

ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ

Введение

Цифровая обработка сигналов, базируясь на математике семнадцатого и восемнадцатого столетий, в настоящее время стала важным инструментом во многих областях науки и техники. Методы и применения цифровой обработки стары, как методы Ньютона и Гаусса, и молоды, как цифровые ЭВМ и интегральные схемы.

При цифровой обработке используется представление сигналов в виде последовательностей чисел или символов. Цель такой обработки может заключаться в оценке характерных параметров сигнала или в преобразовании сигнала в форму, которая в некотором смысле более удобна. Формулы классического численного анализа, такие, как формулы для интерполяции, интегрирования и дифференцирования, безусловно являются алгоритмами цифровой обработки. Наличие быстродействующих цифровых ЭВМ благоприятствовало развитию все более сложных и рациональных алгоритмов обработки сигналов; последние же успехи в технологии интегральных схем обещают высокую экономичность построения очень сложных систем цифровой обработки сигналов.

Цифровая обработка сигналов применяется в таких различных областях, как биомедицина, акустика, звуковая локация, радиолокация, сейсмология, связь, системы передачи данных, ядерная техника, и многих других. Например, при анализе электроэнцефалограмм, электрокардиограмм, а также передаче и распознавании речи требуется выделять некоторые характерные параметры сигнала. Иногда же возникает необходимость отделения помехи типа шума от сигнала или приведения сигнала к виду, который наиболее удобен для пользователя. В качестве другого примера обработки сигналов можно привести случай, когда сигнал, передаваемый по каналу связи, подвергается различным искажениям и приемник компенсирует их.

Обрабатываются не только сигналы одной размерности. Так, в случаях, связанных с обработкой изображений, необходимо использовать методы двумерной обработки сигналов. Это нужно для улучшения рентгеновских изображений, улучшения и анализа изображений при аэрофотосъемке для обнаружения лесных пожаров или повреждения посевов, анализе фотографий, полученных с помощью метеорологических спутников, а также для улучшения телевизионных изображений Луны и дальнего космоса. Для анализа сейсмических данных, необходимого при разведке нефти, измерениях силы землетрясений и контроле ядерных взрывов, также используются методы многомерной обработки сигналов.

До недавнего времени обработка сигналов, как правило, выполнялась при помощи аналоговых устройств. Некоторые исключения имели место в 50-х годах, особенно в областях, где требовалась сложная обработка сигналов. Это требовалось, например, при анализе некоторых геофизических данных, которые могли быть записаны на магнитную ленту для последующей обработки на больших ЭВМ. Анализ геофизических данных был одним из первых примеров обработки сигналов с использованием цифровых ЭВМ. Этот тип обработки сигналов не всегда мог быть выполнен в реальном времени. Например, для обработки данных, записанных на магнитную ленту только в течение нескольких секунд, часто требовались минуты или часы машинного времени. Даже при этом универсальность цифровой ЭВМ обеспечивала высокую эффективность обработки.

В дальнейшем использование цифровых ЭВМ в обработке сигналов шло различными путями. Благодаря своей гибкости цифровые ЭВМ были полезны для моделирования систем обработки сигналов до их технической реализации. При таком подходе новые алгоритмы обработки сигналов или системы могли быть изучены еще в экспериментальных условиях без расходования экономических и технических ресурсов для построения самих систем. Применение цифровых ЭВМ давало большой выигрыш из-за их гибкости и универсальности. Однако обработка не всегда могла быть выполнена в реальном времени. Следовательно,

цифровая ЭВМ использовалась в основном для аппроксимаций или моделирования аналоговых систем обработки. В соответствии с этим в начале задача цифровой фильтрации в основном сводилась к программированию фильтра на цифровой ЭВМ так, чтобы при аналого-цифровом преобразовании сигнала с последующей цифровой фильтрацией и цифро-аналоговым преобразованием система аппроксимировала хороший аналоговый фильтр. Представление о том, что цифровые системы могут в действительности быть практичны для непосредственной обработки сигналов в радиосвязи, радиолокации или во многих других сферах приложений, казалось маловероятным. Быстродействие, стоимость и размеры были, конечно, тремя важными факторами, говорившими в пользу применения аналоговых устройств.

По мере того как обработка сигналов осуществлялась на цифровых ЭВМ, естественной тенденцией было исследование все более сложных алгоритмов обработки сигналов. Некоторые из этих алгоритмов были разработаны с учетом больших возможностей цифровой ЭВМ, однако из-за сложности не реализовывались в аналоговой аппаратуре, т. е. многие из этих алгоритмов оказывались интересными, но до некоторой степени непрактичными. Примером класса алгоритмов этого типа был ряд алгоритмов, названных анализом кепстра и гомоморфной фильтрацией. На цифровых ЭВМ было ясно продемонстрировано, что эти алгоритмы могли быть успешно применены в системах полосового сжатия речи, развертки и устранения эхо-сигналов. Использование этих алгоритмов требует точной оценки обратного преобразования Фурье логарифма преобразования Фурье входного сигнала. При этом требования к точности и разрешающей способности были таковы, что аналоговые анализаторы спектра оказывались непрактичными. Развитие таких алгоритмов обработки сигналов сделало привлекательной идею построения полностью цифровых систем обработки сигналов. Активная работа началась с исследования цифровых вокодеров, цифровых анализаторов спектра и других полностью цифровых систем в предположении, что со временем такие системы станут практичными.

Развитие новой точки зрения на цифровую обработку сигналов в дальнейшем было ускорено открытием в 1965 г. эффективных алгоритмов для вычислений преобразований Фурье. Этот класс алгоритмов стал известен как быстрое преобразование Фурье (БПФ). Возможности БПФ были значительными с нескольких точек зрения. Многие алгоритмы обработки сигналов, полученные на цифровых ЭВМ, требовали времени обработки на несколько порядков больше, чем реальное время. Часто это было связано с тем, что спектральный анализ был важной составной частью обработки сигналов, а эффективные средства для его выполнения не были известны. Алгоритм быстрого преобразования Фурье уменьшил время вычисления преобразования Фурье на несколько порядков. Это позволило создать очень сложные алгоритмы обработки сигналов в реальном времени. Кроме того, с учетом возможностей действительной реализации алгоритма быстрого преобразования Фурье в специализированном цифровом устройстве многие алгоритмы обработки сигналов, бывшие ранее непрактичными, стали находить воплощение в специализированных устройствах.

Другая важная особенность алгоритма быстрого преобразования Фурье связана с тем, что ему внутренне присуща концепция дискретного времени. Эта особенность касается непосредственно вычисления преобразования Фурье дискретного сигнала или последовательности и заключается в ряде свойств и математических операций, строго относящихся к дискретному времени, в связи с чем этот алгоритм не является просто аппроксимацией преобразования Фурье непрерывного сигнала. Это вызвало видоизменение многих понятий и алгоритмов обработки сигналов на основе математических методов для дискретного времени, которые затем привели к формулировке ряда четких соотношений для дискретного времени. Все это явилось отходом от представления, что обработка сигналов на цифровой ЭВМ является лишь аппроксимацией методов аналоговой обработки. При таком изменении точки зрения возник значительный интерес к новой или, как сказано ранее,

Области применения цифровой обработки сигналов стремительно расширялись. Этому способствовало развитие больших интегральных схем и связанное с ним уменьшение стоимости и размеров цифровых устройств при одновременном увеличении их быстродействия. Цифровые фильтры специального назначения сейчас могут работать в мегагерцовом диапазоне тактовой частоты; экономически оправдываются процессоры специального назначения для выполнения быстрого преобразования Фурье при высокой частоте входных данных; несложные цифровые фильтры выполняются на отдельных чипах; в настоящее время почти все вопросы, связанные с системами полосового сжатия речи, рассматриваются в плане построения полностью цифровых систем как наиболее практических; цифровые процессоры также являются неотъемлемой частью многих современных радиолокационных и звуколокационных систем. В дополнение к развитию цифровых специализированных устройств обработки сигналов имеются цифровые программируемые ЭВМ специального назначения, архитектура которых приспособлена к задачам обработки сигналов. Такие ЭВМ находят применение при обработке сигналов в реальном времени так же, как и при моделировании в реальном времени на специализированных цифровых устройствах.

Области применения цифровой обработки сигналов постоянно расширяются. Методы цифровой обработки будут несомненно способствовать существенным изменениям в областях науки и техники, где они будут применяться. Характерным примером является область телефонии, где цифровые методы обещают существенную экономию и гибкость при реализации систем переключения и передачи. Учитывая направление развития цифровой обработки сигналов, очевидно, что ее методы будут применяться скорее по своему прямому назначению, чем для аппроксимации аналоговых систем обработки.

1. Дискретные сигналы и системы

1.1. Классификация сигналов и систем

Сигнал может быть определен как функция, переносящая информацию о состоянии или поведении физической системы. Сигнал может принимать форму колебаний, зависящих от времени или от пространственных координат. Математически сигналы представляются в виде функций одной или более независимых переменных. Так, например, речевой сигнал математически представляется как функция времени, а изображение — как зависимость яркости от двух пространственных переменных. Обычно при математическом представлении сигнала $s(t)$ независимой переменной считают время, и мы будем следовать этому, хотя на самом деле эта переменная может иметь другой смысл.

Независимая переменная в математическом представлении сигнала может быть как непрерывной, так и дискретной. Сигналы в непрерывном времени определяются на континууме моментов времени и, следовательно, представляются как функции от непрерывной переменной. Дискретные сигналы (сигналы в дискретном времени) определяются в дискретные моменты времени и представляются последовательностями чисел. На рис.1.1 графически отображены сигнал непрерывного времени и дискретный сигнал, образованный его эквидистантными выборками. Как мы увидим в дальнейшем, такие сигналы, как речь или изображение, могут быть представлены как с непрерывной, так и с дискретной независимой переменной, и если удовлетворяются определенные условия, эти представления полностью эквивалентны.

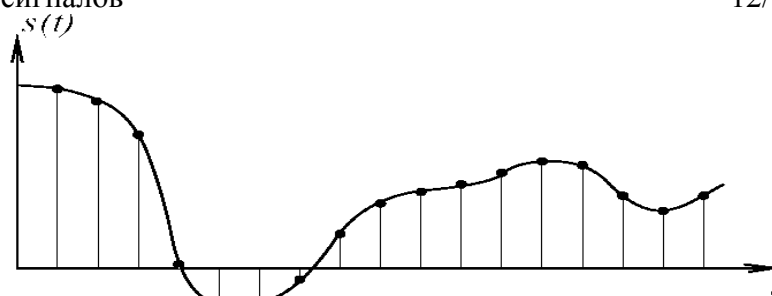


Рис.1.1.

Вдобавок к тому, что независимые переменные могут быть непрерывными или дискретными, амплитуда сигнала также может быть как непрерывной, так и дискретной. Цифровые сигналы — это сигналы, у которых дискретны и время, и амплитуда. Сигналы в непрерывном времени и с непрерывным диапазоном амплитуд также называются аналоговыми сигналами.

Чтобы облегчить извлечение информации, сигналы должны подвергаться обработке. Поэтому весьма важно развитие техники, а также самих систем обработки сигналов. Техника обработки сигналов заключается в преобразовании сигнала в другой сигнал, являющийся более предпочтительным. Например, может понадобиться разделение двух или большего числа сигналов, которые ранее были объединены некоторым образом; выделение некоторой компоненты или параметра сигнала либо оценка одного или нескольких параметров сигнала.

Системы обработки сигналов могут классифицироваться точно так же, как и сами сигналы. Так, системы в непрерывном времени — это системы, у которых на входе и выходе имеются сигналы в непрерывном времени, а дискретные системы (системы в дискретном времени) — это системы, у которых на входе и выходе дискретные сигналы. Точно так же аналоговые системы — это системы с аналоговыми сигналами на входе и выходе, а цифровые системы — системы с цифровыми сигналами на входе и выходе. В таком случае цифровая обработка сигналов имеет дело с преобразованиями сигналов, являющимися дискретными как по амплитуде, так и по времени. Наш курс посвящен скорее дискретным, чем цифровым сигналам и системам. Эффекты, связанные с дискретизацией амплитуды, будут рассмотрены особо.

Дискретные сигналы могут появляться при получении выборок из аналоговых сигналов или же они могут порождаться непосредственно некоторым дискретным во времени процессом. Вне зависимости от происхождения дискретных сигналов цифровые системы обработки таких сигналов обладают рядом полезных качеств. Они могут быть реализованы с большой гибкостью на универсальных цифровых вычислительных машинах или с помощью цифровой аппаратуры. При необходимости их можно использовать для моделирования аналоговых систем или, что более важно, для преобразований сигнала, которые невозможно осуществить на аналоговой аппаратуре. Поэтому, когда требуется сложная обработка сигналов, часто желательно представить их в цифровом виде.

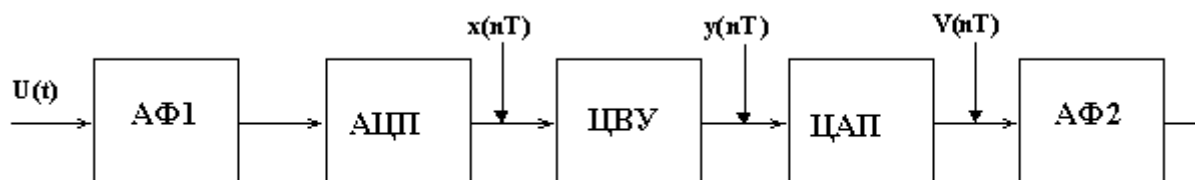


Рис.1.2. Блок-схема цифровой обработки аналогового сигнала.

На рис.1.2 представлена блок-схема цифровой обработки аналогового сигнала, включающая в себя следующие блоки:

- АФ1- аналоговый фильтр нижних частот, ограничивающий ширину спектра

входного сигнала;

- АЦП- аналого-цифровой преобразователь, осуществляющий дискретизацию во времени и квантование по уровню временных отсчетов (выборок) т.е. представление их в форме периодической последовательности двоичных чисел (цифровой сигнал $x(nT)$);
- ЦВУ- цифровое вычислительное устройство, выполняющее соответствующее преобразование сигнала $x(nT)$ в цифровой сигнал $y(nT)$;
- ЦАП- цифро-аналоговый преобразователь, выполняющий преобразование цифрового сигнала $y(nT)$ в сигнал дискретного времени $V(nT)$;
- АФ2- аналоговый фильтр, преобразующий дискретный сигнал $V(nT)$ в аналоговый $V(t)$.

В блок-схеме можно увидеть все упомянутые выше классы систем, в том числе аналоговую, дискретную и цифровую. На рис.1.3а графически отображена процедура преобразования аналогового сигнала сначала в дискретный (дискретизация по времени), затем в цифровой (квантование по уровню), образованный последовательностью десятичных цифр (рис.1.3б) и последовательностью двоичных кодовых групп (1.3в). Рис.1.3г отображает график ошибки квантования, образованной разностью дискретного и квантованного сигналов.

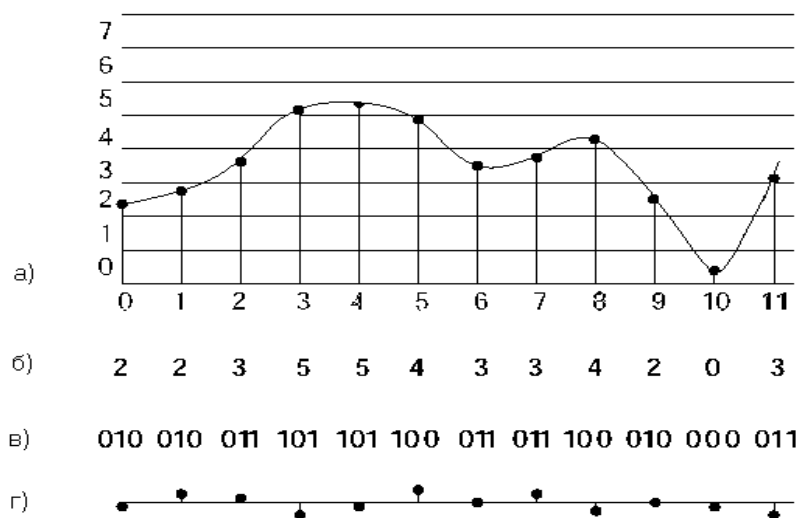


Рис.1.3.

В этом разделе мы рассмотрим основные понятия, связанные с одномерными дискретными сигналами и системами их обработки. Особое внимание уделим классу линейных дискретных систем, инвариантных относительно сдвига. Из этого раздела будет видно, что свойства этих систем похожи на свойства линейных инвариантных во времени аналоговых систем. Действительно, последовательности можно трактовать как аналоговые импульсные сигналы. Однако во многих современных применениях цифровой обработки сигналов не все последовательности получаются путем выборки из непрерывных во времени сигналов. Кроме того, многие дискретные системы не являются просто аппроксимациями соответствующих аналоговых систем. Поэтому мы не будем пытаться применить выводы, следующие из теории для аналоговых систем, а получим их непосредственно для дискретных систем, пользуясь соответствующей этому случаю терминологией и обозначениями. Дискретные сигналы будут соотноситься с аналоговыми сигналами только тогда, когда это будет необходимо.

1.2. Дискретные сигналы (последовательности)

В теории дискретных систем мы будем интересоваться обработкой сигналов, представляемых последовательностями. Последовательность чисел x , в которой n -й член последовательности обозначается как $x(n)$, может быть формально записана в виде

$$x = \{x(n)\}, \quad -\infty < n < \infty. \quad (1.1)$$

Хотя последовательности не всегда получаются путем выборки из аналоговых колебаний, для удобства мы будем называть $x(n)$ « n -й выборкой» последовательности. Хотя, строго говоря, $x(n)$ обозначает n -й член последовательности, запись (1.1) часто слишком громоздка и более удобно говорить о «последовательности $x(n)$ ». Дискретные сигналы (последовательности) часто изображаются графически так, как это показано на рис. 1.4. Хотя абсцисса изображена в виде непрерывной линии, важно сознавать, что $x(n)$ определена только для целых значений n . Неправильно полагать, что $x(n)$ равна нулю для нецелых n , просто $x(n)$ не определена для нецелых значений n .

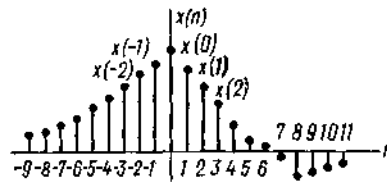


Рис. 1.4. Графическое представление дискретного сигнала.

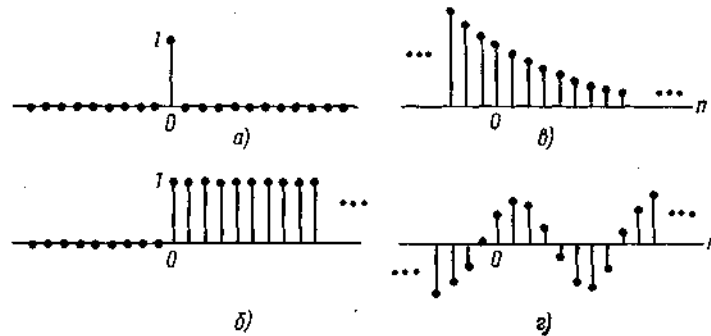


Рис. 1.5. Примеры последовательностей:

а) единичный импульс; б) единичная ступенчатая последовательность; в) действительная экспоненциальная последовательность; г) синусоидальная последовательность.

Некоторые примеры последовательностей, играющих важную роль при дискретной обработке, показаны на рис. 1.5. Единичный импульс $\delta(n)$ определяется как последовательность со значениями

$$\delta(n) = \begin{cases} 1, & n = 0; \\ 0, & n \neq 0. \end{cases}$$

Как мы вскоре увидим, единичный импульс играет для дискретных сигналов и систем ту же роль, какую играет дельта-функция для аналоговых сигналов и систем. Для удобства единичный импульс часто называется просто импульсом. Важно отметить, что с единичным импульсом не связаны те математические затруднения, которые встречаются при использовании дельта-функции. Его определение просто и точно.

Единичная ступенчатая последовательность $u(n)$ имеет значения $u(n) = \begin{cases} 1, & n \geq 0 \\ 0, & n < 0 \end{cases}$ и связана с единичным импульсом соотношением

$$u(n) = \sum_{k=0}^{\infty} \delta(n-k) = \sum_{k=-\infty}^n \delta(k). \quad (1.2)$$

Соответственно единичный импульс связан с единичной ступенчатой последовательностью соотношением

$$\delta(n) = u(n) - u(n-1). \quad (1.3)$$

Действительная экспоненциальная последовательность — это последовательность со значениями вида a^n , где a — действительное число. Эту последовательность можно, например, получить периодическим (с периодом T) взятием отсчетов (выборок) экспоненты непрерывного времени $e^{-\alpha t}/_{t=nT} = e^{-\alpha nT} = a^n$, где $a = e^{-\alpha T}$.

Синусоидальная последовательность имеет вид $A \cos(\omega_0 n + \varphi)$, и ее также можно получить периодическим (с периодом T) взятием отсчетов (выборок) синусоиды непрерывного времени $A \cos(\Omega_0 t + \varphi)/_{t=nT} = A \cos(\Omega_0 nT + \varphi) = A \cos(\omega_0 n + \varphi)$, где $\omega_0 = \Omega_0 T$.

Комплексная экспоненциальная последовательность имеет вид $e^{(\sigma + j\omega_0)n} = e^{\sigma n} (\cos \omega_0 n + j \sin \omega_0 n)$.

Последовательность $x(n)$ по определению называется периодической с периодом N , если $x(n) = x(n+N)$ для всех n . Комплексная экспонента с $\sigma=0$ и синусоидальная последовательность имеют период $2\pi/\omega_0$ только тогда, когда это действительное число является целым. Если $2\pi/\omega_0$ не целое, но рациональное число, то синусоидальная последовательность будет периодической, однако с периодом, большим $2\pi/\omega_0$. Если $2\pi/\omega_0$ не рационально, то синусоидальная и комплексная экспоненциальная последовательности вовсе не будут периодическими. Параметр ω_0 будет называться *цифровой частотой* синусоиды или комплексной экспоненты вне зависимости от того, периодичны они или нет. Частота может быть выбрана в любом непрерывном диапазоне значений. Однако без потери общности можно ограничить этот диапазон, полагая $0 \leq \omega_0 \leq 2\pi$ (или $-\pi \leq \omega_0 \leq \pi$), так как синусоидальные и комплексные экспоненциальные последовательности, получаемые при изменении ω_0 в диапазоне $2\pi k \leq \omega_0 \leq 2\pi(k+1)$, в точности совпадают при любых k с последовательностями, получаемыми при изменении ω_0 в диапазоне $0 \leq \omega_0 \leq 2\pi$.

Иногда удобно пользоваться термином энергии последовательности. Энергия E последовательности $x(n)$ определяется как

$$E = \sum_{n=-\infty}^{\infty} |x(n)|^2.$$

При анализе систем обработки дискретных сигналов приходится производить ряд следующих преобразований последовательностей. Произведение и сумма двух последовательностей x и y определяются как произведение и сумма выборок соответственно: $xy = \{x(n)y(n)\}$, $x+y = \{x(n) + y(n)\}$. Умножение последовательности x на число a определяется как $xa = \{ax(n)\}$.

Говорят, что последовательность y является задержанной или сдвинутой последовательностью x , если y имеет значения $y(n) = x(n-n_0)$, где n_0 — целое число.

Произвольная последовательность может быть представлена как сумма взвешенных и задержанных единичных импульсов. Например, последовательность $p(n)$, изображенную на рис. 1.6, можно записать как $p(n) = a_{-3}\delta(n+3) + a_1\delta(n-1) + a_2\delta(n-2) + a_7\delta(n-7)$. В общем случае

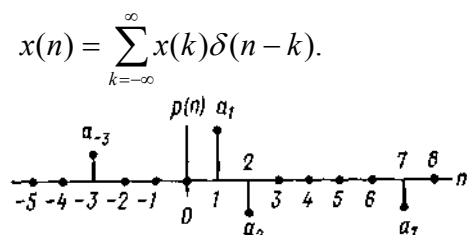


Рис. 1.6. Пример последовательности, представляющей сумму взвешенных и задержанных единичных импульсов.

1.3. Дискретные линейные системы с постоянными параметрами (инвариантные к сдвигу)

Система определяется математически как однозначное преобразование или оператор, отображающий входную последовательность $x(n)$ (вход) в выходную $y(n)$ (выход), что математически записывается в виде $y(n) = T[x(n)]$, а графически часто изображается так, как показано на рис. 1.7.



Рис. 1.7. Представление преобразования, отображающего входную последовательность $x(n)$ в выходную последовательность $y(n)$.

Классы дискретных систем определяются путем наложения ограничений на преобразование $T[]$. В дальнейшем будет широко рассматриваться класс линейных инвариантных относительно сдвига систем, потому что они сравнительно просты в математическом отношении, а также потому, что они дают удобный вид обработки сигналов.

Класс *линейных систем* определяется принципом суперпозиции. Если $y_1(n)$ и $y_2(n)$ —

$$T[ax_1(n) + bx_2(n)] = aT[x_1(n)] + bT[x_2(n)] = ay_1(n) + by_2(n) \quad (1.5)$$

для произвольных постоянных a и b . Мы видели, что произвольная последовательность $x(n)$ может быть представлена в виде задержанной и взвешенной суммы единичных импульсов (1.4). Это представление вместе с (1.5) предполагает, что линейная система может быть полностью охарактеризована откликом на единичный импульс — импульсной характеристикой. А именно, пусть $h_k(n)$ — отклик системы на $\delta(n-k)$ единичный

$$y(n) = T\left[\sum_{k=-\infty}^{\infty} x(k)\delta(n-k)\right]$$

С учетом (1.5) можно записать

$$y(n) = \sum_{k=-\infty}^{\infty} x(k)T[\delta(n-k)] = \sum_{k=-\infty}^{\infty} x(k)h_k(n). \quad (1.6)$$

Таким образом, согласно (1.6) реакцию системы можно выразить через отклики на $\delta(n-k)$. Если накладывается только одно условие — линейность, то $h_k(n)$ будет зависеть как от n , так и от k , и в этом случае польза от выражения (1.6) для вычислений невелика. Более полезный результат получится, если мы наложим дополнительное ограничение, состоящее в инвариантности к сдвигу.

Класс *инвариантных к сдвигу систем* характеризуется следующим свойством: если $y(n)$ — отклик на $x(n)$, то $y(n-k)$ будет откликом на $x(n-k)$, где k — положительное или отрицательное целое число. Когда индекс n связывается со временем, свойству инвариантности к сдвигу соответствует свойство инвариантности во времени. Поэтому такие системы в дальнейшем будем называть дискретными линейными системами с постоянными во времени параметрами (ДЛСПП). Из свойства инвариантности к сдвигу следует, что если $h(n)$ — отклик на $\delta(n)$, то откликом на $\delta(n-k)$ будет просто $h(n-k)$. Поэтому (1.6) принимает вид

$$y(n) = \sum_{k=-\infty}^{\infty} x(k)h(n-k). \quad (1.7)$$

Значит, любая инвариантная к сдвигу система полностью характеризуется импульсной характеристикой $h(n)$. Выражение (1.7) обычно называется сверткой. Если $y(n)$

—последовательность, значения которой связаны со значениями двух последовательностей $h(n)$ и $x(n)$ выражением (1.7), то мы говорим, что $y(n)$ есть свертка $x(n)$ с $h(n)$, и обозначаем $y(n) = x(n) * h(n)$. Заменяя переменную в (1.7), получим другое выражение

$$y(n) = \sum_{k=-\infty}^{\infty} h(k)x(n-k) = h(n) * x(n). \quad (1.8)$$

Поэтому порядок, в котором две последовательности входят в свертку, не важен. Другими словами, линейная инвариантная к сдвигу система со входом $x(n)$ и импульсной характеристикой $h(n)$ будет иметь тот же выход, что и линейная инвариантная к сдвигу система со входом $h(n)$ и импульсной характеристикой $x(n)$.

Две линейные инвариантные к сдвигу системы, включенные каскадно, образуют линейную инвариантную к сдвигу систему с импульсной характеристикой, равной свертке импульсных характеристик исходных систем. Так как порядок в свертке не важен, то импульсная характеристика составной системы не зависит от порядка, в котором включены исходные системы. Это свойство иллюстрируется рис. 1.6, где изображены три системы, имеющие одинаковые импульсные характеристики.

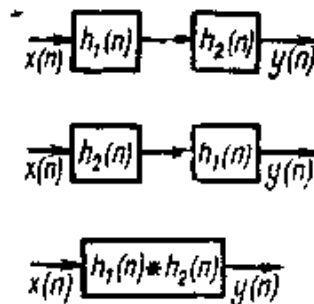


Рис. 1.8. Три линейные инвариантные к сдвигу системы с одинаковыми импульсными характеристиками.

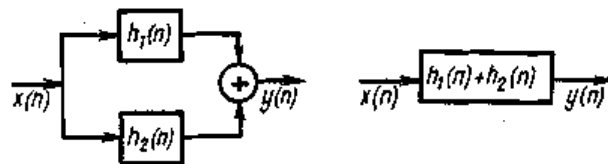


Рис. 1.9. Параллельное включение линейных инвариантных к сдвигу систем и эквивалентная система.

Из (1.7) и (1.8) следует, что две инвариантные к сдвигу системы, включенные параллельно, эквивалентны одной системе с импульсной характеристикой, равной сумме импульсных характеристик исходных систем. Это свойство иллюстрируется рис. 1.9.

Хотя выражение свертки в виде суммы аналогично интегралу свертки в теории линейных аналоговых систем, следует подчеркнуть, что свертку в виде суммы нельзя понимать как приближение к интегралу свертки. В противоположность интегралу свертки, играющему в основном теоретическую роль в применении к аналоговым линейным системам, свертка в виде суммы, как мы увидим в дальнейшем, вдобавок к своей теоретической значимости может служить для реализации дискретной системы. Поэтому важно глубже понять свойства свертки и получить опыт в применении свертки для вычислений.

Пример. Рассмотрим систему с импульсной характеристикой $h(n) = \begin{cases} a^n, & n \geq 0; \\ 0, & n < 0, \end{cases}$

$u(n-N)$, заметим, что в силу (1.7) для получения n -го значения выходной

последовательности нужно сформировать произведение $x(k)h(n-k)$ и просуммировать значения получившейся последовательности.

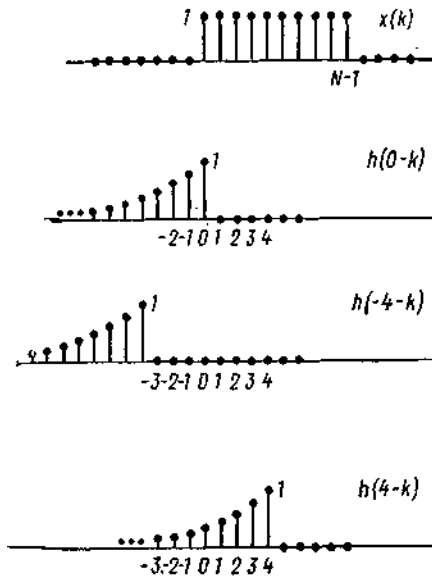


Рис. 1.10. Последовательности, входящие в свертку $[h(n-k)]$, показаны для нескольких значений n .

Две составляющие последовательности показаны на рис. 1.8 как функции k , причем $h(n-k)$ изображена для нескольких значений n . Как видно из рис. 1.8 при $n < 0$ $h(n-k)$ и $x(k)$ не имеют ненулевых перекрывающихся выборок и, следовательно, $y(n)=0$ при $n < 0$. При $0 < n < N$ $h(n-k)$ и $x(k)$ имеют ненулевые перекрывающиеся выборки от $k=0$ до $k=n$, поэтому при $0 \leq n < N-1$

$$y(n) = \sum_{k=0}^n a^{n-k} = a^n [(1 - a^{-(n+1)}) / (1 - a^{-1})] = (1 - a^{n+1}) / (1 - a), 0 \leq n < N;$$

поэтому

$$y(n) = \sum_{k=0}^{N-1} a^{n-k} = a^n [(1 - a^{-N}) / (1 - a^{-1})], n \geq N.$$

Реакция $y(n)$ изображена на рис. 1.11.

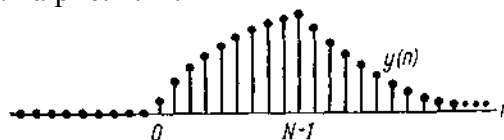


Рис. 1.11. Реакция системы с импульсной характеристикой $h(n)=a^n u(n)$ на входной

1.4. Устойчивость и физическая реализуемость

Мы видели, что требования линейности и инвариантности времени определяют класс систем, которые представляются сверткой. Добавочные ограничения устойчивости и физической реализуемости определяют практически важный, но более узкий класс линейных инвариантных во времени систем.

Устойчивой системой назовем систему, в которой каждый ограниченный входной сигнал создает ограниченный выходной сигнал. Линейная инвариантная к сдвигу система устойчива тогда и только тогда, когда

$$S = \sum_{k=-\infty}^{\infty} |h(k)| < \infty. \quad (1.9)$$

Это можно показать следующим образом. Если (1.9) справедливо и x ограничено, т. е. $|x(n)| < M$ для всех n , то из (1.8) следует

$$|y(n)| = \left| \sum_{k=-\infty}^{\infty} h(k)x(n-k) \right| \leq M \sum_{k=-\infty}^{\infty} |h(k)| < \infty.$$

Поэтому y ограничено. Доказать обратное можно, показав, что если $S = \infty$, то существует ограниченный входной сигнал, который создает неограниченный выходной сигнал. Таким входом является последовательность со значениями

$$x(n) = \begin{cases} \frac{h^*(-n)}{|h(-n)|}, & h(n) \neq 0; \\ 0, & h(n) = 0, \end{cases}$$

где $h^*(n)$ — комплексно-сопряженная к $h(n)$ величина. Ясно, что $x(n)$ ограничена. Значение на выходе при $n=0$ равно

$$y(0) = \sum_{k=-\infty}^{\infty} x(-k)h(k) = \sum_{k=-\infty}^{\infty} |h(k)|^2 / |h(k)| = S.$$

Поэтому при $S = \infty$ выходная последовательность не ограничена.

Физически реализуемая (каузальная) система — это система, у которой изменения на выходе не опережают изменения на входе, т. е. в физически реализуемой системе, если $x_1(n) = x_2(n)$, $n < n_0$, то $y_1(n) = y_2(n)$, $n < n_0$. Линейная инвариантная к сдвигу система физически

Поэтому иногда удобно называть последовательность, которая равна нулю при $n < 0$ физически реализуемой последовательностью, подразумевая под этим то, что она может быть импульсной характеристикой физически реализуемой системы.

инвариантную к сдвигу систему с импульсной характеристикой $h(n) = a^n u(n)$; так как эта импульсная характеристика равна нулю при $n < 0$, то система физически реализуема. Чтобы определить устойчивость, мы должны вычислить сумму

$$S = \sum_{k=-\infty}^{\infty} |h(k)| = \sum_{k=0}^{\infty} |a|^k.$$

Если $|a| < 1$, бесконечная геометрическая прогрессия имеет сумму $S = 1 / (1 - |a|)$, но если $|a| > 1$ — ряд расходится. Следовательно, система устойчива только при $|a| < 1$.

1.5. Линейные разностные уравнения с постоянными коэффициентами

Во многих применениях играет важную роль один подкласс класса линейных инвариантных к сдвигу систем. Это подкласс состоит из систем, для которых вход $x(n)$ и выход $y(n)$ удовлетворяют линейному разностному уравнению N -го порядка с постоянными коэффициентами вида

$$\sum_{k=0}^N a(k)y(n-k) = \sum_{k=0}^M b(k)x(n-k). \quad (1.10)$$

В общем случае системы этого класса не обязательно должны быть физически реализуемыми. Например, рассмотрим разностное уравнение первого порядка

$$y(n) - ay(n-1) = x(n). \quad (1.11)$$

Легко проверить прямой подстановкой, что при $x(n) = \delta(n)$ уравнению (1.11) удовлетворяют как $y(n) = a^n u(n)$, так и $y(n) = -a^n u(-n-1)$. Первое решение соответствует

нереализуемо и устойчиво только при $|a| > 1$. Общепринято полагать, что такое разностное

уравнение, как (1.10), характеризует физически реализуемую систему, и мы будем придерживаться этого положения, если только не будет оговорено противоположное.

Без добавочной информации разностное уравнение вида (1.10) неоднозначно определяет соотношение между входом и выходом системы. Это является следствием того, что разностным уравнениям, как и дифференциальным, удовлетворяет целое семейство решений. Например, если разностному уравнению (1.11) удовлетворяет $y_I(n)$ при $x(n) = x_I(n)$,

I

удовлетворяющую однородному разностному уравнению (т. е. уравнению, у которого правая часть равна нулю), и эта сумма также будет удовлетворять (1.10). Так, решение однородного

$$\sum_{k=0}^N a(k)y(n-k) = 0$$

$$y(n) = \sum_{k=1}^N A_k z_k^n$$

$$\sum_{k=0}^N a(k)z^{N-k} = 0$$

k

записывается иначе.

коэффициентами, будет инвариантна к сдвигу только тогда, когда мы соответствующим образом выберем однородную составляющую. Если, например, система физически реализуема, мы должны выбрать начальные условия так, чтобы если $x(n)=0$ при $n < n_0$, то и $y(n)=0$ при $n < n_0$. Мы будем полагать, что если система удовлетворяет линейному

необходимым для инвариантности к сдвигу условиям до тех пор, пока не оговорено противоположное.

разностное уравнение дает явное соотношение между входом и выходом. Это соотношение можно получить, переписывая (1.10) в виде

$$y(n) = -\sum_{k=1}^N (a(k)/a(0))y(n-k) + \sum_{k=0}^M (b(k)/a(0))x(n-k). \quad (1.12)$$

соответственно N и M прошлых значений выхода и входа. Как и в случае свертки, разностное уравнение не только дает теоретическое описание системы, но может быть также основой для реализации системы.

Пример. Разностное уравнение первого порядка (1.11) дает рекуррентную формулу $y(n)=ay(n-1)+x(n)$. Чтобы получить импульсную характеристику, положим $x(n)=\delta(n)$ при нулевых начальных условиях. Тогда

$$h(n) = 0, n < 0;$$

$$h(0) = ah(-1) + 1 = 1;$$

$$h(1) = ah(0) = a;$$

⋮

$$h(n) = ah(n-1) = a^n.$$

Таким образом, $h(n) = a^n u(n)$.

Чтобы получить другое решение, положим $x(n)=\delta(n)$ и $y(n)=0$ при $n > 0$. Из (1.11) можно записать рекуррентное соотношение $y(n-1) = (1/a)[y(n)-x(n)]$ или $y(n)=(1/a)[y(n+1)-x(n+1)]$. Тогда

$$h(n) = 0, n > 0;$$

$$h(0) = (1/a)[h(1)-x(1)]=0;$$

$$h(-1) = (1/a)[h(0)-x(0)] = -a^{-1};$$

$\vdots$
 $h(n) = (1/a)h(n+1) = -a^n$. Таким образом, $h(n) = -a^n u(-n-1)$.

Рассмотрим еще один пример получения разностного уравнения и его решения. Предположим, что банк предоставил ссуду под недвижимость в сумме 50000 долларов, которая должна быть возвращена через 30 лет равными ежемесячными взносами размером p долларов. Выплачиваемый процент установлен на уровне 15% в год от невозвращенной суммы. Каковы должны быть ежемесячные платежи и общая возвращенная банку сумма денег?

Пусть $P(n)$ - неоплаченная часть ссуды, оставшаяся после выплаты n -го ежемесячного взноса. Тогда будет иметь место следующее соотношение (разностное уравнение):

$$P(n) = (1+r)P(n-1) - p, \text{ для } n = 1, 2, 3, \dots, 360,$$

где $r = 0,15/12 = 0,0125$ - ежемесячная норма процента. Первоначально $P(0) = 50000$ и мы

$$P(1) = (1+r)P(0) - p;$$

$$P(2) = (1+r)P(1) - p = (1+r)^2 P(0) - p[1 + (1+r)];$$

$$P(3) = (1+r)^3 P(0) - p[1 + (1+r) + (1+r)^2];$$

$\vdots$

$$P(n) = (1+r)^n P(0) - p \sum_{k=0}^{n-1} (1+r)^k = (1+r)^n P(0) - \frac{(1+r)^n - 1}{r} p.$$

$$p = \frac{r(1+r)^{360}}{(1+r)^{360} - 1} P(0) = \frac{0,0125 * 1,0125^{360}}{1,0125^{360} - 1} 50000 = 632,22$$

Полная сумма возврата в банк за ссуду в 50000 долларов составит величину $360p = 227599,22$ долларов.

В общем случае линейная система, инвариантная к сдвигу, может иметь импульсную характеристику как конечной, так и бесконечной длительности. В силу определенных свойств цифровой обработки сигналов полезно различать эти два случая. Будем называть системы с конечной импульсной характеристикой коротко *КИХ-системами*, а системы с бесконечной импульсной характеристикой — *БИХ-системами*. Если в (1.10) положить $N=0$,

$$y(n) = (1/a(0)) \left[\sum_{k=0}^M b(k)x(n-k) \right],$$

тогда оно соответствует КИХ-системе. Действительно, сравнение с (1.8) показывает, что это

$$h(n) = \begin{cases} (b(n)/a(0)), & n = 0, 1, \dots, M; \\ 0 & \text{в остальных случаях.} \end{cases}$$

уравнением вида (1.10) с $N=0$. В противоположность этому для БИХ-системы N должно быть больше нуля.

1.6. Представление дискретных сигналов и систем в частотной области

В предыдущих разделах мы ввели несколько основных понятий теории дискретных сигналов и систем. Мы видели, что для линейных инвариантных к сдвигу систем представление входной последовательности в виде взвешенной суммы задержанных единичных импульсов приводит к представлению выхода в виде взвешенной суммы

задержанных импульсных характеристик. Дискретные сигналы могут быть представлены различными способами, причем, как и для аналоговых сигналов и систем, особую роль для дискретных сигналов и систем играют синусоидальные и комплексные экспоненциальные последовательности. Это объясняется тем, что основное свойство линейных инвариантных к сдвигу систем состоит в том, что в установившемся состоянии отклик на синусоидальный входной сигнал является синусоидой той же частоты с амплитудой и фазой, определяемыми системой. Именно это свойство линейных инвариантных к сдвигу систем делает представление сигналов через синусоиды и комплексные экспоненты таким полезным.

Чтобы убедиться в справедливости этого для дискретных систем, предположим, что входная последовательность является комплексной экспонентой круговой частоты ω , $x(n)$

$$y(n) = \sum_{k=-\infty}^{\infty} h(k) e^{j\omega(n-k)} = e^{j\omega n} \sum_{k=-\infty}^{\infty} h(k) e^{-j\omega k}.$$

Если ввести

$$H(e^{j\omega}) = \sum_{k=-\infty}^{\infty} h(k) e^{-j\omega k}, \quad (1.13)$$

$$y(n) = H(e^{j\omega}) e^{j\omega n}. \quad (1.14)$$

Отсюда видно, что $H(e^{j\omega})$ описывает изменение комплексной амплитуды комплексной экспоненты как функции частоты ω . Величина $H(e^{j\omega})$ называется *частотной характеристикой* системы, у которой импульсная характеристика равна $h(n)$. В общем случае $H(e^{j\omega})$ - комплексная функция и может быть выражена через свои действительную и мнимую части:

определяется как взятая со знаком «минус» первая производная фазы по ω .

экспонент, то частотная характеристика также выражает отклик на синусоидальный сигнал.

$$x(n) = A \cos(\omega_0 n + \Phi) = (A/2) e^{j\Phi} e^{j\omega_0 n} + (A/2) e^{-j\Phi} e^{-j\omega_0 n}.$$

Из (1.14) отклик на $(A/2) e^{j\Phi} e^{j\omega_0 n}$ равен $y_1(n) = H(e^{j\omega_0}) (A/2) e^{j\Phi} e^{j\omega_0 n}$. Если $h(n)$ -

$$\begin{aligned} y(n) &= (A/2) [H(e^{j\omega_0}) e^{j\Phi} e^{j\omega_0 n} + H(e^{-j\omega_0}) e^{-j\Phi} e^{-j\omega_0 n}] = (A/2) |H(e^{j\omega_0})| [e^{j[\omega_0 n + \Phi + \theta]} + e^{-j[\omega_0 n + \Phi + \theta]}] = \\ &= \text{Re}\{H(e^{j\omega_0}) A e^{j\Phi} e^{j\omega_0 n}\} = A |H(e^{j\omega_0})| \cos(\omega_0 n + \Phi + \theta), \end{aligned}$$

где $\theta = \arg[H(e^{j\omega_0})]$ — значение фазо-частотной характеристики системы на частоте ω_0 .

Частотная характеристика $H(e^{j\omega})$ является непрерывной функцией частоты. Кроме того, это периодическая функция частоты ω с периодом 2π . Это свойство следует непосредственно из (1.13), так как $e^{j(\omega+2\pi)k} = e^{j\omega k}$. То, что частотная характеристика имеет одинаковые значения на частотах ω и $\omega+2\pi$, означает, что система реагирует одинаково на комплексные экспоненты этих двух частот. Такое поведение понятно, так как эти две экспоненциальные последовательности совпадают друг с другом.

Пример. В качестве примера расчета частотной характеристики рассмотрим систему с импульсной характеристикой

$$h(n) = \begin{cases} 1, & 0 \leq n \leq N-1; \\ 0 & \text{в остальных случаях.} \end{cases} \quad (1.15)$$

(рис. 1.9). Частотная характеристика равна

$$H(e^{j\omega}) = \sum_{n=0}^{N-1} e^{-j\omega n} = \frac{1 - e^{-j\omega N}}{1 - e^{-j\omega}} = \frac{\sin(\omega N / 2)}{\sin(\omega / 2)} e^{-j(N-1)\omega / 2}. \quad (1.16)$$

Модуль и фаза $H(e^{j\omega})$ изображены на рис. 1.10 для случая $N=5$.



Рис. 1.9. Импульсная характеристика системы, для которой рассчитывается частотная характеристика

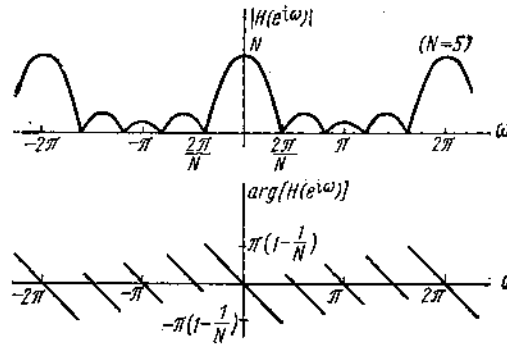


Рис. 1.10. Модуль и фаза частотной характеристики системы с импульсной характеристикой, изображенной на рис. 1.9

Поскольку $H(e^{j\omega})$ — периодическая функция частоты, она может быть представлена в

следует, что $h(n)$ могут быть определены через $H(e^{j\omega})$ как коэффициенты Фурье

$$h(n) = (1/2\pi) \int_{-\pi}^{\pi} H(e^{j\omega}) e^{j\omega n} d\omega, \quad (1.17)$$

где

$$H(e^{j\omega}) = \sum_{n=-\infty}^{\infty} h(n) e^{-j\omega n}. \quad (1.18)$$

Эти равенства можно также трактовать как представление последовательности $h(n)$. А именно, полезно рассматривать (1.17) как представление последовательности $h(n)$ в виде суперпозиции (интеграла) экспоненциальных сигналов, комплексные амплитуды которых определяются выражением (1.18). Таким образом, (1.17) и (1.18) являются парой преобразований Фурье для последовательности $h(n)$, где (1.18) играет роль прямого, а (1.17) — обратного преобразования Фурье. Такое представление существует только тогда, когда ряд в (1.18) сходится.

Представление последовательности преобразованием (1.18) не ограничивается только импульсной характеристикой системы и будет справедливо для любой последовательности при условии, что ряд в (1.18) сходится. Поэтому для произвольной последовательности $x(n)$ мы определим преобразование Фурье дискретного времени (ДВПФ) соотношением

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n) e^{-j\omega n}, \quad (1.19)$$

$$x(n) = (1/2\pi) \int_{-\pi}^{\pi} X(e^{j\omega}) e^{j\omega n} d\omega, \quad (1.20)$$

Ряды (1.19) не всегда сходятся, как, например, в случаях, когда $x(n)$ — единичная ступенчатая

последовательность либо действительная или комплексная экспоненциальная последовательность для всех n . Имеются различные определения и интерпретации

$$\sum_{n=-\infty}^{\infty} |x(n)| < \infty,$$

$$\sum_{n=-\infty}^{\infty} |x(n)|^2$$

неравенства $\sum_{n=-\infty}^{\infty} |x(n)|^2 \leq [\sum_{n=-\infty}^{\infty} |x(n)|]^2$. Однако нельзя считать, что последовательность с конечной энергией абсолютно суммируема. Примером последовательности, имеющей

$$x(n) = \sin \omega_0 n / \pi.$$

что среднеквадратическая ошибка равна нулю.

Возможность представления последовательности как суперпозиции комплексных экспонент является очень важным качеством при анализе линейных систем с постоянными параметрами. Именно вследствие этого факта и принципа суперпозиции реакция такой системы на комплексную экспоненту полностью определяется частотной характеристикой $H(e^{j\omega})$. Если рассматривать (1.20) как суперпозицию комплексных экспонент бесконечно

суперпозицией откликов на каждую экспоненту, входящую в представление сигнала $x(n)$.

$$\begin{aligned} y(n) &= T[x(n)] = T[(1/2\pi) \int_{-\pi}^{\pi} X(e^{j\omega}) e^{j\omega n} d\omega] = (1/2\pi) \int_{-\pi}^{\pi} X(e^{j\omega}) T[e^{j\omega n}] d\omega = \\ &= (1/2\pi) \int_{-\pi}^{\pi} H(e^{j\omega}) X(e^{j\omega}) e^{j\omega n} d\omega. \end{aligned}$$

Поэтому преобразование Фурье выходного сигнала равно

$$Y(e^{j\omega}) = H(e^{j\omega}) X(e^{j\omega}). \quad (1.21)$$

может быть, конечно, получен более строгим образом путем применения преобразования Фурье к свертке $y(n) = \sum_{k=-\infty}^{\infty} x(k)h(n-k)$. Хотя этот более формальный подход дает строгое

подчеркнуть, что (1.21) является прямым следствием особых свойств линейных систем с постоянными параметрами. Чтобы проиллюстрировать полученные результаты, рассмотрим следующий пример.

Пример. Идеальный фильтр нижних частот с дискретным временем имеет частотную

$$H(e^{j\omega}) = \begin{cases} 1, & |\omega| \leq \omega_{cp}; \\ 0, & \omega_{cp} < |\omega| \leq \pi. \end{cases}$$

Так как $H(e^{j\omega})$ является периодической функцией, то это соотношение определяет частотную

диапазоне частот $\omega_{cp} < |\omega| \leq \pi$. Импульсная характеристика $h(n)$ определяется по (1.17):

$$h(n) = (1/2\pi) \int_{-\omega_{cp}}^{\omega_{cp}} e^{j\omega n} d\omega = (\sin \omega_{cp} n / \pi n)$$

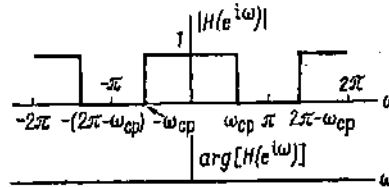


Рис. 1.11. Частотная характеристика идеального дискретного фильтра нижних частот

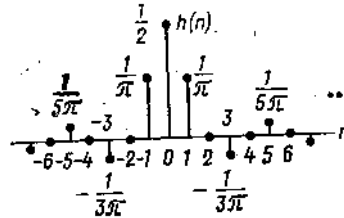


Рис. 1.12. Импульсная характеристика идеального фильтра нижних частот с частотой среза $\omega_{cp} = \pi/2$.

Идеальный фильтр нижних частот является примером системы, которая очень эффективно описывается в частотной области. Легко видеть, что эта система полностью удаляет из входного сигнала компоненты с частотой выше частоты среза ω_{cp} . Ясно, что идеальный фильтр нижних частот не является физически реализуемой системой, более того, можно показать, что он не является устойчивым в смысле определения, данного в 1.4. Тем не менее теоретически этот фильтр является очень важным и мы рассмотрим методы построения систем, близких к идеальному фильтру низких частот.

В заключение приведем некоторые наиболее важные свойства ДВПФ, которые сведены в таблицу 1.1.

Таблица 1.1. Некоторые важные свойства ДВПФ.

Последовательность	ДВПФ
$x(n)$	$X(e^{j\omega})$
$x(n-m)$	$X(e^{j\omega})e^{-j\omega m}$
$x(n)e^{j\omega_0 n}$	$X(e^{j(\omega-\omega_0)})$
$\sum_{k=-\infty}^{\infty} x(k)h(n-k)$	$X(e^{j\omega})H(e^{j\omega})$
$x(n)y(n)$	$\frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{j\theta})Y(e^{j(\omega-\theta)})d\theta$

1.7. Дискретизация сигналов с непрерывным временем

В предыдущих разделах мы старались не сопоставлять дискретные сигналы и системы с аналоговыми сигналами и системами, за исключением указаний на сходство некоторых основных теоретических понятий. Однако часто дискретные сигналы получаются из аналоговых сигналов с помощью периодической дискретизации, поэтому важно понять, как последовательности, полученные таким образом, связаны с исходным сигналом.

Рассмотрим аналоговый сигнал $x(t)$, имеющий спектр - преобразование Фурье (НВПФ) $X(j\Omega)$:

$$x(t) = (1/2\pi) \int_{-\infty}^{\infty} X(j\Omega) e^{j\Omega t} d\Omega; \quad (1.22)$$

$$X(j\Omega) = \int_{-\infty}^{\infty} x(t) e^{-j\Omega t} dt. \quad (1.23)$$

Говорят, что последовательность $x(n)$ со значениями $x(n) = x(nT) = x(t)|_{t=nT}$ получена из $x(t)$ периодической дискретизацией, а T называется *периодом дискретизации*. Величина, обратная T , называется *частотой дискретизации* $F_\delta = 1/T$ или *скоростью дискретизации*.

$X(j\Omega)$ -преобразование Фурье аналогового сигнала $x(t)$ с $X(e^{j\Omega T})$ -преобразованием Фурье последовательности $x(n)$. Для последовательности $x(n)$ пару ДВПФ (1.19) и (1.20) запишем в виде:

$$x(n) = (T/2\pi) \int_{-\pi/T}^{\pi/T} X(e^{j\Omega T}) e^{j\Omega nT} d\Omega,$$

$$X(e^{j\Omega T}) = \sum_{n=-\infty}^{\infty} x(n) e^{-j\Omega nT}$$

Идеализированную процедуру периодической дискретизации аналогового сигнала $x(t)$ представим в виде импульсного сигнала непрерывного времени $x_\delta(t)$:

$$x_\delta(t) = x(t) \sum_{n=-\infty}^{\infty} \delta(t - nT) = \sum_{n=-\infty}^{\infty} x(nT) \delta(t - nT)$$

который содержит только отсчеты $x(n) = x(nT)$. Преобразование Фурье $X_\delta(j\Omega)$ сигнала $x_\delta(t)$

$$X_\delta(j\Omega) = \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x(nT) \delta(t - nT) e^{-j\Omega t} dt = \sum_{n=-\infty}^{\infty} x(nT) \int_{-\infty}^{\infty} \delta(t - nT) e^{-j\Omega t} dt =$$

$$= \sum_{n=-\infty}^{\infty} x(nT) e^{-j\Omega nT} = X(e^{j\Omega T})$$

и, очевидно, совпадает с ДВПФ $X(e^{j\Omega T})$ последовательности $x(n)$. Поэтому $X(e^{j\Omega T})$ можно выразить в виде свертки преобразований Фурье сомножителей $x(t)$ и $\sum_{n=-\infty}^{\infty} \delta(t - nT)$,

$$X(e^{j\Omega T}) = \int_{-\infty}^{\infty} x(t) \sum_{n=-\infty}^{\infty} \delta(t - nT) e^{-j\Omega t} dt = \int_{-\infty}^{\infty} x(t) e^{-j\Omega t} dt \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(t - nT) e^{j\Omega nT} dt =$$

$$= \int_{-\infty}^{\infty} X(j\Omega) \sum_{n=-\infty}^{\infty} \delta(\Omega - n\Omega_\delta) d\Omega = \sum_{n=-\infty}^{\infty} X(j(\Omega - n\Omega_\delta)) \int_{-\infty}^{\infty} \delta(\Omega - n\Omega_\delta) d\Omega =$$

$$= \sum_{n=-\infty}^{\infty} X(j(\Omega - n\Omega_\delta)). \quad (1.27)$$

$$F_\delta \geq 2F_\epsilon$$

$$X(e^{j\Omega T}) = (1/T)X(j\Omega), \quad -\pi/T \leq \Omega \leq \pi/T, \quad (1.28)$$

т.е. спектр $X(e^{j\Omega T})$ совпадает со спектром $X(j\Omega)$ и, следовательно, по дискретной последовательности $x(n)$ можно восстановить исходный аналоговый сигнал $x(t)$. Это утверждение составляет содержание так называемой *теоремы отсчетов* (в отечественной литературе часто именуемой теоремой Котельникова). Общепринятая формулировка теоремы отсчетов такова: произвольный сигнал непрерывного времени, спектр которого не содержит частот выше F_ϕ Гц, может быть полностью восстановлен, если известны отсчеты

$$T \leq 1/2F_\phi$$

Сделаем замечание относительно выбора шкалы частот при спектральных представлениях дискретных последовательностей. В соотношениях (1.24) и (1.25), соответствующих паре ДВПФ, мы использовали аналоговые частоты Ω , измеряемые в радианах/секунду. Аналоговые частоты F , измеряемые в герцах (1/с), определяются как $F = \Omega/2\pi$. В соотношениях (1.19) и (1.20) для ДВПФ использовались *цифровые частоты* $\omega = \Omega T = \Omega/F_\phi$, измеряемые в радианах, которые можно трактовать как аналоговые,

именно цифровые частоты ω вместе с цифровыми безразмерными частотами $f = \omega/2\pi = F/F_\phi$.

дискретизации. На рис. 1.13а и 1.13б использованы шкалы цифровых частот ω .

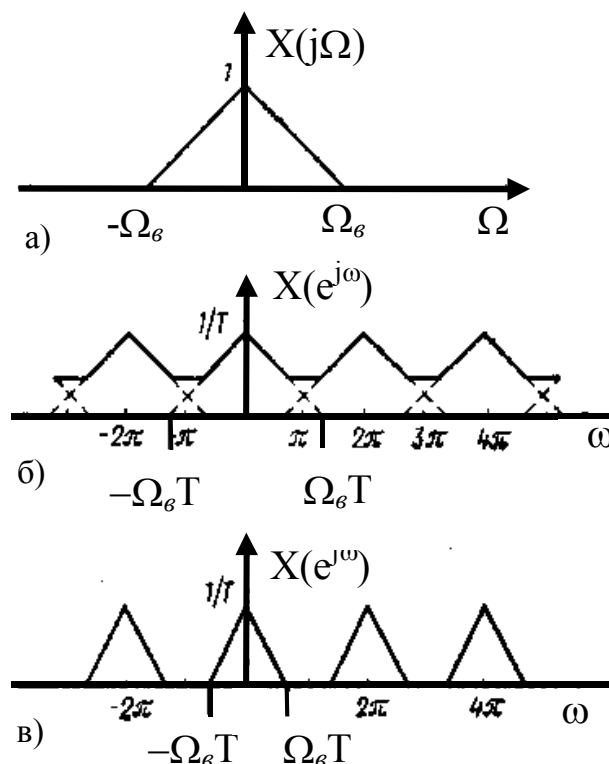


Рис. 1.13. Преобразование Фурье аналогового сигнала (а), преобразование Фурье дискретного сигнала, полученного при периодической дискретизации; период дискретизации велик, и поэтому периодически повторяющиеся преобразования Фурье аналогового сигнала перекрываются (б); период дискретизации мал настолько, что периодически повторяющиеся преобразования Фурье аналогового сигнала не перекрываются (в)

Найдем интерполяционную формулу для восстановления $x(t)$ по $x(n)$. Учитывая соотношения (1.22), (1.25) и (1.28) получим

$$\begin{aligned}
x(t) &= (1/2\pi) \int_{-\infty}^{\infty} X(j\Omega) e^{j\Omega t} d\Omega = (T/2\pi) \int_{-\pi/T}^{\pi/T} X(e^{j\Omega T}) e^{j\Omega t} d\Omega = \\
&= (T/2\pi) \int_{-\pi/T}^{\pi/T} \sum_{n=-\infty}^{\infty} x(n) e^{-j\Omega n T} e^{j\Omega t} d\Omega = \sum_{n=-\infty}^{\infty} x(n) (T/2\pi) \int_{-\pi/T}^{\pi/T} e^{j\Omega(t-nT)} d\Omega = \\
&= \sum_{n=-\infty}^{\infty} x(n) \frac{\sin[(\pi/T)(t-nT)]}{(\pi/T)(t-nT)}.
\end{aligned} \tag{1.29}$$

Представление аналогового сигнала в виде (1.29) справедливо только для функций с ограниченным спектром при достаточно малом T , т. е. отсутствии эффекта наложения. Если $\Omega_0 > \pi/T$ ($F_0 < 2F_0$), из-за наложения спектров восстановление аналогового сигнала без

аналогового сигнала по его отсчетам может быть выполнена пропусканием импульсного сигнала $x_0(t)$ через идеальный фильтр нижних частот, пропускающий без искажений полосу $-\pi/T \leq \Omega \leq \pi/T$.

$|\Omega| > \Omega_0$, можно уменьшить влияние эффекта наложения спектров, если аналоговый сигнал (ФНЧ) с целью ограничения его спектра диапазоном частот $|\Omega| < \pi/T$. При этом останутся лишь искажения, связанные с отбрасыванием хвостов спектра. Эта процедура низкочастотной фильтрации целесообразна и в случае, когда дискретизируемый сигнал смешан с широкополосной помехой, с тем, чтобы после дискретизации не ухудшалось отношение сигнал/помеха вследствие эффекта наложения.

могут быть представлены в виде $x(t) = a(t) \cos[\Omega_0 t + \varphi(t)]$ и спектр которых локализован в полосе частот $\Delta\Omega$ в окрестности средней частоты Ω_0 . Таким образом представляются, например, сигналы, получаемые при модуляции гармонического колебания с частотой Ω_0 по

образом могут быть представлены и любые другие узкополосные сигналы, для которых $(\Delta\Omega/\Omega_0) \ll 1$. Понятно, что такие сигналы могут быть дискретизованы как видеосигналы с верхней

дискретизовать, образуя отсчеты с периодом $T \leq 2\pi/\Delta\Omega$, где $\Delta\Omega$ - ширина спектра

$$\begin{aligned}
x(t) &= a(t) \cos[\Omega_0 t + \varphi(t)] = a(t) \cos \varphi(t) \cos \Omega_0 t - a(t) \sin \varphi(t) \sin \Omega_0 t = \\
&= a_c(t) \cos \Omega_0 t - a_s(t) \sin \Omega_0 t = \operatorname{Re} \{z(t) e^{j\Omega_0 t}\},
\end{aligned}$$

где $z(t) = a_c(t) + ja_s(t) = a(t) e^{j\varphi(t)}$ - комплексная огибающая радиосигнала. При известной

$z(t)$, которая является низкочастотным процессом с шириной спектра $\Delta\Omega$. Найдем связь

$$\begin{aligned}
x(t) &= (1/2) [z(t) e^{j\Omega_0 t} + z^*(t) e^{-j\Omega_0 t}] \\
X(j\Omega) &= (1/2) \{Z[j(\Omega - \Omega_0)] + Z^*[j(-\Omega - \Omega_0)]\}
\end{aligned}$$

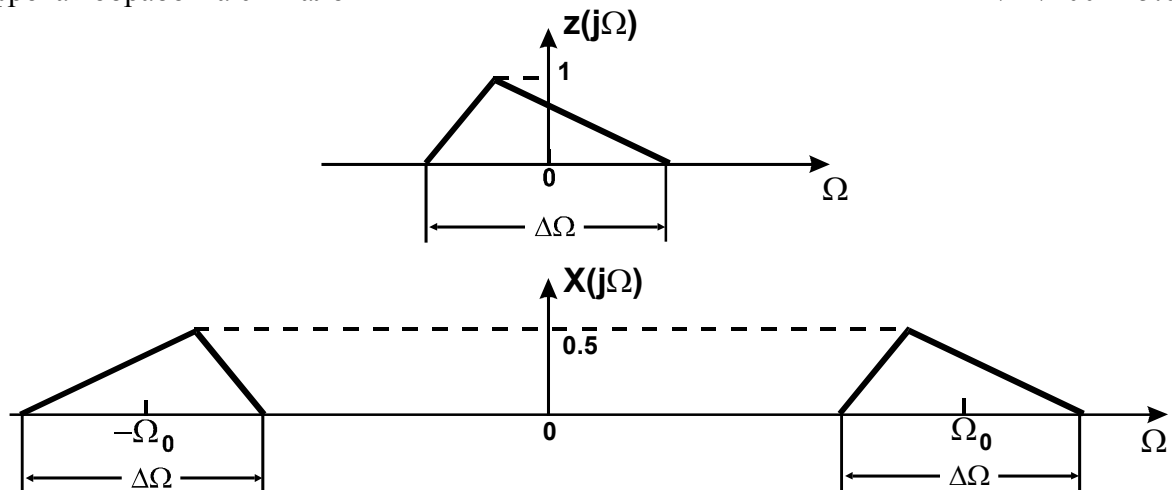


Рис.1.14.

Итак, при известной частоте Ω_0 дискретизация узкополосного сигнала эквивалентна $T \leq 2\pi / \Delta\Omega$

которым этот сигнал может быть восстановлен без искажений. Заметим, что выборки комплексной огибающей, т.е. комплексная последовательность $z(n) = z(t)|_{t=nT} = z(nT) = a_c(nT)$

$$\underset{s}{s} \quad \underset{t=nT}{/t=nT} \quad \underset{s}{s} \quad \underset{c}{c} \quad \underset{c}{c} \quad \underset{t=nT}{/t=nT} \quad \underset{c}{c}$$

квадратного разложения, с последующей их периодической дискретизацией. Каждый из каналов схемы состоит из перемножителя входного сигнала $x(t)$ с опорным – гармоническим колебанием с частотой Ω_0 и фильтра нижних частот Φ с полосой пропускания $0 - \Delta\Omega / 2$,

варианты получения выборок сигналов $a_c(t)$ и $a_s(t)$, например, непосредственной

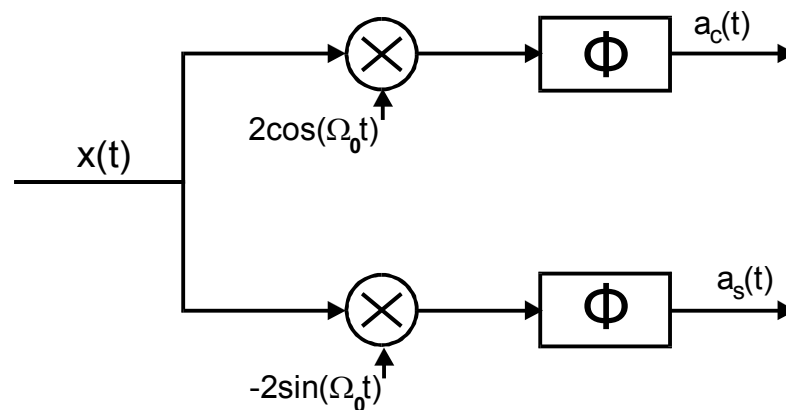


Рис.1.15. Блок квадратурного разложения аналогового радиосигнала.

1.8. Частотно-временные деформации дискретного сигнала

Для сигнала непрерывного времени свойство частотно-временной деформации состоит в следующем. Если спектр (преобразование Фурье) сигнала $x(t)$ равен $X(j\Omega)$, то для сигнала $x(Mt)$ спектр изменяется к виду $(1/M)X(j\Omega/M)$. Другими словами, растяжение сигнала во времени ($M < 1$) приводит к сжатию его спектра по оси частот и наоборот, сжатие сигнала во времени ($M > 1$) растягивает его спектр по оси частот, что также сопровождается соответствующим масштабированием спектров. Аналогичные свойства имеют место и для дискретного сигнала $x(n)$, полученного в результате периодической (с периодом T)

дискретизации сигнала $x(t)$. При этом сжатие и растяжение (удлинение) дискретного сигнала эквивалентны соответственно уменьшению либо увеличению частоты дискретизации $F_d = 1/T$. Рассмотрим эти процедуры подробнее.

1. Уменьшение частоты дискретизации в целое число M раз.

Эта процедура называется также *прореживанием* или *децимацией* дискретного сигнала $x(n)$. При этом новый дискретный сигнал $x_d(n)$, получается прореживанием исходного, т.е. сохранением лишь каждого M -го отсчета сигнала $x(n)$: $x_d(n) = x(Mn)$. Такой сигнал может быть получен в результате периодической (с периодом $T_1 = MT$) дискретизации сигнала $x(t)$, т.е. с частотой дискретизации $F_{d1} = F_d/M$. На рис.1.16 представлено условное графическое отображение прореживателя в M раз.



Рис.1.16

Перепишем соотношение (1.27), связывающее спектры дискретного сигнала $x(n)$ и сигнала непрерывного времени $x(t)$, в виде:

$$X(e^{j\omega}) = \frac{1}{T} \sum_{n=-\infty}^{\infty} X \left[j \left(\frac{\omega}{T} - \frac{2\pi n}{T} \right) \right]. \quad (1.30)$$

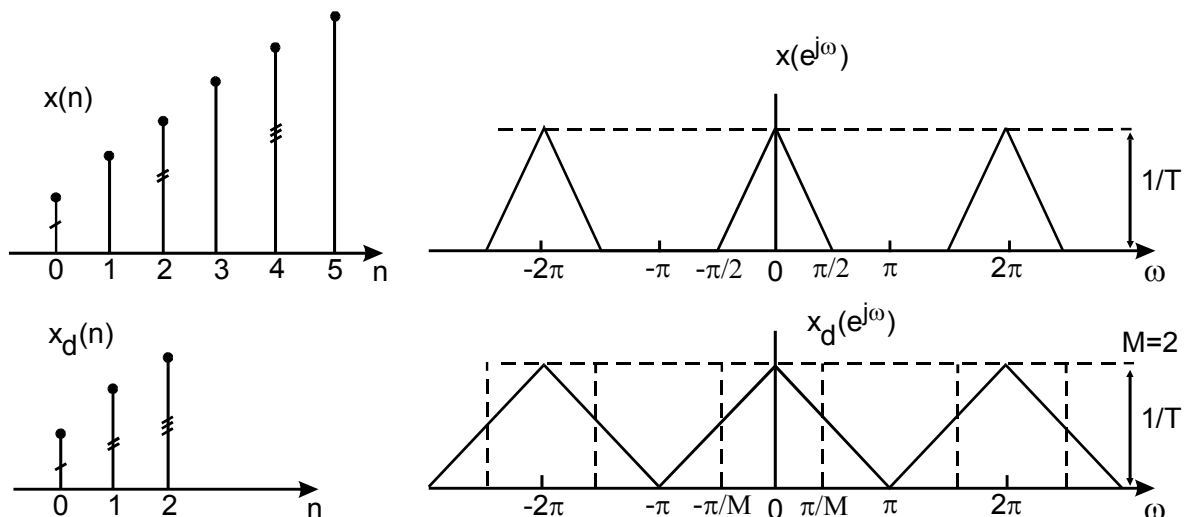


Рис1.17.

На рис.1.17 на основе (1.30) представлены графически спектры дискретных сигналов $x(n)$ и $x_d(n)$ для $M=2$. Чтобы гарантировать отсутствие эффекта наложения в связи с расширением спектра в M раз при прореживании дискретного сигнала, необходимо перед прореживанием пропустить его через дискретный фильтр нижних частот (ДФНЧ) с полосой пропускания $|\omega| \leq \pi/M$. Таким образом совокупность каскадно включенных ДФНЧ с полосой пропускания $|\omega| \leq \pi/M$ и прореживателя в M раз выполнит операцию уменьшения частоты дискретизации в M раз без эффекта наложения. Такую систему (рис.1.18) называют *компрессором* частоты дискретизации.

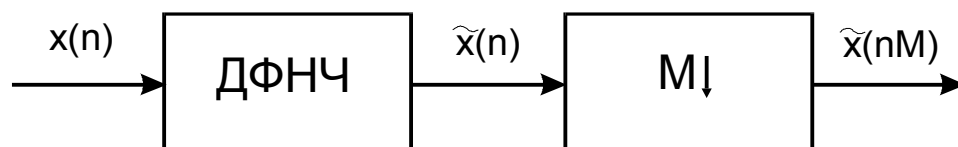


Рис.1.18.

2. Увеличение частоты дискретизации в целое число L раз.

Эту процедуру называют *интерполяцией* или восстановлением отсутствующих отсчетов дискретного сигнала. Для интерполяции дискретного сигнала $x(n)$ его сначала удлиняют в L раз путем вставления между каждыми двумя соседними отсчетами $L-1$ нулевых отсчетов. Такой сигнал $x_{уд}(n)$ можно записать в виде

$$x_{уд}(n) = \begin{cases} x(n/L), & n = kL, \\ 0, & n \neq kL. \end{cases}$$

Следовательно, спектр сигнала $x_{уд}(n)$ примет форму:

$$X_{уд}(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x_{уд}(n) e^{-j\omega n} = \sum_{n=kL} x(n/L) e^{-j\omega n} = \sum_{k=-\infty}^{\infty} x(k) e^{-j\omega Lk} = X(e^{j\omega L}). \quad (1.31)$$

Из (1.31) следует вывод о том, что при описанной выше операции удлинения дискретного сигнала, спектр сигнала $x_{уд}(n)$ сжимается по оси частот в L раз в сравнении со спектром исходного сигнала, а период повторения будет равен $2\pi/L$. Если затем пропустить $x_{уд}(n)$ через идеальный ДФНЧ с полосой пропускания $|\omega| \leq \pi/L$ и усилением L , то лишние спектральные полосы будут удалены, а выходной сигнал фильтра будет точно соответствовать дискретному сигналу $x_{ин}(n)$ с периодом дискретизации T/L , т.е. частота дискретизации его увеличится в L раз, а нужные отсчеты будут восстановлены. На рис.1.19 представлены графически: а) исходный сигнал и его спектр; б) сигнал, удлинённый добавлением нулевых отсчетов и его спектр; в) интерполированный сигнал на выходе ДФНЧ и его спектр. На рис.1.20 представлена структура интерполатора (*экспандера* частоты дискретизации), увеличивающего частоту дискретизации в L раз.

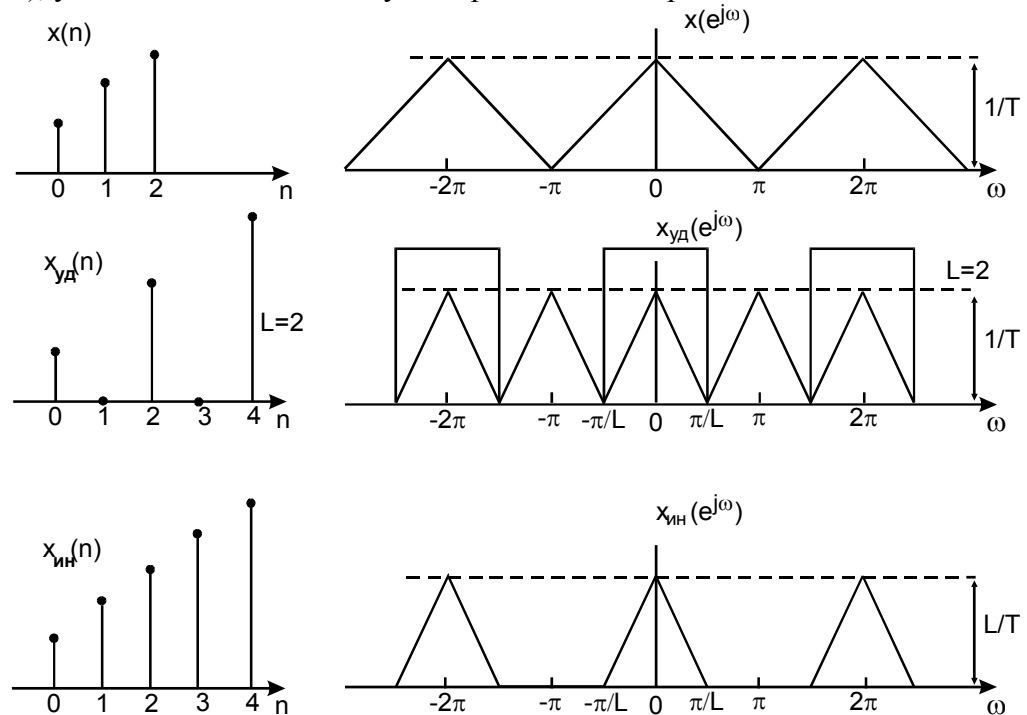


Рис1.19.

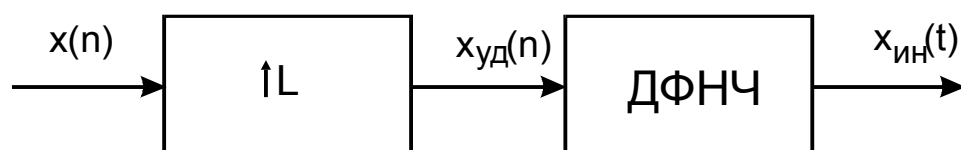


Рис.1.20

Если необходимо изменить частоту дискретизации исходного дискретного сигнала в L/M (рациональное число) раз, то такая операция может быть выполнена каскадным соединением интерполятора в L раз и дециматора в M раз, как это представлено на рис.1.21. Два последовательно включенных дискретных фильтра нижних частот ДФНЧ1 и ДФНЧ2 могут быть заменены одним с меньшей частотой среза.

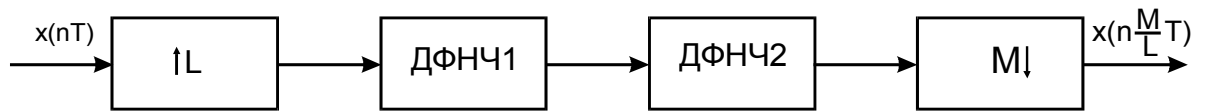


Рис.1.21.

Рассмотрим примеры применения дециматоров и интерполяторов в цифровой аудиотехнике. При преобразовании в цифровую форму аналогового музыкального сигнала полагают, что полоса его соответствует интервалу частот $22\text{кГц} \geq f \geq 0$ и минимальная частота дискретизации равна 44кГц . Перед дискретизацией необходима аналоговая низкочастотная фильтрация (ФНЧ) для исключения эффекта наложения и внеполосного шума. При этом ФНЧ должен иметь высокую равномерность частотной характеристики в полосе пропускания и узкую переходную полосу от полосы пропускания к полосе задерживания. Такие фильтры, как правило, имеют очень нелинейную фазо-частотную (ФЧХ) характеристику к края полосы пропускания (у частоты 22кГц), что считается недопустимым для высококачественного воспроизведения музыки. Распространенный способ решения этой проблемы состоит в повышении вдвое (иногда вчетверо) частоты дискретизации. При этом аналоговый ФНЧ может иметь более широкую переходную полосу в интервале частот $22\text{кГц} - 44\text{кГц}$ и нелинейность ФЧХ оказывается приемлемо малой. На рис.1.22 графически отображены: а) спектр аналогового сигнала; б) требуемая амплитудно-частотная характеристика ФНЧ при частоте дискретизации 44кГц ; в) требуемая амплитудно-частотная характеристика ФНЧ при частоте дискретизации 88кГц .

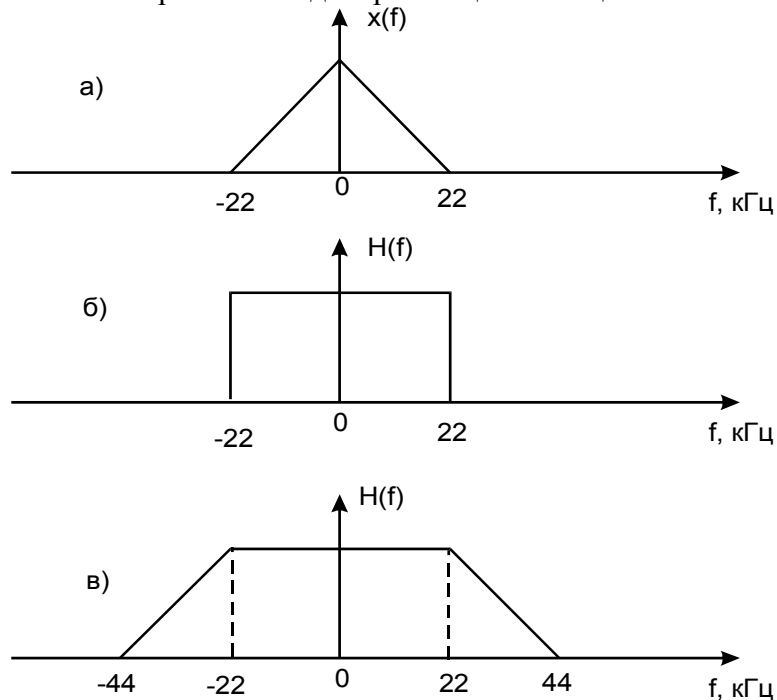


Рис.1.22.

Полученный на выходе аналого-цифрового преобразователя (АЦП) цифровой сигнал с частотой дискретизации 88кГц далее обрабатывается цифровым ФНЧ (ЦФНЧ) с требуемой крутизной переходной полосы и прореживается вдвое для получения нужного сигнала с частотой дискретизации 44кГц , как это представлено на рис.1.23. Заметим здесь, что реализация ЦФНЧ с линейной ФЧХ, как это будет показано в дальнейшем при рассмотрении

характеристик цифровых фильтров, не вызывает принципиальных затруднений. Для этой цели используются КИХ-фильтры с симметричной импульсной характеристикой.

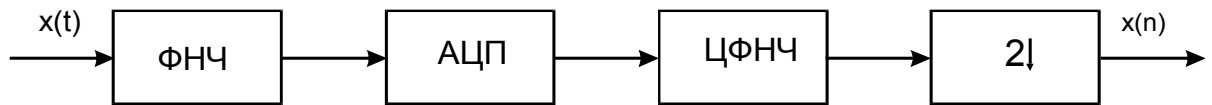


Рис.1.23.

Аналогичная проблема возникает в ступени цифро-аналогового преобразования, когда цифровой музыкальный сигнал должен быть преобразован в аналоговый путем обработки в ФНЧ. Для этой цели нужен аналоговый ФНЧ с частотой среза 22кГц и узкой переходной полосой и, следовательно, он будет иметь сильно нелинейную ФЧХ, что недопустимо. Для решения этой проблемы используют интерполятор, повышающий частоту дискретизации цифрового сигнала вдвое, после чего такой сигнал преобразуется в аналоговый в ЦАП с аналоговым ФНЧ, переходная полоса которого может быть в интервале частот 22кГц – 44кГц, а нелинейность ФЧХ будет приемлемо малой. Блок-схема такой обработки представлена на рис.1.24.

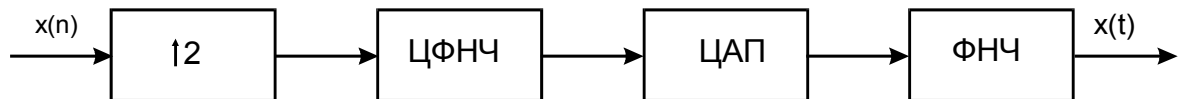


Рис.1.24.

2. Z – преобразование

В теории систем непрерывного времени преобразование Лапласа рассматривается как обобщение преобразования Фурье и широко используется в качестве математического инструмента для вычисления отклика линейных с постоянными параметрами систем умеренной сложности на конкретные относительно простые воздействия. Аналогичную роль для дискретных сигналов и систем играет аппарат Z-преобразования, которое можно считать обобщением дискретного во времени преобразования Фурье.

2.1. Прямое Z – преобразование

Для последовательности $x(n)$, заданной при всех n , z-преобразование определяется следующим образом:

$$X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n}, \quad (2.1)$$

где z - комплексная переменная. Если представить комплексную переменную z в показательной форме $z=re^{j\omega}$, то (2.1) можно интерпретировать как ДВПФ:

$$X(re^{j\omega}) = \sum_{n=-\infty}^{\infty} x(n)(re^{j\omega})^{-n} = \sum_{n=-\infty}^{\infty} x(n)r^{-n}e^{-j\omega n}. \quad (2.2)$$

Действительно, согласно (2.2) z-преобразование $x(n)$ можно интерпретировать как ДВПФ последовательности $x(n)$, умноженной на экспоненциальную последовательность r^{-n} . Очевидно, что для $|z| = r = 1$, т.е. на окружности единичного радиуса в комплексной z -плоскости, z-преобразование $x(n)$ совпадает с ДВПФ последовательности $x(n)$. Соотношение (2.1) называют *двусторонним z-преобразованием*, а в большинстве практических применений используют *одностороннее z-преобразование*, определяемое в виде:

$$X(z) = \sum_{n=0}^{\infty} x(n)z^{-n}. \quad (2.3)$$

При $x(n)=0$, для $n<0$, т.е. для физически реализуемых последовательностей эти преобразования эквивалентны.

Для любой последовательности множество тех значений z , для которых z-преобразование сходится ($|X(z)|<\infty$), называется *областью сходимости*. Ряд (2.1) сходится, если выполняется соотношение

$$\sum_{n=-\infty}^{\infty} |x(n)r^{-n}| < \infty, \quad (2.4)$$

следующее из условия сходимости ДВПФ. Для последовательностей, растущих не быстрее, чем экспонента, z-преобразование будет сходиться для всех z , находящихся вне некоторого круга в комплексной z -плоскости, радиус r_0 которого называют радиусом сходимости. Соответствие между $x(n)$ и $X(z)$ взаимнооднозначное т.е. каждому $x(n)$ соответствует только одно $X(z)$, определенное для $|z|>r_0$ и обратно.

Рассмотрим примеры нахождения z-преобразований некоторых полезных последовательностей.

1). Для единичного импульса $x(n)=\delta(n)$, очевидно, что $X(z)=1$.

2). Найдем z-преобразование единичной ступенчатой последовательности $x(n)=u(n)$.

Поскольку $x(n)=0$ везде, кроме $n \geq 0$, где $x(n)=1$, то

$$X(z) = \sum_{n=0}^{\infty} z^{-n} = \frac{1}{1-z^{-1}} = \frac{z}{z-1},$$

причем $X(z)$ сходится при $|z|>1$, так как $X(z)$ имеет единственную особую точку (полюс) $z=1$, в которой $X(1)=\infty$.

3). Найдем z -преобразование действительной экспоненциальной последовательности $x(n)=a^n u(n)$. Получим

$$X(z) = \sum_{n=0}^{\infty} a^n z^{-n} = \sum_{n=0}^{\infty} (az^{-1})^n = \frac{1}{1 - az^{-1}} = \frac{z}{z - a}.$$

$X(z)$ сходится при $|z| > a$, так как $X(z)$ имеет единственную особую точку (полюс) $z = a$.

4). Найдем z -преобразование комплексной экспоненциальной последовательности $x(n)=e^{j\omega n} u(n)$. Получим

$$X(z) = \sum_{n=0}^{\infty} e^{j\omega n} z^{-n} = \sum_{n=0}^{\infty} (e^{j\omega} z^{-1})^n = \frac{1}{1 - e^{j\omega} z^{-1}} = \frac{z}{z - e^{j\omega}}.$$

$X(z)$ сходится при $|z| > 1$, так как $X(z)$ имеет единственную особую точку (полюс) $z = e^{j\omega}$.

Часто удобно изображать z -преобразование графически с помощью диаграммы нулей и полюсов в z -плоскости. В случае, когда $X(z)$ представляется рациональной функцией т.е. отношением полиномов переменной z , нули – это корни полинома числителя, при которых $X(z)=0$, а полюсы – это корни полинома знаменателя, при которых $X(z)=\infty$. Так, для $X(z)$, полученного в примере 4, нуль соответствует $z = 0$, а полюс $z = e^{j\omega}$. На рис.2.1 нуль обозначен кружком, полюс крестиком, а область сходимости обозначена штриховкой и включает все z на плоскости, удовлетворяющие условию $|z| > 1$.

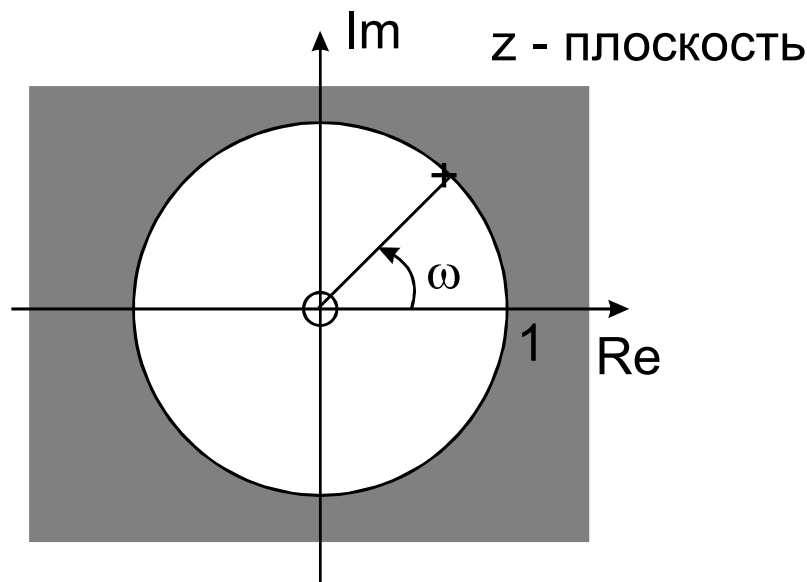


Рис.2.1.

2.2. Обратное Z – преобразование

Обратный переход от z -преобразования $X(z)$ к последовательности $x(n)$ определяется соотношением

$$x(n) = \frac{1}{2\pi j} \oint_C X(z) z^{n-1} dz. \quad (2.5)$$

Это обратное z -преобразование содержит интеграл по любому замкнутому контуру с направлением обхода против часовой стрелки, расположенному в области сходимости и окружающему начало координат.

При выводе соотношения (2.5) используется теорема Коши, согласно которой

$$\frac{1}{2\pi j} \oint_C z^{k-1} dz = \delta(k) = \begin{cases} 1, & k = 0; \\ 0, & k \neq 0. \end{cases} \quad (2.6)$$

Домножая (2.1) на z^{k-1} и интегрируя по замкнутому контуру, окружающему начало координат, получим

$$\frac{1}{2\pi j} \oint_C X(z) z^{k-1} dz = \frac{1}{2\pi j} \oint_C \sum_{n=-\infty}^{\infty} x(n) z^{-n+k-1} dz = \sum_{n=-\infty}^{\infty} x(n) \frac{1}{2\pi j} \oint_C z^{-n+k-1} dz = \sum_{n=-\infty}^{\infty} x(n) \delta(k-n) = x(k),$$

откуда следует (2.5).

Для рациональных z -преобразований контурные интегралы вида (2.5) удобно вычислять с помощью теоремы о вычетах, согласно которой

$$x(n) = \frac{1}{2\pi j} \oint_C X(z) z^{n-1} dz = \sum_i \operatorname{res}_{z=p_i} [X(z) z^{n-1}], \quad (2.7)$$

где $\operatorname{res}_{z=p_i} [X(z) z^{n-1}]$ - вычет подынтегральной функции в полюсе $z=p_i$. В общем случае, если $X(z) z^{n-1}$ - рациональная функция z , то ее можно записать как

$$X(z) z^{n-1} = \frac{\psi(z)}{(z-p_i)^s},$$

когда $X(z) z^{n-1}$ имеет полюс порядка s в точке $z=p_i$, $\psi(z)$ не имеет полюсов в $z=p_i$. Вычет $X(z) z^{n-1}$ в точке $z=p_i$ определяется формулой

$$\operatorname{res}_{z=p_i} [X(z) z^{n-1}] = \frac{1}{(s-1)!} \left[\frac{d^{s-1} \psi(z)}{dz^{s-1}} \right]_{z=p_i}.$$

В частности, если p_i - полюс первого порядка, то $\operatorname{res}_{z=p_i} [X(z) z^{n-1}] = \psi(p_i)$.

Как пример рассмотрим вычисление обратного преобразования от $X(z) = \frac{1}{1-az^{-1}} = \frac{z}{z-a}$, $|z|>a$. Используя (2.7) получим

$$x(n) = \frac{1}{2\pi j} \oint_C \frac{z^{n-1}}{1-az^{-1}} dz = \frac{1}{2\pi j} \oint_C \frac{z^n}{z-a} dz.,$$

где контур интегрирования C является окружностью с радиусом больше a . Тогда при $n \geq 0$ контур интегрирования содержит только один полюс первого порядка в точке $z=a$. Следовательно, при $n \geq 0$ $x(n) = a^n$. При $n < 0$ в точке $z=0$ имеется кратный полюс, порядок которого зависит от n . При $n=-1$ этот полюс имеет первый порядок с вычетом, равным $(-a^{-1})$. Вычет в полюсе $z=a$ равен a^{-1} . Следовательно, сумма вычетов равна нулю и поэтому $x(-1)=0$. Продолжая эту процедуру, можно проверить, что в этом примере $x(n)=0$ при $n < 0$. Поэтому $x(n)=a^n u(n)$ для всех n .

Во многих случаях вычисления по формуле (2.7) оказываются сложными. Часто помогает ряд специальных приемов, которые рассмотрены ниже.

1). Разложение на простые дроби.

Например, для рационального z -преобразования с однократными полюсами $X(z)$ можно представить в форме

$$X(z) = \sum_{i=1}^N \frac{A_i}{1-p_i z^{-1}}, \text{ тогда } x(n) = \sum_{i=1}^N A_i p_i^n u(n) \text{ для всех } n.$$

2). Разложение в степенной ряд.

Если z -преобразование имеет вид степенного ряда, можно заметить, что значение $x(n)$ последовательности есть коэффициент при z^{-n} в этом ряде $X(z) = \sum_{n=-\infty}^{\infty} x(n) z^{-n}$. Если $X(z)$ дается в замкнутом виде, то часто можно вывести соответствующий степенной ряд или использовать известное разложение в ряд.

Например, для $X(z) = \frac{1}{1-az^{-1}} = \frac{z}{z-a}$, $|z|>a$ представление в виде степенного ряда можно получить непосредственным делением числителя z на знаменатель $z-a$:

$$X(z) = \frac{1}{1-az^{-1}} = \frac{z}{z-a} = 1 + az^{-1} + a^2 z^{-2} + a^3 z^{-3} + \dots = \sum_{n=0}^{\infty} a^n z^{-n}.$$

2.3. Свойства Z – преобразования

Рассмотрим некоторые наиболее важные свойства z -преобразования, полезные при его применении.

1. Линейность.

z -преобразование есть линейное преобразование, что означает справедливость для него принципа суперпозиции. Если z -преобразования последовательностей $y(n)$, $x_1(n)$, $x_2(n)$ равны $Y(z)$, $X_1(z)$, и $X_2(z)$ соответственно, то для любых действительных a и b справедливы соотношения: для $y(n)=ax_1(n)+bx_2(n)$, $Y(z)=aX_1(z)+bX_2(z)$.

2. Сдвиг последовательности (задержка).

Если z -преобразования последовательностей $y(n)$, $x(n)$ равны $Y(z)$, $X(z)$ соответственно, то для $y(n)=x(n-n_0)$, где n_0 -целое число, справедливо соотношение:

$$Y(z) = z^{-n_0} X(z) \quad (2.8)$$

Так, при задержке на один такт $y(n)=x(n-1)$, $Y(z)=z^{-1}X(z)$ т.е. z -преобразование исходной последовательности умножается на z^{-1} . Поэтому иногда пользуются оператором задержки на такт $z^{-1}\{\}$, понимая под ним следующее соотношение:

$$z^{-1}\{x(n)\}=x(n-1).$$

Для одностороннего z -преобразования свойство задержки следует рассмотреть подробнее. Пусть $y(n)=x(n-n_0)$, ($n_0>0$ и целое), тогда, если $X(z) = \sum_{n=0}^{\infty} x(n)z^{-n}$, то

$$\begin{aligned} Y(z) &= \sum_{n=0}^{\infty} x(n-n_0)z^{-n} = z^{-n_0} \sum_{n=0}^{\infty} x(n-n_0)z^{-(n-n_0)} = z^{-n_0} \sum_{m=-n_0}^{\infty} x(m)z^{-m} = z^{-n_0} \left[\sum_{m=0}^{\infty} x(m)z^{-m} + \right. \\ &+ \sum_{m=-1}^{-n_0} x(m)z^{-m} \left. \right] = z^{-n_0} \left[X(z) + \sum_{m=-1}^{-n_0} x(m)z^{-m} \right] = z^{-n_0} X(z) + x(-n_0) + x(-n_0+1)z^{-1} + \\ &+ \dots + x(-1)z^{-(n_0-1)}. \end{aligned} \quad (2.9)$$

Если $x(n)=0$ при $n<0$, то $Y(z)=z^{-n_0}X(z)$, т.е. свойство задержки физически реализуемых последовательностей сохраняется и для одностороннего z -преобразования. Однако для опережающего сдвига физически реализуемой последовательности, т.е. для случая, когда $y(n)=x(n+n_0)$, ($n_0>0$ и целое), аналогично (2.8) легко получить соотношение:

$$Y(z) = z^{n_0} \left[X(z) - \sum_{m=0}^{n_0-1} x(m)z^{-m} \right] \quad (2.10)$$

Соотношения (2.8) и (2.9) используются при решении разностных уравнений с применением одностороннего z -преобразования.

3. Свертка последовательностей.

Пусть $y(n) = \sum_{k=-\infty}^{\infty} x(k)h(n-k)$. Если z -преобразования последовательностей $y(n)$, $x(n)$, $h(n)$ равны $Y(z)$, $X(z)$, и $H(z)$ соответственно, то справедливо соотношение:

$$Y(z)=X(z)Y(z),$$

которое означает, что z -преобразование свертки равно произведению z -преобразований свертываемых последовательностей. Действительно,

$$\begin{aligned} Y(z) &= \sum_{n=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} x(k)h(n-k)z^{-n} = \sum_{k=-\infty}^{\infty} x(k) \sum_{n=-\infty}^{\infty} h(n-k)z^{-n} = \sum_{k=-\infty}^{\infty} x(k) \left[\sum_{m=-\infty}^{\infty} h(m)z^{-m} \right] z^{-k} = \\ &= \sum_{k=-\infty}^{\infty} x(k)z^{-k} \sum_{m=-\infty}^{\infty} h(m)z^{-m} = X(z)H(z). \end{aligned} \quad (2.11)$$

4. Перемножение последовательностей.

Пусть $y(n)=x_1(n)x_2(n)$. Если z -преобразования последовательностей $y(n)$, $x_1(n)$, $x_2(n)$ равны $Y(z)$, $X_1(z)$, и $X_2(z)$ соответственно, то справедливо соотношение:

$$Y(z) = \frac{1}{2\pi j} \oint_C X_1(v) X_2(z/v) v^{-1} dv \quad (2.12)$$

Это соотношение называют *комплексной сверткой*. При $z=e^{j\omega}$ и $v=e^{j\theta}$ из (2.12) имеем

$$Y(e^{j\omega}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X_1(e^{j\theta}) X_2(e^{j(\omega-\theta)}) d\theta = X_1(e^{j\omega}) \otimes X_2(e^{j\omega}). \quad (2.12a)$$

Следовательно, ДВПФ произведения последовательностей есть периодическая (круговая) свертка ДВПФ сомножителей.

Доказательство (2.12):

$$\begin{aligned} Y(z) &= \sum_{n=-\infty}^{\infty} x_1(n) x_2(n) z^{-n} = \frac{1}{2\pi j} \sum_{n=-\infty}^{\infty} x_2(n) \oint_C X_1(v) (z/v)^{-n} v^{-1} dv = \\ &= \frac{1}{2\pi j} \oint_C \sum_{n=-\infty}^{\infty} x_2(n) (z/v)^{-n} v^{-1} X_1(v) dv = \frac{1}{2\pi j} \oint_C X_1(v) X_2(z/v) v^{-1} dv \end{aligned}$$

Важным следствием (2.12a) является равенство Парсеваля:

$$\sum_{n=-\infty}^{\infty} |x(n)|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |X(e^{j\omega})|^2 d\omega \quad (2.13)$$

Таблица 2.1. Краткая таблица одностороннего z -преобразования $X(z) = \sum_{n=0}^{\infty} x(n) z^{-n}$

Последовательность	z -преобразование
$x(n), n \geq 0$	$X(z)$
$\delta(n)$	1
$u(n-N)$	$z^{-N}/(1-z^{-1})$
a^n	$1/(1-az^{-1})$
n	$z^{-1}/(1-z^{-1})^2$
na^n	$az^{-1}/(1-az^{-1})^2$
$a^n \cos(n\theta)$	$(1-a\cos\theta z^{-1})/(1-2a\cos\theta z^{-1}+a^2 z^{-2})$
$a^n \sin(n\theta)$	$a\sin\theta z^{-1}/(1-2a\cos\theta z^{-1}+a^2 z^{-2})$

Таблица 2.2. Наиболее важные свойства одностороннего z -преобразования

Последовательность	z -преобразование
$x(n), n \geq 0$	$X(z)$
$y(n), n \geq 0$	$Y(z)$
$ax(n) + by(n)$	$aX(z) + bY(z)$
$x(n+1)$	$z[X(z) - x(0)]$
$x(n-N) u(n-N)$	$z^{-N} X(z), N \geq 0$
$a^n x(n)$	$X(a^{-1} z)$
$nx(n)$	$-z(dX(z)/dz)$
$x^*(n)$	$X^*(z^*)$
$x(-n)$	$X(1/z)$
$x(n) * y(n)$	$X(z) Y(z)$
$x(n)y(n)$	$\frac{1}{2\pi j} \oint_C X(v) Y(z/v) v^{-1} dv$

Заметим, что свойства двустороннего z -преобразования совпадают со свойствами одностороннего z -преобразования за исключением опережающего сдвига.

2.4. Решение разностных уравнений с применением Z – преобразования

Разностные уравнения обычно определены при $n \geq 0$ и имеют набор начальных условий. Рассмотрим процедуру нахождения общего решения с использованием

одностороннего z -преобразования на примере разностного уравнения 1-го порядка

$$y(n) = ay(n-1) + x(n)$$

с начальным условием $y(-1)=k$. Пусть входной сигнал есть экспоненциальная последовательность $x(n) = b^n u(n)$. Найдём одностороннее z -преобразование обеих частей уравнения:

$$\sum_{n=0}^{\infty} y(n)z^{-n} = a \sum_{n=0}^{\infty} y(n-1)z^{-n} + \sum_{n=0}^{\infty} x(n)z^{-n}.$$

Воспользуемся свойством (2.9), связанным с задержкой последовательности, получим:

$$Y(z) = az^{-1}Y(z) + ay(-1) + X(z),$$

откуда

$$Y(z) = \frac{X(z) + ay(-1)}{1 - az^{-1}}.$$

Поскольку для $x(n) = b^n u(n)$, $X(z) = \frac{1}{1 - bz^{-1}}$, то

$$Y(z) = \frac{ak}{1 - az^{-1}} + \frac{1}{(1 - az^{-1})(1 - bz^{-1})}.$$

Разлагая второе слагаемое на простые дроби, получим

$$Y(z) = \frac{ak}{1 - az^{-1}} - \frac{a/(b-a)}{(1 - az^{-1})} + \frac{b/(b-a)}{(1 - bz^{-1})}.$$

Находя обратное z -преобразование, получим

$$y(n) = \left[ka^{n+1} - \frac{a^{n+1}}{b-a} + \frac{b^{n+1}}{b-a} \right] u(n).$$

Первая компонента в скобках представляет составляющую отклика, определяемую начальными условиями, или реакцию при нулевом входном воздействии (РНВ). Остальные компоненты представляют реакцию при нулевом начальном состоянии (РНС), причем второе слагаемое – собственные колебания, вызванные входным воздействием, а третье слагаемое – вынужденные колебания.

При нулевых начальных условиях ($k=0$) и $b=1$ т.е. $x(n)=u(n)$ переходная характеристика системы принимает вид:

$$y(n) = \frac{1 - a^{n+1}}{1 - a} u(n).$$

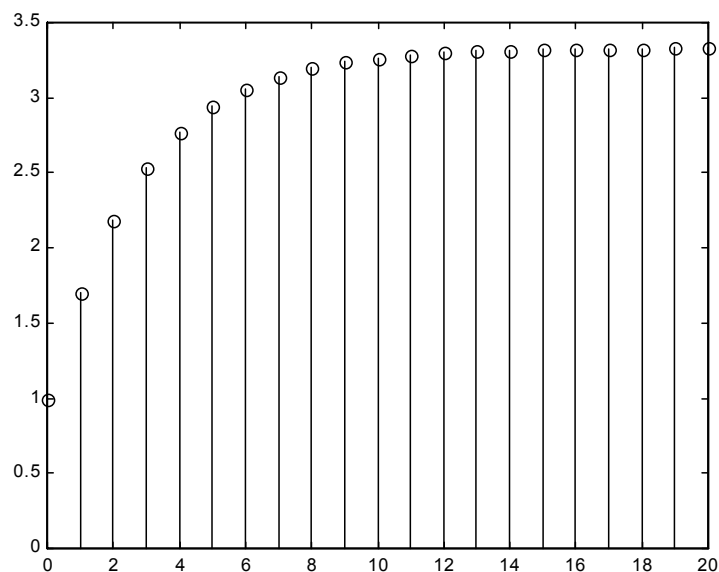


Рис.2.2. Переходная характеристика системы ($a=0,7$)

2.5. Передаточная (системная) функция

Выше мы рассмотрели описание линейных систем с постоянными параметрами с помощью частотной характеристики системы - преобразования Фурье импульсной характеристики. Показано, что в частотной области соотношение между входным и выходным сигналами получается просто умножением преобразования Фурье входного сигнала на преобразование Фурье импульсной характеристики.

Более общим образом можно описать линейные с постоянными параметрами системы с помощью z-преобразования импульсной характеристики. Обозначая $x(n)$, $y(n)$ и $h(n)$ вход, выход и импульсную характеристику соответственно и $X(z)$, $Y(z)$ и $H(z)$ их z-преобразования и используя результаты предыдущего раздела, получим из свертки $y(n) = x(n) * h(n)$ соотношение

$$Y(z) = X(z)H(z). \quad (2.14)$$

Как и в случае преобразования Фурье, соотношение между входом и выходом для линейных инвариантных к сдвигу систем получается умножением z-преобразований входного сигнала и импульсной характеристики.

Часто z-преобразование импульсной характеристики называется *передаточной* или *системной функцией*. Передаточная функция на единичной окружности (т. е. при $|z| = 1$) является частотной характеристикой системы.

Было показано, что необходимым и достаточным условием устойчивости системы является абсолютная суммируемость импульсной характеристики $h(n)$. Область сходимости z-преобразования импульсной характеристики $h(n)$ определяется теми значениями z , при которых $h(n)z^{-n}$ — абсолютно суммируемая

включает единичную окружность, система устойчива и наоборот. Кроме того, мы можем утверждать, что для устойчивой и физически реализуемой системы: область сходимости будет, включать единичную окружность и всю z-плоскость вне единичной окружности, включая $z = \infty$.

коэффициентами, то ее передаточная функция является отношением полиномов. Чтобы убедиться в этом, рассмотрим систему, вход и выход которой удовлетворяют общему разностному уравнению N-го порядка

$$\sum_{k=0}^N a(k)y(n-k) = \sum_{k=0}^M b(k)x(n-k)$$

Применяя z-преобразование к обеим частям разностного уравнения с учетом свойства линейности и свойства задержки, получим

$$\sum_{k=0}^N a(k)z^{-k}Y(z) = \sum_{k=0}^M b(k)z^{-k}X(z).$$

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^M b(k)z^{-k}}{\sum_{k=0}^N a(k)z^{-k}}. \quad (2.15)$$

Эта формула дает конкретное выражение для передаточной функции, и коэффициенты полиномов в числителе и знаменателе являются соответственно коэффициентами в правой и левой частях разностного уравнения.

Так как выражение (2.15) есть отношение полиномов от z^{-1} , то его можно записать в виде

$$H(z) = A \frac{\prod_{i=1}^M (1 - z_i z^{-1})}{\prod_{i=1}^N (1 - p_i z^{-1})} \quad (2.16)$$

Каждый из сомножителей $(1 - z_i z^{-1})$ в числителе (2.16) дает нуль при $z = z_i$ и полюс при

Следовательно, с точностью до скалярного множителя A в (2.16) передаточная функция может быть полностью описана картиной полюсов и нулей в z -плоскости.

крута и область сходимости будет содержать единичную окружность. По этой причине при описании передаточной функции диаграммой полюсов и нулей в z -плоскости удобно изображать также единичную окружность, чтобы было видно расположение полюсов относительно этой окружности.

Пример. В качестве простого примера рассмотрим физически реализуемую систему, описываемую разностным уравнением $y(n) = ay(n-1) + x(n)$. Передаточная функция равна

$$H(z) = 1/(1 - az^{-1})$$

и в силу предположения о физической реализуемости область сходимости определяется неравенством $|z| > |a|$, откуда следует, что импульсная характеристика равна $h(n) = a^n u(n)$.

В частном случае, когда $N=0$ в (2.15) или (2.16), система не имеет полюсов, за исключением точки $z=0$, и ее импульсная характеристика имеет конечную длительность. При $N>0$ система имеет полюсы, каждый из которых прибавляет экспоненциальную последовательность к импульсной характеристике. Таким образом, если передаточная функция имеет полюсы, то импульсная характеристика имеет бесконечную протяженность.

Одним из преимуществ представления передаточной функции посредством ее полюсов и нулей является то, что оно дает полезный геометрический способ получения частотной характеристики системы. Вспомним, что отклик системы на синусоидальное возбуждение описывается частотной характеристикой, т. е. поведением передаточной функции на единичной окружности. В частности, отклик на выходе является синусоидальным с той же частотой, что и на входе, а амплитуда на выходе равна амплитуде на входе, умноженной на модуль передаточной функции на частоте возбуждения. Фазовый сдвиг выхода относительно входа равен аргументу передаточной функции на частоте возбуждения.

Чтобы определить передаточную функцию на единичной окружности, нужно подставить $z = e^{j\omega}$ в (2.16). Таким образом,

$$H(e^{j\omega}) = A \frac{\prod_{i=1}^M (1 - z_i e^{-j\omega})}{\prod_{i=1}^N (1 - p_i e^{-j\omega})}.$$

Представляя $H(e^{j\omega}) = |H(e^{j\omega})| e^{j\arg[H(e^{j\omega})]}$, получим

$$|H(e^{j\omega})| = |A| \frac{\prod_{i=1}^M |1 - z_i e^{-j\omega}|}{\prod_{i=1}^N |1 - p_i e^{-j\omega}|} = |A| \frac{\prod_{i=1}^M |e^{j\omega} - z_i|}{\prod_{i=1}^N |e^{j\omega} - p_i|}, \quad (2.17)$$

$$\arg H(e^{j\omega}) = \arg A + \sum_{i=1}^M \arg(1 - z_i e^{-j\omega}) - \sum_{i=1}^N \arg(1 - p_i e^{-j\omega}). \quad (2.18)$$

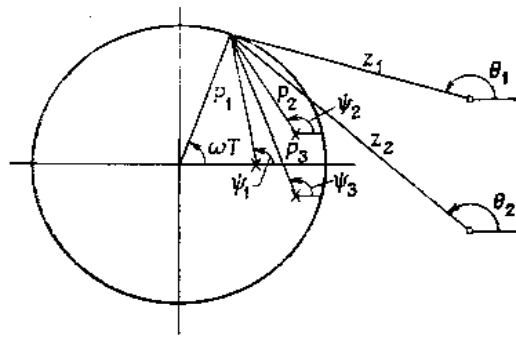


Рис.2.3. Геометрическая интерпретация измерения частотной характеристики.

