Цифровая обработка сигналов

Коржиманов Артём Владимирович

korzhimanov.ru

Понятие сигнала. Аналоговые и цифровые сигналы. Дискретизация

1. Введение в цифровую обработку сигналов (ЦОС)

Мотивация: Почему мы переходим от аналоговой обработки к цифровой?

- **Гибкость:** Одна программа может обрабатывать разные сигналы (аудио, видео, биомедицинские).
- **Точность и воспроизводимость:** Цифровые алгоритмы не подвержены "дрейфу" параметров, как аналоговые компоненты.
- **Сложные алгоритмы:** Возможность реализации методов, недоступных в аналоговой технике (вейвлеты, адаптивная фильтрация).
- **Хранение и передача:** Цифровой сигнал можно сжимать, защищать от ошибок, передавать без потерь.

Области применения:

- системы связи (радиосвязь, мобильная связь, сети интернет)
- аудио/видео кодеки (MP3, H.264)
- медицинская диагностика (ЭКГ, МРТ)
- системы распознавания речи и изображений, радиолокация

Схожие принципы применяются также в:

- **Финансы и алгоритмический трейдинг:** Анализ временных рядов.
- **Data Science и машинное обучение:** Предобработка данных, анализ и выделение признаков из временных рядов и многомерных данных.
- **Теория управления и робототехника:** Дискретное управление системами, анализ и синтез цифровых регуляторов, обработка данных с датчиков.
- **Вычислительная биология и биоинформатика:** Анализ геномных последовательностей, обработка данных спектроскопии, выявление паттернов в биологических временных рядах (например, активность нейронов).
- **Компьютерная графика и физическое моделирование:** Генерация и обработка текстур, создание физически правдоподобных анимаций.

2. Основные понятия

Сигнал — это любая физическая величина, изменяющаяся во времени и несущая информацию.

Цифровой сигнал — это сигнал, представленный в виде **числовой последовательности**, полученной в результате **дискретизации** по времени и **квантования** по уровню.

1. **Информационный носитель:** Исходно сигнал — это изменение во времени или пространстве физической величины (напряжение, давление, интенсивность света, температура и т.д.), которое кодирует информацию (речь, изображение, показания датчика).

2. **Дискретность по времени (или пространству):** В ЦОС работают не с непрерывным сигналом, а с его **отсчётами** (сэмплами), взятыми через равные промежутки времени (**период дискретизации**). Таким образом, сигнал представляется как последовательность `x[n]`, где `n` — целочисленный временной индекс (например, `x[0], x[1], x[2], ...`).

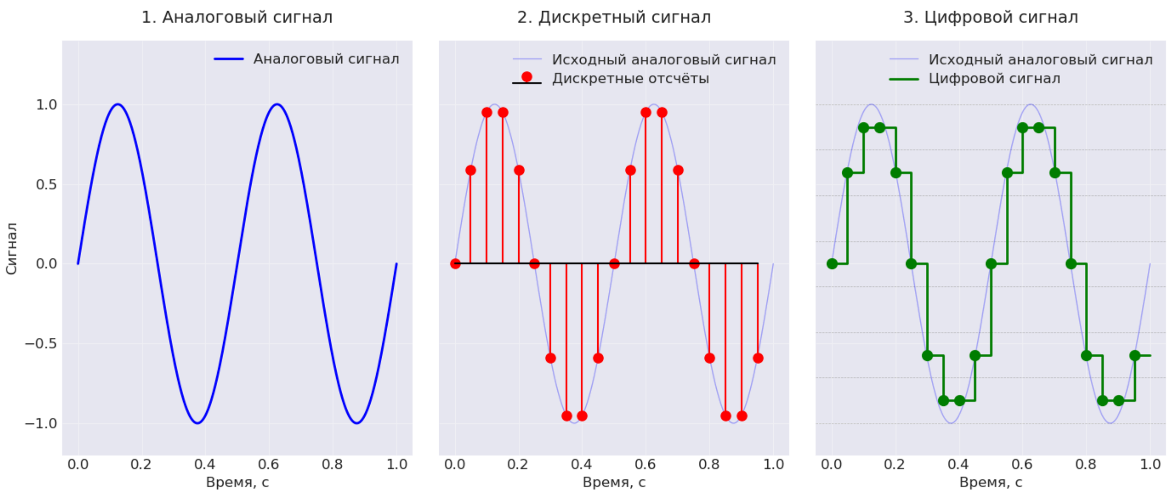
3. **Дискретность по уровню (квантование):** Значение каждого отсчёта не произвольное, а приближённое до ближайшего значения из конечного набора уровней (квантов). Это позволяет хранить и обрабатывать сигнал в цифровой форме.

