

ISSN 1816-0301

ИНФОРМАТИКА

3(47)

ИЮЛЬ-СЕНТЯБРЬ
2015

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, 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(47) • июль-сентябрь 2015

СОДЕРЖАНИЕ

ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ И РЕЧИ

- Дмитрук А.А.** Сравнительный анализ эффективности различных видов дескрипторов в задаче поиска медицинских изображений5
- Ключеня В.В., Петровский Н.А.** Быстрое прототипирование встраиваемых программируемых систем на ПЛИС для мультимедийных приложений13
- Ланкевич Ю.Ю.** Алгоритмы сборки общего изображения топологии СБИС29
- Васильева Ю.Д.** Компьютерное моделирование лица пациента на основе боковых ортогнатических моделей42

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

- Дудин С.А., Дудина О.С.** Исследование многолинейной системы массового обслуживания с абсолютным приоритетом и повторными вызовами51
- Швед О.Л., Абрамов А.А.** Переход от активного нагружения к пассивному при математическом моделировании нелинейно упругопластического процесса62
- Гущинский Н.Н., Баттайа О., Долгий А.** Модели и методы синтеза структуры технологического процесса обработки деталей на поточной линии из станков со стационарным приспособлением72

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

- Стрижнев А.Г., Шихов А.А., Русакович А.Н.** Метод исследования нелинейных объектов управления, работающих в заданном диапазоне выходных координат81

Шашков А.С. Проектирование цифровых систем с пониженным энергопотреблением с применением технологии UPF-описания подсистемы питания.....90

ЗАЩИТА ИНФОРМАЦИИ

**Фисенко В.К., Максимович Е.П., Степанян А.Б., Дмитриев В.А., Ким Д.С.,
Мартинovich Т.С.** Методика экспертной оценки полноты и качества исходных данных системы защиты информации105

Редактор Г.Б. Гончаренко
Корректор А.А. Михайлова
Компьютерная верстка О.Б. Бутевич

Сдано в набор 10.08.2015. Подписано в печать 01.09.2015.
Формат 60×84 1/8. Бумага офсетная. Гарнитура Таймс. Ризография.
Усл. печ. л. 13,5. Уч.-изд. л. 13,2. Тираж 70 экз. Заказ 10.

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.

© Объединенный институт проблем информатики
Национальной академии наук Беларуси, 2015

THE UNITED INSTITUTE OF INFORMATICS PROBLEMS
OF THE NATIONAL ACADEMY OF SCIENCES OF BELARUS

INFORMATICS

PUBLISHED QUATERLY

Issued since 2004

№ 3(47) • Jule-September 2015

CONTENTS

SIGNAL, IMAGE AND SPEECH PROCESSING

- Dmitruk A.A.** Comparative study of descriptors for medical image retrieval.....5
Kliuchenia V.V., Petrovsky N.A. Rapid prototyping of embedded programmable systems on
FPGA for multimedia applications.....13
Lankevich Y.Y. An algorithm for assembling a common image of VLSI layout29
Vasilyeva Y.D. Patient's face computer-assisted simulation on the basis of lateral-view
orthognathic models42

MATHEMATICAL MODELING

- Dudin S.A., Dudina O.S.** Analysis of multi-server queueing system with preemptive priority
and repeated calls51
Shwed O.L., Abramow A.A. Transition from active to passive weighting in nonlinear
elasto-plastic process.....62
Guschinsky N.N., Battaia O., Dolgui A. Models and methods for synthesis of manu-
facturing process structure for machining at flow line of stationary machines.....72

APPLIED INFORMATION TECHNOLOGIES

- Stryzhniou A.G., Shykhau A.A., Rusakovich A.N.** Method of investigation of nonlinear
objects operating within specified range of output coordinates81

Shashkov A.S. Design of low-power electronic systems using UPF power intent specification technology	90
---	----

INFORMATION SECURITY

Fisenko V.K., Maksimovich E.P., Stepanyan A.B., Dmitriev V.A., Kim D.S., Martynovich T.S. Methodological background of expert estimation of initial data completeness and quality according to the certified information security system	105
---	-----

ОБРАБОТКА СИГНАЛОВ, ИЗОБРАЖЕНИЙ И РЕЧИ

УДК 004.93

А.А. Дмитрук

**СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЭФФЕКТИВНОСТИ РАЗЛИЧНЫХ
ВИДОВ ДЕСКРИПТОРОВ В ЗАДАЧЕ ПОИСКА
МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ**

Приводятся результаты сравнительного анализа эффективности различных видов дескрипторов в контексте задачи поиска медицинских изображений по содержанию. Поиск осуществляется по 11 тестовым базам с использованием 137 видов дескрипторов. Наилучшие результаты показывают дескрипторы на основе матриц совместной встречаемости, а также локальных бинарных шаблонов.

Введение

Коллекции медицинских изображений являются ценным источником информации, играющим важную роль в образовании, исследованиях по медицинской тематике и поддержке принятия решений в клинической практике. При этом большую проблему представляет постоянный рост размера цифровых архивов медицинских учреждений в связи с большой доступностью техники, способной производить цифровые снимки. К такой технике относятся рентгеновские аппараты, томографы, ультразвуковые сканеры и микроскопы. Среднестатистическое отделение радиологии в настоящее время производит несколько терабайт данных в год [1]. Указанные причины побуждают исследователей к разработке инструментов для быстрого и эффективного доступа к данным в архивах медицинских изображений. Индексация и каталогизация архивов изображений традиционно выполняются вручную, в связи с чем являются довольно трудоемкими процедурами, которые нередко подвержены ошибкам [2]. Таким образом, существует практическая необходимость в автоматизации индексирования коллекций медицинских изображений для улучшения скорости и качества поиска по ним.

Системы поиска медицинских изображений обычно позволяют осуществлять текстовый поиск по аннотациям или описаниям к изображениям, которые были заранее внесены вручную. В последние десятилетия в дополнение к традиционному текстовому поиску возник поиск изображений по их содержанию (Content-Based Image Retrieval, CBIR). Его отличительной особенностью является отсутствие необходимости в ручной аннотации изображений. Вместо этого содержание изображений автоматически описывается с помощью визуальных признаков (например, характеристик цвета, формы или текстуры). Поиск похожих изображений осуществляется на основе сравнения визуальных признаков (дескрипторов) изображения-образца и остальных изображений коллекции с целью нахождения наиболее близких в пространстве признаков.

В настоящее время существует огромное число видов дескрипторов изображений, применяющихся для широкого круга задач. Однако дескрипторы, положительно показавшие себя для одного типа изображений и решаемых задач, часто оказываются неэффективными для других [3]. Поскольку эффективность поиска зависит от многих причин, включающих тип используемой метрики сравнения дескрипторов, тип применяемого классификатора, способ оценки качества поиска, объективное сравнение дескрипторов возможно лишь при неизменности указанных условий. Существующая проблема объективного сравнения дескрипторов отражена в литературе. В частности, в статье [4] на примере известной базы изображений COREL показано, как, манипулируя процедурой оценки качества поиска, а также самими данными, можно улучшать характеристики системы поиска по содержанию. В итоге делается вывод о том, что объективное сравнение таких систем практически невозможно, поскольку во многих статьях подробное описание процедуры их оценки не приводится. В качестве возможного решения данной проблемы авторы указывают унификацию не только тестовых баз медицинских изображений, но и процедур оценки качества поиска. Эти условия выполняются на международных мероприятиях, например ImageCLEF [5] или PASCAL VOC [6], по сравнению различных алгоритмов на одинаковых дан-

ных и одним и тем же способом. Таким образом, настоящее исследование ставит целью сравнительный анализ различных видов дескрипторов применительно к задаче поиска медицинских изображений по заданному образцу.

1. Тестовые базы изображений

Для тестирования были выбраны общедоступные базы преимущественно медицинских изображений. Краткая характеристика 11 тестовых баз, включающая размер баз и число исковых классов, приведена в табл. 1.

Таблица 1

Тестовые базы изображений			
Тип изображений	Характеристика	Число изображений/ классов	Разрешение, пиксели
Текстуры	Натуральные текстуры из альбома Бродаса [7, 8]	999/111	213×213
Текстуры	База изображений отсегментированных предметов COIL-20 при разных положениях камеры относительно предмета [9]	1440/20	128×128
Биология	База снимков пыльцы различных видов растений Pollen [10]	630/7	25×25
Цитология	Флуоресцентные микроскопические снимки различных органоидов (постоянные специализированные структуры) культуры клеток HeLa, способных делиться бесконечное число раз [11]	862/10	382×382
Цитология	Флуоресцентные микроскопические снимки различных органоидов (постоянные специализированные структуры) культуры клеток яичников китайского хомячка CHO, которые широко используются в различных генетических исследованиях [11]	327/5	512×512
Гистология	Гистологические снимки яичника и щитовидной железы при 200-кратном увеличении (ткани обработаны специальным иммуногистохимическим маркером D2-40, подсвечивающим сеть лимфатических сосудов, окружающих опухоль)	50 000/2	512×512
Рентген	База японской ассоциации радиотехников JSRT, включающая рентгеновские изображения грудной клетки высокого разрешения с узловыми образованиями и без таковых [12]	247/2	2048×2048
Рентген	Снимки грудной клетки здоровых мужчин и женщин [13]	3000/2	950×1000
Компьютерная томография	КТ-изображения узловых образований (подвыборка базы из [14]) и остальных участков легкого	250/2	40×40
Лица	База изображений лиц AT&T с различным выражением эмоций [15]	400/40	92×112
Лица	База изображений лиц Yale Face Database с различным выражением эмоций, условиями освещенности, наличием закрывающих лицо деталей [16]	166/15	320×243

2. Типы дескрипторов изображений

В табл. 2 приведены виды преобразований и типы дескрипторов изображений, участвовавшие в сравнительном анализе. Основная часть дескрипторов подробно описана в [17, 18] и реализована в общедоступной утилите WND-CHARM. В указанной утилите признаки вычислялись как по исходным изображениям, так и по изображениям после их преобразований. Комбинация всех видов преобразований изображений и дескрипторов породила 124 различных типа дескрипторов. Кроме того, были рассчитаны дескрипторы на основе матриц совместной встречаемости, где в качестве характеристик отдельных пикселей элементарных структур рассматривались яркость I , градиент яркости G и анизотропия A (угол между векторами градиентов). Были сформированы две группы дескрипторов, включающие элементарные структуры в виде пар (ID, GGD, AAD) и триплетов (IID, GGGD, AAAD) пикселей. Длина стороны элементарных структур матриц была 1, 3 и 5 пикселей, а диапазон значений параметра пикселя (яркость, градиент, угол) разбивался на 16 интервалов. Подробное описание указанных дескрипторов приведено в работе [19]. Дополнительные дескрипторы включали обычные гистограммы, гистограммы ориентированного градиента и локальные бинарные шаблоны. В итоге в сравнительном анализе участвовало 137 типов дескрипторов. Все дескрипторы были редуцированы при помощи метода главных компонент с сохранением 95 % исходной вариации данных.

Таблица 2

Типы дескрипторов и виды преобразований изображений

Обозначение	Название
<i>Виды преобразований изображений в утилите WND-CHARM [17]</i>	
Chebyshev()	Преобразование на основе полиномов Чебышева, где порядок полинома соответствует размеру изображения
Edge()	Выделение углов с помощью оператора Прюитта
Fourier()	Преобразование Фурье
Wavelet()	Вейвлет-преобразование на основе симплета пятого порядка с уровнем разложения 1
<i>Типы дескрипторов в утилите WND-CHARM [17]</i>	
Chebyshev-Fourier Coefficients	Коэффициенты двухмерного преобразования Чебышева – Фурье
Chebyshev Coefficients	Коэффициенты полиномов Чебышева
Comb Moments	Первые четыре момента, рассчитанные для четырех направлений 0° , 45° , -45° и 90°
Edge Features	Различные статистики, рассчитанные по градиентам Прюитта
Fractal Features	Фрактальные признаки, которые вычислялись как среднее значение абсолютной разности яркостей пар пикселей по горизонтали и вертикали на различных расстояниях (масштабах)
Gabor Texture	Признаки, вычисленные по изображению после преобразования Габора
Gini Coefficient	Коэффициент Джини, который рассчитывался как среднее значение абсолютной разности яркостей всех пикселей изображения и его средней яркости
Haralick Textures	Четырнадцать признаков, вычисленные по классической матрице Харалика
Inverse-Otsu Object Features	Статистические показатели, вычисленные по инвертированному бинаризованному методом Оцу изображению
Multiscale Histograms	Четыре гистограммы яркостей пикселей с 3, 5, 7 и 9 интервалами
Otsu Object Features	Статистические показатели, вычисленные по бинаризованному методом Оцу изображению

Окончание табл. 2

Обозначение	Название
Pixel Intensity Statistics	Пять признаков изображения, включающие минимальную, максимальную и среднюю яркость, а также медиану и стандартное отклонение.
Radon Coefficients	Коэффициенты преобразования Радона, которые вычислялись как гистограммы проекций яркостей пикселей на линии, проходящие через центр изображения под углом 0° , 45° , 90° и 135°
Tamura Textures	Шесть признаков текстур Тамуры, включающие контраст, направленность, гистограмму с тремя интервалами и сумму элементов параметра зернистости текстуры
Zernike Coefficients	Моменты Цернике
<i>Матрицы совместной встречаемости [19]</i>	
IID	Матрица совместной встречаемости пар пикселей на расстояниях 1, 3, 5 с яркостью I
IID	Матрица совместной встречаемости триплетов пикселей с яркостью I
GGD	Матрица совместной встречаемости пар пикселей с абсолютным значением градиента G
GGGD	Матрица совместной встречаемости триплетов пикселей с абсолютным значением градиента G
AAD	Матрица совместной встречаемости пар пикселей с углом между векторами градиентов A
AAAD	Матрица совместной встречаемости триплетов пикселей с углом между векторами градиентов A
<i>Дополнительные виды признаков изображений</i>	
Historam-16	Гистограмма яркости пикселей, 16 интервалов
HoG-16-1	Гистограммы ориентированного градиента (16 интервалов), рассчитанные по исходному изображению [20]
HoG-16-3	Гистограммы ориентированного градиента, рассчитанные по изображению, которое разбито на девять частей (мозаика 3×3)
HoG-16-5	Гистограммы ориентированного градиента, рассчитанные по изображению, которое разбито на 25 частей (мозаика 5×5)
LBP-u2	Дескрипторы на основе однородных локальных бинарных шаблонов в окрестности 1, 3 и 5 пикселей [21; 22, с. 13]
LBP-ri	Дескрипторы на основе вращательно-инвариантных локальных бинарных шаблонов в окрестности 1, 3 и 5 пикселей [21; 22, с. 18]
LBP-riu2	Дескрипторы на основе однородных вращательно-инвариантных локальных бинарных шаблонов в окрестности 1, 3 и 5 пикселей [21; 22, с. 16]

Следует отметить, что рассмотрены далеко не все существующие дескрипторы. Например, по результатам распознавания различных модальностей медицинских изображений на соревнованиях ImageCLEF 2011–2013 SIFT-подобные дескрипторы показали наилучшие результаты [23], однако процедура их построения более сложная, чем для исследуемых дескрипторов. Она зависит от способа вычисления, величины используемых параметров дескрипторов, а также от способа построения так называемого «мешка признаков» (bag-of-features), поэтому указанные дескрипторы не рассматриваются в данной работе.

3. Методика постановки экспериментов и оценки качества поиска изображений

Тестирование выполнялось согласно схеме поиска изображений по образцу. В качестве образца последовательно выступало каждое изображение из тестовой базы, а поиск осуществлялся по всем остальным. Результат поиска считался правильным, если найденное изображение принадлежало к тому же классу, что и образец для поиска.

В литературе при оценке систем поиска по содержанию изображений с заранее заданными классами часто используют следующую оценку [24]:

$$RR(q) = \frac{NF(\alpha, q)}{GT(q)} \times 100 \%,$$

где $RR(q)$ – качество поиска (retrieval rate) по q запросам, принимающее значения от 0 % (не найдено ни одного изображения) до 100 % (все найденные изображения искомого класса) для q запросов;

$GT(q)$ – размер базы изображений заранее заданных искомого классов (ground truth) для запроса q ;

$NF(\alpha, q)$ – число изображений искомого класса, найденных среди первых $\alpha \times GT(q)$ результатов для запроса q .

Таким образом, качество поиска определялось как процент правильно найденных изображений, т. е. изображений, принадлежащих к тому же классу, что и образец для поиска, от их общего числа вне зависимости, на каком месте они находились.

В рассматриваемом случае качество поиска рассчитывалось для первых $q = 5$ найденных изображений, наиболее близких к образцу. Степень близости вычислялась как расстояние L_1 (расстояние городских кварталов Манхэттена) между дескриптором изображения-запроса и дескрипторами остальных изображений.

4. Результаты экспериментов

Для более компактного представления результатов экспериментов использовался суммарный ранг дескриптора, т. е. место, которое занимал определенный тип дескриптора относительно остальных. Поскольку исследовалось 137 типов дескрипторов на 11 тестовых базах, минимально возможный суммарный ранг мог равняться 11, а максимальный – 1507. Первый вариант соответствует ситуации, когда дескриптор во всех случаях занимал первое место (ранг), а второй – когда последнее. В результате была сформирована табл. 3, которая содержит суммарные ранги 30 дескрипторов с наилучшим качеством поиска по всем тестовым базам.

Таблица 3

Суммарный ранг наиболее эффективных дескрипторов

Тип дескриптора	Суммарный ранг
IID	129
LBP-u2	133
Zernike Coefficients (Fourier (Wavelet ()))	155
Historam-16	168
GGD	169
IID	195
Multiscale Histograms	202
Zernike Coefficients (Fourier (Edge ()))	211
Zernike Coefficients (Fourier ())	213
LBP-ri	219
HoG-16-3	263
Fractal Features (Wavelet ())	289
LBP-riu2	306
Fractal Features (Fourier (Chebyshev ()))	325
GGGD	357
Multiscale Histograms (Wavelet ())	367
AAAD	401
HoG-16-5	406
Edge Features	431

Окончание табл. 3

Тип дескриптора	Суммарный ранг
AAD	435
Zernike Coefficients (Edge ())	437
Fractal Features	446
Multiscale Histograms (Fourier (Chebyshev ()))	450
Comb Moments	452
Zernike Coefficients	455
Gabor Textures	457
Pixel Intensity Statistics	464
Radon Coefficients	465
Fractal Features (Chebyshev (Fourier ()))	485
Fractal Features (Fourier (Wavelet ()))	502

Наилучшие результаты показали дескрипторы на основе матриц совместной встречаемости как пар (IID, GGD, AAD), так и триплетов (IID, GGD, AAAD) пикселей. Похожие результаты оказались у дескрипторов на основе локальных бинарных шаблонов (LBP-u2, LBP-gi, LBP-riu2). Процедура вычисления локальных бинарных шаблонов похожа на процедуру вычисления матриц совместной встречаемости: в ней так же используются пары пикселей в некоторой окрестности, которые затем определенным образом кодируются. Возможно, данный факт объясняет их похожие описательные способности. Высокие результаты показали моменты Цернике, вычисленные по изображениям после преобразования Фурье, а также классические гистограммы, гистограммы ориентированного градиента и фрактальные признаки.

Результаты сравнения дескрипторов, полученные для соответствующих семи тестовых баз изображений в [17, 18], в целом согласуются с результатами, представленными в настоящей статье. В указанных статьях точность поиска оценивалась с помощью перекрестной проверки с разбиением набора данных на 50 частей и с использованием заранее отобранных наиболее информативных признаков.

Заключение

В результате сравнительного анализа различных типов дескрипторов при решении задачи поиска медицинских изображений по образцу наилучшие результаты показали дескрипторы на основе матриц совместной встречаемости, а также на основе локальных бинарных шаблонов.

Следует отметить, что результаты сравнительного анализа дескрипторов справедливы лишь для рассмотренных типов дескрипторов и изображений и не могут быть распространены на все типы медицинских изображений в силу их большого разнообразия и широкого круга решаемых на их основе задач.

Список литературы

1. Partik, B. Digital (R) Evolution in Radiology / B. Partik, C. Schaefer-Prokop // Digital radiology in chest imaging. – Springer Vienna, 2001. – P. 189–203.
2. Quality of DICOM header information for image categorization / M.O. Guld [et al.] // Medical Imaging 2002. International Society for Optics and Photonics. – San Diego, 2002. – P. 280–287.
3. Deselaers, T. Features for image retrieval: an experimental comparison / T. Deselaers, D. Keysers, H. Ney // Information Retrieval. – 2008. – Vol. 11, № 2. – P. 77–107.
4. Müller, H. The truth about Corel-evaluation in image retrieval / H. Müller, S. Marchand-Maillet, T. Pun // Image and Video Retrieval. – Springer Berlin Heidelberg, 2002. – P. 38–49.
5. ImageCLEF – The CLEF Cross Language Image Retrieval Track [Electronic resource]. – 2015. – Mode of access : <http://www.imageclef.org>. – Date of access : 15.05.2015.
6. Pattern Analysis, Statistical Modelling and Computational Learning [Electronic resource]. – 2015. – Mode of access : <http://www.pascal-network.org>. – Date of access : 15.05.2015.

7. Brodatz, P. Textures: A Photographic Album for Artists and Designers / P. Brodatz. – N.Y. : Dover Publications, 1966. – 112 p.
8. Brodatz texture database with CASIA V3 Iris database naming scheme [Electronic resource]. – 2012. – Mode of access : <http://staff.neu.edu.tr/~kkilic/prj/lac/brodatz/brodatz.html>. – Date of access : 11.01.2015.
9. Columbia University Image Library (COIL-20) [Electronic resource]. – 2013. – Mode of access : <http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php>. – Date of access : 01.06.2015.
10. Geometric features of Pollen grains [Electronic resource]. – 2013. – Mode of access : <http://ome.grc.nia.nih.gov/iicbu2008/pollen>. – Date of access : 01.06.2015.
11. Identifying Sub-cellular Organelles [Electronic resource]. – 2013. – Mode of access : <http://ome.grc.nia.nih.gov/iicbu2008/hela/index.html>. – Date of access : 01.06.2015.
12. JSRT Database [Electronic resource]. – 2013. – Mode of access : <http://www.jsrt.or.jp/jsrt-db/eng.php>. – Date of access : 01.06.2015.
13. Ковалев, В.А. Влияние мер близости в пространстве признаков на качество поиска медицинских изображений по содержанию / В.А. Ковалев, А.А. Дмитрук // Информатика. – 2011. – № 30. – С. 5–11.
14. Дмитрук, А.А. Обнаружение узловых образований в легком по данным компьютерной томографии / А.А. Дмитрук // Информатика. – 2011. – № 29. – С. 25–33.
15. The Database of Faces [Electronic resource]. – 2013. – Mode of access : <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>. – Date of access : 01.06.2015.
16. Yale Face Database [Electronic resource]. – 2013. – Mode of access : <http://vision.ucsd.edu/content/yale-face-database>. – Date of access : 01.06.2015.
17. WND-CHARM: Multi-purpose image classification using compound image transforms / N. Orlov [et al.] // Pattern recognition letters. – 2008. – Vol. 29, № 11. – P. 1684–1693.
18. General pattern recognition in images using WND-CHARM [Electronic resource]. – 2013. – Mode of access : <http://ome.grc.nia.nih.gov/wnd-charm>. – Date of access : 01.06.2015.
19. Ковалев, В.А. Анализ структуры трехмерных медицинских изображений / В.А. Ковалев. – Минск : Белорус. наука, 2008. – 263 с.
20. Dalal, N. Histograms of oriented gradients for human detection / N. Dalal, B. Triggs // Computer Vision and Pattern Recognition. CVPR 2005. IEEE Computer Society Conf. – Montbonnot-Saint-Martin, 2005. – Vol. 1. – P. 886–893.
21. A general Local Binary Pattern (LBP) implementation for Matlab [Electronic resource]. – 2013. – Mode of access : <http://www.cse.oulu.fi/CMV/Downloads/LBP Matlab>. – Date of access : 01.06.2015.
22. Pietikainen, M. Computer vision using local binary patterns / M. Pietikainen. – Springer, 2011. – Vol. 40. – 207 p.
23. Improved medical image modality classification using a combination of visual and textual features / I. Dimitrovski [et al.] // Computerized Medical Imaging and Graphics. – 2015. – Vol. 39. – P. 14–26.
24. Color and texture descriptors / B.S. Manjunath [et al.] // IEEE Transactions on Circuits and Systems for Video Technology. – 2001. – Vol. 11, № 6. – P. 703–715.

Поступила 25.06.2015

*Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: dmitruk@newman.bas-net.by*

A.A. Dmitruk

**COMPARATIVE STUDY OF DESCRIPTORS
FOR MEDICAL IMAGE RETRIEVAL**

This paper presents results of comparative analysis of image descriptors in the context of content-based medical image retrieval problem. Experiments are carried out on 11 test databases using 137 different types of descriptors. Image descriptors based on co-occurrence matrices and local binary patterns showed the best performance.

УДК 004.031.6

В.В. Ключеня, Н.А. Петровский

БЫСТРОЕ ПРОТОТИПИРОВАНИЕ ВСТРАИВАЕМЫХ ПРОГРАММИРУЕМЫХ СИСТЕМ НА ПЛИС ДЛЯ МУЛЬТИМЕДИЙНЫХ ПРИЛОЖЕНИЙ

Рассматривается процесс быстрого прототипирования мобильных мультимедийных систем трансформационного кодирования изображений по схеме L2L (lossless-to-lossy): сжатие и восстановление изображения как без потерь, так и с контролируемым внесением артефактов. За основу берется отладочный модуль Xilinx ML-401, в качестве основного управляющего ядра используется встроенный софт-процессор MicroBlaze. Процессоры дискретного косинусного преобразования (ДКП) представляются в виде модулей на языке описания аппаратуры VHDL, которые подключаются к софт-процессору MicroBlaze в виде сопроцессора, результат обработки изображений выводится через VGA-интерфейс на экран монитора. Предлагается архитектура рекурсивного процессора вычисления ДКП – обратного ДКП, которая может рассматриваться как базовая при построении кодеров класса L2L.

Введение

Интенсивное развитие информационных технологий в современном мире делает особенно важным решение проблемы быстрой и качественной передачи различного рода информации по цифровым линиям связи. В качестве передаваемой информации часто выступают такие мультимедиаданные, как изображения, видео-, аудио- и речевые сигналы. Алгоритм ДКП играет ключевую роль в приложениях компрессии как аудиоданных [1], так и изображений [2]. Несмотря на большое количество предложений аппаратной реализации алгоритма ДКП, специфические требования прикладных приложений компрессии мультимедиаданных обуславливают актуальность вопроса реализации процессора ДКП, удовлетворяющего ограничениям соответствующего приложения. Большинство решений направлено на получение заказных СБИС для специфических алгоритмов кодирования. Например, в работе [3] показана подсистема обработки мультимедиаданных на основе гибридной схемы «цифровой процессор обработки сигналов – программируемая логическая интегральная схема (ПЛИС)» (DSP-FPGA), реализуемой на печатной плате, которая гарантирует пользователю высокую гибкость настройки аппаратуры на соответствующий алгоритм обработки. Однако недостатком такого решения является повышенная потребляемая мощность (выше, чем в интегральном исполнении), а также то, что производительность системы может быть ниже.

Современные ПЛИС с архитектурой FPGA (Field Programmable Gate Array) обеспечивают уровень интеграции, сопоставимый с текущими специализированными интегральными схемами, и достаточно быстры. Данные характеристики FPGA позволяют сделать доступной реализацию нескольких встроенных процессоров, которые могут быть эффективно задействованы для получения достаточно высокой скорости обработки мультимедиаданных. Это делает реальной возможность реализации программируемой системы-на-кристалле ПЛИС (Field-Programmable System-on-Chip, FPSoC) для компрессии мультимедиаданных. Как и любая система-на-кристалле, FPSoC представляет собой один или сочетание нескольких микропроцессоров, некоторое встроенное программное обеспечение в памяти и другую аппаратную логику, которые с помощью определенных шин связи объединены на кристалле. Для получения эффективной FPSoC при решении соответствующего приложения необходимо такое распределение аппаратно-программных ресурсов, которое бы приводило к высокой производительности системы при решении прикладной задачи [4].

Традиционно к проектированию FPSoC привлекаются многочисленные группы инженеров, включая разработчиков алгоритма, программистов, проектировщиков аппаратной части и группу реализации проекта, причем каждая группа разработчиков задействована только на определенном этапе. Этот традиционный процесс разработки занимает много времени, потому что проектировщики алгоритма часто не имеют доступа к аппаратным средствам, которые фак-

тически уже созданы. Процесс быстрого прототипирования объединяет фазы проектирования, устраняя потенциально узкие места, дает возможность инженерам видеть результаты и быстро повторять решения без использования дорогих аппаратных средств.

Целью настоящей статьи является демонстрация процесса быстрого прототипирования на примере создания процессора ДКП – встраиваемого процессора в программируемую систему-на-кристалле ПЛИС реального времени на основе отладочного модуля Xilinx ML-401 с использованием ПЛИС типа FPGA семейства Virtex 4 (XC4VLX25-FF668-10B). Функциональное назначение FPSoC – компрессия изображений по схеме L2L (lossless-to-lossy) [5, 6]: сжатие и восстановление мультимедиаданных как без потерь, так и с контролируемым внесением артефактов. Система состоит из управляющего процессора, в качестве которого используется встроенный софт-процессор MicroBlaze, выполняющий квантование и энтропийное кодирование, и сопроцессора ДКП, выполняющего преобразование исходного изображения. Основными приложениями данной FPSoC могут быть, например, аэрокосмическое наблюдение и съемка, для которых характерны строгие ограничения по пропускной способности и рассеиваемой мощности. При выполнении подобных проектов по созданию подсистем компрессии изображений во внимание принимаются такие характеристики, как диапазон компрессии; минимальное число кадров, обрабатываемых в секунду; количество пикселей в кадре изображения; обеспечение возможности «легкого апгрейда» алгоритма кодирования.

1. Аппаратная платформа для быстрого прототипирования программируемых систем-на-кристалле ПЛИС

Представители всех серий нового поколения кристаллов программируемой логики могут эффективно применяться для реализации встраиваемых микропроцессорных систем, разрабатываемых на основе конфигурируемых 32-разрядных ядер семейства MicroBlaze фирмы Xilinx или NIOS компании Altera. В рамках решаемой задачи можно использовать отладочную плату Xilinx ML-401, где управляющим процессором выбирается софт-процессор MicroBlaze с широким спектром настраиваемых интерфейсов, периферийных устройств и пользовательских компонентов интеллектуальной собственности (Intellectual Property Core – IP-core), далее IP-компонентов. Встроенное в ПЛИС 32-разрядное софт-процессорное ядро MicroBlaze с RISC-архитектурой оптимизировано для создания FPSoC. Структура ядра софт-процессора MicroBlaze (рис. 1) построена по принципу гарвардской архитектуры, в основе которой лежит концепция использования отдельных шин для данных и команд [7–10]. Организация магистралей микропроцессора согласно этой концепции обеспечивает достижение высоких скоростей выполнения операций. Серым цветом на рисунке отмечены опциональные блоки, наличие которых определяется при конфигурировании системы. Для конфигурации доступны различные параметры софт-процессора: размер кеша, длина конвейера (трех- или пятиуровневый), встроенная периферия, блок управления памятью, шинные интерфейсы и т. д. Кроме того, ключевые компоненты процессора, которые редко используются, но при этом требуют значительных аппаратных затрат (т. е. блоки умножения, деления, операций с плавающей запятой и т. п.), могут быть выборочно добавлены или удалены.

Встроенный в ПЛИС софт-процессор MicroBlaze имеет несколько интерфейсных шин: шину ввода-вывода Core Connect PLB (Processor Local Bus) – универсальное средство связи с периферийными устройствами; специальную шину LMB для доступа к внутренней памяти ПЛИС (Block RAM), позволяющую снизить нагрузку на другие шины. Большинство IP-компонентов как от Xilinx, так и от других производителей подключаются напрямую к PLB или через мост между шинами PLB и OPB (On-Chip Peripheral Bus). Подключение встраиваемого процессора – сопроцессора для решения прикладной задачи возможно средствами специального соединения FIFO – FSL (Fast Simplex Link). Функциональная схема подключения пользовательских IP-блоков показана на рис. 2.

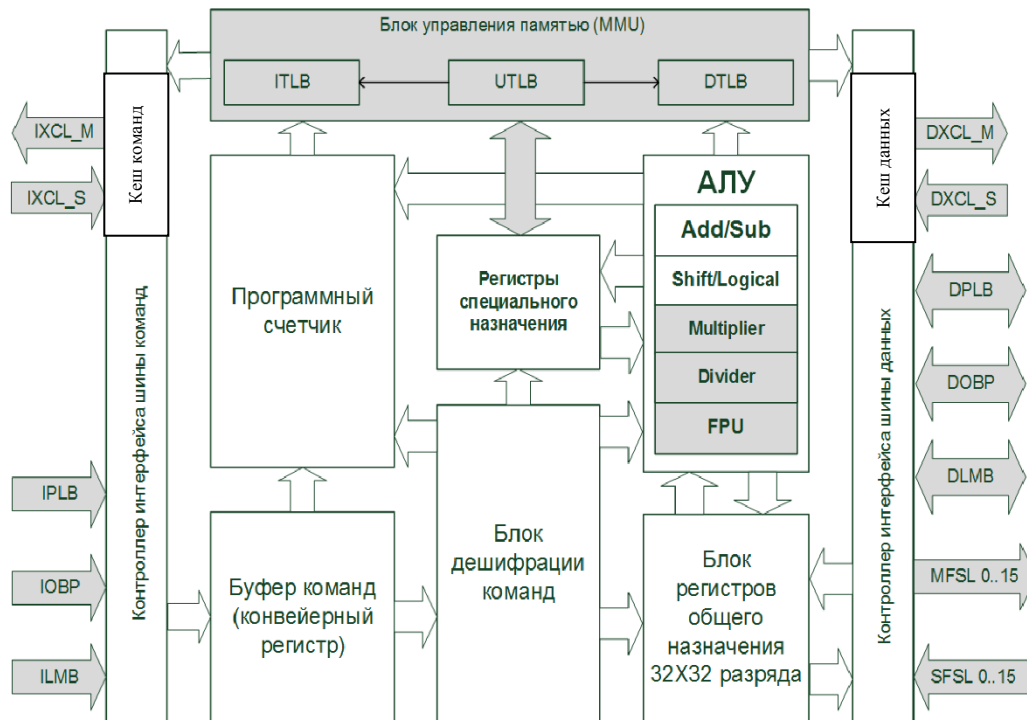


Рис. 1. Обобщенная структура микропроцессорного ядра MicroBlaze [8]

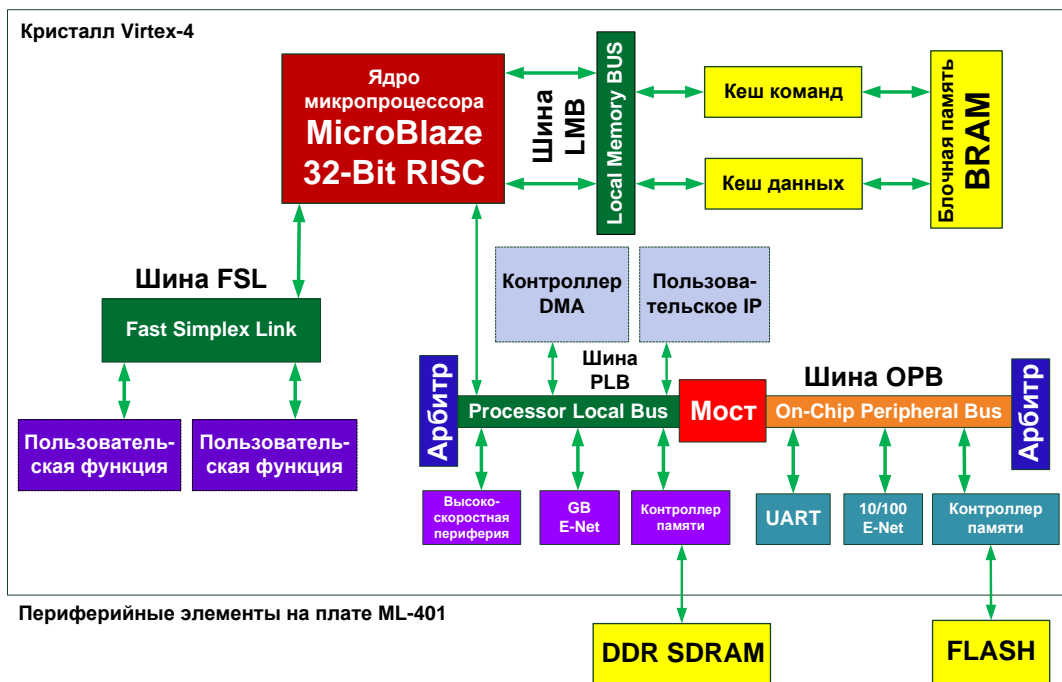


Рис. 2. Функциональная схема подключения периферийных устройств к MicroBlaze

2. Этапы быстрого прототипирования встраиваемых FPGAs

Существует много подходов совместного проектирования аппаратного и программного обеспечения. В данном проекте в качестве методологии для быстрого прототипирования программируемых систем-на-кристалле ПЛИС используется функционально-архитектурное проектирование на основе спиралевидной модели проектирования (рис. 3), которая позволяет осуществлять возврат и корректирование предыдущих этапов процесса проектирования.

Функционально-архитектурное проектирование встраиваемых FPSoC включает в себя следующие уровни абстракции [11, 12]:

- 1) задание спецификации или требований на FPSoC, в которые кроме функциональных требований включаются, например, требования реального масштаба времени (частотный диапазон), максимальная потребляемая мощность, ограничение максимальной стоимости и т. д.;
- 2) моделирование, тестирование и верификацию алгоритмического обеспечения;
- 3) структурный синтез, связанный с выбором и оценкой эффективной архитектуры процессора, соответствующей алгоритмическому обеспечению и требованиям спецификации;
- 4) аппаратную реализацию на выбранной ПЛИС с архитектурой FPGA.

На рис. 3 показаны этапы создания прототипа встраиваемых программируемых систем-на-кристалле ПЛИС, а также инструментальные среды, которые используются для проектирования таких систем. Подобная концепция быстрого прототипирования всей FPSoC на архитектурном уровне является неотъемлемой частью всех современных маршрутов проектирования [9]. Например, при разработке архитектуры FPSoC для определенного приложения может возникнуть необходимость в полной или частичной реализации какого-то прикладного IP-компонента: оценке его аппаратных затрат, потребляемой мощности, производительности, по результатам которой архитектура FPSoC может быть частично или полностью модифицирована.



Рис. 3. Этапы быстрого прототипирования программируемых систем-на-кристалле ПЛИС

3. Проектирование встраиваемого процессора ДКП в FPSoC компрессии изображений по схеме L2L

3.1. Требования к аппаратной реализации алгоритма ДКП

Преобразование, используемое в кодерах изображений по схеме L2L, должно обладать следующими свойствами [13]: обеспечивать перфективную реконструкцию исходного изображения и линейную фазово-частотную характеристику (ФЧХ), иметь конечную импульсную характеристику, быть максимально децимированным и равнополосным. Основопологающими факторами при проектировании преобразователя систем кодирования изображений FPSoC на

ПЛИС с архитектурой FPGA для достижения высокой производительности при программной и аппаратной реализациях являются модульность и регулярность (обеспечение параллельных вычислений), малая емкость буферной памяти и низкая потребляемая мощность. При этом прямое и обратное преобразования осуществляются по схеме обработки «целое в целое» с сохранением перфективной реконструкции исходного изображения (коэффициенты представляются целыми или двоичными рациональными числами; число операций умножения минимально, по возможности они исключаются из алгоритма).

Объективная оценка качества реконструкции исходного изображения осуществляется при помощи часто используемых в области сжатия изображений метрик [5] – среднеквадратической ошибки MSE и пикового отношения сигнал-шум PSNR, что позволяет сравнивать полученные результаты с результатами других систем кодирования:

$$MSE = \frac{\sum_{x,y \in A} (s_{x,y} - s'_{x,y})^2}{A}, \quad PSNR = 20 \lg \frac{255}{\sqrt{MSE}} \text{ dB},$$

где $s_{x,y}$ – значение (x, y) -го пиксела оригинального изображения, $s'_{x,y}$ соответствует значению (x, y) -го пиксела реконструированного изображения, A – общее количество пикселов в изображении.

Основным достоинством данной метрики является охват всего сигнала (в данном случае изображения) при подсчете значения оценки, а не конкретной локальной области (частотной или временной), что важно для выбора метрики при сжатии изображений.

3.2. Моделирование в программной среде MATLAB: выбор преобразования

Целочисленные ДКП (IntDCT) [14], которые реализованы на основе лестничной схемной параметризации, были предложены для кодирования изображений по схеме L2L. Однако известные IntDCT генерируют в восстановленном исходном изображении такой артефакт, как «шахматная доска» при представлении коэффициентов преобразования в формате фиксированной запятой с малой разрядностью слова. В работе [14] предложено преобразование для системы трансформационного кодирования на основе блочной лестничной структурной параметризации прямого и обратного ДКП (ДКП-ОДКП), где блоками лестничных ступеней являются матрицы прямого или обратного ДКП. Математическое описание прямого (рис. 4, а) и обратного (рис. 4, б) преобразований на основе ДКП-ОДКП следующее [14]:

– прямого

$$\begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & -\mathbf{D} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{C} & \mathbf{I} \end{bmatrix};$$

– обратного

$$\begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{C} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{D} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{C} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{0} & -\mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix}.$$

Прямое и обратное ДКП обозначены как матрицы \mathbf{C} и \mathbf{D} соответственно и функционально описываются как ДКП второго типа и ДКП третьего типа [14]:

$$[\mathbf{C}]_{m,n} = \sqrt{\frac{2}{M}} \cdot c_m \cos\left(\frac{m(n+\frac{1}{2})\pi}{M}\right), \quad [\mathbf{D}]_{m,n} = \sqrt{\frac{2}{M}} \cdot c_n \cos\left(\frac{n(m+\frac{1}{2})\pi}{M}\right),$$

где \mathbf{I} – единичная матрица; $\mathbf{D} = \mathbf{C}^{-1} = \mathbf{C}^T$, $0 \leq m, n \leq M-1$; $[\mathbf{C}]_{m,n}$ – матрица прямого ДКП второго типа, а $[\mathbf{D}]_{m,n}$ – матрица ДКП третьего типа, являющаяся обратной для ДКП второго типа;

m – количество столбцов, n – количество строк; $M = 2^n (n \in N)$, $c_m = \begin{cases} \frac{1}{\sqrt{2}} (m=0) \\ 1 (m \neq 0) \end{cases}$, $c_n = \begin{cases} \frac{1}{\sqrt{2}} (n=0) \\ 1 (n \neq 0) \end{cases}$.

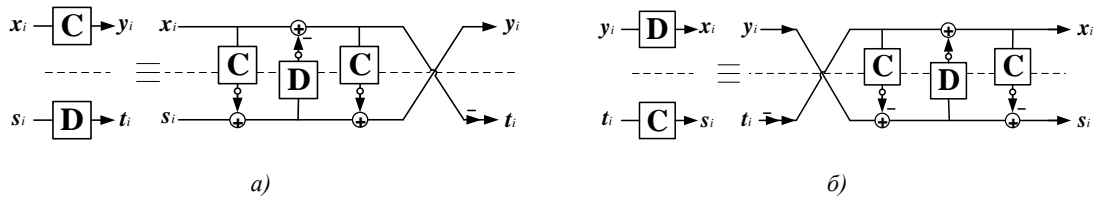


Рис. 4. Прямое (а) и обратное (б) преобразования для системы трансформационного кодирования на основе блочной лестничной структурной параметризации ДКП-ОДКП. Пустые кружки на схеме обозначают округление

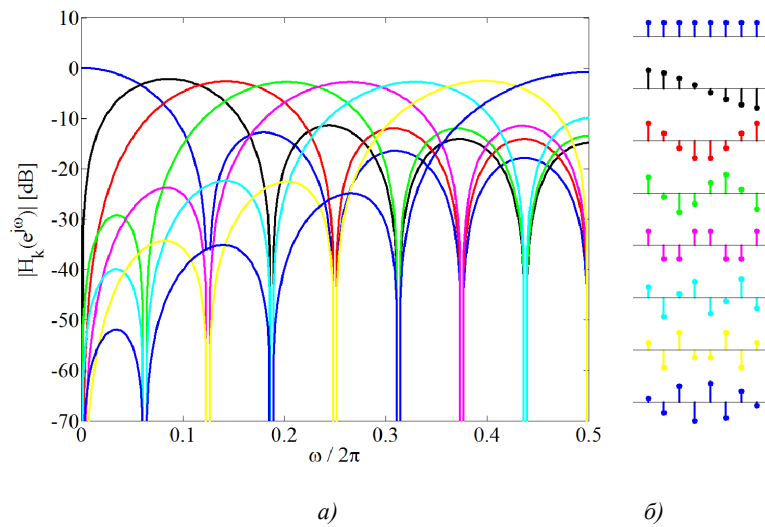


Рис. 5. Амплитудно-частотные (а) и импульсные (б) характеристики прямого ДКП

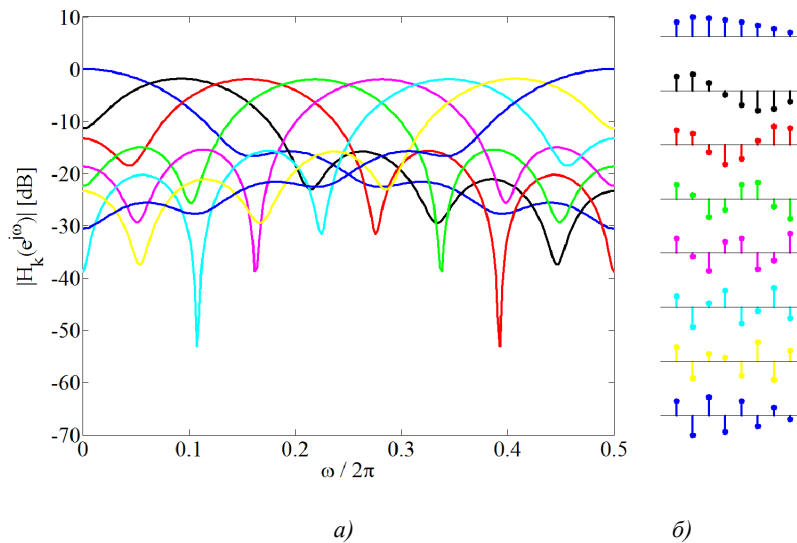


Рис. 6. Амплитудно-частотные (а) и импульсные (б) характеристики обратного ДКП

Амплитудно-частотные и импульсные характеристики восьмиточечного прямого и обратного ДКП представлены на рис. 5 и 6 соответственно. Анализ рисунков показывает, что

преобразование ДКП является равнополосным максимально децимированным банком фильтров с линейной ФЧХ (импульсные характеристики обладают свойством симметрии), постоянная составляющая локализована первым фильтром банка. Однако все фильтры ОДКП имеют постоянную составляющую (см. рис. 6, *a*), т. е. присутствует так называемый эффект DC-leakages.

В программной среде MATLAB с помощью инструментального модуля Fixed Point для моделирования вычислений на арифметике с фиксированной запятой был осуществлен расчет амплитудно-частотных характеристик (АЧХ) системы анализа-синтеза прямого и обратного преобразований ДКП-ОДКП с блочной лестничной структурной параметризацией и преобразований стандарта JPEG [13] для разрядности данных 12 бит, из которых 1 бит знаковый, 1 бит – целая часть слова, оставшиеся 10 бит представляют дробную часть слова. Анализ АЧХ данных преобразований (рис. 7) показывает, что АЧХ системы анализа-синтеза преобразования ДКП-ОДКП с блочной лестничной структурной параметризацией (рис. 7, *a*) представляет собой прямолинейную АЧХ, т. е. система анализа-синтеза данного преобразования вносит только задержку. Между тем АЧХ системы анализа-синтеза прямого и обратного преобразований стандарта JPEG [13] соответствует нелинейной функции с максимальной неравномерностью, равной $-0,03$ дБ. Таким образом, в отличие от алгоритма преобразования, используемого в стандарте сжатия изображений JPEG, преобразование на основе блочной лестничной структурной параметризации прямого и обратного ДКП-ОДКП обладает свойством перфективной реконструкции исходного изображения.