Геометрическая интерпретация соотношений (2.17) — (2.18) дана на рис.2.3. Из точки $z = e^{j\omega}$ находящейся на единичной окружности, во все нули и полюсы проведены векторы. По их величине определяется модуль передаточной функции на заданной частоте ω , а по их углам — фаза. В примере на рис.2.3 имеются три полюса ($N = 3$) и два нуля ($M = 2$), а коэффициент A равняется 1, поэтому (рис.2.3)

$$|H(e^{j\omega})| = \frac{z_1 z_2}{P_1 P_2 P_3},$$

$$\arg H(e^{j\omega}) = \theta_1 + \theta_2 - (\psi_1 + \psi_2 + \psi_3).$$

Для определения передаточной функции на всех частотах $0 \leq \omega \leq \pi$ необходимо перемещать z по единичной окружности против часовой стрелки из точки $z = +1$ до точки $z = -1$.

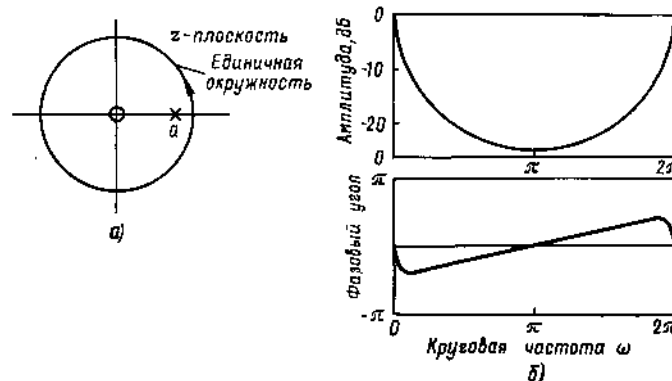


Рис.2.4. Диаграмма полюсов и нулей (а) фильтра первого порядка и соответствующие частотные характеристики (б)

На рис. 2.4 показаны диаграмма полюсов и нулей и частотная характеристика для разностного уравнения первого порядка, соответствующего передаточной функции $H(z) = 1/(1-az^{-1})$ и импульсной характеристике $h(n) = a^n u(n)$. Из картины изменения векторов, соответствующих нулям и полюсам, ясно, что пики частотной характеристики получаются вблизи полюсов. Из этой геометрической картины должно быть понятно, что полюса и нули в начале координат не влияют на модуль частотной характеристики и вводят только линейную компоненту в фазу.

В качестве второго примера рассмотрим случай, когда импульсная характеристика системы является усеченной импульсной характеристикой предыдущего примера, т. е.

$$h(n) = \begin{cases} a^n, & 0 \leq n \leq M-1; \\ 0, & \text{при других } n. \end{cases}$$

$$H(z) = \sum_{n=0}^{M-1} a^n z^{-n} = (1 - a^M z^{-M}) / (1 - a z^{-1})$$

или $H(z) = (z^M - a^M) / z^{M-1} (z - a)$. Так как числитель имеет нули при $z_k = a e^{j(2\pi/M)k}$, $k=0,1,\dots,M-1$, где a считается положительным числом, то полюс при $z=a$

компенсируется нулем в той же точке. Диаграмма полюсов и нулей и соответствующая частотная характеристика для случая $M=8$ показана на рис. 2.5. Заметим наличие пика при $\omega=0$ ($z=1$), где нет нулей, и провалов в частотной характеристике в окрестности каждого нуля. Эти свойства частотной характеристики легко выводятся геометрически из диаграммы полюсов и нулей.

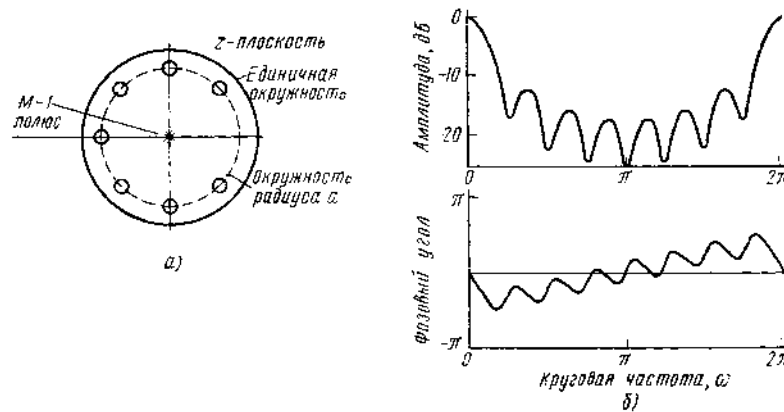


Рис.2.5. Диаграмма полюсов и нулей (а) и частотные характеристики КИХ-системы (б), импульсная характеристика которой является усеченной импульсной характеристикой для примера рис. 2.4.

3. Дискретное преобразование Фурье

Мы рассмотрели два варианта представлений дискретных сигналов и линейных систем с постоянными параметрами: на основе дискретного во времени преобразования Фурье (ДВПФ) в частотной области и z-преобразования в комплексной z-плоскости. Для частного, но практически важного случая, когда представляемая последовательность имеет конечную длительность, т.е. имеет конечное число ненулевых значений, возможно еще одно представление в частотной области, называемое *дискретным преобразованием Фурье (ДПФ)*, которое также является дискретной последовательностью и соответствует равноудаленным по частоте отсчетам ДВПФ. Кроме своей теоретической важности, ДПФ играет центральную роль при реализации ряда алгоритмов обработки сигналов, поскольку разработаны эффективные алгоритмы вычисления ДПФ, получившие название алгоритмов быстрого преобразования Фурье (БПФ).

Имеются различные подходы к выводу и интерпретации представления последовательности конечной длительности с помощью ДПФ. Мы используем представление, основанное на соотношении между последовательностями конечной длительности и периодическими последовательностями, и поэтому рассмотрим сначала ряды Фурье периодических последовательностей. Представление по Фурье периодических последовательностей важно само по себе, и мы используем его также для представления последовательностей конечной длины. Это делается путем создания периодической последовательности, каждый период которой совпадает с последовательностью конечной длины. Как будет показано далее, ряд Фурье этой периодической последовательности соответствует ДПФ последовательности конечной длины.

3.1. Представление периодических последовательностей дискретным рядом Фурье

Рассмотрим периодическую последовательность $x_p(n)$ с периодом в N отсчетов, т.е. $x_p(n) = x_p(n + kN)$ для любого целого значения k . Такие последовательности не могут быть представлены z-преобразованием, т.к. не существует ни одного значения z для которого бы сходилось ее z-преобразование. Однако можно представить $x_p(n)$ рядом Фурье:

$$x_p(n) = \frac{1}{N} \sum_{k=0}^{N-1} X_p(k) e^{j\frac{2\pi}{N}kn}, \quad (3.1)$$

т.е. суммой комплексных экспонент с частотами, кратными основной частоте $2\pi/N$

периодической последовательности: $e_k(n) = e^{j\frac{2\pi}{N}kn}, k = 0, 1, 2, \dots, N-1$. В

противоположность рядам Фурье непрерывных периодических функций имеется только N различных комплексных экспонент $e_k(n)$ с периодом, равным целой части основного периода N . Это является следствием того, что комплексная экспонента $e_k(n)$ периодична по k с периодом N . Так, $e_0(n) = e_N(n)$, $e_1(n) = e_{N+1}(n)$ и т.д. и, следовательно,

множество N комплексных экспонент $e_k(n) = e^{j\frac{2\pi}{N}kn}, k = 0, 1, 2, \dots, N-1$ определяет все различные комплексные экспоненты с частотами, кратными основной частоте $2\pi/N$. Поэтому представление $x_p(n)$ рядом Фурье содержит только N комплексных экспонент. Множитель $1/N$ введен для удобства и, конечно, не влияет на характер представления.

Чтобы выразить коэффициенты $X_p(k)$ через $x_p(n)$, умножим обе части равенства (3.1) на $e^{-j\frac{2\pi}{N}rn}$ и просуммируем результаты по n :

$$\sum_{n=0}^{N-1} x_p(n) e^{-j\frac{2\pi}{N}rn} = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{k=0}^{N-1} X_p(k) e^{j\frac{2\pi}{N}(k-r)n} = \sum_{k=0}^{N-1} X_p(k) \frac{1}{N} \sum_{n=0}^{N-1} e^{j\frac{2\pi}{N}(k-r)n} = X_p(r), \quad (3.2)$$

$$\frac{1}{N} \sum_{n=0}^{N-1} e^{j \frac{2\pi}{N} rn} = \frac{1}{N} \left(\frac{1 - e^{j \frac{2\pi}{N} rN}}{1 - e^{j \frac{2\pi}{N} r}} \right) = \begin{cases} 1, r = mN, m - \text{целое число;} \\ 0, \text{ в остальных случаях.} \end{cases} \quad (3.3)$$

Таким образом, коэффициенты $X_p(k)$ в (3.1) получаются из соотношения:

$$X_p(k) = \sum_{n=0}^{N-1} x_p(n) e^{-j \frac{2\pi}{N} kn}. \quad (3.4)$$

Соотношения (3.1) и (3.4) образуют пару преобразований, называемых *дискретным рядом Фурье (ДРФ)* периодической последовательности $x_p(n)$.

Отметим, что последовательность коэффициентов ДРФ $X_p(k)$, является периодической последовательностью с периодом N : $X_p(k) = X_p(k + lN)$ (l -целое число).

3.2. Представление по Фурье последовательности конечной длительности – дискретное преобразование Фурье

Рассмотрим последовательность $x(n)$ конечной длины N , для которой $x(n)=0$ всюду, за исключением интервала $0 \leq n \leq (N-1)$:

$$x(n) = \begin{cases} x_p(n), & 0 \leq n \leq N-1; \\ 0, & \text{при других } n, \end{cases} \quad (3.5)$$

причем последовательность $x_p(n)$ имеет период в N отсчетов, т. е. $x(n)$ представляет собой один период периодической последовательности $x_p(n)$. z -преобразование $x(n)$ равно

$$X(z) = \sum_{n=0}^{N-1} x(n) z^{-n}, \quad (3.6)$$

Вычисляя сумму (3.6) при $z_k = e^{j \frac{2\pi}{N} k}$, т. е. в точке на единичной окружности с полярным углом $2\pi k/N$, находим

$$X(e^{j \frac{2\pi}{N} k}) = \sum_{n=0}^{N-1} x(n) e^{-j \frac{2\pi}{N} kn}. \quad (3.7)$$

Сравнивая суммы (3.7) и (3.4) и учитывая, что $x_p(n) = x(n)$ на интервале $0 \leq n \leq (N-1)$, получаем

$$X_p(k) = X(e^{j \frac{2\pi}{N} k}). \quad (3.8)$$

Итак, коэффициенты ДРФ последовательности $x_p(n)$ равны значениям z -преобразования последовательности конечной длины $x(n)$ в N точках, равномерно распределенных по единичной окружности. Другими словами коэффициенты ДРФ образуют последовательность отсчетов ДВПФ $X(e^{j\omega})$ последовательности $x(n)$ в точках

$\omega_k = \frac{2\pi}{N} k$. Обозначая $X(k) = X(e^{j\omega_k})$, перепишем (3.7) в виде

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j \frac{2\pi}{N} kn}, k = \overline{0, N-1}. \quad (3.9)$$

Еще более важный вывод состоит в том, что коэффициенты ДРФ однозначно представляют последовательность конечной длины $x(n)$, так как по ним можно точно восстановить исходную последовательность. Используя (3.1), и, учитывая, что $x_p(n)=x(n)$ для $0 \leq n \leq (N-1)$, получим:

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j \frac{2\pi}{N} kn}, n = \overline{0, N-1}. \quad (3.10)$$

Соотношения (3.9) и (3.10) называют соответственно *прямым и обратным дискретным преобразованием Фурье (ДПФ)* последовательности $x(n)$ конечной длины. Таким образом, ДПФ и ДРФ – это по сути дела одни и те же преобразования, используемые

для представления как последовательности конечной длины N , так и ее периодического продолжения с периодом N на всю числовую ось.

Поскольку ДПФ однозначно представляет последовательность конечной длины, появляется возможность найти ее z -преобразование через коэффициенты ДПФ этой последовательности. Из соотношений (3.6) и (3.10) получаем

$$\begin{aligned} X(z) &= \sum_{n=0}^{N-1} x(n)z^{-n} = \sum_{n=0}^{N-1} \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j\frac{2\pi}{N}kn} z^{-n} = \sum_{k=0}^{N-1} \frac{X(k)}{N} \sum_{n=0}^{N-1} [e^{j\frac{2\pi}{N}k} z^{-1}]^n = \\ &= \sum_{k=0}^{N-1} \frac{X(k)}{N} \frac{1 - z^{-N}}{1 - z^{-1} e^{j\frac{2\pi}{N}k}}. \end{aligned} \quad (3.11)$$

Равенство (3.11) показывает, что z -преобразование последовательности непосредственно связано с коэффициентами ее ДПФ. Для точек на единичной окружности равенство (3.11) принимает вид

$$\begin{aligned} X(e^{j\omega}) &= \sum_{k=0}^{N-1} \frac{X(k)}{N} \frac{1 - e^{-j\omega N}}{1 - e^{-j\omega} e^{j\frac{2\pi}{N}k}} = \sum_{k=0}^{N-1} \frac{X(k)}{N} \frac{e^{-j\omega N/2} (e^{j\omega N/2} - e^{-j\omega N/2})}{e^{-j(\frac{\omega}{2} - \frac{\pi}{N}k)} (e^{j(\frac{\omega}{2} - \frac{\pi}{N}k)} - e^{-j(\frac{\omega}{2} - \frac{\pi}{N}k)})} = \\ &= \sum_{k=0}^{N-1} \frac{X(k)}{N} \frac{e^{-j\omega(N-1)/2} \sin(\omega N/2)}{e^{j\frac{\pi}{N}k} \sin\left(\frac{\omega}{2} - \frac{\pi k}{N}\right)}. \end{aligned} \quad (3.12)$$

Следовательно, с помощью формулы (3.12) по коэффициентам ДПФ последовательности конечной длины можно найти ее непрерывный частотный спектр.

Представление конечных последовательностей с помощью ДПФ удобно также для получения значений преобразования Фурье в L -точках ($L > N$), равномерно распределенных по единичной окружности.

Рассмотрим конечную последовательность $x(n)$, $0 \leq n \leq N-1$ с преобразованием Фурье

$$X(e^{j\omega}) = \sum_{n=0}^{N-1} x(n) e^{-j\omega n}.$$

Вычисляя $X(e^{j\omega})$ на частотах $\omega = (2\pi/L)k$, $k=0, 1, \dots, L-1$, получим

$$X(e^{j\frac{2\pi}{L}k}) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{L}kn}, k = \overline{0, L-1}. \quad (3.13)$$

Введем новую последовательность $\hat{x}(n)$ длины L точек ($L > N$):

$$\hat{x}(n) = \begin{cases} x(n), & 0 \leq n \leq N-1; \\ 0, & N \leq n \leq L-1, \end{cases}$$

и найдем ее L -точечное ДПФ:

$$\hat{X}(k) = \sum_{n=0}^{L-1} \hat{x}(n) e^{-j\frac{2\pi}{L}kn} = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{L}kn}, k = \overline{0, L-1}. \quad (3.14)$$

Сравнивая (3.13) и (3.14), получим

$$\hat{X}(k) = X(e^{j\frac{2\pi}{L}k}), k = \overline{0, L-1}. \quad (3.15)$$

Таким образом, простое дополнение последовательности $x(n)$ конечной длины N нулевыми отсчетами до длины $L > N$ с последующим вычислением L -точечного ДПФ позволяет получать отсчеты ДВПФ для совокупности точек, равномерно распределенных по единичной окружности, с частотным интервалом $(2\pi/L)$, т.е. при $L \gg 1$ сколь угодно малым. При спектральном анализе конечных последовательностей эта несложная операция является одной из наиболее полезных.

3.3. Свойства дискретного преобразования Фурье

Некоторые свойства ДПФ играют в практических вопросах обработки сигналов важную роль. Ниже они будут в основном перечислены, детали будут рассмотрены только в случае необходимости.

Свойства ДПФ и ДРФ по существу совпадают, поэтому они будут рассматриваться одновременно применительно как к последовательностям конечной длины, так их периодическим продолжениям на всю числовую ось и именоваться одинаково – ДПФ.

1. Линейность

Если $x_p(n)$ и $y_p(n)$ — периодические последовательности (с периодом в N отсчетов каждая), а $X_p(k)$ и $Y_p(k)$ — их ДПФ, то дискретное преобразование Фурье последовательности $x_p(n) + y_p(n)$ равно $X_p(k) + Y_p(k)$. Это положение справедливо и для последовательностей конечной длины.

2. Сдвиг

Если последовательность $x_p(n)$ периодическая с периодом в N отсчетов, а ее ДПФ равно $X_p(k)$, то ДПФ периодической последовательности вида $x_p(n-n_0)$ будет равно

$$X_p(k)e^{-j\frac{2\pi}{N}kn_0}.$$

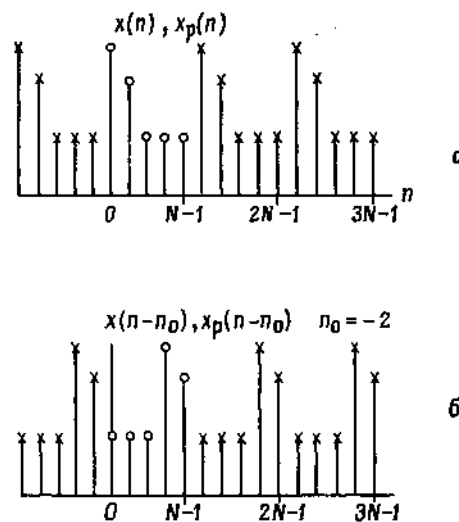


Рис.3.1. К определению ДПФ сдвинутой последовательности.

При анализе последовательностей конечной длины необходимо учитывать специфический характер временного сдвига последовательности. Так, на рис.3.1а изображена конечная последовательность $x(n)$ длиной в N отсчетов. Там же крестиками изображены отсчеты эквивалентной периодической последовательности $x_p(n)$, имеющей то же ДПФ, что и $x(n)$. Чтобы найти ДПФ сдвинутой последовательности $x(n-n_0)$, причем $n_0 < N$, следует рассмотреть сдвинутую периодическую последовательность $x_p(n-n_0)$ и в качестве эквивалентной сдвинутой конечной

последовательности (имеющей ДПФ $X_p(k)e^{-j\frac{2\pi}{N}kn_0}$) принять отрезок последовательности $x_p(n-n_0)$ в интервале $0 \leq n \leq N-1$. Таким образом, с точки зрения ДПФ последовательность $x(n-n_0)$ получается путем *кругового сдвига* элементов последовательности $x(n)$ на n_0 отсчетов, при котором, если одна из выборок в интервале $0 \leq n \leq N-1$ выходит из интервала, такая же выборка входит в интервал с другого конца.

3. Свойства симметрии

Если периодическая последовательность $x_p(n)$ с периодом в N отсчетов является действительной, то ее ДПФ $X_p(k)$ удовлетворяет следующим условиям симметрии:

$$\begin{aligned}
\operatorname{Re}[X_p(k)] &= \operatorname{Re}[X_p(N-k)], \\
\operatorname{Im}[X_p(k)] &= -\operatorname{Im}[X_p(N-k)], \\
|X_p(k)| &= |X_p(N-k)|, \\
\arg X_p(k) &= -\arg X_p(N-k).
\end{aligned}
\tag{3.16}$$

Аналогичные равенства справедливы и для конечной действительной последовательности $x(n)$, имеющей N -точечное ДПФ $X(k)$. Если ввести дополнительное условие симметрии последовательности $x_p(n)$, т. е. считать, что $x_p(n) = x_p(N-n)$, то окажется, что $X_p(k)$ может быть только действительной.

Поскольку чаще всего приходится иметь дело с действительными последовательностями, то, вычислив одно ДПФ, можно получить ДПФ двух последовательностей, используя свойства симметрии (3.16). Рассмотрим действительные периодические последовательности $x_p(n)$ и $y_p(n)$ с периодами в N отсчетов и N -точечными ДПФ $X_p(k)$ и $Y_p(k)$ соответственно. Введем комплексную последовательность $z_p(n)$ вида $z_p(n) = x_p(n) + j y_p(n)$. Ее ДПФ равно $Z_p(k) = X_p(k) + j Y_p(k)$. Выделяя действительную и мнимую части $Z_p(k)$, получим

$$\begin{aligned}
\operatorname{Re}[Z_p(k)] &= \operatorname{Re}[X_p(k)] - \operatorname{Im}[Y_p(k)], \\
\operatorname{Im}[Z_p(k)] &= \operatorname{Im}[X_p(k)] + \operatorname{Re}[Y_p(k)].
\end{aligned}$$

Действительные части $X_p(k)$ и $Y_p(k)$ симметричны, а мнимые — антисимметричны, поэтому их легко разделить, используя операции сложения и вычитания:

$$\begin{aligned}
\operatorname{Re}[X_p(k)] &= \frac{\operatorname{Re}[Z_p(k)] + \operatorname{Re}[Z_p(N-k)]}{2}, \\
\operatorname{Im}[Y_p(k)] &= \frac{\operatorname{Re}[Z_p(N-k)] - \operatorname{Re}[Z_p(k)]}{2}, \\
\operatorname{Re}[Y_p(k)] &= \frac{\operatorname{Im}[Z_p(k)] + \operatorname{Im}[Z_p(N-k)]}{2}, \\
\operatorname{Im}[X_p(k)] &= \frac{\operatorname{Im}[Z_p(k)] - \operatorname{Im}[Z_p(N-k)]}{2}.
\end{aligned}
\tag{3.17}$$

Итак, вычисляя одно N -точечное ДПФ, удастся преобразовать сразу две действительные последовательности длиной по N отсчетов. Если эти последовательности являются еще и симметричными, то число операций, необходимых для получения их ДПФ, можно сократить еще больше.

4. Свертка последовательностей

Если $x_p(n)$ и $h_p(n)$ — две периодические последовательности с периодами по N отсчетов и ДПФ, равными $X_p(k)$ и $H_p(k)$, то N -точечное ДПФ последовательности $y_p(n)$, являющейся круговой (или периодической) сверткой последовательностей $x_p(n)$ и $h_p(n)$, т. е.

$$y_p(n) = \sum_{l=0}^{N-1} x_p(l) h_p(n-l), \tag{3.18}$$

равно

$$Y_p(k) = X_p(k) H_p(k). \tag{3.19}$$

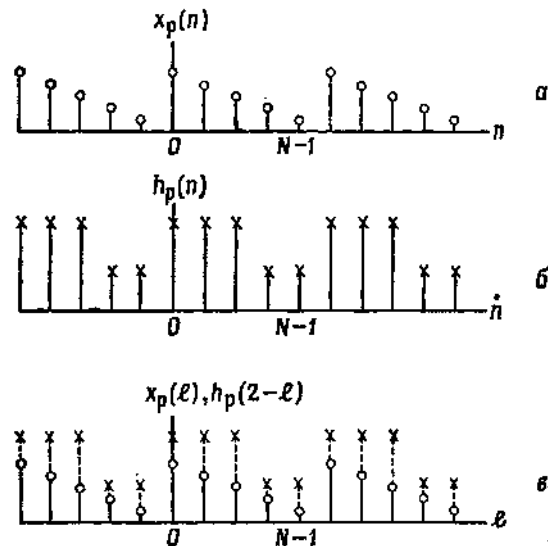


Рис.3.2. Круговая (периодическая) свертка.

Поскольку из формулы (3.19) получаются важные следствия, ниже показано, как она выводится. Сначала необходимо разъяснить понятие круговой свертки. На рис.3.2 а, б изображены периодические последовательности $x_p(n)$ и $h_p(n)$, а на рис.3.2 в показано, как вычисляется значение круговой свертки (3.18) при $n = 2$. В силу периодичности последовательностей $x_p(l)$ и $h_p(n-l)$ достаточно рассматривать их на интервале $0 \leq n \leq N-1$. С изменением n последовательность $h_p(n-l)$ смещается относительно $x_p(l)$. Когда отсчет $h_p(n-l)$ выходит за точку $l = N-1$, точно такой же отсчет появляется в точке $l=0$. Поэтому круговая свертка определяет свертку двух последовательностей, заданных на окружности.

Формулу (3.19) можно получить, найдя N -точечное ДПФ правой части (3.18), т. е.

$$Y_p(k) = \sum_{n=0}^{N-1} \left[\sum_{l=0}^{N-1} x_p(l) h_p(n-l) \right] e^{-j(2\pi/N)nk} = \sum_{l=0}^{N-1} x_p(l) \left[\sum_{n=0}^{N-1} h_p(n-l) e^{-j(2\pi/N)(n-l)k} \right] e^{-j(2\pi/N)lk} =$$

$$= H_p(k) \sum_{l=0}^{N-1} x_p(l) e^{-j(2\pi/N)lk} = H_p(k) X_p(k).$$

Полученная формула справедлива и для конечных последовательностей, если рассматривать $x_p(k)$ и $h_p(k)$ как эквивалентные им периодические последовательности с теми же ДПФ. Однако для конечных последовательностей обычно нужна линейная (ее называют аperiodической), а не круговая свертка, поэтому в приведенные формулы следует внести уточнения.

3.4. Линейная свертка конечных последовательностей

Рассмотрим две конечные последовательности $x(n)$ и $h(n)$ длины по N_1 и N_2 отсчетов, т. е. $x(n)$ отлична от нуля при $0 \leq n \leq N_1-1$, а $h(n)$ - при $0 \leq n \leq N_2-1$. *Линейной или аperiodической сверткой* этих последовательностей называют последовательность $y(n)$, определяемую соотношением

$$y(n) = \sum_{m=0}^n h(m) x(n-m). \quad (3.20)$$

где $h(m)$ и $x(n-m)$ равны нулю вне соответствующих интервалов. На рис.3.3 приведены примеры последовательностей $x(n)$, $h(n)$ и $y(n)$. Ясно, что последовательность $y(n)$ является конечной и имеет длину $(N_1 + N_2 - 1)$ отсчетов.

Выше было показано, что, перемножая ДПФ двух конечных последовательностей и находя обратное ДПФ произведения, получаем такой же результат, как при круговой свертке эквивалентных периодических последовательностей, образованных из данных конечных последовательностей. Исходя

из этого (см. также пример на рис.3.2), можно довольно просто получить линейную свертку двух конечных последовательностей. Свертка периодических последовательностей периодична и имеет тот же период, что и сами последовательности. Поскольку период свертки $y(n)$ (рис.3.3) равен $(N_1 + N_2 - 1)$ отсчетам, то для получения такого периода при круговой свертке необходимо, чтобы $x(n)$ и $h(n)$ содержали по $(N_1 + N_2 - 1)$ отсчетов, что достигается дополнением каждой из двух последовательностей соответствующим числом нулевых отсчетов. После этого можно найти $(N_1 + N_2 - 1)$ -точечные ДПФ дополненных последовательностей,

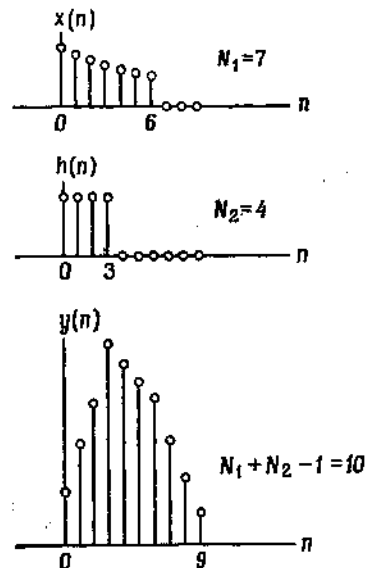


Рис.3.3. Линейная (апериодическая) свертка.

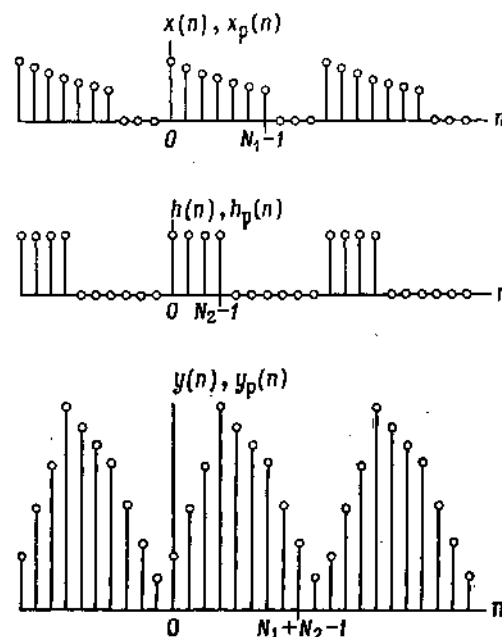


Рис.3.4. Вычисление линейной свертки с помощью круговой свертки.

перемножить их и выполнить обратное ДПФ произведения. В результате получается искомая свертка $y(n)$. На рис.3.4, иллюстрирующем эти операции, изображены эквивалентные периодические последовательности, используемые при вычислении круговой свертки. Ясно, что дополнение исходных последовательностей конечной длины $x(n)$ и $h(n)$ нулевыми отсчетами доводит период до нужной величины и позволяет устранить круговые наложения, характерные для круговой свертки. В результате каждый период последовательности $y_p(n)$ (рис.3.4) совпадает с $y(n)$ (рис.3.4). Рассмотренный метод вычисления свертки двух конечных последовательностей с применением алгоритма ДПФ называется *быстрой сверткой* в противоположность методу прямого вычисления суммы (3.20), называемому прямой или медленной

сверткой. Термин «быстрая» применяется потому, что ДПФ можно вычислить быстро и эффективно, используя один из алгоритмов быстрого преобразования Фурье (БПФ). Можно показать, что даже при умеренных величинах $(N_1 + N_2 - 1)$ (например, порядка 30) быстрая свертка оказывается эффективнее прямой. Поэтому рассмотренная методика является важным вычислительным средством при обработке сигналов.

Для практических приложений важно отметить, что в рассмотренном выше примере размер ДПФ не обязательно ограничивать величиной $(N_1 + N_2 - 1)$. ДПФ можно выполнять по любому числу отсчетов L , удовлетворяющему условию $L \geq N_1 + N_2 - 1$. Если это условие удовлетворяется, то в отличие от вышеописанной методики последовательности $x(n)$ и $h(n)$ дополняются другим числом нулевых отсчетов. В результате эквивалентная периодическая последовательность $y_p(n)$ будет иметь в конце периодов $(L - N_1 - N_2 + 1)$ нулей. Ясно, что эти отличия никак не искажают желаемого результата. Возможность произвольного выбора L существенна, поскольку практические алгоритмы вычисления ДПФ при разных L имеют неодинаковую эффективность. Так, например, для некоторых алгоритмов необходимо, чтобы L равнялось степени 2. В этом случае в качестве L приходится выбирать число, равное степени 2 и не меньшее чем $(N_1 + N_2 - 1)$.

3.5. Секционированные свертки

Во многих практических задачах необходимо вычислять свертку двух конечных последовательностей, когда одна из них гораздо длиннее другой (скажем, $N_1 \gg N_2$ или $N_2 \gg N_1$). Конечно, всегда можно выбрать L равным $(N_1 + N_2 - 1)$, но такой подход неэффективен и по ряду причин неудобен. Во-первых, перед вычислением свертки нужно иметь всю более длинную последовательность. На практике, например в радиолокации или при обработке речевых сигналов, это условие не всегда выполнимо. Во-вторых, поскольку обработка начинается только после приема всей последовательности, то результат получается с большой задержкой. И наконец, при слишком больших $(N_1 + N_2 - 1)$ вычисление ДПФ значительно усложняется, так как для этого требуется большой объем памяти и возникают некоторые другие, чисто практические трудности, связанные с алгоритмами БПФ. От перечисленных недостатков свободны следующие два метода вычисления свертки. Они основаны на разбиении более длинной последовательности на секции и вычислении частичных свертки, из которых затем формируется искомая выходная последовательность.

Первый из них называется методом перекрытия с суммированием. Сущность этого метода иллюстрируется на рис.3.5. Для простоты положим, что последовательность $x(n)$ не ограничена, а $h(n)$ содержит N_2 отсчетов. Разделим последовательность $x(n)$ на смежные секции длиной по N_3 отсчетов (рис.3.5). Выбор N_3 довольно сложен, но хорошие результаты получаются, если N_3 является величиной того же порядка, что и N_2 . Итак, входная последовательность $x(n)$ представляется в виде

$$x(n) = \sum_{k=0}^{\infty} x_k(n),$$

где

$$x_k(n) = \begin{cases} x(n) & \text{при } kN_3 \leq n \leq (k+1)N_3 - 1, \\ 0 & \text{при других } n. \end{cases}$$

Линейная свертка последовательностей $x(n)$ и $h(n)$ равна

$$y(n) = \sum_{m=0}^n h(m) \sum_{k=0}^{\infty} x_k(n-m) = \sum_{k=0}^{\infty} h(n) * x_k(n) = \sum_{k=0}^{\infty} y_k(n). \quad (3.20)$$

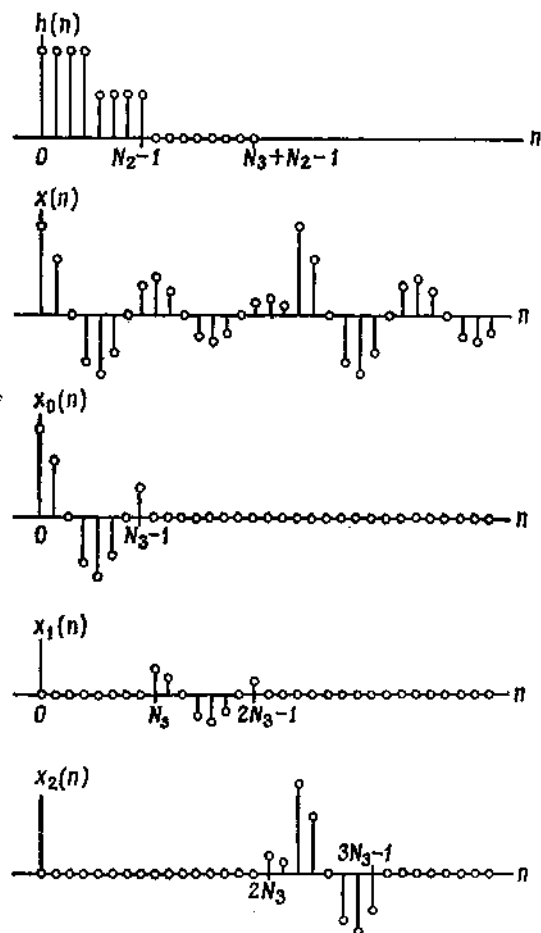


Рис.3.5. Метод перекрытия с суммированием.

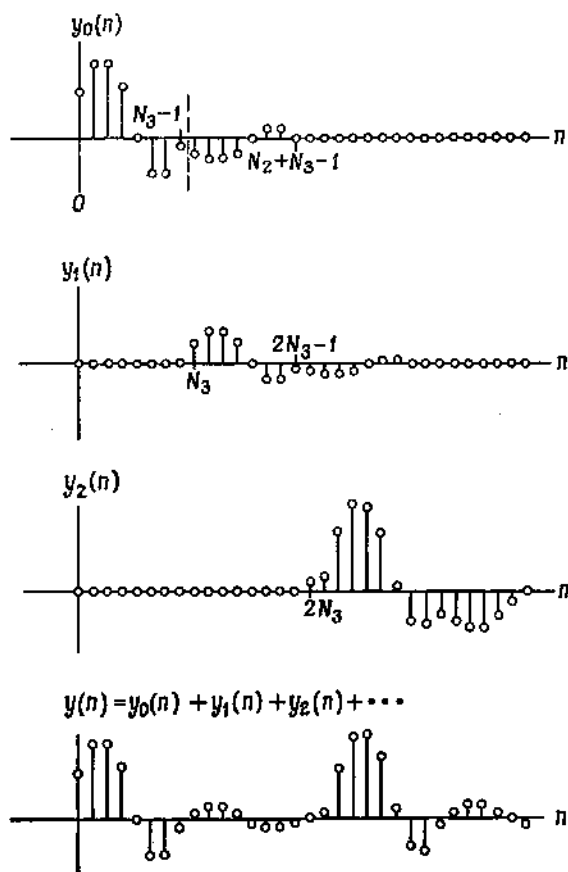


Рис.3.6. Формирование выходных значений свертки при использовании метода перекрытия с суммированием.

Длина каждой из частичных сверток в сумме (3.20) равна $(N_3 + N_2 - 1)$ отсчетам, т. е.

имеется участок длиной в $(N_2 - 1)$ отсчетов, на котором k -я и $(k + 1)$ -я частичные свертки перекрываются, поэтому их отсчеты на участке перекрытия нужно сложить. На рис.3.6 показано, как расположены и как суммируются соседние частичные свертки $y_k(n)$. Каждая из них вычисляется методом быстрой свертки, описанным выше. Рассмотренный метод был назван методом перекрытия с суммированием именно потому, что промежуточные частичные свертки перекрываются и для получения конечного результата их необходимо сложить.

Другой метод вычисления линейной свертки последовательностей, одна из которых значительно длиннее другой, также основан на секционировании более длинной последовательности. Его называют методом перекрытия с накоплением, причем в данном случае перекрываются входные, а не выходные секции. Ошибочные отсчеты круговых сверток отдельных секций отбрасываются. Остальные отсчеты накапливаются, и из них формируется конечный результат.

4. Анализ и проектирование цифровых фильтров

Выше было рассмотрено представление дискретных линейных систем, для которых соотношение между входной и выходной последовательностями определялось разностным уравнением, а передаточная функция – их z -преобразованиями. При этом основное внимание обращалось на характер взаимосвязи между входом и выходом. При построении цифрового фильтра программным путем на универсальной ЦВМ или в виде специализированного устройства эта взаимосвязь между входом и выходом должна быть преобразована в форму алгоритма вычисления. В зависимости от способа построения такой алгоритм будет определяться совокупностями либо основных математических операций, либо элементов схемы. Для построения дискретных систем, описываемых линейными разностными уравнениями с постоянными коэффициентами, в качестве основных операций целесообразно выбрать суммирование, задержку и умножение на константу. В этом случае расчетный алгоритм будет определен либо структурой программы, либо схемой устройства, устанавливающими взаимосвязь этих основных операций. Рассмотрим систему с передаточной функцией

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^M b(k)z^{-k}}{\sum_{k=0}^N a(k)z^{-k}},$$

для которой разностное уравнение, определяющее зависимость между входом и выходом, при $a(0)=1$ имеет вид

$$y(n) = -\sum_{k=1}^N a(k)y(n-k) + \sum_{k=0}^M b(k)x(n-k).$$

Это уравнение можно трактовать как расчетный алгоритм, в котором задержанные величины входной и выходной последовательностей умножаются на коэффициенты $b(k)$ и $a(k)$ соответственно и результаты умножения суммируются. В дальнейшем будет показано, что существует очень много структур, которые в результате будут давать одну и ту же зависимость между входными $x(n)$ и выходными $y(n)$ выборками. Следует однако учитывать, что две структуры могут быть эквивалентны (по отношению к характеристике взаимосвязи между входом и выходом) в случае, когда точность представления коэффициентов и переменных в фильтре неограничена, и они могут иметь существенно различные характеристики, когда указанная точность ограничена. С этой точки зрения одни структуры оказываются предпочтительнее других. Прежде чем обсуждать этот круг вопросов, рассмотрим несколько примеров фильтров первого и второго порядков с анализом их характеристик.

4.1. Примеры КИХ- и БИХ-фильтров первого и второго порядков и их применения

1. КИХ-фильтры первого и второго порядков.

Простейший КИХ-фильтр имеет передаточную функцию:

$$H(z) = b(0) + b(1)z^{-1},$$

частотную характеристику:

$$H(e^{j\omega}) = b(0) + b(1)e^{-j\omega}$$

и разностное уравнение, связывающее выход и вход:

$$y(n) = b(0)x(n) + b(1)x(n-1).$$

Импульсная характеристика такого фильтра имеет только два ненулевых отсчета: $h(n) = b(0)\delta(n) + b(1)\delta(n-1)$.

1). Выберем параметры $b(0)=1$, $b(1)=-1$, тогда передаточная функция будет:

$$H(z) = 1 - z^{-1},$$

с нулем в точке $z=1$, а частотная характеристика примет вид:

$$H(e^{j\omega}) = 1 - e^{-j\omega} = 2je^{-j\omega/2} \frac{e^{j\omega/2} - e^{-j\omega/2}}{2j} = 2je^{-j\omega/2} \sin(\omega/2) = |H(e^{j\omega})|e^{j\varphi(\omega)},$$

где $|H(e^{j\omega})| = 2|\sin(\omega/2)|$ - АЧХ фильтра,

$\varphi(\omega) = \frac{\pi}{2} - \frac{\omega}{2}$, ($0 \leq \omega \leq 2\pi$) - ФЧХ фильтра.

На рис.4.1 представлена структурная схема фильтра, соответствующая его разностному уравнению: $y(n) = x(n) - x(n-1]$, и содержащая элемент задержки на отсчет, обозначенный оператором z^{-1} , умножитель на константу (-1) , обозначенный стрелкой, и сумматор.

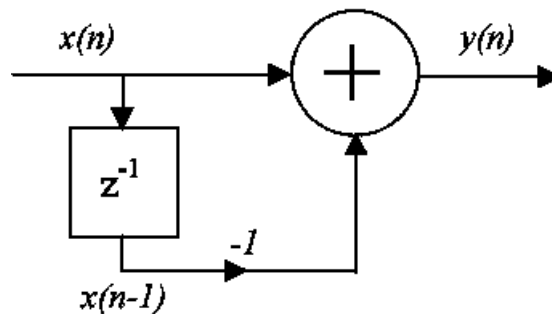


Рис.4.1. Структурная схема КИХ-фильтра.

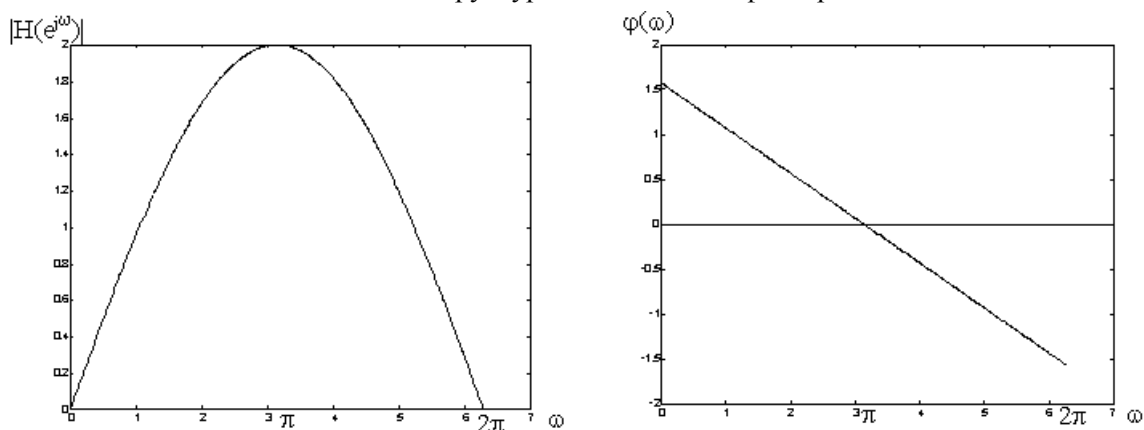


Рис.4.2. АЧХ и ФЧХ КИХ-фильтра.

На рис.4.2 представлены графически АЧХ и ФЧХ фильтра для диапазона частот ($0 \leq \omega \leq 2\pi$). Такой фильтр носит название череспериодного вычитателя (ЧПВ) и широко применяется в радиолокации в устройствах селекции движущихся целей (СДЦ) для борьбы с помехами, образованными отражениями от земли и местных предметов, спектр которых локализован в окрестности нулевой частоты. При этом спектр полезного сигнала, образованного отражением от движущейся цели, будет локализован в окрестности доплеровской частоты, и пропускается на выход фильтра, а спектр помехи существенно подавляется (режектируется).

Лучшее подавление помех от местных предметов обеспечивает каскадное включение двух таких фильтров 1-го порядка, которое эквивалентно КИХ-фильтру 2-го порядка с передаточной функцией:

$$H(z) = 0,5 - z^{-1} + 0,5z^{-2},$$

и АЧХ:

$$|H(e^{j\omega})| = 2\sin^2(\omega/2).$$

Структурная схема такого фильтра, соответствующая разностному уравнению:

$$y(n) = 0,5x(n) - x(n-1) + 0,5x(n-2),$$

представлена на рис.4.3а. На рис.4.3б отображен график АЧХ фильтра для диапазона частот ($0 \leq \omega \leq 2\pi$).

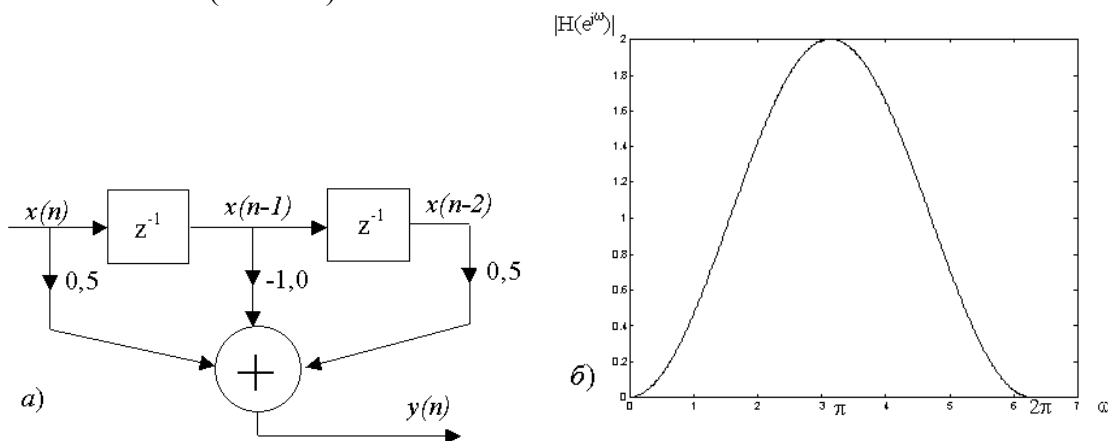


Рис.4.3. Структурная схема (а) и АЧХ (б) фильтра.

Заметим, что фильтр, описываемый уравнением $y(n) = x(n) - x(n-1)$, может считаться аналогом дифференцирующего фильтра для медленных процессов.

2). Выберем теперь параметры КИХ-фильтра $b(0)=1$, $b(1) = -e^{j\omega_0}$, тогда передаточная функция будет равна:

$$H(z) = 1 - e^{j\omega_0} z^{-1},$$

с нулем в точке $z_1 = e^{j\omega_0}$, а частотная характеристика примет вид:

$$H(e^{j\omega}) = 1 - e^{-j(\omega - \omega_0)} = 2je^{-j(\omega - \omega_0)/2} \sin[(\omega - \omega_0)/2] = |H(e^{j\omega})| e^{j\varphi(\omega)}$$

и, следовательно, АЧХ фильтра:

$$|H(e^{j\omega})| = 2|\sin[(\omega - \omega_0)/2]|,$$

а ФЧХ фильтра:

$$\varphi(\omega) = \frac{\pi}{2} - \frac{\omega - \omega_0}{2}.$$

Такой фильтр может подавлять помехи, спектр которых сосредоточен в окрестности частоты ω_0 , что, в частности, имеет место для радиолокационных отражений от местных предметов с учетом скорости ветра. На рис.4.4 представлена графически АЧХ фильтра ($\omega_0 = \pi/4$) для диапазона частот ($0 \leq \omega \leq 2\pi$).

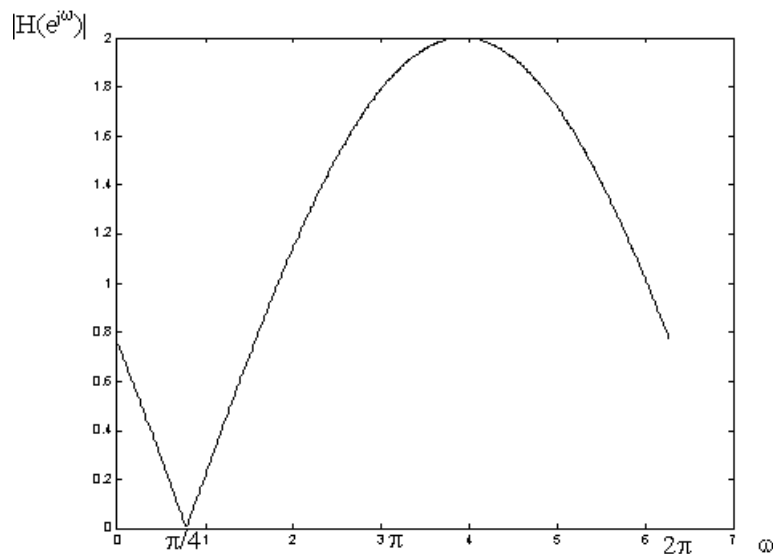


Рис. 4.4. АЧХ комплексного КИХ-фильтра ($\omega_0 = \pi/4$).

Рассмотрим структуру фильтра, описываемого уравнением $y(n) = x(n) + b(1)x(n-1)$ с комплексным коэффициентом $b(1)$. Подставляя в уравнение фильтра комплексные формы сигналов и коэффициента: $x(n) = x_c(n) + jx_s(n)$, $y(n) = y_c(n) + jy_s(n)$, $b(1) = b_c(1) + jb_s(1)$, получим:

$$y_c(n) + jy_s(n) = [x_c(n) + jx_s(n)] + [b_c(1) + jb_s(1)][x_c(n-1) + jx_s(n-1)] = \\ = x_c(n) + b_c(1)x_c(n-1) - b_s(1)x_s(n-1) + j[x_s(n) + b_c(1)x_s(n-1) + b_s(1)x_c(n-1)].$$

Приравнявая порознь действительные и мнимые части, получим эквивалентную систему действительных уравнений:

$$y_c(n) = x_c(n) + b_c(1)x_c(n-1) - b_s(1)x_s(n-1);$$

$$y_s(n) = x_s(n) + b_c(1)x_s(n-1) + b_s(1)x_c(n-1).$$

Структурная схема КИХ-фильтра с комплексным коэффициентом представлена на рис.4.5

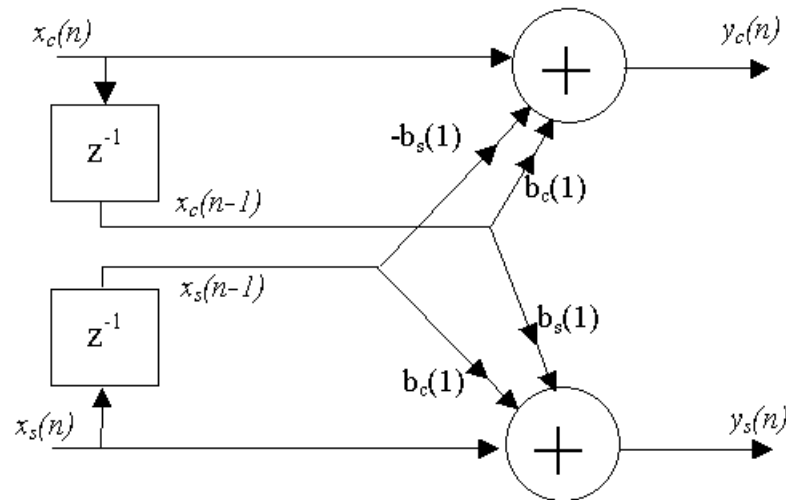


Рис.4.5. Структурная схема КИХ-фильтра с комплексным коэффициентом.

2. БИХ-фильтры первого и второго порядков.

Простейший БИХ-фильтр первого порядка имеет передаточную функцию:

$$H(z) = \frac{1}{1 + a(1)z^{-1}},$$

которая имеет один полюс при $p_1 = -a(1)$. Для устойчивости необходимо, чтобы $|a(1)| < 1$. Разностное уравнение фильтра:

$$y(n) = -a(1)y(n-1) + x(n),$$

а импульсная характеристика:

$$h(n) = [-a(1)]^n u(n).$$

Структурная схема БИХ-фильтра представлена на рис.4.6.

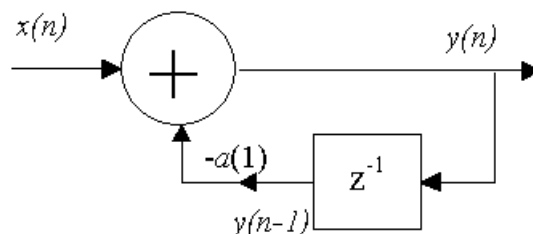


Рис.4.6. Структурная схема БИХ-фильтра первого порядка.

Рассмотрим частные случаи.

1). Пусть действительный коэффициент $a(1) = -r$, ($1 > r > 0$), тогда $H(z) = \frac{1}{1 - rz^{-1}}$ с одним полюсом при $p_1 = r$. Импульсная характеристика будет равна:

$$h(n) = r^n u(n).$$

Частотная характеристика примет вид:

$$H(e^{j\omega}) = \frac{1}{1 - re^{-j\omega}} = \frac{1}{1 - r \cos \omega + jr \sin \omega} = |H(e^{j\omega})| e^{j\varphi(\omega)},$$

и, следовательно, АЧХ фильтра:

$$|H(e^{j\omega})| = \frac{1}{\sqrt{1 + r^2 - 2r \cos \omega}},$$

а ФЧХ фильтра:

$$\varphi(\omega) = -\arctg\left(\frac{r \sin \omega}{1 - r \cos \omega}\right).$$

На рис.4.7 представлены графики АЧХ и ФЧХ БИХ-фильтра первого порядка с действительным коэффициентом ($r=0,7$) для диапазона частот ($0 \leq \omega \leq 2\pi$).

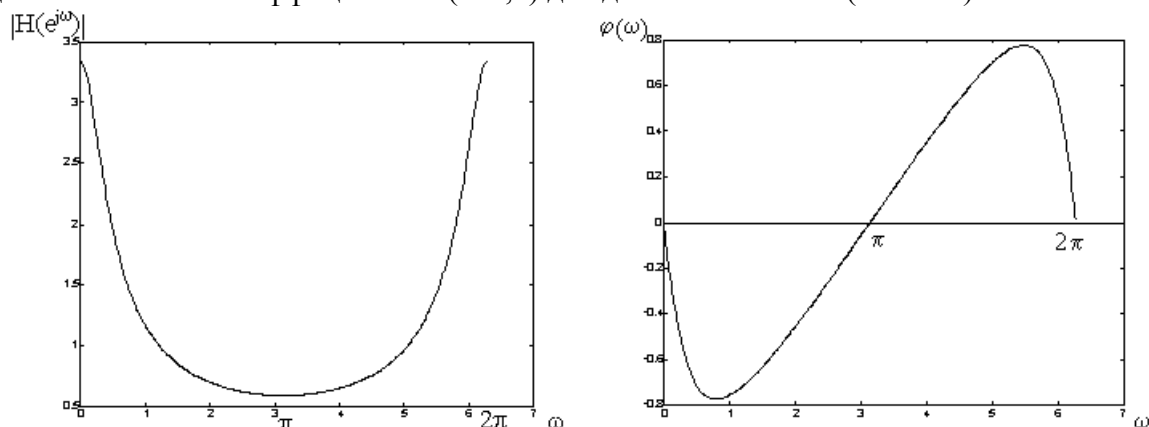


Рис.4.7. АЧХ и ФЧХ БИХ-фильтра первого порядка с действительным коэффициентом ($r=0,7$).

При ω близкой к нулю $\cos \omega \cong 1 - \omega^2 / 2$, поэтому АЧХ может быть описана соотношением:

$$|H(e^{j\omega})| \cong \frac{1}{\sqrt{(1 - 2r + r^2) + r\omega^2}} = \frac{1/(1-r)}{\sqrt{1 + \left(\frac{\sqrt{r}}{1-r}\right)^2 \omega^2}},$$

которое эквивалентно АЧХ аналогового RC-интегратора – фильтра нижних частот с постоянной времени $\tau = RC = \frac{T\sqrt{r}}{1-r}$. Такой фильтр можно использовать для выделения медленных сигналов, наблюдаемых на фоне широкополосных помех. В радиолокации такой фильтр, называемый *рециркулятором*, используется для последетекторного накопления отсчетов с экспоненциальным весом:

$$y(n) = \sum_{k=0}^{\infty} x(n-k)r^k = x(n) + rx(n-1) + r^2x(n-2) + \dots$$

2). Пусть комплексный коэффициент $a(1) = -re^{j\omega_0}$, тогда $H(z) = \frac{1}{1 - re^{j\omega_0} z^{-1}}$ с одним полюсом при $p_1 = re^{j\omega_0}$ будет соответствовать передаточной функции так называемого *комплексного резонатора*. Импульсная характеристика будет равна:

$$h(n) = r^n e^{j\omega_0 n} u(n)$$

и представляет собой затухающую комплексную экспоненту с частотой ω_0 (резонансная частота). Параметр r определяет полосу пропускания (затухание) резонатора.

Частотная характеристика примет вид:

$$H(e^{j\omega}) = \frac{1}{1 - re^{-j(\omega - \omega_0)}} = \frac{1}{1 - r \cos(\omega - \omega_0) + jr \sin(\omega - \omega_0)} = |H(e^{j\omega})| e^{j\varphi(\omega)}$$

и, следовательно, АЧХ фильтра:

$$|H(e^{j\omega})| = \frac{1}{\sqrt{1 + r^2 - 2r \cos(\omega - \omega_0)}},$$

а ФЧХ фильтра:

$$\varphi(\omega) = -\arctg\left(\frac{r \sin(\omega - \omega_0)}{1 - r \cos(\omega - \omega_0)}\right).$$

Таким образом, частотная характеристика БИХ-фильтра с комплексным коэффициентом по отношению к частотной характеристике рециркулятора будет смещена по оси частот на величину ω_0 . На рис.4.8 представлены графики АЧХ БИХ-фильтра первого порядка с комплексными коэффициентами $a(1) = -0,7e^{j\pi/4}$ и $a(1) = -0,9e^{j\pi/4}$ для диапазона частот $(0 \leq \omega \leq 2\pi)$.

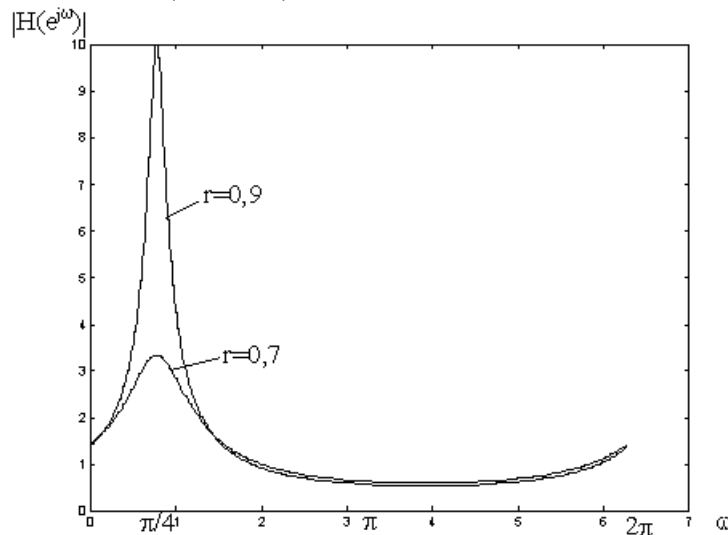


Рис.4.8. АЧХ БИХ-фильтра первого порядка с комплексным коэффициентом

$$a(1) = -0,7e^{j\pi/4} \text{ и } a(1) = -0,9e^{j\pi/4}$$

Комплексный резонатор реализует когерентное (т.е. с компенсацией фазового сдвига) накопление отсчетов комплексного входного сигнала $x(n) = s(n)e^{j\omega_0 n}$:

$$\begin{aligned} y(n) &= \sum_{k=0}^{\infty} x(n-k)h(k) = \sum_{k=0}^{\infty} s(n-k)e^{j\omega_0(n-k)} r^k e^{j\omega_0 k} = e^{j\omega_0 n} \sum_{k=0}^{\infty} e^{-j\omega_0 k} e^{j\omega_0 k} s(n-k) r^k = \\ &= e^{j\omega_0 n} \sum_{k=0}^{\infty} s(n-k) r^k. \end{aligned}$$

Такие фильтры применяются для построения цифровых спектроанализаторов, в РЛС - для реализации когерентного накопления отсчетов. Аналоговый прототип комплексного резонатора – колебательный контур.

Комплексный резонатор первого порядка эквивалентен двум действительным БИХ-фильтрам первого порядка, уравнения которых можно получить, подставляя в уравнение фильтра комплексные представления входного и выходного сигналов $x(n) = x_c(n) + jx_s(n)$, $y(n) = y_c(n) + jy_s(n)$ и коэффициента $a(1) = a_c(1) + ja_s(1)$:

$$y_c(n) + jy_s(n) = [x_c(n) + jx_s(n)] - [a_c(1) + ja_s(1)][y_c(n-1) + jy_s(n-1)] = \\ = x_c(n) - a_c(1)y_c(n-1) + a_s(1)y_s(n-1) + j[x_s(n) - a_c(1)y_s(n-1) - a_s(1)y_c(n-1)].$$

Приравнявая порознь действительные и мнимые части, получим эквивалентную систему действительных уравнений:

$$y_c(n) = x_c(n) - a_c(1)y_c(n-1) + a_s(1)y_s(n-1);$$

$$y_s(n) = x_s(n) - a_c(1)y_s(n-1) - a_s(1)y_c(n-1).$$

Структурная схема комплексного резонатора представлена на рис.4.9.

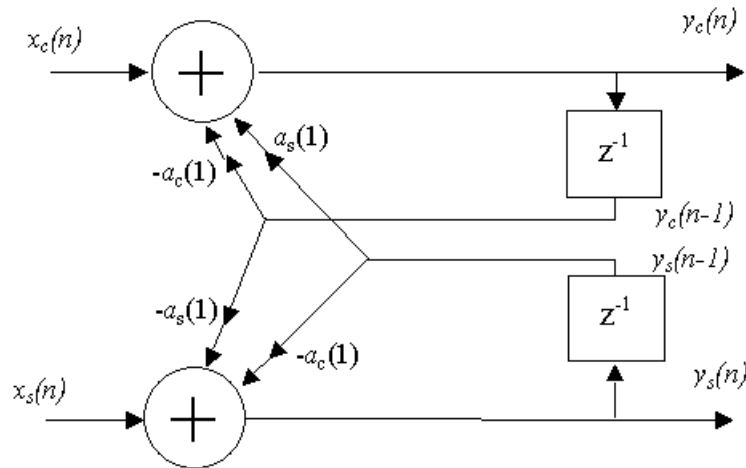


Рис.4.9. Структурная схема комплексного резонатора.

3). Рассмотрим действительный БИХ-фильтр второго порядка, передаточная функция которого имеет два комплексно сопряженных полюса: $p_1 = re^{j\omega_0}$, $p_1^* = re^{-j\omega_0}$:

$$H(z) = \frac{1}{2} \left[\frac{1}{1 - re^{j\omega_0} z^{-1}} + \frac{1}{1 - re^{-j\omega_0} z^{-1}} \right] = \frac{1 - (r \cos \omega_0) z^{-1}}{1 - 2(r \cos \omega_0) z^{-1} + r^2 z^{-2}} = \frac{1 + b(1)z^{-1}}{1 + a(1)z^{-1} + a(2)z^{-2}}$$

Передаточная функция имеет и один нуль $z_1 = r \cos \omega_0$, положение которого жестко связано с положением полюсов. Разностное уравнение фильтра имеет вид:

$$y(n) = -a(1)y(n-1) - a(2)y(n-2) + x(n) + b(1)x(n-1).$$

АЧХ такого действительного резонатора образуется совокупностью АЧХ двух комплексных резонаторов с резонансными частотами ω_0 и $(-\omega_0)$ для диапазона частот $(-\pi \leq \omega \leq \pi)$. На рис.4.10а представлен один из вариантов структурной схемы действительного резонатора, а на рис.4.10б - график его АЧХ при $r=0.95$ и $\omega_0=\pi/2$.

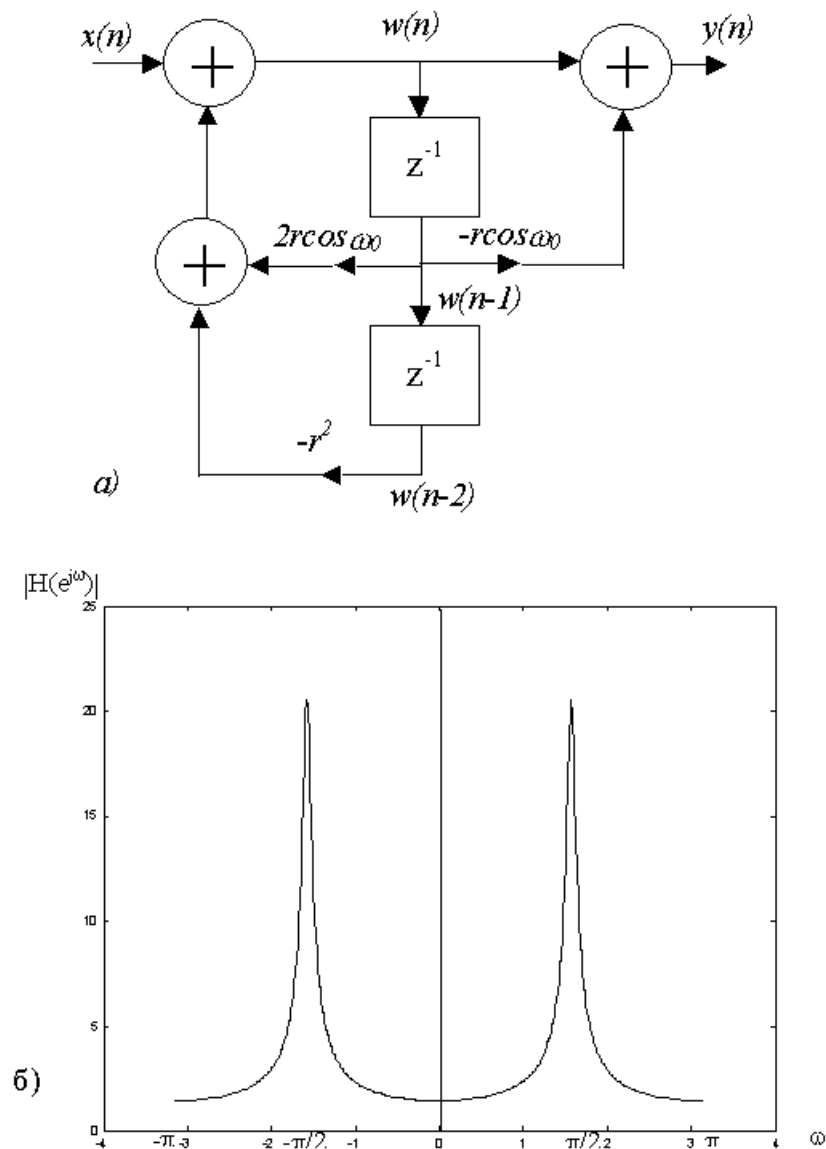


Рис.4.10. Структурная схема (а) и АЧХ действительного резонатора при $r=0.95$ и $\omega_0=\pi/2$ (б).

4.2. Основные структурные схемы при построении БИХ-фильтров

Для построения цифрового фильтра необходимо иметь предшествовавшие значения выходной, входной и промежуточной последовательностей. Это подразумевает необходимость задержки или запоминания этих величин. Наряду с этим следует обеспечить умножение задержанных выборок на коэффициенты и суммирование полученных результатов. Как отмечалось, цифровой фильтр может быть построен либо на базе регистров памяти, арифметического устройства и устройства управления универсальной ЦВМ, либо в виде специализированного цифрового устройства. В первом случае структура фильтра может рассматриваться как детализация расчетного алгоритма, на основе которого строится программа для ЦВМ. Во втором случае ее удобно представлять в виде конкретной схемы специализированного устройства.

В результате анализа представления линейных дискретных систем с постоянными параметрами, показано, что каждой рациональной передаточной функции соответствует большое количество форм построения цепи. Естественно, что одним из важных соображений при выборе форм построения систем является вычислительная

сложность. Это значит, что цепи с наименьшим количеством умножителей на константу и наименьшим числом ветвей задержки часто наиболее желательны, так как умножение является операцией продолжительной по времени и для каждого элемента задержки подразумевается использование регистра памяти. Следовательно, уменьшение количества умножителей означает увеличение быстродействия, а уменьшение числа элементов задержки — снижение требований к памяти. С другой стороны, необходима учитывать, что эффекты конечной длины регистров при построении цифровых фильтров в виде специализированных устройств зависят от структуры последних. Поэтому иногда целесообразнее использовать те структуры, которые, хотя и не имеют минимального количества умножителей и элементов задержки, однако оказываются менее чувствительными к эффектам конечной длины регистров. Таким образом, важно обсудить некоторые из наиболее распространенных форм построения цепей. Рассмотрим сначала БИХ-фильтры.

4.2.1. Прямая форма

Напомним, что если рассматривается рациональная передаточная функция вида

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^M b(k)z^{-k}}{1 + \sum_{k=1}^N a(k)z^{-k}} = \frac{B(z)}{A(z)}, \quad (4.1)$$

то соотношение между входом и выходом такой системы удовлетворяет разностному уравнению

$$y(n) = -\sum_{k=1}^N a(k)y(n-k) + \sum_{k=0}^M b(k)x(n-k). \quad (4.2)$$

Поскольку это разностное уравнение может быть записано непосредственно из выражения для передаточной функции, то форму построения цепи, соответствующей разностному уравнению, называют *прямой формой 1*. Простая структура реализации данного разностного уравнения показана на рис. 4.11.

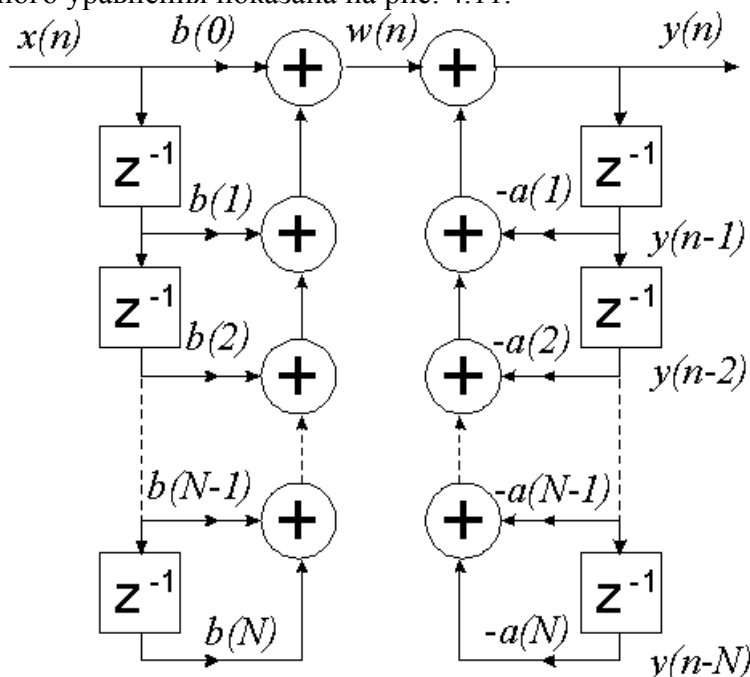


Рис.4.11. Прямая форма 1 реализации разностного уравнения N -го порядка

При построении цепи для простоты принято, что $M=N$. Если это не выполняется, то некоторые из ветвей имеют нулевую передачу. На рис. 4.11 показан такой способ построения, при котором каждый сумматор имеет не больше двух входов. Несмотря на

то, что эта условность приводит к большему числу сумматоров, чем необходимо, она согласуется с тем фактом, что при построении цифровых фильтров (как программным путем на ЦВМ, так и в виде специализированных устройств) операция суммирования нескольких чисел (больше двух) осуществляется на основе формирования отдельных сумм пар чисел. В цифровой аппаратуре в отдельный момент времени, как правило, суммируются только два числа.

Поскольку совокупности коэффициентов $b(k)$ и $a(k)$ соответствуют полиномам числителя $B(z)$ и знаменателя $A(z)$ передаточной функции $H(z)$, то структуру, показанную на рис. 4.11, можно трактовать как каскадное соединение двух цепей. Первая из них реализует нули и имеет передаточную функцию $B(z)$, а вторая — полюсы системы и имеет передаточную функцию $1/A(z)$. Обозначив выходной сигнал первой системы $w(n)$, разностное уравнение можно заменить системой уравнений:

$$y(n) = -\sum_{k=1}^N a(k)y(n-k) + w(n);$$

$$w(n) = \sum_{k=0}^M b(k)x(n-k),$$

которая и реализована структурой, показанной на рис. 4.11.

В линейных дискретных системах с постоянными параметрами общее соотношение между входом и выходом не зависит от порядка каскадного соединения блоков. Из этого свойства вытекает вторая прямая форма построения цепи. А именно, если сначала реализовать полюсы $H(z)$ соответственно правой части структурной схемы рис. 4.11, которая имеет передаточную функцию $1/A(z)$, а затем — нули передаточной функцией $B(z)$, то получим структуру, показанную на рис. 4.12, которая соответствует системе уравнений:

$$w(n) = -\sum_{k=1}^N a(k)w(n-k) + x(n);$$

$$y(n) = \sum_{k=0}^M b(k)w(n-k).$$

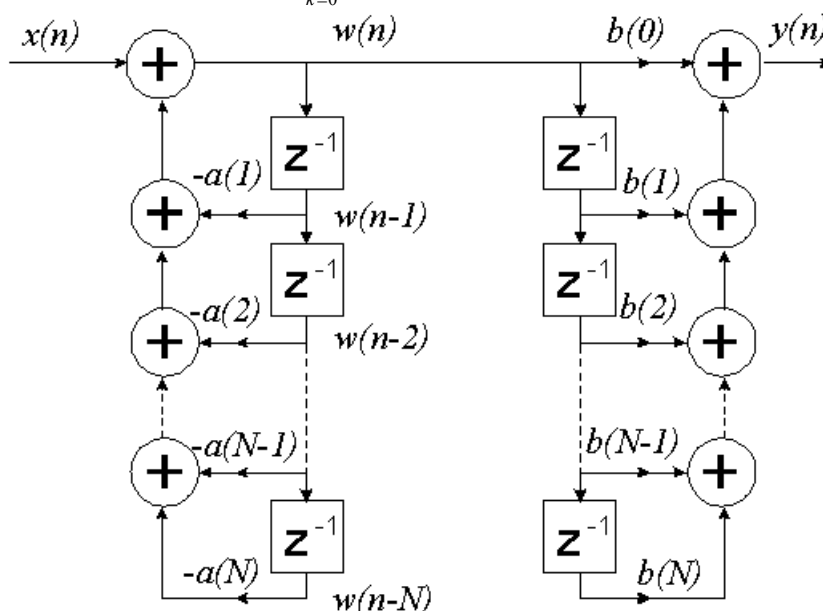


Рис.4.12. Прямая форма 2 (неканоническая)

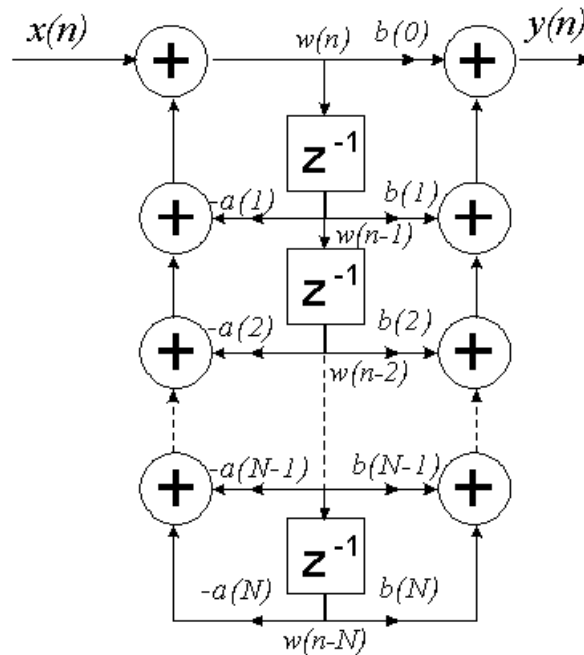


Рис.4.13. Прямая форма 2 (каноническая)

Поскольку в такой схеме имеется два ряда ветвей задержки для одного и того же входа, то можно обойтись одним из них. Таким образом, цепь рис. 4.12 может быть видоизменена так, как показано на рис.4.13. Она часто носит название *прямой формы 2*. Следует отметить, что такая цепь имеет минимальное количество элементов задержки (большее из M или N). Такие структуры часто называют *канонической формой цепи*.

4.2.2. Каскадная форма

Рассмотренная ранее прямая форма построения цепи следовала непосредственно из выражения для ее передаточной функции $H(z)$ вида (4.1). Если в выражении для передаточной функции числитель и знаменатель разложить на произведение линейных или квадратичных множителей, то $H(z)$ можно представить как

$$H(z) = A \frac{\prod_{m=1}^{M_1} (1 - z_m z^{-1}) \prod_{k=1}^{M_2} (1 - z_k z^{-1})(1 - z_k^* z^{-1})}{\prod_{m=1}^{N_1} (1 - p_m z^{-1}) \prod_{k=1}^{N_2} (1 - p_k z^{-1})(1 - p_k^* z^{-1})}, \quad (4.3)$$

где $M = M_1 + 2M_2$ и $N = N_1 + 2N_2$. В этом выражении множители первого порядка представляют действительные нули в точке z_m и действительные полюсы в точке p_m , а множители второго порядка представляют комплексно-сопряженные нули в точках z_k и z_k^* и комплексно-сопряженные полюсы в точках p_k и p_k^* . Этот случай представляет наиболее общее распределение полюсов и нулей, когда все коэффициенты $a(k)$ и $b(k)$ в (4.1) являются действительными. Соотношение (4.3) предполагает множество структур, образованных каскадным соединением блоков первого и второго порядков (рис.4.14):

$$H(z) = \prod_m H_{1m}(z) \prod_k H_{2k}(z), \quad (4.4)$$

где блоки первого порядка с передаточной функцией

$$H_{1m}(z) = \frac{1 - z_m z^{-1}}{1 - p_m z^{-1}} = \frac{1 + b_m(1)z^{-1}}{1 + a_m(1)z^{-1}},$$

а блоки второго порядка с передаточной функцией

$$H_{2k}(z) = \frac{(1 - z_k z^{-1})(1 - z_k^* z^{-1})}{(1 - p_k z^{-1})(1 - p_k^* z^{-1})} = \frac{1 + b_k(1)z^{-1} + b_k(2)z^{-2}}{1 + a_k(1)z^{-1} + a_k(2)z^{-2}}.$$



Рис.4.14. Каскадное соединение блоков первого и второго порядков.

Вполне очевидно, что существует значительная свобода в выборе как формы построения блоков, так и последовательности их расположения. На практике, однако, важно выполнить каскадное построение при минимальном объеме памяти, поэтому аппаратура часто строится с использованием разделения времени или мультиплексирования блоков только второго порядка. Этот подход удобно в общем случае рассмотреть на примере каскадной формы построения цепи, когда передаточная функция имеет вид

$$H(z) = A \prod_{k=1}^{\left\lfloor \frac{N+1}{2} \right\rfloor} \frac{1 + b_k(1)z^{-1} + b_k(2)z^{-2}}{1 + a_k(1)z^{-1} + a_k(2)z^{-2}}.$$

где $\left\lfloor \frac{(N+1)}{2} \right\rfloor$ — наибольшее целое число, содержащееся в $(N+1)/2$ (считается $M \leq N$). Используемая форма записи выражения для $H(z)$ предполагает попарное объединение действительных полюсов и нулей. При этом, если число действительных нулей нечетное, то один из коэффициентов $b_k(2)$ равен нулю. Аналогично, если число действительных полюсов нечетное, то один из коэффициентов $a_k(2)$ равен нулю. Проведенное обсуждение построения структурных схем в прямой форме показало, что можно создать каскадную структуру с минимальной памятью, если каждый блок второго порядка выполнить в прямой форме 2.

Структурная схема блока второго порядка в прямой форме 2 представлена на рис.4.15, а система уравнений, описывающих ее имеет вид:

$$w(n) = x(n) - a(1)w(n-1) - a(2)w(n-2);$$

$$y(n) = w(n) + b(1)w(n-1) + b(2)w(n-2).$$

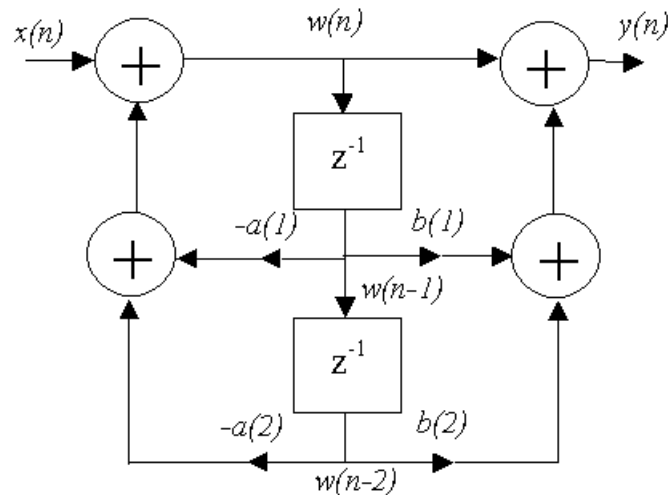


Рис.4.15. Структурная схема блока второго порядка в прямой форме 2.

Как было отмечено, существует значительная гибкость как при выборе способа попарного объединения полюсов и нулей, так и порядка последовательности, в которой следует расположить сформированные блоки второго порядка. В предположении неограниченной точности представления переменных и коэффициентов порядок расположения блоков и способ группирования нулей с полюсами не имеют значения (цепи будут эквивалентны). Однако на практике такие цепи могут значительно различаться из-за эффектов конечной разрядности представления чисел.

4.2.3. Параллельная форма

В отличие от способа разложения полиномов числителя и знаменателя

передаточной функции на множители, выражение для $H(z)$ при $M=N$ можно представить в виде разложения на простые дроби:

$$H(z) = C + \sum_{i=1}^K H_i(z), \quad (4.6)$$

где слагаемые $H_i(z)$ соответствуют или блокам второго порядка

$$H_i(z) = \frac{b_i(0) + b_i(1)z^{-1}}{1 + a_i(1)z^{-1} + a_i(2)z^{-2}}, \quad (4.7)$$

или блокам первого порядка

$$H_i(z) = \frac{b_i(0)}{1 + a_i(1)z^{-1}},$$

причем K равно целой части от $(N+1)/2$, и, как следует из (4.1), $C = b(N)/a(N)$. На рис.4.16 приведена структурная схема, реализующая соотношение (4.6). Ее называют параллельной формой. Блоки первого и второго порядка, описываемые формулами (4.7) и (4.8), строятся по схеме одной из рассмотренных выше прямых форм.

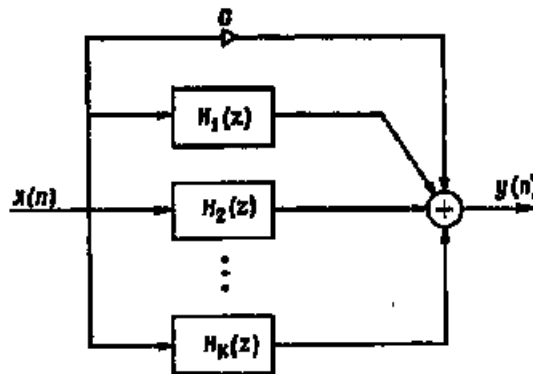


Рис.4.16. Параллельная форма.

Хотя рассмотренные структурные схемы фильтров не исчерпывают всех возможных структур, при моделировании на ЦВМ и аппаратной реализации фильтров они применяются наиболее часто. Другие структуры можно получить множеством способов. Так, например, можно построить параллельно-последовательную структуру, в которой часть передаточной функции реализуется в параллельной форме, а оставшая часть — в последовательной. Кроме того, для всех рассмотренных структур можно получить обращенные схемы, изменив направление прохождения всех сигналов (т. е. направив все стрелки на схемах в обратную сторону) и поменяв местами точки разветвления с точками суммирования сигналов. Полученные таким образом структуры будут иметь те же передаточные функции, но в них будут по-иному проявляться эффекты конечной разрядности чисел.

Выбор наилучшей из этих многочисленных структурных схем как при аппаратной реализации, так и при моделировании на ЦВМ определяется экономическими соображениями. Последние в свою очередь зависят от свойств структур при ограниченной точности представления переменных и коэффициентов фильтров.

4.3. Основные структурные схемы при построении КИХ-фильтров

Предыдущее обсуждение относилось к системам с бесконечными импульсными характеристиками, реализация которых строилась на использовании рекурсивного расчетного алгоритма. В случае физически реализуемых систем с импульсными характеристиками конечной длины для построения систем используют нерекурсивный расчетный алгоритм. Для таких систем передаточная функция имеет вид

$$H(z) = \sum_{n=0}^{N-1} h(n)z^{-n}. \quad (4.9)$$

Это означает, что если длина импульсной характеристики равна N отсчетам, то $H(z)$ является полиномом по z^{-1} степени $N-1$. Поэтому $H(z)$ имеет $N-1$ полюсов в точке $z=0$ и $N-1$ нулей, которые могут быть в любом месте ограниченной z -плоскости. Точно так же, как БИХ-фильтры, КИХ-фильтры могут иметь множество форм построения. Обсудим наиболее важные формы построения цепей КИХ-фильтров

4.3.1. Прямая форма

Прямая форма построения цепи следует непосредственно из соотношения для сверточной суммы, которое имеет вид

$$y(n) = \sum_{k=0}^{N-1} h(k)x(n-k). \quad (4.10)$$

Эта структура идентична показанной на рис. 4.13, когда все коэффициенты $a(k)$ равны нулю. Таким образом, прямая форма построения КИХ-фильтров является частным случаем прямой формы БИХ-фильтров.

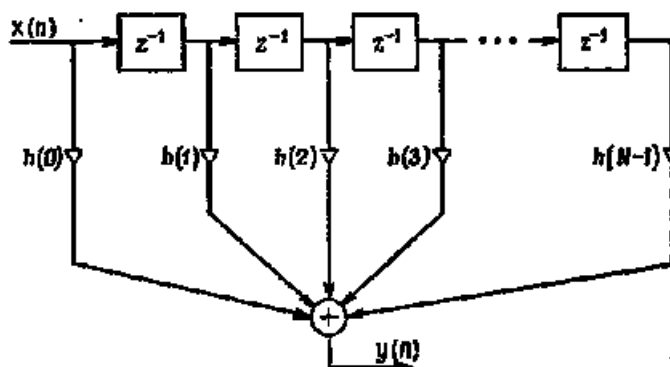


Рис. 4.17. Прямая форма построения КИХ-фильтра.

Структурная схема, показанная на рис. 4.17, является реализацией соотношения (4.10) и соответствует прямому порядку выполнения сложений и умножений. Из-за сходства этой структуры с линией задержки с отводами ее часто называют фильтром с многоотводной линией задержки (или иногда трансверсальным фильтром). Очевидно, что есть много и других способов организации вычисления и, таким образом, много других теоретически эквивалентных структур цепи.

4.3.2. Каскадная форма и структура на основе быстрой свертки

В отличие от прямой формы построения структур при каскадной форме предполагается, что передаточная функция представляется в виде произведения сомножителей второго порядка, т. е.

$$H(z) = \prod_{k=1}^{[N/2]} (b_k(0) + b_k(1)z^{-1} + b_k(2)z^{-2}), \quad (4.11)$$

где если N - четное, то один из коэффициентов $b_k(2)$ будет равен нулю в связи с тем, что для четных N приведенное уравнение передаточной функции $H(z)$ имеет нечетное число действительных корней.

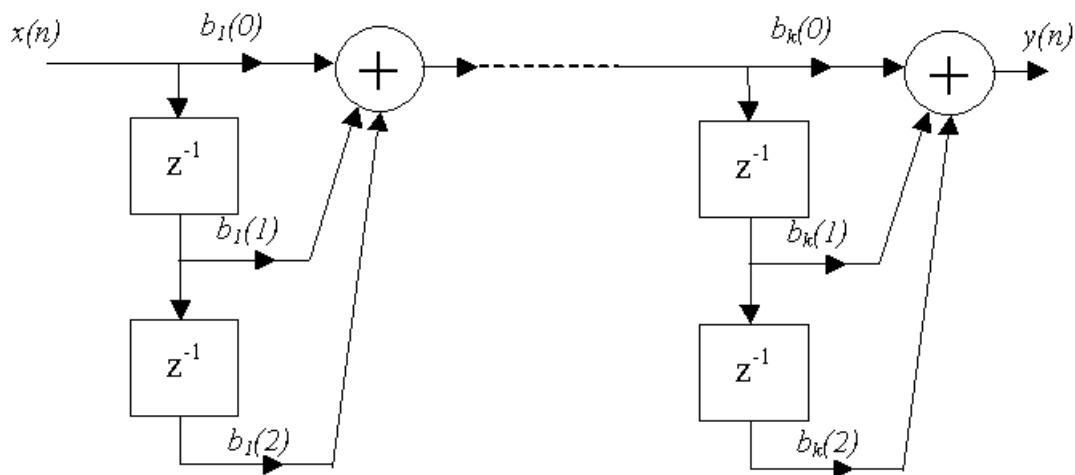


Рис.4.18. Каскадная форма построения КИХ-фильтра

На рис.4.18 показана структурная схема цепи, соответствующая (4.11). Показано, что каждый сомножитель второго порядка имеет прямую форму построения, приведенную на рис.4.17.

Для построения фильтров без полюсов довольно часто применяют еще несколько структур, которые не имеют аналогов с фильтрами общего вида, содержащими и нули, и полюсы. Наиболее распространенная из них основана на так называемом *методе быстрой свертки*, когда свертка вычисляется с помощью обратного преобразования Фурье от произведения преобразований Фурье входной последовательности и импульсной характеристики системы. Структура КИХ-фильтра на основе быстрой свертки представлена на рис.4.19.

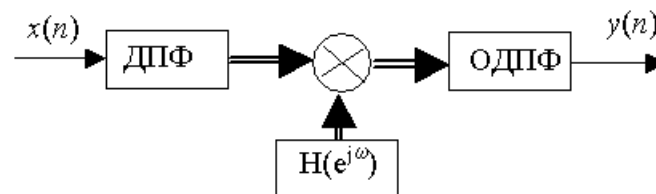


Рис.4.19. Структура КИХ-фильтра на основе быстрой свертки.

4.3.3. Структуры с частотной выборкой

В разделе 3 было показано, что z -преобразование последовательности конечной длины N может быть представлено с помощью N отсчетов, равномерно распределенных на единичной окружности. Для КИХ-фильтра выражение (3.11) подразумевает, что передаточная функция может быть представлена как

$$H(z) = \frac{1 - z^{-N}}{N} \sum_{k=0}^{N-1} \frac{H(k)}{1 - z^{-1} e^{j\frac{2\pi}{N}k}}. \quad (4.12)$$

Значения

$$H(k) = H(e^{j\frac{2\pi}{N}k}) = |H(k)| e^{j\varphi(k)}, \quad k=0,1,\dots,N-1 \quad (4.13)$$

называются *частотными выборками*, так как они являются просто выборками (отсчетами) частотной характеристики системы.

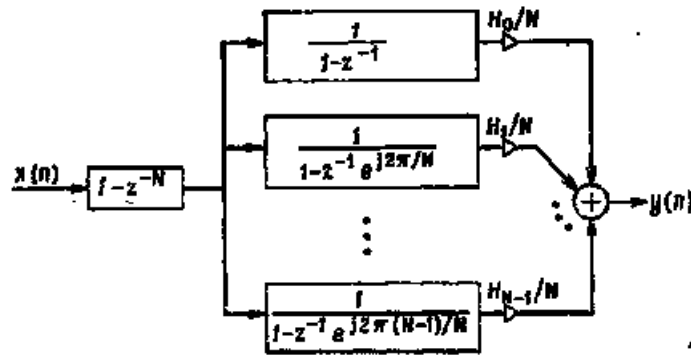


Рис.4.20. Структура КИХ-фильтра с частотной выборкой.

Выражение (4.12) означает, что КИХ-система может быть реализована как каскадное соединение простой КИХ-системы и БИХ-системы, как показано на рис.4.20. Передаточная функция КИХ-системы равна $1 - z^{-N}$ и нули этой системы находятся в точках $z_k = e^{j\frac{2\pi}{N}k}$, $k = 0, 1, \dots, N-1$. Часть структуры (рис.4.20), соответствующая БИХ-системе, состоит из параллельного соединения N комплексных систем первого порядка с полюсами в точках $z_k = e^{j\frac{2\pi}{N}k}$. Эти системы первого порядка имеют полюсы, лежащие строго на единичной окружности, назначение которых заключается в подавлении точно одного из нулей КИХ-системы. На практике трудности обеспечения устойчивости, обусловленные расположением нулей на единичной окружности, исключаются с помощью образования выборок передаточной функции $H(z)$ на окружности радиуса r , где $r = 1 - (2^{-12} \div 2^{-16})$ немного меньше единицы. В этом случае $H(z)$ представляется как

$$H(z) = \frac{1 - r^N z^{-N}}{N} \sum_{k=0}^{N-1} \frac{H(k)}{1 - rz^{-1}e^{j\frac{2\pi}{N}k}}. \quad (4.14)$$

где для точного представления $H(z)$ с помощью (4.14) требуется, чтобы

$$H(k) = H(re^{j\frac{2\pi}{N}k}). \quad (4.15)$$

Если значение r выбрано близким к единице, то появляется небольшая ошибка, когда вместо выражения (4.15) используется (4.13).

В общем случае частотные выборки $H(k)$ являются комплексными, поэтому реализация КИХ-систем, таких, как на рис. 4.20, требует комплексных арифметических операций. Однако, если отсчеты импульсной характеристики $h(n)$ являются действительными, то частотные выборки, выраженные в полярных координатах, удовлетворяют следующим условиям симметрии:

$$\begin{aligned} |H(k)| &= |H(N-k)|; \\ \varphi(k) &= -\varphi(N-k), \quad k = 0, 1, 2, \dots, N-1, \end{aligned}$$

где если $H(0) > 0$, то $\varphi(0) = 0$. Кроме того, поскольку $\left(e^{j\frac{2\pi}{N}k}\right)^* = e^{j\frac{2\pi}{N}(N-k)}$, то цепи

первого порядка на рис. 4.20 образуют комплексно-сопряженные пары, за исключением цепи с полюсом в точке $e^{j(2\pi/N)0} = 1$, и для N четного — цепи с полюсом в точке $e^{j(2\pi/N)N/2} = -1$. Следовательно, комплексные цепи первого порядка могут быть сгруппированы в комплексно-сопряженные пары и выполнены как цепи второго порядка с действительными коэффициентами, как это было показано в примере 3 раздела 4.1.

Существуют два принципиальных преимущества реализаций на основе частотной выборки. Первое заключается в том, что в умножителях на выходах каждой системы второго порядка используются величины, пропорциональные отсчетам

частотной характеристики, равномерно распределенным по углу на единичной окружности. Эти величины могут, конечно, быть получены из ДПФ импульсной характеристики. Если подлежащий выполнению фильтр является частотно-избирательным с одной или более полосой непропускания, то он может быть спроектирован так, чтобы частотные выборки в полосе непропускания были нулевыми, что уменьшает число подлежащих реализации систем второго порядка. Если большинство частотных отсчетов являются нулевыми, как в случае узкополосного низкочастотного или полосового фильтра, то структура на основе частотной выборки может потребовать меньше умножений, чем прямая форма построения. Конечно, реализация на основе частотной выборки будет всегда требовать большую память, чем прямая форма построения.

Второе преимущество следует из того, что полюсы и нули структуры фильтра зависят только от длины импульсной характеристики. Если входной сигнал подлежит обработке с помощью банка КИХ-фильтров (т. е. нескольких различных импульсных характеристик длиной N), то одна реализация множителя $(1 - z^{-N})$ и каждого блока второго порядка будет служить для всех фильтров. Кроме того, структура состоит из одинаковых функциональных блоков второго порядка, что дает возможность реализации этих блоков последовательно во времени.

Заметим в заключение, что структура на основе частотной выборки является частным случаем структуры, получаемой на основе интерполяционной формулы Лагранжа, которая представляет полином $H(z)$ через N его отсчетов в z -плоскости. Предложены и другие структуры КИХ-цепи, основанные на интерполяционных формулах Ньютона и Эрмита и разложении $H(z)$ в ряды Тейлора. Все эти структуры, включая структуру на основе частотной выборки, в общем случае требуют больше умножений и задержек, чем прямая или каскадная форма. Таким образом, полезность подобных структур заключена в возможных преимуществах по чувствительности к эффектам квантования и согласованности расчетной процедуры с реализацией системы.

4.4. Эффекты квантования параметров в цифровых фильтрах

Линейные системы с постоянными параметрами часто используются для выполнения формирования спектра или операции фильтрации. При создании подобных систем существует ряд важных: соображений. Во-первых, мы должны сделать выбор между БИХ- и КИХ-системами. Многие факторы могут влиять на этот выбор. Например, БИХ-система может потребовать наименьшее число задержек и умножителей на константу, чтобы удовлетворять заданной совокупности требований к фильтру. С другой стороны, тот факт, что КИХ-фильтры могут иметь строго линейную фазовую характеристику, оказывается часто важным аргументом в их пользу.

Для выбранного класса системы, т. е. БИХ- или КИХ-системы, мы должны определить передаточную функцию фильтра. В конечном счете для реализации системы в виде программы для ЦВМ или в виде специализированного устройства должна быть выбрана цифровая цепь или структурная схема. Приходится учитывать много соображений при выборе структуры системы: сложность математического обеспечения или конструкции и трудность получения параметров для фильтра. Кроме того, почти все созданные схемы, разработанные вплоть до настоящего времени, приводят к заданию передаточной функции с помощью параметров, которые предполагают неограниченную точность. Если мы выберем реализацию системы на основе программы для универсальной ЦВМ, то точность определения параметров может быть ограничена длиной слова памяти ЦВМ. При аппаратной реализации, конечно, желательно минимизировать длину регистров, которые должны обеспечить хранение параметров фильтра.

Поскольку коэффициенты, используемые при выполнении данного фильтра, в

общем случае не будут точными, то полюсы и нули передаточной функции будут в общем случае отличными от требуемых. Это смещение полюсов и нулей (в случае КИХ-систем только нулей) от требуемых расположений проявится в частотной характеристике, которая будет отличаться от заданной частотной характеристики. Если такие ошибки квантования велики, то система не сможет удовлетворить предъявляемым требованиям. Кроме того, в случае БИХ-системы один или несколько полюсов могут оказаться вне единичного круга, что приведет к неустойчивости системы и ее непригодности. В общем случае эффект квантования коэффициентов в сильной мере зависит от используемой для выполнения системы структуры.

Как мы видели, существует бесконечное множество реализаций цепи, которые обеспечивают данную передаточную функцию, если параметры такой цепи представляются с неограниченной точностью. Некоторые из этих структур будут менее чувствительными к квантованию параметров, т. е. передаточная функция такой реализации будет более близка, в некотором смысле, к требуемой передаточной функции системы. К сожалению, еще не существует систематического метода для определения наилучшей реализации при ограничениях на число умножителей, длину слова и число задержек. На практике выбор, как правило, ограничен формами цепей, рассмотренных выше. Вместо детального математического анализа проблемы чувствительности к значениям параметров обычный практический подход к определению приемлемого квантования параметров данной цепи сводится к моделированию. Ниже мы обсудим некоторые простые результаты, которые обеспечат понимание проблемы, связанной с чувствительностью к значениям параметров.

Эффекты квантования параметров проявляются в отклонениях характеристик фильтра от заданной частотной характеристики или, что эквивалентно, в смещении полюсов и нулей от требуемого положения. Поэтому одной из мер чувствительности данной реализации цепи к квантованию параметров является ошибка в положении полюсов и нулей, обусловленная ошибкой в умножителях на константы в цепи.

Чтобы показать, как квантование параметра влияет на положение полюсов и нулей, рассмотрим передаточную функцию, выраженную в виде

$$H(z) = \frac{\sum_{k=0}^M b(k)z^{-k}}{1 + \sum_{k=1}^N a(k)z^{-k}}.$$

Коэффициенты $a(k)$ и $b(k)$ являются требуемыми коэффициентами при прямой форме построения системы. При квантованных коэффициентах мы, в действительности, реализуем систему с передаточной функцией

$$\hat{H}(z) = \frac{\sum_{k=0}^M \hat{b}(k)z^{-k}}{1 + \sum_{k=1}^N \hat{a}(k)z^{-k}}$$

где $\hat{a}(k) = a(k) + \Delta a(k)$, $\hat{b}(k) = b(k) + \Delta b(k)$.

Допустим, что полюсы $H(z)$ располагаются в точках $z = p_i$, $i = 1, 2, \dots, N$, т. е. знаменатель полинома передаточной функции, представленный в виде сомножителей, равен

$$A(z) = 1 + \sum_{k=1}^N a(k)z^{-k} = \prod_{k=1}^N (1 - p_k z^{-1}). \quad (4.16)$$

Кроме того, определим полюсы $\hat{H}(z)$ при $z = p_i + \Delta p_i$, где $i = 1, 2, \dots, N$. Ошибка Δp_i может быть выражена через ошибки в коэффициентах:

$$\Delta p_i = \sum_{k=1}^N \frac{\partial p_i}{\partial a(k)} \Delta a(k), \quad i = 1, 2, \dots, N.$$

На основании (4.16) и того факта, что

$$\left(\frac{\partial A(z)}{\partial p_i} \right)_{z=p_i} \frac{\partial p_i}{\partial a(k)} = \left(\frac{\partial A(z)}{\partial a(k)} \right)_{z=p_i},$$

следует, что

$$\frac{\partial p_i}{\partial a(k)} = -p_i^{N-k} / \prod_{\substack{l=1 \\ l \neq i}}^N (p_i - p_l). \quad (4.17)$$

Выражение (4.17) является мерой чувствительности i -го полюса к изменению (ошибке) в k -м коэффициенте знаменателя полинома $H(z)$. Этот результат справедлив только для простых полюсов, как это видно из (4.16). Поскольку при прямой форме построения нули зависят только от коэффициентов числителя $b(k)$, аналогичный результат может быть получен для чувствительности нулей к ошибкам в коэффициентах $b(k)$.

Из (4.17) следует, что если полюсы (или нули) плотно сгруппированы, то возможно, что небольшие ошибки в коэффициентах могут обусловить большие смещения полюсов (или нулей). Это можно увидеть при рассмотрении знаменателя выражения (4.17). Каждый множитель $(p_i - p_l)$ можно представить как вектор в z -плоскости, как показано на рис. 4.21

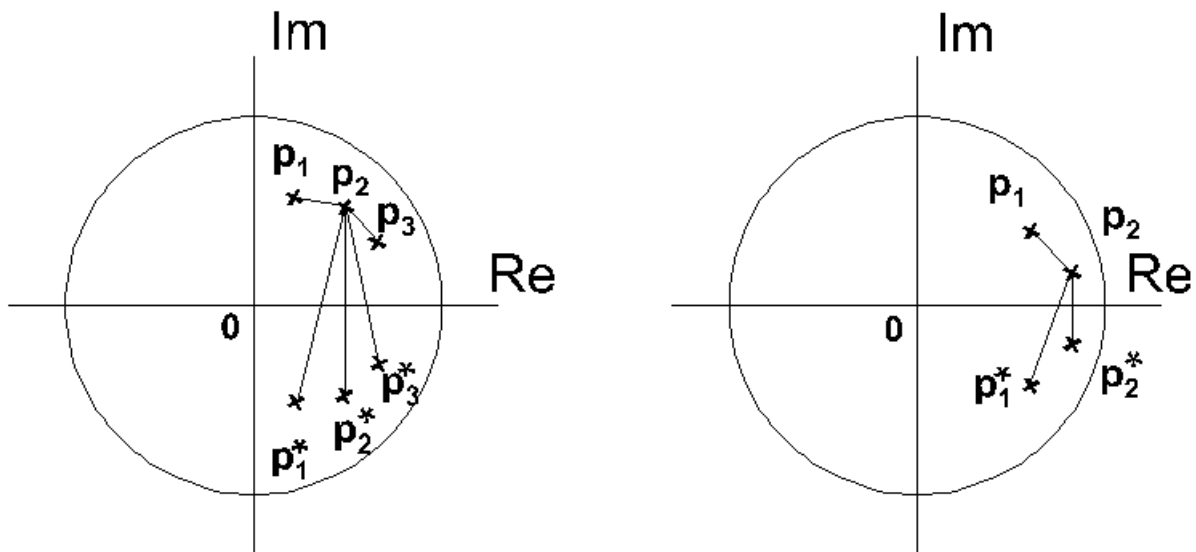


Рис.4.21. Представление сомножителей выражения (4.17) в виде векторов на z -плоскости в случаях: а) фильтра с узкой полосой пропускания; б) узкополосного фильтра нижних частот

Величина знаменателя в (4.17) равна произведению длин векторов из всех оставшихся полюсов до полюса p_i . Поэтому если полюсы (или нули) оказываются плотно сгруппированными, как на рис. 4.21а соответственно фильтру с узкой полосой пропускания или как на рис. 4.21б соответственно узкополосному фильтру нижних частот, то можно ожидать, что полюсы при прямой форме построения будут довольно чувствительны к ошибкам квантования коэффициентов. К тому же очевидно, что чем больше число корней, тем выше чувствительность.

При каскадной и параллельной формах построения систем, с другой стороны, каждая пара комплексно-сопряженных полюсов реализуется отдельно. Таким образом, ошибка для данного полюса является не зависимой от его расстояния до других полюсов системы. Поэтому в общем случае, с точки зрения квантования параметров, каскадная и параллельная формы оказываются предпочтительными перед прямыми

формами. Это особенно справедливо в случае узкополосных частотно-избирательных фильтров с плотно сгруппированными полюсами и нулями.

4.5. Методы проектирования и расчета цифровых фильтров

В наиболее общем смысле цифровой фильтр представляет собой дискретную линейную систему с постоянными параметрами, которая реализуется на основе использования арифметического устройства с ограниченной точностью. Проектирование цифрового фильтра включает три основных этапа: 1) определение требуемых свойств системы; 2) аппроксимацию этих требований на основе использования физически реализуемых дискретных систем; 3) реализацию системы при использовании арифметики с ограниченной точностью. Несмотря на то что эти три этапа не являются независимыми, мы сосредоточим внимание в первую очередь на втором этапе, поскольку первый из них в сильной степени зависит от конкретного применения, а третий требует специального рассмотрения.

При практической постановке задачи наиболее характерен случай, когда требуемый цифровой фильтр должен быть использован для фильтрации цифрового сигнала, который получен из аналогового сигнала путем образования периодических выборок. Требования как для аналоговых, так и для цифровых фильтров часто (но не всегда) задаются в частотной области, как, например, в случае частотно-избирательных фильтров, таких, как фильтры нижних или верхних частот и полосовые фильтры. При заданной частоте дискретизации структура цифрового фильтра может быть получена путем преобразования частотных требований для аналогового фильтра в частотные требования для цифрового фильтра. При этом аналоговые частоты определяются в герцах, а цифровые частоты - в единицах частоты в радианах или углах единичной окружности с точкой $z = -1$, соответствующей половине частоты дискретизации. Существуют, однако, применения, в которых подлежащий фильтрации цифровой сигнал не формируется путем образования периодических выборок аналоговой временной функции, и, как известно, кроме образования периодических выборок существует множество способов представления аналоговых временных функций с помощью последовательностей. К тому же в большинстве методов расчета, которые будут рассмотрены, период дискретизации не играет роли в процедуре аппроксимации.

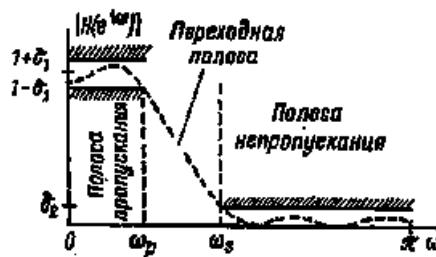


Рис.4.22. Допустимые пределы для аппроксимации идеального фильтра нижних частот

Отдельной проблемой является определение соответствующего набора требований для конкретного цифрового фильтра. Например, в случае фильтра нижних частот такие требования часто принимают вид допусков на ошибки аппроксимации, как показано на рис.4.22. На этом рисунке пределы допустимой ошибки аппроксимации определены горизонтальными линиями со штрихами. Заметим также, что достаточно начертить кривую только для $0 \leq \omega \leq \pi$, поскольку остальная часть может быть выведена на основе свойств симметрии. Пунктирная кривая представляет частотную характеристику системы, удовлетворяющую заданным требованиям.

В данном случае имеется полоса пропускания, в пределах которой характеристика должна аппроксимироваться величиной, равной 1 с ошибкой $\pm \delta_1$, т.е. $1 - \delta_1 \leq |H(e^{j\omega})| \leq 1 + \delta_1$, $|\omega| \leq \omega_p$. Имеется *полоса непропускания*, в которой характеристика должна аппроксимироваться нулевым значением с ошибкой не больше, чем δ_2 , т. е.

$|H(e^{j\omega})| \leq \delta_2$, $\omega_a \leq |\omega| \leq \pi$. Граничные частоты для полосы пропускания ω_p и для полосы непропускания ω_a заданы с помощью угловой частоты z -плоскости (цифровые частоты). Чтобы сделать возможной аппроксимацию идеального фильтра нижних частот таким способом, мы должны также представить *переходную полосу* ненулевой ширины ($\omega_a - \omega_p$), в которой характеристика плавно спадает от значения в полосе пропускания до значения в полосе непропускания. Многие из фильтров, используемых на практике, задаются такой совокупностью допусков (но без ограничений на фазовую характеристику), кроме тех, которые налагаются требованиями устойчивости и физической реализуемости.

При данной совокупности требований следующим этапом является нахождение линейной дискретной системы, характеристика которой изменяется в пределах указанных допусков. С этой точки зрения проблема расчета фильтра становится проблемой аппроксимации. В случае БИХ-систем мы должны аппроксимировать требуемую частотную характеристику с помощью рациональной функции, тогда как в случае КИХ-системы мы имеем дело с полиномиальной аппроксимацией. Для удобства рассмотрение построено так, чтобы провести различие между методами расчета, которые соответствуют БИХ- и КИХ-фильтрам. Будут рассмотрены различные методы расчета для обоих типов фильтров. Эти методы включают как процедуры в замкнутой форме, которые сводятся к подстановке значений заданных параметров в расчетные формулы, так и метод оптимизации, где решение получается с помощью итерационной процедуры.

