

Лекции по теории сигналов

A.A. Григорьев

1 Элементы теории сигналов

1.1 Пространство сигналов с конечной энергией

1.1.1 Основные понятия

В линейном пространстве комплексных сигналов $x(t)$ с конечной энергией

$$\int |x(t)|^2 dt < \infty$$

скалярное произведение

$$\langle x(t), y(t) \rangle = \langle x, y \rangle = \int x(t)y^*(t) dt$$

вводит структуру Гильбертова пространства. Элементы этого пространства (сигналы) становятся векторами.

Свойства скалярного произведения:

Линейность по левой координате

$$\langle \alpha x_1 + \beta x_2, y \rangle = \alpha \langle x_1, y \rangle + \beta \langle x_2, y \rangle$$

Антилинейность по правой координате

$$\langle x, \alpha y_1 + \beta y_2 \rangle = \alpha^* \langle x, y_1 \rangle + \beta^* \langle x, y_2 \rangle$$

Сопряженная симметрия

$$\langle x, y \rangle^* = \langle y, x \rangle$$

Вещественность и неотрицательность нормы

$$\langle x, x \rangle \geq 0$$

Как обычно, векторы считаются ортогональными, если их скалярное произведение $\langle x, y \rangle$ равно нулю. Норма (длина) вектора вводится через скалярное произведение

$$\|x\|^2 = \langle x, x \rangle, \|x\| = \sqrt{\langle x, x \rangle}$$

Квадрат нормы имеет смысл энергии сигнала. Расстояние между сигналами - это норма их разности

$$d(x, y) = \|x - y\|$$

Неравенство треугольника

$$\|x - y\| \leq \|x\| + \|y\|$$

обеспечивает Евклидовость расстояния.

1.1.2 Согласованная фильтрация

Теорема 1.1 (Неравенство Коши-Шварца) Имеет место граница

$$|\langle x, y \rangle| \leq \|x\| \|y\|$$

с равенством если и только если $y = \alpha x$.

Доказательство. Рассмотрим задачу наилучшей аппроксимации вектора y вектором αx по критерию минимальности нормы разности

$$\|y - \alpha x\|^2 = \langle y, y \rangle - \alpha^* \langle y, x \rangle - \alpha \langle x, y \rangle + \alpha \alpha^* \langle x, x \rangle \quad (1)$$

Элементарное решение задачи минимизации по вещественной и мнимой частям α дает $\alpha_{opt} = \langle y, x \rangle / \langle x, x \rangle$, что отвечает условию ортогональности вектора ошибки $y_\perp = y - \alpha x$ вектору x

$$\langle y - \alpha x, x \rangle = \langle y, x \rangle - \alpha \langle x, x \rangle = \langle y, x \rangle - \frac{\langle y, x \rangle}{\langle x, x \rangle} \langle x, x \rangle = 0$$

Подставляя найденное α_{opt} в (1), найдем

$$\|y - \alpha x\|^2 = \langle y, y \rangle - 2 \frac{|\langle x, y \rangle|^2}{\langle x, x \rangle} + \frac{|\langle x, y \rangle|^2}{\langle x, x \rangle^2} \langle x, x \rangle \geq 0,$$

откуда и вытекает неравенство \blacktriangleleft

Обобщение. Пусть имеется ортонормированный базис

$$(e_1, e_2, \dots, e_n); \langle e_i, e_i \rangle = 1, \langle e_i, e_k \rangle = 0$$

Его линейную оболочку будем рассматривать как пространство сигналов. Поставим задачу аппроксимации сигнала (вектора) y вектором $y_\parallel = \sum \alpha_i e_i$ из пространства сигналов.

Наилучшая в смысле минимума нормы разности $y_\perp = y - y_\parallel$ аппроксимация обеспечивается выполнением условия ортогональности вектора ошибки y_\perp пространству сигналов

$$\langle y_\perp, e_i \rangle = 0, i = 1, \dots, n$$

и имеет вид

$$y_\parallel = \sum \langle y, e_i \rangle e_i$$

Компонента y_\perp , ортогональная пространству сигналов, не может повлиять на решения, выносимые по результатам обработки сигнала y внутри пространства. То есть, набор скалярных произведений $\langle x, e_i \rangle$ дает полную информацию о принятом сигнале y .

Процедуру нахождения наилучшей проекции принятого сигнала в подпространство сигналов конечной размерности называют согласованной фильтрацией. Согласованная фильтрация реализуется вычислением скалярных произведений с базисными векторами. В простейшем случае одномерного пространства речь идет о единственном скалярном произведении.

1.2 Спектральная теория

1.2.1 Преобразование Фурье

Преобразование Фурье ставит в соответствие каждому сигналу $x(t)$ его амплитудный спектр

$$X(f) = \int x(t)e^{-2\pi ift} dt = \int x(t)e^{-ft} dt.$$

Преобразование обратимо

$$x(t) = \int X(f)e^{+2\pi ift} df = \int X(f)e^{+ft} df$$

Соглашение. Для лаконичности формул далее везде мы опускаем множитель $2\pi i$ в показателях и множитель 2π в аргументах тригонометрических функций. Таким образом, e^{ft} следует читать как $e^{2\pi ift}$, а $\cos ft$ как $\cos 2\pi ft$.

Сигнал $x(t)$ и его амплитудный спектр $X(f)$ разумно рассматривать как две различные формы представления одной и той же сущности. Он может быть эквивалентно представлен как во временной области, функцией $x(t)$, так и в частотной области - функцией $X(f)$. Иными словами, сигнал рассматривается здесь как пара функций

$$x(t) \leftrightarrow X(f)$$

связанных преобразованиями Фурье.

Упражнение 1 Пусть

$$\text{rect}_X(x) = \begin{cases} 1 & \text{при } x \in (-X/2, X/2) \\ 0 & \text{иначе} \end{cases}$$

симметричный прямоугольный импульс. Показать, что

$$\frac{1}{\sqrt{T}} \text{rect}_T(t) \leftrightarrow \sqrt{T} \frac{\sin \pi T f}{\pi T f} = \sqrt{T} \text{sinc}(\pi T f),$$

$$\sqrt{F} \text{sinc}(\pi F t) = \sqrt{F} \frac{\sin \pi F t}{\pi F t} \leftrightarrow \frac{1}{\sqrt{F}} \text{rect}_F(f),$$

где $FT = 1$

1.2.2 Свойства симметрии

Следующие свойства симметрии вытекают непосредственно из определений.

Фундаментальная симметрия.

Сопряжению сигнала во временной области отвечает сопряжение с отражением в частотной и наоборот

$$x^*(t) \leftrightarrow X^*(-f)$$

$$x^*(-t) \leftrightarrow X^*(f)$$

Сопряженные симметрии.

Если сигнал вещественен ($x^* = x$), то его спектр сопряженно симметричен

$$X(-f) = X^*(f).$$

Наоборот, мнимому сигналу ($x^* = -x$) отвечает сопряженно антисимметричный спектр

$$X(-f) = -X^*(f).$$

Всякий спектр $X(f)$ можно представить суммой сопряженно симметричной и сопряжено антисимметричной компонент

$$X(f) = X_s(f) + X_a(f) = \frac{X(f) + X^*(-f)}{2} + \frac{X(f) - X^*(-f)}{2}$$

и этому представлению будет отвечать разложение комплексного сигнала $x(t) = x_i(t) + ix_q(t)$ в сумму вещественной и мнимой частей

$$x_i(t) \leftrightarrow X_s(f), \quad ix_q(t) \leftrightarrow X_a(f).$$

Наоборот, разложению сигнала на сопряженно симметричную и антисимметричную составляющие во временной области

$$x(t) = x_s(t) + x_a(t) = \frac{x(t) + x^*(t)}{2} + \frac{x(t) - x^*(t)}{2}$$

отвечает разложение спектра $X(f) = X_i + iX_q$ на вещественную и мнимую части

$$x_s(t) \leftrightarrow X_i(f), \quad x_a(t) \leftrightarrow iX_q(f)$$

Двойные симметрии.

Сопряженная симметрия функции означает симметрию ее вещественной и антисимметрию мнимой частей. Сопряжено симметричный вещественный сигнал просто симметричен. Сопряжено симметричный мнимый - антисимметричен. Для сопряженно антисимметричных сигналов все наоборот. По свойству сопряженной симметрии,

1. Если сигнал $x(t)$ вещественен, то его спектр $X(f)$ сопряженно симметричен.
 2. Если сигнал $x(t)$ мним, то его спектр $X(f)$ сопряженно антисимметричен.
 3. Если сигнал $x(t)$ сопряженно симметричен, то его спектр $X(f)$ вещественен.
 4. Если сигнал $x(t)$ сопряженно антисимметричен, то его спектр $X(f)$ мним.
- Согласно (1) и (3), вещественный и симметричный сигнал имеет вещественный симметричный спектр.
 - Согласно (1) и (4), вещественный антисимметричный сигнал имеет мнимый антисимметричный спектр.
 - Согласно (2) и (3), мнимый симметричный сигнал имеет мнимый симметричный спектр.
 - Согласно (2) и (4), мнимый антисимметричный сигнал имеет вещественный антисимметричный спектр.

1.2.3 Синус и косинус преобразования

Имеем

$$X(f) = \int_{-\infty}^{+\infty} (x_s + x_a)(\cos ft - i \sin ft) dt = 2 \int_0^{+\infty} (x_s \cos ft - ix_a \sin ft) dt$$

$$x(t) = \int_{-\infty}^{+\infty} (X_s + X_a)(\cos ft + i \sin ft) df = 2 \int_0^{+\infty} (X_s \cos ft + i X_a \sin ft) dt$$

Отсюда, для случая симметричных сигналов (спектра) получаем косинус-преобразование

$$X(f) = 2 \int_0^{+\infty} x(t) \cos ft dt,$$

$$x(t) = 2 \int_0^{+\infty} X(f) \cos ft dt,$$

а для антисимметричных сигналов (спектра) - синус преобразование

$$X(f) = -2 \int_0^{+\infty} x(t) \sin ft dt,$$

$$x(t) = 2 \int_0^{+\infty} X(f) \sin ft dt.$$

1.2.4 Теормы о сдвиге

Суть утверждений этих очевидных теорем состоит в том, что сдвиг сигнала в одном из представлений эквивалентен умножению на комплексную экспоненту в другом.

$$x(t - t_0) \leftrightarrow X(f) e^{-ft_0}$$

$$x(t) e^{+f_0 t} \leftrightarrow X(f - f_0)$$

Упражнение 2 Опираясь на теорему о сдвиге, найти спектр импульса кода Манчестера (два сдвинутые прямоугольные импульса)

$$M(t) = \text{rect}_{T/2}(t - T/4) - \text{rect}_{T/2}(t + T/4)$$

Предложить алгоритм нахождения спектров серии импульсов.

1.2.5 Теорема о свертке

Сверткой сигналов $x(t)$ и $y(t)$ называют функцию

$$(x * y)(t) = \int x(u) y(t - u) du \quad (2)$$

Элементарно устанавливается факт симметрии свертки относительно перестановки координат

$$(x * y)(t) = (y * x)(t).$$

Теорема 1.2 Свертке временных форм сигналов отвечает перемножение их частотных форм и наоборот.

$$(x * y)(t) \leftrightarrow X(f)Y(f)$$

$$x(t)y(t) \leftrightarrow (X * Y)(f)$$

Доказательство. Заменим в правой части (2) сигнал $x(t)$ его представлением через амплитудный спектр

$$(x * y) = \int \int X(f)e^{fu}y(t-u) du df = \int X(f) \left\{ \int y(t-u)e^{-f(t-u)} du \right\} e^{ft} df = \int X(f)Y(f)e^{ft} df \quad \blacktriangledown$$

В частности, при $t = 0$ отсюда получаем сверточную форму тождества Парсеваля

$$\int x(t)y(-t) dt = \int X(f)Y(f) df, \quad (3)$$

используемую для определения спектров обобщенных функций.

Упражнение 3 Используя теорему о свертке, найти спектры импульсов

1. «гармоническая полуволна» (произведение прямоугольника на сумму комплексных экспонент)

$$p(t) = rect_T(t) \cos(\pi F t), \quad FT = 1$$

2. «приподнятый косинус»

$$p(t) = rect_T(t) \{1 + \cos(2\pi F t)\}, \quad FT = 1$$

Преобразование свертки - это математическая модель операции линейной фильтрации сигналов.

Линейный фильтр характеризуется импульсной реакцией $h(t)$ - откликом фильтра на импульсное воздействие $\delta(t)$. Импульсная реакция связана преобразованием Фурье с комплексным коэффициентом передачи фильтра $H(f)$.

Пусть на входе фильтра $h(t) \leftrightarrow H(f)$ действует сигнал $x(t) \leftrightarrow X(f)$. Тогда отклик фильтра $y(t) \leftrightarrow Y(f)$ может быть найден
во временной области как свертка

$$y(t) = (x * h)(t) = \int x(u)h(t-u) du,$$

или в частотной области как произведение

$$Y(f) = H(f)X(f).$$

1.2.6 Энергетические корреляции и спектры

Взаимно корреляционная функция двух сигналов $x(t)$ и $y(t)$ - это скалярное произведение $x(t)$ на сдвинутую во времени копию сигнала $y(t)$.

$$R_{xy}(t) = \langle x, y(t) \rangle = \langle x(u), y(u-t) \rangle = \int x(u)y^*(u-t) du \quad (4)$$

В случае, когда $x(t)$ и $y(t)$ - один и тот же сигнал, функция $R_x(t) = \langle x, x(t) \rangle$ называется авто корреляционной функцией сигнала $x(t)$.

Очевидно следующее свойство симметрии

$$\langle x, y(-t) \rangle = \langle y, x(t) \rangle^*$$

В частности, авто корреляционная функция комплексного сигнала сопряженно симметрична

$$\langle x, x(-t) \rangle = \langle x, x(t) \rangle^*.$$

Для вещественного сигнала это влечет симметрию авто корреляционной функции.

Из определений свертки и корреляции следует, что взаимно корреляционная функции $\langle x, y(t) \rangle$ есть свертка $x(t)$ с комплексно сопряженным и отраженным во времени сигналом $y^*(-t)$.

$$\langle x, y(t) \rangle = (x(t) * y^*(-t)).$$

По фундаментальному закону симметрии, операции сопряжения с отражением во временной области отвечает комплексное сопряжение в частотной. Поэтому, сигналу $y^*(-t)$ отвечает спектр $Y^*(f)$. С учетом этого, теорема о свертке дает следующее представление для спектра взаимно корреляционной функции.

$$\langle x, y(t) \rangle \leftrightarrow X(f)Y^*(f),$$

или в развернутой форме

$$R_{xy}(t) = \langle x, y(t) \rangle = \int x(u)y^*(u-t) du = \int X(f)Y^*(f)e^{+ft} df \quad (5)$$

$$X(f)Y^*(f) = \int R_{xy}(t)e^{-ft} dt$$

Произведение $X(f)Y^*(f)$ называют взаимным энергетическим спектром сигналов x, y . В частности, для одного сигнала получаем пару

$$R_x(t) = \langle x, x(t) \rangle \leftrightarrow |X(f)|^2$$

авто корреляционная функция \leftrightarrow энергетический спектр.

При $t = 0$ (5) дает корреляционную форму тождества Парсеваля

$$\int x(u)y^*(u) du = \int X(f)Y^*(f) df, \quad (6)$$

которая означает, что скалярное произведение сигналов инвариантно относительно переходов между временной и частотной областями - $\langle x(t), y(t) \rangle = \langle X(f), Y(f) \rangle$.

Скалярное произведение, а следовательно и метрические свойства сигналов во временной и частотной областях совпадают.

В частности, тождество Парсеваля для одного сигнала

$$\langle x(t), x(t) \rangle = \langle X(f), X(f) \rangle$$

энергия сигнала не изменяется при переходе от временной формы представления к частотной.

Упражнение 4 Найти аэто корреляционную функцию и энергетический спектр прямоугольного импульса $\text{rect}_T(t)$.

Упражнение 5 Найти аэто корреляционную функцию и энергетический спектр импульса Найквиста $\text{sinc}(\pi F t) = \frac{\sin(\pi F t)}{\pi F t}$.

Упражнение 6 Найти аэто корреляционные функции и энергетические спектры импульсов из упражнения 3.

1.2.7 Обобщенные функции и их спектры

Опираясь на тождество Парсеваля

$$\int x(t)y(-t) dt = \int X(f)Y(f) df,$$

введем понятие спектра обобщенной функции.

Пусть $g(t)$ - основная функция с конечной энергией, $G(f)$ - ее амплитудный спектр. Спектром обобщенной функции $x(t)$, определенной своим действием на основные функции по правилу

$$\int x(t)g(-t)dt,$$

назовем обобщенную функцию $X(f)$, для которой

$$\int x(t)g(-t)dt = \int X(f)G(f) df.$$

Спектр симметричной единицы, δ -функция

Пусть $x(t) = 1(t) = 1$ - тождественная единичная функция. Имеем

$$\int 1 g(-t)dt = G(0) = \int X(f)G(f) df.$$

Следовательно, $X(f) = \delta(f)$. Таким образом,

$$\begin{aligned} 1(t) &\leftrightarrow \delta(f), \\ \delta(t) &\leftrightarrow 1(f). \end{aligned}$$

Подключив теорему о сдвиге, находим

$$\begin{aligned} e^{+f_0 t} &\leftrightarrow \delta(f - f_0), \\ \delta(t - \tau) &\leftrightarrow e^{-f_0 \tau}. \end{aligned}$$

Далее,

$$\begin{aligned} \cos(f_0 t) &= \frac{e^{+f_0 t} + e^{-f_0 t}}{2} \leftrightarrow \frac{1}{2}\delta(f - f_0) + \frac{1}{2}\delta(f + f_0), \\ \sin(f_0 t) &= \frac{e^{+f_0 t} - e^{-f_0 t}}{2i} \leftrightarrow \frac{1}{2i}\delta(f - f_0) - \frac{1}{2i}\delta(f + f_0), \end{aligned}$$

Упражнение 7 Показать, что $(x(t) * \delta(t - \tau)) = x(t - \tau)$. Проверить выполнение этого тождества в частотной области.

Спектр антисимметричной единицы, главное значение

Рассмотрим антисимметричную единичную функцию

$$sign(t) = \begin{cases} 1 & \text{при } t > 0 \\ 0 & \text{при } t = 0 \\ -1 & \text{при } t < 0 \end{cases}$$

Ее обобщенный спектр может быть найден предельным переходом

$$\int sign(t)e^{-ft} df = \lim_{\alpha \rightarrow 0} \int sign(t)e^{-\alpha|t|} e^{-ft} df = -\frac{i}{\pi f}.$$

Он выражается через обобщенную функцию «главное значение» - $1/x$. Аналогичным образом можно найти временную форму для антисимметричной единицы в частотной области $sign(f)$. В результате получаем пары обобщенных Фурье преобразований

$$sign(t) \leftrightarrow -\frac{i}{\pi f},$$

$$\frac{i}{\pi t} \leftrightarrow sign(f).$$

Упражнение 8 Найти спектр функции

$$sign(t+T/2) - sign(t-T/2).$$

Упражнение 9 Показать, что для функции «единичная ступень»

$$\theta(t) = \begin{cases} 1 & \text{при } t > 0 \\ 0 & \text{при } t \leq 0 \end{cases} = \frac{1}{2} \{1 + sign(t)\}$$

справедливо

$$\begin{aligned} \theta(t) &\leftrightarrow \frac{1}{2}(\delta(f) - i\frac{1}{\pi f}), \\ \theta(f) &\leftrightarrow \frac{1}{2}(\delta(t) + i\frac{1}{\pi t}). \end{aligned}$$

Упражнение 10 Проверить выполнение следующих теорем о свертке обобщенных функций

$$\begin{aligned} (\delta(t) * \delta(t))(\tau) &= \delta(\tau) \\ (\delta(t) * \frac{1}{\pi t})(\tau) &= \frac{1}{\pi t} \\ (\frac{1}{\pi t} * \frac{1}{\pi t}) &= -\delta(t)\tau \end{aligned}$$

Теорема о двух гребенках

Рассмотрим функцию «гребенка», получаемую в результате трансляции δ -функции вдоль оси времени (частоты) с фиксированным шагом T (F)

$$\chi(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT).$$

Эта функция дает простое математическое описание для

- операции равномерной дискретизации $g(t)$ - взятия последовательности выборок $g(nT)$

$$g(t) \Rightarrow g(t) \sum_{n=-\infty}^{\infty} \delta(t - nT) = \sum_{n=-\infty}^{\infty} g(nT) \delta(t - nT)$$

- операции периодического продолжения $g(t)$ вдоль оси времени

$$g(t) \Rightarrow (g(t) * \sum_{n=-\infty}^{\infty} \delta(t - nT)) = \sum_{n=-\infty}^{\infty} g(t - nT)$$

Следующая теорема устанавливает связь между операциями дискретизации и периодического продолжения, имеющую принципиальное значение для приложений, связанных с цифровой обработкой сигналов.

Теорема 1.3 (О двух гребенках) *Фурье образом гребенки с шагом T во временной области является гребенка с шагом F в частотной области и наоборот ($FT = 1$).*

$$\sum_n \delta(t - nT) \leftrightarrow F \sum_m \delta(f - mF).$$

Доказательство. Обобщенный амплитудный спектр суммы в левой части имеет вид

$$\int \left(\sum_n \delta(t - nT) \right) e^{-nft} dt = \sum_n e^{-nTf}.$$

Сумму в правой части будем интерпретировать как периодическую функцию с периодом F и разложим ее в ряд Фурье по системе ортогональных на интервале $(-F/2, F/2)$ функций e^{-kTf}

$$\sum_m \delta(f - mF) = \sum_k c_k e^{-kTf}$$

Коэффициенты Фурье c_k этого ряда

$$Fc_k = \int_{-T/2}^{T/2} \delta(t) e^{kTf} df = 1$$

одинаковы и равны $1/F$, что и доказывает тождество \blacksquare .

По теореме о свертке, перемножению функций во временной области отвечает их свертка в частотной и наоборот. Но, как показано выше, умножение на гребенку описывает дискретизацию, а свертка с ней - периодическое продолжение. Таким образом, результат теоремы приводит к выводам:

- Дискретизации сигнала $x(t)$ во времени с шагом T в частной области отвечает периодическое продолжение спектра $X(f)$ с периодом F , $FT = 1$
- Периодическому продолжению сигнала $x(t)$ во времени с периодом T в частной области отвечает дискретизация спектра $X(f)$ с шагом F , $FT = 1$

Критерий Найквиста обратимости дискретизации

Пусть $x(t)$ - сигнал с амплитудным спектром $X(f)$, финитном на носителе $(-F/2, F/2)$. Осуществим его дискретизацию с шагом T . Спектром дискретизованного сигнала

$$\sum_n x(nT) \delta(t - nT)$$

станет периодическое продолжение $X(f)$ с периодом $F = 1/T$

$$\sum_m X(f - mF).$$

Но, в силу финитности спектра $X(f)$, его транслированные на F реплики не перекрываются на оси частот. Поэтому, исходный спектр $X(f)$ может быть однозначно восстановлен по продолженному. Следовательно, и сигнал $x(t)$ может быть однозначно восстановлен по последовательности своих выборочных значений $x(nT)$.

Теорема 1.4 (Теорема Найквиста о выборках) Всякий сигнал со спектром, сосредоточенным в полосе $(-F/2, F/2)$, без потери информации представляется последовательностью выборочных значений $x(nT)$, взятых с шагом дискретизации $T \leq 1/F$ (или с частотой дискретизации $F_d \geq F$).

Если сигнал вещественен, то его сопряженно симметричный спектр вполне определяется компонентной на положительных частотах с верхней частотой $F_h = F/2$. Тогда минимальная частота дискретизации составляет $F_d = 2F_h$.

Упражнение 11 Пусть спектр сигнала $x(t)$ сосредоточен в полосе $(-F/2, F/2)$. Рассмотрим дискретизованный сигнал

$$x_d(t) = \sum_n x(nT)\delta(t - nT), \quad T \leq 1/F.$$

Найти выражение для сигнала $y(t)$, получающегося в результате фильтрации $x_d(t)$ фильтром с комплексным коэффициентом передачи $H(f) = rect_F(f)$. (Ряд Котельникова).

Упражнение 12 Доказать следующую теорему об интерполяции.

Пусть дана последовательность выборок x_n , конечная или бесконечная.

Найдется интерполирующая функция $x(t)$, такая что $x(nT) = x_n$, а спектр $X(f)$ - финитен в полосе $(-F/2, F/2)$, $F \geq 1/T$.

Намек - одной из интерполирующих функций является ряд Котельникова.

Спектр периодического сигнала

Пусть $x(t)$ - периодический сигнал с периодом T . Умножением на прямоугольный импульс $rect_T(t)$ «вырежем» его реализацию $x_T(t) = x(t)rect_T(t)$ и пусть $X_T(f)$ - амплитудный спектр реализации. Свертка с гребенкой во временной области обеспечивает восстановление $x(t)$ периодическим продолжением реализации $x_T(t)$

$$x(t) = (x_T(t) * \sum \delta(t - nT))$$

В частотной области этому отвечает δ -дискретизация спектра $X_T(f)$. Учитывая это, для спектра периодического сигнала $x(t)$ находим

$$X(f) = F \sum X_T(mF)\delta(f - mF), \quad FT = 1.$$

Таким образом, спектр периодического сигнала имеет вид «гребенки» δ -функций с весами $X_T(mF)$, равными выборочным значениям амплитудного спектра реализации на периоде. Сигнал однозначно восстанавливается по выборкам спектра с шагом $F = 1/T$.

2 Пространство комплексных огибающих

2.1 Фильтр Гильберта для вещественных сигналов

Фильтр с равномерным на всех частотах коэффициентом передачи $1(f) = 1$ назовем тривиальным. Его импульсная реакция $h(t) = \delta(t)$.

Рассмотрим двойственный фильтр с комплексным коэффициентом передачи $\frac{\text{sign}(f)}{i}$. Мнимая единица в знаменателе обеспечивает сопряженную симметричность $\frac{\text{sign}(f)}{i}$, так что импульсная реакция этого фильтра вещественна. Она определяется известным обобщенным Фурье преобразованием функции $\text{sign}(f)$ и имеет вид

$$h(t) = \frac{1}{\pi t}.$$

Этот фильтр называют фильтром Гильберта, а реализуемое им преобразование сигналов

$$\tilde{x}(t) = \left(x(t) * \frac{1}{\pi t} \right) = \frac{1}{\pi} \int \frac{x(u)}{t-u} du$$

преобразованием Гильберта. То же преобразование в частотной области принимает особенно простой вид

$$\tilde{X}(f) = X(f) \frac{\text{sign}(f)}{i}$$

Повторное применение преобразования

$$\tilde{\tilde{X}}(f) = X(f) \frac{\text{sign}(f)}{i} \frac{\text{sign}(f)}{i} = -X(f), \quad \tilde{\tilde{x}}(t) = -x(t)$$

дает исходный сигнал $x(t)$ с обратным знаком. В этом смысле преобразование Гильберта обратимо.

Теорема 2.1 *Преобразование Гильберта сохраняет скалярное произведение, в том числе - метрику.*

$$\langle x, y \rangle = \langle \tilde{x}, \tilde{y} \rangle.$$

Доказательство. Проводится в одну строку переходом в частотную область:

$$\langle \tilde{x}, \tilde{y} \rangle \rightarrow X(f) \left(\frac{\text{sign}(f)}{i} \right) Y^*(f) \left(\frac{\text{sign}(f)}{i} \right)^* = X(f) Y^*(f) \rightarrow \langle x, y \rangle \blacksquare$$

Теорема 2.2 *Имеет место следующее свойство антисимметрии*

$$\langle \tilde{x}, y \rangle = -\langle x, \tilde{y} \rangle.$$

Доказательство. Так же в одну строку:

$$\langle \tilde{x}, y \rangle \rightarrow [X(f) \frac{\text{sign}(f)}{i}] Y^*(f) = -X(f) [Y^*(f) (\frac{\text{sign}(f)}{i})^*] \rightarrow -\langle x, \tilde{y} \rangle \blacksquare$$

Следствие 1 *Сигнал $x(t)$ и его преобразование Гильберта $\tilde{x}(t)$ ортогональны.*

$$\langle \tilde{x}, x \rangle = \langle x, \tilde{x} \rangle = 0.$$

Доказательство. По свойству антисимметрии, $\langle \tilde{x}, x \rangle = -\langle x, \tilde{x} \rangle$. Но, с другой стороны, скалярное произведение вещественных сигналов симметрично, и поэтому $\langle \tilde{x}, x \rangle = \langle x, \tilde{x} \rangle$. Отсюда вытекает равенство нулю \square

Теорема 2.3 (Гильбертовы преобразования гармонических сигналов) Имеем,

$$\widetilde{\cos}(ft) = \sin(ft)$$

$$\widetilde{\sin}(ft) = -\cos(ft)$$

Доказательство. В самом деле,

$$\widetilde{\cos}(f_0t) \rightarrow \frac{1}{2}(\delta(f - f_0) + \delta(f + f_0)) \frac{\text{sign}(f)}{i} = \frac{1}{2i}(\delta(f - f_0) - \delta(f + f_0)) \rightarrow \sin(f_0t)$$

Второе аналогично \square

Теорема 2.4 Пусть $x(t)$ - сигнал со спектром, финитном на $(-f_0, f_0)$. Тогда

$$(x(t)\widetilde{\cos}(f_0t)) = x(t)\sin(f_0t)$$

$$(x(t)\widetilde{\sin}(f_0t)) = -x(t)\cos(f_0t)$$

Доказательство. Начнем с первого тождества. Сравним спектральные представления его левой и правой частей.

$$\begin{aligned} x(t)\widetilde{\cos}(f_0t) &\rightarrow \frac{1}{2}(X(f - f_0) + X(f + f_0)) \frac{\text{sign}(f)}{i} = & \frac{1}{2i}(X(f - f_0) - X(f + f_0)) \\ x(t)\sin(f_0t) &\rightarrow & \frac{1}{2i}(X(f - f_0) - X(f + f_0)) \end{aligned}$$

Совпадение служит доказательством. Второе тождество доказывается аналогичным образом \blacktriangledown

Теорема 2.5 (Ортогональность квадратурных каналов) Пусть $x(t), y(t)$ - два различные сигнала со спектром, финитном на носителе $(-f_0, f_0)$. Тогда сигналы $x(t)\cos(f_0t)$ и $y(t)\sin(f_0t)$ - ортогональны.

Замечание. Если $x(t) = y(t)$, то ортогональность является следствием того, что $x(t)\sin(f_0t)$ есть преобразование Гильберта от $x(t)\cos(f_0t)$. В общем случае это не так, однако ортогональность все же имеет место.

Доказательство. В частотной области скалярное произведение $\langle x(t)\cos(f_0t), y(t)\sin(f_0t) \rangle$ представляется в виде

$$\frac{1}{4} \int [X(f - f_0) + X(f + f_0)] \left[(Y^*(f - f_0) + Y^*(f + f_0)) \frac{\text{sinc}(f)}{i} \right] df$$

Учитывая, что спектры, сдвинутые вправо ($X(f - f_0)$) и влево ($Y(f + f_0)$) не перекрываются, приведем это к виду

$$\frac{1}{4} \int [X(f - f_0)Y^*(f - f_0) + X(f + f_0)Y^*(f + f_0)] \frac{\text{sinc}(f)}{i} df,$$

что равно нулю ввиду антисимметрии подынтегральной функции ∇

В целом, складывается следующая картина.

Пространство сигналов удобно представлять себе расщепленным на два линейные подпространства - **две квадратурные компоненты**. Существует **косинусное** подпространство с сигналами вида $x(t)\cos f_0 t$ и **синусное** подпространство с сигналами $x(t)\sin f_0 t$. Они взаимно ортогональны. Гильбертово преобразование реализует линейное отображение одной квадратурной компоненты на другую с сохранением скалярного произведения.

Сигнал общего вида может быть суммой частей, лежащих в разных квадратурных компонентах. Далее рассматриваются аналитические сигналы, у которых эти части связаны по Гильберту.

2.2 Аналитические сигналы

Пусть $x(t) \leftrightarrow X(f)$ - вещественный сигнал, $\tilde{x}(t) \leftrightarrow X(f) \frac{\text{sign}(f)}{i}$ - его гильбертово преобразование.

Аналитический сигнал $x^+(t)$ определим как

$$x^+(t) = x(t) + i\tilde{x}(t)$$

В спектральной области имеем

$$X^+(f) = X(f) + iX(f) \frac{\text{sign}(f)}{i} = X(f)(1 + \text{sign}(f)) = 2X(f)\theta(f)$$

Таким образом, амплитудный спектр аналитического сигнала тождественно равен нулю на отрицательных частотах, и совпадает с удвоенным спектром исходного сигнала на положительных. Ясно, что ненулевой аналитический сигнал не может быть вещественным, поскольку спектр вещественного сигнала обязан быть сопряжено симметричным.

