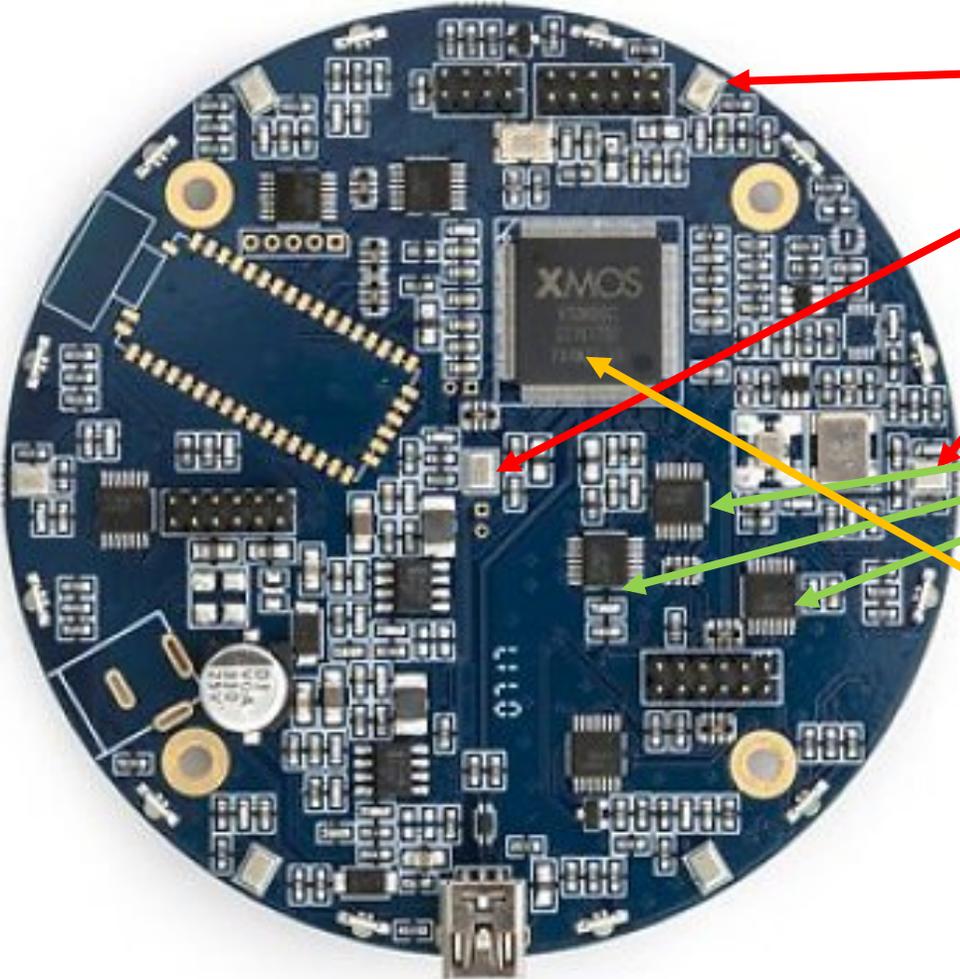


Оцифровка звука

UMA-8



MIC



ADC



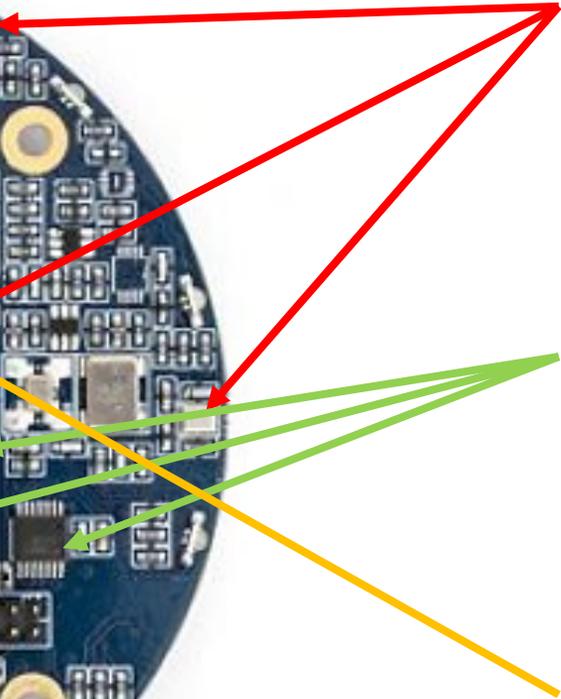
DSP



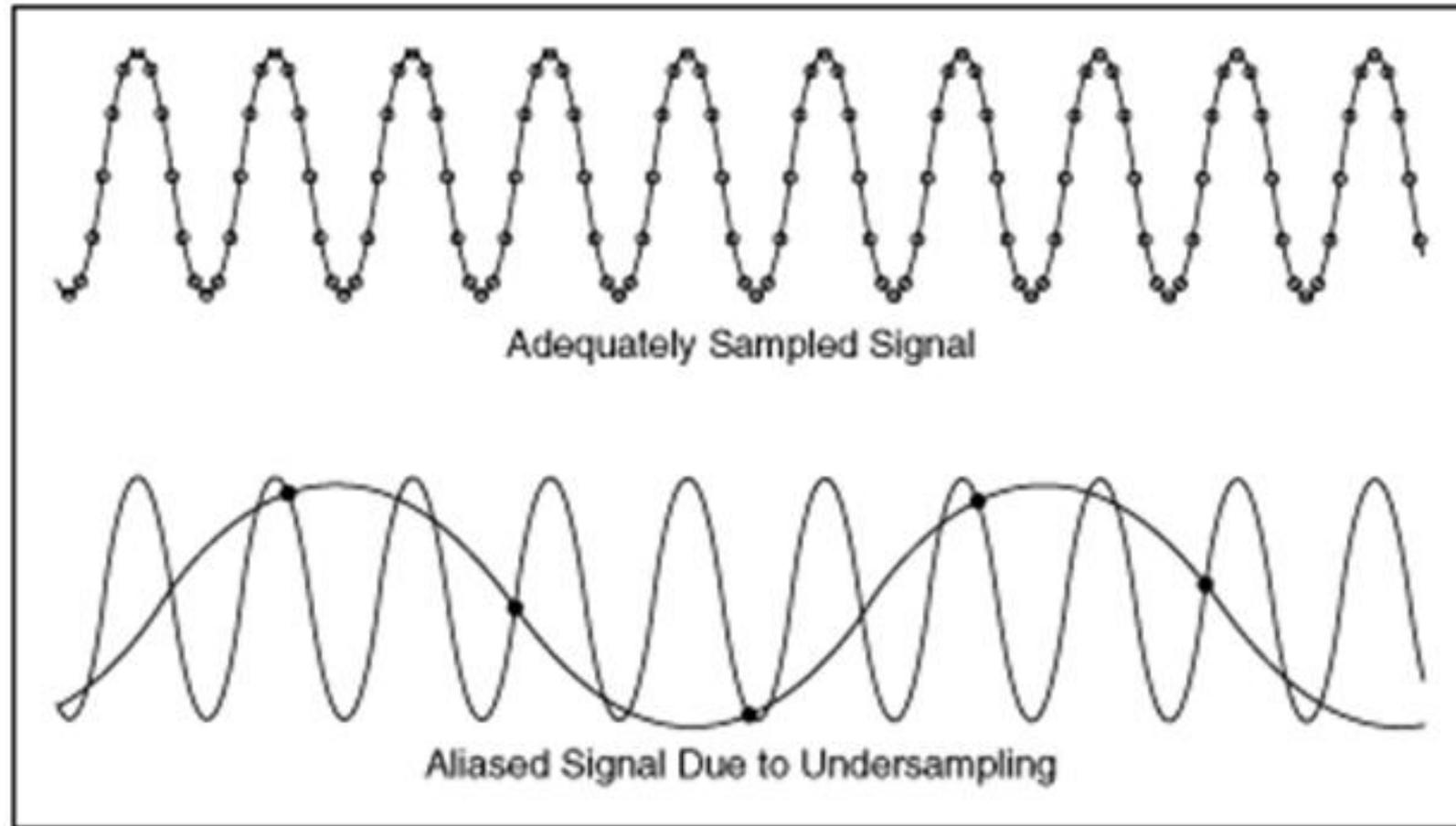
CPU



SRV



Оцифровка сигнала



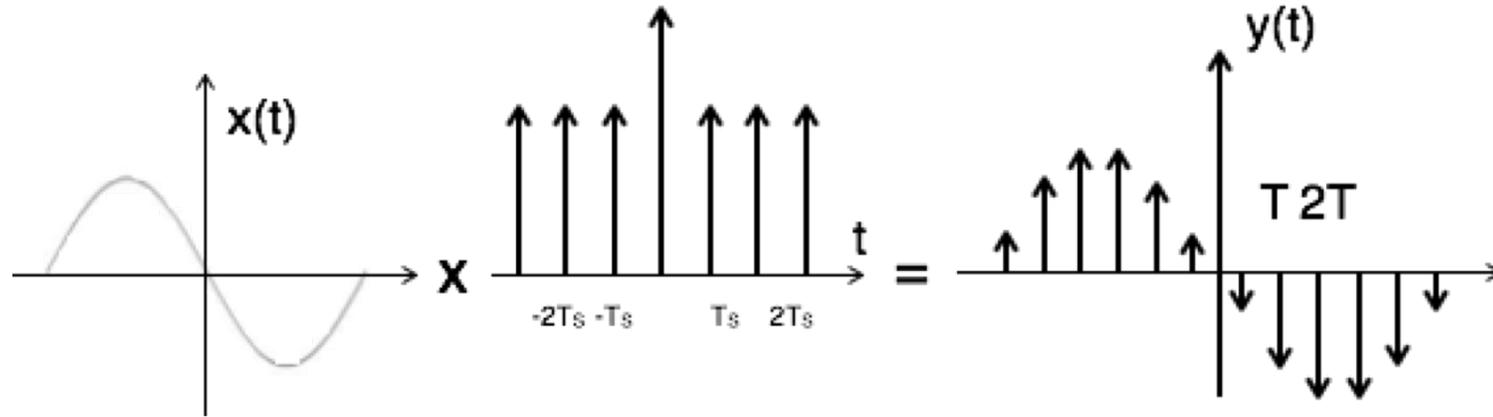
Тоже самое в изображениях



Вопросы

- Что происходит с сигналом при оцифровке
- С какой частотой нужно сэмплировать
- Как восстановить исходный сигнал
- Что делать, если частота сэмплирования недостаточна

