MUS421/EE367B Lecture 9
Multirate, Polyphase, and Wavelet Filter Banks

Julius O. Smith III (jos@ccrma.stanford.edu) , Scott Levine and Harvey Thornburg
Center for Computer Research in Music and Acoustics (CCRMA)
Department of Music, Stanford University
Stanford, California 94305

June 2, 2020

Outline

- Upsampling and Downsampling
- Polyphase Filtering
- Critically Sampled Perfect Reconstruction Filter Banks
- Perfect Reconstruction Filter Banks
- Paraunitary (Orthogonal) Filter Banks
- Multirate Analysis of Filter Banks
- DFT Filter Banks
- Cosine Modulated Filter Banks
- Wavelet Filter Banks

Upsampling and Downsampling

For the DFT, we have the Stretch Theorem (Repeat Theorem) which relates upsampling ("stretch") to spectral copies ("images") in the DFT context (length N signals and spectra).

We also have the Downsampling Theorem (Aliasing Theorem) for DFTs which relates downsampling to aliasing for finite length signals and spectra.

We now look at these relationships in the DTFT case. Thus, the signal length N is extended to infinity, and the spectrum becomes defined continuously over the unit circle in the z plane.

Upsampling (Stretch)

- Diagram:

$$x \longrightarrow \boxed{\uparrow N} \longrightarrow y$$

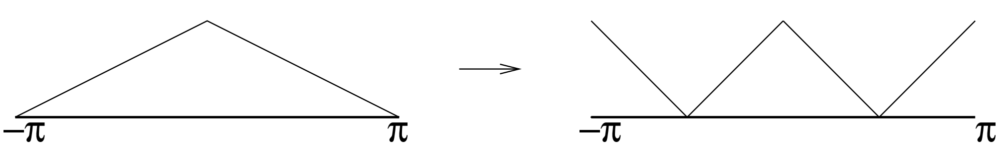
- Basic Idea: To upsample by the integer factor N, insert $N-1$ zeros between $x[n]$ and $x[n+1]$ for all n.
- Time Domain: $y = \textsc{Stretch}_N(x)$, i.e.,

$$y[n] = \begin{cases} x[n/N], & N \text{ divides } n \\ 0, & \text{otherwise.} \end{cases}$$

- Frequency Domain: $Y = \textsc{Repeat}_N(X)$, i.e.,

$$Y(z) = X(z^N),\ z \in \mathbb{C}$$

- Plugging in $z = e^{j\omega}$, we see that the spectrum on $[-\pi, \pi)$ contracts by the factor N, and N *images* appear around the unit circle. For $N = 2$, this is depicted below:



Downsampling (Decimation)

- Diagram:

$$x \longrightarrow \boxed{\downarrow N} \longrightarrow y$$

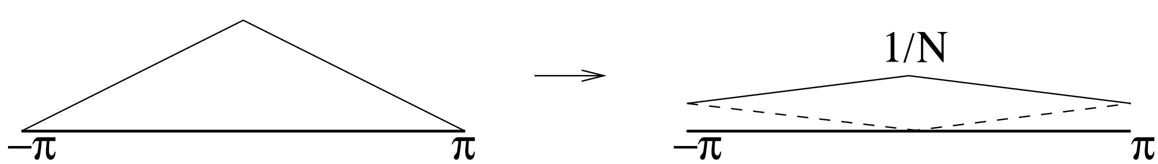
- Basic Idea: Take every N^{th} sample.
- Time Domain: $y = \text{DOWNSAMPLE}_N(x)$, i.e.,

$$y[n] = x[Nn],\ n \in \mathbb{Z}$$

- Frequency Domain: $Y = \text{ALIAS}_N(X)$, i.e.,

$$Y(z) = \frac{1}{N}\sum_{m=0}^{N-1} X\left(z^{\frac{1}{N}} e^{-jm\frac{2\pi}{N}}\right),\ z \in \mathbb{C}$$

Thus, the frequency axis is *expanded* by factor N, wrapping N times around the unit circle and adding. For $N = 2$, two partial spectra are summed, as indicated below:



Twiddle Factor Notation

In FFT terminology, W_N^k denotes the kth "twiddle factor," where W_N is a primitive Nth root of unity:

$$W_N \stackrel{\Delta}{=} e^{-j2\pi/N}.$$

The aliasing expression can therefore be written as

$$\begin{aligned} Y(z) &= \frac{1}{N}\sum_{m=0}^{N-1} X\left(z^{\frac{1}{N}} e^{-jm\frac{2\pi}{N}}\right), \; z \in \mathbb{C} \\ &= \frac{1}{N}\sum_{m=0}^{N-1} X(W_N^m z^{1/N}). \end{aligned}$$

Proof of Downsampling/Aliasing Relationship

$$\boxed{\text{Downsample}_N(x) \leftrightarrow \frac{1}{N}\text{Alias}_N(X)}$$

or $$\boxed{x(nN) \leftrightarrow \frac{1}{N}\sum_{m=0}^{N-1} X\left(e^{j2\pi m/N} z^{1/N}\right)}$$

From the DFT case, we know this is true when x and X are each complex sequences of length N_s, in which case y and Y are length N_s/N. Thus,

$$x(nN) \leftrightarrow Y(\omega_k N) = \frac{1}{N}\sum_{m=0}^{N-1} X\left(\omega_k + \frac{2\pi}{N}m\right),\ k \in \left[0, \frac{N_s}{N}\right)$$

where we have chosen to keep frequency samples ω_k in terms of the original frequency axis prior to downsampling, i.e., $\omega_k = 2\pi k/N_s$ for both X and Y. This choice allows us to easily take the limit as $N_s \to \infty$ by simply replacing ω_k by ω:

$$x(nN) \leftrightarrow Y(\omega N) = \frac{1}{N}\sum_{m=0}^{N-1} X\left(\omega + \frac{2\pi}{N}m\right),\ \omega \in \left[0, \frac{2\pi}{N}\right)$$

Replacing ω by $\omega' = \omega N$ and converting to z-transform notation $X(z)$ instead of Fourier transform notation $X(\omega)$, with $z = e^{j\omega'}$, yields the final result.

Example: Downsampling by 2

As an example, when $N = 2$, $y[n] = x[2n]$, and (since $W_2 \stackrel{\Delta}{=} e^{-j2\pi/2} = -1$)

$$\begin{aligned}
Y(z) &= \frac{1}{2}\left[X\left(W_2^0 z^{1/2}\right) + X\left(W_2^1 z^{1/2}\right)\right] \\
&= \frac{1}{2}\left[X\left(e^{-j2\pi 0/2} z^{1/2}\right) + X\left(e^{-j2\pi 1/2} z^{1/2}\right)\right] \\
&= \frac{1}{2}\left[X\left(z^{1/2}\right) + X\left(-z^{1/2}\right)\right] \\
&= \frac{1}{2}\left[\textsc{Stretch}_2(X) + \textsc{Stretch}_2\left(\textsc{Shift}_\pi(X)\right)\right]
\end{aligned}$$

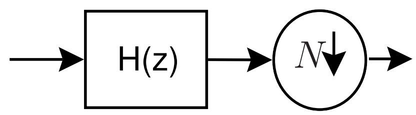
Example: Upsampling by 2

When $N = 2$, $y = [x_0, 0, x_1, 0, \ldots]$, and

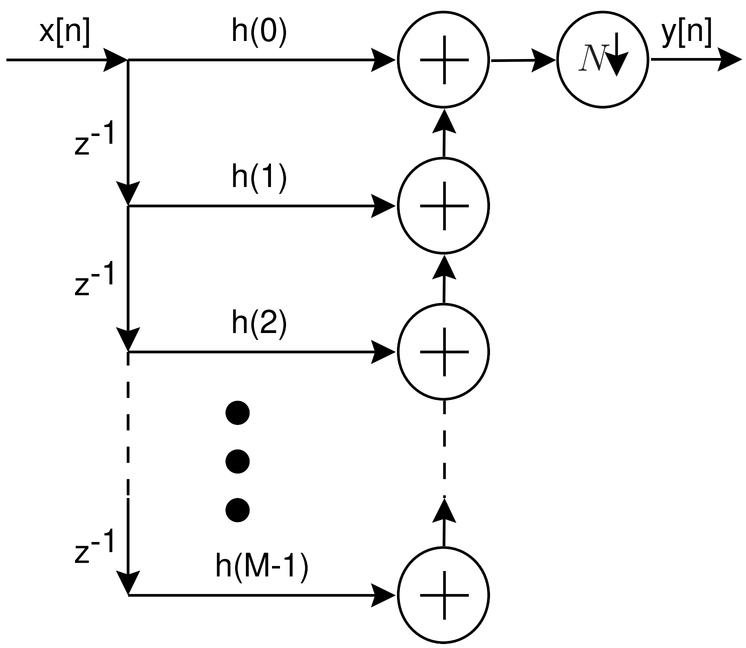
$$Y(z) = X(z^2) = \textsc{Repeat}_2(X)$$

Filtering and Downsampling

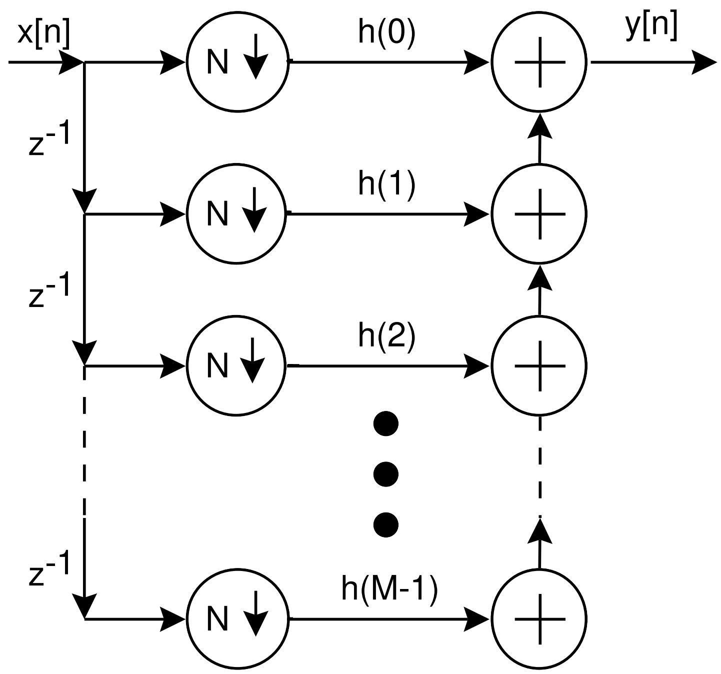
Because downsampling by N will cause aliasing for any frequencies in the original signal above $|\omega| > \pi/N$, the input signal must first be lowpass filtered.



