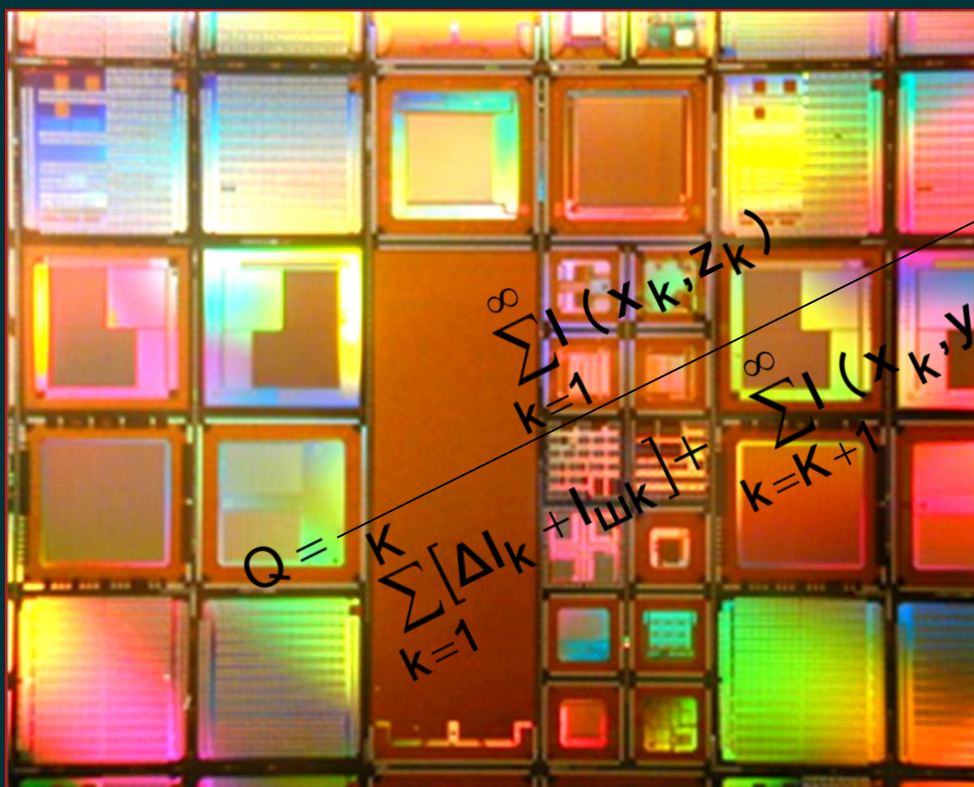


А. К. Цыцулин Д. Ю. Адамов
А. А. Манцветов И. А. Зубакин

ТВЕРДОТЕЛЬНЫЕ ТЕЛЕКАМЕРЫ: НАКОПЛЕНИЕ КАЧЕСТВА ИНФОРМАЦИИ



Оптические измерительные инструменты не являются просто пассивными «приёмниками», которые только регистрируют попадающие на них лучи; они активно участвуют в процессе измерения, оказывая причинное влияние на результат.

Макс Планк



Министерство образования Российской Федерации
Санкт-Петербургский государственный
электротехнический университет «ЛЭТИ»
им. В. И. Ульянова (Ленина)

ТВЕРДОТЕЛЬНЫЕ ТЕЛЕКАМЕРЫ: НАКОПЛЕНИЕ КАЧЕСТВА ИНФОРМАЦИИ

Санкт-Петербург
Издательство СПбГЭТУ «ЛЭТИ»
2014

УДК 621.397.001

ББК 3 945я7

Т 26

Цыцулин А. К., Адамов Д. Ю., Манцветов А. А., Зубакин И. А.

**Т 26 Твердотельные телекамеры: накопление качества информации. –
СПб., Изд-во СПбГЭТУ «ЛЭТИ», 2014. – 234 с.**

ISBN 5-7629-0496-2

Анализируется второй этап твёрдотельной революции в телевидении, определяющийся использованием КМОП-технологии (первый этап – приборы с зарядовой связью), новые результаты по теории информации и связанные с ними новые методы кодирования изображений непосредственно в телекамерах.

На основе принципа доминантной информации, согласно которому информация определяется триединством цели, качества и количества, показано, что качество переданной информации связано с информационным риском, состоящим из потери полезной информации и шумовой информации. Показана связь принципа достаточной слитности изображений с максимумом качества видеоинформации.

С новых позиций рассмотрен принцип накопления в матричных фотоприёмниках. Обоснованы различные методы реализации принципа накопления качества видеоинформации в зависимости от априорной информации и технологических возможностей реализации матриц. Предложены новые методы считывания и новые структуры фотоприёмных матриц, реализующих максимум качества информации в задачах адаптации к динамическим характеристикам сюжета, обнаружения–оценивания, стабилизации изображения, достижения контрастной чувствительности, телевизионно-лазерной локации.

УДК 621.397.001

ББК 3 945 я7

Рецензенты:

академик НАН РК, доктор технических наук, профессор В. Н. Амербаев,
главный научный сотрудник ИППМ РАН,

доктор технических наук, профессор В. В. Александров,
зав. лабораторией СПИИ РАН

ЗАО «Электроника и видеосистемы»

кандидат технических наук А. Н. Куликов, начальник лаборатории

ISBN 5-7629-0496-2

© СПбГЭТУ «ЛЭТИ», 2014

© Цыцулин А. К., Адамов Д. Ю., Манцветов А. А., Зубакин И. А., 2014

ПРЕДИСЛОВИЕ

Материал, изложенный в книге, является развитием идей твёрдотельного телевидения, отражённых в книгах «Твёрдотельное телевидение (телевизионные системы с переменными параметрами на ПЗС и микропроцессорах)» и «Твёрдотельная революция в телевидении (телевизионные системы на основе приборов с зарядовой связью, систем на кристалле и видеосистем на кристалле)», отражает результаты новых исследований авторов в этой области и включает лишь минимум опубликованного в этих книгах, необходимый для связности изложения. Как и последняя из этих книг, предлагаемая вниманию читателей книга посвящена второму этапу твёрдотельной революции в научно-прикладном телевидении. Первый этап характеризовался внедрением рожденных в начале 1970-х гг. достижений твёрдотельной электроники – приборов с зарядовой связью и микропроцессоров. Новый этап твёрдотельной революции в телевидении обусловлен не столько совершенствованием этих приборов, сколько рождением и внедрением новых приборов микроэлектронной техники – «видеосистем на кристалле», которые благодаря единой КМОП-технологии позволили объединить на одном кристалле фотоприёмник, устройства развёртки, оцифровки и цифровой обработки видеосигналов. Современные фотоприёмные КМОП-матрицы уже существенно опередили приборы с зарядовой связью по распространённости – их применяется уже более чем в 10 раз больше, и приближаются к своим идеологическим предшественникам по качеству видеоинформации.

Книга относится к конвергентным технологиям, она отражает сближение (конвергенцию) теории информации и микроэлектроники, как и другие конвергентные технологии, характеризуется междисциплинарностью, изменением парадигмы развития от анализа к синтезу и переходом к наноразмерам. Важной отличительной чертой книги, её целью является попытка построения наброска теории научно-прикладного телевидения на основе нескольких дисциплин – не только физических, но и математических. Авторы взяли за очень трудную задачу, которую ещё решать и решать – они попытались «впрячь в свою телегу» «коня техники» и «трепетную лань философии». Эта цель книги соответствует весьма актуальной проблеме современной прикладной науки – интеграции различных областей знания: не случайно великий математик и физик Анри Пуанкаре

утверждал, что истинная, единственная цель науки – раскрытие не механизма, а единства и что прогресс осуществляется именно благодаря неожиданным сближениям между различными частями науки. Широта охвата материала, включающего теорию информации, системотехнику и технологию проектирования современных телекамер и их сердца – матричных фотоприёмников, – позволила авторам рассматривать телекамеру как устройство кодирования видеоинформации, а не просто как «датчик», что усилило связку теории информации с кодированием непосредственно в телекамере для повышения качества информации. Важно, что авторы, специалисты по видеоинформатике, поставили вопрос о качестве изображений более фундаментально, чем это обычно было принято в телевизионной технике. Это позволило авторам дать новый взгляд на традиционно используемый в телевидении принцип накопления, обычно трактуемый как накопление заряда, и предметно вести речь о накоплении информации.

С точки зрения «социального заказа» книгу можно охарактеризовать как вклад в обоснование путей совершенствования прикладных видеоинформационных систем, путей их оптимизации при проектировании и адаптации в процессе эксплуатации. Технологию твёрдотельного телевидения авторы справедливо трактуют двояко: новая технология создания фотоприёмников дополняется технологией создания систем на них – системой автоматизированного проектирования (САПР). Книга не просто показывает, что отечественная видеоинформационная отрасль характеризуется определённым интеллектуальным заделом при необеспеченности отечественной микроэлектронной базой, но намечает некоторые важные задачи отечественной микроэлектронной науки и промышленности, решение которых позволит сократить наше технологическое отставание от других государств.

Можно надеяться, что знакомство читателей, специалистов трёх уровней – создателей сложных комплексов, разработчиков входящих в них прикладных телекамер и разработчиков микроэлектронных компонентов для них, с достижениями и возможностями нового поколения твёрдотельного телевидения окажет стимулирующее влияние на появление новых радиоэлектронных систем и концепций отечественной прикладной науки.

*Директор института проблем проектирования в микроэлектронике РАН,
академик РАН*

А. Л. Стемковский

ОТ АВТОРОВ

Данное издание лежит в русле книг «Твёрдотельное телевидение» [1], «Видеоинформатика» [2] и «Твёрдотельная революция в телевидении» [3] и отражает изменения в построении телекамер, осознанные авторами за время после издания последней. Как отмечалось в [3], главным содержанием нового этапа твёрдотельной революции стал переход к КМОП-технологии, приведшей к появлению систем и видеосистем на кристалле.

В нашей книге главными героями являются не сами системы, телекамеры и фотоприёмники, а их модели. При этом мы разделяем давно вызревшее мнение о моделировании, сформулированное С. Лемом [4]: *«Моделирование – это подражание Природе, учитывающее немногие её свойства. Почему только немногие? Из-за нашего неумения? Нет. Прежде всего, потому, что мы должны защититься от избытка информации... Сейчас мы уже имеем ситуацию информационного потопа»*. С этим потопом можно бороться только выделяя доминанту из доступной информации, чего возможно достичь двумя способами: *накоплением доминантной информации* для борьбы с шумовой информацией и возможным подавлением (вычитанием, прореживанием) фоновой информации. Поэтому другими главными героями книги являются качество информации и его накопление, к которым мы подходим с позиций теории связи. При этом мы рассматривали телекамеру не столько как источник информации (формирующий «данные» для последующей компьютерной обработки), сколько как кодер-накопитель, ответственный за важнейший этап – рождение видеоинформации. При этом теоретической опорой при описании *сигналов* авторы приняли вероятностную теорию информации, полагая, что алгоритмическая теория информации применима к описанию синтезируемых *устройств* – фотоприёмных матриц и телекамер – и *технологий*. Большинство своих теоретических результатов, полученных в первой части книги, авторы пытались обосновать аналитически и использовали во второй части. В первую очередь найденные оптимальные фильтры рассматривались как эталоны, к которым могут стремиться разработчики фотоприёмников и телекамер.

В книге рассматриваются проблемы, относящиеся к теории проектирования прикладных телевизионных систем. Почему мы говорим о *проблемах*, а не более скромно – о задачах? Потому, что задача – это потребность, и в хозяйстве под проблемой понимают крупную задачу (продовольственная проблема), но *в науке проблема – это конфликт идей* [5], и такие

конфликты рассматриваются как в книге в целом, так и в *каждой главе*. Эти проанализированные конфликты – тот самый «диссонанс, неудовлетворённость – подталкивающий (хотя и не единственный) импульс к творчеству, изобретению» [6]. Для книги в целом первой конфликтующей идеей является утверждение, что телевидение – это не наука, а услуга населению: не только руководители Всесоюзного НИИ телевидения считали его чистой техникой, но и в Академии наук относились к телевидению как к бытовой технике, даже были сравнения телевизора с пылесосом. Конкурирующая идея: телевидение – не только самая информационно-ёмкая отрасль связи, но и основа видеоинформатики, которая в силу квантовой структуры света впервые остро поставила вопрос о качестве информации, чем обозначила потенциальную связь теории информации и теории познания.

К изложенной в книге концепции можно относиться по-разному, и авторы полностью разделяют подход, сформулированный Д. Дойчем [5], от имени Сократа написавшим: *«По нашим обычаям мы выслушиваем любого, кто предлагает честную критику и хочет убедить нас добровольно изменить своё мнение, ведь мы хотим делать то, что правильно... путём поиска мы выучили простые вещи и пришли к согласию в них. А в будущем теми же средствами, а именно отказываясь огораживать любые свои идеи от критики, мы, возможно, узнаем и более сложное»*. Этот подход особенно важен потому, что теоретические результаты, изложенные в книге, получены, конечно, в рамках возможностей авторов – инженеров, а не математиков по образованию, и изложенные идеи ещё ждут подключения математиков к их разработке.

Мы выражаем признательность ветерану НИИ телевидения д-ру техн. наук Л. И. Хромову, чьи идеи развивает эта книга, благодарим за поддержку ген. директора ОАО «НИИТ» д-ра техн. наук А. А. Умбиталиева и зав. каф. телевидения и видеотехники д-ра техн. наук Н. В. Лысенко; благодарим за важные рекомендации рецензентов акад. В. Н. Амербаева, д-ра техн. наук В. В. Александрова и канд. техн. наук А. Н. Куликова; ознакомившихся с рукописью и давших полезные критические замечания д-ра техн. наук Ю. Ф. Адамова, д-ра техн. наук Ю. С. Сагдуллаева, д-ра техн. наук Ю. С. Шинакова и канд. техн. наук В. А. Зимогляда, а также активно взаимодействовавших с нами сотрудников СПбГЭТУ «ЛЭТИ», ОАО «НИИТ» и ООО «Юник Ай Сиз» и всех коллег, чьи работы влились в русло второго, КМОП-го этапа твёрдотельной революции в телевидении.

ВВЕДЕНИЕ

Главная цель всякой науки состоит в слиянии всех возросших в ней теорий в одну-единственную, в которой все научные проблемы занимали бы определённое место и получали бы однозначное решение. Поэтому приходится признать, что наука стоит тем ближе к своей цели, чем сильнее уменьшается число заключающихся в ней теорий.

Макс Планк

В.1. Теория информации – основа теории телевидения

Каждая отрасль техники, и телевидение в том числе, претендует на свою самостоятельность, на то, чтобы иметь свою теорию. Однако всякой прикладной науке необходимо определиться, к какому научному направлению она относится. Известно, что между разными научными направлениями существуют разрывы, и современное состояние науки характеризуется глубинным разрывом между физикой и психологией. Наиболее вероятным мостом между ними может стать теория информации: не случайно Фехнера, автора первого информационного закона о логарифмической связи стимула и реакции организма, известного как закон Вебера–Фехнера, называют «отцом экспериментальной психологии» [6].

Информация столь же фундаментальное понятие для науки, как время и пространство. Ещё в античные времена была осознана идея алгоритма получения информации на основе постановки вопросов, допускающих ответы в форме «да» или «нет» (легенда приписывает эту идею Бар-Кохбе [7]). С. Лем полвека назад, осмысляя и популяризируя идеи информатики, писал, что *«всё сущее есть информация»* и что *«информация существует только тогда, когда имеется её адресат»* [4]. Через четверть века в след ему уже физики, начиная с Дж. Уилера [8], осознали, что информация является фундаментальной концепцией физики. Согласно его доктрине *«It from bit»* («всё из бита») все физические сущности являются информационно-теоретическими в своей основе. Мы познаём мир через посредство информации, и вместо старого принципа отражения «материя есть объективная реальность, данная нам в ощущениях» можно утверждать: **материя дана нам как информация**. Вместе с тем полной ясности в вопросе о том,

что такое информация, нет. *«Вдохновлённая Уилером большая группа учёных – специалистов по информатике, астрономов, математиков, биологов и физиков ... – начала исследовать связи между теорией информации и физикой»*, однако *«Уилер признал, что эти идеи были ещё сырыми и не готовыми для строгого тестирования»*, и хотя теория информации в результате влияния Дж. Уилера *«наслаждается в физике периодом возрождения, ей ещё требуется сделать вклад в эту науку каким-нибудь конкретным способом»* [9]. Наибольшую активность в этом направлении проявляют энтузиасты создания квантового компьютера, которые в рамках науки о вычислениях, идущей от работ Дж. фон Неймана, развивают научное направление *квантовой теории информации* [10]. Оно относится к тематике *computer science*, синтезирует понятия квантовой механики и классической теории информации и исходит из концепции наличия информации в системе даже в отсутствие наблюдателя. Мы исходим из иной концепции, введённой А. Н. Колмогоровым, согласно которой центральным понятием теории информации (теория связи и теория информации понимаются как синонимы) является взаимная информация. Некоторые из таких конкретных способов будут рассмотрены в первой части книги.

К сожалению, термин «информация» возник не в теоретической физике, а в технике связи. В 1928 г. связист Р. Хартли написал статью под названием «Передача информации» [11], где он определил связь именно как передачу информации. Он догадывался, что информация позволит дать нужную, более правильную оценку пропускной способности канала связи, понятие о которой уже существовало. Факт введения в научный обиход понятия информации в теории связи подчёркивает, что информатика – это единство компьютеров и связи. Между ними нет разрыва, и Н. Винер писал: *«Исходя из моих общих идей, я рассматривал автоматические вычислительные машины также как одну из форм систем связи»*. Это, а также само название книги Н. Винера [12] означает, что теория связи – часть информатики и кибернетики.

Информационная сущность закона Вебера–Фехнера означает, что наука об информации есть часть теории познания. Поэтому информатика, как и математика, претендует на то, что она *не служанка, а царица наук*. Математику называют посредником между духом и материей [13]. С гораздо бóльшим основанием можно утверждать, что не столько математика, сколько **информация – посредник между духом и материей** [14]. Вводя

необходимое в таких случаях перерегулирование, можно утверждать, что теория информации имеет оправдание своего существования именно как раздел теории познания. Поэтому необходимо и к понятию информации подходить с позиции теории познания. Такой подход означает, что к информации нельзя подходить только с позиций количества, а необходимо учитывать *цель* системы и *качество* информации [14]. То, что информация имеет отношение к духу, а не только к материи, не ускользнуло от бдительного ока цензоров первого советского издания работы Шеннона [15]. С целью освобождения издания «от ненаучных идеалистических построений» со ссылкой на отсутствие в теории Шеннона учёта смысла редакторы-материалисты заменили термин «информация» на термин «данные». Этот перегиб был исправлен во втором издании работ Шеннона [16], и теория информации стала основой множества систем связи.

Техника телевидения – это техника связи, а теория телевидения – раздел теории связи. Традиционно рассматривая связь как преимущественно радиосвязь, следует помнить, что А. С. Попов создал два вида связи: человек–человек и природа–человек. Если первый вид радиосвязи, имея социальное значение, всеми почитается, то второй вид (ведущий свою родословную от грозоотметчика Попова) незаслуженно находится в тени, хотя и имеет огромное значение не только для теории информации, но и для познания человечеством мира.

Следует воздать должное достаточно длительному этапу развития теории телевидения, основанного создателем электронного телевидения В. К. Зворыкиным [17], концентрирующему внимание на физических аспектах формирования, передачи и воспроизведения телевизионных сигналов. Но наступило время смены парадигмы: фотоприёмник – это не солнечная батарея, его цель, его выход – не энергия, а информация, не джоули, а биты. При этом телекамера – это не преобразователь информации (как компьютер), а устройство, олицетворяющее *рождение* информации, фотоприёмник – это не «датчик», телекамера не формирует «данные», и фотоприёмник, и телекамера – *кодер-накопитель*. Подчеркнём, что для существования информации с точки зрения теории связи не только необходим получатель (наблюдатель, адресат), но её до телекамеры нет – там, по ту сторону объектива, есть только хаос фотонов, из которого волей проектировщика и оператора рождается видеoinформация.

Телевидение сразу попало в поле зрения теории информации, его принципы обсуждали Р. Хартли [11], К. Шеннон [15,] [16] и Б. Оливер [18]. Так, Р. Хартли в 1928 г. обсуждал проблему слитности изображений, понимая её как связь *«требуемой степени расчленения на элементарные площадки и точности передачи интенсивностей этих площадок»* и *«интервалов между последовательными кадрами»* с *«обеспечением непрерывности зрительного восприятия»*. Его идею (*«...можно, например, вообразить систему, более сходную с прямым зрением, где временная функция, представляющая изменение интенсивности каждой отдельной элементарной площадки, передаётся по отдельной линии для получения непрерывно изменяющейся освещённости»*), которую пробовали реализовать в 1960-х гг., но без микроэлектроники эта затея была обречена на неудачу, можно считать предвосхитившей видеосистемы на кристалле с параллельной связью массива пикселей с кодером. Можно сказать, что Хартли предвидел технику телевидения XXI в.

К. Шеннон в 1948 г. с первых шагов изложения математической теории связи выделяет три класса систем: дискретную, непрерывную и смешанную, утверждая, что телевидение – непрерывная система. Б. Оливер в 1952 г., уже во время триумфального шествия теории Шеннона, ядром которой является теорема о блоковом, т. е. цифровом, кодировании, разделяет кодеры на два класса – без задержек и с задержками, и прямо заявляет, что *«телекамеры – пример кодирующих устройств»* (первого класса). Это написано в эпоху аналоговых телекамер, которые Оливер относит к кодерам, осуществляющим *«однозначное преобразование сообщения в новую физическую переменную»*. Он считает, что эти кодеры в отличие от блокового кодирования *«не требуют памяти и обладают малой и постоянной задержкой»*. Первый тезис – разделение Оливером кодеров на два класса – большинством разработчиков был забыт, и рассмотрение телекамер как кодеров в [2] вызывало удивление, потому что в телекамерах в явном виде нет арифметических и логических устройств, а то, что развёртка – это кодирование канала, в глаза не бросается. Вторая мысль Оливера (хотя и чрезмерно упрощает кодирование в телекамере) проясняет причину распространённого им заблуждения: он утверждал, что для правильного учёта статистических свойств сигнала необходимы *память и задержка*. Конечно, современные кодеры цифрового телевидения, используя память и задержку, весьма неплохо учитывают статистические свойства сигналов, до-

полная кодирование в телекамере. Если следовать Оливеру, то телекамера – однозначный преобразователь. Спросим: чего – каждого фотона? Нет. Обладает ли телекамера (согласно признакам, используемым Оливером) памятью и задержкой? Да. Мало того, телекамера осуществляет накопление по полю и по времени, в твёрдотельных фотоприёмниках в явном виде присутствуют элементы памяти – либо в виде отдельных массивов (ПЗС¹ с кадровым переносом), либо в каждом пикселе (КМОП² и ПЗС со строчным переносом). Наличие этих памяти и задержки ведёт к тому, что сжатие информации, осуществляемое в телекамере, на порядки превышает сжатие в последующих цифровых кодерах.

Трудности приложения теории информации к телекамерам проявились даже в том, что К. Шеннон писал [16]: *«В телевидении, например, источником информации является передаваемая сцена, сообщением – выходной сигнал телекамеры и сигналом – выходной сигнал передатчика»*. Получается странная ситуация: на схеме системы связи Шеннон показал, что сообщение формируется на выходе источника. Так куда отнести телекамеру? Если на её выходе сообщение, которое передатчик переделывает в сигнал, – то к источнику, а если источник – сцена, то телекамера входит в состав передатчика, т. е. является *первичным кодером*.

Если при построении теории каналов связи достаточно было считать источник сигнала, в роли которого выступал телеграфный ключ, заданным, то при построении видеоинформационной системы необходимо рассматривать источники принципиально иного типа. Поэтому телекамера, как двуликий Янус, является и кодером, и источником информации. Для потребителей видеоинформации, например, проектирующих канал связи или компьютерную систему обработки изображений, телекамера – источник «данных», для проектировщика телекамеры она, несомненно, – кодер. Мы развиваем подход, согласно которому **телекамера, и в первую очередь фотоприёмник, – это кодер**. Какой кодер – кодер источника или кодер канала? В фотоприёмнике осуществляется ряд воздействий на «сообщение»: фотоэлектрическое преобразование, накопление, дискретизация, развёртка. Первые операции могут трактоваться как кодер источника, последняя – как кодер канала, который традиционно необходим для преобразования светового потока как функции многих аргументов (пространство, вре-

¹ ПЗС – прибор с зарядовой связью.

² КМОП – комплементарный прибор со структурой металл–оксид–полупроводник.

мя, длина волны) в функцию одного аргумента – времени. Развёртка – телевизионное изобретение (про которую Н. Винер [12] говорил, что она может оказаться важнее самого телевидения), но она является следствием идей топологии, про которое Шеннон отмечал [16]: *«Этот тип отображений (отображения плоскости на линию), принадлежащий математику Кантору, может быть легко расширен в направлении сколь угодно большого уменьшения числа измерений... Область пространства большего числа измерений не может быть отображена в область пространства меньшего числа измерений непрерывно»*.

Эти разрывы непрерывности всегда были в телевидении – скачки от строки к строке, от кадра к кадру.

Вместе с тем развёртка, выполняющая задачу приспособления многомерного сигнала к одномерному каналу, позволила применить теорию информации, никак не повлияв на неё, не затронула основ теории связи, которые периодически пересматриваются. Так, Шеннон отмечал, что открытие новых систем связи ЧМ, ФИМ, КИМ *«привело к пересмотру оснований теории связи»*.

Аналогично на современном этапе развития науки рождение твёрдотельного телевидения привело *не только к технологическому прорыву* в виде фотоприёмных КМОП СБИС, объединяющих на одном кристалле и массив накопительных элементов (которые уже сами являются первичными кодерами), и устройства обработки видеoinформации, но и к *пересмотру оснований теории связи в части кодирования зашумлённых сигналов*. Почему именно твёрдотельное телевидение? Потому что во многих телевизионных системах с электронно-лучевыми трубками главным источником шума был каскад связи трубки с усилителем, а не фотонный шум фона (который впервые замаячил как главный источник помех в инфракрасных системах). Почему сразу к пересмотру оснований, а не просто некоторых уточнений, как говорится, в далёком знаке после запятой? Потому что для построения теории кодирования зашумлённых сигналов потребовались *новые понятия* и связывающие эти новые понятия *законы*. Конечно, к понятию «закон» следует относиться критически, о чём предупреждал Р. Карнап [19]: *«Может быть, было бы меньше неясности, если бы слово "закон" вообще не употреблялось в физике. Оно продолжает употребляться потому, что не существует никакого общеупотребительного слова для универсальных утверждений, которые учёные употребляют»*.

ют в качестве основы предсказания и объяснения. Во всяком случае, следует всегда иметь в виду, что, когда учёный говорит о законе, он просто обращается к описанию наблюдаемых регулярностей. Такое описание может быть точным или ошибочным. Если оно неточно, то учёный, а не природа несёт ответственность за это». Здесь уместно привести мнение М. Планка [20]: «Наука находит понятия, с которыми она работает, не готовыми; она впервые их искусственно создаёт и только постепенно совершенствует». Эту концептуальную установку как бы продолжает утверждение А. Пуанкаре [21]: «Часто для обнаружения родства фактов, аналогии, превращающей факт в закон, бывает достаточно изобрести одно новое слово, и это слово становится творцом». Поэтому основания теории связи – динамическая структура, ведь «догматика с течением веков эволюционирует» [6]. Это связано, в частности, с тем, что, согласно Гёделю, любая аксиоматическая система не полна и любая наука обречена на вечное совершенствование и, конечно, дополнение своих оснований.

Суть рассматриваемого пересмотра связана с тем, что Шеннон создал теорию информации [16], не определив, что такое информация, поскольку на самом деле его теория есть *теория количества информации*. Это подчеркивал сам автор, когда писал, что его теория определяет *количество информации, но не понятие информации как таковой*. Про информацию, столь же *фундаментальное* понятие для науки, как время, как и про многие другие сущности, можно сказать словами С. Лема: «...Таинственность этого явления, разгадка которого пока что глумится над нашей любознательностью...» – и заниматься доступным делом – разбираться в отношениях сущностей. Такой подход типичен для математических теорий, например, Н. Винер утверждал, что «*вряд ли стоит вдаваться в подробности относительно того, что стоит за понятием физической величины*», достаточно знать, что «это» можно измерить, а Ньютон на вопрос, что такое сила, прямо говорил: «*Я гипотез не строю*».

Вместе с тем обойтись без гипотез и удовлетвориться традиционным определением информации через какие-нибудь синонимы («данные», «сведения» или «энтропия») не конструктивно, нужно приемлемое «рабочее» определение, пусть и не идеальное. Тем более, что «данные» предполагают «дающего», а представить, кто «даёт» зрению человека или телекамере, трудно. Поэтому дальнейшие построения обязаны не противоречить основному тезису теории передачи информации, как её понимали осново-

положники К. Шеннон, Н. Винер, А. Н. Колмогоров: **информация существует только тогда, когда имеется её адресат.** Эта мысль теории связи находится в соответствии с концепцией квантовой физики: *«Классическая физика основывалась на предположении – или, можно сказать, на иллюзии, – что можно описать мир или по меньшей мере часть мира, не говоря о нас самих»* [22].

Для понимания информации важной подсказкой стала мысль А. Н. Колмогорова: *«Однако так как "информация" по своей природе не обязана быть (и в действительности не является!) скалярной величиной, то никакие аксиоматические исследования указанного направления не могут ответить на вопрос о том, сколь полно характеризует величина $I(\xi, \eta)$ интересующую нас информацию»* [23].

Для понимания информации продуктивной оказалась опора на учение о доминанте А. А. Ухтомского [24], где он писал: *«Господствующий очаг возбуждения, предопределяющий в значительной степени характер текущих реакций центров в данный момент, я стал обозначать термином "доминанта"».*

Сегодня это учение рассматривается как прочная методологическая платформа для понимания законов организации целенаправленного поведения животных и человека. А. А. Ухтомский никогда не ограничивал принцип доминанты областью только физиологических процессов, а увязывал его с фундаментальными вопросами человеческого бытия и познания (предисловие к [24]). Рассмотрение доминанты как цели позволило выдвинуть *принцип доминантной информации*, согласно которому *информация определяется триединством цели, качества и количества* [14]. Количество информации – устоявшееся понятие, другие компоненты информации в совокупности с ним – более новая конструкция. Качество информации – понятие весьма сложное, про которое даже считается, что его невозможно формализовать [25], потому что трудно формализовать понятие смысла. Мы будем в рамках проблематики теории связи говорить о синтаксической информации и понимать качество информации как насыщенность доступной информации доминантой. При этом такое насыщение наблюдателю (получателю) не есть данность, он должен наращивать качество информации в ходе кодирования. Поэтому понятие качества информации необходимо рассматривать в совокупности с его накоплением.

В.2. Принцип накопления качества информации

Ключевым элементом кодирования в телекамере является накопление. Необходимо разобраться: накопление чего? Более полувека господствовало мнение о том, что накопление относится к энергии или заряду. Вокруг этого понятия было много дискуссий начиная с периода перехода от механической развёртки с одноэлементным фотоприёмником к развёртке электронным лучом. Эти дискуссии отражает послесловие С. И. Катаева к книге [17], где он писал: *«Необходимо заметить, что в одном из кардинальных вопросов истории развития телевизионной техники, а именно в вопросе внесения в технику телевидения эффекта накопления или, как его часто не совсем удачно называют, «принципа накопления», в подавляющем большинстве работ нет необходимой точности. Дело в том, что упомянутого эффекта или принципа не изобрели ни В. К. Зворыкин, ни А. П. Константинов, ни С. И. Катаев, которым различные авторы по-разному приписывают эту заслугу. В действительности накопление электрических зарядов на ёмкости при помощи фототока, текущего в содержащей эту ёмкость цепи, использовалось для обнаружения слабых фототоков ещё при первых исследованиях внешнего фотоэффекта в прошлом столетии. Совершенно естественно, что использование этого накопления в телевизионном передающем устройстве возможно лишь при наличии в нём множества независимо работающих накопителей, что уже в 1911 г. в известной мере усматривал Кемпбелл-Свинтон, первым пытавшийся построить рациональную схему передающей трубки с дискретной структурой фотокатода и предусмотревший возможность использования в подобной трубке инерционных фотоэлементов.*

Не исключено, что явление накопления зарядов сыграло бы свою положительную роль при осуществлении этой и некоторых других старых передающих систем, основанных на использовании множества быстро переключаемых фотоэлементов, если бы при конструировании этих систем правильно учитывалось явление накопления зарядов...

Явный, чётко сформулированный учёт эффекта накопления для построения схемы чувствительного телевизионного передатчика был сделан Ч. Ф. Дженкинсом, предложившим в 1928 г. известную схему с накопительными элементами и коммутатором. Дженкинс, однако, не предусматривал сколько-нибудь реальных форм практического осуществления как самой панели накопителей, так и необходимого в этом случае быст-

родействующего коммутатора... Независимые предложения трёх различных авторов – В. К. Зворыкина (США), А. П. Константинова (СССР) и С. И. Катаева (СССР) – внесённые в период 1930–1931 гг., представляют собой различные по своей практической ценности попытки отыскания оптимальной формы осуществления накопительной схемы с электронным коммутатором».

Здесь кроме намёка на необходимость **оптимизации накопления** бросается в глаза разнообразие определений накопления – эффект, принцип, явление, а также скромность С. И. Катаева, который, по его словам, изобретал передающую трубку, а не принцип. Но почему в телевидении всё-таки говорят о принципе накопления? Потому что роль накопления фундаментальна, его не только Катаев называет кардинальным, но и философия отдаёт ему должное. Дело в том, что Катаев прав, накопление никто не изобретал – оно было всегда. Всегда было накопление жира организмом в благоприятный период и трата его в неблагоприятный, что впоследствии в ходе развития цивилизации превратилось в накопление зерна (потенциального жира) и ещё позже – информационной формы жира – денег. И создание библиотек – тоже накопление информации.

Философия приходит к накоплению через принцип экономии: *«Принцип экономии многолик. Накопление знаний, концентрация знаний, изобретательность, ловкость, проницательность, систематизация, ... чувство целостности концепции»* [6]. *«Экономия мысли и экономия труда, к которым... сводятся все стремления науки, являются источниками как красоты, так и практической пользы»* [21]. Комментируя введённый принцип доминанты, А. А. Ухтомский писал: *«Отсюда прямой вывод, что нормальная роль центра в организме есть не неизменное, статически постоянное и единственное его качество, но одно из возможных для него состояний. В других состояниях тот же центр может приобрести существенно другое значение в общей экономии организма»* [24]. Почему накопление сопоставляется с экономией? Потому что количество информации о накопленном определяется логарифмом, требующим существенно меньше знаков для записи, осуществляющим экономию в «бухгалтерии» любой из предметных областей.

Связь накопления и сжатия информации, следующая из рассмотрения накопления как проявления принципа экономии, в приложении к телевидению заставляет обратить внимание на соотношение сжатия информации

в телекамере и вне её, в цифровых кодерах. То, что сжатие в телекамере в ходе накопления в пикселах даёт существенно большее сокращение количества информации, чем сжатие в цифровом кодере [2], не исчерпывает различие этих методов. Дело в том, что в обеих этих телевизионных подсистемах осуществляется **накопление качества информации**. Подробно этот принцип рассмотрен в гл. 2, здесь укажем, что вопрос о качестве информации существенно глубже, чем вопрос о качестве изображений, обычно обсуждаемый в телевидении [26]. Мы будем развивать идею о том, что *качество информации определяется информационным риском, включающим потерю информации о доминантном сигнале и ложную информацию, состоящую из фоновой и шумовой информации*. При этом **принцип накопления качества информации** проявляется двояко: во первых, в ходе кодирования в телекамере качество информации нарастает от каскада к каскаду, от операции к операции (здесь и фотоприёмник, и усилители, и аналого-цифровые преобразователи, и цифровые кодеры); во-вторых, качество информации получает свой главный прирост в ходе накопления в дискретной структуре пикселей матричного фотоприёмника. Примыкает к этой концепции и рост качества информации по мере совершенствования технологии изготовления фотоприёмников в ходе всей истории развития телевидения от механического к электронному и далее к твёрдотельному. При этом технологические революции, обеспечивающие рост качества информации, побеждают не мгновенно. Так, в первые годы внедрения ПЗС со стороны приверженцев старой техники было много нареканий: недостаточная чёткость, большое количество дефектных элементов, недостаточная чувствительность и заметный темновой ток. С годами все эти недостатки были устранены, а явные преимущества остались. Типичное для сторонников вакуумной техники (обеспечивавшей в начале твёрдотельной революции лучшее качество видеоинформации) выражение «побалуются и бросят» кануло в Лету.

На поставленный в начале этого параграфа вопрос о том, что в пикселе накапливается, мы отвечаем так: конечно, сигналы имеют две ипостаси – материальную и информационную, и процедура накопления имеет *форму* накопления *числа* электронов, но цель операции кодирования в фотоприёмнике, её *содержание, сущность* – не материальная, а информационная.

Для телекамеры принцип накопления качества информации связан с понятием *группового счёта фотонов*. А. Роуз [27] поставил вопрос: какова

информация, содержащаяся в конечном числе фотонов? И приходит к выводу о необходимости реализации счётчика фотонов. Это означает, что *максимум количества информации*, извлекаемой из потока фотонов, достигается при их поштучном счёте. Счётчик фотонов реализуется в основном в телевизионной астрономии, когда фотонов очень мало и фоном можно пренебречь [28], а счётчик фотонов рассматривается не изолированно, а в сочетании с внешним накопителем. Это означает, что накопление качества информации осуществляется в несколько этапов:

- фотоэлектрическое преобразование;
- измерение параметров (координат) одноэлектронных событий;
- накопление информации во внешней памяти.

При обычных освещённостях, т. е. при большой плотности потока фотонов, с помощью накопления качества информации реализуется *групповой счёт фотонов*. Стимулом к такому группированию является, во-первых, чрезмерно большая *пропускная способность*, которая требуется для передачи координат большого числа фотонов – порядка единиц и десятков терабайт в секунду. Во-вторых, то, что для закона Пуассона, описывающего поток фотонов, дисперсия числа фотонов, т. е. фотонный шум, равна среднему числу фотонов. Поэтому этот код будет содержать в основном шумовую информацию, а доля полезной информации в нём будет пренебрежимо мала. Принцип математической статистики «чем больше данных, тем лучше» в телевидении трактуется двояко. Первая трактовка имеет положительный смысл, так как «каждый фотон нам интересен, каждый фотон нам дорог»: конечно, из одного фотона картину не построишь, и чем больше фотонов, тем картина точнее. Но вторая трактовка развивает отрицательный взгляд на этот принцип статистики и основана на борьбе с *шумовой информацией*. Сама идея группового счёта фотонов говорит о том, что получателя не интересует значение заряда, его интересует информация о нём.

Групповой счёт генерированных светом электронов, про который можно сказать, что это «**принцип накопления качества информации в действии**», может трактоваться как реализация *принципа минимума информации*, который заложен в определении скорости создания информации источником (по терминологии К. Шеннона *rate distortion*, по терминологии А. Н. Колмогорова *эпсилон-энтропия*). Очевидно, что этот минимум

информации определяется её заданным качеством; встречающиеся опасения в оправданности стремления к минимуму информации (опять «чем больше информации, тем лучше») некорректно абстрагируются от понятия качества информации.

Тот факт, что фотоприёмники накапливают генерированные светом электроны, не противоречит понятию о накоплении энергии. Однако накопление трактуется шире и отражает потребность в накоплении *информации*. Примером может служить задача об измерении координаты точечного объекта. С одной стороны, точность измерения растёт (среднеквадратичная ошибка убывает) пропорционально отношению сигнал/шум. С другой стороны, если проделать измерение несколько раз и усреднить его результаты, то при независимых случайных ошибках точность будет расти пропорционально корню квадратному из числа усредняемых оценок. Это означает, что при *отсутствии иных ошибок и шумов* точность измерения будет одинаковой при двух *различных способах накопления*:

- если накопить энергию светового сигнала за отведённый интервал времени T и измерить координаты объекта;
- если разбить интервал времени T на N частей, накопить энергию в каждой из этих частей, вычислить координаты для каждого из этих измерений (точность в каждом измерении уменьшится в \sqrt{N} раз) и усреднить полученные координаты (точность увеличится в \sqrt{N} раз).

Однако различие указанных способов радикально, так как для реализации второго из них потребуется в N раз бóльшая пропускная способность системы считывания сигнала изображения. Однако равенство точности измерения сигнала в этих двух способах показывает, что в информационных системах речь идёт в первую очередь о накоплении информации и лишь в подчинённом смысле можно говорить о накоплении энергии. Наиболее наглядно принцип *накопления качества информации* проявляется при рассмотрении семантики кадров. В частности, разведчики считают, что данные, собранные в результате разведывательной деятельности, не являются информацией до тех пор, пока они не пройдут аналитической обработки экспертом в соответствующей области.

В рассматриваемой здесь проблеме накопления качества информации необходимо определить, какие новые аспекты следуют из твёрдотельной революции в телевидении.

В.3. Твёрдотельная революция в телевидении

Этапы развития технологии в истории телевидения, включающие механическое, электронно-лучевое и твёрдотельное, подробно рассмотрены [3]. Здесь остановимся на твёрдотельной технологической революции – принципиально новом периоде развития телевидения, связанном с главной тенденцией современной техники связи – тотальным переходом к *«руко-творным кристаллам»*, т. е. большим интегральным схемам (ИС).

Сегодня большинство интегральных схем – цифровые, реализующие обработку информации, в том числе рождаемой телекамерами (а также *цифровыми* фотоаппаратами, которые выполняют и функции видеосъёмки). Авторы данной книги, ни в коей мере не являясь ортодоксами аналоговой техники, исходят из прагматического подхода к распределению функций кодирования информации в телекамере между аналоговыми и цифровыми методами: *если есть возможность выбора*, то кодирование нужно осуществлять в той форме, которая обеспечивает максимум качества информации. При этом в соответствии с принципом доминантной информации важнейшим инструментом повышения качества информации является групповой счёт фотонов, т. е. использование на каком-либо этапе кодирования сигналов в телекамере аналого-цифрового преобразователя (АЦП). При погружении внутрь кристалла фотоприёмника и «размножении» АЦП – сначала был один, затем несколько, после в каждом столбце и как *финал* – в каждом пикселе, но и в последнем случае аналоговое накопление и связь с компаратором остаются в рамках линейного аналогового кодирования.

Первый этап твёрдотельной революции в телевидении связан с появлением ПЗС, ответственных за рождение информации и осуществлявших накопление в аналоговой форме, и микропроцессоров, обрабатывавших информацию в цифровой форме [1]. Революционность изобретения и *реализации* ПЗС признали не только связисты, но и физики – за изобретение 1971 г. Бойл и Смит спустя почти 40 лет были удостоены Нобелевской премии. Фотоэлектрическое преобразование в полупроводнике (в твёрдом теле, кристалле) было освоено ещё в электронно-лучевом телевидении. Знаменитая книга «Телевидение» В. К. Зворыкина и Д. А. Мортонна в ряду физических принципов телевидения подробно отражает значимость свойств полупроводников и их описаний с помощью зонной теории. Параллельное *накопление* зарядового изображения в слое полупроводника тоже

давно было освоено в электронном телевидении. Поэтому твёрдотельное телевидение *на первом этапе* – это новый *технологический* метод реализации другого этапа рождения телевизионного сигнала – развёртки изображения. Поэтому ПЗС вполне могли бы называться *приборами с зарядовой развёрткой*.

Революционность появления твёрдотельного телевидения связана далеко не только с уменьшением габаритов. Огромное значение имеют стабильность параметров и жёсткость раstra телекамеры, сделавшая эффективной межкадровую обработку сигналов изображения.

Приобретя благодаря твёрдотельной технологии такие важные свойства, как надёжность, компактность, жёсткость раstra, прикладное телевидение стало незаменимым в системах безопасности различных объектов, на транспортных средствах, в космосе, армии, авиации и на флоте.

Принцип накопления качества информации не противоречит, а даёт важную «окраску» основной тенденции развития матричных фотоприёмников для прикладных систем – *увеличению количества передаваемой информации* путём увеличения числа элементов и быстродействия. При этом гонка за мегапикселями не отменяет оптимизацию накопления и квантования сигналов изображений.

Завершение первого этапа твёрдотельной революции в телевидении, т. е. переходного этапа от электронно-лучевого телевидения к компьютерному и твёрдотельному, совпало с осознанием человечеством наступления эры *глобальной информатизации*. Прежде количество передающих телевизионных центров было пренебрежимо мало по сравнению с количеством телевизоров. Теперь телевидение и видеотехника с помощью космических средств и интернета охватили всю Землю; число систем, *передающих видеоинформацию*, благодаря мобильным индивидуальным средствам, по-прежнему называемым *телефонами* (разговорное «мобильник» лучше отражает многофункциональность этих устройств), сравнялось с числом принимающих её.

Второй этап твердотельной революции, т. е. резкого изменения технологии, прикладная телевизионная техника переживает в последние годы благодаря переходу к КМОП-фотоприёмникам. *На первый взгляд* главное отличие этих приборов от ПЗС состоит в методе реализации *развёртки*. Координатная адресация к каждому элементу *накопления*, давно освоенная микроэлектроникой при производстве твёрдотельных устройств

памяти, исключает протаскивание заряда каждого элемента сначала по всему столбцу, а потом по всей строке и позволяет избежать искажений, связанных с неэффективностью переноса заряда. Тривиально также то, что благодаря относительной дешевизне по объёмам продаж (в штуках, не говоря о стоимости) и мобильным средствам связи КМОП-фотоприёмники обошли ПЗС уже более чем в 10 раз.

Более пристальный взгляд на отличие ПЗС и КМОП-фотоприёмников выделяет не менее, если не более значимые аспекты:

- новая технология позволила ввести *усиление до воздействия шумов считывания* (по терминологии электронно-лучевого телевидения – *докоммутационное усиление*). Это должно вытеснить не только вакуумные приборы усиления слабых оптических сигналов (электронно-оптические преобразователи, ЭОП), но и ПЗС;

- новая технология цифрового пиксела революционизирует реализацию *накопления в пикселе*, порождая невозможное ранее распределение накопления между аналоговым фотодиодом и цифровым счётчиком непосредственно в пикселе, который становится аналого-цифровым, т. е. примером смешанной системы связи;

- новая технология позволяет реализовать накопление фотоэлектронов, генерированных фотонами разных длин волн, на разной глубине полупроводника, что открывает путь построения цветных твёрдотельных матриц без специальных спектральных фильтров;

- новая технология открыла перспективу заметного повышения *качества* информации благодаря возможности *многоканального считывания* (в пределе – с применением отдельного АЦП в каждом элементе разложения);

- применение КМОП-технологии к реализации матричных фотоприёмников породило новую ветвь твёрдотельных устройств, объединяющих в себе фотоприёмник, устройства развёртки, квантования и обработки изображения, которые названы *видеосистемами на кристалле*.

Совокупность этих черт, проявившихся после изобретения активного пиксела, позволяет говорить о *революционном*, а не эволюционном характере развития уже самой твёрдотельной телевизионной техники. Э. Фоссам «всего лишь» ввёл усилитель в каждый пиксел, технология обращения к которому была известна по устройствам памяти, поэтому не известно, дадут ли ему Нобелевскую премию. Качественный скачок связан с освоением технологии ультрафиолетовой литографии, давшей субмикронную точ-

ность изготовления приборов: ведущими в микроэлектронике странами уже освоена точность 45 нм, ведутся работы по технологии с технологическими нормами 22 нм и менее. Вместе с тем такая высокая точность актуальна при создании цифровых СБИС, в первую очередь памяти, но для фотоприёмных КМОП-матриц, являющихся аналого-цифровыми (по терминологии Шеннона это смешанная система связи), для реализации аналоговых операций с точки зрения динамического диапазона обрабатываемых сигналов компромиссными являются более скромные нормы точности 90...180 нм. Этапы развития твёрдотельной технологии с использованием ультрафиолетовой литографии и их влияние на параметры твёрдотельных фотоприёмных матриц известны [3], закон Мура об экспоненциальном росте числа элементов СБИС распространяется и на фотоприёмники – с определёнными оговорками о «закате эры закона Мура» в силу физических, информационных (необходимый минимум числа электронов на бит) и экономических причин [3].

КМОП-фотоприёмник по сравнению с ПЗС дешевле и позволяет интегрировать на одном кристалле с ним ряд других элементов, в первую очередь – АЦП и интерфейсы каналов связи. Важнейшим фактором является размещение в каждом элементе КМОП-фотоприёмника нескольких транзисторов (при технологии с точностью 0,1 мкм – до 40, а при технологии 45 нм – до 256 [29], [30]). Эти элементы, дополняющие накопление усилением и квантованием, а в специальных структурах цифрового пиксела и распределяющих накопление информации между аналоговым фотодиодом и счётчиком числа превышений порога, иллюстрируют новый технологический этап не просто группового счёта фотонов, но принцип доминанты А. А. Ухтомского, согласно которому фоновые сигналы тормозятся, а доминантные усиливаются.

Важные технолого-экономические преимущества КМОП-фотоприёмников обусловили пристальное внимание к ним, и уже обоснованы перспективы заметного повышения их чувствительности вплоть до достижения лучших характеристик, чем у ПЗС. Идеи о потенциальных преимуществах КМОП-фотоприёмников изобретателем КМОП-матриц с активным пикселом Фоссамом высказывались давно, ещё в 1990 г. он опубликовал статью [30], в которой ПЗС назвал динозаврами. Ультрафиолетовая технология уже сделала КМОП-фотоприёмники конкурентоспособными, а в будущем эти представители нового этапа твёрдотельной революции должны окончательно вытеснить ПЗС [31].

Проектирование видеосистем на кристалле (и многих других сложных функциональных блоков, т. е. систем на кристалле) должно предусматривать совместное участие разработчиков и компонентов, и систем, т. е. технология микроэлектроники превращается из технологии создания компонентов в технологию выпуска систем [32].

Технология изготовления СБИС для прикладных телевизионных систем, в состав которых вошли системы на кристалле и видеосистемы на кристалле, изменила *методы теоретического синтеза и проектирования систем и устройств*. Новый этап порождает новый метод проектирования, который характеризуется высокой скоростью создания систем, резким возрастанием роли системного подхода и значимости программно-алгоритмического обеспечения. Проектировщики компонентов и систем прикладного телевидения вместе с освоением микропроцессоров и программируемых логических матриц должны освоить широкий набор средств САПР для эффективного использования систем на кристалле и видеосистем на кристалле, а иногда они бывают вынуждены сами создавать необходимые средства САПР.

На данном этапе развития наиболее перспективное направление работ видеоинформатики – создание интеллектуальных телевизионно-компьютерных систем, которые позволят реализовать системы технического и искусственного зрения, сопоставимые по характеристикам с биологическим зрением [33]. Сегодня этот прогноз следует читать как «заказ» создания систем, существенно превосходящих биологическое зрение не только по спектральному диапазону, но и по *качеству информации*. Такой прорыв весьма непрост. Сегодня, хотя Дж. фон Нейман и предупреждал о необходимости осторожности при обращении с такими утверждениями, не теряет актуальности его мнение [34]: *«Вполне возможно, что мы уже недалеко от пределов сложности, которую можно достичь в искусственных автоматах, не продвигаясь особенно в глубь теории информации»*. При этом на пути к крупным техническим достижениям необходимы усилия физиков и технологов, системщиков и оптиков, компьютерщиков и специалистов по теории информации. Например, на стыке теории информации и вычислительной математики быстро развивается теория вычислений с использованием модулярной арифметики, которая обеспечит повышение производительности и сбоеустойчивости вычислительной техники [3], [35]. Эти методы позволяют за счёт сокращения вычислительных за-

трат на цифровой раздел кодирования видеоинформации наиболее рационально распределять площадь кристалла между фотоприёмным и вычислительным сложнофункциональными блоками (СФ-блоками) видеосистемы на кристалле.

Не менее важны и работы, которые пытаются максимально широко охватить круг междисциплинарных проблем, включающих и выход на семантическую обработку информации, и взаимодействие человека с видеоинформационными системами, и цифровую технологию [36]–[38].

Таким образом, второй этап твёрдотельной революции в телевидении привёл к рождению нового класса и телевизионных систем с использованием видеосистем на кристалле. Перспективность массового внедрения достижений микроэлектроники в прикладное телевидение очевидна. Конечно, круг нерешённых задач ещё достаточно широк [39]. Для их решения необходимо технологию фотоприёмных КМОП-матриц и видеосистем на кристалле официально, на уровне Федеральных целевых программ, отнести к критическим технологиям федерального уровня, способным существенно укрепить научный, промышленный и оборонный комплексы России.

ЧАСТЬ I

ТЕОРЕТИКО-ИНФОРМАЦИОННЫЕ АСПЕКТЫ ПРОЕКТИРОВАНИЯ ТЕЛЕКАМЕР

Любая наука, будь это наука о природе или духе, начинает понимать свою задачу совсем не сначала, но, так сказать, в середине, и что она должна трудолюбиво искать её без перспективы когда-либо вполне её постигнуть. Наука находит понятия, с которыми она работает, не готовыми; она впервые их искусственно создаёт и только постепенно совершенствует.

Макс Планк

Глава 1. УРАВНЕНИЕ СВЯЗИ

Отмеченные на с. 5, 6 *проблемы*, проявляющиеся как *конфликты идей*, в этой главе будут сосредоточены вокруг двух таких разногласий. Первой конфликтующей идеей является основной результат математической теории связи К. Шеннона, состоящий в обосновании возможности достижения в *идеальной* системе нулевой ошибки передачи при бесконечной сложности кодирования и бесконечной задержке. Конкурирующая с ней идея состоит в необходимости учёта конечных значений сложности и задержки, и не просто учёта, но поиска *оптимального накопления* информации в реальном времени, задачу о котором поставил С. И. Катаев (см. с. 16).

Вторая группа конкурирующих идей относится к значимости критерия оптимизации системы связи. Первая из идей утверждает, что если в системе достигнута пропускная способность, то автоматически в этой системе достигается и минимум ошибки передачи (см. с. 38). Конкурирующая идея не просто констатирует факт о различии методов кодирования для достижения максимума пропускной способности и минимума ошибки передачи, но акцентирует требование оптимизации системы связи при главенстве достижения информационного равновесия между скоростью создания информации источником и пропускной способностью канала связи.

Попытки построения теории телевидения предпринимались многократно. Первые шаги [17] опирались на физические, а не информационные

модели. Это направление имело много продолжений (например [40]). Вместе с тем теория телевидения, претендующая на формализованный синтез систем, взяла на вооружение информационные оценки; различие подходов проявлялось в постановке задач и моделях сигнала, которые приводили к преобладанию методов либо теории решений [41], либо теории информации [42]–[45]. В теории телевидения много внимания уделяется пространственной дискретизации изображений [46], [47]. Хотя в инженерной деятельности важен учёт особенностей двумерных полей (разложение Цернике, анизотропия разрешающей способности зрения, различие корреляционных функций по разным аргументам), однако в первой части книги мы будем рассматривать преимущественно одномерные модели сигналов, предполагая аргумент обобщённым. Особенности кодирования телекамерой трёхмерных полей как функции пространственных аргументов и времени рассмотрены в гл. 5.

Математическая модель связи имеет два основных варианта в зависимости от модели сигнала – случайного сигнала [12], [16] и неслучайного сигнала с параметрической неопределённостью [48]. Обе эти модели используются в информационной теории телевидения. При этом опора только на модель неслучайного сигнала (теория решений) приводит к менее общим результатам, чем опора на модель случайного сигнала (теория информации).

Теория информации, конечно, позволяет получить оценки предельно достижимого качества передачи информации, но она в первую очередь нацелена на *синтез* кодера и декодера, причём акцент делается на кодирование, т. е. на обработку сигналов для противодействия *будущему* влиянию шумов канала. Формальной логической основой теории информации является в основном аксиоматика теории вероятностей, дополненная новыми понятиями, введение которых имеет характер аксиом. Аксиоматика Шеннона включала непрерывность связи вероятности и информации (в его терминологии – энтропии), аддитивность информации и её логарифмическую меру. Относительно опорного понятия шенноновской теории информации А. Н. Колмогоров писал [23]: *«Я настаиваю на той идее, что основным понятием, допускающим обобщение на совершенно произвольные непрерывные сообщения и сигналы, является не непосредственно понятие энтропии, а понятие количества информации $I(\xi, \eta)$ в случайном объекте ξ относительно объекта η »*. С учётом этого утверждения аксиомы Шеннона могут трактоваться так [3]:

1. Существует неотрицательная характеристика информационной связи двух случайных объектов, называемая *взаимной информацией*, которой разумно приписать логарифмическую связь с числом состояний и вероятностной структурой возможных значений случайных объектов.

2. Существует *минимум* взаимной информации, определяющий скорость создания информации источником, называемый *энтропией* источника.

3. Существует *максимум* взаимной информации, определяющий скорость передачи информации через канал – его *пропускную способность*.

Аналогично теория решений позволяет получить оценки потенциальной помехоустойчивости (ограниченной шумом), но нацелена она на *синтез* декодера, т. е. на обработку сигналов для противодействия *уже возникшим* шумам. Поэтому теория решений используется в теории информации при реализации декодера (фильтрация, квантование). В частных задачах обнаружения и оценивания параметров сигналов, когда кодирование до воздействия шума невозможно, теория решений позволяет синтезировать оптимальные фильтры и решающие устройства [12], [48], [49].

Родство теории информации и теории решений распространяется гораздо шире, чем использование общих моделей сигналов и помех и «подчинённой» роли теории решений как фрагмента теории информации, ответственного за построение декодера. Обе эти теории нацелены на построение *оптимальных систем*. К сожалению, часто встречаются мнения, что теория информации «*лишь даёт теоремы о существовании, и не даёт конструктивных рекомендаций*» и что «*практическое осуществление оптимальной системы, предсказываемой теорией статистического синтеза, не всегда может оказаться возможным или экономически оправданным*» [50], что ведёт к конструированию критериев типа «цена – качество». Вместе с тем в теории информации множество конструктивных результатов (и их оптимальность) может быть строго формализована на собственном языке этих теорий без обращения к экономике. Некоторые из таких результатов и аспектов родства этих теорий мы покажем в гл. 2.

Решённая Шенноном задача синтеза идеального универсального кодирования была встречена радистами с восторгом по нескольким причинам. Главной было обещание безошибочной передачи при наличии шумов канала – не случайно радиосвязь считалась важным шагом в развитии почтовой связи. Важна была и полученная для ограничения мощности на выходе кодера оценка пропускной способности, её даже сравнивали по зна-

чению с абсолютным нулём температуры [51]. Любовь связистов к этому решению определяется не только тем, что в *математической* теории связи Шеннона использовано *физическое* ограничение мощности. Вопрос «физичности» этого и других ограничений будет обсуждён в этой главе, но для связистов было важно, что в результате синтеза Шеннон пришёл к выводу об оптимальности равномерного распределения по спектру мощности передаваемого сигнала.

Почему это было безоговорочно принято? Во-первых, потому, что считалось само собой разумеющимся, что максимум пропускной способности соответствует минимуму искажений сигнала при передаче через канал. Это мнение продолжало доминировать (например [52]), хотя быстро стало известно, что минимум среднеквадратической ошибки при линейном кодировании при неравномерном спектре кодируемого сигнала ведёт к неравномерному спектру на выходе кодера [53]. Во-вторых, потому, что не теория связи, а деньги определяют стоимость затрат на аренду полосы частот: раз уж заплачено поровну за каждый герц полосы, то и надо её расходовать так, чтобы каждый герц нёс одинаковую нагрузку. Однако логика теоретика не есть логика «эффективного» менеджера, и в этой главе мы рассмотрим, какие черты приобретает статистический синтез кодера при различных возможных критериях и ограничениях. Кстати, в настоящее время не только исследуются, но и внедряются адаптивные системы связи, распределяющие, в полном соответствии с теорией Шеннона, мощность в отпущенной полосе не равномерно, а в зависимости от уровня помех в различных участках отведённого для передачи спектра, – так называемые системы *когнитивного радио*, использующие технологию динамической автономной коррекции параметров, в том числе неравномерное распределение мощности по спектру.

Практика передачи информации в реальном времени востребовала поиск других эталонов системы кодирования кроме кодирования Шеннона с бесконечными блоками. На этом пути были попытки модифицировать понятие скорости создания информации с введением понятия *эпсилон-энтропии* с задержкой [54], но, главное, было введено **уравнение связи** [25], [55] (см. 1.1). Это уравнение предназначено для нахождения информационного равновесия между скоростью создания информации источником и пропускной способностью канала связи при некоторой заданной совокупности ограничений при *эпсилон-энтропии* (не только среднеквадра-

тическая ошибка, но, например, задержка и/или другой показатель *качества информации*) и при пропускной способности (не только мощность на выходе кодера, но и другие его параметры, например сложность или усиление).

Сформулировать уравнение было важно, но без конкретных примеров, отличающихся от бесконечного блочного кодирования Шеннона, оно не представило бы значительного интереса. Авторам уравнения повезло, они такое решение нашли [56], [57], воспользовавшись несколько более ранней догадкой [58] о возможности использования не только канонизированного ограничения мощности на выходе передатчика, использованного Шенноном *для примера*, а «площади усиления» или широкополосности линейного кодера (суммы квадратов спектральных коэффициентов кодера). Опираясь всего лишь на одно это решение, выходящее за рамки общепринятых представлений, авторы уравнения связи выдвинули гипотезу (можно сказать смелую, а можно – интуитивно прозрачную) о существовании множества решений их уравнения [59]. Впоследствии авторы этой книги догадались, что для поиска решений уравнения связи можно варьировать не ограничение, а критерии качества, в результате чего было найдено решение уравнения связи для смешанной системы (непрерывный источник, дискретный канал, см. 1.3) [60]. Однако и в рамках непрерывной системы, для которой авторами уравнения связи было получено его первое решение, оказалось существует *бесконечное* множество решений. Для того чтобы это обнаружить, надо было догадаться, что можно изменять и критерий качества, и ограничения, которые могут иметь и не очень простую форму, например ограничение на кодер может включать линейную комбинацию мощности на выходе и широкополосности кодера [61], [62].

Здесь мы несколько раз употребили слова «догадка», «догадались». На первый взгляд это может показаться не серьёзным, но, пожалуй, следует прислушаться к авторитетному мнению Р. Фейнмана [63]: «*В догадках нет ничего ненаучного, хотя многие не занимающиеся наукой и думают, что это так*». Эту мысль он поясняет: «*Угадывание законов природы – это действительно искусство... Именно философия помогает нам строить догадки*». Другой физик-теоретик, один из основоположников теории квантовых вычислений Д. Дойч, от имени бога Гермеса, которого он называл богом знаний, посланий и *потока информации*, заявил [5]: «*Догадки, хоть и дают начало всему знанию, также являются источником ошибок,*

и поэтому важно, что случается с идеей после того, как до неё догадались». Схожего мнения придерживался и Д. Пойа [64]: «*Результаты творческой работы математика – доказательное рассуждение, доказательство, но доказательства открывают с помощью правдоподобных рассуждений, с помощью догадки*».

К чему привели эти догадки и насколько убедительны результаты их осмысления – показывают следующие параграфы. Эту убедительность, конечно, следует рассматривать с учётом мнения Р. Л. Добрушина (комментарии в книге [23]), отмечавшего, что работы «*на уровне физической строгости противопоставлены познавательным возможностям среднего математика*», но уж таковы возможности авторов – инженеров, а не математиков по образованию.

1.1. От идеального кодирования к равновесному

Формулирование уравнения связи имело целью распространить задачу Шеннона об информационном согласовании источника и канала связи на системы связи реального времени, т. е. на системы, не удовлетворяющие требованию сколь угодно больших задержек. Логика авторов сводилась к уравниванию ошибок, возникающих при кодировании и в канале связи. Эта идея возникает из весьма странного решения [16], при котором на первом этапе спектр сигнала отсекается по *некоторой (!) частоте W_0 (без оценки потери информации)*, а затем кодируется неограниченно сложным кодером и передаётся по каналу с шумом с *нулевой ошибкой*.

Уравнение связи помогает распространить теорию Шеннона для безошибочной передачи на другие способы кодирования, исходя *не из критерия неискаженной передачи, а из критерия информационного равновесия*, когда количество создаваемой информации равно количеству передаваемой информации [65]. Уравнение сформулировано для максимально обобщённой модели системы связи (рис. 1.1), в рамках которой осуществлён её статистический синтез [25], [55], [66].

Модель стохастического канала без выделения кодера и декодера формализует техническую систему, расположенную между источником и получателем. Звенья триады описываются известными статистическими характеристиками. Сигнал – это случайная величина x (или в общем случае случайная функция $x(t)$), которая характеризуется априорным распределением $p(x)$ (априорной плотностью вероятностей).



Рис. 1.1. Обобщённая модель системы связи

Стохастический фильтр-канал описывается переходной вероятностью $p(z/x)$ получения выходного сигнала z с распределением $Q(z)$, если источник создаёт реализацию сигнала x . Конструктор выбирает форму z выходного сигнала, исходя из функции качества $r(x, z)$, назначаемой получателем, и совокупности ограничений. Принцип создания оптимальных систем передачи информации моделируется триадой «источник – стохастический фильтр – получатель», и главным в этой триаде традиционно является показатель качества информации, определяемый функционалом от трёх аргументов и включающий функцию потерь $r(x, z)$:

$$R(q, p, r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x) p(z|x) r(x, z) dx dz,$$

а требование получателя к качеству характеризуется функционалом R и уровнем ошибки передачи ε :

$$R(q, p, r) \leq \varepsilon, \quad \varepsilon \geq 0.$$

Информационная теория связи появилась с того момента, когда, используя статистическое описание системы связи, Шеннон предложил меру количества средней взаимной информации между сигналами x и z с помощью следующего функционала:

$$I(x, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} q(x) p(z|x) \log \frac{p(z|x)}{Q(z)} dx dy = I(q, p). \quad (1.1)$$

Функционал Шеннона (1.1) «выполняет организующую роль в теории связи, но одного функционала недостаточно для учёта всего разнообразия конкретного проявления качества информации в оптимальной системе связи» [25].

Информационное равновесие в системе означает равенство скорости создания информации источником (эпсилон-энтропии H_ε)

$$H_\varepsilon = \min I(x, z) = \min I(q, p)$$

и скорости передачи информации через канал

$$C = \max I(x, z) = \max I(q, p).$$

Все системы передачи информации синтезируются при наличии некоторой совокупности ограничений $\{E_j\}$, при которых должно достигаться равновесие между скоростью создания информации источником и скоростью её передачи по каналу связи.

Вместе с тем основная теорема Шеннона имеет характер неравенства: если энтропия источника не превышает пропускной способности канала, то благодаря использованию бесконечно длинных кодов с бесконечной задержкой возможна безошибочная передача информации через канал с шумом. Однако проектирование систем связи нуждается в поиске оптимальных систем передачи в реальном времени, в которых информационное равновесие – равенство скорости создания информации источником и пропускной способности канала связи – достигается для заданной совокупности условий абсолютно строго. Такое равновесие может быть достигнуто с помощью решения **уравнения связи**, включающего совокупность ограничений [25], [55], [66]. В качестве примера такого ограничения К. Шеннон использовал ограничение на мощность сигнала в канале связи [16]. Проектировщик должен учитывать большую совокупность ограничений, часть которых $\{E_i\}$ может относиться к источнику, а часть $\{E_n\}$ – к каналу. Поэтому и в уравнение связи [25], [55], [66] кроме количества взаимной информации входят указанные ограничения (варьируемые переменные обозначены обычными буквами, результирующие значения, обеспечивающие экстремальное значение функционала, – буквами со звездочкой):

$$\max_{\{E_i\}; q = \text{var}} I(q, p^*) = \min_{E_n; p = \text{var}} I(q^*, p) = I(q^*, p^*). \quad (1.2)$$

В уравнение связи входит условный минимум информации $I(q^*, p)$ в допущении, что уже выбрано распределение источника q^* . Минимум ищется путём вариации распределения в канале p . Максимум достигается путём вариации распределения источника q . Оба условных экстремума информации $I(q, p)$ при достижении статистического согласования источника и канала совпадают и равны $I(q^*, p^*)$. Поэтому решение уравнения (1.2) – это совокупность распределений сигналов источника и канала. Уравнение связи показывает, что в несингулярном случае (т. е. при конечной задержке) при достижении равновесия энтрон-энтропия H_ϵ источника равна энтрон-пропускной способности C_ϵ канала, т. е. они сходятся в

«седловой» точке и становятся одной и той же величиной $I(q^*, p^*)$ [65]. Эта величина равна минимальному количеству информации, которое формирует согласованная пара, и максимальному количеству информации, которое эта пара способна пропустить. Выбор максимизирующего априорного распределения q^* и оптимального кодера p^* должен производиться так, чтобы обеспечить потребителя информацией необходимого качества.

Уравнение связи, опираясь на функционал (1.1), не противоречит математической теории связи Шеннона, а развивает её, включая в множество решений не только идеальную систему с бесконечными блоками и нулевой ошибкой, но и оптимальные системы с искажающими каналами реального времени. Автор уравнения связи Л. И. Хромов утверждал [25]: *«Главный вывод состоит в том, что оказался возможным эволюционный переход от теории скалярной информации к концепции нескялярной информации. Более того, новая теория переходит в старую в предельном случае. Не понадобилось отказываться ни от количества информации по Шеннону, ни от его важного вывода о существовании конечной пропускной способности канала, вызванной наличием шума».*

Форма записи уравнения связи (1.2) через плотности вероятности акцентирует внимание на том, что проектировщик системы связи, и проектировщик телекамеры в первую очередь, ищет и находит не только кодер и декодер, но и сам источник информации. Вместе с тем в ряде случаев, например при синтезе оптимальной системы кодирования в гауссовском канале, когда вариация относится только к спектральным характеристикам, уравнение связи можно записать с опорой на априорную информацию о спектре сигнала x и на синтезируемый спектр сигнала z в канале:

$$\min_{\{\text{критерий}\}} I(x, z) = \max_{\{\text{ограничения}\}} I(x, z). \quad (1.3)$$

Запись уравнения связи (1.3) акцентирует внимание проектировщика на том, что решение связи должно включать способы формирования кодов на выходе источника и в канале. В первую очередь, как будет показано далее, решение уравнения обеспечивается нахождением оптимальных кодера и декодера в системе связи.

Авторы уравнения отмечали, что оно имеет не одно решение, а целый класс решений, среди которых и идеальная шенноновская пара – источник с равновероятными символами на выходе и неискажающий канал связи,

к которым К. Шеннон приходит на уровне супербуков [25]. Ещё одним из решений уравнения связи, выходящим за рамки теории Шеннона для дискретных систем, является идеальное линейное кодирование [57], но вопрос о существовании каких-либо ещё решений уравнения оставался на уровне декларации.

Обычно термин «оптимальный» понимается как единственный, наилучший для конкретных условий и конкретного критерия качества. Если меняются условия или критерий качества, то меняется и оптимальное решение. Стоит ли говорить при этом о наличии разных оптимальных решений? Мы считаем, что стоит, и для этого есть несколько причин. В первую очередь, синтез системы связи включает не только синтез канала, но и синтез *источника*, в роли которого выступает телекамера. Уравнение связи олицетворяет *неантагонистическую* игру синтеза источника и канала, поэтому и «конкретные» условия фиксированы не жёстко и проектировщик вправе преобразовывать функции распределения сигналов с использованием кодирующих устройств как в канале, так и в источнике. Во-вторых, большинство задач синтеза в статистической радиотехнике относится к оптимальному *приёму* сигналов [48]–[50], [68], [88], где действительно нет такой свободы вариации параметров системы – нет возможности управлять распределениями сигналов *до воздействия шума*. При этом и не возникает вопрос об информационном равновесии. Однако если кодирование до воздействия шума возможно, то можно найти оптимум для некоторых условий, но может остаться открытым вопрос о достижении информационного равновесия. Такой случай известен – это линейное кодирование гауссовских сигналов по критерию минимума среднеквадратической ошибки (о нём подробно далее). Оптимум есть, а равновесия – нет. Может ли проектировщик при этом успокоиться? Полагаем, что нет, и он вынужден искать не просто оптимальную систему связи (*источник плюс канал*), но такую, в которой достигается информационное равновесие.

Для того чтобы найти другие решения уравнения связи, представляется полезным перейти от самой общей схемы рис. 1.1 к более конкретизированной схеме (рис. 1.2), которая явно разграничивает компоненты статистического синтеза системы связи. Далее вместо согласованной пары «источник – канал» или триады «источник – стохастический фильтр – получатель» рассмотрим триаду «априорная информация – критерий качества – совокупность ограничений».

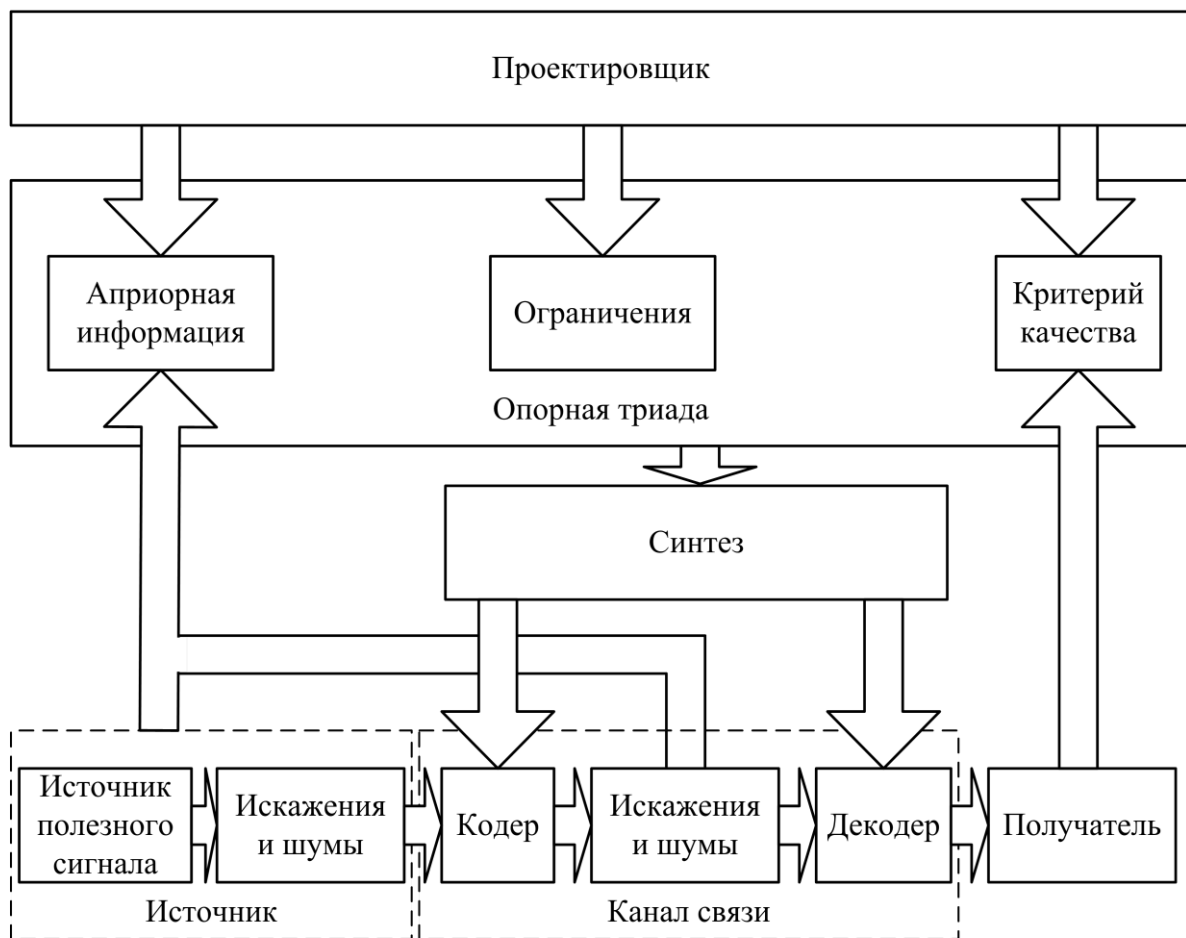


Рис. 1.2. Модель синтеза системы связи на основе опорной триады, которую проектировщик стремится сделать сопряжённой для обеспечения информационного равновесия в скорости создания информации источником и скорости её передачи через канал связи; кодер включает кодер источника и кодер канала, декодер – декодер канала и декодер источника

Создавая математическую теорию связи, Шеннон сразу написал: «Мы можем грубо разделить системы связи на три главные **категории**: дискретные, непрерывные и смешанные. Под дискретной системой будем понимать систему, в которой и сообщение, и сигнал являются последовательностями дискретных символов. Непрерывная система – это такая система, в которой и сообщение, и сигнал рассматриваются как непрерывные функции. Смешанная система – это такая система, в которой встречаются и дискретные, и непрерывные переменные» [16].

Отметим, что существует по крайней мере две группы смешанных систем: дискретный источник при непрерывном канале связи и непрерывный источник при дискретном канале связи. К первой группе относятся системы связи, кодирование источника которых получило название кодово-импульсной модуляции (КИМ) [16]; ко второй – каналы с замираниями и

т. п. (например [67]). На практике комбинаций больше, так как системы связи часто многозвенные.

В ходе поиска решений уравнения связи, т. е. равновесных пар «источник – канал», следует рассматривать необходимую и достаточную для оптимизации системы связи **опорную триаду**, скрытую в уравнении связи, и используемую во всех задачах статистического синтеза, начиная с работ Т. Байеса в XVIII в., в новое время – с работ Н. Винера [12] и К. Шеннона [16]. Она включает:

- **априорную информацию** (плотность вероятности источника и переходную вероятность канала, в частности автокорреляционные и взаимокорреляционные функции сигналов и помех или их спектры);
- **критерий качества** (может трактоваться как ограничение при вычислении энтальпии-энтропии);
- совокупность **ограничений**, накладываемых на кодеры и декодеры (*категория* системы, мощность, задержка, сложность и т. д.).

Такое чёткое разграничение компонентов триады перспективно, во-первых, в силу конструктивности получаемых результатов синтеза; во-вторых, в силу встречающихся невнятных трактовок статистического синтеза. Например, совокупность правил выбора решений и функцию потерь (ядро критерия качества) иногда относят к априорной информации [68]. Кроме того, иногда путают критерии и ограничения, относя последние к «критериям второго рода». Такая инверсия критериев и ограничений часто выводит проектирование из рамок теории связи и ведёт к неконструктивным мнениям о том, что критерий максимума экономической эффективности целесообразно выбрать основным, а остальные критерии взять в качестве ограничивающих. Применение новых технологий, конечно, позволит более эффективно решать задачи кодирования, так как численное значение ограничений (например на ёмкость памяти и быстродействие) по мере совершенствования технологии всё более «играет в пользу» разработчика систем. Это открывает путь к реализации всё более сложных алгоритмов, но всё равно оставляет сложность системы в ранге ограничений.

Можно синтезировать систему связи (найти оптимальные кодер и декодер) при любой комбинации компонентов опорной триады, чем и обусловлено огромное многообразие реализаций систем связи и публикаций о них. Однако далеко не во всех из них достигается равновесие в скорости создания информации источником и скорости её передачи через канал.

Так, в теории связи известны некоторые локальные оптимумы, не обеспечивающие информационного равновесия, понимаемого как равенство энтрон-энтропии источника и пропускной способности канала связи.

Первым важным примером такого локального оптимума, т. е. рассмотрения *несопряжённых* триад, является *тест-канал* Галлагера, формализованный им как способ достижения энтрон-энтропии для двух случаев: одномерной гауссовской случайной величины [52] и гауссовского случайного процесса [52]. Тест-канал Галлагера содержит кодирующий и декодирующий *линейные* фильтры. Их спектральная характеристика обеспечивает минимум взаимной информации (реализует энтрон-энтропию) при заданной полной среднеквадратической ошибке, но не обеспечивает, по мнению автора, равенства энтрон-энтропии и пропускной способности. Он утверждал [52]: «*Даже если имеется соответствующий тест-канал, пригодный для достижения искажения d^* со скоростью $R(d^*)$, то можно достигнуть среднего искажения, меньшего, чем d^* , если использовать дополнительную обработку данных и увеличить среднюю взаимную информацию в канале выше $R(d^*)$* ». Причина такой ситуации кроется в том, что Галлагер и его последователи считали само собой разумеющимся использование следующей опорной триады:

- априорная информация: независимые гауссовский источник и гауссовский шум в канале;
- критерий качества: минимум среднеквадратической ошибки;
- ограничение: дискретный канал с заданной мощностью передаваемого сигнала.

В результате достигаемую энтропию сравнивали с пропускной способностью при ограничениях (категория канала и ограничение мощности), *не согласованных* с априорной информацией и критерием качества.

Вторым важным примером локального оптимума является фильтр Костаса [53] (см. также [69], [70]), обеспечивающий минимум среднеквадратической ошибки при линейном кодировании для ограничения мощности непрерывного сигнала. Аналогично тест-каналу Галлагера кодирующий фильтр Костаса опирается на несогласованные между собой компоненты опорной триады. Вместе с тем фильтр Костаса очень важен в теории связи, так как он показывает, что для конкретного случая гауссовского канала **критерий минимума среднеквадратической ошибки приводит к суще-**

ственному отличию оптимальной спектральной характеристики кодера от «идеального» кодера, обеспечивающего максимум пропускной способности *при том же ограничении*. Этот принципиальный момент делает достаточно странной надежду на достижение минимальной ошибки именно при достижении пропускной способности канала, высказанную Р. Галлагером и ставшую общепринятой. *Разные критерии при одной и той же априорной информации и одинаковом ограничении приводят к разным кодирующим фильтрам*. Это означает, что **можно достичь меньшей ошибки, чем в канале, реализующем пропускную способность**.

Третий важный пример – кодирование в смешанном канале с помощью квантования непрерывной величины – КИМ [16], [71]–[75]. Рассматривая кодирование случайной величины, традиционно анализируют опорную триаду, включающую гауссовскую случайную величину, критерий минимума среднеквадратической ошибки и дискретный канал с ограничением мощности. Эта триада несогласованная, поэтому главным выводом анализа стало утверждение [72], что эpsilon-энтропия не достижима даже при оптимальной расстановке порогов.

Эти примеры стимулируют поиск согласованности в опорной триаде статистического синтеза. *Опорную триаду, обеспечивающую решение уравнения связи, назовём сопряжённой триадой*.

Сопряжённая триада – новое понятие, которого в отличие от понятия опорной триады не было до введения в арсенал статистического синтеза уравнения связи, хотя необходимость согласования априорной информации, критерия качества и ограничений в теории связи известна давно. Во-первых, характеризуя теорию линейной фильтрации Колмогорова–Винера [12], Бодэ и Шеннон писали [16], что абсолютная оптимальность этой фильтрации имеет место для гауссовских сигналов и помех (априорная информация) и при использовании критерия минимума среднеквадратической ошибки. Для других плотностей вероятности оптимальность фильтрации характеризуется ограничением в виде линейности. Этот же смысл имеет и утверждение, что *«использование того или иного критерия качества определяется располагаемыми априорными данными»* [68]. Во-вторых, теория Шеннона [16] для передачи информации по дискретному каналу связи показывает возможность достижения информационного равновесия в дискретном канале связи при ограничении мощности. Вместе с

тем важно, что Шеннон нашел сопряжённую триаду для *дискретной* системы (рассмотренная им смешанная система КИМ не обеспечивает информационного равновесия при выбранной им опорной триаде), но его последователи пытались синтезировать и анализировать системы *других категорий* с помощью тех же компонентов триады, не являющейся для них сопряжённой. Это означает, что для достижения информационного равновесия в произвольной системе связи необходимо видоизменить всю опорную триаду, *приведя её к сопряжённой*.

Рассмотрим решения уравнения связи отдельно для трёх категорий систем в соответствии с классификацией Шеннона [16]. При этом для каждой из категорий остановимся лишь на двух крайних случаях допустимой задержки – не более 1 символа и неограниченной задержки, не останавливаясь на промежуточных вариантах произвольной задержки [54] и на особенностях передачи зашумлённых сигналов, рассматриваемых в гл. 2. Развиваемый тезис о существовании *неограниченного множества сопряжённых триад, обеспечивающих решение уравнения связи*, т. е. строго оптимальных систем связи, только укрепляется при таком сокращении рассматриваемых случаев.

При рассмотрении и непрерывной, и смешанной систем считается, что многомерный сигнал кодируется с использованием преобразования Карунена–Лоэва (обеспечивающего статистическую независимость спектральных отсчётов для гауссовских сигналов) и находятся сопряжённые триады, обеспечивающие *строгое равенство* скорости создания информации и скорости её передачи через канал. Методы векторного квантования (например [73], [74]), использующие не преобразование Карунена–Лоэва из-за *сложности его реализации*, а числовые решётки и блочное кодирование конечной длины, обеспечивающие лишь *приближение* к информационному равновесию, здесь не рассматриваются.

В отличие от идеального кодирования Шеннона, предполагающего единственное решение, обеспечивающее информационное равновесие (при нулевой ошибке передачи) с помощью бесконечного блочного кодирования, уравнение связи определяет оптимальные системы для различных критериев качества и ограничений. Это позволило выдвинуть гипотезу [59] о возможности существования большого количества решений уравнения, т. е. большого количества оптимальных систем связи.

1.2. Дискретная система связи

Основные новые результаты теории информации, изложенные в книге, относятся к непрерывной и смешанной системам связи, но для характеристики общности подхода к обеспечению информационного равновесия, т. е. равенства скорости создания информации источником и скорости её передачи через канал связи, кратко охарактеризуем детально исследованную Шенноном дискретную систему связи. Шеннон дал формулы для вычисления пропускной способности дискретных систем связи в двух случаях – без задержки в кодере (только задержка на символ в декодере) и с неограниченной задержкой в блоковом кодере и декодере – и показал, что в этих системах возможно равенство эpsilon-энтропии и пропускной способности. При этом, не отмечая этого специально, он охарактеризовал для этих случаев введённые ранее сопряжённые триады.

