Experimental Methods

IB Physics A Course

Michaelmas term 2023

Prof Pietro Cicuta

pc245@cam.ac.uk

Lectures on: 9, 11, 16, 18, 25, 30 Oct, 1 Nov

note change from last year: inform DoS & arrange 2 supervisions

(8th lecture on writing skills by Dr Melissa Uchida)

With thanks, based on a course by Prof. Chris Haniff

Physics is an empirical science

- Physicists attempt to infer the "laws of nature" from measurements of the world. Other times perhaps more modestly physicists will discover or understand "an effect".

- Typically, measurements are taken, and their behavior is "captured" in some mathematical form. That form *is* the law, or the effect. In the best cases it is useful, encompasses an understanding, is modular with other knowledge and robust to extrapolation away from the experimental domain.

- If deviations from that mathematical form exist, this implies that some revision is needed – but whatever that is, nature doesn't change.

Devising such measurements, making them, and assessing them are fundamental elements of doing physics. You need to know how to do this.

This course's two main themes

- To introduce you to a toolbox of ideas that are helpful in planning, executing and assessing measurements that physicists often have to make.

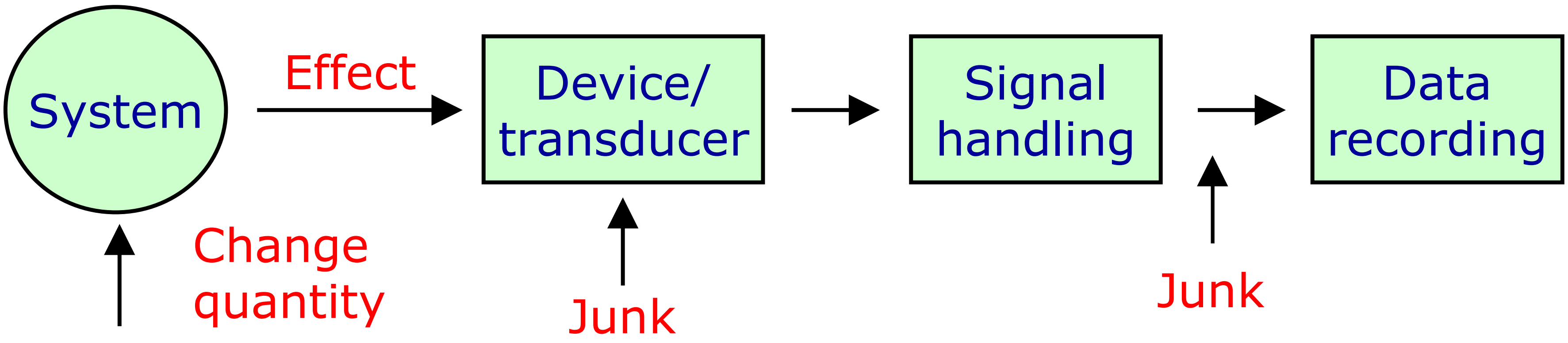
- To provide the background you need to undertake and benefit from the practical classes this term and next:
  - This underlies the sequence/choice of content in the course.
  - But this is a secondary goal – do not lose sight of the first.

Understanding this material will come via the lectures, question sheets, the extra notes and the practical classes.

Topics to include in our "Toolbox"

- Data sampling, transforms and correlation functions (1 lecture)
- Advancing our experiment design (2 lectures):
    - Oscilloscope and Amplification
    - Feedback and PID control
- Processing data (2 lectures):
    - Errors
    - Filtering, lock-in, isolation
- Testing your theory (2 lectures):
    - Distributions
    - Assessing significance: fitting, inference
- Presenting your work (1 lecture).

Progress in physics often involves:



Coalescence of BHs	Gravity wave antenna	Amplify and filter	Analyse (e.g. create spectrum)
Evaporatively cool dilute gas	Laser probe pulse and CCD	Measure phase space distribution, infer temp and density	

- PLUS
  - Statistical methods to analyse the results, and comparison with models.

Lecture 1 –sampling, FFT, LT and correlations

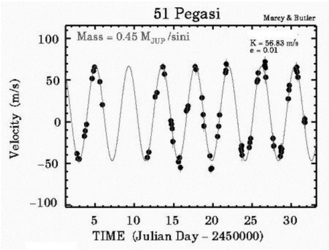
- Digital signal sampling rate:
  - The minimum sampling rate – Nyquist’s criterion.
  - Reconstruction from samples.
  - Quantization and “oversampling”.
  - Spectral resolution.

- Digital signal sampling discretisation:

  ADC

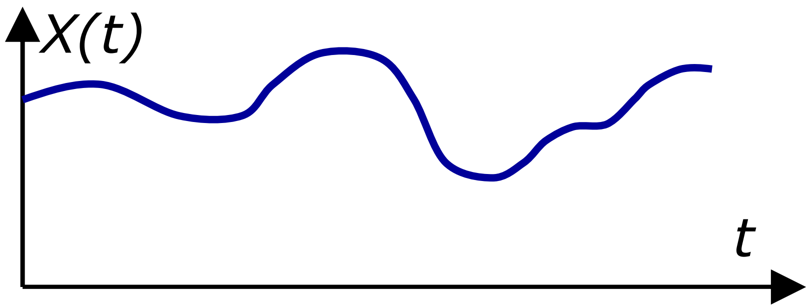
- How to extract features from data:
  - Why/when FFT or LT
  - Correlations in time domain

How do we capture data accurately



- Consider the output of some system measuring a continuous signal.

- Nowadays it is rare to store these values in "analogue" form, i.e. the continuous values of *X(t)*.

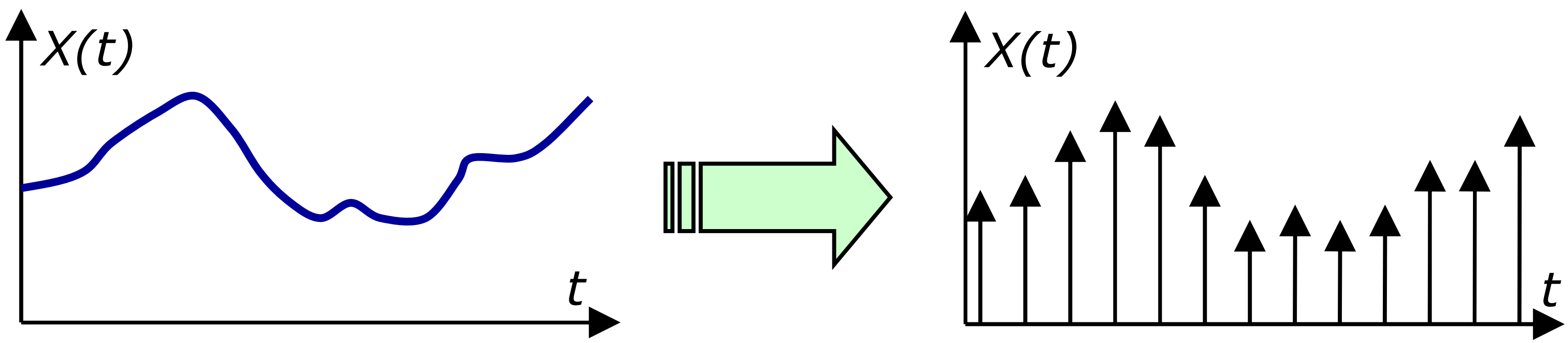


  - There are many reasons:
    - Maintaining precision may be difficult.
    - Media may be non-linear $\Rightarrow$ limitations in dynamic range.
    - We may not have access to the system all the time.
    - Sampling reduces the data volume.
    - Sampled data may be easier to transmit without information loss.
    - Further processing is almost always digital.

