

Detection of Light

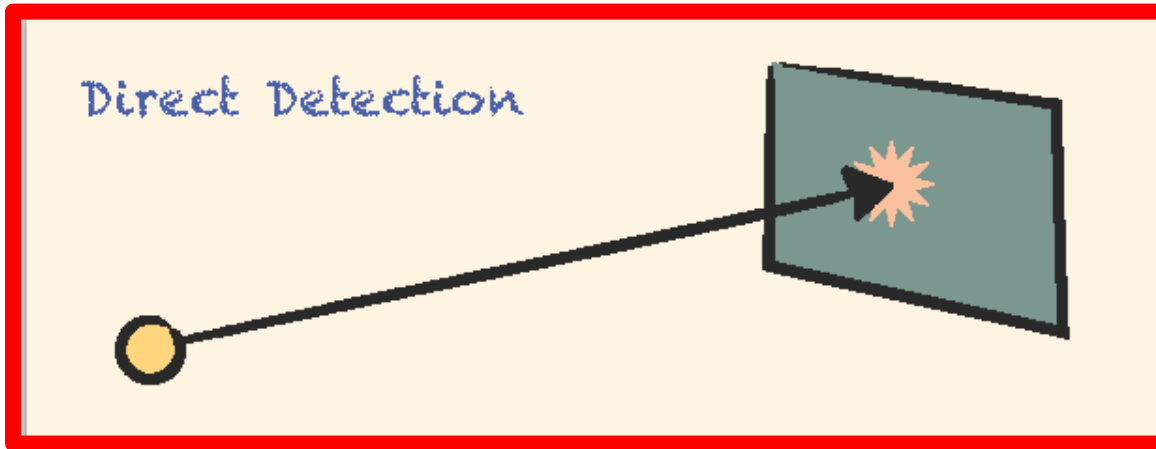


XIII. Principle of Heterodyne Receivers
XIV. Mixers for Heterodyne Detectors
XV. Performance of Heterodyne Det.

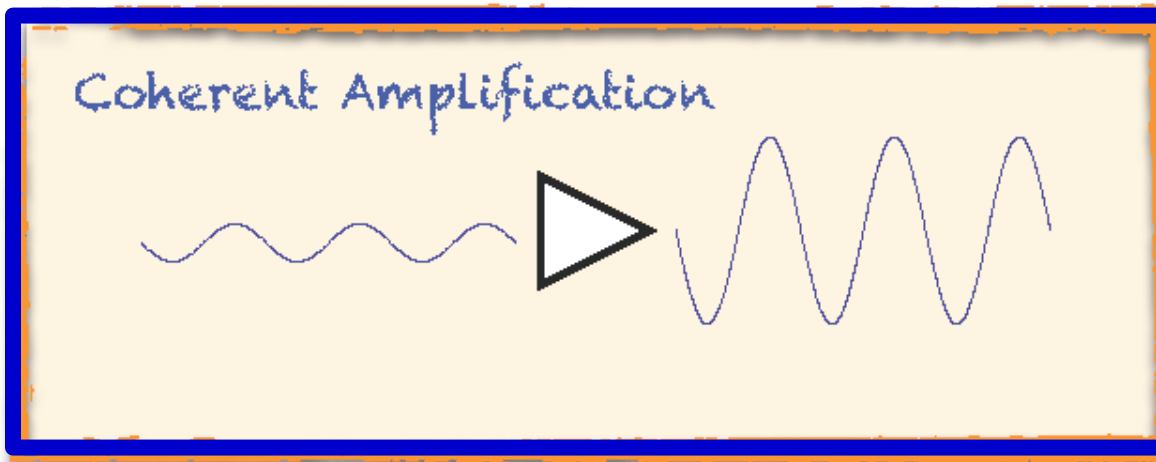
This lecture course follows the textbook "Detection of Light" by George Rieke, Cambridge University Press

Two Fundamental Principles of Detection

Respond to individual photon energy



Photons



Waves

Respond to electrical field strength and preserve phase

Principle of Heterodyne Detectors

Problem with “conventional” Detection of Waves rather than Particles or heat

Consider a detector that responds linearly within a response time τ to measure the resulting field:

If $\tau \gg f$ the **output will be zero** on average

There may not be any detectors which have short enough response times τ

Wavelength	Frequency (GHz)	Frequency (THz)	Response time (s)
10 metres	0.03	0.00003	3.33E-08
1 metre	0.3	0.0003	3.33E-09
10cm	3	0.003	3.33E-10
1cm	30	0.03	3.33E-11
1mm	300	0.3	3.33E-12
0.1mm	3000	3	3.33E-13
10 microns	30000	30	3.33E-14
1 micron	300000	300	3.33E-15
0.1 microns	3000000	3000	3.33E-16

Heterodyne Receivers

Heterodyning is a radio signal processing technique invented in 1901 by Canadian inventor-engineer **Reginald Fessenden**



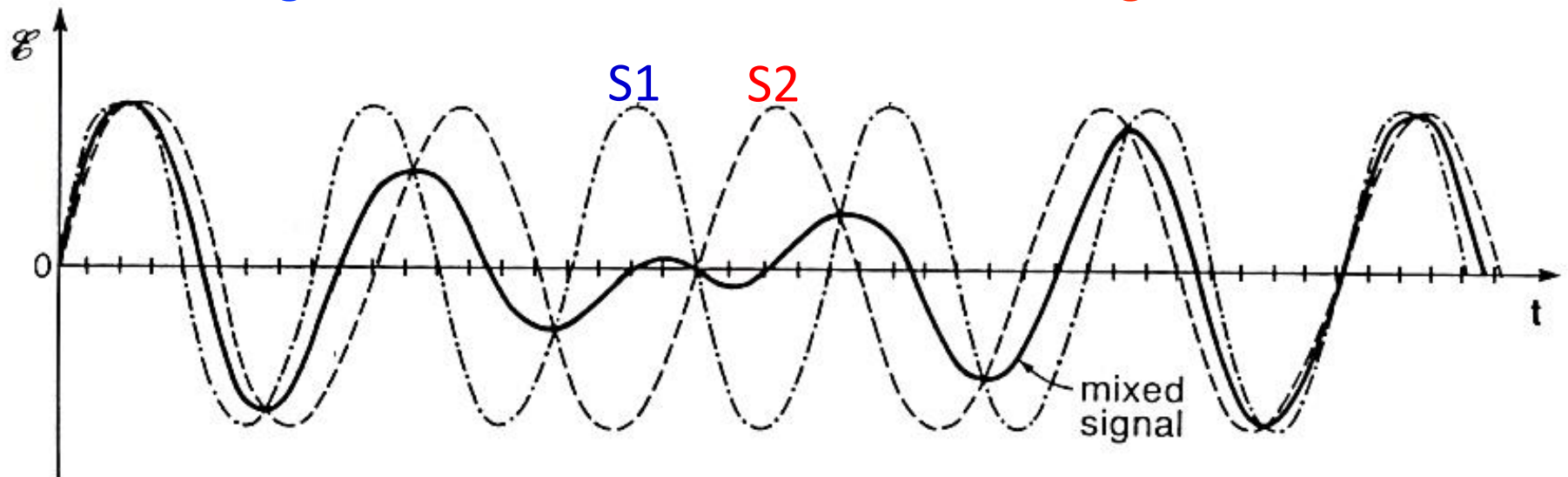
Reginald Aubrey Fessenden
(1866 – 1932)

Heterodyning = new frequencies are created by combining or mixing two frequencies

Mixing

Directly mix incoming light with coherent light of *nearly* the same frequency:

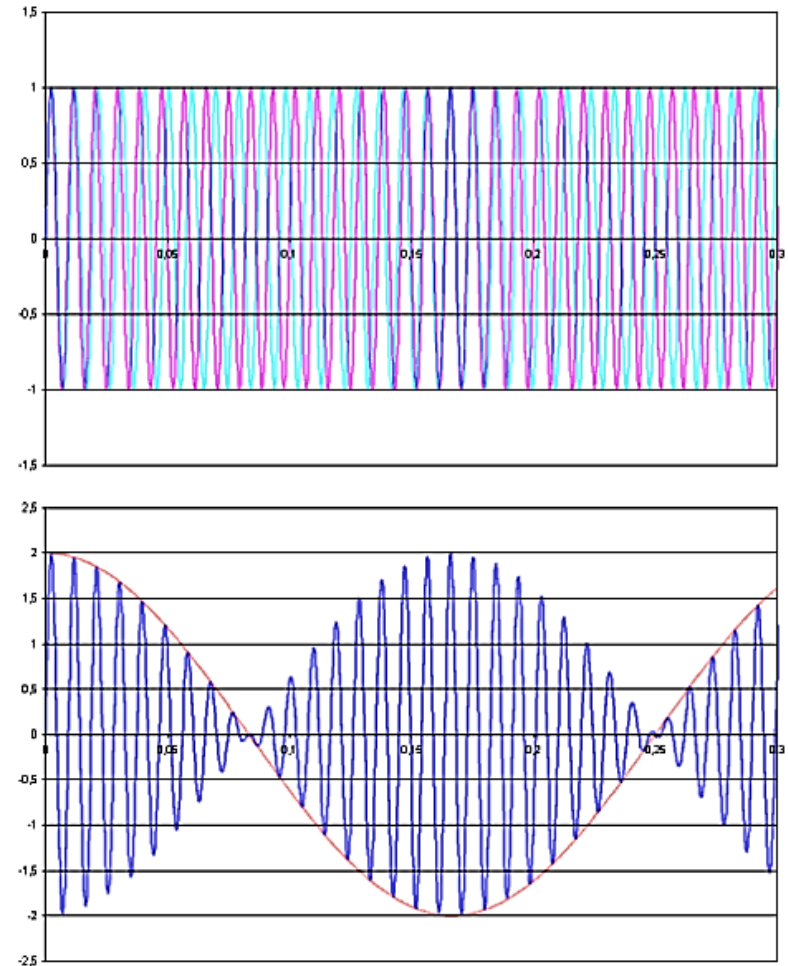
Signal **S1** is mixed with a **local oscillating field S2**



The Intermediate Frequency (IF)

The IF is the “beat frequency” and is the difference between the **local oscillator** and the **signal frequency**

The resultant mix produces a **lower frequency envelope that we can measure**.