The lowpass filter h[n] is an FIR filter of length M with a cutoff frequency of π/N. Let's draw the FIR filter h in direct form:

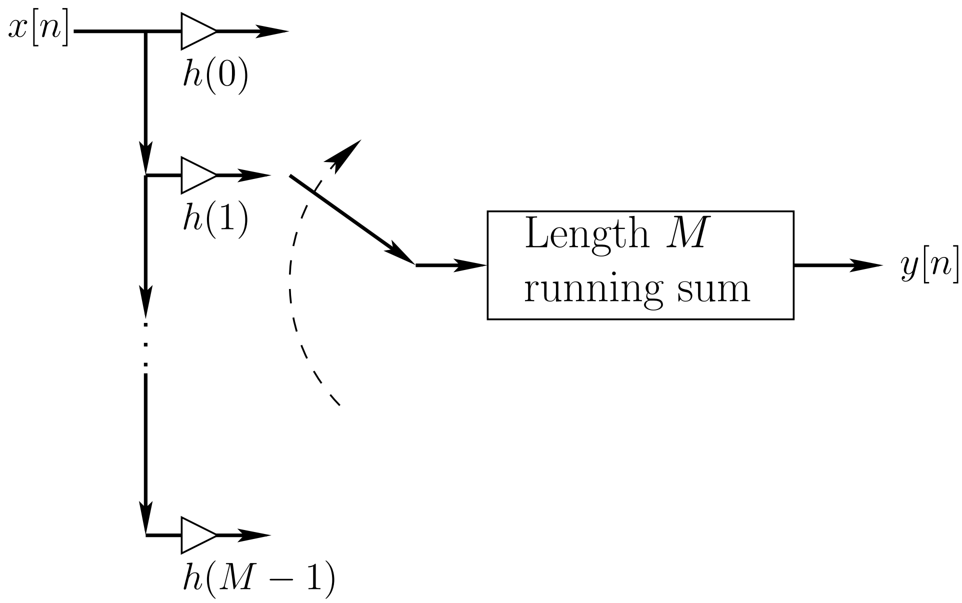


- Note that we do not need $N-1$ out of every N samples due to the $N:1$ downsampler.
- Commute the downsampler through the adders inside the FIR filter:



- The multipliers are now running at $1/N$ times the sampling frequency of the input signal, $x[n]$.
  This reduces the computation requirements by $1/N$.
- The downsampler outputs are called *polyphase signals*
- This is a summed *polyphase filter bank* in which each "subphase filter" is a constant scale factor $h(m)$.

- Interpretation:
  - serial to parallel conversion
    from a stream of scalar samples $x[n]$
    to a sequence of length M buffers every N
    samples, followed by
  - a dot product of each buffer with $h(0 : M - 1)$
- For $N = M$, the overall system is equivalent to a *round-robin demultiplexor*, with a different gain $h(m)$ for each output, followed by an M-sample summer which adds the "de-interleaved" signals together:



Polyphase Processing (Anti-Aliasing Filter)

- Subphase 0,

$$x(nN)\,|_{n=0}^{\infty} = [x_0, x_N, x_{2N}, \ldots]$$

is scaled by $h(0)$

- Subphase 1,

$$x(nN+1)\,|_{n=0}^{\infty} = [x_1, x_{N+1}, x_{2N+1}, \ldots]$$

is scaled by $h(1)$

- $\cdots$

- Subphase m,

$$x(nN+m)\,|_{n=0}^{\infty} = [x_m, x_{N+m}, x_{2N+m}, \ldots]$$

is scaled by $h(m)$.

Polyphase Filtering

In multirate signal processing, it is often fruitful to split a signal or filter into its *polyphase components*.

Let's look at the case $N = 2$:

- Begin with the filter

$$H(z) = \sum_{n=-\infty}^{\infty} h(n) z^{-n}$$

- Separate the even and odd terms:

$$H(z) = \sum_{n=-\infty}^{\infty} h(2n) z^{-2n} + z^{-1} \sum_{n=-\infty}^{\infty} h(2n+1) z^{-2n}$$

- Define the *polyphase component filters*:

$$\begin{aligned} E_0(z) &= \sum_{n=-\infty}^{\infty} h(2n) z^{-n} \\ E_1(z) &= \sum_{n=-\infty}^{\infty} h(2n+1) z^{-n} \end{aligned}$$

  $E_0(z)$ and $E_1(z)$ are the *polyphase components* of the *polyphase decomposition* of $H(z)$ for $N = 2$.

- Now write $H(z)$ as the sum of the odd and even terms:

$$\boxed{H(z) = E_0(z^2) + z^{-1}E_1(z^2)}$$

Example Polyphase Decomposition into 2 Channels

As a simple example, consider

$$H(z) = 1 + 2z^{-1} + 3z^{-2} + 4z^{-3}.$$

Then the even and odd terms are, respectively,

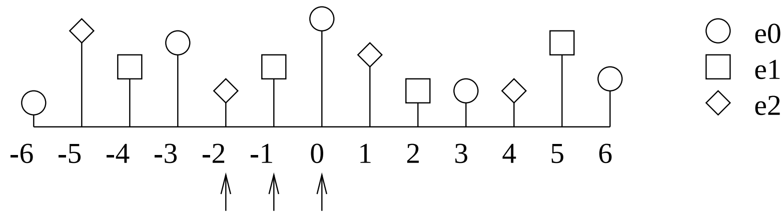
$$\begin{aligned} E_0(z) &= 1 + 3z^{-1} \\ E_1(z) &= 2 + 4z^{-1} \end{aligned}$$

And so $H(z)$ can be written as the sum of the following two polyphase components:

$$\begin{aligned} E_0(z^2) &= 1 + 3z^{-2} \\ z^{-1}E_1(z^2) &= 2z^{-1} + 4z^{-3} \end{aligned}$$

Polyphase Decomposition into N Channels

For the general case of arbitrary N, the basic idea is to decompose $x[n]$ into its periodically interleaved subsequences:



The polyphase decomposition into N channels is given by

$$H(z) = \sum_{l=0}^{N-1} z^{-l} E_l(z^N)$$

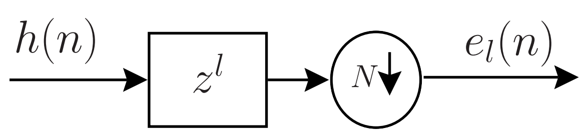
where the subphase filters are

$$E_l(z) = \sum_{n=-\infty}^{\infty} e_l(n) z^{-n},\ l = 0, 1, \ldots, N-1,$$

with

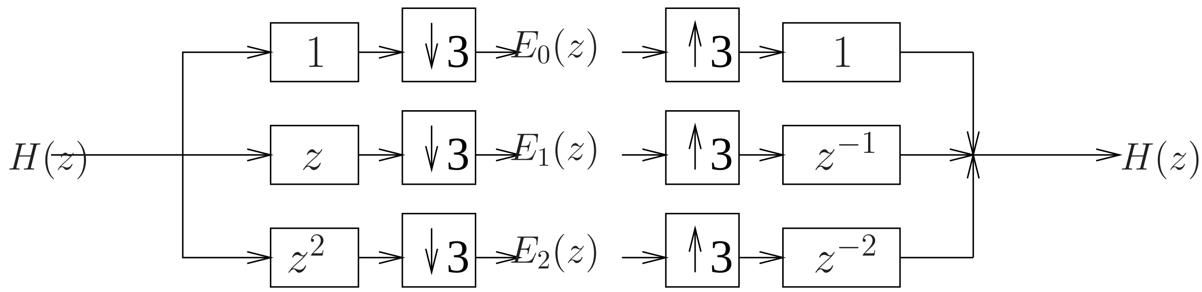
$$e_l(n) \stackrel{\Delta}{=} h(Nn + l). \qquad (l\text{th subphase filter})$$

The signal $e_l(n)$ can be obtained by passing $h(n)$ through an advance of l samples, followed by downsampling by the factor N:



Three-Channel Polyphase Decomposition and Reconstruction

For $N = 3$, we have the following system diagram:



Type II Polyphase Decomposition

The preceding polyphase decomposition of $H(z)$ into N channels

$$H(z) = \sum_{l=0}^{N-1} z^{-l} E_l(z^N)$$

can be termed a "Type I" polyphase decomposition.

In the "Type II", or *reverse* polyphase decomposition, the powers of z progress in the opposite direction:

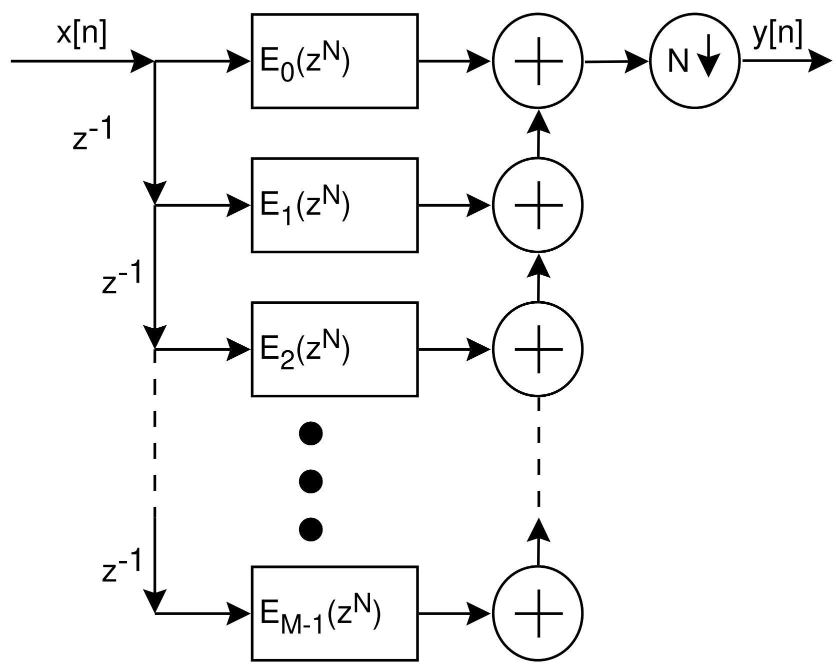
$$H(z) = \sum_{l=0}^{N-1} z^{-(N-l-1)} R_l(z^N).$$

We will see later that we need Type I for *analysis* filters and Type II for *synthesis filters* in a "perfect reconstruction filter bank".