4. **Математическая абстракция:** В рамках алгоритмов цифровой обработки сигнал `x[n]` рассматривается как **математическая функция** целочисленного аргумента (дискретного времени). Это позволяет применять к нему формальные математические операции: фильтрацию, преобразования (Фурье, Z-преобразование), свёртку и другие.

Важно: В широком смысле, в современном понимании сигналом может выступать любая **упорядоченная последовательность данных** (например, ежемесячные продажи компании, последовательность генов в ДНК, пиксели в строке изображения), даже если они не имеют «физического» аналогового прототипа. Ключевое — наличие порядка (например, временного или пространственного) и возможность применения к этим данным стандартного аппарата цифровой обработки сигналов.

Типы сигналов

Тип сигнала	Определение	Пример
Аналоговый (непрерывный)	Непрерывен по времени и амплитуде	Звуковая волна в воздухе, напряжение на выходе микрофона
Дискретный	Определен только в отдельные моменты времени (отсчеты)	Последовательность значений аналогового сигнала в моменты t_1, t_2, ...
Цифровой	Дискретен и по времени, и по амплитуде (квантован)	Файл WAV, массив целых чисел в памяти компьютера



3. Процесс аналого-цифрового преобразования (АЦП)

АЦП состоит из двух основных этапов:

1. **Дискретизация во времени (sampling)** — взятие отсчетов через равные промежутки времени.
   - **Частота дискретизации (f_s)** — количество отсчетов в секунду (Гц).
   - **Период дискретизации (T_s)** — время между отсчетами: $T_s = 1/f_s$.

2. **Квантование по уровню (quantization)** — приведение каждого отсчета к ближайшему значению из конечного набора.
   - **Разрядность (битность)** — количество бит на отсчет (8, 16, 24 бита).
   - **Шум квантования** — ошибка из-за округления амплитуды.

4. Теорема Котельникова (Найквиста — Шеннона)

Если аналоговый сигнал $x(t)$ имеет **ограниченный спектр** (не содержит частот выше некоторой $f_{\max}$), то он может быть **точно восстановлен** из своих дискретных отсчетов, взятых с частотой:

$$f_s > 2f_{\max}$$

где:

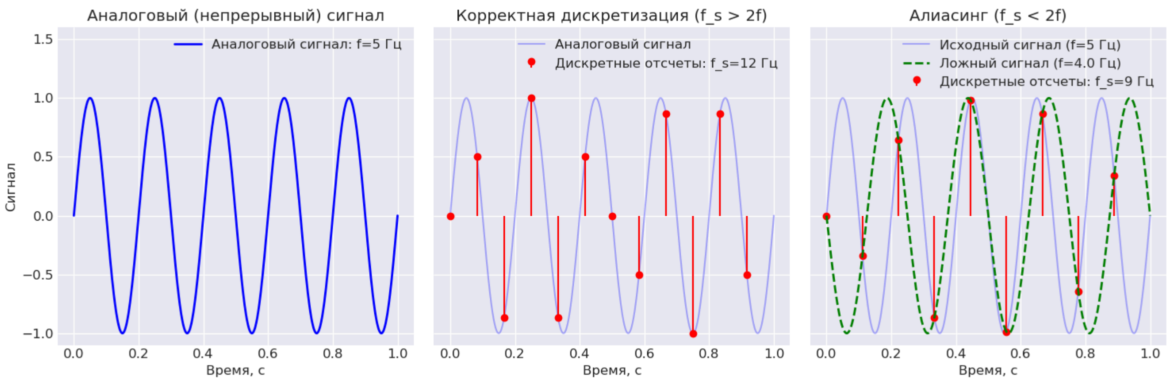
- f_s — частота дискретизации
- $f_{\max}$ — максимальная частота в спектре сигнала
- **Частота Найквиста** $f_N = f_s/2$ — максимальная частота, которая может быть представлена при данной f_s без искажений.