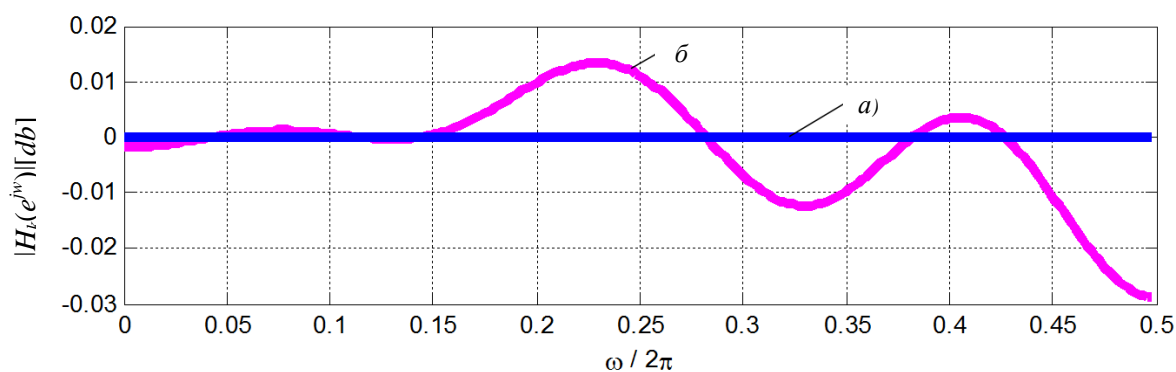


Рис. 7. АЧХ системы анализа-синтеза: *a*) преобразования ДКП-ОДКП с блочной лестничной структурной параметризацией; *б*) преобразования стандарта JPEG [9]

Проявление эффекта DC leakages в системе преобразования ДКП-ОДКП с блочной лестничной структурной параметризацией, обусловленного наличием постоянной составляющей во всех фильтрах ОДКП (см. рис. 6, *a*), что приводит к потере свойства регулярности первого рода¹ [6] банком фильтров ДКП-ОДКП, и в реконструированном исходном изображении проявляется эффект «шахматной доски». Для устранения данного недостатка в работе [14] предлагается преобразователь системы трансформационного кодирования по схеме L2L на основе диагональных матриц ДКП-ОДКП с использованием дополнительного информационного SIB-блока S_i размера $M \times M$. Формулы для прямого и обратного преобразований имеют следующий вид:

$$\begin{bmatrix} y_0 \\ \vdots \\ y_{n-1} \\ s_n \end{bmatrix} = \begin{bmatrix} \mathbf{C} & \cdots & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ \vdots & & \mathbf{C} & \vdots \\ 0 & \cdots & \cdots & \mathbf{D}^n \end{bmatrix} \cdot \begin{bmatrix} x_0 \\ \vdots \\ x_{n-1} \\ s_0 \end{bmatrix}, \quad \begin{bmatrix} x_0 \\ \vdots \\ x_{n-1} \\ s_0 \end{bmatrix} = \begin{bmatrix} \mathbf{D} & \cdots & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ \vdots & & \mathbf{D} & \vdots \\ 0 & \cdots & \cdots & \mathbf{C}^n \end{bmatrix} \cdot \begin{bmatrix} y_0 \\ \vdots \\ y_{n-1} \\ s_n \end{bmatrix},$$

¹ Система отвечает условию регулярности первого рода, когда значения амплитудно-частотных характеристик всех канальных фильтров, кроме первого, равны нулю на нулевой частоте.

где C и D – матрицы прямого и обратного ДКП. Процесс обработки изображения на основе преобразования ДКП-ОДКП с использованием дополнительного информационного SIB-блока изображен на рис. 8. Блоки, показанные пунктирной линией, могут быть факторизованы согласно блочной лестничной структурной параметризации на основе ДКП-ОДКП (см. рис. 4).

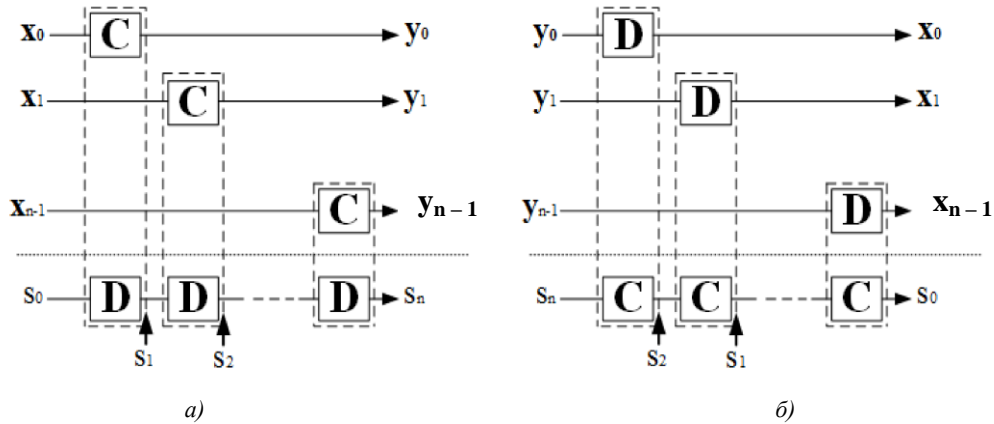


Рис. 8. Прямое (а) и обратное (б) преобразования с использованием SIB-блоков

Моделирование системы анализа-синтеза на основе блочной лестничной параметризации ДКП-ОДКП в программной среде MATLAB с помощью инструментального модуля Fixed Point с использованием арифметики с фиксированной запятой для исходного изображения Lena 512×512 пикселей (рис. 9, а) показало, что без применения информационного SIB-блока S_n наблюдается эффект шахматной доски: PSNR=84,6834 дБ, MSE = 2,2118×10⁻⁴ (рис. 9, б), а с обработкой по схеме с SIB-блоком S_n PSNR=128,888 дБ, MSE = 8,4001×10⁻⁹, т. е. исходное изображение восстанавливается без потерь (рис. 9, в).

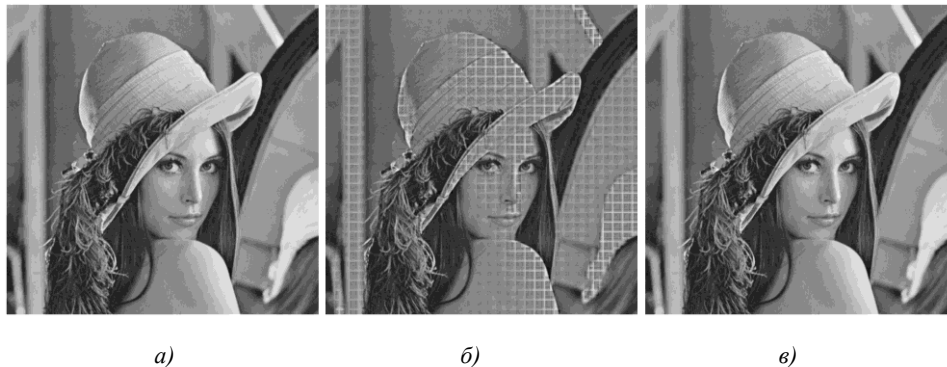


Рис. 9. Реконструкция изображения Lena 512×512 пикселей системой анализа-синтеза преобразования ДКП-ОДКП с блочной лестничной структурной параметризацией

3.3. Выбор архитектурного решения процессора преобразования ДКП-ОДКП с использованием SIB-блоков

3.3.1. Универсальный рекурсивный процессор ДКП-ОДКП

Реализация преобразователя на основе ДКП-ОДКП с использованием дополнительного информационного SIB-блока для системы трансформационного кодирования изображений по схеме L2L предполагает применение рекурсивной структуры процессора (рис. 10), которая позволит за три процессорных цикла в режиме кодирования без потерь выполнить прямое и обратное преобразования блочной лестничной структурной параметризации ДКП-ОДКП согласно рис. 4. В режиме кодирования изображения с потерями используются результаты только первого цикла обработки.

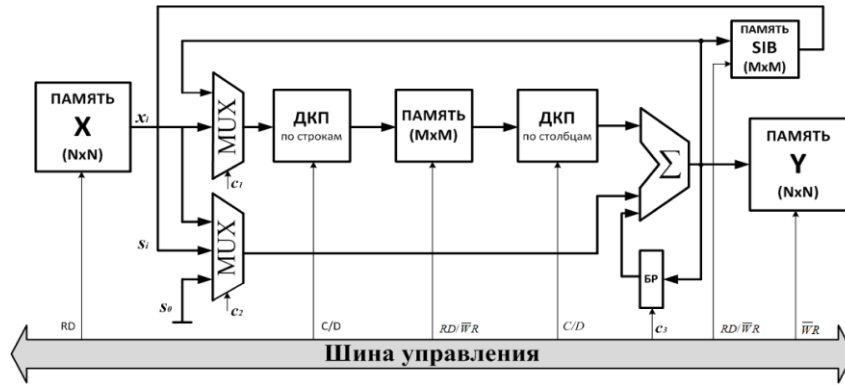


Рис. 10. Архитектура модуля преобразователя на основе ДКП-ОДКП для компрессии изображения по схеме L2L

Вычисление ДКП-ОДКП блока \mathbf{x}_i $M \times M$ входного изображения осуществляется по схеме разделимого преобразования столбцов и строк соответственно для прямого $\mathbf{C}_{2D}\mathbf{x}_i \triangleq (\mathbf{C}(\mathbf{C}\mathbf{x}_i)^T)^T = \mathbf{C}\mathbf{x}_i\mathbf{C}^T$ и обратного $\mathbf{D}_{2D}\mathbf{x}_i \triangleq (\mathbf{D}(\mathbf{D}\mathbf{x}_i)^T)^T = \mathbf{D}\mathbf{x}_i\mathbf{D}^T$ преобразований. Для этого используются два последовательно включенных через буфер памяти размера $M \times M$ процессора ДКП, настраиваемых на выполнение прямого и обратного преобразований [15, 16]. Исходное изображение \mathbf{X} размера $N \times N$ (память X исходного изображения) разделяется на блоки \mathbf{x}_i ($0 \leq i \leq n-1$, $n = (N/M)^2$); для каждого блока \mathbf{x}_i вычисляется ДКП $\mathbf{y}_i = \mathbf{C}_{2D}\mathbf{x}_i$; параллельно с процессом обработки блоков \mathbf{x}_i выполняется ОДКП SIB-блока \mathbf{s}_i : для $i=0$ SIB-блок \mathbf{s}_0 задается нулевой матрицей, а для $i = 1, 2, \dots, n-1$ SIB-блок \mathbf{s}_i будет являться итерационным преобразованием от \mathbf{s}_0 : $\mathbf{s}_i = \mathbf{D}_{2D}\mathbf{s}_{i-1}$ (см. рис. 8, а). Следует заметить, что SIB-блок $\mathbf{s}_n \neq 0$ из-за ошибки округления на каждом шаге преобразования. Реконструкция исходного изображения осуществляется из всего потока ДКП коэффициентов \mathbf{y}_n (память Y размером $N \times N$) и SIB-блока \mathbf{s}_n . Каждый блок \mathbf{x}_i и SIB-блок \mathbf{s}_i восстанавливаются на основе преобразования $\mathbf{x}_i = \mathbf{D}_{2D}\mathbf{y}_i$ и $\mathbf{s}_i = \mathbf{C}_{2D}\mathbf{s}_{i+1}$ соответственно без потерь (см. рис. 8, б). В режиме кодирования с потерями исходное изображение реконструируется на основе потока коэффициентов ДКП \mathbf{y}_n без SIB-блока \mathbf{s}_n .

3.3.2. Модифицированная схема Лофлера ДКП: лестничная схемная параметризация

Сокращение вычислительной сложности алгоритма ДКП очень важно для встраиваемых мультимедиа систем реального времени. Поэтому актуальной задачей является разработка структурных и схемотехнических решений реализации алгоритмов ДКП, которые позволяли бы уменьшить число операций умножения или избавиться от них вовсе [17]. Одним из наиболее эффективных способов реализации алгоритма ДКП является схема Лофлера [18], на основе которой алгоритмы 8-точечного или 16-точечного ДКП могут быть представлены с минимальным количеством операций умножения (рис. 11).

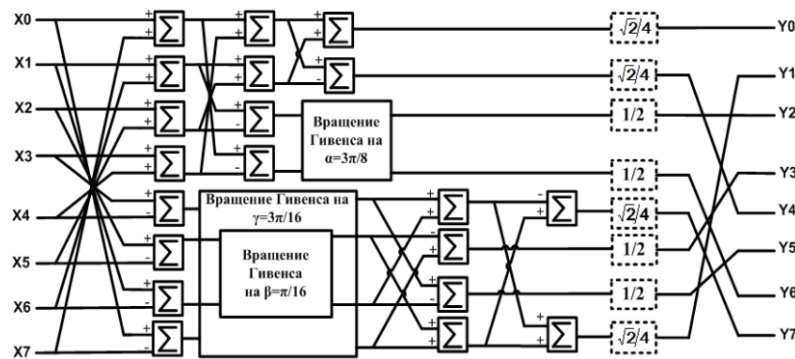


Рис. 11. Схема Лофлера 8-точечного ДКП второго типа

Основной операцией, которая вызывает трудности при аппаратной реализации, является операция умножения на поворачивающиеся множители (вращение Гивенса) – нетривиальные углы вращения. На основе лестничной схемной параметризации матрица вращения Гивенса представляется в виде следующей факторизации матриц:

$$\begin{bmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{bmatrix} = \begin{bmatrix} 1 & \tan(\alpha) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -\sin(2\alpha)/2 & 1 \end{bmatrix} \begin{bmatrix} 1/\cos(\alpha) & 0 \\ 0 & \cos(\alpha) \end{bmatrix}.$$

Например, для вращения на угол $\alpha = 3\pi/8$ (рис. 12) коэффициенты данных матриц являются рациональными числами, которые можно аппроксимировать в виде суммы степеней двойки, что позволит заменить операции умножения на операции сдвига и сложения:

$$\begin{bmatrix} 1 & 77/32 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -45/128 & 1 \end{bmatrix} \begin{bmatrix} 128/49 & 0 \\ 0 & 49/128 \end{bmatrix} = \\ = \begin{bmatrix} 1 & (2^1 + 2^{-2} + 2^{-3} + 2^{-5}) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -(2^{-2} + 2^{-4} + 2^{-5} + 2^{-7}) & 1 \end{bmatrix} \begin{bmatrix} (2^1 + 2^{-1} + 2^{-4} + 2^{-5} + 2^{-6}) & 0 \\ 0 & (2^{-2} + 2^{-3} + 2^{-7}) \end{bmatrix}.$$

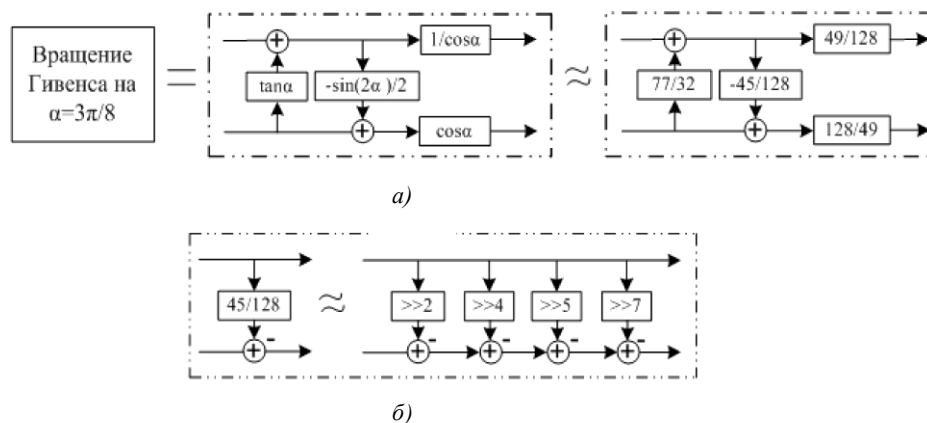


Рис. 12. Двухступенчатая лестничная схемная параметризация матрицы вращения Гивенса для угла $\alpha = 3\pi/8$:
 а) представление коэффициентов лестничной схемы рациональными числами;
 б) пример реализации ступени лестничной схемы на основе регистров сдвига

На рис. 13 показана модифицированная схема Лофлера алгоритма ДКП второго типа с лестничной схемной параметризацией на целочисленной арифметике [19]. Матрица $[C]_{m,n}$ прямого 8-точечного ДКП второго типа с учетом шагов лестничной схемной параметризации вращений Гивенса представляется в виде произведения семи разреженных матриц [11] $[C]_{m,n} = P \cdot G \cdot F \cdot E \cdot D \cdot C \cdot B \cdot A$, каждая из которых описывает структуры ступеней модифицированной схемы Лофлера ДКП. Операции с матрицами **G** и **P** в схеме ДКП могут быть объединены и вынесены в блок квантования, что позволит уменьшить вычислительную сложность алгоритма ДКП (см. рис. 11). Для достижения высокой производительности процессора 8-точечного ДКП второго типа на основе лестничных структур предлагается использовать линейную многоступенчатую конвейерную схему. Каждой матрице **A**, **B**, **C**, **D**, **E**, **F**, **G** и **P** соответствует своя ступень процессора (рис. 13) [11]. В целом архитектура каждой ступени содержит два набора регистров и арифметические модули для вычисления согласно лестничной схемной параметризации матриц вращения Гивенса соответствующих ступеней модифицированной схемы Лофлера. Каждая ступень имеет восемь входов и восемь выходов, а также свои управляющие сигналы (рис. 13, б). Каждая ступень в VHDL-описании представлена в виде конечного автомата. В свою очередь, ступень – это отдельный компонент, выход которого подается на вход следующего компонента.

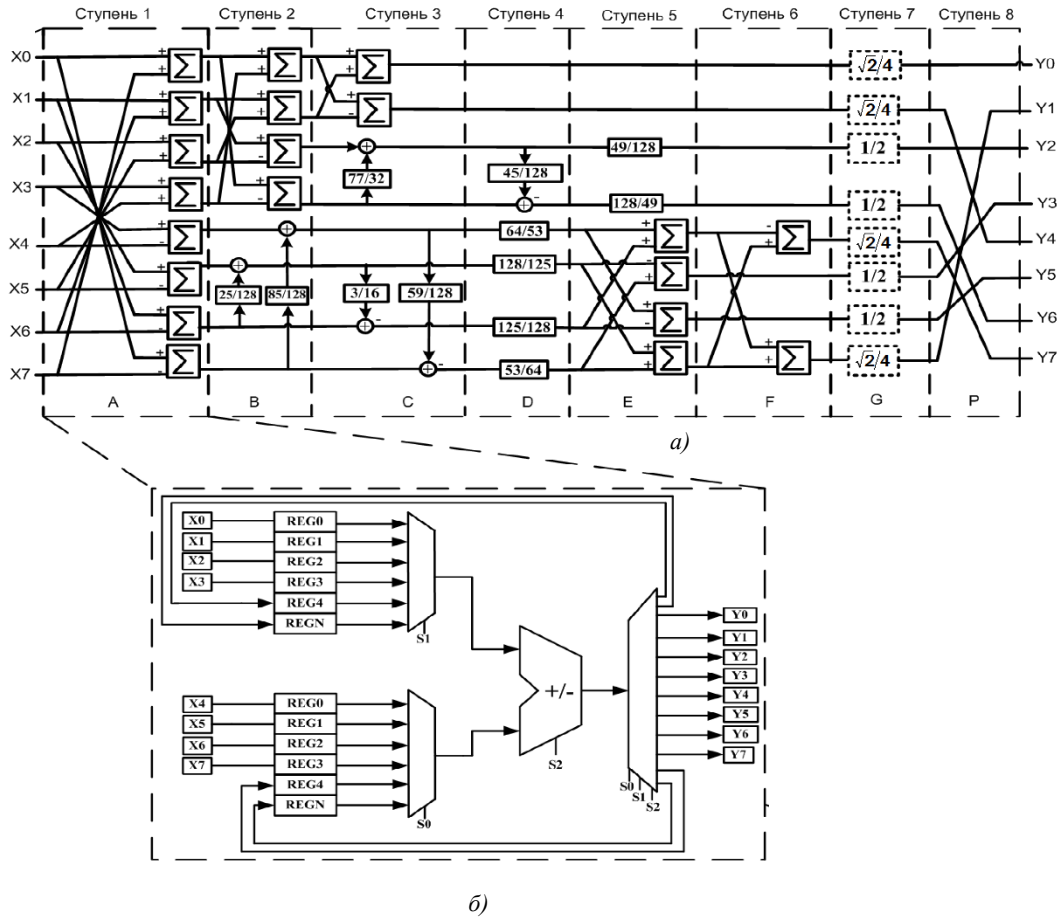


Рис. 13. Модифицированная схема Лофлера 8-точечного ДКП второго типа с лестничной схемной параметризацией на целочисленной арифметике: а) схема алгоритма (лестничная схемная параметризация матриц вращения Гивенса и представление коэффициентов лестничной схемы в виде рациональных коэффициентов); б) структура первой ступени

3.3.3. Реализация оборотов Гивенса на распределенной арифметике

Другим альтернативным решением при построении процессора ДКП, позволяющим исключить операции умножения, является применение распределенной арифметики для реализации вращения Гивенса на нетривиальный угол [20]. Прямой механизм распределенной арифметики предполагает вычисление выражения [20]

$$y = \sum_{k=1}^K a_k x_k,$$

где a_k – постоянные коэффициенты; x_k – входные нормированные переменные $|x_k| \leq 1$, которые в двоичной системе счисления в дополнительном коде представляются следующим образом: $x_k = -b_{k0} + \sum_{l=1}^{B-1} b_{kl} 2^{-l}$. Здесь B – разрядность слова; $b_{kl} \in 0,1$, b_{k0} – бит знака переменной x_k ; b_{kB-1} – наименьший значащий бит. Принимая во внимание двоичное представление входных переменных x_k , выражение для y можно записать следующим образом:

$$y = \sum_{k=1}^K a_k \left[-b_{k0} + \sum_{l=1}^{B-1} b_{kl} 2^{-l} \right] = \sum_{k=1}^K a_k (-b_{k0}) + \sum_{l=1}^{B-1} \left[\sum_{k=1}^K a_k b_{kl} \right] 2^{-l}.$$

Таким образом, сумма $\sum_{k=1}^K a_k x_k$ имеет 2^K возможных значения, которые сохраняются в памяти. Двоичные коды входных переменных x_k могут быть использованы для прямой адресации

памяти. После B итераций в сумматоре-аккумуляторе будет результат выражения для y . Сокращение объема памяти в два раза (2^{K-1}) можно получить, если входные переменные представить в смещенном двоичном дополнительном коде следующим образом [20]:

$x_k = \frac{1}{2} \left[\sum_{l=0}^{B-1} c_{kl} 2^{-l} - 2^{-(B-1)} \right]$, где $c_{kl} = b_{kl} - \bar{b}_{kl}$, $c_{k0} = -(b_{k0} - \bar{b}_{k0})$, \bar{b}_{kn} – инверсные значения битов b_{kn} , $c_{kl} \in -1, 1$. Следовательно, результат $\mathbf{y} = [y_1, y_2]^T$ вычисления операции вращения Гивенса на угол α $y_1 = cx_1 + sx_2$, $y_2 = -sx_1 + cx_2$, где $c = \cos(\alpha)$, $s = \sin(\alpha)$, с помощью распределенной арифметики можно представить в следующем виде:

$$y_1 = \sum_{l=0}^{B-1} (cc_{1l} + sc_{2l}) 2^{l-1} - (c+s) 2^{-B};$$

$$y_2 = \sum_{l=0}^{B-1} (-sc_{1l} + cc_{2l}) 2^{l-1} - (c-s) 2^{-B}.$$

Анализ данного выражения показывает, что для вычисления вектора $\mathbf{y} = [y_1, y_2]^T$ необходимы два модуля памяти (два постоянных запоминающих устройства (ПЗУ)) на два слова каждый, а для схемы Лофлера, где происходят три вращения Гивенса на нетривиальный угол, понадобится шесть ПЗУ, которые могут быть реализованы на регистрах. При последовательном побитном приеме входных данных требуется B итераций. Однако в процессе вычисления формируются все компоненты y_1, y_2 вектора \mathbf{y} , поэтому вычисление с использованием механизма распределенной арифметики является более эффективным по временным затратам, чем применение в этих же целях параллельного умножителя.

Поскольку поступление переменных x_k происходит по одному биту за один такт, такие схемы называются 1-ВААТ (one-bit-at-a-time) [20]. Распределенная арифметика позволяет распараллеливать процесс обработки, когда анализируется L бит за такт. С использованием данного подхода, исходя из условия кратности L и разрядности B представления переменных x_k , можно синтезировать, к примеру, варианты схем 1-ВААТ, 2-ВААТ, 4-ВААТ, ..., B -ВААТ, среди которых можно выбрать оптимальную реализацию одним из методов нелинейного программирования, задав определенные ограничения целевой функции, например минимум потребляемого тока, при выбранной постоянной частоте дискретизации [21, 22].

3.3.4. Анализ эффективности реализации ДКП на FPGA Virtex4 xc4vlx25-10ff668

На ПЛИС FPGA Virtex4 xc4vlx25-10ff668 была осуществлена реализация трех архитектурных решений процессора ДКП:

- модифицированной схемы Лофлера 8-точечного ДКП второго типа с лестничной схемой параметризацией на целочисленной арифметике;
- модифицированной схемы Лофлера 8-точечного ДКП второго типа с применением распределенной арифметики для реализации вращения Гивенса на нетривиальный угол (1ВААТ);
- процессора ДКП-ОДКП с использованием дополнительного информационного SIB-блока (блочной лестничной факторизации) – универсального рекурсивного процессора ДКП-ОДКП для компрессии изображений по схеме L2L. Архитектурные решения реализованы на языке описания аппаратуры VHDL [23].

В таблице приведены результаты синтеза структур ДКП на FPGA Virtex4 xc4vlx25-10ff668, анализ которых показывает, что конвейерная схема процессора ДКП имеет высокую пропускную способность, однако аппаратные затраты выше по сравнению с другими решениями. Это обусловлено тем, что конвейер процессора состоит из шести ступеней, каждая из которых является отдельным компонентом при синтезе процессора ДКП.

Преимущество реализации архитектуры процессора ДКП на распределенной арифметике (вариант 1-ВААТ) заключается в уменьшении аппаратных затрат, но при этом снижается скорость вычислений. Как отмечалось выше, распределенная арифметика позволяет распараллели-

вать процесс обработки (варианты L-ВААТ). При этом будет наблюдаться рост аппаратных затрат, зато можно искать компромиссные решения.

Результаты синтеза структур ДКП на FPGA Virtex4 xc4vlx25-10ff668

Архитектура процессора ДКП	Количество регистров	Количество четырех-входовых таблиц LUT	Максимальная тактовая частота, МГц	Среднеквадратическая ошибка	Производительность процессора, количество 8-точечных ДКП в секунду
Конвейерная схема (лестничная схемная параметризация)	1054	1594	345,781	$1,16 \times 10^{-5}$	$6,65 \times 10^6$
Распределенная арифметика (1ВААТ)	152	507	210,615	$3,43 \times 10^{-5}$	$9,57 \times 10^6$
Процессор ДКП-ОДКП и SIB-блок (блочная лестничная факторизация)	463	619	201,375	$8,40 \times 10^{-9}$	$7,76 \times 10^6$

Реализация универсального рекурсивного процессора ДКП-ОДКП для компрессии изображений по схеме L2L с использованием дополнительного информационного SIB-блока (блочная лестничная факторизация) требует меньше ресурсов кристалла FPGA, чем конвейерная схема процессора ДКП. Отличительной особенностью этого процессора является применение операционного блока, выполняющего операции как прямого, так и обратного ДКП [16], а также то, что данную схему можно использовать для обработки исходных изображений в двух режимах: с потерями и без потерь данных.

3.4. Реализация прототипа FPSoC-кодера изображения на основе софт-процессора MicroBlaze и процессора ДКП-ОДКП

Для реализации прототипа устройства используется пакет программ Xilinx EDK (Embedded Development Kit), применяемый для разработки встроенных систем на основе софт-процессора MicroBlaze. Пакет EDK состоит из двух независимых сред разработки – XPS и SDK. Разработчик использует XPS (Xilinx Platform Studio) для создания и конфигурирования аппаратной спецификации встроенной системы (процессорного ядра, управления памятью, интерфейсов ввода-вывода и др.). Пакет Xilinx SDK (Software Development Kit), основанный на Eclipse, предназначен для разработки и отладки программ на C/C++. Имеется специальный симулятор ISS (Instruction Set Simulator), благодаря которому можно тестировать программу либо в режиме симулятора, либо на основе отладочного модуля Xilinx ML-401 с ПЛИС XC4VLX25-FF668-10.

Процессор ДКП-ОДКП реализован на языке описания аппаратуры VHDL. VHDL-файлы описывают архитектуру преобразования в виде периферийного устройства, которое подключается к ядру софт-процессора MicroBlaze по быстродействующей шине FSL (вариант I) или PLB (вариант II: обмен данными под управлением контроллера DMA) (рис. 14, а) [24, 25]. В структуру FPSoC также включены контроллеры многопортовой памяти, сети Ethernet, дисплея VGA (640×480 пикселей), прерываний, клавиатуры, последовательной передачи данных (UART). Основная задача управляющего софт-процессора MicroBlaze заключается в вводе исходных данных по сети, выполнении этапа квантования и кодирования коэффициентов преобразования

ДКП-ОДКП, программа которого написана на языке C, выводе результатов преобразования на VGA-дисплей, а также в передаче их на персональный компьютер для последующего анализа.

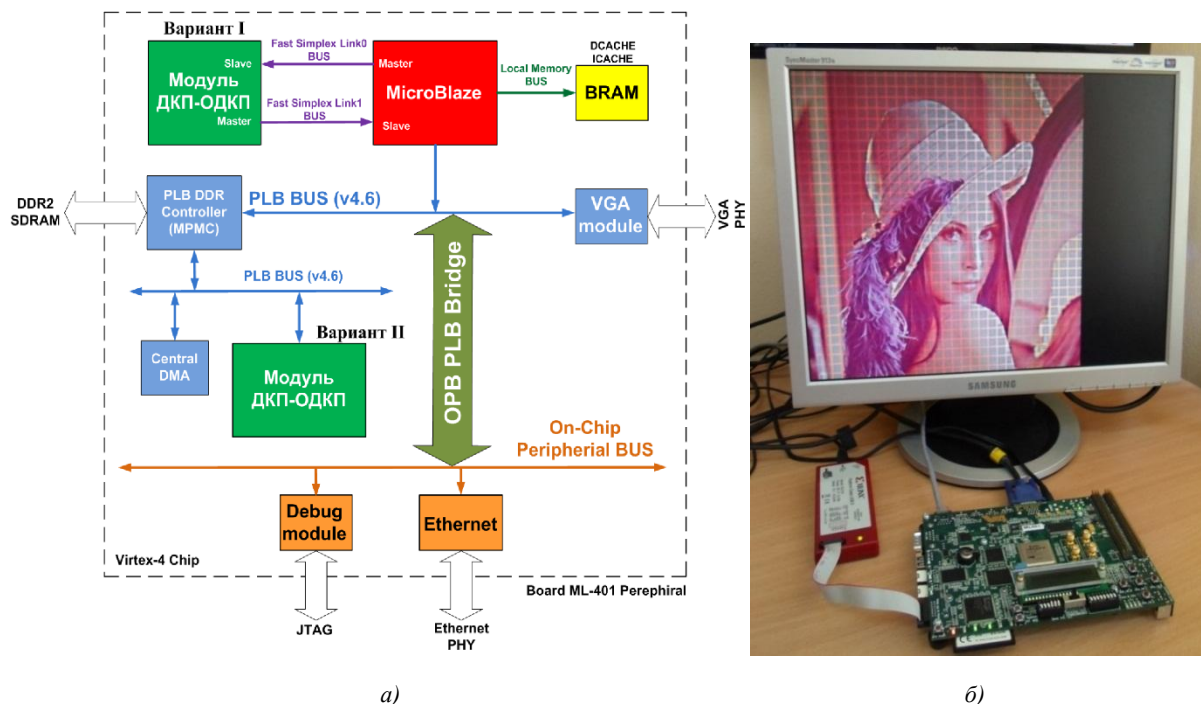


Рис. 14. FPSoC-компрессии изображений на ПЛИС XC4 VLX25-FF668-10 (а) и инструментальный стенд на Xilinx ML-401 (б)

В отладочный модуль Xilinx ML-401 загружается проект FPSoC, подготовленный в виде bit-файла с помощью JTAG-интерфейса. После загрузки bit-файла в ПЛИС стартует софт-процессор под управлением программы, записанной в блочную память. В данном случае там записан стандартный загрузчик, ожидающий управляющую программу через JTAG-интерфейс. Такой подход исключает необходимость разработки собственного загрузчика из флеш-памяти или другого источника, что ускоряет процесс отладки прототипа системы. Управляющая программа, разработанная на языке C, выполняет инициализацию периферийных устройств и ввод исходных данных по сети. Для работы с сетью не требуется полноценный стек TCP/IP (Transmission Control Protocol/Internet Protocol), достаточно реализации транспортного UDP/IP (User Datagram Protocol/Internet Protocol). При этом со стороны персонального компьютера требуется лишь соответствующая запись в таблице ARP (Address Resolution Protocol) с адресом MAC (Media Access Control) платы Xilinx ML-401. Со стороны отладочного модуля Ethernet-фрейм формируется полностью программным образом и для корректной доставки данных потребуется лишь правильный формат и подсчет контрольной суммы UDP-пакета. Таким образом, данное сетевое соединение, несмотря на свои ограничения как по функциональности, так и по скорости, позволяет работать с пользовательскими программами на персональном компьютере и требует меньше усилий по сравнению с полноценной сетевой системой для софт-процессора MicroBlaze. Результаты кодирования изображения передаются по сети и выводятся в буфер динамической памяти, из которого через контроллер TFT передаются в VGA-интерфейс для отображения на экране монитора (рис. 14, б).

Заключение

С помощью функционально-архитектурного проектирования в статье по шагам описана реализация системы быстрого прототипирования процессоров дискретного косинусного преобразования для трансформационного кодирования изображений по схеме L2L на ПЛИС с архитектурой FPGA. Предложен стенд для исследования и быстрого создания прототипа подобных

систем, где разрабатываемые архитектуры процессоров ДКП подключаются в виде модулей периферийных устройств или сопроцессоров и результат преобразования выводится на экран через VGA-интерфейс. Данная технология обеспечивает гибкость в поиске компромиссных решений при реализации программируемых систем на кристалле FPSoC реального времени по таким параметрам, как аппаратная сложность, производительность и точность вычислений. Проектировщик может исследовать ряд архитектурных решений, например, преобразователя трансформационного кодирования изображений по схеме L2L и выбрать более подходящее требованиям спецификации. Предложена архитектура универсального рекурсивного процессора ДКП-ОДКП FPSoC реального времени, работающего в режимах сжатия изображений с потерями и без потерь, которая может рассматриваться как базовая при построении кодеров класса L2L. Данная методология быстрого прототипирования встраиваемых программируемых систем на ПЛИС для мультимедийных приложений реализована как высокоуровневая инструментальная система, которая позволяет получать IP-компоненты пользователя.

Список литературы

1. Spanias, A. Audio Signal Processing and Coding / A. Spanias, T. Painter, V. Atti. – Hoboken : Wiley, 2007. – 486 p.
2. Гонсалес, Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – М. : Техносфера, 2005. – 1072 с.
3. A real time image processing subsystem: GEZGIN / N. Ismailoglu [et al.] // Proc. of the 16th Annual AIAA/USU Conf. on Small Satellites. – Logan, 2002. – Vol. SSC02-VIII-3. – P. 10.
4. Using FPGA and JAVA in rapid prototyping of a real-time H.264/AVC decoder / A. Petrovsky [et al.] // Zeszyty Naukowe Telekomunikacja i Elektrotechnika. – 2009. – Vol. 2009, № 12. – P. 43–55.
5. Suzuki, T. Realization of lossless-to-lossy image coding compatible with JPEG standard by direct-lifting of DCT-IDCT / T. Suzuki, M. Ikehara // 17th IEEE Intern. Conf. on Image Processing (ICIP). – Hong Kong, 2010. – P. 389–392.
6. Vaidyanathan, P.P. The role of lossless systems in modern digital signal processing: A tutorial / P.P. Vaidyanathan, Z. Doğanata // IEEE Trans. Educ. – 1989. – Vol. 32, № 3. – P. 181–197.
7. Зотов, В.Ю. Проектирование встраиваемых микропроцессорных систем на основе ПЛИС фирмы Xilinx / В.Ю. Зотов. – М. : Горячая линия – Телеком, 2006. – 520 с.
8. MicroBlaze Processor Reference Guide Embedded Development Kit EDK 14.5 UG081 (v14.5) [Электронный ресурс]. – 2015. – Режим доступа : http://www.xilinx.com/support/documentation/sw_manuals/mb_ref_guide.pdf. – Дата доступа : 20.06.2015.
9. Немудров, В. Система-на-кристалле. Проектирование и развитие / В. Немудров, Г. Мартин. – М. : Техносфера, 2004. – 216 с.
10. Иванюк, А.А. Проектирование встраиваемых цифровых устройств и систем / А.А. Иванюк. – Минск : Бестпринт, 2012. – 337 с.
11. Петровский, А.А. Быстрое проектирование систем мультимедиа от прототипа / А.А. Петровский, А.В. Станкевич. – Минск : Бестпринт, 2011. – 412 с.
12. FPGA-based Implementation of Signal Processing Systems / R. Woods [et al.]. – Chichester : Wiley, 2008. – 364 p.
13. Rao, K. Techniques and Standards for Image, Video, and Audio Coding / K. Rao, J. Hwang. – Upper Saddle Rive : Prentice Hall, 1996. – 563 p.
14. Suzuki, T. Integer DCT based on direct-lifting of DCT-IDCT for lossless-to-lossy image coding / T. Suzuki, M. Ikehara // IEEE Trans. Image Process. – 2010. – Vol. 19, № 11. – P. 2958–2965.
15. A high-throughput and area-efficient video transform core with a time division strategy / Y.-H. Chen [et al.] // IEEE Trans. VLSI Syst. – 2013. – Vol. PP, № 99. – P. 10.
16. Ключеня, В.В. Модуль вычисления прямого и обратного ДКП для систем сжатия данных / В.В. Ключеня, А.А. Петровский // Труды 16-й Междунар. конф. «Цифровая обработка сигналов и ее применение» (DSPA'2014). – М., 2014. – Т. 2. – С. 642–645.

17. Миано, Д. Форматы и алгоритмы сжатия изображений в действии / Д. Миано. – М. : Триумф, 2003. – 336 с.
18. Loeffler, C. Practical fast 1-D DCT algorithms with 11 multiplications / C. Loeffler, A. Lightenberg, G. Moschytz // Proc. IEEE Intern. Conf. Acoust., Speech, Signal Process. (ICASSP). – Glasgow, 1989. – Vol. 2. – P. 988–991.
19. Kliuchenia, V. FPGA architecture of DCT processor based lifting steps for real-time embedded applications / V. Kliuchenia // Proc. 11th Intern. Conf. Pattern recognition and information processing (PRIP'2011). – Minsk, 2011. – P. 259–262.
20. White, S.A. Applications of distributed arithmetic to digital signal processing: A tutorial review / S.A. White // IEEE ASSP Mag. – 1989. – Vol. 6, № 3. – P. 4–19.
21. Fox, T.W. Rapid prototyping of field programmable gate array-based discrete cosine transform approximations / T.W. Fox, L.E. Turner // EURASIP Journal on Advances in Signal Processing. – 2003. – Vol. 2003, № 6. – P. 543–554.
22. Ключеня, В.В. Выбор оптимальной реализации структурного решения процессора ДКП на распределенной арифметике / В.В. Ключеня, А.А. Петровский // Доклады БГУИР. – 2010. – № 7. – С. 66–75.
23. Бибило, П.Н. Основы языка VHDL : учеб. пособие / П.Н. Бибило. – 6-е изд. – М. : ЛИБРОКОМ, 2014. – 328 с.
24. Rosinger, H.-P. Connecting customized IP to the microblaze soft processor using the fast simplex link (FSL) channel : Tech. Rep. XAPP 529 / H.-P. Rosinger // [Electronic resource]. – 2004. – Mode of access : http://www.xilinx.com/support/documentation/application_notes/xapp529.pdf. – Date of access : 29.06.2015.
25. Петровский, Н.А. Процессор обработки изображения на многополосном вейвлет-преобразовании в алгебре кватернионов / Н.А. Петровский // Доклады БГУИР. – 2011. – Т. 61, № 7. – С. 76–82.

Поступила 05.06.2015

*Белорусский государственный университет
информатики и радиоэлектроники,
Минск, П. Бровки, 6
e-mail: vitaly.kliuchenia@gmail.com,
nick@petrovsky.eu*

V.V. Kliuchenia, N.A. Petrovsky

RAPID PROTOTYPING OF EMBEDDED PROGRAMMABLE SYSTEMS ON FPGA FOR MULTIMEDIA APPLICATIONS

This paper illustrates rapid prototyping method applied for mobile multimedia systems with L2L (lossless-to-lossy) compression scheme. Debug module Xilinx ML-401 based on the FPGA (Filed Programmable Gate Array) is taken as the basis. Microblaze soft-processor is used as the main control unit. Proposed DCT-IDCT cores are developed using VHDL hardware description language as FSL (Fast Simplex Link) coprocessors for Microblaze or standalone PLB (Processor Logical Bus) acceleration peripheral. Result of image processing is displayed on the screen connected via VGA interface. Proposed recursive architecture of DCT-IDCT core can be used as a basis of L2L coder.

УДК 658.512

Ю.Ю. Ланкевич

АЛГОРИТМЫ СБОРКИ ОБЩЕГО ИЗОБРАЖЕНИЯ ТОПОЛОГИИ СБИС

Рассматривается задача сборки общего изображения слоя топологии сверхбольших интегральных микросхем из кадров, полученных путем фотографирования микроскопом технологического слоя микросхемы. Предлагается использование графических ускорителей и возможностей технологии CUDA для обработки большого объема информации. Программно реализуются алгоритмы сборки общего видеоизображения.

Введение

В настоящее время существует необходимость в анализе изображений топологии полупроводниковых сверхбольших интегральных микросхем (СБИС), спроектированных по субмикронным нормам.

Одной из задач, которые необходимо решить, чтобы выполнить анализ микросхемы, является сборка общего изображения топологии СБИС из отдельных кадров. Множество кадров для анализа, которые соответствуют фрагментам слоев микросхемы, получают путем фотографирования кристалла слой за слоем с помощью электронного микроскопа.

Для выполнения анализа видеоизображений топологии СБИС требуются программные средства поддержки указанного процесса. Качество анализа во многом будет определяться набором используемых инструментов и их возможностями [1]. Так как полученные с помощью микроскопа изображения представляют собой небольшие фрагменты топологии схемы, то необходимы программные средства построения общего изображения слоя топологии СБИС. Такие программные средства включают в себя программу поиска оптимального совмещения соседних кадров (по горизонтали и вертикали) и программу объединения кадров в общее изображение слоя топологии СБИС, реализующую алгоритм оценки качества полученного изображения и алгоритм исправления искажений, связанных с ошибкой совмещения.

Обработкой изображений в настоящее время занимается достаточно много специалистов в различных областях: в области картографии поверхности земли и морского дна [2], в медицине [3] и других, в том числе в области обработки изображений СБИС [4, 5].

Обработка изображений СБИС включает в себя получение общего видеоизображения топологии СБИС, выделение контуров объектов [6-10], локализацию и классификацию объектов [7, 11-14].

В настоящей работе рассматривается задача получения общего видеоизображения топологии СБИС, основное отличие предлагаемого подхода заключается в использовании параллельных алгоритмов и их реализации на графических ускорителях, что сильно уменьшает время обработки большого числа изображений.

1. Постановка задачи

Множество кадров изображения топологии СБИС необходимо объединить в одно общее изображение с минимальной ошибкой совмещения. Дополнительным требованием является разработка параллельных алгоритмов решения задачи с использованием технологии параллельных вычислений CUDA (Compute Unified Device Architecture).

Решение задачи предлагается осуществлять в три этапа:

- 1) построение матриц ошибок совмещений;
- 2) анализ матриц ошибок совмещений;
- 3) коррекция общего видеоизображения по результатам анализа и вычисление относительных координат для каждого кадра.

В отличие от работы [4] на этапе 1 предлагается полный перебор возможных совмещений соседних кадров при поиске их наилучшего совмещения, что позволяет получать более качественное решение задачи в целом.

2. Основы создания программ на CUDA

Рассмотрим базовые архитектурные особенности видеочипов фирмы NVIDIA с технологией CUDA [15, 16], на которых реализуются разрабатываемые алгоритмы работы с изображениями. Графический процессор (Graphics Processing Unit, GPU) включает несколько кластеров текстурных блоков (Texture Processing Cluster). Каждый кластер состоит из укрупненного блока текстурных выборок и двух-трех потоковых мультипроцессоров, каждый из которых представляет собой восемь вычислительных устройств и два суперфункциональных блока. Все инструкции выполняются по принципу SIMD (Single Instruction, Multiple Data – одна инструкция, много данных), когда одна инструкция применяется к большому количеству данных. Этот способ выполнения назвали SIMT (Single Instruction, Multiple Threads – одна инструкция, много потоков).

Каждый из мультипроцессоров имеет определенные ресурсы. Так, есть специальная разделяемая память объемом 16 Кбайт на мультипроцессор, но это не кеш, так как программист может использовать ее для любых нужд. Разделяемая память позволяет обмениваться информацией между потоками одного блока. Важно, что все потоки одного блока всегда выполняются одним и тем же мультипроцессором. Потоки из разных блоков обмениваться данными не могут, и необходимо помнить об этом ограничении. Мультипроцессоры могут обращаться и к видеопамяти, но с большими задержками и худшей пропускной способностью. Для ускорения доступа и снижения частоты обращения к видеопамяти у мультипроцессоров есть по 8 Кбайт кеша на константы и текстурные данные.

Мультипроцессор использует 8192–16 384 (для ускорителей G8x/G9x и GT2xx соответственно) регистров, общих для всех потоков всех блоков, выполняемых на нем. Максимальное число блоков на один мультипроцессор для G8x/G9x равно восьми, а число групп по 32 потока – 24 (768 потоков на один мультипроцессор). Всего видеокарты серий GeForce 8 и 9 могут обрабатывать до 12 288 потоков одновременно. GeForce GTX 280 на основе GT200 предлагает до 1024 потоков на мультипроцессор, в нем есть 10 кластеров по три мультипроцессора, обрабатывающих до 30 720 потоков. Знание этих ограничений позволяет оптимизировать алгоритмы под доступные ресурсы.

Первым шагом при переносе существующего приложения на CUDA является его профилирование и определение участков кода, являющихся узким местом, тормозящим работу. Если среди таких участков есть подходящие для быстрого параллельного исполнения, эти функции переносятся на C-расширения CUDA для выполнения на GPU. Программа компилируется при помощи поставляемого фирмой NVIDIA компилятора, который генерирует код и для CPU, и для GPU. При исполнении программы центральный процессор выполняет свои порции кода, а GPU выполняет CUDA-код с наиболее тяжелыми параллельными вычислениями. Эта часть программы, предназначенная для выполнения на GPU, называется ядром (kernel). В ядре определяются операции, которые будут исполнены над данными.

Видеочип получает ядро и создает копии для каждого элемента данных. Эти копии называются потоками (thread). Поток содержит счетчик, регистры и состояние. Для больших объемов данных, таких как обработка изображений, запускаются миллионы потоков. Потоки выполняются группами по 32 штуки, называемыми варпами. Варпам назначается исполнение на определенных потоковых мультипроцессорах. Каждый мультипроцессор состоит из восьми ядер – потоковых процессоров, которые выполняют одну инструкцию за один такт.

Мультипроцессор не является традиционным многоядерным процессором, он отлично приспособлен для многопоточности, поддерживая до 32 варпов одновременно. Каждый такт аппаратное обеспечение выбирает, какой из варпов исполнять, и переключается от одного к другому без потерь в тактах.

GPU (Graphics Processing Unit – графический процессор) является вычислительным устройством, сопроцессором (device) для центрального процессора (host), обладающим собствен-

ной памятью и обрабатывающим параллельно большое количество потоков. Ядром называется функция для GPU, исполняемая потоками (аналогия из 3D-графики – шейдер). Модель программирования в CUDA предполагает группирование потоков. Потоки объединяются в блоки потоков (thread block) – одномерные или двухмерные сетки потоков, взаимодействующих между собой при помощи разделяемой памяти и точек синхронизации. Программа (ядро, kernel) исполняется над сеткой (grid) блоков потоков (рис. 1). Одновременно исполняется одна сетка. Каждый блок может быть одно-, двух- или трехмерным по форме и может состоять из 512 потоков для существующего в настоящее время аппаратного обеспечения.

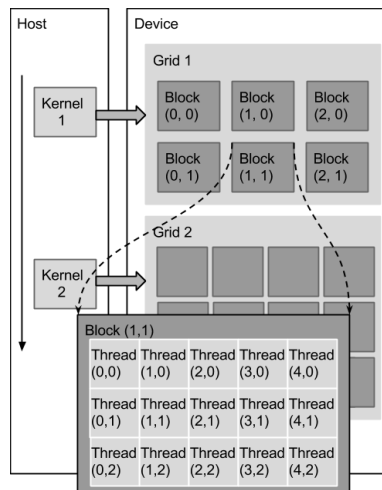


Рис. 1. Структура потоков в CUDA

Блоки потоков выполняются в виде варпов. В связи с тем что не всегда удобно работать с 32 потоками, CUDA позволяет работать и с блоками, содержащими от 64 до 512 потоков.

Группировка блоков в сетки позволяет уйти от ограничений и применить ядро к большему числу потоков за один вызов. Это помогает и при масштабировании. Если у GPU недостаточно ресурсов, он будет выполнять блоки последовательно. В противном случае блоки могут выполняться параллельно, что важно для оптимального распределения работы на видеочипах разного уровня, начиная от мобильных и интегрированных.