By mixing two frequencies – of which one is known in amplitude and frequency – we can **measure an unknown frequency which is much higher than the response time of our detector!**

Why we can measure the Envelope

(1) Take **two waves**: $V_{LO} \sin(\omega_{LO}t)$ and $V_S \sin(\omega_S t)$

(2) **Add them together** and square for Power:

$$V(t)^2 = [V_{LO} \sin(\omega_{LO}t) + V_S \sin(\omega_S t)]^2$$

$$V(t)^2 = V_{LO}^2 \sin^2(\omega_{LO}t) + V_S^2 \sin^2(\omega_S t) + 2V_{LO}V_S \sin(\omega_{LO}t) \sin(\omega_S t)$$

(3) Use the **trigonometric identity**:

$$\sin A \sin B \equiv \frac{1}{2} [\cos(A - B) - \cos(A + B)]$$

...and we get:

$$V(t)^2 = V_{LO}^2 \sin^2(\omega_{LO}t) + V_S^2 \sin^2(\omega_S t) + V_{LO}V_S [\cos(\omega_S - \omega_{LO})t - \cos(\omega_S + \omega_{LO})t]$$

Through a **low pass filter**, **these terms** average to zero

DIFFERENCE **SUM** frequency

Remaining term: $Power(t) \propto V_{LO}V_S \cos(\omega_S - \omega_{LO})t$

Advantages and Applications

Advantages:

- **Direct encoding** of the spectrum of the incoming signal over a given wavelength range
- **Recording of phase** allows for interferometers - very long baseline interferometry (VLBI)
- Signals are down-converted to frequencies where **low noise electronics** can be used

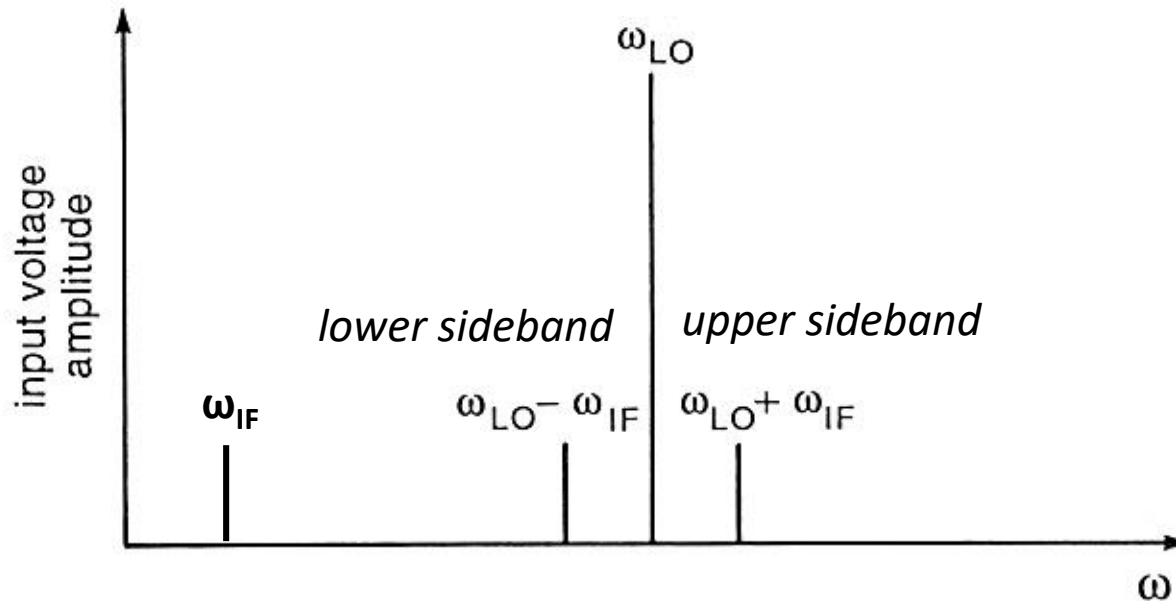
Applications:

- Very common for **sub-mm** receivers, where bolometers provide photometry and heterodyne techniques provide **efficient spectroscopy**
- Possible in **optical/IR** but cannot make large imaging arrays

Sidebands

There are always *two* Sidebands (1)

The mixed signal is amplitude modulated at the intermediate frequency $\omega_{IF} = |\omega_S - \omega_{LO}|$.



But the mixer produces the **same result** at ω_{IF} no matter if $\omega_S > \omega_{LO}$ or $\omega_S < \omega_{LO}$.

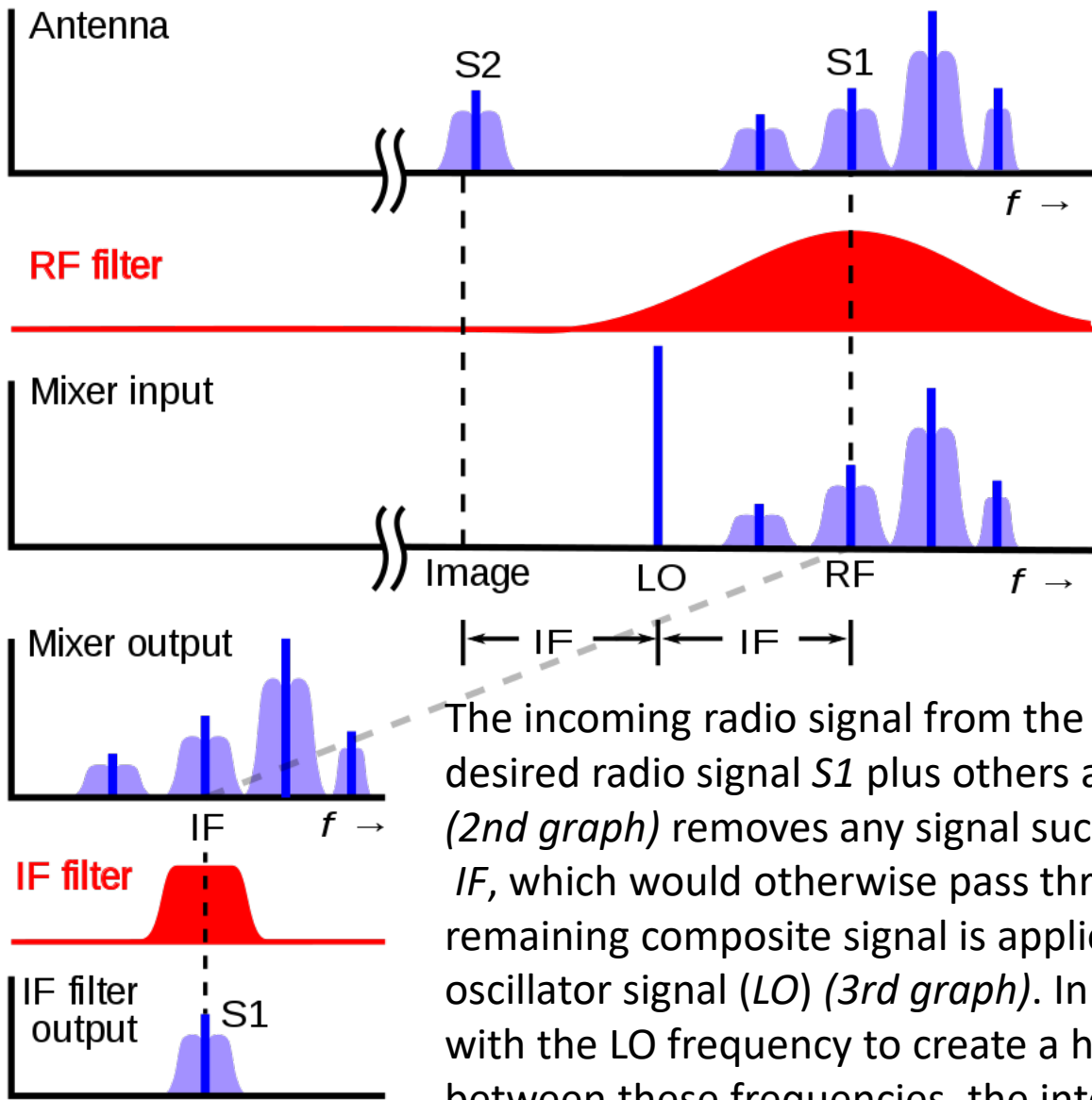
There are always *two* Sidebands (2)

Same result at ω_{IF} for both $\omega_S > \omega_{LO}$ and $\omega_S < \omega_{LO}$ – is this a problem?

- For **continuum sources**: not really, if they vary slowly with wavelength.
- For **spectral lines**: yes, big problem (if ω_{line} is not known) !

SOLUTIONS:

1. tune the mixer to remove the degeneracy
2. “Image rejection” narrowband filter in front of the receiver

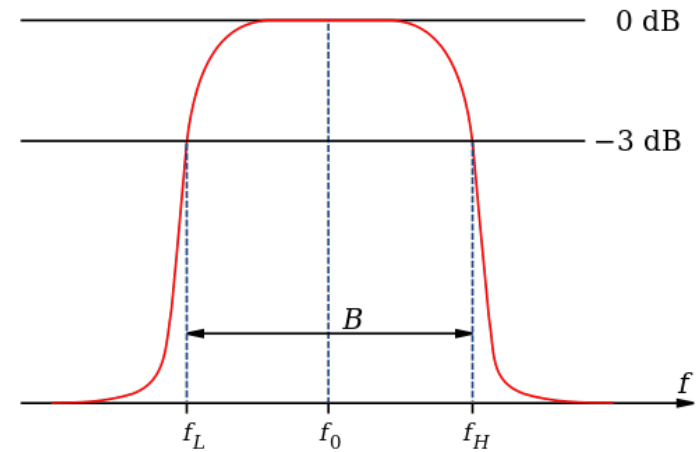


Wikipedia: The horizontal axes are frequency f . The blue graphs show the voltages of the radio signals at various points in the circuit. The red graphs show the transfer functions of the filters in the circuit; the thickness of the red bands shows the fraction of signal from the previous graph that passes through the filter at each frequency.

The incoming radio signal from the antenna (*top graph*) consists of the desired radio signal $S1$ plus others at different frequencies. The RF filter (*2nd graph*) removes any signal such as $S2$ at the image frequency $LO - IF$, which would otherwise pass through the IF filter and interfere. The remaining composite signal is applied to the mixer along with a local oscillator signal (LO) (*3rd graph*). In the mixer the signal $S1$ combines with the LO frequency to create a heterodyne at the difference between these frequencies, the intermediate frequency (IF), at the mixer output (*4th graph*). This passes through the IF bandpass filter (*5th graph*) is amplified and demodulated (demodulation is not shown). The unwanted signals create heterodynes at other frequencies (*4th graph*), which are filtered out by the IF filter.

