которого $k = 5$.

По итогам экспериментов была проведена перекрестная проверка с количеством выборок, равным 5. В качестве примера ниже приведены первые три строки выходной таблицы.

Идентификатор эксперимента	AUC_knn	AUC_forest
IID-16-1 month8 T2 EDL-D	0,7175	0,7879
IID-16-1 month8 T2 EDL-G	0,7928	0,8667
IID-16-1 month8 T2 GASLAT-D	0,9500	0,9733

На основе сравнения количественных признаков сделаны следующие выводы: оценка эффективности признаков на основе анализа ROC-кривых для всех возрастных групп и всех мышц была выше у признаков, основанных на матрицах совместной встречаемости Харалика, чем у интегральных признаков текстур. Также установлено, что эффективность матриц совместной встречаемости сравнима с эффективностью признаков на основе многосортных матриц совместной встречаемости. При этом для всех возрастных групп для мышц GASMED-D, GASMED-G и GASLAT-D качество дифференциации на соответствие группам GRMD и Healthy является наилучшим.

4. Количественные признаки текстур на основе гистограмм анизотропии

Проведены два эксперимента для оценки эффективности признаков на основе гистограмм анизотропии. В первом эксперименте для изображений каждой группы мышц (EDL-G, EDL-D, GASLAT-G, GASLAT-D, GASMED-G, GASMED-D, TC-G, TC-D) всех возрастных групп собак

были вычислены гистограммы анизотропии, а по гистограммам анизотропии – параметры F1, F2 и F3 [7, 8]. Во втором эксперименте по изображениям каждой группы мышц были вычислены вначале матрицы совместной встречаемости типов IID, а затем гистограммы анизотропии матриц совместной встречаемости и те же параметры гистограмм – F1, F2 и F3.

В обоих экспериментах для параметров F1, F2 и F3 каждой группы мышц был проведен поиск параметра с наилучшими характеристиками. Решающим параметром выступала суммарная площадь под ROC-кривой для данного типа параметра. На рис. 5 приведены примеры ROC-кривых для гистограмм анизотропии исходных изображений.

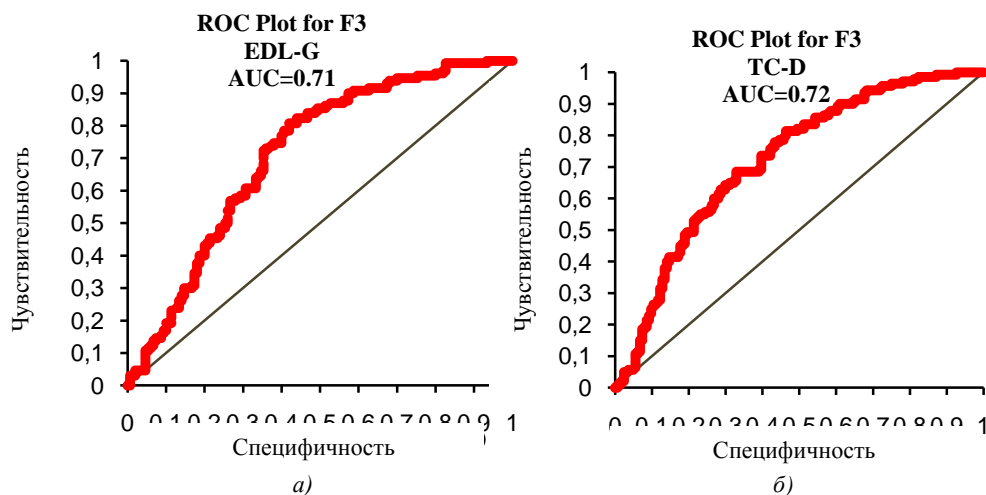


Рис. 5. ROC-кривые для параметра F3 гистограммы анизотропии исходных изображений мышц: а) EDL-G; б) TC-D

Полученные на основе анализа кривых оперативной характеристики надежности результаты показали, что два из трех признаков (F2 и F3) на основе анизотропии яркости элементов изображения обладают дискриминантными свойствами только для некоторых категорий мышц (имеют суммарную площадь под ROC-кривой более 0,7, что соответствует удовлетворительному значению оценки), а один из признаков (F1) ни для одной категории мышц не обладает дискриминантными свойствами. Это же верно и для признаков на основе анизотропии матриц совместной встречаемости исходных изображений.

Также было показано, что лучшие показатели исследованных количественных признаков не достигают значений показателей признаков, приведенных в данной работе ранее.

5. Объединенные количественные признаки текстур МРТ-изображений

В предыдущих разделах был приведен анализ эффективности дискриминации количественных признаков текстур следующих типов для МРТ-изображений мышц:

- интегральных признаков текстур, вычисленных с использованием специализированного программного обеспечения MaZda (313 видов признаков);
- признаков на основе обобщенных матриц совместной встречаемости с использованием значений яркостей элементов изображений (20 видов признаков);
- признаков на основе гистограмм анизотропии (8 видов признаков).

Всего был проанализирован 341 признак. В рамках исследований, представленных в данной статье, была предусмотрена разработка объединенных признаков текстур мышц, которые могли бы повысить эффективность количественной оценки уровня дегенерации мышц. Обычно при анализе большого количества признаков проводится сокращение их числа с помощью метода главных компонент. Данный метод позволяет выделить некоррелированные признаки, называемые главными компонентами. Если имеется одна или несколько компонент (как правило, существенно меньше, чем общее их количество), которые превышают заранее заданный порог дисперсии вектора данных признаков, все остальные компоненты можно не использовать в дальнейшем анализе. С этой целью интегральные признаки текстур были проанализированы с использованием метода главных компо-

нент. Главная компонента PC1 имела значение дисперсии выше 95 % от дисперсии всего множества векторов данных. Был определен интегральный признак $S(2,0)SumAverg$, который имеет наибольший коэффициент корреляции с PC1. Использование метода главных компонент для количественных признаков на основе матриц совместности не проводилось по причине различной размерности указанных признаков. Вместо этого были отобраны два признака (IID-8-1 и IID-16-1) с наибольшим средним значением площади под ROC-кривой по всем мышцам, которые было решено использовать совместно в качестве объединенного признака.

Описанная процедура позволяет существенно сократить пространство признаков для количественной оценки текстур мышц на МРТ-изображениях посредством выбора наиболее информативных. После нормализации значений главных компонент признаков IID-8-1 и IID-16-1, а также значений признака $S(2,0)SumAverg$ получается объединенный набор данных для каждой из групп образцов. Пример объединенного набора значений приведен в табл. 3.

Таблица 3

Значения интегрального признака $S(2,0)SumAverg$ и главных компонент признаков IID-8-1 и IID-16-1 после нормализации для 16 текстур класса GRMD и 26 текстур Healthy образцов мышцы EDL-D, возрастная группа «2-й месяц»

GRMD					Healthy				
0,8357	0,8619	0,1153	0,7966	0,6335	0,0636	0,5151	1,0000	0,5180	0,2952
0,7995	0,8118	0,1500	0,8298	0,5662	0,1625	0,6200	0,9140	0,5195	0,2565
0,7760	0,7475	0,2009	0,6916	0,4953	0,0576	0,5262	0,9452	0,5147	0,2459
0,8296	0,8723	0,1276	0,8559	0,5856	0,0497	0,5832	0,9236	0,5231	0,2112
0,8016	0,8171	0,1448	0,8411	0,5283	0,0000	0,5519	0,9801	0,5193	0,2501
0,8432	0,8200	0,0942	0,4647	0,7403	0,0658	0,4900	1,0000	0,5180	0,1306
0,8822	0,9662	0,0594	0,9715	0,4295	0,1753	0,4493	1,0000	0,5180	0,1050
0,8885	0,8561	0,0579	0,6501	0,4355	1,0000	0,1753	0,0000	0,0000	0,0506
0,8862	1,0000	0,0558	1,0000	0,4198	0,9646	0,0734	0,0232	0,0419	0,0000
0,8598	0,8942	0,0594	0,9855	0,6609	0,8798	0,0000	0,2765	0,2739	0,0510
0,9000	0,7127	0,0708	0,2871	1,0000	0,9677	0,0804	0,1323	0,1256	0,0609
0,8968	0,7337	0,0527	0,3366	0,6460	0,8796	0,8104	0,0923	0,4314	0,1237
0,8805	0,7361	0,0464	0,2324	0,7620	0,2287	0,3929	1,0000	0,5180	0,0347
0,8184	0,8271	0,1313	0,7972	0,6506	0,9821	0,1310	0,0399	0,0577	0,0887
0,7445	0,6860	0,2437	0,6260	0,7481	0,8932	0,0682	0,2451	0,2800	0,0278
0,8481	0,9197	0,0875	0,9660	0,5467	0,8758	0,0352	0,1786	0,2198	0,0854
					0,8058	0,0652	0,4403	0,3504	0,0442
					0,8875	0,0655	0,2057	0,1765	0,0668
					0,3096	0,4353	0,8941	0,5041	0,0967
					0,1509	0,4625	0,9622	0,5136	0,1151
					0,2223	0,5906	0,8292	0,5154	0,3198
					0,3633	0,6406	0,7009	0,5378	0,2886
					0,1966	0,5754	0,8468	0,5180	0,2311
					0,2401	0,5630	0,8182	0,5151	0,2824
					0,3061	0,4307	0,8654	0,4890	0,4472
					0,2663	0,5971	0,6903	0,5269	0,1328

Перекрестная проверка (кросс-валидация) предсказательной эффективности статистической модели, построенной на полученных признаках, осуществлялась с использованием стратегии leave-one-out. Из множества $K = P + N$ (где P – количество векторов, относящихся к классу GRMD, а N – к классу Healthy) векторов поочередно удалялся один вектор v , производилась тренировка на оставшихся $K-1$ векторах, а затем следовало предсказание принадлежности вектора v к классам GRMD и Healthy. После этого строилась матрица неточностей, элементами которой являлись количества истинно положительных (TP), истинно отрицательных (TN), лож-

ноположительных (FP) и ложноотрицательных (FN) случаев. Для рассматриваемого примера (мышца EDL-D, возрастная группа «2-й месяц») матрица неточностей представлена как

$$\begin{aligned} TP &= 16, & FN &= 0, \\ FP &= 1, & TN &= 25, \end{aligned}$$

откуда чувствительность модели (sensitivity) $TP/(TP+FN) = 1$; специфичность (specificity) $TN/(TN+FP) = 0,9615$, точность предсказания (ACC) $(TP + TN)/(P + N) = 0,976$.

Таким образом, были реализованы и протестированы объединенные количественные признаки текстур мышц на МРТ-изображениях (интегральный признак $S(2,0)SumAverg$ и два признака на основе обобщенных матриц совместной встречаемости – IID-8-1 и IID-16-1). В результате тестирования объединенных признаков показано, что их эффективность для дифференциации текстур, относящихся к классам здоровых и GRMD-образцов, выше по сравнению с дифференциацией с использованием каждого признака по отдельности. Эффективность признаков оценивалась по таким параметрам, как специфичность, чувствительность и точность.

6. Тестирование всех разработанных признаков на дополненной выборке МРТ-изображений мышц

Построена таблица ошибок классификации мышц из подвыборок AG_1 и AG_2 по 333 количественным признакам (как интегральным, так и основанным на матрицах совместной встречаемости элементов текстур). Были определены признаки, наиболее эффективные для разделения текстур на классы, принадлежащие подвыборкам AG_1 и AG_2 (табл. 4 и 5).

Таблица 4

Лучшие признаки для каждой из мышц по значению площади под ROC-кривой при дифференциации мышц на подвыборки AG_1 и AG_2

Мышца	Признак	AUC	Мышца	Признак	AUC
AL_G	S(1,1)InvDfMom	0,947044	SM_G	S(0,1)InvDfMom	0,957493
AL_D	S(0,1)InvDfMom	0,960591	SM_D	S(1,1)InvDfMom	0,962506
AM_G	S(1,1)InvDfMom	0,961146	ST_G	S(1,1)InvDfMom	0,982129
AM_D	S(1,-1)InvDfMom	0,954125	ST_D	S(2,0)InvDfMom	0,958039
BF_G	S(0,1)DifEntrp	0,943304	VI_G	S(1,1)InvDfMom	0,889981
BF_D	S(1,1)InvDfMom	0,976403	VI_D	S(2,-2)Correlat	0,876541
RF_G	S(1,-1)InvDfMom	0,864044	VL_G	S(0,3)InvDfMom	0,917637
RF_D	S(1,0)DifEntrp	0,834921	VL_D	S(1,0)DifEntrp	0,862972

Таблица 5

Пять лучших количественных признаков по среднему для всех мышц значению площади под кривой оперативной характеристики надежности для разделения областей изображений мышц на подвыборки AG_1 и AG_2

Признак	AUC
S(1,0)InvDfMom	0,902574
S(1,1)InvDfMom	0,899792
S(0,1)InvDfMom	0,899756
(0,1)DifEntrp	0,892268
S(1,-1)InvDfMom	0,891849

Заключение

На основе проведенных исследований были сделаны следующие выводы:

– для первой выборки изображений наиболее высокой для всех возрастных групп и всех мышц оказалась оценка эффективности признаков на основе матриц совместной встречаемости Харалика и на основе многосортных матриц совместной встречаемости. При этом для всех воз-

растных групп для мышц GASMED-D, GASMED-G и GASLAT-D качество разделения на группы изображений мышц здоровых собак (Healthy) и изображений мышц собак, пораженных дистрофией (GRMD), является наилучшим;

– для второй выборки МРТ-изображений мышц 20 здоровых людей-добровольцев лучшими дискриминационными характеристиками обладает признак «обратный дифференциальный момент», также интерпретируемый как «однородность» текстуры.

Дальнейшие исследования будут направлены на разработку и тестирование автоматизированных методов сегментации мышц на изображениях МРТ для дальнейшего текстурного анализа сегментированных областей изображений.

Данная работа выполнена в рамках проекта БРФФИ – ИЦНИ Ф11Ф-005.

Список литературы

1. Characterization of dystrophic muscle in golden retriever muscular dystrophy dogs by nuclear magnetic resonance imaging / J.L. Thibaud [et al.] // *Neuromuscul Disord.* – 2007. – Vol. 17. – P. 575–584.
2. Mutation segregation and rapid carrier detection of X-linked muscular dystrophy in dogs / R.J. Bartlett [et al.] // *Am J Vet Res.* – 1996. – Vol. 57. – P. 650–654.
3. Mammalian animal models for Duchenne muscular dystrophy / R. Willmann [et al.] // *Neuromuscul Disord.* – 2009. – Vol. 19. – P. 241–249.
4. MaZda – A software package for image texture analysis / P. Szczypinski [et al.] // *Computer Methods and Programs in Biomedicine.* – 2009. – Vol. 94, no. 1. – P. 66–76.
5. Szczypinski P. MaZda – a Software for Texture Analysis / P. Szczypinski, M. Strzelecki, A. Materka // *Proc. of ISITC 2007.* – Republic of Korea, 2007. – P. 245–249.
6. Ковалев, В.А. Поиск медицинских изображений по содержанию в условиях шумов / В.А. Ковалев, А.А. Дмитрук // *Информатика.* – 2009. – № 2 (21). – С. 39–46.
7. Kovalev, V. Texture anisotropy in 3D images / V. Kovalev, M. Petrou, Y. Bondar // *IEEE Transactions on Image Processing.* – Vol. 8, no 3. – 1999. – P. 346–360.
8. Kovalev, V. Texture analysis in three dimensions as a cue to medical diagnosis / V. Kovalev, M. Petrou; ed. by I. Bankman // *Handbook of medical imaging: processing and analysis.* – San Diego : Academic Press, 2000. – P. 231–247.

Поступила 19.06.2014

¹Объединенный институт проблем информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: eduard.snezhko@gmail.com

²Institut Myologie,
Batiment Babinski,
47/83 Bd de L'hôpital, Paris
e-mail: p.carlier@institut-myologie.com

E.V. Snezhko, P. Carlier, V.A. Kovalev, N. Azzabou, A.A. Dmitruk, A.V. Shukelovich

APPLICATION OF TEXTURE ANALYSIS TECHNIQUES TO NMR IMAGES FOR QUANTITATIVE ASSESSMENT OF MUSCLE DISORDERS

Results on development of the basic concept of the medical image analysis methodology are presented. The techniques used in the research are based on adaptation and utilization of the generalized image features for quantitative interpretation of muscles on nuclear magnetic resonance images.

УДК 004

Б.А. Залесский

МЕТОД НЕЧЕТКОЙ КЛАСТЕРИЗАЦИИ k -СРЕДНИХ СО СГЛАЖИВАЮЩЕЙ ШТРАФНОЙ ФУНКЦИЕЙ

Рассматривается новый подход к кластеризации полутоновых, цветных и мультиспектральных изображений на основе метода нечеткой кластеризации k -средних. В нем предполагается локальная однородность кластерного представления в областях, не разделенных градиентом исходного изображения, за счет использования аддитивной штрафной функции гиббсовского типа, задающей степень гладкости решения в соседних пикселях, в зависимости от величины и направления градиента, проходящего между ними. Подход, обобщающий метод нечеткой кластеризации k -средних, позволяет, с одной стороны, получить более однородное кластерное представление исходного изображения, а с другой – предотвращает слияние разных кластеров. Он протестирован на модельных изображениях, аэрофотоснимках и космических изображениях, в частности на восьмиканальных мультиспектральных изображениях спутника LANDSAT 7. Приводятся результаты вычислительных экспериментов, которые подтверждают эффективность нового алгоритма.

Введение

Метод кластеризации k -средних (СМ), предложенный в 1950-х гг. Г. Штейнгаузом [1] и С. Ллойдом [2], широко используется в настоящее время для обработки числовых данных и изображений, в том числе для кластеризации и сегментации полутоновых, цветных и мультиспектральных космических снимков. В 1980-е гг. появилось обобщение метода, полученное Дж. Бездеком [3] и названное методом нечеткой кластеризации k -средних (FCM).

При использовании СМ предполагается, что число кластеров k задано и каждый вектор данных (или пиксел изображения) принадлежит ровно одному кластеру [1, 4]. Действие алгоритма состоит в минимизации среднеквадратичного отклонения векторов кластера от его центра масс. Изначально в качестве центров выбираются произвольные k векторов данных, например, случайным образом. Производится кластеризация данных путем отнесения каждого вектора к ближайшему центру. На каждой итерации переопределяется центр масс для каждого кластера, полученного на предыдущем шаге, затем векторы данных разбиваются на кластеры вновь в соответствии с тем, какой из новых центров оказался ближе к вектору в выбранной метрике. При этом чаще всего используется евклидова метрика. Недостатком СМ является его сходимость в общем случае к локальному экстремуму. На практике это приводит, например, к тому, что при кластеризации изображений алгоритм может объединить разные кластеры в один, даже если их число меньше разрешенного максимального значения k .

Главным отличием FCM от классического СМ является то, что FCM, основанный на понятиях нечеткой логики, при кластеризации векторных данных ставит в соответствие каждому вектору наблюдений не один класс, а набор чисел, которые характеризуют степень принадлежности вектора каждому классу. Нередко в качестве такого набора используется вектор вероятностей принадлежности элемента выборки каждому классу.

На данный момент можно найти большое количество публикаций, посвященных различным модификациям FCM [5–7]. Разнообразие алгоритмов объясняется тем, что на практике невозможно выделить алгоритм, который работал бы одинаково эффективно для всех классов задач.

В настоящей работе предлагается новый метод сегментации полутоновых, цветных и мультиспектральных изображений на основе FCM. В нем задается локальная степень гладкости кластерного представления в каждом пикселе в зависимости от значения и направления градиента исходного изображения. Для этого к целевой функции добавляется новое слагаемое – квадратичная штрафная функция, зависящая от разностей функций принадлежности соседних пикселей и градиента.

Эксперименты по кластеризации спутниковых снимков с целью выделения на них сельскохозяйственных культур и созданных с помощью графического редактора тестовых изображений показали, что предложенный метод превосходит СМ и FCM по точности. Он существенно более устойчив к искажениям и шумам, точнее выделяет границы кластеров.

1. Описание метода

Пусть \mathbf{I} – $m \times n$ полутоновое, трехканальное цветное или d -канальное мультиспектральное изображение с пикселями $p = (p_x, p_y)$ и векторами-столбцами значений яркостей каждого канала $I_p = (I_{1,p}, \dots, I_{d,p})^T$ и пусть k – заранее заданное число кластеров. Принадлежность каждого пикселя p изображению \mathbf{I} кластеру j , $1 \leq j \leq k$, задается матрицей принадлежности $\mathbf{U} = (u_{p,j})$ с элементами $u_{p,j} \in [0, \infty)$, которые можно интерпретировать как степень принадлежности пикселя p классу j . В дальнейшем будем предполагать, что для каждого пикселя p выполняется условие $\sum_{j=1}^k u_{p,j} = 1$, тогда элементы $u_{p,j}$ можно понимать как вероятности принадлежности p каждому из k кластеров. Метод FCM нечеткой кластеризации k -средних заключается в условной EM-минимизации функции

$$F(\mathbf{U}, \mathbf{C}) = \sum_{j=1}^k \sum_p u_{p,j}^\gamma \|I_p - \mathbf{c}_j\|^2 \quad (1)$$

при условиях $\sum_{j=1}^k u_{p,j} = 1$ для всех пикселей p по матрицам \mathbf{U} (размерности $m \times n \times k$) и $\mathbf{C} = (\mathbf{c}_1, \dots, \mathbf{c}_k)$ (размерности $d \times k$), последняя из которых образована центрами кластеров. Параметр $\gamma \in [1, \infty]$ выбирается заранее и остается неизменным при минимизации – он характеризует нечеткость кластеров. EM-минимизация производится пошагово: 1) при текущих значениях элементов \mathbf{U} , которые перед началом вычислений выбираются случайным образом, рассчитываются центры кластеров \mathbf{C} ; 2) затем вычисляется условный минимум функции $F(\mathbf{U}, \mathbf{C})$ по \mathbf{U} при условиях $\sum_{j=1}^k u_{p,j} = 1$; 3) после этого снова находятся центры кластеров. Шаги 2 и 3 повторяются до тех пор, пока не будет достигнута требуемая точность вычисления \mathbf{U} либо выполнено заданное число операций [5].

Для кластеризации изображений предлагается модификация метода FCM, называемая в дальнейшем методом нечеткой кластеризации k -средних со сглаживающей штрафной функцией (Gradient Penalty Fuzzy C-Means – GPFCM). Как и FCM, предлагаемый метод основан на процедуре EM-минимизации целевой функции – критерия качества кластеризации. Новая целевая функция в отличие от старой задает степень локальной связности кластерного представления в каждом пикселе в зависимости от градиента исходного изображения. Она имеет следующий вид:

$$F_{GP}(\mathbf{U}, \mathbf{C}) = \sum_{j=1}^k \sum_p u_{p,j}^2 \|I_p - \mathbf{c}_j\|^2 + \sum_{j=1}^k \sum_p \left[\mu_{p,x} (u_{p,j} - u_{p+(1,0),j})^2 + \mu_{p,y} (u_{p,j} - u_{p+(0,1),j})^2 \right] \quad (2)$$

при условиях $\sum_{j=1}^k u_{p,j} = 1$, которые выполняются для каждого пикселя p изображения.

В формуле (2), как обычно, $\|\cdot\|$ – евклидова норма. Для пикселя p пиксел $p + (1, 0)$ – его правый сосед, а пиксел $p + (0, 1)$ – нижний сосед. Значения параметров $\mu_{p,x}$ и $\mu_{p,y}$ задаются в зависимости от величины соответствующей координаты градиента. Обозначим через $G_p = (G_{p,x}, G_{p,y})$ градиент изображения \mathbf{I} в пикселе p . Параметр $\mu_{p,x}$ выбирается в зависимости от величины $G_{p,x}$, а $\mu_{p,y}$ – в зависимости от величины $G_{p,y}$:

$$\mu_{p,x} = \begin{cases} \mu, & \text{если } |G_{p,x}| \leq \tau, \\ \frac{\mu}{|G_{p,x}|+1}, & \text{если } |G_{p,x}| > \tau, \end{cases} \text{ и } \mu_{p,y} = \begin{cases} \mu, & \text{если } |G_{p,y}| \leq \tau, \\ \frac{\mu}{|G_{p,y}|+1}, & \text{если } |G_{p,y}| > \tau, \end{cases}$$

где $\mu > 0$ задает степень гладкости кластерного представления, а $\tau > 0$ – пороговое значение для градиента, предназначенное для уменьшения влияния на результат кластеризации градиентов малой величины, возникновение которых часто объясняется шумами и искажениями исходного изображения **I**.

Для вычисления разностного градиента G_p использовались 3×3 операторы Превитт или Собеля [4], хотя очевидно, что могут быть применены и другие операторы.

Приведем пошаговое описание GPFCM.

Шаг 1. Задаем матрицу **U** с элементами $u_{p,j} \in [0, 1]$ случайным образом с помощью датчика случайных чисел так, чтобы для каждого пиксела p и целого $i = 1 \div k$ выполнялись условия

$$\sum_{j=1}^k u_{p,j} = 1 \text{ и } \sum_p u_{p,i} > 0.$$

Шаг 2. Вычисляем положение центров кластеров $\mathbf{c}_j = (c_{1,j}, \dots, c_{d,j})^T$ по формуле

$$c_{i,j} = \frac{\sum_p u_{p,j}^2 I_{i,p}}{\sum_p u_{p,j}^2}, \quad i = 1 \div d.$$

Шаг 3. Задаем значения множителей Лагранжа λ_p для условной минимизации целевой функции $F_{GP}(\mathbf{U}, \mathbf{C})$ по **U** при условиях $\sum_{j=1}^k u_{p,j} = 1$ для всех пикселей изображения как

$$\lambda_p = \frac{2}{k} \sum_{j=1}^k u_{p,j} \|I_p - \mathbf{c}_j\|^2.$$

Шаг 4. Определяем элементы матрицы принадлежности **U** для всех пикселей p изображения и всех кластеров j с помощью равенства

$$u_{p,j} = \frac{\frac{1}{2}\lambda_p + \mu_{p,x} u_{p+(1,0),j} + \mu_{p,y} u_{p+(0,1),j} + \mu_{p-(1,0),x} u_{p-(1,0),j} + \mu_{p-(0,1),y} u_{p-(0,1),j}}{\|I_p - \mathbf{c}_j\|^2 + \mu_{p,x} + \mu_{p,y} + \mu_{p-(1,0),x} + \mu_{p-(0,1),y}}.$$

Шаг 5. Проверяем условие окончания вычислений, которое может быть проверкой достижения заданной точности определения матрицы принадлежности **U**, или проверкой количества повторенных итераций, или иным условием. Если условие окончания не выполнено, переходим на шаг 2, иначе STOP.

2. Результаты тестирования и сравнения алгоритмов кластеризации

Исследование свойств предложенного метода кластеризации проводилось на серии искусственно созданных с помощью графического редактора изображений, а также на публично доступных аэрофотоснимках и мультиспектральных снимках со спутников QuickBird и LANDSAT 7.

Тестирование на искусственно созданных изображениях позволило сравнить предложенный GPFCM с такими известными методами, как FCM и CM, а также оценить его точность путем сравнения результатов кластеризации зашумленных изображений с исходными незашумленными. Классические FCM и CM неустойчивы к шумам и помехам: присутствие искажений приводит к «распадению» кластеров на маленькие однородные по характеристикам области, не связанные между собой. Кроме того, тестирование CM, сходящегося в общем случае к одному из локальных экстремумов, показало, что на аэрокосмических изображениях его применение нередко приводит к «пропаданию кластеров» – слиянию больших разнородных областей в один кластер. Результаты CM оказались неустойчивыми относительно начальных условий: повторное применение метода нередко приводило к различающимся кластерным представлениям, поэтому ниже показаны результаты сравнения GPFCM с более надежным методом FCM.

Эксперименты показали, что предложенный метод GPFCM оказался в данном случае существенно более устойчивым. На рис. 1, *а* приведено тестовое изображение, зашумленное гауссовским случайным шумом с параметром $\sigma = 80$ (рис. 1, *б*). Результат работы FCM показан на рис. 1, *в*, а результат работы предложенного GPFCM – на рис. 1, *г*.

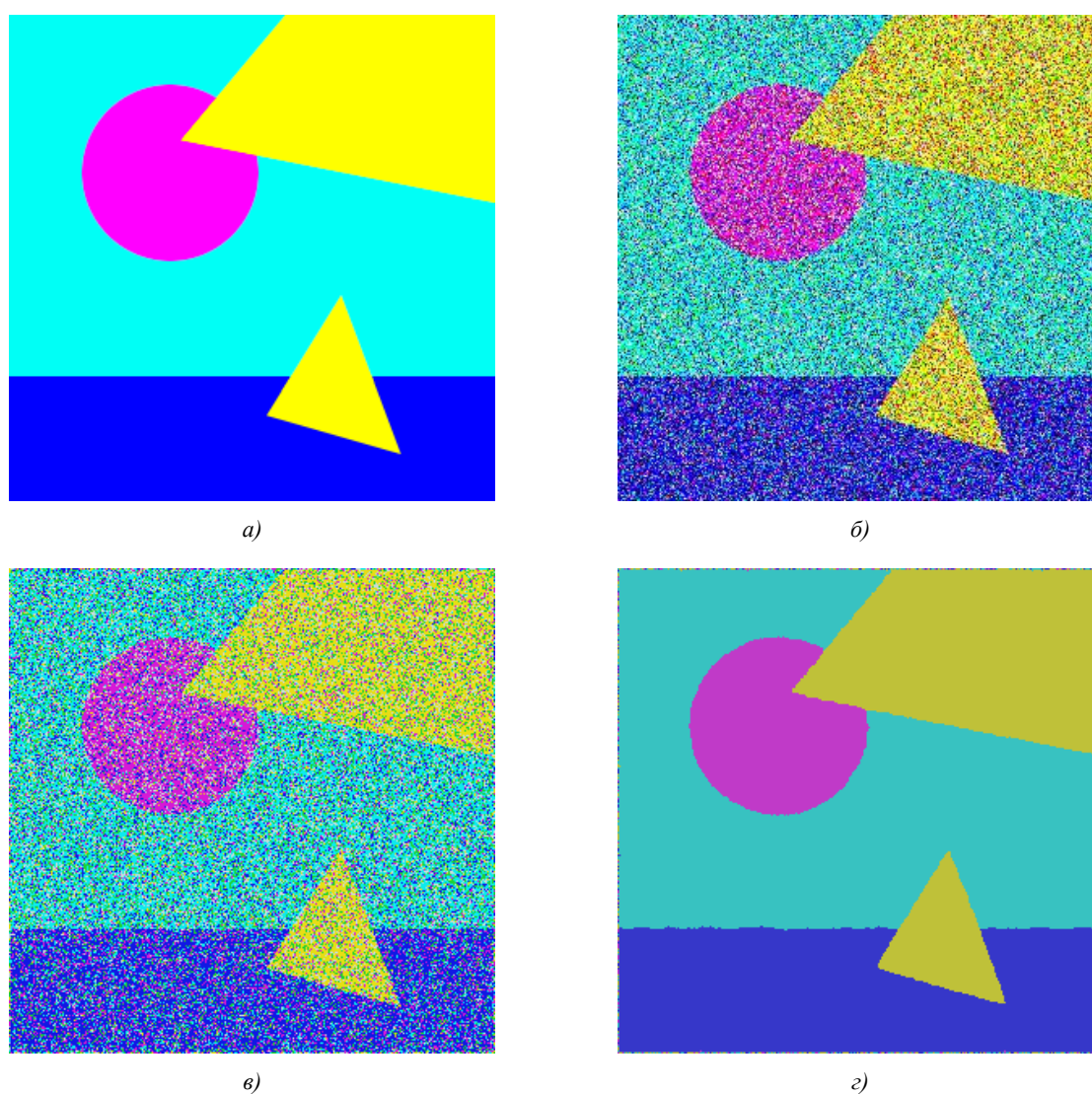


Рис. 1. Пример обработки синтезированного RGB-изображения с шумом: *а*) исходное изображение; *б*) исходное изображение, наблюдаемое с аддитивным гауссовским шумом; *в*) зашумленное изображение, обработанное известным методом FCM; *г*) зашумленное изображение, обработанное предложенным GPFCM

Следует заметить, что предобработка зашумленного изображения на рис. 1, б известными фильтрами для устранения шума с последующим применением FCM дает худший результат, чем непосредственное применение предложенного GPFCM. Постобработка изображения на рис. 1, в – результата применения FCM известными фильтрами с целью получения кластерного представления – также дает худшие результаты, чем непосредственное применение GPFCM.

Заметим также, что после применения FCM и GPFCM цвета кластеров немного отличаются от цветов исходного изображения. На рис. 1 цвета исходного изображения а) и результатов в) и г) отличаются на две-четыре градации яркостей по каждому цветовому каналу. Это отличие вызвано тем, что действие шума в каждом пикселе будет приблизительно симметричным только в том случае, когда значение цвета каждого его канала равно 128.

Для иллюстрации данного эффекта на рис. 2, а приведено однотонное полутоновое изображение с яркостью каждого пиксела, равной 128, искаженное гауссовским случайным шумом с $\sigma=80$ (рис. 2, б). Результаты его фильтрации FCM и предложенным GPFCM изображены на рис. 2, в и г. В данном случае GPFCM действительно дает нулевое смещение, яркость изображения – результата работы GPFCM – действительно равна 128.

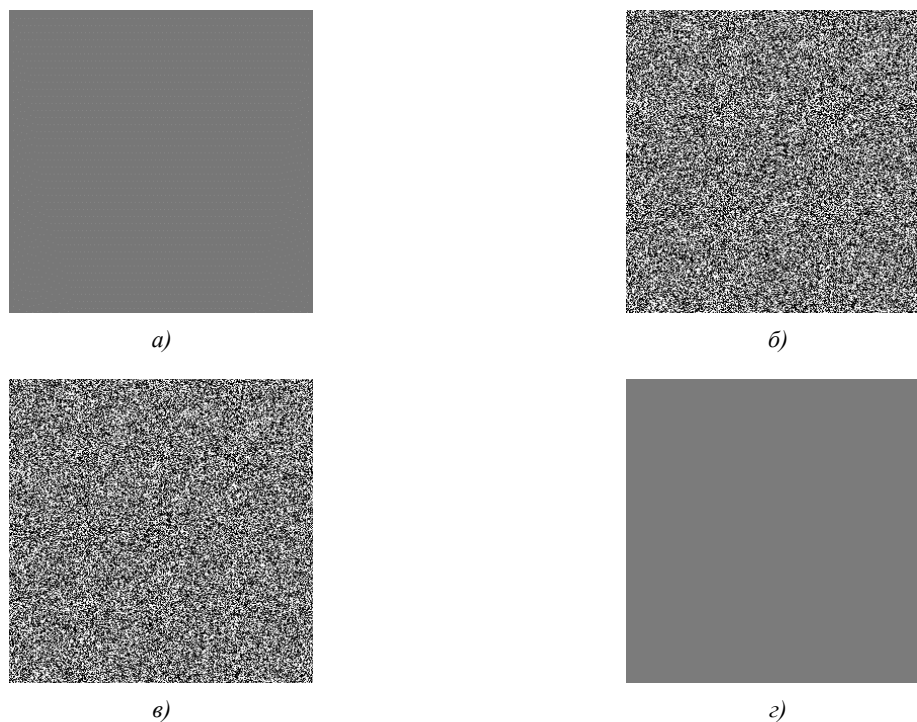


Рис. 2. Пример обработки однородного полутонового изображения: а) исходное однородное полутоновое изображение; б) исходное изображение, наблюдаемое с аддитивным гауссовским шумом; в) зашумленное изображение б), обработанное известным методом FCM; г) зашумленное изображение б), обработанное предложенным GPFCM

На рис. 3 показаны результаты применения FCM и GPFCM к аэрофотоснимку. Кластерное представление, полученное GPFCM, имеет большую регулярность. Далее приводятся результаты кластеризации мультиспектрального снимка с LANDSAT 7 ETM+ с помощью FCM и GPFCM вместе с наложенными на них границами сельскохозяйственных культур, выделенными специалистами РУП «БелНИЦзем». На рис. 4, а показано псевдоцветное RGB-изображение, образованное из четвертого, третьего и второго каналов мультиспектрального снимка с LANDSAT 7. На рис. 4, б приведен результат его кластеризации FCM вместе с наложенными границами сельскохозяйственных культур, на рис. 4, в – результат кластеризации GPFCM с наложенными границами сельскохозяйственных культур.



Рис. 3. Результат применения FCM и предложенного GPFCM к аэрофотоснимку: *а)* часть исходного аэрофотоснимка; *б)* результат применения к нему FCM; *в)* результат применения к исходному изображению *а)* GPFCM; *г)* результат применения к исходному изображению *а)* GPFCM вместе с границами кластеров

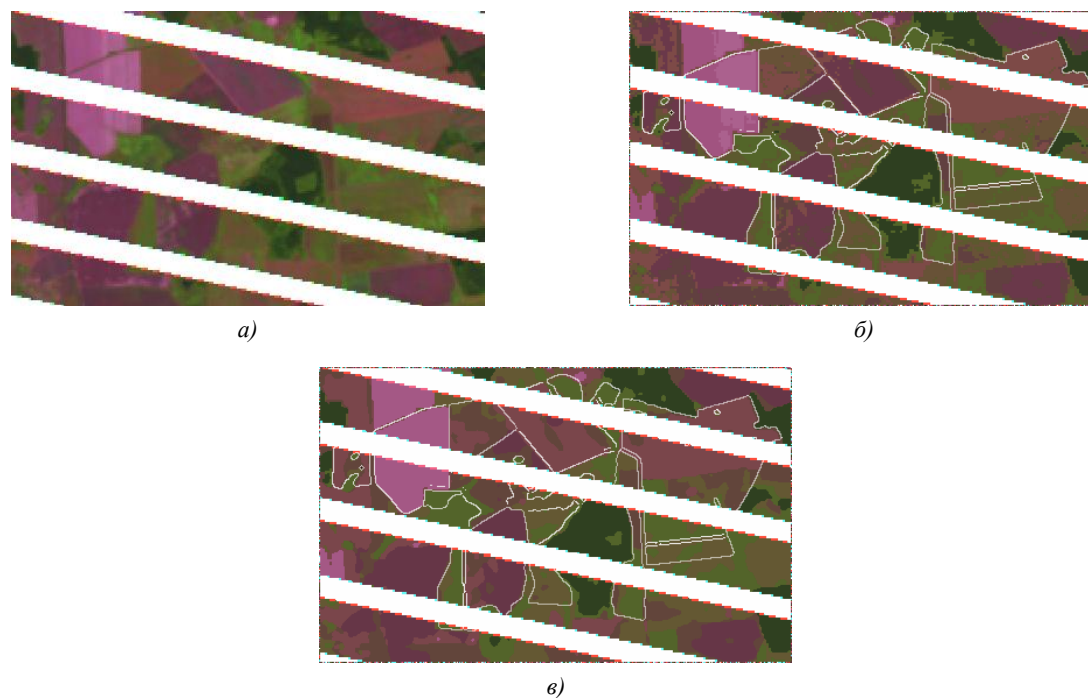


Рис. 4. Результат кластеризации мультиспектрального космического снимка, полученного LANDSAT 7 ETM+, с помощью FCM и предложенного GPFCM: *а)* композитное RGB-изображение, составленное из трех каналов исходного мультиспектрального снимка; *б)* результат кластеризации *а)* с помощью FCM вместе с наложенной на него маской сельскохозяйственных культур, выполненной экспертом; *в)* результат кластеризации *а)* с помощью GPFCM вместе с наложенной на него маской сельскохозяйственных культур, выполненной экспертом

Заключение

Предложенный метод GPFCM показал более высокую точность кластеризации по сравнению с известными методами FCM и CM как на синтетических, так и на реальных аэрофотоснимках и космических мультиспектральных изображениях. Вычислительные эксперименты подтвердили также, что он значительно устойчивее известных методов CM и FCM при наличии на изображении шумов и помех. Параметры новой штрафной функции позволяют контролировать поведение метода для достижения более точных результатов. При использовании нулевых значений для параметров локальной связности метод превращается в FCM.

Список литературы

1. Steinhaus, H. Sur la division des corps materiels en parties / H. Steinhaus // Bull. Acad. Polon. – 1956. – Vol. 4 (12). – P. 801–804.
2. Lloyd, S. Least squares quantization in PCM / S. Lloyd // IEEE Transactions on Information Theory. – 1982. – Vol. 28, no. 2. – P. 129–137.
3. Bezdek, J.C. Pattern Recognition with Fuzzy Objective Function Algorithms / J.C. Bezdek. – MA, USA : Kluwer Academic Publishers Norwell, 1981. – 256 p.
4. Гонсалес, Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – М. : Техносфера, 2005. – 1075 с.
5. Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition / F. Höppner [et al.]. – N. Y. : John Wiley & Sons, 1999. – 300 p.
6. MacQueen, J.B. Methods for classification and Analysis of Multivariate Observations / J.B. MacQueen // Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability. – Berkeley, 1967. – P. 281–297.
7. Genitha, C.H. Classification of satellite images using new fuzzy cluster centroid for unsupervised classification algorithm // C.H. Genitha, K. Vani // Proc. IEEE Conf. on Information and Communication Technologies ICT2013. – JeJu Island, 2013. – P. 203–207.

Поступила 20.05.2014

*Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: zalesky@tut.by*

B.A. Zalesky

METHOD FUZZY CLUSTERING k -MEANS WITH SMOOTHING PENALTY FUNCTION

A new method of clustering of grayscale, color and multispectral images is presented. It is based on conditional optimization of the objective function consisting of the classic fuzzy functional criterion and the penalty function of Gibbs type, which controls local smoothness of the solution. The method provides more smooth solutions that are essentially more precise in comparison with fuzzy c-means results in the case of noisy images.

УДК 004.934.1

А.В. Ткаченя

АДАПТАЦИЯ СКРЫТЫХ МАРКОВСКИХ МОДЕЛЕЙ К РАСПОЗНАВАНИЮ ЭМОЦИОНАЛЬНО ОКРАШЕННОЙ РЕЧИ

Рассматривается алгоритм интерактивной неконтролируемой оценки параметров скрытых марковских моделей (СММ). Решается задача адаптации СММ к эмоционально окрашенной речи. Для увеличения достоверности уточненных параметров СММ предлагается механизм забывания и обновления. Приводятся функциональная блок-схема рассматриваемого алгоритма адаптации СММ, а также полученные результаты улучшения эффективности распознавания эмоциональной речи.

Введение

Скрытая марковская модель является стохастическим конечным автоматом, состоящим из конечного множества состояний $Q = \{q_1, q_2, \dots, q_N\}$ и вероятностей непосредственных переходов между ними $A = \{a_{ij}\}$. Каждое такое состояние q_i связано с вектором признаков o_i при помощи матрицы вероятности наблюдения $B = \{b_i(o_i)\}$, каждый элемент которой является эмиссионной плотностью распределения вероятности $P(o_i|q_i)$. Также немаловажным параметром СММ является вероятность начального распределения состояний $\pi = \{\pi_1, \pi_2, \dots, \pi_N\}$, которое задает вероятность того, что первым будет i -е состояние. Сумма π_i должна быть равна единице. Таким образом, оценив параметры $\lambda = (\pi, A, B)$, СММ можно использовать для моделирования последовательности векторов признаков $O = \{o_1, o_2, \dots, o_T\}$ как кусочно-стационарного процесса, в котором каждый стационарный участок относится к определенному состоянию СММ. Такой подход обеспечивает моделирование динамической структуры речевой единицы (т. е. решает проблему различной длительности сигнала, соответствующего одной и той же фонеме).

Известно, что снижение эффективности распознавания речи связано с несоответствием акустических характеристик обучающих и тестируемых данных. При адаптации моделей на основе классических методов оценки параметров СММ, таких как максимизация вероятности [1–4] и максимизация апостериорной вероятности [5–7], обычно используются алгоритмы, которые требуют обработки всех доступных данных целиком. Тем не менее в ряде работ [6, 8, 9] было предложено использовать интерактивную адаптацию СММ, которая позволяет осуществлять последовательную адаптацию на данных, полученных из тестируемой выборки, на этапе распознавания. Этот подход исключает необходимость предварительного сбора данных для адаптации и позволяет осуществить уточнение параметров СММ для каждого конкретного случая.

Таким образом, с помощью данных из тестируемой выборки и их транскрипции, полученной в результате распознавания эмоциональной речи на основе СММ, может быть произведена адаптация СММ с гауссовым распределением значений матрицы вероятности наблюдения B на основе квазибайесовского обучения [6], алгоритм которого заключается в последовательном обновлении параметров скрытых марковских моделей [8]. Такой тип адаптации называется неконтролируемым в противовес контролируемой адаптации, при которой транскрипция всех данных заведомо известна. Так как полученная транскрипция не проверяется вручную, в ней могут содержаться ошибки распознавания, поэтому было предложено использовать механизм забывания [9] и обновления СММ.

Адаптация СММ позволяет снизить несоответствие между акустическими характеристиками полученных моделей и тестируемых данных, повысив, таким образом, эффективность распознавания речи. Возможность эффективного применения интерактивной неконтролируемой адаптации СММ базируется на следующих условиях:

1. Эффективность распознавания эмоциональной речи на исходных (неадаптированных) моделях должна быть достаточно высока, чтобы проводить корректное уточнение, так как адаптация неконтролируема (ручная проверка правильности распознавания и корректности транскрипции распознанных данных не проводится).

2. Вектор признаков должен меняться плавно, так как адаптация на резко меняющихся данных неэффективна.

3. Изменения акустических характеристик должны иметь макроструктуру, т. е. проследиваться на большом количестве фреймов.

Эмоциональная речь характеризуется изменением пространства информативных признаков по сравнению с нейтральной речью в соответствии с выражаемой эмоцией. Эти изменения прослеживаются на длительном промежутке времени и носят квазистационарный характер, что позволяет считать их макрособытиями. Известно также [10], что кепстральные коэффициенты, которые вычисляются на основе спектральных коэффициентов, рассчитанных по параметрам линейного предсказания, наименее восприимчивы к изменению частоты основного тона и вследствие этого дают хорошую эффективность распознавания эмоциональной речи.

1. Алгоритм адаптации параметров СММ

Для того чтобы решить задачу уточнения параметров СММ, вначале необходимо задать вид распределения вероятности наблюдения $b_i(o_t)$. В большинстве случаев требуется моделировать параметры СММ, используя для вероятности наблюдений многомерное распределение с непрерывной плотностью. В то же время при решении некоторых задач можно работать и с последовательностями наблюдений, в которых распределение вероятностей наблюдения дискретно. Для решения задачи адаптации СММ к эмоционально окрашенной речи будем использовать непрерывную плотность распределения вероятности наблюдения.

Для большинства систем СММ с непрерывной плотностью обычно используется распределение вероятности наблюдения в виде гауссовой смеси. При этом каждый вектор признаков в момент времени t можно разбить на S независимых информативных потоков o_{st} , тогда формула для вычисления $b_i(o_t)$ будет иметь следующий вид:

$$b_i(o_t) = \prod_{s=1}^S \left[\sum_{m=1}^{M_s} c_{ism} N(o_{st}; \mu_{ism}, \Sigma_{ism}) \right], \quad (1)$$

где M_s – число компонент гауссовой смеси в потоке s ; c_{ism} – вес m -й компоненты; $N(o; \mu, \Sigma)$ – многомерное распределение Гаусса с вектором математического ожидания μ и ковариационной матрицей Σ :

$$N(o; \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(o-\mu)^T \Sigma^{-1} (o-\mu)}, \quad (2)$$

где n – размерность o .

