Correspondence to:
P. T. de Boer,
p.t.deboer@utwente.nl

Sideband Asymmetry in Ionospheric Cross Modulation

P. T. de Boer[1] (iD) and **M. Vester**[2] (iD)

[1]Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, The Netherlands, [2]Nuremberg, Germany

Abstract We study the phenomenon of ionospheric cross modulation between LF/MF broadcast stations using modern reception techniques, allowing us to see details that were not seen with the analog receiving techniques from the 1940s. We show experimentally that the sidebands of the transferred modulation are often asymmetric, in amplitude and/or phase. We give qualitative and quantitative explanations for this asymmetry, based on taking contributions from outside the first Fresnel zone into account. The asymmetry is found to be strongest for longer paths, lower carrier frequency, and disturbing transmitter closer to the receiver.

1. Introduction

In January 2013, one of the authors observed amplitude modulation from a medium-wave broadcast station being transferred to an unmodulated longwave utility station (Vester, 2013). Such modulation transfer is due to ionospheric cross modulation, popularly known as the Luxembourg effect, which has been well known since the 1930s (cf. Figure 1a). But exceptionally, in this case most of the upper sideband was missing; see the spectrogram in Figure 1b.

In 2016, an amateur-radio magazine article by the other author (de Boer, 2016) about the Luxembourg effect was mistaken for an April fool's joke by some readers. This prompted him to look for a good example of the phenomenon, and in the ensuing measurements find an asymmetry in the phase of the transferred sidebands.

Through many more measurements, we found that such asymmetries occur in many combinations of transmitter and receiver locations. Furthermore, we constructed an explanation for the asymmetries. The key factor is that different parts of the ionosphere, also outside the first Fresnel zone, contribute significantly. Phase shifts due to propagation delays then cause the different sideband frequencies to effectively reach the receiver from different points of the ionosphere, resulting in the asymmetries.

This paper starts in section 2 by a literature overview and a summary of what amplitude and phase behavior would be expected based on established theory. Then, in section 3, we describe a measurement method of ionospheric cross modulation using software-defined radio (SDR). Whereas previous measurements were done using purposely transmitted test tones after the end of normal programming, our method uses cross correlation to monitor the effect all day long and for all modulation frequencies simultaneously. Results are shown in section 4. Next, in section 5 we describe our model for the phase asymmetry, which shows good agreement with the measurements. Modeling the amplitude asymmetry is much more difficult, but a qualitative understanding can be obtained, in section 6. Finally, in section 7 we put our findings in a historical perspective: why was this asymmetry not discovered much earlier?

2. Background

Ionospheric cross modulation was first mentioned in the scientific literature by Tellegen (1933), reporting that while receiving a medium-wave radio signal from Switzerland in the Netherlands, he also heard modulation from a Luxembourg radio station; after excluding local problems in his receiver, he concluded the modulation was apparently transferred in the ionosphere. This phenomenon was initially called wave interaction, popularly known as the "Luxembourg effect," while a proper scientific name is ionospheric cross modulation, as suggested by Huxley and Ratcliffe (1949).

Tellegen's observation was later found not to be the first; the phenomenon had been reported before by Butt (1933), hearing Radio Paris modulation while listening to Radio Luxembourg in Portsmouth,
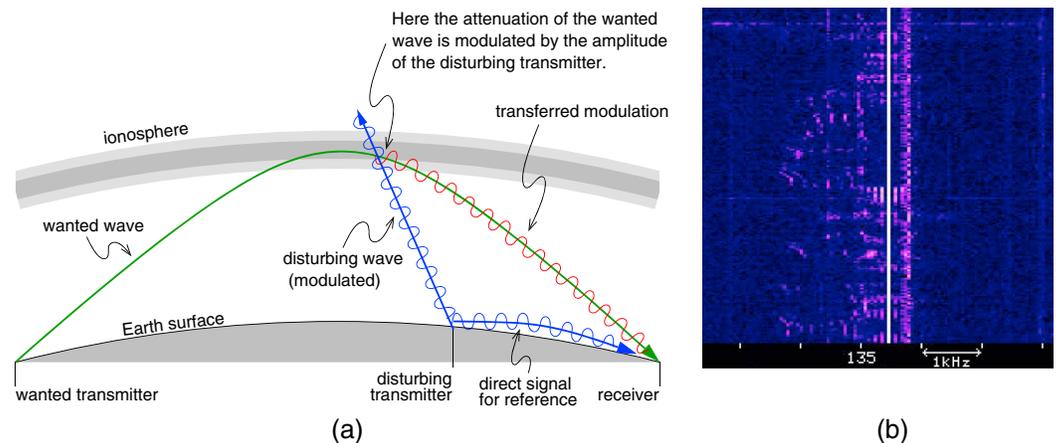
Figure 1. (a) Sketch of how ionospheric cross modulation happens. (b) Spectrogram (from Vester, 2013) on the Twente WebSDR showing the first case of sideband asymmetry; horizontal frequency, vertical time (about 15 s), brightness represents signal strength; the 135.43 kHz transmitter was unmodulated. (The WebSDR is a web-controlled receiver at http://websdr.ewi.utwente.nl:8901/, capable of tuning from 0 to 29 MHz independently by hundreds of simultaneous users; in fact, the Twente receiver referred to later is this same system.)

southwest England. Martin and Jacobs (1956) even state that the effect is "believed to be observed since 1924," unfortunately without giving a source; 1924 seems rather early, given the large transmit powers needed to produce the effect.

Soon after Tellegen's first observations, Bailey and Martyn (1934) came up with a theoretical explanation, the essence of which is that the absorption in the ionosphere of the Swiss signal varies due to the strong Luxembourg signal modulating the average velocity of the free electrons and hence their collision frequency. This theory predicts among others that the cross modulation should have a distinct first-order low-pass filter character due to the relaxation time constant of the electron velocities. Further theoretical development occurred into the 1970s (cf. Maslin, 1976a).

Systematic measurements also started soon after the first detections, as reported by (van der Pol & van der Mark, 1935), followed by many other measurement campaigns, as summarized in (BBC Research Department, 1972). Typical measurements involved switching off the modulation on one transmitter, modulating the other with a test tone, and then using an oscilloscope at the receiver to measure amplitude and phase of the transferred modulation (see, e.g., Newton et al., 1948; Ratcliffe & Shaw, 1948).

Most theoretical and measurement work focused on the magnitude of the cross modulation, since that is of most practical importance when judging the deterioration of reception quality of AM broadcast stations (International Telecommunication Union, 1974). However, by also looking at the phase, and looking at the dependence on the modulation frequency, the theoretical model can be verified, and more can be learned about the ionosphere. Ratcliffe and Shaw (1948) and Huxley et al. (1948) were the first to work on this, crediting Ratcliffe with the idea. They experimentally confirm the low-pass filter predictions from the theoretical model. They also point out that since there is a constant path length difference between the direct path from the disturbing transmitter to the receiver, and the path via the cross modulation in the ionosphere (cf. Figure 1a), an additional phase shift linearly proportional to the modulation frequency is to be expected. Measuring this phase shift reveals, through elementary geometry, where in the ionosphere the cross modulation occurs.

In later years, research into ionospheric nonlinearity has continued; however, this has focused on other aspects than cross modulation, making it less relevant for this paper.

In this paper, we look extensively at the dependence of phase and amplitude of the cross modulation as a function of the modulation frequency. As a baseline, let us summarize what existing theory says. For the amplitude

1. it is expected to follow a first-order low-pass filter of the modulation frequency, with a corner frequency of 160–320 Hz (Huxley & Ratcliffe, 1949);

2. the absolute magnitude is notoriously hard to predict, as it depends on the state of the ionosphere, the angles of incidence, the absorption height, the frequencies, etc. (Maslin, 1976b); in the sequel, we will only use relative comparisons, that is, compare the magnitude of different modulation frequencies for the *same* set of transmit and receive stations at the *same* time.

The phase difference between the ionospheric cross modulation and the direct signal from the disturbing transmitter is expected to consist of four terms:

1. a base phase shift of either 180° or 0° (depending on the state of the ionosphere);
2. a first-order low-pass filter, giving a phase shift increasing from 0° up to 90° for increasing modulation frequencies;
3. a linear increase with the modulation frequency due to the path length difference;
4. possibly an extra phase shift, independent of the modulation frequency, due to the groundwave contributing to the received carrier with an arbitrary phase and magnitude.

