Measuring the Ionosphere at vertical incidence using Hermes,
Alex, and Munin Open HPSDR and Gnuradio.
Tom McDermott, N5EG, n5eg@tapr.org

Key words: Monostatic, Ionosphere, Doppler, Correlation, HPSDR, Gnuradio

Abstract

This paper describes a monostatic method for measuring the vertical virtual height and the vertical velocity of the F-layer of the ionosphere. The equipment is simple and relatively low power, it uses the Open HPSDR Hermes transceiver module, Munin broadband Power Amplifier (PA), and Alex RF filter module. The antennas consist of a 40m dipole and antenna tuner for transmit and an active receive loop antenna. The software real-time processing (reception, windowing, and correlation) is done using Gnuradio on a Linux PC, followed by post-processing using a Python program (multiple sweep integration and plotting).

Ionosphere

Figure 1 shows typical critical frequency for the E- and F layers versus local time of day. While the MUF depends on the angle of incidence to the ionosphere, the critical frequency is defined at vertical incidence. Generally measuring the E-layer requires using the 160m band (except occasionally near noon during the summer at lower latitudes it may be possible to use 80m). Measuring the F2-layer critical frequency can be done on the 80m band much of the day, and sometimes on the 40m band. When measuring the F-layer, the higher the frequency used the less the F-layer echo attenuation caused by transiting the E-layer twice.
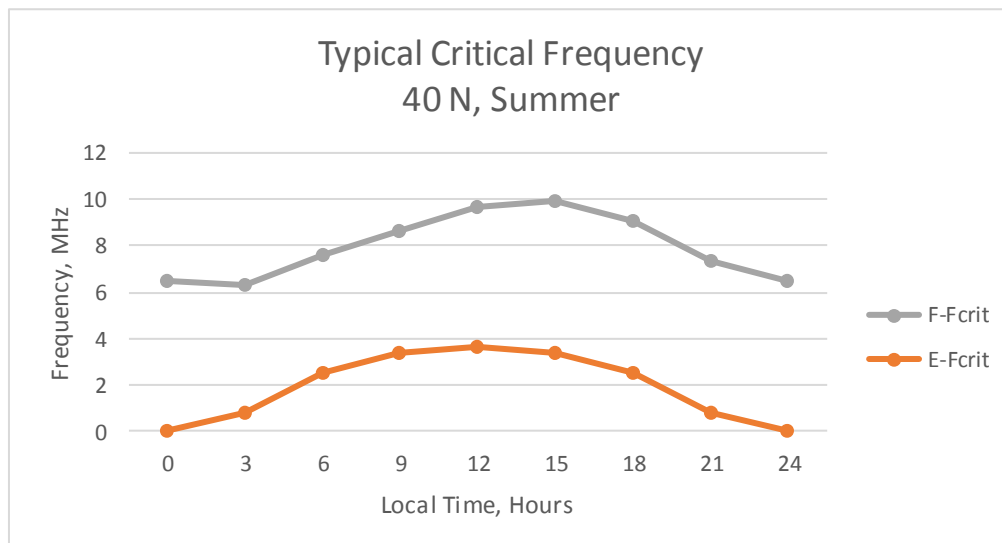

Figure 1 - Typical Critical Frequencies for E- and F- layers, summertime, 40 degrees N, Solar time, Sunspot Number = 86
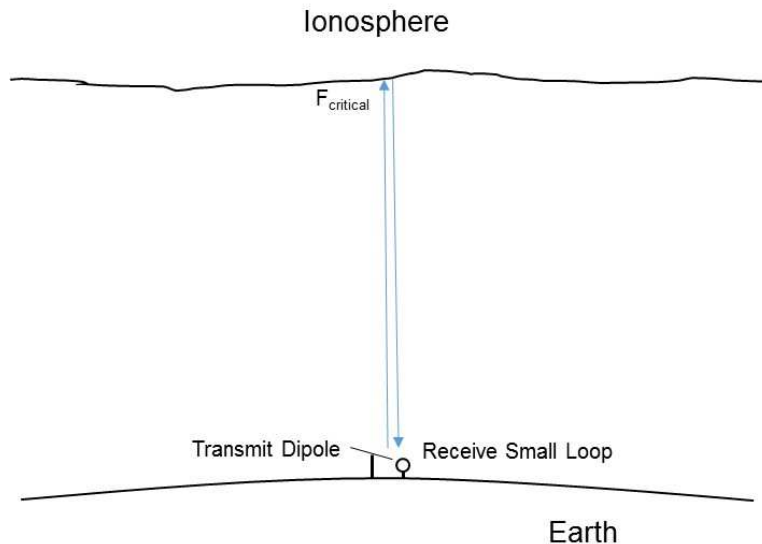
Figure 2 - Basic setup - Transmitter and Receiver Co-located.

The ionosphere reflects vertically incident signals below the critical frequency. The time-of-flight of the go plus return signal indicates how high the ionospheric layer is. Additionally the ionosphere layer may have a vertical velocity (either upwards or downwards) that induces Doppler shift onto the reflected signal. Figure 2 shows the basic setup –transmit and receive antennas are located within about 100 feet of each other – this is known as a monostatic radar configuration.

Chirp Measurement Approach

Practical measurements pose some difficult requirements because the echo is delayed less than one millisecond (E-layer) or about 1.6 milliseconds (F-layer). The monostatic approach also means that the transmit signal will be much stronger than the received signal, thus the receiver dynamic range must be large.