По двойственности вводится анти аналитический сигнал

$$x^-(t) = x(t) - i\tilde{x}(t) = (x^+)^*$$

с нулевым спектром на положительных частотах.

Каждый вещественный сигнал $x(t)$ допускает, таким образом, разложение в сумму

$$x(t) = x^+(t) + x^-(t)$$

аналитического и анти аналитического сигналов. Первый содержит компоненты сигнала на положительных частотах, а второй - на отрицательных.

Пример 1 Пусть $x(t) = \cos(ft)$. Тогда

$$x^+(t) = \cos(ft) + i\sin(ft) = e^{+ft}$$

$$x^-(t) = \cos(ft) - i\sin(ft) = e^{-ft}$$

Таким образом, аналитический гармонический сигнал, это и есть комплексная экспонента, используемая в методе комплексных амплитуд. Концепция аналитического сигнала обобщает метод комплексных амплитуд на сигналы произвольной формы.

Упражнение 13 Пусть $x(t) = \sin(ft)$. Показать, что $x^+(t) = i e^{+ft}$, $x^-(t) = -i e^{-ft}$.

Представив аналитический сигнал через модуль и аргумент

$$x^+(t) = |x^+(t)|e^{i \arg(x^+(t))}$$

получаем общепринятые формальные определения для
мгновенной амплитуды

$$A(t) = |x^+(t)| = \sqrt{\tilde{x}^2(t) + x^2(t)}$$

мгновенной фазы

$$\varphi(t) = \arg(x^+(t)) = \arctg(\tilde{x}(t)/x(t))$$

и мгновенной частоты

$$\varpi(t) = \frac{d\varphi}{dt}.$$

вещественного сигнала $x(t)$ произвольной формы.

2.3 Комплексная огибающая

Под вещественным сигналом (*Passband signal*) будем понимать сигнал $x_p(t)$, спектр которого сосредоточена на $(-2f_0, 0) \cup (0, 2f_0)$. Вещественное линейное пространство таких сигналов будем называть Р-пространством (Р-областью, *Passband space*). Это наши физические сигналы.

Зафиксируем несущую частоту f_0 - точку на частотной оси. Поставим задачу переноса спектрального диапазона Р-сигналов влево на f_0 с тем, чтобы совместить точку несущей частоты с нулем. Чтобы решить проблему с отрицательными частотами, перейдем к аналитическому сигналу

$$x_p^+(t) = x_p(t) + i \tilde{x}_p(t)$$

Умножением на комплексную экспоненту перенесем односторонний спектр аналитического сигнала влево на f_0

$$x(t) = \frac{1}{\sqrt{2}} x_p^+(t) e^{-f_0 t} \quad (7)$$

В результате получим то, что называют комплексной огибающей (*Baseband signal*) реального сигнала $x_p(t)$. Амплитудный спектр комплексной огибающей $x(t)$ сосредоточен в окрестности нулевой частоты на интервале $(-f_0, f_0)$.

Линейное пространство комплексных огибающих будем называть В-пространством (В-областью, *Baseband Space*). В нем лежат виртуальные комплексные сигналы - аналоги используемых в методе комплексных амплитуд.

Алгоритм (7) определяет **РВ-преобразование** физических сигналов $x_p(t)$ в комплексные огибающие $x(t)$. Это вещественно-линейное отображение вещественного Р-пространства в комплексное В-пространство.

Выделим вещественную и мнимую части комплексной огибающей

$$x(t) = x_i(t) + i x_q(t)$$

Прослеживая в (7) обратный переход от $x(t)$ к $x_p(t)$, приходим к следующей форме **ВР-преобразования** из пространства комплексных огибающих в пространство реальных сигналов

$$x_p(t) = \sqrt{2} \operatorname{Re}(x(t) e^{+f_0 t}) = \sqrt{2} \{x_i(t) \cos f_0 t - x_q(t) \sin f_0 t\} \quad (8)$$

Формула (8) дает каноническое представление реального Р-сигнала в виде суммы квадратурных компонент - синфазной (inphase) $x_i(t) \cos f_0 t$ и квадратурной (quadrature) $x_q(t) \sin f_0 t$. Квадратурные компоненты ортогональны в Р-области. В В-области им отвечают вещественная и мнимая части комплексной огибающей $x(t)$.

Подведем итоги. Фиксация несущей делит Р-пространство реальных сигналов на два квадратурные подпространства - косинусное и синусное. Квадратурные подпространства взаимно ортогональны и связаны одно с другим преобразованием Гильберта.

В области комплексных огибающих квадратурным подпространствам отвечают подпространства вещественных и мнимых сигналов. Они связаны умножением на мнимую единицу. Это аналог преобразования Гильберта для В-области.

Преобразования (7) и (8) осуществляют вещественно-линейные отображения Р-области в В-область и наоборот с сохранением разбиения на квадратурные подпространства.

Упражнение 14 Показать, что комплексная огибающая сигнала

$$x_p(t) = A \cos(f_0 t) + B \cos(f_0 + \Delta f) t$$

относительно несущей f_0 имеет вид $x(t) = A + B e^{\Delta f t}$. Найти комплексную огибающую сигнала

$$x_p(t) = A \cos(f_0 + F) t + B \cos(f_0 - F) t$$

Временные сдвиги сигналов в В и Р областях

Рассмотрим сдвинутый во времени вещественный сигнал

$$x'_p(t) = \sqrt{2} (x_i(t - \tau) \cos f_0(t - \tau) - x_q(t - \tau) \sin f_0(t - \tau))$$

Последовательно находя его Гильбертово преобразование

$$\tilde{x}'_p(t) = \sqrt{2} (x_i(t - \tau) \sin f_0(t - \tau) + x_q(t - \tau) \cos f_0(t - \tau)),$$

аналитический сигнал

$$x'^+_p(t) = \sqrt{2} (x_i(t - \tau) e^{f_0(t-\tau)} + i x_q(t - \tau) e^{f_0(t-\tau)}),$$

и комплексную огибающую

$$x'(t) = \frac{1}{\sqrt{2}} x'^+_p(t) e^{-f_0 t} = (x_i(t - \tau) + i x_q(t - \tau)) e^{-f_0 \tau},$$

придем к

$$x'(t) = x(t - \tau) e^{-i\theta}, \quad \theta = 2\pi f_0 \tau.$$

Таким образом, временному сдвигу Р-сигнала в В-области отвечает

- временной сдвиг комплексной огибающей и
- поворот вектора комплексной огибающей на угол $\theta = 2\pi f_0 \tau$, вызванный изменением фазы несущей.

В частности, ошибка в фазе несущей на θ приводит к умножению комплексной огибающей на $e^{-i\theta}$, что эквивалентно ее повороту на угол θ .

Различают два варианта обработки сигнала на стороне приема - когерентный и некогерентный.

Когерентный прием предполагает, что фаза несущей на принимающей стороне известна. Это требует применения систем фазовой синхронизации между передающей и принимающей сторонами.

Формально, задача фазовой синхронизации эквивалента синхронизации шкал времени между передатчиком и приемником. Практически же она решается выделением несущей принимаемого сигнала с помощью петель фазовой авто подстройки (ФАП).

Некогерентный прием не предполагает фазовой синхронизации. При этом могут использоваться только такие методы обработки, которые не зависят от поворота комплексной огибающей на случайный фазовый угол.

2.4 Реализация ВР/РВ преобразований

Преобразования ВР/РВ составляют теоретический фундамент анализа и синтеза сигналов для всевозможных радиоэлектронных систем. Подлежащие передаче сигналы формируются в области комплексных огибающих. Для передачи сформированный комплексный сигнал трансформируют в вещественный эквивалент с конкретной несущей частотой с помощью РВ-преобразования (7). Принятый сигнал, в первую очередь, освобождают от несущей, преобразуя его в комплексную огибающую ВР-преобразованием (8). Это практика позволяет решать задачи формирования и обработки сигналов в области комплексных огибающих безотносительно к конкретному выбору несущей.

Для формирования и обработки сигналов широко применяются цифровые методы, так что сфера применения аналоговых методов в современных радиоэлектронных системах сужается до задач реализации РВ/ВР-преобразований.

Преобразование комплексной огибающей $x(t) = x_i(t) + i x_q(t)$ в вещественный радиосигнал осуществляется квадратурным модулятором, рис. 1.

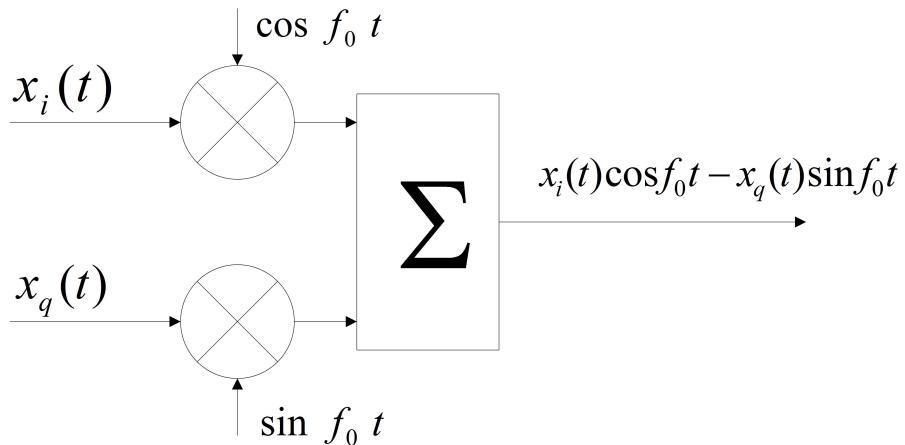


Рис. 1: Квадратурный модулятор.

Структура модулятора не вызывает вопросов. Это устройство, которое реализует ВР-преобразование (8). Технически он содержит два аналоговых перемножителя вещественной и мнимой компонент комплексной огибающей на квадратурные несущие колебания и сумматор.

Полезна следующая точка зрения на процесс модуляции. Комплексный информационный сигнал $x(t)$ следует представлять себе как траекторию движения некоторой сигнальной точки в двухмерной комплексной плоскости - плоскости значений комплексной огибающей. Проекции этой траектории на оси координат и подаются на

входы модулятора. Умножением на комплексную экспоненту $e^{f_0 t}$ модулятор осуществляет переход в быстро вращающуюся систему координат. В этой системе траектория движения сигнальной точки принимает вид

$$\begin{aligned} x(t)e^{-f_0 t} &= \{x_i(t)\cos f_0 t - x_q(t)\sin f_0 t\} \\ &= i\{x_i(t)\sin f_0 t + x_q(t)\cos f_0 t\} \end{aligned}$$

Затем модулятор «забывает» о проекции траектории на мнимую ось, оставляя для передачи только вещественную проекцию. Отбрасывая мнимую проекцию, демодулятор, по своему прав. Дело в том, что проекция на мнимую ось является гильбертовым преобразованием от вещественной проекции. Поэтому ее передача по каналу была бы избыточной. Однако это создает проблему при приеме. Чтобы вернуться в неподвижную систему координат, необходимо восстановить проекцию на мнимую ось. Этим и занимается гильбертов фильтр демодулятора.

Алгоритм (7) формирования комплексной огибающей по принятому сигналу реализуется демодулятором Гильbertа, рис. 2. В состав демодулятора входит фильтр Гильbertа, который восстанавливает утраченную при передаче проекцию на мнимую ось. Принятая вещественная $x_p(t)$ и восстановленная мнимая $\tilde{x}_p(t)$ проекции поступают далее на входы комплексного перемножителя. Здесь, в результате умножения на комплексную несущую $e^{-f_0 t}$, осуществляется переход из вращающейся систему координат в неподвижную. Комплексная огибающая восстановлена.

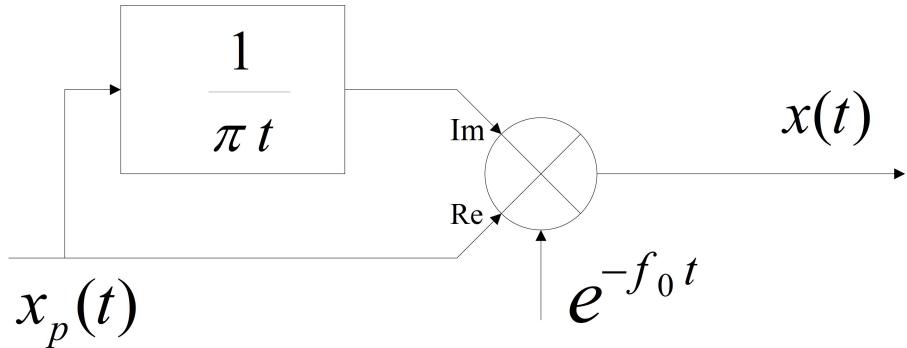


Рис. 2: Демодулятор Гильберта.

Демодулятор Гильберта - достаточно сложное устройство. В его состав, помимо собственно гильбертова фильтра, входит комплексный перемножитель, который включает в себя четыре вещественные перемножители. Существует гораздо более простая схема преобразования Р-сигнала в В-область. Это так называемый квадратурный демодулятор, рис. 3.

Он содержит всего два вещественные перемножителя и два фильтра нижних частот LPF с граничной частотой f_0 .

Рассмотрим работу верхнего (косинусного) плеча этого демодулятора в процессе обработки сигнала

$$x_p(t) = x_i(t)\cos f_0 t - x_q(t)\sin f_0 t.$$

После умножения на косинус получается

$$\begin{aligned} x_p(t)\cos f_0 t &= x_i(t)\cos^2 f_0 t - x_q(t)\sin f_0 t \cos f_0 t \\ &= \frac{1}{2}\{x_i(t) + x_i(t)\cos 2f_0 t - x_q(t)\sin 2f_0 t\}. \end{aligned}$$

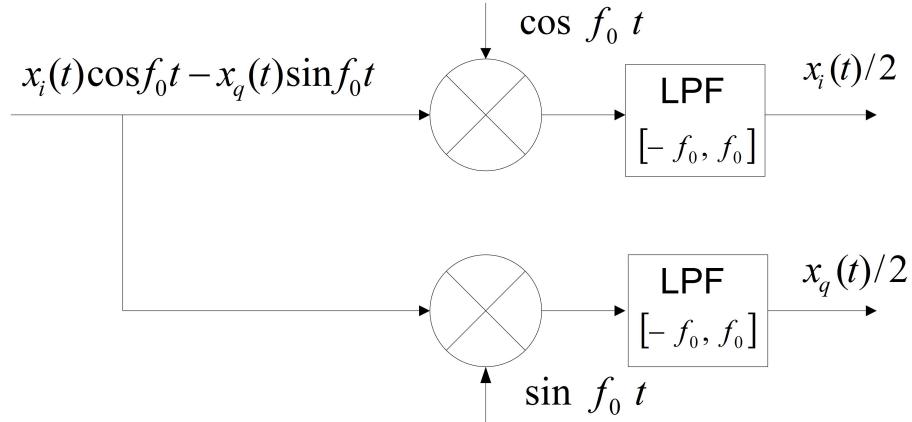


Рис. 3: Квадратурный демодулятор.

Назначение фильтра низких частот состоит в том, чтобы отсечь высокочастотную часть сигнала на удвоенной частоте несущей. После фильтрации остается оценка вещественной части комплексной огибающей.

Упражнение 15 Проанализировать работу синусного плеча квадратурного демодулятора.

Упражнение 16 Проанализировать работу квадратурного демодулятора при наличии ошибки θ в фазе опорной несущей.

Теорема 2.6 Квадратурный демодулятор эквивалентен демодулятору Гильберта.

Доказательство: Заметим, что входящий в квадратурный демодулятор полосовой фильтр

$$rect_{f_0}(f) = \begin{cases} 1, & f \in (-f_0, f_0) \\ 0, & f \notin (-f_0, f_0) \end{cases}$$

допускает представление

$$rect_{f_0}(f) = \frac{1}{2}(sign(f + f_0) - sign(f - f_0))$$

Рассмотрим косинусный канал демодулятора - умножение радиосигнала $x_p(t)$ на косинус + полосовая фильтрация. В частотной области результат этой обработки представляется в виде

$$X(f) = \frac{1}{4} \{(X_p(f + f_0) + X_p(f - f_0)) (sign(f + f_0) - sign(f - f_0))\}$$

Учитывая локализацию спектра $X_p(f)$ в полосах $(-2f_0, 0)$, $(0, +2f_0)$, найдем

$$\begin{aligned} -X_p(f + f_0)sign(f - f_0) &= +X_p(f + f_0), \\ X_p(f - f_0)sign(f + f_0) &= +X_p(f - f_0), \\ X_p(f + f_0)sign(f + f_0) &= i\tilde{X}_p(f + f_0), \\ -X_p(f - f_0)sign(f - f_0) &= -i\tilde{X}_p(f - f_0). \end{aligned}$$

Таким образом,

$$X(f) = \frac{1}{4} \{(X_p + i\tilde{X}_p)(f + f_0) + (X_p - i\tilde{X}_p)(f - f_0)\}$$

Переходя во временную область, получим

$$x(t) = \frac{1}{4} \{x_p^+(t)e^{-f_0 t} + x_p^-(t)e^{+f_0 t}\} = \frac{1}{2} \operatorname{Re}(x_p^+(t)e^{-f_0 t}) = \frac{1}{\sqrt{2}} x_i(t)$$

Таким образом, сигнал на выходе косинусного канала, с точностью до множителя $1/\sqrt{2}$ действительно совпадает с вещественной частью комплексной огибающей \blacktriangledown

Упражнение 17 Завершить доказательство теоремы, проанализировав случай синусного канала. Показать, что этот канал формирует $1/\sqrt{2} x_q(t)$.

2.5 Свертки и корреляции в Р и В областях

Обсудим соотношения между свертками/корреляциями в Р и В областях. Вначале важный предварительный результат.

Лемма 1 (Об умножении на экспоненту) Пусть $a(t)$, $b(t)$ - сигналы с конечной энергией. Рассмотрим результаты их умножения на комплексную экспоненту - $\hat{a}(t) = a(t)e^{-ft}$, $\hat{b}(t) = b(t)e^{-ft}$. Для свертки и корреляции умноженных на экспоненту сигналов справедливы соотношения

$$\begin{aligned} (\hat{a}(t) * \hat{b}(t)) &= e^{-ft} (a(t) * b(t)), \\ \langle \hat{a}(t), \hat{b}(t) \rangle &= e^{-ft} \langle a(t), b(t) \rangle. \end{aligned}$$

Словами - свертка(корреляция) умноженных на экспоненту сигналов равна умноженной на ту же экспоненту свертке(корреляции).

Доказательство. Простой проверкой, исходя из определений свертки и корреляции.

$$\begin{aligned} (a(u)e^{-fu} * b(u)e^{-fu})(t) &= \int a(u)e^{-fu} b(t-u)e^{-f(t-u)} = e^{-ft} (a * b) \\ \langle a(u)e^{-fu}, b(u-t)e^{-f(u-t)} \rangle &= \int a(u)e^{-fu} b^*(t-u)e^{+f(u-t)} = e^{-ft} \langle a, b(t) \rangle \end{aligned}$$

Еще проще заметить, что в частотной области утверждения леммы сводятся к тому, что произведение сдвинутых спектров равно сдвинутому произведению.

Теорема 2.7 (О свертке) Свертке комплексных огибающих отвечает свертка реальных сигналов (с точностью до нормировки).

$$\begin{array}{ccc} x_p(t) & h_p(t) & y_p(t) = (x_p * h_p) \\ \Updownarrow & \Updownarrow & \Updownarrow \\ x(t) & h(t) & y(t) = \frac{1}{\sqrt{2}}(x * h) \end{array} \quad \text{BP/PB - Преобразования}$$

Доказательство. Пусть $x_p(t)$ - реальный сигнал, $h_p(t)$ - импульсная реакция радиочастотного фильтра, $y_p(t) = (x_p * h_p)$ - результат фильтрации. Введем комплексные огибающие названных функций

$$\begin{aligned}\sqrt{2}x(t) &= (x_p + i\tilde{x}_p)e^{-f_0 t} \\ \sqrt{2}h(t) &= (h_p + i\tilde{h}_p)e^{-f_0 t} \\ \sqrt{2}y(t) &= (y_p + i\tilde{y}_p)e^{-f_0 t}\end{aligned}$$

Вычислим свертку в комплексной области

$$2(\mathbf{x}^*\mathbf{h}) = \left((x_p + i\tilde{x}_p)e^{-f_0 t} * (h_p + i\tilde{h}_p)e^{-f_0 t} \right)$$

Согласно лемме 1,

$$\begin{aligned}2(\mathbf{x}^*\mathbf{h}) &= e^{-f_0 t} \left((x_p + i\tilde{x}_p) * (h_p + i\tilde{h}_p) \right) = \\ &= e^{-f_0 t} \left[(x_p * h_p) - (\tilde{x}_p * \tilde{h}_p) + i(x_p * \tilde{h}_p) + i(\tilde{x}_p * h_p) \right]\end{aligned}$$

Далее

$$\begin{aligned}(\tilde{x}_p * \tilde{h}_p) &\rightarrow X_p(f)Y_p(f) \left[\frac{\text{sign}(f)}{i} \right]^2 \rightarrow -(x_p * h_p) \\ (\tilde{x}_p * h_p) &= (x_p * \tilde{h}_p) \rightarrow X_p(f)Y_p(f) \frac{\text{sign}(f)}{i} \rightarrow \widetilde{(x_p * h_p)}\end{aligned}$$

Учитывая это, найдем

$$2(\mathbf{x}^*\mathbf{h}) = 2e^{-f_0 t} (y_p + i\tilde{y}_p) = \sqrt{2}\mathbf{y}$$

или

$$\left(x * \frac{h}{\sqrt{2}} \right) = y. \quad \blacksquare$$

Этот результат сводит задачу радиочастотной фильтрации к фильтрации в области комплексных огибающих. Если словами, то его смысл состоит в следующем.

Вместо того, чтобы применять фильтр $h_p(t)$ к сигналу $x_p(t)$ на частоте несущей, а затем преобразовывать результат фильтрации $y_p(t)$ в комплексную огибающую, можно сразу найти комплексную огибающую сигнала $x_p(t)$, а затем выполнить ее эквивалентную фильтрацию в области огибающих. Импульсная реакция эквивалентного комплексного фильтра $h(t)$ есть результат РВ-преобразования $h_p(t)$, деленный на нормирующий множитель $\frac{1}{\sqrt{2}}$.

Таким образом, для каждого радиочастотного фильтра существует эквивалентный фильтр нижних частот в области комплексных огибающих.

Упражнение 18 Показать, что комплексным аналогом полосового фильтра с коэффициентом передачи

$$H_p(f) = \frac{1}{1 + i2Q \frac{f-f_0}{f_0}}$$

(колебательный контур) является фильтр нижних частот

$$H(f) = \frac{1}{1 + i\tau f}, \quad \tau = \frac{2Q}{f_0}$$

Теорема 2.8 (О корреляции) Взаимные корреляционные функции в Р и В областях связаны соотношением:

$$\langle x, y(t) \rangle = (\langle x_p, y_p(t) \rangle + i \widetilde{\langle x_p, y_p(t) \rangle}) e^{-f_0 t} \quad (9)$$

Доказательство. В тех же обозначениях, что и выше, найдем

$$2 \langle x, y(t) \rangle = \langle (x_p(u) + i \widetilde{x}_p(u)) e^{-f_0 u}, (y_p(u-t) + i \widetilde{y}_p(u-t)) e^{-f_0(u-t)} \rangle$$

Согласно лемме 1,

$$\begin{aligned} 2 \langle x, y(t) \rangle &= e^{-f_0 t} \langle x_p + i \widetilde{x}_p, x_p(t) + i \widetilde{y}_p(t) \rangle = \\ &= e^{-f_0 t} [\langle x_p, y_p(t) \rangle + \langle \widetilde{x}_p, \widetilde{y}_p(t) \rangle + i \langle \widetilde{x}_p, y_p(t) \rangle - i \langle x_p, \widetilde{y}_p(t) \rangle] \end{aligned}$$

Далее

$$\begin{aligned} \langle \widetilde{x}_p, \widetilde{y}_p(t) \rangle &\rightarrow x_p(f) y_p^*(f) \left| \frac{\text{sign}(f)}{i} \right|^2 \rightarrow \langle x_p, y_p(t) \rangle \\ \langle \widetilde{x}_p, y_p(t) \rangle &\rightarrow x_p(f) \frac{\text{sign}(f)}{i} y_p^*(f) \rightarrow \widetilde{\langle x_p, y_p(t) \rangle} \\ \langle x_p, \widetilde{y}_p(t) \rangle &\rightarrow x_p(f) y_p^*(f) \frac{\text{sign}(f)}{i^*} \rightarrow \widetilde{\langle x_p, y_p(t) \rangle} \end{aligned}$$

Используя эти тождества, найдем

$$2 \langle x, y(t) \rangle = 2e^{-f_0 t} [\langle x_p, y_p(t) \rangle + 2i \widetilde{\langle x_p, y_p(t) \rangle}],$$

что и требовалось доказать \blacktriangledown

Стоит удивиться тому, что закон преобразования корреляционных функций между Р и В областями по форме совершенно аналогичен преобразованиям сигналов. В частности, как следует из (9), корреляционную функцию реальных сигналов можно найти по корреляционной функции их комплексных огибающих следующим образом

$$\langle x_p, y_p(t) \rangle = \text{Re} \langle x, y(t) \rangle \cos f_0 t - \text{Im} \langle x, y(t) \rangle \sin f_0 t$$

Положив $t = 0$, найдем соотношение между скалярными произведениями сигналов и их комплексных огибающих

$$\langle x, y \rangle = \langle x_p, y_p \rangle + i \widetilde{\langle x_p, y_p \rangle}$$

Налицо печальный факт. Скалярные произведения в Р и В областях, не совпадают. В вещественной области скалярное произведение всегда вещественно. В области комплексных огибающих оно может быть комплексным. Вещественная часть скалярного произведения огибающих совпадает со скалярным произведением реальных сигналов

$$\langle x_p, y_p \rangle = \text{Re} \langle x, y \rangle.$$

Мнимая же часть - есть значение в нуле Гильбертова преобразования взаимной корреляционной функции реальных сигналов. В силу различия скалярных произведений, имеются различия между понятиями ортогональности в Р в В областях.

В области комплексных огибающих различают

Когерентную ортогональность

$$\text{Re} \langle x, y \rangle = 0$$

Некогерентную ортогональность

$$\langle x, y \rangle = 0$$

Когерентная ортогональность нарушается при сдвиге несущей одного из сигналов. В самом деле

$$Re \langle x, ye^{-i\theta} \rangle = Re \langle x, y \rangle \cos \theta - Im \langle x, y \rangle \sin \theta$$

Некогерентно (по настоящему) ортогональные сигналы сохраняют ортогональность при всех сдвигах несущих.

Ортогональность комплексных огибающих в любом из этих смыслов приводит к ортогональности реальных сигналов. Обратное неверно. Сигналы могут быть ортогональны в радиоканале, но неортогональны в комплексной области. Это случай наличия когерентной, но отсутствия некогерентной ортогональности. Реализуется он просто. Такая ситуация создается ровно тогда, когда радиосигналы ортогональны в силу принадлежности разным квадратурным компонентам - синусной и косинусной. При этом скалярное произведение комплексных огибающих всего лишь чисто мнимо.

Удивительно, но факт. Несмотря на различие в скалярных произведениях нормы (метрики) в пространствах радиосигналов и комплексных огибающих совпадают.

Теорема 2.9 (Теорема о совпадении метрик)

$$\|x\|^2 = \langle x, x \rangle = \langle x_p, x_p \rangle = \|x_p\|^2$$

Доказательство. В вещественной Р-области автокорреляционная функция симметрична. Поэтому, ее Гильбертово преобразование антисимметрично, и, как и всякая антисимметричная функция, принимает значение 0 в нуле. В силу этого,

$$\langle x, x \rangle = \langle x_p, x_p \rangle + i \overline{\langle x_p, x_p(0) \rangle} = \langle x_p, x_p \rangle,$$

что и означает совпадение норм \blacktriangledown

Окончательно. ВР/РВ преобразования - это вещественно-линейные отображения сигнальных пространств, не сохраняющие скалярное произведение (ортогональность), но сохраняющие метрику.

3 Методы модуляции

3.1 Аналоговая модуляция

Любой метода аналоговой модуляции - это отображение полезного сигнала на траекторию движения точки в плоскости значений комплексной огибающей. Проекции этой траектории на вещественную и мнимую оси передаются затем по каналу в виде квадратурных составляющих.

Положение дел слегка осложняет то обстоятельство, что комплексная плоскость двухмерна, а сигнал для передачи у нас имеется только один. Приходится искусственно выбирать какие то специальные классы движений.

Различают два основные вида аналоговой модуляции - амплитудная и фазовая.

При **амплитудной** модуляции в качестве области допустимых движений выбирается какая-либо прямая в комплексной плоскости. Всегда можно считать, что это вещественная ось. Полезный сигнал управляет в этом случае только амплитудой комплексной огибающей.

В случае **фазовой** модуляции движение происходит по окружности в комплексной плоскости. Постоянной остается амплитуда огибающей.

Итак, пусть передаче по радиоканалу подлежит полезный сигнал $s(t)$. Без ограничения общности будем считать его нормированным по уровню к единичному диапазону: $|s(t)| \leq 1$.

Самое простое, что может быть - это **балансная** модуляция (БМ). Полезный сигнал управляет движением по вещественной оси. Комплексная огибающая совпадает с самим сигналом - $x(t) = s(t)$. В радиоканале с несущей f_0 БМ- сигнал принимает вид

$$x_p(t) = s(t) \cos f_0 t.$$

На приеме никаких дополнительных преобразований не требуется. Полезный сигнал $s(t)$ восстанавливается непосредственно квадратурным демодулятором.

В числе недостатков БМ выделяют два.

1. БМ сигнал не допускает некогерентной демодуляции по модулю комплексной огибающей. Дело в том, что сигнал $s(t)$ может менять знак, а модуль огибающей всегда положителен. Некогерентный демодулятор восстановит $|s(t)|$ вместо $s(t)$. Выход прост - передавать нечто, не меняющее знака. Чтобы исключить перемены знака к комплексной огибающей добавляют константу. Получается то, что называют **балансной амплитудной** модуляцией (АМ)

$$x(t) = 1 + ms(t), \quad x_p(t) = (1 + ms(t))\cos f_0 t, \quad 0 \leq m \leq 1 - \text{глубина модуляции}$$

АМ сигнал можно демодулировать по модулю огибающей (некогерентно). В традиционных АМ-приемниках эту задачу успешно решает диодный детектор, работающий непосредственно в Р-области. Это старейший из методов модуляции, известный со времен Попова.

2. БМ модуляция неэффективна спектрально. Радиосигнал, переносящий полезный сигнал $s(t)$ с верхней частотой F , занимает в канале полосу $(f_0 - F, f_0 + F)$ удвоенной ширины $2F$. Это типично для всех методов модуляции, использующих одну квадратурную составляющую.

Проблема с полосной эффективностью решается передачей аналитического сигнала, не содержащего компонент спектра на отрицательных частотах. На передающей стороне

формируется гильбертово преобразование $\tilde{s}(t)$ сигнала $s(t)$ и строится аналитический сигнал $s^+(t) = s(t) + i\tilde{s}(t)$. Он и передается в квадратурах. В эфире это выглядит так:

$$x_p(t) = s(t) \cos f_0 t - \tilde{s}(t) \sin f_0 t$$

Это так называемая **однополосная** модуляция (SSB-single side band). Строго говоря, она не относится ни к чисто амплитудным, ни к фазовым. SSB требует когерентной демодуляции.

1. При **фазовой** модуляции (ФМ) полезный сигнал управляет движением по единичной окружности в комплексной плоскости. Огибающая имеет вид $x(t) = e^{\psi s(t)}$, где ψ -индекс модуляции. В эфире это выглядит как модуляция фазы несущей полезным сигналом

$$x_p(t) = \cos(f_0 t + \psi s(t))$$

В этом виде фазовая модуляция почти не применяется. Зато широкое применение нашла ее модификация - **частотная** модуляция (ЧМ). Отличие небольшое, но принципиальное. При ЧМ полезный сигнал управляет не фазой, а производной от фазы, то есть частотой - $f(t) = Ds(t)$. Параметр D (в герцах) называется девиацией частоты. Он задает границы диапазона $[-D, D]$ перестройки частоты при изменении уровня сигнала в пределах от -1 до +1.

В области комплексных огибающих ЧМ сигнал имеет вид

$$x(t) = e^{\psi(t)} = e^{D \int_0^t s(t) dt}$$

и изображается точкой, вращающейся по единичной окружности с частотой $Ds(t)$. Положительным и отрицательным уровням сигнала отвечают разные направления вращения.