Импульсное семплирование

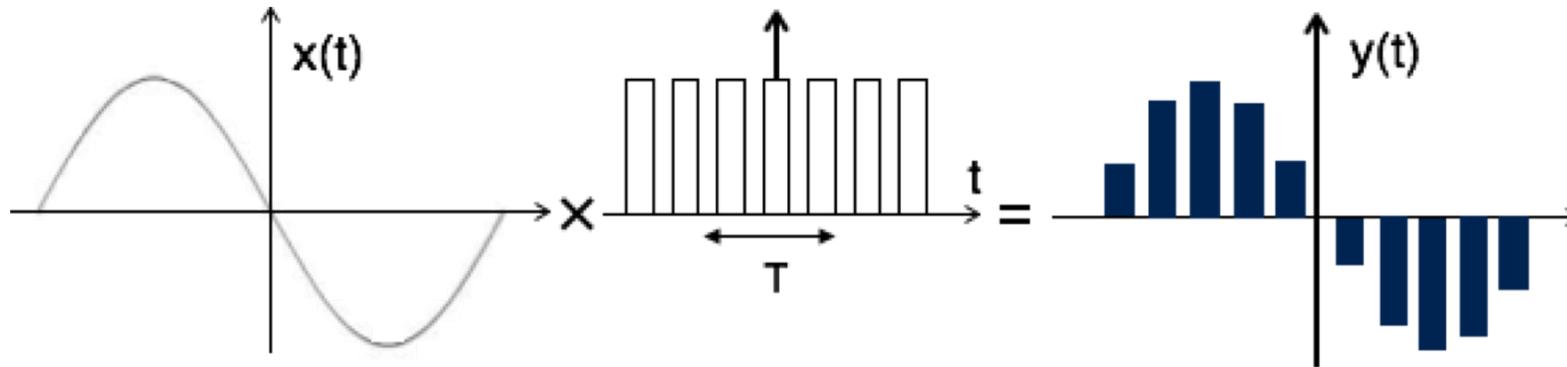


$x(t)$ – исходный сигнал

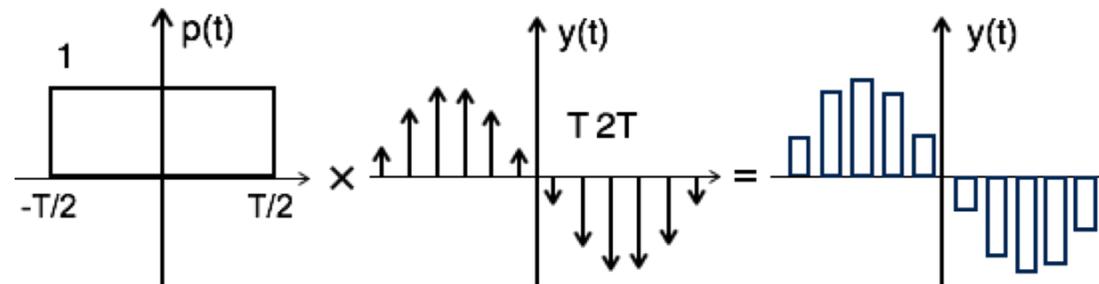
$$p(t) = \sum_{-\infty}^{\infty} \delta(t - nT) \quad \text{– гребень Дирака}$$

$$y(t) = x(t) \times p(t) = \sum_{-\infty}^{\infty} x(nT) \delta(t - nT) = \sum_{-\infty}^{\infty} x[n] \delta(t - nT)$$

Flat-top sampling



Можно заменить импульсным сэмплированием с предварительной фильтрацией.



Импульсное семплирование в Фурье

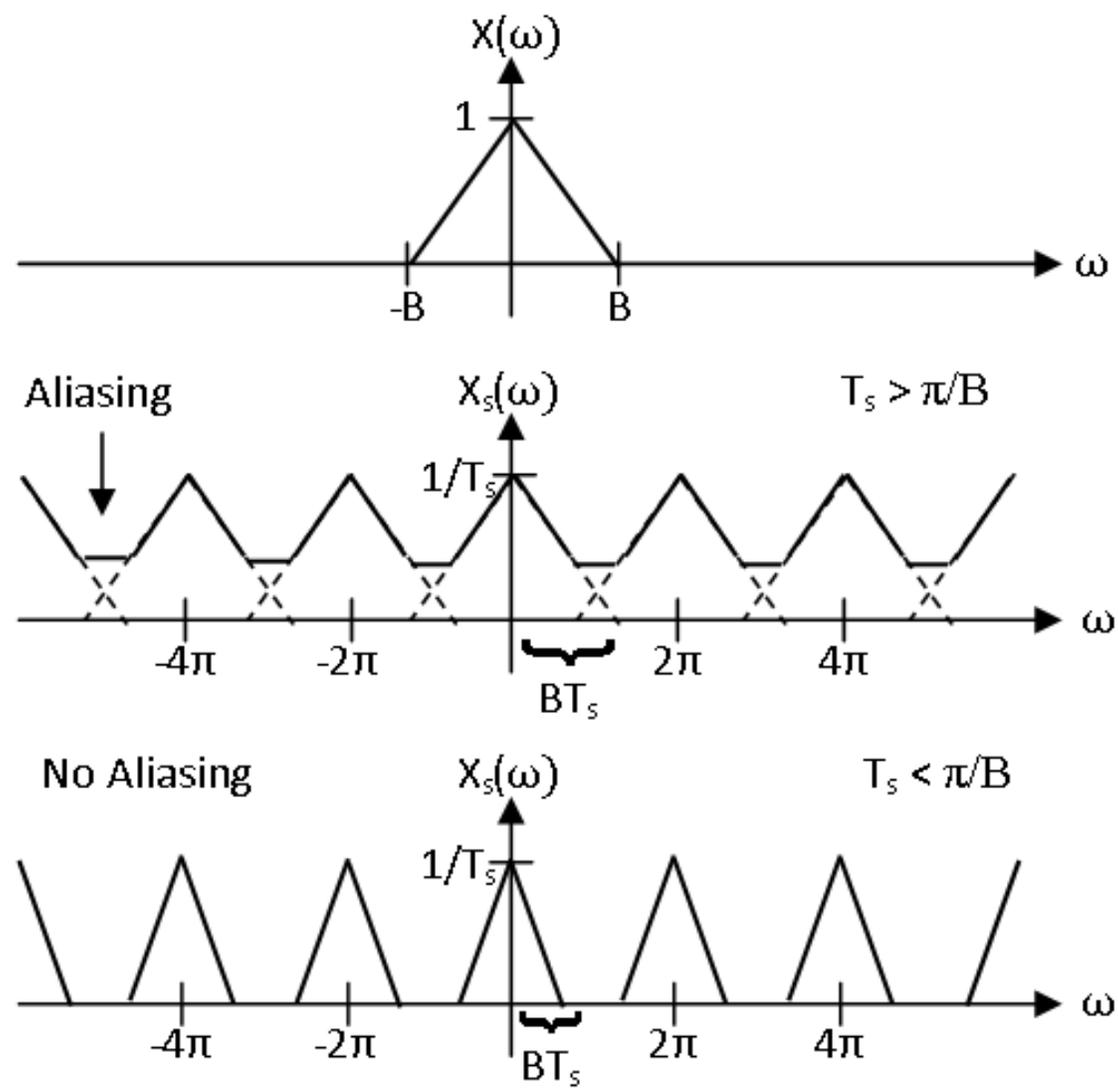
Попробуем понять, что происходит со спектром сигнала при импульсном семплировании?

$$Y(f) = X(f) * P(f)$$

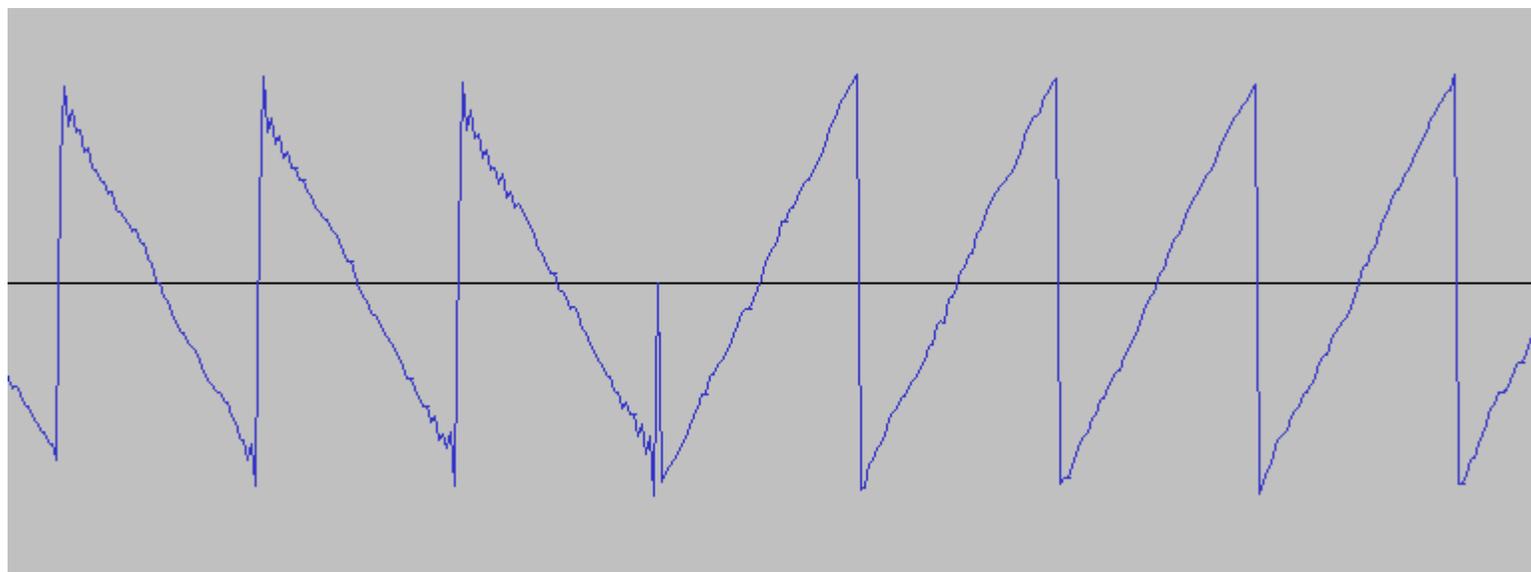
$$\sum_{-\infty}^{\infty} \delta(t - nT) \stackrel{F}{\Leftrightarrow} \frac{1}{T} \sum_{-\infty}^{\infty} \delta\left(t - \frac{n}{T}\right)$$

$$Y(f) = X(f) * \frac{1}{T} \sum_{-\infty}^{\infty} \delta\left(t - \frac{n}{T}\right) = \frac{1}{T} \sum_{-\infty}^{\infty} X\left(t - \frac{n}{T}\right)$$

Алиасинг



Алиасинг



Тоже самое в изображениях



Теорема Котельникова (Nyquist–Shannon)

Теорема II. Любую функцию $F(t)$, состоящую из частот от 0 до f_1 , можно непрерывно передавать с любой точностью при помощи чисел, следующих друг за другом через $1/(2f_1)$ секунд.