The approach chosen for these experiments was to transmit a linear FM chirp signal and correlate the received signal against the signal used to drive the transmitter (matched filter approach). The number of correlation taps is large in order to provide enough dynamic range and time resolution to see echos approximately 100 dB below the transmit signal. This approach requires full-duplex equipment –the transmitter and receiver are operational simultaneously. If the transmitter and receiver are co-located the transmit signal tends to overload the receiver which is listening to the same frequency at the same time as the transmitter is operating. The current experiment uses co-located Tx/Rx (same circuit board) and Tx/Rx antenna having about 20 meters separation. A similar experimental approach using 60 meters of Tx/Rx antenna separation was previously demonstrated by the Institute of Solar-Terrestrial Physics[1]. The low-phase-noise and good ADC dynamic range performance of the Hermes receiver helps minimize receive noise that would otherwise obscure the desired receive echo[2]. Chirp modulation is discussed in some recent amateur radio literature[3].

A linear FM-chirp signal is a constant-amplitude signal that sweeps in frequency at a constant rate. At the end of the sweep the signal returns back to the start frequency and sweeps again. For example, an up-chirp could sweep from -f_d KHz (below the channel center frequency) to +f_d KHz (above the channel

center frequency), then 'snap' back to -f$_d$ kHz and start again. It's also possible to turn off the transmit signal for a short period of time during the retrace. The turn-on and turn-off parts of the signal are amplitude ramped with a raised-cosine waveform in order to prevent spectral transients. A down chirp is just the opposite, it starts at a frequency above the channel center and sweeps down at a constant sweep rate to below the channel center frequency.

The received signal is correlated against a stored version of the transmit signal. In effect the DSP correlation filter is performing as a matched filter of the chirp signal. When using a chirp to measure the ionosphere we are interested in searching for a weak replica of the chirp delayed by some amount of time related to the propagation delay, equipment delay, and frequency shifted due to the Doppler shift induced by the vertical movement of the ionosphere. Doppler shift of the received echo has the effect of appearing to alter the time delay (and thus the virtual height measurement) of the received signal. The effect of Doppler is equal and opposite for an up-chirp signal compared to a down-chirp signal. By transmitting both kinds of chirps, and analyzing them independently, we can compensate for the Doppler-induced range (height) error and additionally, measure the amount of Doppler shift induced thus allowing computation of the vertical velocity of the ionosphere. Figure 3 shows reception of an Up-chirped signal with no Doppler shift. Figure 4 shows reception of an Up-chirped signal with (+) Doppler shift, and Figure 5 shows reception of a Down-chirped signal with (+) Doppler shift.
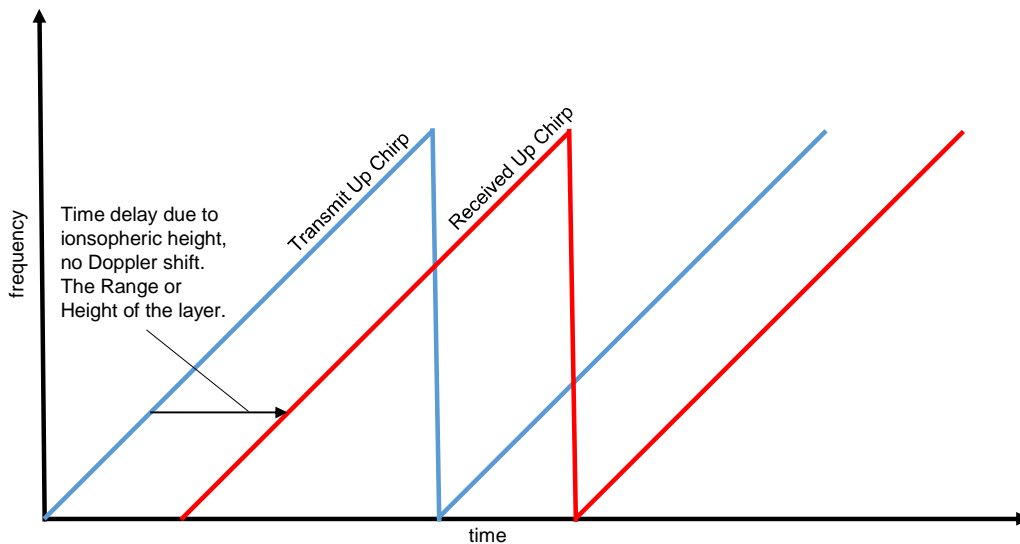

Figure 3 – Up Chirp reception with no Doppler shift.

Notice that the Up-Chirp and Down-chirp exhibit opposite range errors. This allows us to resolve the correct range and the amount of Doppler shift. The amount of range error caused by Doppler shift is dependent on the chirp rate.

The range to the ionosphere (the height) is proportional to half of the round-trip time.

$$Range\ (Height) = c * \frac{t_{echo}}{2}$$

(1)

Where *c* is the speed of light in m/s, and *t* is the time of the echo, in seconds. To compute the range error caused by Doppler, the formula is:

$$Range\ Error = c * \frac{d}{s} \tag{2}$$

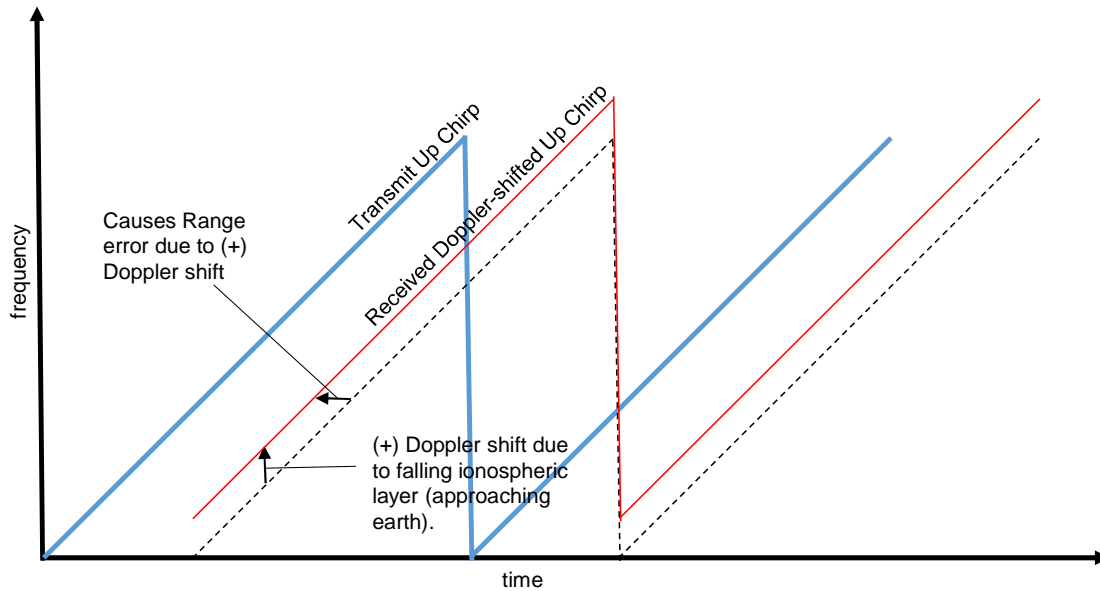Where *d* is the Doppler shift in Hz, and *s* is the chirp sweep rate in Hz / second.