Первый, простейший, случай – двоичный симметричный канал с шумом для передачи сигнала от двоичного симметричного источника (рис. 1.3). Для такого источника вероятности появления на выходе различных символов одинаковы. В дискретном симметричном канале передача отдельного символа характеризуется равными вероятностями ошибочного принятия гипотезы H_1 о передаче символа «1» при справедливости гипотезы H_0 о передаче символа «0» $p(H_1|H_0)$ и ошибочного принятия гипотезы H_0 о передаче символа «0» при справедливости гипотезы H_1 о передаче символа «1» $p(H_0|H_1)$:

$$p(H_1|H_0) = p(H_0|H_1) = p_\epsilon.$$

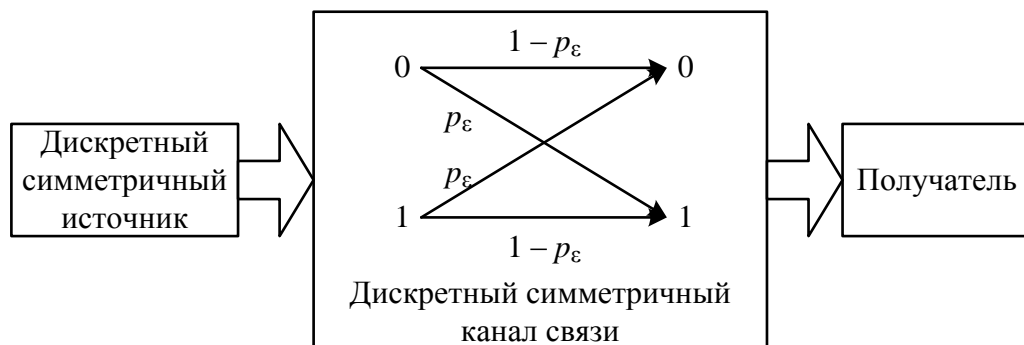


Рис. 1.3. Модель дискретной системы связи без памяти

Задание вероятности ошибки p_ϵ в дискретном симметричном канале в определённом смысле эквивалентно ограничению средней мощности в канале, поэтому вероятность ошибки в таком канале может быть отнесена

к ограничениям. Для априорной информации, состоящей в равенстве вероятностей передачи обоих дискретных символов «0» или «1», Шеннон выбрал согласованный критерий качества в виде допустимой вероятности ошибки передачи символа (т. е. удачно выбрал ограничение при вычислении эpsilon-энтропии) и обнаружил, что информационное равновесие достигается без всякого кодирования [16]; при этом можно записать, что достигается равенство $H_\varepsilon = C_\varepsilon \leq 1$ бит/символ. «Без всякого кодирования» означает, что не требуется универсальное блочное кодирование Шеннона, а следовательно, не требуются задержки при кодировании (в дискретном симметричном канале требуется лишь задержка на символ при декодировании с помощью согласованного фильтра и порогового устройства). Эта ситуация может быть трактуема как дополнительное ограничение на задержку передачи сообщения. Для образованной таким образом сопряжённой триады (априорной информации, критерия качества и ограничений) Шеннон нашёл значение пропускной способности [16] в битах на символ, подчеркнув, что *равенство эpsilon-энтропии и пропускной способности достигается без кодирования*:

$$H_\varepsilon = C_\varepsilon = 1 - [p_\varepsilon \log p_\varepsilon + (1 - p_\varepsilon) \log (1 - p_\varepsilon)].$$

Второй, общий, случай – дискретный канал с аддитивным белым гауссовским шумом для передачи дискретного сигнала. Для него Шеннон сконструировал *сопряжённую триаду* в виде произвольного алфавита и распределения вероятностей символов и заданной спектральной плотностью мощности белого гауссовского шума N_0 (*априорная информация*), заданной мощности в канале и *неограниченной задержки* передачи (*ограничение*) и вероятности ошибки в роли *критерия качества*. Чтобы подчеркнуть специфику такого канала, соответствующую ему пропускную способность обозначим как C_P . В силу сопряжённости этой триады в такой дискретной системе связи может быть достигнуто информационное равновесие $H_\varepsilon = C_P$ [16]. Для *этого* канала Шеннон показал, что существует идеальный кодер, который благодаря использованию блочного кодирования с неограниченной задержкой обеспечивает нулевую вероятность ошибки передачи, если только энтропия (эpsilon-энтропия) не превышает пропускной способности C_P . В силу того, что ошибка передачи становится нулевой, в частном случае дискретного симметричного источ-

ника с энтропией $H = 1$ бит/символ пропускная способность также достигает предельного для такого источника значения $C_P = 1$ бит/символ. Для произвольного источника при передаче через указанный дискретный канал пропускная способность имеет верхний предел в битах на символ (битах на герц) [16], где обозначим отношение мощностей сигнала и шума как $\Psi^2 = P/N_0$:

$$C_P = \frac{1}{2} \log(1 + \Psi^2).$$

Для этого типа каналов Шеннон показал, что в силу линейного увеличения дисперсии шума при увеличении полосы частот пропускная способность C_P не возрастает безгранично, а имеет предел $C_P = \frac{P}{N_0} \log_2 e$ [16]

(e – основание натуральных логарифмов).

Для многомерных случайных сигналов Шеннон показал, что пропускная способность определяется спектральным отношением сигнал/шум Ψ_k^2 и требует вычисления по всему бесконечному ряду. При использовании разложения сигналов в ряд Карунена–Лоэва, обеспечивающего для гауссовских сигналов статистическую независимость спектральных отсчётов, пропускная способность записывается как

$$C_P = \frac{1}{2} \sum_{k=1}^{\infty} \log(1 + \Psi_k^2).$$

Универсальность блочного кодирования Шеннона, опирающегося, в частности, на ограничение мощности, утвердила общее мнение об универсальности именно этого ограничения. Вместе с тем Шеннон писал: *«Если не ограничивать мощность передатчика, то пропускная способность будет бесконечной даже в случае наличия шума... Только в случае, когда имеется шум и мощность передатчика некоторым способом ограничена, получается конечная пропускная способность. Пропускная способность зависит, конечно, от статистической структуры шума так же, как и от того, каким образом ограничена мощность. Простейшим типом шума является белый тепловой шум или шум сопротивления. В этом случае распределение вероятностей амплитуд является гауссовским, а спектр равномерен в указанной полосе частот и может быть положен равным нулю*

вне этой полосы... Простейший способ ограничить мощность передатчика состоит в том, чтобы предположить, что средняя мощность, создаваемая передатчиком (или, более точно, среднеквадратичная амплитуда сигнала), не больше, чем P » [16].

Покажем далее, что для систем связи других категорий сопряжённую триаду образуют элементы, отличные от рассмотренных Шенноном в качестве хотя и важного, но всё-таки *примера*.

1.3. Смешанная система связи

1.3.1. Ограничение максимальной ошибки

Современные телекамеры, как правило, формируют на выходе цифровые сигналы и содержат иногда один АЦП, часто – множество, в том числе по числу столбцов и даже пикселей (см. гл. 3). Поэтому целесообразно в приложении к телекамерам опереться на модель смешанной системы связи, включающей непрерывный источник и дискретный канал.

Шеннон показал, что КИМ проигрывает идеальному блоковому кодированию примерно 8 дБ по требуемому отношению сигнал/шум в канале. Вместе с тем это не единственный показатель отличия КИМ от блокового кодирования дискретного сигнала. Существует проигрыш, измеряемый не по оси абсцисс графиков [16], а по оси ординат, составляющий примерно 2 бит/символ. Это различие складывается из двух компонентов – примерно 1,75 бит/символ из-за отказа от блокового кодирования и примерно 0,25 бит/символ из-за квантования непрерывной величины.

Эти 0,25 бит/символ являются следствием отличия функций плотности вероятности ошибки квантования (равномерная плотность) от сопряжённой с критерием минимума среднеквадратической ошибки гауссовской плотностью вероятности. Известные исследователи [72], опираясь на процитированное ранее мнение Галлагера о недостижимости информационного равновесия при кодировании непрерывных процессов [52], удивляются только тому, как немного можно достичь с помощью сложных систем кодирования независимых гауссовских величин (устранить указанную «избыточность» 0,25 бит/символ), если квантование уже оптимизировано по значениям порогов. Здесь скрыты две ошибки. Первая: *всё* возможное кодирование для передачи непрерывного сигнала по дискретному каналу – квантование по порогам, выравнивающим вероятности всех интервалов, – уже осуществлено, и эти 0,25 бит/символ *не устранимы*. Важно, что эта неустранимость принципиальна и не зависит от функции распределения

квантуемого сигнала. Вторая: удивляться здесь нечему, так как квантование – необходимая и достаточная процедура перевода непрерывной величины в дискретную, и если бы был правильно выбран критерий качества, то обнаружилось бы *абсолютное совпадение* достижимой скорости кода с эpsilon-энтропией.

Отсутствие ясности в проблеме оптимизации смешанной системы связи наглядно демонстрируется в утверждении [76], что при КИМ гауссовского сообщения скорость передачи *«несколько выше, чем это следует из формулы Шеннона»*. Путаница возникает из-за того, что фетишизация среднеквадратического критерия точности ведёт к вычислению среднеквадратической ошибки квантования и подстановке этого выражения

$\varepsilon = \frac{\delta^2}{12}$ в информационные формулы вместо дисперсии гауссовского шума,

тогда как Шеннон показал, что в этих формулах при произвольных плотностях вероятности должны фигурировать не мощности, а энтропийные мощности. Именно различие энтропийных мощностей априорного равномерного и апостериорного гауссовского распределений вероятностей порождает множитель πe в формуле для количества информации о координате видеоимпульса [48]. Этот же множитель в силу равномерной плотности ошибки квантования гауссовского сигнала содержится в формуле для разности эpsilon-энтропии при квадратичном критерии качества и пропускной способности при квантовании непрерывных сигналов

$$\Delta I = \frac{1}{2} \log \frac{\pi e}{6} \approx 0,25 \text{ бит/символ [71]–[75].}$$

Общепринято, что эpsilon-энтропию *«в общем случае трудно вычислить... Вся хитрость состоит в выборе распределения. Например, для квадратично-разностной меры погрешности естественно выбрать гауссовское распределение»* [72]. Применяя указанную «хитрость» при синтезе смешанной системы не надо фетишизировать весьма распространённые и оптимальные для линейных систем гауссовские модели сигналов и помех и критерий минимума среднеквадратической ошибки. Так же как переход от дискретной системы к непрерывной потребовал вслед за изменением априорной информации изменить критерий качества и ограничение, так и при переходе к смешанной системе (задание категории системы – это тоже ограничение) странно было бы ожидать при изменении ограничения сохранения критерия качества и сопряжённой с ним априорной информации.

Для достижения информационного равновесия в смешанной системе «непрерывный источник – дискретный канал» необходимо видоизменить всю опорную триаду. Существует распространенное мнение о недостижимости равенства эpsilon-энтропии и пропускной способности при передаче по дискретному каналу сигнала непрерывного источника, но мы покажем, что есть равновесные триады, обеспечивающие решения уравнения связи для смешанной системы связи с непрерывным источником и дискретным каналом. Решения уравнения связи существуют и для смешанной системы, к этому выводу можно прийти, если рассмотреть усечение числового ряда [59] или учесть финитную функцию распределения сигналов и помех и критерия качества в виде максимальной ошибки [60].

Критерий качества. А. Н. Колмогоров ввёл две формы ε -энтропии, связанные со способом ограничения ошибки (т. е. предложил использовать по крайней мере два критерия качества), – через **единичную вероятность P неперевышения** расстоянием $\rho(\xi, \xi')$ между исходной величиной ξ и её аппроксимацией ξ' заданного значения ε либо через математическое ожидание M дисперсии этого расстояния: «*Кажется естественным потребовать, чтобы $P\{\rho(\xi, \xi') \leq \varepsilon\} = 1$, или чтобы $M\rho^2(\xi, \xi') \leq \varepsilon^2$. Эти два вида « ε -энтропии» распределения P_ξ будем обозначать $H_\varepsilon^0(\xi)$, $H_\varepsilon(\xi)$ » [23].*

А. Н. Колмогоров дал формулу для эpsilon-энтропии при квадратичном критерии, которой пользуются при анализе непрерывных систем (обоснованно) и смешанных систем (без должного основания). Для меры максимального отклонения есть лишь оценки ε -энтропии $H_\varepsilon^0(\xi)$, ограниченные системой неравенств, но она им специально сконструирована для анализа аналого-цифрового преобразования. Мера максимального отклонения, введённая А. Н. Колмогоровым [23], осталась фактически незамеченной (хотя на его работы имеются ссылки в обзоре [75]). Для анализа смешанных систем типично заявление: «*Теория А. Н. Колмогорова допускает применение различных мер точности. Мы будем рассматривать только те случаи, когда мера точности – средний квадрат ошибки*» [73]. Только когда уравнение связи вошло в арсенал средств оптимизации, было намечено [3] применение эpsilon-энтропии с мерой максимума отклонения.

Для упрощения рассуждений видоизменим меру ошибки при вычислении эpsilon-энтропии, записав вместо вероятностного критерия критерий **максимального отклонения** принятого сигналов от передаваемого (сохраняя обозначение H_ε^0 для подчёркивания значения вида ограничения)

и приняв для наглядности ограничение ошибки $\varepsilon/2$ (ε приобретает смысл шага квантования):

$$H_\varepsilon^0 = \min_{\max|x-z| \leq \varepsilon/2} I(X, Z).$$

Остальные два компонента опорной триады рассмотрим так же, как и при рассмотрении дискретной и непрерывной систем, отдельно для двух случаев, различающихся допустимой задержкой передачи: с передачей каждого исходного символа и при допустимости блочного кодирования.

Передача одномерной непрерывной случайной величины. *Априорная информация.* В смешанной системе связи необходимо перейти к использованию источника полезного сигнала с финитной плотностью вероятностей $p(x)$ ($x \in [0, A]$, $A < \infty$), согласованной с мерой максимума отклонения. В простейшем случае – это равномерная плотность вероятностей $p(x) = 1/A$. Нефинитные плотности вероятностей сигнала (в частности гауссовская) из-за неограниченности ошибки на краях диапазона при мере максимального отклонения сопряженных триад не образуют. Шум в канале с посимвольной передачей для образования сопряженной триады также должен иметь нулевое математическое ожидание и финитную плотность вероятностей $p(n)$, $n \in [-N/2, N/2]$. В простейшем случае, как и сигнал, шум может иметь равномерную плотность вероятностей $p(n) = 1/N$.

Ограничения. Кроме ограничений на категорию системы и на посимвольную передачу в сопряженную триаду должно войти ограничение на усиление g перед квантованием.

Аналогично одномерному тест-каналу Галлагера для непрерывной системы в смешанной системе (рис. 1.4) можно найти «тест-квантователь», а на самом деле идеальный кодер, опирающийся на сопряженную триаду. Кодер в системе с посимвольной передачей в качестве первого этапа осуществляет умножение на коэффициент g и квантование сигнала с шагом δ на m интервалов, $m = gA/\delta$. При вычислении требуемой пропускной способности канала общепринято для упрощения записей полагать, что число разрядов квантователя l – целое число ($l = \log_2 m$). Если плотность вероятностей сигнала $p(x)$ равномерна, то после квантования дополнительного кодирования для выравнивания вероятностей символов по Шеннону–Фано не требуется. При произвольной финитной плотности вероятностей такое выравнивание выполняется на втором этапе кодирования.

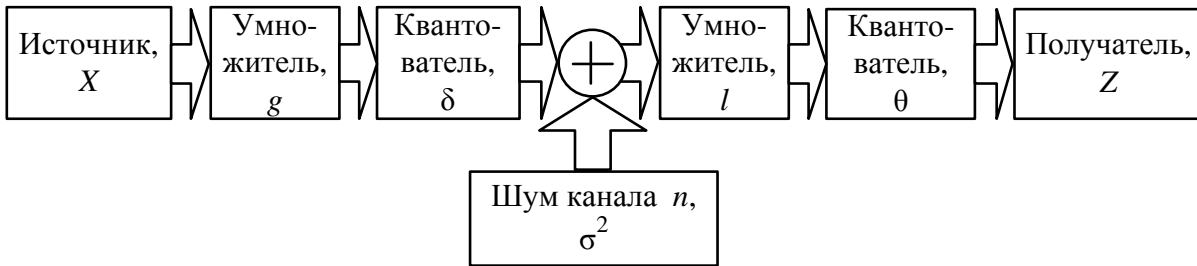


Рис. 1.4. Модель смешанной системы связи без памяти

В равновесной системе шаг квантования в кодере δ выбирается равным размаху шума в канале N . Декодер осуществляет умножение результата на коэффициент $q = 1/(g + N/A)$, квантование принятого сигнала по порогу $\theta = q(\delta + N)/2 = qN$.

При посимвольном кодировании в смешанной системе для обеспечения информационного равновесия шаг квантования в декодере θ определяется из условия равенства эpsilon-энтропии H_ϵ^0 и пропускной способности C_ϵ , вычисляемых в битах на символ [60]:

$$H_\epsilon^0 = \log \frac{A}{\theta} = C_\epsilon = \log \left(1 + \frac{gA}{N} \right).$$

Равенство достигается при коэффициенте усиления кодера и декодера, вычисляемых аналогично усилению в тест-канале Галлагера для непрерывной системы, но не через мощности входящих величин, а через их максимальные значения:

$$g = \frac{N}{\theta} \frac{A - \theta}{A}.$$

Формула для пропускной способности родственна формуле Талера [77], которая отличается от очень похожей формулы Шеннона лишь смыслом входящих величин. Черри [51] инкриминировал Талеру игнорирование того факта, «что шум обладает каким-то амплитудным распределением и имеется определенная вероятность достижения **любого** (выделено Черри) амплитудного уровня». Проведенный анализ показал, что формулы Талера и Шеннона отличаются не просто смыслом букв, а относятся к разным сопряженным триадам: формула Талера опирается на критерий максимальной ошибки, финитные плотности вероятностей сигналов и помех и ограничения на пиковую мощность и задержку передачи, а формула Шеннона – на критерий вероятности ошибки, на гауссовскую плотность вероятностей помехи и ограничение на среднюю мощность (при неограниченной задержке).

Передача многомерной непрерывной случайной величины (случайного процесса). Решение уравнения связи для смешанной системы, осуществляющей передачу непрерывного сигнала по дискретному каналу, получено для следующей комбинации априорной информации и ограничений [60].

Априорная информация. Как и при синтезе непрерывной системы, считаются известными спектр полезного сигнала (совокупность $\{A_k\}$ коэффициентов разложения в ряд Карунена–Лоэва) и спектральная плотность мощности белого (возможно, гауссовского, а возможно, и с иной, например, финитной плотностью вероятностей) шума в канале. Как и при передаче одномерной величины, информационное равновесие возможно при использовании источника полезного сигнала с финитной плотностью вероятностей для каждой спектральной компоненты, $p_k(x)$, $x \in [0, A_k]$, $\sum_{k=1}^{\infty} |A_k| < \infty$, согласованного с критерием качества в виде максимума отличия сигналов.

Ограничения. Сопряженные триады для такой априорной информации можно сформировать для двух комбинаций ограничений при двух различных плотностях вероятностей шума в канале:

1) для финитной плотности вероятностей:

- совокупное усиление кодера перед квантованием $\sum_{k=1}^{\infty} g_k \leq G$;
- задержка передачи;

2) для гауссовской плотности вероятностей шума в канале снимается ограничение на задержку и, кроме ограничения на усиление, возникает второе – на мощность сигнала в канале:

- средняя мощность передаваемого сигнала $\sum_{k=1}^{\infty} g_k^2 A_k^2 \leq P$;
- неограниченная задержка передачи.

Кодер в смешанной системе осуществляет следующие операции:

- разложение сигнала в ряд, обеспечивающий статистическую независимость спектральных отсчётов для гауссовских сигналов (декорреляцию источника);
- умножение спектральных компонент на частотно-зависимые коэффициенты g_k ;

- квантование первичных символов (спектральных коэффициентов) с шагом δ , усекающее в общем случае бесконечное число членов ряда до конечного значения K , каждый из которых передается m_k уровнями;
- цифровое кодирование, выравнивающее вероятности вторичных символов (при неравномерной финитной плотности вероятностей спектральных коэффициентов A_k);
- при гауссовском шуме канала – блочное кодирование.

Структура выражений для эpsilon-энтропии, пропускной способности и ошибки передачи аналогична таковой для непрерывной системы с точностью до меры ошибки и ограничений [60]:

$$H_\varepsilon^0 = \sum_{k=1}^K \log \frac{|A_k|}{\theta}, \quad C_\varepsilon = \sum_{k=1}^K \log \left(1 + \frac{g_k |A_k|}{N} \right).$$

Равенство достигается при коэффициентах усиления кодера и декодера, вычисляемых аналогично усилению в тест-канале Галлагера для непрерывной системы, но не через мощности входящих величин, а через их максимальные значения:

$$g_k = \frac{N |A_k| - \theta}{|A_k|}.$$

Как и в одномерном канале, назначаемая ошибка в полосе пропускания распределяется поровну между ошибкой квантования в кодере и шумом канала и распределена по спектру аналогично линейному кодированию:

$$\varepsilon = K\theta + \sum_{k=K+1}^{\infty} |A_k|.$$

Таким образом, кроме известных ранее четырёх решений уравнения связи (по два для дискретного и непрерывного каналов) существуют решения для смешанного канала – одно для одномерного сигнала и два для многомерного.

К. Шеннон отмечал: «*Основные идеи теории связи не новы... В более поздних исследованиях, однако, учитываются факторы, которые раньше игнорировались, в частности, теперь значительно глубже понимается действие шума в канале и важность статистических свойств сообщений, предназначенных к передаче*» [16]. Аналогично сейчас, после введения в статистический синтез систем передачи информации уравнения связи, как уже видно из оптимизации смешанной системы и будет показано далее в первой части книги, стала глубже пониматься важность **сопряжения**

априорной информации о сигнале и помехах, критерия качества передачи и совокупности ограничений.

1.3.2. Ограничение сложности кодера источника

Сложность кодера источника в телекамере представляет особый интерес в связи с построением видеосистем на кристалле, в которых в рамках одной СБИС размещаются и фотоприёмный массив, и кодер источника [3]. Сложность кодирования традиционно рассматривают в приложении к кодированию канала [72]. Далее делается попытка учёта сложности кодирования источника без учёта сложности кодирования канала.

Существует много подходов к определению сложности, и центральным положением здесь выступает понятие алгоритмической сложности, введённое А. Н. Колмогоровым [23]. Его фундаментальное значение связано с тем, что сложность определяется минимальной длиной программы, необходимой для задания объекта. Это, в отличие от вероятностной теории, позволяет находить сложность индивидуальных объектов. Минимизация программы достигается благодаря опоре на рекурсивные функции (в том числе частично рекурсивные). Алгоритмический взгляд на сложность может быть применен при проектировании твердотельных фотоприёмников – он даёт нижнюю границу количества информации, передаваемой на промежуточных этапах проектирования, так как описание топологии новой СБИС допускает частичную рекурсию в силу использования десятков миллионов однотипных элементов (например пикселей). Однако при реализации самой СБИС на кристалле её сложность, влияющая на площадь кристалла (как следствие, и на цену, и на процент выхода годных кристаллов), определяется совокупностью всех элементов видеосистемы на кристалле. Поэтому сложность СБИС обычно характеризуют использованным числом транзисторов или вентиляей. Далее не будем детально разбираться в структурах кодеров, отметим лишь главные тенденции при их реализации на кристалле.

В круг учитываемых величин, связанных с кодированием источника (показателей качества системы связи), должны войти не только точность передачи (ошибка) и скорость передачи, но и сложность. Известно, что эти величины взаимосвязаны: *«Оптимизацию видеокодека надо делать (минимум) по трём параметрам: по битовой скорости, по искажению и по вычислительной сложности. Все они влияют друг на друга. Например, оптимизация соотношения скорость/искажение достигается за счёт по-*

вышения сложности кодирования, "быстрые" алгоритмы оценки движения часто имеют низкую вычислительную сложность за счёт снижения эффективности кодирования и т. д. Эффективность кодирования и сложность кодирования являются настоящими антиподами» [78].

Обобщённый показатель эффективности кодирования должен учитывать не только меру приближения к энтропии источника, но и то, какими информационными средствами достигнуто данное приближение, в первую очередь – какой сложностью кодера.

В соответствии с методологией оптимизации радиоэлектронных устройств по совокупности показателей качества [41], [79] целесообразно сформировать функционал, включающий взвешенную сумму частных показателей качества. Вектор весовых коэффициентов при показателях качества в системном анализе принято называть «вектором концепции системы» [80]. Очевидно, что в различных системах эти векторы могут существенно различаться, что и приводит к существованию множества различных оптимальных систем [80]. Например, даже при равенстве весовых коэффициентов при скорости передачи и ошибке передачи системы кодирования непрерывных источников в телевизионном вещании и в аэрокосмических системах наблюдения существенно различаются в соотношении весовых коэффициентов при сложности кодера и декодера. В телевизионном вещании в силу огромных тиражей приёмников существенный вес имеет сложность декодера, в аэрокосмических системах связи в силу жёстких ограничений передатчиков основной вес приходится на сложность бортового кодера.

Для обеспечения общего информационного подхода к решению задачи оптимизации системы связи необходимо все частные показатели качества привести к единому виду, имеющему единицу измерения информации (бит и т. п.). В частности, учёт деградации отношения сигнал/помеха следует характеризовать потерей полезной информации ΔI , вычисляемой как логарифм отношения фактической ошибки к минимально возможной, определяемой входным шумом и искажениями сигнала до кодирования.

Следуя методике векторного синтеза системы связи [79], целесообразно сформировать обобщённый показатель эффективности P , включающий «вектор концепции системы» $\{c_i\}$ и совокупность $\{P_i\}$ частных информационных показателей качества системы связи – потери полезной информации ΔI (который может учитывать как пространственные искажения изображения, так и задержку на кодирование, например, увеличиваю-

щую промах при прицеливании), скорости передачи R , сложности W_K кодера и декодера W_D . Так как все частные информационные показатели качества системы связи связаны с ошибкой передачи ε , то далее они будут обозначаться как $P_i(\varepsilon)$:

$$P = c_0 \Delta I(\varepsilon) + c_1 R(\varepsilon) + c_2 W_K(\varepsilon) + c_3 W_D(\varepsilon) \rightarrow \min .$$

Для проектирования видеосистем на кристалле в обобщённом показателе эффективности целесообразно пренебречь сложностью декодера:

$$P = c_0 \Delta I(\varepsilon) + c_1 R(\varepsilon) + c_2 W_K(\varepsilon) \rightarrow \min .$$

Обобщённый показатель эффективности кодирования позволяет объективно сравнивать произвольные кодеры, отличающиеся по всем параметрам – ошибке передачи, скорости передачи и сложности, если зафиксировать выборку испытательных сигналов. Искомый минимум при вариации назначаемой ошибки передачи ε может быть найден аналитически, если известны статистические свойства источника и зависимости сложности и скорости передачи от назначаемой ошибки. Известно [79], что обобщённый показатель качества допускает множество решений. Это разнообразие решений в первую очередь связано с отмеченным взаимообменом скорости передачи и сложности кодера [78].

Исследование взаимообмена скорости передачи и сложности кодера при применении различных методов, как с переходом в спектральную область, так и без него, показало [81] возможность формализации введённого критерия. Такая формализация включает введение нижней границы H_ε^W ошибки передачи при заданной сложности W (которая, естественно, больше границы H_ε эпсилон-энтропии при отсутствии ограничения сложности) и гипотезу об одинаковом влиянии на точность передачи скорости передачи и сложности кодирования непрерывного источника. Это позволило ввести функционал, связывающий эпсилон-энтропию с ограничением сложности H_ε^W и сложность кодирования W . Этот функционал восходит к гармоническому среднему и отличается от него отсутствием множителя числа слагаемых. Эта модификация гармонического среднего хорошо известна в физике: так вычисляются фокусное расстояние линзы по расстояниям до изображения и объекта, сопротивление параллельно включённых резисторов [81]:

$$H_\varepsilon = \frac{H_\varepsilon^W W}{H_\varepsilon^W + W} .$$

Наличие нижней границы скорости передачи при фиксированной ошибке передачи (рис. 1.5) подтверждается экспериментально [81]. Представление взаимоотношения скорости и сложности может быть дано номограммой (рис. 1.6), аналогичной известным в оптике [13].

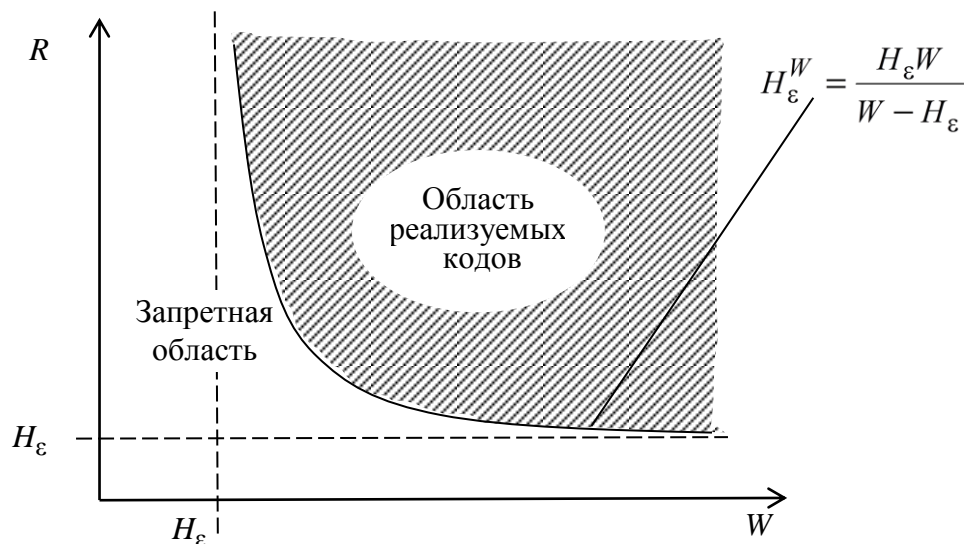


Рис. 1.5. Область реализуемых кодов, обеспечивающих скорость передачи R при сложности W

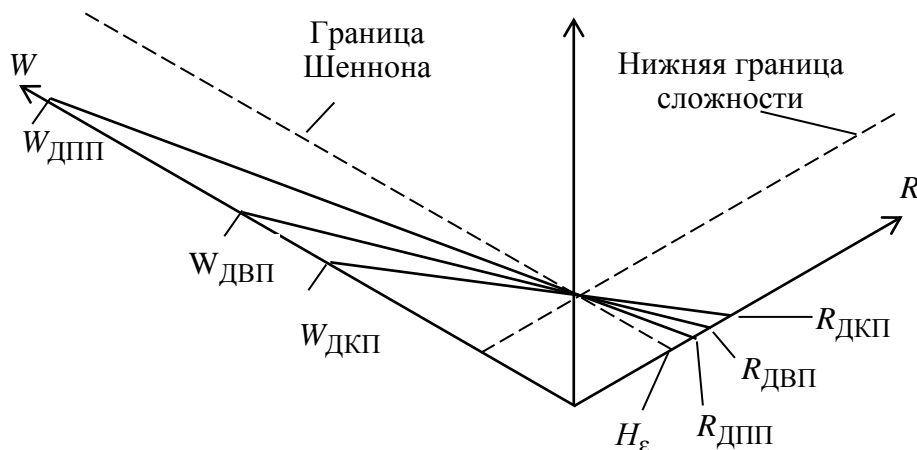


Рис. 1.6. Номограмма взаимоотношения скорости передачи и сложности для различных методов кодирования: ДКП – дискретное косинусное преобразование; ДВП – дискретное вейвлет-преобразование; ДПП – дискретное пространственное преобразование

Её анализ показывает, что численные значения скорости передачи и сложности кодирования в эксперименте хорошо совпадают с предложенным функционалом: линии, соединяющие сложность и скорость передачи для каждого из методов кодирования, пересеклись в одной точке на вертикальной оси номограммы, которая соответствует эpsilon-энтропии источ-

ника. Выявленная формализованная зависимость скорости передачи от сложности кодера наглядно показывает, что при построении видеосистем на кристалле погоня за предельной точностью ведёт к нерациональным тратам сложности и, как следствие, площади кристалла.

Это определяет обоснованный подход к выбору весовых коэффициентов в обобщённом критерии качества кодера.

1.4. Непрерывная система связи и множественность решений уравнения связи для неё

Роль аналоговых (непрерывных) систем связи в телевидении состоит отнюдь не только в использовании в старом телевизионном вещании (о чём так и писали Р. Хартли и К. Шеннон), но и в том, что *самые перспективные фотоприёмные КМОП-матрицы с цифровым выходом и даже с цифровым пикселом всегда будут содержать аналоговый канал связи* фотодиода-накопителя, ответственного за *рождение информации*, с АЦП. Значение этого канала велико, хотя его размеры и измеряются микрометрами, **оптимальный непрерывный канал – эталон для частотных характеристик пиксела**, который пользователи называют «точкой» [36]. Для разработчиков фотоприёмных матриц и телекамер пиксел далеко не точка, его структура достаточно сложна во всех трёх измерениях (см. гл. 3).

Вместе с тем распространено мнение, что люди, занимающиеся аналоговыми системами остались *«во времена Очакова и покоренья Крыма»* [36]. Это мнение разделял и Р. Л. Добрушин, который в личной беседе с одним из авторов говорил, что линейные методы (а значит, и аналоговая связь) сохраняют своё значение только в учебном процессе в силу простоты принципов, но будущее за корректирующими кодами. Такая оценка крупного специалиста по теории информации приводимых в первой части книги результатов не только отражает прорыв в области цифровых технологий, но может считаться достаточно высокой, если исходить из тезиса К. Черри [51]: *«Проблемам кодирования уделяется значительное внимание не всегда с целью решения практических задач, а скорее, подобно другим теоретическим исследованиям в области связи, в поисках общих закономерностей»*.

Конечно, цифровые каналы связи и компьютерные средства обработки информации крайне важны, но они относятся к *использованию информации*, а телекамеры ответственны за её *рождение*. Кстати, цифровые дан-

ные, о которых часто говорят [15], [36], телекамера даёт этим пользователям. А сама телекамера? Ей никто данных не даёт. Отмахнуться от аналогового канала в фотоприёмных матрицах нельзя, даже если пользователи сигналов, порождаемых телекамерами, имеют дело с цифровыми данными.

При анализе и синтезе непрерывных систем часто в качестве эталона рассматриваются гауссовские сигналы и шумы. Так, теория оптимальной фильтрации Колмогорова–Винера использует эти модели совместно с критерием минимума среднеквадратической ошибки. Как мы покажем далее, вполне законно, так как гауссовские сигналы и шумы (априорная информация) и критерий минимума среднеквадратической ошибки входят в сопряжённую триаду синтеза оптимального кодера в канале, который называют гауссовским. Роль этих распределений и критерия задолго до рождения теории связи отмечал А. Пуанкаре [21]: «Закон Гаусса опирается на следующие гипотезы: общая ошибка есть результирующая очень большого числа частных и независимых ошибок; каждая из частных ошибок очень мала и подчиняется закону вероятности – какому угодно, при одном неизменном условии: что вероятность положительной ошибки та же, что и вероятность ошибки, равной и противоположной по знаку... Закон наименьших квадратов является законным не во всех случаях, но из-за незнания статистик ошибок – лучшее, что мы можем сделать, – это принять метод Гаусса».

1.4.1. Информационное равновесие при ограничении широкополосности

Так как кодирование в гауссовском канале при критерии минимума среднеквадратической ошибки и ограничении мощности не позволяет достичь информационного равновесия [70], несмотря на существование локального оптимума (фильтр Костаса [53]), интерес к линейному кодированию быстро иссяк. Вместе с тем использованную триаду можно привести к сопряжённой. При этом в рамках гауссовского канала можно изменять либо критерий качества, либо ограничение, либо и то, и другое.

В данном параграфе рассмотрим путь достижения информационного равновесия сменой ограничения. Как и для дискретной системы связи, рассмотрим отдельно два случая – без задержки в кодере (только задержка на символ в декодере) и с допустимой задержкой в кодере, и покажем, что в этих системах возможно равенство энтальпии и пропускной способности при выборе соответствующей сопряжённой триады.

Для одномерной гауссовской случайной величины с нулевым математическим ожиданием и дисперсией λ , передаваемой по непрерывному каналу с аддитивным гауссовским шумом известной дисперсии σ^2 (*априорная информация*), Галлагер сконструировал сопряжённую с этой априорной информацией триаду и получил тест-канал [52], реализующий эpsilon-энтропию, вычисляемую в соответствии с формулами Шеннона [16] и Колмогорова [23] для среднеквадратической ошибки, не превышающей заданного значения θ (*критерий качества*):

$$H_\varepsilon = \frac{1}{2} \log \frac{\lambda}{\theta}, \quad \varepsilon \leq \theta \leq \lambda.$$

Так как линейный кодер в такой системе есть просто множитель на коэффициент передачи g (предельное значение которого и есть *ограничение*, не позволяющее достичь сколь угодно малого значения ошибки θ), можно считать, что мощность сигнала на выходе кодера есть дисперсия сигнала, умноженная на квадрат коэффициента передачи $P = \lambda g^2$. Тогда пропускная способность, согласно формуле Шеннона, составляет:

$$C = \frac{1}{2} \log \left(1 + \frac{\lambda g^2}{\sigma^2} \right).$$

Информационное равновесие или решение уравнения связи, обеспечивающее равенство эpsilon-энтропии и пропускной способности *в одномерной непрерывной системе* достигается при значении коэффициента передачи [52]

$$g^2 = \frac{\sigma^2}{\theta} \cdot \frac{\lambda - \theta}{\lambda}.$$

Если обозначить отношение дисперсий сигнала и шума $\lambda/\sigma^2 = \Psi^2$, а отношение $\sigma^2/\theta = a^2$ трактовать как квадрат *доступного усиления* (ограничение), то оптимальный линейный одномерный кодер (тест-канал Галлагера) можно записать в виде

$$g^2 = a^2 \frac{\Psi^2 - 1}{\Psi^2}.$$

Для многомерной гауссовской случайной величины (процесса) с нулевым математическим ожиданием и спектром дисперсий $\lambda(\omega)$, передаваемой по непрерывному каналу с аддитивным гауссовским шумом известной дисперсии σ^2 (*априорная информация*), Галлагер сконструировал много-

мерный тест-канал, отличающийся от одномерного лишь наличием зависимости дисперсии передаваемого сигнала от частоты ω [52]. Однако из-за давления аналогии в виде универсальности блочного кодирования Галлагера недооценил свой тест-канал (фильтр), утвердив его неполноценность для реализации равенства эpsilon-энтропии и пропускной способности канала.

Аналогичная ситуация сложилась и с фильтром Костаса [53]. Костас не ставил задачу достижения информационного равновесия, но более поздние исследования показали [70], что хотя условный минимум среднеквадратической ошибки и достигается, равенство эpsilon-энтропии и пропускной способности этот кодер не обеспечивает. Оба эти результата говорят о том, что не была найдена сопряжённая триада.

Сопряжённая триада для линейного кодирования непрерывных процессов была предложена достаточно давно (авторы работы [58] предложили для достижения информационного равновесия сменить тип ограничения). Эта сопряжённая триада, как и для тест-канала Галлагера, включает *априорную информацию* в виде спектров гауссовских сигнала и аддитивного белого шума, *критерий качества* в виде минимума среднеквадратической ошибки, но другое *ограничение*. Позже эта идея была развита и распространена на кодирование случайных процессов конечной длительности T при описании их с помощью совокупности $\{x_k\}$ коэффициентов разложения в ряд Карунена–Лоэва с дисперсиями $\{\lambda_k\}$ с преобразованием сигнала $\{x_k\}$ в код $\{z_k\}$ [56], [57] (шум канала связи традиционно принимается белым с постоянной дисперсией спектральных отсчётов σ^2). Позже было показано, что равенство эpsilon-энтропии источника и пропускной способности канала связи в непрерывной системе связи является одним из решений уравнения связи [25].

При использовании в качестве оптимального декодера фильтра Винера среднеквадратическая ошибка вычисляется следующим образом:

$$\sum_{k=1}^{\infty} \frac{\lambda_k \sigma^2}{\lambda_k g_k^2 + \sigma^2} \leq \varepsilon. \quad (1.4)$$

Вместо ограничения мощности на выходе кодера, использованного Костасом [53] и подразумевавшегося Галлагером [52]:

$$\sum_{k=1}^{\infty} \lambda_k g_k^2 \leq P, \quad (1.5)$$

в сопряжённую триаду для многомерной непрерывной системы связи при ограничении среднеквадратической ошибки входит ограничение широкополосности («площади усиления») кодера:

$$\sum_{k=1}^{\infty} g_k^2 \leq E. \quad (1.6)$$

Очевидно, что мощность, передаваемая в канал связи, при заданной спектральной мощности передаваемого непрерывного сигнала конечной мощности и при использовании ограничения широкополосности, тоже конечна. Поэтому уместно напомнить слова Шеннона, процитированные на с. 43 и 44 о том, что ограничение мощности – простейший способ сделать пропускную способность конечной при наличии шума в канале. Ограничение широкополосности тоже есть некоторый способ ограничения мощности передатчика.

Количество взаимной информации, входящее в уравнение связи, для непрерывной гауссовской системы связи (рис. 1.7) вычисляется по формуле:

$$I(x, z) = \frac{1}{2} \sum_{k=1}^{\infty} \log \left(1 + \frac{g_k^2 \lambda_k}{\sigma^2} \right), \quad (1.7)$$

где g_k – спектральные коэффициенты передачи кодирующего фильтра.

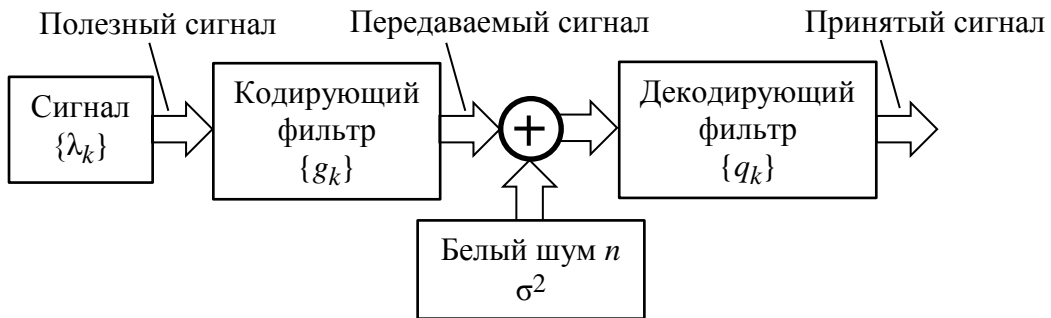


Рис. 1.7. Модель гауссовской системы

Эпсилон-энтропия гауссовского источника информации вычисляется по формуле Колмогорова [23]:

$$H_\varepsilon = \frac{1}{2} \sum_{k=1}^{\infty} \log \left(\frac{\lambda_k}{\theta} \right), \quad \varepsilon = \theta K + \sum_{k=N+1}^{\infty} \lambda_k.$$

Уравнение связи (1.2), (1.3) для линейного кодирования при ограничениях (1.4) и (1.5) принимает вид

$$\min_{\sum_{k=1}^{\infty} \frac{\lambda_k \sigma^2}{\lambda_k g_k^2 + \sigma^2} \leq \varepsilon} \frac{1}{2} \sum_{k=1}^{\infty} \log \left(1 + \frac{g_k^2 \lambda_k}{\sigma^2} \right) = \max_{\sum_{k=1}^{\infty} g_k^2 \leq E} \frac{1}{2} \sum_{k=1}^{\infty} \log \left(1 + \frac{g_k^2 \lambda_k}{\sigma^2} \right). \quad (1.8)$$

Составляя и решая уравнения Эйлера для левой части уравнения (1.8) при ограничении (1.4)

$$\Phi_1(g_k^2) = \frac{1}{2} \sum_{k=1}^{\infty} \log \left(1 + \frac{g_k^2 \lambda_k}{\sigma^2} \right) + \alpha \sum_{k=1}^{\infty} \frac{\lambda_k \sigma^2}{\lambda_k g_k^2 + \sigma^2}; \quad \frac{d\Phi_1(g_k^2)}{dg_k^2} = 0,$$

где α – неопределённый множитель Лагранжа, находим формулу для кодера $g_{k H_\varepsilon}^2$, реализующего эpsilon-энтропию источника:

$$g_{k H_\varepsilon}^2 = a^2 - \frac{\sigma^2}{\lambda_k}, \quad (1.9)$$

где a – усиление кодера.

Составляя и решая уравнения Эйлера для правой части уравнения (1.8) при ограничении (1.6)

$$\Phi_2(g_k^2) = \frac{1}{2} \sum_{k=1}^{\infty} \log \left(1 + \frac{g_k^2 \lambda_k}{\sigma^2} \right) + \alpha \sum_{k=1}^{\infty} g_k^2; \quad \frac{d\Phi_2(g_k^2)}{dg_k^2} = 0,$$

находим формулу для кодера $g_{k C_E}^2$, реализующего пропускную способность канала связи:

$$g_{k C_E}^2 = a^2 - \frac{\sigma^2}{\lambda_k}. \quad (1.10)$$

Сопоставляя (1.9) с (1.10), находим, что при соответствующем выборе усиления a обеспечивается равенство в (1.8), при этом формула для кодера совпадает с формулой для многомерного тест-канала Галлагера.

Сопряжённая триада для непрерывной системы обеспечила достижение равенства эpsilon-энтропии источника и пропускной способности канала связи, которые вычисляются для конечного числа K членов ряда, усечённого кодером g_k^2 [25], [56], [57], следующим образом:

$$H_\varepsilon = C_E = \frac{1}{2} \sum_{k=1}^K \log \Psi_k^2 = \sum_{k=1}^K \log \Psi_k; \quad \varepsilon = K \frac{\sigma^2}{a^2} + \sum_{k=K+1}^{\infty} \lambda_k.$$

Важным для практического применения в ходе дискретизации сигналов в телевидении оказалось отличие этой формулы от формулы для про-

пускной способности, входящей в уравнение связи в форме (1.8), наличие под знаком суммы логарифмов спектрального отношения сигнал/шум при отсутствии единицы, «усохшей» в ходе оптимизации системы. Хотя формула и относится к частному случаю гауссовского канала, она опирается на шумовой порог, которого нет в формуле Шеннона [25].

При этом считалось [25], что линейное кодирование, включающее разложение Карунена–Лоэва (с многократным интегрированием скалярных произведений сигнала на базисные функции на всём интервале T задания сигнала), – безблоковое и именно безблоковость «ответственна» за то, что при усечении бесконечного ряда значение пропускной способности становится несколько меньшим, чем при бесконечном ряде в формуле Шеннона (вместо $\Psi^2 + 1$ фигурирует просто отношение сигнал/шум Ψ^2). Лучшим объяснением этого отличия представляется переход от бесконечного интервала времени к конечному, аналогичный переходу от бесконечного блока к конечному, а не к полному отсутствию блоков.

Роли критерия и ограничения в непрерывной системе иллюстрируются рис. 1.8, из которого видно, что частотная характеристика фильтра Ко-

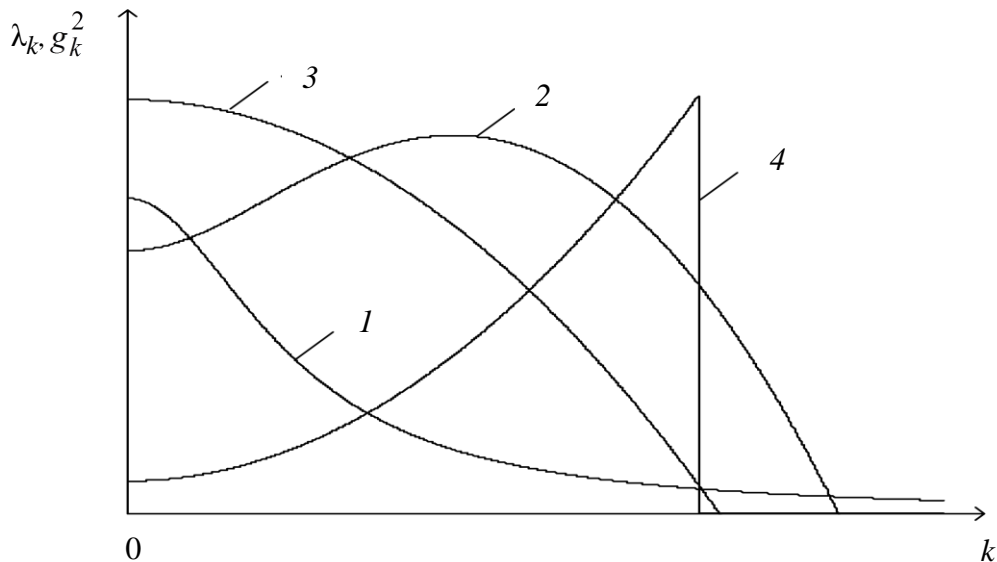


Рис. 1.8. Спектр Баттерворта 1-го порядка 1 и спектральные характеристики оптимальных кодирующих фильтров: $2 - g_{k \varepsilon, P}^2$ для минимизации среднеквадратической ошибки ε при ограничении мощности P (фильтр Костаса, несопряжённая триада); $3 - g_{k \varepsilon, E}^2 = g_{k C, E}^2$ для минимизации среднеквадратической ошибки ε и для максимизации пропускной способности C при ограничении широкополосности E кодирующего фильтра (сопряжённая триада); $4 - g_{k C, P}^2$ для максимизации пропускной способности C при ограничении мощности P (несопряжённая триада)

стаса имеет промежуточный характер между фильтрами, обеспечивающими максимум пропускной способности при ограничениях мощности и широкополосности.

Как в отношении произвольной системы связи, в отношении непрерывной системы справедливо утверждение о *существовании бесконечного множества* сопряжённых триад, обеспечивающих информационное равновесие. Это следует из допустимости в качестве функции потерь $r(x, y)$ в интегральном выражении для ошибки *любой* неубывающей функции. Конечно, такой произвол повлечёт изменение в других компонентах опорной триады и может выглядеть волюнтаристским.

Отметим, что выбор среднеквадратической функции потерь стал общепринятым не в силу его логически обоснованной исключительности для обеспечения информационного равновесия системы связи любой категории, а только в силу удобства анализа – достаточно просто находятся экстремумы искомых функционалов. Кроме того, полезность критерия минимума среднеквадратической ошибки оправдана тем, что этот критерий входит в сопряжённую триаду статистического синтеза и позволяет достичь информационного равновесия при гауссовских сигналах, мнение о которых А. Пуанкаре цитировалось на с. 56.

1.4.2. Информационное равновесие при ограничении мощности

В этом параграфе рассмотрим метод достижения информационного равновесия при смене критерия качества. Приведённое ранее решение уравнения связи достигнуто приведением опорной триады к сопряжённой сменой типа ограничения. Однако это не единственный путь. В рамках линейной системы с гауссовскими сигналом и шумом канала связи ограничению мощности в сопряжённой триаде синтеза соответствует другой критерий качества. Для его нахождения необходима опора на кодирующий фильтр, обеспечивающий максимум скорости передачи информации через канал связи (пропускную способность C_p) при ограничении мощности. Для того чтобы достичь информационного равновесия (привести опорную триаду синтеза к сопряжённой), необходимо и достаточно, чтобы выражение для энтальпии источника совпало с выражением для пропускной способности канала связи. Для этого необходимо в выражение для среднеквадратической ошибки (1.4) ввести весовую функцию φ_k :

$$\sum_{k=1}^{\infty} \frac{\lambda_k \sigma^2 \varphi_k}{\lambda_k g_k^2 + \sigma^2} \leq \varepsilon_{\varphi}. \quad (1.11)$$

Уравнение связи (1.2), (1.3) для линейного кодирования при ограничениях (1.11) и (1.5) примет вид

$$\min_{\sum_{k=1}^{\infty} \frac{\lambda_k \sigma^2 \varphi_k}{\lambda_k g_k^2 + \sigma^2} \leq \varepsilon_{\varphi}} \frac{1}{2} \sum_{k=1}^{\infty} \log \left(1 + \frac{g_k^2 \lambda_k}{\sigma^2} \right) = \max_{\sum_{k=1}^{\infty} \lambda_k g_k^2 = P} \frac{1}{2} \sum_{k=1}^{\infty} \log \left(1 + \frac{g_k^2 \lambda_k}{\sigma^2} \right). \quad (1.12)$$

Составляя и решая уравнения Эйлера для левой части уравнения (1.9) при ограничении (1.8)

$$\Phi_3(g_k^2) = \frac{1}{2} \sum_{k=1}^{\infty} \log \left(1 + \frac{g_k^2 \lambda_k}{\sigma^2} \right) + \alpha \sum_{k=1}^{\infty} \frac{\lambda_k \sigma^2 \varphi_k}{\lambda_k g_k^2 + \sigma^2}; \quad \frac{d\Phi_3(g_k^2)}{dg_k^2} = 0,$$

находим формулу для кодера $g_k^2 H_{\varepsilon\varphi}$, реализующего эпсилон-энтропию источника $H_{\varepsilon\varphi}$ (в терминологии Галлагера – тест-канал, включающий искомый фильтр и источник шума канала связи):

$$g_k^2 H_{\varepsilon\varphi} = \frac{a^2 \lambda_k \varphi_k - \sigma^2}{\lambda_k}. \quad (1.13)$$

Составляя и решая уравнения Эйлера для правой части (1.12) при ограничении (1.2):

$$\Phi_4(g_k^2) = \frac{1}{2} \sum_{k=1}^{\infty} \log \left(1 + \frac{g_k^2 \lambda_k}{\sigma^2} \right) + \alpha \sum_{k=1}^{\infty} g_k^2 \lambda_k; \quad \frac{d\Phi_4(g_k^2)}{dg_k^2} = 0,$$

находим формулу для оптимального кодирующего («выбеливающего») фильтра:

$$g_k^2 C_P = \frac{a^2 - \sigma^2}{\lambda_k} = \frac{P}{K \lambda_k}, \quad (1.14)$$

где K – конечное число членов ряда (который усекается в силу ограничения мощности), эквивалентное конечной полосе частот W_0 в формуле Шеннона.

Подстановка (1.14) в (1.7) показывает, что данный результат, полученный в рамках линейного кодирования гауссовских сигналов, ведёт к формуле Шеннона [16]:

$$C_P = \frac{1}{2} \sum_{k=1}^K \log \left(1 + \frac{P}{K\sigma^2} \right) = \frac{K}{2} \log \left(1 + \frac{P}{K\sigma^2} \right).$$

Сопоставляя (1.13) с (1.14), находим, что при соответствующем выборе весовой функции

$$\varphi_k = \begin{cases} \frac{1}{\lambda_k}, & 1 \leq k \leq K; \\ 1, & k \geq K+1 \end{cases} \quad (1.15)$$

можно достичь равенства в (1.12). Это означает, что весовая функция при среднеквадратической ошибке, обратно пропорциональная спектральной плотности мощности сигнала, модифицирует критерий качества так, что достигается равенство в (1.12):

$$H_{\varepsilon\varphi} = C_P = \frac{K}{2} \log \left(1 + \frac{P}{K\sigma^2} \right); \quad \varepsilon = K \frac{\sigma^2}{a^2} + \sum_{k=K+1}^{\infty} \lambda_k.$$

Концепция взвешенной среднеквадратической ошибки широко обсуждалась в теории связи (например [82]), физиологии [43], [83], телевидении [45], [84]. При этом и физиологи, и телевизионщики сходились во мнении, что актуален именно взвешенный средний квадрат в силу того, что зрительный анализатор человека (глаз плюс мозг) не одинаково реагирует и не одинаково «ценит» различные пространственные частоты.

В качестве аргумента выдвигалась существенная роль контуров изображения, которая побуждала к дифференцированию сигнала, – операция дифференцирования как бы сохраняет важнейшую контурную информацию и подавляет менее важную (ненужную, фоновую).

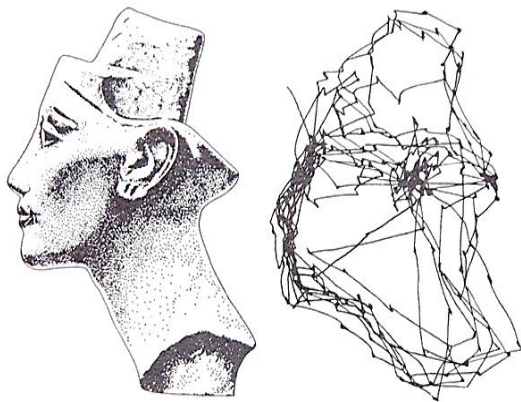


Рис. 1.9. Иллюстрация концентрации внимания зрения на контурах изображения

и подавляет менее важную (ненужную, фоновую). В качестве иллюстрации важности именно контурной информации приводился эксперимент, показывающий траекторию движения взгляда по изображению при его рассматривании (рис. 1.9 [85]).

В соответствии с этими взглядами введение весовой функции (1.15) в выражение для среднеквадратической ошибки является не просто математическим изыском, а логически

оправданной операцией: эта весовая функция для всех реальных изображений, имеющих убывающие спектральные плотности мощности, соответствует операциям дифференцирования и «выбеливания», попытки применения которых известны в аналоговой телевизионной технике [42], [84].

Таким образом, информационное равновесие в непрерывной системе при линейном кодировании возможно и при ограничении мощности на выходе кодера при условии введения в выражение для среднеквадратической ошибки (1.4) весовой функции (1.15).

1.4.3. Информационное равновесие при совместном ограничении мощности и широкополосности

Здесь рассмотрим путь достижения информационного равновесия при вариации и критерия качества, и ограничения. Ранее было рассмотрено ограничение, представляющее взвешенную сумму мощности на выходе кодера и его широкополосности [61]:

$$\beta \sum_{k=1}^{\infty} g_k^2 + \gamma \sum_{k=1}^{\infty} \lambda_k g_k^2 \leq F; \quad \beta + \gamma = 1, \quad (1.16)$$

и показано, что изменение весовых коэффициентов в этих ограничениях даёт «пучок» бесконечного количества спектральных характеристик кодеров, лежащий между кодером тест-канала Галлагера и «выбеливающим» кодером. Покажем, что такой критерий тоже может входить в сопряжённую триаду, обеспечивающую решение уравнения связи, т. е. информационное равновесие. Уравнение связи (1.3) в этом случае, так же как и в предыдущем параграфе, будет включать критерий качества в виде взвешенной среднеквадратической ошибки:

$$\min_{\sum_{k=1}^{\infty} \frac{\lambda_k \sigma^2 \varphi_k}{\lambda_k g_k^2 + \sigma^2} \leq \varepsilon_{\varphi}} \frac{1}{2} \sum_{k=1}^{\infty} \log \left(1 + \frac{g_k^2 \lambda_k}{\sigma^2} \right) = \max_{\substack{\beta \sum_{k=1}^{\infty} g_k^2 + \gamma \sum_{k=1}^{\infty} \lambda_k g_k^2 \leq F; \\ \beta + \gamma = 1}} \frac{1}{2} \sum_{k=1}^{\infty} \log \left(1 + \frac{g_k^2 \lambda_k}{\sigma^2} \right). \quad (1.17)$$

Составляя и решая уравнения Эйлера для левой части (1.16) при ограничении (1.11)

$$\Phi_5(g_k^2) = \frac{1}{2} \sum_{k=1}^{\infty} \log \left(1 + \frac{g_k^2 \lambda_k}{\sigma^2} \right) + \alpha \sum_{k=1}^{\infty} \frac{\lambda_k \sigma^2 \varphi_k}{\lambda_k g_k^2 + \sigma^2}; \quad \frac{d\Phi_5(g_k^2)}{dg_k^2} = 0,$$

находим формулу для кодера $g_k^2 H_{\varepsilon\varphi}$, реализующего эpsilon-энтропию источника $H_{\varepsilon\varphi}$ (в терминологии Галлагера – «тест-канал», включающий искомый фильтр и источник шума канала связи):

$$g_k^2 H_{\varepsilon\varphi} = \frac{a^2 \lambda_k \varphi_k - \sigma^2}{\lambda_k}. \quad (1.18)$$

Составляя и решая уравнения Эйлера для правой части (1.14) при ограничении (1.13):

$$\Phi_6(g_k^2) = \frac{1}{2} \sum_{k=1}^{\infty} \log \left(1 + \frac{g_k^2 \lambda_k}{\sigma^2} \right) + \alpha \left(\beta \sum_{k=1}^{\infty} g_k^2 + \gamma \sum_{k=1}^{\infty} \lambda_k g_k^2 \right); \quad \frac{d\Phi_6(g_k^2)}{dg_k^2} = 0,$$

находим формулу для оптимального кодирующего фильтра:

$$g_k^2 C_F = \frac{a^2}{\beta + \gamma \lambda_k} - \frac{\sigma^2}{\lambda_k}. \quad (1.19)$$

Сопоставляя формулу (1.15) с формулой (1.16), находим, что при соответствующем выборе весовой функции

$$\varphi_k = \begin{cases} \frac{1}{(\beta + \gamma \lambda_k)}, & 1 \leq k \leq K; \\ 1, & k \geq K + 1 \end{cases} \quad (1.20)$$

можно достичь равенства в (1.17). Это означает, что весовая функция при среднеквадратической ошибке, обратно пропорциональная спектральной плотности мощности сигнала, модифицирует критерий качества так, что достигается равенство в (1.17):

$$H_{\varepsilon\varphi} = C_F = \frac{1}{2} \sum_{k=1}^K \log \frac{a^2 \lambda_k}{(\beta + \gamma \lambda_k) \sigma^2}; \quad \varepsilon = K \frac{\sigma^2}{a^2} + \sum_{k=K+1}^{\infty} \lambda_k.$$

Весовая функция (1.17), содержащая «вектор концепции» кодирования $\{\beta, \gamma\}$, характеризует «пучок» бесконечного количества спектральных характеристик кодеров, лежащий между кодером тест-канала Галлагера и «выбеливающим» кодером. Спектральные характеристики оптимальных кодеров, рассчитанные для спектра Баттерворта 1-го порядка (сигнал, обладающий максимальной энтропией), показаны на рис. 1.10.

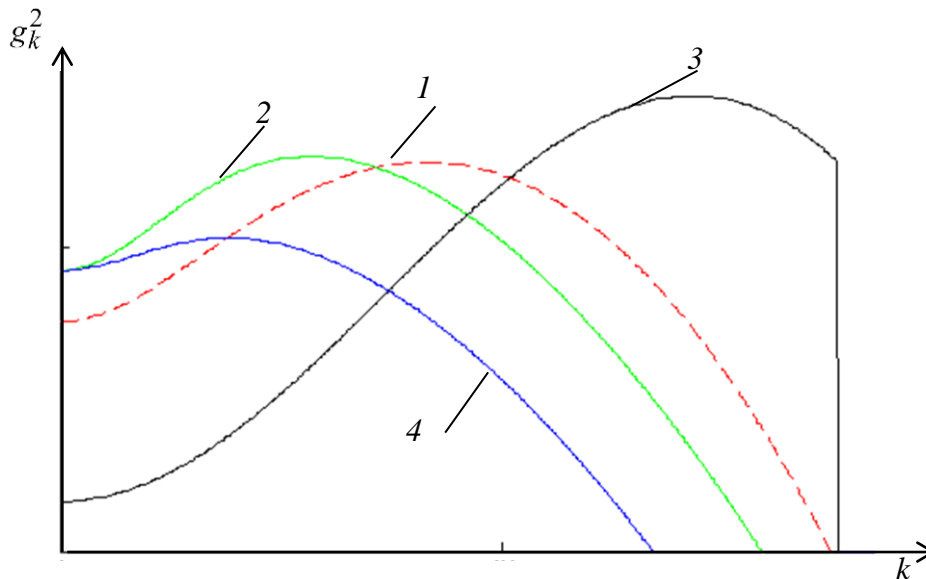


Рис. 1.10. Спектральные характеристики оптимальных фильтров: 1 – фильтра Костаса; 2 – при гибридном ограничении и критерии взвешенного минимума среднеквадратической ошибки, $\beta = \gamma = 0,5$; 3 – то же при $\gamma = 0,98$; 4 – при $\gamma = 0,3$

Показательно, что найденное семейство оптимальных фильтров, обеспечивающих информационное равновесие, имеет при близких значениях параметров ограничения β и γ форму спектральной характеристики, похожую на спектральную характеристику фильтра Костаса [53] (рис. 1.10), который, несмотря на достижение условного (при ограничении мощности) минимума среднеквадратической ошибки, такого равновесия принципиально не обеспечивает.

Это обусловлено тем, что ограничение мощности и традиционная среднеквадратическая ошибка без весовой функции не порождают сопряжённых триад синтеза, и проявляется в том, что в выражении для фильтра Костаса в отличие от формулы (1.18) в знаменатель первого члена входит квадратный корень из спектральной плотности мощности передаваемого сигнала.

1.4.4. Информационное равновесие при ограничении сложности линейного кодера

Рассмотрим решение уравнения связи при ограничении сложности кодера, при этом в качестве ограничения сложности будем рассматривать только объём памяти, необходимый для хранения спектральных коэффициентов кодирующего фильтра:

$$\sum_{k=1}^{\infty} \log \left(\frac{g_k}{\delta} \right) \leq W_k. \quad (1.21)$$

Уравнение связи для линейного кодирования при ограничениях (1.11) и (1.21) примет вид

$$\min_{\sum_{k=1}^{\infty} \frac{\lambda_k \sigma^2 \varphi_k}{\lambda_k g_k^2 + \sigma^2} \leq \varepsilon_{\varphi}} \frac{1}{2} \sum_{k=1}^{\infty} \log \left(1 + \frac{g_k^2 \lambda_k}{\sigma^2} \right) = \max_{\sum_{k=1}^{\infty} \log \left(\frac{g_k^2}{\delta} \right) \leq W_K} \frac{1}{2} \sum_{k=1}^{\infty} \log \left(1 + \frac{g_k^2 \lambda_k}{\sigma^2} \right). \quad (1.22)$$

Составляя и решая уравнения Эйлера для левой части уравнения (1.22) при ограничении (1.11)

$$\Phi_7(g_k^2) = \frac{1}{2} \sum_{k=1}^{\infty} \log \left(1 + \frac{g_k^2 \lambda_k}{\sigma^2} \right) + \alpha \sum_{k=1}^{\infty} \frac{\lambda_k \sigma^2 \varphi_k}{\lambda_k g_k^2 + \sigma^2}; \quad \frac{d\Phi_7(g_k^2)}{dg_k^2} = 0,$$

находим формулу для кодера $g_k^2_{H_{\text{эф}}}$, реализующего эпсилон-энтропию источника $H_{\text{эф}}$ (в терминологии Галлагера – «тест-канал», включающий искомый фильтр и источник шума канала связи):

$$g_k^2_{H_{\text{эф}}} = \frac{a^2 \lambda_k \varphi_k - \sigma^2}{\lambda_k}. \quad (1.23)$$

Составляя и решая уравнения Эйлера для правой части уравнения (1.22) при ограничении (1.21)

$$\Phi_8(g_k^2) = \frac{1}{2} \sum_{k=1}^{\infty} \log \left(1 + \frac{g_k^2 \lambda_k}{\sigma^2} \right) + \alpha \sum_{k=1}^{\infty} \log \left(\frac{g_k^2}{\delta} \right); \quad \frac{d\Phi_8(g_k^2)}{dg_k^2} = 0,$$

находим формулу для оптимального кодирующего («выбеливающего») фильтра:

$$g_k^2_{CW} = \frac{a^2}{\lambda_k}. \quad (1.24)$$

Отсюда видно, что формула для спектральных коэффициентов кодера (1.24) совпадает с формулой для кодера Шеннона (1.14), обеспечивающего решение уравнения связи при средневзвешенном критерии качества.

Сопоставляя (1.23) с (1.24), находим, что при соответствующем выборе весовой функции

$$\varphi_k = \begin{cases} \frac{a^2 + \sigma^2}{a^2 \lambda_k}, & 1 \leq k \leq K; \\ 1, & k \geq K + 1 \end{cases} \quad (1.25)$$

можно достичь равенства в (1.22). Это означает, что весовая функция при среднеквадратической ошибке, обратно пропорциональная спектральной плотности мощности сигнала, модифицирует критерий качества так, что достигается равенство в (1.22):

$$H_{\text{эф}} = C_W = \frac{K}{2} \log \left(1 + \frac{a^2}{\sigma^2} \right); \quad \varepsilon = K \frac{\sigma^2}{a^2} + \sum_{k=K+1}^{\infty} \lambda_k.$$

Таким образом, мы доказали гипотезу [59] о множественности решений уравнения связи и показали, что даже в рамках линейного кодирования существует бесконечное множество оптимальных систем связи.

Выводы по гл. 1. Итак, отмеченный конфликт идей об идеальной связи и оптимальном накоплении разрешается использованием уравнения связи, которое имеет множество решений, обеспечивающих равновесие между скоростью создания информации источником и скоростью её передачи через канал. Мы показали, что существует бесконечное множество оптимальных кодирующих фильтров, среди которых имеется *бесконечное подмножество* кодеров, обеспечивающих информационное равновесие, при этом фильтр Винера и фильтр Костаса, хотя и относятся к оптимальным кодирующим (фильтр Костаса) и декодирующим (фильтр Винера) фильтрам, но не обеспечивают информационного равновесия (рис. 1.11).

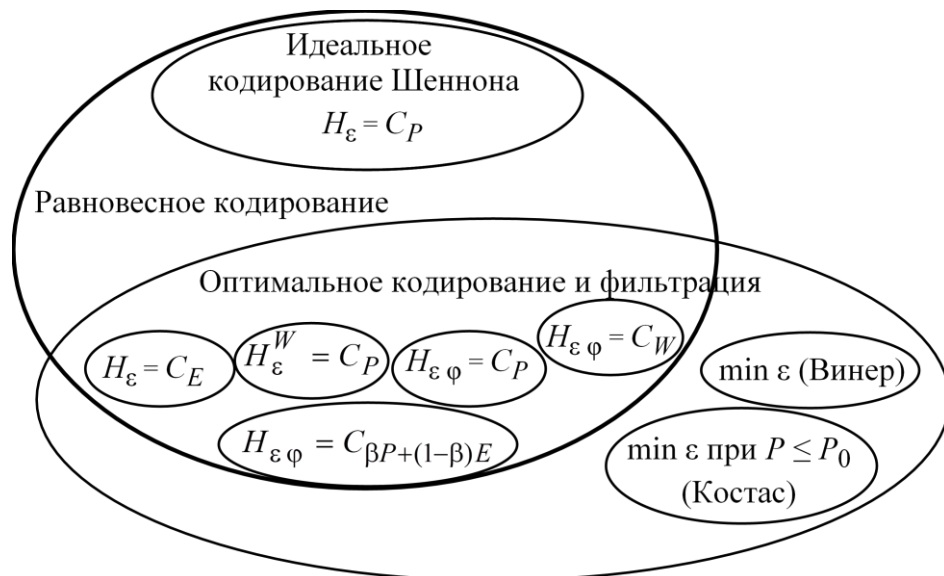


Рис. 1.11. Взаимосвязь идеального кодирования Шеннона, равновесного кодирования и оптимального кодирования и фильтрации

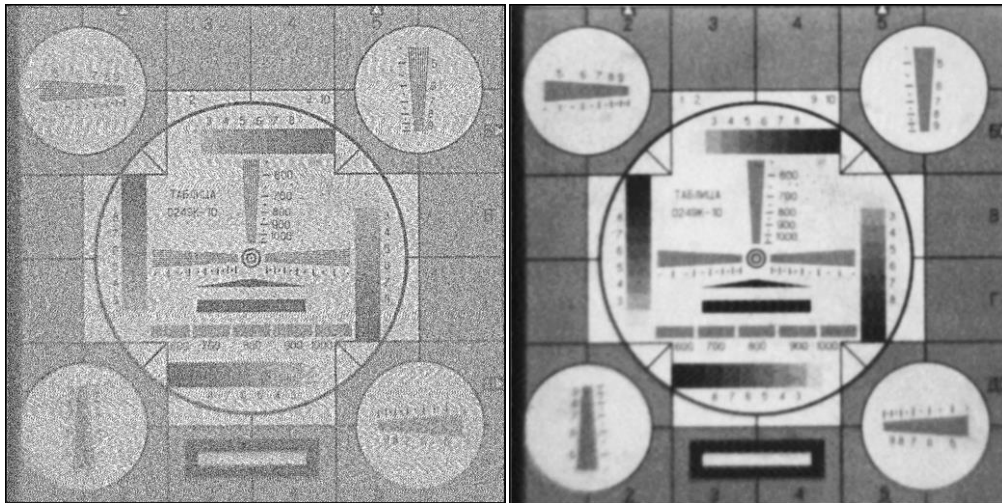
Казалось бы – зачем проектировщику это множество оптимальных кодеров? Ведь очевидно, что если меняются условия или критерии качества, то меняется и оптимальное решение. И стоит ли говорить при этом о нали-

ции разных оптимальных решений? Можно также спросить – в чём преимущества найденных фильтров? Всё дело в том, что проектировщик канала связи, для которого недоступно управление источником, не сможет воспользоваться идеей о множестве оптимальных систем, а создатель телекамеры, управляющий на стадии проектирования и источником, и каналом, – сможет. И ему крайне важно знать о существовании большого количества оптимальных систем, чтобы не фетишизировать ни идеальное блочное кодирование для *передачи* информации, ни оптимальный для *приёма* сигналов фильтр Винера. Огромная роль инженерной интуиции подчёркивается использованием понятия «взвешенной среднеквадратической ошибки», позволившей показать наличие информационного равновесия в линейной системе. Можно сказать, что преимущества найденных фильтров не только в достижении конкретных экстремумов качества, но и в расширении кругозора проектировщиков, которые получают широкий набор оптимальных финитных фильтров (с ограниченной спектральной характеристикой), служащих возможным эталоном при переходе от непрерывных сигналов к дискретным в ходе рождения информации.

Аналогично выводу Шеннона о важности учёта статистических свойств сигнала и шума, процитированному на с. 50, можно сказать, что теперь, благодаря применению уравнения связи и нахождению бесконечного множества его решений, в проблеме достижения информационного равновесия между скоростью создания информации источником и пропускной способностью канала связи значительно глубже понимается не только важность априорной информации, отмеченная Шенноном, но и *важность сопряжения априорной информации, критерия качества и ограничений*. Крайне важно, что *разные критерии при одной и той же априорной информации приводят к разным кодирующим фильтрам, обеспечивающим не просто локальную оптимальность, но и информационное равновесие*. Так, если зафиксировать априорную информацию в виде гауссовских процессов сигнала и шума, а также ограничение мощности на выходе кодера, то очевидно, что ошибка, которую обеспечивает фильтр Костаса, будет меньше, чем выбеливающий фильтр, обеспечивающий максимум взаимной информации, т. е. пропускную способность. Поэтому широко распространённое мнение о том, что достижение предельной пропускной способности автоматически обеспечивает достижение минимума ошибки передачи, следует признать заблуждением.

Введение в арсенал статистического синтеза уравнения связи позволило увидеть, что информационное равновесие (равенство скорости создания информации источником и скорости её передачи через канал связи) возможно не только в дискретной системе, детально исследованной Шенноном. Первое решение уравнения связи, выходящее за пределы рассмотренного Шенноном и относящееся к непрерывной системе, включило тест-канал Галлагера, но потребовало *смены ограничения*.

Этот результат объясняет, почему многие (например [42]) «нутром чуяли», что с критерием среднеквадратической ошибки что-то не так, что она применима только в гауссовском случае, и даже писали, что он в телевидении не применим потому, что и система нелинейна, и шум не белый, и модель линейного искажения не всегда можно правильно учесть – так, авторы книги [42] забыли включить в модель системы линейный искажающий фильтр, из-за чего «потеряли» компенсацию частотных искажений, известную в телевидении как апертурная коррекция. На самом деле критерий минимума среднеквадратической ошибки, конечно, можно применять в любой задаче кодирования, но в *сопряжённую триаду* он входит не всегда. Важность корректной модели системы связи, показанная на примере ошибочных выводов в книге [42], видна также из следующего примера, связанного с визуальным анализом изображений. Конечно, адекватный учёт свойств зрительного анализатора не прост, и даже считается, что «*построение физиологических моделей зрительного анализатора и на их основе математических критериев качества изображения является очень сложной задачей, которая до сих пор не решена*» [86]. При этом считается установленным, «*что субъективные оценки качества изображения склоняются в сторону более зашумлённых, но чётких изображений, чем обесшумленных, но размытых (смазанных)*» [87]. Пример такой пары изображений показан на рис. 1.12. Тезис о более высоком качестве изображения можно считать заблуждением, являющимся следствием некорректности эксперимента: если при анализе *одного* кадра разрешающая способность (традиционная мера качества изображения) для широкой и правильно выбранной узкой полосы частот *одинакова*, то при анализе *серии* кадров наблюдатель в широкополосном сигнале может разглядеть детали, безвозвратно потерянные при сокращении полосы частот. Корректный эксперимент показывает, что предпочтение наблюдателями зашумлённого изображения с широкой полосой частот связано с возможностями дополнительного **накопления** в зрительном анализаторе – указанное разглядывание серии кадров обеспечивает усреднение шума и, как следствие, понижение порога обнаруживаемых пространственных частот.



а

б

Рис. 1.12. Изображения при низкой освещённости (малом отношении сигнал/шум) с одинаковыми разрешением и размахом сигнала:
 а – при широкой полосе частот; б – при сокращённой полосе частот

Хотя авторы уравнения связи и выдвинули гипотезу о существовании множества решений уравнения, второе решение, выходящее за результаты Шеннона и относящееся к смешанной системе передачи непрерывных процессов, было получено авторами [60] существенно позднее.

Для решения второй задачи потребовалось разубить такой же «гордиев узел» расхождения эpsilon-энтропии и пропускной способности, с которым невероятными усилиями надеялись справиться теоретики (Р. Галлагер, А. Витерби и др.). Для смешанной системы нами определена сопряжённая триада, в которую входят не только сигналы с финитной плотностью вероятностей, но и специально предложенная А. Н. Колмогоровым эpsilon-энтропия с мерой максимального (а не среднеквадратического) различия сигналов. Поразительно, что обе эти меры даны им на одной и той же странице, но одну из них, среднеквадратическую (удобную при аналитических расчётах и входящую в сопряжённую триаду для непрерывных систем, но *не входящую в неё для смешанных систем*), все используют, а другую (зная о ней!) – не желают использовать.

Результаты, изложенные в [62], хотя и относятся только к одному из крупных классов систем (с непрерывным источником информации и непрерывным каналом связи), но содержат важный *неожиданный результат*. Даже в частном случае линейного кодирования гауссовских сигналов при гауссовском шуме в канале связи существует бесконечное множество сопряжённых пар ограничений и критериев качества (критериев соответствия исходного и принятого сигналов).

Конечно, можно эти критерии считать малозначимыми для практики – авторы при обсуждении своих результатов на конференциях сталкивались с утверждениями, что ограничение мощности, использованное Шенноном, физично, а что даёт практике ограничение широкополосности – не ясно, а уж гибрид из этих ограничений и вовсе сомнителен. В ответ на эту критику следует дать следующие соображения.

Во-первых, цитированное предупреждение К. Шеннона, что ограничение средней мощности есть лишь *простейший* вид ограничения. Да, конечно, использованное множество ограничений не простейшее, но мощность на выходе кодера «некоторым способом» ограничена.

Во-вторых, в практике проектировщик сам, как правило, интуитивно выбирает опорную триаду. Конечно, проектировщики стремятся к выполнению указания Шеннона, что сигнал в канале связи должен приближаться к белому гауссовскому шуму, но выбор критериев и ограничений остаётся плохо формализованным. Не случайно в применении к непрерывным системам связи появлялись термины типа «предыскажение сигнала» [84]. При этом важнейшим видом «предыскажений» является применявшееся на практике и ориентированное на уменьшение мощности передаваемого сигнала его дифференцирование (трактующееся как «выбеливание» спектра). Однако то, что массовое использование практиками критерия среднеквадратической ошибки при этом не обеспечивает информационного равновесия, оставляло лишь смутные сомнения в окончательной правильности используемого метода кодирования. Поэтому полученный результат о наличии решения уравнения связи (информационного равновесия) в непрерывной линейной системе при ограничении мощности сигнала на выходе кодера (использованного Шенноном), но при введении спектральной весовой функции в выражение для среднеквадратической ошибки, т. е. использование метода «взвешенного среднего квадрата», показывает, что практики изобрели линейный метод «выбеливания», ещё не зная, что он имеет теоретическое обоснование. Таким образом, полученный результат даёт практике обоснование интуитивно развитых ранее методов. В целом приведённые в гл. 1 результаты расширяют понимание важности *сопряжения* априорной информации о сигнале и помехах, критерия качества передачи и совокупности ограничений в каждой конкретной системе. Указанное понимание опирается на следующие выводы:

1. Введённое понятие *сопряжённой триады* – априорной информации, критерия качества и совокупности ограничений – формализует до-

стижение равновесия в уравнении связи, т. е. равенства скорости создания информации источником и пропускной способности канала связи.

2. Для непрерывных систем связи:

- при ограничении мощности сигнала на выходе линейного кодера (линейного аналога кодера Шеннона) информационное равновесие достигается при использовании взвешенной среднеквадратической ошибки, причём не только спектральная характеристика кодирующего фильтра, но и весовая функция в критерии качества обратны спектральной характеристике кодируемого сигнала, который в результате кодирования «выбеливается»;

- существует бесконечное множество сопряжённых триад, обеспечивающих информационное равновесие, которые включают линейную комбинацию ограничений мощности на выходе кодера и его широкополосности и критерий качества в виде взвешенной среднеквадратической ошибки;

- оптимальный линейный кодирующий фильтр, обеспечивающий равенство скорости создания информации и пропускной способности при ограничении сложности, совпадает с оптимальным линейным кодирующим фильтром для ограничения мощности (линейным аналогом кодера Шеннона) при средне взвешенном критерии качества.

3. Для смешанной системы с непрерывным источником и дискретным каналом существуют решения уравнения связи, для достижения которого в опорной триаде смешанной системы связи следует использовать введенную А. Н. Колмогоровым *эпсилон-энтропию с ограничением максимума отклонения* принимаемого сигнала от передаваемого.

4. Дискретный, непрерывный и смешанный каналы связи могут обеспечивать равенство скорости создания информации источником и скорости её передачи через канал, т. е. могут быть *строго оптимальными*, что оставляет проектировщикам свободу выбора типа системы.

5. Существование множества решений уравнения связи, т. е. бесконечного множества строго оптимальных систем, усиливает значение инженерной интуиции при выборе опорной триады статистического синтеза: априорной информации, критерия качества и совокупности ограничений.

Глава 2. ПРИНЦИП ДОМИНАНТНОЙ ИНФОРМАЦИИ

В этой главе мы рассмотрим проблемы, порождённые конфликтом следующих идей. Во-первых, это конфликт идеи Котельникова–Шеннона о том, что спектр сигнала всегда каким-либо образом ограничен, и идеи о том, что непрерывный случайный сигнал имеет бесконечный спектр и содержит бесконечное количество информации.

Во-вторых, это конфликт идеи о том, что понятие информации ассоциируется с понятием пользы, и идеи о необходимости достижения избирательности системы связи, которая должна различать полезные и мешающие сигналы.

В-третьих, это конфликт принципа непрерывности, в явном виде введённого в науку Лейбницем, и принципа дискретного прироста знания, сформулированного Гейзенбергом.

В-четвёртых, это конфликт идеи Добрушина и Цыбакова [89] об оптимальности системы, передающей оптимальную оценку зашумлённого сигнала и противоречия, состоящего в конечном количестве полезной информации при бесконечном количестве передаваемой.

Теория Шеннона посвящена кодированию неискажённых сигналов, и, пользуясь моделями сигналов с бесконечным спектром, произвольно этот спектр усекая, идею о требуемой конечной точности он излагал просто: *«К счастью, нам не нужно передавать непрерывные сообщения в точности»*. Мы эту мысль трактуем так: к несчастью, нам любые сообщения доступны только в искажённом, зашумлённом виде – наблюдения световых сигналов всегда сопровождаются большим количеством искажений, как минимум – фотонным шумом. Однако шенноновский подход привёл к усечению спектра без заботы о потерянной информации и сведению к нулю ошибок в канале связи. Корректный подход к построению системы связи должен привести к обоснованному распределению ошибок, возникающих из-за искажений доступного сигнала и из-за шумов канала связи. Намёк на необходимость решения таких задач известен [88, т. 2], и даже известно решение задачи о построении кодера зашумлённого сигнала, включающего предварительную оптимальную оценку полезного сигнала [89], однако структура дальнейшего кодера оставалась в рамках теории блокового кодирования Шеннона. Не менее странна ситуация, когда проблема избирательности в радиотехнике решается и утверждается государ-

ственными стандартами без опоры на теорию информации. Хотя, казалось бы, именно наличие на любой антенне огромного количества сигналов различных источников (в том числе и имеющих смысл, и помех) является достаточным побудительным мотивом для построения теории, опирающейся на понятие доминантного сигнала.

Как отмечалось во введении, толчком к пониманию информации и предпосылкой разработки теории кодирования искажённых сигналов продуктивной оказалось учение о доминанте А. А. Ухтомского [24]. Это учение обосновывает разделение множества всех воздействий на два класса – доминанту и все остальные фоновые воздействия. При этом доминантный сигнал организмом усиливается, а все остальные тормозятся. Подсказка А. Н. Колмогорова, цитированная на с. 14, о том, что характеристика информации не должна ограничиваться её количеством, и тезис Хартли–Шеннона о бессмысленной информации, которую так же трудно передавать, как и осмысленную, поэтому можно не учитывать смысл, следует отнести к тактическим ходам развития теории. К. Шеннон, кстати, удивлялся: *«Почему-то люди думают, что она <разработанная им математическая теория связи> может что-то сказать им о смысле, но она не может и не предназначалась для этого»* [9]. Конечно, допущение о достаточном приближении бессмысленной информации к осмысленной возможно, если смотреть на информацию с точки зрения канала связи, а не всей системы связи. Хотя смысл действительно плохо поддаётся формализации, даже в рамках теории Шеннона с её функционалом для количества взаимной информации, можно построить важный её раздел для кодирования искажённых сигналов, взяв для опоры понятие о качестве информации. Попытки вывода теории Шеннона из рамок теории количества информации для учёта её смысла предпринимались неоднократно. Согласно концепции кибернетики Н. Винера информация является лишь первым звеном в цепи «информация – управление – поведение». Поэтому в первую очередь предпринимались попытки учесть роль информации в переходе к принятию **решения** и влияние информации на качество решений [90] и достижимость поставленной **цели** [91], [37]. Так, Р. Л. Стратонович ввёл понятие ценности информации $V(I)$, которая *«определяется как максимально возможное уменьшение средних штрафов, которого можно добиться, получив данное количество информации, касающейся принадлежности величины x некоторой области из ограниченного числа $M(I)$ областей ($M(I)$ –*

максимальное целое число, содержащееся в e^I)» [90]. Например, ценность информации представляется как разность максимального риска R_0 (в отсутствие получаемой информации) и риска как функции полученной информации: $V(I) = R_0 - R(I)$. Для обозначений, принятых в гл. 1, эта идея, выдвинутая для произвольных распределений, для гауссовского канала, для которого риск определяется среднеквадратической ошибкой ε , максимум которой есть дисперсия λ передаваемого сигнала, приводит к записи $\varepsilon = \lambda e^{-2I}$. Хотя автор «придерживается другой интерпретации и другой терминологии», его концепция, по сути, совпадает с концепцией эpsilon-энтропии; при этом меняются местами функция и аргумент и делается акцент на качестве решения, но не затрагиваются ни причины введения допустимой ошибки, ни цель системы.

Учёт цели системы связывался с определением ценности информации путём введения вероятностей достижения цели до и после получения информации [91]. Такой подход требует расширения модели системы при помощи блоков, отражающих управление и поведение. В рамках же проблематики теории связи отмечается, что «если при получении информации цель достигается наверняка, то ценность этой информации максимальна, т. е. совпадает с максимальным количеством информации в данном множестве... при этом ценность информации можно понимать как количество ценной информации» [37].

Указанные подходы Р. Л. Стратоновича [90] и А. А. Харкевича [91] исходили из отождествления понятий «информация» и «польза» и содержали в себе лишь ответ на вопрос «Сколько пользы от получения этой информации?», не вникая во внутреннюю структуру самой информации, которую при наличии неконтролируемых искажений до кодирования необходимо учитывать.

На этом пути был выдвинут принцип доминантной информации [14], исходящий из того, что «интересующая нас информация» (слова А. Н. Колмогорова) может, согласно учению А. А. Ухтомского, быть названа доминантной в противовес всегда входящим в доступную информацию другим компонентам – **фоновой и шумовой информации**, мешающим выделению доминанты. Согласно принципу доминантной информации саму информацию следует определить как **триединство цели, качества и количества** [14]. Если количество вполне обоснованно можно вычислять по функционалу Шеннона, то качество информации – понятие достаточно сложное,

оно тесно связано с целью системы. Например, Н. Винер разработал теорию линейной фильтрации и ввёл формулу для *количества* информации в смеси полезного сигнала и шума в ходе работы над системой управления зенитным огнём, для которой *цель* – сбить самолёт противника, а *качество информации* – промах, который проектировщику следует минимизировать. Мы будем в рамках проблематики теории связи аналогично рассмотренному примеру говорить о синтаксической информации и понимать качество информации как ***насыщенность доступной информации доминантой*** [14]. При этом, как указано во введении, эту насыщенность будем понимать не как данность, а как цель проектирования, достигаемую в процессе ***накопления качества информации***.