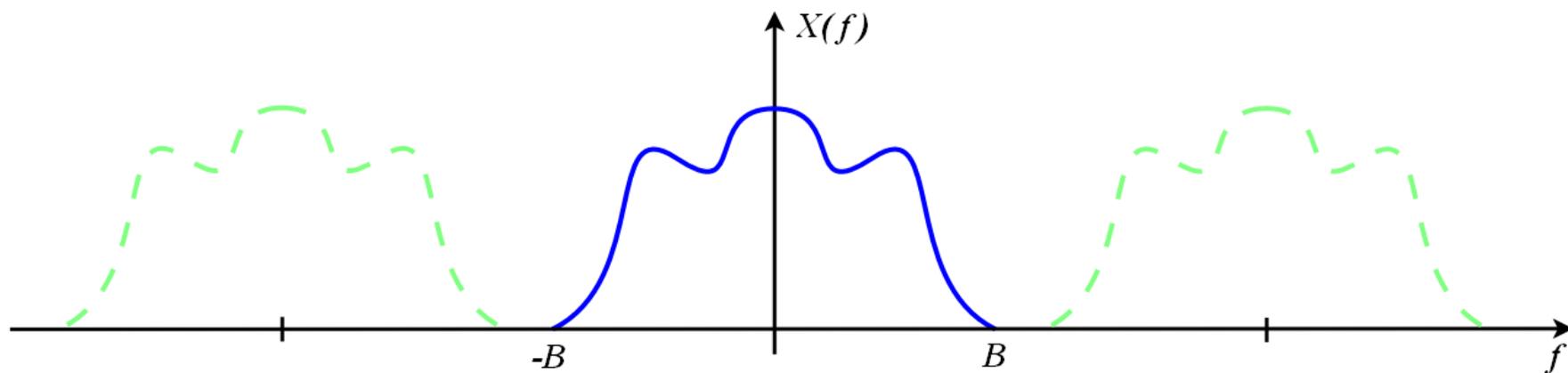
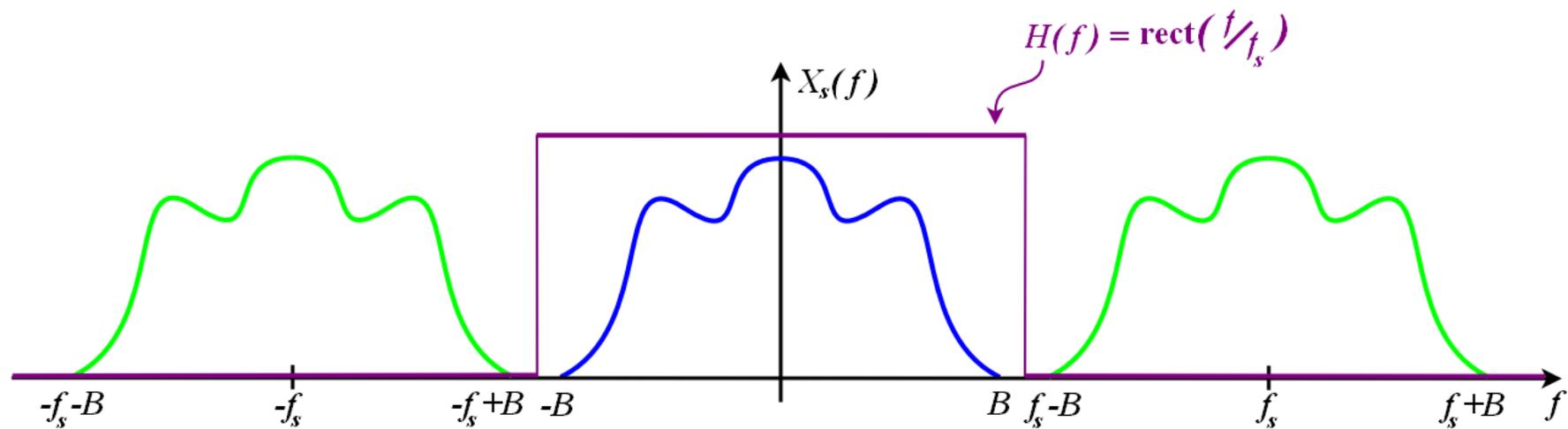
Найквист «Certain topics in telegraph transmission theory» 1928

Котельников «О пропускной способности эфира и проволоки в электросвязи» 1933

Shannon. Communication in the presence of noise. 1949



Реконструкция



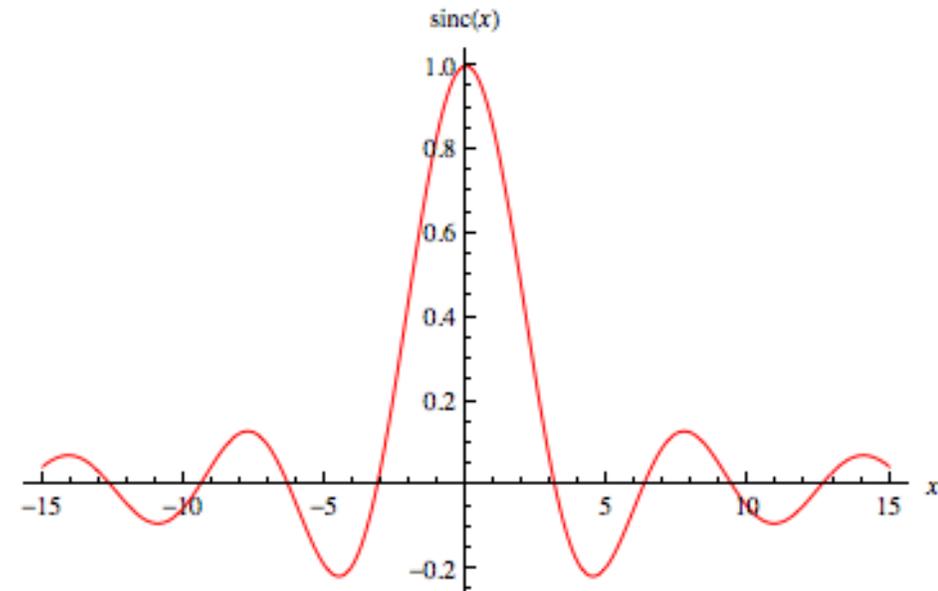
Реконструкция

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

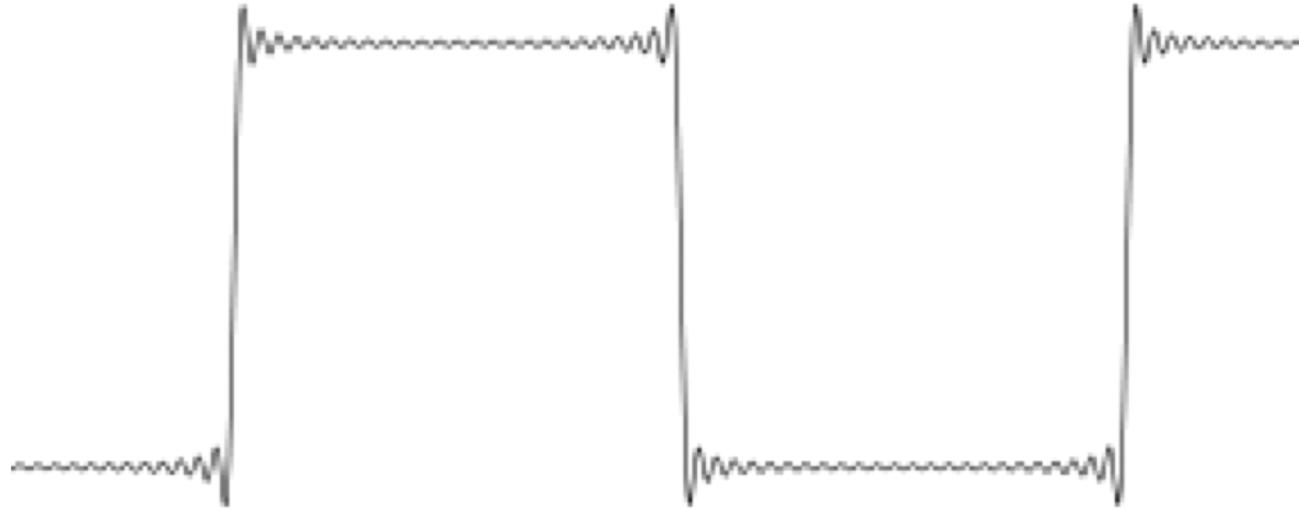
$$H(f) = \begin{cases} T, & -\frac{1}{T} \leq x \leq \frac{1}{T} \\ 0, & \text{else} \end{cases}$$

$$H(f) \stackrel{F}{\Leftrightarrow} T \cdot \text{sinc}\left(\frac{t}{T}\right)$$

$$x(t) = y(t) * h(t) = \sum_{-\infty}^{\infty} y[n] \delta(t - nT) * h(t) = \sum_{-\infty}^{\infty} y[n] h(t - nT)$$



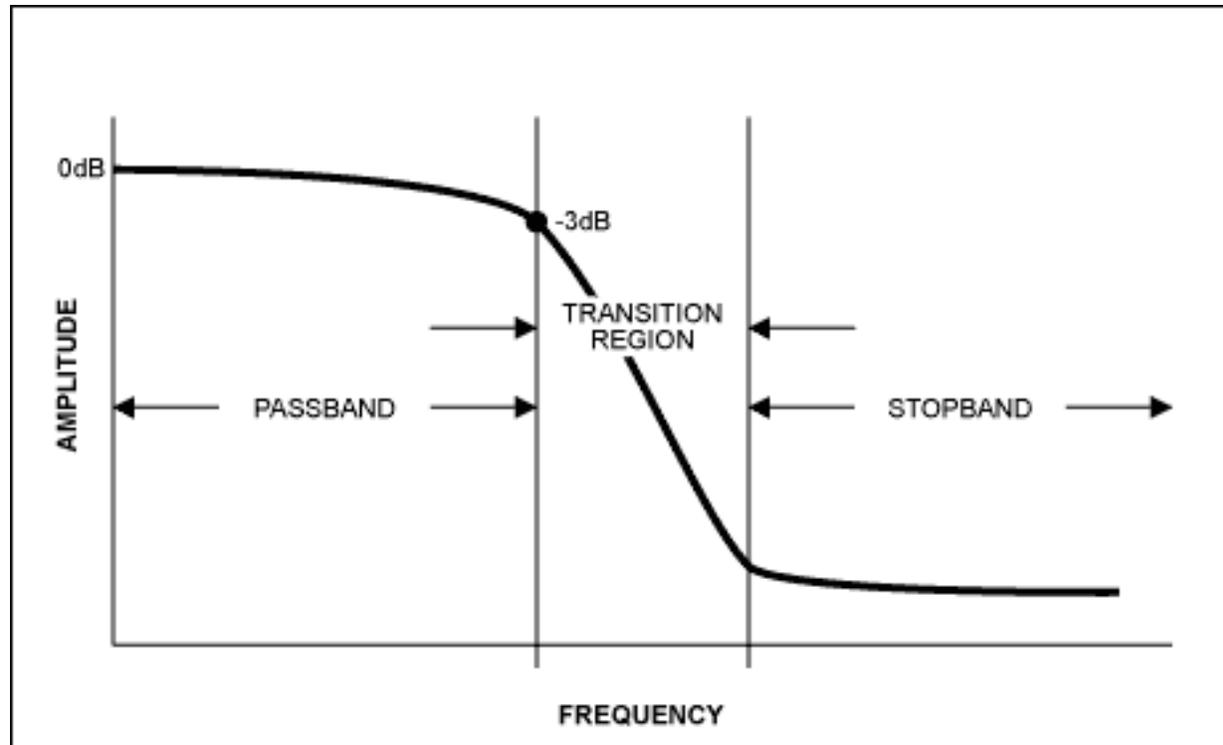
Феномен Гиббса



**Lorem ipsum
turpis vitae v**

44,100 Hz

- Считается, что ухо может слышать 20 kHz
- По теореме Котельникова этого нужно семплирование 40 kHz
- Зачем лишние 4100 семплов?



Отношение сигнал/шум

Бел выражает отношение двух значений громкости звука (субъективно определяемая его энергией) десятичным логарифмом этого отношения.