Figure 4 - Up Chirp Reception with Positive Doppler Shift due to falling ionospheric layer.

For example a chirp rate of 15,000 Hertz/second implies an equivalent range error of 19.919 km per Hertz of Doppler shift. The vertical velocity of the ionosphere is measured by the induced Doppler shift which we infer from the range error.

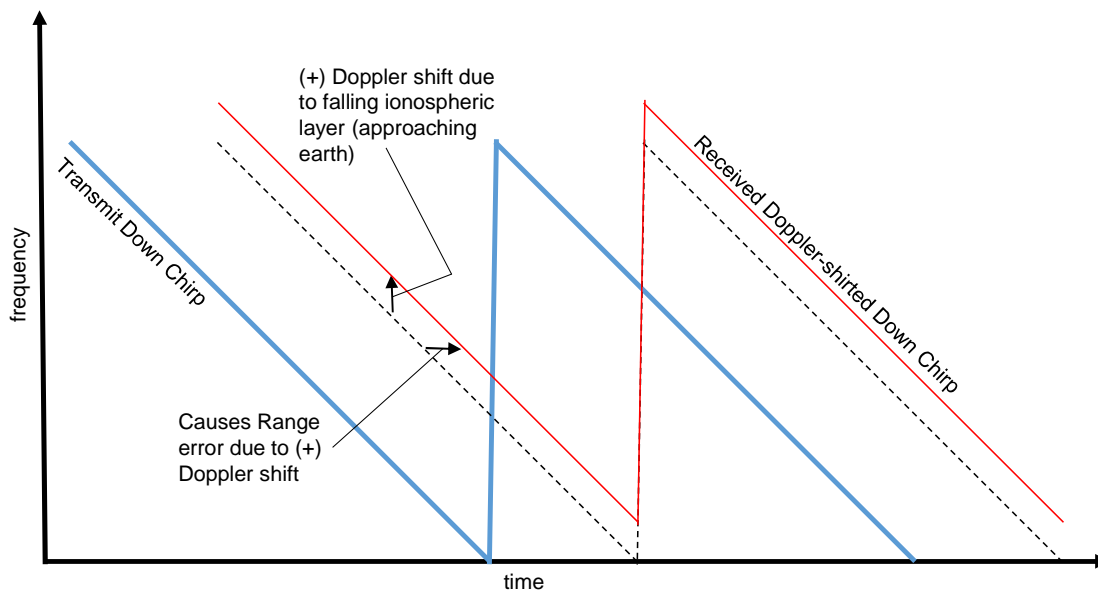$$Doppler\ Shift = s * \frac{Range\ Error}{c} \tag{3}$$

Figure 5 - Down Chirp Reception with Positive Doppler Shift due to falling ionospheric layer.

Note that the vertical movement of the layer induces a doubled Doppler shift. The chirp signal hits a moving ionosphere layer, thus being Doppler shifted as received by the ionospheric layer. Then the signal is re-emitted by the ionosphere and received back at the ground thus inducing another Doppler shift. Doppler shift is related to the layer velocity and the frequency of the radio wave, adding the factor of two for reflection from a moving ionosphere.

$$\Delta f = 2 * \frac{\Delta v}{c} * f_o \tag{4}$$

We compute the velocity of the ionosphere as:

$$\Delta v = \frac{\Delta f * c}{2 * f_o} \tag{5}$$

Relating the velocity to the measured range error:

$$\Delta v = \frac{s * Range\ Error}{2 * f_o} \tag{6}$$

We measure the range error by measuring the difference between the ranges determined by the up-chirp and the down-chirp time measurements. Since each chirp introduces an equal and opposite error, the actual range error is the difference divided by two.

$$\Delta v = \frac{s * (UpRange - DownRange)}{2 * f_o} \tag{7}$$

A slower chirp rate yields higher sensitivity to Doppler shift allowing more resolution of the ionosphere Doppler and thus the ionosphere vertical velocity.

Signal Processing

The basic receive algorithm consists of cross-correlating the received signal against a stored replica copy of the transmit signal. This is known as a matched filter. It provides a couple of benefits:

- The actual transmit signal strongly correlates with its own replica providing a convenient way to compensate for fixed equipment delays.
- A weak echo is easily seen above the background noise even in the presence of a strong transmit signal.
- The delay of the echo signal is the difference in time between the received transmit peak and the received echo peak.

This removes the requirement of knowing the absolute delay through the radio, Ethernet switch, and DSP processing – such errors or unknowns are subtracted out.