Понятие доминантной информации только с виду кажется тривиальным, так как понятия цели, пользы или ценности информации обсуждались давно, хотя и отмечалось, что ценность информации субъективна и не связана с объективной оценкой количества информации [37]. К понятию ценности информации примыкает проблема семантической информации, встречаются попытки переименовать семантическую (доминантную) информацию в просто информацию, а доступную информацию – в «данные» [36], что не даёт способа разделения доступной информации на доминантную и ложную. В телевидении эти вопросы обсуждались полвека назад, например С. Б. Гуревич утверждал [44]: *«Наряду с характеристикой информации, определяющей статистические свойства сообщения, должна быть введена ещё характеристика, определяющая **действенность, ценность сообщения**. С учётом этой характеристики **количество информации определится приобретённой ценностью сведений и достигнутым с их помощью результатом**. Назовём такую информацию, использованную для достижения поставленной в задаче цели, **полезной или целевой информацией**. Учёт статистики объекта приводит к уменьшению количества информации, передаваемого через систему. Учёт действенности, ценности сведений, поступающих от объекта, приводит к дальнейшему увеличению избыточности»*. Однако эта простота понятия доминантной информации – кажущаяся, в полной мере его значение проявляется при формализации и сопоставлении доминантной и ложной информации, состоящей из фоновой информации и шумовой информации.

Ранее, до введения принципа доминантной информации [14], учёт полезности информации сводился к надеждам на её большое сжатие за счёт

декорреляции телевизионных сигналов [43], [44], которые не оправдались, что резюмировалось так [42]: «*В короткой истории применения теории информации в телевидении больше разочарований, чем достижений*». Введение же *принципа доминантной информации* благодаря формализации, как будет показано далее, позволило выявить закон взаимнообмена потери доминантной информации и шумовой информации, ввести *определенные качества информации* и найти *оптимальные фильтры*, обеспечивающие максимум качества информации (как безусловный, так и условный максимумы при выполнении требования информационного равновесия). Более того, принцип доминантной информации с указанными его следствиями устранил важную «занозу» классической теории информации, каковой долго являлась странная для практиков популярная теорема Винера о том, что «*кодирующая система может уменьшить, а в лучшем случае сохранить имеющуюся информацию о зашумлённом сигнале*». Это, конечно, не вело к отказу от обработки сигналов, но часто вело к отказу от использования теории информации. Однако лишь принцип доминантной информации указывает, что в любой системе передачи информации, создаваемой ради передачи доминантной информации, неизбежно существуют другие виды информации, в первую очередь фоновая и шумовая, и сверхзадача любой кодирующей системы, по определению, данному А. А. Ухтомским для биологического субъекта, – не только *усиление доминантного сигнала*, но и *торможение фоновых сигналов*.

Именно эти две операции усиления и торможения и осуществляет любая система кодирования, и лишь принцип доминантной информации даёт им теоретическое обоснование. Эта идея давно известна в теории связи. Так, К. Черри писал [51]: «*Задача приёмника — **извлечь** из принятого сигнала всю **возможную информацию** о переданном сигнале (а следовательно, и о представленном этим сигналом сообщении) и **отбросить** имеющуюся «**фиктивную информацию**» об источнике шума*». Из этой цитаты видно, что выделение доминантной информации из доступной, насыщение доступной информации доминантой – это центральная проблема теории и практики связи, но до недавнего времени она была плохо формализована. В некотором смысле с этой проблемой боролись с помощью фильтрации, но вопрос, что делать с бесконечным количеством информации, которое нужно передать, чтобы доставить получателю конечное

количество информации, содержащееся в формируемой фильтром оценке сигнала, – продолжал «висеть в воздухе».

Определённым приложением к теории доминанты Ухтомского является гештальт-психология [92]. Гештальт (нем. *Gestalt* – форма, образ, структура) – это пространственно-наглядная форма воспринимаемых предметов, чьи существенные свойства нельзя понять при суммировании свойств их частей. При этом основные свойства визуального восприятия: константы, фигура, фон – вступают в отношения между собой и являют новое свойство. Это и есть *гештальт*, или *качество формы*, существование которого обусловлено существованием *качества информации*. Целостность восприятия и его упорядоченность достигаются благодаря принципу *упрощения восприятия*. Важным элементом теории гештальтов, роднящим её с теорией кодирования искажённых сигналов, является формирование восприятия *разделением зрительных ощущений на объект и фон*.

Таким образом, принцип доминантной информации приводит к определению: *информация – атрибут духа и материи, посредник между ними, имеющий количественную меру и обладающий качеством*. Как в человеке неразрывно слиты дух, душа и плоть, так и *в информации неразрывно слиты цель, доминанта и фон*. Фон потенциально может стать полезным (смена доминанты при переключении каналов телевизора, перескок внимания в классическом сюжете – два лица или ваза?), но в каждый конкретный момент он является помехой, приближаясь к шуму, который вреден всегда.

В этом определении участвует понятие качества информации, про которое, воспользовавшись терминологией гештальт-психологии, можно сказать, что качество доступной информации отражает её принадлежность *в той или иной мере* к доминанте или фону. При этом доминантная информация характеризуется наибольшим качеством, а фоновая – минимальным качеством. Как в гештальт-психологии говорят «хороший образ», в теории информации можно говорить «хорошая информация», понимая под этим *насыщенность информации доминантой*. Это же утверждает А. А. Ухтомский [24]: «*Всякий интегральный образ, которым мы располагаем, является остаточным продуктом пережитой нами доминанты*».

Поскольку информация является *посредником* между духом и материей, постольку она определяет *взаимосвязь* материи и образа, создаваемого наблюдателем. Поэтому понятие *взаимной информации* и стало ключевым в теории связи. основополагающая работа Шеннона [15], [16] содержит

несколько аксиом теории информации, относящихся к свойствам энтропии. Позднее произошла переоценка значения информационных величин и, подводя итоги первого этапа развития теории передачи информации, А. Н. Колмогоров подчёркивал ключевое значение не энтропии, а *взаимной информации* (см. цитату на с. 27). Так как мы пользуемся понятиями *доступной, доминантной, фоновой и шумовой информации*, то необходимо рассмотреть их свойства, в первую очередь те, которые позволят формализовать принципы и методы кодирования зашумлённых сигналов, распространив определение «взаимности» информации на эти вводимые величины.

2.1. Проблема кодирования зашумлённых сигналов

Является ли кодирование зашумлённых сигналов проблемой, видно из конфликта идей, первая из которых характеризуется введением понятия коэффициента шума, который во многих приложениях удовлетворял проектировщиков. Вторая идея – коэффициент шума не приводит ни к каким экстремумам качества, его наилучшее значение – единица, т. е. с позиций борьбы со «вторым» шумом, добавляемым к входному «первому» шуму, коэффициент шума толкает проектировщика к достижению температуры усилителей, равной абсолютному нулю. В особо чувствительных системах (и телевизионных, и, например, в радиоастрономии) охлаждение применяется, но существует ли оптимальное охлаждение (оптимальный коэффициент шума) с точки зрения статистической теории связи – вопрос был открытым.

Теория кодирования зашумлённых сигналов востребована многими отраслями информационной техники, в первую очередь – телевидением, для которого неизбежно кодирование световых сигналов, искажённых фотонным шумом, для передачи их по каналу с шумом. Создавая математическую теорию связи, Шеннон сразу написал [16] (см. цитату на с. 36), что можно грубо разделить системы связи на дискретные, непрерывные и смешанные. Для компактного изложения далее рассмотрим основные положения теории лишь линейного кодирования случайных процессов для непрерывных каналов связи. Передача случайных величин, относящаяся к кодированию без задержки, является частным случаем передачи случайных процессов. Передача зашумлённых сигналов в смешанных и дискретных системах требует самостоятельного рассмотрения, хотя общие принципы кодирования едины для всех указанных категорий систем. Важнейшим

объектом рассмотрения в теории линейного кодирования выступает система связи, включающая источник гауссовского процесса и два источника гауссовского шума – искажающий входной сигнал и воздействующий на кодированный сигнал в канале связи. Эта модель не только позволяет получить простые формулы для идеальной, по Шеннону, непрерывной системы связи, но и достаточно близка к практике обработки сигналов. Значение этой модели для телевидения определяется в первую очередь тем, что она отражает обязательное наличие по крайней мере двух источников шума: фотонного и считывания сигнала изображения. Кроме того, эта модель может быть распространена и на многокаскадные системы линейного кодирования.

Кроме отсутствия учёта смысла в классической теории кодирования имеется ещё одна неприятность. Реальные сигналы, всегда известные на конечном интервале времени, принципиально имеют неограниченный спектр. Однако на начальном этапе развития теории передачи непрерывных сигналов предполагалось, что «спектр сигнала *каким-либо образом ограничен*». В. А. Котельников и К. Шеннон прямо начинают построение теории связи словами «если функция времени ограничена по полосе частот». Весьма показательным, что акад. В. А. Котельников, подводя итоги своей долгой научной деятельности, выделял на первое место именно свою теорему 1933 г. о представлении функции с ограниченным спектром дискретной последовательностью отсчётов, обеспечивающей восстановление исходной функции с нулевой ошибкой с помощью интерполирующих функций вида $(\sin x)/x$. Эта же теорема (известная в математике как теорема Уиттекера [93]) в англоязычной литературе упоминается как теорема Шеннона (1948). То, что функции с ограниченным спектром относятся к классу аналитических функций, которые не равны нулю на бесконечном интервале времени, и теорема при этом справедлива для функций с ограниченной энергией, более полувека считалось малозначимым; тщательный анализ применимости теоремы Котельникова [93] не поколебал популярности самой известной среди инженеров техники связи теоремы.

Проблему бесконечного спектра никто не скрывал, и К. Шеннон (совместно с Оливером и Пирсом в работе «The philosophy of P. C. M.» в [16]) даже полагает, что «*входной сигнал ограничивается по спектру так, чтобы исключить все частоты выше W_0 . Отсчёты затем квантуются и кодируются <...> В приёмнике регенерированные кодовые группы декодируются <...> эти импульсы пропускаются через фильтр нижних частот с*

границной частотой W_0 для восстановления первоначального сигнала». Некорректность такой теории в том, что она допускает предварительное **усечение неизвестного (а для незашумлённых непрерывных случайных сигналов – бесконечного!) количества информации** с последующим кодированием, имеющим целью достижение нулевой ошибки передачи. Поэтому восстанавливается отнюдь не «первоначальный сигнал», а его версия, искажённая усечением спектра. Можно сказать, что такая философия отражает попытку распространить на передачу непрерывных сигналов концепцию Шеннона. Её можно назвать концепцией **идеального связиста**, имеющей целью одновременное обеспечение *информационного равновесия* (равенства энтропии источника и пропускной способности канала, $H_\epsilon = C$, достижению которого посвящена гл. 1) и *нулевой ошибки передачи* $\epsilon = 0$.

Странность ситуации усугубляется тем, что введение ошибки усечения спектра сигнала делает невозможным достижение нулевой полной ошибки передачи и при кодировании его усечённой версии осуществляются огромные, в пределе – бесконечные, *затраты сложности и времени* (задержки передачи) для сведения к нулю *части ошибок*.

Не случайно К. Шеннон, переходя от кодирования дискретных источников к кодированию непрерывных источников, отмечает, что «*в непрерывных системах не будет делаться попыток получать результаты с наибольшей общностью или крайней строгостью чистой математики, так как это потребовало бы серьёзного применения абстрактной теории меры и затемнило бы основную нить нашего анализа*» [16]. Аналогично далее рассмотрение ведётся не на уровне математической строгости, а согласно широко известной (благодаря Д. Пойа) концепции правдоподобных рассуждений [64]: «*В высшей степени важным шагом в построении физической теории является её формулировка в математических терминах*».

Противоречивая ситуация общепринятого с давних времён необоснованного усечения спектра сигнала и определила потребность построения теории, формализующей способ ограничения спектра. Для модели линейного кодирования зашумлённых сигналов оказалось возможным найти способ определения *строго оптимальных* спектров кодирующих и декодирующих фильтров.

Основная идея обработки сигналов, искажённых шумом, состоит в достижении компромисса между неизбежными ошибками из-за входного шума и ошибками, связанными с шумом канала и ограничениями на слож-

ность кодирующих устройств. Впервые эта концепция была обоснована ещё до рождения теории связи в рамках методологии приближённых вычислений А. Н. Крыловым, прочитавшим курс лекций в 1906 г. и опубликовавшим его в 1911 г. Этот курс, многократно переизданный, нацелен на практическое применение математики в обработке результатов наблюдений. В его основе лежит утверждение «Для прикладных вопросов нет надобности производить вычисления по абсолютно точным формулам и с совершенной точностью; напротив, можно пользоваться заведомо неточными формулами или приёмами, лишь бы была уверенность, что происходящая от этого погрешность не превышает тех пределов, которые в данном вопросе допускаются». Эти слова характеризуют идею приближённых вычислений, или концепцию **экономного вычислителя**, введённую Кеплером и развивавшуюся Ньютоном, Эйлером, Лапласом, Ляпуновым и Пуанкаре. Опора на эту концепцию имеет перспективы включения в теорию кодирования достижений теории сложности. Концепция А. Н. Крылова используется во множестве инженерных наук, в том числе и в теории связи, включающей раздел кодирования зашумлённых сигналов.

Построение теории кодирования (обработки до передачи) непрерывных зашумлённых сигналов призвано дать *эталон* кодирования и отчасти опирается на известные результаты *теории оптимальной линейной фильтрации* (оптимального приёма) Колмогорова–Винера. Эта теория (в частности, работа Бодэ и Шеннона в [16]) рассматривает выделение доминантного (полезного) сигнала x , имеющего гауссовское распределение со спектром разложения в ряд Карунена–Лоэва $\{\lambda_k\}$ и доступного лишь после линейного искажения $\{B_k\}$ и воздействия аддитивного белого гауссовского шума ξ с равномерным спектром дисперсий σ_1^2 (индекс 1 для дисперсии шума и отношения сигнал/шум означает входной шум в отличие от индекса 2, означающего принадлежность к шуму канала). Для выделения полезного сигнала из смеси $\{y_k\}$, $y_k = B_k x_k + \xi_k$, по критерию минимума среднеквадратической ошибки необходим фильтр, формирующий условное математическое ожидание $\{\hat{x}_k\} = \{M[x_k | y_k]\}$ со спектром дисперсий $\{\hat{\lambda}_k\}$. Этот фильтр, известный как фильтр Винера, согласно изложению Бодэ и Шеннона [16], имеет спектральную характеристику:

$$L_{Bk} = \frac{\lambda_k B_k}{\lambda_k B_k^2 + \sigma_1^2} = \frac{1}{B_k} \frac{\psi_{1k}^2}{\psi_{1k}^2 + 1}, \quad (2.1)$$

$$\Psi_{1k}^2 = \frac{B_k^2 \lambda_k}{\sigma_1^2} \quad (2.2)$$

– спектральное отношение сигнал/шум, т. е. отношение дисперсий *доступного* сигнала (прошедшего искажающий фильтр $\{B_k\}$ доминантного сигнала) и входного шума.

Формируемое фильтром Винера условное математическое ожидание (оно же оптимальная оценка доминантного сигнала и минимальная достаточная статистика) обеспечивает минимальную ошибку оценивания доминантного сигнала (использованную в 1.4, см. (1.1)):

$$\varepsilon_B = \sum_{k=1}^{\infty} \frac{\lambda_k \sigma_1^2}{\lambda_k B_k^2 + \sigma_1^2} = \sum_{k=1}^{\infty} \frac{\lambda_k}{\Psi_{1k}^2 + 1} = \sum_{k=1}^{\infty} \frac{\hat{\lambda}_k}{\Psi_{1k}}, \quad (2.3)$$

а сам фильтр Винера обеспечивает извлечение из смеси y сигнала x с шумом ξ *потенциального количества информации* о доминантном сигнале (см. также (1.4)):

$$I(x, y) = \frac{1}{2} \sum_{k=1}^{\infty} \log \frac{\lambda_k}{\varepsilon_{Bk}} = \frac{1}{2} \sum_{k=1}^{\infty} \log(\Psi_{1k}^2 + 1). \quad (2.4)$$

Значение фильтра Винера в теории линейного кодирования определяется в первую очередь тем, что декодирующий фильтр в оптимальной системе всегда должен определяться по (2.1), вне зависимости от того, доступен ли управлению кодирующий фильтр или не доступен, как в модели фильтрации Колмогорова–Винера.

Первым шагом на пути построения теории кодирования зашумлённых сигналов стало введение понятия скорости создания информации при заданной ошибке воспроизведения К. Шенноном (1949, 1959 [16]) и А. Н. Колмогоровым, использовавшим для этого понятия термин «эпсилон-энтропия» (1956 [23]). К. Шеннон неоднократно подчёркивал, что «*практически при непрерывном источнике нас интересует не точная передача, а передача с определённым допуском*». Именно вопросу распределения ошибок при задании такого допуска и посвящена теория эпсилон-энтропии (*rate distortion theory*). Важнейшим результатом этой теории в приложении к линейному кодированию стала формула Колмогорова для эпсилон-энтропии H_ε гауссовского случайного процесса с дисперсиями $\{\lambda_k\}$ при среднеквадратической мере отклонения передаваемого x и принятого z сигналов, $\varepsilon = M[(x - z)^2]$. Эта формула была использована в гл. 1 при

нахождении оптимальных фильтров, обеспечивающих информационное равновесие.

Среднеквадратическая ошибка θ передачи каждой из спектральных компонент (порог усечения спектра сигнала) определяется из заданной полной среднеквадратической ошибки случайного процесса (использованной в 1.4) [23]:

$$H_\varepsilon = \min_{d(x,z) = \sum_{k=1}^{\infty} (x_k - z_k)^2 \leq \varepsilon} = I(x,z) = \frac{1}{2} \sum_{k=1}^{\infty} \log \frac{\lambda_k}{\theta}, \quad (2.5)$$

$$\varepsilon = \theta K + \sum_{k=K+1}^{\infty} \lambda_k. \quad (2.6)$$

Эти формулы, полученные из условия минимума взаимной информации при заданной ошибке ε , уже описывают *усечение спектра* гауссовского случайного процесса (отбрасывание всех спектральных компонент начиная с λ_{K+1}), которое в теории Шеннона не обосновывалось, хотя неопределённость назначаемой ошибки ещё остаётся. Важным свойством оптимизации (решения уравнения Эйлера для поиска минимума взаимной информации при ограничении ошибки, см. гл. 1) является *равномерное распределение дисперсии ошибок по оставляемым спектральным компонентам*. Как и в иллюстрации работы тест-канала Галлагера для реализации эpsilon-энтропии [52], спектр сигнала может убывать немонотонно, а спектр передаваемого сигнала разбивается на несколько областей. Мы для упрощения иллюстраций всюду принимаем монотонно убывающий спектр сигнала.

Однако этот важный шаг в развитии теории связи оставял без объяснения неизбежность введения допуска ε на точность передачи сигнала и предполагал использование введённой K . Шенноном концепции *идеально-го связиста*.

Важным условием учёта качества информации, как уже указывалось, является рассмотрение модели, включающей два источника шума – фотонный и считывания. Принципиальная необходимость учёта по крайней мере двух источников шума (до кодирования и после) связана с тем, что телевизионная система призвана формировать сигнал из светового стимула в присутствии собственных тепловых шумов. При этом условия рождения

телевизионного сигнала принципиально отличаются от типичных для классической радиотехники. Для телевизионного сигнала энергия светового кванта (кванта Планка) обычно превышает энергию теплового кванта (кванта Больцмана), что вызывает необходимость учёта как квантовой, так и волновой структуры света. Важно, что модель системы связи с одним источником шума не ведёт к усечению полосы частот, как это имеет место в теории фильтрации Винера и теории кодирования Шеннона.

Прямая постановка задачи использования модели с двумя источниками шума, т. е. кодирования зашумлённых сигналов, впервые была сделана Р. Л. Добрушиным и Б. С. Цыбаковым [89] (рис. 2.1).

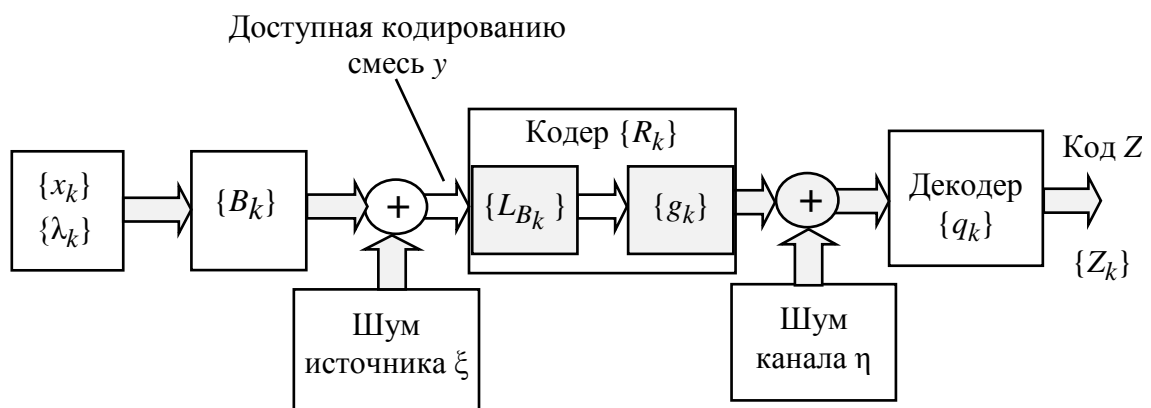


Рис. 2.1. Модель линейного кодирования зашумлённого сигнала

Основными результатами этой работы стали формула для эpsilon-энтропии гауссовского процесса, наблюдаемого в смеси с гауссовским шумом, и обоснование способа передачи зашумлённого сигнала, включающего предварительную операцию формирования условного математического ожидания $\{\hat{x}_k\} = \{M[x_k | y_k]\}$ с помощью фильтра Винера $\{L_{Bk}\}$ (2.1) и операцию фильтрации $\{g_k\}$, реализующую кодирование по заданному критерию качества оптимальной оценки доминантного сигнала [89]:

$$R_k = L_{Bk} g_k. \quad (2.7)$$

При этом эpsilon-энтропия подлежащей передаче *доступной* информации, т. е. оптимальной оценки, зависит не только от назначаемой ошибки ϵ , но и от минимальной ошибки ϵ_B (2.3) (рис. 2.2) [89]:

$$I(y, z) = \frac{1}{2} \sum_{k=1}^{\infty} \log \frac{\hat{\lambda}_k}{\epsilon_k - \epsilon_{Bk}} = \frac{1}{2} \sum_{k=1}^K \log \frac{\lambda}{\epsilon_k - \epsilon_{Bk}} \frac{\Psi_{1k}^2}{\Psi_{1k}^2 + 1}. \quad (2.8)$$

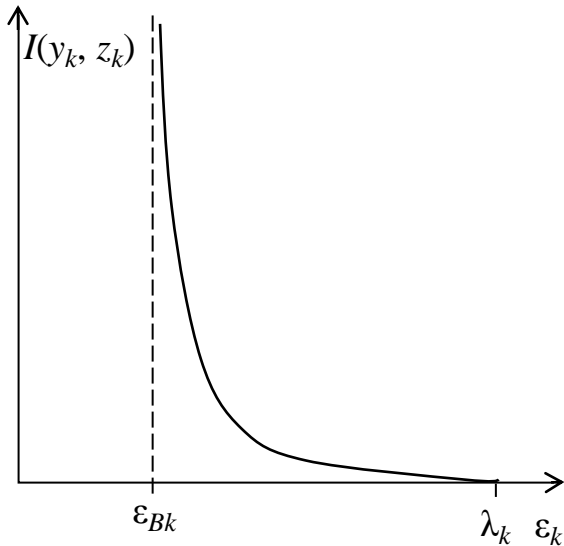


Рис. 2.2. Зависимость ϵ -энтропии (каждой из спектральных компонент кодовой информации (2.7)) от назначаемой среднеквадратической ошибки ϵ

Для достижения ϵ -энтропии (2.8) необходимо ошибку оптимальным образом распределить по спектру (рис. 2.3), что в отличие от (2.6) даёт:

$$\epsilon = \sum_{k=1}^{\infty} \epsilon_k = \theta K + \sum_{k=1}^K \epsilon_{Bk} + \sum_{k=K+1}^{\infty} \lambda_k. \quad (2.9)$$

Однако и этот важный шаг в развитии теории связи предполагал использование введённой К. Шенноном концепции *идеального связиста*. При этом сохранялось мнение, что при линейном кодировании (при ограничении на мощность сигнала в

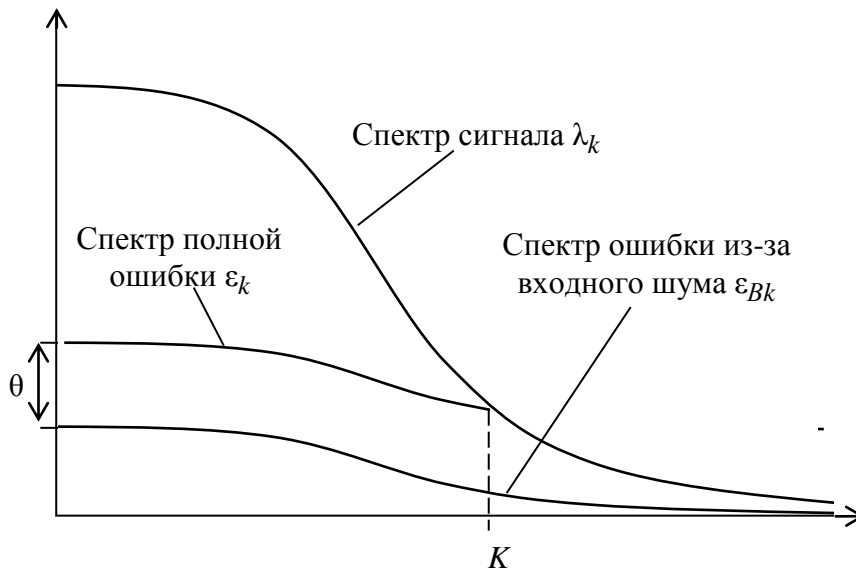


Рис. 2.3. Спектральные характеристики сигнала и ошибки передачи при реализации ϵ -энтропии зашумлённого источника

канале) не достигается равенство энтропии источника и пропускной способности канала. Это было связано с тем, что гауссовский канал передачи незашумлённого сигнала на начальном этапе развития теории линейного кодирования рассматривался с позиций минимума среднеквадратической ошибки при ограничении средней мощности P на выходе линейного кодирующего фильтра $\{G_k\}$ (использованного К. Шенноном лишь в качестве простого примера) и не учитывалась необходимость построения сопряжённых триад, рассмотренных в гл. 1.

Поэтому справедливо считалось, что $H_\varepsilon < C_P$. В гл. 1 мы показали, что построение сопряжённых триад позволяет достичь информационного равновесия при кодировании незашумлённых сигналов, далее подтвердим такую возможность и для кодирования зашумлённых сигналов.

2.2. Аксиоматика информатики рождения и передачи сигналов

Для преодоления трудностей применения теории к практическим системам, имеющим дело с передачей в реальном времени искажённых сигналов, придётся обратиться к аксиоматике теории информации, руководствуясь мнением А. А. Ухтомского: *«Формальная логика есть искусство самоутверждения в своём образе мысли из принятых аксиом. Наше дело – в перепахивании самих аксиом, во вскрытии их природы и тех оснований, которые делают их столь убедительными и неоспоримыми»* [94]. При этом аксиоматические системы, описывающие одни и те же совокупности объектов, могут строиться по-разному: например как предлагаемая далее, где за основу принята взаимная информация [23], или как в квантовой теории информации, где за основу взята собственная информация [10]. В любом случае аксиоматика теории – это не просто недоказанные гипотезы, а гипотезы, являющиеся *следствием накопленного опыта*. Поэтому в силу постоянного накопления опыта аксиоматика периодически изменяется во времени, совершенствуется в процессе исторического развития познания.

Главным вектором развития теории информации должен стать учёт смысла информации на основе системного подхода [14]. Однако эта крупнейшая проблема может иметь ряд промежуточных решений, к которым относится учёт качества информации, на несемантическом уровне достигаемый с помощью *введённого понятия шумовой информации*.

В рамках модели случайного сигнала, лишённого смысла, при отсутствии входного шума все его спектральные компоненты одинаково высокого качества. Желание найти показатель качества информации, учитывающий смысл, выходит за рамки статистической теории: *«Количество информации само является числом, но как оценить качество информации числом? Вот где начинается настоящая вакханалия вредного субъективизма, порождающего море лжи»* [25]. Даже в рамках статистической теории качество информации формализовать не просто: *«Качество информации не удаётся выразить с помощью одного функционала подобно количе-*

ству информации» [25]. Разрешение этой трудности в рамках теории линейной фильтрации случайных сигналов возможно с помощью дополнительного инструментария, в первую очередь – дополнительных аксиом [3], [95], [96], распространяющих теорию линейного кодирования гауссовских сигналов на зашумлённые сигналы.

Обобщение накопленного опыта применения теории информации к кодированию зашумлённых сигналов [1]–[3], [56], [57], [89], [95]–[98] позволило сформулировать ряд дополнительных аксиом для кодирования зашумлённых сигналов. Далее приводятся эти аксиомы и конспективно излагаются следствия из них. При развёрнутом изложении теории большинство приводимых следствий из этих аксиом должно быть оформлено как теоремы, как, например, теорема об идеальном линейном кодировании [57].

Уравнение связи может включать не только количество, но и качество информации. Конечно, учёт качества информации – это ещё не учёт её смысла, но является первым шагом к нему. Панацеи в оценке качества не существует, но можно проиллюстрировать возможность оценки качества на простом примере *линейного кодирования зашумлённых сигналов*. Эта теоретическая задача весьма актуальна, в частности при разработке телекамер. Отмеченная неантагонистическая игра проектировщиков телекамер и каналов связи находит воплощение в реализации в телекамере одновременно элементов кодирования источника и кодирования канала [2]. Так, накопление сигнала в элементе разложения относится к кодированию источника, а развёртка – к кодированию канала. При этом, в частности, вклад в сжатие (компрессию) сигнала в ходе накопления сигнала в телекамере больше, чем вклад последующего кодирования цифровым кодером [2]. Оба вида кодирования – источника и канала – в ходе указанной игры неоднократно чередуются.

В системах передачи зашумлённых сигналов (а таковы все телевизионные сигналы) центральная проблема связи – равновесное согласование источника и канала – усложняется. Обойтись *только* одним понятием взаимной информации не удаётся. Для выхода из «тупика скалярной информации» следует обратить внимание на уже цитированную на с. 14 мысль А. Н. Колмогорова, отмеченную им восклицательным знаком. Соглашаясь с ним в том, что информация не является скалярной величиной, следует найти конструкцию, более полно отражающую *интересующую нас информацию*.

Как и при кодировании неискажённых сигналов, будем использовать сопряжённую триаду априорной информации, критерия и качества, для определённости взяв критерий минимума среднеквадратической ошибки и ограничение широкополосности, рассмотренные в гл. 1. Корректная информационная оценка линейного кодирования зашумлённых сигналов достигается при опоре на принцип доминантной информации. Помня слова А. А. Ухтомского, мы считаем достаточно убедительным и неоспоримым, что *на роль доминанты назначается* один из сигналов, остальные же случайные процессы (которых в общем случае может быть много, например сигналы множества станций, принимаемых антенной радиоприёмника и шумы) относятся к *фону* и с позиций статистического описания *равноправны* с ним. Неравноправие этих сигналов есть лишь следствие вмешательства наблюдателя (получателя информации). Этот подход позволил выдвинуть следующую аксиому [3], [95], [96].

Аксиома 1 – априорного равноправия сигналов: существует взаимная информация между выходным сигналом кодирующей системы и любого из компонентов входного воздействия системы связи – доминантным сигналом, фоновыми сигналами и шумами.

Следствием из аксиомы априорного равноправия является то, что в простейшей модели системы связи (см. рис. 2.1) получателю предъявляется прошедшая канал связи кодированная смесь доминантного сигнала и входного шума (*доминантный сигнал на фоне шума*). Получатель имеет дело не только с доминантной информацией, но и с информацией $I_{\text{ш}}$ о входном шуме, т. е. с **шумовой информацией**.

Шумовая информация – это взаимная информация между входным шумом и суммой шумов в выходном коде, доступном получателю: $I_{\text{ш}} = I(\xi, g\xi + \eta)$. Она равна разности доступной информации (взаимной информации $I(y, z)$ между искажённым шумом сигналом y и получаемым кодом z) и доминантной информации (взаимной информации $I(x, z)$ между доминантным сигналом x и получаемым кодом z):

$$I_{\text{ш}} = I(y, z) - I(x, z). \quad (2.10)$$

Расщепление взаимной информации на две – доступную $I(y, z)$ и доминантную $I(x, z)$ – при кодировании зашумлённых сигналов очевидно. Однако введение понятия шумовой информации в статистическую теорию синтеза кодирующих устройств – существенно более серьёзный шаг. Сте-

реотип восприятия понятия «информация», его идентификация с понятием *пользы*, часто вызывает в среде специалистов техники связи сопротивление синтезу этого понятия с понятием «шум». Конечно, термин устанавливался с трудом, не сразу: авторы понятия, дав формулу (2.10), её значение для гауссовского канала в [56] и даже «говорящее» обозначение $I_{ш}$, отважились лишь на термин «неустранимой избыточности». Позже, в [1], эта величина называлась «ложной информацией», в [57] использовались термины «неустраняемая избыточность», «мусорные биты», «информационный мусор». Начиная со статьи [98] и последовавшей за ней книгой [2] все более поздние публикации нашей научной школы используют термин «шумовая информация».

К понятию шумовой информации вплотную примыкают, правда на интуитивном уровне, без определения понятия и свойств величины, высказывания специалистов по синергетике [37]: *«В эпоху информационного шума... очень немногие знают о последних работах коллег»*. Кстати, журналисты потребность в термине «шумовая информация» прочувствовали раньше многих «технарей», и этот термин можно встретить и в периодической печати, и в телевизионных передачах, и в интернете, где намечаются пути введения цензуры как инструмента борьбы с ложной и шумовой информацией (к которой относят спам). Так, С. Лем отмечал [99]: *«Сейчас мы уже имеем ситуацию информационного потопы... Люди, поставленные под такие информационные ливни, проводят селекцию поступающей информации... Распознавание качества информации, полное неприятие к сведению информации несущественной, рекламной, второразрядной, попросту лишней для человека, является обязательным условием для "удержания на плаву" в нарастающем информационном потоке»*. Далее Лем продолжает дискуссию о том, что *«существует ли "информационный мусор", который из интернетов любой структуры следовало бы выметать?»*, и отвечает, сначала робко: *«Ответ, к сожалению, будет отягощён релятивизмом. Что для одного пользователя является мусором, то для другого представляет информацию со всех сторон питательную, удобоваримую. И даже, возможно, необходимую»*. Далее он усиливает этот взгляд: *«Что же касается "мусора", то издавна, на заре кибернетики, было известно, что то, что для одного человека является "шумом" (заглушающим информацию), для другого, например, инженера-инфор-*

матика, изучающего noise level в канале, передающем информацию, является именно той информацией, которую он разыскивает и измеряет».

Здесь выдающийся философ науки путает стадии передачи информации. Изучение шума – это *этап формирования априорной информации* для передачи *доминантной информации*. Интересно, что когда проектировщик начнёт изучать шум, то в нём окажутся замешаны другие (измерительные) шумы, и *от понятия шумовой информации не уйти*, как бы этого С. Лему ни хотелось. Хотя сама идея о том, что некая структурированная информация может быть для одних полезна, для других вредна, видна из намёка С. Лема о питательной информации. Эта идея восходит к мысли Э. Шрёдингера, что живые организмы питаются не энергией, а информацией (он писал «негэнтропией», но это не меняет сути взгляда). Конечно, отходы жизнедеятельности животных являются питательной средой для растений, но это различие в оценке качества информации лишь подчёркивает тезис о том, что *без получателя нет связи*.

Логическим обоснованием законности термина «шумовая информация» является то, что в статистической теории связи понятие информации с самого начала – с работ Р. Хартли и К. Шеннона – отрывалось от понятия смысла. Более того, как процитировано на с. 27, А. Н. Колмогоров настаивал на главенстве в теории информации понятия «взаимной информации», а не энтропии. При этом «взаимность» может быть между любыми случайными процессами. Поэтому введение понятия «шумовая информация» не только не противоречит понятию «взаимная информация», но и делает первый шаг в наполнении конкретным содержанием утверждения А. Н. Колмогорова об информации, не являющейся скалярной величиной.

Для гауссовской системы связи взаимные информации $I(y, z)$, $I(x, z)$, $I_{ш}$, как и *потенциальное значение информации* $I(x, y)$ о зашумлённом доминантном сигнале, вычисляются достаточно просто [56], [57].

Следствие из аксиомы 1: система передачи зашумлённого сигнала может быть охарактеризована тремя взаимными информациями в доставленном получателю коде z (рис. 2.4):

- о доминантном сигнале x :

$$I(x, z) = \frac{1}{2} \sum_{k=1}^K \log \frac{\lambda_k}{\varepsilon_k}; \quad (2.11)$$

- доступном сигнале y :

$$I(y, z) = \frac{1}{2} \sum_{k=1}^K \log \frac{\lambda_k}{\varepsilon_k - \varepsilon_{Bk}} \frac{\psi_{1k}^2}{\psi_{1k}^2 + 1}; \quad (2.12)$$

- входном шуме ξ :

$$\begin{aligned} I(\xi, g\xi + \eta) &= I(y, z) - I(x, z) = \\ &= \frac{1}{2} \sum_{k=1}^K \log \frac{\varepsilon_k}{\varepsilon_k - \varepsilon_{Bk}} \frac{\psi_{1k}^2}{\psi_{1k}^2 + 1}. \end{aligned} \quad (2.13)$$

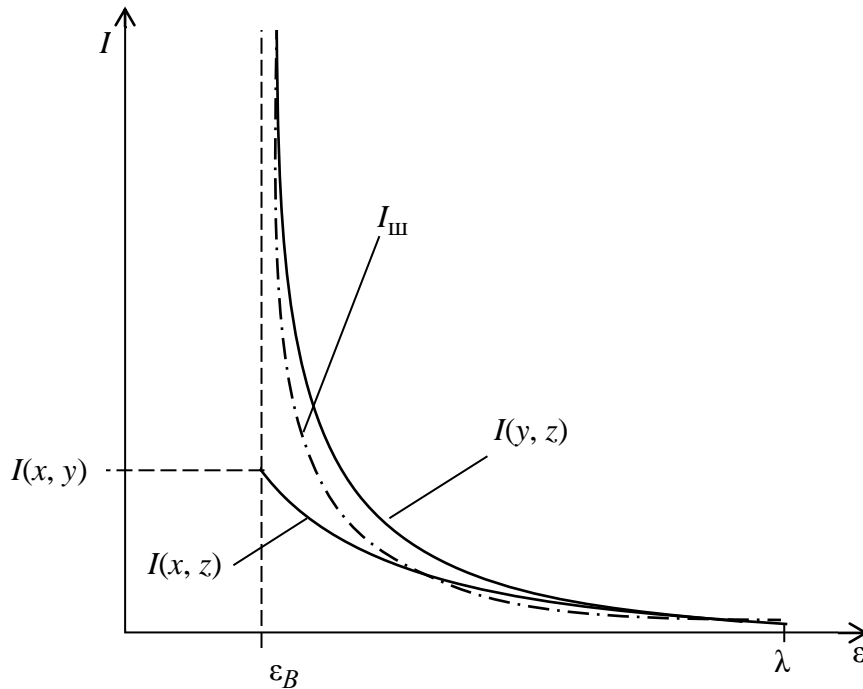


Рис. 2.4. Три взаимных информации: доступная $I(y, z)$; доминантная $I(x, z)$; шумовая $I_{ш}$

Первая взаимная информация (2.11) – **доминантная**. Она вычисляется с учётом полной ошибки (2.9) и не превышает потенциальной взаимной информации (2.4) между доминантным сигналом x и его искажённой шумом версии – доступным сигналом y :

$$I(x, z) = \frac{1}{2} \sum_{k=1}^K \log \frac{\lambda_k}{\varepsilon_k} \leq I(x, y) = \frac{1}{2} \sum_{k=1}^{\infty} \log \frac{\lambda_k}{\varepsilon_{Bk}} = \frac{1}{2} \sum_{k=1}^{\infty} \log (\psi_{1k}^2 + 1). \quad (2.14)$$

Вторая взаимная информация (2.12) может быть названа **доступной**, или **кодовой**; третья (2.13) – **шумовой**, которую далее будем обозначить $I_{ш}$.

Эта методология вскрыла парадоксальную ситуацию: смесь сигнала и шума содержит в себе конечное количество полезной информации о доминантном сигнале, но для её передачи без потерь требуется передача *беско-*

нечного количества смешанной сигнальной и, как следствие, шумовой информации. С позиций восприятия образов в гештальт-психологии шумовой фон является простейшим, так как сам считается не имеющим смысла. Такой подход скрывает вредоносную сущность открытой величины: *шумовая информация является главным «пожирателем» пропускной способности канала связи [65].*

Подчёркивая роль введённого понятия шумовой информации, следует напомнить процитированную на с. 13 мысль А. Пуанкаре, что *«бывает достаточно изобрести одно новое слово, и это слово становится творцом».* Мы покажем далее, что таким творцом в теории доминантной информации стало новое «слово» «шумовая информация». Конечно, формула для вычисления шумовой информации (2.13) важна, но её одной явно недостаточно. Следуя указанию Р. Карнапа, процитированному на с. 12, 13, мы покажем, какие новые законы используют это новое понятие.

Для выявления закона, характеризующего поведение шумовой информации, рассмотрим влияние назначаемой ошибки передачи на три вида взаимной информации (2.11)–(2.13). То, что шум канала ведёт к потере ΔI доминантной информации, тривиально, эту потерю легко понять и вычислить. Для случайных величин она просто определяется разностью кодовой информации (2.12) и доставленной полезной информации (2.14), т. е. логарифмом отношения реальной полной ошибки, вызванной совместным влиянием входного шума и шума канала, к ошибке из-за входного шума, обеспечиваемой фильтром Винера:

$$\Delta I = I(x, y) - I(x, z) = \frac{1}{2} \log \frac{\varepsilon}{\varepsilon_B}. \quad (2.15)$$

Для случайных процессов потеря доминантной информации включает два компонента, определяемые влиянием шума в полосе пропускания и усечением спектральной характеристики оптимальным кодирующим фильтром:

$$\Delta I = \frac{1}{2} \sum_{k=1}^K \log \frac{\varepsilon_k}{\varepsilon_{Bk}} + \frac{1}{2} \sum_{k=K+1}^{\infty} \log(\psi_{1k}^2 + 1). \quad (2.16)$$

Однако более важным свойством систем передачи зашумлённых сигналов является то, что уменьшение потери доминантной информации ΔI может быть достигнуто только ценой передачи большего количества шумовой информации. Уровень шума канала по-разному влияет на три вида вза-

имной информации (2.11)–(2.13). При увеличении шума канала (и назначаемой ошибки ε) они убывают, но с разной скоростью. Для наглядного представления этих различий рассмотрим простой вариант кодирования случайных величин, а не процессов (в (2.11)–(2.13) исчезают знаки сумм и индексы k). Для малых значений назначаемой ошибки передачи (незначительно превышающей ошибку ε_B из-за входного шума) между потерей доминантной информации ΔI и передачей шумовой информации $I_{ш}$, а также между потерей доминантной информации и доступной (кодовой) информацией существует **закон взаимобмена** [2]:

$$\Delta I 2^{2I_{ш}} \approx \frac{\Psi_1^2}{2 \ln 2 (\Psi_1^2 + 1)}, \quad \Delta I 2^{2I(x,y)} \approx \frac{\Psi_1^2}{2 \ln 2}. \quad (2.17)$$

Закон (2.17) показывает *бесконечный рост шумовой информации при стремлении к передаче доминантного сигнала с предельной ошибкой* фильтра Винера ε_B . Конечно, этот закон справедлив только в рамках непрерывных моделей сигналов и шумов, при дискретных распределениях этот рост конечен. Вместе с тем пренебрегать им даже при дискретных распределениях нельзя. Например, при ёмкости пиксела порядка 10^6 электронов и наблюдении случайных полей обнаруживаемого, но невысокого контраста можно говорить, что шумовая информация может превышать доминантную на 2 порядка.

Наличие обменных соотношений (2.17), которые по аналогии с соотношением неопределённостей Гейзенберга, называемым им самим «*соотношением неточностей или принципом неопределённости*» [22], может быть названо **информационным соотношением неопределённости**, является предпосылкой для поиска оптимальной системы связи, в которой уравновешены негативные явления различного вида (потеря полезной информации, чрезмерная требуемая пропускная способность или её загрузка передачей информации о входном шуме). Связь между потерей полезной информации (2.15), кодовой информацией (2.12) и шумовой информацией позволяет построить характеристику системы передачи [97], аналогичную рабочей характеристике приёмника в задачах проверки гипотез [69], [88]. Эту характеристику (рис. 2.5) можно назвать *рабочей характеристикой системы связи*. Она и показывает, что уменьшение потери полезной информации может быть достигнуто только ценой передачи большего количества шумовой информации.

В многомерном случае, т. е. при передаче случайных процессов, при стремлении к минимальной ошибке ε_B , потребуется бесконечная широкополосность кодера, которая будет тратиться на передачу в основном шумовой информации. Это положение обостряет вопрос о качестве переданной информации. Развиваемая концепция позволила выдвинуть следующую аксиому [3], [95], [96].

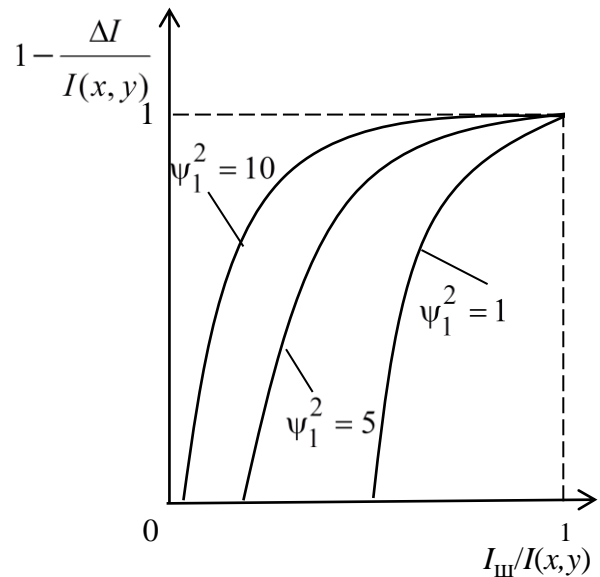


Рис. 2.5. Взаимосвязь потери полезной информации ΔI и шумовой информации $I_{ш}$

Аксиома 2 – повышения качества информации. На высшем (семантическом) уровне качество информации – это достижимость поставленной цели. Этот уровень крайне трудно формализуем, но данное определение позволяет осознанно подойти и к определению качества несемантической информации.

На низшем (несемантическом) уровне качество информации определить проще (см. аксиому 4), потому что априорной самодостаточной целью системы связи является передача получателю (человеку или автомату) информации **максимального качества** при том, что её количество **минимизируется усечением некачественной информации**.

При этом следует различать два подхода к качеству:

- *качество доставленной получателю информации* – то, насколько полученный код при минимизации количества информации соответствует состоянию источника;
- *качество кодирования* (передачи информации) – то, как кодер использует входную информацию.

Качество переданной информации о зашумлённом сигнале перестаёт быть неопределённой величиной, а зависит не только от оптимизации кодера, обеспечивающего информационное равновесие ($H_{\varepsilon} = C$), но и от входного спектрального отношения сигнал/шум (2.2). Развивая цитированную на с. 14 мысль А. Н. Колмогорова об «*интересующей нас информации*», можно утверждать, что получателя интересуют *качественная информация*.

Аксиома повышения качества может быть сформулирована так: **кодирование и декодирование сигналов, в том числе зашумлённых, имеет**

целью повышение качества доставленной получателю информации о доминантном сигнале.

Из неё следует, что качество информации *должно быть оценено количественно* с учётом соотношения различных видов взаимной информации в системе передачи зашумлённых сигналов. В общем случае проблема оценки качества информации необозримо сложна [25], но в простом (несемантическом) случае линейного кодирования зашумлённых сигналов она разрешена. Для решения задачи количественной оценки качества информации потребовалось сформулировать дополнительную аксиому об информационном риске [3], [95], [96].

Аксиома 3 – информационного риска. Она формализует второй из обозначенных в предыдущем пункте подходов – концепцию *качества кодирования* зашумлённых сигналов – то, как кодер распорядился доступной ему информацией. Качество кодирования должно учитывать оба вида искажения доминантного сигнала: потерю информации о нём $\Delta I = I(x, y) - I(x, z)$ и внесение шумовой информации $I_{ш}$. При этом *риск* передачи неправильной информации увеличивается с увеличением любого из этих искажений. В частности, в силу нефинитности спектров реальных сигналов любой кодер будет обязательно вносить искажения в доминантный сигнал и не будет в состоянии подавить до нуля шумовую информацию.

Цель введения аксиомы информационного риска – решительный отход от концепции идеального связиста, предписывающей нулевую ошибку передачи при ненулевой ошибке в доступном сигнале. Напомним, что эта «нулевая» ошибка в теории Шеннона относится к сигналу, искажённому, по крайней мере, произвольным усечением спектра, и покупается бесконечной сложностью и задержкой.

Как и в теории решений, оптимальная стратегия имеет целью минимизацию среднего риска, состоящего из взвешенной вероятности ошибок принятия альтернативных гипотез. В теории кодирования зашумлённых сигналов *оптимальная стратегия* заключается в *минимизации информационного риска*, состоящего из совокупности разнородных взаимных информаций. Как в теории решений критерий среднего риска (Байеса) строится из совокупности разнородных вероятностей [88, т. 1], в теории доминантной информации он строится из совокупности разнородных взаимных информаций. Вероятность – скалярная величина, и с помощью какой-либо одной вероятности нельзя найти оптимальную стратегию принятия решения. Проблему оптимизации в теории решений удалось преодолеть только

введением вектора из двух вероятностей различных видов ошибочных решений. Аналогично проблема оптимизации в теории информации не могла быть решена с помощью только потери одной из информаций (хотя попытки были: например, Д. Миддлтон [49] безуспешно искал систему связи с минимумом потери информации, в теории телевидении тоже были такие попытки [41], [44]), она должна решаться введением вектора по крайней мере из двух величин – потери доминантной информации и шумовой информации.

На семантическом уровне искомый информационный компромисс означает поиск баланса между истиной и ложью. Слова Ф. И. Тютчева о том, что «**мысль изречённая есть ложь**», отражают в первую очередь принципиальную трудность разделения истины и лжи: любое сообщение будет содержать их смесь. То, что ошибки неизбежно присущи любому сообщению, знал и Рабиндранат Тагор, утверждавший, что *заккрытие дверей перед ошибками закрывает их и перед истиной*. В рассматриваемом случае линейного кодирования зашумлённых сигналов это означает предостережение от попытки назначить чрезмерно высокую цену шумовой информации. Вместе с тем *качество кодирования*, т. е. то, насколько рационально распорядился кодер доступной ему смесью доминантной и шумовой информации, характеризуется также тем, сколько доминантной информации было потеряно. Информационный риск, конечно, должен отражать приведённое мнение Рабиндраната Тагора, но мы возьмём на себя смелость дополнить стих писателя-философа двойственной формулировкой:

*Перед ошибками захлопываем дверь –
Ни слова лжи, но всё ж
В смятенье истина: «**Как** я войду теперь?»
Пред истиной распахиваем дверь –
Лавиной хлынет ложь!
В смятенье истина: «**Куда** войду теперь?»*

Информационный риск интуитивно ясен, к нему приближался и Л. Н. Гумилёв, говоривший, что «*Излишние детали знать совершенно ни к чему, это гегелевская "дурная бесконечность", но и недостаток знаний не даёт повода сделать правильный вывод*».

Аксиома информационного риска может быть сформулирована так: ***информационный риск определяется взвешенной суммой потери информации о доминантном сигнале и прохождения информации об остальных компонентах входного воздействия.***

Из неё следует, что в общем случае мы приходим к **критерию минимума информационного риска**, включающему взвешенную сумму потери доминантной информации и ложной информации (фоновой и шумовой). В упрощённой модели кодирования зашумлённых сигналов (другого фона нет) ложная информация состоит только из шумовой и **критерий минимума информационного риска**, определяется взвешенной суммой потери информации о доминантном сигнале (2.15) и дошедшей до получателя шумовой информации (2.13) [2], [95], [96]:

$$R_I = c_0 \Delta I + c_1 I_{\text{ш}}. \quad (2.18)$$

Весовые коэффициенты в критерии минимума информационного риска, как и в критерии Байеса, отражают отношение получателя к ошибкам различного рода. В варианте равенства цен потери доминантной информации и загрузки канала шумовой информацией $c_0 = c_1$ развитая концепция приводит к критерию [2], [56]:

$$R_I = \Delta I + I_{\text{ш}}. \quad (2.19)$$

Критерий (2.19) назван **критерием экономного связиста** [96] по аналогии с **критерием достаточности точности** акад. А. Н. Крылова. Этот критерий имеет аналог в теории решений в виде критерия идеального наблюдателя Котельникова–Зигерта, *уравновешивающего два вида ошибок при равенстве их цен*. Близость смысла и формы записи критериев теории проверки гипотез и доминантной теории информации подчёркивает их методологическое единство.

Информационный риск ровно также эффективен для синтеза систем передачи зашумлённых случайных сигналов, как эффективен средний риск в теории решений – обнаружении и оценивании. Зачем *средний риск*? Для *оптимального* выбора процедур приёма – *фильтров и порогов* при определяющем влиянии шума канала. Зачем **информационный риск**? Для *оптимального* выбора процедур кодирования для *передачи – фильтров и порогов* при определяющем влиянии входного шума.

При использовании критериев (2.18) и (2.19) существуют оптимальные значения полной ошибки ε передачи зашумлённого сигнала, обеспечивающие компромисс между потерей доминантной информации и передачей шумовой информации и выражаемые через значение минимальной ошибки ε_B , формируемой фильтром Винера. Для критерия экономного связиста минимум информационного риска достигается при условии равен-

ства ошибок передачи, вызванных входным шумом и шумом канала: $\varepsilon = 2\varepsilon_B$. Чем больше вес потери доминантной информации в критерии (2.18) по сравнению с весом шумовой информации, тем ближе оптимальное значение ошибки к потенциальному значению ε_B .

В гауссовском одномерном случае критерий экономного связиста вычисляется достаточно просто:

$$R_I = \frac{1}{2} \log \frac{\varepsilon}{\varepsilon_B} + \frac{1}{2} \log \frac{\varepsilon}{\varepsilon - \varepsilon_B} \frac{\psi_1^2}{\psi_1^2 + 1} = \frac{1}{2} \log \frac{\varepsilon^2}{\varepsilon_B(\varepsilon - \varepsilon_B)} \frac{\psi_1^2}{\psi_1^2 + 1}. \quad (2.20)$$

Из аксиомы информационного риска кроме двух форм критериев можно вывести ряд важных следствий.

Следствие 1 из аксиомы 3. Существуют оптимальные значения полной ошибки ε передачи зашумлённого сигнала, обеспечивающие минимум (2.20), т. е. компромисс между потерей полезной информации и передачей шумовой информации для критерия минимума информационного риска и критерия экономного связиста:

$$\varepsilon_R = \arg \min R_I = \varepsilon_B(1 + c_1/c_0); \quad \varepsilon_R = \arg \min R_I = 2\varepsilon_B. \quad (2.21)$$

Простая формула (2.21) имеет простое объяснение: в оптимальной системе кодирования зашумлённых случайных величин *полная ошибка распределяется поровну между влиянием входного шума и шума канала*. Эта концепция давно применялась в телевидении при построении телевизионных трубок с докоммутиационным усилением – усиление первого каскада увеличивали лишь до тех пор, пока заметность шума усилителя не сравняется с заметностью фотонного шума.

Указанное следствие иллюстрируется рис. 2.6.

Следствие 2 из аксиомы 3. Минимальное значение информационного риска зависит от назначаемых цен при потере доминантной информации и при шумовой информации. Чем меньший вес проектировщик придаёт шумовой информации, тем меньший информационный риск может быть достигнут. Вместе с тем при равенстве цен, т. е. при использовании критерия экономного связиста (2.19), существует важная граница: для достаточно больших значений отношения сигнал/шум – лишь несколько менее 1 бит/символ:

$$\min R_I = 1 + \frac{1}{2} \log \frac{\psi_1^2}{\psi_1^2 + 1}; \quad \lim_{\psi_1^2 \rightarrow \infty} \min R_I = 1 \text{ бит/символ}. \quad (2.22)$$

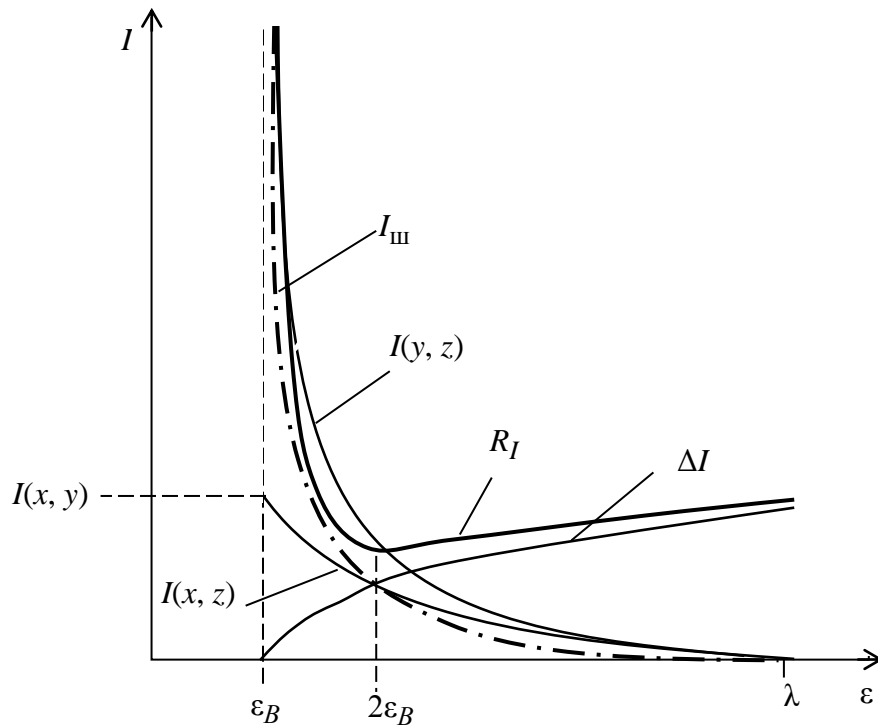


Рис. 2.6. Зависимость информационного риска R_I и его компонентов от назначаемой ошибки передачи ϵ при использовании критерия экономного связиста

Эта важная граница для информационного риска отражает наше *принципиальное незнание*, и в теории доминантной информации играет ту же роль, что и понятие абсолютного нуля в физике, которое носит титул закона Нернста. Не зря С. Лем утверждал [99]: «**Информация без фальсификаций, к сожалению, является фикцией**». Можно спросить – какой эффект от этого утверждения Лема? Оно – стимул для формализации **закона о существовании нижнего предела информационного риска**, для гауссовского канала выражаемого формулой (2.22).

К. Черри (см. цитату на с. 79) мечтал извлечь всю полезную информацию и отбросить всю фиктивную (шумовую). Можно сказать, что один из конструктивных результатов теории доминантной информации – установление **пределов возможного** (в виде закона взаимобмена компонент информационного риска и его нижней границы) на пути намеченного Черри **информационного «вечного двигателя**». К. Черри рассуждал о пропускной способности, но его слова относятся и к найденной границе информационного риска: «**Каждый, кто попытается превысить этот предел, зря потеряет время!**».

При этом минимальном риске количество доставленной получателю полезной информации составляет:

$$I(x, z) = I(x, y) - \Delta I_{R_{\min}} = I(x, y) - \frac{1}{2} = \frac{1}{2} \log \frac{\psi_1^2 + 1}{2}. \quad (2.23)$$

Эта формула означает, что хотя минимальный риск и составляет 1 бит/символ, в оптимальной системе кодирования зашумлённых случайных величин переданная информация о доминантном сигнале всего на полбита меньше потенциальной, т. е. *информационный риск распределён практически поровну между потерей доминантной информации и шумовой информацией.*

Следствие 3 из аксиомы 3. Существует **пороговое отношение сигнал/шум**. При отношении сигнал/шум меньше порогового минимум информационного риска достигается при равенстве ошибки передачи и дисперсии полезного сигнала, что соответствует прекращению связи:

$$\psi_{1\text{пор}}^2 = \frac{c_1}{c_0}. \quad (2.24)$$

Формула (2.24) показывает, что **наличие порога обусловлено существованием шумовой информации**. Пренебрежение ею эквивалентно равенству нулю весового коэффициента c_1 при ней в критерии минимума информационного риска. В частном случае критерия экономного связиста (2.19) пороговое отношение сигнал/шум равно 1: $\psi_{1\text{пор}}^2 = 1$.

Следствие 4 из аксиомы 3. Для многокаскадной системы связи с гауссовскими сигналами и шумами при заданной общей ошибке передачи ε минимум информационного риска достигается не при равномерном распределении ошибки, а при равномерном распределении по каскадам потери доминантной информации [100]:

$$\Delta I_j = \text{const}. \quad (2.25)$$

Так как потеря доминантной информации определяется логарифмом назначаемой ошибки ε , то она должна распределяться по каскадам по экспоненциальному закону – чем ближе каскад ко входу системы связи, тем меньше назначаемая ошибка передачи. Для критерия минимума информационного риска (2.18) в системе передачи гауссовских случайных величин распределение ошибки передачи по каскадам должно осуществляться по правилу [100]:

$$\varepsilon_j = \frac{\varepsilon_B c_1}{c_0} \left(1 + \frac{c_1}{c_0}\right)^{j-1}, \quad j \geq 1, \quad (2.26)$$

а для критерия экономного связиста (2.19) –

$$\varepsilon_j = \varepsilon_B 2^{j-1}, \quad j \geq 1. \quad (2.27)$$

Такое построение многокаскадной системы соответствует инженерной практике, уделяющей наибольшее внимание входным каскадам видеотракта.

Энергетический подход к синтезу системы связи, опирающийся на понятие коэффициента шума, не позволяет найти оптимальную систему, так как рекомендация выбора такого варианта, при котором обеспечивается минимум коэффициента шума, ведёт к нереализуемому коэффициенту шума $K_{\text{ш}} = 1$. Это приводит к конструированию гибридных критериев типа «эффективность – стоимость», а по сути к выходу за рамки информационной теории (см. цитату из [50] на с. 28). Вместе с тем введение на основе аксиомы априорного равноправия сигналов понятия шумовой информации позволяет не просто отбросить понятие коэффициента шума, а адаптировать полученные результаты на уровне терминологии, опирающейся на коэффициент шума. Для схемы рис. 2.1 и обозначений на ней коэффициент шума составляет:

$$K_{\text{ш}} = \frac{L_B^2 g^2 \sigma_1^2 + \sigma_2^2}{L_B^2 g^2 \sigma_1^2}.$$

Все упомянутые взаимные информации – $I(x, z)$, $I(y, z)$, $I_{\text{ш}}$, а также потеря доминантной информации ΔI , могут быть выражены через коэффициент шума $K_{\text{ш}}$ [100]:

$$I(x, z) = \frac{1}{2} \log \left(1 + \frac{\Psi_1^2}{K_{\text{ш}}} \right); \quad I(y, z) = \frac{1}{2} \log \left(\frac{\Psi_1^2 + K_{\text{ш}}}{K_{\text{ш}} - 1} \right); \quad I_{\text{ш}} = \frac{1}{2} \log \left(\frac{K_{\text{ш}}}{K_{\text{ш}} - 1} \right);$$

$$\Delta I = \frac{1}{2} \log \left(K_{\text{ш}} \frac{\Psi_1^2 + 1}{\Psi_1^2 + K_{\text{ш}}} \right); \quad R_I = \frac{1}{2} \log \left(K_{\text{ш}}^2 \frac{\Psi_1^2 + 1}{(\Psi_1^2 + K_{\text{ш}})(K_{\text{ш}} - 1)} \right).$$

При этом бросается в глаза, что шумовая информация определяется коэффициентом шума – и больше ничем, и она наглядно показывает знакомый практикам запрет на значение коэффициента шума, равное 1. Аналогично тому, как из формулы для информационного риска (2.20) для критерия экономного связиста (2.19) дифференцированием и приравниванием производной нулю получено оптимальное значение полной ошибки (2.21) $\varepsilon = 2\varepsilon_B$, можно найти оптимальное значение коэффициента шума, обеспечивающего минимум информационного риска:

$$K_{\text{ш opt}} = 2 \frac{\psi_1^2}{\psi_1^2 - 1}.$$

Отсутствие оптимального коэффициента шума для передачи сигналов с отношением сигнал/шум, меньшим 1, так же как и рабочая характеристика системы передачи на рис. 2.5, показывает наличие порогового отношения сигнал/шум, в приведённых формулах равного 1 и зависящего от отношения цен потери доминантной информации и шумовой информации (2.25).

Таким образом, как это всегда бывает с новыми теориями, которые превращают старые в частный случай, изложенная концепция не отрицает коэффициент шума, а переводит его в ранг понятий, относящихся к *анализу систем*, т. е. делает понятием теории, являющейся частным случаем теории *синтеза*. Поэтому эффективно использовать коэффициент шума при статистическом синтезе можно только в сочетании с изложенной аксиоматикой кодирования зашумлённых сигналов.

Воспользовавшись результатом в виде критериев (2.18) и (2.19), можно корректно подойти к количественной оценке качества несемантической информации, сформулировав следующую аксиому [3], [95], [96].

Аксиома 4 – зависимости качества информации от информационного риска. Принцип двухэкстремального кодирования в уравнении связи предполагает минимизацию качественной информации, вследствие чего в совокупность условий в уравнение должно войти *количественное ограничение качества информации*.

Интуитивно ясно, что качество синтаксической информации должно быть тем больше, чем больше входное отношение сигнал/шум и чем меньше информационный риск. Как и в теории оценивания, качество решения представляют в виде *нормированного среднего риска*, т. е. *отношения дисперсии оцениваемой величины к среднеквадратической ошибке* [88, т. 1], также качество переданной по каналу связи информации должно определяться *нормированным информационным риском* или обратной величиной. Качество переданной через канал связи информации о зашумлённом сигнале следует определять через отношение к информационному риску количества потенциальной доминантной информации $I(x, y)$.

Аксиома зависимости качества информации от информационного риска может быть сформулирована так: *качество информации определяется отношением количества потенциальной доминантной информации к информационному риску*.

Следствием из неё является то, что при кодировании случайных гауссовских величин (одномерных сигналов) при вариации ошибки передачи *существует максимум качества информации*, определяемый минимумом информационного риска. Понятие качества в радиотехнике применяется сто лет, им характеризовали добротность контуров и обозначали символом Q – первой буквой английского слова *quality*, т. е. качество. Мы будем следовать этой традиции и обозначать качество информации буквой Q. Этим мы подчёркиваем, что старинное понятие качества колебательного контура, отвечавшего за *избирательность* радиоприёмника (по соседнему каналу, а при гетеродинном приёме – ещё и по зеркальному каналу), было самым первым интуитивным приближением к понятию доминантной информации и её качеству, связанному и с усилением принимаемого полезного радиосигнала, и с «просачиванием» на выход системы радиосвязи сигнала мешающей радиостанции. Поэтому название «качество» колебательного контура, ответственного за спектральную фильтрацию для усиления доминантного сигнала и торможение мешающего сигнала, лучше соответствует информационной парадигме связи, чем название «добротность».

Подчеркнём, что *качество информации* в приложении к телевизионным сигналам является *наиболее общим*, обобщающим все частные искажения, рассмотренные ранее [26], *охватывающим все возможные причины потери информации* в оптическом звене, фотоприёмнике, видеотракте, цифровых кодерах и каналах связи. Это могут быть и линейные, и нелинейные искажения, с топологическими искажениями и без них. Важно, что введённое определение качества *кроме часто обсуждавшейся потери информации* [26], [41], [44], [49] включает в себя *все компоненты ложной информации – и фоновую, и шумовую информацию*. Качество информации в общем случае, а также для кодирования случайных величин примет вид

$$Q = \frac{I(x, y)}{R_I} = \frac{I(x, y)}{c_0 \Delta I + c_1 (I_{\Phi} + I_{\text{ш}})}; \quad Q = \frac{I(x, y)}{c_0 \Delta I + c_1 I_{\text{ш}}}. \quad (2.28)$$

При кодировании случайных процессов даже при упрощённом критерии экономного связиста и учёте только шумовой информации качество информации вычисляется несколько сложнее, оно может быть записано в трёх эквивалентных формах: с группировкой компонентов риска по при-

знаку отнесения к потере информации, по признаку отнесения к пропускаемым или усекаемым спектральным компонентам [96]:

$$Q = \frac{\sum_{k=1}^{\infty} I(x_k, y_k)}{\left[\sum_{k=1}^K \Delta I_k + \sum_{k=K+1}^{\infty} I(x_k, y_k) \right] + \sum_{k=1}^K I_{\text{шк}}}; \quad Q = \frac{\sum_{k=1}^{\infty} I(x_k, y_k)}{\sum_{k=1}^K [\Delta I_k + I_{\text{шк}}] + \sum_{k=K+1}^{\infty} I(x_k, y_k)};$$

$$Q = \frac{\sum_{k=1}^{\infty} I(x_k, y_k)}{\sum_{k=1}^{\infty} I(x_k, y_k) + \sum_{k=1}^K [\Delta I_k + I_{\text{шк}} - I(x_k, y_k)]}. \quad (2.29)$$

Исходная смесь доминантного сигнала и шума (стимул) при бесконечных спектрах этих случайных процессов, когда в определении (2.29) $K \rightarrow \infty$, в силу конечного количества доминантной информации и бесконечного количества шумовой информации характеризуется *нулевым качеством информации*.

Чем больше отношение сигнал/шум на входе, тем больше количество и качество информации у получателя. Потенциальная возможность достижения предельного качества (бесконечного в смысле (2.29)) имеется только у сигналов, выбираемых из малого дискретного множества (например сигналы светофора). В (2.29) для качества информации входит значение граничной полосы пропускания кодирующего фильтра, которое вместе с формой спектральной характеристики кодера является главным следствием введённой аксиомы зависимости качества информации от информационного риска.

2.3. Оптимальное усечение полосы частот сигналов

Сигналы с ограниченным спектром были введены в теорию связи В. А. Котельниковым и К. Шенноном, без обоснования, но справедливо полагавшими, что спектр «каким-нибудь образом» усечён. Мы с использованием определения (2.29) покажем, что такое *усечение спектра – следствие максимизации качества информации*. При фиксированных параметрах доминантного сигнала и входного шума качество информации определяется информационным риском (в форме критерия экономного связиста (2.19)), поэтому для нахождения кодирующего фильтра, обеспечивающего максимум качества информации, достаточно найти минимум знаменателя

выражения для качества информации (2.29). Для нахождения оптимального кодирующего фильтра учтём, что в качестве декодера в системе связи используется фильтр Винера (2.1) с конечной полосой пропускания K , в выражение для которого входит искомый кодирующий фильтр, а ошибка, вносимая шумом канала, вычисляется при этом по (2.3). Тогда выражение для информационного риска может быть приведено к виду

$$R_I = \frac{1}{2} \sum_{k=1}^{\infty} \log(\psi_{1k}^2 + 1) - \frac{1}{2} \sum_{k=1}^K \log \frac{[\hat{\lambda}_k g_k^2 + \sigma_2^2(\psi_{1k}^2 + 1)]^2}{\sigma_2^2(\psi_{1k}^2 + 1)(\hat{\lambda}_k g_k^2 + \sigma_2^2)}. \quad (2.30)$$

К нахождению оптимального фильтра на основе (2.30) можно подойти двумя способами. Первый состоит в прямом нахождении оптимального кодера дифференцированием и приравниванием к нулю производной $dR_I/dg_k^2 = 0$. Однако такой фильтр, доставляющий максимум качества информации, может не обеспечивать информационного равновесия. Поэтому второй вариант решения задачи поиска оптимального фильтра состоит в нахождении условного максимума качества, которое должно осуществляться решением уравнения связи аналогично записи (1.5) с тем различием, что при поиске энтальпии ограничением является не задаваемая средне-квадратическая ошибка, а максимум качества информации согласно (2.29).

Решение первой задачи даёт оптимальный фильтр:

$$g_{kQ}^2 = \sigma_2^2 \left(\frac{1}{\varepsilon_{Bk}} - \frac{1}{\hat{\lambda}_k} \right) = \frac{\sigma_2^2}{\hat{\lambda}_k} (\psi_{1k}^2 - 1). \quad (2.31)$$

Важнейшей характеристикой оптимального кодирующего фильтра (2.31) является ограниченность передаваемой полосы частот $k \leq K$: в силу неотрицательности квадрата спектральной характеристики её усечение производится на частоте K , на которой спектральное отношение сигнал/(входной) шум (2.2) принимает пороговое значение $\psi_{1k}^2 = 1$. Вообще говоря, следуя интуитивным представлениям, такое усечение полосы передаваемых значений и должно было усекать спектральные компоненты с малым качеством информации, т. е. с малым отношением сигнал/шум. Однако (2.31) даёт *оптимальное* усечение, поэтому такое граничное значение будем обозначать как $K_0 = \arg \max Q$. Для упрощения иллюстрации наличия максимума качества информации на рис. 2.7 построены зависимости переданной доминантной информации и качества доставленной получателю информации для передачи сигнала со спектром Баттерворта 1-го порядка (т. е. имеющего

экспоненциальную автокорреляционную функцию и обладающего максимальной энтропией) в отсутствие искажающего фильтра B_k .

Рис. 2.7 показывает, что сокращение полосы частот относительно K_0 ведёт к снижению качества информации из-за излишних потерь доминантной информации, а расширение полосы частот снижает качество информации из-за загрузки канала и получателя шумовой информацией, что при $K \rightarrow \infty$ ведёт к нулевому качеству информации.

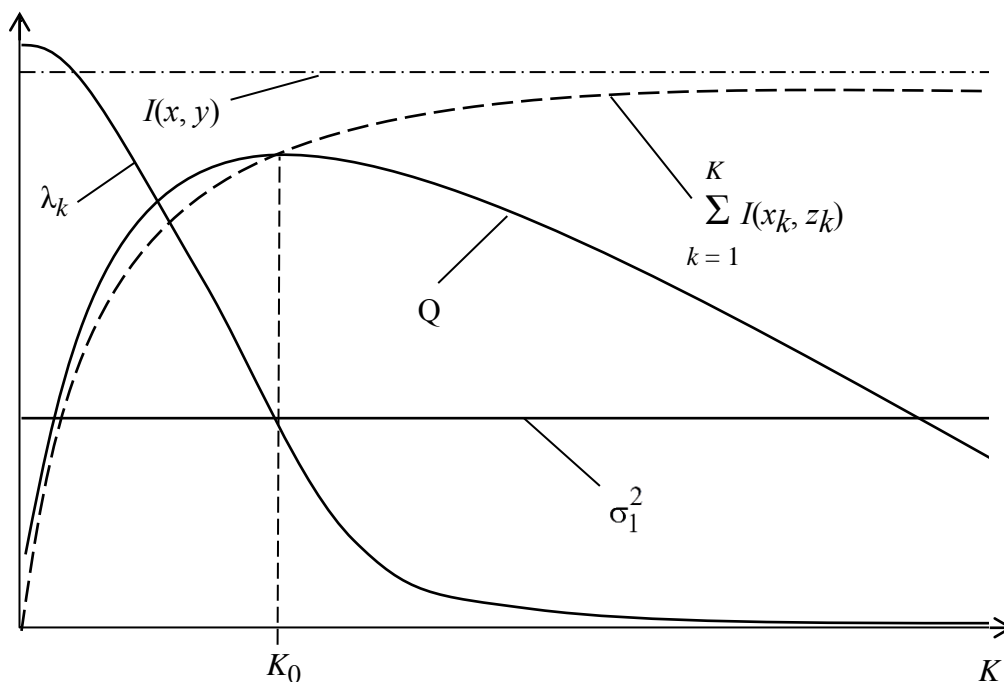


Рис. 2.7. Зависимость качества информации от граничной частоты пропускания кодировочного фильтра

При опоре на критерий минимума информационного риска в виде (2.18) с весовыми коэффициентами при потере доминантной информации и при шумовой информации порог усечения изменяется в соответствии с (2.25). Это может привести к рассмотренной на с. 71 ситуации, когда наблюдатель расценивает потерю доминантной информации выше, чем шумовую информацию, и может предпочесть более зашумлённое изображение.

Отмеченная ранее неразрывная слитность доминанты и фона проявляется в том, что даже при максимизации качества информации доставленный получателю код z не только не доносит всю доминантную информацию, но и содержит некоторое неизбежное количество шумовой информации.

Вторая задача, в полной мере решаемая с помощью уравнения связи аналогично (1.8), может быть упрощена, если воспользоваться уже най-

денным решением для кодирующего фильтра, обеспечивающего информационное равновесие при ограничении широкополосности кодера (1.6). При этом отличие записи спектральной характеристики в рассматриваемом случае от простых формул (1.9), (1.10) состоит в замене дисперсий спектральных отсчётов доминантного сигнала $\{\lambda_k\}$ на дисперсии отсчётов его оптимальной оценки $\{\hat{\lambda}_k\}$:

$$g_k^2 = a^2 - \frac{\sigma_2^2}{\hat{\lambda}_k}.$$

Подстановка этой спектральной характеристики в формулу для информационного риска (2.30) позволяет дифференцированием и приравниванием к нулю производной $dR_I/da^2 = 0$ найти оптимальный кодирующий фильтр:

$$g_{kQ,E}^2 = \sigma_2^2 \left(\frac{1}{\varepsilon_{BK}} - \frac{1}{\hat{\lambda}_k} \right). \quad (2.32)$$

Этот кодирующий фильтр по форме записи близок к кодирующему фильтру (2.31), обеспечивающему безусловный максимум качества информации. Отличие состоит в том, что первый член в скобках включает не переменное значение ошибки фильтра Винера, а её значение на граничной частоте K . Это означает, что усиление выбирается компромиссно, максимум качества обеспечивается не для всех спектральных компонентов, а только на границе полосы пропускания, вблизи K_0 . Фильтр (2.32), как и кодирующий фильтр (2.31), имеет финитную полосу пропускания (см. рис. 2.7), определяемую по пороговому входному отношению сигнал/шум.

На рис. 2.8 представлены относительные спектральные характеристики для передачи сигнала со спектром Баттерворта 1-го порядка в отсутствие искажающего фильтра B_k . Абсолютные значения спектральных характеристик, как и ширина полосы пропускаемых частот K_0 , зависят от входного отношения сигнал/шум. Найденный фильтр (2.32), обеспечивающий условный максимум качества информации, показывает, что при идеальном линейном кодировании *зашумлённых* сигналов ($H_\varepsilon = C_E$) *назначаемая ошибка ε_0 при вычислении энтальпии (2.5) и ограничение широкополосности («площади усиления») кодера E не могут назначаться произвольно. Чем больше входное отношение сигнал/шум, тем ббльшая*

требуется широкополосность кодера, обеспечивающая максимальное

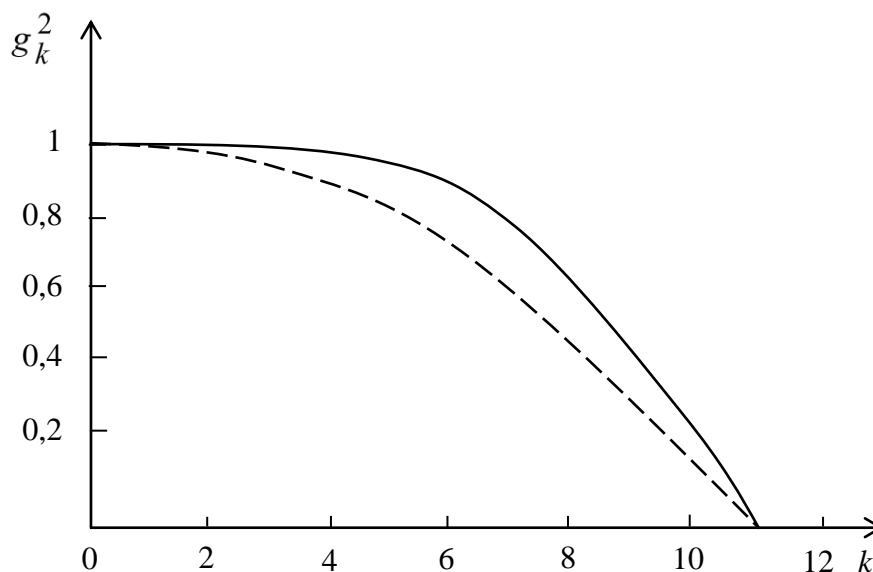


Рис. 2.8. Относительные спектральные характеристики оптимальных кодирующих фильтров, обеспечивающих безусловный (сплошная линия, (2.31)) и условный (пунктирная линия, (2.32)) максимум качества информации

качество переданной информации.

Значение найденных спектральных характеристик для твёрдотельного телевидения определяется тем, что **они должны служить эталоном** при оценке частотно-контрастных характеристик (ЧКХ) телекамер, и в первую очередь – ЧКХ фотоприёмных матриц (см. гл. 3).

Вторым аспектом значимости найденного *оптимального усечения* спектра доминантного сигнала является ещё одно обоснование *оптимального накопления* сигнала, которое соответствует получению из смеси сигнала и входного шума конечного числа отсчётов. То, что в теории речь идёт об отсчётах разложения Карунена–Лоэва, а на практике – о пикселах, не является принципиальным. Так, при возможности аппроксимации автокорреляционной функции сигнала кусочно-линейной функции с шагом, равным размеру пиксела, результат – количество отсчётов – в этих двух вариантах оказывается одинаковым.

В общем случае усечение должно производиться по *пороговому отношению сигнал/шум*. Для критерия минимума информационного риска (2.18) $\psi_{\text{пор}} = c_1/c_0$, для критерия экономного связиста (2.19) $\psi_{\text{пор}} = 1$.

Это означает, что введение понятия шумовой информации дало новую информационную трактовку широко применяемого в теории решений по-

нения «пороговое отношение сигнал/шум»: *порог определяется назначаемым проектировщиком отношением цен шумовой информации и потери доминантной информации.*

Применение принципа доминантной информации, ведущего к учёту качества информации при кодировании зашумлённых сигналов, даёт обоснование *оптимального усечения полосы частот непрерывных сигналов* и конкретные правила для вычисления характеристик кодирующих и декодирующих фильтров. В рамках теории связи эти результаты формализуют фундаментальные закономерности *рождения и передачи* сигналов в реальном времени. В частности, разработанная теория впервые дала формализованное обоснование использованию в теории связи функций с финитным спектром, введённых в теорию связи В. А. Котельниковым и К. Шенноном, но не как априорной данности, а как плода *оптимального кодирования.*

Описанное оптимальное усечение полосы частот при наблюдении доминантного сигнала на фоне шума легко распространяется на проблему *избирательности* систем связи [95], [96]. Для этого достаточно ввести в рассмотрение, кроме спектра доминантного сигнала, ещё и спектры мешающих сигналов (в телевизионном и радиовещании принят термин «избирательность по соседнему каналу») и применить к ним теорию идеального линейного кодирования с учётом максимума качества информации.

2.4. Принцип доминантной информации, накопление качества информации и слитность изображений

Применение принципа доминантной информации обосновало усечение полосы частот непрерывных сигналов с помощью оптимальных кодирующих фильтров. Однако сокращение полосы частот всегда считалось атрибутом накопления – как накопители трактовались и фильтр Винера, и фильтр Норта (согласованный фильтр). Поэтому оптимальное усечение спектра случайных процессов можно трактовать как инструмент, с помощью которого доминантная теория информации формализует накопление качества информации. Вместе с тем для полноценной реализации накопления качества информации оптимального усечения полосы частот сигнала не достаточно, так как сформированный непрерывный код хорош только в рамках рассмотренной двухкаскадной модели, с точки зрения дальнейшей передачи он тоже обладает бесконечной кодовой информацией. Для того чтобы, по выражению Г. Вейля [101], «запрудить поток бесконечного, грозящий затопить в своём течении наш дух», необходимо провести кван-

тование сигнала, подход к которому уже просматривался при определении порогового значения отношения сигнал/шум при оптимальном усечении спектра сигнала. Это пороговое отношение сигнал/шум может трактоваться как *макроквант*. Он не равен биту и вообще не является «атомом» информации. Так как мы исходили из того, что информация – атрибут не только материи, но и духа, то и макроквант информации становится зависим от воли создателя системы: размер макрокванта является функцией не только уровня входного шума, но и отношения назначенных проектировщиком цен потери доминантной информации и шумовой информации (вектора концепции системы).

Количество информации, переносимое макроквантом не однозначно (зависит от цели) и концептуально восходит к софизму Зенона о куче зерна. Макроквант информации – не просто «куча» фотонов, а мера качества информации. *«Куча» фотонов (электронов) составляет макроквант, «куча» макроквантов – пиксел изображения, а «куча» пикселов – образ.* Чем больше макроквант количественно, тем меньше шумовая информация и выше качество образов, их различие между собой, но тем меньше слитность изображений и тем меньшее количество образов может быть сформировано из доступной синтаксической информации. Размеры макрокванта согласно теории доминантной информации зависят от цели системы и вытекающего из неё понятия качества информации. Поэтому макроквант информации оказался полезным для более глубокого понимания принципа накопления – этапа кодирования видеоинформации непосредственно в телекамере.

В телевидении вопрос накопления в фотоприёмниках (сначала в электронно-лучевых трубках, затем в твёрдотельных матрицах) обсуждался давно и активно; принцип накопления качества информации, выросший из принципа накопления заряда, подробно обсуждён во введении к нашей книге. Мы развиваем тезис о том, что накопление в фотоприёмнике имеет только *вид* накопления заряда, а по *сути* своей является *накоплением качества видеоинформации*. Именно накопление качества информации (благодаря квантованию информации на макрокванты) обеспечивает *максимально возможное насыщение доступной информации доминантой за счёт разрушения шумовой информации*.

Важнейшим примером реализации накопления качества информации является групповой счёт фотонов (фотоэлектронов). Идея группового счёта фотонов обсуждалась в телевидении давно, но только на втором этапе

твёрдотельной революции, при реализации цифрового пиксела, которому будет уделено достойное внимание в гл. 3, групповой счёт приобрёл максимальную наглядность.

Размеры макрокванта можно определять по его энергии (по его значению квантуется энергетический спектр сигнала), или, как это имеет место при построении цифровых пикселей фотоприёмников (см. гл. 3), его размеры могут определяться в числе фотогенерированных электронов.

Развитую в 1.3.1 концепцию квантования как оптимизацию смешанной системы здесь необходимо уточнить с тем, что в смешанных системах передачи зашумлённых сигналов бывает актуально выбрать оптимальный шаг квантования даже в отсутствие шума канала. Для этой модели в силу отличия функции распределения от гауссовской анализ проводится с учётом как дисперсии ошибки квантования $D_{\text{КВ}} = \delta^2/12$ с шагом δ , так и её энтропийной мощности $D_{\text{КВ}} H = \delta^2/2\pi e$ [16], [48]. Чаще для определения подходящего размера «кучи» фотоэлектронов опираются на обычную мощность шума $D_{\text{КВ}} = \delta^2/12$, полагая при этом постоянство уровня шума во всём диапазоне полезного сигнала [87], [104]. Важнейшим результатом такого энергетического подхода к квантованию является предложение считать *«систему уравновешенной, если уровень шума квантования равен среднеквадратической сумме всех шумовых сигналов любого происхождения, генерируемых в системе до квантования»* [87]. С позиций теории информации необходимы комментарии к этой гипотезе.

Во-первых, Шенноном дана специальная теорема о пропускной способности канала при произвольной помехе [16]. Согласно этой теореме при произвольной функции распределения помехи этот важнейший показатель системы связи характеризуется не равенством, а двумя неравенствами, в которые входят отношения мощности сигнала к энтропийной мощности помехи и отношение обычной мощности помехи к её энтропийной мощности. Так как энтропийная мощность любой помехи меньше её мощности, то для изобретателей различных способов кодирования появляется некоторый резерв для повышения пропускной способности (помехоустойчивости передачи). Поэтому равенство дисперсий ошибки квантования и шума не обеспечивает информационного равновесия, оставаясь гипотезой в рамках энергетической парадигмы.