Для оценки параметров СММ чаще всего используется алгоритм Баума – Велша, который, по сути, является алгоритмом максимизации среднего (*EM algorithm*) [11]. В свою очередь, алгоритм Баума – Велша может быть эффективно реализован при помощи так называемого *forward-backward*-алгоритма [12].

Тогда определим

$$\gamma_i(t) = P(q_t = i | O, \lambda) = \frac{\alpha_i(t) \beta_i(t)}{\sum_{j=1}^N \alpha_j(t) \beta_j(t)} \quad (3)$$

как вероятность нахождения в i -м состоянии в момент времени t для последовательности векторов признаков O и

$$\xi_{ij}(t) = P(q_t = i, q_{t+1} = j | O, \lambda) = \frac{\gamma_i(t) a_{ij} b_j(o_{t+1}) \beta_j(t+1)}{\beta_i(t)} \quad (4)$$

как вероятность нахождения в i -м состоянии в момент времени t и перехода в j -е состояние в момент времени $t + 1$, где α является *forward*-вероятностью, а β – *backward*-вероятностью.

Цель алгоритма Баума – Велша заключается в оценке новых параметров СММ на основании старых параметров СММ и вновь поступивших данных. Определим для гауссовой смеси вероятность m -й компоненты i -й смеси для независимого информативного потока s как

$$\gamma_{ism}(t) = \gamma_{is}(t) \frac{c_{ism} b_{ism}(o_{st})}{\sum_{m=1}^{M_i} b_{ism}(o_{st})} = P(q_t = i, X_{ist} = m | O, \lambda), \quad (5)$$

где X_{ist} – случайная величина, указывающая компоненту i -й смеси для независимого информативного потока s в момент времени t .

Тогда для гауссовой смеси правила обновления параметров СММ для случая K наблюдаемых последовательностей векторов признаков, где T_k – количество векторов признаков в k -й последовательности, могут быть определены следующим образом:

$$\hat{\pi}_i = \frac{\sum_{k=1}^K \gamma_{is}^k(1)}{K}; \quad (6)$$

$$\hat{c}_{ism} = \frac{\sum_{k=1}^K \sum_{t=1}^{T_k} \gamma_{ism}^k(t)}{\sum_{k=1}^K \sum_{t=1}^{T_k} \gamma_{is}^k(t)}; \quad (7)$$

$$\hat{\mu}_{ism} = \frac{\sum_{k=1}^K \sum_{t=1}^{T_k} \gamma_{ism}^k(t) o_{st}^k}{\sum_{k=1}^K \sum_{t=1}^{T_k} \gamma_{ism}^k(t)}; \quad (8)$$

$$\hat{\Sigma}_{ism} = \frac{\sum_{k=1}^K \sum_{t=1}^{T_k} \gamma_{ism}^k(t) (o_{st}^k - \mu_{ism})(o_{st}^k - \mu_{ism})^T}{\sum_{k=1}^K \sum_{t=1}^{T_k} \gamma_{ism}^k(t)}, \quad (9)$$

а a_{ij} рассчитывается на основании выражений (3) и (4) с учетом (1) как

$$\hat{a}_{ij} = \frac{\sum_{k=1}^K \sum_{t=1}^{T_k} \xi_{ij}^k(t)}{\sum_{k=1}^K \sum_{t=1}^{T_k} \gamma_i^k(t)}. \quad (10)$$

В данном случае для модифицированного алгоритма Баума – Велша, используемого для интерактивного неконтролируемого уточнения параметров СММ с механизмом забывания и обновления, можно записать следующую последовательность шагов:

1. Каждый параметр СММ, требующий уточнения, записать в форме, приведенной в выражениях (6) – (10).
2. С учетом выражения (1) посчитать *forward*- и *backward*-вероятности для всех состояний i и моментов времени t .
3. Для каждого состояния i и момента времени t , используя выражения (3) – (5) и вектор признаков, на котором происходит адаптация, оценить новые параметры СММ.
4. Вычислить апостериорную вероятность $P(O | \lambda^{n+1})$ с новыми параметрами СММ.

5. Прекратить процесс уточнения, если значение $\varphi P(O | \lambda^n) > \psi P(O | \lambda^{n+1})$, где φ и ψ – коэффициенты забывания и обновления соответственно, а $P(O | \lambda^n)$ – апостериорная вероятность, рассчитанная на старых параметрах СММ и вычисленная на втором шаге.

6. Сохранить новые параметры СММ и повторить шаги 1–6 с новыми значениями параметров СММ.

В результате будет получена адаптированная СММ, в которой будут уточнены матрица переходов (A), вероятность начального распределения состояний (π), а также вектор математических ожиданий и ковариационная матрица, соответствующие вероятности наблюдения (B) с учетом коэффициента забывания φ и обновления ψ , влияние которых на эффективность распознавания речи будет исследовано экспериментально.

2. Система распознавания эмоциональной речи с адаптацией СММ

Дополнив систему распознавания эмоциональной речи блоком адаптации СММ, можно добиться улучшения эффективности распознавания эмоциональной речи для последующих речевых данных, так как они уже будут анализироваться на уточненных СММ (рис. 2). Заметим, что для коротких речевых сигналов (до 6–10 слов) эффект адаптации СММ не будет наблюдаться из-за отсутствия достаточного количества данных, необходимых для уточнения параметров СММ.

Стоит отметить, что при быстрой смене эмоционального состояния все же происходят значительные изменения акустических характеристик сигнала. Это может приводить к тому, что все значения апостериорной вероятности распознанных слов во множестве спутывания будут иметь близкие величины. В результате происходит снижение эффективности распознавания эмоциональной речи. Для этого случая следует предусмотреть возможность возврата к исходной модели, полученной на обучающих данных, в которых все эмоции представлены в равной мере. Условием возврата к исходной модели может служить следующее неравенство:

$$\left(\max_{m=1, M_k} (P_k^m) - \frac{\sum_{m=1}^{M_k} P_k^m}{M_k} \right) < thr, \quad (11)$$

где M_k – количество слов в k -м множестве спутывания; P_k^m – апостериорная вероятность m -го распознанного слова в k -м множестве спутывания; thr – это порог, который определяется экспериментальным путем.

Механизм забывания характеризуется коэффициентом $0 < \varphi \leq 1$, при $\varphi = 1$ механизм забывания исходных моделей не задействован. Коэффициент φ нужен для снижения эффекта влияния прошлых наблюдений относительно новых входных данных, чтобы учитывалась изменчивость параметров СММ. Аналогичные механизмы забывания были также предложены в работах [13, 14].

В свою очередь, механизм обновления характеризуется коэффициентом $0 \leq \psi \leq 1$, который задает степень доверия к новым входным данным, чтобы не допускать адаптацию СММ на ложно распознанных словах. Коэффициент ψ было предложено определять исходя из формулы

$$\psi = \begin{cases} 0 & \text{при } P_k < 2BL - 1; \\ \frac{P_k - BL}{1 - BL} + 1 & \text{при } 2BL - 1 \leq P_k < BL; \\ 1 & \text{при } P_k \geq BL, \end{cases} \quad (12)$$

графическое представление которой показано на рис. 1. Здесь BL – среднее значение апостериорной вероятности распознанного слова для корректно распознанных слов. Параметры BL , как и все остальные параметры, которые должны быть исследованы экспериментально, определяются на проверочных данных сразу после обучения системы и при анализе тестовой выборки не меняются.

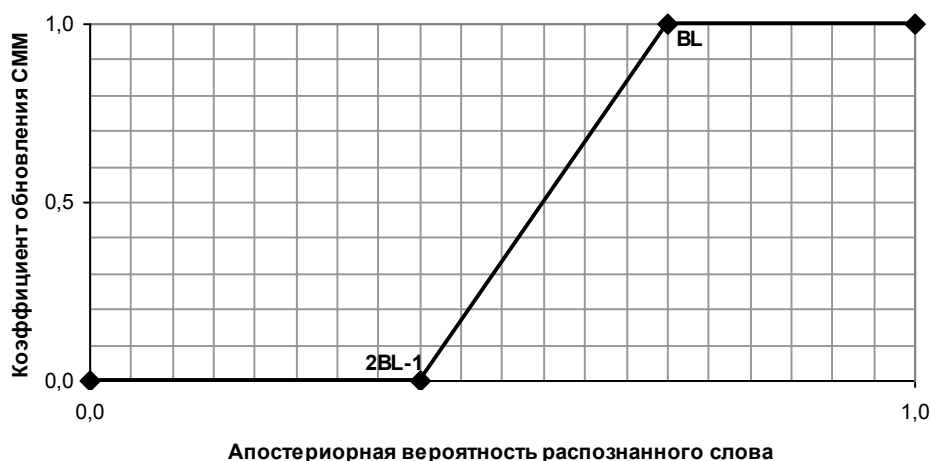


Рис. 1. Зависимость коэффициента обновления СММ от апостериорной вероятности распознанного слова

3. База эмоциональной речи и эффективность распознавания

Распознавание эмоциональной речи и оценка эффективности работы полученной системы будут осуществляться на японском просодическом корпусе MULTEXT (MULTEXT-J) [15]. Выбранная база состоит из 40 текстов (продолжительностью от 10 до 30 с речи), которые записаны при участии шести человек (трех мужчин и трех женщин) в возрасте от 20 до 50 лет. Каждый текст записан в двух вариантах: в нейтральном эмоциональном состоянии и с требуемой эмоцией. Таким образом, размер базы составляет 480 файлов. Запись всех файлов производилась с частотой дискретизации сигнала 16 000 Гц, разрядностью квантования 16 бит, в формате звуковых файлов Waveform Audio File Format (WAV).

Перед обучением и тестированием сигнал преобразовывался в последовательность векторов признаков, представляющих из себя 13 кепстральных коэффициентов, полученных на основе спектральных коэффициентов, рассчитанных по параметрам линейного предсказания. Эти кепстральные коэффициенты распределены по экспоненциально-логарифмической шкале частот, которая снижает изменчивость пространства информативных признаков для эмоциональной речи по сравнению с нейтральной [16]. В состав векторов были включены приближения первой и второй производной каждого коэффициента. Таким образом, общая размерность пространства векторов признаков составила 39.

Обучение и тестирование проводились на основе перекрестной проверки (k-fold cross-validation [17]) с разбиением базы на 10 равных частей. Исходные значения параметров СММ оцениваются на обучающей выборке с учетом ее ручной транскрипции по фонемам.

Для определения величины эффективности распознавания эмоциональной слитной речи была применена формула

$$W_{Acc} = \frac{N - S - D - I}{N},$$

где N – количество слов в распознаваемой речи (правильная транскрипция); S – количество замененных слов в речи при распознавании; D – количество удаленных слов из речи при распознавании; I – количество вставленных слов в речь при распознавании.

Зависимость эффективности распознавания эмоциональной слитной речи от количества распознанных слов для разработанной системы (рис. 2) на базе MULTEXT-J представлена в таблице.

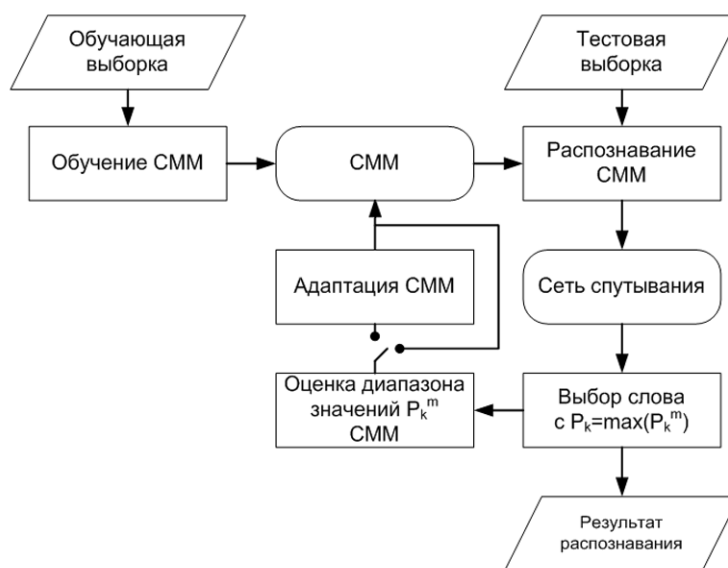


Рис. 2. Система распознавания эмоциональной речи с адаптацией СММ

Эффективность распознавания эмоциональной слитной речи в зависимости от количества распознанных слов

Количество распознанных слов, шт.	Эффективность распознавания речи, %
0	70
5	70
10	71
15	73
20	74
30	75
40	75
50	75

Отметим, что случай для нуля распознанных слов соответствует эффективности распознавания эмоциональной слитной речи при отсутствии адаптации СММ на данных из тестируемой выборки. Для предложенного вектора признака она равняется 70 %.

В ходе экспериментов была определена оптимальная величина коэффициента забывания ϕ , равная 0,75. Величина порога thr в формуле (11) для базы MULTEXT-J оказалась равной 0,1. Значение величины BL для определения коэффициента обновления ψ по формуле (12) составило 0,82.

Заключение

На основе изложенной в статье теории была создана система распознавания эмоциональной речи, дополненная блоком адаптации СММ с механизмом забывания и обновления. Как видно из таблицы, максимальная эффективность распознавания речи, равная 75 %, достигается при анализе 30 слов и более, что соответствует приблизительно 20 с эмоциональной речи.

Таким образом, предложенный дополнительный этап адаптации СММ с механизмом забывания и обновления позволил повысить эффективность распознавания эмоциональной речи на 5 %.

Список литературы

1. Baum, L.E. An inequality and associated maximization techniques in statistical estimation for probabilistic functions of Markov processes / L.E. Baum // Inequalities. – 1972. – № 3. – P. 1–8.
2. A maximization technique occurring in the statistical analysis of probabilistic function of Markov chains / L.E. Baum [et al.] // Ann. Math. Stat. – 1970. – № 41. – P. 164–171.

3. Juang, B.-H. Maximum likelihood estimation for multivariate mixture observations of Markov chains / B.-H. Juang, S.E. Levinson, M.M. Sondhi // *IEEE Trans. Inform. Theory.* – 1993. – № 2. – P. 307–309.
4. Liporace, L.R. Maximum likelihood estimation for multivariate observations of Markov sources / L.R. Liporace // *IEEE Trans. Inform. Theory.* – 1995. – № 28. – P. 729–734.
5. Gauvain, J.-L. Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains / J.-L. Gauvain, C.-H. Lee // *IEEE Trans. Speech Audio Processing.* – 1994. – № 2. – P. 291–298.
6. Huo, Q. Bayesian adaptive learning of the parameters of hidden Markov model for speech recognition / Q. Huo, C. Chan, C.-H. Lee // *IEEE Trans. Speech Audio Processing.* – 1992. – № 5. – P. 334–345.
7. Lee, C.-H. A study on speaker adaptation of the parameters of continuous density hidden Markov models / C.-H. Lee, C.-H. Lin, B.-H. Juang // *IEEE Trans. Signal Processing.* – 1991. – № 39. – P. 806–814.
8. Matsuoka, T. A study of on-line Bayesian adaptation for HMM-based speech recognition / T. Matsuoka, C.-H. Lee // *Proc. EUROSPEECH-93.* – Berlin, Germany, 1993. – P. 815–818.
9. Huo, Q. On-Line Adaptive Learning of the Continuous Density Hidden Markov Model Based on Approximate Recursive Bayes Estimate / Q. Huo, C.-H. Lee // *Speech and Audio Processing.* – 1997. – № 5. – P. 161–172.
10. Рылов, А.С. Анализ речи в распознающих системах / А.С. Рылов. – Минск : Бест-принт, 2003. – 264 с.
11. Bilmes, J. A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models / J. Bilmes // *International Computer Science Institute.* – 1998. – № 1. – P. 164–191.
12. The HTK Book (for HTK v. 3.4) / S. Young [et al.]. – Cambridge University Engineering Department, 2006. – 359 p.
13. Krishnamurthy, V. On-line estimation of hidden Markov model parameters based on the Kullback-Leibler information measure / V. Krishnamurthy, J.B. Moore // *IEEE Trans. Signal Processing.* – 1993. – № 41 (8). – P. 2557–2573.
14. Weinstein, E. Sequential algorithms for parameter estimation based on the Kullback-Leibler information measure / E. Weinstein, M. Feder, A.V. Oppenheim // *IEEE Trans. Acoust, Speech, Signal Processing.* – 1990. – № 38 (9). – P. 1652–1654.
15. MULTEXT-J. Japanese MULTEXT Prosodic Corpus [Electronic resource]. – Mode of access : <http://research.nii.ac.jp/src/en/MULTEXT-J.html>. – Date of access : 30.09.2013.
16. Bou-Ghazale, S.E. A Comparative Study of Traditional and Newly Proposed Features for Recognition of Speech Under Stress / S.E. Bou-Ghazale, J.H.L. Hansen // *Speech and Audio Processing.* – 2000. – № 8. – P. 429–442.
17. K-fold cross-validation. Wikipedia [Electronic resource]. – Mode of access : http://en.wikipedia.org/wiki/Cross-validation_%28statistics%29. – Date of access : 18.08.2014.

Поступила 16.06.2014

*Белорусский государственный университет,
Минск, ул. Курчатова, 5
e-mail: tkachenia@gmail.com*

A.V. Tkachenia

ADAPTIVE LEARNING OF HIDDEN MARKOV MODELS FOR EMOTIONAL SPEECH

An on-line unsupervised algorithm for estimating the hidden Markov models (HMM) parameters is presented. The problem of hidden Markov models adaptation to emotional speech is solved. To increase the reliability of estimated HMM parameters, a mechanism of forgetting and updating is proposed. A functional block diagram of the hidden Markov models adaptation algorithm is also provided with obtained results, which improve the efficiency of emotional speech recognition.

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

УДК 517.958:537.8

В.Т. Ерофеев, В.Ф. Бондаренко

**ЭКРАНИРОВАНИЕ ЭЛЕКТРОМАГНИТНЫХ ПОЛЕЙ ЭКРАНАМИ
ИЗ МАТРИЧНЫХ КОМПОЗИТОВ, СОДЕРЖАЩИХ
БИИЗОТРОПНЫЕ ЧАСТИЦЫ**

Разрабатывается методика решения краевой задачи экранирования электромагнитных полей сосредоточенных источников электрических и магнитных диполей однородными магнитодиэлектрическими экранами, содержащими разреженную систему сферических частиц с комплексными диэлектрическими и магнитными проницаемостями и комплексными параметрами биизотропности. Методика основана на алгоритме вычисления эффективных материальных параметров усредненной биизотропной среды, эквивалентной матричному композиту экрана. Вычисляются компоненты поля за экраном для композита с частицами из метаматериала и киральных частиц.

Введение

Экранирование электромагнитных полей является одной из основных проблем электромагнитной совместимости технических средств радиотехники и электроники. Защита высокочувствительных приборов от воздействия внешних электромагнитных излучений может быть решена с помощью создания эффективных экранирующих структур.

В настоящее время большое внимание в научной литературе уделяется исследованиям электродинамических свойств композитных материалов различных типов, классификация которых проводится в работах [1–7]. Композиты представляют собой структурно-неоднородный материал с большим числом структурных включений, отличающихся геометрией, химическим составом и уровнем линейных размеров. К ним относятся метаматериалы, киральные среды, квадрупольные материалы, гранулированные структуры, наноструктурные композиты и др. Один из подходов моделирования электродинамических свойств таких композитов сводится к замене сложных структурно-неоднородных материалов однородными средами, электродинамические свойства которых близки с некоторой точностью к свойствам исходного материала. Данный подход позволяет разработать модель, которая значительно упрощает решение математических задач электродинамики, основанных на более простой модели. Второй подход моделирования, к примеру, основан на методе автономных блоков. В статьях [8–11] разработаны различные методики моделирования композитов, содержащих идеально проводящие спирали и сферические частицы с винтовой проводимостью поверхности, на основе вычисления эффективных параметров композита с помощью метода Максвелла – Гарнетта [12]. Исследуются композитные структуры, содержащие включения других видов: проволоку, разрезанные кольца, Ω -включения, рулеты, сегнетоэлектрические кубики и др. [2]. В данной работе разрабатываются методы моделирования композитов с биизотропными однородными сферическими частицами, расположенными в магнитодиэлектрической матрице, с учетом однократного рассеяния полей на частицах [13].

Из композитных материалов изготавливаются технические устройства: однослойные и многослойные экраны, композитные покрытия, нанесенные на поверхностях элементов техники. Исследование экранирующих свойств таких экранов проводится аналитическими и численными методами [14–18].

1. Постановка задачи

В пространстве R^3 с электрической и магнитной постоянными ϵ_0, μ_0 размещен экран $D(0 < z < \Delta)$ толщиной Δ , ограниченный плоскостями раздела сред $\Gamma_1(z=0), \Gamma_2(z=\Delta)$. Экран

заполнен магнитоэлектрической однородной средой, называемой матрицей, которая характеризуется диэлектрической и магнитной проницаемостями $\varepsilon_m = \varepsilon_m \varepsilon_0$, $\mu_m = \mu_m \mu_0$. В матрице случайным образом распределено большое число сферических частиц радиуса R , выполненных из биизотропного материала с материальными параметрами $\varepsilon = \varepsilon_r \varepsilon_0$, $\mu = \mu_r \mu_0$, $G = G_r / c$, $Z = Z_r / c$. Композит, состоящий из матрицы и частиц другого материала, будем называть матричным композитом [5]. В полупространстве $D_1 (z < 0)$ в точке $O_1(0, 0, -h)$, $h > 0$, расположен точечный источник, состоящий из суммы электрического и магнитного диполей, ориентированных вдоль оси Oz . Поле источника будем определять следующими соотношениями, приведенными в [19]:

$$\begin{aligned} \vec{E}_D &= A\vec{m}_{01}(r_1, \theta_1, \varphi; k_0) + B\vec{n}_{01}(r_1, \theta_1, \varphi; k_0), \\ \vec{H}_D &= h_0(A\vec{n}_{01}(r_1, \theta_1, \varphi; k_0) + B\vec{m}_{01}(r_1, \theta_1, \varphi; k_0)), \end{aligned} \quad (1)$$

где (r_1, θ_1, φ) – сферические координаты с началом в точке O_1 ; $k_0 = \frac{\omega}{c}$, $\omega = 2\pi f$, – круговая частота; c – скорость света в вакууме; $h_0 = \frac{k_0}{i\omega\mu_0}$; A, B – постоянные.

При $A = 0, B \neq 0$ имеем электрический диполь, при $A \neq 0, B = 0$ – магнитный диполь.

В результате взаимодействия первичного поля (1) с экраном D образуются результирующие электромагнитные поля: в области D_1 – отраженное поле \vec{E}'_1, \vec{H}'_1 и суммарное поле $\vec{E}_1 = \vec{E}_D + \vec{E}'_1$, $\vec{H}_1 = \vec{H}_D + \vec{H}'_1$; в слое D – поле \vec{E}, \vec{H} ; в полупространстве $D_2 (z > \Delta)$ образуется поле \vec{E}_2, \vec{H}_2 , прошедшее через экран D . Экран представляет собой неоднородную среду, поэтому для упрощения модели неоднородную среду экрана заменим на однородную усредненную биизотропную среду с материальными параметрами $\varepsilon_{\text{эф}} = \varepsilon_3 \varepsilon_0$, $\mu_{\text{эф}} = \mu_3 \mu_0$, $G_{\text{эф}} = G_3 / c$, $Z_{\text{эф}} = Z_3 / c$. Для экрана с эффективными параметрами сформулируем краевую задачу экранирования поля (1).

Краевая задача 1. При заданном поле (1) требуется определить поля \vec{E}'_1, \vec{H}'_1 ; \vec{E}_2, \vec{H}_2 ; \vec{E}, \vec{H} , для которых выполнены:

– уравнения поля

$$\begin{aligned} \text{rot } \vec{E} &= i\omega(\mu_{\text{эф}} \vec{H} + Z_{\text{эф}} \vec{E}), \quad \text{rot } \vec{H} = -i\omega(\varepsilon_{\text{эф}} \vec{E} + G_{\text{эф}} \vec{H}) \quad \text{в } D, \\ \text{rot } \vec{E}'_1 &= i\omega\mu_0 \vec{H}'_1, \quad \text{rot } \vec{H}'_1 = -i\omega\varepsilon_0 \vec{E}'_1 \quad \text{в } D_1, \\ \text{rot } \vec{E}_2 &= i\omega\mu_0 \vec{H}_2, \quad \text{rot } \vec{H}_2 = -i\omega\varepsilon_0 \vec{E}_2 \quad \text{в } D_2; \end{aligned} \quad (2)$$

– граничные условия на плоскостях раздела сред Γ_1 и Γ_2

$$\left. (\vec{E}_{1r} - \vec{E}_r) \right|_{\Gamma_1} = 0, \quad \left. (\vec{H}_{1r} - \vec{H}_r) \right|_{\Gamma_1} = 0, \quad \left. (\vec{E}_{2r} - \vec{E}_r) \right|_{\Gamma_2} = 0, \quad \left. (\vec{H}_{2r} - \vec{H}_r) \right|_{\Gamma_2} = 0$$

и условия излучения на бесконечности.

Проблема состоит в вычислении эффективных параметров $\varepsilon_{\text{эф}}, \mu_{\text{эф}}, G_{\text{эф}}, Z_{\text{эф}}$ для матричного композита. Алгоритм вычисления эффективных параметров для сферических биизотропных частиц, расположенных в вакууме, разработан в работе [20].

2. Композит из биизотропных частиц в среде матрицы

В матрице с параметрами ε_m, μ_m , охватывающей пространство R^3 , размещены биизотропные сферические частицы радиуса R , характеризующиеся материальными параметрами

ε, μ, G, Z . Выделим частицу $D_R (0 \leq r < R)$ с центром O в начале декартовых координат $Oxyz$ ($Or\theta\varphi$ – сферические координаты). Обозначим области: $D_P (0 \leq r < P)$ – шаровая область достаточно большого радиуса P ; $D_{RP} (R < r < P)$ – шаровой слой вокруг частицы D_R ; D_0 – область между частицами, $D_0 \subset D_P$; D_S – область внутри s -й частицы, $D_P = \bigcup_{S=1}^N D_S \cup D_0$, $D_1 = D_R$, $D_{SP} = D_P / D_S$. Будем предполагать, что

$$R \ll P \ll \lambda, \quad (3)$$

где λ – длина электромагнитной волны \vec{E}_0, \vec{H}_0 , распространяющейся в вакууме.

Полагаем $f \leq 10$ ГГц, $R \leq 4$ мкм, $P \approx 30R$, $|\varepsilon_m \mu_m| < 10$.

Введем обозначения: N – число частиц в области D_P ; $V_R = \frac{4\pi}{3} R^3$ – объем частицы; V_P – объем области D_P ; $\nu = \frac{N}{V_P}$ – концентрация частиц (число частиц в единице объема матрицы); $\tau = \nu V_R$ – объемный коэффициент заполнения матрицы сферическими частицами (суммарный объем сферических частиц в единице объема матрицы).

Построим математическую модель композита из биизотропных сферических частиц, используя теорию биизотропных сред. Для построения модели композит подвергается воздействию электромагнитного поля \vec{E}_0, \vec{H}_0 и исследуется отраженное поле. Сформулируем вспомогательную краевую задачу дифракции поля \vec{E}_0, \vec{H}_0 на изолированной частице D_R , расположенной в среде матрицы [16, 22]. Обозначим: $D'_R (R < r < \infty)$ – внешняя к шару D_R область; $\Gamma (r = R)$ – сферическая поверхность шара. В D'_R распространяется первичное электромагнитное поле \vec{E}_0, \vec{H}_0 , колеблющееся с круговой частотой ω и временной зависимостью $\exp(-i\omega t)$. Обозначим поля: \vec{E}_R, \vec{H}_R – поле внутри шара D_R ; \vec{E}'_R, \vec{H}'_R – отраженное поле в D'_R ; $\vec{E}_2 = \vec{E}_0 + \vec{E}'_R$, $\vec{H}_2 = \vec{H}_0 + \vec{H}'_R$ – суммарное поле в области D'_R .

Краевая задача 2. Для заданного первичного поля \vec{E}_0, \vec{H}_0 требуется определить электромагнитные поля $\vec{E}_R, \vec{H}_R \in C^1(D_R) \cap C(\bar{D}_R)$, $\vec{E}'_R, \vec{H}'_R \in C^1(D'_R) \cap C(\bar{D}'_R)$, которые удовлетворяют:

– уравнениям Максвелла

$$\text{rot } \vec{E}_R = i\omega(\mu \vec{H}_R + Z \vec{E}_R), \quad \text{rot } \vec{H}_R = -i\omega(\varepsilon \vec{E}_R + G \vec{H}_R), \quad 0 \leq r < R, \quad (4)$$

$$\text{rot } \vec{E}'_R = i\omega \mu_m \vec{H}'_R, \quad \text{rot } \vec{H}'_R = -i\omega \varepsilon_m \vec{E}'_R, \quad r > R; \quad (5)$$

– граничным условиям

$$\vec{E}_{R\tau} \Big|_{r=R} = \vec{E}_{2\tau} \Big|_{r=R}, \quad \vec{H}_{R\tau} \Big|_{r=R} = \vec{H}_{2\tau} \Big|_{r=R} \quad (6)$$

и условиям излучения на бесконечности.

3. Решение задачи дифракции волн на биизотропной сферической частице, расположенной в магнитоэлектрической среде

В качестве первичного поля, действующего на частицы, выберем любое монохроматическое электромагнитное поле, которое представим в виде рядов по базисным сферическим полям [16]:

$$\vec{E}_0 = \sum_{n,m} [a_{mn} \vec{m}_{mn}(\vec{r}; k_m) + b_{mn} \vec{n}_{mn}(\vec{r}; k_m)],$$

$$\vec{H}_0 = h_M \sum_{n,m} [a_{mn} \vec{n}_{mn}(\vec{r}; k_M) + b_{mn} \vec{m}_{mn}(\vec{r}; k_M)], \quad (7)$$

где a_{mn}, b_{mn} – заданные коэффициенты разложений, $\sum_{n,m} = \sum_{n=1}^{\infty} \sum_{m=-n}^n$, $h_M = \frac{k_M}{i\omega\mu_M}$,

$$\vec{n}_{mn} = f_n(k_M r) Y_n^m(\theta, \varphi) \vec{e}_r + g_n(k_M r) \vec{\Pi}_{mn}(\theta, \varphi), \quad \vec{m}_{mn} = j_n(k_M r) \vec{T}_{mn}(\theta, \varphi), \quad Y_n^m = P_n^m(\cos\theta) e^{im\varphi}, \quad f_n(k_M r) = \frac{n(n+1)}{k_M r} j_n(k_M r), \quad g_n(k_M r) = \frac{1}{k_M r} \frac{d}{dr} (r j_n(k_M r)), \quad \vec{\Pi}_{mn} = \left(\frac{\partial}{\partial \theta} P_n^m(\cos\theta) \vec{e}_\theta + \frac{im}{\sin\theta} P_n^m(\cos\theta) \vec{e}_\varphi \right) e^{im\varphi},$$

$$\vec{T}_{mn} = \left(\frac{im}{\sin\theta} P_n^m(\cos\theta) \vec{e}_\theta - \frac{\partial}{\partial \theta} P_n^m(\cos\theta) \vec{e}_\varphi \right) e^{im\varphi}, \quad n=1, 2, \dots, m=0, \pm 1, \pm 2, \dots, \pm n.$$

Вычислим поле (7) в начале координат:

$$\begin{aligned} \vec{E}_0(0) = \vec{E}_0^0 = E_x^0 \vec{e}_x + E_y^0 \vec{e}_y + E_z^0 \vec{e}_z = -\frac{2}{3} b_{11} \vec{e}_+ - \frac{1}{3} b_{-11} \vec{e}_- + \frac{2}{3} b_{01} \vec{e}_z, \\ \vec{H}_0(0) = \vec{H}_0^0 = H_x^0 \vec{e}_x + H_y^0 \vec{e}_y + H_z^0 \vec{e}_z = -\frac{2h_M}{3} a_{11} \vec{e}_+ - \frac{h_M}{3} a_{-11} \vec{e}_- + \frac{2h_M}{3} a_{01} \vec{e}_z, \end{aligned} \quad (8)$$

где

$$\vec{e}_+ = i\vec{e}_y + \vec{e}_x, \quad \vec{e}_- = i\vec{e}_y - \vec{e}_x, \quad a_{01} = \frac{3}{2h_M} H_z^0, \quad a_{11} = \frac{3}{4h_M} (iH_y^0 - H_x^0), \quad a_{-11} = \frac{3}{2h_M} (iH_y^0 + H_x^0),$$

$$b_{01} = \frac{3}{2} E_z^0, \quad b_{11} = \frac{3}{4} (iE_y^0 - E_x^0), \quad b_{-11} = \frac{3}{2} (iE_y^0 + E_x^0).$$

Отраженное электромагнитное поле выразим через базисные сферические поля, которые удовлетворяют условиям излучения и уравнениям (5):

$$\begin{aligned} \vec{E}'_R = \sum_{n,m} [x_{mn}^{(2)} \vec{\tilde{m}}_{mn}(\vec{r}, k_M) + y_{mn}^{(2)} \vec{\tilde{n}}_{mn}(\vec{r}, k_M)], \quad r > R, \\ \vec{H}'_R = h_M \sum_{n,m} [x_{mn}^{(2)} \vec{\tilde{n}}_{mn}(\vec{r}, k_M) + y_{mn}^{(2)} \vec{\tilde{m}}_{mn}(\vec{r}, k_M)], \end{aligned} \quad (9)$$

где

$$\vec{\tilde{n}}_{mn} = f_n^{(1)}(k_M r) Y_n^m(\theta, \varphi) \vec{e}_r + g_n^{(1)}(k_M r) \vec{\Pi}_{mn}(\theta, \varphi), \quad \vec{\tilde{m}}_{mn} = h_n^{(1)}(k_M r) \vec{T}_{mn}(\theta, \varphi),$$

$$f_n^{(1)}(k_M r) = \frac{n(n+1)}{k_M r} h_n^{(1)}(k_M r), \quad g_n^{(1)}(k_M r) = \frac{1}{k_M r} \frac{\partial}{\partial r} (r h_n^{(1)}(k_M r)).$$

Здесь $j_n(\cdot)$, $h_n^{(1)}(\cdot)$ – сферические функции Бесселя, $k_M = \omega\sqrt{\epsilon_M \mu_M}$, $0 \leq \arg k_M < \pi$.

Электромагнитное поле внутри шара D_R представим через базисные сферические поля в композитных средах, которые удовлетворяют уравнениям (4):

$$\begin{aligned} \vec{E}_R = \sum_{n,m} [x_{mn}^{(1)} \vec{K}_{mn}^{(1)}(\vec{r}, k_1) + y_{mn}^{(1)} \vec{K}_{mn}^{(2)}(\vec{r}, k_2)], \quad 0 \leq r < R, \\ \vec{H}_R = \sum_{n,m} [x_{mn}^{(1)} p_1 \vec{K}_{mn}^{(1)}(\vec{r}, k_1) + y_{mn}^{(1)} p_2 \vec{K}_{mn}^{(2)}(\vec{r}, k_2)], \end{aligned} \quad (10)$$

где коэффициенты $x_{mn}^{(1)}, y_{mn}^{(1)}, x_{mn}^{(2)}, y_{mn}^{(2)}$ определены в работе [16, с. 275];
 $\vec{K}_{mn}^{(j)} = \vec{n}_{mn}(\vec{r}, k_j) - q_j \vec{m}_{mn}(\vec{r}, k_j)$, $k_j = \sqrt{g + \frac{1}{2}a^2 + af_j}$, $0 \leq \arg k_j < \pi$, $q_j = \frac{g}{k_j g_j}$, $p_j = \frac{1}{\mu} \left(\frac{ig}{\omega g_j} - Z \right)$,
 $f_j = (-1)^j f_0$, $f_0 = \sqrt{\omega^2 \varepsilon \mu - b^2}$, $0 \leq \arg f_0 < \pi$, $g = \omega^2 (\varepsilon \mu - ZG)$, $g_j = f_j - \frac{1}{2}a$, $a = i\omega(G - Z)$,
 $b = \frac{\omega}{2}(G + Z)$, $j = 1, 2$.

Удовлетворяя граничным условиям (6), получим систему алгебраических уравнений относительно $x_{mn}^{(1)}, y_{mn}^{(1)}, x_{mn}^{(2)}, y_{mn}^{(2)}$:

$$x_{mn}^{(2)} h_n^{(1)}(\xi_M) + a_{mn} j_n(\xi_M) = -x_{mn}^{(1)} q_1 j_n(\xi_1) - y_{mn}^{(1)} q_2 j_n(\xi_2); \quad (11)$$

$$x_{mn}^{(2)} g_n^{(1)}(\xi_M) + a_{mn} g_n(\xi_M) = x_{mn}^{(1)} \bar{p}_1 g_n(\xi_1) + y_{mn}^{(1)} \bar{p}_2 g_n(\xi_2); \quad (12)$$

$$y_{mn}^{(2)} h_n^{(1)}(\xi_M) + b_{mn} j_n(\xi_M) = -x_{mn}^{(1)} \bar{p}_1 q_1 j_n(\xi_1) - y_{mn}^{(1)} \bar{p}_2 q_2 j_n(\xi_2); \quad (13)$$

$$y_{mn}^{(2)} g_n^{(1)}(\xi_M) + b_{mn} g_n(\xi_M) = x_{mn}^{(1)} g_n(\xi_1) + y_{mn}^{(1)} g_n(\xi_2), \quad (14)$$

где $\bar{p}_j = p_j / h_m$, $\xi_j = k_j R$, $\xi_M = k_M R$.

Из уравнения (11) с помощью (12) исключим $x_{mn}^{(2)}$, а из уравнения (13) с помощью (14) исключим $y_{mn}^{(2)}$. Получим систему

$$x_{mn}^{(1)} \left(\bar{p}_1 g_n(\xi_1) h_n^{(1)}(\xi_M) + q_1 j_n(\xi_1) g_n^{(1)}(\xi_M) \right) + y_{mn}^{(1)} \left(\bar{p}_2 g_n(\xi_2) h_n^{(1)}(\xi_M) + q_2 j_n(\xi_2) g_n^{(1)}(\xi_M) \right) = \frac{a_{mn}}{i \xi_M^2}, \quad (15)$$

$$x_{mn}^{(1)} \left(g_n(\xi_1) h_n^{(1)}(\xi_M) + \bar{p}_1 q_1 j_n(\xi_1) g_n^{(1)}(\xi_M) \right) + y_{mn}^{(1)} \left(g_n(\xi_2) h_n^{(1)}(\xi_M) + \bar{p}_2 q_2 j_n(\xi_2) g_n^{(1)}(\xi_M) \right) = \frac{b_{mn}}{i \xi_M^2}.$$

Запишем системы уравнений (11), (13), (15) в матричном виде, полагая $n = 1$:

$$\vec{X}_m^{(2)} = - \left(j_1(\xi_M) \hat{E} \vec{a}_m + \hat{Q} \vec{X}_m^{(1)} \right) / h_1^{(1)}(\xi_M), \quad (16)$$

$$i \xi_M^2 \hat{P} \vec{X}_m^{(1)} = \vec{a}_m, \quad (17)$$

где

$$\vec{X}_m^{(j)} = \begin{pmatrix} x_{m1}^{(j)} \\ y_{m1}^{(j)} \end{pmatrix}, \quad \vec{a}_m = \begin{pmatrix} a_{m1} \\ b_{m1} \end{pmatrix}, \quad \hat{E} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \hat{Q} = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix}, \quad \hat{P} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix},$$

$$q_{11} = q_1 j_1(\xi_1), \quad q_{12} = q_2 j_1(\xi_2), \quad q_{21} = \bar{p}_1 q_1 j_1(\xi_1), \quad q_{22} = \bar{p}_2 q_2 j_1(\xi_2);$$

$$p_{11} = \bar{p}_1 g_1(\xi_1) h_1^{(1)}(\xi_M) + q_1 j_1(\xi_1) g_1^{(1)}(\xi_M), \quad p_{12} = \bar{p}_2 g_1(\xi_2) h_1^{(1)}(\xi_M) + q_2 j_1(\xi_2) g_1^{(1)}(\xi_M), \quad (18)$$

$$p_{21} = g_1(\xi_1) h_1^{(1)}(\xi_M) + \bar{p}_1 q_1 j_1(\xi_1) g_1^{(1)}(\xi_M), \quad p_{22} = g_1(\xi_2) h_1^{(1)}(\xi_M) + \bar{p}_2 q_2 j_1(\xi_2) g_1^{(1)}(\xi_M).$$

Подставив (17) в (16), получим соотношения

$$\vec{X}_m^{(2)} = -\hat{K} \vec{X}_m^{(1)}, \quad \hat{K} = \left(i \xi_M^2 j_1(\xi_M) \hat{P} + \hat{Q} \right) / h_1^{(1)}(\xi_M), \quad m = 0, \pm 1. \quad (19)$$

4. Вычисление специальных интегралов

Вычислим вспомогательные интегралы вида

$$\begin{aligned}\vec{J}_{mn}^Y &= \int_0^{2\pi} \int_0^\pi Y_n^m(\theta, \varphi) \vec{e}_r \sin\theta d\theta d\varphi, \quad \vec{J}_{mn}^\Pi = \int_0^{2\pi} \int_0^\pi \vec{\Pi}_{mn}(\theta, \varphi) \sin\theta d\theta d\varphi, \\ \vec{J}_{mn}^T &= \int_0^{2\pi} \int_0^\pi \vec{T}_{mn}(\theta, \varphi) \sin\theta d\theta d\varphi, \quad n=1, 2, \dots; \quad m=0, \pm 1, \pm 2, \dots \pm n.\end{aligned}$$

Лемма 1. Имеют место формулы

$$\begin{aligned}\vec{J}_{mn}^T &= 0 \quad \text{при } n=1, 2, \dots; \quad m=0, \pm 1, \pm 2, \dots \pm n; \\ \vec{J}_{mn}^{Y, \Pi} &= 0 \quad \text{при } n \geq 2; \quad m=0, \pm 1, \pm 2, \dots \pm n; \\ \vec{J}_{01}^Y &= \frac{4\pi}{3} \vec{e}_z, \quad \vec{J}_{01}^\Pi = \frac{8\pi}{3} \vec{e}_z, \quad \vec{J}_{11}^Y = -\frac{4\pi}{3} \vec{e}_+, \quad \vec{J}_{11}^\Pi = -\frac{8\pi}{3} \vec{e}_+, \quad \vec{J}_{-11}^Y = -\frac{2\pi}{3} \vec{e}_-, \quad \vec{J}_{-11}^\Pi = -\frac{4\pi}{3} \vec{e}_-.\end{aligned}\tag{20}$$

Лемма 2. Объемные интегралы

$$\begin{aligned}\tilde{I}_{mn}^\Pi &= \int_{D_{RP}} \tilde{n}_{mn}(\vec{r}, k_M) dV = \int_R^P \left(f_n^{(1)}(k_M r) \vec{J}_{mn}^Y + g_n^{(1)}(k_M r) \vec{J}_{mn}^\Pi \right) r^2 dr, \\ \tilde{I}_{mn}^T &= \int_{D_{RP}} \tilde{m}_{mn}(\vec{r}, k_M) dV = \int_R^P h_n^{(1)}(k_M r) \vec{J}_{mn}^T r^2 dr\end{aligned}$$

вычисляются с помощью формул

$$\begin{aligned}\tilde{I}_{mn}^T &= 0 \quad \text{при } n=1, 2, \dots; \quad m=0, \pm 1, \pm 2, \dots \pm n; \\ \tilde{I}_{mn}^\Pi &= 0 \quad \text{при } n \geq 2; \quad m=0, \pm 1, \dots \pm n; \\ \tilde{I}_{01}^\Pi &= -2f_M \vec{e}_z, \quad \tilde{I}_{11}^\Pi = 2f_M \vec{e}_+, \quad \tilde{I}_{-11}^\Pi = f_M \vec{e}_-, \end{aligned}\tag{21}$$

где в асимптотическом приближении

$$f_M = \frac{V_R}{\xi_M} h_1^{(1)}(\xi_M) + \frac{4\pi i}{3k_M^3}.$$

Доказательство. Для доказательства используются формулы (20). Для примера вычислим интеграл

$$\tilde{I}_{01}^\Pi = \frac{4\pi}{3} \int_R^P \left(f_1^{(1)}(k_M r) + 2g_1^{(1)}(k_M r) \right) r^2 dr \vec{e}_z = -2f_M \vec{e}_z,$$

где

$$f_M = -\frac{4\pi}{3k_M} \int_R^P \left(h_1^{(1)}(k_M r) + \frac{\partial}{\partial r} \left(r h_1^{(1)}(k_M r) \right) \right) r dr = \frac{4\pi}{3k_M} \left(R^2 h_1^{(1)}(k_M R) - P^2 h_1^{(1)}(k_M P) \right).$$

Преобразуем величину f_M , учитывая допущение (3). Для сферической функции Ханкеля имеем асимптотическую формулу при $k_M P \ll 1$:

$$h_1^{(1)}(k_M P) = -\left(\frac{1}{k_M P} + \frac{i}{(k_M P)^2} \right) e^{ik_M P} \approx -\frac{i}{(k_M P)^2}.$$

В результате $f_m \approx \frac{V_R}{\xi} h_1^{(1)}(\xi_M) + \frac{4\pi i}{3k_M^3}$.

Лемма 3. Объемные интегралы

$$\begin{aligned}\bar{I}_{mn}^{\Pi}(k_j) &= \int_{D_R} \bar{n}_{mn}(\vec{r}, k_j) dV = \int_0^R (f_n(k_j r) \bar{J}_{mn}^Y + g_n(k_j r) \bar{J}_{mn}^{\Pi}) r^2 dr, \\ \bar{I}_{mn}^T(k_j) &= \int_{D_R} \bar{m}_{mn}(\vec{r}, k_j) dV = \int_0^R j_n(k_j r) \bar{J}_{mn}^T r^2 dr\end{aligned}$$

вычисляются с помощью формул

$$\bar{I}_{mn}^T(k_j) = 0 \text{ при } n = 1, 2, \dots; \quad m = 0, \pm 1, \dots, \pm n; \quad (22)$$

$$\bar{I}_{mn}^{\Pi}(k_j) = 0 \text{ при } n \geq 2; \quad m = 0, \pm 1, \dots, \pm n;$$

$$\bar{I}_{01}^{\Pi}(k_j) = 2F_j \vec{e}_z, \quad \bar{I}_{11}^{\Pi}(k_j) = -2F_j \vec{e}_+, \quad \bar{I}_{-11}^{\Pi}(k_j) = -F_j \vec{e}_-,$$

где $F_j = \frac{V_R}{\xi} j_1(\xi_j)$, $j = 1, 2$.