The DSP algorithm that correlates the received signal with the replica is factored through several steps in order to improve the computational efficiency. We define the two signals, f(t) (transmit replica copy) and g(t) (received signal). The circle-plus means convolution, and the circle-cross means correlation. The convolution of f(t) with g(t) is defined as:

$$f(t) \oplus g(t) = \int_{-\infty}^{+\infty} f(\tau) * g(t - \tau) \, d\tau \tag{8}$$

Convolution is implemented in DSP using an FIR filter kernel. A ready-made DSP block exists within Gnuradio that directly implements an FIR filter. Correlation is closely related to convolution, the filter taps need only be time-reversed to implement cross-correlation. The cross-correlation of f(t) with g(t) is defined as:

$$[f \otimes g] \, (t) = \int_{-\infty}^{+\infty} f(-\tau) * g(t - \tau) \, d\tau \tag{9}$$

We need only time-reverse the stored replica copy of the transmit chirp signal before introducing it as the taps of the FIR filter. It's less efficient to time-reverse the receive signal because it would require buffering the received signal first. We don't need to write any code, we can simply use Gnuradio's existing FIR filter and just read in the taps from a file containing the previously stored time-reversed chirp signal.

It is not possibly to actually compute an infinite number of taps, but some finite number of taps. Unfortunately an FIR filter requires on the order of N^2 operations, abbreviated $O(N^2)$. This means that if we try to implement a correlation of 1 million taps (10^6), the correlation operation requires on the order of one trillion (10^{12}) computation cycles which is infeasible. Both the signal and the taps are complex numbers (I and Q), requiring at least four floating-point multiplies and two additions per tap.

Fortunately there is a much more efficient way to implement the FIR filter in the frequency domain using the FFT (Fast Fourier Transform). This is called an FFT filter and it is also a built in block in Gnuradio, so we don't need to write it. The correlation can be implemented by taking the FFT of f(-t) and the FFT of g(t) and then pair-wise multiplying each element of f and g. Finally we take the inverse FFT of the result to get back to the time domain. This sounds more complicated than a direct FIR, but the FFT operation is extremely efficient such that the FFT filter requires far fewer computations, on the order of $O(N \log_2 N)$ operations. For a 1-million element correlation, this is on the order of 18 million operations (2 FFT and 1 IFFT) compared to one trillion operations using the direct FIR filter.

For convolution,

$$f(t) \oplus g(t) = IFFT \, [\, FFT(f(t)) * FFT(g(t)) \,] \tag{10}$$

Similarly to an FIR filter used for correlation, the FFT filter can be used for correlation by time-reversing the waveform used for the filter taps.

$$f(t) \otimes g(t) = IFFT \, [\, FFT(f(-t)) * FFT(g(t)) \,] \tag{11}$$

Properly constructed linear chirp signals have several interesting symmetry properties: an up-chirp signal is the frequency conjugate of the down-chirp signal. An up chirp signal is also the time reverse of a down chirp signal and vice versa. This means we don't even need to bother time-reversing the stored chirp signal. To correlate a receive echo, we just need to load the FFT filter taps (or FIR filter taps) with the opposite type of stored transmit chirp.

$$f(t) \otimes g(t) = IFFT \, [\, FFT(opposite \; chirp(t)) * FFT(g(t)) \,] \tag{12}$$

This series of steps (leading to Equation 12) reduces the computational effort required to correlate the received signal against the stored transmit signal by an extremely large amount. At 384 k samples per second, a Core i7-3770 (3.4 GHz) processor can easily keep up with the 1-million point correlation in real time in Gnuradio. In fact, several can be run in parallel to directly compare various algorithm tradeoffs.

Figure 6 is a block diagram of the DSP steps implemented in the flowgraph. The FFT taps utilize a stored version of the chirp waveform created at 384 ksps, so no decimation is performed in the receive chain of the active flowgraph. There are very few steps required. The time domain output of the correlation filter is stored as a file on disk (File Sink) for later post processing in Python (receive integration).

Receive Lowpass Shaping Filter (Windowing in the Time domain)

In the previous figure the signal received from Hermes is first lowpass filtered in the frequency domain before being applied to the correlation function. This lowpass filter has a gentle shape defined with a Blackman-Harris window. The shaping in the frequency domain results in the time-domain samples in the correlation filter being windowed as though by a time-domain window because a linear chirp is being received (the frequencies at the extreme positive and negative ends of the chirp are the most attenuated). Without this windowing, spectral leakage would obscure the echos. Figure 7 shows the correlator output sidelobes with and without receive lowpass filtering. The time delay of the Blackman-Harris filtered and the unfiltered signals been approximately normalized with a delay element to roughly time-align the two to ease visual comparison in Figure 7.

Receive Integration

While the raw algorithm achieves about 110-120 dB of dynamic range, overloading of the receive antenna amplifier and other parts of the receiver degrades the dynamic range to about 90 dB. In order to bring the received signal up out of the noise, about 10 sweeps of the receive signal are recorded on disk. Then the signals are non-coherently averaged. This brings the F-layer echos clearly up out of the noise level.

Figure 6 is the Gnuradio flowgraph used to capture and real-time process the signal. A custom Gnuradio block was written to generate a programmable chirp signal. A number of parameters were included in that block to permit adjusting the frequency deviation, sweep rate, number of samples per sweep, and providing a raised-cosine start and stop shape. The Chirp block feeds the Hermes transmitter port. The HermesNB module was written to provide access to many features of Hermes and Alex, and has been previously described[4].

The Hermes FPGA code also filters the transmit signal, it limits the maximum frequency response of the transmitter to ±20 KHz of the transmit center frequency.