Простой вывод теоремы (идея)

1. Сигнал с ограниченным спектром ($-f_{max} \leq f \leq f_{max}$) можно представить суммой гармоник (синусоид) с частотами от 0 до f_{max}.

2. Рассмотрим самую быструю гармонику с частотой f_{max}. Её период $T_{min} = 1/f_{max}$.

3. Чтобы однозначно задать синусоиду, нужно как минимум **две точки** за период: например, на максимуме и минимуме. Если взять меньше двух точек, то более высокочастотная синусоида может «маскироваться» под более низкочастотную (это явление — алиасинг или наложение спектров).



4. Следовательно, интервал дискретизации T_s должен быть меньше половины минимального периода: $T_s < T_{min}/2$. Это эквивалентно условию для частоты дискретизации: $f_s = 1/T_s > 2f_{max}$.

5. При $f_s = 2f_{max}$ фаза отсчетов может оказаться неудачной (например, взятие отсчетов только в нулях синусоиды), что сделает восстановление невозможным. Поэтому на практике требуется строгое неравенство $f_s > 2f_{max}$.

То же самое в математической форме

Пусть $X(\omega)$ — спектр сигнала $x(t)$.

Так как предполагается, что $X(\omega)$ равен нулю вне полосы $\left|\frac{\omega}{2\pi}\right| < f_{\max}$, то

$$x(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}X(\omega)e^{i\omega t}\,\mathrm{d}\omega=\frac{1}{2\pi}\int_{-2\pi f_{\max}}^{2\pi f_{\max}}X(\omega)e^{i\omega t}\,\mathrm{d}\omega,$$

Если мы положим $t=\frac{n}{2f_{\max}}=nT_{\min}/2=nT_s$, где n — целое число, то получим:

$$x\left(nT_s\right)=\frac{1}{2\pi}\int_{-2\pi f_{\max}}^{2\pi f_{\max}}X(\omega)e^{i\omega nT_s}\,\mathrm{d}\omega.$$

$$x\left(nT_s\right)=\frac{1}{2\pi}\int_{-2\pi f_{\max}}^{2\pi f_{\max}}X(\omega)e^{i\omega nT_s}\,\mathrm{d}\omega.$$

Слева стоят значения $x(t)$ в точках отсчёта.

Интеграл справа, по сути, представляет собой n-ый коэффициент в разложении функции $X(\omega)$ в ряд Фурье.

Это означает, что значения отсчётов $x(nT_s)=x(n/2f_{\max})$ определяют коэффициенты Фурье в разложении $X(\omega)$.

Таким образом, они определяют и саму функцию $X(\omega)$, поскольку $X(\omega)$ равна нулю для частот выше $f_{\max}$, а для более низких частот $X(\omega)$ определяется, если известны её коэффициенты Фурье.

Но $X(\omega)$ полностью определяет исходную функцию $x(t)$, так как функция определяется, если известен её спектр. Следовательно, исходные отсчёты полностью определяют функцию $x(t)$.

Как точно восстановить сигнал (Формула Уиттекера — Котельникова — Шэннона)

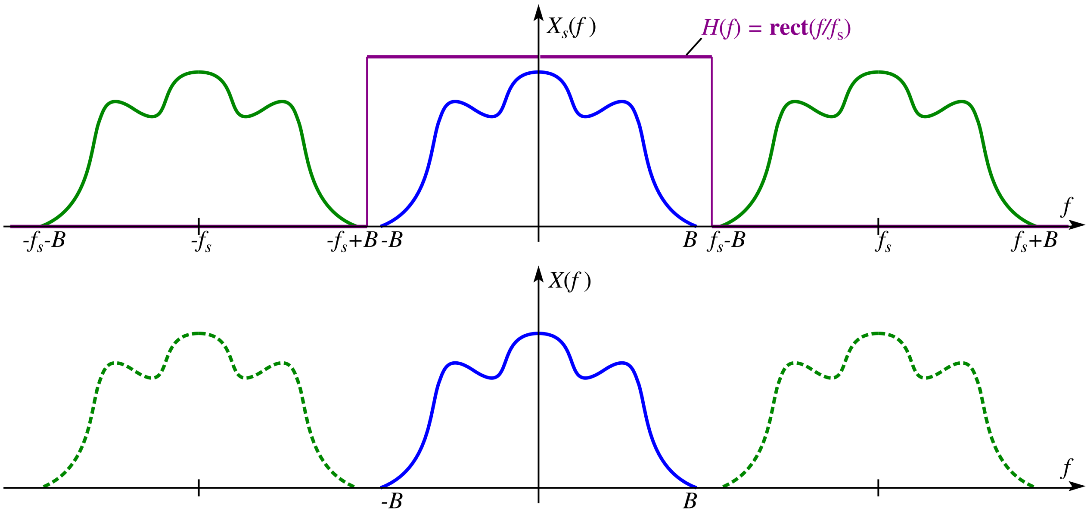
Разложение в ряд Фурье в частотной области