Лемма 4. Имеет место приближенная формула

$$I = \frac{1}{V_P} \sum_{s=1}^N \Phi(\vec{r}_s) \approx \nu G_P, \quad (23)$$

где $G_P = \frac{3}{k_M P} j_1(k_M P)$, $\Phi(\vec{r}) = \exp(ik_M(\vec{p}, \vec{r}))$, $\vec{r}_s = (x_s, y_s, z_s)$ – координаты центра частицы D_s , $\vec{p} = (\cos \varphi_0 \sin \theta_0, \sin \varphi_0 \sin \theta_0, \cos \theta_0)$.

Доказательство. Сумму значений функции будем рассматривать как интегральную сумму. Число точек велико, и они равномерно заполняют всю область D_P . Поэтому

$$I \approx \frac{V}{V_P} \iiint_{D_P} \Phi(\vec{r}) dV = \frac{V}{V_P} \int_0^P \int_0^{\pi} \int_0^{2\pi} \Phi(\vec{r}) r^2 \sin \theta d\theta d\varphi dr.$$

Для вычисления интеграла воспользуемся теоремой сложения [21; с.177; 513.1.3]

$$\Phi(\vec{r}) = \sum_{m=-\infty}^{\infty} \sum_{n=|m|}^{\infty} A_{mn}^{(-)} j_n(k_M r) P_n^m(\cos \theta) e^{im\varphi}.$$

Тогда

$$\begin{aligned}I &= \frac{V}{V_P} \sum_{m=-\infty}^{\infty} \sum_{n=|m|}^{\infty} A_{mn}^{(-)} \int_0^P j_n(k_M r) r^2 dr \int_0^{\pi} P_n^m(\cos \theta) \sin \theta d\theta \int_0^{2\pi} e^{im\varphi} d\varphi = \\ &= 2\pi \frac{V}{V_P} \sum_{n=0}^{\infty} A_{0n}^{(-)} \int_0^P j_n(k_M r) r^2 dr \int_0^{\pi} P_n(\cos \theta) \sin \theta d\theta = \frac{4\pi V}{V_P} \int_0^P j_0(k_M r) r^2 dr = \frac{4\pi V}{k_M V_P} P^2 j_1(k_M P) = \nu G_P.\end{aligned}$$

Здесь использовалось условие ортогональности полиномов Лежандра

$$\int_0^{\pi} P_0(\cos \theta) P_n(\cos \theta) \sin \theta d\theta = 2\delta_{0n}.$$

5. Усредненные электрические и магнитные поля в композите

В результате взаимодействия поля \vec{E}_0, \vec{H}_0 с частицами D_s в матрице образуются отраженные поля \vec{E}'_s, \vec{H}'_s и поля \vec{E}_s, \vec{H}_s , проникшие внутрь частиц D_s . Вычислим суммарные поля в области D_p :

$$\vec{E}_{\text{сум}} = \begin{cases} \vec{E}_0 + \sum_{s=1}^N \vec{E}'_s & \text{в } D_0; \\ \vec{E}_s & \text{в } D_s, \end{cases} \quad \vec{H}_{\text{сум}} = \begin{cases} \vec{H}_0 + \sum_{s=1}^N \vec{H}'_s & \text{в } D_0; \\ \vec{H}_s & \text{в } D_s, \end{cases} \quad (24)$$

где N – число частиц в области D_p .

Усредним поля (24) по объему D_p :

$$\begin{aligned} \vec{E} &= \frac{1}{V_P} \int_{D_p} \vec{E}_{\text{сум}} dV = \frac{1}{V_P} \left[\int_{D_p} \vec{E}_0 dV + \sum_{s=1}^N \left(\int_{D_0} \vec{E}'_s dV + \int_{D_s} \vec{E}_s dV - \int_{D_s} \vec{E}_0 dV \right) \right], \\ \vec{H} &= \frac{1}{V_P} \int_{D_p} \vec{H}_{\text{сум}} dV = \frac{1}{V_P} \left[\int_{D_p} \vec{H}_0 dV + \sum_{s=1}^N \left(\int_{D_0} \vec{H}'_s dV + \int_{D_s} \vec{H}_s dV - \int_{D_s} \vec{H}_0 dV \right) \right]. \end{aligned} \quad (25)$$

Как правило, на практике композит подвергается воздействию плоских электромагнитных волн. Поэтому первичное поле (7) будем моделировать плоской волной, характеризуемой структурой $\vec{E}_0(\vec{r}) = \vec{E}_0 \exp(ik_m(\vec{p} \cdot \vec{r}))$.

В связи с этим

$$\begin{aligned} \int_{D_s} \vec{E}_0 dV &= \Phi(\vec{r}_s) \int_{D_R} \vec{E}_0 dV, & \int_{D_s} \vec{E}_s dV &= \Phi(\vec{r}_s) \int_{D_R} \vec{E}_R dV, \\ \int_{D_s} \vec{H}_0 dV &= \Phi(\vec{r}_s) \int_{D_R} \vec{H}_0 dV, & \int_{D_s} \vec{H}_s dV &= \Phi(\vec{r}_s) \int_{D_R} \vec{H}_R dV. \end{aligned} \quad (26)$$

Между тем частицы расположены на достаточно больших расстояниях друг от друга по сравнению с размерами самих частиц, а отраженное поле в проводящей среде в дальней зоне, как известно, сильно затухает. Поэтому будем предполагать, что отраженные поля \vec{E}'_s, \vec{H}'_s с окружающими частицами не взаимодействуют. В результате положим

$$\int_{D_0} \vec{E}'_s dV = \int_{D_{sp}} \vec{E}'_s dV = \Phi(\vec{r}_s) \int_{D_{RP}} \vec{E}'_R dV, \quad \int_{D_0} \vec{H}'_s dV = \int_{D_{sp}} \vec{H}'_s dV = \Phi(\vec{r}_s) \int_{D_{RP}} \vec{H}'_R dV. \quad (27)$$

Поля (25) с учетом формул (23), (26), (27) примут вид

$$\begin{aligned} \vec{E} &= \frac{1}{V_P} \int_{D_p} \vec{E}_0 dV + vG_P \left(\int_{D_{RP}} \vec{E}'_R dV + \int_{D_R} \vec{E}_R dV - \int_{D_R} \vec{E}_0 dV \right), \\ \vec{H} &= \frac{1}{V_P} \int_{D_p} \vec{H}_0 dV + vG_P \left(\int_{D_{RP}} \vec{H}'_R dV + \int_{D_R} \vec{H}_R dV - \int_{D_R} \vec{H}_0 dV \right). \end{aligned} \quad (28)$$

Используя разложения полей в ряды (7), (9), (10) и учитывая формулы (21), (22), вычислим интегралы, входящие в (28):

$$\begin{aligned} \int_{D_p} \vec{E}_0 dV &= \sum_{m=-1}^1 b_{m1} \int_{D_p} \vec{n}_{m1}(\vec{r}, k_m) dV = -\frac{V_P G_P}{3} (2b_{11} \vec{e}_+ + b_{-11} \vec{e}_- - 2b_{01} \vec{e}_z), \\ \int_{D_p} \vec{H}_0 dV &= h_m \sum_{m=-1}^1 a_{m1} \int_{D_p} \vec{n}_{m1}(\vec{r}, k_m) dV = -h_m \frac{V_P G_P}{3} (2a_{11} \vec{e}_+ + a_{-11} \vec{e}_- - 2a_{01} \vec{e}_z), \end{aligned}$$

$$\begin{aligned}
\int_{D_R} \vec{E}_0 dV &= -F_M (2b_{11} \vec{e}_+ + b_{-11} \vec{e}_- - 2b_{01} e_z), \quad F_M = \frac{V_R}{\xi_M} j_1(\xi_M), \\
\int_{D_R} \vec{H}_0 dV &= -h_M F_M (2a_{11} \vec{e}_+ + a_{-11} \vec{e}_- - 2a_{01} e_z), \\
\int_{D_{RP}} \vec{E}'_R dV &= \sum_{m=-1}^1 y_{m1}^{(2)} \int_{D_{RP}} \tilde{n}_{m1}(\vec{r}, k_m) dV = f_M (2y_{11}^{(2)} \vec{e}_+ + y_{-11}^{(2)} \vec{e}_- - 2y_{01}^{(2)} \vec{e}_z), \\
\int_{D_{RP}} \vec{H}'_R dV &= h_M \sum_{m=-1}^1 x_{m1}^{(2)} \int_{D_{RP}} \tilde{n}_{m1}(\vec{r}, k_m) dV = h_M f_M (2x_{11}^{(2)} \vec{e}_+ + x_{-11}^{(2)} \vec{e}_- - 2x_{01}^{(2)} \vec{e}_z), \\
\int_{D_R} \vec{E}_R dV &= \sum_{m=-1}^1 \left(x_{m1}^{(1)} \int_{D_R} \tilde{n}_{m1}(\vec{r}, k_1) dV + y_{m1}^{(1)} \int_{D_R} \tilde{n}_{m1}(\vec{r}, k_2) dV \right) = \\
&= 2(F_1 x_{01}^{(1)} + F_2 y_{01}^{(1)}) \vec{e}_z - 2(F_1 x_{11}^{(1)} + F_2 y_{11}^{(1)}) \vec{e}_+ - (F_1 x_{-11}^{(1)} + F_2 y_{-11}^{(1)}) \vec{e}_-, \\
\int_{D_R} \vec{H}_R dV &= \sum_{m=-1}^1 \left(x_{m1}^{(1)} p_1 \int_{D_R} \tilde{n}_{m1}(\vec{r}, k_1) dV + y_{m1}^{(1)} p_2 \int_{D_R} \tilde{n}_{m1}(\vec{r}, k_2) dV \right) = \\
&= 2(F_1 p_1 x_{01}^{(1)} + F_2 p_2 y_{01}^{(1)}) \vec{e}_z - 2(F_1 p_1 x_{11}^{(1)} + F_2 p_2 y_{11}^{(1)}) \vec{e}_+ - (F_1 p_1 x_{-11}^{(1)} + F_2 p_2 y_{-11}^{(1)}) \vec{e}_-.
\end{aligned} \tag{29}$$

Теорема 1. Усредненные по области D_p в композитном материале со сферическими биизотропными частицами радиуса R , расположенными в проводящей матрице, электрические и магнитные поля в базисе \vec{e}_+ , \vec{e}_- , \vec{e}_z при выполнении условий (1), (26), (27) определяются формулами

$$\begin{aligned}
\vec{E} &= -E_- \vec{e}_+ - E_+ \vec{e}_- + E_z \vec{e}_z = E_x \vec{e}_x + E_y \vec{e}_y + E_z \vec{e}_z = \\
&= G_P \left[\left(-\frac{2}{3} \bar{t} b_{11} + 2v (f_M y_{11}^{(2)} - F_1 x_{11}^{(1)} - F_2 y_{11}^{(1)}) \right) \vec{e}_+ + \right. \\
&\quad \left. + \left(-\frac{1}{3} \bar{t} b_{-11} + v (f_M y_{-11}^{(2)} - F_1 x_{-11}^{(1)} - F_2 y_{-11}^{(1)}) \right) \vec{e}_- + \right. \\
&\quad \left. + \left(\frac{2}{3} \bar{t} b_{01} - 2v (f_M y_{01}^{(2)} - F_1 x_{01}^{(1)} - F_2 y_{01}^{(1)}) \right) \vec{e}_z \right],
\end{aligned} \tag{30}$$

$$\begin{aligned}
\vec{H} &= -H_- \vec{e}_+ - H_+ \vec{e}_- + H_z \vec{e}_z = H_x \vec{e}_x + H_y \vec{e}_y + H_z \vec{e}_z = \\
&= G_P \left[\left(-\frac{2\bar{t}}{3} h_M a_{11} + 2v (h_M f_M x_{11}^{(2)} - F_1 p_1 x_{11}^{(1)} - F_2 p_2 y_{11}^{(1)}) \right) \vec{e}_+ + \right. \\
&\quad \left. + \left(-\frac{\bar{t}}{3} h_M a_{-11} + v (h_M f_M x_{-11}^{(2)} - F_1 p_1 x_{-11}^{(1)} - F_2 p_2 y_{-11}^{(1)}) \right) \vec{e}_- + \right. \\
&\quad \left. + \left(\frac{2\bar{t}}{3} h_M a_{01} - 2v (h_M f_M x_{01}^{(2)} - F_1 p_1 x_{01}^{(1)} - F_2 p_2 y_{01}^{(1)}) \right) \vec{e}_z \right],
\end{aligned}$$

где $E_+ = \frac{1}{2}(iE_y + E_x)$, $E_- = \frac{1}{2}(iE_y - E_x)$.

Для доказательства достаточно подставить выражения (29) в формулы (28).

Следствие 1. Имеют место матричные соотношения для компонентов полей (30):

$$\vec{V}_+ = \frac{1}{2} G_P \hat{L} \vec{X}_{-1}^{(1)}, \quad \vec{V}_- = G_P \hat{L} \vec{X}_1^{(1)}, \quad \vec{V}_z = G_P \hat{L} \vec{X}_0^{(1)}, \tag{31}$$

где

$$\hat{L} = i \frac{2}{3} \bar{\tau} \bar{\epsilon}_m^2 \hat{H} \hat{P} + 2\nu (f_m \hat{H} \hat{K} + \hat{F}), \quad (32)$$

$$\vec{V}_+ = \begin{pmatrix} E_+ \\ H_+ \end{pmatrix}, \quad \vec{V}_- = \begin{pmatrix} E_- \\ H_- \end{pmatrix}, \quad \vec{V}_z = \begin{pmatrix} E_z \\ H_z \end{pmatrix}, \quad \hat{H} = \begin{pmatrix} 0 & 1 \\ h_m & 0 \end{pmatrix}, \quad \hat{F} = \begin{pmatrix} F_1 & F_2 \\ p_1 F_1 & p_2 F_2 \end{pmatrix}.$$

6. Усредненные электрическая и магнитная индукции

Запишем суммарные индукции в области D_p :

$$\vec{D}_{\text{сум}} = \begin{cases} \epsilon_m \left(\vec{E}_0 + \sum_{S=1}^N \vec{E}'_S \right) & \text{в } D_0, \\ \epsilon \vec{E}_S + G \vec{H}_S & \text{в } D_S; \end{cases} \quad \vec{B}_{\text{сум}} = \begin{cases} \mu_m \left(\vec{H}_0 + \sum_{S=1}^N \vec{H}'_S \right) & \text{в } D_0, \\ \mu \vec{H}_S + Z \vec{E}_S & \text{в } D_S. \end{cases}$$

Усредненные индукции определяются формулами

$$\vec{D} = \frac{1}{V_P} \int_{D_p} \vec{D}_{\text{сум}} dV; \quad \vec{B} = \frac{1}{V_P} \int_{D_p} \vec{B}_{\text{сум}} dV.$$

Получим формулы для индукций, аналогичные формулам (25):

$$\begin{aligned} \vec{D} &= \frac{1}{V_P} \left[\epsilon_m \int_{D_p} \vec{E}_0 dV + \sum_{S=1}^N \left(\epsilon_m \int_{D_0} \vec{E}'_S dV + \int_{D_S} (\epsilon \vec{E}_S + G \vec{H}_S) dV - \epsilon_m \int_{D_0} \vec{E}_0 dV \right) \right]; \\ \vec{B} &= \frac{1}{V_P} \left[\mu_m \int_{D_p} \vec{H}_0 dV + \sum_{S=1}^N \left(\mu_m \int_{D_0} \vec{H}'_S dV + \int_{D_S} (\mu \vec{H}_S + Z \vec{E}_S) dV - \mu_m \int_{D_0} \vec{H}_0 dV \right) \right]. \end{aligned} \quad (33)$$

Терема 2. Усредненные по области D_p в композитном материале со сферическими би-зотропными частицами радиуса R , расположенными в проводящей матрице, электрическая и магнитная индукции в базисе \vec{e}_+ , \vec{e}_- , \vec{e}_z определяются формулами

$$\begin{aligned} \vec{D} &= -D_- \vec{e}_+ - D_+ \vec{e}_- + D_z \vec{e}_z = D_x \vec{e}_x + D_y \vec{e}_y + D_z \vec{e}_z = \\ &= G_p \left\{ \left[-\frac{2}{3} \epsilon_m \bar{\tau} b_{11} + 2\nu (\epsilon_m f_m y_{11}^{(2)} - F_1^{(1)} x_{11}^{(1)} - F_2^{(1)} y_{11}^{(1)}) \right] \vec{e}_+ + \right. \\ &\quad + \left[-\frac{1}{3} \epsilon_m \bar{\tau} b_{-11} + \nu (\epsilon_m f_m y_{-11}^{(2)} - F_1^{(1)} x_{-11}^{(1)} - F_2^{(1)} y_{-11}^{(1)}) \right] \vec{e}_- + \\ &\quad \left. + \left[\frac{2}{3} \epsilon_m \bar{\tau} b_{01} - 2\nu (\epsilon_m f_m y_{01}^{(2)} - F_1^{(1)} x_{01}^{(1)} - F_2^{(1)} y_{01}^{(1)}) \right] \vec{e}_z \right\}; \\ \vec{B} &= -B_- \vec{e}_+ - B_+ \vec{e}_- + B_z \vec{e}_z = B_x \vec{e}_x + B_y \vec{e}_y + B_z \vec{e}_z = \\ &= G_p \left\{ \left[-\frac{2}{3} \mu_m h_m \bar{\tau} a_{11} + 2\nu (\mu_m h_m f_m x_{11}^{(2)} - F_1^{(2)} x_{11}^{(1)} - F_2^{(2)} y_{11}^{(1)}) \right] \vec{e}_+ + \right. \\ &\quad + \left[-\frac{1}{3} \mu_m h_m \bar{\tau} a_{-11} + \nu (\mu_m h_m f_m x_{-11}^{(2)} - F_1^{(2)} x_{-11}^{(1)} - F_2^{(2)} y_{-11}^{(1)}) \right] \vec{e}_- + \\ &\quad \left. + \left[\frac{2}{3} \mu_m h_m \bar{\tau} a_{01} - 2\nu (\mu_m h_m f_m x_{01}^{(2)} - F_1^{(2)} x_{01}^{(1)} - F_2^{(2)} y_{01}^{(1)}) \right] \vec{e}_z \right\}, \end{aligned} \quad (34)$$

где $F_1^{(1)} = F_1(\varepsilon + p_1 G)$, $F_2^{(1)} = F_2(\varepsilon + p_2 G)$, $F_1^{(2)} = F_1(Z + p_1 \mu)$, $F_2^{(2)} = F_2(Z + p_2 \mu)$,

$$D_+ = \frac{1}{2}(iD_y + D_x), \quad D_- = \frac{1}{2}(iD_y - D_x), \quad B_+ = \frac{1}{2}(iB_y + B_x), \quad B_- = \frac{1}{2}(iB_y - B_x).$$

Для доказательства формул (34) достаточно к разложениям индукции (33) последовательно применить формулы (23), (26), (27), (29).

Следствие 2. *Имеют место матричные соотношения для компонентов индукций (34):*

$$\vec{W}_+ = \frac{G_p}{2} \hat{N} \vec{X}_{-1}^{(1)}, \quad \vec{W}_- = G_p \hat{N} \vec{X}_1^{(1)}, \quad \vec{W}_z = G_p \hat{N} \vec{X}_0^{(1)}, \quad (35)$$

где

$$\hat{N} = i \frac{2}{3} \bar{\varepsilon}_{5M}^2 \hat{H}_1 \hat{P} + 2\nu (f_m \hat{H}_1 \hat{K} + \hat{F}_1), \quad (36)$$

$$\vec{W}_+ = \begin{pmatrix} D_+ \\ B_+ \end{pmatrix}, \quad \vec{W}_- = \begin{pmatrix} D_- \\ B_- \end{pmatrix}, \quad \vec{W}_z = \begin{pmatrix} D_z \\ B_z \end{pmatrix}, \quad \hat{H}_1 = \begin{pmatrix} 0 & \varepsilon_m \\ h_m \mu_m & 0 \end{pmatrix}, \quad \hat{F}_1 = \begin{pmatrix} F_1^{(1)} & F_2^{(1)} \\ F_1^{(2)} & F_2^{(2)} \end{pmatrix}.$$

7. Вычисление эффективных материальных параметров композита

Выразим векторы электрической и магнитной индукции (34) через векторы электрической и магнитной напряженностей поля. Воспользуемся формулами (31), получим

$$\vec{X}_{-1}^{(1)} = \frac{2}{G_p} \hat{L}^{-1} \vec{V}_+, \quad \vec{X}_1^{(1)} = \frac{1}{G_p} \hat{L}^{-1} \vec{V}_-, \quad \vec{X}_0^{(1)} = \frac{1}{G_p} \hat{L}^{-1} \vec{V}_z.$$

Подставим данные выражения в (35), тогда

$$\vec{W}_+ = \hat{C} \vec{V}_+, \quad \vec{W}_- = \hat{C} \vec{V}_-, \quad \vec{W}_z = \hat{C} \vec{V}_z, \quad (37)$$

где $\hat{C} = \hat{N} \hat{L}^{-1} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}.$

Запишем (37) покомпонентно в виде

$$D_+ = c_{11} E_+ + c_{12} H_+, \quad D_- = c_{11} E_- + c_{12} H_-; \quad (38)$$

$$B_+ = c_{21} E_+ + c_{22} H_+, \quad B_- = c_{21} E_- + c_{22} H_-; \quad (39)$$

$$D_z = c_{11} E_z + c_{12} H_z, \quad B_z = c_{21} E_z + c_{22} H_z. \quad (40)$$

Складывая и вычитая (38) (аналогично складывая и вычитая (39)), получим

$$\begin{aligned} D_x &= c_{11} E_x + c_{12} H_x, & D_y &= c_{11} E_y + c_{12} H_y; \\ B_x &= c_{21} E_x + c_{22} H_x, & B_y &= c_{21} E_y + c_{22} H_y. \end{aligned} \quad (41)$$

Равенства (40), (41) запишем в векторном виде

$$\vec{D} = c_{11} \vec{E} + c_{12} \vec{H}, \quad \vec{B} = c_{21} \vec{E} + c_{22} \vec{H}. \quad (42)$$

Теорема 3. *Композитный материал, состоящий из случайно распределенных биизотропных сферических частиц радиуса R , описываемых уравнениями Максвелла (3), и заполненный проводящей средой в областях между частицами, является биизотропной средой. Эффективные параметры композита определяются формулами*

$$\varepsilon_{\text{эф}} = c_{11}, \quad G_{\text{эф}} = c_{12}, \quad \mu_{\text{эф}} = c_{22}, \quad Z_{\text{эф}} = c_{21}, \quad (43)$$

где

$$\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} = \frac{1}{d} \hat{N} \hat{M}, \quad \hat{N} = \begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix}, \quad \hat{M} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}.$$

$$n_{11} = \varepsilon_m (p_0 p_{21} + 2\nu F_0 q_{21}) + 2\nu F_1^{(1)},$$

$$n_{12} = \varepsilon_m (p_0 p_{22} + 2\nu F_0 q_{22}) + 2\nu F_2^{(1)},$$

$$n_{21} = \mu_m h_m (p_0 p_{11} + 2\nu F_0 q_{11}) + 2\nu F_1^{(2)},$$

$$n_{22} = \mu_m h_m (p_0 p_{12} + 2\nu F_0 q_{12}) + 2\nu F_2^{(2)},$$

$$m_{11} = h_m (p_0 p_{12} + 2\nu F_0 q_{12}) + 2\nu p_2 F_2,$$

$$m_{12} = -(p_0 p_{22} + 2\nu F_0 q_{22} + 2\nu F_2),$$

$$m_{21} = -h_m (p_0 p_{11} + 2\nu F_0 q_{11}) - 2\nu p_1 F_1,$$

$$m_{22} = p_0 p_{21} + 2\nu F_0 q_{21} + 2\nu F_1,$$

$$F_0 = \frac{V_R}{\xi_m}, \quad p_0 = i \frac{2}{3} \xi_m^2 \left(\bar{\tau} + 3\nu \frac{V_R}{\xi_m} j_1(\xi_m) \right), \quad d = m_{11} m_{22} - m_{12} m_{21}, \quad \xi_m = k_m R.$$

Матричные элементы p_{js}, q_{js} определены в (18).

Доказательство. Сравнивая индукции (42), построенные на основе разработанной методики моделирования, с индукциями уравнений (2), получим формулы (43). Матричные элементы матриц (44) получаем с использованием формул (32), (36), (37), где $L^{-1} = \frac{1}{d} \hat{M}$.

8. Вычислительный эксперимент

На рис.1 показаны зависимости от τ при частоте $f = 0,3$ ГГц относительных эффективных материальных параметров $\varepsilon_s = \varepsilon_{эф}/\varepsilon_0$, $\mu_s = \mu_{эф}/\mu_0$, $G_s = cG_{эф}$, $Z_s = cZ_{эф}$ (см. (43)) композита, представляющего собой следующую структуру: в матрице из дважды левостороннего метаматериала ($\varepsilon_m = -1, \mu_m = -1, G_m = 0, Z_m = 0$) размещены сферические частицы радиуса $R = 4$ мкм из фокусирующего метаматериала с материальными параметрами $\varepsilon_r = -4, \mu_r = 2, G_r = 3i, Z_r = 3i$ для рис. 1, а и материальными параметрами $\varepsilon_r = 2 + 6i, \mu_r = -4 - 6i, G_r = -6 + 3i, Z_r = -6 + 3i$ для рис. 1, б. Метаматериал будем называть фокусирующим, если

$$\varepsilon_r = \chi - 1, \quad \mu_r = -(\chi + 1), \quad G_r = Z_r = i\chi \quad \text{или} \quad \varepsilon_r = -(\chi + 1), \quad \mu_r = \chi - 1, \quad G_r = Z_r = i\chi, \quad (44)$$

где χ – произвольное комплексное число. Коэффициент заполнения матрицы частицами на графиках изменяется в пределах $0,001 < \tau < 0,5$, где $\tau = 0,5$ в случае, когда соседние частицы практически соприкасаются. На рис. 1, а компоненты $\text{Im}\varepsilon_s = \text{Im}\mu_s = \text{Re}G_s = \text{Re}Z_s = 0$.

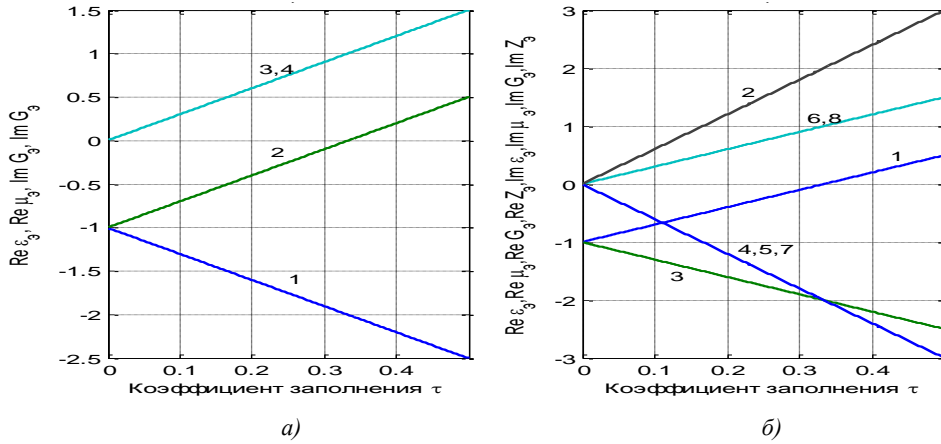


Рис. 1. Эффективные параметры фокусирующих композитов: а) 1 – $\text{Re}\epsilon_3$, 2 – $\text{Re}\mu_3$, 3 – $\text{Im}G_3$, 4 – $\text{Im}Z_3$; б) 1 – $\text{Re}\epsilon_3$, 2 – $\text{Im}\epsilon_3$, 3 – $\text{Re}\mu_3$, 4 – $\text{Im}\mu_3$, 5 – $\text{Re}G_3$, 6 – $\text{Im}G_3$, 7 – $\text{Re}Z_3$, 8 – $\text{Im}Z_3$

Численно показано, что эффективные параметры с достаточно малой погрешностью являются фокусирующими (44).

На рис. 2 изображено поведение на различных частотах эффективных материальных параметров композита, состоящего из магнитодиэлектрической матрицы с относительными диэлектрической и магнитной проницаемостями $\epsilon_m = 2 + i$, $\mu_m = -3$ и сферическими биизотропными частицами радиуса $R = 4$ мкм с комплексными материальными параметрами $\epsilon_r = -3 + 5i$, $\mu_r = 2 - i$, $G_r = 5 + 4i$, $Z_r = 4 - 4i$.

Сравнение графиков на различных частотах показывает, что в пределах $10 \text{ МГц} < f < 2 \text{ ТГц}$ графики практически не изменяются. В пределах частот $2 \text{ ТГц} < f < 4 \text{ ТГц}$ происходит изменение картины поведения линий. Это связано с тем, что отношение длины волны и радиуса частицы при $f = 2 \text{ ТГц} - \frac{\lambda}{R} = 37,5$, при $f = 4 \text{ ТГц} - \frac{\lambda}{R} = 18,75$, т. е. начинает нарушаться условие модели (3), для которого $\frac{\lambda}{R} > 30$.

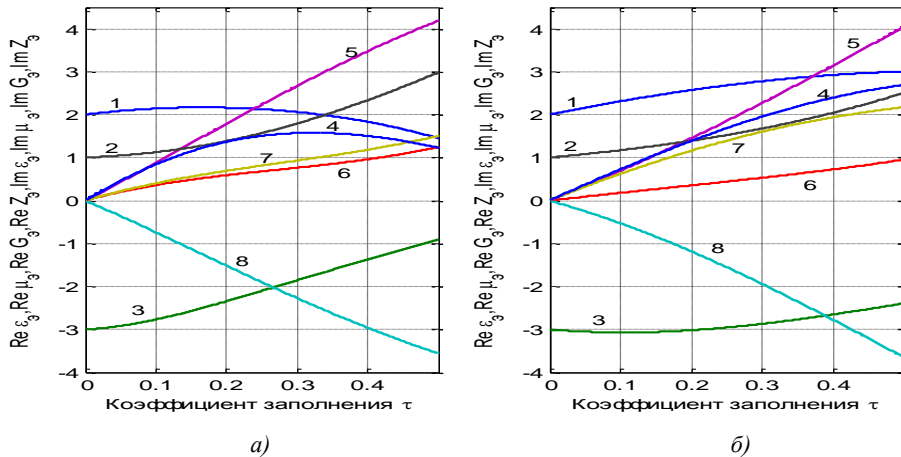


Рис. 2. Эффективные параметры композита с произвольными материальными параметрами матрицы и биизотропных частиц: а) $f = 10 \text{ МГц}$; б) $f = 4 \text{ ТГц}$; 1 – $\text{Re}\epsilon_3$, 2 – $\text{Im}\epsilon_3$, 3 – $\text{Re}\mu_3$, 4 – $\text{Im}\mu_3$, 5 – $\text{Re}G_3$, 6 – $\text{Im}G_3$, 7 – $\text{Re}Z_3$, 8 – $\text{Im}Z_3$

На рис. 3, а показаны зависимости от координаты ρ поля за экраном D с параметрами $\epsilon_m = 1$, $\mu_m = 1$ (вакуум) толщины $\Delta = 0,05$ м, содержащим киральные сферические частицы радиуса $R = 4$ мкм с материальными параметрами

$$\varepsilon_r = 4, \quad \mu_r = 1, \quad G_r = 0,5i, \quad Z_r = -0,5i \quad (45)$$

и с коэффициентом заполнения $\tau = 0,3$. Вычисляются компоненты поля на плоскости $z = 0,4$ м при воздействии на экран поля магнитного диполя (1) с параметрами $f = 3$ ГГц, $A = 1, B = 0, h = 0,1$ м.

Заметим, что на рис. 2 при уменьшении коэффициента заполнения τ эффективные параметры композита приближаются к материальным параметрам матрицы, так как материал матрицы преобладает в композите с малым числом частиц. С другой стороны, в случае увеличения коэффициента τ эффективные параметры приближаются к значениям материальных параметров частиц, так как влияние частиц в композите увеличивается.

На рис. 3, а показано, что график 1 уменьшается на малую величину в сравнении с графиком 1 на рис. 1, а статьи [19], который изображает поле диполя без экрана. Это означает, что поле магнитного диполя практически полностью проходит через экран с частицами, малая часть поля преобразуется в поле электрического диполя за экраном (см. рис. 1, а, графики 2 и 3).

Заметим, что в статье [19, рис. 1] рассмотрен экран D , который полностью заполнен киральной средой с параметрами (45), при воздействии поля магнитного диполя с аналогичными параметрами.

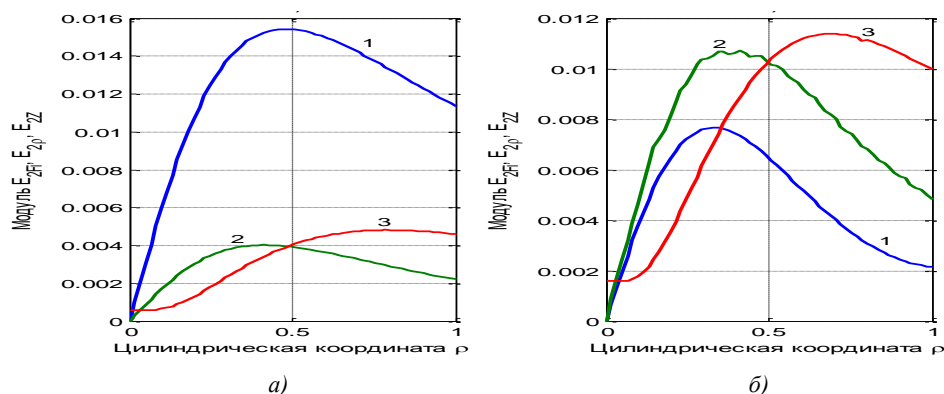


Рис. 3. Прохождение поля магнитного диполя через композитный экран с киральными частицами: а) $G_r = 0,5i, Z_r = -0,5i$; б) $G_r = 1,5i, Z_r = -1,5i$

Обозначим $G_r = ki, Z_r = -ki$, где k – киральность. На рис. 3, б рассмотрен случай, когда киральность частиц увеличена в три раза по сравнению с частицами на рис. 3, а. Показано, что поле магнитного диполя за экраном уменьшилось в два раза (рис. 3, б, график 1), а поле электрического диполя за экраном увеличилось в 2,4 раза (рис. 3, б, графики 2 и 3) по сравнению с рис. 3, а.

Заключение

Разработан алгоритм вычисления эффективных материальных параметров композита, состоящего из случайно распределенных в магнитоэлектрической матрице сферических биизотропных частиц. Эффективные параметры описывают усредненную однородную биизотропную среду, эквивалентную исходному дискретно-неоднородному матричному композиту. Показано, что композит, содержащий в матрице из метаматериала частицы из биизотропных фокусирующих материалов, имеет фокусирующие эффективные параметры среды. На основе использования эффективных параметров разработана методика решения краевой задачи проникновения электромагнитных полей диполей через плоский экран из матричного композита. Численно исследована структура поля, прошедшего через экран, в зависимости от параметров частиц композита.

Список литературы

1. Магнитоэлектрические материалы / М.И. Бичурин [и др.]. – М. : Акад. естествознания, 2006. – 296 с.
2. Лагарьков, А.Н. Радиопоглощающие материалы на основе метаматериалов / А.Н. Лагарьков, В.Н. Кисель, В.Н. Семененко // Радиотехника и электроника. – 2012. – Т. 57, № 10. – С. 1119–1127.
3. Cui, Tie Jun. Metamaterials. Theory, Design and Applications / Tie Jun Cui, D.R. Smith, R. Lui. – Springer, 2009. – 367 p.
4. Неганов, В.А. Современное состояние электродинамики искусственных киральных сред (обзор) / В.А. Неганов, О.В. Осипов // Физика волновых процессов и радиотехнические системы. – 2005. – Т. 8, № 1. – С. 7–33.
5. Наноматериалы и нанотехнологии / В.М. Анищик [и др.]. – Минск : Изд. центр БГУ, 2008. – 373с.
6. Федосюк, В.М. Наноструктурные пленки и нанопроволоки / В.М. Федосюк. – Минск : Издательский центр БГУ, 2006. – 310 с.
7. Памятных, Е.А. Основы электродинамики материальных сред в переменных и неоднородных полях / Е.А. Памятных, Е.А. Туров. – М. : Наука. Физматлит, 2000. – 240 с.
8. Guerin, F. Scattering of electromagnetic waves by helices and application to the modeling of chiral composites / F. Guerin // J. Phys. D: Appl. Phys. – 1995. – Vol. 28. – P. 623–642.
9. Демидчик, В.И. Математическое моделирование характеристик рассеяния проволочных частиц произвольной конфигурации и композитных материалов на их основе / В.И. Демидчик, Р.В. Корнеев, П.Д. Кухарчик // Докл. НАН Беларуси. – 2005. – Т. 49, № 2. – С. 38–41.
10. Шатров, А.Д. Модель биизотропной среды из резонансных сферических частиц с идеальной смешанной проводимостью поверхности вдоль спиральных линий / А.Д. Шатров // Радиотехника и электроника. – 2000. – Т. 45, № 10. – С. 1168–1170.
11. Костин, М.В. К теории киральной среды на основе сферических спирально проводящих частиц / М.В. Костин, В.В. Шевченко // Радиотехника и электроника – 1998. – Т. 43, № 8. – С. 921–926.
12. Виноградов, А.П. К вопросу об эффективных параметрах метаматериалов / А.П. Виноградов, А.В. Дорофеев, С. Зухди // Успехи физических наук. – 2008. – Т. 178, № 5. – С. 514–518.
13. Исимару, А. Распространение и рассеяние волн в случайно неоднородных средах. Однократное рассеяние / А. Исимару. – М. : Мир, 1981. – Т. 1. – 280 с.
14. Иванов, О.В. Распространение электромагнитных волн в анизотропных и бианизотропных слоистых структурах / О.В. Иванов. – Ульяновск : УлГТУ, 2010. – 262 с.
15. Ерофеев, В.Т. Дифракция плоской электромагнитной волны на плоскостойкой структуре из биизотропных материалов / В.Т. Ерофеев, С.В. Малый // Информатика. – 2012. – № 1 (33). – С. 58–65.
16. Ерофеев, В.Т. Аналитическое моделирование в электродинамике / В.Т. Ерофеев, И.С. Козловская. – М. : КД Либроком, 2014. – 304 с.
17. Ерофеев, В.Т. Решение задачи о прохождении плоских электромагнитных волн через бианизотропную неоднородную пластину / В.Т. Ерофеев, А.К. Сеницын // Информатика. – 2013. – № 2. – С. 11–21.
18. Проникновение электромагнитных волн через композитные экраны, содержащие идеально проводящие спирали / В.Т. Ерофеев [и др.] // Инженерно-физический журнал. – 2011. – Т. 84, № 4. – С. 740–746.
19. Ерофеев, В.Т. Численное исследование взаимодействия электромагнитных полей электрического и магнитного диполей с композитным экраном / В.Т. Ерофеев, В.Ф. Бондаренко // Известия НАН Беларуси. Сер. физ.-техн. наук. – 2013. – № 4. – С. 113–120.
20. Ерофеев, В.Т. Электродинамическая модель расчета эффективных параметров композитов из сферических биизотропных частиц / В.Т. Ерофеев // Информатика. – 2014. – № 1 (41). – С. 45–58.

21. Ерофеенко, В.Т. Теоремы сложения / В.Т. Ерофеенко. – Минск : Наука и техника, 1989. – 256 с.

22. Engheta, N. Antenna radiation in the presence of a chiral sphere / N. Engheta, M.W. Kowarz // J. Appl. Phys. – 1990. – Vol. 67 (2). – P. 639–647.

Поступила 15.04.2014

*Учреждение БГУ «НИИ прикладных проблем
математики и информатики»,
Минск, пр. Независимости, 4
e-mail: bsu_erofeenko@tut.by*

V.T. Erofeenko, V.F. Bondarenko

SCREENING ELECTROMAGNETIC FIELDS BY MEANS OF SCREENS FROM MATRIX COMPOSITES, CONTAINING BIISOTROPIC FRACTIONS

A procedure for solution of the boundary-value problem of screening electromagnetic fields of concentrated sources of electrical and magnetic dipoles by means of homogeneous magnetodielectric screens, containing sparse system of spherical fractions with complex permeability and permittivity and complex parameters of biisotropy, is developed. The procedure is based on the algorithm of effective material parameters calculation for averaged biisotropic medium equivalent to matrix composite of the screen. The components of the field beyond the screen for composite with fractions from metamaterials and chiral particles are calculated.

УДК 658.512.2

Г.М. Левин¹, Б.М. Розин¹, А.Б. Долгий²

ЛИНЕЙНАЯ АППРОКСИМАЦИЯ ЗАДАЧИ ОПТИМИЗАЦИИ ИНТЕНСИВНОСТЕЙ ПОСЛЕДОВАТЕЛЬНО-ПАРАЛЛЕЛЬНОГО ВЫПОЛНЕНИЯ ПЕРЕСЕКАЮЩИХСЯ МНОЖЕСТВ ОПЕРАЦИЙ

Предлагаются математическая модель и метод оптимизации интенсивностей последовательно-параллельного выполнения пересекающихся множеств операций. Интенсивность любой операции остается неизменной в составе любого множества. Зависимости материальных и временных затрат на выполнение любой операции в составе конкретного множества от ее интенсивности представимы выпуклыми функциями. Предложенный метод основывается на аппроксимации исходной задачи задачей линейного программирования. Приводится интерпретация в терминах рассмотренной задачи одной из задач оптимизации режимов групповой многоинструментальной обработки деталей на многопозиционном оборудовании.

Введение

Задачам планирования процессов выполнения комплексов операций в системах различного назначения в последние десятилетия уделялось значительное внимание [1–12]. В данной работе исследуется задача оптимизации интенсивностей последовательно-параллельного выполнения операций в составе комплекса пересекающихся между собой множеств операций (далее – работ). Под интенсивностью операции подразумевается время, затрачиваемое на выполнение единицы ее объема. Исследуемая задача является обобщением рассмотренной в работе [12] задачи оптимизации интенсивности последовательно-параллельного выполнения пересекающихся множеств операций. Отличие рассматриваемой постановки задачи заключается в основном в двух аспектах: в допущении различных затрат на выполнение единицы объема одной и той же операции в составе различных работ, причем эти затраты по-прежнему зависят от принимаемой интенсивности операции; и в возможности учета дополнительных затрат времени на восстановление ресурсов, затрачиваемых на выполнение операций, причем затраты ресурсов на выполнение единицы объема каждой операции также зависят как от их интенсивности, так и от работ, в составе которых они выполняются. Рассматриваемая задача сводится к задаче, исследованной в работе [12], если объем и зависимость затрат на единицу объема любой операции от ее интенсивности одинаковы для всех работ, в состав которых она входит, а также нет дополнительных затрат времени на восстановление ресурсов, расходуемых при выполнении работ. В настоящей работе рассматривается частный случай задачи, когда указанные зависимости материальных и временных затрат на выполнение операций в составе различных работ с достаточной для реальных ситуаций точностью могут быть представлены выпуклыми функциями.

1. Общая постановка задачи

В производственной системе планируется последовательное выполнение (однократное либо циклически повторяющееся) всех работ комплекса \mathbf{I} , образованного множеством $I = \{1, \dots, i, \dots, n\}$ различных работ, причем этот комплекс включает n_i идентичных работ $i \in I$. Выполнение очередной работы $i \in I$ комплекса \mathbf{I} заключается в параллельном выполнении всех операций соответствующего подмножества J_i исходного множества J операций, причем в состав работы i также может входить m_{ij} идентичных операций $j \in J_i$. Длительность любой работы равна наибольшей из длительностей входящих в нее операций. Подмножества семейства $\{J_1, \dots, J_i, \dots, J_n\}$, образующие работы комплекса, могут пересекаться и $\bigcup_{i=1}^n J_i = J$ (рис. 1).

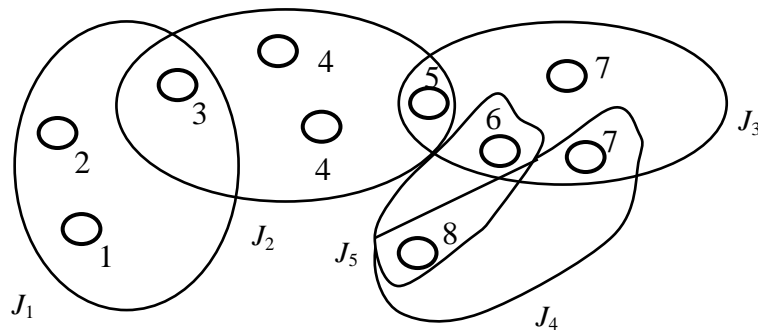


Рис. 1. Пример пересечения работ комплекса

При выполнении каждой операции в составе некоторой работы комплекса расходуются соответствующие ресурсы в количествах, зависящих как от ее объема, так и от интенсивности выполнения. Максимальное количество каждого ресурса, которое может быть в системе, ограничено. Восстановление любого из ресурсов осуществляется после его полного расходования по завершении работы, во время которой это произошло. Выполнение очередной работы может начаться лишь после завершения процесса восстановления соответствующего ресурса.

Необходимо определить интенсивности выполнения всех операций комплекса при условии, что принимаемая интенсивность выполнения операции $j \in J$ должна быть одинакова для всех содержащих ее работ.

Предполагаются заданными следующие исходные данные планируемого процесса:

- объем V_{ij} операции $j \in J$ в содержащей ее работе $i \in I$;
- отрезок $[s_{1j}, s_{2j}]$ возможных значений интенсивности s_j выполнения операции $i \in I$;
- определенные на этом отрезке выпуклые функции $f_{ij}(s_j)$ и $\varphi_{ij}(s_j)$, представляющие зависимости от принимаемой интенсивности s_j выполнения операции $j \in J$ отнесенных к единице объема затрат на ее выполнение (включая затраты на ресурсы и их восстановление) в составе работы $i \in I$ и времени восстановления ресурсов соответственно.

Пусть R – множество видов возобновляемых ресурсов системы, расходуемых при выполнении операций комплекса I . В реальных задачах рассматриваемого класса функции $f_{ij}(s_j)$ и $\varphi_{ij}(s_j)$ обусловлены, в частности, максимально возможными количествами U_r каждого из ресурсов $r \in R$ в системе, затратами (материальными и временными) на их восстановление после полного расходования до допустимого уровня, а также зависимостями $u_{ijr}(s_j)$ количества ресурса $r \in R$, расходуемого на выполнение единицы объема операции j в составе работы i с интенсивностью s_j . Обычно эти параметры являются известными. В этом случае можно принять, что на единицу объема операции j в составе работы i приходится $u_{ijr}(s_j)/U_r$ как общих затрат на U_r единиц ресурса $r \in R$, так и общего времени его восстановления и связанных с этим затрат.

Затраты на выполнение каждой из работ $i \in I$ в целом помимо суммарных затрат на выполнение всех составляющих ее операций включают также дополнительные затраты, пропорциональные длительности ее выполнения. Коэффициенты пропорциональности $E_{1i} > 0$, учитывающие эти дополнительные затраты за единицу времени работы производственной системы при выполнении работ $i \in I$, предполагаются заданными. Затраты времени на выполнение каждой из работ $i \in I$ в целом помимо длительности выполнения самой работы включают также дополнительное время на обслуживание производственной системы. Коэффициенты $E_{2i} > 1$, учитывающие эти дополнительные затраты времени за единицу времени работы производственной системы при выполнении работ $i \in I$, также предполагаются заданными.