Децибел = 1/10 Бел

$$D = \lg \left(\frac{V_1}{V_2} \right)$$

- Увеличение величины на 1 дБ означает её увеличение в $10^{0,1} \approx 1,259$ раза
- «громкость звука составляет 30 дБ» означает, что интенсивность звука в 1000 раз превышает порог слышимости

Отношение сигнал/шум

Signal-to-noise ratio (SNR) — безразмерная величина, равная отношению мощности полезного сигнала к мощности шума

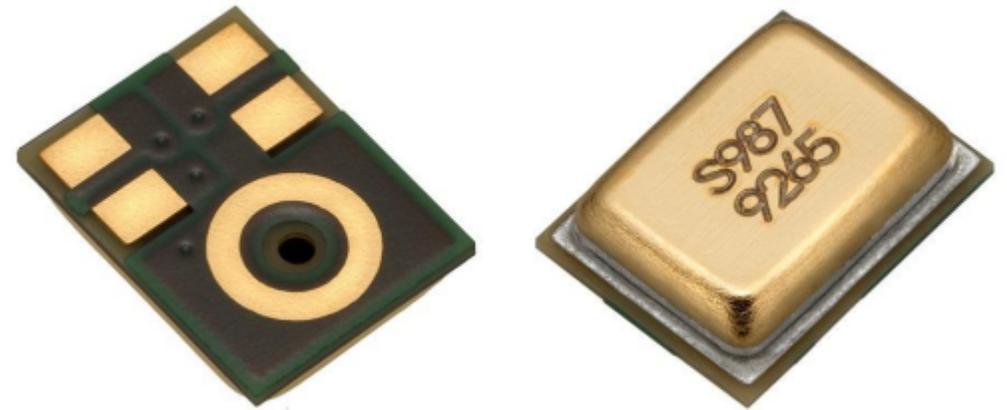
$$\text{SNR(dB)} = 10 \log_{10} \left(\frac{P_{\text{signal}}}{P_{\text{noise}}} \right)$$

- Телефонный разговор: 30 дБ
- Музыка: 75 дБ
- Hi-Fi: 90 дБ

Микрофон

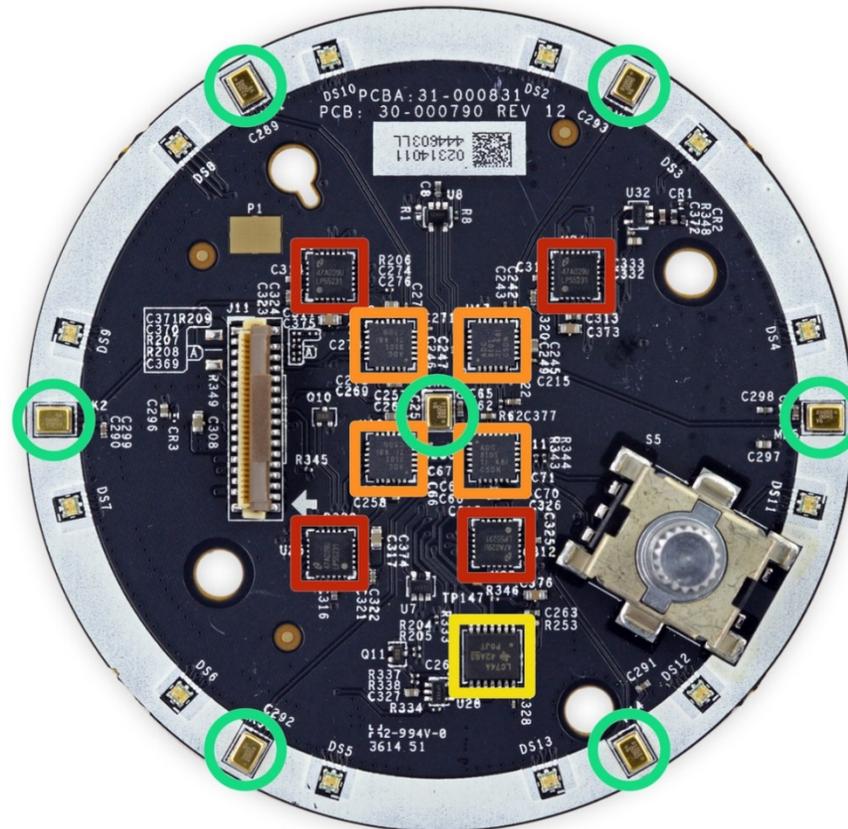
Frequency Range: 100Hz ~ 10kHz

S/N Ratio: 65dB

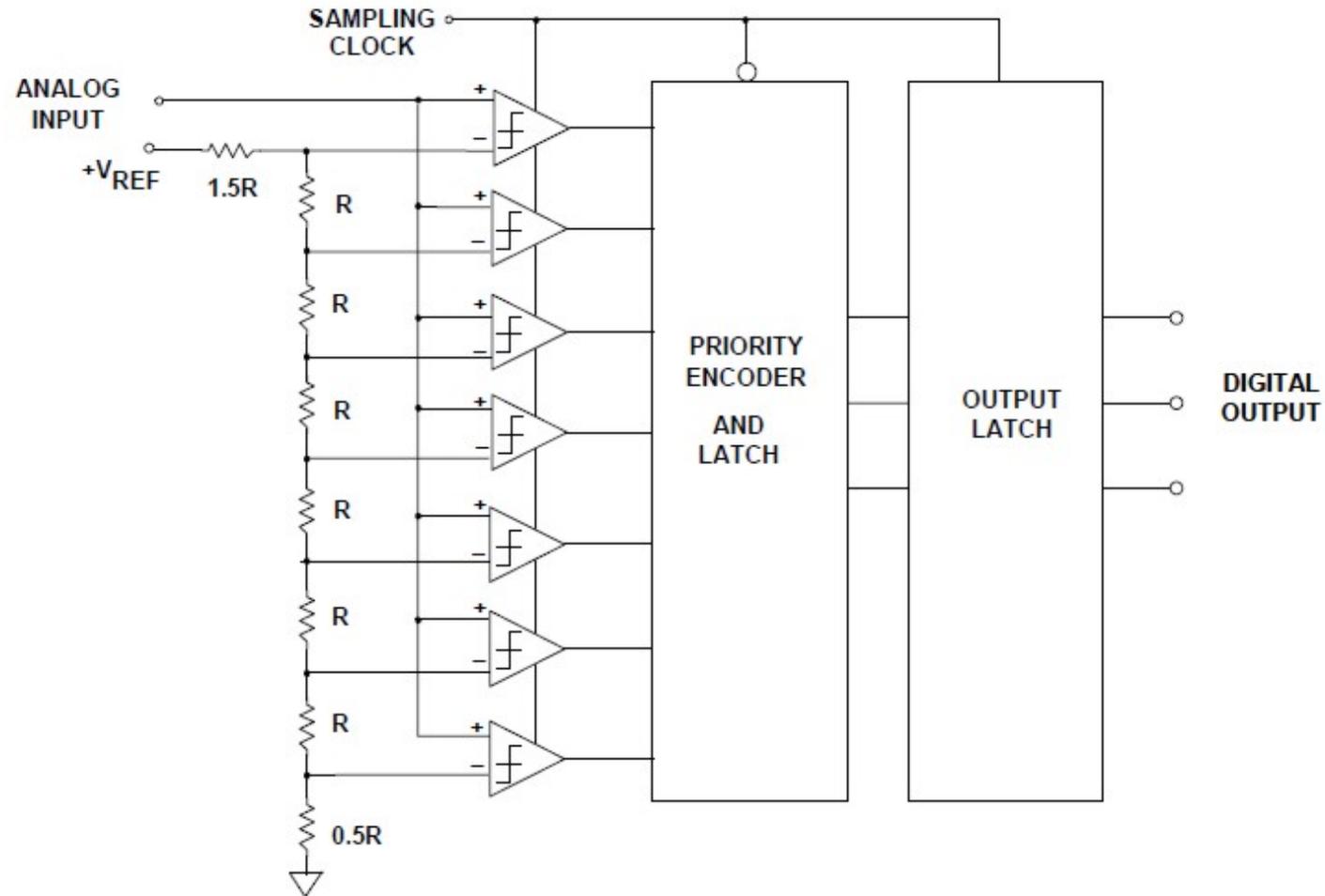


АЦП

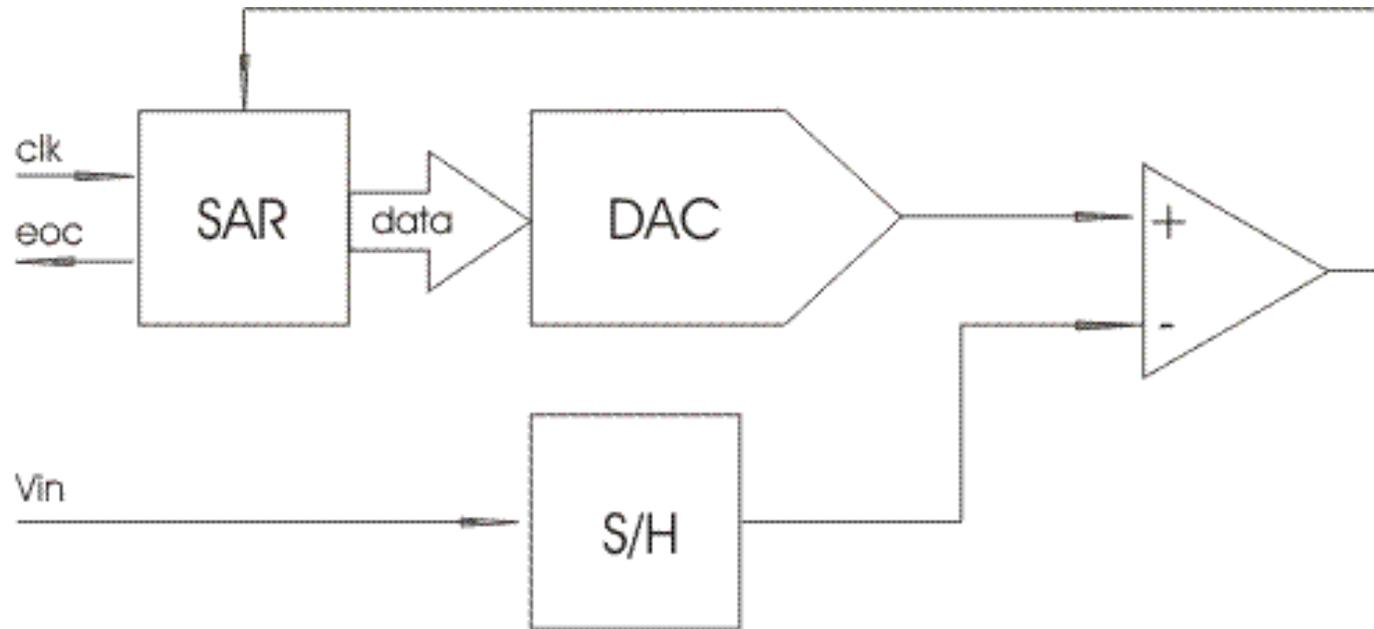
- АЦП – специализированная микросхема для преобразования аналогового сигнала в цифровой вид



