Heterodyne and Homodyne Conversion

Background

The heterodyne technique goes back to the early days of radio (World War I) when amplifiers were in their infancy and all made from vacuum tubes, meaning that it was difficult to boost the amplitude of incoming signals, even at the ~1 MHz or lower carrier frequencies that were being used at that time.[1] Taken from two Greek roots, "hetero" ≡ "different", and "dyne" ≡ "force", the basic idea is to couple the incoming signal to a nonlinear "mixer" that is simultaneously driven by a "local oscillator" (LO), thereby creating a beat note at an intermediate frequency (IF) between the incoming signal and baseband. In the early days, baseband was often just the human audible range since the information being transmitted was imposed by a human voice. So the IF band was usually in the "supersonic" region between ~20 KHz (approximate upper end of audible range) and 1000 KHz, leading to the descriptor "supersonic heterodyne" or super heterodyne for short. Today, superheterodyne continues to be the descriptor, even when the incoming radiation is in the THz region, and the IF band is in the UHF or microwave regions, typically between 0.1 and 10 GHz. Arguably, super heterodyne has been one of the most valuable, if not the most valuable, developments in the history of communications, RF sensors, and more recently THz systems.

Operational Principles

The heterodyne technique generally utilizes a three-port nonlinear device called a mixer. The incoming signal X_{IN} is coupled to one port, and the local oscillator X_{LO} to a second port. The output X_{OUT} is taken from the third (intermediate frequency) port.

$$\begin{array}{c} X_{in} \rightarrow \bigotimes \rightarrow X_{out} \\ \uparrow \\ X_{Lo} \end{array}$$

The mixer design and the amplitude of the LO are chosen to impart on X_{OUT} a beat note at the intermediate frequency through product terms of the form

$$X_{out} = X_{in} X_{L0}$$

A quadratic or "square law" term in the mixer transfer function is an effective way to do this:

$$X_{out} = A X_{in}^2$$

[1] for a fascinating story behind the roots of the heterodyne technique, see Wikipedia entry http://en.wikipedia.org/wiki/Superheterodyne. This is in large part the personal story of the creative genius, Edwin Armstrong, who later invented and developed FM radio in the 1930s.

To see how the beat note is created, we assume coherent input and LO waveforms with a arbitrary phase difference, ϕ :[2]

$$X_{out}(t) = A(X_{in} + X_{LO})^2$$

$$X_{in} = B\cos(\omega_{in} t + \phi)$$

$$X_{LO} = C\cos\left(\omega_{LO} t\right) \tag{1}$$

$$X_{out}(t) = A\left[\frac{B^2}{2}\left(1+\cos(2\omega_{in}t+2\phi)\right)+\frac{C^2}{2}\left(1+\cos(2\omega_{LO}t)\right)+BC\left(\cos[(\omega_{LO}-\omega_{in})t-\phi]\right)+BC\left(\cos[(\omega_{LO}+\omega_{in})t+\phi]\right)\right]$$

The first two terms (with prefactors $B^2/2$ and $C^2/2$, respectively) represent the "self-mixing" of the LO and input waveforms, respectively, with a dc term and second harmonic for each. Generally, the output power is filtered so that neither $2\omega_{in}$ or $2\omega_{LO}$ couple out. There are two product terms with frequencies, $\omega_{LO} - \omega_{in}$, and $\omega_{LO} + \omega_{in}$ respectively. Only the former occurs at an intermediate frequency ($\omega_{IF} = |\omega_{LO} - \omega_{in}|$) between the input band and baseband (assumed to be near dc). In most heterodyne applications, especially THz ones, ω_{in} is relatively close to ω_{LO} , so that the second product term will generally not couple through the output (IF) port. So the output signal simplifies to

$$\boldsymbol{X_{out}(t) = A\left[\frac{B^2}{2} + \frac{C^2}{2} + BC\cos[(\omega_{LO} - \omega_{in})t - \phi]\right]} \tag{2}$$