Задача заключается в определении таких интенсивностей $s_j \in [s_{1j}, s_{2j}]$ выполнения всех операций $j \in J$, которые минимизируют суммарные затраты на выполнение всех работ комплекса I при условии, что общая длительность их выполнения (с учетом затрат времени на восстановление ресурсов) не превосходит заданного значения T^0 , определяемого исходя из требуемой производительности системы.

Производственные системы, планирование функционирования которых включает решение подобных задач, описаны во многих публикациях (например, [2, 3, 6] и др.). Ниже в разд. 3 приводится пример такой системы.

2. Математическая модель и метод решения

Обозначим через I_j подмножество работ i из множества I , содержащих операцию $j \in J$. В дальнейшем под парой ij будем подразумевать операцию $j \in J$, выполняемую в составе работы $i \in I$. Положим $G = \{ij \mid i \in I, j \in J_i\}$.

При сделанных ранее предположениях получаем следующие зависимости интересующих нас характеристик процесса выполнения комплекса \mathbf{I} работ от значения вектора $s = (s_j \mid j \in J) \in \mathbf{S} = \prod_{j \in J} [s_{1j}, s_{2j}]$:

- длительности операции ij и работы i равны $V_{ij}s_j$ и $\max\{V_{ij}s_j \mid j \in J_i\}$ соответственно;
- общие затраты на одну операцию ij равны $V_{ij}f_{ij}(s_j)$;
- затраты времени на восстановление ресурсов, отнесенные к одной операции ij , равны $V_{ij}\varphi_{ij}(s_j)$.

Таким образом, решение исходной проектной задачи может быть получено в результате решения следующей задачи математического программирования:

$$F_1(s) = \sum_{i \in I} n_i (E_{1i} \max_{j \in J_i} (V_{ij}s_j) + \sum_{j \in J_i} m_{ij} V_{ij} f_{ij}(s_j)) \rightarrow \min; \quad (1)$$

$$F_2(s) = \sum_{i \in I} n_i (E_{2i} \max_{j \in J_i} (V_{ij}s_j) + \sum_{j \in J_i} m_{ij} V_{ij} \varphi_{ij}(s_j)) \leq T^0; \quad (2)$$

$$s_j \in [s_{1j}, s_{2j}], \quad j \in J. \quad (3)$$

Здесь функции $F_1(s)$ и $F_2(s)$ представляют общие затраты на выполнение всех работ комплекса \mathbf{I} и общую их продолжительность соответственно.

В силу выпуклости функций $f_{ij}(s_j)$ и $\varphi_{ij}(s_j)$ для всех $i \in I$ и $j \in J_i$ для решения задачи (1)–(3) могут быть использованы известные методы выпуклого программирования, а поскольку вторые слагаемые функций $F_1(s)$ и $F_2(s)$ сепарабельны, то приближенное решение задачи (1)–(3) может быть эффективно получено посредством ее аппроксимации задачей линейного программирования с использованием для решения последней существующих программных средств типа CPLEX. Рассмотрим один из таких подходов более детально.

Построим кусочно-линейные аппроксимации функций $f_{ij}(s_j)$ и $\varphi_{ij}(s_j)$, полагая для всех $ij \in G$:

$$f_{ij}(s_j) \approx \max\{a_{ijk}s_j + b_{ijk} \mid k=1, \dots, v_{ij}\} \quad (4)$$

и

$$\varphi_{ij}(s_j) \approx \max\{c_{ijk}s_j + d_{ijk} \mid k=1, \dots, q_{ij}\}, \quad (5)$$

где a_{ijk} , b_{ijk} , v_{ij} , c_{ijk} , d_{ijk} и q_{ij} являются параметрами этой аппроксимации, параметры v_{ij} и q_{ij} во многом определяют ее точность.

Тогда приближенное решение задачи (1)–(3) может быть получено в результате решения следующей задачи линейного программирования:

$$\sum_{i \in I} \tilde{E}_{1i} t_i + \sum_{i \in I} \sum_{j \in J_i} \tilde{V}_{ij} y_{ij} \rightarrow \min; \quad (6)$$

$$\sum_{i \in I} \tilde{E}_{2i} t_i + \sum_{i \in I} \sum_{j \in J_i} \tilde{V}_{ij} z_{ij} \leq T^0; \quad (7)$$

$$y_{ij} - a_{ijk} s_j \geq b_{ijk}, \quad i \in I, \quad j \in J_i, \quad k=1, \dots, v_{ij}; \quad (8)$$

$$z_{ij} - c_{ijk} s_j \geq d_{ijk}, \quad i \in I, \quad j \in J_i, \quad k=1, \dots, v_{ij}; \quad (9)$$

$$t_i - V_{ij}s_j \geq 0, \quad i \in I, j \in J_i; \tag{10}$$

$$s_j \in [s_{1j}, s_{2j}], \quad j \in J, \tag{11}$$

где $\tilde{E}_{1i} = n_i E_{1i}$, $\tilde{E}_{2i} = n_i E_{2i}$ и $\tilde{V}_{ij} = n_j m_{ij} V_{ij}$. Искомыми в этой задаче являются векторы $t = (t_i \mid i \in I)$, $s = (s_j \mid j \in J)$, $y = (y_{ij} \mid i \in I, j \in J_i)$ и $z = (z_{ij} \mid i \in I, j \in J_i)$. Очевидно, что если (t^*, s^*, y^*, z^*) – ее решение, то вектор s^* может быть принят в качестве приближенного решения исходной задачи (1)–(3). Несовпадение минимальных значений целевых функций этих задач определяется точностью аппроксимации функций $f_{ij}(s_j)$ и $\varphi_{ij}(s_j)$ в окрестности решения исходной задачи.

3. Задача оптимизации режимов групповой многопозиционной обработки

Рассмотрим процесс обработки на многопозиционном многоинструментальном оборудовании последовательности деталей, составленной из следующих друг за другом идентичных подпоследовательностей (групп) $\delta = \{\delta_1, \dots, \delta_n\}$, каждая из которых включает n деталей нескольких наименований $\mathbf{M} = \{1, \dots, m\}$, $n \geq m$. В группе может быть несколько деталей $d \in \mathbf{M}$. Предполагается, что рабочие позиции линейно упорядочены и каждая деталь последовательно в этом порядке обрабатывается на каждой рабочей позиции соответствующим этой позиции и детали набором блоков инструментов, причем в каждый момент времени на каждой позиции может обрабатываться лишь одна деталь. Один такт обработки состоит в одновременной обработке на каждой из рабочих позиций соответствующей (такту и позиции) детали. После завершения любого такта обработки каждая обрабатываемая деталь со своей позиции синхронно перемещается на следующую позицию, деталь с последней рабочей позиции поступает на позицию разгрузки, а на первую рабочую позицию поступает очередная деталь последовательности с загрузочной позиции. Съем и установка деталей на соответствующих позициях совмещены во времени с обработкой деталей на рабочих позициях, причем на оборудовании с кольцевым транспортным устройством загрузочная и разгрузочная позиции обычно совмещены. Таким образом, цикл обработки группы деталей состоит из n тактов. Пусть $I = \{1, \dots, n\}$ – множество номеров тактов в цикле и $\mathbf{P} = \{1, 2, \dots\}$ – множество номеров рабочих позиций. Без ограничения общности можно предположить, что на такте $i \in I$ на позиции $p \in \mathbf{P}$ обрабатывается деталь $d_{pi} = \delta_{\chi(p,i)} \in \mathbf{M}$, где $\chi(p, i) = 1 + \text{mod}((n + i - \text{mod}(p, n)), n)$. Пример многопозиционной сблокированной линии для обработки группы (1, 2, 1, 3) деталей 1, 2, 3 изображен на рис. 2. Для этой линии процесс обработки одной группы деталей состоит из четырех тактов.

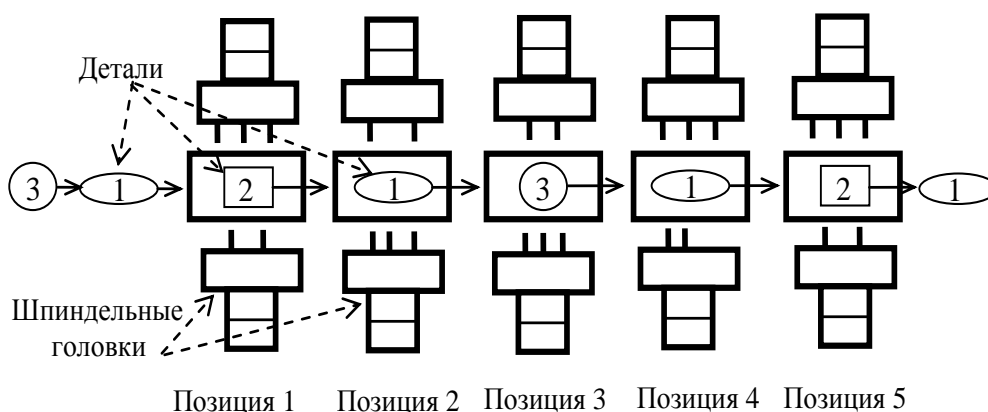


Рис. 2. Многопозиционная сблокированная линия для групповой обработки деталей

В данной работе ограничимся достаточно распространенным на практике случаем, когда:
 – каждая рабочая позиция $p \in \mathbf{P}$ оснащена своим набором J_p инструментальных блоков (далее просто «блоки»), а каждый блок $j \in J_p$ – своим набором U_{pj} инструментов;

– каждой детали $d \in \mathbf{M}$ и позиции $p \in \mathbf{P}$ соответствуют подмножество $J_{pd} \subseteq J_p$ блоков и подмножества $U_{pjd} \subseteq U_{pj}$ инструментов каждого из блоков $j \in J_{pd}$, осуществляющих обработку этой детали, причем все блоки выполняют обработку параллельно;

– для различных деталей из \mathbf{M} как подмножества J_{pd} , так и подмножества U_{pjd} могут пересекаться;

– в процессе обработки различных деталей из \mathbf{M} минутная подача S_{pj} у всех инструментов блока $j \in J_p$ остается неизменной (от детали к детали), а у каждого инструмента $u \in U_{pj}$ неизменной остается скорость резания v_{pju} , в то время как параметры технологических переходов (длина и глубина резания, обрабатываемый материал и др.) могут быть различны для разных деталей;

– блок $j \in J_p$ при обработке детали $d \in \mathbf{M}$ имеет соответствующую длину L_{pjd} рабочего хода;

– используется одно из следующих взаимоисключающих правил смены инструмента: независимая смена инструментов после обработки каждым из них своего расчетного (исходя из его периода стойкости при принимаемых режимах обработки) числа партий деталей или независимая смена блоков инструментов после обработки лимитирующим в соответствующем блоке инструментом расчетного (исходя из тех же условий) числа партий деталей.

Искомыми в данной задаче являются вектор $S=(S_{pj} | p \in \mathbf{P}, j \in J_p)$ минутных подач всех блоков инструментов и векторы $V_{pj}=(v_{pju} | u \in U_{pj})$ скоростей резания всех инструментов каждого из блоков $j \in J_p, p \in \mathbf{P}$. Все остальные параметры группового технологического процесса предполагаются известными. В дальнейшем $V=(V_{pj} | p \in \mathbf{P}, j \in J_p)$; \mathbf{M}_{pj} – множество деталей $d \in \mathbf{M}$, обрабатываемых блоком $j \in J_p$; \mathbf{M}_{pju} – множество деталей $d \in \mathbf{M}_{pj}$, обрабатываемых инструментом $u \in U_{pj}$.

Затраты на обработку группы деталей складываются из стоимостей работ, выполняемых на всех n тактах цикла. Общая стоимость работы на отдельном такте складывается из двух компонентов, первый из которых пропорционален длительности такта (максимальная из длительностей работы всех блоков в этом такте), второй состоит из доли затрат на инструменты, приходящейся на одну группу деталей и зависящей как от принимаемых режимов обработки (S, V) , так и деталей, обрабатываемых в этом такте на каждой из позиций. Аналогична и структура общих затрат времени на выполнение отдельного такта.

Пусть $\mathbf{S}_{pj}=[S_{1pj}, S_{2pj}]$, $[v_{1pju}, v_{2pju}]$ и $[s_{1pju}, s_{2pju}]$ – заданные диапазоны возможных значений искомых параметров S_{pj} , v_{pju} и их отношения S_{pj}/v_{pju} , учитывающего допустимые подачи на оборот, соответственно. Искомые режимы выбираются из этих диапазонов с учетом как требуемой производительности (которая определяет максимально допустимую суммарную длительность T^0 всех n тактов цикла), так и ряда технических ограничений на предельно допустимые по каждому инструменту и детали значения \bar{R}_{pjuk} характеристик процесса обработки (усилия, мощности резания, шероховатости, температуры в зоне резания и др.), зависящих от принимаемых режимов, $k = 1, 2, \dots$. Предполагается, что принятые режимы остаются неизменными в течение всего процесса обработки.

Следуя, в частности, [13], будем предполагать, что зависимости от режимов обработки (минутной подачи S и скорости резания v) периодов стойкости T инструментов и других характеристик R_k процесса для различных инструментов и деталей с достаточной точностью представляются функциями вида (индексы p, j, u и d опущены)

$$T(S, v) = \min\{C / (S^{\eta_w} v^{\mu_w} + G_w) | w = 1, \dots, \bar{w}\}; \quad (12)$$

$$R_k(S, v) = C_k S^{\alpha_k} v^{\beta_k}, \quad k=1, 2, \dots, \quad (13)$$

параметры которых можно считать известными для всех $p \in \mathbf{P}, j \in J_p, u \in U_{pj}$ и $d \in \mathbf{M}_{pju}$. В реальных ситуациях, как правило, $\eta_w, \mu_w > 1$ для всех инструментов и деталей. Таким образом, для любого фиксированного значения $S_{pj} \in \mathbf{S}_{pj}$ и любого инструмента $u \in U_{pj}$ оптимальным является наименьшее значение $v_{pju} \in [v_{1pju}, v_{2pju}] \cap [S_{pj}/s_{2pju}, S_{pj}/s_{1pju}]$, удовлетворяющее техническим ограничениям $R_{pjuk}(S_{pj}, v_{pju}) \leq \bar{R}_{pjuk}, k = 1, 2, \dots$, для всех деталей $d \in \mathbf{M}_{pju}$, поскольку это значение v_{pju} (обозначим его $v_{pju}(S_{pj})$) обеспечивает наибольшую стойкость инструмента для каждой детали $d \in \mathbf{M}_{pju}$, а следовательно, и его общую стойкость по всем деталям из \mathbf{M}_{pju} . Таким образом, опре-

деление оптимальных (т. е. минимизирующих общие затраты на обработку группы деталей при обеспечении требуемой производительности) режимов обработки, по существу, сводится к нахождению соответствующего вектора $S \in \mathbf{S} = \prod_{p \in \mathbf{P}, j \in J_p} \underline{\mathbf{S}}_{pj}$, где $\underline{\mathbf{S}}_{pj} = [\underline{S}_{1pj}, \underline{S}_{2pj}]$ – часть отрезка \mathbf{S}_{pj} ,

для каждой точки которого существует $v_{pju}(S_{pj})$.

Соотношения (12) позволяют (см., в частности, [6] и др.) определить расчетное число $D_{pju}(S_{pj})$ групп деталей, которое может обработать инструмент $u \in U_{pj}$ за период его общей стойкости при минутной подаче $S_{pj} \in \underline{\mathbf{S}}_{pj}$ и скорости резания $v_{pju}(S_{pj})$, а также расчетное число $\underline{D}_{pj}(S_{pj}) = \min\{D_{pju}(S_{pj}) | u \in U_{pj}\}$ групп деталей, которое могут обработать инструменты блока $j \in J_p$ при их одновременной смене исходя из лимитирующего по стойкости инструмента. Таким образом, для каждого из рассматриваемых правил смены инструментов и каждого $p \in \mathbf{P}, j \in J_p$ могут быть определены отнесенные к одной группе деталей суммарные материальные ($f_{pj}(S_{pj})$) и временные ($\varphi_{pj}(S_{pj})$) затраты, связанные со сменой инструментов блока j при их работе на минутной подаче $S_{pj} \in \underline{\mathbf{S}}_{pj}$ и оптимальных скоростях резания $v_{pju}(S_{pj}), u \in U_{pj}$.

При формулировке данной задачи оптимизации режимов групповой обработки на многопозиционном оборудовании в терминах рассмотренной в предыдущих разделах задачи оптимизации интенсивностей последовательно-параллельного выполнения пересекающихся множеств операций можно воспользоваться следующей интерпретацией:

множество работ – множество тактов \mathbf{I} , в каждом из которых одновременно на каждой из позиций обрабатывается соответствующая (такту и позиции) деталь;

операция pj работы $i \in \mathbf{I}$ – обработка в такте i на позиции $p \in \mathbf{P}$ детали d_{pi} блоком $j \in J_{pd_{pi}} \subseteq J_p$.

Таким образом, множество операций определяется множеством $J = \bigcup_{p \in \mathbf{P}} J_p$ всех блоков,

причем обработка различных деталей одним и тем же блоком рассматривается как одна операция, но выполняемая при разных работах;

объем операции pj в работе i – длина рабочего хода $L_{pjd_{pi}}$ блока $j \in J_{pd_{pi}}$ при обработке детали d_{pi} ;

ресурсы, расходуемые на выполнение операции pj , – каждый из инструментов блока $j \in J_p$ в отдельности (при независимой смене инструментов) или все инструменты этого блока в совокупности как один ресурс (при одновременной смене всех инструментов блока);

искомая интенсивность операции pj – величина $s_{pj} = 1/S_{pj}$, однозначно определяющая искомую минутную подачу S_{pj} блока pj ;

зависимости суммарных (по всем работам из \mathbf{I}) материальных и временных затрат на выполнение операции pj от ее интенсивности s_{pj} – функции $f_{pj}(s_{pj}) = f_{pj}(1/S_{pj})$ и $\varphi_{pj}(s_{pj}) = \varphi_{pj}(1/S_{pj})$ соответственно;

параметры E_{1i} и E_{2i} – соответственно стоимость эксплуатации оборудования и коэффициент дополнительных затрат времени на его обслуживание, отнесенные к единице времени работы на такте i .

Положим $\mathbf{s}_{pj} = [1/S_{2pj}, 1/S_{1pj}]$, $\mathbf{s} = (s_{pj} \in \mathbf{s}_{pj} | p \in \mathbf{P}, j \in J_p)$. Для простоты изложения предположим, что в множестве \mathbf{I} идентичных тактов не содержится, т. е. $n_i = 1$ для всех $i \in \mathbf{I} = \mathbf{I}$, а также уникален каждый блок $j \in J_p$, т. е. $m_{ipj} = 1$ для всех $p \in \mathbf{P}, j \in J_p, i \in \mathbf{I}$. Тогда рассматриваемая технологическая задача оптимизации режимов обработки сводится к следующей задаче математического программирования:

$$F_1(\mathbf{s}) = \sum_{i \in \mathbf{I}} E_{1i} \max_{p \in \mathbf{P}} \max_{j \in J_{pd_{pi}}} (L_{pjd_{pi}} s_{pj}) + \sum_{p \in \mathbf{P}} \sum_{j \in J_{pd_{pi}}} f_{pj}(s_{pj}) \rightarrow \min; \quad (14)$$

$$F_2(\mathbf{s}) = \sum_{i \in \mathbf{I}} E_{2i} \max_{p \in \mathbf{P}} \max_{j \in J_{pd_{pi}}} (L_{pjd_{pi}} s_{pj}) + \sum_{p \in \mathbf{P}} \sum_{j \in J_{pd_{pi}}} \varphi_{pj}(s_{pj}) \leq T_0; \quad (15)$$

$$s_{pj} \in \mathbf{s}_{pj}, p \in \mathbf{P}, j \in J_p. \quad (16)$$

Можно показать, что при принятых предположениях функции $f_{pj}(s_{pj})$ и $\varphi_{pj}(s_{pj})$ являются выпуклыми на отрезках s_{pj} для всех $p \in \mathbf{P}$ и $j \in J_p$. Таким образом, для решения задачи (14)–(16) может быть использован предложенный в разд. 2 подход.

Заключение

Предложены математическая модель и метод решения задачи оптимизации интенсивностей последовательно-параллельного выполнения комплекса пересекающихся множеств операций для частного случая, когда интенсивность любой операции остается неизменной в составе любого множества, а зависимости материальных и временных затрат на выполнение любой операции в составе конкретного множества от ее интенсивности представимы выпуклыми функциями. Метод основан на аппроксимации исходной задачи задачей линейного программирования с последующим использованием для решения последней существующих пакетов решения задач этого класса.

В качестве перспективных направлений развития данных исследований можно отметить разработки:

- итерационной схемы решения исходной задачи, предполагающей последовательное уточнение аппроксимационных соотношений (4), (5) по результатам ранее полученных решений;
- эффективных декомпозиционных методов решения большеразмерных задач (6)–(11), учитывающих специфическую структуру их матрицы ограничений;
- эффективных приближенных и точных методов решения исходной задачи с невыпуклыми функциями зависимости материальных и временных затрат на операции от интенсивности их выполнения.

Данная работа была выполнена при поддержке Белорусского республиканского фонда фундаментальных исследований (проект Ф12ФП-001).

Список литературы

1. Alting, L. Computer Aided Process Planning: the state-of-the-art survey / L. Alting, H. Zhang // Intern. J. Prod. Res. – 1989. – Vol. 27, no. 4. – P. 553–585.
2. Halevi, G. Process and Operation Panning / G. Halevi. – Springer, 2003. – 335 p.
3. Bukchin, J. Design of flexible assembly line to minimize equipment cost / J. Bukchin, M. Tzur // IIE Transactions. – 2000. – Vol. 32. – P. 585–598.
4. Gupta, A.K. Optimization of due-date objectives in scheduling semiconductor batch manufacturing / A.K. Gupta, A.I. Sivakumar // Intern. J. of Machine Tools and Manufacture. – 2006. – Vol. 46. – P. 1671–1679.
5. Burkov, V.N. Models and methods of multiprojects' management / V.N. Burkov, D.A. Novikov // Systems Science. – 1999. – Vol. 256, no. 2. – P. 5–14.
6. Dolgui, A. Graph approach for optimal design of transfer machine with rotary table / A. Dolgui, N. Guschinsky, G. Levin // Intern. J. of Prod. Res. – 2009. – Vol. 47, no. 2. – P. 321–341.
7. Dolgui, A. Enhanced mixed integer programming model for a transfer line design problem / A. Dolgui, N. Guschinsky, G. Levin // Computers and Industrial Engineering. – 2012. – Vol. 62, no. 2. – P. 570–578.
8. Левин, Г.М. Оптимизация режимов параллельной многоинструментальной обработки деталей на агрегатном оборудовании с учетом групповой смены инструментов / Г.М. Левин, Б.М. Розин // Информатика. – 2011. – № 3. – С. 33–47.
9. Левин, Г.М. Оптимизация последовательно-параллельного выполнения комплекса взаимосвязанных операций / Г.М. Левин, Б.М. Розин // Весці НАН Беларусі. Сер. фіз.-мат. навук. – 2013. – № 1. – С. 111–116.
10. Levin, G. Optimization of Multi-tool Cutting Modes for Batch Manufacturing in Large Series Machining Environment / G. Levin, B. Rozin, A. Dolgui // Proc. of the 14th IFAC Symposium on Information Control Problems in Manufacturing (INCOM'12). – Bucharest, Romania, 2012. – P. 444–448.

11. Rozin, B. Optimization of Multi-tool Cutting Modes in Multi-item Batch Manufacturing System / B. Rozin, G. Levin, A. Dolgui // Proc. of the IFAC Conf. on Manufacturing Modelling, Management and Control (MIM'2013). – SPb., 2013. – P. 766–771.

12. Долгий, А.Б. Оптимизация интенсивностей последовательно-параллельного выполнения пересекающихся множеств операций / А.Б. Долгий, Г.М. Левин, Б.М. Розин // Междунар. конгресс по информатике: информационные системы и технологии (CSIST'2013). – Минск, 2013. – С. 520–524.

13. Режимы резания металлов : справочник / Ю.В. Барановский [и др.]. – 3-е изд., перераб. и доп. – М. : Машиностроение, 1972. – 408 с.

Поступила 05.05.2014

¹Объединенный институт проблем информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: {levin; rozin}@newman.bas-net.by

²Ecole Nationale Supérieure des Mines,
158, Cours Fauriel, 42 023 Saint-Etienne, France
e-mail: dolgui@emse.fr

G.M. Levin, B.M. Rozin, A.B. Dolgui

**LINEAR APPROXIMATION FOR INTENSITIES OPTIMIZATION PROBLEM
OF SEQUENTIAL-PARALLEL EXECUTION OF INTERSECTING
OPERATION SETS**

The mathematical model and method for the problem of optimization of intensities of sequential-parallel execution of intersecting operation sets are proposed. The proposed method is based on the approximation of the problem by linear programming problem.

УДК 519.6; 621.382

Г.М. Заяц¹, Ф.Ф. Комаров², А.Ф. Комаров²

ЧИСЛЕННОЕ ИССЛЕДОВАНИЕ ВЛИЯНИЯ РАДИАЦИОННЫХ ИЗЛУЧЕНИЙ НА ПАРАМЕТРЫ МДП-ПРИБОРОВ

Рассматривается модель, описывающая пространственно-временную эволюцию заряда, который возникает в диэлектрике структуры металл-диэлектрик-полупроводник (МДП) при воздействии на нее ионизирующих излучений – рентгеновских и гамма-квантов. Система уравнений решается численным методом. Для реализации разностной задачи разрабатывается итерационная процедура. Представляются результаты численного моделирования.

Введение

В настоящее время в военной и космической технике широко применяются приборы, содержащие интегральные схемы различной степени интеграции. Эксплуатация таких устройств осуществляется под воздействием низкоинтенсивного ионизирующего излучения, приводящего к радиационной деградации микросхем и выходу из строя электронной аппаратуры. В этой связи создание радиационно-стойких интегральных схем является одной из актуальных задач электроники. Важнейшим инструментом при этом служит математическое моделирование, предполагающее разработку адекватных диффузионно-кинетических моделей и качественных вычислительных алгоритмов, описывающих процесс воздействия ионизирующих излучений, в частности рентгеновских и гамма-квантов, на рабочие характеристики полупроводниковых приборов.

В данной работе рассматривается МДП-структура с двумя типами ловушечных уровней, учитывающими наличие дефектов в окисле и радиационно-индуцированные поверхностные состояния. Строится физико-математическая модель процессов накопления индуцированного ионизирующей радиацией заряда в диэлектрике структуры SiO₂/Si и поверхностных состояний при облучении гамма-квантами МДП-структуры и последующей релаксации данного заряда за счет туннельной разрядки. Модель представляет собой систему уравнений [1, 2], в которую входят уравнения в частных производных для описания кинетики свободных электронов и дырок; обыкновенные дифференциальные уравнения, отражающие кинетику накопленного дырочного заряда на ловушечных уровнях; уравнение Пуассона, решение которого позволяет определить напряженность электрического поля в диэлектрике. Туннельный механизм разрядки накопленного заряда в слое диэлектрика моделируется обыкновенным дифференциальным уравнением [3]. Распределение потенциала в МДП-структуре при наличии заряда в диэлектрике и на поверхностных состояниях задается уравнением электронейтральности.

Для решения задачи предложен численный метод, основанный на разностном методе [4]. Разработаны алгоритм и итерационная процедура для реализации разностной задачи и проведено численное моделирование радиационно-индуцированного изменения порогового напряжения в зависимости от дозы радиации, распределения свободных и захваченных на ловушечные уровни зарядов в диэлектрике, распределения напряженности электрического поля МДП-структуры в процессе облучения.

1. Физико-математическая модель

В области $\Omega = \{0 < x < d, 0 < t \leq t_f\}$ рассмотрим систему уравнений, описывающую пространственно-временную эволюцию заряда, который возникает в диэлектрике толщиной d МДП-структуры при воздействии на нее ионизирующего излучения, с учетом туннельной разрядки [1–3]:

$$\frac{\partial n}{\partial t} = D_n \frac{\partial^2 n}{\partial x^2} + \mu_n \frac{\partial(n \cdot E)}{\partial x} - R_{n1}(n, E, P_{t1}) - R_{n2}(n, E, P_{t2}) + G(E); \quad (1)$$

$$\frac{\partial p}{\partial t} = D_p \frac{\partial^2 p}{\partial x^2} + \mu_p \frac{\partial(p \cdot E)}{\partial x} - R_{p1}(p, E, P_{t1}) - R_{p2}(p, E, P_{t2}) + G(E); \quad (2)$$

$$\frac{\partial P_{t1}}{\partial t} = R_{p1}(p, E, P_{t1}) - R_{n1}(n, E, P_{t1}), \quad \frac{\partial P_{t2}}{\partial t} = R_{p2}(p, E, P_{t2}) - R_{n2}(n, E, P_{t2}); \quad (3)$$

$$\frac{\partial E}{\partial x} = \frac{q}{\varepsilon_{0x} \varepsilon_0} (P_{t1} + P_{t2} + p - n); \quad (4)$$

$$\frac{\partial P_t}{\partial t} = -\alpha_1 \exp(-\alpha_2 x) P_t; \quad (5)$$

$$V_G = \varphi_{ms} + \psi - \left(\frac{Q_{0t}}{C_{ox}} + \frac{Q_{sc}(\psi)}{C_{ox}} + \frac{Q_{ss}(\psi)}{C_{ox}} \right). \quad (6)$$

На границе области моделирования выполняются условия

$$n(0, t) = n(d, t) = 0, \quad p(0, t) = p(d, t) = 0, \quad 0 < t \leq t_f; \quad Q_{0t}(0) = Q_{ss}(0) = 0; \quad (7)$$

$$n(x, 0) = p(x, 0) = P_{t1}(x, 0) = P_{t2}(x, 0) = 0, \quad E(x, 0) = f(\psi(0)), \quad 0 \leq x \leq d.$$

В системе уравнений (1)–(7) d – толщина диэлектрика; t_f – время моделирования; n, p – концентрации свободных электронов и дырок; E – напряженность электрического поля в диэлектрике; $P_{t1,2}$ – концентрации дырочного заряда, захваченного на «мелких» ловушках (поверхностные состояния в переходных слоях вблизи границы с полупроводником и поликремниевым затвором, индекс $t1$) и «глубоких» ловушках (расположенных в объеме диэлектрика, индекс $t2$); D_n, D_p – коэффициенты диффузии соответственно электронов и дырок; μ_n, μ_p – подвижности электронов и дырок; G – скорость генерации ионизирующим излучением электронно-дырочных пар; $R_{n1,2}, R_{p1,2}$ – скорости захвата электронов и дырок на ловушечные уровни; q – заряд электрона; ε – абсолютная диэлектрическая проницаемость диэлектрика; V_G – напряжение на затворе МДП-структуры; φ_{ms} – разность работ выхода материала затвора и полупроводника; ψ – поверхностный потенциал полупроводника; Q_{0t} – эффективный заряд в диэлектрике, захваченный на ловушечные уровни; Q_{sc} – заряд области пространственного заряда полупроводника; Q_{ss} – заряд на поверхностных состояниях; C_{ox} – емкость слоя диэлектрика; α_1, α_2 – частотный и барьерный факторы.

При моделировании принимаем во внимание следующие зависимости. Скорости захвата электронов и дырок на ловушечные уровни задаются выражениями [1]

$$R_{n1} = nP_{t1}\sigma_n(E)(\mu_n|E| + v_{th}), \quad R_{n2} = nP_{t2}\sigma_n(E)(\mu_n|E| + v_{th}),$$

$$R_{p1} = p(N_{t1} - P_{t1})\sigma_p(E)\left(\mu_p|E| + \frac{\mu_p}{\mu_n}v_{th}\right), \quad R_{p2} = p(N_{t2} - P_{t2})\sigma_p(E)\left(\mu_p|E| + \frac{\mu_p}{\mu_n}v_{th}\right), \quad (8)$$

где $N_{t1,2}(x)$ – концентрации ловушечных уровней (центров дырочного захвата); v_{th} – тепловая скорость носителей заряда; $\sigma_p(E)$ и $\sigma_n(E)$ – полевые зависимости сечения захвата дырки и электрона.

Скорость генерации электронно-дырочных пар ионизирующей радиацией $G(E)$ определяется мощностью радиационной дозы $\dot{D} = dD/dt$, коэффициентом генерации пар носителей k_g и вероятностью $f_y^{x-ray}(E)$ разделения этих пар электрическим полем до их рекомбинации [5, 6]:

$$G(E) = \dot{D} k_g f_y^{x-ray}(E).$$

Величины зарядов в окисле и на поверхностных состояниях описываются соответственно выражениями

$$Q_{0t} = \frac{1}{d} \int_0^d (d-x) \rho_t(x) dx \text{ и } Q_{ss} = q N_{ss} (\Phi_0 - \Psi), \quad (9)$$

где $\rho_t(x)$ – распределение накопленного дырочного заряда на ловушечных уровнях;

$N_{ss} = k_D \frac{Q_{0t}}{q\Phi_0}$ – усредненная по ширине запрещенной зоны полупроводника плотность поверхностного состояния [7, 8] (k_D – экспериментально определяемый параметр).

Заряд области пространственного заряда полупроводника рассчитывается по формуле [9]

$$Q_{sc}(\Psi) = \varepsilon_s \varepsilon_0 E_s = \pm \frac{\sqrt{2} \varepsilon_s \varepsilon_0 kT}{q L_D} F(\Psi, \Phi_0). \quad (10)$$

2. Постановка задачи

С учетом (8) систему уравнений (1)–(10) перепишем в удобной для построения алгоритма решения форме:

$$\frac{\partial n}{\partial t} = D_n \frac{\partial^2 n}{\partial x^2} + \mu_n \frac{\partial(n \cdot E)}{\partial x} - n Q_1(E, P_{t1}, P_{t2}) + G(E); \quad (11)$$

$$\frac{\partial p}{\partial t} = D_p \frac{\partial^2 p}{\partial x^2} + \mu_p \frac{\partial(p \cdot E)}{\partial x} - p Q_2(E, P_{t1}, P_{t2}) + G(E); \quad (12)$$

$$\frac{\partial P_{t1}}{\partial t} = -P_{t1} S_1(p, n, E) + S_2(p, E) N_{t1}, \quad \frac{\partial P_{t2}}{\partial t} = -P_{t2} S_1(p, n, E) + S_2(p, E) N_{t2}; \quad (13)$$

$$\frac{\partial E}{\partial x} = \frac{q}{\varepsilon_{0x} \varepsilon_0} (P_{t1} + P_{t2} + p - n); \quad (14)$$

$$\frac{\partial P_t}{\partial t} = -P_t S_3(E); \quad (15)$$

$$-V_G + \varphi_{ms} + \Psi - \left(\frac{Q_{0t}}{C_{ox}} + \frac{Q_{sc}(\Psi)}{C_{ox}} + \frac{Q_{ss}(\Psi)}{C_{ox}} \right) = 0, \quad (16)$$

где $Q_1(E, P_{t1}, P_{t2}) = (P_{t1} + P_{t2}) \sigma_n(E) (\mu_n |E| + v_{th})$, $Q_2(E, P_{t1}, P_{t2}) = (N_{t1} - P_{t1} + N_{t2} - P_{t2}) \sigma_p(E) \times$

$$\times \left(\mu_p |E| + \frac{\mu_p}{\mu_n} v_{th} \right), \quad S_1(p, n, E) = p \sigma_p(E) \left(\mu_p |E| + \frac{\mu_p}{\mu_n} v_{th} \right) + n \sigma_n(\mu_n |E| + v_{th}),$$

$$S_2(p, E) = p\sigma_p(E) \left(\mu_p |E| + \frac{\mu_p}{\mu_n} v_{th} \right), \quad S_3(E) = \alpha_1(E) \exp(-\alpha_2 x).$$

3. Алгоритм решения

Для задачи моделирования пространственно-временной эволюции заряда в диэлектрике (11)–(16) разработан итерационный алгоритм. На каждом временном шаге $t = t_j$ осуществляются итерации, состоящие из следующих этапов:

1. Решается уравнение непрерывности (11) с соответствующими граничными и начальными условиями для нахождения концентрации свободных электронов $n(x, t_j)$.

2. Находится концентрация дырок $p(x, t_j)$ из уравнения непрерывности (12) с граничными и начальными условиями с учетом этапа 1.

3. Решаются уравнения (13), отражающие кинетику накопленного дырочного заряда на ловушечных уровнях $P_{t1}(x, t_j)$, $P_{t2}(x, t_j)$, с учетом результатов этапов 1 и 2.

4. Решается уравнение (15) для учета процесса туннельной разрядки накопленного заряда в слое диэлектрика.

5. Найденные на этапах 1–4 распределения свободных носителей $n(x, t_j)$, $p(x, t_j)$ и захваченного заряда $P_{t1}(x, t_j)$, $P_{t2}(x, t_j)$ подставляются в уравнение Пуассона (14), решение которого позволяет определить распределение напряженности электрического поля в диэлектрике $E(x, t_j)$.

6. С учетом этапов 1–5 из уравнения электронейтральности (16), характеризующего электрофизические свойства МДП-структуры в целом, находится поверхностный потенциал полупроводника $\psi(t_j)$.

7. Окончание итераций на временном шаге $t = t_j$ определяется выполнением условия $\left| \psi^{s+1} - \psi^s \right| \leq \varepsilon_1 \left| \psi^s \right| + \varepsilon_2$, где $s = 0, 1, 2, \dots$ – номер итерации; $\varepsilon_1, \varepsilon_2$ – некоторые эмпирические константы ($\varepsilon_1 = 0$, если $\left| \psi^s \right| < 1$, и $\varepsilon_2 = 0$, если $\left| \psi^s \right| > 1$).

Алгоритм реализован численно. Разработан численный метод, основанный на разностном методе [4].

4. Разностные схемы

В области моделирования Ω введем неравномерные расчетные сетки по пространственной переменной x :

$$\omega_h = \{x_i \in [0, d], x_i = x_0 + \sum_{k=1}^i h_k, i = 1, 2, \dots, N, x_0 = 0, x_N = d, h_k = x_k - x_{k-1}\}$$

и временной переменной t :

$$\omega_\tau = \{t_0 = 0, t_j = \sum_{k=1}^j \tau_k, j = 1, 2, \dots, j_0, \tau_k > 0\}.$$

Пусть y_1, y_2, y_3, y_4, y_6 ($y_k = y_{k,i}^j = y_k(x_i, t_j)$, $k = 1, 2, 3, 4, 6$, $i = \overline{0, N}$, $j = \overline{0, j_0}$) – сеточные функции, определенные в узлах сетки $\omega = \omega_h \times \omega_\tau$ и соответствующие функциям

$n(x,t)$, $p(x,t)$, $P_{i1}(x,t)$, $P_{i2}(x,t)$, $E(x,t)$; $y_5 = y_5^j = y_5(t_j)$ ($j = \overline{0, j_0}$) – сеточная функция, заданная на ω_τ и соответствующая функции $\psi(t)$.

Для аппроксимации уравнений (11), (12) с соответствующими краевыми условиями используем неявную разностную схему

$$\begin{aligned} \frac{y_i^j - y_i^{j-1}}{\tau_j} = D \frac{1}{h_i} \left(\frac{y_{i+1}^j - y_i^j}{h_{i+1}} - \frac{y_i^j - y_{i-1}^j}{h_i} \right) + \\ + 0,5\mu \left(\frac{y_{6,i+1}^j y_{i+1}^j - y_{6,i}^j y_i^j}{h_{i+1}} + \frac{y_{6,i-1}^j y_{i-1}^j - y_{6,i}^j y_{k,i}^j}{h_i} \right) - y_i^j Q_i^j + G_i^j, \quad i = \overline{1, N-1}, \\ y_0^j = 0, \quad y_N^j = 0, \quad j = \overline{1, j_0}, \end{aligned} \quad (17)$$

где $G_i^j = G(x_i, t_j)$, $h_i = 1/2(h_{i+1} + h_i)$; $y_i^j = y_{1,i}^j$, $D = D_n$, $\mu = \mu_n$, $Q_i^j = (y_{3,i}^j + y_{4,i}^j) \sigma_n(y_{6,i}^j) (\mu_n |y_{6,i}^j| + v_{th})$ при аппроксимации уравнения (11); $y_i^j = y_{2,i}^j$, $D = D_p$, $\mu = \mu_p$, $Q_i^j = (N_{i1} + N_{i2} - y_{3,i}^j - y_{4,i}^j) \sigma_p(y_{6,i}^j) \left(\mu_p |y_{6,i}^j| + \frac{\mu_p}{\mu_n} v_{th} \right)$ в случае аппроксимации уравнения (12).

Дифференциальные уравнения (13), (15) аппроксимируем неявной разностной схемой

$$\frac{y_i^j - y_i^{j-1}}{\tau_j} = -y_i^j S_{1,i}^j + S_{2,i}^j N_i, \quad j = \overline{1, j_0}, \quad i = \overline{0, N}, \quad (18)$$

где при аппроксимации уравнений (13) $S_{1,i}^j = S_{1,i}^j(y_{1,i}^j, y_{2,i}^j, y_{6,i}^j) = y_{2,i}^j \sigma_p(y_{6,i}^j) \left(\mu_p |y_{6,i}^j| + \frac{\mu_p}{\mu_n} v_{th} \right) + y_{1,i}^j \sigma_n(y_{6,i}^j) (\mu_n |y_{6,i}^j| + v_{th})$, $S_{2,i}^j = S_{2,i}^j(y_{2,i}^j, y_{6,i}^j) = y_{2,i}^j \sigma_p(y_{6,i}^j) \left(\mu_p |y_{6,i}^j| + \frac{\mu_p}{\mu_n} v_{th} \right)$, $N_i = N_{i1}$ для первого уравнения из (13) и $N_i = N_{i2}$ для второго уравнения; при аппроксимации уравнения (15) $S_{1,i}^j = S_{3,i}^j(y_{6,i}^j) = \alpha_1(y_{6,i}^j) \exp(-\alpha_2 x_i)$, $S_{2,i}^j = 0$.

Дифференциальную задачу для уравнения Пуассона (14) заменяем разностной задачей

$$\frac{y_{6,i+1}^j - y_{6,i}^j}{h_{i+1}} = \frac{q}{\varepsilon_{0x} \varepsilon_0} (y_{4,i}^j + y_{3,i}^j + y_{2,i}^j - y_{1,i}^j), \quad i = \overline{N-1, 0}, \quad y_{6,N}^j = E_N^j(y_5^j) = E(y_5^j), \quad j = \overline{1, j_0}. \quad (19)$$

Уравнение электронейтральности аппроксимируем уравнением

$$-V_G + \Phi_{ms} + y_5^j - \left(\frac{Q_{0t}}{C_{ox}} + \frac{Q_{sc}(y_5^j)}{C_{ox}} + \frac{Q_{ss}(y_5^j)}{C_{ox}} \right) = 0. \quad (20)$$

5. Численное решение

В соответствии с алгоритмом, представленным в разд. 3, построен итерационный численный алгоритм для реализации разностной задачи (17)–(20), состоящий из следующих шагов.

В начальный момент времени $t = t_0 = 0$:

Шаг 1°. С учетом начальных условий (7) методом деления отрезка пополам решаем уравнение электронейтральности (20) и находим значение сеточной функции $y_5^0 = y_5(0)$.

Шаг 2°. Из соотношения $y_{6,N}^0 = \pm \frac{Q_{sc}(y_5^0)}{\epsilon_{ox}\epsilon_o}$ находим граничное условие для уравнения Пуассона в диэлектрике, решаем разностную задачу (19) и находим значения сеточной функции $y_{6,i}^0, i = \overline{0, N}$.

В момент времени $t = t_j (j = \overline{1, j_0})$:

Шаг 1. Полагаем $s = 0$ (s – номер итерации). В качестве значений искомых функций на нулевой итерации ($s = 0$) выбираем значения данных функций, полученные на предыдущем временном слое: $y_{k,i}^{j,s=0} = y_{k,i}^{j-1} (i = \overline{0, N}, k = 1, 2, 3, 4, 6), y_{5,j}^{s=0} = y_5^{j-1}, j = \overline{2, j_0}$. При $j = 1$ – это начальные условия задачи: $y_{k,i}^{1,s=0} = y_{k,i}^0 (i = \overline{0, N}, k = 1, 2, 3, 4, 6), y_{5,j}^{1,s=0} = y_5^0$.

Шаг 2. Находим значения сеточных функций $y_{1,i}^{j,s+1}$ и $y_{2,i}^{j,s+1} (i = \overline{0, N})$ на $(s+1)$ -й итерации, решая разностные задачи

$$\frac{y_{k,i}^{j,s+1} - y_{k,i}^{j-1}}{\tau_j} = D_k \frac{1}{h_i} \left(\frac{y_{k,i+1}^{j,s+1} - y_{k,i}^{j,s+1}}{h_{i+1}} - \frac{y_{k,i}^{j,s+1} - y_{k,i-1}^{j,s+1}}{h_i} \right) +$$

$$+ 0,5\mu_k \left(\frac{y_{6,i+1}^{j,s} y_{k,i+1}^{j,s+1} - y_{6,i}^{j,s} y_{k,i}^{j,s+1}}{h_{i+1}} + \frac{y_{6,i-1}^{j,s} y_{k,i-1}^{j,s+1} - y_{6,i}^{j,s} y_{k,i}^{j,s+1}}{h_i} \right) - y_{k,i}^{j,s} Q_{k,i}^j + G_i^j, i = 1, \dots, N-1,$$

$$y_{k,0}^{j,s+1} = 0, y_{k,N}^{j,s+1} = 0, k = 1, 2,$$

где

$$Q_{1,i}^j = \left(y_{3,i}^j + y_{4,i}^j \right) \sigma_n (y_{6,i}^j) \left(\mu_n |y_{6,i}^j| + v_{th} \right), Q_{2,i}^j = \left(N_{t1} + N_{t2} - y_{3,i}^j - y_{4,i}^j \right) \sigma_p (y_{6,i}^j) \left(\mu_p |y_{6,i}^j| + \frac{\mu_p}{\mu_n} v_{th} \right).$$

Для фиксированного $j (j = \overline{1, j_0})$ система алгебраических уравнений (21) решается методом прогонки [4].