Рассмотрим сигнал $x(t)$ с **ограниченным спектром**: его Фурье-образ $X(f)=0$ для всех $|f|\geq f_{\max}$.

Введём функцию $\tilde{X}(f)$, являющуюся периодическим продолжением функции $X(f)$ с периодом $2f_{\max}$ за пределы интервала $[-f_{\max},f_{\max}]$, тогда:

$$X(f)=\tilde{X}(f)\cdot\Pi\left(\frac{f}{f_{\max}}\right)$$

где $\Pi(x)$ — прямоугольная функция, равная 1 при $|x|<1$ и 0 при $|x|\geq 1$.



Поскольку функция $\tilde{X}(f)$ периодическая, то её можно разложить в ряд Фурье:

$$\tilde{X}(f) = \sum_{n=-\infty}^{\infty} C_n \cdot e^{-j2\pi n \frac{f}{2f_{\max}}} = \sum_{n=-\infty}^{\infty} C_n \cdot e^{-j2\pi f n T_s}, \quad |f| < f_{\max}$$

Коэффициенты Фурье C_n вычисляются как:

$$C_n = \frac{1}{2f_{\max}} \int_{-f_{\max}}^{f_{\max}} \tilde{X}(f) e^{j2\pi f n T_s} df = \frac{1}{2f_{\max}} \int_{-f_{\max}}^{f_{\max}} X(f) e^{j2\pi f n T_s} df$$

Связь коэффициентов C_n с отсчётами сигнала

Применим обратное преобразование Фурье к $X(f)$, получаем:

$$x(t) = \int_{-\infty}^{+\infty} X(f) e^{j2\pi f t} df = \int_{-\infty}^{+\infty} \tilde{X}(f) \Pi\left(\frac{f}{f_{\max}}\right) e^{j2\pi f t} df = \int_{-f_{\max}}^{f_{\max}} \tilde{X}(f) e^{j2\pi f t} df$$

Подставим разложение $\tilde{X}(f)$ в ряд Фурье:

$$x(t) = \int_{-f_{\max}}^{f_{\max}} \left(\sum_{n=-\infty}^{\infty} C_n e^{-j2\pi f n T_s} \right) e^{j2\pi f t} df$$

Поменяем порядок суммирования и интегрирования:

$$x(t) = \sum_{n=-\infty}^{\infty} C_n \int_{-f_{\max}}^{f_{\max}} e^{j2\pi f (t - nT_s)} df$$

Вычислим интеграл:

$$\int_{-f_{\max}}^{f_{\max}} e^{j2\pi f (t - nT_s)} df = 2f_{\max} \cdot \operatorname{sinc}\left(\frac{t}{T_s} - n\right)$$

где $\operatorname{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$.

Таким образом:

$$x(t) = 2f_{\max} \sum_{n=-\infty}^{\infty} C_n \cdot \operatorname{sinc}\left(\frac{t}{T_s} - n\right)$$

Выражение коэффициентов C_n через отсчёты сигнала

Вычислим C_n при $t = nT_s$:

$$x\left(nT_s\right) = 2f_{\max} \sum_{k=-\infty}^{\infty} C_k \cdot \operatorname{sinc}(n-k)$$

Учтём, что $\operatorname{sinc}(n-k) = 1$ при $n = k$ и 0 при $n \neq k$, получаем:

$$x\left(nT_s\right) = 2f_{\max} C_n$$

Отсюда:

$$C_n = \frac{1}{2f_{\max}} x\left(nT_s\right)$$

То есть коэффициенты ряда Фурье C_n — это просто значения сигнала $x(t)$ в точках отсчёта nT_s (с нормировкой, но это непринципиально).