$$\equiv \eta\Re\{P_{in} + P_{LO} + 2(P_{in}P_{LO})^{1/2}\cos[(\omega_{LO} - \omega_{in})t - \phi]\} \tag{3}$$

where η is the fraction of input signal and LO powers usefully absorbed. Eqns (2) and (3) illustrate three very important aspects of heterodyne conversion that are critical to its success in RF and THz systems alike: (1) X_{out} depends linearly on the input amplitude B, (2) the phase factor ϕ is preserved in the process, and (3) when C >> B as is generally the case, the LO effective amplifies the signal compared to what it would be without the heterodyne effect. The first two aspects are tantamount to "linearity", and the last one gives heterodyne a distinct advantage in sensitivity over "direct detection" (i.e., rectification) when amplifiers are not available at the input frequency.

One complexity of heterodyne conversion is that two possible input frequencies can generate the same ω_{IF} : (1) $\omega_{in} = \omega_{LO} + \omega_{IF}$, and (2) $\omega_{in} = \omega_{LO} - \omega_{IF}$. In the first case, the input is part of an "upper sideband", and in the second case it is part of a "lower sideband", both relative to ω_{LO}.. Heterodyne is usually implemented for "single-sideband" conversion, meaning that the

[2] Note: this is fundamentally different than an analog multiplier. The quadratic term of the mixer sums and then squares, whereas an ideal multiplier just takes the product.

input signal consists of an upper sideband, a lower sideband, but not both. In this case, any unwanted signals, as well as background noise, lying in the opposite sideband (commonly called the "image" band) can be down-converted along with the desired signal. Special types of mixers can be constructed that, by phase cancellation techniques, down-convert one sideband but not the other. Such "image rejection" mixers are quite common at microwave and even millimeter-wave frequencies, but have not yet appeared at THz frequencies because of their added circuit complexity.

Average Output Power

In the heterodyne case with the input power all contained in a single sideband, we compute the IF output power by taking a long time average compared to the ω_{IF} period:

$$\boldsymbol{P_{out} = D \cdot \overline{(X_{out})^2}} = D\eta^2\Re^2 \cdot 2\, P_{in}\, P_{LO} \qquad (4)$$

where D is a matching factor between the mixer and the IF impedance [the $\cos^2(\omega_{IF}t)$ term averages to ½]. For practical reasons, it is easier to combine all the mixer parameters into one quantity, the mixer conversion gain:

$$G_{mix} = 2D\eta^2\Re^2\, P_{LO}$$

so that

$$P_{out} = G_{mix}\, P_{in}$$

G_{mix} represents the fraction of absorbed incident power in one sideband that is converted to IF power. In other words, it is an easily measurable quantity that doesn't require knowing the detailed behavior of the mixer itself. Usually $G_{mix} < 1$, but at microwave frequencies there are "active" mixers that have $G_{mix} > 1$. Such mixers have not yet appeared at THz frequencies, at least not with room-temperature operation. For a mixer made of passive components, such as Schottky diodes, and assuming single-sideband conversion, it can be shown (HW Problem) that the highest possible conversion gain is $G_{mix} = 0.5$ (-3 dB). This can be understood physically from Eqns (1) whereby the input power is divided equally between the sum and difference frequencies ($\omega_{LO} - \omega_{in}$, and $\omega_{LO} + \omega_{in}$), and the sum term is generally never used.

In the special case that $\omega_{in} = \omega_{LO}$, we still get frequency down-conversion but ω_{IF} degenerates to 0 (i.e., "dc"). This is called "homodyne" conversion. In going from Eqn (3) to Eqn (4), the time-average has no effect on the dc term so an extra factor of two appears in the output power:

$$P_{out} = D\eta^2\Re^2 \cdot 4\, P_{in}\, P_{LO} = 2G_{mix}\, P_{in}$$

G_{mix} retains its single-sideband-to-IF definition. So the factor-of-two increase in power can be attributed to the coincidence of the upper- and lower-sideband that only occurs in homodyne conversion.