So preference is to sample digitally.

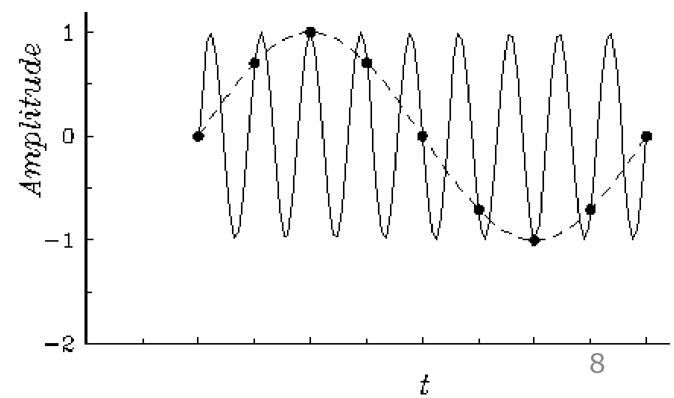
What determines how we should <u>sample</u> the signal?

- Usually we record the instantaneous values of *X(t)* at regular intervals.



- How fast should we sample the underlying function?
- For how long is each recording (each point) averaged? (we assume instantaneous in this lecture)
- How accurately should we sample the underlying function?
- How long in total should we sample the signal for?

> There is a <u>minimum</u> rate of sampling required for a given rate of signal variation.
>
> Why? Because an under-sampled high frequency is indistinguishable from a properly sampled low frequency.

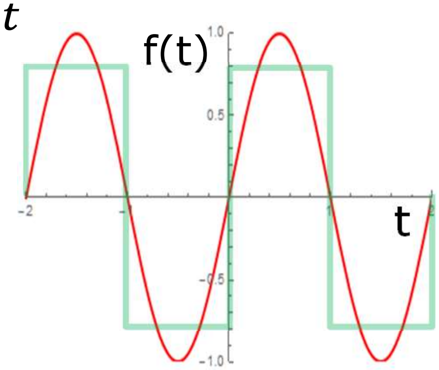


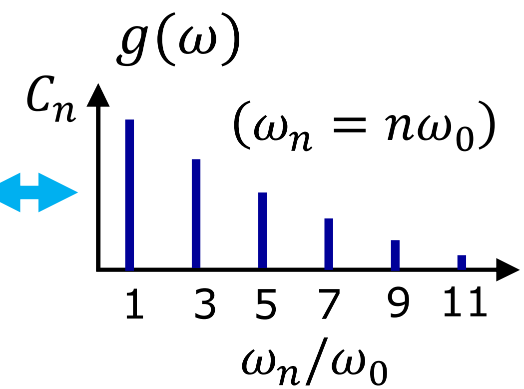
Fourier Transforms in data/experiment

- Fourier series is a decomposition of a function with period *T*:

$$f(t) = \sum_{n=-\infty}^{\infty} C_n e^{i2\pi nt/T} = \sum_{n=-\infty}^{\infty} C_n e^{in\omega_0 t}$$

e.g. a square wave $C_n = \begin{cases} \frac{1}{n}, & n = \text{odd} \\ 0, & n = \text{even} \end{cases}$





- For an aperiodic fn, take limit $T \to \infty$ (Fourier transform)
  - Sum over discrete frequencies becomes integral over continuous variable ω (the ω_n are infinitesimally spaced).

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(\omega) e^{i\omega t}\, d\omega$$

  - $g(\omega)$ characterises the amplitudes and phases of the constituent complex exponentials making up the signal.

$$g(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i\omega t}\, dt$$

Note: the prefactors on direct and inverse are completely arbitrary, but their product needs to be $1/2\pi$

An example

$$g(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i\omega t}\, dt$$

$$f(t) = A\cos\omega_0 t$$

$$g(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} A\cos\omega_0 t\ e^{-i\omega t} dt$$

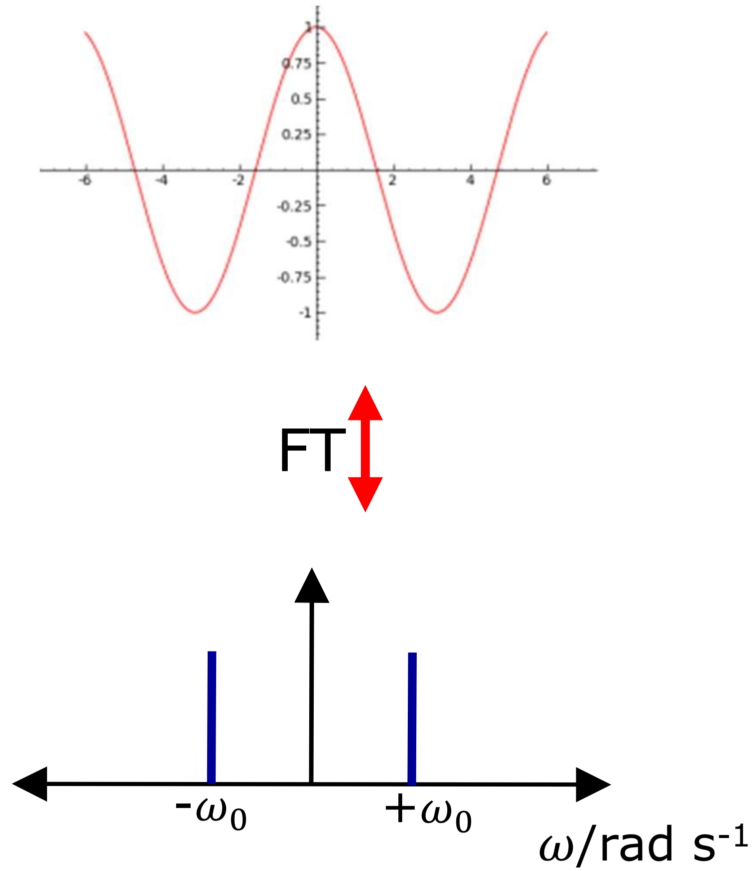
$$g(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{A}{2}\left(e^{i\omega_0 t} + e^{-i\omega_0 t}\right) e^{-i\omega t} dt$$

$$g(\omega) = \frac{1}{\sqrt{2\pi}} \frac{A}{2} \int_{-\infty}^{\infty} e^{-i(\omega+\omega_0)t} + e^{-i(\omega-\omega_0)t} dt$$

Which is only non-zero when:

$$\omega = \pm\omega_0$$

$$g(\omega) = \frac{1}{\sqrt{2\pi}} \frac{A}{2}\left[\delta(\omega+\omega_0) + \delta(\omega-\omega_0)\right]$$



(δ = Dirac ‘delta function’)

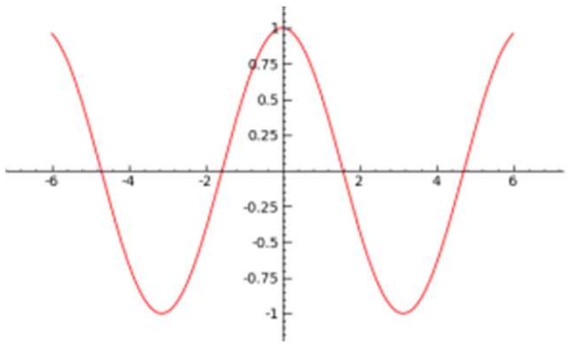
And if you have “perfect” $g(\omega)$ you can recover $f(t)$ with the “inverse” transform.