Bandwidth

Bandwidth Basics

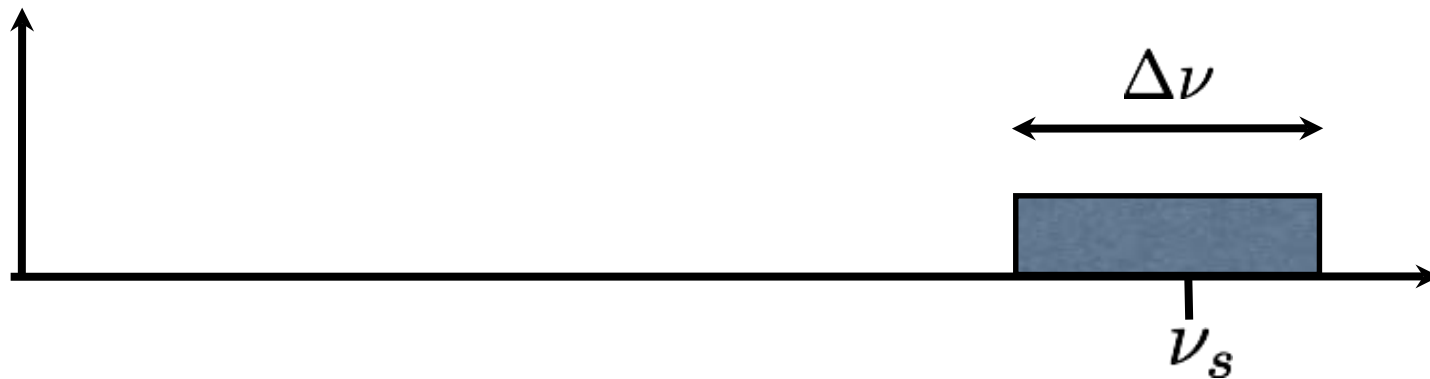


from Wikipedia:

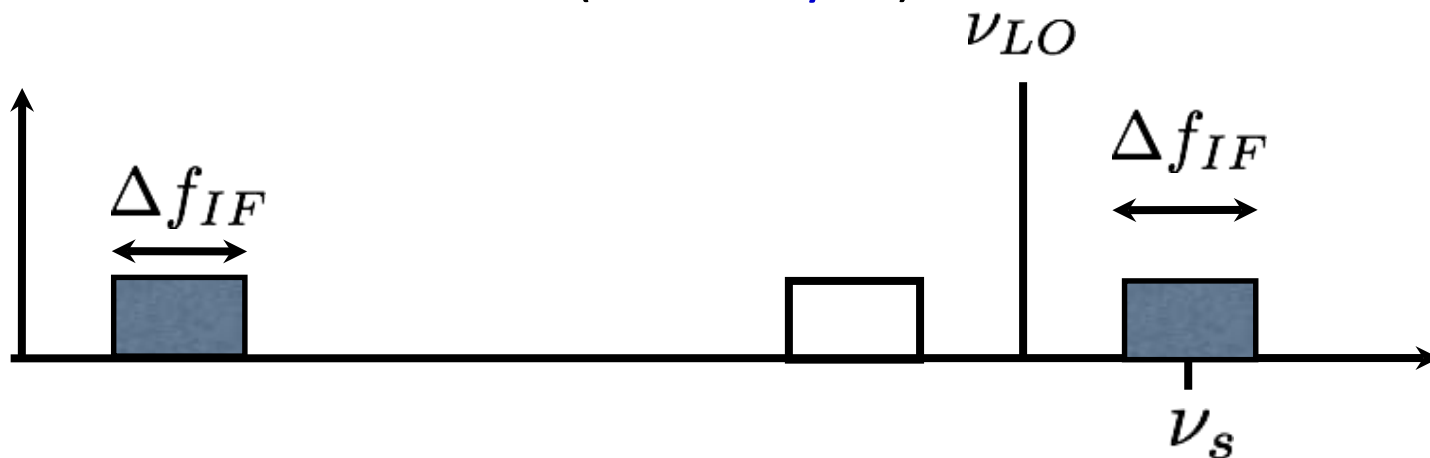
- Bandwidth is the **difference between high and low frequencies f_H and f_L** in a continuous set of frequencies.
- Bandwidth is typically measured in **Hertz [s^{-1}]**
- **any band of a given width can carry the same amount of information**, regardless of where that band is located in the frequency spectrum
- The equivalent to spectral resolution⁻¹, $R^{-1} = \Delta\lambda / \lambda$, would be the **percent bandwidth $(f_H - f_L) / f_C$** , which can be 200% at max.

Comparison of Receiver Technologies

INCOHERENT receivers (bolometer, photoconductor)



COHERENT receivers (heterodyne)



The intermediate Frequency Bandwidth Δ_{IF}

- The bandwidth of even the best photodiode mixers is usually **small compared to the signal frequency**, typically one part in 100 or 1000
- Heterodyne receivers operating at short wavelengths have **poor S/N on continuum sources**
- **...so their best use is spectral line measurement at extremely high resolutions**
- If the bandwidth Δ_{IF} is sufficiently wide the *IF* output can be sent to a set of **parallel narrowband filters**
- The **time response** of a heterodyne receiver is $1/f_{IF}$ and can be as short as a few nanoseconds.

The IF Bandwidth Δ_{IF} (2)

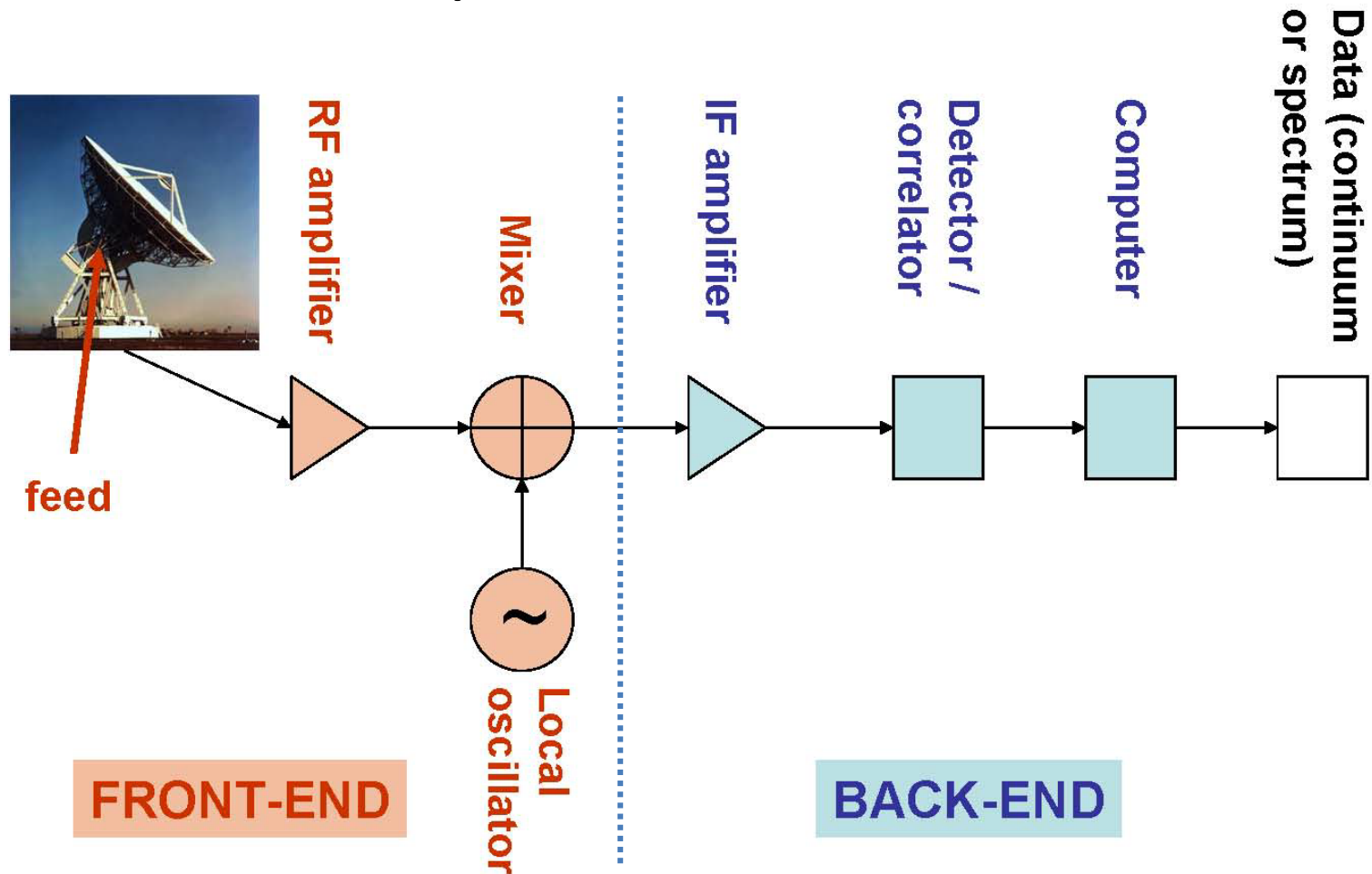
- The IF bandwidth Δf_{IF} depends on:
 1. frequency response of the mixer
 2. signal amplifier
 3. signal filter
- The Δf_{IF} provided by photoconductor mixers in the radio/sub-mm are usually **narrow** $\Delta f_{IF} < \text{few} \times 10^9$ Hz – limited by the carrier recombination time:
 - for Ge: $\Delta f_{IF} < 10^8$ Hz
 - for InSb (hot electron bolometers): $\Delta f_{IF} < 10^6$ Hz
- Δf_{IF} is even **narrower in the infrared**. Example: mixer at $10\mu\text{m}$
 $\rightarrow \nu = 3 \cdot 10^{13}$ Hz, $\Delta f_{IF} \sim 10^9$ Hz \rightarrow bandwidth is only 0.01% of λ

System Components

I. Overview

Setup of a Coherent Detection System

I. Radio/Sub-mm Receivers

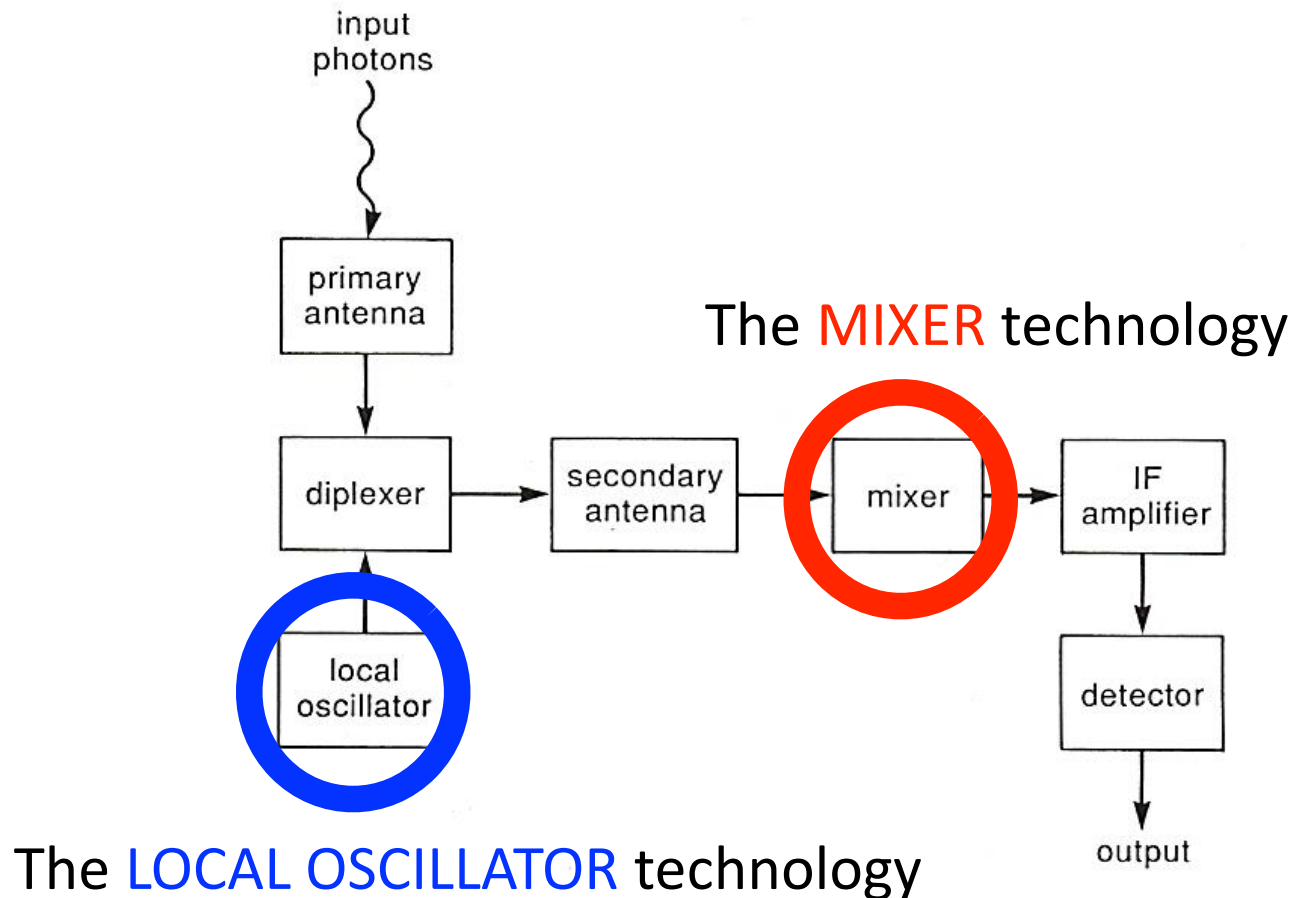


taken from Sandor Frey's Summer School presentation at <http://www.vlti.org/events/assets/4/documents/RadiolInterferometry-Frey.pdf>