Шаг 3. Используя рассчитанные на шаге 2 значения $y_{1,i}^{j,s+1}$ и $y_{2,i}^{j,s+1} (i = \overline{0, N})$, из разностной задачи (18) определяем функции $y_{3,i}^{j,s+1}$ и $y_{4,i}^{j,s+1} (i = \overline{0, N})$:

$$y_{k,i}^{j,s+1} = \frac{y_{k,i}^{j-1} + \tau_j S_{2,i}^j (y_{2,i}^{j,s+1}, y_{6,i}^j) N_{tk-2}}{1 + \tau_j S_{1,i}^j (y_{1,i}^{j,s+1}, y_{2,i}^{j,s+1}, y_{6,i}^j)}, j = \overline{1, j_0}, y_{k,i}^{0,s+1} = 0, i = \overline{0, N}, k = 3, 4. \quad (22)$$

Шаг 4. Для учета процесса туннельной разрядки полагаем $z_{3,i}^{j-1} = y_{3,i}^{j,s+1}, z_{4,i}^{j-1} = y_{4,i}^{j,s+1}$ и, используя схему (18), определяем «уточненные» функции $y_{3,i}^{j,s+1}$ и $y_{4,i}^{j,s+1} (i = \overline{0, N})$:

$$y_{k,i}^{j,s+1} = \frac{z_{k,i}^{j-1}}{1 + \tau_j S_{3,i}^j (y_{6,i}^j)}, j = \overline{1, j_0}, y_{k,i}^{0,s+1} = 0, i = \overline{0, N}, k = 3, 4. \quad (23)$$

Шаг 5. Учитывая полученные на шагах 2–4 значения функций $y_{k,i}^{s+1}$ ($k=1, 2, 3, 4, i=\overline{0, N}$), методом деления отрезка пополам решаем уравнение (20) и определяем значение y_5^{s+1} .

Шаг 6. Проверяем условие сходимости итераций:

$$\left| y_5^{s+1} - y_5^s \right| \leq \varepsilon_1 \left| y_5^s \right| + \varepsilon_2 \quad (\varepsilon_1 = 0, \text{ если } \left| y_5^s \right| < 1, \text{ и } \varepsilon_2 = 0, \text{ если } \left| y_5^s \right| > 1). \quad (24)$$

Шаг 7. В случае выполнения условия (24) находим граничное условие для разностной задачи (19), затем решаем задачу (19) и находим значения сеточной функции $y_{6,i}^{s+1}$, $i=\overline{0, N}$. Вычисляем характеристики МДП-структуры в момент времени $t=t_j$. Полагаем $j=j+1$ и переходим к шагу 1 для решения задачи на следующем временном слое.

Шаг 8. Если условие (24) не выполняется, полагаем $y_{k,i}^s = y_{k,i}^{s+1}$, $i=\overline{0, N}$ ($k=1, 2, 3, 4, 6$), $y_5^s = y_5^{s+1}$ ($k=5$), $s=s+1$ и переходим к шагу 2 для выполнения новой итерации.

6. Результаты численного моделирования

На основе предложенного алгоритма проведено численное моделирование радиационно-индуцированного изменения характеристик МДП-структуры в процессе ее облучения гамма-квантами в зависимости от дозы радиации. Некоторые результаты моделирования показаны на рис. 1–3. При расчетах полагали, что интегральная доза ионизирующего излучения $D = 5 \cdot 10^5 \text{P}$, мощность радиационной дозы $\dot{D} = dD/dt = 10^2 \text{P/c}$, концентрация легирующей примеси в кремнии $N_B = 10^{15} \text{см}^{-3}$, температура $T = 300^\circ \text{K}$, $\varphi_{ms} = -0,5 \text{ В}$, подвижность электронов в диэлектрике $\mu_n = 10^2 \text{см}^2 \text{В}^{-1} \text{с}^{-1}$, дырок $\mu_p = 0,6 \cdot 10^{-3} \text{см}^2 \text{В}^{-1} \text{с}^{-1}$. Коэффициент генерации электронно-дырочных пар рентгеновским излучением в SiO_2 равен $k_g = 8 \times 10^{12} \text{см}^{-3} \text{рад}^{-1} \text{ пар}$ [10], $\varepsilon_{ox} = 1,6$, $\varepsilon_s = 11,5$.

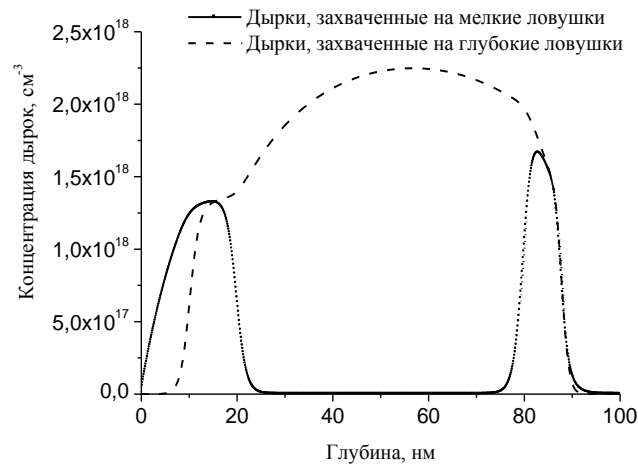


Рис. 1. Распределение дырочного заряда на ловушечных уровнях после облучения

Рис. 1 иллюстрирует распределение дырочного заряда на «мелких» и «глубоких» ловушечных уровнях в диэлектрике толщиной $d = 100 \text{ нм}$, рассчитанное в предположении, что $V_G = -0,9 \text{ В}$, $k_D = 1,03$, распределение по толщине диэлектрика «мелких» N_{l1} и «глубоких» N_{l2} ловушечных уровней имеет вид [1] $N_{l1} = 5 \times 10^{18} \left((e^{(x-20)} + 1)^{-1} + (e^{(80-x)} + 1)^{-1} \right)$, $N_{l2} = 5 \times 10^{18} \left((e^{(x-90)} + 1)^{-1} - (e^{(x-10)} + 1)^{-1} \right)$.

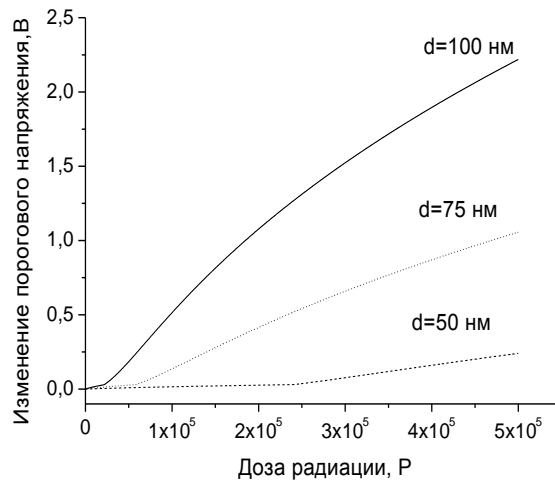


Рис. 2. Дозовая зависимость изменения порогового напряжения МДП-структуры при различной толщине диэлектрика

На рис. 2 и 3 представлены результаты моделирования в предположении, что $V_G = -1,0$ В, $k_D = 1,15$, распределение по толщине диэлектрика «мелких» N_{t1} и «глубоких» N_{t2} ловушечных уровней равномерное, $N_{t1} = 5 \cdot 10^{18} \text{ см}^{-3}$, $N_{t2} = 10^{19} \text{ см}^{-3}$. Рис. 2 иллюстрирует изменение порогового напряжения МДП-структуры в процессе облучения гамма-квантами в зависимости от толщины диэлектрика. Уменьшение результирующего изменения порогового напряжения в процессе облучения МДП-структуры с уменьшением толщины диэлектрика обусловливается уменьшением концентрации накопленного дырочного заряда в нем, а также распределением его на «мелких» и «глубоких» ловушечных уровнях данного диэлектрика. В рассматриваемом случае уменьшение толщины подзатворного окисла на 25 нм (с 75 до 50 нм) приводит к снижению величины порогового напряжения на 1,0 В при интегральной дозе облучения $D = 5 \cdot 10^5$ Р рентгеновскими квантами с энергией 20 КэВ и мощностью дозы $dD/dt = 10^2$ Р/с.

На рис. 3 изображено распределение дырочного заряда на «мелких» и «глубоких» ловушечных уровнях в диэлектрике различной толщины.

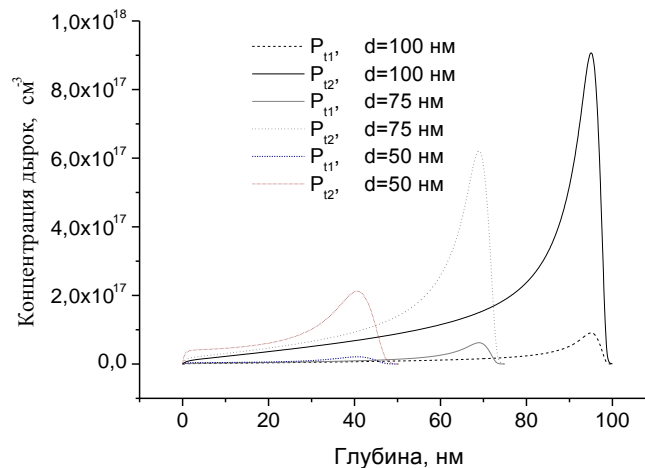


Рис. 3. Распределение дырочного заряда на ловушечных уровнях в зависимости от толщины диэлектрика

Заключение

Модель описывает процессы зарядовой деградации кремниевых структур под действием ионизирующего излучения различной интенсивности. Разработанный на основе предло-

женного численного алгоритма программный модуль позволяет находить изменение порогового напряжения МДП-структуры в зависимости от ряда параметров: концентрации ловушек в окисле, их распределения по толщине окисла, величин подвижностей и сечений захвата для электронов и дырок, а также рассчитывать распределение свободных и захваченных на ловушечные уровни зарядов в диэлектрике, распределение напряженности электрического поля, изменение поверхностного потенциала на границе диэлектрик-полупроводник в процессе облучения МДП-структуры.

Сравнение полученных результатов с результатами работ [1, 11] свидетельствует о достаточной степени адекватности выбранной модели и об эффективности разработанного численного метода.

Список литературы

1. Моделирование процессов рентгеновской корректировки пороговых напряжений МДП-интегральных схем / М.Н. Левин [и др.] // Микроэлектроника. – 2006. – Т. 35, № 5. – С. 382–391.
2. Численное моделирование влияния низкоинтенсивного ионизирующего излучения космического пространства на параметры МДП-приборов / Г.М. Заяц [и др.] // XI Белорусская математическая конф. : тез. докл. Междунар. науч. конф. Минск, 5–9 ноября 2012 г. Ч. 3. – Минск : Институт математики НАН Беларуси, 2012. – С. 42–43.
3. MCWhorter, P.J. Modelling the anneal of radiation-induced trapped holes in a varying thermal environment / P.J. MCWhorter, S.L. Miller, W.M. Miller // IEEE Trans. Nuclear Physics. – 1990. – Vol. 37, no. 6. – P. 1682–1689.
4. Самарский, А.А. Теория разностных схем / А.А. Самарский. – М. : Наука, 1989. – 616 с.
5. Ausman, G.A. Electron-hole pair creation energy in SiO / G.A. Ausman, F.B. McLean // Appl. Phys. Lett. – 1975. – No. 26. – P. 173.
6. Benedetto, J.M. The relationship between Co₆₀ and 10-keV X-ray damage in MOS devices / J.M. Benedetto, H.E. Boesch // IEEE Trans. Nuclear Science. – 1986. – NS-33, no. 6. – P. 1318–1323.
7. Релаксационные процессы в МДП-элементах интегральных схем, вызванные ионизирующим излучением и импульсным магнитным полем / А.Г. Кадменский [и др.] // Письма в ЖТФ. – 1993. – Т. 19, № 3. – С. 43–45.
8. Levin, M.N. Relaxation processes induced in Si–SiO₂ systems by ionizing radiation and pulsed magnetic field treating / M.N. Levin, V.M. Maslovsky // Solid State Communication. – 1994. – Vol. 90, no. 12. – P. 813–816.
9. Гуртов, В.А. Твердотельная электроника : учеб. пособие / В.А. Гуртов. – Петрозаводск : ПетрГУ, 2004. – 312 с.
10. Поверхностные радиационные эффекты в интегральных схемах / Сокоян А.В. [и др.] // Модель космоса : в 2 т. / Моск. гос. ун-т им. М. В. Ломоносова, Науч.-исслед. ин-т ядерной физики им. Д. В. Скобельцына. – 8-е изд. – М. : КДУ, 2007. – Т. 2. – С. 466–493.
11. Прогнозирование радиационной стойкости МОП ИС в условиях низкоинтенсивного облучения / М.Н. Левин [и др.] // Конденсированные среды и межфазные границы. – 2010. – Т. 12, № 3. – С. 226–232.

Поступила 14.03.2014

¹Институт математики НАН Беларуси,
Минск, Сурганова, 11
e-mail: zayats@im.bas-net.by

²НИИ прикладных физических проблем
им. А.Н. Севченко БГУ,
Минск, ул. Курчатова, 7
e-mail: KomarAF@bsu.by

G.M. Zayats, F.F. Komarov, A.F. Komarov

**NUMERICAL SIMULATION OF THE INFLUENCE OF RADIATION
ON THE MOS DEVICES PARAMETERS**

A model describing the space-time evolution of the charge which arises in the dielectric structure of metal-insulator-semiconductor under ionizing radiation of X-ray and gamma-rays is considered. The system of equations is solved by the numerical method. For realization of the difference problem an iterative algorithm is developed. The results of numerical modeling are presented.

УДК 004.925.8

А.П. Побегайло

МОДЕЛИРОВАНИЕ СПЛАЙН-ПОВЕРХНОСТЕЙ БЭЗЬЕ НА ЧЕТЫРЕХУГОЛЬНЫХ СЕТКАХ КРИВЫХ

Рассматривается подход к моделированию интерполяционных сплайн-поверхностей на четырехугольных сетках кривых. Клетка поверхности моделируется посредством смешивания своих границ при помощи специальных полиномов, что влечет локальную зависимость геометрических свойств поверхности от границ клеток. Если границы клетки поверхности определяются посредством кривых Бэзье, то клетка поверхности является поверхностью Бэзье. Требуемая непрерывность поверхности обеспечивается выбором полинома подходящей степени. Представленный подход предназначен для моделирования сплайн-поверхностей в таких приложениях, как компьютерная графика и геометрическое моделирование.

Введение

Моделирование сплайн-поверхностей на четырехугольных сетках кривых находит применение в различных областях прикладной математики. В научной литературе широко освещены вопросы, связанные с данной тематикой. Так, геометрические основы методов построения интерполяционных кривых и поверхностей представлены в работе [1]. Классификация алгоритмов для моделирования интерполирующих поверхностей с локальными свойствами приведена в работе [2]. Обзор методов моделирования параметрических поверхностей посредством смешивания их границ дан в работе [3]. Метод построения куска поверхности посредством билинейной интерполяции его границ предложен Кунсом в работе [4]. Современное изложение этого метода дано в работе Форреста [5] и рассмотрено также в монографии Фокса и Пратта [6]. Построение интерполяционных сплайн-поверхностей с использованием сеток кривых впервые было предложено Гордоном в работе [7]. Построение интерполирующих поверхностей посредством линейной интерполяции также рассмотрено в статье [8]. Представленный подход к моделированию сплайн-поверхностей может рассматриваться как развитие методов Кунса и Гордона.

1. Сглаживающие полиномы

Для моделирования сплайн-поверхностей с локальными свойствами будут использоваться полиномы

$$w_n(u) = \sum_{i=n}^{2n-1} b_{2n+1,i}(u), \quad u \in [0,1], \quad (1)$$

где $b_{n,m}(u)$ обозначают полиномы Бернштейна $b_{n,m}(u) = C_n^m (1-u)^{n-m} u^m$.

Из формулы (1) следует, что полиномы $w_n(u)$ имеют граничные значения

$$w_n(0) = 0, \quad w_n(1) = 1, \quad (2)$$

а их производные удовлетворяют граничным условиям

$$w_n^{(m)}(0) = w_n^{(m)}(1) = 0, \quad \forall m \in \{1, 2, \dots, n-1\}. \quad (3)$$

Из свойства полиномов Бернштейна

$$\sum_{m=0}^n b_{n,m}(u) = 1, \quad \forall u \in [0, 1],$$

следует, что

$$w_n(u) + w_n(1-u) = 1, \quad w_n(u) + w_n(1-u) = 1.$$

Кроме того, для любого значения $n \in N$ полином $w_n(u)$ является минимумом функционала

$$J_n(f) = \int_0^1 |f^{(n)}(u)|^2 du,$$

где функция $f(u)$, $u \in [0, 1]$, удовлетворяет граничным условиям (2) и (3). Подробно полиномы $w_n(u)$ и их свойства рассмотрены в статье [9].

2. Моделирование кусков поверхностей по их границам

Моделирование четырехугольных кусков поверхностей по их границам с использованием билинейной интерполяции впервые было предложено Кунсом [4]. Представляемый в данной статье подход к моделированию четырехугольных кусков поверхности по их границам может рассматриваться как обобщение этого метода. Концептуально такой подход был представлен автором на конференции [10].

Рассмотрим четыре параметризованные кривые $p_0(u)$, $p_1(u)$, $u \in [0, 1]$, и $q_0(v)$, $q_1(v)$, $v \in [0, 1]$ (рис. 1), которые имеют общие граничные точки

$$p_0(0) = q_0(0) = r_{0,0}, \quad p_0(1) = q_1(0) = r_{1,0}, \quad p_1(0) = q_0(1) = r_{0,1}, \quad p_1(1) = q_1(1) = r_{1,1}. \quad (4)$$

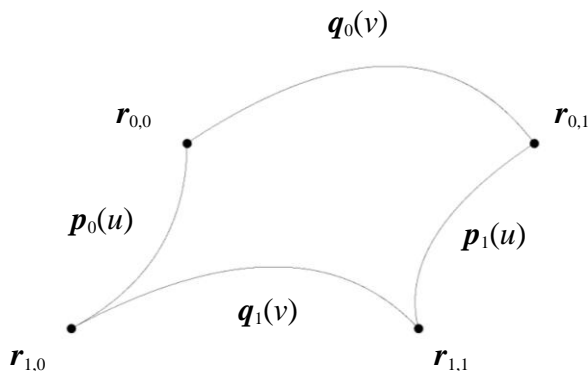


Рис. 1. Конфигурация границ четырехугольного куска поверхности

Задача состоит в том, чтобы построить кусок параметризованной поверхности $r(u, v)$, $u, v \in [0, 1]$, границами которой являются заданные кривые:

$$r(u, 0) = p_0(u), \quad r(0, v) = q_0(v), \quad r(u, 1) = p_1(u), \quad r(1, v) = q_1(v). \quad (5)$$

Кроме того, эта поверхность удовлетворяет следующим условиям на частные производные в угловых точках поверхности:

$$\frac{\partial^m r(u, v)}{\partial u^m}(0, 0) = (p_0^{(m)}(u))(0), \quad \frac{\partial^m r(u, v)}{\partial v^m}(0, 0) = (q_0^{(m)}(v))(0); \quad (6)$$

$$\frac{\partial^m r(u, v)}{\partial u^m}(0, 1) = (p_1^{(m)}(u))(0), \quad \frac{\partial^m r(u, v)}{\partial v^m}(0, 1) = (q_0^{(m)}(v))(1); \quad (7)$$

$$\frac{\partial^m \mathbf{r}(u, v)}{\partial u^m} (1, 0) = (\mathbf{p}_0^{(m)}(u))(1), \quad \frac{\partial^m \mathbf{r}(u, v)}{\partial v^m} (1, 0) = (\mathbf{q}_1^{(m)}(v))(0); \quad (8)$$

$$\frac{\partial^m \mathbf{r}(u, v)}{\partial u^m} (1, 1) = (\mathbf{p}_1^{(m)}(u))(1), \quad \frac{\partial^m \mathbf{r}(u, v)}{\partial v^m} (1, 1) = (\mathbf{q}_1^{(m)}(v))(1); \quad (9)$$

$$\frac{\partial^m \mathbf{r}(u, v)}{\partial u^r \partial v^s} (0, 0) = \frac{\partial^m \mathbf{r}(u, v)}{\partial u^r \partial v^s} (0, 1) = \frac{\partial^m \mathbf{r}(u, v)}{\partial u^r \partial v^s} (1, 0) = \frac{\partial^m \mathbf{r}(u, v)}{\partial u^r \partial v^s} (1, 1) = 0 \quad (10)$$

для всех $m \in \{1, 2, \dots, n\}$, $n \in N$, где $s + r = m$.

Для решения поставленной задачи определим требуемую поверхность $\mathbf{r}(u, v)$ следующим образом:

$$\mathbf{r}(u, v) = \mathbf{s}(u, v) - \tilde{\mathbf{r}}(u, v), \quad u, v \in [0, 1], \quad (11)$$

где

$$\begin{aligned} \mathbf{s}(u, v) &= (1 - w_{n+1}(v))\mathbf{p}_0(u) + w_{n+1}(v)\mathbf{p}_1(u) + (1 - w_{n+1}(u))\mathbf{q}_0(v) + w_{n+1}(u)\mathbf{q}_1(v), \\ \tilde{\mathbf{r}}(u, v) &= (1 - w_{n+1}(u))(1 - w_{n+1}(v))\mathbf{r}_{0,0} + w_{n+1}(u)(1 - w_{n+1}(v))\mathbf{r}_{1,0} + \\ &+ (1 - w_{n+1}(u))w_{n+1}(v)\mathbf{r}_{0,1} + w_{n+1}(u)w_{n+1}(v)\mathbf{r}_{1,1} \end{aligned}$$

и полиномы $w_{n+1}(u)$ определены при помощи равенства (1).

Покажем, что поверхность $\mathbf{r}(u, v)$ имеет требуемые границы, другими словами, удовлетворяет условиям, заданным равенствами (5). Подставляя граничные значения полиномов $w_{n+1}(u)$, заданные равенствами (2), и граничные значения кривых, заданные равенствами (4), в равенство (11), получим

$$\mathbf{s}(u, 0) = \mathbf{p}_0(u) + (1 - w_{n+1}(u))\mathbf{r}_{0,0} + w_{n+1}(u)\mathbf{r}_{1,0}, \quad \tilde{\mathbf{r}}(u, 0) = (1 - w_{n+1}(u))\mathbf{r}_{0,0} + w_{n+1}(u)\mathbf{r}_{1,0}$$

и, следовательно,

$$\mathbf{r}(u, 0) = \mathbf{s}(u, 0) - \tilde{\mathbf{r}}(u, 0) = \mathbf{p}_0(u).$$

Аналогично

$$\mathbf{s}(0, v) = (1 - w_{n+1}(v))\mathbf{r}_{0,0} + w_{n+1}(v)\mathbf{r}_{0,1} + \mathbf{q}_0(v), \quad \tilde{\mathbf{r}}(0, v) = (1 - w_{n+1}(v))\mathbf{r}_{0,0} + w_{n+1}(v)\mathbf{r}_{0,1}$$

и поэтому

$$\mathbf{r}(0, v) = \mathbf{s}(0, v) - \tilde{\mathbf{r}}(0, v) = \mathbf{q}_0(v).$$

Таким образом, первые два из равенств (5) выполняются. Аналогично можно доказать справедливость оставшихся двух равенств.

Покажем, что поверхность $\mathbf{r}(u, v)$ имеет требуемые частные производные в угловых точках, т. е. удовлетворяет условиям, заданным равенствами (6)–(10). Для этого вычислим частные производные поверхности $\mathbf{r}(u, v)$. Получим

$$\frac{\partial^m \mathbf{r}(u, v)}{\partial u^m} = \frac{\partial^m \mathbf{s}(u, v)}{\partial u^m} - \frac{\partial^m \tilde{\mathbf{r}}(u, v)}{\partial u^m}, \quad (12)$$

где

$$\frac{\partial^m \mathbf{s}(u, v)}{\partial u^m} = (1 - w_{n+1}(v))\mathbf{p}_0^{(m)}(u) + w_{n+1}(v)\mathbf{p}_1^{(m)}(u) + (1 - w_{n+1}(u))^{(m)}\mathbf{q}_0(v) + w_{n+1}^{(m)}(u)\mathbf{q}_1(v),$$

$$\begin{aligned} \frac{\partial^m \tilde{\mathbf{r}}(u, v)}{\partial u^m} &= (1 - w_{n+1}(u))^{(m)} (1 - w_{n+1}(v)) \mathbf{r}_{0,0} + w_{n+1}^{(m)}(u) (1 - w_{n+1}(v)) \mathbf{r}_{1,0} + \\ &+ (1 - w_{n+1}(u))^{(m)} w_{n+1}(v) \mathbf{r}_{0,1} + w_{n+1}^{(m)}(u) w_{n+1}(v) \mathbf{r}_{1,1} \end{aligned}$$

и

$$\frac{\partial^m \mathbf{r}(u, v)}{\partial v^m} = \frac{\partial^m \mathbf{s}(u, v)}{\partial v^m} - \frac{\partial^m \tilde{\mathbf{r}}(u, v)}{\partial v^m}, \quad (13)$$

где

$$\frac{\partial^m \mathbf{s}(u, v)}{\partial v^m} = (1 - w_{n+1}(v))^{(m)} \mathbf{p}_0(u) + w_{n+1}^{(m)}(v) \mathbf{p}_1(u) + (1 - w_{n+1}(u)) \mathbf{q}_0^{(m)}(v) + w_{n+1}(u) \mathbf{q}_1^{(m)}(v),$$

$$\begin{aligned} \frac{\partial^m \tilde{\mathbf{r}}(u, v)}{\partial v^m} &= (1 - w_{n+1}(u)) (1 - w_{n+1}(v))^{(m)} \mathbf{r}_{0,0} + w_{n+1}(u) (1 - w_{n+1}(v))^{(m)} \mathbf{r}_{1,0} + \\ &+ (1 - w_{n+1}(u)) w_{n+1}^{(m)}(v) \mathbf{r}_{0,1} + w_{n+1}(u) w_{n+1}^{(m)}(v) \mathbf{r}_{1,1} \end{aligned}$$

для всех $m \in \{1, 2, \dots, n\}$, $n \in N$. Подставляя в эти уравнения значения полиномов $w_{n+1}(u)$ и их производных в граничных точках, которые определены соответственно равенствами (2) и (3), получим

$$\begin{aligned} \frac{\partial^m \mathbf{r}(u, v)}{\partial u^m} (0, 0) &= \frac{\partial^m \mathbf{s}(u, v)}{\partial u^m} (0, 0) - \frac{\partial^m \tilde{\mathbf{r}}(u, v)}{\partial u^m} (0, 0) = (\mathbf{p}_0^{(m)}(u))(0), \\ \frac{\partial^m \mathbf{r}(u, v)}{\partial v^m} (0, 0) &= \frac{\partial^m \mathbf{s}(u, v)}{\partial v^m} (0, 0) - \frac{\partial^m \tilde{\mathbf{r}}(u, v)}{\partial v^m} (0, 0) = (\mathbf{q}_0^{(m)}(v))(0). \end{aligned}$$

Таким образом, равенства (6) выполняются. Аналогично можно доказать справедливость равенств (7)–(9).

Теперь покажем, что поверхность $\mathbf{r}(u, v)$ имеет требуемые смешанные частные производные в угловых точках, другими словами, удовлетворяет условиям, заданным равенствами (10). Для этого вычислим смешанные частные производные поверхности $\mathbf{r}(u, v)$. Получим

$$\frac{\partial^m \mathbf{r}(u, v)}{\partial u^r \partial v^s} = \frac{\partial^m \mathbf{s}(u, v)}{\partial u^r \partial v^s} - \frac{\partial^m \tilde{\mathbf{r}}(u, v)}{\partial u^r \partial v^s}, \quad (14)$$

где

$$\begin{aligned} \frac{\partial^m \mathbf{s}(u, v)}{\partial u^r \partial v^s} &= (1 - w_{n+1}(v))^{(s)} \mathbf{p}_0^{(r)}(u) + w_{n+1}^{(s)}(v) \mathbf{p}_1^{(r)}(u) + \\ &+ (1 - w_{n+1}(u))^{(r)} \mathbf{q}_0^{(s)}(v) + w_{n+1}^{(r)}(u) \mathbf{q}_1^{(s)}(v), \\ \frac{\partial^m \tilde{\mathbf{r}}(u, v)}{\partial u^r \partial v^s} &= (1 - w_{n+1}(u))^{(r)} (1 - w_{n+1}(v))^{(s)} \mathbf{r}_{0,0} + (1 - w_{n+1}(u))^{(r)} w_{n+1}^{(s)}(v) \mathbf{r}_{1,0} + \\ &+ w_{n+1}^{(r)}(u) (1 - w_{n+1}(v))^{(s)} \mathbf{r}_{0,1} + w_{n+1}^{(r)}(u) w_{n+1}^{(s)}(v) \mathbf{r}_{1,1} \end{aligned}$$

для всех $m \in \{1, 2, \dots, n\}$, $n \in N$, где $s + r = m$. Подставляя в эти уравнения значения производных полиномов $w_{n+1}(u)$ в граничных точках, которые определены равенствами (3), получим

$$\frac{\partial^m s(u, v)}{\partial u^r \partial v^s}(0, 0) = 0, \quad \frac{\partial^m \tilde{r}(u, v)}{\partial u^r \partial v^s}(0, 0) = 0$$

и, следовательно,

$$\frac{\partial^m \mathbf{r}(u, v)}{\partial u^r \partial v^s}(0, 0) = \frac{\partial^m s(u, v)}{\partial u^r \partial v^s}(0, 0) - \frac{\partial^m \tilde{r}(u, v)}{\partial u^r \partial v^s}(0, 0) = 0.$$

Аналогично можно доказать справедливость остальных равенств (10).

В результате получили, что поверхность $\mathbf{r}(u, v)$, определенная при помощи равенства (11), удовлетворяет заданным граничным условиям и имеет требуемые частные производные в угловых точках (рис. 2).

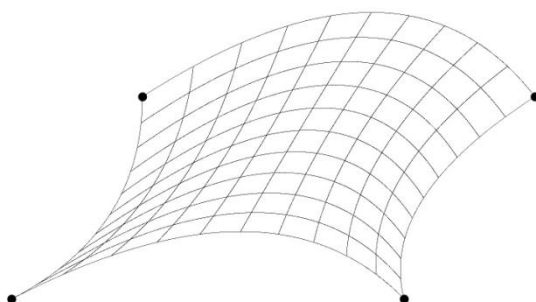


Рис. 2. Кусок поверхности $\mathbf{r}(u, v)$, построенный при помощи равенства (11)

3. Моделирование сплайн-поверхностей на четырехугольных сетках кривых

Интерполяция сеток кривых без ограничений на кручение поверхности была рассмотрена в работе [11]. Условия, обеспечивающие G^1 -непрерывность для поверхности из четырехугольных и треугольных кусков, обсуждались в статье [12]. Генерация поверхностей посредством бикубической интерполяции с использованием схемы Кунса рассмотрена в статье [13]. В данном разделе представлен подход к интерполяции четырехугольной сетки кривых кусками поверхности, которые построены смещением своих границ, как описано в разд. 2. Этот подход обеспечивает C^n -непрерывность моделируемой поверхности.

Рассмотрим двухмерный массив узловых точек $\mathbf{r}_{i,j}$, $i \in \{0, 1, \dots, k\}$, $j \in \{0, 1, \dots, l\}$, на котором построена сеть кривых $\mathbf{p}_{i,j}(u)$, $u \in [0, 1]$, и $\mathbf{q}_{i,j}(v)$, $v \in [0, 1]$, где $i \in \{0, 1, \dots, k-1\}$, $j \in \{0, 1, \dots, l-1\}$, имеющих следующие общие границы:

$$\mathbf{p}_{i,j}(0) = \mathbf{q}_{i,j}(0) = \mathbf{r}_{i,j}, \quad \mathbf{p}_{i,j}(1) = \mathbf{p}_{i+1,j}(0) = \mathbf{r}_{i+1,j}, \quad \mathbf{q}_{i,j}(1) = \mathbf{q}_{i,j+1}(0) = \mathbf{r}_{i,j+1}.$$

Кроме того, предположим, что параметризованные кривые $\mathbf{p}_{i,j}(u)$ и $\mathbf{p}_{i+1,j}(u)$, а также $\mathbf{q}_{i,j}(v)$ и $\mathbf{q}_{i,j+1}(v)$ соединены в узловых точках с непрерывностью C^n , другими словами, производные этих кривых удовлетворяют граничным условиям

$$(\mathbf{p}_{i,j}^{(m)}(u))(1) = (\mathbf{p}_{i+1,j}^{(m)}(u))(0), \quad (\mathbf{q}_{i,j}^{(m)}(v))(1) = (\mathbf{q}_{i,j+1}^{(m)}(v))(0)$$

для всех $m \in \{1, 2, \dots, n\}$, $n \in \mathbb{N}$. Задача состоит в том, чтобы построить на заданной сетке кривых интерполирующую поверхность $\mathbf{r}(u, v)$, которая также имеет степень непрерывно-

сти C^n . Определим клетку поверхности $\mathbf{r}(u, v)$, ограниченную параметризованными кривыми $\mathbf{p}_{i,j}(u)$, $\mathbf{p}_{i,j+1}(u)$, $\mathbf{q}_{i,j}(v)$ и $\mathbf{q}_{i+1,j}(v)$, используя равенство (11), следующим образом:

$$\mathbf{r}_{i,j}(u, v) = \mathbf{s}_{i,j}(u, v) - \tilde{\mathbf{r}}_{i,j}(u, v), \quad u, v \in [0, 1], \quad (15)$$

где

$$\mathbf{s}_{i,j}(u, v) = (1 - w_{n+1}(v))\mathbf{p}_{i,j}(u) + w_{n+1}(v)\mathbf{p}_{i,j+1}(u) + (1 - w_{n+1}(u))\mathbf{q}_{i,j}(v) + w_{n+1}(u)\mathbf{q}_{i+1,j}(v),$$

$$\tilde{\mathbf{r}}_{i,j}(u, v) = (1 - w_{n+1}(u))(1 - w_{n+1}(v))\mathbf{r}_{i,j} + w_{n+1}(u)(1 - w_{n+1}(v))\mathbf{r}_{i+1,j} +$$

$$+ (1 - w_{n+1}(u))w_{n+1}(v)\mathbf{r}_{i,j+1} + w_{n+1}(u)w_{n+1}(v)\mathbf{r}_{i+1,j+1}.$$

Из результатов, полученных в разд. 2, следует, что клетка поверхности $\mathbf{r}_{i,j}(u, v)$ имеет своими границами параметризованные кривые $\mathbf{p}_{i,j}(u)$, $\mathbf{p}_{i,j+1}(u)$, $\mathbf{q}_{i,j}(v)$ и $\mathbf{q}_{i+1,j}(v)$. Кроме того, поверхность $\mathbf{r}(u, v)$ имеет непрерывность C^n в узловых точках $\mathbf{r}_{i,j}$. Покажем, что поверхность $\mathbf{r}(u, v)$ имеет также степень непрерывности C^n на границах клеток. Для этого вычислим частные производные поперек границ клетки, используя равенства (12), (13). Получим

$$\begin{aligned} \frac{\partial^m \mathbf{r}_{i,j}(u, v)}{\partial u^m}(1, v) &= (1 - w_{n+1}(v))(\mathbf{p}_{i,j}^{(m)}(u))(1) + w_{n+1}(v)(\mathbf{p}_{i,j+1}^{(m)}(u))(1) = \\ &= (1 - w_{n+1}(v))(\mathbf{p}_{i+1,j}^{(m)}(u))(0) + w_{n+1}(v)(\mathbf{p}_{i+1,j+1}^{(m)}(u))(0) = \frac{\partial^m \mathbf{r}_{i+1,j}(u, v)}{\partial u^m}(0, v) \end{aligned}$$

и аналогично

$$\frac{\partial^m \mathbf{r}_{i,j}(u, v)}{\partial v^m}(u, 1) = \frac{\partial^m \mathbf{r}_{i,j+1}(u, v)}{\partial v^m}(u, 0)$$

для всех $m \in \{1, 2, \dots, n\}$. Используя равенство (14), вычислим смешанные частные производные поперек границ клетки. Получим

$$\begin{aligned} \frac{\partial^m \mathbf{r}_{i,j}(u, v)}{\partial u^r \partial v^s}(1, v) &= (1 - w_{n+1}(v))^{(s)}(\mathbf{p}_{i,j}^{(r)}(u))(1) + w_{n+1}^{(s)}(v)(\mathbf{p}_{i,j+1}^{(r)}(u))(1) = \\ &= (1 - w_{n+1}(v))^{(s)}(\mathbf{p}_{i+1,j}^{(r)}(u))(0) + w_{n+1}^{(s)}(v)(\mathbf{p}_{i+1,j+1}^{(r)}(u))(0) = \frac{\partial^m \mathbf{r}_{i+1,j}(u, v)}{\partial u^r \partial v^s}(0, v) \end{aligned}$$

и аналогично

$$\frac{\partial^m \mathbf{r}_{i,j}(u, v)}{\partial u^r \partial v^s}(u, 1) = \frac{\partial^m \mathbf{r}_{i,j+1}(u, v)}{\partial u^r \partial v^s}(u, 0)$$

для всех $m \in \{1, 2, \dots, n\}$, где $s + r = m$.

Таким образом, поверхность $\mathbf{r}(u, v)$, построенная при помощи равенства (15), является C^n -непрерывной. На рис. 3 показана сплайн-поверхность, построенная предложенным методом на заданной сетке кривых.

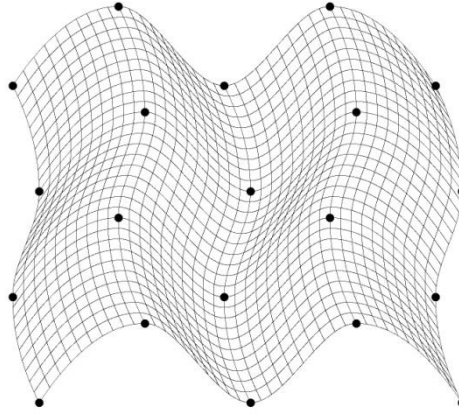


Рис. 3. Сплайн-поверхность, построенная на заданной сетке кривых

4. Моделирование сплайн-поверхностей Бэ́зы

Рассмотрим применение предложенного подхода к построению сплайн-поверхностей Бэ́зы на четырехугольных сетках кривых. Представление поверхностей Гордона [7] в форме поверхностей Бэ́зы впервые было рассмотрено в работе [14]. Локальная интерполяция поверхностей с использованием кусков Бэ́зы на сетке кубических кривых описана в работе [15]. Этот метод обеспечивает G^1 -непрерывность на границах кусков поверхности. В данном разделе представлен метод, который позволяет моделировать сплайн-поверхность Бэ́зы требуемой непрерывности C^n . Для выполнения этого условия требуется, чтобы прямоугольная сетка кривых, на которой моделируется сплайн-поверхность, также была C^n -непрерывной. Для обеспечения данного свойства четырехугольной сетки кривых (учитывая, что поверхность моделируется локально) границы куска поверхности должны быть представлены кривыми Бэ́зы, по крайней мере, степени $2n+1$. В таком случае границы четырехугольного куска поверхности $\mathbf{r}_{i,j}(u, v)$ можно представить следующим образом:

$$\mathbf{p}_{i,j}(u) = \sum_{k=0}^{2n+1} b_{2n+1,k}(u) \mathbf{p}_{i,j,k}(u), \quad u \in [0, 1], \quad \mathbf{q}_{i,j}(v) = \sum_{l=0}^{2n+1} b_{2n+1,l}(v) \mathbf{q}_{i,j,l}(v), \quad v \in [0, 1],$$

где граничные точки кривых Бэ́зы $\mathbf{p}_{i,j}(u)$ и $\mathbf{q}_{i,j}(v)$ являются узловыми точками четырехугольной сетки кривых:

$$\mathbf{p}_{i,j}(0) = \mathbf{p}_{i,j,0} = \mathbf{r}_{i,j}, \quad \mathbf{p}_{i,j}(1) = \mathbf{p}_{i,j,2n+1} = \mathbf{r}_{i+1,j},$$

$$\mathbf{q}_{i,j}(0) = \mathbf{q}_{i,j,0} = \mathbf{r}_{i,j}, \quad \mathbf{q}_{i,j}(1) = \mathbf{q}_{i,j,2n+1} = \mathbf{r}_{i+1,j}.$$

Тогда четырехугольный кусок поверхности $\mathbf{r}_{i,j}(u, v)$ может быть описан, используя равенства (15), (16), следующим образом:

$$\begin{aligned} \mathbf{r}_{i,j}(u, v) &= \mathbf{s}_{i,j}(u, v) - \tilde{\mathbf{r}}_{i,j}(u, v) = \\ &= \sum_{l=0}^n b_{2n+1,l}(v) \sum_{k=0}^{2n+1} b_{2n+1,k}(u) \mathbf{p}_{i,j,k} + \sum_{l=n+1}^{2n+1} b_{2n+1,l}(v) \sum_{k=0}^{2n+1} b_{2n+1,k}(u) \mathbf{p}_{i,j+1,k} + \end{aligned}$$

$$\begin{aligned}
 & + \sum_{k=0}^n b_{2n+1,k}(u) \sum_{l=0}^{2n+1} b_{2n+1,l}(v) \mathbf{q}_{i,j,l} + \sum_{k=n+1}^{2n+1} b_{2n+1,k}(u) \sum_{l=0}^{2n+1} b_{2n+1,l}(v) \mathbf{q}_{i+1,j,l} - \\
 & - \sum_{k=0}^n b_{2n+1,k}(u) \sum_{l=0}^n b_{2n+1,l}(v) \mathbf{r}_{i,j} - \sum_{k=n+1}^{2n+1} b_{2n+1,k}(u) \sum_{l=0}^n b_{2n+1,l}(v) \mathbf{r}_{i+1,j} - \\
 & - \sum_{k=0}^n b_{2n+1,k}(u) \sum_{l=n+1}^{2n+1} b_{2n+1,l}(v) \mathbf{r}_{i,j+1} - \sum_{k=n+1}^{2n+1} b_{2n+1,k}(u) \sum_{l=n+1}^{2n+1} b_{2n+1,l}(v) \mathbf{r}_{i+1,j+1}.
 \end{aligned}$$

Приводя подобные, получим представление четырехугольного куска поверхности Бэ́зье, которая имеет непрерывности C^n :

$$\begin{aligned}
 \mathbf{r}_{i,j}(u,v) & = \sum_{k=0}^n b_{2n+1,k}(u) \sum_{l=0}^n b_{2n+1,l}(v) (\mathbf{p}_{i,j,k} + \mathbf{q}_{i,j,l} - \mathbf{r}_{i,j}) + \\
 & + \sum_{k=0}^n b_{2n+1,k}(u) \sum_{l=n+1}^{2n+1} b_{2n+1,l}(v) (\mathbf{p}_{i,j+1,k} + \mathbf{q}_{i,j,l} - \mathbf{r}_{i,j+1}) + \\
 & + \sum_{k=n+1}^{2n+1} b_{2n+1,k}(u) \sum_{l=0}^n b_{2n+1,l}(v) (\mathbf{p}_{i,j,k} + \mathbf{q}_{i+1,j,l} - \mathbf{r}_{i+1,j}) + \\
 & + \sum_{k=n+1}^{2n+1} b_{2n+1,k}(u) \sum_{l=n+1}^{2n+1} b_{2n+1,l}(v) (\mathbf{p}_{i,j+1,k} + \mathbf{q}_{i+1,j,l} - \mathbf{r}_{i+1,j+1}).
 \end{aligned}$$

В геометрических приложениях главным образом используются сплайн-поверхности с непрерывностью C^1 и C^2 . Четырехугольные куски этих поверхностей имеют следующее представление Бэ́зье:

– для поверхности с непрерывностью C^1

$$\mathbf{r}_{i,j}(u,v) = \sum_{k=0}^3 b_{3,k}(u) \sum_{l=0}^3 b_{3,l}(v) \mathbf{p}_{k,l},$$

где точки $\mathbf{p}_{k,l}$ могут рассматриваться как элементы матрицы

$$\mathbf{P}_{k,l} = \begin{bmatrix} \mathbf{r}_{i,j} & \mathbf{q}_{i,j,1} & \mathbf{q}_{i,j,2} & \mathbf{r}_{i,j+1} \\ \mathbf{p}_{i,j,1} & \mathbf{p}_{i,j,1} + \mathbf{q}_{i,j,1} - \mathbf{r}_{i,j} & \mathbf{p}_{i,j+1,1} + \mathbf{q}_{i,j,2} - \mathbf{r}_{i,j+1} & \mathbf{p}_{i,j+1,1} \\ \mathbf{p}_{i,j,2} & \mathbf{p}_{i,j,2} + \mathbf{q}_{i+1,j,1} - \mathbf{r}_{i+1,j} & \mathbf{p}_{i,j+1,2} + \mathbf{q}_{i+1,j,2} - \mathbf{r}_{i+1,j+1} & \mathbf{p}_{i,j+1,2} \\ \mathbf{r}_{i+1,j} & \mathbf{q}_{i+1,j,1} & \mathbf{q}_{i+1,j,2} & \mathbf{r}_{i+1,j+1} \end{bmatrix},$$

– для поверхности с непрерывностью C^2

$$\mathbf{r}_{i,j}(u,v) = \sum_{k=0}^5 b_{3,k}(u) \sum_{l=0}^5 b_{3,l}(v) \mathbf{p}_{k,l},$$

где точки $\mathbf{p}_{k,l}$ могут рассматриваться как элементы блочной матрицы

$$\mathbf{P}_{k,l} = \begin{bmatrix} \mathbf{B}_{0,0} & \mathbf{B}_{0,1} \\ \mathbf{B}_{1,0} & \mathbf{B}_{1,1} \end{bmatrix}.$$

Здесь

$$\begin{aligned}
 \mathbf{B}_{0,0} &= \begin{bmatrix} \mathbf{r}_{i,j} & \mathbf{q}_{i,j,1} & \mathbf{q}_{i,j,2} \\ \mathbf{p}_{i,j,1} & \mathbf{p}_{i,j,1} + \mathbf{q}_{i,j,1} - \mathbf{r}_{i,j} & \mathbf{p}_{i,j,1} + \mathbf{q}_{i,j,2} - \mathbf{r}_{i,j} \\ \mathbf{p}_{i,j,2} & \mathbf{p}_{i,j,2} + \mathbf{q}_{i,j,1} - \mathbf{r}_{i,j} & \mathbf{p}_{i,j,2} + \mathbf{q}_{i,j,2} - \mathbf{r}_{i,j} \end{bmatrix}, \\
 \mathbf{B}_{0,1} &= \begin{bmatrix} & \mathbf{q}_{i,j,3} & & \mathbf{q}_{i,j,4} & & \mathbf{r}_{i,j+1} \\ \mathbf{p}_{i,j+1,1} + \mathbf{q}_{i,j,3} - \mathbf{r}_{i,j+1} & & \mathbf{p}_{i,j+1,1} + \mathbf{q}_{i,j,4} - \mathbf{r}_{i,j+1} & & \mathbf{p}_{i,j+1,1} & \\ \mathbf{p}_{i,j+1,2} + \mathbf{q}_{i,j,3} - \mathbf{r}_{i,j+1} & & \mathbf{p}_{i,j+1,2} + \mathbf{q}_{i,j,4} - \mathbf{r}_{i,j+1} & & \mathbf{p}_{i,j+1,2} & \end{bmatrix}, \\
 \mathbf{B}_{1,0} &= \begin{bmatrix} \mathbf{p}_{i,j,3} & \mathbf{p}_{i,j,3} + \mathbf{q}_{i+1,j,1} - \mathbf{r}_{i+1,j} & \mathbf{p}_{i,j,3} + \mathbf{q}_{i+1,j,2} - \mathbf{r}_{i+1,j} \\ \mathbf{p}_{i,j,4} & \mathbf{p}_{i,j,4} + \mathbf{q}_{i+1,j,1} - \mathbf{r}_{i+1,j} & \mathbf{p}_{i,j,4} + \mathbf{q}_{i+1,j,2} - \mathbf{r}_{i+1,j} \\ \mathbf{r}_{i+1,j} & \mathbf{q}_{i+1,j,1} & \mathbf{q}_{i+1,j,2} \end{bmatrix}, \\
 \mathbf{B}_{1,1} &= \begin{bmatrix} \mathbf{p}_{i,j+1,3} + \mathbf{q}_{i+1,j,3} - \mathbf{r}_{i+1,j+1} & \mathbf{p}_{i,j+1,3} + \mathbf{q}_{i+1,j,4} - \mathbf{r}_{i+1,j+1} & \mathbf{p}_{i,j+1,3} \\ \mathbf{p}_{i,j+1,4} + \mathbf{q}_{i+1,j,3} - \mathbf{r}_{i+1,j+1} & \mathbf{p}_{i,j+1,4} + \mathbf{q}_{i+1,j,4} - \mathbf{r}_{i+1,j+1} & \mathbf{p}_{i,j+1,4} \\ & \mathbf{q}_{i+1,j,3} & \mathbf{q}_{i+1,j,4} & \mathbf{r}_{i+1,j+1} \end{bmatrix}.
 \end{aligned}$$

Аналогично могут быть организованы в блочную матрицу точки для моделирования куска четырехугольной сплайн-поверхности с непрерывностью C^n . На рис. 4 показана сплайн-поверхность Бэзе непрерывности C^2 , построенная на заданной сетке кривых Бэзе.