4.5.1. Расчет цифровых БИХ-фильтров по данным аналоговых фильтров

Традиционный подход к расчету цифровых БИХ-фильтров включает преобразование аналогового фильтра в цифровой фильтр, удовлетворяющий заданным требованиям. Это является разумным подходом, потому что

1) способы расчета аналоговых фильтров являются хорошо отработанными и, поскольку могут быть получены полезные результаты, оказывается выгодным использовать уже развитые для аналоговых фильтров процедуры расчета;

2) многие полезные методы расчета аналоговых фильтров имеют сравнительно простые расчетные процедуры в замкнутой форме. Поэтому методы расчета цифрового фильтра, основанные на таких аналоговых расчетных формулах, являются довольно простыми;

3) во многих приложениях представляет интерес использование цифрового фильтра для моделирования работы аналогового линейного с постоянными параметрами фильтра.

Рассмотрим передаточную функцию аналоговой системы

$$H_a(p) = \frac{\sum_{k=0}^M d_k p^k}{\sum_{k=0}^N c_k p^k} = \frac{Y_a(p)}{X_a(p)},$$

где $x_a(t)$ и $y_a(t)$ — соответственно входной и выходной сигналы, а $X_a(p)$ и $Y_a(p)$ — их преобразования Лапласа. Предполагается, что $H_a(p)$ была получена с помощью одного из установившихся методов аппроксимации, используемых в расчетах аналогового фильтра. (Примеры рассмотрены ниже.) Входной и выходной сигналы такой системы связаны между собой с помощью интеграла свертки

$$y_a(t) = \int_{-\infty}^{\infty} x_a(\tau) h_a(t - \tau) d\tau,$$

где $h_a(t)$ — импульсная характеристика, являющаяся обратным преобразованием Лапласа от $H_a(p)$. С другой стороны, аналоговая система, имеющая передаточную функцию $H_a(p)$, может быть описана дифференциальным уравнением

$$\sum_{k=0}^N c_k \frac{d^k y_a(t)}{dt^k} = \sum_{k=0}^M d_k \frac{d^k x_a(t)}{dt^k}$$

Соответствующая рациональная передаточная функция для цифровых фильтров имеет вид

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^M b(k)z^{-k}}{\sum_{k=0}^N a(k)z^{-k}}$$

Входной и выходной сигналы связаны между собой с помощью сверточной суммы

$$y(n) = \sum_{k=-\infty}^{\infty} x(k)h(n-k)$$

или, что то же, с помощью разностного уравнения

$$\sum_{k=0}^N a(k)y(n-k) = \sum_{k=0}^M b(k)x(n-k).$$

При преобразовании аналоговой системы в цифровую мы должны поэтому получить либо $H(z)$, либо $h(n)$ из расчета аналогового фильтра. При таких преобразованиях, как правило, требуется, чтобы существенные свойства аналоговой частотной характеристики сохранялись в частотной характеристике получающегося в результате преобразования цифрового фильтра. Под этим подразумевается, что необходимо мнимую ось из s -плоскости отобразить в единичную окружность на z -плоскости. Второе требование состоит в том, что устойчивый аналоговый фильтр должен быть преобразован в устойчивый цифровой фильтр. Это означает, что если аналоговая система имеет полюсы только в левой половине s -плоскости, то цифровой фильтр должен иметь полюсы только внутри единичного круга. Эти ограничения являются основными для всех методов, которые будут рассмотрены ниже.

4.5.1.1. Метод инвариантности импульсной характеристики

Одна из процедур преобразования данных расчета аналогового фильтра для расчета цифрового фильтра соответствует выбору в качестве импульсной характеристики цифрового фильтра равномерно распределенных отсчетов импульсной характеристики аналогового фильтра. Иначе говоря, $h(n) = h_a(nT)$, где T - период дискретизации.

Частотная характеристика цифрового фильтра связана с частотной характеристикой аналогового фильтра соотношением

$$H(e^{j\omega}) = \frac{1}{T} \sum_{n=-\infty}^{\infty} K_a \left[j \left(\frac{\omega}{T} - n \frac{2\pi}{T} \right) \right].$$

Очевидно, что если $K_a(j\Omega) = 0$, $|\Omega| \geq \pi/T$, то $H(e^{j\omega}) = (1/T)K_a(j\omega/T)$, $|\omega| \leq \pi$. К сожалению, любой практический аналоговый фильтр не будет обладать резко ограниченной полосой, и, следовательно, имеет место явление наложения между соседними спектральными полосами, как показано на рис. 4.23.

Из-за явления наложения, которое возникает в процессе дискретизации, частотная характеристика результирующего цифрового фильтра не будет идентична исходной аналоговой частотной характеристике.

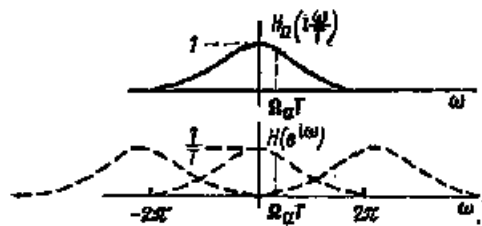


Рис.4.23. Графическое представление эффектов наложения при расчете фильтра методом инвариантности импульсной характеристики

Чтобы исследовать интерпретацию метода расчета на основе импульсной инвариантности с помощью соотношения между p - и z -плоскостями, рассмотрим передаточную функцию аналогового фильтра, представленную в виде разложения на простые дроби так, что

$$K_p(p) = \sum_{k=1}^N \frac{A_k}{p - p_k}.$$

Соответствующая импульсная характеристика имеет вид

$$h_a(t) = \sum_{k=1}^N A_k e^{p_k t} u(t).$$

где $u(t)$ - аналоговая единичная ступенчатая функция. Тогда импульсная характеристика цифрового фильтра равна

$$h(n) = h_a(nT) = \sum_{k=1}^N A_k e^{p_k nT} u(n) = \sum_{k=1}^N A_k (e^{p_k T})^n u(n).$$

Передаточная функция цифрового фильтра $H(z)$ поэтому определяется выражением

$$H(z) = \sum_{k=1}^N \frac{A_k}{1 - e^{p_k T} z^{-1}}.$$

При сравнении $K_a(p)$ и $H(z)$ замечаем, что полюс в точке $p = p_k$ из p -плоскости преобразуется в полюс в точке $e^{p_k T}$ на z -плоскости, а коэффициенты в разложении на простые дроби $K_a(p)$ и $H(z)$ равны. Если аналоговый фильтр является устойчивым (соответственно чему действительная часть p_k меньше нуля), то величина $e^{p_k T}$ будет меньше единицы, так что соответствующий полюс в цифровом фильтре находится внутри единичного круга и, следовательно, цифровой фильтр является также устойчивым. Заметим, что передаточная функция цифрового фильтра $H(z)$ соответствует параллельной форме структуры, каждая из ветвей которой в общем случае есть комплексный резонатор.

Несмотря на то, что полюсы p -плоскости отображаются в полюсы z -плоскости в соответствии с соотношением $z_k = e^{p_k T}$, важно отметить, что процедура на основе инвариантности импульсной характеристики не полностью соответствует отображению p -плоскости на z -плоскость согласно такому или фактически любому другому соотношению. В частности, нули цифровой передаточной функции являются функцией полюсов и коэффициентов A_k в разложении на простые дроби, и они в общем случае не будут отображаться таким же образом, как отображались полюсы.

Пример. В качестве примера определения цифрового фильтра из аналогового фильтра на основе инвариантности импульсной характеристики рассмотрим аналоговую передаточную функцию $H_a(p)$, заданную в виде

$$H_a(p) = \frac{p + a}{(p + a)^2 + b^2} = \frac{1}{2} \left(\frac{1}{p + a + jb} + \frac{1}{p + a - jb} \right).$$

Соответствующая передаточная функция цифрового фильтра, полученного методом импульсной инвариантности, равна

$$H(z) = \frac{1}{2} \left(\frac{1}{1 - e^{-aT} e^{-jbT} z^{-1}} + \frac{1}{1 - e^{-aT} e^{jbT} z^{-1}} \right) = \frac{1 - (e^{-aT} \cos bT) z^{-1}}{(1 - e^{-aT} e^{-jbT} z^{-1})(1 - e^{-aT} e^{jbT} z^{-1})} =$$

$$= \frac{1 - (e^{-aT} \cos bT) z^{-1}}{1 - (e^{-aT} \cos bT) z^{-1} + e^{-2aT} z^{-2}}.$$

Цифровой фильтр, следовательно, имеет один нуль в начале координат и нуль в точке $z = e^{-aT} \cos bT$.

На рис.4.24 показано расположение полюсов и нулей для $H_a(p)$ в p -плоскости и для $H(z)$ в z -плоскости совместно с аналоговой и цифровой частотными характеристиками. В данном случае частотная характеристика аналоговой системы спадает довольно медленно по отношению к частоте дискретизации, и поэтому эффекты наложения в цифровой частотной характеристике становятся очевидными.

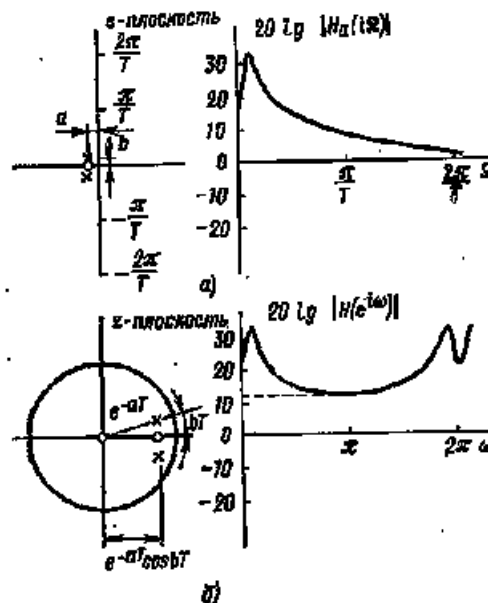


Рис.4.24. Расположение полюсов и нулей и частотная характеристика:

а) аналоговой системы второго порядка; б) дискретной системы, полученной путем дискретизации импульсной характеристики аналоговой системы

Необходимо отметить, что когда аналоговый фильтр является «достаточно ограниченным по полосе», то вышеприведенная процедура дает цифровой фильтр, частотная характеристика которого имеет вид $H(e^{j\omega}) \approx (1/T)K_a(j\omega/T)$, $|\omega| \leq \pi$. Таким образом, при высоких частотах дискретизации (малом T) цифровой фильтр может иметь чрезвычайно большое усиление. По этой причине в общем случае целесообразно использовать выражение вида

$$H(z) = \sum_{k=1}^N \frac{TA_k}{1 - e^{p_k T} z^{-1}}.$$

Это означает, что импульсной характеристикой является $h(n) = Th_a(nT)$.

Основой для метода импульсной инвариантности, как было указано выше, является выбор импульсной характеристики для цифрового фильтра, которая будет подобной в некотором смысле импульсной характеристике аналогового фильтра. Использование этой процедуры часто мотивируется не столько желанием сохранить форму импульсной характеристики, сколько знанием того факта, что если аналоговый фильтр является ограниченным по полосе частот, то частотная характеристика цифрового фильтра будет точно аппроксимировать аналоговую частотную характеристику. Тем не менее в некоторых случаях расчета фильтров основной задачей может оказаться создание некоторых свойств отклика, таких, как импульсная характеристика или реакция на единичный скачок. В подобных случаях естественный подход должен быть связан с расчетом цифрового фильтра на основе импульсной

инвариантности или *процедуры инвариантности к скачку*. В последнем случае в качестве отклика цифрового фильтра на дискретизованную функцию единичного скачка принимаются отсчеты отклика аналогового фильтра на единичный скачок. Если аналоговый фильтр обладает хорошими параметрами переходной характеристики, такими, как малое время нарастания и небольшой выброс, то эти параметры должны быть сохранены в цифровом фильтре. Очевидно, эта концепция инвариантности формы сигнала может быть распространена на случаи сохранения формы выходного сигнала для многих входных сигналов.

Хотя в процедуре расчета на основе импульсной инвариантности вводятся искажения в частотную характеристику из-за эффекта наложения, соотношение между аналоговыми и цифровыми частотами является линейным и, следовательно, за исключением эффекта наложения, форма частотной характеристики сохраняется. Это отличает изложенное от процедур, подлежащих дальнейшему рассмотрению, в которых используются алгебраические преобразования. Необходимо еще отметить, что методы импульсной инвариантности оказываются подходящими только в случае фильтров, существенно ограниченных по полосе частот. Например, фильтры высоких частот или режекторные фильтры потребовали бы дополнительного ограничения полосы частот для того, чтобы избежать сильных искажений из-за эффекта наложения.

Другая процедура дискретизации аналогового фильтра заключается в замене производных в дифференциальном уравнении аналоговой системы на конечные разности [2,3]. Этот метод является стандартной процедурой при численном анализе и при цифровом моделировании аналоговых систем. Он основан на интуитивном представлении о том, что производная аналоговой временной функции может аппроксимироваться разностью между последовательными отсчетами ее, и по мере возрастания частоты дискретизации отсчеты оказываются ближе друг к другу, точность аппроксимации производной будет возрастать. Показано [2,3], что при таком преобразовании характеристики аналогового фильтра не сохраняются, а требуемые высокие частоты дискретизации приводят к весьма неэффективному варианту цифрового фильтра, поэтому для дискретизации аналоговых фильтров применяются иные методы.

Наибольшее распространение в практике синтеза БИХ- фильтров стандартных типов, к которым относятся фильтры нижних частот (ФНЧ), верхних частот (ФВЧ), полосно-пропускающие (ФПП) и полосно-заграждающие (ФПЗ) получил так называемый метод билинейного преобразования.

4.5.1.2. Метод билинейного преобразования

Процедура, приводящая к билинейному преобразованию, основана на интегрировании дифференциального уравнения и последующей численной аппроксимации интеграла. Например, рассмотрим уравнение первого порядка

$$c_1 y'_a(t) + c_0 y_a(t) = d_0 x(t), \quad (4.18)$$

для которого передаточная функция имеет вид

$$K_a(p) = \frac{d_0}{c_1 p + c_0}. \quad (4.19)$$

Представим $y_a(t)$ в виде интеграла от $y'_a(t)$, т.е.

$$y_a(t) = \int_{t_0}^t y'_a(t) dt + y_a(t_0).$$

В частности, при $t = nT$ и $t_0 = (n-1)T$ имеем

$$y_a(nT) = \int_{(n-1)T}^{nT} y'_a(t) dt + y_a[(n-1)T].$$

Аппроксимируя интеграл формулой трапеции, можно записать

$$y_a(nT) = y_a[(n-1)T] + \frac{T}{2} [y'_a(nT) + y'_a(n-1)T]. \quad (4.20)$$

Подставляя в (4.20) $y'_a(nT) = -\frac{c_0}{c_1} y_a(nT) + \frac{d_0}{c_1} x_a(nT)$, выраженное из (4.18), получим соотношение

$$y(n) - y(n-1) = \frac{T}{2} \left\{ -\frac{c_0}{c_1} [y(n) + y(n-1)] + \frac{d_0}{c_1} [x(n) + x(n-1)] \right\}. \quad (4.21)$$

Нахождение z -преобразования от (4.21) и его решение для $H(z)$ дает

$$H(z) = \frac{Y(z)}{X(z)} = \frac{d_0}{c_1 \left(\frac{2}{T} \right) \frac{1-z^{-1}}{1+z^{-1}} + c_0}. \quad (4.22)$$

Сравнивая (4.22) и (4.19), заметим, что $H(z)$ получается из $K_a(p)$ заменой

$$p = \frac{2}{T} \cdot \frac{1-z^{-1}}{1+z^{-1}}, \quad (4.23)$$

т.е.

$$H(z) = K_a(p) \Bigg|_{p = \frac{2}{T} \cdot \frac{1-z^{-1}}{1+z^{-1}}}. \quad (4.24)$$

Можно показать, что это справедливо и в общем случае, т.к. дифференциальное уравнение n -го порядка можно записать как систему n уравнений первого порядка вида (4.18). Из (4.23) следует, что

$$z = \frac{2/T + p}{2/T - p}. \quad (4.25)$$

Преобразование (4.24), использующее замену (4.23), называется билинейным. Характер этого преобразования проще уяснить, если обратиться к рис.4.25, где показано, каким образом p -плоскость отображается в z -плоскость.

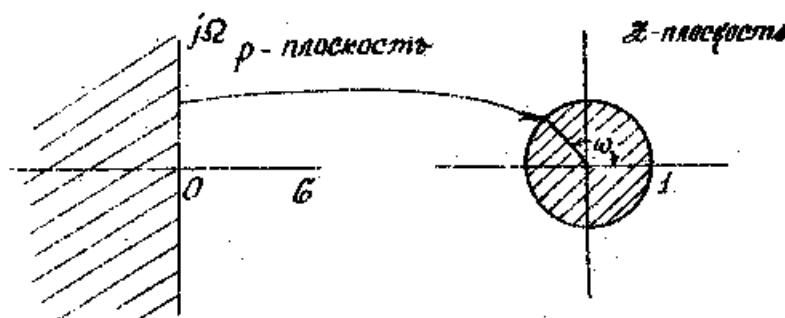


Рис. 4.25

Видно, что мнимая ось $j\Omega$ из p -плоскости отображается в окружность единичного радиуса на z -плоскости, левая полуплоскость отображается во внутреннюю область единичного круга. Это можно увидеть и из (4.25).

При $p=j\Omega$

$$z = \frac{2/T + j\Omega}{2/T - j\Omega} = e^{j\omega},$$

следовательно, $|z| = 1$.

При $\omega = 0, z = 1 (\omega = 0)$, и при $\Omega \rightarrow \infty, z \rightarrow -1 (\omega \rightarrow \pi)$.

При $p = \sigma + j\Omega$ имеем

$$z = \frac{\frac{2}{T} + \sigma + j\Omega}{\frac{2}{T} - \sigma - j\Omega}, \quad |z| = \frac{\sqrt{\left(\frac{2}{T} + \sigma\right)^2 + \Omega^2}}{\sqrt{\left(\frac{2}{T} - \sigma\right)^2 + \Omega^2}},$$

и, следовательно, при $\sigma < 0$ (для левой полуплоскости p) $|z| < 1$, т.е. точки располагаются внутри единичного круга в z -плоскости. Поэтому устойчивый аналоговый фильтр преобразуется в устойчивый цифровой фильтр.

Чтобы найти связь между частотами аналогового фильтра Ω и цифрового фильтра ω , положим в (3.6) $z = e^{j\omega}$, $p = j\Omega$ и получим

$$j\Omega = \frac{2}{T} \frac{1 - e^{-j\omega}}{1 + e^{-j\omega}} = \frac{2}{T} \frac{e^{j\frac{\omega}{2}} - e^{-j\frac{\omega}{2}}}{e^{j\frac{\omega}{2}} + e^{-j\frac{\omega}{2}}} = j \frac{2}{T} \operatorname{tg} \frac{\omega}{2}.$$

Следовательно,

$$\Omega = \frac{2}{T} \operatorname{tg} \frac{\omega}{2}, \quad (4.26)$$

что означает существенно нелинейный характер связи аналоговых Ω и цифровых ω частот. Поэтому расчет цифровых фильтров на основе билинейного преобразования оказывается полезным только тогда, когда искажение шкалы частот и связанная с ним деформация АЧХ могут быть скомпенсированы и для аппроксимации АЧХ выбрана идеальная кусочно-постоянная характеристика. К ним относятся частотно-избирательные фильтры стандартных типов: фильтр нижних частот (ФНЧ), фильтр верхних частот (ФВЧ), фильтр полосно-пропускающий (ФПП), фильтр полосно-заграждающий (ФПЗ).

Например, если необходимо рассчитать цифровой ФНЧ, то отыскивается аппроксимация для идеальной АЧХ, представленной на рис.4.26. Если рассчитан идеальный аналоговый ФНЧ, частота среза которого равна $\Omega_c = \frac{2}{T} \operatorname{tg} \frac{\omega_c}{2}$, то в результате билинейного преобразования (4.24) получится требуемый идеальный цифровой ФНЧ (рис.4.26).

Конечно, реализовать идеальный фильтр этого типа невозможно ни в аналоговом, ни в цифровом варианте. В общем случае мы аппроксимировали бы подобную характеристику фильтра, допуская



Рис.4.26. Частотная характеристика идеального фильтра нижних частот

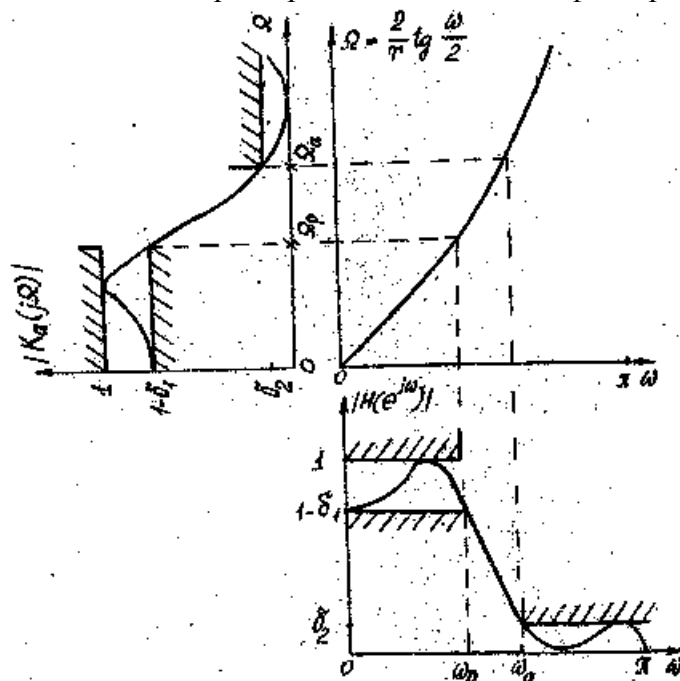


Рис.4.27. Деформация шкалы частот при преобразовании аналогового фильтра нижних частот в цифровой фильтр нижних частот

некоторое отклонение от единицы в полосе пропускания и некоторое отклонение от нуля в полосе непропускания с переходной полосой ненулевой ширины. На рис.4.27. показаны отображение аналоговой АЧХ в соответствующую цифровую АЧХ и допуски для их параметров.

Если частоты среза аналогового фильтра предварительно скорректированы в соответствии с соотношениями:

$$\Omega_p = \frac{2}{T} \operatorname{tg} \frac{\omega_p}{2}, \quad \Omega_a = \frac{2}{T} \operatorname{tg} \frac{\omega_a}{2},$$

где ω_p и ω_a - частоты среза цифрового фильтра, то при преобразовании (4.24) аналогового фильтра в цифровой, последний будет удовлетворять заданным требованиям.

Типичными примерами частотно-избирательных аналоговых фильтров являются фильтры Баттерворта, Чебышева и эллиптические (Золотарева-Кауэра) [2].

Аналоговый фильтр Баттерворта обладает монотонной характеристикой в полосах пропускания и непропускания. Чебышевский фильтр имеет характеристику с равновеликими пульсациями в полосе пропускания и монотонную в полосе непропускания. Эллиптический фильтр имеет равновеликие пульсации как в полосе пропускания, так и в полосе непропускания. Очевидно, что эти свойства будут сохраняться, когда аналоговый фильтр отображается в цифровой с помощью билинейного преобразования. Это иллюстрируется частотными характеристиками на рис.4.27.

Заметим в заключение, что хотя билинейное преобразование может быть эффективно использовано для отображения АЧХ из p -плоскости на z -плоскость,

деформация шкалы частот будет проявляться в искажениях фазо-частотной характеристики фильтра. При этом и импульсные характеристики аналогового и цифрового фильтров будут различаться.

Прежде чем перейти к методике проектирования цифровых БИХ-фильтров, рассмотрим расчет аналоговых фильтров для некоторых широко используемых видов аппроксимаций.

4.5.2 Обзор методов расчета аналоговых ФНЧ

Как показано выше, расчет цифровых БИХ-фильтров на основе аналогового прототипа требует понимания методов расчета фильтров непрерывного времени. Поэтому ниже мы приведем основные расчетные соотношения для аналоговых ФНЧ, использующих аппроксимации Баттерворта, Чебышева (типа 1 и 2) и эллиптическую (Золотарева-Кауэра).

1. Фильтры Баттерворта

Эти фильтры характеризуются тем, что их АЧХ являются максимально плоской в полосе пропускания. Выражение для квадрата фильтра Баттерворта n -го порядка имеет вид:

$$|K_a(j\Omega)|^2 = \frac{1}{1 + (\Omega/\Omega_c)^{2n}}, \quad (4.27)$$

а при введении нормированной к частоте среза Ω_c частоты $x = \Omega / \Omega_c$

$$|K_a(jx)|^2 = \frac{1}{1 + x^{2n}}. \quad (4.28)$$

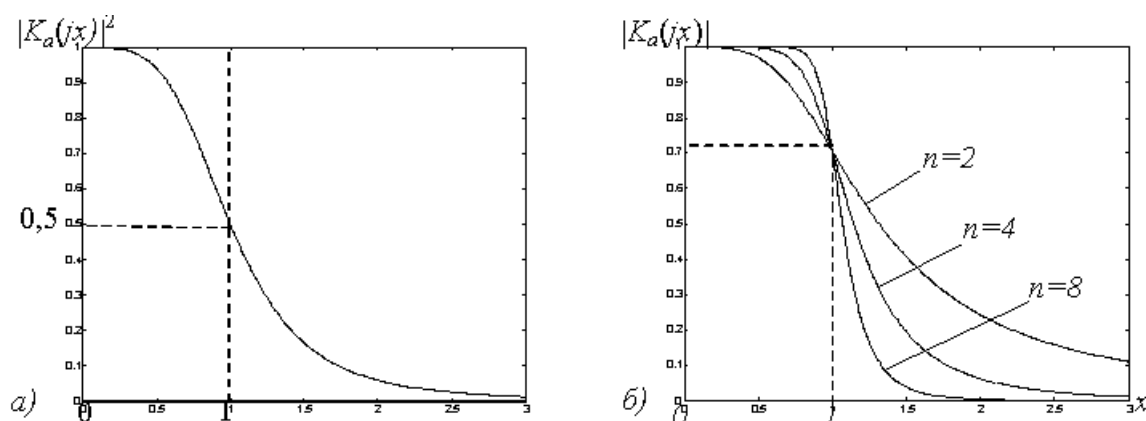


Рис. 4.28

На рис.4.28а представлен график квадрата АЧХ, а рис.4.28б отображает зависимость АЧХ фильтра Баттерворта от его порядка n . Аналитически продолжая $K_a(jx)$ на комплексную плоскость p , заменив в (4.28) $jx = p$, $x^2 = (-1)p^2$ получим

$$|K_a(p)|^2 = \frac{1}{1 + (-1)^n p^{2n}}. \quad (4.29)$$

Полюсы этой функции находятся из уравнения:

$$p^{2n}(-1)^n + 1 = 0, \quad p^{2n} = (-1)^{-(n-1)}. \quad (4.30)$$

Поскольку

$$(-1) = e^{j\pi}, \quad (-1)^{n-1} = e^{j\pi(n-1)} = e^{-j2k\pi},$$

где k -любое целое число, то для k -го корня уравнения (4.30) получаем соотношение

$$p_k = e^{j\pi[(n-1)+2k]/2n}, \quad k = 1, 2, \dots, 2n, \quad (4.31)$$

причем число корней равно $2n$. Очевидно, что все полюсы p_k находятся на единичной окружности в p -плоскости, на одинаковом угловом расстоянии друг от друга, поскольку модули p_k равны 1, а аргументы

$$\varphi_k = \pi[n + (2k - 1)]/2n = \pi[1/2 + (2k - 1)/2n].$$

На рис.4.29 представлены расположения полюсов для $n = 3$ и $n = 4$.

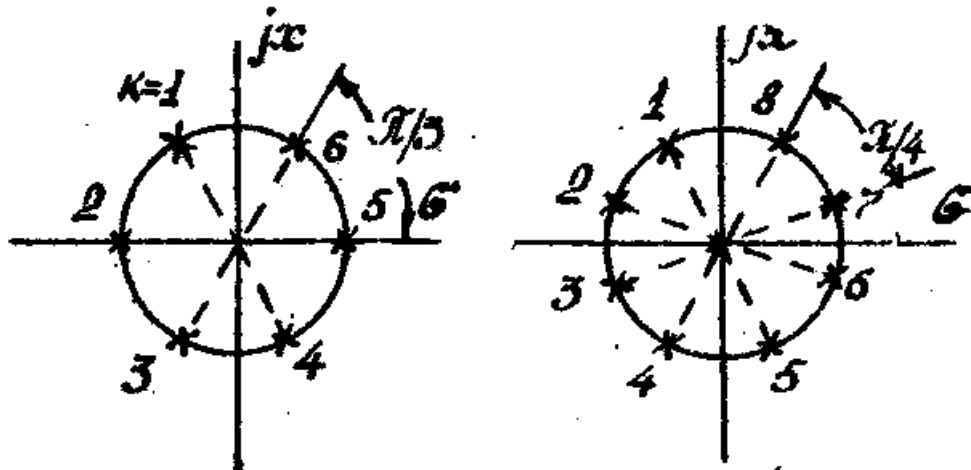


Рис. 4.29.

Мы видим, что полюсы квадрата АЧХ образуют пары так, что для полюса $p = p_k$ имеется полюс $p = -p_k$. Следовательно, для получения $K_a(p)$ на основе $|K_a(p)|^2$ нужно выбрать по одному полюсу из каждой такой пары. Если ограничить фильтр условиями устойчивости и физической реализуемости, которые, как правило, требуются, то нужно выбрать полюсы левой полуплоскости. Поэтому передаточную функцию фильтра можно представить в виде

$$K_a(p) = \frac{1}{\prod_{k=1}^n (p - p_k)}, \quad (4.32)$$

где

$$p_k = e^{j\pi[1/2 + (2k-1)/2n]}, \quad k = 1, 2, \dots, n.$$

или в тригонометрической форме

$$p_k = -\sin \frac{2k-1}{2n} \pi + j \cos \frac{2k-1}{2n} \pi, \quad k = 1, 2, \dots, n.$$

Заметим также, что все полюсы в (4.32) образуют комплексно-сопряженные пары, кроме одного полюса на вещественной оси, соответствующего $k=(n+1)/2$ при нечетном n . Объединяя множители с комплексно-сопряженными полюсами в (4.32) можно получить соотношение:

$$K_a(p) = \begin{cases} \frac{1}{\prod_{k=1}^{n/2} (p^2 - 2\alpha_k p + 1)}, & n - \text{четное}, \\ \frac{1}{(p+1)^{(n-1)/2} \prod_{k=1}^{(n-1)/2} (p^2 - 2\alpha_k p + 1)}, & n - \text{нечетное}, \end{cases} \quad (4.33)$$

где

$$\alpha_k = -\sin\left[\frac{(2k-1)}{2n}\pi\right].$$

Из приведенных выше соотношений очевидно, что порядок фильтра n полностью определяет весь фильтр. На практике порядок фильтра рассчитывают, исходя из условия обеспечения определенного ослабления

$$A = \frac{1}{|K_a(jx_3)|}$$

на некоторой заданной частоте $x_3 > 1$. При исчислении ослабления в децибелах (дб) имеем

$$A_{db} = 20 \lg(1 + x_3^{2n})^{1/2} = 10 \lg(1 + x_3^{2n}).$$

откуда

$$1 + x_3^{2n} = 10^{0.1A}, \quad x_3^{2n} = 10^{0.1A} - 1$$

и, следовательно,

$$n = \frac{\lg(10^{0.1A} - 1)}{2 \lg x_3} \approx \frac{A_{db}}{20 \lg x_3}. \quad (4.34)$$

Например, фильтр Баттерворта, обеспечивающий ослабление на частоте $x_3 = 2$ равное $A = 66 \text{ дб}$, должен иметь порядок $n \approx 66 / (20 \lg 2) \approx 10,96$. Следовательно, требуемые параметры обеспечит фильтр с $n = 11$.

2. Фильтры Чебышева

Отличительной чертой фильтров Чебышева является наименьшая величина максимальной ошибки аппроксимации характеристики идеального ФНЧ в заданной полосе частот. В действительности ошибка аппроксимации представляется в заданной полосе равновеликими пульсациями, т.е. она изменяется между максимумами равной величины. В зависимости от того, минимизируется ошибка аппроксимации в полосе пропускания или в полосе непропускания различают фильтры Чебышева типа 1 (рис.4.30а) и типа 2 (рис.4.30б).

Фильтры Чебышева типа 1 имеют только полюсы и обеспечивают равновеликие пульсации АЧХ в полосе пропускания и монотонное изменение ослабления в полосе непропускания.

Квадрат АЧХ фильтра Чебышева типа 1 n -го порядка имеет вид:

$$|K_a(jx)|^2 = \frac{1}{1 + \varepsilon^2 T_n^2(x)}, \quad (4.35)$$

где $x = \Omega / \Omega_p$ - частота, нормированная к частоте среза Ω_p , а $T_n(x)$ - полином Чебышева n -го порядка, по определению равный

$$T_n(x) = \begin{cases} \cos(n \cdot \arccos x), & |x| \leq 1, \\ \operatorname{ch}(n \cdot \operatorname{arch} x), & |x| > 1, \end{cases} \quad (4.36)$$

ε - параметр, ограничивающий амплитуду пульсаций.

Для полиномов Чебышева справедлива рекуррентная формула

$$T_{m+1}(x) = 2xT_m(x) - T_{m-1}(x), \quad T_0(x) = 1, \quad T_1(x) = x. \quad (4.37)$$

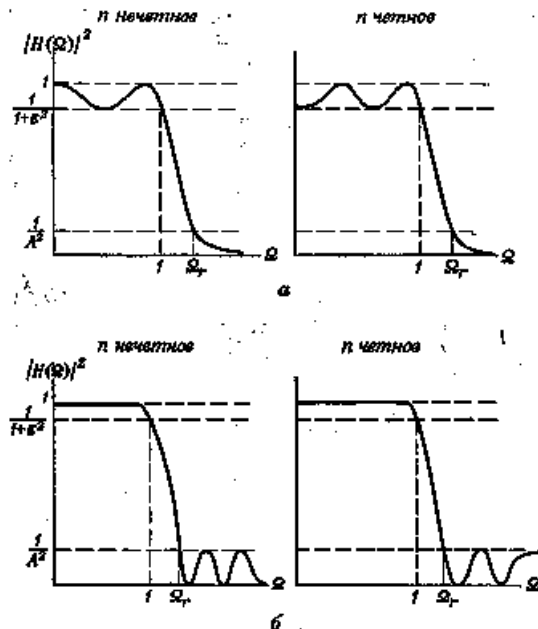


Рис.4.30. Общий вид квадрата АЧХ фильтров Чебышева типа 1 (а) и типа 2 (б).

Свойство оптимальности фильтров Чебышева типа 1 заключается в том, что не существует какого-либо другого фильтра порядка n , содержащего только полюсы, который имел бы такие же или лучшие характеристики и в полосе пропускания, и полосе непропускания.

Определим передаточную функцию фильтра подстановкой в (4.35) $x = p/j$:

$$|K_a(p)|^2 = \frac{1}{1 + \varepsilon^2 T_n^2(p/j)},$$

с последующим нахождением полюсов из уравнения

$$1 + \varepsilon^2 T_n^2(p_k/j) = 0.$$

Решение уравнения для полюсов в левой полуплоскости p имеет вид

$$p_k = -\operatorname{sh}\Phi_1 \sin\left[\frac{(2k-1)}{2n}\pi\right] + j\operatorname{ch}\Phi_1 \cos\left[\frac{(2k-1)}{2n}\pi\right], \quad (4.38)$$

где $\Phi_1 = \frac{1}{n} \operatorname{arsh}(1/\varepsilon)$, $k = 1, 2, \dots, n$.

При этом передаточная функция примет вид:

$$K_a(p) = \frac{K_0}{\prod_{k=1}^n (p - p_k)}, \quad (4.39)$$

где нормировочный коэффициент K_0 определяется из условий $K_a(0) = \frac{1}{\sqrt{1 + \varepsilon^2}}$

для нечетных n и $K_a(0)=1$ для четных n :

$$K_0 = \begin{cases} \prod_{k=1}^n (-p_k), & n - \text{нечетное}, \\ \frac{1}{\sqrt{1+\varepsilon^2}} \prod_{k=1}^n (-p_k), & n - \text{четное}. \end{cases}$$

Заметим, что полюсы фильтра Чебышева типа 1 находятся через полюсы фильтра Баттерворта того же порядка, если абсциссу каждого полюса умножить на $sh\Phi_1$, а ординату на $ch\Phi_1$. При выполнении расчетов следует пользоваться соотношениями

$$arshx = \ln(x + \sqrt{x^2 + 1}), \quad archx = \ln(x + \sqrt{x^2 - 1}).$$

Фильтры Чебышева типа 2 (их еще называют инверсными фильтрами Чебышева) обеспечивают монотонное изменение ослабления в полосе пропускания и равновеликие пульсации в полосе непропускания. Квадрат АЧХ такого фильтра n -го порядка представляется в виде

$$|K_a(jx)|^2 = \frac{1}{1 + \varepsilon^2 [T_n(x_a)/T_n(x_a/x)]^2}, \quad (4.40)$$

где x_a - граничная частота полосы непропускания (рис.4.30б). Фильтры Чебышева типа 2 имеют и полюсы, и нули. Нули являются чисто мнимыми и находятся в точках

$$p_{0k} = j \frac{x_a}{\cos\left(\frac{2k-1}{2n}\pi\right)}, \quad k = 1, 2, \dots, n. \quad (4.41)$$

Полюсы левой полуплоскости p находятся по формулам

$$p_k = \sigma_k + j\Omega_k, \quad \sigma_k = \frac{x_a \alpha_k}{\alpha_k^2 + \beta_k^2}, \quad \Omega_k = -\frac{x_a \beta_k}{\alpha_k^2 + \beta_k^2}, \quad (4.42)$$

где

$$\alpha_k = -sh\Phi \sin\left[\frac{(2k-1)}{2n}\pi\right], \quad \beta_k = ch\Phi \cos\left[\frac{(2k-1)}{2n}\pi\right], \quad \Phi = \frac{1}{n} archA, \quad k=1, 2, \dots, n,$$

$$A = \frac{1}{|K_a(x_a)|}.$$

Передаточная функция такого фильтра может быть представлена в форме

$$K_a(p) = K_0 \prod_{k=1}^n \frac{p - p_k}{p - p_{0k}}, \quad (4.43)$$

где нормировочный множитель

$$K_0 = \prod_{k=1}^n \frac{p_k}{p_{0k}}$$

определяется из условия $K_a(0) = 1$. Фильтры Чебышева полностью определяются любыми тремя из следующих четырех параметров:

- 1) n - порядок фильтра;

- 2) ε - параметр, характеризующий пульсации;
- 3) x_a - граничная частота полосы не пропускания;
- 4) A - параметр, характеризующий величину ослабления в полосе не пропускания.

Порядок фильтра, необходимый для обеспечения заданных ε , A , x_a определяется формулой

$$n = \frac{\lg\left(g + \sqrt{g^2 - 1}\right)}{\lg\left(x_a + \sqrt{x_a^2 - 1}\right)}, \quad g = \sqrt{\frac{A^2 - 1}{\varepsilon^2}}. \quad (4.44)$$

Пример. Рассчитаем фильтр Чебышева минимального порядка, удовлетворяющий следующим условиям:

- пульсации в полосе пропускания равны 2дБ;
- переходное отношение $1/x_a=0,781$;
- ослабление в полосе не пропускания 30дБ.

Используя рис.4.30, найдем параметры фильтра ε , A , x_a по заданным характеристикам

$$-2\delta = 20 \lg \frac{1}{\sqrt{1 + \varepsilon^2}}, \quad \varepsilon = 0,764,$$

$$-30\delta = 20 \lg(1/A), \quad A = 31,62,$$

$$x_a = 1/0,781 = 1,28.$$

Затем по (4.44) найдем $g=41,33$, и $n=6,03$.

3. Эллиптические фильтры (Золотарева-Кауэра)

Эти фильтры характеризуются тем, что их АЧХ имеет равновеликие пульсации как в полосе пропускания, так и в полосе не пропускания. Можно показать, что с точки зрения минимальной ширины переходной полосы эллиптические фильтры являются наилучшими, т.е. для заданных порядка фильтра и уровня пульсаций не существует других фильтров с более быстрым переходом от полосы пропускания к полосе не пропускания.

$$|K_a(jx)|^2 = \frac{1}{1 + \varepsilon^2 R_n^2(x, x_a)},$$

где $R_n^2(x, x_a)$ - эллиптическая функция Якоби. На рис.4.31 представлен типичный вид графиков функций $R_n^2(x, x_a)$ и $|K_a(jx)|^2$. Передаточную функцию эллиптического фильтра можно представить через его нули p_{0k} и полюсы p_k в форме

$$K_a(p) = K_0 \prod_{k=1}^n \frac{p - p_{0k}}{p - p_k} \quad k = 1, 2, \dots, n,$$

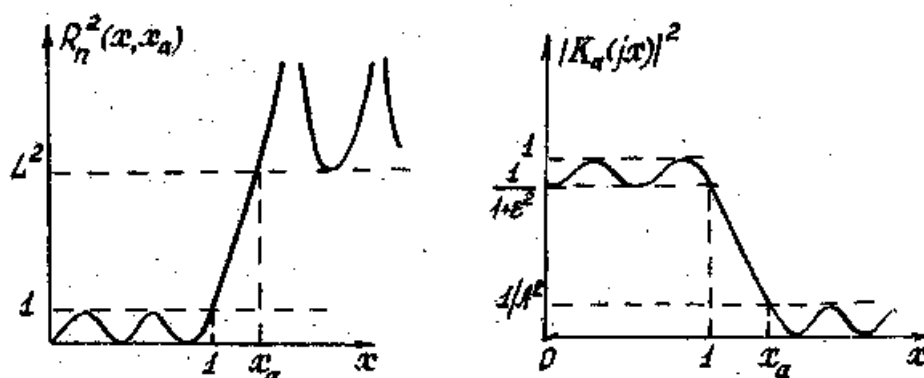


Рис. 4.31.

либо после объединения комплексно-сопряженных нулей и полюсов в форме

$$K_a(p) = \begin{cases} K_0 \prod_{k=1}^{n/2} \frac{p^2 + a_{0k}}{p^2 + b_{1k}p + b_{0k}}, & n - \text{четное}, \\ \frac{K_0}{p + \sigma_0} \prod_{k=1}^{(n-1)/2} \frac{p^2 + a_{0k}}{p^2 + b_{1k}p + b_{0k}}, & n - \text{нечетное}. \end{cases} \quad (4.45)$$

Расчетные соотношения для параметров функции (4.45) приведены в [6].

Обычно расчет параметров аналоговых фильтров рассмотренных типов выполняют с использованием специальных справочников.

Расчеты рассмотренные выше, касались ФНЧ с нормированной частотой среза $\tilde{\Omega}_p$. Обычно $\tilde{\Omega}_p = 1p/c$. Существует много различных способов преобразования такого фильтра в ФНЧ с другой частотой среза Ω_p , в фильтр верхних частот, полосно-пропускающий и полосно-заграждающий фильтры.

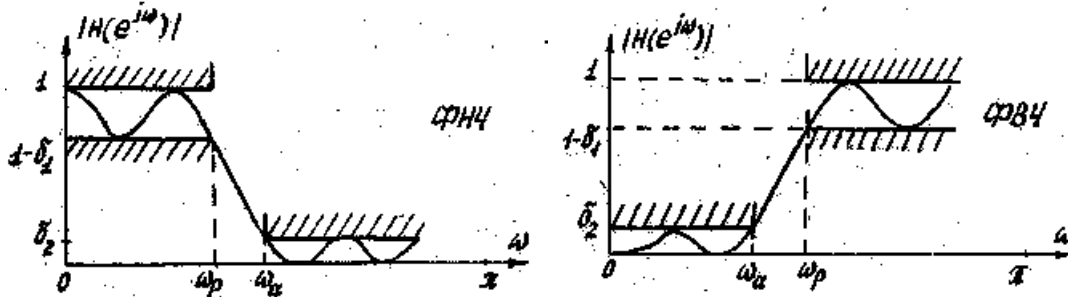
Наиболее простые преобразования, связанные с деформацией шкалы частот, получаются соответствующей заменой $p=f(p)$ [6]. Основные соотношения этих преобразований приведены в табл.4.1. Нелинейный характер этих преобразований приводит лишь к изменению взаимного расположения максимумов и минимумов пульсаций в преобразованной АЧХ, но не влияет на амплитуду этих пульсаций.

Таблица 4.1.

Требуемое преобразование	Формула замены	Значения параметров
ФНЧ $\rightarrow$ ФНЧ	$p \rightarrow \lambda p$	$\lambda = \tilde{\Omega}_p / \Omega_p$
ФНЧ $\rightarrow$ ФВЧ	$p \rightarrow \lambda/p$	$\lambda = \tilde{\Omega}_p \Omega_p$
ФНЧ $\rightarrow$ ФПП	$p \rightarrow \frac{1}{B} \left(p + \frac{\Omega_0^2}{p} \right)$	$\Omega_0^2 = \Omega_{p_1} \Omega_{p_2}$ $B = (\Omega_{p_2} - \Omega_{p_1}) / \tilde{\Omega}_p$
ФНЧ $\rightarrow$ ФПЗ	$p \rightarrow \frac{Bp}{p^2 + \Omega_0^2}$	$\Omega_0^2 = \Omega_{p_1} \Omega_{p_2}$ $B = (\Omega_{p_2} - \Omega_{p_1}) \tilde{\Omega}_p$
$\tilde{\Omega}_p$ - частота среза нормированного ФНЧ		
$\Omega_{p_1}, \Omega_{p_2}$ - нижняя и верхняя граничные частоты среза требуемого фильтра		

4.5.3. Расчет БИХ-фильтров стандартных типов методом обобщенного билинейного преобразования

Фильтры стандартных типов включают четыре варианта форм их частотных характеристик: фильтр нижних частот (ФНЧ), фильтр верхних частот (ФВЧ), фильтр полосно-пропускающий (ФПП), фильтр полосно-заграждающий (ФПЗ). Требования к ним, заданные в форме допусков на АЧХ, представлены на рис.4.32.



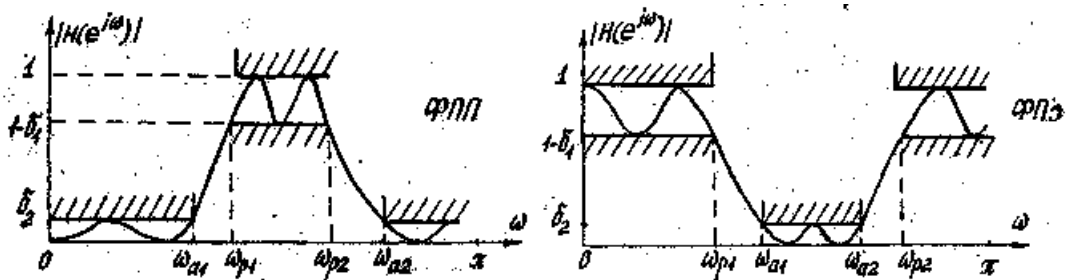


Рис.4.32. Требования к АЧХ фильтров стандартных типов.

Итак, к фильтру соответствующего типа предъявляются требования на максимальное значение неравномерности (пульсации) в полосе пропускания δ_1 и на максимальное отклонение от нуля в полосе задерживания δ_2 . Часто задаются требованиями не к АЧХ, а к характеристике затухания, выраженной в децибеллах:

$$A(\omega) = 20 \lg \frac{1}{|H(e^{j\omega})|} = -20 \lg |H(e^{j\omega})| (\text{дБ}). \text{ При этом задаются максимальное затухание в}$$

полосе пропускания $A_p = -20 \lg(1 - \delta_1) (\text{дБ})$ и минимальное затухание в полосе задерживания $A_a = -20 \lg \delta_2 (\text{дБ})$. Задаются также граничные частоты полос пропускания ω_{pi} и задерживания ω_{ai} и частота дискретизации $\Omega_0 = 2\pi/T$.

Порядок расчета цифрового фильтра может быть следующим.

1). Для граничных частот цифрового фильтра ω_{pi} и ω_{ai} определяются соответствующие граничные частоты аналогового прототипа Ω_{pi} , Ω_{ai} по формулам

$$\Omega_{pi} = \frac{2}{T} \operatorname{tg} \frac{\omega_{pi}}{2}, \quad \Omega_{ai} = \frac{2}{T} \operatorname{tg} \frac{\omega_{ai}}{2}.$$

2). Требования к АЧХ цифрового фильтра адресуются к АЧХ $|K_a(j\Omega)|$ аналогового прототипа в соответствующих полосах. По этим требованиям выбирается аналоговый фильтр того или иного типа (Баттерворта, Чебышева (типа 1 и 2) или эллиптический) и определяется его передаточная функция $K_a(p)$ как это было описано выше. Заметим, что обычно этот расчет выполняется с помощью справочников. (Христиан Э., Эйзенман Е. Таблицы и графики по расчету фильтров (справочник). М., Связь, 1975.)

3). Рассчитывается передаточная функция цифрового фильтра применением билинейного преобразования, т.е. в выражение $K_a(p)$ делается подстановка

$$p = \frac{2}{T} \cdot \frac{1 - z^{-1}}{1 + z^{-1}};$$

$$H(z) = K_a(p) \left| p = \frac{2}{T} \cdot \frac{1 - z^{-1}}{1 + z^{-1}} \right.$$

4). Рассчитываются коэффициенты, например, каскадной (или параллельной) формы реализации передаточной функции

$$H(z) = H_0 \prod_{i=1}^N \frac{1 + b_{1i}z^{-1} + b_{2i}z^{-2}}{1 + a_{1i}z^{-1} + b_{2i}z^{-2}}.$$

Поскольку в справочной литературе приводятся, как правило, расчетные сведения о передаточных функциях нормированных аналоговых ФНЧ (с частотой среза $\tilde{\Omega}_p = 1 \text{ p/c}$), поэтому для получения передаточной характеристики соответствующего аналогового прототипа, необходимо использовать одно из преобразований (табл.4.1), а затем билинейное преобразование (4.23), которое определит в конечном итоге

передаточную функцию требуемого цифрового фильтра. Этой процедуры двукратного расчета аналогового фильтра можно избежать с учетом следующих соображений.

1. Пусть мы проектируем цифровой ФНЧ с частотой среза ω_p . Если использовать билинейное преобразование в форме

$$p = \gamma \frac{1 - z^{-1}}{1 + z^{-1}},$$

где $\gamma = \text{const}$, то соотношение между аналоговыми Ω и цифровыми ω частотами примет вид $\Omega = \gamma \text{tg} \frac{\omega}{2}$. Выберем константу γ так, чтобы частота среза ω_p цифрового фильтра получилась в результате билинейного преобразования аналогового ФНЧ с частотой среза Ω_p , тогда

$$\Omega_p = \gamma \text{tg} \frac{\omega_p}{2}, \quad \gamma = \Omega_p \text{ctg} \frac{\omega_p}{2}.$$

Таким образом, для преобразования нормированного аналогового ФНЧ ($\Omega_p = 1\text{p/c}$) в цифровой ФНЧ с частотой среза ω_p достаточно использовать билинейное преобразование в форме

$$p = \text{ctg} \left(\frac{\omega_p}{2} \right) \cdot \frac{1 - z^{-1}}{1 + z^{-1}}. \quad (4.46)$$

При расчете аналогового ФНЧ граничная частота Ω_a полосы задерживания определится по граничной частоте ω_a цифрового фильтра соотношением

$$\Omega_a = \gamma \text{tg} \frac{\omega_a}{2}.$$

2. Для получения передаточной функции цифрового ФВЧ из нормированного аналогового ФНЧ с учетом промежуточной замены p на λ/p достаточно единого преобразования вида

$$p = \gamma \frac{1 + z^{-1}}{1 - z^{-1}},$$

для которого связь аналоговых и цифровых частот примет вид $\Omega = \gamma \text{ctg} \frac{\omega}{2}$. Константу γ определим из условия преобразования граничной частоты Ω_p аналогового ФНЧ в граничную частоту ω_p цифрового ФВЧ:

$$\Omega_p = \gamma \text{tg} \frac{\omega_p}{2}, \quad \gamma = \Omega_p \text{tg} \frac{\omega_p}{2}$$

При $\Omega_p = 1\text{p/c}$, $\gamma = \text{tg} \frac{\omega_p}{2}$, поэтому такой нормированный аналоговый ФНЧ преобразуется в цифровой ФВЧ с частотой среза ω_p при использовании одного преобразования

$$p = \text{tg} \left(\frac{\omega_p}{2} \right) \frac{1 + z^{-1}}{1 - z^{-1}}. \quad (4.47)$$

При расчете аналогового ФНЧ граничная частота Ω_a полосы задерживания

определится по граничной частоте ω_a цифрового фильтра соотношением

$$\Omega_a = \gamma \operatorname{ctg} \frac{\omega_a}{2}.$$

3. Аналогичные рассуждения приводят к преобразованию аналогового ФНЧ с частотой среза Ω_p в полосно-пропускающий цифровой фильтр с граничными частотами среза ω_{p1} и ω_{p2} при использовании подстановки

$$p = \gamma \frac{z^{-2} - 2\alpha z^{-1} + 1}{1 - z^{-2}}, \quad (4.48)$$

для которой соотношение между аналоговыми Ω и цифровыми ω частотами имеет вид

$$\Omega = \gamma \frac{\alpha - \cos \omega}{\sin \omega}. \quad (4.49)$$

Константы γ и α определяются из условий преобразования граничных частот:

$$-\Omega_p = \gamma \frac{\alpha - \cos \omega_{p1}}{\sin \omega_{p1}}, \quad \Omega_p = \gamma \frac{\alpha - \cos \omega_{p2}}{\sin \omega_{p2}},$$

из которых следует, что

$$\alpha = \frac{\cos[(\omega_{p1} + \omega_{p2})/2]}{\cos[(\omega_{p2} - \omega_{p1})/2]}, \quad \gamma = \Omega_p \operatorname{ctg}[(\omega_{p2} - \omega_{p1})/2]. \quad (4.50)$$

При расчете аналогового ФНЧ граничная частота Ω_a полосы задерживания определится по граничным частотам ω_{a1} и ω_{a2} цифрового фильтра соотношением

$$\Omega_a = \min(|\Omega_a^{(1)}|, |\Omega_a^{(2)}|), \text{ где } \Omega_a^{(1)} = \gamma \frac{\alpha - \cos \omega_{a1}}{\sin \omega_{a1}}, \quad \Omega_a^{(2)} = \gamma \frac{\alpha - \cos \omega_{a2}}{\sin \omega_{a2}}.$$

4. Наконец, для получения полосно-заграждающего цифрового фильтра с граничными частотами полос пропускания ω_{p1} и ω_{p2} из аналогового ФНЧ с частотой среза Ω_p следует использовать преобразование

$$p = \gamma \frac{1 - z^{-2}}{z^{-2} - 2\alpha z^{-1} + 1}, \quad (4.51)$$

для которого соотношение частот имеет вид

$$\Omega = \gamma \frac{\sin \omega}{\cos \omega - \alpha}. \quad (4.52)$$

Константы γ и α находятся из условий

$$\Omega_p = \gamma \frac{\sin \omega_{p1}}{\cos \omega_{p1} - \alpha}, \quad -\Omega_p = \gamma \frac{\sin \omega_{p2}}{\cos \omega_{p2} - \alpha},$$

откуда следуют соотношения

$$\alpha = \frac{\cos[(\omega_{p1} + \omega_{p2})/2]}{\cos[(\omega_{p2} - \omega_{p1})/2]}, \quad \gamma = \Omega_p \operatorname{tg}[(\omega_{p2} - \omega_{p1})/2], \quad (4.53)$$

При расчете аналогового ФНЧ граничная частота Ω_a полосы задерживания определится по граничным частотам ω_{a1} и ω_{a2} цифрового фильтра соотношением

$$\Omega_a = \min(|\Omega_a^{11}|, |\Omega_a^{11}|), \text{ где } \Omega_a^1 = \gamma \frac{\sin \omega_{a1}}{\alpha - \cos \omega_{a1}}, \quad \Omega_a^{11} = \gamma \frac{\sin \omega_{a2}}{\alpha - \cos \omega_{a2}}.$$

Соотношения (4.46)-(4.53), полученные выше, сведены в табл.4.2.

Таблица 4.2.

Требуемый цифровой фильтр	Граничные цифровые частоты	Формула замены р на	Значения параметров
ФНЧ	ω_p	$\gamma \frac{1-z^{-1}}{1+z^{-1}}$	$\gamma = \Omega_p \operatorname{ctg} \frac{\omega_p}{2}$
ФВЧ	ω_p	$\gamma \frac{1+z^{-1}}{1-z^{-1}}$	$\gamma = \Omega_p \operatorname{tg} \frac{\omega_p}{2}$
ФПП	ω_{p1}, ω_{p2}	$\gamma \frac{z^{-2} - 2\alpha z^{-1} + 1}{1 - z^{-2}}$	$\gamma = \Omega_p \operatorname{ctg} \left[(\omega_{p2} - \omega_{p1})/2 \right]$ $\alpha = \frac{\cos \left[(\omega_{p1} + \omega_{p2})/2 \right]}{\cos \left[(\omega_{p2} - \omega_{p1})/2 \right]}$
ФПЗ	ω_{p1}, ω_{p2}	$\gamma \frac{1 - z^{-2}}{z^{-2} - 2\alpha z^{-1} + 1}$	$\gamma = \Omega_p \operatorname{tg} \left[(\omega_{p2} - \omega_{p1})/2 \right]$ $\alpha = \frac{\cos \left[(\omega_{p1} + \omega_{p2})/2 \right]}{\cos \left[(\omega_{p2} - \omega_{p1})/2 \right]}$

4.5.4. Примеры расчета фильтров на основе билинейного преобразования

Рассмотрим примеры расчета цифровых фильтров на основе билинейного преобразования аналогового нормированного ФНЧ, для которого использовались различные аппроксимации, включая аппроксимации Баттерворта, Чебышева и эллиптическую. При одних и тех же требованиях к фильтру нижних частот для каждого метода аппроксимации приведем расчетные соотношения для коэффициентов передаточной функции фильтра, выражение для передаточной функции, графики АЧХ и характеристики затухания.

1). Аппроксимация Баттерворта. Параметры рассчитанного фильтра шестого порядка:

коэффициенты передаточной функции				
A1	A2	B1	B2	
-1.3143182	0.714895368	2	1	
-1.05406201	0.375318443	2	1	
-0.945920026	0.234217004	2	1	
H0 = 0.000579693109				

Передаточная функция фильтра:

$$H(z) = \frac{0,0005797 (1 + 2z^{-1} + z^{-2}) (1 + 2z^{-1} + z^{-2}) (1 + 2z^{-1} + z^{-2})}{(1 - 1,3143z^{-1} + 0,7149z^{-2}) (1 - 1,0541z^{-1} + 0,3753z^{-2}) (1 - 0,9459z^{-1} + 0,2342z^{-2})}$$

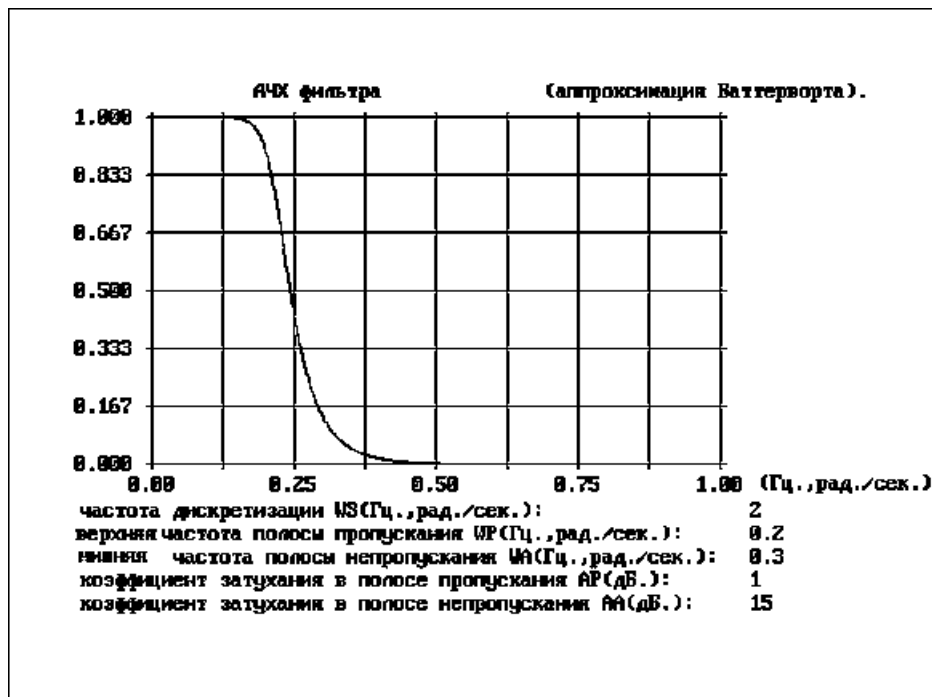


Рис.4.33а. АЧХ цифрового ФНЧ (аппроксимация Баттерворта).

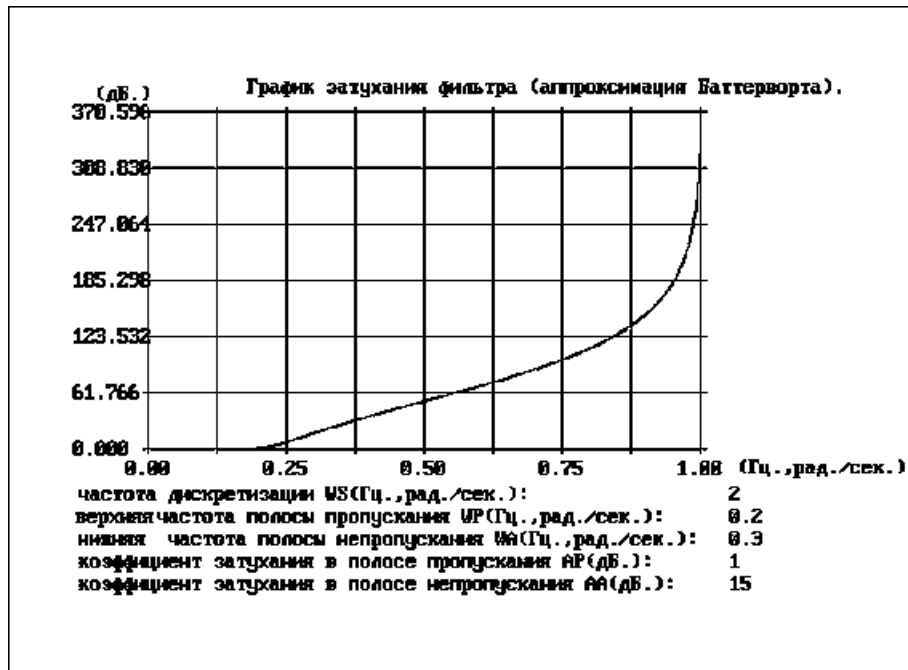


Рис.4.33б. Характеристика затухания цифрового ФНЧ (аппроксимация Баттерворта).

2). Аппроксимация Чебышева. Параметры рассчитанного фильтра четвертого порядка:

коэффициенты передаточной функции				
A1	A2	B1	B2	
-1.4995545	0.848218682	2	1	
-1.55478518	0.649295438	2	1	
$H0 = 0.00183555037$				

Передаточная функция фильтра:

$$H(z) = \frac{0,001836 (1 + 2z^{-1} + z^{-2}) (1 + 2z^{-1} + z^{-2})}{(1 - 1,4946z^{-1} + 0,8482z^{-2}) (1 - 1,5548z^{-1} + 0,6493z^{-2})}$$

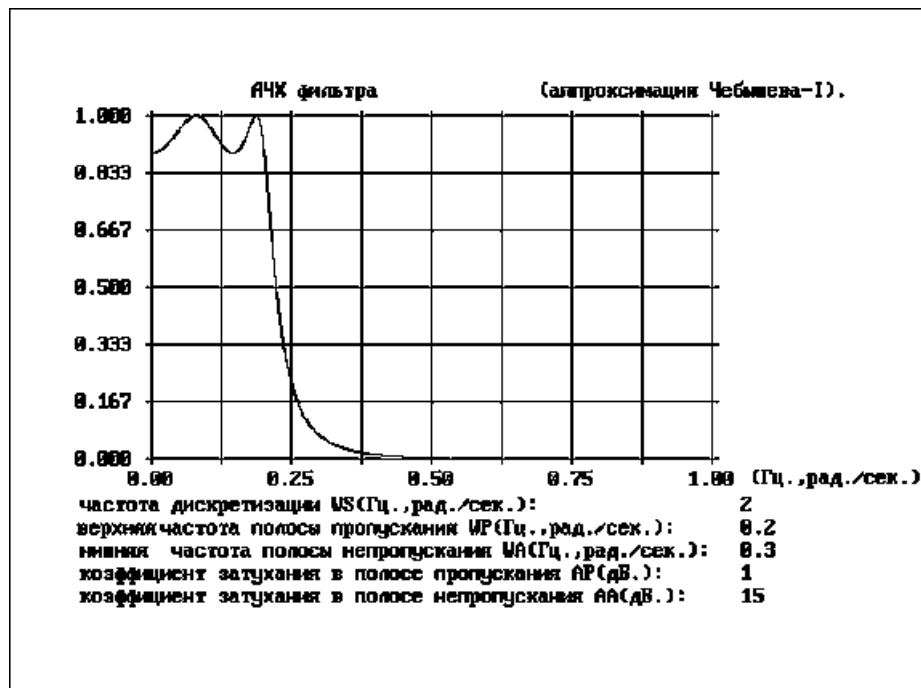


Рис.4.34а. АЧХ цифрового ФНЧ (аппроксимация Чебышева).

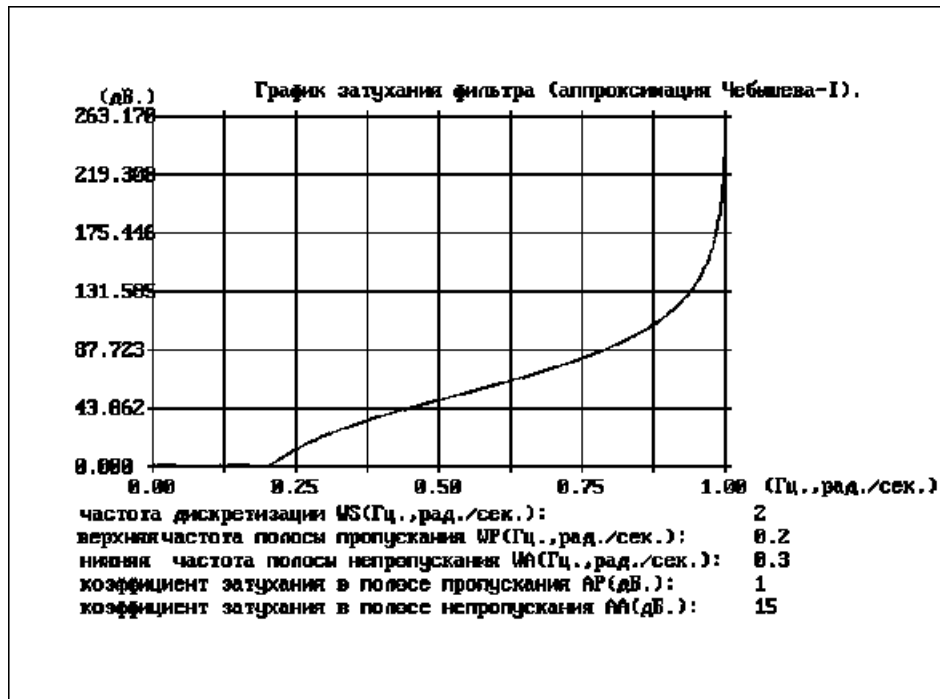


Рис.4.34б. Характеристика затухания цифрового ФНЧ (аппроксимация Чебышева).

3). Аппроксимация эллиптическая. Параметры рассчитанного фильтра третьего порядка:

коэффициенты передаточной функции			
A1	A2	B1	B2
-1.44609085	0.795643438	-1.01660573	1
-0.682944373	0	1	0
N0 = 0.0563495328			

Передаточная функция фильтра:

$$H(z) = \frac{0,05635(1 - 1,0166z^{-1} + z^{-2})(1 + z^{-1})}{(1 - 1,4461z^{-1} + 0,7956z^{-2})(1 - 0,6829z^{-1})}$$

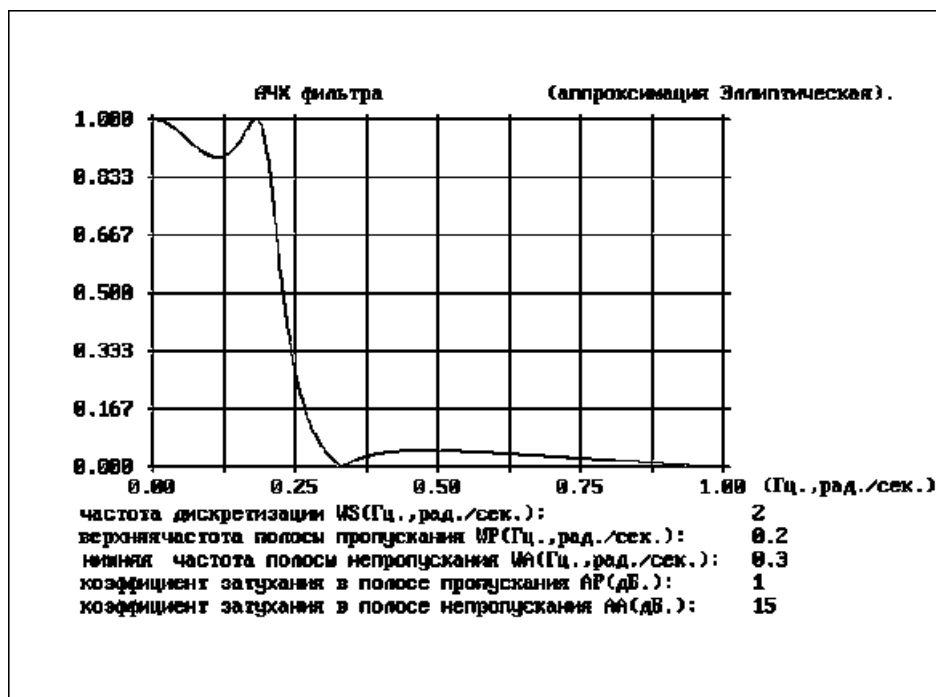


Рис.4.35а. АЧХ цифрового ФНЧ (аппроксимация эллиптическая).

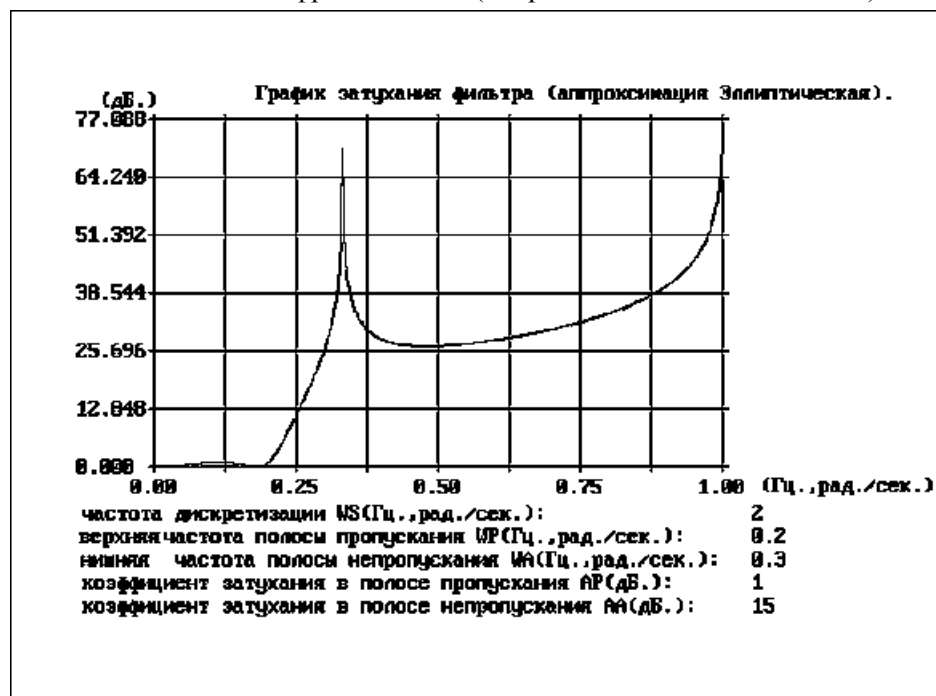


Рис.4.35б. Характеристика затухания цифрового ФНЧ (аппроксимация эллиптическая).

Приведем результаты расчета цифрового фильтра верхних частот на основе билинейного преобразования аналогового нормированного ФНЧ, для которого использовалась аппроксимация Чебышева типа 1.

Параметры рассчитанного фильтра четвертого порядка:

коэффициенты передаточной функции				
A1	A2	B1	B2	
0.556147453	0.76471379	-2	1	
1.04156854	0.401948754	-2	1	
H0 = 0.0242611537				

Передаточная функция фильтра:

$$H(z) = \frac{0,02426(1 - 2z^{-1} + z^{-2})(1 - 2z^{-1} + z^{-2})}{(1 + 0,5561z^{-1} + 0,7647z^{-2})(1 + 1,0416z^{-1} + 0,4019z^{-2})}$$

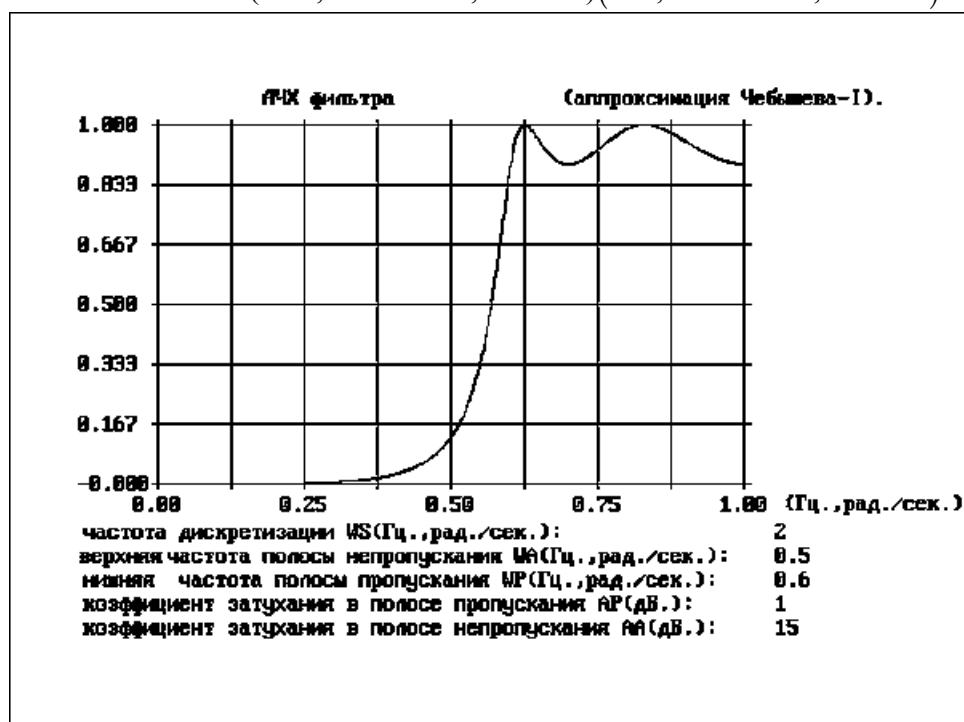


Рис.4.36а. АЧХ цифрового ФВЧ (аппроксимация Чебышева).

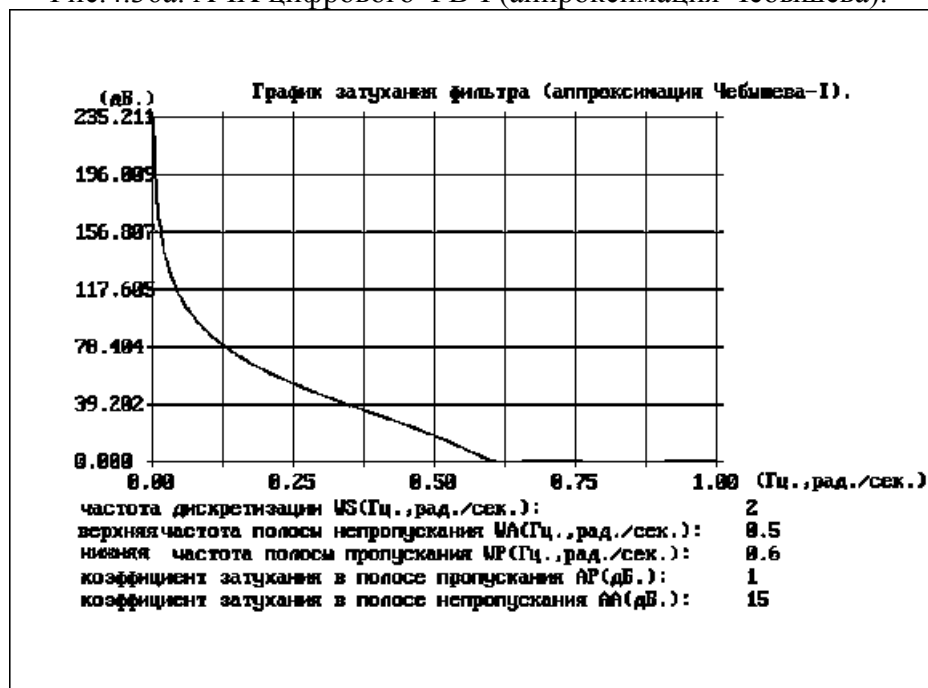


Рис.4.36б. Характеристика затухания цифрового ФВЧ (аппроксимация Чебышева).

4.5.5. Машинное проектирование цифровых БИХ-фильтров

В общем случае не существует аналитических процедур для расчета аналоговых или цифровых фильтров, удовлетворяющих требованиям произвольной частотной характеристики или другим типам требований. В этих случаях применяются методы оптимизации, которые основаны на использовании ЦВМ для решения систем линейных или нелинейных уравнений для коэффициентов фильтра.

Рассмотрим одну из процедур машинного расчета цифровых БИХ-фильтров, предложенную Штейнглицем и основанную на минимизации среднеквадратической ошибки в частотной области [2,3,4]. При этой процедуре необходимо, чтобы требуемая

частотная характеристика $H_d(e^{j\omega})$ была определена для дискретного ряда частот $\{\omega_j\}$, $j=1,2,\dots,M$ в диапазоне $0 \leq \omega \leq \pi$. Среднеквадратическая ошибка на этих частотах определяется как

$$E = \sum_{j=1}^M \left[\left| H(e^{j\omega_j}) - H_d(e^{j\omega_j}) \right|^2 \right].$$

Предполагается, что передаточная функция фильтра имеет вид

$$H(z) = A \prod_{k=1}^K \frac{1 + a_k z^{-1} + b_k z^{-2}}{1 + c_k z^{-1} + d_k z^{-2}}$$

соответствующий каскадной форме, поскольку она менее чувствительна к точности коэффициентов и для удобства в расчете производных, необходимых в процедуре оптимизации.

Обозначим символом $\theta = \{a_k, b_k, c_k, d_k\}$ совокупность $4K$ неизвестных коэффициентов, $H_j = H(e^{j\omega_j}) / A$, $H_{dj} = |H_d(e^{j\omega_j})|$ и представим ошибку в виде:

$$E(\theta, A) = \sum_{j=1}^M \left[A |H_j| - H_{dj} \right]^2.$$

Поскольку необходимо найти значения этих параметров, минимизирующих ошибку E , то нужно взять частные производные от E по каждому параметру θ и A и приравнять эти производные к нулю. Таким образом, получается $(4K+1)$ уравнений с $(4K+1)$ неизвестными.

Уравнение для A оказывается особенно простым, поскольку

$$\frac{\partial E}{\partial A} = \sum_{j=1}^M 2A^* [A^* |H_j| - H_{dj}] |H_j| = 0, \quad A^* = \frac{\sum_{j=1}^M |H_j| H_{dj}}{\sum_{j=1}^M |H_j|^2}.$$

Оптимальные значения θ^* , минимизирующие ошибку E , найдутся из решения системы нелинейных уравнений:

$$\frac{\partial E(\theta, A^*)}{\partial \theta_k} = 2A^* \sum_{j=1}^M [A^* |H_j| - H_{dj}] \frac{\partial |H_j|}{\partial \theta_k} = 0, \quad k = 1, 2, \dots, 4K,$$

где θ_k – k -я компонента θ , соответствующая либо a_k , либо b_k , либо c_k , либо d_k .

Эти уравнения можно решить с помощью методов оптимизации, используя, например, алгоритм Флетчера-Пауэла [4]. Отметим, что при этой процедуре учитывается только амплитудная характеристика. Алгоритм оптимизации может в итоге дать значения параметров, соответствующих неустойчивому фильтру, т.е. полюсы и нули каждого блока второго порядка окажутся за пределами единичного круга. Кроме ограничений на расположение нулей и полюсов, Штейглиц проверял корни каждого множителя второго порядка после завершения процедуры минимизации, и, если полюс (или нуль) оказывался вне единичного круга, он заменял его на зеркальное отображение, оставляя таким образом амплитудно-частотную характеристику неизменной. Например, при наличии действительного полюса $z=p_1$, он заменяется на полюс $z=1/p_1$, что эквивалентно умножению передаточной функции на множитель $(z - p_1)/(z - 1/p_1)$, который при $|z|=1$ имеет постоянную величину равную p_1 . Это означает, что такая трансформация полюса не изменяет формы АЧХ. Штейглиц нашел, что продолжение процедуры оптимизации иногда приводит к дальнейшему уменьшению ошибки.

Ниже приведены результаты расчетов на основе рассмотренной процедуры, когда идеальный фильтр нижних частот определялся частотой среза равной 0.1, а число каскадов второго порядка изменялось от $K=1$ до $K=3$. На рис.4.37 – 4.39 представлены

графики АЧХ рассчитанных цифровых ФНЧ.

1. Число каскадов K=1. Параметры рассчитанного фильтра:

коэффициент затухания в полосе пропускания: 2.89009737			
коэффициент затухания в полосе задерживания: -9.21151763			
H0 = 0.0782083834			
коэффициенты передаточной функции			
A1	A2	B1	B2
-1.54822209	0.707216933	-0.218597931	0.998194379

Передаточная функция фильтра:

$$H(z) = \frac{0.07821(1 - 0.2186z^{-1} + 0.9982z^{-2})}{(1 - 1.5482z^{-1} + 0.7072z^{-2})}.$$

номер итерации:

64

значение оптимизируемой функции: 1.04407138

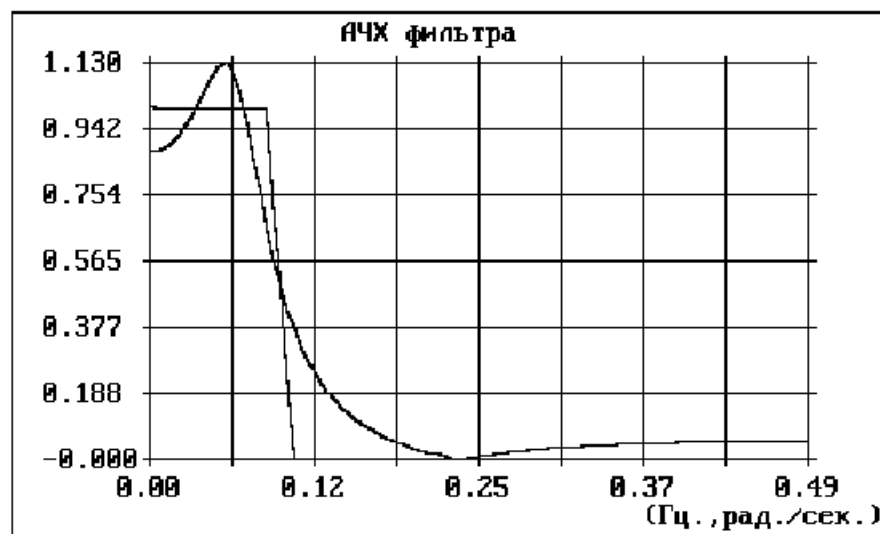


Рис.4.37. АЧХ однокаскадного фильтра.

2. Число каскадов K= 2. Параметры рассчитанного фильтра:

коэффициент затухания в полосе пропускания: 0.50143646			
коэффициент затухания в полосе задерживания: -19.9277534			
H0 = 0.0252692647			
коэффициенты передаточной функции			
A1	A2	B1	B2
-1.48245652	0.600753211	-1.42975046	0.999943444
-1.59515438	0.894142518	0.40645954	0.984880758

Передаточная функция фильтра:

$$H(z) = \frac{0.02527(1 - 1.4297z^{-1} + 0.9999z^{-2})(1 + 0.4065z^{-1} + 0.9849z^{-2})}{(1 - 1.5952z^{-1} + 0.8941z^{-2})(1 - 1.4825z^{-1} + 0.6007z^{-2})}.$$

номер итерации: 276
 значение оптимизируемой функции: 0.0443938605

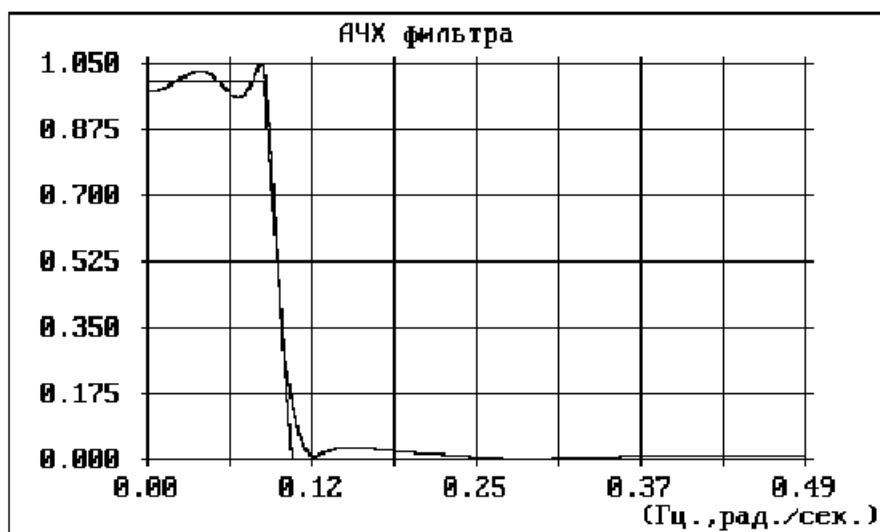


Рис.4.38. АЧХ двухкаскадного фильтра.

3. Число каскадов K= 3. Параметры рассчитанного фильтра:

коэффициент затухания в полосе пропускания: 0.107626967
 коэффициент затухания в полосе задерживания: -31.7442555
 H0 = 0.00843185725

коэффициенты передаточной функции			
A1	A2	B1	B2
-1.52390702	0.742620345	0.736747261	0.989142137
-1.62005669	0.934697906	-1.40489400	0.991828949
-1.39633742	0.512312106	-1.40200775	0.993191902

Передаточная функция фильтра:

$$H(z) = \frac{0.008432(1 + 0.7367z^{-1} + 0.9891z^{-2})(1 - 1.4049z^{-1} + 0.9918z^{-2})(1 - 1.402z^{-1} + 0.9932z^{-2})}{(1 - 1.5239z^{-1} + 0.7426z^{-2})(1 - 1.6201z^{-1} + 0.9347z^{-2})(1 - 1.3963z^{-1} + 0.5123z^{-2})}$$

номер итерации:

369

значение оптимизируемой функции: 0.00219643739

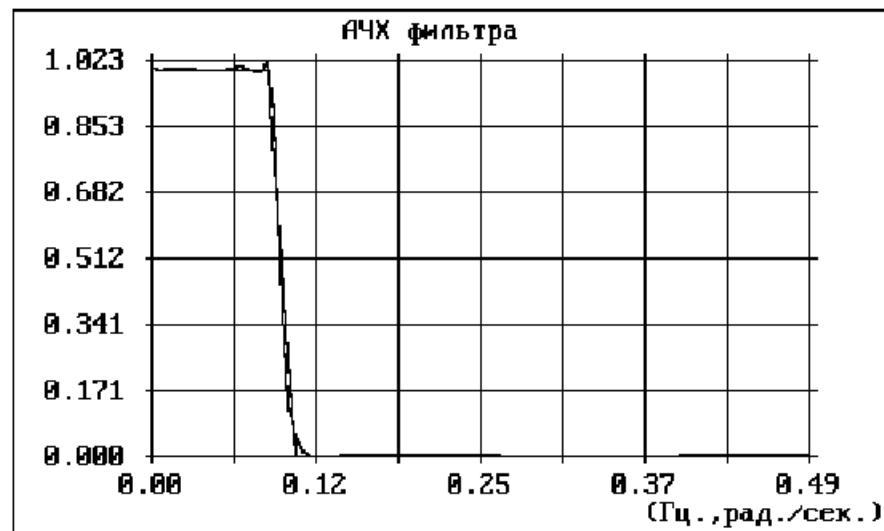


Рис.4.39. АЧХ трехкаскадного фильтра.

4.5.6. Свойства цифровых КИХ-фильтров

Как было показано выше, класс последовательностей конечной длины обладает некоторыми свойствами, желательными с точки зрения построения фильтров. Например, никогда не возникает вопрос об устойчивости и физической реализуемости фильтров, поскольку КИХ-последовательности гарантируют устойчивость, а при введении соответствующей конечной задержки и реализуемость. Более того, ниже будет показано, что КИХ-последовательности можно выбрать так, чтобы фильтры имели строго *линейные фазовые характеристики*. Поэтому, используя КИХ-последовательности, можно проектировать фильтры с произвольной амплитудной характеристикой.

Интересно отметить, что до появления алгоритма быстрого преобразования Фурье (БПФ) реализация КИХ-фильтров считалась, как правило, нереальной, поскольку для достаточно хорошей аппроксимации фильтров с острыми срезами требуются весьма длинные последовательности. Разработка на основе высокоэффективного алгоритма БПФ методов быстрой свертки изменила это положение, и в настоящее время КИХ-фильтры успешно конкурируют с БИХ-фильтрами, имеющими острые срезы в частотной характеристике.

Перечислим основные достоинства КИХ-фильтров:

1. Легко создавать КИХ-фильтры со строго линейной фазовой характеристикой. Во многих случаях, когда проектируется фильтр с произвольной амплитудной характеристикой, это упрощает задачу аппроксимации. Фильтры с линейной фазовой характеристикой особенно важны в случаях, когда приходится учитывать дисперсионные искажения, связанные с нелинейностью фазовой характеристики (например, при обработке речи и передаче данных).

2. КИХ-фильтры можно эффективно строить как по рекурсивной, так и по нерекурсивной схемам.

3. КИХ-фильтры, реализуемые нерекурсивно, т. е. с помощью прямой свертки, всегда устойчивы.

Основной недостаток КИХ-фильтров состоит в том, что для аппроксимации фильтров, частотные характеристики которых имеют острые срезы, требуется импульсная характеристика с большим числом отсчетов N . Поэтому при использовании обычной свертки необходимо выполнять большой объем вычислений.

Передаточная функция физически реализуемого КИХ-фильтра имеет вид

$$H(z) = \sum_{n=0}^{N-1} h(n)z^{-n}.$$

Это означает, что если длина импульсной характеристики равна N отсчетам, то $H(z)$ является полиномом по z^{-1} степени $N-1$. Поэтому $H(z)$ имеет $N-1$ полюсов в точке $z=0$ и $N-1$ нулей, которые могут быть в любом месте z -плоскости.

Частотная характеристика представляет собой тригонометрический полином

$$H(e^{j\omega}) = \sum_{n=0}^{N-1} h(n)e^{j\omega n} = |H(e^{j\omega})|e^{j\varphi(\omega)}.$$

С фазо-частотной характеристикой фильтра $\varphi(\omega)$ связаны такие понятия как фазовая τ_p и групповая τ_g задержки, определяемые соотношениями:

$$\tau_p(\omega) = -\frac{\varphi(\omega)}{\omega} \quad \text{и} \quad \tau_g(\omega) = -\frac{\partial \varphi(\omega)}{\partial \omega}.$$

Для обеспечения постоянства фазовой τ_p и групповой τ_g задержек ФЧХ фильтра должна быть линейной функцией частоты ω , т.е. $\varphi(\omega) = -\tau\omega$. Во многих приложениях требуется постоянство только групповой задержки и, следовательно, ФЧХ может иметь вид $\varphi(\omega) = \varphi_0 - \tau\omega$, где φ_0 – постоянная величина.

Если импульсная характеристика удовлетворяет условию симметрии или антисимметрии, т.е.

$$h(n) = \pm h(N-1-n), \quad n = 0, 1, 2, \dots, N, \quad (4.54)$$

то КИХ-фильтр имеет линейную ФЧХ. Для центрированной импульсной характеристики $h^0(n)$, удовлетворяющей соотношению $h^0(n) = h(n + (N-1)/2)$ или $h(n) = h^0(n - (N-1)/2)$, условие линейности ФЧХ фильтра можно записать в виде:

$$h^0(n) = \pm h^0(-n), \quad (4.55)$$

т.е. необходима четная или нечетная симметрия центрированной импульсной характеристики $h^0(n)$. Целесообразно рассмотреть условия (4.54) отдельно для четного и нечетного N . Типичные импульсные характеристики фильтра с линейной ФЧХ представлены на рис.4.40 ($\alpha = (N-1)/2$): а) нечетное $N=11$ (четная симметрия); б) четное $N=10$ (четная симметрия); в) нечетное $N=11$ (нечетная симметрия); г) четное $N=10$ (нечетная симметрия).

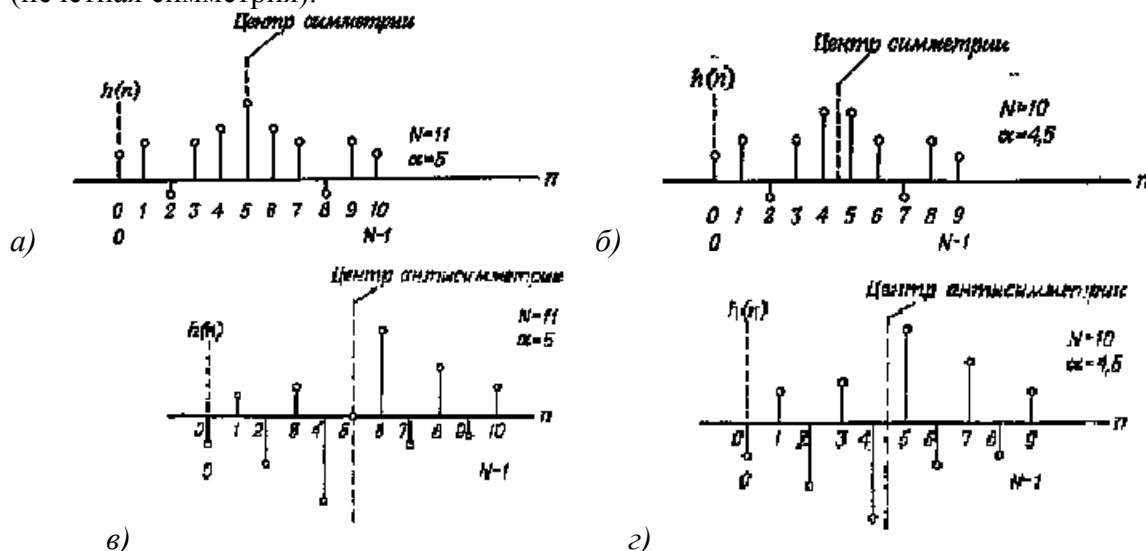


Рис.4.40. Типичные импульсные характеристики: а) нечетное N (четная симметрия); б) четное N (четная симметрия); в) нечетное N (нечетная симметрия); г) четное N (нечетная симметрия).

Достаточность этого условия следует из соотношения для частотной характеристики:

$$H(e^{j\omega}) = \sum_{n=0}^{N-1} h^0\left(n - \frac{N-1}{2}\right) e^{-j\omega n} = e^{-j\omega\left(\frac{N-1}{2}\right)} \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} h^0(n) e^{-j\omega n} = e^{-j\omega\left(\frac{N-1}{2}\right)} H^0(e^{j\omega}),$$

$$\text{где } H^0(e^{j\omega}) = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} h^0(n) e^{-j\omega n}.$$

(4.56)

Покажем, что если $h^0(n)$ обладает четной симметрией, то $H^0(e^{j\omega})$ - действительная функция, а при нечетной симметрии $h^0(n)$ - $H^0(e^{j\omega})$ - мнимая функция.

Понятно, что в обоих случаях ФЧХ будет линейной: в первом - $\varphi(\omega) = -\frac{N-1}{2}\omega$, во

втором - $\varphi(\omega) = \pm \frac{\pi}{2} - \frac{N-1}{2}\omega$.

Фильтр вида 1. Симметричная импульсная характеристика, нечетное N . Для этого случая $h^0(n) = h^0(-n)$, тогда

$$\begin{aligned} H^0(e^{j\omega}) &= \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} h^0(n) e^{-j\omega n} = h^0(0) + 2 \sum_{n=1}^{\frac{N-1}{2}} h^0(n) \frac{e^{j\omega n} + e^{-j\omega n}}{2} = \\ &= h^0(0) + 2 \sum_{n=1}^{\frac{N-1}{2}} h^0(n) \cos(\omega n) = \sum_{n=0}^{\frac{N-1}{2}} a(n) \cos(\omega n), \end{aligned} \quad (4.57)$$

$$\text{где } a(0) = h^0(0) = h\left(\frac{N-1}{2}\right), \quad a(n) = 2h^0(n) = 2h\left(n + \frac{N-1}{2}\right), \quad n = 1, 2, \dots, \frac{N-1}{2}.$$

Таким образом, $H^0(e^{j\omega})$ - действительная и четная функция частоты ω .

Фильтр вида 2. Симметричная импульсная характеристика, четное N . Для этого случая $h^0(n) = h^0(-n)$ или $h(n) = h(N-1-n)$, $n = 0, 1, \dots, (N/2)-1$, тогда

$$H^0(e^{j\omega}) = \sum_{n=-\frac{N}{2}}^{\frac{N}{2}-1} h^0(n) e^{-j\omega n} = 2 \sum_{n=0}^{\frac{N}{2}-1} h(n) \cos\left[\omega\left(\frac{N-1}{2} - n\right)\right] = \sum_{n=1}^{\frac{N}{2}} b(n) \cos\left[\omega\left(n - \frac{1}{2}\right)\right], \quad (4.58)$$

$$\text{где } b(n) = 2h\left(\frac{N}{2} - n\right), \quad n = 1, 2, \dots, \frac{N}{2}.$$

Таким образом, и в этом случае $H^0(e^{j\omega})$ - действительная и четная функция частоты ω .

Необходимо отметить, что всегда для этой функции $H^0(e^{j\pi}) = 0$. Это означает, что нельзя использовать фильтры этого вида для аппроксимации частотной характеристики, отличной от нуля при $\omega = \pi$ (например, при проектировании фильтров верхних частот).

Фильтр вида 3. Антисимметричная импульсная характеристика, нечетное N . Для этого случая $h^0(n) = -h^0(-n)$, $h^0(0) = 0$, тогда

$$\begin{aligned}
 H^0(e^{j\omega}) &= \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} h^0(n) e^{-j\omega n} = 2j \sum_{n=1}^{\frac{N-1}{2}} h^0(n) \frac{e^{j\omega n} + e^{-j\omega n}}{2j} = \\
 &= 2j \sum_{n=1}^{\frac{N-1}{2}} h^0(n) \sin(\omega n) = j \sum_{n=1}^{\frac{N-1}{2}} c(n) \sin(\omega n) = e^{j\frac{\pi}{2}} H^*(e^{j\omega}),
 \end{aligned} \quad (4.59)$$

$$\text{где } c(n) = 2h^0(n) = 2h\left(n + \frac{N-1}{2}\right), \quad n=1, 2, \dots, \frac{N-1}{2}, \quad H^*(e^{j\omega}) = \sum_{n=1}^{\frac{N-1}{2}} c(n) \sin(\omega n).$$

Таким образом, в этом случае $H^0(e^{j\omega})$ - чисто мнимая функция частоты ω . Кроме того, всегда $H^0(e^{j0}) = H^0(e^{j\pi}) = 0$. Поэтому этот вид фильтров наиболее пригоден для проектирования преобразователей Гильберта и дифференциаторов.

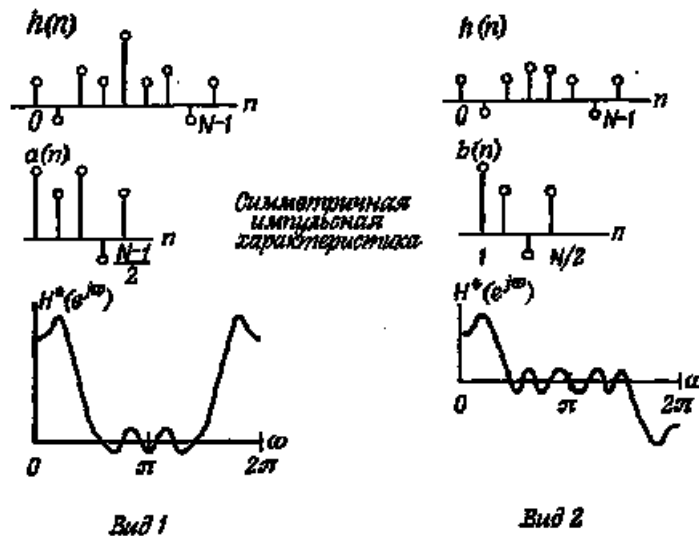
Фильтр вида 4. Антисимметричная импульсная характеристика, четное N . Для этого случая $h^0(n) = -h^0(-n)$, или $h(n) = -h(N-1-n)$, $n=0, 1, \dots, (N/2)-1$, тогда

$$\begin{aligned}
 H^0(e^{j\omega}) &= \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} h^0(n) e^{-j\omega n} = 2j \sum_{n=0}^{\frac{N}{2}-1} h(n) \sin\left[\omega\left(\frac{N-1}{2} - n\right)\right] = \\
 &= j \sum_{n=1}^{\frac{N}{2}} d(n) \sin\left[\omega\left(n - \frac{1}{2}\right)\right] = e^{j\frac{\pi}{2}} H^*(e^{j\omega}),
 \end{aligned}$$

$$\text{где } d(n) = 2h\left(\frac{N}{2} - n\right), \quad n=1, 2, \dots, \frac{N}{2}, \quad H^*(e^{j\omega}) = \sum_{n=1}^{\frac{N}{2}} d(n) \sin\left[\omega\left(n - \frac{1}{2}\right)\right]. \quad (4.60)$$

Таким образом, и в этом случае $H^0(e^{j\omega})$ - чисто мнимая функция частоты ω . Кроме того, всегда $H^0(e^{j0}) = 0$. Поэтому этот вид фильтров наиболее пригоден для проектирования преобразователей Гильберта и дифференциаторов.

На рис.4.41 графически представлены типичные импульсные характеристики $h(n)$, соответствующие им последовательности (от $a(n)$ до $d(n)$ для каждого конкретного случая) и типичные частотные характеристики $H^*(e^{j\omega})$ для каждого из четырех видов КИХ-фильтров с линейной ФЧХ.



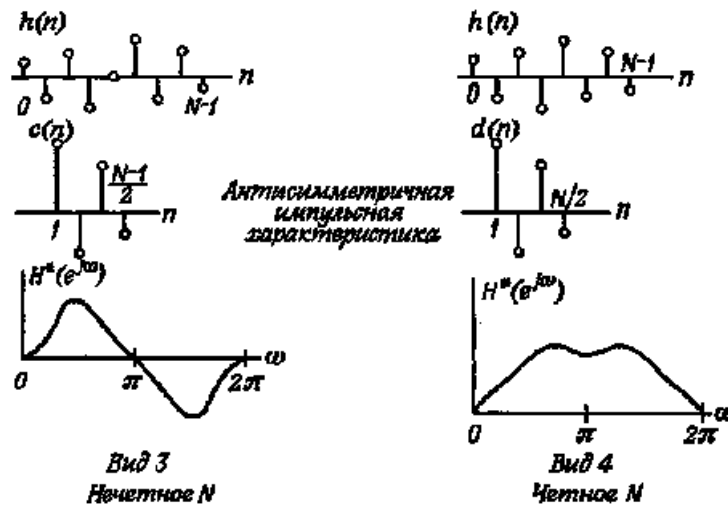


Рис.4.41. Четыре вида фильтров с линейной ФЧХ.

4.5.7. Расчет КИХ-фильтров с линейной ФЧХ по методу взвешивания (метод «окна»)

Известны три класса методов расчета КИХ-фильтров с линейной фазой: методы взвешивания с помощью окна, методы частотной выборки, а также методы расчета оптимальных (по Чебышеву) фильтров. Трудно рекомендовать какой-либо один из этих методов, поскольку в каждом конкретном случае выбор метода расчета определяется большим числом факторов. Поэтому ниже будут рассмотрены преимущества и недостатки каждого из этих методов с тем, чтобы проектировщик сам смог решить, какой метод лучше всего использовать, чтобы удовлетворить заданным требованиям.

Начнем рассмотрение с метода взвешивания (метода окна), суть которого сводится к получению импульсной характеристики конечной длины путем усечения последовательности импульсной характеристики бесконечной длины. Поскольку частотная характеристика требуемого идеального цифрового фильтра $H_d(e^{j\omega})$ является периодической функцией частоты, ее можно представить рядом Фурье:

$$H_d(e^{j\omega}) = \sum_{n=-\infty}^{\infty} h_d(n) e^{-j\omega n}, \quad (4.61)$$

где $h_d(n)$ - соответствующая последовательность отсчетов импульсной характеристики, т.е.

$$h_d(n) = (1/2\pi) \int_{-\pi}^{\pi} H_d(e^{j\omega}) e^{j\omega n} d\omega, \quad (4.62)$$

Использование соотношений (4.61) и (4.62) для проектирования КИХ-фильтров связано с двумя трудностями. Во-первых, импульсная характеристика фильтра в общем случае имеет бесконечную длину, поскольку суммирование в (4.61) производится в бесконечных пределах. Во-вторых, фильтр физически нереализуем, так как импульсная характеристика начинается в $-\infty$, т. е. никакая конечная задержка не сделает фильтр физически реализуемым. Итак, фильтр, рассчитываемый на основе представления функции $H_d(e^{j\omega})$ рядом Фурье, оказывается физически нереализуемым БИХ-фильтром.

Один из возможных методов получения КИХ-фильтра, аппроксимирующего заданную функцию $H_d(e^{j\omega})$, заключается в усечении бесконечного ряда Фурье (4.61) за пределами $n = \pm(N-1/2)$. Однако простое усечение ряда приводит к хорошо известному явлению Гиббса, которое проявляется в виде выбросов и пульсаций определенного уровня до и после точки разрыва в аппроксимируемой частотной характеристике. Так, например, при аппроксимации стандартных фильтров типа идеального фильтра нижних частот или полосового фильтра максимальная амплитуда пульсаций частотной характеристики составляет около 9% и не уменьшается с увеличением длины

импульсной характеристики, т. е. учет все большего числа членов ряда Фурье не приводит к уменьшению максимальной амплитуды пульсаций. Вместо этого по мере увеличения N уменьшается ширина выброса.

Лучшие результаты дает метод проектирования КИХ-фильтров, основанный на использовании весовой последовательности конечной длины $w(n)$, называемой окном, для модификации коэффициентов Фурье $h_d(n)$ в формуле (4.61) с тем, чтобы управлять сходимостью ряда Фурье. Метод взвешивания иллюстрируется на рис.4.42. Сверху показаны заданная периодическая частотная характеристика $H_d(e^{j\omega})$, и ее коэффициенты Фурье $h_d(n)$. Ниже изображена весовая последовательность конечной длины $w(n)$ и ее преобразование Фурье $W(e^{j\omega})$. Для большинства приемлемых окон функция $W(e^{j\omega})$ имеет главный лепесток, содержащий почти всю энергию окна, и боковые лепестки, которые обычно быстро затухают. Чтобы получить КИХ-аппроксимацию функции $H_d(e^{j\omega})$, формируется последовательность $h(n) = h_d(n)w(n)$, в точности равная нулю за пределами интервала $-(N-1)/2 \leq n \leq (N-1)/2$. Третья пара графиков на рис.4.42 представляет последовательность $h(n)$ и ее преобразование Фурье $H(e^{j\omega})$, равное, очевидно, круговой свертке функций $H_d(e^{j\omega})$ и $W(e^{j\omega})$, поскольку $h(n)$ является произведением $h_d(n)w(n)$:

$$H(e^{j\omega}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_d(e^{j\theta}) W(e^{j(\omega-\theta)}) d\theta. \quad (4.63)$$

Наконец, внизу на рис.4.42 приведена физически реализуемая последовательность $h'(n) = h(n - (N-1)/2)$, которая равна задержанной последовательности $h(n)$ и может быть использована в качестве искомой импульсной характеристики фильтра. При этом частотная характеристика физически реализуемого фильтра будет равна $H(e^{j\omega})e^{jj\omega(N-1)/2}$ и при соблюдении условий симметрии импульсной характеристики: $h'(n) = \pm h'(N-1-n)$ будет иметь линейную ФЧХ.

На примере, представленном на рис.4.42, можно проследить влияние операции взвешивания коэффициентов Фурье фильтра на его частотную характеристику.

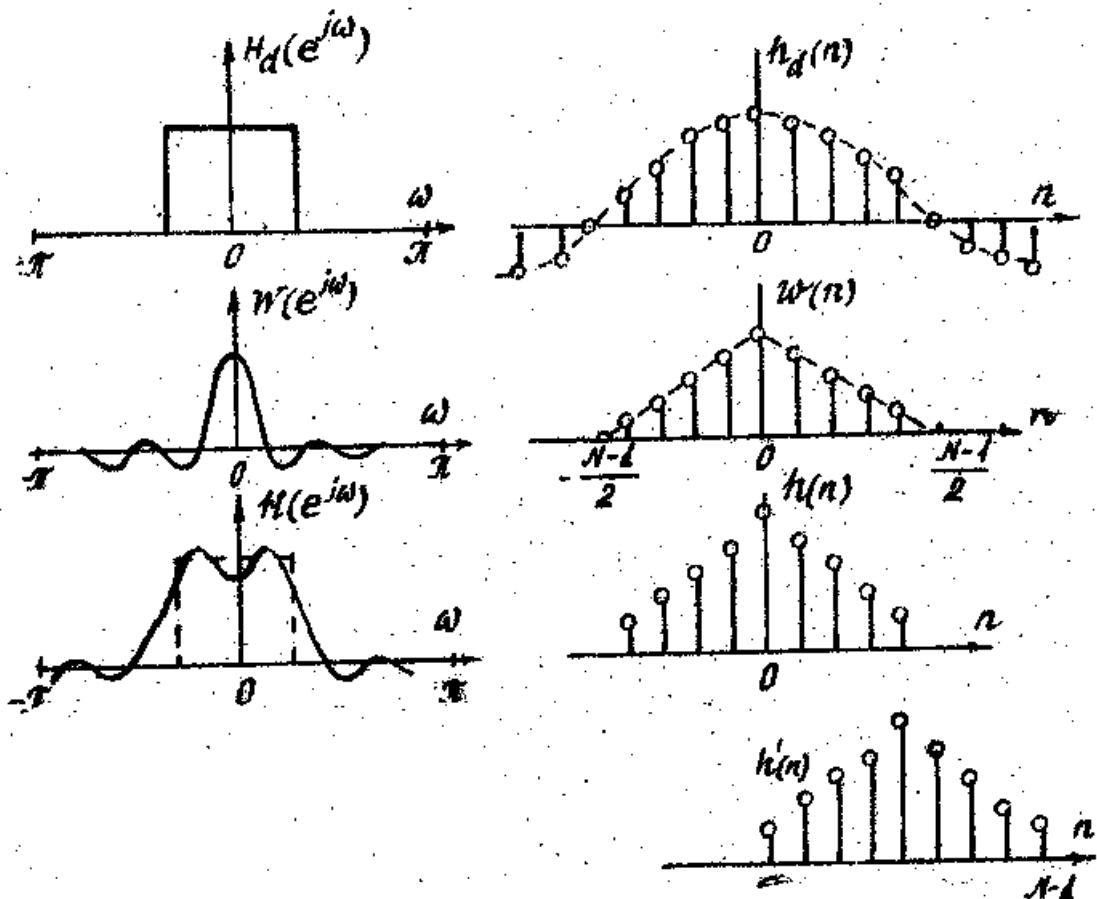


Рис.4.42. Иллюстрация метода взвешивания.

