Максимычев А.В.

Физические методы исследования. Конспект лекций. 1.Погрешности измерений.

Предмет, важность и цель курса.

Полное название данного курса: «Физические методы исследования химических и биологических систем». Под физическими методами исследования понимают обычно методы, разработанные физиками и в химии и биологии с использованием современной аппаратурной базы. Тем самым, физические методы исследования противопоставляются рутинным методам исследования, таким как взвешивание, определение температуры плавления, разделение веществ путем перегонки и т.п. Физические методы составляют применяемые экспериментальный арсенал науки, необходимый для исследования молекулярных механизмов химико-биологических процессов. Примером классического феноменологического описания процесса является запись химической реакции в виде уравнения

$$+2B \rightarrow P \tag{1}$$

Молекулярный механизма процесса предполагает знание того, сколько элементарных стадий включает реакция (1), каковы константы скоростей каждой стадии, какие промежуточные структуры и в каких конформационных состояниях образуются, как изменяется энергия реагирующей системы частиц, каково распределение электронной плотности на каждом шаге и другая информация, которая может быть необходимой для понимания молекулярной динамики процесса. На эти вопросы призваны отвечать физические методы исследования.

А



Методы исследования вообще принято подразделять на теоретические и эмпирические (Табл. 1), последние, в свою очередь, делятся на измерения и наблюдения. Пунктиром показаны взаимоотношения между различными подходами: наблюдения стимулирую проведение количественных измерений, результаты измерений подтверждают или опровергают теоретические построения, теоретические выводы предсказывают – что может быть измерено. Обращает на себя внимание, что количественные измерения, составляющие суть физических методов исследования, занимают ключевое место в приведенной схеме (Табл. 1).

Развитие методов исследования позволяет наполнить физическим содержанием понятия, которые ранее носили довольно абстрактный характер. Например, представление об активированном комплексе долгие годы использовалось в химической кинетике, но

определялось это состояние довольно туманно: нечто – «не исходное вещество, но и не продукты превращения». Развитие методов лазерной спектроскопии с фемтосекундным временным разрешением позволило количественно исследовать молекулярную динамику промежуточных состояний некоторых химических реакций, эти работы были удостоены Нобелевской премии в 1999 г (Zewail).

Нобелевский лауреат Н.Н. Семенов внес ясность в сложнейшую область химических исследований, связанную с резким изменением скорости протекания некоторых реакций при малейшем изменении внешних условий. Его предсказания о наличии активных частиц, ведущих цепочку химических превращений, были подтверждены позднее с помощью физических методов, позволяющих исследовать короткоживущие радикалы.

Изучая передачу признаков от родительских к дочерним особям, биологи пришли к выводу о существовании «фактора наследственности», но только применение метода рентген-структурного анализа позволило в 1969 г. (Уотсон. Крик) установить структуру ДНК и вывело исследования в этой области на уровень молекулярных механизмов.

Физические методы позволяют исследовать не только те явления и объекты, существование которых предполагалось ранее на основании теоретических рассуждений, но и открывают совершенно новые области науки. Методом спектроскопии Рамзаем Р.У в 1894-99 гг было открыто существование инертных газов, что имело огромное значение для «достраивания» периодической системы элементов и формирования современных квантовых представлений о строении атомов и молекул. Очевидно, чисто химическими методами обнаружить инертные элементы было бы затруднительно. Подобные примеры несложно привести и из биологии: изобретение электронного микроскопа привело к открытию вирусов, размер которых составляет 15-30 нм, возникла большая новая область исследований.

Возможности, предоставляемые современными физическими методами исследования, поражают воображение даже подготовленного специалиста, способного понять «как это работает». Метод пэтч-клямп дает возможность наблюдать функционирование единичного ионного канала клетки, сканирующие зондовые методы позволяют визуализовать и манипулировать отдельными молекулами и атомами на поверхности твердого тела, можно создавать оптические импульсы, длина которых меньше длины одного колебания световой волны. Возникли разговоры о том, что физические методы подошли к «естественной» границе своих возможностей. История учит, что такие утверждения – верный признак близости нового качественного скачка в данной области исследований.

Признанием ведущей роли новых методов в развитии науки является история присуждения Нобелевских премий -значительное количество наград было присуждено за разработку именно новых методов исследования. Из примеров последних лет можно привести: Дж. Фенн, К. Танаки, К. Вютрех – лауреаты 2002 г., награжденные за разработку методов масс-спектрометриии и ЯМР для исследования биомолекул; П. Лотесберг и П. Мэнсфилд (2003) – метод ЯМР-томографии для исследования внутренних органов; Дж. Хилл (2005) – интерференционные методы измерения, обеспечивающие точность ~10⁻¹¹ м. Признанием важнейшей роли физических методов в химии является то, что по правилам ИЮПАК (International Union of Pure and Applied Chemistry) структура вновь синтезированных или выделенных веществ должна быть подтверждена с использованием масс-спектроскопии, ЯМР, ИК-спектроскопии и др. физических методов.

Отметим, что все упомянутые в этом введении физические методы будут изучаться в рамках данного курса.

Какова цель настоящего курса? Обилие и сложность существующих методов исследования и постоянное развитие делают нереальным их детальное изучения за какието разумные сроки. Поэтому основные задачи настоящего курса видятся в следующем:

1. Познакомиться с физическими принципами, лежащими в основе базовых, наиболее распространенных физических методов исследования;

2. Выработать самостоятельный подход к изучению новых методов, который сводился бы к получению четких ответов на вопросы: что лежит в основе метода, какая информация об исследуемой системе может быть получена в результате, каковы ограничения в применении;

3. Решая поставленную экспериментально задачу, ориентироваться в существующих подходах и выбирать оптимальные методы из множества возможных.

Анализ применимости того или иного метода и планирование экспериментов невозможны без количественной основы. Отсюда вытекает важнейшее требование данного курса: умение на качественном уровне понимать суть физических явлений, строить простейшие физические модели, позволяющие «на пальцах» получать количественные оценки по порядку величины. Рассматривается метод размерностей как средство контроля полученных результатов и способ получения решений для «трудных» задач.

Виды измерений и виды эталонов.

Измерения – это всегда сравнение измеряемой величины с эталоном. Принято все измерения делить на две группы: прямые и косвенные. Прямое измерение предполагает непосредственное сравнение измеряемой величины с эталоном. Например, измеряя длину стола с помощью рулетки, мы сравниваем объект (стол) с эталоном, в качестве которого выступает рулетка. В современной лабораторной практике измерения такого рода встречаются довольно редко.

При косвенных измерениях об интересующей нас величине А мы судим на основании результатов измерения величины В, при этом функциональная связь между величинами А и В предполагается точно известной. Измерения такого рода наиболее часто используются в современной аналитической практике. Например, массспектрометрия является распространенным косвенным методом измерения массы молекул. Одна из разновидностей процедуры измерения может выглядеть так. Вначале ионизуем молекулу М путем столкновения с достаточно энергичным электроном:

$$M + e \rightarrow M^+ + 2e \tag{2}$$

Образовавшийся ион M⁺ ускорим разностью электрических потенциалов U и направим в область, в которой создано постоянное магнитное поле с напряженностью H, при этом H нормально вектору скорости v. Основные соотношения, описывающие процедуру измерения, включают закон сохранения энергии

$$\frac{mv^2}{2} = Ue \tag{3}$$

(m – масса иона M⁺, e – единичный заряд), силу Лоренца F_L, действующую в постоянном магнитном поле на движущийся заряд

$$F_L = e[v \times H] \tag{4}$$

и уравнение Ньютона, связывающее центростремительную силу и ускорение

$$\frac{mv^2}{R} = evH \tag{5}$$

(R – радиус окружности, по которой будет двигаться ион M⁺ в магнитном поле). Из выражений (3) и (5) получается соотношение, связывающее радиус траектории с отношением масса/заряд:

$$(m/e) = \frac{H^2 r^2}{2U} \tag{6}$$

Таким образом, измеряя R, можно судить о величине m. Достоверность результатов косвенных измерений решающим образом зависит от надежности заложенных предположений, в данном случае соотношения (3)-(5) сомнения не вызывают.

В качестве второго примера косвенных измерений рассмотрим метод определения межъядерного расстояния в двухатомной молекуле на основании анализа вращательных оптических спектров поглощения. Энергия E_j вращательного уровня, характеризуемого вращательным квантовым числом j равна для жесткого ротатора

$$E_j = \frac{\hbar^2}{2J}j(j+1) \tag{7}$$

 $(J - момент инерции, выражаемый через приведенную массу молекулы m* и межъядерное расстояние r: J = m*r²). На основании (7) нетрудно убедиться, что чисто вращательный спектр, отвечающий переходам между соседними уровнями (<math>\Delta j = \pm 1$) должен представлять собой систему равноотстоящих линий в частотном представлении. Измерение разности частот Δv между соседними линиями позволяет измерить момент инерции и, тем самым, величину г. На практике, однако, может оказаться, что линии в спектре неэквидистантны. Это может означать, что модель жесткого ротатора (7) не применима, и необходимо учитывать увеличение г (растяжение молекулы) при сильном вращательном возбуждении.

Вернемся к утверждению о том, что в процессе измерения всегда присутствует эталон, в случае прямого измерения его присутствие очевидно (рулетка в руках), в случае косвенных измерений он присутствует не столь явным образом. Использование рулетки в качестве эталона оправдано тем, что в процессе своего изготовления она сравнивалась с заданной степенью точности с другим эталоном более высокого класса, который, в свою очередь, калибровался по более совершенному образцу. Существует стандартизованная государством иерархия эталонов, наверху которой находятся наиболее точные Национальные эталоны, воспроизводящие принятые в данное время первичные Международные эталоны.

Необходимость единых международных эталонов была осознана научным и промышленным сообществом в конце 18 века, в России проблемами стандартизации измерений активно занимался Д.И. Менделеев. Первичные эталоны существенно менялись на протяжении 200 лет, работа по выработке эталонов продолжается и сейчас.

Можно указать две важные тенденции в современной метрологии. Первая состоит в децентрализации эталонов, вторая – стремление воспроизводить физические величины через фундаментальные мировые константы, такие как скорость света, постоянная Планка и т.п. Децентрализация означает, что эталон не воспроизводит какое-то уникальное изделие или явление, но может быть воссоздан в любом месте. Пример децентрализации – отказ от метра как от расстояния между штрихами на Парижской линейке и выражение его в длинах волн излучения Cd при заданных условиях.



$$L_{pl} = \sqrt{\frac{G\hbar}{c^3}} = 1.6 \cdot 10^{-33} \ cm$$
$$m_{pl} = \sqrt{\frac{c\hbar}{G}} = 2.2 \cdot 10^{-5} \ c$$
$$t_{pl} = \frac{L_{pl}}{c} = 0.5 \cdot 10^{-43} \ c$$

Вторая тенденция заключается в стремлении воспроизводить эталонные единицы через фундаментальные мировые константы, эта область исследований получила название

квантовой метрологии. В теории известны такие комбинации гравитационной постоянной G, константы Планка h, скорости света c, которые имеют размерности длины L, массы m, времени t, соответствующие комбинации известны как «планковские». В таблице 2 приведены выражения и численные значения некоторых планковских величин. Задачей ученого-метролога является нахождение таких экспериментальных условий, в которых интересующая эталонная величина воспроизводилась бы через мировые константы. Примером квантово-метрологического подхода является создание эталона вальта (Puc.1)



Рис. 1. Эталон вольта на основе джозефсоновского перехода.

Джосефсоновский контакт представляет собой два сверхпроводника (Pb), разделенных туннельно тонким (толщина ~10Å) слоем диэлектрика (PbO). К туннельному переходу прикладывается переменное напряжение с частотой ω , между частотой переменного напряжения и постоянной разностью потенциалов V на контакте существует зависимость

$$n\hbar\omega = 2eV; \tag{8}$$

Сканирование частоты ω приводит к возникновению на вольт-амперной характеристике скачков потенциала, на которых значение V составляет

$$V = n(\frac{\hbar\omega}{2e}) \tag{9}$$

(n –целое квантовое число) Связь между частотой переменного поля и величиной постоянного напряжения на контакте задается постоянной

$$\frac{2e}{\hbar} = 4.83 \cdot 10^{14} \ \Gamma \mu / B \quad ,$$

выражаемой через универсальные мировые константы. Относительная погрешность воспроизведения вольта этим эталоном составляет 2·10⁻⁹.

