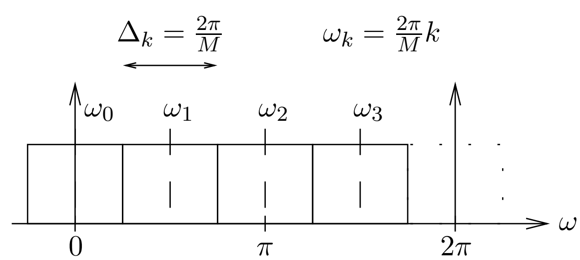
ECE-700 Filterbank Notes

Phil Schniter

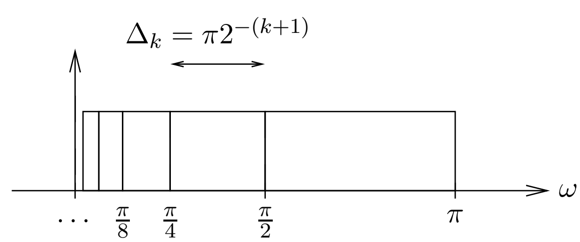
March 10, 2008

1 Why Filterbanks?

- *Sub-band Processing:* There exist many applications in modern signal processing where it is advantageous to separate a signal into different frequency ranges called sub-bands. The spectrum might be partitioned in the uniform manner illustrated below, where the sub-band width $\Delta_k = \frac{2\pi}{M}$ is identical for each sub-band and the band centers are uniformly spaced at intervals of $\frac{2\pi}{M}$:

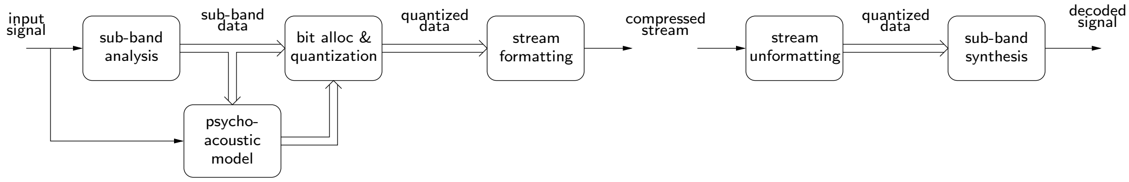
  Alternatively, the sub-bands might have a logarithmic spacing like that shown below:

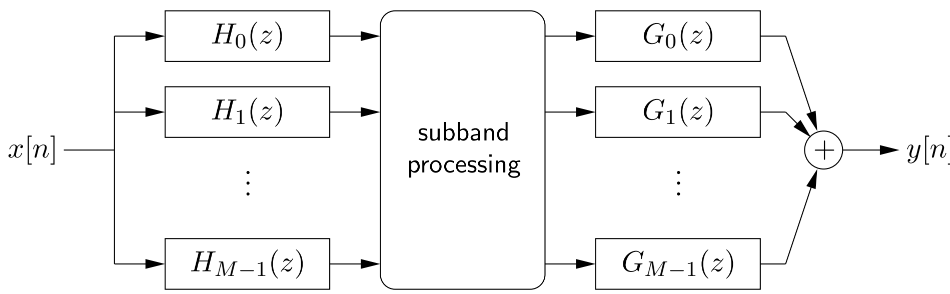
  For most of our discussion, we will focus on uniformly spaced sub-bands.

  The separation into sub-band components is intended to make further processing more convenient. Some of the most popular applications for sub-band decomposition are audio and video source coding (with the goal of efficient storage and/or transmission).

  The figure below illustrates the use of sub-band processing in MPEG audio coding. There a psychoacoustic model is used to decide how much quantization error can be tolerated in each sub-band while remaining below the hearing threshold of a human listener. In the sub-bands that can tolerate more error, less bits are used for coding. The quantized sub-band signals can then be decoded and recombined to reconstruct (an approximate version of) the input signal. Such processing allows, on average, a 12-to-1 reduction in bit rate while still maintaining "CD quality" audio. The psychoacoustic model takes into account the "spectral masking" phenomenon of the human ear, which says that high energy in one spectral region will limit the ear's ability to hear details in nearby spectral regions. Therefore, when the energy in one sub-band is high, nearby subbands can be coded with less bits without degrading the perceived quality of the audio signal. The MPEG standard specifies a 32-channels of sub-band filtering. Some psychoacoustic models also take into account "temporal masking" properties of the human ear, which say that a loud bursts of sound will temporarily overload the ear for short time durations, making it possible to hide quantization noise in the time interval after a loud sound burst.



In typical applications, non-trivial signal processing takes places between the bank of analysis filters and the bank of synthesis filters, as shown below. We will focus, however, on filterbank design rather than on the processing that occurs between the filterbanks.

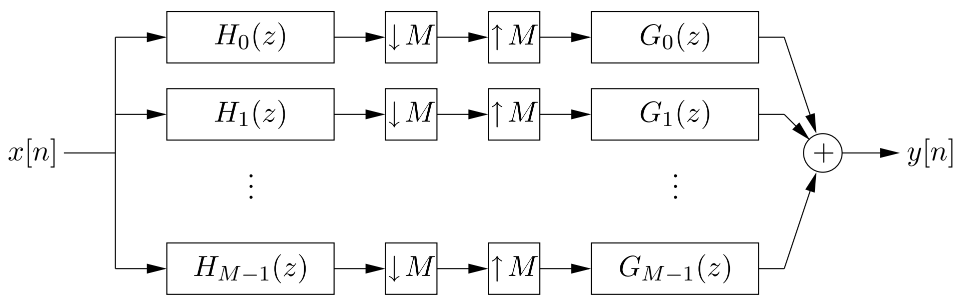


Our goals in filter design are

1. Good sub-band frequency separation (i.e., good "frequency selectivity").
2. Good reconstruction (i.e., $y[n] \approx x[n-d]$ for some integer delay d) when the sub-band processing is lossless.

The first goal is driven by the assumption that the sub-band processing works "best" when it is given access to cleanly separated sub-band signals, while the second goal is motivated by the idea that the sub-band filtering should not limit the reconstruction performance when the sub-band processing (e.g., the coding/decoding) is lossless or nearly lossless.

- *Uniform Filterbanks:* With M uniformly spaced sub-bands, the sub-band width is $\frac{2\pi}{M}$ radians, implying that the sub-band signal can be downsampled by factor M (but not more than M) without loss of information. This is referred to as a "critically sampled" filterbank. This maximal level of downsampling is advantageous when storing or further processing the sub-band signals. With critical sampling, the total number of downsampled sub-band output samples equals the total number of input samples. Assuming lossless sub-band processing, the critically-sampled synthesis/analysis procedure is illustrated below.



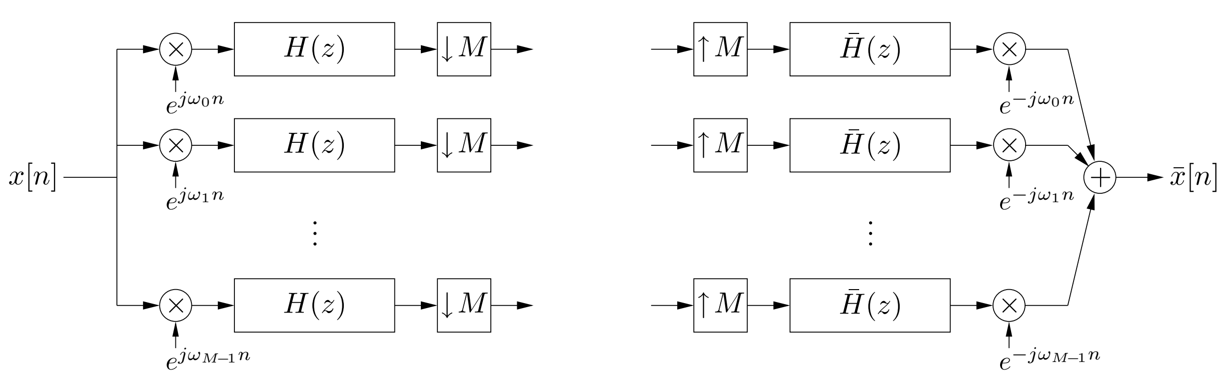
Recall that one of our goals in filter design is to ensure that $y[n] \approx x[n-d]$ for some integer delay d. From the block diagram above, one can see that imperfect analysis filtering will contribute aliasing errors to the sub-band signals. This aliasing distortion will degrade $y[n]$ if it is not canceled by the synthesis filterbank. Though ideal brick-wall filters $H_k(z)$ and $G_k(z)$ could easily provide perfect reconstruction (i.e., $y[n] = x[n-d]$), they would be un-implementable due to their doubly-infinite impulse responses. Thus, we are interested in the design of causal FIR filters that give near-perfect reconstruction or, if possible, perfect reconstruction.