The latter term is specific to our single-sideband measurements; with double-sideband demodulation, the groundwave can only cause a 180° phase shift, as pointed out in Ratcliffe and Shaw (1948).

3. Measurement Setup

Figure 2 shows a block diagram of the measurement setup we use. Signals from the antenna are digitized and fed to three AM demodulators, each preceded by a filter. Digitization by a common wideband ADC ensures that the data from all channels are perfectly synchronous. The first AM demodulator is set to the frequency of the "disturbing" station, with a filter wide enough to pass both sidebands. The second and third AM demodulators are set to the "wanted station", with filters passing the lower and upper sidebands (LSB and USB), respectively. These filters have a flat passband and very steep flanks (−50 dB in 200 Hz). The AM demodulators in most cases are synchronous AM demodulators, which regenerate the carrier locally using a phase-locked loop, but in phase with the incoming carrier. One transmitter (TDF on 162 kHz) carries too much phase modulation, so in experiments involving this station, envelope detectors were used. All three demodulators have an automatic gain control, controlled by the carrier level, so their output is proportional to the modulation index.

The output of the AM demodulators is at a sample rate between 7 and 16 kHz (depending on details of which SDR hardware and filter settings are in use) and is fed to a filter bank, implemented by a raised-cosine windowed 1024-point fast Fourier transform (FFT), giving bin sizes between 7 and 16 Hz, with 50% overlap between consecutive FFTs, resulting in a sample rate at the output of the filter bank between 14 and 31 Hz.

The outputs of the filter banks are fed to a cross correlation, which is performed separately for each audio frequency bin. The cross-correlation sums 2,048 consecutive samples. The magnitude and phase are plotted using brightness and color, respectively, as a function of modulation frequency (vertical) and time (horizontal, one column of pixels every 1 to 2.5 min).

Denoting by $x_{i,j}$ the complex amplitude of the ith frequency bin from the jth FFT from the "disturbing wave" (which serves as reference), and $y_{i,j}$ the same from one of the sidebands of the "wanted wave," the normalized cross correlation for the ith frequency bin is computed as $z_i = \frac{\sum_{j=0}^{2047} y_{i,j} x_{i,j}^*}{\sum_{j=0}^{2047} x_{i,j} x_{i,j}^*}$, with $*$ denoting complex conjugation. We only compute this cross correlation for time offset 0, since we have no reason to expect correlations at larger offsets than the duration corresponding to an FFT bin, 64 to 146 ms.

The amplitude plotted is $|z_i|$. Due to the use of automatic gain control, this amplitude can be interpreted as the ratio of the modulation depth of the disturbing wave, and the modulation depth it induces onto the wanted wave, with some caveats: multipath (e.g., groundwave) may change the net carrier power, and modern AM transmitters adapt the carrier power to the modulation for energy efficiency (called Dynamic Carrier Control).

The phase plotted is $\arg(z_i)$ for the upper and $-\arg(z_i)$ for the lower sideband. The choice of reversing the sign for the LSB may seem odd at first. However, this choice is natural in terms of complex signals, if one interprets the LSB as the mixing product of the carrier with negative modulation frequencies. In fact, as we will see in the graphs in the sequel, this choice of signs provides a smooth continuation between LSB and USB. As the cross-modulation lags, because it takes a detour compared to the reference signal, we will see a negative phase in the USB, and positive in the LSB.
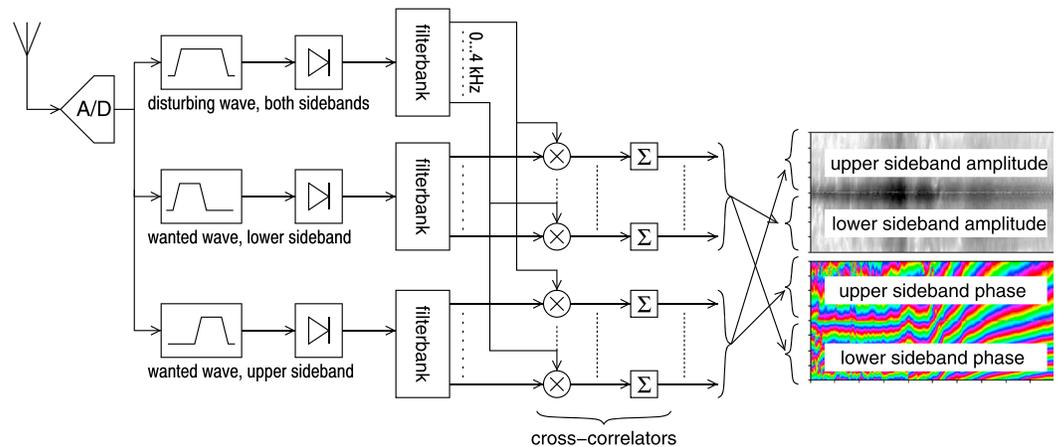
Figure 2. Measurement setup.

In the above, a number of parameters were given somewhat arbitrary values, such as the sample rates and the FFT size. These values are not critical. The goal here is to get a qualitative impression, and the effects are strong enough that they are clearly visible with the given parameter values. The signal-to-noise ratio of the plots could be improved at the cost of time or frequency resolution, which is not desirable for initial exploration; furthermore, the human eye is quite good at doing spatial averaging to find patterns in these plots even if they are noisy. For more quantitative work in sections 5 and 6, a further step of integration over time is done, after using the qualitative plots to select a time interval during which propagation is stable enough.

We have described the procedure as a filter bank followed by a cross correlation for each frequency bin. However, one can also, but equivalently, see it as measuring the transfer function of a (supposedly) linear system with much noise added. The linear system's input is the disturbing station's modulation, and the output is the demodulator output of one of the single-sideband receivers for the wanted station; the wanted station's own modulation acts as added noise. Determining the transfer function, while averaging out the noise, can be done by computing the Fourier transform of the cross correlation between the system's input and output, and dividing this by the Fourier transform of the input. However, due to the typical spectrally sparse modulation content, direct spectral division before averaging would fail because many Fourier bins would frequently get close to divide-by-zero. This problem is much less severe in our approach where the divisor is a long-term autocorrelation.

4. Exploratory Measurement Results

The setup described in the previous section has been used to study the per-sideband phase and magnitude in a large number of cases of ionospheric cross modulation. The transmitters involved were mostly longwave (153–279 kHz) and medium-wave (531–1602 kHz) broadcast transmitters, and a few longwave utility transmitters, spread out over Europe, with various transmit powers and antenna configurations. Most reception was done using a fixed receiver at the University of Twente, in Enschede, the Netherlands; this system allows for round-the-clock monitoring. Using a portable receive setup, measurements have also been done at various other locations; these measurements were limited to about 1 hour at a time. Figure 3 gives an overview of transmit and receive locations.

A few representative examples of the received signals are presented in Figure 4. Let us first focus on Figure 4a, showing how the Luxembourg (RTL) 234 kHz transmitter affects the 216 kHz signal from southern France (RMC), as received at Twente; we concisely denote this as RMC/RTL/Twente, that is, wanted transmitter / disturbing transmitter / receiver.