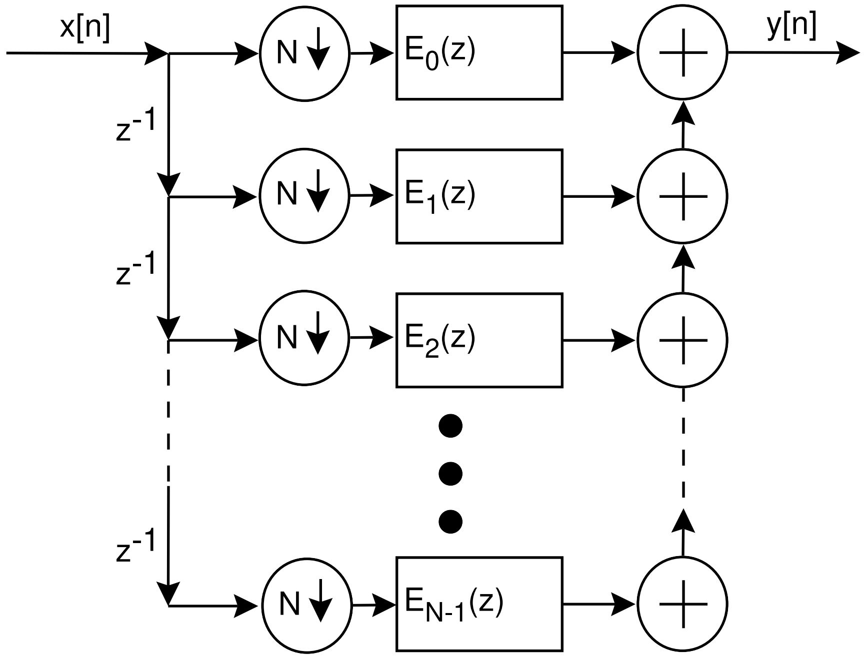
Filtering and Downsampling, Revisited

As another example of polyphase filtering, we return to the previous example about downsampling and filtering. This time,

- Let the FIR lowpass filter h[n] be of length $M = LN, L \in \mathbb{Z}$
- The N polyphase filters, $e_l[n]$, are each length L.
- Recall, $H(z) = E_0(z^N) + z^{-1}E_1(z^N) + z^{-2}E_2(z^N) + \cdots + z^{-(N-1)}E_{N-1}(z^N)$:



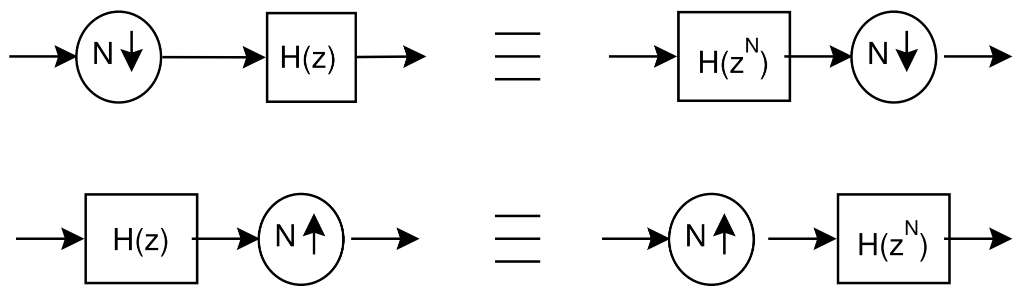
- Now commute the $N:1$ downsampler through the adders and through the upsampled polyphase filters, $E_l(z^N)$:



Commuting the downsampler through the subphase filters $E_l(z^N)$ to get $E_l(z)$ is an example of a "multirate noble identity".

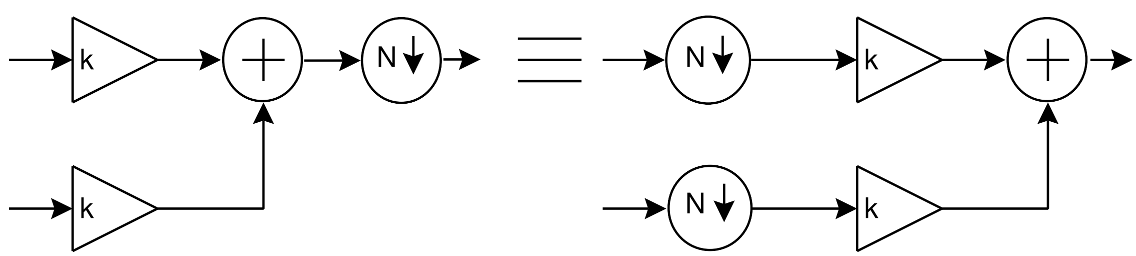
Multirate Noble Identities

Downsamplers and upsamplers are linear, time-varying operators. Therefore, operation order is very important.



Multirate noble identities

It is also important to note that adders or multipliers (any memoryless operators) can commute across downsamplers and upsamplers:



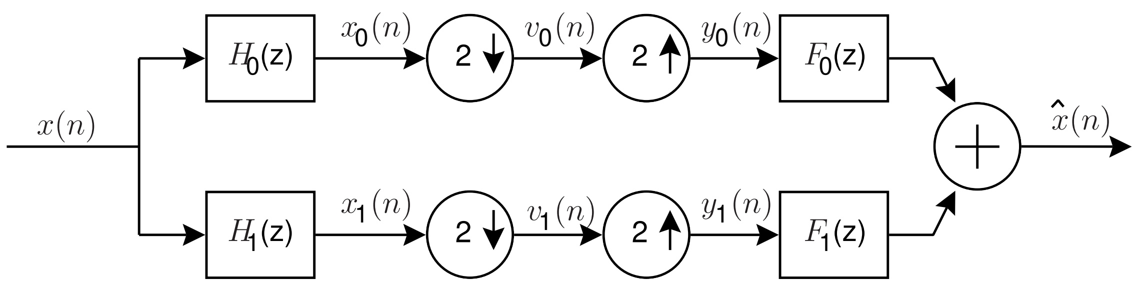
Critically Sampled Perfect Reconstruction Filter Banks

- A *perfect reconstruction (PR) filter bank* is any filter bank whose reconstruction is the original signal, possibly delayed, and possibly scaled by a constant.
- In this context, *critical sampling* (also called "maximal downsampling") means that the downsampling factor is the same as the number of filter channels. For the STFT, this implies $R = M = N$ (with $M > N$ for Portnoff windows).
- The short-Time Fourier transform (STFT) is a PR filter bank whenever the constant-overlap-add (COLA) condition is met by the analysis window w and the hop size R. However, *only the rectangular window case with no zero-padding is critically sampled* (OLA hop size = FBS downsampling factor = N).
- Advanced audio compression algorithms ("perceptual audio coding") are based on critically sampled filter banks, for obvious reasons.
- **Important Point:** *We normally do not require critical sampling for audio analysis, effects, and music*

applications. We normally only need it when *compression* is a requirement.

Two-Channel Critically Sampled Filter Banks

Let's begin with a simple two-channel case, with lowpass analysis filter $H_0(z)$, highpass analysis filter $H_1(z)$, lowpass synthesis filter $F_0(z)$, and highpass synthesis filter $F_1(z)$:



The outputs of the two analysis filters are then

$$X_k(z) = H_k(z)X(z), \quad k = 0, 1$$

After downsampling, the signals become

$$V_k(z) = \frac{1}{2}\left[X_k(z^{1/2}) + X_k(-z^{1/2})\right], \; k = 0, 1$$

After upsampling, the signals become

$$\begin{aligned} Y_k(z) = V_k(z^2) &= \frac{1}{2}[X_k(z) + X_k(-z)] \\ &= \frac{1}{2}[H_k(z)X(z) + H_k(-z)X(-z)], \; k = 0, 1 \end{aligned}$$

After substitutions and rearranging, the output $\hat{x}$ is a filtered replica plus an aliasing term:

$$\begin{aligned}\hat{X}(z) &= \frac{1}{2}[H_0(z)F_0(z) + H_1(z)F_1(z)]X(z) \\ &\quad + \frac{1}{2}[H_0(-z)F_0(z) + H_1(-z)F_1(z)]X(-z) \\ &\quad \text{(Filter Bank Reconstruction)} \qquad (1)\end{aligned}$$

We require the second term (the *aliasing term*) to be zero for perfect reconstruction. This is arranged if we set

$$\begin{aligned}F_0(z) &= \quad H_1(-z) \\ F_1(z) &= -H_0(-z) \\ &\quad \text{(Aliasing Cancellation Constraints)} \qquad (2)\end{aligned}$$

Thus,

- The synthesis lowpass filter $F_0(z)$ is the rotation by π of the analysis highpass filter $H_1(z)$ on the unit circle. If $H_1(z)$ is highpass, cutting off at $\omega = \pi/2$, then $F_0(z)$ will be lowpass, cutting off at $\pi/2$.
- The synthesis highpass filter $F_1(z)$ is the negative of the π-rotation of the analysis lowpass filter $H_0(z)$.

Note that aliasing is completely canceled by this choice of synthesis filters F_0, F_1, for any choice of analysis filters H_0, H_1.

For perfect reconstruction, we additionally need

$$c = H_0(z)F_0(z) + H_1(z)F_1(z)$$
$$\text{(Filtering Cancellation Constraint)} \qquad (3)$$

where $c = Ae^{-j\omega D}$ is any constant $A > 0$ times a linear-phase term corresponding to D samples of delay.

Choosing F_0 and F_1 to cancel aliasing,

$$c = H_0(z)H_1(-z) - H_1(z)H_0(-z)$$
$$\text{(Filtering and Aliasing Cancellation)} \qquad (4)$$

Perfect reconstruction thus also imposes a constraint on the analysis filters, which is of course true for any band-splitting filter bank.

Let $\tilde{H}$ denote $H(-z)$. Then both constraints can be expressed in matrix form as

$$\begin{bmatrix} H_0 & H_1 \\ \tilde{H}_0 & \tilde{H}_1 \end{bmatrix} \begin{bmatrix} F_0 \\ F_1 \end{bmatrix} = \begin{bmatrix} c \\ 0 \end{bmatrix}$$

Amplitude-Complementary 2-Channel Filter Bank

Perhaps the most natural choice of analysis filters for our two-channel, critically sampled filter bank, is an *amplitude-complementary* lowpass/highpass pair, i.e.,

$$H_1(z) = 1 - H_0(z)$$

where we impose the unity dc gain constraint $H_0(1) = 1$.

Amplitude-complementary thus means *constant overlap-add* (COLA) on the unit circle in the z plane.