Блок схема ЧМ-демодулятора показана на рис. 4. Это петля фазовой авто подстройки (ФАП). Ее основа - управляемый напряжением генератор VCO, который выступает в роли частотного модулятора, пытаясь сформировать местную копию $e^{\psi'(t)}$ принимаемого сигнала $e^{\psi(t)}$. Местному модулятору приходится довольствоваться сигналом $\tilde{s}(t)$, который пока лишь претендует на роль оценки полезного. Фазовый дискриминатор FD измеряет величину текущей фазовой ошибки $\theta(t) = \psi(t) - \psi'(t)$. После усиления в K раз сигнал ошибки используется для управления VCO.

Имеем

$$\frac{d\theta}{dt} = D(s(t) - \tilde{s}(t)),$$

но, с другой стороны, $\tilde{s}(t) = K\theta(t)$. Приходим к следующему уравнению для $\tilde{s}(t)$

$$\tau \frac{d\tilde{s}(t)}{dt} + \tilde{s}(t) = s(t); \tau = 1/KD$$

Можно проверить, что это в точности то уравнение, которое описывает прохождение сигнала $s(t)$ через интегрирующую цепь с постоянной времени τ .

Таким образом, петля ФАП действительно дает оценку $\tilde{s}(t)$ полезного сигнала $s(t)$ с точностью до фильтрации эквивалентной интегрирующей цепью. Увеличением коэффициента усиления петли K постоянная времени $1/KD$ интегрирующей цепи может быть сделана произвольно малой, так что фильтрация становится незаметной. Блок LPF служит фильтром низких частот, отсекающим возможные помехи вне полосы полезного сигнала.

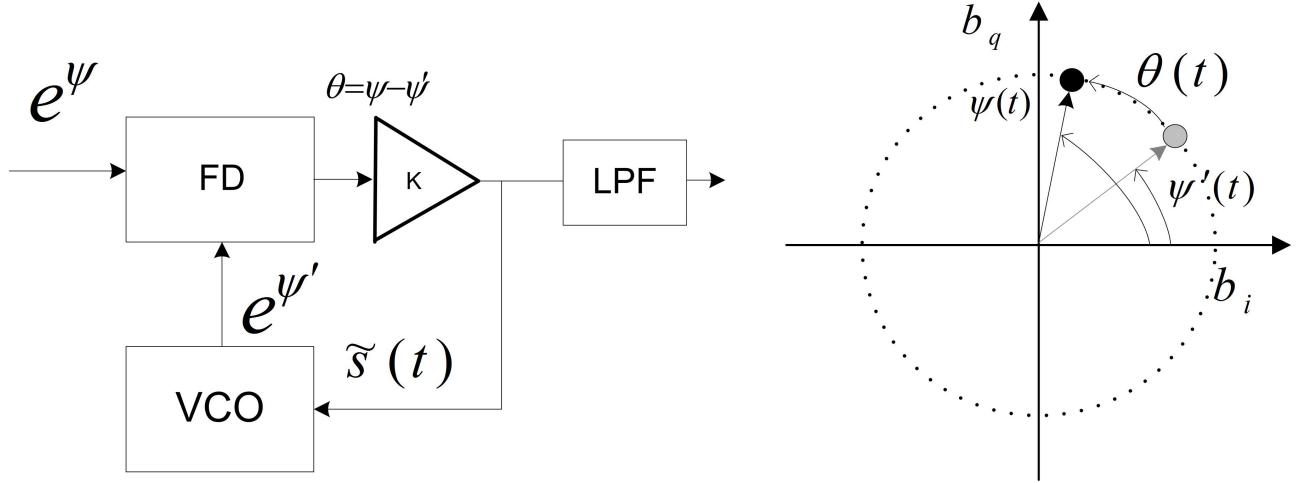


Рис. 4: Частотный демодулятор на петле ФАП

3.2 Цифровая линейная модуляция

Наиболее простые технические решения для передачи цифровых данных относятся к классу линейных методов модуляции.

Пусть передаче подлежит некоторый цифровой информационный поток. Будем описывать его бесконечной в обе стороны последовательностью комплексных символов $c_n \in \mathbb{C}$, случайно выбираемых из некоторого конечного набора \mathbb{C} (созвездия) мощности M .

В пространстве комплексных огибающих выберем некоторый вещественный сигнал $p(t)$ - базовый импульс. Без ограничения общности можно принять его энергию $\langle p, p \rangle$ равной единице. Общая форма линейно модулированного сигнала в В-области имеет вид взвешенной суммы копий базового импульса, сдвинутых во времени с шагом T

$$x(t) = \sum_{n=-\infty}^{+\infty} c_n p(t - nT) \quad (10)$$

Это и есть комплексная огибающая сигнала линейной модуляции. Каждый отдельный импульс переносит один информационный символ. Скорость передачи составляет $1/T$ символов в секунду или $\log_2 M / T$ битов в секунду.

Реальный радиосигнал $x_p(t)$ формируется из комплексной огибающей по стандартной схеме квадратурного модулятора. На стороне приема квадратурный демодулятор воссоздает копию огибающей $x(t)$ для обработки.

Замечание 1 Заметим, что сигнал (10) можно рассматривать как результат фильтрации серии δ -импульсов

$$c(t) = \sum_{n=-\infty}^{+\infty} c_n \delta(t - nT)$$

линейным фильтром с импульсной реакцией $h = p(t)$. Отсюда слово «линейная» в названии. С этих позиций, модуляция - это нечто противоположное дискретизации. В одном случае речь идет о представлении сигнала потоком выборочных значений, а во втором - о представлении потока символов сигналом.

Оптимальный приемник реализует согласованную фильтрацию. Для этого он вычисляет корреляционную функцию $R(t) = \langle x, p(t) \rangle$ принятой реализации с базовым импульсом и берет ее выборочные значения в дискретные моменты времени $t = mT$.

Пусть базовый импульс ортогонален всем своим сдвигам на время, кратное T

$$\langle p(t), p(t - mT) \rangle = \delta_{m,0}.$$

Тогда значение отсчета коррелятора в момент $t = mT$ оказывается равным c_m , давая оценку m -го символа, не зависящую от других информационных символов. Условие ортогональности сдвигов базового импульса называют условием *отсутствия межсимвольной интерференции*. Отсутствие интерференции автоматически обеспечивается требованием финитности импульса на временном интервале $(-T/2, T/2)$, но может выполняться и для импульсов, не финитных во времени.

Схема линейной модуляции вполне определяется выбором

- базового импульса $p(t)$
- сигнального созвездия \mathbb{C} .

Линейная модуляция относится к одномерным - в комплексном пространстве сигналов ею используется ровно одно измерение. Каждый отдельный передаваемый сигнал $c_n p(t)$ коллинеарен базовому импульсу. Поэтому согласованную фильтрацию удается реализовать применением единственного коррелятора.

Передача сигнала $c_n p(t)$ связана с затратой энергии $\langle c_n p, c_n p \rangle = |c_n|^2 \langle p, p \rangle = |c_n|^2$. Энергетику созвездия в целом характеризует средняя энергией сигнала

$$E_s = \frac{1}{M} \sum_{n=1}^M |c_n|^2$$

или энергия на бит информации $E_b = E_s / \log_2 M$.

Степень надежности различия двух сигналов $c_n p(t)$, $c_m p(t)$ на приеме определяется Евклидовым расстоянием между ними

$$d_{n,m} = \|c_n p(t) - c_m p(t)\| = |c_n - c_m|.$$

В этом контексте задача выбора созвездия ставится следующим образом.

В комплексной плоскости выбрать заданное число M точек так, чтобы обеспечить максимум расстояния между парами при ограниченной средней энергии созвездия. В качестве разумного критерия качества созвездия используют энергетическую эффективность - отношение квадрата минимального Евклидова расстояния к средней энергии сигнала.

$$\eta = \frac{\min_{n \neq m} d_{n,m}^2}{E_s}$$

Существует два популярные класса созвездий. Это созвездия фазовой манипуляции PSK (Phase Shift Keying), рис. 5, и созвездия квадратурной амплитудной модуляции QAM (Quadrature Amplitude Modulation), рис. 6.

Простейший и наиболее популярный вид линейной модуляции - это двоичная фазовая манипуляция PSK-2, которая использует созвездие всего из двух точек. Этот вид модуляции имеет специальные названия. Можно встретить такие как BPSK (Binary Phase Shift Keying), PAM (Pulse Amplitude Modulation).

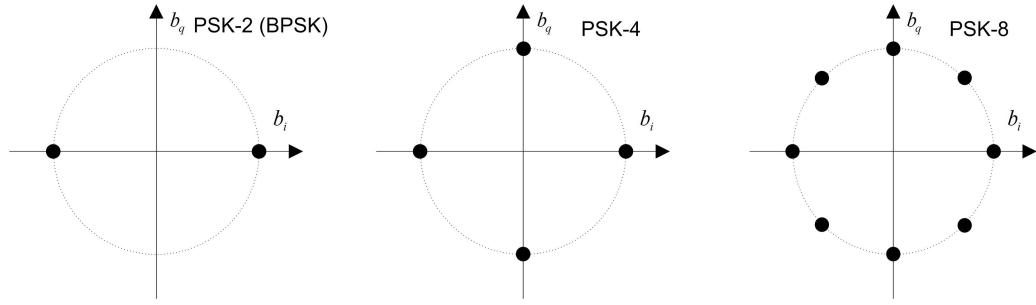


Рис. 5: Созвездия PSK

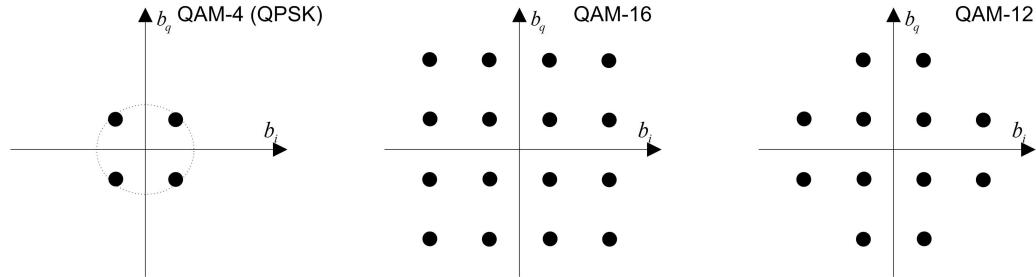


Рис. 6: Созвездия QAM

При выборе прямоугольного импульса двоичная фазовая манипуляция сводится к изменению знака несущей в соответствии со знаками символов бинарного информационного потока.

$$x_p(t) = \sqrt{\frac{2}{T}} \sum_n rect(t - nT) c_n \cos f_0 t, \quad c_n = \pm 1$$

BPSK заведомо неэффективна спектрально, так как использует всего один квадратурный канал. В той же полосе частот можно разместить ортогональный квадратурный канал с другим BPSK сигналом, переносящим независимый бинарный поток $\hat{c}_n = \pm 1$.

$$x_p(t) = \sqrt{\frac{2}{T}} \sum rect(t - nT) (c_n \cos f_0 t - \hat{c}_n \sin f_0 t)$$

В результате получается квадратурная фазовая манипуляция QPSK (Quadrature Phase Shift Keying). QPSK в точности совпадает с QAM-4 и отличается от PSK-4 лишь поворотом сигнального созвездия на $\pi/4$. Так что все эти виды модуляции, по существу, эквивалентны.

При анализе QPSK ее чаще всего интерпретируют как два ортогональных BPSK канала, использующие единую полосу. На приеме косинусная и синусная компоненты QPSK сигнала преобразуются квадратурным демодулятором в вещественную и мнимую части комплексной огибающей, которые могут обрабатываться независимо как самостоятельные BPSK-сигналы.

При этом приходится, правда, встать на ту точку зрения, что перекрывающиеся во времени импульсы, переносящие символы c_n , \hat{c}_n , - это разные импульсы. В P-области эти импульсы ортогональны в силу принадлежности разным квадратурным компонентам. В области комплексных огибающих они ортогональны лишь когерентно. Первый - вещественен, второй - мним.

При такой трактовке оказывается, что каждый импульс переносит один бит, энергия на символ совпадает с энергией на бит, вероятности ошибки в битах можно анализировать независимо.

QPSK является простейшим спектрально эффективным видом цифровой модуляции. Модель QPSK проста и до конца анализируется аналитическими методами. Поэтому она получила статус референсной модели, с которой сравнивают характеристики других схем модуляции.

3.2.1 Спектральная эффективность и межсимвольная интерференция

Полоса частот, занимаемая линейно модулированным сигналом, определяется шириной спектра базового импульса. Произведение эффективной двухсторонней полосы F на эффективную длительность T называют базой импульса $B = FT$. Обратная величина $\rho_s = 1/B = \frac{1/T}{F}$ имеет смысл отношения символьной (бодовой) скорости передачи к полосе. Она измеряется в числе комплексных измерений (Бод) в секунду на герц двухсторонней полосы.

Битовой спектральной эффективностью ρ называют скорость передачи в битах в секунду на герц двухсторонней полосы. Битовая эффективность $\rho = \frac{\log_2 M}{T} \frac{1}{F}$ больше бодовой ρ_s в $\log_2 M$ раз, где M -мощность созвездия. В отличие от бодовой, битовая эффективность ρ не относится к чисто спектральным характеристикам, поскольку она учитывает также и энергетические характеристики созвездия.

Для реальных импульсов конечной длительности оценить спектральную эффективность можно лишь приближенно, приняв некие разумные соглашения относительно понятия эффективной полосы.

Ниже даны примеры спектров $P(f)$ ряда типовых импульсов $p(t)$. Здесь $FT = 1$.

Импульс Найквиста

$$\sqrt{F} \frac{\sin \pi F t}{\pi F t} \leftrightarrow \frac{1}{\sqrt{F}} \text{rect}_F(f)$$

Прямоугольный импульс

$$\frac{1}{\sqrt{T}} \text{rect}_T(t) \leftrightarrow \sqrt{T} \frac{\sin \pi T f}{\pi T f}$$

Импульс кода Манчестера

$$\frac{1}{\sqrt{T}} \{ \text{rect}_{T/2}(t - T/2) - \text{rect}_{T/2}(t + T/2) \} \leftrightarrow -i \sqrt{T} \frac{\sin^2 \pi \frac{T}{2} f}{\pi \frac{T}{2} f}$$

Гармонический полупериод

$$\sqrt{\frac{2}{T}} \cos\left(\pi \frac{t}{T}\right) \text{rect}_T(t) \leftrightarrow \frac{2\sqrt{2T}}{\pi} \frac{\cos \pi T f}{1 - (2Tf)^2}$$

Принципиальный косинус

$$\sqrt{\frac{2}{3T}} \left(1 + \cos\left(2\pi \frac{t}{T}\right)\right) \text{rect}_T(t) \leftrightarrow \sqrt{\frac{2T}{3}} \frac{\sin \pi T f}{\pi T f} \frac{1}{1 - (Tf)^2}$$

Эффективная полоса прямоугольного импульса, определенная по расстоянию между первыми нулями $P(f)$, составляет $2F$, $F = 1/T$, что отвечает базе $B = 2FT = 2$, или спектральной эффективности $\rho_s = 1/2$ измерений на герц. Для сравнения, база B импульса Найквиста составляет FT , что дает вдвое большую бодовую спектральную эффективность - одно измерение на герц.

Определенное представление о непростой связи спектральных свойств с межсимвольной интерференцией дает следующий результат.

Теорема 3.1 (Критерий Найквиста) *Межсимвольная интерференция при взятии отсчетов с шагом T отсутствует если и только если спектр $P(f)$ базового импульса $p(t)$ удовлетворяет условию*

$$\sum_{m=-\infty}^{m=+\infty} |P(f - mF)|^2 \equiv T, \quad FT = 1. \quad (11)$$

Доказательство. Отсутствие межсимвольной интерференции эквивалентно требованию к автокорреляционной функции: $R(nT) = \langle p, p(nT) \rangle = \delta_{n,0}$. Рассмотрим дискретизацию корреляционной функции во времени

$$R(t) \sum \delta(t - mT) = \sum R(mT) \delta(t - mT) = \delta(t)$$

В частотной области этому отвечает периодическое продолжение энергетического спектра

$$F \sum |P(f - mF)|^2$$

Но, с другой стороны, спектр δ -функции - это тождественная единица \blacktriangledown

Положив $f = 0$ в (11), получаем следующее

Следствие 2 *Межсимвольная интерференция при взятии отсчетов с шагом T отсутствует если и только если спектр $P(f)$ базового импульса $p(t)$ удовлетворяет условию*

$$\sum_{m=-\infty}^{m=+\infty} |P(mF)|^2 = T, \quad FT = 1. \quad (12)$$

Согласно критерию Найквиста, для отсутствия межсимвольной интерференции необходимо, чтобы

- Периодическое продолжение энергетического спектра по частоте с шагом $F = 1/T$ равномерно покрывало всю ось.
- Сумма выборок энергетического спектра в точках mF была равна T .

Выборка $|P(f)|^2$ в точке 0 может служить оценкой энергии главного лепестка спектра. Сумма же выборок в точках $mF, m \neq 0$ дает представление о суммарной внеполосной энергии. Особенно интересен уровень первой внеполосной выборки на границе F «разрешенного» частотного диапазона.

Если уровень главного лепестка достигает максимально возможного значения T , то согласно (12), внеполосные выборки спектра должны обратиться в нуль. По мере снижения уровня главного лепестка, суммарная внеполосная энергия растет. *Красоту картины несколько портит особый случай прямоугольного импульса, когда спектр явно плох, но это маскируется его «случайным» обращением в нуль во всех точках дискретизации $mF, m \neq 0$, где берутся оценки внеполосного излучения.*

Из критерия Найквиста вытекает, что прямоугольный в частотной области импульс Найквиста оптимален по критерию компактности энергетического спектра. Периодическое продолжение его спектра равномерно покрывает ось частот, внеполосное излучение отсутствует. Таким образом, импульса Найквиста дает теоретический предел для спектральной эффективности линейной модуляции без межсимвольной интерференции. Это значение составляет одно комплексное измерение (Бод) в секунду на герц двухсторонней полосы.

Прямоугольный импульс имеет вдвое большую эффективную ширину спектра по расстоянию между первыми нулями, что дает спектральную эффективность $1/2$. К тому же, его спектр медленно спадает с частотой - скорость спада всего $1/f$. В сравнении с ним, импульс «гармонический полупериод» имеет несколько более широкий спектр. Относительный уровень первой внеполосной выборки по энергии составляет $1/9$. Это плата за большую скорость спада - $1/f^2$. Скорость спада спектра импульса «приподнятый косинус» еще выше - $1/f^3$, но компактность спектра хуже - уровень первой внеполосной выборки - $1/4$. Таким образом, использование в линейной модуляции неперекрывающихся во времени импульсов приводит к спектральной эффективности порядка $1/2$, что вдвое ниже теоретического предела.

Теория рекомендует применять в системах линейной модуляции нефинитный во времени импульс Найквиста. Практически же он не используется совсем. И дело тут не только в сложности его формирования. Основная проблема связана с тем, что этот импульс обеспечивает отсутствие интерференции только в том случае, когда отсчеты коррелятора берутся в точности в моменты nT . Любая ошибка во временном положении точек дискретизации приводит к катастрофической межсимвольной интерференции. Скорость спада импульса Найквиста во времени составляет всего $1/t$. Поэтому, при наличии ошибки синхронизации, импульсы, удаленные по времени на nT , начинают давать вклад в отсчет коррелятора с уровнем порядка $1/nT$. Ряд же $\sum 1/nT$ - расходится.

Удаётся построить нефинитные во времени импульсы с большей скоростью спада, скажем, порядка $1/t^2$. Однако это достигается за счет ухудшения компактности спектра.

3.3 Варианты линейной модуляции с короткой памятью

При линейной модуляции комплексный символ созвездия $c = c_i + ic_q$ фактически передается по частям - вещественная часть в косинусном канале, а мнимая в синусном.

$$x_p(t) = \sqrt{2}\{c_i p(t)\cos f_0 t - c_q p(t)\sin f_0 t\}$$

При приеме сигналу $c_i p(t)$ отвечает вещественная часть выхода коррелятора, а сигналу $c_q p(t)$ -мнимая.

Эксплуатируется идея разнесения этих двух вещественных подканалов на половину длительности символьного интервала T .

$$x_p(t) = \sqrt{2}\{c_i p(t)\cos f_0 t - c_q p(t - T/2)\sin f_0 t\}$$

Естественно, что и решения в вещественном и мнимом подканалах также должны приниматься со сдвигом - в моменты nT и $nT + T/2$.

Так возникают системы модуляции с короткой памятью. В течение каждого $T/2$ -интервала сигнал зависит сразу от двух символов созвездия. Влияние же данного символа распространяется ровно на два $T/2$ -интервала.

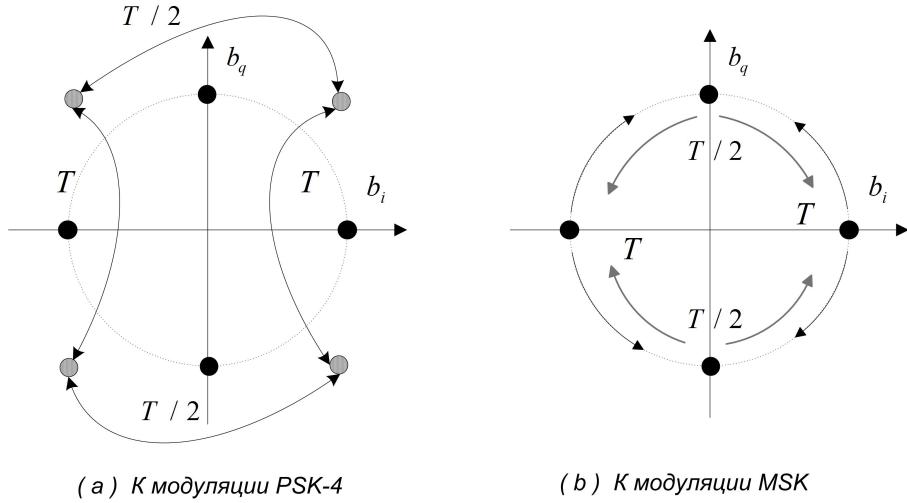


Рис. 7: К обсуждению схем модуляции с памятью.

Начнем с более простого случая PSK-4 с прямоугольным импульсом, рис. 7(a).

При стандартной схеме модуляции сигнальная точка пребывает в одной из точек созвездия (черные точки на рисунке). В моменты, кратные T , она либо остается на месте, либо совершает переход в любую другую точку. Этому отвечают сдвиги несущей на $0, +\pi/2, -\pi/2, \pi$.

При модуляции с $T/2$ -сдвигом полный сигнал может пребывать в одной из 4-х точек, показанных на рисунке серыми. В T -моменты меняет знак вещественная компонента сигнала. Этому отвечает один из возможных переходов по горизонтали. В $T/2$ -моменты меняет знак мнимая часть, что вызывает вертикальный переход. Фокус состоит в том, что при этом фаза сигнала никогда не изменяется больше чем на $\pi/2$.

Вид модуляции, который получается в результате, часто интерпретируют как спектрально эффективную двоичную фазовую манипуляцию BPSK. Очередной бит информации передается каждые $T/2$ секунд. В четные моменты времени он определяет амплитуду косинусной компоненты сигнала, в нечетные - синусной. Модуляция получается

чисто двоичной, поскольку на приеме решения о четных и нечетных символах выносятся независимо на основании обработки вещественной и мнимой компонент отклика коррелятора.

Небольшая модификация этой схемы ведет к знаменитой системе манипуляции минимальным частотным сдвигом (Minimal Shift Keying). Модификация заключается просто в том, что прямоугольный базовый импульс заменяется на «гармоническую полуволну»

$$p(t) = \cos\left(\frac{\pi}{2} \frac{t}{T/2}\right), \quad t \in (-T/2, T/2).$$

В Т-моменты импульс $p(t)$ максимальен, а сдвинутый на $T/2$ импульс равен нулю и наоборот. Таким образом, в Т-моменты сигнал находится в одной из точек на горизонтальной оси, рис. 7(b). Здесь решается вопрос о знаке очередного символа в мнимом канале. Результатом этого решения становится плавное, по закону нарастания синуса, движение по мнимой оси вверх или вниз от нуля. Движение же вдоль вещественной оси просто продолжается - в сторону нуля. В очередной $T/2$ момент сигнал достигает одной из двух возможных точек на мнимой оси. Здесь решается вопрос о направлении следующего вещественного движения.

На каждом $T/2$ -интервале сложение двух гармонических движений вдоль вещественной и мнимой осей порождает движение по окружности. За время $T/2$ проходится ровно четверть окружности - от точки до точки. Вот список четырех возможных траекторий движения в комплексной форме

$$x(t) = \pm e^{\pm i \frac{\pi}{2} \frac{t}{T/2}}, \quad t \in [0, T/2] \quad (13)$$

Каждое новое движение начинается точке созвездия, достигнутой на предыдущем полуинтервале. Направление движения определяется очередным битом информации. Равномерное движение по окружности влево/вправо эквивалентно сдвигу по частоте на

$$2\pi\Delta f = \frac{\pm \pi/2}{T/2}; \quad \Delta f = \pm 1/2T = F/2; \quad FT = 1$$

Известно, что $1/2T$ - это минимальный сдвиг по частоте, который обеспечивает когерентную ортогональность гармонических колебаний на отрезке длительности T . Отсюда название - манипуляция с минимальным частотным сдвигом.

Упражнение 19 Проверить некогерентную ортогональность сигналов (13)

3.4 Решетчатая модуляция

В отсутствие межсимвольной интерференции схемы линейной модуляции не обладают памятью. Сигнал на данном T -интервале, зависит только от текущего информационного символа. Марковские методы модуляции, которые будут обсуждаться далее, подключают зависимость от предыстории. Говорят, что модуляторы этого класса обладают памятью.

Модулятор с памятью удобно описывать как конечный автомат.

Формально, конечный автомат - это конечное множество состояний $\{s \in \mathbb{S}, |\mathbb{S}| = S\}$, входной алфавит $\{b \in \mathbb{B}, |\mathbb{B}| = B\}$ и выходной алфавит $\{c \in \mathbb{C}, |\mathbb{C}| = C\}$.

Автомат функционирует в дискретном времени n . Если в момент n автомат находится в состоянии s_n , а на его входе присутствует символ b_n , то в следующий момент он окажется в состоянии $s_{n+1} = F_s(s_n, b_n)$ и выдаст на выход символ $c_{n+1} = F_c(s_n, b_n)$. Отображения

следования $F_s : \mathbb{S} \times \mathbb{B} \mapsto \mathbb{S}$ и вывода $F_b : \mathbb{S} \times \mathbb{B} \mapsto \mathbb{C}$ вполне определяют работу автомата при заданной последовательности входов.

В процессе работы автомат порождает последовательность выходных символов $c_n \in \mathbb{C}$. Каждый новый символ определяется текущими состоянием и входом. Текущее состояние зависит от предыстории работы автомата, в том числе, от последовательности предыдущих входных символов. В итоге автомат порождает поток символов марковского типа, каждый новый символ которого статистически зависит от предыдущих.

Работа автомата описывается диаграммой состояний или решетчатой (trellis) диаграммой, рис. 8. На первой изображаются состояния-точки и переходы-стрелки. Решетчатая диаграмма разворачивает последовательность работы автомата по времени. Состояниям на ней отвечают вершины, переходам - ребра.

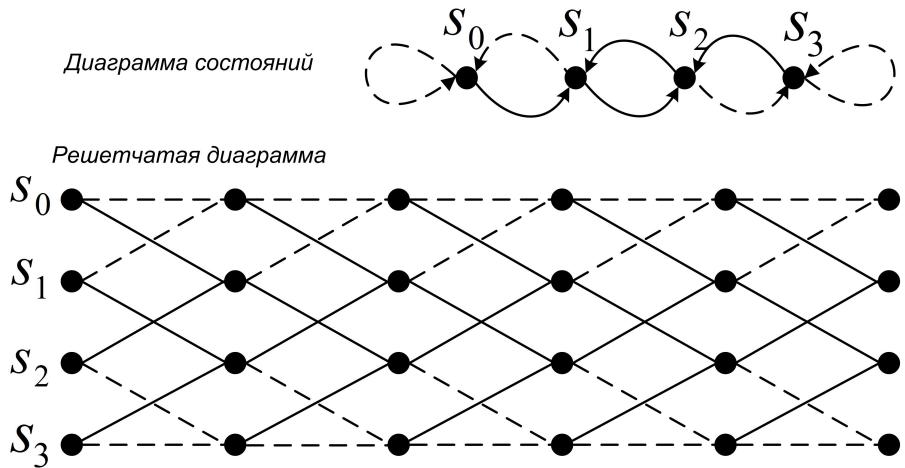


Рис. 8: Диаграмма состояний и ее решетка.

Показанные на рисунке диаграммы представляют автомат с четырьмя состояниями $\mathbb{S} = \{s_0, s_1, s_2, s_3\}$ и двоичным входным алфавитом $\mathbb{B} = [0, 1]$. При двоичных входах из каждого состояния имеется ровно два перехода. На диаграмме состояний они показаны сплошными ($b = 1$) и пунктирными ($b = 0$) стрелками. Соответственно, на решетке из каждой вершины выходит два ребра. Переход (ребро) однозначно определяется парой - текущее состояние/вход. Функция вывода $F_c : \mathbb{S} \times \mathbb{C} \mapsto \mathbb{B}$ задает маркировку ребер символами выходного алфавита. Наш автомат насчитывает 8 переходов(ребер). Для их маркировки можно использовать любой конечный алфавит. Если есть желание присвоить разным ребрам разные символы, что не обязательно, то хватит 8-символьного алфавита.

Идея решетчатого кодирования проста. Она заключается в том, чтобы использовать информационный поток для управления автоматом, а по каналу передавать маркировку ребер.

В самом общем случае можно считать, что каждому из ребер присвоен некоторый уникальный сигнал. В реальности, как правило, ребра маркируют символами из подходящего сигнального созвездия, применяя для их передачи линейную модуляцию. К примеру, множество ребер нашего автомата взаимно однозначно отображается на созвездие PSK-8. В принципе, можно применить и отображение на PSK-2. При этом некоторым из ребер придется присвоить одинаковые сигналы.

Решетчатая модуляция открывает возможность проводить демодуляцию не локально, принимая независимые решения по каждому символу, а распределенно, анализируя

длинные последовательности ребер - пути на решетчатой диаграмме. Для этого применяется, алгоритм Витерби, появление которого вызвало революцию в технике связи.

Алгоритм Витерби.

Алгоритм Витерби использует решетчатую диаграмму автомата и поставляемые демодулятором метрики путей.

Рассмотрим процесс формирования метрик. Пусть, для конкретности, пути нашего автомата маркованы символами $\{c_0, \dots, c_7\}$ созвездия PSK-8. Каждому пути отвечает тогда посланный по каналу сигнал $c_i p(t)$. Приемник получает этот сигнал с аддитивной помехой $y(t) = c_i p(t) + n(t)$. Природа помехи для нас сейчас несущественна. Согласованный фильтр приемника вычисляет скалярное произведение

$$\hat{y} = \langle y(t), p(t) \rangle = c_i + \langle n(t), b(t) \rangle.$$

Результат фильтрации \hat{y} , вообще говоря, не совпадает ни с одним из ожидаемых сигналов. Вместо того, чтобы, тем не менее, принять жесткое решение о переданном символе, демодулятор ограничивается вычислением набора из 8-и расстояний до них - $d_i = |\hat{y} - b_i|$. Это называют демодуляцией с мягким решением. Мягкое решение, то есть, набор расстояний до сигналов, и образует набор реберных метрик для данного тактового интервала.

После демодуляции с мягкими решениями мы получаем не данные, а всего лишь решетчатую диаграмму с реберными метриками. Время выделять данные. Здесь и начинается работа алгоритма Витерби.

Алгоритм Витерби исходит из того, что для выделения данных достаточно найти правильный путь на решетке и решает именно эту задачу. Правильный путь ищется по критерию наименьшей длины пути - наименьшего значения суммы реберных метрик.

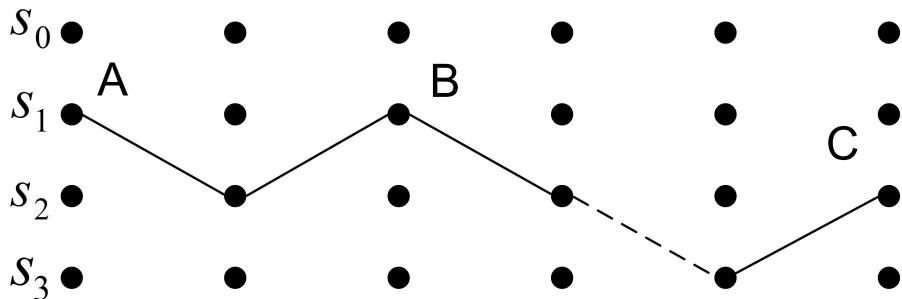


Рис. 9: К поиску минимального пути.