Other important Fourier Transforms

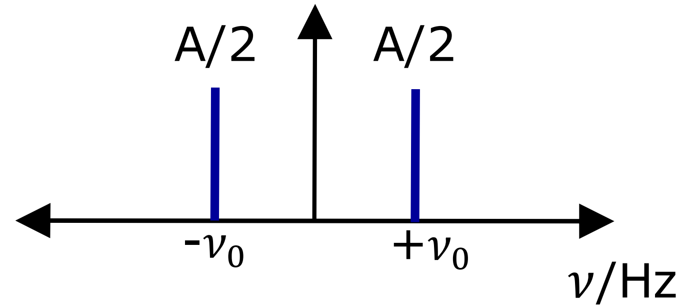
$$g(\nu) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{-i2\pi\nu} \; dt$$

$f(t)$

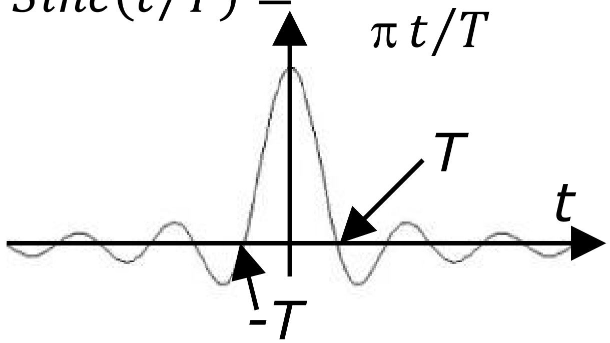
$g(\nu)$



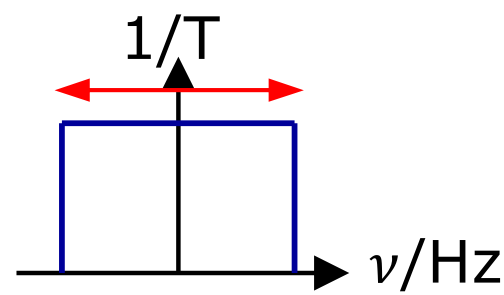
FT

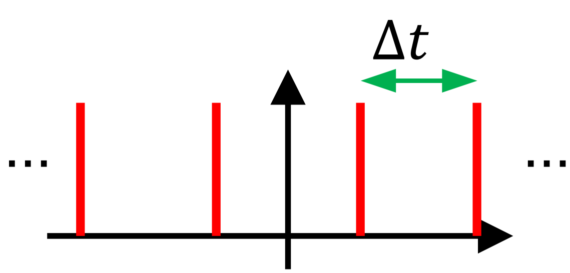


$$Sinc(t/T) = \frac{Sin(\pi\, t/T)}{\pi\, t/T}$$

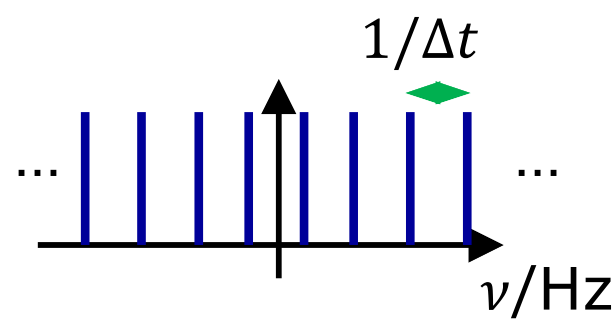


FT





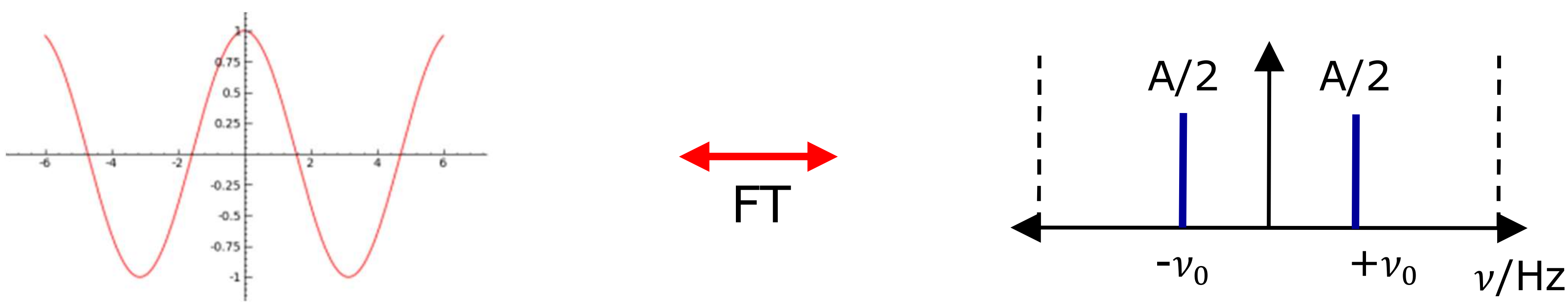
FT



Minimum sampling rate: Nyquist

- Nyquist's criterion – basic version:
  - For a band-limited function, you need to sample at a minimum rate of twice the highest frequency Fourier component present in the signal.
  - If the sampling is noiseless, then you can recover the signal perfectly from its samples.