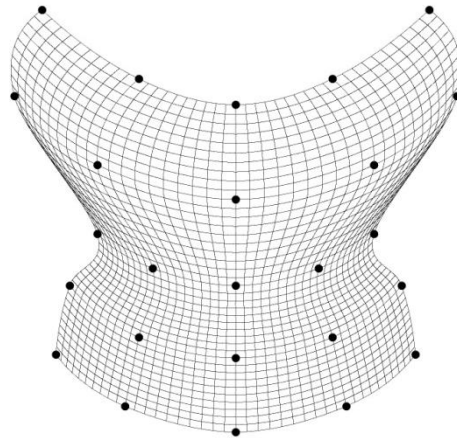


Рис. 4. Сплайн-поверхность Бэзе непрерывности C^2

Заключение

В статье представлен подход к моделированию интерполяционных сплайн-поверхностей с непрерывностью C^n на четырехугольных сетках кривых. Клетка поверхности строится посредством смешивания своих границ при помощи весовых полиномов. Доказано, что требуемая параметрическая непрерывность поверхности обеспечивается выбором полинома подходящей степени. Представленный подход обеспечивает локальные геометрические свойства моделируемой поверхности. Если сетка кривых задана кривыми Бэзе, то куски моделируемой сплайн-поверхности являются поверхностями Бэзе. Такой подход к моделированию поверхностей может быть использован для быстрого моделирования набросков поверхностей, а также, учитывая локальное свойство поверхностей, может применяться в геометрических приложениях, требующих вычислений в режиме реального времени.

Список литературы

1. Robin, J.Y. Geometry and Interpolation of Curves and Surfaces / J.Y. Robin, M. McLeod, L. Baart. – Cambridge University Press, 2011. – 430 p.
2. Peters, J. Local smooth surface interpolation: a classification / J. Peters // Computer Aided Geometric Design. – 1990. – Vol. 7, no. 1–4. – P. 191–195.
3. Vida, Q.J. A survey of blending methods that use parametric surfaces / Q.J. Vida, R.R. Martin, T. Várady // Computer-Aided Design. – 1994. – Vol. 26, no. 5. – P. 341–365.
4. Coons, S.A. Surfaces for computer-aided design of space forms / S.A. Coons // Report MAC-TR-41, Project MAC. – Massachusetts Institute of Technology, 1967.
5. Forrest, A.R. On Coons and other methods for the representation of curved surfaces / A.R. Forrest // Computer Graphics and Image Processing. – 1972. – Vol. 1, no. 4. – P. 341–359.
6. Фокс, А. Вычислительная геометрия. Применение в проектировании и производстве / А. Фокс, М. Пратт. – М. : Мир, 1982. – 304 с.
7. Gordon, W.J. Spline-blended surface interpolation through curve networks / W.J. Gordon // J. Math. and Mech. – 1969. – Vol. 18, no. 10. – P. 931–957.
8. Juhász, I. Surface interpolation with local control by linear blending / I. Juhász, M. Hoffmann // Annales Mathematicae et Informaticae. – 2009. – Vol. 36. – P. 77–84.
9. Побегайло, А.П. Полиномиальная деформация кривых и ее применение в геометрическом моделировании / А.П. Побегайло // Вестник БГУ. Сер. 1. Физика. Математика. Информатика. – 2009. – № 3. – С. 106–109.
10. Pobegailo, A.P. Interpolation of Rectangular Grids Using Deformation of Curves / A.P. Pobegailo // Pattern Recognition and Information Processing (PRIP)'2009 : proc. of the 10th Intern. Conf. (19–21 May, 2009, Minsk, Belarus). – Minsk : Publ. Center of BSU, 2009. – P. 334–338.
11. Gregory, J. Smooth interpolation without twist constraints / J. Gregory ; ed. R.E. Barnhill, R.F. Riesenfeld // Computer Aided Geometric Design. – Academic Press, 1974. – P. 71–88.
12. Farin, G. Agnostic G1 Gregory Surfaces / G. Farin, D. Hansford // Graphical Models. – 2012. – Vol. 76, no. 4. – P. 346–350.
13. Comninos, P. An interpolating piecewise bicubic surface with shape parameters / P. Comninos // Computers & Graphics. – 2001. – Vol. 25, no. 3. – P. 463–481.
14. Chiyokura, H. Design of solids with free-form surfaces / H. Chiyokura, F. Kimura // Computer Graphics. – 1983. – Vol. 17, no. 3. – P. 289–298.
15. Shirman, L.A. Local surface interpolation with B_ezier patches / L.A. Shirman, C.H. Sequin // Computer Aided Geometric Design. – 1987. – Vol. 4, no. 4. – P. 279–295.

Поступила 25.06.2014

*Белорусский государственный университет,
Минск, пр. Независимости, 4
e-mail: pobegailo@bsu.by*

A.P. Pobegailo

**DESIGN OF BEZIER SPLINE SURFACES
OVER BIVARIATE NETWORKS OF CURVES**

The paper presents an approach to construct interpolating spline surfaces over a bivariate network of curves with rectangular patches. Patches of the interpolating spline surface are constructed by means of blending their boundaries with special polynomials. In order to ensure a necessary parametric continuity of the designed surface the polynomials of the corresponding degree must be used. The constructed interpolating spline surfaces have a local shape control. If the surface frame is determined by means of Bezier curves, then patches of the interpolating spline surface are Bezier surfaces. The presented approach to surface modeling can be used in such applications as computer graphics and geometric design.

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 004.921+004.415.2

О.Г. Казанцева, Е.А. Барановский, Е.А. Ландарский

ШЕЙДЕРНАЯ МЕТАМОДЕЛЬ КАК КОМПОНЕНТ АРХИТЕКТУРЫ
ИНТЕРАКТИВНОГО ПРИЛОЖЕНИЯ

Предлагается метамодель рендеринга в качестве основы для архитектуры сложных графических приложений. Данная модель позволяет отделить абстрактные логические и алгоритмические подходы, используемые в теории современной компьютерной графики, от вопросов управления вычислительными ресурсами и предоставляет большие возможности по управлению сложностью программного кода.

Введение

Одними из ключевых элементов компьютерной графики стали программируемые графические шейдеры – программы, которые могут быть загружены для исполнения на графическом процессоре [1].

Использование таких устоявшихся парадигм, как объектно-ориентированное и процедурное программирование, сегодня не может в полной мере удовлетворить потребность в управлении постоянно растущей сложностью программного обеспечения. Новые подходы и технологии наподобие аспектно-ориентированной парадигмы и концепции IoC (Inversion of Control)-контейнера уже доказали свою эффективность в сфере заказных программных продуктов (интернет-магазинов, систем автоматизации документооборота и т. д.). Между тем опасения потерь производительности замедлили внедрение этих технологий при разработке программных ядер для приложений, насыщенных графикой: игр, систем автоматизации проектирования и моделирования, что является серьезным заблуждением, так как мощность даже мобильных устройств сегодня значительно возросла. Более того, любой инструмент, упрощающий контроль над программным обеспечением, всегда можно использовать разумно, чтобы обойти его недостатки.

В работе анализируются существующие подходы к управлению программируемого сценария рендеринга (programmable graphics pipeline) и проектируются более гибкие механизмы рендеринга с учетом лучших их черт.

1. Стратегия управления шейдерами

Управление шейдерами порождает целый ряд проблем при разработке приложений, использующих аппаратное ускорение графики. При этом многие из них являются общими для всех графических API (application programming interface), используемых на сегодняшний день как на мобильных, так и на настольных платформах. Общий сценарий работы с шейдерами представляется как следующая последовательность шагов, теоретически не вызывающих явных трудностей [2–4]:

– инициализация:

- подготовка исходного кода составляющих подпрограмм (вершинный шейдер, пиксельный шейдер, геометрический шейдер и т. д.);
- компиляция каждой подпрограммы;
- объявление параметров, используемых в каждой подпрограмме;
- связывание при необходимости подпрограмм в единую программу;

– использование в цикле рендеринга:

- активация ресурсов (например, текстуры) в памяти GPU (graphics processing unit);
- передача параметров в шейдерные подпрограммы;

- активация шейдерного кода;
 - освобождение ресурсов, занимаемых шейдерным кодом или ассоциированными с ним, когда шейдер не нужен.

Следует отметить, что порядок перечисленных выше шагов зависит от специфики используемого API. Как видно из общего сценария работы с шейдерами, с применением шейдеров сопряжено достаточно интенсивное использование вызовов методов графического API, а также резервирование нескольких различных ресурсов в памяти GPU.

Предположим, что у нас имеется некоторая сцена, состоящая, как минимум, из нескольких десятков объектов, обладающих набором различных визуальных характеристик – состояний визуализации (Render State). Далее будем называть данное понятие *RS*. Предположим, что эти объекты представляют несколько категорий, которые различаются конфигурациями их *RS*. Объекты могут иметь различные материалы (разную степень прозрачности, реакцию на освещение, излучение света, текстуру и т. п.). Различия могут быть и более существенными. Например, для некоторых категорий объектов определенный набор эффектов, таких как динамическое затенение, может быть проигнорирован из соображений достижения приемлемого баланса производительности и качества. Более того, в целях оптимизации вышеперечисленные свойства могут изменяться в зависимости от расстояния до точки наблюдателя независимо от категории объекта. Для визуализации такой сцены подходят шейдеры. Будучи далеко не самой сложной, данная ситуация уже порождает ряд проблем. Так как программная логика в этом случае допускает динамические изменения *RS* любого из объектов сцены, может возникнуть необходимость переключиться на другой шейдерный код. Вполне возможно, что данный код может понадобиться не сразу или не понадобиться вообще. Также не исключено, что уже загруженный шейдерный код не будет использоваться достаточно длительный промежуток времени. Здесь возникает вопрос о времени инициализации и загрузки шейдеров в память GPU, вопрос их совместного использования несколькими объектами, а также эффективного отслеживания неиспользуемого кода. Вместе с общей нетривиальностью приложений, пользующихся аппаратным ускорением графики, и необходимостью изолировать всю полезную для пользователя логику от управления ресурсами системы, эти вопросы представляют достаточно серьезный барьер для проектирования и написания качественных и гибких программ.

Для решения вышеописанных проблем можно выбирать различные пути: от применения усложненных структур управления на стороне GPU до организации специальной логики на стороне CPU.

Два самых простых и распространенных подхода:

- использование так называемых убершейдеров;
- подготовка всевозможных вариантов шейдеров по отдельности и их предварительное кэширование.

Первый подход подразумевает написание одной или нескольких больших шейдерных программ, содержащих множество операторов ветвления и макроопределения. Каждая такая программа затем используется для работы с несколькими вариантами *RS* в зависимости от значений достаточно обширного набора входных параметров. Преимущества такого подхода состоят в существенном сокращении числа дескрипторов ресурсов, выделенных в памяти GPU, в централизации описания шейдерной модели, а также в реализации достаточно простой логики конечного автомата для управления сценарием рендеринга. Однако здесь имеется и ряд серьезных недостатков. В первую очередь это удар по производительности при работе с убершейдерами из-за необходимости выполнения большого числа проверок на стороне GPU, что наиболее заметно на этапе вычисления цвета фрагментов изображения для пиксельных шейдеров и для любых шейдеров в целом при работе с мобильными платформами. Кроме того, при необходимости радикальных изменений в сценарии рендеринга приходится существенно усложнять либо переписывать весь шейдерный код.

Второй подход состоит в разбиении каждой из шейдерных подпрограмм на секции, отвечающие за тот или иной этап вычисления, написании реализации каждой из таких секций для всех частных случаев и сборки всевозможных комбинаций из полученных составляющих кода. На этапе выполнения программы в этом случае все вариации, которые могут потребо-

ваться, предварительно компилируются и загружаются в память GPU. Преимущество данного подхода состоит в естественности разбиения сценария рендеринга на логические блоки, которые легко изменять, добавлять и удалять, а также в простоте этапа инициализации. В остальном это решение приносит избыток использования памяти GPU, а следовательно, и снижение производительности (за счет дополнительных действий, предотвращающих исчерпание ресурсов GPU). Кроме того, оно обычно требует использования дополнительных инструментов для проверки всех фрагментов кода и генерации из них готовых шейдеров.

Описанные подходы действуют абсолютно противоположно, но на практике чаще всего требуется компромисс. Более того, они слишком прямолинейны и обладают целым рядом серьезных недостатков.

Наличие отдельной иерархии классов, инкапсулирующих работу с конфигурируемой шейдерной моделью, основанной на метаданных, позволило бы упростить разработку сложных сценариев рендеринга и гибко контролировать баланс производительности и качества на каждом его этапе.

2. Общий сценарий рендеринга сцены

Опишем общие концепции логики рендеринга сцены с учетом возможности программирования на стороне GPU.

Интерфейс *Skippable* определяет поведение классов, экземпляры которых имеют аспект логического состояния, позволяющий определить, следует ли пропустить данный объект на итерации некоторого алгоритма.

Интерфейс *Stateful* определяет поведение классов, экземпляры которых имеют составное состояние (куда и входит *RS*) и предоставляют сервисы для его изменения.

Интерфейс *Transformable* определяет поведение классов, экземпляры которых имеют положение и границы в пространстве и предоставляют сервисы по их изменению.

Интерфейс *Renderable* расширяет интерфейсы *Skippable*, *Stateful* и *Transformable* и определяет поведение классов, экземпляры которых поддаются процессу рендеринга.

Интерфейс *Node* расширяет интерфейсы *Skippable*, *Stateful* и *Transformable* и определяет поведение классов, экземпляры которых могут быть добавлены в граф сцены с целью группировки нескольких экземпляров типа *Renderable*.

Интерфейс *Controller* определяет поведение классов, экземпляры которых отвечают за обновление определенных аспектов состояния у наблюдаемых экземпляров типа *Stateful*.

Интерфейс *Renderer* определяет поведение классов, экземпляры которых осуществляют рендеринг экземпляров типа *Renderable*.

Интерфейс *GraphicsPipeline* определяет поведение класса, экземпляр которого предоставляет сервис по рендерингу графа сцены, представленного экземпляром типа *Node*. Этот экземпляр является корневым узлом графа сцены.

Общая логика инициализации *GraphicsPipeline* представлена на диаграмме последовательности (рис. 1). Объекты *models*, загружаемые в самом начале, соответствуют набору так называемых *метамodelей*. Под метамodelью далее будем понимать конфигурируемую сущность, соответствующую некоторым правилам описания и содержащую метаданные, по которым приложение во время исполнения может восстановить состояние набора объектов, связанного общей целью функционирования. Одной из таких сущностей является шейдерная модель, которую рассмотрим ниже. Момент восстановления необходимого состояния объектов, скрытых под абстракцией *GraphicsPipeline* и *Renderer*, обозначен событием *configurePipeline(..)*.

Далее рассмотрим логику итерации рендеринга (рис. 2). Вначале поток приложения, выделенный под рендеринг, иницирует итерацию над заданным графом сцены, передавая объекту *graphicsPipeline* его корневой узел (экземпляр *Node*). Далее *graphicsPipeline* собирает коллекцию всех контроллеров *controllers*, ассоциированных со всеми элементами графа сцены, после чего обновляет состояние этих элементов. Затем следует этап предварительной обработки графа сцены (например, сортировки и фильтрации), в результате чего получается упорядоченный список объектов, которые должны быть видны в текущем кадре.

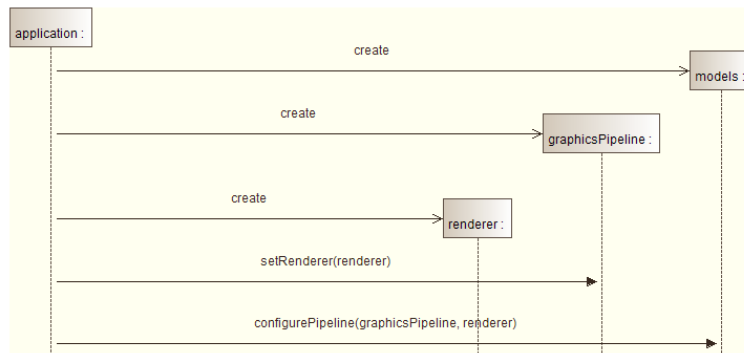


Рис. 1. Общая логика инициализации сценария рендеринга

Одна итерация рендеринга с учетом таких графических техник, как отложенное освещение и эффект адаптации глаз, может состоять из нескольких субитераций – проходов (render passes). Поэтому завершающий этап рендеринга одного кадра состоит в выполнении всех таких проходов в требуемой последовательности. В свою очередь, для каждой субитерации выполняются следующие действия:

- учет параметров графического контекста – объекта, через который осуществляется связь с GPU;
- определение набора из отобранных ранее объектов, который должен быть отображен на текущем проходе;
- подготовка экземпляра *Renderer*, связанного с *graphicsPipeline*. Это необходимо, если, например, некоторые аспекты *RS* должны быть проигнорированы на текущем проходе;
- запрос у экземпляра *Renderer* на заполнение целевого буфера текущего прохода.

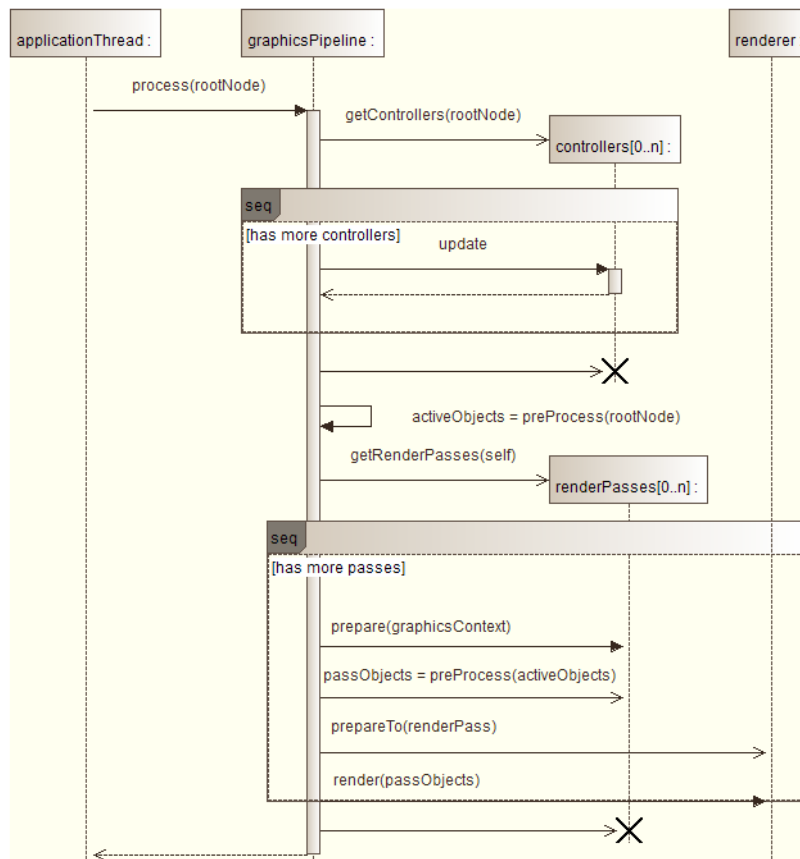


Рис. 2. Итерация рендеринга

Рассмотрим более детально действия, выполняемые экземпляром *Renderer* (далее будем называть его *визуализатор*) в рамках одного прохода. Из диаграммы (рис. 3) видно, что визуализатор для каждого объекта из переданной для отображения коллекции перебирает все активные на текущем проходе аспекты *RS* и обрабатывает их с помощью подходящих аппликаторов – вспомогательных объектов, инкапсулирующих работу с конкретным типом сущности состояния.

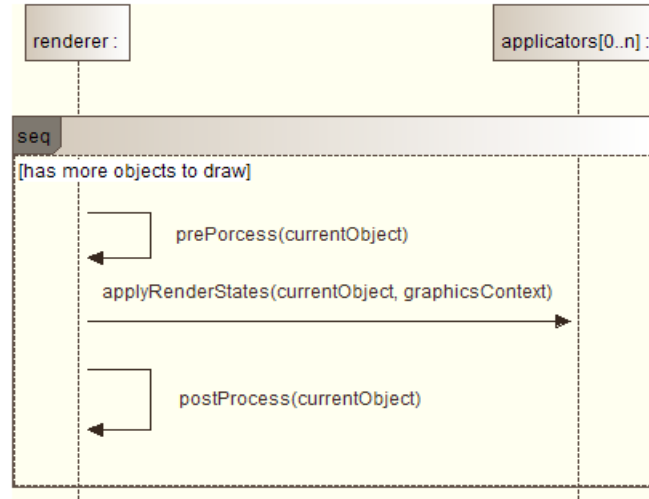


Рис. 3. Действия *Renderer* в рамках одной субитерации

Объекты, которые нужно отобразить в текущем кадре, передаются визуализатору с информацией о *RS* в такой форме, что все параметры, которые нужны при выполнении или подготовке шейдера, содержатся в одной сущности состояния – *shader state*.

Рассмотрим теперь возможную логику действия аппликатора для *shader state* (рис. 4). Вначале данный аппликатор проверяет, была ли изменена сущность для *shader state* так, что для нее необходимо сменить шейдерную программу. В зависимости от результата проверки он будет вынужден обратиться к менеджеру шейдеров за подходящей программой, иначе будет использована текущая. После разрешения шейдерной программы аппликатор активирует ее и применяет к ней все параметры *shader state* в соответствии со спецификой используемого графического API.

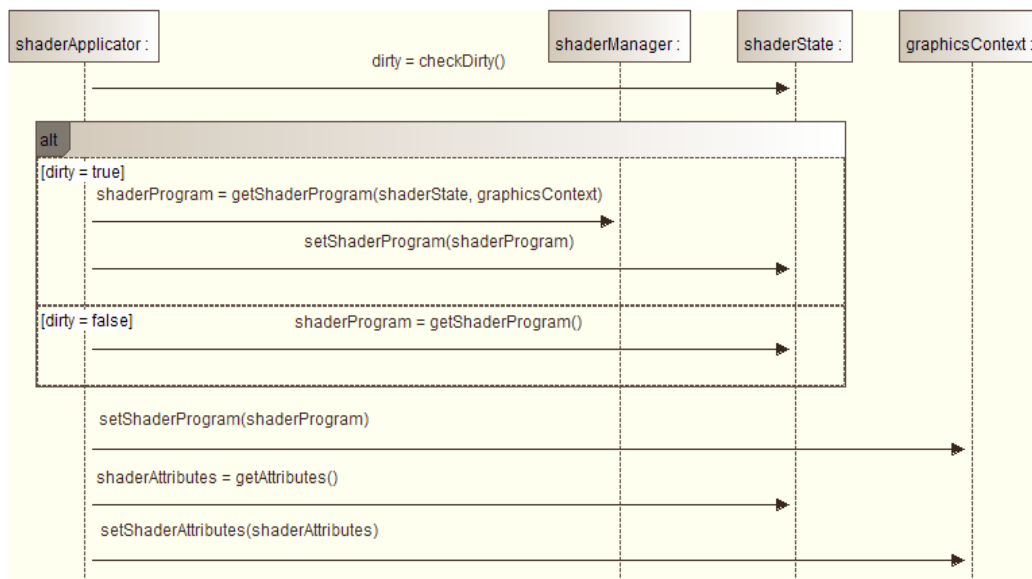


Рис. 4. Диаграмма последовательности действий аппликатора *shader state*

- Примерный алгоритм работы менеджера шейдеров (рис. 5) состоит из следующих шагов:
- извлечение из сущности *shader state* ключа, идентифицирующего подходящую программу;
 - проверка на существование скомпилированной программы, соответствующей данному ключу;
 - в случае отсутствия такой программы:
 - построение кода программы;
 - его компиляция;
 - сохранение скомпилированной программы в реестре;
 - возврат программы.

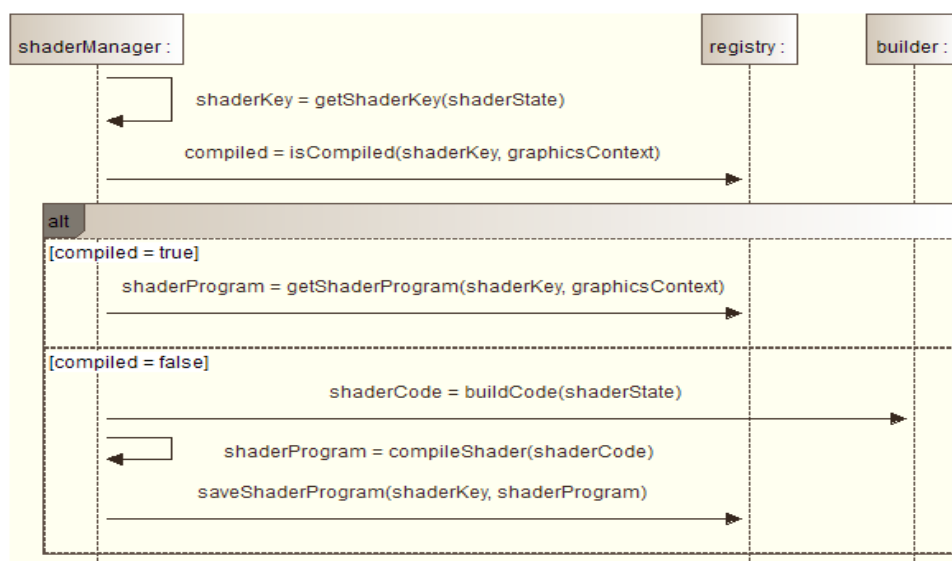


Рис. 5. Принцип действия менеджера шейдеров

3. Шейдерная модель

Ранее было отмечено, что на этапе инициализации приложения значительная часть действий по восстановлению объектной структуры и связыванию компонентов выполняется в соответствии с металогики и метаданными, содержащимися в предварительно загруженных моделях. В случае организации сценария рендеринга одной из таких моделей является шейдерная модель, отвечающая за его программируемую часть (*programmable pipeline*). Как и любая другая метасущность, шейдерная модель имеет персистентное представление, предназначенное для возможности хранения между запусками приложения и конфигурирования, и представление во время исполнения приложения. Персистентное представление логично реализовать в одном из структурированных форматов, таких как XML, YAML или JSON.

Представление шейдерной модели времени исполнения приложения удобно разбить на две составляющие: ранее упомянутую сущность *shader state*, являющуюся частью состояния визуализации отображаемого объекта, и вспомогательные объекты, упрощающие работу для классов, реализующих логику рендеринга (т. е. для *GraphicsPipeline*, *Renderer*, аппликатора и менеджера шейдеров).

Сущность *shader state* для классов логики рендеринга должна быть представлена простым интерфейсом, состоящим из методов доступа к следующей информации:

- данным, идентифицирующим загруженную в память GPU шейдерную программу (например, в случае OpenGL это целочисленный идентификатор);
- значениям параметров, передаваемым в шейдерную программу.

Следует отметить, что предоставляемые данным интерфейсом значения перечисленных атрибутов должны быть готовы к непосредственному использованию с конкретным графическим API. Однако скрытая под ним реализация может иметь дополнительные возможности, упрощающие работу с сущностью для вспомогательных классов, описанных ниже.

После загрузки из персистентного представления шейдерная модель создает все необходимые вспомогательные объекты и производит их связывание с общей логикой рендеринга на основе своих метаданных.

Типичным набором вспомогательных объектов в данном случае являются:

– объекты, описывающие все требуемые для построения каждого кадра проходы (render passes);

– объект-валидатор, инжектируемый в аппликатор сущности *shader state* и определяющий, нужно ли сменить шейдерную программу, привязанную к данной сущности;

– объект, инжектируемый в менеджер шейдеров и вычисляющий по полям сущности *render state* ключ, идентифицирующий необходимую для нее шейдерную программу в рамках объекта – хранилища шейдеров;

– объект, инжектируемый в менеджер шейдеров и отвечающий за сборку и кэширование кода требуемой шейдерной программы в случае ее отсутствия в памяти GPU.

Следует отметить, что наиболее удобной и гибкой реализацией связывания любой из метамodelей, в том числе и шейдерной модели, с остальной логикой приложения является инъекция зависимостей в среде, подобной IoC-контейнеру.

Как было отмечено ранее, любые конфигурируемые метамodelи, описывающие позднее связывание компонентов приложения, удобно хранить в структурированном представлении. Ниже приведем примерный шаблон персистентного представления простейшей шейдерной модели освещения по Фонгу для OpenGL ES 2.x-3.x в формате XML.

Листинг-шаблон XML-представления шейдерной модели освещения по Фонгу имеет следующий вид:

```
<?xml version="1.0" encoding="UTF-8"?>
<shaderModel>
  <shader type="vertex">
    <declarations>
      <declaration
        name="u_MVMatrix"
        category="uniform"
        dataType="mat4"/>
      <declaration
        name="u_MVPMatrix"
        category="uniform"
        dataType="mat4"/>
      <declaration
        name="a_Position"
        category="attribute"
        dataType="vec4"/>
      <declaration
        name="v_Position"
        category="varying"
        dataType="vec3"/>
      <if target="lightingEnabled" value="true">
        <declaration
          name="a_Normal"
          category="attribute"
          type="vec4"/>
        <declaration
          name="v_Normal"
          category="varying"
          type="vec4"/>
      </if>
    </declarations>
  </body>
```

```
<if target="lightingEnabled" value="true">
  v_Normal = vec3(u_MVMatrix * vec4(a_Normal, 0.0));
</if>
v_Position = vec3(u_MVMatrix * a_Position);
gl_Position = u_MVPMatrix * a_Position;
</body>
</shader>
<shader type="fragment">
  <!-- аналогичные объявления declarations и body -->
</shader>
</shaderModel>
```

Заключение

В качестве компромиссного решения в работе предложена метамодель рендеринга, позволяющая достигать требуемого в конкретной ситуации баланса в отношении объема потребляемой памяти, расхода машинного времени, централизации управления родственными компонентами и обеспечения модульности. Концепция этой метамодели легко распространяется на любую функциональность при наличии минимальной поддержки ядром приложения возможностей ЮС-контейнера. Данное решение упрощает контроль над архитектурой сложных приложений, не имеет привязки к конкретной платформе и языку программирования. Что касается производительности, то здесь следует отметить возможность разработки дополнительных инструментов, которые на определенном этапе жизненного цикла приложения (например, перед интеграционным тестированием или перед релизом) генерируют обычный код, не использующий металогику.

Список литературы

1. Bailey, M. Graphics shaders: Theory and practice / M. Bailey, S. Cunningham. – 2nd ed. – A K Peters/CRC Press, 2011. – 518 p.
2. Zink, J. Practical rendering and computation with Direct3D 11 / J. Zink, M. Pettineo, J. Hoxley. – CRC Press, 2011. – 637 p.
3. Sellers, G. OpenGL superBible: comprehensive tutorial and reference / G. Sellers, R.S. Wright, N. Haemel. – 6th ed. – Crawfordsille : Addison-Wesley, 2013. – 848 p.
4. Smithwick, M. Pro OpenGL ES for Android / M. Smithwick, M. Verma. – N.Y. : Apress, 2012. – 308 p.

Поступила 08.04.2014

*Витебский государственный университет
им. П.М. Машерова,
Витебск, пр. Московский, 33
e-mail: o.g.kazantseva@gmail.com,
judzin.baranovsky@gmail.com*

V. Kazantsava, Y. Baranouski, Y. Landarski

SHADER METAMODEL AS A COMPONENT OF AN INTERACTIVE APPLICATION ARCHITECTURE

This article introduces the rendering metamodel as an architecture basis for complicated graphical applications. The model separates abstract logical and algorithmic approaches used in modern computer graphics theory from problems of computational resource management. It also yields wide possibilities for program code complexity management.

УДК 519.67; 004.896

В.А. Сычев

РЕАЛИЗАЦИЯ ДИСКРЕТНЫХ НЕЛИНЕЙНЫХ ДИНАМИЧЕСКИХ СИСТЕМ С ХАОТИЧЕСКИМИ РЕЖИМАМИ НА АРИФМЕТИКЕ С ФИКСИРОВАННОЙ ЗАПЯТОЙ

Рассматривается проблема практической реализации нелинейных систем с хаотической динамикой для целенаправленной генерации хаотических последовательностей в цифровых устройствах. Исследуется возможность применения в данной задаче арифметики с фиксированной запятой для обеспечения идентичности результатов, полученных на различных программно-аппаратных платформах. Описывается реализация логистического отображения, проводится анализ полученных результатов. Предлагается применение полученных результатов для решения различных задач из области мобильной робототехники.

Введение

В последние три десятилетия возрос интерес исследователей к системам, демонстрирующим хаотические режимы работы. Такие известные ученые, как И. Пригожин, Э. Лоренц, а также Г. Хакен, В. Эбелинг и др., показали, что новый порядок (самоорганизация) в сложной системе возникает через динамический хаос – хаотический режим функционирования сложной системы [1, 2].

Новые знания о хаотической динамике нашли свое применение для решения ряда практических задач, которые можно подразделить на две основные группы. Первая – это задача стабилизации нелинейных динамических систем (НДС), включая и сложные НДС, т. е. предотвращения возникновения нежелательных хаотических колебаний в них [3]. Вторая задача обратна первой и состоит в целенаправленной генерации хаотических колебаний. Вслед за теоретическими исследованиями появляются практические примеры использования знаний о хаотических режимах НДС для решения инженерных задач [4]. Методы стабилизации НДС нашли применение для управления электромеханическими приводами, оптическими системами, нелинейными электронными цепями. Хаотические процессы используются в задачах прогнозирования, обработки и защиты информации, управления и планирования [2–7].

Задача целенаправленной генерации хаотических колебаний возникает и в аналоговой, и в цифровой технике. В то время как в аналоговых электронных приборах хаотические сигналы генерируются специальными устройствами, в числе которых наиболее известна схема Чуа [8], в вычислительной технике хаотические последовательности генерируются путем решения одного из известных уравнений или системы уравнений с параметрами, обеспечивающими хаотическую динамику [4, 5].

Важным условием при генерации хаотических последовательностей, особенно в задачах коммуникации, является точность вычислений. Недостаточно высокая точность приводит к постепенному переходу НДС из хаотического в периодический режим либо вовсе препятствует возникновению хаотического режима. Минимально необходимый уровень точности определяется для каждой НДС в отдельности.

Существует множество способов представления в памяти ЭВМ вещественных чисел и алгоритмов выполнения вычислений. В результате вычисления, произведенные на различных платформах, могут давать неодинаковый результат. Следовательно, важна переносимость решений на различные программно-аппаратные платформы, также называемая кроссплатформенностью.

Одним из способов достижения идентичности результатов на разных платформах является стандартизация алгоритмов представления и обработки данных. Однако в процессе разработки кроссплатформенных алгоритмов нередко возникает необходимость в использовании программно-аппаратных платформ, не соответствующих существующим стандартам. Напри-

мер, большинство микропроцессоров, выпускаемых в мире, не имеют аппаратной поддержки вычислений с плавающей запятой. Следовательно, алгоритм работы с числами с плавающей запятой реализуется программным путем. Чаще всего программная реализация ориентирована на достижение высокой производительности в ущерб точности вычислений и переносимости программ.

Общепринятым подходом к достижению требуемой точности и переносимости разработок является применение арифметики с фиксированной запятой [9] (в англоязычной литературе используется термин *fixed-point arithmetic*). Алгоритм генерации хаотических последовательностей, реализованный на целочисленной арифметике, позволит получить результат, воспроизводимый на различных программно-аппаратных платформах с требуемой точностью. Помимо прочего, такой алгоритм обеспечит большую скорость генерации хаотических последовательностей.

Целью данной работы является исследование моделей дискретных НДС с хаотическими режимами, реализованных на арифметике с фиксированной запятой, для их использования в системах управления робототехническими аппаратами.

1. Свойства дискретных нелинейных систем с хаотической динамикой

В отличие от случайных процессов динамику системы принято называть хаотической, если на нее не оказывают влияние случайные или непредсказуемые величины и параметры. Помимо этого, хаотическая динамика отличается рядом особых свойств, в числе которых чувствительность к начальным условиям, потеря информации о начальных условиях, образование фрактальных структур, способность к самосинхронизации [5]. Именно эти свойства хаотической динамики находят применение в инженерных задачах [10].

Хаотические режимы работы обнаруживаются в различных НДС с непрерывным и дискретным временем. К числу широко известных и всесторонне исследованных НДС с непрерывным временем относятся аттракторы Лоренца, Ресслера и др., представляющие собой систему дифференциальных уравнений, решение которой с определенными параметрами позволяет наблюдать хаотическую динамику [5]. Данные аттракторы могут реализовываться на аналоговой элементной базе и использоваться как источник хаотических сигналов. Однако с точки зрения практической реализации НДС в качестве источника хаотических последовательностей наибольший интерес представляют дискретные системы как не требующие дополнительных вычислительных затрат и дополнительных устройств для получения хаотических последовательностей.

Одним из примеров систем с хаотической динамикой является логистическое отображение [10]

$$x_i = a - x_{i-1}^2. \quad (1)$$

Другой вариант дискретной НДС с хаотической динамикой представлен аттрактором Энона [5]:

$$\begin{aligned} x_{i+1} &= 1 + y_i - ax_i^2; \\ y_{i+1} &= bx_i. \end{aligned} \quad (2)$$

Хаотические режимы наблюдаются и в такой дискретной НДС, как кубическое отображение (отображение Холмса), что также описано в [5]:

$$\begin{aligned} x_{i+1} &= y_i; \\ y_{i+1} &= -bx_i + dy_i - y_i^3. \end{aligned} \quad (3)$$

Параметры данных НДС, при которых наблюдаются хаотические режимы работы, и измеренные показатели Ляпунова (старшие экспоненты Ляпунова) для каждой из них приведены в табл. 1.

Таблица 1

Дискретные отображения и их характеристики

Отображение	Параметры хаотизации	Показатель Ляпунова
Логистическое	$a = 1,99$	0,6607
Аттрактор Энона	$a = 1,4; b = 0,3$	0,4192
Кубическое	$b = 0,2; d = 2,77$	0,5323

Для вычисления показателя Ляпунова был использован алгоритм М.Т. Розенштейна [11], реализованный в системе компьютерной математики (СКМ) Matlab [12].

Исследуя отображение (1), можно определить, что ряд значений x_i , вычисленный несколько раз на компьютере при равных начальных условиях, но с использованием различных микропроцессоров, компиляторов, языков программирования, может различаться. Так как отображение (1) в соответствии с общим свойством хаотических систем обладает высокой чувствительностью к начальным условиям, малейшее различие в результатах вычислений в итоге приводит к значительному расхождению результатов.

Иллюстрацией может служить ряд значений, полученных итерационным решением формулы (1) с начальными условиями $x_1 = 0,5$ и $a = 1,99$. В первом случае значения были получены с помощью программы на языке Си, откомпилированной компилятором *icc*, во втором случае – компилятором *gcc* с параметрами, установленными по умолчанию при инсталляции. В остальных условиях исполнения программ оставались идентичными. Вычисления проводились с использованием чисел с плавающей запятой в формате *double*. На графике (рис. 1, а) показана разность между полученными в обоих случаях значениями.

На 59-м шаге разница достигает уровня, при котором значения полностью расходятся (рис. 1, б).

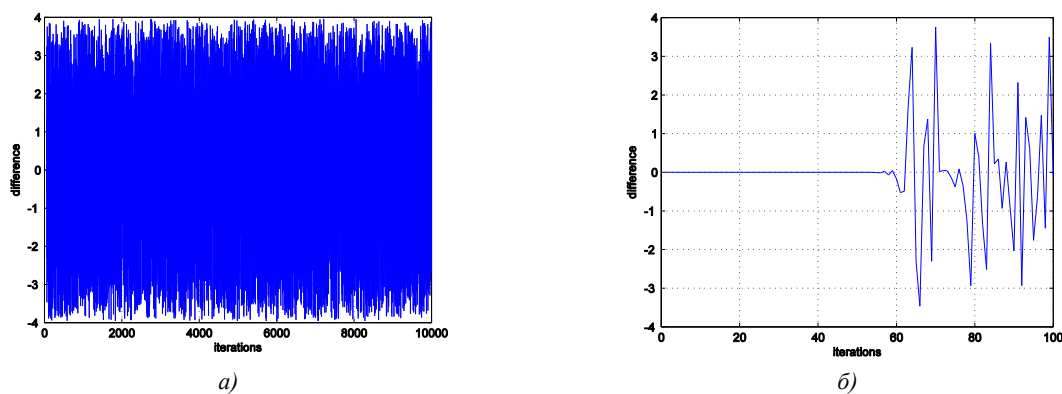


Рис. 1. Разность между значениями, полученными с использованием компиляторов *icc* и *gcc*: а) на всем диапазоне значений; б) начало расхождения в значениях, сгенерированных с использованием данных компиляторов

Таким образом, при использовании того или иного отображения в качестве источника хаотических последовательностей на различных программно-аппаратных платформах необходимо принять дополнительные меры, обеспечивающие идентичность получаемых результатов.

2. Особенности применения арифметики с фиксированной запятой

Современные микропроцессоры, применяемые в персональных компьютерах, позволяют на аппаратном уровне выполнять вычисления с плавающей запятой. Способы представления и обработки таких чисел описываются стандартом IEEE754 [9, 13], однако не все компиляторы точно соответствуют требованиям данного стандарта, что объясняет расхождение результатов вычислений, показанное на рис. 1.

В то же время в системах управления мобильными роботами применяются энергоэкономичные микропроцессоры с ограниченными вычислительными ресурсами. Вычисления с пла-

вающей запятой такими процессорами на аппаратном уровне могут не поддерживаться или поддерживаться ограниченно. Компиляторы позволяют использовать данные типа *float* и *double*, однако соответствие стандарту IEEE754 ими не гарантируется.

Для решения всех описанных выше задач, в которых используются хаотические процессы, необходимо применение генератора хаотических последовательностей, обеспечивающего повторяемость результата на различных программно-аппаратных платформах. Эта цель может быть достигнута реализацией генератора хаотических последовательностей на основе вычислений с фиксированной запятой [9].

Однако вследствие ограниченной разрядности регистров ЭВМ при выполнении операций в формате с фиксированной запятой возникает ряд нежелательных эффектов, в числе которых шум усечения или образование предельных циклов [9, 14, 15]. Шум усечения может быть как коррелирован, так и не коррелирован с генерируемой хаотической последовательностью чисел. Образование же предельных циклов происходит как вследствие влияния шума усечения, так и по причине переполнения регистров [9].

В настоящей работе для обозначения чисел с фиксированной запятой будет использоваться так называемая *Q*-нотация вида *Qm.n*, где *m* – число бит, используемых для хранения целой части числа; *n* – число бит, хранящих дробную часть. Числа в формате с фиксированной запятой в соответствии с [16] будут обозначаться индексом *fi*.

Для перевода вещественного десятичного числа в двоичный формат с фиксированной запятой целую часть вещественного числа переводят в двоичную систему счисления методом последовательного деления, а дробную часть – методом последовательного умножения [18]. Полученные результаты объединяют.

Число $a = 1,99$, преобразованное в двоичный формат с одним битом для хранения целой части и 17 битами для хранения дробной части, будет иметь вид $a_2 = 1,11111101011100001$. В данном преобразовании отброшены биты младше 17-го, т. е. произведено усечение числа. Обратное преобразование уже не даст изначальное число $a = 1,99$ по причине потери точности вследствие усечения исходного числа. Разница между числом до усечения и после является шумом усечения [15]:

$$\Delta_y = a - \hat{a}, \quad (4)$$

где Δ_y – шум усечения; a – число до усечения; \hat{a} – число после усечения.

Изменяя величину q , удается достичь приемлемого уровня шума. При этом существуют числа, которые не могут быть точно представлены в двоичном формате.

Помимо усечения, может быть использовано округление чисел. Усечение и округление применяются к числам как с фиксированной, так и с плавающей запятой.

Математическое ожидание шума округления равно нулю [15]. Дисперсия шума квантования при округлении вычисляется по формуле

$$D(\Delta_o) = \frac{2^{-2l}}{12}, \quad (5)$$

где l – разрядность регистра, хранящего число.

При усечении дисперсия шума квантования увеличивается:

$$D(\Delta_y) = \frac{2^{-2l}}{3}. \quad (6)$$

Операции над числами с фиксированной запятой производятся так же, как и над любыми другими числами, представленными в экспоненциальной форме.

Таким образом, при реализации кроссплатформенных генераторов хаотических последовательностей на базе цифровых ЭВМ может быть использована арифметика с фиксированной запятой. Влияние же ограниченной разрядности регистров ЭВМ и значительной дисперсии шума (8) оценивается экспериментально.

3. Экспериментальное исследование дискретных отображений в арифметике с фиксированной запятой

Исследование возможности применения целочисленной арифметики в задачах генерации хаотических последовательностей было проведено на примере отображений, представленных в табл. 1. Исследование включало два этапа. На первом этапе каждое отображение было реализовано в виде компьютерной программы на языке Си и языке СКМ Matlab. При этом были использованы только стандартные целочисленные типы данных длиной от 8 до 64 бит. Специализированные функции для выполнения операций над числами, представленными в формате с фиксированной запятой, не использовались. Однако при данных ограничениях логистическое отображение переходило из хаотического режима в периодический после 334-го шага, аттрактор Эннона – после шестого шага, кубическое отображение стабилизировалось после 39-го шага. В то же время расширенные возможности, предоставляемые функциями из набора *Fixed-point toolbox* СКМ Matlab или аналогичных наборов функций иных языков программирования, позволяют реализовать хаотические режимы отображений, представленных в табл. 1, на значительном числе отсчетов. Таким образом, на втором этапе исследования были применены функции из набора *Fixed-point toolbox* СКМ Matlab, позволившие определить такие длины целой и дробной частей переменных, при которых хаотический режим наблюдался на достаточно большом – до одного миллиона – числе отсчетов для каждого из отображений.

Отображения были реализованы в виде компьютерной программы на арифметике с фиксированной запятой на языке СКМ Matlab.