Plugging the COLA constraint into the Filtering and Aliasing Cancellation constraint (4) gives

$$\begin{aligned}
c &= H_0(z)[1 - H_0(-z)] - [1 - H_0(z)]H_0(-z) \\
&= H_0(z) - H_0(-z) \quad \longleftrightarrow \\
A\delta(n - D) &= h_0(n) - (-1)^n h_0(n) \\
&= \begin{cases} 0, & n \text{ even} \\ 2h_0(n), & n \text{ odd} \end{cases}
\end{aligned}$$

- Even-indexed terms of the impulse response are *unconstrained*, since they subtract out in the constraint.
- For perfect reconstruction, *exactly one odd-indexed term must be nonzero* in the lowpass impulse response $h_0(n)$. The simplest choice is $h_0(1) \neq 0$.

- Thus, the lowpass-filter impulse response can be anything of the form

  $$h_0 = [h_0(0), \mathbf{h_0(1)}, h_0(2), 0, h_0(4), 0, h_0(6), 0, \ldots]$$

  or

  $$h_0 = [h_0(0), 0, h_0(2), \mathbf{h_0(3)}, h_0(4), 0, h_0(6), 0, \ldots]$$

  etc.
- The corresponding highpass-filter impulse response is

  $$h_1(n) = \delta(n) - h_0(n).$$

- The first example above corresponds to the highpass-filter

  $$h_1 = [1{-}h_0(0), -h_0(1), -h_0(2), 0, -h_0(4), 0, -h_0(6), 0, \ldots]$$

  etc.

The above class of amplitude-complementary filters can be characterized as follows:

$$\begin{aligned} H_0(z) &= E_0(z^2) + h_0(o)z^{-o}, \quad E_0(1) + h_0(o) = 1, \; o \text{ odd} \\ H_1(z) &= 1 - H_0(z) = 1 - E_0(z^2) - h_0(o)z^{-o} \end{aligned}$$

In summary, we have shown that an amplitude-complementary lowpass/highpass analysis filter pair yields perfect reconstruction (aliasing and filtering cancellation) when there is exactly one odd-indexed term in the impulse response of $h_0(n)$.

Problem:

- $E_0(z^2)$ repeats twice around the unit circle.
- Since we assume real coefficients, the frequency response, $E_0(e^{j2\omega})$ is magnitude-symmetric about $\omega = \pi/2$ as well as π.
- This is not good since we only have one degree of freedom, $h_0(o)z^{-o}$, with which we can break the $\pi/2$ symmetry to reduce the high-frequency gain and/or boost the low-frequency gain.
- In other words, this class of filters cannot be expected to give us high quality lowpass or highpass behavior.

To enable the use of high-quality lowpass and highpass channel filters, we must relax the amplitude-complementary constraint (and/or filtering cancellation and/or aliasing cancellation) and find another approach.

Haar Example

Before we leave this case (amplitude-complementary, two-channel, critically sampled, perfect reconstruction filter banks), let's see what happens when $H_0(z)$ is the simplest possible lowpass filter having unity dc gain, i.e.,

$$H_0(z) = \frac{1}{2} + \frac{1}{2}z^{-1}$$

This case is obtained above by setting $E_0(z^2) = 1/2$, $o = 1$, and $h_0(1) = 1/2$.

The polyphase components of $H_0(z)$ are clearly

$$E_0(z^2) = E_1(z^2) = 1/2.$$

Choosing $H_1(z) = 1 - H_0(z)$ and choosing $F_0(z)$ and $F_1(z)$ for aliasing cancellation, the four filters become

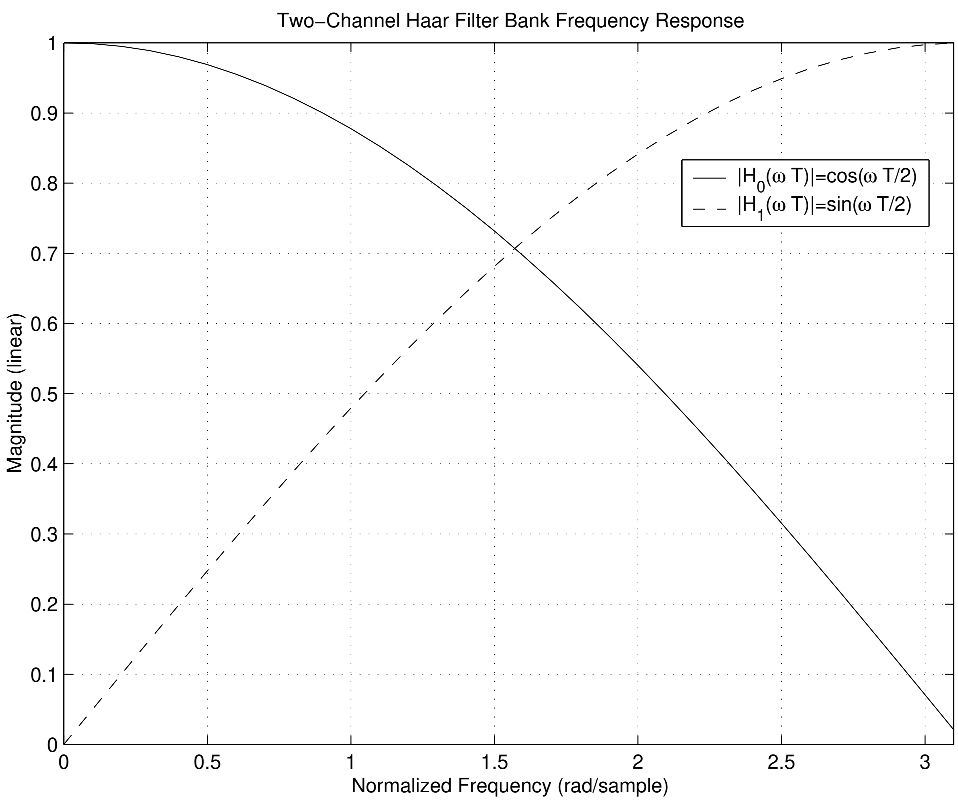
$$\begin{aligned}
H_0(z) &= \frac{1}{2} + \frac{1}{2}z^{-1} = E_0(z^2) + z^{-1}E_1(z^2) \\
H_1(z) &= = 1 - H_0(z) = \frac{1}{2} - \frac{1}{2}z^{-1} = E_0(z^2) - z^{-1}E_1(z^2) \\
F_0(z) &= H_1(-z) = \frac{1}{2} + \frac{1}{2}z^{-1} = H_0(z) \\
F_1(z) &= -H_0(-z) = -\frac{1}{2} + \frac{1}{2}z^{-1} = -H_1(z)
\end{aligned}$$

Thus, both the analysis and reconstruction filter banks are scalings of the familiar *Haar filters* ("sum and difference" filters $(1 \pm z^{-1})/\sqrt{2}$).

The frequency responses are

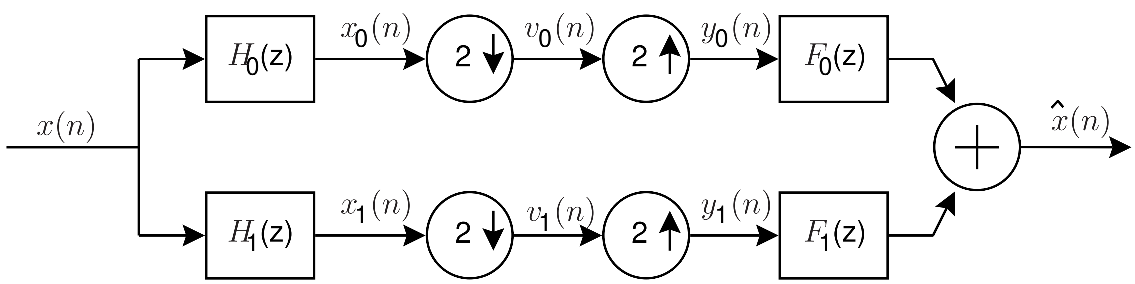
$$\begin{aligned} H_0(e^{j\omega}) &= \quad F_0(e^{j\omega}) = \frac{1}{2} + \frac{1}{2}e^{-j\omega} = e^{-j\frac{\omega}{2}} \cos\left(\frac{\omega}{2}\right) \\ H_1(e^{j\omega}) &= -F_0(e^{j\omega}) = \frac{1}{2} - \frac{1}{2}e^{-j\omega} = je^{-j\frac{\omega}{2}} \sin\left(\frac{\omega}{2}\right) \end{aligned}$$

which are plotted below:



Polyphase Haar Example

Let's look at the polyphase representation for this example. Starting with the filter bank and its reconstruction,



the polyphase decomposition of $H_0(z)$ is

$$H_0(z) = E_0(z^2) + z^{-1}E_1(z^2) = \frac{1}{2} + \frac{1}{2}z^{-1}$$

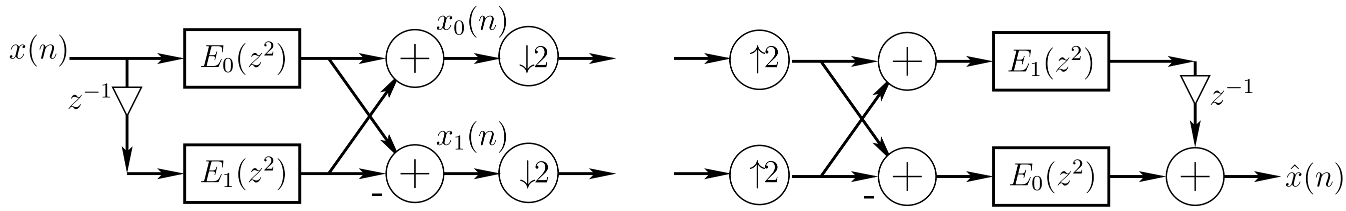
Thus, $E_0(z^2) = E_1(z^2) = 1/2$, and therefore

$$H_1(z) = 1 - H_0(z) = E_0(z) - z^{-1}E_1(z)$$

We may derive polyphase synthesis filters as follows:

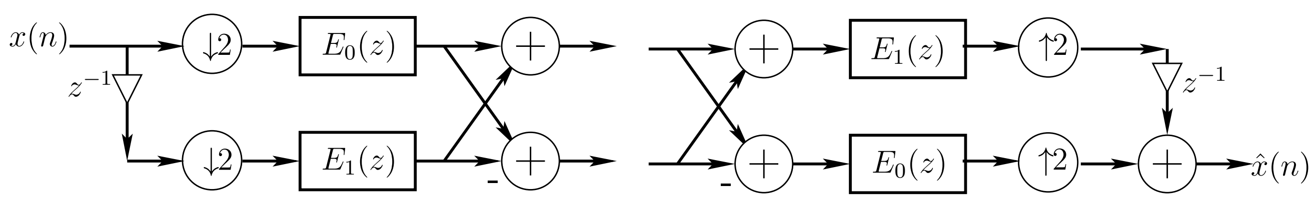
$$\begin{aligned}\hat{X}(z) &= \left[F_0(z)H_0(z) + F_1(z)H_1(z)\right]X(z)\\ &= \left[\left(\frac{1}{2} + \frac{1}{2}z^{-1}\right)H_0(z) + \left(-\frac{1}{2} + \frac{1}{2}z^{-1}\right)H_1(z)\right]\\ &= \frac{1}{2}\left\{\left[H_0(z) - H_1(z)\right] + z^{-1}\left[H_0(z) + H_1(z)\right]\right\}\end{aligned}$$

The polyphase representation of the filter bank and its reconstruction can now be drawn as below:



Notice that the reconstruction filter bank is formally the *transpose* of the analysis filter bank.