Во-вторых, в рамках принятой модели квантования смеси сигнала с аддитивным гауссовским шумом декларация об «уравновешенности» си-

стемы является интуитивной гипотезой. Для её законного применения необходимо обращение к критерию минимума информационного риска (2.18) или критерию экономного связиста (2.19). Аналогично записи информационного риска (2.20) можно получить решение, по смыслу эквивалентное решению (2.21): влияние ошибок из-за входного шума и шума канала должно быть одинаковым. В первом приближении этот аналитический результат обосновывает гипотезу [87] об уравниваемости системы, подтверждая, что если опираться на обычную мощность ошибки квантования $D_{\text{кв}} = \delta^2/12$, то минимум информационного риска в критерии экономного связиста (2.19) достигается при её равенстве ошибке винеровской фильтрации ε_B . Лучшим приближением к информационному равновесию является равенство винеровской ошибки и энтропийной мощности ошибки квантования $D_{\text{кв}} H = \delta^2/2\pi e$. Это соответствует оптимальному значению шага квантования, большому среднеквадратической ошибки σ из-за входного шума не в $\sqrt{12} \approx 3,56$ раз [87], а в $\sqrt{2\pi e} \approx 4,13$ раз. Это позволяет считать, что все три предложенных в [87] варианта соотношения шага квантования и среднеквадратического значения шума соответствуют различным соотношением потери полезной информации и шумовой информации в критерии минимума информационного риска (2.18). Так, отношение $\delta/\sigma = 3,56$ соответствует отношению цен в критерии (2.18) $c_1/c_0 \approx 0,86$.

Определение размера группы фотоэлектронов с учётом мультипликативного фотонного шума и её связь с пороговым контрастом цифрового пиксела рассмотрены в гл. 3.

В любом случае размеры макрокванта зависят от цели системы, весьма наглядным примером применения макрокванта информации является синтез оптимальной системы обнаружения–оценивания в телевизионной астрономии (см. гл. 4). В такой системе фотоэлектроны группируются в макрокванты («кучи») *с переменными параметрами*: с увеличением интенсивности сигнала объекта группы фотоэлектронов укрупняются, кроме этого геометрия оптимальной зоны накопления изменяется в зависимости от интенсивности как доминантного сигнала, так мешающего фона. Это означает, что важным лозунгом (парадигмой) принципа доминантной информации является утверждение В. Гейзенберга [22]: **«Наше знание, несомненно, изменяется прерывно».**

Эта прерывность нарастания знания вступает в определённое противоречие с концепцией непрерывности, введённой в науку Лейбницем с це-

лью объяснения явлений природы. Хотя Лейбниц выдвинул принцип слитности для анализа природы, т. е. строил физическую теорию, идею непрерывности развивали и обосновывали не физики, а многие выдающиеся математики (практически на этот принцип опирались, не формализуя, до Лейбница – Кеплер, одновременно с ним – Ньютон).

Далее будем рассматривать прикладное телевидение с позиций не чисто математического понятия непрерывности, а с позиций его более типичного для технических наук понятия слитности, которой придаётся большое значение в современном вещательном телевидении и при создании компьютерных мониторов. Вместе с тем *проблема пространственно-временной слитности изображений выходит за рамки создания иллюзии у зрителя*. Она актуальна и имеет важные научные аспекты в научно-прикладном телевидении, где получателя – компьютер и/или профессионала-дешифровщика, имеющего время и технические возможности для детального анализа (пространственная и временная «лупа») – можно считать «идеальным».

Математическая идея непрерывности, ведущая к понятию бесконечно малой величины, в физике имеет ограниченное применение, так как нулевая погрешность измерений требует бесконечных затрат энергии. При этом отмечается, что *«нужно сделать упор на фундаментальное различие между математическими и физическими теориями»* и что следует *«особо подчеркнуть невозможность физического определения континуума в пространстве и времени»* [102].

В телевидении иллюзия слитности интерпретировалась отдельно для пространства и времени. Так, В. К. Зворыкин писал [17]: *«Количество информации, воспринимаемое наблюдателем, ограничивается двумя физиологическими характеристиками глаза: конечной разрешающей способностью глаза и инерционностью зрительного восприятия <...> в результате инерционности зрительного восприятия создаётся впечатление непрерывного движения <...> Глаз не ощущает разрывности движения при частоте около 15 периодов в секунду, однако мелькания оказываются заметными для глаза при гораздо большей частоте»*.

Здесь в явном виде говорится об **иллюзии** слитности, т. е. **достаточности** некоторой совокупности дискретных данных для того, чтобы считать сигнал слитным: изображение состоит из дискретных кадров и дискретных строк (в твёрдотельном телевидении – и элементов), но зритель,

владелец большого числа ограничений своего зрения, *принимает* это изображение за слитное как в пространстве, так и по времени. В кино и телевизионном вещании требуемую частоту кадров и чёткость определяли эмпирически и привели её к частоте, на которой не заметны ни скачки движения, ни пульсации яркости равномерного фона. Иллюзия слитности изображения, строящегося в мозгу человека по дискретному множеству отсчётов, связана не только с отмеченной В. К. Зворыкиным конечной разрешающей способностью зрительного анализатора в пространстве и времени, но и с отмеченной Анри Пуанкаре *«верой в непрерывность, которую трудно было бы оправдать убедительным рассуждением, но без которой никакая наука не была бы возможна»*.

Шагом к отходу от упрощённого понимания слитности телевизионных сигналов стала идея С. И. Катаева о сокращении полосы частот при передаче малоподвижных изображений. Сущность этого метода, названного малокадровым, трактовалась как накопление и передача только тех изображений, которые существенно отличаются по смысловому содержанию [45]. Поэтому малокадровый метод заключается в обеспечении **семантической слитности изображений**.

Наличие разрывов, т. е. нарушение слитности, неизбежно при переходах от пиксела к пикселу, от строки к строке и от кадра к кадру. Слитность сигнала вдоль строки в электронно-лучевом телевидении не отменяет общей проблемы нарушения слитности, которая обусловлена обязательным переходом от многомерного сигнала (как функции пространства и времени) к одномерному как функции времени (см. цитату из [16] на с. 12).

Применение принципа слитности обусловлено стремлением к *экономной модели*, когда явление природы описывается аналитической функцией, имеющей бесконечное число точек, но определяемой *конечным числом параметров* – чем меньше параметров, тем лучше. Иллюзию слитности можно понимать как приложение задачи табулирования аналитических функций [23]. Там центральная идея – представление аналитических функций дискретными рядами их отсчётов с таким интервалом, чтобы внутри него *с заданной точностью* функция *интерполировалась прямой линией*. Отсюда, с одной стороны, связь идеи слитности табулирования с эпсилон-энтропией случайных процессов [23], с другой – зависимость количества выборок функции от требуемой точности. Например, долгое время инженеры использовали пятизначные таблицы.

Принцип Гельвеция «знание немногих принципов заменяет знание многих фактов» относится к этой же идее экономии памяти при создании образов. Именно экономная модель (*компактный образ*) делает науку способной к развитию. Нельзя бесконечно усложнять модель, и Н. Винер предупреждал, что окончательной моделью кошки может быть только кошка. Р. Беллман уточнял: чрезмерное упрощение модели вредно, и моделирование – это искусство прохождение между болотом переусложнения и капканом переупрощения.

Набор чисел, структурирующих макрокванты, формируемый телекамерой, – это тоже модель. При этом число двоичных единиц, характеризующих изображение, на порядки меньше числа фотонов, участвовавших в построении этой модели. Стремление к сжатию видеoinформации в цифровых кодерах источника, включаемых в состав телевизионной системы, тоже отражает потребность в экономной модели. Сколько можно сэкономить при построении изображений, какие группы порождённых светом электронов следует формировать в прикладных телевизионно-компьютерных системах? Ответ зависит от цели системы. Если система предназначена для поиска нового в природе, то макрокванты должны формироваться по уровню шума.

Вырождение сигналов. Так как В. А. Котельников и К. Шеннон излагали теорию связи, исходя из необоснованного предположения о том, что сигналы имеют ограниченную полосу частот, создатели систем связи вынуждены были обратиться не к адекватной теории, а к аналогии для дискретизации в виде модуляции [103]. Подчеркнём, что подход к вырождению числа степеней свободы с позиций теоремы отсчётов в телевидении не корректен по двум принципиальным причинам. Во-первых, реальные *входные сигналы* на конечном интервале поля зрения и времени наблюдения принципиально имеют *неограниченный спектр* пространственных и временных частот. Во-вторых, ключевой вывод из теоремы отсчётов о желательности взятия выборок сигнала с помощью дельта-функции ведёт к нулевым площади накопления (площади пиксела) и времени накопления. Это означает нулевую чувствительность телевизионной системы.

Подход к дискретизации как к модуляции долго использовался и используется многими проектировщиками телевизионных систем (например [46], [47] [104]), хотя известен более корректный для этой цели математический аппарат – теория *вырожденных операторов* [69]. Здесь вырождение –

это переход от бесконечномерного континуального сигнала к конечномерному, число измерений которого равно числу дискретных отсчётов. При этом вырожденным оператором отсчёты формируются с использованием операции *интегрирования* произведений сигнала и набора базисных функций, т. е. накопления, столь важного для обеспечения чувствительности телекамеры.

Вырождение ведёт к ошибкам усечения спектра на частоте Найквиста, обратной удвоенному интервалу дискретизации Δ , и возникновению побочных спектров (в англоязычной литературе – *aliasing*). Однако традиционные модели [42], [46] абстрагируются от влияния шумов. Основанием для корректного подхода к дискретному представлению непрерывных зашумлённых изображений является понятие качества информации (2.29), определяющее *оптимальное вырождение*, т. е. *усечение полосы частот с граничной частотой K_0* . Такой выбор граничной частоты сигнала означает, что ошибки усечения спектра, а значит, и ошибки из-за возникновения побочных спектров «тонут» в шуме (рис. 2.9).

Оптимальное вырождение по частоте K_0 имеет аналогию в теории линейной фильтрации в виде компромисса между линейной и шумовой ошиб-

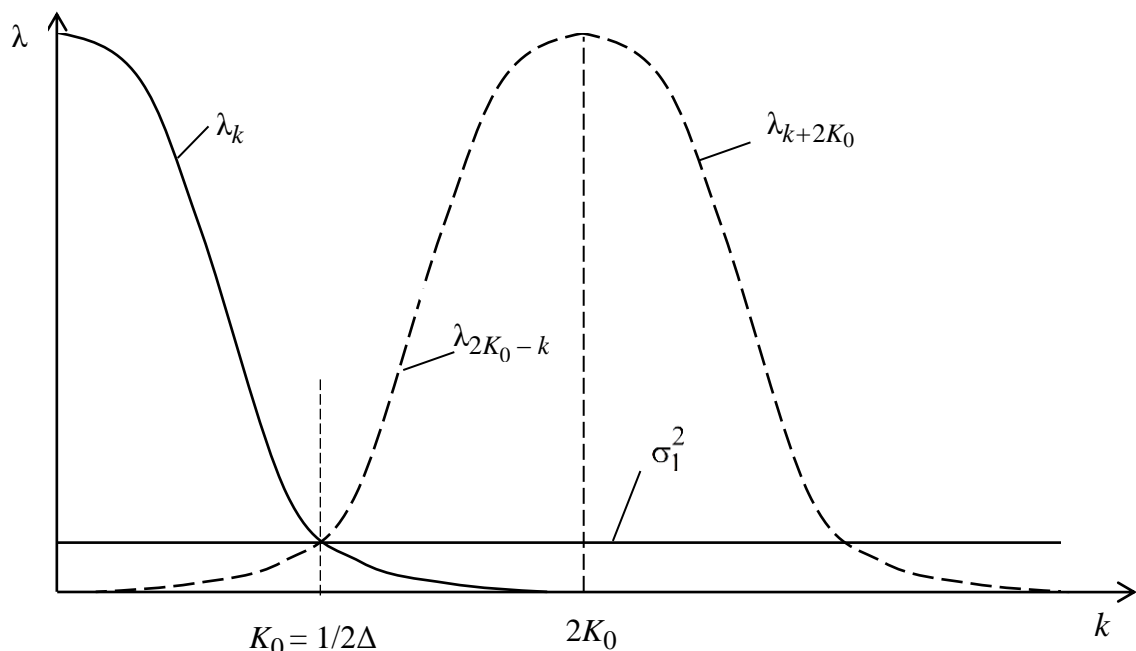


Рис. 2.9. Маскирование шумом ошибок дискретизации, возникающих из-за появления побочных спектров, обеспечивающее максимум качества информации при дискретизации; граничная частота $K_0 = 1/2\Delta$ – частота Найквиста; показан побочный спектр только первого порядка

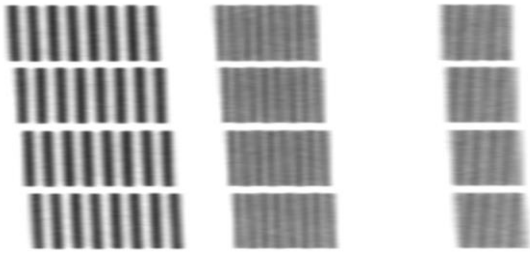


Рис. 2.10. Изображения штриховых мир с различными пространственными частотами и фазами

ками в формуле Бодэ–Шеннона [16]: максимум качества информации концептуально соответствует минимуму вреда от линейных ошибок дискретизации (включающих потерю доминантной информации в полосе пропускания и из-за усечения спектра сигнала на частоте K_0) и шумовых

ошибок. Важным аспектом такого минимума является **отказ от старой концепции существования «оригинала»** изображения (подробно иллюстрированной В. К. Зворыкиным [17]), согласно которой чем меньше размер пиксела, тем лучше.

Здесь уместно подчеркнуть, что учёт ложной информации, возникающей из-за появления побочных спектров сигнала при дискретизации, и включение спектра этой ложной информации в формулы (2.29) и (2.30) ведёт к снижению качества информации и увеличению информационного риска, но *не влияет на спектральные характеристики оптимальных фильтров* (2.31) и (2.32).

Важным проявлением усечения спектра сигнала является подробно исследованный и хорошо известный не только в телевидении, но и в полиграфии эффект *муара* (например [42], [46], [104]), означающий появление комбинационных частот испытательного сигнала и растровой структуры фотоприёмника и/или монитора. Муаровая картина зависит от фазы периодического входного сигнала относительно дискретной структуры растра, что означает неинвариантность отклика на выходе вырожденного оператора к сдвигу входного воздействия [2]. Неинвариантность к сдвигу входного сигнала многолика. Это вариации не только муаровой картины (рис. 2.10: в каждой группе по 8 штрихов; каждый последующий ряд сдвинут относительно предыдущего на четверть периода; в левом столбце муара нет; в среднем столбце – пространственная частота на границе возникновения муара; в правом столбце фаза муарового сигнала зависит от фазы испытательной миры), но и отклика от точечного объекта (изображение звезды может попасть и в центр пиксела, и на перекрестье между пикселами), частотно-контрастной характеристики при изменении фазы штриховой миры относительно дискретной структуры пикселов.

Наличие муара в изображениях, формируемых системами научно-прикладного телевидения, можно трактовать как несогласованность параметров дискретизации с априорной информацией о входном воздействии. При этом весьма неконкретен подход, идущий от работы Оливера, Пирса и Шеннона «Принципы импульсно-кодовой модуляции» [16]: будто бы надо сначала исказить сигнал усечением спектра, а лишь затем дискретизировать.

Вопрос выбора граничной частоты без учёта шума при этом оставался открытым. Волонтаризм в этом вопросе доходит до рекомендаций вида «обычно она <частота дискретизации> выбирается в 5–12 или более раз выше эффективной полосы частот» [105]. Однако для получения информации наилучшего качества нельзя волонтаристски усекать спектр сигнала, а следует (насколько позволяет априорная информация) выбирать частоту дискретизации сигнала вырожденными операторами с учётом входного спектрального отношения сигнал/шум.

Вырождение сигналов связано с определением параметров разложения изображения: «Выбор параметров разложения является наиболее ответственной операцией в расчёте оптимальных телевизионных систем» [44], но он должен вестись не на основе «учёта потери пространственной и временной информации» [44], а на основе максимизации качества видеоинформации, в частности обеспечения достаточной слитности. Примеры корректного выбора параметров разложения приведены в гл. 5.

Достаточность слитности – маскирование ошибок дискретизации шумом. Развитие прикладного телевидения, пережившего твёрдотельную революцию и ставшего разделом видеоинформатики, привело к технологическому и методологическому слиянию телевизионной и компьютерной техники [3]. В духе этого методологического слияния следует подойти к оценке слитности изображения аналогично оценке качества искусственного интеллекта согласно тезису Тьюринга: если человек, задавая вопросы, по ответам не может обнаружить, что его собеседником является не человек, а компьютерная программа, то эта программа олицетворяет искусственный интеллект. Здесь уместно такое сравнение: известно различие в визуальном восприятии геометрических фигур (окружностей и прямоугольников со сторонами, не параллельными растру), синтезированных компьютером, и переданных телекамерой изображений этих же реальных фигур. Первые при внимательном рассмотрении воспринимаются как имеющие разрывы, вторые в силу влияния размытия линий объективом и

шума обычно воспринимаются (как телезрителем, так и профессионалом-дешифровщиком) как слитные. Конечно, в ряде экспериментов при наблюдении специальных тестовых сигналов эффекты дискретизации в телекамере обнаруживаются в виде муара. Тезис Тьюринга можно трактовать как развитие известного операционального принципа Бриджмена [102], согласно которому для определения физической величины необходимо формализовать процедуру её измерения. Можно связать слитность изображения с *измерением* (с проверкой статистических гипотез, т. е. с обнаружением и различением) эффектов, вызванных дискретизацией изображения по полю и по времени. При этом *оптимальное усечение* полосы частот обеспечивает маскирование шумом ошибок дискретизации (вырождения), что *обобщает* формулировку В. К. Зворыкина о незаметности для глаза ошибок дискретизации. Кстати, в оптике известно правило, что для принятия решения о том, что данная оптическая система *разрешает* данную штриховую мишу с данной пространственной частотой, необходимо не только обнаружить провалы между линиями миры, но и правильно определить количество штрихов в ней. При обнаружении провалов между штрихами, но неправильном определении числа штрихов разрешение называют *ложным*. Поэтому в отношении слитности можно выдвинуть следующий тезис: *если наблюдатель (человек или компьютер) из-за шума не может определить, сформирован предъявленный ему сигнал телекамерой из непрерывного изображения или выбран из дискретного множества существенно меньшей мощности, то этот сигнал признаётся слитным (в пространстве и/или времени).*

Синтез прикладной телевизионно-компьютерной системы должен опираться на этот тезис, который упрощённо выглядит так: *следует так выбирать размеры макроквантов, на которые разбивается входной оптический сигнал в процессе накопления информации, чтобы изображение было неотличимо от слитного, т. е. следует стремиться к маскированию шумом ошибок дискретизации.*

Рассмотрение дискретизации сигнала в твёрдотельных фотоприёмниках с учётом накопления информации для борьбы с входным шумом позволяет найти наглядное условие для оптимальной дискретизации и без перехода в спектральную область. Это условие, устанавливающее оптимальное соотношение между математическими ожиданиями приращения полезного сигнала $S(t)$ на интервале дискретизации Δ и ошибок ε (суммы

шумовых и линейных ошибок), прямо трактовалось [1] как **условие практической слитности** дискретного сигнала, наблюдаемого компьютерным получателем:

$$M\{[S(t) - S(t - \Delta)]^2\} = \mu\varepsilon,$$

где μ – численный коэффициент, имеющий смысл порогового отношения сигнал/шум и зависящий от степени n интерполирующего полинома (сплайна), при $n = 0$ $\mu \approx 2$; при $n = 1$ $\mu \approx 2,4$; при $n \rightarrow \infty$ $\mu \rightarrow e$. Этому же числу e (основание натуральных логарифмов) равно отношение математических ожиданий дисперсии приращений сигнала на интервале эквивалентной ширины импульсной характеристики *непрерывного* фильтра Винера и ошибки оценивания на его выходе.

Теоретико-информационный подход к системам прикладного телевидения приводит к отходу от прежних взглядов на системы наблюдения подвижных изображений. На смену старой парадигмы *пассивного учёта* снижения разрешающей способности из-за движения изображения [40] пришёл *принцип слитности* телевизионного изображения в прикладных системах, отличающийся от подхода в вещании, опирающегося лишь на инерционность зрения. Подход, базирующийся на определяющем влиянии входного (фотонного) шума изображения привёл к выбору *оптимальных параметров разложения (пространственно-временной дискретизации)*, ведущих к **оптимальному смазу** движущегося изображения на фотоприёмнике. На первом этапе развития космической видеоинформатики считалось, что оптимальный смаз составляет половину элемента разрешения [45]. Позднее было осознано, что максимум качества изображения достигается лишь при оптимальном времени кадра, зависящем от скорости движения изображения: чем больше скорость, тем меньше оптимальное время кадра и (при постоянной плотности потока принимаемых фотонов) тем меньше информации содержит кадр изображения [98].

Оптимальный смаз должен быть больше нуля в любой прикладной телевизионно-компьютерной системе. Стремление свести смаз к нулю означает веру проектировщика в детерминированность движения объектов. Конкретное значение оптимального смаза зависит от модели движения сигнала. При известной форме траектории объекта с параметрической неопределённостью смаз должен иметь достаточно большое значение (несколько пикселей) (рис. 2.11) и определяется допустимой задержкой получения информации о параметрах движения объекта. При случайной траек-



Рис. 2.11. Изображение экрана монитора телевизионной системы контроля космического пространства с сигналом от движущегося спутника на фоне звёзд; смаз на порядок больше размера объекта

нии) обеспечивается тем, что *отношение межкадрового разностного сигнала к шуму должно быть равно пороговому отношению сигнал/шум*. Это значение оптимального смаза логически связано с потенциальной разрешающей способностью, вычисляемой для случая *большой априорной информации о форме* сигналов, когда решение о наличии двух объектов принимается по незначительному увеличению длительности сигнала [45].

Если в постановке задачи разрешения Рэлеем требуется провал между сигналами двух соседних объектов, то при априори известных размерах объектов и равенстве интенсивностей сигналов от них для принятия решения о наличии двух объектов достаточно увеличения длительности (ширины) сигнала на относительно малое значение, обратное отношению сигнал/шум. Поэтому обе эти задачи (для интервала разрешения и оптимального смаза) по формулам и по логике связаны с формулой Вудворда [48] для потенциальной точности измерения координаты импульса известной формы.

В научно-прикладном телевидении, в частности в космосе, широко применяются телевизионные системы на основе и линейных, и матричных твёрдотельных фотоприёмников. Кадровые системы имеют самую большую чувствительность и благодаря ей возможную высокую кадровую частоту, обеспечивающую высокую слитность передачи движения. Кадровые системы «подвергли отрицанию», т. е. вытеснили, механические системы телевидения в вещании и многих других применениях (кроме офисных сканеров, систем панорамного наблюдения, систем дистанционного зондирования Земли).

Для разграничения сфер применения кадровых и строчных систем на начальном этапе развития космического телевидения было сформулирова-

тории движения малоразмерного объекта *оптимальный смаз меньше размера объекта (и меньше размера пиксела); он тем меньше, чем больше отношение сигнал/шум* (см. 5.3).

Изложенная концепция требует, чтобы в оптимальной системе изображение было *практически слитным*. Практическое приближение к слитности (непрерывности траектории сигнала во време-

но золотое правило малокадрового телевидения [106]: если нужны широкий угол зрения и большая *разрешающая способность* на местности, то следует применить сканирующую систему (строчную или одноэлементную); если нужна высокая *чувствительность*, то следует применить систему с кадровым накоплением.

Под давлением слияния телевизионной и компьютерной техники, породившего цель получения всё большего количества и всё большего качества информации, особенно при межкадровой обработке видеoinформации, это правило включило понятие слитности [107]: *если высокая слитность изображения по полю важнее слитности во времени, то следует применить сканирующую систему; если важнее высокая слитность воспроизведения движения объектов, то следует применить систему с кадровым накоплением.*

Это правило корректно применимо при условии «исключающего или». Вместе с тем, если, исходя из априорной информации и ограничений, можно обеспечить и то, и другое, то при равенстве времени кадра система с кадровым накоплением окажется чувствительнее, т. е. сможет обеспечить информацию лучшего качества.

Контрастная чувствительность и качество видеoinформации. При оценке качества формируемых изображений кроме количества пикселей в матрицах учитывают пороговую чувствительность системы, но акцент обычно делают на динамический диапазон. Вместе с тем в современных КМОП-матрицах весьма большой динамический диапазон достигается при нелинейной (близкой к логарифмической) свет-сигнальной характеристике [3], [108] (см. гл. 3). Однако в нелинейных системах динамический диапазон перестаёт быть исчерпывающей характеристикой светопередачи, и при оценке качества видеoinформации необходимо учитывать контрастную чувствительность системы, т. е. способность обнаруживать объекты, мало отличающиеся от фона. Контрастная чувствительность – понятие, относящееся к известной физической концепции: «*Надо определять вероятность того, что значение некоторой физической величины находится в данном интервале, но не претендовать на то, что это значение можно знать точно*» [102]. Контрастная чувствительность тесно связано с понятием макрокванта и через его размер определяет *слитность видеосигнала по амплитуде.*

Контрастная чувствительность изучалась в физиологии намного раньше, чем в телевидении. Фундаментальный закон Вебера–Фехнера

о логарифмической связи ощущений и воздействий (первый в науке информационный закон) для широкого диапазона яркости фона устанавливает постоянство порогового контраста обнаруживаемого сигнала. Пороговый контраст (отношение минимального обнаруживаемого сигнала к фону по соседству с ним) вместе с количеством элементов в изображении отражает качество видеoinформации.

В отличие от зрения, которому присущ ряд важных ограничений, прикладное телевидение может поставить цель достижения *потенциальной контрастной чувствительности*, ограниченной фотонным шумом фона. За прошедшие полвека научно-прикладное телевидение превратилось в видеoinформатику, охватившую и собственно традиционно понимаемое телевидение, и компьютерную обработку изображений, и телекамеры и офисные сканеры, и космическое телевидение, и IP-телевидение. Совершенствование технологий привело также к тому, что доминировавшие ранее в телекамерах шумы считывания перестали определять чувствительность, которая теперь в первую очередь определяется фотонным шумом в изображении. Это обстоятельство порождает новый взгляд на связь слитности изображений с контрастной чувствительностью.

Акцент на ошибках дискретизации проводился в предположении безошибочной передачи значений яркостных отсчётов пикселей. Вместе с тем наличие шума, в первую очередь фотонного, ставит второй предел на пути к стремлению к математической непрерывности сигналов. Наличие доверительных интервалов как по абсциссе, так и по ординате сигнала изображения ведёт к понятию **достаточной слитности** [107]. Конечно, проектировщик и пользователь системы стремятся к уменьшению этих доверительных интервалов, т. е. макроквантов, для увеличения *количества* получаемой *качественной* информации. Если при дискретизации шум ведёт к *вырождению числа отсчётов*, то при квантовании амплитуд сигналов (размер макроквантов, или шаг квантования, определяется контрастной чувствительностью) он ведёт к *вырождению разрядности сигнала* [55]. Важным отличием квантования амплитуды сигнала от дискретизации его аргументов в твёрдотельном компьютерном телевидении по сравнению с вещательным является опора не на контрастную чувствительность зрения, а на **потенциальную контрастную чувствительность**. Здесь подчеркнём, что повсеместно применяемый в твёрдотельном телевидении метод автоматической регулировки времени накопления в широком диапазоне

освещённости [1], [108] обеспечивает пороговый контраст, равный пороговому отношению сигнал/шум, делённому на квадратный корень из ёмкости пиксела. Это примерно соответствует контрастной чувствительности зрения. То, что *при автоматической регулировке времени накопления большая часть фотонов превращается в тепло*, а не в информационные фотогенерированные электроны, не позволяет достичь потенциальной контрастной чувствительности. Вместе с тем в научно-прикладном компьютерном телевидении актуально достижение потенциальной контрастной чувствительности, лимитируемой фотонным шумом фона, а не ёмкостью пиксела, выраженной в числе сигнальных электронов. Эта ясная концепция была сформулирована А. Роузом как требование учёта каждого фотона для извлечения из светового стимула максимума информации [27].

В рамках технологии приборов с зарядовой связью, характерной для первого этапа твёрдотельной революции в телевидении, цель группового счёта фотонов достигалась в два этапа с разнесением их на разные СБИС – фотоприёмную матрицу и вынесенный АЦП. На втором этапе твёрдотельной революции в телевидении, когда на первые роли вышли КМОП-матрицы, стало возможным размещение в пикселе большого количества дополнительных элементов [3], [29], [30]. Современные методы формирования сигналов в фотоприёмных КМОП-матрицах позволяют реализовать контрастную чувствительность, близкую к потенциальной, благодаря размещению в каждом пикселе многоразрядных счётчиков количества превышения накопленным сигналом выбранного порога за время формирования кадра изображения (см. гл. 3).

Технологические достижения привели к тому, что в погоне за практической *слитностью по полю* в интересах научно-прикладного компьютерного телевидения созданы матрицы со 120 мегапикселями и наборные поля матриц в 1 гигапиксел, а также сканеры с линейными фотоприёмниками по 12 тыс. элементов, что при круговом обзоре по азимуту даёт кадр размером более 1 гигапиксела.

В борьбе за *слитность во времени* была создана первая отечественная телевизионная система с кадровой частотой, существенно превышающей вещательную – 250 Гц при достаточном пространственном разрешении [109] (см. 5.2). В настоящее время создаются матрицы с *кадровой частотой в единицы и десятки килогерц* [110]– [112]. При этом кадрковая частота

наращивается считыванием окна, т. е. *обменом слитности по пространству на слитность во времени* (см. 5.1).

Таким образом, использование в *моделях* принципа слитности в *результатах наблюдения* дополняется **принципом достаточной слитности, при котором размер макрокванта определяется шумом**, и относящимся к упоминаемым Гельвецием и необходимым для прикладного телевидения «*некоторым общим правилам проектирования и использования оптимальных телевизионных систем*» [44]. Опора на принцип слитности при передаче многомерных полей в телевидении не просто ведёт к маскированию шумом ошибок дискретизации, но и соответствует передаче видеoinформации максимального качества.

Выводы по гл. 2. Цитированная на с. 13 и 95 мысль А. Пуанкаре о *новом слове, которое становится творцом*, получила хорошую иллюстрацию на примере понятия **шумовой информации**, без которой не формализуется **информационный риск и качество информации**. *Доминантная информация*, введение понятия которой стимулировано принципом доминанты А. А. Ухтомского и утверждением А. Н. Колмогорова о нескальярности информации, и *информационный риск*, введение понятия которого стимулировано понятием среднего риска Т. Байеса и растерянностью исследователей из-за неэффективности опоры только на потерю информации, *наполнены конкретикой* благодаря понятию *шумовой информации*.

В ходе обсуждений теории кодирования зашумлённых сигналов встречались мнения о том, что шумовая информация не опасна, она только в нереальном случае континуальных сигналов страшна, а в телевидении при низких освещённостях фотонов мало и все они полезны, а при больших освещённостях шумов на изображении не видно. Однако это означает только то, что фотонный шум не виден благодаря рациональному подавлению шумовой информации с помощью накопления и квантования.

Шумовая информация на интуитивном уровне уже признана, и её формализация востребована не только с позиций теории связи, но и с позиций экспериментальной физики, в которой всегда актуален вопрос *округления* результатов с *непременными значащими цифрами (доминанта)* и *доверительным интервалом* – это отголосок остаточной шумовой информации. Почему только отголосок? Потому что при конечном с точки зрения экспериментатора доверительном интервале *точность задания*

этого доверительного интервала может быть малой или бесконечной. Эффект от теории кодирования зашумлённых сигналов – тот же, что и от разумного округления. Физика создавала *правила округления* для измерения одной *неизвестной величины*. Теория решений и теория информации создают эти *правила для случайных величин и процессов*. Поэтому **правила накопления, усечения спектров и округления** – основной результат теории кодирования зашумлённых сигналов.

Теория доминантной информации благодаря опоре на понятие шумовой информации позволила установить важную границу для информационного риска, которая отражает принципиальное незнание и играет ту же роль, что и понятие абсолютного нуля температуры. Тезис С. Лема [99]: **«Информация без фальсификаций, к сожалению, является фикцией»** – формализован в виде **закона о нижнем пределе информационного риска**. К. Черри (см. цитату на с. 79) мечтал извлечь всю полезную информацию и отбросить всю фиктивную (шумовую). Можно сказать, что один из конструктивных результатов теории доминантной информации – установление **пределов возможного** (в виде **законов** взаимобмена компонент информационного риска и его нижней границы) на пути **намеченного Черри «информационного вечного двигателя»**. К. Черри рассуждал о пропускной способности, но его слова относятся и к найденной границе информационного риска: **«Каждый, кто попытается превысить этот предел, зря потеряет время!»**.

Изложенное в этой главе можно рассматривать как иллюстрацию к тезисам философии познания [5]: *«Проблема – это конфликт идей. Проблемы неизбежны. Проблемы можно решить. Новые объяснения приводят к новым проблемам»*. Например, проблема обоснования оптимального усечения полосы сигнала часто является следствием конфликта идеи В. А. Котельникова и К. Шеннона о том, что полоса частот всегда каким-либо образом ограничена, и идеи бесконечного количества информации в непрерывном случайном сигнале. Теория кодирования зашумлённых сигналов решает проблему обоснования *оптимального усечения* полосы частот непрерывных сигналов в *процессе их рождения* и даёт конкретные правила для вычисления характеристик кодирующих и декодирующих фильтров. Принцип доминантной информации обосновывает представление видеоинформации в виде совокупности макроквантов, размеры которых определяются шумом. Важная философская нагрузка изложенной теории связана с устранением неопределённости, имевшейся в теории связи из-за опоры

на функции с ограниченным спектром, введённые В. А. Котельниковым и К. Шенноном без должных обоснований. Основанием, на которое опирается изложенная теория кодирования зашумлённых сигналов, служит *принцип минимума количества информации при максимуме её качества*.

Каждый шаг в развитии теории включает прежнее знание, которое делает своей составной частью. Так, мнение К. Шеннона, процитированное на с. 50, характеризует его вклад в теорию связи, основанный на учёте статистических свойств сигнала и шума. Сейчас, после введения в статистический синтез систем связи принципа доминантной информации, стало **значительно глубже пониматься действие входного шума, влияние которого на качество информации значительно сильнее, чем влияние шума канала.**

Изложенная теория линейного кодирования при передаче зашумлённых гауссовских сигналов имеет перспективы широкого применения и *развития* с целью учёта произвольных функций распределения сигналов и шумов, влияния на качество принимаемых сигналов не только входного шума, но и априорной неопределённости и нестационарности источников сигнала и источников шумов, а также формализованного ограничения на вычислительную сложность кодера и декодера. Трудности приложения существующей теории информации, в частности к телевизионной технике, конечно, имеются. Они состоят и в отсутствии решений упомянутых проблем, и в специфике субъективной оценки качества телевизионных изображений (связанной с многомерностью сигнала и его *смыслом*). Однако уже сейчас многочисленное сообщество изобретателей способов и устройств кодирования изображений может перестать относиться к теории информации только как к теории о существовании некоторых оптимальных методов кодирования. На самом деле теория информации – конструктивная теория, дающая фундаментальные решения задач кодирования сигналов. Одним из таких конструктивных результатов является теория линейного кодирования зашумлённых сигналов.

Применение принципа доминантной информации, насыщенного конкретикой в виде изложенных аксиом, закона взаимообмена потери доминантной информации и шумовой информации и формул для оптимальных кодирующих фильтров, в соответствии с идеей А. Н. Колмогорова о том, что *информация не является скалярной величиной*, является первым шагом

на пути построения теории доминантной информации, опирающейся на учёт количества, качества и цели информации.

Можно прогнозировать применение теории доминантной информации и при разработке государственных стандартов, относящихся к избирательности радиотехнических систем, в которых до настоящего времени (хотя и привлекают децибелы, т. е. логарифмы, вплотную приближающие к информационным оценкам) нет опоры на теорию информации.

В силу того, что *«новые объяснения приводят к новым проблемам»* [5], можно прогнозировать, что следующие проблемы, связанные с учётом качества информации, будут порождены конфликтом идей о том, что статистическая теория связи опирается на модели случайных процессов и что реальные системы связи строятся для передачи семантической информации. Эффект от большинства теорий – лучшее понимание, пусть и всегда обречённое на продолжение поиска ещё лучшего понимания. Это означает, что каждый этап в развитии понимания конкретной предметной области важен не только в узком смысле, но и имеет важный философский эффект. Р. Фейнман отмечал [63]: *«Именно философия помогает нам строить догадки»*. Изложенная теория доминантной информации – не только формулы и законы, но и философия, которая поможет строить дальнейшие догадки лучше, чем предшествовавшие теории.

ЧАСТЬ II

УПРАВЛЕНИЕ НАКОПЛЕНИЕМ КАЧЕСТВА ВИДЕОИНФОРМАЦИИ В ТЕЛЕКАМЕРАХ

Оптические измерительные инструменты не являются просто пассивными «приёмниками», которые только регистрируют попадающие на них лучи; они активно участвуют в процессе измерения, оказывая причинное влияние на результат.

Макс Планк

Глава 3. ОРГАНИЗАЦИЯ НАКОПЛЕНИЯ И СЧИТЫВАНИЯ В МАТРИЧНЫХ КМОП-ФОТОПРИЁМНИКАХ

В этой главе мы не только будем использовать результаты первой части в качестве эталонов для характеристик фотоприёмников, но и рассмотрим проблемы, порождённые конфликтом следующих идей. Во-первых, это конфликт идеи математической статистики, формулируемой как «чем больше данных, тем лучше», и «телевизионной» идеи о том, что чем больше интервал накопления по полю и по времени, тем больше отношение сигнал/шум.

Во-вторых, это конфликт идеи А. Роуза [27] о том, что для извлечения максимума информации нужна предельная квантовая эффективность, нужно учитывать *каждый фотон*, и идеи о необходимости расширения светового диапазона фотоприёмника с помощью автоматической регулировки времени накопления (АРВН), в результате работы которой световой диапазон расширяется на несколько порядков, но подавляющее *большинство фотонов превращается не в информацию, а в тепло*.

В-третьих, это конфликт идеи о том, что пошаговое накопление информации, называемое режимом временной задержки и накопления, реализуется только в приборах с зарядовой связью (ПЗС), и идеи о том, что стратегический путь развития фотоприёмной микроэлектроники – видеосистемы на кристалле, которые реализуются с использованием КМОП-технологии и не могут быть реализованы в рамках технологии ПЗС.

Развитие микроэлектроники продолжает соответствовать закону Мура, гласящего, что количество транзисторов, размещаемых на кристалле микросхемы, удваивается каждые 18...24 мес., что обеспечивается КМОП-

технологией. Развитие идёт по пути повышения уровня интеграции и воспроизводимости параметров полупроводниковых приборов с применением новых элементов обработки аналоговых сигналов, в первую очередь по пути перехода к субмикронным технологиям и системам на кристалле, когда в отличие от ПЗС на кристалле располагается целая система, состоящая из сложнофункциональных блоков [113]. Попытки использования КМОП-технологии для изготовления фотоприёмников предпринимались очень давно, но реальные успехи были получены с уровня интеграции 3...5 мкм. Однако по качеству получаемого сигнала, чувствительности и однородности такие фотоприёмники существенно уступали развивавшимся параллельно фотоприёмникам по технологии приборов с зарядовой связью. С выходом на уровень интеграции 0,25...0,35 мкм у разработчиков наряду с существенно возросшей стабильностью параметров интегральных полупроводниковых приборов появилась возможность использовать активные пиксели с усилением сигнала [29], а обработку сигналов фотоприёмников проводить непосредственно на кристалле микросхемы [3], [31]. При этом КМОП-технология позволила создавать фотоприёмные матрицы с уникальными потребительскими характеристиками информационной производительности – вплоть до гигапиксельных форматов или килогерцевых кадровых частот [110]–[112], [114]–[118]. В дальнейшем при переходе на уровень КМОП-технологии 0,11...0,18 мкм стало возможным проектирование законченных видеосистем на едином кристалле, включающих в себя наряду с фотоприёмниками высокого разрешения усилители и преобразователи сигнала, оперативную память, источники питания, блоки интерфейсов передачи информации потребителю и разнообразные цифровые блоки, в том числе и встроенные кодеры информации. Потребитель получил возможность использовать однокристалльные камеры для самых разных применений, в том числе для мобильных устройств и систем технического зрения. Количество видеоинформации в мировом масштабе совершило гигантский скачок исключительно за счёт массового использования современных видеосистем на кристалле, изготовленных по КМОП-технологии.

Ключевой частью фотоприёмников является фотоприёмная ячейка, традиционно называемая пикселом, под которым мы имеем в виду один элемент фотоприёмной матрицы. Он по световому потоку $E(x, y, t, \lambda)$ с учётом квантовой эффективности η фотоэлектрического преобразования формирует сигнал S в i -м элементе j -й строки k -го кадра в n -м интервале длин

волн $\Delta\lambda_n$, реализуя накопление информации в соответствии с многомерным интегралом, охватывающим интегрирование по пространственным координатам x и y , времени t и длине волны λ , причём и при проектировании фотоприёмников, и при оптимизации телекамер (см. гл. 4 и 5) важнейшей операцией становится выбор пределов интегрирования:

$$S_{ijkn} = \int_{\lambda_n}^{\lambda_{n+1}} d\lambda \int_{i\Delta x}^{(i+1)\Delta x} dx \int_{j\Delta y}^{(j+1)\Delta y} dy \int_{k\Delta t}^{(k+1)\Delta t} \eta E(\lambda, x, y, t) dt. \quad (3.1)$$

В задачу оптимизации накопления и считывания в фотоприёмнике входит в первую очередь достижение наилучшей квантовой эффективности (см. [27]), во вторую – такая организация накопления, которая обеспечивает максимум качества информации, учитывающего как потери доминантной информации о распределении $E(x, y, t, \lambda)$, так и ложные сигналы – темновых токов, шумов считывания и т. п.

3.1. КМОП-технология фотоприёмников

Не вдаваясь в детали КМОП-технологического процесса (см., например, [119], [120]), необходимо отметить, что для формирования фотоприёмных областей на кристалле необходима частичная *модификация* стандартного цифрового технологического маршрута. На этапе формирования активных областей требуется проведение дополнительных имплантаций примесей в глубокие и поверхностные слои кремния для формирования встроенных полей, буферных областей и каналов проводимости с целью сбора фотоэлектронов с требуемой зоны; отсечки ИК-сигнала, возникающего в более глубоких слоях кремния и уменьшения шума поверхностной генерации в составе полезного сигнала. Наличие большого процента площади, занятой фотоприёмными областями, зачастую нарушает требования по плотности заполнения, что, в свою очередь, требует специальных мер для обеспечения качества химико-механической полировки при многоуровневой системе металлизации. Также необходимо использовать логические операции по отмене формирования защитных пассивных элементов в слоях металлизации, препятствующих прохождению светового потока к фоточувствительным областям в кремнии. В конце концов, необходимо убедиться, не являются ли слои изолирующего оксида кремния или защитных стекол на поверхности кристалла поглощающими фильтрами для интересующих нас длин волн. Это также может потребовать изменения технологии изготовления в части толщин и состава этих слоёв. Таким обра-

зом, мы видим, что поддержание возможности использования фотоприёмных приборов в составе КМОП-технологического процесса изготовления микросхем существенно увеличивает его сложность, повышает требования ко многим параметрам и ведёт в итоге к удорожанию технологии изготовления. Поэтому возможность использования фотоприёмников в составе КМОП-микросхем есть далеко не на каждом производстве, а наличие такой возможности характеризует технологический процесс с точки зрения качества, воспроизводимости и широких возможностей разработчика в обработке не только цифровых, но и аналоговых сигналов.

Организация структуры пиксела влияет на квантовую эффективность и пороговую чувствительность фотоприёмного устройства. Схема считывания для более передовых технологий может быть размещена под шинами передачи данных (рис. 3.1).

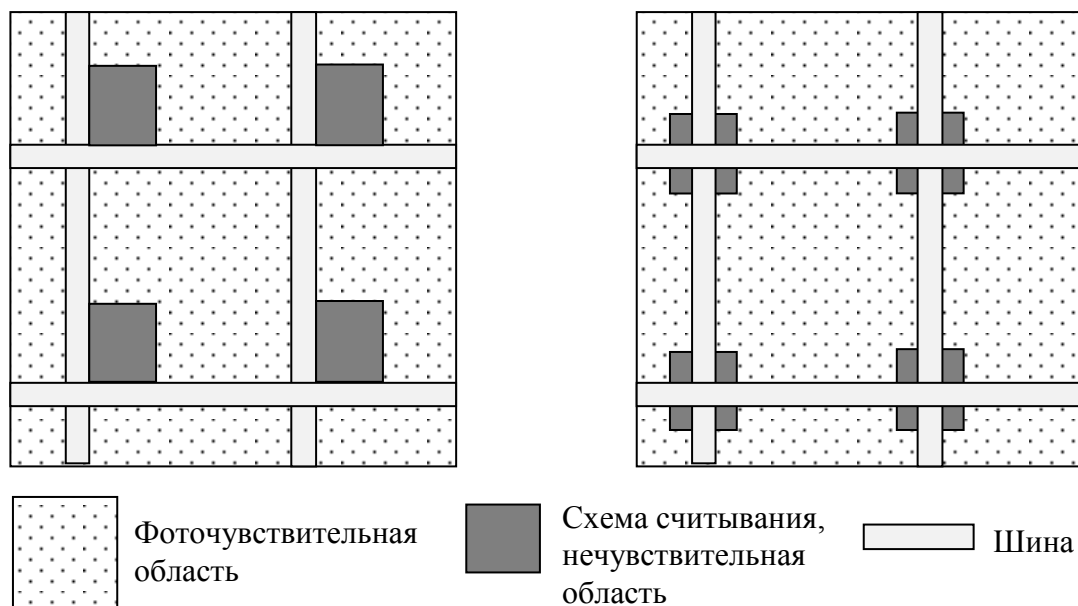


Рис. 3.1. Типовые структуры размещения схем считывания в пикселах

Для более эффективного использования площади кристалла микросхемы шаг пикселей следует выбирать кратным шагу ячеек цифровых библиотек. Исходя из этого принципа, используются типовые размеры 5,04, 10,08 и 15,12 мкм.

В технологии изготовления микросхем с использованием шести слоёв металлизации есть возможность вертикального совмещения нечувствительных областей. Две вертикальные шины могут быть выполнены в первом слое металла, четыре горизонтальные – в других слоях: две во втором и две в четвертом. При небольшом размере пиксела фоточувствительная

область может оказаться на глубине порядка 10 мкм от поверхности кристалла, при наклонном падении света металлизация может затенять фоточувствительную область. Поэтому разработчики стараются использовать минимально достаточное количество металлов (рис. 3.2). Важным примером служит технология цветных КМОП-фотоприёмников, реализованная в составе цифроаналогового КМОП-процесса 0,18 мкм фабрики «SilTerra» (Малайзия).

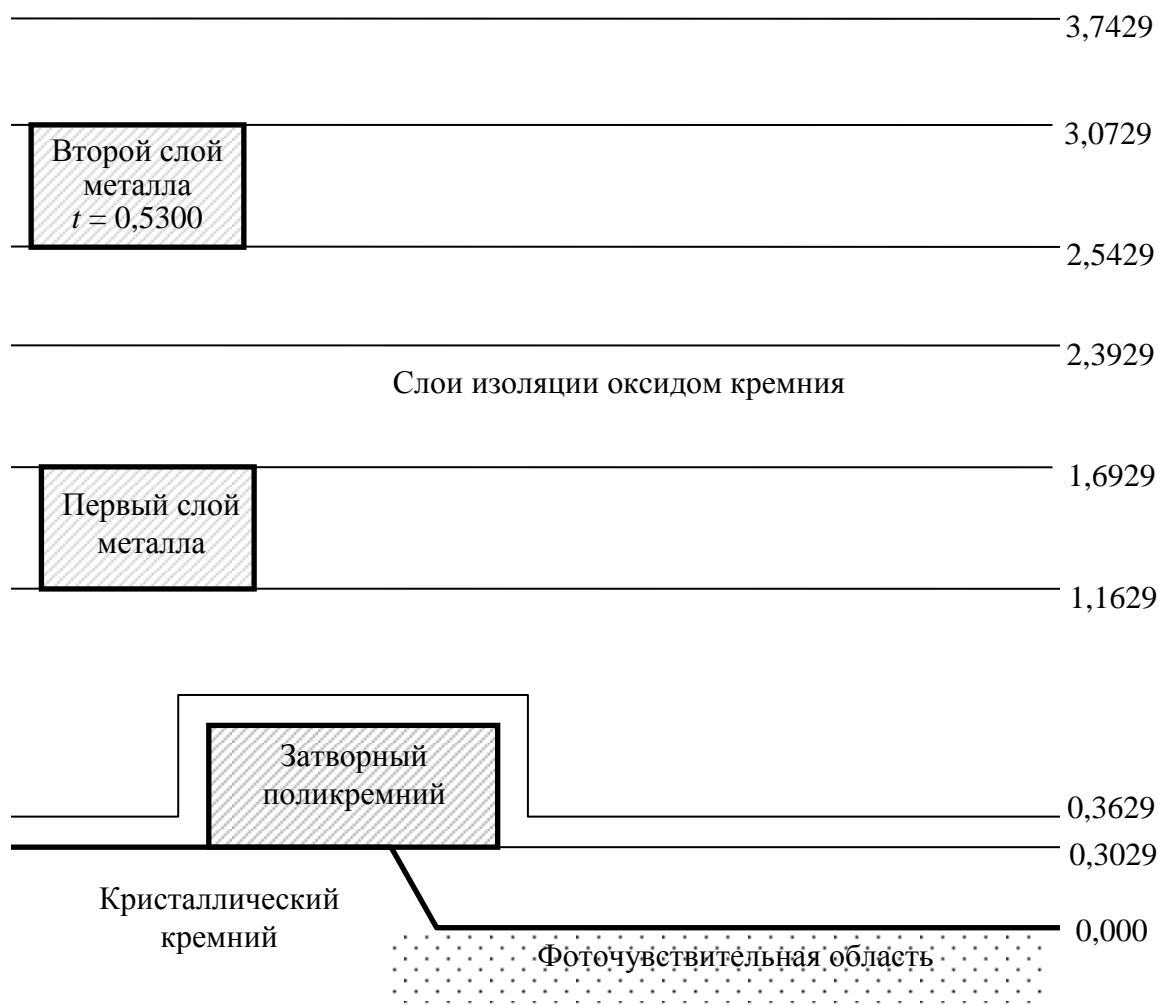


Рис. 3.2. Сечение структуры фоточувствительной области и первых двух слоёв металлизации; толщина слоёв указана в микрометрах

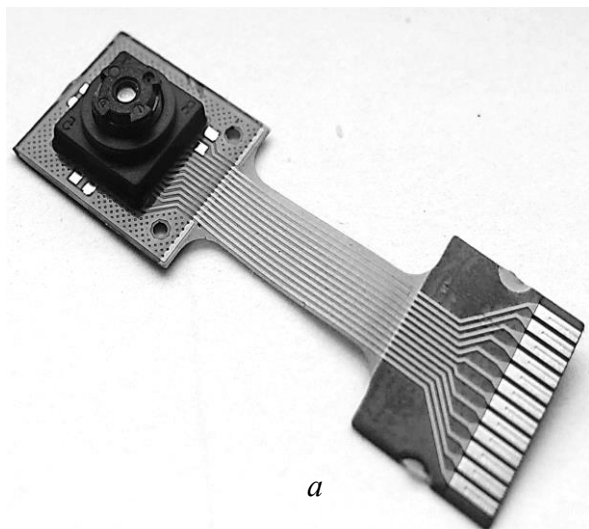
Традиционно отечественные фотоприёмники разрабатывались в Зеленоградском НИИ физических проблем им. Ф. В. Лукина. Начиная с 1970-х гг. велись работы по новым структурам приборов с зарядовой связью. В стенах НИИФП была создана первая технология производства отечественных ПЗС-фотоприёмников.

В начале 1990-х гг. при исследовании процессов генерации фотоэлектронов группой исследователей под руководством проф. Ю. И. Тишина

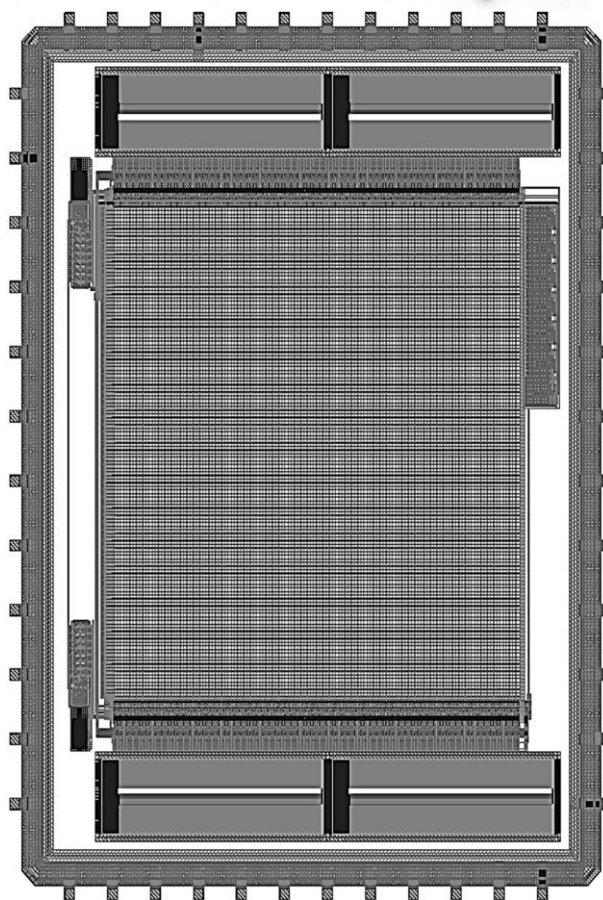
были получены результаты, позволяющие определить глубину максимальной плотности генерации электронов, зависящую от длины волны света. В 2000 г. группа разработчиков перешла в ООО «Юник Ай Сиз», где появилась возможность продолжения исследований на зарубежной технологической базе. В 2002 г. было заключено соглашение с малайской фабрикой по производству КМОП-микросхем, в настоящее время называющейся «XFAB Sarawak», о совместной разработке технологии цветных КМОП-фотоприёмников без внешних цветных фильтров в составе КМОП-технологического процесса с проектными нормами 0,25 мкм. После трёх лет работы были получены первые КМОП-фотоприёмники, реализующие разделение фотоэлектронов по глубине их генерации на основе системы глубокой многоуровневой имплантации. Однако невозможность стабилизации параметров исключила возможность коммерческого использования этой разработки. В 2006 г. было принято решение о переходе на КМОП-технологиями 0,18 мкм фабрики «SilTerra», которая смогла обеспечить требуемую воспроизводимость параметров приборов для коммерческой реализации данных фотоприёмников, а также заложила основы для развития серии фотоприёмных структур для лавинных фотоприёмников и видеосистем на кристалле. В настоящее время данный технологический процесс позволил создать более десятка КМОП-фотоприёмников и видеосистем на кристалле, применяемых в различных областях, в том числе в бортовой аппаратуре видеосистем космических аппаратов.

Примером видеосистемы на кристалле может служить микросхема КМОП-фотоприёмника с функцией датчика координат [121] (рис. 3.3). Этот проект возник в рамках создания технологии КМОП-фотоприёмников по КМОП-технологии с проектными нормами 0,25 мкм на фабрике «XFAB Sarawak» в Малайзии. Авторским коллективом была решена задача технологической отработки процесса изготовления фоточувствительных элементов, усиления, преобразования и обработки полученного видеосигнала, а также объединения различных функциональных, интерфейсных и служебных блоков в единую видеосистему на кристалле.

Первая версия этой микросхемы была аттестована и запущена в массовое производство в декабре 2004 г. До настоящего времени она используется в промышленных и игровых устройствах, позволяя создавать трёхосевые бесконтактные и безынерционные оптические датчики.



a



б

Рис. 3.3. Бескорпусная сборка с КМОП-видеосистемой на кристалле UIC1201 на гибкую печатную плату с интегрированной оптической системой (а) и её топология (б)

Однако современные технологии и рост технических требований привели к необходимости разработки изделия с качественно новыми параметрами. Базовой для новой версии микросхемы была выбрана технология КМОП 0,18 мкм малайской фабрики «SilTerra», позволяющая увеличить степень интеграции и скорость обработки сигнала. Структура микросхемы (рис. 3.4) включает монохромный приёмник изображения формата 128×128 пикселей со встроенным блоком вычислителя координат центра тяжести изображения специально формируемого оптического сигнала и является типичным примером реализации *видеосистемы на кристалле*. Топологический рисунок кристалла на рис. 3.3 характеризуется расположением блоков, аналогичным их расположению на функциональной схеме рис. 3.4.

Для расширения функциональных возможностей – определения не только двух ортогональных координат, но и угла поворота – предусмотрено определение координат двух поочередно включаемых оптических сигналов, формирующих световые метки на

фоточувствительном поле. В микросхеме реализован режим выделения оптического сигнала на фоне паразитных засветок с помощью вычитания промежуточных кадров, которые хранятся во встроенных блоках памяти.

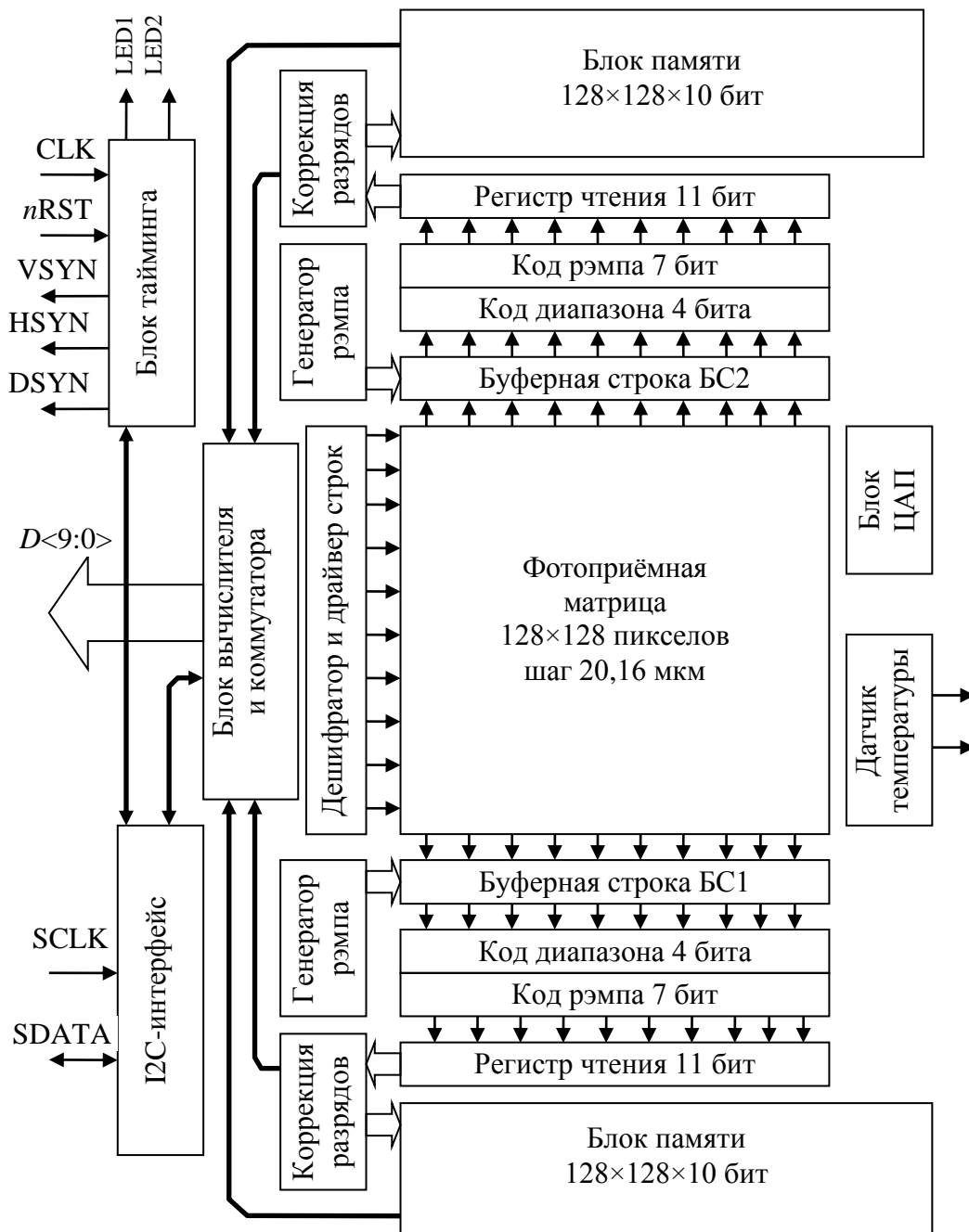


Рис. 3.4. Функциональная схема построения видеосистемы

Синхронное с экспозицией кадров включение оптических сигналов осуществляется по командам блока тайминга, выдаваемым на выходы LED1 и LED2.

Блок тайминга – управления и синхронизации – программируемый по интерфейсу I2C, управляет режимами работы микросхемы. На него приходят внешний тактовый сигнал CLK опорной частотой 4...48 МГц и сигнал начальной установки (сброса) nRST. Этот блок формирует сигнал сопровождения видеоданных DSYN, кадровый и строчный импульсы VSYN,

HSYN. В случае необходимости соответствующей командой можно отключить выдачу видеоданных по параллельной 10-разрядной шине $D\langle 9:0 \rangle$, уменьшив таким образом энергопотребление.

Важной характеристикой приёмника изображения для технического зрения является динамический диапазон (см. 3.5). В этой видеосистеме на кристалле для расширения динамического диапазона использовано относительно простое решение: пиксел снабжён двумя последовательно расположенными плавающими узлами FN1 и FN2 и, соответственно, двумя активными транзисторами T1 и T2 (рис. 3.5).

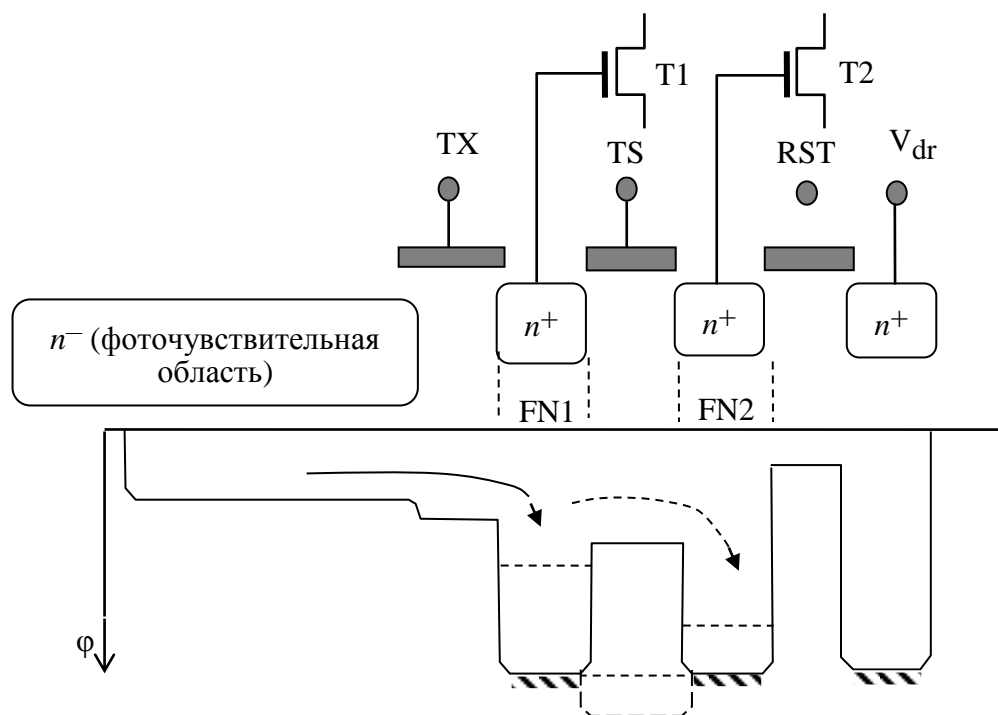


Рис. 3.5. Схематический разрез и диаграмма потенциалов пиксела с двумя последовательными плавающими узлами

Слабые сигналы в пикселах с небольшой освещённостью детектируются на высокочувствительном первом плавающем узле с характерной ёмкостью 3 фФ и коэффициентом преобразования примерно 50 мкВ/электрон.

В сильно освещённых пикселах фотогенерированный заряд после переполнения первого плавающего узла перетекает во второй плавающий узел, ёмкость которого в 16 раз больше, а коэффициент преобразования, соответственно, в 16 раз меньше.

Считывание сигналов с первого и второго плавающих узлов производится последовательно, каждый в режиме двойной коррелированной выборки (ДКВ): сначала считываются уровни предустановки плавающих узлов, затем открывается затвор переноса TX и считывается сигнал с первого плавающего узла в нижнюю буферную строку БС1; после этого открыва-

ется затвор TS и считывается второй суммарный сигнал в верхнюю буферную строку БС2. В буферных строках уровни предустановки вычитаются из сигналов в технике переключаемых конденсаторов. Необходимые для работы пикселей и буферной строки постоянные напряжения вырабатываются блоком ЦАП по командам блока тайминга.

Параметры фотоприёмника в значительной мере зависят от способов организации (архитектуры) управления фотоприёмным массивом пикселей и считывания информации с них. Хорошо известны недостатки традиционной архитектуры приёмников изображения, в которых сигналы с выбранной строки двумерного массива пикселей записываются с помощью ДКВ в буферную строку, а затем элементы буферной строки с хранящейся в них информацией поочередно подключаются к длинной сигнальной шине, по которой аналоговый сигнал подаётся на программируемый усилитель и далее на аналого-цифровой преобразователь. Передача аналогового сигнала по длинной шине требует большой мощности от каждого элемента буферной строки; велико также потребление скоростного многоуровневого АЦП, что приводит к нагреву кристалла и высокому уровню помех.

Для преодоления упомянутых недостатков в современных конструкциях приёмников изображения применяется архитектура с аналого-цифровым преобразованием в каждом столбце. В столбцах используются АЦП последовательного приближения (*successive approximation, SAR*), циклические АЦП, однако наилучшие результаты по однородности видеосигнала, скорости вывода и энергопотреблению демонстрирует архитектура параллельного преобразования с общей шиной пилообразного напряжения опорного плеча всех компараторов буферной строки – рэмпом (*ramp*) [121]. Лучше всего такой способ аналого-цифрового преобразования подходит для приёмников изображения большого формата, поскольку занимает число тактов, равное числу отсчётов, т. е. для 10-разрядного преобразования потребуется 1024 такта. Преобразование во всех столбцах производится одновременно с выводом цифровых данных предыдущей строки, так что длина строки фотоприёмника должна быть не меньше числа отсчётов.

Для уменьшения числа тактов преобразования предложена конструкция с мультиплицированным рэмпом (*multiple-ramp single-slope, MRSS*) [121]. Суть её в том, что полная шкала изменения входного сигнала разбивается на m диапазонов, у каждого своя шина рэмпа. На первом этапе оцифровки тем или иным методом (последовательным перебором или методом стягивающихся хорд) определяется, в каком из диапазонов находится

сигнал в данном столбце, на что потребуется $n1 = m$ (последовательный перебор) или $n1 = \log 2m$ (метод хорд) тактов. Двоичный номер диапазона записывается в старшую часть регистра результата. На втором этапе линейным изменением напряжения одновременно на всех шинах диапазонов определяются младшие разряды результата. Например, если при 10-разрядном преобразовании определить $2^4 = 16$ диапазонов, то на рэмп в каждом диапазоне останется $2^6 = 64$ уровня (ступеньки пила). Весь процесс оцифровки займет $16 + 64 = 80$ (перебор) или $4 + 64 = 68$ (хорды) тактов, а не 1024, как в случае одиночного рэмпа.

В описываемой микросхеме (см. рис. 3.4) полная шкала видеосигнала амплитудой 1 В в блоке «Генератор рэмпа» резистивной цепочкой делится на 16 диапазонов. К буферной строке через повторители подключены 16 шин с соответствующих выходов генератора рэмпа. Буферная строка в каждом столбце содержит компаратор-интегратор со схемой двойной выборки, набор ключей и логику выбора диапазона рэмпа. Совмещение компаратора и интегратора в одном каскаде позволяет компенсацию сдвига плеч компаратора выполнять в аналоговой форме на этапе двойной коррелированной выборки.

При практической реализации необходимо учитывать возможную нестыковку краев диапазонов [121], поэтому «пила» рэмпа в каждом диапазоне имеет 96 ступеней: 64 ступени сам диапазон и по 16 ступеней с каждой стороны на перекрытие соседних диапазонов.

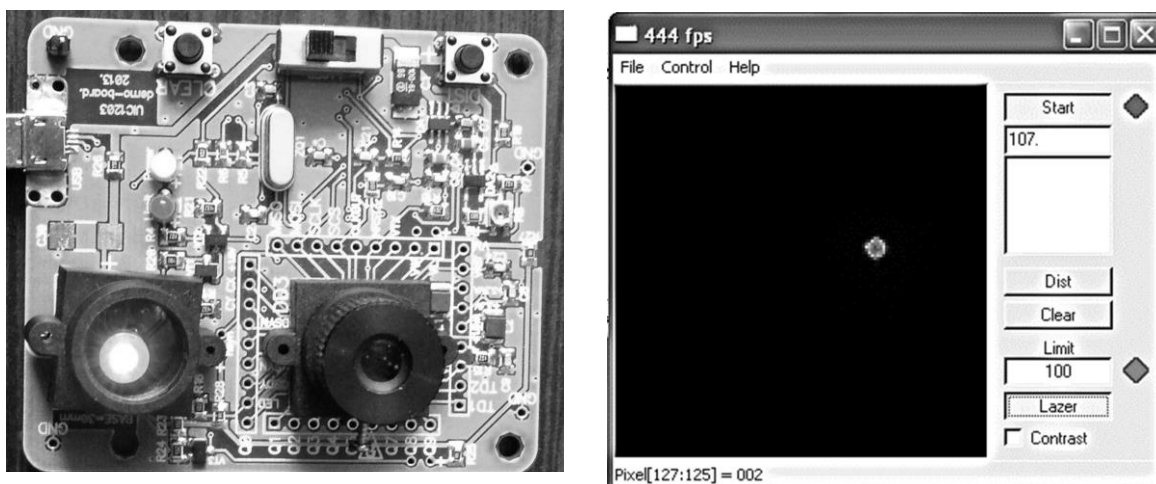
Преобразование хранящихся в буферной строке аналоговых сигналов в цифровую форму происходит следующим образом. Сначала переключением шин диапазонов в каждом столбце методом стягивающихся хорд определяется диапазон, в котором находится сигнал данного столбца, и код найденного диапазона (4 бита) записывается в старшие биты регистра записи (на рис. 3.4 блок «Код диапазона 4 бита»); шина выбранного диапазона остается подключенной к компаратору данного столбца. Далее на все шины рэмпа подается «пила» в 96 ступеней. Синхронно с началом тактирования рэмпа запускается инкрементный набор кода в 7 младших разрядов регистра кода сигнала. При достижении рэмпом уровня компарируемого сигнала и переключении компаратора останавливается инкрементная запись кода в регистр. Таким образом, в 11-битном регистре оказывается число, соответствующее амплитуде сигнала пиксела.

В конце процедуры аналого-цифрового преобразования 11-разрядные коды сигналов из регистра записи переписывается в регистр чтения. Из регистра чтения коды сигналов попадают в блок «коррекции разрядов».

С помощью таблицы поправочных сдвигов, хранящейся в регистрах коррекции, формируется 10-разрядный код сигналов, при этом удаляются области перекрытия диапазонов. Таблица поправочных сдвигов формируется при фабричном тестировании и паспортизации конкретного образца микросхемы. После блока коррекции полные 10-разрядные сигналы нижней и верхней буферных строк, соответствующие фотогенерированным зарядам первого и второго плавающих узлов пикселей, подаются на блоки памяти и блок вычислителя и коммутатора. 14-разрядные координаты центров тяжести изображений оптических сигналов вычисляются в каждом рабочем кадре и передаются по интерфейсу I2C. Основные характеристики микросхемы сведены в таблицу (световые измерения проведены в ОАО «Геофизика-Космос» А. М. Щекиным).

Параметр	Значение
Формат	128×128
Шаг пикселей	20,16×20,16 мкм ²
Габариты кристалла	3,8×5 мм ²
Технология	0,18 мкм CMOS «Silterra»
Напряжения питания ядра	1,8 В
Напряжение питания интерфейсов	3,3 В
Ток потребления ядра	Не более 12 мА
Ток потребления интерфейсов	3...40 мА, в зависимости от нагрузки и частоты шины выходных данных
Интерфейс управления	I2C
Тип пиксела	Заглублённый диод с полным обеднением и двумя последовательными плавающими узлами
Спектральный диапазон	0,4...1,05 мкм
Коэффициент преобразования первого плавающего узла	~50 мкВ/электрон
Коэффициент преобразования второго плавающего узла	~3 мкВ/электрон
Шина выходных видеоданных	Мультиплексированная из двух 10-разрядных параллельных, со сдвигом на четыре разряда и внешней сшивкой
Частота шины выходных данных	2...24 МГц, задается тактовым сигналом CLK 4...48 МГц
Режимы экспозиции	Мгновенный (<i>Global Shutter</i>); бегущая шторка (<i>Rolling Shutter</i>)
Длительность экспозиции	Мгновенная – от 0 до 16 384 строк; бегущая шторка – от 128 до 16 512 строк
Монохроматическая пороговая экспозиция (длина волны 700 нм)	$5,6 \cdot 10^{-12}$ Дж/см ²
Интегральная пороговая экспозиция (источник типа «А»)	$1,1 \cdot 10^{-5}$ лк · с

На основе первой отечественной видеосистемы на кристалле UIC1201 может быть создан ряд приборов, примером которых является разработанный в ООО «Юник Ай Сиз» дальномер (рис. 3.6 [122]).



а *б*
Рис. 3.6. Дальномер на основе видеосистемы на кристалле UIC1201:
а – внешний вид; *б* – окно пользователя на экране компьютера

Дальномер работает на основе принципа измерения паралакса и при работе на частоте кадров до 500 Гц обеспечивает высокую точность измерения расстояний на умеренных дальностях (до 10 м).

Примером современного КМОП-фотоприёмника является матрица формата 1024×1024 высокочувствительных высокоскоростных пикселей, расположенных с шагом 15,12×15,12 мкм (рис. 3.7).

Коэффициент заполнения (*Fill Factor*) этой матрицы больше 75 %, спектральный диапазон – 0,36...1,05 мкм. Матрица разбита на 4 квадранта – секции *A*, *B*, *C* и *D* форматом по 512×512 пикселей. Верхняя половина матрицы (секции *A* и *B*) сканируется от середины вверх с выносом видеосигнала в буферные строки *A* и *B*. Нижняя половина матрицы (секции *C* и *D*) сканируется от середины вниз с выносом видеосигнала в буферные строки *C* и *D*. Драйверы строк для каждого из квадрантов *A*, *B*, *C*, *D* формируют необходимые уровни сигналов управления пикселями выбранной строки (сигналы RST, TX, SR). Буферные строки предназначены для приёма сигналов и подавления шумов от пикселей выбранной строки, а также схемы подготовки к аналого-цифровому преобразованию, которая состоит из усилителей с ДКВ и выходных компараторов. Блок распределенных АЦП с архитектурой «RAMP» преобразует аналоговые сигналы выбранной строки пикселей в 12-разрядный цифровой код. Быстрый поиск положения видеосигнала методом стягивающихся хорд и соответствующее аналого-цифровое

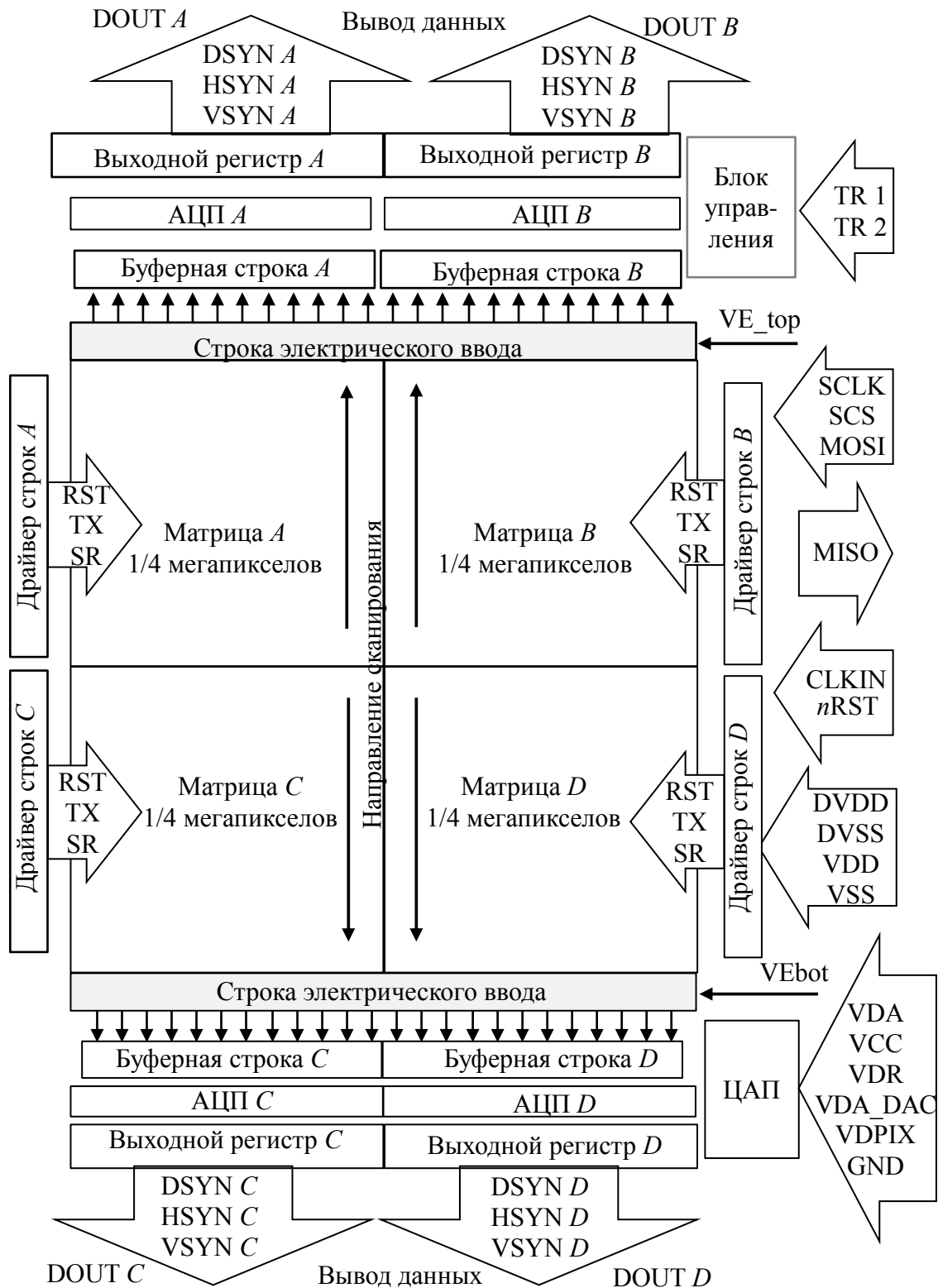


Рис. 3.7. Функциональная блок-схема КМОП-матричного фотоприёмника (не показаны кольцевые шины земли и питания): DVDD – питание 3,3 В; VDD – питание 1,8 В цифровых блоков; VDA – питание 1,8 В аналоговых блоков; VDD – питание 1,8 В цифровых блоков; GND – общая шина; VSS, DVSS – общая шина цифровых блоков; VCCA – питание 3,3 В аналоговых блоков; VSYN – кадровый синхросигнал; HSYN – строчный синхросигнал; DSYN – сигнал сопровождения данных; DOUT – выходная шина данных

преобразование выполняются независимо в каждом столбце. Распараллеливание повышает скорость аналого-цифрового преобразования при малом энергопотреблении. Фотоприёмное поле также содержит строки электрического ввода для проверки тракта обработки сигнала фотоприёмника. Блок управления (блок тайминга) через систему внутренних регистров и блоков выполняет ряд задач:

- принимает сигналы внешнего управления по последовательному 4-проводному двунаправленному SPI-интерфейсу;
- управляет взаимодействием и синхронизацией блоков сенсора;
- формирует сигналы управления накоплением, чтением, разверткой матрицы пикселей;
- формирует сигналы управления буферной строкой, оцифровкой и выводом видеосигнала;
- формирует сигналы управления блока ЦАП;
- организует сигналы выходных данных на шинах DOUT (A, B, C, D) с соответствующими сигналами сопровождения и синхронизации.

Техника извлечения информации из светового потока давно освоила передачу не только интенсивности света, но и цвета. Реализацию принципа накопления качества информации с учётом различных спектральных компонентов можно рассматривать с позиций либо визуального наблюдения, либо автоматического различения объектов, к которому методологически примыкает анализ многоспектральных изображений дешифровщиком, использующим различные методы компьютерного анализа изображений. Общим для этих различных классов систем является реализация накопления в различных участках спектра и *развёртки* изображения. Вне зависимости от числа спектральных зон (в традиционном цветном телевидении их три, в спектральных системах может быть больше – даже на два порядка) для передачи в канал связи часто используют мультиплексирование (перемежение) спектральных сигналов с соответствующим *прореживанием накопления*. Так, во времена электронно-лучевого телевидения применялись системы с последовательной передачей красного (R), зелёного (G) и синего (B) полей; в настоящее время широко применяются фотоприёмные матрицы с цветными фильтрами, обеспечивающими различную спектральную чувствительность соседних пикселей (наиболее распространены фильтры Байера, использующие основные цвета R, G, B, но известно и много других типов фильтров). С точки зрения реализации принципа

накопления важно, что оба эти типа передачи спектральной информации связаны с потерей разрешающей способности либо по времени, либо по полю. Не останавливаясь на всём многообразии методов формирования сигналов цветного и спектрального телевидения, рассмотрим лишь один важный фрагмент этой задачи: фотоприёмные матрицы, позволяющие накапливать информацию о распределении света в нескольких *параллельных* каналах и в пределах замкнутых телевизионных систем параллельно передавать информацию о спектральном составе изображения. Эта цель достигается в *фотоприёмниках с цветоделением на потенциальных барьерах в кремнии*. Подавляющая часть (~99 %) матричных цветных фотоприёмников содержит органические цветоделительные фильтры, наносимые на поверхность кремния в fotocувствительных областях (рис. 3.8, а).

Обычно наносятся три типа фильтров, каждый из которых пропускает голубой, зелёный или красный компоненты спектра. Фильтры организованы в квадратные блоки по четыре элемента, имеющие один голубой, два зелёных и один красный фильтры (RGGB). Каждый такой блок обеспечивает приём цветного сигнала практически всего видимого спектра. Все блоки организованы в мозаичную конструкцию, называемую схемой Байера. Фотоприёмники с такими фильтрами имеют хорошую воспроизводимость параметров, простую конструкцию кремниевой части пиксела, но требуют тщательного обращения из-за возможности повреждения органического фильтра, подвержены временной деградации параметров в результате старения фильтра (особенно под воздействием радиации), имеют существенно более высокую себестоимость по сравнению с обычными интегральными схемами (ИС).

Разработка многоканальных ПЗС в 1970-е гг. привела к мысли о возможности использования фундаментального свойства полупроводников – зависимости глубины поглощения света от его длины волны, что дало толчок к созданию цветных фотоприёмников с внутренними спектральными фильтрами (рис. 3.8, б).

Многоканальные ПЗС не позволяли достаточно хорошо разделять носители заряда, возбуждённые светом на разной глубине от поверхности кремния. Достоинствами многоканальных КМОП-фотоприёмников являются не только возможность разделения сигналов по длине волны, но и временная стабильность параметров и лучшая чувствительность.

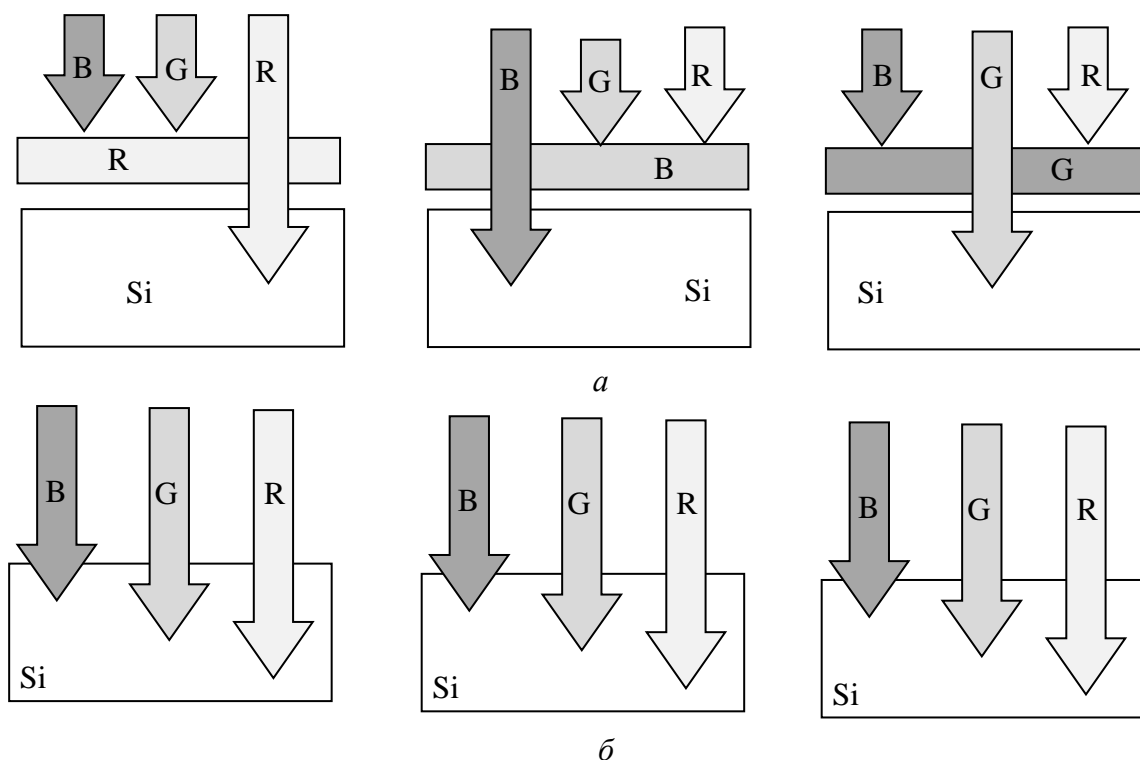


Рис. 3.8. Сравнение принципов цветоделения: *a* – с помощью спектральных фильтров; *б* – с помощью разделения по глубине поглощения фотонов разной длины волны

Первое коммерческое изделие, использующее принцип внутреннего цветоделения, создала компания «Foveon» (см., например, [111], [124]). Она разработала пиксел, состоящий из нескольких вертикально расположенных *p-n*-переходов, разделяющих носители, генерированные на разной глубине под поверхностью кремния. При этом каждый пиксел выдает три аналоговых сигнала, которые после соответствующей обработки преобразуются в стандартные цветовые компоненты. Компания «ЮникАйСиз» в 2001 г. начала разработку другого варианта КМОП-фотоприёмника с внутренними фильтрами на основе энергетических барьеров, созданных полупроводниковыми слоями с разными уровнями легирующей примеси. Была разработана конструкция пиксела, представленная на рис. 3.9 и защищённая рядом патентов (например [124]). Интерес к этому направлению в мировой фотонике к настоящему времени не ослаб [125].