Модель памяти в CUDA имеет возможность побайтной адресации. Доступно довольно большое количество регистров на каждый потоковый процессор – до 1024 штук. Доступ к ним очень быстрый, хранить в них можно 32-битные целые числа или числа с плавающей точкой. Каждый поток имеет доступ к различным типам памяти (рис. 2).

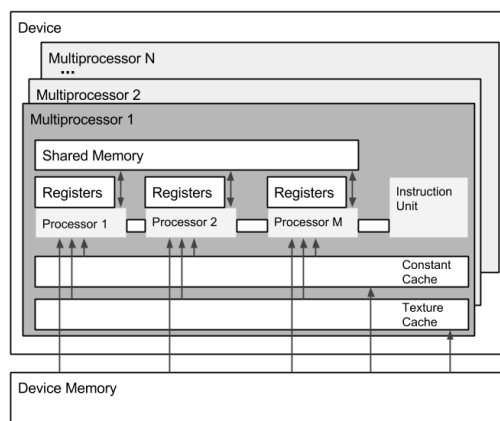


Рис. 2. Структура памяти в CUDA

Глобальная память (Global memory) – самый большой объем памяти, доступный для всех мультимикропроцессоров на GPU. Ее размер составляет от 256 Мбайт до 8 Гбайт. Обладает высокой

пропускной способностью (более 100 Гбайт/с для современных решений NVIDIA) и имеет очень большие задержки в несколько сот тактов.

Локальная память (Local memory) – небольшой объем памяти, к которому имеет доступ только один потоковый процессор. Она относительно медленная – такая же, как и глобальная.

Разделяемая память (Shared memory) – 16-килобайтный блок памяти с общим доступом для всех потоковых процессоров в мультипроцессоре. Эта память имеет такую же, как и регистры, скорость доступа. Она обеспечивает взаимодействие потоков, управляется разработчиком напрямую и имеет низкие задержки. Преимущества разделяемой памяти: использование в виде управляемого программистом кеша первого уровня, снижение задержек при доступе исполнительных блоков к данным, сокращение количества обращений к глобальной памяти.

Память констант (Constant memory) – область памяти объемом 64 Кбайт, доступная только для чтения всеми мультипроцессорами. Она кешируется по 8 Кбайт на каждый мультипроцессор. Имеет задержку в несколько сот тактов при отсутствии нужных данных в кеше.

Текстурная память (Texture memory) – блок памяти, доступный для чтения всеми мультипроцессорами. Выборка данных осуществляется при помощи текстурных блоков видеочипа, поэтому предоставляются возможности линейной интерполяции данных без дополнительных затрат. Кешируется по 8 Кбайт на каждый мультипроцессор. Медленная, как и глобальная: имеет сотни тактов задержки при отсутствии данных в кеше.

Естественно, что глобальная, локальная, текстурная и память констант – это физически одна и та же память, известная как локальная видеопамять видеокарты. Их отличия состоят в различных алгоритмах кеширования и моделях доступа. Центральный процессор может обновлять и запрашивать только внешнюю память: глобальную, константную и текстурную (рис. 3).

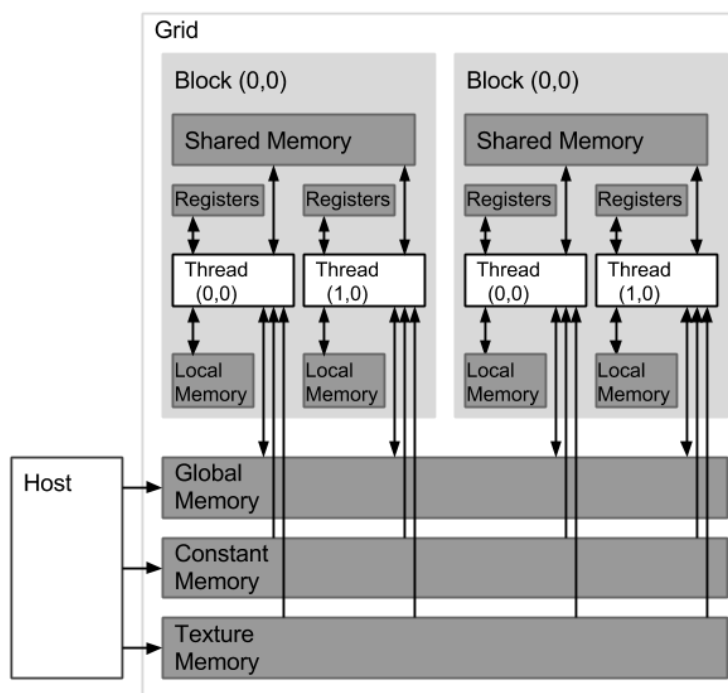


Рис. 3. Структура потоков в CUDA

При работе с графическими ускорителями нужно помнить о разных типах памяти, в частности о том, что локальная память и глобальная не кешируются и задержки при доступе к ним гораздо выше, чем у регистровой памяти, так как она физически находится в отдельных микросхемах. Таким образом, технология CUDA предполагает специальный подход к разработке – не совсем такой, как принят в программах для CPU.

3. Алгоритм построения матриц ошибок совмещений соседних кадров (этап 1)

Прежде чем сформулировать задачу, решаемую на этапе 1, определим некоторые понятия: перекрытие – общая область соседних кадров, совмещение – позиция одного кадра относительно другого.

Входными данными для алгоритма построения матриц ошибок совмещений соседних кадров являются:

- множество кадров (все кадры имеют одинаковые размеры) изображения топологии СБИС в формате *.dds или *.bmp;
- упорядоченный список имен файлов кадров, начиная с первого кадра первой строки и заканчивая последним кадром последней строки;
- количество кадров в строке и в столбце;
- среднее перекрытие для правого и для нижнего совмещений;
- максимальное отклонение от среднего перекрытия.

Алгоритм основывается на том, что оценить ошибку каждого совмещения для перекрытия двух кадров можно, вычислив сумму разниц значений соответствующих пикселей двух кадров по компонентам цвета. Кадры представляются в рамках цветовой модели RGB (Red, Green, Blue), т. е. каждый пиксел имеет три компонента цвета. Каждый элемент результирующей матрицы ошибок представляет собой суммарную ошибку для одного совмещения двух соседних кадров.

В качестве примера выберем один из возможных форматов обрабатываемых изображений BMP (Bitmap Picture). В формате BMP левый нижний угол изображения находится в начале массива данных (рис. 4). Далее все выражения выводятся для этого случая.

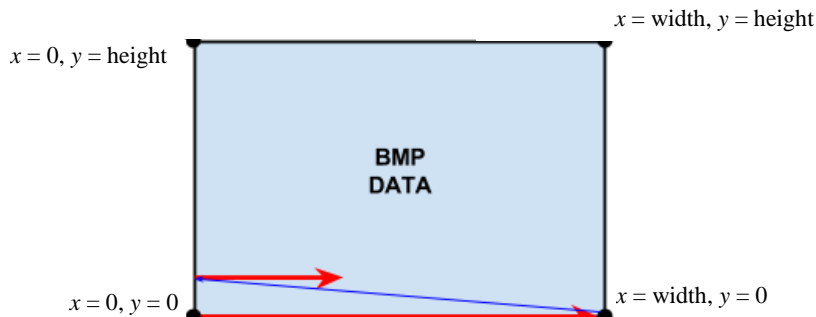


Рис. 4. Координаты пикселей в BMP-изображении

Рассмотрим пример горизонтального совмещения соседних кадров. На рис. 5 $xmargin$ представляет собой наибольшее перекрытие кадров по горизонтали. Оно равно сумме среднего перекрытия по горизонтали и максимально допустимого отклонения перекрытия от среднего ($offset$).

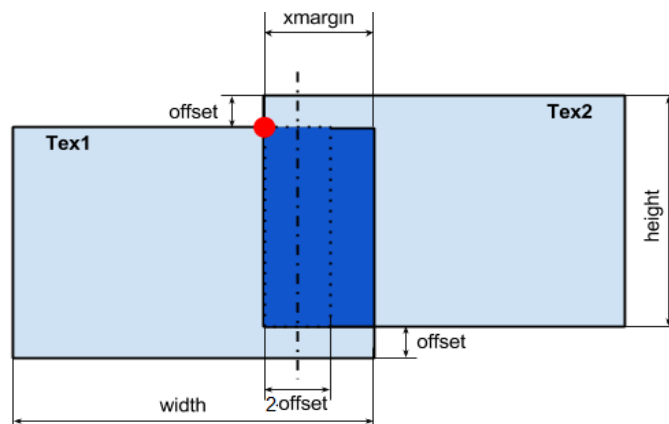


Рис. 5. Горизонтальное совмещение соседних кадров

Далее будем использовать следующую сокращенную запись условных выражений:

$$Y = A ? B : C,$$

интерпретируемых так: если условие A выполняется, то $Y = B$, если же условие A не выполняется, то $Y = C$.

При горизонтальном сравнении координаты левого верхнего угла для левого кадра Tex1 при заданном смещении Δx и Δy вычисляются по формулам

$$\begin{aligned} X_{\text{Tex1Pixel}} &= \text{width} - 1 - \text{xmargin} + \Delta x; \\ Y_{\text{Tex1Pixel}} &= (\Delta y < \text{offset}) ? \text{height} - 1 : (\text{height} - 1) - (\Delta y - \text{offset}), \end{aligned}$$

где Δx и Δy – смещение перекрытия по горизонтали и вертикали, которое изменяется в дискретном диапазоне от 0 до $2 * \text{offset} - 1$, единица дискретизации равна одному пикселу; width – ширина изображения кадра; height – высота изображения кадра; offset – максимально допустимое отклонение перекрытия от среднего.

Координаты левого верхнего угла для правого кадра Tex2 при заданном смещении Δx и Δy вычисляются по формулам

$$\begin{aligned} X_{\text{Tex2Pixel}} &= 0; \\ Y_{\text{Tex2Pixel}} &= (\Delta y < \text{offset}) ? (\text{height} - 1) - (\text{offset} - \Delta y) : \text{height} - 1. \end{aligned}$$

Ширина xwidth и высота yheight окна перекрытия кадров для заданных $(\Delta x, \Delta y)$ определяются выражениями

$$\begin{aligned} \text{xwidth} &= \text{xmargin} - \Delta x; \\ \text{yheight} &= (\Delta y < \text{offset}) ? \text{height} - \text{offset} + \Delta y : \text{height} - (\Delta y - \text{offset}). \end{aligned}$$

Ошибка совмещения кадров для заданного смещения $(\Delta x, \Delta y)$

$$\text{Err}(\Delta x, \Delta y) = \frac{\sum_{i=0; j=0}^{i=\text{xwidth}-1; j=\text{yheight}-1} (R_{\text{Tex1Pixel}} - R_{\text{Tex2Pixel}}) + (G_{\text{Tex1Pixel}} - G_{\text{Tex2Pixel}}) + (B_{\text{Tex1Pixel}} - B_{\text{Tex2Pixel}})}{\text{yheight} \cdot \text{xwidth}},$$

где $R_{\text{TexXPixel}}$, $G_{\text{TexXPixel}}$, $B_{\text{TexXPixel}}$ ($X = 1, 2$) – три компонента цвета пиксела в левом и правом изображении кадров для заданных смещений i, j в окне.

Задача совмещения, решаемая на этапе 1: необходимо рассчитать значения ошибок совмещения для всех возможных перекрытий $(\Delta x, \Delta y)$ соседних кадров, т. е. всего $4 * \text{offset}^2$ (Δx и Δy могут принимать значения от 0 до $2 * \text{offset}$), после чего определить смещение $(\Delta x, \Delta y)$, для которого $\text{Err}_{(\Delta x, \Delta y)}$ будет минимальным.

Для вертикального совмещения (рис. 6) координаты левого верхнего угла для верхнего изображения Tex1 при заданном смещении Δx и Δy вычисляются по формулам

$$\begin{aligned} X_{\text{Tex1Pixel}} &= (\Delta x < \text{offset}) ? 0 : \Delta x - \text{offset}; \\ Y_{\text{Tex1Pixel}} &= \text{ymargin} - \Delta y, \end{aligned}$$

где ymargin – наибольшее перекрытие кадров по вертикали.

Координаты левого верхнего угла для нижнего кадра Tex2 при заданном смещении Δx и Δy вычисляются по формулам

$$\begin{aligned} X_{\text{Tex2Pixel}} &= (\Delta x < \text{offset}) ? \text{offset} - \Delta x : 0; \\ Y_{\text{Tex2Pixel}} &= \text{height} - 1. \end{aligned}$$

Ширина и высота окна перекрытия кадров определяются выражениями

$$xwidth = (\Delta x < offset) ? width - offset + \Delta x : width - (\Delta x - offset);$$

$$yheight = ymargin - \Delta y.$$

Расчет матрицы ошибок Егг для вертикального совмещения выполняется с учетом приведенных выше выражений.

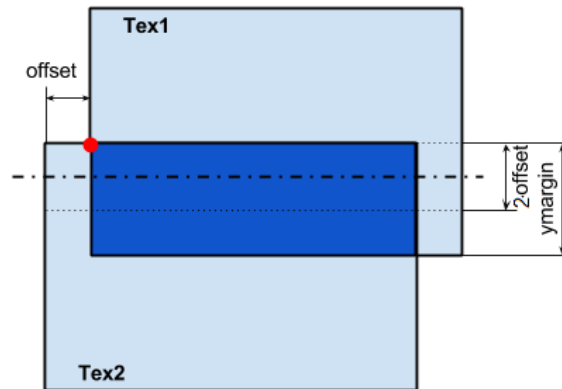


Рис. 6. Вертикальное совмещение соседних кадров

Отдельный поток выполнения программы называется *thread*. Потоки (не более 1024 штук) объединяются в блоки. Идентифицировать поток в блоке можно по трем координатам *threadIdx.x*, *threadIdx.y*, *threadIdx.z*, максимальные значения (*blockDim.x*, *blockDim.y*, *blockDim.z*) которых задаются пользователем перед запуском задачи на GPU. Блоки объединяются в сеть, в которой блок можно определить также по трем координатам *blockIdx.x*, *blockIdx.y*, *blockIdx.z*, максимальные значения которых задаются параметрами *gridDim.x*, *gridDim.y*, *gridDim.z*. На рис. 7 показано разделение блоков на потоки, используемое для реализации алгоритма нахождения лучшего совмещения.

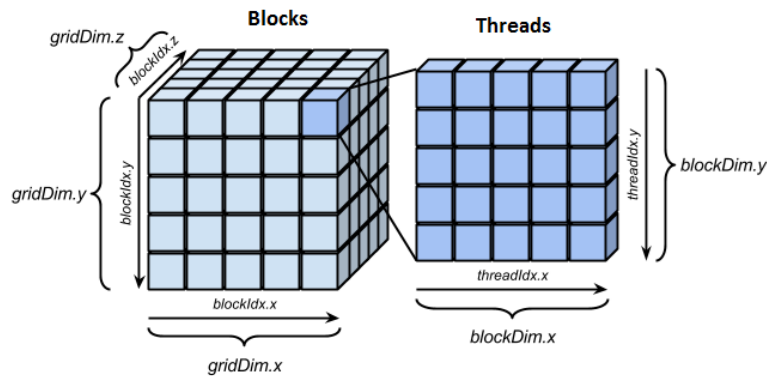


Рис. 7. Разделение параллельных потоков на блоки и потоки

По координатам *blockIdx.x*, *blockIdx.y* задаются различные смещения (Δx , Δy) по горизонтали и вертикали перекрытия соседних кадров, в этом случае $blockDim.x = 2 \times offset$, $blockDim.y = 2 \times offset$. Окно перекрытия соседних кадров (рис. 8) делится на квадратные области по 16×16 пикселей, которые представляют один блок по координате *blockIdx.z*. На рис. 8 показан пример для случая, когда $n_x = 5$, $n_y = 10$. Таким образом, для расчета координат пиксела в окне перекрытия кадров используются выражения

$$divx = (blockIdx.z \% n_x) * blockDim.x + threadIdx.x;$$

$$divy = (blockIdx.z / n_x) * blockDim.y + threadIdx.y.$$

Каждый поток рассчитывает значение ошибки (разницы между цветами пикселей) для одного отдельного пикселя, сохраняя это значение в разделяемой памяти, общей для одного блока. После того как все потоки в блоке рассчитают значения ошибки, она суммируется с помощью стандартного параллельного алгоритма суммирования путем редукции [15, с. 75]. Синхронизация потоков в блоке осуществляется с помощью функции `__syncthreads()`. Значение суммарной ошибки по блоку 16×16 пикселей записывается в глобальную память GPU. Результирующий массив ошибок снова суммируется с помощью редукции, и результат сохраняется в файл. Формат файла матриц ошибок представляет собой двухмерный массив, содержащий $4 * \text{offset}^2$ элементов типа `float`.

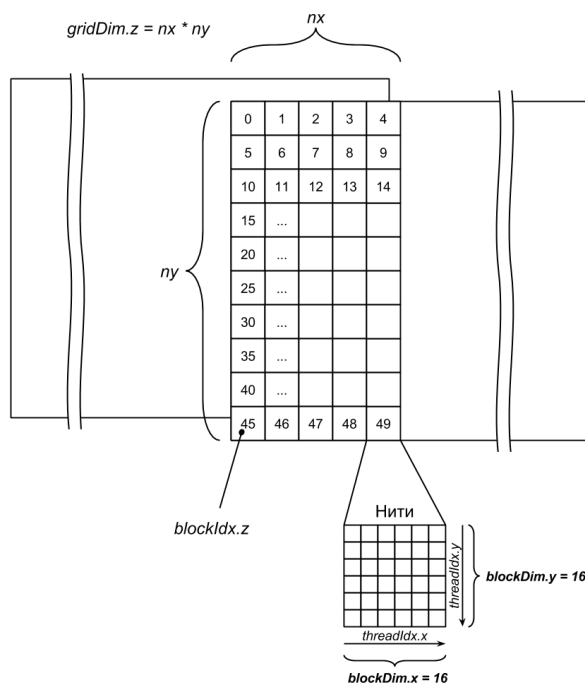


Рис. 8. Разбиение вычислений на блоки и потоки при горизонтальном совмещении соседних кадров

4. Алгоритм анализа матриц ошибок совмещений (этап 2)

Задача анализа матриц ошибок (задача этапа 2) состоит в том, чтобы по матрицам ошибок получить матрицы относительных координат соседних кадров и матрицы приоритетов, которые соответствуют полученным относительным координатам.

Алгоритм анализа матриц ошибок совмещений обрабатывает исходную информацию, заданную в виде матриц ошибок совмещений соседних кадров. Результатом выполнения алгоритма решения задачи анализа матриц ошибок совмещений являются матрицы относительных координат соседних кадров и матрицы приоритетов, которые соответствуют полученным относительным координатам. Понятие приоритета будет рассмотрено далее.

Алгоритм анализа ошибок совмещений основывается на следующих утверждениях:

1. Каждый кадр (если он не крайний справа или снизу) имеет соседний кадр справа и снизу.
2. Для каждого соседнего кадра существует матрица совмещений M , каждый элемент которой соответствует позиции совмещения этого кадра с соседним, а значение соответствует ошибке совмещения (сумме разниц значений соответствующих пикселей совмещаемых кадров).
3. Среднее перекрытие для всех соседних кадров зависит от шага движения микроскопа, величина шага является примерно одинаковой в пределах одного слоя.
4. Каждая матрица совмещений M нормализуется так, что все ее элементы принимают значения от 0 до 1.
5. Для всех матриц M существует отклонение от минимального значения, в пределах которого находится искомое значение относительных координат взаимосовмещения соседних кадров, которым соответствует эта матрица.

б. Чем больше разброс значений, которые находятся в пределах заданного отклонения от минимального значения, тем менее достоверной является эта матрица и тем больше вероятность того, что относительная позиция соседнего кадра, соответствующая элементу из M с минимальным значением, не будет соответствовать искомой.

На рис. 9 в графическом виде представлены некоторые виды матриц совмещений в форме изображений в градациях серого. Черная точка соответствует нулю, а белая – единице. На рис. 9, а показаны наименее достоверные матрицы совмещений: на значения, полученные с помощью этих матриц, полагаться нельзя, на рис. 9, б – наиболее достоверные матрицы.

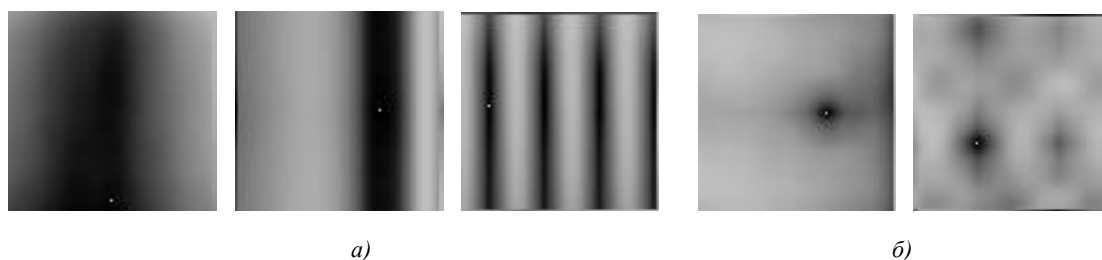


Рис. 9. Виды матриц совмещений: а) менее достоверные матрицы; б) более достоверные матрицы

Минимум находится в наиболее темной части каждого изображения, соответствующего матрице совмещений M . Поэтому чем больше эта область, тем менее достоверной будет относительная координата соседнего кадра, которая соответствует найденному минимуму. Если таких областей в одной матрице несколько и они разбросаны по всему изображению, найденная относительная координата соседнего кадра будет недостоверной.

Введем понятие *веса относительной координаты* – чем меньше вес, тем больше вероятность того, что значение относительной координаты является верным. Значение веса относительной координаты в дальнейшем будет выступать в роли приоритета в порядке пересчета полученных на первом этапе значений относительных координат всех кадров изображения.

Значение веса, или приоритета, вычисляется следующим образом:

1. Выбирается пороговое значение, ниже которого все значения из матрицы ошибок считаются равнозначными.

2. Находятся наименьшая и наибольшая позиции (по горизонтали и по вертикали отдельно) элементов матрицы ошибок из множества элементов, значения которых меньше порогового значения.

3. Весу присваивается разница наименьшей и наибольшей позиций.

Потом вычисляются значение относительной координаты, которое соответствует минимуму в заданных границах матрицы ошибок, и среднее значение относительных координат с наименьшим значением веса относительных координат для ряда кадров и производится пересчет относительных координат в соответствии с правилом трех сигм [17].

5. Алгоритм коррекции общего видеоизображения по результатам анализа (этап 3)

Основная задача коррекции общего видеоизображения (задача этапа 3): по результатам анализа требуется пересчитать значения относительных координат всех совмещенных кадров общего видеоизображения.

Прежде чем описывать алгоритм решения задачи коррекции общего видеоизображения, рассмотрим общий случай (рис. 10) расположения четырех соседних кадров видеоизображений.

На этапе 2 были определены матрицы относительных координат (Xh и Xv – для правого и нижнего кадра соответственно по координате x ; Yh и Yv – для правого и нижнего кадра соответственно по координате y) исходя из информации матриц ошибок совмещений.

Для четырех соседних кадров должны выполняться следующие условия:

$$Xv[i][j] + Xh[i][j] = Xv[i][j + 1] + Xh[i + 1][j]; \quad (1)$$

$$Yv[i][j] + Yh[i][j] = Yv[i][j + 1] + Yh[i + 1][j], \quad (2)$$

где i – позиция кадра в строке; j – позиция кадра в столбце; $Xv[i][j]$ – относительная координата по x для нижнего соседнего кадра; $Yv[i][j]$ – относительная координата по y для нижнего соседнего кадра; $Xh[i][j]$ – относительная координата по x для правого соседнего кадра; $Yh[i][j]$ – относительная координата по y для правого соседнего кадра.

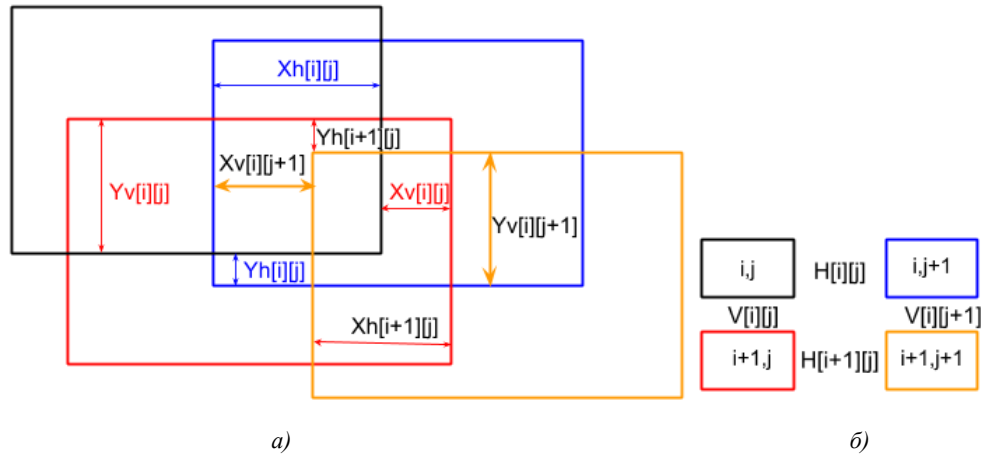


Рис. 10. Относительные координаты для четверки кадров (i – номер строки, j – номер столбца):
а) кадры с перекрытием; б) положение кадров относительно друг друга

На этапе 2 были также определены веса относительных координат ($matrix_pr_d[i][j][0]$ и $matrix_pr_r[i][j][0]$ – для нижнего и правого кадра соответственно по координате x ; $matrix_pr_d[i][j][1]$ и $matrix_pr_r[i][j][1]$ – для нижнего и правого кадра соответственно по координате y) и средние допустимые значения смещений по x ($M_value_x_r[i]$ и $M_value_x_d[i]$ для правых и нижних кадров соответственно) и по y ($M_value_y_r[i]$ и $M_value_y_d[i]$ для правых и нижних кадров соответственно) для каждой строки кадров.

Алгоритм коррекции общего видеоизображения выполняется отдельно по координате x и отдельно по координате y .

Алгоритм для координаты x включает следующие шаги:

1. Определить четверку кадров с наименьшим суммарным весом относительных координат исходя из матриц $matrix_pr_d$ и $matrix_pr_r$, где $i:=0..rows - 1$, $j:=0..cols - 1$. Четверка с минимальным (ненулевым) значением суммы

$$sum = matrix_pr_r[i][j][0] + matrix_pr_d[i][j][0] + matrix_pr_r[i + 1][j][0] + matrix_pr_d[i][j + 1][0]$$

будет соответствовать искомой четверке. Обозначим найденную позицию min_i_x и min_j_x . Если такой четверки не нашлось, то закончить.

2. Вычислить значение для каждой координаты из четырех, участвующих в пересчете, исходя из трех остальных, пользуясь (1), и вычислить разницу E_{gr} между полученным значением и средним по ряду кадров ($M_value_x_r[i]$ или $M_value_x_d[i]$).

3. Сравнить значения E_{gr} , вычисленные в п. 2, со значением $offset/4$ ($offset$ – максимальное отклонение от среднего перекрытия) по модулю и в зависимости от результата сравнения выполнить следующие действия:

– если нет ни одного значения из полученных разностей, которое меньше $offset/4$, то увеличить значения весов всех участвующих в пересчете относительных координат и перейти к п. 1;

– в противном случае для всех относительных координат, которые участвуют в пересчете, в порядке убывания соответствующих им значений $|E_{gr}|$, полученных в п. 2, выполнить следующее: если фиксация текущей координаты не мешает дальнейшему пересчету, то зафиксировать ее, присвоив соответствующему элементу из матриц приоритетов нуль.

4. Перейти к п. 1.

После присвоения какому-либо элементу из матриц приоритетов значения 0 производится пересчет всех возможных относительных координат с помощью выражения (1), исходя из того, что значение каждой относительной координаты, которой соответствует элемент из матрицы приоритетов, имеющий значение 0, понимается как константа. Соответствующему каждой пересчитанной относительной координате элементу матрицы приоритетов присваивается значение 0.

На шаге проверки, помешает ли фиксация координаты дальнейшему пересчету, сложность вызывает только проверка, будет ли иметь решение система линейных уравнений, которая определяется выражением (1) и нулевыми элементами из матриц приоритетов.

На рис. 11 шестигранник соответствует определенному кадру; позиция i (по строке), j (по столбцу) соответствует относительной координате, которая фиксируется в текущем цикле алгоритма. Сплошной линией на рисунке обозначены относительные координаты, которые были зафиксированы ранее (на предыдущих шагах алгоритма); пунктирной линией из точек – относительные координаты, которые пересчитываются в текущем состоянии; пунктирной линией из дефисов – относительная координата, которая фиксируется в текущем состоянии без пересчета.

Количество уравнений в системе можно подсчитать исходя из количества четверок соседних кадров. Количество переменных соответствует количеству пунктирных линий из точек.

В ситуации, представленной на рис. 11, *а*, решается система двух линейных уравнений с двумя переменными. В этом случае решение будет найдено всегда.

В ситуации, которая показана на рис. 11, *б*, решается система четырех линейных уравнений с тремя переменными: в этом случае решение может не существовать. Такую ситуацию необходимо избегать. Поэтому в алгоритме предусмотрена дополнительная проверка, основанная на анализе разрешимости системы линейных уравнений, приводимой ниже. В результате проверки меняется последовательность установки относительных координат.

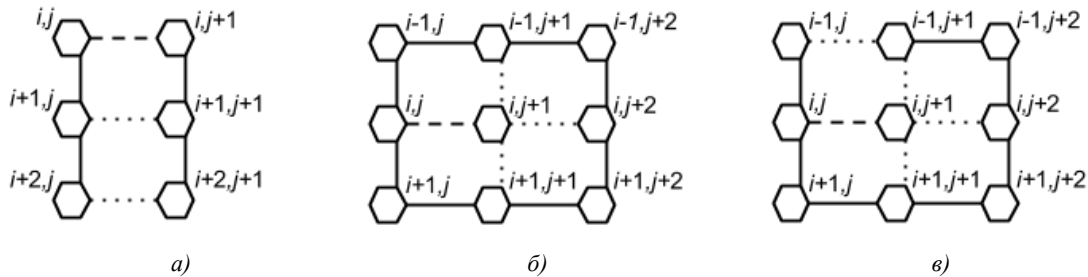


Рис. 11. Графическое представление системы линейных уравнений:
а) два уравнения с двумя переменными; *б*) четыре уравнения с тремя переменными;
в) четыре уравнения с четырьмя переменными

Фактически это означает, что рассматриваемая на текущем шаге алгоритма относительная координата не фиксируется и на последующих шагах вместо данной ситуации будет возникать ситуация, показанная на рис. 11, *в*. В этом случае решается система четырех линейных уравнений, где присутствуют четыре переменные, т. е. система уравнений будет иметь одно решение. Подчеркиванием обозначены те относительные координаты, которые пересчитываются, остальные считаются константами.

Система линейных уравнений на рис. 1, *а* имеет вид

$$Xv[i][j] + Xh[i][j] = Xv[i][j + 1] + \underline{Xh[i + 1][j]};$$

$$Xv[i + 1][j] + \underline{Xh[i + 1][j]} = Xv[i + 1][j + 1] + \underline{Xh[i + 2][j]}.$$

Система линейных уравнений на рис. 11, *б* описывается как

$$Xv[i - 1][j] + Xh[i - 1][j] = \underline{Xv[i - 1][j + 1]} + Xh[i][j];$$

$$Xv[i][j] + Xh[i][j] = \underline{Xv[i][j+1]} + Xh[i+1][j];$$

$$Xv[i-1][j+1] + Xh[i-1][j+1] = Xv[i-1][j+2] + \underline{Xh[i][j+1]};$$

$$\underline{Xv[i][j+1]} + \underline{Xh[i][j+1]} = Xv[i][j+2] + Xh[i+1][j+1].$$

В случае, показанном на рис. 11, в, решается следующая система линейных уравнений:

$$Xv[i-1][j] + \underline{Xh[i-1][j]} = \underline{Xv[i-1][j+1]} + Xh[i][j];$$

$$Xv[i][j] + Xh[i][j] = \underline{Xv[i][j+1]} + Xh[i+1][j];$$

$$\underline{Xv[i-1][j+1]} + Xh[i-1][j+1] = Xv[i-1][j+2] + \underline{Xh[i][j+1]};$$

$$\underline{Xv[i][j+1]} + \underline{Xh[i][j+1]} = Xv[i][j+2] + Xh[i+1][j+1].$$

Алгоритм для координаты y имеет такие же шаги, что и для x , только вместо координаты x рассматривается координата y .

6. Экспериментальные данные

Все упомянутые в предыдущих разделах алгоритмы были программно реализованы. В течение наибольшего времени выполняется программа, реализующая алгоритм построения матриц ошибок, все остальные программы выполняются за сравнительно короткое время.

Приведем результаты выполнения программы построения матриц ошибок для различного числа кадров и различных параметров. В качестве ускорителя использовалась видеокарта GeForce Titan Black.

Пример 1. Количество кадров – 280, среднее перекрытие по x – 84, среднее перекрытие по y – 79, offset – 32.

Время выполнения для кадров в формате bmp – 1 мин 22 с, время выполнения для кадров в формате dds – 58 с.

Пример 2. Количество кадров – 6141, среднее перекрытие по x – 150, среднее перекрытие по y – 150, offset – 40.

Время выполнения для кадров в формате bmp – 66 мин 29 с, время выполнения для кадров в формате dds – 60 мин 45 с.

Заключение

В статье описаны алгоритмы для объединения кадров в общее изображение топологии СБИС. Отличительной особенностью предложенных алгоритмов является то, что они полностью ориентированы на возможности графических ускорителей и технологии CUDA для увеличения быстродействия. Использование параллельных алгоритмов и графических ускорителей позволило ускорить вычисления, что является первоочередным требованием для задач обработки большого объема информации, решение которых на обычных процессорах занимает слишком много времени. Разработанное на основе предложенных алгоритмов программное обеспечение является частью комплекса программ, необходимых для обеспечения технологического цикла производства СБИС как на этапе повторного проектирования, так и на этапе проверки полученных изделий. Экспериментальные данные свидетельствуют о высоком быстродействии разработанных программ.

Список литературы

1. НТЦ «Белмикросистемы» [Электронный ресурс]. – 2011. – Режим доступа : <http://www.integral.by/files/files/bms-2011.pdf>. – Дата доступа : 01.03.2015.

2. Камаев, А.Н. Создание панорамных карт подводного дна на основе больших массивов изображений / А.Н. Камаев. – Владивосток, 2013. – С. 298–301.
3. Недзьведь, А.М. Анализ изображений для решения задач медицинской диагностики / А.М. Недзьведь, С.В. Абламейко. – Минск : ОИПИ НАН Беларуси, 2012. – 240 с.
4. Дудкин, А.А. Обработка изображений в проектировании и производстве интегральных схем / А.А. Дудкин, Р.Х. Садыхов. – Минск : ОИПИ НАН Беларуси, 2008. – 270 с.
5. Абламейко, С.В. Обработка изображений: технология, методы, применение / С.В. Абламейко, Д.М. Лагуновский. – Минск : ОИПИ НАН Беларуси, 1999. – 300 с.
6. Прэтт, У. Цифровая обработка изображений / У. Прэтт. – М. : Мир, 1982. – Кн. 2. – 480 с.
7. Введение в контурный анализ; приложения к обработке изображений и сигналов / Я.А. Фурман [и др.]. – М. : Физматлит, 2003. – 592 с.
8. Yuen, P.C. A contour detection method: Initialization and contour model / P.C. Yuen, G.C. Feng, J.P. Zhou // Pattern Recognition Letters. – 1999. – Vol. 20. – P. 141–148.
9. Koshchan, A. A Comparative Study on Color Edge Detection / A. Koshchan // Pattern Recognition Letters. – 2001. – Vol. 22, № 13. – P. 1419–1429.
10. Rosin, P.L. Edges: saliency measures and automatic thresholding / P.L. Rosin. – Mach. Vision, 1997. – P. 139–159.
11. Методы компьютерной обработки изображений / под. ред. В.А. Сойфера. – М. : Физматлит, 2003. – 784 с.
12. Kerfoot, I.B. Theoretical analysis of multispectral image segmentation criteria / I.B. Kerfoot, Y. Bresler // IEEE Trans. Image Processing. – 1999. – Vol. 8, № 6. – P. 768–820.
13. Coleman, G.B. Image Segmentation by Clustering / G.B. Coleman, H.C. Andrews // Proc. IEEE. – 1979. – Vol. 67. – P. 773–785.
14. Zhang, Z. A survey on evaluation methods for image segmentation / Z. Zhang // Pattern Recognition. – 1996. – Vol. 29(8). – P. 1335–1346.
15. Сандерс, Д. Технология CUDA в примерах: введение в программирование графических процессоров / Д. Сандерс, Э. Кэндрот. – М. : ДМК Пресс, 2011. – 232 с.
16. Боресков, А.В. Основы работы с технологией CUDA / А.В. Боресков, А.А. Харламов. – М. : ДМК Пресс, 2010. – 232 с.
17. Боровиков, В. STATISTICA. Искусство анализа данных на компьютере: для профессионалов / В. Боровиков. – СПб. : Питер, 2003. – 688 с.

Поступила 14.05.2015

*Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: yury.lankevich@newman.bas-net.by*

Y.Y. Lankevich

AN ALGORITHM FOR ASSEMBLING A COMMON IMAGE OF VLSI LAYOUT

We consider problem of assembling a common image of VLSI layout. Common image is composed of frames obtained by electron microscope photographing. Many frames require a lot of computation for positioning each frame inside the common image. Employing graphics processing units enables acceleration of computations. We realize algorithms and programs for assembling a common image of VLSI layout. Specificity of this work is to use abilities of CUDA to reduce computation time. Experimental results show efficiency of the proposed programs.

УДК 004.921

Ю.Д. Васильева

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ ЛИЦА ПАЦИЕНТА НА ОСНОВЕ БОКОВЫХ ОРТОГНАТИЧЕСКИХ МОДЕЛЕЙ

Приводятся методы прогнозирования профиля лица пациента после лечения зубочелюстных аномалий. Они основываются на поиске оптимального профиля из базы данных ортогнатических моделей, ближайшего к профилю лица пациента после планируемого лечения. Практическая реализация методов позволяет на стадии предварительной диагностики показать пациенту его лицо после планируемого лечения и может быть полезна специалистам-ортодонтам для составления и корректировки плана лечения.

Введение

Зубочелюстные аномалии и нарушения строения зубочелюстной системы встречаются очень часто. Они отражаются на гармонии лицевого скелета, их следствием могут быть эстетические отклонения и развитие пародонтоза [1]. Лечат зубочелюстные аномалии с помощью различных ортодонтических аппаратов, а в сложных случаях прибегают к хирургическому вмешательству.

Существует ряд параметров пропорций лица, которые определяют гармонию лицевого профиля. Термин «гармония профиля» используется в настоящей работе в соответствии с основными понятиями о пропорции лица [2].

Перед лечением зубочелюстных аномалий (ортодонтическим лечением) специалист в зависимости от характера заболевания строит предварительный прогноз положения скелетных тканей лица. В процессе лечения важно предоставить пациенту визуальную информацию о его будущей внешности, о том, как будет выглядеть его лицо (граница мягких тканей) после изменения положения скелетных тканей (границы лицевого скелета).

Задача состоит в разработке метода и алгоритма, которые показывают, как будет выглядеть профиль лица пациента после планируемого ортодонтического лечения, имея только боковую телерентгенограмму черепа пациента [3] и базу данных боковых телерентгенограмм ортогнатических моделей (с гармоничным профилем).

Аналогичная тема затрагивается в работе В.Н. Трезубова [4], который при лечении больных с зубочелюстными аномалиями использует для прогнозирования мягких тканей лица фотографию пациента и программу двухмерного редактора. Прогноз строится специалистом вручную, а основное внимание уделяется построению и планированию скелетной границы.

В Берлине велась работа над проектом CAS [5], цель которого заключалась в точном моделировании хирургического вмешательства на трехмерных моделях лица пациента. В реальном времени хирург может изменять участок зубочелюстной системы и все изменения отражаются на лицевом профиле мягких тканей. Похожая работа была и у Erwin Keeve, Sabine Girod и Bernd Girod, которые визуализируют внешний вид пациента после челюстно-лицевой хирургии [6].

1. Постановка задачи

Необходимо разработать компьютерные методы и алгоритмы, которые в двухмерной среде показывают, как будет выглядеть профиль лица пациента после планируемого ортодонтического лечения. Рассматриваемые методы и алгоритмы основаны на использовании боковой телерентгенограммы (далее ТРГ) головы пациента и базы данных ортогнатических точечных моделей, которая создается заранее и постоянно пополняется. ТРГ представляет собой данные в двухмерной среде, и все расчеты и визуализация выполняются на боковой проекции.

Решение поставленной задачи включает следующие этапы:

1) структурирование базы данных ортогнатических моделей и формирование данных о пациенте;

- 2) выбор критериев подобия скелетных кривых;
- 3) поиск ближайшего профиля из базы данных боковых телерентгенограмм головы ортогнатических моделей к профилю лицевого скелета головы пациента;
- 4) корректировку предварительного положения кожного профиля лица пациента.

При выполнении исследований использована база данных ортогнатических моделей, представленных в виде точечных моделей.

Будем считать, что каждая ТРГ без существенной потери данных может быть заменена точечной моделью. Точечная модель представляет собой набор характерных точек профильной ТРГ. Рассматриваются только характерные точки, расположенные в области зубочелюстной системы, так как профиль лица после лечения зубочелюстных аномалий подвергается значительным изменениям только в этой части.

В ОИПИ НАН Беларуси разработаны методы построения индивидуальных и обобщенных норм лицевого скелета пациента после планируемого лечения зубочелюстных аномалий. Методы были предложены в рамках выполнения работы по автоматизированной диагностике и планированию лечения зубочелюстных аномалий Г.М. Левиным, Л.Н. Ламбиным, И.В. Токаревичем, Э.М. Мельниченко, О.П. Чудаковым и А.В. Глинником [7]. При прогнозировании была использована полученная авторами индивидуальная норма пациента в виде точечной модели, которая отражает приближенную форму скелетной границы профиля лица при прогнозируемом лечении. Задача сводится к формированию прогноза положения мягких тканей лица пациента относительно прогнозируемой формы индивидуальной нормы.

2. Этапы прогнозирования

Рассмотрим этапы прогнозирования более подробно.

1. *Структурирование базы данных ортогнатических моделей и формирование данных о пациенте.* Предлагаемый в работе метод прогнозирования положения мягких тканей лица пациента основан на поиске профиля в базе данных ортогнатических моделей (далее БД), похожего на индивидуальную норму лицевого скелета пациента. С этой целью была использована экспериментальная БД, которая создана заведующим кафедрой ортодонтии Белорусского государственного медицинского университета профессором И.В. Токаревичем. БД состоит из боковых ТРГ ортогнатических моделей, которые заменяются точечными моделями на лицевом скелете и кожном профиле.

Предполагается, что уже имеется некоторый набор ортогнатических моделей, образующих БД. Для ее корректировки предусматривается возможность добавления или удаления данных. Пополнение БД положительно влияет на получение конечного результата, так как при этом увеличивается вероятность нахождения лучшей ортогнатической модели, по форме приближенной к индивидуальной норме скелетного профиля лица пациента.

После формирования БД собирается информация о пациенте, которая включает боковую ТРГ головы, скелетную кривую до лечения (рис. 1, кривая 1), индивидуальную норму пациента (рис. 1, кривая 3), а также кожную кривую до лечения (рис. 1, кривая 2).

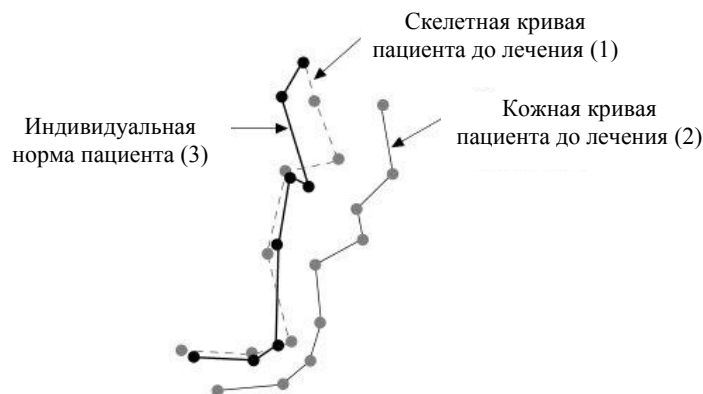


Рис. 1. Данные о пациенте

2. *Выбор критериев подобия скелетных кривых.* После формирования БД ортогнатических моделей и данных о пациенте необходимо в БД найти объект, который меньше всего отличался бы по «упитанности» (толщине мягких тканей) от модели лица пациента и скелетная кривая которого наименее отличалась бы от индивидуальной нормы пациента, представленной в виде кривой.

Для отбора критериев, приемлемых для решения поставленной задачи, были рассмотрены 27 скелетных кривых из БД. Затем совместно с группой экспертов были выбраны четыре критерия, оказывающих основное влияние на принятие решения: площадь между скелетной кривой из БД и индивидуальной нормой пациента; максимальное расстояние между соответствующими точками скелетных кривых; отношение площадей фигур, образованных кожной и скелетной кривыми пациента до лечения и кожной и скелетной кривыми из БД; сумма квадратов расстояний между соответствующими точками скелетных кривых.

Перед вычислением значений критериев предварительно выполняются преобразования данных, позволяющие устранить возможные погрешности, возникающие при подготовке данных о пациенте, а также данных в БД.

Таким образом, степень близости между скелетными кривыми лица пациента и лица из БД решено вычислять с применением следующих критериев:

- площади между скелетной кривой из БД и индивидуальной нормой пациента;
- максимального расстояния между соответствующими точками скелетных кривых из БД и индивидуальной нормой пациента;
- отношения площадей между скелетной и кожной кривыми пациента и скелетной и кожной кривыми из БД;
- суммы квадратов расстояний между соответствующими точками скелетных кривых из БД и индивидуальной нормой пациента.

В приведенной формулировке рассматриваемая задача является многокритериальной с использованием четырех критериев и решается с помощью метода последовательных уступок.

Для отыскания в БД кривой, ближайшей к индивидуальной норме пациента, на первом шаге было выполнено совмещение положения конечных вершин рассматриваемых кривых. Однако такое положение скелетной кривой из БД не обеспечивает ее положения относительно индивидуальной нормы пациента, при котором площадь между ними является минимальной. В связи с этим на следующем этапе применяется метод случайного поиска положения скелетной кривой из БД относительно индивидуальной нормы пациента с вычислением площади, заключенной между этими кривыми. Наименьшее из полученных значений площади для всех построенных случайных положений скелетной кривой из БД принимается за приближенное значение искомой площади. Скелетная кривая из БД, для которой найденное этим методом приближенное значение площади является минимальным, считается ближайшей к индивидуальной норме пациента.

Преобразования координат применяются к ближайшей скелетной кривой и кожной кривой из БД.

Для вычисления площади фигуры, образованной скелетной кривой объекта из БД и индивидуальной нормой пациента, используется известный метод вычисления площади произвольной односвязной или многосвязной области, ограниченной ломаной линией, вершины которой упорядочены вдоль границы области:

$$S_j = \frac{1}{2} \left| \sum_{i=1}^{n-1} (y_i + y_{i+1}) \times (x_{i+1} - x_i) \right|, \quad (1)$$

где i – номера вершин кривой, ограничивающей общую область; j – номер кривой из БД.

Предварительно вычисляются все точки самопересечения граничной кривой, и все вершины ломаных вместе с точками их самопересечения упорядочиваются вдоль ориентированной границы области таким образом, чтобы область всегда оставалась по одну сторону от ее границы. Среди всех полученных результатов выбирается то положение j -й кривой из БД, для которого вычисленная площадь S_j окажется минимальной.

Значение второго критерия представляет собой максимальное расстояние между соответствующими вершинами скелетных кривых из БД и индивидуальной нормой пациента:

$$d_j = \max_i \sqrt{(x_i - x_{i,j})^2 + (y_i - y_{i,j})^2}, \quad (2)$$

где i – номера вершин скелетных кривых; j – номер кривой из БД; (x_i, y_i) – координаты i -й вершины скелетной кривой пациента; $(x_{i,j}, y_{i,j})$ – координаты i -й вершины j -й скелетной ломаной из БД.

Третий критерий рассчитывается исходя из сформированных точечных моделей и определяется как отношение площади фигуры, ограниченной индивидуальной нормой пациента и кожной кривой пациента до лечения $S_{\text{пац}}$, к площади фигуры, ограниченной кожной и скелетной кривыми из БД $S_{\text{бд},j}$:

$$K_j = S_{\text{пац}} / S_{\text{бд},j}. \quad (3)$$

Если полученное значение оказывается меньшим единицы, то в качестве значения третьего критерия принимается обратная величина:

$$K_j = S_{\text{бд},j} / S_{\text{пац}}. \quad (4)$$

Четвертый критерий определяется как сумма квадратов расстояний между соответствующими вершинами скелетной кривой пациента и j -й кривой из БД:

$$D_j = \sum_{i=1}^n ((x_i - x_{i,j})^2 + (y_i - y_{i,j})^2). \quad (5)$$

3. Поиск ближайшего профиля из базы данных боковых ТРГ головы ортогнатических моделей к профилю лицевого скелета головы пациента. Выбор ближайшей кривой из БД решается как многокритериальная задача с помощью метода последовательных уступок [8]. Для этого вначале все частные критерии упорядочиваются в порядке их относительной важности. Затем многокритериальная задача сводится к поочередной оптимизации частных критериев и выбору величин уступок.