Прежде всего по обе стороны от точек разрыва заданной функции $H_d(e^{j\omega})$ появляются переходные полосы. Ясно, что поскольку результирующая частотная характеристика фильтра равна круговой свертке идеальной частотной характеристики и частотной характеристики окна, то ширина переходных полос зависит от ширины главного лепестка функции $W(e^{j\omega})$. Кроме того, на всех частотах ω возникают ошибки аппроксимации, имеющие вид пульсаций частотной характеристики, которые обусловлены боковыми лепестками функции $W(e^{j\omega})$. Ясно, наконец, и то, что получаемые фильтры ни в каком смысле не являются оптимальными (даже если окна и удовлетворяют тому или иному критерию оптимальности), поскольку их частотные характеристики рассчитываются через свертку.

После общего рассмотрения метода взвешивания возникают два вопроса: какими свойствами должны обладать окна и насколько точно они могут быть реализованы на практике? Ответ на первый вопрос относительно прост. Желательно, чтобы окно обладало следующими свойствами:

1. Ширина главного лепестка частотной характеристики окна, содержащего по возможности большую часть общей энергии, должна быть малой.
2. Энергия в боковых лепестках частотной характеристики окна должна быстро уменьшаться при приближении ω к π .

Было предложено много окон, аппроксимирующих заданные характеристики [33]. В качестве примеров мы рассмотрим три окна, а именно прямоугольное окно, «обобщенное» окно Хэмминга и окно Кайзера. Эти окна обладают свойствами всех возможных видов окон и позволяют достаточно хорошо понять преимущества и недостатки метода взвешивания.

1. N -точечное прямоугольное окно, соответствующее простому усечению (без модификации) ряда Фурье, описывается весовой функцией

$$w_R(n) = \begin{cases} 1, & -\frac{N-1}{2} \leq n \leq \frac{N-1}{2}, \\ 0, & \text{при других } n. \end{cases} \quad (4.64)$$

(Здесь и в следующих разделах, посвященных окнам, предполагается, что N - нечетное. С помощью простой модификации аналогичные результаты могут быть получены и для четного N . Предполагается также, что последовательность окна имеет нулевую задержку.) Частотная характеристика прямоугольного окна описывается соотношением

$$W(e^{j\omega}) = \sum_{n=-(N-1)/2}^{(N-1)/2} e^{-j\omega n} = \frac{e^{j\omega(N-1)/2} (1 - e^{-j\omega N})}{(1 - e^{-j\omega})} = \frac{e^{j\omega N/2} - e^{-j\omega N/2}}{e^{j\omega/2} - e^{-j\omega/2}} = \frac{\sin(\omega N/2)}{\sin(\omega/2)}. \quad (4.65)$$

График частотной характеристики $|W(e^{j\omega})| = \left| \frac{\sin(\omega N/2)}{\sin(\omega/2)} \right|$ представлен на рис.4.43 для случая $N=7$.

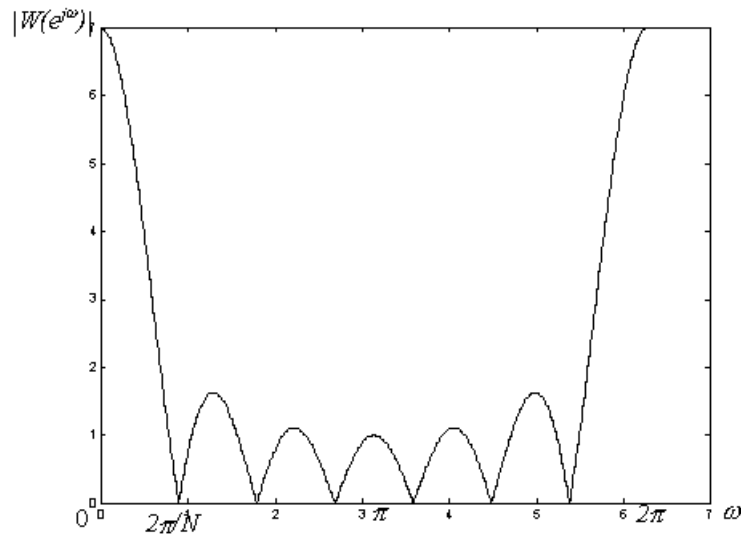


Рис.4.43. Частотная характеристика прямоугольного окна ($N=7$).

2. Окно, называемое обобщенным окном Хэмминга, имеет вид

$$w_H(n) = \begin{cases} \alpha + (1-\alpha) \cos\left(\frac{2\pi n}{N-1}\right), & -\frac{N-1}{2} \leq n \leq \frac{N-1}{2}, \\ 0, & \text{при других } n, \end{cases} \quad (4.66)$$

причем α лежит в пределах $0 \leq \alpha \leq 1$. Случай $\alpha=0,5$ соответствует окну Ханна (или окну с хэннингом), случай $\alpha=0,54$ - окну Хэмминга. Частотную характеристику рассматриваемого окна легко получить, если учесть, что оно может быть представлено в виде произведения прямоугольного окна и окна, определяемого формулой (4.66), но для всех n , т. е.

$$w_H(n) = w_R(n) \left[\alpha + (1-\alpha) \cos\left(\frac{2\pi n}{N-1}\right) \right].$$

Следовательно, частотная характеристика обобщенного окна Хэмминга может быть записана в виде

$$W_H(e^{j\omega}) = \alpha W_R(e^{j\omega}) + \frac{1-\alpha}{2} W_R(e^{j[\omega - (2\pi/(N-1))]} + \frac{1-\alpha}{2} W_R(e^{j[\omega + (2\pi/(N-1))]}).$$

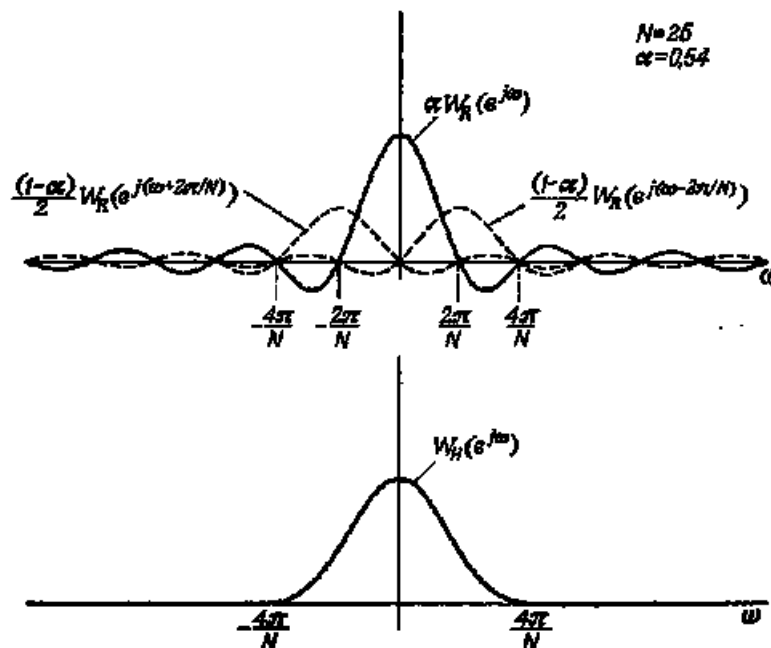


Рис.4.44. Частотная характеристика окна Хэмминга при $\alpha=0,54$.

На рис.4.44 наверху изображены графики трех компонент характеристики $W_H(e^{j\omega})$, а внизу - результирующая частотная характеристика (здесь принято $\alpha = 0,54$ и $N=25$). Создается впечатление, что частотная характеристика окна Хэмминга не имеет пульсаций на частотах выше $\omega = 4\pi/N$, однако на самом деле это не так. Пульсации настолько малы, что при линейной шкале они не видны. Из рис.4.44 видно, что ширина главного лепестка частотной характеристики окна Хэмминга в два раза больше, чем для прямоугольного окна. Однако уровень боковых лепестков в случае окна Хэмминга значительно ниже, чем у характеристики прямоугольного окна. При $\alpha = 0,54$, т. е. для обычного окна Хэмминга, 99,96% общей энергии спектра содержится в главном лепестке, а максимумы боковых лепестков на 43 дБ ниже главного максимума. В отличие от окна Хэмминга максимум боковых лепестков в спектре прямоугольного окна ниже главного максимума всего на 13,3 дБ.

Из рис.4.44 хорошо видно, каким образом при использовании окна Хэмминга достигается подавление боковых лепестков при одновременном расширении главного лепестка: боковые лепестки функций $W_R[e^{j(\omega \pm 2\pi(N-1))}]$ находятся в противофазе с боковыми лепестками $W_R(e^{j\omega})$, поэтому общий уровень боковых лепестков значительно уменьшается. В то же время пропорционально увеличивается ширина главного лепестка частотной характеристики. При расчете фильтра нижних частот расширение главного лепестка соответствует расширению переходной полосы между полосами пропускания и непропускания, тогда как уменьшение уровня боковых лепестков соответствует меньшим пульсациям в полосе пропускания и лучшему подавлению в полосе непропускания фильтра.

3. Окно Кайзера. Задача расчета хороших окон фактически сводится к математической задаче отыскания ограниченных во времени функций, преобразования Фурье которых наилучшим образом аппроксимируют функции, ограниченные по частоте, т. е. имеют минимальную энергию за пределами заданного интервала частот. При решении этой задачи в замкнутой форме для непрерывных функций времени был введен класс так называемых *вытянутых сфероидальных волновых функций*. Эти функции имеют достаточно сложный вид. Поэтому Кайзер в качестве наилучшего окна предложил относительно простую аппроксимацию этих функций. Эта аппроксимация, названная окном Кайзера, записывается в виде

$$w_K(n) = \begin{cases} \frac{I_0(\alpha \sqrt{1 - [2n/(N-1)]^2})}{I_0(\alpha)}, & -\frac{N-1}{2} \leq n \leq \frac{N-1}{2}, \\ 0, & \text{при других } n, \end{cases} \quad (4.67)$$

где α - константа, определяющая компромисс между максимальным уровнем боковых лепестков и шириной главного лепестка (или долей общей энергии в главном лепестке) частотной характеристики окна, а $I_0(x) = 1 + \sum_{k=1}^{\infty} \left[\frac{1}{k!} \left(\frac{x}{2} \right)^k \right]^2$ - модифицированная функция Бесселя первого рода нулевого порядка. Как именно величина α оказывает влияние на этот компромисс, будет проиллюстрировано ниже с помощью табл. 4.3.

Таблица 4.3

α	$D = \Delta F N$ ΔF - ширина переходной полосы, N - длина окна	Пульсация в полосе пропускания в дБ	Пульсация в полосе непропускания в дБ
2,120	1,50	$\pm 0,27$	-30
3,384	2,23	$\pm 0,0864$	-40
4,538	2,93	$\pm 0,0274$	-50
5,658	3,62	$\pm 0,00868$	-60
6,764	4,32	$\pm 0,00275$	-70
7,865	5,0	$\pm 0,000868$	-80
8,960	5,7	$\pm 0,000275$	-90
10,056	6,4	$\pm 0,000087$	-100

Частотная характеристика дискретного окна Кайзера в замкнутой форме не получена. Окно Кайзера является по существу оптимальным окном в том смысле, что оно представляет последовательность конечной длины, которая имеет минимум энергии спектра за пределами некоторой заданной частоты. Еще одним оптимальным окном является окно Дольфа—Чебышева, обеспечивающее минимальную ширину главного лепестка частотной характеристики при фиксированном уровне боковых лепестков. Все боковые лепестки в спектре этого окна имеют одинаковый уровень. Однако, как уже говорилось, ни одно из этих окон не позволяет получить оптимальную в минимаксном смысле аппроксимацию произвольной идеальной частотной характеристики, поскольку в действительности характеристика фильтра является результатом свертки частотных характеристик окна и идеального фильтра. Таким образом, несмотря на наличие оптимальных окон, оптимальных фильтров, которые рассчитывались бы с их помощью, не существует.

Как отмечено выше, метод взвешивания предполагает предварительное вычисление импульсной характеристики $h_d(n)$ (4.62) требуемого фильтра как коэффициентов ряда Фурье (4.61) его частотной характеристики. В тех случаях, когда аналитическое вычисление затруднено, можно осуществить дискретизацию по частоте функции $H_d(e^{j\omega})$ с очень малым интервалом дискретизации $2\pi/M$ ($M \gg N$), а затем вычислить обратное дискретное преобразование Фурье полученной последовательности выборок:

$$h(n) = \frac{1}{N} \sum_{k=0}^{M-1} H_d(e^{j(2\pi k/M)}) e^{j(2\pi kn/M)}.$$

Тогда функция $h(n) \approx h_d(n)$ в диапазоне $|n| \leq (N-1)/2$.

Аналитическое определение импульсных характеристик не представляет труда для большинства широко используемых типов фильтров, т.е. для ФНЧ, ФВЧ, ФПП, ФПЗ, дифференциаторов (ФД) и преобразователей Гильберта (ФПГ). В таблицу 4.4

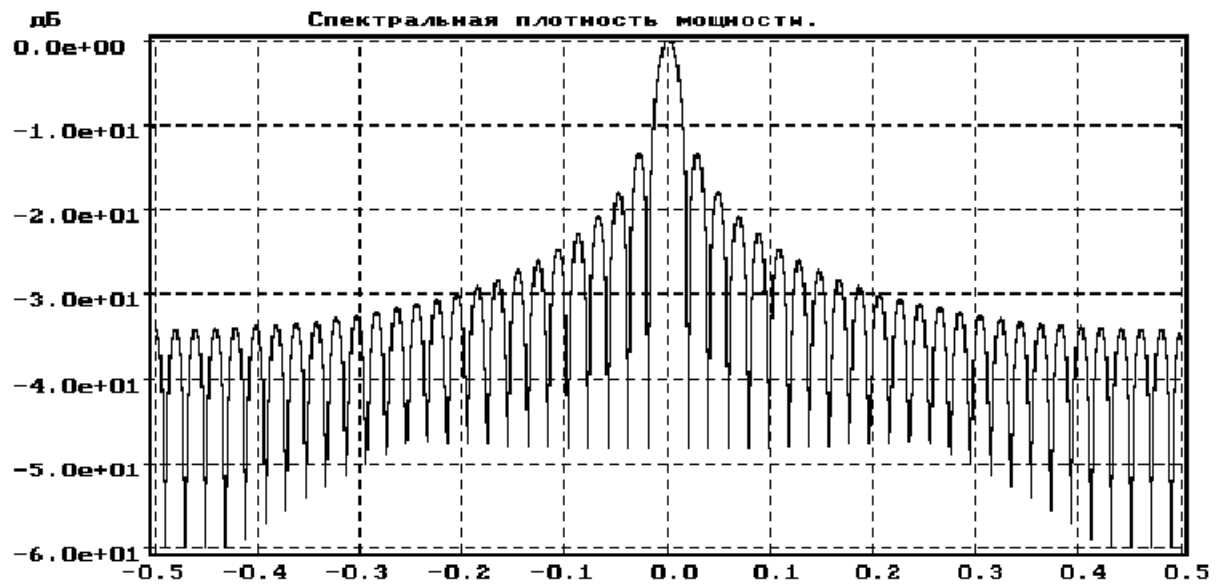
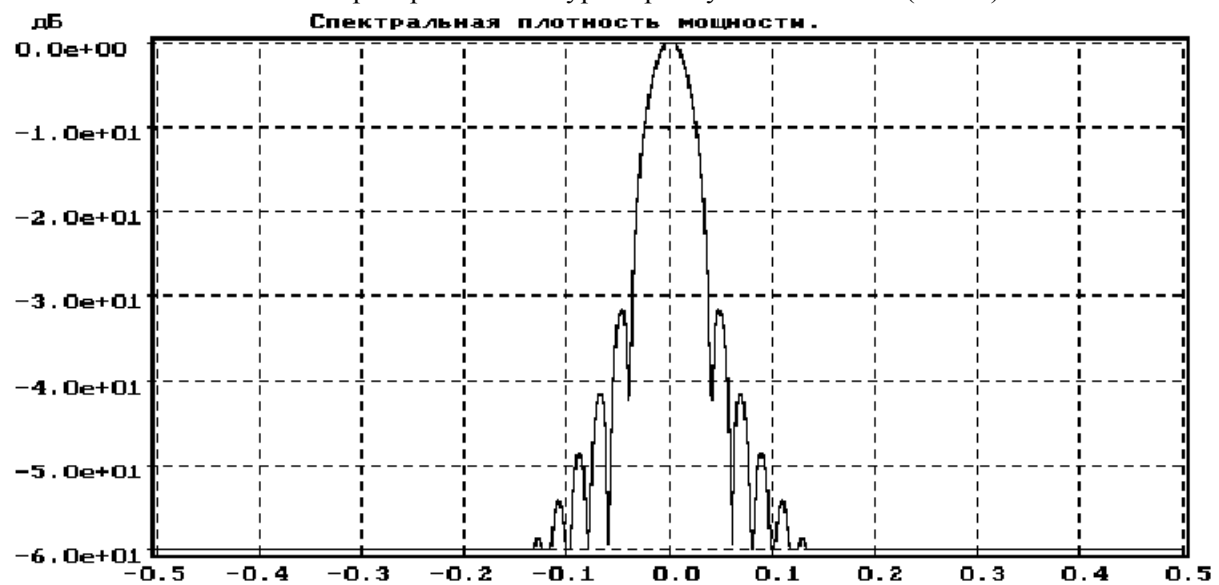
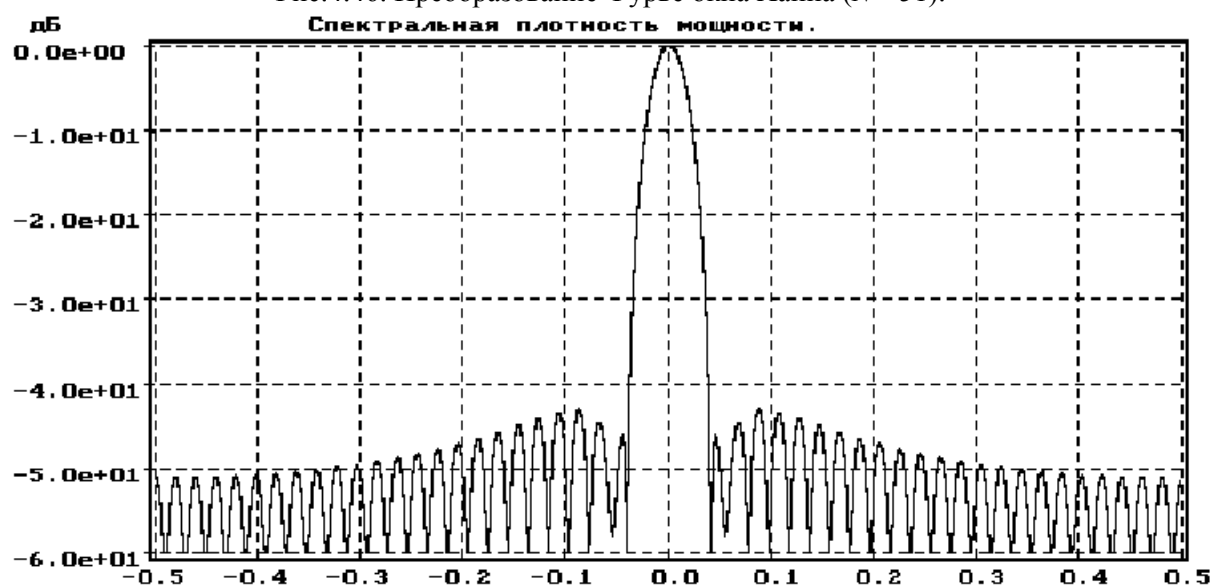
сведены импульсные характеристики перечисленных выше идеальных, физически нереализуемых фильтров, для которых f_c обозначает безразмерную частоту среза для ФНЧ, ФВЧ и ФД, а f_1 и f_2 – граничные частоты полосы пропускания для ФПП и ФПГ, либо граничные частоты полосы задерживания (непропускания) для ФПЗ.

Таблица 4.4.

Тип фильтра	N -нечетное	N -четное
ФНЧ	$h(0) = 2f_c$ $h(n) = \frac{\sin(2\pi n f_c)}{n\pi}$	$h(n) = \frac{\sin 2\pi(n - 0,5)f_c}{(n - 0,5)\pi}$
ФВЧ	$h(0) = 1 - 2f_c$ $h(n) = -\frac{\sin(2\pi n f_c)}{n\pi}$	$h(n) = \frac{1}{(n - 0,5)\pi} [\sin \pi(n - 0,5) - \sin 2\pi(n - 0,5)f_c]$
ФПП	$h(0) = 2(f_2 - f_1)$ $h(n) = \frac{1}{n\pi} [\sin(2\pi n f_2) - \sin(2\pi n f_1)]$	$h(n) = \frac{1}{(n - 0,5)\pi} [\sin 2\pi(n - 0,5)f_2 - \sin 2\pi(n - 0,5)f_1]$
ФПЗ	$h(0) = 1 - 2(f_2 - f_1)$ $h(n) = \frac{1}{n\pi} [\sin(2\pi n f_1) - \sin(2\pi n f_2)]$	$h(n) = \frac{1}{(n - 0,5)\pi} [\sin \pi(n - 0,5)f_c - \sin 2\pi(n - 0,5)f_c]$
ФД	$h(0) = 0$ $h(n) = -\frac{1}{n^2 \pi^2} [\sin(2\pi n f_c) - 2\pi n f_c \cos(2\pi n f_c)]$	$h(n) = \frac{1}{(n - 0,5)^2 \pi^2} [\sin(2\pi(n - 0,5)f_c) - 2\pi(n - 0,5)f_c \cos(2\pi(n - 0,5)f_c)]$
ФПГ	$h(0) = 0$ $h(n) = \frac{1}{n\pi} [\cos(2\pi n f_1) - \cos(2\pi n f_2)]$	$h(n) = \frac{1}{(n - 0,5)\pi} [\cos(2\pi(n - 0,5)f_1) - \cos(2\pi(n - 0,5)f_2)]$

4.5.8. Примеры расчета КИХ-фильтров методом взвешивания (окна)

Рассмотрим практические примеры использования окон для расчета идеального фильтра нижних частот. Будут рассмотрены окна: прямоугольное, Ханна, Хэмминга и Кайзера. На рис.4.45, 4.46, 4.47 представлены соответственно частотные характеристики прямоугольного окна, окна Ханна и Хэмминга для случая $N = 51$, из сравнения которых следует, что максимальные уровни боковых лепестков частотной характеристики равны: для прямоугольного окна $-13,3$ дБ (в логарифмическом масштабе), для окна Ханна $-31,5$ дБ, для окна Хэмминга -43 дБ. Однако ширина главного лепестка частотной характеристики окна Ханна и Хэмминга примерно вдвое больше ширины главного лепестка частотной характеристики прямоугольного окна. Таким образом, при аппроксимации разрывов частотной характеристики идеального фильтра (как показано ниже на примере фильтра нижних частот) с использованием окна Хэмминга ширина переходной полосы у точки разрыва будет вдвое больше, чем для прямоугольного окна.

Рис.4.45. Преобразование Фурье прямоугольного окна ($N=51$).Рис.4.46. Преобразование Фурье окна Ханна ($N=51$).Рис.4.47. Преобразование Фурье окна Хэмминга ($N=51$).

На последующих нескольких рисунках изображены характеристики фильтров нижних частот, рассчитанных с использованием различных окон. На рис.4.48, 4.49, 4.50

и 4.51 показаны полученные частотные и импульсные характеристики фильтра. Проектируемый идеальный фильтр нижних частот имеет коэффициенты ряда Фурье (т. е. отсчеты импульсной характеристики) вида

$$h(n) = \frac{\sin(2\pi f_c n)}{\pi n}, \quad -\infty < n < \infty \quad (4.68)$$

(во всех трех примерах $f_c = 0,15$). На рис.4.48б представлен результат умножения $h(n)$ на прямоугольное окно длиной $N=128$ отсчетов. Видно, что характерный вид импульсной характеристики фильтра (4.68) сохранился. Использование окон Хэмминга, Ханна и Кайзера (рис.4.49, 4.50 и 4.51) приводит к ослаблению далеко отстоящих членов $h(n)$. На рис.4.48а показана частотная характеристика фильтра нижних частот для случая прямоугольного окна. Отчетливо наблюдается эффект Гиббса, причем максимум первого бокового лепестка равен -21 дБ. Переходная полоса является в данном случае наиболее узкой ($\approx 2/N=1/64$). Однако из-за больших пульсаций, связанных с явлением Гиббса, для многих приложений этот фильтр непригоден. На рис.4.49а показана частотная характеристика фильтра нижних частот с окном Хэмминга. Для этого фильтра максимум амплитуды пульсаций в полосе не пропускания -53дБ. Ширина переходной полосы более чем в два раза больше ширины переходной полосы для прямоугольного окна. На рис.4.50а показана частотная характеристика фильтра нижних частот с окном Ханна. Для этого фильтра максимум амплитуды пульсаций в полосе не пропускания -44дБ. Ширина переходной полосы примерно в два раза больше ширины переходной полосы для прямоугольного окна. Наконец, для фильтра нижних частот с окном Кайзера (рис.4.51) максимум амплитуды пульсаций в полосе не пропускания составляет -50 дБ, а ширина переходной полосы примерно в три раза больше ширины переходной полосы для прямоугольного окна.

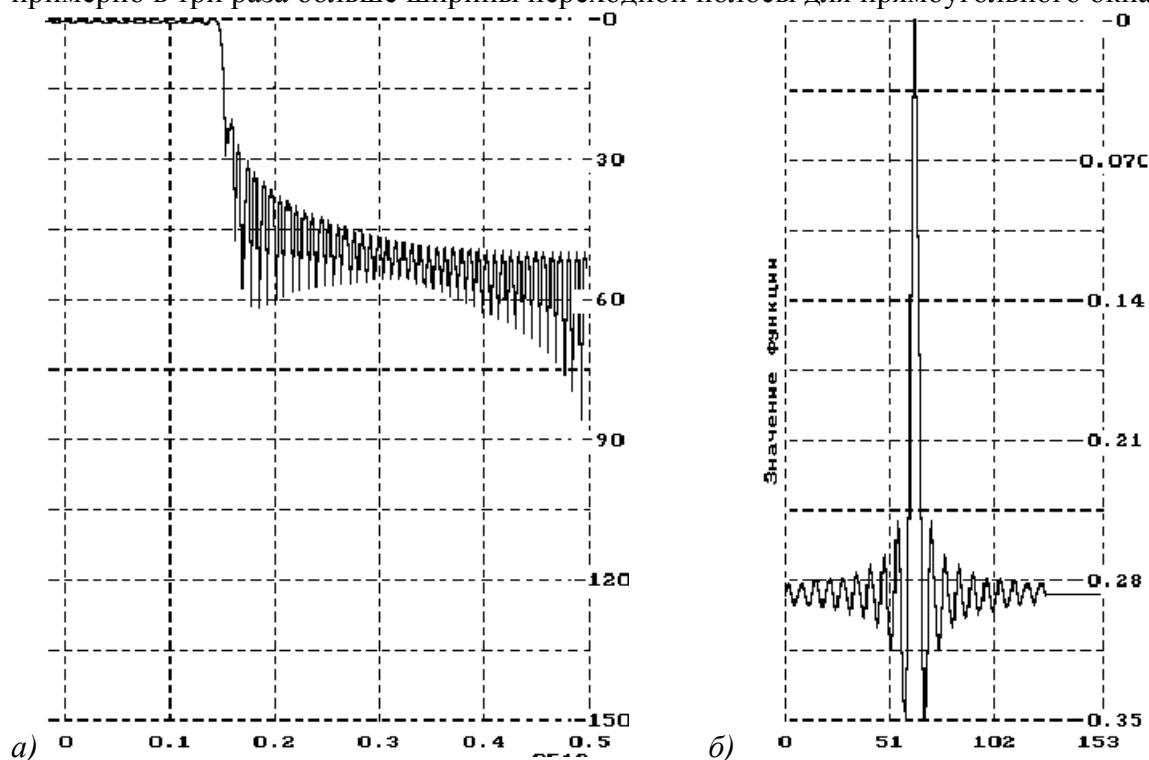


Рис.4.48. Частотная характеристика (а) и импульсная характеристика (б) фильтра нижних частот с прямоугольным окном ($N=128$).

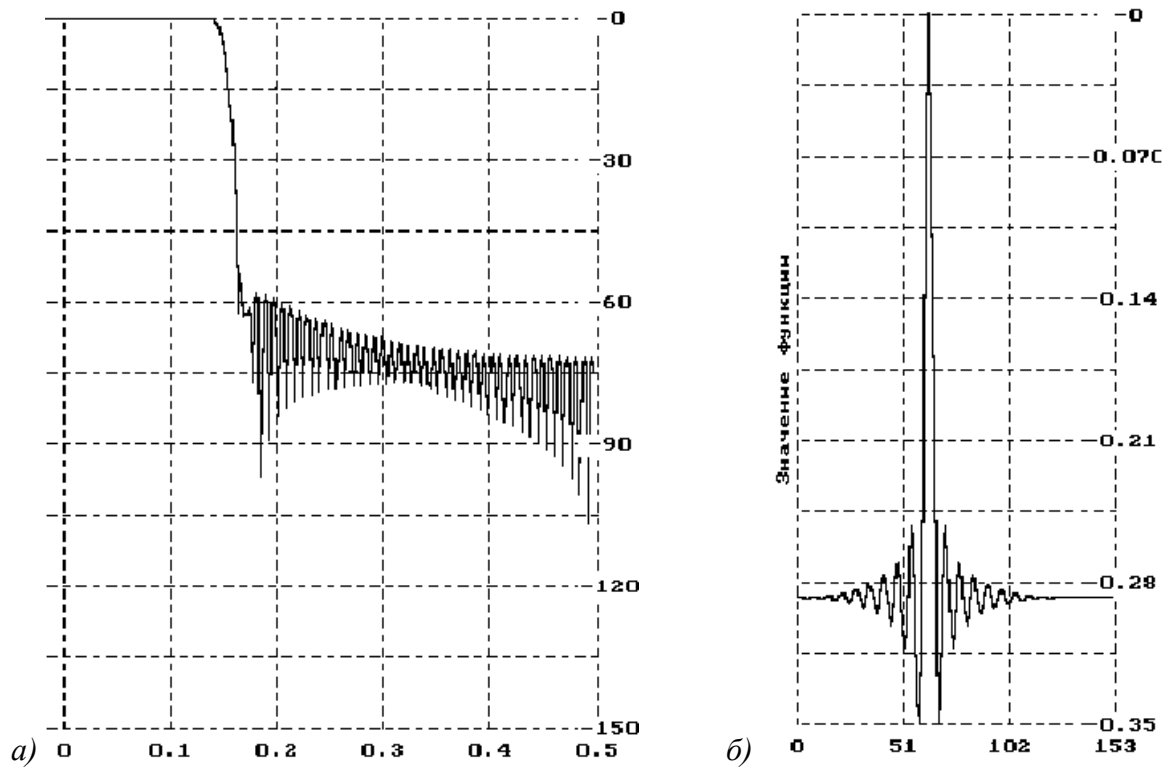


Рис.4.49. Частотная характеристика (а) и импульсная характеристика (б) фильтра нижних частот с окном Хэмминга ($N=128$).

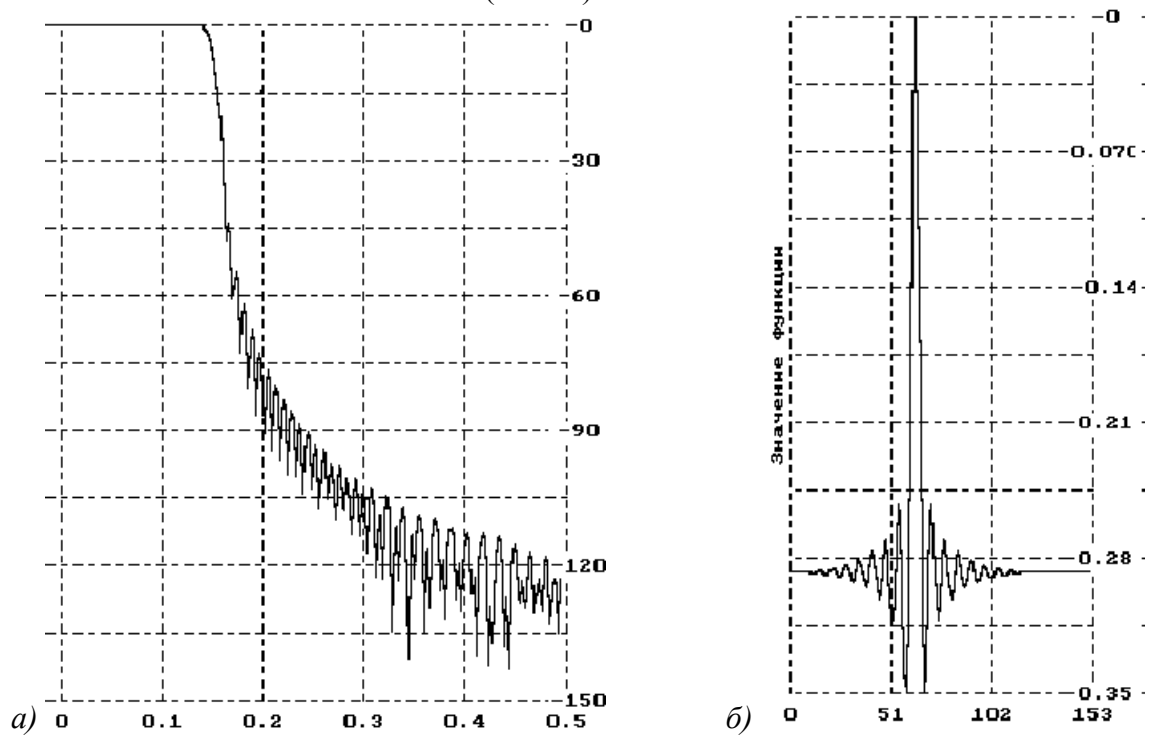


Рис.4.50. Частотная характеристика (а) и импульсная характеристика (б) фильтра нижних частот с окном Ханна ($N=128$).

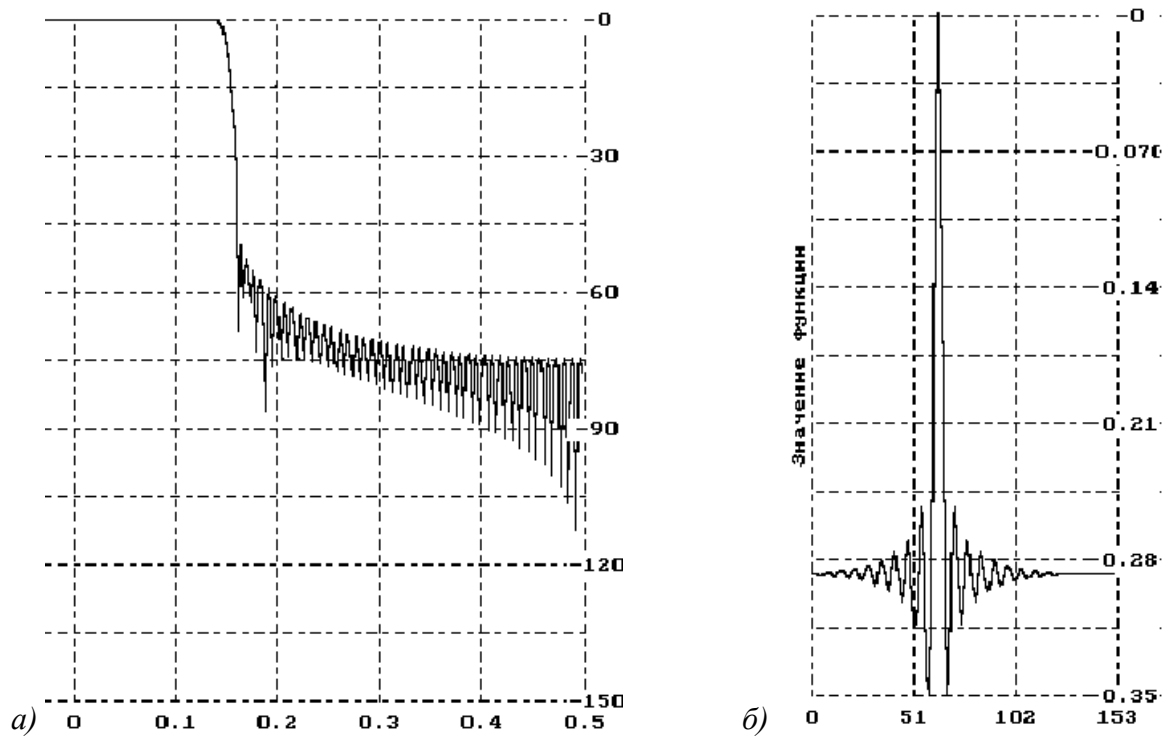


Рис.4.51. Частотная характеристика (а) и импульсная характеристика (б) фильтра нижних частот с окном Кайзера ($\alpha=4,5$, $N=128$).

Ниже приведены еще два примера, иллюстрирующие типичные КИХ-фильтры, рассчитанные методом взвешивания. На рис 4.52 представлена амплитудно-частотная характеристика фильтра верхних частот, рассчитанного с использованием окна Ханна при длине импульсной характеристики 45 отсчетов и частоте среза идеальной характеристики, равной 0,35. На рис 4.53 представлена амплитудно-частотная характеристика полосового фильтра, рассчитанного с использованием окна Кайзера ($\alpha=4,5$) при длине импульсной характеристики 46 отсчетов и верхней и нижней частотах среза идеальной характеристики, равных соответственно 0,35 и 0,15.

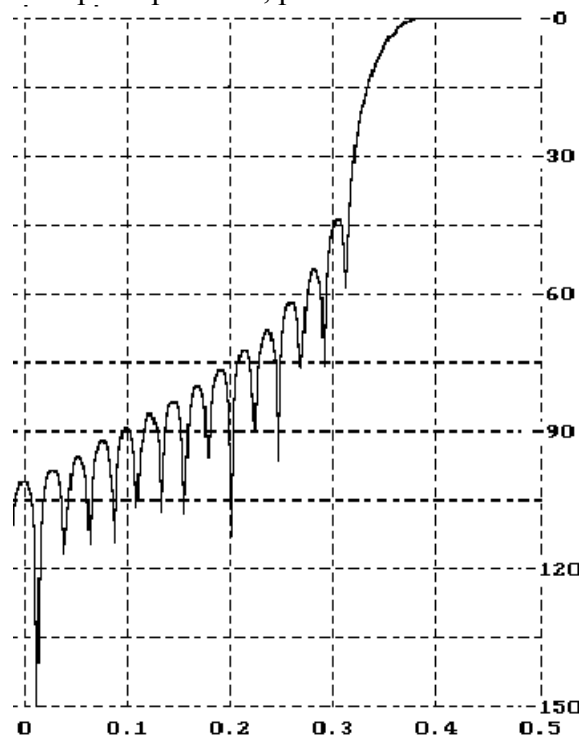
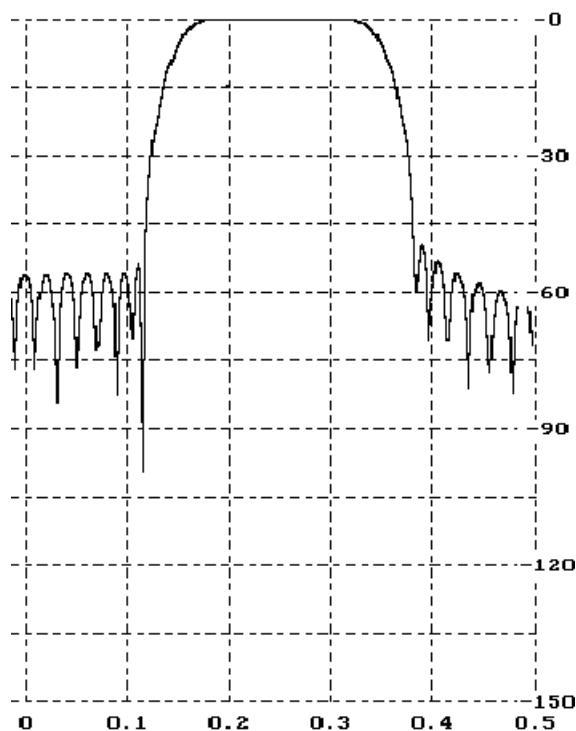


Рис.4.52. Частотная характеристика фильтра верхних частот с окном Ханна ($N=45$).

Рис.4.53. Частотная характеристика полосового фильтра с окном Кайзера ($\alpha=4,5$, $N=46$).

Приведенные примеры иллюстрируют основные принципы метода использования окон при проектировании КИХ-фильтра. За счет выбора формы окна и его длины можно осуществить некоторое управление процессом расчета. Например, для заданного ослабления в полосе непропускания, как правило, оказывается справедливо условие типа $N=A/\Delta\omega$, где $\Delta\omega$ - ширина переходной полосы [приблизительно ширина главного лепестка $W(e^{j\omega})$] и A - постоянная, зависящая от формы окна. Как было показано, форма окна является существенной при определении минимального ослабления в полосе непропускания. Для окон, которые мы рассмотрели, основные параметры для расчета фильтра нижних частот сведены в табл.4.5. Следует отметить, что величины в табл.4.5 являются приближенными; они зависят до некоторой степени от N и частоты среза требуемого фильтра.

Таблица 4.5

Окно	Амплитуда пика бокового лепестка, дБ	Ширина переходной полосы главного лепестка, рад	Минимальное затухание в полосе непропускания, дБ
Прямоугольное	-13,3	$4\pi/N$	-21
Ханна	-31,5	$8\pi/N$	-44
Хемминга	-43	$8\pi/N$	-53

Приведенные примеры показывают, что для улучшения аппроксимации идеального фильтра нижних частот приходится увеличивать ширину переходной полосы с тем, чтобы уменьшить максимальное значение ошибки аппроксимации (пульсаций). Для окна Кайзера, как уже было сказано выше, параметр α позволяет разработчику находить компромиссное решение для ширины переходной полосы ΔF и максимального уровня пульсаций. В табл.4.3 для нескольких α даны результирующие значения $D=\Delta FN$ и пульсаций в полосах пропускания и непропускания (в дБ). Данные получены путем идеального интегрирования непрерывного окна Кайзера и при больших N являются достаточно хорошим приближением к дискретному случаю. Кайзер аналитически описал эмпирическую зависимость максимального уровня пульсаций АЧХ от параметра α и связь уровня пульсаций АЧХ с шириной переходной полосы и порядком фильтра. В результате предложена инженерная методика

проектирования КИХ-фильтров с окном Кайзера, когда по заданным величинам пульсации в полосе пропускания A_p , минимального затухания в полосе непропускания A_a и ширине переходной полосы $\Delta\omega$ рассчитывается необходимый порядок N фильтра [6].

Рассмотренные примеры являются результатом аналитического вычисления бесконечных импульсных характеристик идеальных фильтров (см. таблицу 4.4) и последующей весовой обработки с применением соответствующих окон. Одним из главных преимуществ метода взвешивания является относительная простота расчета импульсных характеристик с минимальным числом вычислительных операций. Основным недостаток метода состоит в том, что получаемые КИХ-фильтры не удовлетворяют всем известным критериям оптимальности и, следовательно, в большинстве случаев они могут быть существенно улучшены.

4.5.9. Расчет КИХ-фильтров по методу частотной выборки

КИХ-фильтр может быть однозначно задан как коэффициентами импульсной характеристики $h(n)$, так и коэффициентами ДПФ импульсной характеристики $H(k)$. Напомним, что обе эти последовательности связаны соотношениями

$$\begin{aligned} H(k) &= \sum_{n=0}^{N-1} h(n) e^{-j\frac{2\pi}{N}kn}, k = \overline{0, N-1}. \\ h(n) &= \frac{1}{N} \sum_{k=0}^{N-1} H(k) e^{j\frac{2\pi}{N}kn}, n = \overline{0, N-1}. \end{aligned} \quad (4.69)$$

Кроме того, известно, что коэффициенты ДПФ КИХ-последовательности, равные $H(k)$, можно рассматривать как значения z -преобразования импульсной характеристики фильтра, найденные в N равноотстоящих точках на единичной окружности, т. е.

$$H(k) = H(e^{j\frac{2\pi}{N}k}).$$

Как было показано ранее (3.11), z -преобразование импульсной характеристики КИХ-фильтра можно выразить через коэффициенты ДПФ его импульсной характеристики соотношением:

$$H(z) = \frac{1 - z^{-N}}{N} \sum_{k=0}^{N-1} \frac{H(k)}{1 - z^{-1} e^{j\frac{2\pi}{N}k}}. \quad (4.70)$$

Из соотношения (4.70) следует, что для аппроксимации произвольной непрерывной частотной характеристики $\hat{H}(e^{j\omega})$ следует произвести ее *дискретизацию по частоте* в N равноотстоящих точках на единичной окружности (взять частотную выборку) $H(k) = \hat{H}(e^{j\frac{2\pi}{N}k})$ и найти непрерывную частотную характеристику, интерполируя отсчеты частотной характеристики. В этом случае ошибка аппроксимации на частотах взятия выборки будет в точности равна нулю и иметь конечную величину в промежуточных точках. Чем более гладкой является аппроксимируемая частотная характеристика, тем меньше ошибка аппроксимации между частотными отсчетами. Пример такой аппроксимации показан нарис.4.54. На рис.4.54а изображена заданная частотная характеристика (сплошная линия) и выборка из частотных отсчетов (точки). На рис.4.54б представлен результат непрерывной интерполяции частотных отсчетов. Описанную процедуру можно было бы использовать непосредственно для расчета КИХ-фильтров, однако для улучшения качества аппроксимации, т. е. для уменьшения ошибки аппроксимации, часть частотных отсчетов целесообразно сделать независимыми переменными. Значения этих независимых переменных обычно рассчитывают методами оптимизации на

вычислительной машине таким образом, чтобы минимизировать некоторую простую функцию ошибки аппроксимации (например, наибольшую ошибку аппроксимации). В качестве независимых переменных можно выбрать, например, частотные отсчеты, расположенные в переходной полосе между двумя полосами, внутри которых частотная характеристика определена (т. е. в случае фильтра нижних частот между полосами пропускания и непропускания).

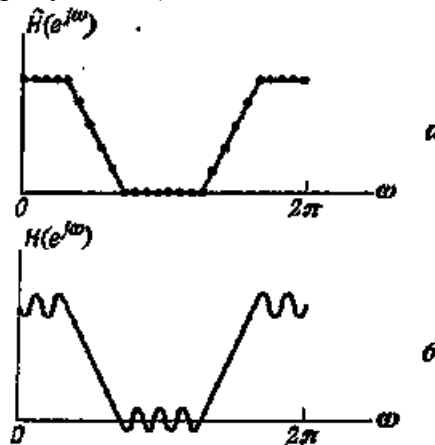


Рис.4.54. Иллюстрация метода частотной выборки.

Чтобы понять, почему при такой методике оптимизации часто лишь несколько из N частотных отсчетов могут существенно уменьшить ошибку аппроксимации, необходимо вычислить значения $H(z)$ на единичной окружности. Получаемая при этом интерполяционная формула для расчета частотной характеристики фильтра в функции непрерывной частоты имеет вид

$$H(e^{j\omega}) = \frac{1 - e^{-j\omega N}}{N} \sum_{k=0}^{N-1} \frac{H(k)}{1 - e^{j\frac{2\pi}{N}k} e^{-j\omega}} = \frac{e^{-j\omega\left(\frac{N-1}{2}\right)}}{N} \sum_{k=0}^{N-1} H(k) \frac{\sin N\left(\frac{\omega}{2} - \frac{\pi k}{N}\right)}{\sin\left(\frac{\omega}{2} - \frac{\pi k}{N}\right)} e^{j\pi k\left[1 - \frac{1}{N}\right]}. \quad (4.71)$$

Заметим, что (4.71) представляет другую форму полученного ранее соотношения (3.12).

Отсюда следует, что частотная характеристика фильтра является линейной комбинацией частотных интерполирующих функций вида

$$S_k(\omega) = \frac{\sin N\left(\frac{\omega}{2} - \frac{\pi k}{N}\right)}{N \sin\left(\frac{\omega}{2} - \frac{\pi k}{N}\right)} e^{j\pi k\left[1 - \frac{1}{N}\right]} \quad (4.72)$$

со значениями частотных отсчетов $H(k)$ в качестве коэффициентов. Таким образом, вклад каждого частотного отсчета в общую частотную характеристику пропорционален его значению $H(k)$, умноженному на функцию $\sin(N\omega/2)/\sin(\omega/2)$, смещаемую по частоте на $\pi k/N$. Оказалось, что интерполирующие функции [т. е. $\sin(N\omega/2)/\sin(\omega/2)$], связанные с частотными отсчетами из переходной полосы, обеспечивают хорошее подавление пульсаций в примыкающих частотных полосах. Таким образом, оптимизируя значения только тех незадаанных частотных отсчетов, которые лежат в предварительно выбранных переходных полосах, можно получить фильтры с очень хорошими характеристиками.

4.5.9.1. Формулировка и решение задачи оптимизации

Чтобы найти оптимальные значения незадаанных частотных отсчетов, нужно составить и решить систему уравнений, математически описывающих задачу

оптимизации. На рис.4.55 иллюстрируется типичный способ задания фильтра при расчете его методом частотной выборки. Частотная характеристика фильтра задана в полосах 1 и 2 и не задана в переходной полосе между ними. Сплошной кривой на рис.4.55 представлена желаемая частотная характеристика $\hat{H}(e^{j\omega})$, а точками отмечены частотные отсчеты $H(k) = \hat{H}(e^{j\frac{2\pi}{N}k})$. Обозначим для удобства частотные отсчеты в переходной полосе через T_1 и T_2 . Именно эти отсчеты необходимо оптимизировать.

На рис.4.55 показана только половина частотных отсчетов, так как, чтобы импульсная характеристика фильтра была действительной, последовательность частотных отсчетов $H(k)$ должна иметь относительно своего центра комплексно сопряженную симметрию. Кроме того, чтобы фильтр обладал строго линейной фазовой характеристикой, на значения $H(k)$ накладываются дополнительные ограничения. Предположим, что $H(e^{j\omega})$ можно выразить в виде

$$H(e^{j\omega}) = e^{-j\omega(N-1)/2} \sum_{k=0}^{N-1} H(k) S_k(\omega) = e^{-j\omega(N-1)/2} H^*(e^{j\omega}), \quad (4.73)$$

где $S_k(\omega)$ — результирующая частотная интерполирующая функция,

$$H^*(e^{j\omega}) = \sum_{k=0}^{N-1} H(k) S_k(\omega). \quad (4.74)$$

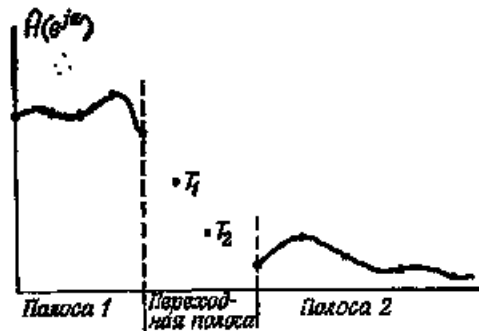


Рис.4.55. Задание фильтра при расчете методом частотной выборки

Ниже при составлении уравнений будет использована действительная функция $H^*(e^{j\omega})$ поскольку множитель с линейным изменением фазы в (4.73) при проектировании фильтра можно не учитывать.

Функцию $H^*(e^{j\omega})$ можно представить в более простой форме, выделяя в явном виде незадаанные частотные отсчеты T_m в переходной полосе:

$$H^*(e^{j\omega}) = \sum_{k=0}^{N-1} H(k) S_k(\omega) = B(e^{j\omega}) + \sum_{m=L_1}^{L_2} T_m S_m(\omega), \quad (4.75)$$

где $B(e^{j\omega}) = \sum_{k=0}^{L_1-1} H(k) S_k(\omega) + \sum_{k=L_2+1}^{N-1} H(k) S_k(\omega)$ — функция, учитывающая вклад всех

задаваемых частотных отсчетов, L_1 и L_2 — номера первой и последней выборок в переходной полосе соответственно, амплитуды которых T_m требуется рассчитать.

Чтобы найти эти незадаанные частотные отсчеты T_m , необходимо для частот в пределах полос 1 и 2 составить систему ограничивающих уравнений. Типичными ограничениями для такой системы уравнений могут быть следующие: минимизируется максимум ошибки

$$\left| W(e^{j\omega}) [H^*(e^{j\omega}) - \hat{H}(e^{j\omega})] \right|, \quad (4.76)$$

где ω изменяется в полосах 1 и 2, $W(e^{j\omega})$ — известная весовая функция ошибки аппроксимации частотной характеристики, позволяющая устанавливать различный вес

ошибки на различных участках аппроксимации.

Обозначая через δ максимальную ошибку в областях 1 и 2 и проводя выборки на плотной сетке частот ω_k , $k=0,1,\dots,N_s$, из (4.76) получаются следующие ограничения:

$$\begin{aligned} W(e^{j\omega_k})[H^*(e^{j\omega_k}) - \hat{H}(e^{j\omega_k})] &\leq \delta, \\ -W(e^{j\omega_k})[H^*(e^{j\omega_k}) - \hat{H}(e^{j\omega_k})] &\leq \delta, \quad k=0,1,\dots,N_s. \end{aligned} \quad (4.77)$$

После подстановки (4.75) в (4.77) задача формулируется следующим образом. Найти коэффициенты T_m , удовлетворяющие ограничениям

$$\begin{aligned} W(e^{j\omega_k}) \sum_{m=L_1}^{L_2} T_m S_m(\omega) - \delta &\leq W(e^{j\omega_k}) \hat{H}(e^{j\omega_k}) - W(e^{j\omega_k}) B(e^{j\omega_k}), \\ -W(e^{j\omega_k}) \sum_{m=L_1}^{L_2} T_m S_m(\omega) - \delta &\leq -W(e^{j\omega_k}) \hat{H}(e^{j\omega_k}) + W(e^{j\omega_k}) B(e^{j\omega_k}), \quad k=0,1,\dots,N_s. \end{aligned} \quad (4.78)$$

и максимизирующие величину $(-\delta)$. Эта формулировка соответствует так называемой двойственной задаче линейного программирования, которая в общем случае может быть сформулирована так: найти вектор $(y_1, y_2, \dots, y_M)$, удовлетворяющий ограничениям

$$\sum_{j=1}^M a_{ij} y_j \leq c_i, \quad i=1,2,\dots,N$$

и максимизирующий сумму $\sum_{j=1}^M b_j y_j$. Обычно на ЭВМ реализуется симплекс-

метод решения задачи линейного программирования, для которого разработаны эффективные программы.

В общем случае, когда частотная характеристика задана на нескольких участках, разделенных переходными полосами, в которых частотные отсчеты не задаются, можно составить систему линейных неравенств относительно этих неизвестных частотных отсчетов и решить ее методами линейного программирования. Итак, при расчете фильтров методом частотной выборки используются лишь те коэффициенты ДПФ импульсной характеристики фильтра (образующие частотную выборку), которые находятся в интересующих нас полосах, а остальные коэффициенты ДПФ, попадающие в переходные полосы, считаются незадаанными. Относительно этих неизвестных коэффициентов составляется система линейных неравенств, описывающая ограничения, накладываемые на частотную характеристику. Решая эту систему методами линейного программирования, получают значения незадаанных частотных отсчетов.

При расчете фильтров методом частотной выборки используют два способа выбора частотных отсчетов. Первый способ состоит в выборе N равноотстоящих отсчетов заданной частотной характеристики на единичной окружности в точках

$$\omega_k = \frac{2\pi k}{N}, \quad k=0,1,\dots,N, \quad (4.79)$$

соответствующих N частотам, для которых вычисляется N -точечное ДПФ. Возможен другой набор равноотстоящих частот, также пригодный для расчета фильтров методом частотной выборки:

$$\omega_k = \frac{2\pi(k+1/2)}{N}, \quad k=0,1,\dots,N. \quad (4.80)$$

Точное расположение обоих наборов частот, заданных соотношением (4.79) для фильтров вида 1 и соотношением (4.80) для фильтров вида 2, показано на рис.4.56 для случаев четного и нечетного N . Видно, что для фильтров вида 1 точка отсчета частот

выборки соответствует $\omega = 0$, а для фильтров вида 2 она соответствует $\omega = \theta/2$, причем величина $\theta = 2\pi/N$ для фильтров обоих видов равна угловому расстоянию между соседними отсчетами.

Использование фильтров вида 2 с частотной выборкой предоставляет разработчику дополнительные возможности при расчете фильтров с заданной частотной характеристикой. Так, граничная частота полосы фильтра может оказаться намного ближе к точке, используемой в фильтре вида 2, чем в фильтре вида 1, так что в этом случае для решения задачи оптимизации предпочтительнее фильтр вида 2 с частотной выборкой. Поскольку при оптимизации несущественно, как расположены отсчеты, то для вычисления коэффициентов фильтра можно использовать фильтр любого вида при условии, что для каждого из них имеется действительная функция $H^*(e^{j\omega})$.

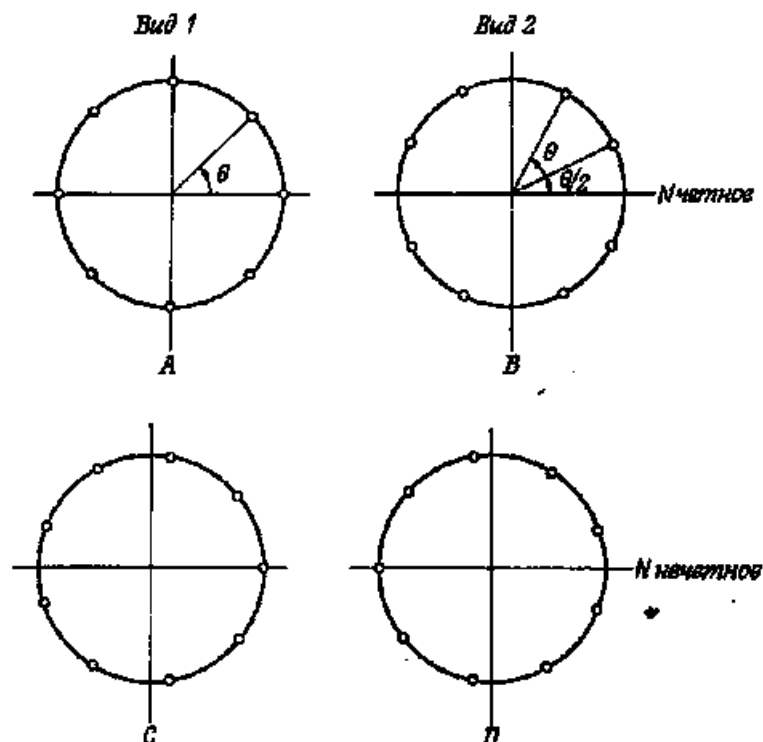


Рис.4.56. Четыре способа расположения отсчетов частотной выборки.

Выше было показано, что частотная характеристика фильтра вида 1 с частотной выборкой описывается выражением (4.71). Найдем выражение для частотной характеристики $H^*(e^{j\omega})$, необходимое для решения задачи оптимизации проектируемого фильтра с учетом ограничений, которые обеспечат получение действительной импульсной характеристики $h(n)$ и частотной характеристики с линейной фазой.

Для действительных фильтров с линейной фазой [с задержкой на $(N-1)/2$ отсчетов] частотные отсчеты $H(k)$ можно записать в виде

$$H(k) = |H(k)|e^{j\varphi(k)}, k = 0, 1, \dots, N-1, \quad (4.81)$$

причем

$$|H(k)| = |H(N-k)|, k = 0, 1, \dots, N. \quad (4.82)$$

Кроме того, при четном N

$$\varphi(k) = \begin{cases} -\frac{2\pi}{N} k \left(\frac{N-1}{2} \right), & k = 0, 1, \dots, \frac{N}{2} - 1, \\ -\frac{2\pi}{N} (N-k) \left(\frac{N-1}{2} \right), & k = \frac{N}{2} + 1, \dots, N-1, \\ 0, & k = \frac{N}{2}, \end{cases} \quad (4.83)$$

$$H\left(\frac{N}{2}\right) = 0,$$

а при нечетном N

$$\varphi(k) = \begin{cases} -\frac{2\pi}{N} k \left(\frac{N-1}{2} \right), & k = 0, 1, \dots, \frac{N-1}{2} \\ -\frac{2\pi}{N} (N-k) \left(\frac{N-1}{2} \right), & k = \frac{N+1}{2}, \dots, N-1. \end{cases} \quad (4.84)$$

Уравнение (4.83) отражает тот факт, что, как ранее было показано, для фильтра с линейной фазой и четным N функция $H(e^{j\omega})=0$ при $\omega=\pi$.

Подстановка выражения (4.81) в (4.71) с учетом (4.82) и (4.83) приводит к соотношению (для четного N):

$$H(e^{j\omega}) = e^{-j\omega\left(\frac{N-1}{2}\right)} \left(\frac{|H(0)|}{N} \frac{\sin(\omega N / 2)}{\sin(\omega / 2)} + \sum_{k=0}^{\frac{N}{2}-1} \left| \frac{H(k)}{N} \right| \left\{ \frac{\sin[N(\omega / 2 - \pi k / N)]}{\sin(\omega / 2 - \pi k / N)} + \frac{\sin[N(\omega / 2 + \pi k / N)]}{\sin(\omega / 2 + \pi k / N)} \right\} \right) = e^{-j\omega\left(\frac{N-1}{2}\right)} H^*(e^{j\omega}), \quad (4.8)$$

5)

где

$$H^*(e^{j\omega}) = \frac{|H(0)|}{N} \frac{\sin(\omega N / 2)}{\sin(\omega / 2)} + \sum_{k=0}^{\frac{N}{2}-1} \left| \frac{H(k)}{N} \right| \left\{ \frac{\sin[N(\omega / 2 - \pi k / N)]}{\sin(\omega / 2 - \pi k / N)} + \frac{\sin[N(\omega / 2 + \pi k / N)]}{\sin(\omega / 2 + \pi k / N)} \right\} \quad (4.86)$$

Аналогичным путем для фильтра вида 1 и нечетного N можно получить:

$$H^*(e^{j\omega}) = \frac{|H(0)|}{N} \frac{\sin(\omega N / 2)}{\sin(\omega / 2)} + \sum_{k=0}^{\frac{N-1}{2}} \left| \frac{H(k)}{N} \right| \left\{ \frac{\sin[N(\omega / 2 - \pi k / N)]}{\sin(\omega / 2 - \pi k / N)} + \frac{\sin[N(\omega / 2 + \pi k / N)]}{\sin(\omega / 2 + \pi k / N)} \right\} \quad (4.87)$$

Аналогичным путем для фильтра вида 2 могут быть получены соответствующие выражения для $H^*(e^{j\omega})$ [2]. Таким образом, приведенные выражения (4.86) и (4.87) для $H^*(e^{j\omega})$ могут быть использованы в процедуре оптимизации. Выбор вида фильтра с частотной выборкой, четного или нечетного N производится разработчиком и зависит прежде всего от назначения рассчитываемого фильтра.

4.5.9.2. Примеры расчета КИХ-фильтров по методу частотной выборки

Приведем некоторые результаты расчета КИХ-фильтров по методу частотной выборки. На рис.4.57-4.59 представлен АЧХ типичных фильтров нижних частот, рассчитанных этим методом. Критерий оптимизации во всех приводимых примерах состоял в минимизации максимума пульсаций в полосе непропускания. Рис.4.57

соответствует фильтру вида 1 с частотной выборкой ($N = 256$) и с тремя подбираемыми частотными отсчетами в переходной полосе. Максимум пульсаций в полосе непропускания составляет около 0,05. На рис.4.58 показана характеристика узкополосного фильтра нижних частот ($N = 65$) с частотой среза, равной 0,0306. Здесь подбираются три отсчета в переходной полосе; максимум пульсаций в полосе непропускания близок к 0,00002 (или -93 дБ). На рис.4.59 изображена характеристика широкополосного фильтра нижних частот ($N = 64$) с частотой среза, равной 0,4355. Здесь также подбираются три частотных отсчета; в результате максимум пульсаций в полосе непропускания составляет около 0,000002 (или -115 дБ).

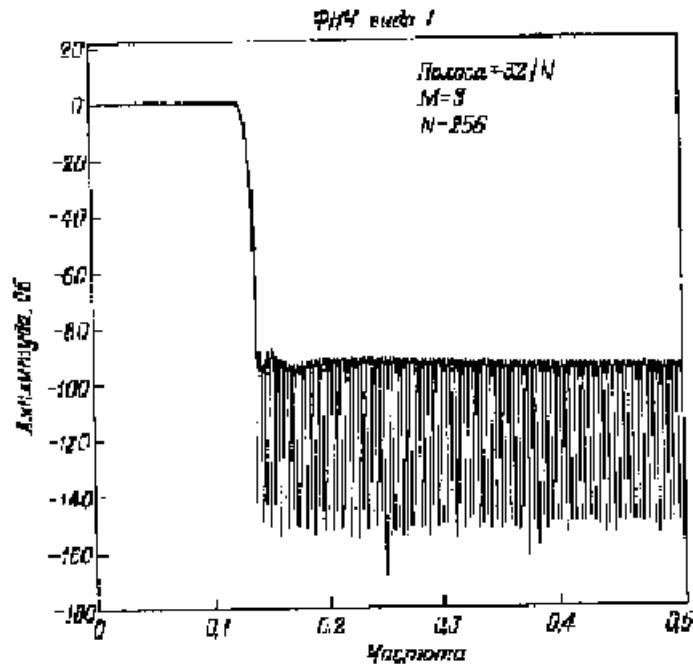


Рис.4.57. Частотная характеристика фильтра нижних частот, рассчитанного методом частотной выборки.

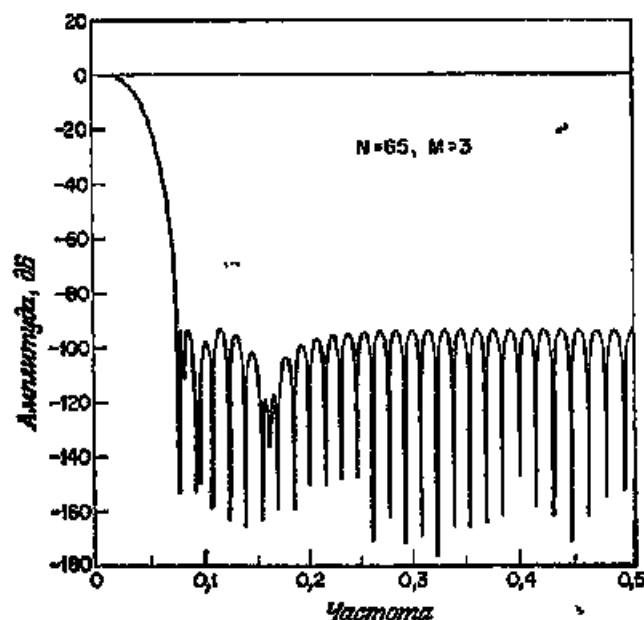


Рис.4.58. Частотная характеристика узкополосного фильтра нижних частот, рассчитанного методом частотной выборки.

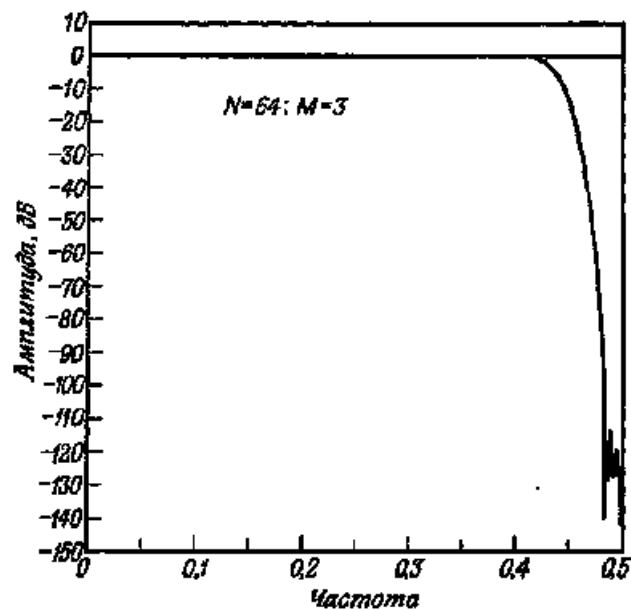


Рис.4.59. Частотная характеристика широкополосного фильтра нижних частот, рассчитанного методом частотной выборки.

В общем случае при проектировании фильтров нижних частот с одним подбираемым отсчетом в переходной полосе (с минимизацией максимума пульсаций в полосе не пропускания) можно достичь подавления вне полосы пропускания от 44 до 54 дБ. С двумя подбираемыми отсчетами в переходной полосе можно достичь подавления от 65 до 75 дБ, а возможное подавление при трех подбираемых отсчетах в переходной полосе составляет 85—95 дБ. Интересно отметить, что как для узкополосных, так и для широкополосных фильтров результаты обычно *лучше* (т. е. подавление в полосе не пропускания сильнее), чем для фильтров со средними значениями полос пропускания. Это можно объяснить следующим образом. При малой ширине нормализованной полосы приходится подавлять очень маленькие пульсации, поэтому отсчеты в переходной полосе обеспечивают отличное подавление этих пульсаций. При большой величине нормализованной полосы остается очень небольшая часть диапазона частот, в котором приходится подавлять пульсации, поэтому отсчеты в переходной полосе вновь обеспечат эффективное подавление в полосе не пропускания.

Полосовые фильтры можно рассчитывать практически так же, как фильтры нижних частот. На рис.4.60 изображена характеристика полосового фильтра вида 1 ($N = 128$) с тремя подбираемыми частотными отсчетами, расположенными симметрично по обе стороны от полосы пропускания. Максимум пульсаций в полосе не пропускания составляет около 0,000025 (или -91 дБ), а огибающая пульсаций спадает до уровня 10^{-6} (или до уровня -120 дБ) на нулевой частоте и частоте, равной половине частоты дискретизации.

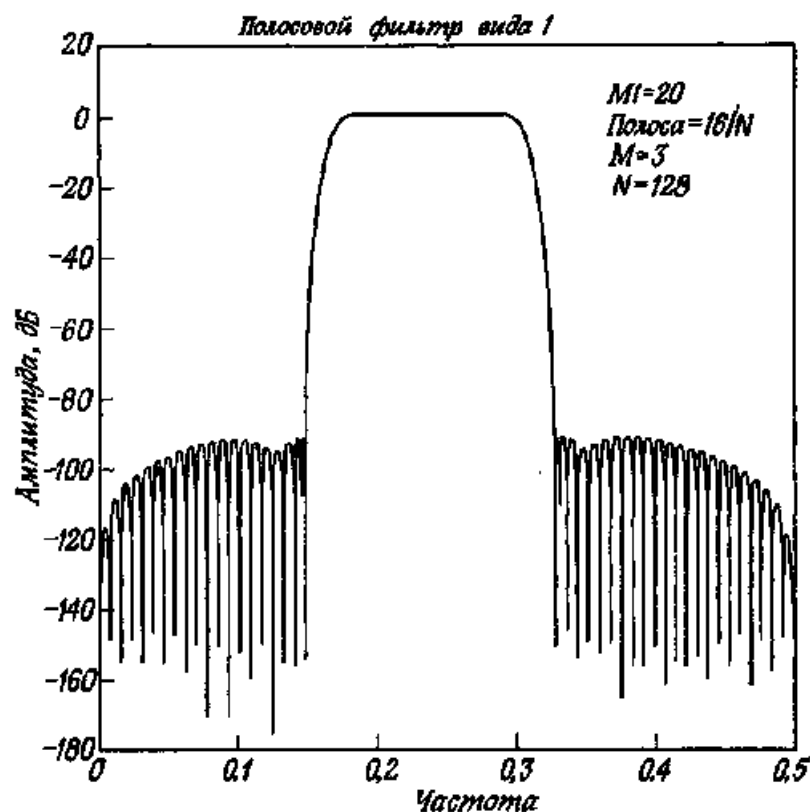


Рис.4.60. Частотная характеристика полосового фильтра, рассчитанного методом частотной выборки.

Расчеты методом частотной выборки особенно эффективны в случае узкополосных частотно-избирательных фильтров, когда лишь несколько отсчетов частотной характеристики являются ненулевыми. В таких случаях, как отмечалось выше, реализация на основе частотной выборки может оказаться значительно более эффективной, чем метод прямой свертки или свертки на основе ДПФ. В общем случае, даже если количество ненулевых отсчетов не очень мало, метод расчета на основе частотной выборки дает прекрасные результаты. Однако этому методу не хватает гибкости в точном определении частот среза полос пропускания и непропускания, поскольку положение отсчетов ограничено целочисленными значениями, кратными $2\pi/N$. Выбирая N достаточно большим, можно получить отсчеты произвольно близкими к любой заданной частоте, однако это является неэффективным подходом. По этой причине, особенно если фильтр не должен строиться на основе структуры с частотной выборкой, были разработаны другие методы оптимизации с более эффективными свойствами для расчета частотно-избирательного фильтра в общем виде.

4.5.10. Расчет оптимальных КИХ-фильтров с минимаксной ошибкой

При рассмотрении свойств КИХ-фильтров с линейной фазой было показано (соотношения (4.57)–(4.60)), что частотная характеристика фильтров четырех различных видов может быть представлена в виде

$$H(e^{j\omega}) = e^{-j\omega \frac{N-1}{2}} e^{jL \frac{\pi}{2}} H^*(e^{j\omega}). \quad (4.88)$$

Значения L и выражения $H^*(e^{j\omega})$ для каждого из четырех видов фильтров приведены в таблице 4.6

Таблица 4.6

Вид фильтра	L	$H^*(e^{j\omega})$
1. N -нечетное, симметричная импульсная характеристика	0	$\sum_{n=0}^{(N-1)/2} a(n) \cos(\omega n)$
2. N -четное, симметричная импульсная характеристика	0	$\sum_{n=1}^{N/2} b(n) \cos[\omega(n - 1/2)]$

3. N -нечетное, антисимметричная импульсная характеристика	1	$\sum_{n=1}^{(N-1)/2} c(n) \sin(\omega n)$
4. N -четное, антисимметричная импульсная характеристика	1	$\sum_{n=1}^{N/2} d(n) \sin[\omega(n - 1/2)]$

Используя простые тригонометрические тождества, каждое выражение для $H^*(e^{j\omega})$ из табл. 4.6 можно записать в виде произведения фиксированной функции ω [обозначим ее через $Q(\omega)$] и члена, представляющего собой сумму косинусов [обозначим его через $P(e^{j\omega})$]. Итак,

$$H^*(e^{j\omega}) = Q(\omega)P(e^{j\omega}),$$

где выражения для $Q(\omega)$ и $P(e^{j\omega})$ для различных видов фильтров представлены в таблице 4.7.

Таблица 4.7

Вид фильтра	$Q(\omega)$	$P(e^{j\omega})$
1	1	$\sum_{n=0}^{(N-1)/2} \tilde{a}(n) \cos(\omega n)$
2	$\cos(\omega/2)$	$\sum_{n=0}^{(N/2)-1} \tilde{b}(n) \cos(\omega n)$
3	$\sin(\omega)$	$\sum_{n=0}^{(N-1)/2} \tilde{c}(n) \cos(\omega n)$
4	$\sin(\omega/2)$	$\sum_{n=0}^{(N/2)-1} \tilde{d}(n) \cos(\omega n)$

Коэффициенты сумм в таблицах 4.6 и 4.7 связаны соотношениями:

$$a(n) = \tilde{a}(n), n = 0, 1, \dots, N-1;$$

$$\begin{cases} b(1) = \tilde{b}(0) + \tilde{b}(1)/2 \\ b(k) = [\tilde{b}(k-1) + \tilde{b}(k)]/2, k = 2, 3, \dots, N-1; \\ b(N) = \tilde{b}(N-1); \end{cases}$$

$$\begin{cases} c(1) = \tilde{c}(0) - \tilde{c}(2)/2, \\ c(k) = [\tilde{c}(k-1) - \tilde{c}(k+1)]/2, k = 2, 3, \dots, N-1, \\ c(N-1) = \tilde{c}(N-2)/2, \\ c(N) = \tilde{c}(N-1)/2; \end{cases}$$

$$\begin{cases} d(1) = \tilde{d}(0) - \tilde{d}(1)/2, \\ d(k) = [\tilde{d}(k-1) - \tilde{d}(k)]/2, k = 2, 3, \dots, N-1, \\ d(N) = \tilde{d}(N-1)/2. \end{cases} \quad (4.89)$$

Чтобы показать, как задачу расчета оптимального КИХ-фильтра с линейной фазой сформулировать в виде задачи чебышевской аппроксимации, необходимо ввести заданную (действительную) частотную характеристику фильтра $D(e^{j\omega})$ и весовую функцию ошибки аппроксимации $W(e^{j\omega})$, что позволяет разработчику выбирать различную величину ошибки для разных частотных полос. Взвешенная ошибка аппроксимации $E(e^{j\omega})$ по определению равна

$$E(e^{j\omega}) = W(e^{j\omega}) [D(e^{j\omega}) - H^*(e^{j\omega})]. \quad (4.90)$$

Записав $H^*(e^{j\omega})$ в виде произведения $P(e^{j\omega})$ и $Q(\omega)$, представим $E(e^{j\omega})$ как

$$E(e^{j\omega}) = W(e^{j\omega}) [D(e^{j\omega}) - P(e^{j\omega})Q(\omega)] = W(e^{j\omega})Q(\omega)[D(e^{j\omega})/Q(\omega) - P(e^{j\omega})] \quad (4.91).$$

Формула (4.91) справедлива во всех точках, за исключением, возможно, точки $\omega = 0$ и (или) точки $\omega = \pi$. Вводя обозначения

$$\hat{W}(e^{j\omega}) = W(e^{j\omega})/Q(\omega),$$

$$\hat{D}(e^{j\omega}) = D(e^{j\omega})/Q(\omega).$$

выражение для функции ошибки можно записать в виде

$$E(e^{j\omega}) = \hat{W}(e^{j\omega})[\hat{D}(e^{j\omega}) - P(e^{j\omega})]. \quad (4.91)$$

Теперь задачу чебышевской аппроксимации можно сформулировать как задачу поиска таких коэффициентов $a(n)$, $b(n)$, $c(n)$ или $d(n)$, которые минимизируют максимум модуля ошибки $E(e^{j\omega})$ в тех частотных полосах, где выполняется аппроксимация. Используя символ $\|E(e^{j\omega})\|$ для обозначения минимальной ошибки, задачу чебышевской аппроксимации математически можно сформулировать следующим образом:

$$\|E(e^{j\omega})\| = \min_{\text{коэффициенты } \omega \in A} [\max |E(e^{j\omega})|] \quad (4.92)$$

где A — совокупность всех интересующих нас частотных полос. Для получения решения уравнения (4.92) можно использовать хорошо известное свойство этого класса задач чебышевской аппроксимации, описываемое следующей обобщенной теоремой Чебышева.

Теорема. Если $P(e^{j\omega})$ представляет собой линейную комбинацию из r косинусных функций, т. е.

$$P(e^{j\omega}) = \sum_{n=0}^{r-1} \alpha(n) \cos(\omega n),$$

то необходимое и достаточное условие того, что $P(e^{j\omega})$ является единственной и наилучшей аппроксимацией со взвешиванием непрерывной функции $D(e^{j\omega})$ в компактной подобласти из области $(0, \pi)$, состоит в том, что взвешенная функция ошибки $E(e^{j\omega})$ имеет по крайней мере $(r+1)$ экстремальных частот в подобласти A , т. е. в этой подобласти должно существовать $(r+1)$ точек ω_i , таких, что $\omega_1 < \omega_2 < \dots < \omega_r < \omega_{r+1}$ и $E(e^{j\omega_i}) = -E(e^{j\omega_{i+1}})$, $i = 1, 2, \dots, r$, и $|E(e^{j\omega_i})| = \max_{\omega \in A} |E(e^{j\omega})|$.

Сформулированная выше обобщенная теорема Чебышева чрезвычайно важна, поскольку дает необходимые и достаточные условия для получения решения, оптимального в чебышевском смысле. В настоящее время на основе той или иной интерпретации этой теоремы разработан ряд методов получения оптимального решения. Наиболее эффективный метод, для которого созданы машинные программы, основан на итерационной процедуре, получившей название алгоритма многократной замены Ремеза (см. Ланнэ А.А. Оптимальный синтез линейных электронных схем. М., Связь, 1978.). Программа, реализующая этот метод, разработана Д.Мак-Клеланом и Т.Парксом и опубликована в [2,4,5].

Суть алгоритма замены Ремеза состоит в следующем. Теорема о чебышевском альтернансе утверждает, что в случае оптимального решения ошибка имеет по крайней мере $r+1$ экстремумов. Пусть множество экстремальных частот $\{\omega_k\}$ $k=0, 1, \dots, r$ в области аппроксимации таково, что можно записать следующую систему соотношений:

$$\hat{W}(e^{j\omega_k})[\hat{D}(e^{j\omega_k}) - P(e^{j\omega_k})] = (-1)^k \delta,$$

состоящую из $r+1$ линейных уравнений с $r+1$ неизвестными, из которых r переменных являются коэффициентами $\alpha(n)$ аппроксимирующей функции $P(e^{j\omega})$, а $r+1$ -я переменная есть неизвестная ошибка δ . После нахождения неизвестных, анализируют ошибку $E(e^{j\omega})$, и если $|E(e^{j\omega})| \leq \delta$ во всей области аппроксимации, полученное решение является оптимальным. Если в некоторых точках $|E(e^{j\omega})| \geq \delta$, то

выбирают новое множество экстремальных частот путем рассмотрения $r+1$ точек, где ошибка максимальная и имеет чередующийся знак. Можно показать, что в этой процедуре δ на каждом шаге возрастает и в конце концов сходится к своей верхней грани, которая и является решением задачи. Начальное множество частот можно выбирать произвольно, например, эквидистантно. Однако в таком виде процедура не очень эффективна, поскольку на каждом шаге приходится решать систему линейных уравнений.

В действительности нет необходимости решать систему, если оценить δ по формуле

$$\delta = \frac{a_0 \hat{D}(e^{j\omega_0}) + a_1 \hat{D}(e^{j\omega_1}) + \dots + a_r \hat{D}(e^{j\omega_r})}{a_0 / \hat{W}(e^{j\omega_0}) - a_1 / \hat{W}(e^{j\omega_1}) + \dots + (-1)^r a_r / \hat{W}(e^{j\omega_r})},$$

$$\text{где } a_k = \prod_{\substack{i=0 \\ i \neq k}}^r \frac{1}{(\cos \omega_k - \cos \omega_i)}.$$

После вычисления δ значения $P(e^{j\omega})$ в r точках $\omega_0, \omega_1, \dots, \omega_{r-1}$ можно вычислить по формуле

$$P(e^{j\omega_k}) = \hat{D}(e^{j\omega_k}) - (-1)^k \delta / \hat{W}(e^{j\omega_k}), k = 0, 1, \dots, r-1$$

и получить интерполяционную формулу для вычисления $P(e^{j\omega})$ и, следовательно, $E(e^{j\omega})$ на плотной сетке частот путем построения полинома, проходящего через r точек на основе интерполяционной формулы Лагранжа:

$$P(e^{j\omega}) = \frac{\sum_{k=0}^{r-1} [\beta_k / (\cos \omega - \cos \omega_k)] P(e^{j\omega_k})}{\sum_{k=0}^{r-1} [\beta_k / (\cos \omega - \cos \omega_k)]},$$

$$\text{где } \beta_k = \prod_{\substack{i=0 \\ i \neq k}}^{r-1} \frac{1}{(\cos \omega_k - \cos \omega_i)}.$$

Последний этап проектирования оптимальных фильтров, заключающийся в расчете импульсной характеристики, сводится к вычислению отсчетов $P(e^{j\omega})$ на 2^M равноотстоящих частотах ($2^M \geq N$) и использованию ДПФ для получения последовательности коэффициентов $\alpha(n)$, по которым находятся коэффициенты импульсной характеристики для каждого из четырех видов фильтров с линейной фазой.

Расчет оптимальных по Чебышеву КИХ-фильтров можно выполнить методами линейного программирования. В оптимальном КИХ-фильтре с линейной фазой максимум ошибки аппроксимации $E(e^{j\omega})$ минимизируется на всех частотах ω . Обозначив максимальную ошибку через δ , можно записать систему линейных неравенств, описывающих эту минимаксную задачу:

$$-\delta \leq \hat{W}(e^{j\omega_k}) [\hat{D}(e^{j\omega_k}) - P(e^{j\omega_k})] \leq \delta, \omega_k \in F \quad (4.93)$$

Здесь F — густая сетка частот в полосах, где выполняется аппроксимация.

Поскольку $P(e^{j\omega}) = \sum_{n=0}^{r-1} \alpha(n) \cos(\omega n)$, является линейной комбинацией r косинусоидальных функций, неравенства (4.93) можно формально записать в виде задачи линейного программирования: минимизировать δ при условиях

$$\left. \begin{aligned} -\hat{W}(e^{j\omega_k}) \sum_{n=0}^{r-1} \alpha(n) \cos(\omega_k n) - \delta &\leq -\hat{W}(e^{j\omega_k}) \hat{D}(e^{j\omega_k}) \\ \hat{W}(e^{j\omega_k}) \sum_{n=0}^{r-1} \alpha(n) \cos(\omega_k n) - \delta &\leq \hat{W}(e^{j\omega_k}) \hat{D}(e^{j\omega_k}) \end{aligned} \right\} \omega_k \in F.$$

Для решения этой системы неравенств можно использовать методы линейного

программирования. Но так как метод линейного программирования весьма близок к методу однократной замены и в то же время значительно более громоздок по сравнению с методом Ремеза (который является методом многократной замены), то для рассматриваемого класса задач он почти не используется. С другой стороны, в случаях, когда приходится учитывать ограничения и во временной области, именно линейное программирование может оказаться единственным простым методом решения задачи проектирования.

Приведем некоторые результаты расчетов оптимальных фильтров. На рис.4.61 приведены частотная (а) и импульсная (б) характеристики фильтра нижних частот с параметрами: длина импульсной характеристики $N = 99$, граничные частоты полосы пропускания и непропускания $f_p = 0.0808$, $f_a = 0.1111$ соответственно, и равных весовых коэффициентах в обеих полосах. На рис.4.62 приведены частотная (а) и импульсная (б) характеристики полосно-пропускающего фильтра с параметрами: длина импульсной характеристики $N = 32$, граничные частоты: 0 и 0,1 для нижней полосы непропускания, 0,2 и 0,35 для полосы пропускания, 0,425 и 0,5 для верхней полосы непропускания, весовой коэффициент для полосы непропускания равен 10, а для полосы пропускания - 1. Максимум ошибки равен 0,00151 (54дБ) в полосе непропускания и 0,0151 в полосе пропускания. На рис.4.63 приведены частотная (а) и импульсная (б) характеристики полосно-заграждающего фильтра с параметрами: длина импульсной характеристики $N = 31$, граничные частоты: 0 и 0,1 для нижней полосы пропускания, 0,15 и 0,36 для полосы непропускания, 0,41 и 0,5 для верхней полосы пропускания, весовой коэффициент для полосы непропускания равен 50, а для полосы пропускания - 1. Максимум ошибки равен 0,144 в полосе пропускания и 0,00288 (50,8дБ) в полосе непропускания. На рис.4.64 приведены частотная (а) и импульсная (б) характеристики полосно-пропускающего фильтра ($N = 55$) с двумя полосами пропускания и с тремя полосами непропускания граничные частоты: 0 и 0,05 для нижней полосы непропускания (весовой коэффициент 10), 0,1 и 0,15 для первой полосы пропускания (весовой коэффициент 1), 0,18 и 0,25 для средней полосы непропускания (весовой коэффициент 3), 0,3 и 0,36 для второй полосы пропускания (весовой коэффициент 1), 0,41 и 0,5 для верхней полосы непропускания (весовой коэффициент 20). Максимум ошибки в полосах пропускания равен 0,034, в полосах непропускания соответственно - 49,3дБ, -38,8дБ, -55,3дБ. На рис.4.65 приведены частотная (а) и импульсная (б) характеристики преобразователя Гильберта с параметрами: длина импульсной характеристики $N = 20$, граничные частоты 0,05 и 0,5. Максимум ошибки равен 0,0205.

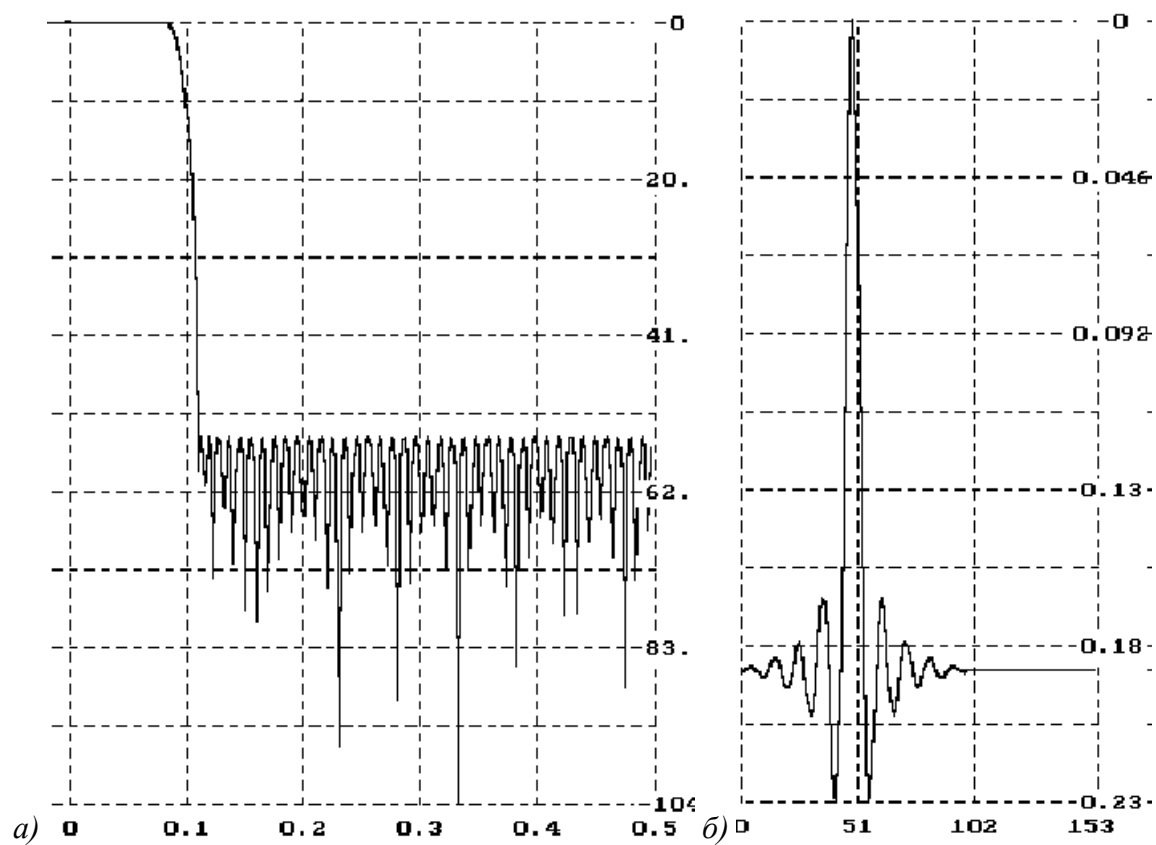


Рис.4.61. Частотная (а) и импульсная (б) характеристики фильтра нижних частот.

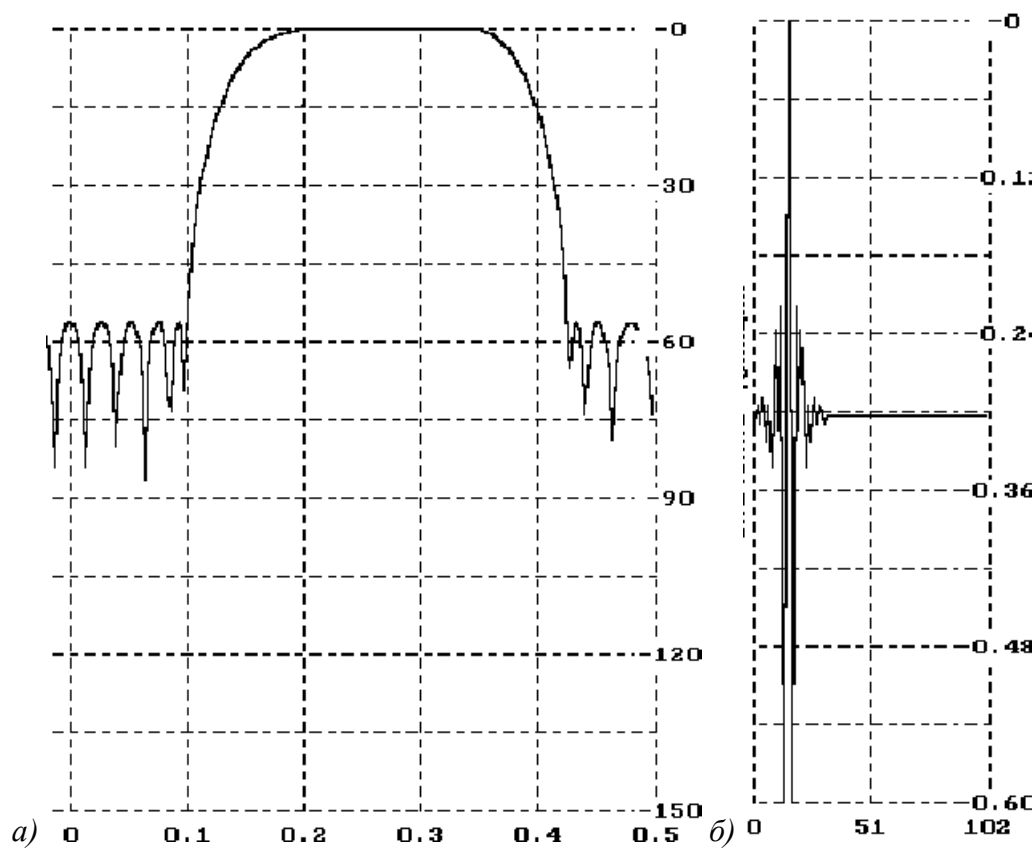


Рис.4.62. Частотная (а) и импульсная (б) характеристики полосно-пропускающего фильтра.

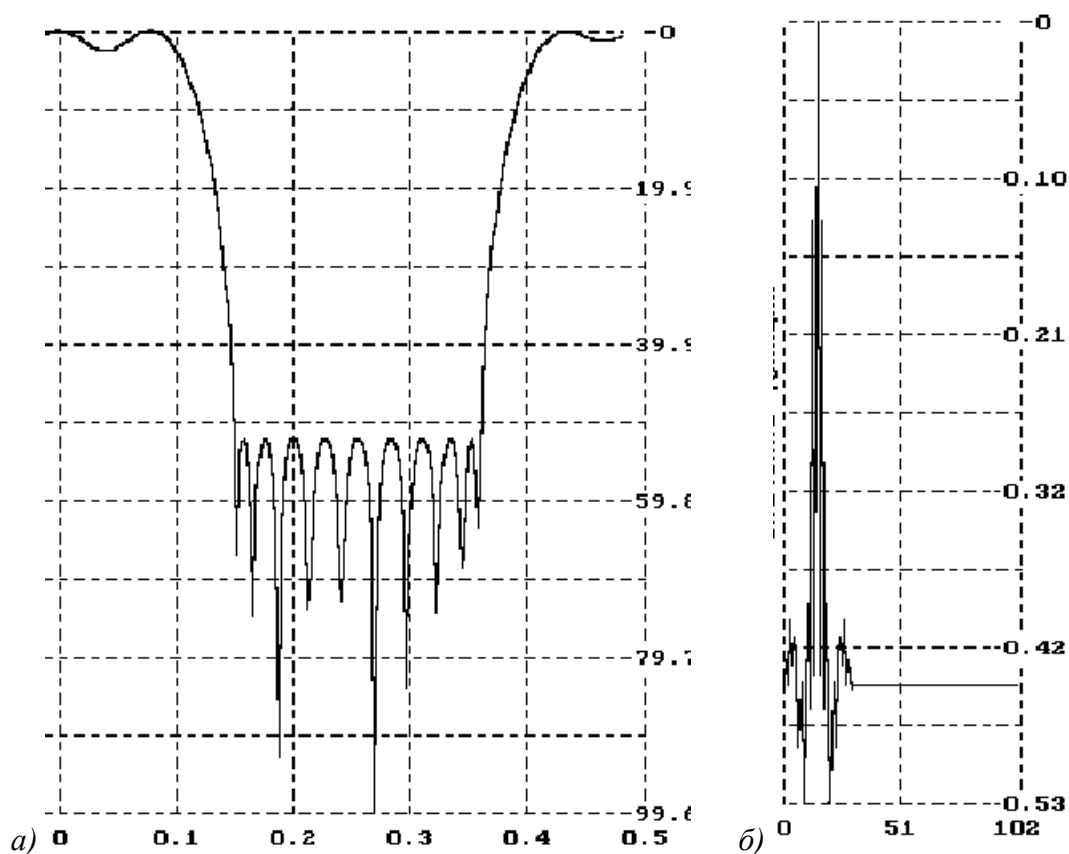


Рис.4.63. Частотная (а) и импульсная (б) характеристики полосно-заграждающего фильтра.

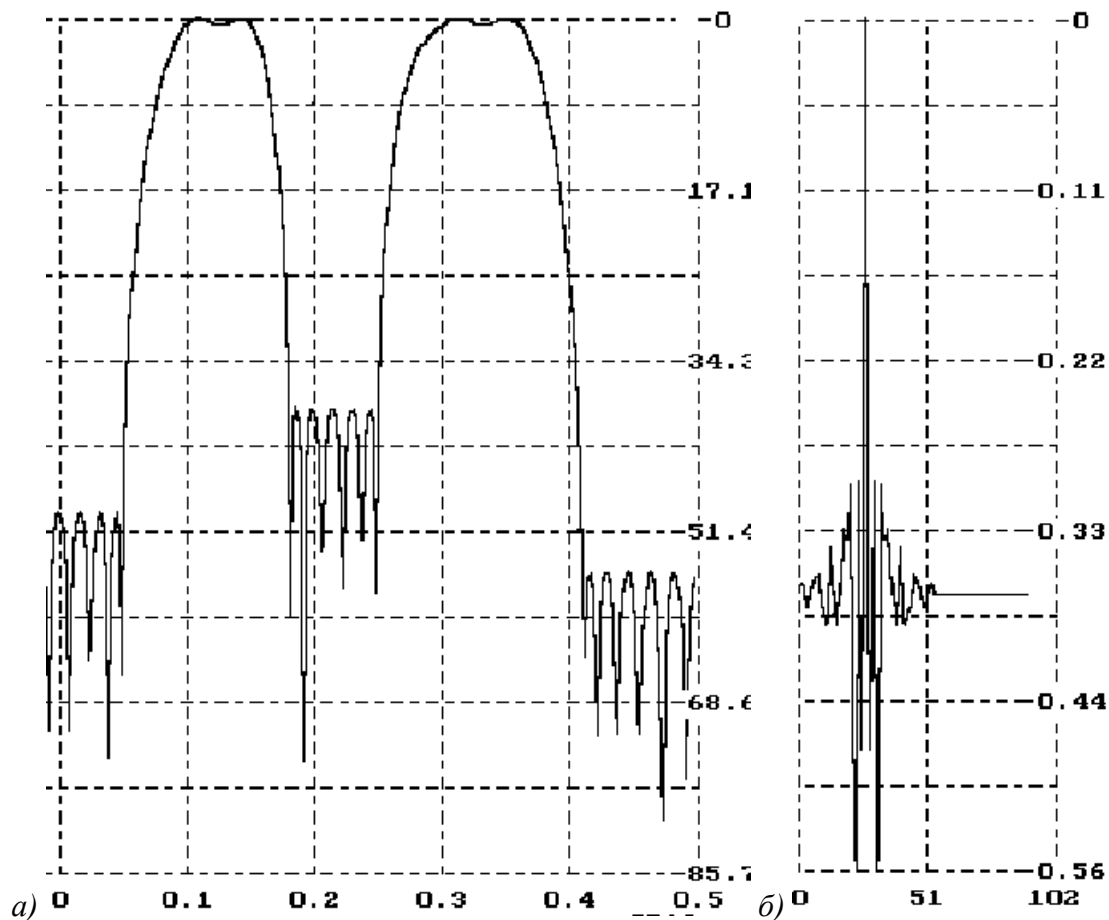


Рис.4.64. Частотная (а) и импульсная (б) характеристики многополосного фильтра.

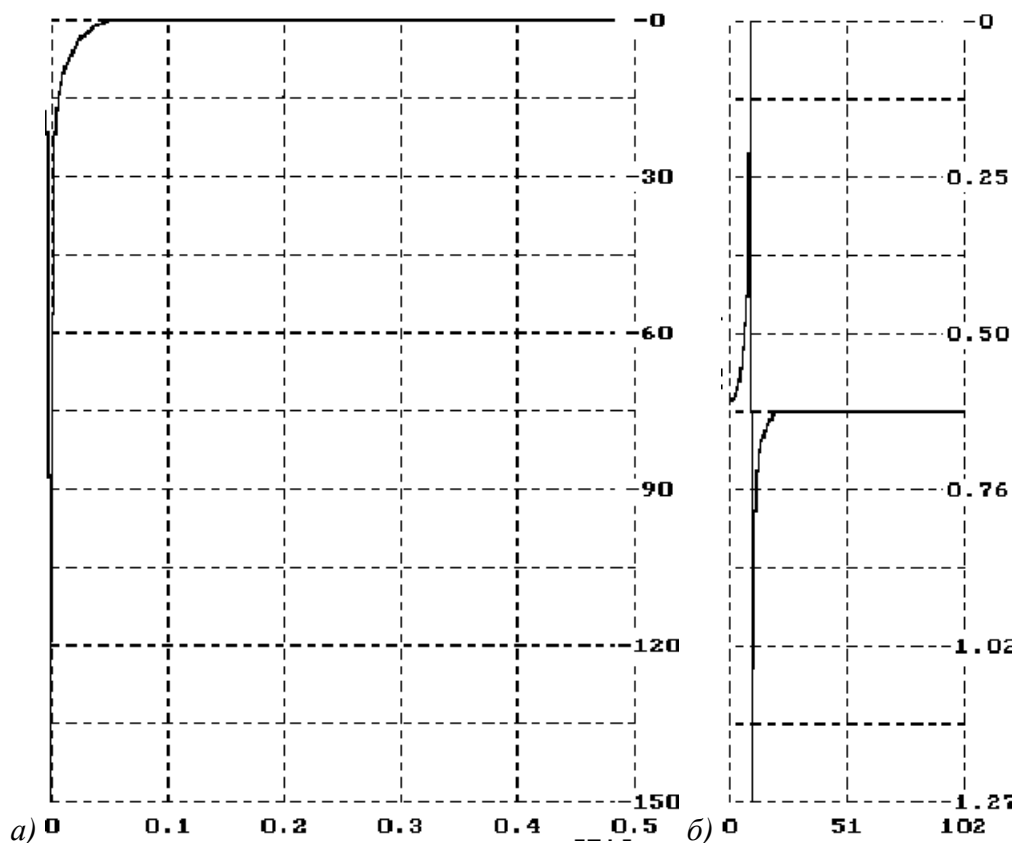


Рис.4.65. Частотная (а) и импульсная (б) характеристики преобразователя Гильберта.

4.6. Сравнение цифровых БИХ- и КИХ-фильтров

Выше были рассмотрены методы расчета линейных инвариантных к сдвигу цифровых фильтров, обсужден широкий диапазон методов расчета фильтров с импульсными характеристиками как конечной, так и бесконечной длины. Естественно возникающими вопросами являются: какой тип системы является лучшим — БИХ- или КИХ-тип? Почему дается так много методов расчета? Какой метод дает наилучшие результаты? Ответом на эти вопросы является обсуждение большого числа различных методов расчета как БИХ-, так и КИХ-фильтров, учитывая, что ни один из типов фильтра, ни один из методов расчета не является наилучшим при всех обстоятельствах.

Выбор КИХ- или БИХ-фильтра зависит от достоинств и недостатков фильтра каждого типа. Например, БИХ-фильтры обладают тем достоинством, что множество частотно-избирательных фильтров может быть рассчитано при использовании расчетных формул замкнутой формы. Это значит, что сначала определяются требования, соответствующие данному типу фильтра (например, Баттерворта, Чебышева или эллиптического), а затем коэффициенты (полюсы и нули) нужного цифрового фильтра получаются путем прямой подстановки в систему расчетных уравнений. Этот вид простоты расчетной процедуры является эффективным в том случае, если лишь несколько фильтров подлежат расчету или если в распоряжении имеются вычислительные средства, обладающие ограниченными возможностями.

В случае КИХ-фильтров расчетные уравнения замкнутой формы не существуют. Несмотря на то что метод окна может применяться довольно прямым путем, некоторая итерация может быть необходима для удовлетворения заданным требованиям. Большинство других методов расчета КИХ-фильтров являются итерационными процедурами, требующими для своего выполнения довольно мощных вычислительных средств. В противоположность этому зачастую возможно произвести расчет частотно-избирательного цифрового БИХ-фильтра, используя лишь ручной калькулятор и таблицы рассчитанных параметров аналогового фильтра. Однако из-за простоты такой

расчетной процедуры происходит уменьшение гибкости получаемой частотной характеристики фильтра. Расчеты БИХ-фильтров на основе замкнутых формул имеют ограниченное применение и предназначены прежде всего для расчета фильтров нижних и верхних частот, полосовых и других. Кроме того, в этих расчетах, как правило, не учитывается фазовая характеристика фильтра. Таким образом, несмотря на то что мы можем получить эллиптический фильтр нижних частот с прекрасными амплитудными характеристиками с помощью относительно простой расчетной процедуры, его фазовая характеристика будет существенно нелинейной (особенно на границе полосы).

В противоположность этому КИХ-фильтры могут иметь строго линейную фазовую характеристику. К тому же метод окна и большинство из методов оптимизации обеспечивают возможность аппроксимации более произвольных частотных характеристик при несколько большей сложности, чем те, которые встречаются при расчете фильтров нижних частот. Процедура расчета КИХ-фильтров в большей мере поддается управлению, чем процедура расчета БИХ-фильтров, так как для КИХ-фильтров существует теорема оптимальности, которая оказывается очень важной в широком диапазоне практических ситуаций.

В конечном счете существуют вопросы экономичности выполнения цифрового фильтра. Экономические показатели обычно оцениваются с помощью аппаратной сложности или скорости вычисления. Оба этих фактора в большей или меньшей степени прямо связаны с порядком фильтра, необходимым для удовлетворения заданным требованиям. Если не учитывать фазовую характеристику, то, как правило, заданные требования к амплитудной характеристике будут обеспечиваться более эффективно БИХ-фильтром. Однако во многих случаях линейная фазовая характеристика, характерная для КИХ-фильтра, может быть получена в результате дополнительных затрат. Таким образом должно быть рассмотрено множество компромиссных вариантов при проектировании цифрового фильтра. Очевидно, что окончательный выбор чаще будет производиться на основании инженерного опыта в таких вопросах, как формулирование требований, методов выполнения и вычислительных средств, доступных для расчета.

5. Цифровой спектральный анализ (ЦСА)

Спектральный анализ - это один из методов обработки сигналов, который позволяет охарактеризовать частотный состав измеряемого сигнала. Преобразование Фурье является математической основой, которая связывает временной или пространственный сигнал (или же некоторую модель этого сигнала) с его представлением в частотной области. Существенный вклад в развитие цифровых методов спектрального анализа внесли эффективные алгоритмы, предназначенные для вычисления дискретного преобразования Фурье, предложенные Д.Кули и Д.Тьюки в 1965 году. Набор алгоритмов, называемых алгоритмами быстрого преобразования Фурье (БПФ), включает разнообразные методы уменьшения времени вычисления дискретного преобразования Фурье (ДПФ). Поскольку вычисление ДПФ является основной операцией в большинстве задач спектрального анализа, то использование БПФ в некоторых встречающихся на практике случаях, позволяющее ускорить вычисление ДПФ в 100 и более раз по сравнению с методом прямого вычисления ДПФ, имеет чрезвычайно важное значение и должно рассматриваться как неотъемлемая часть применения методов цифровой обработки сигналов для спектрального анализа. Возможно, именно алгоритмы БПФ более чем какие-либо другие методы существенно расширили область применения методов спектрального анализа как средства обработки сигналов. Поэтому начнем рассмотрение вопросов ЦСА с алгоритмов БПФ, включающих алгоритмы с основанием 2 и прореживанием по времени и по частоте.

5.1. Введение в алгоритмы БПФ с основанием 2

Напомним, что прямое ДПФ конечной последовательности $x(n)$, $0 \leq n \leq N-1$ определяется выражением

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j \frac{2\pi}{N} kn}, k = \overline{0, N-1}$$

или в более удобном виде как

$$X(k) = \sum_{n=0}^{N-1} x(n) W_N^{kn}, k = \overline{0, N-1}, \quad (5.1)$$

где $W_N = e^{-j(2\pi/N)}$,

а обратное ДПФ имеет вид

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) W_N^{-kn}, n = \overline{0, N-1}. \quad (5.2)$$

Выражения (5.1) и (5.2) различаются только знаком экспоненты от W_N и коэффициентом $1/N$, поэтому рассуждения, касающиеся вычислительных процедур для (5.1), применимы с очевидными изменениями к (5.2).

Из соотношения (5.1) следует, что в случае, когда последовательность $x(n)$ является комплексной, при прямом вычислении N -точечного ДПФ нужно выполнить $(N-1)^2$ комплексных умножений и $N(N-1)$ комплексных сложений. Таким образом, для достаточно больших N (порядка 1000) прямое вычисление ДПФ требует выполнения чрезмерного количества вычислительных операций. Основная идея БПФ состоит в том, чтобы разбить исходную N -точечную последовательность на две более короткие последовательности, ДПФ которых могут быть скомбинированы таким образом, чтобы получилось ДПФ исходной N -точечной последовательности. Так, например, если N четное, а исходная N -точечная последовательность разбита на две $(N/2)$ -точечные последовательности, то для вычисления искомого N -точечного ДПФ потребуется порядка $(N/2)^2 \cdot 2 = N^2/2$ комплексных умножений, т. е. вдвое меньше по сравнению с прямым вычислением. Здесь множитель $(N/2)^2$ дает число умножений, необходимое для прямого вычисления $(N/2)$ -точечного ДПФ, а множитель 2 соответствует двум ДПФ, которые должны быть вычислены. Эту операцию можно повторить, вычисляя вместо $(N/2)$ -

точечного ДПФ два $(N/4)$ -точечных ДПФ (предполагая, что $N/2$ четное) и сокращая тем самым объем вычислений еще в два раза. Выигрыш в два раза является приближенным, поскольку не учитывается, каким образом из ДПФ меньшего размера образуется искомое N -точечное ДПФ.