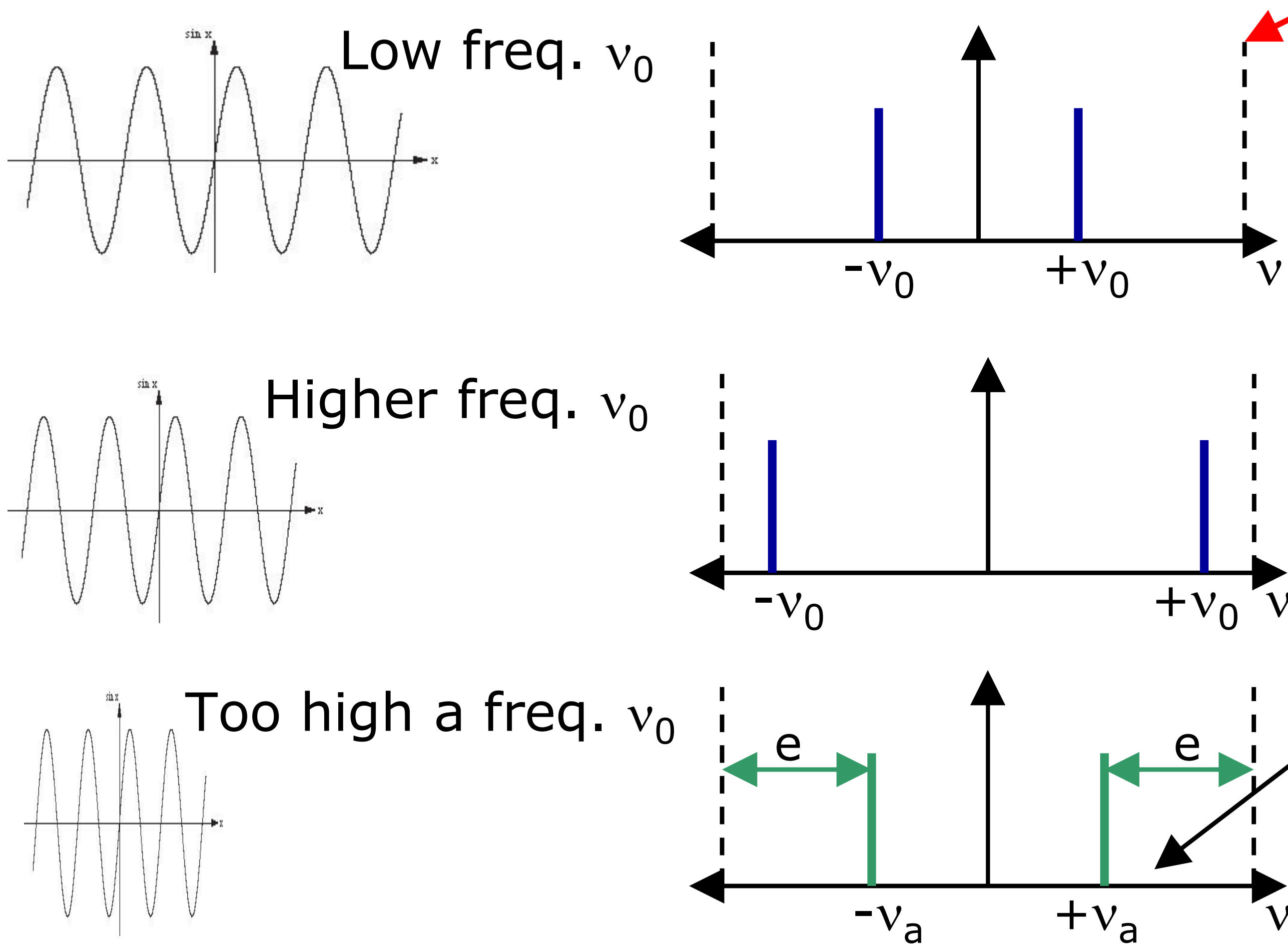
- Frequency content of the signal: examine its FT
  - This will reveal the amplitudes and phases of the constituent sine and cosines (or complex exponentials) that make it up.
  - These will have frequencies between $\pm$ infinity.

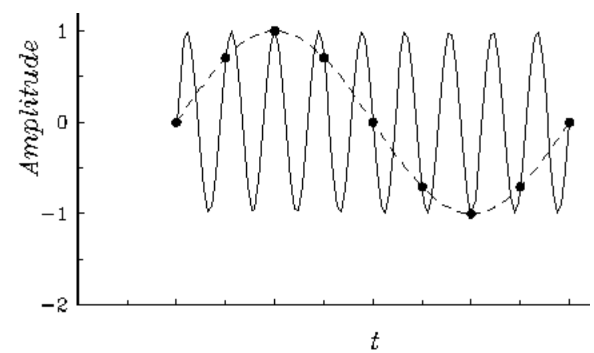


What happens if we try and recover an undersampled signal? we get what is called "aliasing"

- We can illustrate this by sampling waveforms then asking what was the frequency content of what was measured.

Half the sampling frequency





ν_a is the false, aliased frequency seen:

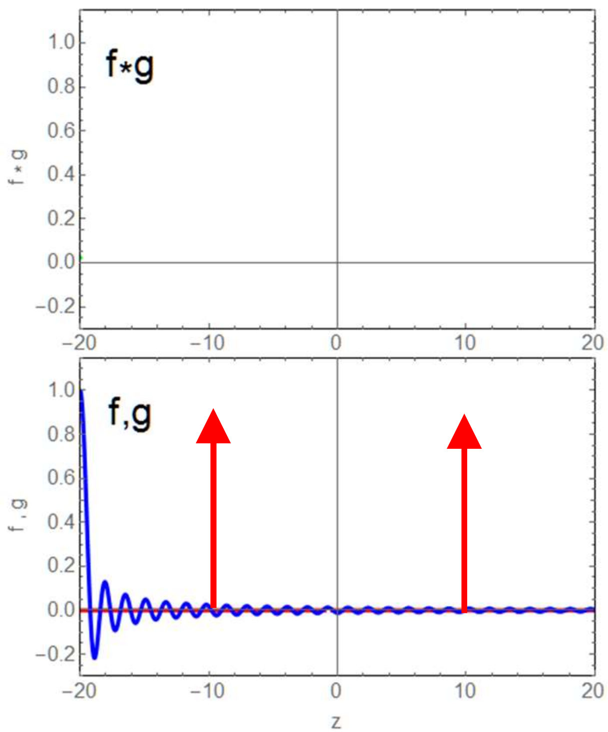
$$\text{True freq} = \frac{\text{Sampling freq}}{2} + \text{e}$$

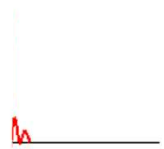
We can understand Nyquist's limit by examining the frequency spectrum of a sampled signal

- First we need to review two key results from Fourier theory

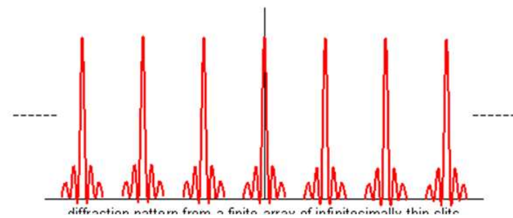
$$h(z) = f * g = \int_{-\infty}^{+\infty} f(t)g(z-t)\,dt$$

- The convolution of two functions:





=



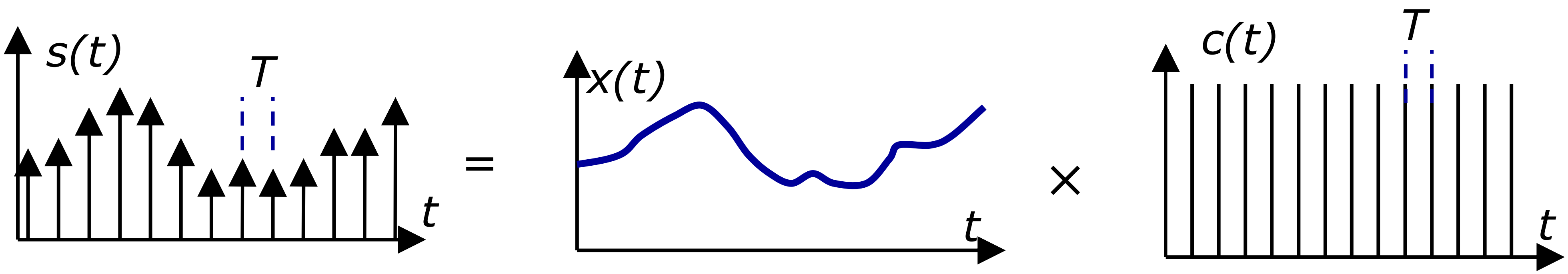
- eorem":

$$t)\cdot g(t)) \propto F(\omega) * G(\omega)$$

$$FT(f * g) \propto F(\omega) \times G(\omega)$$

We can understand Nyquist's limit by examining the frequency spectrum of a sampled signal

- The sampled signal *s(t)* is given by the product of the true signal *x(t)* and a comb of unit impulse "δ-functions", *c(t)*:



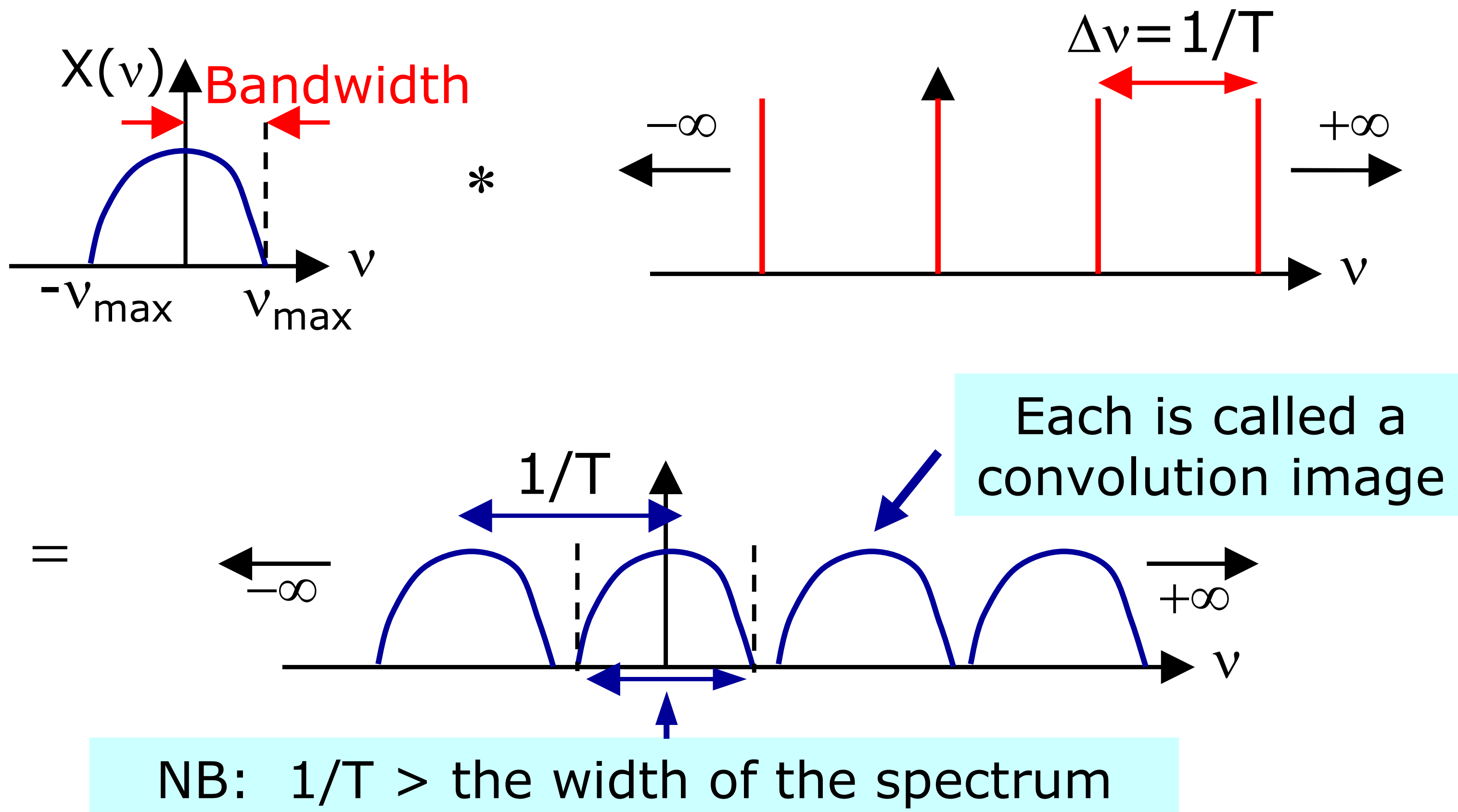
- What is the frequency spectrum of *s(t)*?

The convolution theorem says it's the Fourier spectrum of the true signal convolved with the Fourier spectrum of the "comb" we sampled it with.

$\widetilde{S(\nu)} = \widetilde{X(\nu)}$ * array of δ-functions with spacing $\Delta\nu = 1/T$

What does S(ν) look like?

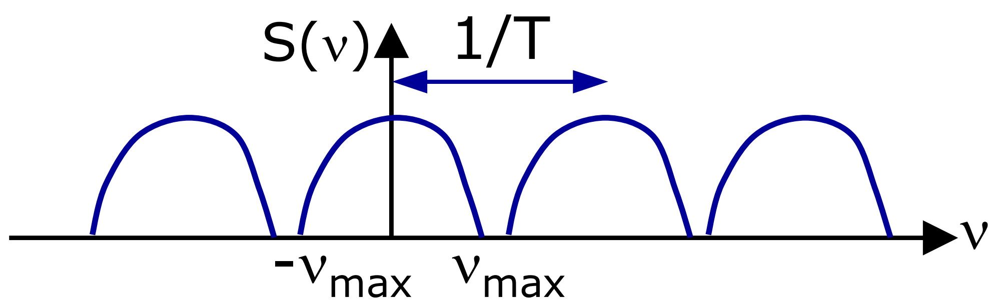
$\widetilde{S(\nu)} = \widetilde{X(\nu)}$ * array of δ-functions with spacing $\Delta\nu = 1/T$



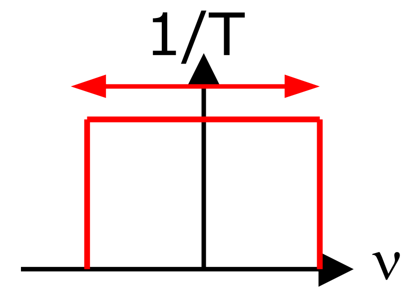
- So, the Fourier content of the analog signal is maintained without “mixing”.

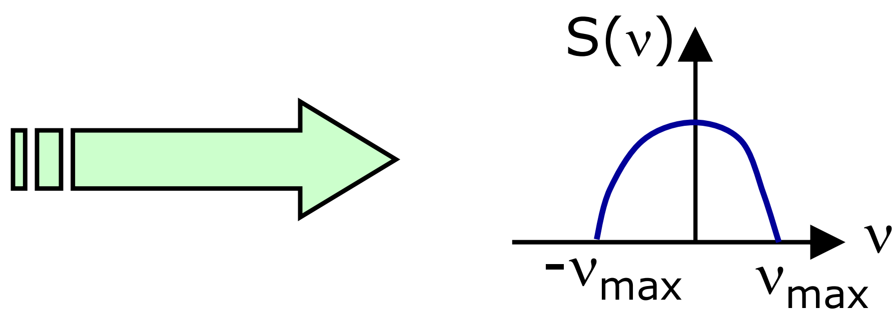
Can we recover the signal from its sampled representation?

- Step 1: FT the sampled signal and multiply it with a top-hat.