There are two principle approaches to the design of filterbanks:

1. *Classical*: Approximate ideal brick wall filters to ensure good sub-band isolation (i.e., frequency selectivity) and accept (a hopefully small amount of) aliasing and thus reconstruction error.
2. *Modern*: Constrain the filters to give perfect (or near-perfect) reconstruction and hope for good sub-band isolation.

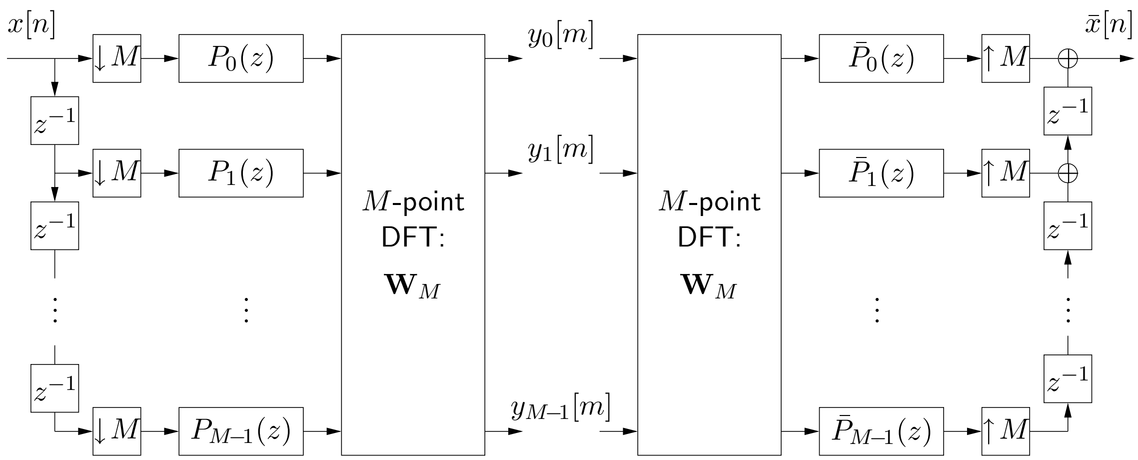
2 Classical Filterbanks

- <u>*Uniform Modulated Filterbank:*</u> A modulated filterbank is composed of analysis branches which (i) modulate the input to center the desired sub-band at DC, (ii) lowpass filter the modulated signal to isolate the desired sub-band, and (iii) downsample the lowpass signal. The synthesis branches interpolate the sub-band signals by upsampling and lowpass filtering, then modulate each sub-band back to its original spectral location.



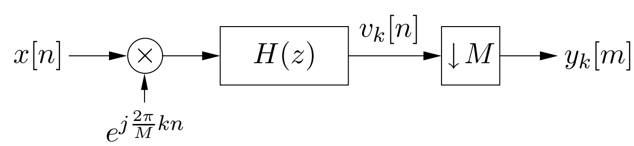
In an M-branch critically-sampled uniformly-modulated filterbank, the k^{th} analysis branch extracts the sub-band signal with center frequency $\omega_k = \frac{2\pi}{M}k$ via modulation and lowpass filtering with a (one-sided) bandwidth of $\frac{\pi}{M}$ radians, and then downsamples the result by factor M.

- *Polyphase/DFT Implementation of Uniform Modulated Filterbank:* The uniform modulated filterbank can be implemented using polyphase filterbanks and DFTs, resulting in huge computational savings. The figure below illustrates the equivalent polyphase/DFT structures for analysis and synthesis. The impulse responses of the polyphase filters $P_\ell(z)$ and $\bar{P}_\ell(z)$ can be defined in the time domain as $\bar{p}_\ell[m] = \bar{h}[mM+\ell]$ and $p_\ell[m] = h[mM+\ell]$, where $h[n]$ and $\bar{h}[n]$ denote the impulse responses of the analysis and synthesis lowpass filters, respectively.



Recall that the standard implementation performs modulation, filtering, and downsampling, in that order. The polyphase/DFT implementation reverses the order of these operations; it performs downsampling, then filtering, then modulation (if we interpret the DFT as a two-dimensional bank of "modulators"). We derive the polyphase/DFT implementation below, starting with the standard implementation and exchanging the order of modulation, filtering, and downsampling.

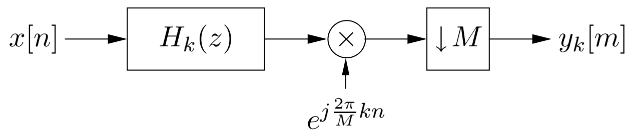
We start by analyzing the k^{th} filterbank branch, analyzed below.



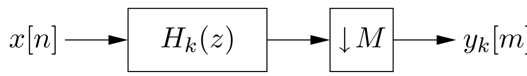
The first step is to reverse the modulation and filtering operations. To do this, we define a "modulated filter" $H_k(z)$:

$$
\begin{aligned}
v_k[n] &= \sum_i h[i]x[n-i]e^{j\frac{2\pi}{M}k(n-i)} \\
&= \Big(\sum_i \underbrace{h[i]e^{-j\frac{2\pi}{M}ki}}_{h_k[i]} x[n-i]\Big)e^{j\frac{2\pi}{M}kn} \\
&= \Big(\sum_i h_k[i]x[n-i]\Big)e^{j\frac{2\pi}{M}kn}
\end{aligned}
$$

The equation above indicates that $x[n]$ is convolved with the modulated filter and that the filter output is modulated. This is illustrated in the figure below.



Notice that the only modulator outputs not discarded by the downsampler are those with time index $n = mM$ for $m \in \mathbb{Z}$. For those outputs, the modulator has the value $e^{j\frac{2\pi}{M}kmM} = 1$, and thus it can be ignored. The resulting system is portrayed below.



Next we would like to reverse the order of filtering and downsampling. To apply the Noble identity, we must decompose $H_k(z)$ into a bank of upsampled polyphase filters. The technique used to derive polyphase decimation can be employed here:

$$H_k(z) \;=\; \sum_{n=-\infty}^{\infty} h_k[n] z^{-n} \;=\; \sum_{\ell=0}^{M-1} \sum_{m=-\infty}^{\infty} h_k[mM+\ell] z^{-mM-\ell}$$

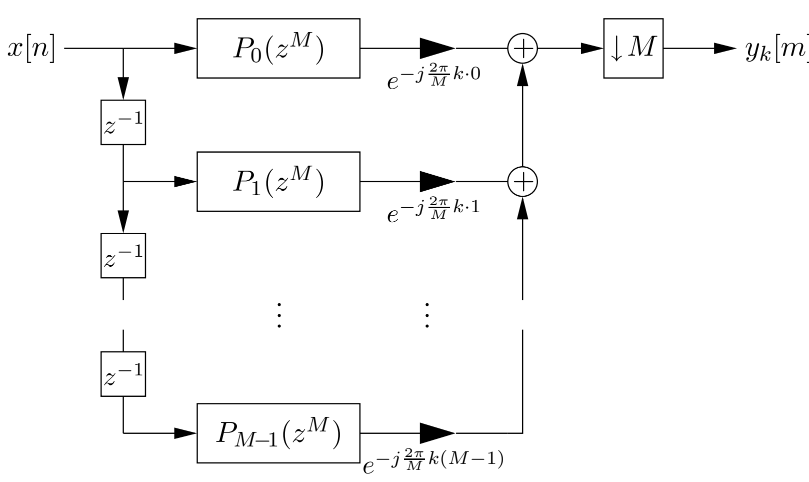
Noting that the ℓ^{th} polyphase filter has impulse response

$$h_k[mM+\ell] \;=\; h[mM+\ell] e^{-j\frac{2\pi}{M}k(mM+\ell)} \;=\; h[mM+\ell] e^{-j\frac{2\pi}{M}k\ell} \;=\; p_\ell[m] e^{-j\frac{2\pi}{M}k\ell},$$

