Optimum Bandpass Filter Bandwidth for a Rectangular Pulse

Mark A. Richards
July 2015

1 Introduction

It is well-known that for a radar waveform $x(t)$ in additive white noise of power σ_w^2, the optimum receiver frequency response, in the sense of maximizing the signal-to-noise ratio (SNR) in the receiver output signal $y(t)$ at a particular instant T, is the *matched filter* impulse response $h(t) = \alpha \cdot x^*(T-t)$. The corresponding frequency response is $H(F) = \alpha \cdot X^*(-F)$. The constant α is arbitrary; it affects only the overall gain, but has no impact on the SNR achieved or the shape of the receiver frequency response. Henceforth we will assume α has been chosen to make the peak gain of $H(F)$ equal to 1.

The definition of SNR is the square of the peak signal voltage at $t = T$, divided by the output noise power σ_{out}^2 obtained by integrating the filtered noise power spectral density over all frequency:

$$\chi \equiv \frac{|y(T)|^2}{\sigma_{\text{out}}^2} = \frac{|y(T)|^2}{\sigma_w^2 \int_{-\infty}^{\infty} |H(F)|^2 \, dF} \tag{1}$$

When the matched filter is used, the peak SNR achieved can be shown to be $\chi_{\text{MF}} = E/\sigma_w^2$, where E is the signal energy

$$E = \int_{-\infty}^{\infty} |x(t)|^2 \, dt = \int_{-\infty}^{\infty} |X(F)|^2 \, dF \tag{2}$$

If $x(t)$ is a simple rectangular pulse of unit amplitude and duration τ, $E = \tau$ and $\chi_{\text{MF}} = \tau/\sigma_w^2$. See [1] for details of the derivation of these and related results.

The problem of interest here is what happens when the signal is a rectangular pulse, but the receiver filter is, as may be more likely, a bandpass filter (BPF) with cutoff frequency $\pm B$ Hz? As the BPF cutoff B increases from zero, noise is added linearly to the output, $\sigma_{\text{out}}^2 = (\sigma_w^2)(2B) = 2B\sigma_w^2$. However, signal energy is *not* added linearly. As B increases, the energy in the mainlobe of the pulse's sinc-shaped spectrum increases signal output energy fairly rapidly, but after a point, only the sinc sidelobes are passed to the output, and the rate of signal energy growth will slow. Consequently, there is some choice of B that produces the maximum SNR at the output. A narrower cutoff filters out too much signal energy; a wider cutoff allows too much noise energy to pass. The goal of this memo is to determine this optimum BPF cutoff B_{opt} analytically, and to confirm the results via simulation. The solution is not new. In fact, it was published in the earliest days of radar by North [2]. Our Figure 3(b), to appear later in this memo, duplicates the U_1 curve in North's Figure 3. Skolnik [3] cites North's result and reproduces the key figure, and also gives a table of similar optimum parameters for various combinations of rectangular and Gaussian pulse shapes, and rectangular, Gaussian, or "RLC" tuned circuit filters.[1] In this memo, only the rectangular pulse and ideal (rectangular) BPF frequency response are considered.

2 Output SNR for a Rectangular Pulse Passed Through an Ideal Bandpass Frequency Response

The input signal pulse of interest is

$$x(t) = \begin{cases} 1, & -\tau/2 \le t \le +\tau/2 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

The input noise is complex white Gaussian noise (CWGN) of variance σ_w^2. The Fourier transform of $x(t)$ is

$$X(F) = \frac{\sin(\pi F \tau)}{\pi F} \quad (F \text{ in hertz}) \tag{4}$$

Note that the zeroes of this sinc function occur at the nonzero integer multiples of $1/\tau$ Hz, and the peak amplitude of τ occurs at F = 0 Hz. The frequency response of the receiver is modeled as an ideal rectangular bandpass filter with gain of 1 and a cutoff frequency of B Hz:

$$H(F) = \begin{cases} 1, & -B \le F \le +B \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

The corresponding receiver impulse response is

$$h(t) = \frac{\sin(2\pi B t)}{\pi t} \tag{6}$$

with peak amplitude $2B$ at time t = 0 and zeroes at the nonzero integer multiples of $1/2B$ seconds.