The reason that the echos are non-coherently integrated is that the Doppler shift induced onto the receive signal means that the relative phase of the receive echo changes each sweep of the chirp. If coherently integrated the echos would average towards zero. Non-coherently we just integrate the magnitude of each echo (neglecting phase). A 3 dB improvement in SNR should be possible through echo phase de-rotation and coherent integration.
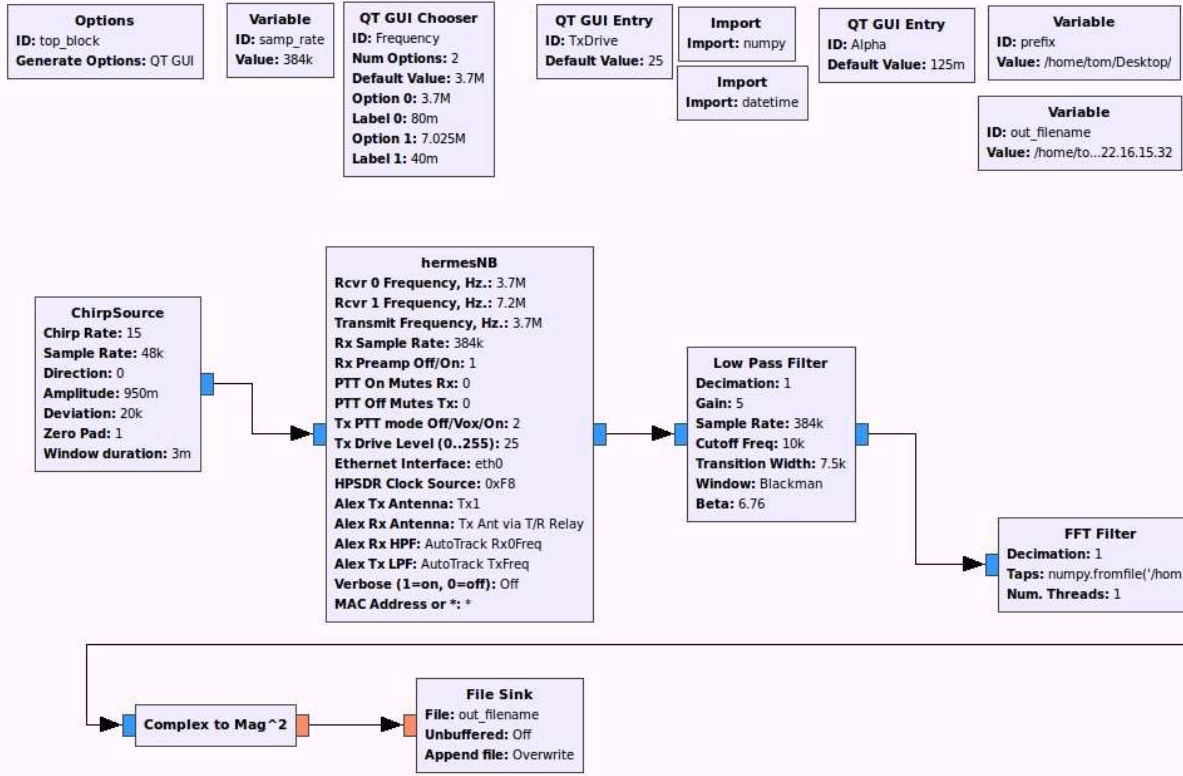
Figure 6 - Block diagram of the Gnuradio chirp transmit & receive data processing flowgraph.
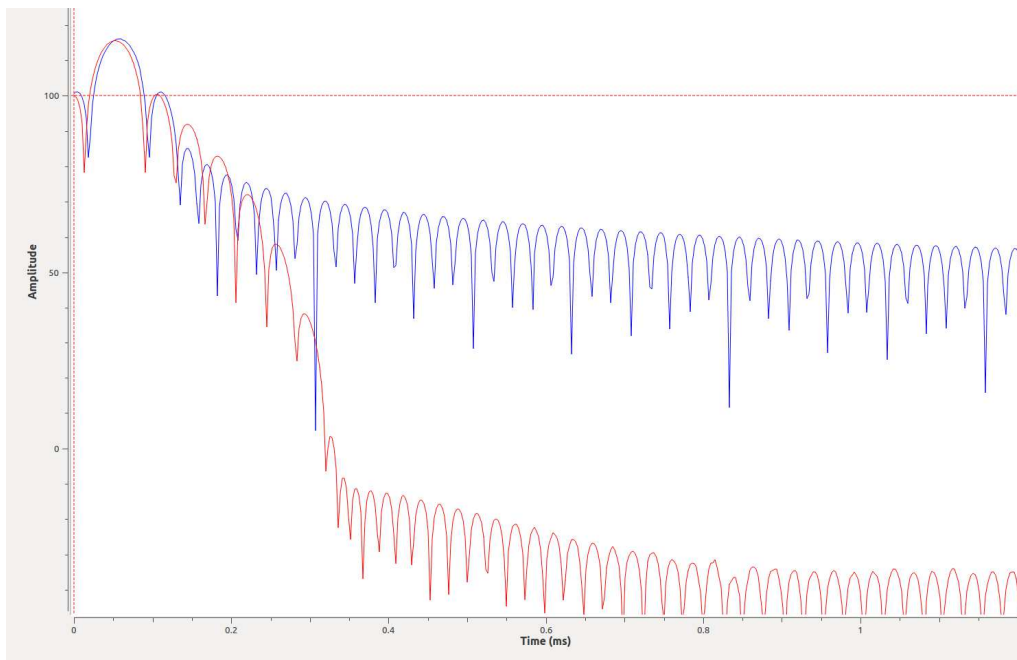

Figure 7 - Rectangular (no low pass filter) has high residual correlation sidelobes. A Blackman-Harris lowpass filter reduces the sidelobes out past about 0.3 milliseconds. The vertical scale is -50 to +120 dB.