Наиболее точным из используемых в настоящее время эталонов является квантовый эталон частоты, основанный на переходах между состояниями с различным значением орбитального квантового числа в атоме ¹³³Cs (относительная погрешность ~10⁻¹⁴). Погрешности воспроизведения упомянутых квантовых эталонов разительно отличаются от погрешности воспроизведения килограмма (2·10⁻⁶), в качестве которого до сих пор используется «рукотворная» гиря.

Всего к настоящему времени стандартизовано около 130 эталонов, воспроизводящих различные физические величины.

Общие для всех физических измерений составляющие представлены на Рис. 2 в виде т.н. «скелетной схемы измерения». Рассмотрим общую схему на примере исследования спектра излучения молекул азота, возбуждаемых в тлеющем разряде. Объектом исследования является газообразный азот, возбуждаемый электрическим разрядом (некоторые объекты могут не требовать специального возбуждения). Выходной сигнал – интенсивность излучения на определенной длине волны I_{λ} - преобразуется с помощью фотоэлектронного умножителя (ФЭУ) в электрический сигнал, величина

которого измеряется амперметром. Результат измерения отображается индикатором (в данном случае – монитором) для визуального восприятия, параллельно с индикацией происходит регистрация результатов на бумаге или на магнитном диске для хранения и последующей обработки. Преобразование выходного сигнала объекта в электрическую форму используется в большинстве современных измерительных устройств.



Рис. 2. Принципиальная схема измерения на примере исследования спектра излучения разряда в $N_{\rm 2}.$

Методы измерений.

Рассматривая различные по физическому содержанию методы измерений, можно выделить три основных подхода к их проведению – метод отклонений, разностный и нулевой метод измерений.

1. Метод отклонений состоит в том, что измеряется вся величина. Например, с помощью пружинных весов измеряется вес груза целиком, специальной рейкой измеряется полный рост человека рулеткой и т.п.

2. Разностным методом измеряется отклонение интересующей нас величины от какого-либо стандарта. Например, коромысловые весы показывают отличие веса груза от веса гири при неполном уравновешивании системы (рис. 3).



Рис. 3. Разностный метод измерений.

Разностный метод используется в схеме двухлучевой спектрофотометрии, когда спектр раствора исследуемого соединения, помещенного в кювету, регистрируется относительно точно такой же кюветы, заполненной растворителем без вещества. В ряде случаев это помогает исправить спектры поглощения, исключив смещение нулевой линии (показано пунктиром на рис. 3), вызванное рассеянием света

Разностный метод лежит в основе исследования периодических процессов или структур путем сравнения их с периодическими эталонами. Известно, что при сложении двух звуковых волн с различными периодами возникают биения, период которых Т зависит от разности частот звуковых волн:

$$T = \frac{2\pi}{\Delta\omega} \tag{10}$$

При наложении периодических пространственных структур со слабо различающимися периодами, возникают пространственные биения, называемые муаровым узором (рис. 4)



Рис. 4. Возникновение картины муара при наложении двух пространственных периодических структур.

Пространственный период L в этом случае выражается аналогично (10):

$$L = \frac{2\pi}{\Delta k}; \qquad (k = \frac{2\pi}{\lambda}) \tag{11}$$

k носит название волнового числа. Трехмерным вариантом разностного метода является рентген-структурный анализ, когда периодическая структура кристалла сравнивается с электромагнитной волной в качестве второго периодического объекта.

Одним из главных достоинств разностного метода измерений является перенос масштабов, на которых производятся измерения. В соответствии с (10) и (11) возникающие масштабы Т и L тем больше, чем меньше различия в периодах исследуемой и эталонной структурой. Пространственный масштаб L дифракционной картины, возникающий при рассеянии рентгеновских лучей на кристалле с периодичностью ~1Å, составляет несколько миллиметров.

3. Нулевой метод измерения состоит в полной компенсации измеряемой величины путем приложения внешнего воздействия. Регистрируется при этом не сама величина, а факт отсутствия сигнала. Классическим примером нулевого метода является использование моста Уитстона для измерения электрических величин (Рис. 5). Примером нулевого метода в механических измерениях служит упругая мембрана, деформация которой измеряется по изменению емкости конденсатора, одной из пластин которого

мембрана является. Конденсатор включают в резонансный контур и подстройкой переменного конденсатора добиваются возвращения контура в резонанс.



Рис. 5. Нулевой метод измерений: конденсаторный метод измерения механических смещений, яркостный пирометр, мост Уитстона.

Пример оптических нулевых измерений – определение температуры светящегося тела с помощью яркостного пирометра. Яркость объекта сравнивают с яркостью нагреваемой током нити, совпадение яркостей («исчезновение» нити на фоне изображения) свидетельствует о равенстве температур.

Принято считать, что нулевой метод обеспечивает наивысшую точность измерений. Например, в классическом мостике Уитстона мы можем использовать очень чувствительный нуль-гальванометр для того, чтобы зафиксировать факт отсутствия тока. Кроме того, полная компенсация устраняет погрешность, обусловленную нагревом резисторов протекающим током.

Основные характеристики измерительных систем.

Важнейшими метрологическими характеристиками любой измерительной системы являются чувствительность, разрешающая способность, динамический диапазон, пространственное разрешение, время отклика. Обсудим содержание этих понятий. Предварительно условимся обозначать сигнал на входе в измерительное устройство х, выходной сигнал у.

Чувствительностью системы S называют отношение

$$S = \frac{y}{x} \quad S_{dif} = \frac{dy}{dx} \tag{12}$$

Поскольку часто чувствительность непостоянна в пределах возможного диапазона измеряемой величины, используется понятие дифференциальной чувствительности $S_{dif} = S(x)$. Например, при подаче напряжения x=40 мВ на отклоняющие пластины осциллографа наблюдается смещение луча на экране на y=8см. Чувствительность осциллографа в этом случае S=8/40=0,2 см/мВ, на практике не менее часто используют обратную величину 5 мВ/см.

Разрешающей способностью называют отношение

$$R = \frac{x}{\delta x} \tag{13}$$

δх – минимальное изменение входного сигнала, которое может быть обнаружено. Важно подчеркнуть, что значение R не является абсолютной характеристикой данного метода

измерений. Если предположить, что бх является константой во всем диапазоне измеряемых значений x, то разрешающая способность возрастает в области больших x.

Динамический диапазон D

$$D = \frac{x_{\max}}{x_{\min}} \to \frac{x_{\max}}{\delta x}$$
(14)

определяет ширину области, в которой сигналы могут быть измерены. Минимальный сигнал по сути дела совпадает с величиной минимально обнаружимого сигнала δх, что подчеркнуто в соотношении (14). Максимальный входной сигнал x_{max} лимитируется обычно не разрушением измерительного устройства, а величиной нелинейных искажений, возникающих, как правило, в случае больших сигналов на входе.

Важной характеристикой является пространственное разрешение метода измерений. При любых измерениях входной сигнал является величиной, усредненной по некоторому объему, поверхности или телесному углу. Например, измерение температуры ртутным термометром дает результат, усредненный по объему порядка объема концевого резервуара с ртутью. При измерении интенсивности свечения происходит усреднение измеряемой интенсивности в пределах телесного угла, вырезаемого входным окошком устройства и т.д. Пространственное dV или поверхностное dS разрешение можно ввести с помощью следующего соотношения:

$$\delta x = \int_{S} x_{S} dS = \int_{V} x_{V} dV \tag{15}$$

Соотношение (15) не следует понимать так, что пространственное разрешение всегда ограничивается величиной входного сигнала, в ряде случаев предел пространственному разрешению ставят технические возможности изготовления датчика требуемого размера.

Любое физическое измерение требует конечного времени, поскольку в процессе измерения объект и измерительное устройство обмениваются энергией. Для измерительной системы вводят понятие времени отклика т. Символически понятие времени отклика можно ввести соотношением (16):

$$\delta x = \int x_t dt \tag{16}$$

Констатация того, что проведение измерений требует конечного времени, кажется тривиальной. Но внимательный анализ временного аспекта измерений оказывается физически весьма содержательным и требует обстоятельного обсуждения.

Обратим внимание на то, что во все соотношения (12)-(16) входит величина δx , которую называют порогом обнаружения. Порог обнаружения – это минимальный входной сигнал, который может быть обнаружен с заданной степенью достоверности. Для создания «хорошей» во всех отношениях измерительной системы нужно устремить $\delta x \rightarrow 0$, но это стремление вступает в противоречие с достоверностью обнаружения сигнала. Надежность регистрации сигнала ограничивают погрешности измерений.

Погрешности измерений.

Основные виды экспериментальных погрешностей суммированы в Таблице 3.

Погрешности принято делить на случайные и систематические. Систематическая погрешность обусловлена влиянием измерительного устройства на измеряемую величину. Как пример, рассмотрим измерение температуры с помощью термопары (рис. 6). Разность температур между левым и правым контактами термопары создает ЭДС величиной V_0 , которая измеряется вольтметром с входными сопротивлением R_i , сопротивление проводников обозначим R_0 . Из рассмотрения эквивалентной электрической схемы получаем соотношение между V_0 и разностью потенциалов на входе вольтметра V_i :



Таблица 3. Погрешности измерений

$$V_i = V_0 \frac{R_i}{R_0 + R_i} \tag{17}$$

Относительная погрешность тем меньше, чем выше входное сопротивление измерительного вольтметра. Особенностью систематических погрешностей является то, что величина их может быть оценена (в данном случае – из соотношения R_i и R_0) и при необходимости результат может быть скорректирован на величину систематической ошибки.





Источники случайной погрешности измерений можно разделить на шумы и помехи. Принципиальное отличие между ними состоит в том, что шумы, в отличие от помех, являются неотъемлемой составляющей процесса измерения и полностью устранены быть не могут, тогда как принципиального запрета на полное устранение помех не существует. Подробнее о шумах поговорим позднее, а вначале рассмотрим источники помех и возможные способы минимизации их влияния. Если ограничиться только электрическими сигналами, то можно выделить три основных источника помех (см. табл. 3): космические, атмосферные, промышленные.

Космические помехи создаются потоками галактического излучения, солнечной активностью, порождающей потока заряженных частиц.

Источниками атмосферных помех являются движением паров воды, аэрозолей, выпадение осадков, образование инея и многое другое. Земная атмосфера насыщена электричеством, Земля несет электрический заряд с плотностью ~ 10^{-9} K/м², напряженность электрического поля вблизи поверхности земли составляет ~100 B/м. В спокойной атмосфере протекают токи ~ 10^{-16} A/см², при выпадении снега или дождя плотность тока увеличивается до $10^{-11} - 10^{-10}$ A/см², при ливнях и граде наблюдаются токи ~ 10^{-8} A/см²; ток в молнии может достигать 500 кA.

Промышленные помехи вызваны работой мощных моторов, сварочных аппаратов, радиостанций и других электрических устройств большой мощности.

Организационные меры устранения помех могут состоять в изменении места и времени проведения измерений. Например, промышленные помехи, как правило, значительно уменьшаются в ночное время, состояние атмосферы ночью обычно тоже более спокойно. Высокочастотные помехи можно снизить, выбрав, если возможно, частотный диапазон, не используемый для радиовещания и мобильной связи. Большие города являются источниками разнообразных по своей природе помех – электромагнитных, сейсмических, световых, акустических и др.

Рассмотрим технические приемы защиты от помех. Влияние помехи на процесс измерения в общем виде можно представить в виде эквивалентной электрической схемы, включающей два связанных контура (рис. 7). Контур 1 является измерительным, с ним через импеданс Z_{12} связан контур 2, процессы в котором являются для нас неконтролируемыми. В качестве меры связанности контуров вводится коэффициент передачи мощности k_{21} из контура 2 в контур 1

$$k_{21} = \frac{P_{12}}{P_1} \tag{18}$$

где P₁₂ и P₁ – мощность, выделяющаяся на импедансах Z₁₂ и Z₁ соответственно.



Рис.7. Связанные электрические контуры как причина возникновения помех.