×

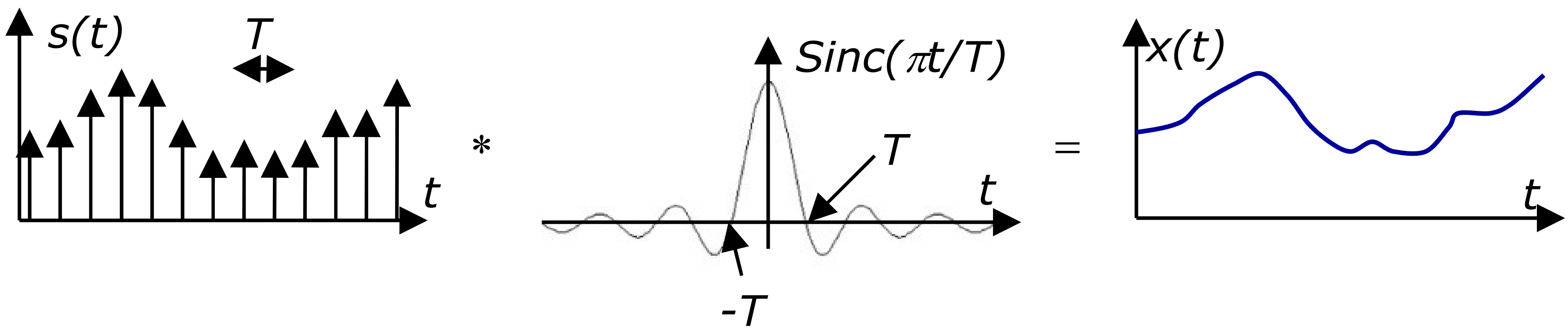




- Step 2: Inverse FT this and you get *x(t)*.
- This now explains Nyquist's criterion:
  - This process only works if the convolution images don't overlap, i.e. if $1/T$ (the sampling frequency) $\geq 2\,\nu_{max}$
  - If this criterion is not met, then you get severe distortion in the recovered signal – aliasing.

What is the real-space version of this reconstruction (e.g. in a digital music player)?

- Just replace the Fourier operations with their equivalents in the time domain
  - NB: multiplying the FT of a function with a top-hat in frequency space ≡ convolving the function with a Sinc function in time.

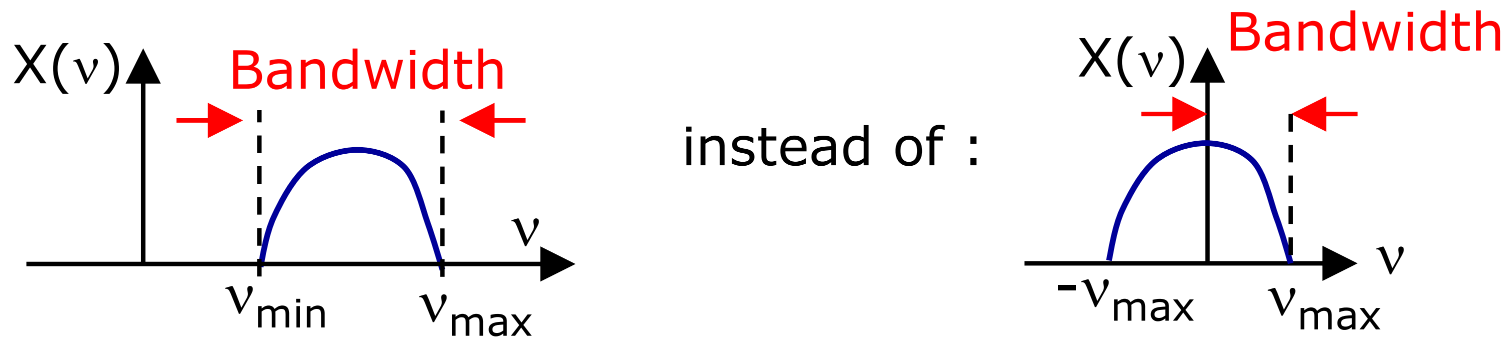


- So all we do to recover the signal is convolve the sample values with a special smoothing kernel, a sinc function.
  - Note the sinc function extends to infinity…

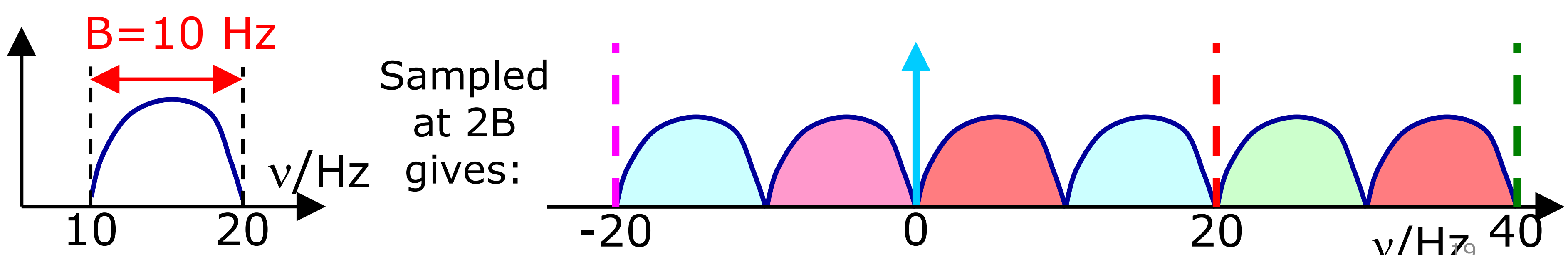
Nyquist's criterion – a more subtle version

- In understanding Nyquist's theorem we have assumed the signal contains frequencies from 0 Hz to $\pm\nu_{max}$ (and their negative frequency partners).
- However, in some cases we <u>know</u> a signal <u>only</u> contains power over a range of frequencies that does not include 0 Hz, i.e. over a known bandwidth, *B*, with $\nu_{min} > 0$

E.g.



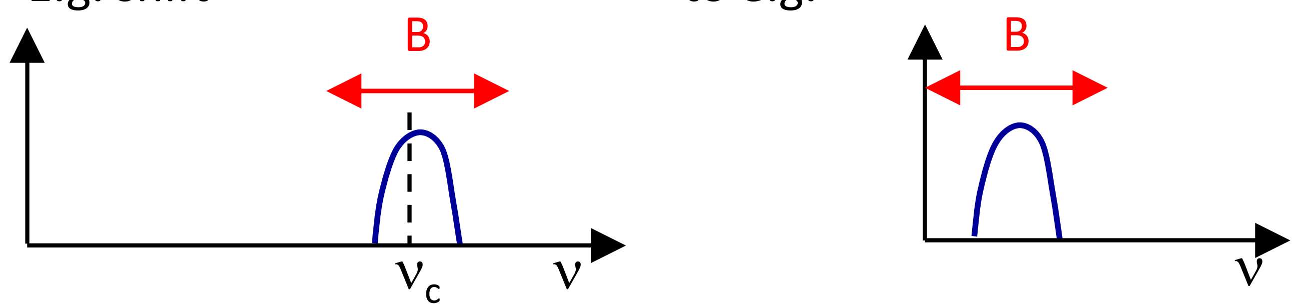
- If you sample at $\geq 2B$ (rather than $\geq 2\nu_{max}$), the convolution images need not overlap, implying that a sampling rate of $\geq 2B$ is OK.



Sub-Nyquist sampling

- So, if you want to sample at sub-Nyquist rates, you need to consider what the spectrum of the sampled function will look like to see if overlap of the convolution images will occur.
    - If there is overlap then you should either:
        - Sample faster, or
        - Lower the signal frequency, while preserving the information. This wont change B, but it may be easier to sample faster at these lower signal frequencies:

E.g. shift to e.g.



- NB: this approach of "shifting" the signal you are interested in to a different temporal frequency is a very common technique in experimental physics, for other reasons too.

Laplace and Fourier transforms – why one or the other?

Fourier FT

$$g(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}\, dt$$

Laplace LT

$$g(s) = \int_{0}^{\infty} f(t)e^{-st}\, dt$$

You can see some relation… if we do the two-sided Laplace, and set $s=i\omega$, and if the integrals exist, then they are the same.