Setup of a Coherent Detection System

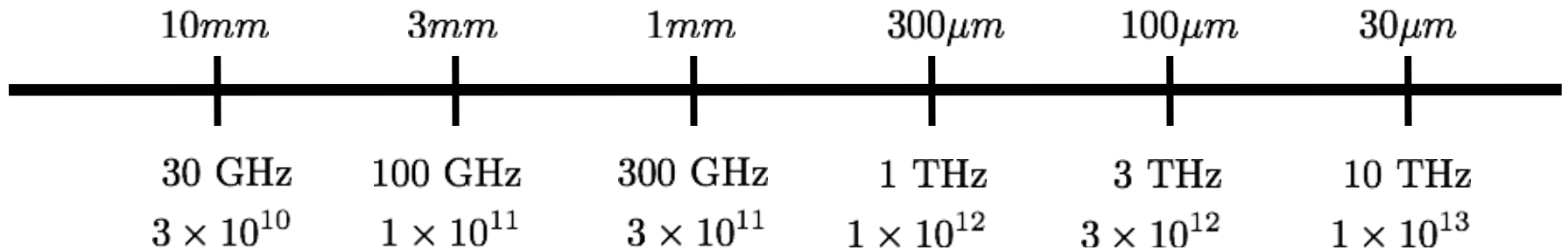
II. Visible/Infrared Heterodyne Detection

The same principles from IR heterodyne apply to sub-mm, apart from...:



Differences in LO Technology

The LO power can be fed to the mixer via a second waveguide or from a diplexer:



Low frequencies, use an
electronic LO

+ easily tunable in frequency

- low output power at high
frequencies

High frequencies, use a
continuous wave laser

+ high output power

- discrete frequencies

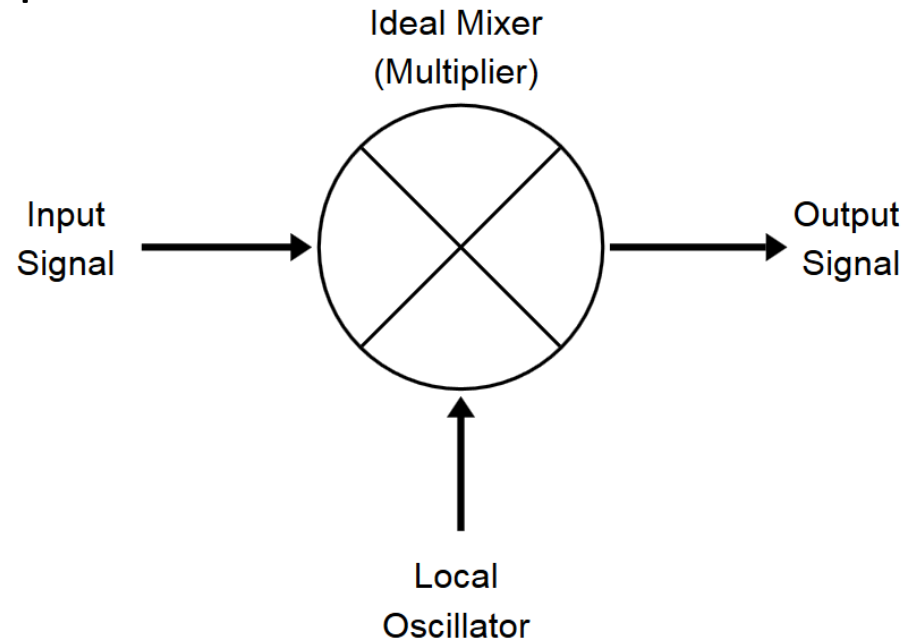
System Components

II. Mixers

Mixer Basics

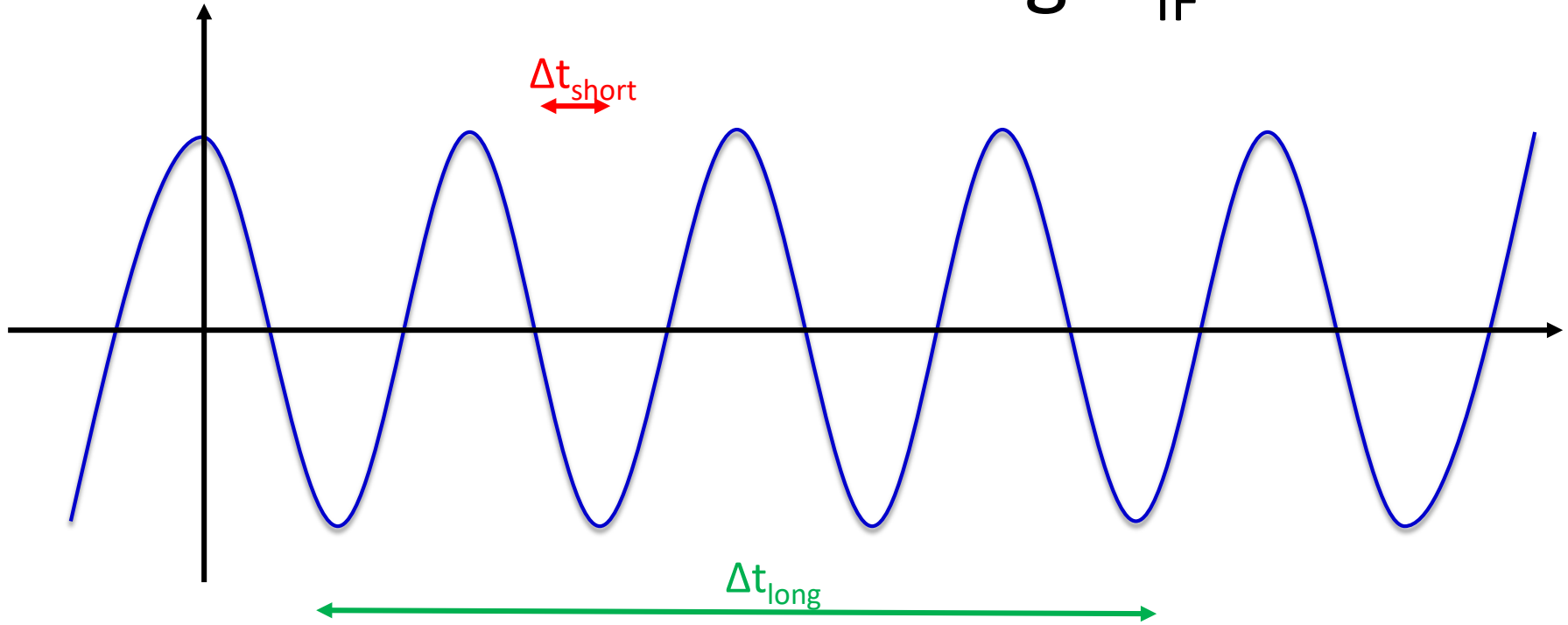
Wikipedia again:

a mixer is a **nonlinear electrical circuit** that creates new frequencies from two signals applied to it.



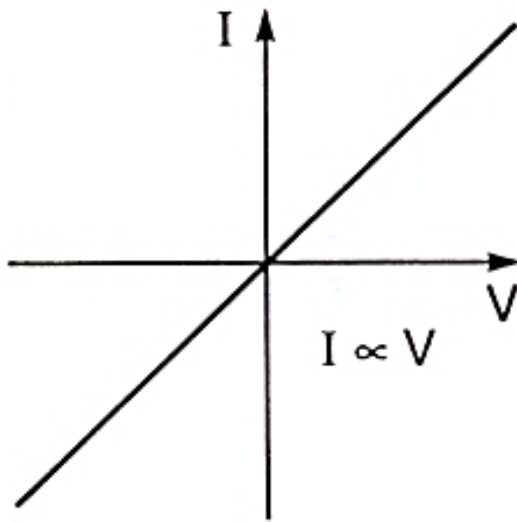
Why a nonlinear electrical circuit ?

Problem measuring ω_{IF}

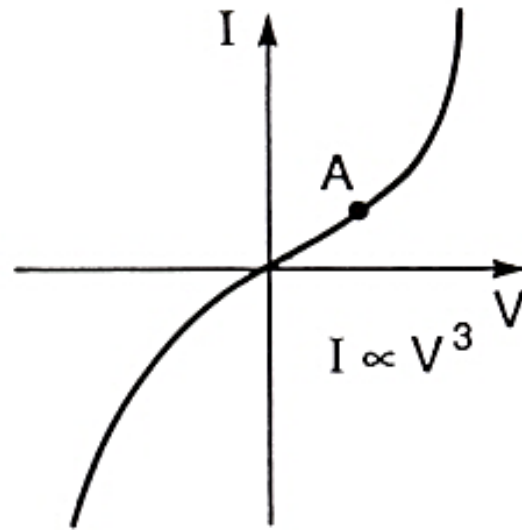


- With a **long time constant**, the outer envelope is **symmetric** about zero and averages out to zero.
- With **short time constant**, you could trace the signal at **high frequency** but a device with linear response will also give zero response.

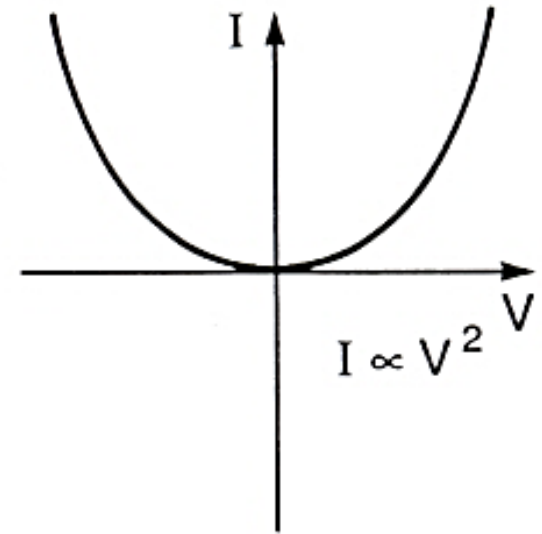
Linear and non-linear Devices



(a) linear



(b) cubic



(c) quadratic

- As shown on the previous slide, if the mixer is **linear (a)** then the conversion efficiency is ZERO
- Even if it's an **odd function** of voltage about the origin **(b)** the conversion efficiency is zero (but biased above zero at A can work)
- The quadratic case **(c)** is called a **square law device**.