Figure 1 illustrates the relationship between the magnitudes of $X(F)$ and $H(F)$. The frequency scale in this figure covers the first three sidelobes of $X(F)$ to each side of the mainlobe. Note that the vertical (ordinate) scales are different. The maximum magnitude of $X(F)$ is τ (left ordinate scale), while that of $H(F)$ is 1 (right ordinate scale).

[1] Skolnik's book is currently in its third edition, but I cite the second edition on purpose because the third does not include the figure from North.

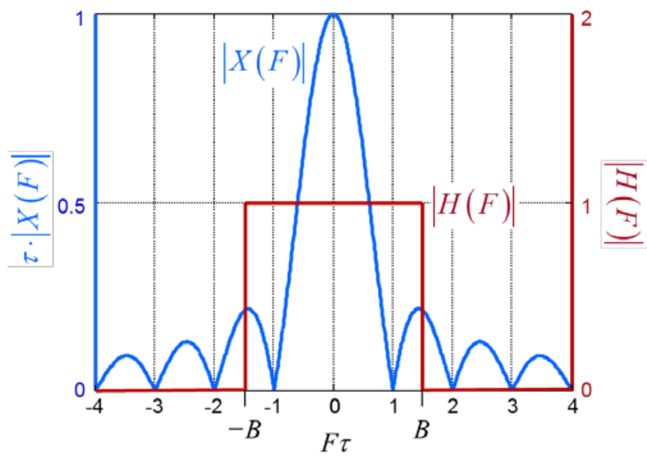


Figure 1. Magnitude of signal spectrum (left-side ordinate scale) and bandpass receiver frequency response (right-side ordinate scale).

The output signal component can be found either as the inverse Fourier transform of the output signal spectrum $Y(F)=H(F)\cdot X(F)$ or as the convolution of the pulse and the receiver impulse response. We take the latter path, though both proceed similarly. The output time domain signal is

$$y(t)=\int_{-\infty}^{\infty} h(z)\cdot x(t-z)\,dz=\int_{t-\tau/2}^{t+\tau/2}\frac{\sin(2\pi Bz)}{\pi z}\,dz \tag{7}$$

The relationship of the signals in this convolution is sketched in Figure 2.

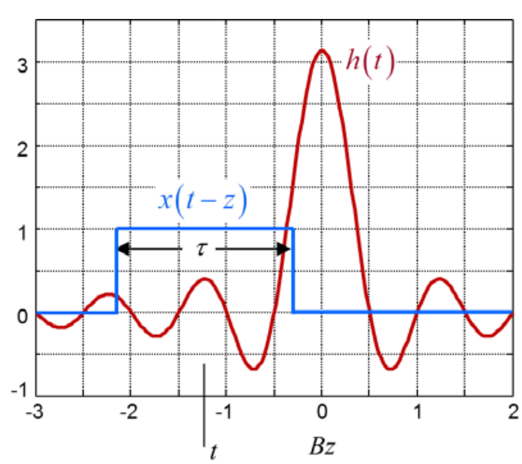


Figure 2. Sketch of the relationships of the functions in the convolution of Eq. (7).