- They both allow to simplify/solve differential equations.
- Depending on the type of signal and/or maths, one will give simpler expressions than the other.
- Much of physics is done with FT, much of engineering with LT.
- LT is better suited to initial value problems, and very natural for exponential decaying functions.

Some properties of signals follow from FT

- If we have a continuous signal *f(t)* the total "energy" is

$$E_{Tot} = \int_{-\infty}^{\infty} |f(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |g(\omega)|^2 d\omega$$

"Parseval's Theorem"

So $|g(\omega)|^2$ is the "spectral energy density", in Hertz.

- If we have a continuous (real) signal *f(t)* we can ask how similar it remains to itself over time, this is the autocorrelation function:

$$C_{ff}(\tau) = \int_{-\infty}^{\infty} f(t) f(t+\tau) dt \quad \text{which is an energy. } \tau \text{ is a delay time.}$$

Interesting properties: even in τ ; has $C_{ff}(0) > 0$ and $C_{ff}(0)$ is the max.

- If we have two continuous (real) signals *f(t)* and *g(t)* we can ask how similar they are to each other over time, this is the crosscorrelation function:

$$C_{fg}(\tau) = \int_{-\infty}^{\infty} f(t) g(t+\tau) dt \quad \text{which is also an energy.}$$

Across Frequency and Time domains

Via two steps of calculus that will be easy once you see in maths.*
**Note: this is easy only for finite energy signals.*

- Take the FT of the autocorrelation function

$$FT[C_{ff}(\tau)] = FT\left[\int_{-\infty}^{\infty} f(t)f(t+\tau)dt\right] = |g(\omega)|^2$$

"Wiener-Khinchin Theorem"

$$\text{corollary:} \quad C_{ff}(0) = E_{Tot}$$

So autocorrelation function and power spectrum are related by FT.
Common situation is:

$$C_{ff}(\tau) = C_{ff}(0)\exp(-\tau)$$

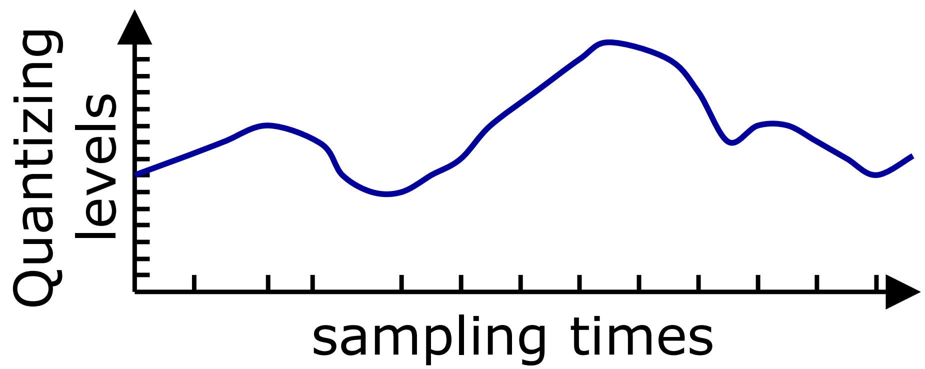
$$FT[\exp(-a|t|)] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(-a|t|)\exp(-i\omega t)dt =$$

$$\frac{1}{\sqrt{2\pi}}\int_0^{\infty}\exp(-t(a+i\omega))dt + \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0}\exp(-t(-a+i\omega))dt = \sqrt{\frac{2}{\pi}}\frac{a}{a^2+\omega^2}$$

the "Lorentzian" function

How accurately must we sample our function?

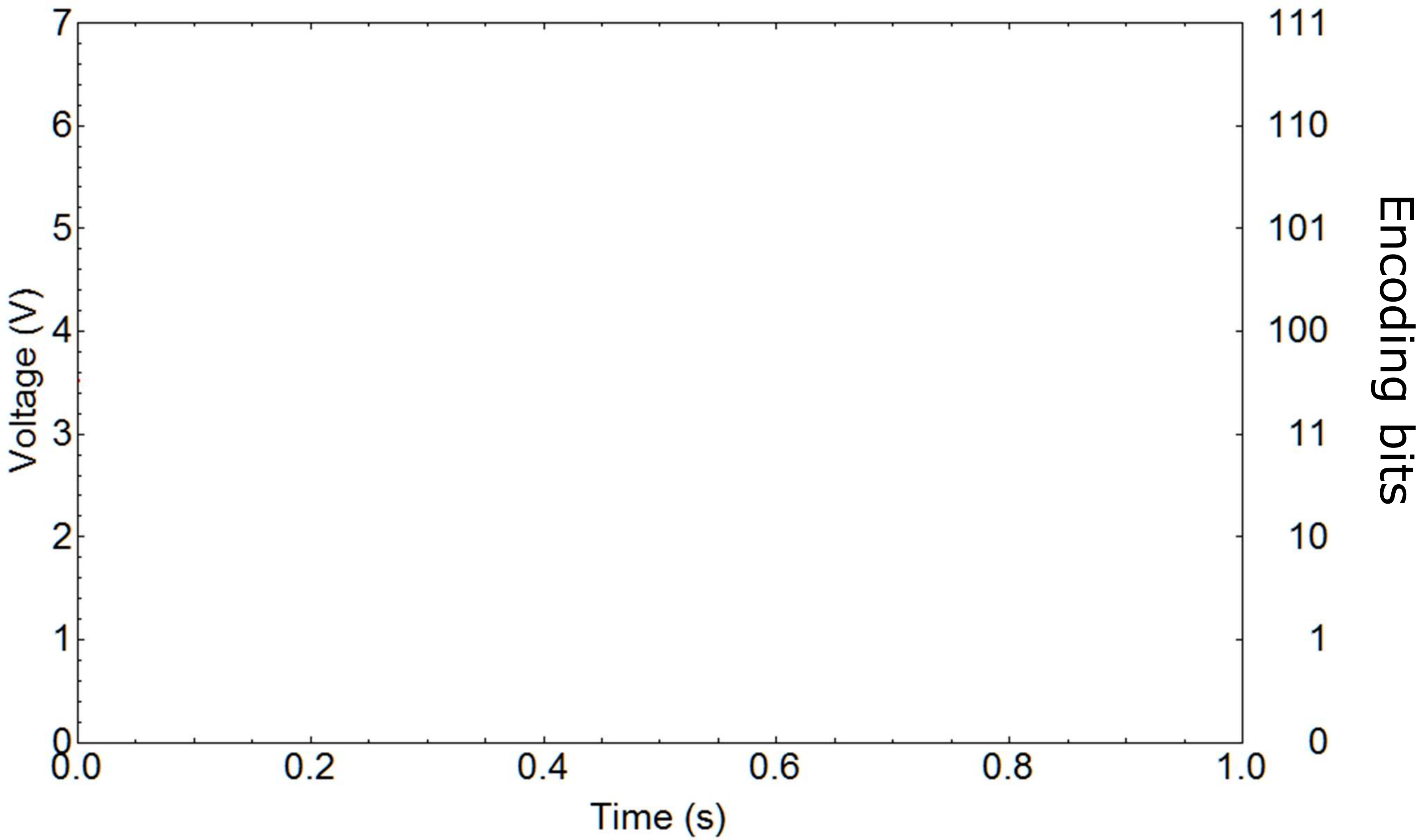
- Digitizing is a 2-step process involving sampling & quantizing.
  - The quantization levels are not continuous



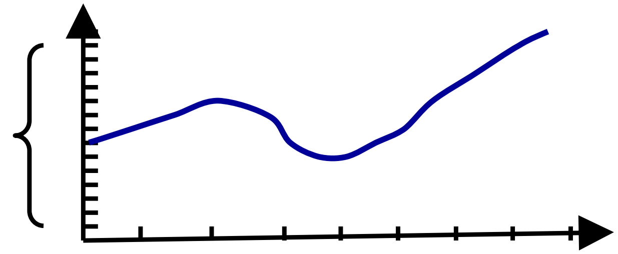
- In practice different terms are used interchangeably:
  - Hence “sampler” = “digitizer” = “analogue-to-digital convertor” = “A-to-D convertor” = “A to D”
  - Similarly, “digitization error” = “quantization error”.