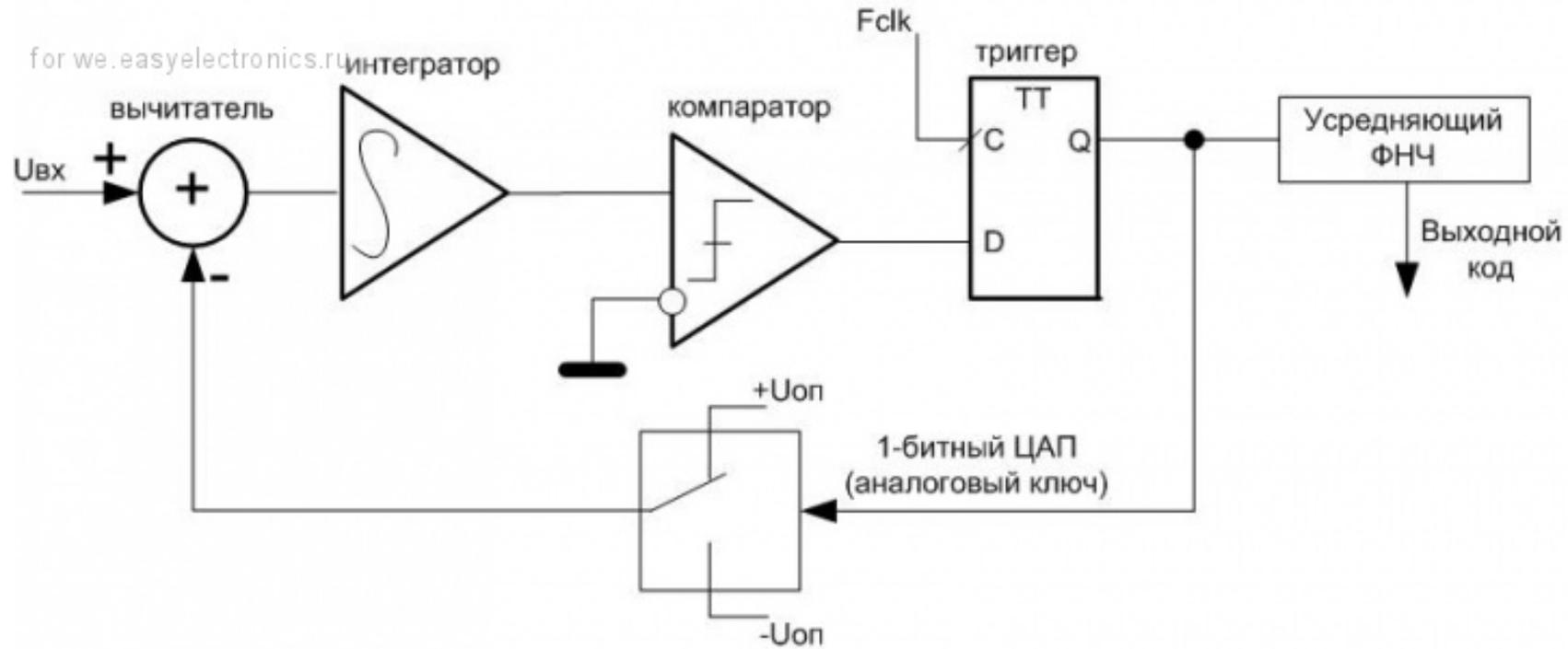
АЦП прямого преобразования



АЦП последовательного приближения

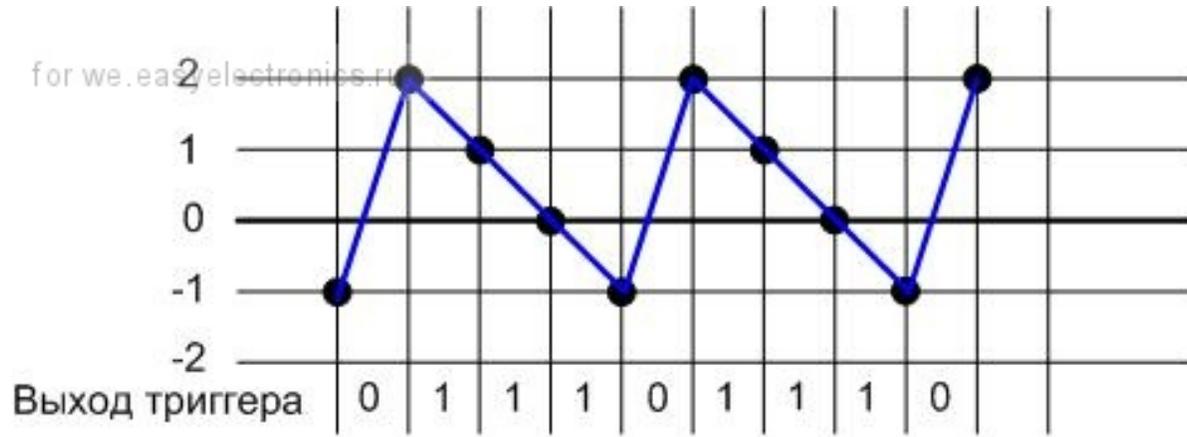


Дельта-сигма АЦП

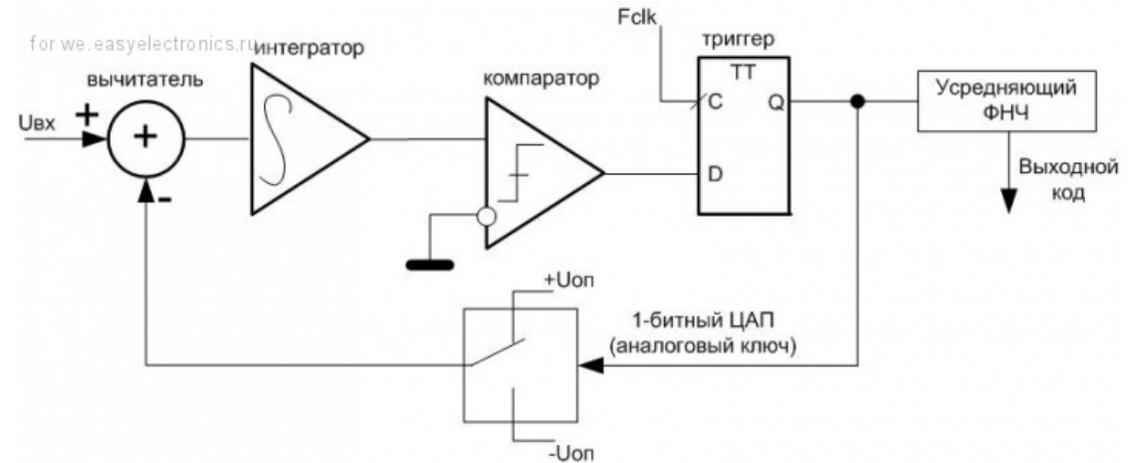


при $U_{вх} = -U_{оп}$ будут одни нули, при $U_{вх} = +U_{оп}$ – одни единицы

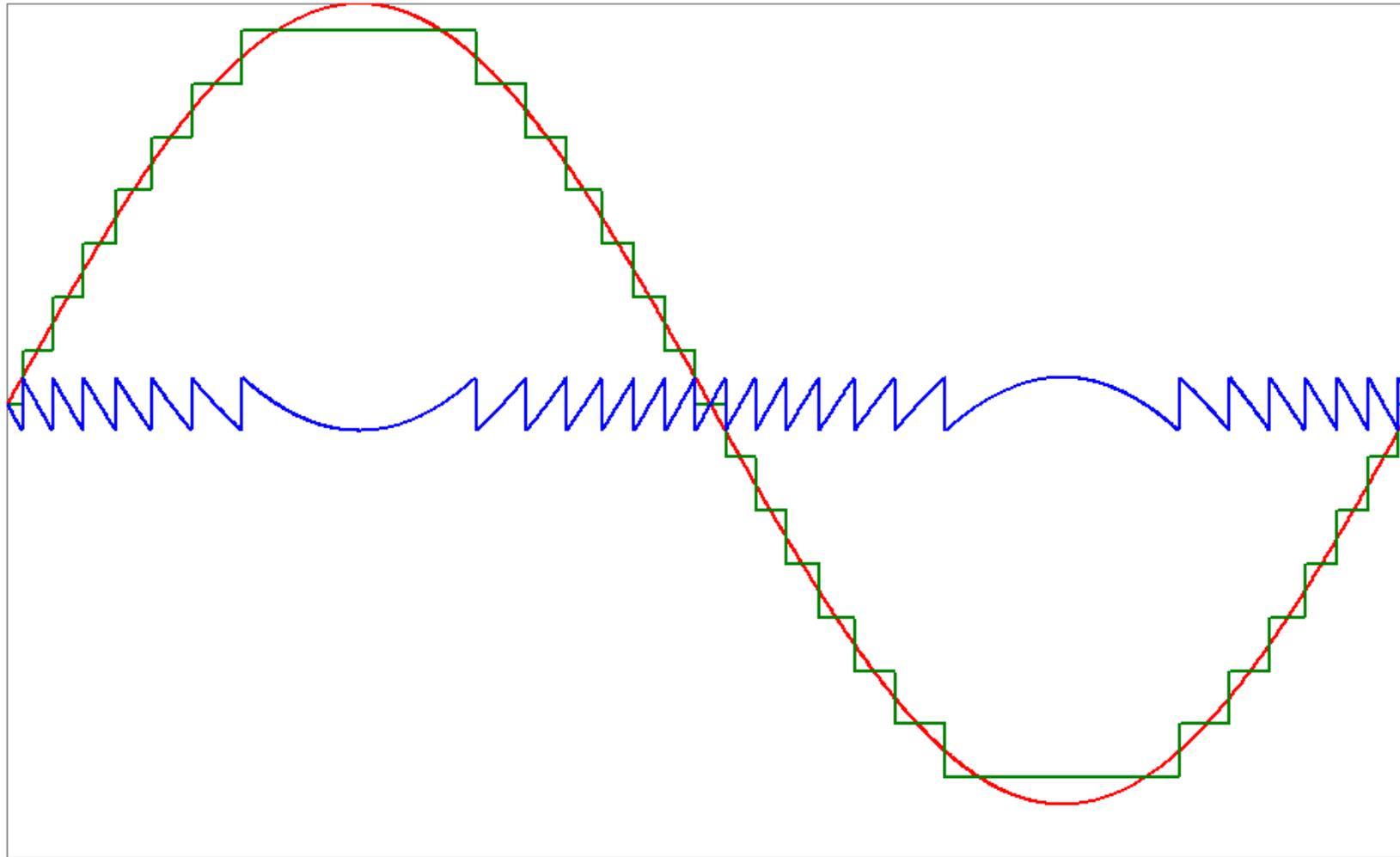
Дельта-сигма АЦП



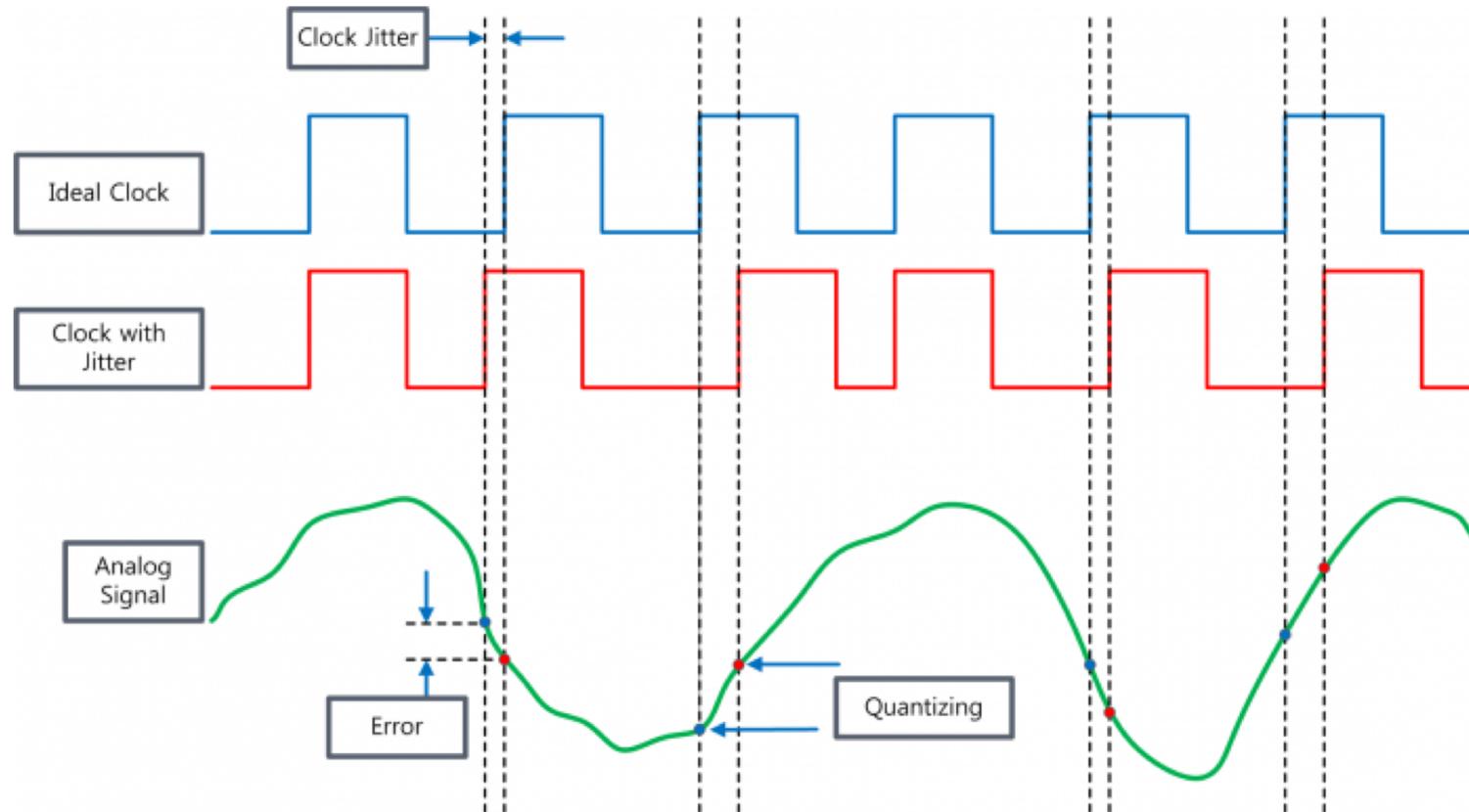
$$U_{оп} = 2В; U_{вх} = 1В$$



Шум квантования



Апертурная погрешность (джиттер)



Дизеринг

Дизеринг – подмешивание случайного шума в исходный сигнал для оптимизации ошибки квантования.