В зависимости от характера импеданса Z_{12} можно выделить три «чистых» типа связи: резистивный ($Z_{12} = R$), емкостной ($Z_{12} = 1/j\omega C$), индуктивный ($Z_{12} = j\omega L$). Рассмотрим эти три случая по отдельности, но при этом будем иметь в виду, что в реальных условиях вклад в помеху может создаваться любой комбинацией этих составляющих.

Резистивный тип связи представляет собой токи утечки, обусловленные несовершенством изоляции. Цепи с высоким значением импеданса в большей степени подвержены влиянию токов утечки. Увеличению токов утечки способствуют

неблагоприятные внешние условия: высокая влажность, химически агрессивная среда, наличие поверхностных загрязнений. Например, сопротивление, измеренное между двумя точками на поверхности печатной платы, находящимися на расстоянии 1 см, становится со временем равным 10^8 Ом. Диэлектрические материалы подвержены процессу старения, поэтому качество изоляции требует периодического контроля. Наряду с пассивными мерами, которые сводятся к поддержанию высокого качества изоляции, для уменьшения токов утечки применяется активная защита (рис. 8).

Рис. 8. Активное подавление токов утечки в измерительном кабеле.

Активная защита состоит в том, что токоведущая жила окружается экраном, на который через операционный усилитель подается потенциал. незначительно отличающийся от потенциала измерительного проводника V_i. Пусть потенциал охранного экрана составляет $V_g = V_i(1-\varepsilon)$, причем $\varepsilon \ll 1$. Если обозначить Z_i импеданс утечки между токоведущей жилой и заземленным проводником линии при отсутствии защитного экрана, то включение активной защиты приведет к тому, что эффективный импеданс утечки станет равным $Z_l^* = Z_l/\varepsilon$. Аналогичный прием используется и на печатных платах, в этом случае роль защитного экрана выполняют две проводящие дорожки, расположенные по обеим сторонам от токоведущей.

Условия возникновения емкостной наводки и эквивалентная схема связанных контуров для этого случая представлены на рис. 9.

Рис.9. Емкостная наводка и эквивалентная схема связи контуров.

Обозначим V_0 и Z_0 измеряемый сигнал и импеданс объекта измерения, Z_i - входной импеданс измерителя, V_d напряжение источника помехи, C_p - паразитная емкость, связывающая измерительный контур с источником помехи. Напряжение V_i , создаваемое помехой V_d на входе в измеритель равно:

$$V_{i} = \frac{V_{d} \cdot (Z_{i} || Z_{o})}{Z_{p} + (Z_{i} || Z_{o})} = V_{d} \cdot \frac{1}{1 + \frac{Z_{p}}{Z_{i} || Z_{o}}}$$
(19)

Несложно убедиться в том, что условие $V_i \rightarrow 0$ достигается при $Z_p \rightarrow \infty$ ($C_p \rightarrow 0$), либо при $Z_i \rightarrow 0$. Другими словами, нужно удалиться от источника помехи и по возможности использовать измеритель с меньшим входным импедансом. Хорошей защитой от емкостной наводки служит заземленный экран, охватывающий объект исследования и измеритель.

Условия, при которых в измерительном контуре может появиться индуктивная помеха, иллюстрирует рис. 10.

Рис. 10. Индуктивная помеха, наводимая контуром с переменным током I(t).

Величина наведенного сигнала помехи V_d пропорциональна скорости изменения магнитного потока Ф, пронизывающего измерительный контур:

$$V_d = -\frac{d\Phi}{dt} = -\frac{d}{dt} \int_A H \, dA \tag{20}$$

Напряжение наводки на входных клеммах измерителя составит:

$$V_{id} = V_d \frac{Z_i}{Z_i + Z_0} \tag{21}$$

«Полезный сигнал», т.е. напряжение, создаваемое источником V₀ на входных клеммах, равно:

$$V_i = V_d \frac{Z_i}{Z_i + Z_0} \tag{22}$$

Из сопоставления (22) и (21) видно, что в случае индуктивной помехи отношение «сигнал/шум» нельзя улучшить изменением импедансов измерителя и источника.

Пути уменьшения помехи подсказывает анализ соотношения (20). Можно уменьшить напряженность магнитного поля H(t) в контуре, для этого измерительный необходимо удалить от источника наводки. Иногда эффективным оказывается всего лишь разворот измерительных проводов, т.к. при этом уменьшается «видимая» площадь контура. Если скрутить измерительные провода в тугую косичку, то достигается двоякая цель: максимально уменьшается площадь A, пронизываемая магнитным потоком, кроме того, на одном полном витке косички нормаль к площади контура также совершает полный оборот, что приводит к изменению направления магнитного потока относительно контура и компенсации наводимой в витке ЭДС. Следует устранить вибрации измерительного контура, т.к. изменение площади контура dA/dt дает свой вклад в изменение магнитного потока. Если мягкие меры – подвигать, скрутить и закрепить провода – оказались безуспешными, то приходится использовать экранирование всей измерительной схемы и объекта. Причем, в отличие от электростатической наводки, в этом случае необходимо использовать толстые экраны из ферромагнитных материалов.

Бывают случаи, когда помеха занимает узкий диапазон частот, например, создается сетевым током с частотой 50 Гц. В этом случае относительно простым способом устранения наводки может оказаться включение в измерительную цепь соответствующих фильтров.

Рис. 11. Возникновение термо-ЭДС в измерительной системе.

Кроме неконтролируемой связи контуров, источником помехи может оказаться наличие в цепи нежелательного источника ЭДС, например, термоэлектрической природы. Если два спая разнородных проводников находятся при разной температуре, то между ними возникает разность потенциалов ΔV , пропорциональная разности температур: $\Delta V = \alpha \Delta T$. Этот эффект используется в термопарах для измерения температуры, но он же может служить источником помех, как показано на рис. 11.

Величина термоэлектрического коэффициента для пары Cu/Pb-Sn составляет $\alpha = 3$ мкB/K, в случае контакта металла с полупроводником величина α может быть существенно больше, например, $\alpha = 1000$ мкB/K для пары Cu/CuO. Данные примеры выбраны из-за высокой вероятности их присутствия, поскольку сплав Pb-Sn обычно используется при пайке медных проводов, а окисная пленка присутствует на поверхности медных контактов.

Завершая тему помех, нужно сказать несколько слов о заземлении.

Измерительные цепи всегда имеют заземление, поскольку заземление позволяет снизить помехи и повышает безопасность работы. Перечислим кратко основные требования, которые должны соблюдаться при заземлении.

1. В измерительном контуре должна быть *одна* точка заземления. Если точек заземления несколько, то в образующихся петлях возможно наведение индуктивной помехи. Кроме того, по земляной шине нередко протекают значительные токи (см. рис. 12), в этом случае потенциал «земли» в разных точках измерительной цепи будет различным. Заземление в одной точке в ряде случаев ведет к удлинению заземляющих

Рис. 12 Заземление измерительной схемы.

проводников, поэтому сечение их должно быть достаточно большим, чтобы соответствовать требованиям техники безопасности (максимальное сопротивление заземляющего провода не более 7 Ом).

2. Экраны следует заземлять на том конце, который соединен с минимальным импедансом (у объекта измерения, если измеряется потенциал, у измерителя, если измерения проводят в токовом режиме).

3. Экраны не должны иметь щелей, зазоров, должны отсутствовать вибрации, в особенности, если по цепи передаются высокочастотные сигналы. С точки зрения целостности экрана наиболее уязвимы места соединения кабеля с клеммами и разъемами.

Надо отметить, что устранение наводок и других помех является трудной экспериментальной задачей, т.к. отсутствует строгий алгоритм, гарантирующий успешный результат; помимо знания физической сути явлений, большое значение имеют практический опыт и интуиция.

Время измерения.

Прежде, чем перейти к рассмотрению шумов как источнику случайных ошибок измерения (см. Табл. 3), необходимо подробнее остановиться на временных характеристиках измерения. Любое измерение включает в себя процесс обмена энергией между объектом и измерительной системой и требует конечного времени, характерное время измерения называют временем отклика измерительной системы.

Вернемся к примеру с измерением температуры с помощью термопары. Зависимость от времени напряжения на входе вольтметра V_i приведена на рис. 13. (V_0 – термоЭДС).

Рис. 13. Эквивалентная схема измерения температуры термопарой и зависимость показаний вольтметра от времени.

Более реалистичная (по сравнению с приведенной на рис. 6) эквивалентная схема измерительного процесса должна включать емкость С_i. Добавление емкости учитывает, во-первых, входную емкости вольтметра и, во-вторых, «тепловую емкость» термопары: после установления контакта с измеряемым объектом в термопаре происходит формирование некоторого стационарного распределения температуры вдоль проводников, которое можно представить как заполнение некоторой тепловой емкости тепловым потоком.

Покажем, что эквивалентная электрическая схема на рис. 13 позволяет описать кинетику установления потенциала $V_i(t)$). Обозначим через I_0 , I_R и I_C токи, протекающие через резисторы R_0 , R_i и конденсатор C соответственно.

$$I_{R} = I_{0} - I_{C}$$

$$q = CV_{i}$$

$$I_{C} = \frac{dq}{dt} = C\frac{dV_{i}}{dt}$$

$$V_{i} = I_{r}R_{i} = R_{i}(I_{0} - I_{C}) = V_{i\infty} - R_{i}C\frac{dV_{i}}{dt}$$

$$\frac{dV_{i}}{V_{i} - V_{i\infty}} = -\frac{dt}{R_{i}C}$$
(23)

 $V_{i\infty}$ - стационарное значение входного напряжения при t $\to\infty$. Решением последнего уравнения является экспонента, которая с учетом начальных данных имеет вид

$$V_{i} = V_{i\infty} (1 - e^{-t/\tau})$$
(24)

Характерное время нарастания сигнала равно $\tau = R_i C$. (В качестве самостоятельного упражнения предлагается убедиться, что произведение RC действительно имеет размерность времени). Полученная временная зависимость $V_i(t)$ представляет собой отклик измерительной системы на входной сигнал x в виде ступеньки $x = V_0 \cdot \Theta(t)$ и называется *переходной характеристикой*, время $\tau = R_i C$ является масштабом *времени отклика* измерительной системы.

Из этого примера видно, что сколько-нибудь разумная точность измерения достигается лишь в том случае, когда время измерения не меньше величины τ. Реально время измерения выбирается больше τ, и вопрос – насколько больше – это вопрос приемлемой точности измерения.

Выделим измерительный фрагмент схемы – RC-цепочку, включающую входные элементы R_i и C (рис. 14) и рассмотрим *чувствительность* такого измерителя S =y/x. Входным сигналом будем считать ток (x = I), выходным - падение напряжения на RC-цепочке (y = U) - см. рис. 14. Обозначая через Z импеданс RC-цепочки, получим:

Рис. 14. RC-цепочка как измеритель

$$U = ZI = I \frac{Z_R Z_C}{Z_R + Z_C}$$

$$S = \frac{R \cdot \frac{1}{j\omega C}}{R + \frac{1}{j\omega C}} = \frac{R}{1 + j\omega RC} = \frac{R(1 - j\omega\tau)}{1 + \omega^2 \tau^2}$$
(25)

Чувствительность $S(\omega)$ оказывается комплексной величиной. Один из способов задания комплексного числа состоит в использовании полярной системы координат: задается модуль числа (длина изображающего вектора) и его аргумент (полярный угол). Именно такой способ представления принят в теории передачи сигналов, при этом используется следующие определения. Модуль $S(\omega)$ носит название амплитудночастотной характеристики (АЧХ), аргумент $S(\omega)$ называется фазо-частотной характеристикой, в целом $S(\omega)$ представляет амплитудно-фазовую характеристику системы. Нас будет интересовать, прежде всего, АЧХ.

По определению АЧХ:

$$S(\omega) \models \sqrt{\operatorname{Re}^2 S + \operatorname{Im}^2 S}$$

Из соотношения (25) следует:

$$\operatorname{Re} S = \frac{R}{1 + \omega^{2} \tau^{2}} \qquad \operatorname{Im} S = -\frac{Rw\tau}{1 + \omega^{2} \tau^{2}} \qquad (\tau = \operatorname{RC})$$

Откуда для АЧХ получаем

$$|S(\omega)| = \frac{R}{\sqrt{1 + \omega^2 \tau^2}}$$
(26)

Вид зависимости $|S(\omega)|$ приведен на рис. 15

Рис.15.Амплитудно-частотная характеристика RC-цепочки.