The amplitude picture (top half) is essentially symmetric here. One clearly sees that the cross modulation is strongest for low modulation frequencies, demonstrating the low-pass effect due to thermal inertia discussed in section 2. Note that the effect is only present between about 03 and 22 h UTC, simply because the 216 kHz transmitter is switched off during the night.
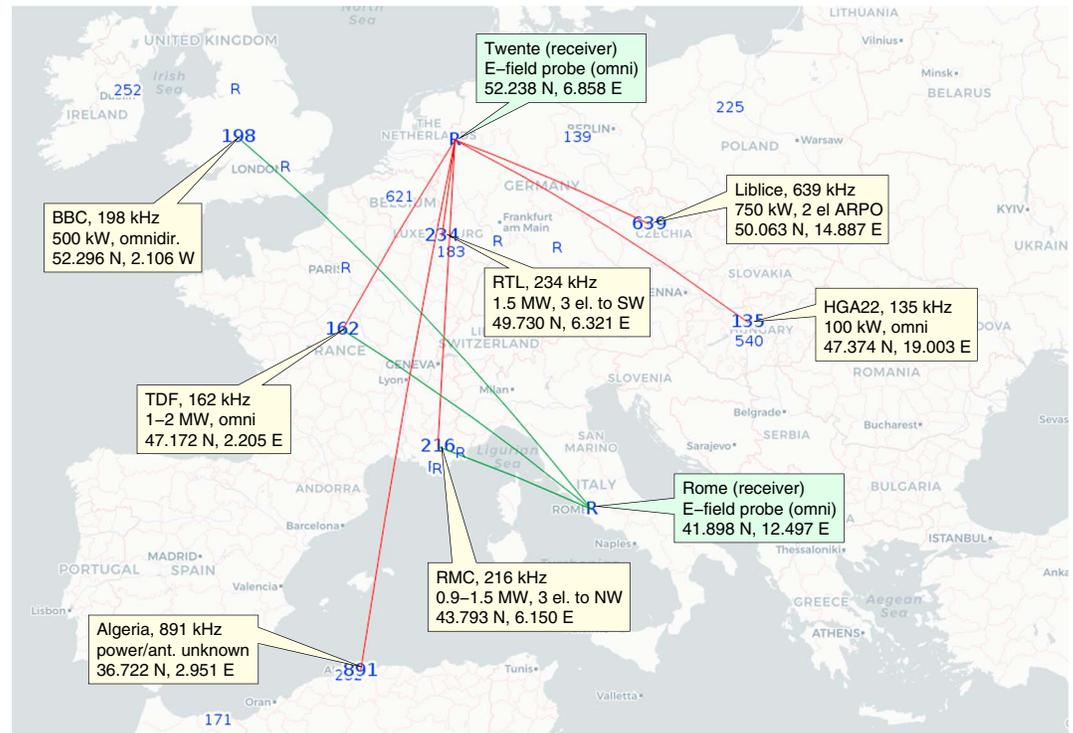
Figure 3. Map showing the stations involved; numbers are transmit frequencies in kilohertz; "R" denote receiving sites; colored lines are great-circle paths of measurements shown in this paper.

The phase plot (lower half) is more revealing. The colored bands in the phase diagram get wider as we go from the LSB into the USB. This is the unexpected phase asymmetry: as mentioned in section 2, if there were a simple constant delay between the reference and cross-modulation paths, the phase should rotate by a constant number of degrees per kilohertz of modulation frequency, leading to equally wide color bands.

Clearly, the colored bands are not completely straight; they drift a bit, randomly, and occasionally even undergo jumps, and these patterns also vary from day to day. Presumably, these effects are caused by diurnal, seasonal, and random changes in the ionosphere. Also, reception of the carrier via both undisturbed groundwave and the ionosphere, with potentially deep fading and carrier phase inversions, may play a role, as discussed in section 2.

Next, consider Figure 4b, showing how the same Luxembourg 234-kHz transmitter affects the 162-kHz signal from central France, again received in Twente. The change in widths of the phase bands is now much less pronounced. We do see that during daylight hours the LSB is stronger. The lower ionosphere at midday also shows less skywave delay and thus wider phase bands.

Figure 4c relates to the utility transmitter on 135 kHz in Hungary affected by a medium-wave transmitter in the Czech Republic, of which we already saw a spectrogram in Figure 1b. The amplitude diagram is heavily asymmetric, with the lower sideband being much stronger than the upper.

Finally, Figures 4d and 4e give results from a 1-hour measurement in Rome, with RMC's 216-kHz transmitter disturbing signals from central France (TDF, 162 kHz) and the UK (BBC on 198 kHz). Interestingly, in the former case the LSB is strongest, as opposed to the USB in the latter case. We will return to this in section 6.

5. Modeling of the Phase Behavior

Usually, one treats the signals as rays propagating from the wanted transmitter to the receiver, and from the disturbing transmitter to some place in the ionosphere, as sketched in Figure 1a. This seems incompatible with our measurements, since it predicts a constant delay of the cross modulation, thus a linear phase curve. Effectively, the ray picture assumes only the center of the first Fresnel zone is of relevance.
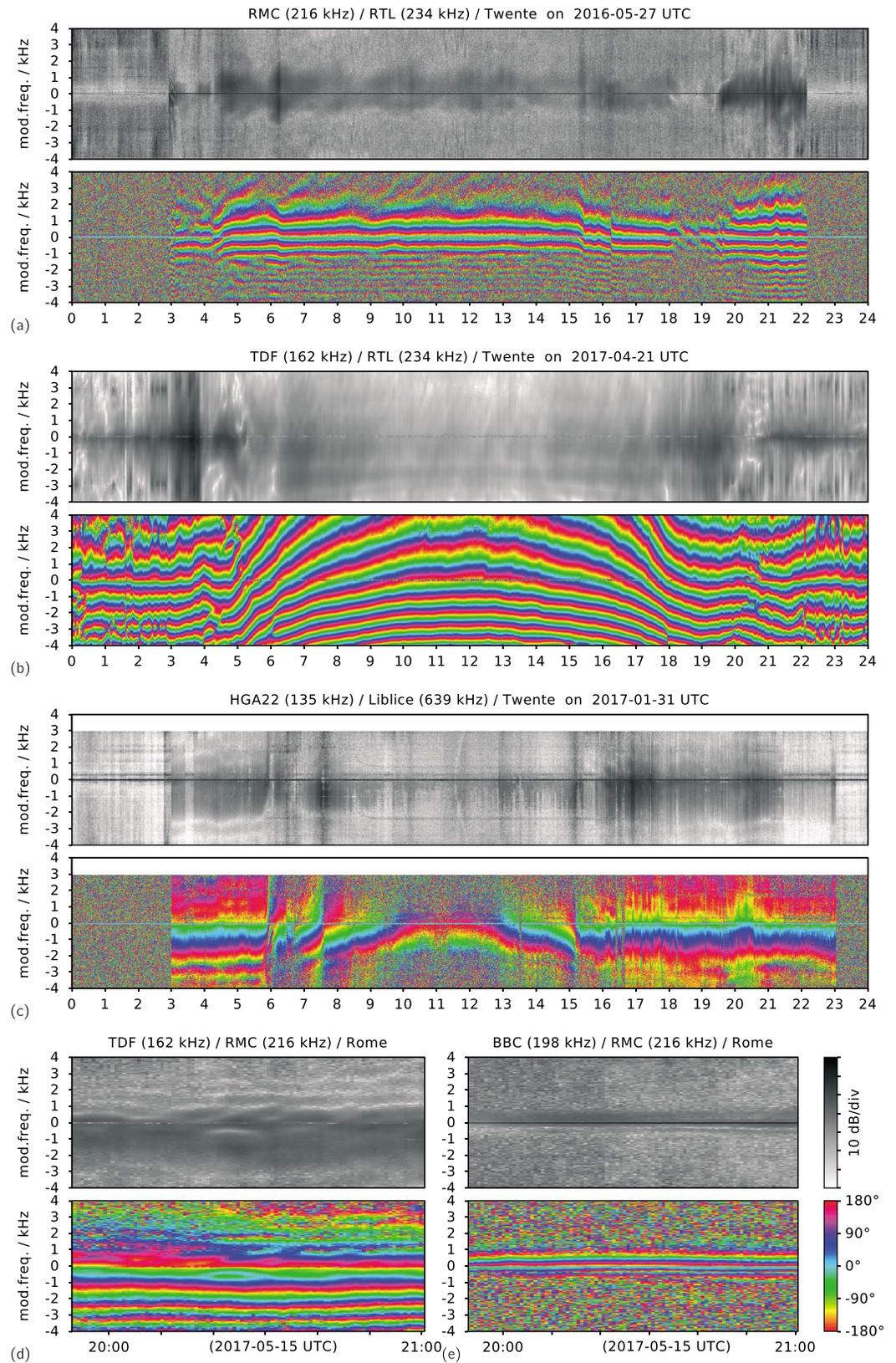
Figure 4. Typical exploratory measurement results, showing the magnitude (darker = larger) and phase (color scale at bottom right) of the ionospheric cross modulation separately for both sidebands. (a)–(e) are for several combinations of transmitters and receivers.