Original
(8 bits)

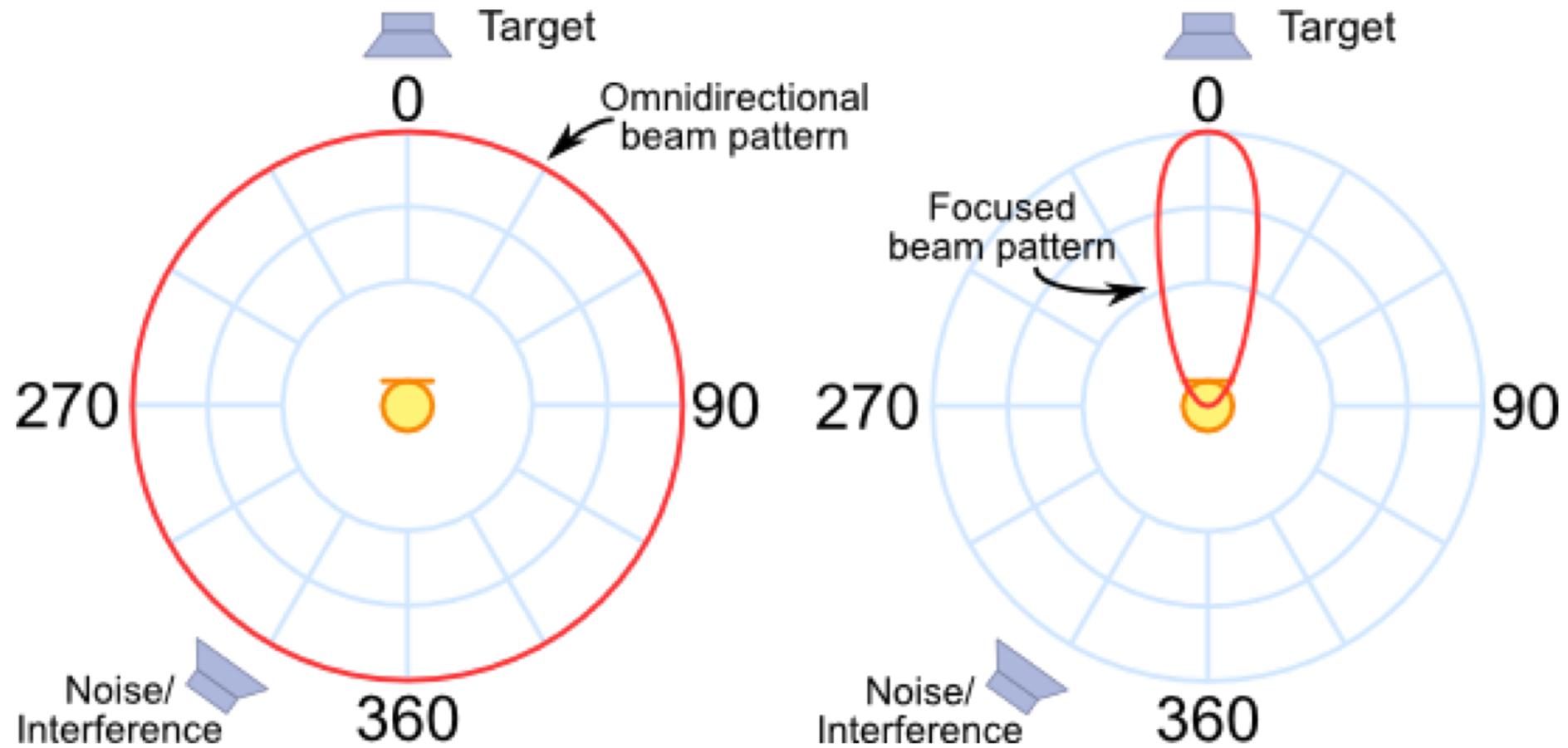


Uniform
Quantization
(1 bit)



Floyd-Steinberg
Dither
(1 bit)

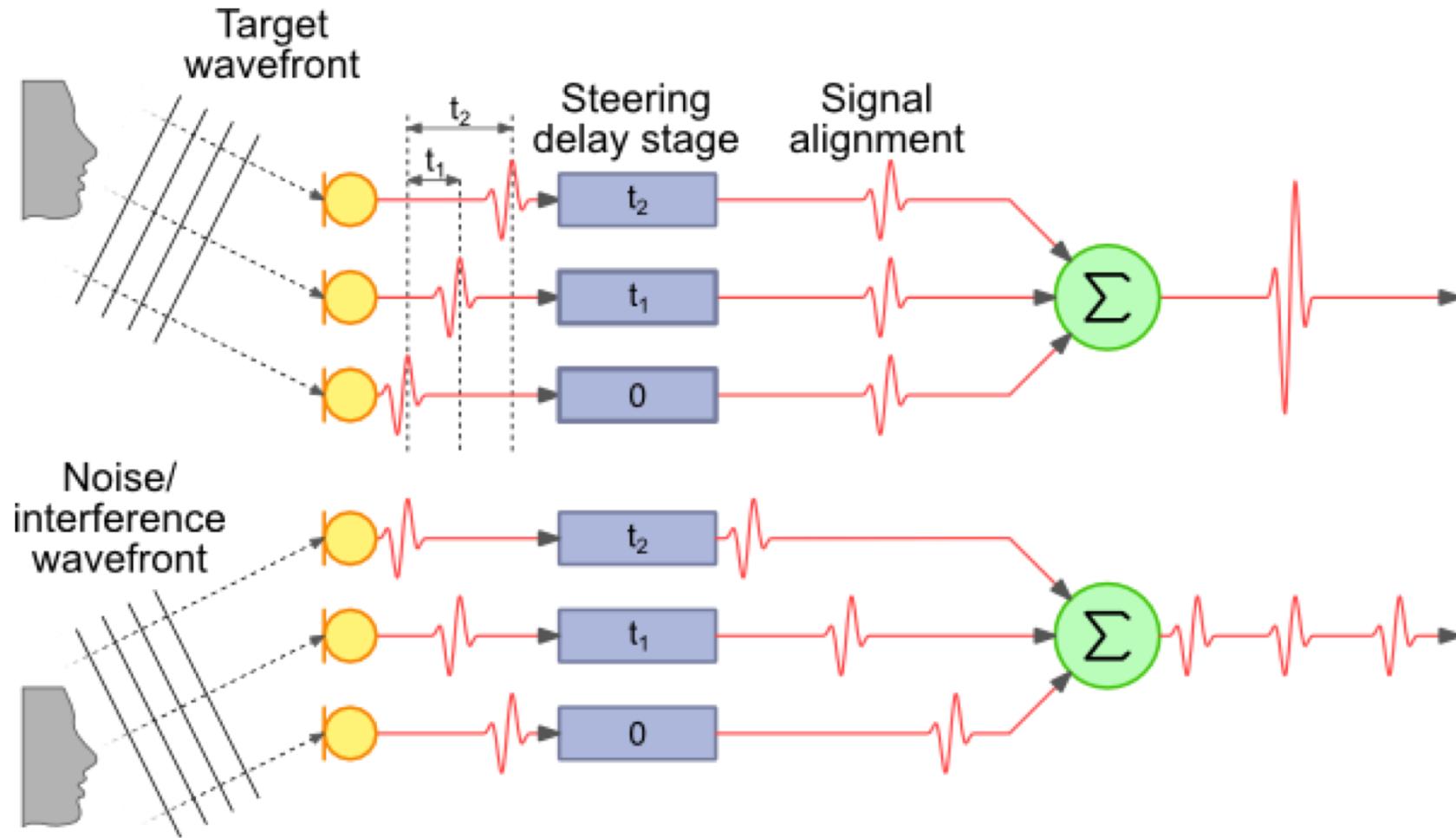
Beamforming



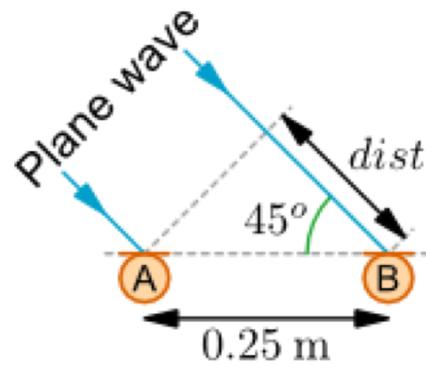
Beamforming

- Идея взята от фазированных антенных решеток
- В случае с аудио все сложнее:
 - Узкополосный радио сигнал vs широкополосный аудио
 - Сферический волновой фронт
 - Сложная конфигурация шума и искажений сигнала
- 2 основных вида:
 - **fixed beamforming**: data-independent, fixed filters F_m
e.g. delay-and-sum, filter-and-sum
 - **adaptive beamforming**: data-dependent adaptive filters F_m
e.g. LCMV-beamformer, Generalized Sidelobe canceler

Delay-and-sum



Delay-and-sum



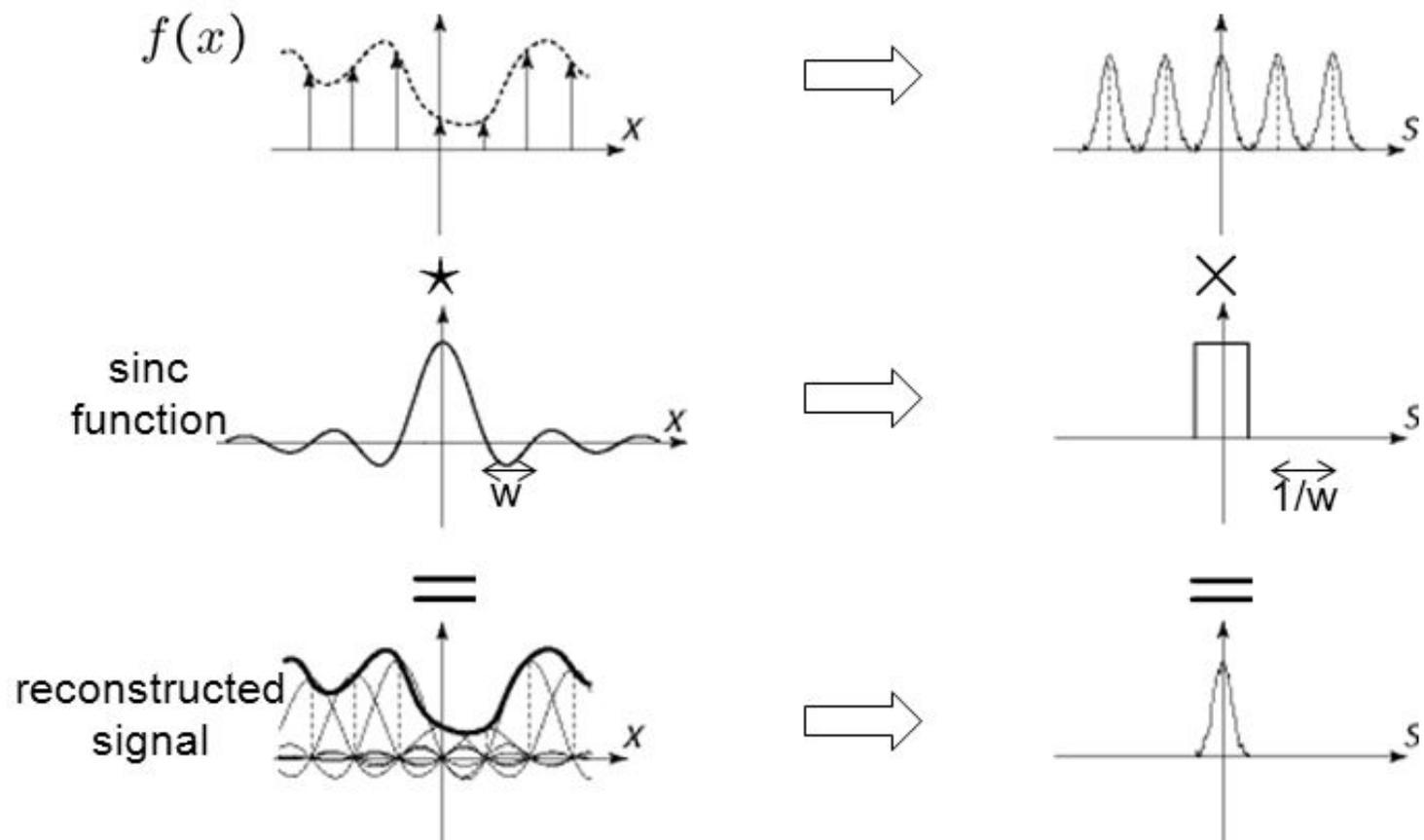
$$dist = 0.25 \times \cos(45) = 0.1768 \text{ m} \quad s = 48000 \text{ samples/second}$$