Поиск ближайшего профиля осуществляется в следующей последовательности:

1. Решается задача однокритериальной оптимизации по главному критерию. Для этого из БД выбирается та скелетная кривая, у которой площадь S_j между ней и скелетной кривой пациента является минимальной:

$$Y_1 = \min \{S_j \mid j = 1, \dots, N\}, \quad (6)$$

где N – число объектов в БД.

Назначается допустимая уступка Δ_1 по главному критерию. Данная уступка обычно выбирается в пределах от нескольких процентов до 20–30 % значения главного критерия. После решения этой задачи для последующего рассмотрения остаются только те объекты из БД, для которых площадь между скелетными кривыми оказывается меньше значения $Y_{1,\text{max}} = Y_1 \times (1 + \Delta_1)$. Пусть остается N_1 объектов, образующих множество J_1 объектов. В экспериментальных расчетах уступка Δ_1 по главному критерию была назначена в пределах 20 %.

2. Решается задача однокритериальной оптимизации по второму частному критерию и получается ее значение Y_2 . Им будет минимальное значение максимального расстояния d_{j_1} между соответствующими точками скелетных кривых:

$$Y_2 = \min d_{j_1}; \quad j_1 = 1, \dots, N_1; \quad j_1 \in J_1. \quad (7)$$

Дополнительным ограничением для решения второй оптимизационной задачи является то, что значение площади между скелетными кривыми не превышает $Y_{1,\max}$.

Назначается допустимая уступка Δ_2 по второму критерию и для последующего рассмотрения остаются только те объекты из ранее оставшихся, для которых значение второго критерия не превышает $Y_{2,\max} = Y_2 \times (1 + \Delta_2)$. Пусть остается N_2 объектов, образующих множество J_2 объектов. Значение второй уступки также назначается в пределах 20 %.

3. Решается задача однокритериальной оптимизации по третьему критерию и получается ее значение Y_3 . Им будет минимальное значение отношения K_{j_2} площадей мягких тканей лица пациента и объекта из БД:

$$Y_3 = \min K_{j_2}, \quad j_2 = 1, \dots, N_2; \quad j_2 \in J_2. \quad (8)$$

Назначается допустимая уступка Δ_3 по третьему критерию и для последующего рассмотрения остаются только те объекты из ранее оставшихся, для которых значение третьего критерия не превышает $Y_{3,\max} = Y_3 \times (1 + \Delta_3)$. Пусть остается N_3 объектов, образующих множество J_3 объектов. Значение третьей уступки назначается в пределах 25 %.

4. Решается последняя задача однокритериальной оптимизации по четвертому критерию и получается ее значение Y_4 . Им будет минимальное значение суммы квадратов расстояний D_{j_3} между соответствующими точками скелетной кривой пациента и скелетными кривыми из БД, оставшимися для рассмотрения:

$$Y_4 = \min D_{j_3}, \quad j_3 = 1, \dots, N_3; \quad j_3 \in J_3. \quad (9)$$

Назначается допустимая уступка Δ_4 по четвертому критерию. Из последующего рассмотрения исключаются те из оставшихся объектов, для которых сумма квадратов расстояний D_{j_3} между скелетной кривой пациента и скелетными кривыми из БД превышает $Y_{4,\max} = Y_4 \times (1 + \Delta_4)$. Значение последней уступки назначается в пределах 30 %.

В таблице представлены значения всех приведенных выше критериев для одного пациента и 27 объектов из БД.

Значения критериев для одного пациента

Площадь между скелетными кривыми S_j	Максимальное расстояние d_j	Отношение толщин K_j	Сумма квадратов расстояний D_j
1002,505	21,642	1,068 781	1256,573
958,642	20,855	1,096 564	1172,294
545,193	9,676	1,032 931	266,473
736,525	12,625	1,291 805	444,535
726,321	13,483	1,049 883	483,408
1605,011	19,682	1,256 068	1116,911
873,733	25,05	1,38 589	1607,09
1007,205	23,798	1,358 522	1214,599
428,858	15,015	1,006 286	546,411
365,42	6,721	1,061 035	211,832
415,202	20,408	1,056 051	687,433
416,465	18,297	1,13 687	962,511
553,415	15,107	1,138 043	571,674
903,687	18,536	1,087 324	1202,061
762,275	15,878	1,045 207	511,406
1001,965	14,786	1,537 603	633,654
1124,439	22,241	1,046 383	1464,742
983,494	13,926	1,212 628	588,067
825,625	21,214	1,020 796	1416,755
745,422	17,779	1,079 085	594,597

Окончание таблицы

Площадь между скелетными кривыми S_i	Максимальное расстояние d_i	Отношение толщин K_i	Сумма квадратов расстояний D_i
723,247	10,003	1,203 471	486,251
589,028	10,789	1,131 483	424,922
360,426	6,074	1,093 208	230,528
413,682	7,039	1,174 628	369,207
1042,225	22,336	1,100 846	1634,359
372,647	7,184	1,363 723	340,492
1260,348	19,353	1,530 603	1219,581

Оптимальной считается любая стратегия, которая будет получена после вычисления оптимального значения последнего частного критерия. Если, по мнению специалиста, результат является не вполне удовлетворительным, то он может изменить значения уступок и повторить вычисления методом уступок.

После нахождения в БД объекта, ближайшего к индивидуальной норме пациента, определяется кожный профиль лица (рис. 2).

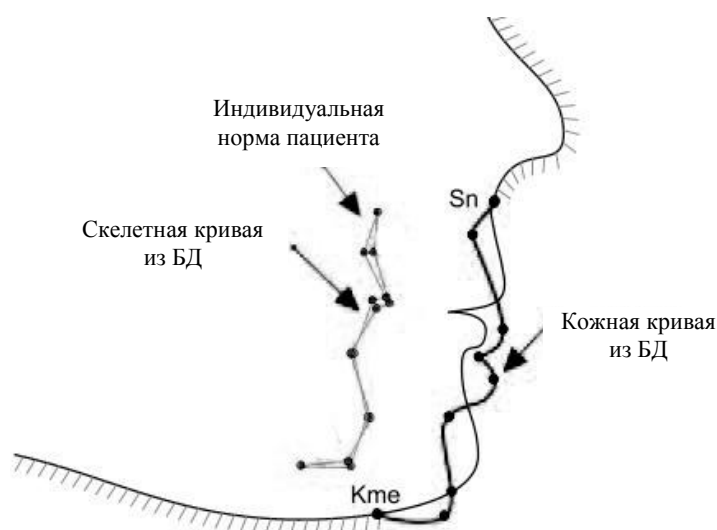


Рис. 2. Совмещение выбранных из БД кожной и скелетной кривых с индивидуальной нормой и кожным профилем пациента

Выбранный профиль подвергается преобразованиям в соответствии с преобразованиями, примененными к скелетной кривой объекта из БД. Преобразованный профиль принимается за предварительное положение кожного профиля лица пациента после планируемого лечения.

4. *Корректировка предварительного положения кожного профиля лица пациента.* Индивидуальная норма пациента отличается от выбранной скелетной кривой объекта из БД. Поэтому найденное положение мягких тканей лица пациента должно быть скорректировано в соответствии с его индивидуальной нормой.

Корректировка осуществляется в два этапа. На первом этапе вычисляются векторы между соответствующими точками на скелетной кривой объекта из БД и индивидуальной нормой пациента. Затем определяется соответствующая им точка на кожной кривой из БД и смещается на тот же вектор. Таким образом получается прогнозируемое положение точек на кожной кривой пациента после лечения.

На втором этапе смещаются промежуточные точки на кожном профиле лица пациента до лечения, расположенные между соответствующими точками на кожной кривой (рис. 3). Величина смещения определяется величиной вектора, рассчитанной по формуле

$$\bar{V}_i = \bar{V}_1 + \frac{i-1}{n_1-1}(\bar{V}_2 - \bar{V}_1), \quad 1 \leq i \leq n_1, \quad (10)$$

$$\bar{V}_j = \bar{V}_i,$$

где \bar{V}_i – вектор, на который смещается i -я промежуточная точка на скелетном профиле из БД для получения соответствующей промежуточной точки индивидуальной нормы пациента. На этот же вектор следует сместить соответствующую промежуточную точку профиля лица из БД (рис. 3, вектор \bar{V}_j).

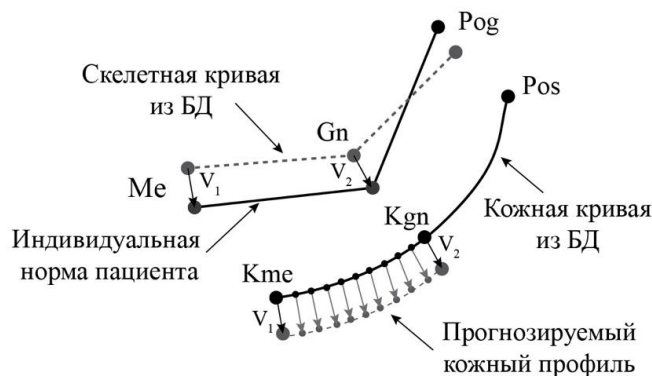


Рис. 3. Корректировка положения кожной кривой из БД

Найденный лицевой профиль считается прогнозируемым профилем мягких тканей лица пациента после планируемого лечения.

3. Результаты экспериментальных исследований

Экспериментальные исследования проводились на кафедре стоматологии Белорусского государственного медицинского университета. Были рассмотрены десять пациентов, которые уже прошли лечение и после лечения имели гармоничный профиль лица. Оценка проводилась путем наложения прогнозируемого профиля лица, полученного в результате моделирования, на профиль реального лица после лечения (рис. 4). Таким образом была проверена результативность методики прогнозирования на боковой ТРГ головы. Проведенные исследования показали достаточно высокую степень совпадения результатов моделирования с результатами лечения.



Рис. 4. Визуальная оценка соответствия прогнозируемого положения мягких тканей на примере одного пациента: а) до лечения; б) после лечения; в) сравнение результатов

Заключение

В работе предлагается метод поиска профиля из БД ортогнатических моделей, ближайшего к профилю лица пациента после планируемого лечения. Для этого анализируется индивидуальная норма пациента и кожные и скелетные кривые из БД. БД состоит из боковых ТРГ головы и соответствующего им набора скелетных и кожных точечных моделей. Поиск ближайшего профиля выполняется с помощью метода последовательных уступок с многократным применением следующих критериев: площади между скелетными кривыми; максимального расстояния между соответствующими точками скелетных кривых; суммы квадратов расстояний между соответствующими точками скелетных кривых; отношения площадей фигур, образованных скелетными и кожными кривыми пациента и объекта из БД.

Разработан также метод прогнозирования положения мягких тканей лица пациента на основе боковой ТРГ головы. Найденный в БД ближайший профиль накладывается на лицо пациента с дополнительными корректировками для индивидуализации формы, что позволяет специалисту на стадии диагностики показать пациенту его будущую внешность. Методы ориентированы на пациентов, при лечении которых используются аппаратные методы без хирургического вмешательства.

Список литературы

1. Методы лечения зубочелюстных аномалий / Стоматологический информационный портал [Электронный ресурс]. – 2009. – Режим доступа : <http://stomfak.ru/ortodontiya/metody-lecheniya-zubochelyustnyh-anomalij.html?ite mid=76>. – Дата доступа : 25.04.2013.
2. Талалаева, Е.В. Оценка эстетики лица у лиц с физиологической окклюзией зубных рядов при помощи 3D-сканер-системы : автореф. дис. ... канд. мед. наук : 14.01.14 / Е.В. Талалаева; Московский гос. медико-стоматол. ун-т. – М., 2012. – 22 с.
3. Палов, Н.А. Рентгенодиагностика / Н.А. Палов, А.З. Гингольд, К.А. Москачева. – М. : МЕД пресс-информ, 1999. – 556 с.
4. Трезубов, В.Н. Планирование и прогнозирование лечения больных с зубочелюстными аномалиями / В.Н. Трезубов, Р.А. Фадеев – М. : МЕД пресс-информ, 2005. – 224 с.
5. Computer aided planning for orthognatic surgery / M. Chabanas [et al.] // Proc. of the 16th Intern. congress and exhibition «Computer Assisted Radiology and Surgery». – Italy, 2002. – P. 12–14.
6. Anatomy-based facial tissue modeling using the finite element method / E. Keeve [et al.] // Proc. of the 7th conf. on visualization '96. – USA, 1996. – P. 21–28.
7. Автоматизированная диагностика и планирование лечения зубочелюстных аномалий / Л.Н. Ламбин [и др.]. – Минск : Ин-т техн. кибернетики АН Беларуси, Минск. гос. мед. ин-т, 1996. – 241 с.
8. Метод последовательных уступок [Электронный ресурс]. – 2006. – Режим доступа : <http://ref.by/refs/49/32807/1.html>. – Дата доступа : 15.01.2014.

Поступила 28.07.2015

*Институт современных знаний
им. А.М. Широкова,
Минск, ул. Филимонова, 69
e-mail: vyud@tut.by*

Y.D. Vasilyeva

**PATIENT'S FACE COMPUTER-ASSISTED SIMULATION ON THE BASIS
OF LATERAL-VIEW ORTHOGNATHIC MODELS**

Methods of predicting patient's facial profile after malocclusion treatment are considered. Methods are based on the orthognathic models data base search for the profile, demonstrating the best patient's face after the planned treatment. Practical application of these methods provides an opportunity to demonstrate the face after the planned treatment to the patient at the stage of preliminary diagnosis and can be useful for orthodontists in treatment planning and correction.

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

УДК 519.872

С.А. Дудин, О.С. Дудина

ИССЛЕДОВАНИЕ МНОГОЛИНЕЙНОЙ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ С АБСОЛЮТНЫМ ПРИОРИТЕТОМ И ПОВТОРНЫМИ ВЫЗОВАМИ

Исследуется многолинейная система массового обслуживания с повторными вызовами без буфера и двумя типами запросов как модель системы когнитивного радио. Запросы первого типа имеют абсолютный приоритет. Процессы поступления запросов первого и второго типов моделируются марковскими входными потоками. Время обслуживания запроса имеет экспоненциальное распределение с зависимым от типа запроса параметром. Запросы второго типа принимаются на обслуживание, если число занятых каналов меньше заданного порогового значения. Запросы второго типа могут совершать повторные попытки попасть на обслуживание. Находится условие существования стационарного режима. Приводятся формулы для вычисления основных характеристик производительности системы.

Введение

В настоящее время ресурсы большинства проводных телекоммуникационных систем не являются дефицитными, в то время как ресурсы беспроводных сетей связи строго дефицитные. Это обусловлено тем, что доступный спектр радиочастот ограничен физической природой. В связи с быстрым ростом числа смартфонов и других мобильных устройств передача данных в беспроводных сетях может столкнуться с катастрофической нехваткой частот. Мобильные операторы и научно-исследовательские лаборатории прилагают значительные усилия для решения задачи оптимального использования спектра радиочастот. Одним из возможных решений является применение технологии когнитивного радио. В системах когнитивного радио предлагается давать возможность нелицензированным пользователям сети временно «захватывать» неиспользуемый спектр частот, не создавая при этом помех для лицензированных пользователей (см., например, [1]). Проблема оптимизации совместного доступа первичных и вторичных пользователей может быть эффективно решена с помощью теории систем массового обслуживания.

На сегодняшний день существует достаточно большое количество работ, посвященных исследованию систем когнитивного радио. Как правило, в литературе рассматриваются модели с двумя типами запросов, где запросы первого типа имеют абсолютный приоритет над запросами второго типа. В случае если запрос первого типа поступает в систему в момент, когда все приборы заняты, но на обслуживании есть запросы второго типа, обслуживание одного запроса второго типа прекращается (запрос выбивается с обслуживания) и запрос первого типа занимает освободившийся прибор. Таким образом, запрос первого типа получает отказ только тогда, когда в момент его прихода все приборы заняты запросами первого типа, т. е. запросы второго типа не оказывают влияние на обслуживание запросов первого типа. Обслуживание запросов второго типа может прерываться, что может приводить к бесполезной трате пропускной способности системы и низкому качеству обслуживания запросов второго типа. Чтобы минимизировать вероятность наступления таких прерываний, предлагается осуществлять управление доступом в систему запросов второго типа, т. е. остановить прием вторичных запросов, когда число занятых приборов превышает некоторое пороговое значение и риск выбивания вторичных запросов является высоким. Таким образом, разумно резервировать некоторое число каналов только для обслуживания первичных пользователей. Сведения о состоянии дел в данной предметной области можно получить, например, из статей [1–4].

Практически все модели систем когнитивного радио, рассматриваемые в литературе, имеют следующие существенные недостатки. В частности, предполагается, что потоки запросов лицензированных и нелицензированных пользователей описываются стационарным пуассоновским входным потоком, в то время как входные потоки могут быть коррелированными. Также предполагается, что нелицензированные (вторичные) пользователи, которые получают отказ в обслуживании в момент прихода, покидают систему навсегда, в то время как на самом деле вторичный пользователь может повторять попытки получить обслуживание.

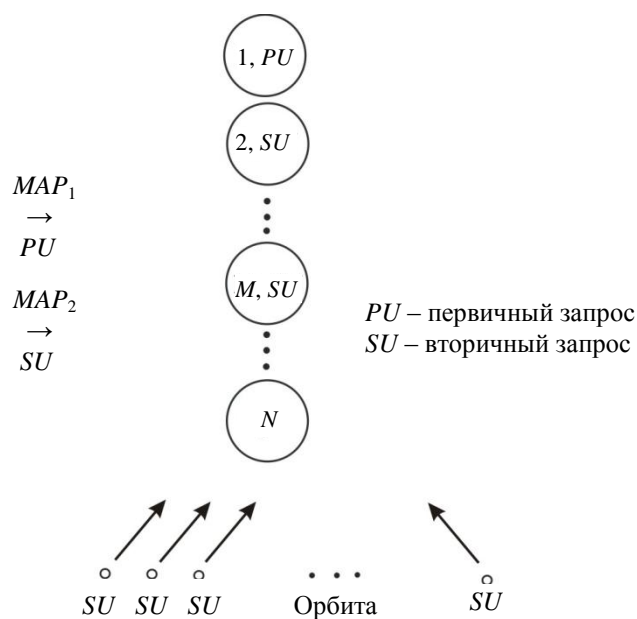
Преимущества модели, исследуемой в данной статье, по сравнению с моделями, посвященными анализу систем когнитивного радио посредством теории массового обслуживания, заключаются в следующем:

1. Предполагается, что входные потоки первичных и вторичных запросов описываются марковскими входными потоками (МАР, от англ. Markovian Arrival Process). МАР-поток коррелированный, поэтому он идеально подходит для моделирования коррелированного или пульсирующего трафика в современных телекоммуникационных сетях. Стационарный пуассоновский процесс является простейшим случаем МАР-потока. Если попытаться описать некоторый реальный поток с помощью стационарного пуассоновского входного потока, то модель будет отражать только среднюю скорость поступления, но не дисперсию или более высокие моменты распределения времени между моментами поступления запросов и возможную корреляцию между этими временами.

2. Предполагается, что вторичный запрос, который не получает доступ в систему в момент поступления или обслуживания которого прервано, имеет возможность покинуть систему навсегда или пойти на так называемую в литературе орбиту (некоторое виртуальное место) и повторить попытку получить доступ к ресурсам системы через случайный временной промежуток. Считается, что вторичные запросы могут быть ненастойчивыми (могут покинуть систему после неудачной попытки) и (или) нетерпеливыми (могут покинуть систему через некоторое время пребывания на орбите).

1. Математическая модель

Рассмотрим систему массового обслуживания без буфера, с двумя типами запросов и повторными вызовами, состоящую из N приборов (рисунок).



Структура системы

В систему поступают два МАР-потока запросов. Запросы первого типа (первичных пользователей) поступают в соответствии с потоком, обозначаемым как $МАР_1$ и заданным неприводимой цепью Маркова $w_t, t \geq 0$, с непрерывным временем и конечным пространством состояний $\{0, 1, \dots, W\}$. Время пребывания цепи в состоянии w экспоненциально распределено с положительным параметром λ_w . Когда время пребывания в состоянии w истекло, с вероятностью $p_{w,w'}^{(k)}$ процесс w_t переходит в состояние w' и при этом генерируется k запросов, $k = 0, 1$; $w, w' = \overline{0, W}$. Поведение МАР-потока полностью характеризуется матрицами $D_k, k = 0, 1$, элементы которых определяются следующим образом:

$$(D_k)_{w,w'} = \lambda_w p_{w,w'}^{(k)}, \quad k = 1, \quad k = 0, \quad w \neq w';$$

$$(D_0)_{w,w} = -\lambda_w, \quad w = \overline{0, W}.$$

Матрица $D(1) = D_0 + D_1$ представляет собой инфинитезимальный генератор цепи $w_t, t \geq 0$. Средняя интенсивность поступления запросов первого типа λ_1 имеет вид

$$\lambda_1 = \theta D_1 e,$$

где θ – вектор стационарного распределения цепи Маркова $w_t, t \geq 0$. Вектор θ является единственным решением системы линейных алгебраических уравнений

$$\theta D(1) = \mathbf{0}, \quad \theta e = 1.$$

Здесь и далее e – вектор-столбец, состоящий из единиц; $\mathbf{0}$ – вектор-строка, состоящая из нулей.

Коэффициент вариации c_{var} длин интервалов между моментами поступления запросов первого типа определяется формулой

$$c_{var}^2 = 2\lambda_1 \theta (-D_0)^{-1} e - 1.$$

Коэффициент корреляции c_{cor} длин двух соседних интервалов между поступлением запросов первого типа вычисляется следующим образом:

$$c_{cor} = (\lambda_1 \theta (-D_0)^{-1} (D(1) - D_0) (-D_0)^{-1} e - 1) / c_{var}^2.$$

Запросы второго типа (вторичных пользователей) поступают в систему в соответствии с марковским входным потоком $МАР_2$, который управляется случайным процессом $v_t, t \geq 0$, с конечным пространством состояний $\{0, 1, \dots, V\}$ и описывается квадратными матрицами H_0 и H_1 размером $\bar{V} = V + 1$. Средняя интенсивность λ_2 поступления запросов второго типа вычисляется как $\lambda_2 = \chi H_1 e$, где χ – вектор стационарного распределения процесса $v_t, t \geq 0$, – определяется как единственное решение системы линейных алгебраических уравнений

$$\chi (H_0 + H_1) = \mathbf{0}, \quad \chi e = 1.$$

Запросы первого типа считаются приоритетными, а запросы второго типа – неприоритетными. Полагаем, что запросы второго типа принимаются на обслуживание, если число занятых каналов меньше чем $M, N \geq M > 0$. Если в момент прихода запроса второго типа число заня-

тых приборов больше либо равно M , то запрос идет на орбиту с вероятностью q_1 , а с дополнительной вероятностью покидает систему.

Если в момент прибытия запроса первого типа есть свободный прибор, то этот запрос принимается в систему. Если в момент прибытия запроса первого типа все приборы заняты, но на обслуживании есть запрос второго типа, то этот приоритетный запрос «выбивает» неприоритетный запрос с обслуживания, т. е. приоритетный запрос начинает обслуживание, а выбитый неприоритетный запрос идет на орбиту с вероятностью q_2 или покидает систему с дополнительной вероятностью. Если в момент прибытия запроса первого типа все приборы заняты и на обслуживании нет запросов второго типа, то пришедший запрос покидает систему.

Запросы, находящиеся на орбите, совершают повторные попытки попасть на обслуживание через экспоненциально распределенное с параметром α , $\alpha > 0$, время. В случае если попытка попасть на обслуживание осуществляется, когда число занятых приборов больше либо равно M , то запрос возвращается на орбиту с вероятностью q_1 , а с дополнительной вероятностью покидает систему, т. е. запросы с орбиты могут проявлять ненастойчивость.

Запросы с орбиты также могут проявлять нетерпеливость. Считаем, что запрос покидает орбиту через экспоненциально распределенное время с параметром γ , $\gamma > 0$, с момента попадания на нее.

Время обслуживания запросов первого и второго типов имеет экспоненциальное распределение с параметрами μ_1 и μ_2 соответственно.

Условие, что запросы второго типа принимаются в систему, когда число занятых каналов меньше чем M , $M > 0$, по сути, означает, что для приоритетных запросов зарезервировано $N - M$ приборов. Это может показаться бессмысленным, так как приоритетные запросы могут выбивать неприоритетные с обслуживания и не нуждаются в резервировании каналов. На самом деле резервирование имеет смысл для улучшения качества обслуживания запросов второго типа. Резервирование может помочь избежать частых выбиваний запросов второго типа с обслуживания и тем самым увеличить пропускную способность системы. В случае частых выбиваний ресурсы системы на обслуживание выбитых запросов второго типа теряются впустую, а резервирование приборов может помочь использовать эти ресурсы оптимальным образом.

2. Процесс изменения состояний системы и стационарное распределение числа запросов в системе

Введем в рассмотрение следующие процессы:

$i_t, i_t \geq 0$, – число запросов на орбите;

$n_t, n_t = \overline{0, N}$, – число занятых приборов;

$l_t, l_t = \overline{0, \min\{n_t, M\}}$, – число запросов второго типа на обслуживании;

$w_t, w_t = \overline{0, W}$, – состояние управляющего процесса MAP_1 ;

$v_t, v_t = \overline{0, V}$, – состояние управляющего процесса MAP_2

в момент времени $t, t \geq 0$.

Процесс $\xi_t = \{i_t, n_t, l_t, w_t, v_t\}$, $t \geq 0$, является неприводимой цепью Маркова с непрерывным временем.

Примем следующие обозначения:

I – единичная матрица, O – нулевая матрица соответствующего размера;

\oplus и \otimes – символы кронекеровой суммы и кронекерова произведения матриц (см., например, [5]);

$\bar{W} = W + 1$;

$C_l = \text{diag}\{0, 1, \dots, l\}$, $l = \overline{0, M}$;

$\bar{C}_l = \text{diag}\{l, l-1, \dots, 0\}$, $l = \overline{0, M}$;

$\tilde{C}_l = \text{diag}\{l, l-1, \dots, l-M+1, l-M\}$, $l = \overline{M, N}$;
 $\text{diag}\{A_1, \dots, A_l\}$ – блочно-диагональная матрица с диагональными блоками A_1, \dots, A_l ;
 E_l^+, \tilde{E}_l^+ , $l = \overline{0, M-1}$, – матрицы размером $(l+1) \times (l+2)$, которые определяются как

$$E_l^+ = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}, \quad \tilde{E}_l^+ = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix};$$

E_l^-, \tilde{E}_l^- , $l = \overline{1, M}$, – матрицы размером $(l+1) \times l$, которые определяются как

$$E_l^- = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 0 \end{pmatrix}, \quad \tilde{E}_l^- = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix};$$

\tilde{I} – диагональная матрица размером $(M+1)(N-M/2+1)$ с диагональными элементами $\{0, \dots, 0, 1, \dots, 1\}$;
 $\overbrace{M(M+1)/2}^{M(M+1)/2}$

E^-, \tilde{I} – квадратные матрицы размером $M+1$, которые имеют вид

$$E^- = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}, \quad \tilde{I} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}.$$

Перенумеруем состояния цепи ξ_t в лексикографическом порядке компонент (i, n, l, w, v) . Множество состояний, имеющих значение (i, n) двух первых компонент цепи, будем называть макросостоянием (i, n) .

Пусть A – генератор цепи Маркова ξ_t , $t \geq 0$, который сформирован из блоков $A_{i,j}$, состоящих из матриц $(A_{i,j})_{n,n'}$ интенсивностей переходов цепи ξ_t , $t \geq 0$, из макросостояния (i, n) в макросостояние (j, n') , $n, n' = \overline{0, \min\{i, N\}}$. Диагональные элементы матрицы $A_{i,i}$ отрицательны и их модули определяют интенсивность выхода из соответствующего состояния цепи Маркова.

Теорема 1. Генератор A имеет следующую блочно-трехдиагональную структуру:

$$A = \begin{pmatrix} A_{0,0} & A_{0,1} & O & O & \dots \\ A_{1,0} & A_{1,1} & A_{1,2} & O & \dots \\ O & A_{2,1} & A_{2,2} & A_{2,3} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Не нулевые блоки $A_{i,j}$, $i, j \geq 0$, имеют вид

$$A_{i,i} = \begin{pmatrix} Q_i^{(0)} & B^{(0)} & O & \dots & O & O \\ F^{(1)} & Q_i^{(1)} & B^{(1)} & \dots & O & O \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ O & O & O & \ddots & Q_i^{(N-1)} & B^{(N-1)} \\ O & O & O & \dots & F^{(N)} & Q_i^{(N)} \end{pmatrix} +$$

$$+(1-q_1)\tilde{I} \otimes I_{\bar{W}} \otimes H_1 + I_{(M+1)(N-M/2+1)} \otimes (D_0 \oplus H_0), \quad i \geq 0,$$

$$A_{i,i+1} = A^+ = \text{diag}\{H^{(0)}, \dots, H^{(N)}\}, \quad i \geq 0,$$

$$A_{i,i-1} = \begin{pmatrix} L_i^{(0)} & \tilde{B}_i^{(0)} & O & \dots & O & O \\ O & L_i^{(1)} & \tilde{B}_i^{(1)} & \dots & O & O \\ O & O & L_i^{(2)} & \dots & O & O \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ O & O & O & \ddots & L_i^{(N-1)} & \tilde{B}_i^{(N-1)} \\ O & O & O & \dots & O & L_i^{(N)} \end{pmatrix}, \quad i \geq 1,$$

где

$$Q_i^{(n)} = \begin{cases} -(\mu_2 C_n + \mu_1 \bar{C}_n + i(\alpha + \gamma)I_{n+1}) \otimes I_{\bar{W}\bar{V}}, & n < M, \\ -(\mu_2 C_M + \mu_1 \tilde{C}_n + i((1-q_1)\alpha + \gamma)I_{M+1}) \otimes I_{\bar{W}\bar{V}}, & M \leq n < N, \\ -(\mu_2 C_M + \mu_1 \tilde{C}_n + i((1-q_1)\alpha + \gamma)I_{M+1}) \otimes I_{\bar{W}\bar{V}} + \\ + ((1-q_2)E^- + \tilde{I}) \otimes D_1 \otimes I_{\bar{V}}, & n = N, \end{cases} \quad i \geq 0;$$

$$B^{(n)} = \begin{cases} E_n^+ \otimes I_{\bar{W}} \otimes H_1 + \tilde{E}_n^+ \otimes D_1 \otimes I_{\bar{V}}, & n < M, \\ I_{M+1} \otimes D_1 \otimes I_{\bar{V}}, & M \leq n < N; \end{cases}$$

$$F^{(n)} = \begin{cases} (\mu_1 \bar{C}_n E_n^- + \mu_2 C_n \tilde{E}_n^-) \otimes I_{\bar{W}\bar{V}}, & n \leq M, \\ (\mu_1 \tilde{C}_n + \mu_2 C_M E^-) \otimes I_{\bar{W}\bar{V}}, & M < n \leq N; \end{cases}$$

$$H^{(n)} = \begin{cases} O, & n < M, \\ q_1 I_{(M+1)\bar{W}} \otimes H_1, & M \leq n < N, \\ q_2 E^- \otimes D_1 \otimes I_{\bar{V}} + q_1 I_{M+1} \otimes I_{\bar{W}} \otimes H_1, & n = N; \end{cases}$$

$$\tilde{B}_i^{(n)} = \begin{cases} i\alpha E_n^+ \otimes I_{\bar{W}\bar{V}}, & n < M, \\ O, & n \geq M, \end{cases} \quad i \geq 1;$$

$$L_i^{(n)} = \begin{cases} i\gamma I_{(n+1)\bar{W}\bar{V}}, & n < M, \\ i(\gamma + (1-q_1)\alpha) I_{(M+1)\bar{W}\bar{V}}, & M \leq n \leq N, \end{cases} \quad i \geq 1.$$

Замечание. Цепь Маркова ξ_t , $t \geq 0$, принадлежит классу асимптотически квазитеплицевых цепей Маркова (АКТЦМ) с непрерывным временем [6].

Как следует из [6], необходимым условием существования стационарного распределения АКТЦМ ξ_t , $t \geq 0$, является условие

$$\mathbf{y}Y_0\mathbf{e} > \mathbf{y}Y_2\mathbf{e}, \quad (1)$$

где вектор-строка \mathbf{y} – единственное решение системы линейных алгебраических уравнений

$$\mathbf{y}(Y_0 + Y_1 + Y_2) = \mathbf{y}, \mathbf{y}\mathbf{e} = 1, \quad (2)$$

матрицы Y_0 , Y_1 и Y_2 определяются выражениями

$$Y_0 = \lim_{t \rightarrow \infty} R_i^{-1} A_{t,i-1}, \quad Y_1 = \lim_{t \rightarrow \infty} R_i^{-1} A_{t,i} + I, \quad Y_2 = \lim_{t \rightarrow \infty} R_i^{-1} A_{t,i+1},$$

а R_i – диагональная матрица, диагональные элементы которой определяются как модули соответствующих диагональных элементов матрицы $A_{t,i}$, $i \geq 0$.

Легко убедиться, что матрица R_i является блочно-диагональной матрицей с диагональными блоками $T_i^{(n)}$, $n = \overline{0, N}$, $i \geq 0$, которые определяются следующим образом:

$$T_i^{(n)} = \begin{cases} -A_i^{(n)} + I_{n+1} \otimes (\Sigma_1 \oplus \Lambda_1), & n = \overline{0, M-1}; \\ -A_i^{(n)} - (1-q_1)I_{M+1} \otimes I_{\bar{w}} \otimes \Lambda_2 + I_{M+1} \otimes (\Sigma_1 \oplus \Lambda_1), & n = \overline{M, N-1}; \\ (\mu_1 \tilde{C}_N + \mu_2 C_M + i((1-q_1)\alpha + \gamma)) \otimes I_{\bar{w}\bar{v}} - \\ -\tilde{I} \otimes \Sigma_2 \otimes I_{\bar{v}} - (1-q_1)I_{(M+1)\bar{w}} \otimes \Lambda_2 + I_{M+1} \otimes (\Sigma_1 \oplus \Lambda_1), & n = N, \end{cases} \quad i \geq 0,$$

где Σ_1 , Σ_2 , Λ_1 и Λ_2 – диагональные матрицы, диагональные элементы которых являются диагональными элементами матриц $-D_0$, D_1 , $-H_0$ и H_1 соответственно.

В рассматриваемом случае матрицы Y_0 , Y_1 и Y_2 определяются как

$$Y_0 = \begin{pmatrix} \frac{\gamma}{\gamma+\alpha} & \frac{\alpha}{\gamma+\alpha} E_0^+ & \dots & O & O & O & \dots & O \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ O & O & \dots & \frac{\gamma}{\gamma+\alpha} I_M & \frac{\alpha}{\gamma+\alpha} E_M^+ & O & \dots & O \\ O & O & \dots & O & I_{M+1} & O & \dots & O \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ O & O & \dots & O & O & O & \dots & O \\ O & O & \dots & O & O & O & \dots & I_{M+1} \end{pmatrix} \otimes I_{\bar{w}\bar{v}}, \quad Y_1 = O, \quad Y_2 = O.$$

Очевидно, что при этом условие (1) может быть переписано в виде $\mathbf{y}Y_0\mathbf{e} > 0$, а система (2) – в виде $\mathbf{y}Y_0 = \mathbf{y}$, $\mathbf{y}\mathbf{e} = 1$. Исходя из этого, получаем неравенство

$$\mathbf{y}Y_0\mathbf{e} = \mathbf{y}\mathbf{e} = 1 > 0,$$

т. е. условие (1) выполняется для любых параметров системы.

Поскольку условие эргодичности системы выполнено для любых параметров системы, существуют следующие пределы (стационарные вероятности):

$$p(i, n, l, w, v) = \lim_{t \rightarrow \infty} P\{i_t = i, n_t = n, l_t = l, w_t = w, v_t = v\}, \quad i \geq 0, \quad n = \overline{0, \min\{i, N\}},$$

$$l = \overline{0, \min\{n, M\}}, \quad w = \overline{0, W}, \quad v = \overline{0, V}.$$

Сформируем вектор-строки \mathbf{p}_i :

$$\mathbf{p}(i, n, l, w) = (p(i, n, l, w, 0), p(i, n, l, w, 1), \dots, p(i, n, l, w, V)), \quad w = \overline{0, W};$$

$$\mathbf{p}(i, n, l) = (\mathbf{p}(i, n, l, 0), \mathbf{p}(i, n, l, 1), \dots, \mathbf{p}(i, n, l, W)), \quad l = \overline{0, \min\{n, M\}};$$

$$\mathbf{p}(i, n) = (\mathbf{p}(i, n, 0), \mathbf{p}(i, n, 1), \dots, \mathbf{p}(i, n, \min\{n, M\})), \quad n = \overline{0, \min\{i, N\}};$$

$$\mathbf{p}_i = (\mathbf{p}(i, 0), \mathbf{p}(i, 1), \dots, \mathbf{p}(i, \min\{i, N\})), \quad i \geq 0.$$

Известно, что векторы \mathbf{p}_i , $i \geq 0$, удовлетворяют следующей системе линейных алгебраических уравнений:

$$(\mathbf{p}_0, \mathbf{p}_1, \dots)A = \mathbf{0}, \quad (\mathbf{p}_0, \mathbf{p}_1, \dots)\mathbf{e} = 1, \quad (3)$$

где A – инфинитезимальный генератор цепи Маркова ξ_t , $t \geq 0$.

С учетом того что генератор A в данной модели имеет блочно-трехдиагональную форму, для решения системы (3) может быть использован численно устойчивый алгоритм, приведенный в работе [7].

3. Характеристики производительности

Найдя векторы стационарных вероятностей \mathbf{p}_i , $i \geq 0$, можно вычислить различные характеристики производительности системы:

– распределение вероятностей числа запросов на орбите

$$\lim_{t \rightarrow \infty} P\{i_t = i\} = \mathbf{p}_i \mathbf{e}, \quad i \geq 0;$$

– среднее число запросов в системе

$$L = \sum_{i=0}^{\infty} \sum_{n=0}^N (i+n) \mathbf{p}_i \mathbf{e};$$

– среднее число запросов на орбите

$$L_{orbit} = \sum_{i=1}^{\infty} i \mathbf{p}_i \mathbf{e};$$

– среднее число занятых приборов

$$N_{server} = \sum_{i=0}^{\infty} \sum_{n=1}^N n \mathbf{p}_i \mathbf{e};$$

– среднее число занятых приборов, обслуживающих запросы первого типа,

$$N_{server-1} = \sum_{i=0}^{\infty} \sum_{n=1}^N \sum_{l=0}^{\min\{n,M\}} (n-l) \mathbf{p}(i, n, l) \mathbf{e};$$

– среднее число занятых приборов, обслуживающих запросы второго типа,

$$N_{server-2} = \sum_{i=0}^{\infty} \sum_{n=1}^N \sum_{l=1}^{\min\{n,M\}} l \mathbf{p}(i, n, l) \mathbf{e} = N_{server} - N_{server-1};$$

– среднюю интенсивность выходящего потока обслуженных запросов первого типа

$$\lambda_{out}^{(1)} = \mu_1 N_{server-1};$$

– среднюю интенсивность выходящего потока обслуженных запросов второго типа

$$\lambda_{out}^{(2)} = \mu_2 N_{server-2};$$

– среднюю интенсивность выходящего потока обслуженных запросов

$$\lambda_{out} = \lambda_{out}^{(1)} + \lambda_{out}^{(2)} \cdot \frac{1}{n};$$

– вероятность потери запроса первого типа

$$P_1^{(loss)} = \lambda_1^{-1} \sum_{i=0}^{\infty} \mathbf{p}(i, N, 0) (D_1 \otimes I_{\bar{V}}) \mathbf{e} = 1 - \frac{\lambda_{out}^{(1)}}{\lambda_1};$$

– вероятность потери запроса второго типа

$$P_2^{(loss)} = 1 - \frac{\lambda_{out}^{(2)}}{\lambda_2};$$

– вероятность потери произвольного запроса

$$P^{(loss)} = 1 - \frac{\lambda_{out}}{\lambda_1 + \lambda_2};$$

– вероятность потери запроса второго типа на входе в систему из-за занятости более $M - 1$ приборов

$$P^{(ent-loss)} = (1 - q_1) \lambda_2^{-1} \sum_{i=0}^{\infty} \sum_{n=M}^N \mathbf{p}(i, n) (I_{(M+1)\bar{W}} \otimes H_1) \mathbf{e};$$

– вероятность того, что запрос второго типа поступит в систему, когда число занятых приборов больше $M - 1$, и пойдет на орбиту,

$$P^{(ent-to-orbit)} = q_1 \lambda_2^{-1} \sum_{i=0}^{\infty} \sum_{n=M}^N \mathbf{p}(i, n) (I_{(M+1)\bar{W}} \otimes H_1) \mathbf{e};$$

– вероятность того, что запрос второго типа будет выбит запросом первого типа и покинет систему,

$$P^{(knock-out-loss)} = (1 - q_2) \lambda_2^{-1} \sum_{i=0}^{\infty} \sum_{l=1}^M p(i, N, l) (D_1 \otimes I_{\bar{V}}) e;$$

– вероятность потери запроса с орбиты

$$P^{(loss-from-orbit)} = P_2^{(loss)} - P^{(ent-loss)} - P^{(knock-out-loss)};$$

– вероятность того, что запрос с орбиты совершит повторную попытку, когда число занятых приборов больше $M - 1$, и вернется на орбиту,

$$P^{(return-to-orbit)} = \frac{q_1}{\alpha L_{orbit}} \sum_{i=1}^{\infty} \sum_{n=M}^N i \alpha p(i, n) e.$$

Заключение

В статье изучена многолинейная система массового обслуживания с повторными вызовами и запросами разных приоритетов. Построен процесс изменения состояний системы, найдено условие его эргодичности, приведены формулы для нахождения основных характеристик производительности системы. Результаты исследования могут применяться для оценивания производительности и оптимизации функционирования систем когнитивного радио.

Данная работа выполнена при частичной поддержке Белорусского республиканского фонда фундаментальных исследований Ф15М-005.

Список литературы

1. Feasibility analysis of vehicular dynamic spectrum access via queueing theory model / S. Chen [et al.] // IEEE Communications Magazine. – 2011. – Vol. 49. – P. 156–163.
2. Next generation dynamic spectrum access cognitive radio wireless networks: A survey / I.F. Akyildiz [et al.] // Computer Networks. – 2006. – Vol. 50(13). – P. 2127–2159.
3. Performance analysis of dynamic spectrum handoff scheme with variable bandwidth demand on secondary users for cognitive radio networks / Y. Konishi [et al.] // Wireless Networks. – 2013. – Vol. 19. – P. 607–617.
4. Zahed, S. Analytical modeling for spectrum handoff decision in cognitive radio networks / S. Zahed, I. Awan, A. Cullen // Simulation modelling practice and theory. – 2013. – Vol. 38. – P. 98–114.
5. Graham, A. Kronecker Products and Matrix Calculus with Applications / A. Graham. – Chichester : Ellis Horwood, 1981. – 130 p.
6. Klimenok, V.I. Multi-dimensional asymptotically quasi-Toeplitz Markov chains and their application in queueing theory / V.I. Klimenok, A.N. Dudin // Queueing Systems. – 2006. – Vol. 54. – P. 245–259.
7. Dudina, O. Retrial queueing system with Markovian arrival flow and phase-type service time distribution / O. Dudina, Ch. Kim, S. Dudin // Computers and Industrial Engineering. – 2013. – Vol. 66. – P. 360–373.

Поступила 12.05.2015

Белорусский государственный университет,
Минск, пр. Независимости, 4
e-mail: dudin85@mail.ru,
dudina_olga@email.com

S.A. Dudin, O.S. Dudina**ANALYSIS OF MULTI-SERVER QUEUEING SYSTEM
WITH PREEMPTIVE PRIORITY AND REPEATED CALLS**

Multi-server retrial queueing system with no buffer and two types of customers is analyzed as the model of cognitive radio system. Customers of type 1 have a preemptive priority. Customers of both types arrive according to Markovian Arrival Processes. Service times have exponential distribution with parameter depending on the customer type. Type 2 customers are admitted for service only if the number of busy servers is less than the predefined threshold. The rejected type 2 customers retry for the service. Existence condition of the stationary mode of system operation is derived. Formulas for computing key performance measures of the system are presented.

УДК 539.3

О.Л. Швед, А.А. Абрамов

ПЕРЕХОД ОТ АКТИВНОГО НАГРУЖЕНИЯ К ПАССИВНОМУ ПРИ МАТЕМАТИЧЕСКОМ МОДЕЛИРОВАНИИ НЕЛИНЕЙНО УПРУГОПЛАСТИЧЕСКОГО ПРОЦЕССА

Предлагается модель нелинейно упругопластического материала. Формулируется условие непрерывности при переходе от активного нагружения к пассивному, которое обеспечивается в регулярной точке поверхности текучести с неположительной удельной мощностью деформации. В противном случае непрерывность перехода отсутствует. Рассматривается возможность описания явления «запирания» области высокого давления при переходе материала из пластического состояния в упругое.

Введение

В работах [1–3] представлена модель нелинейно упругопластического материала. Она предназначена прежде всего для математического моделирования процессов обработки металлов давлением, которые не описываются существующими теориями. Идея построения состоит в обобщении модели нелинейной упругости на упругопластичность на основе фундаментального закона упругости Мурнагана [4, 5]. При применении менее сложного закона упругости в данном подходе эффективность модели резко снижается. Несмотря на простоту идеи, она обладает новизной и является нестандартной. Используются принцип сохранения потенциальной природы упругой деформации (существования потенциалов напряжений и их скоростей) и принцип попеременного чередования упругого и пластического состояний элементов деформируемого твердого тела. В пластическом состоянии материал предполагается несжимаемым. Требование потенциальности в скоростях напряжений с учетом вида дифференциального определяющего уравнения позволяет определить кусочно-вогнутые девиаторные сечения поверхности текучести. Они получаются путем соединения в сингулярных точках представителей двух семейств вогнутых поверхностей, отвечающих за процессы растяжения и сжатия. При этом просто учитывается эффект Баушингера. Поверхность текучести образуется своими девиаторными сечениями с учетом экспериментальных данных. Понятие пластической деформации не является рабочим и в общем случае не определяется, поскольку в условиях геометрической нелинейности однозначно это сделать невозможно [6]. Поэтому для описания процесса течения вводятся два независимых параметра упрочнения по типу классического параметра Р. Шмидта, совпадающих с ним в частных случаях. При однородной деформации пластическую деформацию можно определить путем полной разгрузки материала.

В работе [3] естественным образом описываются вызывающие затруднения при моделировании следующие экспериментальные факты: одноосное растяжение больше действует в смысле упрочнения на последующее растяжение в ортогональном направлении и меньше – на последующее сжатие; одноосное сжатие меньше действует в смысле упрочнения на последующее сжатие в ортогональном направлении и больше – на последующее растяжение. В [7] описывается явление увеличения пластичности материала под действием дополнительного всестороннего сжатия. Требование существования потенциалов напряжений позволяет описать важный экспериментальный факт роста анизотропии при пластической деформации. Ниже будут рассматриваться возможности описания необычного с точки зрения механики материалов явления «запирания» области высокого давления, связанного с переходом материала от активного нагружения к пассивному – разгрузкой [8]. Так, при сжатии тонкого диска на наковальне Бриджмена [9], начиная с некоторой величины усилия, в центре диска образуется и расширяется упругая область, и никаким сколь угодно большим давлением не удастся изменить остаточную толщину. В первую очередь приведем необходимые соотношения и рассмотрим вопрос о возможном учете условия непрерывности материала при разгрузке в данной модели для ее небольшого уточнения.

1. Модель нелинейно упругопластического материала

Геометрически нелинейная модель упругости [1–3] обобщается на геометрически нелинейную модель упругопластичности. Транспонированный градиент деформации $\overset{0}{\nabla} \mathbf{R}^T$ в упругом законе заменяется неособенным тензором \mathbf{F}_e , который при течении находится из решения системы определяющих уравнений. Удельная потенциальная энергия упругой деформации представляется в виде

$$\begin{aligned} \vartheta = \vartheta_0 + \vartheta_2 + \vartheta_3 + c, \quad \vartheta_0 = 4^{-1}(4^{-1}(-12\lambda - 8\mu + 9v_1 + 18v_2 + 8v_3)I_1 + 4^{-1}(2\lambda + 4\mu - 3v_1 - 10v_2 - 8v_3)I_1^2 + \\ + (-2\mu + 3v_2 + 4v_3)I_2 - (v_2 + 2v_3)I_1I_2 + 12^{-1}(v_1 + 6v_2 + 8v_3)I_1^3 + 2v_3I_3), \end{aligned} \quad (1)$$

где ϑ_0 – изотропный потенциал; $\vartheta_2 = \delta_1(\mathbf{c}_1 \cdot \mathbf{C} \cdot \mathbf{c}_1)^2 + \dots$; $\vartheta_3 = \delta_{22}(\mathbf{c}_1 \cdot \mathbf{C} \cdot \mathbf{c}_1)^3 + \dots$ – анизотропные структуры второй и третьей степени по компонентам тензора упругой деформации Коши – Грина \mathbf{C} ; c – минимальная постоянная, обеспечивающая условие $\vartheta \geq 0$, начальное ее значение $c = 8^{-1}(9\lambda + 6\mu - 2^{-1}9v_1 - 9v_2 - 4v_3)$; λ, μ – постоянные Ляме второго и v_1, v_2, v_3 – третьего порядков; \mathbf{G} – мера упругой деформации Коши – Грина ($\mathbf{G} = 2\mathbf{C} + \mathbf{E}$); I_1, I_2, I_3 – первый, второй и третий главные инварианты меры $\mathbf{G} = \mathbf{F}_e^T \cdot \mathbf{F}_e$ и меры упругой деформации Фингера $\mathbf{F} = \mathbf{F}_e \cdot \mathbf{F}_e^T$, $\mathbf{F} = \mathbf{V}^2$; \mathbf{V} – мера упругих искажений. Для изотропного материала все параметры анизотропии являются нулевыми.