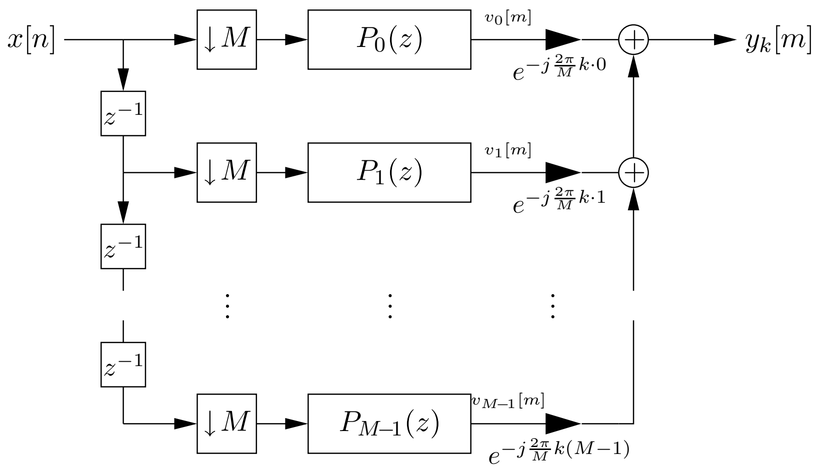
where $p_\ell[m]$ is the ℓ^{th} polyphase filter defined by the original (unmodulated) lowpass filter $H(z)$, we obtain

$$\begin{aligned} H_k(z) &= \sum_{\ell=0}^{M-1} \sum_{m=-\infty}^{\infty} p_\ell[m] e^{-j\frac{2\pi}{M}k\ell} z^{-mM-\ell} \\ &= \sum_{\ell=0}^{M-1} e^{-j\frac{2\pi}{M}k\ell} z^{-\ell} \sum_{m=-\infty}^{\infty} p_\ell[m] (z^M)^{-m} \\ &= \sum_{\ell=0}^{M-1} e^{-j\frac{2\pi}{M}k\ell} z^{-\ell} P_\ell(z^M). \end{aligned}$$

The k^{th} filterbank branch (now containing M polyphase branches) is illustrated below.



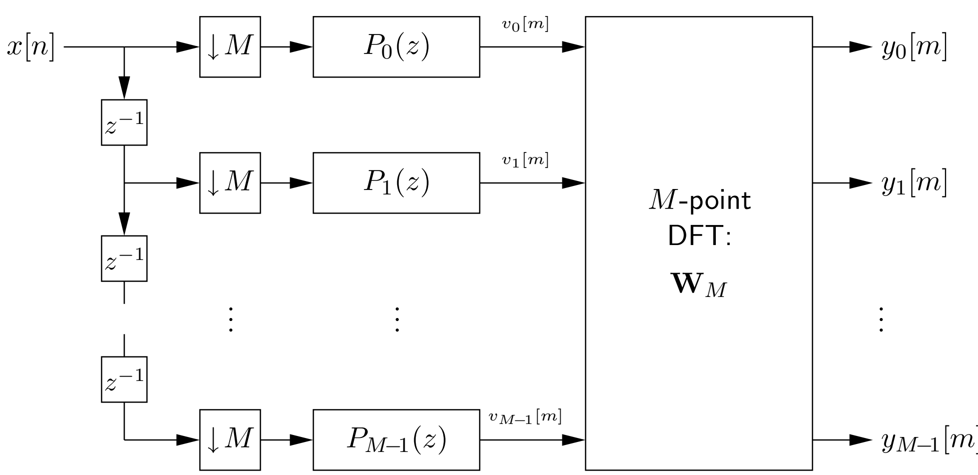
Because it is a linear operator, the downsampler can be moved through the adders and the (time-invariant) scalings $e^{-j\frac{2\pi}{M}k\ell}$. Finally, the Noble identity is employed to exchange the filtering and downsampling. The k^{th} filterbank branch becomes:



Observe that the polyphase outputs $\{v_\ell[m], \ell = 0...M-1\}$ are identical for each filterbank branch, while the scalings $\{e^{-j\frac{2\pi}{M}k\ell}, \ell = 0...M-1\}$ are different for each filterbank branch since they depend on the filterbank branch index k. Thus, we only need to calculate the polyphase outputs $\{v_\ell[m], \ell = 0...M-1\}$ once. Using these outputs we can compute the branch outputs via

$$y_k[m] = \sum_{\ell=0}^{M-1} v_\ell[m] e^{-j\frac{2\pi}{M}k\ell}$$

From the previous equation it is clear that $y_k[m]$ corresponds to the k^{th} DFT output given the M-point input sequence $\{v_\ell[m], \ell = 0...M-1\}$. Thus the M filterbank branches can be computed in parallel by taking an M-point DFT of the M polyphase outputs.



The polyphase/DFT synthesis bank can be derived in a similar manner.

- *Computational Savings of the Polyphase/DFT Modulated Filterbank Implementation:* Here we consider the analysis bank only; the synthesis bank can be treated similarly. Assume that the lowpass filter $H(z)$ has impulse response length N. To calculate the sub-band output vector $\{y_k[m], k = 0...M-1\}$ using the standard structure, we have

$$M \, \tfrac{\textsf{decimator outputs}}{\textsf{vector}} \times M \, \tfrac{\textsf{filter outputs}}{\textsf{decimator output}} \times (N+1) \, \tfrac{\textsf{multiplications}}{\textsf{filter output}} \;=\; M^2(N+1) \, \tfrac{\textsf{multiplications}}{\textsf{vector}},$$

where we have included one multiply for the modulator. The calculations above pertain to standard (i.e., not polyphase) decimation. If we implement the lowpass/downsampler in each filterbank branch with a polyphase decimator,

$$M \, \tfrac{\textsf{branch outputs}}{\textsf{vector}} \times (N+M) \, \tfrac{\textsf{multiplications}}{\textsf{branch output}} \;=\; M(N+M) \, \tfrac{\textsf{multiplications}}{\textsf{vector}}.$$

To calculate the same output vector for the polyphase/DFT structure, where we assume a radix-2 FFT algorithm is used to implement the DFT, we have approximately

$$1 \, \tfrac{\textsf{DFT}}{\textsf{vector}} \times \Big(\tfrac{M}{2} \log_2 M \, \tfrac{\textsf{multiplications}}{\textsf{DFT}} + M \, \tfrac{\textsf{polyphase outputs}}{\textsf{DFT}} \times \tfrac{N}{M} \, \tfrac{\textsf{multiplications}}{\textsf{polyphase output}} \Big) \;=\; \Big(N + \tfrac{M}{2} \log_2 M \Big) \, \tfrac{\textsf{multiplications}}{\textsf{vector}}.$$

The table below gives some typical numbers. Recall that the filter length N will be linearly proportional to the decimation factor M, so that the ratio $\frac{N}{M}$ determines the passband and stopband performance of the filtering.

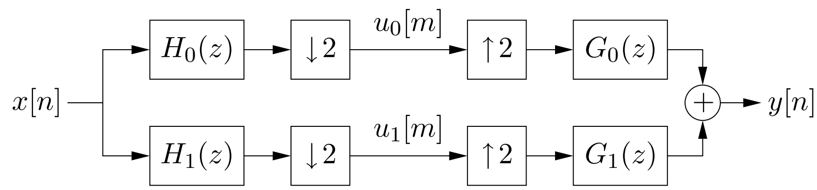
	$M = 32, \frac{N}{M} = 10$	$M = 128, \frac{N}{M} = 10$
standard	328704	20987904
standard with polyphase	11264	180224
polyphase/DFT	400	1728

- *Drawbacks of Classical Filterbank Designs:* The "classical" filterbanks that we have considered so far give perfect reconstruction performance only when the analysis and synthesis filters are ideal. With non-ideal (i.e., implementable) filters, aliasing will result from the downsampling/upsampling operation and corrupt the output signal. Since aliasing distortion is inherently non-linear, it may be very undesirable in certain applications. Thus, long analysis/synthesis filters might be required to force aliasing distortion down to tolerable levels. The cost of long filters is somewhat offset by the efficient polyphase implementation, though.

  That said, clever filter designs have been proposed which prevent aliasing in *neighboring* sub-bands [1, 2, 3]. As neighboring-subband aliasing typically constitutes the bulk of aliasing distortion, these designs give significant performance gains. In fact, such filter designs are used in the MPEG high-performance audio compression standards.

3 Modern Filterbanks

- *Aliasing-Cancellation Conditions:* It is possible to design combinations of analysis and synthesis filters such that the aliasing from downsampling/upsampling is completely canceled. Below we derive aliasing-cancellation conditions for two-channel filterbanks. Though the results can be extended to M-channel filterbanks in a rather straightforward manner, the two-channel case offers a more lucid explanation of the principle ideas.