Формула восстановления (ряд Уиттекера — Котельникова — Шеннона)

Подставим C_n в выражение для $x(t)$, получаем окончательную формулу:

$$x(t) = \sum_{n=-\infty}^{\infty} x\left(nT_s\right) \cdot \operatorname{sinc}\left(\frac{t}{T_s} - n\right)$$

Примечание

Можно сделать быстрее, если воспользоваться свойством преобразования Фурье: умножение спектров соответствует свёртке оригиналов, и знать, что обратное преобразование Фурье от прямоугольной функции — функция sinc:

$$\mathcal{F}^{-1}\left[\Pi\left(\frac{f}{f_{\max}}\right)\right] = 2f_{\max} \cdot \operatorname{sinc}(2f_{\max} t)$$

Важно

- Для точной реконструкции сигнала требуется применение специального преобразования (интерполяции, ресепмплинга) по формуле Уиттекера — Котельникова — Шеннона
- При $f_s = 2f_{\max}$ фаза отсчётов для правильной реконструкции амплитуды должна попадать точно в максимум синуса, поэтому надо брать $f_s > 2f_{\max}$

- Если спектр сигнала ограничен и снизу частотой $f_{\min}$, то достаточно $f_s = 2(f_{\max} - f_{\min})$. Но для восстановления требуется знать центральную частоту $f_0 = (f_{\max} + f_{\min})/2$.

5. Алиасинг (наложение спектров)

Что происходит при нарушении теоремы?

Если $f_s < 2f_{\max}$, возникает **алиасинг** (*alias* — псевдоним, выдающий себя за другого).

Механизм: Высокочастотные компоненты сигнала "маскируются" под низкочастотные после дискретизации.