Зависимость $|S(\omega)|$ показывает, что амплитуда выходного сигнала U монотонно уменьшается с ростом частоты входного тока (при неизменной амплитуде последнего). Уменьшение выходного напряжения U обусловлено шунтирующим влиянием конденсатора, импеданс которого в пределе $\omega \to \infty$ стремится к нулю. Зависимость $|S(\omega)|$ позволяет ввести понятие частотной *полосы пропускания* измерительной системы. Границей пропускания, по определению, считают частоту, при которой максимальное значение $|S(\omega)|$ уменьшается в $\sqrt{2}$ раз. Для RC-цепочки максимум расположен при $\omega = 0$, при этом |S(0)| = R (рис. 15), граничная частота ω^* , при которой $|S(\omega)| = R/\sqrt{2}$, равна $\omega^* = 1/RC$ и полоса пропускания в этом случае $\Delta \omega = \omega^* - 0 = \omega^*$ (заштрихована на рис. 15).

Определение, эквивалентное данному, состоит в ограничении полосы пропускания по уровню сигнала –3 дБ по отношению к максимальному. Напомним, что Бел – логарифмическая единица для сравнения относительной *мощности* двух сигналов:

$$\Pi(Бел) = lg(P_2/P_1)$$

Децибел равен 1/10 Бел, поэтому

$$\Pi(дБ) = 10 \cdot lg(P_2/P_1)$$

Поскольку мощность пропорциональна квадрату тока или напряжения, то при сравнении этих величин

$$\Pi(\mathbf{A}\mathbf{B}) = 10 \cdot \lg(V_2/V_1)^2 = 20 \cdot \lg(V_2/V_1)$$
$$\Pi(\mathbf{A}\mathbf{B}) = 10 \cdot \lg(I_2/I_1)^2 = 20 \cdot \lg(I_2/I_1)$$

В качестве самостоятельного упражнения предлагается показать, что определения полосы пропускания по уровню -3дБ и $1/\sqrt{2}$ по отношению к максимуму $|S(\omega)|$ эквивалентны.

Если дополнить RC-цепочку параллельно включенной индуктивностью, то получим параллельный колебательный контур, для которого AЧX имеет хорошо известный вид резонансной кривой (рис. 16). Пользуясь данным выше определением полосы пропускания, можем выделить диапазон частот Δω, «пропускаемых» контуром. В

случае колебательного контура полоса пропускания имеет как верхнюю, так и нижнюю ненулевую граничные частоты (шунтирование емкостью и индуктивностью соответственно).

Рис. 16. Схема и амплитудно-частотная характеристика параллельного колебательного контура

Порядок системы и переходные процессы.

Имеется непосредственная связь между переходной характеристикой измерительной системы (т.е. откликом на входной сигнал в виде ступеньки $x = \Theta(t)$), видом АЧХ и динамическим поведением в переходный период.

Измерительные системы принято относить к нулевому, первому или второму порядку в соответствии с порядком дифференциального уравнения, описывающим их поведение. Системы порядка 0 описываются алгебраическими уравнениями (нулевая производная по времени). Эквивалентная электрическая схема такой системы не включает реактивных элементов, отклик в точности повторяет временной ход входного сигнала, а это означает, что АЧХ постоянна в бесконечном диапазоне частот (одинаково хорошо пропускаются все частоты).

Переходная характеристика системы первого порядка, которую моделирует RCцепочка, представляет собой монотонную экспоненциальную функцию. АЧХ такой системы имеет максимум при $\omega = 0$ и монотонно уменьшается при $\omega \to \infty$.

Из курса физики известно, что поведение колебательного RLC-контура описывается уравнением второго порядка, решением которого являются затухающие колебания. АЧХ имеет вид колоколообразной кривой (1) на рис. 17. При сильном затухании (в этом случае говорят, что колебательная система «передемпфирована) поведение системы второго порядка практически не отличается от поведения системы первого порядка (кривые (2) на рис. 17). Величина затухания зависит от относительного вклада активной составляющей R в суммарный импеданс контура.

Рис. 17 Эквивалентная схема, функция отклика и АЧХ измерительных систем.

Следует отметить, что отнесение системы к тому или иному порядку не носит абсолютного характера, а зависит от условий эксперимента. Например, при обычных измерениях температуры воздуха жидкостным термометром поведение измерительной системы соответствует нулевому порядку. Однако, если измерения проводятся в среде, температура которой меняется быстро, то можно заметить, что термометр обладает инерционностью и для установления показаний требуется некоторое время, необходимое для переноса тепла через стенку капилляра. В этом случае более адекватным будет приближение первого порядка. Можно заключить, что нулевой порядок возникает в том случае, когда характерное время изменения измеряемого параметра значительно превышает время переходных процессов в измерительной системе.

Полоса пропускания и время измерения.

Пользуясь понятием полосы пропускания, измерительные системы можно разделить на две группы - широкополосные и узкополосные. Ширина полосы пропускания определяет избирательные свойства системы и ее временные характеристики. Под избирательностью системы понимают ее способность воспринимать сигналы в узком диапазоне частот. Настраивая радиоприемник, мы используем избирательные (резонансные) свойства колебательного контура. Чем шире полоса пропускания, тем, очевидно, менее избирательной является система. Для того, чтобы обсудить связь между избирательностью и временными характеристиками измерительной системы, напомним некоторые результаты из теории Фурье-преобразования функций.

Зависящий от времени сигнал f(t) с помощью преобразования Фурье (27) может быть представлен в виде функции $F(\omega)$, зависящей от частоты, которая называется спектром сигнала.

$$F(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt$$
(27)

Обратное преобразование Фурье (28) позволяет по спектру восстановить временную зависимость f(t).

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega) e^{i\omega t} d\omega$$
(28)

Периодические сигналы представляются в виде ряда Фурье и имеют дискретный спектр, непериодические сигналы представимы в виде интеграла Фурье, в этом случае $F(\omega)$ непрерывна. (В качестве упражнения ответьте на вопрос: являются ли процессы излучения на Солнце периодическими или непериодическими?). Из теории спектров известно, что характерная длительность процесса Δt и ширина его спектра $\Delta \omega$ связаны между собой соотношением:

$$\Delta \omega \cdot \Delta t \sim 1 \tag{29}$$

Приведенное соотношение можно рассматривать как своего рода соотношение неопределенностей, а умножение обеих частей на \hbar приводит его к знакомому виду $\Delta t \cdot \Delta E \sim \hbar$. Для иллюстрации соотношения (29) рассмотрим два предельных случая (рис. 18).

Рис. 18. Временные функции $y = \sin \omega t$ и $y = \delta(t)$ и их Фурье-спектры.

Синусоида бесконечна во времени и обладает предельно локализованным спектром в виде функции $\delta(\omega)$, противоположный пример – временной импульс в виде $\delta(t)$ имеет бесконечно широкий спектр. Отметим, что способность коротких временных импульсов, приближающихся к δ-функции, нести в себе широкий непрерывный набор частот активно используется в современных спектральных методах исследования (ИК-спектроскопия, ЯМР, масс-спекктроскопия).

На рис. 19 показано, что ограниченная временным интервалом (- τ , τ) «ступенька» обладает спектром, сосредоточенным на частотном интервале примерно $4\pi/\tau$. Спектр гауссовой кривой в целом сохраняет форму оригинала.

Рис. 19. Фурье-спектры конечных во времени функций.

Из соотношения между длительностью и спектральной шириной сигнала (29) вытекает, в частности, невозможность получения коротких монохроматических импульсов. Если считать, что достигаемая сегодня длительность фемтосекундных импульсов составляет по порядку величины 10^{-14} с, то энергия такого импульса «размазана» по частотному интервалу $\Delta \omega \sim 10^{14}$ или $\Delta f = \Delta \omega / 2\pi \sim 10^{13}$ Гц. Таким образом, одно из хрестоматийных достоинств лазерного излучения – высокая монохроматичность – может быть реализовано лишь для достаточно длительных импульсов или в непрерывном режиме генерации.

Принципиальный вывод из теории спектров в применении к измерительным системам состоит в том, что такие свойства как избирательность и время отклика являются противоречивыми. «Быстрые» системы, пригодные для разрешения малых временных интервалов Δt , по необходимости являются широкополосными, и наоборот, избирательные системы, характеризуемые узкой полосой $\Delta \omega$, будут обладать относительно большим временем отклика. Этот вывод иллюстрирует рис. 20.

Рис. 20. Переходные характеристики широкополосной (1) и узкополосной (2) измерительных систем.

Неэлектрические системы.

Важные характеристики измерительных систем, такие как время отклика, кинетический порядок системы, переходная характеристика, полоса пропускания обсуждались нами на примере электрических цепочек, и могло сложиться впечатление, что эти понятия применимы только к электрическим контурам. Это неверно. Электрические цепи выбраны исключительно из соображений удобства и привычности термина «импеданс» в применении к ним. Перечисленные понятия характеризуют свойства *любого* измерительного процесса, независимо от физической природы измерительной системы (механическая, тепловая и др.). Для иллюстрации этого утверждения воспользуемся вначале аналогией между механическими и электрическими величинами, замеченной Лапласом'

На рис. 21 слева приведены элементы электрической схемы и соответствующие соотношения, связывающие падение напряжения на каждом элементе с зарядом. Справа приведены идеализированные элементы, из которых может быть построена механическая система, и соотношения, связывающие силу и пространственную координату х. Механическим аналогом емкости конденсатора при таком рассмотрении является коэффициент упругости пружины, аналогом сопротивления является коэффициент динамической вязкости, аналогом индуктивности – масса.

Рис. 21. Аналогия между электрическими и механическими величинами

Таким образом, если произвести замены

$$U \leftrightarrow F$$

$$q \leftrightarrow x$$

$$1/C \leftrightarrow k$$

$$R \leftrightarrow \eta$$

$$L \leftrightarrow m$$

то все «электрические» соотношения можно дословно переписать на «механическом» языке, а для любой электрической цепочки построить эквивалентную механическую схему. Все рассуждения о переходных процессах, полосе пропускания, АЧХ и т.п., а

также о фазовых сдвигах между током и напряжением могут быть в точности повторены для механических систем.

Электромеханическая аналогия может быть расширена на системы, в которых протекают процессы различной физической природы. При рассмотрении электрических явлений возникают две характерные величины: падение потенциала U и ток I. Частное этих величин дает электрический импеданс, их произведение дает мгновенную мощность P (см. табл. 4). В механике мгновенная мощность представляет собой произведение силы на скорость перемещения, по аналогии, частное этих величин носит название механического импеданса. (Это вполне согласуется с соответствием q \sim x, поскольку и электрический ток, и механическая скорость являются производными по времени от соответствующих величин).

Подобные пары сопряженных величин обнаруживаются в процессах теплопередачи – разность температур ΔT и тепловой поток I_h , в процессах гидродинамики – разность давлений ΔP и объемный поток жидкости I_v . Используя эти величины, вводятся понятия теплового и гидродинамического импеданса (табл. 4). Примеры можно множить, но и приведенных достаточно, чтобы выявить закономерность.

В каждой паре одна из сопряженных величин является потоковой или «сквозной» величиной (through-value или I-value), такая величина может быть измерена в одной точке, при этом измеритель надо поместить в поток – включить «последовательно», как амперметр. Вторая величина в каждой паре является «продольной» или «градиентной» величиной (across-value или V-value), она измеряется как разность значений в двух точках, при этом измеритель подключается «параллельно» потоку, как вольтметр. Потоковые величины (I-value) являются производными по времени от некоторой физической величины ξ, а «градиентные» величины являются производными от энергии Е по ξ:

 $I = d\xi/dt;$ $V = dE/d\xi;$ $V \cdot I = dE/dt \equiv P$ (30) (физический смысл импеданса как отношения V/I объяснить несколько сложнее).

Ясно, что каждое физическое явление переноса порождает свои потоковые и «продольные» величины, являющиеся аналогами электрического тока и разности электрических потенциалов, а это означает, что понятия времени отклика, полосы пропускания, АЧХ, фазового сдвига и др. с необходимостью возникают при изучении динамики самых различных процессов.