Commuting the downsamplers (by the noble identities), we obtain



Since $E_0(z) = E_1(z) = 1/2$, this is simply the OLA form of an *STFT filter bank* for $N = 2$, with $N = M = R = 2$, and rectangular window $w = [1/2, 1/2]$. That is, the DFT size, window length, and hop size are all 2, and both the DFT and its inverse are simply sum-and-difference operations.

Quadrature Mirror Filterbanks (QMF)

The well studied subject of Quadrature Mirror Filters (QMF) is entered by imposing the following symmetry constraint on the analysis filters:

$$H_1(z) = H_0(-z) \quad \text{(QMF Symmetry Constraint)} \quad (5)$$

That is, the filter for channel 1 is constrained to be a π-rotation of filter 0 along the unit circle. In the time domain, $h_1(n) = (-1)^n h_0(n)$, i.e., all odd-index coefficients are negated.

Two-channel QMFs have been around since at least 1976 (see Croisier et al. in Music 421 Citations), and appear to be the first critically sampled perfect reconstruction filter banks.

If H_0 is a lowpass filter cutting off near $\omega = \pi/2$ (as is typical), then H_1 is a complementary highpass filter. The exact cut-off frequency can be adjusted along with the roll-off rate to to provide a maximally constant frequency-response sum.

Historically, the term QMF applied only to two-channel filter banks having the QMF symmetry constraint (5). Today, the term "QMF filter bank" may refer to more general PR filter banks with any number of channels and not obeying (5).

Combining the QMF symmetry constraint with the aliasing-cancellation constraints, given by

$$\begin{aligned} F_0(z) &= H_1(-z) = H_0(z) \\ F_1(z) &= -H_0(-z) = -H_1(z), \end{aligned}$$

the perfect reconstruction requirement reduces to

$$\begin{aligned} \text{constant} &= H_0(z)F_0(z) + H_1(z)F_1(z) = H_0^2(z) - H_0^2(-z) \\ & \text{(QMF Perfect Reconstruction Constraint)} \qquad (6) \end{aligned}$$

Now, all four filters are determined by $H_0(z)$.

It is easy to show using the polyphase representation of $H_0(z)$ (see Vaidyanathan) that the only causal FIR QMF analysis filters yielding exact perfect reconstruction are *two-tap FIR filters* of the form

$$\begin{aligned} H_0(z) &= c_0 z^{-2n_0} + c_1 z^{-(2n_1+1)} \\ H_1(z) &= c_0 z^{-2n_0} - c_1 z^{-(2n_1+1)} \end{aligned}$$

where c_0 and c_1 are constants, and n_0 and n_1 are integers.

- Only weak channel filters are available in the QMF case ($H_1(z) = H_0(-z)$), as we saw in the amplitude-complementary case.

- On the other hand, very high quality IIR solutions are possible. See Vaidyanathan for details (pp. 201–204).
- In practice, approximate "pseudo QMF" filters are common, which only give approximate perfect reconstruction. We'll return to this topic later.

The *Haar filters*, which we saw gave perfect reconstruction in the amplitude-complementary case, are also examples of a QMF filter bank:

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

In this example, $c_0 = c_1 = 1$, and $n_0 = n_1 = 0$.

Linear Phase Quadrature Mirror Filter Banks

It is generally desirable to use *linear phase* filters whenever possible in audio work. This is because linear phase filters delay all frequencies by equal amounts.

A filter phase response is linear in ω whenever its impulse response $h_0(n)$ is *symmetric*, i.e.,

$$h_0(L - n) = h_0(n)$$

in which case the frequency response can be expressed as

$$H_0(e^{j\omega}) = e^{-j\omega N/2} \left| H_0(e^{j\omega}) \right|$$

Substituting this into the QMF perfect reconstruction constraint (6) gives

$$\text{constant} = e^{-j\omega N}\left[\left|H_0(e^{j\omega})\right|^2 - (-1)^N\left|H_0(e^{j(\pi-\omega)}\right|^2\right]$$

When N is even, the right hand side of the above equation is forced to zero at $\omega = \pi/2$. Therefore, we will only consider odd N, for which the perfect reconstruction constraint reduces to

$$\text{constant} = e^{-j\omega N}\left[\left|H_0(e^{j\omega})\right|^2 + \left|H_0(e^{j(\pi-\omega)}\right|^2\right]$$

We see that perfect reconstruction is obtained in the linear-phase case whenever the analysis filters are *power complementary*. Since FIR QMF filters are constrained to the two-tap case, this is best accomplished using IIR filters. See Vaidyanathan for details.

Conjugate Quadrature Filters (CQF)

A class of causal, FIR, two-channel, criticially sampled, exact perfect-reconstruction filter-banks is the set of so-called *Conjugate Quadrature Filters* (CQF).

In the z-domain, the CQF relationships are

$$H_1(z) = z^{-(L-1)} H_0(-z^{-1})$$

In the time domain, the analysis and synthesis filters are given by

$$\begin{aligned} h_1[n] &= -(-1)^n h_0[L-1-n] \\ f_0[n] &= h_0[L-1-n] \\ f_1[n] &= -(-1)^n h_0(n) = -h_1(L-1-n) \end{aligned}$$

That is, $f_0 = \textsc{Flip}(h_0)$ for the lowpass channel, and the highpass channel filters are a modulation of their lowpass counterparts by $(-1)^n$.

- Again, all four analysis and synthesis filters are determined by the lowpass analysis filter $H_0(z)$.
- It can be shown that this is an *orthogonal filter bank*.

- The analysis filters $H_0(z)$ and $H_1(z)$ are *power complementary*, i.e.,

$$\left|H_0 e^{j\omega}\right|^2 + \left|H_1 e^{j\omega}\right|^2 = 1 \qquad \text{(Power Complementary)}$$

or

$$\tilde{H}_0(z)H_0(z)+\tilde{H}_1(z)H_1(z) = 1 \qquad \text{(Power Complementary)}$$

where $\tilde{H}_0(z) \stackrel{\Delta}{=} \overline{H}_0(z^{-1})$ denotes the *paraconjugate* of $H_0(z)$ (for real filters H_0). The paraconjugate is the analytic continuation of $\overline{H_0(e^{j\omega})}$ from the unit circle to the z plane.

- Moreover, the analysis filters $H_0(z)$ are *power symmetric*, e.g.,

$$\tilde{H}_0(z)H_0(z)+\tilde{H}_0(-z)H_0(-z) \qquad \text{(Power Symmetric)}$$

- The power symmetric case was introduced by Smith and Barnwell in 1984.

With the CQF constraints, (1) reduces to

$$\hat{X}(z) = \frac{1}{2}[H_0(z)H_0(z^{-1}) + H_0(-z)H_0(-z^{-1})]X(z) \quad (7)$$

Let $P(z) = H_0(z)H_0(-z)$, such that $H_0(z)$ is a spectral factor of the half-band filter $P(z)$ (i.e., $P(e^{j\omega})$ is a nonnegative power response which is lowpass, cutting off near $\omega = \pi/4$). Then, (7) reduces to

$$\hat{X}(z) = \frac{1}{2}[P(z) + P(-z)]X(z) = -z^{-(L-1)}X(z) \quad (8)$$

- The problem of the PR filter design has been reduced to designing one half-band filter, $P(z)$.
- It can be shown that any half-band filter can be written in the form $p[2n] = \delta[n]$. That is, all non-zero even-idexed values of $p[n]$ are set to zero.

A simple design of an FIR half-band filter would be to window a sinc function:

$$p[n] = \frac{\sin[\pi n/2]}{\pi n/2}w[n] \quad (9)$$

where $w[n]$ is any suitable window, such as the Kaiser window.

- Note that as a result of (7), the CQF filters are power complementary. That is, they satisfy:

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

- Also note that the filters H_0 and H_1 are not linear phase.
- It can be shown that there are no two-channel perfect reconstruction filter banks that have all three of the following characteristics (except for the Haar filters):
  - FIR
  - orthogonality
  - linear phase

  In this design procedure, we have chosen to satisfy the first two and give up the third.
- By relaxing "orthogonality" to "biorthogonality", it becomes possible to obtain FIR linear phase filters in a critically sampled, perfect reconstruction filter bank. (See later section on Wavelet filter banks.)

Orthogonal Two-Channel Filter Banks

Recall the reconstruction equation for the two-channel, critically sampled, perfect-reconstruction filter-bank:

$$\begin{aligned}\hat{X}(z) &= \frac{1}{2}[H_0(z)F_0(z) + H_1(z)F_1(z)]X(z) \\ &+ \frac{1}{2}[H_0(-z)F_0(z) + H_1(-z)F_1(z)]X(-z)\end{aligned}$$

This can be written in matrix form as

$$\hat{X}(z) = \frac{1}{2}\begin{bmatrix} F_0(z) \\ F_1(z) \end{bmatrix}^T \begin{bmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{bmatrix} \begin{bmatrix} X(z) \\ X(-z) \end{bmatrix}$$

where the above 2×2 matrix, $\mathbf{H}_m(z)$, is called the *alias component matrix* (or analysis modulation matrix). If

$$\tilde{\mathbf{H}}_m(z)\mathbf{H}_m(z) = 2\mathbf{I}$$

where $\tilde{\mathbf{H}}_m(z) \stackrel{\Delta}{=} \mathbf{H}_m^T(z^{-1})$ denotes the *paraconjugate* of $\mathbf{H}_m(z)$, then the alias component (AC) matrix is *lossless*, and the (real) filter bank is *orthogonal*.

It turns out orthogonal filter banks give perfect reconstruction filter banks for any number of channels.

Orthogonal filter banks are also called *paraunitary* filter banks, which we'll study shortly in polyphase form.