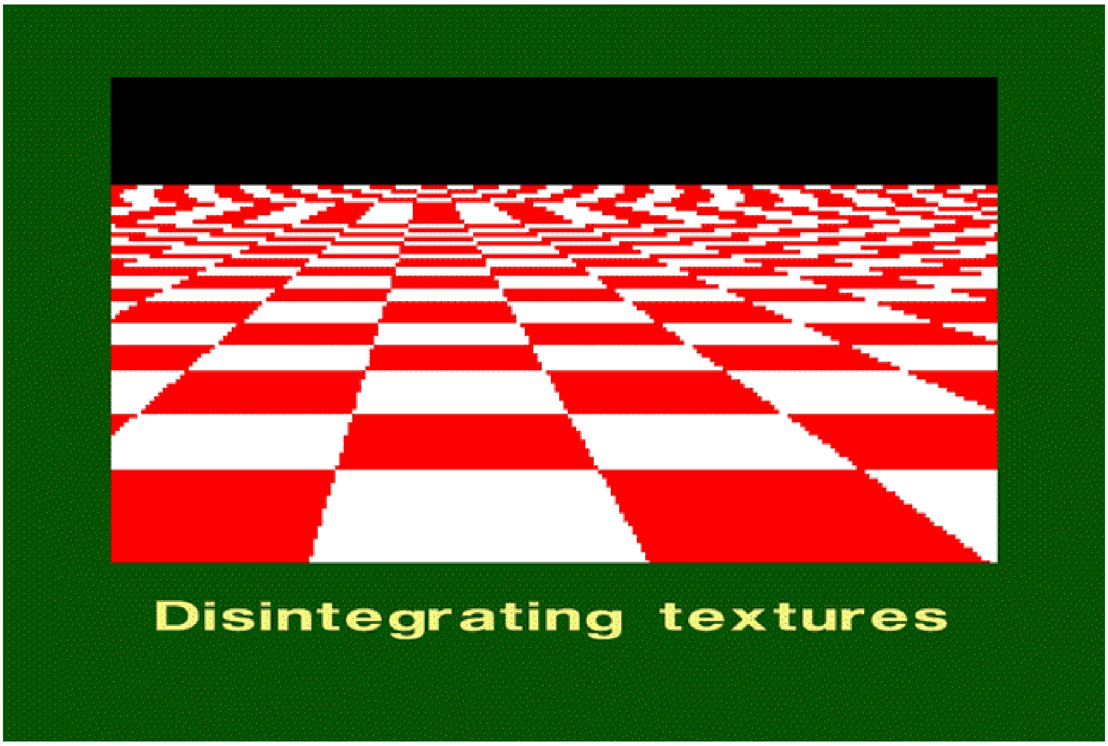
Примеры

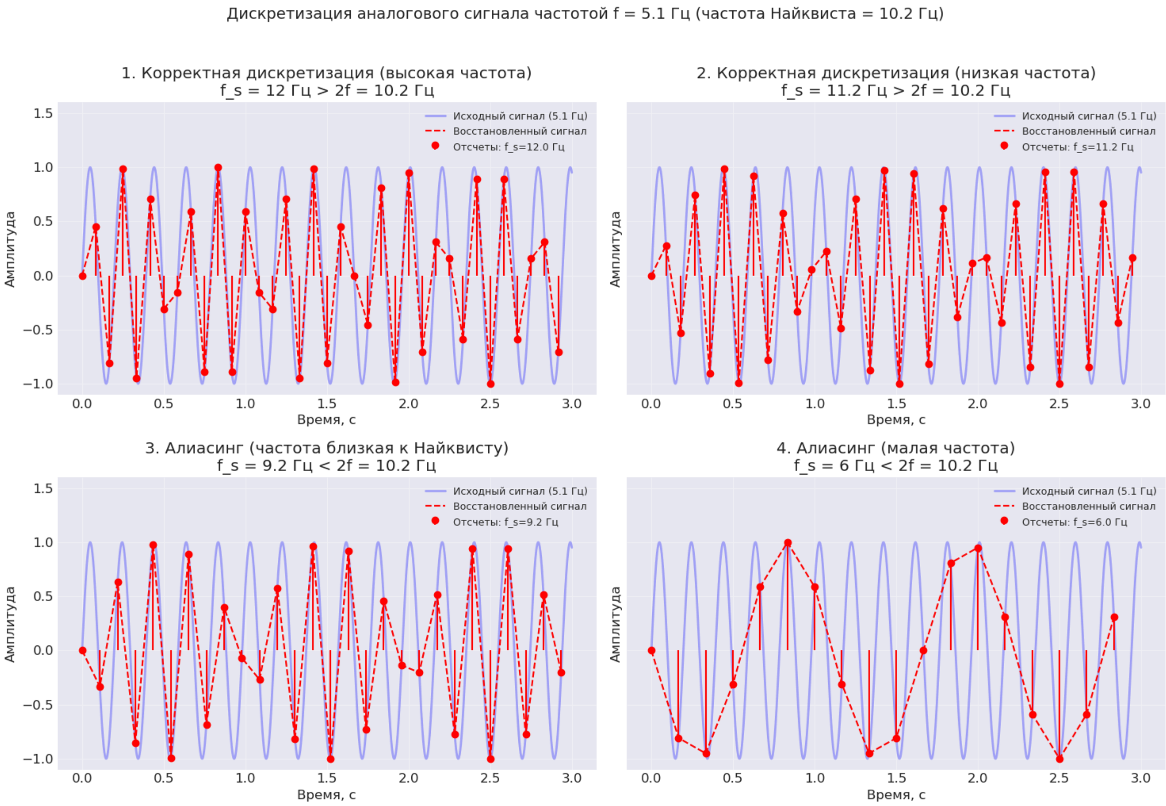
"Эффект колеса" (стробоскопический эффект) в кино