The aliasing cancellation conditions follow directly from the input/output equations derived below. Let $i \in \{0,1\}$ denote the filterbank branch index. Then

$$\begin{aligned}
U_i(z) &= \frac{1}{2}\sum_{p=0}^{1} H_i(z^{\frac{1}{2}}e^{-j\pi p})X(z^{\frac{1}{2}}e^{-j\pi p}) \\
Y(z) &= \sum_{i=0}^{1} G_i(z)U_i(z^2) \\
&= \sum_{i=0}^{1} G_i(z)\frac{1}{2}\sum_{p=0}^{1} H_i(ze^{-j\pi p})X(ze^{-j\pi p}) \\
&= \frac{1}{2}\sum_{i=0}^{1} G_i(z)\big(H_i(z)X(z) + H_i(-z)X(-z)\big) \\
&= \frac{1}{2}\begin{bmatrix} X(z) & X(-z)\end{bmatrix} \underbrace{\begin{bmatrix} H_0(z) & H_1(z) \\ H_0(-z) & H_1(-z)\end{bmatrix}}_{\mathbf{H}(z)} \begin{bmatrix} G_0(z) \\ G_1(z)\end{bmatrix}
\end{aligned}$$

$\mathbf{H}(z)$ is often called the "aliasing component matrix." For aliasing cancellation, we need to ensure that $X(-z)$ does not contribute to the output $Y(z)$. This requires that

$$0 = \begin{bmatrix} H_0(-z) & H_1(-z)\end{bmatrix} \begin{bmatrix} G_0(z) \\ G_1(z)\end{bmatrix} = H_0(-z)G_0(z) + H_1(-z)G_1(z),$$

which is guaranteed by

$$\frac{G_0(z)}{G_1(z)} = -\frac{H_1(-z)}{H_0(-z)}.$$

The last statement can be expressed as

$$\left\{\begin{array}{rcl} G_0(z) &=& C(z)H_1(-z) \\ G_1(z) &=& -C(z)H_0(-z) \end{array}\right. \quad \text{for any rational } C(z).$$

Under these aliasing-cancellation conditions, we get the input/output relation

$$Y(z) = \underbrace{\frac{1}{2}\Big(H_0(z)H_1(-z) - H_1(z)H_0(-z)\Big)C(z)}_{T(z)}\, X(z)$$

where $T(z)$ represents the system transfer function. We say that "perfect reconstruction" results when $y[n] = x[n-\ell]$ for some $\ell \in \mathbb{N}$, or equivalently when $T(z) = z^{-\ell}$.

Note that the aliasing-cancellation conditions remove one degree of freedom from our filterbank design; originally we had the choice of four transfer functions $\{H_0(z), H_1(z), G_0(z), G_1(z)\}$, whereas now we choose three: $\{H_0(z), H_1(z), C(z)\}$.

- *Quadrature Mirror Filterbank:* The quadrature mirror filter (QMF) bank is an aliasing-cancellation filterbank with the additional design choices:

$$\begin{cases} H_0(z) & : & \text{causal real-coefficient FIR} \\ H_1(z) & = & H_0(-z) \\ C(z) & = & 2 \end{cases}$$

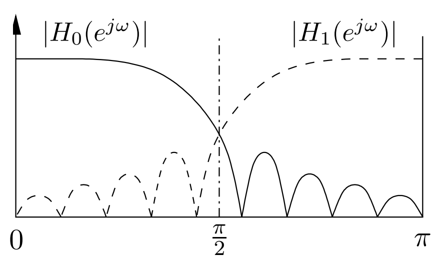
  Combining the various design rules, it is easy to see that all filters will be causal, real-coefficient, and FIR. The QMF choices yield the system transfer function

$$T(z) \;=\; H_0^2(z) - H_1^2(z) \;=\; H_0^2(z) - H_0^2(-z).$$

  The name "QMF" is appropriate for the following reason. Note that

$$|H_1(e^{j\omega})| \;=\; |H_0(-e^{j\omega})| \;=\; |H_0(e^{j(\omega-\pi)})| \;=\; |H_0(e^{j(\pi-\omega)})|$$

  where the last step follows from the DTFT conjugate-symmetry of real-coefficient filters. This implies that the magnitude responses $|H_0(e^{j\omega})|$ and $|H_1(e^{j\omega})|$ form a mirror-image pair, symmetric around $\omega = \frac{\pi}{2} = \frac{2\pi}{4}$ (the "quadrature frequency"), as illustrated below.



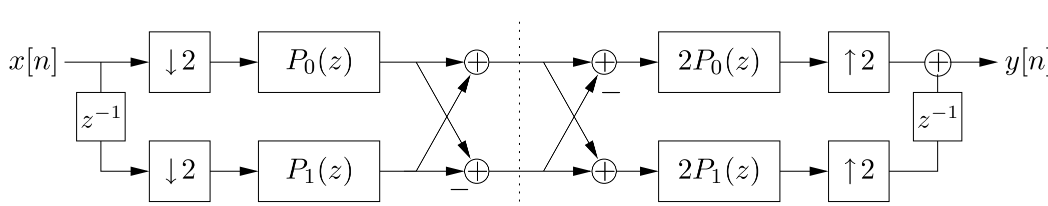
  The QMF design rules imply that all filters in the bank are directly related to the "prototype" filter $H_0(z)$, and thus we might suspect a polyphase implementation. In fact, one exists. Using the standard polyphase decomposition of $H_0(z)$, we have

$$H_0(z) \;=\; P_0(z^2) + z^{-1}P_1(z^2)$$

  so that

$$\begin{aligned} H_1(z) &= H_0(-z) \;=\; P_0(z^2) - z^{-1}P_1(z^2) \\ G_0(z) &= 2H_1(-z) \;=\; 2P_0(z^2) + 2z^{-1}P_1(z^2) \\ G_1(z) &= -2H_0(-z) \;=\; -2P_0(z^2) + 2z^{-1}P_1(z^2) \end{aligned}$$

  Application of the Noble identity results in the polyphase structure below.



The QMF choice $C(z) = 2$ implies that the synthesis filters have twice the DC gain of the corresponding analysis filters. Recalling that decimation by 2 involves anti-alias lowpass filtering with DC gain equal to one, while interpolation by 2 involves anti-image lowpass filtering with DC gain equal to 2, the polyphase diagram suggests an explanation for the choice $C(z) = 2$.

It is interesting to note that the polyphase output pair is processed using the matrix operation $\left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$ on the analysis side and $\left[\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right]$ on the synthesis side. The matrix $\left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$ is a 2×2 DFT matrix, providing an interesting connection to the classical polyphase/DFT structure studied earlier.

- *Perfect Reconstruction QMF:* The system transfer function for a QMF bank is

$$\begin{aligned} T(z) &= H_0^2(z) - H_1^2(z) \\ &= 4z^{-1}P_0(z^2)P_1(z^2) \end{aligned}$$

For perfect reconstruction, we need $T(z) = z^{-\ell}$ for some $\ell \in \mathbb{N}$, which implies the equivalent conditions

$$\begin{aligned} 4z^{-1}P_0(z^2)P_1(z^2) &= z^{-\ell} \\ P_0(z^2)P_1(z^2) &= \tfrac{1}{4}z^{-(\ell-1)} \\ P_0(z)P_1(z) &= \tfrac{1}{4}z^{-\frac{\ell-1}{2}} \end{aligned}$$

For FIR polyphase filters, this can only be satisfied by

$$\left\{ \begin{array}{rcl} P_0(z) &=& \beta_0 z^{-n_0} \\ P_1(z) &=& \beta_1 z^{-n_1} \end{array} \right. \quad \text{where} \quad \left\{ \begin{array}{rcl} n_0 + n_1 &=& \frac{\ell-1}{2} \\ \beta_0\beta_1 &=& \frac{1}{4} \end{array} \right.$$

In other words, the polyphase filters are trivial, so that the prototype filter $H_0(z)$ has a two-tap response. With only two taps, $H_0(z)$ cannot be a very good lowpass filter, meaning that the sub-band signals will not be spectrally well-separated. From this we conclude that two-channel[1] perfect reconstruction QMF banks exist but are not very useful.

- *Johnston's QMF Banks:* Two-channel perfect-reconstruction QMF banks are not very useful because the analysis filters have poor frequency selectivity. The selectivity characteristics can be improved, however, if we allow the system response $T(e^{j\omega})$ to have magnitude-response ripples while keeping it linear phase.