Пиксел имеет три отдельных выхода для каждого из компонентов цветового сигнала – голубого, зелёного, красного. Разделение цветовых компонентов близко к цветочувствительности человеческого глаза, о чём свидетельствуют относительные спектральные характеристики квантовой эффективности η (рис. 3.10, *a*). На рис. 3.10, *б* показана относительная характеристика свет–сигнал, свидетельствующая о достаточной линейности и близких уровнях насыщения для всех трёх цветовых компонентов сигнала.

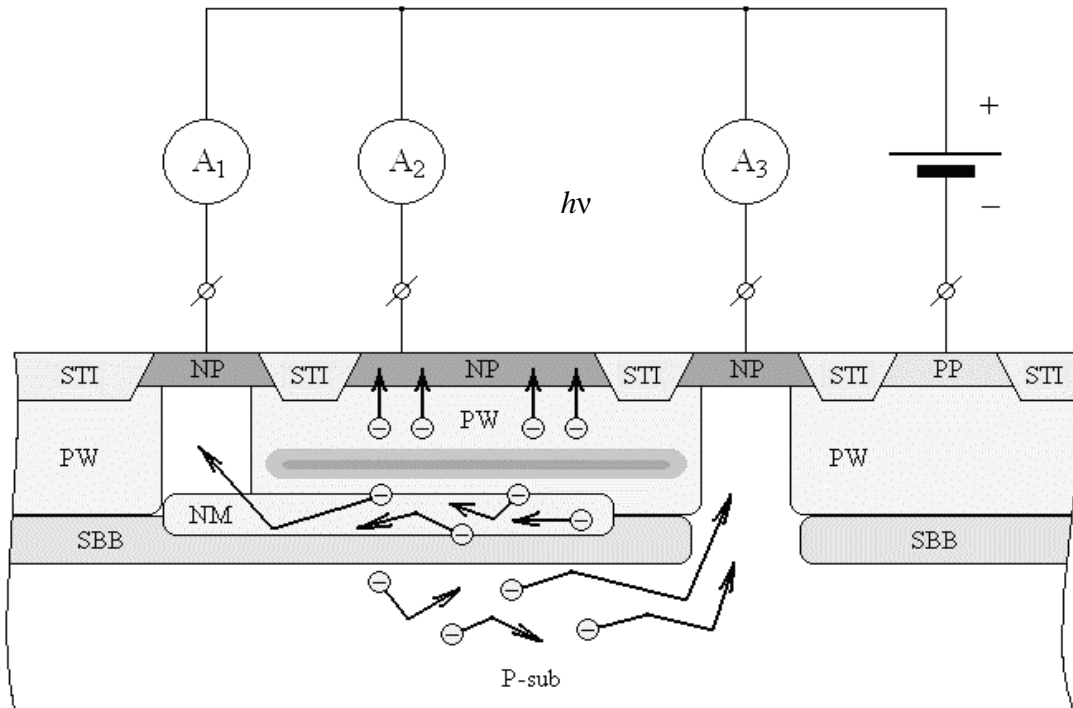


Рис. 3.9. Разрез фотоприёмной структуры с внутренним цветodelением: STI – поверхностная изоляция SiO_2 ; NP – зона легирования канала сбора поверхностных носителей; PP – зона легирования контакта к охранной зоне пиксела; PW – легирование охранной зоны по границе пиксела; SBB – зона легирования барьера между средними и глубокими носителями; NM – легирование канала сбора средних носителей

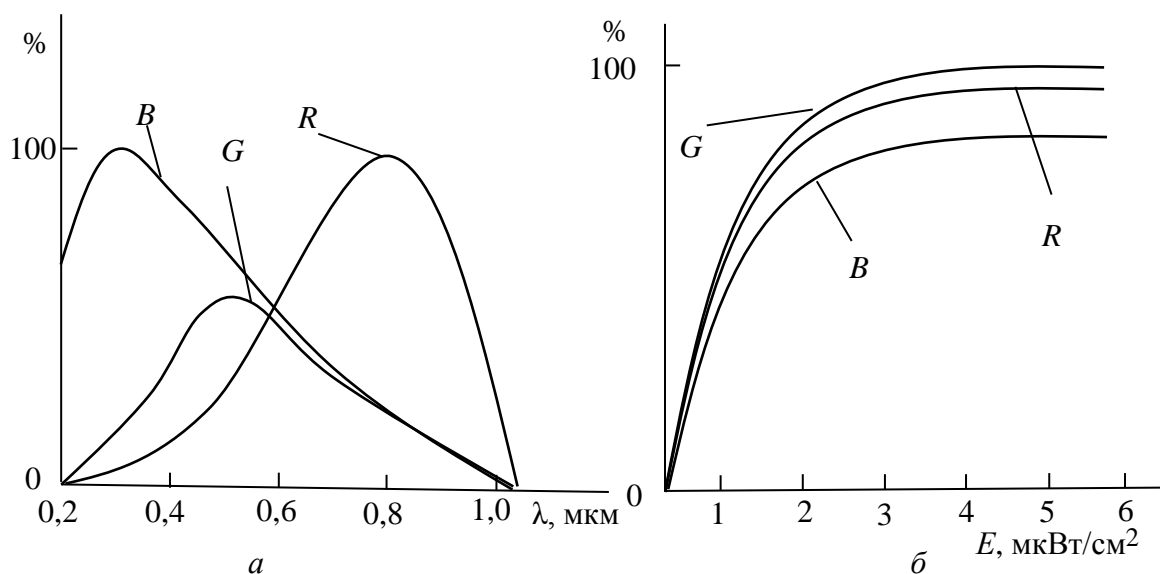


Рис. 3.10. Относительные характеристики матрицы: а – спектральные; б – свет-сигнал для $\lambda = 540$ нм

На основе разработанной конструкции пиксела созданы несколько вариантов фотоприёмников с разрешением от $320 \times 240 \times 3$ до $800 \times 612 \times 3$, что эквивалентно разрешению от 0,3 до 2,1 мегапиксела с байеровским фильтром. Все они имеют встроенный пиксел-процессор, производящий пред-

варительную обработку видеосигнала и доведение его до стандартного формата. Исследования этих фотоприёмников выявили следующие их преимущества по сравнению со стандартными с байеровским фильтром: сильное ослабление муара на периодических рисунках; лучшая цветопередача, особенно заметная на близких цветовых оттенках; простая реализация черно-белого режима за счет суммирования аналоговых сигналов непосредственно в пикселах, что обеспечивает повышенную чувствительность в этом режиме. Небольшое количество дополнительных технологических операций, требуемое для изготовления внутренних энергетических барьеров, а также возможность использования стандартной технологии сборки кристаллов в корпуса без опасности повреждения цветных фильтров обеспечивают низкую себестоимость этих фотоприёмников.

Накопление информации с цветоделением по глубине поглощения фотонов могут быть использованы для систем технического зрения с регулируемой спектральной чувствительностью за счёт изменения коэффициентов усиления по каждому из спектральных каналов. Рассмотренная структура пиксела, основанная на независимой обработке сигнала от каждой из вертикально совмещённых областей поглощения, при минимальных вычислительных затратах позволяет *динамически управлять спектральной характеристикой* системы без использования спектральных фильтров и исключать или усиливать отдельную область спектрального диапазона для практических приложений со спектрально-зависимыми шумовыми или фоновыми свойствами. Типичным является адаптивное к состоянию атмосферы исключение коротковолновой части спектра для увеличения контрастности (см. 4.6), а также анализ изображений в узкой спектральной зоне.

Применение цветоделения по глубине поглощения фотонов имеет важные особенности. Во-первых, для коротковолновой УФ и голубой частей спектра, имеющих наименьшую глубину поглощения, существенным является влияние «встроенных» поверхностных состояний на поверхности кремния или границе кремния с оксидом. Центры генерации заряда в точках нарушения кристаллической структуры дают постоянную засветку отдельных пикселей, так называемые белые точки, которые на полноцветном изображении на самом деле синие.

Во-вторых, большое количество медленно перезаряжающихся состояний в запрещённой зоне вызывает смаз движущихся объектов и послесвечение изображения источников света. Для подавления этих эффектов используется дополнительное поверхностное легирование. Для красной и

ИК-области чувствительности кремния характерен эффект растекания заряда между соседними пикселями вследствие большей глубины поглощения, сравнимой с размером пиксела. Для достижения максимальной контрастности необходимо глубоким легированием сформировать потенциальный барьер в виде сетки по границам пикселей, а также отсечь носители, генерируемые глубже определённого уровня, позволяющего рассчитывать на их попадание в конкретный пиксел, и собираемые подложкой. Таким способом можно достичь максимального расчётного уровня контрастности, пожертвовав долей процента самой коротковолновой и длинноволновой частей светового потока. Подобные вертикально совмещённые структуры характеризуются и максимальной разрешаемой пространственной частотой, в то время как для традиционного цветного фотоприёмника с внешними светофильтрами по схеме Байера кроме 30 % эффективности использования светового потока ещё и пространственное разрешение в 2 раза хуже по синей и красной областям спектра, а по зелёной в $\sqrt{2}$ ниже.

3.2. Технология проектирования фотоприёмных матриц

Разработка оборудования для телевизионной и оптоэлектронной техники требует развитой элементной базы в виде микросхем фотоприёмников, видеосистем на кристалле и прочих электронных компонентов, а создание специализированной аппаратуры для работы в космическом пространстве выдвигает специфические требования. Для специальной аппаратуры, действующей в условиях дестабилизирующих факторов космического пространства, добавляется требование радиационной стойкости, как это отражено в Федеральной целевой программе «Развитие электронной компонентной базы и радиоэлектроники» на 2008–2015 гг. В ней для разработки комплекса моделей расчёта радиационной стойкости электронной компонентной базы и для определения технически обоснованных норм испытаний требуется создание и совершенствование методов моделирования и проектирования радиационно стойких ИС средствами современных систем автоматизированного проектирования (САПР) [113], [126].

В самом начале процесса проектирования видеосистемы разработчик сталкивается с необходимостью выбора её внутренних технических составляющих, которые определяют её общие технические характеристики и качество предоставляемой пользователю информации. Выбор осуществляется между фотоприёмниками различных типов, рабочими спектральными

диапазонами видеосигнала, методами реализации накопления и развёртки изображения, способами оптимизации параметров пространственно-временной дискретизации изображения и используемыми типами каналов связи. Все эти компоненты влияют на техническую реализацию и конкретные схемотехнические и технологические решения, применяемые в видеосистеме на кристалле для обеспечения максимального качества информации для потребителя. В процессе проектирования перед разработчиком стоит задача объединения аналоговых, цифровых, силовых и конструктивных элементов микросхемы в непротиворечивую систему в рамках множества ограничений электротехнического, технологического и конструкционного характера. Преемственность технологических решений и возможность использования отработанных технических решений определяет сроки создания и финансовые затраты, надежность и, в конечном счёте, качество.

САПР является ключевым элементом при проектировании современных ИС, естественно, кроме знаний и опыта самого разработчика. САПР – это высокоинтеллектуальный «станок», который должен обеспечить полный маршрут проектирования, начиная от замысла и заканчивая выводом топологии интегральной схемы для передачи на фабрику изготовителя кристаллов. Большинство пользователей слабо представляют внутреннее устройство САПР, но они чётко знают, что хотят от неё получить. Современные методы и способы автоматизированного проектирования настолько сложны, что профессиональное освоение возможностей системы может занимать от 2 до 5 лет. Зачастую даже профессиональные пользователи с многолетним стажем не знают до конца всех возможностей САПР, с которой они работают. Выбор интуитивно понятной системы, которая позволила бы разработчику, в том числе не обладающему большим опытом работы, быстро и без «лишних» вопросов создавать законченные устройства с предельными характеристиками, – это, пожалуй, одна из основных задач, которую необходимо решить при начале проектирования микросхем.

САПР является концентрированным опытом разработки изделия, включая все аспекты его работы: цель, физику функционирования, математические методы моделирования, различные методики проектирования, графические интерфейсы и т. д. Возможность создания отечественной САПР [126], как и любого штучного высокотехнологичного станка, – основа технологической безопасности страны и показатель высокого уровня российских инженеров.

Для разработки микросхем КМОП-фотоприёмников и видеосистем на кристалле требуется модификация стандартного процесса проектирования микросхем, предлагаемого основными производителями САПР. Для каждой конкретной технологии изготовления кроме комплекта библиотек стандартных цифровых элементов и набора аналоговых блоков необходимо получить модели фоточувствительных приборов, фотодиодов и фототранзисторов. Структура и модели интегральных полупроводниковых приборов рассчитываются на основе физического моделирования в программах типа T-CAD (рис. 3.11) и проходят длительную опытную апробацию в составе тестовых партий изделий, необходимых для настройки перечисленных технологических параметров, влияющих на их электрические характеристики. Качество получаемой информации – спектральная характеристика, шумы, разброс коэффициента преобразования – во многом определяются структурой приборов, которую мы имеем возможность изучить при физическом моделировании.

При выборе инструментария разработчики могут использовать и российскую систему SYMICA (www.symica.ru), позволяющую применять модели элементов различных производителей. Разработанная в Российской Федерации САПР SYMICA позволяет проектировать и моделировать ана-

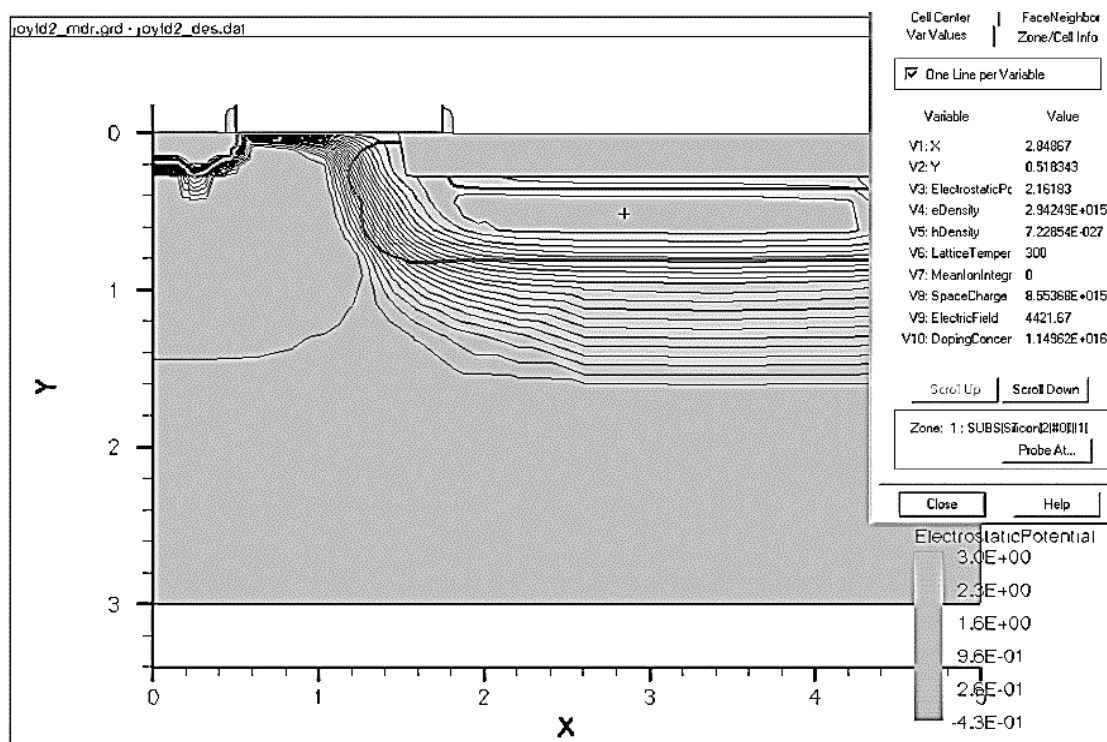


Рис. 3.11. Этап физического моделирования распределения потенциала в структуре пиксела в САПР T-CAD

лого-цифровые СБИС, в том числе и микросхемы фотоприёмников. Рассмотрим типовой маршрут проектирования микросхемы на основе САПР SYMICA, однако реализованный аналогично в САПР любого поставщика (рис. 3.12).

Новый проект создается в подсистеме графического редактора Capture и имеет знакомую для разработчика микросхем архитектуру: библиотека – ячейка – вид. Такая структура позволяет организовать смешанный проект прозрачным образом, когда разработчик в любой момент может заменить одно представление элемента на другое. Capture является полнофункциональным редактором электронных схем с полным набором инструментов для построения сложных проектов. Разработчик может использовать библиотеки элементов в различном представлении, а также транслировать части других



Рис. 3.12. Маршрут проектирования цифроаналоговых ИС в САПР SIMICA

проектов, которые могут быть использованы в новой разработке. Графический редактор позволяет работать как с иерархическим логическим представлением, так и проводить дискретизацию до библиотечных элементов или на транзисторном уровне.

На зарубежном рынке основными потребителями крупных САПР являются большие корпорации. Однако любой разработчик микросхем обычно вынужден приобретать минимальный пакет САПР от крупного поставщика, что обусловлено необходимостью использовать комплект разработчика (*Process Design Kit, PDK*) от фабрики-производителя кристалла. Кроме того, стандартные комплекты САПР от Cadence, Synopsys и Mentor Graphics не поддерживают опции фотоприёмников для КМОП-проектирования; производители микросхем передают разработчикам данные по технологии фотоприёмников под соглашение о неразглашении информации, так как она является коммерческим преимуществом данных производителей и для освоения фотоприёмников в производстве были затрачены существенные средства.

Выбор какой-либо фабрикой в качестве основного САПР для разработки PDK какого-либо крупного зарубежного производителя автоматически заставляет всех клиентов этой фабрики использовать данную САПР. Стоимость одного рабочего места разработчика в этом случае начинается от нескольких сотен тысяч долларов и может превысить 1 млн дол. Приобрести такие лицензии в российских условиях могут позволить себе только крупные компании, которые практически все работают на государственный заказ. Мелкие и средние компании вынуждены пользоваться «серыми» лицензиями на ПО.

Значительную долю в прибыли САПР-компаний составляют дополнительные услуги по обслуживанию и настройке программного обеспечения; выпуску на заказ инструментов, учитывающих специфические требования заказчика; разработке IP-блоков и др. Фактически данный сервис является не дополнительным, а необходимым для любого поставщика САПР.

Определив систему внешних воздействий, разработчик переходит к этапу схемотехнического моделирования в подсистеме Spice. Для обеспечения моделирования сложных смешанных аналого-цифровых систем на кристалле в приемлемое для разработчиков время SYMICA предоставляет возможность так называемого смешанного моделирования, когда цифровые блоки микросхемы могут быть описаны на языке логического моделирования Verilog, а аналоговые подсистемы представлены на транзисторном уровне или описаны на языке Verilog-A. Этот подход также позволяет реа-

лизовать методологию проектирования «сверху вниз», когда сначала отдельные блоки системы описываются на языке высокого уровня, отрабатывается работа системы в целом с нужными характеристиками, а затем проводится реализация отдельных блоков на транзисторном уровне, что сокращает сроки и повышает надежность проектирования. Для обеспечения этого в систему проектирования SYMICA встроены поддержка разных вариантов представления подсхем (ячеек), поддержка языков Verilog и Verilog-A, автоматическое формирование интерфейса между цифровой и аналоговой частями схемы и поддержка совместной работы цифрового и аналогового симуляторов.

Для визуализации и анализа полученных результатов моделирования в графическом виде используется подсистема отображения Probe. Поскольку при моделировании современных микросхем результаты моделирования могут занимать десятки гигабайт, при разработке Probe большое внимание было уделено скорости работы с данными и их последующей обработке. Также в Probe предусмотрено более 50 различных функций обработки аналогового сигнала, включая быстрое преобразование Фурье, индикаторные (EYE) диаграммы, гистограммы и другие виды анализа.

По завершении итерационного процесса оптимизации параметров электрической схемы разработчик переходит в подсистему топологического проектирования Layout, позволяющую в полуавтоматическом режиме на основе библиотечных элементов и с учётом правил проектирования для конкретного технологического процесса провести создание задания фабрике для изготовления микросхем в виде стандартного топологического файла. На этом этапе (после получения топологической схемы будущей СБИС) проводится моделирование паразитных емкостей и индуктивностей полученной структуры. Проект, учитывающий эти явления, возвращается в подсистему моделирования для проверки функционирования схемы и соответствия требуемым параметрам. Проведя проверки и оптимизацию топологической схемы, разработчик может формировать задание на изготовление комплекта фотошаблонов для процесса фотолитографии изготовления ИС.

Такой процесс проектирования микросхем может занять значительное время и требовать больших вычислительных ресурсов. Благодаря применению оригинальных математических методов SYMICA обеспечивает более высокую скорость моделирования при сохранении заданной точности.

В настоящее время технические требования, предъявляемые к аппаратуре систем управления и контроля различного назначения (ракетно-

космической и военной техники, связи и энергетики, а также научного физического эксперимента), определяют возможность их эксплуатации в условиях воздействия ионизирующих излучений. Выделяют два типа ионизирующих излучений – стационарные и импульсные. Среди радиационных факторов естественного и искусственного происхождения наибольшим поражающим действием по отношению к микроэлектронной элементной базе аппаратуры по энергетическому критерию обладает импульсное ионизирующее излучение. В ряде применений разработчику достаточно учитывать воздействие стационарного ионизирующего излучения.

Задача проектирования радиационно стойких схем требует использования специализированных методов и программ моделирования, учитывающих поведения схем при воздействии ионизирующего излучения. Доступные импортные САПР позволяют оценить влияние стационарного ионизирующего излучения только косвенным образом через процедуру подгонки параметров транзисторов под экспериментальные данные. Однако, как отмечают сами разработчики, этого часто бывает недостаточно, поскольку, получив отрицательный результат, разработчик не находит ответа на вопрос: «Как надо модифицировать схему, чтобы соответствовать предъявляемым требованиям?». В результате процесс проектирования по существу является методом «проб и ошибок», который может и не завершиться в намеченные сроки. Таким образом, можно утверждать, что в настоящее время в России отсутствуют полноценные современные средства моделирования поведения ИС с учётом ионизирующих излучений. Зарубежные САПР ИС общего назначения не обладают специальными инструментами для схемотехнического моделирования радиационных эффектов. Такие средства разрабатываются специализированными фирмами и не предназначены для продажи. В сложившейся ситуации поддерживаемая отечественными разработчиками САПР SYMICA может служить основой для проведения таких работ.

3.3. Учет влияния различных параметров на этапе проектирования фотоприёмников

Формирование зоны сбора фотоэлектронов является основой, определяющей характеристики фотоприёмника. В отличие от употребляемого в других главах термина «площади накопления» при проектировании фотоприёмной СБИС необходимо рассматривать все процессы, происходящие в

объёме как кремния, так и всех формируемых фотолитографией поверхностных слоёв [127], [128].

О необходимости отсечки шумов поверхностной генерации и наводок, генерируемых остальными частями микросхемы, которые имеют тенденцию распространяться через подложку и шины питания, а также шумов пространственной фильтрации ИК-сигнала необходимо помнить на этапе разработки конструкции самой фоточувствительной области. На этапе проектирования перед разработчиком стоит задача обеспечения требуемого быстродействия видеосистемы, эффективности работы тракта обработки видеосигнала, энергоэффективности системы на кристалле при сохранении качества информации. В зависимости от конкретных приложений у разработчика появляется возможность многофакторной оптимизации как по динамическому управлению параметрами захвата видеоизображения, так и по архитектурной компоновке видеосистемы на кристалле. Например, контрастность изображения напрямую связана с кадровой частотой. Ёмкости перезарядки в пикселе зачастую определяют максимальную скорость считывания, но ограничивают максимальный заряд фотоэлектронов, определяющий уровень насыщения, а цифроаналоговые преобразователи выбираются из возможности достижения практического динамического диапазона, определяющегося по минимально достижимому уровню шумов в видеосистеме. Таким образом, перед разработчиком стоит задача оптимизации схемы считывания пиксела в части ёмкости накопления фотоэлектронов и схемы преобразования заряд–напряжение. На следующем этапе, оценив получаемый сигнал от пиксела, разработчик выбирает АЦП для преобразования видеосигнала. Уровень шумов определяет шаг преобразования, а максимальное значение сигнала – его динамический диапазон. Для видеосистем на кристалле в зависимости от требуемого быстродействия типичным является диапазон 8...12 бит.

Отдельная область повышения качества информации – её защита от различных внешних воздействующих на микросхему факторов. В первую очередь это электростатические разряды и радиационное воздействие. Наиболее полно комплекс внешних воздействующих факторов можно рассмотреть на примере разработки фотоприёмных электронных компонентов для аппаратуры космического назначения.

Рассматривая комплекс дестабилизирующих факторов космического пространства [129]–[133], мы видим целый набор воздействий, различными способами влияющих на фотоприёмную аппаратуру и ведущих как к

потере полезной информации, так и к возникновению ложной информации, что критически снижает качество передаваемой информации. Космические аппараты подвергаются воздействию ионизирующего излучения Солнца; на низких орбитах необходимо учитывать воздействие разреженной плазмы, создаваемой атмосферой Земли; для всех космических аппаратов существенным является воздействие галактического излучения.

Большая часть энергии солнечного излучения сосредоточена в видимом диапазоне. Суммарная мощность излучения Солнца в оптическом диапазоне составляет 1400 Вт/м^2 . Кроме оптического излучения от Солнца идет поток продуктов термоядерной реакции. В составе потока 90 % – протоны, 7 % – альфа-частицы, 1 % – атомы и ионы тяжелее гелия, 1 % – электроны. Также в излучении присутствуют и гамма-кванты. Нейтронов в потоке почти нет. На Солнце происходят вспышки и крупные извержения, так что во времени поток частиц неоднороден. Вспышки имеют регулярный характер и повторяются в среднем через 20 дней. Извержения происходят реже одного раза в год. При вспышке поток частиц возрастает в несколько тысяч раз, а при извержении – в миллионы раз. Средняя энергия протонов, излучаемых спокойным Солнцем, – $600 \dots 800 \text{ эВ}$; электронов – $0,3 \dots 0,4 \text{ эВ}$. Во время вспышек энергия частиц возрастает и достигает 50 МэВ для протонов и 5 МэВ для электронов. Во время извержений энергия частиц достигает 10^{13} эВ [130].

Если говорить о галактическом излучении, то его интенсивность очень мала и сравнима с яркостью звезд по отношению к Солнцу. Состав галактического излучения: протоны, электроны, гамма-кванты и тяжелые атомы. Это излучение постоянно во времени и однородно во всех направлениях [130]. Проблема галактического излучения определяется его спектральным составом. В галактическом потоке присутствуют частицы с энергией до $3 \cdot 10^{21} \text{ эВ}$, что соответствует 50 Дж и эквивалентно энергии пули мелкого калибра. Плотность потока частиц с энергией больше 10^{10} эВ не превышает $3 \cdot 10^4 \text{ 1/(м}^2 \cdot \text{с)}$, а для частиц больше 10^{20} эВ – не более $10 \text{ 1/(м}^2 \cdot \text{с)}$.

Говоря о воздействии атмосферы Земли, отметим, что химически активными компонентами плазмы являются ионы и атомы кислорода. Плотность кислорода на орбите $10^6 \dots 10^8 \text{ атомов/см}^3$. Атомарный кислород разрушает полимерные покрытия и окисляет металлы. Разрушение полимер-

ных покрытий стимулируется ультрафиолетовым излучением Солнца, а неорганические продукты реакции защищают поверхность полимерных слоев и тормозят дальнейшее разрушение [133].

Среди прочих факторов отметим термоциклы, которые оказывают общее негативное воздействие на аппаратуру космических аппаратов за счёт создания циклически меняющихся термомеханических напряжений в электронных приборах и элементах конструкции. Фотоприёмная аппаратура выполняется в термостабилизированных корпусах, однако есть опасный процесс, вызывающий общие отказы электронной аппаратуры. Это рост нанокристаллов из оловянных сплавов на электронных платах и компонентах. Термомеханические напряжения являются основным источником энергии для роста нанокристаллов, которые прорастают сквозь изолирующие полимерные покрытия шин питания и кабелей связи и вызывают короткое замыкание проводников или становятся слабым местом для электростатического пробоя. Поэтому в бортовой аппаратуре космических аппаратов всячески избегают лужения контактов и выводов, применяя покрытие свинцом или золотом.

В космосе и, в частности, на околоземной орбите присутствует много частиц твёрдого вещества – микрометеоритов. Их характерные размеры – единицы и десятки микрометров, и они редко превышают 100 мкм. Скорости, на которых микрометеориты сталкиваются с космическими аппаратами, достигают 70 км/с [131]. При столкновении кинетическая энергия превращается в тепловую, а вещество метеорита и участок поверхности космического аппарата испаряются и ионизируются. Эффективность ионизации зависит от скорости столкновения и оценивается средним значением 500 Кл/кг. Энергия плазмы частично преобразуется в электромагнитное излучение с шириной спектра от рентгеновского до радиочастотного с максимальной плотностью мощности *в оптическом диапазоне*. Попадание микрометеорита регистрируется фотоприёмной аппаратурой как яркая вспышка. На поверхности образуется кратер размером в сотни микрометров. Это особенно существенно для оптических систем, со временем значительно снижается их качество.

Следующим фактором являются статические электрические разряды. До 30 % отказов в работе космических аппаратов на высоких орбитах связано с их электризацией. В ионосфере Земли существует два уровня ионизирующих частиц. Нижний уровень – это поток протонов, верхний – поток электронов. Ультрафиолетовое излучение Солнца также вызывает внеш-

ний фотоэффект, т. е. излучение электронов с внешней поверхности космических аппаратов. Равномерное распределение заряда по металлической поверхности космического аппарата не опасно для электронной аппаратуры. В реальной ситуации поверхность аппарата покрыта слоем полимерного теплозащитного материала, а солнечные батареи – прозрачным экраном. Поток ионов и протонов совместно с фотоэффектом создают на освещённой поверхности положительный потенциал. На теневой стороне аппарат приобретает отрицательный потенциал. Такая разность потенциалов между элементами конструкции из различных материалов может достигать 10 кВ, что приводит к возникновению электростатических разрядов на поверхности аппарата. В результате по конструкции спутника протекают импульсные токи, которые генерируют в проводниках помехи в единицы и десятки вольт, что вызывает сбои в работе электронной аппаратуры и необратимые отказы электронных систем. Частицы с высокой энергией и вторичные нейтроны проникают внутрь корпуса и вызывают электризацию элементов конструкции электронных систем и устройств. В этом случае электростатические разряды воздействуют непосредственно на полупроводниковые приборы электронных систем. Космические аппараты на эллиптических орбитах постоянно меняют полярность электризации, что увеличивает частоту и мощность электростатических разрядов [132]. В некоторых областях околоземного космического пространства концентрация заряженных частиц такова, что эти частицы способны создавать абсолютный и дифференциальные потенциалы на поверхности и внутри космического аппарата до уровня напряжения пробоя. Заряды освобождаются и возбуждают в кабелях электромагнитные помехи. Импульсы проникают в корпус и распространяются по проводникам печатных плат, достигают активных элементов, переключают триггеры, насыщают усилители, плавят проводники в микросхемах, пробивают $p-n$ -переходы и диэлектрики. Эти процессы приводят к сбоям или необратимым отказам бортовых видеосистем. При разработке микросхем следует учитывать возможность электростатических разрядов между выводами корпуса, предусматривая соответствующие блоки защиты, а разработка оптических систем должна включать специализированное антистатическое покрытие оптических компонентов. Воздействие ионизирующего излучения на оптические стекла вызывает их окрашивание и люминесценцию. Первый эффект определяется суммарной поглощенной дозой, причем он слабо зависит от вида воздействующего излучения, а второй – мощностью дозы [132], [133].

Радиационное окрашивание стекол объясняется накоплением в их объёме центров окраски, образующихся в результате захвата электронов и дырок структурными дефектами разных типов. В процессе воздействия ионизирующего излучения ловушки, имеющиеся в стекле и ответственные за образование центров окраски, заполняются электронами и дырками. Это приводит к росту оптической плотности образца, характеризующей суммарное поглощение света во всей его толщине. С увеличением мощности дозы равновесный уровень приращения оптической плотности стекол повышается. Этот эффект становится менее выраженным при увеличении температуры образца за счёт отжига центров окраски. На радиационное окрашивание стекол большое влияние оказывают примеси. Оптические стекла без специальных добавок заметно окрашиваются при дозах $\sim 10^3$ Гр. Защитные стекла солнечных батарей, содержащие оксиды церия (0,5...1,5 %), сохраняют высокую прозрачность при поглощенных дозах до 10^6 Гр. Введение оксида церия в состав стекла перемещает спектр центров окраски в ультрафиолетовую область.

Радиация порождает в фотоприёмниках ложные сигналы, главные из которых – пиксели с аномально большим темновым сигналом («горячие» пиксели) и однокадровые сигналы. Аномально большие темновые сигналы бывают двух видов – точечные дефекты и «столбы» (рис. 3.13) [134].

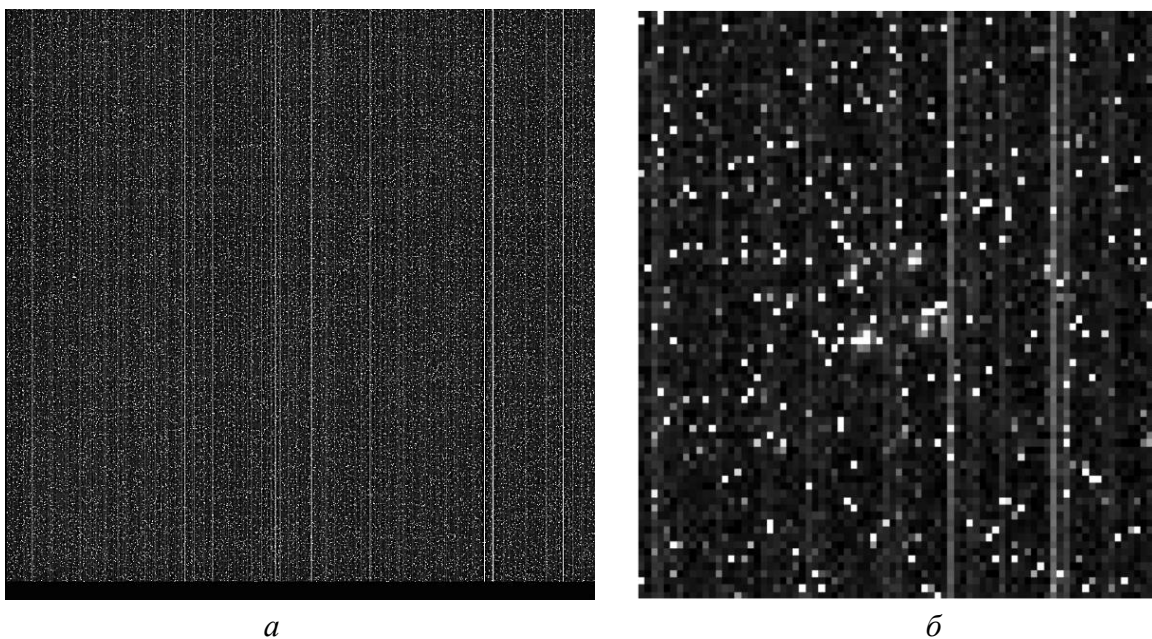


Рис. 3.13. Темновой сигнал после воздействия облучения протонами с энергией 1 ГэВ при дозе, эквивалентной 8 годам эксплуатации телекамеры в космосе:
а – матрица 1 мегапиксел; *б* – фрагмент матрицы с сигналом тестового объекта

«Столбцовые» радиационные дефекты более типичны для матричных ПЗС, но встречаются в КМОП-матрицах. Однокадровые события из-за радиации в основном имеют вид точечных объектов, но могут иметь вид и протяжённых треков (рис. 3.14). Повышение качества информации при воздействии радиации достигается межкадровой обработкой видеосигнала. При этом в системах наблюдения сложных сюжетов, например при контроле стыковки космических аппаратов (см. 5.1), требуется периодическое запоминание темновых сигналов с перекрытием света специальной шторкой.

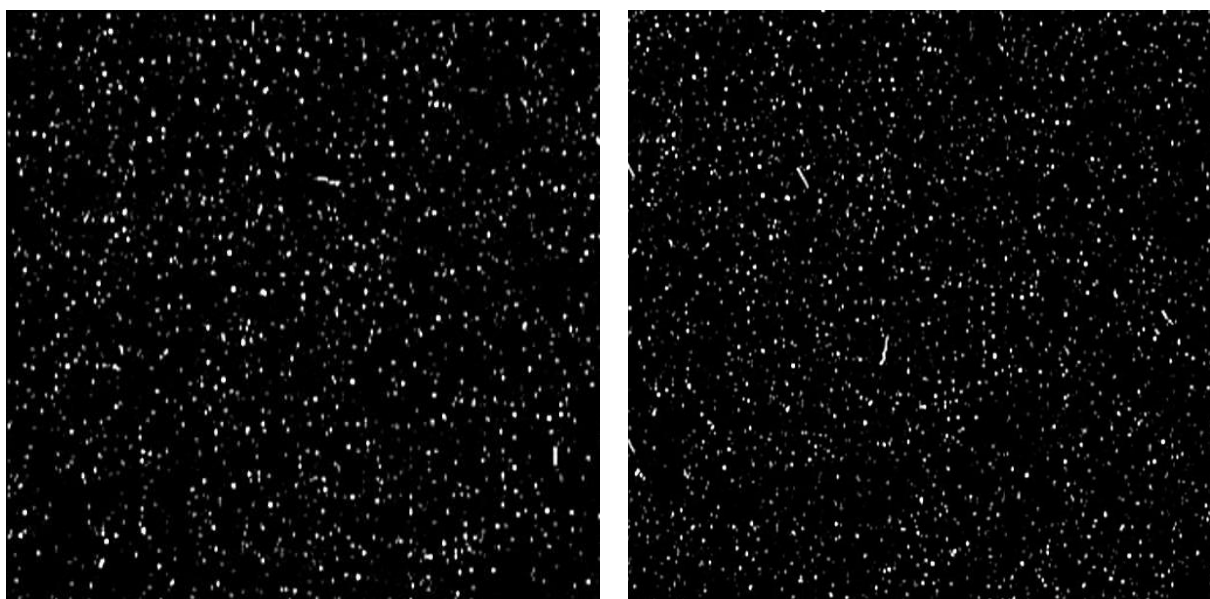


Рис. 3.14. Однокадровые события (точечные и треки) на фоне темнового сигнала во время облучения протонами с энергией 1 ГэВ, ориентировочная плотность потока протонов $10^7 \dots 10^8$ $1/(\text{см}^2 \cdot \text{с})$, два соседних по времени кадра (фрагмент матрицы)

В астродатчиках (см. 4.2, 4.3) имеется возможность подавления темновых сигналов без перекрытия светового потока. Алгоритм подавления ложной информации из-за однокадровых событий в астродатчиках при видимой скорости звёзд менее пиксела за кадр может иметь простой вид дискретного нелинейного фильтра (дискрет – время кадра), имеющего нулевую импульсную характеристику и неискажённую переходную характеристику [134]:

$$Y_{i-1} = \min\{S_{i-1}, \max\{S_i, S_{i+1}\}\}. \quad (3.2)$$

Эффективность алгоритма (3.2) иллюстрируется рис. 3.15, 3.16: постоянные дефекты сохраняются, однокадровые подавляются. При видимых скоростях звёзд, заметно бóльших пиксела за кадр, селекция сигналов звёзд на фоне радиационных сигналов включает внутрикадровую обработку.

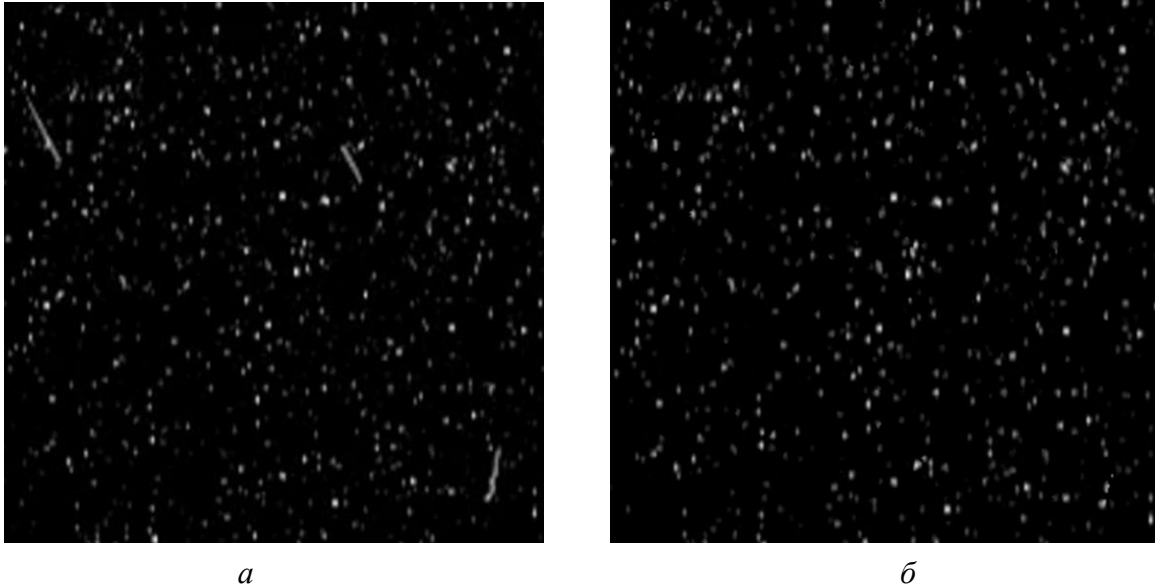


Рис. 3.15. Два соседних по времени кадра: *а* – один из группы анализируемых алгоритмом (3.2); *б* – выходной сигнал алгоритма (фрагмент матрицы)

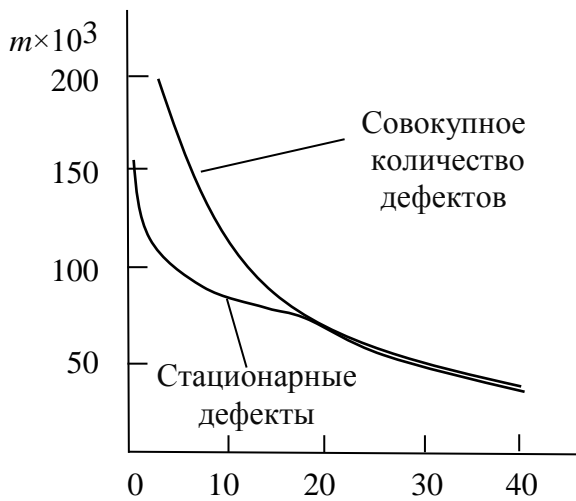


Рис. 3.16. Зависимость количества ложных сигналов от порога

Алгоритм подавления ложной информации из-за стационарных дефектов в астродатчиках (рис. 3.17) может иметь вид дискретного нелинейного фильтра (дискрет – время кадрового интервала), имеющего единичную импульсную характеристику и переходную характеристику, единичную в интервале группы k кадров, достаточную для смещения изображения звёздного поля до полного несовпадения сигналов звёзд в

поле зрения (порядка 10^3 кадров или 1...2 мин). Для снижения коэффициента шума такого режекторного фильтра сравниваемые сигналы оценки фона усредняются по группе n кадров (порядка 8–16 кадров) [134]:

$$Z_i = S_i - \min \left\{ \frac{1}{n} \sum_{i=1}^{-n} S_i, \frac{1}{n} \sum_{i=1}^{-n} S_{i-k} \right\}. \quad (3.3)$$

Вместе с тем первичная обработка видеосигнала для подавления столбцовых ложных сигналов может быть реализована достаточно простыми средствами, в том числе в виде видеосистемы на кристалле, в которой

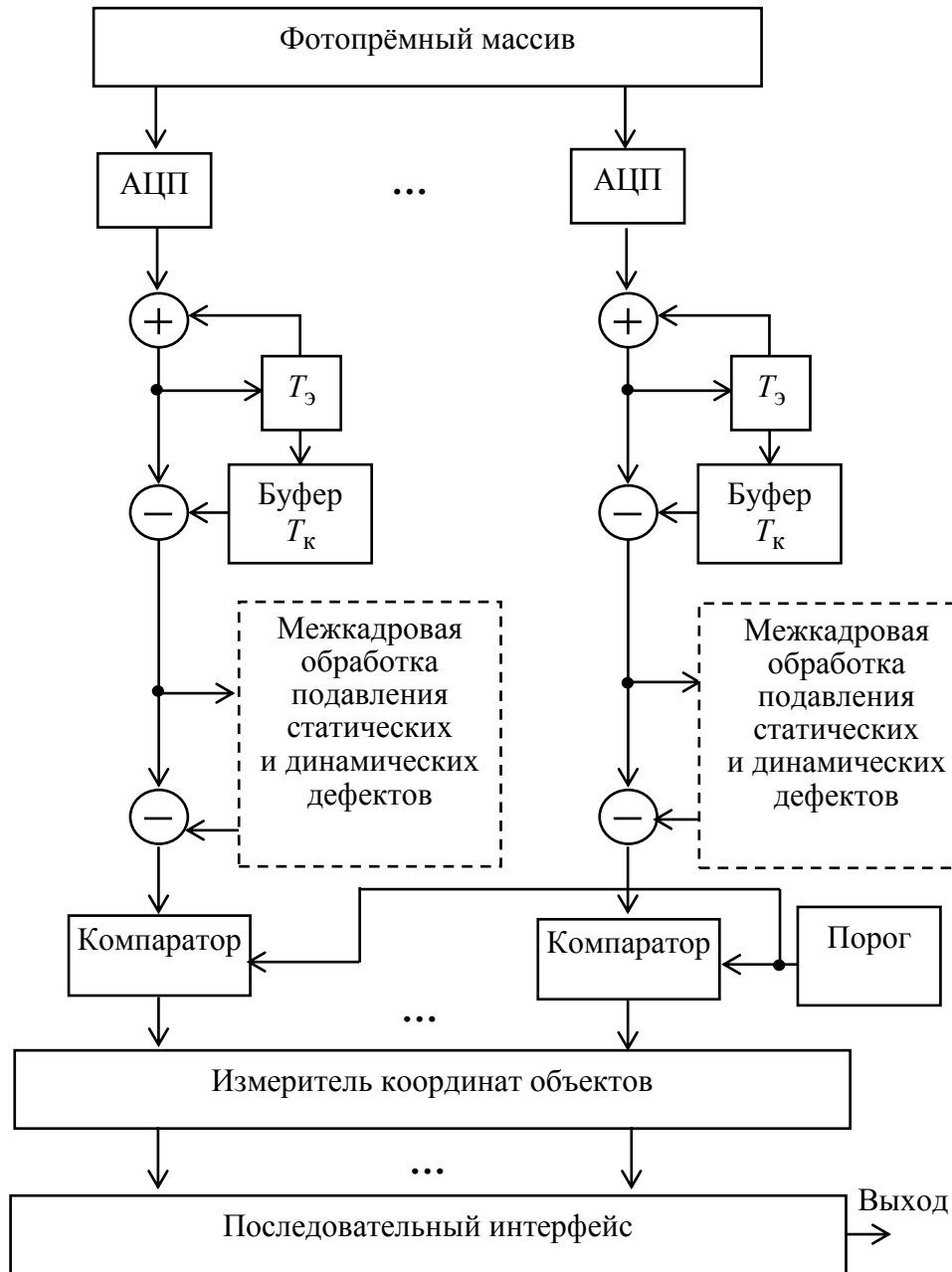


Рис. 3.17. Структурная схема видеосистемы на кристалле с параллельной обработкой сигнала; подавление возникающих постоянных дефектов и однокадровых сигналов выполняется на отдельном кристалле

вычислительная подсистема связана с фотоприёмным массивом множеством параллельных каналов аналогично известному принципу [135]–[137], используемому при кодировании изображений (см. рис. 3.17). Реализация алгоритма последовательного подавления столбцовых, точечных постоянных и однокадровых ложных сигналов [134] позволяет повысить качество информации при выделении сигналов звёзд в астродатчике (рис. 3.18).

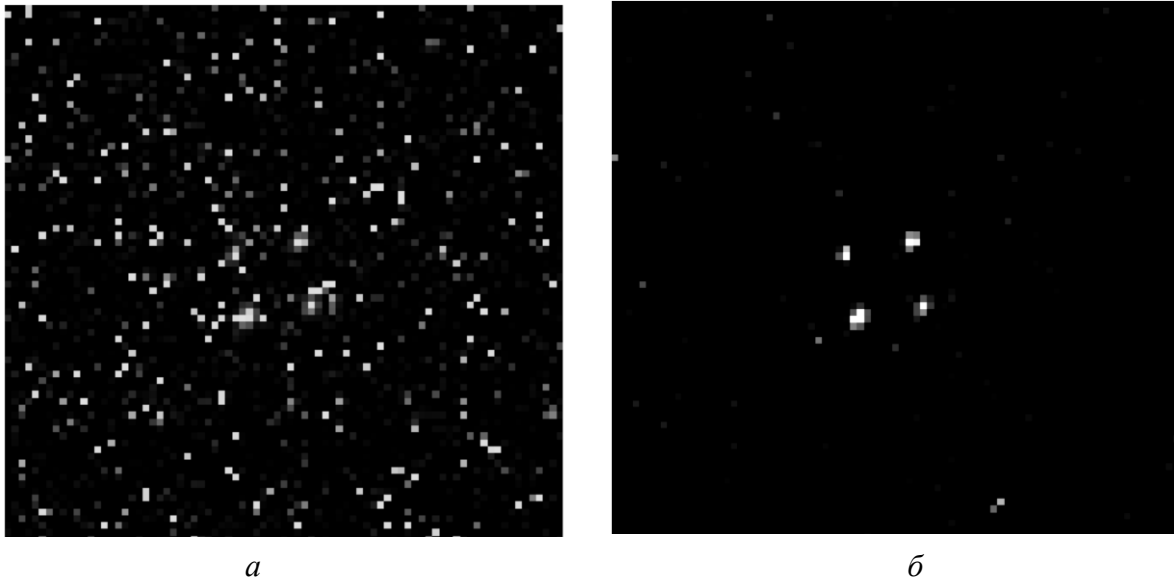


Рис. 3.18. Тестовый сигнал на фоне стационарных радиационных помех после: *а* – подавления столбцовых дефектов с помощью метода (3.3); *б* – подавления стационарных точечных дефектов межкадровой обработкой

Корпус фотоприёмной матрицы, представляющий собой трёхмерную электрическую схему, как и корпус телекамеры, ослабляет внешние воздействия, порождающие ложную информацию. Например, в ходе эксплуатации аппаратуры на околоземной орбите корпус телекамеры при массовой защите $1,5 \text{ г/см}^2$ уменьшает суммарную поглощённую дозу на три порядка.

3.4. Управление параметрами считывания видеоинформации на этапе проектирования фотоприёмной матрицы

Качество фотоэлектрического преобразования принято характеризовать внутренней квантовой эффективностью, определяемой как отношение числа накопленных фотоэлектронов к числу упавших фотонов. Энергетический баланс фотоэффекта описывается формулой Эйнштейна и определяет равенство энергии фотона сумме работы выхода электрона и его кинетической энергии. Квантовый выход при внешнем фотоэффекте из-за различия в значении работы выхода обычно заметно меньше, чем квантовая эффективность внутреннего фотоэффекта.

Так как для внешнего фотоэффекта требуется бóльшая энергия кванта, «красная граница» чувствительности приборов с внутренним фотоэффектом может быть значительно дальше продвинута в инфракрасную область. Используя фоточувствительные структуры на основе кремния, разработчик ограничен практическими значениями внутренней квантовой эффективности, которые не превышают 0,7 для некоторых конструкций фотоприёмни-

ков на лавинных эффектах и на практике находятся в диапазоне $0,1 \dots 0,5$. Таким образом, отношение сигнал/шум не достигает своего теоретического предела.

В фотоприёмниках с непосредственным преобразованием фотонов в электроны фотоэлектрический эффект описывается на основе понятий зонной теории твёрдого тела. В полупроводнике фотоприёмной мишени имеются определённые разрешённые энергетические уровни, на которых могут находиться электроны. Верхняя заполненная электронами зона называется валентной. Следующая за ней разрешённая зона – зона проводимости – может быть и пустой, но электроны в ней появляются под воздействием излучения или нагрева. Такой беспримесный полупроводник имеет достаточно большую работу выхода электрона в зону проводимости, и, соответственно, его «красная граница» лежит в области весьма коротких волн. Фотоэлектрический эффект в кремнии имеет «красную границу» на длине волны $1,1 \text{ мкм}$; для более длинных волн кремний прозрачен и используется в качестве материала для линз инфракрасных объективов.

Все КМОП-сенсоры можно разделить на две группы. В первой из них под действием светового потока идёт накопление потенциала. Фотогенерированные заряды формируют фототок, который разряжает ёмкость фотодиода, заряженную до напряжения питания. Таким образом, потенциал на фотодиоде во время накопления является переменным.

Во втором случае фотодиод находится под постоянным потенциалом и в нём под действием светового потока идёт накопление фотогенерированных зарядов. При считывании накопленный заряд переносится в ёмкость плавающей диффузионной области, где и преобразуется в потенциал. Второй тип КМОП-сенсоров подразумевает использование двух технологий: как КМОП, так и ПЗС.

Первое поколение КМОП-сенсоров носит название «сенсор с пассивным пикселом» (*Passive Pixel Sensor, PPS*). Их особенностью является наибольшая связь с породившими их микросхемами динамической или статической памяти. Данный вид сенсора помимо матрицы фотодиодов содержит регистр, осуществляющий выбор строки; ключевые полевые транзисторы, осуществляющие коммутацию фотодиодов выбранной строки к вертикальным видеопинам; аналоговый мультиплексор, осуществляющий подключение видеосигнала с выбранного фотодиода к выходному усилителю (рис. 3.19).

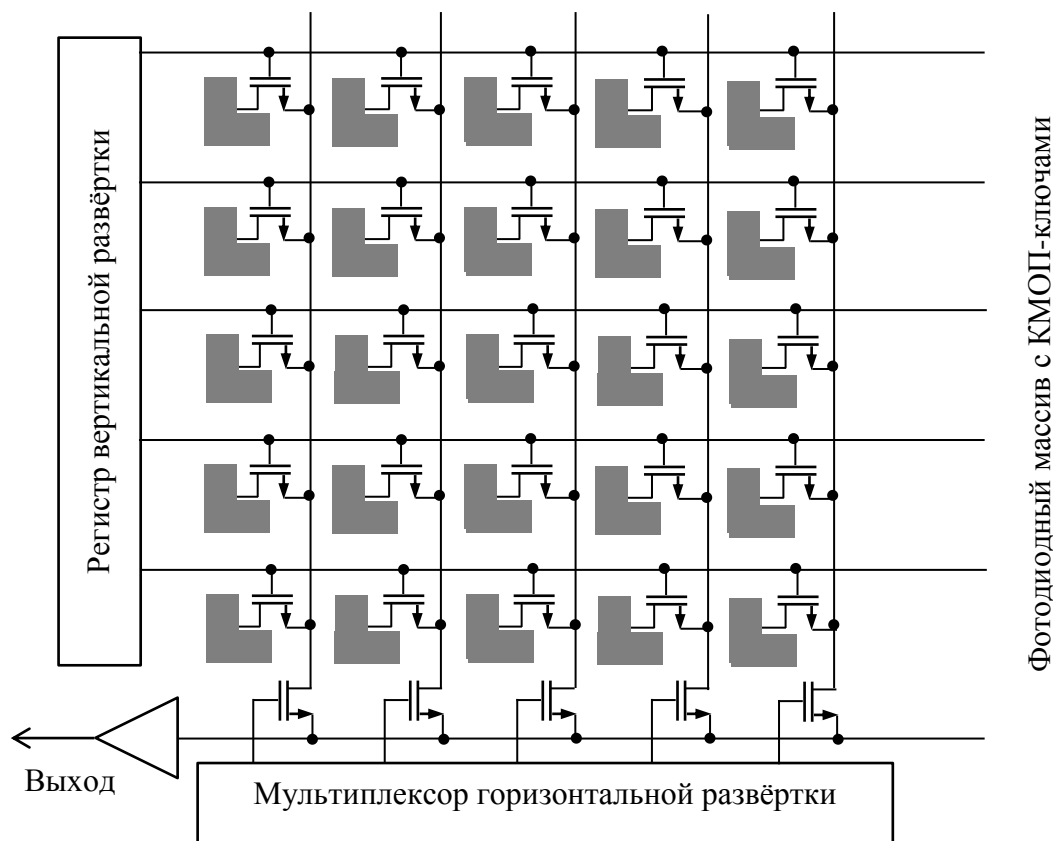


Рис. 3.19. КМОП-фотоприёмная матрица с пассивным пикселом

Высокая ёмкость вертикальной видеошины, суммирующаяся с ёмкостью горизонтальной шины, не позволяет получить приемлемый уровень видеосигнала при ограниченном объёме зарядового пакета, как следствие, данный вид сенсора обеспечивает низкое отношение сигнал/шум. Для снижения влияния ёмкостей введены усилители видеосигнала на каждый столбец (рис. 3.20). Такое решение позволило устранить влияние на амплитуду видеосигнала ёмкости горизонтальной видеошины. Влияние ёмкости вертикальной шины на видеосигнал было преодолено в КМОП-сенсоре с активным пикселом (*Active Pixel Sensor, APS*, рис. 3.21), изобретённого Э. Фоссамом [30] и применяемым и в настоящее время [154].

Революционность этой технологии связана с тем, что усилитель (в простейшем случае однотранзисторный истоковый повторитель) размещается непосредственно в каждом пикселе фотоприёмника. Тем самым ёмкость считывания и ёмкость вертикальной видеошины оказываются разделёнными этим усилителем. При этом ёмкость считывания может иметь тот же порядок, что и в ПЗС – единицы фемтофард. С учётом усилителя коэффициент преобразования заряда в напряжение оказывается не ниже, а даже

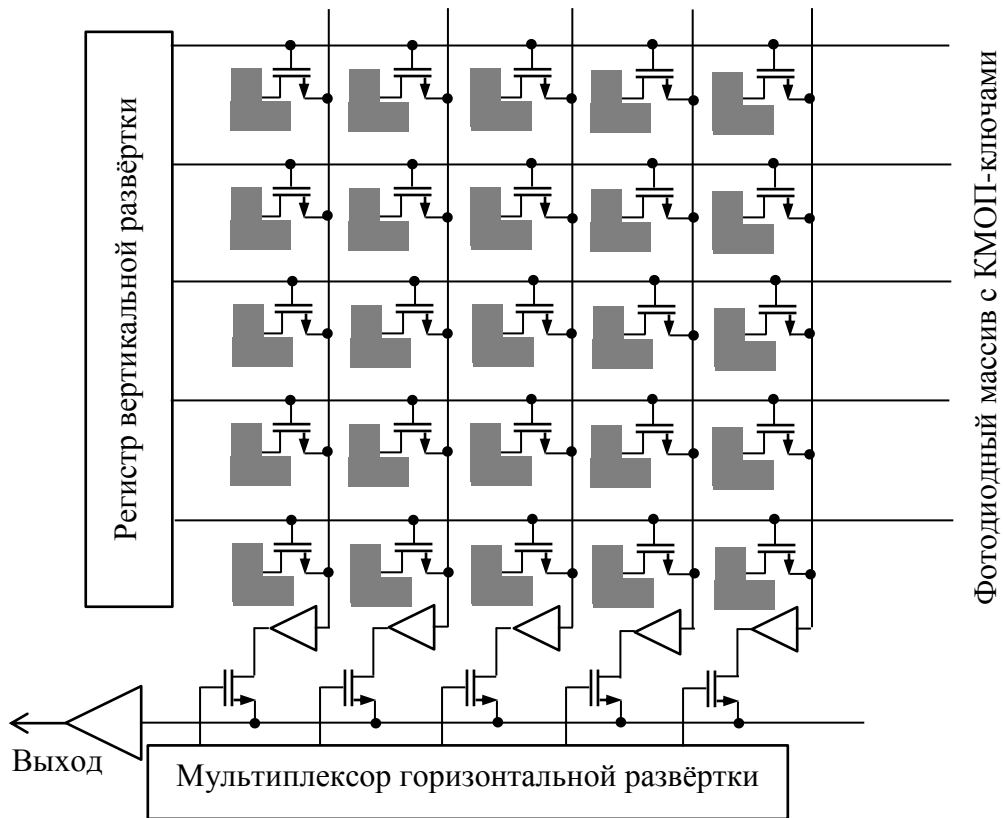


Рис. 3.20. КМОП-фотоприёмная матрица с пассивным пикселом и активным столбцом

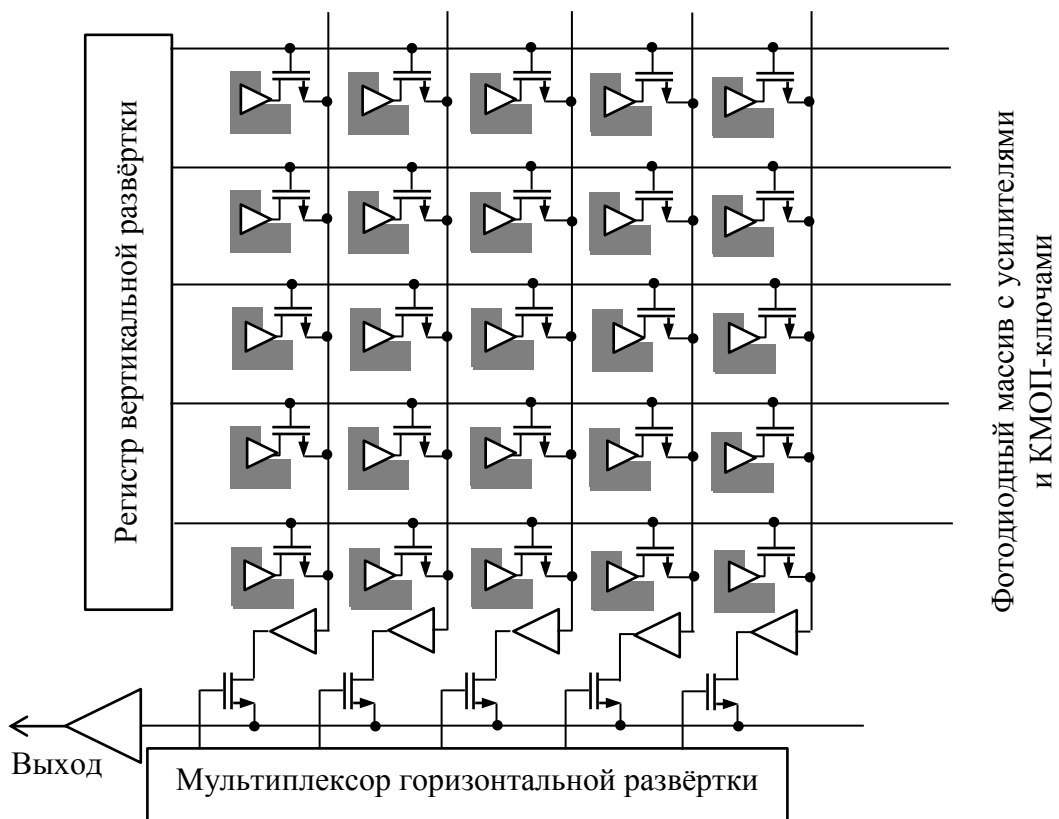


Рис. 3.21. КМОП-фотоприёмная матрица с активным пикселом и активным столбцом

выше, чем в фотоприёмниках на ПЗС. Это в корне решает проблему чувствительности и отношения сигнал/шум. Развитие технологии «активного пиксела» в дальнейшем шло по пути улучшения качественных характеристик фотодиодов и транзисторов, а также по пути увеличения транзисторов в каждом пикселе. Другое важное направление развития КМОП-сенсоров было связано с тем, что, как уже говорилось ранее, КМОП-технология позволяет интегрировать различные узлы телевизионной камеры на тот же кристалл и даёт разработчику возможность использования очень большого числа полевых транзисторов для реализации тех или иных узлов.

Следующим шагом на этом пути была реализация аналого-цифрового преобразователя на том же кристалле (рис. 3.22).

Одной из особенностей КМОП-сенсоров является возможность реализации высокой частоты опроса, превышающей частоту считывания зарядовых пакетов из ПЗС. В связи с этим требования по быстродействию к АЦП, установленному в КМОП-сенсоре, оказываются весьма жесткими. Высокая частота считывания приводит к широкой полосе частот видеосигнала и худшим шумовым характеристикам, а большая частота преобразования АЦП – к большому выделению мощности.

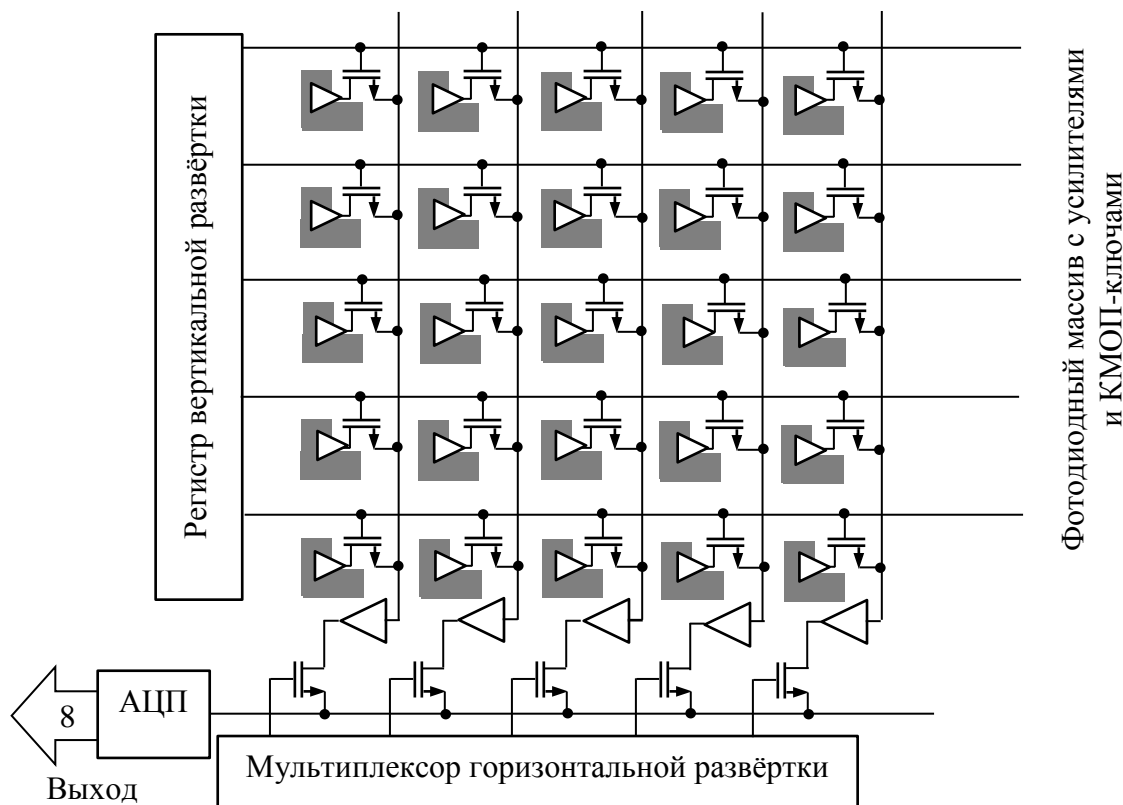


Рис. 3.22. КМОП-фотоприёмная матрица с активным пикселом, активным столбцом и одним АЦП

Для преодоления этого недостатка в КМОП-сенсорах АЦП стали располагать в каждом столбце [115], [116] и коммутацию осуществлять в цифровом виде (рис. 3.23). Это позволило снизить частоту преобразования каждого из АЦП в число раз, равное числу столбцов в фоточувствительной секции, а также в соответствующее число раз уменьшить полосу частот видеосигнала.

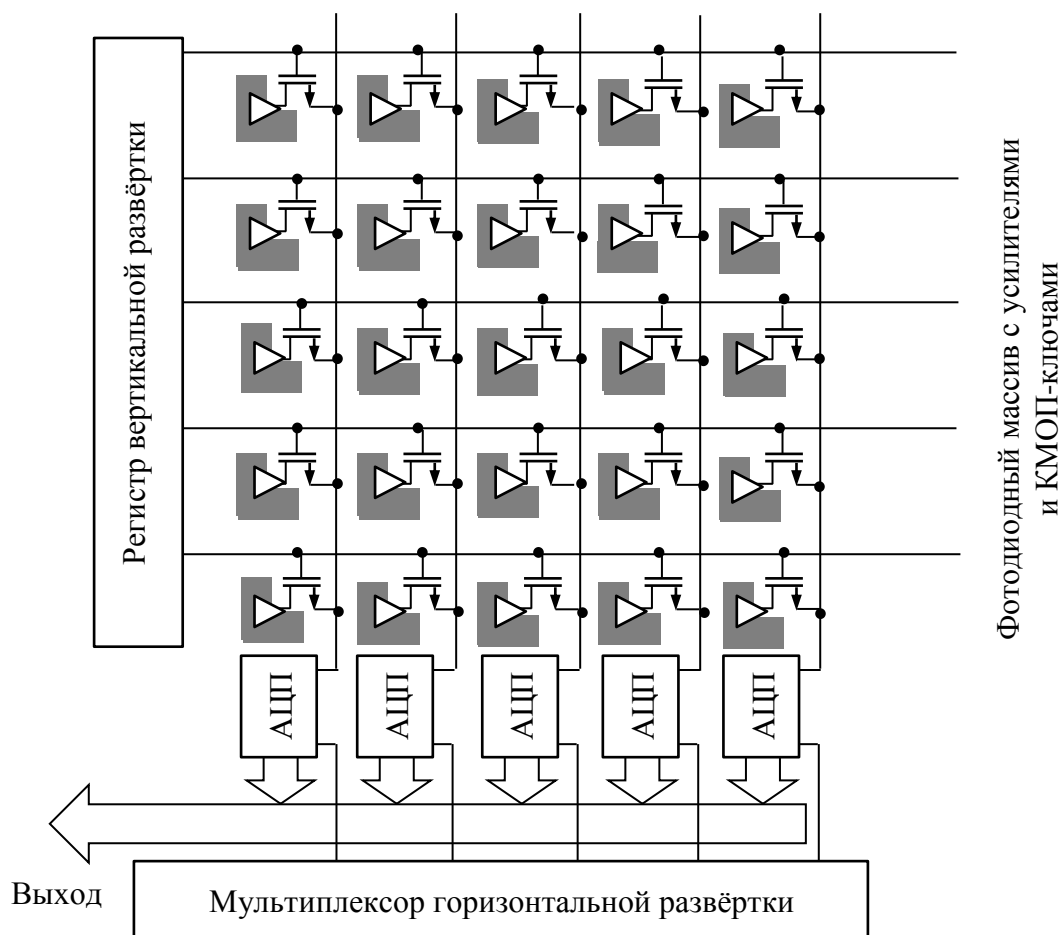


Рис. 3.23. КМОП-фотоприёмная матрица с активным пикселем, и столбцовыми АЦП

Дальнейшее развитие пошло по пути движения к цифровому пикселу (*Digital Pixel Sensor, DPS*) [29], [138], [151], [152] (рис. 3.24). Особенностью этих преобразователей является интеграция в каждый пиксел не только активного усилителя, но и непосредственно АЦП. При этом частота преобразования АЦП становится равной кадровой частоте фотоприёмника, а полоса частот видеосигнала уменьшается до минимальных значений.

Известны КМОП-сенсоры с цифровым пикселем размером $9,8 \times 9,8$ мкм при технологии 0,18 мкм, содержащие в каждом пикселе АЦП, состоящий из 37 транзисторов [29]. Остановимся на реализации наиболее распространенных структур пикселей современных КМОП-сенсоров. Как и в матрич-

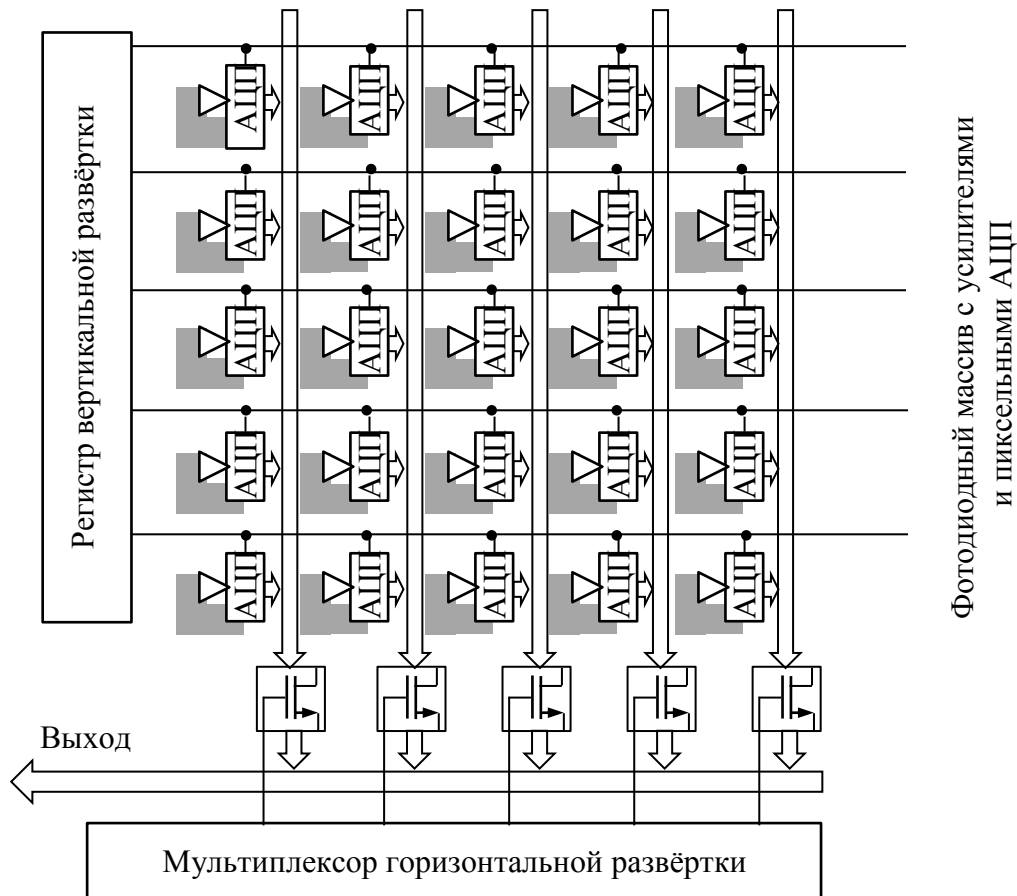


Рис. 3.24. КМОП-фотоприёмная матрица с цифровым пикселем

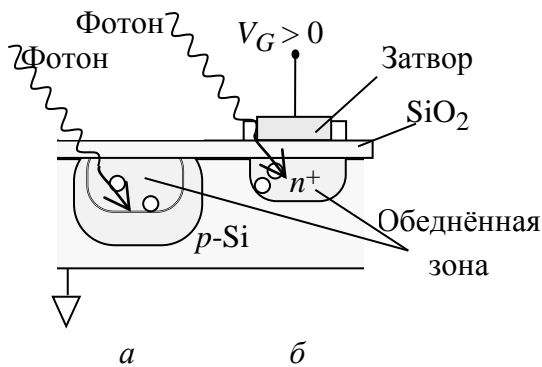


Рис. 3.25. Структура элемента твёрдотельного фотоприёмника

ных ПЗС, светочувствительным пикселем в КМОП-сенсоре может являться обеднённая область МОП-ёмкости, возникающая при подаче на фазный электрод обедняющего напряжения (рис. 3.25, б), либо же обеднённая область обратносмещённого фотодиода (рис. 3.25, а). Светочувствительный элемент второго типа предпочтительнее хотя бы с той точ-

ки зрения, что фотодиодная структура имеет существенно более высокий коэффициент сбора светового потока из-за отсутствия слоёв поликремния, поглощающего световой поток.

Простейшей является структура пассивного пикселя (рис. 3.26). Помимо фотодиода она содержит ключевой транзистор VT , который по сигналу RS (*Row Select*, выбор столбца) подключается к вертикальной видеошине. Таким

образом, к этому пикселу подходят две шины – видеошина и шина, по которой подается сигнал считывания RS. Достоинством такого типа пиксела является высокий коэффициент использования площади, так как площадь транзистора много меньше площади фотодиода, а недостатком – высокий уровень шумов, связанный с тем, что фотодиод подключается непосредственно к вертикальной видеошине, имеющей бóльшую ёмкость [110].

Структура активного пиксела содержит уже как минимум три полевых транзистора и четыре шины – вертикальную видеошину, шины, по которым подаются сигналы RS и RST, а также шину питания V_{DD} (рис. 3.27).

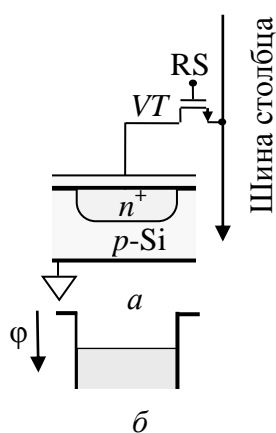


Рис. 3.26. Структура пассивного пиксела (а); соответствующая диаграмма потенциала (б)

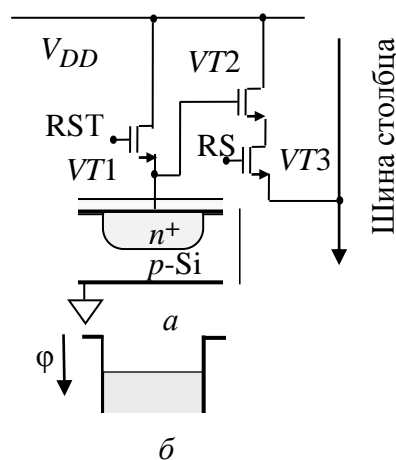


Рис. 3.27. Структура активного пиксела (а); соответствующая диаграмма потенциала (б)

Транзистор $VT3$, управляемый сигналом RS, выполняет ту же функцию, что и в пассивном пикселе. Транзистор $VT1$, управляемый сигналом RST (*Reset* – сброс), служит для сброса накопленного заряда, а гальванометрический транзистор $VT2$ выполняет роль истокового повторителя, осуществляющего усиление видеосигнала по току. Нагрузкой данного транзистора является источник тока, расположенный на конце столбца за пределами фото-чувствительной секции. Коэффициент использования площади у такого пиксела значительно ниже за счёт большего числа транзисторов, однако шумовые характеристики значительно лучше, чем в пассивном пикселе.

Структура активного пиксела (рис. 3.27, а) позволяет организовать *неразрушающее* считывание видеосигнала (рис. 3.28). Для этого сигнал сброса накопленного потенциала фотодиода RST подаётся не в каждый кадр длительностью T_K , а один раз в несколько кадров, что обеспечивает время накопления больше длительность одного кадра (на рис. 3.28 – время накоп-

ления равно $3T_K$). Так как считывание видеосигнала осуществляется в моменты выбора строки RS, то за время накопления будет считано три последовательных значения видеосигнала. Разности между этими значениями ΔU будут соответствовать уровню видеосигнала в соответствующем кадре.

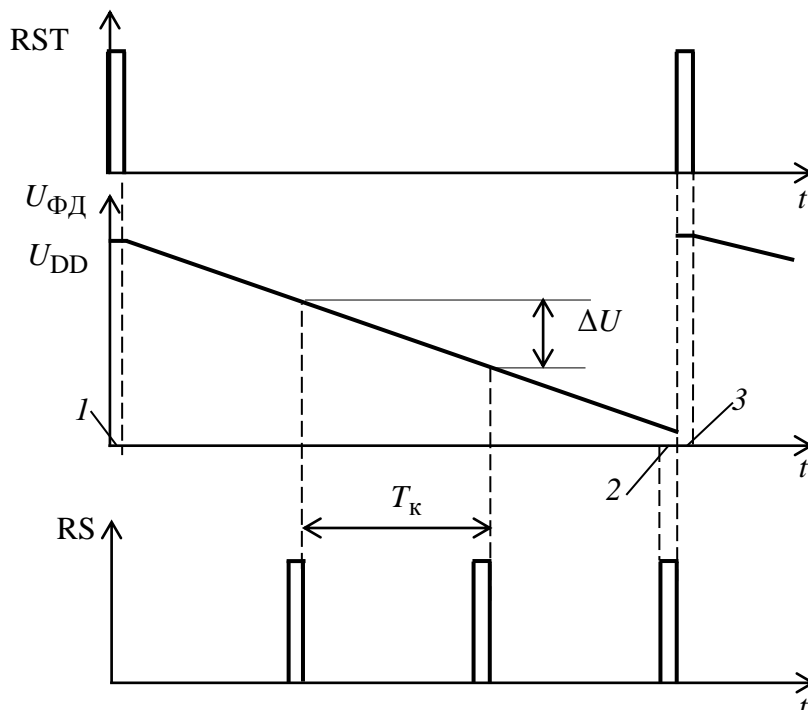


Рис. 3.28. Временные диаграммы управления КМОП-матрицей при неразрушающем считывании

Неразрушающее считывание позволяет реализовать высокую чувствительность фотоприёмника и, в то же время, – высокую кадровую частоту для передачи изображений динамических объектов. Неразрушающее считывание можно трактовать как расширение динамического диапазона фотоприёмника, поскольку максимальная освещённость будет определяться

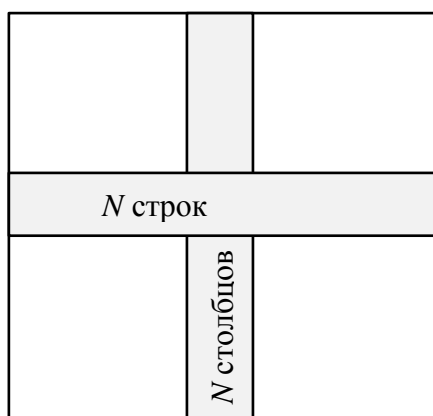


Рис. 3.29. Зоны скоростного неразрушающего считывания для оценки общего смещения изображения

малым временем накопления, равным длительности одного кадра T_K , а минимальная – временем нескольких кадров.

Данный метод накопления совместно с режимом считывания заданного окна может быть использован для реализации на КМОП-сенсоре датчика смещения для системы стабилизации изображения (рис. 3.29; см. 5.3). Центральные строки и

столбцы могут считываться с частотой, в несколько раз превышающей частоту кадров, с которой формируется основное изображение. Таким образом, информация о смещении оптического изображения относительно фотоприёмника по ортогональным осям может формироваться несколько раз за кадр, что позволит точнее реализовать оптическую стабилизацию изображения. Так, при считывании с более высокой частотой падает чувствительность фотоприёмника. Для компенсации этого явления может быть использовано усреднение видеосигнала по N центральным строкам и столбцам.

Одним из существенных недостатков КМОП-сенсоров с активным пикселом является режим бегущего считывания, или бегущего электронного затвора (*rolling shutter*). Это связано с тем, что при непосредственном последовательном выборе строк времени накопления в них оказываются равными, но смещёнными друг относительно друга на время длительности строки. Такой режим накопления приводит к геометрическим искажениям движущихся объектов [139]. Так, объекты, движущиеся по вертикали, растягиваются или сжимаются, а по горизонтали – наклоняются. Для борьбы с этим явлением используется режим накопления, называемый глобальным считыванием, или глобальным электронным затвором (*global shutter*) [116]. Так как для реализации глобального электронного затвора требуется введение в структуру пиксела дополнительных элементов памяти, этот режим характеризуется снижением площади фотодиода и чувствительности матрицы. Это стимулировало разработку большого количества алгоритмов обработки видеосигналов для коррекции геометрических искажений подвижных объектов при бегущем считывании [140]–[142].

На рис. 3.30 представлена структура активного пиксела с глобальным считыванием, для реализации которого в структуру каждого активного пиксела вводится запоминающий аналоговый видеосигнал элемент – ёмкость C . Транзистор $VT1$ является транзистором сброса, обеспечивающим заряд ёмкости фотодиода до напряжения питания V_{DD} . Видеосигнал с фотодиода снимается истоковым повторителем на транзисторе $VT2$, нагруженном на источник тока на транзисторе $VT3$. Далее следует схема выборки-хранения, реализованная на ключевом транзисторе $VT4$ и ёмкости C . Транзистор $VT4$ открывается импульсом выборки S/H и запоминается на ёмкости C . На транзисторе $VT5$ собран ещё один истоковый повторитель, работающий непосредственно на шину столбца, к которой он подключается при считывании ключевым транзистором $VT6$, открываемым импульсом выбора строки RS .

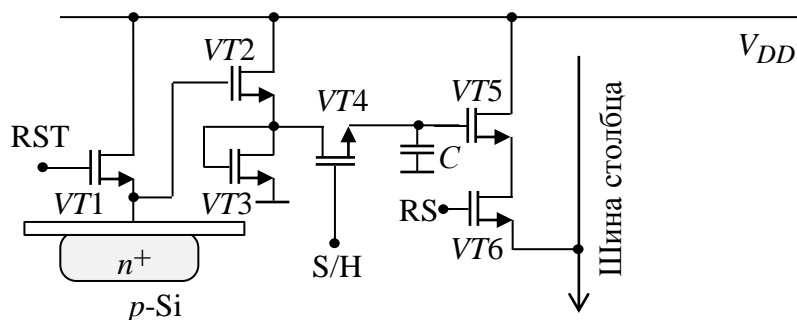


Рис. 3.30. Структура активного пиксела с глобальным считыванием

Запоминание значений видеосигнала со всех фотодиодов происходит одновременно, т. е. импульс выборки S/H является глобальным. В силу этого геометрические искажения динамических объектов, характерные для режима бегущего считывания, в данном случае отсутствуют. Значение ёмкости запоминания C выбирается из следующих компромиссных условий: для запоминания в течение короткого времени выборки (длительность импульса S/H) оно должно быть достаточно малым, но и достаточно большим для хранения значения видеосигнала за время кадра.

Структуры пикселей, сходные со структурами рис. 3.30 и 3.31, использующими накопление заряда, а не потенциала, также позволяют организовать режим глобального считывания. Для этого перенос заряда из фотодиода в плавающую диффузионную область должен происходить для всех пикселей КМОП-сенсора одновременно.

Дальнейшим развитием активного пиксела было введение в него плавающей диффузионной области (рис. 3.31 и 3.32). В данном виде пиксела накопленный заряд переносится из фотодиода в плавающую диффузионную область подачей положительного смещения на затвор переноса TX.

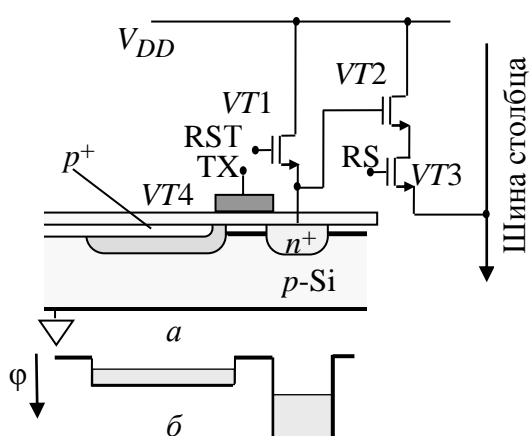


Рис. 3.31. Структура активного пиксела с дополнительным затвором (а), соответствующая диаграмма потенциала (б)

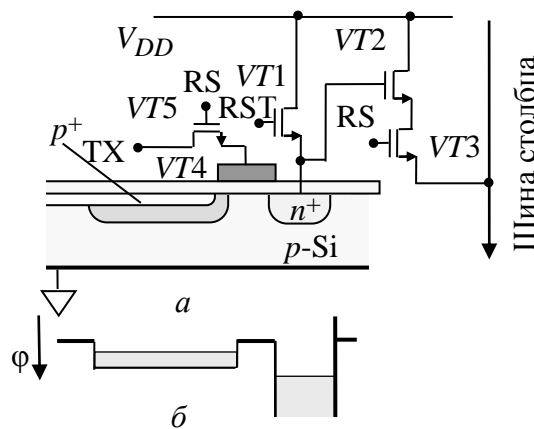


Рис. 3.32. Структура активного пиксела с дополнительным затвором и усилителем в цепи затвора (а), соответствующая диаграмма потенциала (б)

Возможность реализации плавающей диффузионной области с очень малой ёмкостью позволяет увеличить коэффициент преобразования заряда в напряжение и разделить процессы накопления и считывания по пространству. С учётом затвора переноса такой пиксел содержит четыре транзистора и пять шин. По сравнению с предыдущим пикселем в данном случае возникает необходимость в шине управления затвором переноса.