Block Diagram

Figure 8 is a block diagram of the test setup. The Alex module is used as the transmit bandpass filter, however it is not used in the receiver path due to insufficient isolation between the transmitter and the separate receive connector on the Alex module. Input to the Hermes receiver had to completely bypass the Alex module. Gnuradio sends a FM chirp signal of constant amplitude to the transmit section of Hermes, and then to Munin (broadband PA) where it is amplified to about 20 watts output power. The amplifier output is filtered by the Alex RF filters, and sent to an antenna tuner and ladderline and a 40m dipole. For F-layer echos, the transmit signal is about 3.6 MHz.
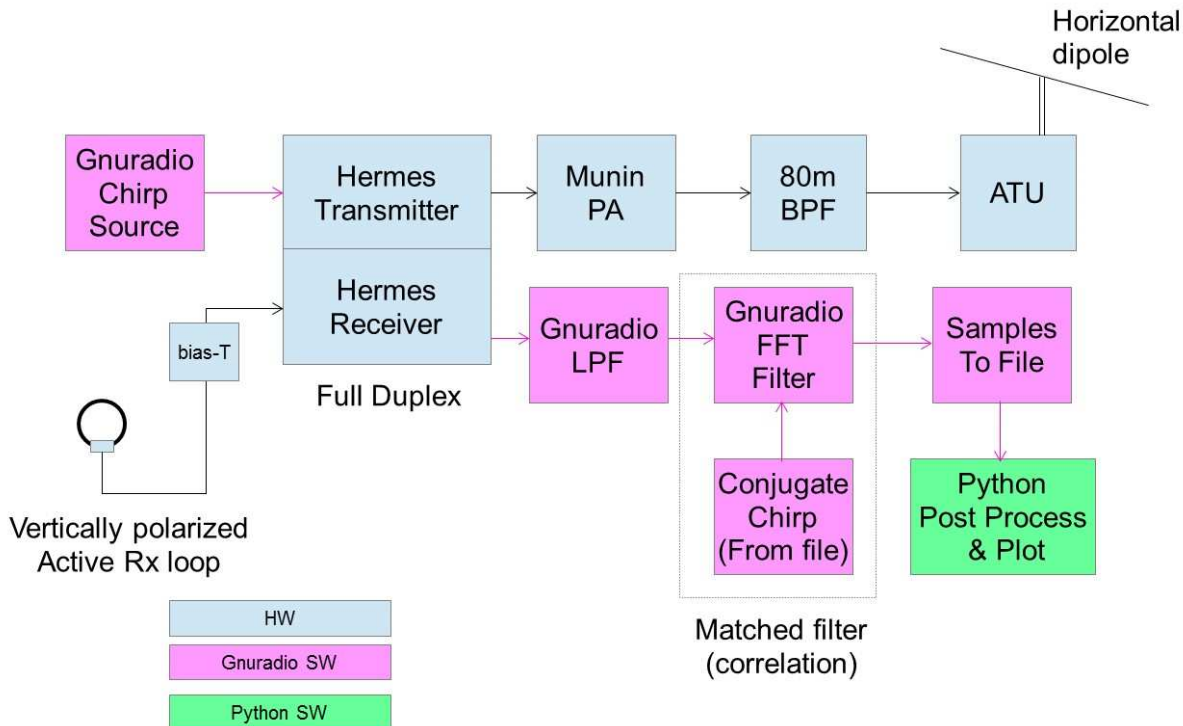

Figure 8 - Block Diagram of Test Setup. Box colors indicate functions implemented in Hardware, Gnuradio software, and Python software.

Due to the high SWR on the ladderline, about 6 watts of transmit power is actually radiated by the antenna, while 14 watts is absorbed by the feedline loss. On receive a homebrew active loop antenna 1 meter in diameter feeds a differential amplifier and balanced-to-differential transformer through a common mode choke. The antenna is remotely powered over the RG-6 feedline. After reception the Hermes signal is filtered in Gnuradio through a baseband lowpass shaping filter (to window the samples in time) before being sent to the FFT correlator.

Figure 9 is a photograph of some of the components of the experimental setup, while the RF Alex bandpass filters and the Core i7 Linux computer running Gnuradio are not shown in the photograph.

Figure 10 is a photograph of the active receive loop antenna. Only one of the two loops are used in this experiment. The loop antenna is vertically polarized, helping to reduce coupling to the horizontally polarized transmit antenna. The antenna was constructed with future experiments in mind where it

should be possible to discriminate between the Ordinary-ray (O) and eXtraordinary-ray (X) reflected by the ionosphere.
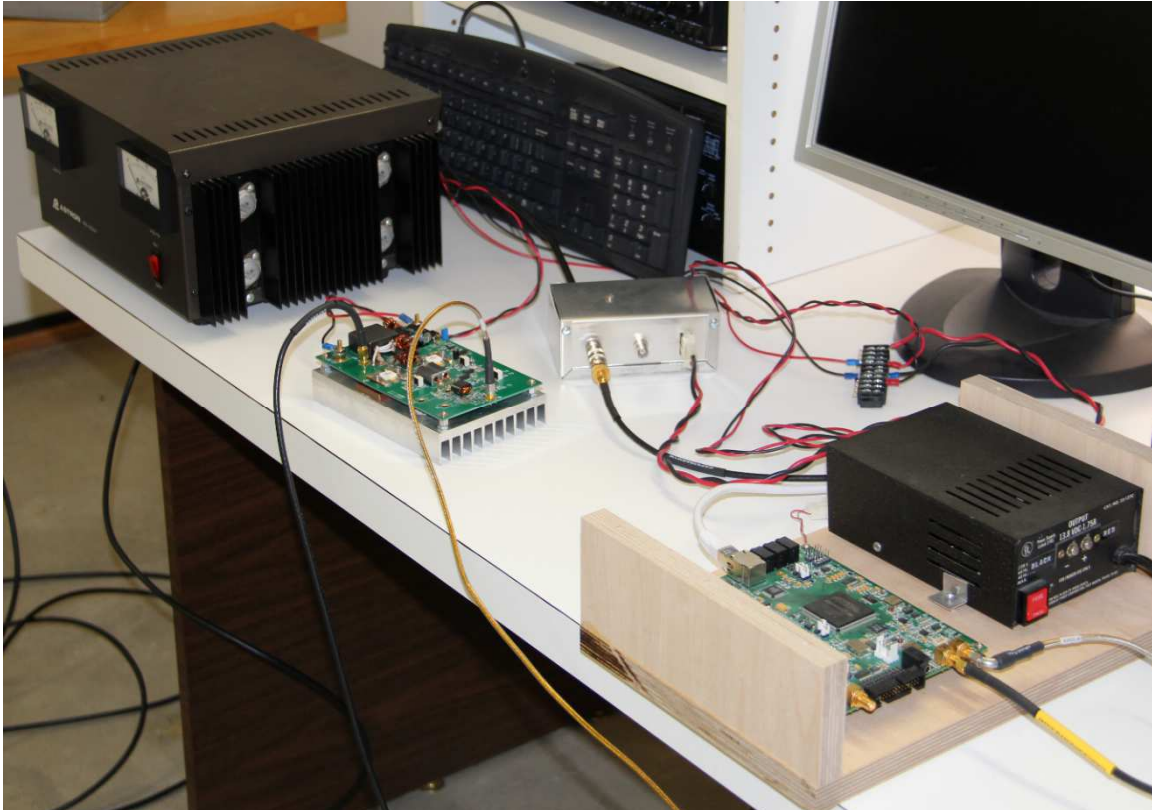