Say that $H_0(z)$ is causal, linear-phase, and has impulse response length N. Then it is possible to write $H_0(e^{j\omega})$ in terms of a real-valued zero-phase response $\tilde{H}_0(e^{j\omega})$, so that

$$\begin{aligned} H_0(e^{j\omega}) &= e^{-j\omega\frac{N-1}{2}}\tilde{H}_0(e^{j\omega}) \\ T(e^{j\omega}) &= H_0^2(e^{j\omega}) - H_0^2(e^{j(\omega-\pi)}) \\ &= e^{-j\omega(N-1)}\tilde{H}_0^2(e^{j\omega}) - e^{-j(\omega-\pi)(N-1)}\tilde{H}_0^2(e^{j(\omega-\pi)}) \\ &= e^{-j\omega(N-1)}\left(\tilde{H}_0^2(e^{j\omega}) - e^{j\pi(N-1)}\tilde{H}_0^2(e^{j(\omega-\pi)})\right) \end{aligned}$$

Note that if N is odd, $e^{j\pi(N-1)} = 1$, implying that

$$T(e^{j\omega})\big|_{\omega=\frac{\pi}{2}} = 0.$$

[1] It turns out that M-channel perfect reconstruction QMF banks have more useful responses for larger values of M.

A null in the system response would be very undesirable, and so we restrict N to be an even number. In that case,

$$\begin{aligned} T(e^{j\omega}) &= e^{-j\omega(N-1)}\left(\tilde{H}_0^2(e^{j\omega}) + \tilde{H}_0^2(e^{j(\omega-\pi)})\right) \\ &= e^{-j\omega(N-1)}\left(|H_0(e^{j\omega})|^2 + |H_0(e^{j(\omega-\pi)})|^2\right) \end{aligned}$$

Note that the system response is linear phase, but will have amplitude distortion if $|H_0(e^{j\omega})|^2 + |H_0(e^{j(\omega-\pi)})|^2$ is not equal to a constant.

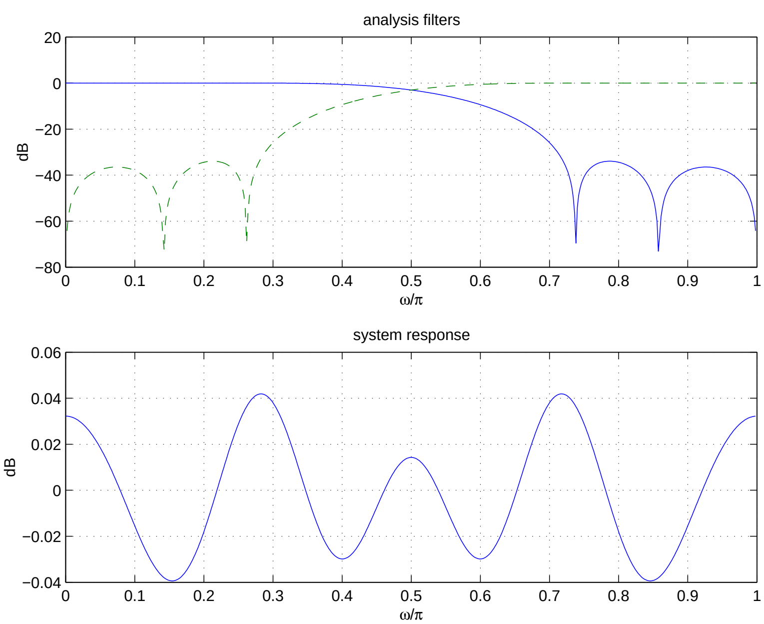
Johnston's idea was to minimize a cost function which penalizes deviation from perfect reconstruction as well as deviation from an ideal lowpass filter with cutoff ω_0. Specifically, real symmetric coefficients $h_0[n]$ are chosen to minimize

$$J \;=\; \lambda \int_{\omega_0}^{\pi} |H_0(e^{j\omega})|^2 d\omega + (1-\lambda) \int_0^{\pi} \left|1 - |H_0(e^{j\omega})|^2 - |H_0(e^{j(\omega-\pi)})|^2\right| d\omega$$

where $0 < \lambda < 1$ balances between the two conflicting objectives. Numerical optimization techniques can be used to determine the coefficients, and a number of popular coefficient sets have been tabulated [2, 4, 5]. As an example, consider the "12B" filter from [4]:

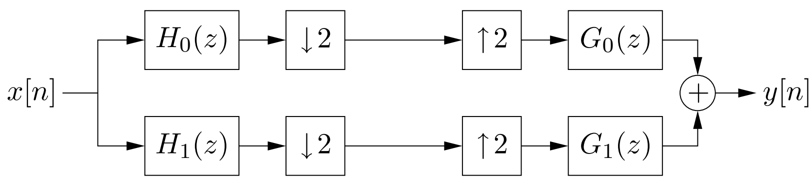
$$\begin{aligned} h_0[0] &= \text{-0.006443977} &= h_0[11] \\ h_0[1] &= 0.02745539 &= h_0[10] \\ h_0[2] &= \text{-0.00758164} &= h_0[9] \\ h_0[3] &= \text{-0.0913825} &= h_0[8] \\ h_0[4] &= 0.09808522 &= h_0[7] \\ h_0[5] &= 0.4807962 &= h_0[6] \end{aligned}$$

which gives the following DTFT magnitudes.



- *FIR Perfect-Reconstruction Conditions:* The QMF design choices prevented the design of a useful (i.e., frequency selective) perfect-reconstruction (PR) FIR filterbank. This motivates us to re-examine PR filterbank design without the overly-restrictive QMF conditions. However, we will still require causal FIR filters with real coefficients.

Recalling that the two-channel filterbank



has the input/output relation

$$Y(z) \;=\; \frac{1}{2}\begin{bmatrix} X(z) & X(-z) \end{bmatrix}\begin{bmatrix} H_0(z) & H_1(z) \\ H_0(-z) & H_1(-z) \end{bmatrix}\begin{bmatrix} G_0(z) \\ G_1(z) \end{bmatrix},$$

we see that delay-ℓ perfect reconstruction requires

$$\begin{bmatrix} 2z^{-\ell} \\ 0 \end{bmatrix} \;=\; \underbrace{\begin{bmatrix} H_0(z) & H_1(z) \\ H_0(-z) & H_1(-z) \end{bmatrix}}_{\mathbf{H}(z)}\begin{bmatrix} G_0(z) \\ G_1(z) \end{bmatrix},$$

or, equivalently, that

$$\begin{aligned} \begin{bmatrix} G_0(z) \\ G_1(z) \end{bmatrix} &= \mathbf{H}^{-1}(z)\begin{bmatrix} 2z^{-\ell} \\ 0 \end{bmatrix} \\ &= \frac{1}{\det(\mathbf{H}(z))}\begin{bmatrix} H_1(-z) & -H_1(z) \\ -H_0(-z) & H_0(z) \end{bmatrix}\begin{bmatrix} 2z^{-\ell} \\ 0 \end{bmatrix} \\ &= \frac{2}{\det(\mathbf{H}(z))}\begin{bmatrix} z^{-\ell}H_1(-z) \\ -z^{-\ell}H_0(-z) \end{bmatrix} \end{aligned}$$

where

$$\det(\mathbf{H}(z)) \;=\; H_0(z)H_1(-z) - H_0(-z)H_1(z).$$

For FIR $G_0(z)$ and $G_1(z)$, we require [2] that

$$\det(\mathbf{H}(z)) \;=\; cz^{-k} \quad \text{for } c \in \mathbb{R} \text{ and } k \in \mathbb{Z}^+.$$

Under this determinant condition, we find that

$$\begin{bmatrix} G_0(z) \\ G_1(z) \end{bmatrix} \;=\; \frac{2z^{-(\ell-k)}}{c}\begin{bmatrix} H_1(-z) \\ -H_0(-z) \end{bmatrix}$$

Assuming that $H_0(z)$ and $H_1(z)$ are causal with non-zero initial coefficient, we choose $k = \ell$ to keep $G_0(z)$ and $G_1(z)$ causal and free of unnecessary delay. Summarizing the two-channel FIR-PR conditions:

$$\left\{ \begin{array}{rcl} H_0(z) \;\&\; H_1(z) & : & \text{causal real-coefficient FIR} \\ \det(\mathbf{H}(z)) & = & cz^{-\ell} \;\; \text{for } c \in \mathbb{R} \text{ and } \ell \in \mathbb{Z}^+ \\ G_0(z) & = & \frac{2}{c}H_1(-z) \\ G_1(z) & = & -\frac{2}{c}H_0(-z) \end{array} \right.$$

[2] Since we cannot assume that FIR $H_0(z)$ and $H_1(z)$ share a common root.