To proceed, we need to identify the specific time at which $y(t)$ achieves its maximum, and evaluate y at that time to get the peak signal amplitude. One suspects that $t = 0$ may give the maximum output for a given pulse width τ since it appears to integrate the largest part of the sinc curve, but this is not certain since the sidelobes are both positive and negative. Simulations show that the peak output does occur at $t = 0$ for values of $B\tau < 1.42$, but that it occurs for some other positive value of t for larger $B\tau$. Fortunately, the restriction $B\tau < 1.42$ includes the optimal choice of bandwidth, as will be seen shortly. Therefore, assume for the time being that the maximum occurs at $t = 0$. Then

$$y_{\max} = y(0) = \int_{-\tau/2}^{\tau/2} \frac{\sin(2\pi Bz)}{\pi z}\,dz = \frac{2}{\pi}\mathrm{Si}(\pi B\tau) \tag{8}$$

where $\mathrm{Si}(x)$ is the "sine integral" function,

$$\mathrm{Si}(x) = \int_0^x \frac{\sin(t)}{t}\,dt \tag{9}$$

The peak signal power at the filter output is then $y_{\max}^2 = \dfrac{4}{\pi^2}\mathrm{Si}^2(\pi B\tau)$.

To get the SNR, we now need the noise power at the BPF output, which was noted earlier to be σ_{out}^2 = $2B\sigma_w^2$. The SNR at the output of the BPF becomes

$$\chi_{\text{BPF}} = \left(\frac{4\mathrm{Si}^2(\pi B\tau)/\pi^2}{2B\sigma_w^2}\right) = \frac{2}{\pi^2 B\sigma_w^2}\mathrm{Si}^2(\pi B\tau) \tag{10}$$

For a unit amplitude pulse of length τ the energy $E = \tau$, so the matched filter peak SNR is

$$\chi_{\text{MF}} = \frac{E}{\sigma_w^2} = \frac{\tau}{\sigma_w^2} \tag{11}$$

We can now define the *efficiency* of the BPF receiver for a rectangular pulse as the ratio of the SNR obtained with the bandpass receiver to the matched filter maximum SNR. The result is

$$\eta = \frac{\dfrac{2}{\pi^2 B\sigma_w^2}\mathrm{Si}^2(\pi B\tau)}{\tau/\sigma_w^2} = \frac{2}{\pi^2 B\tau}\mathrm{Si}^2(\pi B\tau) \tag{12}$$

Equation (12) is the desired result. If one defines the rescaled variable $\alpha = 2B\tau$ and renames $\eta \rightarrow U_1$, it is identical to Eq. (30) of [2].

3 Discussion

Figure 3 plots the efficiency of Eq. (12) versus the ***two***-sided bandwidth $2B$ in multiples of the sidelobe null spacing, $1/\tau$. (Because the matched filter output SNR does not depend on B, the BPF output SNR as a function of B follows these same curves, differing only in the absolute scale.) The abscissa range of part (a) of the figure encompasses 14 sidelobes to each side of the pulse spectrum mainlobe. Not surprisingly, the efficiency drops severely after the first couple of sidelobes. Parts (b) and (c) of the figure expand the axis to cover just the mainlobe and first sidelobe to each side. It is seen that a maximum efficiency of 82% of the matched filter SNR is obtained, corresponding to a loss relative to the matched filter of 0.84 dB. The peak efficiency occurs when the two-sided bandwidth of the receiver equals $1.37/\tau$ Hz. This is about 50% wider than the 3 dB two-sided bandwidth of $0.89/\tau$ Hz. The optimum cutoff frequency is half this, or $B_{\text{opt}} = 0.685/\tau$ Hz. This is less than the frequency of the first spectral null ($1/\tau$ Hz), so the optimum bandwidth includes only about two-thirds of the signal mainlobe. Note also that $B_{\text{opt}}\tau = 0.685 < 1.42$, so the peak output fo the receiver filter does occur at $t = 0$, a key assumption made in deriving Eq. (12).

Figure 3(d) reproduces Figure 3 from [1]. The curve labeled U_1 is for the case of a rectangular pulse and ideal BPF receiver frequency response. Comparing this curve to our Figure 3(b) shows that the results are the same. (Note that the span of the frequency scale is somewhat smaller in North's figure.)