Из (1) получаем соотношение для тензора напряжений Коши:

$$\begin{aligned} \mathbf{T} = 2(\sqrt{I_3})^{-1} \mathbf{F}_e \cdot \frac{\partial \vartheta}{\partial \mathbf{G}} \cdot \mathbf{F}_e^T = \mathbf{T}_0 + \sum_j \delta_j \mathbf{T}_j, \\ \mathbf{T}_0 = 2(\sqrt{I_3})^{-1}(\varphi_0 \mathbf{E} + \varphi_1 \mathbf{F} + \varphi_2 \mathbf{F}^2)(\varphi_0 = a_0 I_3, \varphi_1 = b_0 + b_1 I_1 + b_2 I_1^2 + b_3 I_2, \varphi_2 = c_0 + c_1 I_1), \quad (2) \\ a_0 = 2^{-1}v_3, \quad b_0 = 16^{-1}(-12\lambda - 8\mu + 9v_1 + 18v_2 + 8v_3), \quad b_1 = 8^{-1}(2\lambda - 3v_1 - 4v_2), \\ b_2 = 16^{-1}(v_1 + 2v_2), \quad b_3 = -4^{-1}(v_2 + 2v_3), \quad c_0 = 4^{-1}(2\mu - 3v_2 - 4v_3), \quad c_1 = -b_3. \end{aligned}$$

Явный вид скаляров ϑ_2, ϑ_3 и тензоров \mathbf{T}_j для ортотропного материала приведен в [10]. Уравнения в конечном виде для потенциала напряжений (1) и тензора напряжений (2) справедливы для упругого и пластического состояний материала.

Введем критериальный девиатор-оператор, через который будет записываться критерий текучести. Обозначим: $\overset{\Omega}{\mathbf{T}} = \mathbf{T} - \mathbf{\Omega} \cdot \mathbf{T} + \mathbf{T} \cdot \mathbf{\Omega}$ ($\mathbf{\Omega} = \mathbf{O}^T \cdot \mathbf{O}$ – тензор упругого спина [11], точка над символом означает материальную производную) – объективная О-производная тензора \mathbf{T} ; \mathbf{D} – тензор скорости деформаций; \mathbf{W} – тензор вихря; $\nabla \mathbf{v}$, $\nabla \mathbf{v}^T$ – градиенты скорости. Определим девиатор-оператор $\mathbf{Q} = \mathbf{Q}(\mathbf{D})$ как О-производную девиатора $\text{dev} \mathbf{T}$, вычисленную при условии несжимаемости $\nabla \cdot \mathbf{v} = 0$ по соотношению нелинейной упругости $\dot{\mathbf{F}}_e = \nabla \mathbf{v}^T \cdot \mathbf{F}_e$ [4]. Последнее уравнение задает тензор \mathbf{F}_e в упругом состоянии.

Вычислим материальную производную тензора \mathbf{T} . Имеем

$$\begin{aligned} \dot{\mathbf{G}} = (\mathbf{F}_e^T \cdot \dot{\mathbf{F}}_e) = \mathbf{F}_e^T \cdot (\nabla \mathbf{v}^T + \nabla \mathbf{v}) \cdot \mathbf{F}_e = 2\mathbf{F}_e^T \cdot \mathbf{D} \cdot \mathbf{F}_e, \quad \dot{\mathbf{F}} = (\mathbf{F}_e \cdot \dot{\mathbf{F}}_e^T) = \nabla \mathbf{v}^T \cdot \mathbf{F} + \mathbf{F} \cdot \nabla \mathbf{v}, \\ \dot{I}_1 = \mathbf{E} \cdot \dot{\mathbf{F}} = 2\mathbf{F} \cdot \mathbf{D}, \quad \dot{I}_2 = (\mathbf{E} I_1 - \mathbf{F}) \cdot \dot{\mathbf{F}} = 2(I_1 \mathbf{F} \cdot \mathbf{D} - \mathbf{F}^2 \cdot \mathbf{D}), \quad \dot{I}_3 = I_3 \mathbf{F}^{-1} \cdot \dot{\mathbf{F}} = 2I_3 \nabla \cdot \mathbf{v}, \end{aligned}$$

$$((\sqrt{I_3})^{-1})' = -2^{-1}((\sqrt{I_3})^{-3})2I_3\nabla \cdot \mathbf{v} = -(\sqrt{I_3})^{-1}\nabla \cdot \mathbf{v},$$

$$(\varphi_0 \mathbf{E} + \varphi_1 \mathbf{F} + \varphi_2 \mathbf{F}^2)' = \dot{\varphi}_0 \mathbf{E} + \dot{\varphi}_1 \mathbf{F} + \dot{\varphi}_2 \mathbf{F}^2 + \varphi_1 (\nabla \mathbf{v}^T \cdot \mathbf{F} + \mathbf{F} \cdot \nabla \mathbf{v}) + \varphi_2 (\nabla \mathbf{v}^T \cdot \mathbf{F}^2 + \mathbf{F}^2 \cdot \nabla \mathbf{v} + \mathbf{F} \cdot \mathbf{D} \cdot \mathbf{F}),$$

$$\dot{\varphi}_1 = (b_0 + b_1 I_1 + b_2 I_1^2 + b_3 I_2)' = 2(d \mathbf{F} \cdot \mathbf{D} + c_1 \mathbf{F}^2 \cdot \mathbf{D}) \quad (d = b_1 + (2b_2 - c_1)I_1),$$

$$\dot{\varphi}_0 = (a_0 I_3)' = 2a_0 I_3 \nabla \cdot \mathbf{v} = 0, \quad \dot{\varphi}_2 = (c_0 + c_1 I_1)' = 2c_1 \mathbf{F} \cdot \mathbf{D},$$

$$\begin{aligned} \mathbf{T} &= \mathbf{T}_0 + \sum_j \delta_j \mathbf{S}_j, \quad \mathbf{T}_j = 2(\sqrt{I_3})^{-1} \mathbf{S}_j, \quad \dot{\mathbf{T}}_j = -\nabla \cdot \mathbf{v} \mathbf{T}_j + \nabla \mathbf{v}^T \cdot \mathbf{T}_j + \mathbf{T}_j \cdot \nabla \mathbf{v} + \mathbf{S}_j, \quad \dot{\mathbf{T}}_0 = -\nabla \cdot \mathbf{v} \mathbf{T}_0 + \\ &+ \mathbf{T}_0 \cdot \nabla \mathbf{v} + \nabla \mathbf{v}^T \cdot \mathbf{T}_0 + 4L_3^{-1}(a_0 I_3 \nabla \cdot \mathbf{v} \mathbf{E} + (d \mathbf{F} \cdot \mathbf{D} + c_1 \mathbf{F}^2 \cdot \mathbf{D}) \mathbf{F} + c_1 \mathbf{F} \cdot \mathbf{D} \mathbf{F}^2 - \varphi_0 \mathbf{D} + \varphi_2 \mathbf{F} \cdot \mathbf{D} \cdot \mathbf{F}). \end{aligned} \quad (3)$$

Учитывая, что $\nabla \mathbf{v} = \mathbf{D} - \mathbf{W}$, $\nabla \mathbf{v}^T = \mathbf{D} + \mathbf{W}$, и используя (2), из (3) получим

$$\begin{aligned} \mathbf{Q}(\mathbf{D}) &= (\mathbf{W} - \mathbf{\Omega}) \cdot \mathbf{T} - \mathbf{T} \cdot (\mathbf{W} - \mathbf{\Omega}) + \mathbf{Q}_0(\mathbf{D}) + \mathbf{Q}_1(\mathbf{D}) + \mathbf{Q}_2(\mathbf{D}), \quad \mathbf{Q}_0 = \text{dev} 4I_3^{-\frac{1}{2}} ((d \mathbf{F} \cdot \mathbf{D} + \\ &+ c_1 \mathbf{F}^2 \cdot \mathbf{D}) \mathbf{F} + c_1 \mathbf{F} \cdot \mathbf{D} \mathbf{F}^2 + \varphi_1 2^{-1} (\mathbf{F} \cdot \mathbf{D} + \mathbf{D} \cdot \mathbf{F}) + \varphi_2 (2^{-1} (\mathbf{F}^2 \cdot \mathbf{D} + \mathbf{D} \cdot \mathbf{F}^2) + \mathbf{F} \cdot \mathbf{D} \cdot \mathbf{F})). \end{aligned} \quad (4)$$

Соотношения для тензоров \mathbf{S}_j , $\mathbf{Q}_1(\mathbf{D})$, $\mathbf{Q}_2(\mathbf{D})$ в (3) и (4) являются громоздкими и представлены в работе [9].

Критерий текучести в точках поверхности текучести запишем как

$$\mathbf{Q} \cdot \mathbf{N} \geq 0 \quad (\mathbf{D} \neq 0). \quad (5)$$

Переход от активного нагружения к пассивному в пластическом состоянии происходит при нарушении неравенства (5). Ранее в (5) использовалось строгое неравенство [1–3], поэтому ограничение $\mathbf{D} \neq 0$ было излишним.

О п р е д е л е н и е 1. Если в соотношении (5) выполняется $\mathbf{Q} \cdot \mathbf{N} = 0$, нагружение назовем нейтральным.

В силу положительной определенности оператора $\mathbf{Q} \cdot \mathbf{D}$ неравенство (5) можно заменить неравенством $\mathbf{D} \cdot \mathbf{N} \geq 0$ и равенство $\mathbf{Q} \cdot \mathbf{N} = 0$ равносильно равенству $\mathbf{D} \cdot \mathbf{N} = 0$. При векторной интерпретации симметричных тензоров второго ранга направления векторов \mathbf{Q} , \mathbf{D} являются близкими.

Уравнения в дифференциальном виде для напряжений, потенциала напряжений и параметров анизотропии при течении в регулярной точке поверхности текучести получены в [1, 2]:

$$\begin{aligned} \dot{\mathbf{T}} &= K(\mathbf{Q} - \mathbf{Q} \cdot \mathbf{N} \mathbf{N}), \quad ((\sqrt{I_3})^{-1})' = P \mathbf{T} \cdot \mathbf{D}, \quad \dot{\delta}_j = k_j \beta \mathbf{N} \cdot \mathbf{T}_j (\sqrt{\mathbf{T}_j \cdot \mathbf{T}_j})^{-1} \\ &(\mathbf{T}_j \neq 0, \quad \beta \geq 0, \quad \beta \rightarrow \min, \quad k_j = \pm 1), \end{aligned} \quad (6)$$

где \mathbf{N} – вектор внешней нормали к поверхности девиаторного сечения ($\mathbf{N} \cdot \mathbf{N} = 1$). Он выбирается из двух собственных векторов оператора $\mathbf{Q}(\mathbf{D})$. Скаляр β характеризует скорость роста анизотропии. При значении $\beta = 0$ параметры анизотропии не изменяются.

В уравнениях (6) возможны два случая. В первом (основном) случае величина удельной мощности деформации является положительной: $\mathbf{T} \cdot \mathbf{D} > 0$, во втором – неположительной: $\mathbf{T} \cdot \mathbf{D} \leq 0$. В зависимости от этого задаются скаляры K и P .

В первом случае не зависящий от девиатора \mathbf{D} малый скаляр $K > 0$ определяется в экспериментах по двухосному растяжению ($K = K_1$) и по двухосному сжатию ($K = K_2$):

$$K = K_i, \quad P = 1 - \alpha_i \quad (i = 1, 2). \quad (7)$$

Близкая к единице величина α_i определяется в одноосных растяжении и сжатии и имеет смысл относительной части рассеиваемой работы деформации.

Во втором случае полагаем

$$K = K_0 + \theta K_i, \quad P = 1, \quad \theta = \mathbf{D} \cdot \mathbf{N} / \sqrt{\mathbf{D} \cdot \mathbf{D}} \sqrt{1 - (\text{dev } \mathbf{T} \cdot \mathbf{N})^2 / \text{dev } \mathbf{T} \cdot \text{dev } \mathbf{T}}. \quad (8)$$

Значения скаляра θ равномерно изменяются в промежутке $0 \leq \theta \leq 1$ при значениях величины скорости деформаций от $\mathbf{D} \cdot \mathbf{N} = 0$ до $\mathbf{D} \cdot \mathbf{T} = 0$. Ранее в работе [1] выбиралось $\theta = 1$. Скаляр K_0 определяется следующим образом: надо положить в (6) $P = 1, \beta = 0$, затем из системы уравнений (1), (2), (6) для неизвестных $\overset{\Omega}{\mathbf{V}}, K_0$ определить значение $K_0 = 0$. Для изотропного материала можно выбирать в качестве неизвестных $\overset{\Omega}{\mathbf{F}}, K_0$. При $\mathbf{D} \cdot \mathbf{T} = 0$ получаем $K_0 = 0$. Теряется потенциальность в скоростях напряжений, но первое уравнение в (6) остается однородным первой степени относительно компонентов тензора \mathbf{D} . Необходимое условие по Р. Хиллу для определяющего уравнения теории пластичности выполняется.

Неубывающий параметр упрочнения Р. Шмидта определяется как $\dot{\chi} = \mathbf{T} \cdot \mathbf{D}$, где под тензором \mathbf{D} понимается тензор скорости пластических деформаций. Скалярная величина в правой части последнего уравнения является неотрицательной. Начальное значение для изотропного материала $\chi = 0$. В локальном базисе собственных векторов критериального оператора справедливы соотношения $\text{dev } \mathbf{T} = \mathbf{T} \cdot \mathbf{N}_i \mathbf{N}_i + \dots, \mathbf{D} = \mathbf{D} \cdot \mathbf{N}_i \mathbf{N}_i + \dots$, где \mathbf{N}_i – вектор нормали к поверхности девиаторного сечения. Учитывая, что выполняются неравенства $\mathbf{T} \cdot \mathbf{N}_i > 0, \mathbf{D} \cdot \mathbf{N}_i \geq 0$, два неотрицательных и неубывающих параметра упрочнения можно задать следующим образом:

$$\dot{\chi}_i = \mathbf{T} \cdot \mathbf{N}_i \mathbf{D} \cdot \mathbf{N}_i, \quad \dot{\chi}_j = 0 \quad (i, j = 1, 2; i \neq j). \quad (9)$$

Начальные значения будут $\chi_1 = 0, \chi_2 = 0$. При одноосных нагружениях активный параметр упрочнения, который соответствует вектору нормали \mathbf{N}_i , в силу пропорциональности девиаторов $\text{dev } \mathbf{T}, \mathbf{D}, \mathbf{N}_i$ совпадает с параметром Р. Шмидта.

2. Условие непрерывности при переходе от пластического состояния к упругому

В линейной теории пластичности с поверхностью текучести существует понятие нейтрального нагружения. Если приращение вектора напряжений направлено в наружную сторону поверхности текучести, оно сопровождается пластической деформацией и называется активным. Если приращение направлено внутрь объема, ограниченного данной поверхностью, нагружение называют пассивным, или разгрузкой. Промежуточный случай называют нейтральным нагружением. В этом случае должна отсутствовать пластическая деформация. Так, в деформационной теории ее наличие явилось серьезным дефектом.

В рассматриваемой нелинейной модели материала при течении согласно этому определению все нагружения являются нейтральными и должны сопровождаться пластической дефор-

мацией за исключением предельного случая. Поэтому будем использовать указанную выше формулировку нейтрального нагружения. Условие непрерывности можно сформулировать следующим образом.

Определение 2. Пусть в соотношении (5) выполняется $\mathbf{Q} \cdot \mathbf{N} \rightarrow 0$. Следовательно, выполняется и $\mathbf{D} \cdot \mathbf{N} \rightarrow 0$. Тогда величина прироста пластической деформации должна стремиться к нулю и должно иметь место затухание роста параметров анизотропии, т. е. в (6) $\beta \rightarrow 0$.

Как будет показано ниже, полностью устранить появление пластических деформаций не удастся в силу особенностей модели. Для пояснения дальнейших рассуждений обратимся к примеру [1, 3].

Пример 1. Рассмотрим ортотропный идеально упругопластический материал ($\mathbf{\Omega} = \mathbf{W} = 0$), который первоначально является изотропным (рис. 1, а). Используем данные по вольфраму. Рассчитывалось нагружение типа растяжение и сжатие по трем осям, когда удельная мощность деформации становится неположительной. Вектор \mathbf{D} направлен в створ угла, обозначенного двумя дугами. На начальной кривой пластичности, соответствующей одноосному растяжению, отрезком указана точка, в которой находится точка процесса (рис. 1). В эксперименте это соответствует растяжению по первой, второй оси и сжатию по третьей оси с искажениями 1,0001, 1,00136 и 0,99913. При течении происходило перемещение точки процесса по кривой пластичности влево в точку, соответствующую ее положению в одноосном растяжении, при малом приросте анизотропии. Согласно (9) значения параметров упрочнения остаются нулевыми ($\chi_1 = 0$, $\chi_2 = 0$) при нейтральном нагружении.

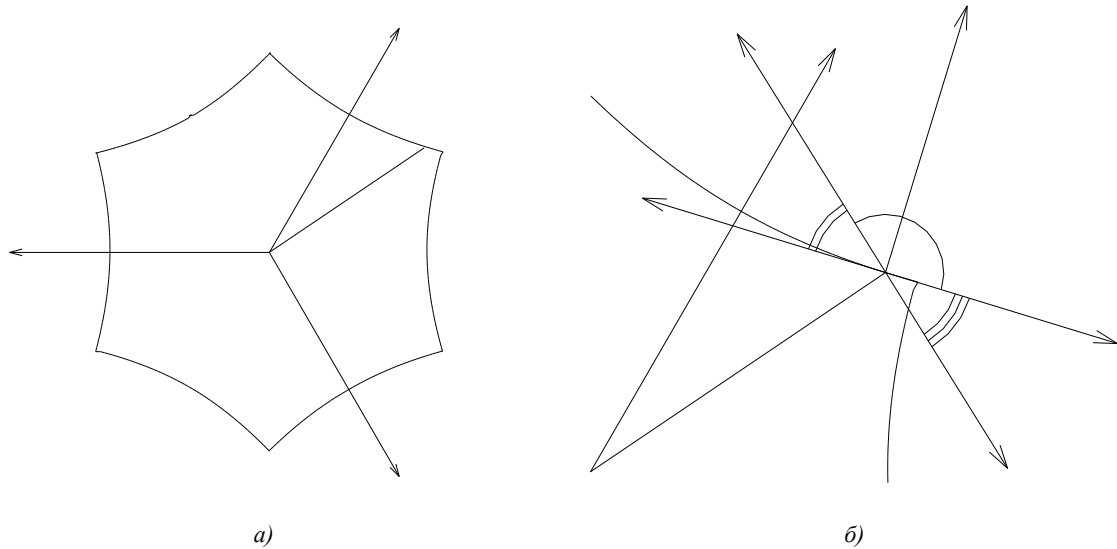


Рис. 1. Кривая пластичности (а) и ее увеличенный участок (б)

Далее, вычисляя из (2) \mathbf{T} , а из (1) $\dot{\epsilon}$, получаем систему одного тензорного и одного скалярного уравнений для определения \mathbf{F} и K_0 :

$$\begin{aligned}
 & -2^{-1} \mathbf{F}^{-1} \cdot \mathbf{F} \mathbf{T} + 2\sqrt{I_3} (\varphi_0 \mathbf{F}^{-1} \cdot \mathbf{F} \mathbf{E} + (d \mathbf{E} \cdot \mathbf{F} + c_1 \mathbf{F} \cdot \mathbf{F}) \mathbf{F} + c_1 \mathbf{E} \cdot \mathbf{F} \mathbf{F}^2 + \varphi_1 \mathbf{F} + \\
 & + \varphi_2 (\mathbf{F} \cdot \mathbf{F} + \mathbf{F} \cdot \mathbf{F})) - K_0 \mathbf{Q} = 0, \quad (\sqrt{I_3})^{-1} (\varphi_1 \mathbf{E} \cdot \mathbf{F} + \varphi_2 \mathbf{F} \cdot \mathbf{F} + (\varphi_0 - 2^{-1} \epsilon_0) \mathbf{F}^{-1} \cdot \mathbf{F}) = \mathbf{D} \cdot \mathbf{T}.
 \end{aligned} \tag{10}$$

Умножаем в (10) первое уравнение на $2^{-1} \sqrt{I_3}$, второе – на $\sqrt{I_3}$. Пусть $\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3$ – ортонормированный триэдр главных осей анизотропии.

При течении выполняется условие несжимаемости в скоростях $\mathbf{D} \cdot \mathbf{E} = 0$. В упругом состоянии для ортотропного материала справедливо соотношение $\dot{\mathbf{F}} = \mathbf{D} \cdot \mathbf{F} + \mathbf{F} \cdot \mathbf{D}$. Вероятно, при нейтральном нагружении должно выполняться условие $K_0 = 1$ в (6), (8) из соображений непрерывности перехода из пластического состояния в упругое.

Следовательно, получаем $\mathbf{F}^{-1} \cdot \dot{\mathbf{F}} = \mathbf{F}^{-1} \cdot (\mathbf{D} \cdot \mathbf{F} + \mathbf{F} \cdot \mathbf{D}) = 2\mathbf{D} \cdot \mathbf{E} = 0$. В первом уравнении (10) девиаторная часть обращается в нуль по определению девиатора \mathbf{Q} . Система сводится к двум скалярным уравнениям

$$A\mathbf{E} \cdot \mathbf{F} + B\mathbf{F} \cdot \mathbf{F} = 0, \quad \varphi_1 \mathbf{E} \cdot \mathbf{F} + \varphi_2 \mathbf{F} \cdot \mathbf{F} = 0 \quad (A = dI_1 + c_1 d_2 + \varphi_1, \quad B = c_1 I_1 + 2\varphi_2). \quad (11)$$

С точностью до малых величин $d'_2 = \mathbf{E} \cdot \text{dev} \mathbf{F}^2$, $I_1 - 3$ справедливы оценки [9]

$$I_1 = 3, \quad d_2 = \mathbf{E} \cdot \mathbf{F}^2 = 3, \quad I_3 = 1, \quad c_1 = 4^{-1}(\nu_2 + 2\nu_3), \quad d = 4^{-1}(\lambda - 2\nu_2 - 6\nu_3), \\ \varphi_1 = -2^{-1}\mu - \nu_3, \quad \varphi_2 = 2^{-1}(\mu + \nu_3).$$

Находим определитель системы (11):

$$\Delta = A\varphi_2 - B\varphi_1 = 8^{-1}(3\lambda\mu + 2\mu^2 + 3\lambda\nu_3 + 3\nu_2\nu_3 + 4\nu_3^2).$$

Согласно данным по вольфраму находим $\frac{\Delta}{\mu^2} = 1,76$. Таким образом, решения системы (11) будут только нулевыми. Переходим к системе

$$\mathbf{F}^{-1} \cdot \dot{\mathbf{F}} = 0, \quad \mathbf{E} \cdot \mathbf{F} = 0, \quad \mathbf{F} \cdot \mathbf{F} = 0 \quad (12)$$

относительно тензора скорости деформаций. Определитель системы уравнений (12) будет нулевым тогда и только тогда, когда выполняется по крайней мере одно из соотношений $\mathbf{F} \cdot \mathbf{c}_i \mathbf{c}_i = \mathbf{F} \cdot \mathbf{c}_j \mathbf{c}_j$ ($i, j = 1, 2, 3; i \neq j$). Это условие равносильно выполнению соотношений $\text{dev} \mathbf{T} \cdot \mathbf{c}_i \mathbf{c}_i = \text{dev} \mathbf{T} \cdot \mathbf{c}_j \mathbf{c}_j$. Следовательно, искомая точка должна находиться на пересечении кривой пластичности и проекций (с учетом ее продолжения) базисных диад $\mathbf{c}_j \mathbf{c}_j$, которые показаны стрелками на рис. 1, а. Однако в данных точках имеет место условие положительности удельной мощности деформации при течении. Значит, рассматриваемый случай не реализуется. Заданное перемещение по кривой пластичности при пассивном нагружении возможно только при условии $\mathbf{D} \cdot \mathbf{E} \neq 0$. Поэтому в расчетах не достигается значение $K_0 = 1$, а получается величина $K_0 = 1 \pm \varepsilon$, где согласно расчетам $\varepsilon = 0,0013$. Наблюдается расхождение величин скоростей общих и упругих искажений порядка 0,1 %. Поэтому неизбежно накапливается пренебрежительно малая величина пластических деформаций.

В первом случае при течении (6) соблюдение условия непрерывности представляется искусственным. В третьем случае (в сингулярной точке) [1] перейти к разгрузке можно только тривиальным образом при условии $\mathbf{D} = 0$. В упругом состоянии непрерывность перехода соблюдается. Разгрузка происходит при перемещении точки процесса внутрь области, ограниченной поверхностью текучести.

Отметим, что обеспечение непрерывности не является целью исследования, гораздо важнее создать условия для описания проблемных течений, в частности таких, при которых имеет место явление «запирания» области высокого давления. Важным обстоятельством при этом является кусочная вогнутость полученной из обоснованных предположений поверхности текучести. Если вектор \mathbf{Q} направлен в створ угла, обозначенного на рис. 1 тремя дугами, то происходит разгрузка. Именно такой случай имеет место в следующем примере.

3. Возможности описания явления «запирания» области высокого давления

Пример 2. Рассмотрим модельный пример, заключающийся в наложении нагружения сжатия по третьей оси на нагружение простого сдвига по первым двум осям [12]. Расчет процесса простого сдвига проведен для вольфрама в предположении идеальной упругопластичности материала. В промежуточном состоянии, близком к конечному, по третьей оси прикладывается нагружение сжатия при помощи гладких подвижного и неподвижного штампов. Тензор скорости деформаций при этом имеет вид $\mathbf{D} = -3^{-1} \dot{p} \mathbf{c}_1 \mathbf{c}_1 - 3^{-1} \dot{p} \mathbf{c}_2 \mathbf{c}_2 + 3^{-1} 2 \dot{p} \mathbf{c}_3 \mathbf{c}_3 + \dot{s} 2^{-1} (\mathbf{c}_1 \mathbf{c}_2 + \mathbf{c}_2 \mathbf{c}_1)$, где \dot{p} – скорость подвижного штампа, \dot{s} – скорость изменения параметра сдвига. В результате вычислительных экспериментов установлено, что при условии $\dot{p} = -k\dot{s}$, $k \geq 0,6$, имеет место нарушение критерия текучести, приводящее к разгрузке. Оно реализуется, как указано в конце разд. 2. При этом величина удельной мощности деформации является положительной (см. рис. 1, б). Расчет модельного примера 2 также выполнен с использованием системы LS-DYNA для материала Баммана [13]. Установлено, что разгрузка не реализуется.

Вероятно, явление «запирания» области высокого давления будет описываться таким же образом. Численная реализация рассматриваемой модели материала для решения краевой задачи расчета процесса сжатия с кручением тонкого диска на наковальне Бриджмена пока не разработана. Используем систему LS-DYNA для модели Баммана. В качестве материала возьмем свинец, необходимые данные для расчетов получены в работе [14]. Для наглядного представления процесса осадки с кручением первоначально рассмотрим нагружение толстого диска.

Пример 3. Осадка круглого диска производится двумя бойками, нижний из которых с некоторого момента времени начинает вращаться относительно верхнего. Ось вращения бойка совпадает с осью осаживаемой заготовки, трение между ними принималось максимальное, т. е. выбрано условие прилипания. Размеры дисков составляли: радиус $R = 0,1$ м, высота $H = 0,1$ м и $H = 0,005$ м. Модель пластичности Баммана реализована в системе LS-DYNA в виде материала MAT_BAMMAN [15]. Параметры материала приведены в таблице.

Параметры материала (свинца) для модели Баммана

Параметр	Обозначение	Значение
Плотность	RO	11 340 кг/м ³
Модуль Юнга	E	1,6000E+10 Па
Коэффициент Пуассона	PR	0,4
Параметры модели Баммана	c1	3,9000E+5 Па
	c3	1,0000E+7 Па
	c5	0,002
	c7	5,8390E-7 Па
	c9	6,7460E+7 Па
	c15	1,1340E+7 Па
	c2, c9, c4, c6, c8...	0

Угловая скорость вращения нижнего бойка взята 0,314 1/с, общая степень обжатия составляла 10 % от высоты заготовки, отношение скорости вращения к относительной скорости осадки получалось 31,4. Вращение нижнего бойка начинается в момент, когда осадка выполнена на половину общего обжатия. Оба бойка рассматривались как абсолютно твердые тела. Геометрия заготовки моделировалась с помощью объемных элементов с равномерным распределением напряжения в форме гексаэдров. Задача решалась в неявной постановке с автоматическим выбором шага и нелинейным методом решения. Для учета больших деформаций и связанных с ними больших перемещений узлов задавалось подавление искажения формы элементов в картах HOURGLASS и CONTROL_HOURGLASS, благодаря чему программа более корректно рассчитывает усилия скручивания и снижает вероятность критического искажения форм элементов.

Результаты моделирования толстого диска приведены на рис. 2 и 3. Изображена половина диска с двух точек зрения (рис. 2). Наблюдается бочкообразование при осадке и сильное равномерное скручивание элементов. На рис. 3 показана зависимость усилия осадки от времени. В момент времени $t = 5$ с (когда начинается кручение заготовки) усилие осадки резко падает. Этот факт соответствует опытным данным [16]. К сожалению, в последней работе имеются некоторые математические неточности.

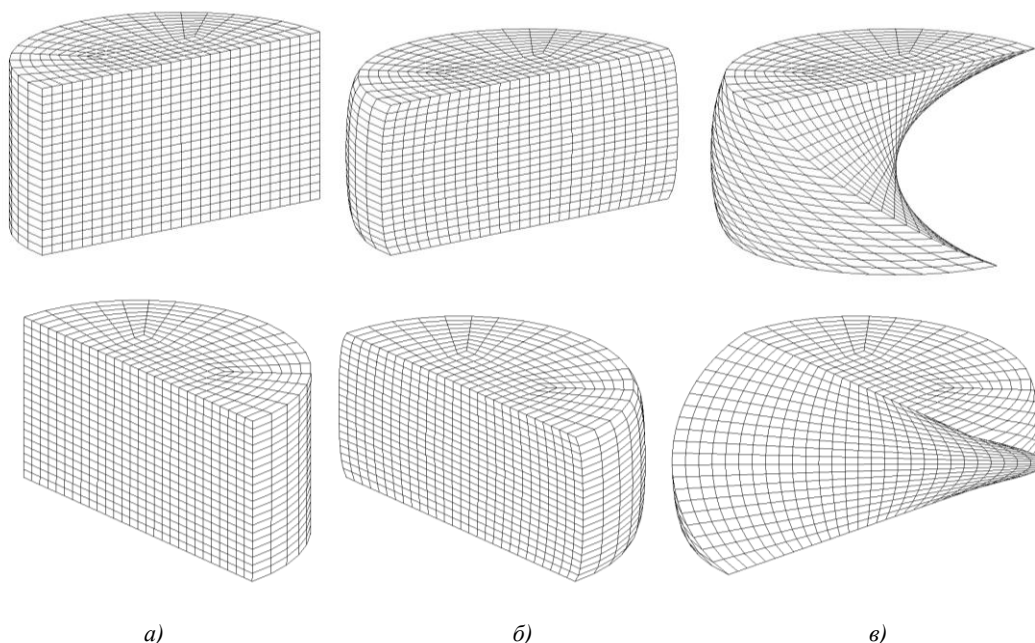


Рис. 2. Результаты моделирования формоизменения толстого диска: а) исходное состояние; б) после осадки на 5 % без кручения; в) после осадки на 10 % с поворотом бойка на 90°

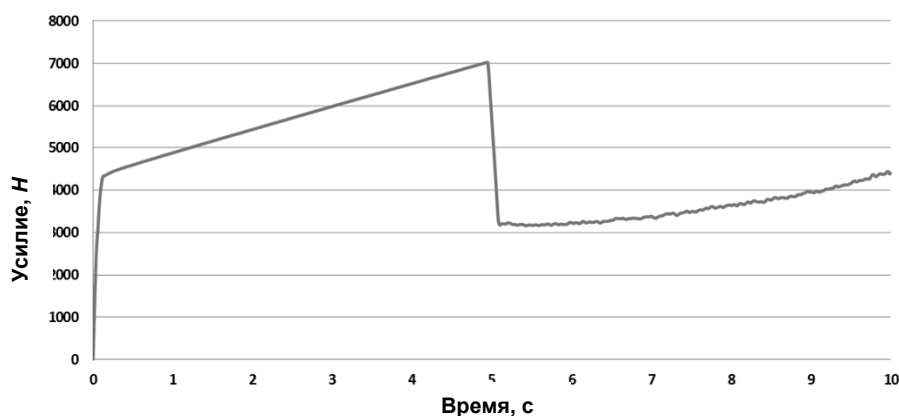


Рис. 3. Зависимость усилия осадки от времени

Пример 4. На рис. 4 показаны результаты вычислительного эксперимента нагружения тонкого диска на наковальне Бриджмена. Как и следовало ожидать, явление разгрузки в объеме диска не наблюдается и явление «запирания» области высокого давления не описывается.

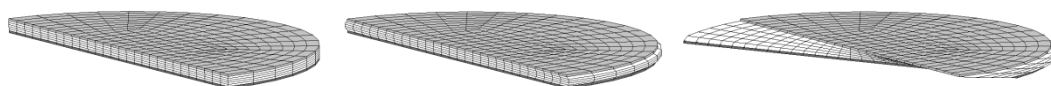


Рис. 4. Результаты моделирования формоизменения тонкого диска

Заключение

Сформулировано условие непрерывности при переходе от активного нагружения к пассивному. Установлено, что оно может быть обеспечено при состоянии течения в случае отрицательности удельной мощности деформации. Рассмотрен модельный пример для ортотропного материала, соответствующий этому случаю. В предельном состоянии задания скорости деформаций при указанном переходе параметры анизотропии остаются постоянными, обобщенный упругий закон не изменяется, однако наблюдается пренебрежительно малый рост пластической деформации. Причина состоит в том, что перемещение по кривой пластичности в упругом состоянии материала возможно только при сжимаемых полях скоростей перемещений. В активном нагружении поля скоростей перемещений являются несжимаемыми, что приводит к появлению малой пластической деформации.

В случае положительной удельной мощности деформации непрерывность при переходе отсутствует и не может быть соблюдена естественным образом. Исследовались возможности описания для этого случая явления «запирания» области высокого давления. Основанием для такого предположения является невыпуклость полученной поверхности текучести. Рассмотрен соответствующий модельный пример. С помощью системы LS-DYNA показано, что в данном случае разгрузка не происходит.

Для материала Баммана выполнено моделирование системой LS-DYNA процесса сжатия тонкого свинцового диска на наковальне Бриджмена. Установлено, что перехода материала из пластического состояния в упругое не наблюдается и явление «запирания» области высокого давления не описывается.

Список литературы

1. Швед, О.Л. Модель нелинейно упругопластического материала / О.Л. Швед // Весті АН Беларусі. Сер. фіз.-мат. навук. – 2014. – № 1. – С. 63–68.
2. Швед, О.Л. О возможных определяющих соотношениях нелинейной упругопластичности / О.Л. Швед // Тр. VII Всерос. конф. по механике деформируемого твердого тела, Ростов-на-Д., 2013. – Ростов-на-Д., 2013. – Т. II. – С. 219–223.
3. Швед, О.Л. Определяющие соотношения ортотропного упругопластического материала / О.Л. Швед // Тр. VII Всерос. конф. по механике деформируемого твердого тела. – Ростов-на-Д., 2013. – Т. II. – С. 224–228.
4. Лурье, А.И. Нелинейная теория упругости / А.И. Лурье. – М. : Наука, 1980. – 512 с.
5. Murnagan, F.D. Finite deformation of an elastic solid / F.D. Murnagan. – N.Y. : Dover, 1967. – 140 p.
6. Naghdi, P.M. A critical review of the state of finite plasticity / P.M. Naghdi // ZAMP. – 1990. – Vol. 41, № 3. – P. 315–394.
7. Швед, О.Л. Численное моделирование эффекта увеличения пластичности металла при растяжении под действием высокого гидростатического давления / О.Л. Швед // Весті АН Беларусі. Сер. фіз.-тэхн. навук. – 2014. – № 4. – С. 18–23.
8. Левитас, В.И. Большие упругопластические деформации материалов при высоком давлении / В.И. Левитас. – Киев : Наукова думка, 1987. – 232 с.
9. Бриджмен, П. Исследование больших пластических деформаций и разрыва / П. Бриджмен. – М. : Изд-во иностр. лит., 1955. – 444 с.
10. Швед, О.Л. Определение девиаторного сечения поверхности текучести при математическом моделировании упругопластического поведения материалов / О.Л. Швед // Информатика. – 2014. – № 2(41). – С. 49–57.
11. Швед, О.Л. Определение тензора упругого спина в нелинейной теории пластичности / О.Л. Швед // Весті АН Беларусі. Сер. фіз.-мат. навук. – 2009. – № 1. – С. 52–58.
12. Поздеев, А.А. Большие упругопластические деформации: теория, алгоритмы, приложения / А.А. Поздеев, П.В. Трусов, Ю.И. Няшин. – М. : Наука, 1986. – 232 с.

13. Bammann, D.J. A model for finite deformation plasticity / D.J. Bammann, E.C. Aifantis // *Acta Mechanica*. – 1987. – Vol. 69. – P. 97–117.

14. Швед, О.Л. Исследование эффекта двойного бочкообразования при обработке металлов давлением с использованием системы LS-DYNA / О.Л. Швед, А.А. Абрамов // *Информатика*. – 2009. – № 1(21). – С. 17–24.

15. LS-DYNA: keyword user's manual. – Livermore : Livermore Software Technology Corporation, 2001. – Vol. 3. – 1421 p.

16. Воронцов, А.Л. Теория осадки с кручением заготовки / А.Л. Воронцов // *Кузнечно-штамповочное производство. Обработка металлов давлением*. – 2007. – № 6. – С. 3–9.

Поступила 05.06.2015

*Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: swed@newman.bas-net.by*

O.L. Shwed, A.A. Abramow

TRANSITION FROM ACTIVE TO PASSIVE WEIGHTNING IN NONLINEAR ELASTO-PLASTIC PROCESS

A model of nonlinear elasto-plastic material is proposed. Continuity condition is formulated in transition from an active to passive weighting, which is provided in a regular point of the flow surface with a non-positive specific power of deformation. If the latter condition is not satisfied, there is no transition continuity. In this case, possibility of description of the phenomenon of «locking» the range of high pressure is considered in transition of material from plastic to elastic state.

УДК 621-235:004:519.8

Н.Н. Гушинский¹, О. Баттайа², А. Долгий³

МОДЕЛИ И МЕТОДЫ СИНТЕЗА СТРУКТУРЫ ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА ОБРАБОТКИ ДЕТАЛЕЙ НА ПОТОЧНОЙ ЛИНИИ ИЗ СТАНКОВ СО СТАЦИОНАРНЫМ ПРИСПОСОБЛЕНИЕМ

Рассматривается задача проектирования поточной линии, комплектуемой из однопозиционных агрегатных станков со стационарным приспособлением, для обработки деталей одного наименования. Предлагаются математические модели и методы синтеза структуры технологического процесса обработки на ней. Критерием качества искомого решения принимается оценка стоимости оборудования линии. Учитываются основные технологические и конструктивные ограничения. Предлагается декомпозиционная схема решения исходной задачи и ее сведения к задаче смешанного линейного программирования. Приводятся результаты вычислительных экспериментов.

Введение

Специальные агрегатные металлорежущие станки [1, 2] широко используются в серийном и крупносерийном производстве, поскольку благодаря высокой степени концентрации операций и уровню автоматизации технологического процесса они обеспечивают высокую производительность и точность обработки при сравнительно небольших требованиях к квалификации станочников-операторов, невысокой себестоимости выпускаемой продукции и занимаемой площади. Важным этапом разработки агрегатного оборудования является аванпроектирование, в процессе которого формируются основные проектные решения (структура и основные параметры технологического процесса обработки детали, компоновки станков и т. п.), во многом определяющие эффективность создаваемого оборудования. Этот этап достаточно трудоемок, связан с решением целого ряда сложных технико-экономических задач и требует привлечения высококвалифицированных специалистов.

Аванпроектирование агрегатного оборудования существенно усложняется при проектировании поточной линии из нескольких станков для обработки детали в случае, когда эта обработка не может быть реализована на одном станке. В работах [3, 4] предложены математическая модель и декомпозиционная схема решения задачи синтеза структуры технологического процесса обработки деталей на поточной линии, комплектуемой из станков со стационарным приспособлением, приспособлением на поворотном столе или на двухпозиционном передвижном столе. Важной особенностью этих станков являлось отсутствие возможности применения параллельно-последовательной схемы совмещения технологических переходов на позициях. Вопросы использования такой схемы для станков с приспособлением на поворотном столе рассматривались в [5].

В настоящей работе возможность применения параллельно-последующей схемы исследуется для станков со стационарным приспособлением.

1. Постановка задачи

Рассматривается задача синтеза структуры технологического процесса (ТП) обработки деталей на поточной линии, состоящей из однопозиционных агрегатных станков со стационарным приспособлением. Предполагается, что обработка на станке может производиться с трех из четырех возможных (сверху, слева, сзади, справа) сторон детали последовательно несколькими блоками инструментов из установленных на соответствующих револьверных бабках (рис. 1 и 2). На рис. 1 обработка сверху производится с помощью револьверной бабки, на которой установлены два блока по два инструмента, обработка слева – блоком из трех инструментов (шпиндельной коробкой), а обработка справа – блоком из одного инструмента (бабкой). На рис. 2 обработка слева производится с помощью двухшпиндельной коробки, а обработка справа и сзади – с помощью револьверных бабок, на которых установлены два блока по одному инструменту.

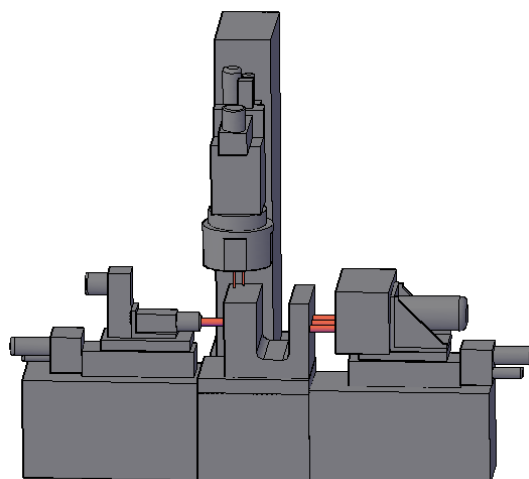


Рис. 1. Станок с вертикальной револьверной бабкой

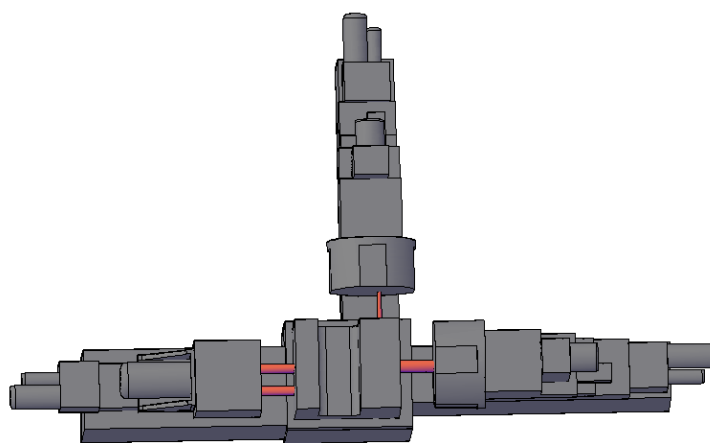


Рис. 2. Станок с двумя горизонтальными револьверными бабками

Предполагается, что заданы множества M_s технологических переходов по обработке элементов, расположенных на s -й стороне детали, $s=1, \dots, n_0$. Множество возможных положений детали задается матрицей \mathbf{H} размерности $r_0 \times n_0$, элементы r -й строки которой $(h_{r1}, h_{r2}, \dots, h_{rn_0})$ определяют способ обработки каждой стороны детали. При этом $h_{rs} \in \{0, 1, 2, 3, 4\}$, $r=1, \dots, r_0$, $s=1, \dots, n_0$, где $h_{rs}=1, 2, 3, 4$, если при ориентации r детали элементы ее стороны s можно обрабатывать сверху, слева, сзади и справа соответственно, а $h_{rs}=0$ означает, что обработка детали со стороны s при этой ориентации детали невозможна.

Положим $M = \bigcup_{s=1}^{n_0} M_s$. Для каждого перехода $u \in M$ определены следующие параметры:

- длина $\lambda(u)$ рабочего хода инструмента;
- сторона $s(u)$ детали, на которой расположен обрабатываемый элемент;
- диапазон $[\gamma_1(u), \gamma_2(u)]$ возможных значений значений минутной подачи;
- множество допустимых $h(u) \subseteq J = \{1, 2, 3, 4\}$ ориентаций инструмента при выполнении перехода u .

Искомыми в рассматриваемой задаче являются следующие параметры проектируемого ТП:

- число m станков в линии;
- положение H_k детали на k -м станке, $k=1, \dots, m$;
- количество b_{kj} блоков инструментов (шпиндельных головок), устанавливаемых на k -м станке для обработки сверху ($j=1$), слева ($j=2$), сзади ($j=3$) и справа ($j=4$), $k=1, \dots, m$;

– множество N_{kjl} переходов из M по обработке детали, выполняемых на k -м станке l -й шпиндельной головкой для обработки сверху ($j=1$), слева ($j=2$), сзади ($j=3$) и справа ($j=4$), $k=1, \dots, m$, $l=1, \dots, b_{kj}$;

– минутная подача $\Gamma_{kjl} \in [\max\{\gamma_1(u) | u \in N_{kjl}\}, \min\{\gamma_2(u) | u \in N_{kjl}\}]$ при выполнении множества N_{kjl} переходов единым блоком инструментов, $k=1, \dots, m$, $j \in J$, $l=1, \dots, b_{kj}$.

Этот набор параметров должен удовлетворять ряду технологических и конструктивных требований, учитывающих:

- возможный порядок выполнения переходов при обработке детали;
- необходимость и (или) невозможность выполнения отдельных групп переходов из M одной шпиндельной коробкой, одной револьверной головкой, в одном станке;
- предельно допустимое число m_0 станков в линии и число b_0 шпиндельных головок, устанавливаемых на револьверной бабке;
- ориентацию детали, принимающую во внимание возможность выполнения каждого из переходов при определенных ориентациях детали на станке;
- режимы обработки, обусловленные как взаимосвязями режимов обработки для переходов, выполняемых от одного силового узла, так и экономически целесообразными диапазонами этих параметров;
- производительность, регламентирующую время работы линии с учетом заданного коэффициента ее загрузки.

Ограничения, регламентирующие возможный порядок выполнения технологических переходов по обработке деталей, определяют отношение строгого частичного порядка на всем множестве переходов M , которое может быть задано орграфом $G=(M, A)$, в котором дуга $(u, u') \in A$ тогда и только тогда, когда переход $u \in M$ должен выполняться перед переходом $u' \in M$. Следует отметить, что если переходы u и u' обрабатывают элементы, расположенные на разных сторонах детали, то переход u должен быть выполнен на одном из предыдущих станков. Если эти переходы выполняют обработку с одной и той же ориентацией инструмента, то переход u может быть выполнен и одной из предыдущих шпиндельных головок.

Регламенты обязательности выполнения отдельных пар переходов из M в одной шпиндельной головке, одной и той же револьверной бабкой и на одном станке могут быть заданы неориентированными графами $G^{SB}=(M, E^{SB})$, $G^{SG}=(M, E^{SG})$ и $G^{SS}=(M, E^{SS})$, в которых ребро $e=\{u, u'\} \in E^{SB}$ ($e \in E^{SG}$, $e \in E^{SS}$) тогда и только тогда, когда переходы u и u' должны выполняться в одной шпиндельной головке, одной и той же револьверной бабкой и на одном станке соответственно.