При выполнении программы была получена хаотическая последовательность данных. Показатель Ляпунова, измеренный для этой последовательности, как описано в разд. 2, и состоящей из 100 значений, составил $\lambda_1 = 0,6712$.

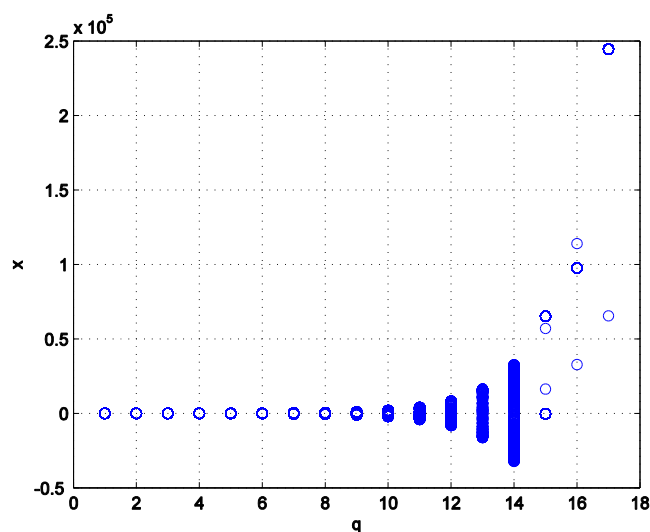


Рис. 2. Влияние величины q на результат вычислений значений логистического отображения

При указанных выше параметрах вычислений хаотический режим наблюдается при величине дробной части числа $q=13$ и $q=14$. Недостаточно высокая точность вычислений ниже указанных значений позволяет данной НДС сохранять стабильность. При значениях $q=15$ и выше и при использовании стандартных типов данных происходит переполнение регистров ЭВМ, вызывающее возникновение предельных циклов (рис. 2). Сравнение результатов вычислений, проведенных в форматах с фиксированной и плавающей запятой, показано на рис. 3.

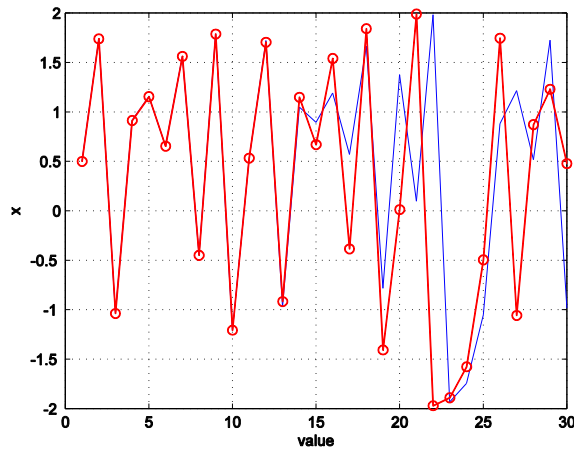


Рис. 3. Результат вычисления логистического отображения с плавающей запятой (красная линия с круглыми маркерами) и фиксированной запятой (сплошная синяя линия)

Как видно из рис. 3, до 13-го шага результаты полностью совпадают, после чего начинается расхождение. В формате $Q1.14$ после 334-го шага происходит переход от хаотического к периодическому режиму вследствие недостаточно высокой точности.

Бифуркационная диаграмма отображения (1), полученная при выполнении вычислений с фиксированной запятой (рис 4, а), напоминает классическую бифуркационную диаграмму логистического отображения [10, 17]. Виден переход к хаотическому режиму через удвоение периода. При каждой смене значения a новое значение x устанавливается не сразу (рис. 4, б).

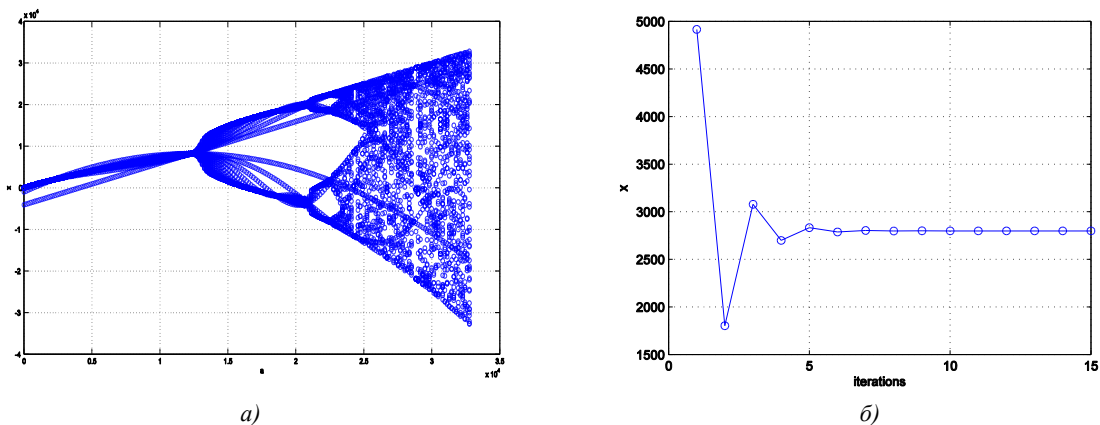


Рис. 4. Влияние управляющего параметра a на значение x : а) бифуркационная диаграмма логистического отображения; б) установление значения x при изменении величины a

Программа на языке Си, реализующая вычисления логистического отображения в формате с фиксированной запятой, позволяет получить повторяемый результат с использованием различных компиляторов не только на персональном компьютере, но и на микроконтроллерных и встраиваемых системах. В процессе испытаний она была откомпилирована с использованием компиляторов *icc* и *gcc*, выполнялась на персональном компьютере 64-битным процессором и восьмибитном микроконтроллере. Во всех случаях результат точно соответствовал тому, что был получен в СКМ Matlab.

Недостатком данной программы, как и описанной выше программы на языке СКМ Matlab, является переход от хаотического к периодическому режиму после 334-й итерации по причине недостаточно высокой точности. Преодолеть данный недостаток можно путем использования в программе нестандартных типов данных длиной более 64 бит. Пример програм-

мы на языке СКМ Matlab, использующей 128 бит для хранения числа и построенной на основе *Fixed-point toolbox*, приведен в табл. 2.

Таблица 2
Текст программы на языке СКМ Matlab с использованием *Fixed-point toolbox* для вычисления логистического отображения

Номер строки	Текст программы
1	<code>clear all; close all;</code>
2	<code>wordLength = 128;</code>
3	<code>fractionLength = 126;</code>
4	<code>a = fi(1.99,1,wordLength,fractionLength);</code>
5	<code>X = fi(1:50000,1,wordLength,fractionLength);</code>
6	<code>startX = fi(0.5,1,wordLength,fractionLength);</code>
7	<code>X(1) = startX;</code>
8	<code>for i = 2:50000</code>
9	<code> X(i) = a - X(i-1)^2;</code>
10	<code>end</code>

Графики по переменной x , иллюстрирующие выход из хаотического режима аттрактора Энона и кубического отображения при вычислении в формате $Q1.14$, показаны на рис. 5.

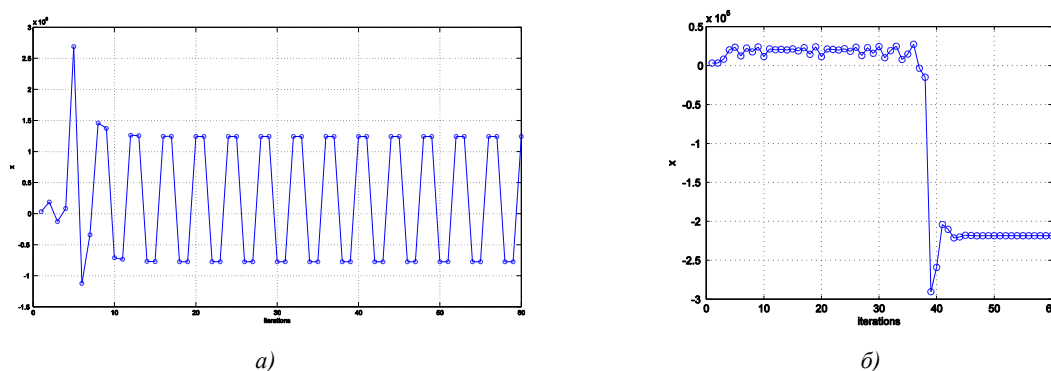


Рис. 5. Графики последовательностей x : а) для аттрактора Энона; б) кубического отображения

При $Q1.14$ аттрактор Энона демонстрирует возникновение предельного цикла по причине переполнения регистров на шестом шаге при вычислении нового значения x . При вычислении кубического отображения переполнение происходит на 39-м шаге.

Для достижения хаотического режима аттрактором Энона и кубическим отображением при реализации на арифметике с фиксированной запятой требуется повышение точности вычислений. Достижение хаотического режима возможно также путем применения метода динамического изменения формата данных, при котором количество бит, хранящих целую и дробную части, изменяется в процессе выполнения программы для исключения переполнений.

По результатам моделирования логистического отображения с использованием *Fixed-point toolbox* хаотический режим на одном миллионе отсчетов наблюдался при $Q2.41$.

Заключение

В работе описана реализация логистического отображения с использованием арифметики с фиксированной запятой, проведено исследование полученных результатов. Отличительной особенностью данного отображения является высокая чувствительность к начальным условиям (старшая экспонента Ляпунова равна 0,6607). Другая особенность состоит в том, что диапазон значений x не превышает пределов от -2 до 2 , благодаря чему удалось использовать только два бита для кодирования целой части получаемых результатов.

Целочисленная реализация других типов дискретных НДС, описанных выше, с теми же параметрами точности и типами данных показала, что для решения таких отображений требуется более высокая точность вычислений, чем при решении логистического отображения.

Полученный результат может быть использован в ряде практических задач. Так, в частности, в работе [19] описывается общая схема поискового адаптивного поведения с инерционным переключением между поисковыми тактиками. Как было указано в данной работе, эффективное поисковое движение может быть реализовано сочетанием перемещения на значительные расстояния, частыми случайными переменами в направлении движения и инерционностью переключения между каждым из видов движения. Данный способ поискового поведения может быть эффективно использован в области мобильной робототехники для управления мобильным роботом, выполняющим задачу поиска на местности некоторых предметов, явлений или условий. При этом описываемый способ поискового поведения удобно реализовать в бортовой системе управления мобильного робота именно на основе целочисленного генератора хаотических последовательностей.

Интерес представляет также возможность криптозащиты информационного обмена по радиоканалу между роботами и системой управления вышестоящего уровня с помощью хаотических последовательностей. В этом случае важную роль играет кроссплатформенность целочисленного подхода, что позволит организовать взаимодействие между вычислительными системами различных типов.

Работа выполнена при поддержке гранта БРФФИ–ГФФИУ № Ф13К-144 «Разработка методов оперативной обработки и передачи информации для эффективного управления мобильными роботами и подвижными системами».

Список литературы

1. Пригожин, И. Порядок из хаоса / И. Пригожин, И. Стенгерс. – М. : Прогресс, 1986. – 431 с.
2. Малинецкий, Г.Г. Хаос. Структуры. Вычислительный эксперимент: введение в нелинейную динамику / Г.Г. Малинецкий. – М. : Эдиториал УРСС, 2000. – 256 с.
3. Handbook of Chaos Control / E. Schöll [et al.]. – 2nd ed. – Wiley-VCH Verlag GmbH&Co.KGaA, 2008. – 819 p.
4. Chaos in Automatic Control / W. Perruquetti [et al.] ; ed. W. Perruquetti. – CRC Press, 2005. – 564 p.
5. Moon, F. Chaotic Vibrations: An Introduction for Applied Scientists and Engineers / F. Moon. – John Wiley&Son, 2004. – 309 p.
6. Basic principles of direct chaotic communications / A.S. Dmitriev [et al.] // Nonlinear Phenomena in Complex Systems. – 2003. – Vol. 6, no. 1. – P. 488–501.
7. Хаотические процессоры / Ю.В. Андреев [и др.] // Успехи современной радиоэлектроники. – 1997. – № 10. – С. 50–79.
8. Nonlinear Dynamics in Circuits / T. Carroll [et al.] ; ed. T. Carroll. – World Scientific, 1995. – 344 p.
9. Рабинер, Л. Теория и применение цифровой обработки сигналов / Л. Рабинер, Б. Гоулд. – М. : Мир, 1978. – 848 с.
10. Gilmore, R. The Topology of Chaos / R. Gilmore, M. Lefranc. – John Wiley & Sons, 2008. – 518 p.
11. A practical method for calculating largest Lyapunov exponents from small data sets / M.T. Rosenstein [et al.] // Physica D 65 1993. – P. 117–134.
12. Largest Lyapunov Exponent with Rosenstein's Algorithm // Matlab Central [Electronic resource]. – Mode of access : <http://www.mathworks.com/matlabcentral/fileexchange/38424-largest-lyapunov-exponent-with-rostensteins-algorithm/content/lyarosenstein.m>. – Date of access : 30.03.2014.
13. Standard for Binary Floating-Point Arithmetic : IEEE 754–2008. – Institute of Electrical and Electronics Engineers, 2008. – 23 p.
14. Оппенгейм, А. Цифровая обработка сигналов / А. Оппенгейм, Р. Шафер. – М. : Техносфера, 2006. – 856 с.

15. Передача дискретных сообщений : учебник для высших учебных заведений / В.П. Шувалов [и др.] ; под ред. В.П. Шувалова. – М. : Радио и связь, 1990. – 464 с.
16. Wilkinson, J.H. Rounding Errors in Algebraic Processes / J.H. Wilkinson. – Dover Publications, 1994. – 161 p.
17. Korsch, H.J. Chaos. A program collection for PC / H.J. Korsch, H-J. Jodl, T. Hartmann. – 3rd ed. – Springer, 2008. – 341 p.
18. Путков, В.Н. Электронные вычислительные устройства : учеб. пособие для радиотехн. спец. вузов / В.Н. Путков, И.И. Обросов, С.В. Бекетов. – Минск : Вышэйшая школа, 1981. – 333 с.
19. Непомнящих, В.А. Бионическая модель адаптивного поискового поведения / В.А. Непомнящих, Е.Е. Попов, В.Г. Редько // Известия РАН. Теория и системы управления. – 2008. – № 1. – С. 85–93.

Поступила 03.05.2014

*Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: vsychyov@robotics.by*

U.A. Sychou

REALIZATION OF DISCRETE NONLINEAR DYNAMICAL SYSTEMS WITH CHAOTIC REGIMES BASED ON FIXED-POINT ARITHMETIC

In this article, the problem of the practical realization of nonlinear systems with chaotic dynamics for targeted generation of chaotic sequences in digital devices is considered. The possible application in this task with using fixed-point arithmetic to ensure the identity of the obtained results on different hardware and software platforms is studied. The implementation of logistic mapping is described; carry out the analysis of the results. This article proposes using the obtained results for the various tasks of the field of mobile robotics.

УДК 658.512.22.011.56:004(076.5)

А.В. Бородуля¹, А.Н. Соловьев², В.В. Напрасников¹, М.А. Мирзаванд¹

ВЫБОР РАЦИОНАЛЬНЫХ ПАРАМЕТРОВ КОНСТРУКЦИИ КАРКАСА МОРСКОЙ ПЛАТФОРМЫ

Рассматривается создание параметрической геометрической модели конструкции каркаса морской нефтедобывающей платформы с использованием специальных встроенных языков систем конечно-элементного моделирования, а также конечно-элементной модели этой конструкции с учетом эксплуатационных нагрузок. Описывается подготовка оптимизационной модели конструкции, выполняются оптимизационные расчеты и даются рекомендации по выбору рациональных параметров конструкции.

Введение

Потребности современного индустриального общества в углеводородном сырье постоянно растут. Одной из возможностей их удовлетворения являются разведка и добыча углеводородов на морском шельфе. В настоящее время в Исламской Республике Иран ведутся активные работы по разведке и добыче углеводородов на морском шельфе. При выполнении проектирования оборудования для таких работ возникают задачи, специфика которых предполагает необходимость разработки специальных моделей, позволяющих оценивать работоспособность, долговечность, ремонтпригодность сложных технических систем добычи и транспортировки углеводородов с учетом современных требований экологии и снижения их материалоемкости [1–3].

Цель настоящей работы – создание конечно-элементной модели конструкции каркаса морской нефтедобывающей платформы с учетом эксплуатационных нагрузок, на основе которой проектировщик получит возможность принимать обоснованные решения по выбору варианта конструкции минимальной материалоемкости.

Задачами исследования являются:

- создание параметрической геометрической модели конструкции каркаса морской нефтедобывающей платформы с использованием специальных встроенных языков систем конечно-элементного моделирования;
- создание конечно-элементной модели этой конструкции с учетом эксплуатационных нагрузок;
- подготовка оптимизационной модели конструкции;
- выполнение оптимизационных расчетов и выработка рекомендаций по выбору рациональных параметров конструкции.

1. Моделирование и решение задачи

Расчетная схема конструкции каркаса морской нефтедобывающей платформы изображена на рис. 1. Каркас остова изготовлен из шести типов труб, диаметры которых 4,26–8,208 м, а соответствующая толщина труб 0,12–0,206 м.

Для моделирования каркаса конструкции используется конечный элемент, расчетная схема которого показана на рис. 2, а, а для учета массы рабочего оборудования на верхнем уровне каркаса платформы – конечный элемент, расчетная схема которого показана на рис. 2, б.

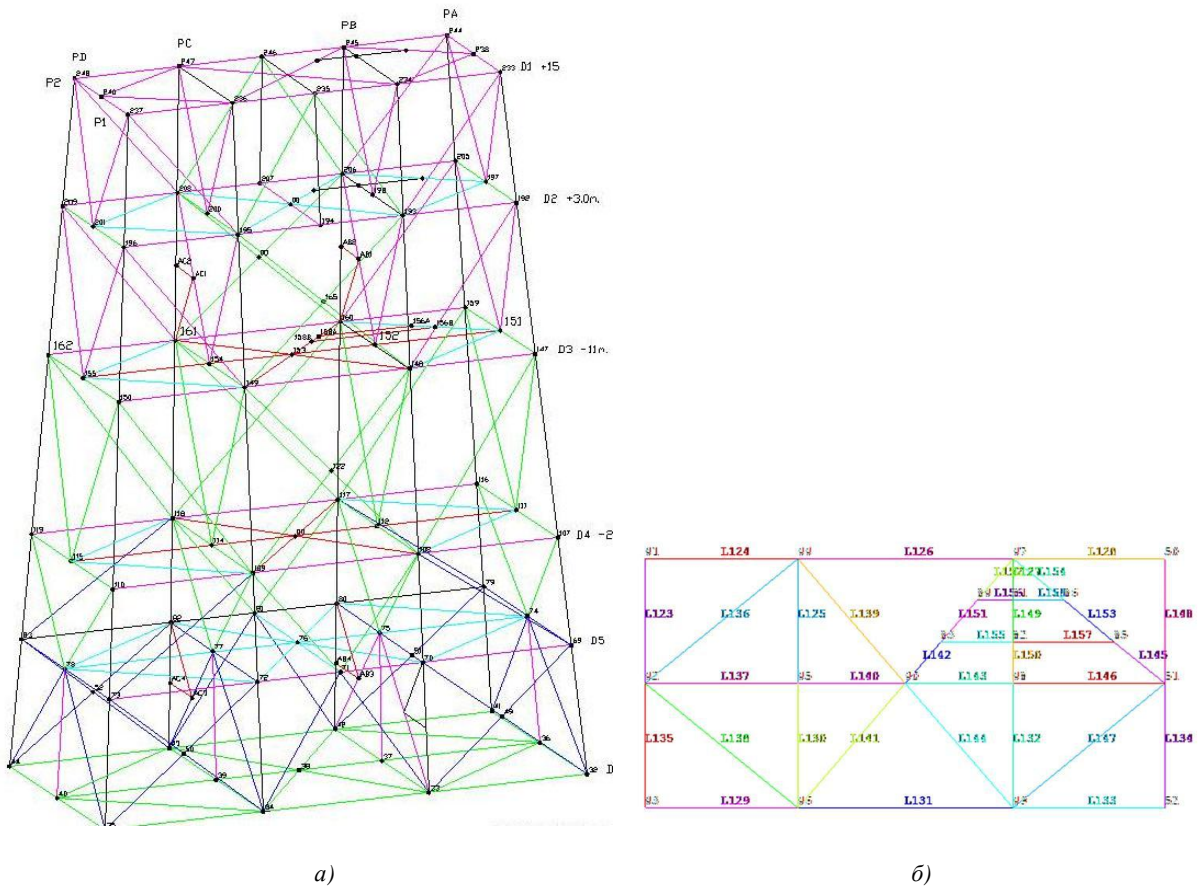


Рис. 1. Параметрическая геометрическая модель: а) остова; б) его подводного четвертого уровня

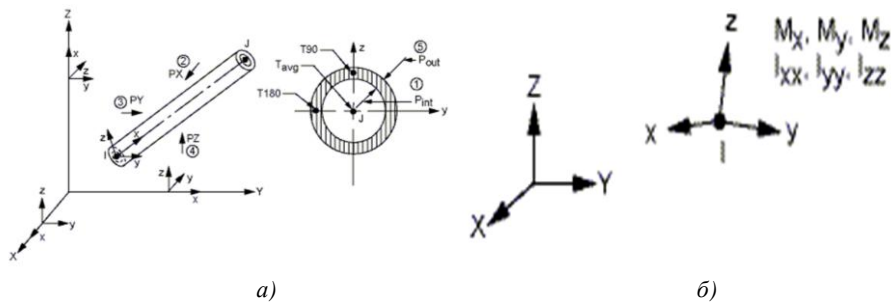


Рис. 2. Расчетные схемы конечных элементов: а) трубы; б) сосредоточенной массы

В модели данные величины и длина базовой части, ширина нулевого и пятого уровней, добавочная длина на пятом уровне, высоты всех уровней являются параметрами.

Численные эксперименты на основе предложенной модели показали:

– максимальные напряжения по Мизесу не превышают 25 МПа (рис. 3, а), что существенно ниже предела текучести материала конструкции. При этом наиболее нагруженными оказываются стойки каркаса;

– максимальные перемещения не превышают 0,0144 м (рис. 3, б), что допустимо по технологическим соображениям;

– коэффициенты запаса по устойчивости для первых шести форм составляют 54,908; 58,03; 60,085; 60,496; 61,214 и 61,398 соответственно.

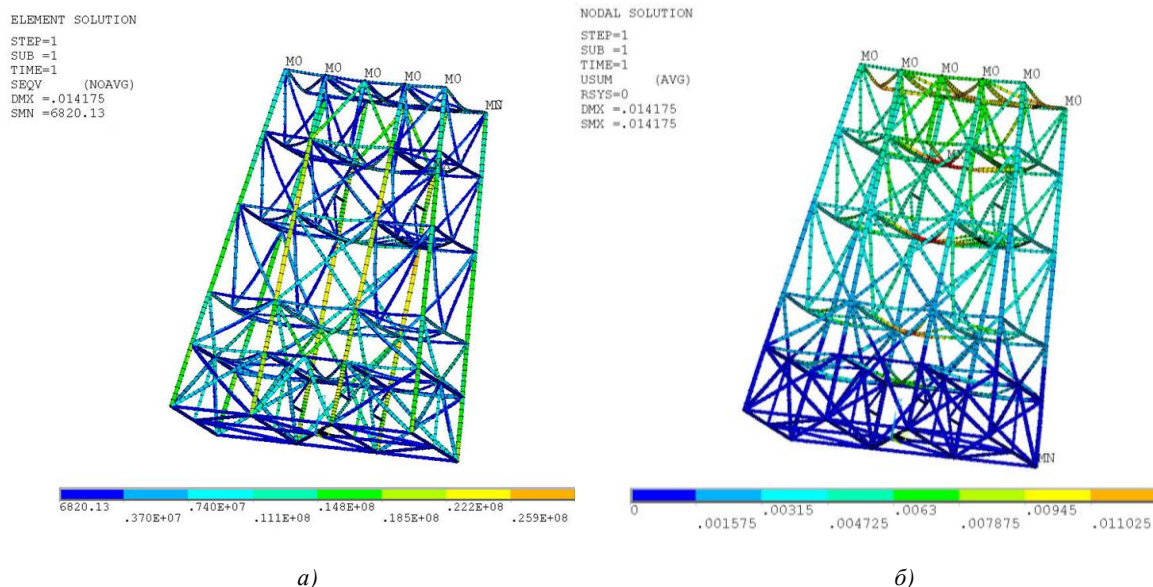


Рис. 3. Результаты моделирования: а) напряженное состояние конструкции; б) деформированное состояние конструкции

Таким образом, выявлен значительный запас прочности в материале конструкции. Это позволяет перейти к поиску более рационального варианта по критерию материалоемкости.

2. Оптимизация

Задачу оптимизации будем ставить следующим образом: найти $\min_{X \in D} F(X)$, где D – область пространства, в которой ведется поиск; X – m -мерный вектор переменных; $F(X)$ – заданная функция n переменных, определенная в области D . Предполагается, что функция $F(X)$ непрерывна в области D . Будем считать, что область D является m -мерным гиперпараллелепипедом с гранями, параллельными координатным осям:

$$D = \{X \in R^m \mid x_{i \text{ inf}} \leq x_i \leq x_{i \text{ sup}}, \quad i = \overline{1, m}\}. \quad (1)$$

Наложим ограничения на множество векторов X , ограничив их m -мерным гиперпараллелепипедом. В общем случае оптимальное значение X в зависимости от вида вектор-функции $F(X)$ может быть точкой либо локального, либо глобального минимума.

Определим вектор X^* как точку локального минимума, если для всех точек X^* , принадлежащих ε -окрестности точки X^* , выполняется следующее неравенство: $F(X^*) \leq F(X)$, где $X \in D_i$, $D_i \subset D$, $D_i = \{X \in R^m \mid x_i - \varepsilon \leq x_i \leq x_i + \varepsilon, \quad i = \overline{1, m}\}$.

Определим глобальный минимум $F(X)$ как наименьший из всех локальных, т. е. точка X^* является точкой глобального минимума на области, определенной выше, если справедливо неравенство

$$F(X^*) \leq F(X), \quad \forall X \in D.$$

Анализ литературы и опыт разработок показали, что наиболее эффективным методом поиска экстремума для данных моделей с точки зрения машинной реализации является расчет функционала на сетке точек с последующим уточнением полученного экстремума. Для построения сетки точек, равномерно заполняющих данную область, удобно использовать

$LP\tau$ -последовательности [4]. Следует отметить, что среди известных в настоящее время равномерно распределенных последовательностей наилучшие характеристики при $M \rightarrow \infty$, где M – количество точек, имеют $LP\tau$ -последовательности. Подробно методы генерации и доказательство равномерности распределения $LP\tau$ -последовательностей описаны в [4–6].

Рассмотрим область C – m -мерный единичный гиперкуб:

$$C = \{X \in R^m \mid 0 \leq x_i \leq 1, \quad i = \overline{1, m}\}.$$

Пусть M – количество точек ω_j , $(j = \overline{1, M})$, принадлежащих области C . Из свойств равномерного распределения $LP\tau$ -последовательностей следует, что выполняется условие $\lim_{M \rightarrow \infty} \frac{M_G}{M} = \sqrt{G}$, где G – любая m -мерная область, подобласть C ; M_G – количество точек $LP\tau$ -последовательности, принадлежащих области G ; M – количество точек $LP\tau$ -последовательности; \sqrt{G} – объем m -мерной области G [4]. При больших значениях M количество точек, попавших в произвольную область G , пропорционально объему этой области.

Сохраняя равномерность расположения точек, с помощью линейного преобразования точки ω_j , $j = \overline{1, M}$, равномерно заполняющие m -мерный единичный гиперкуб C , можно преобразовать в точки, равномерно заполняющие произвольный гиперпараллелепипед D , который удовлетворяет условиям (1).

Сформулируем оптимизационную задачу следующим образом:

критерий оптимизации (целевая функция) $F(X)$ – объем материала конструкции;

варьируемые параметры (переменные конструкции) X – m -мерный вектор переменных. Его компоненты – диаметры трубы первого типа (стойки) и трубы второго типа (наиболее массивной перемычки) – могут изменяться в пределах от 0,5 до 1,0 м;

функциональное ограничение (переменная состояния) – максимальное напряжение по Мизесу в материале конструкции – не должно превышать заданный уровнем 200 МПа.

Результаты поиска представлены ниже. Здесь в качестве примера отображены вычисленные значения для вариантов 1, 2, 3, 17, 18, 19. Из 21 рассмотренного варианта наилучшим оказался вариант 17:

	SET 1 (FEASIBLE)	SET 2 (FEASIBLE)	SET 3 (FEASIBLE)
STRESS_MAX (SV)	0.24544E+08	0.22654E+08	0.27719E+08
D1 (DV)	0.82080	0.92359	0.69400
D2 (DV)	0.72000	0.72859	0.84618
TOTAL_VOL (OBJ)	110.64	115.52	107.75
	SET 17 (FEASIBLE)	SET 18 (FEASIBLE)	SET 19 (FEASIBLE)
STRESS_MAX (SV)	0.27701E+08	0.26867E+08	0.28737E+08
D1 (DV)	0.68390	0.72657	0.65904
D2 (DV)	0.51099	0.92147	0.80708
TOTAL_VOL (OBJ)	99.604	110.96	105.26

На рис. 4 представлена динамика изменения объема конструкции в зависимости от номера итерации. При этом максимальное напряжение по Мизесу в материале конструкции увеличилось до 27 МПа, максимальные перемещения увеличились до 0,0156 м, а объем уменьшился до 99,604 м³ против 110,64 м³ в исходном варианте.

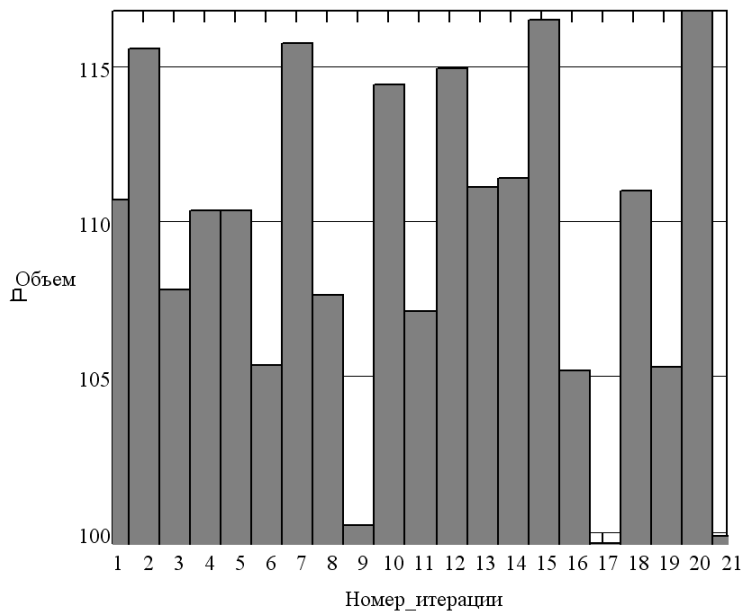


Рис. 4. Динамика изменения объема конструкции

Проверим коэффициенты запаса устойчивости для первых шести форм потери устойчивости. В оптимальном варианте они составляют 38,141; 40,167; 40,469; 41,389 и 41,788 соответственно, что обеспечивает достаточную устойчивость конструкции (рис. 5).

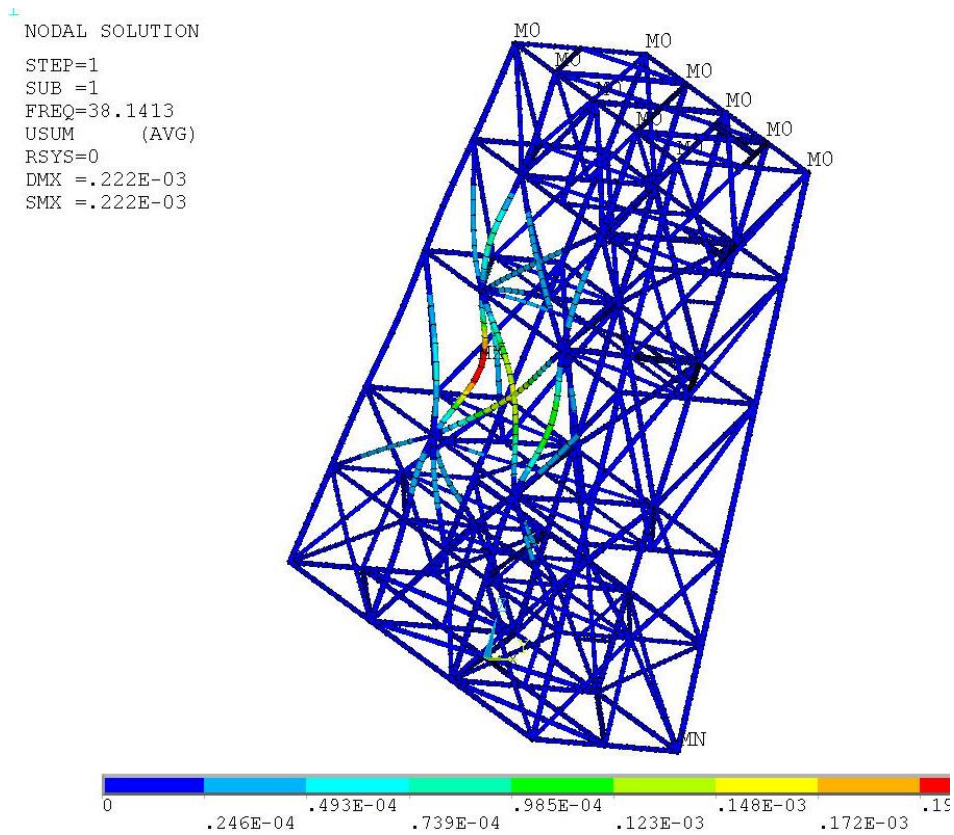


Рис. 5. Первая форма потери устойчивости оптимальной конструкции

Заключение

В ходе выполнения работы получены следующие результаты:

– создана параметрическая геометрическая модель конструкции каркаса остова морской нефтедобывающей платформы с использованием языка APDL системы конечно-элементного моделирования ANSYS;

– создана конечно-элементная модель данной конструкции с учетом эксплуатационных нагрузок, при этом для моделирования каркаса конструкции используется конечный элемент PIPE20, а для учета массы рабочего оборудования на верхнем уровне каркаса платформы – конечный элемент MASS21;

– подготовлена оптимизационная модель конструкции. Из 23 параметров конечно-элементной модели в качестве варьируемых выбраны два, к которым наиболее чувствительна результирующая масса конструкции, – это диаметры трубы первого типа (стойки) и трубы второго типа (наиболее массивной перемычки);

– выполнены оптимизационные расчеты и выработаны рекомендации по выбору рациональных параметров конструкции. При этом оказалось, что диаметр трубы первого типа следует выбрать размером 0,68390 м, а диаметр трубы второго типа – размером 0,51099 м;

– выполнена проверка оптимального варианта на устойчивость. Установлено, что коэффициент запаса устойчивости для первой формы составляет 38,141;

– экономия материала (стали) в предложенном варианте конструкции составляет более 10 м³.

В дальнейшем предполагается выполнить проверку оптимального варианта конструкции на сопротивление сейсмическим нагрузкам, характерным для регионов Ирана, в которых предполагается использование этих систем.

Представленный спектр конечно-элементных моделей позволяет обоснованно выбирать рациональные варианты подобных проектов сложных технических систем добычи и транспортировки углеводородов на морском шельфе. Полученные результаты используются в учебном процессе кафедры САПР БНТУ.

Список литературы

1. Информационный программно-технический комплекс для дистанционного решения сложных прикладных задач на основе использования суперкомпьютерных систем / В.А. Кочуров [и др.] // Весці НАН Беларусі. Сер. фіз.-тэхн. навук. – 2010. – № 2. – С. 86–96.

2. Построение спектра конечно-элементных моделей для принятия рациональных инженерных решений при ремонте морских платформ / А.В. Бородуля [и др.] // Весці НАН Беларусі. Сер. фіз.-тэхн. навук. – 2013. – № 4. – С. 101–107.

3. Mirzavand, M. Construction of the spectrum of finite element models for the rational design of complex technical production systems and transportation of hydrocarbons offshore / M. Mirzavand, H. Izadneshan // Pensee J. – 2014. – Vol. 76, no. 2. – P. 348–356.

4. Соболев, И.М. Точки, равномерно заполняющие многомерный куб / И.М. Соболев. – М. : Знание, 1985. – 32 с.

5. Соболев, И.М. Выбор оптимальных параметров в задачах со многими критериями / И.М. Соболев, Р.Б. Статников. – М. : Наука, 1981. – 193 с.

6. Методы оптимизации / под ред. В.С. Зарубин, А.П. Крищенко. – М. : Изд-во МГТУ им. Н.Э. Баумана, 2003. – 440 с.

Поступила 02.07.2014

¹Белорусский национальный технический университет,
Минск, пр. Независимости, 65
e-mail: n_v_v@tut.by

mohsen.mirzavand@yahoo.co.uk

²Донской государственный технологический университет,
Ростов-на-Дону, пл. Гагарина, 1
e-mail: solovievarc@gmail.com

A.V. Borodulia, A.N. Soloveev, V.V. Naprasnikov, M.A. Mirzavand

**CHOICE OF RATIONAL PARAMETERS OF THE FRAME CONSTRUCTION
OF OFFSHORE PLATFORM**

Creation of a parametric geometrical model of the frame construction of offshore oil platform using special built-in languages of finite-element modeling systems and finite-element model of the construction subject to operational loads is considered. Preparation of an optimization construction model is described, optimization calculations are carried out and recommendations on choosing rational construction parameters are given.

УДК 681.511

А.А. Несенчук

ИССЛЕДОВАНИЕ ДИНАМИЧЕСКИХ СИСТЕМ ТРЕТЬЕГО ПОРЯДКА С НЕОПРЕДЕЛЕННОСТЬЮ НА ОСНОВЕ КОРНЕВЫХ ПОРТРЕТОВ

Рассматривается вопрос об устойчивости динамической системы третьего порядка с интервальной неопределенностью. Устанавливаются закономерности поведения функции параметра корневого портрета системы на границе устойчивости, характер зависимости конфигурации портрета от значений изменяющихся параметров. На этой основе формулируются корневое и алгебраическое условия устойчивости и разрабатывается метод параметрического синтеза системы. Приводится алгоритм для реализации процедуры параметрического синтеза.

Введение

Исследование корневых портретов систем с неопределенностью [1–3], формирующих эти портреты полей и корневых годографов [2, 3] имеет большое значение, поскольку они позволяют наблюдать и наглядно представить реакцию системы на параметрические вариации, установить фактическую область расположения семейства корней системы и, следовательно, сделать вывод о том, где и каким образом следует размещать корни для достижения робастной устойчивости и робастного качества [1, 3] системы.

Динамические системы, описываемые характеристическими уравнениями третьей степени, достаточно широко используются на практике. Устойчивость подобных систем с интервальной неопределенностью рассматривается в [4] и ряде других работ, где, как правило, применяется алгебраический подход, который позволяет решить вопрос об устойчивости, однако не дает возможности решать задачи синтеза и получать представление о динамике системы в условиях параметрических вариаций. Корневой подход к проблеме, который специально предназначен для решения задач синтеза, позволяет получить полное представление о поведении системы в условиях неопределенности и на этой основе выполнить синтез.

1. Общие положения

Рассмотрим динамическую систему, описываемую характеристическим уравнением

$$s^n + a_1 s^{n-1} + \dots + a_{n-1} s + a_n, \quad (1)$$

где a_j – действительные коэффициенты полинома, $j = \overline{1, n}$.

Определение 1. *Параметром* или *свободным параметром* корневого годографа алгебраического уравнения (полинома) назовем коэффициент этого уравнения или параметр динамической системы, описываемой этим уравнением, который варьируется по определенному закону с целью построения корневого годографа.

Сделав в уравнении (1) замену переменного, $s = \sigma + i\omega$, и выделив мнимую и действительную части, перепишем его относительно свободного члена в форме функции комплексного переменного

$$a_n = -\frac{\varphi(s)}{\psi(s)} = f(s) = u(\sigma, \omega) + iv(\sigma, \omega), \quad (2)$$

где $u(\sigma, \omega)$ и $v(\sigma, \omega)$ – действительные функции двух независимых переменных σ и ω ; $\varphi(s)$ и $\psi(s)$ – полиномы комплексного переменного s .

С помощью функции отображения (2) осуществляется отображение некоторых заданных в плоскости свободного параметра [2] a_n образов на плоскость комплексного переменного s .

Определив в качестве свободного (варьируемого) параметра годографа параметр a_n , и при условии $-\infty < a_n < +\infty$ запишем в общем виде уравнение корневого годографа полинома (1)

$$v(\sigma, \omega) = 0 \quad (3)$$

и уравнение параметра годографа

$$a_n = u(\sigma, \omega). \quad (4)$$

Под корневым годографом будем понимать корневой годограф Теодорчика – Эванса (КГТЭ) полинома [2].

Определение 2. *Свободным корневым годографом* алгебраического уравнения (полинома) назовем корневой годограф этого уравнения (полинома), параметром которого является его свободный член.

Сформулируем понятие поля корневых траекторий [5]. Пусть

$$\varphi(u, v) = 0 \quad (5)$$

есть кривая в плоскости свободного параметра, которая является образом корневого годографа. Тогда, используя соотношение (2), получим уравнение

$$F(\sigma, \omega) = C, \quad (6)$$

где $F(\sigma, \omega) = \varphi(u(\sigma, \omega), v(\sigma, \omega)) + C$; C – некоторая постоянная, которую назовем *параметром поля корневого годографа*.

Тогда функция

$$F = F(\sigma, \omega) \quad (7)$$

представляет собой функцию поля корневых траекторий [5]. Очевидно, что уравнение (6) является уравнением линий уровня поля корневых траекторий.

Определение 3. *Поле корневых траекторий динамической системы* с характеристическим уравнением (1) назовем совокупность кривых (6), когда C изменяется в промежутке $(-\infty, +\infty)$.

В качестве параметра поля может быть определен как параметр образа корневого годографа, заданного в плоскости свободного параметра, так и коэффициент характеристического полинома или параметр динамической системы.

2. Постановка задачи

Запишем характеристическое уравнение динамической системы третьего порядка в виде

$$s^3 + a_1 s^2 + a_2 s + a_3 = 0, \quad (8)$$

где a_1, \dots, a_3 – коэффициенты, которые изменяются в пределах заданных интервалов действительных значений,

$$a_j \in [a_j, \bar{a}_j], j \in \{1, 2, 3\}, \quad (9)$$

и могут принимать любые значения из этих интервалов.

Система, описываемая уравнением (8), представляет собой систему с интервальной неопределенностью (интервальную динамическую систему (ИДС)), т. е. фактически интервальное семейство систем.

При решении задачи необходимо:

- 1) исследовать динамические свойства и установить закономерности поведения системы, описываемой уравнением (8), при вариации параметров в заданных интервалах (9);
- 2) сформулировать корневое условие устойчивости интервальной системы, описываемой уравнением (8);
- 3) с целью выполнения параметрического синтеза системы разработать алгоритм расчета интервалов вариации коэффициентов характеристического уравнения системы (8), в пределах которых сохраняется асимптотическая устойчивость, т. е. система является робастно устойчивой.

Входными данными алгоритма являются характеристическое уравнение (8), описывающее динамические свойства системы, и исходные граничные значения интервалов (9) вариации коэффициентов уравнения (8).

Выходными данными алгоритма являются граничные значения интервалов вариации коэффициентов уравнения (8), при которых система является робастно устойчивой.

3. Закономерности поведения интервальной системы на границе устойчивости

Исследуем поведение семейства, описываемого уравнением (8), на границе устойчивости при положительных коэффициентах уравнения, используя корневой портрет системы [3].

Определение 4. *Корневым портретом* или *корневым годографом динамической системы с параметрической неопределенностью* назовем семейство корневых годографов, порождаемых семейством функций отображения вида (2) этой системы.

Определение 5. Свободным корневым портретом динамической системы с параметрической неопределенностью назовем корневой портрет, сформированный семейством свободных корневых годографов этой системы.

С целью проведения исследования, приняв в качестве параметра годографа параметр a_3 , запишем функцию отображения вида (2) для системы третьего порядка, переписав уравнение (8) относительно a_3 в форме

$$s^3 + a_1 s^2 + a_2 s = -a_3 \quad (10)$$

и подставив вместо комплексного переменного s его значение $s = \sigma + i\omega$. Тогда функция отображения примет следующий вид:

$$\sigma^3 - 3\sigma\omega^2 + a_1\sigma^2 - a_1\omega^2 + a_2\sigma + i(3\sigma^2\omega - \omega^3 + 2a_1\sigma\omega + a_2\omega) = u(\sigma, \omega) + iv(\sigma, \omega) = -a_3.$$

На основании последнего выражения запишем уравнение свободного КГТЭ алгебраического уравнения (8)

$$3\sigma^2\omega - \omega^3 + 2a_1\sigma\omega + a_2\omega = v(\sigma, \omega) = 0 \quad (11)$$

и уравнение (функцию) параметра корневого годографа

$$g(\sigma, \omega) = -\sigma^3 + 3\sigma\omega^2 - a_1\sigma^2 + a_1\omega^2 - a_2\sigma = u(\sigma, \omega) = a_3. \quad (12)$$

Для определения динамических свойств системы на границе устойчивости (т. е. при $\sigma = 0$) перепишем соответственно уравнение (11) корневого годографа в виде

$$\omega^3 - a_2\omega = 0 \quad (13)$$

и функцию параметра (12) в виде

$$g(\omega) = a_1\omega^2 = a_3. \quad (14)$$

Выражение (13) используется для вычисления координат ω точек пересечения границы устойчивости (мнимой оси $i\omega$) ветвями годографов корневого портрета, а выражение (14) – для вычисления значений параметра a_3 в этих точках.

Поведение системы на оси σ описывается соответственно следующими уравнениями параметра и корневого годографа:

$$3\sigma^2 + 2a_1\sigma + a_2 = 0; \quad (15)$$

$$\sigma^2 + a_1\sigma + a_2 = a_3. \quad (16)$$

Согласно формулам (13)–(16) при положительных значениях коэффициентов a_1 и a_2 , $a_1 > 0$ и $a_2 > 0$, граница асимптотической устойчивости пересекается только положительными ветвями свободных корневых годографов семейства (рис. 1) и начальные точки (полюсы) годографов (на рис. 1 показаны крестиками) располагаются в левой полуплоскости. Соответствующие точки пересечения обозначены на рис. 1 буквой t . При $a_1 < 0$ эта закономерность противоположна, т. е. граница устойчивости пересекается только отрицательными ветвями свободных корневых годографов семейства и начальные точки располагаются в правой полуплоскости. При $a_1 = 0$ все начальные точки семейства располагаются на мнимой оси $i\omega$. Эти варианты будут подробно рассмотрены ниже.

Свободный корневой портрет системы определим семейством полей корневых траекторий с функцией (7):

$$F(\sigma, \omega) = 3\sigma^2 - \omega^2 + 2a_1\sigma \quad (17)$$

и линиями уровня (6):

$$F(\sigma, \omega) = 3\sigma^2 - \omega^2 + 2a_1\sigma = -a_2, \quad (18)$$

что позволяет свести исследование интервального корневого портрета к исследованию поля корневых траекторий как носителя всех свойств портрета.