- *Orthogonal Perfect Reconstruction FIR Filterbanks:* The FIR perfect-reconstruction (PR) conditions leave some freedom in the choice of $H_0(z)$ and $H_1(z)$. "Orthogonal PR" filterbanks are defined by causal real-coefficient even-length-N analysis filters which satisfy

$$\begin{cases} 1 &= H_0(z)H_0(z^{-1}) + H_0(-z)H_0(-z^{-1}) \\ H_1(z) &= \pm z^{-(N-1)} H_0(-z^{-1}) \end{cases}$$

To verify that these design choices satisfy the FIR-PR requirements for $H_0(z)$ and $H_1(z)$, we evaluate $\det(\mathbf{H}(z))$ under the second condition above. This yields

$$\begin{aligned} \det(\mathbf{H}(z)) &= H_0(z)H_1(-z) - H_0(-z)H_1(z) \\ &= \mp z^{-(N-1)}\big(H_0(z)H_0(z^{-1}) + H_0(-z)H_0(-z^{-1})\big) \\ &= \mp z^{-(N-1)}. \end{aligned}$$

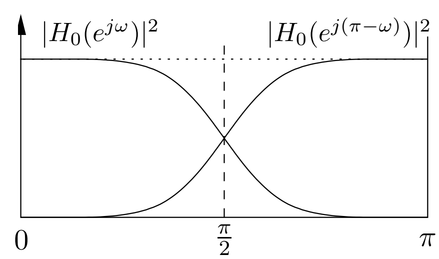
which corresponds to $c = \mp 1$ and $\ell = N-1$ in the FIR-PR determinant condition $\det(\mathbf{H}(z)) = cz^{-\ell}$. The remaining FIR-PR conditions then imply that the synthesis filters are given by

$$\begin{cases} G_0(z) &= \mp 2H_1(-z) &= 2z^{-(N-1)}H_0(z^{-1}) \\ G_1(z) &= \pm 2H_0(-z) &= 2z^{-(N-1)}H_1(z^{-1}) \end{cases}$$

The orthogonal PR design rules imply that $H_0(e^{j\omega})$ is "power symmetric" and that $\{H_0(e^{j\omega}), H_1(e^{j\omega})\}$ form a "power complementary" pair. To see the power symmetry, we rewrite the first design rule using $z = e^{j\omega}$ and $-1 = e^{\pm j\pi}$, which gives

$$\begin{aligned} 1 &= H_0(e^{j\omega})H_0(e^{-j\omega}) + H_0(e^{j(\omega-\pi)})H_0(e^{-j(\omega-\pi)}) \\ &= |H_0(e^{j\omega})|^2 + |H_0(e^{j(\omega-\pi)})|^2 \\ &= |H_0(e^{j\omega})|^2 + |H_0(e^{j(\pi-\omega)})|^2. \end{aligned}$$

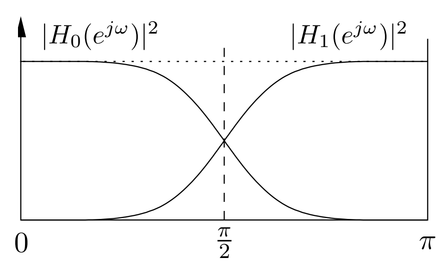
The last two steps leveraged the fact that the DTFT of a real-coefficient filter is conjugate-symmetric. The power-symmetry property is illustrated below.



Power complementarity follows from the second orthogonal PR design rule, which implies $|H_1(e^{j\omega})| = |H_0(e^{j(\pi-\omega)})|$. Plugging this into the previous equation, we find

$$1 = |H_0(e^{j\omega})|^2 + |H_1(e^{j\omega})|^2.$$

The power-complementary property is illustrated below.



- *Design of Orthogonal PR-FIR Filterbanks via Halfband Spectral Factorization:* Recall that analysis-filter design for orthogonal PR-FIR filterbanks reduces to the design of a real-coefficient causal FIR prototype filter $H_0(z)$ which satisfies the power-symmetry condition

$$|H_0(e^{j\omega})|^2 + |H_0(e^{j(\pi-\omega)})|^2 \; = \; 1.$$

How can we design such an $H_0(z)$? Below we give a procedure which uses halfband filters and spectral factorization.

Power-symmetric filters are closely related to real-valued zero-phase "halfband" filters. A zero-phase halfband filter $F(z)$ has the property

$$F(z) + F(-z) \; = \; 1.$$

Note that the zero-phase halfband property forces the 0^{th} order coefficient to $\frac{1}{2}$ and all other even coefficients to zero:

$$\begin{array}{rcl} F(z) & = & \cdots \; f[-2]z^2 \; + f[-1]z^1 + \; f[0] \; + f[1]z^{-1} + \; f[2]z^{-2} \; \cdots \\ F(-z) & = & \cdots \; f[-2]z^2 \; - f[-1]z^1 + \; f[0] \; - f[1]z^{-1} + \; f[2]z^{-2} \; \cdots \\ F(z)+F(-z) & = & \cdots \; 2f[-2]z^2 + \quad 0 \quad + 2f[0] + \quad 0 \quad + 2f[2]z^{-2} \cdots \end{array}$$

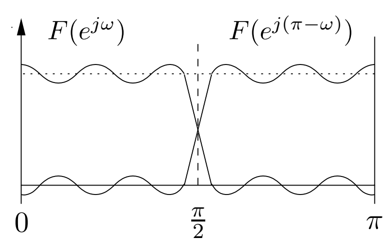
though it does not constrain the odd coefficients. Recalling that zero-phase filters have coefficients which are conjugate symmetric about the origin, a real-valued zero-phase halfband will have the properties:

$$\begin{cases} f[n] = \frac{1}{2} & \text{if } n = 0, \\ f[n] = 0 & \text{if } n \text{ even and } n \neq 0, \\ f[n] = f[-n] \in \mathbb{R} & \text{if } n \text{ odd.} \end{cases}$$

Recalling that real-valued coefficients imply a conjugate-symmetric DTFT, zero-phase halfband filters will have an "amplitude-symmetric" DTFT:

$$F(e^{j\omega}) + F(e^{j(\pi-\omega)}) \; = \; 1.$$

An example of amplitude-symmetry is illustrated below.



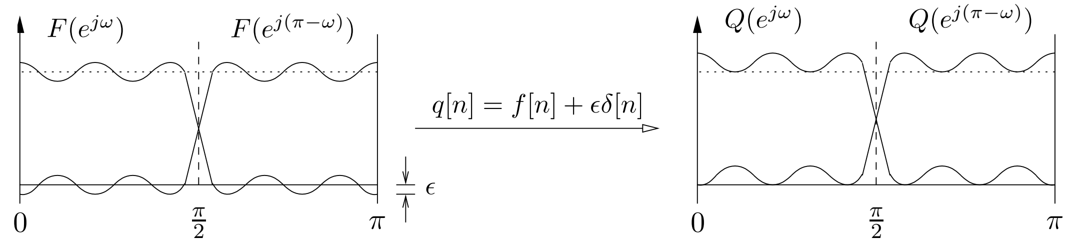
If, in addition to being real-valued,[3] $F(e^{j\omega})$ was non-negative, then $F(e^{j\omega})$ would make a valid power response (like $|H_0(e^{j\omega})|^2$). Furthermore, if we could find $H_0(z)$ such that

$$|H_0(e^{j\omega})|^2 = F(e^{j\omega}),$$

then this $H_0(z)$ would satisfy the desired power-symmetry property $1 = |H_0(e^{j\omega})|^2 + |H_0(e^{j(\pi-\omega)})|^2$. This is the approach that we take below.

[3] Recall that zero-phase filters have real-valued DTFTs.

First, realize $F(e^{j\omega})$ is easily modified to ensure non-negativity: construct $q[n] = f[n] + \epsilon\delta[n]$ for sufficiently large ϵ, which will raise $F(e^{j\omega})$ by ϵ uniformly over ω.