Mixer Math

Generally, a useful mixer has an I-V curve that can be approximated by a Taylor series around the operating voltage V_0 :

$$I(V) = I(V_0) + \left(\frac{dI}{dV}\right)_{V=V_0} dV + \frac{1}{2!} \left(\frac{d^2I}{dV^2}\right)_{V=V_0} dV^2 + \frac{1}{3!} \left(\frac{d^3I}{dV^3}\right)_{V=V_0} dV^3 + \frac{1}{4!} \left(\frac{d^4I}{dV^4}\right)_{V=V_0} dV^4 + \dots$$

DC Voltage

Zero response

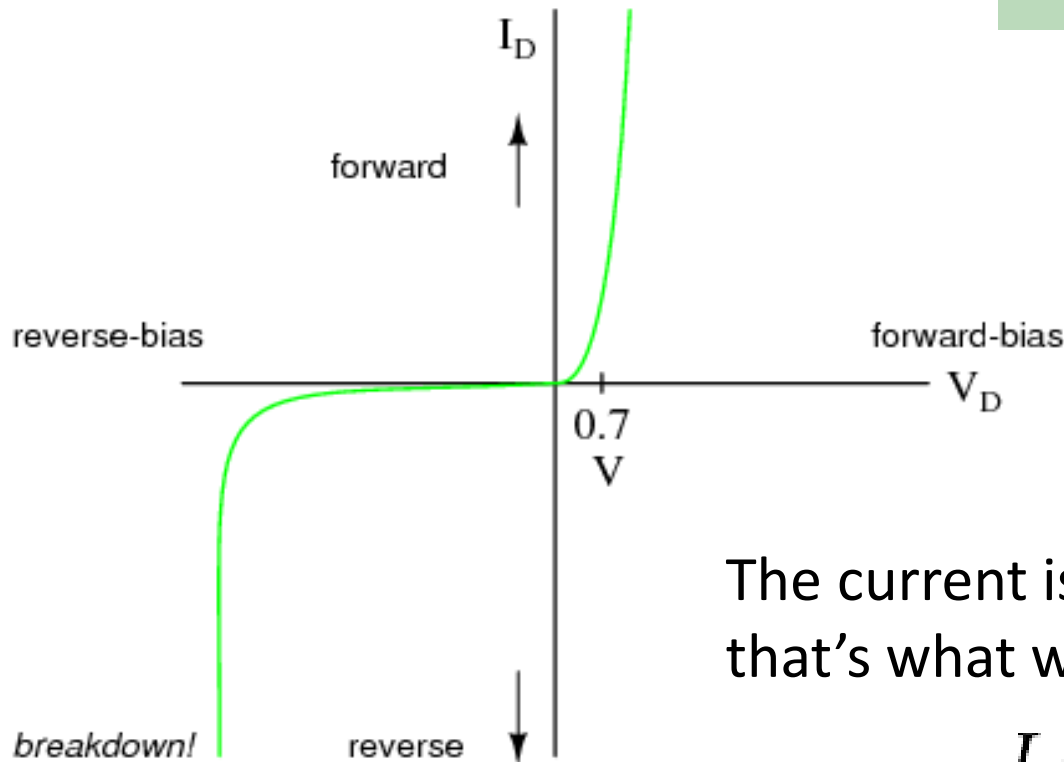
Square law mixer

Negligible if $dV = V - V_0$ is small

A Diode as a Mixer

Let's try a diode as a mixer $I = I_0(e^{qV_B/kT} - 1)$, which can be

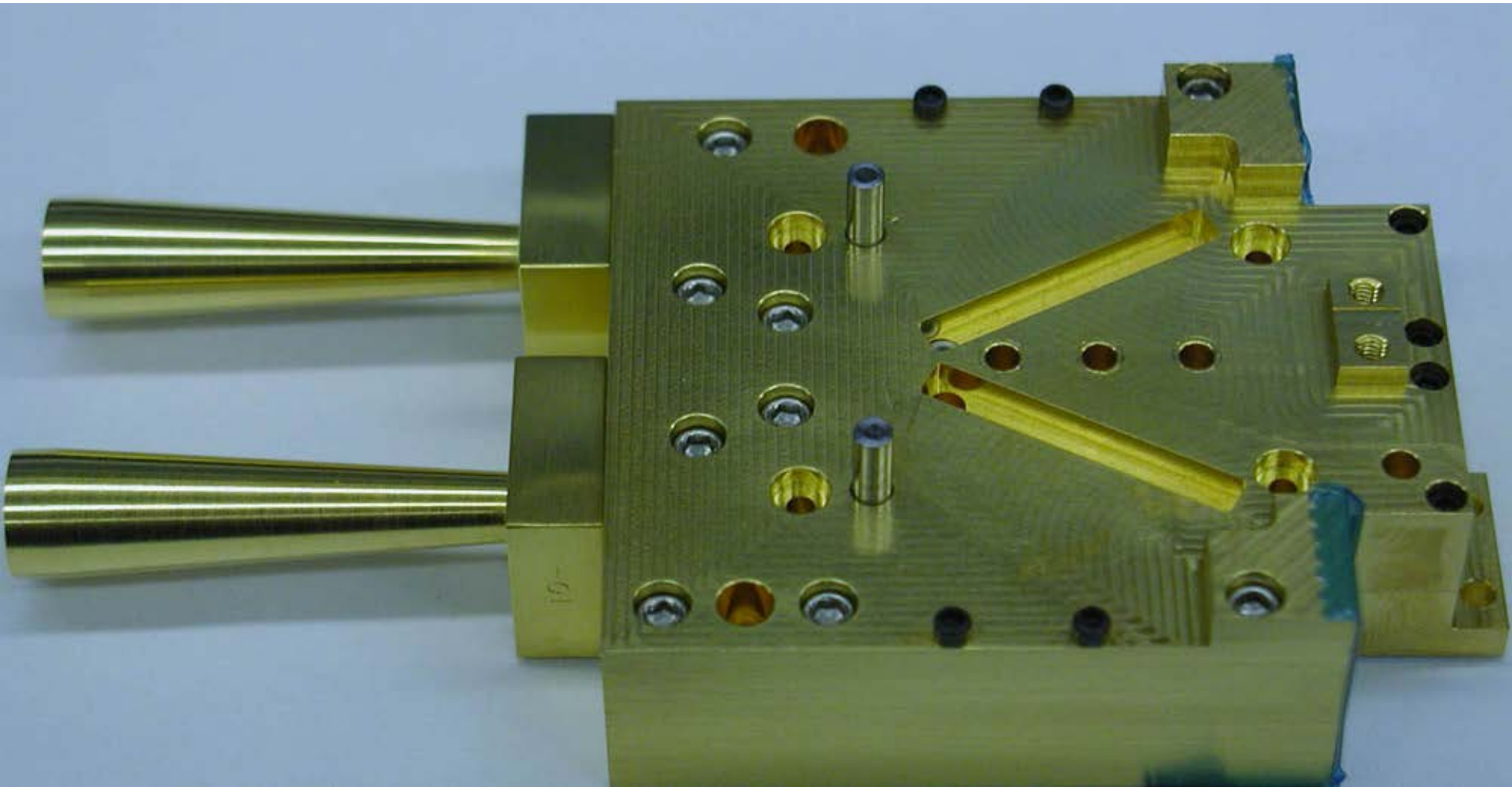
expanded as: $e^{qV_B/kT} - 1 \approx \frac{qV_B}{kT} + \frac{1}{2} \left(\frac{qV_B}{kT} \right)^2 + \dots$



The current is proportional to (voltage)² – that's what we want:

$$I \propto V^2 \propto E^2 \propto P$$

Example: 230 GHz Balanced Mixer

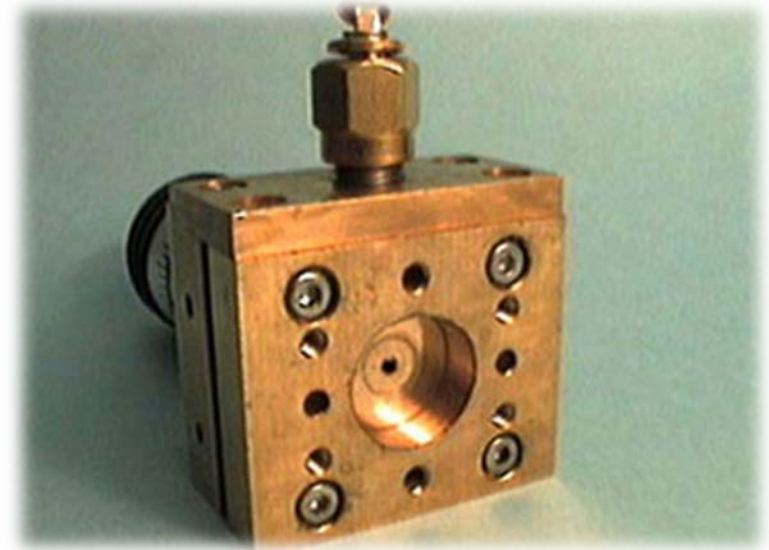


Mixer block hardware of the 180-280 GHz Balanced Mixer

<http://www.submm.caltech.edu/cso/receivers/>

More Example: Mixers

SIS front-end receiver for balloon heterodyne receiver TELIS



2.5THz Schottky diode mixer

183 GHz fixed-tuned sub-harmonic mixer



560 GHz micro-machined sub-harmonic mixer

http://www.sstd.rl.ac.uk/mmt/components_mixers.php

Mixer Choices for Wavelengths $\lambda > 40\mu\text{m}$

Fast photon detectors *do not* exist for wavelengths $\lambda > 40\mu\text{m}$

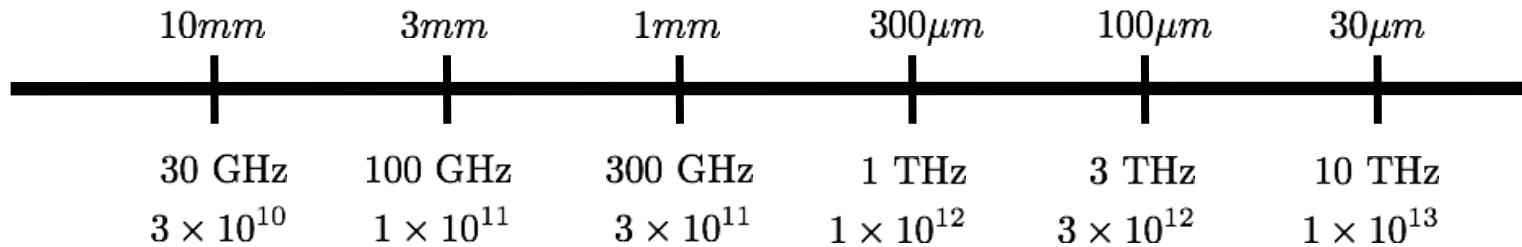
PHOTODIODE mixers have a frequency response limited to less than 1 GHz due to the recombination time of the charge carriers that have crossed the junction!