Проиллюстрируем описанную методику для N -точечной последовательности $\{x(n)\}$, считая, что N равно степени 2. Введем две $(N/2)$ -последовательности $\{x_1(n)\}$ и $\{x_2(n)\}$ из четных и нечетных членов $x(n)$ соответственно, т. е.

$$\begin{aligned} x_1(n) &= x(2n), \quad n = 0, 1, \dots, (N/2) - 1, \\ x_2(n) &= x(2n + 1), \quad n = 0, 1, \dots, (N/2) - 1. \end{aligned} \quad (5.3)$$

N -точечное ДПФ последовательности $\{x(n)\}$ можно записать как

$$\begin{aligned} X(k) &= \sum_{n=0}^{N-1} x(n) W_N^{kn} = \sum_{n=0}^{N/2-1} x(2n) W_N^{2nk} + \sum_{n=0}^{N/2-1} x(2n+1) W_N^{(2n+1)k} = \\ &= \sum_{n=0}^{N/2-1} x_1(n) W_{N/2}^{kn} + W_N^k \sum_{n=0}^{N/2-1} x_2(n) W_{N/2}^{kn} = X_1(k) + W_N^k X_2(k), \end{aligned} \quad (5.4)$$

где учтено, что $W_N^2 = [e^{j(2\pi/N)}]^2 = e^{j[2\pi/(N/2)]} = W_{N/2}$,

и введены обозначения для $(N/2)$ -точечных ДПФ последовательностей $x_1(n)$ и $x_2(n)$:

$$X_1(k) = \sum_{n=0}^{N/2-1} x_1(n) W_{N/2}^{nk}, \quad X_2(k) = \sum_{n=0}^{N/2-1} x_2(n) W_{N/2}^{nk}.$$

Из формулы (5.4) следует, что N -точечное ДПФ $X(k)$ может быть разложено на два $(N/2)$ -точечных ДПФ, результаты которых объединяются согласно (5.4). Если бы $(N/2)$ -точечные ДПФ вычислялись обычным способом, то для вычисления N -точечного ДПФ потребовалось бы, очевидно, $(N^2/2 + N)$ комплексных умножений. При больших N (когда $N^2/2 \gg N$) это позволяет сократить время вычисления на 50%. Поскольку $X(k)$ определено при $0 \leq k \leq N-1$, а $X_1(k)$ и $X_2(k)$ определены при $0 \leq k \leq N/2-1$, необходимо доопределить формулу (5.4) для $k \geq N/2$. Это определение достаточно очевидно и может быть записано следующим образом:

$$X(k + \frac{N}{2}) = X_1(k + \frac{N}{2}) + W_N^{k+N/2} X_2(k + \frac{N}{2}) = X_1(k) - W_N^k X_2(k), \quad k = \overline{0, (N/2) - 1}, \quad (5.5)$$

где учтено, что $W_N^{k+N/2} = e^{-j\frac{2\pi N}{N^2}} W_N^k = e^{-j\pi} W_N^k = -W_N^k$.

Таким образом, вычисление $X(k)$ по $X_1(k)$ и $X_2(k)$ можно представить в виде

$$\begin{aligned} X(k) &= X_1(k) + W_N^k X_2(k), \\ X(k + \frac{N}{2}) &= X_1(k) - W_N^k X_2(k), \quad k = \overline{0, (N/2) - 1}. \end{aligned} \quad (5.6)$$

На рис.5.1 с помощью направленного графа представлена последовательность операций при вычислении восьмиточечного ДПФ с использованием двух четырехточечных ДПФ. Незачерненный кружок графа означает операцию сложения/вычитания, причем верхний выход соответствует сумме, а нижний – разности. Стрелка обозначает операцию умножения на значение множителя a , указанного над стрелкой. Входная последовательность $x(n)$ сначала разбивается на две последовательности $x_1(n)$ и $x_2(n)$ из четных и нечетных членов $x(n)$, после чего рассчитываются их преобразования $X_1(k)$ и $X_2(k)$. Затем в соответствии с формулой (5.6) получают $X(k)$.

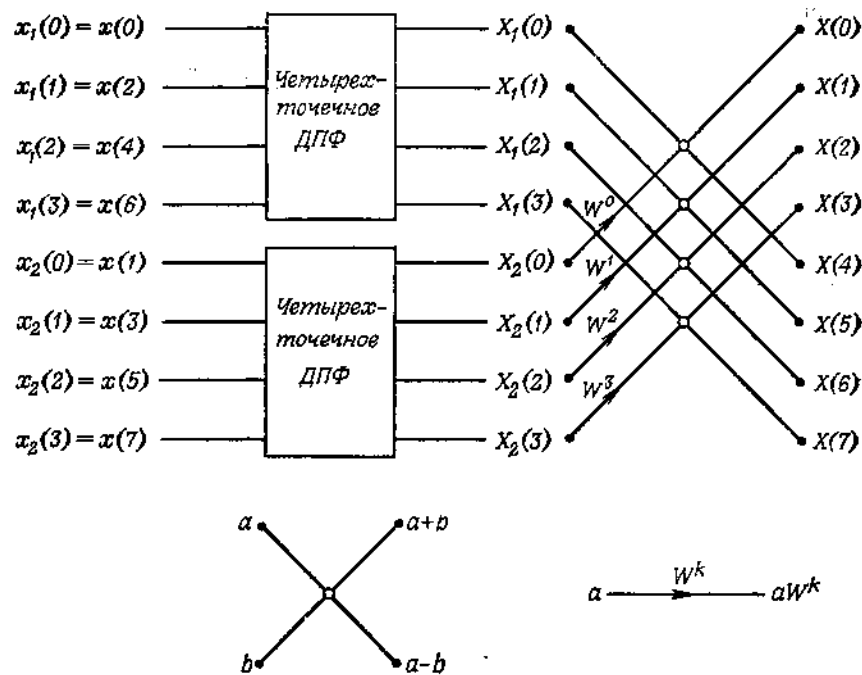


Рис.5.1. Вычисление восьмиточечного ДПФ через два четырехточечных ДПФ.

Выражение (5.6) соответствует разбиению исходного N -точечного вычисления ДПФ на два $N/2$ -точечных вычисления. Если N – четное число, что имеет место всегда, когда N равно степени 2, то можно вычислять каждое $N/2$ -точечное ДПФ в (5.6) путем разбиения сумм на два $N/4$ -точечных ДПФ, которые затем объединяются, давая $N/2$ -точечное ДПФ. Каждая из последовательностей $x_1(n)$ и $x_2(n)$ разбивается на две последовательности, состоящие из четных и нечетных членов. Аналогично $N/2$ -точечные ДПФ могут быть записаны как комбинации двух $N/4$ -точечных ДПФ, т. е.

$$\begin{aligned} X_1(k) &= A(k) + W_{N/2}^k B(k), \\ X_1(k + N/2) &= A(k) - W_{N/2}^k B(k); \\ X_2(k) &= C(k) + W_{N/2}^k D(k), \\ X_2(k + N/2) &= C(k) - W_{N/2}^k D(k); \end{aligned} \quad (5.7)$$

или

$$\begin{aligned} X_1(k) &= A(k) + W_N^{2k} B(k), \\ X_1(k + N/4) &= A(k) - W_N^{2k} B(k); \\ X_2(k) &= C(k) + W_N^{2k} D(k), \\ X_2(k + N/4) &= C(k) - W_N^{2k} D(k); \end{aligned} \quad (5.8)$$

где $0 < k < N/4 - 1$, $A(k)$ и $B(k)$ – $N/4$ -точечные ДПФ соответственно четных и нечетных членов $x_1(n)$, $C(k)$ и $D(k)$ – $N/4$ -точечные ДПФ соответственно четных и нечетных членов $x_2(n)$. На рис.5.2 показан результирующий направленный граф, в котором четырехточечные ДПФ из рис.5.1 рассчитываются согласно (5.8).

Процесс уменьшения размера ДПФ от L до $L/2$, где L равно степени 2, может быть продолжен до тех пор, пока не останутся только двухточечные ДПФ. Двухточечное ДПФ, например, $A(k)$, $k = 0, 1$, может быть рассчитано без использования умножений по формулам

$$\begin{aligned} A(0) &= a(0) + W_2^0 a(1) = a(0) + W_8^0 a(1) = a(0) + a(1), \\ A(1) &= a(0) + W_2^1 a(1) = a(0) + W_8^4 a(1) = a(0) - a(1). \end{aligned} \quad (5.9)$$

Таким образом, восьмиточечное ДПФ (рис.5.1 и 5.2) в итоге сводится к алгоритму, описываемому направленным графом, представленным на рис.5.3.

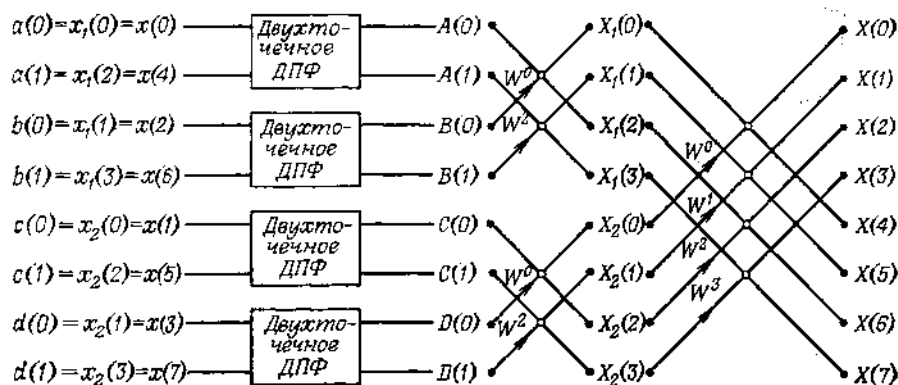


Рис.5.2. Вычисление восьмиточечного ДПФ через два четырехточечных ДПФ, которые в свою очередь вычисляются через четыре двухточечных ДПФ.

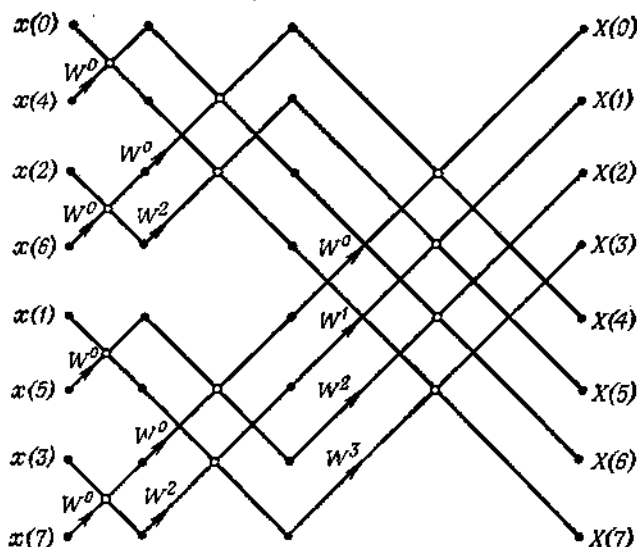


Рис. 5.3. Восьмиточечное ДПФ, полученное последовательным прореживанием в 2 раза.

Анализ графа на рис. 5.3 и процедуры последовательного сокращения вдвое размеров преобразований показывает, что на каждом этапе БПФ (т. е. при каждом сокращении размеров ДПФ) необходимо выполнить $N/2$ комплексных умножений. Поскольку общее количество этапов равно $\log_2 N$, то число комплексных умножений, необходимое для нахождения N -точечного ДПФ, приблизительно равно $N/2 \log_2 N$. Слово *приблизительно* использовано по той причине, что умножения на W_N^0 , $W_N^{N/2}$, $W_N^{N/4}$ и $W_N^{3N/4}$ в действительности сводятся просто к сложениям и вычитаниям комплексных чисел. Так, например, на рис.5.3 первый этап БПФ содержит только сложения и вычитания комплексных чисел. Даже на втором этапе используются только сложения и вычитания комплексных чисел. Фактически, как следует из направленного графа на рис.5.3, вместо ожидаемых 12 (т. е. $4 \log_2 8$) достаточно выполнить всего два нетривиальных умножения. Однако для больших значений N фактическое число нетривиальных умножений хорошо аппроксимируется выражением $N/2 \log_2 N$.

Описанный выше алгоритм был назван алгоритмом с прореживанием по времени, поскольку на каждом этапе входная (т.е. временная) последовательность разделяется на две обрабатываемые последовательности меньшей длины, т. е. входная последовательность прореживается на каждом этапе. Другая форма алгоритма БПФ (с прореживанием по частоте) будет описана ниже, а сейчас целесообразно обсудить некоторые общие свойства алгоритмов БПФ.

Базовая операция алгоритма с прореживанием по времени (так называемая «бабочка») состоит в том, что два входных числа A и B объединяются для получения

двух выходных чисел X и Y следующим образом:

$$\begin{aligned} X &= A + W_N^k B, \\ Y &= A - W_N^k B. \end{aligned} \quad (5.10)$$

На рис.5.4 изображен направленный граф базовой операции (5.10).

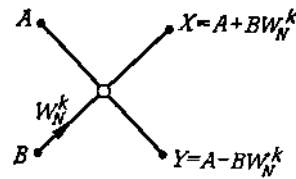


Рис.5.4. Базовая операция алгоритма БПФ.

Внимательное рассмотрение направленного графа на рис.5.3 показывает, что каждый из этапов содержит $N/2$ базовых операций. В случае когда множитель W_N^k нетривиальный, для каждой базовой операции необходимо выполнить только одно умножение, поскольку величину BW_N^k можно вычислить и запомнить. Таким образом, структура базовых операций такова, что для выполнения БПФ N -точечной последовательности, размещенной в памяти, достаточно иметь лишь одну дополнительную ячейку памяти. Результаты всех промежуточных этапов БПФ можно размещать в те же ячейки памяти, где находились исходные данные. Поэтому для хранения и входной, и выходной последовательностей можно использовать один и тот же массив ячеек памяти. Алгоритм, в котором для размещения входной и выходной последовательностей используются одни и те же ячейки памяти, называется алгоритмом БПФ с замещением.

Еще одной особенностью алгоритма с прореживанием по времени (как, впрочем, и большинства других алгоритмов БПФ) является необходимость такой перестановки элементов входной последовательности, чтобы выходная последовательность $X(k)$ имела естественный (прямой) порядок расположения, т. е. $k = 0, 1, \dots, N-1$. В примере на рис.5.3 для этого требовался следующий порядок размещения входной последовательности: $x(0)$, $x(4)$, $x(2)$, $x(6)$, $x(1)$, $x(5)$, $x(3)$ и $x(7)$. Характер перестановки элементов входной последовательности может быть описан сравнительно просто. Ниже будет показано, что в случае, когда N является степенью 2, входная последовательность должна быть расположена в памяти в двоично-инверсном порядке для того, чтобы выходная последовательность получалась в прямом порядке. Двоично-инверсный порядок определяется следующим образом. Если записать порядковые номера элементов входной последовательности в двоичном коде, используя L двоичных разрядов, причем $N = 2^L$, а затем инвертировать порядок следования разрядов, то получаемые при этом числа и будут номерами элементов входной последовательности после их перестановки.

Таблица 5.1

Номер	Двоичное представление	Двоичная инверсия	Двоично-инверсный номер
0	000	000	0
1	001	100	4
2	010	010	2
3	011	110	6
4	100	001	1
5	101	101	5
6	110	011	3
7	111	111	7

Так, для случая $N = 8 = 2^3$ прямой порядок номеров приведен в табл. 6.1 слева, а двоично-инверсный порядок - справа. Таким образом, для двоичной инверсии входной последовательности необходим соответствующий алгоритм. Простой и эффективный

двоично-инверсный счетчик предложен Рейдером [2]

Из сказанного выше ясно, что перестановку входной последовательности можно произвести с замещением, меняя в парах местами числа с прямым и двоично-инверсным номерами и используя для этого лишь одну вспомогательную ячейку памяти. На рис.5.5 показана схема перестановки данных, представленных в табл. 5.1.

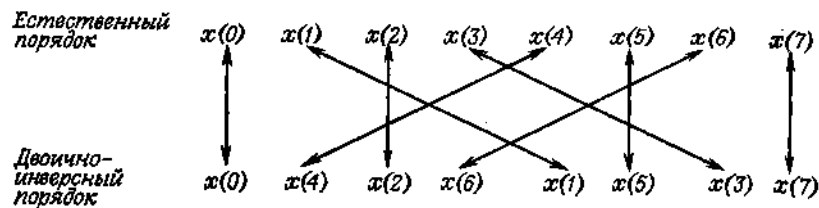


Рис. 5.5. Перестановка данных с замещением.

Простая, составленная на ФОРТРАНе, программа расчета БПФ с прореживанием по времени (авторы ее Кули, Льюис, Уэлч) [2] приведена на рис.5.6.

```

SUBROUTINE FFT(A,M,N)
COMPLEX A(N),U,W,T
N=2**M
NV2=N/2
NM1=N-1
J=1
DO 7 I=1,NM1
IF(I .GE. J) GO TO 5
T=A(J)
A(J)=A(I)
A(I)=T
5 K=NV2
6 IF(K .GE. J) GO TO 7
J=J-K
K=K/2
GO TO 6
7 J=J+K
PI=3.141592653589793
DO 20 L=1,M
LE=2**L
LE1=LE/2
U=(1.0,0.)
W=CMPLX(COS(PI/LE1),SIN(PI/LE1))
DO 20 J=1,LE1
DO 10 I=J,N,LE
IP=I+LE1
T=A(IP)*U
A(IP)=A(I)-T
10 A(I)=A(I)+T
20 U=U*W
RETURN
END

```

Рис.5.6. Программа БПФ с основанием 2 и прореживанием по времени

Входная комплексная последовательность представляет собой массив A , размером $N=2^M$, причем M указывается в операторе обращения к подпрограмме. Все операторы цикла DO 7 предназначены для выполнения двоично-инверсной перестановки элементов входного массива. Остальные операторы используются для вычисления непосредственно БПФ и образуют три вложенных цикла. С помощью первого (внешнего) цикла выполняется M этапов, другой цикл предназначен для выполнения

базовых операций в пределах каждого этапа, а третий цикл (внутренний) необходим для вычисления степеней W , используемых при выполнении базовых операций в пределах одного этапа. Для вычисления степеней W используется рекуррентное соотношение $W_N^k = (W_N^{k-L})W_N^L$, причем синусы и косинусы непосредственно вычисляются лишь для нахождения W по приращениям.

5.2. Алгоритм БПФ с прореживанием по частоте

Другая распространенная форма алгоритма БПФ (при условии, что N равно степени 2) - так называемый алгоритм БПФ с прореживанием по частоте. В этом варианте алгоритма БПФ входная последовательность $\{x(n)\}$ разбивается на две последовательности, содержащие по $N/2$ отсчетов каждая следующим образом: первая последовательность $\{x_1(n)\}$ состоит из первых $(N/2)$ отсчетов $\{x(n)\}$, а вторая $\{x_2(n)\}$ — из остальных $(N/2)$ отсчетов $\{x(n)\}$, т. е.

$$\begin{aligned} x_1(n) &= x(n), \quad n = 0, 1, \dots, (N/2) - 1, \\ x_2(n) &= x(n + N/2), \quad n = 0, 1, \dots, (N/2) - 1. \end{aligned} \quad (5.11)$$

При таком разбиении N -точечное ДПФ последовательности $x(n)$ можно записать в виде

$$\begin{aligned} X(k) &= \sum_{n=0}^{N/2} x(n)W_N^{nk} + \sum_{n=N/2}^{N-1} x(n)W_N^{nk} = \sum_{n=0}^{N/2-1} x_1(n)W_N^{nk} + \sum_{n=0}^{N/2-1} x_2(n)W_N^{(n+\frac{N}{2})k} = \\ &= \sum_{n=0}^{N/2-1} [x_1(n) + (-1)^k x_2(n)]W_N^{nk}, \end{aligned}$$

где учтено, что $W_N^{kN/2} = e^{-j\frac{2\pi N}{N2}k} = e^{-j\pi k} = (-1)^k$.

Запишем выражения отдельно для четных и нечетных отсчетов ДПФ:

$$X(2k) = \sum_{n=0}^{N/2-1} [x_1(n) + x_2(n)](W_N^2)^{nk} = \sum_{n=0}^{N/2-1} [x_1(n) + x_2(n)]W_{N/2}^{nk}, \quad (5.12)$$

$$X(2k+1) = \sum_{n=0}^{N/2-1} [x_1(n) - x_2(n)]W_N^{n(2k+1)} = \sum_{n=0}^{N/2-1} \{[x_1(n) - x_2(n)]W_N^n\}W_{N/2}^{nk}. \quad (5.13)$$

Из выражений (5.12) и (5.13) видно, что четные и нечетные отсчеты ДПФ можно получить из $(N/2)$ -точечных ДПФ последовательностей $f(n)$ и $g(n)$, равных

$$\begin{aligned} f(n) &= x_1(n) + x_2(n), \quad n = 0, 1, \dots, N/2 - 1, \\ g(n) &= [x_1(n) - x_2(n)]W_N^n, \quad n = 0, 1, \dots, N/2 - 1. \end{aligned} \quad (5.14)$$

Таким образом, снова вычисление N -точечного ДПФ удалось свести к вычислению двух $(N/2)$ -точечных ДПФ. На фиг. 5.7 эта методика иллюстрируется для случая $N = 8$.

Описанную методику можно применить повторно и представить каждое из $(N/2)$ -точечных ДПФ в виде комбинации двух $(N/4)$ -точечных ДПФ. На фиг. 5.8 и 5.9 показан переход от четырехточечных ДПФ (рис.6.9) к двухточечным ДПФ с последующим прямым вычислением двухточечных ДПФ.

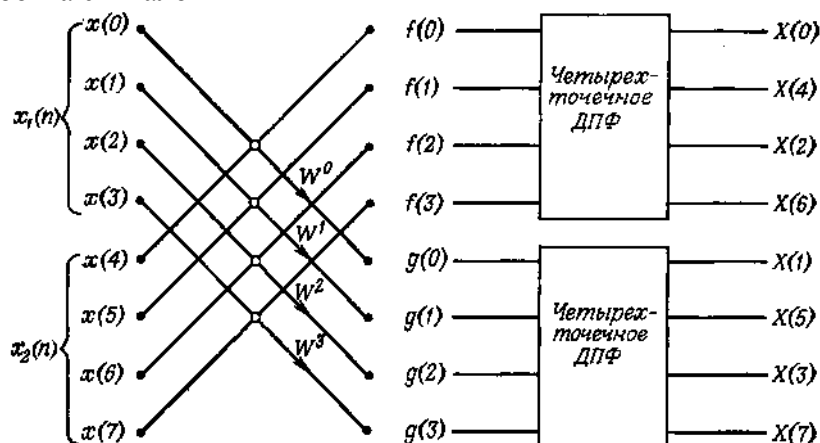


Рис.5.7. Переход от восьмиточечного ДПФ к двум четырехточечным ДПФ при прореживании по частоте.

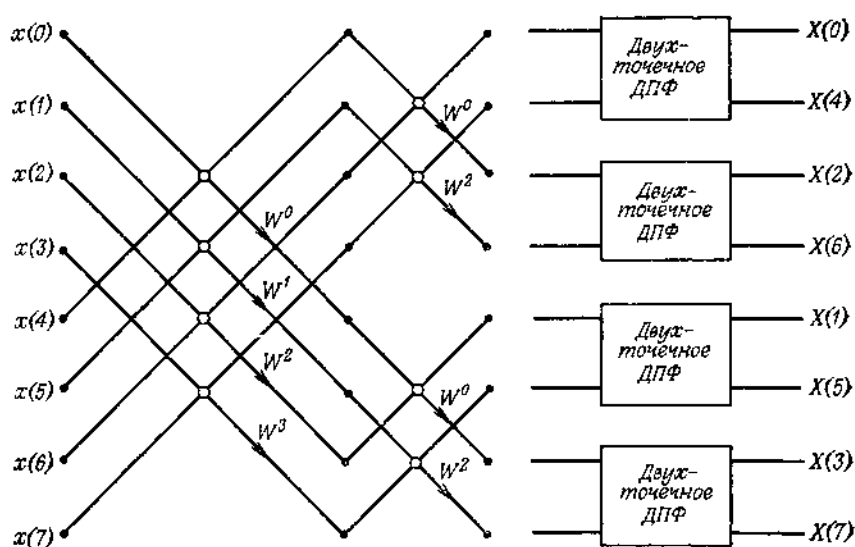


Рис.5.8. Переход от четырехточечных ДПФ на рис.5.7 к двухточечным ДПФ.

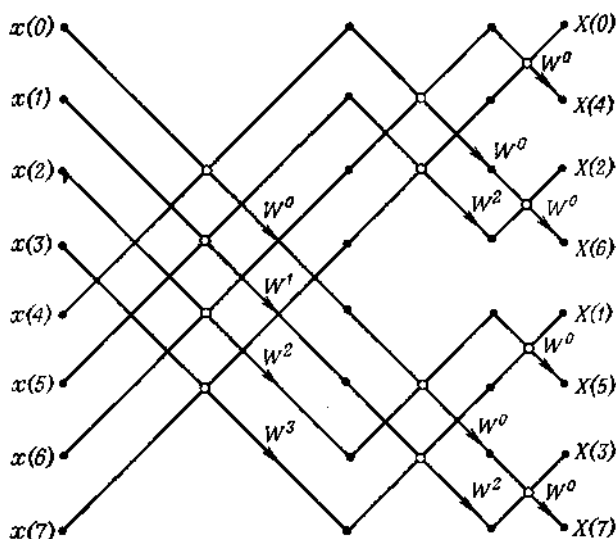


Рис.5.9. Полный направленный граф восьмиточечного ДПФ с замещением и прореживанием по частоте.

Сравнение алгоритмов, иллюстрированных на рис. 5.3 и 5.9, позволяет выявить два очевидных различия между ними. Во-первых, при прореживании по времени порядок следования входных отсчетов двоично-инверсный, а выходных — прямой и наоборот при прореживании по частоте (рис.5.9). Второе отличие заключается в

несколько иным выполнении базовой операции (см. рис.5.10 и 5.4): при прореживании по частоте комплексное умножение выполняется *после* сложения - вычитания.

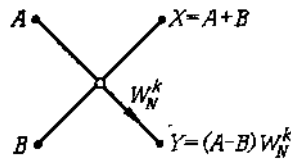


Рис.5.10. Базовая операция алгоритма БПФ с прореживанием по частоте.

Легко заметить и сходство между алгоритмами с прореживанием по времени и по частоте. В обоих случаях при вычислении ДПФ требуется около $N \log_2 N$ операций, вычисления могут быть проведены с замещением и должно быть предусмотрено выполнение двоичной инверсии. Отметим еще одно сходство между вычислениями, соответствующим базовым операциям («бабочкам»), и между рис.5.3 и 5.9 т.е. то, что рис.5.9 можно получить из рис.5.3 изменив направление стрелок (сигналов) и поменяв вход и выход.

Как уже отмечалось, для вычисления обратного ДПФ (5.2) можно использовать БПФ-алгоритм, если разделить результат на N и использовать вместо степеней W_N степени W_N^{-1} . В приложении 5.1 приведена программа на ФОРТРАНе, реализующая вычисления прямого и обратного ДПФ по основанию 2 с прореживанием по частоте.

5.3. Алгоритмы БПФ для составного значения N

Рассмотренные выше алгоритмы предполагали, что N является целой степенью 2, т.е. $N=2^L$. В более общем случае эффективное вычисление ДПФ связано с представлением N в виде сомножителей:

$$N = p_1 p_2 \dots p_L. \quad (5.15)$$

Если N представлено в виде произведения одинаковых сомножителей r , то соответствующий алгоритм называют алгоритмом БПФ по основанию r . Понятие «смешанное основание» означает, что не все сомножители N одинаковы. N -точечные алгоритмы по основанию 2 особенно просты для реализации, поэтому на практике выгодно всегда иметь дело с последовательностями длины $N=2^L$. Это можно сделать во многих случаях, просто дополняя последовательность конечной длины нулями, если это необходимо. В случаях, когда невозможно выбрать $N=2^L$, приходится рассматривать более общую ситуацию (5.15). Рассмотрим применение принципа прореживания по времени, когда N является произведением сомножителей не все из которых равны 2. Пусть $q_1 = p_2 p_3 \dots p_L$, так что $N = p_1 q_1$. Если N является степенью 2, можно выбрать $p_1 = 2$, а $q_1 = N/2$. Используя прореживание по времени, можно разложить $x(n)$ на две последовательности длины $N/2$, состоящие из четных и нечетных выборок соответственно, как это мы делали выше. Если $N = p_1 q_1$, то можно разделить входную последовательность на p_1 последовательностей длины q_1 , так, что каждая p_1 -я выборка попадает в одну последовательность. Например, если $p_1 = 3$, а $q_1 = 4$, так что $N = 12$, то можно разложить $x(n)$ на три последовательности длины 4, причем первая последовательность состоит из выборок $x(0), x(3), x(6), x(9)$, вторая – из $x(1), x(4), x(7), x(10)$, а третья – из $x(2), x(5), x(8), x(11)$. В общем случае можно записать $X(k)$ в виде

$$\begin{aligned} X(k) &= \sum_{n=0}^{N-1} x(n) W_N^{kn} = \sum_{r=0}^{q_1-1} x(p_1 r) W_N^{p_1 r k} + \sum_{r=0}^{q_1-1} x(p_1 r + 1) W_N^k W_N^{p_1 r k} + \dots + \\ &+ \sum_{r=0}^{q_1-1} x(p_1 r + p_1 - 1) W_N^{(p_1-1)r} W_N^{p_1 r k} = \sum_{l=0}^{p_1-1} W_N^{lk} \sum_{r=0}^{q_1-1} x(p_1 r + l) W_N^{p_1 r k}. \end{aligned} \quad (5.16)$$

Внутренние суммы можно представить как q_1 -точечные ДПФ:

$$G(k) = \sum_{r=0}^{q_1-1} x(p_1 r + l) W_{q_1}^{rk},$$

потому что, как легко проверить, $W_N^{p_1 r k} = W_{q_1}^{rk}$, при $N = p_1 q_1$.

Таким образом, (5.16) представляет $X(k)$ в виде p_1 ДПФ последовательностей длины q_1 . Чтобы определить число комплексных умножений и сложений для вычисления ДПФ по (5.16), будем считать, что q_1 -точечные ДПФ получаются путем прямого вычисления. Из (5.16) видно, что нужно рассчитать p_1 q_1 -точечных ДПФ. Поэтому общее число требуемых комплексных сложений и умножений равно $p_1 q_1^2$. Внешняя сумма в (5.16) получается путем умножения q_1 -точечных ДПФ на коэффициенты W_N^{lk} и сложения результатов. Так как двойное суммирование в (5.16) выполняется для N значений k , то для объединения p_1 q_1 -точечных ДПФ требуется $N(p_1-1)$ комплексных сложений и умножений. Следовательно, общее число комплексных умножений и сложений для вычисления ДПФ по (5.16) равно $N(p_1-1) + p_1 q_1^2$. Теперь q_1 -точечное ДПФ может быть разложено аналогичным образом. В частности, если представить q_1 в виде $q_1 = p_2 q_2$, то q_1 -точечные последовательности во внутренней сумме (5.16) могут быть разбиты на p_2 последовательностей, каждая из которых состоит из q_2 точек, так, что внутренняя сумма в (5.16) может быть заменена на двойную сумму тем же способом, с которого мы начали. Тогда число операций требуемых для расчета q_1 -точечных ДПФ в (5.16) вместо q_1^2 станет равным $q_1(p_2-1) + p_2 q_2^2$. Следовательно, общее число комплексных умножений и сложений для вычисления ДПФ станет равным

$$N(p_1-1) + N(p_2-1) + p_1 p_2 q_2^2.$$

Если продолжить эту процедуру, разлагая далее q_2 -точечные ДПФ, то, в конце концов, общее число комплексных умножений и сложений для вычисления ДПФ станет равным

$$N(p_1 + p_2 + \dots + p_L - L). \quad (5.17)$$

Из (5.17) видно, что лучше производить разложение на максимально возможное число сомножителей и лучше выбирать простые сомножители.

В [3] рассмотрены примеры описанной процедуры с прореживанием по времени для составных N , а в [2] рассмотрен единый подход к различным алгоритмам БПФ на основе последовательного применения операции представления одномерного массива чисел двумерным. Большой набор методов вычисления ДПФ рассмотрен в [26].

5.4. Реализация ДПФ на основе цифровой фильтрации

Покажем, что ДПФ можно рассматривать как отклик цифрового фильтра и найдем структуру фильтра, определяющего один спектральный отсчет $X(k)$ в точке $\omega_k = 2\pi k/N$:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\omega_k n} = \sum_{n=0}^{N-1} x(n) h_k(N-1-n) = y_k(N-1). \quad (5.18)$$

В такой форме записи (5.18) можно рассматривать $X(k)$ как выходной сигнал $y_k(N-1)$ фильтра с импульсной характеристикой $h_k(n)$ в момент $(N-1)$. Из (5.18) импульсная характеристика $h_k(n)$ определится соотношением:

$$\begin{aligned} h_k(N-1-n) &= e^{-j\omega_k n}; \\ h_k(n) &= e^{j\omega_k n} e^{-j\omega_k(N-1)} = e^{j\omega_k n} e^{j\omega_k}, \quad 0 \leq n \leq N-1. \end{aligned} \quad (5.19)$$

Таким образом, цифровой фильтр – это КИХ-фильтр (комплексный), передаточная функция которого имеет вид:

$$H_k(z) = \sum_{n=0}^{N-1} h_k(n) z^{-n} = e^{j\omega_k} \sum_{n=0}^{N-1} (e^{j\omega_k} z^{-1})^n = e^{j\omega_k} \frac{1 - e^{j\omega_k N} z^{-N}}{1 - e^{j\omega_k} z^{-1}} = (1 - z^{-N}) \frac{e^{j\omega_k}}{1 - e^{j\omega_k} z^{-1}}. \quad (5.20)$$

Эта структура (5.20) уже встречалась – один канал фильтра с частотной выборкой. Для вычисления всех значений $X(k)$ ($k=0, 1, \dots, N-1$) необходимо N таких фильтров – частотных каналов анализатора спектра. Каждый частотный канал анализатора спектра представляет собой комплексный резонатор – когерентный накопитель отсчетов комплексной гармоники $e^{j\omega_k n}$. На рис. 5.11 представлена структурная схема анализатора спектра в виде гребенки фильтров. Достоинством

такого выполнения является возможность вычисления «скользящего» $X_n(k)$ спектра по N предшествующим текущему моменту n отсчетам: $x(n), x(n-1), x(n-2), \dots, x(n-N+1)$. Заметим, что при скользящем анализе на основе БПФ для каждого нового входного отсчета приходится вычислять полное БПФ, т.е. $N/2 \log_2 N$ базовых операций на отсчет. При анализе с помощью гребенки фильтров (рис.5.11) число комплексных умножений на отсчет равно N . Таким образом, в большинстве представляющих интерес случаев гребенка фильтров оказывается эффективнее скользящего БПФ.

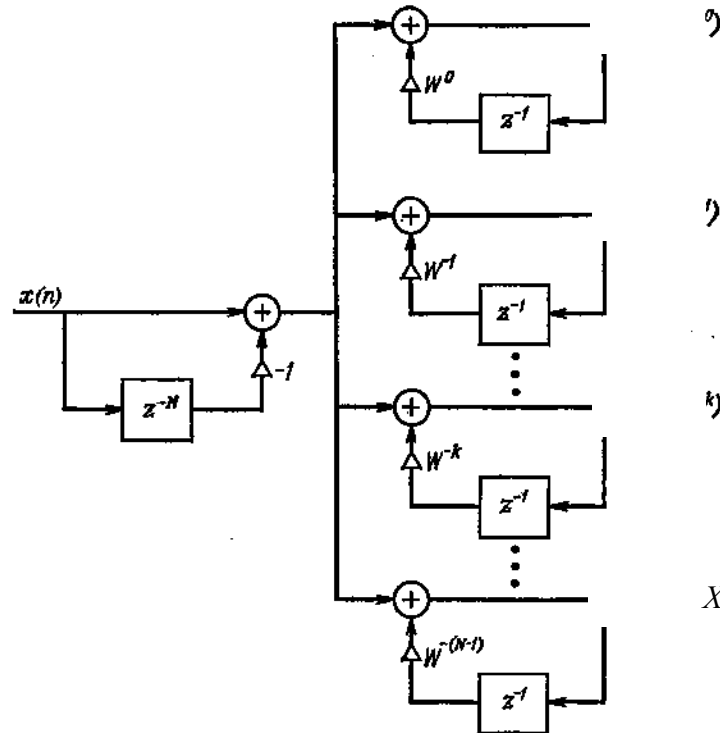


Рис.5.11. Анализатор спектра в виде гребенки фильтров.

Определим амплитудно-частотную характеристику одного канала анализатора спектра, вычислив $H_k(z)$ на единичной окружности:

$$|H_k(e^{j\omega})| = \left| \frac{1 - e^{-jN\omega}}{1 - e^{-j(\omega - \omega_k)}} \right| = \left| \frac{1 - e^{-jN(\omega - \omega_k)}}{1 - e^{-j(\omega - \omega_k)}} \right| = \left| e^{-j\frac{N-1}{2}(\omega - \omega_k)} \frac{\sin \frac{N(\omega - \omega_k)}{2}}{\sin \frac{(\omega - \omega_k)}{2}} \right| = \left| \frac{\sin \frac{N(\omega - \omega_k)}{2}}{\sin \frac{(\omega - \omega_k)}{2}} \right| =$$

$$= |D_N(\omega - \omega_k)|, \text{ где } D_N(\omega) = \frac{\sin \frac{N\omega}{2}}{\sin \frac{\omega}{2}}.$$

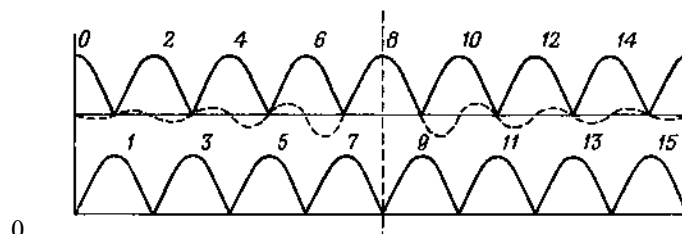


Рис.5.12. АЧХ гребенки фильтров, эквивалентной скользящему ДПФ с размером $N=16$.

На рис.5.12 изображены графики функций $D_N(\omega - \omega_k)$ для 16-точечного ДПФ, реализованного гребенкой фильтров в соответствии с рис.5.11. Частотные характеристики фильтров с четными номерами k показаны на рис.5.12 сверху, а с нечетными номерами k – внизу, при этом для всех фильтров, за исключением 8-го, изображены лишь главные лепестки. Видно, что скользящее ДПФ эквивалентно довольно грубому набору фильтров с АЧХ, имеющими относительно большие боковые

лепестки и существенным перекрытием между соседними фильтрами. При поступлении на вход анализатора спектра комплексной гармоники с частотой $\omega = \omega_k = 2\pi k/N$, все отсчеты $X(i) \equiv 0$, ($i \neq k$), а $X(k) = N$. Если же $\omega \neq 2\pi k/N$, то отличны от нуля все $X(i)$, т.е. происходит «растекание» (просачивание) энергии входной гармоники по всем N спектральным отсчетам. Чтобы улучшить частотную характеристику канальных фильтров анализатора можно ввести весовую обработку входной последовательности $x(n)$ умножением ее на весовую функцию (окно) $w(n)$. Аналогично проделанной выше процедуре, можно показать, что АЧХ канальных фильтров с весовой обработкой совпадает с амплитудно-частотным спектром окна, сдвинутого на частоту ω_k . Выбор функций окна обсудим ниже в специальном разделе.

На основе цифровых фильтров предложен более экономный по числу операций, чем прямой алгоритм вычисления $X(k)$, получивший название *алгоритма Герцеля* [3]. Заметим, что в этих алгоритмах нет необходимости вычислять все N значений $X(k)$, т.е. в общем случае можно вычислять $X(k)$ для любых M значений k . Поэтому, при $M < \log_2 N$ прямой метод или метод Герцеля могут быть наиболее эффективными методами вычисления ДПФ.

Приложение П5.1

```
SUBROUTINE FFT2(X,N,INV)
C   программа быстрого преобразования Фурье
C   X-комплексный массив отсчетов функции или спектра
C   N-размер массива X, равный целой степени двойки
C   INV-указатель преобразования (=0-прямое, =1-обратное)
C
  COMPLEX X(1),W,T,CMLPX
  ITER=0
  IREM=N
10  IREM=IREM/2
  IF(IREM.EQ.0) GO TO 20
  ITER=ITER+1
  GO TO 10
20  CONTINUE
  SIGN=-1.
  IF(INV.EQ.1) SIGN=1.
  NXP2=N
  DO 50 IT=1,ITER
  NXP=NXP2
  NXP2=NXP/2
  DEL=FLOAT(NXP2)
  WPWR=3.141592/DEL
  DO 40 M=1,NXP2
  ARG=FLOAT(M-1)*WPWR
  W=CMPLX(COS(ARG),SIGN*SIN(ARG))
  DO 40 MXP=NXP,N,NXP
  J1=MXP-NXP+M
  J2=J1+NXP2
  T=X(J1)-X(J2)
  X(J1)=X(J1)+X(J2)
40  X(J2)=T*W
50  CONTINUE
  N2=N/2
  N1=N-1
  J=1
  DO 65 I=1,N1
  IF(I.GE.J) GO TO 55
  T=X(J)
  X(J)=X(I)
  X(I)=T
55  K=N2
60  IF(K.GE.J) GO TO 65
  J=J-K
  K=K/2
  GO TO 60
65  J=J+K
  IF(INV.NE.1) GO TO 75
  DO 70 I=1,N
70  X(I)=X(I)/FLOAT(N)
75  CONTINUE
  RETURN
  END
```

5.5. Спектральные представления детерминированных сигналов

Напомним определения понятий, связанных со спектральными представлениями детерминированных сигналов непрерывного и дискретного времен.

Для непрерывных во времени детерминированных сигналов $x(t)$ понятие спектральной плотности амплитуды $X(f)$ связано с парой преобразований Фурье (НВПФ):

$$\begin{aligned} X(f) &= \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt; \\ x(t) &= \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} df. \end{aligned} \quad (5.21)$$

Энергия E сигнала $x(t)$ определяется соотношением

$$E = \int_{-\infty}^{\infty} x^2(t) dt = \int_{-\infty}^{\infty} |X(f)|^2 df = \int_{-\infty}^{\infty} E(f) df,$$

где $E(f) = |X(f)|^2$ – спектральная плотность энергии (СПЭ), описывающая распределение энергии сигнала по частоте. Для сигналов с ограниченной мощностью вводят спектральную плотность мощности, описывающую распределение мощности (энергии, рассеиваемой в единицу времени) по частоте.

В случае, когда мы располагаем лишь отсчетами (выборками) непрерывной функции времени $x(n) = x(t)|_{t=nT}$, взятыми с периодом T , понятие спектральной плотности амплитуды последовательности отсчетов определяется парой дискретно-временного преобразования Фурье (ДВПФ)

$$\begin{aligned} X_p(f) &= T \sum_{n=-\infty}^{\infty} x(n) e^{-j2\pi fnT}, \\ x(n) &= \int_{-1/2T}^{1/2T} X_p(f) e^{j2\pi fnT} df, n = 0, \pm 1, \dots \end{aligned} \quad (5.22)$$

Заметим, что $X_p(f)$ определяет периодическую (с периодом $1/T$) функцию частоты, которая является периодическим продолжением спектральной плотности $X(f)$ сигнала $x(t)$. Если $X(f)$ имеет ограниченную локализацию по оси частот, например, в интервале $-f_s \leq f \leq f_s$, а период отсчетов удовлетворяет соотношению $T \leq 1/2f_s$, ($f_s = \frac{1}{T} \geq 2f_s$), то спектральные полосы функции $X_p(f) = \sum_{k=-\infty}^{\infty} X(f - kf_s)$ не перекрываются, поэтому

$$X_p(f) = X(f), -1/2T \leq f \leq 1/2T.$$

Следовательно, спектр дискретной последовательности $x(n)$ повторяет форму спектра функции $x(t)$, которая может быть восстановлена по ее отсчетам. В противном случае имеет место эффект наложения спектров и восстановление функции $x(t)$ по ее отсчетам $x(n)$ оказывается невозможным.

Заметим также, что $X_p(f)$ связано с z -преобразованием последовательности

$$X(z) = \sum_{n=-\infty}^{\infty} x(n) z^{-n}:$$

$$X_p(f) = T X(z) \Big|_{z=e^{j2\pi fT}}.$$

Поэтому пару соотношений ДВПФ можно записать с введением шкалы частот, нормированной к частоте дискретизации $f/f_s = fT$ в виде

$$X(e^{j2\pi f}) = T \sum_{n=-\infty}^{\infty} x(n) e^{-j2\pi f n},$$

$$x(n) = \frac{1}{T} \int_{-0.5}^{0.5} X(e^{j2\pi f}) e^{j2\pi f n} df. \quad (5.23).$$

Для сигнала $x(t)$, ограниченного по длительности интервалом $0-NT$ секунд, спектральные представления иногда связывают с непрерывно-временным рядом Фурье (НВРФ):

$$X(k) = \int_0^{NT} x(t) e^{-j2\pi kt/NT} dt,$$

$$x_p(t) = \frac{1}{NT} \sum_{k=-\infty}^{\infty} X(k) e^{j2\pi kt/NT}, \quad (5.24)$$

который по существу на всей временной оси представляет периодически (с периодом NT) продолженную функцию $x(t)$, т.е. $x_p(t) = \sum_{k=-\infty}^{\infty} x(t - kNT)$. Для $0 \leq t \leq NT$ $x_p(t) = x(t)$, а $X(k) = X(f_k) \Big|_{f_k = k/NT}$, т.е. коэффициенты ряда Фурье суть отсчеты непрерывного по частоте спектра $X(f)$ в дискретной последовательности точек на оси частот $f_k = k/NT$.

Наконец для последовательности $x(n)$ конечной длины в N отсчетов спектральные представления связываются с дискретно - временным рядом Фурье (ДВРФ):

$$X(k) = T \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N}, 0 \leq k \leq N-1;$$

$$x(n) = \frac{1}{NT} \sum_{k=0}^{N-1} X(k) e^{j2\pi kn/N}, 0 \leq n \leq N-1. \quad (5.25)$$

Обе последовательности $X(k)$ и $x(n)$ периодичны с периодом N и на интервале в один период представляют соответственно совокупность отсчетов спектра $X(f)$ в точках $f_k = k/NT$ и отсчетов сигнала $x(t)$ в точках $t = nT$.

Соотношения (5.25) называют еще парой дискретного преобразования Фурье (ДПФ), их отличия от традиционно используемых, вызванные наличием множителей T и $1/T$, связаны с желанием обеспечить корректность масштабов при вычислении энергии и мощности. Теорема об энергии для ДПФ имеет вид

$$E = T \sum_{n=0}^{N-1} |x(n)|^2 = \frac{1}{NT} \sum_{k=0}^{N-1} |X(k)|^2.$$

Поэтому спектральная плотность энергии примет вид

$$E(k) = |X(k)|^2 = \left| T \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N} \right|^2, \quad (5.26)$$

а соотношение для спектральной плотности мощности (СПМ) примет форму

$$P(k) = \frac{1}{NT} E(k) = \frac{T}{N} \left| \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N} \right|^2, 0 \leq k \leq N-1. \quad (5.27)$$

Таким образом, ДПФ (5.25) позволяет преобразовать N отсчетов $x(n)$ непрерывного во времени сигнала $x(t)$ со спектральной плотностью $X(f)$ в N отсчетов $X(k)$ его спектральной плотности. Для получения большего количества отсчетов спектра $L > N$, т.е. для частот $f_k = k/L$, следует доопределить исходную последовательность $x(n)$, заданную на интервале $0 \leq n \leq N-1$, необходимым числом нулевых отсчетов, полагая $x(n) = 0$, $N \leq n \leq L-1$. Алгоритм ДПФ, реализуемый одним

из вариантов БПФ, является основным при реализации спектрального анализа сигналов на основе цифровых вычислительных устройств.

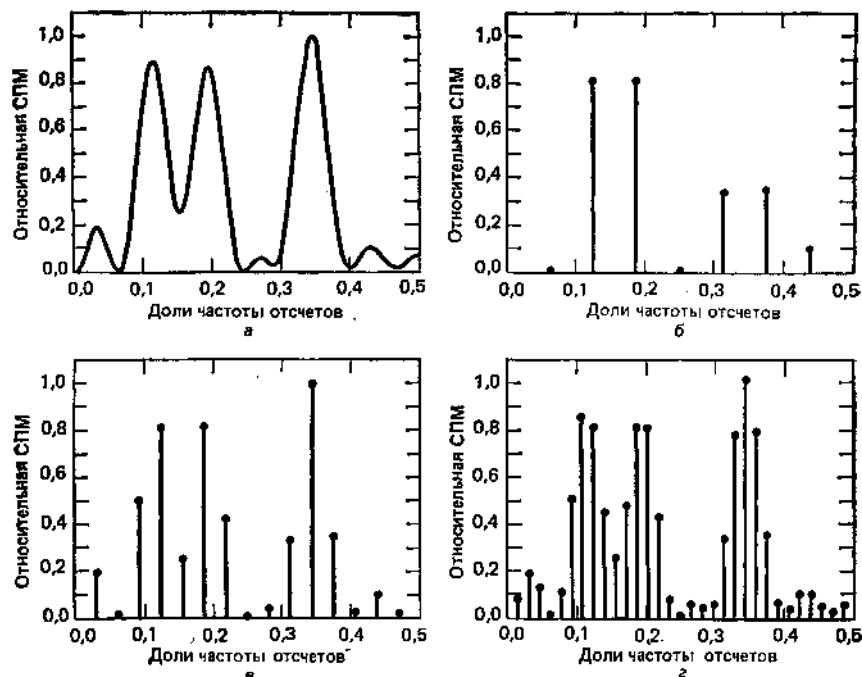


Рис. 5.13. Интерполяция за счет дополнения нулями: *а* - модуль ДВПФ 16-точечной записи данных, содержащих три синусоиды; *б* - модуль ДПФ той же последовательности данных без дополнения нулями (неопределенности не разрешены); *в* - модуль ДПФ той же последовательности после двукратного увеличения ее отсчетов за счет дополнения нулями (неопределенности разрешены, так как различимы все три синусоиды); *г* - модуль ДПФ той же последовательности данных после восьмикратного увеличения числа ее отсчетов за счет дополнения нулями.

Рис.5.13 показывает как влияет дополнение нулями исходной 16-точечной последовательности данных, содержащих три синусоиды, с соответствующим увеличением объема ДПФ на воспроизведение ее спектра.

5.6. Разрешение и произведение длительности на ширину спектра

Одним из важных параметров устройства спектрального оценивания является его разрешающая способность, под которой понимается способность разрешать (раздельно измерять) спектральные отклики двух синусоидальных сигналов, близких по частоте и амплитуде. Считается, что спектральные пики разрешены, если провал между ними составляет величину не менее 3 дБ.

Обычно полагают, что разнесение частот двух синусоид не может быть меньше эквивалентной ширины полосы частот B_e спектра окна, через которое наблюдаются отрезки этих синусоид. Поэтому говорят, что разрешение в герцах приблизительно равно величине, обратной времени наблюдения T_e в секундах, поскольку $T_e B_e \approx 1$. Обосновать это утверждение можно на основе следующих соображений.

Сигнал $x(n)$ не может быть одновременно быть ограниченным по длительности и по ширине полосы спектра. Однако его можно охарактеризовать некоторым интервалом T_e секунд, в котором сосредоточена большая часть его энергии при представлении во временной области, и некоторым интервалом B_e герц, в котором сосредоточена большая часть его энергии при представлении в частотной области. Для количественного описания временной концентрации энергии дискретно-временной последовательности отсчетов сигнала и соответствующей частотной концентрации ее ДВПФ предложены несколько различных мер. Определим две из них.

Эквивалентная длительность T_e сигнала $x(n)$ определяется как

$$T_e = \frac{T \sum_{n=-\infty}^{\infty} x(n)}{x(0)}, \quad (5.28)$$

т. е. иными словами, как «площадь» этого дискретного сигнала, поделенная на его центральное значение. Эквивалентная длительность сигнала равна длительности сигнала с прямоугольной огибающей, высота которого равна значению $x(n)$ в начале координат, т. е. $x(0)$, а площадь равна площади исходного сигнала; см. рис. 5.14а и 5.14б. Заметим, что в общем случае T_e не будет целочисленно кратным интервалу отсчетов T . Эквивалентная ширина полосы B_e дискретно-временного преобразования Фурье $X(f)$ сигнала $x(n)$ определяется аналогично как

$$B_e = \frac{\int_{-1/2T}^{1/2T} X(f) df}{X(0)} \quad (5.29)$$

см. рис. 5.14 в и 5.14 г. Эти две меры временной и частотной концентраций применимы лишь к действительным симметричным сигналам с максимальным значением в начале координат, интеграл от которых имеет конечное ненулевое значение. Эти условия точно выполняются для весовых функций (окон), которые рассматриваются в связи с обсуждением классических спектральных оценок.

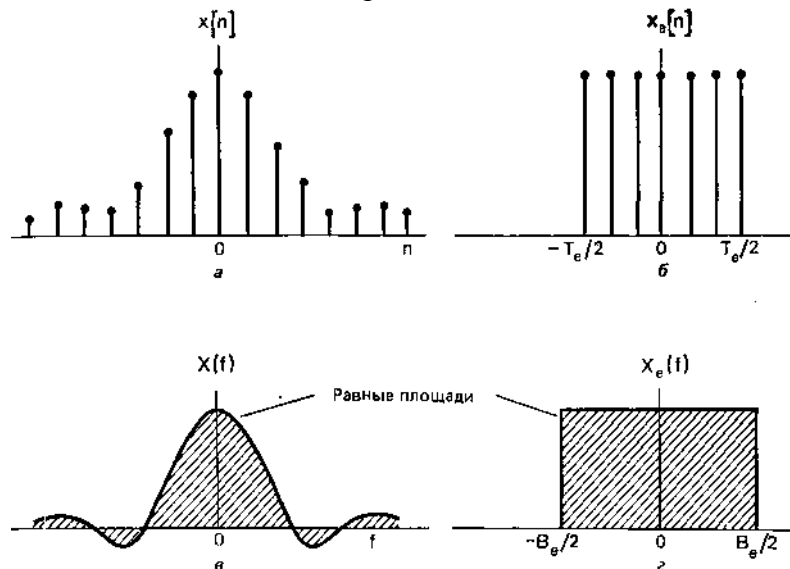


Рис.5.14. а – последовательность отсчетов исходного сигнала; б – эквивалентный сигнал с прямоугольной огибающей; в – ДВПФ исходного сигнала; г – эквивалентное ДВПФ прямоугольной формы.

Из соотношений ДВПФ с очевидностью следует, что

$$X(0) = T \sum_{n=-\infty}^{\infty} x(n),$$

$$x(0) = \int_{-1/2T}^{1/2T} X(f) df,$$

откуда получаем, что произведение длительности и ширины полосы равно

$$T_e B_e = \frac{T \sum_{n=-\infty}^{\infty} x(n)}{x(0)} \frac{\int_{-1/2T}^{1/2T} X(f) df}{X(0)} = 1$$

а это означает, что эквивалентная длительность сигнала и эквивалентная ширина его преобразования являются взаимно обратными величинами. Именно это равное единице произведение, в котором T_e полагается равным интервалу наблюдения, и положено в основу эмпирического правила, определяющего частотное разрешение.

Возможно другое определение произведения длительности и ширины полосы,

основанное на понятии среднеквадратичной длительности (ширины). Среднеквадратичная ширина является мерой дисперсии (среднеквадратичного отклонения) некоторой функции от ее среднего значения. Для того чтобы получить возможность оперировать с более широким классом функций, чем это возможно при использовании эквивалентных длительности (5.28) и ширины полосы (5.29), можно использовать определения среднеквадратичной ширины для функций вида $|x(n)|^2$ и $|X(f)|^2$. Такой подход позволяет получать значимые среднеквадратичные величины ширины для комплексных функций, осциллирующих функций и функций, которые имеют нулевую интегральную площадь[9].

На рис. 5.15 – 5.18 показано, как влияет на величину частотного разрешения длина N окна данных на примере вычисления ДПФ объемом 512 точек суммы двух комплексных синусоид с равными амплитудами и близкими частотами 0,2 и 0,24. Длина N окна данных выбиралась равной 10, 30, 100 и 300 отсчетов соответственно и обрабатываемая реализация дополнялась необходимым количеством нулевых отсчетов

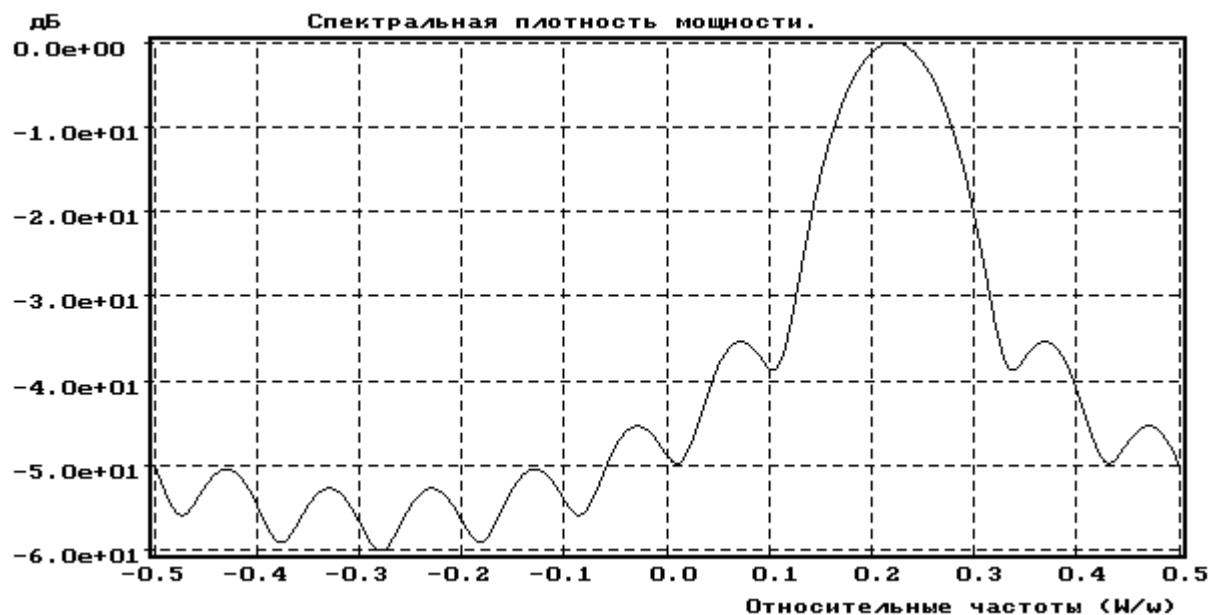


Рис. 5.15. Спектр суммы двух комплексных синусоид с равными амплитудами и близкими частотами 0,2 и 0,24, $N=10$.

На рис.5.15 синусоиды не разрешены, поскольку разнос их частот составляет величину $\Delta f=0,04$, а частотное разрешение $1/N=1/10$. На рис.5.16 синусоиды уже разрешены, поскольку частотное разрешение $1/N=1/30 \approx 0.33$. Рис.5.17 и 5.18 иллюстрируют спектры разрешенных синусоид при $N=100$ и $N=300$ соответственно.

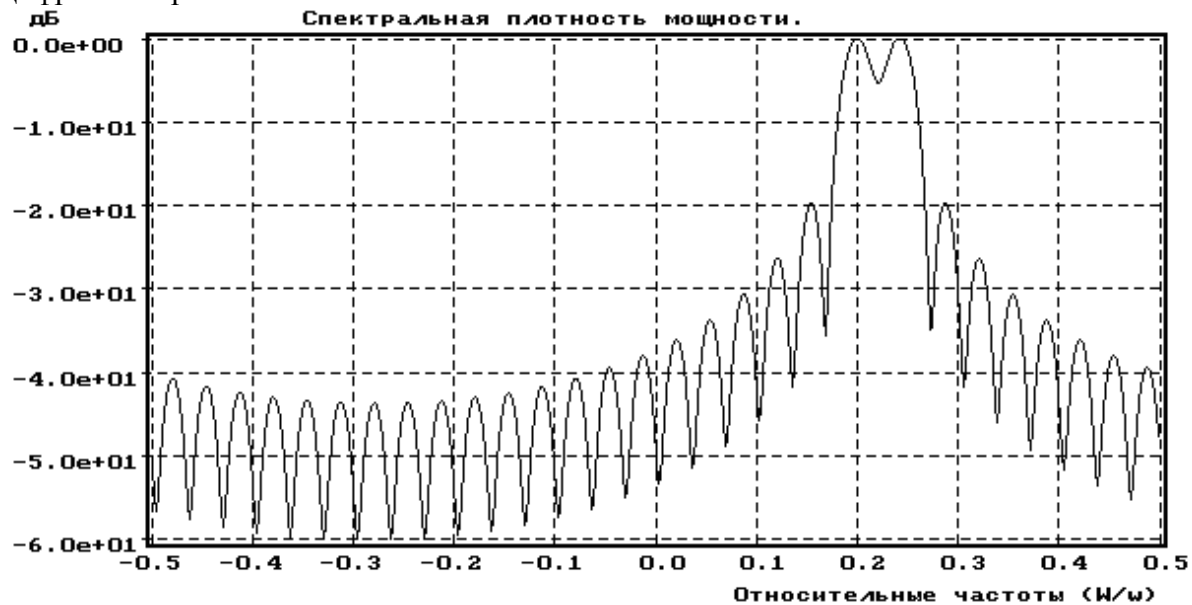


Рис. 5.16. Спектр суммы двух комплексных синусоид с равными амплитудами и близкими частотами 0,2 и 0,24, $N=30$.

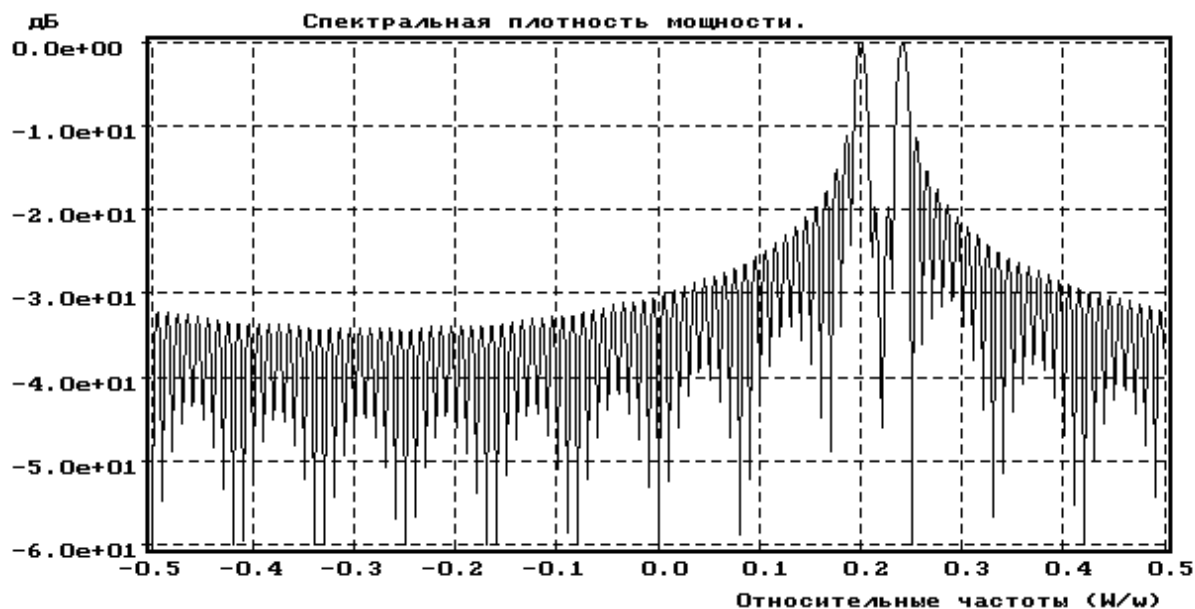


Рис. 5.17. Спектр суммы двух комплексных синусоид с равными амплитудами и близкими частотами 0,2 и 0,24, $N=100$.

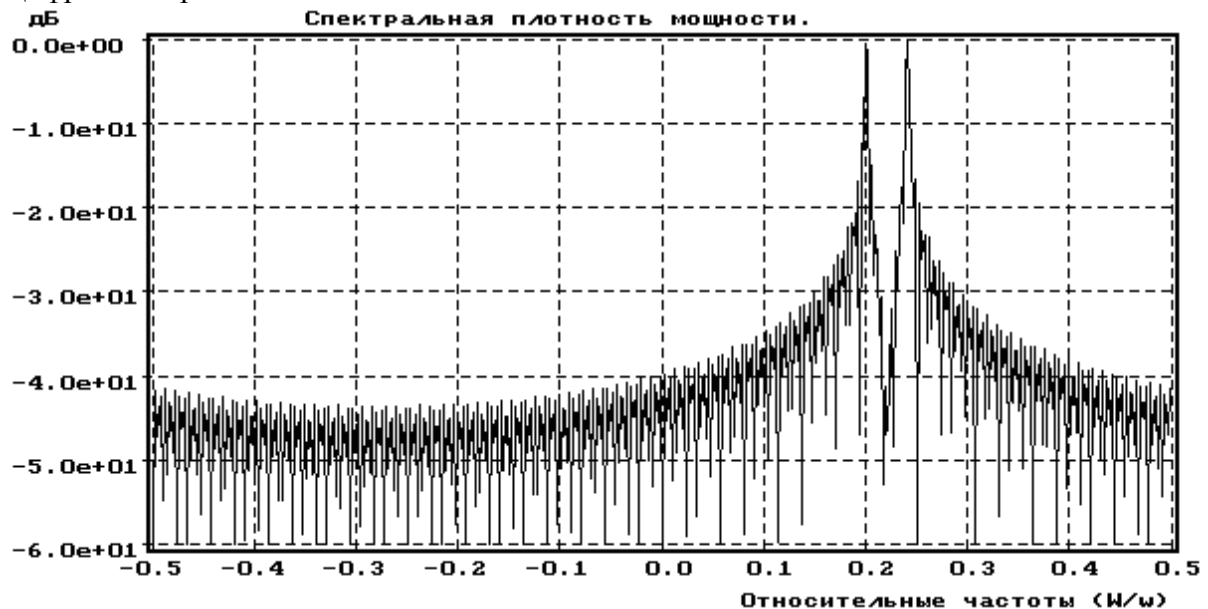


Рис. 5.18. Спектр суммы двух комплексных синусоид с равными амплитудами и близкими частотами 0,2 и 0,24, $N=300$.

5.7. Спектральные представления случайных процессов

Дискретный случайный процесс можно рассматривать как некоторую совокупность или ансамбль действительных или комплексных дискретных временных последовательностей $\{x_i(n)\}, i = 0, 1, 2, 3, \dots$, каждую из которых можно было бы наблюдать как результат проведения некоторого эксперимента. При фиксации временного индекса n в сечениях случайного процесса образуются случайные величины, при задании распределений плотностей вероятностей которых можно определить соответствующие моменты распределений.

Среднее значение (момент 1-го порядка) определяется выражением

$$\langle x(n) \rangle = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{i=0}^{M-1} x_i(n) = \int_{-\infty}^{\infty} x p_1(x, n) dx.$$

Автокорреляция случайного процесса в два различных момента времени n_1 и n_2 определяется выражением

$$r_{xx}(n_1, n_2) = \langle x(n_1) x^*(n_2) \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2^* p_2(x_1, n_1, x_2^*, n_2) dx_1 dx_2^*,$$

где $*$ - знак комплексного сопряжения. Автокорреляция центрированного случайного процесса, т.е. с удаленным средним значением, называется автоковариацией:

$$\begin{aligned} c_{xx}(n_1, n_2) &= \langle (x(n_1) - \langle x(n_1) \rangle) (x^*(n_2) - \langle x^*(n_2) \rangle) \rangle = \\ &= r_{xx}(n_1, n_2) - \langle x(n_1) \rangle \langle x^*(n_2) \rangle. \end{aligned}$$

Случайный процесс называется стационарным в широком смысле, если его среднее значение не зависит от времени, а автокорреляционная последовательность (АКП) зависит только от разности индексов времени $m = n_2 - n_1$.

Спектральная плотность мощности (СПМ) стационарного случайного процесса определяется как дискретно - временное преобразование Фурье автокорреляционной последовательности

$$P_{xx}(f) = T \sum_{m=-\infty}^{\infty} r_{xx}(m) e^{-j2\pi f m T}. \quad (5.31)$$

СПМ дискретного случайного процесса периодически (с периодом $1/T$ Гц)

размножает по оси частот СПМ исходного непрерывного во времени процесса в предположении, что она локализована в интервале частот $-1/2T \leq f \leq 1/2T$.

Обратное ДВПФ имеет вид

$$r_{xx}(m) = \int_{-1/2T}^{1/2T} P_{xx}(f) e^{j2\pi f m T} df, \quad (5.32)$$

поэтому при $m=0$ имеем

$$r_{xx}(0) = \langle x^2(n) \rangle = \int_{-1/2T}^{1/2T} P_{xx}(f) df.$$

Последнее соотношение оправдывает трактовку $P_{xx}(f)$ как распределение средней мощности процесса по частоте. Пару преобразований Фурье (5.31) и (5.32) называют соотношениями Винера - Хинчина для случая дискретного времени. Поскольку $r_{xx}(-m) = r_{xx}^*(m)$, то СПМ должна быть действительной и неотрицательной функцией. Если АКП – действительная функция, то $r_{xx}(-m) = r_{xx}(m)$, и СПМ можно представить в форме

$$P_{xx}(f) = 2T \sum_{m=0}^{\infty} r_{xx}(m) \cos(2\pi f m T), \quad (5.33)$$

а это означает четность СПМ как функции частоты, т.е. $P_{xx}(f) = P_{xx}(-f)$.

Если стационарный случайный процесс обладает свойством эргодичности, то процедуру усреднения по ансамблю при вычислении среднего значения и автокорреляции можно заменить усреднением по времени с использованием одной единственной реализации процесса.

Допущение об эргодичности позволяет ввести еще одно определение СПМ:

$$P_{xx}(f) = \lim_{M \rightarrow \infty} \left\langle \frac{1}{(2M+1)T} \left| T \sum_{n=-M}^M x(n) e^{-j2\pi f n T} \right|^2 \right\rangle. \quad (5.34)$$

Эта эквивалентная (5.32) форма СПМ получается путем статистического усреднения квадрата модуля ДВПФ взвешенной совокупности данных, поделенного на длину записи данных, при увеличении числа отсчетов данных до бесконечности. Метод определения СПМ на основе оценки автокорреляционной последовательности называют косвенным (коррелограммным), поскольку случайный процесс $x(n)$ непосредственно не используется для оценки СПМ. Определение СПМ по формуле (5.34) называют прямым методом (периодограммным), т.к. процесс $x(n)$ непосредственно используется в расчете СПМ.

5.8. Коррелограммный метод оценки СПМ.

Рассмотрим подробнее практическую реализацию коррелограммного метода оценивания СПМ. Автокорреляционная последовательность $r_{xx}(m)$ эргодического процесса как предел среднего по времени определяется соотношением

$$r_{xx}(m) = \lim_{M \rightarrow \infty} \frac{1}{2M+1} \sum_{n=-M}^M x(n+m) x^*(n). \quad (5.35)$$

На практике эта последовательность, как правило, не известна и поэтому должна оцениваться по имеющейся конечной последовательности данных.

Предположим, что мы располагаем N отсчетами данных $x(n)$, $n = 0, \dots, N-1$, тогда одним из вариантов соотношения для оценки АКП может быть, например, следующее:

$$\hat{r}_{xx}(m) = \frac{1}{N-m} \sum_{n=0}^{N-m-1} x(n+m) x^*(n), \quad (5.36)$$

применимое при $0 \leq m \leq N-1$.

При отрицательных значениях индекса m значения АКП получаются из свойства сопряженной симметрии $\hat{r}_{xx}(-m) = \hat{r}_{xx}^*(m)$. Таким образом, $(2N-1)$ значений АКП могут быть оценены по N отсчетам данных. Дискретная последовательность $\hat{r}_{xx}(m)$ является несмещенной оценкой истинной АКП, поскольку

$$\langle \hat{r}_{xx}(m) \rangle = \frac{1}{N-m} \sum_{n=0}^{N-m-1} \langle x(n+m)x^*(n) \rangle = r_{xx}(m).$$

При увеличении N значение дисперсии оценки $\hat{r}_{xx}(m)$ стремится к нулю, поэтому это статистически состоятельная оценка АКП [3].

Другой вариант соотношения для оценки АКП имеет вид

$$\check{r}_{xx}(m) = \frac{1}{N} \sum_{n=0}^{N-m-1} x(n+m)x^*(n), \quad 0 \leq m \leq N-1. \quad (5.37)$$

Эта оценка отличается от $\hat{r}_{xx}(m)$ только нормирующим множителем $\check{r}_{xx}(m) = \frac{N-|m|}{N} \hat{r}_{xx}(m)$, поэтому при конечном N она является смещенной, поскольку

$$\langle \check{r}_{xx}(m) \rangle = \left(1 - \frac{|m|}{N}\right) r_{xx}(m).$$

Для типичных приложений средний квадрат ошибки будет, как правило, больше для оценки $\hat{r}_{xx}(m)$, чем для оценки $\check{r}_{xx}(m)$. По этой и ряду других причин [9] предпочтение часто отдается смещенной оценке АКП.

Суммирование, предусмотренное в выражениях (5.36) и (5.37) можно записать как линейную свертку

$$\sum_{n=0}^{N-m-1} x(n+m)x^*(n) = x(n) * x^*(-n).$$

Следовательно, для вычисления оценок дискретной автокорреляции можно применить эффективные в вычислительном отношении процедуры, такие как «быстрая свертка», в которых используются алгоритмы БПФ.

Коррелограммный метод оценивания СПМ есть подстановка в соотношение (5.31) конечной последовательности значений оценки автокорреляции (коррелограммы) на место бесконечной последовательности неизвестных истинных значений АКП. Так, при подстановке несмещенной оценки АКП $\hat{r}_{xx}(m)$ с максимальными значениями временного сдвига $\pm L$ получается оценка СПМ в виде

$$\hat{P}_{xx}(f) = T \sum_{m=-L}^L \hat{r}_{xx}(m) e^{-j2\pi f m T}, \quad -\frac{1}{2T} \leq f \leq \frac{1}{2T}. \quad (5.38)$$

Максимальный индекс временного сдвига L обычно ограничивают величиной $L \approx N/10$. Причина выбора такого максимального значения – стремление устранить большие значения дисперсии, связанные с оценками автокорреляции при больших временных сдвигах.

Среднее значение оценки $\hat{P}_{xx}(f)$ определяется соотношением

$$\begin{aligned} \langle \hat{P}_{xx}(f) \rangle &= T \sum_{m=-L}^L \langle \hat{r}_{xx}(m) \rangle e^{-j2\pi f m T} = \\ &= T \sum_{m=-L}^L r_{xx}(m) e^{-j2\pi f m T} = T \sum_{m=-\infty}^{\infty} w(m) r_{xx}(m) e^{-j2\pi f m T}, \end{aligned} \quad (5.39)$$

где прямоугольная функция (окно) единичной амплитуды

$$w(m) = \begin{cases} 1, & -L \leq m \leq L, \\ 0, & m < -L; m > L. \end{cases}$$

Соотношение (5.39) есть ДВПФ произведения последовательностей АКП $r_{xx}(m)$ и $w(m)$, которое выражается через свертку ДВПФ сомножителей:

$$\langle \hat{P}_{xx}(f) \rangle = P_{xx}(f) * D_{2L+1}(f), \quad (5.40)$$

где ДВПФ прямоугольного окна

$$D_{2L+1}(f) = \frac{\sin(2L+1)\pi fT}{\sin \pi fT}.$$

Таким образом среднее значение оценки $\hat{P}_{xx}(f)$ отличается от истинного спектра $P_{xx}(f)$, поскольку это свертка истинного спектра и спектра прямоугольного окна.

Для уменьшения эффекта просачивания (растекания) спектра из-за влияния прямоугольного окна и уменьшения смещения оценки в соотношение (5.38) следует ввести весовую функцию (корреляционное окно) $w(m)$, $-L \leq m \leq L$, симметричное относительно начала отсчета. Тогда оценка СПМ, получаемая на основе коррелограммного метода, принимает форму

$$P_{xx}(f) = T \sum_{m=-L}^L w(m) \hat{r}_{xx}(m) e^{-j2\pi f m T}. \quad (5.41)$$

Среднее значение этой оценки будет сверткой истинного спектра и спектра окна $W(f)$:

$$\langle \hat{P}(f) \rangle = T \sum_{m=-L}^L w(m) r_{xx}(m) e^{-j2\pi f m T} = P_{xx}(f) * W(f).$$

Правильный выбор окна позволит уменьшить растекание спектра и его смещение.

Для вычисления оценки СПМ, определяемой на сетке из K частот $f_i = i/KT$, где $0 \leq i \leq K-1$, можно использовать алгоритм БПФ. Обычно значение $K=2^M \gg L$, а это требует доопределения $\hat{r}_{xx}^*(m) = 0$, при $L < m \leq K-L-1$ и $\hat{r}_{xx}(K-m) = \hat{r}_{xx}^*(m)$.

5.9. Периодограммный метод оценки СПМ

Как уже отмечалось выше, для эргодического случайного процесса эквивалентным определением СПМ является следующая дискретно-временная форма:

$$P_{xx}(f) = \lim_{N \rightarrow \infty} \left\langle \frac{1}{(2N+1)T} \left| T \sum_{n=-N}^N x(n) e^{-j2\pi f n T} \right|^2 \right\rangle.$$

При наличии конечного множества данных $x(n)$, $0 \leq n \leq N-1$ и единственной реализации это соотношение преобразуется в СПМ выборки или периодограмму:

$$\tilde{P}_{xx}(f) = \frac{1}{NT} \left| T \sum_{n=0}^{N-1} x(n) e^{-j2\pi f n T} \right|^2 = \frac{T}{N} \left| \sum_{n=0}^{N-1} x(n) e^{-j2\pi f n T} \right|^2. \quad (5.42)$$

Несложно показать [3], что среднее значение оценки $\tilde{P}_{xx}(f)$ такое же, как и у коррелограммной оценки, дисперсия оценки не стремится к нулю при $N \rightarrow \infty$, а среднеквадратичное значение оценки при любом N сравнимо со средним значением. Следовательно, выборочный спектр не является состоятельной (статистически устойчивой) оценкой СПМ. На рис.5.19 – 5.22 показаны периодограммы, вычисленные по (5.42), для суммы комплексной синусоиды с амплитудой, равной 1,0 и частотой 0,2 и белого нормального (гауссова) шума с дисперсией, равной 1,0 при $N=50, 100, 200, 500$

соответственно. Из сравнения рис. 5.19 – 5.22 видно что с ростом N дисперсия спектральной оценки не уменьшается, а значения периодограммы начинают быстрее флуктуировать.

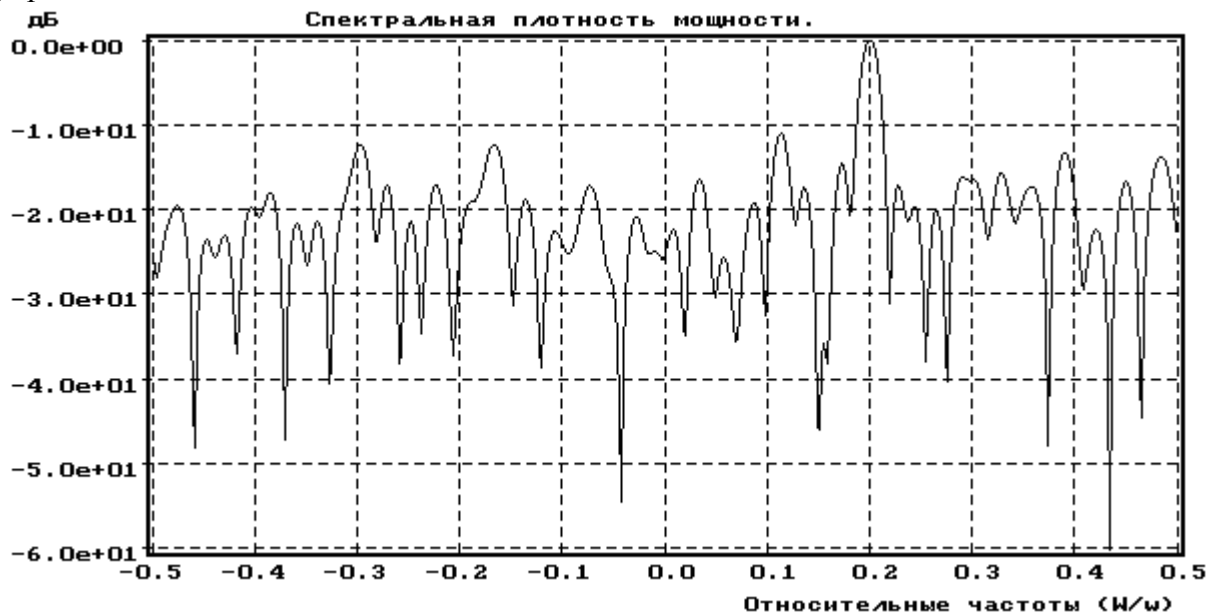


Рис. 5.19. Периодограмма суммы комплексной синусоиды с амплитудой, равной 1,0 и частотой 0,2 и белого нормального (гауссова) шума с дисперсией, равной 1,0 при $N=50$.

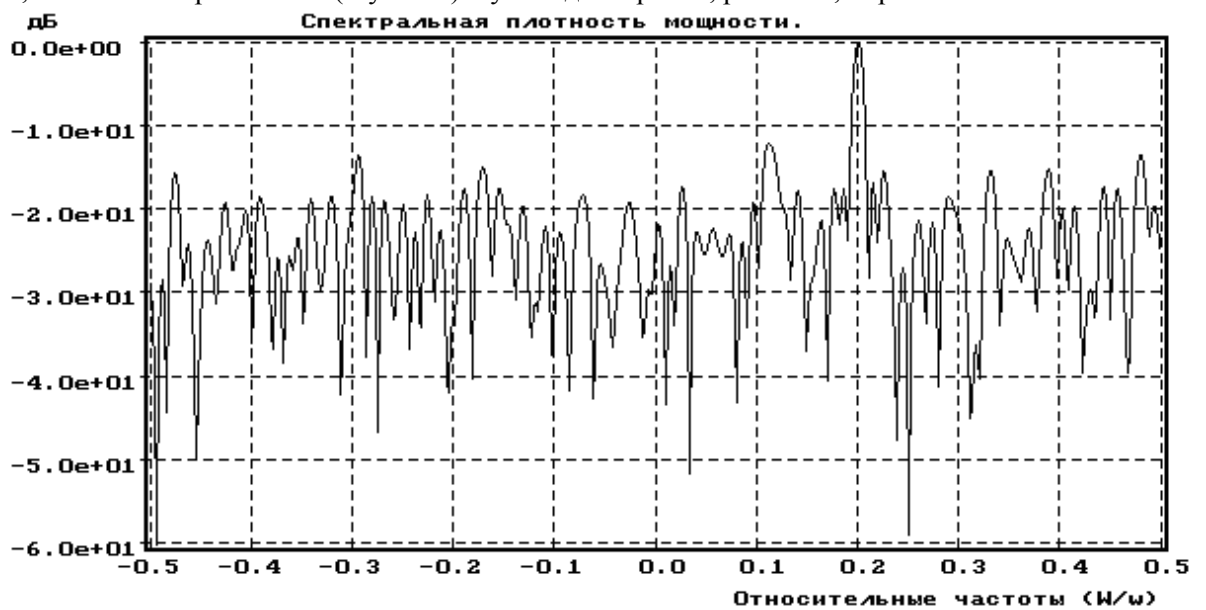


Рис. 5.20. Периодограмма суммы комплексной синусоиды с амплитудой, равной 1,0 и частотой 0,2 и белого нормального (гауссова) шума с дисперсией, равной 1,0 при $N=100$.

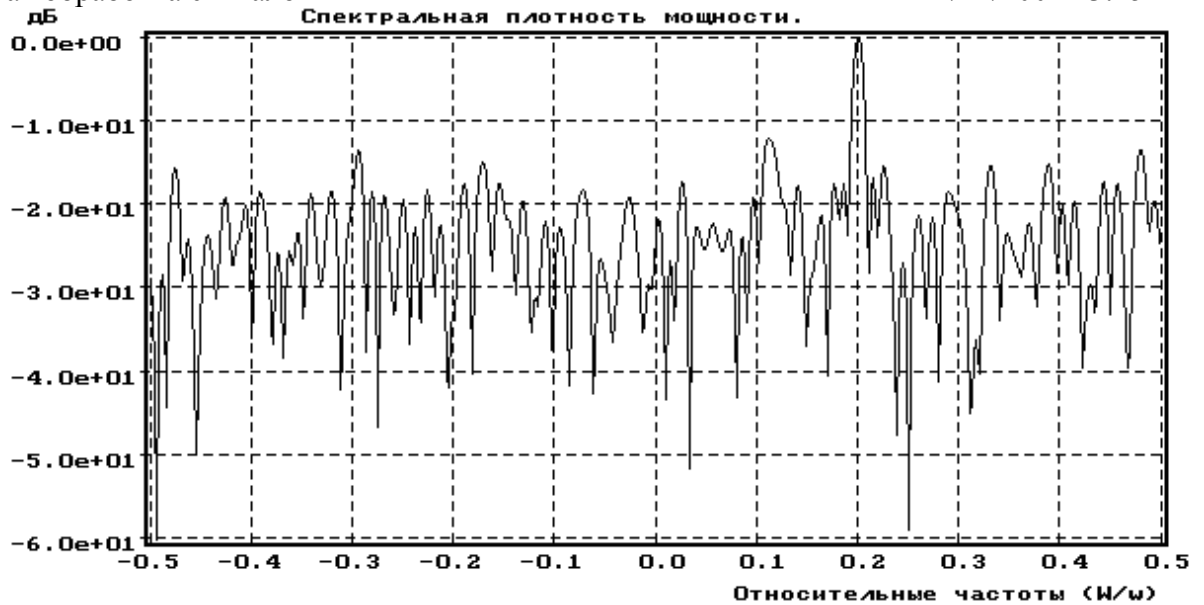


Рис. 5.21. Периодограмма суммы комплексной синусоиды с амплитудой, равной 1,0 и частотой 0,2 и белого нормального (гауссова) шума с дисперсией, равной 1,0 при $N=200$.

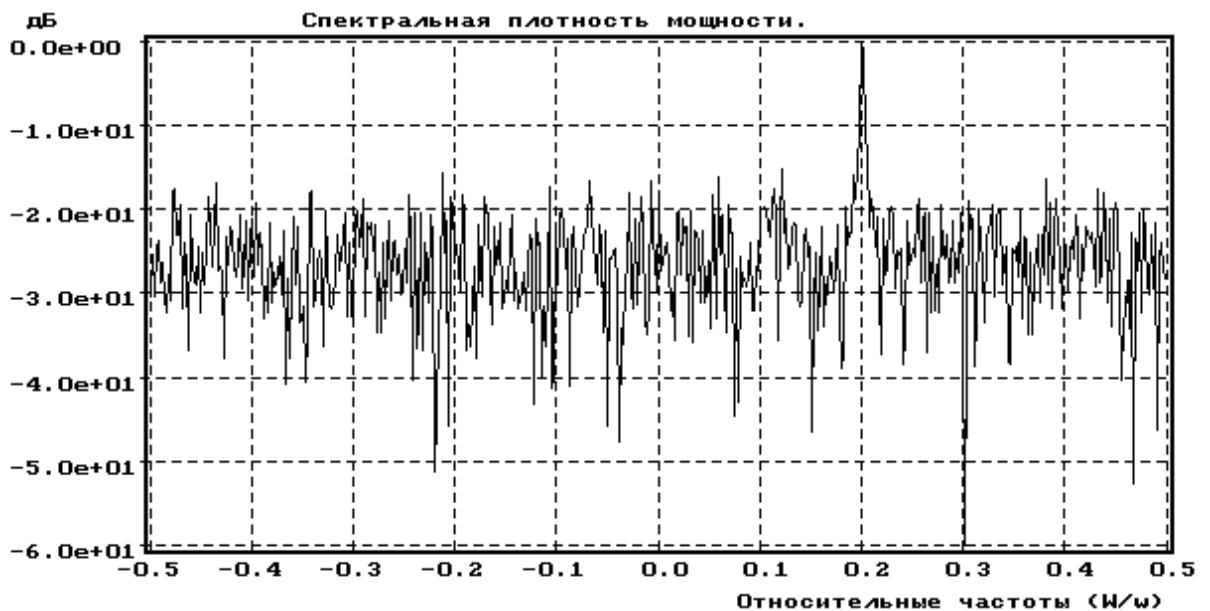


Рис. 5.22. Периодограмма суммы комплексной синусоиды с амплитудой, равной 1,0 и частотой 0,2 и белого нормального (гауссова) шума с дисперсией, равной 1,0 при $N=500$.

Предложен ряд модификаций соотношения (5.42), суть которых сводится к сглаживанию оценки путем введения псевдоусреднения по некоторому ансамблю [9]. Наибольшую популярность в настоящее время получил метод Уэлча, идея которого состоит в образовании псевдоансамбля периодограмм за счет разбиения наблюдаемых данных на перекрывающиеся сегменты и обработки каждого сегмента с помощью окна данных.

Если запись комплексных данных $x(n)$, $0 \leq n \leq N-1$ разбита на P сегментов по D отсчетов в каждом со сдвигом S отсчетов между соседними сегментами ($S \leq D$), то максимальное число сегментов P есть целая часть величины $(N-D)/S+1$. После взвешивания окном $w(n)$ сегмент с номером p ($0 \leq p \leq P-1$) будет содержать отсчеты

$$x^{(p)}(n) = w(n)x(n+ps), 0 \leq n \leq D-1.$$

Выборочный спектр P -го сегмента определяется выражением

$$\hat{P}_{xx}^{(p)}(f) = \frac{1}{UDT} X^{(p)}(f) [X^{(p)}(f)]^* = \frac{1}{UDT} |X^{(p)}(f)|^2, \quad (5.43)$$

где $X^{(p)}(f) = T \sum_{n=0}^{D-1} x^{(p)}(n) e^{-j2\pi f n T}$ - ДВПФ сегмента, $U = T \sum_{n=0}^{D-1} w^2(n)$ - энергия окна.

Среднее значение сегментных периодограмм определяет оценку СПМ:

$$\tilde{P}_{xx}(f) = \frac{1}{P} \sum_{p=0}^{P-1} \tilde{P}_{xx}^{(p)}(f). \quad (5.44)$$

Уэлч, в частности, предложил использовать окно Ханна и 50%-ное перекрытие сегментов, которое обеспечивало очень эффективные реализации его метода на основе алгоритма БПФ. Кроме того, при 50%-ном перекрытии сегментов все данные используются дважды, за исключением $D/2$ отсчетов на каждом конце исходной N -точечной последовательности данных, а это выравнивает обработку большинства отсчетов данных, поскольку те отсчеты, которые имели малые веса на одном сегменте, получают большие веса на следующем сегменте. Анализ поведения периодограммы Уэлча для гауссовских процессов показал, что минимальная дисперсия для окна Ханна достигается при 65%-ном перекрытии, при этом величина дисперсии увеличивается приблизительно на 8% при использовании 50%-ного перекрытия сегментов. Дисперсия периодограммы Уэлча примерно обратно пропорциональна числу сегментов, т.е.

$$\text{var}\{\tilde{P}_{xx}(f)\} = \frac{\tilde{P}_{xx}^2(f)}{P} \quad (5.45)$$

в предположении независимости сегментов (хотя перекрытие сегментов приводит, конечно, к некоторой их взаимозависимости). Благодаря перекрытию по заданной записи данных можно сформировать большее число сегментов и уменьшить величину дисперсии. На рис. 5.23 представлена периодограмма Уэлча суммы комплексной синусоиды с амплитудой, равной 1,0 и частотой 0,2 и белого нормального (гауссова) шума с дисперсией, равной 1,0 при $N=500$, $D=50$, $S=25$ (19 сегментов).

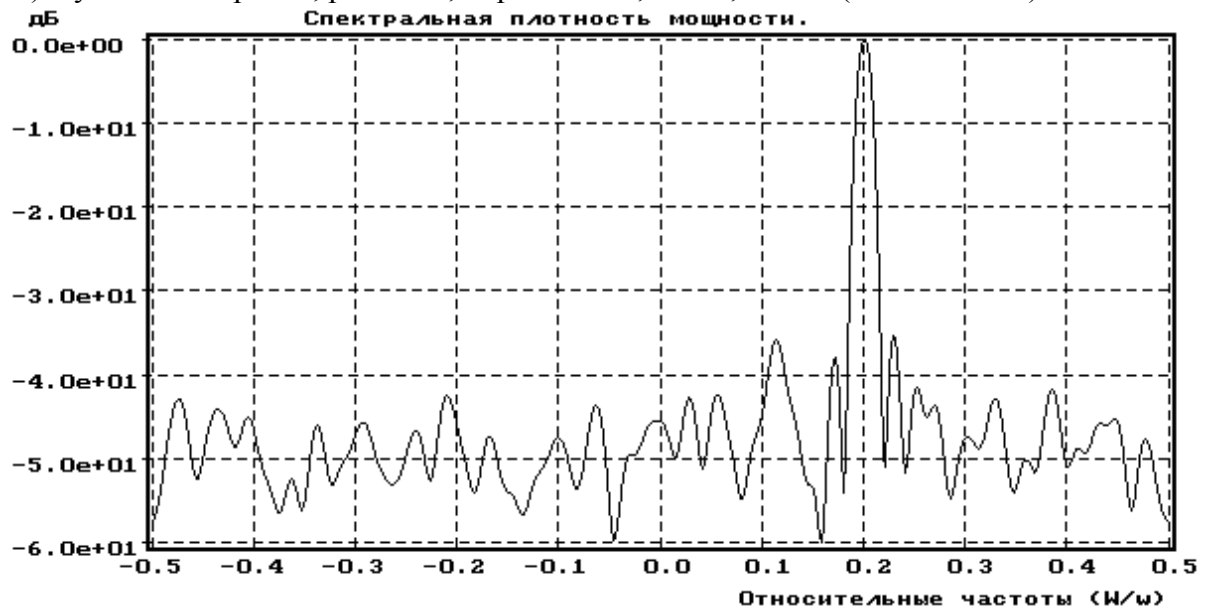


Рис. 5.23. Периодограмма Уэлча суммы комплексной синусоиды с амплитудой, равной 1,0 и частотой 0,2 и белого нормального (гауссова) шума с дисперсией, равной 1,0 при $N=500$, $D=50$, $S=25$ (19 сегментов).

5.10. Использование окон при спектральных измерениях

Один из важных вопросов, который является общим для всех классических методов спектрального оценивания, связан с применением функций окна.

Заметим, что конечную запись данных $x(n)$ из N отсчетов можно представить как некоторую часть исходной бесконечной последовательности $x_u(n)$, видимую через прямоугольное окно $w(n)$, в форме произведения:

$$x(n) = x_u(n)w(n),$$

где

$$w(n) = \begin{cases} 1, & 0 \leq n \leq N-1, \\ 0, & \text{при других } n. \end{cases}$$

При этом мы полагаем, что все ненаблюдаемые отсчеты равны нулю независимо от того, так ли это на самом деле или нет.

ДВПФ взвешенной окном последовательности, выраженное через преобразования последовательности $x_u(n)$ и прямоугольного окна $w(n)$, равно свертке этих преобразований:

$$X(f) = X_u(f) * D_N(f),$$

где $D_N(f) = T e^{-j\pi f T(N-1)} \frac{\sin(\pi f T N)}{\sin(\pi f T)}$ - ядро Дирихле, представляющее ДВПФ

прямоугольной функции. ДВПФ наблюдаемой конечной последовательности является искаженной версией ДВПФ бесконечной последовательности. Влияние прямоугольного окна на дискретно-временную синусоиду с частотой f_0 иллюстрирует рис.5.24, из которого видно, что острые спектральные пики ДВПФ исходной синусоидальной последовательности расширились из-за воздействия ДВПФ окна. При этом минимальная ширина спектральных пиков конечной последовательности ограничена шириной, определяемой главным лепестком ДВПФ окна, и не зависит от исходных данных. Боковые лепестки ДВПФ окна, называемые растеканием [просачиванием (spectral leakage)] будут изменять амплитуды соседних спектральных пиков, приводя к смещению спектральных оценок. Аналогичные искажения будут наблюдаться и в случае несинусоидальных сигналов. Просачивание приводит не только к появлению амплитудных ошибок, но может также маскировать присутствие слабых сигналов и, следовательно, затруднять их обнаружение.

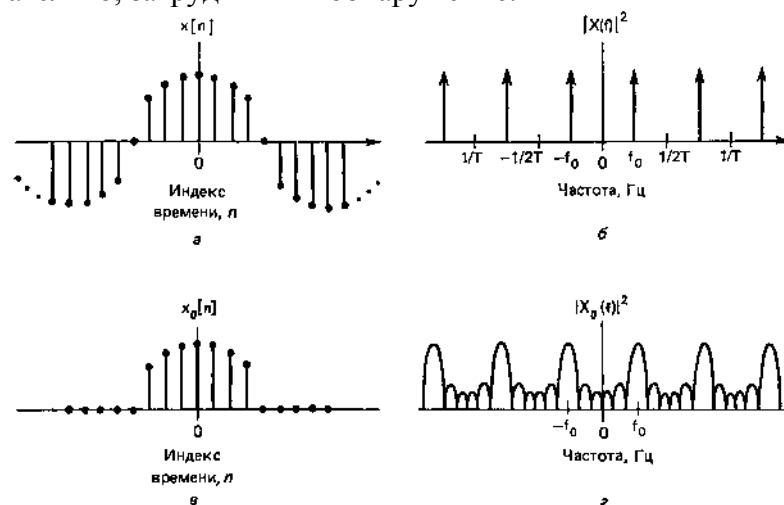


Рис.5.24. Иллюстрация эффекта растекания спектра из-за взвешивания данных:

a – исходная синусоидальная последовательность; $б$ – модуль периодического ДВПФ синусоидальной последовательности; $в$ – взвешенная синусоидальная последовательность; $г$ – модуль ДВПФ взвешенной синусоидальной последовательности.

Предложен ряд других функций окна, имеющих меньший уровень боковых лепестков их спектра, чем в случае прямоугольного окна [33]. Однако, как правило, это достигается ценой расширения главного лепестка спектра окна. Следовательно, окно должно выбираться с учетом компромисса между шириной главного лепестка и уровнем подавления боковых лепестков.

В таблице 5.2 представлены сравнительные характеристики некоторых оконных функций. Из всех приведенных окон самый узкий главный лепесток имеет частотная характеристика прямоугольного окна, однако у него самый высокий уровень боковых лепестков.

Таблица 5.2. Характеристики оконных функций

Окно	Дискретно-временная функция $w(n)$, $-N/2 \leq n \leq N/2$	Максимальный уровень боковых лепестков (дБ)	Ширина полосы по уровню 0,5 мощности относительно $1/NT$
Прямоугольное	1	-13,3	0,89
Треугольное (Бартлетта)	$1 - \frac{2 n }{N}$	-26,5	1,28
Ханна	$0,5 + 0,5 \cos(\frac{2\pi n}{N})$	-31,5	1,44
Хэмминга	$0,54 + 0,46 \cos(\frac{2\pi n}{N})$	-43	1,3

На рис.5.25 представлены типичные дискретно-временные функции окна (слева) и логарифм модуля их ДВПФ (справа) для прямоугольного окна, треугольного окна, окна Ханна и окна Хэмминга.

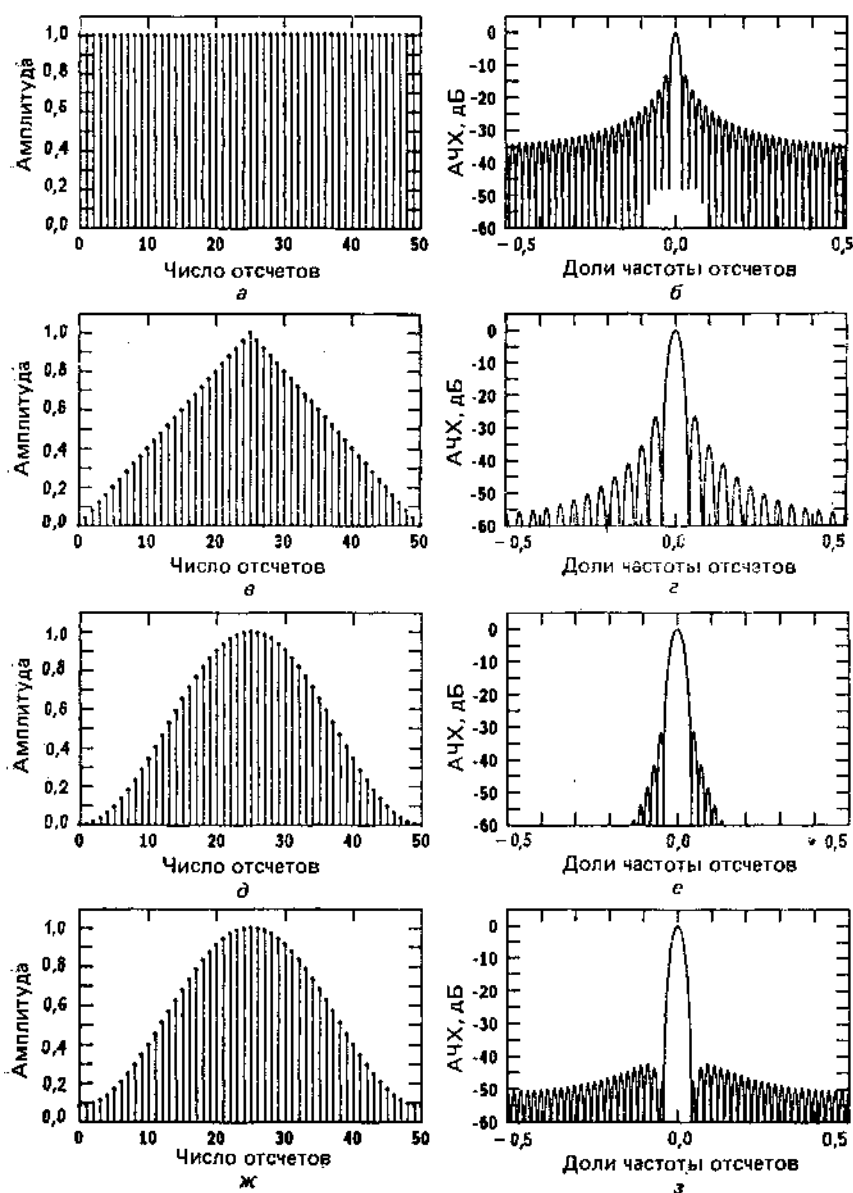


Рис.5.25. Типичные дискретно-временные функции окна (слева) и логарифм модуля их ДВПФ (справа): а-б -прямоугольное окно; в-г -треугольное окно; д-е -окно Ханна; ж-з- окно Хэмминга.

В таблице 5.2 приведены определения некоторых используемых N -точечных функций окна симметричных относительно $n=0$. Именно такие окна используются в качестве корреляционных окон в коррелограммном методе оценки СПМ. При этом

используются только окна нечетной длины с $w(0)=1$. Для окон данных (в периодограммном методе оценки СПМ) применяются окна четной и нечетной длины, начинающиеся с $n=0$ и имеющие симметрию относительно точки $(N-1)/2$. При этом у окон четной длины точка симметрии находится посередине между отсчетами.

5.11. Разрешение и произведение «устойчивость×длительность×ширина полосы»

Для детерминированного сигнала с длительностью $T_e = NT$ секунд произведение длительности на ширину полосы удовлетворяет условию $T_e B_e = 1$.

Поэтому разрешение спектральной плотности энергии, полученной по конечной последовательности отсчетов детерминированного сигнала, равно B_e герц, т.е. приблизительно равно величине, обратной интервалу записи данных. Для того чтобы определить разрешение спектральной плотности мощности, получаемой по конечной последовательности отсчетов случайного сигнала, произведение длительности на ширину полосы необходимо несколько видоизменить, с тем чтобы учесть случайный характер сигнала, влияющий на статическое «качество» спектральной оценки.

Одной из подходящих модификаций этого произведения является произведение «устойчивость×длительность×ширина полосы» $QT_e B_e$, где Q -статистический показатель качества, определяемый как отношение дисперсии оценки СПМ к квадрату математического ожидания этой оценки

$$Q = \frac{\text{var}\{\hat{P}(f)\}}{[E\{\hat{P}(f)\}]^2}. \quad (5.45)$$

Показатель качества Q -это инвертированное отношение сигнал/шум (SNR), которое непосредственно связано со статистической устойчивостью спектральной оценки. Значения Q , много меньшие единицы, соответствуют гладким спектральным оценкам с малой дисперсией. Значения Q , много большие единицы, соответствуют весьма зашумленным спектральным оценкам с большой дисперсией. Эквивалентная ширина полосы B_e замеряется теперь величиной B_s - эффективной статистической шириной полосы, которая в случае спектрального окна $W(f)$ определяется выражением

$$B_s = \frac{\left[\int_{-0.5T}^{0.5T} W(f) df \right]^2}{\int_{-0.5T}^{0.5T} W^2(f) df}. \quad (5.46)$$

Величина B_s характеризует ширину полосы эквивалентного прямоугольного окна с тем же значением отношения дисперсии к квадрату среднего значения на его выходе, как и у окна $W(f)$ в том случае когда на его вход подан белый шумовой процесс. Однако обе эти величины можно связать соотношением $B_s = \alpha B_e$, где α выбирается в зависимости от окна, так для прямоугольного окна $\alpha=0.8$, для окна Ханна-1.3 и для окна Хемминга -1.4. Ширина полосы B_s позволяет грубо судить о разрешении спектральной оценки

Показано, что для гауссовского случайного процесса условие

$$QT_e B_s \geq 1. \quad (5.47)$$

применимо ко всем классическим процедурам спектрального оценивания. Если $T_e B_s \cong 1$, то, как правило $QT_e B_s > 1$. С другой стороны, если значение произведения $T_e B_s$ выбрано значительно большим единицы, то для всех классических спектральных оценок будет выполняться приближенное равенство

$$QT_e B_s \approx 1. \quad (5.48)$$

Произведение «устойчивость×длительность×ширина полосы» устанавливает взаимосвязь между тремя фундаментальными параметрами, от которых зависят характеристики спектральных оценок. Для заданной записи данных длительностью T_e невозможно получить оценки, которые одновременно обладают высоким разрешением (малыми значениями B_s) и высокой устойчивостью (малыми значениями Q). Например, если необходимо, чтобы дисперсия спектральной оценки составляла десятую часть от ее среднего значения, то $Q=0,1$ и достижимое разрешение будет равно $B_e \approx 10/T_e$ герц. Это значение в 10 раз больше эмпирического значения $1/T_e$ для разрешения детерминированного сигнала. Этот пример подчеркивает важность принятия такого компромиссного решения, которое не нарушало бы баланс между статистической устойчивостью и разрешением, которые требуются от всех методов спектрального оценивания. Заметим, что увеличение числа отсчетов за счет увеличения частоты дискретизации при неизменном интервале T_e не будет влиять на максимально достижимое разрешение, так как на него влияет только длина записи данных, но не число отсчетов. На рис.5.26 и 5.27 показаны периодограммы Уэлча суммы двух комплексных синусоид с амплитудами, равными 1,0 и частотами 0,2 и 0,24 и белого нормального (гауссова) шума с дисперсией, равной 1,0 при $N=500$, $D=50$, $S=25$ (19 сегментов) и при $N=500$, $D=250$, $S=125$ (3 сегмента) соответственно, которые иллюстрируют соотношение между устойчивостью и разрешением. Рис. 5.26 соответствует спектральной оценке с большей устойчивостью, но меньшим разрешением, а рис. 5.27 соответствует спектральной оценке с меньшей устойчивостью, но большим разрешением.

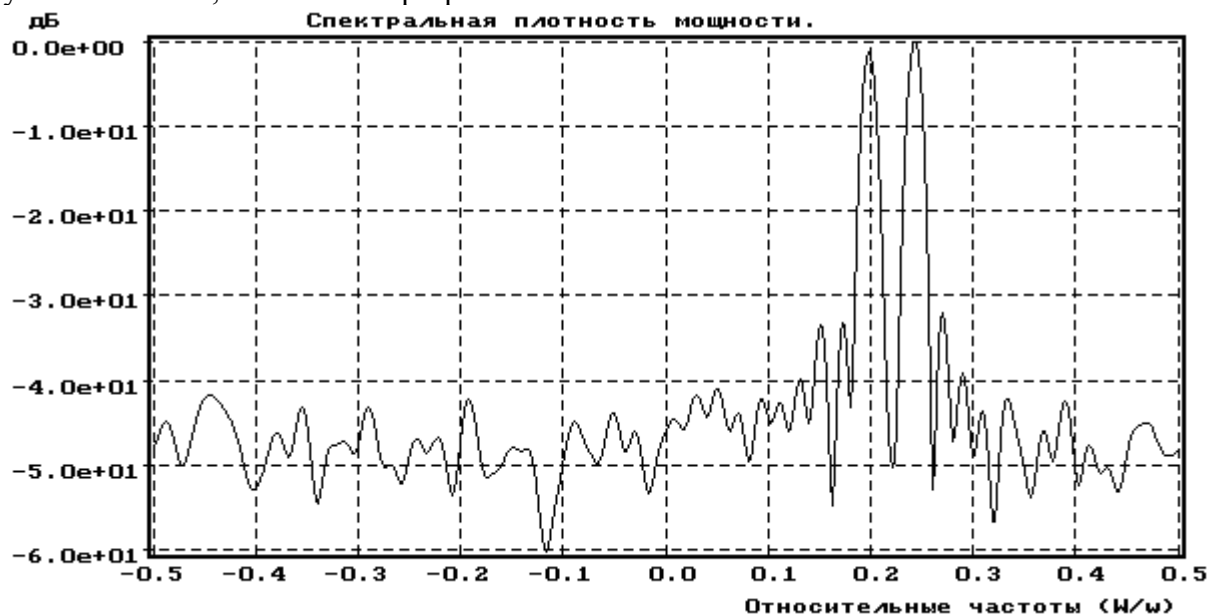


Рис. 5.26. Периодограмма Уэлча суммы двух комплексных синусоид с амплитудами, равными 1,0 и частотами 0,2 и 0,24 и белого нормального (гауссова) шума с дисперсией, равной 1,0 при $N=500$, $D=50$, $S=25$ (19 сегментов).