As we will show in this section, an explanation consistent with our measurements can be found by taking contributions from other parts of the ionosphere than just the first Fresnel zone into account. These parts contribute due to two aspects:

1. the heating due to the disturbing wave does not occur simultaneously in the whole Fresnel zone, and
2. the propagation delay and phase from different parts of the heated ionosphere to the receiver varies with their location, also near the center of the Fresnel zone.

In section 5.1 we calculate precisely what sidebands this results in, using Kirchhoff's diffraction integral, and in section 5.2 we show that this matches our measurements. To complement the precise but uninsightful integral, we give a qualitative phenomenological discussion in section 5.3.

An alternative view is developed in section 5.4; here we do treat the waves as rays again, but these rays are deflected upon acquiring cross modulation in the ionosphere, in a manner reminiscent of Bragg reflection in crystallography. The resulting ray tracing diagrams are elucidating, and also useful for amplitude asymmetry considerations in section 6. Furthermore, we use the ray view to derive an explicit expression for the quadratic term of the phase behavior in section 5.5.

Throughout, we assume that the ionosphere consists of a single layer at an altitude of 90 km (the E layer), which serves both as a reflector and as the place where the cross modulation happens. Essentially, we model the ionosphere as a spatially distributed scatterer for the wanted wave, whose reflectivity (or, conversely, absorption) is modulated by the disturbing wave. During nighttime, this simplification is justified by the steep gradient of the ionization with height (cf. Maslin, 1976c; Figure 3). In section 5.6, we qualitatively discuss the daytime situation.

5.1. Kirchhoff Integral Explanation

Our precise approach to model the phenomenon consists of applying the Kirchhoff diffraction formula to the E layer surface. Thus, one integrates over the entire surface, computing for each point what its contribution to the field strength at the receiver is, given the incoming field from the wanted transmitter, and the modification this field undergoes under the influence of the disturbing transmitter. From this, we calculate a complex, frequency dependent transfer function (or equivalently, a complex impulse response in time domain) from the disturbing station's modulation to the induced modulation as seen by the receiver.

Apart from normalizing factors, the integral is as follows (cf., e.g., Barton, 1989).

$$E_{RX} = \int_S \frac{e^{j\omega(r_{TX}+r_{RX})+j\Omega(r_H+r_{RX}-r_{ref})}}{r_{TX} \cdot r_{RX}} \cdot \frac{\cos(\vec{n},\vec{r}_H)}{r_H^2} \cdot \left(\cos(\vec{n},\vec{r}_{TX}) + \cos(\vec{n},\vec{r}_{RX}) \right) dS \tag{1}$$

where the integral is over the entire surface S of the E layer; j denotes the imaginary unit, r_X denotes the distance of the point under consideration to object X (wanted transmitter (TX), disturbing transmitter (H, for Heater), or receiver (RX)), $\vec{r}$ same but as a vector, $\cos(\vec{n},\vec{r})$ the cosine of the angle between $\vec{r}$ and the normal $\vec{n}$ to the surface patch dS, ω the angular frequency of the wanted wave, and Ω the angular frequency of the modulation of the disturbing wave. Finally, r_{ref} is the length of the direct path from disturbing transmitter to receiver; its inclusion in the modulation phase term represents the fact that the receiver uses the directly received signal as a reference for the modulation phase. Note the division by r_H^2, since the cross modulation is proportional to the disturbing wave's power, that is, quadratic in its field strength (Huxley & Ratcliffe, 1949), and the factor $\cos(\vec{n},\vec{r}_H)$ because the patch of ionosphere area intercepts the disturbing wave's power at this angle. If the antennas are not omnidirectional, factors representing their radiation diagrams can easily be inserted. Except where noted otherwise, we will use the radiation diagram of a short vertical monopole for all antennas in our evaluations of the Kirchhoff integral (i.e., field strength proportional to cosine of the elevation angle).

We evaluate the formula numerically, by discretizing the ionospheric layer into a two-dimensional array of patches; a 0.5-km grid was found to work well, and the integration was limited to the part of the ionosphere that contributes significantly (found experimentally by increasing until convergence). Then for each modulation frequency, we calculate and sum the complex contributions from all patches. An alternative, numerically efficient variant of the calculation works directly in the time domain, by binning the contributions from the patches by their delay into an array for the impulse response, and then computing the FFT of this.
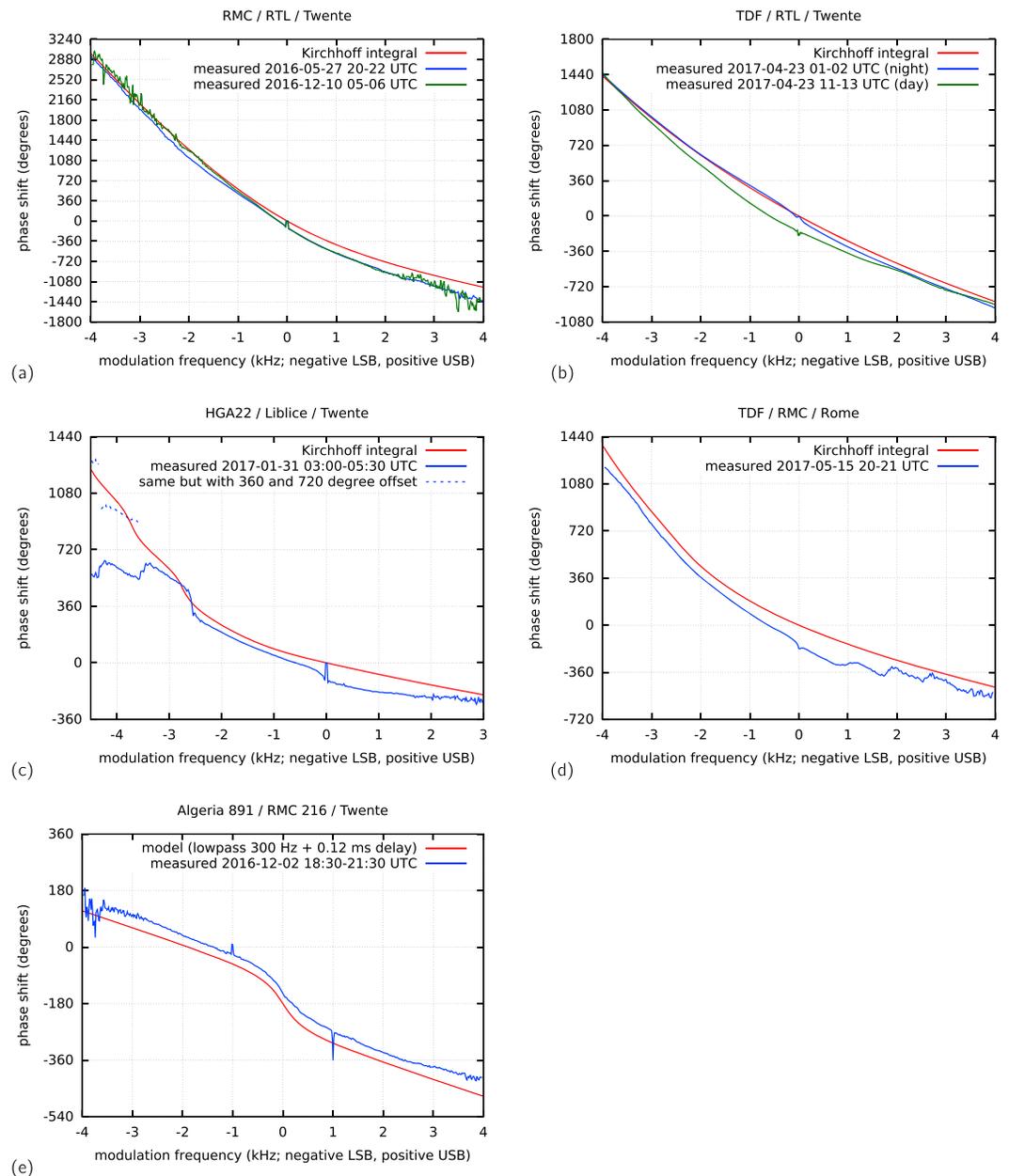
Figure 5. Comparison of phase as measured and as calculated using the Kirchhoff integral in (a)–(d); (e) shows a case without asymmetry for reference. LSB = lower sideband; USB = upper sideband.