Такой пиксел имеет малый коэффициент использования полезной площади, однако обеспечивает малый уровень флуктуационных и геометрических шумов. С целью уменьшения числа шин управления и увеличения коэффициента использования полезной площади сигналы RS и TX можно объединить (рис. 3.32). Для структур пиксела с плавающей диффузионной областью формируемый видеосигнал имеет трёхуровневую форму, характерную для фотоприёмников на основе ПЗС. Наличие попиксельного сброса приводит к появлению шума установки узла детектирования заряда, так называемого КТС-шума. Как следствие, для обработки такого видеосигнала в структуру КМОП-сенсора необходимо интегрировать схему двойной коррелированной выборки, т. е. узел, характерный для всех линейных и матричных ПЗС.

3.5. Чувствительность КМОП-фотоприёмников

Определим чувствительность КМОП-сенсора с активным пикселем с накоплением потенциала (см. рис. 3.27). Будем считать, что фотогенерированные носители заряда образуются непосредственно в фотодиоде и диффузия носителей, образовавшихся за его пределами, в фотодиод отсутствует. Пусть под действием импульса сброса RS ёмкость фотодиода C_{PD} зарядилась до потенциала питания U_{DD} (см. рис. 3.28). Заряд на ёмкости фотодиода составит $Q_{PD} = U_{DD}C_{PD}$. Под действием освещённости E , выраженной в люксах, за время накопления T_H в фотодиоде образуется заряд [143]:

$$Q_S = E \frac{KA_P T_H}{k} \frac{\int_{\lambda_{\min}}^{\lambda_{\max}} S(\lambda) \varphi'(\lambda) d\lambda}{\int_{\lambda_{\min}}^{\lambda_{\max}} v(\lambda) \varphi'(\lambda) d\lambda},$$

где K – коэффициент усиления аналогового усилителя; A_P – площадь пиксела; $k = 683$ лм/Вт – фотометрический коэффициент; λ_{\max} и λ_{\min} – макси-

мальная и минимальная длина волны чувствительности сенсора соответственно; $v(\lambda)$ – кривая видности зрительного анализатора человека; $\phi'(\lambda)$ – нормированная спектральная плотность мощности источника света; $S(\lambda) = = \lambda q \theta(\lambda) / hc$ – спектральная чувствительность сенсора, А/Вт (λ – длина волны света; q – заряд электрона; $\theta(\lambda)$ – спектральная зависимость квантовой эффективности ФП; h – постоянная Планка; c – скорость света).

Полезный видеосигнал U_S будет равен $U_S = |Q_S / C_{PD}|$ (знак модуля применён из-за того, что видеосигнал под действием света снижается, см. рис. 3.28). Теперь учтём, что $C_{PD} = C_{уд} A_P$, где $C_{уд}$ – удельная ёмкость фотодиода, обычно составляющая единицы фемтофарада на квадратный микрометр. Тогда выражение для полезного видеосигнала примет вид

$$U_S = E \frac{KT_H}{k C_{уд}} \frac{\int_{\lambda_{\min}}^{\lambda_{\max}} S(\lambda) \phi'(\lambda) d\lambda}{\int_{\lambda_{\min}}^{\lambda_{\max}} v(\lambda) \phi'(\lambda) d\lambda},$$

т. е. видеосигнал КМОП-фотоприёмника с активным пикселом не зависит от площади фоточувствительного элемента – фотодиода. Полученное выражение даёт возможность фирмам-производителям уменьшать размер фотодиода вплоть до 1 мкм практически без потери чувствительности.

Для нахождения пороговой чувствительности, т. е. значения освещённости, реализующей заданное отношение сигнал/шум, необходим анализ составляющих шумов КМОП-фотоприёмника с активным пикселом.

Рассмотрим составляющие шумов КМОП-фотоприёмника, образующиеся в процессе накопления сигнала. Данные составляющие аналогичны компонентам, образующимся при накоплении в фотодиодах матричных ПЗС со строчным переносом [144]. В литературе принято составляющие шумов $n_{ш}$ характеризовать числом шумовых электронов e , однако в КМОП-фотоприёмниках с накоплением потенциала логичным будет использование среднеквадратического отклонения потенциала $U_{ш} = q n_{ш} / C_{PD}$, В.

В процессе накопления полезного сигнала неминуемо образуется фотонный шум [3] $n_{ш.ф} = \sqrt{n_S}$, где n_S – число накопленных фотоэлектронов. Соответственно, $U_{ш.ф} = q \sqrt{n_S} / C_{PD}$. Данный компонент шума определя-

ется непосредственно эмиссией фотонов из источника света и является глобальным ограничением чувствительности фотоприёмника.

Помимо накопления полезных, фотогенерированных носителей в фотодиоде образуются носители, генерированные темновым током полупроводника. Количество темновых электронов находится по формуле $n_{Т.Т} = j_{Т.Т} A_P T_H / q$; формируемый ими сигнал темнового тока: $U_{Т.Т} = j_{Т.Т} A_P T_H / C_{PD}$; соответствующие им шумы равны: $n_{ш.Т.Т} = \sqrt{n_{Т.Т}} = \sqrt{j_{Т.Т} A_P T_H / q}$, $U_{ш.Т.Т} = \sqrt{j_{Т.Т} A_P T_H q / C_{PD}}$. С учётом выражения для ёмкости фотодиода $U_{ш.Т.Т} = \sqrt{j_{Т.Т} T_H q / C_{уд}^2 A_P}$, т. е. шумы темнового тока, выраженные в вольтах, с увеличением размера пиксела уменьшаются.

Сброс потенциала фотодиода после каждого цикла накопления приводит к появлению шума сброса, или так называемого КТС-шума [118], [144]. Среднеквадратическое отклонение шума сброса может быть найдено как $n_{сбр} = \sqrt{kTC_{PD}} / q$, где T – абсолютная температура, $U_{ш.сбр} = \sqrt{kT / C_{PD}}$. В фотоприёмнике на основе ПЗС шум сброса эффективно подавляется схемой ДКВ, однако в КМОП-фотоприёмнике технологически сложно реализовать ДКВ и значительно чаще используется схема двойной выборки, когда отсчёты берутся не в моменты времени 1 и 2 (см. рис. 3.28), а в моменты времени 2 и 3. Это позволяет подавить разброс напряжения смещения транзистора $VT2$ в различных пикселах, т. е. уменьшить геометрический шум, однако приводит к увеличению дисперсии шума сброса в 2 раза и, соответственно, среднеквадратического значения шума в $\sqrt{2}$ раз.

При считывании через включённый по схеме истокового повторителя транзистор $VT2$ вносит две составляющие шумов – фликкер-шум со спектральной плотностью вида $1/f$ и тепловой шум канала. Фликкер-шум эффективно подавляется схемой ДКВ. Тепловые шумы канала транзистора подробно рассмотрены в [118], [144].

При открывании ключевого транзистора $VT3$ также образуется шум сброса, так как при этом происходит перезаряд ёмкости видеошины C_{VB} . Значение шума сброса в числе электронов составляет $n_{ш.сбр} = \sqrt{kTC_{VB}} / q$, в вольтах – $U_{ш.сбр} = \sqrt{kT / C_{VB}}$.

На основе полученных выражений для полезного видеосигнала и составляющих шумов может быть получено значение пороговой чувствительности КМОП-фотоприёмника.

Исследование составляющих шумов КМОП-фотоприёмника показало, что их зависимость от размера пиксела может быть как нарастающей, так и спадающей функцией и в каждом конкретном случае (спектральный диапазон, наличие ДКВ, нормы точности) должна определяться отдельно.

Проведённый анализ [145] позволяет сделать **вывод** о том, что **оптимальной стратегией построения активного пиксела** является использование фотодиода минимальных размеров (ёмкости) с микролинзой, максимально собирающей световой поток. Данная стратегия открывает путь к созданию КМОП-фотоприёмника с цифровым пикселом, а также с обработкой видеосигнала непосредственно в пикселе, так как в нём образуется достаточно места для размещения АЦП, схемы ДКВ и других устройств обработки видеосигнала.

3.6. Борьба за динамический диапазон

В технике детектирования частиц традиционно использовались приборы на лавинных эффектах, позволяющие зафиксировать единичные события. Современный уровень развития технологии позволяет создавать фотоприёмники на лавинном эффекте, обеспечивающие огромный динамический диапазон сигнала, но вследствие высокой чувствительности применимые только для сигналов низкой интенсивности. Для условий малой освещённости такой фотоприёмник, использующий высоковольтную структуру фоточувствительной области и эффект лавинного фотоэлектронного умножения, работает фактически в счётном режиме, его динамический диапазон определяется скоростью работы регистров счётчика. На

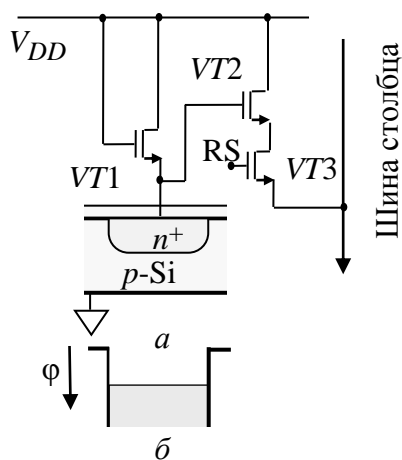


Рис. 3.33. Структура активного пиксела с логарифмической шкалой (а), соответствующая диаграмма потенциала (б)

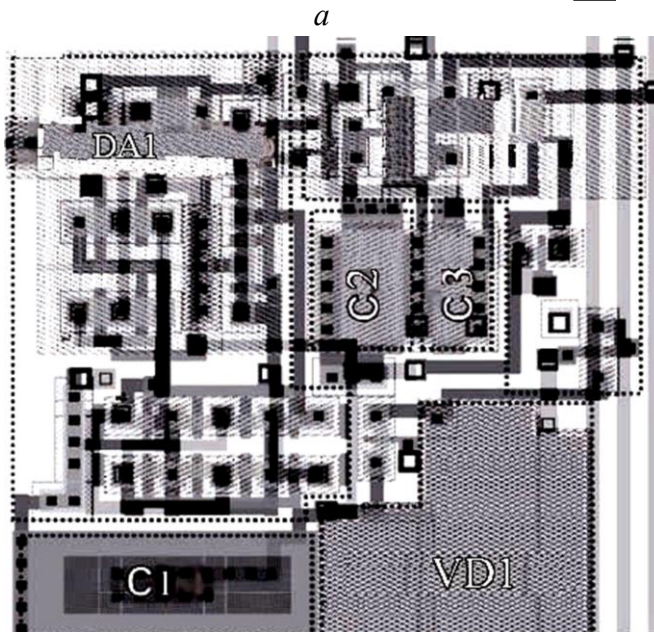
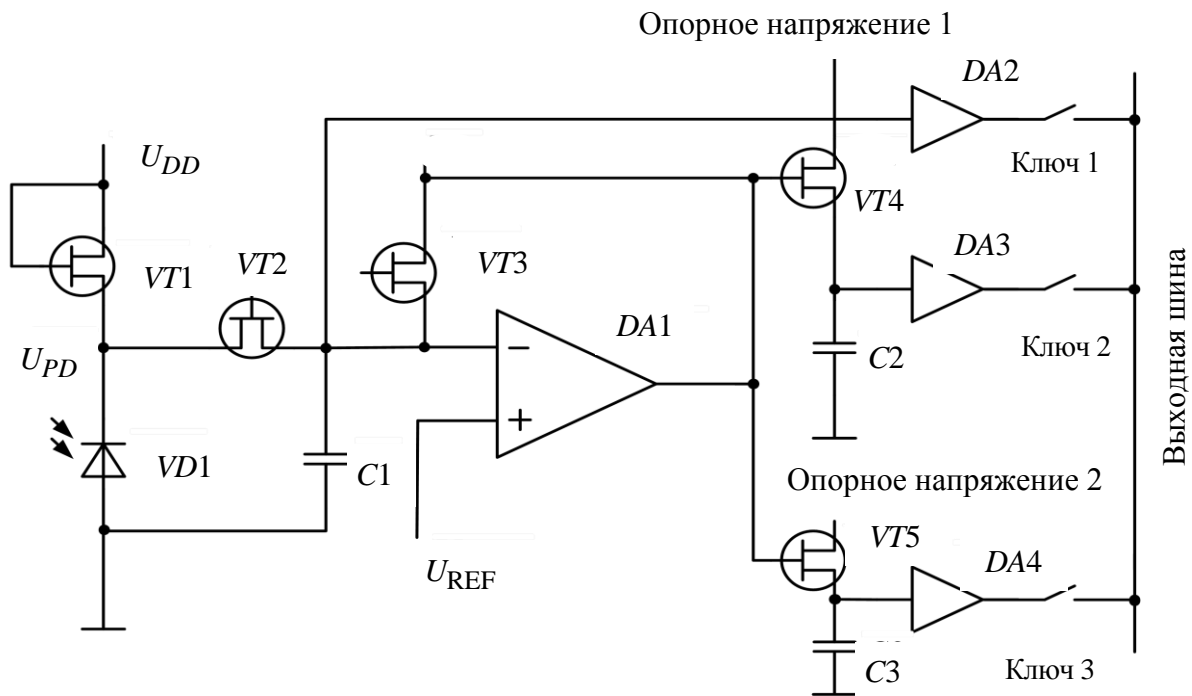
практике подобные системы позволяют получить до 20 бит динамического диапазона при использовании некоторых дополнительных алгоритмов пересчёта потока на единичные события.

Особый интерес представляют пиксели с логарифмической характеристикой накопления. При большом динамическом диапазоне аналогового видеосигнала разработчик может использовать фоточувствительный транзистор во включении, показанном на рис. 3.33,

при котором соотношение между фототоком и напряжением на затворе носит логарифмический характер. Указанное свойство позволяет получить фотоприёмник с расширенным динамическим диапазоном по свету и характеристикой свет–сигнал, аналогичной таковой у человеческого глаза, которая согласно закону Вебера–Фехнера имеет логарифмический характер и учитывается в телевидении с самого его зарождения. Важность принципа накопления в телевидении неоспорима, именно с его помощью осуществляется групповой счёт фотонов (см. 2.4). КМОП-телекамеры в отличие от фотоприёмников предыдущего поколения на ПЗС могут обеспечить сверхширокий линейный динамический диапазон.

Технической основой увеличения динамического диапазона является усложнение структуры пиксела введением элементов аналоговой или цифровой памяти. На рис. 3.34, а приведена принципиальная схема пиксела, обеспечивающего линейный динамический диапазон по входной освещенности 132 дБ – *более миллиона раз* [146], [118], а на рис. 3.34, б – его топология. Пиксел состоит из фотодиода $VD1$, первичной накопительной ёмкости $C1$, компараторов $DA1$, дополнительных емкостей хранения $C2$ и $C3$ и логики считывания.

Прикладной задачей по достижению максимального динамического диапазона является дистанционное зондирование земли (ДЗЗ). По мере нарастания требований к разрешению по Земле до 1...2 м и менее время накопления в классических строчных (линейных) фотоприёмниках (определяемое временем строки) становится всё меньше и возникает потребность дополнительного увеличения времени накопления во внешней памяти. В настоящее время для систем ДЗЗ видимого диапазона применяются исключительно фотоприёмники, традиционно реализуемые на ПЗС-технологии. Эта технология позволяет расположить аналоговую память вплотную с основной строкой и реализовать режим временной задержки и накопления (ВЗН). В ПЗС с режимом ВЗН отношение сигнал/шум возрастает пропорционально корню квадратному из числа складываемых строк. В ПЗС возможность суммирования зарядов ограничивается динамическим диапазоном аналоговых схем и обычно не превышает 128, что соответствует увеличению отношения сигнал/шум на порядок. Вместе с тем именно режим ВЗН считается ключевой технологией, обеспечивающей стратегическое преимущество ПЗС перед КМОП-фотоприёмниками, делая ПЗС ВЗН незаменимыми [147].



б

Рис. 3.34. Принципиальная схема аналогового пиксела с расширенным динамическим диапазоном (с тремя линейными участками характеристики свет–сигнал) (а); его топология (б)

КМОП-технология, однако, открывает возможность осуществления режима ВЗН не в аналоговой форме, как в ПЗС, а в цифровой. При этом, так как скрытая ранее идея накопления информации, а не заряда, становится явной, очевидно, что цифровая реализация режима ВЗН ведёт к нарастанию разрядности выходного сигнала. Например, при 10-разрядном квантовании видеосигнала и 128 строках накопления выходной сигнал может иметь разрядность до 17.

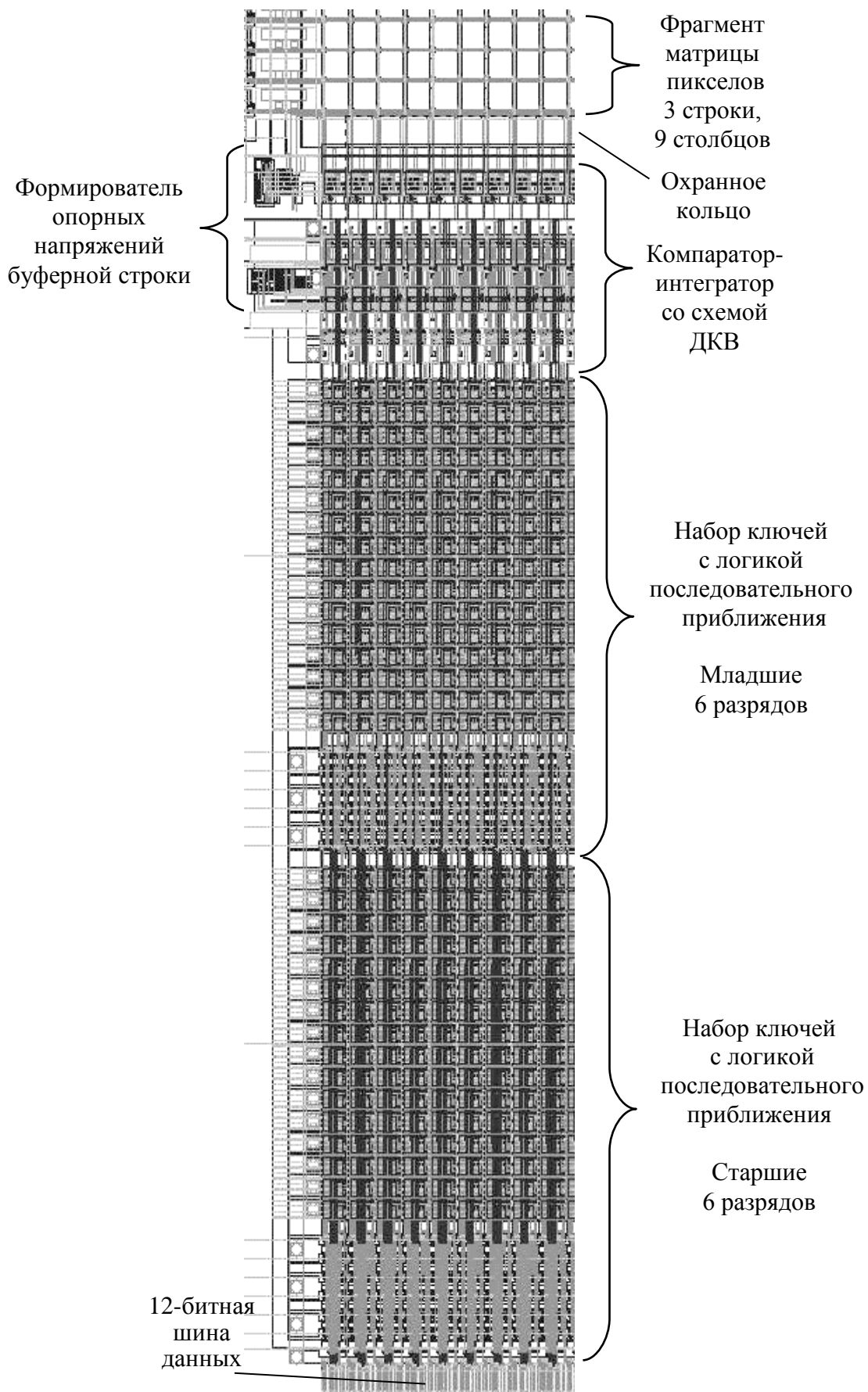


Рис. 3.35. Фрагмент топологии фотоприёмной КМОП-матрицы со столбцовыми АЦП

Реализация режима ВЗН в КМОП-фотоприёмнике наглядно демонстрирует новую концепцию видеосистем на кристалле, где в рамках одной СБИС кроме фотоприёмного массива расположены не только коммутирующие схемы (управление развёрткой), но и устройства обработки видеоинформации. В видеосистемах на кристалле в дополнение к *параллельному накоплению* информации в фотоприёмном массиве открывается возможность *параллельного считывания и параллельной обработки* видеоинформации [135]–[137] (см., например, рис. 3.17).

КМОП-технология, позволяющая реализовать фотоприёмные матрицы со столбцовыми АЦП (рис. 3.35; [115], [116], [149]), обеспечивает первый шаг в создании крупноформатных видеосистем на кристалле. При этом открываются возможности структурирования памяти (недоступна в ПЗС) и введения устройств для борьбы с ложной информацией, порождаемой космической радиацией, алгоритмическими методами [134]–[136]. Для реализации режима ВЗН в КМОП-фотоприёмниках необходимо решить две задачи: во-первых, перейти от суммирования заряда к цифровому сдвиговому регистру; во-вторых, решить проблему аппаратных шумов, которые, включая шум квантования, должны быть ниже дробового шума заряда. В настоящее время разработки в области КМОП ВЗН ведутся рядом компаний: «Surgus Semiconductor», «CMOSIS» и др. [149], [150].

3.7. Борьба за потенциальную контрастную чувствительность

Достижение большого динамического диапазона методом изменения крутизны характеристики свет–сигнал не позволяет достичь предельной контрастной чувствительности из-за уменьшения крутизны преобразования при больших освещённостях. Поэтому для достижения предельной контрастной чувствительности требуются иные подходы. К аппаратуре видеосистем потребители предъявляют множество требований как по качеству получаемой информации, так и по общетехническим параметрам. Размеры используемых оптических систем зачастую определяются не только разрешающей способностью, но и компромиссом между низкой освещённостью исследуемого объекта и уровнем реальной чувствительности фотоприёмника. Кроме того, передача информации до потребителя упирается в необходимость проведения сжатия видеопотока без потери качества информации. Для этого требуются существенные вычислительные мощности, реализующие эффективные алгоритмы кодирования для компрессии потока информации в соответствии с пропускной способностью каналов передачи данных. С энергетической и

размерной точек зрения разработчик видеосистем должен использовать возможности современных полупроводниковых технологий для максимального объединения всех требуемых функциональных блоков в составе единого кристалла микросхемы. Это требование зависит не только от удобства применения полученной видеосистемы в конечной аппаратуре пользователя. Необходимость интеграции различных функциональных блоков видеосистем на одном кристалле микросхемы вызвана целым рядом факторов, определяющих в итоге контрастную чувствительность видеосистемы. Для передачи сигналов между микросхемами требуется обеспечение помехозащищенности и быстродействия. При этом на длинных проводниках на печатных платах возникают наводки от посторонних сигналов и окружающей аппаратуры, а энергия тратится на перезарядку паразитных емкостей выводов корпуса микросхемы и межсоединений. Таким образом, уровень токов при передаче аналоговых сигналов внутри и вне системы на кристалле зачастую отличается на два-три порядка, возрастая с десятков микроампер до единиц миллиампер для каждого канала передачи информации. Это не только сказывается на общей энергоэффективности системы, неравномерности нагрева, но и увеличивает уровень шумов, возникающих за счёт эмиссии носителей в подложку кристалла микросхемы со стороны формирователей выходных сигналов. Другим принципиальным подходом по повышению контрастной чувствительности видеосистемы является использование распределенных блоков аналого-цифровой обработки видеосигнала.

В традиционных видеосистемах предыдущего поколения вследствие технологических ограничений использовался единый блок АЦП, который последовательно обрабатывал коммутируемый видеосигнал, поступающий от блоков строчных усилителей фотоприёмной области. При этом уровень шумов определялся электромагнитными наводками в линиях связи из-за влияния помех окружающих компонентов микросхемы. По мере развития технологии изготовления и уменьшения проектных норм у разработчика появилась возможность создания распределенных систем аналого-цифрового преобразования сигнала. На первом этапе это было реализовано для каждой строки фотоприёмника, что позволило значительно увеличить быстродействие системы, до этого момента ограниченное максимальной тактовой частотой единого АЦП. При этом можно отметить существенное снижение доли наводок в первую очередь цифровых блоков в уровне шумов видеосигнала. Рассматривая физические эффекты в пикселе, мы можем пойти по пути известных допущений. Накопление отражает передачу информации о простран-

ственно-временных координатах не каждого воспринятого фотона (или возникшего из-за него электрона), а их группы. С помощью накопления реализуется групповой счёт фотонов, рассмотренный в гл. 2. Стимулом к такому группированию является, во-первых, чрезмерно большая пропускная способность, которая требуется для передачи координат каждого фотона (при освещённости 1 лк на 1 см² в 1 с случайным образом падает примерно 10¹² фотонов); во вторых, подавляющее большинство из них несёт лишь шумовую информацию. Таким образом, возможность группового счёта может быть реализована в составе пиксела со встроенной системой аналого-цифрового преобразования напряжения и счётчиком [151]–[153].

Для достижения потенциальной контрастной чувствительности представляется естественным отказаться от сокращения времени накопления при интенсивных световых воздействиях и от логарифмической характеристики свет–сигнал. Первым шагом в этом направлении стал «цифровой» пиксел в матрицах КМОП, в состав которого был включен АЦП с числом разрядов, достигающим 10–12. Обычный цифровой пиксел предполагает разделение во времени накопления фотогенерированных электронов и аналого-цифрового преобразования. Этот способ не обеспечивает потенциальной контрастной чувствительности из-за трудностей реализации быстродействующих АЦП большой разрядности. Поэтому необходимо процесс группового счёта фотонов в пикселе совместить во времени с процессом накопления, т. е. ввести в состав пиксела КМОП-фотоприёмной матрицы компаратор и счетчик превышений накопленным сигналом заданного порога [151]–[153] (рис. 3.36). Диаграммы работы такого пиксела приведены на рис. 3.37, топология – на рис. 3.38.

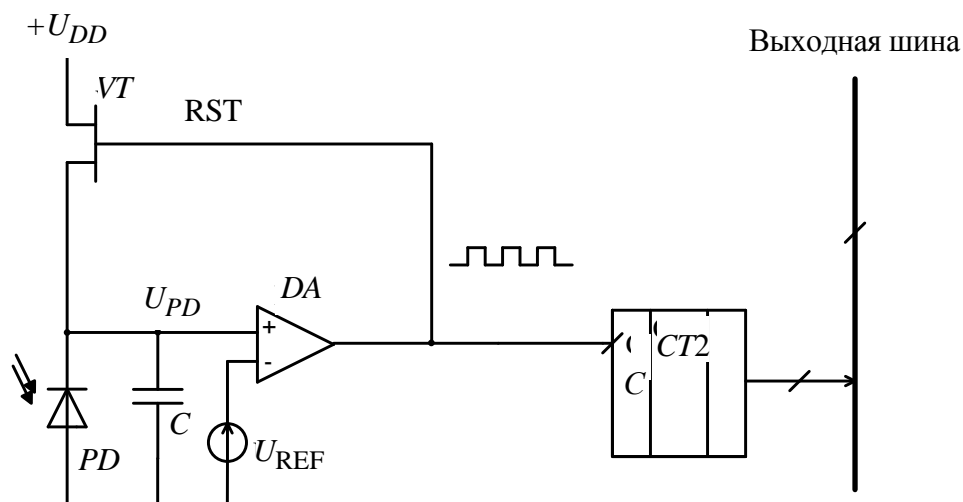


Рис. 3.36. Цифровой пиксел с преобразованием сигнал–частота

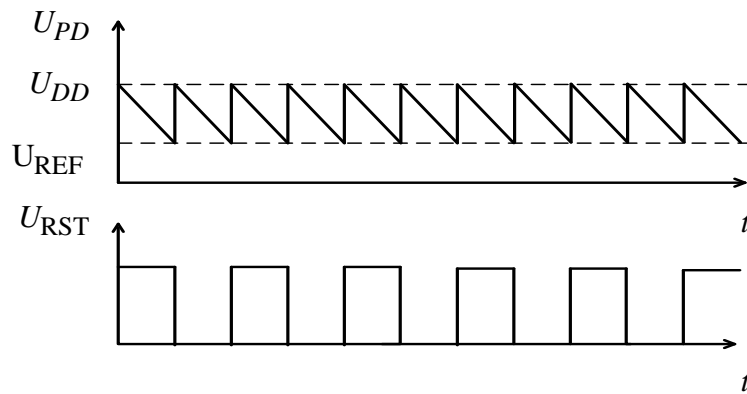


Рис. 3.37. Диаграммы работы цифрового пиксела с преобразованием сигнал–частота

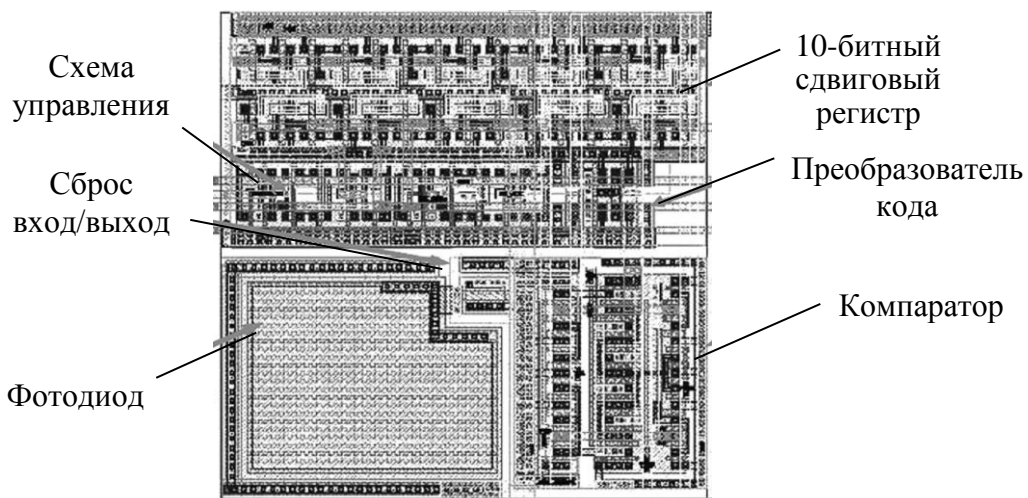


Рис. 3.38. Топология пиксела (технология с нормой точности 0,18 мкм) с 10-разрядным счетчиком групп фотогенерированных электронов

В такой цифровой пиксел с объединением процессов накопления и счёта групп фотогенерированных электронов целесообразно ввести усилитель с управляемым коэффициентом усиления, позволяющим не только регулировать размер группы фотогенерированных электронов, но и с высокой точностью осуществлять выравнивание чувствительности пикселей по полю (см. рис. 3.33) [151]–[153] с целью подавления так называемого геометрического шума (*Fixed Pattern Noise, FPN*), выражающегося в различных коэффициентах преобразования светового потока в различных пикселах.

Работа фотоприёмной КМОП-матрицы такого типа (рис. 3.39) иллюстрируется рис. 3.40, на котором показано, что при достижении видеосигналом напряжения U_{REF} вырабатывается импульс сброса RST, который также изменяет значение цифрового кода, записанного в двоичный счётчик *CT2*.

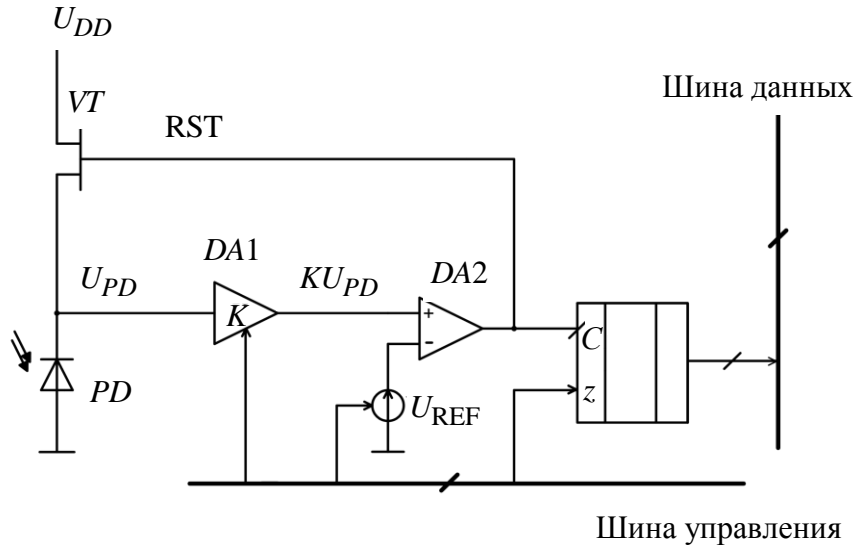


Рис. 3.39. Цифровой пиксел с преобразованием сигнал–частота и регулируемым размером групп фотогенерированных электронов (регулируемым коэффициентом усиления)

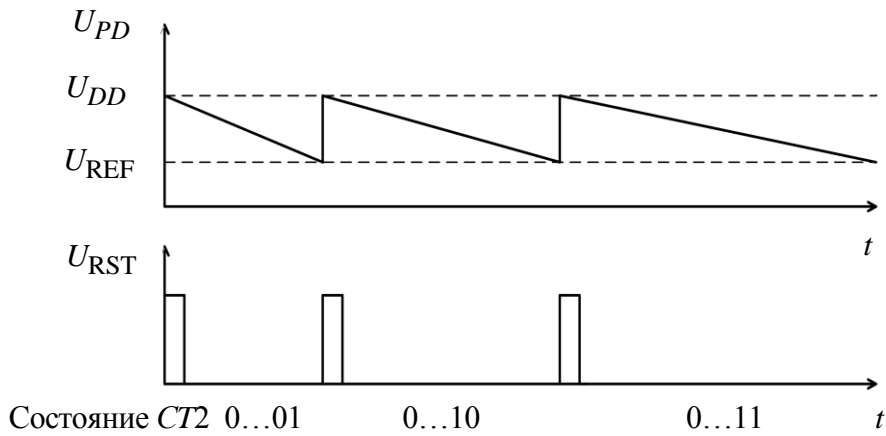


Рис. 3.40. Диаграммы работы цифрового пиксела с преобразованием сигнал–частота с регулируемым коэффициентом усиления

Оптимальный размер группы фотоэлектронов n_k и соответствующая разрядность внутрипиксельного счётчика могут определяться исходя из двух различных подходов: с учётом и без учёта оставшегося в конце интервала накопления сигнала, меньшего уровня срабатывания компаратора в схеме рис. 3.36 и 3.39. В обоих случаях следует исходить из того, что за время накопления в пикселе генерируется N электронов, причём максимальное значение N много больше ёмкости пиксела n_e . При использовании обычного активного пиксела пороговый контраст определяется известной формулой [27]:

$$K_{\text{пор}} = \frac{\Psi_{\text{пор}}}{\sqrt{n_e}}. \quad (3.4)$$

Так как к каждой группе n_k накапливаемых электронов добавляется шум считывания со среднеквадратическим значением $n_{\text{сч}}$ электронов, результирующая дисперсия D шума в выходном сигнале будет обусловлена дробовым шумом с дисперсией n_k , равной числу электронов в группе, шумом считывания и числом групп сигнальных электронов N/n_k :

$$D = \left(n_k + n_{\text{сч}}^2 \right) \frac{N}{n_k} = N \left(1 + \frac{n_{\text{сч}}^2}{n_k} \right).$$

Это определяет зависимость контрастной чувствительности от размера n_k группы накапливаемых электронов:

$$K_{\text{пор}} = \frac{\Psi_{\text{пор}}}{\sqrt{N}} \sqrt{1 + \frac{n_{\text{сч}}^2}{n_k}}. \quad (3.5)$$

Отсюда следует, что при передаче кода числа срабатываний счётчика (порядок числа) и оставшегося сигнала (мантиссы числа, которая может квантоваться отдельным АЦП) [1] чем больше размер n_k группы накапливаемых электронов, тем ближе контрастная чувствительность к предельному значению, определяемому дробовым шумом электронов. Естественно, что при таком подходе размеры группы n_k приближаются к ёмкости пиксела n_e . Так как ёмкость пиксела n_e имеет порядок $2^{15} \dots 2^{16}$, а, например, при наблюдении из космоса поверхности дневной Земли в типовом пикселе с площадью 100 мкм^2 за 1 с может генерироваться до 2^{30} электронов, в системе с передачей остаточной части сигнала (мантиссы) внутрипиксельный счётчик должен иметь примерно 15 разрядов (2 байта) плюс примерно 12 разрядов традиционного АЦП для передачи мантиссы.

Если всё квантование сосредоточено в пикселе и реализуется с использованием схем рис. 3.36 или 3.39, то разрядность счётчика должна определяться с учётом деградации контрастной чувствительности согласно формуле (3.5) как из-за уменьшения размера n_k электронов по отношению к шуму считывания, так и в силу потери полезной информации из-за увеличения размера n_k . Оптимальный размер группы n_k можно определить исходя из минимума информационного риска (2.19) при её формировании, учитывая, что, как и для сигнала в целом, дисперсия шума включает дро-

бовый шум сигнальных электронов и шум считывания. Так как при числе электронов в группе, большем 10, их распределение приближённо можно считать гауссовским, минимум информационного риска будет определяться энтропийной мощностью ошибки квантования сигнала на группы и достигаться при размере группы накапливаемых электронов, определяемом из соотношения, аналогичного рассмотренному в 2.4 (с. 115):

$$n_k = \sqrt{2\pi e(n_k + n_{сч}^2)}, \quad \text{откуда} \quad n_k = \pi e \left(1 + \sqrt{1 + \frac{2n_{сч}^2}{\pi e}} \right). \quad (3.6)$$

Контрастная чувствительность связана с размером группы электронов тройко: увеличение размеров группы не только увеличивает потерю полезной информации и уменьшает шумовую информацию, но и ведёт к снижению требуемого быстродействия элементов цифрового пиксела (компаратора и счётчика). Например, при среднеквадратическом значении шума $n_{сч} = 10$ электронов оптимальный размер группы составит $n_k \approx 51$ электрон, а требуемая разрядность внутрипиксельного счётчика $m = \log_2 N/n_k \approx 23\text{--}25$ разрядов (3 байта), требуемое быстродействие 21 МГц. Такое быстродействие компаратора и счётных устройств вполне достижимо, главной особенностью их проектирования является достижение наилучшей энергетической эффективности.

Разрядность пиксела 3 байта соответствует деградации контрастной чувствительности согласно (3.5) в 1,7 раза. Это означает, что *выигрыш* в контрастной чувствительности при ёмкости пиксела $n_e = 2^{15}$ электронов и времени накопления 1 с может составить не 180 раз, как при передаче порядка и мантиссы сигнала (15-разрядный счётчик плюс АЦП на 12 разрядов), а в 100 раз. Это означает, что при пороговом отношении сигнал/шум, равном 5 (принятом А. Роузом [27]), обычный фотоприёмник с активным пикселом для принятых параметров в силу ограниченной ёмкости пиксела n_e и превращения большинства фотонов не в информацию, а в тепло, будет иметь пороговый контраст примерно 2,7 %; фотоприёмник с комбинированным цифровым преобразованием и двухбайтным внутрипиксельным счётчиком – $\approx 0,015$ %; с трёхбайтным внутрипиксельным счётчиком – 0,03 %. При шуме считывания 20 электронов оптимальная группа $n_k \approx 92$ электрона, разрядность счётчика 23–24 разрядов, быстродействие 16,7 МГц, деградация контрастной чувствительности относительно предельного зна-

чения 2,3 раза, предельный контраст 0,035 %. В гипотетическом случае отсутствия шума считывания и наличия только дробового шума оптимальный размер «кучи» электронов составляет $2\pi e$, ближайшее целое число к которому равно 17. При большом уровне шума асимптотическое значение размера группы составляет, как это показано в гл. 2, $n_k = n_{сч} \sqrt{2\pi e}$.

С технологической точки зрения введение такой внутрипиксельной обработки потребует использования более высоких норм точности изготовления СБИС и достаточно большого размера пиксела. Последнее согласуется с общими тенденциями создания систем обнаружения малоразмерных малоконтрастных объектов.

Сравнение динамического диапазона и порогового контраста для активного, логарифмического и цифрового пикселей (рис. 3.41) показывает, что первые два типа пикселей найдут применение в простейших телекамерах; серьёзные системы должны строиться с использованием фотоприёмных матриц с цифровым пикселем высокой разрядности.

Таким образом, потенциальная контрастная чувствительность, лимитируемая шумом флуктуаций фона, а не накопительной способностью пиксела, может быть достигнута в КМОП-фотоприёмной матрице с групповым

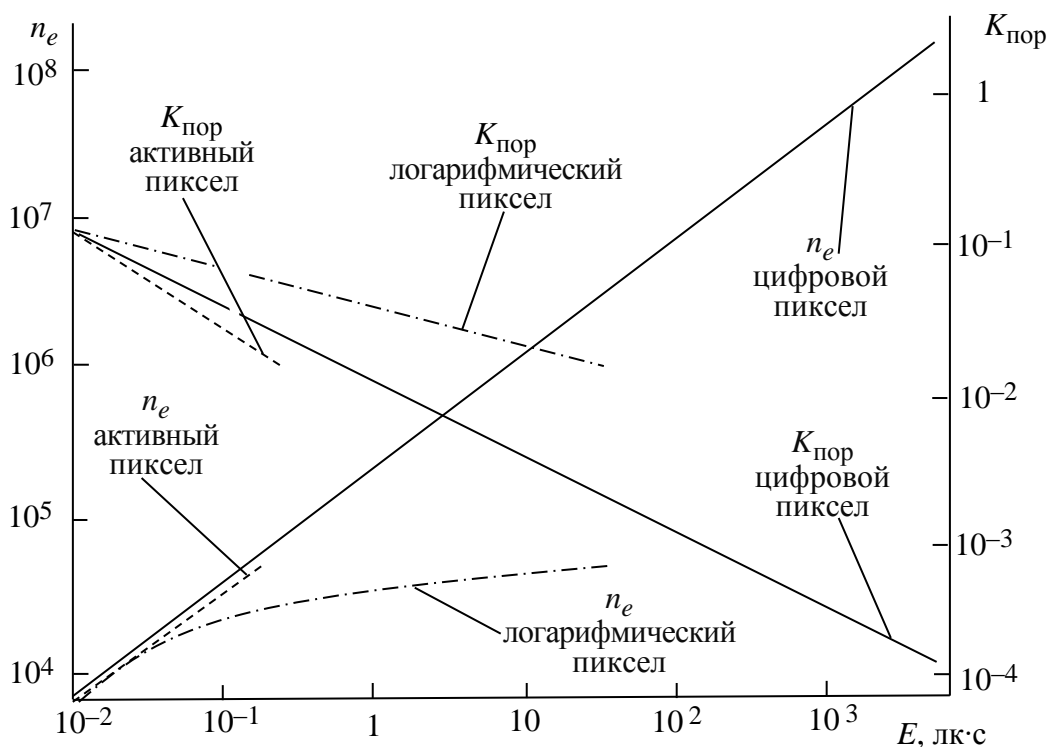


Рис. 3.41. Зависимость числа накапливаемых электронов и порогового контраста для активного, логарифмического и цифрового пикселей при одинаковых площадях фотодиода

счетом фотонов, в которой каждый пиксел включает регулируемый усилитель, компаратор и многоразрядный счетчик. Поэтому следующим шагом развития систем приёма видеоинформации станет применение цифрового пиксела, который за счёт различных механизмов преобразования фотогенерированного заряда в цифровой сигнал позволит исключить влияние внешних компонентов на видеосигнал, а за счёт регистровой обработки видеосигнала значительно расширит динамический диапазон и контрастную чувствительность полученной информации.

Контрастную чувствительность обычно рассматривают по отношению к крупной детали, однако хорошо известно [27], что пороговый контраст при обнаружении зрительным анализатором существенно зависит от площади обнаруживаемого объекта или от пространственной частоты наблюдаемого сигнала. Поэтому при рассмотрении контрастной чувствительности целесообразно учесть её зависимость от пространственной частоты и дополнительных факторов, в первую очередь от кадровой частоты фотоприёмника, которая влияет на предельный контраст изображения.

Частотно-контрастная характеристика (ЧКХ) фотоприёмной матрицы определяется рядом параметров: размером пиксела, соотношением размера фоточувствительной зоны пиксела (фотодиода) и шага расположения пикселов, длиной волны поглощаемого излучения. Аналитические зависимости глубины модуляции от пространственной частоты для матричных фотоприёмников достаточно подробно исследованы [1], [144], [155]. Здесь мы отмечаем лишь результирующий вид ЧКХ, измеряемой по стандартной испытательной таблице (рис. 3.42) с помощью оптической системы, в которой размеры функции рассеяния точки (кружка рассеяния) существенно меньше размера зоны накопления в пикселе, что позволяет измеряемую ЧКХ считать характеристикой не системы, а собственно фотоприёмной матрицы.

Измеренная ЧКХ простейшего трёхтранзисторного пиксела КМОП-матрицы имеет монотонно спадающую форму, которая характеризуется граничной частотой, зависящей от уровня освещённости (отношения сигнал/шум). Форма ЧКХ несколько отличается от идеальных спектральных характеристик, обеспечивающих максимальное качество информации и рассмотренных в гл. 2 (рис. 3.43; внешнее отличие графиков оптимальных кодирующих фильтров от показанных на рис. 2.8 связано с тем, что при теоретическом рассмотрении мы пользовались квадратом спектральной характеристики, здесь – первой степенью).

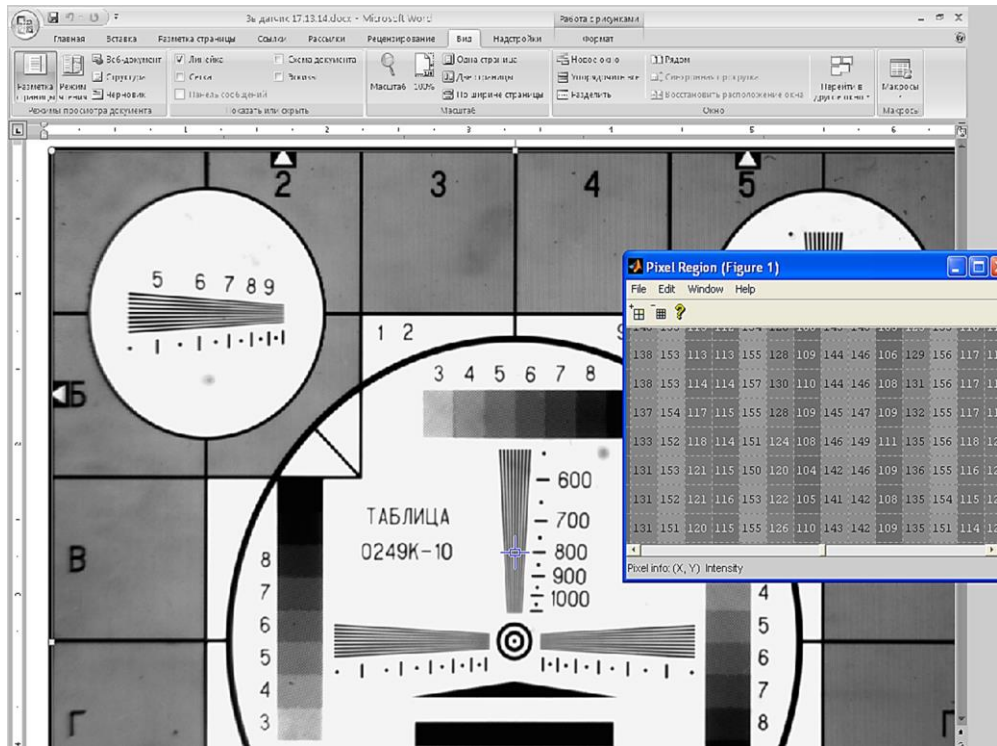


Рис. 3.42. Измерение ЧКХ фотоприёмной КМОП-матрицы (1 мегапиксел) по 1000-строчной квадратной испытательной таблице 0249К-10: курсор наведён на уровень 800 телевизионных линий; справа – окно уровней сигнала выделенного фрагмента испытательной матрицы

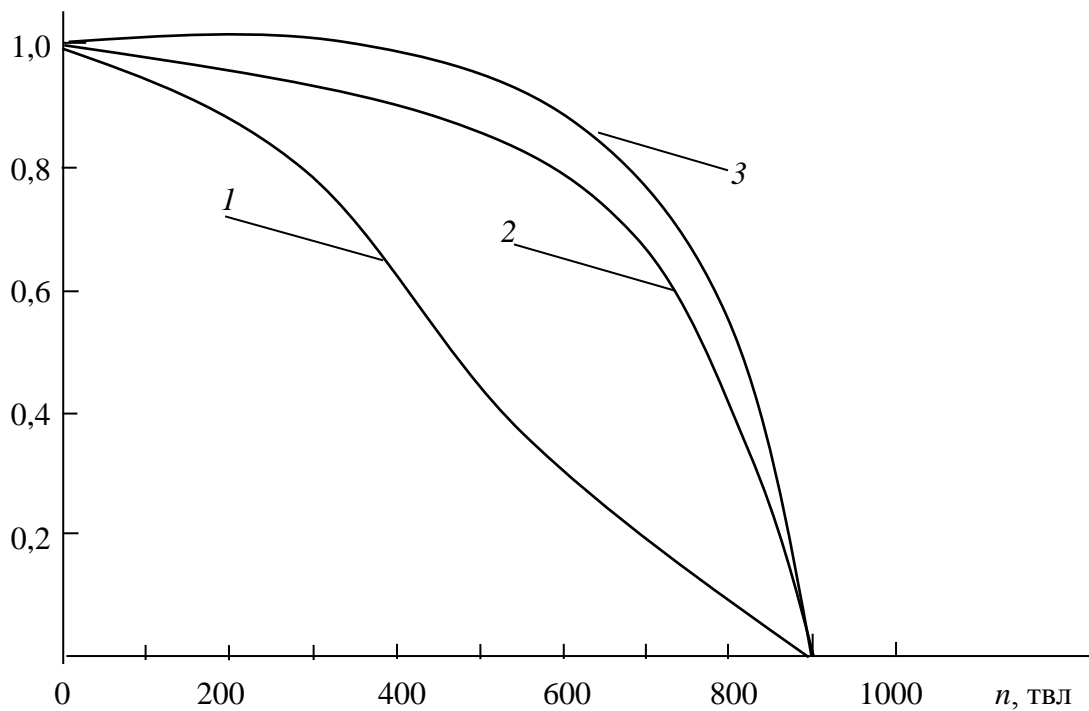


Рис. 3.43. Частотно-контрастная характеристика фотоприёмной матрицы (1 мегапиксел) (1); оптимальная характеристика, обеспечивающая условный максимум качества информации при ограничении широкополосности кодера (2); оптимальная характеристика, обеспечивающая максимум качества информации (3)

Для приведения ЧКХ фотоприёмной матрицы к виду, доставляющему максимум качества информации, традиционно ещё со времён электронно-лучевого телевидения используют апертурную коррекцию. Принципиальное отличие применения апертурной коррекции от оптимизации спектральной характеристики пиксела состоит в том, что кодирование сигнала в пикселе осуществляется до воздействия шума считывания, а апертурная коррекция после него.

Во многих конкретных применениях видеосистем на кристалле возникает противоречие между необходимостью обеспечивать высокое разрешение и кадровую частоту фотоприёмника с требованиями по обеспечению контрастной чувствительности. В этих случаях в составе системы должна быть предусмотрена система оконного или многооконного считывания, позволяющая в общем поле матричного фотоприёмника высокого разрешения выделить области изображения, представляющие наиболее важную информацию. При этом видеосистема может в широком диапазоне изменять параметры считывания небольшой области, достигая оптимального соотношения кадровой частоты, времени экспозиции и контрастности получаемого сигнала. Пример частотно-контрастной характеристики при освещенности 100 лк для единичного пиксела размером 10×10 мкм приведён на рис. 3.44. Вид типовой зависимости характеризуется областью насыщения при низкой кадровой частоте, линейным участком, соответствующим зависимости от накопленного заряда, и спадом в области высоких частот, определяемым ограничениями по переносу заряда в схеме считывания.

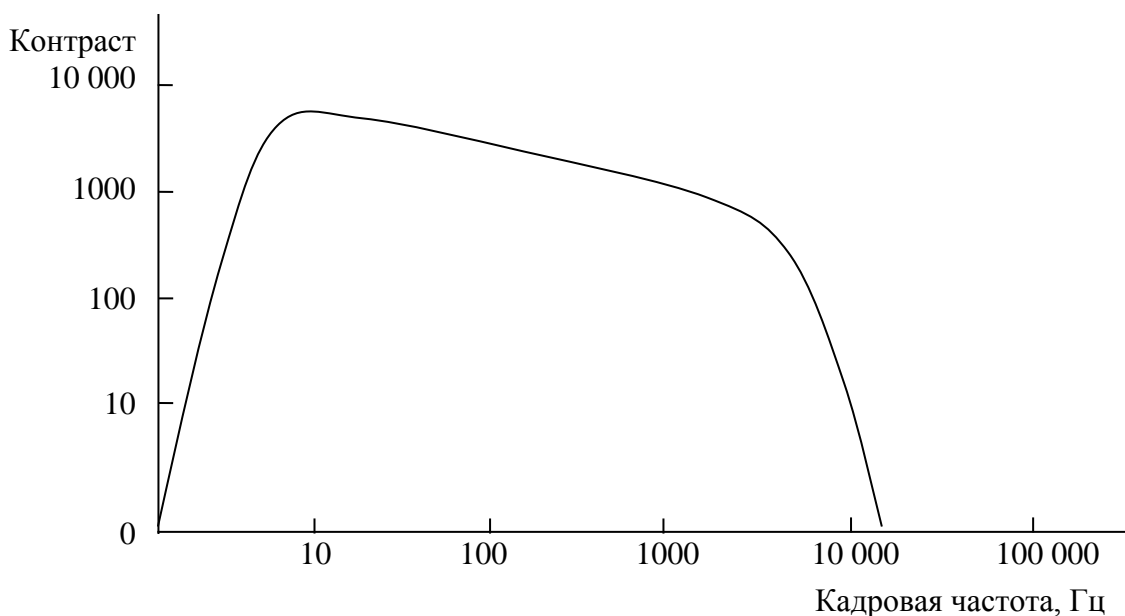


Рис. 3.44. Характеристика зависимости предельного контраста от кадровой частоты КМОП-фотоприёмника

Таким образом, квантовая эффективность рассматриваемых систем достаточно высока, а в случае увеличения количества схем обработки в самой пикселе может быть использована структура фотоприёмника с обратной засветкой с противоположной стороны кристалла. Оптимизация фотоприёмников в зависимости от цели системы должна рассматривать не только организацию пиксела, но и систему считывания с управлением окнами интереса, возможности временного накопления заряда с помощью цифрового пиксела, оптимизацию топологии межсоединений и примыкающих блоков микросхемы.

Выводы по гл. 3. Центральной идеей оптимизации накопления, к которому стремятся все разработчики телевизионных систем, начиная с В. К. Зворыкина и С. И. Катаева, является поиск компромисса между желанием учесть влияние каждого фотона, несущего полезную информацию, и стремлением подавить многочисленные источники помех, в первую очередь фотонных и тепловых, а также ложных сигналов от возможной ионизирующей радиации. Конфликт идеи математической статистики «чем больше данных, тем лучше» и идеи о том, что чем больше интервал накопления по полю и по времени, тем больше отношение сигнал/шум, разрешается с учётом формализованного в гл. 2 понятия качества информации в ходе оптимизации размеров группы фотоэлектронов при реализации цифрового пиксела.

Реализация цифрового пиксела большой разрядности позволяет отказаться от расточительного превращения большинства фотонов в тепло в ходе автоматической регулировки времени накопления и достичь потенциальной – ограниченной фотонным шумом – помехоустойчивости при обнаружении малоконтрастных объектов, обеспечивая широкий динамический диапазон световых сигналов.

Простой пример реализации режима ВЗН в КМОП-фотоприёмнике показывает, что современные фотоприёмники – это принципиально новый класс приборов (видеосистемы на кристалле). Второй, КМОП-й этап твердотельной революции в телевидении породил уникальную возможность – впервые реализованное В. К. Зворыкиным параллельное накопление изображения дополнить параллельным считыванием и выйти на параллельную обработку видеoinформации на том же рукотворном кристалле, на котором находится массив фоточувствительных пикселов.

Видеосистемы на кристалле найдут своё развитие в дальнейшей интеграции фоточувствительной области с трактом обработки видеoinформации. Использование двустороннего совмещения и повышение степени интеграции позволят создать эффективные фотоприёмники с цифровым пикселем и высоким коэффициентом заполнения поверхности кристалла светочувствительной частью пикселей.

Проводя выбор технических решений при разработке новых фотоприёмников для телевизионных камер, разработчик вынужден оперировать огромным множеством параметров. Упростить эту задачу можно, сведя её к достижению максимального качества получаемой информации в рамках ресурсно-технологических ограничений. Рассмотренный подход к анализу и синтезу решений позволяет провести декомпозицию проблемы и оценивать независимо конкретные технические решения.

Особенностью фотоприёмных структур является необходимость достаточно высокого напряжения питания для блоков аналого-цифровой обработки, что требует совмещения на одном кристалле не только различных функциональных блоков, но и различных технологий изготовления транзисторов. Перспективные фотоприёмники будут компоноваться из фотоприёмных массивов на транзисторах с напряжением питания 1,8...1,2 В и блоков цифровой обработки на транзисторах с нанометровыми затворами и напряжением питания 0,7...0,45 В. Динамическое управление питанием и тактовыми частотами позволит сэкономить до половины требуемой энергии и минимизировать нагрев кристалла, ведущий к росту ложной информации из-за темновых сигналов; возможность поддержки эффективно-го кодирования видеопотока решит проблемы пропускной способности каналов связи и помех в сложных условиях применения.

Всё это позволит потребителю получить высокоэффективные видеосистемы на кристалле, соответствующие возможностям оптических систем, проводящие обработку изображения и формирующие требуемый заказчику поток информации на основе эффективного использования доступных ресурсов, способствующие разрешению обозначенных в преамбуле к этой главе конфликтов идей.

Глава 4. ОПТИМИЗАЦИЯ НАКОПЛЕНИЯ КАЧЕСТВА ИНФОРМАЦИИ В ТВЁРДОТЕЛЬНЫХ ТЕЛЕКАМЕРАХ

В этой главе с учётом найденных в первой части оптимальных спектральных характеристик кодирующей системы в ходе накопления сигнала мы рассмотрим проблемы, порождённые конфликтом следующих идей. Во-первых, как и в гл. 2, это конфликт идеи ассоциации понятий информации и пользы и идеи о том, что кроме доминантного объекта в поле зрения телекамеры всегда есть некие фон и шумы.

Во-вторых, это конфликт идеи построения телекамеры на все случаи жизни (типична для телевизионного вещания) и идеи о целесообразности для прикладных телекамер копирования свойств зрительного анализатора, имеющего переменную по полю зрения чёткость и чувствительность.

В-третьих, это конфликт идеи о возможности реализации совместного обнаружения и оценивания параметров сигналов в непрерывных системах и идеи о том, что в аналого-дискретных системах оптимизация обнаружения и оценивания приводит к различным оптимальным значениям размеров зоны накопления.

Накопление качества информации в телекамере можно рассматривать с нескольких сторон. Во-первых, это накопление фотогенерированных электронов в пикселах за время кадра, рассмотренное в гл. 3. Во-вторых, это наращивание качества информации от каскада к каскаду телекамеры. При этом каскады следует понимать обобщённо, включая в них всё, что преобразует информацию, начиная с объектива и кончая цифровым выходом телекамеры, который может отображать видеосигнал и/или результат его обработки специализированным вычислителем, как это имеет место в астродатчиках, передающих координаты и интенсивности звёзд в поле зрения телекамеры [156]. В некотором смысле можно говорить и о накоплении качества видеoinформации в ходе исторического развития технологии создания телекамер.

В телекамерах в процессе накопления качества информации используются стандартные операции: видеосигнал обрабатывается схемой двойной коррелированной выборки, устраняющей шум сброса, фликкер-шум, уменьшающей тактовую наводку, снижающей внешние низкочастотные наводки и восстанавливающей постоянную составляющую. Обработанный схемой ДКВ видеосигнал поступает на управляемый усилитель, коэффициент усиления которого может изменяться в широких пределах как по

командам, так и в автоматическом режиме. Схема фиксации по защищенным от света пикселям в начале каждой строки осуществляет фиксацию уровня чёрного для корректного аналого-цифрового преобразования. Подготовленный видеосигнал оцифровывается аналого-цифровым преобразователем. Цифровой сигнал обрабатывается спецпроцессором для выделения доминантной информации, например путём вычисления координат центров тяжести звёзд, измерения их яркости и количества.

Наращивание качества информации в телекамерах осуществляется оптимизацией параметров их подсистем. При этом параметрический синтез, критерии оптимизации при котором определяются целью системы, должен производиться для всей системы, включающей оптическую, фотоэлектронную и вычислительную подсистемы. Примерами различных целей и соответствующих им критериев являются обнаружение объектов и оценка их параметров, например координат. Широко используемый принцип декомпозиции для синтеза телекамер неприемлем, так как попарное сопряжение оптика–фотоприёмник и фотоприёмник–вычислитель может не обеспечить глобального оптимума качества информации, поскольку выбор конкретного алгоритма вычисления оцениваемых параметров наблюдаемых объектов влияет на оптимальное соотношение размеров площади пикселя и импульсной характеристики объектива (функции рассеяния точки или кружка рассеяния) [3]. Далее мы не будем останавливаться на вопросах повышения качества информации на этапах проектирования объективов, а рассмотрим лишь вопросы согласования объективов с фотоприёмными матрицами и оптимизации размеров зон накопления в матрицах.

4.1. Согласование размера пикселя с сюжетом

Согласование размера пикселя с сюжетом при наблюдении случайных полей – достаточно подробно исследованная задача [1], [43], [157], [158] – здесь охарактеризовано конспективно. Эта задача характерна, в частности, для оптимизации телевизионных систем дистанционного зондирования Земли (ДЗЗ). Априорная информация о доминантном сигнале, фоне и шумах в практических системах часто считается малозначимой, так как опираться на имеющуюся богатую статистику сюжетов при синтезе системы наблюдения не принято. Это связано с тем, что проектировщики стремятся к построению универсальной системы, синтезируемой на основе принципа минимакса: желательно найти условие минимума ошибок при максималь-

но неблагоприятных условиях. Под такими неблагоприятными условиями часто понимают полезные сигналы с максимальной энтропией как в смысле функции распределения, так и в смысле ширины полосы пространственных частот. Время накопления в задаче согласования размера пиксела с ребром Δ и пространственным спектром считается фиксированным. Граничным значением для энергетического спектра сигнала часто принимают квадрат спектральной характеристики объектива, в этом случае необходимость рассмотрения пространственного спектра сигнала считается необязательной, а согласование зоны накопления с сюжетом заменяют её согласованием с импульсной характеристикой объектива (функцией рассеяния точки) [157], [158]. Такой подход соответствует гипотезе о независимости накопленных отсчётов случайного поля, разнесённых на интервал разрешения оптической системы. При этом исследуются различные варианты модели: для дифракционного качества оптики [157] и для абберационной оптики [158].

Основная идея такого согласования сводится к рассуждениям, в которых шумы участвуют, но нет явного критерия, который позволяет найти оптимальный вариант согласования оптической и фотоприёмной подсистем [157]:

1. Для достижения *предельной* разрешающей способности необходимо уменьшать размер пиксела до половины радиуса кружка рассеяния ($\Delta \approx \lambda f / 2D$, где λ – длина волны; f – фокусное расстояние объектива; D – диаметр входного зрачка объектива). При этом признаётся, что такой выбор ведёт к возрастанию фокусного расстояния и габаритов оптической системы примерно в 2 раза, а также к потере части энергии сигнала, в результате чего требуется увеличение времени накопления сигнала для сохранения нужного отношения сигнал/шум.

2. В случае если размер пиксела превышает радиус кружка рассеяния, отношение сигнал/шум при фиксированном времени накопления возрастает, однако пропорционально снижается разрешающая способность, т. е. возможности оптической системы используются не полностью. Этот вариант используется в радиометрических системах, когда разрешение не так важно и требуется увеличить отношение сигнал/шум.

3. Для большинства существующих космических систем оптико-электронного детального наблюдения принят компромиссный вариант равенства размера пиксела и радиуса дифракционного кружка рассеяния.

Рассмотрение абберационной оптики с моделированием функции рассеяния точки гауссовской поверхностью со стандартным отклонением σ (рис. 4.1, а) даёт результат в виде оптимального размера зоны накопления (ребра пиксела), обеспечивающего максимум количества информации как функции отношения сигнал/шум. При этом при изменении отношения сигнал/шум от 10 до 10^3 оптимальный размер зоны накопления изменяется от $1,45\sigma$ до σ . Если традиционно принять размер гауссовской функции рассеяния точки по основанию, соответствующему 3σ , то это результат рекомендует в 2–3 раза более подробную дискретизацию, чем следует из рассмотрения безабберационной оптики и абстрагирования от влияния шума.

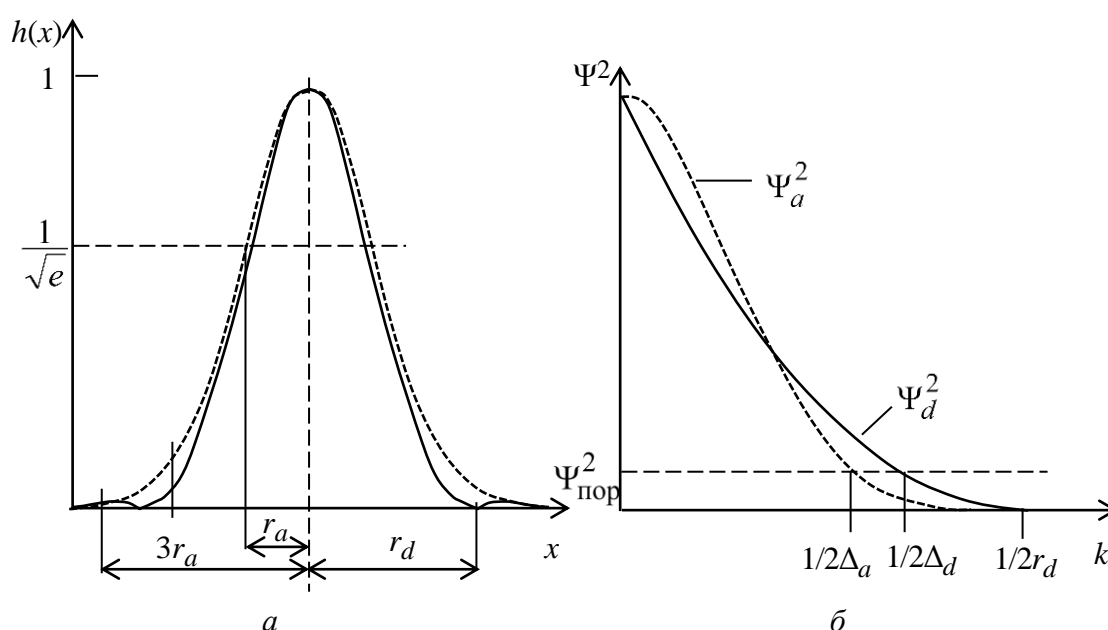


Рис. 4.1. Импульсные характеристики объектива (а) с дифракционным качеством (радиус r_d по первому нулю, сплошные линии) и с абберационными искажениями (радиус r_a по уровню $1/\sqrt{e} \approx 0,606$, пунктирные линии)

и спектральные характеристики (б) для этих импульсных характеристик; оптимальный интервал дискретизации имеет границы Δ_d и Δ_a , определяемые по пересечению спектральных характеристик с пороговым отношением сигнал/шум

С методологической точки зрения опора на поиск максимального размера зоны накопления соответствует главенству оптической системы над фотоприёмником. Вместе с тем согласование параметров зоны накопления с оптической системой – это не антагонистическая игра, обе эти подсистемы, как и вычислительная подсистема измерения параметров наблюдаемых объектов, работают на одну цель – накопление качества информации.

В отличие от этих моделей развитая в гл. 2 (см., в частности, рис. 2.8) концепция усечения спектра случайных сигналов учитывает два источника

шума и может быть применена при любой модели спектра сигнала, в том числе при его аппроксимации спектральной характеристикой объектива – как дифракционного качества, так и абберационного. С точки зрения практики оценки соотношений размеров кружка рассеяния и пиксела, даваемых на основе рассмотренной в гл. 2 модели, получаемая граничная частота спектра $K = 1/2\Delta$ зависит от входного отношения сигнал/шум (рис. 4.1, б). Поэтому вывод о равенстве размеров пиксела и радиуса кружка рассеяния можно рассматривать как границу самой подробной дискретизации, т. е. предел, к которому может стремиться оптимальный размер пиксела при *двух условиях*: 1) спектр полезного сигнала определяется частотно-контрастной характеристикой объектива, сглаживающей сигнал с существенно бóльшими пространственными частотами; 2) входное отношение сигнал/шум стремится к бесконечности. Это означает, что рассмотренная в гл. 2 модель с двумя источниками шума даёт *оптимальное* соотношение размеров пиксела к радиусу кружка рассеяния объектива несколько большее, чем компромиссное значение [157], и заметно большее, чем учитывающее только один из источников шума [158].

4.2. Согласование размера пиксела с сигналом точечного объекта

Эта задача [1]–[3], [43], [159], [160] типична для оптимизации телекамер при простейшей форме априорной информации о доминантном сигнале, фоне и шумах. Общая формула (3.1) здесь рассматривается упрощённо, время накопления в задаче согласования размера пиксела с ребром Δ и функции рассеяния точки объективом считается фиксированным. Форма полезного сигнала обычно не известна абсолютно точно; бывают известны диапазоны изменения её основных характеристик. Можно считать, что априорная неопределённость не функциональная, а параметрическая. В этом случае вариации формы полезного сигнала могут быть сведены к вариациям некоторых параметров: амплитуды, радиуса (зависящего от качества фокусировки объектива), эксцентриситета (зависящего от наличия комы объектива) и т. п. Априори считается известной форма сигнала $S(x, y) = \alpha s(x - x_0, y - y_0)$, часто моделируемая двумерной гауссовской поверхностью с круговой симметрией и радиусом r по уровню $1/\sqrt{e} \approx 0,606$ с параметрической неопределённостью о координатах объекта x_0, y_0 , которые считаются случайными величинами, равномерно распределёнными на всём интервале наблюдения. Сигнал формируется за счёт генерации пада-

ющим потоком фотонов N_s электронов. Фон, аддитивный, с полезным сигналом, считается равномерным с плотностью потока фотонов, генерирующих B электронов на единицу площади, т. е. $N_b = B\Delta^2$ электронов в каждом пикселе. Шум содержит три независимых компонента: шум сигнала объекта (звезды) и шум фона, имеющие пуассоновское распределение с дисперсиями, соответственно, N_s и N_b , и шум считывания, для которого принимается гауссовское распределение с дисперсией D .

Обнаружение сигналов точечных объектов. При проверке гипотезы о наличии сигнала минимум среднего риска достигается максимизацией отношения сигнал/шум Ψ . Однако при описанной априорной информации и заданной модели системы, осуществляющей дискретно-аналоговое преобразование, решение уже не сводится к согласованному фильтру, обеспечивающему максимум отношения сигнал/шум в непрерывной системе при аддитивном гауссовском шуме. Зависимость отношения сигнал/шум от числа сигнальных N_s фотоэлектронов, числа фоновых N_b фотоэлектронов и дисперсии D шума считывания составляет:

$$\Psi^2 = \frac{N_s^2}{N_s + N_b + D}.$$

Усреднённое по фазе полезного сигнала относительно дискретной структуры твёрдотельного фотоприёмника отношение сигнал/шум вычисляется с учётом шумов фона и считывания, дисперсии которых не зависят от номера отсчёта. При оценке амплитуды сигнала и проверке гипотезы о его наличии обычно рассматривается отсчёт, содержащий максимум сигнала; при оценке координаты сигнала – группа отсчётов, содержащая максимальный сигнальный отсчёт. При обнаружении сигнала полагается, что решение о его наличии принимается путём сравнения амплитуд сигналов без дополнительной фильтрации. Это означает, что согласование площади зоны накопления с размером кружка рассеяния (полезным сигналом звезды) является исчерпывающей процедурой оптимизации. Моделирование с использованием вычисления значений отсчётов полезного сигнала с помощью табулированной функции интеграла вероятностей и с усреднением значений сигнала по фазе координаты центра сигнала относительно дискретной структуры раstra (рис. 4.2) показывает, что существует оптимальное значение соотношения Δ/r размера пиксела и радиуса импульсной характеристики объектива (рис. 4.3).

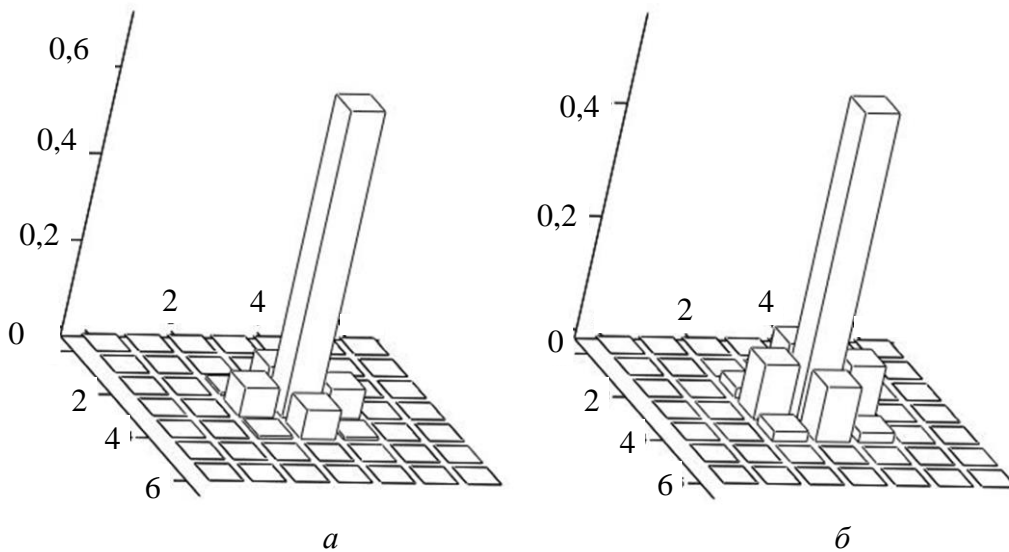


Рис. 4.2. Распределение сигнала звезды по пикселям при относительном размере кружка рассеяния Δ/r , равном: $a - 4$, $б - 3$

В простой бинарной задаче обнаружения качество информации определяется средним риском, включающим согласно критерию Байеса взвешенную сумму ошибок двух видов – пропуска сигнала и ложной тревоги. Анализ показывает, что качество информации (максимум помехоустойчивости) при обнаружении зависит от интенсивности сигнала звезды α и интенсивности фона B (рис. 4.3).

Как и в модели непрерывного сигнала и аддитивного белого гауссовского шума, параметр Δ/r не зависит от интенсивности сигнала объекта α . Однако в связи с новыми элементами модели возникает зависимость относительного размера зоны накопления от отношения потенциального контраста объекта относительно фона α/B , который определяет не только достижимое отношение сигнал/шум, но и оптимальный размер зоны накопления.

В широком диапазоне изменения соотношения шумов фона и считывания оптимальный по критерию отношения сигнал/шум нормирован-

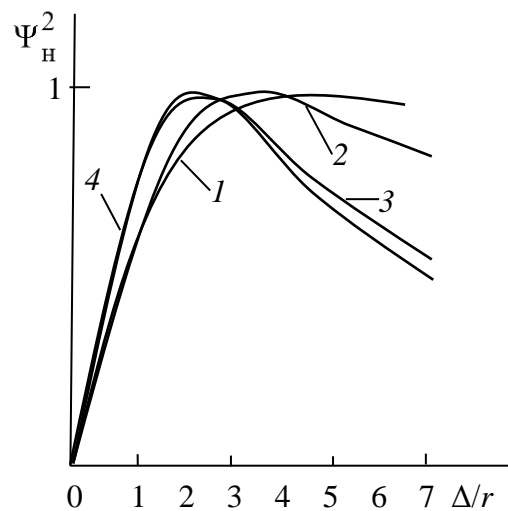


Рис. 4.3. Зависимость нормированного отношения сигнал/шум $\Psi_n^2 = \Psi^2 / \Psi_{\max}^2$ от соотношения размеров зоны накопления и функции рассеяния точки объективом Δ/r при постоянных уровнях полезного сигнала $\alpha = 100$ и шума считывания $D = 3$ и различных значениях фона B :
 $1 - 1$; $2 - 10$; $3 - 100$, $4 - 1000$

ный на радиус сигнала звезды размер элемента разложения изменяется в пределах от 3 до 7 [3] (рис. 4.3).

Такая методика позволяет осуществить оценку помехоустойчивости обнаружения и найти ориентировочное значение оптимального соотношения размеров зоны накопления и кружка рассеяния объектива в условиях действия трёх источников шума, т. е. в широком диапазоне изменения контраста звезды относительно фона.

Оценивание координат сигналов точечных объектов. Построение оптимальной телевизионной камеры для второй задачи астрономических телевизионных систем – оценивания координат изображения объекта – распадается на *взаимосвязанные* задачи:

- вычисление оптимального размера зоны накопления Δ ;
- выбор оптимального (приемлемого по вычислительным затратам) алгоритма вычисления оценки;
- определение оптимального шага квантования результатов вычисления координат объектов (размера макрокванта информации), связанного со среднеквадратической ошибкой измерения координат.

Связь этих задач проявляется в том, что значение оптимального размера зоны накопления зависит от используемого алгоритма. Это подчёркивает единство телевизионной системы, в которой три важнейших звена – объектив, фотоприёмник и устройство обработки – не могут просто сопрягаться попарно, а должны оптимизироваться совместно [3].

Компромиссный с позиций точности и сложности алгоритм центра тяжести (центроида) проанализируем для одномерного дискретного сигнала, когда оценка координаты вычисляется по формуле, следующей из определения центра тяжести [1]–[3], [156], [159], [160]:

$$\hat{x} = \sum_i i \Delta U_i / \sum_i U_i .$$

Метод центра тяжести (центроида) обосновывают трактовкой отклика оптической системы на излучение точечного источника как плотности вероятности его координат. При этом считается, что *«процедура получения оценок координат по центру тяжести не зависит ни от вида, ни от масштаба реакции оптической системы на излучение точечного источника»* [159]. Для алгоритма центроида вычисления по каждой оси ведут по трём отсчётам, т. е. оптимальный по критерию минимума среднеквадратической ошибки размер элемента разложения считают близким к удвоенному ради-

усу r [1], [159]. Этот результат, полученный в предположении о преобладании собственных шумов фотоприёмной матрицы, уточняется в соответствии с моделью, учитывающей влияние фотонного шума сигнала и фона [3]. При этом изменение модели не влияет на важнейшие два проявления шума [48]: «Шум вызывает два явления: одно – обычное уменьшение точности, другое – неоднозначность. Оценка может быть отброшена от своего правильного положения настолько, что она будет казаться соответствующей совершенно ложному значению координаты». Поэтому вариация соотношения размеров зоны накопления и импульсной характеристики объектива ведёт к изменению соотношений двух видов ошибки – нормальной, которая уменьшается при уменьшении размера зоны накопления, и аномальной (назначение на роль пиксела с максимумом сигнала соседнего с ним пиксела), увеличивающейся при уменьшении размера зоны накопления из-за распределения конечного числа фотоэлектронов по всё большему числу зон накопления (рис. 4.4).

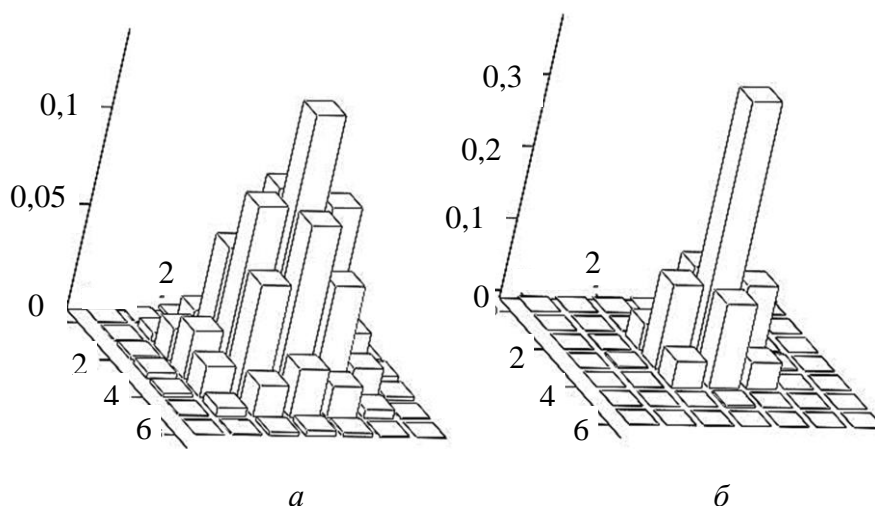


Рис. 4.4. Распределение сигнала звезды по пикселям при относительном размере кружка рассеяния Δ/r , равном: $a - 1$, $b - 2$

Расчёт показывает, что в широком диапазоне изменения контраста объекта (соотношения интенсивности сигнала и шумов фона α/B) и шума считывания имеется оптимальное значение размера элемента $\Delta/r_{\text{opt}} \approx 1,2 \dots 1,5$ (рис. 4.5). При этом качество информации, формируемой телекамерой астродатчика, определяется согласно общей формуле (2.28) через полученное количество информации о координате объекта и информационный риск, определяемый потерей информации при квантовании результатов измерения координат и шумовой информацией, определяемой шумовой ошибкой, входящей в передаваемый результат измерения.

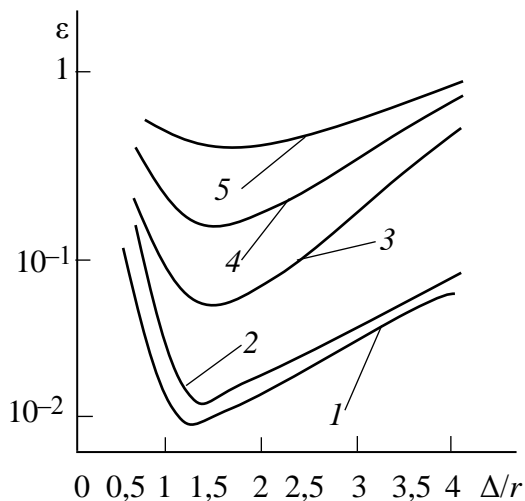


Рис. 4.5. Зависимость среднеквадратической ошибки измерения координат звезды от соотношения размеров элемента разложения и функции рассеяния точки объективом при различных соотношениях шума фона и собственного шума телекамеры: 1: $\alpha = 100, B = 5, D = 5$; 2: $\alpha = 100, B = 5, D = 10$; 3: $\alpha = 100, B = 50, D = 5$; 4: $\alpha = 100, B = 100, D = 5$; 5: $\alpha = 10, B = 1, D = 1$

Потенциальное значение полезной информации в этой задаче согласно формуле Вудворда [48] определяется логарифмом отношения априорного интервала (площади фотоприёмной матрицы) и «площади неопределённости», соответствующей произведению ошибок по двум координатам матрицы.

Это означает, что расчёт, учитывающий три источника шума (полезного сигнала, фона и считывания) и влияние двух типов ошибок при использовании алгоритма измерения центра тяжести, конкретизировал правило дискретизации сигнала звезды на три элемента по каждому направлению для дости-

жения минимума ошибки измерения координат. При этом, как и при доминировании шумов телекамеры, точность измерения координат растёт с увеличением отношения сигнал/шум. При соответствующей калибровке системы (например [156]) для коррекции неоднородности чувствительности элементов матрицы ошибка измерения координат при типовой ёмкости элемента разложения порядка 10^5 электронов может достигать нескольких сотых долей элемента разложения.

Сопоставление зависимостей отношения сигнал/шум и среднеквадратической ошибки оценки координат от соотношения размеров зоны накопления и функции рассеяния точки объективом показывает, что их экстремумы достигаются при существенно различных значениях аргумента. Самым важным является то, что интервал дискретизации, оптимальный для оценивания координаты, практически постоянен $\Delta/r_{\text{opt } \varepsilon} \approx \text{const}$, а оптимальный для обнаружения — зависит от контраста и соотношения шумов фона и камеры $\Delta/r_{\text{opt } \psi} = \text{var}$. Важным конструктивным результатом такого сопоставления является тезис: *изменение критерия помехоустойчивости при переходе от обнаружения к оцениванию координат ведёт к переменным размерам зоны накопления.*

4.3. Совместное обнаружение и оценивание сигналов с переменным размером зоны накопления

Кроме простейших задач обнаружения и оценивания параметров сигнала на практике часто возникает потребность совместного обнаружения объектов и оценивания их параметров [161]. Эта ситуация характерна, например, для датчиков звёздной ориентации, целью которых является определение ориентации космических аппаратов по результатам сравнения наблюдений с каталогами звёзд. *Априорная информация* в этой задаче заключена, так же как и в задачах обнаружения и оценивания координат звёзд, в форме сигнала, определяемой функцией рассеяния точки объективом (часто принимаемой гауссовской), диапазоном яркостей звёзд, измеряемых через создаваемую звездой освещённость E в звёздных величинах $m = -13,89 - \log_{2,5} E = -13,89 - 2,5 \lg E$. Распределение яркостей звёзд по яркостям экспоненциальное (количество звёзд в поле зрения экспоненциально растёт по мере увеличения звёздной величины m). Распределение звёзд по полю зрения считается равномерным, как это принято и в радиолокационных системах обнаружения–оценивания [48].

Доминантной информацией в этих приборах выступают координаты наблюдаемых звёзд, *качество информации* определяется количеством звёзд и точностью измерения их координат. Важно, что обнаружено звёзд в поле зрения телекамеры должно быть много, при малом их количестве (недостаточной чувствительности телекамеры) точность опознавания участка звёздного неба катастрофически падает. Это означает, что хотя точность измерения координат слабых звёзд и существенно хуже, чем ярких звёзд, добавление их координат заметно улучшает надёжность измерения ориентации космического аппарата.

Развивавшаяся в первой части книги идея накопления качества информации группированием фотогенерированных электронов в макрокванты в приложении к астродатчикам приобретает вполне конкретное выражение – изменение размеров макроквантов в зависимости от сюжета. В любых фотоприёмных матрицах – с цифровым или аналоговым пикселем – макрокванты должны формироваться не всюду одинаково, а в зависимости от интенсивности сигнала звезды. В фотоприёмной матрице астродатчика фотоэлектроны группируются в макрокванты («кучи») *с переменными параметрами*: с увеличением интенсивности сигнала объекта

группы фотоэлектронов укрупняются в силу роста дисперсии фотонного шума. Важнейшей чертой такого группового счёта фотоэлектронов является изменение геометрических размеров зоны накопления: при слабых сигналах необходимо увеличивать площадь накопления для повышения вероятности обнаружения сигнала звезды; при интенсивных сигналах площадь зоны накопления уменьшается для уменьшения среднеквадратической ошибки измерения координат звёзд, т. е. для повышения качества информации. Это изменение в определённом смысле близко к адаптации, но отличается от неё тем, что параметры телекамеры не зависят от времени (что рассматривается в гл. 5). Вводимое различие размеров зоны накопления от интенсивности сигнала в силу постоянства параметров перехода от одних размеров площади зоны накопления к другим роднит эту процедуру с нелинейной обработкой сигналов изображений, известной как гамма-коррекция. Такое изменение площади зоны накопления имеет также аналогию в виде АРУ «по точечному объекту».