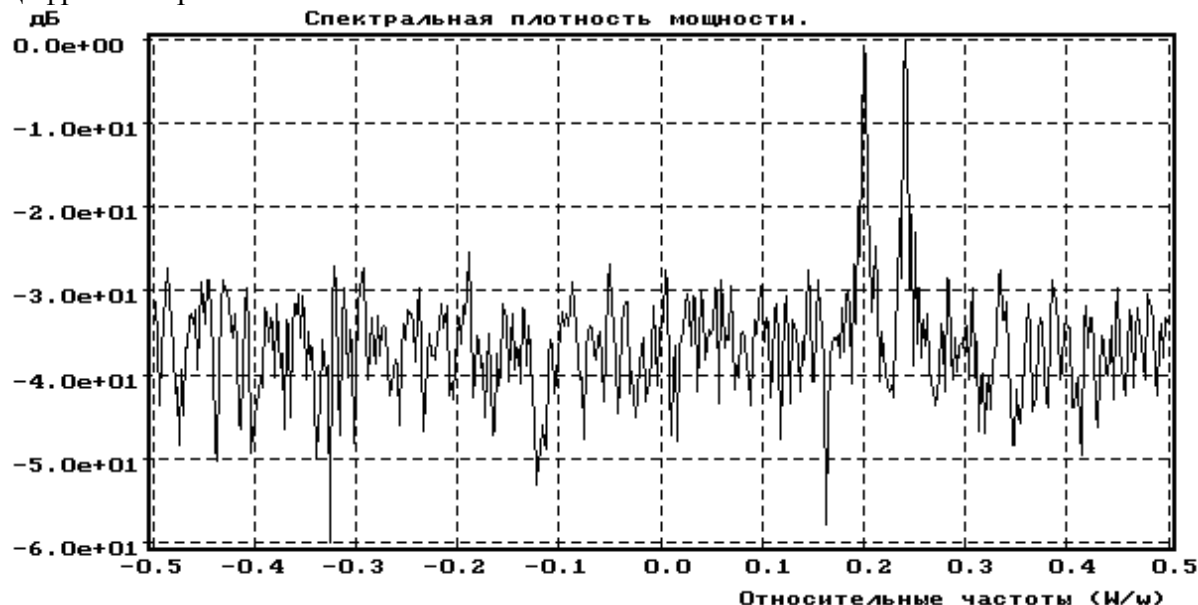


Рис. 5.27. Периодограмма Уэлча суммы двух комплексных синусоид с амплитудами, равными 1,0 и частотами 0,2 и 0,24 и белого нормального (гауссова) шума с дисперсией, равной 1,0 при $N=500$, $D=250$, $S=125$ (3 сегмента).

5.12. Классические спектральные оценки тест-последовательности

Для того, чтобы охарактеризовать поведение каждой спектральной оценки при анализе короткой последовательности данных Марпл [9] предложил использовать искусственно синтезированную последовательность данных с заранее известными свойствами. Эта тест-последовательность данных содержит 64 отсчета некоторого комплекснозначного процесса, состоящего из 4-х комплексных синусоид и комплексного аддитивного окрашенного шума. Истинный спектр тест-последовательности показан на рис. 5.28.

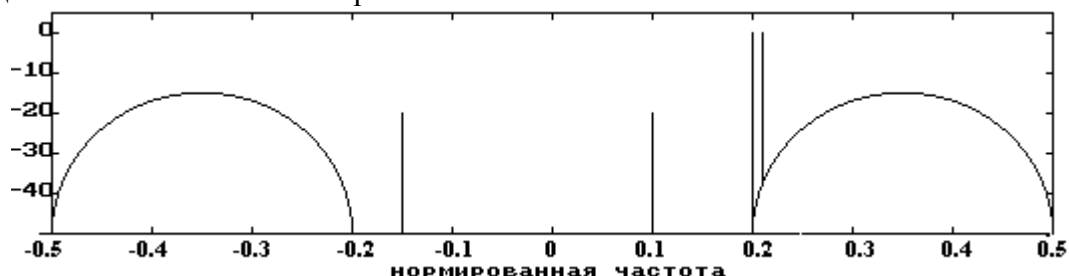


Рис. 5.28. Истинный спектр процесса, соответствующего тест-последовательности данных.

Частоты двух синусоид выбраны очень близкими друг к другу (0,2 и 0,21) с тем, чтобы проверить разрешающую способность той или иной спектральной оценки. Мощность двух более слабых синусоид с частотами 0,1 и -0,15 на 20 дБ меньше мощности двух более сильных. Эти слабые синусоиды используются для испытания способности спектральных оценок обеспечить обнаружение слабых компонент сигналов на фоне сильных сигналов. Окрашенный шумовой процесс был сформирован пропусканием двух независимых белых шумов через идентичные фильтры с частотными характеристиками типа приподнятой косинусоиды для раздельного получения действительной и мнимой составляющих шумового процесса в тест-последовательности. Высота линий, характеризующих синусоидальные составляющие выбрана так, чтобы отображать мощность каждой синусоиды относительно мощности окрашенного шума. Следовательно, мощность каждой из сильных синусоид превышает мощность шума, а мощность каждой из 2-х слабых синусоид меньше мощности шума.

Приведем примеры классических периодограммного и коррелограммного методов спектрального оценивания СПМ тест-последовательности.

На рис. 5.29 приведены две периодограммы тест-последовательности Марпла.

Периодограмма этой последовательности, соответствующая максимальному разрешению и минимальной устойчивости, показана на рис.5.29а. Она получена для случая только одного сегмента (полной записи данных) без применения окна. Заметим, что близкие по частоте синусоиды с разном $0,01$ не разрешены, поскольку предельное разрешение в данном случае составляет величину $1/64=0,0156$. Слабая синусоида с частотой $0,1$ маскируется боковыми лепестками сильных синусоид. На рис. 5.29 б показана гладкая усредненная периодограмма с низким разрешением, которая получена разбиением записи данных на три сегмента по 32 отсчета в каждом, взвешенным окном Хэмминга, с перекрытием сегментов, равным 16 отсчетов. На этой периодограмме отчетливо видны слабые синусоиды.

Коррелограммные оценки СПМ тест-последовательности показаны на рис.5.30 а, б для 16 и 32 значений автокорреляции соответственно и взвешивания окном Хэмминга. Графики на рис 5.30 дают полное представление о взаимосвязи между разрешением и гладкостью оценки.

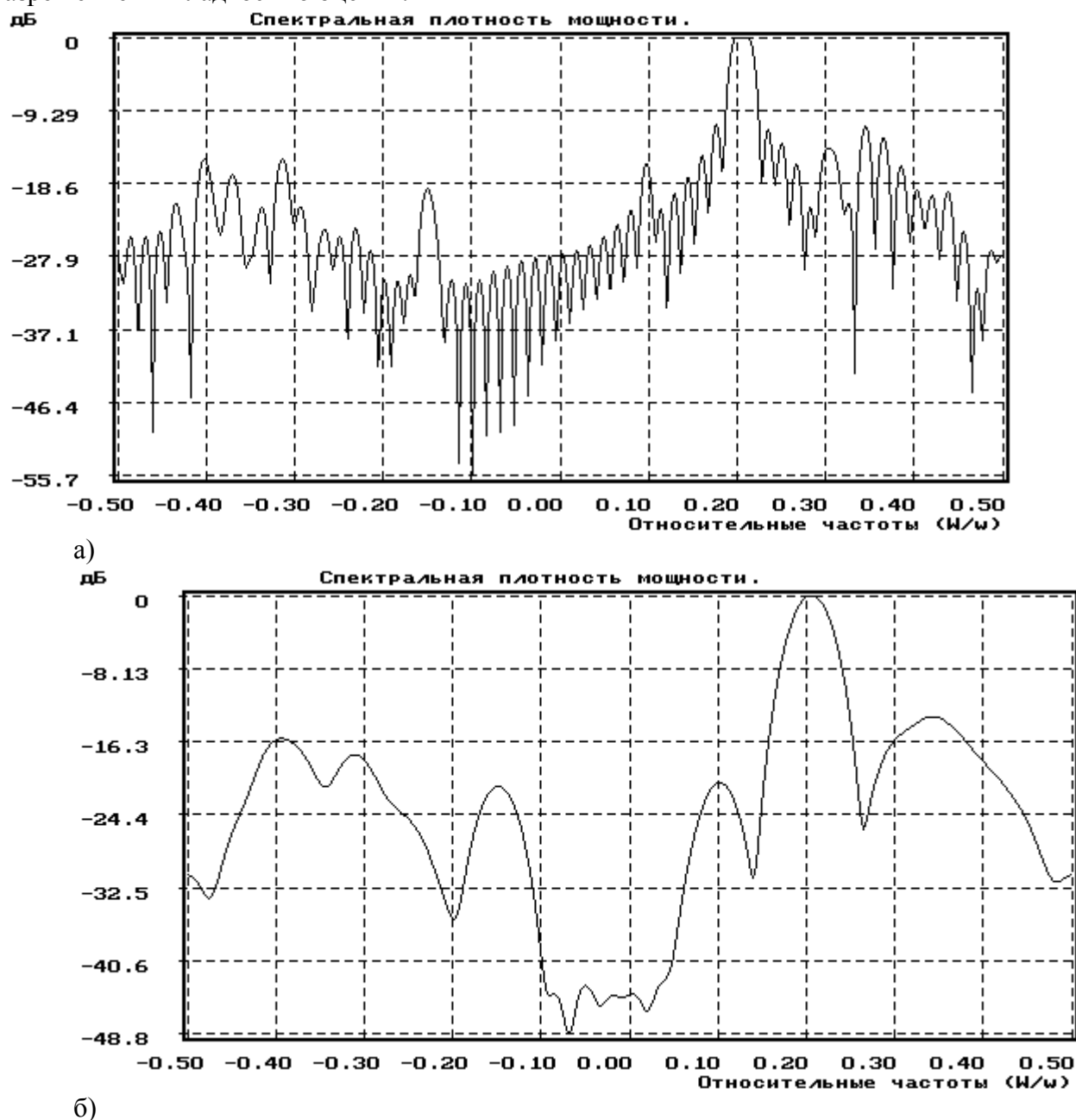


Рис. 5.29. Периодограммные оценки СПМ для 64-точечной тест-последовательности: а- один сегмент, окно не применялось; б – три сегмента, окно Хэмминга.

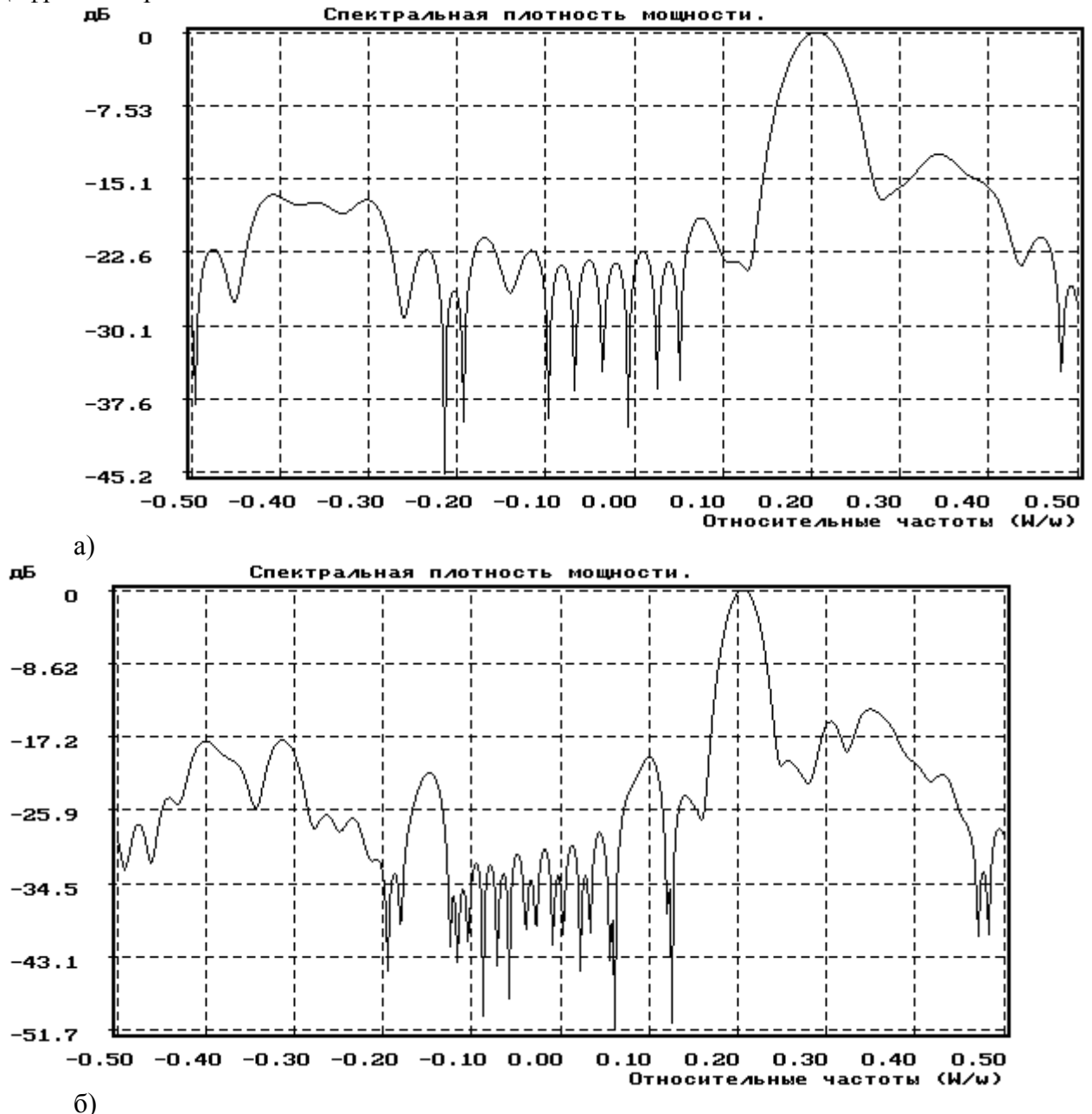


Рис. 5.30. Коррелограммные оценки СПМ для 64-точечной тест-последовательности: а – 16 значений автокорреляции, окно Хэмминга; б – 32 значения автокорреляции, окно Хэмминга.

Заметим, что если не известно как выбирать компромиссное решение относительно устойчивости и разрешения оценок, то рекомендуется начать с выбора максимальной устойчивости (наиболее гладкой оценки), а это значит – с минимального разрешения, а затем постепенно увеличивать длину сегментов (в случае периодограммы) или число корреляционных членов (в случае коррелограммы). Эта процедура будет увеличивать дисперсию оценки и уменьшать ширину спектральных пиков (повышать разрешение). Следует продолжать указанную процедуру до тех пор, пока достигнутое разрешение не позволит передать наиболее тонкие и важные детали спектра и дальнейшее увеличение разрешения к существенным изменениям в спектральных оценках уже не приводит.

6. Современные методы цифрового спектрального анализа высокого разрешения

6.1. Введение

Оценивание спектральной плотности мощности (СПМ) или просто спектра дискретизованных детерминированных и случайных процессов обычно выполняется с помощью процедур, использующих быстрое преобразование Фурье (БПФ). Такой подход к спектральному анализу эффективен в вычислительном отношении и обеспечивает получение приемлемых результатов для большого класса сигнальных процессов. Однако, несмотря на указанные достоинства, подходу, основанному на вычислении БПФ, присущ ряд принципиальных ограничений. Наиболее важное из них - это ограничение частотного разрешения, т.е. способности различать спектральные линии двух и более сигналов. Частотное разрешение в герцах примерно равно величине, обратной временному интервалу в секундах, на котором можно получить отсчеты сигнала. Второе ограничение обусловлено неявной весовой обработкой данных при вычислении БПФ. Взвешивание проявляется в виде «утечки» в частотной области, т.е. энергия главного лепестка спектральной линии «утекает» в боковые лепестки, что приводит к наложению и искажению спектральных линий других присутствующих сигналов. При этом спектральные линии слабых сигналов могут маскироваться боковыми лепестками спектральных линий более сильных сигналов. Правильный выбор функции окна, значения которой спадают на краях, позволяют ослабить утечку в боковые лепестки, однако лишь за счет снижения разрешающей способности.

Два указанных ограничения подходов на основе БПФ особенно сильно проявляются при анализе коротких записей данных, с которыми чаще всего и приходится иметь дело на практике. Многие измеряемые процессы обладают малой длительностью или же медленно изменяющимися во времени спектрами, которые можно считать постоянными только на коротких участках записей данных. Например, для изучения характеристик внутриимпульсной модуляции в радиолокаторах в пределах очень короткого радиолокационного импульса можно осуществить лишь несколько временных отсчетов. В случае гидролокатора можно сделать большее число подобных отсчетов, но движение цели заставляет ограничиваться короткими интервалами наблюдения, с тем чтобы гарантировать неизменность статистик цели на интервале анализа.

В течение трех последних десятилетий предложено большое число самых различных процедур спектрального оценивания, которые разработаны для того, чтобы так или иначе ослабить ограничения, присущие подходу на основе БПФ. Часто о процессе, из которого берутся отсчеты, известна некоторая совокупность сведений, позволяющая выбрать модель процесса, которая является хорошей его аппроксимацией. В этом случае можно, как правило, получить более точную спектральную оценку, определяя параметры выбранной модели по результатам измерений. Так называемый моделирующий подход к спектральному оцениванию становится трехэтапной процедурой. Первый этап состоит в выборе модели исследуемого временного ряда. Второй этап состоит в оценивании параметров принятой модели либо с использованием имеющихся отсчетов данных, либо значений автокорреляционной функции (известной или оцениваемой по имеющимся данным). Наконец, третий этап состоит в получении спектральной оценки путем подстановки оценок параметров модели в расчетное выражение для СПМ, соответствующее этой модели. Ценность такого подхода в том, что при хорошем соответствии выбранной модели наблюдаемым данным получаются более точные оценки СПМ при более высоком разрешении, чем при традиционном подходе на основе БПФ, поскольку отпадает необходимость в функциях окна и устраняются связанные с ними искажения.

Платой за улучшение оценок СПМ оказывается возрастание вычислительной сложности алгоритмов получения оценок, которая стала успешно преодолеваться только на современном уровне развития вычислительной техники. Современный цифровой спектральный анализ - это оценка СПМ на основе параметрических моделей случайных процессов. Основной интерес к методам параметрического спектрального оценивания связан с высокой разрешающей способностью, достигаемой с их помощью при обработке последовательностей данных, содержащих очень малое число отсчетов.

6.2. Параметрические модели случайных процессов

Модель временного ряда, которая пригодна для аппроксимации многих встречающихся на практике случайных процессов с дискретным временем, описывается выходом фильтра, выражаемым следующим линейным разностным уравнением с комплексными коэффициентами:

$$x[n] = -\sum_{k=1}^p a[k]x[n-k] + \sum_{k=0}^q b[k]u[n-k] = \quad (6.1)$$

$$= \sum_{k=0}^{\infty} h[k]u[n-k], \quad (6.2)$$

где $x[n]$ - последовательность на выходе физически реализуемого фильтра,

$u[n]$ - входная возбуждающая последовательность (белый шум с нулевым средним и дисперсией ρ_w), $h[n]$ - импульсная характеристика фильтра ($h[k] = 0$, при $k < 0$, $h[0] = 1$), $a[k]$ - коэффициенты авторегрессии, $b[k]$ - коэффициенты скользящего среднего ($b[0] = 1$). Заметим, что здесь и в дальнейшем изложении в обозначениях функций дискретный аргумент заключен в квадратные [дискретные] скобки, а непрерывный аргумент - в круглые (непрерывные) скобки, в соответствии с обозначениями, принятыми в книге С.Л.Марпла [9].

Системная функция, связывающая вход и выход этого фильтра, имеет рациональную форму (отношение полиномов) :

$$H(z) = \frac{B(z)}{A(z)}, \quad (6.3)$$

где
$$A(z) = 1 + \sum_{k=1}^p a[k]z^{-k}, \quad (6.4)$$

$$B(z) = 1 + \sum_{k=1}^q b[k]z^{-k}, \quad (6.5)$$

$$H(z) = 1 + \sum_{k=1}^{\infty} h[k]z^{-k}. \quad (6.6)$$

При этом полагается, что нули $A(z)$, $B(z)$ расположены внутри единичной окружности в Z - плоскости с тем, чтобы гарантировать принадлежность функции $H(z)$ устойчивому минимально - фазовому каузальному фильтру. Процесс на выходе фильтра (6.1) называют соответствующим модели авторегрессии - скользящего среднего (АРСС), параметры $a[k]$ характеризуют авторегрессионную часть этой модели порядка p , а параметры $b[k]$ - ее часть, соответствующую скользящему среднему порядка q .

Спектральная плотность мощности для АРСС-модели имеет вид

$$P_{APCC}(f) = T\rho_w |B(f)/A(f)|^2, \quad (6.8)$$

где

$$A(f) = 1 + \sum_{k=1}^p a[k] \exp(-j2\pi kTf),$$

$$B(f) = 1 + \sum_{k=1}^q b[k] \exp(-j2\pi kTf),$$
(6.9)

$$-\frac{1}{2T} \leq f \leq \frac{1}{2T}, \quad T - \text{период взятия отсчетов.}$$

Можно заметить, что значения AP -параметров $a[k]$ ($k=1,2,\dots,p$), CC -параметров $b[k]$ ($k=1,2,\dots,q$) и дисперсии белого шума ρ_w полностью характеризуют спектральную плотность мощности $APCC$ -процесса $x[n]$.

Найдем соотношения, связывающие параметры $APCC$ -процесса с его автокорреляционной последовательностью (АКП) $r_{xx}[m]$. Домножая обе части уравнения (6.1) на $x^*[n-m]$, где $*$ - знак комплексного сопряжения и определяя математическое ожидание, получим соотношение

$$r_{xx}[m] = -\sum_{k=1}^p a[k] r_{xx}[m-k] + \sum_{k=0}^q b[k] r_{ux}[m-k], \quad (6.10)$$

Поскольку $u[k]$ - белый шум, то функция взаимной корреляции между входным и выходным процессами фильтра примет вид

$$r_{ux}[i] = \begin{cases} 0, & i > 0; \\ \rho_w, & i = 0; \\ \rho_w h^*[-i], & i < 0, \end{cases}$$

Отсюда получаем окончательное выражение, связывающее параметры $APCC$ -модели и автокорреляционную последовательность процесса $x[k]$:

$$r_{xx}[m] = \begin{cases} r_{xx}^*[-m], & m < 0, \\ -\sum_{k=-1}^p a[k] r_{xx}[m-k] + \rho_w \sum_{k=m}^q b[k] h^*[k-m], & 0 \leq m \leq q, \\ -\sum_{k=1}^p a[k] r_{xx}[m-k], & m > q, \end{cases} \quad (6.11)$$

где учтено, что $h[0] = 1$.

AP -параметры $APCC$ -модели и автокорреляционная последовательность связаны системой линейных уравнений. Выражение (6.11) можно, например, записать для p значений индекса временного сдвига $q+1 \leq m \leq q+p$ и затем представить в матричной форме

$$\begin{bmatrix} r_{xx}[q] & r_{xx}[q-1] & \dots & r_{xx}[q-p+1] \\ r_{xx}[q+1] & r_{xx}[q] & \dots & r_{xx}[q-p+2] \\ \vdots & \vdots & \ddots & \vdots \\ r_{xx}[q+p-1] & r_{xx}[q+p-2] & \dots & r_{xx}[q] \end{bmatrix} \begin{bmatrix} d[1] \\ d[2] \\ \vdots \\ d[p] \end{bmatrix} = \begin{bmatrix} r_{xx}[q+1] \\ r_{xx}[q+2] \\ \vdots \\ r_{xx}[q+p] \end{bmatrix} \quad (6.12)$$

Таким образом, если задана автокорреляционная последовательность для $q-p+1 \leq m \leq q+p$, то AP -параметры можно найти отдельно от CC -параметров как решение системы линейных уравнений (6.12). Уравнения (6.12) называются *нормальными уравнениями Юла-Уолкера для $APCC$ -процесса*; иногда их также

называют *модифицированными уравнениями Юла-Уолкера*. Следует заметить, что CC -параметры не являются, к сожалению, решением системы линейных уравнений, поскольку в (6.12) они входят в виде свертки с коэффициентами импульсной характеристики $h[k]$, что приводит к нелинейной связи с автокорреляционной последовательностью.

Соотношения (6.12) позволяют сделать вывод о том, что для $APCC(p, q)$ -процесса задание p последовательных значений автокорреляционной функции позволяет однозначно продолжить ее до бесконечности с помощью рекуррентного соотношения. Именно этим обстоятельством можно объяснить высокую разрешающую способность оценок СПМ на основе $APCC$ -модели в сравнении с классическими процедурами получения оценок СПМ, в которых АКП вне интервала наблюдения полагается нулевой или периодически повторяемой, что, как правило, не соответствует поведению реальной АКП.

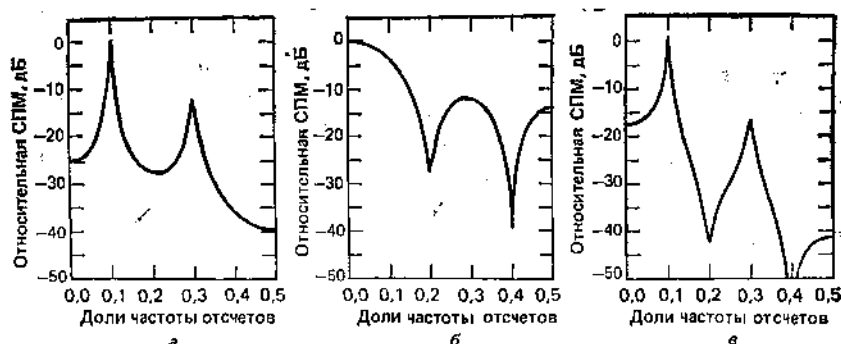


Рис.6.1. Типичные параметрические модели спектра: а — AP (4)-спектр; б — CC (4)-спектр; в — APCC (4, 4) -спектр.

Если все AP -параметры, за исключением $a(0)=1$, положить равными нулю, то уравнение (6.1) преобразуется к виду

$$x[n] = \sum_{k=1}^q b[k]u[n-k] + u[n],$$

и будет соответствовать модели процесса скользящего среднего порядка q , или $CC(q)$ -процессу. Спектральная плотность мощности CC -процесса получается из (6.8) при $p=0$ в виде:

$$P_{CC}(f) = T\rho_w |B(f)|^2.$$

Соотношения связи CC -параметров $b(k)$ и АКП получатся из (6.12) при $p=0$ в виде следующих нелинейных уравнений

$$r_{xx}[m] = \begin{cases} r_{xx}^*[-m], & m < 0, \\ \rho_w \sum_{k=0}^q b[k]b[n-k], & 0 < m < q, \\ 0, & |m| > q. \end{cases}$$

Наконец, если в уравнении (6.1) все CC -параметры, кроме $b(0)=1$, положить равными нулю, получим модель процесса авторегрессии порядка p или $AP(p)$ -процесс, рассмотрению которого посвящен следующий раздел. На рис.6.1 показаны спектры типичных $APCC$ -, CC - и AP -процессов. Отметим острые пики, характерные для AP -спектров, и глубокие провалы, характерные для CC -спектров. $APCC$ -спектр, показанный на рис.6.1в, представляет собой результат объединения AP - и CC -спектров, показанных на рис.6.1а и 6.1б. $APCC$ -спектр пригоден для моделирования как острых пиков, так и глубоких провалов.

6.3. Авторегрессионный процесс и свойства его спектра

Из всех описанных выше моделей временных рядов наибольшее внимание в

технической литературе уделяется AP -процессам по двум причинам. Во-первых, AP -модель применяется для спектрального оценивания, если необходимы спектры с острыми пиками, что часто связывается с высоким частотным разрешением. Кроме того, оценки AP -параметров получаются из решения системы линейных уравнений, в отличие от других моделей. Итак, AP -процесс описывается следующим линейным разностным уравнением с комплексными коэффициентами, которое получается, если в уравнении (6.1) все CC -параметры, за исключением $b[0]=1$, положить равными нулю :

$$x[n] = -\sum_{k=1}^p a[k]x[n-k] + u[n], \quad (6.13)$$

где $x[n]$ - AP - последовательность на выходе каузального фильтра, который формирует наблюдаемые данные;
 $u[n]$ - входная возбуждающая последовательность, соответствующая белому шуму с нулевым средним и дисперсией ρ_w .

Если в уравнении (6.8) положить $q = 0$, то получим спектральную плотность мощности AP -процесса:

$$P_{AP}(f) = \frac{T\rho_w}{|A(f)|^2} = \frac{T\rho_w}{\mathbf{e}_p^H(f) \mathbf{a} \mathbf{a}^H \mathbf{e}_p(f)}, \quad (6.14)$$

где полином $A(f) = 1 + \sum_{k=1}^p a(k) \exp(-2j\pi f k T)$, вектор комплексных синусоид $\mathbf{e}_p(f)$ и вектор параметров $\mathbf{a}$ имеют следующий вид:

$$\mathbf{e}_p(f) = \begin{bmatrix} 1 \\ \exp(2j\pi f T) \\ \vdots \\ \exp(2j\pi f p T) \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} 1 \\ a[1] \\ \vdots \\ a[p] \end{bmatrix},$$

а надстрочный символ «^H» означает эрмитово сопряжение (или эрмитово транспонирование) вектора, получаемое в результате комплексного сопряжения его элементов с последующей их транспозицией, т.е. образованием вектор-строки.

Полагая в (6.12) $q = 0$, получаем уравнение, связывающее автокорреляционную последовательность с параметрами автокорреляционной модели:

$$r_{xx}[m] = \begin{cases} -\sum_{k=1}^p a[k] r_{xx}[m-k], & m > 0, \\ -\sum_{k=1}^p a[k] r_{xx}[-k] + \rho_w, & m = 0, \\ r_{xx}^*[-m], & m < 0. \end{cases} \quad (6.15)$$

Это выражение можно записать для $p+1$ значений индекса временного сдвига $0 \leq m \leq p$, затем представить в матричной форме

$$\begin{bmatrix} r_{xx}[0] & r_{xx}[-1] & \cdot & \cdot & \cdot & r_{xx}[-p] \\ r_{xx}[1] & r_{xx}[0] & \cdot & \cdot & \cdot & r_{xx}[-p+1] \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ r_{xx}[p] & r_{xx}[p-1] & \cdot & \cdot & \cdot & r_{xx}[0] \end{bmatrix} * \begin{bmatrix} 1 \\ a[1] \\ \cdot \\ \cdot \\ \cdot \\ a[p] \end{bmatrix} = \begin{bmatrix} \rho_w \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}. \quad (6.16)$$

Таким образом, если задана автокорреляционная последовательность для $0 \leq m \leq p$, то AP -параметры можно найти в результате решения уравнения (6.16), которые называются *нормальными уравнениями Юла-Уолкера для AP -процесса*. Автокорреляционная матрица в (6.16) является теплицевой и эрмитовой, поскольку $r_{xx}^*[k] = r_{xx}[-k]$. Очевидно, что для $СПМ$ AP -процесса справедливы следующие эквивалентные выражения:

$$P_{AP}(f) = \frac{T\rho_w}{|A(f)|^2} = T \sum_{k=-\infty}^{\infty} r_{xx}[k] \exp(2j\pi f k T). \quad (6.17)$$

Заметим, что значения автокорреляции, соответствующие индексам временного сдвига от 0 до p , позволяют определить из уравнения Юла-Уолкера дисперсию белого шума ρ_w и AP -параметры $a[1], a[2], \dots, a[p]$, а затем по (6.17) вычислить AP $СПМ$. Можно также рассчитать значения автокорреляции для $n > p$ по соотношению

$$r_{xx}[m] = -\sum_{k=1}^p a[k] r_{xx}[m-k], \quad (6.18)$$

и далее воспользоваться второй частью (6.17) для вычисления AP $СПМ$, хотя это и не всегда эффективно на практике. Здесь уместно сравнить AP $СПМ$ с оценкой $СПМ$, полученной классическим коррелограммным методом. Напомним, что этот метод позволяет по $p+1$ значениям автокорреляции получить оценку $СПМ$ в виде

$$P_{кор}(f) = T \sum_{k=-p}^p r_{xx}[k] \exp(-2j\pi f k T).$$

Можно заметить, что в коррелограммном методе значения $АКП$ вне интервала суммирования, то есть для $|k| > p$, полагаются нулевыми, в то время как для AP -оценки они экстраполируются в соответствие с (6.18). Применением этой ненулевой экстраполяции $АКП$ при вычислении AP $СПМ$ с помощью (6.17) и объясняется то высокое разрешение, которое характерно для оценок AP $СПМ$. Поскольку при получении оценок AP $СПМ$ не используется обработка $АКП$ с помощью функции окна, им не свойственны эффекты, вызванные наличием боковых лепестков, всегда присутствующих в классических спектральных оценках.

6.3.1. Связь с анализом, основанным на линейном предсказании

Уравнения, соответствующие линейному предсказанию, по своей структуре идентичны уравнениям Юла-Уолкера для авторегрессионного процесса, а потому существует тесная связь между фильтром линейного предсказания и AP -процессом.

Рассмотрим оценки линейного предсказания вперед:

$$\hat{x}^f[n] = -\sum_{k=1}^m a^f[k] x[n-k] \quad (6.19)$$

где крышка « $\wedge$ » обозначает оценку, надстрочный индекс f используется для обозначения оценки вперед. Предсказание вперед понимается в том смысле, что оценка, соответствующая временному индексу n , вычисляется по m предыдущим временным отсчетам. Комплексная ошибка линейного предсказания вперед:

$$e^f[n] = x[n] - \hat{x}^f[n], \quad (6.20)$$

имеет действительную дисперсию:

$$\rho^f = \langle |e^f[n]|^2 \rangle, \quad (6.21)$$

Коэффициенты линейного предсказания вперед, минимизирующие дисперсию ошибки (6.21), определяются из системы нормальных уравнений в матричном виде:

$$\begin{bmatrix} r_{xx}[0] & r_{xx}^*[1] & \cdot & \cdot & \cdot & r_{xx}^*[m] \\ r_{xx}[1] & r_{xx}[0] & \cdot & \cdot & \cdot & r_{xx}^*[m-1] \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{xx}[m] & r_{xx}[m-1] & \cdot & \cdot & \cdot & r_{xx}[0] \end{bmatrix} * \begin{bmatrix} 1 \\ a^f[1] \\ \cdot \\ \cdot \\ \cdot \\ a^f[m] \end{bmatrix} = \begin{bmatrix} \rho^f \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}. \quad (6.22)$$

Можно заметить, что эти матричные уравнения по своей структуре идентичны уравнениям Юла-Уолкера (6.16) для авторегрессионного процесса. Если выражение (6.20) переписать в виде

$$x[n] = -\sum_{k=1}^m a[k]x[n-k] + e^f[n], \quad (6.23)$$

то можно заметить его подобие уравнению (6.13) для авторегрессионного процесса. В уравнении (6.13) последовательность $u[n]$ соответствует белому шумовому процессу, который используется в качестве входного воздействия авторегрессионного фильтра, а $x[n]$ - представляет собой выходной сигнал фильтра. В отличие от (6.13) в уравнении (6.23) последовательность значений ошибки $e^f[n]$ представляет собой выход фильтра линейного предсказания ошибки вперед, а $x[n]$ - входное воздействие фильтра предсказания ошибки. Если последовательность $x[n]$ генерируется как $AP(p)$ - процесс с $m=p$, то последовательность значений ошибки будет белым шумовым процессом, коэффициенты линейного предсказания вперед будут идентичны AP -параметрам ($a^f[k] = a[k]$), а фильтр предсказания ошибки можно рассматривать как фильтр, отбеливающий процесс $x[n]$. На рис.6.2 показаны фильтр, формирующий AP -процесс из белого шума, и фильтр, отбеливающий AP -процесс.

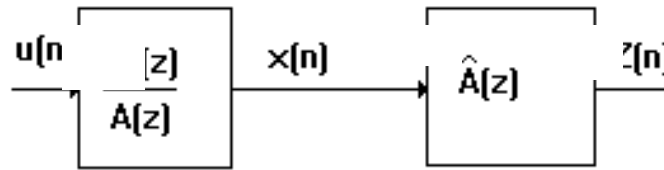


Рис. 6.2.

Аналогичные рассуждения можно провести относительно оценки линейного предсказания назад

$$x^b[n] = -\sum_{k=1}^m a^b[k]x[n+k], \quad (6.24)$$

которая определяется по m последующим временным отсчетам, ввести ошибку линейного предсказания назад

$$e^b[n] = x[n-m] - \hat{x}^b[n-m] = x[n-m] + \sum_{k=1}^m a^b[k]x[n-m+k] \quad (6.25)$$

и показать, что коэффициенты линейного предсказания назад $a^b[k]$ минимизирующие дисперсию ошибки $\rho^b = \langle |e^b[n]|^2 \rangle$, будут комплексно - сопряженными величинами коэффициентам линейного предсказания вперед

$a^b[k] = (a^f[k])^*$, где $1 \leq k \leq m$, а дисперсии ошибок одинаковы $\rho^b = \rho^f$ [9].

6.4. Методы авторегрессионного спектрального оценивания

Рассмотренные свойства AP -модели, позволяют рассчитать ее параметры и, следовательно, функцию $СПМ$ по известным значениям $АКП$, исследуемого случайного процесса. При практических измерениях эта функция ($АКП$) обычно неизвестна, поэтому разработано большое количество методов нахождения AP $СПМ$, по имеющимся отсчетам данных. Все эти методы можно разбить на два класса: алгоритмы для обработки *блоков* данных и алгоритмы для обработки *последовательных* данных. Ниже описаны методы, предназначенные для обработки целых блоков накопленных отсчетов данных некоторого фиксированного объема. Блочные методы можно описать как алгоритмы с фиксированным временем, рекурсивные относительно порядка в том смысле, что они применяются к фиксированным блокам временных отсчетов данных и позволяют рекурсивным образом получать оценки параметров AP - модели более высокого порядка по оценкам параметров AP - модели более низкого порядка. С другой стороны, последовательные методы можно рассматривать как алгоритмы с фиксированным порядком, рекурсивные относительно времени в том смысле, что они применяются для последовательной обработки данных с целью обновления оценок параметров AP - модели фиксированного порядка. Применение таких алгоритмов целесообразно для слежения за спектрами, медленно изменяющимися во времени.

6.4.1. Метод Юла-Уолкера

Наиболее очевидный подход к AP - оцениванию $СПМ$ состоит в решении уравнений Юла-Уолкера (6.16), в которые вместо значений неизвестной автокорреляционной функции подставляются их оценки. Так для отсчетов данных $x[0], x[1], \dots, x[N-1]$ можно получить оценки автокорреляции в форме:

$$\hat{r}_{xx}[m] = \begin{cases} \frac{1}{(N-m)} \sum_{n=0}^{N-m-1} x[n+m]x^*[n], & 0 \leq m \leq N-1; \\ \frac{1}{(N-|m|)} \sum_{n=0}^{N-|m|-1} x^*[n+|m|]x[n], & 1-N \leq m < 0. \end{cases} \quad (6.26)$$

Эти оценки являются несмещенными, поскольку $\langle \hat{r}_{xx}[m] \rangle = r_{xx}[m]$ и состоятельными, поскольку при неограниченном возрастании N , дисперсия оценки стремится к нулю. Другой вариант получения оценок автокорреляции, имеет вид:

$$\tilde{r}_{xx}[m] = \begin{cases} \frac{1}{N} \sum_{n=0}^{N-m-1} x[n-m]x^*[n], & 0 \leq m \leq N-1; \\ \frac{1}{N} \sum_{n=0}^{N-|m|-1} x^*[n-|m|]x[n], & 1-N \leq m < 0. \end{cases} \quad (6.27)$$

При конечном N , эта оценка будет смещенной, поскольку $\langle \tilde{r}_{xx}[m] \rangle = (1 - \frac{|m|}{N})r_{xx}[m]$. При использовании смещенных оценок автокорреляции получаемая оценка AP -параметров, всегда соответствуют устойчивому AP - фильтру, что для несмещенных оценок не всегда имеет место.

Поскольку автокорреляционная матрица в системе уравнений (6.16) по своей структуре является теплоцевой, так как элементы любой ее диагонали одинаковы, и эрмитовой, так как $r_{xx}[-k] = r_{xx}^*[k]$, то для получения решения $\rho_w, a[1], a[2], \dots, a[p]$ при подстановке оценок автокорреляции в (6.16) можно использовать весьма эффективный рекурсивный алгоритм Левинсона. Рекурсивное решение уравнений Юла-Уолкера методом Левинсона связывает AP - параметры порядка p с параметрами

порядка $(p-1)$ соотношением:

$$a_p[n] = a_{p-1}[n] + K_p a_{p-1}^*[p-n], \quad (6.28)$$

где n изменяется от 1 до $(p-1)$. Коэффициент K_p , получивший название коэффициента отражения, определяется по значениям автокорреляции, соответствующие сдвигам от 0 до $(p-1)$:

$$K_p = a_p[p] = -\frac{1}{\rho_{p-1}} \sum_{n=0}^{p-1} a_{p-1}[n] r_{xx}[p-n], \quad (6.29)$$

Рекурсивное уравнение для дисперсии белого шума имеет вид:

$$\rho_p = \rho_{p-1}(1 - |K_p|^2), \quad (6.30)$$

с начальным условием $\rho_0 = r_{xx}[0]$.

В рекурсии Левинсона без дополнительных вычислительных затрат находятся AP -коэффициенты всех моделей, порядок которых $m=1, 2, \dots, p$. Можно также заметить, что коэффициенты $AP(p)$ -модели могут быть определены по известным (вычисленным) величинам $r_{xx}(0)$ и коэффициентам отражения $K_1, K_2, \dots, K_p$. Поэтому эти коэффициенты также полностью определяют $AP(p)$ -процесс.

Таким образом, метод Юла-Уолкера отвечает совокупности соотношений (6.26) - (6.30) для определения AP -параметров, по которым в соответствии с (6.17), определяется оценка $СПМ$. В случае длинных записей данных, метод Юла-Уолкера, может давать вполне приемлемые спектральные оценки, однако для коротких записей данных, получаемые с его помощью, спектральные оценки имеют худшее разрешение, по сравнению с оценками, получаемыми другими AP -методами. В [9] приведена программа YULEWALKER, реализующая метод Юла-Уолкера.

6.4.2. Метод Берга

Один из первых алгоритмов, послужившим толчком к активному исследованию методов авторегрессионного спектрального оценивания, был предложен Бергом [9], его иногда называют алгоритмом максимальной энтропии. Идея алгоритма использует тот факт, что в рассмотренных выше формулах (6.28) - (6.30) только коэффициент отражения K_p непосредственно зависит от автокорреляционной функции $АКП$, а это означает, что одна из процедур получения AP -оценки $СПМ$ в том случае, когда имеется некоторый блок отсчетов данных, может быть основана на оценивании коэффициента отражения по этим отсчетам на каждом шаге рекурсии Левинсона.

Подставляя в уравнения (6.20) и (6.25) значения $a_p[n]$, определяемые выражением (6.28), получаем рекурсивные соотношения:

$$e_p^f[n] = e_{p-1}^f[n] + K_p e_{p-1}^b[n-1], \quad (6.31)$$

$$e_p^b[n] = e_{p-1}^b[n-1] + K_p^* e_{p-1}^f[n], \quad (6.32)$$

которые связывают ошибки предсказания порядка p с ошибками предсказания порядка $(p-1)$.

В алгоритме Берга используется оценка коэффициента отражения, определяемая по методу наименьших квадратов. При каждом значении порядка p в нем минимизируется среднее арифметическое мощности ошибок линейного предсказания вперед и назад (выборочная дисперсия ошибки предсказания):

$$\rho_p^{fb} = \frac{1}{2N} \left[\sum_{n=p+1}^N |e_p^f[n]|^2 + \sum_{n=p+1}^N |e_p^b[n]|^2 \right], \quad (6.33)$$

ρ_p^{fb} - является функцией только одного параметра комплексного коэффициента отражения K_p . Приравнявая комплексную производную от ρ_p^{fb} к нулю:

$$\frac{d\rho_p^{fb}}{d\text{Re}\{K_p\}} + j \frac{d\rho_p^{fb}}{d\text{Im}\{K_p\}} = 0,$$

и решая полученное уравнение относительно K_p , получаем следующее выражение для оценки по методу наименьших квадратов:

$$\hat{K}_p = \frac{-2 \sum_{n=p+1}^N e_{p-1}^f[n] e_{p-1}^{b*}[n-1]}{\sum_{n=p+1}^N |e_{p-1}^f[n]|^2 + \sum_{n=p+1}^N |e_{p-1}^b[n-1]|^2}, \quad (6.34)$$

В (6.34) предполагается, что имеется N отсчетов данных $x[1], \dots, x[N]$ и ошибки предсказания формируются в диапазоне индексов от $n=p+1$ до $n=N$, поскольку используются только имеющиеся отсчеты данных. Таким образом, алгоритм Берга использует рекурсивный алгоритм Левинсона, в котором вместо K_p , вычисляемого по АКП используется его оценка (6.34). Базовый алгоритм Левинсона дополняется уравнениями (6.31) и (6.32), вычисления по которым начинаются с $e_0^f[n] = e_0^b[n] = x[n]$, $1 \leq n \leq N$. Начальное значение дисперсии ошибки предсказания равно

$$\rho = \frac{1}{N} \sum_{n=1}^N |x[n]|^2.$$

Оценка коэффициента отражения (6.34) представляет собой гармоническое среднее коэффициентов частной корреляции ошибок предсказания вперед и назад. Рекурсивная формула, которая упрощает вычисление знаменателя в выражении для оценки (6.34)

$$\begin{aligned} DEN_p &= \sum_{n=p+1}^N \left\{ |e_{p-1}^f[n]|^2 + |e_{p-1}^b[n-1]|^2 \right\}; \\ DEN_p &= (1 - |\hat{K}_p|^2) DEN_{p-1} - |e_{p-1}^f[n]|^2 - |e_{p-1}^b[N]|^2 \end{aligned} \quad (6.35)$$

Гармонический метод дает несколько смещенные оценки частоты синусоид. Для уменьшения этого смещения предложено взвешивание среднего квадрата ошибки предсказания:

$$\rho_p^{fb} = \frac{1}{2N} \sum_{n=p+1}^N W_p[n] \left\{ |e_p^f[n]|^2 + |e_p^b[n]|^2 \right\}, \quad (6.36)$$

что приводит к следующей оценке коэффициента отражения:

$$\hat{K}_p = \frac{-2 \sum_{n=p+1}^N W_{p-1}[n] e_{p-1}^f[n] e_{p-1}^{b*}[n-1]}{\sum_{n=p+1}^N W_{p-1}[n] (|e_{p-1}^f[n]|^2 + |e_{p-1}^b[n-1]|^2)}, \quad (6.37)$$

где $W_{p-1}[n]$ - определяет весовую функцию. Показано [9], что частотное смещение уменьшается при использовании окна Хэмминга. В [9] приведена программа BURG, реализующая метод Берга.

6.4.3 Ковариационный метод

Налагая на AP - коэффициенты ограничения, с тем чтобы они удовлетворяли рекурсивному соотношению Левинсона, Бергу удалось осуществить оптимизацию по методу наименьших квадратов единственного параметра - коэффициента отражения. Другой подход состоит в минимизации в методе наименьших квадратов одновременно по всем коэффициентам линейного предсказания, что позволяет полностью устранить ограничение, налагаемое рекурсией Левинсона. Такой подход будет несколько улучшать характеристики спектральной оценки.

Предположим, что для оценивания AP - параметров порядка p используется N -

точечная последовательность данных $x[1], \dots, x[N]$. Оценка линейного предсказания вперед для отсчета $x[n]$ будет иметь форму

$$\hat{x}^f[n] = -\sum_{k=1}^p a^f[k]x[n-k]. \quad (6.38)$$

Ошибка линейного предсказания вперед определяется выражением:

$$e_p^f[n] = x[n] - \hat{x}^f[n] = x[n] + \sum_{k=1}^p a_p^f[k]x[n-k], \quad (6.39)$$

Ошибку линейного предсказания вперед можно определить в диапазоне временных индексов от $n=1$ до $n=N+p$, если предположить, что данные до первого и после последнего отсчетов равны нулю (т.е. $x[n]=0$, при $n < 1, n > N$). $N+p$ - членов ошибки линейного предсказания вперед, определяемых выражением (6.39), можно записать, используя матрично-векторное обозначение, в следующем виде:

$$\begin{bmatrix} e_p^f[1] \\ \vdots \\ e_p^f[p+1] \\ \vdots \\ e_p^f[N-p] \\ \vdots \\ e_p^f[N] \\ \vdots \\ e_p^f[N+p] \end{bmatrix} = \begin{bmatrix} x[1] & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x[p+1] & \cdot & \cdot & \cdot & x[1] \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x[N-p] & \cdot & \cdot & \cdot & x[p+1] \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x[N] & \cdot & \cdot & \cdot & x[N-p] \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & x[N] \end{bmatrix} * \begin{bmatrix} 1 \\ \cdot \\ \cdot \\ a[1] \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a[p] \end{bmatrix},$$

X_p

где X_p - прямоугольная теплица $(N+p) \times (p+1)$ - матрица данных. Модуль среднего квадрата ошибки линейного предсказания вперед, который необходимо минимизировать, это величина:

$$\rho_p^f = \sum_n |e_p^f[n]|^2. \quad (6.41)$$

Поделив (6.41) на N , получим выборочную дисперсию. Выбор диапазона суммирования в (6.41) зависит от конкретного применения. Выбирая полный диапазон суммирования от $e_p^f[1]$ до $e_p^f[N+p]$, получаем так называемый взвешенный случай, поскольку отсутствующие значения данных приравниваются к нулю. Выбирая диапазон суммирования от $e_p^f[1]$ до $e_p^f[N]$, получаем предвзвешенный случай, поскольку при этом полагается, что значения данных, предшествующие отсчету $x[1]$, равны нулю. Диапазон суммирования от $e_p^f[p+1]$ до $e_p^f[N]$ соответствует невзвешенному случаю, поскольку используются только имеющиеся отсчеты данных. Взвешенный случай получил название *автокорреляционного* метода линейного предсказания. Случай отсутствия взвешивания называется *ковариационным* методом линейного предсказания.

Соотношение между ошибками линейного предсказания вперед и коэффициентами линейного предсказания для ковариационного (т.е. без взвешивания) метода, можно в матричной форме записать в следующем виде:

$$\begin{bmatrix} e_p^f[p+1] \\ \vdots \\ e_p^f[N-p] \\ \vdots \\ e_p^f[N] \end{bmatrix} = \begin{bmatrix} x[p+1] & \cdot & \cdot & \cdot & x[1] \\ \cdot & \cdot & & & \cdot \\ \cdot & & & \cdot & \cdot \\ x[N-p] & \cdot & \cdot & \cdot & x[p+1] \\ \cdot & \cdot & & & \cdot \\ \cdot & & & \cdot & \cdot \\ x[N] & \cdot & \cdot & \cdot & x[N-p] \end{bmatrix} * \begin{bmatrix} 1 \\ \cdot \\ \cdot \\ a^f[1] \\ \cdot \\ \cdot \\ a^f[p] \end{bmatrix}. \quad (6.42)$$

Нормальные уравнения, минимизирующие средний квадрат ошибки:

$$\rho_p^f = \sum_{n=p+1}^N |e_p^f[n]|^2 \quad (6.43)$$

порядка p , имеют вид:

$$\mathbf{R}_p \begin{pmatrix} 1 \\ \mathbf{a}_p^f \end{pmatrix} = \begin{pmatrix} \rho_p^f \\ \mathbf{0}_p \end{pmatrix}. \quad (6.44)$$

Элементы эрмитовой $(p+1) \times (p+1)$ матрицы $\mathbf{R}_p$ имеют вид корреляционных форм

$$r_p[i, j] = \sum_{n=p+1}^N x^*[n-i]x[n-j], \quad 0 \leq i, p \leq j. \quad (6.45)$$

Элементы матрицы $\mathbf{R}_p$ в корреляционном методе не могут быть записаны как функции разности $(i-j)$, а это означает, что $\mathbf{R}_p$ не является теплицевой матрицей. Однако тот факт, что матрица является произведением теплицевых матриц, все же обеспечивает возможность построения быстрого алгоритма, аналогичного алгоритму Левинсона [9]. Необходимым, но недостаточным условием того, чтобы матрица была невырожденной, является условие $N-p \geq p$ или $p \leq N/2$. Отсюда следует, что выбранный порядок модели не должен превышать половины длины записи данных. Аналогичное рассмотрение можно провести применительно и к оценке линейного предсказания назад. В [9] приведена программа COVAR, реализующая ковариационный метод. Быстрый алгоритм для ковариационного метода одновременно решает нормальные уравнения относительно коэффициентов линейного предсказания вперед и назад при всех промежуточных значениях порядка модели, поэтому оба набора коэффициентов получаются здесь без дополнительных вычислительных затрат.

Коэффициенты линейного предсказания вперед и назад, определяемые с помощью ковариационного метода, вообще говоря, не гарантируют получение устойчивого фильтра. Однако, это не приводит к каким-либо затруднениям, если их значения используются только для целей спектрального оценивания. В действительности спектральные оценки, получаемые по оценкам AP -коэффициентов с помощью ковариационного метода обычно имеют меньшие искажения, чем спектральные оценки, получаемые с помощью методов, гарантирующих устойчивость фильтра.

6.4.4 Модифицированный ковариационный метод

Для стационарного случайного процесса авторегрессионные коэффициенты линейного предсказания вперед и назад представляют собой комплексно-сопряженные величины, поэтому ошибку линейного предсказания назад можно записать в следующем виде:

$$e_p^b = x[n-p] + \sum_{k=1}^p a_p^{f*}[k]x[n-p+k]. \quad (6.46)$$

Поскольку оба направления предсказания обеспечивают получение одинаковой статистической информации, представляется целесообразным объединить статистики ошибок предсказания вперед и назад с тем, чтобы получить большее число точек, в которых определяются ошибки, и улучшить оценку AP - параметров.

Минимизируя среднее значение квадратов ошибок предсказания вперед и назад:

$$\rho_p^{fb} = \frac{1}{2} \left[\sum_{n=p+1}^N |e_p^f[n]|^2 + \sum_{n=p+1}^N |e_p^b[n]|^2 \right]$$

по коэффициентам линейного предсказания, получаем систему нормальных уравнений :

$$\mathbf{R}_p \begin{pmatrix} 1 \\ \mathbf{a}_p^{fb} \end{pmatrix} = \begin{pmatrix} 2\rho_p^{fb} \\ \mathbf{0}_p \end{pmatrix}, \quad (6.47)$$

где элементы матрицы $\mathbf{R}_p$ имеют вид:

$$r_p[i, j] = \sum_{n=p+1}^N (x^*[n-i]x[n-j] + x[n-p+i]x^*[n-p+j]), \quad (6.48)$$

где $0 \leq i, j \leq p$. Процедура, основанная на совместном использовании ошибок линейного предсказания вперед и назад по методу наименьших квадратов, получила название *модифицированного ковариационного метода*.

Модифицированный ковариационный метод и гармонический метод Берга основаны на минимизации средних квадратов ошибок линейного предсказания вперед и назад. В первом из них минимизация выполняется по всем коэффициентам предсказания, во втором выполняется условная (т.е. с наложенным ограничением) минимизация только по одному коэффициенту предсказания $a_p[p]$ (т.е. по коэффициенту отражения K_p). При использовании метода Берга возникает ряд проблем, включая расщепление спектральных линий и смещение частотных оценок, которые устраняются при использовании модифицированного ковариационного метода.

Необходимым условием невырожденности матрицы $\mathbf{R}_p$ является условие $2(N-p) > p$ или $p \leq \frac{2N}{3}$, т.е. порядок модели не должен превышать две трети длины записи данных. В [9] приведена программа MODCOVAR, реализующая модифицированный ковариационный метод.

6.4.5. Выбор порядка модели

Поскольку наилучшее значение порядка модели заранее, как правило, не известно, на практике приходится испытывать несколько порядков модели. При слишком низком порядке модели получаются сильно сглаженные спектральные оценки, при излишне высоком - увеличивается разрешение, но в спектре появляются ложные пики. Интуитивно ясно, что следует увеличивать порядок AP - модели до тех пор, пока вычисляемая ошибка предсказания не достигнет минимума. Однако во всех процедурах оценивания по методу наименьших квадратов мощности ошибок предсказания монотонно уменьшаются с увеличением порядка модели p . Так, например, в алгоритме Берга и в уравнениях Юла-Уолкера используется соотношение

$$\rho_p = \rho_{p-1} (1 - |a_p[p]|^2).$$

До тех пор, пока величина $a_p[p]$ отлична от нуля (она должна быть равной или меньше 1), мощность ошибки предсказания уменьшается. Следовательно, сама по себе мощность ошибки предсказания не может служить достаточным критерием окончания процедуры изменения порядка модели.

Для выбора порядка AP -модели предложено несколько целевых критериев. Акаике предложил два критерия. Первым из них является величина окончательной ошибки предсказания ($ООП$). Согласно этому критерию, порядок AP -процесса выбирается таким образом, чтобы средняя дисперсия ошибки на каждом шаге предсказания была минимальна. Акаике рассматривал ошибку как сумму мощностей в непредсказуемой (или не обновляемой) части процесса и как некоторую величину, характеризующую неточность оценивания AP -параметров. Окончательная ошибка предсказания для AP -процесса определяется

$$ООП_p = \hat{\rho}_p \left(\frac{N + p + 1}{N - p + 1} \right),$$

где N - число отсчетов данных, p - порядок AP -процесса, $\hat{\rho}_p$ - оценочное значение дисперсии шума (дисперсии ошибки предсказания). Член в круглых скобках увеличивает окончательную ошибку предсказания по мере того, как p приближается к N , характеризуя тем самым увеличение неопределенности оценки $\hat{\rho}_p$ для дисперсии ошибки предсказания. Выбирается порядок p , при котором величина окончательной ошибки предсказания минимальна. Критерий на основе окончательной ошибки предсказания исследовался в различных приложениях и для идеальных AP -процессов он обеспечивает хорошие результаты. Однако при обработке реальных сигналов этот критерий приводит к выбору слишком малого порядка модели.

Второй критерий Акаике основан на методе максимального правдоподобия и получил название информационного критерия Акаике ($ИКА$). Согласно этому критерию, порядок модели выбирается посредством минимизации некоторой теоретико-информационной функции. Если исследуемый процесс имеет гауссовы статистики, то $ИКА$ определяется выражением

$$ИКА[p] = N \ln(\hat{\rho}_p) + 2p.$$

Второй член $2p$ характеризует плату за использование дополнительных AP -коэффициентов, но это не приводит к значительному уменьшению дисперсии ошибки предсказания. И здесь выбирается порядок модели, при котором $ИКА$ минимален.

Третий метод выбора критерия предложен Парзенем и получил название авторегрессионной передаточной функции критерия ($АПФК$). Порядок модели p выбирается в этом случае равным порядку, при котором оценка разности среднеквадратичных ошибок между истинным фильтром предсказания ошибки (его длина может быть бесконечной) и оцениваемым фильтром минимальна. Парзен показал, что эту разность можно вычислить, даже если истинный предсказывающий ошибку фильтр точно не известен:

$$АПФК[p] = \left(\frac{1}{N} \sum_{j=1}^p \frac{1}{\rho_j} \right) - \frac{1}{\rho_p},$$

где $\overline{\rho_j} = (N/(N-j))\rho_j$. И здесь p выбирается так, чтобы минимизировать $АПФК$.

Результаты оценивания спектра при использовании трех указанных критериев мало отличаются друг от друга, особенно в случае реальных данных, а не моделируемых процессов. Исследования показали, что в случае коротких записей данных ни один из этих критериев не обеспечивает удовлетворительных результатов. Для гармонических процессов в присутствии шума использование первого и второго критериев приводит к заниженной оценке порядка модели, если отношение *сигнал / шум* велико. При анализе коротких сегментов данных предложено выбирать порядок модели равным от $N/3$ до $N/2$, что во многих случаях приводит к удовлетворительным результатам. Заметим, что выбор порядка модели для данных, полученных из реальных процессов, пока еще носит субъективный, а не строго научный характер.

6.4.5. Характеристики оценок

Свойства спектров, получаемых по оценкам авторегрессионных параметров, рассмотренным выше, подробно исследованы в многочисленных опубликованных работах. Так как метод Берга является одним из первых и наиболее широко используемых алгоритмов, то проверке и сравнению чаще всего подвергались результаты, получаемые именно с его помощью. Появление каждой новой процедуры спектрального оценивания было, как правило, вызвано необходимостью устранить те или иные аномалии в спектральных оценках, получаемых с помощью гармонического алгоритма Берга. К такого рода аномалиям, которые будут рассмотрены ниже, относятся ложные спектральные пики, смещения частотных оценок и расщепление спектральных линий.

Если выбран большой порядок AP - модели относительно имеющегося числа отсчетов данных, то в авторегрессионных спектральных оценках могут появляться ложные пики. Из-за ошибок оценивания матрица нормальных уравнений для большинства AP - методов будет иметь полный ранг, равный большим значениям порядка моделей, так что решения для AP - параметров получаются даже тогда, когда истинная модель имеет значительно меньший порядок. Дополнительные полюсы, порождаемые лишними AP - параметрами, приводят к появлению ложных спектральных пиков. Для уменьшения числа ложных пиков следует использовать методы определения порядка модели. Следует заметить, что уменьшение порядка модели с

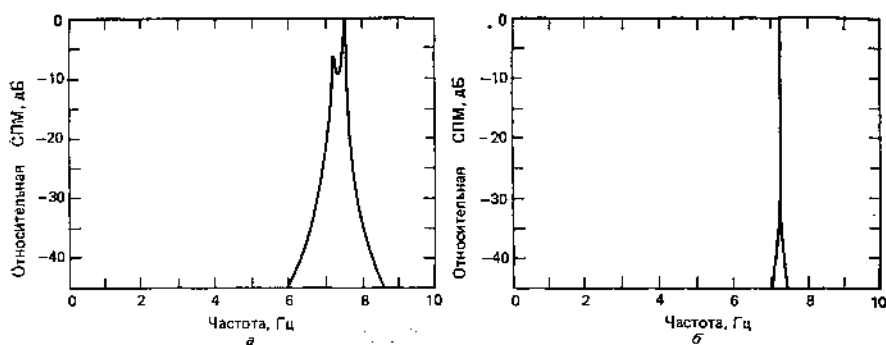


Рис. 6.3. Две спектральные AP - оценки, полученные по 101 отсчету процесса, состоящего из синусоиды с частотой 7,25 Гц и аддитивного белого шума (отношение сигнал/шум = 50, частота дискретизации—100 Гц, начальная фаза равна 45°): а — оценка методом Берга с расщепленной спектральной линией; б — оценка модифицированным ковариационным методом, расщепление спектральной линии отсутствует.

целью борьбы с ложными пиками снижает также и разрешение спектральной оценки. Из описанных выше AP -методов спектрального оценивания более высокое разрешение при заданном порядке модели обеспечивают метод Берга, ковариационный метод и модифицированный ковариационный метод. Обусловлено это главным образом отсутствием в них эффектов, связанных с применением окна. Именно по этой причине наихудшее разрешение из всех описанных здесь методов имеет автокорреляционный метод.

При использовании метода Берга и автокорреляционного метода было замечено, что при некоторых условиях в спектральной оценке могут появляться два близко расположенных спектральных пика там, где должен присутствовать только один спектральный пик. Это явление, названное *расщеплением спектральной линии*, иллюстрирует рис. 6.3, на котором показаны спектральные оценки, полученные с помощью метода Берга и модифицированного ковариационного метода. Оба спектра на этом рисунке построены по 25 значениям оценок, полученным по 101 отсчету процесса, состоящего из синусоиды единичной амплитуды с начальной фазой 45° и частотой 7,25 Гц и аддитивного шума с дисперсией 0,01 (отношение сигнал/шум равно 50). Частота отсчетов равна 100 Гц. Спектральная оценка по методу Берга, показанная на рис. 6.3, а,

имеет расщепленный пик на частоте примерно 7,25 Гц, что создает ложное представление о наличии частот двух синусоид. Спектр, полученный по тем же данным с помощью модифицированного ковариационного метода, имеет только один пик на правильной частоте. При использовании метода Берга расщепление спектральных линий наиболее вероятно в тех случаях, когда 1) велико отношение сигнал/шум, 2) начальные фазы синусоидальных компонент нечетно кратны углу 45° , 3) протяженность последовательности данных во времени такова, что синусоидальные компоненты имеют нечетное число четвертей периодов, и 4) процентное соотношение между числом оцениваемых *AP*-параметров и числом используемых для этой цели отсчетов данных относительно велико. Расщепление спектральных линий отмечалось при использовании как действительных, так и комплексных данных, причем спектры, характеризующиеся расщеплением линий, как правило, содержат много ложных спектральных пиков. С увеличением длины записи данных вероятность расщепления спектральных линий быстро уменьшается. Теоретический анализ причин расщепления спектральных линий в случае автокорреляционного метода Юла — Уолкера, показал, что используя несмещенную оценку автокорреляции, можно ослабить или даже полностью устранить расщепление спектральных линий, свойственное этому методу. Аналогичный анализ для метода Берга показал, что использование в данном случае взвешенных квадратов ошибок позволяет снизить вероятность расщепления спектральных линий, появление которого, по всей видимости, обусловлено смещением между положительными и отрицательными спектральными компонентами действительных синусоид. Однако наилучшее средство от этого - модифицированный ковариационный метод, при использовании которого еще ни разу не отмечалось расщепления спектральных линий, особенно в тех случаях, когда применение метода Берга дает оценки с расщепленными спектральными линиями.

В работах ряда исследователей [9] отмечается, что в случае процесса, состоящего из смеси одной или двух синусоид и аддитивного белого шума, спектральные пики авторегрессионной спектральной оценки по методу Берга оказываются сдвинутыми, причем величина их сдвига зависит от начальной фазы этих синусоид. В одном из экспериментов Ульриха и Клейтона с помощью метода Берга и модифицированного ковариационного метода определялись спектральные оценки по ансамблям из 15 отсчетов процесса, состоящего из действительной синусоиды единичной амплитуды с частотой 1 Гц и аддитивного белого шума при отношении сигнал/шум, равном 10 (интервал отсчетов был равен 0,05 с).

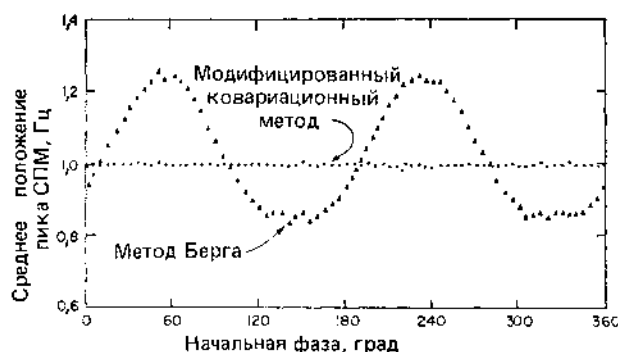


Рис. 6.4. Частотное смещение в случае двух *AP*-методов спектрального оценивания.

На рис. 6.4 показан график зависимости среднего положения пика этих спектральных оценок от начальной фазы синусоиды. Каждая точка на этом графике характеризует среднюю частоту спектрального пика, вычисленную по ансамблю из 50 независимых реализаций данных. Из приведенного рисунка с очевидностью следует, что в случае модифицированного ковариационного метода эта средняя частота очень слабо зависит от начальной фазы синусоиды и является точной оценкой ее частоты. В то же время метод Берга характеризуется достаточно сильным смещением частотной

оценки, величина которого с изменением начальной фазы меняется примерно по синусоидальному закону. Теоретическое обоснование такого характера изменения частотного смещения дано в работах Суинглера [9], где, в частности, показано, что это смещение может достигать 16% величины элемента (или ячейки) разрешения $1/NT$ Герц. Он же показал, что использование взвешенных квадратов ошибок [таких, например, как в (6.36)] ослабляет фазовую зависимость частотных оценок по методу Берга. Эффекты, связанные с этим смещением, уменьшаются также и при использовании аналитического (т. е. комплексного) сигнала.

Наттолл [9], используя усреднение по ансамблю из большого числа наборов данных, проанализировал дисперсию оценок СПМ, получаемых с помощью различных *АР*-методов. Полученные им результаты показывают, что в случае несинусоидальных процессов из всех этих методов лишь метод Берга и модифицированный ковариационный метод дают, как правило, оценки СПМ и частоты с минимальной дисперсией.

Хотя выше основное внимание было уделено характеристикам авторегрессионных спектральных оценок для коротких последовательностей отсчетов данных, следует также кратко упомянуть и об их асимптотических статистических свойствах. Так, Сакаи [9] экспериментально показал, что в случае процесса, состоящего из синусоид и аддитивного шума, дисперсия частоты авторегрессионной спектральной оценки оказывается обратно пропорциональной длине записи данных и квадрату отношения сигнал/шум. Килер [9] представил экспериментальное доказательство того, что в случае несинусоидальных процессов дисперсия обратно пропорциональна длине записи данных и отношению сигнал/шум (а не квадрату этого отношения, как в случае синусоидальных процессов). Акаике, Кроумер и Берк [9] показали, что с ростом числа отсчетов данных асимптотические характеристики *АР*-оценки СПМ, основанной на автокорреляционном методе, асимптотически приближаются к характеристикам гауссовского распределения вероятностей; иными словами, среднее значение *АР*-оценки СПМ оказывается в пределе равным истинному среднему значению этой оценки, а ее дисперсия стремится к значению, пропорциональному величине $(4p/N)P_{AP}^2(f)$.

6.5. Спектральное оценивание на основе моделей авторегрессии - скользящего среднего

АРСС-модель имеет больше степеней свободы, чем *АР*-модель, поэтому с ее помощью при меньшем числе параметров можно точнее аппроксимировать *СПМ* реальных процессов. Однако в отличие от большого числа разработанных линейных алгоритмов вычисления *АР*-оценок *СПМ*, для получения *АРСС*-оценок *СПМ* в настоящее время используется лишь несколько алгоритмов, а поиск новых алгоритмов - задача весьма актуальная. Сложность оценки параметров *АРСС*-модели - в нелинейных соотношениях (6.11) связи их с *АКП*. Решение нелинейных уравнений, подобных (6.11), с использованием итерационных алгоритмов требует больших вычислительных затрат и не гарантирует их сходимости. Для уменьшения вычислительных затрат предложены варианты раздельной оценки *АР*- и *СС*-параметров, использующие линейные процедуры и позволяющие получать решения, близкие к оптимальным.

Раздельный способ оценивания параметров *АРСС*-модели заложен в соотношениях (6.11), из которых следуют линейные уравнения (6.12) для определения *АР*-параметров $a[k]$, $k = \overline{1, p}$. После определения *АР*-параметров на основе модифицированных уравнений Юла-Уолкера (6.12) производится оценка *СС*-параметров в предположении, что *АР*-параметры известны. Рассмотрим подробнее основные этапы раздельной оценки *АР*- и *СС*-параметров.

При известной *АКП* *АРСС* (p, q)-процесса соотношения

$$r_{xx}[m] = -\sum_{k=1}^p a[k] r_{xx}[m-k], \quad (6.49)$$

где $q+1 \leq m \leq q+p$, образуют систему p уравнений, решая которую можно определить AP -параметры $a[k]$, $k = \overline{1, p}$.

В практических задачах спектрального оценивания $AKП$ неизвестна, а обрабатывается лишь некоторая совокупность отсчетов данных, по которой находятся оценки $AKП$. Таким образом, один из возможных вариантов нахождения оценок AP -параметров - решение системы p уравнений (6.49), в которую вместо истинных значений $AKП$ $r_{xx}[m]$ подставляются ее оценки $\hat{r}_{xx}[m]$, вычисляемые по наблюдаемым данным $x[n]$, $n = \overline{1, N}$. И хотя с вычислительной точки зрения этот подход представляется вполне оправданным, недостаток его в том, что он использует для определения параметров $a[k]$, $k = \overline{1, p}$ лишь часть вычисленных значений $AKП$ $\hat{r}_{xx}[m]$, $q-p \leq m \leq q+p$. Если предположить, что оценки $AKП$ $\hat{r}_{xx}[m]$ вычислены надежно для корреляционных сдвигов от 0 до M , то можно использовать больше, чем минимальное число p уравнений Юла-Уолкера (переопределенную систему), а именно $(M-q)$ уравнений ($M-q > p$) в форме

$$\hat{r}_{xx}[m] = -\sum_{k=1}^p a[k] \hat{r}_{xx}[m-k] + \varepsilon[m], \quad (6.50)$$

где $q+1 \leq m \leq M$, а $\varepsilon[m]$ - ошибка оценивания. Необходимо использовать несмещенные автокорреляционные оценки, с тем, чтобы гарантировать, что смещение ошибки будет нулевым. Затем сумма квадратов ошибки

$$\rho = \sum_{m=q+1}^M |\varepsilon[m]|^2$$

минимизируется относительно авторегрессионных параметров. Эта процедура названа *модифицированным методом наименьших квадратов Юла-Уолкера*.

Получаемые в результате нормальные уравнения оказываются идентичными уравнениям, которые получаются при использовании ковариационного метода линейного предсказания, описанного выше, т.е. они имеют следующий вид:

$$[\mathbf{T}_p^H \mathbf{T}_p] \begin{bmatrix} 1 \\ a[1] \\ \vdots \\ a[p] \end{bmatrix} = \begin{bmatrix} \rho \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (6.51)$$

где

$$\mathbf{T}_p = \begin{bmatrix} \hat{r}_{xx}[q+1] & \cdot & \cdot & \cdot & \hat{r}_{xx}[q-p+1] \\ \cdot & \cdot & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ \hat{r}_{xx}[M-p] & & & & \hat{r}_{xx}[q+1] \\ \cdot & & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ \hat{r}_{xx}[M] & \cdot & \cdot & \cdot & \hat{r}_{xx}[M-p] \end{bmatrix}$$

- прямоугольная теплицева матрица, состоящая из автокорреляционных оценок, $\mathbf{T}_p^H$ - эрмитово транспонированная матрица $\mathbf{T}_p$, получаемая в результате комплексного сопряжения всех элементов матрицы $\mathbf{T}_p$ и последующей их транспозиции. Для решения матричного уравнения (6.51) можно использовать подпрограммы (например, COVAR), которые реализуют один из быстрых алгоритмов, предназначенных для AP -оценивания $СПМ$, с тем отличием, что вместо последовательности данных $x[n], n = \overline{1, N}$ следует использовать последовательность оценок автокорреляции $\hat{r}_{xx}[q-p+1], \dots, \hat{r}_{xx}[M]$. Выбор порядка AP -составляющей $APCC$ -процесса может быть основан на методах, применяемых к выбору порядка чистого AP -процесса, которые описаны в п. 6.4.5.

Для того, чтобы завершить решение задачи идентификации модели $APCC$, необходимо определить значения CC -параметров. Это можно сделать различными процедурами. Одна из них предполагает обработку исходного временного ряда KIX -фильтром p -го порядка с системной функцией

$$\hat{A}(z) = 1 + \sum_{k=1}^p \hat{a}[k]z^{-k}, \quad (6.52)$$

где $\hat{a}[k]$ - оценки AP -параметров, определенные из решения системы уравнений (6.51). Каскадное (последовательное) соединение фильтра с системной функцией $H(z) = B(z)/A(z)$, формирующей $APCC$ -процесс, и фильтра с системной функцией $\hat{A}(z)$ эквивалентно фильтру, системная функция которого равна

$$\frac{B(z)}{A(z)} \hat{A}(z) \approx B(z). \quad (6.53)$$

Таким образом, фильтрация превращает исходный временной ряд в так называемый остаточный ряд $z[n]$, который является процессом скользящего среднего порядка q со спектральной плотностью мощности $|B(f)|^2$ (рис 6.5.).

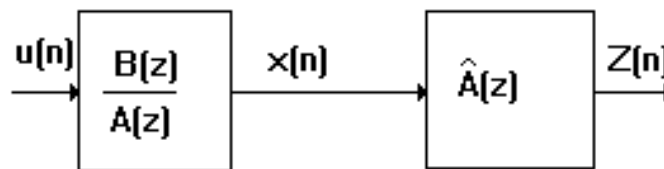


Рис 6.5. Формирование остаточного ряда.

Фильтрованная последовательность длиной $N-p$ будет определяться соотношением:

$$z[n] = x[n] + \sum_{k=1}^p \hat{a}[k]x[n-k], \quad (6.54)$$

где $p+1 \leq n \leq N$.

Последний этап состоит в оценке CC -параметров по фильтрованной последовательности $z[n]$, аппроксимирующей процесс скользящего среднего.

Оценка параметров CC -процесса также может быть реализована различными вариантами. Рассмотрим некоторые из них. Пусть временной ряд $z[n]$, соответствующий CC -модели порядка q , описывается уравнением

$$z[n] = \sum_{k=1}^q b[k]u[n-k] + u[n]. \quad (6.55)$$

Тогда коэффициенты $b[k]$ и $AKП$ оказываются связанными нелинейными

соотношениями

$$r_{zz}[m] = \begin{cases} \rho_w \sum_{k=0}^q b[k] b^*[k-m], & 0 < m < q; \\ r_{zz}^*[-m], & -q < m < 0; \\ 0, & |m| > q. \end{cases} \quad (6.56)$$

Таким образом, АКП СС-процесса имеет конечную длину $(2q+1)$, определяемую порядком q СС-процесса.

Как можно определить параметры $b[k]$ фильтра, порождающего заданную АКП $r_{zz}[n]$, $-q \leq n \leq q$?

1. Проанализируем z - преобразование заданной корреляционной последовательности, т.е. спектральную характеристику

$$\begin{aligned} S_z[z] &= \sum_{n=-q}^q r_{zz}[n] z^{-n} = \rho_w \sum_{n=-q}^q \sum_{k=0}^q b[k] b^*[k-n] z^{-n} = \\ &= \rho_w \sum_{k=0}^q b[k] z^{-k} \sum_{m=0}^q b^*[m] z^m \end{aligned} \quad (6.57)$$

Но, так как коэффициенты конечного степенного ряда $S_z(z)$ являются комплексно-сопряженными ($r_{zz}[-n] = r_{zz}^*[n]$), его нули должны образовывать взаимно-обратные пары. Следовательно, всегда можно осуществить факторизацию этой спектральной характеристики и представить ее в виде

$$S_z(z) = \alpha^2 \prod_{k=1}^q (1 - z_k z^{-1})(1 - z_k^* z), \quad (6.58)$$

где α^2 - вещественный параметр. Сравнивая (6.57) и (6.58), получим соотношения для определения параметров $b[k]$:

$$\sqrt{\rho_w} \sum_{k=0}^q b[k] z^{-k} = \alpha \prod_{k=1}^q (1 - z_k z^{-1}). \quad (6.59)$$

Параметры $b[k]$ находятся приравнованием коэффициентов при одинаковых степенях в обеих частях уравнения (6.59). Машинная программа факторизации комплексного полинома приведена в [9].

2. Другой вариант оценивания $b[k]$ основан на аппроксимации СС- процесса AP - моделью высокого порядка и использует только линейные операции.

Пусть

$$B(z) = 1 + \sum_{k=1}^q b[k] z^{-k}$$

системная функция фильтра, формирующего СС(q)-процесс из белого шума, и пусть $1/A_\infty(z)$, где

$$A_\infty(z) = 1 + \sum_{k=1}^{\infty} a(k) z^{-k}$$

системная функция фильтра, формирующего из белого шума AP_∞ - процесс бесконечного порядка, эквивалентный СС(q) - процессу. Тогда имеют место соотношения

$$B(z) = \frac{1}{A_\infty(z)}, B(z) A_\infty(z) = 1. \quad (6.60)$$

Обратное z - преобразование от обеих частей соотношения (6.60) дает уравнение, левая часть которого есть свертка СС - параметров с AP - параметрами, а правая часть - единичный импульс:

$$a[m] + \sum_{n=1}^q b[n]a[m-n] = \delta[m] = \begin{cases} 1, & m=0, \\ 0, & m \neq 0, \end{cases} \quad (6.61)$$

где по определению $a[0] = 1, a[k] = 0, k < 0$.

Таким образом, CC -параметры $b[k]$ можно определить по параметрам $a[k]$ некоторой эквивалентной AP -модели бесконечного порядка путем решения подсистемы q линейных уравнений, полученных из (2.61). На практике вычисляют по конечной записи данных оценки параметров $AP(M)$ -модели высокого порядка, такого, что $M \gg q$. Используя $AP(M)$ -параметры $\hat{a}_M[1], \dots, \hat{a}_M[M]$, можно записать следующую систему уравнений

$$\hat{a}_M[m] + \sum_{n=1}^q b[n]\hat{a}_M[m-n] = e_{CC}[m]. \quad (6.62)$$

В идеальном случае, в соответствии с (6.61), ошибка $e_{CC}[m]$ должна быть равна нулю для всех m , кроме $m=0$. Однако для конечной записи данных и ограниченного порядка $AP(M)$ -модели эта ошибка не будет равна нулю. Поэтому оценки CC -параметров должны определяться минимизацией суммы квадратов ошибок

$$\hat{\rho}_q = \sum_m |e_{CC}[m]|^2 / M. \quad (6.63)$$

Уравнение (6.62) по своей форме идентично выражению для ошибки линейного предсказания вперед с тем отличием, что вместо отсчетов данных в нем фигурируют оценки параметров $\hat{a}_m[k]$. В соотношении (6.63) можно использовать два интервала суммирования: интервал $0 \leq m \leq M+q$, который соответствует автокорреляционному методу линейного предсказания, и интервал $-q \leq m \leq M$, который соответствует ковариационному методу линейного предсказания. Решение нормальных уравнений в случае автокорреляционного метода можно выполнить с помощью подпрограммы **YULEWALKER**, а в случае ковариационного метода - с помощью подпрограммы **COVAR**, приведенных в [9].

3. Наконец, обратим внимание на следующее обстоятельство. Хотя CC -параметры и необходимы для оценивания параметров $APCC$ -модели временного ряда, для

получения оценки $СПМ$ нужна лишь оценка величины $T\rho_w \left| \hat{B}(f) \right|^2$, которую можно получить, минуя оценку CC -параметров. Действительно, находя коррелограммную оценку $СПМ$ фильтрованной последовательности $z[n]$, получим

$$T\rho_w \left| \hat{B}(f) \right|^2 = T \sum_{m=-q}^q \hat{r}_{zz}[m] \exp\{-j2\pi f m T\}, \quad (6.64)$$

где $\hat{r}_{zz}[m]$ - оценка автокорреляции фильтрованной последовательности $z[n]$ (6.54) может быть вычислена например, по соотношению

$$\hat{r}_{zz}[n] = \begin{cases} \sum_{k=0}^p \sum_{m=0}^p \hat{a}[k] \hat{a}[m] r_{xx}[n+m-k], & -q \leq n \leq q, \\ 0, & |n| > q. \end{cases} \quad (6.65)$$

на основе оценок $АКП$ исходного временного ряда.

6.6 Экспериментальное исследование алгоритмов спектрального оценивания на основе АР-моделей и моделей АРСС

6.6.1. Измерение характеристик разрешения алгоритмов

Для определения частотного разрешения алгоритмов ЦСА используют тест - сигнал, состоящий из пары синусоид равной амплитуды с близкими частотами в сумме с шумом наблюдения. Частотная расстройка, при которой спектральная оценка позволяет зафиксировать два пика с провалом между ними не менее 3 дБ, зависит как от способа получения оценки, так и от совокупности параметров сигнала: длины реализации, отношения сигнал / шум, порядка модели. Для классических методов ЦСА процедурой закрытия окна добиваются нужного компромисса между частотным разрешением и устойчивостью (дисперсией) получаемой спектральной оценки при фиксированной длине реализации. При этом, увеличивая длину временного окна, достигают нужного частотного разрешения за счет возрастания дисперсии флюктуаций спектральной оценки, которое вызвано соответствующим уменьшением числа сегментов в процедуре усреднения оценки СПМ. Предельное частотное разрешение, обеспечиваемое классическими методами, не превосходит величины, обратной длине записи данных, и не зависит от характеристик анализируемых данных. Авторегрессионные методы оценки СПМ требуют правильного выбора важного параметра - порядка АР-модели. Увеличение порядка модели приводит к возрастанию частотного разрешения, однако при этом в спектральных оценках могут появляться ложные пики, что ассоциируется с ухудшением устойчивости оценки (большей ее дисперсией). Предельное значение порядка не должно, как правило, превышать половины длины записи данных. Для определения наиболее подходящего порядка АР-модели приходится испытывать много различных порядков, и эта процедура, называемая «закрытием» порядка модели, сходна с процедурой закрытия окна в классических методах.

Временное разрешение как характеристика анализатора спектра, а также понятие частотно - временной разрешающей способности возникают при оценке качества измерения нестационарных (изменяющихся во времени) спектров. При скользящих спектральных измерениях временное разрешение отождествляется с минимальной длиной временного окна, перемещающегося по длине реализации обрабатываемого сигнала, при котором отдельно измеряются спектральные пики двух близких по частоте синусоид.

Измерения параметров разрешения различных алгоритмов АР-оценивания СПМ показали, что наихудшее разрешение имеет алгоритм Юла-Уолкера, а лучшее - реализует модифицированный ковариационный метод, основанный на линейном предсказании вперед и назад. Метод Берга и ковариационный метод имеют сравнимые характеристики разрешения и существенно лучшие, чем у классических методов оценки СПМ. В качестве примера сравнительных измерений частотного и временного разрешений алгоритмов АР-оценивания СПМ приведем численные значения параметров для пары комплексных синусоид в шуме. Так синусоиды с относительными частотами $f_1 = 0.2$ и $f_2 = 0.25$ и равными амплитудами ($A_1 = A_2 = 1.0$) удавалось разрешить при минимально возможном числе временных отсчетов $N = 4$ и порядке модели $p = 2$ только модифицированным ковариационным методом при отношении сигнал / шум $SNR > 40$ дБ, что соответствовало частотно - временному разрешению $\Delta f N = 0.05 * 4 = 0.2$. При увеличении числа отсчетов до $N = 6$ ($p = 2$, $SNR = 40$ дБ) разрешает синусоиды и алгоритм Берга, а при $N = 7$ ($p = 2$, $SNR = 40$ дБ) - ковариационный алгоритм. Алгоритм Юла - Уолкера начинает разрешать эти синусоиды при $N = 14$ ($p = 7$, $SNR = 40$ дБ). Заметим, что периодограммный метод при односегментном анализе и прямоугольном окне разрешает синусоиды при $N = 16$. При увеличении числа отсчетов до $N = 8$ ($p = 3, 4$) методы Берга и модифицированный

ковариационный начинают разрешать синусоиды при $\text{SNR} = 25$ дБ. Аналогичные измерения для пары синусоид равной амплитуды с частотами $f_1 = 0.2$ и $f_2 = 0.21$ при относительной расстройке $\Delta f = f_2 - f_1 = 0.01$ и отношении сигнал / шум = 60 дБ показали, что модифицированный ковариационный метод разрешает их при $N = 6$ ($p = 2,3$), что соответствует частотно - временному разрешению $\Delta f N = 0.06$. Метод Юла - Уолкера разрешает эти синусоиды при $N > 70$ ($p = 35$), периодограммный метод при односегментном анализе - при $N > 80$.

В таблицу 6.1 сведены значения числа отсчетов N , порядка модели p и соответствующего отношения сигнал / шум (SNR) (дБ), при которых модифицированный ковариационный метод разрешает синусоиды с частотами 0.2 и 0.21. При увеличении отношения сигнал / шум от 10 дБ до 60 дБ частотно - временное разрешение возрастает от величины 0.48 до 0.06.

Таблица 6.1. Характеристики разрешения модифицированного ковариационного метода AP - оценивания СПМ

SNR (дБ)	0	10	20	30	40	50	60
N	56	48	32	24	16	8	6
p	28	18	16	12	6	4	2
$\Delta f N$	0.56	0.48	0.32	0.24	0.16	0.08	0.06

На рис. 6.6, 6.7 и 6.8 представлены спектрограммы, отображающие оценки СПМ суммы двух синусоид в шуме для предельных значений параметров разрешения, соответственно для периодограммного, Юла-Уолкера и модифицированного ковариационного методов.

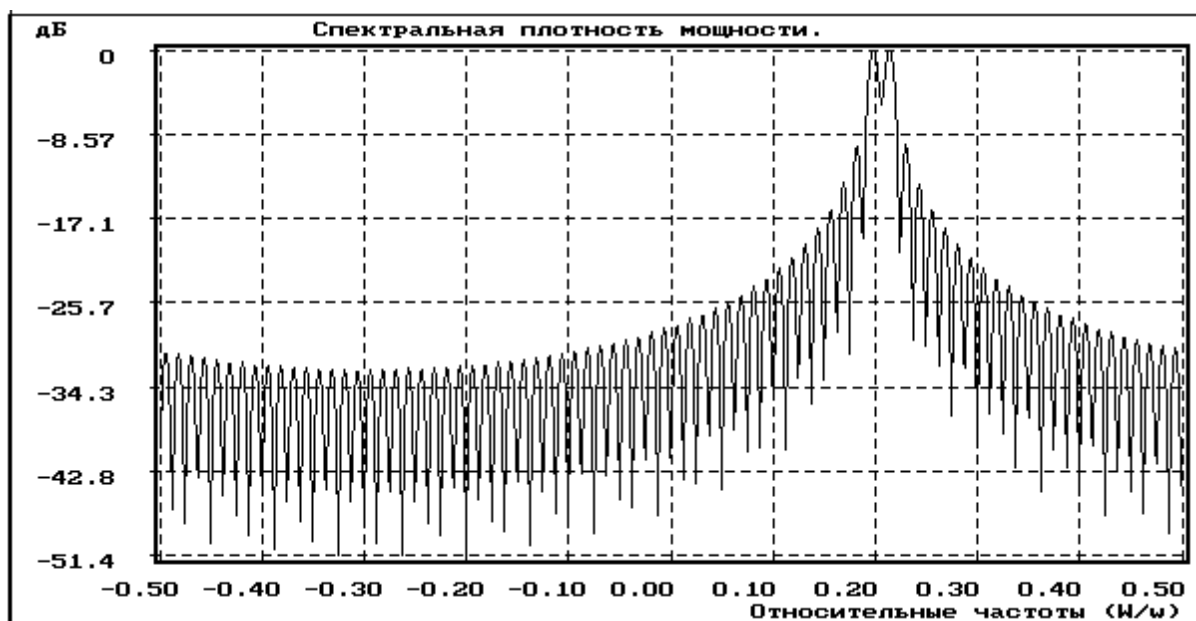


Рис.6.6. Периодограмма суммы двух синусоид в шуме (80 отсчетов, SNR =60дБ)

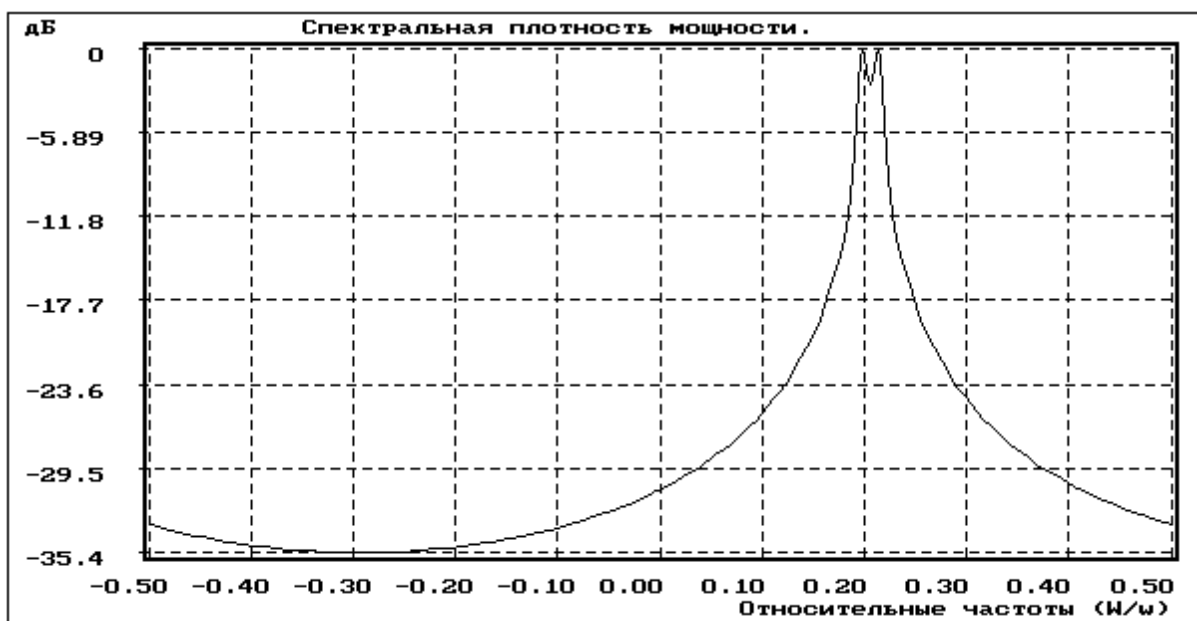


Рис. 6.7. Спектральная AP(35)-оценка суммы двух синусоид в шуме, полученная методом Юла-Уолкера (70 отсчетов, SNR =60 дБ)

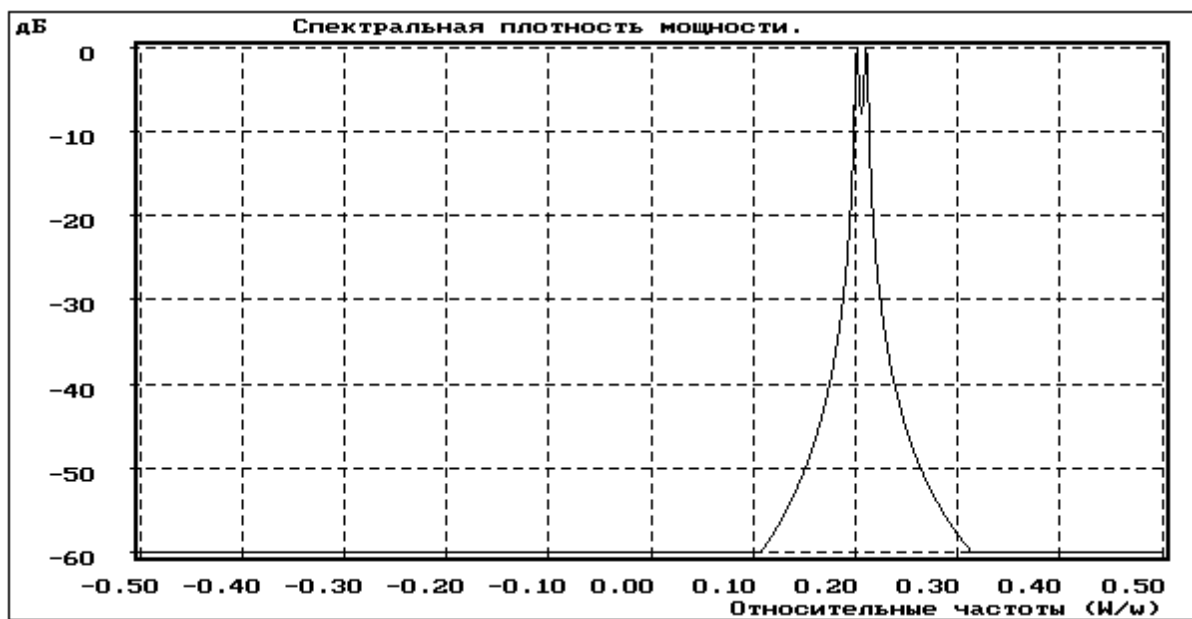


Рис. 6.8. Спектральная AP(3)-оценка суммы двух синусоид в шуме, полученная модифицированным ковариационным методом (6 отсчетов, SNR = 60 дБ)

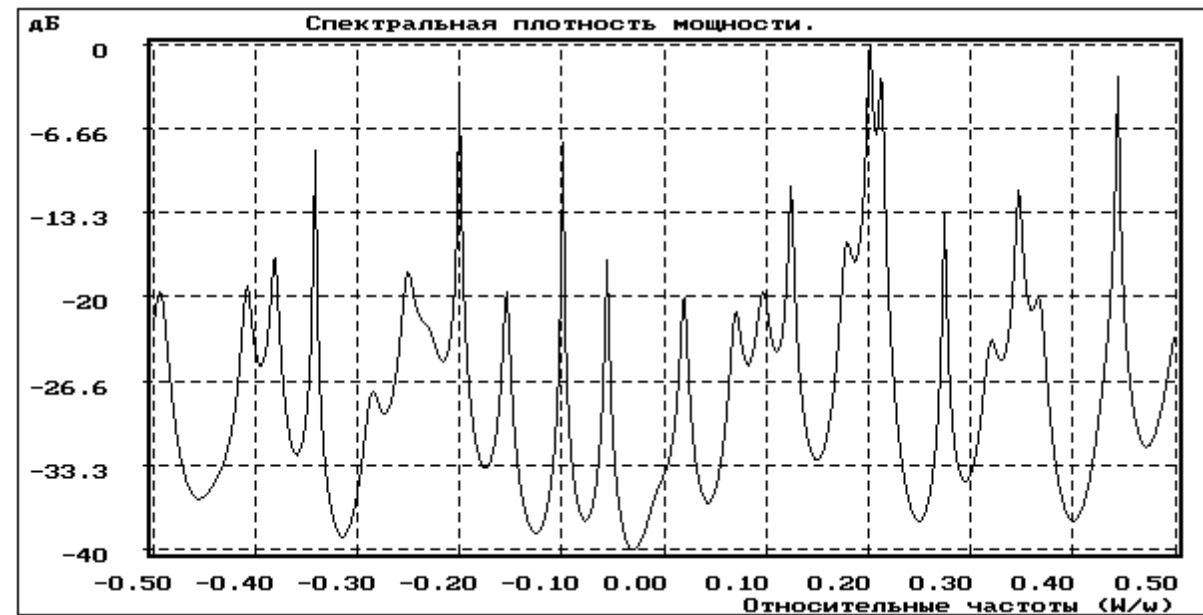


Рис. 6.9. Спектральная AP(28)-оценка суммы двух синусоид в шуме, полученная модифицированным ковариационным методом (56 отсчетов, SNR = 0 дБ)

Для малых отношений сигнал / шум высокое частотное разрешение AP-оценок СПМ может быть реализовано только при больших порядках модели и, следовательно, больших длинах последовательности данных. Так, пара синусоид с частотами 0.2 и 0.21 при отношении сигнал / шум SNR = 0 дБ ($A_{1,2} = 1$, дисперсия шума = 0.5) разрешается модифицированным ковариационным методом при $N = 56$ и $p = 28$, что отображено спектрограммой, представленной на рис. 6.9. Однако большое число ложных пиков, связанное с большим порядком модели и эквивалентное большой дисперсии спектральной оценки, затрудняют правильную ее интерпретацию.

Для сравнения на рис.6.10 приведена APCC(9, 9)-оценка той же последовательности при $N = 56$ и SNR = 0 дБ при существенно меньшем эквивалентном порядке модели, равном $p + q = 18$, и, следовательно, меньшей дисперсии оценки с разрешенными синусоидами. Для этого случая AP- параметры оценивались алгоритмом Берга, СС-параметры оценивались на основе длинной AP-модели ($p = 30$), параметры которой определялись также алгоритмом Берга, с последующей оценкой СС-параметров ковариационным методом. Из сравнения спектрограмм можно сделать вывод о том, что APCC-модель позволяет получить лучшие оценки СПМ и реализовать лучшее разрешение, чем AP-модель для сильно зашумленных синусоид. Этот весьма важный для практики вывод в пользу APCC-оценок СПМ подтвержден и многими другими вариантами полученных APCC-оценок.

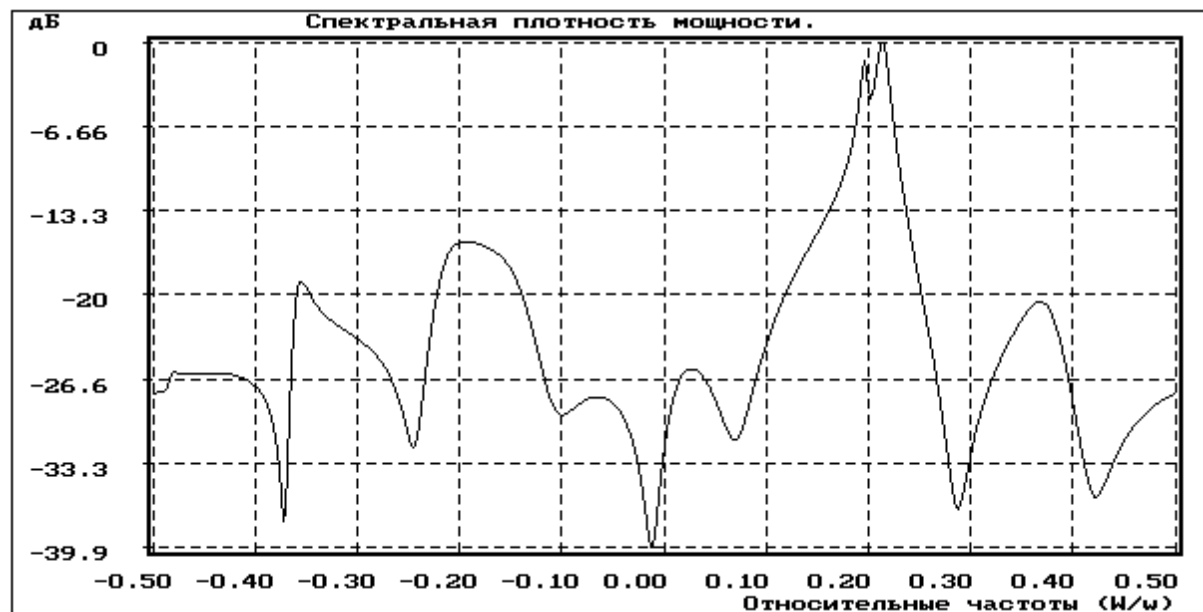


Рис. 6.10. Спектральная АРСС (9, 9)-оценка суммы двух синусоид в шуме (56 отсчетов, SNR = 0 дБ) с раздельным оцениванием АР- и СС-параметров

6.6.2. Исследование спектральных оценок моделей стационарных процессов

Примеры четырех авторегрессионных оценок СПМ 64-точечной тест-последовательности Марпла [9], полученных соответственно по алгоритмам Юла-Уолкера, Берга, ковариационного и модифицированного ковариационного, приведены на рис. 6.11, 6.12, 6.13, 6.14. Для всех методов использовался АР-процесс 13-го порядка. Все алгоритмы, кроме Юла-Уолкера, хорошо разрешают все четыре комплексные синусоиды и примерно одинаково отображают спектры коррелированных шумов. Алгоритм Юла-Уолкера имеет худшее разрешение из всех описанных здесь методов АР-оценивания СПМ, превосходящее однако разрешение классических методов. Синусоиды тест-последовательности разрешаются методом Юла-Уолкера только для максимального порядка модели, равного 32, классические методы близкие по частоте синусоиды не разрешают, ввиду малой длины последовательности.

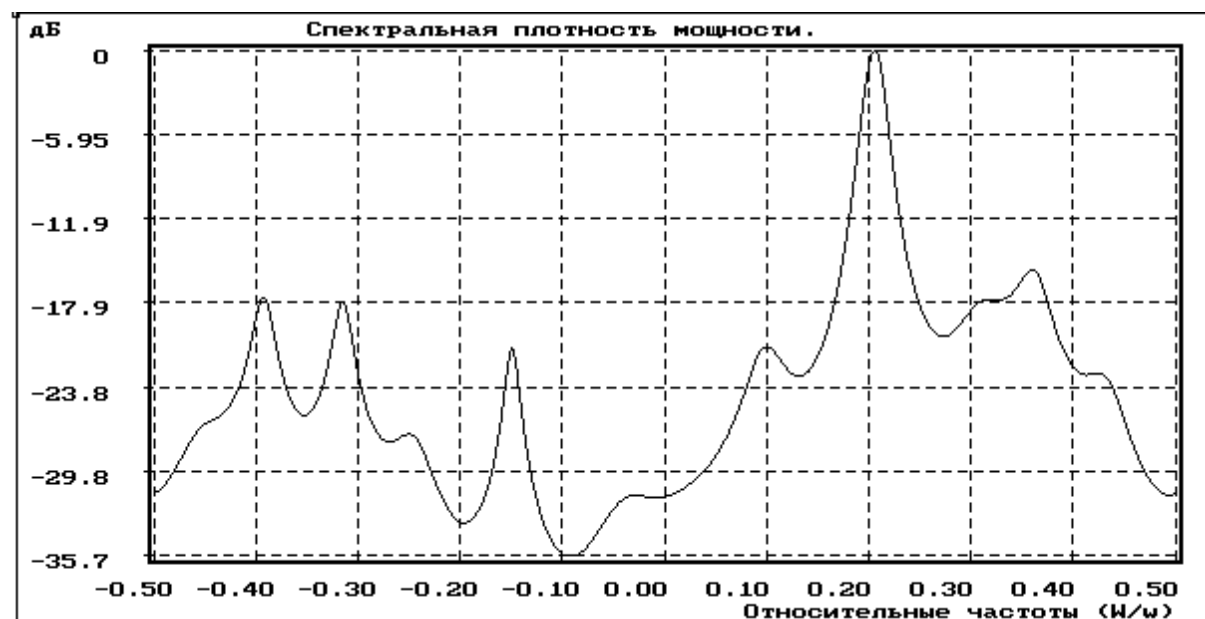


Рис. 6.11. АР(13)-оценка СПМ тест-последовательности Марпла, полученная методом Юла-Уолкера.

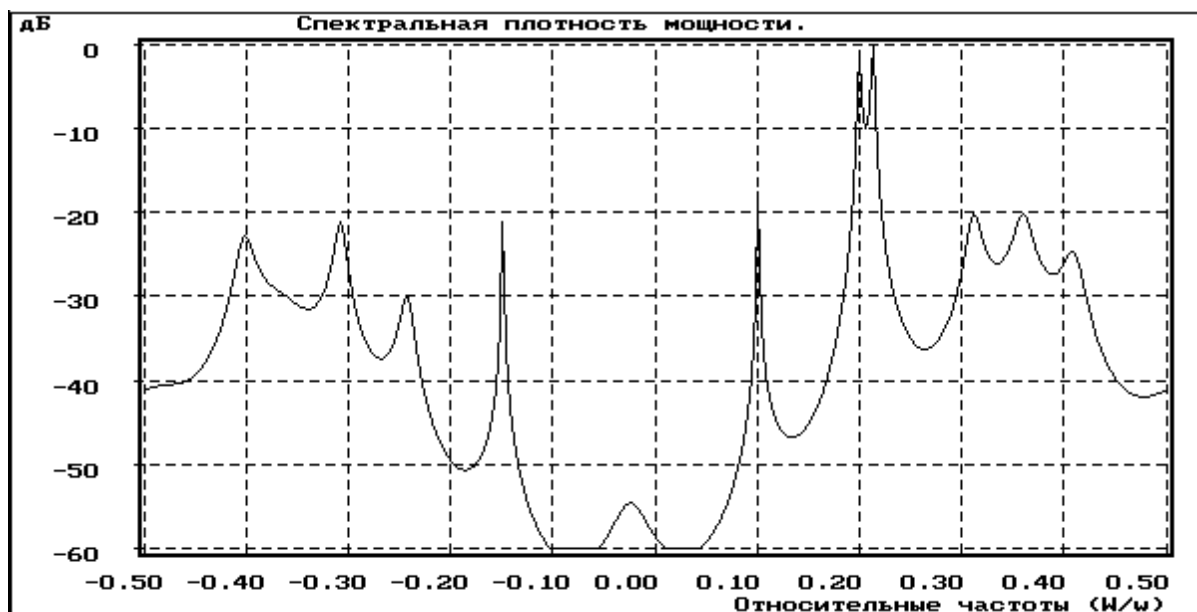


Рис. 6.12. AR(13)-оценка СПМ тест-последовательности Марпла, полученная методом Берга

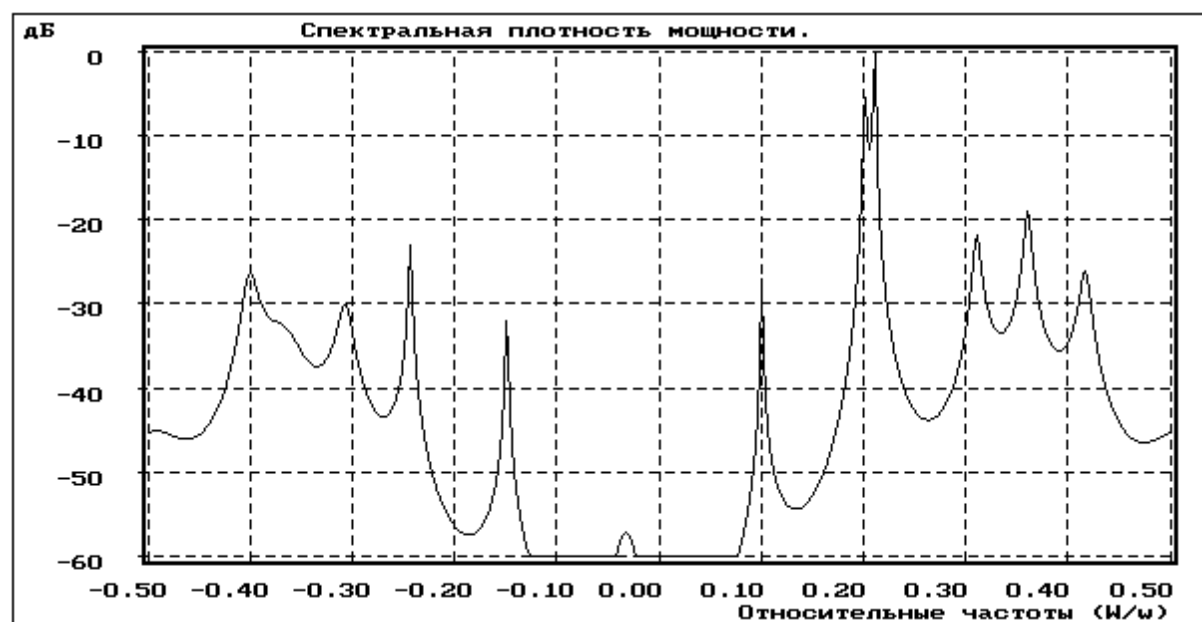


Рис. 6.13. AR(13)-оценка СПМ тест-последовательности Марпла, полученная ковариационным методом

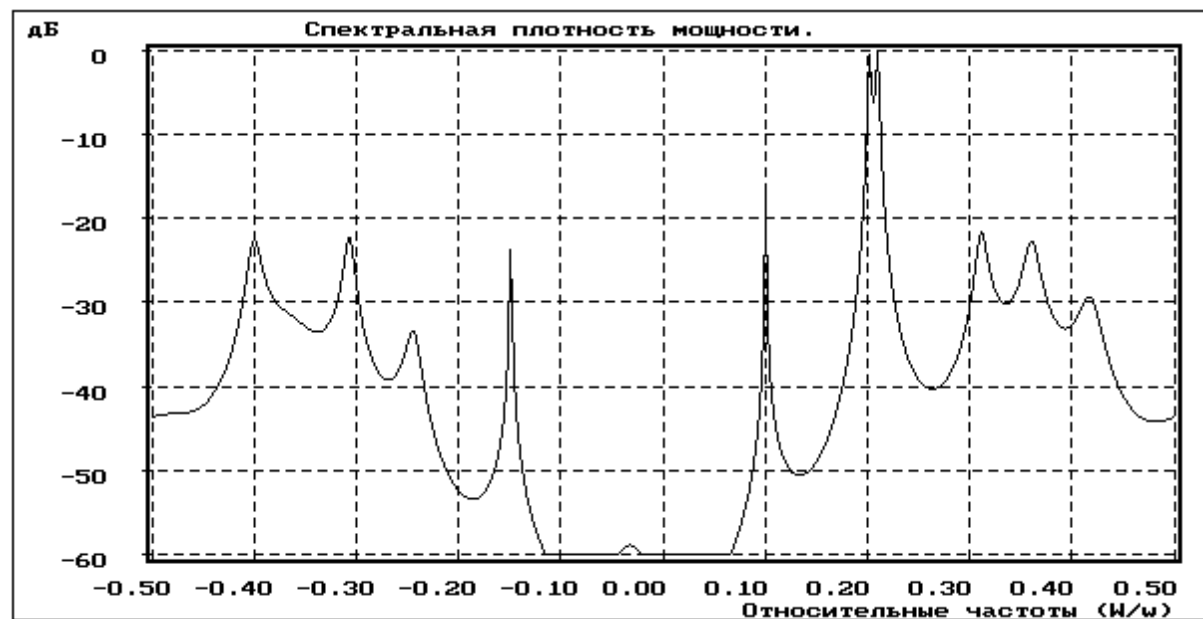


Рис. 6.14. AP(13)-оценка СПМ тест-последовательности Марпла, полученная модифицированным ковариационным методом

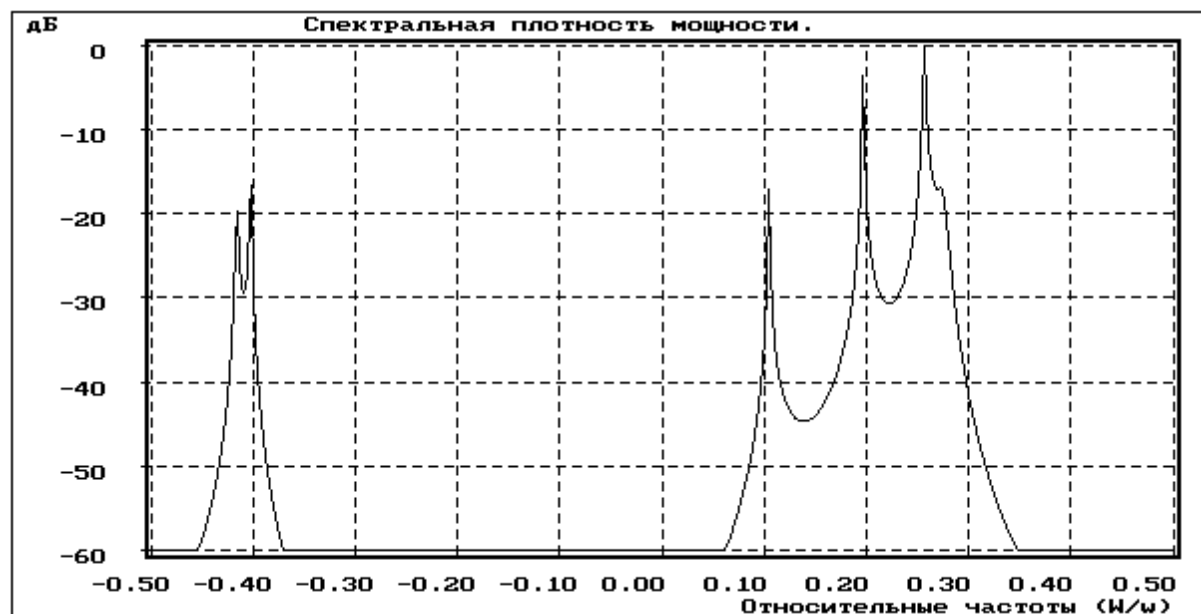


Рис. 6.15. AP(28)-оценка СПМ суммы четырех синусоид, полученная методом Берга, иллюстрирующая эффект расщепления спектральной линии на частоте -0.4

Измерения характеристик точности оценки частоты синусоидального сигнала в шуме в зависимости от длины реализации, отношения сигнал / шум для различных методов AP-оценивания СПМ показали, что дисперсия и смещение оценки частоты синусоиды в шуме примерно обратно пропорциональны длине реализации и отношению сигнал / шум. Абсолютные значения смещения и дисперсии оценки частоты синусоиды с частотой 0.2 при отношении сигнал / шум = -10 дБ составили соответственно величины 0.075 и 0.04 для реализации длиной 5 отсчетов и 0.01 и 0.025 - для реализации длиной 15 отсчетов. Для алгоритма Берга зафиксирован эффект расщепления спектральных линий, отображенный на рис. 6.15, где представлена спектральная AP(28)-оценка суммы четырех синусоид с частотами -0.4, 0.1, 0.2, 0.25, по реализации длиной 56 отсчетов при дисперсии шума равной 10^{-6} . Расщепленным оказался спектральный пик, соответствующий частоте -0.4. Эффект расщепления наблюдался, как правило, для малых длин реализаций, большого порядка модели и малой дисперсии шума при использовании алгоритма Берга.