Figure 9 - Photograph of the test setup, showing DC Power Supply, Munin amplifier, Active loop bias-T, Hermes board. Not shown: Alex RF bandpass filter, Computer (outside the photo).

Figure 10 - Homebrew dual active-loop receive antenna (vertically polarized). Only one of the two loops is used in this experiment.

Results

So far measurements have been made on the F-layer at several times of day. The tests were conducted at about 3.6 MHz channel center frequency. Before dawn in fall / winter this is above the critical frequency, and no vertical reflections were received. In the evening local time the critical frequency is usually higher than 3.6 MHz, and vertical reflections have been received.

Figure 11 is a graph of the up-chirp (blue) and down-chirp (green) signals. It will be noticed that the Up- and Down- chirps exhibit range difference due to the Doppler shift of the moving ionospheric F-layer. This figure represents about ten sweeps that have been non-coherently integrated. The vertical axis is the magnitude of the correlation, in dB while the horizontal axis is time in seconds (spanning 0 to 5 milliseconds). The transmit peak has been adjusted to zero time by the post-processing Python integration program. The primary reflections are seen at about 1.7 milliseconds after the transmit peak. There is a spurious peak at 4 milliseconds (and multiples of 4 milliseconds) for both sweeps, the cause has not yet been determined.
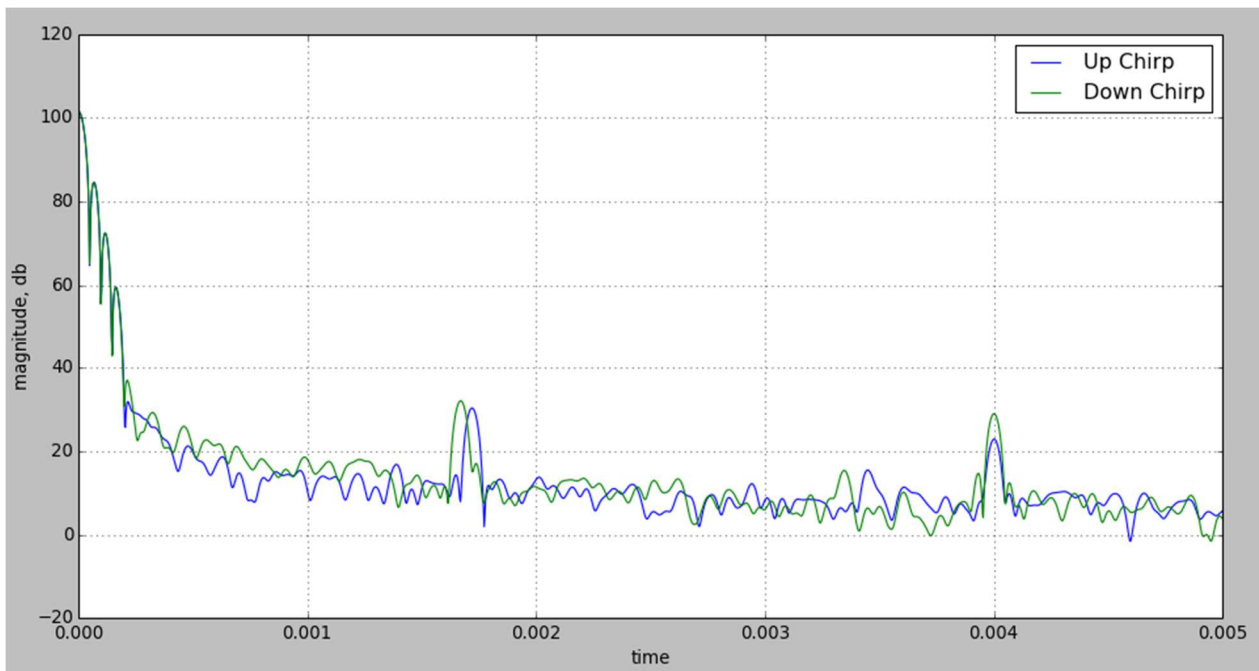

Figure 11 - F-layer reflection at 3.6 MHz in the evening local time. This figure is the correlation integral output after post-processing and integrating about 10 sweeps with Python software. The vertical axis is dB, the horizontal axis is delay time in seconds. The signals near 1.7 milliseconds are the F-layer reflections. Signals near 3.4 milliseconds are double-transit reflections. The signals at 4.0 milliseconds are system artifacts.

At about 3.4 milliseconds, small peaks in the Up- and Down-chirps can be seen. These are double-transit reflections – the signals traveled up to the ionosphere, down to ground, reflected by the ground back up to the ionosphere, and reflected back downward a second time. Notice that the Doppler range error is doubled as well.

The Doppler induced range errors indicate that the F-layer of the ionosphere is ascending at the time of this measurement.