Пусть требуется найти путь минимальной длины из точки А в точку С, рис. 9. Предположим, что такой путь уже найден, и это путь, показанный на рисунке. Тогда его начальный участок от А до В также минимален по длине среди всех возможных путей из А в В. В самом деле, пусть это не так. Тогда существует более короткий путь из А в В. Пойдя по нему мы, заодно, сократим и путь из А в С, что невозможно. В теории оптимизации это простое наблюдение служит основой методов динамического программирования. В нашем случае оно позволяет решать задачу построения оптимальных путей рекурсией по глубине решетки.

Шаг рекурсии Витерби.

Предположим, что задача построения кратчайших путей от некоторой исходной вершины до всех возможных вершин в момент N уже решена. Эти пути называют *выжившими* до момента N .

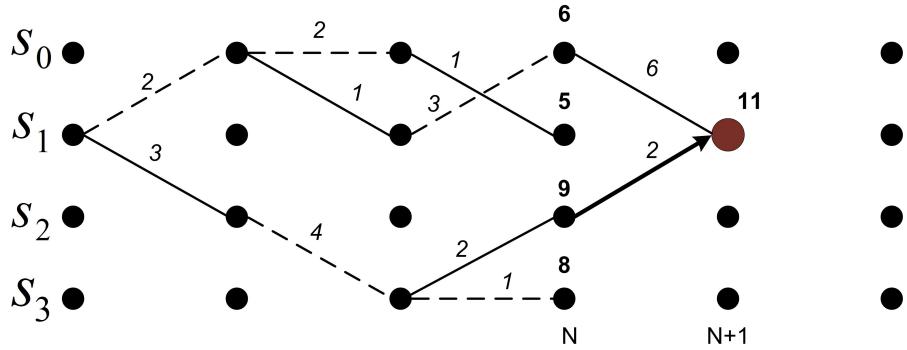


Рис. 10: К поиску минимального пути.

На рис.10 все пути из исходной вершины s_1 , выжившие до момента N , показаны с их длинами. Для простоты взяты целочисленные реберные метрики. Видим, что выживший путь в вершину s_0 имеет длину 6, выживший путь в s_1 -длину 5, и т.д.

Список выживших путей с их суммарными длинами - это набор исходных данных для шага итерации алгоритма Витерби. Итерация продлевает выжившие пути на один шаг, до момента $N + 1$. Реализуется она так:

Будем рассматривать все возможные вершины в следующий момент $N + 1$ поочередно.

Выделим одну из них. В выделенную вершину ведет некоторый набор ребер из предшествующих вершин. Каждое из них, удлиняя уже известный оптимальный путь на один шаг, дает путь из исходного состояния в выделенное. Имеем список таких путей. Можем вычислить их длины. Для этого к известной длине выжившего до N пути прибавляем метрику удлиняющего его ребра. Осталось выбрать из этого списка путей наиболее короткий. Это и будет новым путем, выжившим до момента $N + 1$. Его длина также известна.

К примеру, на рисунке в вершину s_2 приводят два ребра, начинающиеся в s_0 и s_2 . Длина пути, проходящего через s_0 составляет $6+6=12$. Длина пути через s_2 составляет $9+2=11$. Выживает второй путь с длиной 11.

Остается проблема выбора начальной вершины. Есть два варианта ее решения.

Первый основывается на использовании путей фиксированной длины N с раз и навсегда выбранными начальной и конечной вершинами, к примеру, вершинами s_0 . Путь начинают вести из s_0 и продолжают его до тех пор, пока не наступает пора заканчивать блок. Тогда организуют искусственную вставку данных, приводящую автомат снова в s_0 . К сожалению, это требует синхронизации границ блоков между сторонами.

Второй путь таков. Декодер Витерби начинает работу из состояния со списком выживших путей нулевой длины. Опыт показывает, что он достаточно быстро выходит на штатный режим, возможно совершая несколько ошибок на первых символах.

Упражнение 20 Относительная фазовая телеграфия.

Пусть для передачи двоичного потока $\{b\}$ используется автомат с двумя состояниями $\{s\}$, функцией переходов $s_{n+1} = s_n \oplus b_n$ и функцией выходов $c_{n+1} = s_n \oplus b_n$.

Показать, что информационные символы могут быть восстановлены по правилу

$$b_n = c_n \oplus c_{n-1},$$

что обеспечивает инвариантность информации относительно инвертирования символов принятого потока $\{c\}$.

Построить диаграмму состояний, решетчатую диаграмму. Показать, что знание двух соседних ребер на решетчатой диаграмме однозначно определяет один информационный бит, причем, значение так определенного бита не изменяется при инвертировании разметки ребер.

Выбрав трибуальную реберную метрику - 0 для правильного ребра и 1 для ошибочного, проанализировать работу алгоритма Витерби.

Упражнение 21 Построить диаграмму состояний и решетку для MSK-модулятора , рис. 7.

К решетчатой модуляции автоматически приводит использование сверточного кодирования данных. Пример сверточного кодера показан на рис. 11. Это автомат, который преобразует двоичную информационную последовательность b_n в последовательность двоичных пар (c_n^0, c_n^1) . Пары используются затем для маркировки точек созвездия. К примеру, четыре двоичные пары (c_n^0, c_n^1) идеально отображаются на созвездие QPSK.

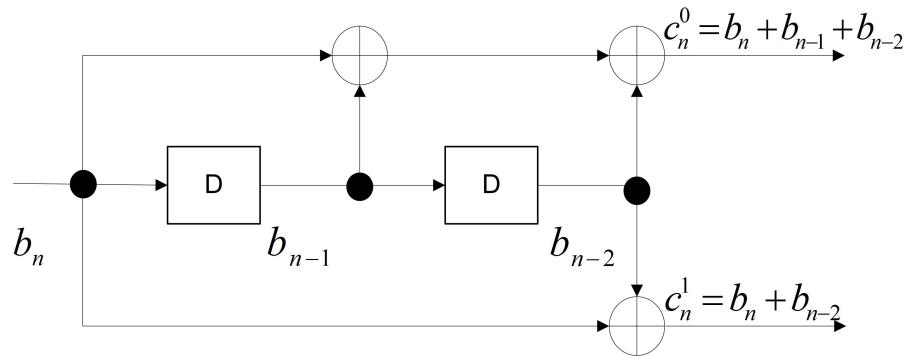


Рис. 11: Сверточный кодер.

Упражнение 22 Построить диаграмму состояний и решетчатую диаграмму для сверточного кодера на рис. 11

Решетчатая модуляция широко применяется для управления формой спектра линейно модулированных радиосигналов за счет должной организации корреляций между передаваемыми символами. Использование мягкой демодуляции с последующим применением алгоритма Витерби - это основной путь борьбы с последствиями межсимвольной интерференции, возникающей как из-за фильтрации сигнала в радиотракте, так и вследствие многолучевости распространения.

3.5 Ортогональная модуляция, временное, частотное и кодовое разделение

Рассмотренные выше методы модуляции используют одно комплексное измерение. Это предельно упрощает обработку на приеме. Требуется всего один коррелятор, вычисляющий наилучшую проекцию принятого сигнала на вектор базового импульса. Дальнейшее развитие методов модуляции предполагает выход в сигнальные пространства большей размерности. Это приводит к схемам модуляции с ортогональными сигналами.

3.5.1 Ортогональные сигналы с финитным спектром

Рассмотрим комплексное линейное пространство сигналов со спектром, финитным в частотной области на интервале $(-F/2, F/2)$. Набор функций

$$\phi_k(f) = \frac{1}{\sqrt{F}} \text{rect}_F(f) e^{-ikTf}, \quad TF = 1, \quad k \in [-\infty, +\infty]$$

образует полную ортонормированную систему на этом интервале, так что любая функция $X(f)$ с финитным спектром может быть представлена рядом

$$X(f) = \sum_{k=-\infty}^{+\infty} x_k \phi_k(f), \quad (14)$$

в котором легко опознается обычный ряд Фурье с коэффициентами x_k . Перенесем это разложение во временную область. Базисным функциям $\phi_k(f)$ во временной области отвечают сдвинутые во времени преобразования Фурье прямоугольного импульса $\text{rect}_F(f)$

$$\phi_k(t) = \sqrt{F} \frac{\sin \pi F(t - kT)}{\pi F(t - kT)} = \sqrt{F} \text{sinc} \pi F(t - kT), \quad FT = 1, \quad k \in [-\infty, +\infty],$$

которые также образуют ортонормированную систему. Временная форма ряда (14) принимает вид

$$x(t) = \sqrt{F} \sum_{k=-\infty}^{+\infty} x_k \text{sinc} \pi F(t - kT)$$

Обратим внимание, что $\text{sinc}(\pi FmT) = \delta_{m,0}$. Оценивая значение ряда $x(t)$ в точке $t = mT$, найдем: $x(mT) = \sqrt{F}x_m$. В итоге получим следующий **ряд Котельникова** для пространства комплексных огибающих с финитным спектром

$$x(t) = \sum_{k=-\infty}^{+\infty} x(kT) \text{sinc} \pi F(t - kT) \quad (15)$$

Он дает представление любой функции $x(t)$ с двухсторонней полосой F по дискретному набору ее выборочных значений $x(kT)$, взятых с шагом $T = 1/F$. В области реальных сигналов ряд Котельникова принимает вид

$$x_p(t) = \sqrt{2} \sum_{k=-\infty}^{+\infty} x_i(kT) \text{sinc} \pi F(t - kT) \cos f_0 t - x_q(kT) \text{sinc} \pi F(t - kT) \sin f_0 t,$$

где $x_i(nT)$, $(x_q(nT))$ - вещественная (мнимая) части выборочных значений комплексной огибающей.

Полученный результат достоин того, чтобы быть сформулированным в качестве отдельного предложения.

Теорема 3.2 (Теорема Котельникова) Сигнал с финитным в двухсторонней полосе F спектром, вполне определяется последовательностью комплексных выборочных значений, взятых с временным шагом $T = 1/F$. Минимальная частота дискретизации без потерь информации составляет F комплексных или $2F$ вещественных выборок в секунду.

Сейчас важно то, что для пространства функций времени с финитным спектром имеется ортонормированная система $\{\phi_k(t)\}$, которая позволяет представить любую функцию в виде разложения (15). Эта система бесконечна. Выберем из нее конечный базис, включающий N функций с индексами $k = [0, N - 1]$ и рассмотрим натянутое на него конечномерное пространство

$$x(t) = \sqrt{F} \sum_{k=0}^{N-1} x_k \frac{\sin \pi F(t - kT)}{\pi F(t - kT)}.$$

Все базисные функции имеют строго финитный спектр в полосе $(-F/2, F/2)$. С другой стороны, они практически финитны также и во времени, поскольку базисные импульсы $\text{sinc} \pi F t$ быстро спадают. Эффективную временную длительность функций, представимых в выбранном конечном базисе, разумно оценить как

$$(N - 1)T + 2T (N + 1)T \simeq NT = T_s.$$

В оценку включены $(N - 1)$ расстояний между N sinc -импульсами плюс 2 защитные Т-интервала на границах. Получаем следующий принципиально значимый результат:

Теорема 3.3 (О размерности пространства сигналов) Пространство сигналов, одновременно финитных в частотной области в полосе F и во временной области на интервале длительности T_s **конечномерно**. Его комплексная размерность составляет $B = FT_s$, вещественная $2FT_s$.

Как правило, при конструировании сигналов имеются ограничения на базу $B = FT_s$ - произведение разрешенной длительности сигналов на выделенную полосу частот. При таких ограничениях мы всегда оказываемся в конечномерном сигнальном пространстве комплексной размерности $N = B$. Здесь и разворачивается игра.

Двойственность между временной и частотной областями, естественным образом приводит к построению системы ортогональных сигналов, строго финитных во времени. На интервале $(-T/2, T/2)$ следует выбрать базис в виде

$$\phi_k(t) = \frac{1}{\sqrt{T}} \text{rect}_T(t) e^{kFt}, \quad TF = 1, \quad k \in [-\infty, +\infty]$$

Это даст ряд Котельникова в частотной области - представление спектра сигнала через выборочные значения $X(kF)$

$$X(f) = \sum_{k=-\infty}^{+\infty} X(kF) \text{sinc} T(f - kF)$$

суммой sinc -импульсов, сдвинутых по частоте с шагом F . В конечном итоге все приведет к той же размерности пространства сигналов - $B = TF_s$.

Упражнение 23 Восстановить опущенные детали анализа для частотной области.

Помимо системы ортогональных гармонических функций e^{kFt} с шагом F на интервале $(-T/2, T/2)$ можно построить когерентно ортогональный базис, образованный гармоническими функциями с вдвое меньшим шагом по частоте.

Теорема 3.4 *Функции*

$$\psi_k(t) = (i)^k e^{k\frac{F}{2}t}, \quad FT = 1$$

когерентно ортогональны на $(-T/2, T/2)$.

Доказательство. Рассмотрим скалярное произведение $\langle \psi_j, \psi_k \rangle$. Пусть $n = j - k$. Имеем,

$$\langle \psi_j, \psi_k \rangle = \int_{-T/2}^{T/2} i^n e^{n\frac{F}{2}t} = i^n T \frac{\sin \pi n \frac{F}{2} T}{\pi n \frac{F}{2} T}$$

При четных $n = 2m$ это равно нулю в силу обращения в нуль синуса. Это некогерентная (настоящая) ортогональность. При нечетных $n = 2m + 1$ скалярное произведение чисто мнимо. А это когерентная ортогональность, которая становится настоящей в P-области ▀

Разумеется, по когерентно ортогональным функциям можно рассматривать только ряды с вещественными коэффициентами x_n . Такой ряд

$$x(t) = \sum_n \hat{x}_n \psi_n = \sum_m x_{2m} e^{mFt} + i \sum_m x_{2m+1} e^{(m+\frac{1}{2})Ft}, \quad (16)$$

где $x_{2m} = (-1)^m \hat{x}_{2m}$, $x_{2m+1} = (-1)^m \hat{x}_{2m+1}$.

дает представление комплексного $(-T/2, T/2)$ -сигнала $x(t)$ набором вещественных амплитуд x_n гармоник с шагом $F/2$ по частоте. Вещественная и мнимая части $x(t)$ представляются при этом рядами

$$Re(x(t)) = \sum_m x_{2m} \cos mFt - \sum_m x_{2m+1} \sin(m + \frac{1}{2})Ft,$$

$$Im(x(t)) = \sum_m x_{2m} \sin mFt + \sum_m x_{2m+1} \cos(m + \frac{1}{2})Ft.$$

Упражнение 24 Преобразовать ряд (16) в частотную область. Найти форму соответствующего ряда Котельникова.

Таким образом, на интервале T_s комплексный сигнал представляется либо N комплексными коэффициентами при гармонических сигналах с шагом $F = 1/T_s$ частоте, либо $2N$ вещественными коэффициентами при гармониках с шагом $F/2$. Разумеется, обе формы представления дают одну и ту же вещественную размерность сигнального пространства $B = 2F_s T_s$, $F_s = NF$.

3.5.2 Модуляция ортогональными сигналами

Для канала с заданной базой B выберем ту ли иную ортонормированную систему сигналов e_1, \dots, e_N размерности $N = B$ и будем передавать данные, выбирая один сигнал из этого набора. Так получаются ортогональные системы модуляции. Они обладают следующими замечательными характеристиками.

Все сигналы имеют равную, единичную энергию. Они эквидистантны. Евклидово расстояние между любой парой сигналов составляет

$$\|e_i - e_k\| = \sqrt{2}.$$

Обеспечивается скорость передачи в $\log_2 B$ битов на сигнал. Спектральная эффективность ρ составляет при том $\log_2 B / B$ битов в секунду на герц полосы и стремится к нулю с ростом базы.

Ортогональные сигналы хороши в смысле энергетики, но никуда не годятся в плане спектральной эффективности. Область применения - дальняя космическая связь, где практически нет ограничений на полосу.

Низкая спектральная эффективность обусловлена тем, что мы выбираем для использования слишком малую долю сигналов. Базис - это далеко не все в N -мерном пространстве. Встает проблема выбора в пространстве N -измерений как можно большего числа точек с заданными ограничениями на энергию и минимальное расстояние. Отметим, что именно на этом пути достигается известная граница Шеннона пропускной способности канала связи.

Упражнение 25 По ортонормированной системе e_i , $i = [1, N]$ строят систему комплексных сигналов

$$s_i = e_i - \frac{\sum_{k=1}^N e_k}{N}$$

Показать, что при том же расстоянии эти сигналы обладают несколько меньшей энергией $1 - 1/N$. Объяснить геометрический смысл построения.

Рассмотренные выше системы ортонормированных функций дают начало двум двойственным системам ортогональной модуляции. Это системы с разделением по времени OTDM (Orthogonal Time Division Modulation) и OFDM (Orthogonal Frequency Division Modulation). Первая использует *sinc*-подобные импульсы, ортогонально сдвинутые во времени. Вторая - гармонические сигналы, ортогонально сдвинутые по частоте.

Возможно безумное количество разных ортогональных систем модуляции - по числу ортонормированных базисов в N -мерном комплексном пространстве. Вопрос в выборе системы, допускающей простую аппаратную реализацию.

Упражнение 26 (Ортогональная система Уолша) Натуральные числа k от 0 до $2^m - 1$ в двоичной позиционной системе представляются m -разрядными двоичными кодами.

Пусть $\bar{k} = (k_0, k_1, \dots, k_{m-1})$ - двоичный позиционный код числа k . Введем скалярное произведение

$$(\bar{j}, \bar{k}) = \sum_0^{m-1} j_i k_i$$

и рассмотрим систему целочисленных функций Уолша

$$\zeta_j(k) = e^{i\pi(\bar{j}, \bar{k})}, \quad j = [0, 2^m - 1]$$

Проверить, что они образуют ортогональную систему функций

$$\sum_{k=0}^{2^m-1} \zeta_i(k) \zeta_j^*(k) = 2^m \delta_{i,j}$$

Ортогональная система сигналов Уолша $x_j(t)$ строится из любой удобной ортонормированной системы $\phi_k(t)$ ортогональным преобразованием базиса

$$x_j(t) = \sum_{k=0}^{2^m-1} \zeta_j(k) \phi_k(t), \quad j = [0, 2^m - 1]$$

В качестве $\phi_k(t)$ могут быть взяты представители из систем OTDM

$$\phi_k(t) = \sqrt{F} \frac{\sin \pi F(t - kT)}{\pi F(t - kT)}$$

или OTFM

$$\phi_k(f) = \frac{1}{\sqrt{T}} \text{rect}_T(t) e^{+ikFt}$$

3.5.3 Принципы разделения радиоканала

Радиоканал является общим ресурсом. Необходимо как-то организовать его разделение его между многими пользователями. Представление об ортогональных системах сигналов дает концептуальную основу для обсуждения путей организации множественного доступа в канал.

Как мы знаем, радиоканал с ограниченными частотным и временными ресурсами является конечномерным линейным пространством комплексной размерности $N = F_s T_s$. Естественный способ его разделения - это ввести в нем ортогональную систему сигналов в количестве N штук и раздавать эти сигналы пользователям по потребности. Ортогональность гарантирует, что пользователи не будут мешать друг другу, если каждый будет фильтровать эфир на свою систему корреляторов.

Системы ортогональных сигналов OTDM и OTFM ведут к давно используемым методам разделения. Системе OTDM отвечает разделение временного ресурса канала. Абоненту выделяется временное окно, в течение которого имеет право использовать весь частотный ресурс. Это метод OTDMA (Orthogonal Time Division Multiple Access). Система сигналов OFDM приводит к схеме разделения канала по частоте. Это система OFDMA (Orthogonal Frequency Division Multiple Access). Пользователю навсегда выделяется часть частотного ресурса канала. На сегодняшний день это самый распространенный вариант разделения.

Можно делить ресурсы канала иначе. К примеру, возьмем систему сигналов Уолша, каждый из которых использует частотный и временной ресурс в полной мере, и поделим эти сигналы между пользователями. Придем с системе кодового разделения канала OCDMA (Orthogonal Code Division Multiple Access).

3.6 Спектральный анализ радиосигналов бесконечной энергии

Линейно модулированный сигнал

$$x(t) = \sum_{n=-\infty}^{\infty} c_n p(t - nT)$$

не имеющий ни начала, ни конца во времени, не обладает конечной энергией. Становится неясно, как вообще может идти речь о спектральных свойствах таких сигналов.

Проблема решается введением новых понятий - спектральной плотности мощности и мощностной корреляционной функции.

Рассмотрим «урезанный» вариант сигнала $x(t)$

$$x_W(t) = \sum_{n=0}^{N-1} c_n p(t - nT), \quad W = NT \quad (17)$$

Это уже вполне нормальный сигнал с конечной энергией. Для него корректно определены корреляционная функция $R_W(t) = \langle x_W, x_W(t) \rangle$ и энергетический спектр $|X_W(f)|^2$, связанные преобразованиями Фурье. Теперь следует перейти к пределу по длительности окна: $W \rightarrow \infty$. Чтобы этот предел не уходил в бесконечность, приходится все нормировать на длительность окна W . В результате приходим к понятиям

1. Мощностоной автокорреляционной функции

$$R(t) = \lim_{W \rightarrow \infty} \frac{R_W(t)}{W} \quad (18)$$

2. Спектральной плотности мощности

$$X(f) = \lim_{W \rightarrow \infty} \frac{|X_W(f)|^2}{W} \quad (19)$$

Новые понятия не следует путать с ранее введенными энергетическим спектрами и корреляциями. Мощностные корреляции и спектр - это совершенно новые категории, только с похожими названиями. У них даже размерности другие.

Таблица 1. Физические размерности в теории сигналов

Величина	Обозначение	Размерность	Комментарий
Время	t	$[T] = 1/[F]$	Секунда
Частота	f	$[F] = 1/[T]$	Герц
Уровень	$x(t)$	$[V]$	Вольты (амперы, ...)
Спектр амплитудный	$X(f)$	$[V][T] = [V]/[F]$	Вольты на герц
Квадрат уровня	$x^2(t)$	$[V]^2$	Мощность
Корреляция по энергии	$\langle x, y \rangle$	$[V]^2[T]$	Энергия
Спектр энергии	$ X(f) ^2$	$[V]^2[T]/[F]$	Энергия на герц
Корреляция по мощности	$R(\tau)$	$[V]^2$	Мощность
Спектр мощности	$X(f)$	$[V]^2[T] = [V]^2/[F]$	Энергия

Исследуем введенные мощностные понятия более пристально. Найдя квадрат модуля преобразование Фурье усеченного сигнала (17), придем к следующему выражению для энергетического спектра

$$|X_W(f)|^2 = |P(f)|^2 \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} e^{-(j-k)Tf} c_j c_k^*,$$

где $|P(f)|^2$ - энергетический спектр базового импульса $p(t)$, а суммирование ведется по $N \times N$ -квадрату. Переидем к суммированию по диагонали $n = j - k$. Получим

$$|X_W(f)|^2 = \sum_{n=-(N-1)}^{+(N-1)} e^{-nTf} \left\{ \sum_{[0,N-1]} c_{n+k} c_k^* \right\} \quad (20)$$

Чтобы не возиться с пределами во внутренней сумме, мы ввели новое обозначение $\sum_{[0,N-1]}$, имея в виду, что в сумму входят только те члены, у которых оба индекса $(k, n + k)$

принадлежат диапазону $[0, N - 1]$. Фактически, тем самым мы рассматриваем блок информационных символов $[c_0, c_1, \dots, c_{N-1}]$ как дополненный нулями слева и справа до бесконечности. Внутренняя сумма хорошо известна специалистам по цифровой обработке сигналов. Это авто корреляционная функция символьного блока

$$R_c^W(n) = \frac{1}{N} \sum_{[0, N-1]} c_{n+k} c_k^*.$$

Формально введем также и авто корреляционную функцию бесконечного потока символов c_n как предел

$$R_c(n) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=-\infty}^{\infty} c_{n+k} c_k^*. \quad (21)$$

Переходя к пределу по N в (20), найдем

$$X(f) = \frac{|P(f)|^2}{T} \sum_{n=-\infty}^{\infty} R_c(n) e^{-nTf} \quad (22)$$

представление спектра мощности сигнала через нормированный на T энергетический спектр базового импульса и авто корреляционную функцию символьного потока.

Упражнение 27 Проверьте справедливость аналогичного представления для корреляционной функции сигнала $x(t)$

$$R(t) = \frac{1}{T} \sum_{n=-\infty}^{\infty} R_p(t - nT) R_c(n) \quad (23)$$

где $R_p(t) = \langle p, p(t) \rangle$ - корреляционная функция импульса.

Упражнение 28 Показать, что авто корреляционная функция свертки двух сигналов есть свертка их авто корреляционных функций. (Перейти в частотную область). Дать интерпретацию результата (23) в свете этого факта.

Теорема 3.5 Если определение (21) автокорреляционной функции символьного потока корректно, то спектр мощности (22) и корреляционная функция (23) сигнала $x(t)$ корректно определены и связаны преобразованием Фурье.

Смысл суммы в правой части (22) таков. Представим себе, что корреляционная функция символьного потока $R_c(n)$, определенная только для целых n , на самом деле существует как функция непрерывного времени $R_c(t)$. Просто мы имеем дело с ее дискретизированной версией $R_c(n) = R_c(nT)$. Пусть $X_c(f)$ спектр гипотетической корреляционной функции $R_c(t)$. В частотной области дискретизации функции отвечает периодическое продолжение ее спектра

$$\sum R_c(nT) \delta(t - nT) \leftrightarrow \sum R_c(n) e^{-nTf} = F \sum X_c(f - mF)$$

Окончательно, выражение (22) для спектра мощности представляется в виде

$$X(f) = \frac{|P(f)|^2}{T^2} \sum_{n=-\infty}^{\infty} X_c(f - mF) \quad (24)$$

через энергетический спектр импульса и периодически продолженный спектр корреляционной функции символного потока.

Этот результат демонстрирует возможность формирования спектра мощности модулированного сигнала $X(f)$ выбором спектра корреляционной функции символного потока - $X_c(f)$. Возможность эта реализуется методами решетчатой модуляции. Разумным выбором автомата решетчатого модулятора формируются символные потоки, функции корреляции которых отличаются компактным спектром. Это приводит к повышению компактности спектра мощности радиосигнала в целом.

Упражнение 29 Пусть $c(t)$ - сигнал конечной энергии, $c_d(t)$ - результат его дискретизации.

$$c_d(t) = c(t) \sum_n \delta(t - nT) = \sum_n c(nT) \delta(t - nT).$$

Проверьте, что

$$\langle c_d, c_d(\tau) \rangle = \sum_n \left\{ \sum_k c_n c_{n+k}^* \right\} \delta(nT - \tau) = \sum_n R_c(n) \delta(nT - \tau)$$

Упражнение 30 Корреляционную функцию символного потока $R_c(t)$ всегда можно выбрать так, чтобы при периодическом продолжении реплики спектра $X_c(f)$ не перекрывались. Почему? Намек: покажите, что для всякой последовательности выборочных значений с шагом T существует интерполирующая функция со спектром, финитным в полосе F .

Упражнение 31 Проверьте, что в представлениях (22) и (24) соблюдены размерности.

С корреляционной функцией бесконечного символного потока (21) все не просто. И дело здесь не в проблеме существования предела, а в том, что вычисление этой функции подразумевает наличие какого-либо алгоритмического описания бесконечной последовательности символом (типа формулы общего члена ряда). Такое описание, разумеется, существует только для очень ограниченного класса последовательностей. Вообще, детерминистские модели оказываются малопригодными применительно к задаче описания символьных потоков.

Это обуславливает актуальность вероятностных моделей. Будем рассматривать символы информационного потока как случайные. Это предполагает описание потока некоторыми распределениями вероятностей символов. Пусть для каждой пары c_k, c_{k+n} задано совместное распределение $Q_n(c_k, c_{k+n})$, зависящие только от меры удаленности символов n (условие стационарности). Введем вероятностную корреляционную функцию

$$\tilde{R}_c(n) = \sum_{c_0, c_n} Q_n(c_0, c_n) c_0 c_n^*$$

как среднее по ансамблю.

Случайную последовательность называют эргодической если для нее среднее по ансамблю $\tilde{R}_c(n)$ совпадает со средним по времени R_c (21). Свойство эргодичности позволяет подменить в представлениях (22) и (24) результат усреднения по времени эквивалентным средним по ансамблю. В результате мощностная корреляция и спектр мощности оказываются определенными для всех случайных стационарных эргодических последовательностей.

Таким образом, использование стохастических моделей информационных потоков, переводя сигналы в ранг случайных процессов, с неизбежностью приводит к использованию мощностных категорий корреляции и спектра. Фактически, эти категории составляют фундамент корреляционной теории случайных процессов.

Упражнение 32 Пусть символы бинарного информационного потока $c_i = \pm 1$ случайны, равновероятны ($Q(1) = Q(-1) = 1/2$) и независимы

$$Q(c_k, c_{k+n}) = Q(c_k)Q(c_{k+n}).$$

Покажите, что для статистической корреляционной функции справедливо

$$\tilde{R}_c(n) = \delta_{n,0}.$$

Используя (22), покажите, что спектральная плотность сигнала $x(t)$, линейно модулированного этим потоком имеет вид

$$X(f) = \frac{|P(f)|^2}{T},$$

где $P(f)$ - амплитудный спектр базового импульса $p(t)$

4 Корреляционная теория шумов

4.0.1 Экскурс в теорию вероятностей

Описание вещественной случайной величины x плотностью вероятностей $p(x)$ позволяет для произвольной функции $F(x)$ найти ее математические ожидание (*статистическое среднее*)

$$E(F(x)) = \int F(x)p(x) dx.$$

В частности, вводится среднее значение самой случайной величины $m = E(x)$ и ее момент $E((x - m)^n)$ порядка n .

Соглашение. Для простоты далее будем считать, что все рассматриваемые случайные величины обладают нулевым средним. Это не ограничивает общности, поскольку случайная величина с ненулевым средним m всегда может быть представлена в виде $x' = x + m$ с распределением $p(x' - m)$.

В корреляционной теории ведущую роль играют моменты второго порядка. Для единственной случайной величины такой момент всего один - это ее дисперсия $\sigma^2 = E(x^2)$.

Для нас случайная величина - это выборочное значение шума, среднеквадратичное (эффективное) значение $\sigma = \sqrt{E(x^2)}$ - это его уровень, а дисперсия σ^2 - мощность.

Случайную величину x с плотностью вероятностей

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (25)$$

называют гауссовской. Дисперсия x составляет σ^2 и вполне определяет плотность вероятностей.

4.0.2 Вещественные случайные векторы

Конечный набор случайных величин будем рассматривать как вектор

$$\mathbf{x} = (x_1, x_2, \dots, x_n).$$

Для n случайных величин x_j существует n^2 моментов второго порядка $E(x_j x_k)$. Все они упаковываются в матрицу корреляций

$$\mathbb{C} = E(\mathbf{x}^T \mathbf{x}) = \left\{ E(x_j x_k) \right\}.$$

Здесь \mathbf{x}^T - вектор-столбец - результат транспонирования \mathbf{x} . (*Умножение столбца на строку дает матрицу*).

Из определения ясно, что матрица корреляций симметрична.

По аналогии, для двух случайных векторов \mathbf{x} и \mathbf{y} вводится матрица взаимных корреляций

$$\mathbb{C}_{xy} = E(\mathbf{x}^T \mathbf{y}) = \left\{ E(x_j y_k) \right\}.$$

Ясно, что

$$\mathbb{C}_{yx} = \mathbb{C}_{xy}^T$$

Матрица корреляций позволяет найти дисперсию произвольной линейной формы

$$y = \sum a_j x_j = \mathbf{x} \mathbf{a}^T \quad (26)$$

в виде

$$E(yy) = \langle y|y \rangle = \mathbf{a} \mathbb{C} \mathbf{a}^T \quad (27)$$

Отсюда следует положительная определенность матрицы корреляций

$$\mathbf{a} \mathbb{C} \mathbf{a}^T \geq 0, \quad \forall \mathbf{a} \neq 0$$

Определение 1 Случайный вектор \mathbf{x} назовем гауссовским, если любая линейная форма от его компонент (26) дает Гауссовскую случайную величину с дисперсией (27). Будем говорить, что компоненты гауссовского случайного вектора образуют набор совместно гауссовых случайных величин.

Простым следствием этого определения становится следующий результат о сохранении свойства гауссности при линейных преобразованиях.

Теорема 4.1 Пусть \mathbf{x} - гауссовский вектор, A - невырожденная матрица. Тогда вектор $\mathbf{y} = \mathbf{x}A$ также гауссовский с матрицей корреляций

$$\mathbb{C}_y = A^T \mathbf{x}^T \mathbf{x} A = A^T \mathbb{C}_x A$$

Следующий результат показывает, что для гауссовского случайного вектора плотность вероятностей определяется матрицей корреляций.