→ Common mixer devices:

- SIS junctions
- Schottky diodes
- Hot electron bolometers (HEB)

Material	$\tau_{\text{recombination}}$
Si	100 μs
Ge	10000 μs
PbS	20 μs
InSb	0.1 μs
GaAs	1 μs
InP	~ 1 μs

Mixer Technologies



SIS Pb

150GHz bandwidth

SIS NbTiN

SIS, Schottky diodes and HEB all become less effective above 1THz

Schottky

HEB

150GHz bandwidth

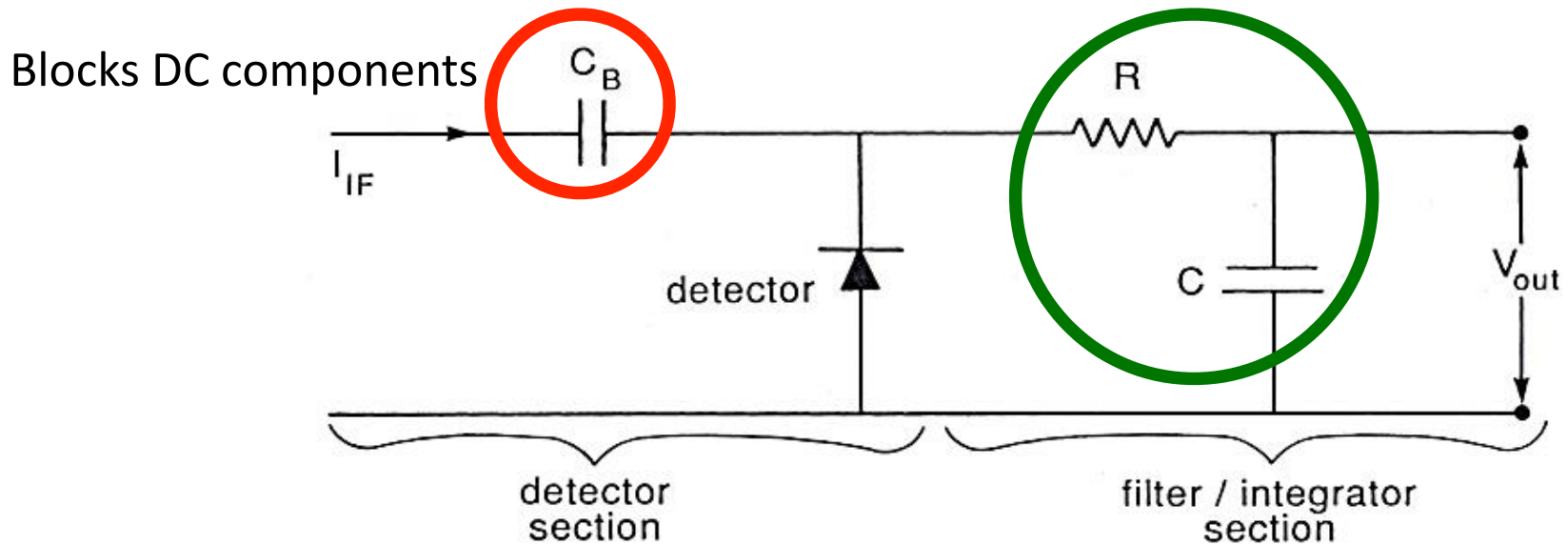
Superconducting HEB

System Components

III. Detectors

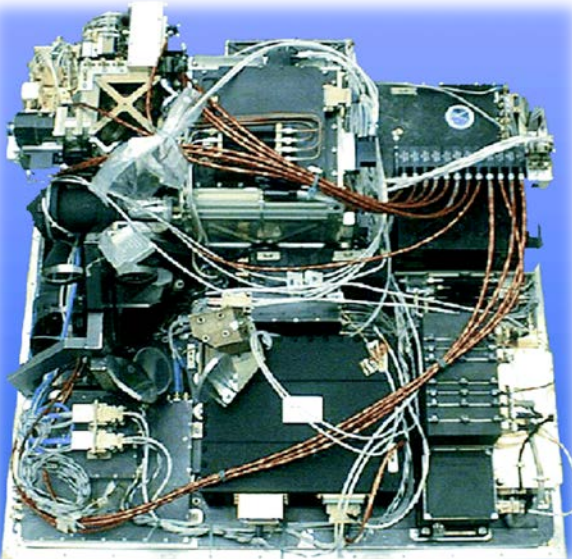
Tasks of a Detector Stage

The detector sends the signal through a low pass filter so that the output is a smooth, 'slowly' varying output.



Filter Banks

- If the IF signal contains important **frequency (“color”) components** it should not be smoothed directly.
- Instead, the signal can be sent to a **bank of narrow-band electronic filters**, operating in parallel – with a smoothing detector for each filter output.
- Hence, the **filter bank can provide a spectrum of the source**. (A back-end spectrometer could consist of several filters tuned to different frequencies with detectors on their outputs.)



This spectral multiplexing is one of the most useful features of heterodyne receivers.

Throughput

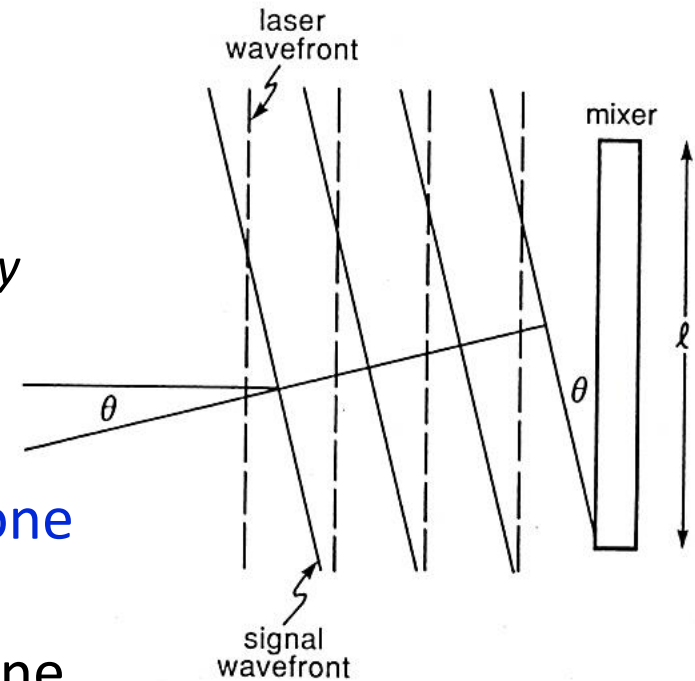
Throughput (1)

Factors that limit the throughput of a heterodyne system:

1. Only components of the signal electric field vector **parallel to the laser field** can interfere (incl. polarization!)

The signal beam will strike the mixer in a range of angles relative to the LO (or laser) beam. The conditions for interference limit the maximum angular displacement. Full cancellation occurs only when the offset $\sim \lambda$

Since the LO / laser field is polarized only **one polarization component** of the source can interfere and produce a signal so heterodyne receivers = **single-mode detectors**.



$$l \sin \theta_{\max} = \lambda \approx l \theta_{\max}$$

Throughput (2) – Antenna Theorem

The **angular diameter of the FOV** on the sky is given by:

$$\Phi \approx \frac{\lambda}{D} \approx \text{Rayleigh criterion}$$

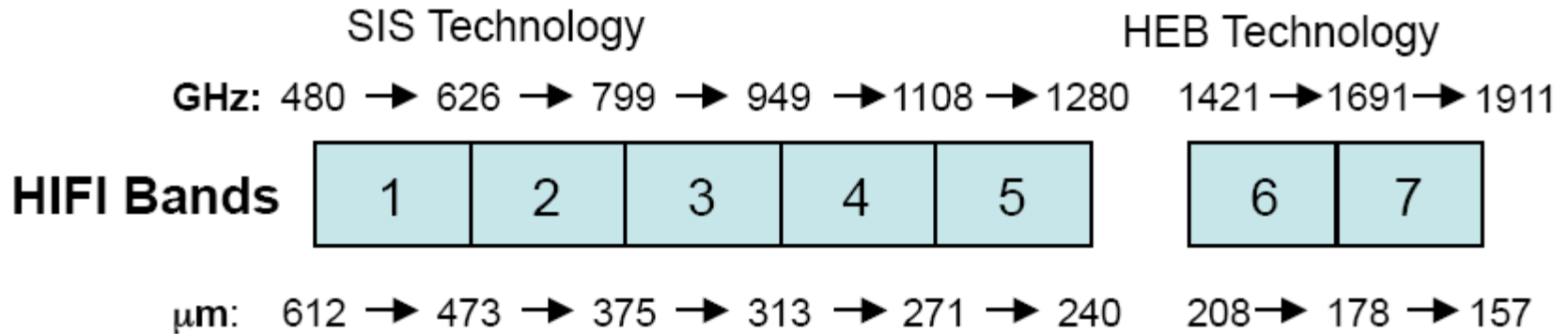
A coherent receiver should operate at the diffraction limit of the telescope. ← This is the “**Antenna theorem**” (applies to all heterodyne detectors).

- If the receiver only accepts a smaller FOV there is significant loss.
- If the receiver accepts much more it leads to a higher background and limited throughput (factor 1 above).