Подробное описание алгоритма перехода к различным размерам зоны накопления дано в [3], там же детально обосновывается конкретное количественное наполнение изложенной идеи, дающее вычисление оптимальных соотношений размеров пиксела и кружка рассеяния. Соотношение оптимального размера элемента и радиуса кружка рассеяния объектива при измерении координат и обнаружении позволяет строить систему так, чтобы при измерении координат размер элемента составлял $1,2 \dots 1,5$ радиуса кружка рассеяния (при его гауссовской аппроксимации), а при обнаружении увеличивался в 2–3 раза [3].

Согласование зоны накопления с кружком рассеяния объектива может производиться группированием зарядовых пакетов до передачи через выходное устройство, т. е. методом переменной по полю чёткости ([2], см. 4.4), уменьшающим влияние шумов канала считывания зарядового изображения. Переменная чёткость, реализуемая непосредственно в фотоприёмной матрице, принципиально отличается от таковой, организованной компьютерной обработкой считанного обычным способом изображения. Указанное отличие состоит в том, что шум считывания в обычной системе добавляется к каждому элементу разложения, а в системе с переменной чёткостью он добавляется один раз на группу элементов.

Вместе с тем при создании твёрдотельных систем реального времени, например астродатчиков или систем контроля космического пространства,

заранее разделить кадр на фрагменты с полной и сокращённой чёткостью, как это описано в [2], удастся редко. Изменение площади накопления информации от точечных объектов в зависимости от интенсивности сигнала позволяет достичь доминантной цели – потенциальной помехоустойчивости и для обнаружения, и для оценивания координат. При интенсивных сигналах звезды дискретизация должна быть подробной (предельной для матрицы), относительный размер элемента $\Delta \approx 1,5r$. При малой интенсивности сигнала должно осуществляться увеличение размера зоны накопления. Такая перестройка требует включения в состав телекамеры порогового устройства, которое обычно используется в вычислителе оценок координат точечных объектов. Значение порога определяется через функцию $\Phi^{-1}(p_{л.т.})$ (обратную интегралу вероятности) от заданной вероятности ложной тревоги $p_{л.т.}$. Превышение порога сигналом конкретного элемента переводит считывающее устройство в режим сброса после каждого элемента, а при сигналах, меньших порога, считывание производится с группированием зарядовых пакетов [3]. Такое группирование сопровождается управлением нормировкой сигнала в видеотракте, учитывающей размер зоны накопления. При сложении двух зарядовых пакетов устанавливается коэффициент передачи $K = 0,5$, а при сложении трёх зарядовых пакетов $K = 0,33$. Реализация алгоритма работы телекамеры с переменной площадью накопления позволяет при сохранении требований к уровню ложной тревоги повысить чувствительность к слабым источникам. Возможный выигрыш в качестве информации определяется соотношением значения порогов перехода от одной площади зоны накопления к другой.

В зависимости от соотношения уровней шумов фона и считывания соотношение порогов, а следовательно, и выигрыш в чувствительности изменяются. При практической реализации описанной системы уровень шума считывания измеряется по защищённым от воздействия оптического излучения элементам фотоприёмной матрицы, а уровень шума фона (как и собственно уровень фона) – в процессоре обработки видеосигнала. В зависимости от уровня фона расчётный выигрыш в чувствительности, учитывающий и изменение порога обнаружения (уровня шума), составляет для группирования двух зарядовых пакетов от 1,1 до 1,6, а при группировании трёх зарядовых пакетов – от 1 до 1,8 [3].

Экспериментальная проверка метода переменной площади накопления производилась на специально разработанной телекамере [3] на охлаждаемом с помощью двухкаскадного термоэлектрического холодильника матричном ПЗС с прогрессивной развёрткой и строчным переносом ICX415AL фирмы «Sony». Для повышения точности быстрого измерения среднего значения темнового сигнала матриц применён метод сложения непосредственно в матрице зарядовых пакетов закрытых от света элементов считывающего регистра. Формат матриц ПЗС – 1/2" со светочувствительной поверхностью 4,8×6,4 мм; переключаемое время накопления:

- режим 1: кадровая частота 25 Гц (8 бит/элемент);
- режим 2: время считывания 0,2 с, 12 бит/элемент; время накопления 1, 3, 8, 16, 32, 64, 128, 180 с устанавливается программно.

Результаты эксперимента подтвердили расчёты и показали, что метод переменной площади накопления реализуем и даёт выигрыш в помехоустойчивости, иллюстрируемый рис. 4.6, на котором:

а – при малой площади накопления и заданной вероятности ложной тревоги в элементе разложения $p_{л.т} < 10^{-6}$ обнаруживается только два объекта из трёх имеющих;

б – при той же малой площади зоны накопления для обнаружения третьего объекта необходимо понизить порог до уровня, приводящего к большой вероятности ложной тревоги;

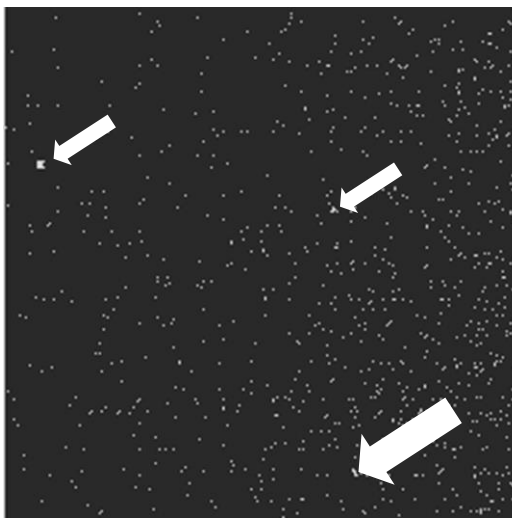
в – при переменной площади зоны накопления два интенсивных сигнала дискретизируются подробно, но дополнительно обнаружен грубодискретизированный слабый сигнал третьего объекта.

Показанные на рис. 4.7 пороги 1 и 2 обеспечивают при различных режимах дискретизации одинаковое значение ложной тревоги, меньшее 1 за кадр. Выигрыш в чувствительности, как показывают расчёт и эксперимент при различных условиях наблюдения (при изменениях интенсивности внешнего фона и температуры матрицы ПЗС), уменьшается при увеличении как собственных шумов телекамеры, так и шумов внешнего фона.

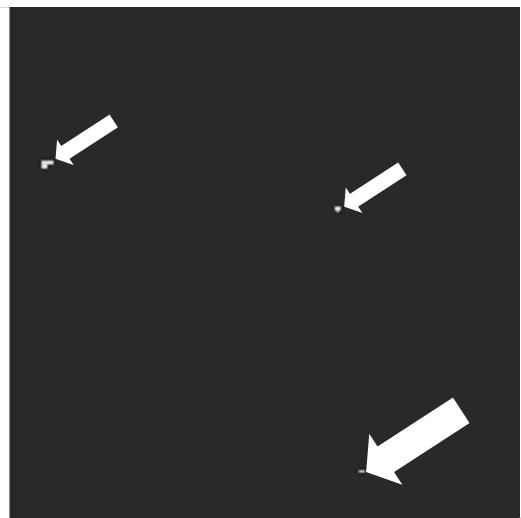
При использовании разработанного метода адаптации для сигналов, уровень которых ниже значения порога 1, достигаемая точность измерения координат в силу более грубой дискретизации, конечно, хуже (уже не «субпиксельная»), чем для сигналов большой интенсивности.



a

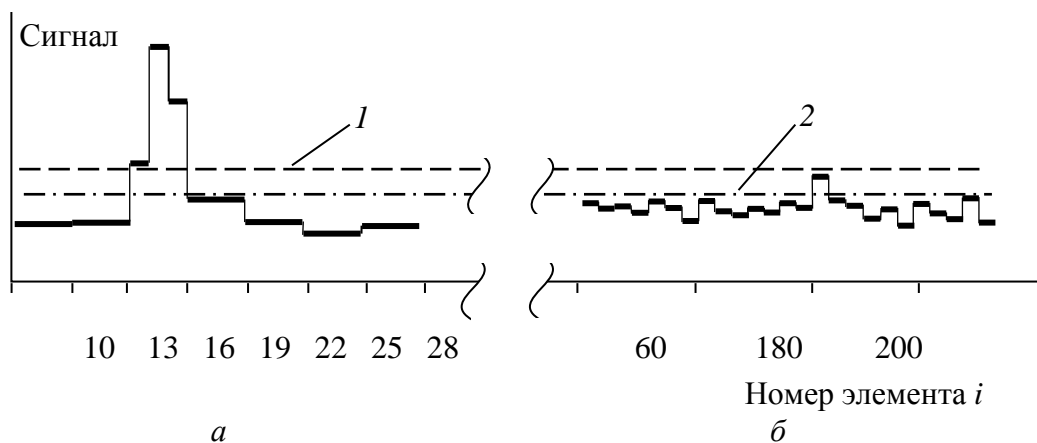


б



в

Рис. 4.6. Иллюстрация эффективности метода переменной площади зоны накопления



a

б

Рис. 4.7. Осциллограммы сигнала при адаптивной дискретизации. Интервал дискретизации для слабых сигналов, не превышающих порог I , в 3 раза больше, чем для интенсивных сигналов (*a*); слабый сигнал, не превышающий порог I при подробной дискретизации, превышает порог 2 и обнаруживается при адаптивной дискретизации (*б*)

Вместе с тем достигнутый выигрыш в чувствительности обеспечит наблюдение дополнительного числа слабых объектов в поле зрения при том же времени накопления.

В системах астроориентации требования к точности измерения координат звёзд экспоненциально уменьшаются при увеличении количества наблюдаемых звёзд. Поэтому, как уже указывалось, измерение координат слабых звёзд даже с пониженной точностью (до одного элемента разложения) позволяет увеличить достоверность идентификации наблюдаемого участка звёздного неба.

Значимость, казалось бы, незначительного выигрыша в энергетике сигналов определяется существенным приростом информации при классификации созвездий в системах астроориентации, что иллюстрирует роль информационной парадигмы при проектировании твёрдотельных телевизионных систем обнаружения–оценивания сигналов.

Таким образом, синтез дискретно-аналоговой системы оптимального совместного обнаружения и оценивания координат объектов в условиях влияния нескольких источников шума показал, что, несмотря на различие в оптимальных размерах зоны накопления при обнаружении объектов и оценивании их координат, согласно известной концепции [161] можно построить систему, формирующую эти различные решения по одной реализации процесса на входе решающего устройства, в роли которого выступает телекамера.

4.4. Группирование пикселов для повышения контрастной чувствительности

Согласование параметров разложения изображения с пространственно-временными характеристиками сюжета в соответствии с (3.1) подчинено накоплению качества информации. При этом широко применяются методы обмена разрешающей способности на чувствительность группированием сигналов соседних пикселов [1], [3] [162] и подчёркивается возможность снижения уровня пороговой освещённости благодаря росту отношения сигнал/шум для крупных деталей изображения. Эта идея нашла воплощение в широко используемых телекамерах с «ночным» режимом, в которых по мере уменьшения освещённости группируются сигналы соседних пикселов и кадров. Все демонстрации эффективности такого режима работы телекамер при низкой освещённости проводятся на примере объектов большого оптического контраста.

Вместе с тем группирование сигналов соседних пикселей имеет важное значение ещё в одном классе прикладных телевизионных систем, предназначенных для обнаружения малококонтрастных объектов. В этих системах, например при горизонтном обнаружении объектов [163], яркость фона достаточно велика, но контраст объектов из-за рассеяния света в атмосфере в соответствии с законом Бугера падает по экспоненте от расстояния. Специально введённая для этих условий наблюдения величина называется метеорологической дальностью видности (МДВ) – это дальность, на которой при данных погодных условиях контраст чёрного объекта относительно светлого неба у горизонта падает до уровня 2 %, выбранного исходя из контрастной чувствительности зрения.

Возможности обнаружения малококонтрастных объектов телевизионными системами во многом связаны с их контрастной чувствительностью, определяемой как пороговый обнаруживаемый контраст. Он определяется фотонным шумом фона и вычисляется через ёмкость n_e пикселя матрицы в электронах, а также принятым пороговым отношением сигнал/шум $\Psi_{\text{пор}}$ [27] по (3.2). В разных системах пороговое отношение сигнал/шум $\Psi_{\text{пор}}$ принимается различным, как правило, в пределах 2...10; А. Роуз [27] приходит к значению $\Psi_{\text{пор}} = 5$. Обычно эта величина определяется заданными вероятностями обнаружения и ложной тревоги; например, при гауссовских шумах и вероятности обнаружения $p = 0,9$ и вероятности ложной тревоги не более одной в кадре $\Psi_{\text{пор}} \approx 7,5...8$.

Ёмкость пикселя в электронах для матриц, изготовленных по разным технологиям, различна, но имеет ориентировочное значение 1000 электронов/мкм². Для типовых матриц ёмкость пикселя составляет 30 000...60 000 электронов. Эти размеры пикселей при $\Psi_{\text{пор}} = 8$ дают значение порогового контраста 3,3–4,6 %. Это означает, что телекамеры на типовых матрицах в стандартных режимах не обеспечивают дальности обнаружения, определяемой метеорологической дальностью видности.

Для повышения дальности обнаружения малококонтрастных объектов необходимо использовать матрицы с большой ёмкостью пикселя. Радикальным методом повышения контрастной чувствительности является использование цифрового пикселя большой разрядности (см. гл. 3). В рамках технологии аналогового активного пикселя важнейший путь повышения контрастной чувствительности – увеличение активной площади пикселя. Для серийных матриц с умеренной площадью пикселя этот путь означает обращение к сложению сигналов соседних пикселей (так называемому

«бинированию»). Эффективность группирования сигналов соседних пикселей несколько различается при реализации её в аналоговой и в цифровой формах, так как накопление после квантования ведёт к накоплению ошибок квантования.

Конечно, бинирование приводит к уменьшению чёткости изображения, формируемого матрицей, и в простых системах охранного телевидения с этим злом принято мириться. В системах пригоризонтного обнаружения малоконтрастных объектов, где необходимо распознавание классов обнаруживаемых объектов, деградация чёткости недопустима. Поэтому основным путём повышения дальности обнаружения является совместное использование трёх положений:

- группирование сигналов соседних пикселей по 2×2 , 3×3 , 4×4 с соответствующим улучшением контрастной чувствительности в 2, 3 и 4 раза;
- формирование сборок матриц по 2×2 , 3×3 , 4×4 ;
- увеличение фокусного расстояния объектива в 2, 3 и 4 раза.

Эксперимент с использованием телекамеры на КМОП-матрице с размером пиксела $3,5 \times 3,5$ мкм² (размер матрицы 3002×2210 элементов, объ-

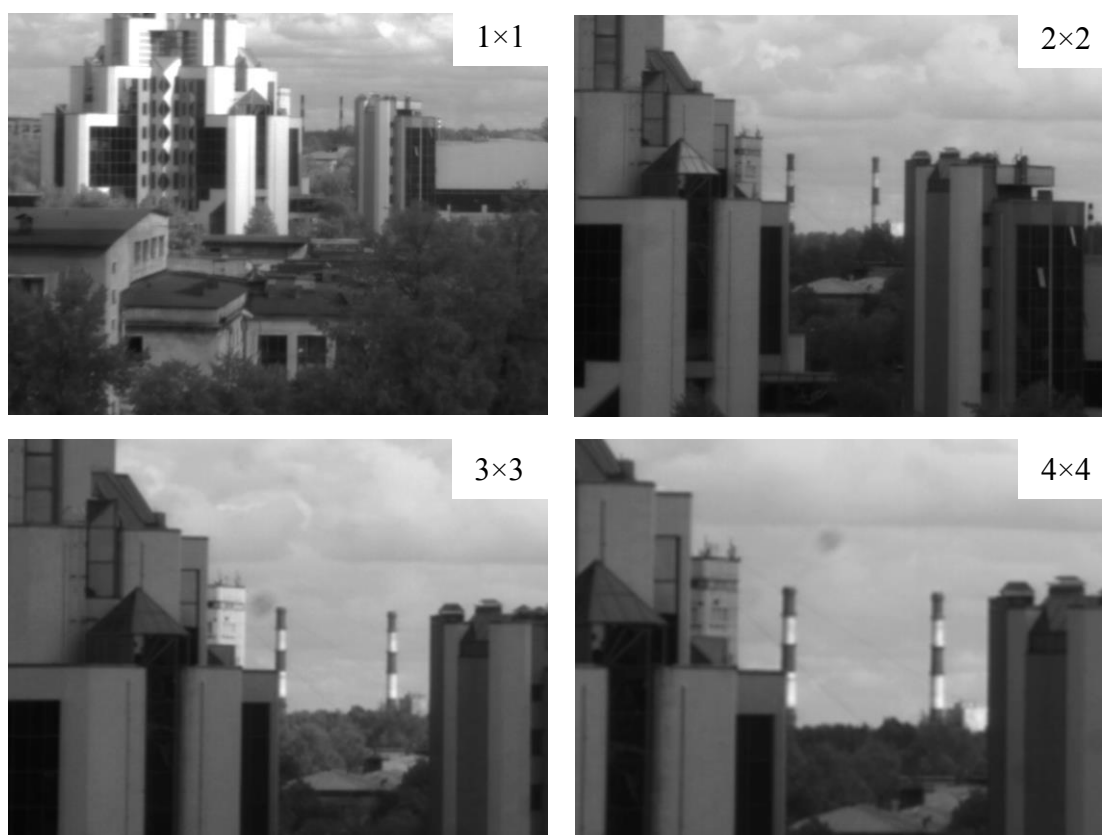


Рис. 4.8. Влияние сложения сигналов соседних пикселей на отношение сигнал/шум и контраст удалённых объектов; тестовый объект – трубы с участками светлее и темнее фона на дистанции 4,3 км при МДВ 6 км

ектив «Sony» с переменным фокусным расстоянием, изменяемым для поддержания постоянной угловой разрешающей способности при сложении пикселов [163] (рис. 4.8)) показывает, что отношение сигнал/шум в перепаде между удалённым объектом и фоном можно увеличить по сравнению с передачей изображения полной чёткости.

Анализ результатов наблюдения (рис. 4.9) показал [163], что по мере увеличения числа группируемых пикселов отношение сигнал/шум малококонтрастного объекта возрастает, а повышение контрастной чувствительности приближается к корню квадратному от числа суммируемых сигналов;

различия в качестве видеoinформации для объектов с разными знаками контраста наблюдаются при разных метеоусловиях в силу заложенных в КМОП-матрицу неуправляемых коррекций уровней «белого» и «чёрного». Это подтверждает тезис о том, что группирование сигналов соседних пикселов серийных матриц позволяет достичь дальности пригоризонтного обнаружения, близкой к МДВ. Для заметного увеличения дальности, как уже указывалось, необходим переход к цифровому пикселу большой разрядности (см. гл. 3).

При различных метеоусловиях контраст объектов не только меняется по абсолютному значению, но и может в зависимости от угла «Солнце–объект–наблюдатель» приобретать знакопеременный контраст. Это ведёт к использованию в устройстве обнаружения известной структуры обнаружения случайных сигналов «фильтр–квадратор–интегратор» [88].

4.5. Переменная чёткость в твёрдотельных телекамерах

Разработчики прикладных телевизионных камер при решении многих вопросов сверяют свои решения со свойствами зрительного анализатора человека. Так как чёткость зрения убывает к периферии и максимальна только в небольшой центральной зоне (называемой *fovea*), то были приняты многочисленные попытки создания прикладных телевизионных

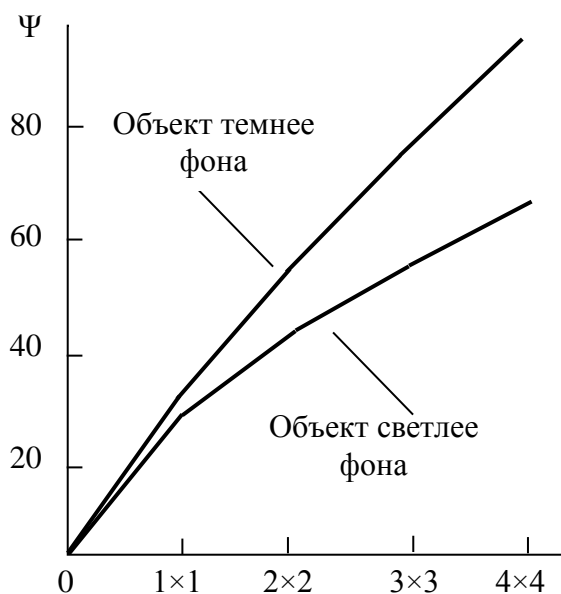


Рис. 4.9. Экспериментальная зависимость отношения сигнал/шум в тестовом объекте от числа слагаемых пикселов при постоянной угловой разрешающей способности телекамеры

систем с переменной по полю чёткостью. В период электронного телевидения такой режим удавалось практически реализовать только при группировании сигналов соседних элементов разложения во внешнем по отношению к телекамере устройстве цифровой обработки изображений.

В твёрдотельном телевидении эта идея не одинаковой по всему полю зрения чёткости стала реализуемой, а её практическая потребность обосновывалась наличием априорной информации о неоднородной смысловой структуре изображения, желанием повышения помехоустойчивости и быстродействия телевизионной системы. В частности, чёткость может быть сокращена во всех участках кадра, кроме доминанты изображения («зон особого внимания» – «*region of interest*» [164], [165]), определяемой целью системы. Разнообразие способов опроса зон накопления в КМОП-матрице возводит этап развёртки изображений в видеосистемах на кристалле в ранг микропрограммирования эксперимента. Наличие таких зон характерно для телевизионных охранных систем, в которых пользователь настраивает систему на максимальную чувствительность (в том числе к движению объектов) в некотором множестве зон, но не во всём кадре. Переменная по полю чёткость актуальна не только в охранных системах, но и в системах наведения на движущиеся удалённые объекты, используемых в космонавтике и оборонных системах.

Для обеспечения малого текущего промаха (показатель *качества информации*) при наведении на подвижный объект необходимо быстро отслеживать малые флуктуации его траектории. Однако при значительных ускорениях наблюдаемых объектов текущий промах характеризуется двумя значениями: средней ошибкой в режиме сопровождения движущегося объекта и вероятностью срыва слежения при манёвре объекта слежения. В идеале обе эти характеристики должны быть как можно меньше. Малое среднее значение промаха позволяет достичь *цели системы* при меньших энергетических затратах. Малая вероятность срыва сопровождения позволяет повысить вероятность выполнения системой своей цели. Для уменьшения среднего значения промаха требуются высокая угловая разрешающая способность и высокая кадровая частота, а для уменьшения вероятности срыва сопровождения – высокая кадровая частота и большое поле зрения. Одновременно достичь в системах наведения высокую угловую разрешающую способность, большое поле зрения и высокую кадровую частоту затруднительно, а главное, и не нужно, этого нет в цели системы. Кроме того, следует считаться с тем, что чем больше считывается элементов с

матричного фотоприёмника, тем шире полоса частот и выше уровень шума, увеличивающий и средний промах, и вероятность срыва слежения.

Решение проблемы наведения визирной оси телевизионной системы на подвижный объект достигается использованием концепции переменных параметров системы в зависимости от этапов наведения, которое разделяется на грубое наведение (в широком поле зрения) и точное наведение (в узком поле зрения). Ключевыми пунктами в такой перестройке является изменение структуры считываемого сигнала в зависимости от значения текущего промаха. При малом промахе сигнал должен формироваться таким образом, чтобы минимизировать шумовую ошибку измерения координат объекта. При этом можно пожертвовать качеством изображения вне сопровождаемого объекта. При большом промахе, возникающем при манёвре объекта, нужно максимизировать отношение сигнала объекта к шуму, а разрешающей способностью, т. е. точностью измерения его координат, можно пренебречь. В результате телекамера может обладать наилучшими показателями качества и по точности измерения координат объекта в режиме сопровождения, и по вероятности срыва сопровождения при его манёвре. При этом в системах слежения и наведения доминантных зон может быть несколько; они могут организовываться как следящие окна с малой площадью зоны накопления (высокой чёткости) на общем поле фотоприёмной матрицы с увеличенной площадью зоны накопления (зоны с сокращённой чёткостью).

Концепция переменной чёткости реализуется в матричных фотоприёмниках с зарядовой связью и с координатной адресацией с помощью различных алгоритмов, так как принципиальное отличие в способах развёртки изображения в них ведёт к различию в возможных методах обработки сигнала непосредственно в матрице.

Сложность достижения малого значения текущего промаха при использовании матричных ПЗС состоит в том, что естественная задержка на накопление фотонов (равная в среднем половине времени кадра T_k) усугубляется дополнительной задержкой на считывание накопленного зарядового изображения объекта. В результате в матрице ПЗС суммарная задержка (при равенстве времени накопления и времени считывания) зависит от координаты элемента разложения в кадре и изменяется от $0,5T_k$ для самого близкого к считывающему узлу элемента до $1,5T_k$ для самого удалённого от считывающего узла элемента.

Реализация переменной чёткости в телекамерах на КМОП по сравнению с телекамерами на ПЗС имеет две важные особенности. Во-первых, в однокристалльную КМОП-телекамеру может быть заложен режим управления скоростью считывания в выбранном фрагменте, благодаря которому в нём может быть существенно повышена кадровая частота по сравнению с остальным полем зрения. Во-вторых, в КМОП-матрицах возможна реализация всевозможных видов развёртки, открывающих путь существенного упрощения операции увеличения площади зоны накопления, необходимой для обмена разрешающей способности системы на чувствительность.

В частности, для дополнительной пространственной обработки, например усреднения нескольких элементов изображения, при традиционной прогрессивной развёртке требуются дополнительные элементы аналоговой или цифровой памяти. Так называемая развёртка Z-сканирования (рис. 4.10, [164], [165]) позволяет проводить пространственное усреднение групп элементов 2×2 , 4×4 , 8×8 без применения дополнительных запоминающих устройств. Такая Z-развёртка в КМОП-телекамерах, невозможная в приборах с зарядовой связью,

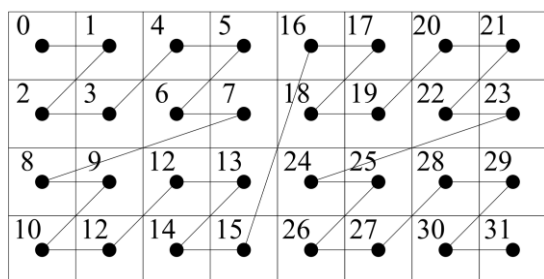
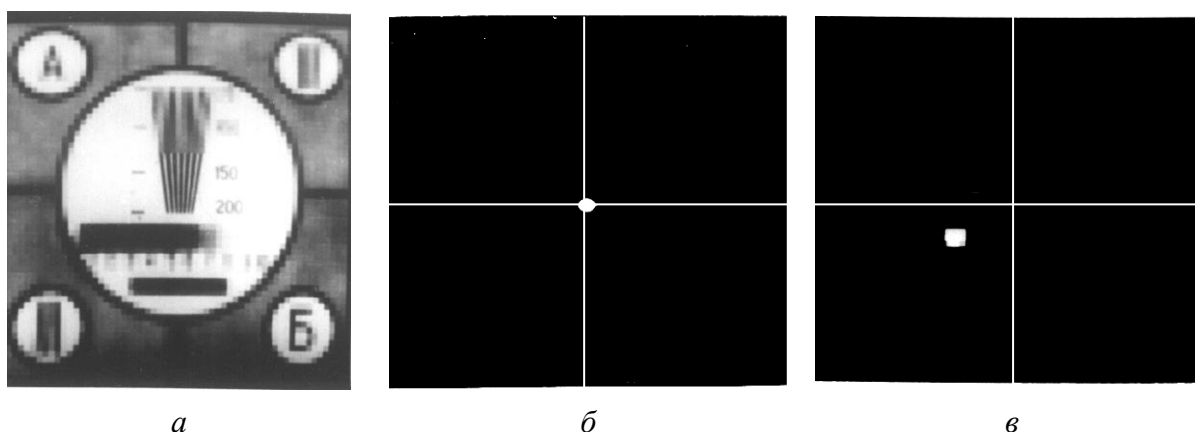


Рис. 4.10. Траектория Z-развёртки

делает достаточно просто реализуемыми *множественные выходы* телекамеры, отличающиеся различной (от «страницы к странице») чёткостью, требующейся при анализе сложных сюжетов.



а

б

в

Рис. 4.11. Иллюстрация метода переменной по полю чёткости:

- а – изображение испытательной таблицы, полученное при полной чёткости в центре поля зрения и сокращённой в 4 раза чёткости по вертикали и горизонтали;
- б – изображение удалённого объекта в центре поля зрения с предельной чёткостью;
- в – изображение удалённого объекта вне центральной зоны при увеличенной площади зоны накопления

Выделение доминантной зоны высокой чёткости в изображении, формируемом твёрдотельной телекамерой (рис. 4.11), может быть дополнено уменьшением скорости считывания выделенного фрагмента для уменьшения шума считывания [2]. Первая в мире экспериментальная аппаратура с переменной по полю чёткостью, реализованной непосредственно в матрице ПЗС, была создана во ВНИИ телевидения [2], [166] и использовалась в ходе исследований по созданию космических телевизионных систем наведения.

В этой системе в соответствии с различием оптимальных зон накопления при обнаружении и измерении координат в периферийной зоне кадра чёткость сокращалась в 4 раза по сравнению с центральной (фовеальной) зоной изображения (см. рис. 4.11).

За рубежом этот результат был повторен позже [164], [165], [167] и получил там официальное наименование *foveal vision systems*, т. е. систем со структурой сетчатки глаза. К сожалению, российские специалисты не всегда знают об отечественном приоритете в данной области твёрдотельного телевидения [168].

4.6. Оптимизация спектральной характеристики телекамеры

Накопление качества информации от падающего потока фотонов определяется интегрированием по выбираемым областям всех четырёх аргументов – двух пространственных, времени и длины волны. В этом параграфе рассмотрим примеры оптимизации спектральной характеристики телекамеры в предположении о фиксированных (оптимальных) параметрах зон накопления по остальным аргументам.

Задаче оптимизации характеристики спектральной чувствительности фотоприёмника посвящено много публикаций, они опираются на модель наблюдения соседствующих на изображении доминантного объекта и фона, различающихся спектральными характеристиками излучения (апликативная смесь сигнала и фона). Целью системы является принятие решения о наличии сигнала от объекта со спектром $S_o(\lambda)$ при наличии фона со спектром $S_f(\lambda)$, или различение сигналов доминантного объекта и фона. Качество информации определяется средним риском принятия неверного решения – отнесения конкретной пространственно-временной зоны накопления не к истинной категории (принятия решения о наличии объекта при наличии только фона и пропуске объекта). Максимизация качества инфор-

мации в этой постановке сводится к максимизации отношения сигнал/шум в разностном сигнале между объектом и фоном. При этом в случае малого контраста объекта относительно фона решение задачи упрощается в силу близости функций распределения фотонного шума по обеим гипотезам.

В пионерской работе [169] показано, что при известных спектральных зависимостях светового потока от доминантного объекта $S_o(\lambda)$ и фона $S_\phi(\lambda)$ характеристика пропускания оптимального светофильтра K_0 будет иметь прямоугольный вид. В [170] предложены пути получения значений длин волн, определяющих границы оптимальных прямоугольных светофильтров. В [45] данный вопрос получил развитие, введены понятия дифференциального и интегрального контрастов, описывающих отношения как спектральных плотностей световых потоков фона и объекта, так и интегралов по этим плотностям. Связь интегрального контраста с отношением сигнал/шум в перепаде между фоном и объектом при доминирующей роли фотонного шума подчёркнута в [170]. Решения задач оптимизации спектральных характеристик спектрально-зональной системы, в том числе для обнаружения точечных объектов на изменяющемся мешающем фоне и при априорной неопределённости спектральных характеристик объекта и фона обобщены в [170]. Оптимизация спектральных диапазонов при учёте всех компонентов шума – собственных шумов фотоприёмника, фотонных шумов фона и полезного сигнала – рассмотрена в [171].

Краткая сводка результатов для модели с преобладающим влиянием фотонного шума сводится к следующему. В выбранной пространственно-временной зоне накопления от сигнала и фона накапливаются фотоэлектроны в количестве

$$N_o = \int N_o(\lambda) K_0(\lambda) d\lambda; \quad N_\phi = \int N_\phi(\lambda) K_0(\lambda) d\lambda.$$

В первом приближении (при не слишком большой разнице между числом накопленных фотонов по каждой гипотезе) в качестве оценки дисперсии шума можно воспользоваться полусуммой дисперсий шума по каждой гипотезе. Так как распределение плотности потока фотонов является пуассоновским, со средним значением, равным дисперсии, то отношение сигнал/шум в перепаде между объектом и фоном составит

$$\Psi^2 = \frac{2(N_1 - N_2)^2}{(N_1 + N_2)}.$$

Помехоустойчивость обнаружения объектов можно связать не только с отношением сигнал/шум, но и с контрастом сигналов, определяемым как нормированная разность числа накопленных по двум гипотезам фотонов:

$$K = \frac{N_1 - N_2}{N_1 + N_2}.$$

Отношение сигнал/шум зависит от контраста и среднего по обеим гипотезам числа накопленных фотонов $\bar{N} = (N_1 + N_2)/2$:

$$\Psi^2 = 4K^2\bar{N}; \Psi = 2K\sqrt{\bar{N}}.$$

Спектральная характеристика $K_0(\lambda)$ оптимального фильтра, обеспечивающего максимум отношения сигнал/шум для заданных спектральных плотностей, имеет дискретный вид: она равна своему максимальному значению на тех длинах волн, где спектральный контраст превышает некоторый порог, и равна нулю на прочих длинах волн (рис. 4.12).

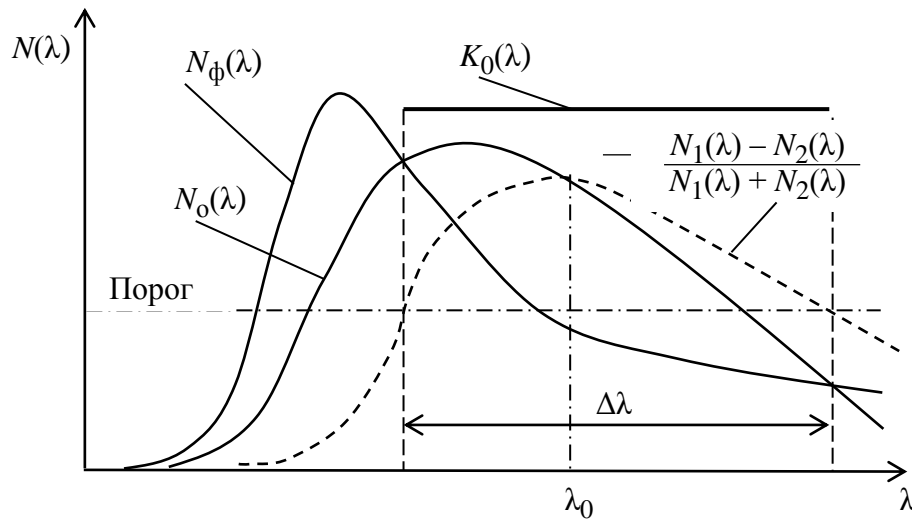


Рис. 4.12. Спектральные характеристики плотности потока фотонов объекта $N_1(\lambda)$, фона $N_2(\lambda)$ и оптимального фильтра $K_0(\lambda)$

Для достижения максимума отношения сигнал/шум необходимо (но не достаточно) выбрать тот участок спектра, в котором расположена длина волны λ_0 , на которой контраст обнаруживаемого сигнала максимален.

Кроме контраста в отношении сигнал/шум входит ещё и среднее число накопленных фотонов \bar{N} . Если бы помехоустойчивость определялась только контрастом, то ширина полосы пропускания оптимального фильтра $\Delta\lambda$ равнялась бы нулю. Однако, хотя при удалении от точки λ_0 различие между сигналами уменьшается, продолжается увеличение отношения сиг-

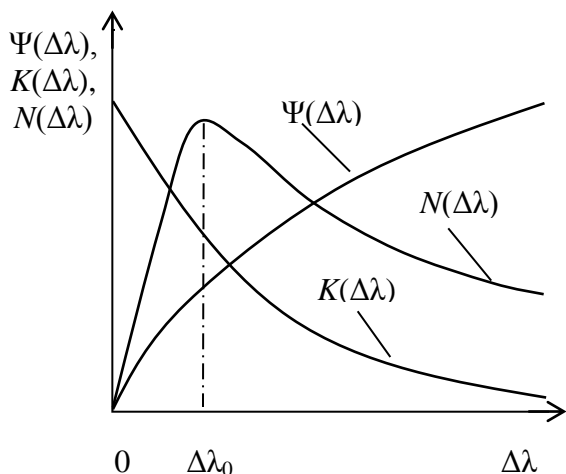


Рис. 4.13. Зависимость числа накопленных фотонов в перепаде между сигналом и фоном N , отношения сигнал/шум Ψ , контраста K от ширины полосы пропускания $\Delta\lambda$ оптического фильтра

нал/шум вплоть до его максимального значения при оптимальной ширине полосы пропускания фильтра (рис. 4.13).

Априорная неопределённость в спектрах сигналов существенно усложняет задачу, в результате чего оптимальные фильтры теряют свойство разрывности [170], а реализуемые спектральные фильтры с конечной крутизной среза спектральной характеристики становятся не грубым приближением к оптимуму, а им самим.

Практическое применение оптимизации спектральной характеристики телекамеры иллюстрируется следующими примерами.

Пригоризонтное обнаружение удалённых объектов. Удалённые на большие расстояния от телевизионной системы объекты на изображении теряют детали и контраст, что приводит к изменению их формы и затрудняет процесс обнаружения и опознавания. Данное явление связано с механизмами рассеяния света в атмосфере, и при решении задач пригоризонтного наблюдения объектов это необходимо учитывать.

Существует много факторов, ослабляющих излучение объектов при прохождении через земную атмосферу. Можно выделить два принципиально различающихся типа. Первый – рассеяние света, при котором частота пришедшего кванта не изменяется, но меняется его направление. В этом случае часть света от объекта приходит к наблюдателю в результате многократных фактов переотражения. Рассеяние делят на молекулярное (рэлеевское) и рассеяние на аэрозолях – более крупных, чем молекулы, частицах, находящихся в атмосфере. При ведении наблюдения, например, за морской обстановкой и распознавании пригоризонтных объектов на фоне неба очень часто рассеяние света обусловлено присутствием тумана, который представляет собой скопление мелких частиц воды. Второй тип – истинное поглощение. В этом случае квант излучения поглощается молекулой, а затем энергия этого кванта может быть излучена и на других частотах, и в других направлениях.

Частичная компенсация фоновой помехи, вызванной рассеянием света преимущественно в коротковолновой части спектра (из-за чего небо имеет голубой цвет), возможна с помощью подавления этой коротковолновой части спектра. Спектр объекта оценивается по его огибающей, соответствующей спектральной характеристике излучения Солнца, аппроксимируемой спектральной функцией излучения абсолютно чёрного тела по формуле Планка для температуры 6000 К с максимумом на длине волны 0,55 мкм (рис. 4.14).

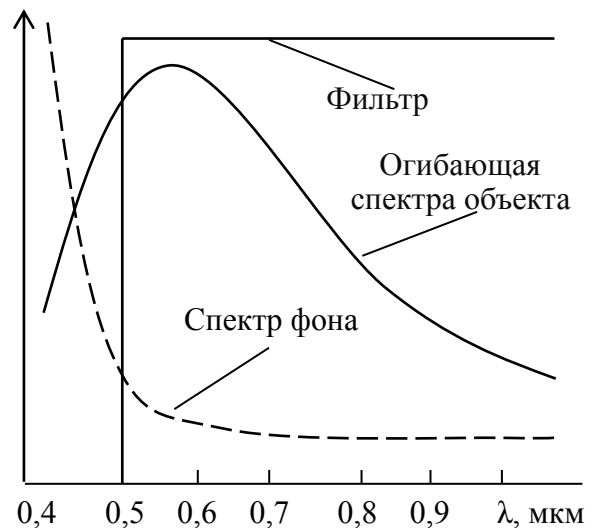


Рис. 4.14. Спектральные характеристики объекта и фона (плотности мощности) и оптимального светофильтра (коэффициент передачи) при атмосферном рассеянии света

Можно экспериментально найти оптимальные параметры спектральных светофильтров, выбираемые из стандартного набора и обеспечивающие максимум отношения сигнал / шум в перепаде между объектом и фоном в зависимости от метеорологической дальности видности. При этом чем меньше МДВ, тем большую часть спектра в коротковолновой области должен подавлять оптимальный спектральный фильтр.

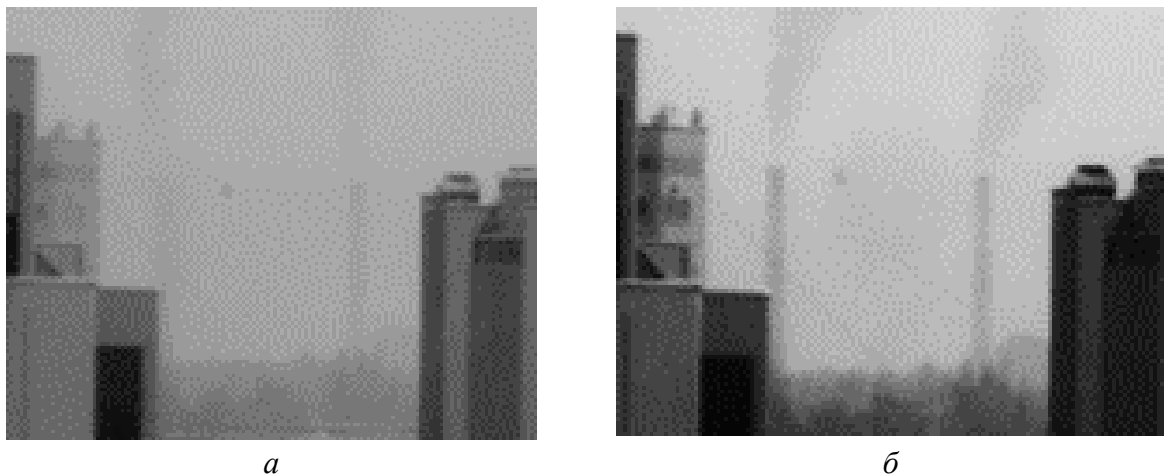


Рис. 4.15. Оптимизация спектральной характеристики телекамеры: а – без светофильтра, б – с использованием красного светофильтра КС-16 контраст удалённых объектов увеличен

Эксперимент [163] (рис. 4.15, 4.16) при МДВ 6 км с использованием телекамеры на КМОП-матрице с размером пиксела $3,5 \times 3,5$ мкм² (размер матрицы 3002×2210 элементов, объектив Pentax С7528-М с фокусным расстоя-

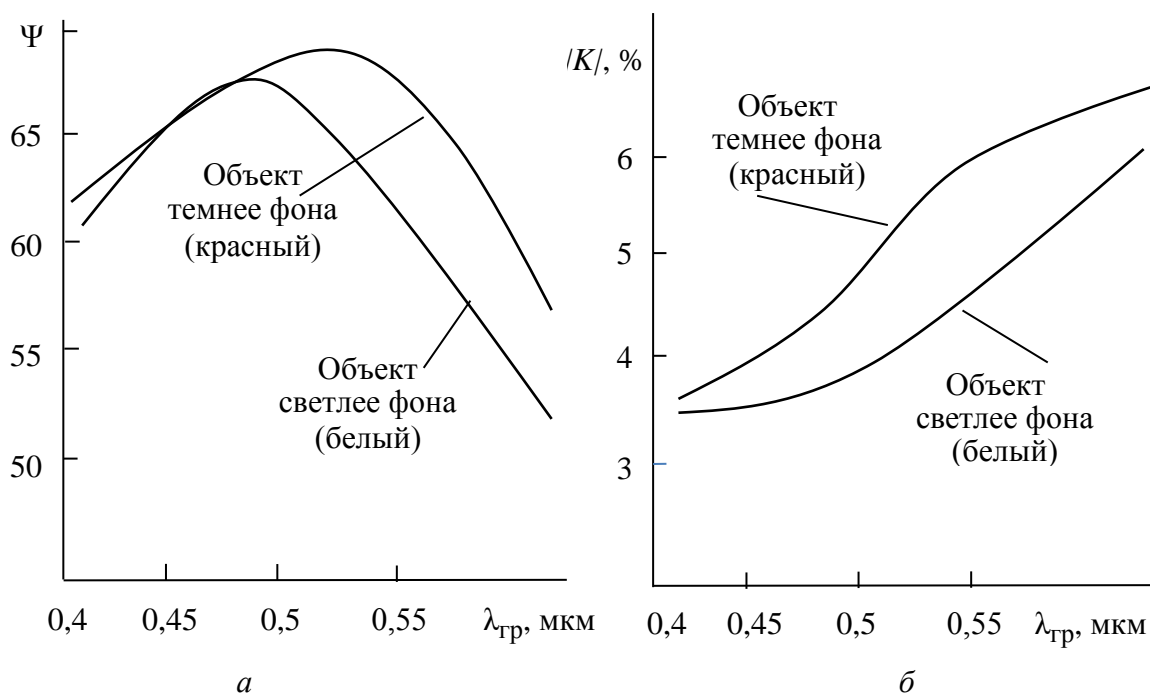


Рис. 4.16. Экспериментальная зависимость: *a* – отношения сигнал/шум; *б* – контраста в тестовых объектах от «синей границы» спектрального светофильтра (тестовый объект – тот же, что и на рис. 4.8)

нием 75 мм) показывает, что отношение сигнал/шум в перепаде между удалённым объектом и фоном можно увеличить по сравнению с передачей всего спектрального диапазона чувствительности фотоприёмной матрицы.

Анализ результатов наблюдения показал [163], что по мере сокращения пропускаемых длин волн (смещения «синей границы» в длинноволновую область) контраст объекта монотонно возрастает, а отношение сигнала удалённого объекта к шуму (качество информации) имеет максимум, который и определяет оптимальный спектральный фильтр, длина волны среза («синяя граница») которого зависит от МДВ.

Обнаружение пламени. Эта задача спектральной селекции имеет много применений [172], [173]. Максимальная чувствительность обнаружения пламени на фоне освещённых Солнцем объектов достигается в среднем инфракрасном диапазоне (1,5...3 мкм) [172]. Вместе с тем вопрос выбора спектрального диапазона часто оказывается дискуссионным [174] вследствие существенного различия в стоимости телекамер инфракрасного диапазона и стандартных телекамер, фотоприёмники которых изготовлены по кремниевой технологии и имеют красную границу около 1,1 мкм. Не останавливаясь на возможностях обнаружения пламени по его динамическим характеристикам, проиллюстрируем возможность повышения качества информации о доминантном объекте с помощью спектральной фильтрации.

Компенсация фоновой помехи, вызванной отражением солнечного света (эта же кривая спектральной плотности излучения на рис. 4.14 характеризует доминантный объект), возможна с помощью подавления коротковолновой части спектра, где спектр пламени оценивается по его огибающей, аппроксимируемой фрагментом спектральной функции излучения абсолютно чёрного тела по формуле Планка для типовой температуры пламени 1500 К [172], [173] (рис. 4.17).

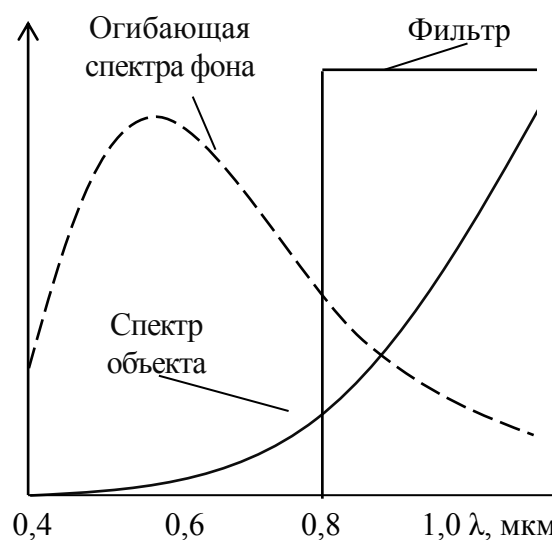


Рис. 4.17. Спектральные характеристики объекта и фона (плотности мощности) и оптимального светофильтра (коэффициент передачи) при обнаружении пламени

Эксперимент [175] (рис. 4.18) показал, что из стекла наилучший результат в смысле отношения полезного сигнала к сумме сигналов фона и шума даёт совместное использование светофильтров КС-18 и КС-19. В этом случае фон удаётся практически полностью отсечь, при этом изображение пламени практически не подавляется.



Рис. 4.18. Изображения пламени на фоне дневного леса, полученные с использованием светофильтров: а – СЗС-17 с синей границей 0,5 мкм; б – КС-19 с синей границей 0,75 мкм

Эксперимент [175] показал наличие ряда компонентов линейчатых спектров в пламени, по которым с помощью узкозонных интерференционных фильтров может идентифицироваться горящий материал. Проиллюстрированные методы спектральной селекции объектов далеко не исчерпывают её возможности. Так, широко применяются спектрально-зональные методы в геологии, землепользовании, контроле ледовой обстановки и т. д.

Выводы по гл. 4. Опираясь на «указание» С. И. Катаева, мы показали, как в различных задачах в твёрдотельных телекамерах по-разному решается проблема оптимизации накопления качества информации. При этом важной идеей является отказ от принципа декомпозиции и одновременная оптимизация всех ключевых узлов телекамеры – объектива, фотоприёмной матрицы и вычислителя параметров сигналов. Наиболее ярко принцип поиска глобального оптимума проявляется при реализации телекамер на базе единой СБИС – видеосистемы на кристалле.

Так, проблема согласования размера пиксела с функцией рассеяния точки имеет различные решения при наблюдении случайных полей (для поиска объектов при малой априорной информации о них), обнаружении точечных объектов и измерении их координат.

Группирование зарядовых пакетов в поисках оптимальных пределов интегрирования в (3.1) может осуществляться либо по всему полю – для повышения как абсолютной чувствительности, так и контрастной чувствительности, либо с переменной чёткостью в зависимости от интенсивности обнаруживаемых сигналов при совместном обнаружении и оценивании параметров сигналов, либо с переменной чёткостью для выделенной «зоны интереса», т. е. для доминантного сигнала и окружающего его фона. При этом группирование сигналов пикселей для повышения контрастной чувствительности для сохранения угла зрения телекамеры требует применения сборки матриц и является временной мерой повышения качества видеoinформации (предельная контрастная чувствительность должна реализовываться с помощью цифровых пикселей высокой разрядности, рассмотренных в гл. 3).

Оптимизация спектральных характеристик телекамеры определяется целью системы и априорной информацией о спектральных характеристиках объекта, фона и среды распространения света. При этом главным критерием качества видеoinформации становится отношение сигнал/шум в перепаде объект–фон, максимум которого достигается при спектральных характеристиках, отличных от обеспечивающих максимальный контраст объектов относительно фона.

Таким образом, обозначенные в преамбуле к этой главе конфликты идей в основном разрешены.

Глава 5. АДАПТАЦИЯ НАКОПЛЕНИЯ КАЧЕСТВА ИНФОРМАЦИИ В ТВЁРДОТЕЛЬНЫХ ТЕЛЕКАМЕРАХ

В этой главе мы рассмотрим проблемы, порождённые конфликтом следующих идей. Во-первых, это конфликт идеи о борьбе с априорной неопределённостью о статистиках сигналов с помощью адаптации параметров накопления по разным аргументам и идеи о том, что раз скорость считывания информации с матричного фотоприёмника ограничена, то перестройка параметров разложения по разным аргументам должна производиться совместно и компромиссно.

Во-вторых, это конфликт идеи о том, что из-за движения изображения возникающий смаз ведёт к потере разрешающей способности и поэтому является злом, и идеи о том, что при наличии движения нулевой смаз означает нулевое время накопления и, следовательно, нулевое отношение сигнал/шум.

Основными трудностями приложения теории оптимальных систем приёма и передачи информации являются априорная неопределённость параметров доминантных, фоновых и шумовых сигналов и их нестационарность. Известно, что *«первая и наиболее очевидная трудность заключается в том, что статистика источника может быть неизвестной. В системах связи можно предполагать, что она поддаётся определению путём анализа используемого языка или кода, но в системах наблюдения это не всегда возможно»* [49]. При этом считалось, что априорная трудность является фундаментальным препятствием на пути создания «полностью удовлетворительной теории приёма». Для преодоления этой априорной трудности применяют несколько подходов, среди которых главные [176], [177]: минимакс, инвариантность к изменению статистик сигналов и адаптация. При цифровом кодировании телевизионных сигналов, сформированных телекамерой, априорная трудность усугубляется их нестационарностью [178], методы учёта которой различны для пространственной и временной координат. Пространственная нестационарность обычно учитывается выбором постоянного размера кодируемых фрагментов (устойчивые алгоритмы), нестационарность во времени часто ведёт к использованию обнаружителей смены сюжета (адаптивные алгоритмы). Применяемые методы преодоления указанных двух трудностей показывают важную роль адаптации при синтезе телевизионных кодирующих устройств. В системах реального времени адаптация телекамер ещё более важна, чем при

цифровом кодировании. Адаптация телекамер к изменениям освещённости достаточно подробно освещена в [1], [3], [108], [111], поэтому далее рассматриваются вопросы адаптации параметров телекамер к *динамическим* характеристикам сюжета при различных моделях пространственно-временных характеристик сигналов и назначениях систем. Важной проблемой видеоинформатики является получение максимума качественной информации при наличии движения изображения. При этом отношение к различным видам движения объекта и наблюдателя изменялось в ходе развития *технологий и концепций*. Можно выделить несколько характерных (перекрывающихся во времени) концепций учёта динамики сюжета:

- учёт семантической смены сюжета, так называемая концепция малокадрового телевидения, предложенная С. И. Катаевым и вошедшая в практику космического телевидения [45], позже получившая развитие в виде адаптации кадровой частоты к динамике сюжета для максимизации количества информации в кадре [2], [98];

- учёт влияния подвижности изображения на разрешающую способность телевизионной системы [38] (в том числе при наблюдении Земли из космоса [45], [179]), позволивший формализовать требования к стабилизаторам изображений, в частности, в современных цифровых фотоаппаратах, где реализуются косвенные методы оценки возмущения визирной оси с помощью датчиков ускорений [180];

- обработка изображений для компенсации смаза [47], [104], выдвинувшая требования к программно-аппаратным средствам обработки видеосигнала в реальном времени;

- электронная стабилизация изображения для облегчения работы оператора, примиряющаяся с потерей информации из-за смаза и использующая прямой метод оценки возмущений, но эффективная при медленных возмущениях [181];

- адаптивная оптика, парирующая возмущения турбулентной атмосферы, имеющая множество подвижных сегментов зеркала в силу локальных различий искажения волнового фронта [182];

- электромеханическая стабилизация изображения в астрономии, использующая прямой метод оценки возмущений и работающая при медленных возмущениях по ограниченному классу изображений при наличии в поле зрения опорной звезды с яркостью, на порядки большей, чем у исследуемой в основном поле [183];

- электромеханическая стабилизация изображения, использующая прямой метод оценки возмущений и работающая при быстрых возмущениях по широкому классу изображений. Необходимый запас энергии сигнала достигается не за счёт яркости объектов, а за счёт площади вспомогательного изображения [184]–[187].

В рамках старой парадигмы [40] движение изображения расценивалось лишь как мешающий фактор, ведущий к потере разрешающей способности телевизионной системы. *Новая парадигма накопления качества информации ведёт к поиску оптимальных структуры и параметров камеры, которые обеспечивают максимум качества информации.*

Пространственно-временная дискретизация изображения – главный инструмент для достижения потенциального качества передачи телевизионных сигналов и измерения их параметров. В отличие от вещания в научно-прикладном телевидении вследствие меньшей априорной неопределённости передаваемых сигналов такая оптимизация получает аналитическое решение. В литературе часто рассматривается задача оптимизации либо пространственной, либо временной дискретизации, но в общем случае разделить эти процессы нельзя. Далее рассматриваются варианты пространственно-временной дискретизации для различных моделей сигнала: случайной по полю и по времени, случайной по времени (характерна, например, при изменении в широких пределах дальности до наблюдаемого подвижного объекта) и известной по полю (последняя характерна для гартмановских анализаторов динамических искажений волнового фронта [182]).

5.1. Адаптация к динамике сюжета: обмен чёткости и кадровой частоты

На практике адаптация телекамер к динамике сюжета востребована, например, в процессе контроля стыковки космических аппаратов. Адаптация телекамеры призвана дополнить итерационный принцип контроля стыковки, сформулированный С. П. Королёвым, включив в него не только последовательное применение различных радиотехнических навигационных приборов, но и перестройку параметров телекамер.

Кодирование оптического стимула с преобразованием его в электрический сигнал в телекамере при изменениях пространственно-временных статистик может оптимизироваться для различных моделей сигнала – в виде случайного поля (задаваемого совокупностью корреляционных функций

по полю и по времени) и в виде сигнала от объекта известной формы (с неопределённостью размеров, ракурса, ориентации, смещения).

Согласно принципу доминантной информации, рассмотренному в гл. 2, *цель* телевизионной системы в таком применении состоит в минимизации промаха активного космического аппарата относительно узла стыковки; *доминантной информацией* являются пространственно-временные координаты космического аппарата и его стыковочного узла; *качество информации*, понимаемое как насыщенность доминантой передаваемой информации, определяется отношением количества информации о доминанте к общему количеству информации, передаваемому телекамерой.

Повышение точности измерения пространственно-временных координат в таких системах, характеризующихся очень широким диапазоном *видимой* скорости объекта, может быть достигнуто с помощью перестройки параметров разложения изображения таким образом, что на больших дальностях при малой видимой скорости объекта предпочтение отдаётся высокому пространственному разрешению, на малых дальностях при большой видимой скорости объекта возможно повышение кадровой частоты в ущерб уже не столь актуальному пространственному разрешению. Эта идея, реализованная лишь недавно благодаря твёрдотельной революции в телевидении [188], [189], концептуально была понятна полвека назад [44]: *«В светоинформационной системе может быть проведено чёткое разделение информации, характеризующей распределение света вдоль оси x, от информации, характеризующей распределение вдоль оси y, пространственной информации от временной и в некоторых случаях от спектральной. Возможность обмена различных компонентов информации – одно из средств создания оптимальной системы, передающей наибольшее количество ценной информации, необходимой для достижения поставленной цели»*.

Уменьшение количества передаваемой информации для повышения качества информации, как и во многих других приложениях, предполагает реализацию группового счёта фотоэлектронов, при котором размер групп (шаг квантования в АЦП) определяется согласно критерию минимума информационного риска, включающего потерю доминантной информации из-за линейных ошибок пространственно-временной дискретизации и шумовую информацию из-за входного фотонного шума.

Адаптация пространственной и временной дискретизации к динамике сюжета (управление кадровой частотой и зоной накопления с помощью

сложения сигналов пикселей (так называемого «бининга» [1], [3]) должна осуществляться **взаимосвязанно** [188], [189]. Эта взаимосвязь определяется тем, что все изображения являются пространственно-временными процессами, дискретизируемыми фотоприемником на N^2 пикселей и на временные кадровые интервалы T_K **при ограниченной частоте считывания**. С учётом общепринятой оценки скорости считывания изображения по полосе частот видеосигнала в соответствии с теоремой отсчётов В. А. Котельникова ($C = N^2/2T_K$) *критерий качества* может быть записан в виде *условного минимума* среднеквадратической ошибки при ограничении скорости C_0 считывания изображения с фотоприёмной матрицы:

$$\varepsilon = \min_{N, T} \quad \text{при} \quad C = \frac{N^2}{2T_K} \leq C_0. \quad (5.1)$$

Формализация взаимосвязи накопления по полю и по времени возможна при моделировании сигналов как случайными полями, так и неслучайными движущимися объектами с неизвестными пространственно-временными параметрами. При моделировании изображений случайными процессами традиционно считают известными (оцененными) автокорреляционные функции по полю $K_x(x) = K_y(y)$ (изотропное изображение) и по времени $K_t(t)$. Дисперсии приращений сигнала на интервалах пикселя и кадра составляют [69]:

$$D_{\Delta x} = 2K_x(0)[1 - K_x(\Delta x)/K_x(0)]; \quad D_{\Delta t} = 2K_t(0)[1 - K_t(\Delta t)/K_t(0)]. \quad (5.2)$$

Примем для конкретности модель сигнала с экспоненциальными автокорреляционными функциями: $K(x) = K_x(0) \exp(-|x|/\tau_x)$ с интервалом корреляции τ_x и $K(t) = K_t(0) \exp(-|t|/\tau_t)$ с интервалом корреляции τ_t . Эта модель соответствует большинству изображений, к тому же сигнал с такими свойствами обладает максимальной энтропией: его энергетический спектр убывает наиболее медленно – пропорционально квадрату частоты (при сколь угодно меньших степенях интеграл от спектра расходится и мощность сигнала становится бесконечной).

При независимости случайных процессов по разным аргументам дисперсии их приращений складываются, и при не слишком больших интервалах дискретизации ошибка передачи выражается через дисперсии межэлементной $D_{\Delta x}$ и межкадровой $D_{\Delta t}$ разностей (5.2):

$$\varepsilon = 2D_{\Delta x} + D_{\Delta t} = 4K_x(0)\Delta x/\tau_x + 2K_t(0)\Delta t/\tau_t. \quad (5.3)$$

Линейный размер пиксела определяется линейным размером фотоприёмной матрицы l : $\Delta x = l/N$, и показатель качества изображения (интервал дискретизации по времени равен времени кадра $\Delta t = T_k$) примет вид

$$\varepsilon = 4K_x(0)l/N\tau_x + K_t(0)N^2/C_0\tau_t. \quad (5.4)$$

Взяв производную от суммарной ошибки (5.4) по числу N пикселей в строке и приравняв её нулю, найдём оптимальные значения полного числа пикселей N^2 и времени кадра [188]:

$$N_{\text{opt}}^2 = \left(2lC_0 \frac{K_x(0)\tau_t}{K_t(0)\tau_x} \right)^{2/3}; \quad T_{k\text{opt}} = \frac{1}{(2C_0)^{1/3}} \left(l \frac{K_x(0)\tau_t}{K_t(0)\tau_x} \right)^{2/3}. \quad (5.5)$$

Отсюда следует, что $N_{\text{opt}}^2/2T_{k\text{opt}} = C_0D_{\Delta x}/D_{\Delta t}$. Это обусловлено правилом адаптации, требующим обеспечения равенства дисперсий межэлементной и межкадровой разностей.

Аналогичное решение для оптимального взаимобмена чёткости и кадровой частоты при ограниченной скорости считывания C_0 можно получить при решении задачи измерения пространственно-временных параметров движущегося объекта. В этой модели полная ошибка также содержит две составляющие: статическую $\varepsilon_{\text{ст}}$ с дисперсией $\varepsilon_{\text{ст}} = 2l^2/N^2$ и динамическую $\varepsilon_{\text{дин}}$ (которая определяется дальностью L до объекта, фокусным расстоянием f объектива и модулем вектора скорости v объекта) с дисперсией $\varepsilon_{\text{дин}} = v^2T_k^2f^2/L^2$. Принятый критерий минимума суммы ошибок по трём аргументам приводит к аналогичным по структуре формулам [188]:

$$N_{\text{opt}}^2 = \left(\frac{2C_0lL}{vf} \right)^{2/3}, \quad T_{k\text{opt}} = \frac{1}{(2C_0)^{1/3}} \left(\frac{lL}{vf} \right)^{1/3}. \quad (5.6)$$

Полученные формулы *предписывают адаптацию параметров накопления к дальности до объекта и его скорости*: по мере уменьшения дальности и увеличения скорости количество элементов убывает, а кадровая частота возрастает. Важно, что для обеих моделей сигнала – случайной и неслучайной – (5.5) и (5.6) ведут к *правилу равенства дисперсий межэлементной и межкадровой разностей*.

При достаточной освещённости сюжета (при работе системы автоматической регулировки времени накопления) скорость считывания должна быть близкой к максимально возможному значению C_0 . Если освещённость фотоприёмника мала $E < E_0$, то она может трактоваться как дополнительное ограничение, а показатель качества усложняется и включает не только ошибки дискретизации, но и шумовую ошибку, обусловленную дискретной структурой света. Приняв для определённости плотность вероятности освещённости равномерной, можно считать, что в силу распределения потока фотонов по закону Пуассона с учётом квантовой эффективности η фотоприёмника и количества n фотонов в единице освещённости ($1,3 \times 10^{12}$ фотонов/(см²·с) [17]) нормированная к освещённости *дисперсия ошибки* оценивания случайного поля изображения будет состоять из двух линейных (по полю и по времени) и шумовой составляющих:

$$E = 4K_x(0)l/N\tau_x + 2K_t(0)l/N\tau_t + N^2/2n\eta ET_k l^2. \quad (5.7)$$

Взяв производные от этой суммарной ошибки по числу N пикселей в строке и по времени кадра T_k и приравняв их нулю, найдём оптимальные значения полного числа пикселей и времени кадра [188]:

$$N_{\text{opt}}^2 = \frac{2l^2 K_x(0)}{\tau_x} \sqrt{\frac{n\eta E \tau_t}{K_t(0)}}; T_{\text{kopt}} = \left(\frac{\tau_t}{K_t(0)} \right)^{3/4} \left(\frac{K_x(0)}{2\tau_x} \right)^{1/2} \frac{1}{(n\eta E)^{1/4}}. \quad (5.8)$$

Отсюда следует, что количество элементов разложения и время кадра должны зависеть и от статистических свойств сигнала по полю и по времени, и от освещённости. Лишь при достаточно большой освещённости, превышающей критическую E_0 , вместо (5.8) используются расчётные формулы (5.5).

Однако главный вывод из (5.8) состоит в *правиле адаптации* телекамеры, аналогичном полученному для случая больших освещённостей: для достижения условного минимума ошибки должно быть обеспечено равенство дисперсий межэлементной и межкадровой разностей [188]:

$$D_{\Delta x} = D_{\Delta t} = \left(\frac{K_t(0)}{\tau_t n \eta E} \right)^{1/4} \left(\frac{K_x(0)}{2\tau_x} \right)^{1/2}. \quad (5.9)$$

При обеих моделях сигнала по измеренному отношению дисперсий межэлементной и межкадровой разностей $D_{\Delta x}$ и $D_{\Delta t}$ оценивается соотношение ошибок пространственной и временной дискретизации (рис. 5.1).

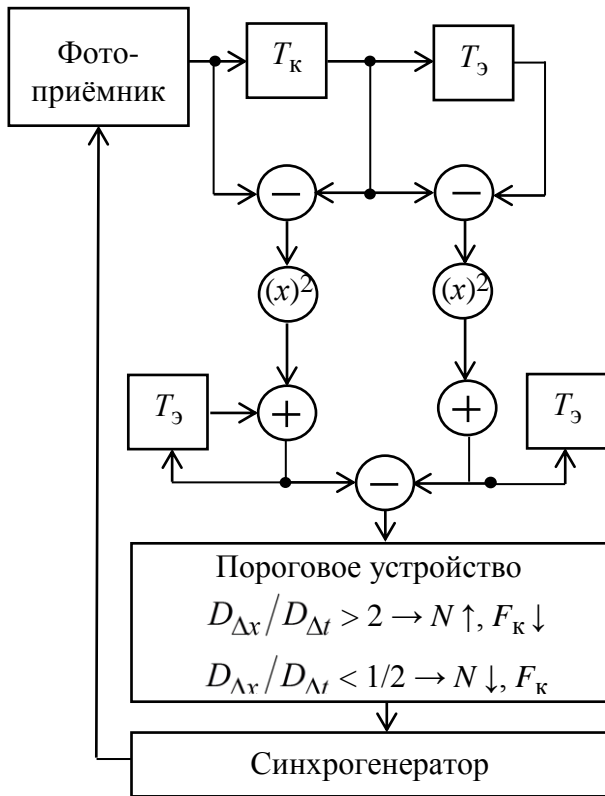


Рис. 5.1. Структурная схема адаптации телекамеры

управляемым числом складываемых сигналов пикселей (сначала по 2×2 , затем по 3×3 и, на последнем этапе, по 4×4), обеспечивая дискретный переход от одной пары «чёткость – кадровая частота» к другой. Существование *оптимального сочетания* чёткости и кадровой частоты, зависящего от скорости движения объекта (скорости изменения сюжета во времени) и обеспечивающего максимум качества пространственно-временной оценки изображения, подтверждается экспериментально. Для проверки эффективности предложенного метода управления накоплением был проведён эксперимент с подвижными объектами

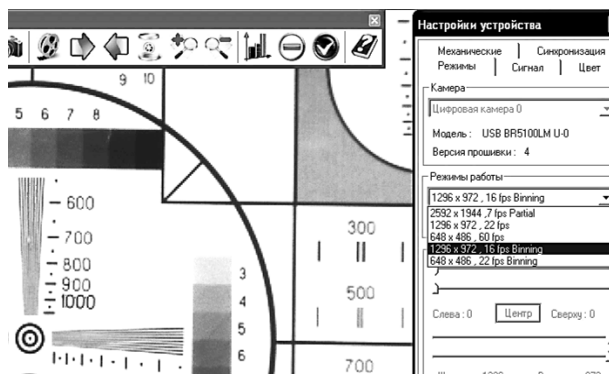


Рис. 5.2. Фрагмент изображения: движущаяся тест-таблица и панель управления параметрами разложения

Когда это отношение превышает порог γ (например при автоматической регулировке времени накопления применяют двукратное отношение), подаётся команда на изменение параметров разложения. При этом если усреднённая межкадровая разность оказывается больше усреднённой межэлементной, то время кадра уменьшается, а число элементов увеличивается. Если же межкадровая разность оказывается меньше усреднённой межэлементной, то время кадра увеличивается, а число элементов уменьшается. Практическое приложение полученных формул к адаптации накопления в телекамере опирается на известную операцию «бининга» с

подвижными объектами (в том числе с *подвижной тест-таблицей* 0249-10, рис. 5.2) с использованием телекамеры на матричном КМОП-фотоприёмнике, позволяющей реализовать три сочетания чёткости и кадровой частоты: 5 мегапикселей при кадровой частоте 7 Гц; 1,25 мегапикселя при кадровой частоте 16 Гц; 0,3 мегапикселя при кадровой частоте 22 Гц [188].

Эти сочетания не являются предельными, но для доступных фотоприёмных матриц в режиме постоянной чувствительности они близки к предельным. Эксперимент показал, что сумма статической и динамической ошибок (измеренных как дисперсии межэлементной и межкадровой разностей) увеличивается при увеличении скорости движения, измерявшейся в пересчёте на фокальную плоскость объектива.

Эксперимент (рис. 5.3) [188] подтвердил существование оптимального сочетания чёткости и кадровой частоты, зависящего от скорости движения объекта. Измерения разрешающей способности по подвижной тест-таблице, конечно, не противоречат старым представлениям – чем быстрее движется объект, тем меньше разрешающая способность телекамеры. Однако важнее то, что эксперимент по **совместной адаптации чёткости и кадровой частоты** подтвердил наличие **максимума качества информации**

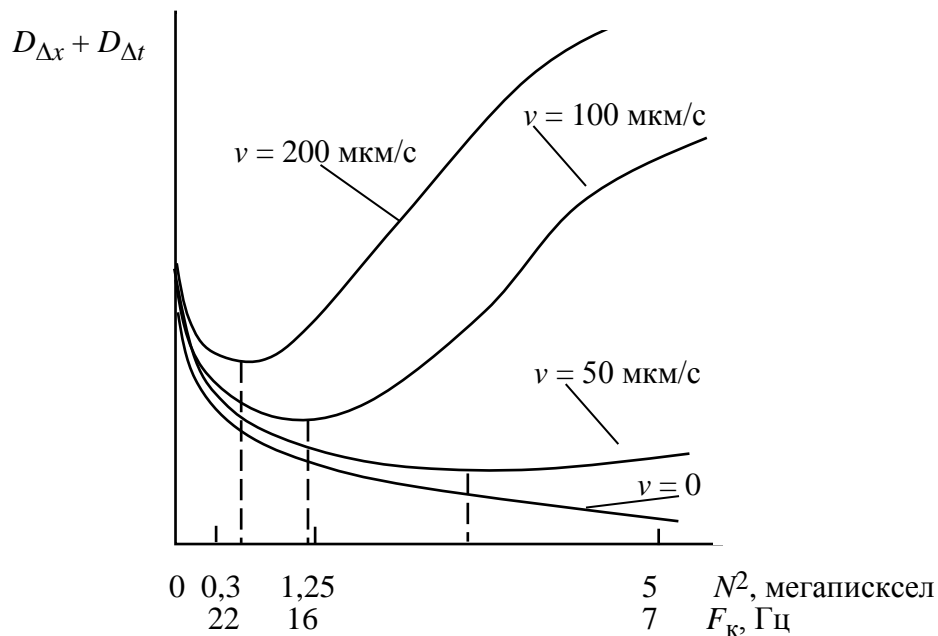


Рис. 5.3. Экспериментальная зависимость суммы дисперсий межэлементной $D_{\Delta x}$ и межкадровой $D_{\Delta t}$ разностей от эффективного числа элементов фотоприёмной матрицы для различных скоростей объекта

Показанный на рис. 5.3 минимум ошибок, т. е. минимум потери информации, обеспечивает максимум качества информации и отражает важный эффект *влияния твёрдотельной технологии на изменение методов системного проектирования* в видеоинформатике.

5.2. Адаптация к состоянию атмосферы

Кодирование оптического стимула при изменениях пространственно-временных статистик в ряде систем может опираться на модель сигнала от объекта известной формы и размеров, движущегося по *случайной траек-*

тории. При этом часто оптимизация пространственной дискретизации в телекамере должна осуществляться на основе согласования размеров пиксела фотоприёмника с кружком рассеяния объектива, а оптимизация временной дискретизации – согласованием кадровой частоты со скоростью движения проекции объекта на фотоприёмную матрицу.

На практике адаптация телекамер к динамике сюжета востребована, например, в системах анализа волнового фронта в астрономической адаптивной оптике, предназначенной для парирования динамических возмущений, возникающих в атмосфере. В результате работы адаптивной системы удаётся компенсировать основные искажения волнового фронта, приблизив функцию рассеяния точки телескопа к форме, определяемой дифракцией.

Эту весьма сложную междисциплинарную задачу каждый специалист излагает с использованием собственной терминологии. Например, физик может написать [5]: «В некоторых телескопах с помощью компьютеров постоянно меняется форма зеркала, чтобы компенсировать флуктуации земной атмосферы», мы же будем выделять в качестве ключевого звена не управляющий компьютер, а телекамеру, которая рождает сигнал, из которого извлекается информация об искажениях волнового фронта. Для скоростного измерения искажений волнового фронта используются различные методы, среди которых видное место занимает метод Гартмана или его современная модификация, называемая методом Гартмана–Шака. Этот метод состоит в дискретизации волнового фронта в пределах входного зрачка телескопа на ряд парциальных пучков с помощью линзовой маски, при плоском фронте формирующей в фокальной плоскости группу пятен (функций

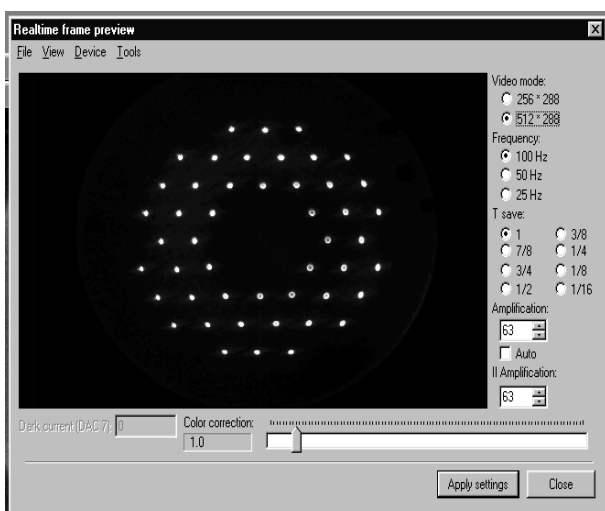


Рис. 5.4. Изображение теста Гартмана–Шака и панель управления параметрами разложения

рассеяния точки), повторяющую расположения линз в маске (рис. 5.4 [97], [109]). Об искажениях волнового фронта (локальных углах наклона) судят по относительному случайному смещению этих пятен относительно среднего положения, соответствующего идеально спокойной атмосфере. Случайное смещение измеряется в реальном времени в телевизионно-компьютерной системе, формиру-

ющей в каждом кадре группу разностей координат пятен с точностью до сотых долей элемента разложения.

Согласно принципу доминантной информации, рассмотренному в гл. 2, *цель* телевизионной системы в гартмановском датчике волнового фронта состоит в минимизации искажений волнового фронта; *доминантной информацией* в такой телевизионной системе являются пространственно-временные координаты специально сформированных точечных объектов – пятен гартманограммы; *качество информации*, понимаемое как насыщенность доминантой передаваемой информации, определяется отношением количества информации о доминанте к общему количеству информации, передаваемому телекамерой.

Интенсивность флуктуаций волнового фронта увеличивается по мере отклонения от зенита к горизонту, а их спектр (зависящий от конкретного состояния атмосферы) простирается до частот в сотни герц, хотя доминирующие возмущения (нравящееся поэтам и раздражающее астрономов мерцание звёзд) бывают сосредоточены в области частот до 10 Гц. Для компенсации искажений адаптивная оптическая система должна обладать широкой полосой частот, что требует от измерительной системы кадровой частоты, существенно превышающую используемую в телевизионном вещании, – до 200 [109] и даже до 5000 Гц [112]. Такая высокая частота обусловлена двумя причинами:

- высокой граничной частотой спектра флуктуаций волнового фронта;
- необходимостью экстраполяции состояния волнового фронта (экстраполяции положения пятен гартманограммы) с учётом задержки, возникающей в процессах накопления светового сигнала в фотоприёмнике, считывания зарядового изображения и передачи электрических и механических сигналов от телекамеры к управляемым сегментам зеркала (рис. 5.5).

Число парциальных пучков (число пятен динамического теста Гартмана–Шака) в адаптивной оптической системе и в системе анализа искажений волнового фронта обычно определяется типичным радиусом корреляции волнового фронта и бывает тем больше, чем больше диаметр сегментированного зеркала. Так, относительно небольшие адаптивные зеркала имеют семь сегментов, при диаметре зеркала 10 м число сегментов составляет 54, а в перспективных системах 25-метрового класса число сегментов превышает 600.

При анализе волнового фронта методом Гартмана–Шака (когда изображение одной звезды сначала преобразуется в параллельный световой пучок, а затем с помощью группы микролинз – в множество изображений точечных объектов, положение которых отражает искажения волнового фрон-

та) с применением телевизионных систем пространственно-временная дискретизация волнового фронта производится в два этапа.

Пространственная дискретизация на парциальные световые пучки осуществляется в оптическом тракте, временная – в телекамере. При этом подход к оптимизации кадровой частоты несколько различен в системах накопления статистик волнового фронта, где задержка не существенна, и в системах реального времени (рис. 5.6).

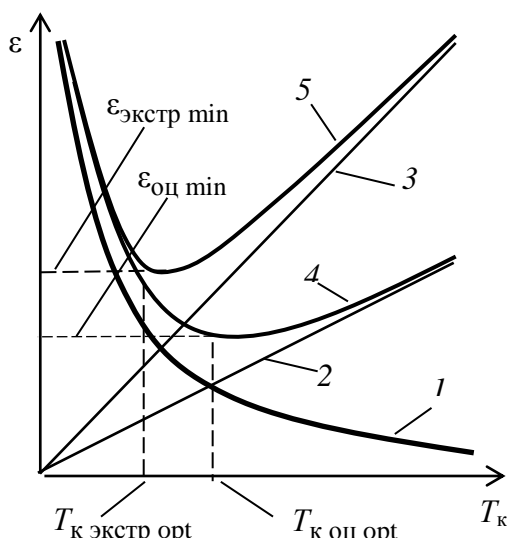


Рис. 5.5. Дисперсии ошибок: шумовой (1), линейной для оценивания с задержкой (2), линейной для экстраполяции координат (3), суммарной для оценивания (4) и суммарной для экстраполяции (5)

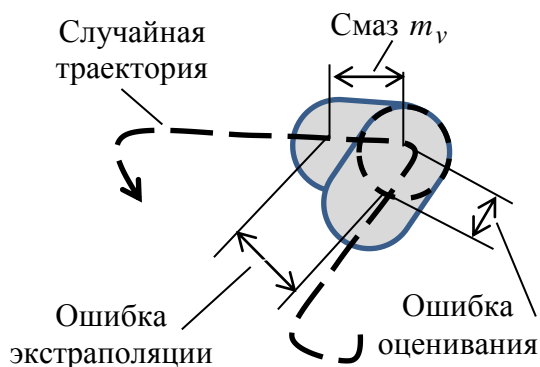


Рис. 5.6. Фрагмент случайной траектории объекта, смаз за время кадра и ошибки оценивания и экстраполяции координат объекта; пунктирная окружность с радиусом ρ – положение объекта в конце предшествующего и в начале последующего кадров

Оптимизацию пространственной и временной дискретизации часто осуществляют раздельно: число парциальных пучков N (микролинз в маске Гартмана–Шака) выбирают, исходя из соотношения радиуса телескопа R и радиуса корреляции волнового фронта r_F . При этом временную дискретизацию часто считают уже выбранной, так что она, опираясь на временной интервал корреляции τ , обеспечивает заданное быстродействие системы.

Гипотеза о возможности рассмотрения волнового фронта как случайного поля с разделимыми временной и пространственной корреляционными функциями позволяет осуществить параметрический синтез анализатора волнового фронта новым методом. Он приводит к совместной оптимизации пространственно-временной дискретизации волнового фронта по крите-

рию минимума среднеквадратической ошибки оценивания пространственно-временного случайного поля при ограничении точности фотонными шумами светового сигнала. Такой подход позволяет выделить три вида ошибок измерения волнового фронта (координат пятен гартманогаммы):

- линейную пространственную ошибку, уменьшающуюся пропорционально числу N парциальных пучков;
- линейную временную ошибку, уменьшающейся пропорционально кадровой частоте $f_k = 1/T_k$;
- шумовую ошибку, увеличивающуюся пропорционально кадровой частоте и числу N парциальных пучков раstra Гартмана (числу микролинз маски Гартмана–Шака).

Результатом параметрического синтеза, осуществляемого приравниванием нулю производных суммарной ошибки измерения координат световых пятен гартманогаммы, являются формулы для оптимального числа N_{opt} парциальных пучков и оптимального времени кадра $T_{k\ opt}$. Коэффициенты пропорциональности в данных формулах различны при учёте задержки на считывание изображения в телекамере и в случае исследования волнового фронта без учёта задержек на считывание. В обоих случаях имеют место функциональные зависимости [97]:

$$T_{k\ opt} \sim \tau^{3/4} (R/r_F)^{1/2}; \quad N_{opt} \sim (R/r_F) \tau^{1/2}.$$

Важным аспектом этих формул является следствие совместной оптимизации пространственной и временной дискретизации – зависимость оптимальной пространственной дискретизации не только от пространственных параметров волнового фронта, но и от временных, а также зависимость оптимальной кадровой частоты не только от временных параметров волнового фронта, но и от пространственных.

Возможен вариант оценки близости выбранного времени кадра к оптимальному без измерения временного интервала корреляции волнового фронта на основе сравнения дисперсий первой и второй конечных разностей координат пятен гартманогаммы [109].

Одна из основных характеристик пространственно-временной дискретизации случайного поля – значение смаза пятна гартманогаммы из-за его движения в процессе накопления. Смаз удобно характеризовать не абсолютным значением, а относительным: отношением значения *смаза* m_v

изображения пятна гартманогаммы к его эффективному радиусу ρ . На этапе становления космического телевидения [45] считалось, что пространственно-временная дискретизация телевизионного изображения должна приводить к значению смаза порядка элемента разрешения (размера точечного объекта). Приведённые соотношения позволяют получить иной вывод [97]: при наблюдении случайно движущегося сигнального пятна размером ρ оптимальная по критерию минимума среднеквадратической ошибки дискретизация даёт существенно меньшее значение смаза.

Оптимальный смаз обратно пропорционален отношению сигнал/шум Ψ в сигнальном пятне. Это означает, что при оптимальной дискретизации смаз случайно движущегося пятна настолько мал, что дисперсия межкадровой разности $D_{\Delta t}$ из-за движения объекта должна быть равна дисперсии σ^2 шума (амплитуда межкадровой разности равна среднеквадратическому значению шума), а движение пятна по фотоприёмнику характеризуется **слитностью траектории** [97]:

$$m_v = \frac{\rho}{\Psi}; \quad D_{\Delta t} = \sigma^2.$$

Этот результат является частным случаем применения в прикладном телевидении принципа слитности телевизионных пространственно-временных сигналов, рассмотренного в гл. 3.

5.3. Стабилизация изображения с прямым измерением смещения

Из широкой проблематики учёта движения изображений в данном параграфе выделим частную, но достаточно актуальную задачу компенсации смаза изображения, вызываемого колебаниями визирной оси телекамеры при воздействии механических возмущений типа вибрации основания.

Далее рассмотрен структурный и параметрический синтез телекамеры, реализующей прямой метод измерения возмущений [184]–[187]. Прямой метод измерения смещения может быть реализован, если наблюдаемый сюжет содержит объекты с переменной яркостью как в центральном поле, так и на периферии. Требование скоростного совмещения изображений ведёт к необходимости использования прямого метода измерения смещения изображений, т. е. введения в фокальную плоскость кроме основного матричного фотоприёмника двух дополнительных линейных фотоприёмников (рис. 5.7). При этом дополнительные линейные массивы могут быть сформированы как топологически, так и виртуально, программным способом

выделения из общего массива фоточувствительных элементов. С точки зрения компьютерной системотехники этот метод переводит твёрдотельную телекамеру в новый класс приборов с *множественным потоком данных*, способствующим извлечению доминантной информации с минимальными вычислительными затратами. С точки зрения микроэлектроники на рис. 5.7 представлена уже не просто фотоприёмная матрица, а *видеосистема на кристалле*.

В соответствии с принципом доминантной информации, рассмотренным в гл. 2, *цель* системы стабилизации изображения состоит в минимизации смаза изображения из-за вибраций телекамеры; *доминантная информация* для основного массива определяется целью системы (смыслом, качеством и количеством видеoinформации). Для дополнительных линейных фотоприёмников *доминантой* является оценка смещения изображения, её точность и задержка измерения; *качество информации* понимается как отношение количества информации о смещении изображения к общему количеству информации, передаваемому измерительной подсистемой.

