Современные методы распознавания и синтеза речи
Лекция 1

Лекция 1 Дискретные сигналы

§1 Введение в курс

Опр. Сигнал — формальное описание феномена развивающегося в пространстве и времени.

Обработка сигнала — любая операция, меняющая, анализирующая или иным образом взаимодействующая с сигналом

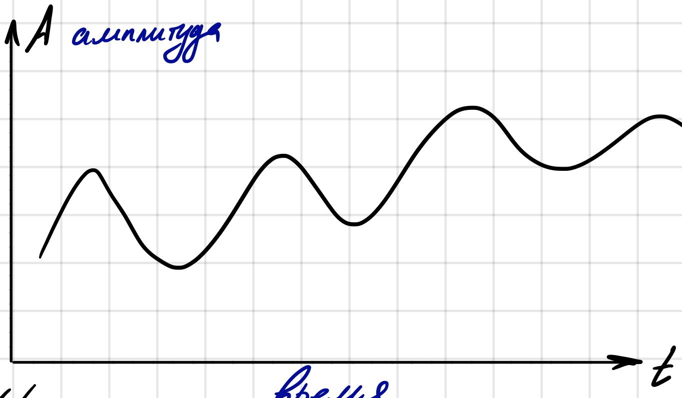
E.g. - померили температуру. Нарисовали график / Посчитали среднее.

- записали голос. Провели идентификацию

Цифровой (lat. digitus, integer) — можно представить целым числом.

Что дискретно?

- время
- амплитуда



Th (Найквист-Шеннон / Котельников)

NB: на Wiki есть история по Котельникова, Найквиста и Шеннона

Пусть максимальная частота в спектре сигнала — w_0. Тогда он полностью описывается своими значениями, следующими на удалении $1/2w_0$ сек. друг за другом.

=> однозначное соответствие между сигналами с непр. временем и дискретным.

Квантование - разбиение значений непрерывной или дискретной величины на конечное число интервалов.

E.g. $x \sim U[0,1]$. Можем записать только 4 бита => 16 интервалов: $[\frac{i}{16}, \frac{i+1}{16})$

неравномерное: коды Хэмминга.

В курсе: дискретное время непрерывная амплитуда (кроме 4^й лекции)

§ 2 Базовые операции с сигналами

Потеряем на некоторое время связь с реальностью...

Опр. Сигнал — последовательность комплексных чисел:

$x[n] \in \mathbb{C}, \quad n \in \mathbb{Z}$

"время", безразмерное

Примеры:

1. Импульс : $\delta[n] = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases}$

   Будет нужен для анализа фильтров. Будем смотреть как фильтр реагирует на такой сигнал. Например, фильтр — помещение. Как помещение разносит эхо?

2. Ступенька: $u[n] = \begin{cases} 1, & n \geq 0 \\ 0, & n < 0 \end{cases}$

   Пример: резко выключили ток в системе. Изменение тока $\Rightarrow$ переменное магнитное поле $\Rightarrow$ пер. электр. поле $\Rightarrow$ как себя будет вести система?

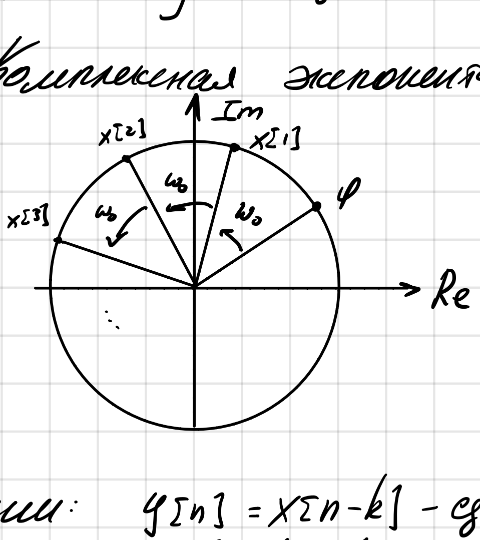
3. Экспоненциальное затухание (exponential decay)