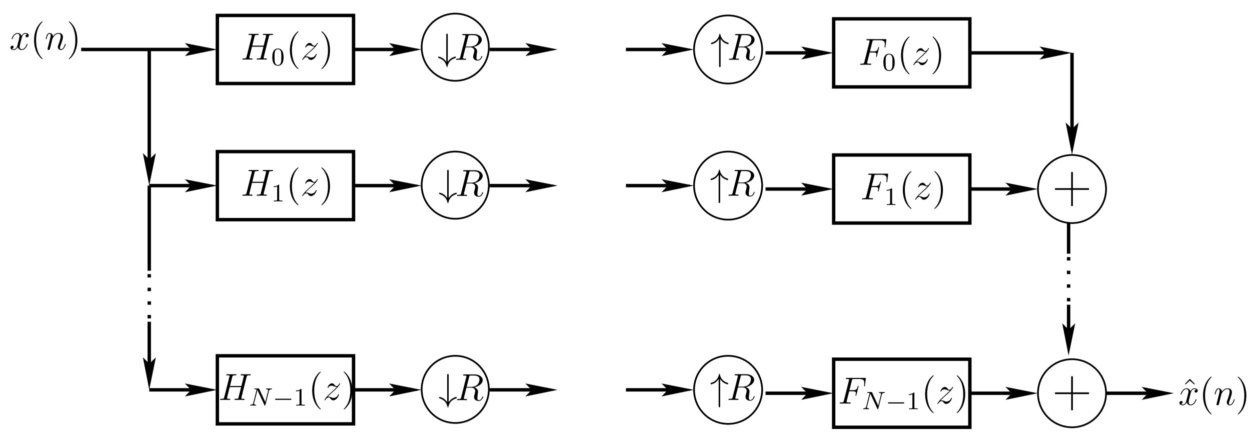
The AC matrix is paraunitary if and only if the *polyphase matrix* is paraunitary. (See Vaidyanathan.)

Perfect Reconstruction Filter Banks

We now consider filter banks with an arbitrary number of channels, and ask under what conditions do we obtain a perfect reconstruction filter bank?

Polyphase analysis will give us the answer readily.

Let's begin with the N-channel filter bank below:



- The downsampling factor is $R \leq N$.
- For critical sampling, $R = N$.

The next step is to expand each analysis filter $H_k(z)$ into its N-channel "Type 1" polyphase representation:

$$H_k(z) = \sum_{l=0}^{N-1} z^{-l} E_{kl}(z^N)$$

or

$$\underbrace{\begin{bmatrix} H_0(z) \\ H_1(z) \\ \vdots \\ H_{N-1}(z) \end{bmatrix}}_{\mathbf{h}(z)} = \underbrace{\begin{bmatrix} E_{0,0}(z^N) & E_{0,1}(z^N) & \cdots & E_{0,N-1}(z^N) \\ E_{1,0}(z^N) & E_{1,1}(z^N) & \cdots & E_{1,N-1}(z^N) \\ \vdots & \vdots & \cdots & \vdots \\ E_{N-1,0}(z^N) & E_{N-1,1}(z^N) & \cdots & E_{N-1,N-1}(z^N) \end{bmatrix}}_{\mathbf{E}(z^N)} \underbrace{\begin{bmatrix} 1 \\ z^{-1} \\ \vdots \\ z^{-(N-1)} \end{bmatrix}}_{\mathbf{e}(z)}$$

which we can write as

$$\mathbf{h}(z) = \mathbf{E}(z^N)\mathbf{e}(z).$$

Similarly, expand the synthesis filters in a Type II polyphase decomposition:

$$F_k(z) = \sum_{l=0}^{N-1} z^{-(N-l-1)} R_{lk}(z^N)$$

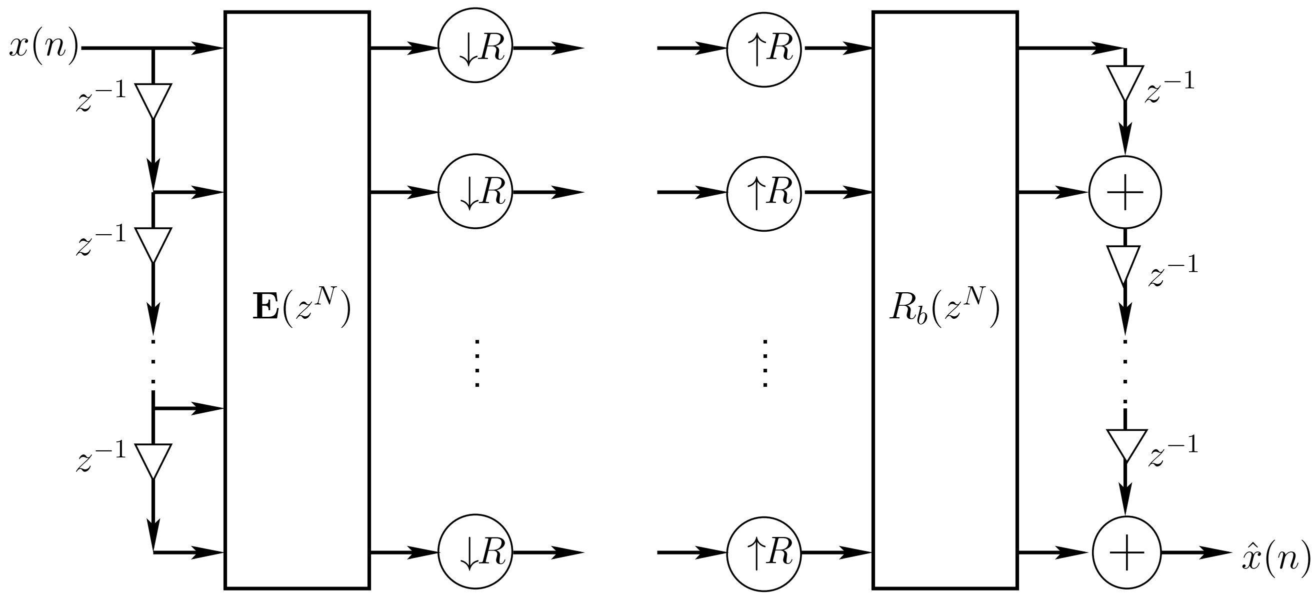
or

$$\underbrace{\begin{bmatrix} F_0(z) \\ F_1(z) \\ \vdots \\ F_{N-1}(z) \end{bmatrix}^T}_{\mathbf{f}^T(z)} = \underbrace{\begin{bmatrix} z^{-(N-1)} \\ z^{-(N-2)} \\ \vdots \\ 1 \end{bmatrix}^T}_{\tilde{\mathbf{e}}(z)} \underbrace{\begin{bmatrix} R_{0,0}(z^N) & R_{0,1}(z^N) & \cdots & R_{0,N-1}(z^N) \\ R_{1,0}(z^N) & R_{1,1}(z^N) & \cdots & R_{1,N-1}(z^N) \\ \vdots & \vdots & \cdots & \vdots \\ R_{N-1,0}(z^N) & R_{N-1,1}(z^N) & \cdots & R_{N-1,N-1}(z^N) \end{bmatrix}}_{\mathbf{R}(z^N)}$$

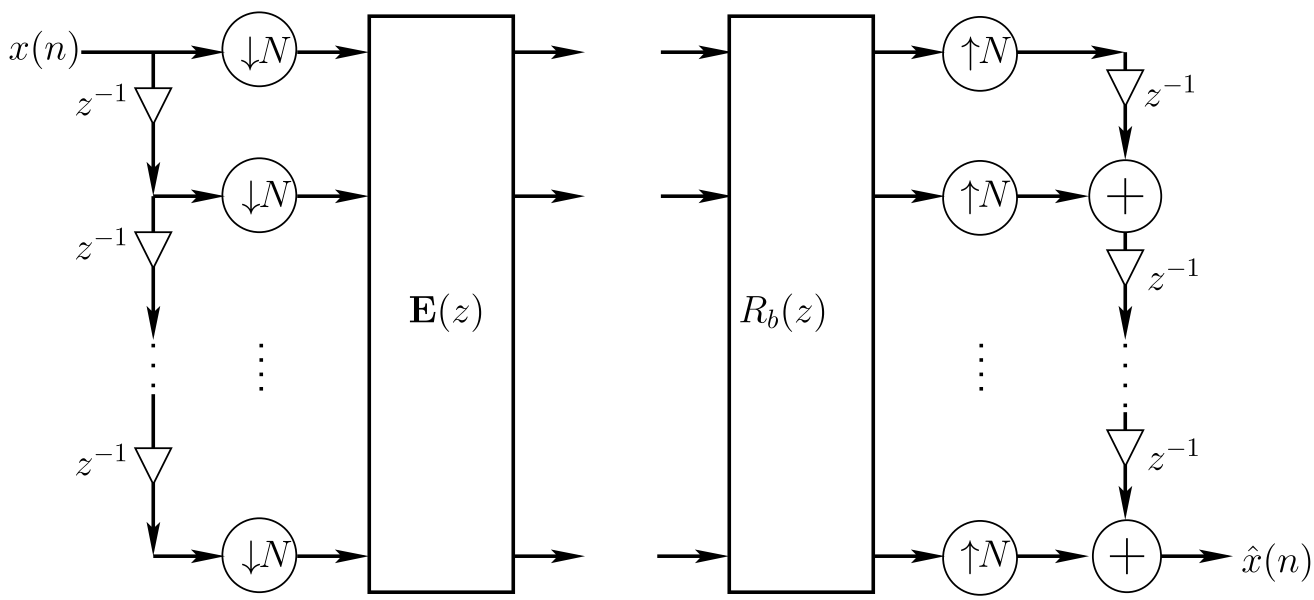
which we can write as

$$\mathbf{f}^T(z) = \tilde{\mathbf{e}}(z)\mathbf{R}(z^N).$$

The polyphase representation can now be depicted as



When $R = N$, commuting the up/downsamplers gives



We call $\mathbf{E}(z)$ the *polyphase matrix*.