It can also be seen from Figure 3(a) and Figure 3(b) that for cutoffs less than B_{opt} the increase in the efficiency and thus the output SNR is approximately proportional to B, while for cutoffs greater than B_{opt} they decrease approximately as $1/B$.[2] Heuristically, for small cutoffs, the output spectrum $Y(F)$ is approximately rectangular with a cutoff of B. The resulting output waveform $y(t)$ is then approximately a sinc function with a peak amplitude proportional to B, and therefore a peak power proportional to B^2. Since the output noise power is proportional only to B, the SNR increases proportional to B. For cutoffs greater than B_{opt}, the noise power still increases as B. However, once the cutoff includes all of the mainlobe, increasing it adds only increasingly small sidelobes to the output spectrum. The output spectrum therefore approximates the input spectrum, so that $y(t)$ approximates the input pulse. The output signal amplitude is then independent of B, so that the SNR becomes proportional to $1/B$.

Finally, Figure 4 compares the theoretical efficiency of Eq. (12) to the results of a simulation[3] that explicitly computes the output signal component and finds its maximum value without assuming that the maximum occurs at $t = 0$, since this is known to be incorrect for $B\tau > 1.42$ ($2B\tau > 2.84$ on the two-sided bandwidth axis used in these figures). In the region of the figure where the maximum efficiency occurs, the peak output does indeed occur at $t = 0$ and the agreement between simulation and theory is excellent. For larger values of $B\tau$, the actual time of maximum output power moves away from $t = 0$ and the curves diverge somewhat, with the maximum difference being 1.87 dB at $2B\tau = 2$, and the difference being less than 1 dB for most larger values of $B\tau$. Notice that the theoretical result *under*estimates the efficiency because it computes SNR using the signal output at $t = 0$, even when a bigger peak occurs later in time.

---

[2] Thanks to Dr. William A. Holm of Georgia Tech for pointing out these trends and their explanation.

[3] The simulation used to create these figures is included in the appendix to this memo.