Случайные величины.

Результат измерения из-за влияния случайных помех и шумов, становится случайной величиной. Вспомним основные понятия, которые понадобятся для описания экспериментальных данных, содержащих случайную составляющую. Полезно также сопоставить отличия в подходе теоретиков и экспериментаторов к описанию случайных величин – см. таблицу 5.

Таблица 5.

ТеорияЭкспериментЗадана
$$\rho(x)$$
 $\rho(x)$ неизвестна $\bar{x} = \int_{-\infty}^{+\infty} x \rho(x) dx$ $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ $D = \sigma_x^2 = \int_{-\infty}^{+\infty} (x - \bar{x})^2 \rho(x) dx$ $\sigma_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$

Случайная величина характеризуется, прежде всего, *средним значением*. В теории вероятностей для случайной величины х вводят функцию распределения плотности вероятности $\rho(x)$, смысл которой состоит в том, что вероятность P обнаружить случайную величину в пределах интервала (x₁, x₂) задается соотношением

 $D = \overline{(x - \overline{x})^2} = \overline{(x^2)} - (\overline{x})^2$

$$P = \int_{x_1}^{x_2} \rho(x) dx , \qquad (31)$$

функцию ρ(x) удобно выбирать нормированной, когда

$$\int_{-\infty}^{+\infty} \rho(x) dx = 1$$
(32)

В теории вероятностей среднее значение <u>x</u> является *математическим ожиданием* случайной величины x и вычисляется как первый момент функции распределения (см. табл. 5). Экспериментатору обычно функция распределения неизвестна (если $\rho(x)$ известна, проведение измерений теряет смысл), поэтому среднее <u>x</u> рассчитывается по результатам дискретных измерений и является в этом случае э*мпирическим средним*.

Рис. 22. Попадания в мишень двух стрелков.

Для характеристики случайной величины в большинстве случаев недостаточно только ее среднего значения. Предположим, проходят соревнования по стрельбе, результаты стрельбы двух спортсменов приведены на рис. 22.

В среднем пули каждого из стрелков попадали в самый центр мишени, но у одного все они легли в пределах «десятки» и он показал чемпионский результат, а второй вряд ли прошел даже отборочный тур. Очень важную дополнительную информацию о случайной величине дает ее рассеяние вокруг среднего значения. Величину рассеяния характеризует *дисперсия*. Дисперсией D называют среднее значение квадрата разности между случайной величиной и ее средним значением, соответствующие формулы для расчетов приведены в табл. 5. В теории дисперсию находят с использованием функции распределения, в эксперименте по результатам отдельных измерений вычисляется эмпирическая дисперсия.

Следует обратить внимание на то, что дисперсия имеет размерность квадрата размерности измеряемой величины. Таким образом, физический смысл дисперсии - это площадь, на которой происходит рассеяние. С одной стороны, это удобно, поскольку обычно интересна величина отклонения, но не его знак. С другой стороны, чтобы сравнивать величину отклонения со средним значением (т.е. оценивать *относительное отклонение*), необходимо охарактеризовать разброс величиной той же размерности, что и измеряемая величина. Поэтому на практике используется *среднеквадратическое отклонение* (синоним *стандартное отклонение*) о, которое есть квадратный корень из дисперсии:

$$\sigma_x = \sqrt{D} \tag{33}$$

Стандартное отклонение – наиболее часто употребляемая мера погрешности измерений, применяемая «по умолчанию».

Очень полезным для решения практических задач является соотношение, приведенное в нижней части табл. 5, которое строго доказывается в курсе теории вероятностей. Часто оказывается так, что среднее значение и дисперсию определить проще, чем среднее значение квадрата случайной величины, в этом случае используем соотношение:

$$\overline{(x^2)} = D + (\overline{x})^2 \tag{34}$$

Распределение Гаусса.

Часто оказывается, что функция Гаусса хорошо описывает плотность распределения случайной величины. *Распределение Гаусса*, иначе называемое *нормальным* распределением, задается соотношением:

$$\rho(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-\frac{(x-a)^2}{2\sigma^2}]$$
(35).

Плотность распределения (35) зависит от двух параметров: σ и *a*, это распределение является нормированным. Нормирующий множитель равен $\frac{1}{\sigma \sqrt{2\pi}}$, поскольку

 $\int_{-\infty}^{+\infty} \exp(-x^2) dx = \sqrt{\pi}$. Нормальное распределение имеет вид симметричной колоколообразной кривой (рис. 23), значение в максимуме:

$$\rho_C = \rho_{\max} = \frac{1}{\sigma\sqrt{2\pi}} \tag{36}.$$

Рис. 23. Распределение Гаусса.

Математический смысл параметров σ и *a*, задающих распределение Гаусса, состоит в следующем: *a* есть математическое ожидание ($\bar{x} = a$), σ - среднее квадратическое отклонение нормального распределения (доказательство можно найти в любом курсе теории вероятностей). Чем больше параметр σ , тем шире «колокол» и тем меньше значение ρ_{max} .

Значение функции в точках $x = a \pm \sigma$ (точки A и B на рис. 23)

$$\rho_A = \rho_B = \frac{\rho_{\text{max}}}{\sqrt{e}} = \frac{1}{\sigma\sqrt{2e\pi}} = 0,606 \cdot \rho_{\text{max}}, \qquad (37).$$

точки А и В являются точками перегиба.

Возникает вопрос, почему распределение Гаусса столь часто оказывается пригодным для описания случайных величин? Российским математиком Ляпуновым была доказана *центральная предельная теорема (ЦПТ)*, физический смысл которой сводится к следующему: если а) распределение некоторой случайной величины стационарно т.е. $\underline{x} \neq \underline{x}(t)$ и б) случайная величина подвержена влиянию большого числа взаимно независимых флуктуаций, влияние каждой из которых ничтожно, то эта величина будет распределена по Гауссу. Очевидно, что этим условиям отвечает очень большое число измеряемых физических величин.

Предположим, задачей является исследование содержания вредных примесей в атмосфере. Схема эксперимента в таком случае может выглядеть, например, как на рис.24.

Рис. 24. Оптический метод исследования содержания загрязнений в атмосфере.

Имеется источник излучения (например, лазер), луч от которого, проходя через анализируемый газ, отражается от зеркала и возвращается в приемник. В таком эксперименте источниками погрешностей являются: 1) турбулентность атмосферы, 2) рассеяние на аэрозольных частицах, 3) неоднородное распределение примесей, 4) механические вибрации источника и зеркала, 5) электрические флуктуация в приемнике и многое другое. Таким образом, есть все основания предполагать, что результаты измерений удовлетворяют ЦПТ и их распределение будет нормальным.

Кроме того, равновесные тепловые флуктуации, присутствующие в любой измерительной системе, также создают случайный разброс, который распределен по

Гауссу. Рассмотрим в качестве измерительного устройства пружинку с коэффициентом жесткости *ξ*.. Потенциальная энергия пружины W:

$$W = \frac{\xi x^2}{2} \tag{38}$$

(х – величина деформации). Пружинка находится в равновесии с окружающим тепловым резервуаром, следовательно, вероятность Р пребывания ее в состояниях с различной энергией W подчиняется распределению Больцмана:

$$P(W) \propto \exp(-\frac{W}{kT}) = \exp(-\frac{\xi x^2}{2kT})$$
 (39)

Соотношение (39) представляет собой (с точностью до нормировки) распределение Гаусса

для координаты x, параметры распределения: $\bar{x} = 0$; $\sigma^2 = \frac{kT}{\xi}$. Таким образом, координата x (отклонение от длины недеформированной пружины) будет флуктуировать

вблизи нуля и среднеквадратическая величина деформации пружины составит $\sigma_x = \sqrt{\frac{kT}{z}}$.

Рассмотренная ранее аналогия между механическими, электрическими и другими физическими величинами позволяет распространить полученный результат на флуктуации в различных системах.

Рисунок 25 иллюстрирует влияние параметров a и σ на вид гауссова распределения. Слева на рисунке показано влияние систематической погрешности на вид функции распределения. Систематическая погрешность изменяет среднее значение результатов измерений, что приводит к смещению кривой вдоль оси х на величину Δx . Справа приведен вид функции Гаусса при различных значениях параметра σ . Меньшие значения σ характеризуют меньший случайный разброс результатов, что означает более высокую точность измерений.

Рис. 25. Вид нормального распределения при различных значениях а и о.

Функция Гаусса наиболее часто используется для описания распределения случайных величин, но не является единственно возможной. В зависимости от методики проведения измерений, разумной аппроксимацией может оказаться другая функция. Обычно в качестве негауссовой формы распределения выбирают какую-нибудь простую геометрическую фигуру – прямоугольник, треугольник или трапецию, что позволяет легко вычислять средние значения и дисперсию (см. табл. 5). Пример равномерного на интервале («прямоугольного») распределения случайной величины будет рассмотрен при анализе шумов в цифровых устройствах.

ЦПТ формулирует достаточные условия того, что случайная величина имеет гауссово распределение, и эти условия реализуются во многих физических системах.

Однако, детальное изучение распределения экспериментальных результатов в целом ряде физических, химических и биохимических систем, проводимое в течение многих лет под руководством российского ученого С.Э. Шноля, показало негауссов, а в ряде случаев и нестационарный характер разбросов. Реальная действительность оказывается разнообразнее простых моделей. Общепринятого мнения по поводу этих результатов в настоящее время еще не сложилось, это область дискуссий, одна из «точек роста» современной науки.

Закон распространения ошибок

Пусть *z* – случайная величина, представляющая собой сумму двух независимых случайных величин: *z* = *x* + *y*, плотности распределения вероятности которых $\rho(x)$ и $\rho(y)$ – нормальны. Распределения $\rho(x)$ и $\rho(y)$ и $\rho(y)$ характеризуются средними значениями <u>x</u> = *a*, <u>y</u> = *b* и дисперсиями σ_x^2 и σ_y^2 соответственно. Можно показать, что в этом случае распределение $\rho(z)$ будет также гауссовым, его среднее значение и дисперсия определяются через параметры распределений $\rho(x)$ и $\rho(y)$:

$$\underline{z} = a + b, \tag{40}.$$

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2 \tag{41},$$

Соотношения (40) и (41) выражают правило сложения случайных погрешностей. Число слагаемых в (40) и (41), очевидно, может быть любым («любое» для экспериментатора означает заведомо конечное число). Отметим «векторный» характер выражения (41): суммарная погрешность ведет себя как модуль некоторого вектора, компонентами которого являются независимые погрешности, и число независимых погрешностей задает размерность «пространства ошибок».

Из правила (41) вытекает очень простое, но важное для практики следствие. Представим себе, что на вход измерительного устройства поступает электрический сигнал, напряжения случайные флуктуации которого характеризуется величиной среднеквадратического отклонения $\sigma_1 = 15$ мкВ, устройство вносит собственный шум, величина которого $\sigma_2 = 7$ мкВ, шумы приемника не зависят от шумов сигнала. Среднеквадратическое значение суммарного шума на выходе измерителя составит $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2} = \sqrt{225 + 49} \approx 16,5$ мкВ. Добавление к собственному шуму сигнала шумов измерительного устройства, которые вполне сравнимы по величине с шумом сигнала, приводит к увеличению суммарного шума всего на 10%. В этой ситуации практически бессмысленно было бы брать для измерений более совершенный измеритель, а, если стоит задача снижения шумов, то сосредоточиться на «очищении» сигнала. На практике крайне редко случается так, чтобы различные источники случайных ошибок давали одинаковый вклад в суммарную погрешность, отсюда правило: находить и минимизировать, в первую очередь, основной источник погрешностей.