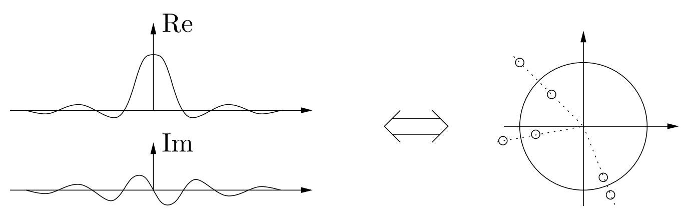


The resulting $Q(z)$ is non-negative and satisfies the amplitude-symmetry condition $Q(e^{j\omega})+Q(e^{j(\pi-\omega)}) = 1+2\epsilon$. We will make up for the additional "2ϵ" gain later. The procedure by which $H_0(z)$ can be calculated from the raised halfband $Q(z)$, known as "spectral factorization," is described next.

The fact that $f[n]$ is conjugate symmetric implies that $q[n]$ will also be conjugate-symmetric. Conjugate-symmetric coefficients imply that the roots of $Q(z)$ come in pairs $\{a_i, \frac{1}{a_i^*}\}$. This can be seen by writing $Q(z)$ in the factored form below, which clearly corresponds to a polynomial with coefficients conjugate-symmetric around the 0^{th}-order coefficient.

$$Q(z) \;=\; \sum_{n=-(N-1)}^{N-1} q[n]z^{-n} \;=\; A\prod_{i=1}^{N-1}(1-a_i z^{-1})(1-a_i^* z), \quad A \in \mathbb{R}^+$$

Note that the pair of complex numbers $\{a_i, \frac{1}{a_i^*}\}$ is symmetric across the unit circle in the z-plane. Thus, for every root of $Q(z)$ inside the unit circle, there exists a root outside of the unit circle.

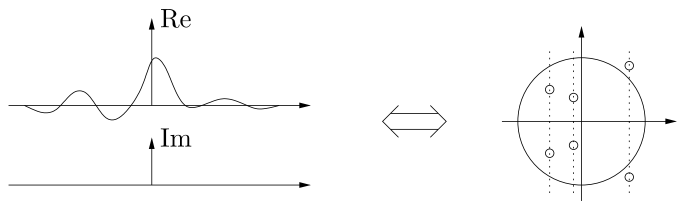


Let us assume, without loss in generality, that each $|a_i| < 1$. If we form $H_0(z)$ from the roots of $Q(z)$ with magnitude less than one, as in

$$H_0(z) \;=\; \sqrt{\frac{A}{1+2\epsilon}}\prod_{i=1}^{N-1}(1-a_i z^{-1}),$$

then it is apparent that $|H_0(e^{j\omega})|^2 = \frac{1}{1+2\epsilon}Q(e^{j\omega})$, so that $|H_0(e^{j\omega})|^2 + |H_0(e^{j(\pi-\omega)})|^2 = 1$, which was our goal. This $H_0(z)$ is the so-called "minimum-phase spectral factor" of $Q(z)$. Note: To handle the case where some $|a_i| = 1$, we can shown that the unit-magnitude roots actually come in pairs, and we would allocate one member of each pair to $H_0(z)$.

To make $|H_0(e^{j\omega})|^2 = Q(e^{j\omega})$, we are not actually required to choose all roots inside the unit circle; it is enough to choose one root from every unit-circle-symmetric pair. However, we do want to ensure that $H_0(z)$ has real-valued coefficients. For this, we must ensure that roots used to construct $H_0(z)$ come in conjugate-symmetric pairs, i.e., pairs having symmetry with respect to the real axis in the complex plane.



Because $Q(z)$ has real-valued coefficients, we know that its roots come in conjugate-symmetric pairs. Then, forming $H_0(z)$ from the roots of $Q(z)$ that are strictly inside (or strictly outside) the unit circle ensures that the roots of $H_0(z)$ will come in conjugate-symmetric pairs.

Finally, we say a few words about the design of the halfband filter $F(z)$. The "window design" method is one technique that could be used in this application. The window design method starts with an ideal lowpass filter and windows its doubly-infinite impulse response using a window function with finite time-support. The ideal real-valued zero-phase halfband filter has impulse response

$$\bar{f}[n] = \frac{\sin(\pi n/2)}{\pi n}, \quad n \in \mathbb{Z},$$

which obeys the zero-phase halfband properties listed earlier. Note that windowing the ideal halfband (with a real-valued origin-symmetric window) does not disturb these properties. It turns out that many of the other popular design methods (e.g., LS and equiripple) also produce halfband filters when their cutoff is specified to be $\pi/2$ radians and all passband/stopband specifications are symmetric with respect to $\omega = \pi/2$. These latter designs might be advantageous over the window design since they are "optimal" in certain respects.

We now summarize the design procedure for a length-N analysis lowpass filter that generates an orthogonal perfect-reconstruction FIR filterbank:

1. Design a zero-phase real-coefficient halfband lowpass filter $F(z) = \sum_{n=-(N-1)}^{N-1} f[n]z^{-n}$ where N is a positive even integer (via, e.g., window, LS, or equiripple designs).
2. Calculate ϵ, the maximum value of $-F(e^{j\omega})$. (Recall that $F(e^{j\omega})$ is real-valued for all ω because it has a zero-phase response.) Then create "raised halfband" $Q(z)$ via $q[n] = f[n] + \epsilon\delta[n]$, ensuring that $Q(e^{j\omega}) \geq 0 \;\; \forall\omega$.
3. Compute the roots of $Q(z)$, and verify that they come in unit-circle-symmetric pairs $\{a_i, \frac{1}{a_i^*}\}$. Then collect the roots with magnitude less than one into the filter $\hat{H}_0(z)$.
4. $\hat{H}_0(z)$ is the desired prototype filter except for a scale factor. Recall that we desire $|H_0(e^{j\omega})|^2 + |H_0(e^{j(\pi-\omega)})|^2 = 1$. Using Parseval's Theorem, we see that $\{\hat{h}_0[n]\}$ should be scaled to give $\{h_0[n]\}$ for which $\sum_{n=0}^{N-1} h_0^2[n] = 1/2$.

- *Bi-orthogonal Perfect Reconstruction FIR Filterbanks:* Due to the minimum-phase spectral factorization, orthogonal PR-FIR filterbanks will not have linear-phase analysis and synthesis filters. Non-linear phase may be undesirable in certain applications. "Bi-orthogonal" designs are closely related to orthogonal designs, yet give linear-phase filters. The analysis-filter design rules for the bi-orthogonal case are

$$\begin{cases} F(z) & : \text{ zero-phase real-coefficient halfband such that} \\ & \quad\; F(z) = \sum_{n=-(N-1)}^{N-1} f[n]z^{-n}, \quad N \text{ even} \\ z^{-(N-1)}F(z) & = H_0(z)H_1(-z) \end{cases}$$

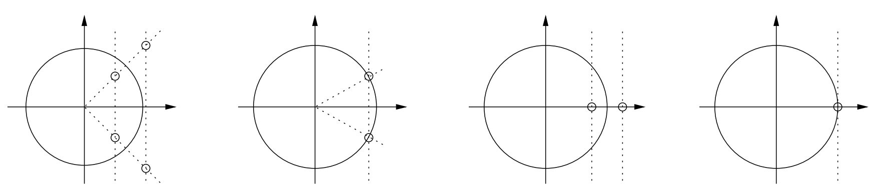
It is straightforward to verify that these design choices satisfy the FIR perfect reconstruction condition $\det(\mathbf{H}(z)) = cz^{-\ell}$ with $c = 1$ and $\ell = N-1$:

$$\begin{aligned}\det(\mathbf{H}(z)) &= H_0(z)H_1(-z) - H_0(-z)H_1(z)\\ &= z^{-(N-1)}F(z) \underbrace{-(-1)^{-(N-1)}}_{+ \text{ since } N \text{ even}} z^{-(N-1)}F(-z)\\ &= z^{-(N-1)} \underbrace{\big(F(z) + F(-z)\big)}_{=1 \text{ since zero-phase halfband}}\\ &= z^{-(N-1)}.\end{aligned}$$

Furthermore, note that $z^{-(N-1)}F(z)$ is causal with real coefficients, so that both $H_0(z)$ and $H_1(z)$ can be made causal with real coefficients. (This was another PR-FIR requirement.) The choice $c = 1$ implies that the synthesis filters should obey

$$\begin{cases} G_0(z) = 2H_1(-z) \\ G_1(z) = -2H_0(-z) \end{cases}.$$

Essentially, in bi-orthogonal analysis filter design, we factor the causal halfband filter $z^{-(N-1)}F(z)$ into $H_0(z)$ and $H_1(-z)$ that have both real coefficients *and* linear-phase. Earlier we saw that real-valued coefficients imply complex-conjugate root symmetry and a linear-phase response implies unit-circle root symmetry.[4] Simultaneous satisfaction of these two properties can be accomplished by *quadruples* of roots. However, there are special cases in which a root pair, or even a single root, can simultaneously satisfy these two types of symmetry. Examples are illustrated below.