Регламенты невозможности выполнения пар переходов из M на одном станке, одной и той же револьверной бабкой или шпиндельной головкой могут быть также заданы неориентированными графами $G^{DS}=(M, E^{DS})$, $G^{DG}=(M, E^{DG})$ и $G^{DB}=(M, E^{DB})$, в которых ребро $e=\{u, u'\} \in E^{DS}$ ($e \in E^{DG}$, $e \in E^{DB}$) тогда и только тогда, когда переходы u и u' не могут быть выполнены совместно на одном станке, одной и той же револьверной бабкой или шпиндельной головкой соответственно.

Предполагается, что матрица \mathbf{H} и множества $h(u)$, $u \in M$, согласованы с ограничениями обязательности выполнения некоторых пар переходов одной шпиндельной головкой, т. е. $h(u) = h(u')$ для всех $\{u, u'\} \in E^{SB}$ и матрица \mathbf{H} не содержит строк r , в которых $h_{rs(u)} \notin h(u)$.

Положим $P_{kj}=(P_{kjl}=(N_{kjl}, \Gamma_{kjl}) | l=1, \dots, b_{kj})$, $P_k=(P_{kj} | j \in J)$, $P=(P_k | k=1, \dots, m)$. Определим некоторые временные характеристики набора P .

Время выполнения непустого множества N_{kjl} переходов одной шпиндельной головкой при ее минутной подаче Γ_{kjl} находится по формуле $t^p(P_{kjl})=L(N_{kjl})/\Gamma_{kjl}+\tau^a$, где $L(N_{kjl})=\max\{\lambda(u) | u \in N_{kjl}\}$ и τ^a – время, связанное с подводом и отводом блока инструментов. Если для j -го типа обработки

установлена револьверная бабка, то время обработки детали $t^p(P_{kj})=\sum_{l=1}^{b_{kj}} t^p(P_{kjl})+\tau^g b_{kj}$, где τ^g – время,

связанное с одним поворотом револьверной бабки, в противном случае $t^p(P_{kj})=t^p(P_{kj1})$. Время обработки детали на k -м станке $t^p(P_k)=\max\{t^p(P_{kj}) | j \in J\}$, а на всей линии $T(P)=\max\{t^p(P_k) | k=1, \dots, m\}+\tau^r$,

где τ^r – время, связанное с перемещением детали транспортным устройством по станкам, их базированием и закреплением на станке. Для обеспечения требуемой производительности это время не должно превышать некоторого заданного значения T_0 .

Пусть C_1, C_2, C_3 и C_4 – «весовые» коэффициенты относительной стоимости оборудования одного станка, одной револьверной бабки, одной шпиндельной головки, устанавливаемой на револьверной бабке, и одной шпиндельной коробки соответственно. Тогда стоимость оборудования для j -го типа обработки вычисляется по формуле

$$C(b_{kj}) = \begin{cases} 0, & \text{если } b_{kj} = 0; \\ C_4, & \text{если } b_{kj} = 1; \\ C_2 + C_3 b_{kj}, & \text{если } b_{kj} > 1. \end{cases}$$

Задача отыскания наилучшего (по предполагаемой оценке суммарной стоимости оборудования линии) набора P может быть представлена следующими соотношениями:

$$\text{Min } Q(P) = C_1 m + \sum_{k=1}^m \sum_{j=1}^4 C(b_{kj}); \quad (1)$$

$$T(P) \leq T_0; \quad (2)$$

$$\bigcup_{k=1}^m \bigcup_{j=1}^4 \bigcup_{l=1}^{b_{kj}} N_{kjl} = M; \quad (3)$$

$$N_{k'j'l'} \cap N_{k''j''l''} = \emptyset, \quad k', k'' = 1, \dots, m, \quad j', j'' = 1, \dots, 4, \quad l', l'' = 1, \dots, b_{k'j'}, \quad l' \neq l''; \quad (4)$$

$$\sum_{j=1}^4 \text{sign}(|N_{kj}|) \leq 3, \quad k = 1, \dots, m; \quad (5)$$

$$u \in \hat{N}_{k-1}, (u, u') \in A, u' \in N_{kjl}, u \notin N^j, \quad k = 1, \dots, m, \quad j \in J, \quad l = 1, \dots, b_{kj}; \quad (6)$$

$$u \in \hat{N}_{k-1} \cup \bigcup_{l'=1}^{l-1} N_{kjl'}, (u, u') \in A, u' \in N_{kjl}, u \cup N^j, \quad k = 1, \dots, m, \quad j \in J, \quad l = 1, \dots, b_{kj}; \quad (7)$$

$$|N_{kjl} \cap \{u, u'\}| \neq 1, \{u, u'\} \in E^{SB}, \quad k = 1, \dots, m, \quad j \in J, \quad l = 1, \dots, b_{kj}; \quad (8)$$

$$|\underline{N}_{kj} \cap \{u, u'\}| \neq 1, \{u, u'\} \in E^{SG}, \quad k = 1, \dots, m, \quad j \in J; \quad (9)$$

$$|\underline{N}_k \cap \{u, u'\}| \neq 1, \{u, u'\} \in E^{SS}, \quad k = 1, \dots, m; \quad (10)$$

$$|N_{kjl} \cap \{u, u'\}| \neq 2, \{u, u'\} \in E^{DB}, \quad k = 1, \dots, m, \quad j \in J, \quad l = 1, \dots, b_{kj}; \quad (11)$$

$$|\underline{N}_{kj} \cap \{u, u'\}| \neq 2 \text{ или } b_{kj} = 1, \{u, u'\} \in E^{DG}, \quad k = 1, \dots, m, \quad j \in J; \quad (12)$$

$$|\underline{N}_k \cap \{u, u'\}| \neq 2, \{u, u'\} \in E^{DS}, \quad k = 1, \dots, m; \quad (13)$$

$$s(u) = s(u'), \{u, u'\} \in N_{kjl}, \quad k = 1, \dots, m, \quad j \in J, \quad l = 1, \dots, b_{kj}; \quad (14)$$

$$\Gamma_{kjl} \in [\min\{\gamma_2(u) | u \in N_{kjl}\}, \max\{\gamma_1(u) | u \in N_{kjl}\}], \quad k = 1, \dots, m, \quad j \in J, \quad l = 1, \dots, b_{kij}; \quad (15)$$

$$\exists r_k \in \{1, \dots, r_0\}, h_{r_k s(u)} = j, \quad j \in h(u), \quad u \in N_{kjl}, \quad k = 1, \dots, m, \quad l = 1, \dots, b_{kj}; \quad (16)$$

$$m \leq m_0, \quad b_{kj} \leq b_0, \quad k = 1, \dots, m, \quad j \in J. \quad (17)$$

В этих соотношениях $\text{sign}(a)=1$, если $a>0$, и $\text{sign}(a)=0$, если $a\leq 0$; $\underline{N}_{kj} = \bigcup_{l=1}^{b_{kj}} N_{kjl}$; $\underline{N}_k = \bigcup_{j=1}^4 \underline{N}_{kj}$;

$$\hat{N}_k = \bigcup_{k'=1}^k \underline{N}_{k'}; N^j = \bigcup_{k=1}^m \bigcup_{l=1}^{b_{kj}} N_{kjl}.$$

Соотношения (3) и (4) обеспечивают выполнение каждого перехода из множества M ровно одной шпиндельной головкой; соотношение (5) гарантирует, что на каждом станке будет производиться обработка не более чем с трех сторон; соотношения (6) и (7) обеспечивают соблюдение ограничений предшествования; соотношения (8)–(13) гарантируют выполнение ограничений обязательности и невозможности совмещения пар переходов; соотношение (14) запрещают совмещение для одного типа обработки на станке переходов, соответствующих элементам, расположенным на разных сторонах детали; соотношения (15)–(17) обеспечивают выбор допустимых режимов обработки, положения детали и конфигурации линии соответственно.

2. Метод решения

Решение задачи (1)–(17) может быть получено с помощью следующей трехуровневой декомпозиционной схемы [3, 4]. На верхнем уровне решается задача A_1 отыскания наилучшего разбиения $\langle \underline{N}_1, \underline{N}_2, \dots, \underline{N}_m \rangle$ всего заданного множества переходов M по станкам. На среднем уровне решается задача A_2 выбора положения детали для выполнения множества переходов \underline{N}_k . На нижнем уровне решается задача A_3 определения возможности распределения множества переходов N_k по блочным переходам станка при фиксированном положении детали H_k на загрузочной позиции станка с помощью модификации графового подхода [5]. Эта модификация заключается в независимом распределении подмножеств технологических переходов, соответствующих различным сторонам обрабатываемой детали. Задача A_2 решается полным перебором по допустимым положениям детали. Для решения задачи A_1 используется теоретико-графовый подход, основанный на сведении исходной задачи к поиску оптимального ограниченного пути в специальном орграфе [4].

В настоящей работе для решения задачи (1)–(17) предлагается подход, основанный на ее сведении к задаче смешанного линейного программирования.

Введем следующие переменные:

X_{ukjl} – булевы переменные, равные 1, если переход u из M выполняется на k -м станке l -й шпиндельной головкой с ориентацией j , и равные 0 в противном случае, $u \in M$, $k=1, \dots, m_0$, $j \in J$, $l=1, \dots, b_0$;

Y_{kjl} – вспомогательные булевы переменные, равные 1, если на k -м станке производится обработка детали l -й шпиндельной головкой ориентации j , и равные 0 в противном случае, $k=1, \dots, m_0$, $j \in J$, $l=1, \dots, b_0$;

Y_{kj}^s – вспомогательные булевы переменные, равные 1, если на k -м станке производится обработка шпиндельной головкой при ориентации j элемента, расположенного на стороне s детали, и равные 0 в противном случае, $s=1, \dots, n_0$, $k=1, \dots, m_0$, $j \in J$;

Y_{kj} – вспомогательные булевы переменные, равные 1, если на k -м станке производится обработка шпиндельной головкой ориентации j , и равные 0 в противном случае, $s=1, \dots, n_0$, $k=1, \dots, m_0$, $j \in J$, $l=1, \dots, b_0$;

Z_k – вспомогательные булевы переменные, равные 1, если на k -м станке производится обработка, и равные 0 в противном случае, $k=1, \dots, m_0$;

h_{rk} – вспомогательные булевы переменные, равные 1, если на k -м станке деталь может быть обработана при ее ориентации r , и равные 0 в противном случае, $k=1, \dots, m_0$, $r=1, \dots, r_0$;

F_{kjl} – вспомогательные действительные неотрицательные переменные, используемые для определения времени обработки детали на k -м станке l -й шпиндельной головкой ориентации j , $k=1, \dots, m_0$, $j \in J$, $l=1, \dots, b_0$;

F_k – вспомогательные действительные неотрицательные переменные, используемые для определения времени обработки детали на k -м станке, $k=1, \dots, m_0$.

Для определения значения F_{kj} введем также вспомогательный параметр $f(u, u') = \max[\lambda(u), \lambda(u')] / \min[\gamma_2(u), \gamma_2(u')] + \tau^a$ для всех $j \in J$ и таких пар $(u, u') \in (M \times M) \setminus E^{DB}$, что $\max[\gamma_1(u), \gamma_1(u')] \leq \min[\gamma_2(u), \gamma_2(u')]$. При этом предполагается, что $(u, u') \in E^{DB}$, если $\max[\gamma_1(u), \gamma_1(u')] > \min[\gamma_2(u), \gamma_2(u')]$.

Тогда целевая функция (1) может быть представлена выражением

$$\text{Min } C_1 \sum_{k=1}^{m_0} Z_k + C_4 \sum_{k=1}^{m_0} \sum_{j=1}^4 Y_{kj1} + (C_2 + 2C_3 - C_4) \sum_{k=1}^{m_0} \sum_{j=1}^4 Y_{kj2} + C_3 \sum_{k=1}^{m_0} \sum_{j=1}^4 \sum_{l=3}^{b_0} Y_{kjl} . \quad (18)$$

Вспомогательные переменные Y_{kjl} , Y_{kjl}^s , Y_{kj} и Z_k определяются следующими соотношениями:

$$Y_{kj}^s \leq \sum_{u \in M_s} X_{ukj1}, \quad s=1, \dots, n_0, \quad k=1, \dots, m_0, \quad j \in J; \quad (19)$$

$$\sum_{u \in M_s} X_{ukj1} \leq |M_s| Y_{kj}^s, \quad s=1, \dots, n_0, \quad k=1, \dots, m_0, \quad j \in J; \quad (20)$$

$$\sum_{j=1}^4 Y_{kj}^s \leq 1, \quad s=1, \dots, n_0, \quad k=1, \dots, m_0; \quad (21)$$

$$\sum_{s=1}^{n_0} Y_{kj}^s \leq 1, \quad k=1, \dots, m_0, \quad j \in J; \quad (22)$$

$$Y_{kjl} \leq \sum_{u \in M} X_{ukjl}, \quad k=1, \dots, m_0, \quad j \in J, \quad l=1, \dots, b_0; \quad (23)$$

$$\sum_{u \in M} X_{ukjl} \leq M |Y_{kjl}|, \quad k=1, \dots, m_0, \quad j \in J, \quad l=1, \dots, b_0; \quad (24)$$

$$Y_{kjl} \leq Y_{kjl+1}, \quad k=1, \dots, m_0, \quad j \in J, \quad l=1, \dots, b_0-1; \quad (25)$$

$$\sum_{j=1}^4 Y_{kj} \leq 3, \quad k=1, \dots, m_0; \quad (26)$$

$$\sum_{j=1}^4 Y_{kj} \leq 3Z_k, \quad k=1, \dots, m_0; \quad (27)$$

$$Z_k \leq \sum_{j=1}^4 Y_{kj}, \quad k=1, \dots, m_0; \quad (28)$$

$$Z_k \geq Z_{k+1}, \quad k=1, \dots, m_0-1. \quad (29)$$

При выборе ориентации $(h_{r1}, h_{r2}, \dots, h_{rm_0})$ детали булевы переменные $Y_{kh_{rs}}^s$ принимают значение 1, если h_{rs} не равно 0, и соответственно $Y_{kj}^s = 0$ для $s=1, \dots, n_0, \quad k=1, \dots, m_0, \quad j \in J, \quad j \neq h_{rs}$. Это

условие равносильно тому, что $\sum_{s=1}^{n_0} \sum_{j=1, j \neq h_{rs}}^4 Y_{kj}^s = 0$. Таким образом, вспомогательные булевы

переменные h_{rk} могут быть определены из следующих соотношений:

$$h_{rk} \geq 1 - \sum_{s=1}^{n_0} \sum_{j=1, j \neq h_{rs}}^4 Y_{kj}^s, \quad r=1, \dots, r_0, \quad k=1, \dots, m_0; \quad (30)$$

$$\sum_{r=1}^{r_0} h_{rk} \geq 1, \quad k=1, \dots, m_0. \quad (31)$$

Следует отметить, что соотношения (30)–(31) могут выполняться для нескольких $r \in \{1, \dots, n_0\}$.

Искомые булевы переменные X_{ukjl} должны удовлетворять следующей системе соотношений:

$$\sum_{k=1}^{m_0} \sum_{j=1}^4 \sum_{l=1}^{n_0} X_{ukjl} = 1, \quad u \in M; \quad (32)$$

$$\sum_{k'=1}^{k-1} \sum_{j=1}^4 \sum_{l=1}^{b_0} X_{uk'j'l'} + \sum_{l'=1}^{l-1} X_{ukjl'} \geq X_{u'kjl}, \quad (u, u') \in A, \quad k=1, \dots, m_0, \quad j \in J, \quad l=1, \dots, b_0; \quad (33)$$

$$X_{ukjl} = X_{u'kjl}, \quad \{u, u'\} \in E^{SB}, \quad k=1, \dots, m_0, \quad j \in J, \quad l=1, \dots, b_0; \quad (34)$$

$$\sum_{l=1}^{b_0} X_{ukjl} = \sum_{l=1}^{b_0} X_{u'kjl}, \quad \{u, u'\} \in E^{SG}, \quad k=1, \dots, m_0, \quad j \in J; \quad (35)$$

$$\sum_{j=1}^4 \sum_{l=1}^{b_0} X_{ukjl} = \sum_{j=1}^4 \sum_{l=1}^{b_0} X_{u'kjl}, \quad \{u, u'\} \in E^{SS}, \quad k=1, \dots, m_0; \quad (36)$$

$$X_{ukjl} + X_{u'kjl} \leq 1, \quad \{u, u'\} \in E^{DB}, \quad k=1, \dots, m_0, \quad j \in J, \quad l=1, \dots, b_0; \quad (37)$$

$$\sum_{l=1}^{b_0} X_{ukjl} + \sum_{l=1}^{b_0} X_{u'kjl} + Y_{kj2} \leq 2, \quad \{u, u'\} \in E^{DG}, \quad k=1, \dots, m_0, \quad j \in J; \quad (38)$$

$$\sum_{j=1}^4 \sum_{l=1}^{b_0} X_{ukjl} + \sum_{j=1}^4 \sum_{l=1}^{b_0} X_{u'kjl} \leq 1, \quad \{u, u'\} \in E^{DS}, \quad k=1, \dots, m_0; \quad (39)$$

$$X_{ukjl} = 0, \quad u \in M, \quad k=1, \dots, m_0, \quad j \in J, \quad j \notin h(u), \quad l=1, \dots, b_0. \quad (40)$$

Вспомогательные переменные F_{kjl} и F_k определяются соотношениями

$$F_{kjl} \geq f(u, u) X_{ukjl}, \quad u \in M, \quad k=1, \dots, m_0, \quad j \in J, \quad l=1, \dots, b_0; \quad (41)$$

$$F_{kjl} \geq f(u, u')(X_{ukjl} + X_{u'kjl} - 1), \quad u, u' \in M, \quad k=1, \dots, m_0, \quad j \in J, \quad l=1, \dots, b_0; \quad (42)$$

$$F_k \geq \sum_{l=1}^{b_0} F_{kjl} + \tau^g (2Y_{kj2} + \sum_{l=3}^{b_0} Y_{kjl}), \quad k=1, \dots, m_0, \quad j \in J; \quad (43)$$

$$F_k \leq T_0; \quad (44)$$

$$F_{kjl} \in [0, T_0 - \tau^r], \quad k=1, \dots, m_0, \quad j \in J, \quad l=1, \dots, b_0; \quad (45)$$

$$F_k \in [0, T_0 - \tau^r], \quad k=1, \dots, m_0. \quad (46)$$

Если на k -м станке установлена револьверная бабка с b_{kj} шпиндельными головками с ориентацией инструмента j , то согласно (43) $F_k \geq \sum_{l=1}^{b_0} F_{kjl} + b_{kj} \tau^g$. Если на k -м станке установлена шпиндельная коробка с ориентацией инструмента j , то $F_k \geq F_{kj1}$, поскольку $Y_{kjl} = 0$ для $l=2, \dots, b_0$.

Для решения задачи (18)–(46) могут быть использованы известные пакеты для решения задач смешанного линейного программирования (CPLEX, XPRESS-MP и др.).

3. Вычислительные эксперименты

Эксперименты проводились на модифицированных задачах из [3] на ПК с объемом оперативной памяти 2 Гб и процессором Intel 2.33 ГГц с использованием академической версии CPLEX 12.2. Модификации заключаются в применении при проектировании линии только станков со стационарным приспособлением вместо станков с приспособлением на поворотном или передвижном столе. Кроме того, для модифицированных задач возможна установка нескольких шпindelных головок на револьверной бабке, в то время как в задачах из [4] это не допускается. Для «весовых» коэффициентов C_1, C_2, C_3 и C_4 были приняты значения 25, 5, 1, 2 соответственно. Параметры $\tau^a = \tau^s = \tau^r = 0,1$ мин. Характеристики модифицированных задач приведены в табл. 1. В этой таблице O^{OR} – плотность графа предшествования G (отношение числа дуг в транзитивном замыкании графа G к $|M||M-1|/2$). Параметры $O^{DB}, O^{DG}, O^{DS}, O^{SB}, O^{SG}$ и O^{SS} определяются соответственно отношением числа ребер в графах $G^{DB}, G^{DG}, G^{DS}, G^{SB}, G^{SG}$ и G^{SS} к $|M||M-1|/2$. Для задач (6)–(10) в качестве допустимого рассматривалось только одно положение детали, при котором обработка производится справа и слева, поскольку обрабатываемые элементы расположены на противоположных сторонах (сверху и снизу).

Таблица 1

Характеристики задач

Задача	$ M $	m_0	n_0	r_0	O^{OR}	O^{DB}	O^{DG}	O^{DS}	O^{SB}	O^{SG}	O^{SS}
1	24	2	13	3	0,134	0,355	0,207	0,489	0,025	0,000	0,000
2	24	8	7	3	0,152	0,388	0,076	0,772	0,033	0,000	0,000
3	32	10	20	4	0,101	0,470	0,050	0,784	0,022	0,000	0,000
4	34	4	16	4	0,103	0,455	0,194	0,693	0,023	0,000	0,000
5	34	3	16	4	0,103	0,455	0,232	0,608	0,011	0,000	0,000
6	59	10	2	1	0,106	0,364	0,001	0,798	0,020	0,000	0,000
7	61	9	2	1	0,442	0,349	0,001	0,781	0,016	0,000	0,000
8	64	8	2	1	0,493	0,390	0,000	0,790	0,018	0,000	0,000
9	68	11	2	1	0,302	0,401	0,004	0,863	0,017	0,000	0,004
10	68	10	2	1	0,340	0,400	0,000	0,869	0,017	0,000	0,000

В табл. 2 представлены результаты решения модифицированных задач: с помощью декомпозиционного подхода (ДП) и модели смешанного линейного программирования (СЛП).

Таблица 2

Результаты решения задач

Задача	ДП				СЛП			
	Стоимость линии	Время решения, с	Число вершин	Число дуг	Стоимость линии	Время решения, с	Число переменных	Число ограничений
1	77	33,1	412	2043	77	1,8	725	10 824
2	205	10,1	248	885	205	71,7	2697	36 290
3	246	46,7	2173	11 847	246	51,9	4371	80 243
4	142	226,5	1407	8877	142	5,5	1925	42 617
5	118	548,4	1023	6465	118	2,6	1540	32 175
6	250	69,2	15 322	84 064	250	477,6	6181	238 439
7	244	27,6	2452	10 145	244	43,1	6283	229 976
8	227	37,9	4528	21 404	227	10,9	5329	227 735
9	316	51,4	16 312	115 301	316	25,3	7855	338 788
10	291	65,3	15 579	100 651	291	13,9	7141	307 429

Заключение

В статье предложены математическая модель и трехуровневая декомпозиционная схема решения задачи синтеза структуры технологического процесса для обработки деталей одного наименования на поточной линии, комплектуемой из однопозиционных агрегатных станков со стационарным приспособлением. Эта схема предполагает решение на верхнем уровне задачи отыскания наилучшего разбиения всего заданного множества переходов по станкам, на среднем уровне – задачи выбора оптимального положения детали на станке, а на нижнем уровне – задачи определения минимальной стоимости оборудования станка для выполнения распределенного множества переходов при фиксированном положении детали. Разработана также модель СЛП, которая позволяет для решения исходной проектной задачи использовать пакеты CPLEX, XPRESS-MP и др. Проведенные вычислительные эксперименты на ряде реальных проектных задач показали преимущества использования модели СЛП.

Список литературы

1. Автоматизация процессов машиностроения / под ред. А.И. Дашенко. – М. : Высшая школа, 1991. – 480 с.
2. Дашенко, А. Конструкции агрегатных станков / А.И. Дашенко, А.И. Шмелев. – М. : Высшая школа, 1982. – 176 с.
3. Комплекс программ автоматизации аванпроектирования поточных линий из агрегатных металлорежущих станков / Г.М. Левин [и др.] // Информационные технологии программы Союзного государства «ТРИАДА»: основные результаты и перспективы : сб. науч. тр. – Минск : ОИПИ НАН Беларуси, 2010. – С. 143–158.
4. Integrated configurable equipment selection and line balancing for mass production with serial-parallel machining systems / O. Battaia [et al.] // Engineering Optimization. – 2014. – Vol. 46, № 10. – P. 1369–1388.
5. Combinatorial techniques to optimally customize an automated production line with rotary transfer and turrets / O. Battaia [et al.] // IE Transactions. – 2014. – Vol. 46, № 9. – P. 867–879.

Поступила 06.07.2015

¹*Объединенный институт проблем информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: gyshin@newman.bas-net.by*

²*Высшая национальная школа аэронавтики и космоса,
Тулуза, Белина, 10
e-mail: battaia@emse.fr*

³*Горная школа Нанта,
Нант, Кастлера, 4
e-mail: Alexandre.Dolgui@mines-nantes.fr*

N.N. Guschinsky, O. Battaia, A. Dolgui

MODELS AND METHODS FOR SYNTHESIS OF MANUFACTURING PROCESS STRUCTURE FOR MACHINING AT FLOW LINE OF STATIONARY MACHINES

A problem of design of a flow line composed of stationary machines is considered. Mathematical models and methods for synthesis of manufacturing process structure are proposed. The models take into account main technological and structural constraints with the criterion of minimizing the line equipment cost. A decomposition scheme for solving the initial problem and its formulation in terms of a mixed linear problem are developed. Computational results are also presented.

ПРИКЛАДНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 621.382.3

А.Г. Стрижнев¹, А.А. Шихов¹, А.Н. Русакович²**МЕТОД ИССЛЕДОВАНИЯ НЕЛИНЕЙНЫХ ОБЪЕКТОВ
УПРАВЛЕНИЯ, РАБОТАЮЩИХ В ЗАДАННОМ ДИАПАЗОНЕ
ВЫХОДНЫХ КООРДИНАТ**

Предлагается метод исследования нелинейных объектов управления (ОУ), работающих в заданном диапазоне значений выходных координат. Рассматривается влияние существующих нелинейностей на работу ОУ, которое может быть определено по различным регулировочным характеристикам (РХ). Описываются новые подходы к снятию и построению регулировочных и логарифмических амплитудно-фазовых частотных характеристик (ЛАФЧХ). Разрабатывается специальный алгоритм тестирования, позволяющий провести полный цикл идентификации ОУ, построить комбинированную ЛАФЧХ, восстановить полную РХ и переменный коэффициент передачи, получить модель, близкую к поведению реального ОУ. Предложенный метод использовался при исследованиях реальных ОУ и показал свою высокую эффективность.

Введение

Большинство технических ОУ являются нелинейными. К наиболее часто встречающимся нелинейностям относят зону нечувствительности, зону насыщения, нелинейный коэффициент усиления, неуравновешенность (несбалансированность) механизмов, люфты и зазоры зубчатых передач и др. [1–4]. Определение характера нелинейностей и получение аналитических выражений, которыми они могут быть описаны, является важной задачей, позволяющей в дальнейшем не только проводить высокоточное моделирование работы всей системы, но и осуществлять синтез регуляторов, удовлетворяющих заданным критериям качества управления.

Выявить существующие нелинейности [5] проще всего по статической РХ [6], которая представляет собой зависимость установившихся значений выходной величины y от входной величины x . На практике наиболее часто используются РХ по положению, скорости и ускорению. Для ОУ, работающих в широком диапазоне выходных координат, снятие РХ не представляет большой проблемы. Однако для ОУ, работающих в узком диапазоне значений выходных координат, этот процесс может вызывать серьезные трудности. Дело в том, что при подаче ступенчатых входных сигналов на ОУ, работающие в узком диапазоне, выходная величина y не успевает достичь установившегося значения. Кроме того, при подаче входных сигналов большой амплитуды могут происходить удары ОУ об ограничители (упоры) и, как следствие, поломка техники.

Для преодоления указанных трудностей и решения ряда других задач был разработан новый метод исследования нелинейных ОУ, работающих в заданном диапазоне значений выходных координат. Метод содержит несколько последовательно выполняемых этапов: снятие и построение РХ и ЛАФЧХ, определение передаточной функции ОУ. Для уяснения сути данного метода необходимо раскрыть содержание этапов и определить условия его практического применения.

1. Снятие и построение РХ

При снятии РХ (рис. 1, а) на вход ОУ подают ступенчатые сигналы x_i равномерно изменяемой амплитуды и, не доходя 5–10 % до максимума заданного диапазона $|y_{\max}|$ выходной координаты y , в точке $|D|$ осуществляют изменение знака входных сигналов x_i (координаты точки $|D|$ в процессе снятия РХ не изменяют). При этом образуются входные сигналы x_i ,

имеющие форму меандра, с периодом повторения T_i , а выходные сигналы y_i оказываются «привязанными» к рабочему диапазону $[-y_{\max}; y_{\max}]$. Изменение амплитуды ступенчатых сигналов осуществляют через каждый цикл сканирования $T_{\text{цс}}$, минимальное время которого получается при движении в одну и другую сторону. Шаг изменения амплитуды входных сигналов x_i выбирают таким, чтобы получить достаточное число (обычно 20–30) точек РХ в диапазоне изменения $[-x_{\max}; x_{\max}]$ входной координаты. При реверсировании движения в точке $|D|$ возникают выбеги y_e [7], которые могут приводить к ударам механизмов ОУ об ограничители при подаче максимальных $|x_{\max}|$ входных сигналов. Этим и объясняется уменьшение координаты точки переключения $|D|$ на 5–10 % относительно максимума $|y_{\max}|$.

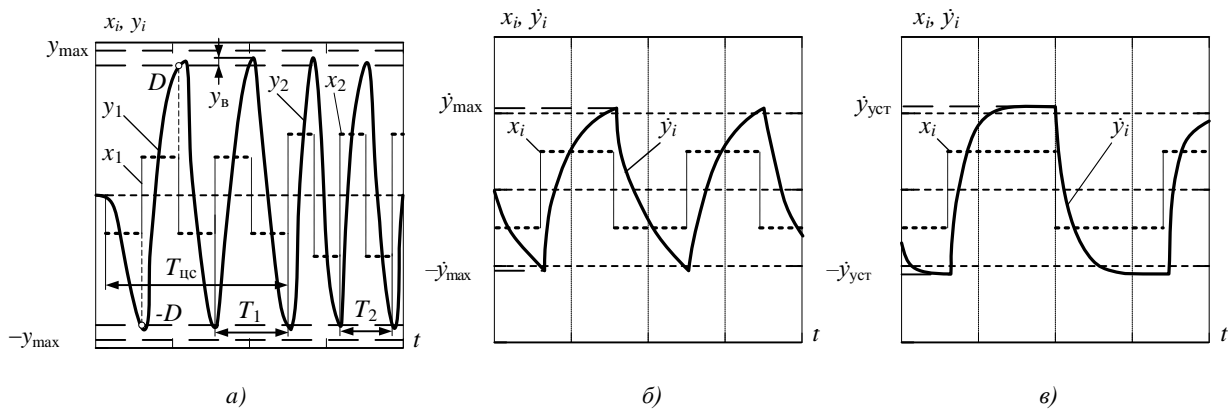


Рис. 1. Входные и выходные сигналы при снятии РХ: а) выбор точки реверсирования; б) определение максимальной скорости; в) определение установившейся скорости

В зависимости от диапазона выходной координаты $|y_{\max}|$ и инерционных свойств ОУ на его выходе могут образоваться максимальные или установившиеся значения выходной величины, которые затем используют для построения соответствующих РХ. В качестве примера на рис. 1, б приведены максимальные \dot{y}_{\max} , а на рис. 1, в – установившиеся $\dot{y}_{\text{уст}}$ значения выходной величины (скорости) астатического ОУ. Следует заметить, что максимальные значения отличаются от установившихся тем, что при условном увеличении координаты точки $|D|$ реверсирования максимальные значения \dot{y}_{\max} продолжали бы возрастать до установившейся величины $\dot{y}_{\text{уст}}$. Полученные таким образом максимальные или установившиеся значения выходных координат используют для построения соответствующих РХ и определения существующих нелинейностей. При этом помимо ранее перечисленных нелинейностей целесообразно проверить несимметричность положительной и отрицательной ветвей РХ, появление которой может быть вызвано, например, различной степенью износа шестерен или неуравновешенностью исполнительного механизма.

Для оценки несимметричности можно использовать выражение приведенной погрешности [8]:

$$N = \frac{y_3 - y_n}{y_{\max, \varepsilon}} \cdot 100 \% , \quad (1)$$

где $y_3 = y^+$ – эталонное значение РХ; $y_n = |y^-|$ – измеренное значение РХ; $y_{\max, \varepsilon} = y_{\max, \text{РХ}}^+$ – максимальное эталонное значение РХ. Знак плюс указывает на положительную ветвь РХ, а знак минус – на отрицательную. Применительно к вращательному движению принято считать, что положительная ветвь РХ образуется при вращении ОУ по часовой стрелке, а отрицательная – против часовой стрелки. Следует заметить, что окончательная оценка несимметричности осу-

ществляется по максимальному N_{\max} и среднеквадратичному $N_{\text{скв}}$ значениям несимметричности [8]:

$$N_{\max} = \max(|N_i|); N_{\text{скв}} = \sqrt{\frac{1}{n} \sum_{i=1}^n N_i^2}, \quad (2)$$

где N_i – текущее значение несимметричности; n – количество измеренных точек.

Несимметричность (1) считается допустимой, если среднеквадратичное значение N не превышает 3–5 %. При таких отклонениях N РХ считается симметричной и используется только ее положительная ветвь, в противном случае необходимо проводить исследование обеих ветвей РХ.

Для дальнейших исследований полученную ограниченную (с применением максимальных значений рис. 1, б) или полную (с применением установившихся значений рис. 1, в) РХ необходимо нормировать относительно максимального значения используемой РХ с помощью выражения

$$y_{\text{норм}} = y / y_{\max.\text{РХ}}, \quad (3)$$

где y , $y_{\max.\text{РХ}}$ – текущее и максимальное значения РХ.

После нормирования нужно провести аппроксимацию используемой РХ. Обычно применяется полиномиальная (рис. 2, а) либо кусочно-линейная (рис. 2, б) аппроксимация.

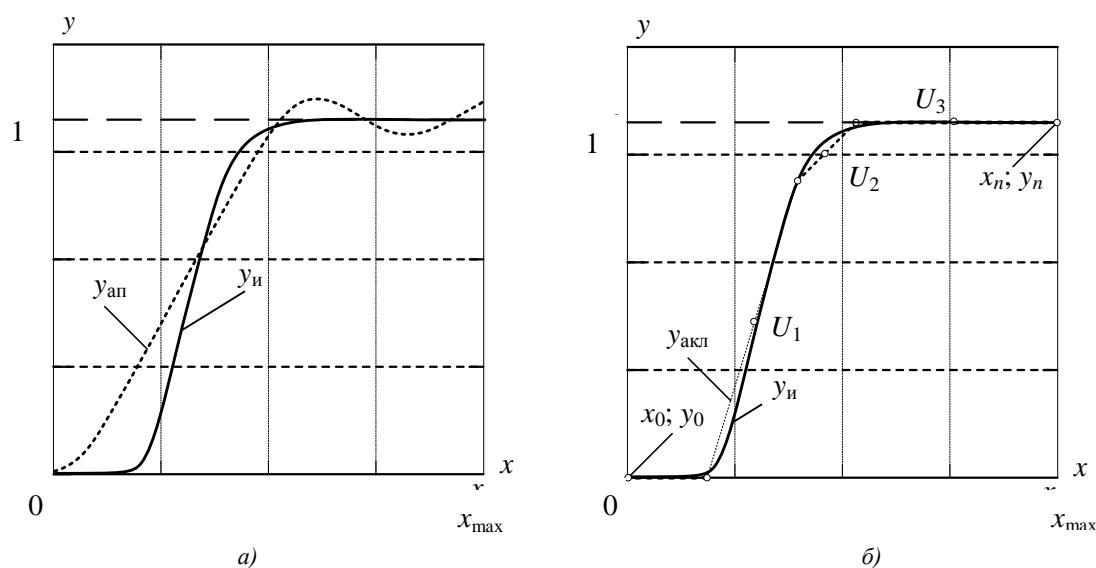


Рис. 2. Аппроксимация РХ: а) полиномиальная; б) кусочно-линейная

Выражение полиномиальной аппроксимации имеет вид [9]

$$y_{\text{ап}} = c_0 x^0 + c_1 x^1 + c_2 x^2 + \dots + c_n x^n, \quad (4)$$

где x – текущее значение входного сигнала; c_i – коэффициенты полинома.

Степень n полинома (4) может быть любой и выбирается исходя из требуемой точности аппроксимации, которая определяется по выражению, аналогичному (1):

$$p = \frac{y_i - y_a}{y_{i.\max}} \cdot 100\%, \quad (5)$$

где p – приведенное отклонение; y_i – измеренное значение; $y_{i, \max} = 1$ – максимальное измеренное значение; y_a – аппроксимированное значение.

Окончательная оценка аппроксимации осуществляется по максимальному и среднеквадратичному значениям, которые вычисляются по выражениям, аналогичным (2).

В случаях когда РХ имеет сложную форму, она может быть аппроксимирована с требуемой точностью только полиномами высоких порядков, которые сложны в вычислении. В такой ситуации предпочтительнее применять кусочно-линейную аппроксимацию, которую обычно представляют отдельными формулами на каждом из интервалов регулирования, а если РХ непрерывная, то целесообразно использовать выражение [10]

$$y_{\text{акт}} = \frac{1}{2} \left[y_0 + \frac{y_1 - y_0}{x_1 - x_0} (x - x_0) + y_{n-1} + \frac{y_n - y_{n-1}}{x_n - x_{n-1}} (x - x_{n-1}) \right] + \frac{1}{2} \sum_{k=1}^{n-1} \left(\frac{y_{k+1} - y_k}{x_{k+1} - x_k} - \frac{y_k - y_{k-1}}{x_k - x_{k-1}} \right) |x - x_k|, \quad (6)$$

где n – количество линейных участков РХ (см. рис. 2, б).

Количество участков кусочно-линейной функции (6) может быть любым. Оно выбирается исходя из того, чтобы среднеквадратичное значение отклонения (2), (5) не превышало 3–5 %.

Кусочно-линейная аппроксимация используется для выбора амплитуды U_i тестовых гармонических сигналов $x = U \sin \omega t$ изменяемой частоты ω , которые в дальнейшем применяют для снятия и построения семейства ЛАФЧХ [11]. В качестве U_i предлагается выбирать середины $U_1 - U_3$ линейных участков аппроксимированной РХ.

2. Снятие и построение ЛАФЧХ

Как было отмечено ранее, из-за наличия нелинейностей, главной из которых является неуравновешенность нагрузки, неизбежно приводящая к произвольному уходу ОУ на одну из границ выходного диапазона и столкновению его с ограничительными упорами, процесс получения РХ становится затруднительным или невозможным. По этой же причине нельзя на низких частотах использовать гармонические тестовые сигналы для снятия и построения ЛАФЧХ. Решить данную проблему можно, изменяя координаты точки реверсирования D_i (рис. 3, а) от минимальной до максимально возможной величины и тем самым обеспечивая регулирование частоты ($f_i = 1/T_i$) входных и выходных сигналов ОУ. Однако данный метод в сравнении с классическим [6] обеспечивает получение характеристик в более узком диапазоне частот. Это объясняется тем, что нижняя (низкочастотная) граница диапазона ограничена выходным сектором перемещения ОУ, который определяет максимально возможное значение точки D_i . Верхняя же частотная граница определяется минимально возможным значением точки D_i , которое ограничено минимальным значением выходной координаты, измеряемой датчиком. Для решения данной проблемы предлагается комбинация предложенного и классического методов получения ЛАФЧХ с учетом того, что уходы ОУ на одну из границ диапазона при снятии характеристик наблюдаются преимущественно в низкочастотной области. На высоких частотах времена циклов сканирования $T_{\text{цс}}$ и амплитуды колебаний u достаточно малы и вследствие уходов ОУ не успевают достичь границ сектора перемещения. Таким образом, предлагается в низкочастотной области использовать предложенный метод снятия характеристик, а в высокочастотной – классический метод [6]. Для получения результирующих характеристик требуется совместить ЛАФЧХ (L_1, φ_1), полученные с помощью предложенного метода, и ЛАФЧХ (L_2, φ_2), полученные с помощью классического метода (рис. 3, б).

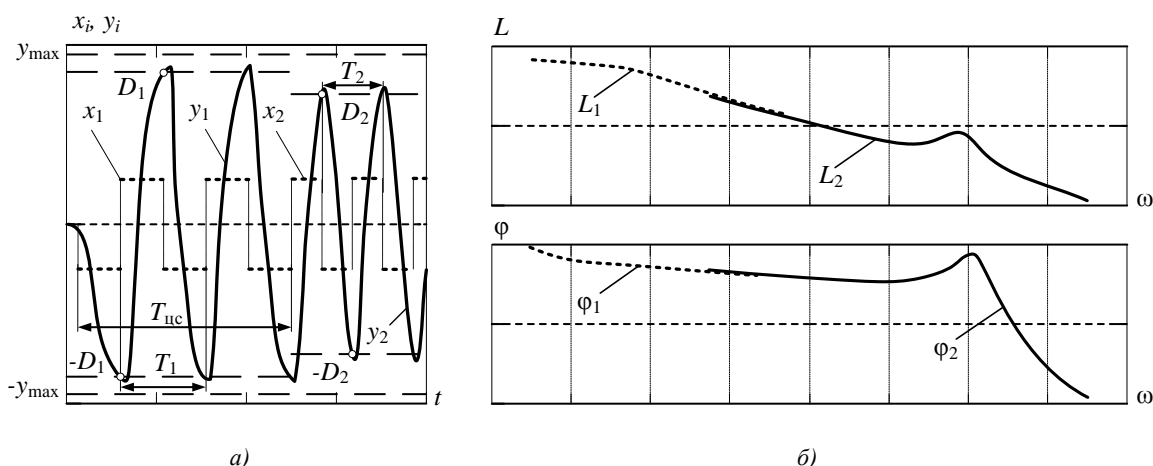


Рис. 3. Частотные характеристики:

а) входные и выходные сигналы; б) результирующая ЛАФЧХ

Следует отметить, что при использовании предложенного метода образуются входные и выходные сигналы, которые не являются гармоническими, и для построения ЛАФЧХ требуется определять параметры их первых гармоник.

Полученные таким образом результирующие ЛАФЧХ позволяют не только более точно определить передаточную функцию ОУ, но и восстановить полную РХ.

Передаточную функцию можно определить в упрощенном виде, используя графический метод либо метод узловых частот [12]. Если требуется более точная передаточная функция, можно, задаваясь порядком или структурой ОУ, воспользоваться известными методами параметрической идентификации [13–16] либо автоматизированными методами, которые предлагаются, например, в пакете MATLAB [17].

Восстановление полной РХ рассмотрим на примере астатического ОУ, имеющего передаточную функцию $G(s) = K[s(Ts + 1)]^{-1}$, где K – коэффициент передачи, T – постоянная времени. Следует заметить, что коэффициент передачи K характеризует отношение установившегося значения выходного сигнала к амплитуде тестового входного синусоидального сигнала на частоте $\omega = 1$, т. е. $K = y_{уст.U_i} / U_i |_{\omega=1}$ или $y_{уст.U_i} = KU_i |_{\omega=1}$.

Следовательно, можно записать выражение для восстановления полной РХ:

$$y_B = y_{норм} \frac{y_{уст.U_i}}{y_{норм.U_i}}, \tag{7}$$

где $y_{норм}$ – текущее значение нормированной РХ, которое определяется по выражению (3); $y_{уст.U_i} = KU_i$ – установившееся значение РХ при входном сигнале U_i ; $y_{норм.U_i}$ – нормированное значение РХ при входном сигнале U_i ; K – коэффициент передачи ОУ при входном сигнале U_i .

С помощью выражения (7) можно восстановить переменный коэффициент передачи ОУ во всем диапазоне входных сигналов:

$$K_B = \frac{y_B}{x}, \tag{8}$$

где y_B – восстановленное значение выходной величины, рассчитанное по выражению (7); x – амплитуда входного сигнала.

Выражения (7), (8) позволяют восстановить полную РХ, определить переменный коэффициент передачи и получить модель, адекватную по поведению реальному нелинейному ОУ.

Предложенный метод и его этапы характеризуются сложным вычислительным процессом и рассчитаны на применение современных информационных технологий.

3. Разработка специального алгоритма тестирования

Для автоматизации процессов исследования нелинейных ОУ, работающих в заданном диапазоне выходных координат, и реализации этапов предложенного метода был разработан специальный алгоритм тестирования (рис. 4).

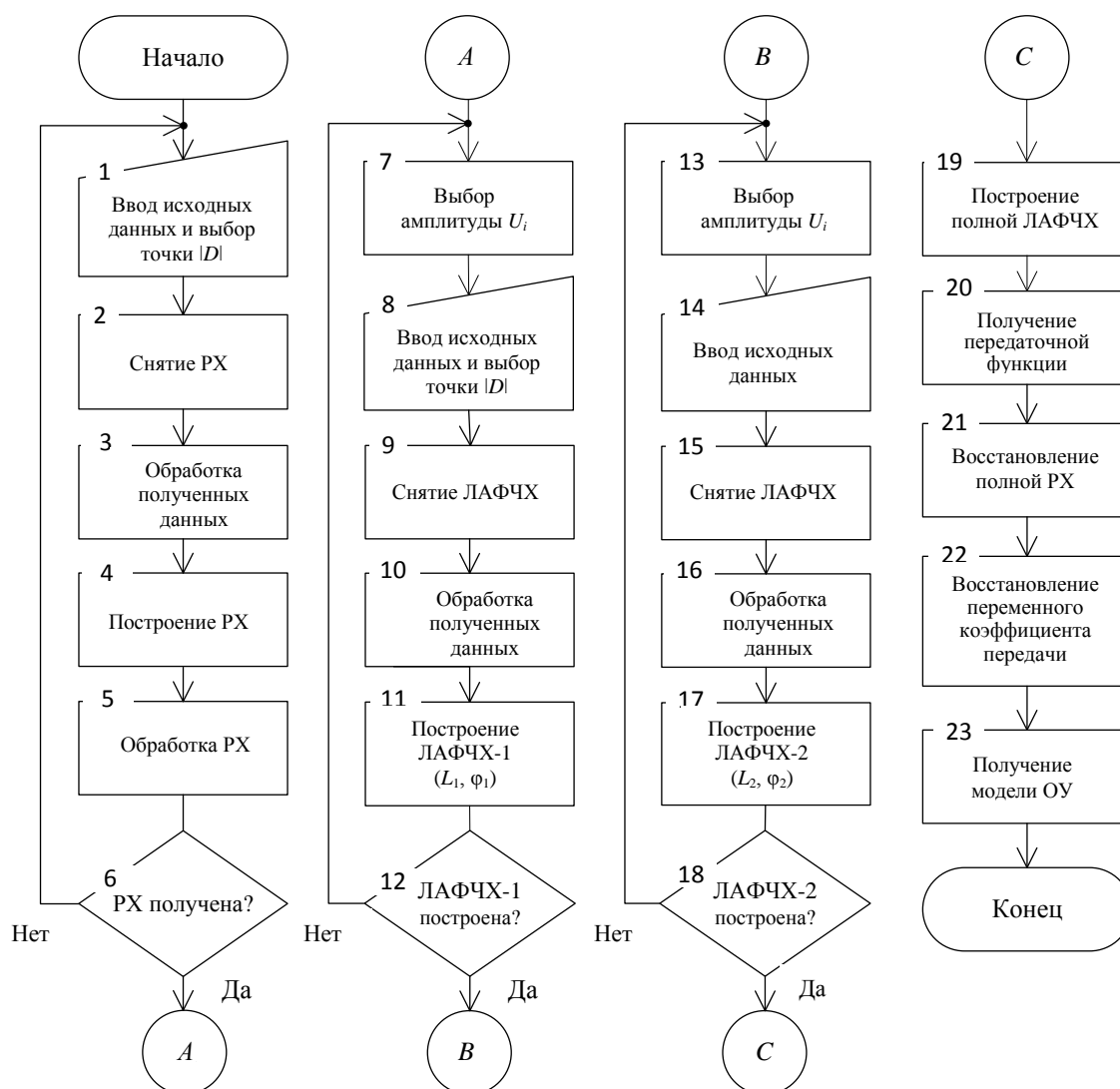


Рис. 4. Алгоритм исследования нелинейных ОУ, работающих в заданном диапазоне выходных координат

Алгоритм состоит из четырех этапов: снятие и построение РХ, снятие и построение ЛАФЧХ-1, снятие и построение ЛАФЧХ-2, построение комбинированной ЛАФЧХ и получение модели ОУ.

В начале этапа снятия и построения РХ в блоке 1 осуществляется выбор точки $|D|$ реверсирования и ввод исходных данных: диапазона и шага изменения входных сигналов x_i , количества движений в одном цикле сканирования $T_{ис}$, допустимого диапазона выходных координат y_{max} . Значение точки $|D|$ может быть установлено оператором или определено

автоматически. При автоматическом определении подается максимальный входной сигнал с начальной координатой точки $|D|$, примерно равной 50 % от максимального диапазона y_{\max} , а затем значение точки $|D|$ постепенно увеличивается на 10 % до значения, исключающего выход ОУ за границу заданного диапазона. Для снятия РХ в блоке 2 производится формирование, подача и запись входных и выходных сигналов ОУ. Полученные входные и выходные данные подаются в блок 3, где осуществляются дифференцирование, фильтрация, усреднение и определение максимальных и установившихся значений выходной величины. В блоке 4 происходит построение снятой РХ, которая затем подвергается проверке на несимметричность, нормированию и аппроксимации (блок 5). Полученная РХ является базовой характеристикой для перехода к этапам построения ЛАФЧХ-1 и ЛАФЧХ-2.