5.2. Comparing Measurements and Theory

In Figures 5a–5d we compare the phase as calculated above and as measured as a function of modulation frequency, for the examples from Figures 4a–4d. The phase in these plots has been "unrolled" heuristically to resolve the 360° ambiguity inherent in phase measurements.

The phase as calculated is purely the outcome of the Kirchhoff integral. As pointed out in section 2 there may be an additional constant phase shift due to the presence of a groundwave carrier. Furthermore, there is an additional phase shift of up to 90° for higher-modulation frequencies due to the low-pass filter effect. Due to the choice of sign, this is positive for the LSB and negative for the USB. Thus, when comparing the measurement and the calculation, we should expect an arbitrary constant offset, plus an offset which increases by 180° over the entire span of LSB and USB modulation frequencies. All of this is clearly visible when there is no asymmetry (cf. Figure 5e).
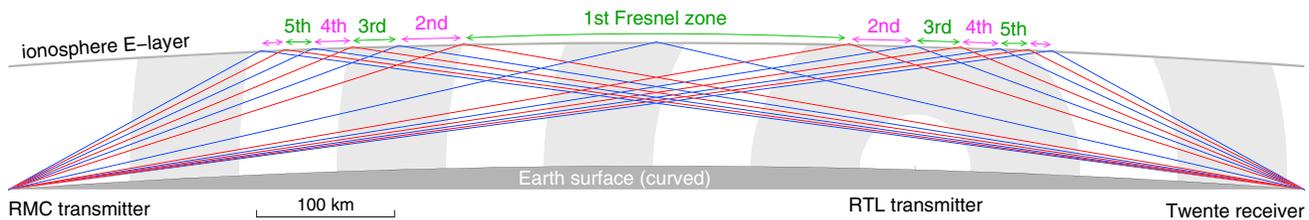
Figure 6. Sketch illustrating a qualitative explanation of the sideband asymmetry.

Figure 5a relates to our main example, RMC/RTL/Twente, corresponding to Figure 4a. Clearly, the match between measurements (at two different times of year) and calculation is quite good (taking into account the above remarks about the expected arbitrary offset).

Figure 5b relates to the 162 kHz transmitter in central France received in Twente and disturbed by Luxembourg. The nighttime measurement matches well with the calculation. The daytime measurement does not; we will get back to this in section 5.6.

Figure 5c shows the phase in the case of the Hungarian and Czech transmitters. The severe deviations at the left can well be explained by phase-unroll errors, cf. the dotted lines 360° and 720° higher. Tracking the phase for unrolling was hampered in this case by several deep amplitude dips (cf. Figure 10).

Finally, Figure 5d relates to the TDF/RMC/Rome case. The asymmetry is very pronounced here, with only about 360° of phase shift over the entire USB, and 4 times as much over the LSB. The phase curve in the USB has some irregularities, presumably due to too low signal-to-noise ratio there, and due to a phase drift in the course of the 1 hr measurement (cf. Figure 4d).

5.3. Qualitative Explanation

Although the results from evaluating Kirchhoff's integral clearly match well with the measurements, this does not provide much understanding, other than that contributions beyond the direct ray apparently need to be taken into account. In this section we therefore reason qualitatively to show how the contribution from the first Fresnel zone diminishes, and how higher Fresnel zones can contribute.

Figure 6 shows, to scale, the situation for RMC/RTL/Twente. The central blue line is the shortest path from RMC to Twente (for normal single-hop propagation, without the ionospheric cross modulation). The red lines beside it are paths that are one-half wavelength longer, thus contributing in opposite phase. The next blue lines are again half a wavelength longer, and so on. Together, they delineate the Fresnel zones on the ionosphere: the odd-numbered (green) zones contribute in phase which each other, while the even-numbered (pink) zones are in opposite phase to them. Effectively, the contributions from the higher-numbered zones largely cancel each other, leaving the first zone as the dominant contributor. Note that due to the slant incidence, Fresnel zones are highly elongated ellipses; thus, their spatial scale can be on the order of the wavelength of the *modulation* of the disturbing transmitter.

Now consider what happens if the absorption (or the reflectivity) of the ionosphere varies with the modulation of the RTL transmitter. If the variation is slow, the signal strength received in Twente will simply vary synchronously with it. But if the modulation frequency is higher, this changes. Let us assume the RTL transmitter is modulated with a pulse, which, upon reaching the ionosphere will briefly enhance the reflection (decrease the absorption) of that part of the ionosphere. Clearly, this "ionospheric heating pulse" will not reach all parts of the first Fresnel zone simultaneously. And even if it did, the distance from each part of this zone to the receiver is different, so the briefly enhanced reflectivity from all parts will not be received simultaneously. In other words, the first Fresnel zone's total contribution to the ionospheric cross modulation as seen by the receiver, is smeared out in time.

In contrast, interesting things start to happen farther out, near the border of the first Fresnel zone and into the higher Fresnel zones, causing them to start contributing to the sidebands. Consider for example the second and third Fresnel zones right above the RTL transmitter, where the pulse arrives almost simultaneously. Normally, the contributions of these zones would mostly cancel each other. However, from right to left the distance to the receiver increases. So when the reflectivity of this piece of ionosphere temporarily increases,
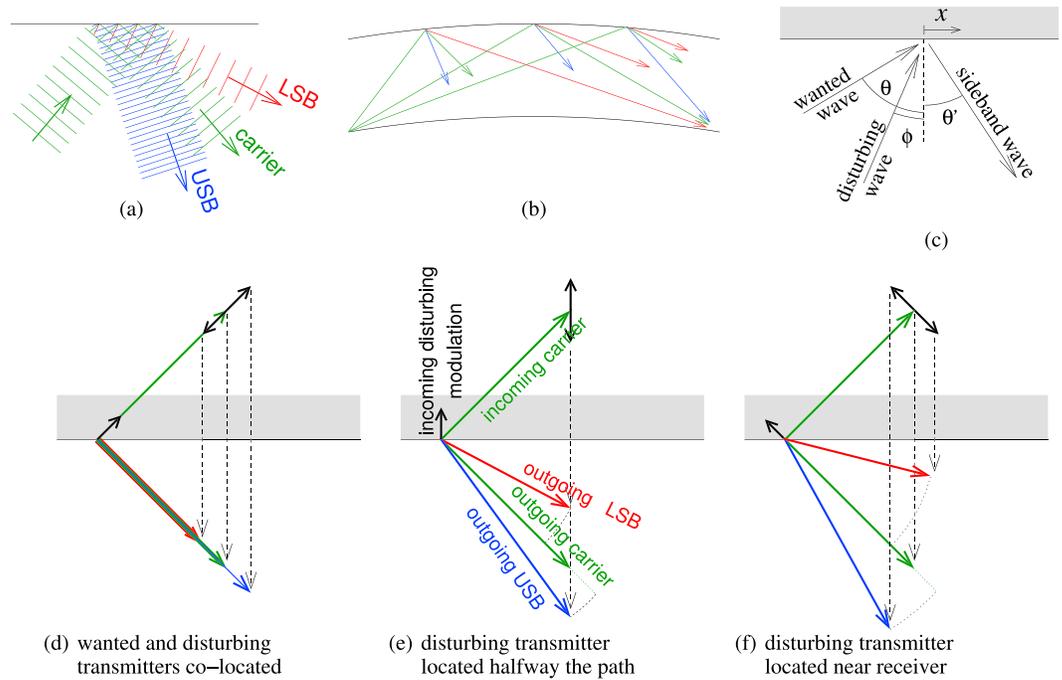
Figure 7. Essence of the deflected-ray explanation. (a) Reflection at one point of the ionosphere; (b) reflections as seen by one receiver; (c) calculation of reflection angle; (d)–(f) wave vectors in three cases.

then first the increased contribution from the rightmost part of the third Fresnel zone (which lags the center of the first Fresnel zone by 540°) will reach the receiver, followed by that from the middle (lagging by 360°), followed finally by the leftmost part of the second Fresnel zone (lagging by 180°). So we see the contribution cycle through 360° in a small amount of time. This is effectively a Doppler effect, and if after precisely this time the pulse (and thus the cycling through 360°) repeats, the result is a new signal at a slightly higher frequency: the USB comes into existence. Similarly, for other pulse repetition frequencies, other parts of Fresnel zones can produce an appropriate USB, and parts of the ionosphere left of the first Fresnel zone give LSB signals by the same mechanism.