С целью установления динамических свойств поля проводится исследование поведения функции параметра $g(\omega)$ (14) для поля на границе устойчивости $i\omega$ с помощью первой производной $g'(\omega)$. В этом случае точки экстремума функции определяются выражением

$$g'(\omega) = 2a_1\omega = 0, \quad (19)$$

которое дает основание утверждать, что единственный экстремум функции (двухкратный нуль) располагается в начале координат, где $\omega = 0$. Поэтому очевидно, что функция (14) является непрерывно возрастающей и, следовательно, значения параметра траектории a_3 при положительных значениях коэффициентов характеристического уравнения возрастают от начала координат, где $g(\omega) = 0$, по направлению оси $i\omega$.

Для значений $a_1 < 0$ установленная закономерность противоположна, т. е. имеет место не возрастающий, а убывающий характер распределения значений функции параметра вдоль границы устойчивости (см. формулу (14)).

При $a_1 = 0$ и положительных a_2 функция параметра на границе устойчивости всегда равна нулю.

Соответствующие диаграммы распределения значений параметра годографа a_3 при изменении параметра поля a_2 в пределах бесконечного интервала значений $-\infty < a_2 < +\infty$ изображены на рис. 2. Кривые на диаграммах для простоты заменены прямыми линиями.

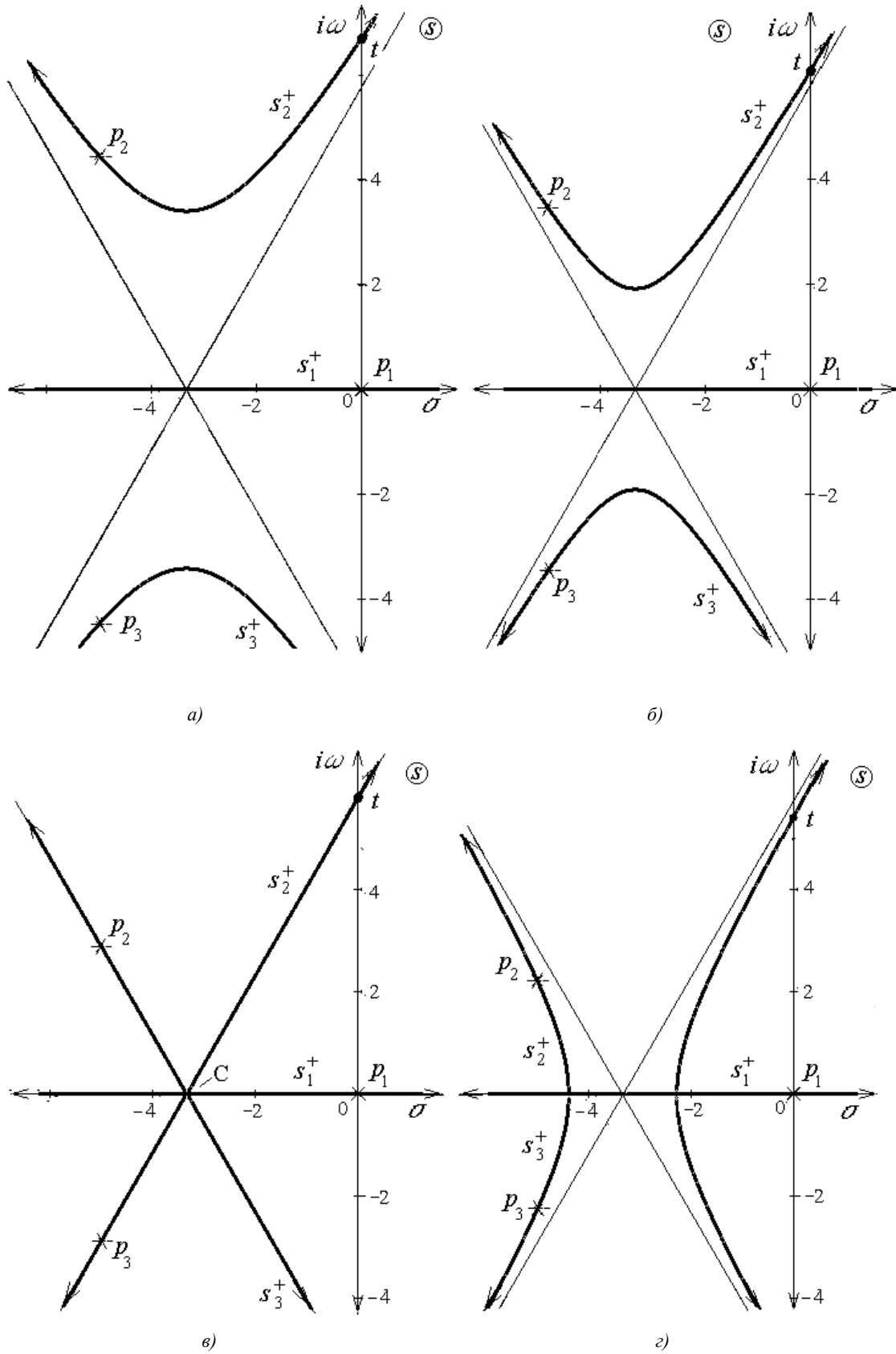


Рис. 1. Свободные КГТЭ ИДС, описываемой характеристическим уравнением $s^3 + 10s^2 + a_2s + a_3 = 0$, построенные при различных значениях a_2 : а) $a_2 = 40$; б) $a_2 = 33,3$; в) $a_2 = 30$; г) $a_2 = 25$

На основании изложенного выше сформулируем следующее утверждение.

Утверждение. Функция параметра траектории (14) для поля корневых траекторий (17) динамической системы третьего порядка, описываемой характеристическим уравнением (8) с положительными или отрицательными коэффициентами, является соответственно монотонно возрастающей или убывающей функцией; в единственной точке экстремума функции, располагающейся в начале координат плоскости корней, параметр траектории равен нулю.

На рис. 2, а показано поле корневых траекторий динамической системы третьего порядка в плоскости корней s при изменении параметра поля a_2 в пределах некоторого заданного интервала значений: $a_{2\min} < a_2 < a_{2\max}$. Поле представлено двумя ограничивающими его годографами (линиями уровня), построенными при $a_2 = a_{2\min}$ (первый) и при $a_2 = a_{2\max}$ (второй). Точки пересечения границы устойчивости ветвями данных годографов обозначены соответственно буквами t_1 и t_2 . Начальные точки годографов, т. е. точки, в которых параметр годографа a_3 равен нулю, показаны крестиками. Начальные точки, принадлежащие первому годографу, обозначены цифрой 1, второму – цифрой 2. Координаты ω точек экстремума функции параметра обозначены как ω_{\min}^f (точка минимума на рис. 2, а) и ω_{\max}^f (точка максимума на рис. 2, б). Область пересечений D_{ω}^f ветвями годографов поля границы устойчивости $i\omega$ при изменении параметра поля a_2 в пределах бесконечного интервала значений $-\infty < a_2 < +\infty$ разделена на две комплексно-сопряженные части: верхнюю $D_{\omega t}^f$ и нижнюю $D_{\omega b}^f$. Реальная область пересечений D_{ω} границы устойчивости $i\omega$ ветвями годографов поля при изменении параметра поля a_2 в пределах указанного выше заданного интервала значений $a_{2\min} < a_2 < a_{2\max}$ ($\underline{a}_2 < a_2 < \bar{a}_2$) располагается в интервале между точками t_1 и t_2 . Комплексно-сопряженная часть интервала $[t_1, t_2]$, расположенная в нижней полуплоскости на рис. 2, не обозначена и не рассматривается, поскольку аналогична расположенной в верхней полуплоскости. Исходя из установленных выше закономерностей, значение функции параметра в области пересечений возрастает от начала координат до точки t_1 и от начала координат, где параметр траектории $a_3 = 0$, в направлении оси $i\omega$.

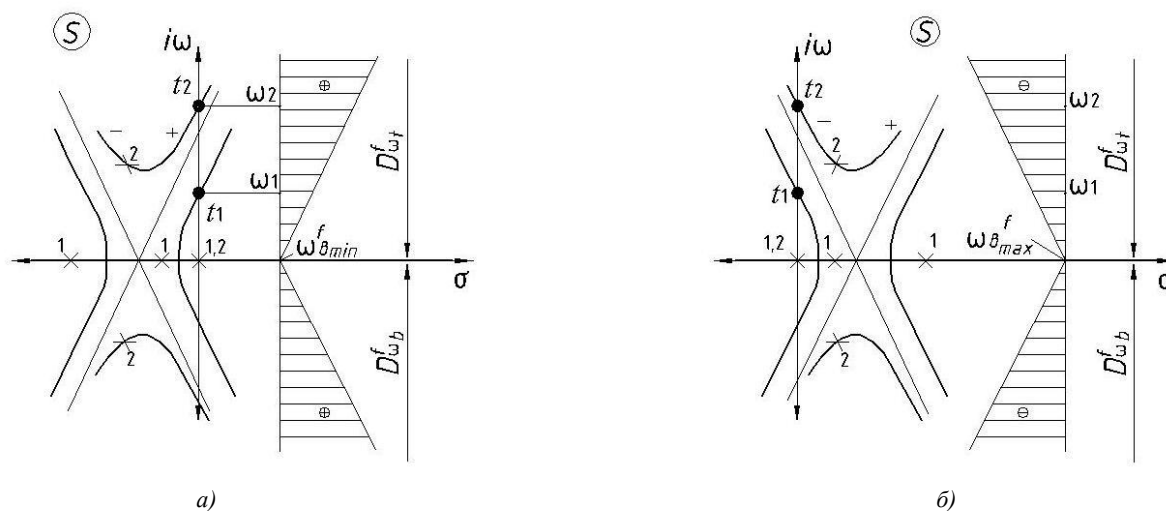


Рис. 2. Динамика функции параметра траекторий на границе устойчивости для системы третьего порядка: а) $a_1 > 0$; б) $a_1 < 0$

При изменении знака коэффициента a_1 на противоположный картина (рис. 2, б) аналогична приведенной на рис. 2, а, но располагается симметрично рассмотренной относительно границы устойчивости, что следует из формулы (14), поскольку при $a_1 < 0$ согласно формуле (14) значения параметра a_3 на границе устойчивости также будут отрицательны,

т. е. граница будет пересекаться только отрицательными ветвями годографов поля (портрета), а в таком случае все положительные комплексно-сопряженные ветви годографов поля (портрета) располагаются в правой полуплоскости корней. Симметричность портрета, построенного при $a_1 < 0$ (рис. 2, б), относительно мнимой оси плоскости корней портрету, построенному при $a_1 > 0$ (рис. 2, а), также вытекает из уравнения КГТЭ (11). Следовательно, при прочих равных условиях при непрерывном уменьшении параметра a_1 корневой портрет интервального семейства постепенно и непрерывно перемещается вправо в сторону неустойчивого состояния (рис. 3).

На рис. 2 дана наглядная исчерпывающая информация о характере изменения динамических свойств системы в пределах бесконечных интервалов изменения параметров a_1 , a_2 и a_3 .

Рассмотренные корневые портреты позволяют получить полное представление о динамике исследуемой системы. Они наглядно подтверждают, что в случае отрицательности хотя бы одного из коэффициентов система является неустойчивой. Рассмотрим каждый коэффициент отдельно.

Предположим, что $a_1 < 0$ или $a_2 < 0$. На рис. 2 видно, что в этом случае комплексные положительные и правые действительные ветви всего семейства всегда расположены в правой полуплоскости, т. е. ни в каком из возможных состояний система не является устойчивой.

Предположим, что $a_3 < 0$. Невозможность достижения устойчивого состояния в этом случае доказывается тем, что правая крайняя действительная ветвь каждого из годографов всего семейства, которая, как известно, является отрицательной, всегда совпадает с положительной («неустойчивой») полуосью σ , расположенной в правой полуплоскости и направленной вправо от начала координат (рис. 2).

На рис. 3 представлена динамика поля свободных корневых траекторий интервальной системы для случая $a_1 > 0$, где показан ряд характерных последовательных состояний системы (для некоторых фиксированных значений a_1 и a_2) при непрерывном уменьшении значений a_2 от положительных значений (рис. 3, а–в) к отрицательным (рис. 3, г). Крестиками отмечены начальные точки. Каждый из графиков соответствует определенной линии уровня поля. Рисунок наглядно иллюстрирует описанные выше результаты.

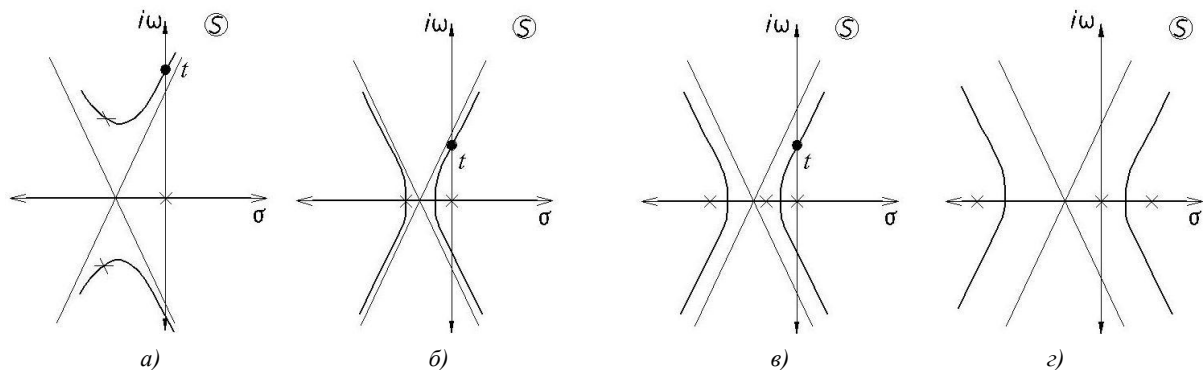


Рис. 3. Динамика поля свободных корневых траекторий системы третьего порядка:

а) $a_2 \gg 0$; б) $a_2 > 0$; в) $a_2 > 0$; г) $a_2 < 0$

На рис. 3, а две начальные точки располагаются в комплексной плоскости корней, затем с уменьшением a_2 они постепенно сближаются и сходятся на оси σ (рис. 3, б), затем начинают расходиться в разные стороны (рис. 3, в), устремляясь одна в $+\infty$, другая в $-\infty$. Соответственно при уменьшении a_2 от $+\infty$ до $-\infty$ две комплексные ветви корневого портрета сближаются в направлении оси σ , затем сходятся на этой оси в одной точке и далее расходятся в противоположные стороны по оси σ , приближая тем самым систему к неустойчивому состоянию. После пересечения одной из начальных точек оси $i\omega$ (когда коэффициент a_2 становится отрицательным) и перехода ее в правую полуплоскость s (рис. 3, г) система теряет устойчивость. С увеличением a_1 весь портрет сдвигается влево, т. е. в сторону более устойчивого состояния.

В результате проведенного исследования динамических свойств интервальных систем третьего порядка на основе использования корневых портретов, представленных семейством полей корневых траекторий, выявлены и математически доказаны следующие основные закономерности:

1. Функция $g = g(\omega)$ параметра траектории на границе устойчивости для поля корневых траекторий интервальной динамической системы третьего порядка при положительных коэффициентах характеристического уравнения системы является монотонно возрастающей от начала координат, где $g(\omega) = 0$, в направлении оси $i\omega$ комплексной плоскости корней.

2. При положительных значениях коэффициентов a_1 и a_2 характеристического уравнения интервальной динамической системы третьего порядка граница асимптотической устойчивости пересекается только положительными ветвями свободных корневых годографов семейства (см. рис. 1) и начальные точки (полюсы) годографов располагаются в левой полуплоскости. При $a_1 < 0$ эта закономерность противоположна, т. е. граница устойчивости пересекается только отрицательными ветвями свободных корневых годографов семейства и начальные точки располагаются в правой полуплоскости. При $a_1 = 0$ все начальные точки семейства располагаются на мнимой оси $i\omega$.

3. Корневой портрет интервальной динамической системы третьего порядка, описываемой характеристическим уравнением с заданными значениями коэффициентов a_1 , a_2 и a_3 , при изменении знака a_1 на противоположный будет располагаться симметрично исходному портрету относительно мнимой оси $i\omega$ плоскости корней s .

4. При прочих равных условиях при непрерывном уменьшении коэффициента a_1 характеристического уравнения корневой портрет интервального семейства постепенно и непрерывно перемещается вправо в сторону неустойчивого состояния.

Выполнена интерпретация положения о неустойчивости системы при отрицательности хотя бы одного коэффициента характеристического уравнения на основе закономерностей локализации корневого портрета системы в плоскости корней.

4. Синтез характеристического полинома и условия устойчивости

4.1. Синтез устойчивого характеристического полинома

На основании результатов исследования, описанного в разд. 3, сформулируем корневое условие устойчивости ИДС третьего порядка.

Теорема 1. Для асимптотической устойчивости интервальной динамической системы, описываемой семейством характеристических уравнений вида (8), необходимо и достаточно, чтобы выполнялось условие

$$\bar{a}_3 < a_{r\min}, \quad (20)$$

где \bar{a}_3 – верхняя заданная граница интервала изменения параметра траектории a_3 ; $a_{r\min}$ – минимальное значение параметра a_3 в области D_ω пересечения границы устойчивости годографами корневого портрета системы:

$$a_{r\min} = \inf A_{\omega r}. \quad (21)$$

Здесь $A_{\omega r}$ – множество значений параметра a_3 в точках области пересечений D_ω .

Доказательство. Начальные точки всех ветвей корневых годографов портрета интервальной системы третьего порядка, описываемой характеристическим уравнением с положительными коэффициентами, располагаются в левой полуплоскости комплексной плоскости корней s , что следует из уравнения (8). Это означает, что интервалы этих ветвей от начальных точек до точек пересечения с границей асимптотической устойчивости (осью $i\omega$) также располагаются в левой полуплоскости, т. е. являются устойчивыми. Следовательно, если заданный интервал изменения параметра a_3 на нижней, доминирующей ветви [3] портрета полностью располагается в левой полуплоскости, т. е. выполняется условие (20), то на оставшихся двух

ветвях каждого годографа портрета этот интервал также будет наверняка располагаться в левой полуплоскости. Теорема доказана.

Поэтому для выполнения параметрического синтеза устойчивой интервальной системы третьего порядка достаточно вычислить только значение параметра траектории $a_3 = a_{r\min}$ (см. формулу (21)) и выполнить сравнение заданного значения \bar{a}_3 с полученным значением $a_{r\min}$ в соответствии с условием (20).

На основании сформулированного в разд. 3 утверждения можно заключить, что минимальное значение параметра $a_{r\min}$ на оси $i\omega$ будет в точке области пересечений D_ω . Значение координаты ω

$$\omega_{r\min} = \inf W_r, \quad (22)$$

где W_r – множество значений координат ω точек пересечения в области пересечений D_ω ; $\omega_{r\min}$ – минимальное значение координаты ω в области D_ω .

На рис. 2, а значению $\omega_{r\min}$ соответствует точка t_1 с координатой ω_1 .

Значение $\omega_{r\min}$ для корневого портрета системы вычисляется с использованием уравнения корневого годографа (13):

$$\omega_{r\min_{1,2}} = \pm\sqrt{a_2}. \quad (23)$$

Значение параметра в полученных точках $\omega_{r\min_{1,2}}$ вычисляется с использованием формулы параметра (14):

$$a_{r\min_{1,2}} = \underline{a}_1(\omega_{r\min_{1,2}})^2. \quad (24)$$

Очевидно, что $a_{r\min_1} = a_{r\min_2} = a_{r\min}$.

4.2. Алгебраическое условие устойчивости системы

На основе сделанных выше заключений сформулируем алгебраическое условие устойчивости для рассматриваемой системы.

Теорема 2. Для асимптотической устойчивости интервальной динамической системы третьего порядка, описываемой семейством характеристических полиномов вида (8), необходимо и достаточно, чтобы был устойчив только один следующий полином семейства:

$$h(s) = s^3 + \underline{a}_1 s^2 + \underline{a}_2 s + \bar{a}_3. \quad (25)$$

Доказательство. Данный полином составлен из коэффициентов \bar{a}_3 , \underline{a}_2 и \underline{a}_1 , фигурирующих соответственно в формулах (11), (13) и (14). Поэтому на основании теоремы 1 можно утверждать, что этому полиному принадлежит доминирующая ветвь корневого портрета, которая пересекает границу устойчивости в точке с координатой $\omega_{r\min}$ (13), где значение параметра равно $a_{r\min}$ (14), т. е. устойчивость полинома (25) определяет устойчивость всего полиномиального семейства. Теорема доказана.

Полином (25) совпадает с полиномом, полученным Андерсоном и Джури [4]. Однако по сравнению с условием Андерсона и Джури полученное условие (20) имеет серьезное преимущество: оно позволяет не только проверить устойчивость интервальной системы, но и найти значения интервалов изменения коэффициента a_3 , обеспечивающие ее устойчивость в случаях, когда проверка исходной системы на устойчивость показала отрицательный результат.

5. Алгоритм параметрического синтеза

На основании описанного выше метода определим алгоритм параметрического синтеза интервальной системы третьего порядка, состоящий из следующих шагов:

Шаг 1. Определение входных данных. В качестве входных данных определяются исходные граничные значения интервалов (9) вариации коэффициентов уравнения (8).

Шаг 2. Вычисление координаты $\omega_{r\min}$ (22) по формуле (23).

Шаг 3. Вычисление минимального значения параметра годографа $a_3 = a_{r\min}$ (21) по формуле (24).

Шаг 4. Проверка условия устойчивости (20) и корректировка предельного значения \bar{a}_3 в соответствии с этим условием в случае необходимости.

Шаг 5. Определение выходных данных. Интервал вариации параметра a_3 , $0 < a_3 < \bar{a}_3$, обеспечивающий робастную устойчивость системы.

Проверка на устойчивость заданного интервального полинома (8) может быть также осуществлена сразу после выполнения шага 1 посредством проверки устойчивости доминирующего полинома (25) согласно условию, сформулированному в теореме 2. Если полином (25) окажется устойчивым, то работа алгоритма завершается, если же он неустойчив, то выполняются шаги 2–5 алгоритма.

Заключение

В статье рассмотрен вопрос об устойчивости динамических систем третьего порядка с интервальной неопределенностью. Использована математическая модель системы в форме корневого портрета, представленного семейством полей корневых траекторий. Установлены свойства и закономерности поведения функции параметра корневого портрета системы на границе асимптотической устойчивости, показан характер зависимости конфигурации портрета от значений изменяющихся параметров. Сформулированы корневое и алгебраическое условия устойчивости. На этой основе разработан алгоритм параметрического синтеза интервальной системы, который позволяет рассчитать параметры системы, обеспечивающие ее робастную устойчивость. В отличие от имеющихся подходов к данной проблеме представленный корневой подход отличается большой наглядностью, позволяет не только осуществлять проверку системы на устойчивость, но и наблюдать непрерывную картину изменения динамических свойств системы в ответ на параметрические вариации, что дает возможность выполнять параметрический синтез систем, обладающих робастными свойствами.

Список литературы

1. Дорф, Р. Современные системы управления / Р. Дорф, Р. Бишоп. – М. : Лаборатория базовых знаний, 2009. – 832 с.
2. Римский, Г.В. Автоматизация исследований динамических систем / Г.В. Римский, В.В. Таборовец. – Минск : Наука и техника, 1978. – 336 с.
3. Несенчук, А.А. Анализ и синтез робастных динамических систем на основе корневого подхода / А.А. Несенчук. – Минск : ОИПИ НАН Беларуси, 2005. – 234 с.
4. Anderson, B.D.O. On robust Hurwitz polynomials / B.D.O. Anderson, E. Jury, M. Mansour // IEEE Trans. on Automatic Control. – 1987. – Vol. AC-32, no. 8. – P. 909–913.
5. Несенчук, А.А. Размещение корней динамической системы в областях произвольной формы / А.А. Несенчук // Доклады НАН Беларуси. – 2008. – Т. 52, № 5. – С. 31–35.

Поступила 02.08.2014

A.A. Nesenчук

**RESEARCH OF UNCERTAIN DYNAMIC SYSTEMS
OF THE THIRD ORDER BASED ON THE ROOT LOCUS PORTRAITS**

The problem of stability of uncertain interval dynamic systems of the third order is considered. Properties and regularities of the system root locus portrait parameter function behavior at the system stability bound have been discovered, the dependency of the portrait configuration from the values of variable parameters has been shown. The root locus and algebraic stability conditions have been formulated. On this basis the algorithm for parametric synthesis of interval systems has been worked out, which allows to calculate the system parameters ensuring its robust stability.

ЗАЩИТА ИНФОРМАЦИИ

УДК 681.324.067

А.С. Поляков, В.Е. Самсонов

АНАЛИЗ ВОЗМОЖНОСТЕЙ АЛГОРИТМОВ МЕЖДУНАРОДНОГО
СТАНДАРТА «ОБЛЕГЧЕННАЯ КРИПТОГРАФИЯ» – ISO/IEC 29192-2:2012

Приводятся данные о характеристиках алгоритмов международного стандарта «Облегченная (легковесная) криптография» при их аппаратной реализации в базисе микросхем типа FPGA. Дается сравнение характеристик этих алгоритмов с характеристиками нескольких широко используемых стандартных алгоритмов шифрования и оцениваются возможности алгоритмов легковесной криптографии.

Введение

В 2007 г. коллектив авторов [1] представил алгоритм шифрования под интригующим названием «ультралегкий», имея в виду, что алгоритм обладает высокой производительностью (быстродействием), сравнительно неплохой криптостойкостью в сравнении с используемыми стандартными алгоритмами, а главное – требует небольших затрат (т. е. является ультралегким) при его аппаратной реализации.

Немного позже, в 2009 г., Аксель Пошманн в своей работе [2] привел аргументацию и обоснование необходимости развития направления «облегченная криптография» с учетом повсеместного и бурного развития компьютеризации всех аспектов жизнедеятельности общества: с одной стороны, внедрения компьютерных технологий в широкие сферы социальной жизни, характеризующегося требованием реализации функций шифрования с использованием малых объемов ресурсов (памяти, логических элементов и т. п.); с другой стороны, обеспечения необходимой защиты информации пользователей открытых сетей типа Интернет от несанкционированного доступа.

К облегченной криптографии относятся алгоритмы, разрабатываемые специально для устройств с ограниченными или крайне малыми ресурсами [3, 4]. Общим свойством таких алгоритмов являются низкие требования:

- к требуемой площади кристалла, на котором алгоритм может быть аппаратно реализован;
- вычислительной мощности микропроцессора, на котором выполняются вычисления;
- оперативной памяти вычислительного устройства и т. п.

Видимо, проблема перехода, по крайней мере в отдельных областях применения криптографической защиты данных, к малоресурсным (в смысле высказанных выше требований) алгоритмам шифрования созрела настолько, что организации ISO и IEC очень быстро приняли и утвердили в 2012 г. международный стандарт «Lightweight Cryptography» [5], в русскоязычном переводе – «Облегченная (легковесная) криптография».

К настоящему времени уже разработано много алгоритмов, относящихся к облегченной криптографии. Некоторые из них рассмотрены в работе [4]. В данной статье внимание будет уделено двум алгоритмам, включенным в стандарт [5]:

PRESENT – блочный шифр с размером блока данных 64 бита и размером ключа 80 бит – PRESENT-80 или размером ключа 128 бит – PRESENT-128;

CLEFIA – блочный шифр с размером блока данных 128 бит и размерами ключей шифрования 128, 192 или 256 бит – CLEFIA-128, CLEFIA-192, CLEFIA-256 соответственно.

Данные алгоритмы, основанные на концепции малых затрат ресурсов на их реализацию и получение высокого быстродействия, вызывают повышенный интерес специалистов в части соответствия продекларированных и реальных характеристик этих алгоритмов, а также сравнения их с широко применяемыми на практике стандартными криптографическими алгоритмами. Имеются исследования (см., например, [6]), которые показывают, что при реализации на неко-

торых аппаратных платформах алгоритмы PRESENT и CLEFIA по производительности уступают алгоритму AES.

Поскольку рассматриваемые алгоритмы изначально были ориентированы на аппаратную реализацию, в данной работе исследованы их основные характеристики: быстродействие (временные) и сложность реализации (объемные), а также проведено сравнение с характеристиками некоторых широко применяемых стандартных алгоритмов при их реализации в базе микросхем типа FPGA.

1. Краткое описание исследуемых алгоритмов

Авторы алгоритма *PRESENT* подчеркивают, что разработали его для узкоспециальных применений, для которых не подходит более универсальный, но и более ресурсоемкий алгоритм AES. По их мнению, алгоритм PRESENT рассчитан на аппаратную реализацию и его предлагается применять прежде всего в микрочипах в тех случаях, когда требуются низкие затраты ресурсов и допускаются пониженные требования к криптографической стойкости шифра.

Алгоритм PRESENT включает два алгоритма: PRESENT-80 и PRESENT-128, которые по составу операций и порядку их применения идентичны. Разница заключается лишь в том, что в алгоритме PRESENT-80 раундовый ключ длиной 64 бита формируется с помощью соответствующих операций из первоначально заданного образующего ключа длиной 80 бит, а в алгоритме PRESENT-128 в качестве образующего используется ключ длиной 128 бит. Правила формирования раундовых ключей в алгоритмах немного отличаются, но при этом используется одинаковый набор операций.

Алгоритм PRESENT основан на принципе SP-сети и предусматривает выполнение 31 раунда шифрования. Перед выполнением операции шифрования данных производится процедура генерации 32 64-битовых раундовых ключей.

Шифрование блока данных длиной 64 бита выполняется за 31 раунд, на каждом из которых производятся следующие операции:

- сложение по mod 2 текущего состояния блока данных с очередным раундовым ключом;
- преобразование усложнения, состоящее в замене полубайтов блока данных соответствующими значениями четырехбитовых S-блоков подстановки;
- перемешивающее преобразование, предусматривающее перестановку значений битов в текущем состоянии блока данных в соответствии с заданными правилами;
- завершающая операция – сложение по mod 2 результата, полученного после выполнения 31-го раунда алгоритма, с последним раундовым ключом.

Алгоритм *CLEFIA* разработан корпорацией Sony [7], является симметричным блочным шифром и соответствует требованиям к шифру AES [8]: размер блока – 128 бит, поддерживаемые длины ключей шифрования – 128, 192 и 256 битов, что соответствует алгоритмам CLEFIA-128, CLEFIA-192, CLEFIA-256. Число раундов зависит от длины ключа и равно 18, 22, 26 соответственно.

Структура алгоритма представляет собой обобщенную сеть Фейстеля и предусматривает выполнение следующих операций:

1) при генерации раундовых ключей используется заданный начальный ключ шифрования и специальная таблица 32-битовых констант. В зависимости от размера заданного начального ключа – 128, 192 или 256 битов – создается массив 32-битовых раундовых ключей в количестве 36, 44 или 52 ключа соответственно;

2) для алгоритма CLEFIA-128 блок данных и исходный ключ шифрования разделяются на четыре 32-разрядных слова, которые обрабатываются в течение 18 раундов с помощью функций F_0 и F_1 , предусматривающих выполнение следующих операций: суммирование по mod 2 слова блока данных с очередным раундовым ключом, подстановка байтов полученного результата с использованием соответственно таблиц S_0 и S_1 ; умножение двух полиномов степени меньше 8, вычисление модуля результата умножения полиномов по неприводимому полиному $z^8 + z^4 + z^3 + z^2 + 1$. При выполнении очередного раунда i используются раундовые ключи с номерами $(2i - 1)$ и $2i$.

Особенностью алгоритма CLEFIA является использование операции «отбеливание»: перед выполнением первого раунда алгоритма второе и четвертое слова блока данных суммируются по mod 2 со вторым и четвертым словами исходного ключа, а после выполнения последнего раунда полученные результаты шифрования, соответствующие первому и третьему словам блока данных, суммируются по mod 2 с первым и третьим словами исходного ключа соответственно.

Из множества алгоритмов CLEFIA исследование характеристик проводилось только для самого простого варианта – CLEFIA-128, поскольку на основании изучения описания алгоритмов CLEFIA-192 и CLEFIA-256 был сделан вывод о нецелесообразности проведения их исследований в связи с повышенной сложностью этих алгоритмов в сравнении с алгоритмом CLEFIA-128.

2. Объемно-временные характеристики алгоритмов PRESENT и CLEFIA

Анализ показателей аппаратной реализации рассматриваемых алгоритмов производился следующим образом: для каждого из алгоритмов с помощью системы проектирования фирмы XILINX были разработаны реализующие их проекты в базисе микросхем типа FPGA, затем произведено выполнение проектов (этапы Synthesize, Translate, Map), в результате чего получены данные о количестве (объеме) оборудования, необходимого для реализации рассматриваемых алгоритмов.

С помощью моделирующей системы ModelSim производилось логическое моделирование разработанных проектов, результаты которого позволили определить количество тактов, необходимых для шифрования одного блока информации каждым из алгоритмов. Логическое моделирование позволило также проверить правильность реализации алгоритмов, для чего использовались тестовые примеры, приведенные в приложении к стандарту [5].

Объемные характеристики аппаратной реализации алгоритмов PRESENT-80, PRESENT-128 и CLEFIA-128 в базисе микросхем типа FPGA серий Spartan 3 и Virtex 4 представлены в табл. 1, в которой указаны абсолютные значения показателей характеристик, а в скобках приведено процентное соотношение количества требуемых элементов оборудования к количеству оборудования, имеющегося в соответствующих микросхемах.

Таблица 1

Объемные характеристики алгоритмов PRESENT-80, PRESENT-128, CLEFIA-128

Характеристика	Тип и серия микросхемы					
	Spartan 3, xc3s2000-4fg456			Virtex 4, xc4vlx15-10sf363		
	PRESENT-80	PRESENT-128	CLEFIA-128	PRESENT-80	PRESENT-128	CLEFIA-128
Slice	146 (0,8 %)	316 (1,6 %)	2061 (10 %)	147 (2,4 %)	316 (5,0 %)	2081 (34%)
Триггеры	227 (0,6 %)	275 (0,7 %)	1964 (4,8 %)	227 (1,9 %)	275 (2,3 %)	1967 (16%)
Четырехвходовой LUT	292 (0,7 %)	341 (0,9 %)	2712 (6,7 %)	293 (2,4 %)	341 (2,6 %)	2743 (22%)
BRAM	35 (88 %)	36 (90 %)	18 (45 %)	35 (73 %)	36 (75 %)	18 (38%)

Примечание: LUT (look-up table) – логическая таблица, представляющая собой однобитовое ОЗУ на 16 ячеек; Slice – единица оборудования, состоящая из двух триггеров и двух LUT; BRAM – блок памяти размером 2 Кбит.

3. Сравнение характеристик алгоритмов PRESENT и CLEFIA с характеристиками стандартных алгоритмов

Считается, что PRESENT является одним из самых компактных криптоалгоритмов, существует оценка, что для его аппаратной реализации требуется приблизительно в 2,5 раза меньше логических элементов, чем для AES или CLEFIA [4].

Для сравнения характеристик были выбраны известные и широко применяемые стандартные алгоритмы: AES [8], ГОСТ 28147-89 [9] и алгоритм Belt [10], недавно введенный в действие в качестве стандарта Республики Беларусь [11]. Выбор этих алгоритмов обусловлен следующим:

- ГОСТ 28127-89, имеющий размер ключа 256 бит, считается одним из самых криптостойких алгоритмов шифрования [12] и используется во многих странах СНГ;
- AES (Advanced Encryption Standard) является стандартом шифрования США, выбранным в результате конкурса из множества алгоритмов как наиболее соответствующий современным требованиям к криптографическим алгоритмам и широко используемый во многих странах мира;
- алгоритм Belt с размером ключа 256 бит представляет интерес для сравнения в качестве нового современного алгоритма шифрования, имеющего высокую криптостойкость.

Данные по объему оборудования и быстродействию алгоритмов PRESENT-80, PRESENT-128, CLEFIA-128 и сравниваемых с ними стандартных алгоритмов представлены в табл. 2 и 3. При этом использованы результаты исследования характеристик аппаратной реализации стандартных алгоритмов шифрования, приведенные в работе [13].

Таблица 2

Затраты оборудования на реализацию алгоритмов

Алгоритм	Количество Slice	Количество триггеров	Количество четырехходов. LUT	Количество BRAM
ГОСТ 28147-89	349	233	479	9
Belt	1070	302	2050	28
AES-128	2107	504	3461	35
PRESENT-80	146	227	292	35
PRESENT-128	316	275	341	36
CLEFIA-128	2081	1967	2743	18

Таблица 3

Затраты времени на шифрование данных, число тактов

Алгоритм	Размер блока данных, бит	Количество тактов на один блок данных	Количество тактов на блок данных размером 64 бита
ГОСТ 28147-89	64	129	129
Belt	128	211	106
AES-128	128	97	49
PRESENT-80	64	97	97
PRESENT-128	64	97	97
CLEFIA-128	128	419	210

При оценке и сравнении показателей характеристик алгоритмов, приведенных в табл. 2 и 3, следует учитывать, что проекты всех алгоритмов были реализованы на одинаковом аппаратном базисе – микросхемах типа FPGA серий Spartan 3 и Virtex 4, все проекты были максимально оптимизированы (насколько это удалось авторам) путем распараллеливания всех операций, допускающих распараллеливание. Поэтому можно считать, что сравнение показателей характеристик алгоритмов производится корректно.

В табл. 3 не приведены данные о затратах времени на подготовку раундовых ключей, которые составляют:

- ГОСТ 28147-89 – подготовка раундовых ключей не требуется;
- AES-128 – 210 тактов;
- PRESENT-80 – 130 тактов;
- PRESENT-128 – 130 тактов;
- CLEFIA-128 – 460 тактов.

Belt – дополнительное время на подготовку раундовых ключей не требовалось, так как в представленном в табл. 2 и 3 варианте реализации алгоритма Belt подготовка ключей для по-

следующего раунда производилась одновременно (параллельно) с выполнением очередного раунда шифрования.

Затраты времени на шифрование блока данных алгоритмами PRESENT-80 и PRESENT-128 одинаковы, а затраты на оборудование отличаются. Это объясняется тем, что подготовка 64-битовых раундовых ключей в алгоритме PRESENT-80 производится с использованием начального ключа размером 80 бит, а в алгоритме PRESENT-128 начальный ключ имеет размер 128 бит, что потребовало дополнительных затрат аппаратных ресурсов.

Заключение

Из табл. 1 видно, что микросхемы типа FPGA для реализации алгоритмов PRESENT, исходя из экономических критериев, не совсем пригодны, поскольку при незначительных затратах логических элементов на реализацию алгоритмов требуются большие объемы памяти, которые в дешевых микросхемах типа FPGA отсутствуют.

Анализ табл. 2 показывает, что для реализации алгоритмов PRESENT-80 и PRESENT-128 требуется не намного меньше оборудования, чем для сравниваемых стандартных алгоритмов, но зато необходимо значительно больше элементов памяти.

По быстродействию алгоритмы PRESENT лишь немного превосходят алгоритмы Belt и ГОСТ 28147-89, но вдвое уступают алгоритму AES-128.

Возможно, что на других аппаратных платформах алгоритмы PRESENT могут иметь значительные преимущества перед обычными криптографическими алгоритмами, но в базе FPGA они явно неконкурентоспособны.

Анализ характеристик алгоритма CLEFIA-128 и сравниваемых с ним стандартных алгоритмов показывает, что алгоритм CLEFIA-128 по всем показателям (ресурсоемкости, быстродействию) значительно уступает известным стандартным алгоритмам и поэтому ни по каким показателям не может быть отнесен к «облегченным».

Исследование характеристик алгоритмов CLEFIA-192 и CLEFIA-256 не проводилось, поскольку показатели для этих алгоритмов будут значительно хуже, чем у CLEFIA-128, из-за увеличения количества раундов (22 и 26 соответственно вместо 18).

Возможно, что на других аппаратных платформах (например, специально разработанных интегральных схемах), характеристики алгоритмов CLEFIA будут лучше, чем на платформе микросхем типа FPGA, но существенных изменений ожидать не приходится, поскольку и сложность алгоритмов и набор выполняемых операций остаются неизменными при реализации на любой платформе.

Список литературы

1. PRESENT: An ultra-lightweight Block Cipher / A. Bogdanov [et al.] // Proc. of CHES 2007. – Springer-Verlag, 2007. – P. 450–466.

2. Poschmann, A. Lightweight Cryptography – Cryptographic Engineering for a Pervasive World : dissertation for the degree Doktor-Ingenieur Faculty of Electrical Engineering and Information Technology Ruhr-University Bochum, Germany [Electronic resource]. – Mode of access : eprint.iacr/2009/516.pdf. – Date of access : 16.06.2014.

3. Панасенко, С. Облегченные алгоритмы шифрования / С. Панасенко, С. Смагин // Мир ПК. – 2011. – № 7. – С. 50–52.

4. Агафьин, С.С. LW-криптография: шифры для RFID-систем / С.С. Агафьин // Безопасность информационных технологий. – 2011. – № 4.

5. ISO/IEC 29192-2:2012. Information technology – Security techniques – Lightweight Cryptography – Part 2: Block ciphers.

6. Masanobu, Katagi. Lightweight Cryptography for the Internet of Things // Masanobu Katagi, Shibo Moriai [Electronic resource]. – Mode of access : <http://www.iab.org/wp-content/IAB-uploads/2011/030Kaftan.pdf>. – Date of access : 10.01.2014.

7. The 128-Bit blockcipher CLEFIA / T. Shirai [et al.] // Proceedings of Fast Software Encryption – FSE'07 / ed. A. Biryukov. – Springer-Verlag, 2007. – Vol. 4593. – P. 181–195.

8. Announcing the Advanced Encryption Standard (AES) // Federal Information Processing Standards Publication [Electronic resource]. – Mode of access : <http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf>, <http://www.nist.gov/CryptoToolkit>. – Date of access : 12.12.2013.

9. Системы обработки информации. Защита криптографическая. Алгоритм криптографического преобразования : ГОСТ 28147-89. – М. : Изд-во стандартов, 1989. – 28 с.

10. Алгоритм шифрования Belt // С.В. Агиевич [и др.]. – Управление защитой информации. – 2002. – № 4. – С. 407–412.

11. Информационные технологии. Защита информации. Криптографические алгоритмы шифрования и контроля целостности : СТБ 34.101.31–2011. – Введ. 31.01.2011. – Минск : Госстандарт, 2011. – 31 с.

12. Пудовченко, Ю.Е. Когда наступит время подбирать ключи / Ю.Е. Пудовченко // Защита информации. – Конфидент. – 1998. – № 3. – С. 65–71.

13. Поляков, А.С. Характеристики аппаратной реализации некоторых симметричных алгоритмов шифрования / А.С. Поляков, В.Е. Самсонов // Информатика. – 2011. – № 1. – С. 89–94.

Поступила 14.05.2014

*Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: sveby@mail.ru
alexpolja@tut.by*

A.S. Poljakov, V.E. Samsonov

ANALYSIS OF THE CHARACTERISTICS OF INTERNATIONAL STANDARD ALGORITHMS «LIGHTWEIGHT CRYPTOGRAPHY» – ISO/IEC 29192-3:2012

The data on the characteristics of international standard algorithms «lightweight cryptography» while application in hardware implementation based on microchips of FPGA are provided. A comparison of the characteristics of these algorithms with the characteristics of several widely-used standard encryption algorithms is made and possibilities of lightweight cryptography algorithms are evaluated.

ПРАВИЛА ДЛЯ АВТОРОВ

1. Статьи принимаются в редакцию через электронную систему подачи по адресу <http://jinfo.bas-net.by> в формате файлов текстовых редакторов Microsoft Word 97 и Word 2000 для Windows. Основной текст статьи набирается с переносами шрифтом Times New Roman 11 пт, интервал между строками – одинарный, абзацный отступ 1 см, поля по 2,5 см со всех сторон.

2. Объем статьи не должен превышать 12 страниц (включая таблицы, иллюстрации, список литературы), количество иллюстраций – не больше пяти. Допускаются краткие сообщения до трех страниц.

3. Статья должна иметь индекс УДК (универсальная десятичная классификация).

4. Название статьи, фамилии всех авторов и аннотация должны быть переведены на английский язык. Для каждого из авторов приводится развернутое название учреждения с полным почтовым адресом, а также номер телефона и электронный адрес (e-mail) для связи с редакцией.

5. Формулы, иллюстрации, таблицы, встречающиеся в статье, должны быть пронумерованы в соответствии с порядком цитирования в тексте. Ссылки на рисунки и таблицы в тексте обязательны. Необходимо избегать повторения одних и тех же данных в таблицах, графиках и тексте статьи.

Рисунки должны быть выполнены с хорошим разрешением в масштабе, позволяющем четко различать надписи и обозначения. Подрисовочные подписи с расшифровкой всех позиций, представленных на рисунке, набираются шрифтом гарнитуры основного текста, размер символов 9 пт. Цветные иллюстрации печатаются только в том случае, когда это необходимо для понимания излагаемого материала.

6. Набор формул выполняется в формульных редакторах Microsoft Equation или Math Type и должен быть единообразным по применению шрифтов и знаков по всей статье.

Прямо () набираются: греческие и русские буквы; математические символы (\sin , \lg , ∞); символы химических элементов (C, Cl, CHCl_3); цифры (римские и арабские); векторы; индексы (верхние и нижние), являющиеся сокращениями слов.

Курсивом (~) набираются: латинские буквы – переменные, символы физических величин (в том числе и в индексе).

7. Сокращения в тексте статьи (за исключением единиц измерения) могут быть использованы только после упоминания полного термина. Единицы измерения физических величин следует приводить в Международной системе СИ.

8. Литература приводится автором общим списком в конце статьи. Ссылки на литературу в тексте идут по порядку и обозначаются цифрой в квадратных скобках. Ссылаться на неопубликованные работы не допускается. С примерами оформления библиографического описания в списке литературы можно ознакомиться в приложении 2 к *Инструкции по оформлению диссертации, автореферата и публикаций по теме диссертации* на сайте Высшей аттестационной комиссии Республики Беларусь <http://vak.org.by>.

9. Поступившие в редакцию статьи направляются на рецензирование специалистам. Основным критерием целесообразности публикации является новизна и информативность статьи. Если по рекомендациям рецензента статья возвращается автору на доработку, а переработанная рукопись вновь рассматривается редколлегией, датой поступления считается день получения редакцией ее окончательного варианта. Статьи не по профилю журнала возвращаются авторам после заключения редколлегии.

10. Статьи, направляемые на доработку, должны быть возвращены в исправленном виде с ответами на все вопросы.

11. Редакция журнала предоставляет возможность первоочередного опубликования статей, представленных лицами, которые осуществляют послеузовское обучение (аспирантура, докторантура, соискательство) в год завершения обучения.

12. Авторы несут ответственность за направление в редакцию статей, уже опубликованных ранее, или статей, принятых к публикации другими изданиями.

13. Редакция оставляет за собой право на редакционные изменения, не искажающие основное содержание статьи.

Журнал «Информатика» включен Высшей аттестационной комиссией Республики Беларусь в список научных изданий для опубликования результатов диссертационных исследований.

Индексы

00827

для индивидуальных
подписчиков

008272

для предприятий и
организаций