Из соотношений (40) и (41) вытекает закон уменьшения относительной погрешности при многократном повторении однотипного измерения. Пусть N раз измеряется какая-то физическая величина x, среднее значение которой <u>x</u> = a, дисперсия при однократном измерении равна σ^2_1 . Относительная погрешность однократного измерения при этом составит $\varepsilon_1 = \sigma_1/a$. Проведение N независимых измерений эквивалентно измерению величины x, среднее значение a_N и дисперсия σ^2_N которой даются соотношениями:

$$a_{N} = a_{1} + a_{2} + \dots = Na$$

$$\sigma_{N}^{2} = \sigma_{1}^{2} + \sigma_{2}^{2} + \dots = N\sigma_{1}^{2}$$
(42)

Относительная погрешность серии однотипных измерений составит:

$$\varepsilon_N = \frac{\sigma_N}{a_N} = \frac{\sqrt{N\sigma_1}}{Na} = \frac{\varepsilon_1}{\sqrt{N}}$$
(43)

Обобщением правила сложения случайных погрешностей является соотношение, которое в литературе носит название закон распространения ошибок. Область применения этого закона – косвенные измерения. Пусть интересующая нас величина x связана известной функциональной зависимостью f с величинами p, q, r,..., которые подлежат измерению.

$$x = f(p,q,r,...)$$

Если бы случайные погрешности измерений отсутствовали, то *точное* значение x вычислялось бы после проведения измерения при фиксированных значениях аргументов p_0 , q_0 , r_0 . Наличие случайных погрешностей приводит к тому, что значения переменные фиксированы только *в среднем*, а величина x находится с погрешностью Δf

$$x = f(p, q.r, ...) + \Delta f \tag{44}$$

Приращение Δf в окрестности точки (p_0, q_0, r_0) находим по известному правилу разложения в ряд функции многих переменных. Ограничиваясь первым порядком разложения, получаем:

$$\Delta f = \frac{\partial f}{\partial p} \Delta p + \frac{\partial f}{\partial q} \Delta q + \frac{\partial f}{\partial r} \Delta r + \dots$$
(45)

(все производные берутся в точке (p_{θ} , q_{θ} , r_{θ})). Если принять, что масштаб отклонения от среднего значения задается среднеквадратическим отклонением $\Delta \xi = \sigma_{\xi}$, то

$$\sigma_x^2 = \left(\frac{\partial f}{\partial p}\right)^2 \sigma_p^2 + \left(\frac{\partial f}{\partial q}\right)^2 \sigma_q^2 + \left(\frac{\partial f}{\partial r}\right)^2 \sigma_r^2 + \dots$$
(46).

Соотношение (46) выражает закон распространения ошибок и является обобщением (41) на случай, когда разные источники случайных погрешностей имеют различную функциональную значимость.

В качестве примера косвенных измерений ранее был приведен метод массспектрометрии со статическим магнитным анализатором, позволяющий оценить отношение массы иона к его заряду. Используя закон распространения ошибок, можно оценить суммарную погрешность измерений $\delta(m/e)$, выполняемых данным методом. Применяя (46) к соотношению (6), получим:

$$(\delta(\frac{m}{e}))^{2} = (\frac{2rH^{2}}{2U})^{2}(\delta r)^{2} + (\frac{2r^{2}H}{2U})^{2}(\delta H)^{2} + (\frac{r^{2}H^{2}}{2U^{2}})^{2}(\delta U)^{2}$$
(47)

Выражение (47) можно использовать для оценки относительного вклада каждого из слагаемых и определения главного источника погрешности измерений. Как и в случае простого сложения погрешностей (41), правило первоочередной минимизации главного вклада остается в силе.

В качестве резюме сформулируем два важных правила экспериментальной работы:

1). Стремясь к точности, начинать с основного источника погрешности.

2). Измерения проводить не столь точно, как возможно, а столь точно, как необходимо.

функция ошибок.

Вопрос, с которым часто приходится сталкиваться, имея дело со случайными величинами, состоит в том, какова вероятность P(a,b). обнаружить случайную величину *x* в заданном интервале значений (a < x < b). Пусть функция $\rho(x)$ представляет плотность распределения вероятности случайной величины *x*, в этом случае:

$$P(a,b) = \frac{\int_{-\infty}^{b} \rho(x) dx}{\int_{-\infty}^{+\infty} \rho(x) dx}$$
(48).

Если $\rho(x)$ нормирована, то знаменатель (48) равен 1 и соотношение (48) упрощается:

$$P(a,b) = \int_{a}^{b} \rho(x) dx$$
(49)

Используем выражение (49) чтобы оценить вероятность того, что нормально распределенная случайная величина отличается от своего среднего значения на величину, не превышающую δ. Поскольку распределение Гаусса нормировано, то из (49) получаем:

$$P_{\delta} = P(\overline{x} - \delta, \overline{x} + \delta) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x-\delta}^{x+\delta} \exp\left[-\frac{(x-x)^2}{2\sigma^2}\right] dx$$
(50).

Используем стандартную замену переменных, которая задает величину σ в качестве масштаба переменной *x*

$$z = \frac{x - x}{\sigma};\tag{51}$$

Откуда следует $dx = \sigma dz$.

Соотношение (50) в новых переменных имеет вид

$$P_{\delta} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\delta/\sigma}^{+\delta/\sigma} \exp(-\frac{z^2}{2}) \,\sigma dz = \frac{1}{\sqrt{2\pi}} \int_{-\delta/\sigma}^{+\delta/\sigma} \exp(-\frac{z^2}{2}) \,dz = erf(\frac{\delta}{\sigma})$$
(52).

Интеграл (52) не выражается через элементарные функции, но может быть рассчитан. Выражение (52) определяет специальную функцию - «функцию ошибок». Вид функции erf(x) приведен на рис. 26 (ордината указана в процентах), там же даны значения функции для некоторых значений аргумента.

Рис. 26. Вид и таблица значений функции ошибок.

Функция erf(x) широко используется в теории вероятности и математической статистике, таблицу ее значений можно найти практически в каждой книге на эту тему. В качестве примера использования функции ошибок вернемся к примеру с флуктуациями длины пружинки. Было показано, что флуктуации длины пружинки имеют нормальное распределение, которое характеризуется величиной среднеквадратического отклонения

 $\sigma_x = \sqrt{\frac{kT}{\xi}}$. Предположим, что температура термостата комнатная, пружинка имеет длину

в нерастянутом состоянии 1,5 Å, а ее коэффициент жесткости равен 700 H/м (такими параметрами характеризуется «типичная» двухатомная молекула). Используя функцию ошибок, мы можем ответить на такой вопрос: какова доля молекул, отклонение которых от равновесной длины не превышает $\delta = 0.05$ Å? Найдем масштаб флуктуаций молекулярного осциллятора, задаваемый величиной σ :

$$\sigma_x = \sqrt{\frac{kT}{\xi}} = \sqrt{\frac{1.4 \cdot 10^{-23} \cdot 300}{700}} = 2.45 \cdot 10^{-12} \ m = 0.0245 \ A^{\circ}$$

Заданное нами отклонение, выраженное в единицах σ_x , составляет

$\xi = \delta/\sigma_x = 0.05/0.0245 \approx 2.$

Теперь поставленная задача формулируется так: какова вероятность обнаружить отклонение от среднего значения на величину 2σ для нормально распределенной величины? Ответ дает значение erf(ξ) \approx erf(2) = 0,9544. Для 95% молекул относительное удлинение не превышает $\varepsilon = 0,05/1,5 \approx 3\%$, столь слабая «раскачка» позволяет говорить, что при комнатной температуре молекулярные колебания практически не возбуждены.

Вид функции erf(x) (рис. 26) позволяет понять «эмпирическое правило 3σ », которое гласит, что отклонение случайной величины от среднего значения не превышает 3σ . По таблице находим значение erf(3) = 0,9973, а это означает, что вероятность обнаружить отклонение свыше 3σ для гауссовой случайной составляет всего 0,27% - событие, с практической точки зрения, весьма маловероятное.

Вернемся к данному ранее определению порога обнаружения измеряемой величины как *минимального сигнала, который может быть обнаружен с заданной степенью достоверности*. Пусть нам необходимо зарегистрировать сигнал на фоне случайного шума, распределение интенсивности которого является гауссовым. В отсутствие полезного сигнала на входе в измерительное устройство будет регистрироваться шум, распределенный вокруг нулевого среднего значения. Допустим, в какой-то момент отклонение от нуля сигнала на входе составило 1 σ . Вероятность обнаружить такое отклонение в отсутствие сигнала равна примерно 32% (см. табл. на рис. 26), а это означает, что с вероятностью 68% можно предполагать наличие полезного сигнала. Таким образом, если нас удовлетворяет 68%-ная достоверность, то порогом обнаружения будет величина сигнала, равная 1 σ от шумового распределения. Если требуется достоверность 95%, то, согласно таблице, порогом обнаружения была бы «назначена» величина 2 σ и т.д. Таблица значений erf(x) позволяет соотнести минимальный сигнал, вероятность его обнаружения и характеристики шума.

Шумы.

Шумы в качестве источника случайных ошибок измерения являются неотъемлемым атрибутом любой физической системы при конечной температуре, а также при отсутствии равновесия. Возвращаясь к табл. 3, суммирующей данные об источниках погрешности измерений, отметим, что по своей природе шумы делятся на *равновесные* и *неравновесные*.

Равновесные шумы обусловлены_флуктуациями термодинамических величин и являются следствием молекулярного движения. Величина равновесного шума зависит от температуры системы. Равновесный шум задает минимальный уровень шумов, который

достижим при данной температуре. Примерами равновесных шумов являются броуновское движение для механических систем (его аналог – флуктуации пружинки, рассмотренные ранее) и шум Найквиста (Nyquist) для электрических цепей.

При отсутствии в системе равновесия, в дополнение к равновесным шумам, возникают разнообразные *неравновесные шумы*, обусловленные флуктуациями потоков – дробовой шум в электрических цепях, пульсации турбулентном потоке, вибрации в механических системах.

Шумы, независимо от их физической природы, принято характеризовать спектральной плотностью мощности шума (СПМШ) - мощностью шумового сигнала P_N, приходящейся на единичный интервал частоты f:

$$N \equiv \frac{dP_N}{df}$$
(53)

СПМШ электрических шумов выражается в В²/Гц или А²/Гц.

СПМШ удобна тем, что позволяет связать мощность шума, например, на входе измерительного устройства, с полосой его пропускания. Пусть на входе присутствует шум, спектр мощности которого $N(\omega)$ приведен на рис. 29, а полоса пропускания измерителя составляет $\Delta\omega$.

Рис. 27. Связь мощности шума с шириной полосы пропускания.

Мощность шума на входе P_N в этом случае составит:

$$P_N = \int_{\Delta\omega} N(\omega) \, d\omega \tag{54}$$

Шум, для которого $N(\omega)$ не является константой, называют спектральнозависимым, если $N(\omega) \equiv \text{const}$, то шум является спектрально-независимым. В последнем случае шум иногда называют «белым», по аналогии с белым цветом, включающим все частоты видимого спектрального диапазона.

Примером *равновесного электрического шума* является шум Найквиста. Шум Найквиста обусловлен флуктуациями плотности заряда, которые создают флуктуирующую разность потенциалов. Если подключить вольтметр к резистору, как показано на рис. 28, то в отсутствие задаваемого внешним источником тока зависимость измеряемой разности потенциалов V от времени будет представлять собой случайную функцию, среднее значение которой равно нулю, а среднеквадратическое отклонение потенциала от среднего дается формулой Найквиста

$$\sigma_N = \sqrt{4RkT\Delta f} \tag{55}$$

где R - сопротивление резистора, k - постоянная Больцмана, T - температура, Δf - полоса частот, в которой происходит регистрация. Соотношение (55) применимо при T >> 0 для

частотной области $f \ll 10^{12}$ Гц (вывод соотношения (55) рассмотрим позже при обсуждении передачи сигналов по длинным линиям). При наличии источника ЭДС флуктуации потенциала будут наблюдаться относительно средней величины, задаваемой падением напряжения на резисторе.

Соотношение (55) означает, что средний квадрат флуктуаций напряжения на резисторе составит

$$U_N^2 = 4RkT\Delta f , \qquad (56)$$

откуда для СПМШ равновесного теплового шума получим:

$$N = \frac{dP_N}{df} = \frac{dU_N^2}{df} = 4RkT$$
(57)

СПМШ Найквиста представляет собой константу, зависящую только от величины сопротивления и температуры и, таким образом, равновесный шум является частотнонезависимым ("белым"). Интегральная мощность шума будет тем больше, чем более широкополосное устройство используется для измерений. Напомним, что широкой полосой пропускания характеризуются устройства с малым временем отклика. Это означает, что за уменьшение времени измерения платить приходится большим уровнем шума.