Теорема 4.2 Совместная плотность вероятностей распределения компонент гауссовского случайного вектора имеет вид

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det \mathbb{C}}} e^{-\mathbf{x} \mathbb{C}^{-1} \mathbf{x}^T}$$

Часто эту форму плотности вероятностей используют для определения понятия совместной гауссности. Тогда свойство гауссности линейных форм (26) становится теоремой.

Упражнение 33 Компоненты случайного вектора не коррелированы, если матрица корреляций - диагональна. Показать, что для гауссовского случайного вектора некоррелированность компонент эквивалентна их статистической независимости в смысле

$$p(\mathbf{x}) = p(x_1)p(x_2) \dots p(x_n).$$

4.0.3 Комплексные случайные векторы

Имея пару вещественных случайных векторов \mathbf{x} и \mathbf{y} , можем рассмотреть комплексный вектор $\mathbf{z} = \mathbf{x} + i\mathbf{y}$. Для комплексного случайного вектора можно ввести матрицу корреляций

$$\mathbb{C}_z = E(\mathbf{z}^+ \mathbf{z}) = E(\mathbf{x}^T \mathbf{x}) + E(\mathbf{y}^T \mathbf{y}) + i E(\mathbf{x}^T \mathbf{y}) - i E(\mathbf{y}^T \mathbf{x}),$$

где \mathbf{z}^+ - сопряженный и транспонированный вектор, и матрицу псевдокорреляций

$$\tilde{\mathbb{C}}_z = E(\mathbf{z}^T \mathbf{z}) = E(\mathbf{x}^T \mathbf{x}) - E(\mathbf{y}^T \mathbf{y}) + i E(\mathbf{x}^T \mathbf{y}) + i E(\mathbf{y}^T \mathbf{x}).$$

Определение 2 Случайный вектор \mathbf{z} назовем **нормальным**, если его матрица псевдокорреляций равна нулю.

Для нормального комплексного вектора матрицы корреляций вещественной и мнимой компонент удовлетворяют условиям

$$E(\mathbf{x}^T \mathbf{x}) = E(\mathbf{y}^T \mathbf{y}), \quad E(\mathbf{x}^T \mathbf{y}) = -E(\mathbf{y}^T \mathbf{x}),$$

а комплексная матрица корреляций приобретает вид

$$\mathbb{C}_z = 2\mathbb{C}_z^i + 2i\mathbb{C}_z^q; \quad \mathbb{C}_z^i = E(\mathbf{x}^T \mathbf{x}) = E(\mathbf{y}^T \mathbf{y}), \quad \mathbb{C}_z^q = E(\mathbf{x}^T \mathbf{y}) = -E(\mathbf{y}^T \mathbf{x})$$

Упражнение 34 Проверить, что нормальность вектора \mathbf{z} эквивалента инвариантности вещественных матриц корреляций

$$E(\mathbf{x}^T \mathbf{x}), \quad E(\mathbf{y}^T \mathbf{y}), \quad E(\mathbf{y}^T \mathbf{x})$$

относительно поворота вектора \mathbf{z} на угол θ

$$\mathbf{z}' = \mathbf{z}e^{i\theta}$$

$$\mathbf{x}' = \mathbf{x} \cos\theta - \mathbf{y} \sin\theta, \quad \mathbf{y}' = \mathbf{x} \sin\theta + \mathbf{y} \cos\theta.$$

Упражнение 35 Проверить, что свойство нормальности сохраняется при несъроожденных линейных преобразованиях $\mathbf{z}' = \mathbf{z}A$ с комплексной матрицей A .

Дальнейшее для нормальных случайных векторов отличается от вещественного случая лишь в незначительных деталях.

Матрица корреляций нормального вектора сопряженно симметрична (Эрмитова), ее вещественная часть - симметрична, а мнимая - антисимметрична

$$\mathbb{C}_z^+ = \mathbb{C}_z, \quad (\mathbb{C}_z^i)^T = \mathbb{C}_z^i, \quad (\mathbb{C}_z^q)^T = -\mathbb{C}_z^q$$

В частности, для одномерного нормального вектора - комплексной случайной величины $z = x + iy$, имеем

$$\begin{aligned} E(xx) &= E(yy) = \sigma^2, \quad E(xy) = -E(yx) = 0, \\ E(zz^*) &= 2\sigma^2 \end{aligned}$$

Нормальная комплексная случайная величина называется гауссовской, если ее плотность вероятностей имеет

$$p(z) = \frac{1}{2\pi\sigma^2} e^{-\frac{|z|^2}{2\sigma^2}}$$

Вещественная и мнимая части гауссовского случайного комплексного числа независимы и распределены по гауссовскому закону с дисперсией σ^2 . В полярных координатах $z = |z|e^{i\varphi}$ получаем равномерное распределение для фазы φ и распределение Релея для амплитуды $|z|$.

$$p(|z|) = \frac{|z|}{\sigma^2} e^{-\frac{|z|^2}{2\sigma^2}}.$$

Линейная форма

$$y = \sum a_j z_j = \mathbf{z} \mathbf{a}^T$$

от компонент комплексного вектора \mathbf{z} дает нормальную комплексную случайную величину y с дисперсией

$$E(y^*y) = \langle y|y \rangle = \mathbf{a} \mathbb{C}_z \mathbf{a}^+ \tag{28}$$

Это обеспечивает положительную определенность матрицы корреляций.

Определение 3 Случайный комплексный вектор \mathbf{z} назовем гауссовским, если любая линейная форма от его компонент дает комплексную гауссовскую случайную величину с дисперсией (28).

Теорема 4.3 Совместная плотность вероятностей распределения компонент комплексного гауссовского случайного вектора вполне определяется матрицей корреляций и имеет вид

$$p(\mathbf{z}) = \frac{1}{\pi^n \det \mathbb{C}} e^{-\mathbf{z}^* \mathbb{C}^{-1} \mathbf{z}}$$

Упражнение 36 Показать, что матрица корреляций двухмерного нормального вектора $(z_1, z_2) = (x_1 + i y_1, x_2 + i y_2)$ имеет вид

$$\mathbb{C}_z = \begin{vmatrix} \sigma_1^2 & \varrho^2 + i \gamma^2 \\ \varrho^2 - i \gamma^2 & \sigma_2^2 \end{vmatrix},$$

где

$$\begin{aligned} \sigma_1^2 &= E(x_1 x_1) = E(y_1 y_1), & \sigma_2^2 &= E(x_2 x_2) = E(y_2 y_2) \\ \varrho^2 &= E(x_1 x_2) = E(y_1 y_1), \\ \gamma^2 &= E(x_1 y_2) = E(x_2 y_1). \end{aligned}$$

4.0.4 Случайные процессы

Случайный процесс $x(t)$ (*шум*) - это комплексная функция времени, значение которой в каждой точке есть комплексная случайная величина. Не ограничивая общности, будем рассматривать процессы с нулевым средним.

Полное описание случайного процесса предполагает, что для каждого n точек (t_1, t_2, \dots, t_n) задана совместная плотность вероятностей распределения выборочных значений. Для целей корреляционной теории достаточно ограничиться случаем $n = 2$.

Итак, пусть для каждого (t_1, t_2) задана двухмерная плотность вероятностей $p(x(t_1), x(t_2)) = p(x_1, x_2)$, удовлетворяющая естественным условиям согласованности

$$p(x) = \int p(x_1, x_2) dx_1 = \int p(x_1, x_2) dx_2.$$

Случайный процесс **стационарен**, если плотность вероятностей p не изменяется при сдвигах во времени, то есть зависит только от разности $\tau = t_2 - t_1$,

$$p(x(t_1), x(t_2)) = p_\tau(x_1, x_2) = p_{-\tau}(x_2, x_1)$$

Эта двухмерная плотность определяет 2×2 -матрицу корреляций случайного вектора (x_1, x_2) . Единственный ее недиагональный компонент дает статистическую корреляционную функцию случайного процесса $x(t)$

$$R_x(\tau) = \langle x_1 | x_2^*(\tau) \rangle = E(x_1 x_2^*) = \int \int x_1 x_2^* p_\tau(x_1, x_2) dx_1 dx_2$$

Преобразование Фурье корреляционной функции назовем спектральной плотностью $X(f)$ случайного процесса (шума) $x(t)$. Итак, для случайного процесса определены **корреляционная функция и спектральная плотность**

$$R_x(t) \leftrightarrow X(f).$$

Это базовые понятия для корреляционной теории шумов.

Свойства симметрии.

Из определения следует, что корреляционная функция сопряженно симметрична

$$R_x(-\tau) = R_x^*(\tau).$$

Отсюда заключаем, что спектральная плотность $X(f)$ всегда вещественна (*см. свойства симметрии преобразования Фурье*).

Если шум *вещественен*, то корреляционная функция также вещественна, а следовательно, симметрична. Тогда симметрична также и спектральная плотность. То есть, в вещественном случае и корреляционная функция и спектральная вещественны и симметричны. При этом они оказываются связанными косинус-преобразованием.

В *комплексном* случае корреляционная функция $R_x(t) = R_i(t) + iR_q(t)$, вообще говоря, комплексна. Однако, ее сопряженная симметричность гарантирует симметрию вещественной и антисимметрию мнимой частей,

$$R_i(-t) = R_i(t), \quad R_q(-t) = -R_q(t).$$

Вещественная спектральная плотность $X(f)$ может быть представлена суммой симметричной и антисимметричной компонент

$$X(f) = X_s(f) + X_a(f) = \frac{X(f) + X(-f)}{2} + \frac{X(f) - X(-f)}{2}$$

Симметричная часть - есть Фурье образ вещественной части корреляционной функции, а антисимметричная - мнимой,

$$R_i(t) \leftrightarrow X_s(f), \quad R_q(t) \leftrightarrow X_a(f).$$

Значение корреляционной функции в нуле всегда вещественно. Оно дает оценку мощности шума. Этую же мощность можно оценить как интеграл от спектральной плотности

$$\sigma^2 = R_x(0) = \int |X(f)|^2 df$$

По аналогии вводится взаимная корреляционная функция двух случайных процессов

$$R_{xy}(\tau) = \langle x|y^*(\tau)\rangle = E(xy^*).$$

В расчет берется выборочное значение одного шума в момент времени t_1 и другого - в момент t_2 , $t_2 - t_1 = \tau$. Для вычисления среднего требуется совместное распределение $p(x(t_1), y(t_2))$. Взаимная корреляционная функция не отличается особой симметрией. Верно лишь, что

$$R_{xy}(-\tau) = R_{yx}^*(\tau).$$

Фурье образ взаимной корреляционной функции - это взаимная спектральная плотность

$$R_{xy}(\tau) \leftrightarrow X_{xy}(f)$$

со свойством симметрии

$$X_{xy}^*(f) = X_{yx}(f)$$

Сигналы называют некоррелированными при временном сдвиге τ , если $R_{xy}(\tau) = 0$. Чаще всего приходится иметь дело с некоррелированностью в данный момент ($\tau = 0$) и некоррелированностью при всех сдвигах. В частотной области первому условию отвечает равенство нулю интеграла от взаимной спектральной плотности, а второму - обращение в нуль самой этой плотности.

Подключаем условие гауссовости.

Определение 4 Случайный процесс $x(t)$ называют гауссовским, если для любого набора моментов времени (t_1, t_2, \dots, t_n) вектор выборочных значений $(x(t_1), x(t_2), \dots, x(t_n))$ является гауссовским случайным вектором.

Выполнение условия гауссности не является обязательным в корреляционной теории. Предположение о гауссности шумов дает лишь дополнительную возможность восстанавливать распределения вероятностей по корреляционному описанию.

К примеру, для просто шума знание значения корреляционной функции в нуле $R_x(0)$ - это оценка дисперсии σ^2 и только. Если же шум гауссовский, то дисперсия определяет плотность вероятностей целиком, что дает значительно больше, чем знание единственного момента второго порядка.

В Гауссовском случае корреляционная функция дает **исчерпывающее описание шума**, позволяя находить плотности вероятностей произвольной размерности.

В самом деле. Рассмотрим n -вектор выборочных значений

$$(x(t_1), x(t_2), \dots, x(t_n)) = (x_1, x_2, \dots, x_n).$$

Взаимная корреляция между любыми двумя его компонентами $E(x_i x_k^*)$ определяется значением корреляционной функции в точке $t_k - t_i$

$$E(x_i x_k^*) = R_x(t_k - t_i).$$

Следовательно, корреляционная функция позволяет найти все компоненты матрицы корреляций. Для гауссовского вектора это полностью определяет совместное распределение вероятностей.

Любая линейная форма от гауссовского вектора - это гауссовская случайная величина. Поэтому, предположение о гауссовой природе шума позволяет делать выводы о гауссости любых результатов его линейной обработки.

Согласно предельным теоремам теории вероятностей, сумма большого числа независимых случайных величин имеет тенденцию становиться гауссовой. С другой стороны, линейная обработка шума свойство гауссности сохраняет. Поэтому модель гауссовского шума оказывается совершенно естественной в областях, связанных с линейной обработкой сигналов.

Усреднение по времени, эргодический переход.

Введем временное окно $w(t) = \text{rect}_T(t)$ и с его помощью «вырежем» финитную во времени реализацию шума $x_T(t) = w(t)x(t)$. Это обычный сигнал с конечной энергией и для него определены корреляционная функция $R_w(\tau) = \langle x_w, x_w(\tau) \rangle$ и энергетический спектр $|X_w(f)|^2$. Переходя к пределу по длительности окна, введем мощностную корреляционную функцию

$$\mathbb{R}(\tau) = \lim_{T \rightarrow \infty} \frac{R_w(\tau)}{T}$$

и соответствующий ей спектр мощности

$$\mathbb{X}(f) = \lim_{T \rightarrow \infty} \frac{|X_w|^2(f)}{T}$$

Определение 5 Случайный процесс назовем **эргодическим**, если для него моменты (до второго порядка включительно), найденные усреднением по распределению вероятностей, совпадают с найденными усреднением по времени.

Свойство эргодичности играет ключевую роль в корреляционной теории, позволяя заменять корреляции $\mathbb{R}(\tau)$, введенные через интегрирование по времени, на статистические корреляционные функции $R(\tau)$. Эту замену будем называть эргодическим переходом.

Замечание 2 Вот простой пример не эргодического случайного процесса.

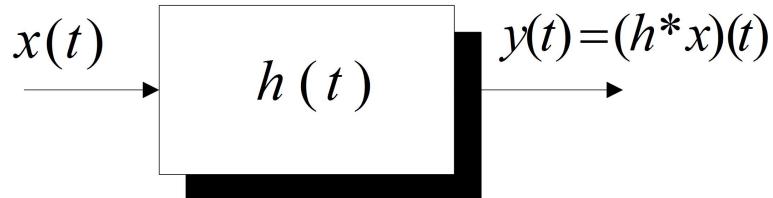
Пусть некто бросает монету, а затем, в зависимости от исхода, предъявляет либо реализацию шума $n(t)$, либо реализацию того же шума, но с аддитивно добавкой - $a+n(t)$. Статистическое среднее значение шума в этой модели равно $1/2 \times 0 + 1/2 \times a = a/2$. Результат же усреднения по времени случаен и принимает значение 0 или a с равной вероятностью.

Все не эргодические процессы устроены подобным образом. Средние по времени для них сами оказываются случайными величинами.

4.0.5 Фильтрация шумов, теорема Винера-Хинчина

Обратимся к задаче прохождения сигнала с конечной энергией через линейный фильтр с импульсной реакцией $h(t)$, рис. 12. Результатом фильтрации становится сигнал $y(t)$, который определяется сверткой $y(t) = (h * x)(t)$. В частотной области этому отвечает умножение амплитудного спектра сигнала на комплексный коэффициент передачи $H(f)$.

Замечание 3 Чтобы сигналы на входе и выходе фильтра были одной размерности, импульсную реакцию следует считать имеющей размерность обратного времени (частоты). Тогда комплексный коэффициент передачи оказывается безразмерным.



$$Y(f) = H(f)X(f)$$

Рис. 12: К фильтрации

Теорема 4.4 (О фильтрации корреляций.) Авто корреляционная функция сигнала $y(t)$ на выходе фильтра есть свертка авто корреляционной функции входа $x(t)$ с авто корреляционной функцией импульсной реакции фильтра.

$$\langle y, y(\tau) \rangle = (\langle h, h(\tau) \rangle * \langle x, x(\tau) \rangle)$$

Доказательство. Перейдем в частотную область. В левой части получим энергетический спектр. В правой имеем свертку двух корреляционных функций. В частотной области ей будет отвечать произведение их спектров

$$|Y(f)|^2 = \langle h, h(\tau) \rangle (f) \langle x, x(\tau) \rangle (f).$$

Но спектр корреляционной функции есть модуль квадрат спектра сигнала. Окончательно имеем,

$$Y(f)Y^*(f) = |H(f)|^2|X(f)|^2 = (H(f)X(f))(H^*(f)X^*(f)),$$

что верно, поскольку $Y(f)=H(f)X(f)$ ▼

Теперь перейдем к фильтрации шума. Пусть, как и выше, $x_w(t)$ - реализация шума, клиппированная окном длительности T . Поскольку это сигнал с конечной энергией, можем применить теорему о фильтрации корреляций

$$\langle y, y(\tau) \rangle = (\langle h, h(\tau) \rangle * \langle x_w, x_w(\tau) \rangle)$$

Переходя здесь пределу при $T \rightarrow \infty$, найдем

$$\mathbb{R}_y(\tau) = (\langle h, h(\tau) \rangle * \mathbb{R}_x(\tau))$$

Наконец, привлекая эргодичность шума, заменим фигурирующие здесь корреляционные функции с усреднением по времени корреляциями с усреднением по плотности вероятностей (эргодический переход). В итоге придем к следующей теореме.

Теорема 4.5 (Теорема Винера-Хинчина)

Во временной области.

Корреляционная функция сигнала на выходе линейного фильтра есть свертка корреляционной функции входного сигнала с корреляционной функцией импульсной реакции.

$$R_y(\tau) = (\langle h, h(\tau) \rangle * R_x(\tau)). \quad (29)$$

В частотной области.

Спектральная плотность шума на выходе фильтра есть произведение спектральной плотности входного шума на модуль квадрат коэффициента передачи фильтра

$$Y(f) = |H(f)|^2 X(f). \quad (30)$$

Теорема 4.6 (О паре фильтров)

Пусть имеется пара фильтров $h_1(t) \leftrightarrow H_1(f)$ и $h_2(t) \leftrightarrow H_2(f)$ и на их входах действуют шумы $x_1(t)$ и $x_2(t)$ со взаимной корреляционной функцией $R_{x_1 x_2}(\tau)$ (взаимной спектральной плотностью $X_{x_1 x_2}(f)$). Тогда сигналы на выходах $y_1(t) = (h_1 * x_1)(t)$ и $y_2(t) = (h_2 * x_2)(t)$:

Во временной области.

Обладают взаимно корреляционной функцией

$$R_{y_1 y_2}(\tau) = (\langle h_1, h_2(\tau) \rangle * R_{x_1 x_2}(\tau)), \quad (31)$$

равной свертке взаимной корреляционной функции входов со взаимной корреляционной функцией импульсных реакций фильтров или

В частотной области.

$$X_{y_1 y_2}(f) = H_1(f)H_2^*(f) X_{x_1 x_2}(f). \quad (32)$$

Имеют взаимную спектральную плотность, равную произведению взаимной спектральной плотности входов на взаимный энергетический спектр фильтров.

Упражнение 37 Провести доказательство теоремы о двух фильтрах по аналогии с теоремой Винера-Хинчина.

Фильтры называют некогерентно ортогональными, если их взаимный энергетический спектр тождественно равен нулю. Ортогональными в этом смысле будут, к примеру, два полосовые фильтра с неперекрывающимися полосами пропускания. Более слабое требование когерентной ортогональности предполагает лишь равенство нулю интеграла от взаимного спектра. Некогерентная ортогональность сохраняется при временных сдвигах импульсных реакций, когерентная - нет.

Упражнение 38 Показать, что если фильтры некогерентно ортогональны, то шумы на их выходе не коррелированы при любой природе входных шумов. Какие выводы о взаимно корреляционных свойствах выходных сигналов позволяет сделать когерентная ортогональность?

4.0.6 Белый шум. шумовая полоса

При анализе шумов крайне часто используется модель белого шума. Начнем с вещественного случая.

Определение 6 Белым шумом со спектральной плотностью $\mathcal{N}_0/2$ называют такой шум, который, действуя на входе любого линейного фильтра, дает на его выходе шум со спектральной плотностью

$$X(f) = \mathcal{N}_0/2 |H(f)|^2,$$

Результат фильтрации белого шума обладает, таким образом, корреляционной функцией

$$R(\tau) = \mathcal{N}_0/2 \int |H(f)|^2 e^{+ft} df,$$

и мощностью (дисперсией)

$$\sigma^2 = \mathcal{N}_0/2 \int |H(f)|^2 df$$

Из (30) следует, что спектральная плотность самого белого шума постоянна и составляет $\mathcal{N}_0/2$ на всех частотах. Соответственно, корреляционной функцией белого шума является δ -функция

$$\frac{\mathcal{N}_0}{2} \delta(\tau) \leftrightarrow \frac{\mathcal{N}_0}{2} 1(f)$$

Белый шум - это не более чем удобная абстракция хотя бы уже потому, что его мощность бесконечна. Если непременно хотят рассматривать шум на выходе фильтра как гауссовский, то в оправдание добавляют эпитет «гауссовский» к названию белого шума на входе. Получается знаменитый аддитивный белый гауссовский шум (АБГШ). Гауссовость здесь - не более чем игра слов.

Инженерам ближе односторонняя спектральная плотность \mathcal{N}_0 , которая вдвое больше двухсторонней. Она позволяет им лукаво оценивать мощности шума на выходе фильтра с полосой F как $\sigma^2 = \mathcal{N}_0 F$, «забывя» о наличии такой же полосы пропускания фильтра на отрицательных частотах. На самом деле $\sigma^2 = \frac{\mathcal{N}_0}{2} 2F$.

Комплексный белый шум мало чем отличается от вещественного. У него те же по форме спектральная плотность и корреляционная функция. Только понятие

односторонней плотности для него бессмысленно - в комплексной области спектры не симметричны. Потому пишут просто

$$\mathcal{N}_0 \delta(\tau) \leftrightarrow \mathcal{N}_0 1(f),$$

и оказываются правы, потому что спектральная плотность комплексного эквивалента белого шума как раз вдвое больше плотности вещественного.

Упражнение 39 *Пусть два фильтра когерентно ортогональны в том смысле, что интеграл по частоте от их взаимного спектра равен нулю.*

Покажите, что это условие эквивалентно ортогональности импульсных реакций.

Подключим входы фильтров к одному и тому же источнику белого шума.

Покажите, что шумы на выходах фильтров не коррелированы при нулевом сдвиге по времени.

Отсюда следует вывод о некоррелированности шумов на выходах ортогональных корреляторов.

Самая типовая инженерная задача - это оценивание мощности шума на выходе фильтра. Ввиду ее важности придумано специальное понятие - шумовая полоса.

Определение 7 *Шумовая полоса фильтра с коэффициентом передачи $H(f)$ это ширина прямоугольника единичного уровня, площадь которого равна площади под нормированной кривой $|H(f)|^2$*

$$F_n = \frac{\int |H(f)|^2 df}{\max_f |H(f)|^2}$$

Мощность шума на выходе оценивают, умножая спектральную плотность входного (белого) шума на шумовую полосу. Бывает не лишне учесть также квадрат коэффициента усиления фильтра.

Упражнение 40 *Оценить шумовую полосу интегрирующей цепи $H(f) = \frac{1}{1+i(f/f_0)}$ с частотой среза f_0 (Ответ - $\frac{\pi f_0}{2}$).*

Упражнение 41 *Оценить шумовую полосу полосового фильтра $H(f) = \frac{1}{1+i2Q(f-f_0/f_0)}$ с резонансной частотой f_0 и добротностью Q (Ответ - $\frac{\pi}{2} \frac{f_0}{Q}$).*

4.0.7 Корреляционная алгебра

Сложность освоения корреляционной теории шумов заключается в том, что люди сплошь и рядом говорят и сигналах, а имеют в виду их спектральные плотности или, что тоже самое, корреляционные функции. Положение усугубляется тем, что при обсуждении прикладных вопросов недвусмысленные обозначения типа «шум $n(t)$, его спектральная плотность $X_n(f)$, корреляционная функция $R_n(\tau)$ » быстро надоедают. На их место приходит некий «птичий язык», не всегда адекватно понимаемый новичком.

Поэтому, приступая к реальным задачам, от строгих обозначения лучше отойти сразу, введя специально оговоренный лаконичный язык.

Удобнее всего обозначать все сущности - шум, корреляционную функцию, спектральную плотность - одной буквой. Это уменьшает нагрузку на память, что при решении задач важно. Корреляции и спектральные плотности отличаются от шумов тем, что они квадратичны. Сами же корреляции и спектральные плотности находятся в

родстве. Это просто две формы представления одной сущности во временной и частотной областях.

Разумно поступать так. Шумы обозначать как есть - $n(t)$, а для спектральных плотностей (корреляций) ввести символические обозначения - $n^2(f), n^2(t)$. Символический квадрат присутствует всегда, отличая шум от не шума. А вот аргумент t, f можно опускать, если не важно о чём идет речь - о корреляции или плотности. Поставив аргумент, мы без лишних слов оказываемся во временной или частотной области. Использование этих обозначений ведет, правда, к нелепому на вид выражению для дисперсии

$$\sigma_n^2 = \int n^2(f) df = n^2(0)$$

На самом деле это всего лишь означает, что дисперсию шума можно оценить либо как интеграл от спектральной плотности $n^2(f)$, либо как значение в нуле корреляционной функции $n^2(t)$.

Будут встречаться также взаимные корреляции и взаимные спектральные плотности. Для них оставим скобочные обозначения - $\langle e_i | e_k \rangle(t), \langle e_i | e_k \rangle(f)$.

Задача корреляционной алгебры ставится так.

Имеется некоторая электрическая схема, содержащая, наряду с LRC -компонентами, источники шума $e(t)$ с известным корреляционным описанием. Требуется найти корреляционные характеристики шума $n(t)$ на выходе.

Предлагаемая процедура решения состоит в следующем.

На первом этапе считаем все источники обычными сигналами. Если цепь не содержит реактивных компонентов, то можно работать с постоянными напряжениями, если же таковые имеются, придется применить метод комплексных амплитуд. Цель первого этапа - найти решение в виде

$$n = \sum H_j(f) e_j \quad (33)$$

взвешенной суммы источников с коэффициентами передачи $H_i(f)$. Принцип суперпозиции гарантирует получение решения в этом виде для любой линейной системы.

Версия решения (33) для временной области имеет вид

$$n(t) = \sum (h_j(t) * e_j(t)),$$

суммы сверток временных форм сигналов с импульсными реакциями h_j фильтров, отвечающих за коэффициенты передачи H_j . (*Вычислять импульсные реакции реально не придется*).

Теперь приступаем к нахождению корреляционной функции $n(t)$. Имеем

$$n^2(t) = \langle n, n(t) \rangle = \sum_{j,k} \langle (h_j * e_j), (h_k * e_k)(t) \rangle = \sum_{j,k} (\langle h_j, h_k(t) \rangle * \langle e_j, e_k(t) \rangle)$$

Мы воспользовались здесь тем, что скалярное произведение сверток есть свертка скалярных произведений. В правой части получились временные корреляционные функции шумов. Проговариваем стандартные слова, сопровождающие эргодический переход, и приходим к тому же выражению, но уже со средними по плотностям вероятностей

$$n^2(t) = \sum_{j,k} (\langle h_j, h_k \rangle(t) * \langle e_j | e_k \rangle(t))$$

В частотной области это принимает более приятный вид

$$n^2(f) = \sum_{j,k} H_j(f) H_k^*(f) \langle e_j | e_k \rangle (f) \quad (34)$$

В принципе ответ, готов. Спектральная плотность результирующего шума представлена в виде взвешенной суммы взаимных спектральных плотностей всех пар источников. Весовыми коэффициентами являются взаимные энергетические спектры пар фильтров.

Пока еще все довольно сложно. К счастью, в задачах обычно предполагается, что разные источники шума не коррелированы. Тогда все недиагональные члены в (34) обращаются в нуль и мы получаем простой окончательный результат

$$n^2(f) = \sum_j |h_j(f)|^2 e_j^2(f), \quad (35)$$

который выражает тот факт, что спектральная плотность шума на выходе есть взвешенная сумма плотностей источников. В качестве весовых коэффициентов выступают квадраты модулей коэффициентов передачи. Это просто обобщение теоремы Винера-Хинчина на случай нескольких некоррелированных источников.

При решении задач переход от (33) к (35) обычно сопровождается ссылкой на очевидность.

Вот несколько простых примеров применения корреляционной алгебры.

Пример 2 Последовательное соединение двух шумовых источников ЭДС.

$$e(t) = e_1(t) + e_2(t).$$

Имеем форму (33) с единичными коэффициентами передачи. Привлекая (35), находим

$$e^2(f) = e_1^2(f) + e_2^2(f).$$

Это закон сложения спектральных плотностей некоррелированных источников.

Для общего случая коррелированных источников из (34) получим

$$\begin{aligned} e^2(f) &= e_1^2(f) + e_2^2(f) + \langle e_1 | e_2 \rangle (f) + \langle e_2 | e_1 \rangle (f) = \\ &= e_1^2(f) + e_2^2(f) + 2\operatorname{Re}\{\langle e_1 | e_2 \rangle (f)\} \end{aligned}$$

сумму спектральных плотностей плюс удвоенная вещественная часть взаимной спектральной плотности.

Пример 3 Параллельное соединение двух реальных шумовых источников ЭДС с внутренними сопротивлениями R_1, R_2 . Имеем

$$e(t) = h_1 e_1(t) + h_2 e_2(t)$$

где $h_1 = R_2/(R_1 + R_2)$, $h_2 = R_1/(R_1 + R_2)$. Следовательно,

$$e^2(f) = h_1^2 e_1^2(f) + h_2^2 e_2^2(f)$$

Пример 4 Закон Ома для шумов. Пусть по резистору R протекает шумовой ток $i(t)$. Для напряжения на резисторе имеем

$$e(t) = Ri(t)$$

Коэффициент R - это и есть коэффициент передачи в (33), просто он размерный. Поэтому

$$e^2(f) = R^2 i^2(f).$$

Спектральная плотность напряжения на резисторе есть произведение спектральной плотности тока на квадрат сопротивления.

Упражнение 42 По аналогии вывести шумовые законы Ома для LC компонентов, показав что

$$\begin{aligned} e^2(f) &= (1/2\pi f C)^2 i^2(f) && \text{для емкости } C, \\ e^2(f) &= (2\pi f L)^2 i^2(f) && \text{для индуктивности } L. \end{aligned}$$

4.0.8 Шумы в электронных схемах

Физические источники шума.

Естественными источниками шума в электронных схемах являются

- Джонсоновские тепловые шумы, связанные с диссинацией энергии. Двухсторонняя спектральная плотность белого шума на выводах резистора R составляет

$$e^2(f) = \mathcal{N}_0 = 4kTR \left([V]^2/[F] \right),$$

где $k = 1.38 \cdot 10^{-23}$ - постоянная Больцмана, T - температура в градусах Кельвина.

- Дробовые шумы, связанные с прохождением потока носителей через потенциальный барьер. Двухсторонняя спектральная плотность дробового шумового тока составляет

$$i^2(f) = \mathcal{N}_0 = 2qI \left([I]^2/[F] \right),$$

где $q = 1.6 \cdot 10^{-19}$ - заряд электрона, I - постоянная составляющая тока через барьер.

- Фликкер шумы не вполне понятной природы со спектральной плотностью

$$e^2(f) = \frac{N}{f},$$

где N - константа размерности мощности. Для фликкер шумов характерно свойство равной распределенности мощности по октавам (декадам) - мощность шума в полосе от $(F, 2F)$ такая же, как и полосе от $(2F, 4F)$.

Пример 5 При комнатной температуре среднеквадратичное значение шума резистора с сопротивлением $R(\Omega)$ в полосе $F(\text{Герц})$ составляет

$$\sigma \simeq 1.27 \cdot 10^{-4} \sqrt{RF} \text{ мкв}$$

Пример 6 Средний квадрат напряжения теплового шума резистора в двухсторонней полосе F составляет $e^2 = 2kTRF$. Этому отвечает мощность $P = e^2/R = 2kTF$ или энергия $E = 2kTFT$ за время t . Пространство сигналов полосы F и длительности t имеет вещественную размерность $N = 2Ft$. Приходим с следующей оценкой энергии теплового шума на вещественное измерение $E/N = kT$ (kT - это энергия на осциллятор при термодинамическом равновесии).