$$delay = \frac{dist}{v} = \frac{0.25 \times \cos(45)}{343.3} = 0.5149 \text{ ms}$$

$$t_s = \frac{1}{48000} = 20.8333 \mu\text{s}$$

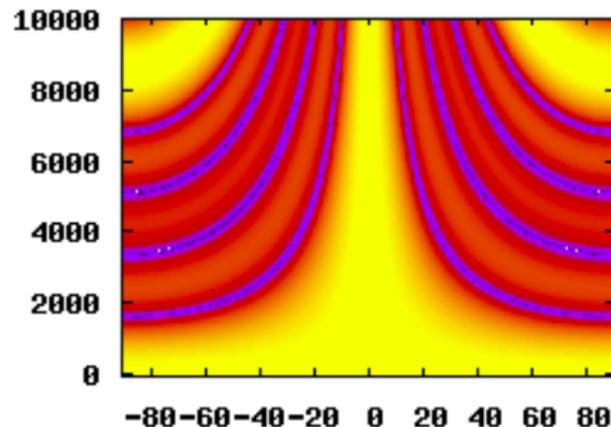
$$delay_s = \frac{0.5149 \text{ ms}}{20.8333 \mu\text{s}} = \frac{48000 \times 0.25 \times \cos(45)}{343.3} = \underline{24.717 \text{ samples}}$$