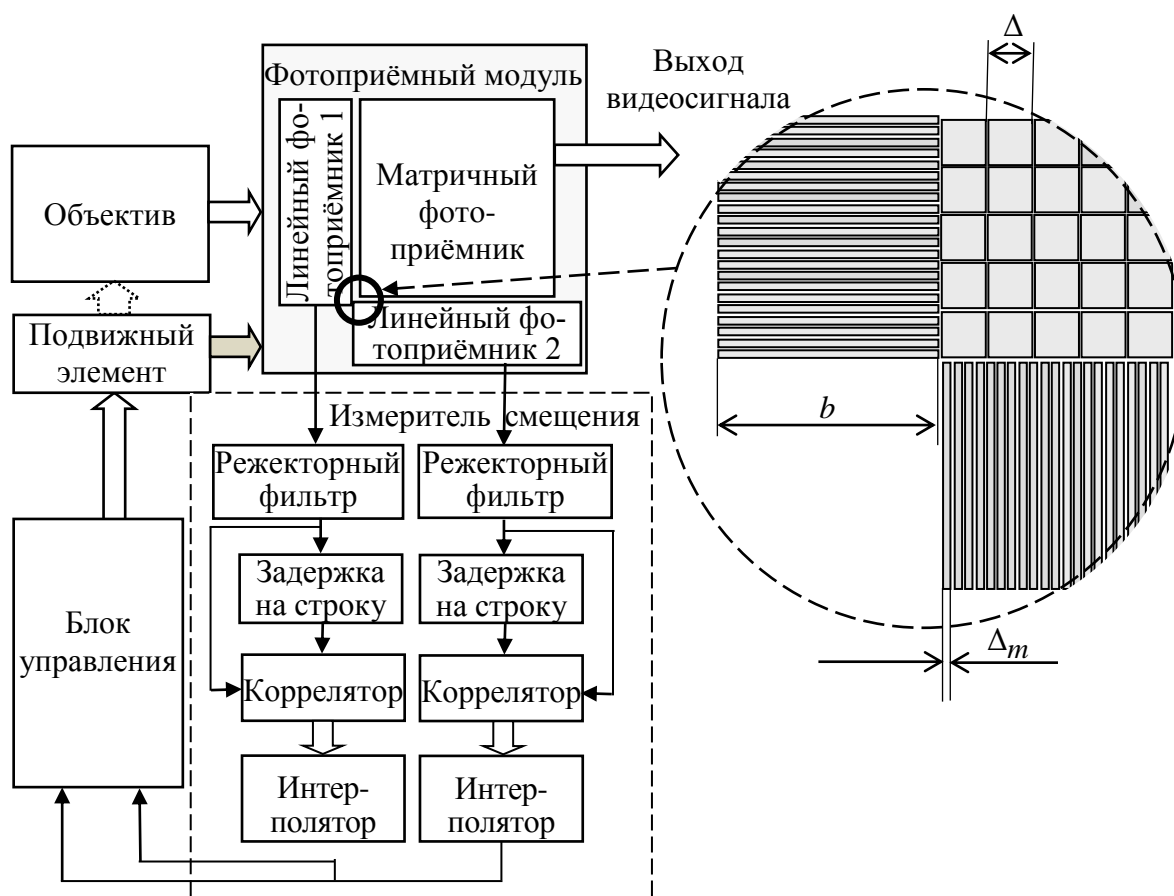


Рис. 5.7. Функциональная схема системы стабилизации изображения с прямым измерением смещения изображений дополнительными линейными фотоприёмниками

Основной положительный эффект рассматриваемого метода, состоящий в расширении на два порядка диапазона скоростей компенсируемого смаза из-за возмущений визирной оси, достигается за счёт повышения на два порядка частоты считывания информации с измерительных линейных фотоприёмников по сравнению с кадровой частотой основного массива.

На рис. 5.7 показано, что размеры и форма пикселей вспомогательных линейных фотоприёмников отличаются от таковых для основного массива. Размер пикселя вспомогательного фотоприёмника Δ_m по стороне, примыкающей к основному массиву, делается меньше размера пикселя Δ основного фотоприёмника; размер пикселя вспомогательного линейного фотоприёмника в ортогональном направлении b делается существенно больше размера пикселя Δ основного фотоприёмника. Эти различия обусловлены необходимостью повышения точности измерения смещения изображения при времени накопления строки T_c во вспомогательных фотоприёмниках, много меньшем времени кадра T_k основного фотоприёмника. Существенно большая площадь пикселя измерительной строки $b\Delta_m$ по сравнению с площадью пикселя основной матрицы Δ^2 при равенстве удельной ёмкости c (вычисляемой в электронах на единицу площади) обеспечивает необходимый для повышения частоты считывания запас в накопительной способности пикселей: $T_k N_e = c\Delta^2 \approx T_c N_{e\ m} = cb\Delta_m$. Различие в режимах управления чувствительностью в трактах обработки видеoinформации основного массива и во вспомогательных линейных фотоприёмниках состоит в том, что в основном массиве чрезмерное увеличение освещённости парируется системой автоматического регулирования времени накопления (АРВН) при постоянной частоте кадров, а во вспомогательных линейных фотоприёмниках целесообразно повышение строчной частоты при времени накопления, равном времени строки.

Оптимальная структура измерителя смещения изображений включает:

- режекторный фильтр (декоррелятор сигнала) для контрастирования взаимно корреляционной функции (ВКФ) [47];
- коррелятор – вычислитель ВКФ для дискретных значений сдвига (на целое число пикселей) [47];
- интерполятор – вычислитель смещения, как целочисленного относительно пикселей изображения [47], так и дробного (так называемая субпиксельная точность измерения) [190], [191].

Эти три блока могут проектироваться как самостоятельные СФ-блоки, и система в целом может быть реализована как видеосистема на кристалле. Сложность вычислителя, слагаемая из сложности трёх указанных блоков, определяется исходными параметрами сигнала и телекамеры, а также выбранными параметрами этих блоков.

Структурный синтез системы стабилизации изображения подчинён цели такого скоростного высокоточного парирования возмущений визирной оси, чтобы ошибки неточности стабилизации маскировались шумом изображения. Структурный и параметрический синтез входящего в систему стабилизации вычислителя подчинён цели достижения потенциальной точности оценки смещения изображения в системе стабилизации с дополнительными линейными фотоприёмниками при следующей опорной триаде синтеза:

1. *Априорная информация* для параметрического синтеза о типовых изображениях, формируемых фотоприёмными массивами телекамеры:

- телекамера с кадровым периодом T_K формирует последовательность сигналов изображения основного массива $N \times N$ квадратных пикселей площадью $\Delta \times \Delta$ и две дополнительные последовательности $y_{i \rightarrow}$ и $y_{i \uparrow}$ с ортогональных линейных массивов по J пикселей каждый с размерами $b\Delta_m$; время считывания T_c этих строк существенно меньше времени кадра T_K основного массива: $T_K \gg T_c$;

- диаметр импульсной характеристики объектива (функции рассеяния точки, кружка рассеяния) согласован с размером Δ пиксела основного массива;

- автокорреляционная функция (АКФ) изображений, анализируемых линейными фотоприёмниками, при достаточно больших аргументах в силу свойств наблюдаемых сюжетов моделируется экспоненциальной функцией $K(x) = K(0) \exp(-|x|/\tau)$ с интервалом корреляции $J\Delta \gg \tau \gg \Delta$;

- АКФ изображений, анализируемых линейными фотоприёмниками, в силу свёртки оптического сигнала с импульсной характеристикой объектива $h_{об} \approx \exp(-x^2/2r^2)$ с эффективным диаметром $2r$ при малых аргументах $x \leq \Delta$ описывается не линейной функцией (как принято в [47]), а более сложной (расчёт [190] показывает зависимость формы АКФ от отношения r/Δ и отношения сигнал/шум). Эксперимент устанавливает, что АКФ при малых аргументах описывается квадратичной функцией (рис. 5.8);

- все сигналы наблюдаются на фоне белого (некоррелированного) гауссовского шума ξ_i , включающего шум считывания и фотонный шум, т. е. наблюдению доступны сигналы $y_i = s_i + \xi_i$.

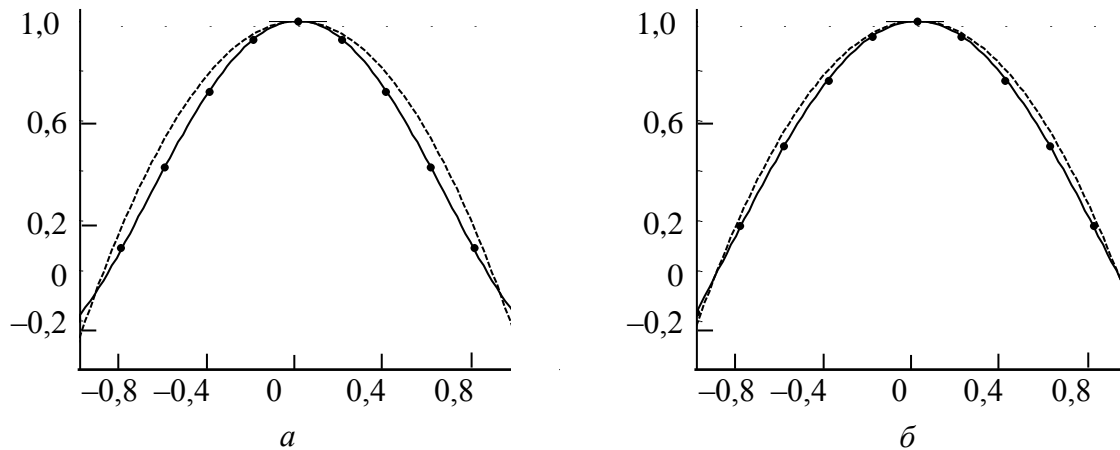


Рис. 5.8. Пример нормированной АКФ изображения по результатам эксперимента при $r/\Delta < 1$ до контрастирования режекторным фильтром (а) и после (б) при аппроксимации дискретных замеров (показаны точками) кривой 2-го (пунктир) и 4-го (сплошные линии) порядков

2. *Критерий качества*: максимум качества информации (минимум информационного риска, включающего потери информации из-за искажений смаза и загрузку получателя шумовой информацией) в сигнале изображения, формируемом основным фотоприёмным массивом. Критерий качества достигается:

- при максимально возможной частоте считывания со вспомогательных линейных фотоприёмников;
- при минимуме среднеквадратической ошибки измерения смещения сигналов соседних строк вспомогательными линейными фотоприёмниками.

3. *Ограничения*: сложность вычислителя смещения изображения, стабильность вычислений, требуемые для реализации вычислителя в виде отдельной цифровой СБИС или в виде СФ-блоков, которые могут быть включены в состав видеосистемы на кристалле.

Варьируемыми (искомыми) параметрами являются:

- размеры пиксела измерительной строки;
- разрядность АЦП;
- число членов импульсной характеристики режекторного фильтра;
- алгоритм вычисления корреляции и его сложность;
- алгоритм интерполяции и его сложность.

При нахождении размеров пиксела измерительной строки введём отношение $\beta = \Delta/\Delta_m$. Аналитически найти оптимальное значение β трудно, так как формально ошибка измерения смещения (с учётом интерполяции ВКФ) не зависит от β в силу взаимно компенсирующего влияния ёмкости пикселей N_{em} и их числа J в измерительной строке. Вместе с тем достижимая точность измерения связана с методологической ошибкой интерполяции ВКФ: чем меньше β , тем устойчивее система к отклонениям АКФ реальных сигналов от принятой модели. С другой стороны, уменьшение β ведёт к возрастанию сложности вычислителя, его требуемого быстродействия и тепловыделения, а также к необходимости использования существенно более дорогой технологии изготовления с лучшими топологическими нормами. Поэтому приходится находить компромиссное значение β , можно прогнозировать, что оно должно лежать в интервале 2...5.

Для выбора оптимальных параметров сначала зафиксируем связь сложности блоков с разрядностью m АЦП сигналов вспомогательных линейных фотоприёмников. Предварительная фильтрация с помощью режекторных фильтров обеспечивает не только известное [47] обострение ВКФ, но и уменьшение сложности накапливающего сумматора вычислителя ВКФ пропорционально $\log_2 \tau/\Delta_m$.

Основными требованиями к режекторному фильтру, обеспечивающему обострение ВКФ, являются симметрия импульсной характеристики $\{h_k\}$ и равенство нулю суммы всех её членов:

$$\sum_{k=-l}^{k=l} h_k = 0; \quad h_0 = 1; \quad h_l = 0 \quad \text{при } k > l.$$

Синтез импульсной характеристики режекторного фильтра опирается на критерий минимума функционала, включающего взвешенную сумму потери полезной информации ΔI об измеряемом параметре сдвига и сложности фильтра W [97]. Потеря полезной информации связана с коэффициентом изменения отношения сигнал/шум $K_\Psi = K_{\text{ш}}/K_s$ из-за возрастания мощности шума на выходе фильтра пропорционально коэффициенту шума

$$K_{\text{ш}} = \sum_{k=-\infty}^{k=\infty} h_k^2$$

и возможного уменьшения энергии сигнала одноэлементных

фрагментов и перепадов яркости, в первом приближении определяемого отношением полуширины импульсной характеристики фильтра к длитель-

ности фронта видеоимпульса $K_s = [\min \{l/\beta; 1\}]^2$. Это означает, что в рассматриваемой системе полуширина l импульсной характеристики режекторного фильтра должна быть не меньше отношения β размеров пиксела основного и вспомогательных фотоприёмников.

При оценке сложности блоков вычислителя смещения изображений будем считать, что

- сложность характеризуется числом вентиляей;
- минимальная сложность ячейки памяти – 2 вентиля;
- минимальная сложность одноразрядного сумматора 5 вентиляей.

При оценке сложности режекторного фильтра можно считать, что она определяется только требуемыми элементами задержки со сложностью $4lm$ и $2l$ сумматорами со сложностью $4(m + l)$, так как применяется только умножение на целочисленные степени числа 2 ($1/2, 1/4, 1/8, \dots$), осуществляемое простым сдвигом кода, ведущим к увеличению разрядности кода на l . Далее при количественной оценке параметров будем полагать $\beta = 4$.

Простейшим фильтром, отвечающим поставленным условиям, является фильтр с импульсной характеристикой $\{h_{-1} = -1/2; h_0 = 1, h_1 = -1/2\}$ (рис. 5.9, а). Этот фильтр, известный в радиолокации как череспериодный компенсатор второго порядка, рекомендуется к применению в задаче обострения ВКФ [47, кн. 2]. Так как применение этого фильтра при $\beta > 1$ ведёт к заметной потере информации, необходимо обратиться к фильтрам с большим числом членов импульсной характеристики, примеры которых показаны на рис. 5.9, б–г, а параметры сведены в таблицу.

Фильтр	l	$K_{ш}$	K_s	K_{ψ}	Δl	W
Рис. 5.9, а	1	1,5	0,0625	24	2,3	$8(m + 1)$
Рис. 5.9, б	2	1,25	0,25	5	1,16	$16(m + 1)$
Рис. 5.9, в	2	1,266	0,25	5,06	1,17	$16(m + 2)$
Рис. 5.9, г	4	1,125	1	1,125	0,08	$32(m + 1,5)$
–	8	1,0625	1	1	0,04	$64(m + 1,5)$

Эти фильтры имеют следующие импульсные характеристики:

- рис. 5.9, б: $\{h_{-2} = h_{-1} = h_1 = h_2 = -1/4; h_0 = 1\}$;
- рис. 5.9, в: $\{h_{-1} = h_1 = -0,3125; h_{-2} = h_2 = -0,1875; h_0 = 1\}$;
- рис. 5.9, г: $\{h_{-4} = h_{-3} = h_{-2} = h_{-1} = h_1 = h_2 = h_3 = h_4 = -1/8; h_0 = 1\}$.

Фильтр на рис. 5.9, в является оптимальным для обнаружения одноэлементных сигналов на фоне с экспоненциальной АКФ, хотя и требует несколько бóльших вычислительных затрат.

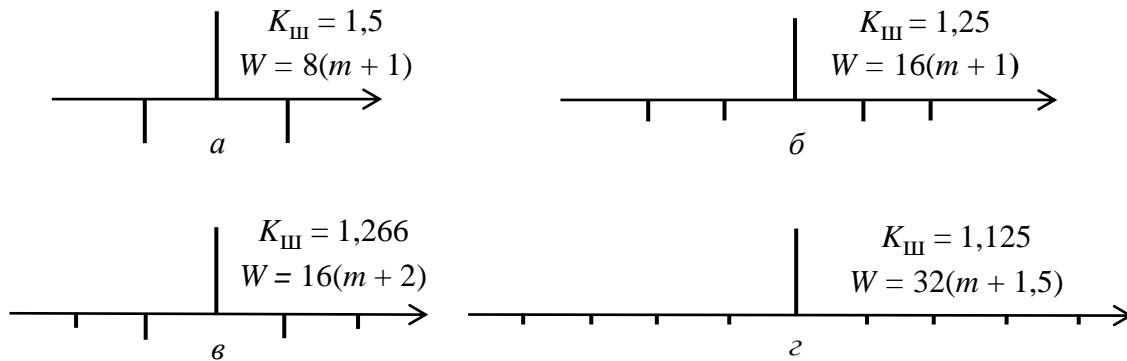


Рис. 5.9. Импульсные характеристики режекторных фильтров и их параметры

Таким образом, наблюдается уменьшение потери информации о смещении изображения ценой увеличения сложности фильтра. Весовые коэффициенты при этих величинах следует определять с учётом сложности остальной части вычислителя. Для широкого диапазона значений этих коэффициентов при выбранном значении $\beta = 4$ наилучший компромисс между точностью и сложностью обеспечивает фильтр с импульсной характеристикой, показанной на рис. 5.9, г.

Рассмотрим *разрядность квантователя видеосигнала и результат измерения смещения*. При параметрическом синтезе измерителя смещения изображений важен вопрос: какова потеря информации о параметре сдвига χ из-за равномерного, оптимального с точки зрения информационного риска квантования [3]? Для видеосистемы на кристалле, построенной по рассматриваемому способу, площадь пиксела измерительной строки на два порядка превосходит площадь пиксела основного фотоприёмного массива, что соответствует ёмкости пиксела $N_{em} = 10^6$ электронов и более. Для такой ёмкости среднее значение дисперсии флуктуационного шума $\sigma_{фл}^2$ составляет $N_{em}/2$, а оптимальное при оценке *формы сигнала* равномерное квантование [3] требует числа разрядов $m_{opt} \approx 1/2 \log(2N_{em} \log N_{em}) = 12$. Это значение соответствует внесению дополнительного шума квантования с дисперсией $\sigma_{кв}^2 = N_{em} 2^{-2m} / 12$ и коэффициенту шума $K_{III} = (\sigma_{кв}^2 + \sigma_{фл}^2) / \sigma_{фл}^2$, который соответствует потере информации о параметре смещения 0,007 бит при $m = 12$; 0,1 бит при $m = 10$ и 0,9 бит при $m = 8$. Эта потеря информации существенно меньше потери информации о *форме* сигнала, которая при

$m = 12$ составляет 1/4 бит/пиксел, что для измерительной строки в 10^3 пикселов составит примерно 250 бит. Можно показать, что для типового значения интервала корреляции сигнала 16 пикселов среднеквадратическое значение ошибки измерения смещения составляет $\varepsilon \approx 8 \cdot 10^{-3} \Delta_m = 2 \cdot 10^{-3} \Delta$. Оцененное значение ошибки совмещения должно обеспечить накопление в основном массиве фотоприёмных элементов без потери разрешающей способности, т. е. обеспечить смаз не более $m_v \leq \Delta / N_e^{1/2}$, что для типовых матричных фотоприёмников составляет $2 \dots 7 \cdot 10^{-3} \Delta$. Так как рассматриваемая система имеет целью маскирование шумом ошибок, вызванных смазом, можно считать, что априорный интервал a возможных сдвигов за время считывания измерительной строки $a = \Delta_m / 2$. Тогда оптимальная с точки зрения минимума информационного риска разрядность кода оценки смещения (на выходе вычислителя) $m_\chi = \log_2^*(a/\varepsilon)$ [3], что для приведённого примера составляет 7 разрядов (* обозначает ближайшее большее целое число). На практике рационально использование вычислителя с разрядностью оценки смещения $m_\chi = 8$. Критерий минимума информационного риска в этой задаче применяется в виде (2.18), т. е. цена потери информации весьма высока и коэффициент при потере информации должен быть на порядок больше, чем при шумовой информации. При таком подходе оптимальным значением разрядности квантователя сигнала измерительных строк (на входе вычислителя) является $m = 10$. При этом потеря информации о параметре сдвига составляет 0,1 бита. Таким образом, рассмотренная структура и определённые параметры обеспечивают совпадение требуемой и достижимой точности измерения смещения.

Коррелятор является самой сложной частью исследуемой системы измерения смещения сигнала. Будем считать, что априорное значение смещения изображения от измерения к измерению составляет $a\Delta_m$, т. е. a пикселов. При этом количество значений вычисляемых коэффициентов корреляции $n = 2a^* + 1$. В прецизионной *следающей* системе стабилизации изображения частота измерения должна выбираться достаточно большой, для того чтобы смаз изображения не проявлялся. Для этого необходимо, чтобы выполнялось условие $a < 1$. Вычисления заметно упрощаются, если обеспечивается условие $a < 0,5$. Отметим, что значение смещения изображения на полпиксела за время кадра было зафиксировано как эмпирическая оценка [45] без учёта обработки изображений и стабилизации раstra.

Малое значение сдвига сигнала за время считывания измерительной строки при интерполяции полиномом 2-й степени ведёт к необходимости и достаточности вычисления взаимной корреляции в трёх точках: с нулевым сдвигом и со сдвигом на пиксел влево и вправо. При интерполяции полиномом 4-й степени необходимо измерение взаимной корреляции в пяти точках. Далее для примера рассмотрим сложность простого коррелятора с интерполяцией полиномом 2-й степени, корректной при больших значениях отношения размеров кружка рассеяния и элемента накопления r/Δ .

Сложность вычисления взаимной корреляции соседних строк во многом связана с операцией умножения. Для облегчения аппаратной реализации коррелятора целесообразно перейти от прямого вычисления произведений разных сигналов к несколько более просто организуемой операции возведения в квадрат. Возможность такой реализации определяется известной формулой [69], использованной в 5.1:

$$M\{[y(x) - y(x - \chi)]^2\} = 2K(0) \left[1 - \frac{K(\chi)}{K(0)} \right]. \quad (5.10)$$

При любом из этих способов вычисления корреляции в схему коррелятора входят устройство задержки на строку с J пикселями и три устройства задержки на элемент с совокупной сложностью $(J + 3)m$.

Сложность коррелятора при прямом методе вычислений определяется сложностью входящих в него умножителя и накапливающего сумматора:

$$W_K \approx (J + 3)m + 3(m^2 + m) + 4(2m + \log J).$$

Анализ в пакете Quartus II фирмы «Altera» для семейства микросхем программируемой логики Cyclone IV GX показывает, что не прямое вычисление корреляции по (5.10) для принятых ранее параметров изображения требует примерно в 2,7 раза меньше вычислительных ресурсов, чем прямое вычисление. Это обусловлено тем, что возведение в квадрат проще, чем умножение, и тем, что благодаря вычитанию коррелированных сигналов уменьшается разрядность накапливающих сумматоров. При этом общая сложность меньше, чем требуемая для вычисления ВКФ за счёт отказа от нормировки накопленных статистик. Такая возможность связана с целью измерения – поиском не функции корреляции изображения, а лишь аргумента её максимума. Формула (5.10) изменяет структуру вычислителя, который вычисляет три статистики: не корреляцию, а значения дисперсий разностных сигналов D_{-1} , D_0 , D_1 , разрядность которых $2m + \log J$:

- предыдущей строки и сдвинутой на пиксел влево текущей строки D_{-1} ;

- предыдущей строки и текущей строки D_0 ;
- предыдущей строки и сдвинутой на пиксел вправо текущей строки D_1 .

Вычислитель-интерполятор по полученным значениям статистик D_{-1} , D_0 , D_1 вычисляет значение оценки смещения на основании интерполяции ВКФ при малых смещениях. При использовании параболической интерполяции формула для оценки смещения достаточно проста:

$$\hat{\chi} = \Delta_m(D_{-1} - D_1) / [2(D_{-1} - 2D_0 + D_1)].$$

Сложность такого интерполятора превышает сложность коррелятора в силу большей разрядности входных сигналов:

$$W_{\text{и}} \approx 3[(m + \log J)^2 + m + \log J] + 4(2m + \log J).$$

При малых значениях отношения размеров кружка рассеяния и элемента накопления r/Δ в силу перехода вида АКФ сигнала от параболического к экспоненциальному меньшие алгоритмические ошибки обеспечивает аппроксимация АКФ на интервале $[-\Delta, \Delta]$ полиномом 4-й степени (см. рис. 5.8). По полученным трём значениям статистик D_{-1} , D_0 , D_1 полином 4-й степени может быть построен только при предварительной калибровке системы с учётом конкретного значения отношения r/Δ размеров кружка рассеяния и пиксела фотоприёмника.

Параболическая (степени 2 или 4) интерполяция статистик D_{-1} , D_0 , D_1 обеспечивает потенциальную точность измерения смещения и, в конечном счёте, стабилизацию изображений с маскированием шумом смаза изображения, несмотря на возмущения визирной оси телекамеры.

Освоенные мировой микроэлектроникой топологические нормы точности позволяют реализовать процессор вычисления корреляции с элементами памяти, занимающий площадь порядка десятой части площади всего кристалла, на котором, как и в обычной фотоприёмной КМОП-матрице, основную часть площади занимает массив фоточувствительных элементов. Для повышения быстродействия вычисления измерения смещения целесообразно использовать распараллеливание считывания сигнала с дополнительных фотоприёмных модулей [186]. Итак, изложенное в данном параграфе можно обобщить следующим образом:

1. Осуществлённый структурный и параметрический синтез быстродействующего прецизионного измерителя смещения изображений является приложением новой парадигмы видеоинформатики, которая выдвигает на первый план извлечение максимума качественной информации из потока

фотонов при наличии возмущений, а не просто констатирует деградацию разрешения из-за смаза.

2. Рассмотренный метод стабилизации открывает путь к практической реализации СБИС класса «видеосистема на кристалле», обеспечивающей формирование изображений *без потери разрешающей способности* в условиях интенсивных возмущений визирной оси, т. е. позволяет обеспечить *смаз, маскируемый шумом*.

3. Прецизионное скоростное совмещение изображений – условие стабилизации для неискажённого накопления в прикладном телевидении (а не просто для удобства зрителя) – достигается *прямым измерением* (а не косвенным с датчиками ускорения) с использованием *дополнительных фотоприёмников*, переводящих такую ВСнК в класс приборов с *множественным потоком данных*.

4. Прецизионное измерение смещения изображений должно учитывать квадратичный вид автокорреляционной функции при малых смещениях; в зависимости от соотношения размеров светочувствительного элемента и кружка рассеяния объектива оптимальными могут быть интерполирующие полиномы 2-й или 4-й степеней.

5. Предельное быстродействие и обеспечиваемое им расширение диапазона ускорений парируемых возмущений определяются доступным увеличением площади пиксела вспомогательных линейных приёмников; реально использование длины пиксела, на два порядка большей его ширины. Это значит, что верхняя граница частоты компенсируемых искажений на *два порядка* выше кадровой частоты.

5.4. Телевизионно-лазерная локация

Телевизионно-лазерная локация предназначена для определения пространственно-временных координат объектов, например при контроле сближения и стыковки космических аппаратов. Согласно принципу доминантной информации, рассмотренному в гл. 2, *цель* телевизионно-лазерного локатора – обнаружение и оценка координат объектов; *доминантной информацией* являются пространственно-временные координаты объекта и их производные; *качество информации* определяется отношением количества информации о доминанте к общему количеству информации, передаваемому телекамерой. Кроме этой явной задачи телевизионно-лазерный локатор выполняет ещё одну, пусть и промежуточную, но концептуально крайне важную функцию – *выделение доминантного сигнала подавлением*

(торможением) фонового сигнала от объектов, находящихся на иных дальностях. Это позволяет говорить о том, что в ходе функционирования телевизионно-лазерный локатор не просто накапливает фотоны, отражённые от объектов на определённой дальности, а выполняет функцию накопления качества информации.

Принципиальным отличием телевизионно-лазерного локатора от традиционного лазерного локатора является сканирование не по углу, а по дальности при поиске доминантного объекта. Это обуславливает достижение тактического преимущества скорости обнаружения объекта ценой потери дальности обнаружения. В результате телевизионно-лазерная локация перспективна в ближней и средней зонах и не применяется при больших дальностях. Сканирование по дальности имеет две характерные стадии – поиск объекта и сопровождение по дальности, методология которого традиционна для радиолокации. Телевизионно-лазерная локация по сравнению с обычной лазерной локацией может обеспечить выигрыш во времени обнаружения объекта, достигаемый за счёт излучения энергии света в широком угле, равном углу зрения телекамеры. Для достижения указанного выигрыша кроме расширения угла излучения подсветки необходимо применение итерационных методов, позволяющих за счёт сокращения точности измерения дальности на первых шагах существенно сократить их количество. Для этого применим важнейший метод сокращения вычислительных операций – реализация накопления информации по принципу дихотомии, широко используемому в АЦП поразрядного уравнивания.

В активных телевизионно-лазерных системах используются телевизионные камеры со временами накопления, составляющими доли микросекунды. Для реализации столь малых времен накопления использовались стробируемые электронно-оптические преобразователи (ЭОП). Времена включения/выключения ЭОП могут составлять единицы наносекунд. В то же время использование ЭОП имеет целый ряд недостатков: невысокий квантовый выход фотокатода, трудности сочленения с твёрдотельным фотоприёмником, в том числе потерю преимущества жесткого раstra, большие массу и габаритные размеры, очень высокую стоимость и малый ресурс работы.

С этой точки зрения представляет интерес реализация малых времен накопления непосредственно на матричном фотоприёмнике. Разработанный метод апробирован на матричном ПЗС [188], но может быть модифицирован и для применения в КМОП-матрицах с глобальным затвором.

В стандартном режиме работы матричного ПЗС со строчным переносом время накопления определяется интервалом от момента окончания импульса сброса зарядовых пакетов в подложку SUB (рис. 5.10, *а*) до окончания импульса считывания накопленных зарядовых пакетов из фотодиода в вертикальный ПЗС-регистр U_H . Уровень U_M соответствует потенциальной яме в регистре, уровень U_L – потенциальному барьеру. Подача высокого потенциала U_H формирует глубокую потенциальную яму под электродом, достаточную для переноса зарядовых пакетов из фотодиодов. Необходимо отметить, что накопление пакетов во время действия импульса U_H продолжается. В разработанном методе импульс SUB формируется таким образом, что он уничтожает все заряды, образовавшиеся до и после интервала накопления (рис. 5.10, *б*). В этом случае минимальная длительность интервала накопления будет определяться длительностью фронтов импульса SUB.

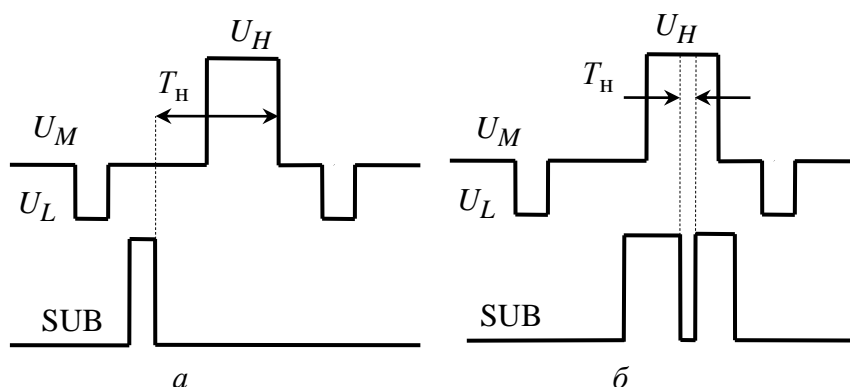


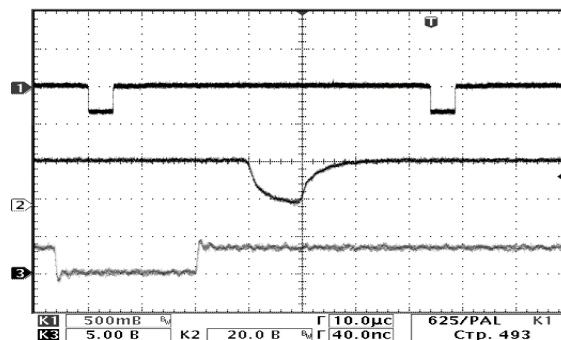
Рис. 5.10. Временные диаграммы импульсов управления накоплением и считыванием в режиме электронного затвора:
а – общепринятый режим для длительного накопления;
б – режим стробирования

В разработанном методе импульс *SUB* формируется таким образом, что он уничтожает все заряды, образовавшиеся до и после интервала накопления (рис. 5.10, *б*). В этом случае минимальная длительность интервала накопления будет определяться длительностью фронтов импульса SUB.

На рис. 5.11 представлены результаты эксперимента по получению времени накопления, равного 40 нс в матричном ПЗС ICX659AL фирмы «Sony». В процессе эксперимента импульс маломощного лазера по времени сдвигался относительно интервала накопления. На рис. 5.11 показаны различные случаи временного положения импульса лазера относительно интервала накопления: с перекрытием по времени, до интервала накопления и после него. Рядом с различными случаями приведены изображения пучка лазера, полученные на экране видеомонитора.



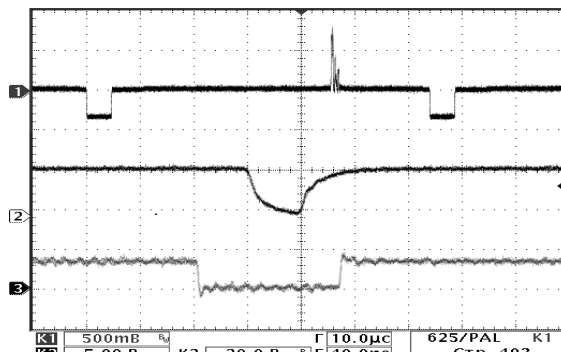
a



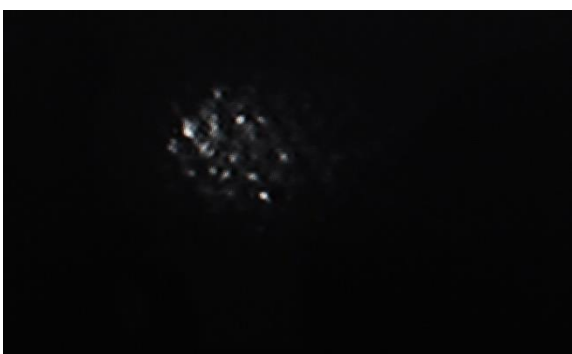
2



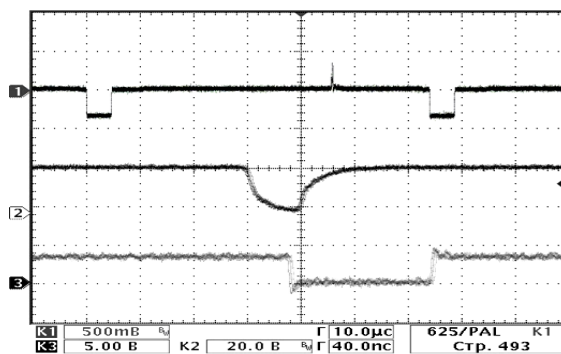
б



д



в



e

Рис. 5.11. Результаты экспериментального исследования макета активной телевизионно-лазерной импульсной системы. Изображения катафота, полученные: *a* – при несовпадении отражённого светового импульса с интервалом накопления; *б* – при полном совпадении; *в* – при частичном перекрытии, и осциллограммы: *2* – видеосигнала; *д* – интервала накопления; *e* – импульса лазера

Экспериментально подтвержденное время накопления в 40 нс позволяет измерять дальность до объекта с точностью 6 м [188]. При длительности импульса лазера, большем времени накопления, взвешивание сигналов соседних кадров позволяет на порядок повысить точность – до долей метра. Проведённый энергетический расчёт показал, что для уверенного измерения дальности при размере катафота 0,1 м поверхностная плотность энергии должна составлять 0,1 Дж/км². Такую плотность энергии может обеспечить твёрдотельный лазер на алюмоиттриевом гранате.

Указанный метод стробирования накопления может быть использован для повышения контрастов изображений в активных телевизионно-лазерных импульсных системах, работающих в рассеивающих средах, например в воде. Стробирование накопления информации может отсеять нежелательные отражения от дальностей, больших и меньших интересующей, оно может оказаться полезным при наблюдении различного вида физических экспериментов, например взрывов, пролётов заряженных элементарных частиц в искровых камерах и т. д.

Полученные для активной телевизионно-лазерной импульсной системы результаты говорят о существенно большей эффективности твёрдотельного решения по сравнению с использовавшимися ранее системами с ЭОП.

Выводы по гл. 5. Предложенные новые методы управления накоплением качества видеоинформации, обеспечивающие оптимальную слитность воспроизведения изображения как пространственно-временного процесса по его дискретным отсчётам (маскирование шумом артефактов – смаза, ошибок дискретизации), отражают важный эффект влияния твёрдотельной технологии на изменение методов системного проектирования в видеоинформатике. Если в эпоху электронно-лучевого телевидения господствовала парадигма пассивного учёта ухудшения разрешающей способности из-за движения изображения, то твёрдотельная технология способствовала выдвиганию на первый план максимизации не столько количества, сколько *качества информации* адаптацией параметров телекамер, в том числе параметров разложения изображения. Таким образом:

1. Многообразие сюжетов и целей прикладных видеоинформационных систем порождает многообразие методов управления накоплением качества информации, которые опираются на общий для них принцип максимума качества информации и его частную форму – *практическую слитность* изображений.

2. Предложен метод управления накоплением, реализующий *принцип совместной адаптации* параметров разложения к пространственно-временным характеристикам сюжета *обменом чёткости на кадровую частоту* по балансу межэлементных и межкадровых приращений сигнала.

3. Предложен метод *совместной* оптимизации пространственной и временной дискретизации при синтезе динамического теста Гартмана–Шака для анализа волнового фронта, приведший к правилу равенства дис-

персии межкадровой разности и дисперсии шума и обосновавший существование *оптимального значения смаза* изображения случайно движущегося объекта.

4. Предложен метод прямого измерения смещения изображения с помощью дополнительных быстродействующих линейных фотоприёмников, позволяющих на два порядка расширить диапазон возмущений визирной оси без деградации качества информации.

5. Предложен метод управления накоплением в матричном фотоприёмнике в составе активной телевизионно-лазерной импульсной системы, позволяющий благодаря управлению длительностью и фазой времени накопления в дополнение к определению направления на объект реализовать быстрый алгоритм высокоточного измерения дальности до объекта.

Все рассмотренные методы опираются на *принцип накопления качества информации*, внедрение которого обусловлено совместным применением достижений твёрдотельной технологии и теории информации.

Изложенное в этой главе позволяет считать, что основные конфликты идей, отмеченные в преамбуле к ней, если не разрешены, то заметно сглажены. Важнейшим шагом можно считать реализацию в ряде задач *принципа совместной адаптации* параметров разложения, в первую очередь чёткости и кадровой частоты. Конечно, проблема априорной неопределённости настолько широка, что разрешить её окончательно можно только на понятийном уровне, отметив, что существует два основных подхода – создание адаптивных методов и методов, устойчивых к отклонению от априори принятых статистик сигналов. При этом иногда оказывается, что лучшими характеристиками обладают адаптивные методы, а иногда – устойчивые, и эти предпочтения зависят от уровня технологии. В приложении к рассматриваемой проблематике накопления качества информации в телекамерах примерами такой альтернативы могут служить адаптивный метод регулировки времени накопления для расширения светового диапазона телекамеры и метод расширения динамического диапазона с помощью использования цифрового пиксела большой разрядности (см. гл. 2).

Огромное разнообразие *целей* прикладных телевизионных систем позволяет считать проблему адаптации телекамер неисчерпаемой.

ЗАКЛЮЧЕНИЕ

Начало эпохи твёрдотельного телевидения связывают с рождением приборов с зарядовой связью в 1971 г. Этот технологический прорыв не только вытеснил электронно-лучевые приборы, но и наполнил новыми идеями технику связи, породив телевизионно-компьютерные системы. Вместе с тем отличия технологии приборов с зарядовой связью и КМОП-технологии вычислительных систем первые тридцать лет твёрдотельной революции в телевидении обусловили во многом самостоятельное развитие теории и техники.

Теория и практика твёрдотельного телевидения развиваются, питая друг друга идеями и возможностями. Эта тенденция на современном языке обозначается как конвергенция (сближение) теории информации и микроэлектроники, благодаря междисциплинарности обеспечивающая изменение парадигмы развития от анализа к синтезу и переходу к нанотехнологиям. Хотя и теоретики, и технологи часто пытаются обходиться друг без друга, такие «сепаратистские» позиции постепенно изживаются. Так, изложенный в данной книге материал охватывает широкий спектр вопросов от теории информации до физики и технологии твёрдотельных фотоприёмников. Это не случайность: в любой области науки и техники, кроме узких специальных работ, должны быть работы широкого охвата проблематики. Здесь авторы следовали концепции С. Лема, который в книге «Сумма технологии» писал: *«Сущностью "эмпирической эпохи" в технологии было не столько отсутствие теории, сколько её вторичность... Я бы рискнул утверждать даже, что технологическая эмпирика стремится развиваться до тех пор, пока это вообще возможно. <...> Следует обратить внимание на различный подход к модели со стороны ученого и со стороны технолога. Технолог, получив возможность "синтеза живого организма" – если такова была его цель, – удовлетворится "конечным продуктом". Учёный – по крайней мере, учёный в классическом понимании – стремится детально изучить "теорию синтеза организмов". Учёный жаждет алгоритма, технолог же скорее походит на садовника, который, сажая дерево и срывая яблоки, не заботится о том, "как яблоня это сделала". Учёный считает такой узкоутилитарный, прагматический подход прегрешением против канонов полного познания. Нам кажется, что в будущем обе эти позиции изменятся».* Насколько сблизились эти позиции теории и техники твёрдотельного телевидения на страницах этой книги – судить читателю.

Второй этап твёрдотельной революции, определяемый доминированием КМОП-фотоприёмных матриц и КМОП-видеосистем на кристалле, является в определённом смысле даже более революционным, чем её первый этап. Лишь второй этап твёрдотельной революции во весь голос заявил о повсеместном внедрении *рукотворных «зрячих» кристаллов*, которые вошли не просто в каждый дом, но и почти в каждый компьютер и мобильный телефон, который теперь только в силу привычки можно называть *телефоном*.

Вместе с тем уже первый, ПЗС-ный, этап твёрдотельной революции стал важным для теории телевидения, так как обратил пристальное внимание на роль входного – фотонного – шума. Важность достижения предельной чувствительности, ограниченной фотонным шумом, до этого отмечалась только в инфракрасных системах, тогда как в электронно-лучевых приборах доминировали шумы усилителей, и лишь в экзотических системах с вакуумными электронно-оптическими умножителями удавалось приблизиться к счёту фотонов малыми группами. Стимулированная достижениями технологии теория кодирования зашумлённых сигналов в реальном времени охватила не только передачу сигналов, но и их рождение. На этом пути принцип доминантной информации, опирающийся на учение А. А. Ухтомского о доминанте и насыщенный конкретикой в виде аксиом и формул о доминантной и шумовой информации, в соответствии с идеей А. Н. Колмогорова, вывел информацию из разряда скалярных величин и явился первым шагом на пути построения теории не скалярной информации. В теорию информации внесена новая парадигма: информация – не атрибут материи, а посредник между духом и материей; теория информации – не теория каналов, не теория количества информации, а теория, имеющая аксиоматику, введённые уравнение, принципы, законы и критерии, отражающие тройственную структуру информации, имеющую *количество, качество и цель*. Конечно, изложенная концепция – не истина в конечной инстанции. Авторы разделяют мнение [192] о том, что *«принимаемая в качестве парадигмы теория должна казаться лучшей, чем конкурирующие с ней другие теории, но она вовсе не обязана (и фактически этого никогда не бывает) объяснять все факты, которые могут встретиться на её пути... ещё Фрэнсис Бэкон писал: "Истина всё же скорее возникает из заблуждения, чем из неясности"»*. В силу теоремы Гёделя полная формализация человеческого знания невозможна, но, опираясь на введённый прин-

цип доминантной информации и предложенную аксиоматику, можно конкретизировать вектор развития теории информации в направлении формализации информационно-образной связи, т. е. принципов и законов рождения и передачи *семантической информации*. Можно прогнозировать, что будущие исследования, отталкиваясь от достигнутых результатов развития теории информации, охватившей рассмотрение вопросов не только передачи, но и *рождения* информации, включают в круг решаемых проблем рождение и кодирование семантической информации. Такой вектор развития теории связан с переходом от концепции Дж. Уиллера «*Всё из бита!*» (*It from bit*) к лозунгу «*Всё из образа!*» (*It from pattern*).

Теория доминантной информации в первую очередь опирается на принцип накопления качества информации, родословную которого можно проследить от принципа экономии, известного Аристотелю. Об этом принципе С. Лем писал: «*Мы стремимся открыть не Порядок вообще, а лишь некоторый определенный порядок, наиболее экономный ("бритва" Оккама!)*». Принцип накопления качества информации находит важные пункты плодотворного взаимодействия с практикой твёрдотельного телевидения, обосновывая идею макроквантов, характеризуемых адаптивными размерами зоны накопления, т. е. переменным числом электронов в группе-«куче» при групповом счёте фотонов. Поэтому такие важные, казалось бы, чисто технические продвижения, как цифровой пиксел или переменная площадь зоны накопления, не «висят в воздухе», а являются следствием оптимизации кодирования видеоинформации в твёрдотельной телекамере в ходе *рождения* информации.

Методология проектирования твёрдотельных фотоприёмников и телекамер, опирающаяся на принцип накопления качества информации, позволила ввести в практику ряд новых методов управления накоплением, в том числе обобщить принцип параллельной обработки информации, включив в зону его влияния не только фотоприёмную и вычислительную подсистемы, но и связь между ними. Важными приложениями принципа накопления качества информации являются совместная адаптация чёткости и кадровой частоты, совместная оптимизация и/или адаптация пространственных и временных характеристик телевизионных гартмановских датчиков, астродатчиков, стабилизаторов и дальномеров.

Твёрдотельная революция в телевидении началась тогда, когда уровень точности изготовления СБИС составлял 3...5 мкм; её второй этап,

связанный с появлением видеосистем на кристалле, начался при освоении ультрафиолетовой литографии и проектных норм точности 0,35 мкм.

Сегодня ведущие производители СБИС владеют технологиями уровня 65, 45 и даже 22 нм. Использование технологии с проектными нормами 180 нм с соответствующими им САПР характеризует уровень, *минимально достаточный* для создания отечественных видеосистем на кристалле класса «активный пиксел». В настоящее время цивилизация переживает переход к шестому технологическому укладу, который оформляется в 2010–2020 гг., фаза зрелости прогнозируется на 2040-е гг. Он нацелен на наукоёмкие («высокие») технологии, главными среди которых станут нанотехнологии, фотоника, микромеханика, появятся квантовые компьютеры. Это означает, что в 2020–2025 гг. произойдёт новая научно-техническая революция. Не случайно именно на это время прогнозируется окончание действия закона Мура об экспоненциальном росте числа элементов в одной микросхеме [193]. Однако микроэлектроника уже давно преодолела границу норм точности 100 нм, официально считающуюся границей нанотехнологий, и заняла лидирующие позиции в переходе к шестому технологическому укладу. Дальнейшее продвижение в область «высоких» технологий (десятки и единицы нанометров) открывает совершенно новые возможности в видеоинформатике. Так, ранее научно-прикладное телевидение сделало видимым многое недоступное глазу, во многом благодаря включению невидимых глазом диапазонов длин волн, возможности длительного накопления в твёрдом теле фотоприёмника и компьютерной обработке изображений. Сейчас благодаря возможности размещения в каждом пикселе большого числа элементов открылась возможность радикального повышения контрастной чувствительности, что позволит сделать для телекамер видимым то, что ранее было невидимым. Другим примером прорыва в микроэлектронике должны стать видеосистемы на кристалле с множественными параллельными каналами связи фотоприёмного массива пикселов с расположенными на этом же кристалле устройствами кодирования и принятия решений.

В ближайшей перспективе развития отечественная твёрдотельная видеоинформационная техника создаст:

- ВСнК с множеством параллельных каналов связи фотоприёмного массива с вычислительной подсистемой;
- ВСнК с множественными потоками информации от многоспектрального фотоприёмного массива к вычислительной подсистеме;

- КМОП-матричные фотоприёмники большого формата с множеством выходных устройств для повышения кадровой частоты;
- КМОП-матричные фотоприёмники с цифровым пикселем высокой разрядности, обеспечивающие предельную контрастную чувствительность и попиксельную адаптацию ВСнК к свойствам сюжета;
- КМОП-матричные фотоприёмники с логарифмическим пикселем с умеренным количеством разрядов;
- КМОП-линейные фотоприёмники с режимом ВЗН;
- КМОП-фотоприёмники с предельной квантовой эффективностью, достигаемой за счёт утонения и обратного освещения матрицы;
- технологии изготовления матричных фотоприёмников с возможностью беззазорной стыковки с одной, двух, трёх и четырёх сторон, необходимых для создания крупноформатных фокальных мозаик.

Для решения этих задач дальней и ближайшей перспективы потребуется концентрация научных сил, финансовых и материально-технических средств, которая обеспечит прорыв в области элементной базы, перспективных технологий для создания средств связи, обнаружения, наведения. При этом следует ориентироваться на обеспечение не просто сокращения разрыва в уровне микроэлектроники с ведущими странами, но и на выделение ряда перспективных изделий, в которых можно получить новое, отсутствующее в зарубежных системах *качество систем и информации*, доставляемой ими.

Норберт Винер утверждал, что уже есть много специалистов, которые в области радиоэлектроники «знают как», но очень мало людей, которые «знают что» (что делать). Сегодня ведущие отечественные специалисты видеоинформационной техники уже знают «*что* делать» и знают «*как* делать». Конечно, сегодняшнее научное знание не абсолютно, и так же, как на первом этапе твёрдотельной революции нельзя было угадать свойства техники её второго этапа, сейчас трудно предсказать методы совершенствования телевизионной техники будущего. Возможно, благодаря объединению усилий в области теории информации и твёрдотельной технологии научно-прикладное телевидение сольётся с направлением «нанотехнологий» и породит «нано-видеоинформатику» или «нанофотонику». Можно только с уверенностью сказать, что никто никогда не откажется развивать телевизионную технику и видеоинформатику – основу «зрячих» машин, многократно усиливающих возможности человеческого познания.

СПИСОК ЛИТЕРАТУРЫ

1. Твёрдотельное телевидение / Л. И. Хромов, Н. В. Лебедев, А. К. Цыцулин, А. Н. Куликов. М.: Радио и связь, 1986. 184 с.
2. Хромов Л. И., Цыцулин А. К., Куликов А. Н. Видеоинформатика. М.: Радио и связь, 1991. 192 с.
3. Твёрдотельная революция в телевидении. Телевизионные системы на основе приборов с зарядовой связью, систем на кристалле и видеосистем на кристалле / В. В. Березин, А. А. Умбиталиев, Ш. С. Фахми и др.; под ред. А. А. Умбиталиева и А. К. Цыцулина. М.: Радио и связь, 2006. 312 с.
4. Лем С. Сумма технологии. М.: ООО «Изд-во АСТ»; СПб.: Terra Fantastica, 2004. 668 с.
5. Дойч Д. Начало бесконечности: Объяснения, которые меняют мир. М.: Альпина нон-фикшн, 2014. 581 с.
6. Лапшин И. И. Философия изобретения и изобретение в философии. М.: Республика, 1999. 399 с.
7. Реньи А. Записки студента по теории информации. Трилогия о математике. М.: Мир, 1980. 376 с.
8. Wheeler J. A. Information, physics, quantum: The search for links, // Complexity, Entropy, and the Physics of Information. W. Zurek (ed.). Redwood City, CA: Addison-Wesley, 1990.
9. Хорган Дж. Конец науки: Взгляд на ограниченность знания на закате Века Науки. СПб.: Амфора, 2001. 479 с.
10. Холево А. С. Квантовые системы, каналы, информация. М.: МЦНМО, 2010. 328 с.
11. Хартли Р. Передача информации. Теория информации и её приложения: сб. пер. / под ред. А. А. Харкевича. М.: ГИФМЛ, 1959. 328 с.
12. Винер Н. Кибернетика или управление и связь в животном и машине. М.: Сов. радио, 1968. 328 с.
13. Штейнгауз Г. Математика – посредник между духом и материей. М.: БИНОМ. Лаб. знаний, 2005. 351 с.
14. Хромов Л. И., Цыцулин А. К. Основания космической видеоинформатики // Вопр. радиоэлектроники. Сер. Техника телевидения. 2011. Вып 1. С. 6–31.
15. Теория передачи электрических сигналов при наличии помех: сб. пер. / под ред. Н. А. Железнова. М.: Иностран. литер., 1953. 288 с.
16. Шеннон К. Работы по теории информации и кибернетике. М.: Иностран. литер., 1963. 832 с.
17. Зворыкин В. К., Мортон Д. А. Телевидение. М.: ГИФМЛ, 1956. 780 с.
18. Оливер Б. М. Эффективное кодирование. Теория информации и её приложения: сб. пер. / под ред. А. А. Харкевича. М.: ГИФМЛ, 1959. 328 с.
19. Карнап Р. Философские основания физики: Введение в философию науки. М.: Изд-во ЛКИ, 2008. 360 с.
20. Планк М. Единство физической картины мира: сб. ст. М.: Наука, 1966. 288 с.
21. Пуанкаре А. О науке. М.: ГИФМЛ, 1983. 560 с.
22. Гейзенберг В. Физика и философия. Часть и целое. М.: Наука, 1989. 400 с.
23. Колмогоров А. Н. Теория информации и теория алгоритмов. М.: Наука, 1987. 304 с.
24. Ухтомский А. А. Доминанта. СПб.: Питер, 2002. 448с.
25. Хромов Л. И. Теория информации и теория познания. СПб.: РФО, 2006. 200 с.

26. Антипин М. В. Интегральная оценка качества телевизионного изображения. Л.: Наука, 1970. 154 с.
27. Роуз А. Зрение человека и электронное зрение. М.: Мир, 1977. 216 с.
28. Цуккерман И. И. Дискретная оптика и методы обработки сверхслабых телевизионных изображений // Телевизионная техника и связь. 1995. С. 38–44.
29. Стемповский А., Шилин В. КМОП-фотодиодные СБИС. Перспективная элементная база однокристалльных систем приёма и обработки информации // Электроника: наука, технология, бизнес. 2003. № 2. С. 14–20.
30. Fossum E. R. Active pixel sensor – Are CCD's dinosaurs? Charge-Coupled Devices and Optical Sensors III // Proc. of the SPIE. 1993. Vol. 1900. P. 2–14.
31. Theoretical basis and experimental confirmation: why a CMOS is superior to a CCD / L. J. Kozovski, D. Stanley, J. Luo at al. // Proc. of the SPIE. 1999. Vol. 3698. P. 388–396.
32. Немудров В. Г., Мартин Г. Системы на кристалле. Проблемы проектирования и развития. М.: Техносфера, 2004. 216 с.
33. Алфёров Ж. И. Перспективы электроники в России. Гетероструктурная электроника и акустоэлектроника // Электроника: наука, технология, бизнес. 2004. № 6. С. 90–93.
34. Нейман фон Дж. Теория самовоспроизводящихся автоматов. М.: Мир, 1971. 384 с.
35. Семповский А. Л., Амербаев В. М., Соловьёв Р. А. Принципы рекурсивных модулярных вычислений // Информ. технологии. 2013. № 2. С. 22–27.
36. Александров В. В., Кулешов С. В., Цветков О. В. Цифровая технология инфокоммуникации. Передача, хранение и семантический анализ текста, звука, видео. СПб.: Наука, 2008. 244 с.
37. Чернавский Д. С. Синергетика и информация. Динамическая теория информации. М.: ЛИБРОКОМ, 2009. 304 с.
38. Лысенко Н. В. Основы видеоинформатики. СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2002.
39. Люхин А. В., Умбиталиев А. А. Задачи космических оборонных видеоинформационных систем // Вопр. радиоэлектроники. Сер. Техника телевидения. 2013. Вып. 2. С. 3–14.
40. Рыфтин Я. А. Телевизионная система. Теория. М.: Сов. радио, 1967. 272 с.
41. Левит А. Б. Введение в общую теорию телевидения. М.: Сов. радио, 1967. 116 с.
42. Лебедев Д. С., Цуккерман И. И. Телевидение и теория информации. М.–Л.: Энергия, 1965.
43. Глезер В. Д., Цуккерман И. И. Информация и зрение. М.–Л.: Изд-во АН СССР, 1961. 184 с.
44. Гуревич С. Б. Теория и расчёт не вещательных систем телевидения. Л.: Энергия, 1970. 236 с.
45. Брацлавец П. Ф., Росселевич И. А., Хромов Л. И. Космическое телевидение. М.: Связь, 1973. 248 с. (1-е изд. 1967).
46. Цифровое кодирование телевизионных изображений / И. И. Цуккерман, Б. М. Кац, Д. С. Лебедев и др.; под ред. И. И. Цуккермана. М.: Радио и связь, 1981. 240 с.
47. Прэтт У. Цифровая обработка изображений. М.: Мир, 1982. Кн. 1. 312 с.; кн. 2. 480 с.
48. Вудворд Ф. М. Теория вероятностей и теория информации с применениями в радиолокации. М.: Сов. радио, 1955. 128 с.
49. Миддлтон Д. Введение в статистическую теорию связи: в 2 т. М.: Сов. радио, 1962. Т. 1. 783 с.; т. 2. 832 с.
50. Статистическая теория связи и её практические приложения / под ред. Б. Р. Левина. М.: Связь, 1979. 288 с.
51. Черри К. Человек и информация. М.: Связь, 1972. 368 с.
52. Галлагер Р. Теория информации и надёжная связь. М.: Сов. радио, 1974. 720 с.

53. Costas J. P. Coding with linear systems // Proc. IRE. 1952. Vol. 40. P. 1101–1103.
54. Горбунов А. К., Пинскер М. С. Эпсилон-энтропия с задержкой гауссовского зашумлённого сигнала // Проблемы передачи информации. 1988. Т. 24, № 3. С. 18–23.
55. Хромов Л. И., Ковригин А. Б., Мартынихин А. В. Принцип равновесного согласования в теории информации // Докл. АН. 1995. Т. 344, № 1. С. 30, 31.
56. Теория кодирования непрерывного сигнала и её применение в прикладном телевидении / Л. И. Хромов, А. Б. Ковригин, А. К. Цыцулин, А. В. Мартынихин // Техника средств связи. Сер. Техника телевидения. 1985. Вып. 1. С. 3–11.
57. Идеальное линейное кодирование телевизионных сигналов / Л. И. Хромов, А. Б. Ковригин, А. В. Мартынихин, А. К. Цыцулин // Техника средств связи. Сер. Техника телевидения. 1987. Вып. 4. С. 3–10.
58. Хромов Л. И., Литвинчук Л. А. Оценка пропускной способности телевизионного канала с учётом линейных искажений и нескольких источников шумов // Техника средств связи. Сер. Техника телевидения. 1981. Вып. 1. С. 3–10.
59. Хромов Л. И., Ковригин А. Б., Мартынихин А. В. Принцип двойственности в теории информации // Техника средств связи. Сер. Техника телевидения. 1991. Вып. 3. С. 3–11.
60. Зубакин И. А., Фахми Ш. С., Цыцулин А. К. Решения уравнения связи // Вопр. радиоэлектроники. Сер. Техника телевидения. 2008. № 2. С. 9–27.
61. Зубакин И. А. Ограничения в теории линейного кодирования // Вопр. радиоэлектроники. Сер. Техника телевидения. 2011. Вып. 2. С. 127–134.
62. Цыцулин А. К., Зубакин И. А. Множественность решений уравнения связи // Вопр. радиоэлектроники. Сер. Техника телевидения. 2013. Вып. 2. С. 47–59.
63. Фейнман Р. Характер физических законов. М.: АСТ: Астрель, 2012. 252 с.
64. Пойа Д. Математика и правдоподобные рассуждения. М.: Наука, 1975. 464 с.
65. Хромов Л. И. Поиск смысла теории виртуальной связи // Вопр. радиоэлектроники. Сер. Техника телевидения. 2007. Вып. 2. С. 82–91.
66. Хромов Л. И. Теоретические аспекты передачи информации в реальном масштабе времени // Телевизионная техника и связь. 1995. С. 21–27.
67. Николаев Б. И. Последовательная передача дискретных сообщений по непрерывным каналам с памятью. М.: Радио и связь, 1988. 264 с.
68. Левин Б. Р., Шварц В. Вероятностные модели и методы в системах связи и управления. М.: Радио и связь, 1985. 312 с.
69. Френкс Л. Теория сигналов. М.: Мир, 1974. 344 с.
70. Цыбаков Б. С. Линейное кодирование сообщений // Радиотехника и электроника. 1962. Т. 7, вып. 1. С. 375–385.
71. Кошелёв В. Н. Квантование с минимальной энтропией // Проблемы передачи информации. 1963. Вып. 3. С. 151–156.
72. Витерби А. Д., Омура Дж. К. Принципы цифровой связи и кодирования. М.: Радио и связь, 1982. 536 с.
73. Величкин А. И. Передача аналоговых сообщений по цифровым каналам связи. М.: радио и связь, 1983. 240 с.
74. Gray R. M., Neuhoff D. L. Quantization // IEEE Transaction on inform. Theory. Vol. 44, № 6. P. 2325–2383.
75. Verdu S. Fifty years Shannon theory // IEEE Transaction on inform. theory. 1998. Vol. 44, № 6. P. 2057–2078.
76. Быховский М. А. Пионеры информационного века. История развития теории связи. М.: Техносфера, 2006. 376 с.
77. Талер В. Теоретические ограничения скорости передачи информации // Теория информации и её приложения. М.: ГИФМЛ, 1959. 328 с.

78. Ричардсон Я. Видеокодирование. H.264 и MPEG-4 – стандарты нового поколения. М.: Техносфера, 2005. 366 с.
79. Гуткин Л. С. Оптимизация радиоэлектронных устройств по совокупности показателей качества. М.: Сов. радио, 1985. 368 с.
80. Моисеев Н. Н. Математические задачи системного анализа. М.: Наука, 1981. 487 с.
81. Цыцулин А. К., Фахми Ш. С., Зубакин И. А. Начальный этап проектирования кодера источника непрерывного сигнала // Вопр. радиоэлектроники. Сер. Техника телевидения. 2010. Вып. 2. С. 17–34.
82. Сакрисон Д. Лекции по аналоговой связи / пер. с англ. М.: Мир, 1974. 168 с.
83. Марр Д. Зрение. Информационный подход к изучению представления и обработки зрительных образов / пер. с англ. М.: Радио и связь, 1987. 400 с.
84. Маригодов В. К., Бабуов Э. Ф. Синтез оптимальных радиосистем с адаптивным предсказанием и корректированием сигналов. М.: Радио и связь, 1985. 248 с.
85. Демидов В. Е. Как мы видим то, что видим. М.: Знание, 1979. 208 с.
86. Василенко Г. И., Тараторин А. М. Восстановление изображений. М.: Радио и связь, 1986. 304 с.
87. Обработка изображений и цифровая фильтрация / под ред. Т. Хуанга. М.: Мир, 1979. 319 с.
88. Ван Трис Г. Теория обнаружения, оценивания и модуляции: в 3 т. М.: Радио и связь. Т. 1. 1972. 744 с.; т. 2. 1975. 245 с.; т. 3. 1977. 664 с.
89. Добрушин Р. Л., Цыбаков Б. С. Передача информации с дополнительным шумом // Проблемы передачи информации. 1963. № 14. С. 21–42.
90. Стратонович Р. Л. Теория информации. М.: Сов. радио, 1975. 424 с.
91. Харкевич А. А. О ценности информации // Проблемы кибернетики. 1960. Вып. 4. С. 53–58.
92. Гештальт психология. URL: www.geshtaltpsy.ru
93. Хургин Я. И., Яковлев В. П. Фinitные функции в физике и технике. М.: Наука, 1971. 408 с.; 2-е изд. М.: Либроком, 2010. 414 с.
94. Ухтомский А. А. Доминанта души. Из гуманитарного наследия. Рыбинск: Рыбинское подворье, 2000. 608 с.
95. Цыцулин А. К. Избирательность и теория информации // Вопр. радиоэлектроники. Сер. Техника телевидения. 2007. Вып. 1. С. 3–9.
96. Цыцулин А. К. Теория линейного кодирования зашумлённых сигналов // Вопр. радиоэлектроники. Сер. Техника телевидения. 2009. Вып. 2. С. 16–40.
97. Цыцулин А. К. Телевидение и космос. СПб.: Изд-во СПбГЭТУ «ЛЭТИ», 2014. 240 с.
98. Цыцулин А. К. Построение телевизионных систем для наблюдения подвижных изображений // Техника средств связи. Сер. Техника телевидения. 1988. Вып. 5. С. 3–11.
99. Лем С. Молох. М.: АСТ Транзит-книга, 2005. 781 с.
100. Цыцулин А. К. Распределение ошибок по каскадам в многокаскадной системе связи // Техника средств связи. Сер. Техника телевидения. 1991. Вып. 3 С. 49–57.
101. Вейль Г. О философии математики. М.: КомКнига, 2010. 128 с.
102. Бриллюэн Л. Научная неопределённость и информация. М.: Мир, 1966. 272 с.
103. Linvill W. K. Sampled-data control systems studied through comparison of sampling with amplitude modulation // Trans. AIEE. 1951. Part II. Vol. 70. P. 1779–1788.
104. Ярославский Л. П. Цифровая обработка сигналов в оптике и голографии. М.: Радио и связь, 1987. 296 с.
105. Кайзер Д. Цифровые фильтры: прил. к кн. Голд Б., Рэйдер Ч. Цифровая обработка сигналов. М.: Сов. радио, 1973. 368 с.

106. Росселевич И. А., Зайцев В. П., Хромов Л. И. Выбор варианта системы на начальном этапе проектирования в малокадровом телевидении // Техника средств связи. Сер. Техника телевидения. 1982. Вып. 5. С. 3–8.
107. Манцветов А. А., Умбиталиев А. А., Цыцулин А. К. Принцип слитности в прикладном телевидении // Вопр. радиоэлектроники. Сер. Техника телевидения. 2012. Вып. 1. С. 3–15.
108. Смелков В. М. Иду на растр. Эссе об изобретениях по классу H04N. Великий Новгород, 2007. 176 с.
109. Быстродействующая телевизионная измерительная система для оценки искажений волнового фронта методом Гартмана // В. В. Войцехович, С. Н. Анкудинов, А. А. Манцветов и др. // Опт. журн. Т. 67, № 2. 2000. С. 113–119.
110. Манцветов А. А., Цыцулин А. К. Телекамеры на КМОП-фотоприёмниках // Вопр. радиоэлектроники. Сер. Техника телевидения. 2006. Вып. 2. С. 70–89.
111. Гридин В. Н., Титов В. С., Труфанов М. И. Адаптивные системы технического зрения. М.: Наука, 2009. 434 с.
112. Krymski A. I., Tu N. A 9-V/lux-s 5000-frames/s 512×512 CMOS sensor // IEEE Trans. Electron Devices. 2003. Vol. 50, № 1. P. 136–143.
113. Комаров А. С. Актуальные вопросы развития отечественной электроники. М.: ОАО «Росэлектроника», 2014. 204 с.
114. A Multiresolution 100-GOPS 4-Gpixels/s Programmable Smart Vision Sensor for Multisense Imaging / L. Lindgren, J. Melander, R. Johansson, B. Möller // IEEE J. of Sol.-St. Circuits. 2005. Vol. 40, № 6. P. 1350–1359.
115. A Wide Dynamic Range CMOS Image Sensor With Multiple Exposure-Time Signal Outputs and 12-bit Column-Parallel Cyclic A/D Converters / M. Mase, S. Kawahito, M. Sasaki et al. // IEEE J. of Sol.-St. Circuits. 2005. Vol. 40, № 12. P. 2787–2795.
116. A High-Speed, High-Sensitivity Digital CMOS Image Sensor With a Global Shutter and 12-bit Column-Parallel Cyclic A/D Converters / M. Furuta, Y. Nishikawa, I. Toru, S. Kawahito // IEEE J. of Sol.-St. Circuits. 2007. Vol. 42, № 4. P. 766–774.
117. Multiple-Ramp Column-Parallel ADC Architectures for CMOS Image Sensors / M. F. Snoeij, A. J. P. Theuwissen, K. A. A. Makinwa, J. H. Huijsing // IEEE J. of Sol.-St. Circuits. 2007. Vol. 42, № 12. P. 2968–2977.
118. High performance silicon imaging. Fundamentals and applications of CMOS and CCD image sensors / Ed. by D. Durini // Woodhead Publishing series in electronic and optical materials. № 60. Elsevier, 2014. 551 p.
119. Нано-КМОП-схемы и проектирование на физическом уровне // Б. П. Вонг, А. Миттал, Ю. Цао, Г. Старр. М.: Техносфера, 2014. 432 с.
120. Адамов Ю. Ф. Проектирование систем на кристалле. М., 2005. 112 с.
121. Микросхема КМОП-приёмника изображения с функцией датчика координат / В. А. Зимогляд, А. В. Лепендин, Д. Ю. Адамов и др. // Вопр. радиоэлектроники. Сер. Техника телевидения. 2011. Вып. 1. С. 83–92.
122. Адамов Д. Ю. Основные вопросы проектирования видеосистем на кристалле // Вопр. радиоэлектроники. Сер. Техника телевидения. 2014. Вып. 2. С. 3–16.
123. Low-Leakage-Current and Low-Operating-Voltage Buried Photodiode for a CMOS Imager / Inoue I., Tanaka N., Yamashita H. et al // IEEE Trans. on electron devices. 2003. Vol. 50, № 1. P. 43–47.
124. Пат. РФ № 2297074. Фотоприёмная ячейка с разделением цветов / А. В. Лепендин, В. А. Гергель, В. А. Зимогляд и др. Опубл. 10.04. 2007 // Бюл. № 10.
125. Пат. РФ № 2525654. Мультиспектральное фоточувствительное устройство / Х. Сяопин. Опубл. 20.08. 2014 // Бюл. № 23.

126. САПР АВОКАД для проектирования КМОП-фотоприёмников и аналого-цифровых интегральных схем / В. В. Аблогин, Д. Ю. Адамов, В. А. Зимогляд и др. // *Вопр. радиоэлектроники. Сер. Техника телевидения*. 2010. Вып. 2. С.83–92.
127. Адамов Д. Ю., Зимогляд В. А., Сомов О. А. Повышение стойкости цифровых схем на базе стандартной КМОП-технологии // *Тр. XIII Междунар. НТК «Электронная компонентная база космических систем»*, Сочи, ГК «Бридж Ресорт», 18–24 сентября 2014 г.
128. Адамов Д. Ю. Фотоприёмные системы на кристалле для космического применения // *Тр. XIII Междунар. НТК «Электронная компонентная база космических систем»*, Сочи, ГК «Бридж Ресорт», 18–24 сентября 2014 г.
129. Гецелев И. В., Зубарев А. И., Пудовкин О. Л. Радиационная обстановка на борту космических аппаратов. М.: ЦИПК, 2001. 316 с.
130. Космические лучи и естественный радиационный фон у поверхности Земли / А. А. Бирюков, А. Ф. Крутов, А. Г. Пузырный, И. С. Цирова. Самара: Изд-во Самарского ун-та, 2012. С. 118.
131. Моделирование воздействия микрометеоритов и фрагментов космического мусора на космические аппараты / А. А. Воробьев, Т. С. Зыкова, Д. Д. Спицын и др. // *Вопр. электромеханики*. 2011. Т. 120. С. 27–30.
132. Природа электростатических разрядов на внешней поверхности космических аппаратов / А. Н. Дорофеев, А. П. Тютнев, В. С. Саенко, Е. Д. Пожидаев // *Физика и химия обработки материалов*. 2004. № 5. С. 32–37.
133. Новиков Л. С. Взаимодействие космических аппаратов с окружающей плазмой: учеб. пособие. М.: Университетская кн., 2006. 120 с.
134. Обработка видеосигнала астродатчика для подавления радиационных эффектов / Г. В. Левко, А. А. Манецветов, Д. Ю. Адамов и др. // *Вопр. радиоэлектроники. Сер. Техника телевидения*. 2014. Вып. 2. С. 24–32.
135. Mendis S., Kemeny S. E., Fossum E. R. A 128 × 128 CMOS active pixel image sensor for highly integrated imaging systems // *IEEE IEDM Tech. Dig. Dec.* 1993.
136. Zhang M., Bermak A. CMOS Image Sensor with On-Chip Image Compression: A Review and Performance Analysis // *J. of Sensors*. 2010. Vol. 2010. P. 1–17.
137. Березин В. В., Фахми Ш. С., Цыцулин А. К. Начальный этап проектирования видеосистем на кристалле // *Опт. журн*. 2012. № 11. С. 76–83.
138. Ракович Н. Н. CameraChip: полная видеосистема на кристалле // *Компоненты и технологии*. 2004. № 1. С. 128–130.
139. Пилипко Н. Е., Рычажников А. Е. Особенности работы КМОП-фотоприёмника в режиме бегущего затвора // *Изв. СПбГЭТУ «ЛЭТИ»*. 2008. Вып. 1. С. 40–54.
140. Ringaby E. Geometric computer vision for rolling-shutter and push-broom sensors // *Link-ping studies in science and technology*. 2012. 30 P. URL: www.liu/diva-portal.org/smash/get/diva2:526675/FULLTEXT01.
141. Jung-Bum Chun, Hunjoon Jung, Chong-Min Kyung. Suppressing rolling-shutter distortion of CMOS image sensors by motion vector detection // *IEEE Transaction on Consumer Electronics*. 2008. Vol. 54, № 4. P. 1479–1487.
142. Digital video stabilization and rolling shutter correction using gyroscopes / A. Karpenko, D. Jacobs, J. Back at al. // *Stanford Tech Report CTSR*. 2011. Vol. 03. P. 1–7.
143. Вахромеева О. С., Манецветов А. А., Шиманская К. А. Характеристики чувствительности телевизионных камер на матричных приборах с зарядовой связью / *Изв. вузов России. Радиоэлектроника*. 2004. Вып. 4. С. 16–26.
144. Носов Ю. Р., Шилин В. А. Основы физики приборов с зарядовой связью. М.: Наука, 1986. 320 с.

145. Манцветов А. А. Фотоэлектрические характеристики КМОП-фотоприёмников с активным пикселом // *Вопр. радиоэлектроники. Сер. Техника телевидения*. 2014. Вып. 2. С. 17–23.
146. Novel CMOS image sensor with a 132-dB dynamic range / D. Stoppa, A. Simoni, L. Gonzo, M. Gottardi // *IEEE J. of Sol.-St. Circuits*. 2002. Vol. 37, № 12. P. 1846–1852.
147. Молодяков С. А. Методика использования в цифровых камерах пульсарных процессоров кадровых ПЗС-фотоприёмников в режиме временной задержки и накопления // *Наука и образование*. 2013. № 5. С. 163–182.
148. A High-Speed Low-Noise CMOS Image Sensor With 13-b Column-Parallel Single-Ended Cyclic ADCs. / Park Jong-Ho, S. Aoyama, T. Watanabe et al. // *IEEE Trans. Electron devices*. 2009. Vol. 56. P. 2524–2533.
149. US Pat. № 7675561. Time Delayed Integration CMOS image sensor with zero desynchronization / Lepage G. B2, 2010.
150. Зимогляд В. А., Цыцулин А. К., Левко Г. В. Перспективный КМОП-фотоприёмник с функцией ВЗН // *Матер. XI конф. «Системы наблюдения, мониторинга и дистанционного зондирования Земли "ДЗЗ 2014"»*. Сочи, 13–18 сент. 2014. М., 2014. С. 351–352.
151. Wang X., Wong W., Hornsey R. A High Dynamic Range CMOS Image Sensor With Inpixel Light-to-Frequency Conversion // *IEEE Transactions on Electron Devices*. 2006. Vol. 53, № 12. P. 2988–2992.
152. Пятков В. В., Умбиталиев А. А. КМОП-фотосенсор с многоканальной схемой опроса // *Вопр. радиоэлектроники. Сер. Техника телевидения*. 2011. Вып. 2. С. 34–42.
153. Умбиталиев А. А., Манцветов А. А., Цыцулин А. К. КМОП-матричный фотосенсор предельной контрастной чувствительности // *НТК «Современные научно-технические проблемы совершенствования систем и средств РКО»*. 20 окт. 2011. М.: МАК «Вымпел», 2011.
154. КМОП-матричные фотоприёмники видимого диапазона форматов 1024×1024 и 256×256 ячеек / Д. В. Бородин, Ю. В. Осипов, А. С. Скрылёв и др. // *Матер. XI конф. «Системы наблюдения, мониторинга и дистанционного зондирования Земли "ДЗЗ-2014"»*. Сочи, 13–18 сент. 2014. М. 2014., С. 337, 338.
155. Грязин Г. Н. Системы прикладного телевидения. М.: Политехника, 2000. 277 с.
156. Звездные координаторы систем ориентации космических аппаратов / Г. А. Аванесов, С. В. Воронков, А. А. Форш, М. И. Куделин // *Изв. вузов. Приборостроение*. 2003. Вып. 4. С. 66–69.
157. Занин К. А. Выбор параметров оптико-электронной космической системы наблюдения по качеству изображения // *Полёт*. 2007. № 11. С. 30–37.
158. Ресовский В. А. Сопряжение оптики и фотоприёмников в оптикоэлектронных системах // *Вопр. радиоэлектроники. Сер. Техника телевидения*. 2011. Вып. 1. С. 54–62.
159. Захаренков В. Ф., Раковский Ю. Н., Шибяев С. Н. Об ошибках определения координат точечного источника методом центраида // *Опт. журн*. 2001. № 8. С. 61–66.
160. Методика оценки потенциальных характеристик обнаружения малоразмерных аэрокосмических целей многоспектральной оптико-электронной аппаратурой на матричных фотоприёмниках / В. А. Арутюнов, В. Г. Иванов, А. А. Каменев, А. Е. Прокофьев // *Вопр. радиоэлектроники. Сер. Техника телевидения*. 2006. Вып. 2. С. 47–69.
161. Трифонов А. П., Шинаков Ю. С. Совместное различение сигналов и оценка их параметров на фоне помех. М.: Радио и связь, 1986. 254 с.
162. Сайт ЗАО «Электроника и видеосистемы». URL: www.evs.ru
163. Оптимизация телевизионной системы пригоризонтного наблюдения / Г. В. Левко, А. А. Манцветов, А. А. Чиркунова и др. // *Вопр. радиоэлектроники. Сер. Техника телевидения*. 2014. Вып. 2. С. 33–41.

164. Morton (Z) Scan Based Real-Time Variable Resolution CMOS Image Sensor / E. Artyomov, Y. Rivenson, G. Levi, O. Yadid-Pecht // IEEE Transactions on circuits and systems for Video Technology. 2005. Vol. 15, № 7. P. 947–952.
165. Fossum E. R. Active-Pixel Image Sensors With Programmable Resolution (NPO-19510) // NASA Tech Briefs. 1996. Vol. 20, № 5. P. 26.
166. Переменная чёткость в твёрдотельных телевизионных системах / С. А. Иванов, А. Н. Куликов, Д. А. Довжиков и др. // Техника средств связи. Сер. Техника телевидения. 1989. Вып. 7. С. 27–33.
167. Шэндл Д. Датчики изображения со структурой сетчатки для систем технического зрения // Электроника. 1993. № 17. С. 7, 8.
168. Пахомов А. Н. Состояние и перспективы систем активного зрения // Заруб. радиоэлектроника. 1999. № 2. С. 57–65.
169. Kleinhans W. A. Optimum spectral filtering for background-limited infrared system // J. Opt. Soc. Amer. 1965. Vol. 55, № 1. P. 104.
170. Зубарев Ю. Б., Сагдуллаев Ю. С., Сагдуллаев Т. Ю. Видеоинформационные технологии систем связи. М.: Спутник+, 2011. 296 с.
171. Иванова А. В., Манцветов А. А. Оптимизация характеристики спектральной чувствительности фотоприёмника при наблюдении малоконтрастных изображений // Вопр. радиоэлектроники. Сер. Техника телевидения. 2011. Вып. 2. С. 26–33.
172. Сафронов Ю. А., Андрианов Ю. Г. Инфракрасная техника и космос. М.: Сов. радио, 1978. 248 с.
173. Космическая оптико-электронная аппаратура обнаружения очагов лесных пожаров / Л. А. Мирзоева, В. Д. Стариченкова, Г. А. Маковцов и др. // Опт. журн. 1992. № 8. С. 17–21.
174. Рукин М. В. Инновационные решения в области обнаружения пламени // Системы безопасности. 2009. № 4. С. 61.
175. Экспериментальное исследование спектральных и динамических характеристик пламени / А. В. Иванова, Н. Г. Мордасова, А. А. Манцветов и др. // Вопр. радиоэлектроники. Сер. Техника телевидения. 2012. Вып. 2. С. 55–58.
176. Репин В. Г., Тартаковский Г. П. Статистический синтез при априорной неопределённости и адаптация информационных систем. М.: Сов. радио, 1977.
177. Левин Б. Р. Теоретические основы статистической радиотехники. М.: Сов. радио, 1976. Кн. 3. 286 с.
178. Зубакин И. А., Козлов В. В., Цыцулин А. К. Методика тестирования видеокодеков с учётом меры нестационарности изображений // Телекоммуникации. 2012. № 8. С. 29–33.
179. Бузников А. А., Купянский А. В. Динамическое совмещение полутоновых аэрокосмических и графических изображений // Изв. вузов. Сер. Геодезия и аэрофото-съемка. 1993. № 3. С. 102–107.
180. Ушольд А. Системы стабилизации изображения // Foto & Video. 2007. № 7. С. 76–81.
181. Березин В. В., Манцветов А. А. Система цифровой стабилизации телевизионного изображения // Приборы и техника эксперимента. 1995. Вып. 6. С. 74–78.
182. Воронцов М. А., Корябин А. В., Шмальгаузен В. И. Управляемые оптические системы. М.: Наука, 1988.
183. Автоматическая стабилизация оптического изображения / Д. Н. Еськов, Ю. П. Ларионов, В. А. Новиков и др.; под ред. Д. Н. Еськова, В. А. Новикова. Л.: Машиностроение, 1988. 240 с.
184. Пат. РФ № 2384967. Способ стабилизации изображений / Ш. С. Фахми, А. К. Цыцулин, А. В. Переспелов. Опубл. 20.03. 2010 // БИ. № 8.

185. Пат. РФ № 2486688. Видеосистема на кристалле для стабилизации изображения / А. К. Цыцулин., Ш. С. Фахми., Д. О. Малашин. Оpubл. 27.06 2013 // БИ. № 18.
186. Пат. РФ №2517347. Устройство стабилизации изображения / А. К. Цыцулин, И. А. Зубакин, Д. О. Малашин, А. А. Манцветов. Оpubл. 27.05. 2014 // БИ. № 15.
187. Автономная стабилизация изображений в процессе накопления видеoinформации / А. К. Цыцулин, Ш. С. Фахми, А. А. Манцветов и др. // Опт. журн. 2012. № 11. С. 67–75.
188. Управление режимом накопления в твёрдотельных фотоприёмниках / А. А. Умбитаев, А. К. Цыцулин, А. А. Манцветов и др. // Опт. журн. 2012. № 11. С. 84–92.
189. Адаптация космической телевизионной системы к этапам наблюдения объектов / А. А. Манцветов, И. А. Зубакин, В. В. Козлов, А. К. Цыцулин // Вопр. радиоэлектроники. Сер. Техника телевидения. 2012. Вып 1. С. 29–43.
190. Голушко М. Н. Интерполяция дискретного видеосигнала в автоматических телевизионных системах // Техника средств связи. Сер. Техника телевидения. 1983. Вып. 5. С. 63–71.
191. Robin A., Moisan L., Hagarat-Masclé Le S. An A-Contrario approach for subpixel detection in satellite imagery // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2010. Vol. 32. P. 1977–1993.
192. Кун Т. Структура научных революций. М.: АСТ: АСТ МОСКВА, 2009. 317 с.
193. Каку М. Физика будущего. М.: Альпина нон-фикшн, 2013, 584 с.

Оглавление

ПРЕДИСЛОВИЕ.....	3
ОТ АВТОРОВ.....	5
ВВЕДЕНИЕ.....	7
В.1. Теория информации – основа теории телевидения.....	7
В.2. Принцип накопления качества информации.....	15
В.3. Твёрдотельная революция в телевидении.....	20
Часть I. ТЕОРЕТИКО-ИНФОРМАЦИОННЫЕ АСПЕКТЫ ПРОЕКТИРОВАНИЯ ТЕЛЕКАМЕР.....	26
Глава 1. УРАВНЕНИЕ СВЯЗИ.....	26
1.1. От идеального кодирования к равновесному.....	31
1.2. Дискретная система связи.....	41
1.3. Смешанная система связи.....	44
1.4. Непрерывная система связи и множественность решений уравнения связи для неё.....	55
Глава 2. ПРИНЦИП ДОМИНАНТНОЙ ИНФОРМАЦИИ.....	75
2.1. Проблема кодирования зашумлённых сигналов.....	81
2.2. Аксиоматика информатики рождения и передачи сигналов.....	89
2.3. Оптимальное усечение полосы частот сигналов.....	107
2.4. Принцип доминантной информации, накопление качества информации и слитность изображений.....	112
Часть II. УПРАВЛЕНИЕ НАКОПЛЕНИЕМ КАЧЕСТВА ВИДЕОИНФОРМАЦИИ В ТЕЛЕКАМЕРАХ.....	132
Глава 3. ОРГАНИЗАЦИЯ НАКОПЛЕНИЯ И СЧИТЫВАНИЯ В МАТРИЧНЫХ КМОП-ФОТОПРИЁМНИКАХ.....	132
3.1. КМОП-технология фотоприёмников.....	134
3.2. Технология проектирования фотоприёмных матриц.....	151
3.3. Учёт влияния различных параметров на этапе проектирования фотоприёмников.....	157
3.4. Управление параметрами считывания видеоинформации на этапе проектирования фотоприёмной матрицы.....	166
3.5. Чувствительность КМОП-фотоприёмников.....	177
3.6. Борьба за динамический диапазон.....	180
3.7. Борьба за потенциальную контрастную чувствительность.....	184
Глава 4. ОПТИМИЗАЦИЯ НАКОПЛЕНИЯ КАЧЕСТВА ИНФОРМАЦИИ В ТВЁРДОТЕЛЬНЫХ ТЕЛЕКАМЕРАХ.....	197
4.1. Согласование размера пиксела с сюжетом.....	198
4.2. Согласование размера пиксела с сигналом точечного объекта.....	201
4.3. Совместное обнаружение и оценивание сигналов с переменным размером зоны накопления.....	207
4.4. Группирование пикселов для повышения контрастной чувствительности.....	212
4.5. Переменная чёткость в твёрдотельных телекамерах.....	215
4.6. Оптимизация спектральной характеристики телекамеры.....	219
Глава 5. АДАПТАЦИЯ НАКОПЛЕНИЯ КАЧЕСТВА ИНФОРМАЦИИ В ТВЁРДОТЕЛЬНЫХ ТЕЛЕКАМЕРАХ.....	227
5.1. Адаптация к динамике сюжета: обмен чёткости и кадровой частоты.....	229
5.2. Адаптация к состоянию атмосферы.....	235
5.3. Стабилизация изображения с прямым измерением смещения.....	240
5.4. Телевизионно-лазерная локация.....	251
ЗАКЛЮЧЕНИЕ.....	257
СПИСОК ЛИТЕРАТУРЫ.....	262

Научное издание

Цыцулин Александр Константинович

Адамов Денис Юрьевич

Манцветов Андрей Александрович

Зубакин Игорь Александрович

Твердотельные телекамеры: накопление качества информации

Редактор Т. А. Лунаева

Подписано в печать 16.12.2014 Формат 60×84 1/16. Бумага офсетная.
Печать цифровая. Гарнитура «Times New Roman». Печ. л. 17,0.
Тираж 500 экз. Заказ177

Издательство СПбГЭТУ «ЛЭТИ»
197376, С.-Петербург, ул. Проф. Попова,5



Цыцулин А. К.
заместитель генерального ди-
ректора ОАО «НИИТ»
по научной работе, профессор
кафедры телевидения
и видеосистем
СПбГЭТУ «ЛЭТИ»
доктор технических наук
(1991)
закончил СЗПИ (1973)



Адамов Д. Ю.
генеральный директор
ООО «Юник Ай Сиз»
кандидат технических наук (2003)
закончил МИЭТ (1996)



Манцветов А. А.
доцент кафедры телевидения
и видеотехники
СПбГЭТУ «ЛЭТИ»,
ведущий научный сотрудник
ОАО «НИИТ»
кандидат технических наук (1990)
закончил ЛЭТИ (1982)



Зубакин И. А.
доцент кафедры телевидения
и видеотехники
СПбГЭТУ «ЛЭТИ»
кандидат технических наук (2010)
закончил СПбГЭТУ «ЛЭТИ» (2007)