Проведенные измерения спектральных оценок для процессов, соответствующих АР-моделям, при различных длинах реализаций, уровне аддитивного шума наблюдения, порядке модели показали, что в случае, когда измеряемый процесс и его модель соответствуют друг другу, большинство методов АР-оценивания СПМ дают хорошие спектральные оценки, дисперсия которых уменьшается обратно пропорционально длине реализации и отношению сигнал / шум. На рис. 6.16 и 6.17 в качестве примера приведены соответственно истинная СПМ и ее АР(4)-оценка методом Берга последовательности, состоящей из 40 отсчетов следующего АР(4)-процесса:

$$x(n) = 2.7607 x(n-1) - 3.8106 x(n-2) + 2.6535 x(n-3) - 0.9238 x(n-4) + u(n),$$

где $u(n)$ - белый гауссовский шумовой процесс. На рис.6.18 эти спектрограммы совмещены для лучшего уяснения их различий.

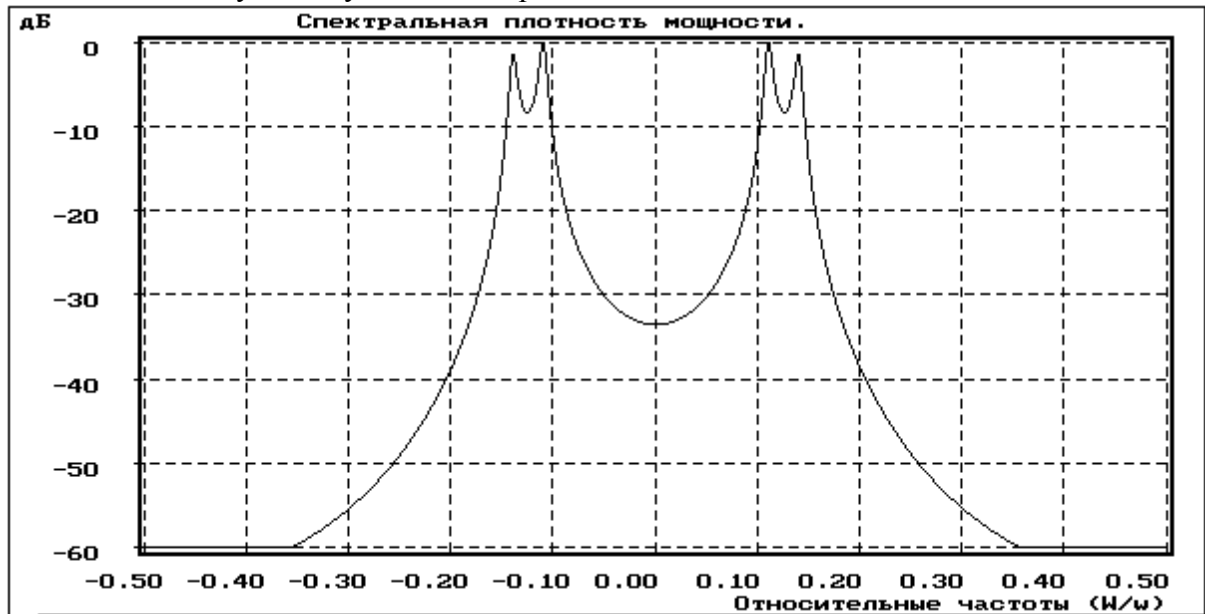


Рис. 6.16. Истинный спектр АР(4)-процесса

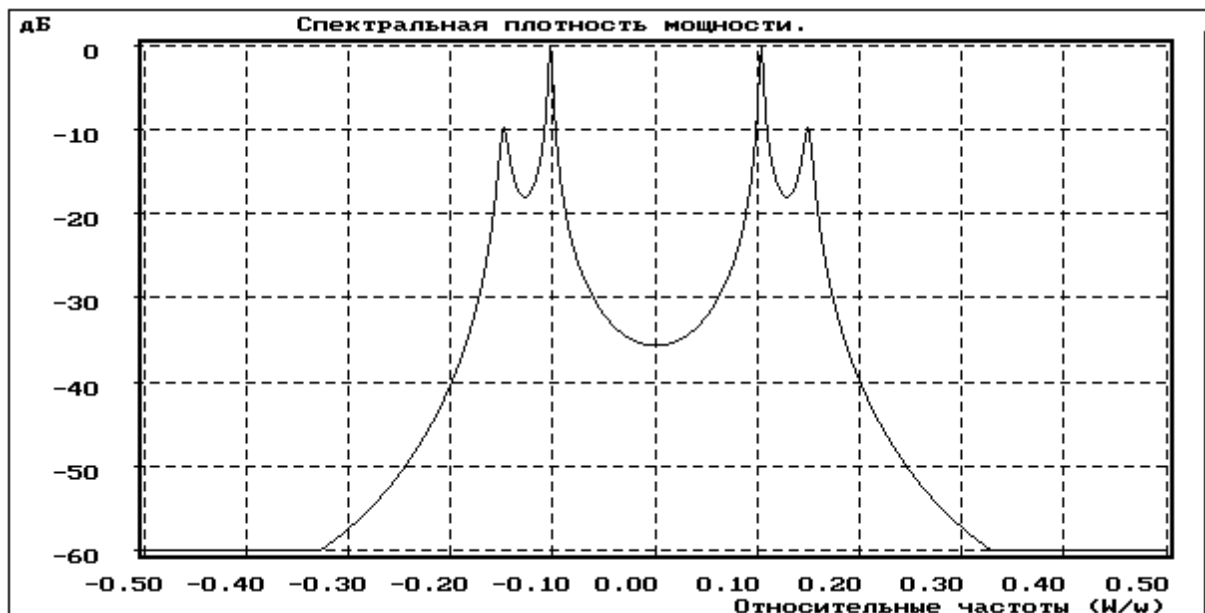


Рис. 6.17. АР(4)-оценка СПМ методом Берга модельного АР(4)-процесса

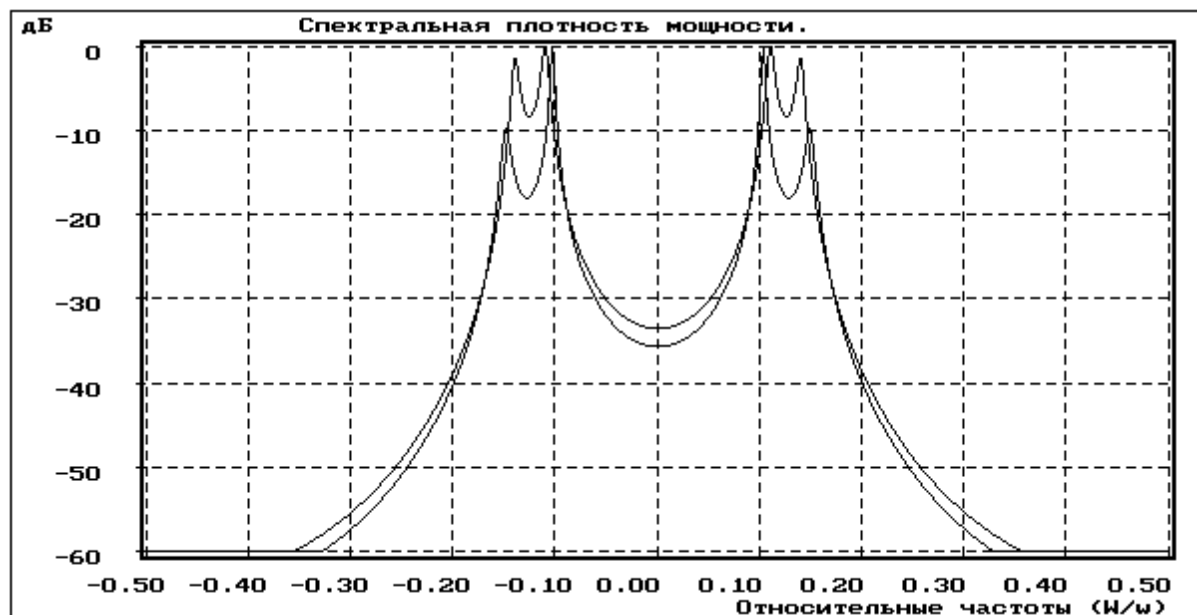


Рис. 6.18. Совмещенные спектрограммы, представленные на рис. 6.16 и 6.17.

На рис. 6.19 представлен истинный спектр АРСС(2, 2)-процесса, удовлетворяющего следующему разностному уравнению:

$$x(n) = 1.2727922 x(n-1) - 0.81 x(n-2) + u(n) - u(n-1),$$

где $u(n)$ - белый гауссов шум. Пики спектра этого процесса соответствуют относительным частотам 0.125 и -0.125, а провалы (нули) - частотам 0 и 0.5. На рис. 6.20 представлена АРСС(2, 2)-оценка СПМ этого процесса без шума наблюдения для реализации длиной 100 отсчетов при оценке АР-параметров ковариационным методом, СС-параметров - методом Юла - Уолкера с использованием эквивалентной длинной АР(20)-модели, параметры которой оценивались ковариационным алгоритмом. Очевидно хорошее соответствие оценки истинному спектру.

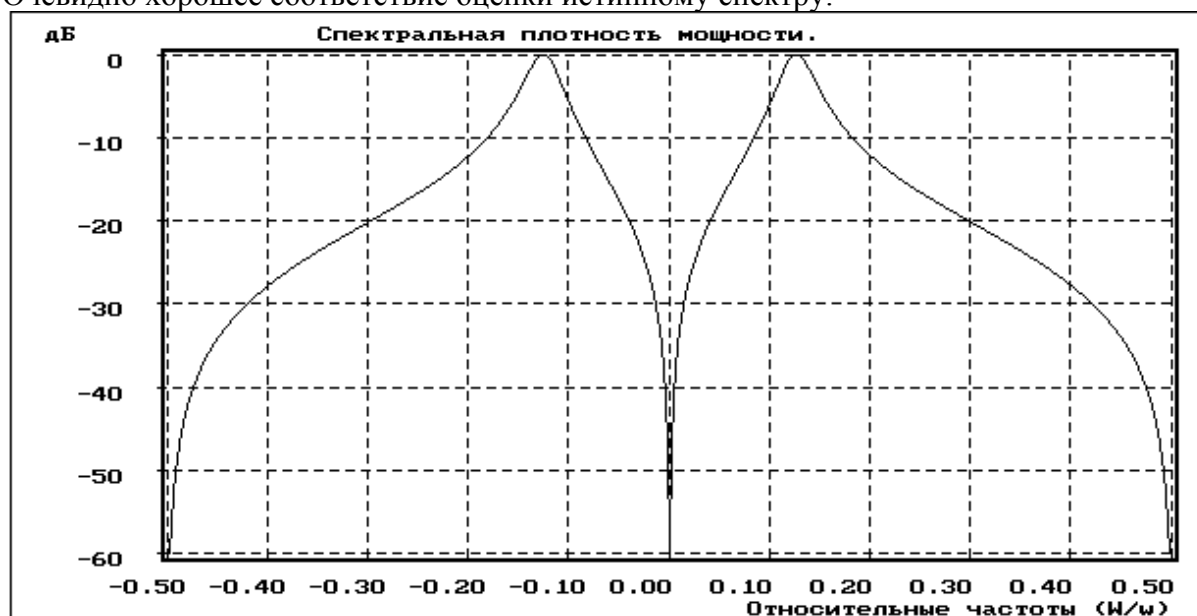


Рис. 6.19. Истинный спектр АРСС (2, 2)-процесса

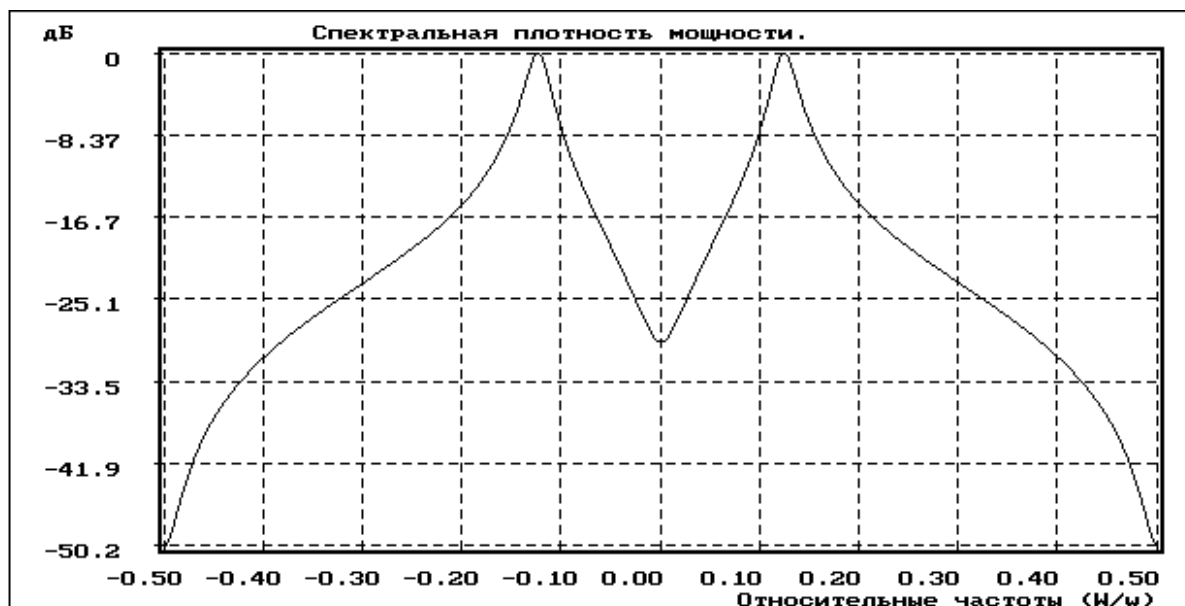


Рис. 6.20. APCC (2, 2)-оценка СПМ процесса, спектр которого отображен на рис. 6.19, по реализации длиной 100 отсчетов.

Обработка реальных временных рядов иллюстрируется на примерах спектрального анализа чисел солнечных пятен (чисел Вольфа), наблюдения за которыми имеют давнюю историю и связаны с поиском их периодичностей. Сведения о среднемесячных числах солнечных пятен за период с 1700 по 1997 г. получены из Internet'a и представляют собой реализацию длиной 2975 значений. AP(75)-оценка методом Берга 1000 значений этого процесса представлена на рис. 6.21, где на фоне сильного пика постоянной составляющей, обусловленного положительностью чисел солнечных пятен, замечен пик на частоте, близкой к 0.09 цикла на год, соответствующий классическому 11-летнему периоду активности солнечных пятен. Для более точного измерения параметров этого пика и других выполнена децимация (прореживание) временного ряда в 6 раз с тем, чтобы частота отсчетов снизилась до двух в год, при этом относительная частота, равная 0.5, на спектрограмме будет соответствовать одному циклу в год и, следовательно, пик, соответствующий частоте 0.1 цикла на год, будет на относительной частоте, равной 0.05. Для устранения эффекта наложения исходная реализация перед децимацией обрабатывалась фильтром низких частот (ФНЧ), представляющим собой КИХ - фильтр с линейной ФЧХ, имеющий 49 коэффициентов и переходную зону в пределах от 0.042 до 0.083 относительной частоты. Прореженные в 6 раз числа солнечных пятен обрабатывались далее фильтром верхних частот (ФВЧ) с тем, чтобы устранить мешающее действие постоянной составляющей (эффект смещения близкого информационного пика). ФВЧ имел 51 коэффициент и переходную зону шириной от 0 до 0.004 относительной частоты.

На рис. 6.22 представлена AP(40)-оценка СПМ фильтрованных чисел солнечных пятен, полученная модифицированным ковариационным методом по реализации длиной 555 отсчетов. На этой спектрограмме видны сильные пики на частотах 0.04589 (0.09178 цикла на год - 10.9-летний цикл), 0.09277 (0.1855 цикла на год - 5.39-летний цикл), а также пик на частоте 0.0244 (0.0488 цикла на год - 20.49-летний цикл).

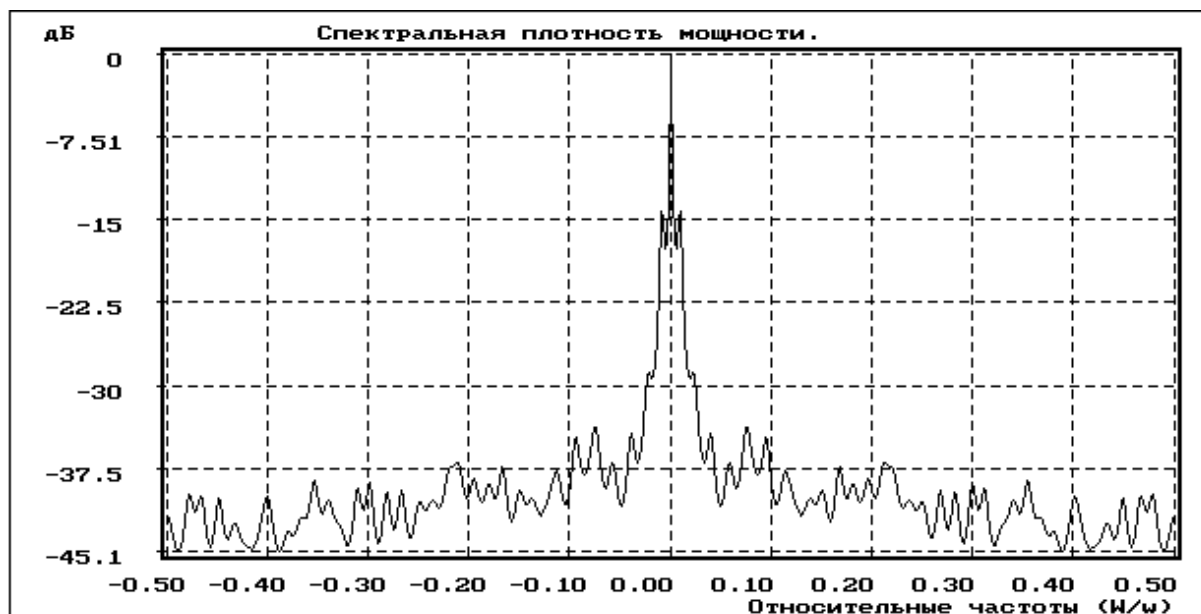


Рис.6.21. AP(75)-оценка СПМ среднемесячных чисел солнечных пятен, полученная методом Берга



Рис.6.22. AP(40)-оценка СПМ фильтрованных и прореженных чисел солнечных пятен, полученная модифицированным ковариационным методом.

6.6.3. Измерение спектров нестационарных сигналов

Поскольку частотно-временное разрешение параметрических алгоритмов спектрального оценивания в сравнении с классическими существенно выше, представляется целесообразным использование их для измерения переменных во времени спектров таких, например, как спектры частотно-модулированных, импульсных и других сигналов с целью определения областей локализации их энергии в плоскости частота-время, измерений законов модуляции и т. д. Измерение переменного во времени (активного) спектра сигнала можно выполнить, анализируя спектры его сегментов, наблюдаемых через перемещающееся по длине реализации временное окно и связывая спектр сегмента с временным положением (началом) окна. Совокупность полученных спектров образует в координатах интенсивность, частота, время частотно-временное распределение энергии (мощности) сигнала (ЧВРМ), по которому и определяются соответствующие характеристики сигнала. Так, например, для ЧМ-сигнала проекция максимумов распределения на плоскость частота-время (так

называемый скелет максимумов) позволяет определить закон изменения во времени мгновенной частоты, вдоль которого в достаточно узкой полосе концентрируется энергия (мощность) сигнала.

На рис. 6.23 отображен закон изменения во времени мгновенной частоты комплексного ЧМ-сигнала, модулированного синусоидой с периодом 50 отсчетов (несущая частота - 0.1, девиация частоты - 0.3), измеренный с помощью скользящей AP(1)-оценки СПМ методом Берга (окно длиной 6 отсчетов, шаг перемещения окна 1 отсчет, длина реализации 100 отсчетов, отношение сигнал / шум 42 дБ). Именно вдоль этой кривой в весьма узкой области локализовано ЧВРМ данного ЧМ-сигнала.

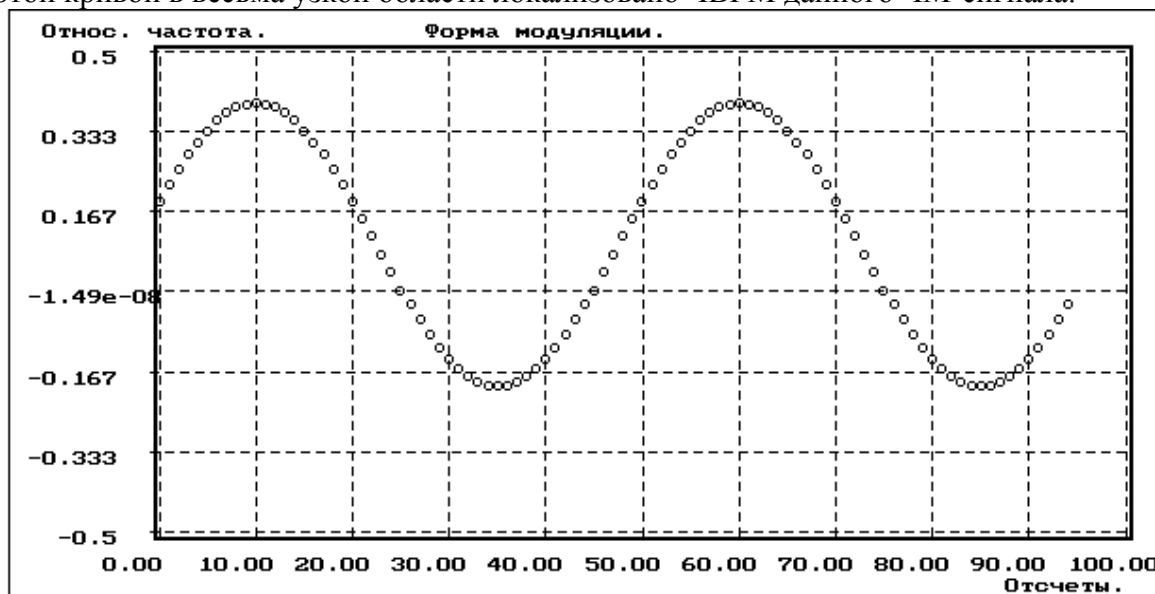


Рис. 6.23. Зависимость мгновенной частоты от времени ЧМ-сигнала, измеренная с помощью скользящей AP(1) спектральной оценки методом Берга

AP(20)-оценка СПМ методом Берга данного сигнала представлена на рис. 6.24, который отображает по сути дела проекцию ЧВРМ на вертикальную плоскость, проходящую через ось частот, т.е. обычный (пассивный) спектр широкополосного ЧМ-сигнала с базой равной $2 \cdot 0.3 \cdot 100 = 60$.

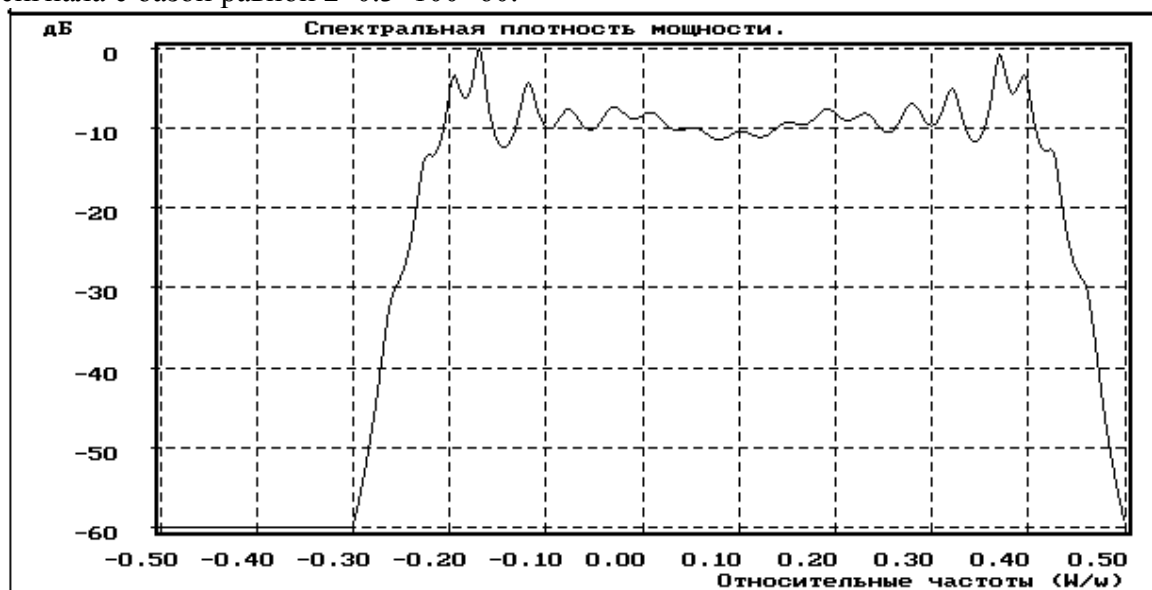


Рис. 6.24. AP(20)-оценка СПМ методом Берга ЧМ-сигнала с синусоидальным законом модуляции

На рис. 6.25 представлена частотно-временная картина, отображающая зависимости, вдоль которых в узкой области локализовано ЧВРМ суммы двух комплексных ЛЧМ-сигналов с противоположными законами модуляции (длина

реализации - 100 отсчетов, диапазон изменения частоты от - 0.3 до + 0.3), полученная с помощью скользящей AP(2)-оценки СПМ модифицированным ковариационным методом (длина окна - 6 отсчетов, шаг перемещения окна - 1 отсчет). AP(20)-оценка СПМ методом Берга данного широкополосного сигнала представлена на рис. 6.26.

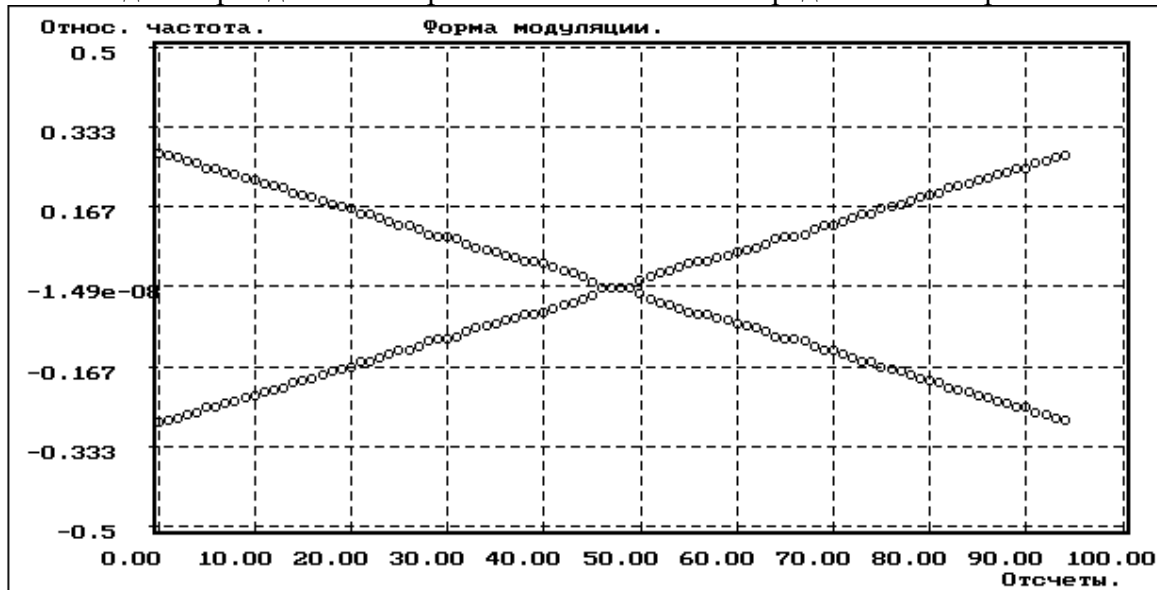


Рис. 6.25. Зависимости мгновенных частот от времени суммы двух ЛЧМ-сигналов, измеренные с помощью скользящей AP(2)-спектральной оценки модифицированным ковариационным методом.

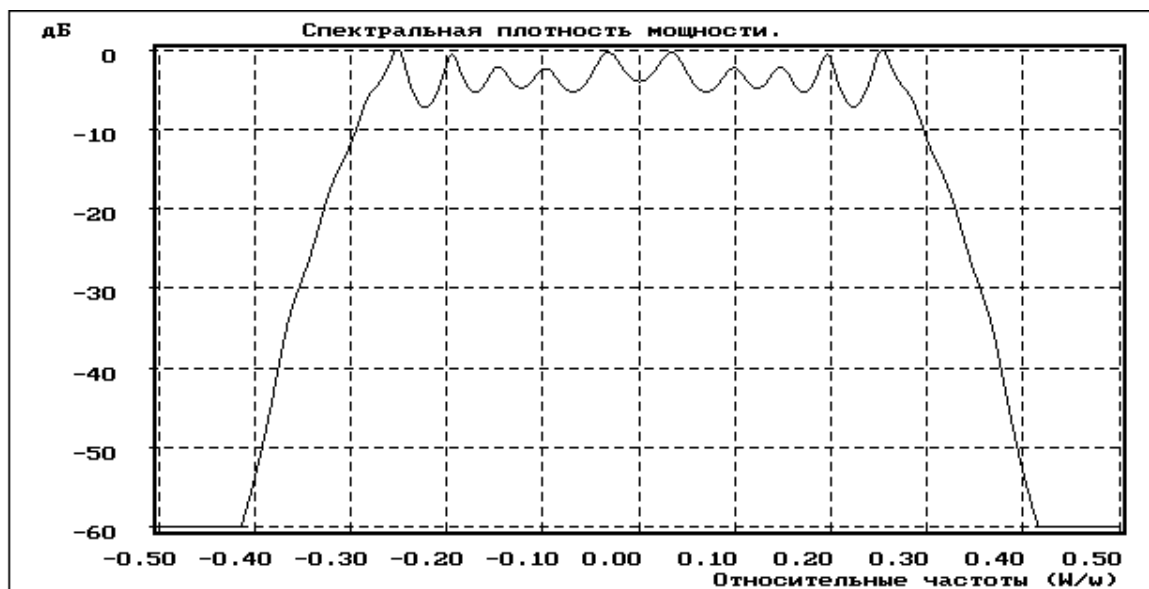


Рис. 6.26. AP(20)-оценка СПМ методом Берга суммы двух ЛЧМ-сигналов с встречными законами модуляции по реализации длиной 100 отсчетов.

На рис. 6.27 представлена частотно-временная зависимость, вдоль которой в узкой области локализовано ЧВРМ комплексного сигнала длиной 100 отсчетов, состоящего из суммы двух синусоид с частотами 0.15 и 0.25 со сдвигом моментов их появления на 50 отсчетов и шума (отношение сигнал / шум 28 дБ), полученная с помощью скользящей AP(2)-оценки СПМ ковариационным методом (длина окна - 6 отсчетов, шаг перемещения окна - 1 отсчет).

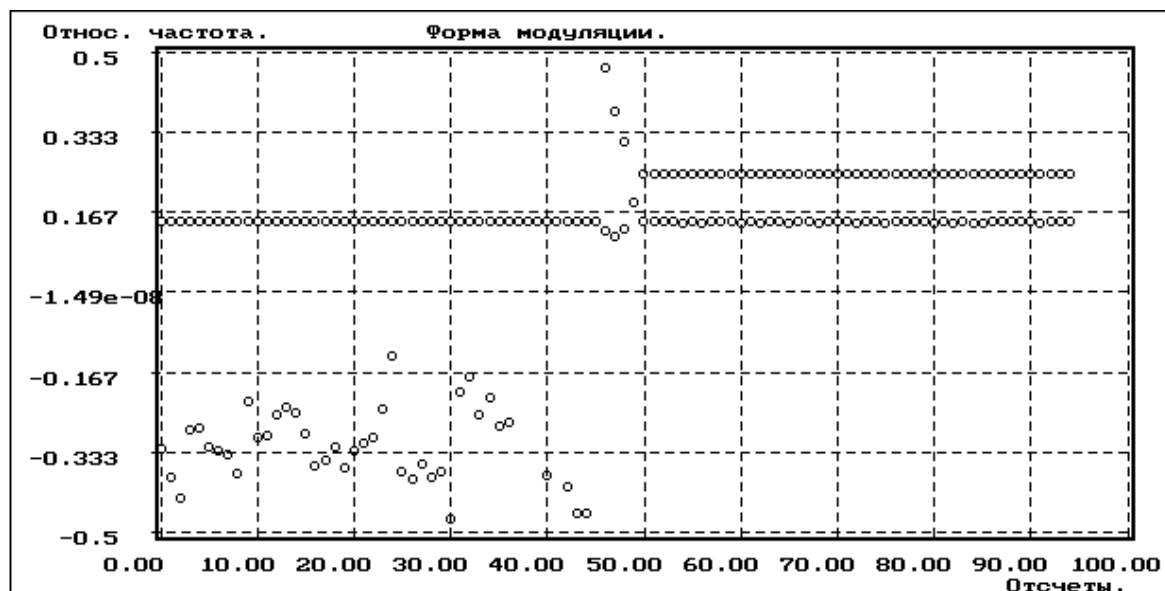


Рис. 6.27. Частотно-временная зависимость, соответствующая сумме двух синусоид с частотами 0.15 и 0.25, сдвинутых во времени на 50 отсчетов, полученная на основе скользящей AP(2)-оценки ковариационным методом.

На рис. 6.28 отображена частотно-временная зависимость положений спектральных пиков сигнала, соответствующего фильтрованной и прореженной последовательности ежемесячных чисел солнечных пятен длиной 555 отсчетов, полученной с помощью скользящей AP(8)-оценки СПМ модифицированным ковариационным методом (длина окна - 24 отсчета, шаг перемещения окна - 6 отсчетов), которая иллюстрирует нестационарность основного 11-летнего периода солнечной активности.

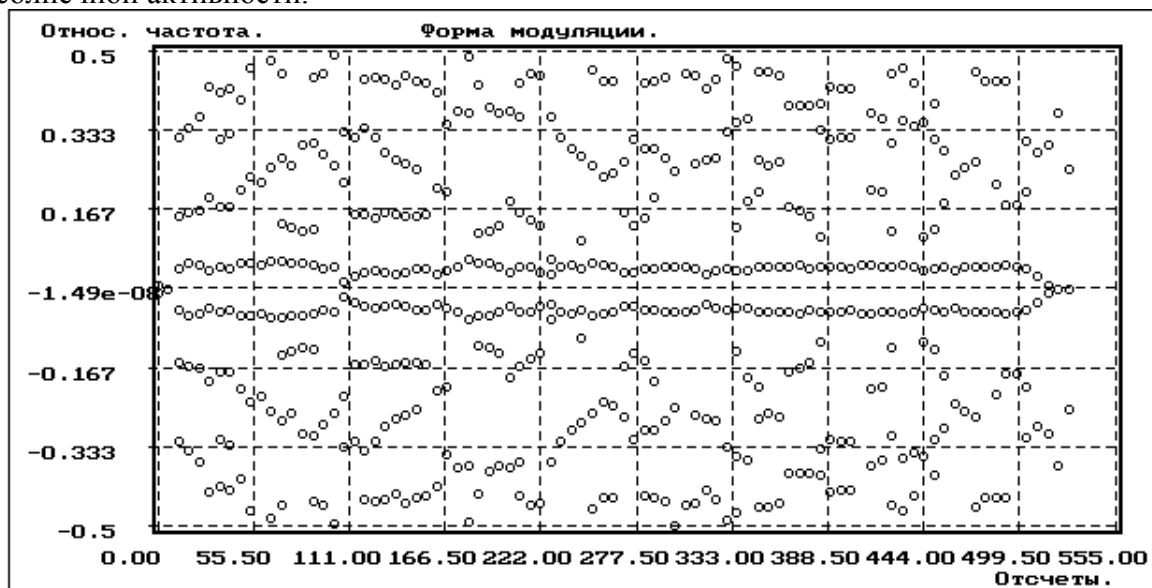


Рис. 6.28. Частотно-временная зависимость положений спектральных пиков фильтрованных чисел солнечных пятен, полученная на основе скользящей AP(8)- спектральной оценки модифицированным ковариационным методом

На рис. 6.29 отображена частотно-временная зависимость, соответствующая закону частотной модуляции комплексного ЛЧМ-сигнала с шумом наблюдения, вдоль которой в узкой полосе локализовано ЧВРМ этого сигнала (длина 100 отсчетов, начальная частота 0.1, девиация частоты 0.3, отношение сигнал / шум 42 дБ), полученная с помощью скользящей AP(1)-оценки СПМ методом Берга (длина окна - 6 отсчетов, шаг перемещения окна - 1 отсчет). На рис. 6.30 представлена AP(20)-оценка СПМ методом Берга этого широкополосного сигнала с базой равной $0.3 \cdot 100 = 30$.

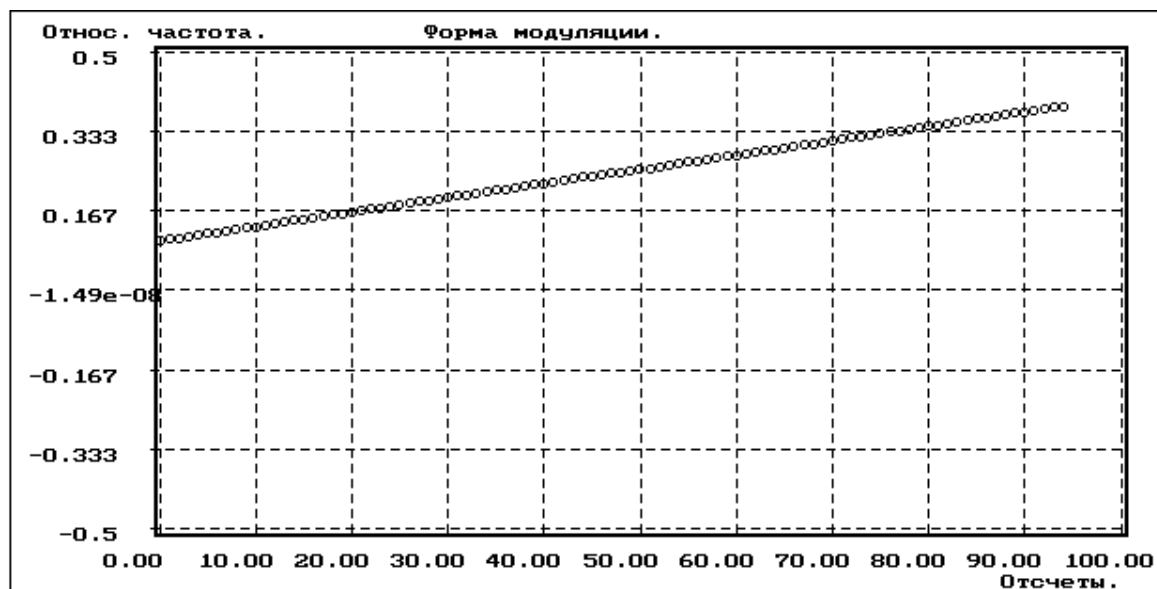


Рис. 3.29. Закон модуляции ЛЧМ-сигнала, измеренный с помощью скользящей AP(1)-оценки СПМ методом Берга

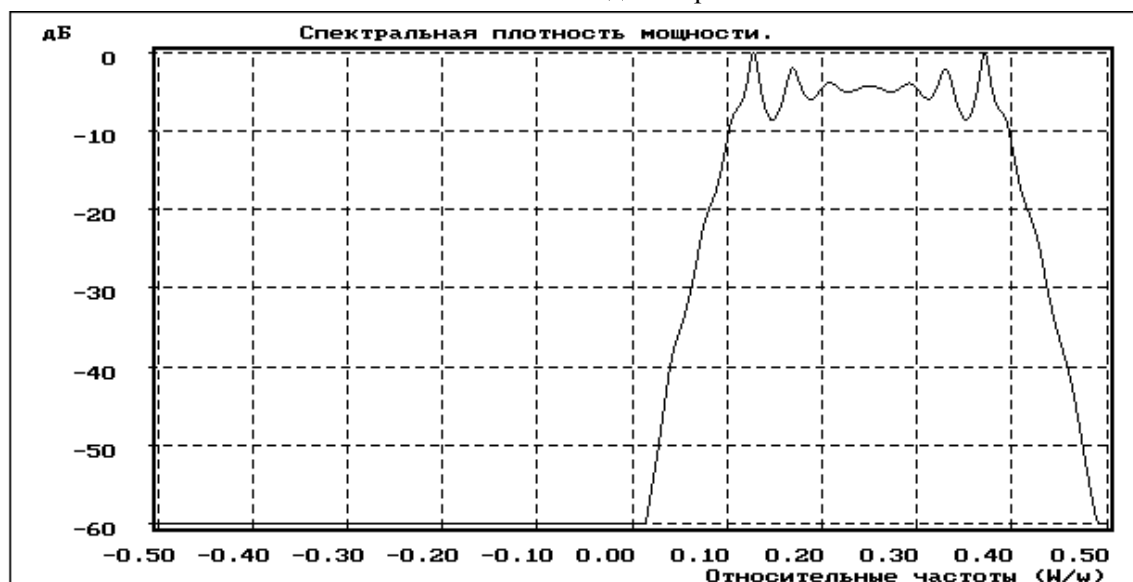


Рис. 6.30. AP(20)-оценка СПМ методом Берга ЛЧМ-сигнала с начальной частотой 0.1 и девиацией - 0.3 по реализации длиной 100 отсчетов.

Анализируя результаты измерений переменных во времени спектров и построения на их основе ЧВРМ сложных многокомпонентных сигналов, можно отметить следующие моменты. Для однокомпонентного, например, ЧМ-сигнала при большом отношении сигнал / шум длина скользящего окна может быть выбрана предельно малой (4-6 отсчетов), что позволяет осуществить весьма точные измерения ЧВРМ и закона модуляции сигнала. Увеличение числа компонент сигнала, а также уменьшение отношения сигнал / шум требует соответствующего увеличения порядка модели, что, в свою очередь, вызывает необходимость расширения окна и уменьшает точность измерения законов модуляции. При скользящих измерениях активных спектров многокомпонентных сигналов на основе параметрических методов не возникают перекрестные компоненты, затрудняющие интерпретацию ЧВРМ, что выгодно отличает их от других нелинейных алгоритмов СВА, таких, например, как алгоритм Вигнера-Вилля [34]. Измерения пассивных спектров сложных широкополосных ЧМ-сигналов и других на основе параметрических моделей показали возможность точного воспроизведения их формы при правильном выборе порядка моделей.

7. Аналого-цифровое и цифро-аналоговое преобразование

7.1. Двоичное представление сигналов

Как Вы уже знаете, цифровые сигналы представляют собой последовательности чисел. Сейчас мы поговорим о том, как кодируются числа.

Цифровые компьютеры - большие специалисты в хранении и обработке чисел; к несчастью, этот процесс не обходится без ошибок. Например, Вы указываете вашему компьютеру сохранить число: 1.41421356. Компьютер делает все возможное, сохраняя ближайшее число, которое может быть им представлено: 1.41421354. В некоторых случаях эта ошибка совсем незначительная, в то время как в других случаях - катастрофическая. В качестве другой иллюстрации напомним о классической ошибке в вычислениях, происходящей при сложении двух чисел с весьма разными величинами, например, 1 и 0.00000001. Мы хотели бы иметь ответ 1.00000001, но компьютер отвечает 1. Подобные проблемы сплошь и рядом возникают при реализации устройств и алгоритмов цифровой обработки сигналов, в частности невозможность представить с заданной точностью, например, коэффициенты цифрового фильтра, искажает частотную характеристику системы ЦОС (реальная АЧХ может существенно отличаться от расчетной).

Корень проблемы – ограниченное число двоичных разрядов (битов), выделяемых для представления числа. Обычно это 8, 16, 32 или 64 бита, хотя в устройствах ЦОС бывает также и 6, 10, 12, 14 бит (таким числом бит часто представляются результаты аналого-цифрового преобразования). Вообще N битами можно представить не более 2^N разных чисел. Это фундаментальное ограничение. Мы можем лишь менять диапазон представляемых чисел (только положительные, только положительные и отрицательные, только целые, целые и дробные), договариваясь об алгоритме или формуле пересчета двоичного представления числа в его действительное значение, т.е. о типе кодирования.

Наиболее широко применяются так называемые коды с «фиксированной» и «плавающей» точкой.

К кодам с фиксированной точкой относятся *прямой двоичный код*, *смещенный код*, *код модуль со знаком* и *дополнительный код*.

N -разрядный прямой двоичный код это *строка бит*

$B_{N-1} B_{N-2} B_{N-3} \dots B_0$, представляющая целое число

$$D = 2^{N-1} \cdot B_{N-1} + 2^{N-2} \cdot B_{N-2} + 2^{N-3} \cdot B_{N-3} + \dots + 2^0 \cdot B_0.$$

Диапазон представляемых чисел $0 \div 2^N - 1$ (для $N=8$ - $0 \div 255$).

Числа с дробной частью также могут быть представлены прямым двоичным кодом путем добавления бит, являющихся весами отрицательных степеней двойки

$$D = 2^{N-1} \cdot B_{N-1} + 2^{N-2} \cdot B_{N-2} + 2^{N-3} \cdot B_{N-3} + \dots + 2^0 \cdot B_0 + 2^{-1} \cdot B_{-1} + 2^{-2} \cdot B_{-2} + \dots,$$

или перераспределением бит

$$D = 2^{N-3} \cdot B_{N-1} + 2^{N-4} \cdot B_{N-2} + \dots + 2^0 \cdot B_2 + 2^{-1} \cdot B_1 + 2^{-2} \cdot B_0.$$

При представлении отрицательных чисел возможны варианты: смещенный код, модуль со знаком и дополнительный код.

Прямой двоичный		Смещенный		Модуль со знаком		Дополнительный	
Десятичн.	Двоичн.	Десятичн.	Двоичн.	Десятичн.	Двоичн.	Десятичн.	Двоичн.
15	1111	8	1111	7	0111	7	0111
14	1110	7	1110	6	0110	6	0110
13	1101	6	1101	5	0101	5	0101
12	1100	5	1100	4	0100	4	0100
11	1011	4	1011	3	0011	3	0011
10	1010	3	1010	2	0010	2	0010
9	1001	2	1001	1	0001	1	0001

8	1000	1	1000	0	0000	0	0000
7	0111	0	0111	-1	1001	-1	1111
6	0110	-1	0110	-2	1010	-2	1110
5	0101	-2	0101	-3	1011	-3	1101
4	0100	-3	0100	-4	1100	-4	1100
3	0011	-4	0011	-5	1101	-5	1011
2	0010	-5	0010	-6	1110	-6	1010
1	0001	-6	0001	-7	1111	-7	1001
0	0000	-7	0000	-8	-	-8	1000

Смещенный код часто используется при аналого-цифровом и цифро-аналоговом преобразовании. Так, например, диапазон входных напряжений от -5В до 5В может отображаться в цифровых кодах 0 - 4095 для 12 битового преобразования.

При вычислениях наиболее употребительным является дополнительный код.

Кодирующая схема для чисел с плавающей точкой более сложная, нежели для чисел с фиксированной точкой. Основная идея та же, что использована в научной нотации, где мантисса умножается на десять в некоторой степени. Например, $5.4321 \cdot 10^6$, где 5.4321 - мантисса и 6 - показатель. Данный способ позволяет представлять как очень большие, так и очень маленькие числа. Обратите внимание, что числа представленные в научной нотации нормируются так, что есть только единственная ненулевая цифра слева от десятичной точки. Это достигается соответствующей «подстройкой» показателя степени.

Наиболее широко распространенный формат представления чисел с плавающей точкой определяется стандартом ANSI/IEEE Std. 754-1985.

Этот стандарт определяет формат для 32 битовых чисел одинарной точности, а также 64 битовых чисел двойной точности.

Биты 0 - 22 формируют мантиссу, биты 23 - 30 показатель степени, и бит 31 - бит знака. Значение числа определяется следующим соотношением:

$$D = (-1)^S \times M \times 2^{E-127}.$$

Здесь S - знаковый разряд, S = 0 для положительного числа и 1 для отрицательного номера. Переменная, E, - число между 0 и 255 представляет восемь бит показателя степени 2 со смещением 127. Мантисса M формируется 23-мя битами.

7.2. Дискретизация и квантование

Большинство сигналов, встречающихся в повседневной жизни человека, научной и инженерной деятельности, являются непрерывными: интенсивность света, которая изменяется с расстоянием; напряжение, которое изменяется над временем; скорость химической реакции, которая зависит от температуры, и т.п.. *Аналого-цифровое преобразование* (АЦП) и *Цифро-аналоговое преобразование* (ЦАП) - процессы, которые позволяют цифровым компьютерам взаимодействовать с этими повседневными сигналами и, соответственно, использовать широчайшие возможности цифровых алгоритмов для обработки аналоговых сигналов.

Переход к цифровому представлению сигнала состоит из двух составляющих: *дискретизации*, производимой путем взятия *выборок* из аналогового сигнала, и *квантования*. Обе этих составляющих ограничивают количество информации, которую цифровой сигнал может содержать. Информация, которую нужно сохранить при переходе от аналогового к цифровому представлению сигнала и обратно диктует выбор частоты дискретизации, количество битов, и тип аналоговых фильтров.

Рисунок 7.1 показывает “осциллограммы” типичного аналого-цифрового преобразования. Для простоты допустим, что напряжение может измениться от 0 до 4.095 вольт, соответствуя цифровым кодам от 0 до 4095, т.е. 12 двоичным разрядам. Обратите внимание на то, что аналого-цифровое преобразование разбито на два блока: выборку-

хранение (УВХ) и собственно аналого-цифровое преобразование (АЦП).

Разделение оцифровки на эти два этапа - важная теоретическая модель для понимания процессов и преобразований сигнала, хотя для некоторых типов аналого-цифровых преобразователей действительно необходимо на время преобразования поддерживать уровень входного напряжения постоянным.

На рис.7.1 показаны сигналы, иллюстрирующие процесс оцифровки. Преобразование разбивается на два этапа, чтобы подчеркнуть эффекты дискретизации, выделив их из эффектов квантования. После первого этапа преобразования сигнал становится дискретным, но не квантованным, точнее ИЗМЕНЕНИЯ сигнала происходят в дискретные моменты времени. На втором этапе напряжение преобразуется в ЦЕЛОЧИСЛЕННЫЙ цифровой код. Округление до ближайшего целого приводит к появлению ошибки квантования величиной до половины напряжения, соответствующего изменению кода на единицу (LSB). Таким образом, квантование всегда сопровождается добавлением шума к сигналу (обратите внимание на плотность вероятности и спектральную плотность мощности шума квантования). СПМ уменьшается при увеличении частоты дискретизации!

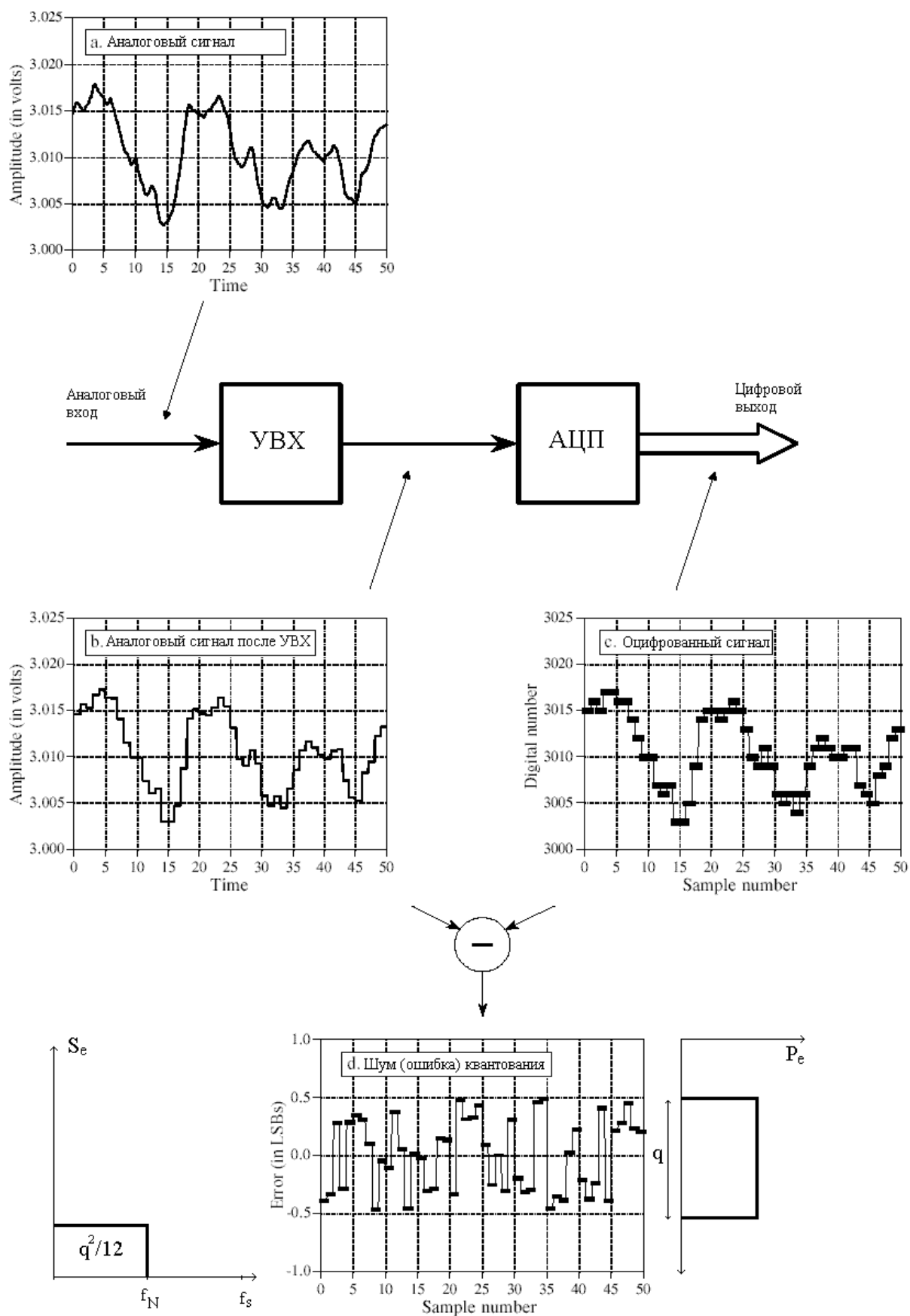


Рис. 7.1

Эффективное напряжение шума квантования равно $q/\sqrt{12} = 0.29 \cdot q = 0.29 \cdot (U_{\max} - U_{\min}) / 2^N$.

Пусть $U_{\max} - U_{\min} = 5\text{В}$. Тогда эффективное напряжение шума квантования будет примерно равно:

- для 8-разрядного АЦП - 5.7 мВ;
- для 12-разрядного - 0.35 мВ;
- для 16-разрядного - 22 мкВ.

Причем это напряжение «размазано» по полосе частот от 0 до половины частоты выборки.

Когда эта модель квантования не верна? Только когда ошибка квантования не может рассматриваться как случайная. Единственный случай - это когда аналоговые сигнальные остатки примерно одинаковы для многих последовательных выборок, как иллюстрируется на рис.7.2а. Выходные коды оказываются одинаковыми для многих выборок подряд, даже если аналоговый сигнал изменяется вплоть до $\pm 1/2$ младшего значащего разряда. Вместо случайного дополнительного шума, ошибка квантования теперь проявляется как пороговый эффект или нелинейное искажение сигнала.

Для улучшения квантования таких слабо изменяющихся сигналов применяется их *зашумление (dithering)*. Соответствующие иллюстрации приведены на рис.7.2.

Это довольно странная ситуация: добавление шума обеспечивает получение большей информации о сигнале! Процедуру можно усложнить: генерировать шум в компьютере (сигнальном процессоре), выводить через ЦАП, подмешивать к сигналу, а после оцифровки вычитать.

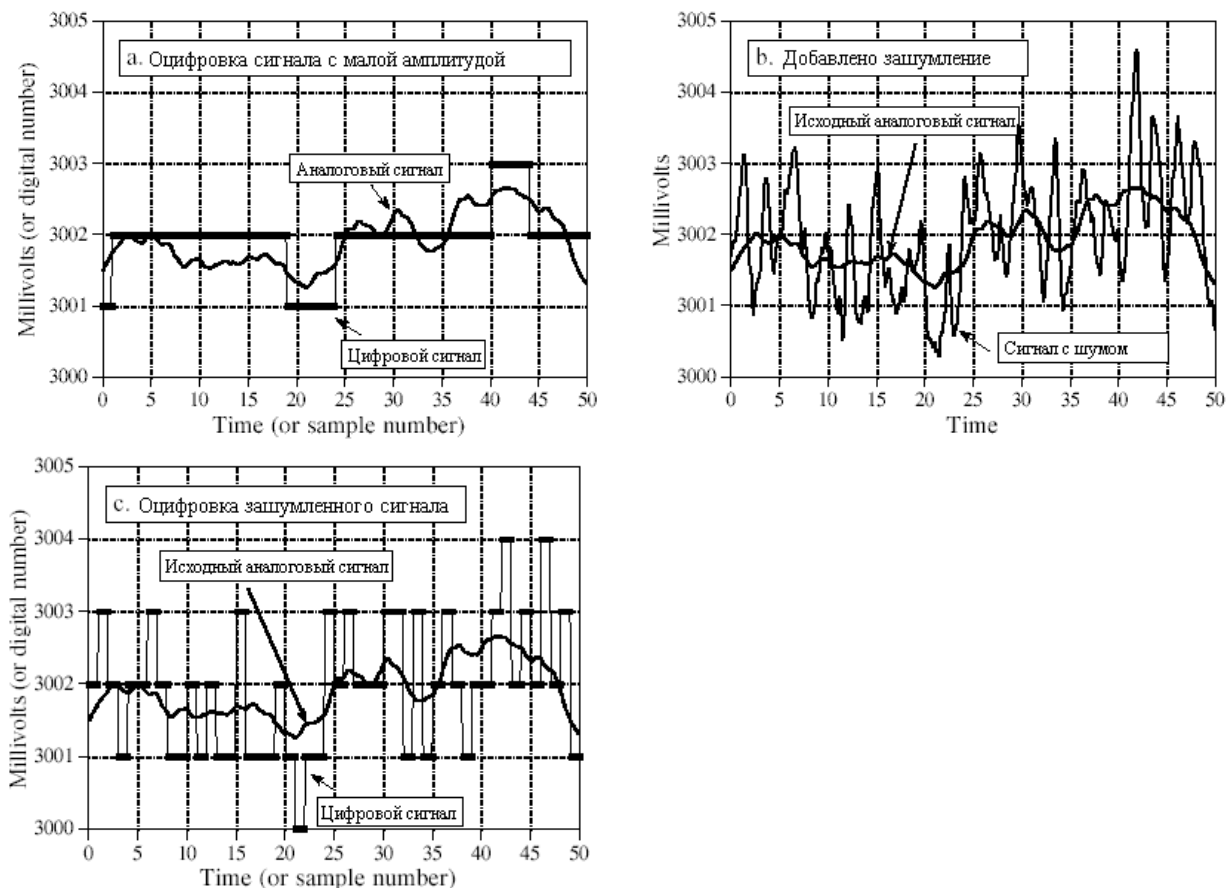


Рис. 7.2

Квантуются не только отсчеты сигнала, но и другие постоянные и переменные величины, участвующие в вычислениях в соответствии с алгоритмом цифровой обработки сигналов, например, весовые коэффициенты цифровых фильтров. На следующих рисунках проиллюстрировано влияние квантования коэффициентов на частотную характеристику трансверсального цифрового фильтра низкой частоты 255-го порядка. На

рисунке 7.3 весовые коэффициенты имеют разрядность, равную 16, на рисунке 7.4 – 10.

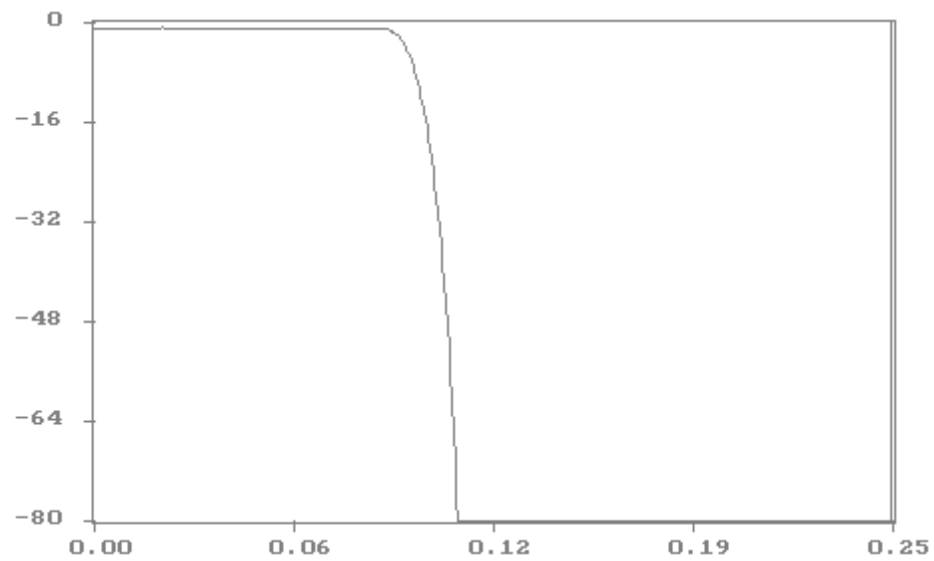


Рис. 7.3

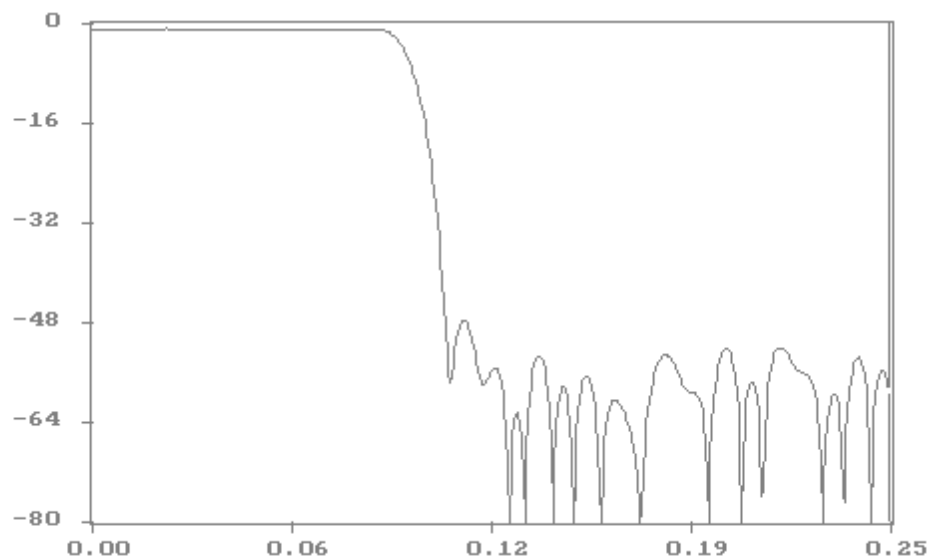


Рис. 7.4

Кратко напомним об эффектах, сопровождающих дискретизацию сигнала (рис. 7.5, 7.6).

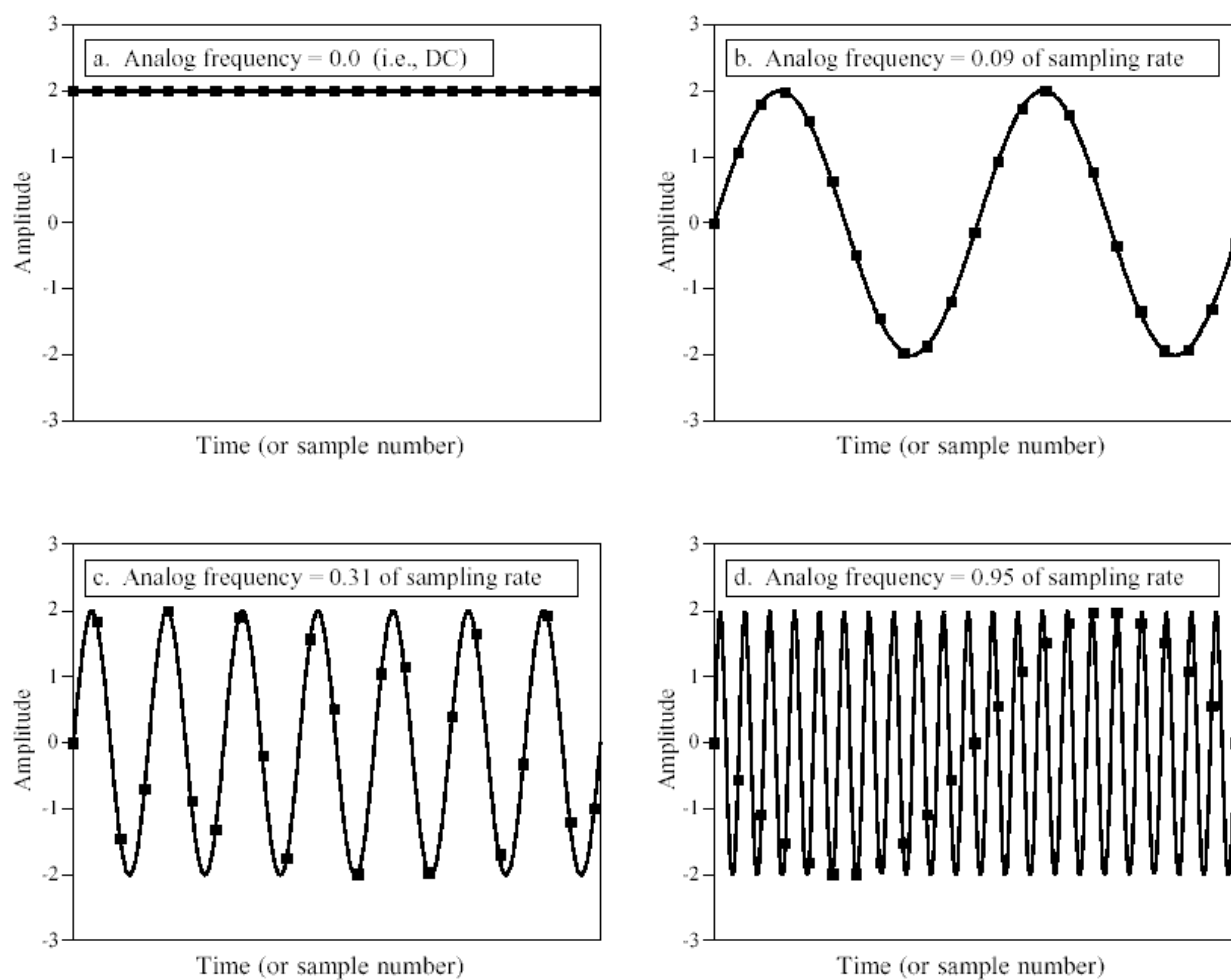


Рис.7.5

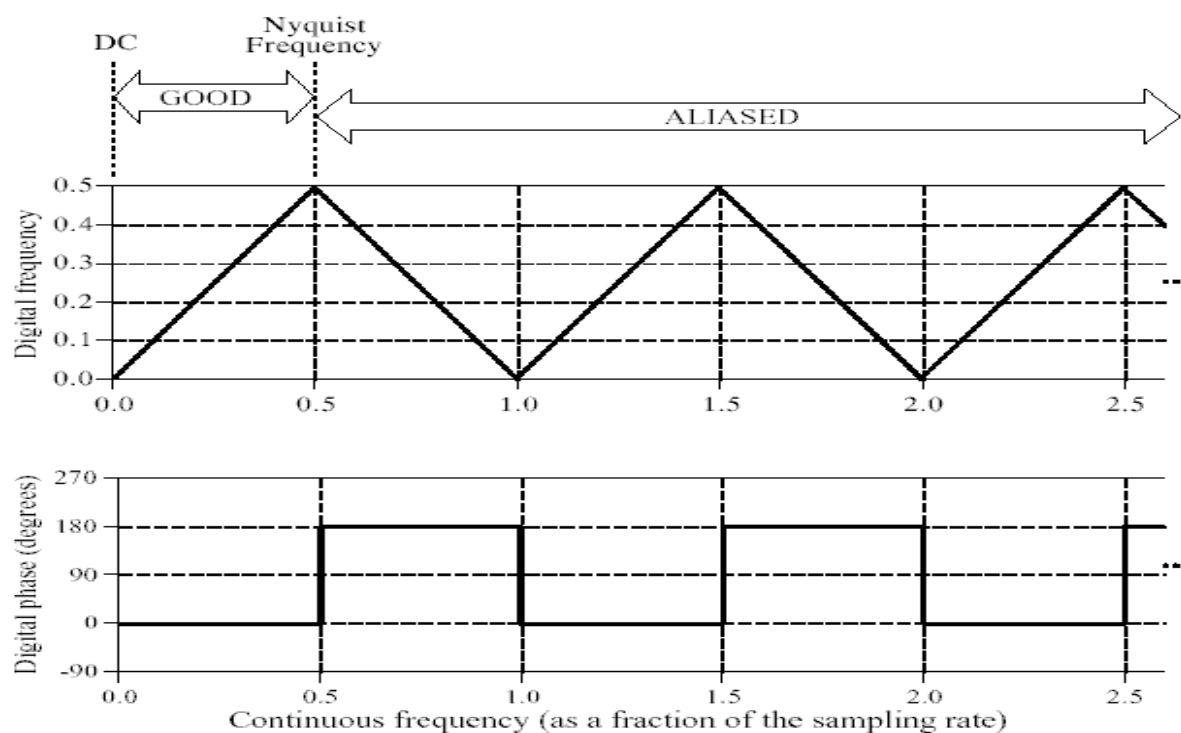


Рис. 7.6

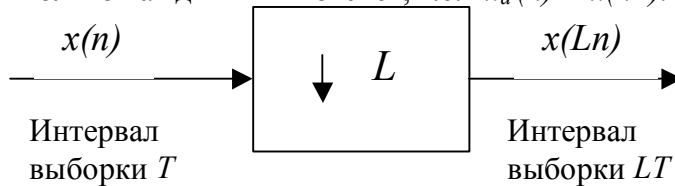
7.3. Децимация и интерполяция

Часто при реализации систем цифровой обработки сигналов возникает необходимость изменять (понижать или повышать) частоту дискретизации. Процесс понижения частоты дискретизации называется децимацией или прореживанием, обратный процесс – интерполяцией. В частности, так называемые сигма-дельта аналого-цифровые и цифро-аналоговые преобразователи, принцип действия которых мы рассмотрим ниже, используют децимацию и интерполяцию.

Здесь мы рассмотрим наиболее простой случай – изменение частоты выборки в целое число раз.

Уменьшение частоты дискретизации в целое число раз

Для выполнения операции прореживания в L раз из последовательности $x(n)$ берется только каждый L -ый отсчет, т.е. $x_d(n) = x(nL)$.



Как Вам известно, чтобы избежать наложения частот сигнал $x(t)$ и, соответственно, $x(n)$ должен иметь полосу частот $f_M < 1/T$ (быть отфильтрован фильтром с частотой среза $f_C < 1/T$). Для сигнала, прореженного в L раз, полоса частот должна быть, соответственно, $f_{L,M} < 1/LT$.

Рассмотрим ДВПФ прореженного сигнала

$$X_d(f) = \frac{1}{L} \sum_{i=0}^{L-1} X\left(\frac{f-i}{L}\right).$$

Мы видим, что он представляет собой сумму промасштабированных и сдвинутых копий спектра исходной последовательности $x(n)$. Вспомним, что подобное преобразование спектра происходит при дискретизации сигнала.

Условие отсутствия наложения спектров при децимации есть $f_{L,M} < 1/LT$ (для круговых частот $\omega_M < \pi / L$). На следующем рисунке представлен случай, когда данное условие выполняется.

$$f_{L,M} < 1/LT.$$

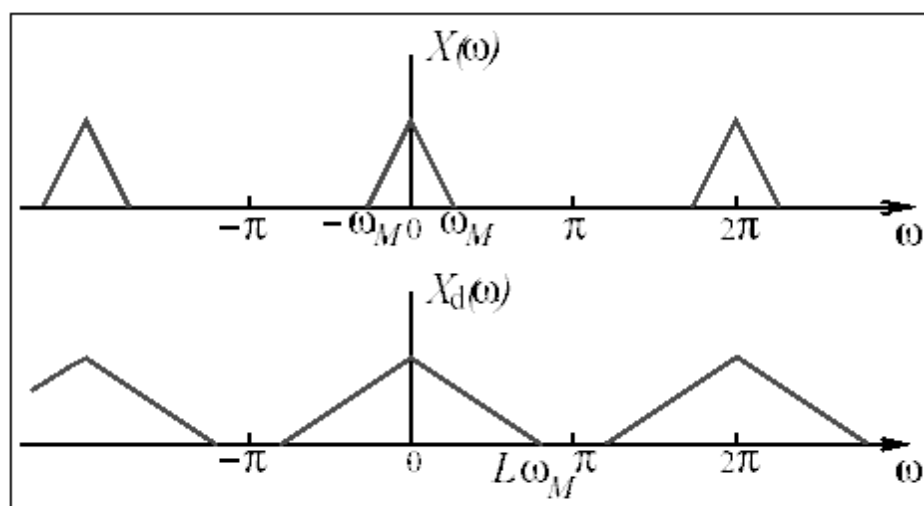


Рис. 7.7

Далее представлено прореживание сигнала, сопровождаемое наложением частот, поскольку $\omega_M > \pi / L$.

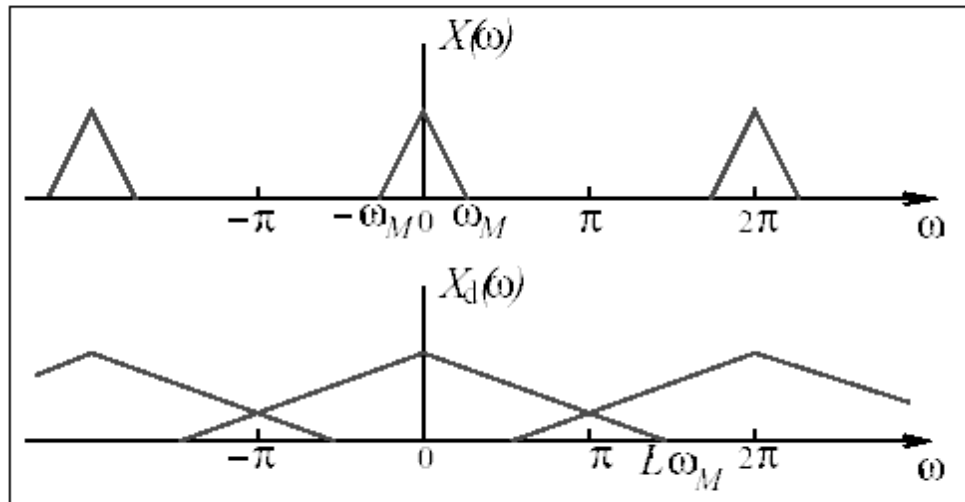
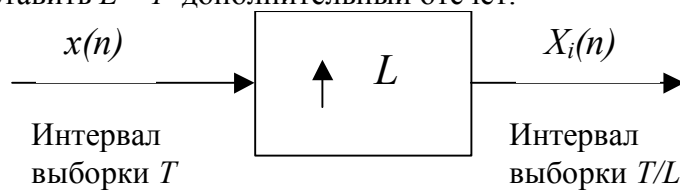


Рис. 7.8

Увеличение частоты дискретизации в целое число раз

$$x_i = x(nT'), T' = T/L$$

Для выполнения операции увеличения частоты выборки в L раз необходимо произвести интерполяцию: между соседними отсчетами исходной последовательности $x(n)$ вставить $L - 1$ дополнительных отсчетов.



Для выполнения операции используется интерполяционная формула

$$\begin{aligned} x_i(n) &= \sum_{k=-\infty}^{\infty} x(k) \frac{\sin[\pi(n-kL)/L]}{\pi(n-kL)/L} = \\ &= \sum_{m=-\infty}^{\infty} x_p(m) \frac{\sin[\pi(n-m)/L]}{\pi(n-m)/L}, \end{aligned}$$

где

$$\{x_p(m)\} = \{\dots, x(-1), 0, \dots, 0, x(0), 0, \dots, 0, x(1), \dots\}.$$

Т.е. повышение частоты дискретизации в L раз может быть выполнено путем добавления между отсчетами сигнала $L-1$ нулей с последующей фильтрацией полученной последовательности низкочастотным фильтром с частотой среза π/L :

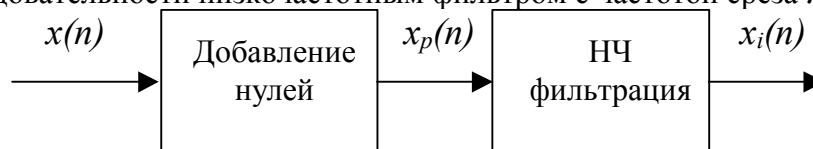


Иллюстрация во временной и частотной областях

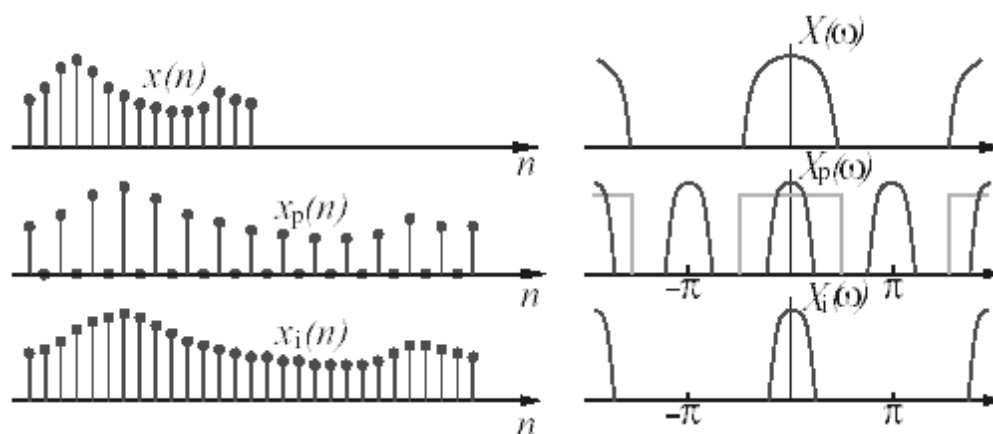
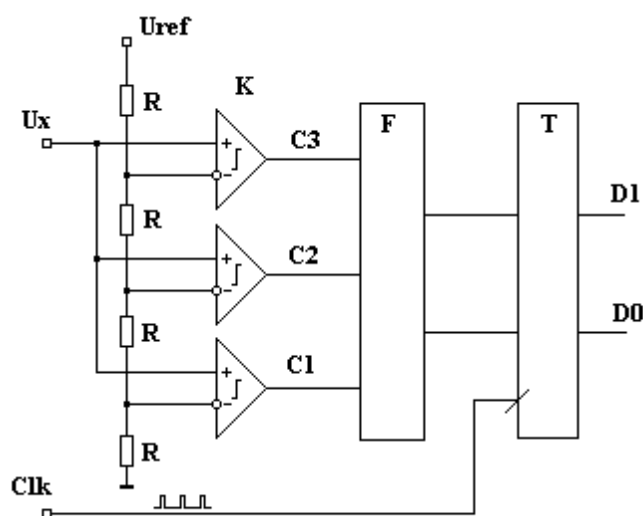


Рис. 7.9

7.4. Аналого-цифровые преобразователи

Самый быстрый АЦП – параллельный (*Flash, Direct conversion*). Структурная схема такого преобразователя представлена на рис.7.10. Параллельный АЦП проводит непосредственное сравнение входного напряжения с напряжениями, сформированными цепочкой резисторов с одинаковыми номиналами. Чем входное напряжение больше, тем у большего числа компараторов выход находится в состоянии логической единицы. На выходе линейки компараторов **К** формируется код «с заполнением единицами». Шифратор (логическая схема) **F** преобразует этот код в прямой двоичный код. Регистр **T** фиксирует полученный код в моменты прихода импульсов синхронизации **Clk**. Таким образом, частота повторения импульсов **Clk** является частотой выборки.



	$0 < U_x < U_{ref}/4$	$U_{ref}/4 < U_x < U_{ref}/2$	$U_{ref}/2 < U_x < 3U_{ref}/4$	$U_{ref}/4 < U_x$
C1	0	1	1	1
C2	0	0	1	1
C3	0	0	0	1
D0	0	1	0	1
D1	0	0	1	1

Рис.7.10

Для реализации главного преимущества - высокого быстродействия – необходимо использовать быстродействующие компараторы, рассеивающие значительную мощность. В N-разрядном АЦП должно быть $2^N - 1$ компараторов. Компаратор – аналоговое устройство, достаточно сложное. Большая рассеиваемая мощность ограничивает разрядность параллельных АЦП шестью – восемью битами. Максимальная частота выборки определяется быстродействием компараторов и составляет в настоящее время сотни МГц.

Наиболее популярный и широко распространенный тип АЦП – АЦП последовательного приближения, принцип действия которого сходен с решением алгебраического уравнения методом деления отрезка пополам. Структурная схема АЦП последовательного приближения и диаграмма, поясняющая его работу представлены на рис.7.11.

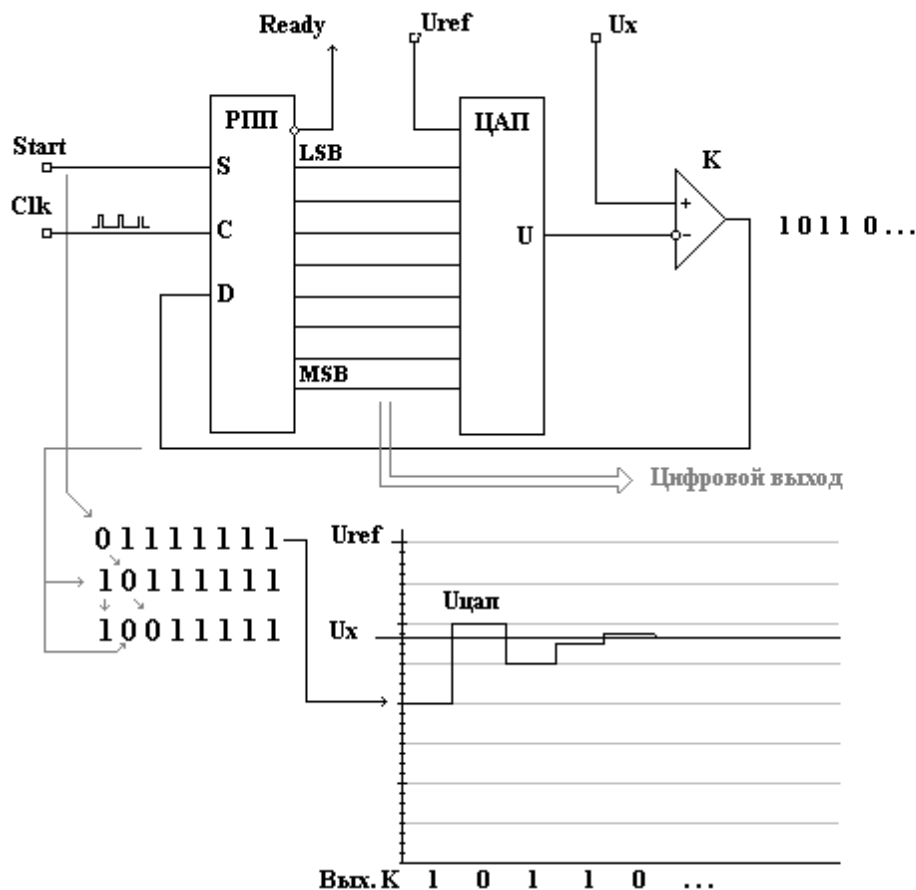


Рис.7.11

Полное N-разрядное преобразование занимает N+1 такт. Длительность такта определяется временем установления выходного напряжения ЦАП и быстродействием компаратора.

Как видно из приведенного рисунка, в течение всех тактов преобразования входной сигнал должен оставаться постоянным. Для обеспечения этого применяются специальные схемы - устройства выборки-хранения (УВХ, SHA – Sample-Hold Amplifier, ТНА - Track – Hold Amplifier). Устройство и работа этих приборов представлены на следующих рисунках. Часто схема УВХ выполняется на том же кристалле, что схема АЦП. В этом случае АЦП называют аналого-цифровым преобразователем с выборкой (sampling ADC), а параметром, характеризующим быстродействие, становится не время преобразования, а максимальная частота выборки. Минимальный период выборки равен сумме времени выборки и времени преобразования.

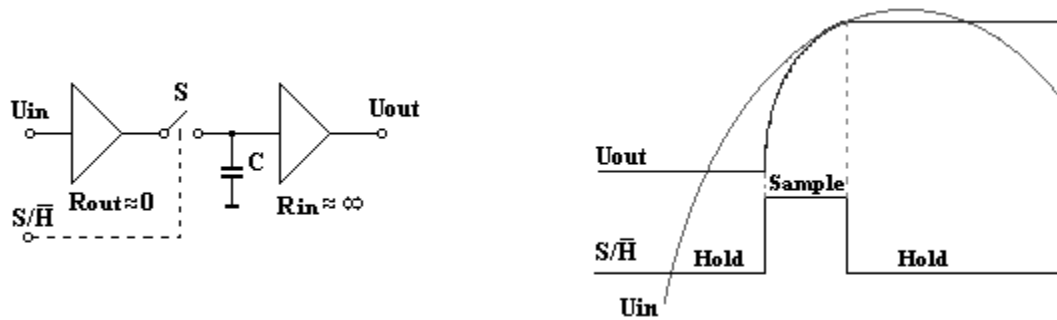
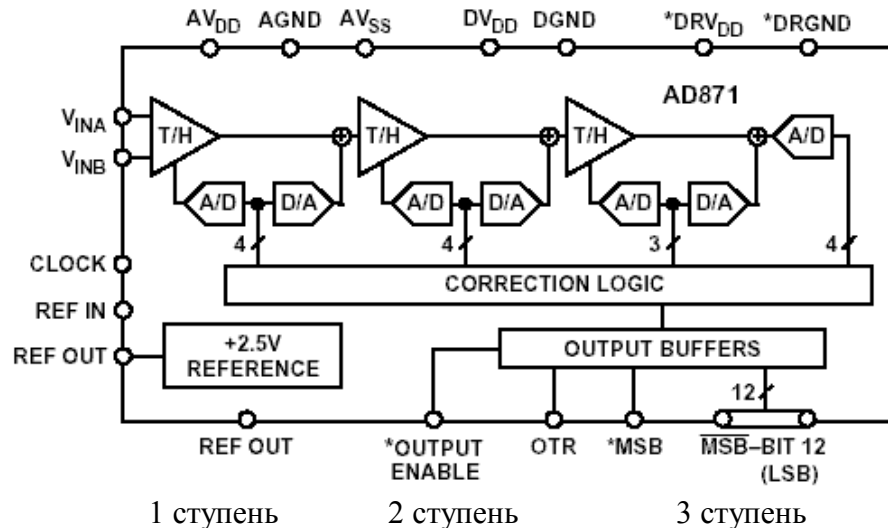


Рис.7.12

Максимальная частота выборки АЦП последовательного приближения зависит от разрядности преобразователя и составляет в настоящее время сотни кГц для 12-ти разрядных АЦП. Время преобразования для АЦП без УВХ – единицы микросекунд.

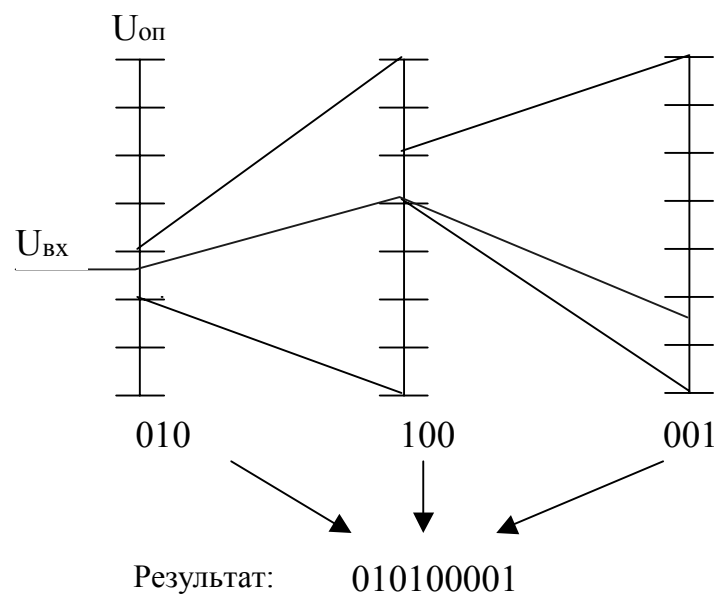
Заметное увеличение быстродействия с сохранением приемлемого потребления мощности обеспечивают *многоступенчатые АЦП (Subranging, +Pipe-lined ADC)*. Одна из возможных схем такого АЦП приведена ниже. Здесь же представлен рисунок, поясняющий принцип действия преобразователя (рис.7.13).



1 ступень

2 ступень

3 ступень



Результат: 010100001

Рис.7.13

Параметры, характеризующие аналого-цифровой преобразователь:

Время преобразования/Максимальная частота выборки – *Conversion time/Conversion rate. [μs/ksps].*

Дифференциальная нелинейность – *Differential Nonlinearity (DNL) [LSB].*

Интегральная нелинейность – *Integral Nonlinearity (INL) [LSB].*

Искажения (Гармоники) – *Distortion, Total Harmonics (THD) [dB]*

$$THD = 10 \log[(V_2^2 + V_3^2 + V_4^2 + V_5^2 + V_6^2) / V_1^2]$$

Интермодуляционные искажения – *Intermodulation (IMD) [dB]*

$$IMD = 10 \log[V_{Fa+Fb}^2 / V_{Fa}^2] \cdot \{Fa-Fb; 2Fa-Fb; Fa+2Fb; \dots\}$$

Отношение сигнал/шум – *Signal-to-Noise Ratio (SNR) [dB]*

Идеальный $SNR = (6.02N + 1.76) \text{ dB}$. - отсюда можно найти эффективное число разрядов реального АЦП.

Отношение сигнал/шум+искажения – *Signal-to-Noise and Distortion (SNDR) [dB]*

Spurious-Free Dynamic Range (SFDR) [dB]

Типичный вид преобразования Фурье сигнала на выходе АЦП с указанием значений основных параметров представлен на рис.7.14.

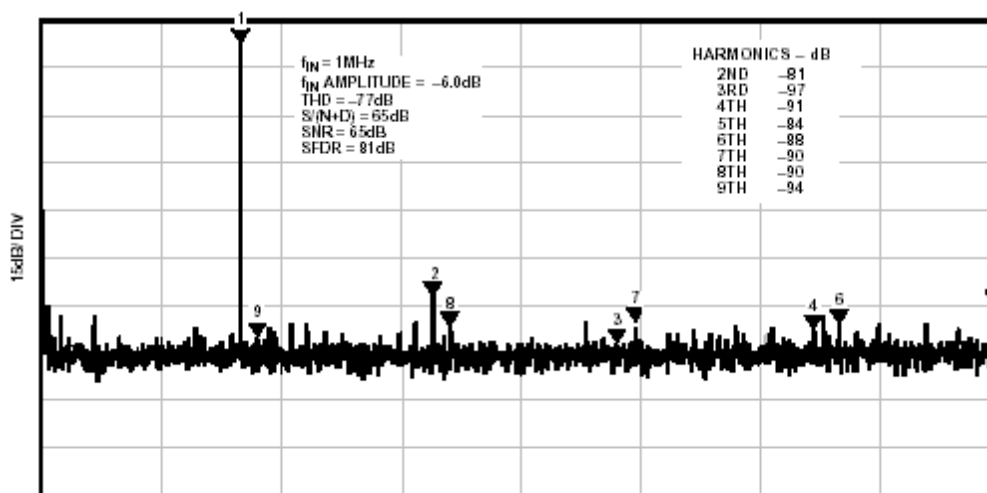


Figure 5. AD872A Typical FFT. $f_{IN} = 1 \text{ MHz}$. f_{IN} Amplitude = -6 dB

Рис.7.14

Для измерений постоянных или медленно меняющихся напряжений разработан специальный тип АЦП, позволяющий избавиться от помех, обусловленных сетевыми наводками, - АЦП с двойным интегрированием (рис.7.15). Интегрирование входного сигнала на интервале времени, кратном периоду помехи (сети переменного тока), позволяет ослабить помеху на десятки децибел.

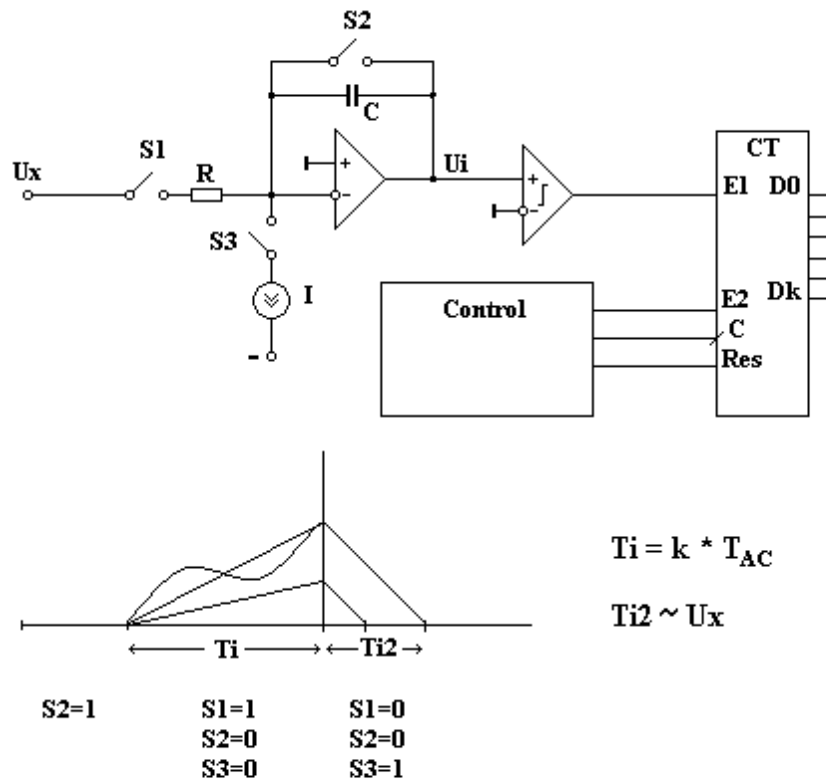


Рис.7.15

При классическом подходе к процессу дискретизации эффективное значение шума квантования, попадающего в полосу от 0 до $F_s/2$, составляет $Q/\sqrt{12}$ (Q — вес младшего разряда, F_s — частота следования выходных отсчетов). Здесь значительная часть шума квантования попадает в рабочую полосу частот. При соблюдении условия теоремы Котельникова (полоса частот полезного сигнала меньше либо равна $F_s/2$) аналоговый фильтр на входе преобразователя должен обладать высокой крутизной спада АЧХ за полосой пропускания, для эффективного ослабления высокочастотных шумов и помех, которые могут проникать в рабочую полосу в результате интерференции с гармониками частоты дискретизации. В подавляющем большинстве случаев это активный ФНЧ. Но добиться удовлетворительного коэффициента гармоник и малых фазовых искажений у таких фильтров — весьма непростая задача. Тут можно столкнуться с глубоким противоречием.

Другой способ улучшения разрешения преобразователя — *передискретизация*. При этом входной сигнал квантуется с частотой $K \cdot F_s$ (K — отношение передискретизации), а выходной цифровой поток следует уже с темпом F_s (рис.7.16).

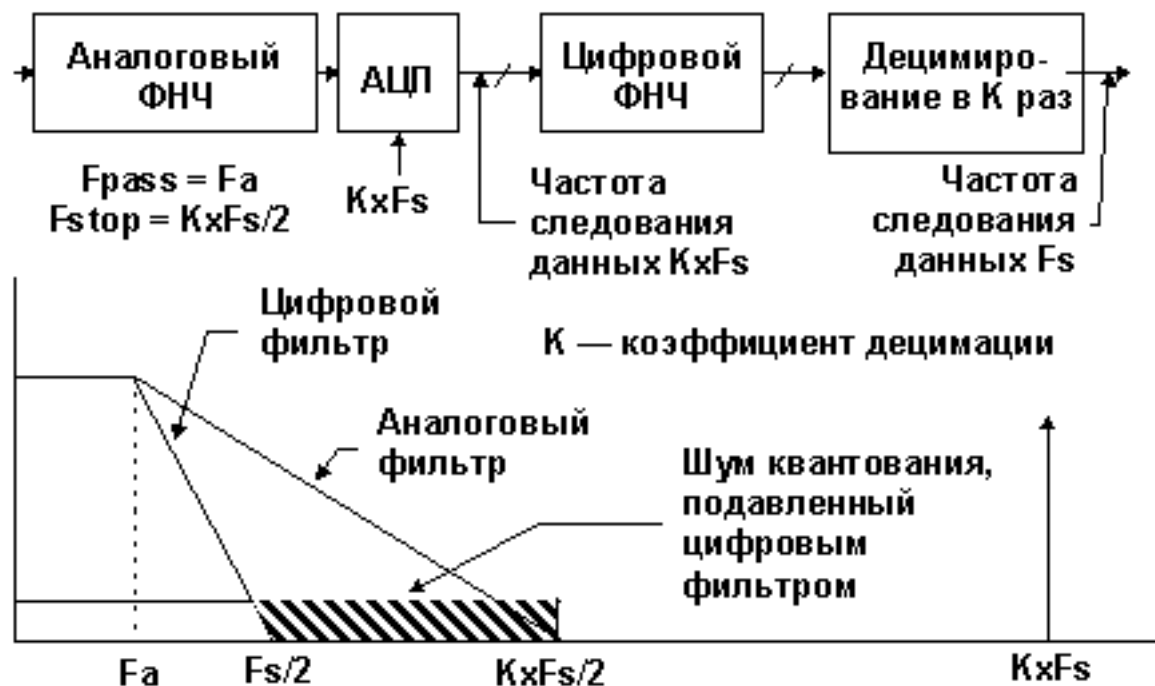


Рис.7.16

Мы можем видеть два новых элемента схемы: цифровой фильтр и дециматор (устройство снижения темпа следования отсчетов). Шум квантования в полосе частот от F_s до $(K \cdot F_s/2)$ подавляется цифровым фильтром в выходном потоке. Это приводит к улучшению отношения сигнал-шум на величину, равную $10 \cdot \log_{10}(K)$. Кроме того, можно добиться малой неравномерности АЧХ и ФЧХ цифрового фильтра и высокой линейности. Сам же аналоговый фильтр вырождается в простое RC-звено. К сожалению, цена за сверхразрешение высока, потому что для улучшения отношения сигнал-шум на 6 дБ (1 бит) требуется соответственно увеличить коэффициент передискретизации в 4 раза. Для сохранения значения этого коэффициента в разумных пределах можно формировать спектр шума квантования так, чтобы основной шум был между $F_s/2$ и $K \cdot F_s/2$ и только малая часть в области между 0 и $F_s/2$. Эту функцию выполняет *сигма-дельта модулятор*. После такого формирования цифровой фильтр легко подавит значительную часть энергии шума квантования, и общее отношение сигнал-шум, определяющее динамический диапазон, ощутимо возрастет.

Блок-схема сигма-дельта модулятора первого порядка представлена на рисунке 7.17. Сигма-дельта модулятор преобразует входной сигнал в последовательный непрерывный поток нулей и единиц, следующих с темпом $K \cdot F_s$.

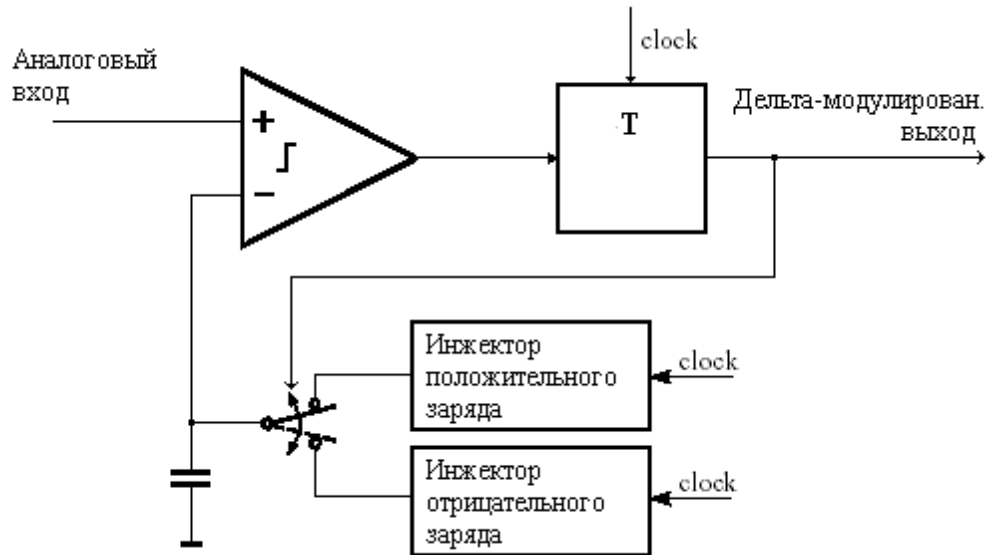


Рис.7.17

На рис.7.18 показаны диаграммы, иллюстрирующие работу сигма-дельта модулятора.

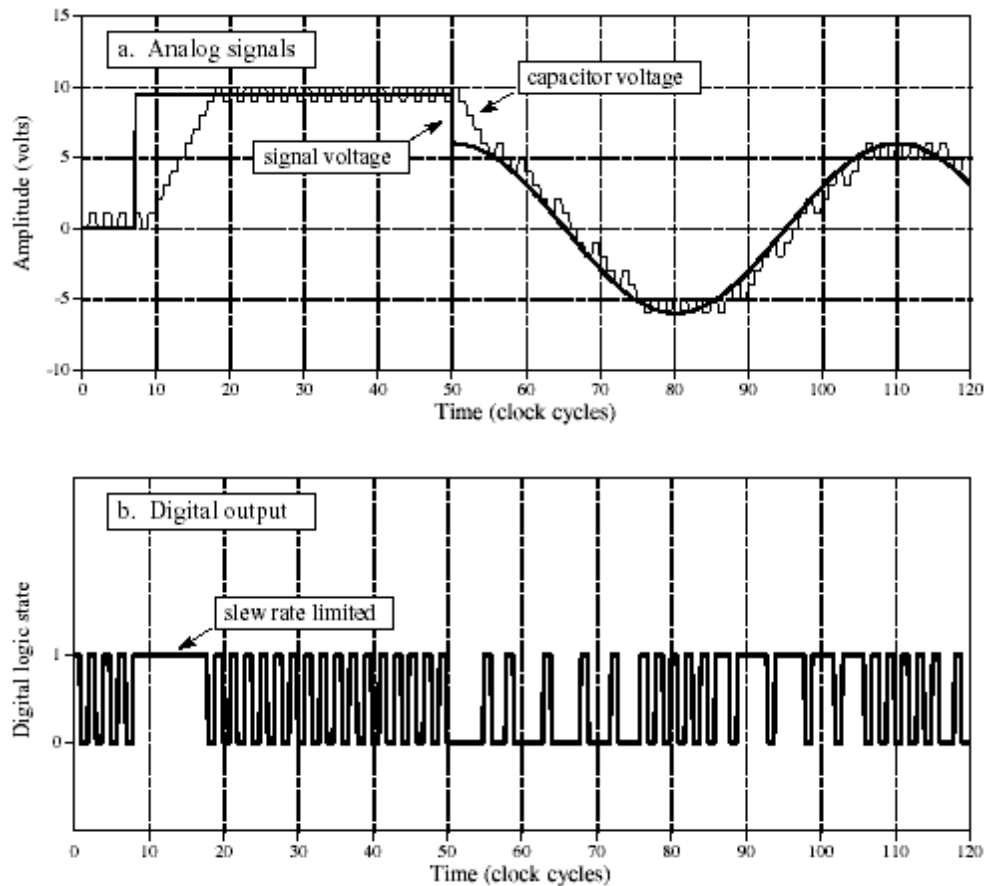


Рис.7.18

На рис.7.17 мы видим замкнутую цепь обратной связи: вычитающее устройство, интегратор, компаратор (1-битовый АЦП), 1-битовый ЦАП. Этот ЦАП принимает последовательный поток данных, а выходное воздействие ЦАП вычитается из входного сигнала. Из теории обратной связи следует, что средняя величина напряжения на выходе ЦАП может достигать того же значения, что и на входе модулятора, при достаточном

петлевого усилении. Интегратор может быть представлен в рабочем диапазоне частот как фильтр, амплитуда отклика которого пропорциональна $1/f$, где f — частота входного воздействия. Компаратор синхронизируется тактовыми импульсами, следующими с частотой ($K \times F_s$), преобразуя медленный входной сигнал в сигнал переменного тока высокой частоты, которая меняется в зависимости от среднего значения напряжения на входе. Таким образом, эффективное значение шума квантования на низких частотах пренебрежимо мало, а интегратор выступает в роли "фильтра высоких частот" для шума квантования. Распределение спектра результирующего шума сильно зависит от скорости квантования, постоянной времени интегратора и точности балансировки обратной связи по напряжению.

На рисунках 7.19, 7.20 представлена блок-схема сигма-дельта АЦП, имеющего на входе сигма-дельта модулятор с несколько модифицированной схемой.