With this factorization in mind, the design procedure for the analysis filters of a bi-orthogonal perfect-reconstruction FIR filterbank is summarized below:

1. Design a zero-phase real-coefficient halfband filter $F(z) = \sum_{n=-(N-1)}^{N-1} f[n]z^{-n}$ where N is a positive even integer (via, window, LS, or equiripple design).
2. Compute the roots of $F(z)$ and partition into a set of root groups $\{\mathcal{G}_0, \mathcal{G}_1, \mathcal{G}_2, ...\}$ that have *both* complex-conjugate and unit-circle symmetries. Thus a root group may have one of the following forms:

$$\begin{array}{ll} \mathcal{G}_i = \{a_i, a_i^*, \frac{1}{a_i}, \frac{1}{a_i^*}\} & \\ \mathcal{G}_i = \{a_i, a_i^*\} & \text{if } |a_i| = 1, \\ \mathcal{G}_i = \{a_i, \frac{1}{a_i}\} & \text{if } a_i \in \mathbb{R}, \\ \mathcal{G}_i = \{a_i\} & \text{if } a_i = \pm 1. \end{array}$$

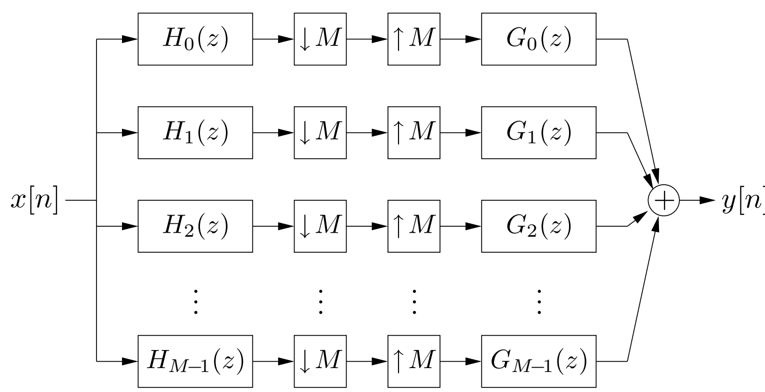
[4]Actually, we saw that zero-phase implies unit-circle root symmetry. But a zero-phase filter is transformed into a linear-phase filter by simply multiplying the z-transform by $z^{-(N-1)}$, which corresponds to adding $N-1$ roots at infinity.

Choose a few of the root groups and construct $\hat{H}_0(z)$ from the roots they contain.[5] Then construct $\hat{H}_1(-z)$ from the roots in the remaining groups. Finally, construct $\hat{H}_1(z)$ by changing the signs on the odd-indexed coefficients of $\hat{H}_1(-z)$. You may have to experiment with different group selections in order to get filters with good frequency selectivity.

3. $\hat{H}_0(z)$ and $\hat{H}_1(z)$ are the desired analysis filters up to a scaling. To take care of the scaling, first create $\check{H}_0(z) = a\hat{H}_0(z)$ and $\check{H}_1(z) = b\hat{H}_1(z)$ where a and b are selected so that $\sum_n |\check{h}_0[n]|^2 = 1 = \sum_n |\check{h}_1[n]|^2$. Then create $H_0(z) = c\check{H}_0(z)$ and $H_1(z) = c\check{H}_1(z)$ where c is selected so that the property $z^{-(N-1)}F(z) = H_0(z)H_1(-z)$ is satisfied at DC (i.e., $z = e^{j0} = 1$). In other words, find c so that $\sum_n h_0[n] \sum_m h_1[m](-1)^m = 1$.

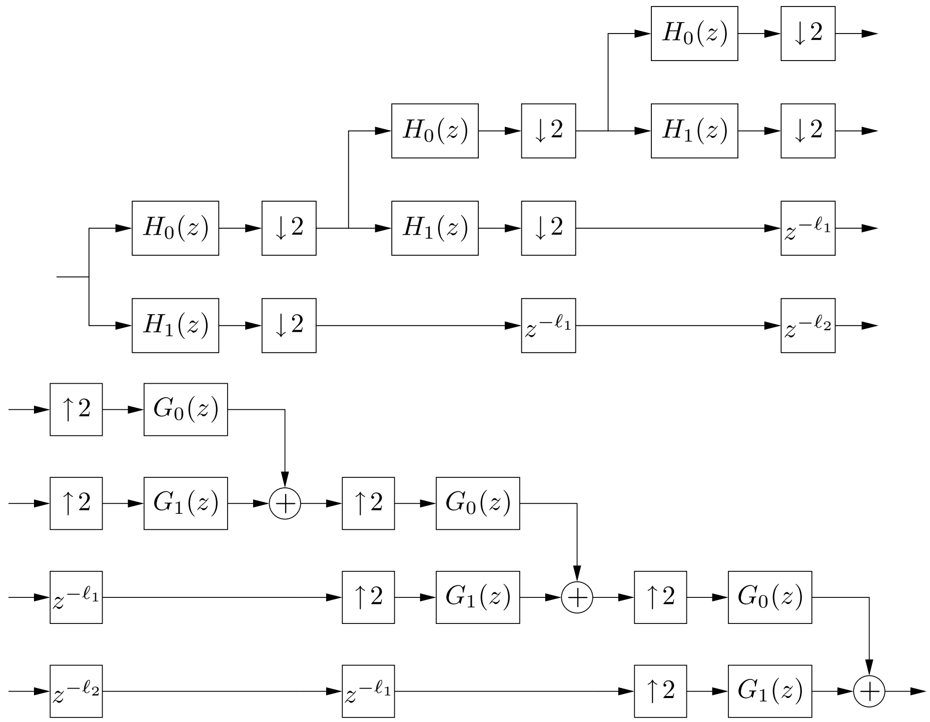
- <u>*Filterbanks with > 2 Branches:*</u> Thus far we have concentrated on "modern" filterbanks with only two branches. There are two standard ways by which the number of branches can be increased.

1. *M-band Filterbanks:*
The ideas used to construct two-branch PR-FIR filterbanks can be directly extended to the M-branch case [3, 6]. This yields, for example, a polynomial matrix $\mathbf{H}(z)$ with M rows and M columns. For these M-band filterbanks, the sub-bands will have uniform widths of $\frac{2\pi}{L}$ radians (in the ideal case).



2. *Multi-level (Cascaded) Filterbanks:*
The two-branch PR-FIR filterbanks can be cascaded to yield PR-FIR filterbanks whose sub-band widths equal $2^{-k}\pi$ radians for non-negative integers k (in the ideal case). If the magnitude responses of the filters are not well behaved, however, the cascading will result in poor effective frequency-selectivity. Below we show a filterbank in which the low-frequency sub-bands are narrower than the high-frequency subbands. Note that the number of input samples equals the total number of sub-band samples.

[5] Note that $\hat{H}_0(z)$ and $\hat{H}_1(z)$ will be real-coefficient linear-phase regardless of which groups are allocated to which filter. Their frequency selectivity, however, will be strongly influenced by group allocation. Thus, you may need to experiment with different allocations to find the best highpass/lowpass combination. Note also that the length of $H_0(z)$ may differ from the length of $H_1(z)$.



We shall see similar structures when we study the discrete wavelet transform.

References

[1] J. H. Rothweiler, "Polyphase quadrature filters—A new subband coding technique," in *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing*, (Boston, MA), pp. 1280–1283, 1983.

[2] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing.* Englewood Cliffs, NJ: Prentice Hall, 1983.

[3] P. P. Vaidyanathan, *Multirate Systems and Filter Banks.* Englewood Cliffs, NJ: Prentice Hall, 1993.

[4] J. D. Johnston, "A filter family designed for use in quadrature mirror filter banks," in *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing*, pp. 291–294, April 1980.

[5] R. Ansari and B. Liu, "Multirate signal processing," in *Handbook for Digital Signal Processing* (S. K. Mitra and J. F. Kaiser, eds.), ch. 14, pp. 981–1084, New York: Wiley-Interscience, 1993.

[6] S. K. Mitra, *Digital Signal Processing.* New York: McGraw-Hill, 2nd ed., 2001.