Thus, we see how the receiver effectively sees the upper and lower sidebands of the transferred modulation arriving from different parts of the ionosphere.

5.4. Deflected Ray Explanation

Another way of looking at the phenomenon is by going back to the usual explanation of why a plane wave reflects off a flat surface with equal angles of incidence and reflection. Normally, one demands that for each incoming wave front, an outgoing wave front leaves the reflecting surface; continuity of the wave fronts then leads to the angles of incidence and reflection being equal. With ionospheric cross modulation, at each point of the ionosphere sum and difference frequencies arise (wanted carrier $\pm$ disturbing modulation frequency), which have a different wavelength than the incoming carrier wave. If we still demand continuity of the departing wave fronts, they have to leave in different directions, as sketched with exaggeration in Figure 7a. (This is in fact comparable to *refraction* happening at a border between regions with different wave propagation speeds.) The net result of this is that as seen from a receiver, the sidebands seem to be received from different directions, as shown in Figure 7b.

Using Figure 7c, we can compute the angle under which the sideband wave will leave. We assume the wanted wave, with angular frequency ω, arrives under angle of incidence θ, and the disturbing wave, carrying modulation with angular frequency Ω, under angle of incidence ϕ. Note that ϕ may be negative if the disturbing station is more to the right, that is, closer to the receiver, and that negative Ω corresponds to producing the lower sideband. Consider the situation at point x on the ionosphere surface (presumed flat at this scale). The complex incoming signal from the wanted wave, assuming unity amplitude, is $e^{j\omega t - j\sin(\theta)\frac{x\omega}{c}}$. The complex amplitude of the disturbing wave's modulation is $e^{j\Omega t - j\sin(\phi)\frac{x\Omega}{c}}$. The resulting sideband is the product of these;
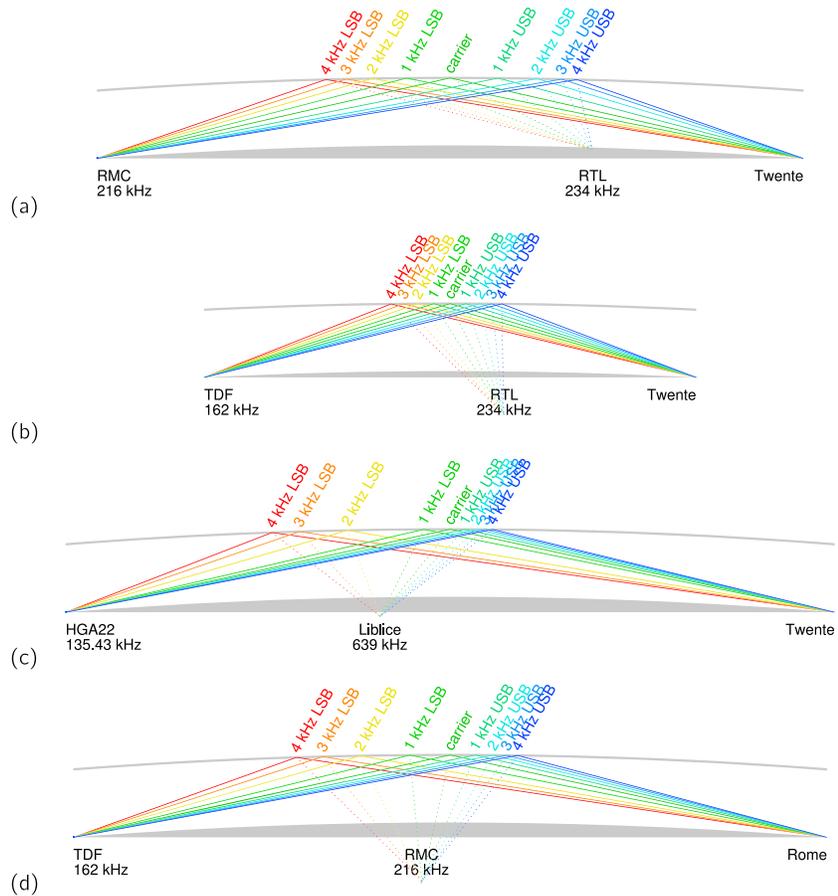
Figure 8. Ray tracing for the various examples. (a)–(d) correspond to the cases of Figures 4a–4d and 5a–5d.

but this product must be the sideband wave of angular frequency $\omega+\Omega$ departing under angle θ' to the right, so we must have, independently of x:

$$e^{j\omega t - j\sin(\theta)\frac{x\omega}{c}} \cdot e^{j\Omega t - j\sin(\phi)\frac{x\Omega}{c}} = e^{j(\omega+\Omega)t - j\sin(\theta')\frac{x(\omega+\Omega)}{c}}.$$

Many terms cancel, leaving

$$\omega\sin(\theta) + \Omega\sin(\phi) = (\omega+\Omega)\sin(\theta')$$

Since $|\Omega| \ll |\omega|$, we expect $\theta' \approx \theta$, so let us linearize around θ; then an elementary calculation gives

$$\theta' - \theta = -\frac{\Omega}{\omega} \cdot \frac{\sin\theta - \sin\phi}{\cos\theta} + O(\Omega^2). \tag{2}$$

Figures 7d–7f show the deflection in a graphical way. The arrows here are the wave vectors, that is, vectors pointing in the direction of the wave propagation and with a length proportional to the frequency. The cross-modulation process is a multiplication of the complex field strengths at the ionosphere surface, which boils down to addition (subtraction) for USB (LSB) of the incoming wanted wave vector and the incoming disturbing station's modulation wave vector, projected onto that surface. The horizontal component of the outgoing wave vector must match this projection. The vertical component of the outgoing wave vector, and thus its direction, is then fully determined by the requirement that its total length represents its frequency, which equals the sum (difference) of the incoming frequencies. In short, the outgoing vector must match the sum (difference) of the incoming wave vectors both in total length, and in projection onto the ionosphere surface.

Figure 7d shows the special case where both transmitters are colocated. Then the three outgoing vectors point in the same direction, as can also be seen by substituting $\theta = \phi$ in (2); consequently, there will be no
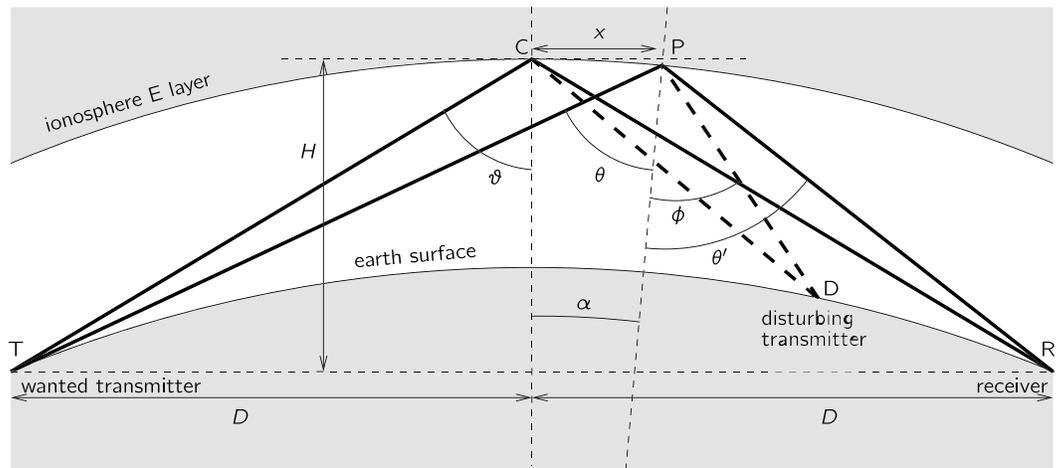
Figure 9. Geometry of the deflected-ray explanation.

sideband asymmetry. This may explain why in the case of Algeria/RMC/Twente no asymmetry occurred, as shown in Figure 5e: a two-hop path from Algeria to Twente would have its ground reflection near the RMC transmitter, effectively making them colocated for the second hop.

Having derived formula (2) for the change of direction of a ray in ionospheric cross modulation, we can make correctly scaled drawings of the rays for our various examples: Figure 8. These figures represent vertical planes through the wanted transmitter and the receiver. Correctly showing the disturbing transmitter is only possible if it happens to be in this same plane; in other cases, it is drawn such that the distance from it to the ionosphere is still correct, leading to an apparent underground location.