- Key issues are:
  - At any given time, you want there to be a quantization level close to the actual value of the signal.
  - If the signal is noisy, then you don’t need to sample it so finely.

Quantization with too few bits



Quantitative details



- If you use 2^N quantizing bins this is called N "bit" sampling.

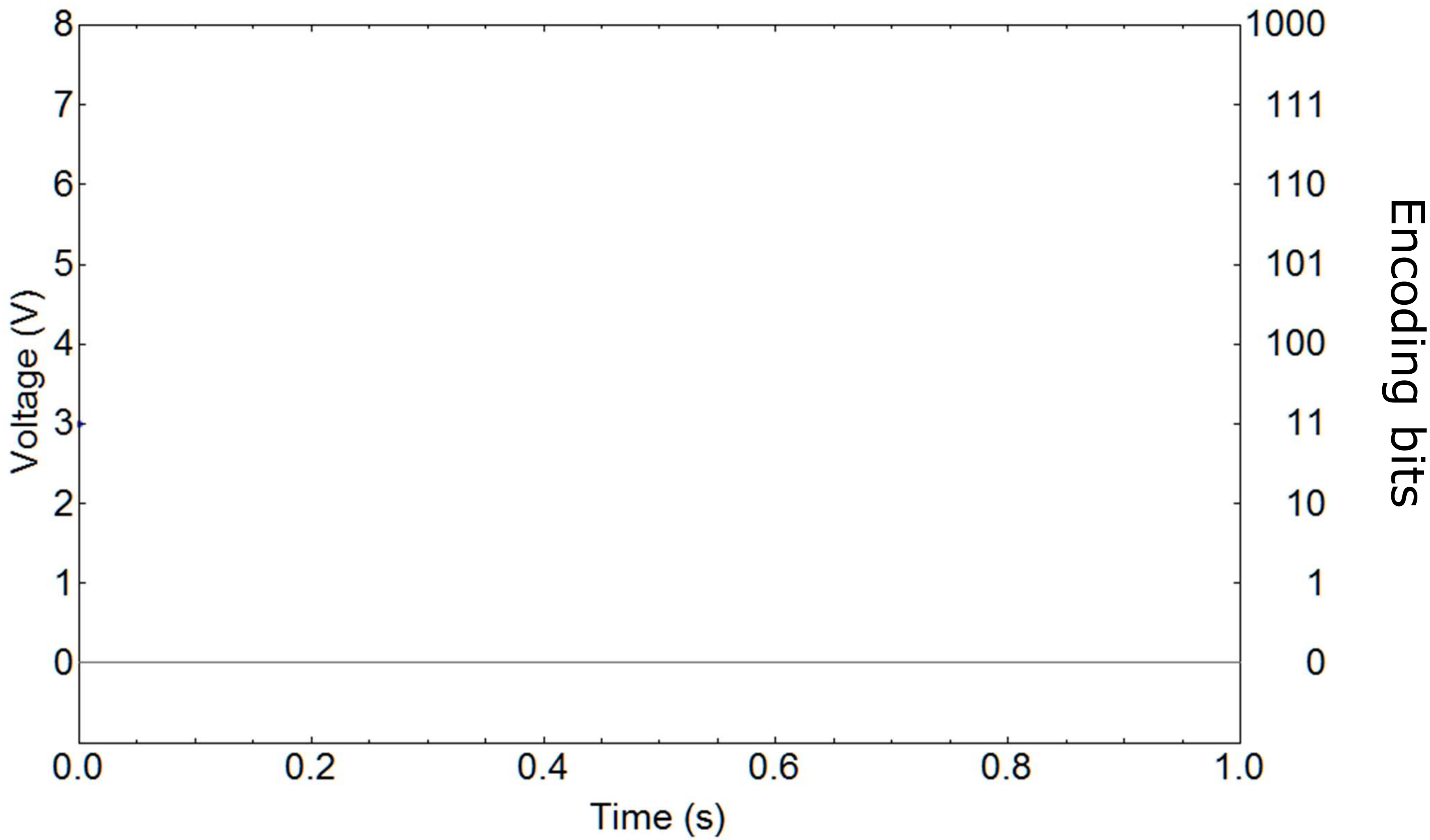
  Quantising a 1m ruler in mm → 1000 bins (~10 bit sampling)

- In a real digitizer, the separation between levels is not constant, and typically will depend on frequency. Good fast A-to-Ds are expensive.

- Oversampling reduces the quantization error for a finite number of quantizing levels:
  - E.g. sampling at 4 times the Nyquist rate, allows averaging of 4 successive samples of the signal, and hence improves the resolution of the average sample by a factor of 2.
  - This emulates using a digitizer with 2x finer resolution

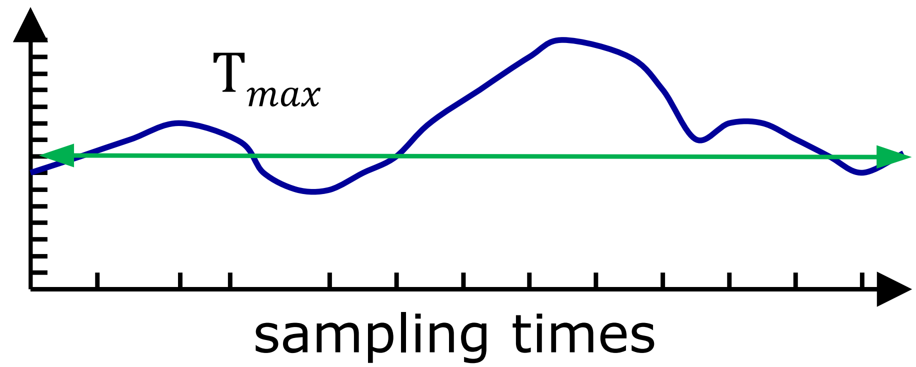
- If the signal contains random noise, then faster sampling $\Rightarrow$ taking more measurements, and so averaging of N multiple samples improves the S/N by a factor of $\sqrt{N}$.

Averaging oversampled data



How long should we sample our function for?

- Nyquist tells us that we need to sample twice per cycle.



- This means that the lower the frequency we are interested in the longer we have to sample the signal.
- A consequence of this is that our spectral resolution will be given by $\Delta f \sim 1/T_{max}$

- Key issues are:
  - If you want to be sensitive to long-period variations you have to wait a long time to see them.
  - High spectral resolution is not related to how fast the signal is sampled.

Summary so far

- Sampling:
  - Rationale.
  - Fundamental issues to do with:
    - Sampling.
    - Quantisation.
    - Spectral resolution.
  - Restoration of signals from samples.
  - Benefits of oversampling.
  - FT and LT
  - Autocorrelation and power spectrum