Example: Herschel/HIFI

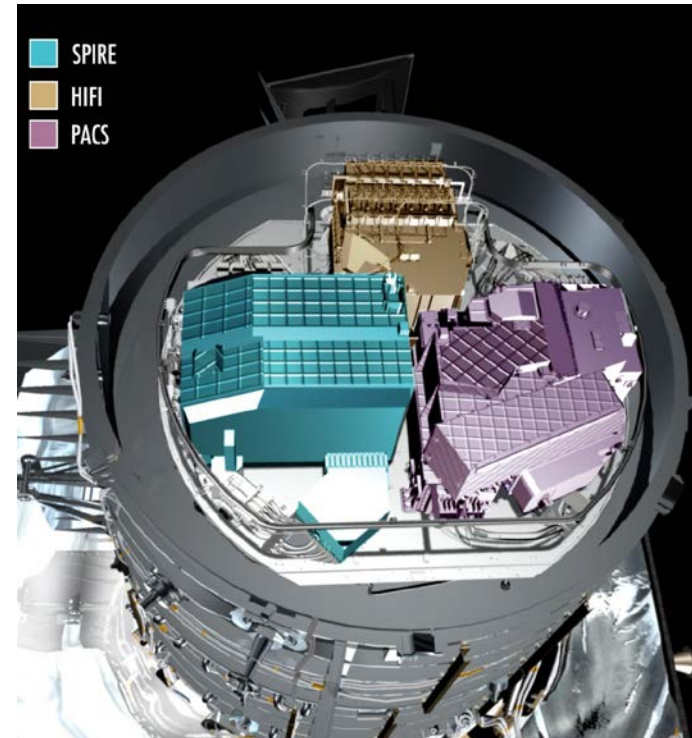
Herschel / HIFI

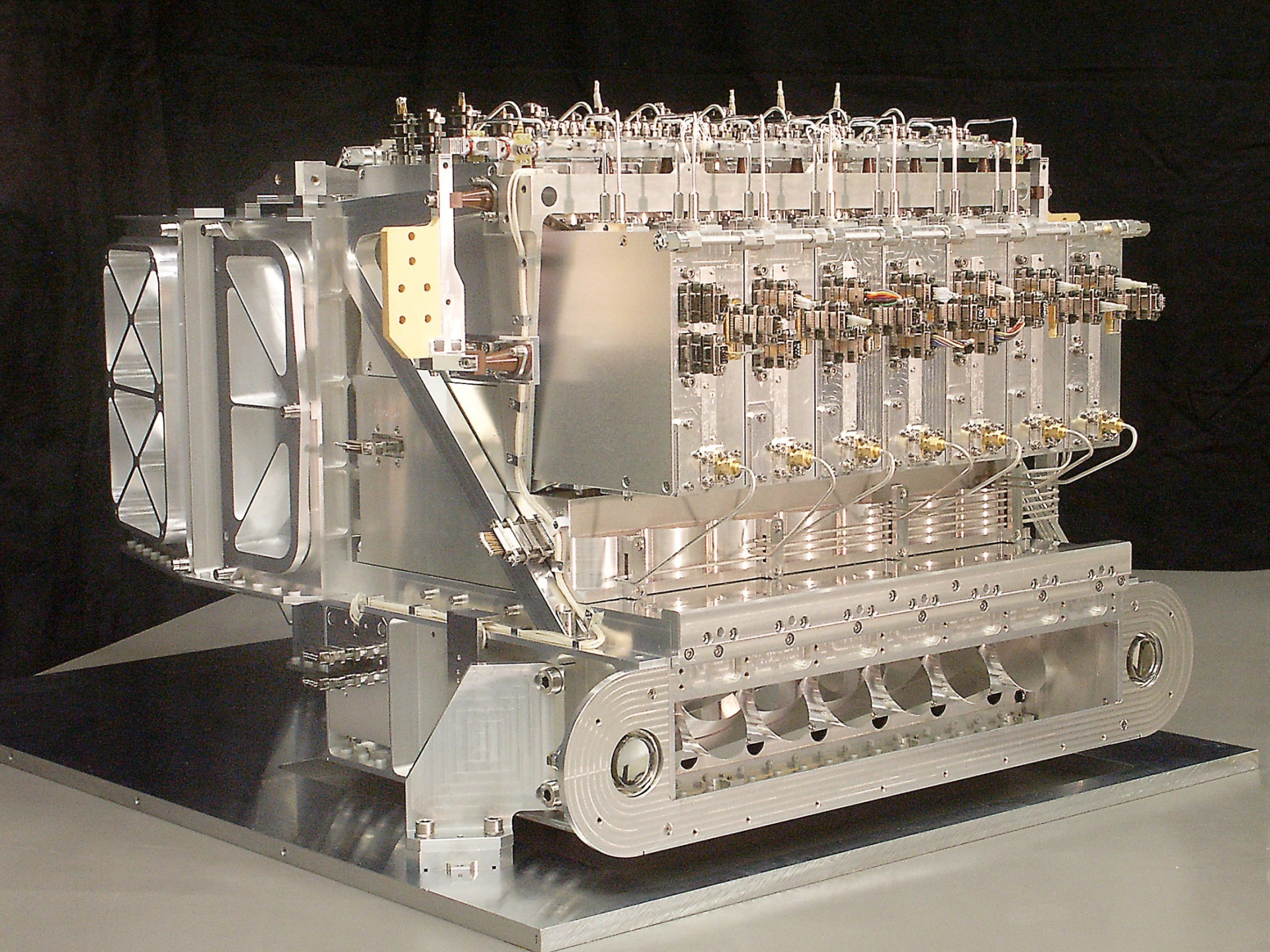
- Seven spectral bands:

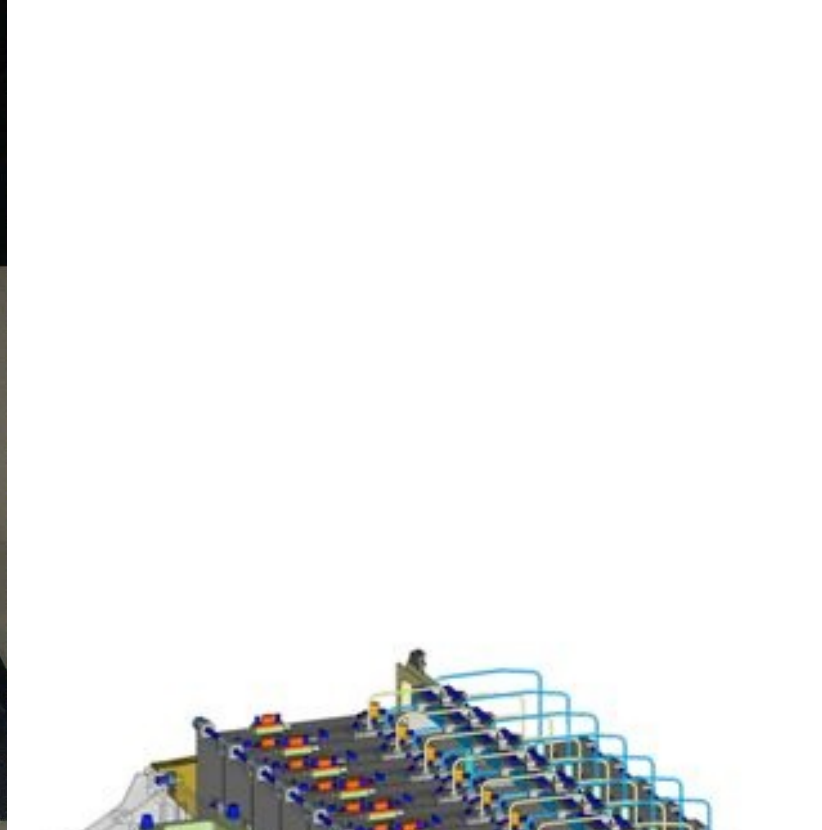
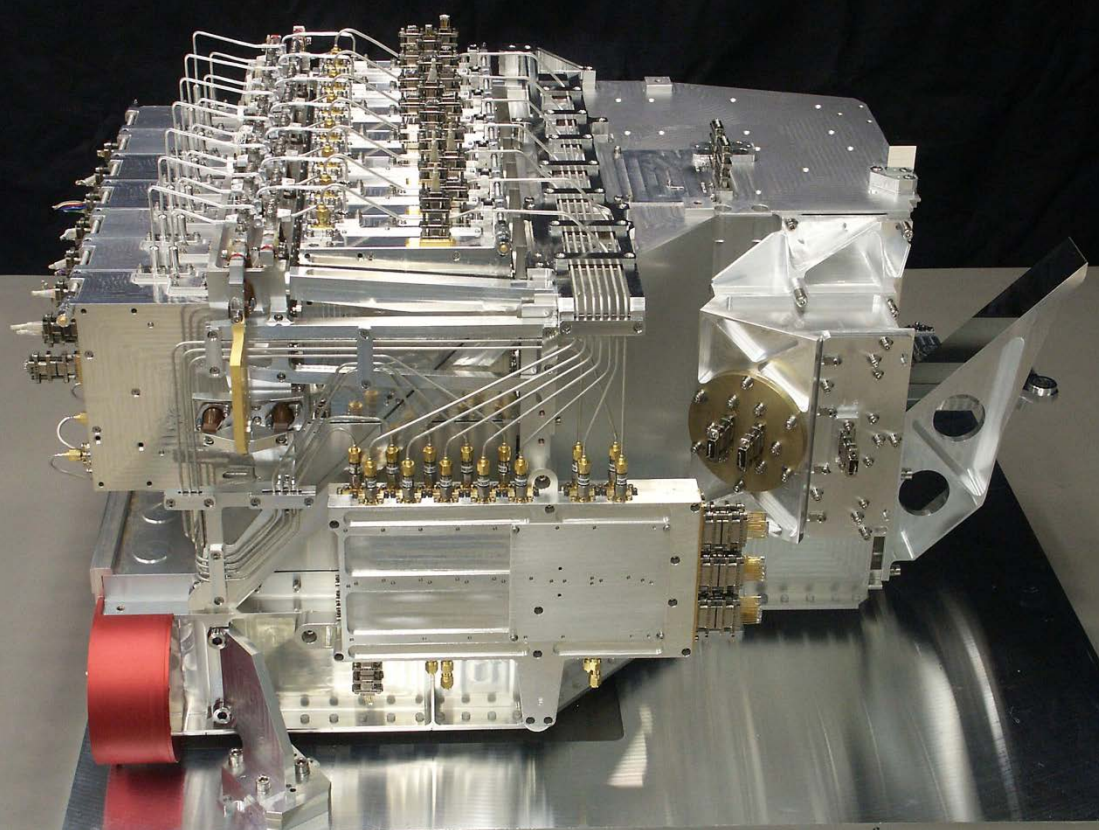


- two polarization components each
- resolving power up to 10^7
- down-converted ω_{IF} is centered at 6 GHz
- bandwidth $\Delta\omega_{IF} = 4 \text{ GHz}$

$$d\lambda = \frac{c}{\nu^2} = \frac{3 \times 10^8 \text{ m.s}^{-1}}{(1920 \times 10^9 \text{ Hz})^2} = 4 \times 10^9 \text{ Hz} = 0.33 \mu\text{m}$$