Пример 7 Ток дробового шума, проходя по резистору R (Ом), создает на нем шумовое напряжение со спектральной плотностью $e^2(f) = 2qIR^2$. При токе в I (мА) в полосе F (Герц) это дает среднеквадратичное значение напряжения

$$\sigma \simeq 1.8 \cdot 10^{-5} R\sqrt{IF} \text{ мкв.}$$

Упражнение 43 Показать, что спектральная плотность шума Джонсона, порождаемого двумя последовательно соединенными резисторами R_1 , R_2 , равна $4kT(R_1 + R_2)$. То же для параллельного соединения - $4kT(\frac{R_1R_2}{R_1+R_2})$.

Упражнение 44 Рассмотреть резистор нагрузки R_c в коллекторной цепи усилителя на биполярном транзисторе с начальным током коллектора I_c . Сравнить спектральную плотность шума Джонсона и спектральную плотность напряжения, создаваемого протеканием по резистору нагрузки дробового коллекторного тока. Показать, что эти плотности одинаковы, если $I_cR_c = kT/q \simeq 50$ мВ. При больших токах превалирует дробовой шум.

Упражнение 45 Показать, что среднеквадратичное шумовое напряжение на резисторе R , шунтированном емкостью C составляет kT/C .

Упражнение 46 Найти среднеквадратичное шумовое напряжение на параллельном колебательном RLC -контуре.

Шумы усилителей.

Рассмотрим усилитель с коэффициентом усиления K , ко входу которого подключен источник сигнала e_s со внутренним сопротивлением R , рис. 13

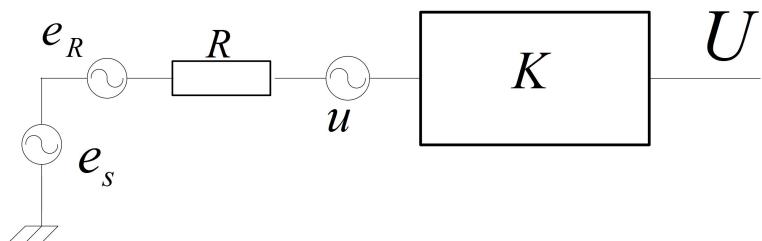


Рис. 13:

Для начала исключим все, что мешает. Положим $e_s = 0$ и охладим резистор R до абсолютного нуля. Теперь на входе у нас чистый нуль и на выходе усилителя присутствуют только его собственные шумы. Пусть $U^2(f)$ - наблюдаемая в этих условиях спектральная плотность выходного шума.

Удобно приводить шум усилителя к его входу. Начинают считать, что выходной шум возник не сам по себе, а является результатом усиления шума некоего виртуального

источника $u(t)$, добавленного на вход усилителя последовательно с R . Спектральная плотность виртуального источника меньше плотности шума на выходе в квадрат коэффициента усиления раз - $u^2(f) = U^2(f)/K^2$.

Теперь станем считать, что виртуальный источник - это как бы шум Джонсона, созданный сопротивлением источника.

$$u^2(f) = 4kT_n(f)R.$$

В итоге придем к понятию шумовой температуры усилителя

$$T_n(f) = \frac{u^2(f)}{4kR} = \frac{U^2(f)}{4kRK^2}.$$

Шумовая температура дает разумный способ описания шумовых свойств усилителя применительно к данному сопротивлению источника сигнала. Вообще говоря, она зависит от частоты, поскольку шум на выходе усилителя вовсе не обязан быть белым.

Теперь вынем резистор R из жидкого гелия. Он быстро нагреется до комнатной температуры T и создаст дополнительный Джонсоновский шум $e_R(t)$ со спектральной плотностью $e_R^2(t) = 4kTR$. Нет никаких оснований считать, что этот шум как-то коррелирован с шумом виртуального источника. Поэтому полная плотность шума на выходе усилителя составит

$$4kTR + u^2(f) = 4k(T + T_n)R.$$

Коэффициент шума усилителя вводят как отношение в децибелах суммарной плотности к той, которая имела бы место в случае не шумящего усилителя.

$$K_n(\text{дБ}) = 10 \lg \frac{4kTR + u^2(f)}{4kTR} = 10 \lg \left(1 + \frac{T_n(f)}{T} \right)$$

Это второй общепринятый способ описания шумовых свойств усилителя. Важно понимать, что коэффициент шума определен применительно к данному сопротивлению источника, а его значение зависит от частоты. Для ориентировки, если шумовая температура равна комнатной, то значение коэффициента шума составляет 3 дБ. Это плохой усилитель.

Наконец, включим источник сигнала. В результате полный сигнал на выходе усилителя станет суммой трех компонент

$$e_s(t) + e_R(t) + u(t),$$

из которых только третья связана с шумовыми свойствами самого усилителя.

На этом этапе у нас возникает проблема. Заключается она в том, что в данной сумме последние два слагаемые - это шумы, а первое слагаемое - нет. Единственно корректный выход из положения - начать считать сигнал $e_s(t)$ шумом.

Это не так уж глупо. Известно, например, что использование вероятностных моделей информационных потоков делает линейно модулированные сигналы случайными процессами, для которых определено понятие спектра мощности. Даже обычный синусоидальный сигнал всегда можно рассматривать как шум, если принимать во внимание наличие в нем фазовых шумов, вызывающих флюктуации частоты. Рассматриваемый как шум, сигнал обладает спектральной плотностью $e_s^2(t)$. Для синусоидального сигнала эта плотность сосредоточена в узкой полосе вокруг основной частоты. Для линейно модулированного сигнала она занимает некоторую полосу частот. Если считать, что шум сигнала некоррелирован с двумя другими шумами, то для спектральной плотности шума на выходе усилителя найдем

$$e_s^2(f) + e_R^2(f) + u^2(f)$$

Отношение сигнал шум.

Теперь нам предстоит ввести понятие отношения сигнал/шум (SNR - Signal to Noise Ratio). Определять его как отношение спектральных плотностей сигнала и шума не вполне разумно, поскольку такое определение учитывало бы конкретную форму кривой спектральной плотности сигнала, о которой мало что известно. Разумный инженерный подход состоит в том, чтобы выбрать некоторую полосу частот F , в которой заведомо лежит почти вся энергия сигнала, и перейти от спектральных плотностей к мощностям в этой полосе

$$\int_F e_s^2(f)df + \int_F e_R^2(f)df + \int_F u^2(f)df$$

Здесь все интегралы берутся по полосе частот сигнала. Первый дает полную мощность сигнала σ_s^2 . Другие два - оценки мощностей шумов в полосе F . Если считать шумы белыми, то для мощности смеси сигнала с шумами найдем

$$\sigma_s^2 + e_R^2(f)F + u^2(f)F$$

Отношение сигнал/шум введем теперь как отношение полной мощности сигнала к мощности шума в полосе F .

Замечательно, что так введенное SNR оказывается не зависящим от конкретной формы спектральной плотности сигнала. При измерении SNR в качестве тестового сигнала можно использовать как узкополосный синусоидальный сигнал, так и приближенный к реальному сигнал с полосой F . В определении SNR участвует только полная мощность сигнала σ_s^2 и его полоса F .

Упражнение 47 Предложить схему эксперимента по измерению отношения сигнал/шум. (Сигнал выключить можно, шум - нет).

Представляют интерес два отношения сигнал шум:

- SNR на выходе источника сигнала

$$SNR_s = 10 \lg \frac{\sigma_s^2}{e_R^2(f)F} = 10 \lg \frac{\sigma_s^2}{4kTRF}.$$

- SNR на эквивалентном входе шумящего усилителя сигнала

$$SNR_u = 10 \lg \frac{\sigma_s^2}{e_R^2(f)F + u^2(f)F} = 10 \lg \frac{\sigma_s^2}{4kTRF + u^2(f)F}.$$

Таким же оно окажется и на его выходе, поскольку сигнал и шумы усиливаются в одинаковое число K раз.

Имеем,

$$SNR_s = 10 \lg \frac{\sigma_s^2}{e_R^2(f)F} = 10 \lg \frac{\sigma_s^2}{e_R^2(f)F + u^2(f)F} \frac{e_R^2(f)F + u^2(f)F}{e_R^2(f)F} = SNR_u + K_n.$$

Это раскрывает смысл введения коэффициента шума. Оказывается, что отношение сигнал/шум на выходе шумящего усилителя хуже отношения сигнал/шум источника сигнала ровно на коэффициента шума.

$$SNR_u = SNR_s - K_n.$$

Отношение сигнал/шум для согласованного фильтра.

Для систем с согласованной фильтрацией отношение сигнал/шум вводится несколько иначе. В расчет берется отношение сигнала и шума не в полосе частот, а на выходе конкретного согласованного фильтра в фиксированный момент времени - момент принятия решения.

Формально, согласованный фильтр - это коррелятор, вычисляющий скалярное произведение - интеграл от произведения принятой реализации на местную (опорную) копию принимаемого сигнала. Между операциями вычисления корреляции и согласованной фильтрацией имеется тонкое различие. С точки зрения чистой алгебры, в обоих случаях выполняются одинаковые операции. Различие в размерностях. Коррелятор - это квадратичное устройство. Его отклик, вычисляемый как интеграл от произведения сигналов, имеет размерность энергии. Напротив, фильтр - устройство линейное. Его отклик имеет размерность уровня сигнала.

Разберемся с калибровками на примере построения согласованного фильтра для импульса $p(t)$ некоторой системы линейной модуляции. Наша цель состоит в том, чтобы построить фильтр, импульсная реакция которого пропорциональна $p^*(-t)$. Все очевидно, но налицо проблема с размерностями. Размерность сигнала - это уровень (вольты). Размерность импульсной реакции - обратное время.

Проблема с размерностью уровня сигнала решается просто. Введем амплитуду импульса c и будем считать что в вольтах измеряется именно она, а сам импульс $p(t)$ - безразмерен. Придем к представлению сигнала в виде $cp(t)$. Для энергии этого сигнала найдем

$$E_s = \langle cp, cp \rangle = |c|^2 \langle p, p \rangle = |c|^2 \int p(t)p^*(t) dt = |c|^2 T.$$

Появившуюся нормировочную константу T можно рассматривать как меру эффективной длительности импульса. Тогда квадрат амплитуды $|c|^2 = E_s/T$ приобретает смысл средней мощности сигнала.

Импульсную реакцию согласованного фильтра выберем теперь в виде $h(t) = p^*(-t)/T$. Все требования к размерности импульсной реакции соблюdenы. Фильтрация сигнала $cp(t)$ фильтром $h(t)$ в момент принятия решения $\tau = 0$ дает отсчет уровня сигнала.

$$(h * cp) = c(h * p) = \frac{c}{T} (p(-t) * p(t)) = \frac{c}{T} \langle p, p \rangle = \frac{c}{T} T = c$$

Комплексный коэффициент передачи $H(f)$ фильтра $h(t)$, как и должно быть, безразмерен. Имеем

$$\langle H(f), H(f) \rangle = \int |H(f)|^2 df = \int |h(t)|^2 dt = \langle h, h \rangle = \frac{1}{T^2} \langle p, p \rangle = \frac{1}{T}$$

Пусть сигнал $cp(t)$ принимается на фоне аддитивного шума со спектральной плотностью $n^2(f)$. На выходе фильтра шум приобретет спектральную плотность $|H(f)|^2 n^2(f)$.

Отношение сигнал/шум (SNR) введем теперь как отношение мощности сигнала к дисперсии шума в децибелах,

$$SNR = 10 \lg \frac{|c|^2}{\int |H(f)|^2 n^2(f)}.$$

Это отношение аналогично введенному ранее. Единственное отличие состоит том, что теперь мощность шума берется не в абстрактной полосе сигнала F , а на выходе вполне конкретного фильтра с комплексным коэффициентом передачи $H(f)$.

Сейчас мы имеем классическую калибровку плоскости комплексных огибающих. Размерностью сигнала является уровень (вольты), дисперсия шума имеет размерность мощности. Слабым местом в этой калибровке оказывается то, что понятие мощности финитного во времени сигнала введено несколько искусственно, через понятие эффективной длительности T .

Умножим числитель и знаменатель SNR на T .

$$SNR = 10 \lg \frac{|c|^2 T}{\int T |H(f)|^2 n^2(f)} = 10 \lg \frac{E_s}{\int T |H(f)|^2 n^2(f)} \quad (36)$$

В числителе имеем теперь энергию сигнала - $E_s = |c|^2 T$. Это вполне корректно определенная величина. Более того, она определена именно как атрибут сигнала, безотносительно к каким бы то ни было фильтрам. Однако проблемный нормировочный множитель T переместился в знаменатель.

Хорошей идеей оказывается нормирование согласованного фильтра на единичную шумовую полосу. Введем нормированный фильтр

$$\widetilde{H}(f) = \sqrt{T} H(f),$$

у которого

$$\int |\widetilde{H}(f)|^2 df = \int |\tilde{h}(t)|^2 dt = 1. \quad (37)$$

Тогда для SNR получим

$$SNR = 10 \lg \frac{E_s}{\int |\widetilde{H}(f)|^2 n^2(f)}. \quad (38)$$

В этой нормировке размерностью плоскости комплексных огибающих оказывается корень квадратный из энергии. Квадрат сигнала приобретает размерность энергии. Интеграл по частоте от $|H(f)|^2$ безразмерен. Поэтому интеграл от спектральной плотности $e^2(f)$ (дисперсия шума) оказывается той же размерности энергии, что и спектральная плотность. Все это несколько непривычно, поскольку обычно спектральная плотность - это энергия, а дисперсия шума - мощность.

Пусть, наконец, спектральная плотность шума равномерна в полосе фильтра $H(f)$ и составляет $n^2(f_0) = N_0$. Тогда отношение сигнал/шум оказывается не зависящим от от свойств фильтра во всем, кроме нормировки. В результате приходим к следующей трактовке отношения сигнала/шума

Отношение сигнал/шум на выходе согласованного фильтра есть отношение энергии сигнала к спектральной плотности шума.

$$SNR = 10 \lg \frac{E_s}{N_0} \quad (39)$$

И численно и по существу это отношение сигнал/шум тождественно классическому. Так что шумы в усилительных устройствах влияют на его ровно так, как это было описано выше.

Применение согласованной фильтрации часто обосновывают ссылкой на следующее свойство оптимальности.

Теорема 4.7 *Согласованный фильтр оптимален в классе линейных фильтров по критерию максимума отношения сигнал/шум на выходе.*

Доказательство. Будем вести обработку сигнала линейным фильтром с импульсной реакцией $h(t)$. Без ограничения общности можем считать фильтр нормированным на единицу

$$\langle h(t), h(t) \rangle = \langle H(f), H(f) \rangle = 1.$$

Для отношения сигнал/шум на выходе фильтра найдем

$$SNR = \frac{c \langle p(t), h^*(-t) \rangle}{N_0}$$

Нормировка фильтра обеспечила независимость знаменателя от его выбора. Для скалярного произведения в числителе имеет место неравенство Коши-Шварца

$$\langle p(t), h^*(-t) \rangle \leq \|p(t)\| \|h^*(-t)\|,$$

которое достигается (выполняется со знаком равенства), если и только если $h^*(-t) = \alpha p(t)$. Это означает, что отношение сигнал/шум максимизирует именно согласованный фильтр ∇ .

Применительно к согласованной фильтрации калибровка размерности сигналов на корень из энергии наиболее популярна, но не единственна возможна.

Замечание 4 От проблем, связанных с преобразованием коррелятора в согласованный фильтр, можно уйти совсем. Для этого достаточно принять, что сигналом является отклик коррелятора как такового. Тогда сигнал приобретет размерность энергии $E_s = \int p^2(t)dt$. В единицах же энергии будет измеряться и уровень шума. А вот дисперсия шума

$$\sigma^2 = \int e^2(f)|H(f)|^2df = N_0 \int |H(f)|^2df = N_0 E_s$$

окажется имеющей «дикую» размерность квадрата энергии. И это нормальная точка зрения. Отношение квадрата уровня сигнала к дисперсии шума $E_s^2/N_0 E_s = E_s/N_0$ останется все тем же отношением сигнал/шум.

Стоит сказать пару слов о калибровках в теории сигналов. За размерностями здесь бывает весьма сложно уследить. Проблемой становятся размерные нормирования. К примеру, по ходу дела говорят: "далее без ограничения общности будем считать, что $\langle p, p \rangle = 1$ ". Так появляется единица, имеющая размерность. Если забыть о размерности и перестать писать эту единицу в формулах, что всегда и происходит, то анализ размерностей получающихся выражений будет вызывать сильное изумление.

Существует две крайности. Математики предпочитают думать о размерностях в последнюю очередь. Их цель - подметить алгебраические закономерности, ведущие к тем или иным полезным приемам обработки сигналов. Учет размерностей здесь только вредит. Наоборот, в прикладных областях, в особенности связанных энергетикой радиотрактов, анализ размерностей приобретает принципиальное значение. Разумный подход состоит в том, чтобы не сосредоточиваться на размерностях при промежуточных преобразованиях, откладывая их анализ до получения значимых результатов.

При анализе влияния шумов в системах обработки сигналов любой физической природы обязательно встает вопрос о привязке используемых величин к термодинамике. В теории электрических сигналов эту привязку как раз и обеспечивает понятие отношения сигнал/шум (39), связывая энергию сигналов со спектральной плотностью физического теплового шума.

4.0.9 Шумы в Р и В областях

Результатом прохождения реального шума $n_p(t)$ через демодулятор Гильберта становится эквивалентный комплексный шум в В-области.

$$n(t) = \frac{1}{\sqrt{2}}(n_p(t) + i\tilde{n}_p(t))e^{-f_0t},$$

где $\tilde{n}_p(t)$ - результат прохождения шума $n_p(t)$ через линейный фильтр Гильберта с импульсной реакцией $\frac{1}{\pi t}$ и комплексным коэффициентом передачи $\frac{\text{sign}(f)}{i}$.

Мост между шумами в Р и В областях наводит результат теоремы о корреляции, устанавливающий связь между корреляциями сигналов и их комплексных огибающих

$$\langle x, y(t) \rangle = (\langle x_p, y_p(t) \rangle + i \widetilde{\langle x_p, y_p(t) \rangle}) e^{-f_0 t} \quad (40)$$

На случай шумов это тождество переносится по стандартной схеме эргодического перехода. В результате получается аналог (40) для статистических корреляций

$$\langle n|n(t) \rangle = (\langle n_p|n_p(t) \rangle + i \widetilde{\langle n_p|n_p(t) \rangle}) e^{-f_0 t}.$$

Этим устанавливается связь между корреляционной функцией физического шума $R_p(t) = \langle n_p|n_p(t) \rangle$, ее Гильбертовым преобразованием $\tilde{R}_p(t) = \widetilde{\langle n_p|n_p(t) \rangle}$ и корреляционной функцией $R(t) = \langle n|n(t) \rangle$ комплексной огибающей.

$$R(t) = (R_p(t) + i \tilde{R}_p(t)) e^{-f_0 t}, \quad (41)$$

где

$$\tilde{R}_p(t) = \frac{1}{\pi} \int \frac{R_p(u)}{u-t} du.$$

Это же соотношение в частотной области, выраженное через спектральные плотности \mathcal{N} , принимает вид

$$\mathcal{N}(f) = \mathcal{N}_p(f + f_0) + i \tilde{\mathcal{N}}_p(f + f_0), \quad (42)$$

где

$$\tilde{\mathcal{N}}_p = \mathcal{N} \frac{\text{sign}(f)}{i}.$$

Вот сводка сведений о довольно многочисленных симметриях.

- Корреляционная функция $R_p(t)$ вещественного шума вещественна и симметрична.
- Поэтому спектральная плотность $\mathcal{N}_p(f) \leftrightarrow R_p(t)$ также симметрична и вещественна.
- Функция $R(t)$ комплексна, но сопряженно симметрична - $R(-t) = R^*(t)$.
- Поэтому ее Фурье преобразование $\mathcal{N}(f) \leftrightarrow R(t)$ вещественно.
- Если корреляционную функцию представить суммой вещественной и мнимой частей $R(t) = R_i(t) + iR_q(t)$, а ее спектральную плотность - суммой симметричной и антисимметричной компонент

$$\mathcal{N} = \mathcal{N}_s + \mathcal{N}_a = \frac{\mathcal{N}(f) + \mathcal{N}(-f)}{2} + \frac{\mathcal{N}(f) - \mathcal{N}(-f)}{2},$$

то окажется, что

$$R_i(f) \leftrightarrow \mathcal{N}_s(f), \quad R_q(f) \leftrightarrow \mathcal{N}_a(f)$$

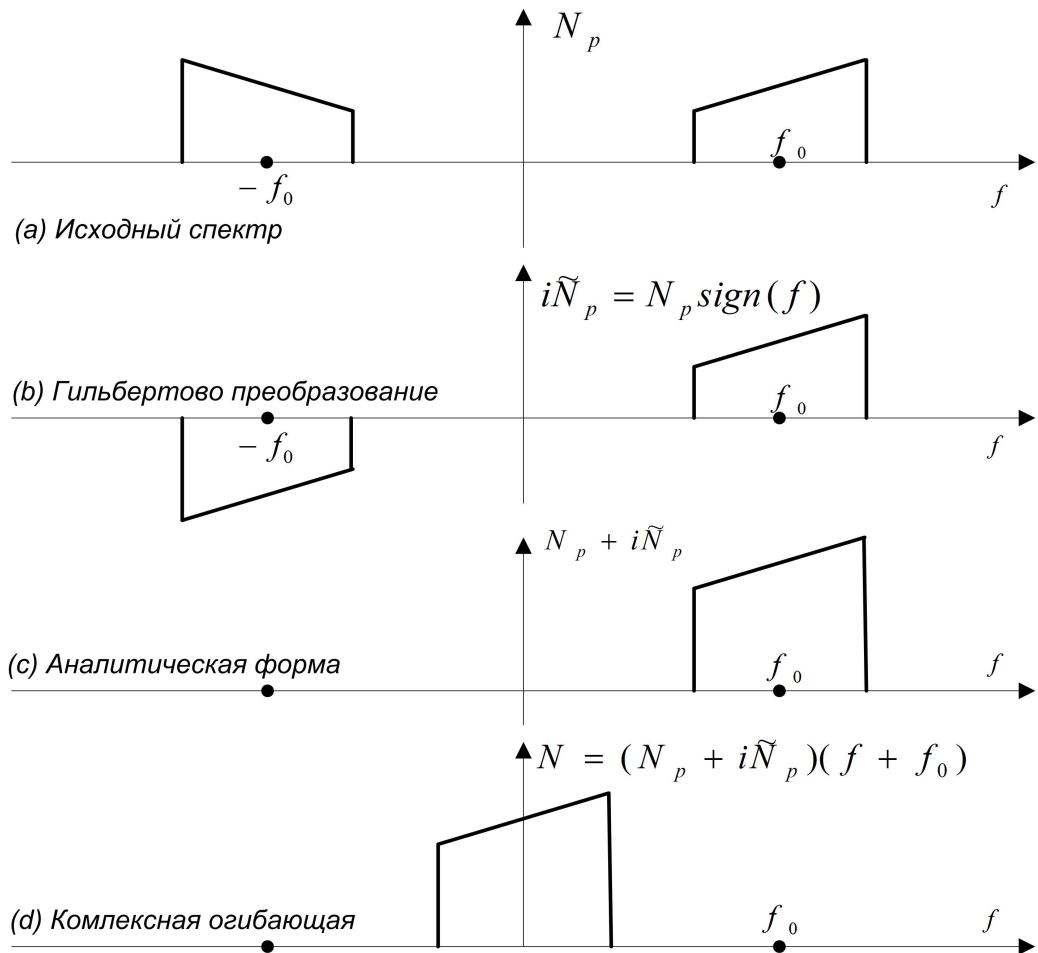


Рис. 14: Формирование комплексной огибающей шума

- И, следовательно,

$$R_i(-t) = R_i(t), \quad R_q(-t) = -R_q(t). \quad \text{В частности, } R_q(0) = 0$$

Теорема 4.8 Комплексная огибающая любого вещественного шума нормальна.

Доказательство. Для нормальности шума

$$n(t) = n_i(t) + n_q(t)$$

требуется, чтобы

$$\langle n_i | n_i \rangle (t) = \langle n_q | n_q \rangle (t), \quad \langle n_i | n_q \rangle (t) = -\langle n_q | n_i \rangle (t).$$

Но это в точности последнее из перечисленных свойств симметрии ∇ .

Замечание 5 Нормальность комплексного шума означает инвариантность его вторых моментов, а в Гауссовом случае и распределения, относительно поворотов плоскости комплексных огибающих - $n(t) \rightarrow n(t)e^\theta$.

Процедуру (42) построения спектральной плотности комплексной огибающей вещественного шума иллюстрирует рис. 14. Рис. (а) показывает спектральную плотность вещественного шума $N_p(f)$, сосредоточенную в окрестности частоты f_0 . Как и должно быть, эта плотность показана вещественной и симметричной. На первом этапе, рис. (б), строится спектральная плотность Гильбертова преобразования корреляционной функции шума. В частотной области это эквивалентно умножению на $\frac{\text{sign} f}{i}$. В результате получается мнимая антисимметричная плотность $\tilde{N}_p(f)$. Рисунок представляет график $i\tilde{N}_p(f)$, который вещественен и антисимметричен. Результатом сложения $N_p(f) + i\tilde{N}_p(f)$ становится уничтожение компонент спектра на отрицательных частотах и их удвоение - на положительных, рис. (с). Это спектральная плотность аналитического шума. Последнее - это перенос спектра влево умножением на $e^{-f_0 t}$ во временной области.

Результатом оказывается спектральная плотность комплексной огибающей шума. Она сосредоточена в окрестности нулевой частоты, вещественна, но, вообще говоря, не симметрична. Ее разложению на симметричную и антисимметричную части отвечает разложение корреляционной функции на вещественную и мнимую компоненты. В частном случае, когда спектральная плотность симметрична, корреляционная функция шума окажется вещественной, а вещественная и мнимая компоненты шума - некоррелированными.

Теорема 4.9 (О белом шуме) *Комплексная огибающая вещественного белого шума со спектральной плотностью $N_0/2$ является комплексным белым шумом со спектральной плотностью N_0 .*

$$\begin{aligned} \text{В } P\text{-области: } \quad & \frac{N_0}{2} \leftrightarrow \frac{N_0}{2} \delta(t), \\ \text{В } B\text{-области: } \quad & N_0 \leftrightarrow N_0 \delta(t). \end{aligned}$$

Доказательство. На рис. 14(а) заменить кривую спектральной плотности на равномерную - $N_0/2$ и повторить все этапы преобразований ∇ .

Если к вещественному белому шуму добавить эпитет «гауссовский», то эта характеристика перейдет и на комплексную огибающую.

Теорема 4.10 *Любой шум, который может быть получен в результате фильтрации белого шума, имеет нормальную комплексную огибающую.*

Доказательство. Комплексный белый шум нормален. Но линейная обработка нормального случного процесса всегда дает нормальный процесс. Поэтому все результаты фильтрации белого шума в комплексной области нормальны.

Каждый вещественный фильтр имеет комплексный эквивалент. Поэтому комплексные огибающие всех результатов фильтрации белого шума в Р-области также нормальны ∇ .

Проблемы наивного подхода к квадратурным преобразованиям шума

Переход от реальных сигналов к комплексным огибающим через гильбертово преобразование дает закон (41) преобразования корреляционных функций шумов между Р и В областями. Это позволяет обойти не заметив некоторые подводные камни, с которыми приходится сталкиваться при наивных попытках квадратурного преобразования шумов. Рассмотрим вопрос о поведении шумов при их умножении на гармонические функции более тщательно.

Пусть $n(t)$ -шум с корреляционной функцией $n^2(\tau) = \langle n, n(\tau) \rangle$ и спектральной плотностью $n^2(f)$.

Умножение шума на комплексную гармоническую функцию не приводит к трудностям. Легко проверяется, что

$$\langle n(t)e^{f_0 t}, n(t - \tau)e^{f_0(t-\tau)} \rangle = e^{f_0\tau} \langle n, n \rangle = e^{f_0\tau} n^2(\tau).$$

Это означает, что умножение шума на комплексную экспоненту ведет к умножению на ту же экспоненту корреляционной функции, что эквивалентно сдвигу спектральной плотности $e^2(f)$ на f_0 .

Трудности возникают при умножении шума на вещественную гармоническую функцию. Умноженный на косинус шум представляется в виде суммы двух компонент

$$2n(t) \cos f_0 t = n(t)e^{+f_0 t} + n(t)e^{-f_0 t} = n^+(t) + n^-(t).$$

Его корреляционная функция оказывается суммой четырех корреляций - двух авто $\langle n^+, n^+ \rangle$, $\langle n^-, n^- \rangle$, и двух взаимных - $\langle n^+, n^- \rangle$, $\langle n^-, n^+ \rangle$.

С автокорреляциями все просто. Они выражаются через умноженные на комплексные экспоненты корреляционные функции $\langle n, n \rangle$ и дают компоненты спектральной плотности, сдвинутые влево и вправо на f_0 . Если допустить некоррелированность компонент - $\langle n^+, n^- \rangle = \langle n^-, n^+ \rangle = 0$, то для спектральной плотности $N(f)$ умноженного на косинус шума получается ожидаемый результат

$$N(f) = \frac{1}{4}(n^2(f - f_0) + n^2(f + f_0)).$$

Если $n(t)$ - белый шум с равномерной спектральной плотностью $\mathcal{N}_0/2$, то

$$N(f) = \frac{1}{2} \mathcal{N}_0/2,$$

где $\mathcal{N}_0/2$ - половина спектральной плотность эквивалентного комплексного шума, а нормировочный множитель $1/2$ можно трактовать как эффективный коэффициент передачи умножителя на косинус - среднее значение $\cos^2 f_0 t$.

Сложность состоит в обосновании обращения в нуль перекрестных корреляций. Вот развернутая форма одной из них:

$$\langle n^+, n^- \rangle = e^{-f_0\tau} \int n(t)n^*(t - \tau)e^{2f_0t} dt.$$

Требуется обосновать обращение в нуль интеграла.

Выхода из затруднительной ситуации находится на пути обобщения. Введем в рассмотрение начальную фазу несущей - $\cos(f_0 t + \theta)$. До настоящего момента она всегда выбиралась нулевой. Легко проверить, что обе авто корреляции от θ не зависят. В перекрестную же корреляцию начальная фаза несущей войдет явно.

$$\langle n^+, n^- \rangle = e^{2\pi i \theta} e^{-f_0\tau} \int n(t)n^*(t - \tau)e^{2f_0t} dt. \quad (43)$$

Усреднение по начальной фазе θ и дает желаемое обращение перекрестных корреляций в нуль.

Это техническая сторона дела. Суть же состоит в следующем. Умноженный на косинус шум нельзя считать стационарным. Плотность $p(u(t))$ распределения вероятностей его значения $u(t) = n(t)\cos f_0 t$ в момент t , разумеется, зависит от значения косинуса в этот момент. Можно констатировать только выполнение условия периодической стационарности - $p(u(t + mT)) = p(u(t))$, $T f_0 = 1$.

Подобные шумы называют циклически стационарными. К циклически стационарным шумам приводят, например, все вероятностные модели модулированных сигналов на базе случайных потоков символов.

Для циклически стационарных шумов корреляционную теорию удается «спасти» введением искусственной рандомизации начала отсчета времени. Вводится случайный равномерно распределенный на периоде начальный момент τ и по нему производится дополнительное усреднение функции корреляции. В результате получается не зависящая от τ стационарная корреляционная функция, через которую и вводится понятие спектральной плотности.

В примере с шумом, умноженным на косинус, можно положить, что шум имеет ненулевое начало отсчета времени $\tau = 2\pi\theta$ и усреднение в (43) ведется именно по τ , а фаза несущей - нулевая. Тогда оказывается, что построенная усреднением по времени корреляционная функция шума зависит от выбора начала отсчета времени для шума. Усреднение по равномерно распределенному началу отсчета делает шум стационарным и, заодно, исключает зависимость корреляционной функции от начала отсчета времени.

Важно понять, что случайная фаза θ в (43) не является случайным процессом. То есть, она не флюктуирует «все время». Это случайная величина, значение которой раз и навсегда была выбрано когда-то в начало отсчета времен, после чего оно не изменяется.

Упражнение 48 По аналогии найти спектральную плотность шума на выходе умножителя на синус.

Проанализировать умножение шума на $\cos f_0 t + i \sin f_0 t$. Убедиться в том, что в этом случае перекрестные корреляции благополучно выбывают из игры вместе со всеми связанными с ними проблемами.

Узкополосный вещественный шум

Для описания шумов на выходе узкополосных фильтров широко используется модель узкополосного шума. Проще всего к ней прийти преобразованием низкочастотного комплексного шума в Р-область.