   $x[n] = a^n u[n] \quad a \in \mathbb{C}, \ |a| < 1$

   Закон радиоактивного распада (decay): $\frac{dN}{dt} = -\lambda N$, скорость распада проп-а количеству вещества. $N(t) = N_0 e^{-\lambda t}$

4. Комплексная экспонента: $x[n] = e^{j(\omega_0 n + \varphi)} = \cos(\omega_0 n + \varphi) + j \sin(\omega_0 n + \varphi)$

   у инженеров i – ток



Время дискретно $\Rightarrow$ частота тоже

безразмерно

радианы $[0, 2\pi)$

Операции:

$y[n] = x[n-k]$ – сдвиг

$y[n] = \alpha x[n]$ – масштабирование

$y[n] = x[n] w[n]$ – произв-е

$y[n] = \sum_{k=-\infty}^{n} x[k]$ – интегр-е

$y[n] = x[n] - x[n-1]$ – диф-е

$x[n] = \sum_{k=-\infty}^{+\infty} x[k]\,\delta[n-k]$ — воспроизводящая формула

напряжение

Энергия: $E_x = \|x\|_2^2 = \sum_{n=-\infty}^{+\infty} |x[n]|^2$ ← энергия тока через резистор 1Ω.

$E = UI = \frac{U^2}{R}$ для периодических сигналов м.б. $+\infty$

Мощность: $P_x = \lim_{N\to\infty} \frac{1}{2N} \sum_{n=-N}^{N} |x[n]|^2$

Сигналы

Конечные | Периодические $\xrightarrow{\text{частный случай}}$ Апериодические

$x[n],\ n=\overline{0,N-1}$ $\xrightarrow{x[n]=x[n \% N]}$ $x[n] = x[n+kN]\ \forall k\in\mathbb{Z}$ | $x[n]$

дополнить нулями

§4 Сигналы как объекты гильбертова пространства

Опр. Гильбертово пространство V — лин. векторное пр-во с введенным на нем скалярным произведением (допустима ∞ разм.)

- $\|x\|^2 = \langle x,x\rangle$
- $\rho(x,y) = \|x-y\|$

Опр. Ортонормированный базис: $\{b_k\}_{k=1}^{\infty} \in V$:

$$\forall g\in V\ \exists \{\alpha_k\}_{k=1}^{\infty}:\ g = \sum_{n=1}^{K} \alpha_n b_n$$

$$\langle b_i, b_j\rangle = \begin{cases} 0, & i\neq j \\ 1, & i=j \end{cases}$$

Ряд Фурье: $\sum_{k=1}^{\infty} \underbrace{\langle g, b_k\rangle}_{= g_k} b_k \overset{?}{=} g$ коэф. Фурье

Тождество Бесселя

лучшее приближение

$$\Big\|\sum_{k=1}^{n} c_k b_k - g\Big\|^2 = \sum_{k=1}^{n} c_k^2 - 2\sum c_k \langle g, b_k\rangle + \|g\|^2 =$$

$$= \sum_{k=1}^{n} (c_k - \langle g_k, b_k\rangle)^2 + \|g\|^2 - \sum_{k=1}^{n} \langle g, b_k\rangle^2 = \{ c_k = \langle g_k, b_k\rangle = g_k \} = \|g\|^2 - \sum_{k=1}^{n} g_k^2$$

$\Rightarrow \sum_{k=1}^{n} g_k^2 \leq \|f\|^2 \quad n \to \infty$

$\sum_{k=1}^{\infty} g_k^2 \leq \|f\|^2$ ← неравенство Бесселя ($\Rightarrow$ ряд Фурье сходится)