Signal-to-Noise Ratio

(1) Quantum-Limit

The fluctuations in X_{out} caused by quantum-mechanical variations in the self-mixing terms P_{in} and P_{LO} must be considered. This is the so-called photon "shot-noise" effect, which is very commonly seen in infrared and visible photoelectric mixers and detectors, but rarely observed in THz mixers except those operating at cryogenic temperatures. Nevertheless, we include it in the analysis since it comprises the ultimate quantum-mechanical limit on the sensitivity of any mixer. The quadratic term in the mixer transfer characteristic that creates the dc terms in Eqns (1) and (2) leads to the following expression for output (IF port) fluctuations caused by input (input or LO) fluctuations:

$$\langle(\Delta X_{out})^2\rangle = 2\Re^2 \langle(\Delta P)^2\rangle\, B_{IF}/\Delta\nu$$

Where $\Delta\nu$ is the input bandwidth and B_{IF} is the IF bandwidth. In almost all mixers, $P_{LO} >> P_{in}$ so that $\Delta P \approx \Delta P_{LO}$. Then, if the LO is a sinusoidal oscillator (generally true at THz frequencies), we have $\langle(\Delta P)^2\rangle = \eta h\nu_{LO} P_{LO}\, \Delta\nu$ from the quantum mechanical theory of a "coherent state." This leads to

$$\langle(\Delta X_{out})^2\rangle = 2\Re^2\, \eta h\nu_{LO} P_{LO} B_{IF}$$

The mean-square fluctuations in output power are

$$\langle(\Delta P_{out})^2\rangle = 2D\eta\Re^2\, h\nu_{LO} P_{LO} B_{IF} \equiv G_{mix}\, \eta\, h\nu_{LO}\, B_{IF}$$

Taking the ratios between average signal power and the noise power, we can compute the SNR for the quantum-noise-limited heterodyne and homodyne cases in the IF port just after the mixer:

Heterodyne: $$SNR = \frac{\eta^2 \cdot G_{mix} \cdot P_{in}}{G_{mix} \cdot \eta \cdot h\nu_{LO} \cdot B_{IF}} = \frac{\eta \cdot P_{in}}{h\nu_{LO} B_{IF}}$$

Homodyne: $$SNR = \frac{2\eta \cdot P_{in}}{h\nu_{LO} B_{IF}}$$

These both represent fundamental limits on the SNR of any mixer, at least one with a sinusoidal local oscillator waveform. Although a mixer does not carry out "detection" as we have learned it, one can still define a (quantum-limited) noise equivalent power (NEP) for the IF port by setting the SNR to unity:

Heterodyne: $$NEP = h\nu_{LO}\, B_{IF}/\eta$$

Or by normalizing out the IF bandwidth:

$$NEP' = h\nu_{LO}/\eta$$

This should look familiar to students familiar with photonic receivers. The same expression occurs there for a "photon-shot-noise" limited optical receiver and for the same physical reason: that even sinusoidal oscillators such as lasers exhibit a fluctuation in power associated with quantum mechanical fluctuations in the photon arrival rate. The big difference between the THz quantum limit and the photonic one is the much smaller magnitude of $h\nu$, which makes the quantum-limited NEP practically unattainable at THz frequencies, at least at room temperature.

(2) Electrical-Noise Limit

In all room-temperature THz mixers developed to date, the signal-to-noise ratio is farl lower than the values predicted above because the output fluctuations are dominated by thermal and shot noise effects in the mixer itself, and in the electronics before and after the mixer. On first thought, this might seem like a complicated situation and very specific to the mixer type and the surrounding electronics. However, we harken back to key issue discussed earlier that if the input (signal) power is much less than the LO power, as is generally the case, then the mixer behaves linearly with respect to signal amplitude and phase. And therefore, we can invoke the concepts of noise factor, noise temperature, and noise figure. We recall the general definition addressed in Notes#7, but now applied to the mixer:

$$F = \left(\frac{(S/N)_{IN}}{(S/N)_{OUT}} \right) = \frac{S/(k_B T_0 B_N)}{GS/[G(k_B T_0 + k_B T_N) B_N]}$$