На втором этапе реализуется новая методика снятия ЛАФЧХ-1 в низкочастотной области, изложенная в разд. 2. При этом необходимо соблюдать следующую последовательность действий. Используя полученную на первом этапе аппроксимированную РХ, в блоке 7 определяются амплитуды входных тестовых сигналов U_i , которые поступают в блок 8. Кроме того, в блоке 8 вводятся исходные данные: диапазон и шаг изменения точки $|D|$, количество периодов движения в одном цикле сканирования. В блоке 9 осуществляются формирование, подача и запись входных и выходных сигналов ОУ, которые обрабатываются в блоке 10 с выделением первых гармоник, а затем используются в блоке 11 для построения ЛАФЧХ-1.

На третьем этапе реализуется классическая методика снятия ЛАФЧХ-2 в высокочастотной области. При этом методе снятия частотной характеристики последовательность действий в блоках 13–18 практически аналогична действиям в блоках 7–12. В блоке 13 определяются амплитуды входных тестовых сигналов U_i , эквивалентные амплитудам, определяемым в блоке 7. В блоке 14 вводятся исходные данные: диапазон и шаг изменения частоты, количество периодов движения в одном цикле сканирования. В блоке 15 осуществляются формирование, подача и запись гармонических входных и выходных сигналов ОУ, которые обрабатываются в блоке 16, а затем используются в блоке 17 для построения ЛАФЧХ-2.

На четвертом этапе построенные ЛАФЧХ-1 и ЛАФЧХ-2 поступают в блок 19, где происходит их объединение и образование результирующей (комбинированной) ЛАФЧХ. С помощью полученной ЛАФЧХ в блоке 20 осуществляются структурная и параметрическая идентификация и получение передаточной функции ОУ. Затем с помощью блоков 21–23 последовательно происходит восстановление полной РХ, переменного коэффициента передачи K_i и получение более точной модели ОУ.

Данный алгоритм позволяет проводить полный цикл идентификации ОУ, работающих в заданном диапазоне выходных координат, и может быть реализован в виде программного комплекса с помощью современных языков программирования высокого уровня.

Заключение

Разработан метод исследования ОУ, работающих в заданном диапазоне выходных координат, который позволяет осуществлять снятие и построение различных РХ, их нормирование и аппроксимацию, а также снятие и построение ЛАФЧХ в заданном диапазоне выходных координат без опасности уходов и ударов механизмов ОУ об ограничители. Для более полного изучения свойств ОУ в широком диапазоне частот предложена комбинированная амплитудно-фазовая частотная характеристика, получаемая с помощью предложенного и классического методов. Предложенный метод также позволяет восстановить полную РХ, определить переменный коэффициент передачи ОУ и получить модель, адекватную по поведению реальному нелинейному объекту.

Для автоматизации процесса исследования нелинейных ОУ, работающих в заданном диапазоне выходных координат, был разработан специальный алгоритм тестирования, который может быть реализован с помощью современных языков программирования высокого уровня. В дальнейшем планируется использовать разработанный метод для исследования различных нелинейных ОУ, сформировать правила и рекомендации практического характера.

Список литературы

1. Попов, Е.П. Теория нелинейных систем автоматического регулирования и управления : учеб. пособие / Е.П. Попов. – 2-е изд. – М. : Наука, 1988. – 256 с.
2. Способ и цифровая следящая система для определения и компенсации влияний неуравновешенности нагрузки на работу следящих систем и приводов : евраз. пат. 016669, МПК G05B 11/00, G05B 19/406 / А.Г. Стрижнев, Г.В. Ледник, А.А. Шихов, С.Л. Ботеновский, И.А. Гончарук, С.М. Виноградов, Г.В. Кирячок; заявитель ООО «Техносоюзпроект». – № 201001761; заявл. 07.10.10; опубл. 29.06.12 // Бюл. евраз. пат. ведомства. – 2012. – № 6. – С. 1–5.
3. Стрижнев, А.Г. Повышение точности работы цифровой следящей системы, содержащей неуравновешенную нагрузку / А.Г. Стрижнев, Г.В. Ледник, А.А. Шихов // Наука и техника. – 2013. – № 5. – С. 10–14.
4. Лурье, Б.Я. Классические методы автоматического управления / Б.Я. Лурье, П.Дж. Энрайт. – СПб. : БХВ-Петербург, 2004. – 640 с.
5. Справочное пособие по теории систем автоматического регулирования и управления / В.Д. Громыко [и др.]. – Минск : Выш. школа, 1973. – 584 с.
6. Пупков, К.А. Методы классической и современной теории автоматического управления : в 5 т. / К.А. Пупков, Н.Д. Егупов. – М. : Изд-во МГТУ им. Н.Э. Баумана, 2004. – Т. 2. – 640 с.
7. Машиностроение : энциклопедия : в 40 т. / под ред. К.С. Колесникова. – М. : Машиностроение, 1995. – Т. 1–3. – 624 с.
8. Новицкий, П.В. Оценка погрешностей результатов измерений / П.В. Новицкий, И.А. Зограф. – Л. : Энергоатомиздат, 1985. – 248 с.
9. Бахвалов, Н.С. Численные методы : учеб. пособие для вузов / Н.С. Бахвалов, Н.П. Жидков, Г.М. Кобельков. – М. : Лаборатория базовых знаний, 2002. – 632 с.
10. Доля, П.Г. Об одном способе представления кусочных полиномов в системах символьной математики / П.Г. Доля // Вестник Харьк. нац. ун-та. – 2008. – Т. 833, № 10. – С. 110–120.
11. Борцов, Ю.А. Автоматизированный электропривод с упругими связями / Ю.А. Борцов, Г.Г. Соколовский. – СПб. : Энергоатомиздат, 1992. – 288 с.
12. Солодовников, В.В. Основы теории и элементы систем автоматического управления : учеб. пособие для вузов / В.В. Солодовников, В.Н. Плотников, А.В. Яковлев. – М. : Машиностроение, 1985. – 536 с.
13. Марков, А.В. Параметрическая идентификация динамических объектов по фазочастотным характеристикам / А.В. Марков, В.И. Симаньков // Доклады БГУИР. – 2015. – № 3(89). – С. 29–35.
14. Гроп, Д. Методы идентификации систем / Д. Гроп. – М. : Мир, 1979. – 302 с.
15. Ljung, L. System Identification: Theory for the User / L. Ljung. – New Jersey : Prentice Hall, 1999. – 672 p.
16. Эйкхофф, П. Основы идентификации систем управления / П. Эйкхофф. – М. : Мир, 1975. – 686 с.
17. Ljung, L. System Identification Toolbox: User's Guide / L. Ljung. – Natick : Mathworks, 2005. – 886 p.

Поступила 19.05.2015

¹НПООО «ОКБ Техносоюзпроект»,
Минск, пр. Независимости, 115
e-mail: skb@tspb.com

²Белорусский государственный университет
информатики и радиоэлектроники,
Минск, П. Бровки, 6
e-mail: shihovalex@gmail.com,
aliaksei.rusakovich@gmail.com

A.G. Stryzhniou, A.A. Shykhau, A.N. Rusakovich**METHOD OF INVESTIGATION OF NONLINEAR OBJECTS OPERATING
WITHIN SPECIFIED RANGE OF OUTPUT COORDINATES**

This article presents a method of investigation of nonlinear objects operating within specified range of output coordinates. The effect of nonlinear behavior of objects is considered, which is detected by measurable characteristics. New approaches to decibel-log frequency response measurement are proposed. Special testing method that realizes complete cycle of object identification is developed. The method is applied for investigation of real objects with high efficiency.

УДК 004.31

А.С. Шашков

ПРОЕКТИРОВАНИЕ ЦИФРОВЫХ СИСТЕМ С ПОНИЖЕННЫМ ЭНЕРГОПОТРЕБЛЕНИЕМ С ПРИМЕНЕНИЕМ ТЕХНОЛОГИИ UPF-ОПИСАНИЯ ПОДСИСТЕМЫ ПИТАНИЯ

В рамках проблемы проектирования энергетически эффективных цифровых систем приводятся различные методы понижения энергопотребления. Обосновывается необходимость дополнительных средств формального поведенческого описания доменной подсистемы питания. Рассматриваются элементы доменной подсистемы питания и этапы проектирования цифровой системы с учетом ее разделения на домены питания. Дается пример описания системы с применением формата UPF.

Введение

Проблема проектирования энергетически эффективных (с пониженным энергопотреблением) вычислительных систем всегда была одной из основных в индустрии электронных систем. В настоящее время эта проблема стоит наиболее остро [1]. Необходимость в удешевлении микросхем и увеличении их компактности, обусловленная как широким распространением все усложняющейся мобильной электроники, так и потребностью в высокоэффективных центрах обработки данных, поставила проблему энергосбережения на новый уровень. Так, производительность систем на кристалле стала существенно ограничиваться возможностями автономных источников питания для мобильных систем, в то время как многочисленные центры обработки данных затруднительно и затратно эксплуатировать и расширять ввиду ограниченного ресурса электроэнергии [2]. Одной из актуальных проблем является также проектирование эффективных и компактных средств отвода тепла как в мобильных, так и в стационарных системах. В таких условиях среди всех целей проектирования компьютерных систем на одном из первых мест стоит увеличение производительности, приходящейся на единицу рассеиваемой мощности. Все это способствовало активному развитию методов и технологий, позволяющих снижать энергопотребление интегральных схем при сохранении их высокой производительности [2–6].

В настоящей обзорной статье рассматриваются существующие методы снижения энергопотребления электронных систем, однако особое внимание уделяется одному из новейших методов снижения энергопотребления – использованию доменной системы питания, а также способу поведенческого проектирования системы питания с использованием средств формата, специально для этого предназначенного и названного Unified Power Format (UPF).

1. Обзор методов понижения энергопотребления для однокристалльных систем

Рассеяние мощности интегральной микросхемы состоит из двух составляющих: динамической и статической мощностей. Динамическая мощность пропорциональна квадрату напряжения питания, частоте переключений элементов и эффективной емкости элементов. Статическое энергопотребление характеризуется величиной токов утечки, которая возрастает с ростом напряжения питания, уменьшением пороговых напряжений переключения транзисторов и с увеличением ширины транзисторов [1, 5]. С другой стороны, понижение напряжения питания, увеличение пороговых напряжений, уменьшение частоты переключений приводят к снижению вычислительной производительности микросхемы.

При технологических процессах производства микросхем более 90 нм динамическая составляющая мощности существенно преобладает. Для снижения динамической мощности были разработаны следующие методы [1]:

- оптимизация цепей синхросигнала и управление пропуском синхросигнала;
- изоляция операндов;
- реструктуризация логики;
- оптимизация размеров транзисторов;
- использование буферов для минимизации времени переключения вентиляей;
- обмен выводами (выбор выводов с наименьшей емкостной нагрузкой).

Исследования показывают [1], что оптимизация цепей синхросигнала позволяет снизить динамическую мощность до 20 %. Другие методы позволяют снизить динамическую мощность до 5 % каждый.

Методы снижения динамической мощности эффективно реализуются в САПР при синтезе проекта в автоматическом режиме. Часть задач по оптимизации логики может быть решена разработчиком на уровне RTL-описаний. Таким образом, использование методов снижения динамической мощности предполагает наличие следующих входных данных:

- RTL-описание проекта;
- задания параметров оптимизации энергопотребления для САПР.

С уменьшением технологических процессов уменьшается и пороговое напряжение переключения транзисторов, что обуславливает существенный рост токов утечки, а значит, и увеличение статической мощности [3]. Современные массовые технологические процессы производства микросхем уже давно перешагнули пороговый рубеж в 65 нм, после которого статическая мощность микросхем стала превышать динамическую мощность [1]. Также с уменьшением технологических процессов значительно возросла и плотность рассеиваемой мощности кристаллов [6], что затруднило отвод тепла. В связи с этим возникла острая потребность в методах, позволяющих наряду с уменьшением динамической мощности в первую очередь уменьшать статическую мощность всей системы на кристалле. К ним относятся:

- использование нескольких доменов (областей кристалла) с различными напряжениями питания;
- использование отключаемых доменов питания;
- динамическое изменение напряжений питания;
- динамическое изменение частоты сигнала синхронизации;
- распределение памяти на блоки, расположенные в разных доменах питания (часть памяти может отключаться или переводиться в режим удержания данных без возможности чтения и записи);
- использование библиотек логических вентиляей с разными порогами переключения (реализуется автоматически в САПР);
- использование технологии смещения напряжения подложки транзисторов в режимах высокого и низкого энергопотребления (требует дополнительного источника питания).

Исследования показывают [1], что для технологических процессов от 65 нм и ниже при комбинировании всех современных технологий энергосбережения систем на кристалле наибольший вклад в сохранение энергии вносит технология отключаемых доменов питания, позволяющая сократить среднюю мощность токов утечки до 50 раз. Технологии динамического изменения напряжения питания и частоты синхросигнала позволяют снизить динамическую мощность до 70 %, а статическую мощность – до трех раз.

Доменные методы энергосбережения не представляется возможным эффективно реализовать автоматически ввиду того, что фактически эти методы являются задачами логического проектирования активной доменной сети или задачами, сопряженными с логической составляющей проекта. Таким образом, использование вышеперечисленных методов предполагает следующие входные данные для синтеза проекта:

- RTL-описание логической иерархии проекта;
- описание доменов и других элементов сети питания;
- описание логики управления активной частью сети питания (также реализуется средствами RTL).

Анализируя входные данные, необходимые для описания энергетически эффективных систем на кристалле, можно заключить, что часть описания энергетически эффективной системы можно выполнить привычными средствами HDL (Hardware Description Language, язык описания аппаратуры), такими как VHDL [7] или System Verilog [8], дополняя их выбором настроек оптимизации проекта в САПР. Однако описание активной сети доменов питания и других элементов доменной системы невозможно с помощью только VHDL или System Verilog ввиду того, что данные языки изначально не были предназначены для этой цели. Поэтому возникла проблема формального описания проектов цифровых систем с учетом обеспечения их энергетической эффективности уже на начальном этапе проектирования, т. е. на уровне разработки исходных HDL-описаний.

2. Элементы доменных систем питания

В общем случае для описания системы с управляемым питанием на поведенческом уровне можно выделить следующие основные элементы [1, 9, 10].

Описание логической составляющей проекта. Представляет собой RTL-описание модулей, составляющих логическую часть иерархии проекта.

Порты и цепи питания (Supply Ports, Supply Nets). Цепи питания и земли, подключенные к соответствующим портам питания и земли, обеспечивают питание для соответствующих элементов проекта: как логические элементы (RTL), так и элементы обеспечения энергетической эффективности имеют свои источники питания. Цепи питания могут логически характеризоваться не только параметром выключения-включения, но и уровнем напряжения. Недостаточное напряжение питания неизбежно приводит к ненормальной работе питаемого компонента системы.

Домены питания (Power Domains). Логические элементы проекта (RTL-модули) распределяются в соответствующие обособленные домены питания. Домены могут быть логически вложенными в другие домены, несколько доменов могут быть вложены в один домен. Таким образом, проект будет иметь не только иерархию RTL-модулей, но и иерархию доменов питания. Разделение системы на домены питания позволяет выделить участки системы с общей стратегией питания (с единым набором правил питания). Так, разные домены могут питаться от различных цепей питания, имеющих различные напряжения. Домены могут иметь несколько различных источников питания: основной для логических элементов и специальные для элементов обеспечения энергетической эффективности.

Переключатели питания (Power Switches). Одним из элементов стратегии питания для домена может являться возможность отключения всех логических элементов домена с помощью переключателя (переключателей). В этом случае как динамическое, так и статическое потребление энергии элементами домена сводится к минимуму. Переключатели также могут иметь несколько входных портов питания при одном выходном. Переключатели питания управляются специальными внешними логическими сигналами.

Регистры сохранения состояния для выключаемых доменов (State Retention). Отключение питания от элемента памяти (регистра) приводит к потере его состояния. Регистры сохранения по сигналам управления сохраняют и восстанавливают содержимое целевых регистров в процессе выключения-включения основного питания. Регистры сохранения состояния имеют отдельные источники питания, отличные от основных (и отключаемых) источников питания для домена, а также позволяют ускорить процесс инициализации модуля после его очередного включения. Проектировщик конкретного RTL-модуля должен выделить некоторое подмножество регистров, для которых необходимы регистры сохранения. После очередного включения домена и восстановления данных из регистров сохранения должна быть гарантирована нормальная логическая работоспособность подсистем этого домена.

Средства изоляции сигналов, исходящих из отключаемых доменов (Isolation). В процессе отключения питания все сигналы, выходящие из домена, переходят в неопределенное (плавающее) состояние. Принимающий данный сигнал модуль, входящий в состав одного из включенных доменов, может нарушить свою логическую работу из-за случайного переключения этого входного для себя сигнала. Нежелательный промежуточный уровень сигнала также мо-

жет привести к высокому потреблению тока по данной линии сигнала. Для решения этих проблем выходные сигналы отключаемого домена пропускаются через блоки изоляции. По управляющему сигналу включения блок изоляции выдает вместо входного неопределенного сигнала задаваемое разработчиком выходное значение сигнала 0, 1, Z или «защелкнутое» в регистр предыдущее определенное значение входного сигнала.

Средства изменения уровня напряжения сигналов, проходящих между доменами с различными уровнями напряжения (Level Shifters). На границах доменов, имеющих различные напряжения питания, устанавливаются блоки изменения уровня напряжения сигналов. Таким образом, напряжение единицы для выходного сигнала одного домена будет однозначно восприниматься как напряжение единицы в любом другом домене с другим напряжением питания, который принимает данный сигнал.

Средства логического управления элементами обеспечения энергетической эффективности. Как было описано выше, такие элементы доменной системы питания, как переключатели питания доменов, блоки изоляции доменов, регистры сохранения, управляются некоторыми специальными сигналами. За формирование этих сигналов может отвечать либо отдельное RTL-устройство (модуль управления питанием), либо несколько устройств. Таким образом, система управления питанием может быть централизованной или распределенной, что задается разработчиком. Для перевода домена из включенного состояния в выключенное может потребоваться определенная разрозненная во времени последовательность управляющих сигналов. Так, сначала домен необходимо изолировать от внешних для него доменов. Затем необходимо сохранить содержимое определенных регистров в соответствующие регистры сохранения. Только после этого на переключатель питания домена можно подать сигнал выключения питания. Процедура включения домена должна идти в логически обратном порядке: включение питания; возможно, сброс логических элементов домена, затем восстановление состояния регистров из регистров сохранения и отключение блоков изоляции. Также управляющая сменой режима питания домена последовательность может быть дополнена сигналами включения-выключения пропуска синхросигнала.

Список энергетических состояний системы. Стратегия управления состояниями энергопотребления характеризуется списком всех возможных для конкретной системы глобальных состояний питания (включено, сон, выключено и т. п.). Каждое состояние питания определяется неизменным во времени набором состояний всех доменов системы для такого глобального состояния. Состояние домена определяется состоянием портов и цепей питания домена (включено или выключено, текущее напряжение). Изменение списка глобальных энергетических состояний может потребовать изменения набора необходимых элементов обеспечения энергетической эффективности. Например, нет необходимости в блоках изоляции между доменами, которые отключаются одновременно для любого глобального состояния энергопотребления системы.

Представленное описание элементов системы с разделением на домены предполагает некоторую обособленность и независимость логической составляющей проекта от элементов, обеспечивающих работу подсистемы питания. Однако если разработка логических компонентов проекта осуществляется с учетом последующего внедрения в доменную структуру питания, то конечная структура логического компонента, изначально целостного, может быть впоследствии разделена на составляющие, предназначенные для разных доменов: часть структуры компонента может быть отключаемой или может питаться от другого источника питания.

В общем случае разработка поведенческого описания энергетически эффективной доменной системы включает несколько этапов [10]:

1. Спецификация логики средствами RTL и индивидуальная верификация RTL-компонентов.
2. Определение доменов питания и распределение логических элементов в них.
3. Спецификация списка режимов питания.
4. Спецификация необходимых элементов обеспечения энергетической эффективности для доменов (переключателей, изоляции, регистров сохранения и т. п.).
5. Спецификация компонентов логики управления (описанной на RTL) для элементов обеспечения энергетической эффективности.
6. Верификация всей системы с учетом всех компонентов как логической, так и энергетической составляющих проекта.

3. Основные положения технологии UPF

Представленные выше элементы, позволяющие описать доменную структуру проекта, являются, по сути, поведенческими элементами, описывающими логическую структуру энергетической подсистемы проекта. Эта структура, в свою очередь, тесно связана с логической составляющей проекта: с одной стороны, подсистема питания управляется логической составляющей, с другой стороны, состояния элементов подсистемы питания влияют на работоспособность логической составляющей проекта. Однако текущие версии языков проектирования логических схем VHDL и System Verilog не имеют никаких специальных средств для описания (а значит, верификации и синтеза) проекта с учетом цепей питания. Ввиду тесной взаимосвязанности логической (HDL) и энергетической составляющих уже на поведенческом уровне описания системы очевидна необходимость введения специального расширения языка, позволяющего описать такую систему уже на одном уровне с RTL.

Язык (формат) описания UPF, решающий данную проблему, был впервые представлен в стандарте UPF 1.0 [11] организацией Accellera Systems Initiative в 2007 г. В 2009 г. организация IEEE дополнила формат UPF в стандарте IEEE 1801–2009 (UPF 2.0) [12]. Последней версией стандарта является IEEE 1801–2013 (UPF 2.1) [13].

Формат описания UPF – это специальный формат, предназначенный для поведенческого описания энергетической составляющей проекта электронной системы или ее компонентов. Формат предоставляет возможность описания сети питания проекта: цепей питания, энергетических доменов, переключателей, средств изоляции, средств сохранения состояния и других элементов, относящихся к энергетической составляющей проекта, непосредственно задаваемой разработчиком. UPF позволяет формально описать элементы подсистемы питания в их взаимосвязи с логическими элементами, описанными на языках проектирования VHDL и System Verilog. Важно отметить, что стандарт UPF изначально не предназначен для описания всей совокупности схемотехнических и электрических подробностей цепей питания, необходимых для аналоговой симуляции и верификации синтезированного проекта: UPF описывает структуру питания на поведенческом уровне. Однако UPF-описания в связке с HDL-описаниями могут быть использованы в качестве входных данных как при симуляции и верификации проекта [9, 14], так и при его синтезе (рис. 1) [13]. При синтезе проекта абстрактные модели объектов подсистемы питания отображаются на соответствующие физические ячейки.

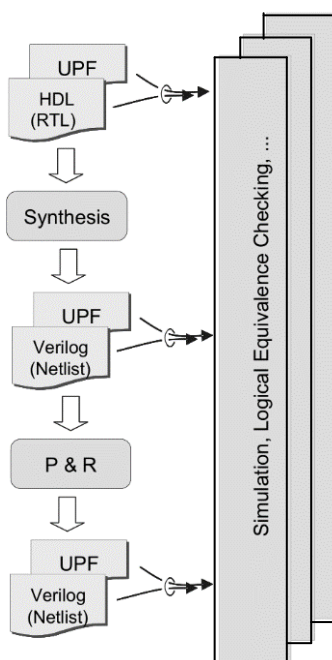


Рис. 1. Процесс проектирования системы с применением UPF-описаний

Стандарт UPF описывает четыре основные составляющие [10]:

1. *Команды описания энергетической составляющей (Commands)*. Команды языка UPF представляют собой расширение сценарного языка программирования Tcl [15]. Последовательность компактных команд-директив, записанная в файл UPF-описания с расширением .upf, поэтапно (инкрементально) описывает все элементы архитектуры энергетической составляющей проекта. Ключевой особенностью UPF-описаний является то, что данные описания никаким образом не изменяют HDL-описания логических элементов системы. Логические HDL-компоненты можно по-прежнему верифицировать отдельно от энергетической составляющей. С другой стороны, для одного и того же HDL-компонента или набора компонентов можно иметь несколько возможных UPF-описаний для различных случаев использования.

2. *Команды запроса информации о состоянии энергетической архитектуры проекта (Queries)*. Данные Tcl-команды, позволяющие облегчить процесс отладки UPF-описаний, генерируют детальные отчеты о состоянии UPF-элементов проекта.

3. *Пакет UPF (Package UPF)*. Пакет UPF, разработанный как для VHDL, так и для System Verilog, определяет типы и функции, необходимые для описания тестов, которые управляют как логическими портами, так и портами питания тестируемого модуля. Эти типы и функции также применяются для проектирования HDL-моделей, описывающих поведение энергетической составляющей проекта. Например, используя типы и функции из пакета UPF, можно описать поведенчески специфическую VHDL-модель переключателя питания, входными и выходными портами которого являются порты питания наравне с логическими портами.

Пакет UPF определяет два основных типа данных, позволяющих моделировать источники питания:

state := (OFF | UNDETERMINED | PARTIAL_ON | FULL_ON) – состояние источника питания;

supply_net_type := (state, voltage) – состояние (state) цепи питания в совокупности с уровнем напряжения (voltage).

Кроме того, пакет UPF определяет ряд функций для управления портами и цепями питания и контроля над ними.

Некоторые функции управления состоянием и напряжением порта (Supply Port) питания из теста (заголовки функций отображены в Си-подобном псевдокоде): supply_on(string port_name, real voltage), supply_off(string port_name).

Некоторые функции опроса состояния и напряжения порта питания:

get_supply_value(string port_name) – возвращает тип supply_net_type;

get_supply_state(supply_net_type arg) – возвращает тип state;

get_supply_voltage(supply_net_type arg) – возвращает тип voltage.

4. *Формат SAIF (Switching Activity Interchange Format)*. Предназначен для стандартизации представления информации об активности переключений элементов, генерируемой симулятором по окончании симуляции. Результат помещается в «обратный» backward-SAIF-файл. Этот файл представляет входные данные для программных средств анализа и оптимизации энергопотребления. Программа автоматизации проектирования может предварительно сгенерировать «прямой» forward-SAIF-файл, содержащий некоторые директивы, определяющие содержимое выходного backward-SAIF-файла. Важно отметить, что формат SAIF, являющийся частью стандарта IEEE 1801, служит лишь типовым шаблоном представления информации о переключательной активности, при этом в стандарте IEEE 1801 не указаны алгоритмы или правила для эффективного определения действительной переключательной активности: данная задача выходит за рамки стандарта IEEE 1801. Предполагается, что при разработке блока будет разработана тестовая входная последовательность, соответствующая типовым условиям функционирования этого блока; после окончания данной последовательности SAIF-совместимая САПР должна сгенерировать информацию о переключательной активности в виде файла формата SAIF. Эта информация позволит оценить динамическую мощность, рассеиваемую блоком,

с помощью любого SAIF-совместимого пакета программных средств анализа и оптимизации энергопотребления.

Важнейшей особенностью процесса симуляции проекта с UPF-описаниями является то, что все процессы HDL-логики домена с выключенным основным питанием особым образом завершают процессы симуляции: содержимое регистров становится неопределенным, а все внутренние и выходные сигналы по цепочке переходят в неопределенное состояние (X).

На сегодняшний день язык UPF поддерживается во всех ведущих системах автоматизированного проектирования и симуляции от таких компаний, как Mentor Graphics, Synopsys, Cadence Design Systems и др. Однако на данный момент UPF не является единственным поддерживаемым языком описания энергетически эффективных систем. Конкуренцию UPF составляет стандарт Common Power Format [16], который во многом совпадает с UPF, но имеет от него и некоторые отличия [3, 17].

4. Группы команд UPF

Стандарт IEEE 1801–2013 [13] содержит исчерпывающее описание команд UPF с примерами их использования, однако для лучшего понимания возможностей языка UPF целесообразно проанализировать данные команды и сгруппировать их по смысловому назначению. Так, команды описания энергетической составляющей проекта (Commands) UPF можно условно разделить по функциональному назначению на несколько групп (рис. 2) [10].



Рис. 2. Группы команд UPF

Команды навигации определяют, к какой части HDL-иерархии проекта будут относиться последующие UPF-команды.

Команды доменов используются для сегментирования проекта на отдельные участки питания, определяемые совокупностью логических компонентов (экземпляров RTL-компонентов) и совокупностью основного и специальных источников питания.

Команды описания разрешенных состояний энергопотребления используются для автоматизированной верификации энергетической системы.

Команды описания стратегий определяют правила размещения и управления элементами изоляции и элементами изменения уровня напряжения сигналов на границе домена питания, правила размещения и управления элементами сохранения состояния в пределах домена питания.

Команды имплементации определяют, каким образом абстрактные поведенческие элементы подсистемы питания отображаются на заданные библиотечные ячейки элементов питания в процессе автоматического синтеза проекта.

Команды сопряжения с HDL позволяют добавлять в проект UPF-осведомленные проверочные утверждения. Эти команды предназначены также для операций преобразования типов портов питания в логические типы HDL и наоборот.

Команды статуса симуляции (Simstate-команды) добавлены в UPF 2.0 для расширения возможностей по детальному определению статуса объектов симуляции (добавлены частично рабочие состояния доменов и сетей питания).

Команды групп и атрибутов цепей питания (Supply-Sets-команды) добавлены в UPF 2.0 для оптимизации описания и подключения сетей питания.

Команды описания логики управления добавлены в UPF 2.0 для некоторых ситуаций моделирования, в которых удобнее описывать логические сигналы управления элементами энергетической составляющей проекта в пределах UPF-файлов вместо HDL-файлов (команды не предназначены для полной замены HDL-описаний).

Важно отметить, что часть команд, представленных или дополненных в UPF 2.0, замещает по функциональности некоторые устаревшие команды из подмножества команд, представленных в UPF 1.0. В UPF 2.1 данные устаревшие команды полностью исключаются, в связи с чем обратная совместимость с предыдущими версиями невозможна. Ввиду новизны стандарта UPF 2.1 актуальной версией можно считать UPF 2.0 [10], однако использовать команды UPF 2.0, исключенные из UPF 2.1, не рекомендуется.

5. Пример разработки проекта с применением технологии UPF

Для того чтобы показать принципы проектирования энергетически эффективной системы с применением UPF-описаний, был разработан демонстрационный проект. Файлы RTL-описаний (System-Verilog-файлы `upf_demo.sv` и `tb.sv`) и UPF-описания (`upf_demo.upf`), а также файл сценария запуска `run.do` для среды Mentor Graphics QuestaSim 10.2 [18] доступны по ссылке [19]. Схемы логического проекта показаны на рис. 3. При разработке проекта использовалась методика, представленная в разд. 3. Моделирование проекта с учетом энергетической составляющей осуществлено на поведенческом уровне.

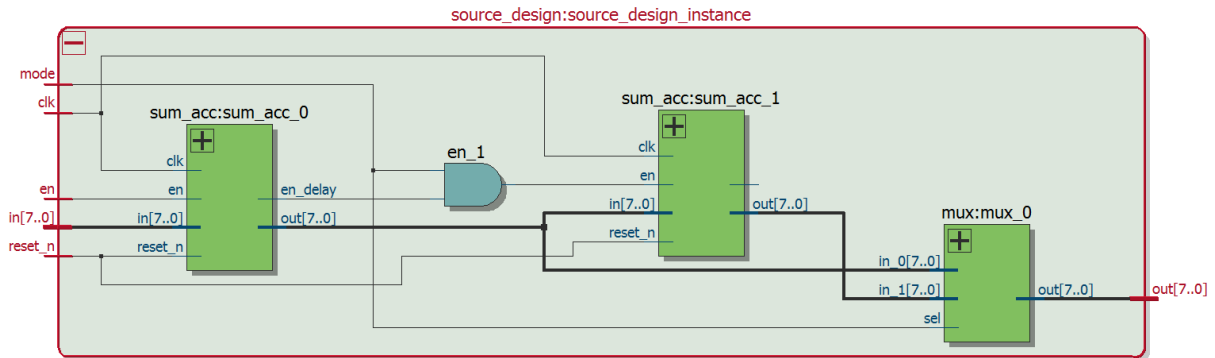
Под регистром здесь и далее понимается элемент памяти произвольной разрядности.

Основой энергетически эффективной электронной системы является иерархия ее логической части, которая некоторым оптимальным образом распределяется в отдельные домены с различными стратегиями питания. Логическая структура исходного проекта (без энергетической составляющей – системы питания) представлена модулем `source_design` (экземпляр `source_design_instance`) и вложенными в него компонентами (рис. 3, *a*). Основным логическим компонентом демонстрационного проекта выбран модуль `sum_acc`, который по фронту сигнала синхронизации при `en = 1` прибавляет входные данные порта `in` к текущему содержимому внутреннего регистра `acc[7:0]`. Содержимое регистра `acc` выводится в порт `out[7:0]`. Вывод `en_delay` является задержанным на один такт значением входного порта `en` (внутренний регистр `en_d`). Экземпляры модуля `sum_acc` `sum_acc_0` и `sum_acc_1` поставлены в каскад, где выход `sum_acc_0.out` подключен к входному порту `in` экземпляра `sum_acc_1`, а выходной порт `sum_acc_0.en_delay` подключен (через элемент И) к входному порту `en` экземпляра `sum_acc_1`. Мультиплексор `mux_0` по сигналу от порта `source_design_instance.mode` выбирает (через порт `mux_0.sel`), выход `out` какого из экземпляров `sum_acc` будет направлен на общий вывод `source_design_instance.out[7:0]`.

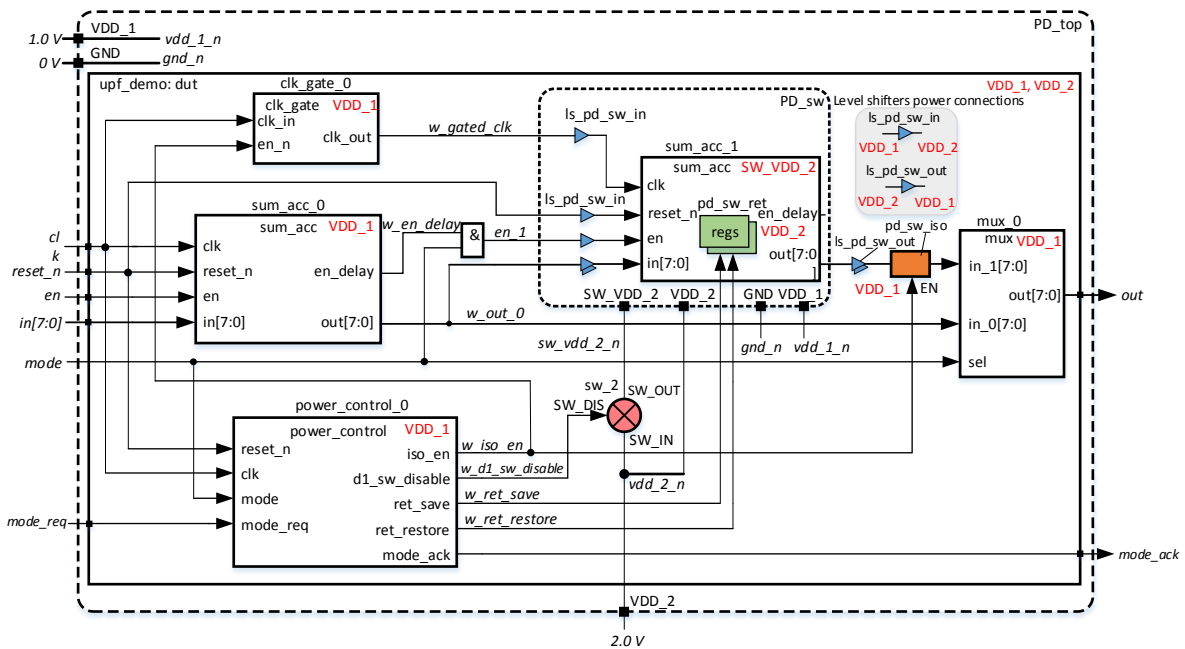
Таким образом, при `source_design_instance.mode = 1` и после изменения регистра `acc` экземпляра `sum_acc_0` в следующем такте ввиду `w_en_delay = 1` значение регистра `acc` экземпляра `sum_acc_1` будет увеличено на очередное значение регистра `sum_acc_0.acc`, а новое значение `sum_acc_1.acc` будет подано на вывод `source_design_instance.out`.

В режиме `source_design_instance.mode = 0`, при котором на вывод `source_design_instance.out` передается значение `sum_acc_0.out`, элемент И, к которому подключен

sum_acc_0.en_delay, обеспечивает сохранение состояния регистра sum_acc_1.acc. Таким образом, в режиме mode = 0 экземпляр sum_acc_1 полностью не задействован, а значит, sum_acc_1 является частным случаем компонента, который может быть отключен от питания в одном из режимов работы (при условии сохранения содержимого sum_acc_1.acc).



а)



б)

Рис. 3. Логические схемы: а) без энергетической составляющей; б) с учетом системы питания

После получения основной логической части RTL-описания проекта, представленной в модуле source_design на рис. 3, а, и соблюдения дополнительных условий его функционирования проект можно разделить на следующие домены (рис. 3, б): выключаемый домен PD_sw, содержащий экземпляр sum_acc_1, и невыключаемый домен PD_top, содержащий все остальные логические компоненты source_design. В демонстрационных целях (для обеспечения необходимости установки блоков изменения уровня сигнала) домены питания проекта с энергетической составляющей используют различные источники напряжения питания: 1.0 V для PD_top (порт питания VDD_1) и 2.0 V для PD_sw (порт питания VDD_2). Для PD_sw используется переключатель питания (ключ) sw_2 с выходным портом питания SW_OUT. Входной порт ключа SW_IN питается от напряжения 2.0 V (VDD_2). SW_OUT является основным источником питания для PD_sw, а VDD_1 – основным источником питания для PD_top. Оба домена используют общий порт земли GND.

После выделения доменов питания системы необходимо составить таблицу всех возможных одновременных состояний доменов системы, выраженных через состояния портов питания. По таблице можно выделить полный набор необходимых элементов обеспечения энергетической эффективности: блоки изоляции, сохранения, изменения уровня напряжения сигналов, блоки отключения синхросигнала. Выключение цепей питания в таблице также сигнализирует о необходимости постановки переключателей питания.

Состояние питания	VDD_1 (PD_top)	VDD_2	sw_2/SW_OUT (PD_sw)	GND
FULL_ON	ON 1.0 V	ON 2.0 V	ON 2.0 V	ON 0 V
PART_ON	ON 1.0 V	ON 2.0 V	OFF	ON 0 V
FULL_OFF	OFF	OFF	OFF	ON 0 V

В режиме питания FULL_ON домены PD_top и PD_sw питаются от различных источников напряжения питания. Для правильной интерпретации логических сигналов, связывающих эти домены (ON 1.0 V \rightarrow ON 2.0 V и ON 2.0 V \rightarrow ON 1.0 V), необходимо ввести блоки изменения уровня напряжения сигнала как для входных, так и для выходных сигналов PD_sw (ls_pd_sw_in и ls_pd_sw_out на рис. 3, б). Входы и выходы модулей изменения уровня сигнала питаются от источников питания VDD_1 и VDD_2 в соответствии с направлением изменения уровня напряжения сигналов.

В режиме питания PART_ON домен PD_sw находится в отключенном состоянии (OFF). Согласно описанным выше требованиям состояние всех внутренних регистров модуля sum_acc_1 должно сохраняться, поэтому в домен PD_sw необходимо добавить регистры сохранения для sum_acc_1.acc и sum_acc_1.en_d (pd_sw_ret на рис. 3, б). Регистры сохранения питаются от включенного порта питания VDD_2.

В режиме питания PART_ON сигналы домена PD_top передаются в отключенный домен PD_sw. Во избежание негативных эффектов от такого подключения выходной порт sum_acc_1.out подключается к блоку изоляции (сигнал от порта sum_acc_1.out принимает значение, которое предшествовало включению изоляции). Выходной порт sum_acc_1.en_delay является неподключенным, а значит, изоляция для него не требуется. Блоки изоляции питаются от порта VDD_1.

Ввиду незадействованности экземпляра sum_acc_1 в режиме PART_ON сигнал синхронизации clk для этого компонента отключается с помощью блока отключения clk_gate_0 (RTL-схема отключения пропускания синхросигнала с защелкой).

После того как для проекта определены все переключатели питания, элементы изоляции и блоки сохранения, а также блоки отключения сигнала синхронизации, необходимо описать логику управления этими элементами посредством RTL. Модуль power_control_0, показанный на рис. 3, а, является модулем управления элементами питания для всего upf_demo:dut. По сигналу mode_req конечный автомат компонента power_control_0 начинает в соответствии с таблицей изменять состояние сигнала выключения переключателя питания w_d1_sw_disable, включения изоляции w_iso_en и сигналов w_ret_save и w_ret_restore для сохранения и восстановления данных в блоках сохранения. В процессе выключения домена PD_sw модуль управления питанием сначала включает изоляцию, затем сохраняет данные в регистры сохранения и только потом выключает sw_2. По завершении процесса смены режима питания модуль сигнализирует dut.mode_ack = 1. Процесс включения PD_sw идет в обратном порядке с учетом восстановления состояний регистров sum_acc_1 из регистров сохранения. Для упрощения схемы в качестве сигнала выключения синхросигнала w_gated_clk выбран сигнал w_iso_en. Важно отметить, что для реальных проектов процессы включения и выключения домена могут занимать длительное время ввиду инертности реальных физических компонентов системы питания (в демонстрационном проекте это не показано).

```

# Определение главного модуля:
set_design_top upf_demo
set_scope .

# Создание доменов PD_top: все элементы главного
модуля
# включаются PD_top:
create_power_domain PD_top \
  -include_scope
# Создание домена PD_sw,
# перенос sum_acc_1 в домен PD_sw:
create_power_domain PD_sw \
  -elements {sum_acc_1}

# Создание портов питания:
create_supply_port VDD_1
create_supply_port VDD_2
create_supply_port GND

# Создание цепей питания:
create_supply_net vdd_1_n \
  -domain PD_top
create_supply_net vdd_2_n \
  -domain PD_top
create_supply_net gnd_n \
  -domain PD_top

# Подключение цепей питания к портам:
connect_supply_net vdd_1_n \
  -ports VDD_1
connect_supply_net vdd_2_n \
  -ports VDD_2
connect_supply_net gnd_n \
  -ports GND

# Группировка цепей питания в набор pwr_1_ss:
create_supply_set pwr_1_ss \
  -function {power vdd_1_n} \
  -function {ground gnd_n}
# Назначение группы pwr_1_ss в
# качестве основного набора питания для PD_top:
associate_supply_set pwr_1_ss \
  -handle PD_top.primary

# Группировка цепей питания в набор pwr_2_ss:
create_supply_set pwr_2_ss \
  -function {power vdd_2_n} \
  -function {ground gnd_n}

# Создание группы цепей питания
# для подключения к PD_sw:
create_supply_net sw_vdd_2_n \
  -domain PD_top
create_supply_set sw_pwr_2_ss \
  -function {power sw_vdd_2_n} \
  -function {ground gnd_n}
# Назначение группы sw_pwr_2_ss в качестве
# основного набора питания для PD_sw:
associate_supply_set sw_pwr_2_ss \
  -handle PD_sw.primary

# Создание модели ключа sw_2, управляемого
# RTL-сигналом w_d1_sw_disable:
create_power_switch sw_2 \
  -domain PD_sw \
  -input_supply_port \
    {SW_IN pwr_2_ss.power} \
  -output_supply_port \
    {SW_OUT sw_pwr_2_ss.power} \
  -control_port \
    {SW_DIS w_d1_sw_disable} \
  -on_state \
    {ON_STATE SW_IN {!SW_DIS}} \
  -off_state {OFF_STATE {SW_DIS}}

# Создание стратегии по автоматическому
# добавлению входных для PD_sw блоков изменения
# уровня напряжения сигналов:
set_level_shifter ls_pd_sw_in \
  -domain PD_sw \
  -applies_to inputs \
  -rule low_to_high \
  -location self

# Подключение питания к входным блокам изменения
# уровня напряжения сигналов:
associate_supply_set pwr_1_ss \
  -handle PD_sw.ls_pd_sw_in.input
associate_supply_set pwr_2_ss \
  -handle PD_sw.ls_pd_sw_in.output
# Создание стратегии по автоматическому
# добавлению выходных для PD_sw блоков изменения
# уровня напряжения сигналов:
set_level_shifter ls_pd_sw_out \
  -domain PD_sw \
  -applies_to outputs \
  -rule high_to_low \
  -location parent

# Подключение питания к выходным блокам изменения
# уровня напряжения сигналов:
associate_supply_set pwr_2_ss \
  -handle PD_sw.ls_pd_sw_out.input
associate_supply_set pwr_1_ss \
  -handle PD_sw.ls_pd_sw_out.output

# Создание стратегии по добавлению блоков изоляции
# с точным указанием изолируемых элементов:
set_isolation pd_sw_iso \
  -domain PD_sw \
  -clamp_value latch \
  -isolation_signal w_iso_en \
  -isolation_sense high \
  -location parent \
  -elements {sum_acc_1/out}
# Подключение питания к блокам изоляции:
associate_supply_set pwr_1_ss \
  -handle PD_sw.default_isolation

# Создание стратегии по автоматическому добавлению
# блоков сохранения для всех регистров
# домена PD_sw (управление RTL-сигналами
# w_ret_save и w_ret_restore):
set_retention pd_sw_ret \
  -domain PD_sw \
  -save_signal {w_ret_save posedge} \
  -restore_signal \
    {w_ret_restore posedge}
# Подключение питания к регистрам сохранения:
associate_supply_set pwr_2_ss \
  -handle PD_sw.default_retention

# Определение состояний портов питания
# (согласно таблице 1):
add_port_state VDD_1 \
  -state {ON_1 1.0} \
  -state {OFF_ST OFF}
add_port_state VDD_2 \
  -state {ON_2 2.0} \
  -state {OFF_ST OFF}
add_port_state sw_2/SW_OUT \
  -state {ON_2 2.0} \
  -state {OFF_ST OFF}
add_port_state GND \
  -state {ON_0 0}
# Описание таблицы 1 в формате UPF (таблица
# используется для верификации работы блока
# управления питанием):
create_pst DEMO_PST \
  -supplies {VDD_1 VDD_2 sw_2/SW_OUT GND}
add_pst_state FULL_ON -pst DEMO_PST \
  -state {ON_1 ON_2 ON_2 ON_0}
add_pst_state PART_ON -pst DEMO_PST \
  -state {ON_1 ON_2 OFF_ST ON_0}
add_pst_state FULL_OFF -pst DEMO_PST \
  -state {OFF_ST OFF_ST OFF_ST ON_0}

```

Рис. 4. Описание энергетической составляющей проекта. Файл upf_demo.upf

Следующим шагом является формальное поведенческое описание полученной выше информации об энергетической составляющей на языке UPF. Результат поведенческого проектирования (описания) доменной системы для демонстрационного проекта представлен в файле upf_demo.upf [19] (рис. 4). Важно отметить, что в RTL-описании модуля upf_demo:dut (файл upf_demo.sv) сигналы w_d1_sw_disable, w_iso_en, w_ret_save и w_ret_restore, выходящие из модуля power_control_0, являются неподключенными, но они явно подключаются к соответствующим элементам обеспечения энергетической эффективности уже в файле upf_demo.upf.

Моделирование демонстрационного проекта было осуществлено на поведенческом уровне в среде Mentor Graphics QuestaSim 10.2 в режиме Power Aware Simulation [20]. Временная диаграмма функционирования проекта показана на рис. 5. В ходе теста (System-Verilog-файл tb.sv [19]) система отработала в составе полного каскада (все домены включены). Затем был послан запрос на выключение домена PD_sw (сигнал mode = 0 → 1, mode_req = 1), после чего система отработала с отключенным от питания модулем sum_acc_1. Регистры sum_acc_1.acc и sum_acc_1.en_d также были отключены от питания (на диаграмме это показано диагональной решеткой). Перед выключением питания состояния этих регистров были сохранены по сигналу w_ret_save, а после включения питания были восстановлены по сигналу w_ret_restore. Сплошная прозрачная заливка для сигнала от порта sum_acc_1.out на диаграмме показывает включение изоляции: вывод sum_acc_1.out изолируется от всех изменений регистра sum_acc_1.acc в процессе включения и выключения питания.