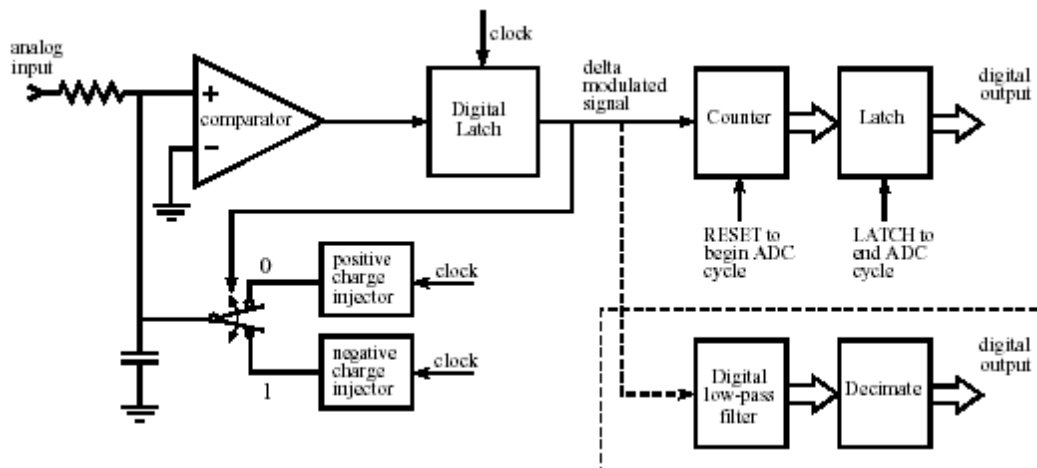


Рис.7.19

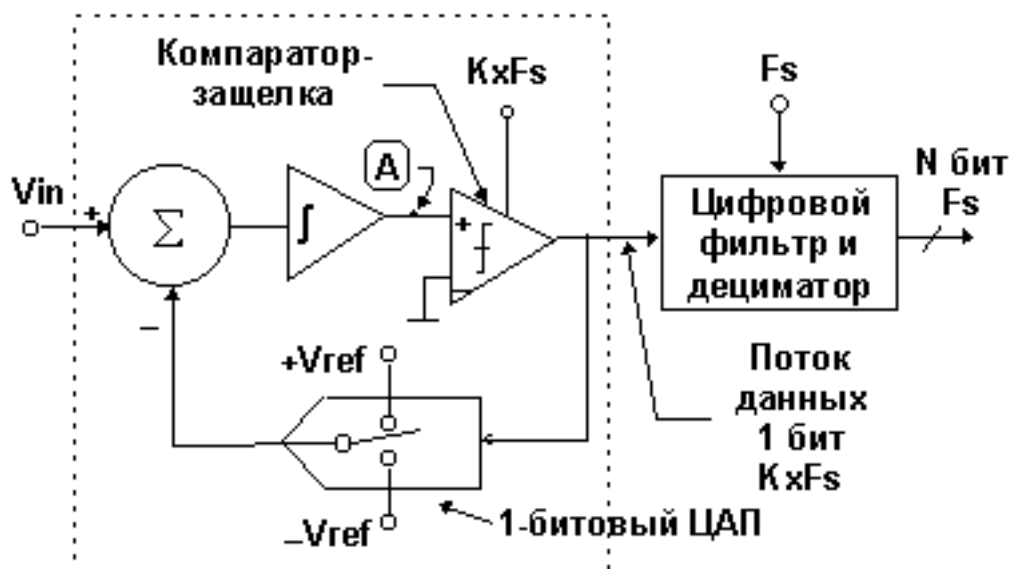


Рис.7.20

Дальнейший анализ сигма-дельта АЦП лучше всего производить в частотном аспекте, используя линейную модель. Отметим, что здесь интегратор представлен как аналоговый фильтр с заданной передаточной характеристикой $H(f)$. Эта характеристика имеет амплитудную зависимость, которая обратно пропорциональна частоте. Квантователь на рис.7.21 показан как каскад усиления, предшествующий сумматору шума квантования.

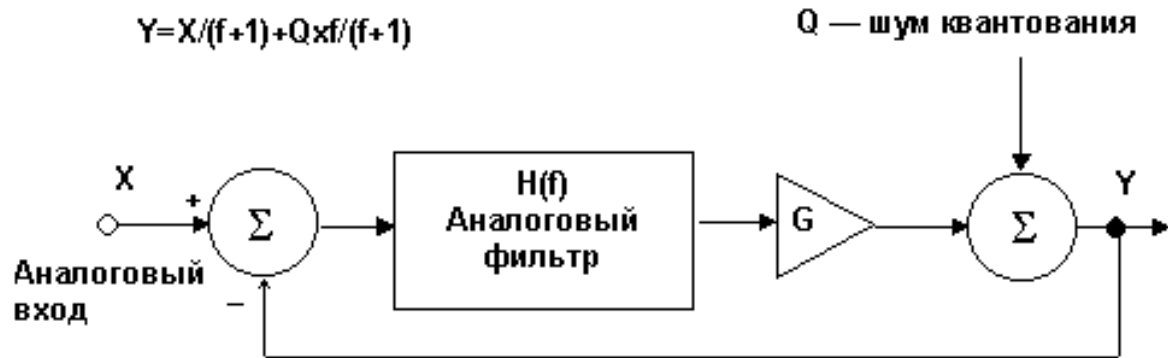


Рис.7.21

Одним из преимуществ частотного подхода является то, что для описания поведения сигналов можно пользоваться просто алгеброй. Выходная величина Y может быть представлена как разность $(X-Y)$, умноженная на передаточную функцию аналогового фильтра, умноженная на коэффициент передачи усиливающего звена и затем сложенная с шумом квантования Q . Если положить коэффициент передачи, равным 1, а передаточную функцию представить как $1/f$, то в результате математических преобразований получим:

$$Y = (X-Y)/f + Q = X/(f+1) + Qf/(f+1)$$

Отсюда следует, что на частоте f , равной 0, Y будет равен только X . При увеличении частоты величина X уменьшается, а величина шумовой компоненты увеличивается. Так как аналоговый фильтр действует как ФНЧ на сигнал и как ФВЧ на шум квантования, такие модуляторы с фильтрами часто называют шумообразующими. Это иллюстрирует следующий рисунок 7.22.

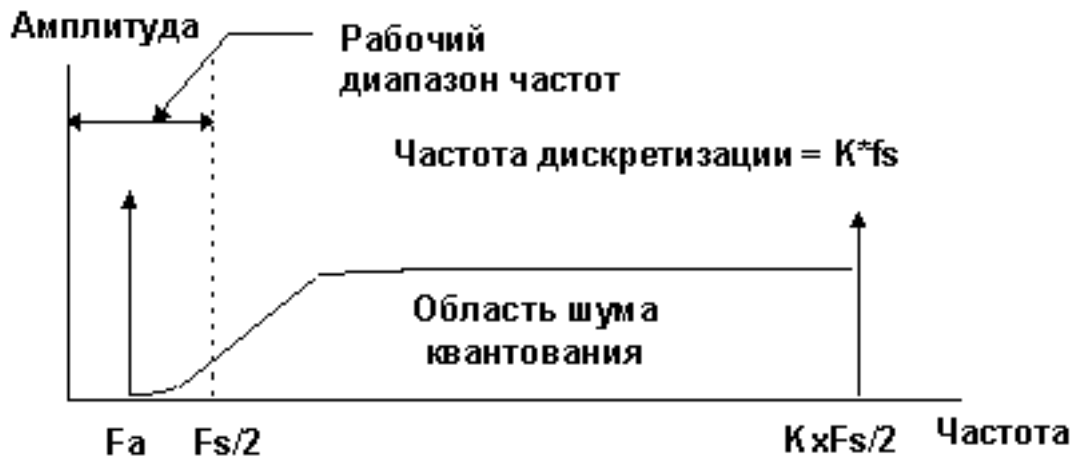


Рис.7.22

После того как шум квантования был сформирован квантователем в полосе частот выше рабочего диапазона, необходимо подавить продукты этого шума с помощью цифровой фильтрации (рис.7.23).

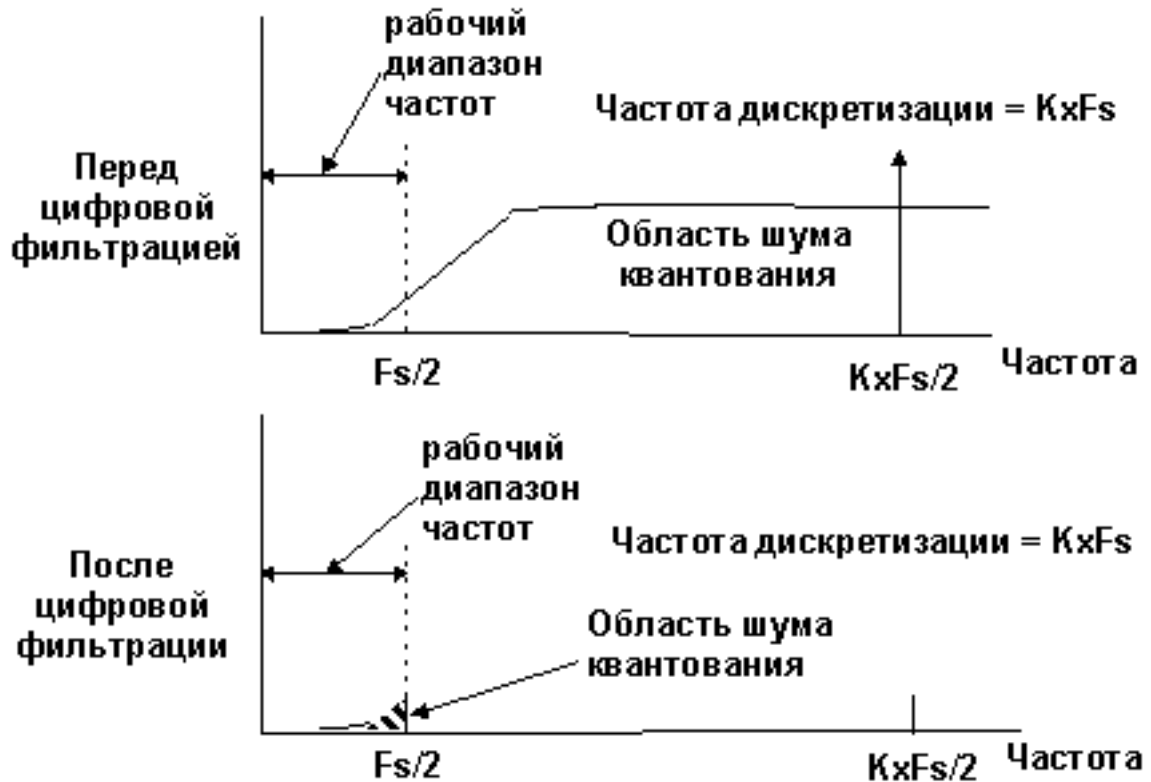


Рис.7.23

На рисунке 7.24 представлена реальная частотная характеристика одного из сигма-дельта АЦП, выпускаемых фирмой Analog Devices.

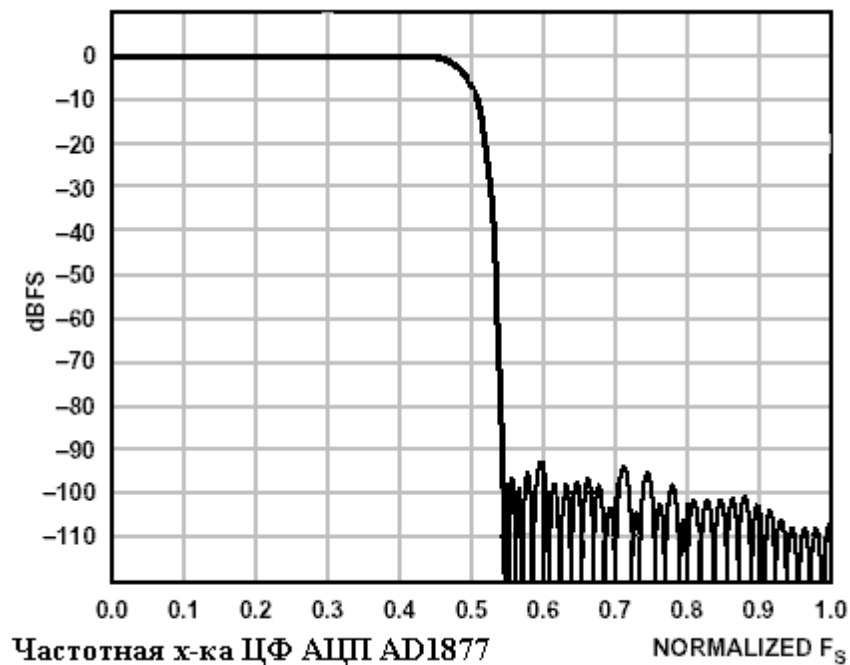


Рис.7.24

При использовании сигма-дельта АЦП не следует забывать о большой групповой задержке сигнала, обусловленной цифровой фильтрацией.

7.5 Цифро-аналоговые преобразователи

Теоретически самый простой метод восстановления непрерывного аналогового

сигнала – преобразование последовательности чисел, представляющей цифровой сигнал, в последовательность импульсов соответствующей амплитуды с последующей фильтрацией идеальным фильтром низких частот. При этом цифровой сигнал, последовательность импульсов и восстановленный непрерывный сигнал имеют идентичные спектры на частотах ниже частоты Найквиста (половины частоты дискретизации). На высших частотах, последовательность импульсов содержит периодические частотные компоненты, такие же, как у цифрового сигнала, и которых нет у непрерывного колебания.

В реальных устройствах почти все ЦАП удерживают последнюю величину выходного напряжения до момента прихода следующей выборки. Значение выходного напряжения как бы интерполируется постоянной величиной на интервал времени между выборками. Интерполяция постоянной называется интерполяцией нулевого порядка и приводит к формированию выходного сигнала в виде лестницы (рис.7.25). Интерполяция нулевого порядка математически может быть описана сверткой исходного импульсного сигнала с импульсной характеристикой интерполятора, представляющей собой прямоугольный импульс длительностью, равной периоду дискретизации.

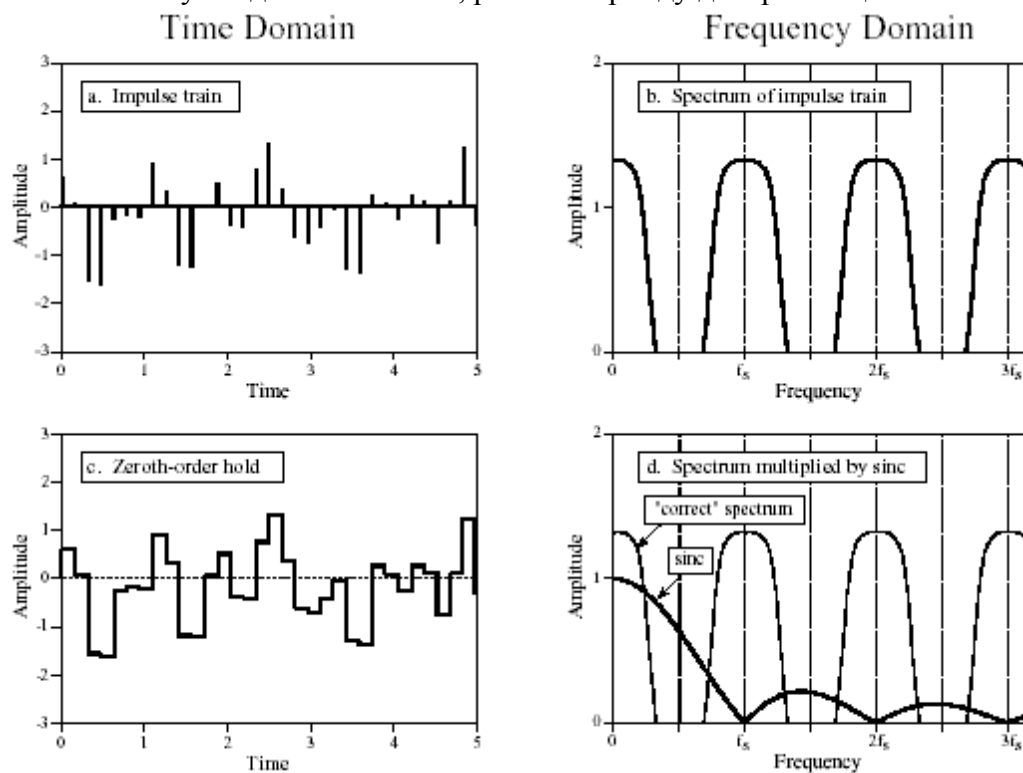


Рис.7.25

В частотной области интерполяция приводит к взвешиванию периодизированного спектра сигнала функцией $\text{sinc}(x)$, являющейся Фурье-преобразованием импульсной характеристики интерполятора.

Таким образом, выходной аналоговый фильтр, используемый для восстановления аналогового сигнала, должен делать две вещи: удалять все частоты выше половины частоты дискретизации, и компенсировать частотные искажения, обусловленные интерполяцией, т.е. в полосе частот от 0 до частоты Найквиста иметь частотную характеристику $1/\text{sinc}(x)$ (рис.7.26).

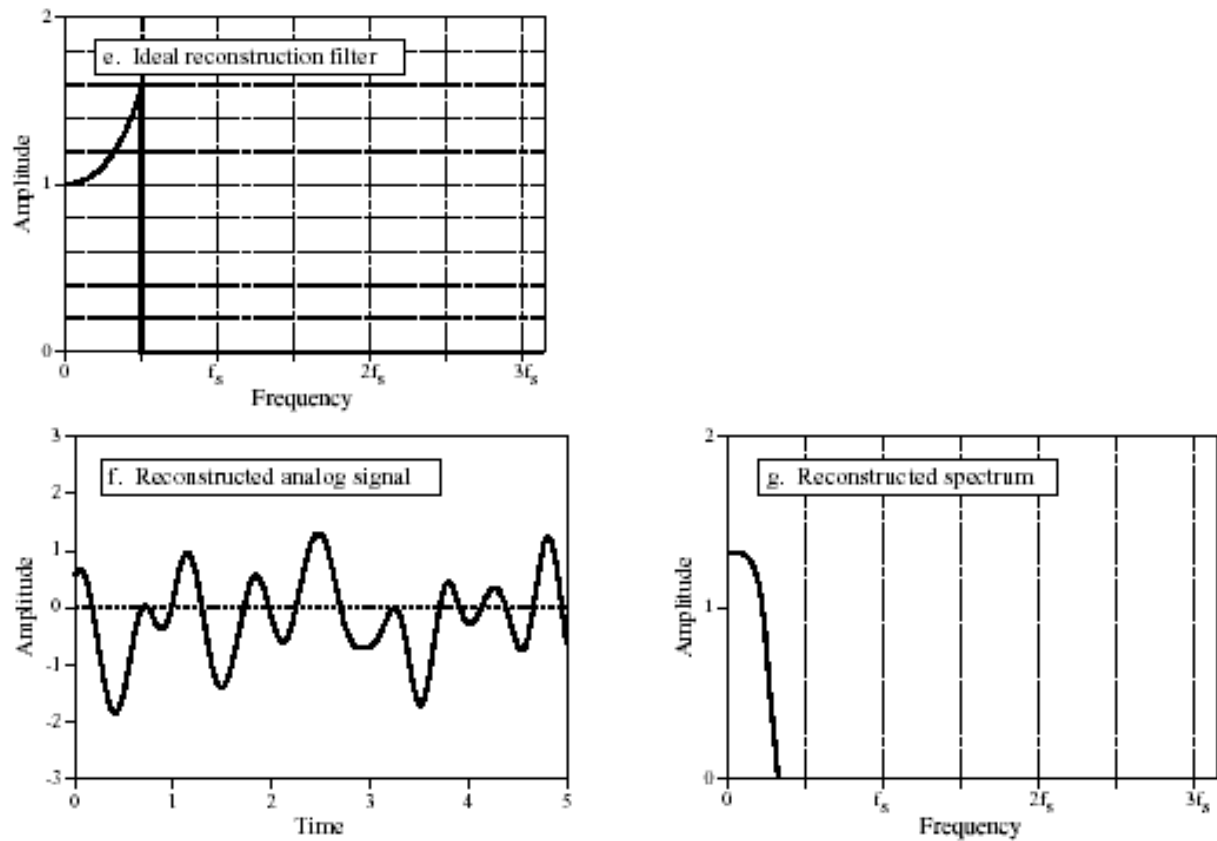


Рис.7.26

Проще эту восстанавливающую характеристику заложить в цифровой алгоритм обработки сигнала.

Теперь кратко об устройстве цифро-аналоговых преобразователей.

Простейший ЦАП можно построить по схеме суммирования токов (рис.7.27).

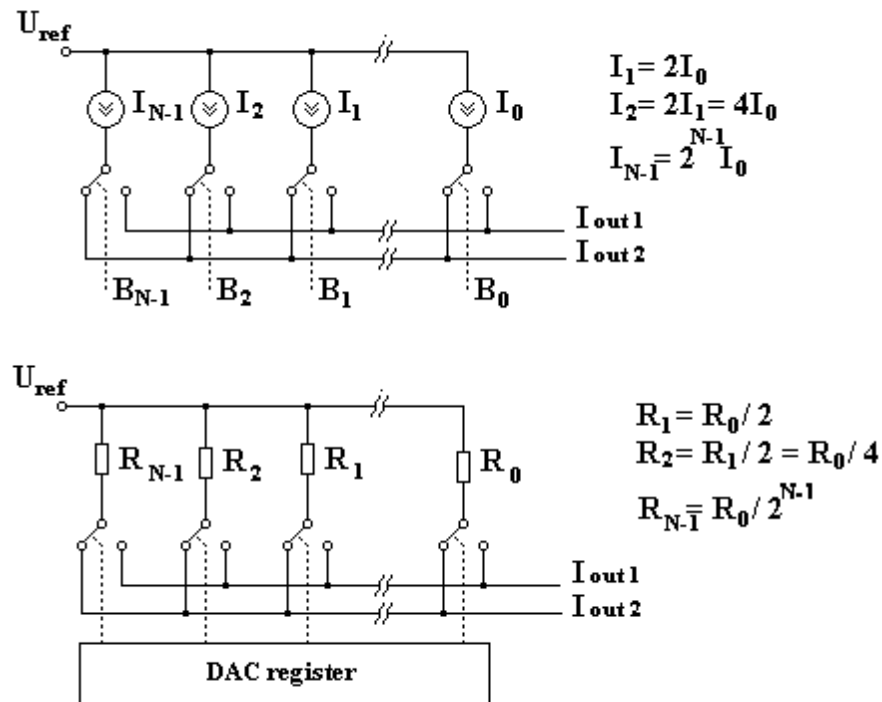


Рис.7.27

Схема проста и понятна, однако реализации подобного ЦАП большой разрядности препятствует широкий диапазон требуемых значений номиналов резисторов, поскольку отношение $R_{\max}/R_{\min} = 2^{N-1}$.

Абсолютное большинство выпускаемых ныне ЦАП реализовано по схеме матрицы $R - 2R$ (рис.7.28).

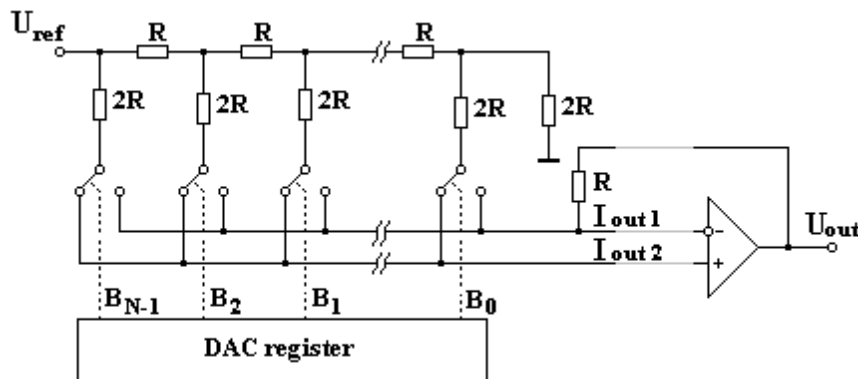


Рис.7.28

При реализации ключей ЦАП по комплементарной МОП технологии, допускающей широкий диапазон переключаемых токов и изменение полярности приложенных напряжений, в качестве U_{ref} может быть использован сигнал переменного тока. Такой ЦАП называется перемножающим. Он может быть использован для построения программируемых аттенуаторов и усилителей с переменным коэффициентом усиления.

В традиционных архитектурах ЦАП на основе матрицы сопротивлений $R/2R$ цифровая природа внутренних ключей порождает переходные процессы - *глитчи*, зависящие от кода. В результате в выходном спектре появляются паразитные гармонические составляющие. Для достижения чистоты спектра основную энергию глитчей можно уменьшить с помощью схемы выборки-хранения, которая фиксирует напряжение на выходе ЦАП в области, свободной от переходных процессов (рис.7.29).



Рис.7.29

Использование передискретизации и переход к однобитовому преобразованию позволяют исключить глитчи и кардинально упростить фильтрацию выходного сигнала ЦАП. Сигма-дельта цифро-аналоговое преобразование можно рассматривать как аналогово-цифровое, только в обратном порядке, где все основные функции цифровых фильтров и сигма-дельта модуляторов остаются неизменными. Сигма-дельта ЦАП проявляют те же превосходные свойства, что и сигма-дельта АЦП. Благодаря высокому коэффициенту передискретизации фильтрация высокочастотных компонент спектра сигнала на выходе ЦАП может быть достигнута простыми средствами.

На рисунке 7.30 представлены основные элементы сигма-дельта ЦАП

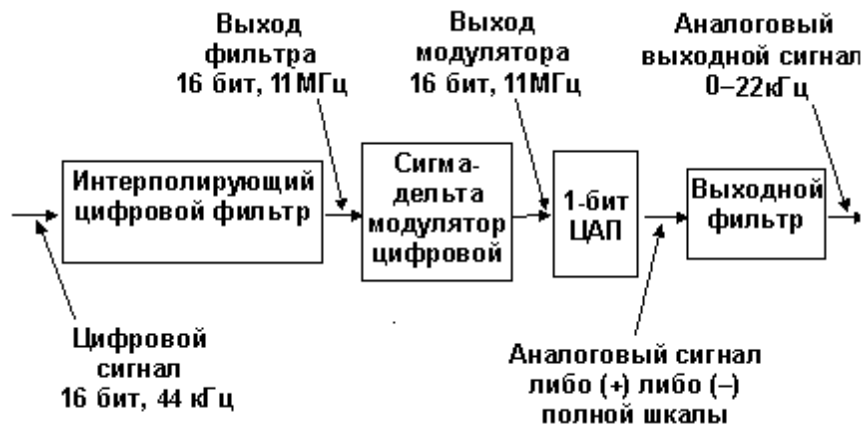


Рис.7.30

8. Цифровые процессоры для обработки сигналов

8.1. Особенности архитектуры сигнальных процессоров

ЦОС производится посредством выполнения математических операций, в то время как стандартная задача, выполняемая текстовыми процессорами, например, - простая перестановка хранимых данных. Несомненно, универсальные компьютеры могут производить вычисления, но их архитектура и система команд не очень подходит для решения задач цифровой обработки сигналов. Далее мы увидим, что это действительно так. *Сигнальные процессоры (ЦПС)* – микропроцессоры, специально предназначенные для решения задач ЦОС. В последние десятилетия наблюдается потрясающее развитие как самих ЦПС, так и расширение областей их применения: от сложнейших измерительных приборов, научного и военного оборудования до сотовых телефонов. Далее мы рассмотрим особенности архитектуры ЦПС, отличающие их от стандартных микропроцессоров и микроконтроллеров, и познакомимся с некоторыми примерами применения ЦПС.

В 60-е годы прошлого века появилось мнение, что искусственный интеллект революционизирует общение человека с компьютерами и другими машинами. Многие верили, что к концу века роботы будут делать домашнюю работу, управлять автомобилем и т.п. Этого не случилось, поскольку эти задачи оказались более сложными, чем ожидалось, и очень трудными для «пошаговой» логики компьютеров. Тем не менее, последние сорок лет показали, что компьютеры чрезвычайно способны в двух широких областях: обработка данных (1), например, текстовая обработка и управление базами данных, и (2) математические вычисления, используемые в науке, проектировании, и Цифровой Обработке Сигналов. Все микропроцессоры могут выполнить оба задания; тем не менее, трудно (дорого) сделать устройство, которое оптимизировано для обеих задач. Приходится идти на компромиссы в аппаратной разработке, касающиеся, например, длины команд и обработки прерываний. Еще более важно то, что в процессе разработки учитываются маркетинговые вопросы: стоимость разработки и конечного продукта, конкурентоспособность, срок службы продукта, и так далее. Так разработаны традиционные микропроцессоры, например, Pentium, «выросший» из микропроцессора 8080, первоначально предназначенного для обработки данных. Цифровые сигнальные процессоры (DSPs) оптимизированы для выполнения математических вычислений и иных операций, нужных для цифровой обработки сигналов.

	Обработка данных	Математические вычисления
Типичные приложения	Текстовые процессоры, Управление базами данных, Операционные системы	ЦОС, Научные и инженерные вычисления, Моделирование
Выполняемые операции	Перенос данных (A->B), Проверка условий (if A=B then ...)	Сложение (A+B=C) Умножение (A*B=C)

Приведенная таблица иллюстрирует наиболее важные различия между этими двумя категориями. Обработка данных включает хранение и сортировку информации.

Например, рассмотрим программу обработки текстов. Основное задание - загрузить информацию (введенную оператором), организовать информацию (копировать, вставлять, переносить текст, проверить правописание, сформатировать страницу и т.п.), и затем сохранить информацию на диске или на бумаге посредством печати. Эти задания выполняются перемещением данных из одной позиции на другое, и тестированием условий (проверкой неравенств $A=B$, $A<B$, и т.п.). Как пример, представьте себе сортировку списка слов в алфавитный порядок. Каждое слово представлено 8 битовым числом, величиной ASCII-кода первой буквы в слове. Упорядочение по алфавиту означает перестановку слов по порядку возрастания ASCII-кодов. После завершения анализа первых букв программа переходит к аналогичному анализу вторых букв в словах и т.п.

Математические вычисления мало используются в этом типе приложений, и значительно не влияют на общую скорость выполнения заданий.

В отличие от вышесказанного, скорость выполнения большинства алгоритмов DSP почти полностью определяется количеством требуемых умножений и сложений. Например, следующий рисунок 8.1 показывает реализацию цифрового КИХ-фильтра, наиболее общую задачу DSP. Реализуемое разностное уравнение имеет вид:

$$y[n] = a_0 x[n] + a_1 x[n-1] + a_2 x[n-2] + a_3 x[n-3] + a_4 x[n-4].$$

Это уравнение просто говорит, что входной сигнал свернут с импульсной характеристикой фильтра, представляющей собой последовательность a_0, a_1, a_2 и т.д. В зависимости от требований к фильтру коэффициентов может быть от двух до многих тысяч. Хотя здесь есть некоторая пересылка данных и оценка неравенства, математические операции доминируют в определении времени выполнения.

Дополнительно к очень быстрому выполнению математических вычислений, ЦПС должны также иметь предсказуемое время выполнения. Предположим, что Вы запустите ваш настольный компьютер на некоторое задание, скажем, форматирования текстового документа. Не имеет значения, займет обработка десять миллисекунд или одну секунду. Вы просто ждете завершения действия прежде чем дать компьютеру следующее задание.

Большинство же ЦПС используются в приложениях, где обработка непрерывная, не имеющая определенного начала или конца. Система ЦОС должна обеспечить ввод и обработку заданного числа выборок в секунду (скажем около 10000 для обработки речи в слуховом аппарате). Причем есть важные причины не делать обработку быстрее чем необходимо.

По увеличению скорости увеличивается также потребляемая мощность, стоимость разработки, производства и цена процессора. Это делает знание точного времени выполнения операций необходимым для выбора соответствующего устройства, а также алгоритмов, которые могут быть выполнены.

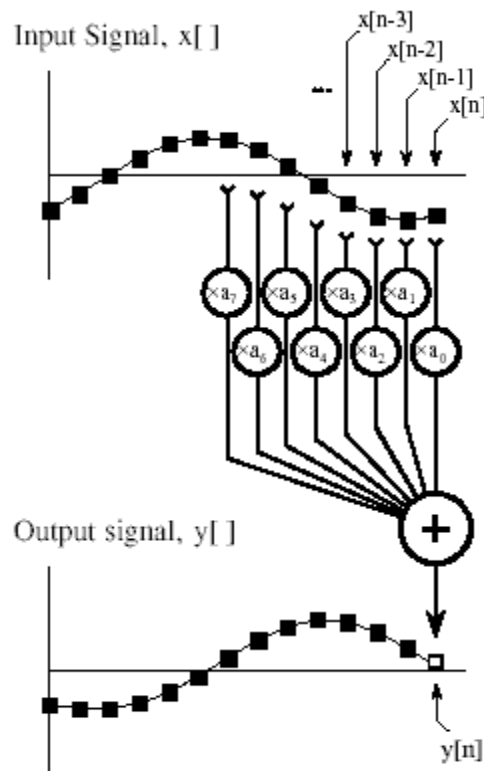


Рис.8.1

Одно из самых узких мест в выполнении DSP алгоритмов - передача информации из памяти. Это информация включает данные, например, выборки входного сигнала и коэффициенты фильтра, а также программные инструкции. Предположим, что нам нужно умножить два числа, которые находятся где-нибудь в памяти. Для того, чтобы делать это, мы должны выбрать три двоичных величины из памяти, числа, которые должны умножаться, плюс программная инструкция, описывающая, что нужно делать.

На рисунке 8.2 показана архитектура традиционного микропроцессора. Ее часто называют Фон Неймановской архитектурой.

a. Von Neumann Architecture (*single memory*)

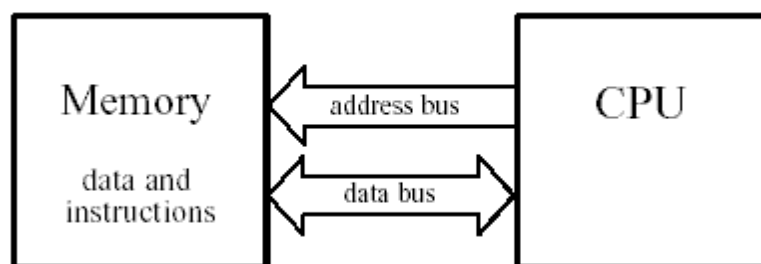


Рис.8.2

Как показано на рис.8.3, Фон Неймановская архитектура содержит единственную память и единственную шину для передачи данных в и из центрального устройства обработки (CPU). Умножение двух чисел требует по крайней мере три машинных цикла, один на выборку команды и два для передачи данных из памяти в CPU.

Мы не считаем время, требуемое чтобы передать результат обратно в память, из-за того, что мы допускаем что, это результат произведения остается в CPU для дополнительной обработки (например, суммирования произведений в КИХ-фильтре). Фактически, большинство компьютеров сегодня имеют Фон Неймановскую архитектуру,

что подтверждает ее достаточность для решения задач. Для ускорения решения задач ЦОС нам нужна другая архитектура, за которую придется расплачиваться возрастающей сложностью проектирования и реализации процессора. Это ведет нас к Гарвардской архитектуре, показанной на рис.8.3.

b. Harvard Architecture (*dual memory*)

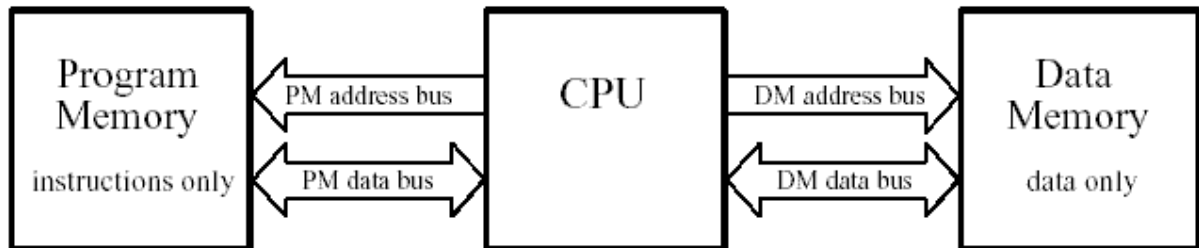


Рис.8.3

Эта архитектура разработана в Гарвардском Университете в 1940-е годы под руководством Howard Aiken (1900-1973). Как показано на этой иллюстрации, Aiken ввел отдельные блоки памяти с отдельными шинами для данных и программных инструкций. Теперь шины действуют независимо, инструкции и данные могут быть выбраны в одно и то же время, что увеличивает скорость обработки по сравнению с архитектурой с единственной шиной. Большинство современных ЦПС используют эту архитектуру с двумя шинами.

Рисунок 8.4 иллюстрирует следующий уровень усложнения: Супер Гарвардскую архитектуру. Этот термин был введен компанией Analog Devices для того, чтобы описать внутреннее устройство их ЦПС семейств ADSP-2106x и ADSP-211xx. Их назвали **SHARC®** DSPs, от Super Harvard Architecture. Из десятка усовершенствований, имеющих место в SHARC DSPs, отметим два важнейших, отраженных на рисунке 8.4: кэш инструкций и контроллер ввода-вывода.

c. Super Harvard Architecture (*dual memory, instruction cache, I/O controller*)

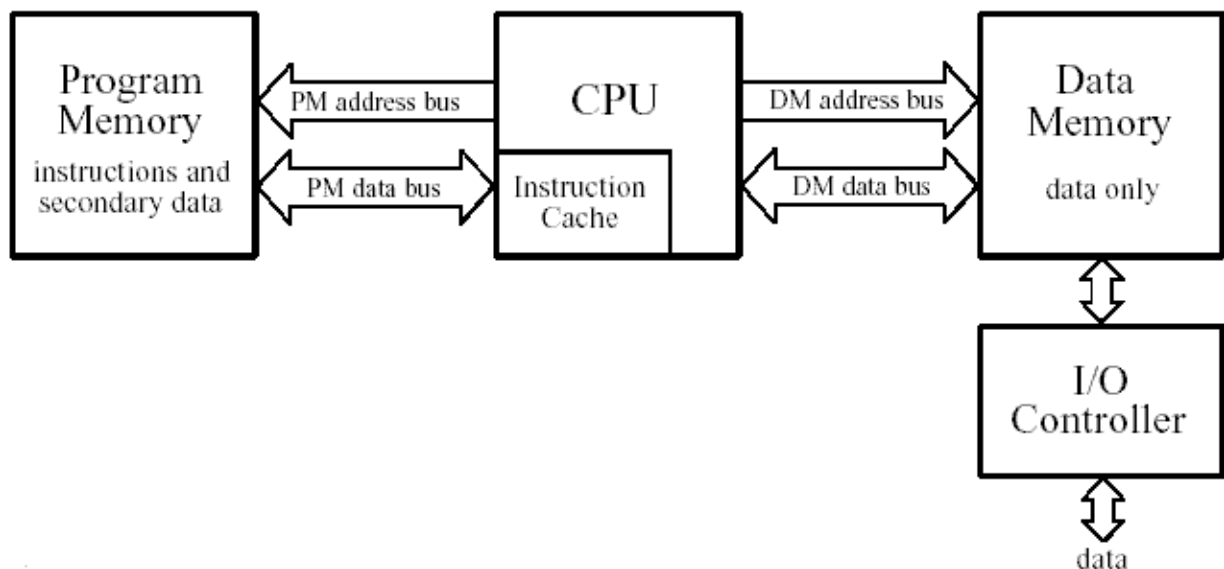


Рис.8.4

В первую очередь посмотрим, как кэш инструкций улучшает характеристики Гарвардской архитектуры. Недостаток основной Гарвардской архитектуры в том, что шина памяти данных более занята, чем шина памяти программ. Когда два числа умножаются, две двоичных величины (числа) должны быть пропущены через шину

памяти данных, в то время как только одна двоичная величина (программная инструкция) должна быть передана через шину памяти программ. Чтобы улучшить эту ситуацию, можно переместить часть "данных" в программную память. Например, мы можем записать в памяти программ коэффициенты цифрового фильтра, сохраняя входной сигнал в памяти данных. (Эти перемещенные данные названы "второстепенные данные" в иллюстрации).

На первый взгляд, это кажется, не помогает ситуации; теперь мы должны передать одну переменную по шине памяти данных (выборку входного сигнала), но две величины по шине программной памяти (программная инструкция и коэффициент). Фактически, при выполнении произвольных инструкций ситуация несколько не улучшается. Однако, алгоритмы DSP обычно тратят большинство времени на выполнение циклов. Это означает, что один и тот же набор программных инструкций периодически пересылается из программной памяти в центральный процессор. Супер-Гарвардская архитектура имеет преимущество в этой ситуации, имея в ЦП кэш инструкций. Это - небольшая память, которая содержит около 32 наиболее последних программных инструкций. При первом выполнении цикла программные инструкции должны быть пропущены через программную шину памяти. В результате этого первый «проход» цикла выполняется медленнее из-за конфликта с коэффициентами, которые должны также выбираться по той же шине. Тем не менее, в дополнительных выполнениях цикла, программные инструкции могут выбираться из кэша инструкций. Это означает, что вся информация, необходимая ЦПУ может передаваться в него в единственном цикле: выборка входного сигнала приходит по шине памяти данных, коэффициент фильтра приходит по программной шине, а программная инструкция - из кэш инструкций.

На следующем рисунке 8.5 более подробно показана SHARC-архитектура. Здесь I/O-контроллер подключен непосредственно к памяти данных в силу того, что ввод/вывод может осуществляться с помощью ПДП - прямого доступа к памяти. SHARC ЦПС обеспечивает как последовательные так и параллельные чрезвычайно быстрые порты связи. Например, при 40 МГц тактовой частоте процессора два последовательных порта работают на скорости до 40 Мбит/секунду каждый, кроме того, каждый из шести параллельных портов обеспечивает передачу 40 Мбайт/секунду.

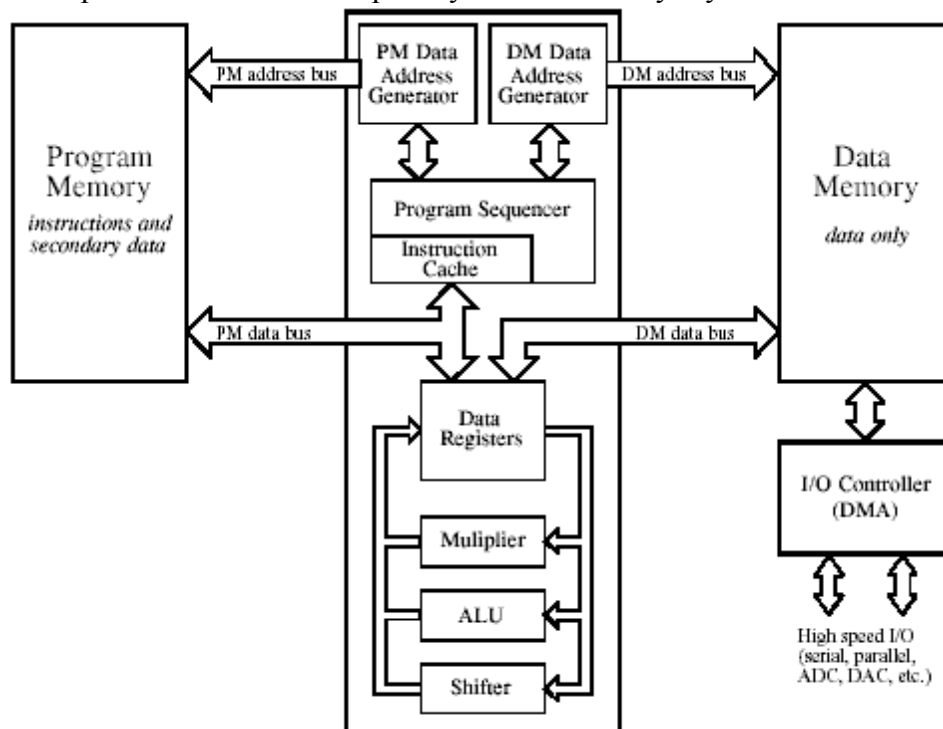


Рис.8.5

Обратите внимание на то, что специальные аппаратные средства управляют потоками данных, направляя их непосредственно в память (Прямой Доступ В Память, или DMA), минуя регистры ЦП. Другими словами, ввод и вывод сигналов могут производиться независимо и одновременно с другими заданиями; никакие циклы ЦП не теряются на ввод-вывод. Основные шины доступны и вне кристалла, предоставляя возможность подключения дополнительных интерфейсов и внекристалльной памяти до 4 Гигаслов (16 Мбайт) со скоростью обмена 160 Мбайт/сек.

Наличие высокоскоростного I/O является ключевой особенностью ЦПС. Некоторые ЦПС имеют на кристалле аналого-цифровые и цифро-аналоговые преобразователи. Такие ЦПС иногда называют процессорами обработки смешанных сигналов (*mixed signal*). Тем не менее, любой ЦПС может связаться с внешними преобразователями через последовательный или параллельный порты.

Теперь давайте посмотрим внимательнее на центральный процессор. На верху диаграммы - два блока, помеченных как Генератор Адреса Данных (DAG), для каждой из двух памяти. Они управляют адресами, выставляемыми на адресные шины памяти программ и памяти данных, определяя ячейку, где информация должна быть прочитана или записана. Кроме прочего, генераторы адресов данных обеспечивают аппаратную поддержку *циклических буферов* (в процессорах Analog Devices SHARC – до 8 буферов; для работы каждого нужно хранить и управлять 4-мя переменными), очень удобных для организации цифровых фильтров и асинхронного ввода-вывода (FIFO). DAG обеспечивают также двоично-инверсную адресацию, применяемую в алгоритме FFT.

Регистры данных ЦП используются так же как в традиционных микропроцессорах. В ADSP-2106x SHARC DSPs, например, есть 16 40-битовых универсальных регистров. Они могут содержать временные переменные, подготовить данные для математического процессора, послужить в качестве буфера для пересылки данных, содержать флаги для программного управления, и так далее. Если необходимо, эти регистры могут также использоваться, чтобы управлять циклами и счетчиками; тем не менее, SHARC DSPs имеет дополнительные аппаратные регистры, чтобы выполнить многие эти функции.

Математическая обработка разбивается на три раздела, умножитель, арифметическое логическое устройство (ALU), и циклическое сдвигающее устройство. Умножитель берет данные из двух регистров, умножает их, и устанавливает результат в третий регистр. ALU выполняет сложение, вычитание, вычисляет абсолютное значение, выполняет логические операции (И, ИЛИ, ИСКЛЮЧАЮЩЕЕ ИЛИ, НЕТ), преобразования между форматами чисел с фиксированной и плавающей точкой, и аналогичные функции. Элементарные битовые операции выполняются циклическим сдвигающим устройством, как например, сдвиг вправо/влево, циклический сдвиг. «Мощное» свойство ЦПС состоит в том, что умножитель и ALU могут быть доступны параллельно. В единственном такте, данные из регистров 0-7 могут быть переданы в умножитель, данные от регистров 8-15 могут быть переданы в ALU, и два результата могут быть возвращены в любой из 16 регистров.

Есть также много важных характеристик SHARC-архитектуры, которые не показаны на этой упрощенной иллюстрации. Например, 80 битовый аккумулятор, встроенный в умножитель, не только позволяет осуществлять умножение-накопление (см. алгоритм цифровой фильтрации), но уменьшает вероятность переполнения и уменьшает ошибку округления, возникающую при выполнении математических операций с числами с фиксированной запятой (при «коротком» аккумуляторе нужно сдвигать результат каждого умножения, а не накопленной суммы). Другое интересное свойство ЦПС - использование *теневых* регистров для всех ключевых регистров ЦП. Это двойные регистры, которые могут переключаться с их аналогами в единственном такте. Они используются для быстрого переключения контекста, обеспечивая способность быстро обрабатывать прерывания. Когда прерывание происходит в традиционных

микропроцессорах, все внутренние данные должны быть сохранены прежде, чем прерывание может быть обработано. Это сохранение обычно обеспечивается занесением всех используемых в обработке прерывания регистров в стек поочередно. В отличие от вышесказанного, при прерывании в SHARC-семействе переключение всех регистров на теньевые регистры производится в единственном такте. Когда программа обработки прерываний завершается, регистры также быстро восстанавливаются. Эта свойство позволяет очень быстро обрабатывать прерывание, скажем по готовности новой выборки сигнала.

Теперь посмотрим, сколько операции в пределах цикла цифровой фильтрации может быть выполнено в то же самое время. Из-за своей параллельной природы, SHARC DSP может одновременно выполнить все эти задания. В пределах единственного такта он может выполнить чтение двух величин (выборки сигнала из кольцевого буфера и коэффициента фильтра), их умножение, накопление частичной суммы, коррекцию двух указателей циклических буферов, и, наконец, управление циклом. Дополнительные циклы конечно потребуются на выполнение операций, связанных с началом и завершением цикла; тем не менее, эти задания также выполняются очень эффективно. Если цикл выполнен более чем несколько раз, потери будут незначительными. Например, предположим, что выполняется программа, реализующая КИХ-фильтр со 100 коэффициентами. На нее потребуется около 105 - 110 циклов на выборку (то есть, 100 циклов умножения-накопления плюс потери). Это очень впечатляющее; традиционный микропроцессор требует многие тысячи циклов для этого алгоритма.

Для иллюстрации того, насколько неэффективно решает универсальный процессор задачу цифровой фильтрации, рассмотрим реализацию алгоритма на процессоре i286.

Казалось бы короткая функция, реализующая цифровой фильтр, написанная на языке «C»

```
int filter(int inp,int len,int *w,int *x)
{
    int i;
    long a=0;
    for(i=len-1;i>0;i--) x[i]=x[i-1];
    x[0]=inp;
    for(i=0;i<len;i++) a+=(x[i]*w[i]);
    return a>>15;
},
```

компилируется в длинный перечень ассемблерных команд, причем большинство из них нужно выполнить для обработки КАЖДОЙ выборки сигнала:

```
@filter$qiipit3      proc    near
    enter    4,0
    push     si
    push     di
    mov     di,word ptr [bp+10]
    mov     word ptr [bp-2],0
    mov     word ptr [bp-4],0
    mov     ax,word ptr [bp+6]
    dec     ax
    mov     si,ax
    jmp     short @1@114
@1@58:
    lea     bx,word ptr [si-1]
    add     bx,bx
```

```

        mov     ax,word ptr [bx+di]
        mov     bx,si
        add     bx,bx
        mov     word ptr [bx+di],ax
        dec     si
@1@114:
        or      si,si
        jg      short @1@58
        mov     ax,word ptr [bp+4]
        mov     word ptr [di],ax
        xor     si,si
        jmp     short @1@226
@1@170:
        mov     bx,si
        add     bx,bx
        mov     ax,word ptr [bx+di]
        mov     bx,si
        add     bx,bx
        add     bx,word ptr [bp+8]
        imul    word ptr [bx]
        cwd
        add     word ptr [bp-4],ax
        adc     word ptr [bp-2],dx
        inc     si
@1@226:
        cmp     si,word ptr [bp+6]
        jl      short @1@170
        mov     dx,word ptr [bp-2]
        mov     ax,word ptr [bp-4]
        mov     cl,15
        call    near ptr N_LXRSH@
        jmp     short @1@282
@1@282:
        pop     di
        pop     si
        leave
        ret
@filter$qiipit3      endp

```

Любой ЦПС реализует алгоритм фильтрации «в одно касание»

(приведено выполнение свертки без ввода-вывода выборок для процессора Texas Instruments TMS320)

RPT N

MACD W, *AR-

Компактная запись, одна команда и всего один процессорный такт на отвод фильтра!

8.2. Фиксированная точка против плавающей

Процессоры ЦОС разделяются на две категории: с фиксированной точкой и с плавающей точкой. Тип процессора зависит от формата чисел, с которыми он манипулирует. ЦПС с фиксированной точкой обычно представляет каждое число минимумом 16-ю битами, хотя может быть использована и другая длина. Например,

Motorola производит семейство ЦПС с фиксированной точкой, которые используют 24 бита.

О форматах представления чисел с фиксированной точкой мы уже говорили. Обычно используется дополнительный код, представляющий числа от -1 до $+1$ ($1/2, 1/4, 1/8 \dots$).

В сравнении, ЦПС с плавающей точкой естественно использует минимум 32 бита, чтобы представить каждую величину. В наиболее часто используемом формате (ANSI/IEEE Std. 754-1985), самые большие и минимальные числа - $\pm 3.4 \times 10^{38}$ и $\pm 1.2 \times 10^{-38}$, соответственно. Представимые числа в указанном диапазоне распределяются неравно: промежуток между любыми двумя числами – примерно одна десятиллионная от величины числа.

Формат устанавливает большие промежутки между большими числами и малые промежутки между малыми числами.

Все ЦПС с плавающей запятой могут также оперировать числами с фиксированной точкой числа, необходимыми для организации счетчиков циклов, представления сигналов приходящих от АЦП и выходящих на ЦАП. Обычно ЦПС оптимизируются для выполнения операций как с плавающей так и фиксированной точкой с равной эффективностью.

Ранее мы говорили о том, что в процессорах общего назначения математические операции над числами с фиксированной точкой проводятся быстрее, чем над числами с плавающей точкой. В ЦПС скорость примерно одинакова. Внутренняя архитектура процессора с плавающей точкой более сложная. Все регистры и шины данных должны быть шириной 32 бита вместо 16; умножитель и ALU должны уметь быстро выполнять арифметику с плавающей запятой, система команд должна быть расширенной (нужно кодировать операции с обоими форматами), и так далее.

Процессор с плавающей точкой (32 бит) имеет лучшую точность и больший динамический диапазон. Кроме того, цикл разработки ПО для процессора с плавающей запятой часто бывает короче, так как программисту обычно не нужно беспокоиться о, например, переполнении, потере значимости и ошибках округления.

С другой стороны, ЦПС с фиксированной точкой традиционно более дешевы, чем устройства с плавающей запятой. Хотя ничто не изменяется быстрее, чем цена электроники, различие в стоимости останется в силу различия сложности приборов (порядок величин - \$5-50 для фикс. точки и \$15-200 для плавающей). (Не следует также забывать о потребляемой мощности и доступных размерах корпусов, особенно при проектировании мобильных устройств).

Обратим теперь наше внимание на то, что 32-битовая система ЦОС с плавающей запятой может делать, а 16-битовая с фиксированной точкой не может? Чем системы отличаются друг от друга. Ответ на этот вопрос – отношением сигнал-шум. Предположим, что мы сохраняем результат некоторых вычислений в виде числа в 32 битовом формате с плавающей запятой. Как мы упоминали прежде, промежуток между соседними числами в этом формате примерно равен одной десятиллионной от величины числа. Чтобы сохранить результат нужно его округлить в большую или меньшую сторону (изменить максимум на половину размеров указанного промежутка).

Другими словами, всякий раз, когда мы сохраняем число в форме с плавающей запятой, мы добавляем шум к сигналу.

То же самое происходит, когда мы оперируем с 16-битовыми числами с фиксированной точкой, за исключением того, что дополнительный шум - значительно больше. Дело в том, что промежутки между смежными числами - значительно больше. Например, предположим, что мы будем сохранять число 10000.5 как целое число со знаком (интервал допустимых значений от -32,768 - 32,767). Промежуток между числами - одна десятитысячная от величины числа. Если мы захотим сохранить число 1000.5,

промежуток между числами - только одна тысячная величины.

Для описания величины шума в сигналах обычно используется его среднеквадратичное отклонение. Вспомним, что среднеквадратичное отклонение этого шума квантования примерно равно одной трети размера промежутка между соседними числами. Это означает, что отношение сигнал-шум для чисел с плавающей запятой - около 30 миллионов, для чисел с фиксированной запятой это только десять тысяч. Другими словами, обработка чисел с плавающей запятой сопровождается генерацией шума квантования приблизительно в 30000 раз менее интенсивного, чем в случае с фиксированной точкой.

Предположим, что мы будем осуществлять КИХ-фильтрацию в формате с фиксированной точкой. Для того, чтобы делать это, мы, перебирая последовательно каждый коэффициент, умножаем его на соответствующую выборку входного сигнала и добавляем результат произведения к аккумулятору. Вот тут и возникает проблема. В традиционных микропроцессорах этот аккумулятор - просто другой 16-ти битный регистр или ячейка памяти, хранящая число с фиксированной точкой. Для того, чтобы избежать переполнения, нам нужно масштабировать добавляемые к содержимому аккумулятора величины, и, соответственно, добавлять шум квантования на каждом шаге (для каждого коэффициента фильтра), что существенно уменьшит отношение сигнал/шум системы. Например, в КИХ-фильтре с 500 коэффициентами, шум в каждой выходной выборке может быть 500 раз меньше шума в каждой входной выборке (на самом деле в $\sqrt{500}$, потому что добавляемые шумы квантования неоррелированы). Хотя это - экстремальный случай, он иллюстрирует основную проблему.

Для ее решения сигнальные процессоры используют аккумулятор повышенной разрядности. Это - специальный регистр, который имеет разрядность в 2-3 раза большую, чем основная память и регистры общего назначения. Например, в 16-ти битных ЦПС аккумулятор может иметь 32 - 40 бит, в то время как в SHARC DSPs он содержит 80 бит для накопления чисел с фиксированной точкой. Использование расширенного аккумулятора фактически устраняет шум округления во время накопления (для 32 бит не устраняет, для 40 - устраняет частично). Остается единственная ошибка округления, когда аккумулятор масштабируется для сохранения в 16-ти битовой памяти. Эта стратегия работает очень хорошо, хотя она ограничивает возможности некоторых алгоритмов, например БПФ. Системы с плавающей запятой имеют такой низкий шум квантования, который обычно даже не является необходимым.

В дополнение к более низкому шуму квантования для систем с плавающей запятой гораздо легче разрабатывать алгоритмы. Большинство алгоритмов ЦОС базируются на регулярных умножениях и сложениях. Для систем с фиксированной точкой возможность переполнения или потери значимости должна оцениваться после каждой операции. Программисту нужно непрерывно следить за «амплитудой» чисел, понимать, как накапливаются ошибки квантования, и решать, какое и где применить масштабирование. За исключением редких случаев эти вопросы не возникают при проектировании систем с плавающей запятой.

Как выбрать какой ЦПС использовать? Вот некоторые моменты, которые необходимо рассматривать при выборе используемого формата чисел. Сначала посмотрите, сколько битов используются в АЦП и ЦАП. В многих приложениях, например, телевидении и другом видео используются 8-ми битные ADC и ЦАП и здесь точность фиксированной точки приемлема. В сравнении, профессиональные звуковые приложения могут работать с выборками разрядностью 18, 20 или даже 24 бита, и здесь почти несомненно нужен процессор с плавающей запятой, чтобы обеспечить требуемый динамический диапазон.

Следующая вещь, на которую нужно посмотреть - сложность алгоритма, который будет выполняться.

Если алгоритм сравнительно прост, применяйте фиксированную точку; если это что-то более сложное, применяйте процессор, выполняющий операции с плавающей точкой. Например, КИХ-фильтрация требуют только несколько строк текста программы, делая их пригодными для фиксированной точки (при наличии аккумулятора достаточной разрядности). На контрасте, алгоритмы в частотной области, как, например, БПФ, очень трудно запрограммировать при использовании только чисел с фиксированной точкой. Время разработки существенно уменьшится, если использовать числа с плавающей точкой. Дело не только во времени кодирования, но и в результатах работы программы.

Обсудим влияние типа коэффициентов (поворачивающих множителей) быстрого преобразования Фурье на оценку спектра сигнала.

Очевидно, шумы квантования коэффициентов Фурье вносят дополнительные искажения в оценку спектра сигнала.

Для оценки этого влияния удобно ввести понятие эффективного числа разрядов сигнала, вычисляемое на основе отношения сигнал/шум. Отношение сигнал/шум квантованного сигнала вычисляется путем деления мощности максимального сигнала, представимого N-разрядным двоичным кодом, к мощности шума квантования (см. рис.8.6).

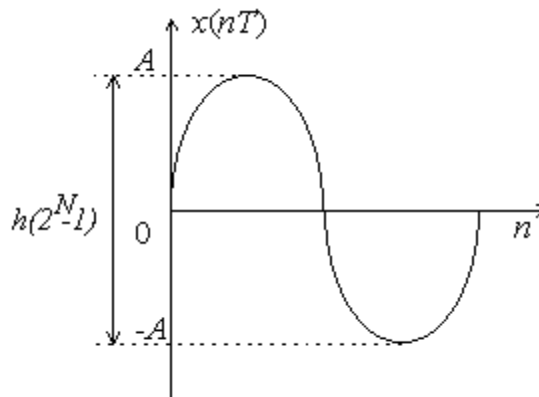


Рис.8.6

$$R_{S/N} = N \cdot 6.02 + 1.761 \text{ [dB]}$$

Отсюда эффективная разрядность, которая позволяет описать влияние на сигнал не только его квантования, но и шума, сопровождающего преобразования сигнала

$$N_{eff} = \frac{10 \cdot \lg\left(\frac{S_x}{S_e}\right) - 1.761}{6.02}.$$

Эффективная двоичная разрядность сигнала связана с относительным уровнем шума квантования. Чем ниже относительный уровень шума квантования, тем больше N_{eff} . Выражение $10 \cdot \lg(S_x/S_e)$ является отношением сигнал/шум выраженным в децибелах. Следовательно, один эффективный двоичный разряд гармонического сигнала соответствует, приблизительно, шести децибелам отношения сигнал/шум. При отсутствии цифровой обработки N_{eff} совпадает с теоретическим значением. Теоретический анализ приводит к следующей зависимости эффективного числа бит сигнала, подвергнутого преобразованию Фурье, от амплитуды сигнала (его исходной разрядности).

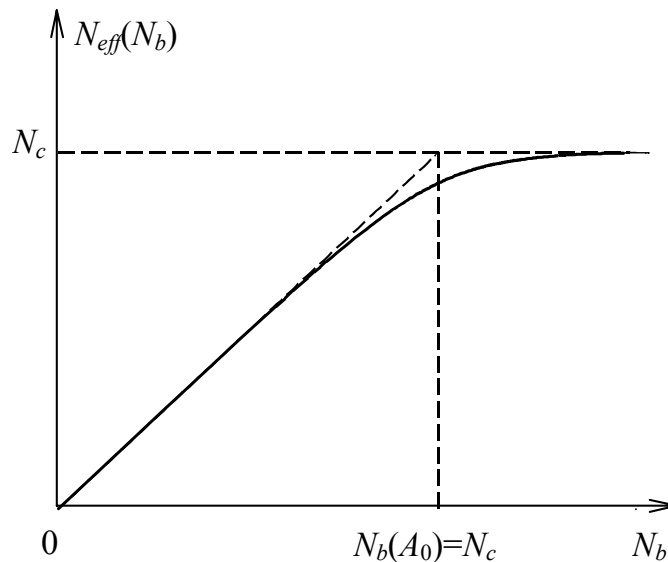


Рис.8.7

При малых амплитудах, $A \ll A_0$, эффективная разрядность совпадает с исходной разрядностью сигнала на выходе АЦП:

$$N_{eff} \approx \log_2(2A/h_x) = N_b.$$

В этом случае отношение сигнал/шум определяется мощностью (аддитивного) шума квантования в АЦП. Шумом квантования весовых коэффициентов можно пренебречь.

При большой амплитуде, $A \gg A_0$, эффективная разрядность стремится к постоянному значению, равному двоичной разрядности весовых коэффициентов преобразования Фурье:

$$N_{eff} \approx \log_2(2/h_c) = N_c.$$

Когда амплитуда существенно превышает пороговое значение, отношение сигнал/шум определяется, в основном, мультипликативным шумом квантования весовых коэффициентов. Мощность данного шума пропорциональна мощности сигнала. Следовательно, отношение сигнал/шум не зависит от амплитуды сигнала и принимает постоянное значение, которое определяется шагом квантования весовых коэффициентов. В этом случае эффективная разрядность сигнала после преобразования Фурье не зависит от амплитуды и определяется разрядностью весовых коэффициентов.

Были проведены численные эксперименты с различными типами БПФ и определена эффективная разрядность гармонического сигнала для каждого типа. Для целочисленного БПФ отсчеты сигнала представлены массивом целочисленных переменных (тип integer), амплитуда гармонического сигнала 2^{14} . Быстрое преобразование Фурье с плавающей точкой применяются к сигналам, представленным переменными типа long ($h_x=1$). Частота гармонического сигнала одинаковая в каждой реализации. Общее количество отсчетов спектра равно 512. Для уменьшения дополнительных ошибок округления применено БПФ типа double.

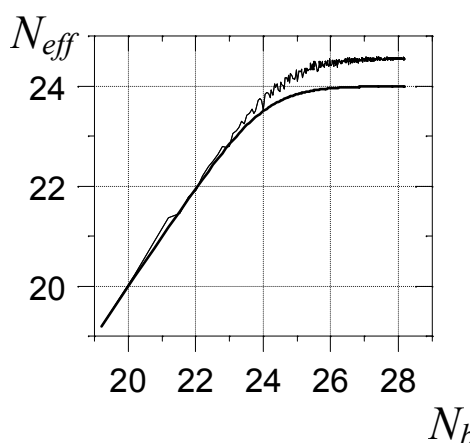


Рис.8.8

Из рисунка 8.8 видно, что при малой амплитуде сигнала A (то есть при низкой разрядности оцифрованного сигнала) эффективная разрядность практически совпадает с исходной. Данная часть зависимости $N_{eff}(N_b)$ определяется шумом квантования сигнала.

При большом значении исходной разрядности (то есть при большой амплитуде сигнала) эффективная разрядность принимает постоянное значение, определяемое разрядностью весовых коэффициентов: $N_{eff}=N_c$. Зависимость от амплитуды исчезает, что обусловлено мультипликативностью шума квантования указанных коэффициентов. Отсюда можно экспериментально оценить разрядность весовых коэффициентов.

Из проведенного численного эксперимента следует, что эффективная разрядность весовых коэффициентов, квантованных переменной типа float, равна примерно 24,5.

В таблице приведены экспериментальные данные для двоичной разрядности коэффициентов БПФ разных типов: целочисленного (integer) и с плавающей запятой одинарной (float) точности. Разумеется, двоичная разрядность переменных, используемых в промежуточных вычислениях БПФ, не должна быть меньше разрядности весовых коэффициентов.

В данном численном эксперименте тип промежуточных переменных БПФ и тип соответствующих весовых коэффициентов выбраны одинаковыми.

Тип БПФ	N_c
Integer	11
Float	23
Double (не эмулируется целочисленными ЦПС)	44

Принимая во внимание полученные результаты можно определить оптимальный тип весовых коэффициентов БПФ в зависимости от разрядности АЦП, которым оцифрован сигнал.

АЦП\БПФ	Integer	float	double
≤ 11	+	–	–
11-23	×	+	–
> 23	×	×	+

Здесь использованы следующие обозначения: “плюс” – рекомендуется, “минус” – не эффективно, “косой крест” – не допустимо.

Например, при оценке спектра сигнала, преобразованного 16-ти разрядным АЦП, следует использовать БПФ с весовыми коэффициентами одинарной точности (float), который дает 23 эффективных разряда в спектре.

В системах цифровой обработки сигналов часто при 16-ти разрядном АЦП применяется целочисленное БПФ типа integer, что приводит к потере пяти разрядов сигнала и ухудшению отношения сигнал/шум в спектре. Применение в данном случае весовых коэффициентов с плавающей запятой двойной точности типа double, которые имеют не менее 44 эффективных разрядов, непрактично, поскольку, займет больше времени и не улучшит оценку спектра.

Исследовалось также влияние длины БПФ на эффективную разрядность преобразованного сигнала. Характерная зависимость представлена на рисунке 8.9.

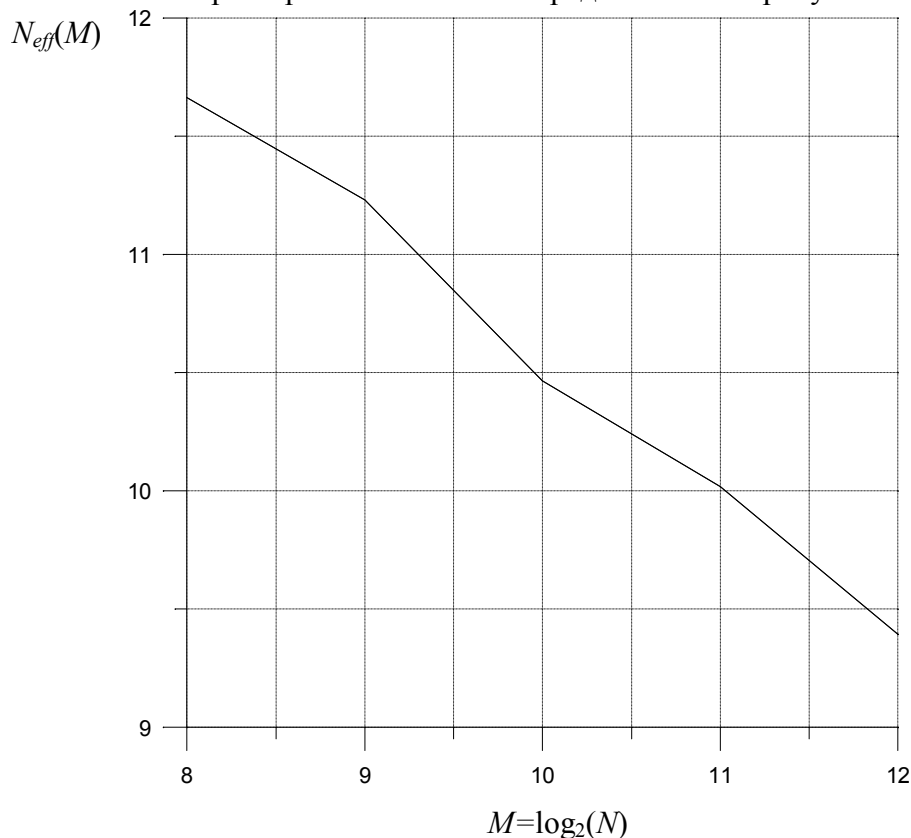


Рис.8.9

Хотя процессоры с плавающей точкой обладают явными преимуществами, не следует забывать о том, что далеко не последнюю роль играют размеры процессора, потребляемая мощность и, наконец, его цена.

Следующий рисунок (8.10) показывает тенденции в использовании ЦПС с фиксированной и плавающей точкой (данные 1998г.). Диаграмма (а) иллюстрирует долю, которую ЦПС имеют на рынке встроенных приложений. Это приложения, которые используют микропроцессор, чтобы непосредственно обслуживать и управлять некоторой большой системой, как например, сотовый телефон, микроволновая печь, или автомобильная панель приборов. Используемые здесь процессоры часто называют "микроконтроллерами", чтобы различить их с микропроцессорами, используемыми в персональных компьютерах. Как показано на (а), около 38% разработчиков встроенных систем уже начали использовать DSPs и другие 49% рассматривают возможность перехода на ЦПС. Высокая производительность и вычислительная мощность ЦПС часто делает их идеальным выбором для встроенных систем.

Как проиллюстрировано диаграммой (b), количество инженеров, к настоящему времени использующих процессоры с фиксированной точкой, примерно вдвое больше, чем использующих процессоры с плавающей точкой. Как мы уже говорили, использование того или иного типа процессора существенно зависит от приложения. Фиксированная точка более популярна в конкурирующих потребительских товарах, где

стоимость электроники должна держаться на очень низком уровне. Хороший пример это - сотовые телефоны. Когда Вы конкурируете, чтобы продать миллионы экземпляров вашего продукта, различие стоимости только на несколько долларов может быть различием между успехом и неудачей. В сравнении, процессоры с плавающей запятой используются когда нужна большая производительность, а стоимость не столь критична.

Например, предположим, что Вы разрабатываете медицинскую систему формирования изображения - томограф. Только несколько сот экземпляров может быть когда-либо продано по цене несколько сотен тысяч долларов каждый. Для этого приложения, стоимость ЦПС незначительная, но производительность - критическая характеристика. Несмотря на большое количество используемых процессоров с фиксированной точкой, рынок процессоров с плавающей запятой является самым быстрорастущим. Как показано на (с), свыше половины инженеров, использующих 16-битовые устройства, планируют перейти на процессоры с плавающей запятой в ближайшем будущем.

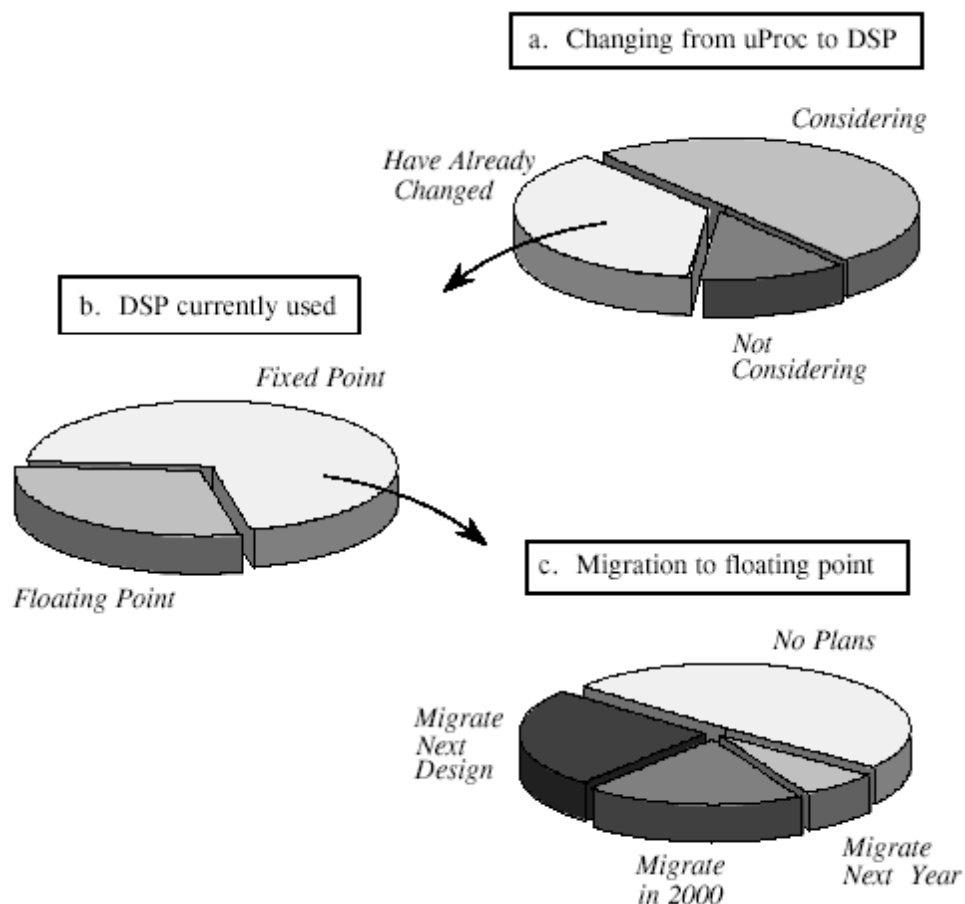


Рис. 8.10

Прежде, пойти дальше отметим, что обычно процессоры с плавающей и фиксированной точкой обычно используют 32 бита и 16 бит, соответственно, но не всегда. Например, процессоры семейства SHARC могут представить числа в 32-битовом формате фиксированной точки, режиме, который часто используется, скажем, в цифровых звуковых приложениях. Числа имеют 2^{32} уровней квантования, равномерно распределенных в интервале, скажем, между -1 и 1. В сравнении, запись в форме с плавающей запятой устанавливает 2^{32} уровней квантования, логарифмически распределенных в огромном интервале $\pm 3.4 \cdot 10^{38}$.

Это дает лучшую точность представления числа для 32-битовой фиксированной точки и уменьшение ошибки квантования. Тем не менее, 32-битовые числа с плавающей запятой имеют больший динамический диапазон, то есть большее отношение

максимального и минимального представимого числа.

8.3. Как быстры ЦПС?

Первичная причина использования ЦПС вместо традиционного микропроцессора - скорость, способность быстро вводить выборки, выполнять необходимые математические операции и выводить обработанные данные. Это вызывает вопрос: как быстры ЦПС? Обычный способ ответить на этот вопрос – использование тестов, оценивающих скорость микропроцессора как некое число. Например, скорость процессора с фиксированной точкой часто оценивается в MIPS - миллионах целочисленных операций в секунду (миллионах инструкций в секунду). Подобно, производительность устройств с плавающей точкой может определяться в MFLOPS (миллионах операций с плавающей точкой в секунду).

Сто пятьдесят лет тому назад Британский Премьер-министр Бенджамин Дизраэли сказал, что есть три типа лжи: ложь, ложь под присягой и статистика. Если Дизраэли был бы жив сегодня и работал с микропроцессорами, он должен был бы добавлять benchmarks как четвертую категорию. Идея тестов производительности в том, чтобы сравнить разные системы при выполнении одних и тех же операций, чтобы показать, которая из систем является лучшей. К несчастью, это часто терпит неудачу в практике, каждый из микропроцессоров хорош в своей области. Какой автомобиль лучше, Кадиллак или Феррари? Ответ зависит от того, для чего Вам нужен автомобиль!

Неразбериха с тестами усугубляется конкурентной природой электронной промышленности. Изготовители хотят показать свои продукты в наилучшем свете, и они будут использовать любую неоднозначность в испытательной процедуре для демонстрации преимуществе своих изделий. Лучший способ оценки производительности – получение информации о скорости выполнения именно тех алгоритмов, которые Вы планируете выполнять. Например, если ваша система представляет собой КИХ-фильтр, найдите точное число процессорных циклов, требуемых для выполнения этого конкретного задания. Используя эту стратегию, давайте посмотрим на время требуемое, чтобы выполнить различные алгоритмы на рассматриваемых ЦПС семейства SHARC. Имейте в виду, что скорость микропроцессоров удваивается каждые три года. Это означает, что нужно уделять особое внимание методам, которое мы используем в этом примере. Фактические значения всегда изменяются, и вам понадобится повторять вычисления каждый раз, который вы начинаете новый проект. В мире технологий двадцать первого века миг - и данные устарели!

Как и большинство современных ЦПС семейства SHARC процессоров может выполнить умножение-накопление в единственном такте. Поскольку большинство КИХ-фильтров имеют 25 - 400 коэффициентов, 25 - 400 циклов требуются, соответственно, для каждой обрабатываемой выборки. Как мы говорили прежде, есть небольшие накладные расходы на организацию цикла, незначительные, когда порядок фильтра большой. Чтобы оценить производительность процессора, мы можем разделить тактовую частоту процессора (40 МГц в настоящее время) на число циклов, требующихся на обработку одного отвода фильтра. Это дает нам максимальную скорость КИХ фильтрации от 100k до 1.6М выборок /секунду для фильтра длиной 400 – 25, соответственно.

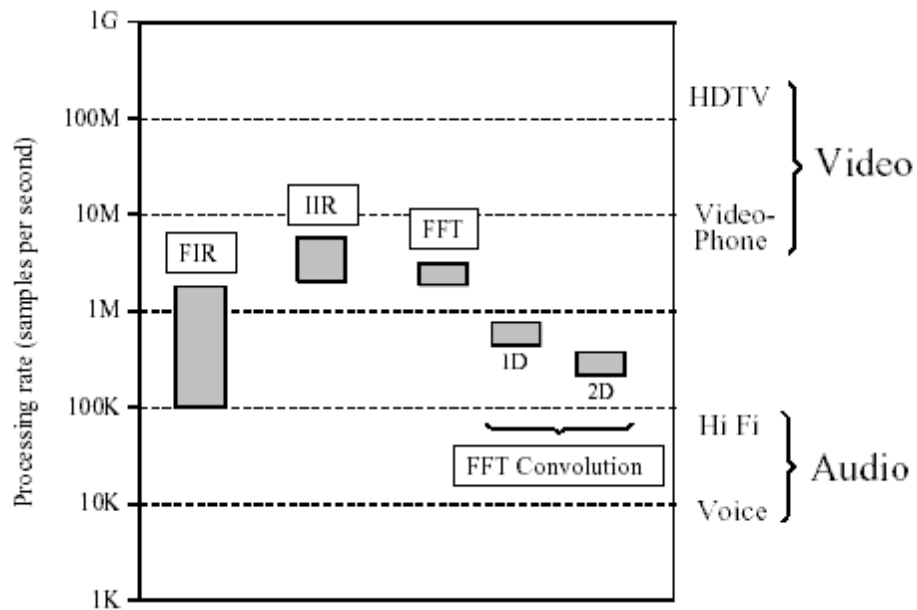


Рис.8.11

Производительность конкретного алгоритма ЦОС может определяться делением тактовой частоты на число циклов, расходуемых на обработку каждой выборки. Эта иллюстрация показывает область производительности для ряда общих алгоритмов, выполняемых на SHARC DSP при тактовой частоте 40 МГц.

Также легко произвести вычисление производительности для рекурсивных фильтров. Типичные БИХ-фильтры (IIR) используют около 5 - 17 коэффициентов. Поскольку циклов, требуемых на умножение-накопление совсем немного, мы добавим небольшие накладные расходы, скажем, 3 цикла на выборку. Итого получим 8 - 20 циклов, требуемых на одну выборку обрабатываемого сигнала. Для 40 МГц тактовой частоты производительность БИХ-фильтра составит 1.8М - 3.1М выборок/секунду (см. рисунок).

Теперь перейдем в технику обработки сигналов в частотной области, базирующуюся на быстром преобразовании Фурье. Подпрограммы БПФ почти всегда предусматриваются изготовителем ЦПС. Это очень оптимизированные программы, написанные на ассемблере. Спецификация ADSP-21062 SHARC DSP указывает, что выполнение 1024-точечного комплексного БПФ требует 18221 циклов, или около 0.46 миллисекунд при 40 МГц тактовой частоте. Чтобы вычислить производительность, это удобнее рассматривать как 17.8 циклов на выборку. Этот параметр "на выборку" слабо изменяется с увеличением или уменьшением длины БПФ. Например, 256-ти точное БПФ требует около 14.2 циклов на выборку, а 4096-ти точечное БПФ требует 21.4 циклов на выборку. Действительные БПФ могут вычисляться примерно на 40% быстрее чем комплексные. Таким образом, производительность ЦПС при вычислении БПФ составляет от 1.8М до 3.3М выборок/секунду.

«Быстрая» свертка, вычисляемая с помощью БПФ, - лучший способ реализации КИХ- фильтра. В типичном случае, 512-точечный сегмент данных, взятый из последовательности входных выборок, дополняется 512 нулями и преобразуется в спектр, используя 1024-точечное БПФ. После умножения этого спектра на желаемую частотную характеристику применяется 1024-точечное обратное БПФ для возврата во временную область. Полученные 1024 точки объединяются с наложением со смежными обработанными сегментами. Сколько на эту операцию нужно циклов? Каждый 512-выборочный сегмент для обработки требует два 1024-точечных БПФ плюс небольшое количество тактов на перемножение спектра сигнала на АЧХ и накладные расходы. Округленно это раз в 5 больше, чем необходимо для выполнения 512-точечного БПФ.

Поскольку для выполнения действительного 512-точного БПФ требуется около 12 циклов на выборку, «быстрая» свертка может выполняться примерно за 60 циклов на выборку. Для 2106x SHARC DSP с тактовой частотой 40 МГц, это соответствует производительности обработки данных приблизительно 660k выборок/секунду.

Обратите внимание, что 60 циклов это столько же, сколько требуется для реализации КИХ-фильтра 60 порядка обычным способом во временной области. Другими словами, если КИХ-фильтр имеет менее чем 60 коэффициентов, более быстрым он будет при стандартной реализации свертки. Если он имеет больше чем 60 коэффициенты, «быстрая» свертка предпочтительнее. Ключевое преимущество «быстрой» свертки в том, что время выполнения увеличивается только как логарифм количества коэффициентов.

Быстрая свертка с применением БПФ может также использоваться для обработки двумерных сигналов, например, для обработки изображений. Для 40 МГц ADSP-2106x получится производительность около 260k выборок/секунду.

Теперь, когда мы понимаем, как быстро ЦПС могут обрабатывать цифровые сигналы, давайте подумаем, а как быстро нам нужно обрабатывать данные? Конечно, это зависит от приложения. Мы рассмотрим две наиболее общие задачи, обработку звука и видео.

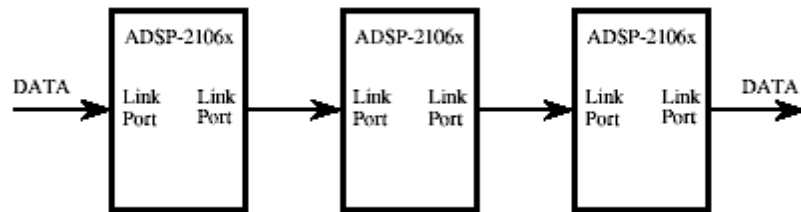
Скорость, необходимая для обработки звукового сигнала зависит от необходимого качества воспроизводимого звука. Так речь «телефонного» качества требует обработки сигнала в полосе частот от 100 Гц до 3.2 кГц, т.е. частоты дискретизации в 8k выборок/секунду. В сравнении, для воспроизведения музыки с высокой точностью в диапазоне, соответствующем чувствительности человеческого уха 20 Гц - 20 кГц, необходима частота выборки около 44 кГц. Эта частота дискретизации используется для как левого, так для и правого канала, вследствие чего получается частоты выборки сигнала 88.2k выборок/секунду. Насколько процессоры семейства SHARC удовлетворяют этим требованиям? Как следует из наших расчетов, результаты которых представлены на рисунке, процессоры легко могут оперировать звуковым сигналом высокой верности, или обрабатывать параллельно десятки голосовых сигналов.

Видео сигналы являются более широкополосными. Скорость их передачи (и, следовательно, оцифровки) по крайней мере в тысячу раз превышает скорость передачи звуковых сигналов. Хороший пример видео низкого качества - CIF (Common Interface Format) - стандарт для видеофонов. Используется разрешение 352*288 пикселей, с 3 цветами на пиксель, и 30-ю кадрами в секунду; общая скорость передачи данных - 9.1 миллионов выборок в секунду. Стандарт, описывающий видео высокого качества, HDTV (high-definition television), предусматривает использование разрешения 1920*1080 (high-definition television) при 30 кадрах в секунду. Это требует скорости передачи данных свыше 186 миллионов выборок в секунду. Эти показатели превышают возможности одного SHARC DSP. Есть и другие приложения, которые также требуют очень высокой производительности, такие, например, как радио- и гидролокация, управление ракетным оружием.

Для того, чтобы оперировать с заданной производительностью, несколько ЦПС могут объединяться в единую систему. Это называется многопроцессорной или параллельной обработкой. SHARC DSPs разрабатывались с учетом возможности реализации многопроцессорной обработки и включают специальные средства, позволяющие упростить построение многопроцессорных систем. Так, например, не требуется никакой внешней дополнительной аппаратной логики для объединения внешних шин нескольких SHARC DSPs вместе; вся логика арбитража шин уже содержится внутри каждого процессора. Как альтернатива, специальные порты связи (параллельные, 4 бита) могут использоваться, чтобы соединить многочисленные процессоры в различных конфигурациях. Рисунок 8.12 показывает типичные способы соединения SHARC DSPs в системах многопроцессорной обработки. На рисунке (a)

алгоритм разбивается на последовательные шаги, и каждый процессор выполняет один из шагов в конвейере.

a. Data flow multiprocessing



b. Cluster multiprocessing

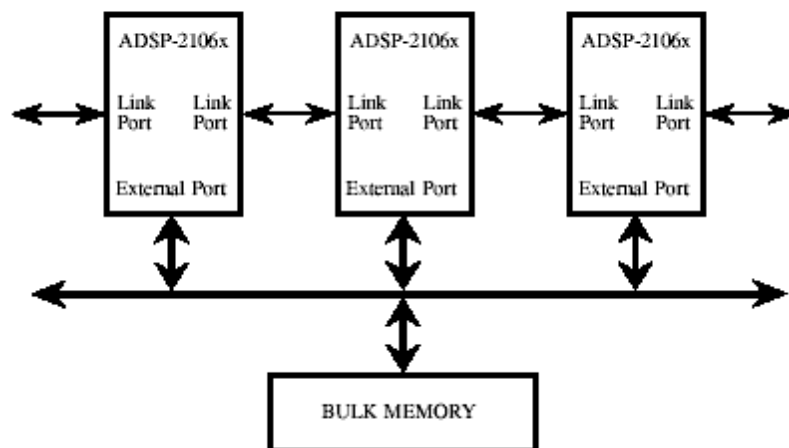


Рис.8.12

Конфигурации многопроцессорной обработки. Многопроцессорные системы естественно используют одну из двух схем для передачи данных между узлами системы, (a) – специально предназначенные каналы связи, или (b) - коллективную глобальную память, доступную через параллельную шину.

На рисунке (b) процессоры взаимодействуют через единственную коллективную глобальную память, доступную через параллельную шину. Следующий рисунок 8.13 показывает другой способ соединения многих процессоров в единую систему, двумерную или трехмерную «сетку». Каждая из этих конфигураций будет иметь относительные преимущества и недостатки для конкретного приложения.

Для того, чтобы сделать жизнь программиста легче, SHARC семейство использует унифицированное адресное пространство. Это означает, что 4-х Гигасловное адресное пространство, доступное через 32-битные шины адреса, делится среди различных процессоров, которые работают вместе. Для того, чтобы передавать данные из одного процессора в другой просто нужно прочитать или записать их в соответствующих областях памяти.

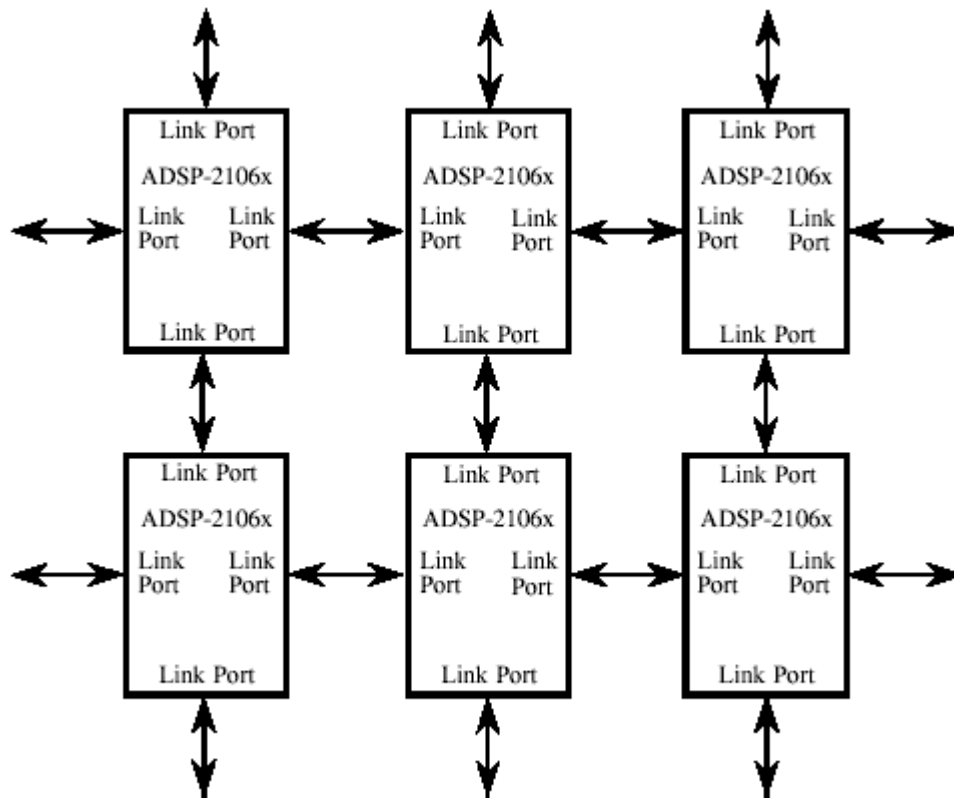


Рис.8.13

Многопроцессорная Обработка в конфигурации "сетка". Наиболее подходит для выполнения пространственно-временной обработки сигналов в радио- и гидролокации.

Более подробно со структурной схемой и особенностями реальных ЦПС мы познакомимся на примере семейства целочисленных процессоров C54xx фирмы Texas Instruments – основного конкурента Analog Devices.

8.4 Оптимизированная архитектура ядра C5000

Высокий уровень производительности процессоров семейства C5000 стал возможен благодаря их высокопроизводительной архитектуре, специально разработанной с учетом требований широкого круга приложений (рис.8.14). Так, например, в состав ядра процессоров включен Витерби-ускоритель, который сокращает выполнение операции “butterfly update”, применяемой в декодировании GSM канала, до 4-х команд процессора.

Другие важные особенности архитектуры ядра C54xx таковы:

- **Четыре** внутренние шины и два генератора адреса позволяют параллельно выполнять операции с несколькими операндами и снижают критические параметры доступа к памяти.
- 40-разрядный сумматор и два 40-разрядных аккумулятора позволяют производить параллельную обработку инструкций в одном цикле команды.
- Второй дополнительный 40-разрядный аккумулятор на выходе умножителя позволяет выполнять неконвейерные MAC операции типа двух суммирований и умножения в параллель.
- Одноцикловая нормализация и экспоненциальное декодирование дает поддержку операций с плавающей точкой, которые часто используются при кодировании речи.
- Умножитель 17х17 разрядов позволяет выполнять 16-разрядное знаковое или беззнаковое умножение с округлением и контролем насыщения - и все это в одном

командном цикле.

- Новые одноктактные команды эффективно выполняют общие ЦОС операции, такие как обработка звена симметричного КИХ фильтра.
- 40-разрядное АЛУ может работать в режиме двойного 16-разрядного, что позволяет выполнять двойные одноктактные операции.

Восемь дополнительных регистров и программный стек позволяют применить наиболее развитый и эффективный среди процессоров с плавающей точкой С-компилятор.

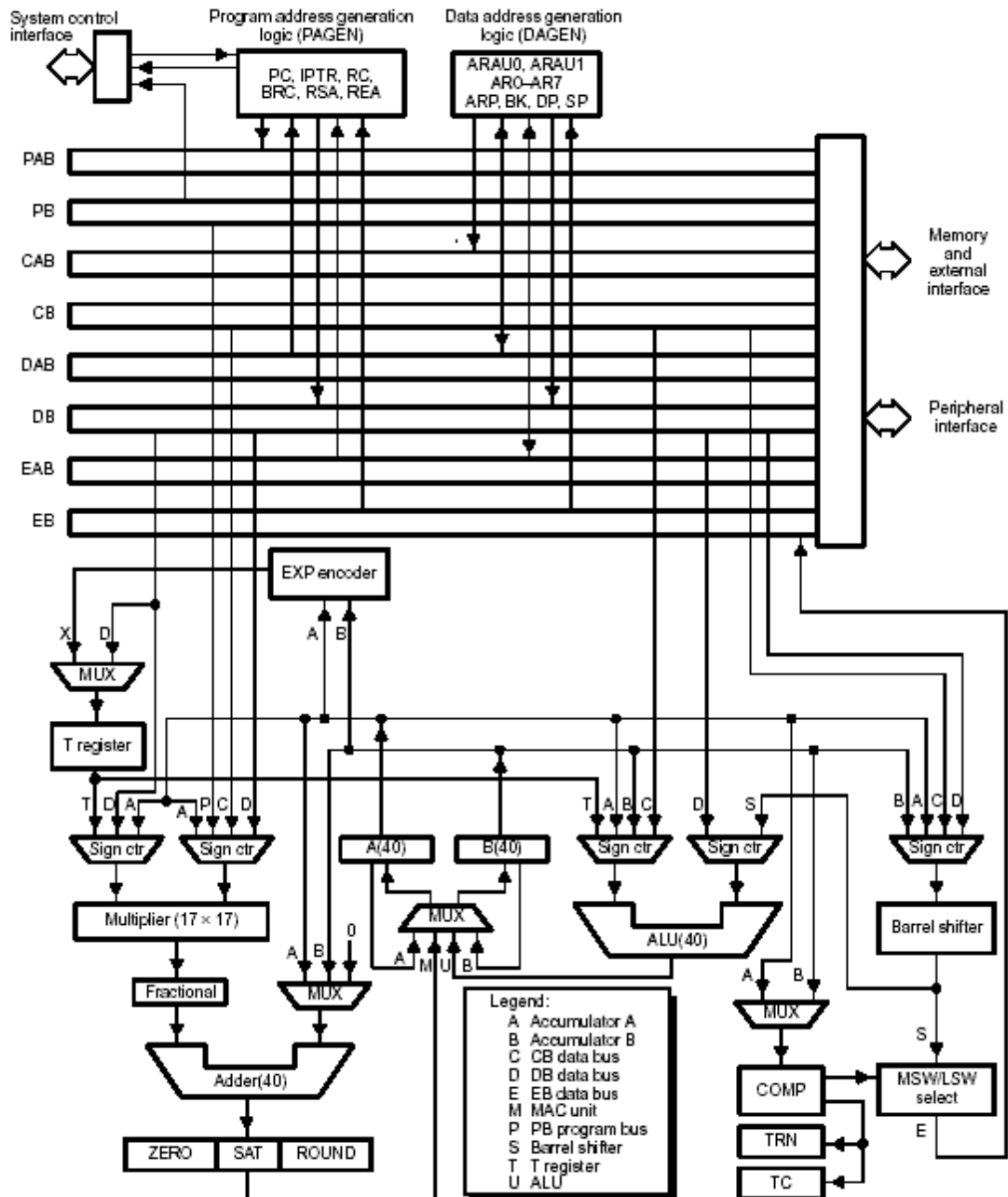


Рис.8.14

Семейство C5000 построено по модифицированной Гарвардской архитектуре. В нем используется 3 отдельные системы шин для данных и одна для команд. Разделение памяти команд и памяти данных позволяет выполнять одновременный доступ к командам и данным, что повышает параллелизм выполнения. Например, в одном цикле могут быть

выполнены две операции чтения и одна операция записи. Команды с параллельной записью и специальные команды полностью используют все преимущества данной архитектуры. Кроме того, данные могут пересылаться между областями команд и данных. Такая степень параллелизма дает возможность поддерживать мощный набор арифметических, логических операций и операций с битами, которые выполняются за один машинный цикл. В дополнении к этому, ядро C54xx включает в себя эффективные механизмы управления прерываниями, повторения операций и вызовов подпрограмм.

Структура шин. Архитектура C54xx состоит из 8 основных шин:

Одна шина чтения команд (PB), которая предназначена для выборки команд и непосредственных операндов.

Две шины чтения данных (CB и DB) и одна шина записи данных (EB), которые соединяют большинство элементов системы. Через шины CB и DB операнды читаются из памяти данных, а через шину EB - пишутся.

Четыре адресных шины (PAB, CAB, DAB и EAB), на которые выставляются адреса, необходимые для выполнения команды.

Устройства C54xx имеют возможность генерации одновременно до двух адресов данных, которые сохраняются в двух дополнительных регистрах (ARAU0 и ARAU1).

Шина PB может переносить операнд, хранимый в памяти программ (для примера, таблицу коэффициентов), на вход умножителя для выполнения операций умножения/суммирования или в область данных для операций передачи данных. Это дает возможность использовать одноктактовые команды с тремя операндами.

Устройства C54xx также имеют внутреннюю двунаправленную шину для доступа к размещенной на кристалле периферии. Эта шина соединяется с шинами DB и EB через мост.

Через этот же мост к внутренней структуре шин подключается и внешняя шина процессора, для интерфейса с внешней памятью и устройствами.

Доступ к устройствам через эту шину может требовать более чем двух тактов для чтения и записи - это зависит от внешнего устройства.

Устройства семейства C54xx имеют систему поддержки шины данных и шины адреса, исключаящую на ней плавающие уровни в то время, когда шина находится в третьем состоянии.

Память. Полное адресуемое пространство для процессоров семейства C54xx составляет 192К 16-разрядных слов. Память делится на три сегмента по 64 К Слов: сегмент команд, сегмент данных и сегмент ввода-вывода. Сегмент команд служит для хранения программы, также в нем могут содержаться некоторые данные, например, таблицы коэффициентов. Сегмент данных содержит данные, используемые при выполнении программы. Сегмент ввода-вывода используется для доступа к внешним периферийным устройствам и может также использоваться как расширение сегмента данных.

Параллельная архитектура ядра процессора позволяет выполнять 4 конкурирующих операции с памятью в одном такте: выборка команды, чтение двух операндов и запись результата. Для этого используются четыре параллельных системы шин: шина чтения команд (PB), шина записи данных (EB) и две шины чтения данных (CB и DB). Каждая шина имеет доступ к разным областям памяти для разных видов операций. Такая архитектура позволяет выполнять параллельное чтение двух операндов, 32-разрядный доступ и одновременное чтение и запись операндов.

На кристалле C5000 располагаются два вида внутреннего ОЗУ - с двойным доступом и с одинарным доступом. ОЗУ с двойным доступом делится на блоки по 2Кслов. Данные в каждом из этих блоков могут быть считаны или записаны дважды за один машинный такт. Основное назначение данной памяти - хранение данных, однако она может быть сконфигурирована и как память команд.

Старшие модели семейства могут поддерживать до 8М 16 -разрядных слов (см. таблицу характеристик) и имеют эффективные механизмы работы со страничной адресацией, позволяющие производить автоматическое переключение старших разрядов адреса. Эта память организована как 128 страниц длиной по 64Кслов. При этом в процессоре добавляются дополнительные 7 линий адреса и расширенный регистр счетчика команд, а также ряд дополнительных команд.

Начальный загрузчик. Начальный загрузчик - это программа, размещенная во внутреннем ПЗУ процессора. Он используется для переписывания программы из внешнего источника в ОЗУ процессора или внешнее ОЗУ, где она будет исполняться. Начальный загрузчик вызывается автоматически при включении питания или при сбросе процессора, если установлен соответствующий режим. Режим устанавливается переводом вывода МР/МС в нулевое состояние. Если при сбросе процессора этот вывод находится в состоянии 0, то процессор начинает выполнять команды из внутреннего ПЗУ процессора. Начальный загрузчик может выполнять загрузку программы из большого набора источников:

- Параллельное 8- или 16-разрядное ПЗУ

- Параллельная загрузка из регистра ввода-вывода

- Загрузка из последовательного порта с 8- или 16-разрядными данными

- Загрузка через Хост-порт интерфейс

- Мягкий старт

Режимы загрузки программы из последовательного или из хост-порта дают возможность отказаться от отдельного ПЗУ для ЦСП и загружать программу их хост процессора без каких либо дополнительных затрат. При этом не требуется подведения громоздкой параллельной шины и для загрузки программы используется тот-же интерфейс, что и для связи между ЦСП и хост-процессором.

Внутренняя периферия.

Все процессоры семейства C54xx имеют множество размещаемых на кристалле периферийных устройств. Конкретный набор зависит от модели процессора. В этот набор входят

- программируемый генератор задержек

- программируемый переключатель банков памяти

- параллельный порт ввода-вывода

- широкий выбор последовательных портов

- таймер

- тактовый генератор

Программируемый генератор задержек служит для увеличения длительности цикла обмена по шине до 7 машинных тактов при обмене с медленными внешними устройствами или с медленной внешней памятью. Он работает без использования какого-либо внешнего аппаратного обеспечения. Количество тактов задержки задается для набора областей памяти через специальные регистры.

Программируемый переключатель банков памяти служит для автоматической вставки одного дополнительного такта при пересечении границ банка памяти или при переходе от памяти команд к памяти данных. Этот дополнительный цикл позволяет одному устройству памяти освободить шину перед тем, как ее займет другое устройство памяти, что исключает аппаратные конфликты на шине. Размер банка памяти задается специальным регистром.

Параллельные порты ввода-вывода. Каждый процессор семейства C5000 может адресовать 64К портов ввода-вывода. Доступ к портам осуществляется специальными командами ввода-вывода. Через пространство портов ввода-вывода организуется наиболее простой интерфейс к внешним устройствам, требующий минимального количества внешней логики декодирования адреса.

Хост-порт интерфейса. ХПИ это параллельный 8-разрядный (или 16-разрядный) интерфейс для связи ЦСП с хост-процессором (управляющим контроллером). Обмен информацией между ЦСП и хост-процессором идет через блок во внутреннем двухпортовом ОЗУ ЦСП размером в 2Кслов, к которому имеют доступ как ЦСП, так и хост-процессор. Возможна организация 16 разрядного доступа через обмен двойными байтами. ХПИ может работать в двух режимах - в первом из них хост-процессор и ЦСП имеют конкурентный доступ к этому блоку памяти с приоритетом хост-процессора. Во втором режиме хост-процессор имеет доступ к внутренней памяти ЦСП, когда ЦСП находится в спящем режиме или в состоянии сброса. Обмен с ХПИ возможен и тогда, когда процессор находится в состоянии IDLE2, то есть остановлены все тактовые сигналы как ядра процессора, так и периферии. При этом не происходит выхода процессора из этого режима и, соответственно, увеличения потребления питания. Скорость обмена через ХПИ в 16-разрядном режиме может достигать 20 МБайт в секунду.

Тактовый генератор. Процессоры семейства C5000 имеют гибкую систему формирования тактовых импульсов с переменными параметрами. Эта система позволяет как делить входную частоту, так и умножать ее, что дает возможность гибкого выбора частоты внешнего генератора и рабочей частоты процессора. Возможность умножения частоты позволяет использовать для генерации быстрого внутреннего тактового сигнала низкую частоту. Это уменьшает количество проблем с наводками внешней тактовой частоты, с разводкой ее по плате, позволяет использовать более медленные генераторы и резонаторы и снизить энергопотребление системы. Кроме того, старшие модели семейства имеют возможность программной подстройки частоты работы процессора. Коэффициент умножения частоты при программной подстройке может меняться от 0,25 до 16. Это позволяет менять тактовую частоту программно, в процессе работы, в зависимости от требуемого режима работы. Вместе с тактовой частотой меняется и ток потребления процессора. Таким образом, можно, меняя ток потребления, добиться компромисса между временем активной работы процессора и потребляемой им в активном режиме мощностью, что особенно важно для мобильных систем с лимитированной емкостью источника питания. При этом для каждого режима работы устройства можно устанавливать оптимальное соотношение между временем выполнения задачи и током потребления процессора.

Контроллер прямого доступа к памяти (ПДП). Позволяет осуществлять передачу данных из одной области памяти в другую без участия ЦПУ. Контроллер ПДП позволяет пересылать данные в и из памяти команд, памяти данных, внутренней периферии, такой как МКБПП, или внешней памяти, параллельно с работой ЦПУ. Всего контроллер ПДП имеет 6 независимых каналов с программируемым приоритетом.

Последовательные порты. Процессоры семейства C5000 имеют несколько видов последовательных портов.

Стандартный последовательный порт - синхронный последовательный порт с независимыми каналами приема и передачи, двойной буферизацией, программируемыми тактовыми частотами и длиной данных.

TDM порт. Позволяет обмениваться данными в режиме временного разделения. Всего через этот порт могут быть соединены до 7 устройств. При этом каждому устройству выделяется свой временной интервал для обмена.

Буферизованный последовательный порт - имеет возможность прямого чтения и записи данных во внутреннюю память процессора. Размеры и положение буфера в памяти программируются пользователем. При работе буферизованного порта прерывание вызывается не на каждый прием или передачу данных, а при заполнении или исчерпании буфера данных. Максимальная скорость обмена по буферизованному последовательному порту может достигать 50 мегабит/сек.

Многоканальный буферизованный последовательный порт (МКБПП).

Это скоростной синхронный последовательный порт. Он имеет возможность читать/записывать данные в память без участия процессора через контроллер ПДП. Кроме того, у него существуют многоканальные расширения, совместимые со стандартами E1, T1, SCMA и MVRP.

Как и стандартный порт, МКБСП имеет следующие возможности:

Полнодуплексная работа

Двойная буферизация данных - позволяет поддерживать непрерывность потока.

Независимые тактовые частоты и схемы синхронизации для приема и передачи

Прямое подключение стандартных промышленных кодеков, микросхем аналоговых интерфейсов и микросхем ЦАП/АЦП с последовательным интерфейсом.

В дополнение к уже перечисленным МКБСП имеет следующие возможности:

Прямое подключение к шинам

E1/T1

ST-BUS

IOM-2

AC97

IIS

SPI

Многоканальный обмен при количестве каналов до 128.

Переменный размер данных 8, 12, 16, 20, 24 и 32 бита

Встроенное u-Law и A-Law компандирование

Возможность передачи первым старшего или младшего разряда данных

Программируемая полярность сигналов синхронизации и тактовых сигналов данных

Гибкое программирование внутренних тактовых импульсов и синхронизации.

МКБСП состоит из канала приема и канала передачи, которые работают полностью независимо. При работе в многоканальном режиме используется временное разделение каналов. При этом возможно указание каналов, для которых осуществляется прием и передача данных. Максимальная тактовая частота МКБСП равна половине тактовой частоты процессора.

Программируемый таймер. Каждый процессор семейства C5000 имеет 16-разрядный программируемый таймер с 4-разрядным предварительным делителем частоты. Счетчик таймера уменьшается на единицу в каждом такте процессора; когда содержимое счетчика становится равным 0, вызывается прерывание. Таймер можно запустить, остановить, сбросить или замаскировать через управляющий регистр.

Снижение энергопотребления.

Для мобильных телекоммуникационных систем очень критичным параметром является потребление питания. При этом развитие систем приводит к усложнению алгоритмов обработки, соответственно требуются все большие ресурсы на их выполнение, и, соответственно, возрастает потребляемая мощность. Требуется найти компромисс между потребляемой мощностью устройства и скоростью его работы. При разработке таких систем очень важно свести до минимума время активной работы процессора и подобрать оптимальный режим его работы, чтобы не превышать предельное потребление в активном режиме. Процессоры C5000, как ориентированные на рынок мобильных телекоммуникационных систем, имеют широкий набор средств для решения этой проблемы.

Семейство процессоров C5000 имеет три режима снижения энергопотребления. Они устанавливаются командами IDLE1, IDLE2 и IDLE3. В этих режимах различные части процессора переходят в неактивное состояние, и процессор рассеивает значительно меньше мощности, чем в активном режиме. При выполнении команды IDLE1 происходит останов ядра процессора. Однако при этом на периферийные модули продолжает

подаваться тактовый сигнал, и они остаются активными. Команда IDLE2 останавливает ядро процессора и периферийные модули, а команда IDLE3 используется для полного останова всех блоков процессора, включая схему тактового генератора и умножителя частоты.

Специальное внимание при разработке C5000 было уделено быстрому выходу из режимов пониженного энергопотребления, что позволяет использовать этот режим как можно чаще.

Теперь рассмотрим измерение потребления процессора. В данном случае Милливатт на операцию - наиболее правильный метод измерения рассеиваемой мощности.

Измерение производительности в MIPS важно, но не учитывает того, что из-за различий в архитектуре разные ЦСП с одинаковой производительностью в MIPS могут существенно различаться по времени выполнения реальных задач.

Когда ЦСП выполняет операции более эффективно, то у проектировщиков появляется возможность использовать незадействованные MIPS для выполнения тех задач, которые обычно выполняются внешними ASIC или микроконтроллерами. Такая интеграция экономит не только место, занимаемое системой, но и потребляемую мощность. Кроме того, когда операции выполняются ЦСП более эффективно, он может больше времени пребывать в режиме низкого энергопотребления - IDLE режиме.

Таким образом, милливатты на операцию - более точный и правильный метод оценки эффективности ЦСП.

ЦСП семейства C5000 допускают снижение напряжения питания до 2,5 вольт при полной производительности в активном режиме и имеют специальные особенности архитектуры для подбора оптимальных параметров активного режима работы.

Например, процессору C5000 требуется около 12,7 MIPS для обработки полноскоростного GSM, 26.2 MIPS для обработки GSM с половинной скоростью в канале, только 2.3 MIPS для речевого кодека GSM и всего 12.8 MIPS для IS-54/136 VSELP.

Перевод этого в милливатты за операцию дает потребление в 31.1 милливатт для IS-54/136 VSELP и всего 5.6 милливатт для GSM кодека. Таким образом, эффективность архитектуры C5000 дает TI возможность предоставлять минимальное потребление в милливаттах за операцию для рынка систем беспроводной связи.

Новые ЦСП на базе ядра C55xx позволяют существенно улучшить эти показатели, доведя соотношение производительность/ потребление для ядра процессора до 0.05 милливатта на MIPS.