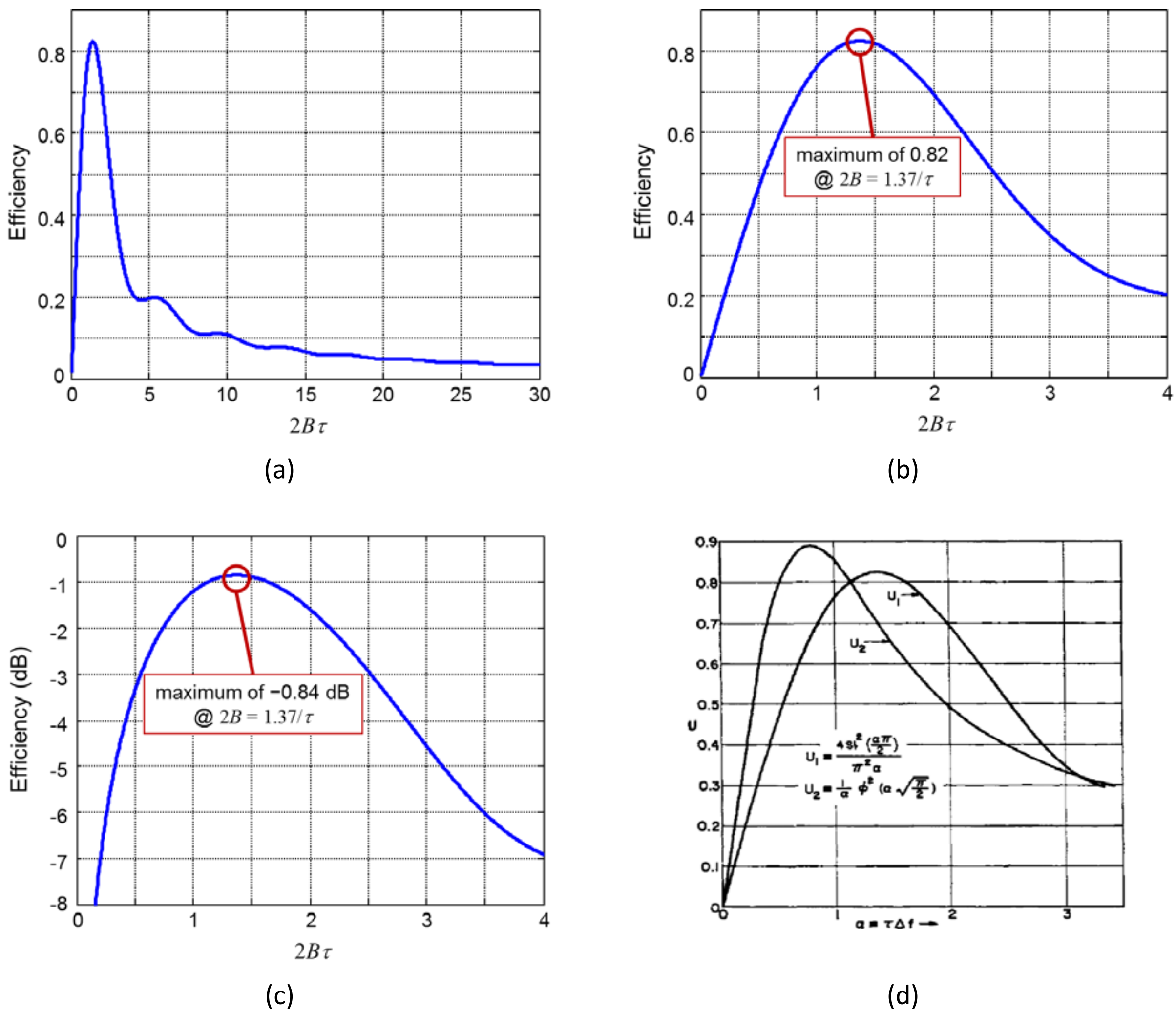
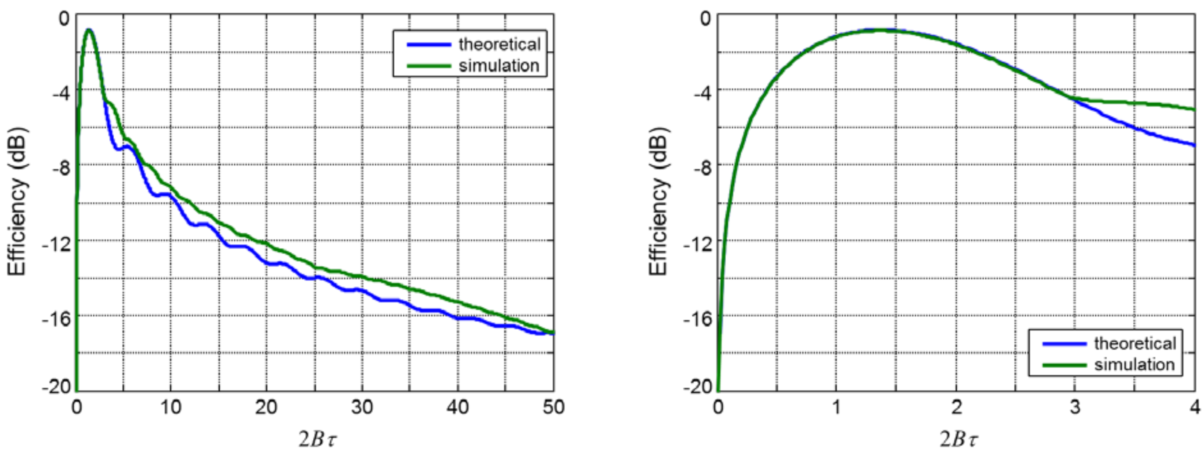


Figure 3. Ideal bandpass receiver efficiency for a rectangular input pulse. (a) Frequency scale includes 14 sidelobes to either side of the signal spectrum mainlobe. (b) Expanded frequency scale showing the region of maximum efficiency. (c) Same as (b), but on a decibel scale. (d) Reproduction of Figure 3 from [2]. Compare the "U_1" curve to part (b).