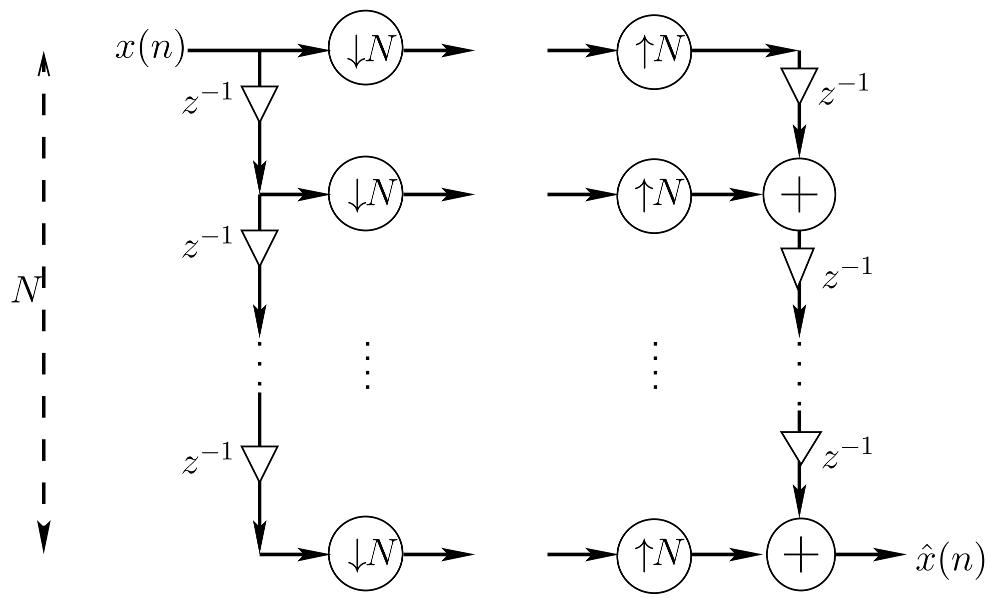
As we will show below, the above simplification can be carried out more generally whenever R *divides* N (e.g., $R = N/2, \ldots, 1$). In these cases $\mathbf{E}(z)$ becomes $\mathbf{E}(z^{N/R})$ and $\mathbf{R}(z)$ becomes $\mathbf{R}(z^{N/R})$.

Simple Examples of Perfect Reconstruction

If we can arrange to have

$$\boxed{\mathbf{R}(z)\mathbf{E}(z) = \mathbf{I}_N}$$

then the filter bank will reduce to the simple system below:



When $R = N$, we have a simple parallelizer/serializer (or de-multiplexor/multiplexor, or de-interleaver/re-interleaver), which is perfect-reconstruction by inspection:

- Think of the input samples $x(n)$ as "filling" a length $N-1$ delay line over $N-1$ sample clocks.

- At time 0, the downsamplers and upsamplers "fire", transferring $x(0)$ (and $N-1$ zeros) from the delay line to the output delay chain, summing with zeros.
- Over the next $N-1$ clocks, $x(0)$ makes its way toward the output, and zeros fill in behind it in the output delay chain.
- Simultaneously, the input "buffer" (delay line) is being filled with samples of $x(n)$.
- At time $N-1$, $x(0)$ makes it to the output.
- At time N, the downsamplers fire again, transferring a length N buffer $[x(1:N)]$ to the upsamplers.
- On the same clock pulse, the upsamplers also fire, transferring N samples to the output delay chain.
- The bottom-most sample $[x(n-N+1)=x(1)]$ goes out immediately at time N.
- Over the next $N-1$ sample clocks, the length $N-1$ output buffer will be "drained" and refilled by zeros.
- Simultaneously, the input buffer will be replaced by new samples of $x(n)$.
- At time $2N$, the downsamplers and upsamplers fire, and the process goes on, repeating with period N.

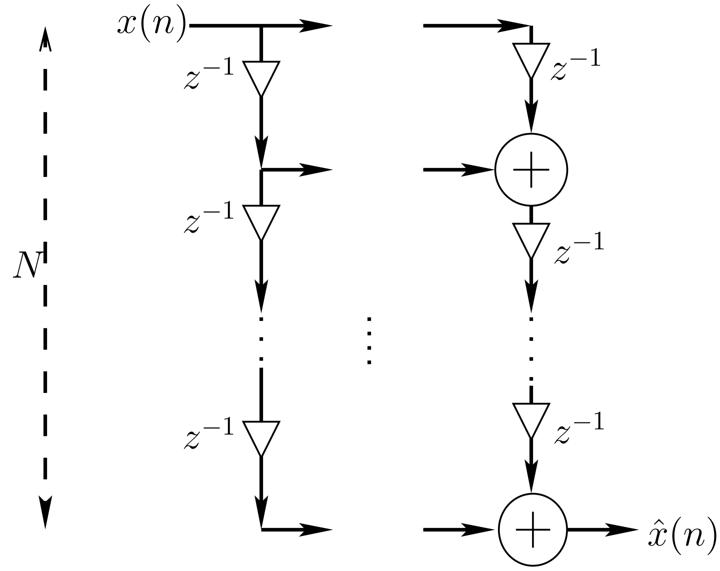
The output of the N-way parellelizer/serializer is therefore

$$\hat{x}(n) = x(n - N + 1)$$

and we have perfect reconstruction.

Sliding Polyphase Filter Bank

When $R = 1$, there is no downsampling or upsampling, and the system further reduces to the case below:



Working backward along the output delay chain, the output sum can be written as

$$\begin{aligned}\hat{X}(z) &= \Big[z^{-0}z^{-(N-1)} + z^{-1}z^{-(N-2)} + z^{-2}z^{-(N-3)} + \cdots \\ &\qquad + z^{-(N-2)}z^{-1} + z^{-0}z^{-(N-1)}\Big] X(z) \\ &= N z^{-(N-1)} X(z)\end{aligned}$$

Thus, when $R = 1$, the output is

$$\hat{x}(n) = Nx(n - N + 1)$$

and we again have perfect reconstruction.

Hopping Polyphase Filter Bank

When $1 < R < N$ and R divides N, we have, by a similar analysis,

$$\hat{x}(n) = \frac{N}{R} x(n - N + 1)$$

which is again perfect reconstruction.

Note the built-in overlap-add when $R < N$.

Sufficient Condition for Perfect Reconstruction

Above, we found that, for any integer $1 \leq R \leq N$ which divides N, a *sufficient* condition for perfect reconstruction is

$$\mathbf{P}(z) \stackrel{\Delta}{=} \mathbf{R}(z)\mathbf{E}(z) = \mathbf{I}_N$$

and the output signal is then

$$\hat{x}(n) = \frac{N}{R} x(n - N + 1)$$

More generally, we allow any nonzero scaling and any additional delay:

$$\mathbf{P}(z) \stackrel{\Delta}{=} \mathbf{R}(z)\mathbf{E}(z) = cz^{-K}\mathbf{I}_N$$

$$\text{(Perfect Reconstruction Constraint)} \quad (10)$$

where $c \neq 0$ is any constant and K is any nonnegative integer. In this case, the output signal is

$$\hat{x}(n) = c\frac{N}{R}x(n - N + 1 - K)$$

Thus, given any polyphase matrix $\mathbf{E}(z)$, we can attempt to compute $\mathbf{R}(z) = \mathbf{E}^{-1}(z)$:

- If it is stable, we can use it to build a perfect-reconstruction filter bank.
- However, if $\mathbf{E}(z)$ is FIR, $\mathbf{R}(z)$ will typically be IIR.
- In the next section, we will look at *paraunitary filter banks*, for which $\mathbf{R}(z)$ is FIR and paraunitary whenever $\mathbf{E}(z)$ is.

Necessary and Sufficient Conditions for Perfect Reconstruction

It can be shown (see Vaidyanathan '93) that the most general conditions for perfect reconstruction are that

$$\boxed{\mathbf{R}(z)\mathbf{E}(z) = cz^{-K}\begin{bmatrix} \mathbf{0}_{(N-L)\times L} & z^{-1}\mathbf{I}_{N-L} \\ \mathbf{I}_L & \mathbf{0}_{L\times(N-L)} \end{bmatrix}}$$

for some constant c and some integer $K \geq 0$, where L is any integer between 0 and $N-1$.

Note that the more general form of $\mathbf{R}(z)\mathbf{E}(z)$ above can be regarded as a (non-unique) *square root of a vector unit delay*, since

$$\begin{bmatrix} \mathbf{0}_{(N-L)\times L} & z^{-1}\mathbf{I}_{N-L} \\ \mathbf{I}_L & \mathbf{0}_{L\times(N-L)} \end{bmatrix}^2 = z^{-1}\mathbf{I}_N.$$

Thus, the general case is the same thing as

$$\mathbf{R}(z)\mathbf{E}(z) = cz^{-K}\mathbf{I}_N.$$

except for some channel swapping and an extra sample of delay in some channels.

Polyphase View of the STFT

As a familiar special case, set

$$\mathbf{E}(z) = \mathbf{W}_N^*$$

where $\mathbf{W}_N^*$ is the *DFT matrix:*

$$\mathbf{W}_N^*[kn] = \left[e^{-j2\pi kn/N}\right].$$

The inverse of this polyphase matrix is then simply the *inverse DFT matrix*:

$$\mathbf{R}(z) = \frac{1}{N}\mathbf{W}_N$$

We see that the STFT can be seen as the simple special case of a perfect reconstruction filter bank for which the polyphase matrix is *constant*. It is also *unitary* when $\mathbf{E}(z) = \mathbf{W}_N^*/\sqrt{N}$ and $\mathbf{R}(z) = \mathbf{W}_N/\sqrt{N}$.

The channel analysis and synthesis filters are, respectively,

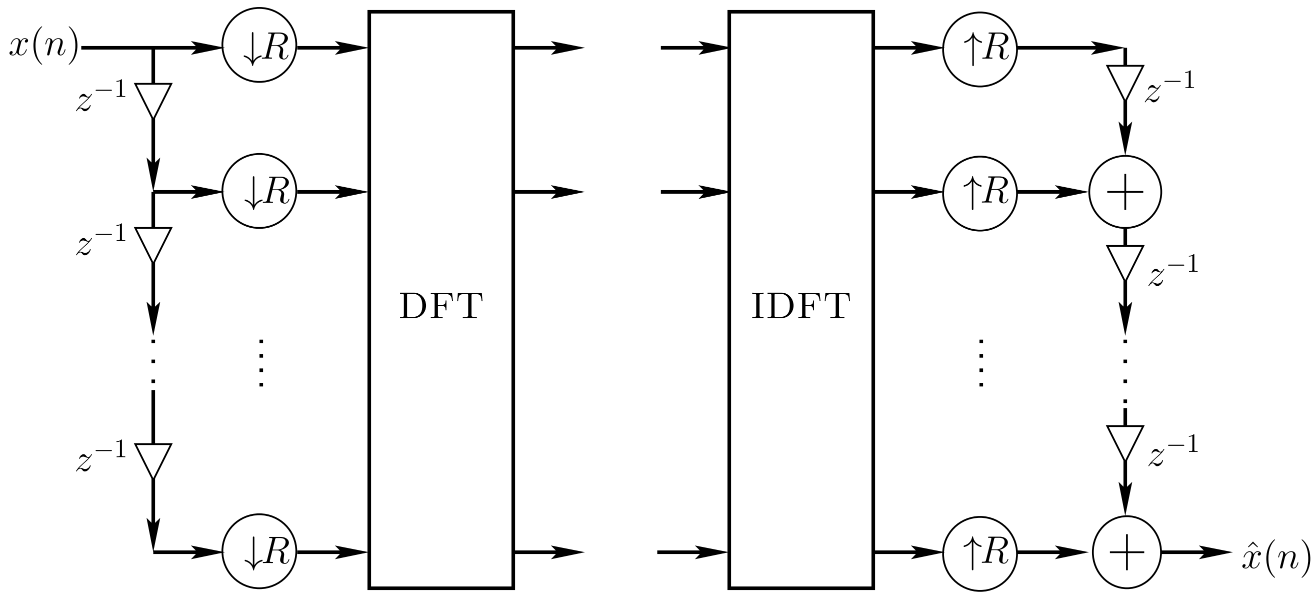
$$\begin{aligned} H_k(z) &= H_0(zW_N^k) \\ F_k(z) &= F_0(zW_N^{-k}) \end{aligned}$$

where $W_N \stackrel{\Delta}{=} e^{-j2\pi/N}$, as usual, and

$$F_0(z) = H_0(z) = \sum_{n=0}^{N-1} z^{-n} \leftrightarrow [1, 1, \ldots, 1]$$

corresponding to the *rectangular window*.