Опр $\{b_k\}$ – замкнутая ортонормированная система (ОНС), если

$\forall \varepsilon > 0 \ \forall g \in V \ \exists N \in \mathbb{N}, \ \exists c_1 \dots c_N : \|\sum_{k=1}^{n} c_k b_k - f\| < \varepsilon$

Th (тождество Парсеваля)

В замкнутой ОНС $\|g\|^2 = \sum_{k=1}^{\infty} g_k^2$

Proof $\forall \varepsilon > 0 \ \exists n_0 : \|g - \sum_{k=1}^{n_0} c_k b_k\|^2 < \varepsilon^2$ но min. в фикс. базисе при $c_k = g_k$

$0 \leq \|g\|^2 - \sum_{k=1}^{n_0} g_k^2 \leq \|g - \sum_{k=1}^{n_0} c_k b_k\|^2 \leq \varepsilon^2 \quad \varepsilon \to 0 \quad \Rightarrow$ ч.т.д.

1. Интерпретация: сохраняется энергия при переходе между базисами

2. Теперь ряд Фурье $\sum g_k b_k$ имеет смысл

Теор. $\{b_k\}$ – замк. ОНС $\Rightarrow \|\sum_{k=1}^{n} g_k b_k - g\|^2 \xrightarrow{n \to \infty} 0$

Д-во $\|\sum_{k=1}^{n} g_k b_k - g\|^2 = \|g\|^2 - \sum_{k=1}^{n} g_k^2 \xrightarrow[n \to \infty]{} 0$

Примеры пространств

	Простр-во	Скал. пр-е
Конечные сигналы	$\mathbb{C}^N$	$\langle x, y \rangle = \sum_{n=0}^{N-1} x[n]^* y[n]$
Периодические сигналы	$\tilde{\mathbb{C}}^N$	$\langle x, y \rangle = \sum_{n=0}^{N-1} x[n]^* y[n]$
Апериодические сигналы	$l_2 \ (\sum \lvert x[i] \rvert^2 < \infty)$	$\langle x, y \rangle = \sum_{n=-\infty}^{+\infty} x[n]^* y[n]$

зачем нам сигналы ∞ длины? Конечные приближают бесконечные!

Примеры базисов сигналов ($\mathbb{C}^N$)

$\delta_n^{(k)} = \delta[n-k] = \begin{cases} 1, & n = k \\ 0, & \text{иначе} \end{cases}$ – канонический

$w_n^{(k)} = \frac{1}{\sqrt{N}} e^{-j \frac{2\pi}{N} nk}$ – базис Фурье

Матрица перехода: $M_{nk} = \left(\frac{1}{\sqrt{N}} e^{-j \frac{2\pi}{N} nk}\right)^* = \frac{1}{\sqrt{N}} e^{j \frac{2\pi}{N} nk}$

$\alpha = Mx$

Упр $w_n^{(k)}$ – ортонорм. базис

Д-во $\langle w^{(t)}, w^{(k)} \rangle = \frac{1}{N}\sum_{n=0}^{N-1} e^{j\frac{2\pi}{N}kn} e^{-j\frac{2\pi}{N}tn} = \frac{1}{N}\sum_{n=0}^{N-1} \underbrace{e^{j\frac{2\pi}{N}(k-t)n}}_{=1,\ k=t \ \Rightarrow \frac{1}{N}\Sigma 1 = 1} = \{k \neq t\}$

$= \frac{1}{N}\frac{1-e^{j\frac{2\pi}{N}(k-t)N}}{1-e^{j\frac{2\pi}{N}(k-t)}} = \{ e^{j\cdot 2\pi k} = \cos(2\pi k) + j\sin(2\pi k) = 1 \} = 0$

Преобразование Фурье — матричное произведение! Можем использовать в нейросетях: $y = Mx \rightarrow \frac{\partial L}{\partial x} = M^{H}\frac{\partial L}{\partial y}$

обратное преобразование Фурье! Forward и Backward вычисляются быстро!