(a) (b)

Figure 4. Effect of the maximum output time assumption. (a) Comparison of simulation and theoretical results over a wide bandpass receiver cutoff frequency range. (b) Expanded frequency scale showing the region of maximum efficiency.

4 References

[1] M. A. Richards, *Fundamentals of Radar Signal Processing*, second edition. McGraw-Hill, 2014.

[2] D. O. North, "An Analysis of the Factors Which Determine Signal/Noise Discrimination in Pulsed-carrier Systems", RCA Technical Report PTR-6C, June 25, 1943. Reprinted in *Proceedings IEEE*, vol. 51, pp. 1016-1027, July 1963.

[3] M. I. Skolnik, *Introduction to Radar Systems*, second edition. McGraw-Hill, 1980.

5 Appendix: MATLAB® Simulation Code

```
% rect_BPF_output
%
% M-file for considering the effect of passing a rectangular pulse (in
% time) through a BPF with an ideal rectangular passband in frequency, and
% seeing how the peak amplitude of the result, and thus the peak
% instantaneous power vs. noise PSD version of SNR, will vary with the
% cutoff B.
%
% M. A. Richards, July 2015
%

% We'll work with positive frequency only because everything is even
% symmetric.
%
% Spectrum of a pulse of amplitude A and length tau is
% X(F) = A*sin(pi*tau*F)/(pi*F).
```

```
%
% Noise PSD is sig2 W/Hz.
%
% We pass the signal and noise through an ideal BPF with cutoff B Hz and
% gain = 1 (because gain doesn't matter). The one-sided noise power at the
% output will be B*sig2.
%

clear all
close all

A = 3.7;  % pulse amplitude; doesn't really matter
% Time-domain noise variance, used to define SNRs. I don't actually include
% noise here, it's not necessary, this is really just to let me check
% analytic formulas.
sigt2 = 4.3e-5;

% Set up the pulse length in samples. The number of sidelobes in the pulse
% spectrum (one side) will be ceil((Ntau-2)/2)
Ntau = 51;  % this has to be odd so we can  make it symmetric about the origin
K = 5001;  % DFT size; make it odd for symmetry convenience
Klast = (K+1)/2;

kz = K/Ntau;  % spacing of the DFT zeroes; not necessarily integer
Ftau = (0:K-1)/kz;  % frequency in multiples of (1/tau), i.e of the null spacing

% The time domain pulse. It has an energy of (A^2)*Ntau. Center it on time
% zero.
x = zeros(K,1);
Ntau2 = (Ntau-1)/2;
x(1:Ntau2+1) = A;
x(K-Ntau2+1:K) = A;

% The pulse spectrum.  It has a peak of A*Ntau.
X = fft(x,K);

% The SNR of the matched filter pulse would be this value.  This is peak
% output voltage of the matched filter signal component, squared (which is
% A^4*Ntau^2) and divided by average noise power of the matched filter
% output, which will be A^2*Ntau.  See (Richards, 2014) or similar.
SNRmatched = A^2*Ntau/sigt2;

figure(1)
plot(Ftau,abs(X));
grid
title('Spectrum of Pulse')
xlabel('frequency*tau')
ylabel('amplitude')
shg

% Now step though the possible values of BPF cutoff, compute the output
% waveform, and find its peak power.  We do NOT assume the peak output
% voltage occurs at t = 0 (k=1 in this code). The BPF is given a gain of 1.
peak = zeros(Klast-1,1);
index = zeros(Klast-1,1);
for k = 1:Klast
    Y = X;
    Y(k+1:K-k+1) = 0;  % zero out Y in the BPF's stopband region
    y = ifft(Y);
    % This commented-out section is used to examine the output waveform
    % under various circumstances if desired.
%     if k >= 130
%         figure(2)