Равновесный тепловой шум возникает только в активных элементах электрического контура, в которых происходит диссипация энергии, диссипативные процессы играют роль «мостика», связывающего резистор с окружающим термостатом. В случае контура, содержащего наряду с активными и реактивные элементы, соотношение (55) следует записать в виде

$$\sigma_N = \sqrt{4kT\Delta f \cdot \operatorname{Re} Z} \tag{58}$$

где *Re Z* – действительная часть импеданса.

Оценим масштаб тепловых электрических флуктуаций при комнатной температуре для измерительного устройства, которое имеет входное сопротивление $R_{in} = 10^6$ Ом и время отклика $\tau \sim 1$ нс. Ширина полосы пропускания $\Delta f = 1/2\pi\tau \sim 10^8$ Гц,

$$\sigma_N = \sqrt{4 \cdot 1, 4 \cdot 10^{-23} \cdot 300 \cdot 10^8 \cdot 10^6} = 10^{-3} B = 1 \ \text{MB}$$

В электрических устройствах встречается целый ряд неравновесных шумов (см. рис.29), в частности, *дробовой шум*. Дробовой шум возникает в электровакуумных приборах и связан с флуктуационным характером переноса дискретных зарядов через активационный барьер. Для удаления электрона с поверхности катода необходимо преодолеть энергетический барьер, называемый *работой выхода электрона* из металла, преодоление барьера носит термоактивационный характер и сопровождается флуктуациями числа эмиттируемых электронов. Термин «дробовой шум» исходит из аналогии с шумом,

производимым падающими дробинками. Шотки (W. Schottky) было получено соотношение:

$$I^2{}_s = 2eI_0\Delta f \tag{59}$$

где I_0 – номинальное значение протекающего тока, e – заряд электрона. Равновесный тепловой

Рис. 29 Шумы в электрических цепях

Соотношение (59) применимо, если температура катода не слишком велика (A/kT >> 1, A - работа выхода электрона из катода) и рассматриваемые частоты удовлетворяют неравенству $f \ll 1/\tau$, где τ - время пролета электрона через рабочую область, например между электродами вакуумной лампы.

Дробовой шум является белым – из (59) видно, что СПМШ не зависит от частоты, на неравновесный характер шума указывает зависимость его интенсивности от неравновесного параметра – средней величины протекающего тока *I*₀.

Одним из основных видов шума в полупроводниковых устройствах является *генерационно-рекомбинационный шум*, в частности, этот шум ограничивает интенсивность измеряемых световых потоков в полупроводниковых фотодетекторах. Возникновение этого шума обусловлено механизмом электропроводимости полупроводников, включающем стадию генерации электронно-дырочной пары из нейтрального состояния M:

$$M \rightarrow p^+ + e$$

Образовавшиеся свободный электрон e и дырка p^+ дрейфуют под действием электрического поля в зоне проводимости и валентной зоне соответственно, обеспечивая протекание тока. Состояние с разделенными зарядами существуют в течение некоторого времени τ , после чего происходит гибель носителей p^+ и e в результате либо их рекомбинации, либо из-за попадания в ловушки – дефекты кристалла. Разброс времен жизни τ приводит к флуктуациям мгновенного числа носителей заряда и, следовательно, мгновенной величины тока.

Средний квадрата флуктуаций тока в случае генерационно-рекомбинационного шума определяется величиной протекающего тока I_0 и временем τ :

$$\overline{I_{G}^{2}} = \beta \frac{I_{0}^{2}}{1 + (\omega \tau)^{2}}$$
(60)

 $(\beta - \kappa o \circ \phi \phi u quent пропорциональности). Мощность генерационно-рекомбинационного шума зависит от частоты, СПМШ образует «ступеньку» в области частот <math>\omega \sim 1/\tau$. В полупроводнике параллельно могут протекать различные процессы, приводящие к гибели носителей заряда, каждый из которых характеризуется своим временем τ_i . В этом случае в спектре шума $I^2(\omega)$ обнаруживается каскад ступенек при соответствующих частотах ω_I (см. 29).

В электрических цепях нередко наблюдается *импульсный шум* как результат микропробоев в конденсаторах или на контактах. Этот шум выглядит как ряд случайных по времени появления, по длительности и по величине всплесков тока (рис. 29) С импульсным шумом невозможно связать какую-то определенную форму спектра, поскольку результат Фурье-преобразования случайного временного ряда может быть любым.

В 1925 году Джонсон (Johnson) при исследовании эмиссии с подогреваемых катодов обнаружил флуктуации, частотная зависимость которых выражается соотношением:

$$\overline{I^2} = \frac{A}{f^{\alpha}} \tag{61}$$

А – константа, $\alpha \sim 1$. Это явление получило название *фликкер--или 1/f-шума*. Последующие исследования на различных системах показали, что показатель α обычно заключен в пределах от 0,6 до 3, кроме того, не удалось обнаружить нижней частотной границы применимости соотношения (61). Во всяком случае, в области частот $f \sim 10^{-7}$ Гц запределивания мощности фликкер-шума не обнаружено. (Может быть, нелишне будет обратить внимание на то, что диапазон частот $\sim 10^{-7}$ Гц соответствует времени проведения эксперимента порядка одного года!)

Первоначальное объяснение природы фликкер-шума предполагало, что в твердом теле имеется широкий набор времен релаксации и наблюдаемый спектр шума представляет множества случаю собой суперпозицию «ступенек», подобно генерационнорекомбинационного шума. Однако позже низкочастотные шумы типа 1/f были обнаружены в системах самой различной природы, не только физических. Закономерность $\sim 1/f$ в геофизике носит название закона Гутенберга-Рихтера и связывает мощность и частоту землетрясений, в экономике аналогичное соотношение известно как закон Ципфа-Парето, связывающий количество предприятий и их капитал, в социологии зависимость типа 1/f обнаружена между числом городов и их населением. Флуктуации типа 1/f обнаружены в медицине для фаз сердечного ритма, в музыке, в структуре литературных произведений, в распределении галактик во Вселенной – примеры можно множить бесконечно. В статистике широко известно так называемое «эмпирическое правило 80/20», суть которого состоит в том, что 20% событий обеспечивают 80% результата (на бытовом уровне: «20% людей выпивают 80% пива»). Такой характер зависимости обеспечивает самоподобие фликкер шума на любых частотных интервалах, что позволяет отнести 1/f-шум к категории временных (частотных) фракталов.

Очевидно, говорить об универсальном физическом механизме возникновения фликкер-шума во всех системах невозможно и следует согласиться с высказыванием Ю.Л. Климонтовича: «Каждый исследователь вправе иметь свою теорию фликкер-шума». Можно, однако, предположить, что в ряде случаев имеется общая причина формирования 1/f-зависимости, которая состоит в корреляции последовательных флуктуаций. В неравновесных системах возможна ситуация, когда вероятность возникновения ϕ луктуации в момент времени *t* зависит от событий, которые происходили в момент *t*- τ - в этом случае флуктуации не являются статистически независимыми. В качестве простейшей иллюстрации рассмотрим эксперимент, в котором наблюдатель считает автомобили, проезжающие по дороге мимо него за определенный промежуток времени (реально проводимые подсчеты такого рода действительно обнаруживают фликкер-шум). Если на каком-то расстоянии от наблюдателя возникнет ситуация, затрудняющая движение автомобилей, то спустя некоторое время наблюдатель заметит, что частота следования автомобилей уменьшилась, но вслед за этим произошел всплеск количества проезжающих машин – когда волна сгущения, образовавшаяся перед препятствием, докатится до него. Таким образом, коррелированный характер неравновесных флуктуаций является в ряде случаев весьма вероятной причиной возникновения 1/f-шума.

Разнообразие неравновесных шумов в различных физических системах не поддается никакому описанию и даже перечислению. Следует отметить лишь два обстоятельства:

1) динамический шум присутствует в любой неравновесной системе, и по мере удаления от состояния равновесия роль флуктуаций в определении характера поведения системы возрастает;

2) шумы являются не только источником погрешности измерений, но в ряде случаев позволят получать уникальную информацию о протекающих процессах. Примерами являются изучение свойств ловушек заряда в полупроводниках на основании спектров генерационно-рекомбинационного шума, о котором шла речь выше, а также оценка состояния здоровья человека на основании шумов, прослушиваемых в сердце и бронхах.

Шумы цифровой записи

Широкое применение вычислительной техники в экспериментах делает актуальным обсуждение специфических шумов, возникающих при цифровой обработке сигналов.

Сигнал называют *аналоговым*, если он принимает непрерывный ряд значений в непрерывные моменты времени. *Цифровой* сигнал может принимать дискретный ряд значений в дискретные моменты времени. Дискретизация значений аналогового сигнала называется квантованием, о дискретизации во времени говорят как о взятии выборки (sampling). Для преобразования аналоговых сигналов, обычно имеющихся на выходе измерительных устройств, в цифровые, с которыми работают компьютеры, служат аналого-цифровые преобразователи (АЦП).

Почему преобразование сигнала из аналоговой в цифровую форму сопровождается возникновением шума? На рис. 30 показано «наложение» непрерывного аналогового сигнала на сетку значений, которые может принимать цифровой сигнал. (Будем для определенности считать, что измеряемый сигнал представляет собой напряжение). АЦП превращает гладкую функцию в ступенчатую, выделенную на рисунке жирной линий. Разность между точным значением непрерывного аналогового сигнала и его цифровым приближением представляет собой сигнал, который АЦП добавляет к исходному, добавляемый сигнал является по существу случайным шумом, определение характеристик которого (СПМШ) и является нашей задачей.

Рис. 30. Цифровое представление аналогового сигнала с помощью АЦП.

Будем предполагать, что вносимая погрешность обусловлена исключительно ошибкой квантования, а моменты взятия выборки задаются абсолютно точно. Такое предположение обосновано тем, что поддержание стабильности частоты АЦП – вопрос чисто технический, тогда как квантование принципиально сопряжено с ошибками.

Для кодирования значения сигнала в дискретный момент времени в АЦП используется слово длиной *n* бит. Это означает, что величина измеряемого сигнала V_i разбивается на 2ⁿ равных отрезка и минимальное фиксируемое приращение («квант») V_0 измеряемой величины, равно:

$$V_0 = \frac{V_i}{2^n} \tag{62}$$

Можно предположить, что ошибка квантования Q при каждом измерении составит величину порядка

$$|Q| \cong V_0 / 2 = \frac{V_i}{2^{n+1}}$$
(63)

Таким образом, мы рассматриваем Q как случайную переменную, среднее значение которой Q = 0 (положительные и отрицательные ошибки квантования в среднем компенсируют друг друга). Для вычисления дисперсии D_Q воспользуемся определением:

$$D = \sigma_Q^2 = \int_{-\infty}^{+\infty} (Q - \overline{Q})^2 \rho(Q) dQ$$
(64)

Чтобы использовать (64) для расчетов, необходимо знать функцию плотности распределения вероятности $\rho(Q)$. Вид функции неизвестен, но можно предположить, что

значения функции $\rho(Q)$ равномерно распределены на интервале $(-V_0/2, +V_0/2)$ – это означает, что сетка возможных значений цифрового сигнала случайным образом наложена на аналоговый сигнал и АЦП округляет текущее значение сигнала до ближайшей величины V_0 . Вид функции распределения $\rho(Q)$ приведен на рис. 31, значение функции $\rho(Q = 1/V_0)$ выбирается из условий нормировки.

С учетом $\overline{Q} = 0$ из соотношения (64) для дисперсии получаем:

$$D = \int_{-\infty}^{+\infty} Q^2 \rho(Q) dQ = \int_{-V_0/2}^{+V_0/2} Q^2 \rho(Q) dQ = \frac{Q^3}{3V_o} \Big|_{-V_0/2}^{+V_0/2} = \frac{V_0^2}{12}$$
(65)

Среднеквадратическое значение шумового сигнала

$$\sigma_{Q} = \frac{V_{0}}{2\sqrt{3}} = \frac{V_{i}}{2^{n} \cdot 2\sqrt{3}} = \frac{V_{i}}{2^{n+1} \cdot \sqrt{3}}$$
(66)

Рис. 31. Плотность функции распределения ошибки квантования АЦП.