Как выглядит эта матрица? $\underline{N=4}$

$$M = \frac{1}{2}\begin{bmatrix} i^0 & i^0 & i^0 & i^0 \\ i^0 & i^1 & i^2 & i^3 \\ i^0 & i^2 & i^4 & i^6 \\ i^0 & i^3 & i^6 & i^9 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix} \qquad e^{j\frac{2\pi}{N}nk}$$

Наивно: $O(N^2)$. Можно $O(N\log N)$! (см. лекцию 2.)

§5 Преобразование Фурье (DFT, DFS, DTFT)

Почему синусоиды?
- звуковые волны
- уравнение колебания маятника
- уравнения Максвелла
- Linear Time-invariant systems: (LTI)
  $\sin(\omega t) \rightarrow A\sin(\omega t + \varphi)$

Опр. LTI

- $f(ax[n] + by[n]) = af(x[n]) + bf(y[n])$
- $f(x[n-T]) = f(x)[n-T]$

Формула анализа: $\alpha_k = \langle f, b_k \rangle$

Формула синтеза: $f = \sum \alpha_k b_k$

Экспонента: $Ae^{j(\omega n + \varphi)} = A\cos(\omega n + \varphi) + jA\sin(\omega n + \varphi)$

нач. фаза

амплитуда частота

Когда периодичная функция?

$\rightarrow \omega = 2\pi(M/N)$

Конечные сигналы:

$w_k[n] = e^{j\omega_k n}$. Хотим полное число периодов в 0..N-1:

$w_k[0] = w_k[N] \Leftrightarrow \left(e^{j\omega_k}\right)^N = 1 \Rightarrow \omega_k = \frac{2\pi}{N}k \quad k = 0..N-1$

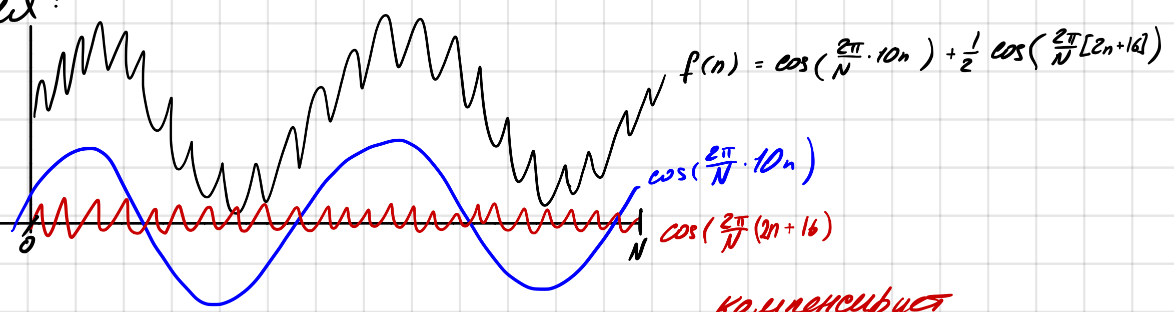
$w_k[n] = e^{j\frac{2\pi}{N}k}$ $\qquad \|w_k\|^2 = N$, но будем использовать нормировку только в синтезе