Рассмотрим комплексный шум $n(t) = n_i(t) + i n_q(t)$. Будем считать, что его спектральная плотность сосредоточена в относительно узкой полосе частот в окрестности нуля.

Преобразование этого шума в Р-область на частоту несущей f_0 дает

$$n_p(t) = \sqrt{2} \{ n_i(t) \cos f_0 t - n_q(t) \sin f_0 t \}. \quad (44)$$

Это и есть узкополосный шум. Он содержит две случайные квадратурные компоненты n_i, n_q , статистические свойства которых полностью характеризуют свойства шума.

Функция авто корреляции узкополосного шума определяется из (41) через вещественную и мнимую компоненты корреляционную функцию комплексной огибающей $R(t)$

$$R_p(t) = \operatorname{Re}\{R(t)e^{f_0 t}\} = R_i(t) \cos f_0 t - R_q(t) \sin f_0 t \quad (45)$$

и оказывается выраженной через корреляции квадратурных составляющих

$$R_i(t) = 2 \langle n_i, n_i \rangle (t) = 2 \langle n_q, n_q \rangle (t), \quad R_q(t) = 2 \langle n_q, n_i \rangle (t) = -2 \langle n_i, n_q \rangle (t). \quad (46)$$

Упражнение 49 Получить выражение для корреляционной функции (45) непосредственно из формы (44) с использованием свойств (46).

При традиционном построении модели узкополосного шума свойства симметрии корреляций (46) просто постулируются, а корреляционная функция вычисляется так, как это предложено сделать в упражнении. Если не знать, откуда растут уши, то оправдать введение симметрий бывает довольно сложно. Привлекаются соображения вплоть до равно распределенности шумовой энергии по степеням свободы. В какой то степени это так и есть. Но фактически эти симметрии просто выражают требование нормальности комплексного шума. Их невыполнение привело бы к модели узкополосного шума с комплексной огибающей, не инвариантной относительно вращений комплексной плоскости. Такой шум, в частности, не мог бы быть получен никакой фильтрацией белого шума.

Упражнение 50 Проверить выполнение свойств узкополосного шума (45). (Все это прямые следствия свойства нормальности комплексного шума).

- Квадратурные компоненты имеют одинаковые дисперсии

$$\sigma^2 = \langle n_i, n_i \rangle (t) = \langle n_q, n_q \rangle (t).$$

- Квадратурные компоненты не коррелированы при нулевом временном сдвиге -

$$\langle n_i, n_i \rangle = R_q(0) = 0.$$

- Дисперсия узкополосного шума равна сумме дисперсий квадратурных составляющих и составляет $2\sigma^2$.
- Если шум - Гауссовский, то квадратурные компоненты независимы и каждая из них распределена по Гауссовскому закону с дисперсией σ^2 .
- Корреляционные функции

$$\langle n_i, n_i \rangle (t) = \langle n_q, n_q \rangle (t) = R_i(t)/2$$

симметричны.

- Перекрестные корреляционные функции

$$\langle n_q, n_i \rangle (t) = -\langle n_i, n_q \rangle (t) = R_q(t)/2$$

антисимметричны.

- Если спектральная плотность шума симметрична относительно f_0 , то перекрестные корреляционные функции - нулевые. При этом квадратурные шумы не коррелированы при любых сдвигах.

5 Оптимальная демодуляция

5.1 Статистическое различение гипотез

Рассмотрим схему линейной модуляции с нормированным базовым импульсом $p(t)$, $\langle p, p \rangle = 1$ и созвездием $\mathbb{C} = \{c_j, j = 1 \dots M\}$, рис.15(а). Переданному сигналу $x(t) = c_j p(t)$ отвечает принятая реализация $y(t) = c_j p(t) + n(t)$ с аддитивным шумом. Результатом ее согласованной фильтрации оказывается комплексная случайная величина

$$y = c_j \langle p, p \rangle + \langle n, p \rangle = c_j + \langle n, p \rangle.$$

Распределение вероятностей шума задает условную плотность вероятностей $p(y/c_j)$ отклика y при данном переданном символе. Зададим также некоторое априорное распределение вероятностей передаваемых символов созвездия

$$\pi(c) = \{\pi(c_1), \dots, \pi(c_M)\}, \quad \sum_j \pi(c_j) = 1.$$

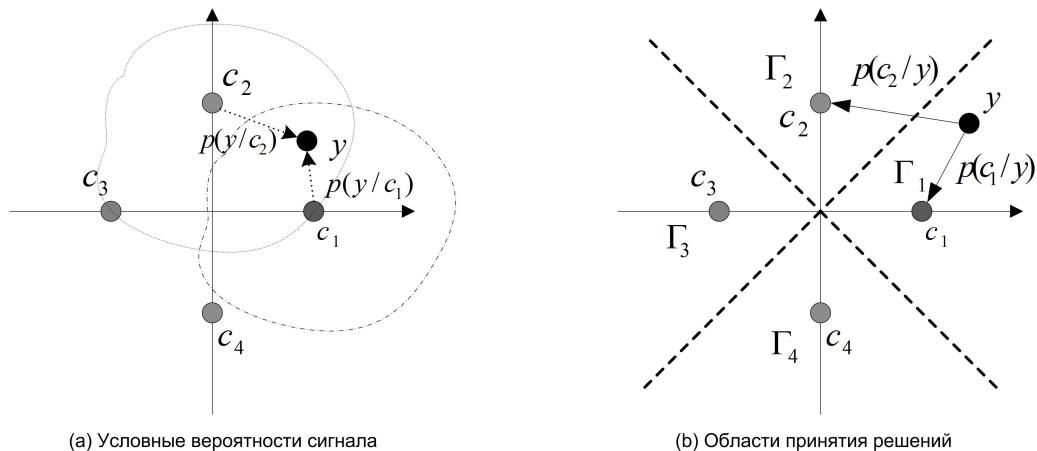


Рис. 15: К задаче различения гипотез

Перед нами классическая статистическая задача различения гипотез.

Имеем M гипотез $\{c_j, j = 1 \dots M\}$. Известно априорное распределение их вероятностей $\pi(c_j)$ и набор условных распределений $p(y/c_j)$, $j = 1 \dots M$. Требуется по наблюдению y принять решение в пользу одной из гипотез так, чтобы вероятность ошибочного решения оказалась минимальной.

Исходные данные позволяют построить совместное распределение $p(c, y) = \pi(c)p(y/c)$ и апостериорное распределение гипотез при заданном наблюдении

$$p(c/y) = \frac{p(c, y)}{p(y)} = \frac{\pi(c)p(y/c)}{\sum_c \pi(c)p(y/c)}.$$

Предположим, что по результату наблюдения y выносится случайное решение \hat{c} о переданном символе в соответствии с некоторым распределением вероятностей $\gamma(\hat{c}/y)$. В этой модели получаем следующее совместное распределение для вероятности того, что передавался символ c , наблюдалось значение y и было принято решение \hat{c}

$$p(c, y, \hat{c}) = \pi(c)p(y/c)\gamma(\hat{c}/y) = p(y)p(c/y)\gamma(\hat{c}/y). \quad (47)$$

Усредняя по y , придем к следующей оценке средней вероятности ошибочного решения

$$\begin{aligned} P_e &= \sum_{c \neq \hat{c}} \int p(c, y, \hat{c}) dy = 1 - \sum_{c=\hat{c}} \int p(c, y, \hat{c}) dy = \\ &= 1 - \int p(y) \sum_c p(c/y) \gamma(c/y) dy \end{aligned} \quad (48)$$

Поскольку $p(y)$ неотрицательно, задача минимизации средней вероятности ошибки P_e оказывается эквивалентной максимизации по распределению γ суммы

$$\sum_c p(c/y) \gamma(c/y) \quad (49)$$

для каждого значения y . Здесь оказывается полезным следующий простой факт.

Лемма 2 (О вреде совместительства) *Пусть некая работа i приносит доход $w_i \geq 0$ при полной занятости. Требуется найти распределение*

$$\{\gamma_1, \dots, \gamma_M\}, \quad \sum_{i=1}^M \gamma_i = 1$$

своего рабочего времени между M местами работы, максимизирующее суммарный доход

$$W = \sum_{i=1}^M w_i \gamma_i$$

Оптимальное распределение имеет вид

$$\gamma_i = \begin{cases} 1 & i = \arg \max_j w_j \\ 0 & \text{иначе} \end{cases}$$

Иными словами, все рабочее время следует проводить там, где большие платят.

Упражнение 51 Предложить формальное доказательство леммы.

Согласно лемме, сумму (49) максимизирует алгоритм принятия решений γ , который по каждому результату наблюдения y с вероятностью 1 выдает решение \hat{c} , максимизирующее вес $p(c/y)$

$$\hat{c} = \arg \max_c p(c/y) \quad (50)$$

Это известный критерий **максимума апостериорной вероятности**, который гарантирует получение минимальной средней ошибки при принятии той гипотезы, которая имеет наибольшую апостериорную вероятность при данных результатах наблюдений.

Применение критерия разбивает область значений y на M подмножеств, рис.15(b)

$$\Gamma_{\hat{c}} = \{y : \arg \max_c p(c/y) = \hat{c}\}$$

Подмножество $\Gamma_{\hat{c}}$ имеет смысл области принятия решения в пользу гипотезы \hat{c} . Разбиение пространства наблюдений на области $\Gamma_{\hat{c}}$ позволяет привести выражение (48) для средней вероятности ошибки к виду

$$P_e = 1 - \sum_c \int_{\Gamma_c} p(y) p(c/y) dy = 1 - \sum_c \pi(c) \int_{\Gamma_c} p(y/c) dy.$$

Пусть

$$P_{\hat{c}}(\Gamma_c) = \int_{\Gamma_c} p(y/\hat{c}) dy$$

вероятностная мера области Γ_c , вычисленная по условной плотности вероятностей $p(y/\hat{c})$. В этих обозначениях для вероятности ошибки найдем

$$P_e = 1 - \sum_c \pi(c) P_c(\Gamma_c) = \sum_c \pi(c) P_c(\bar{\Gamma}_c) = \sum_c \pi(c) \sum_{\hat{c} \neq c} P_c(\Gamma_{\hat{c}}) \quad (51)$$

где

$$\bar{\Gamma}_c = \bigcup_{\hat{c} \neq c} \Gamma_{\hat{c}}.$$

Критерию максимума апостериорной вероятности (50) можно придать следующую форму

$$\hat{c} = \arg \max_c p(c/y) = \arg \max_c \frac{\pi(c)p(y/c)}{p(y)} = \arg \max_c \pi(c)p(y/c).$$

В отсутствии разумных предположений об априорном распределении вероятностей гипотез его принимают равномерным - $\pi(c) = 1/M$. В результате получается следующий критерий **максимального правдоподобия**

$$\hat{c} = \arg \max_c p(y/c), \quad (52)$$

согласно которому решение следует принимать в пользу той гипотезы, которая максимизирует условную вероятность наблюдения.

Подведем итоги. Выносить решение о переданном символе c по отсчету y на выходе согласованного фильтра следует по критерию максимума апостериорной вероятности (50) (*при наличии априорной информации*) или по критерию максимума правдоподобия (52) (*в ее отсутствии*).

Любой из этих критериев приводит к разбиению пространства наблюдений y на области принятия решения Γ_c . При попадании отсчета y в область Γ_c выносится решение в пользу символа c . Такое поведение гарантирует минимизацию средней вероятности ошибки.

Формула (51) предлагает два пути построения оценок средней вероятности ошибки.

- Через построение оценки средней вероятности правильного решения. Вероятность $P_c(\Gamma_c)$ правильного решения при передаче данного символа c находится интегрированием условного распределения $p(y/c)$ по области Γ_c .
- Прямыми вычислением средней вероятности ошибки. Вероятность $P_c(\bar{\Gamma}_c)$ ошибки при передаче символа c находится интегрированием условной вероятности $p(y/c)$ по дополнению $\bar{\Gamma}_c$ к множеству Γ_c .

Для не совсем тривиальных созвездий оба варианта ведут к достаточно сложным вычислениям даже в предположении о гауссовском характере условных распределений. Это придет смысл построению простых приближенных оценок.

5.2 Граница суммы попарных вероятностей

Пусть против переданного сигнала c_j выставлен всего один кандидат c_k , рис. 16(а). Любое из оптимальных правил принятия решения разобьет пространство наблюдений на две области - область $\Gamma_{j/k}$ принятия решения в пользу c_j и ее дополнение, $\Gamma_{k/j}$. Граница областей показана на рисунке пунктирной линией.

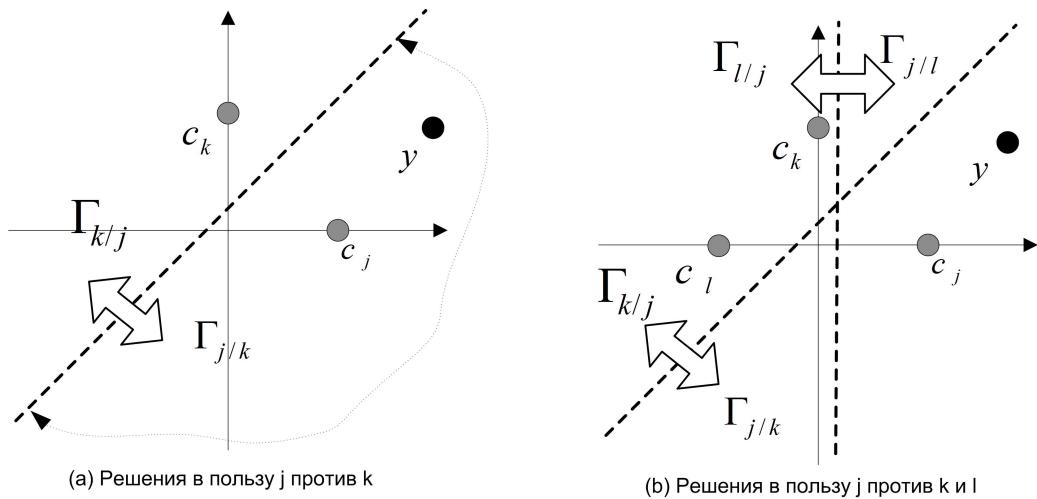


Рис. 16: Области решения в пользу c_j

Упражнение 52 Сформулировать требования к условным плотностям вероятностям $p(y/c)$, выполнение которых гарантирует прямолинейность границы областей Γ .

Добавление еще одного оппонента c_l , рис. 16(b), вводит новое разбиение пространства на области $\Gamma_{j/l}$, $\Gamma_{l/j}$. Сигнал c_j выстаивает против обоих оппонентов, если наблюдение y оказывается принадлежащим пересечению $\Gamma_{j/k} \cap \Gamma_{j/l}$. Обобщая, легко прийти к выводу, что при наличии множества оппонирующих точек областью принятия правильного решения в пользу c_j оказывается пересечение

$$\Gamma_j = \bigcap_{k \neq j} \Gamma_{j/k}.$$

Вероятность правильного решения в пользу c_j есть мера этого пересечения по распределению с плотностью $p(y/c_j)$. В силу этого, для вероятности ошибочного решения при приеме c_j найдем

$$P_{e,j} = 1 - P_j\left(\bigcap_{k \neq j} \Gamma_{j/k}\right) = P_j\left(\bigcup_{k \neq j} \bar{\Gamma}_{j/k}\right) = P_j\left(\bigcup_{k \neq j} \Gamma_{k/j}\right).$$

Оценив сверху вероятностную меру объединения множеств суммой их мер, получим

$$P_{e,j} \leq \sum_{k \neq j} P_j(\Gamma_{k/j}) = \int_{\Gamma_{k/j}} p(y/s_j) dy. \quad (53)$$

Это граница суммы для вероятности ошибки при демодуляции символа c_j .

Замечание 6 Согласно (51), точное выражение для вероятности $P_{e,j}$ имеет вид суммы мер областей Γ_k .

$$P_{e,j} \leq \sum_{k \neq j} P_j(\Gamma_k)$$

Переход к границе (53) эквивалентен замене мер оптимальных областей Γ_k мерами полупространств $\Gamma_{k/j}$. Так что граница суммы оказывается строгой только для созвездия из двух точек.

Применив оценку (53) в (51), получим следующую границу суммы попарных вероятностей для средней вероятности ошибки

$$P_e \leq \sum_{j=1}^M \pi(c_j) \sum_{k \neq j} P_j(\Gamma_{k/j}) = \sum_{j=1}^M \pi(c_j) \sum_{k \neq j} \int_{\Gamma_{k/j}} p(y/s_j).$$

Это граница сводит задачу оценивания средней вероятности ошибки для созвездия к вычислению попарных вероятностей для всех пар сигнальных точек.

Более того, пусть $Q(\mathbb{C})$ - вероятность ошибки для наихудшей пары точек созвездия \mathbb{C} ,

$$Q(\mathbb{C}) = \max_{j \neq k} P_j(\Gamma_{k/j}).$$

Тогда

$$P_e \leq (M-1)Q(\mathbb{C})$$

где M - число точек созвездия.

5.3 Оценка вероятности ошибки для двух точек

Займемся оцениванием вероятности ошибки для двухточечного созвездия на рис. 17. Будем предполагать, что аддитивный шум на выходе согласованного фильтра является комплексным гауссовским с дисперсией $2\sigma^2$.

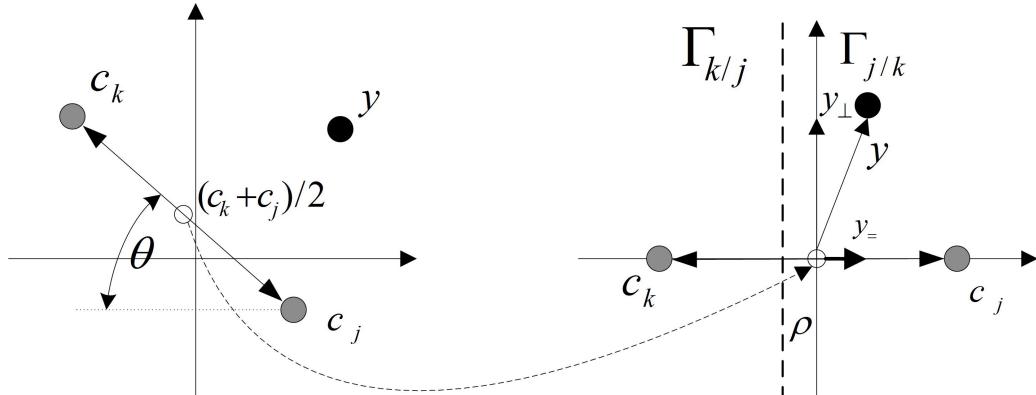


Рис. 17: К оцениванию вероятности ошибки

Тогда условная плотность вероятности отклика y при приеме сигнала c_j принимает вид

$$p(y/c_j) = \frac{1}{2\pi\sigma^2} e^{-\frac{|y-c_j|^2}{2\sigma^2}}.$$

Гауссовская плотность инвариантна по отношению к преобразованиям координат, сохраняющим расстояние между точками. Это дает возможность упростить геометрию задачи.

Параллельным сдвигом комплексной плоскости перенесем в начало координат середину отрезка между сигнальными точками, а затем повернем плоскость на угол θ , так чтобы обе сигнальные точки оказались лежащими на вещественной оси, рис.17. Никакие расстояния при этих преобразованиях не изменились. Далее будем работать в новой системе отсчета.

Построим границы оптимальных областей принятия решений $\Gamma_{k/j}$, $\Gamma_{j/k}$. Критерий максимума апостериорной вероятности предписывает выносить решение в пользу c_j , если наблюдение y удовлетворяет условию

$$\pi(c_j)p(y/c_j) \geq \pi(c_k)p(y/c_k)$$

Введение функции правдоподобия

$$L(y) = \ln \frac{\pi(c_j)p(y/c_j)}{\pi(c_k)p(y/c_k)}$$

позволяет представить это условие в более простой форме - $L(y) > 0$. Вычисление функции правдоподобия для гауссовых распределений дает

$$|y - c_j|^2 - |y - c_k|^2 \leq \sigma^2 \ln \frac{\pi(c_j)}{\pi(c_k)}.$$

Границу раздела между двумя областями принятия решений образуют те точки y , для которых это условие выполняется со знаком равенства. Ясно, что этой границей является вертикальная прямая. При равных априорных вероятностях она просто совпадает с мнимой осью. В общем случае граница оказывается смещенной от начала координат на

$$\rho = \frac{\sigma^2}{2d} \ln \frac{\pi(c_k)}{\pi(c_j)}, \quad d = |c_k - c_j|.$$

Формально области принятия решений в пользу c_j , c_k можно описать так

$$\Gamma_{j/k} = \{y : Re(y) \geq \rho\}, \quad \Gamma_{k/j} = \{y : Re(y) < \rho\}.$$

Заметим, что при такой форме областей решение не зависит от мнимой части наблюдения y_\perp . То есть, вещественная часть y_\parallel оказывается достаточной статистикой.

Оценим, наконец, вероятность ошибки при приеме символа c_j . Имеем

$$P_{e,j} = \int_{\Gamma_{k/j}} p(y/c_j) dy = \frac{1}{2\pi\sigma^2} \int_{\Gamma_{k/j}} e^{-\frac{|y-c_j|^2}{2\sigma^2}} dy_\parallel dy_\perp.$$

Двухмерную плотность $p(y/c_j)$, $y = y_\parallel + iy_\perp$ представим в виде произведения плотностей

$$p(y/c_j) = \frac{1}{2\pi\sigma^2} e^{-\frac{|y-c_j|^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{|y_\parallel - c_j|^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{|y_\perp|^2}{2\sigma^2}}$$

Второй сомножитель - это просто одномерное гауссовское распределение с дисперсией σ^2 . Интеграл от него по dy_\perp дает единицу. Поэтому выражение для вероятности ошибки приводится к виду

$$P_{e,j} = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\rho} e^{-\frac{|y_\parallel - c_j|^2}{2\sigma^2}} dy_\parallel \quad (54)$$

Функция ошибок. Вводится как

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{x^2}{2}} dx$$

Дает оценку площади под гуссовой кривой с единичной дисперсией на интервале (x, ∞) . Ясно, что $Q(0) = 1/2$. Справедлива верхняя граница

$$Q(x) \leq \frac{1}{2}e^{-\frac{x^2}{2}}, \quad x \geq 0.$$

При больших x выполняется более сильная асимптотическая оценка

$$Q(x) \sim e^{-\frac{x^2}{2}},$$

верная в следующем смысле

$$\lim_{x \rightarrow \infty} \frac{-\ln Q(x)}{x^2/2} = 1$$

Через функцию ошибок $Q(x)$ формула (54) для вероятности ошибки представляется в виде

$$P_{e,j} = Q\left(\frac{d}{2\sigma}\left(1 + \frac{\sigma^2}{d^2} \ln \frac{\pi(c_j)}{\pi(c_k)}\right)\right),$$

где $d = |c_k - c_j|$ расстояние между сигнальными точками.

Вероятность ошибки экспоненциально падает с ростом отношения расстояния d к дисперсии шума. Учет априорной информации дает дополнительное снижение вероятности ошибки для символа с большей априорной вероятностью. Наоборот, для априорно менее вероятного символа вероятность ошибки возрастает. Средняя вероятность ошибки в результате снижается.

Принятию решений по максимуму правдоподобия отвечает предположение о равенстве априорных вероятностей - $\pi(c_j) = \pi(c_k)$. В том случае выражение для вероятности ошибки упрощается

$$P_{e,j} = Q\left(\frac{d}{2\sigma}\right) = Q\left(\frac{1}{2}\sqrt{\frac{d^2}{\sigma^2}}\right). \quad (55)$$

Упражнение 53 Пусть априорная вероятность одного из символов стремится к 1, а второго к нулю. Показать, что в этом предельном случае средняя вероятность ошибочного решения по максимуму апостериорной вероятности стремится к нулю, в то время как средняя вероятность ошибочного решения по максимуму правдоподобия остается неизменной.

Формула (55) дает выражение для вероятностей попарных ошибок при демодуляции символов произвольного созвездия по критерию максимального правдоподобия. Попарная вероятность оказывается одинаковой для обоих символов пары и зависит только от расстояния d между сигнальными точками. В каждом созвездии можно выбрать пару точек, находящихся на минимальном расстоянии d_{min} . Тогда, используя границу суммы попарных вероятностей, для средней вероятности ошибки демодуляции созвездия мощности M найдем

$$P_e < (M-1)Q\left(\frac{1}{2}\sqrt{\frac{d_{min}^2}{\sigma^2}}\right).$$

Вероятность ошибки для BPSK.

Простейший вид цифровой модуляции - это двухуровневая фазовая манипуляция BPSK с созвездием из двух точек. Для нее формула (55) дает точное выражение для средней

вероятности ошибки демодуляции символа (бита). Подставив в нее уровни сигналов $c = d/2$, найдем

$$P_e = Q\left(\frac{c}{\sigma}\right) = Q\left(\sqrt{\frac{c^2}{\sigma^2}}\right).$$

Здесь $c^2 = E_s = E_b$ - это энергия сигнала (она же энергия на бит информации), σ^2 - это половина дисперсии комплексного шума на выходе нормированного согласованного фильтра. Полная дисперсия комплексного шума $2\sigma^2$ составляет \mathcal{N}_0 . Поэтому

$$P_e = Q\left(\frac{c}{\sigma}\right) = Q\left(\sqrt{\frac{2E_s}{\mathcal{N}_0}}\right) \simeq e^{-\frac{E_s}{\mathcal{N}_0}}.$$

Это точное выражение для вероятности средней ошибки демодуляции BPSK, выраженное через отношение сигнал/шум E_s/\mathcal{N}_0 .

Замечание 7 То же самое выражение оказывается справедливым для вероятности ошибки на бит при демодуляции QPSK-сигнала.

5.4 Шенноновский предел

Представление о предельно достижимых возможностях передачи цифровой информации дает теорема Шеннона о пропускной способности канала. Этот фундаментальный для теории связи результат позволяет уяснить место известных методов цифровой модуляции и дает представление о путях их совершенствования.

Обсудим результат Шеннона на примере модели канала с аддитивным белым гауссовским шумом (АБГШ).

Пусть в радиоканале с полосой F действует АБГШ со спектральной плотностью \mathcal{N}_0 . Исследуем вопрос о предельном объеме информации, который может быть надежно передан по этому каналу в течение ограниченного времени T .

Канал будем рассматривать как линейное пространство передаваемых сигналов. Как известно, канал с ограничениями по времени и полосе конечномерен. Его комплексная размерность N определяется базой канала FT . Выберем некоторую ортонормированную систему базисных функций

$$\varphi_n(t), \quad \langle \varphi_n(t), \varphi_m(t) \rangle = \delta_{n,m}, \quad n, m \in [1, N].$$

Тогда любой передаваемый сигнал окажется представимым в форме разложения по базису

$$x(t) = \sum_{n=1}^N c_n \varphi_n(t)$$

с вектором комплексных коэффициентов (символов) $\mathbf{c} = (\mathbf{c}_1, \dots, \mathbf{c}_N)$. Для энергии сигнала найдем

$$E = \langle x, x \rangle = \sum c_n c_m^* \langle \varphi_n, \varphi_m \rangle = \sum c_n c_n^* = N E_s,$$

где E_s - средняя энергия на символ.

Принятый с аддитивным шумом сигнал

$$y(t) = x(t) + n(t)$$

будем обрабатывать с помощью N нормированных фильтров, согласованных с базисными функциями. Откликом j -го фильтра

$$y_j = \langle y, \varphi_j \rangle = c_j + \langle n, \varphi_j \rangle = c_j + n_j$$

окажется сумма сигнала c_j и шума n_j с дисперсией $2\sigma^2 = N\sigma^2$. В силу ортогональности фильтров шумы на их выходах можно считать некоррелированными. В результате получим векторный канал

$$\mathbf{y} = \mathbf{c} + \mathbf{n},$$

по которому передаются N -мерные комплексные сигнальные векторы с энергией NE_s . В канале к ним добавляются N -векторы шума с некоррелированными компонентами дисперсии $2\sigma^2$.

Попытаемся оценить количество M сигналов \mathbf{c} с заданной средней энергией E_s , которые могут быть надежно различены в присутствии шума. Для этого нам понадобится следующий результат из теории вероятностей.

Лемма 3 *Пусть x - случайная величина с матожиданием $E(x)$, $\mathbf{x} = (x_1, \dots, x_N)$ - вектор из N ее независимых реализаций. Тогда эмпирическое среднее*

$$\frac{1}{N} \sum x_j$$

сходится к матожиданию в смысле

$$\lim_{N \rightarrow \infty} P\left(\left|\frac{1}{N} \sum x_j - E(x)\right| < \varepsilon\right) = 1.$$

Таким образом, при больших N эмпирическое средние оказывается равным матожиданию с вероятностью единица.

Шум \mathbf{n} - это случайный вектор в линейном пространстве комплексной размерности N (вещественной размерности $2N$). Матожидание квадрата его длины составляет

$$R_n^2 = \sum_j \sum_k E(n_j n_k) = \sum_j E(n_j n_j) = N2\sigma^2.$$

Это означает, что при $N \rightarrow \infty$ эмпирическая длина вектора шума будет совпадать с R_n с вероятностью единица. То есть, при больших N значения случайного шумового вектора с вероятностью единица лежат в $2N$ -мерном пространстве на сфере радиуса $R_n = \sqrt{N2\sigma^2}$.

Рассмотрим теперь принятый вектор \mathbf{y} . Матожидание квадрата его длины составляет

$$R_y^2 = N(E_s + 2\sigma^2).$$

Таким образом, векторы y с вероятностью единица лежат на сфере радиуса $R_y = \sqrt{N(E_s + 2\sigma^2)}$.

В $2N$ -мерном пространстве поверхность сферы радиуса R оценивается величиной $\gamma_N R^{2N}$, где γ_N - некий безразмерный множитель. Оценим число M «шумовых» сфер радиуса R_n , которые могут быть без пересечений упакованы на «сигнальной» сфере радиуса R_y .

$$M \leq \frac{R_y^{2N}}{R_n^{2N}} = \left(\frac{\sqrt{N(E_s + 2\sigma^2)}}{\sqrt{N2\sigma^2}} \right)^{2N} = \left(1 + \frac{E_s}{2\sigma^2} \right)^N.$$

В результате придем к следующей границе сферической упаковки для числа сигналов, различимых в присутствии шума

$$\frac{\log_2 M}{N} \leq \log_2 \left(1 + \frac{E_s}{N_0} \right).$$

Заметим, что $\log_2 M$ - это число битов информации, которые могут быть закодированы набором из M сигналов. Введем скорость передачи $R = \frac{\log_2 M}{T}$ битов в секунду. Припомнив, что $N = FT$, найдем

$$R \leq F \log_2 \left(1 + \frac{E_s}{N_0} \right). \quad (56)$$

Это и есть граница Шеннона для пропускной способности канала с АБГШ.

Теорема 5.1 (Теорема Шеннона) Для всех скоростей передачи, удовлетворяющих границе (56) существует набор сигналов, который обеспечивает сходимость к нулю средней вероятности ошибки демодуляции сигнала P_e с ростом размерности канала ($N \rightarrow \infty$).

Теорема 5.2 (Обращение) Для скоростей передачи, превышающих границу (56) любой способ передачи дает отличную от нуля среднюю вероятность ошибки P_b на бит информации.

Теорема Шеннона устанавливает предел совершенствования систем цифровой модуляции. Причем, этот предел устанавливается абсолютно, для любых отличных от нуля вероятностей ошибки. С шенноновских позиций, схема передачи данных с заданными характеристиками либо не возможна вообще, либо, если возможна, то способна обеспечить сколь угодно малую вероятность ошибки при достаточной размерности канала.

При сопоставлении конкретных систем модуляции с Шенноновским пределом вместо скорости передачи используется инвариантное относительно калибровки размерностей понятие спектральной эффективности

$$\varrho = \frac{\log_2 M}{FT} = \frac{R}{F}.$$

Это безразмерная величина, измеряемая в числе битов в секунду на герц полосы канала, или в числе битов на комплексное измерение.

Выраженный через калибровочно-инвариантные величины - спектральную плотность и отношение сигнал/шум, - шенноновский предел приобретает вид

$$\varrho \leq \log_2 \left(1 + \frac{E_s}{N_0} \right) = \log_2 (1 + SNR)$$

или, что то же самое,

$$SNR \geq 2^\varrho - 1.$$

В этой форме он устанавливает границу снизу на отношение сигнал/шум, требуемое для достижения заданной спектральной эффективности. К примеру, для систем модуляции со спектральной эффективностью $\rho = 2$ (QPSK) граница гарантирует достижение сколь угодно малой вероятности ошибки при $SNR \geq 3/2 = 1.76$ дБ.

Форма границы позволяет выделить два класса систем передачи данных

- Нишу малых спектральных эффективностей ρ занимают системы с высокими требованиями по энергетике, но без особых ограничений на полосу канала. Это область дальней космической связи. Для систем этого рода предел Шеннона приобретает вид

$$SNR \geq \rho \ln 2.$$

Для увеличения скорости передачи здесь требуется лишь пропорциональное увеличение энергии сигнала.