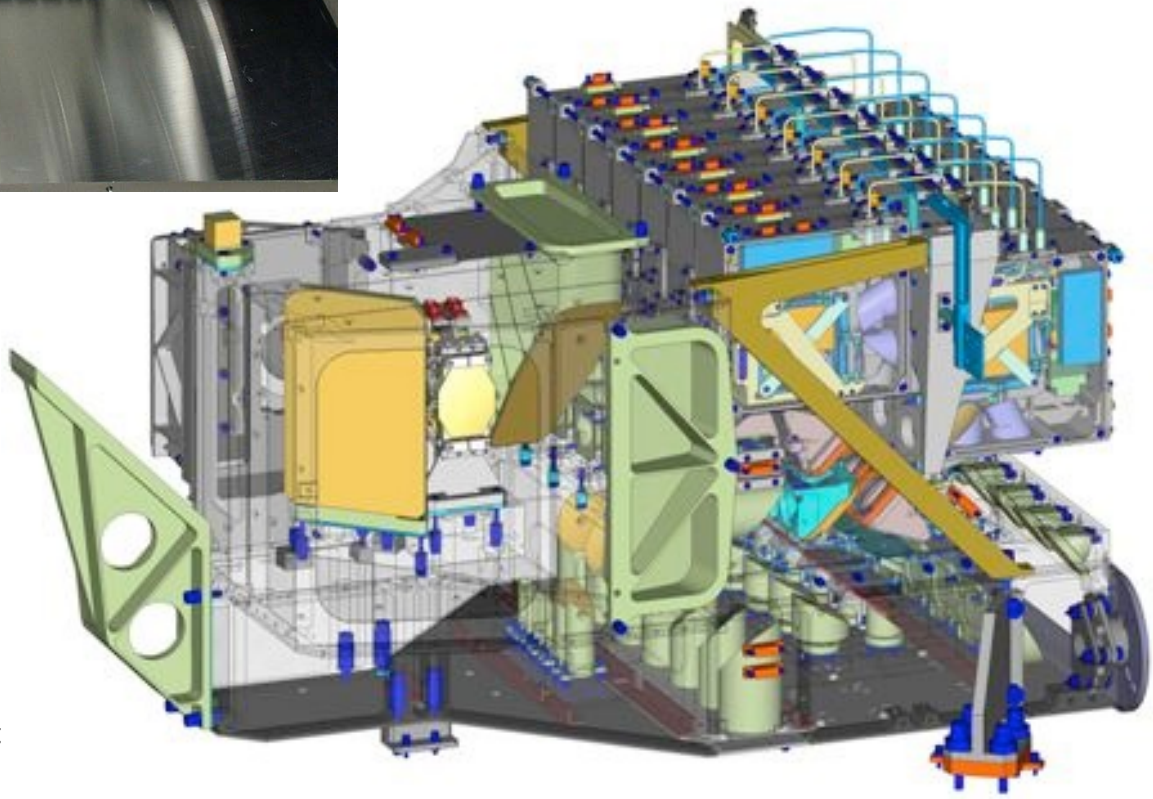






5-4-2018

Det



Signal-to-Noise

Noise Sources

There are *two types* of noise in heterodyne receivers:

1. Noise **independent** of the LO generated current I_{LO}

*Fundamental noise from the **thermal background** detected by the system*

2. Noise **dependent** on the LO generated current I_{LO} , which are **fundamental noise limits** for heterodyne receivers

*Noise **in the mixer** from the generation of charge carriers by the **LO power***

S/N, Quantum and Thermal Limit

The S/N is given by (Rieke, p. 291):

$$\left(\frac{S}{N}\right)_{IF} = \frac{\eta P_S}{h\nu\Delta_{IF} \left[\frac{a}{G} + \frac{2\eta\varepsilon}{e^{h\nu/kT_B} - 1} \right]}$$

a ≈ 1 for a photodiode mixer and a ≈ 2 for a photoconductor; G = gain, ε = BB emissivity

We can distinguish two cases:

1. **QUANTUM LIMIT** ($h\nu \gg kT_B$): G-R noise in the mixer dominates.
2. **THERMAL LIMIT** ($h\nu \ll kT_B$): noise from thermal background dominates.

The **dividing line between the two cases** is roughly at:

$$\frac{a}{G} \approx \frac{2\eta\varepsilon}{e^{h\nu/kT_B} - 1}$$

NEP

Remember: The **noise equivalent power (NEP)** is the signal that can be detected at a S/N of unity within unity frequency bandwidth Δf :

The **NEP in the quantum limit** is:

$$\text{NEP}_{ql} = \frac{P}{\left(\frac{S}{N}\right)_{out} (\Delta f)^{1/2}} = \frac{h\nu a}{\eta G} (2\Delta f_{IF})^{1/2}$$

The **NEP in the thermal limit** is:

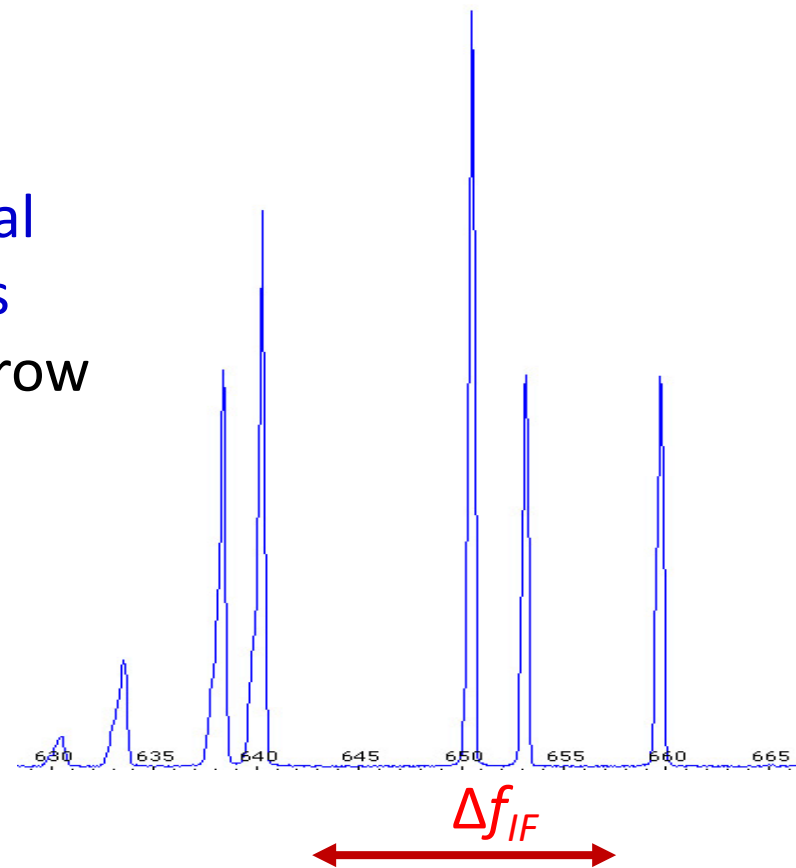
$$\text{NEP}_{th} = \frac{P}{\left(\frac{S}{N}\right)_{out} (\Delta f)^{1/2}} = \frac{2h\nu\epsilon}{e^{h\nu/kT_B} - 1} (2\Delta f_{IF})^{1/2}$$

Noise Temperature

NEP and Bandwidth

So far: $NEP_H \propto \sqrt{2\Delta f_{IF}}$ implies that the NEP decreases (i.e., the S/N increases) when narrowing down the bandwidth!

However, this is only correct if all signal power falls within an interval, which is smaller than Δf_{IF} . This is given for narrow emission lines but not for continuum sources.



The Noise Temperature

Let's define a **noise temperature** T_N such, that a matched blackbody at the receiver input at a temperature T_N produces a $S/N = 1$.

The concept of noise temperatures offers a convenient means to quantify the LO-independent components, such as **amplifier noise**.

This amplifier noise is usually **Johnson noise**: $\langle I_A^2 \rangle = \frac{4kT_N \Delta f_{IF}}{R_A}$,

where R_A and T_N are the amplifier input resistance and noise temperature, respectively.

The **lower limit** for the noise temperature is given by $T_N \approx \frac{h\nu}{k}$

- For an amplifier operating at 32 GHz $T_N \sim 1.5\text{K}$.
- For a good HEMT amplifier $T_A \sim 10\text{K}$.

Noise Temperatures at the Limits

First we estimate the noise temperature T_N in the *thermal* limit.

- If the BB emissivity $\varepsilon = 1$ then: $T_N = T_B$
- If the BB emissivity $\varepsilon < 1$ then: $T_N = \frac{h\nu}{k} \frac{1}{\ln(\varepsilon - 1 + e^{h\nu/kT_B}) - \ln \varepsilon}$

Similarly, the noise temperature in the quantum limit (double

sideband) is:
$$T_N = \frac{h\nu}{k \ln \left(1 + \frac{2G\eta}{a} \right)} \approx \frac{h\nu}{k}$$

Ideally: $G = \eta = a = 1$

How to measure Noise Temperatures

Take **two blackbody emitters** with well spaced different temperatures T_{hot} and T_{cold} .

If V is the output voltage of the receiver we can **define a “Y” factor**:

$$Y = \frac{V_{hot}}{V_{cold}}$$

...which can be measured by **alternately placing** the blackbodies over the receiver input:

$$Y = \frac{T_{hot} + T_N}{T_{cold} + T_N}$$

...and solve it for the **receiver noise temperature** T_N :

$$T_N = \frac{T_{hot} - YT_{cold}}{Y - 1}$$

Antenna or Source Temperature

Just like the noise temperature T_N describes the strength of the noise background, we can assign the **source flux** an **antenna temperature** T_S :

We get for a blackbody-type source in the Rayleigh-Jeans approximation ($h\nu \ll kT$):

$$P_S = L_\nu T_S A \Omega \Delta f_{IF} = \frac{2kT_S \nu^2}{c^2} A \Omega \Delta f_{IF} = 2\Delta f_{IF} kT_S$$

$A\Omega = \lambda^2 = \frac{c^2}{\nu^2}$
↓

where $2\Delta f_{IF}$ is the frequency bandpass for a double sideband receiver.

→ The antenna temperature is linearly related to the input flux density: $P_s \sim T_S$

Coherent Incoherent Receivers

Performance Ratio of In/coherent Receivers

The achievable S/N for a **coherent receiver** in terms of antenna and system noise temperatures is given by the **Dicke radiometer equation**:

$$\left(\frac{S}{N}\right)_{coh} \approx \frac{T_S}{T_N^{sys}} (\Delta f_{IF} \Delta t)^{1/2}$$

....so the signal to noise for an **incoherent receiver** operating at the diffraction limit is:

$$\left(\frac{S}{N}\right)_{inc} = \frac{2kT_S \Delta \nu_{inc} \Delta t^{1/2}}{NEP_{inc}}$$

Hence, the **performance ratio** between these two types of receivers is:

$$\frac{(S/N)_{coh}}{(S/N)_{inc}} = \frac{NEP_{inc} (\Delta f_{IF})^{1/2}}{2kT_N^{sys} \Delta \nu_{inc}}$$

Operation at the thermal (Background) Limit

Consider a bolometer operating at the **background limit (BLIP)** and a heterodyne receiver operating in **the thermal limit**:

$$\frac{(S/N)_{coh}}{(S/N)_{inc}} = \left[\left(\frac{1}{\eta} \right) \left(\frac{\Delta f_{IF}}{\Delta \nu_{inc}} \right) \left(\frac{h\nu}{kT_B} \right) \right]^{1/2}$$

→ the bolometer will perform better unless $\Delta f_{IF} \gg \eta \Delta \nu$

The *latter case* ($\Delta f_{IF} \gg \eta \Delta \nu$) will be given for **measurements at high spectral resolution**, much higher than the *IF* bandwidth.

Operation at the Quantum Limit

Consider a **detector noise-limited** bolometer and a heterodyne receiver operating at **the quantum limit**:

$$\frac{(S/N)_{coh}}{(S/N)_{inc}} = \frac{NEP_{inc}(\Delta f_{IF})^{1/2}}{2h\nu\Delta\nu_{inc}}$$

In the case of **narrow bandwidth and high spectral resolution**, the heterodyne receiver will outperform the bolometer

→ *heterodyne receivers are best for high spectral resolution applications in the sub-mm!*

If you keep the spectral resolution $\nu/\Delta\nu$ constant (typically given) in the above equation, then the relative **figure of merit goes as $1/\nu^2$** → transition from case favoring incoherent over coherent detectors is relatively abrupt.