The figures clearly show that as one goes from USB to LSB, the path length from disturbing transmitter via the ionosphere to the receiver increases, explaining the nonlinear phase behavior. In fact, if we measure the lengths of these paths and calculate the associated phase shifts, we get curves that are practically indistinguishable from the Kirchhoff integral results (less than 2% error in the unrolled phase), with one exception: for HGA22/Liblice/Twente the "wobbles" in the LSB in Figure 5c are not reproduced by the ray tracing.

As we will see in section 6, these ray tracing plots are also useful for understanding the amplitude asymmetry.

5.5. Deflected Ray Explanation: Quantitative
Based on ray deflection, we can derive an analytical expression for the quadratic term of the phase as a function of modulation frequency. This quadratic term is the lowest order of sideband asymmetry, as a linear term is always expected (cf. section 2).

Refer to Figure 9. We define C to be the point on the ionosphere halfway between wanted transmitter and receiver, where the carrier reflects. We assume the sideband for modulation frequency Ω is received from a point P, located a distance x to the right of C. Let H denote the height of the ionosphere at this point above the straight (subterranean) line connecting wanted transmitter and receiver, and D half the length of this line; denote by $\alpha = x/R + O(x^2)$ the angle spanned by the x displacement as seen from the earth center, with R the radius of the ionosphere. Then $\tan(\theta + \alpha) = (D + x)/H$ and $\tan(\theta' - \alpha) = (D - x)/H$. A straightforward derivation using the Taylor expansion of the tan function results in

$$x = \frac{\Omega}{\omega} \cdot \frac{H}{2} \cdot \frac{\sin\theta - \sin\phi}{\cos^3\theta} \cdot \rho^{-1} + O(\Omega^2),$$ (3)

where $\rho = 1 - H/R\cos^2\theta$. For a flat Earth, $\rho = 1$, and this is also often a good approximation on our spherical Earth: ρ^{-1} initially grows slowly with D, reaching 1.1 at $D = 200$ km, and 1.3 at 400 km; only at larger distances, it becomes large, reaching about 20 at 1,000 km, the limit for single-hop propagation.

The phase of the output of the receiver's demodulator is the difference of the phase of the sideband signal, generated at P, and the carrier itself reflecting at C. If there were no displacement, that is, $x = 0$ and P = C, then the received phase would vary linearly with the modulation frequency, as argued in section 2. We only need to calculate the extra phase shift due to P ≠ C. Again referring to Figure 9, some straightforward but
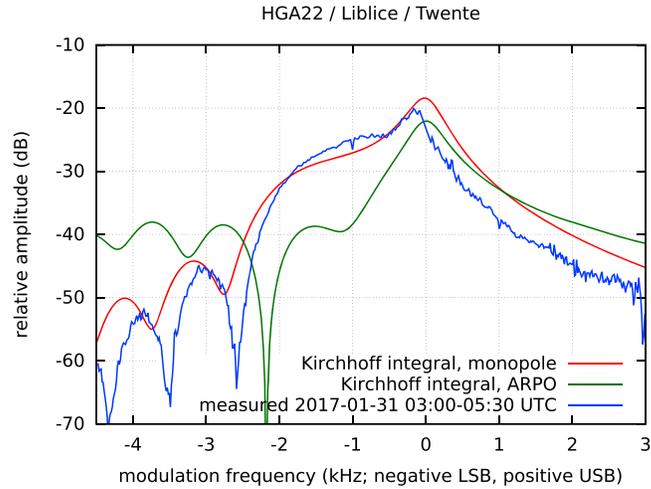
Figure 10. Measured and modeled amplitude behavior in the HGA22/Liblice/Twente case. LSB = lower sideband; USB = upper sideband.

tedious calculations show that the displacement x increases TP (distance from wanted transmitter to point P) by, where $a = 1/2R$,

$$\Delta_{TP} = \sqrt{(D+x)^2 + (H - ax^2)^2 + O(x^4)} - \sqrt{D^2 + H^2} = \sin\vartheta \cdot x + \frac{\cos^3\vartheta}{2H}\rho x^2 + O(x^3).$$

By symmetry, the same holds for PR with the sign of x reversed. Adding them, the linear terms in x cancel (as they should, C being the center of the first Fresnel zone), so the total increase in the TPR path length is

$$\Delta_{TP} + \Delta_{PR} = \frac{\cos^3\vartheta}{H}\rho x^2 + O(x^3) = \frac{\cos^3\theta}{H}\rho x^2 + O(x^3).$$

Note that $\vartheta - \theta = O(x)$; we henceforth do not need to distinguish between ϑ and θ anymore. Next, consider the path DPR, that is, from disturbing transmitter via ionosphere to receiver. Due to the displacement by x, its length increases by

$$\Delta_{DP} + \Delta_{PR} = -(\sin\theta - \sin\phi)x + O(x^2).$$

By substituting x from (3), we find that due to the displacement the phase out of the demodulator advances by

$$-\Delta_{TP}\frac{\omega}{c} - \Delta_{DP}\frac{\Omega}{c} - \Delta_{PR}\frac{\omega + \Omega}{c} = \frac{(\sin\theta - \sin\phi)^2}{\cos^3\theta} \cdot \frac{1}{1 - H/R\cos^2\theta} \cdot \frac{H}{4c} \cdot \frac{\Omega^2}{\omega} + O(\Omega^3). \tag{4}$$

Note that in principle θ and ϕ depend on x and thus on Ω, but these are higher-order effects. We see that the quadratic term in Ω (i.e., the curvature of the phase curve) is larger:

1. for lower carrier frequencies ω;
2. for larger distance between receiver and wanted transmitter, due to the $\cos^3\theta$ factor in the numerator since larger distance translates into larger θ;
3. when the disturbing transmitter is closer to the receiver (ϕ negative, $|\theta - \phi|$ large).

5.6. Daytime

So far, we modeled the ionosphere as if it were a simple surface at which both reflection and intermodulation happen, justified by the sudden increase of the electron density with height. But during daytime, the electron density increase is more gradual (cf. Figure 3 in Maslin, 1976c, and references therein). As a consequence, the wanted wave will be bent gradually rather than reflected, on average reach a lower height, and the cross modulation will be induced along the path rather than only at the reflection point.

Indeed, our measurements during daytime show a significant deviation from the Kirchhoff integral predictions; see Figure 5b. Around $\Omega = 0$, the gradient of the phase is smaller, hinting at a smaller path length difference for the modulation, which makes sense given the lower starting altitude of the daytime ionosphere. Also, the shape of the curve is different. We did not try to model this in more detail.

6. Amplitude Considerations

In many cases the measurements do not just show phase asymmetry but also amplitude asymmetry. Given the considerations from the previous section, it is easy to understand how such amplitude asymmetry can arise. We saw that for each modulation frequency in each sideband, a different path provides the main contribution. Each such path reflects off the ionosphere at a different place, so in a different direction as seen from the antennas. Since the antennas are not isotropic, this will lead to different amplitudes for different (cross-modulation) sideband frequencies. Also, due to many other factors (cf. section 2) at different places in the ionosphere, the magnitude of the cross modulation can be different. Furthermore, the geometry may cause rays to become "focused," effectively distributing the sideband energy unevenly over the Earth surface, like a nonflat mirror would do.

Quantitatively, this can be taken into account in principle by inserting appropriate amplitude factors into the Kirchhoff integral (1). Unfortunately, however, it is hard to find the correct factors. The elevation radiation diagrams of the transmit antennas involved are often unknown, and the amount of cross modulation in the ionosphere is also hard to predict. As a consequence, attempts at getting quantitative agreement between calculation results from equation (1) and measurements have not been successful, so we limit ourselves to a qualitative discussion here.

Qualitatively, we can explain a few of the asymmetries using the ray tracing diagrams from Figure 8.

In the case of RMC/RTL/Twente (Figure 8a), no amplitude asymmetry was observed. One might expect the USB to be stronger because it arises much closer to the disturbing transmitter; so apparently, this is offset by its radiation diagram, which presumably favors low elevations and thus the parts of the ionosphere that contribute more to the LSB.