- В области высоких спектральных эффективностей располагаются связные системы с жесткими ограничениями по полосе. Для этой области характерна экспоненциальная зависимость требуемой энергии от скорости передачи. Каждое удвоение скорости связано с увеличением энергии сигнала в 4 раза.

Общепринят следующий универсальный подход к сопоставлению характеристик реальных схем модуляции с шенноновским пределом. Вводится понятие нормированного отношения сигнал/шум $SNR_n = SNR / (2^\rho - 1)$. Для него предел Шеннона принимает вид

$$SNR_n = \frac{SNR}{2^\rho - 1} \geq 1 = 0 \text{dB}.$$

Для конкретной системы модуляции аналитически или моделированием строят зависимость средней вероятности ошибки на символ P_e от нормированного отношения сигнал/шум.

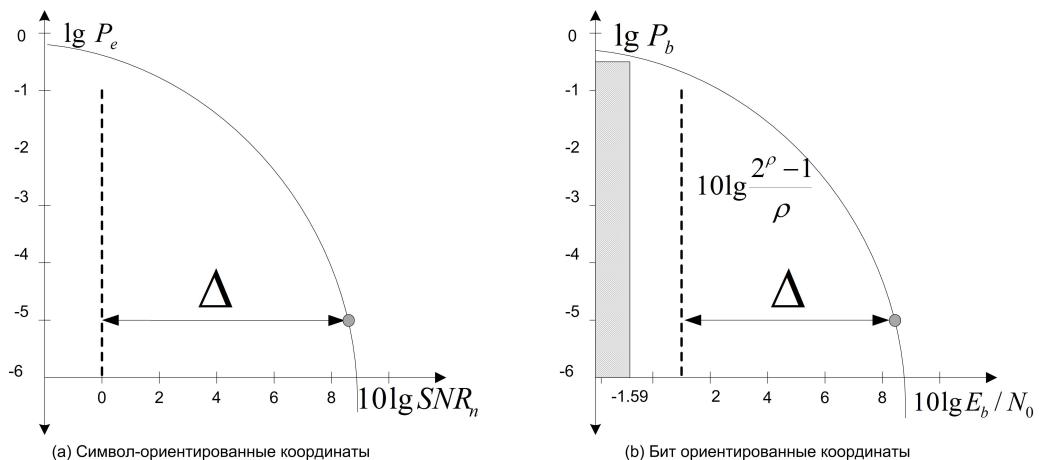


Рис. 18: Сопоставление с пределом Шеннона

Типичная кривая этого рода показана на рис. 18(a). Шенноновский предел на этом рисунке - это вертикальная линия на уровне 0 дБ. При нормированном отношении сигнал/шум меньше 0 дБ никакая передача данных невозможна. Пусть Ваша система модуляции обеспечивает вероятность ошибки на символ $P_e = 10^{-5}$ при $SNR_n \sim 8$ дБ. Тогда Вы проигрываете Шеннону $\Delta = 8$ децибел и можно думать об улучшении системы. Проигрыш этого порядка характерен для всех систем линейной модуляции.

Универсальный подход предполагает нормирование всех величин - отношения сигнал/шум, спектральной плотности, вероятности ошибки, - на символ (комплексное измерение). В области малых спектральных эффективностей популярна несколько другая нормировка. Здесь каждый символ переносит немного битов информации.

Поэтому возникает желание перейти от символ-ориентированной нормировки к бит-ориентированной.

Вводится новое понятие энергии на бит. Передача N символов канала связана с затратой энергии NE_s и переносит $\log_2 M$ битов информации. Тогда энергия на бит составляет

$$E_b = \frac{NE_s}{\log_2 M} = \frac{E_s}{\rho}, \quad E_s = E_b \rho.$$

В этих координатах шенноновский предел представляет в виде

$$\varrho \leq \log_2 \left(1 + \frac{E_b}{N_0} \varrho \right),$$

или

$$\frac{E_b}{N_0} \geq \frac{2^\varrho - 1}{\varrho}.$$

При $\varrho \rightarrow 0$ отсюда получается «окончательный» шенноновский предел

$$\frac{E_b}{N_0} \geq \ln 2 = -1.57 \text{дБ.}$$

Он устанавливает ограничение на величину энергии E_b , которую необходимо затратить на передачу одного бита информации вообще, безотносительно к скорости. При $E_b < N_0$ никакая передача невозможна в принципе.

Сравнение характеристик методов модуляции в этих координатах проводят при фиксированной спектральной эффективности, рис. 18(b). Проводится вертикальная линия, обозначающая шенноновский предел для данной эффективности ϱ . Для конкретного метода модуляции так или иначе строится кривая зависимости вероятности ошибки на бит от отношения сигнал/шум, выраженного в E_b/N_0 . Тогда становится ясно, что, скажем, при вероятности ошибки 10^{-5} мы проигрываем Шеннону около 7 дБ. При меньшей вероятности ошибки проигрыш будет значительнее.

Замечание 8 Сравнение в бит-ориентированных координатах требует вычисления вероятности ошибки на бит. Для много точечных созвездий типа QAM – 128 задача пересчета средней вероятности ошибки на символ созвездия в среднюю вероятность ошибки на бит становится не простой. Более того, ответ оказывается зависящим от выбора конкретной маркировки точек созвездия наборами битов.

Упражнение 54 Оценить расстояние до шенноновского предела для QPSK при $P_b = 10^{-6}$. Принять $\rho = 2$, $E_b = E_s$, $P_b = e^{-\frac{E_b}{N_0}}$.

6 Фазовая синхронизация, петли ФАП

Когерентная демодуляция предполагает наличие на принимающей стороне опорной несущей, синхронизованной с несущей принимаемого сигнала с точностью до фазы. Для формирования когерентных опорных несущих применяются петли фазовой автоподстройки (ФАП).

6.1 Идеальная петля ФАП

Идеальная петля ФАП, рис. 19(а), включает фазовый дискриминатор FD, управляемый напряжением генератора VCO и, в общем случае, петлевой фильтр с импульсной реакцией $Kh(t)$.

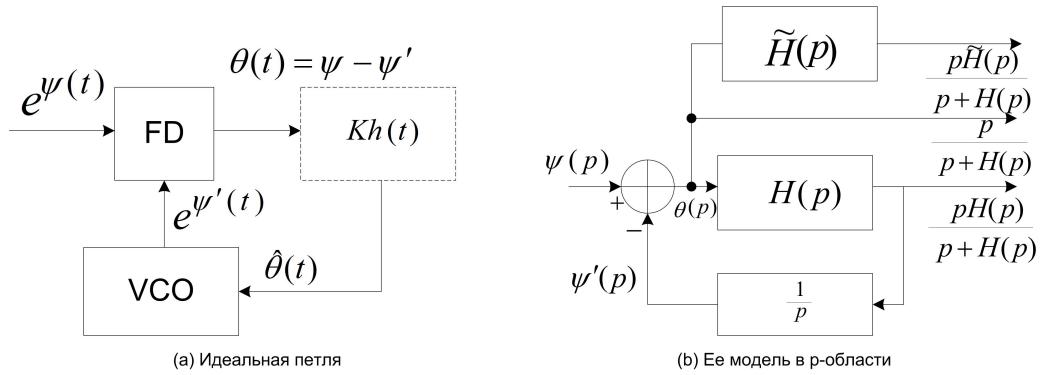


Рис. 19: Идеальная петля и ее p -модель

На вход петли поступает гармоническое колебание с фазой $\psi(t)$. VCO формирует его местный аналог - гармоническое колебание с фазой $\psi'(t)$. Задача петли - обеспечить совпадение фаз $\psi(t)$ и $\psi'(t)$. Она решается типовой петлей автоматического регулирования. Фазовый дискриминатор FD измеряет текущую фазовую ошибку $\theta(t) = \psi - \psi'$, которая, после должного усиления (K) и фильтрации ($h(t)$), используется для управления местным генератором в сторону уменьшения текущей ошибки.

VCO как интегратор

Специфика петли ФАП как системы автоматического регулирования определяется наличием в ней идеального интегратора - VCO. Фаза сигнала управляемого генератора есть интеграл от управляющего сигнала.

$$\psi'(t) = D \int_0^t \hat{\theta}(u) du.$$

Комплексный коэффициент передачи интегратора, рассматриваемого как линейный фильтр, имеет вид

$$K(f) = \frac{D}{i f} = \frac{1}{i f/D}.$$

Он вполне определяется единственным конструктивным параметром D размерности частоты. Параметр этот можно трактовать как меру девиации (отклонения) частоты при единичном воздействии. С другой стороны, D - это та частота, на которой модуль комплексного коэффициента передачи равен единице. В этой трактовке D называют частотой единичного усиления интегратора.

Отклик интегратора на единичное воздействие неограниченно возрастает Dt . Это говорит о физической не реализуемости интегратора как фильтра. Другим указанием на не реализуемость служит бесконечность коэффициента передачи на нулевой частоте. VCO не является линейным фильтром и реализуемый им интегратор идеален в смысле отсутствия физических ограничений на рост фазы ψ' . Правда, фазу как таковую он нам не предъявляет, ограничиваясь выдачей гармонического сигнала $e^{\psi'}$, который позволяет судить о фазе только с точностью до числа оборотов.

Уравнение ФАП

Согласно рис. 19(а), имеем $\theta = \psi - \psi'$. Но

$$\psi'(t) = D \int_0^t \hat{\theta}(u) du = KD \int_0^t (h * \theta)(u) du, \quad \text{где } (h * \theta)(u) \text{- свертка.}$$

В итоге приходим к следующему уравнению идеальной петли ФАП

$$KD \int_0^t (h * \theta)(u) du + \theta(t) = \psi(t).$$

Параметр KD имеет смысл коэффициента передачи разомкнутой петли. Он характеризует величину сдвига частоты VCO при единичной фазовой ошибке. Этот коэффициент имеет размерность частоты, что отражает факт присутствия в петле идеального интегратора.

Разумно ввести параметр $F_1 = KD$ - частоту единичного усиления петли, который объединяет в себе как меру D девиации частоты VCO, так и все возможные в петле частотно-независимые коэффициенты усиления. Удобно нормировать физические времена t_p и частоту f_p на частоту единичного усиления, положив

$$t = F_1 t_p, \quad f = f_p / F_1.$$

В этой нормировке в уравнении петли размерный множитель KD станет единицей. Теперь, применив преобразование Лапласа, перейдем в p -область. Получим

$$\left(\frac{H(p)}{p} + 1 \right) \theta(p) = \psi(p).$$

Отсюда найдем ключевое для нас выражение для передаточной функции петли от фазы сигнала ψ до фазовой ошибки θ

$$\mathbb{H}(p) = \frac{\theta(p)}{\psi(p)} = \frac{p}{p + H(p)}. \quad (57)$$

Петля ФАП, 19(б), оказывается эквивалентной охваченному обратной связью усилителю. Единственное отличие - это наличие идеального интегратора $1/p$ в обратной связи. Если сигнал фазовой ошибки имеет самостоятельное прикладное значение, как это бывает в случае частотного демодулятора на ФАП, то может представлять интерес результат его фильтрации либо непосредственно петлевым фильтром $H(p)$, либо отдельным фильтром $\tilde{H}(p)$. Соответствующие передаточные функции показаны на рис.19(б).

6.2 О передаточных функциях

Передаточная функция $\mathbb{H}(p)$ позволяет судить

- О динамических свойствах петли - характере протекающих в ней переходных процессов.

- О асимптотическом поведении петли при $t \rightarrow 0$ и $t \rightarrow \infty$.

Переходные процессы

Прежде всего, всякая передаточная функция есть рациональная функция комплексной частоты p

$$\mathbb{H}(p) = \frac{Q(p)}{R(p)},$$

представимая в виде отношения двух многочленов с вещественными коэффициентами.

С точностью до множителя рациональная функция вполне определяется набором своих **нулей** (корней $Q(p)$) и **полюсов** (корней $R(p)$). У многочлена с вещественными коэффициентами корни могут быть либо вещественными, либо образовывать комплексно сопряженные пары.

Представление о переходных процессах в петле дают полюсы передаточной функции - собственные частоты автономной системы. Вещественному полюсу $-\sigma$ отвечает собственное решение автономной системы вида $e^{-\sigma t}$, паре комплексно сопряженных полюсов $-\sigma \pm if$ - колебательное собственное решение $e^{-\sigma t} \cos(ft + \varphi)$. Любой переходный процесс в петле есть линейная форма от собственных решений, коэффициенты которой задаются начальными условиями.

Петля **устойчива**, если и только если все полюсы ее передаточной функции лежат в левой полуплоскости. У устойчивой петли переходные процессы экспоненциально затухают во времени.

Скорость затухания переходных процессов определяется значениями вещественных частей полюсов.

С вещественными полюсами все просто - чем левее полюс, тем быстрее затухание.

Пара комплексно сопряженных полюсов - это решения некоторого квадратного уравнения вида $p^2 + 2\delta f_0 p + f_0^2$. Нормировка частоты p на f_0 приводит это уравнение к стандартному виду

$$p^2 + 2\delta p + 1,$$

единственным параметром которого оказывается постоянная затухания δ . Обратную величину $Q = \frac{1}{2\delta}$ называют добротностью пары полюсов.

Полюсы устойчивы (лежат в левой полуплоскости), если $\delta > 0$.

При малых $\delta < 1$ имеется пара комплексно-сопряженных полюсов $p_{1,2} = -\delta \pm i\sqrt{1 - \delta^2}$, которые лежат на единичной окружности, приближаясь к мнимой оси при $\delta \rightarrow 0$. При $\delta > 1$ полюсы становятся вещественными - $p_{1,2} = -\delta \pm \sqrt{\delta^2 - 1}$. С ростом δ один из них уходит в бесконечность, а второй стремится к нулю как $1/2\delta$. Критическому затуханию $\delta = 1$ отвечает кратный вещественный полюс.

С точки зрения времени переходных процессов плохим оказывается выбор как малых, так и больших затуханий. Малым затуханиям отвечают продолжительные переходные процессы колебательного характера. При больших затуханиях появляется «ленивое» собственное решение, отвечающее полюсу $1/2\delta$ в окрестности нуля. Возникает задача оптимизации затухания по критерию минимума продолжительности переходных процессов. Оптимум находится где-то в окрестности критического затухания $\delta \simeq 1$, $Q \simeq 1/2$.

Асимптотика решений, установившиеся значения

Основу для анализа асимптотики по виду передаточной функции дает следующий результат

Теорема 6.1 Пусть $y(t)$, $t > 0$ функция, $Y(p)$ - ее преобразование Лапласа. Тогда

$$\lim_{t \rightarrow 0} y(t) = \lim_{p \rightarrow \infty} pY(p),$$

$$\lim_{t \rightarrow \infty} y(t) = \lim_{p \rightarrow 0} pY(p).$$

Пусть на вход системы с передаточной функцией $\mathbb{H}(p)$ подан тестовый сигнал в виде единичной ступени $1(t) = 1$, $t > 0$ с преобразованием Лапласа $1(p) = 1/p$. Отклик системы на единичную ступень - переходная характеристика $h(t)$, в p -области будет иметь вид $h(p) = \frac{\mathbb{H}(p)}{p}$. Привлекая предыдущую теорему к нахождению асимптотик поведения переходной характеристики, найдем

$$\lim_{t \rightarrow 0} h(t) = \lim_{p \rightarrow \infty} \mathbb{H}(p),$$

$$\lim_{t \rightarrow \infty} h(t) = \lim_{p \rightarrow 0} \mathbb{H}(p).$$

Значение переходной характеристики при $t \rightarrow 0$ оказывается связанным с усилением системы в бесконечности. Если усиление $\mathbb{H}(f)$ не стремится к нулю при $f \rightarrow \infty$, то на переходной характеристике наблюдается резкий скачок при $t = 0$. Если же $\mathbb{H}(f)$ имеет нуль в бесконечной удаленной точке, то $h(t)$ нарастает плавно.

Наоборот, установившееся значение переходной характеристики - $h(\infty)$ определяется значением коэффициента передачи в нуле. Если $\mathbb{H}(f)$ имеет нуль в нулевой точке p -плоскости, то установившееся значение оказывается нулевым. Петля отрабатывает скачок на входе с нулевой ошибкой.

Помимо единичной ступени, при анализе систем автоматического регулирования рассматривают тестовые сигналы $1_k(t)$, являющиеся k -кратными интегралами от $1(t)$ по времени

$$1_1(t) = \int_0^t 1(u)du = t, \quad 1_2(t) = \int_0^t 1_1(u)du = t^2/2, \quad \dots,$$

с преобразованиями Лапласа

$$h_k(t) = \frac{1}{p^{k+1}}$$

Асимптотическое поведение соответствующих импульсных реакций $h_k(t)$ характеризуют соотношения

$$\lim_{t \rightarrow 0} h_k(t) = \lim_{p \rightarrow \infty} \frac{\mathbb{H}(p)}{p^k},$$

$$\lim_{t \rightarrow \infty} h_k(t) = \lim_{p \rightarrow 0} \frac{\mathbb{H}(p)}{p^k}.$$

Применительно к системам ФАП $h(t)$ - это реакция на единичный скачок в фазе, $h_1(t)$ - реакция на скачок по частоте, $h_2(t)$ - реакция на скачок в скорости изменения частоты и т.д.

Существует эффекта Допплера, который приводит к сдвигу частоты сигнала, пропорциональному скорости движения излучателя, $\Delta f/f_0 = v/c$. Он позволяет придать нашим тестовым сигналам следующую физическую трактовку: $1(t)$ - это скачок дальности, $1_1(t)$ - скачок скорости, $1_2(t)$ - скачок ускорения.

При связи с подвижными объектами актуальной становится задача синхронизации несущей с учетом допплеровских сдвигов, вызванных изменениями скорости излучающего

объекта. Требование нулевого установившегося значения фазовой ошибки в ответ на скачок по частоте эквивалентно выполнению условия

$$\lim_{t \rightarrow \infty} h_1(t) = \lim_{p \rightarrow 0} \frac{\mathbb{H}(p)}{p} = 0.$$

Иными словами, передаточная функция петли должна иметь в точке $p = 0$ нуль кратности, как минимум, два.

6.3 Петля ФАП без фильтра

В отсутствие петлевого фильтра передаточная функция петли приобретает вид

$$\mathbb{H}(p) = \frac{p}{1 + p}.$$

Напомним, что время и частота нормированы у нас на частоту единичного усиления петли F_1 .

Имеем, $\mathbb{H}(\infty) = 1$. Это означает, что реакцией петли на скачок фазы ψ является такой же скачок фазовой ошибки θ . С другой стороны, поскольку $\mathbb{H}(0) = 0$, установившееся значение фазовой ошибки $\theta(\infty)$ оказывается нулевым. Впрочем, для этой петли форма импульсной реакции на фазовый скачок легко записывается в явном виде

$$h(t) = 1(t)e^{-t} = 1(t)e^{-F_1 t_p},$$

где t_p -физическое время, а $t = F_1 t_p$ - безразмерное.

Рассмотрим единичный скачок по частоте $1_1(t)$. Установившимся значением импульсной реакции $h_1(t)$ оказывается

$$\lim_{p \rightarrow 0} \frac{\mathbb{H}(p)}{p} = \lim_{p \rightarrow 0} \frac{1}{1 + p} = 1$$

То есть, скачок по частоте на ΔF приведет к установившейся ошибке

$$\Delta\theta = \frac{\Delta F}{F_1} \text{ оборотов или } 2\pi \frac{\Delta F}{F_1} \text{ радианов.}$$

Упражнение 55 В физических единицах выпишать формулу для импульсной реакции $h_1(t)$ на ошибку по частоте.

Увеличением усиления петли $F_1 \rightarrow \infty$, в принципе, достигаются все цели. Реакция на ошибку по фазе становится сколь угодно короткой, а установившаяся ошибка по частоте - сколь угодно малой.

В отсутствие шума больше желать нечего. Проблема только в том, что наша петля крайне широкополосна - частота единичного усиления разомкнутой петли составляет F_1 .

Шум эфира, воздействуя на фазовый дискриминатор, приводит к появлению шумовой компоненты фазовой ошибки. Далее весь этот шум напрямую поступает на вход управления VCO, что вызывает флюктуации фазы местной несущей. Для систем выделения несущей такая ситуация неприемлема. В когерентной системе связи фазовый шум опорного генератора несущей должен, по возможности, отсутствовать.

Осознав это, пробуем сузить полосу петли введением петлевого фильтра.

6.4 Петли с фильтром первого порядка

Начнем с простого.

Интегрирующая цепь

Введем в петлю фильтр нижних частот первого порядка в виде простой интегрирующей цепи

$$H = \frac{1}{1 + i(f_p/F_h)} = \frac{1}{1 + i(f_p/F_1)(F_1/F_h)} = \frac{1}{1 + i(f/\alpha)} = \frac{1}{1 + p/\alpha}.$$

Здесь F_h - верхняя частота интегрирующей цепи, а $\alpha = F_h/F_1$ - она же в безразмерных единицах.

Подставляя это в выражение (57) для передаточной функции петли найдем

$$\mathbb{H}(p) = \frac{p(1 + p/\alpha)}{p^2/\alpha + p + 1}.$$

Перенормируем частоту, положив $s^2 = p^2/\alpha$. Получим,

$$\mathbb{H}(s) = \frac{s^2 + \sqrt{\alpha}s}{s^2 + \sqrt{\alpha}s + 1}.$$

Легко проверить, что в новой нормировке единицей частоты стала величина $\sqrt{F_h F_1}$, а сами нормировочные соотношения приняли вид

$$f = \frac{f_p}{\sqrt{F_h F_1}}, \quad \alpha = \frac{F_h}{F_1}$$

Результат неутешителен. Чтобы попасть в область критических затуханий, необходимо иметь $\delta = \sqrt{\alpha}/2 \simeq 1$ или $\alpha = \frac{F_h}{F_1} \simeq 4$. То есть, интегрирующая цепь должна быть примерно в 4 раза широкополоснее петли. Это бессмысленно. Сужение полосы петли ($\alpha \rightarrow 0$) дает пару полюсов с низким затуханием. Это ведет к продолжительному переходному процессу колебательного характера.

Упражнение 56 Показать, что скорость затухания переходных процессов в рассматриваемой петле имеет порядок

$$e^{-\sqrt{\frac{F_h}{F_1}} F_h t_p}$$

Асимптотически при $\alpha \rightarrow 0$ получаем передаточную функцию

$$\mathbb{H}(s) = \frac{s^2}{s^2 + 1}$$

с двумя полюсами на мнимой оси, переходные процессы в которой не затухают вовсе.

Пропорционально интегрирующая цепь

Передаточная функция пропорционально интегрирующей цепи имеет вид

$$H = \frac{1 + \gamma \frac{p}{\alpha}}{1 + \frac{p}{\alpha}},$$

отличаясь от передаточной функции интегрирующей цепи одним дополнительным параметром γ , имеющим смысл коэффициента передачи на бесконечной частоте. Коэффициент передачи цепи изменяется от 1 на нулевой частоте до уровня γ на высоких частотах, падая до уровня $1/\sqrt{2}$ на частоте порядка $F_h = \alpha F_1$.

Подставляя это в выражение (57) и применив ту же перенормировку, что и выше, найдем

$$\mathbb{H}(s) = \frac{s^2 + \sqrt{\alpha}s}{s^2 + \sqrt{\alpha} \left(1 + \frac{\gamma}{\alpha}\right)s + 1}.$$

Для этой передаточной функции постоянная затухания имеет вид

$$\delta = \frac{1}{2} \left(\sqrt{\alpha} + \frac{\gamma}{\sqrt{\alpha}} \right)$$

и оказывается растущей как при больших, так и при малых значениях α .

Задача выбора параметров оптимального пропорционально интегрирующего фильтра для петли ФАП решается в следующем контексте. Задаемся некоторым приемлемым значением γ уровня подавления шума в дальней зоне, скажем, полагаем $\gamma = 10^{-3}$. Тогда попадание постоянной затухания δ в область критических значений обеспечивается при значениях $\alpha = \frac{F_h}{F_1} \simeq \gamma^2 = 10^{-6}$. При заданной полосе фильтра F_h этого всегда можно добиться выбором достаточно большого коэффициента передачи разомкнутой петли F_1 . Ввиду плавности фигурирующих в оптимизации зависимостей, большой точности в выборе параметров не требуется. Достаточно, чтобы порядки величин были выбраны разумно.

6.5 Петли второго порядка - два интегратора.

Нулевой уровень установившегося значения ошибки по фазе в ответ на скачок частоты обеспечивают только петли с дополнительным интегратором в качестве петлевого фильтра. Наивная подстановка передаточной функции интегратора $1/p$ в (57) приводит к передаточной функции неустойчивой петли

$$\mathbb{H}(p) = \frac{p^2}{1 + p^2}$$

с двумя полюсами на мнимой оси. Усовершенствование состоит во введении пропорциональной компоненты. Передаточную функцию петлевого фильтра выбирают в виде

$$H(p) = \frac{\alpha}{p} + \gamma,$$

где $\alpha = F_h/F_1$ - частота единичного усиления второго интегратора, нормированная на усиление петли F_1 . Подстановка этого в (57), после нашей стандартной перенормировки, дает

$$\mathbb{H}(s) = \frac{s^2}{s^2 + \frac{\gamma}{\sqrt{\alpha}}s + 1}.$$

С переходными процессами здесь все в порядке. При любой заданной полосе единичного усиления петлевого интегратора F_h выбором постоянной $\gamma \simeq 2\sqrt{\alpha}$ делаем передаточную функцию около критической. Передаточная функция имеет в нуле нуль кратности два, что обеспечивает равенство нулю установившегося значения переходной характеристики $h_1(t)$ - реакции на единичную ступень по частоте $1_1(t)$.

Основную сложность здесь представляет проблема реализации «настоящего» интегратора с реальным полюсом передачи на нулевой частоте. Никакие компромиссы здесь не проходят. Любая попытка приближенной реализации интегратора в виде интегрирующей цепи с «очень большой» постоянной времени немедленно приводит к варианту петли с пропорционально интегрирующей цепью, для которой установившееся значение ошибки по частоте равно 1.

Петли этого класса, в основном, реализуются в цифровой форме. Цифровая реализация идеального интегратора, как раз, тривиальна. В его роли может выступать, например, обычный счетчик.

6.6 Фазовые дискриминаторы. Эффекты неидеальности.

Основной источник проблем с петлями ФАП скрыт в реализации фазового дискриминатора. В идеальном варианте фазовый дискриминатор строится на базе квадратурного демодулятора, который восстанавливает косинусную и синусную компоненты сигнала относительно местной несущей, рис. 20. Для почти гармонического сигнала эти компоненты оказываются пропорциональными косинусу и синусу текущей ошибки по фазе. Значение фазовой ошибки определяется как арктангенс отношения квадратурных компонент.

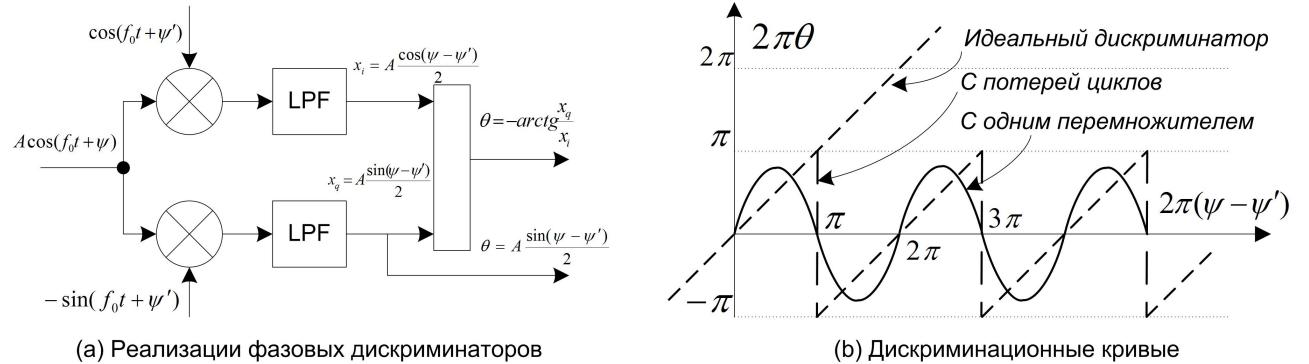


Рис. 20: К обсуждению фазового дискриминатора

Первая сложность связана с периодичностью гармонических функций. Обработка конкретной пары значений квадратурных составляющих дает возможность определить значение фазы только с точностью до периода (оборота).

Идеальный фазовый дискриминатор должен вести счет числа полных оборотов и включать его в оценку текущей фазовой ошибки. В результате получается линейная зависимость выходного сигнала дискриминатора от ошибки по фазе.

Если дискриминатор не ведет счет оборотов, его сигнал оказывается «намотанным на цилиндр», становясь периодической функцией фазовой ошибки. При разумном выборе начала отсчета сигнал дискриминатора линейно изменяется в пределах от $-\pi$ до $+\pi$ - пол оборота в сторону опережения и пол оборота в сторону отставания. В этих пределах дискриминационная кривая остается линейной. Превышение значением фазовой ошибки границы $\pm\pi$ вызывает проскальзывание цикла. Информация о фазовой ошибке на один оборот теряется.

Наиболее распространенная реализация фазового дискриминатора просто использует в качестве сигнала фазовой ошибки выход синусного канала квадратурного демодулятора. При этом диапазон допустимых изменений величины фазовой ошибки составляет всего

$\pm\pi/2$, а зависимость сигнала дискриминатора от фазовой ошибки становится нелинейной. Петля с таким дискриминатором остается линейной только при малых ошибках по фазе.

Вторая важная особенность дискриминаторов этого типа состоит в появлении зависимости сигнала фазового дискриминатора от уровня A сигнала на входе. Это приводит к зависимости от уровня сигнала коэффициента передачи разомкнутой петли. При малых сигналах полоса единичного усиления петли сужается, при больших - пропорционально расширяется. Колебания петлевого усиления оказывают влияние на динамику переходных процессов в петле.

Полоса удержания

Следствием периодичности дискриминационной кривой становится ограниченность полосы удержания.

Пусть частота входного сигнала петли плавно увеличивается. Петля отслеживает это увеличение. В ходе отслеживания в петле растет значение установившейся фазовой ошибки. Но, при периодической кривой FD, диапазон возможных значений фазовой ошибки ограничен. Когда он оказывается исчерпанным, происходит срыв слежения. К примеру, в петлях первого порядка установившаяся ошибка по фазе при сдвиге частоты на ΔF составляет $2\pi \frac{\Delta F}{F_1}$. При пилообразной дискриминационной кривой срыв слежения произойдет при $\Delta F > F_1/2$. Ширина двухсторонней полосы удержания оказывается равной полосе единичного усиления F_1 .

Упражнение 57 Оценить ширину двухсторонней полосы слежения для петли с синусной дискриминационной кривой (Ответ: $\Delta F > AF_1/2$).

В петлях второго порядка установившееся значение фазовой ошибки при слежении за частотой равно нулю. Поэтому, теоретически, полоса слежения безгранична. В практических ситуациях она ограничивается диапазоном перестройки по частоте управляемого генератора.

Упражнение 58 Задав жесткие границы диапазона перестройки управляемого генератора $[-D, +D]$, оценить ширину полосы слежения для петель первого и второго порядков (один и два интегратора).

Полоса захвата

Понятие полосы захвата относится к анализу процесса вхождения петли в синхронизм с внезапно появившимся сигналом.

Пусть петля временно разомкнута. На входе управления VCO-нулевой сигнал и он формирует свою номинальную частоту f_0 . При наличии на входе сигнала с частотой $f_0 + \Delta F$ на выходе фазового дискриминатора будет наблюдаться периодический сигнал частоты ΔF - непрерывное проскальзывание циклов.

Подключим этот сигнал ко входу VCO. Одна из его полуволн будет сдвигать частоту VCO в «правильную» сторону, другая в противоположную. Результатом «правильного» сдвига станет снижение ошибки по частоте, то есть, уменьшение скорости расхождения фаз. Это приведет к увеличению продолжительности полуволны. Продолжительности полуволн обратного знака, наоборот, сокращаются. Это означает появление в сигнале FD постоянной составляющей, которая смещает частоту VCO в нужную сторону. Процесс заканчивается захватом фазы.

В петлях без фильтра полоса захвата имеет ту же ширину, что и полоса удержания. В петлях с полосовым фильтром это не так. Объяснение заключается в том, что узкополосный петлевой фильтр может почти полностью подавить сигнал ошибки на

частоте ΔF . В результате на вход VCO ничего не поступает, постоянная составляющая в сигнале дискриминатора не образуется. Практически полоса захвата узкополосных петель всего лишь в несколько раз шире полосы петлевого фильтра.