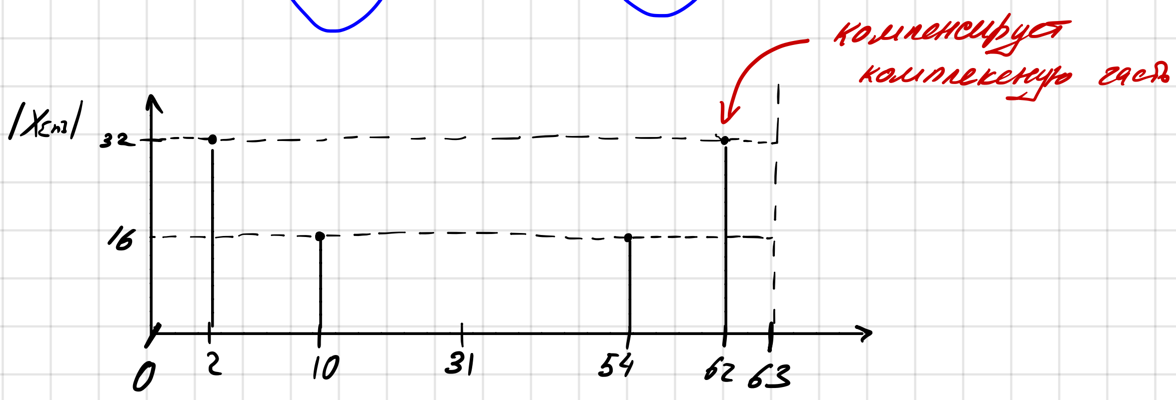
Сигнал	Преобр.	Анализ	Синтез
Конечный	DFT (discrete Fourier transform)	$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi}{N}kn}$	$x[n] = \frac{1}{N}\sum_{k=0}^{N-1} X[k] e^{j\frac{2\pi}{N}nk}$
Периодический	DFS (discrete Fourier series)	$\tilde{X}[k] = \sum_{n=0}^{N-1} \tilde{x}[n] e^{-j\frac{2\pi}{N}kn}$ $k \in \mathbb{Z}$	$\tilde{x}[n] = \frac{1}{N}\sum_{k=0}^{N-1} \tilde{X}[k] e^{j\frac{2\pi}{N}kn}$ $n \in \mathbb{Z}$
Апериодический, с конечной энергией	DTFT (discrete-time Fourier transform)	$X(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} x[n] e^{-j\omega n}$ показывает периодичность $X(\cdot)$	$x[n] = \frac{1}{2\pi}\int_0^{2\pi} X(e^{j\omega}) e^{j\omega n} d\omega$

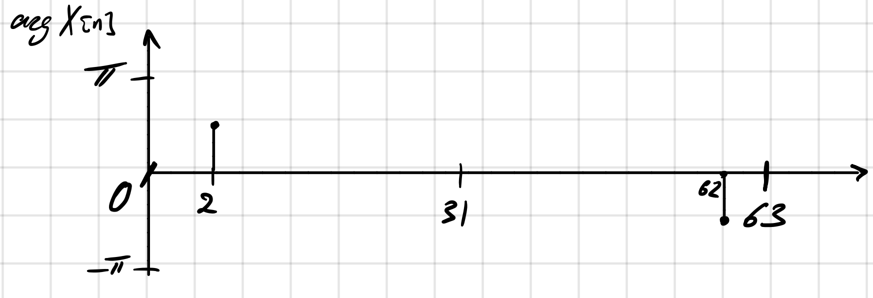
Почему 3 преобр-я? Бесконечное эхо ⇒ беск. сигнал (а на входе был конечный). Продолжение: период. или нули.

Интерпретация:

$N = 64$







Существование и сходимость DTFT:

$$X(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} x[n]e^{-j\omega} = \lim_{M\to\infty} \sum_{n=-M}^{M} x[n]e^{-j\omega} = \lim_{M\to\infty} X_M(e^{j\omega})$$

$$|X(e^{j\omega})| \le \sum_{n=-\infty}^{+\infty} |x[n]||e^{-j\omega}| = \sum_{n=-\infty}^{+\infty} |x[n]|$$ $\Rightarrow$ достаточно абс. суммируемости.

