Walter's original email about the pcal extraction algorithm is copied in below.

There are also a couple of old documents from Sergei Pogrebenko which detail it EVN Doc 2 EVN Doc 8

A key parameter is the greatest common factor of the sample frequency and the frequency offset. Ths number leads to a number I will call N that tells you how many baseband data samples are required before the phase cal time series repeats itself exactly yielding the same phase.

The algorithm I envision works as follows:

1. create a real-valued accumulator array, A, with dimension N.

2. for the duration of an accumulation period, fold the voltage data, V, into this array:

for(i = 0; i < nsamples; i++) A[i % N] += V[i]

3. Perform a real-to-complex Fourier transform on this array into an N/2 element complex array B

B = FFT[A]

4. Extract the tones wanted. The B array has N/2 points, but one will only be interested in a very small subset of those points – the ones the correspond to the pulse cal tone frequencies. Note that it _might_ be faster to do a DFT only for the desired points rather than performing a full FFT to generate the entire spectrum.

As a concrete example, lets consider the VLBI case with a single baseband channel, bandwidth bw = 2^n MHz, with 1 tone per MHz, with the first tone nu = 10*d kHz offset from the DC edge of the channel for integer d (note this is a standard case for VLBI).

The condition for return-to-phase is:

delta phi = 2*pi*nu*deltaT = R*2*pi for integer R

deltaT must be an integral number of samples which we call N (the same N as above). Each sample is 2^(-n-1) microseconds in duration. Thus

d/100 MHz * N*2^(-n-1) mus = R

or

d * N = 200 * 2^n * R

This expression is where the greatest common divisor concept comes in. It is easy to see that N = 200 * 2^n will meet the required condition for all values of d, meaning for all possible tones in the baseband. For 8 MHz this implies N = 1600 points which forms a spectrum with 800 complex points. There will be 8 tones to recover, each separated by 1 MHz = 100 points in the spectrum. The position of the first tone is at spectrum bin 800*nu/bw . The remaining 792 points of the spectrum are to be discarded.

Note the key aspect of this is that all of the averaging is done in the time domain and a single transform to the frequency domain is needed. There might be some logistical reason for doing the FFT and tone selection once per sub-integration in order to minimize data transfer from core processes to the manager process.