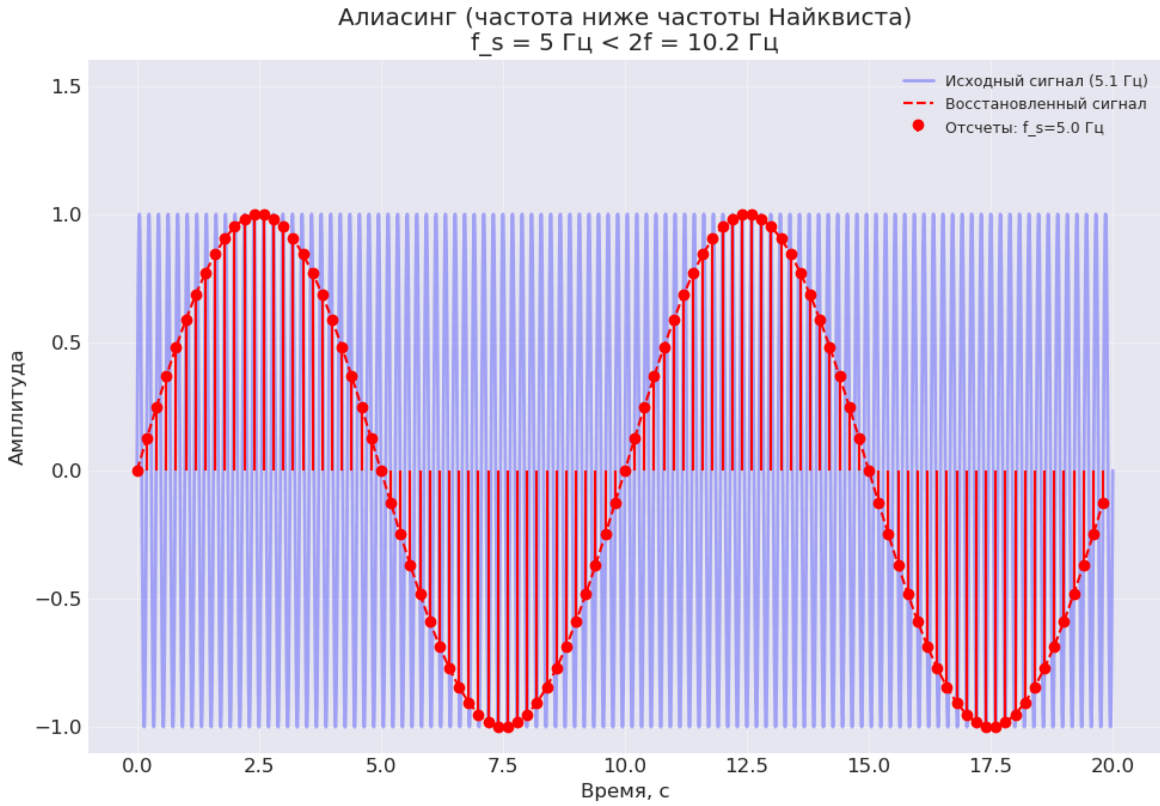
Классический пример: колесо или винт в кино вращается назад, хотя на самом деле вращается вперед. Частота кадров (f_s) недостаточна для корректного представления частоты вращения колеса (f).

Муар на изображениях

Искажения в компьютерной графике







Важно

Алиасинг это не просто искажение высоких частот. Он приводит к "ложным" низким частотам.

6. Антиалиасинговые фильтры

На практике перед АЦП всегда ставят **антиалиасинговый фильтр (фильтр нижних частот)**, который:

1. Пропускает частоты ниже $f_{\max}$
2. Подавляет частоты выше $f_{\max}$ (чтобы они не создавали алиасинг)

Схема реального АЦП:

[Аналоговый сигнал] → [Антиалиасинговый фильтр] → [АЦП] → [Цифровой сигнал]

7. Практические рекомендации

1. **Выбор f_s:** Учитывайте максимальную частоту в сигнале и добавляйте запас 10-20%.

2. **Примеры:**
   - Аудио CD: $f_{\max}$ = 20 кГц, f_s = 44.1 кГц
   - Телефонная связь: $f_{\max}$ = 3.4 кГц, f_s = 8 кГц
   - Видео: частота кадров 60 Гц позволяет передавать частоты колебаний и вращений до 30 Гц

3. **Антиалиасинговый фильтр:** Перед дискретизацией накладывайте на сигнал фильтр нижних частот с $f < f_s/2$.

4. **Проверка на алиасинг:** Визуализируйте спектр сигнала до и после дискретизации.

8. Домашнее задание

Задача: Музыкальный фрагмент содержит частоты до 22 кГц. Для его оцифровки доступны АЦП со следующими частотами дискретизации: 32 кГц, 44.1 кГц, 48 кГц, 96 кГц.

1. Какие из этих частот теоретически пригодны для корректной оцифровки?
2. Какую частоту вы выберете и почему?
3. Рассчитайте, сколько мегабайт займет 3-минутная стерео запись (16 бит на отсчет) при выбранной вами частоте дискретизации.