При абс. суммируемости:

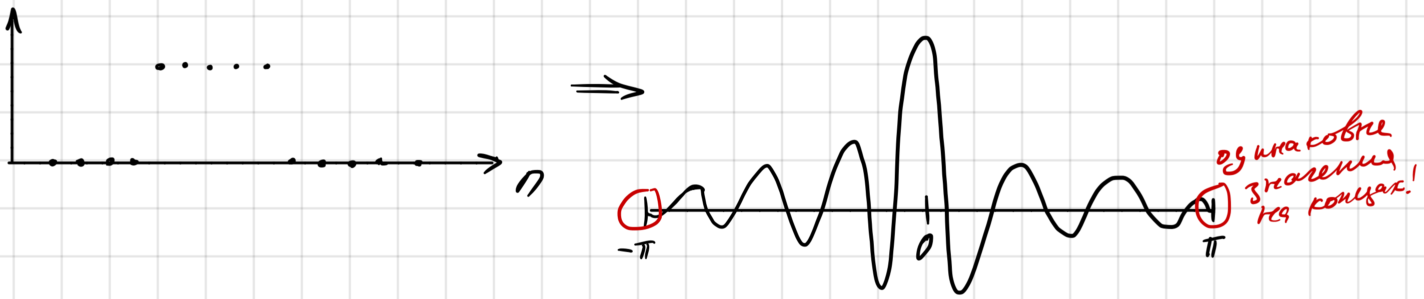
- $X(e^{j\omega}) \overset{[0,2\pi]}{\rightrightarrows} X_M(e^{j\omega})$, т.е. $\lim\limits_{M\to\infty} \sup\limits_{\omega} |X(e^{j\omega}) - X_M(e^{j\omega})| = 0$
- $X(e^{j\omega})$ — непр. на $\mathbb{R}$ ($\Rightarrow$ можно считать в отдельных точках и интерполировать)

Более слабое требование: $x[n] \in l_2$, т.е. $\sum\limits_{n=-\infty}^{+\infty} x[n]^2 < \infty$. Тогда:

- $X(e^{j\omega}) \xrightarrow{\text{ср. кв.}} X_M(e^{j\omega})$, т.е. $\lim\limits_{M\to\infty} \int_0^{2\pi} |X(e^{j\omega}) - X_M(e^{j\omega})|^2 d\omega = 0$ ← среднекв.
- $X(e^{j\omega})$ мб разрывной

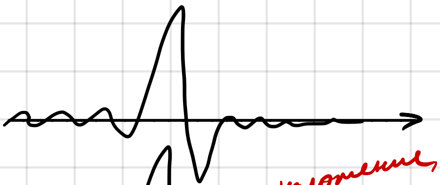
Пример:
$$x[n] = \begin{cases} 1, & -N \le n \le N \\ 0, & \text{иначе} \end{cases}$$

(что происходит при дополнении нулями)



Связь между DTFT и DFS

Рассмотрим такой сигнал (апериодический, бесконечный).



Абс. суммируется $\Rightarrow$ рано или поздно затухает.
Повторим сигнал с задержкой N:

$$\tilde{x}[n] = \sum_{i=-\infty}^{+\infty} x[n+iN] \qquad \lim_{N\to\infty} \tilde{x}[n] = x[n]$$

$$\underbrace{e^{-j\frac{2\pi}{N}k(n+iN)}}_{=}$$

DFS от $\tilde{x}[n]$: $\tilde{X}[k] = \sum_{n=0}^{N-1} \tilde{x}[n] e^{-j\frac{2\pi}{N}kn} = \sum_{i=-\infty}^{+\infty}\left(\sum_{n=0}^{N-1} x[\underbrace{n+iN}_{\text{пробегает все значения}}] e^{-j\frac{2\pi}{N}kn}\right) =$

$$= \sum_{i=-\infty}^{+\infty} x[m] e^{-j\frac{2\pi}{N}mk} = X(e^{j\omega})\Big|_{\omega=\frac{2\pi}{N}k}$$

$\Rightarrow$ совпадает в "целых" точках. Чем больше точек, тем ближе к DTFT.

$$\tilde{x}[n] = \frac{1}{N}\sum_{n=0}^{N-1} X(e^{j\frac{2\pi}{N}k}) e^{j\frac{2\pi}{N}k} = \left\{\Delta = \frac{2\pi}{N}\right\} = \frac{1}{2\pi}\sum_{n=0}^{N-1} X(e^{j(k\Delta)}) e^{j(k\Delta)n}\Delta$$