In the case of TDF/RMC/Rome (Figure 8d), the LSB is much stronger (cf. Figure 4d). This can be understood from the ray tracing in Figure 8d and the fact that RMC uses a directional antenna: the USB parts of the ionosphere are at the "back" of the antenna.

In the case of BBC/RMC/Rome, the effect is only visible between about 0- and 1-kHz USB (cf. Figure 4e). With ray tracing (not shown here), it turns out that the parts of the ionosphere that would contribute the other sideband parts are all below the horizon from the point of view of either the transmitter or the receiver, hence their absence.

Finally, consider the case of HGA22/Liblice/Twente, in which the sideband asymmetry was first noticed (Figures 1b and 4c). The ray tracing picture for this case (Figure 8c) shows that a relatively large part of the ionosphere is active for 1- to 2-kHz LSB. This is an example of the focusing effect mentioned above and could well explain the enhancement of the LSB. However, we can calculate that as seen from the Liblice transmitter, this part of the ionosphere appears at an elevation of about 45°. The Liblice antenna is a so-called ARPO antenna, the elevation radiation diagram of which is known (Hausky, 1979) and has a null at about 45° of elevation. Figure 10 shows an attempt at calculating the amplitude behavior of the cross modulation using the Kirchhoff integral, using either the ARPO antenna's radiation diagram or a simple vertical monopole, and assuming a low-pass filter with a 250-Hz corner frequency; also shown is the measured amplitude. Comparing the diagrams (up to a constant amplitude offset, since we cannot calculate the absolute amplitude in the Kirchhoff integral), we see that the measurements do not match the ARPO antenna calculation at all. They do, at least qualitatively, match the vertical monopole calculation: the LSB is enhanced, and there are some dips at high LSB frequencies. It is unknown why the measurements do not match the prediction based on the ARPO antenna; one hypothesis is that the Liblice antenna's radiation diagram somehow deviates from the published ARPO diagram.

7. Historical Perspective

Ionospheric cross modulation has been known for over 80 years and has been studied theoretically and with detailed measurements from the 1930s to the 1970s. Still, the sideband asymmetry was (apparently) never noticed until a few years ago, even though it turns out to be a rather common occurrence when one looks for it. Why has it not been noticed before?

One obvious reason is that nobody looked for it. The natural way of studying ionospheric cross modulation is by using both sidebands in an envelope detector. That is how the effect was first discovered in the 1930s, and with the standard explanation of the desired signal undergoing a time-varying attenuation due to the disturbing station's modulation, one expects the effect to also be a pure amplitude modulation. Haberkant and Vogt (1966) report using a spectrum analyzer for their measurements, so they could have spotted amplitude asymmetry; however, in their case the disturbing and wanted transmitter were near each other, in which case we now understand not much asymmetry will occur (cf. section 5.4).

Only the fact that spectrum diagrams have become ubiquitous due to the use of SDR allowed the first case of sideband asymmetry to be observed by chance; this is the case of HGA/Liblice/Twente, first reported in (Vester, 2013) and shown in Figure 1b. This observation in turn prompted a look at the sidebands independently, when irregular phase behavior was seen using double-sideband detectors in the RMC/RTL/Twente case in 2016.

Another reason is related to the frequencies and path geometries used in the 1940s studies. In most cases where the phase as a function of modulation frequency was studied (Huxley, 1950; Huxley et al., 1948; Ratcliffe & Shaw, 1948), the wanted signal was in the medium-wave range (500–1,600 kHz), while in the current research we focus on wanted signals in the longwave range (130–280 kHz); also, the path lengths were shorter (up to about 500 km, as opposed to 900 km in our main example). As we found in section 5.5, the effect is expected to be stronger with lower carrier frequency and larger distance. Thus, the effect simply was much less pronounced in the 1940s experiments, and apparently too small to be accidentally found.

On the other hand, the asymmetry could also have been predicted theoretically, using reasoning as in sections 5.3 and 5.4. In fact, Maslin (1976b) comes tantalizingly close in his section 4(c), where he discusses the first Fresnel zone, noting the following: "In most practical cases therefore, the region where the transfer of modulation mainly occurs will extend over a horizontal area less than 1000 km^2. Throughout a region as small as this it may be assumed that the strength of the disturbing wave is roughly constant." Thus, he did consider that different parts of the ionosphere contribute, but too quickly concluded that this did not matter.

8. Conclusion

In this paper, we have revisited the phenomenon of ionospheric cross modulation among AM broadcast stations using modern receiver technology. We have demonstrated experimentally that the cross-modulation sidebands are often asymmetric, in amplitude and/or phase, and we have provided an explanation for this phenomenon. The explanation is based on taking into account the contributions from many parts of the ionosphere, and demonstrating that the dominant contributions to the sidebands are received from different places in the ionosphere.

Possible future work includes trying further to understand the anomaly in the HGA22/Liblice/Twente case, and studying the phenomenon at VLF, where the D layer becomes important.

References

Bailey, V., & Martyn, D. (1934). Interaction of radio waves. *Nature*, *133*(3354), 218.

Barton, G. (1989). *Elements of Green's functions and wave propagation*. Clarendon Press.

BBC Research Department (1972). L.F. And M.F. propagation: A study of ionospheric cross-modulation measurements (Report no. 1972/23).

Butt, A. G. (1933). Letters to the editor — Luxembourg: Background of Radio-Paris. *World-Radio*, 556.

de Boer, P. T. (2016). Technische notities van PA3FWM: Maxwell en de ionosfeer. *Electron*, *71*, 146–150.

Haberkant, E., & Vogt, K. (1966). Measurement of ionospheric cross-modulation affecting two high-power LF broadcast transmitters. *European Broadcasting Union Review*, *98*, 154–161.

Hausky, M. (1979). Klasický antifading ci ARPO? *Telekomunikace*, *6*, 83–86.

Huxley, L. G. H. (1950). Ionospheric cross-modulation at oblique incidence. *Proceedings of the Royal Society of London A: Mathematical, Physical, and Engineering Sciences*, *200*(1063), 486–511.

Huxley, L. G. H., Foster, H. G., & Newton, C. C. (1948). Measurements of the interaction of radio waves in the ionosphere. *Proceedings of the Physical Society*, *61*(2), 134.

Huxley, L. G. H., & Ratcliffe, J. A. (1949). A survey of ionospheric cross-modulation (wave interaction or Luxembourg effect). *Proceedings of the IEE – Part III: Radio and Communication Engineering*, *96*(43), 433–440.

International Telecommunication Union (1974). Ionospheric cross-modulation (Report 460-1). First Session of the LF/MF Broadcasting Conference, Geneva.

Martin, E. T., & Jacobs, G. (1956). Ionospheric cross modulation from a 1000 kW long wave broadcast transmitter. In *1956 IRE International Convention Record* (Vol. 4, pp. 9–13).

Maslin, N. (1976a). The effect of a time varying collision frequency on a radio wave obliquely incident on the lower ionosphere. *Proceedings of the Royal Society of London, Series A*, *348*(1653), 245–263.

Maslin, N. (1976b). Estimating ionospheric cross-modulation. *Proceedings of the Royal Society of London, Series A*, *351*, 277–293.

Maslin, N. (1976c). Theory of the modulation imposed on an obliquely incident radio wave reflected in a disturbed region of the lower ionosphere. *Proceedings of the Royal Society of London, Series A, 349*(1659), 555–570.

Newton, C. C., Hyde, F. J., & Foster, H. G. (1948). Ionospheric cross-modulation: Techniques of measurement. *Proceedings of the Physical Society*, *61*(2), 134.

Ratcliffe, J., & Shaw, I. (1948). A study of the interaction of radio waves. *Proceedings of the Royal Society of London, Series A, 193*(1034), 311–343.

Tellegen, B. D. H. (1933). Interaction between radio waves. *Nature*, *131*(3319), 840.

van der Pol, B., & van der Mark, J. (1935). Interaction of radio waves. *Tijdschrift van het Nederlandsch Radiogenootschap*, *7*, 12–17.

Vester, M. (2013). Asymmetric Luxembourg effect on HGA: Posting to rsgb-lf-group Internet mailing list, 2013-01-06. Retrieved from https://uk.groups.yahoo.com/neo/groups/rsgb_lf_group/conversations/topics/2189