where T_N becomes the mixer noise temperature and G is the single-sideband mixer conversion gain. As before, this leads to the noise factor

$$F = 1 + T_N/T_0$$

In general, we expect T_N for mixers to be rather high compared to high-quality amplifiers because (passive) mixers generally display $G < 1$, and in the THz region $G \ll 1$. To understand this better, we define an excess noise factor τ to describe the noise contributed by the mixer at the output in excess of the noise one would expect in thermal equilibrium at T_0 (according to the Nyquist generalized theorem). In thermal equilibrium, the noise should just be the input reference noise $k_B T_0 B_N$ (since the mixer is passive), so that the excess should be written $(\tau - G_{mix}) k_B T_0 B_N$. We thus the noise factor as

$$\boldsymbol{F = \frac{S/(k_B T_0 B_N)}{G_{mix} S/[G_{mix} k_B T_0 B_N + k_B(\tau - G_{mix}) T_0 B_N]} = 1 + \frac{\tau}{G_{mix}} - 1 = \frac{\tau}{G_{mix}}}$$

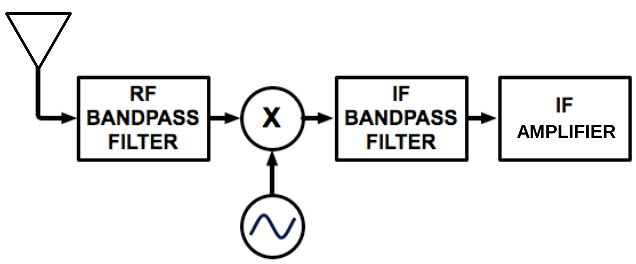
By definition, the mixer noise temperature is then

$$T_{N,mix} = (F - 1)\, T_0 = (\tau/G_{mix} - 1) T_0$$

Typically the excess noise factor is at least $\tau = 2$ for mixers since they usually generate shot noise (associated with dc current induced by the LO power attenuation) in addition to thermal noise, and G_{mix} around 1.0 THz is about 0.25 (-6 dB) at best. So a ballpark estimate of the noise factor is F = 8 (NF = 9.0 dB), and $T_{N,mix}$ is 7*290 = 2030 K. As we shall see, this is about as good as room-temperature mixers ever get at these frequencies.

Typical THz Superheterodyne Front End

In the THz region, there are still no commercially available low-noise amplifiers. So most heterodyne receivers are constructed as shown below where the only component separating the mixer from the antenna might be a bandpass filter to block infrared or visible radiation. The LO is usually at a fixed frequency, phase- or frequency-locked. The IF amp/filter is fixed and is designed to have sharp cutoffs, which allows for high rejection of all other signals except for the desired signal. By splitting the effort of amplification and high rejection, the IF filter and amplifier are individually much easier to design. A key point is that the linearity of the mixer and single-sideband operation means that there is a simple translation in frequency space between the IF passband and the THz (i.e., "RF") passband. So the THz passband is centered at ν_{LO} +/- $\nu_{IF,0}$, where $\nu_{IF,0}$ is the center of the IF passband, the + corresponds to upper-sideband operation, and the – corresponds to lower-sideband operation. Furthermore, the THx bandwidth $\Delta\nu$ is equal to the IF bandwidth B_{IF} assuming that the latter is well defined by a sharp-skirted filter.



Superheterodyne Basic Receiver

Example. Suppose we have a mixer whose single-sideband conversion gain is 0.25 (-6 dB) and whose noise figure NF = 9 dB. It is followed by an IF bandpass filter and IF amplifier whose combined noise figure at the center of the IF band is 3.0 dB. What is the overall noise figure and equivalent noise temperature at the input port of the mixer, assuming it is impedance matched at this port ? (solution worked out in class). Clue: the key to solving this problem is

$$F_{TOT} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 \cdot G_2} + \ldots\ldots \frac{F_n - 1}{\prod_{i=1}^{n-1} G_i} \quad .$$