```

```
%         subplot(1,3,1)
%         plot(Ftau,abs(X)); xlim([0,Ntau]); grid
%         subplot(1,3,2)
%         plot(Ftau,abs(Y)); xlim([0,Ntau]); grid
%         subplot(1,3,3)
%         plot(y); xlim([0,200]); grid
%         pause
%     end
    yy = abs(y).^2;
    peak(k) = max(yy);
    index(k) = find(yy==peak(k),1,'first');
end

% At this point, the array 'peak' contains the peak instantaneous power in
% the signal component of the filter output for each value of cutoff
% frequency.  Now we need a noise calculation to turn that into an SNR.

% Let's assume the noise variance (power) in the time domain is sigt2 (from
% earlier) and that noise is present in all K implied input samples of the
% DFT.  Of course we assume white noise. Then the average noise power in
% each  DFT sample is K*sigt2, and the total noise power in the spectrum is
% K^2*sigt2.  Now, if the BPF cuts off the spectrum at some index +/-k (as
% in the loop above), the total noise power in the spectrum will be reduced
% by the factor (2*k/K), cutting it to 2*k*K*sigt2. The corresponding total
% time-domain noise power is 2*k*sig2t, which means the time-domain noise
% power per sample is now 2*(k/K)*sig2t.  And that is the noise power that
% should be compared to the signal power to get an SNR, I think.

SNRbpf = peak./(2*((1:Klast)'/K)*sigt2);

% SNR  of filter output, divided by matched filter SNR, is an "efficiency"
% measure of the BPF compared to the matched filter.
eff = SNRbpf/SNRmatched;

figure(3)
plot(Ftau(1:Klast),eff)
grid
title('Efficiency (Relative SNR) of BPF vs. Matched Filter')
xlabel('1-sided bandwidth*tau')
ylabel('efficiency')
shg

% Now, North in Fig. 3 of his paper plots the efficiency vs. the quantity
% delF*tau.  His delF appears to me to be the same as twice my cutoff
% frequency (i.e., it is the two-sided bandwidth of the BPF). My Ftau
% variable above is the  cutoff frequency, i.e. the one-sided bandwidth.
% To match his figure, I therefore need to plot against 2*Ftau, thusly:
figure(4)
plot(2*Ftau(1:Klast),eff)
grid
title('Efficiency (Relative SNR) of BPF vs. Matched Filter')
xlabel('2-sided bandwidth*tau')
ylabel('efficiency')
shg

% Same thing on a decibel scale.
figure (5)
plot(2*Ftau(1:Klast),db(eff,'power'))
ylim([-5,0])
grid
title('Efficiency (Relative SNR) of BPF vs. Matched Filter')
xlabel('2-sided bandwidth*tau')
ylabel('efficiency (dB)')
```

```
shg

% Now let's do new plots in linear and dB space comparing the simulation
% and the theoretical result, and in terms of two-sided bandwidth again.
% The theoretical is based on an upper bound for signal value which we know
% is not met for larger values of F*tau, but it is true throughout the
% region where the overall max of the efficiency occurs. See the memo for
% details.
eta = 2*(sinint(Ftau*pi).^2)/pi^2./Ftau;  % theoretical formula from the memo

figure(98)
plot(2*Ftau(1:Klast),[eta(1:Klast);eff'])
legend('theoretical','simulation')
grid
title('Theoretical vs. Simulation Efficiency of BPF')
xlabel('2-sided bandwidth*tau')
ylabel('Efficiency')
shg

% repeat on dB scale
figure(99)
plot(2*Ftau(1:Klast),db([eta(1:Klast);eff'],'power'))
legend('theoretical','simulation')
grid
title('Theoretical vs. Simulation Efficiency of BPF')
xlabel('2-sided bandwidth*tau')
ylabel('Efficiency (dB)')
shg
```