Получили Римановскую сумму $f(\omega) = X(e^{j\omega})e^{j\omega}$

При $N\to\infty$: $\tilde{x}[n] \to x[n]$ и $\tilde{x}[n] \longrightarrow \frac{1}{2\pi}\int_0^{2\pi} X(e^{j\omega}) e^{j\omega} d\omega$

Свойства преобразования

$$x[n] \longleftrightarrow X(e^{-j\omega})$$

- Симметрии: $x[-n] \longleftrightarrow X(e^{-j\omega})$

$$x^*[n] \longleftrightarrow X^*(e^{-j\omega})$$

- Сдвиг: $x[n-n_0] \longleftrightarrow e^{-j\omega n_0} X(e^{j\omega})$

$$e^{j\omega_0 n} x[n] \longleftrightarrow X(e^{j(\omega-\omega_0)})$$

- Тождество Парсеваля:

$$\sum_{n=-\infty}^{+\infty} |x[n]|^2 = \frac{1}{2\pi}\int_0^{2\pi} |X(e^{j\omega})|^2 d\omega$$

Аналогичные св-ва у DFT и DTFT.

§6 Быстрое преобразование Фурье

$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j\frac{2\pi}{N}kn}$, $k = 0, 1, \dots, N-1$ $\quad \underline{O}(N^2)$. Сейчас научи

Вычисления можно сократить за счёт симметричности формулы.

Рассмотрим случай $N = 2^m$ (алгоритм Кули-Тьюки):

Пусть $W_N = e^{-j\frac{2\pi}{N}}$. Тогда $X[k] = \sum_{n=0}^{N-1} x[n] W_N^{kn}$

Разобьем суммирование на четные и нечетные индексы:

$$X[k] = \sum_{n-\text{четно}} x[n] W_N^{nk} + \sum_{n\ \text{неч.}} x[n] W_N^{nk} = \sum_{z=0}^{N/2-1} x[2z] W_N^{(2z)k} + \sum_{z=0}^{N/2-1} x[2z+1] W_N^{(2z+1)k} =$$

$$= \left\{ W_N^2 = e^{-j\frac{2\pi}{N}\cdot 2} = e^{-j\frac{2\pi}{N/2}} = W_{N/2} \right\} = \underbrace{\sum_{z=0}^{N/2-1} x[2z] W_{N/2}^{zk}}_{\substack{\text{DFT от четной части} \\ = E_{N/2}[k]}} + W_N^k \underbrace{\sum_{z=0}^{N/2-1} x[2z+1] W_{N/2}^{zk}}_{\substack{\text{DFT от нечетной части} \\ = O_{N/2}[k]}} =$$

$$= \begin{cases} E_{N/2}[k] + W_N^k O_{N/2}[k], & k < N/2 \\ E_{N/2}[k - N/2] + W_{N/2}^k O_{N/2}[k - N/2], & k \geq N/2 \end{cases}$$

$$\underbrace{\cancel{W_{N/2}^{N/2}}}_{1} \sum_{z=0}^{N/2-1} x[2z] W_{N/2}^{k-N/2}$$

Итоговая сложность: спускаемся рекуррентно на $\log_2 N$ шагов. На каждом суммарно $O(N)$ операций $\Longrightarrow O(N\log N)$

В общем случае можно дополнить нулями до степени двойки. Но так делать не нужно! Получим краевые эффекты! Есть алгоритмы для произвольного N.

<u>Алгоритм</u> FFT(X):

```
if |X| = 1:
    return X
else:
    E = FFT(X[::2])
    O = FFT(X[1::2])
    W_k = [exp(-j 2π/N k)   k = 0..N-1]

    return (E + W_k[:N/2]·O || E + W_k[N/2:]·O)
```