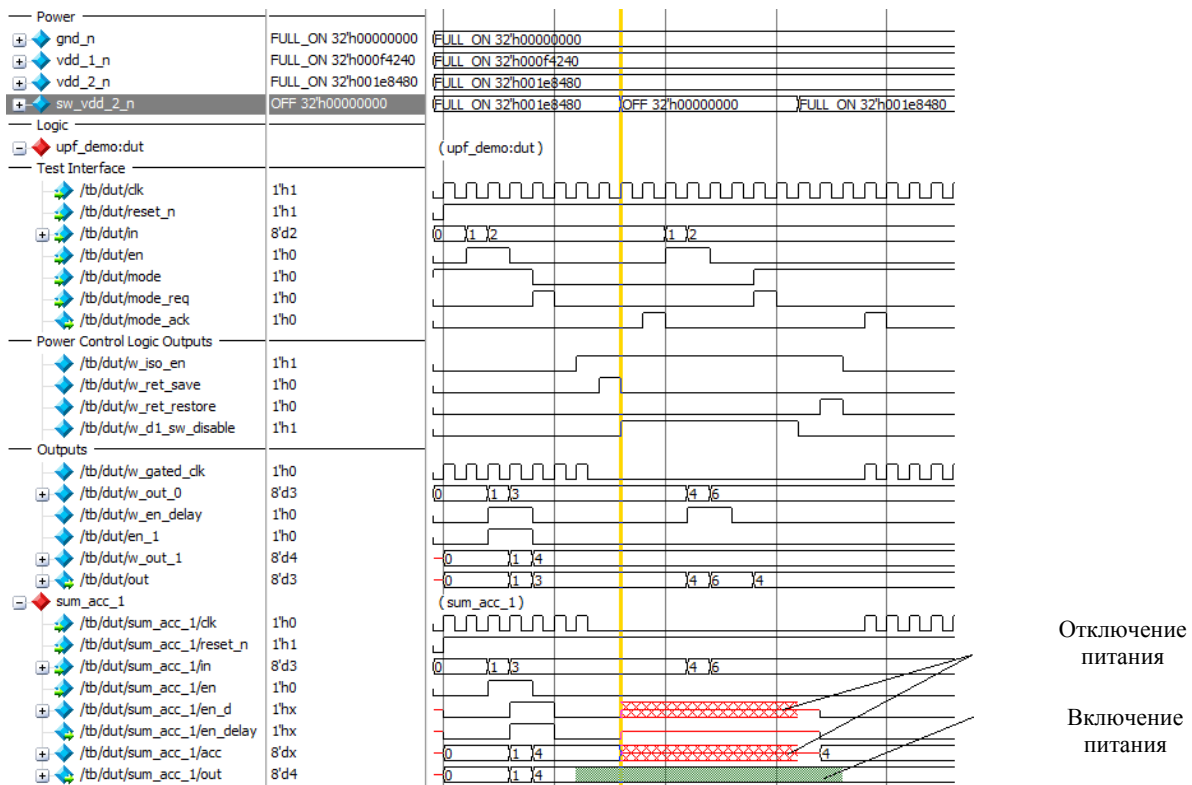


Рис. 5. Временная диаграмма симуляции проекта

Основная логика демонстрационного проекта была спроектирована таким образом, чтобы обусловить возможность разделения проекта на домены питания, причем один из доменов питания сделан отключаемым. Было осуществлено разделение системы на конкретные домены питания с различными напряжениями питания, определены все возможные состояния этих доменов через сети питания. Из таблицы состояний получен список всех переключателей питания, блоков изоляции, сохранения и изменения уровня сигнала. После этого было разработано формальное поведен-

ческое описание энергетической подсистемы проекта в виде UPF-файла. Временная диаграмма, полученная в среде симуляции, показывает специфику изменений сигналов при переключениях энергетических состояний проекта.

Дальнейшая разработка проекта предполагает синтез RTL-логики и отображение абстрактных моделей переключателей питания, изоляции, регистров сохранения и изменения уровня на реальные физические библиотечные ячейки (UPF-команды имплементации “map_”). Важно отметить, что оценить эффект от разбиения системы на домены с различными источниками питания, часть из которых является отключаемой, можно только после синтеза проекта с использованием конкретных библиотек логических элементов. Действительно, для рассмотренного демонстрационного проекта энергия токов утечки, сохраненная во время отключения домена PD_sw, для какого-то из результатов синтеза может быть меньше, чем энергия, затраченная на постоянную работу модуля управления power_control и других элементов обеспечения энергетической эффективности, которые были добавлены в систему для управления выключаемым доменом PD_sw. Однако, как правило, в реальных системах отключаемые домены питания содержат гораздо больше компонентов логики, в то время как затраты на элементы поддержки и управления отключаемыми доменами остаются незначительными [2]. Общая же методология проектирования для реальных проектов будет соответствовать методологии, показанной в данном демонстрационном проекте.

Заключение

Одной из самых эффективных мер снижения энергопотребления для систем на кристалле, производимых по современным технологическим процессам, является проектирование таких систем с активной доменной системой питания. Необходимы дополнительные средства поведенческого описания доменной подсистемы питания ввиду недостаточности возможностей языков VHDL и System Verilog. В качестве средства описания подсистемы питания был рассмотрен формат UPF, приведен пример описания системы с применением средств этого формата.

На примере схемы с отдельными доменами питания было показано, как исходное RTL-описание логической структуры проекта можно дополнить элементами доменной системы питания, используя средства формата UPF. Моделирование демонстрационного проекта было осуществлено в UPF-совместимой САПР с учетом работы компонентов системы питания.

Формат UPF помогает описывать системы с активной доменной системой питания, однако выделение доменов питания и управление ими в системе является сложной научной и инженерной задачей, лежащей на плечах разработчиков [2]. Так, для современных технологических процессов производства микросхем многие существующие сложные вычислительные блоки могут быть разделены на отдельные домены питания, что потребует тщательного пересмотра структуры таких компонентов для максимального энергосбережения. В связи с этим формат UPF является инструментом, позволяющим значительно облегчить процесс проектирования новейшего класса цифровых систем – систем с активной доменной системой питания.

Список литературы

1. A Practical Guide to Low-Power Design. User Experience with CPF [Electronic resource]. – 2009. – Mode of access : http://www.si2.org/events_dir/2009/PowerForward/LowPowerGuide09232009/pfi_lpg_chapters/lpg_sect1_06052009.pdf. – Date of access : 19.03.2015.
2. Low Power Methodology Manual: For System-on-Chip Design / D. Flynn [et al.]. – N.Y. : Springer, 2011. – 300 p.
3. Chadha, R. An ASIC Low Power Primer: Analysis, Techniques and Specification / R. Chadha, J. Bhasker. – N.Y. : Springer, 2015. – 232 p.

4. Rabaey, J. Low Power Design Essentials / J. Rabaey. – N.Y. : Springer, 2009. – 288 p.
5. Рабаи, Ж.М. Цифровые интегральные схемы / Ж.М. Рабаи, А. Чандракасан, Б. Николич. – 2-е изд. – М. : Вильямс, 2007. – 912 с.
6. Power-efficient System Design / P.R. Panda [et al.]. – N.Y. : Springer, 2011. – 253 p.
7. IEEE Standard VHDL Language Reference Manual. IEEE 1076 – 2002 [Electronic resource]. – 2002. – Mode of access : <http://ieeexplore.ieee.org/servlet/opac?punumber=7863>. – Date of access : 19.03.2015.
8. IEEE Standard for SystemVerilog: Unified Hardware Design, Specification, and Verification Language. IEEE 1800 – 2012 [Electronic resource]. – 2012. – Mode of access : <http://standards.ieee.org/findstds/standard/1800-2012.html>. – Date of access : 19.03.2015.
9. Low Power Verification Methodology Using UPF / F. Bembaron [et al.] // DVCon 2011 [Electronic resource]. – 2011. – Mode of access : http://2011.dvcon.org/file/11.3_Paper.pdf. – Date of access : 19.03.2015.
10. Marschner, E. Power Aware Verification / E. Marschner, C. Seeley // Mentor Graphics Verification Academy [Electronic resource]. – 2013. – Mode of access : <https://verificationacademy.com/courses/power-aware-verification>. – Date of access : 19.03.2015.
11. Unified Power Format (UPF) Standard. Version 1.0 [Electronic resource]. – 2007. – Mode of access : <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=B47D579C08C9E70D33FF5F2A9DA15DEC?doi=10.1.1.133.6194&rep=rep1&type=pdf>. – Date of access : 19.03.2015.
12. IEEE Standard for Design and Verification of Low Power Integrated Circuits: IEEE Std 1801 – 2009 [Electronic resource]. – 2009. – Mode of access : <http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?reload=true&punumber=4809843>. – Date of access : 19.03.2015.
13. IEEE Standard for Design and Verification of Low Power Integrated Circuits: IEEE Std 1801 – 2013 [Electronic resource]. – 2013. – Mode of access : <http://standards.ieee.org/getieee/1801/download/1801-2013.pdf>. – Date of access : 19.03.2015.
14. Verification Methodology Manual for Low Power / S. Jadcherla [et al.]. – N. Y. : Springer, 2011. – 226 p.
15. Tcl/Tk Documentation // Tcl Developer Xchange [Electronic resource]. – 2015. – Режим доступа : <http://www.tcl.tk/doc/>. – Дата доступа : 19.03.2015.
16. Si2 Common Power Format (CPF) // Si2 [Electronic resource]. – 2009. – Mode of access : http://www.si2.org/openeda.si2.org/project/showfiles.php?group_id=51. – Date of access : 19.03.2015.
17. Power Intent Formats: Light at the End of the Tunnel? / S. Dobre [et al.] // EETimes [Electronic resource]. – 2012. – Mode of access : http://www.eetimes.com/document.asp?doc_id=1279413. – Date of access : 19.03.2015.
18. Questa Advanced Simulator [Electronic resource]. – 2015. – Mode of access : <http://www.mentor.com/products/fv/questa/>. – Date of access : 19.03.2015.
19. UPF Demo Project [Electronic resource]. – 2015. – Mode of access : https://www.dropbox.com/s/ogty44li1w67pxz/upf_demo.zip?dl=0. – Date of access : 19.03.2015.
20. Power Aware Simulation User's Manual. Questa SIM. Software Version 10.2c [Electronic resource]. – 2013. – Mode of access : https://www.dropbox.com/s/r43dfuqkgada6r9/questa_sim_pa.pdf?dl=0. – Date of access : 19.03.2015.

Поступила 09.06.2015

*Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: alexey.shashkov@gmail.com*

A.S. Shashkov

**DESIGN OF LOW-POWER ELECTRONIC SYSTEMS USING
UPF POWER INTENT SPECIFICATION TECHNOLOGY**

In this review paper, the problem of the design of energy-efficient electronic systems is considered. Different energy-saving methods are presented. The necessity of additional tools of formal behavioral description of the power subsystem is explained. The elements of the power domain energy subsystem are explained, the design stages for the system with power domains are presented. Unified Power Format (UPF) is presented as a tool to describe power intent in a low-power system. An example system with the UPF-description is presented.

ЗАЩИТА ИНФОРМАЦИИ

УДК 681.3; 004.056.53

**В.К. Фисенко, Е.П. Максимович, А.Б. Степанян,
В.А. Дмитриев, Д.С. Ким, Т.С. Мартинович**

**МЕТОДИКА ЭКСПЕРТНОЙ ОЦЕНКИ ПОЛНОТЫ И КАЧЕСТВА
ИСХОДНЫХ ДАННЫХ СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ**

Анализируется проблема аттестации систем защиты информации, рассматриваются задачи анализа исходных данных по аттестуемой системе защиты информации. Формулируются цели, показатели и критерии принятия решений. Показывается, что для повышения качества, снижения сроков и затрат на подготовку к аттестации целесообразно использовать программный комплекс автоматизации процесса анализа исходных данных, предоставляемых владельцем информационной системы.

Введение

Аттестация системы защиты информации (СЗИ) информационной системы (ИС) – комплекс организационно-технических мероприятий, в результате которых документально подтверждается соответствие СЗИ ИС требованиям законодательства об информации, информатизации и защите информации [1]. Реализация мероприятий, предусмотренных в соответствии с процедурой аттестации, связана с необходимостью рассмотрения целого ряда нетривиальных задач оценки СЗИ в реальных условиях эксплуатации ИС и, как следствие, приводит к возникновению общей проблемы аттестации. Актуальность данной проблемы обусловлена требованиями законодательства Республики Беларусь [2] и необходимостью обеспечения информационной безопасности ИС, хранящих, обрабатывающих или передающих важные научно-технические и народнохозяйственные данные, а также сложностью и многоаспектностью процесса проведения аттестации. В частности, быстрое развитие и появление все более крупных и сложных ИС, непрерывное увеличение объемов внутренних и внешних потоков информации, с одной стороны, и непрерывное совершенствование методов и средств информационных атак, с другой стороны, требуют периодического проведения повторной аттестации СЗИ. Вместе с тем необходимо учитывать при аттестации СЗИ и специфику ИС: архитектурные особенности, сложность, многокомпонентность, гетерогенность, возможную распределенность, а также наличие в ИС разных по критичности доменов, требующих разного уровня защиты, наличие сложного прикладного программного обеспечения по обработке данных, содержащих важную информацию, крупных баз данных, удаленных пользователей и т. д.

Основные положения аттестации СЗИ ИС изложены в приказах Оперативно-аналитического центра при Президенте Республики Беларусь, в которых определены этапы подготовки и проведения аттестации [3]. На этапе подготовки к аттестации предусмотрено представление владельцем ИС комплекта исходных данных по аттестуемой СЗИ ИС в специализированную организацию, которая проводит оценку исходных данных по показателям полноты и качества. Согласно результатам оценки принимается одно из решений:

1. Представленные исходные данные удовлетворяют установленным требованиям. СЗИ ИС может быть представлена на аттестацию.
2. Представленные исходные данные не удовлетворяют установленным требованиям. Отказ в аттестации СЗИ ИС.

Подготовка к аттестации – это не прихоть отдельного человека, а объективная необходимость. Государство затрачивает огромные средства на обеспечение информационной безопасности ИС, на периодическую подготовку и аттестацию систем защиты информации и заинтересовано в том, чтобы эти важные и дорогостоящие мероприятия обеспечивали эффективное

и защищенное функционирование ИС. Это может быть достигнуто лишь тогда, когда оценка полноты и качества исходных данных по аттестуемой СЗИ проводится не только на базе достаточно обоснованных показателей, критериев принятия решений и методик оценки, но и автоматизации процессов обследования, оценки и принятия решений путем создания соответствующего программного комплекса.

Создание программных комплексов – это научная, методологическая и практическая задача, требующая детального знания ИС, а также показателей и критериев принятия решений, возможных программных решений и т. д.

Целью настоящего исследования является описание подхода к созданию программного комплекса, автоматизирующего процесс экспертной оценки полноты и качества исходных данных по аттестуемой СЗИ ИС, в частности:

- 1) разработка методики экспертной оценки полноты и качества исходных данных по аттестуемой СЗИ ИС;
- 2) создание укрупненной блок-схемы алгоритма экспертной оценки полноты и качества исходных данных;
- 3) подготовка структуры программного комплекса, реализующего методику экспертной оценки полноты и качества исходных данных;
- 4) создание программного комплекса автоматизации процесса экспертной оценки полноты и качества исходных данных.

1. Методика оценки полноты и качества исходных данных

Методика экспертной оценки полноты и качества исходных данных по аттестуемой СЗИ ИС предназначена для экспертов специализированной организации, имеющей специальную лицензию на проведение работ по аттестации СЗИ ИС и содержит описание порядка проведения оценки по каждому из нижеприведенных показателей.

Показатель «полнота исходных данных» определяет соответствие совокупности исходных данных, представленных заявителем на аттестацию, по наименованию и назначению установленным требованиям.

Показатель «качество исходных данных» характеризуется двумя частными показателями: адекватностью отображения и достаточностью уровня детализации. Адекватность отображения определяет соответствие характеристик исходных данных по аттестуемой СЗИ ИС, представленных владельцем ИС, характеристикам, представленным в документации реальной СЗИ ИС. Достаточность уровня детализации исходных данных – это предоставление по каждому типу исходных данных информации, достаточно детальной для принятия решения по четкости, избыточности, непротиворечивости, однозначности описания, целевому назначению, реализуемости представленных положений.

Показатель «соответствие исходных данных требованиям нормативных правовых актов в области защиты информации, в том числе технических нормативных правовых актов» определяет соответствие исходных данных требованиям нормативных правовых актов (НПА) в области защиты информации, в том числе требованиям технических нормативных правовых актов (ТНПА), определенных в Национальной системе подтверждения соответствия в области технического нормирования и стандартизации.

1.1. Экспертная оценка полноты исходных данных

Для проведения экспертной оценки исходных данных по аттестуемой СЗИ ИС по показателю «полнота» на соответствие требованиям нормативных документов за основу взята вербально-числовая шкала Харрингтона, которая широко применяется для характеристики степени выраженности критериального свойства. Шкала имеет универсальный характер и может использоваться в соответствующих модификациях для оценки показателей качественного характера [4]. В рассматриваемом случае показатель «полнота» определяется лингвистической и интервальной оценками и характеризуется составом исходных данных и назначением каждого документа (табл. 1).

Таблица 1

Оценки исходных данных по показателю «полнота»

Лингвистическая оценка	Интервал оценок
Комплект исходных данных по составу и назначению соответствует установленным требованиям	0,8–1,0
В комплекте исходных данных допущены незначительные отклонения от установленных требований	0,63–0,79
По некоторым (не менее одного) исходным данным допущены отклонения от установленных требований	0,37–0,62
Комплект исходных данных по составу и назначению не соответствует установленным требованиям	0,01–0,36

Входные данные – совокупность исходных данных по аттестуемой СЗИ, представляемых заявителем в специализированную организацию.

Цель анализа – установить, соответствует ли состав представленных исходных данных требованиям приложения 2 Положения о порядке аттестации СЗИ ИС, предназначенных для обработки информации, распространение и (или) предоставление которой ограничено (приказ Оперативно-аналитического центра при Президенте Республики Беларусь, № 3 от 16.01.2015), а также отвечает ли каждый отдельный тип исходных данных по своему общему назначению требованиям нормативных документов Республики Беларусь.

Критерий принятия решений формулируется следующим образом: представленные исходные данные обладают свойством полноты в том случае, если они по составу соответствуют установленным требованиям, а по назначению – нормативным документам Республики Беларусь.

Количественная оценка исходных данных по показателю «полнота» определяется согласно типу совокупности количественных оценок, вычисляемых для каждого отдельного типа данных по формуле

$$c_i = \Theta_i \cdot g_i, \quad 1 \leq i \leq K,$$

$$\Theta_i = \begin{cases} 0, & \text{если заявитель не представил исходные данные под номером } i; \\ 1, & \text{если заявитель представил исходные данные под номером } i, \end{cases}$$

где K – количество единиц исходных данных, представляемых в соответствии с перечнем из приложения 2 Положения о порядке аттестации СЗИ ИС; $i, 1 \leq i \leq K$, – номер оцениваемой единицы исходных данных в этом перечне; g_i – количественная экспертная оценка соответствия единицы данных под номером i по своему общему назначению нормативным документам в области информационной безопасности.

Здесь и далее все экспертные оценки формируются в следующем порядке: сначала эксперт выставляет более грубую лингвистическую оценку в соответствии с лингвистической шкалой, указанной в левом столбце соответствующей шкалы (для оценок типа g_i – по показателю «полнота»), а затем уточняет ее количественно в соответствии с числовым интервалом, отвечающим выставленной лингвистической оценке.

Экспертное заключение по показателю «полнота» формируется на основе следующих правил:

1. Если для всех единиц исходных данных количественная оценка по показателю «полнота» находится в пределах 0,63–1,0, то по указанному показателю СЗИ ИС признается пригодной для проведения аттестации и эксперты приступают к оценке данных по другим показателям.

2. Если для некоторой единицы исходных данных количественная оценка по показателю «полнота» меньше либо равна 0,62, соответствующие данные возвращаются на доработку, а эксперты продолжают оценку данных, признанных пригодными по показателю «полнота», по другим показателям.

1.2. Экспертная оценка качества исходных данных

Задача экспертной оценки – провести анализ исходных данных по показателю «качество». Качество исходных данных определяется адекватностью отображения и достаточностью уровня детализации.

Цель анализа – установить, соответствует ли совокупность исходных данных требованиям нормативных документов к адекватности отображения и достаточности уровня детализации.

1.2.1. Оценка качества совокупности исходных данных по показателю «адекватность отображения»

Задача оценки: провести анализ исходных данных по показателю «адекватность отображения».

Входные данные: совокупность исходных данных, представляемых заявителем по аттестуемой СЗИ ИС.

Действия эксперта: эксперт последовательно анализирует каждый документ по показателю «адекватность отображения».

Цель: установить степень соответствия характеристик исходных данных по аттестуемой СЗИ ИС, представленных владельцем ИС, характеристикам, представленным в документации реальной ИС.

Для показателя «адекватность отображения» применяется следующий критерий принятия решений: представленные исходные данные обладают свойством надлежащей адекватности отображения в том случае, если все приведенные в них конкретные характеристики ИС в полной мере соответствуют характеристикам, представленным в документации реальной ИС.

Степень соответствия по показателю «адекватность отображения» определяется лингвистической и интервальной оценками (табл. 2).

Таблица 2

Оценки по показателю «адекватность отображения»

Лингвистическая оценка	Интервал оценок
Строгое соответствие	0,8–1,0
Высокая степень соответствия	0,63–0,79
Средняя степень соответствия	0,37–0,62
Несоответствие	0,01–0,36

Количественная оценка исходных данных по показателю «адекватность отображения» определяется по совокупности количественных оценок, которые вычисляются для каждого отдельного типа данных, содержащих конкретные характеристики информационной системы, согласно формуле

$$A_i = \frac{\sum_{t=1}^{m(i)} p_{it} \cdot A_{it}}{\sum_{t=1}^{m(i)} p_{it}}, \quad 1 \leq i \leq K,$$

где $m(i)$ – количество оцениваемых характеристик для типа исходных данных под номером i в перечне из приложения 2 Положения о порядке аттестации СЗИ ИС; p_{it} – вес характеристики t , отражающий ее важность с точки зрения оценки СЗИ ИС; A_{it} – количественная экспертная оценка степени соответствия характеристики t i -го типа данных аналогичной характеристике, представленной в документации реальной ИС.

Для вычисления весов p_{it} может использоваться метод парных сравнений [5], основанный на заполнении экспертом таблиц парного сравнения важности всех характеристик ИС между собой (рис. 1).

Номера оцениваемых характеристик	1	2	...	$m(i)-1$	$m(i)$
1	×				
2		×			
...			...		
$m(i)-1$				×	
$m(i)$					×

Рис. 1. Пример таблицы парного сравнения

В каждой ячейке верхней части данной таблицы, находящейся на пересечении строки t , $1 \leq t \leq m(i)-1$, и столбца j , $2 \leq j \leq m(i)$, эксперт указывает номер t или j более важной, по его мнению, характеристики.

Вес p_{it} характеристики t определяется по формуле

$$P_{it} = \frac{M_{it}}{C},$$

где M_{it} – общее число указаний номера t в ячейках таблицы (предпочтений экспертом характеристики t в ходе проведения парных сравнений); C – количество произведенных экспертом сравнений:

$$C = \frac{m(i)(m(i)-1)}{2}.$$

Экспертное заключение по показателю «адекватность отображения» формируется на основе следующих правил:

1. Если для типов исходных данных, содержащих конкретные характеристики ИС, количественная оценка по показателю «адекватность отображения» находится в пределах 0,63–1,0, то по указанному показателю СЗИ ИС признается пригодной для проведения аттестации.

2. Если для некоторого типа исходных данных, содержащих конкретные характеристики ИС, количественная оценка показателя «адекватность отображения» находится в пределах 0,01–0,62, то по указанному показателю данный тип исходных данных признается несоответствующим требованиям соответствия. Анализируемый документ возвращается на доработку.

1.2.2. Оценка качества совокупности исходных данных по показателю «достаточность уровня детализации»

Задача оценки: провести анализ исходных данных по аттестуемой СЗИ ИС согласно показателю «достаточность уровня детализации».

Входные данные: совокупность исходных данных, представляемых заявителем по аттестуемой СЗИ ИС.

Действия эксперта: эксперт последовательно анализирует каждый документ по показателю «достаточность уровня детализации».

Цель: установить достаточность детализации представленной информации для принятия решения по четкости и однозначности описания, целевому назначению, реализуемости представленных положений.

Для показателя «достаточность уровня детализации» применяется следующий критерий принятия решений. Уровень детализации информации, представленной в элементе исходных данных или документе, является достаточным, если представленная информация:

- содержит минимальный, но достаточный набор показателей для принятия экспертом правильного решения относительно целевого назначения и реализуемости представленных положений;

- определяет все основные концептуальные положения, на базе которых сформулированы основные положения о функциональных возможностях аттестуемой СЗИ ИС;

- содержит обоснование принятых основных положений или ссылку на ранее реализованное положение, программу или устройство;

- является избыточной, непротиворечивой и с этой точки зрения не может привести к ошибкам в принятии решений.

Достаточность уровня детализации отдельных элементов исходных данных определяется лингвистической и интервальной оценками (табл. 3).

Таблица 3

Оценки по показателю «достаточность уровня детализации»

Лингвистическая оценка	Интервал оценок
Положительная оценка по всем составляющим показателя	0,8–1,0
Допущены незначительные отклонения по некоторым составляющим показателя	0,63–0,79
Допущены отклонения хотя бы по одной составляющей показателя	0,37–0,62
Допущены значительные отклонения по нескольким составляющим показателя	0,01–0,36

Количественная оценка исходных данных по показателю «достаточность уровня детализации» определяется по совокупности количественных оценок, вычисляемых для каждого отдельного типа данных, согласно формуле

$$D_i = \frac{\sum_{t=1}^{h(i)} p_{it} \cdot D_{it}}{\sum_{t=1}^{h(i)} p_{it}}, \quad 1 \leq i \leq K,$$

где $h(i)$ – количество характеристик, которые, по мнению эксперта, являются информативными для принятия решения относительно качества, целевого назначения и реализуемости представленных положений для исходных данных под номером i в перечне из приложения 2 Положения о порядке аттестации СЗИ ИС; p_{it} – вес характеристик t , отражающий их важность с точки зрения оценки СЗИ ИС; D_{it} – количественная экспертная оценка достаточности, неизбыточности и непротиворечивости описания характеристик t для данных типа i .

Экспертное заключение по показателю «достаточность уровня детализации» формируется на основе следующих правил:

1. Если для всех типов исходных данных количественная оценка по показателю «достаточность уровня детализации» находится в пределах 0,63–1,0, то по указанному показателю СЗИ ИС признается пригодной для проведения аттестации.

2. Если для некоторого типа исходных данных количественная оценка по показателю «достаточность уровня детализации» находится в пределах 0,01–0,62, то документ (документы) возвращаются на доработку.

Обобщенное экспертное заключение по качеству исходных данных формируется на основе следующего правила: исходные данные по аттестуемой СЗИ ИС признаются пригодными для проведения аттестации по показателю «качество» в том и только в том случае, если для всех типов исходных данных количественные оценки по показателям «адекватность отображения» и «достаточность уровня детализации» находятся в пределах 0,63–1,0.

1.3. Методика оценки соответствия исходных данных требованиям нормативных правовых актов в области защиты информации, в том числе технических нормативных правовых актов

С учетом имеющейся нормативно-правовой базы Республики Беларусь в области защиты информации и типа конкретной СЗИ ИС оценивается соответствие исходных данных и документов требованиям НПА в области защиты информации, в том числе требованиям ТНПА, определенных в Национальной системе подтверждения соответствия в области технического нормирования и стандартизации.

Под оценкой соответствия документа понимается деятельность эксперта по определению соответствия назначения и структуры разработанного документа требованиям НПА и ТНПА в области защиты информации.

Степень соответствия характеризуется такими же лингвистической и интервальной оценками, как и показатель «адекватность отображения» (см. разд. 1.2.1).

Количественная оценка исходных данных по показателю «соответствие исходных данных требованиям ТНПА в области защиты информации» определяется по совокупности количественных оценок, вычисляемых для каждого отдельного типа данных, согласно формуле

$$E_i = \frac{\sum_{l=1}^{T(i)} E_{il}}{T(i)}, \quad 1 \leq i \leq K,$$

где E_i – оценка соответствия исходных данных типа i требованиям ТНПА; $T(i)$ – количество ТНПА, требованиям которых должны удовлетворять исходные данные типа i ; E_{il} – количественная экспертная оценка степени соответствия исходных данных типа i требованиям ТНПА l .

Экспертное заключение о соответствии пакета исходных данных и документов требованиям НПА и ТНПА в области защиты информации формируется на основе следующих правил:

1. Если для всех типов исходных данных количественная оценка по показателю «соответствие исходных данных требованиям ТНПА в области защиты информации» находится в пределах 0,63–1,0, то по указанному показателю СЗИ ИС признается пригодной для проведения аттестации.

2. Если для некоторого типа исходных данных количественная оценка по показателю «соответствие исходных данных требованиям ТНПА в области защиты информации» находится в пределах 0,01–0,62, то документ (документы) возвращается на доработку.

Обобщенное экспертное заключение по всей совокупности исходных данных формируется следующим образом:

1. Если для всей совокупности показателей все количественные оценки находятся в пределах 0,63–1,0, то СЗИ ИС признается пригодной для проведения аттестации.

2. Если хотя бы по одному типу исходных данных для любого показателя количественная оценка находится в пределах 0,01–0,62, а их доработка не привела к соответствию установленным требованиям, то принимается решение об отказе в аттестации СЗИ ИС.

2. Укрупненная блок-схема алгоритма экспертной оценки полноты и качества исходных данных

На основе принятых показателей и правил принятия решений создана укрупненная блок-схема алгоритма экспертной оценки полноты и качества исходных данных по аттестуемой СЗИ ИС (рис. 2).

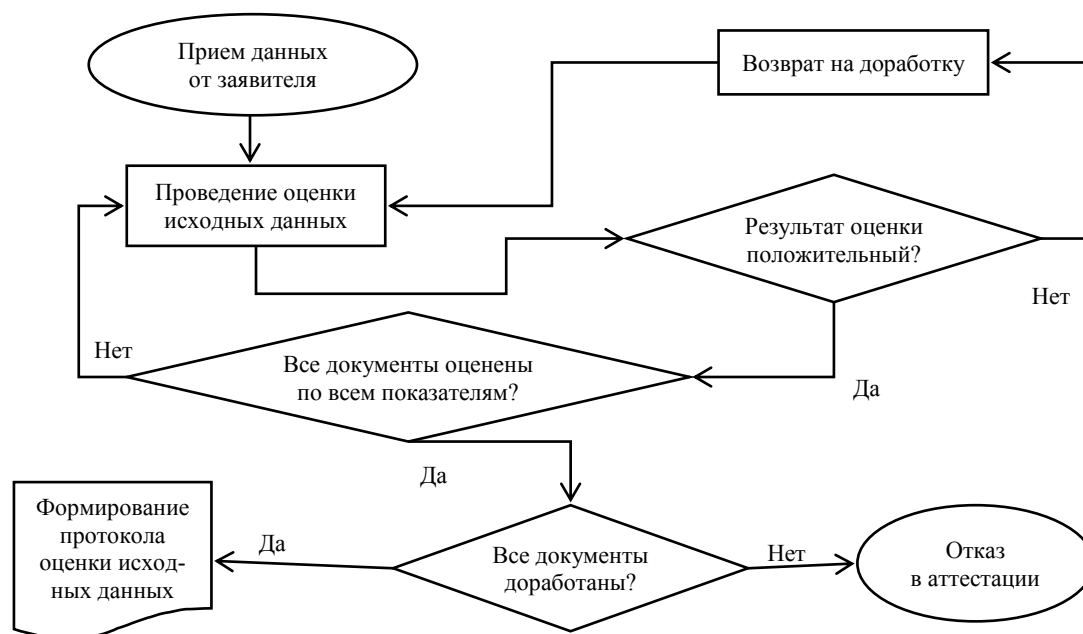


Рис. 2. Укрупненная блок-схема алгоритма экспертной оценки исходных данных

3. Структура программного комплекса, реализующего методику экспертной оценки полноты и качества исходных данных

Рассмотрим структурную схему программного комплекса оценки исходных данных по аттестуемой СЗИ ИС (рис. 3) и подсистему их оценки (рис. 4).

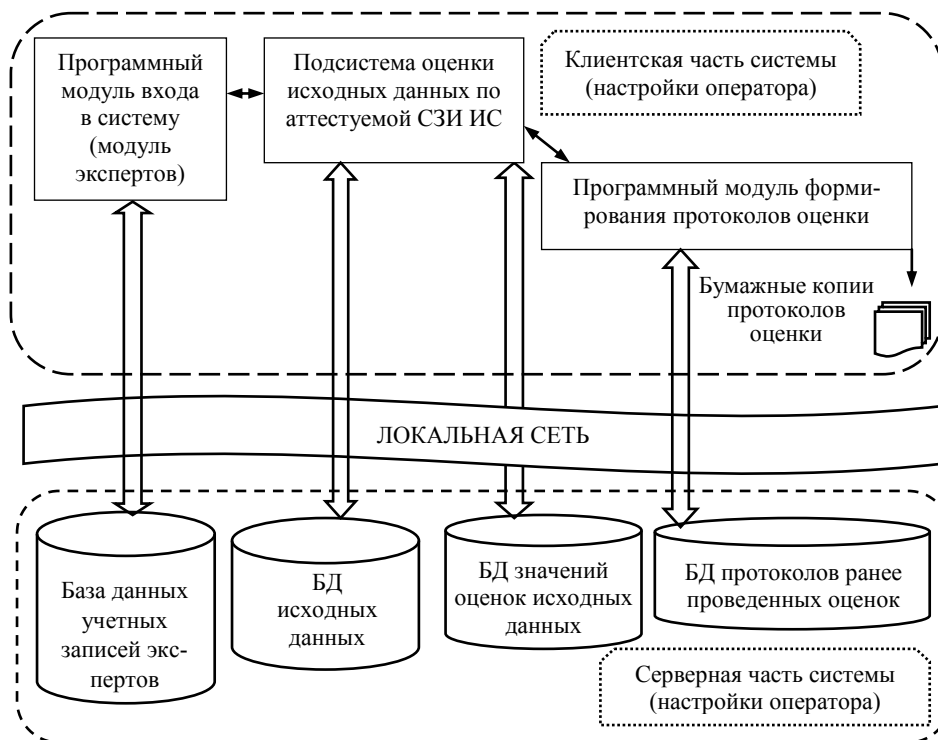


Рис. 3. Структурная схема программного комплекса оценки исходных данных по аттестации СЗИ ИС

Состав программного комплекса оценки и назначение его основных структурных компонентов представлены ниже.

Программный модуль входа в систему. С помощью модуля обеспечиваются регистрация экспертов с определенной степенью доступа, их идентификация и аутентификация, редактирование учетных записей, регистрация комплекта исходных данных для аттестуемой СЗИ ИС.

Программный модуль оценки исходных данных. С помощью модуля формируются результаты оценки каждого типа исходных данных по показателям «полнота», «адекватность отображения», «достаточность уровня детализации» и «соответствие исходных данных требованиям нормативных правовых актов и технических нормативных правовых актов в области защиты информации», а также экспертное заключение по каждому показателю и по их совокупности.

Программный модуль формирования протоколов оценок. С помощью модуля осуществляются корректировка, просмотр и сохранение результатов оценки исходных данных в соответствии с решениями экспертов, формирование протоколов оценки в формате «doc» и формирование бумажной копии документа.

База данных учетных записей экспертов. В базе хранятся данные о правах доступа экспертов, их пароли и другие реквизиты, предназначенные для аутентификации и входа экспертов в систему.

База данных результатов оценки исходных данных. В базе хранятся оценки, выставленные экспертами в ходе анализа исходных данных, а также сделанные при этом экспертные замечания. Сведения из базы данных используются для формирования отчетных документов.

База данных протоколов оценок исходных данных по аттестованным СЗИ ИС. В базе аккумулируется накопленный опыт оценки исходных данных – хранятся электронные версии протоколов оценки исходных данных, сформированные в ходе проведенных ранее аттестаций СЗИ ИС.

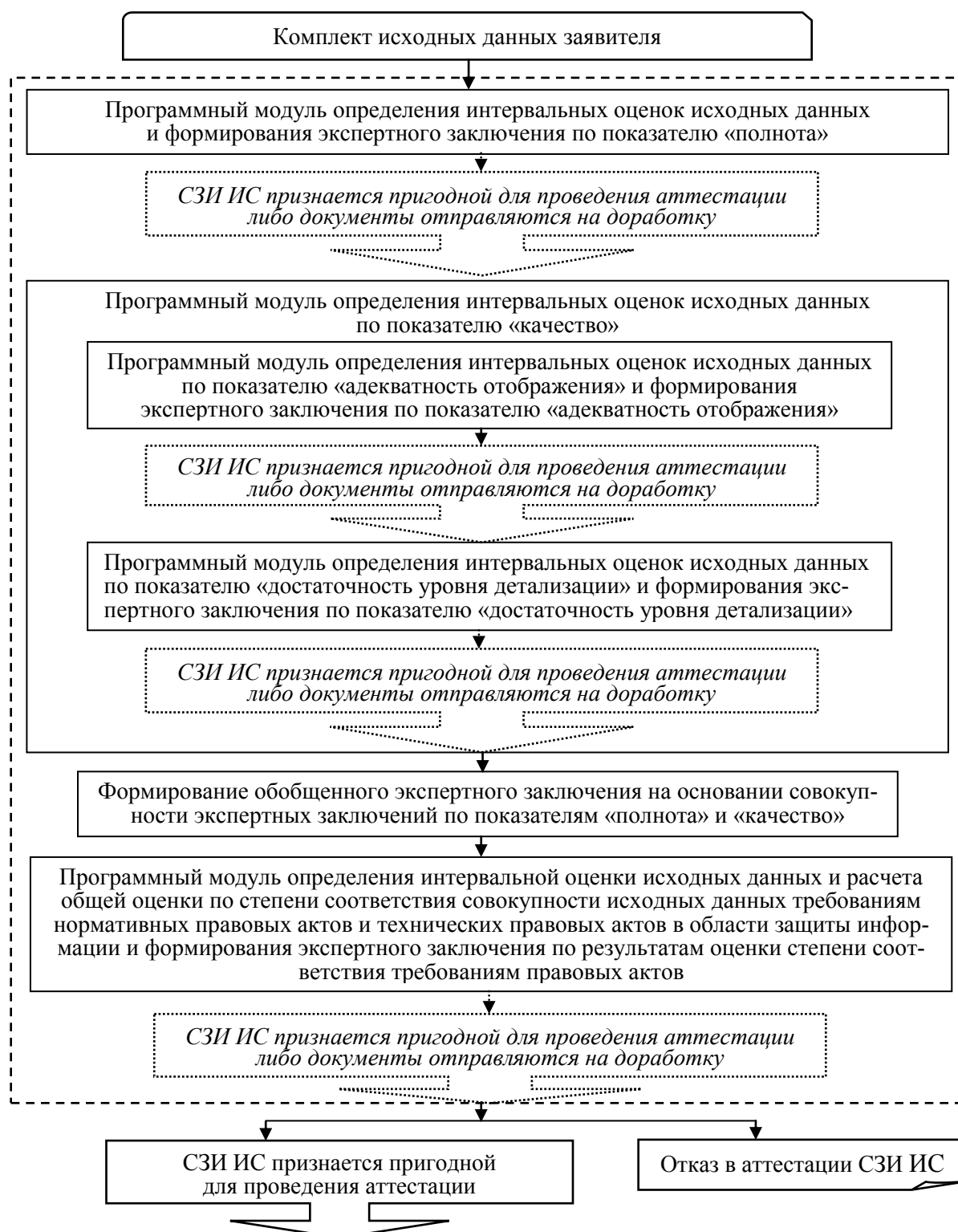


Рис. 4. Структурная схема подсистемы оценки исходных данных по аттестуемой СЗИ ИС

4. Программная реализация методики оценки исходных данных по аттестации СЗИ ИС

На основе методик оценки полноты и качества исходных данных создана их программная реализация, позволяющая значительно сократить время оценки исходных данных, представленных заявителем, а следовательно, и сроки подготовки к проведению аттестации.

Программная реализация представляет собой программу, использующую библиотеки Framework Qt. Программа предназначена для платформы Windows. Минимальные требуемые характеристики: Windows XP, 256 МБ RAM, 50 МБ ROM.

Программа включает в себя три основных этапа работы с исходными данными заявителя.

На первом этапе эксперт проводит подготовительные работы для объекта оценки, а именно заполняет информацию о заказчике, наименовании объекта оценки. Во время выполнения программы существует возможность изменения интервальных и лингвистических шкал оценок при изменении критериев оценки.

В основной части программы проводится оценка представленных заявителем исходных данных по четырем критериям: «полнота исходных данных», «адекватность отображения», «достаточность уровня детализации», «соответствие нормативным документам». На рис. 5 в верхней части представлена вкладка оценки показателей, которая визуально разбита на четыре основных блока. Автоматизация процессов выставления оценок и написания замечаний позволяет значительно ускорить проведение оценки.

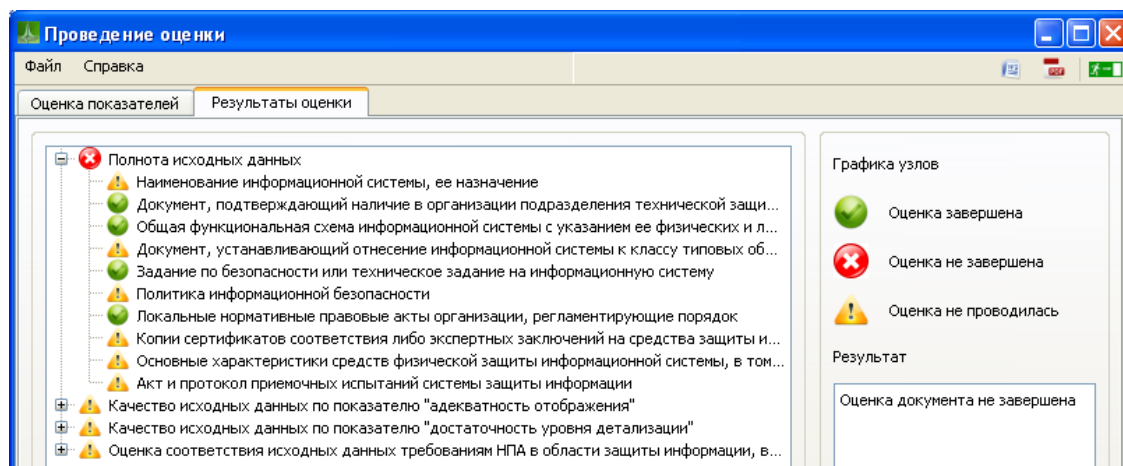
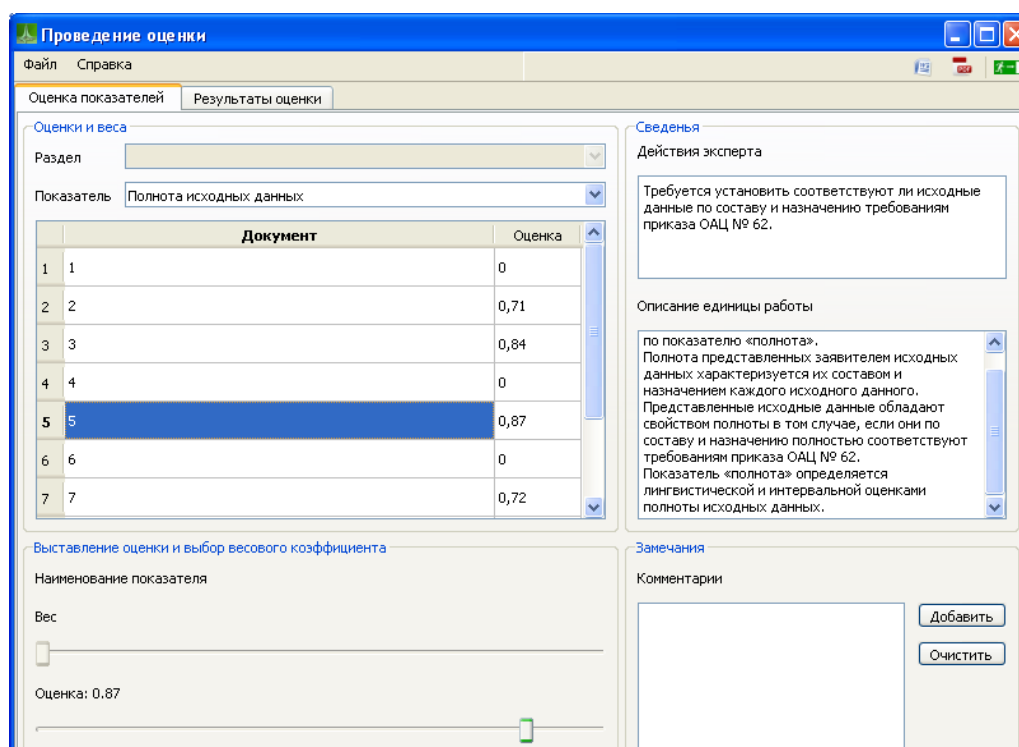


Рис. 5. Этап выполнения оценки

На завершающем этапе после проведения оценки эксперту не требуется время для написания отчета о проделанной работе, так как данный отчет формируется автоматически на основании данных и оценок, выставленных экспертом. Кроме того, во второй вкладке (см. рис. 5) можно просмотреть ход проведения оценки по типу графа с иконками в его узлах, которые наглядно показывают статус документа по просматриваемому показателю. Также в текстовом поле Результат выводится результат проведения оценки.

Практическая значимость программной реализации методик определяется следующими фактами:

– за счет автоматизации процессов аттестации СЗИ значительно (в два-три раза) сокращается время аттестации;

– в процессе аттестации выявляются уязвимые места СЗИ ИС, что позволяет владельцам систем своевременно принимать меры по их устранению и тем самым способствует выполнению таких свойств защиты, как конфиденциальность, целостность и доступность.

Структура программы состоит из следующих окон: Вход в систему, Настройка пользователя/Изменение эксперта, Документы оценки, Объект оценки, Шкалы оценок, Заказчики, Добавление/Изменение заказчика, Проведение оценки. Их логическая связка объединяет программу в единое целое: из окна Вход в систему можно перейти к добавлению либо настройке пользователя и к окну Документы оценки, из окна Документы оценки – к добавлению либо редактированию документа, к окну Заказчики (из которого можно перейти к добавлению либо изменению заказчика), а также непосредственно к окну оценки Проведение оценки.

Входными данными в программе являются оцениваемые экспертом документы, а также наименование заказчика, его реквизиты, наименование объекта оценки, шкала лингвистических и интервальных оценок. В ходе проведения оценки эксперт с помощью программы формирует оценки документов по показателям.

Выходными данными являются результаты оценки исходных данных, которые представляются в виде документа «Протокол оценки полноты и качества исходных данных по аттестуемой системе защиты информации».

Заключение

В статье изложена методика и общий алгоритм реализации процесса последовательного обследования исходных данных по аттестации СЗИ ИС с позиции информационной безопасности по показателям полноты, адекватности отображения, достаточности уровня детализации, а также по степени соответствия исходных данных требованиям НПА в области защиты информации, в том числе ТНПА.

На основе разработанных методик обследования исходных данных по аттестации создана их программная реализация, позволяющая за счет автоматизации сократить время обследования и тем самым значительно уменьшить сроки подготовки к проведению аттестации. Благодаря этому отпадает необходимость написания протокола обследования, так как данный документ формируется автоматически на основании данных и оценок, выставленных экспертами.

Список литературы

1. Jet Info. Аттестация автоматизированных систем. – М. : Джет Инфо Паблшер. – 2000. – № 11(90). – 20 с.
2. Закон Республики Беларусь от 10 ноября 2008 г. № 455-3 «Об информации, информатизации и защите информации» // Национальный правовой интернет-портал Республики Беларусь [Электронный ресурс]. – 2008. – Режим доступа : <http://www.pravo.by/main.aspx?guid=3871&p0=h10800455&p2={NRP}>. – Дата доступа : 23.06.2015.
3. Приказ Оперативно-аналитического центра при Президенте Республики Беларусь от 16 января 2015 г. № 3 «О внесении дополнений и изменений в приказ Оперативно-аналитического центра при Президенте Республики Беларусь от 30 августа 2013 г. № 62 [Электронный ресурс]. – 2015. – Режим доступа : <http://oac.gov.by/law/49.html>. – Дата доступа : 03.07.2015.

4. Кириллова, О.В. Сбор и агрегирование экспертных оценок для ранжирования научных журналов / О.В. Кириллова, О.В. Федоренко // Тез. докл. 12-й Междунар. конф. LIVCOM–2008. – М. : ГПНТБ России, 2008. – С. 32–35.

5. Саати, Т. Принятие решений. Метод анализа иерархий / Т. Саати. – М. : Радио и связь, 1993. – 278 с.

Поступила 04.05.2015

*Объединенный институт проблем
информатики НАН Беларуси,
Минск, Сурганова, 6
e-mail: fisenko@newman.bas-net.by*

**V.K. Fisenko, E.P. Maksimovich, A.B. Stepanyan,
V.A. Dmitriev, D.S. Kim, T.S. Martinovich**

**METHODOLOGICAL BACKGROUND OF EXPERT ESTIMATION
OF INITIAL DATA COMPLETENESS AND QUALITY ACCORDING
TO THE CERTIFIED INFORMATION SECURITY SYSTEM**

Problem of information security systems certification is analyzed and the tasks of initial data analysis are carried out. The objectives, indices and decision making criteria, as well as the challenges to be addressed are formulated. It is shown that, in order to improve quality, reduce time and cost of preparation for certification, it is reasonable to use software system for automatization of the process of initial data analysis, presented by the owner of the information system.

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. Статья должна иметь индекс УДК (универсальная десятичная классификация).

3. 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.

4. 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.

5. 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).

6. 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.

7. 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>.

8. 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.

9. Статьи, направляемые на доработку, должны быть возвращены в исправленном виде с ответами на все вопросы.

10. Редакция журнала предоставляет возможность первоочередного опубликования статей, представленных лицами, которые осуществляют послевузовское обучение (аспирантура, докторантура, соискательство) в год завершения обучения.

11. Авторы несут ответственность за направление в редакцию статей, уже опубликованных ранее, или статей, принятых к публикации другими изданиями.

12. 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