Восстановление дискретного сигнала

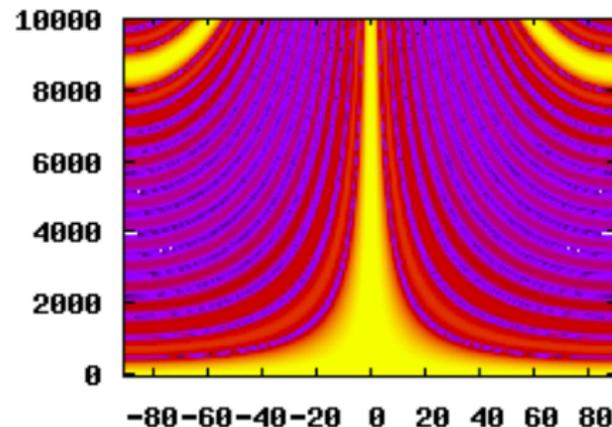


Зависимость от частоты и направления

More Elements

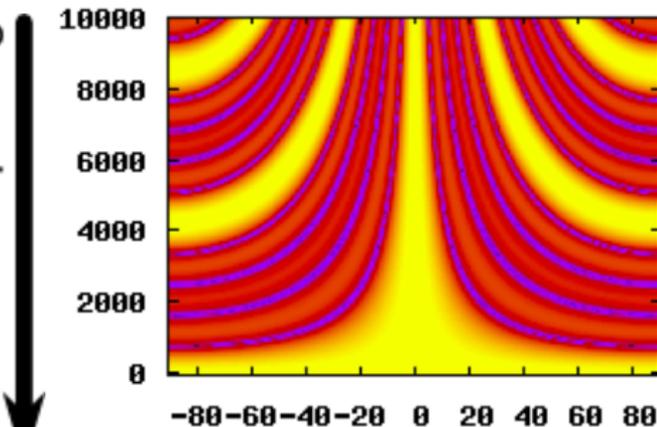


5 Elements, 0.04m Spacing, 0.2m Aperture

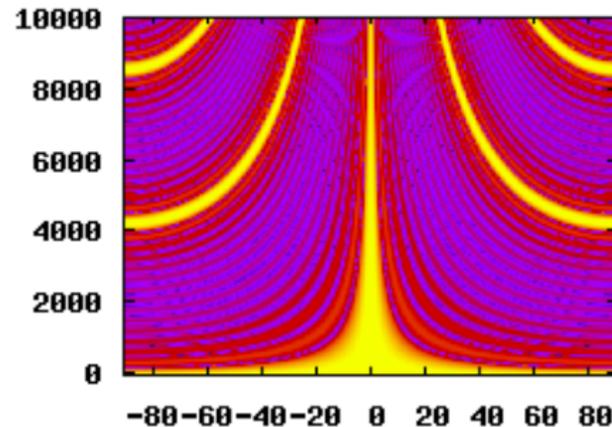


15 Elements, 0.04m Spacing, 0.6m Aperture

Increased Spacing

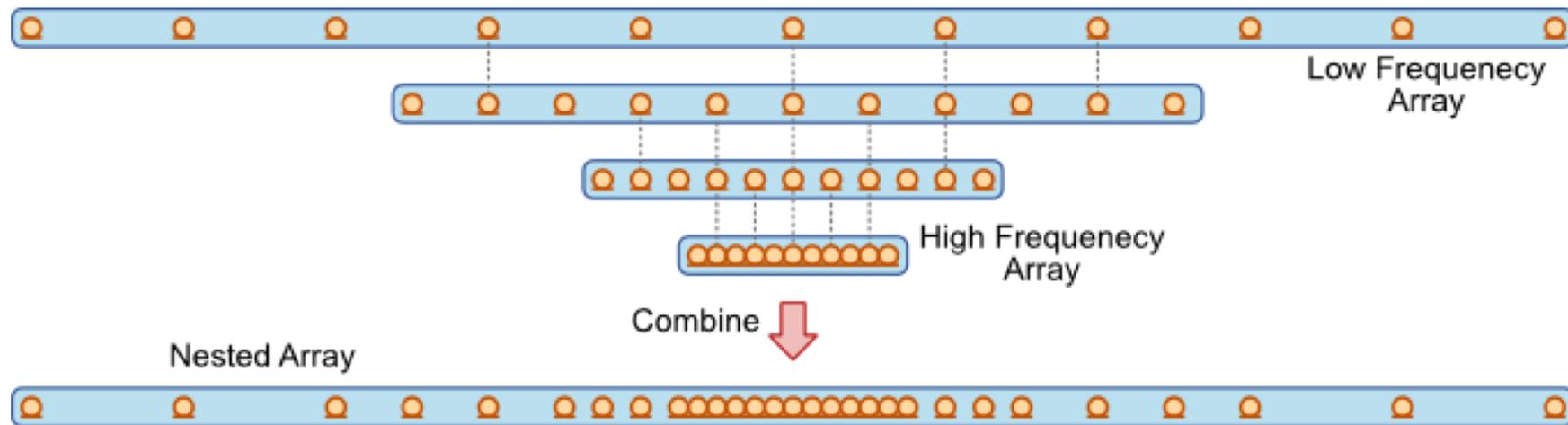


5 Elements, 0.08m Spacing, 0.4m Aperture

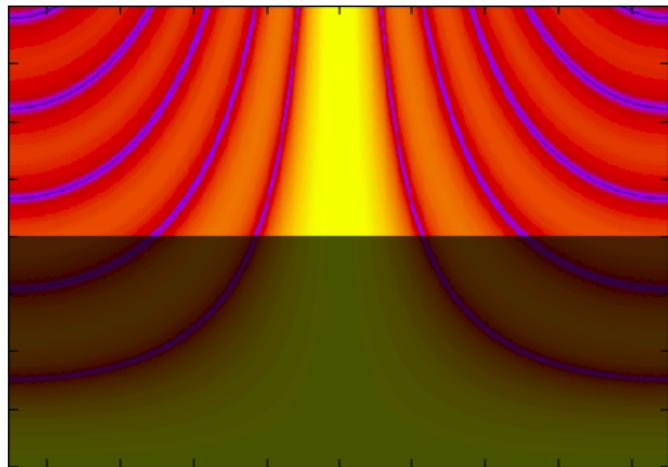


15 Elements, 0.08m Spacing, 1.2m Aperture

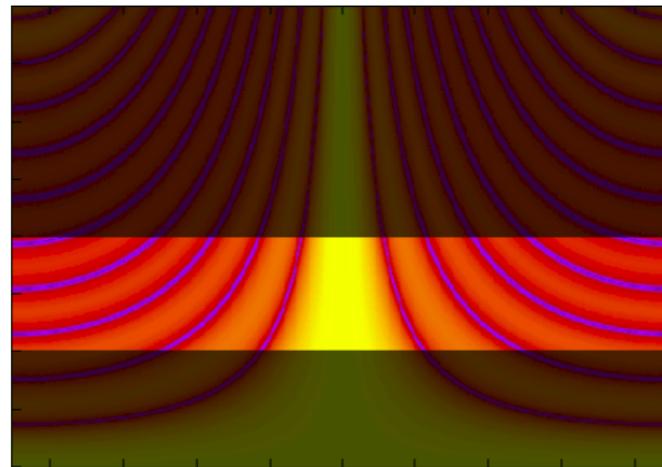
Вложенные массивы



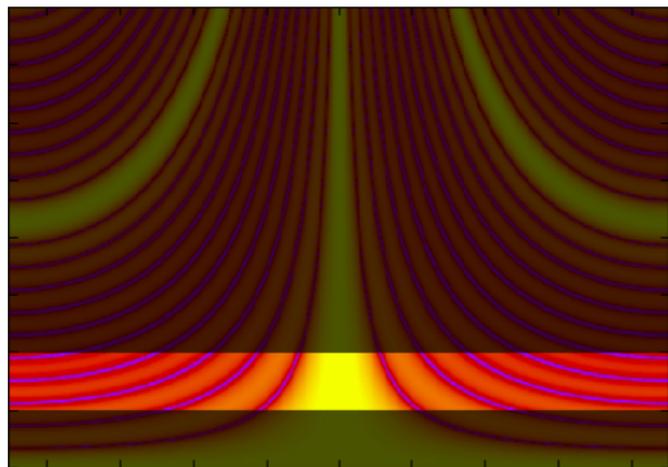
Вложенные массивы



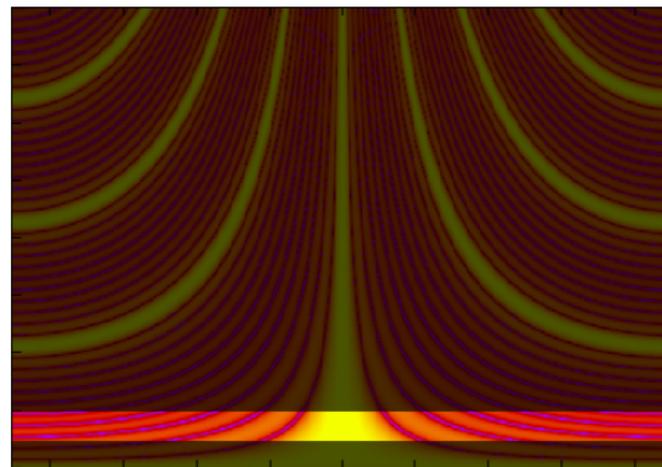
Sub Array, 11 elements, 0.02m spacing



Sub Array, 11 elements, 0.04m spacing

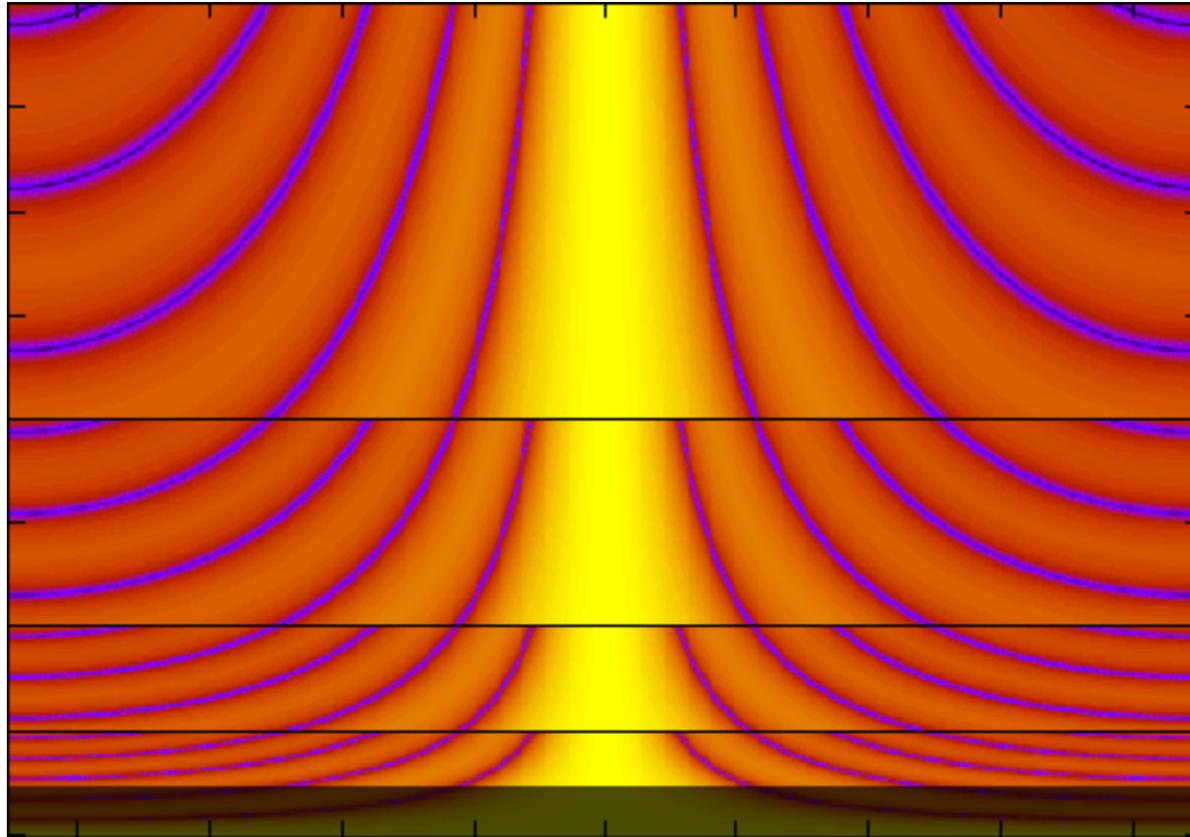


Sub Array, 11 elements, 0.08m spacing

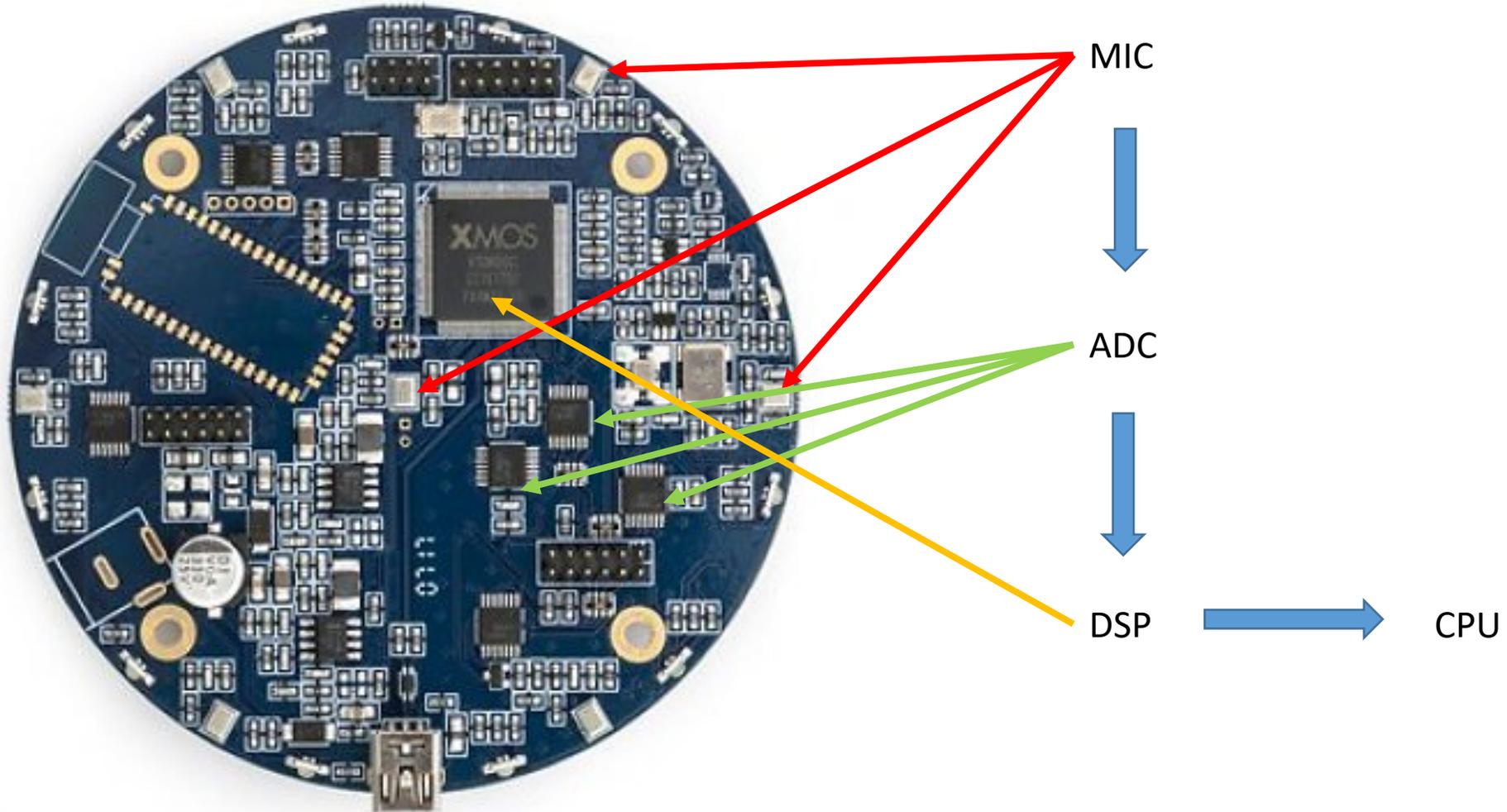


Sub Array, 11 elements, 0.16m spacing

Вложенные массивы



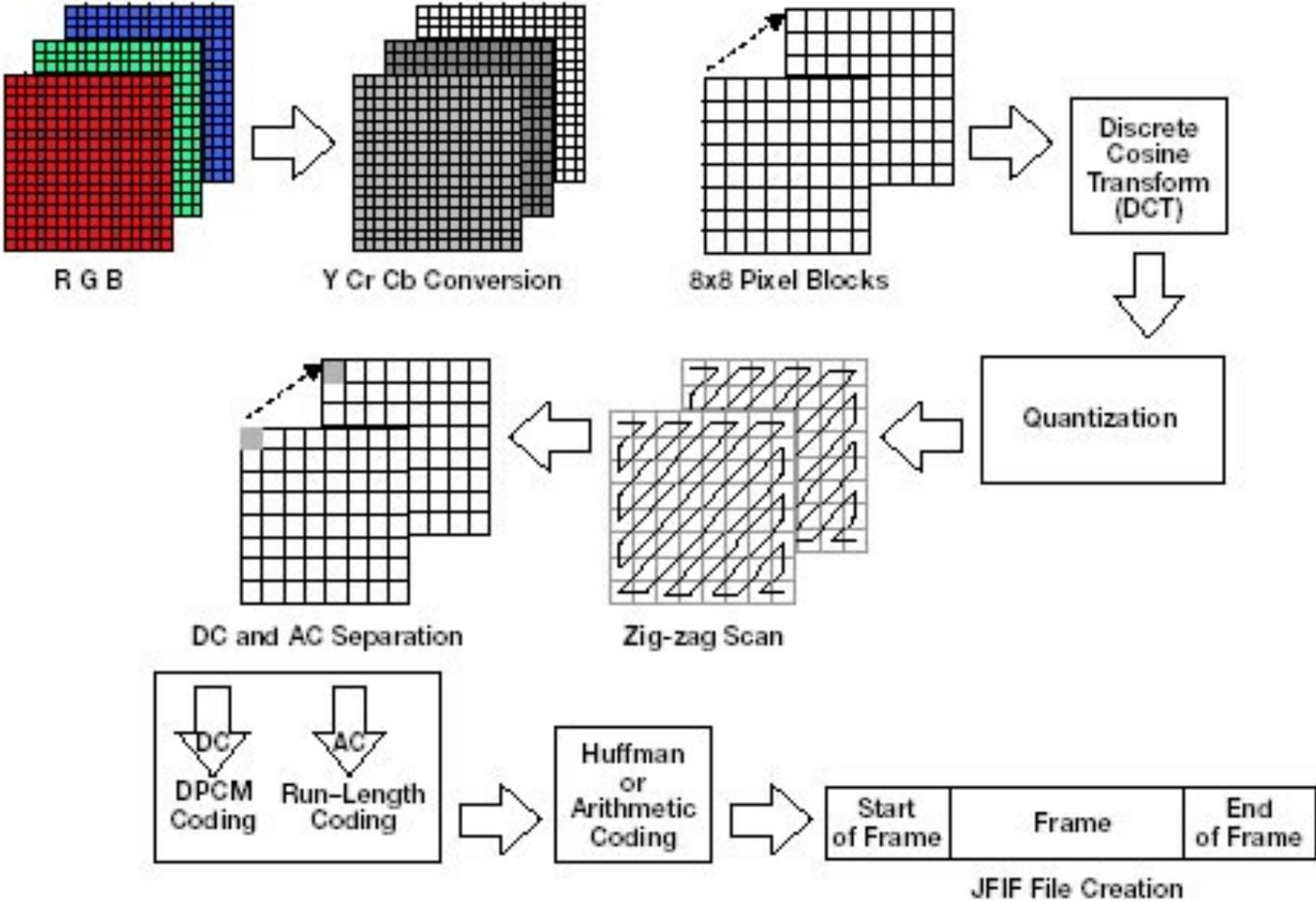
UMA-8



Передача данных

- Как правило, полноценный ASR на устройстве сделать не возможно.
- Решение – детекция ключевой фразы и речи на устройстве, ASR на сервере.
- Решение:
 - оптимизация сигнала в частотном диапазоне а-ля jpeg
 - использование психоакустической модели восприятия человеком

Jpeg



Mpeg / Mp3