Выборка измерений, сделанная в течение одной секунды с частотой f, включает N значений, N = 1/f. По правилу сложения независимых ошибок (41) для погрешности измерения входного сигнала $D(V_i)$ получаем:

$$D(V_i) = \sigma_{Q1}^2 + \sigma_{Q2}^2 + \dots + \sigma_{QN}^2 = N\sigma_Q^2$$
(67)

Подставляя выражение (66) для σ_Q единичного измерения и выражая число измерений через частоту, получаем:

$$D(V_i) = \frac{\sigma_Q^2}{f} = \frac{V_i^2}{3 \cdot 2^{2(n+1)} \cdot f}$$
(68)

Величина $D(V_i)$ в полученном выражении имеет размерность В²/Гц и представляет собой СПМШ преобразования аналогового сигнала в цифровой. Можно заключить, что мощность шума увеличивается с ростом величины входного сигнала и уменьшается при возрастании частоты взятия выборки и длины слов, используемых для цифровой записи.

Соотношение (68) показывает, что минимизация шума при цифровой записи сигнала может достигаться путем выбора двух переменных — частоты f и длины слова п кодирующей последовательности. Однако, существуют ограничения, которые не позволяют произвольно выбирать частоту взятия выборки f, но ставят ее в зависимость от спектральных свойств сигнала V_i .

В теории передачи информации доказпнп *теорема Котельникова-Шеннона*, которая утверждает, что сохранение полной информации о форме исходного аналогового сигнала возможно лишь в том случае, если частота взятия выборки f_s по крайней мере вдвое превышает f_{max} – максимальную частоту в спектре аналогового сигнала:

$$f_s \ge 2f_{\max} \tag{69}$$

Рисунок 32 иллюстрирует применение теоремы Котельникова-Шеннона. Предположим, в цифровую форму необходимо преобразовать сигнал f(t), его спектр $F(\omega)$ локализован на конечном частотном интервале $\Delta \omega$. Если сигнал имеет длительность τ , то соотношение $\Delta \omega \cdot \tau \cong 1$ определяет минимальную частоту ω_{\min} , максимальная частота ω_{\max} определяется характерными временами изменения f(t) на самых крутых фронтах сигнала. В соответствии с теоремой Котельникова-Шеннона для частоты взятия выборки получаем:

Рис. 32. Временной сигнал f(t) и его спектр $F(\omega)$.

$$f_s \ge \frac{2\omega_{\max}}{2\pi} \tag{70}.$$

Таким образом, если необходимо иметь полную информацию о форме сигнала, то выбор частоты диктуется условием (70) и минимизация шума при цифровом преобразовании может быть достигнута выбором длины кодирующей последовательности *n*. На практике часто приходится искать компромисс между техническими возможностями АЦП и его доступностью со всеми вытекающими последствиями в виде частичной потери информации о высокочастотных составляющих сигнала и возрастания шума.

Измерения при различном соотношении ширины полосы пропускания сигнала и измерительного устройства.

Условие, выполнение которого гарантирует отсутствие искажений при регистрации сигнала, связывает полосу пропускания измерительного $\Delta f_{u_{3M}}$ и спектральную ширину Δf_{cure} анализируемого сигнала: $\Delta f_{u_{3M}} \ge 2\Delta f_{cure}$. Рассмотрим некоторые возможности, которые возникают, если данное соотношение не выполняется.

Пусть $\Delta f_{u_{3M}} \ll \Delta f_{curh}$, в этом случае форму *строго периодического* сигнала можно исследовать методом *когерентных выборок*. Суть метода иллюстрирует рис. 33.

Анализируемый сигнал имеет период *T*; можно непосредственно изучать форму сигнала, беря выборки через достаточно малый интервал времени δ . Возьмем первый отсчет в момент t = 0, пользуясь строгой периодичностью сигнала, второй отсчет возьмем в момент $T + \delta$, следующий - в момент $2T + 2\delta$ и т.д. Данный прием ведет к расширению эффективной полосы пропускания измерительного устройства до величины $\Delta f_{3\phi}$:

$$\Delta f_{s\phi} = \Delta f_{usm} \cdot \frac{T+\delta}{\delta} \approx \Delta f_{usm} \cdot \frac{T}{\delta}$$
(71)

Рис. 33. Метод когерентных выборок.

Логическим развитием этого подхода будет взятие отсчетов не через один, а через N периодов, в этом случае эффективная ширина полосы пропускания увеличится в N раз:

$$\Delta f_{g\phi} = \Delta f_{ugm} \cdot \frac{NT}{\delta}$$
(72)

На этом принципе основано использование стробоскопического осциллографа. Насколько велико может быть значение N – зависит от степени периодичности сигнала. Например, стробоскопический осциллограф с полосой пропускания $\Delta f_{u_{3M}} = 20 \text{ к}\Gamma \text{ц}$ может использоваться для анализа быстрых процессов, характеризуемых шириной спектра $\Delta f_{cuch} = 15 \Gamma\Gamma \text{ц}$. В этом случае для величины N получаем оценку: N = $15 \cdot 10^9 / (20 \cdot 10^3) \sim 10^6$, т.е. каждый следующий отсчет берется примерно через миллион периодов, что отвечает эффективному расширению полосы измерителя в 10^6 раз. Очевидно, реализовать такой подход можно лишь для строго периодических сигналов.

При выполнении условия $\Delta f_{u_{3M}} >> \Delta f_{cu_{2H}}$ появляется возможность регистрировать одновременно несколько сигналов, используя метод частотного мультиплексирования. Применение метода частотного мультиплексирования иллюстрирует рис. 34. Рассмотрим три сигнала S₁(t), S₂(t) и S₃(t), характерные времена изменения которых равны τ_1 , τ_2 и τ_3 соответственно.

В соответствии с «принципом неопределенностей» (29) каждый сигнал занимает полосу частот $\Delta \omega_i \cong 1/\tau_i$. Для одновременной передачи сигналов можно использовать несущие частоты ω_1 , ω_2 и ω_3 , расстояния между которыми выбираются такими, чтобы полосы $\Delta \omega_i$ не перекрывались. Число сигналов, которые могут одновременно передаваться, зависит от соотношения между шириной полосы пропускания измерителя $\Delta \omega$ и шириной полосы $\Delta \omega_i$, необходимой для передачи отдельного сигнала (рис. 34). Параллельная передача сигналов с использованием частотного мультиплексирования широко используется в каналах передачи информации, например, в телефонных линиях.

Звуковые частоты, воспринимаемых ухом человека, занимают диапазон от 20 Гц до 20 кГц, удовлетворительное восприятие речи по телефону – возможность узнавать собеседника по голосу, воспринимать эмоциональные оттенки тембра – сохраняется при сужении полосы передаваемых частот до 5 кГц. Это означает, что если мы намерены по одному проводу передавать, например, 20 телефонных разговоров, то требуемая полоса пропускания должна составлять ~100 кГц. Для организации мобильной связи используется диапазон частот ~900 МГц, число каналов для голосовой связи может при этом достигать $n = 9 \cdot 10^8 / 5 \cdot 10^3 \sim 10^5$.

Рис. 34. Частотное мультиплексирование.

Особенно впечатляют в этом отношении возможности волоконно-оптических линий. В оптических волокнах для передачи сигналов служит излучение с длиной волны ~ 1 мкм, что соответствует частоте $3 \cdot 10^{14}$ Гц. Это означает, что одновременно по оптическому волокну может быть передано $\sim 10^{11}$ телефонных разговоров (можно сопоставить это число с населением Земли – $6,5 \cdot 10^9$ человек). Разумеется, по линиям связи передаются не только телефонные разговоры, но и значительно более широкополосные сигналы (например, телевизионные изображения), но общий вывод о высокой пропускной способности оптоволоконных линий остается в силе.

Управление полосой пропускания измерительного устройства может использоваться в ряде случаев для подавления шумов и помех. Рассмотрим два предельных случая: измеряемый сигнал почти периодический, регистрируемый сигнал почти постоянный.

В случае почти синусоидального сигнала с частотой ω эффективным с точки зрения подавления шумов оказывается максимальное сужение полосы пропускания детектора. Выбор допустимой ширины полосы пропускания Δω диктуется спектральными свойствами полезного сигнала S(t) – отклонением от «чистой» синусоиды.

В случае медленно меняющегося во времени сигнала U_s(t) хорошие результаты дает метод синхронного детектирования, суть которого состоит в следующем (рис. 35).

Рис. 35. Метод синхронного детектирования сигнала. 1 – модулятор, 2- детектор, 3 – смеситель (гетеродин), 4 – частотный фильтр.

Сигнал $U_s(t) \approx \text{const}$, который необходимо зарегистрировать, модулируется с частотой ω_0 , так что на выходе модулятора 1 сигнал имеет вид $U_s(t)\sin\omega_0 t$, пусть далее модулированный сигнал попадает на детектор 2, где он приобретает шумовую составляющую U_N (будем считать, что шумы обусловлены измерителем), на выходе

детектора сигнал имеет вид $U_s(t)\sin\omega_0 t + U_N$ и в таком виде он поступает на вход смесителя – гетеродина 3. В смесителе происходит смешивание измеряемого сигнала с опорным синусоидальным сигналом $V_0 \sin(\omega_0 t + \phi)$, частота которого совпадает с частотой модуляции, между фазами опорного сигнала и модулированного измеряемого сигнала имеется постоянная разность ϕ . После смешивания измеряемого сигнала с опорным получим на выходе гетеродина:

$$U \sim [U_{s}(t)\sin\omega_{0}t + U_{N}(t)]V_{0}\sin(\omega_{0}t + \phi) =$$

= $\frac{U_{s}(t)V_{0}}{2}[\cos\phi - \cos(2\omega_{0}t + \phi)] + V_{0}U_{N}(t)\sin(\omega_{0}t + \phi)$ (73)

Если отфильтровать переменную составляющую сигнала (73), то получим медленно меняющийся сигнал

$$U = \frac{U_s(t)V_0}{2}\cos\phi \tag{74},$$

свободный от шумов. Метод синхронного детектирования оказывается весьма эффективным и в ряде случаев позволяет выделять из шума полезный сигнал, мощность которого значительно ниже мощности шума.

Практическое применение метода синхронного детектирования иллюстрирует рис. 36. Измеряется температура нагретой поверхности путем измерения интенсивности ИК-излучения (пирометрический метод измерения температуры). Между датчиком ИКизлучения (помещенном в фокусе собирающего зеркала) и источником излучения помещен модулятор, который представляет собой непрозрачный диск, вращающийся с постоянной скоростью. В диске имеются симметрично расположенные радиальные прорези, расположение которых определяет характер модуляции измеряемого и опорного сигналов. Последний создается специальным источником света, расположенным перед вращающимся диском (см. рис. 36) и регистрируемым фотодиодом.

Рис. 36. Применение синхронного усиления в пирометрическом измерении интенсивности излучения от нагретой поверхности.

Далее в синхронном усилителе происходит смешивание сигналов и последующая фильтрация шумов.

При практическом проведении измерений постоянные во времени сигналы, как правило, превращают путем модуляции в периодические. Связано это с тем, что детекторы и усилители переменных сигналов обладают значительно большей стабильностью. Недостатком усилителей и других устройств постоянного тока обычно является значительный дрейф нулевого уровня сигнала.

Литература.

Стариковская С.М. Физические методы исследования. Семинарские занятия. 1.1. Учет погрешностей при обработке результатов измерений: М: МФТИ, **2003**

Тейлор Дж. Введение в теорию ошибок. М. Мир, 1985.

Гмурман В.Е. Теория вероятностей и математическая статистика. М. Высшая школа, **2002.**

Клаассен К.Б. Основы измерений. Электронные методы и приборы в измерительной технике. М. Постмаркет, **2000**.

Дубнищев Ю.Н. Колебания и волны. Новосибирск, Сибирское университетское издательство, **2004** (гл.1, 2).

Харт Х. Введение в измерительную технику. М. Мир, 1999.

Максимычев А.В. Физические методы исследования. Задачи (часть 1). Погрешности эксперимента, длинные линии, измерение давления, температуры, потоков излучения и частиц. М., МФТИ, **2003**