The average of the two time measurements is the actual F-layer height, while the difference between the two is proportional to 4 times the Doppler shift. The moving ionosphere reflects the signal, thus doubling the induced Doppler shift, and the up and down chirps have opposite range errors induced. Thus the measured range difference represents quadruple the Doppler shift.

Calculations from Fig 11 show the F-layer height at 254 km, and the layer velocity at +15.4 meters per second (upward). The Doppler shift is about -0.38 Hertz. Both range and Doppler resolution is limited by the filters and the correlation width. The half-lobe width is about ±17 microseconds, implying a range bin size of roughly 5 km.

Further Work

Figure 12 shows F-layer reflections that include reception of both the Ordinary (O) wave and the Extraordinary (X) waves. With a single receiver and linearly polarized receive antenna it is not possible to know which is O and which is X. The current single linearly polarized receive antenna sums the Right Hand Circular (RHC) and Left Hand Circular (LHC) components into one received signal. The dual-receive crossed linear loops (previously shown) plus two phase-coherent receivers (not available to the author at this time) could be used to capture two signals, one per antenna. Then the Gnuradio DSP software would be able to synthesize the right-hand and left-hand circular signals from the two linearly polarized receive signal components.
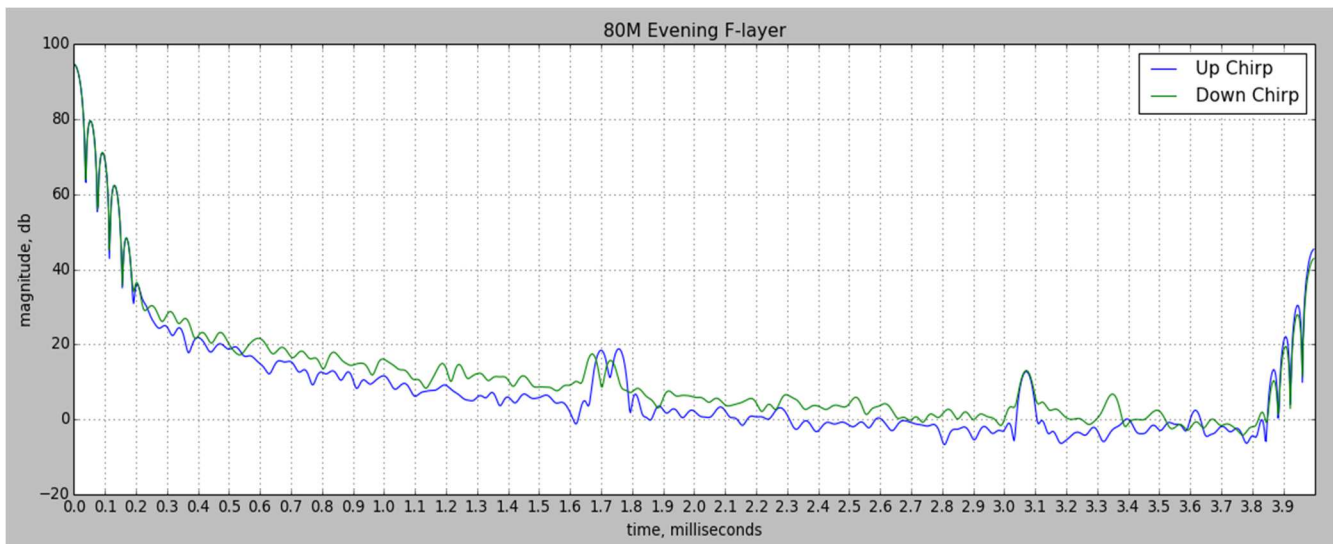

Figure 12 - F-layer reflection at 3.6 MHz showing Ordinary (O-wave) and Extraordinary (X-wave) reflection components from the F-layer near 1.7 milliseconds. The signals at 3.06 and 4 milliseconds are spurious artifacts.

Circularly polarized signals identify which reflection is O and which is X. Since the transmit antenna is linearly polarized, it emits a RHC plus a LHC signal simultaneously. These two signals remain coupled as one linearly polarized signal until encountering the ionosphere undergoing magnetic bias from the Earth's magnetic field. This causes them to de- couple into independent LHC and RHC components, which are the O and X waves. The effective ionospheric refractive index is different for the RHC and LHC components, which thus propagate with different characteristics in the ionosphere and are received as two different reflections.

We can construct the RHC and LHC components from the two received signals. Designating these two received signals as RX_a and RX_b (each complex valued) then:

$$RHC = RX_a + e^{j\pi/2}RX_b \qquad\qquad\qquad (13)$$

$$LHC = RX_a + e^{-j\pi/2}RX_b \qquad\qquad\qquad (14)$$

LHC and RHC can be generated at baseband using standard Gnuradio complex multiply blocks.

E-layer reflections have not been received at this time due to lack of appropriate transmit antenna.

Thanks to Andrew Martin, VK3OE, John Petrich, W7FU, and Phil Harman, VK6PH for helpful review and comments.

[1] "Vertical ionosphere sounding using continuous signals with linear frequency modulation" A.V. Podlesny, V.I. Kurkin, A.V. Medvedev, K.G. Ratovsky, Institute of Solar Terrestrial Physics, Irkutsk, Russia, Conference General Assembly and Scientific Symposium 2011, Istanbul, IEEE 2011. INSPEC Accession Number: 12354018

[2] Phil Harman, VK6APH. 2010. "Software defined radio: The Hermes state of the art single board SDR transceiver". RadCom 86(05):28-29.

[3] Phil Harman, VK6APH. 2012. "Chirp Modulation: A sophisticated radar-like technique for propagation study that makes 100W act like 100 megawatts" RadComm 2012(3):32-38

[4] "Gnuradio Companion module for openHPSDR Hermes / Metis SDR Radio", Tom McDermott, N5EG, Proceedings of the 32nd ARRL and TAPR Digital Communications Conference (2013), pp 36-42.