Looking again at the polyphase representation of the N-channel filter bank with hop size R, $\mathbf{E}(z) = \mathbf{W}_N^*$, $\mathbf{R}(z) = \mathbf{W}_N$, R dividing N, we have



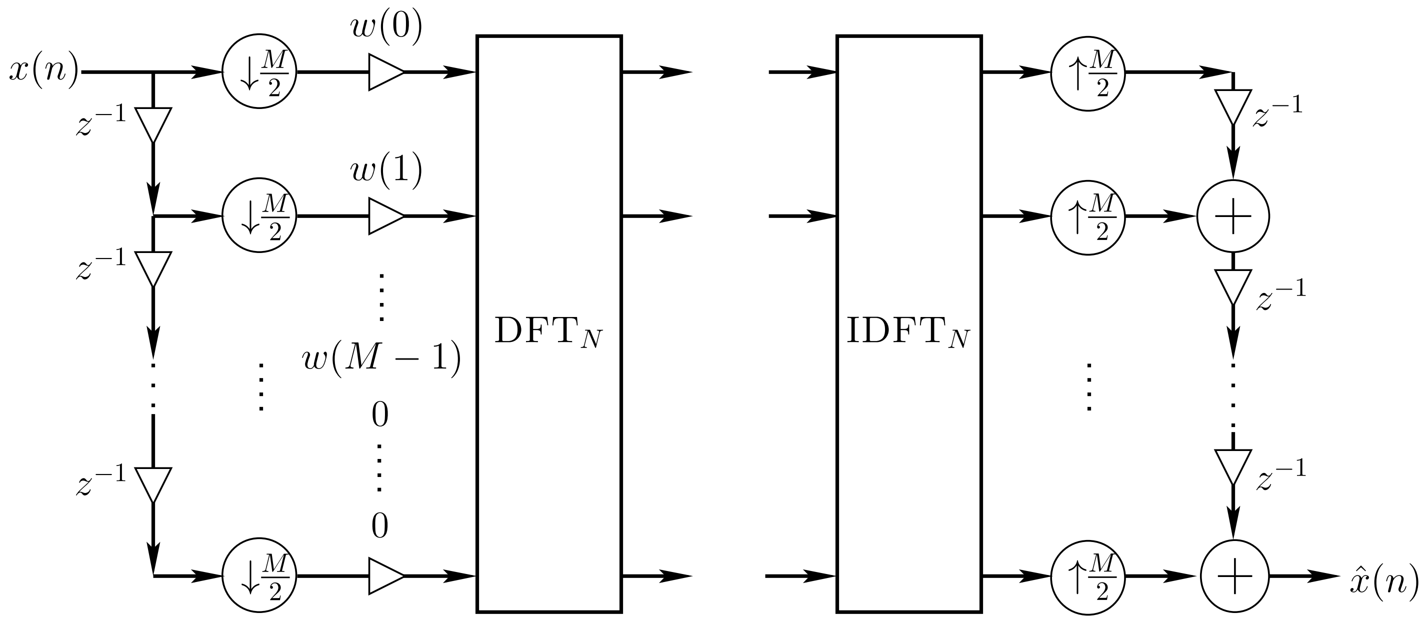
Thus,

$$\boxed{\text{The polyphase representation is an } \textit{overlap-add} \text{ representation}}$$

- Our analysis showed that the STFT using a *rectangular window* is a perfect reconstruction filter bank for all *integer* hop sizes in the set $R \in \{N, N/2, N/3, \ldots, N/N\}$.
- The same type of analysis can be applied to the STFT using the other windows we've studied, *including Portnoff windows*.

Example: Polyphase Analysis of the STFT with 50% Overlap, Zero-Padding, and a Non-Rectangular Window

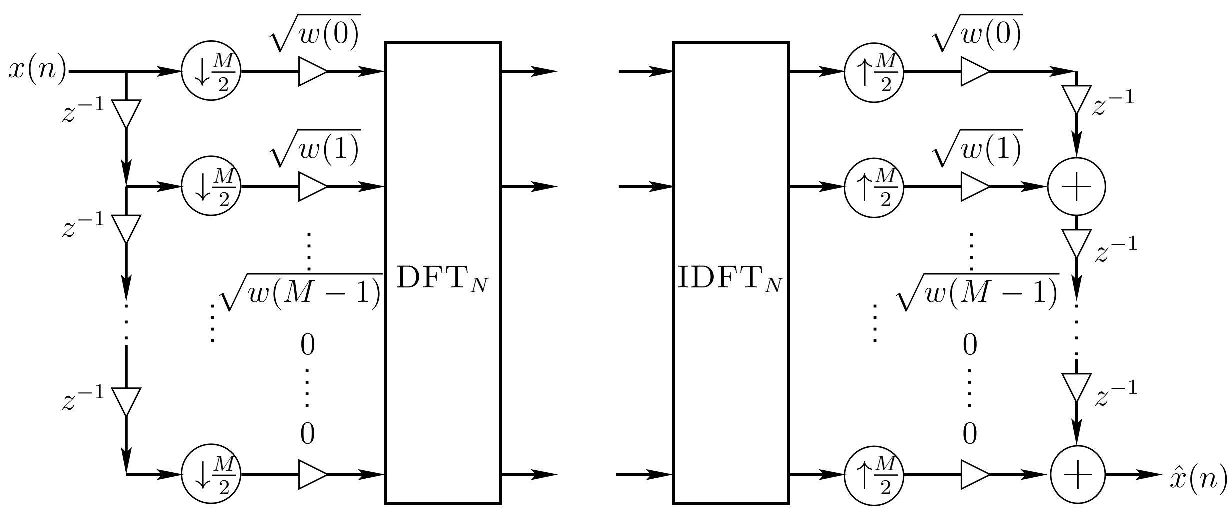
The figure below illustrates how a window and a hop size other than N can be introduced:



- The constant-overlap-add of the window $w(n)$ is implemented in the synthesis delay chain (which is technically the transpose of a tapped delay line).
- The downsampling factor and window must be selected *together* to give constant overlap-add, independent of the choice of polyphase matrices $\mathbf{E}(z)$ and $\mathbf{R}(z)$ (shown here as the DFT and IDFT).

Example: Polyphase Analysis of the Weighted Overlap Add Case: 50% Overlap, Zero-Padding, and a Non-Rectangular Window

We may convert the previous example to a *weighted overlap-add* (WOLA) filter bank by replacing each $w(m)$ by $\sqrt{w(m)}$ and introducing these gains also between the IDFT and upsamplers:



Paraunitary Filter Banks

Paraunitary filter banks form an interesting subset of perfect reconstruction (PR) filter banks:

- We saw above that we get a PR filter bank whenever the analysis polyphase matrix $\mathbf{E}(z)$ times the synthesis polyphase matrix $\mathbf{R}(z)$ is the identity matrix, i.e., when

$$\mathbf{P}(z) \stackrel{\Delta}{=} \mathbf{E}(z)\mathbf{R}(z) = \mathbf{I}$$

- In particular, if $\mathbf{R}(z)$ is the *paraconjugate* of $\mathbf{E}(z)$, we say the filter bank is *paraunitary*.

So what is paraconjugation?

The short answer is that it is the generalization of the complex conjugate transpose operation from the unit circle to the entire z plane. A paraunitary filter bank is therefore a generalization of an *orthogonal* filter bank. Recall that an orthogonal filter bank is one in which $\mathbf{E}(e^{j\omega})$ is an orthogonal (or unitary) matrix, to within a constant scale factor, and $\mathbf{R}(e^{j\omega})$ is its transpose (or Hermitian transpose).

Lossless Filters

To motivate the idea of paraunitary filters, let's first review some properties of lossless filters, progressing from the simplest cases up to paraunitary filter banks:

- A linear, time-invariant filter $H(z)$ is said to be *lossless* (or *allpass*) if it *preserves signal energy*. That is, if the input signal is $x(n)$, and the output signal is $y(n) = (h * x)(n)$, then we have

  $$\sum_{n=-\infty}^{\infty} |y(n)|^2 = \sum_{n=-\infty}^{\infty} |x(n)|^2$$

  In terms of the L_2 signal norm $\| \cdot \|_2$, this can be expressed more succinctly as

  $$\| y \|_2^2 = \| x \|_2^2$$

- Notice that only stable filters can be lossless since, otherwise, $\| y \| = \infty$. We further assume all filters are causal for simplicity.

- It is straightforward to show that losslessness implies

  $$\left| H(e^{j\omega}) \right| = 1, \quad \forall \omega.$$

  That is, the frequency response must have magnitude 1 everywhere on the unit circle in the z plane.

Another way to express this is to write

$$\overline{H(e^{j\omega})}H(e^{j\omega}) = 1, \quad \forall\omega,$$

and this form generalizes to $\tilde{H}(z)H(z)$ over the entire the z plane.

- The *paraconjugate* of a transfer function may be defined as the *analytic continuation of the complex conjugate* from the unit circle to the whole z plane:

  $$\tilde{H}(z) \stackrel{\Delta}{=} \overline{H}(z^{-1})$$

  where $\overline{H}(z)$ denotes complex conjugation of the *coefficients only* of $H(z)$ *and not the powers of* z. For example, if $H(z) = 1 + jz^{-1}$, then $\overline{H}(z) = 1 - jz^{-1}$. We can write, for example,

  $$\overline{H}(z) \stackrel{\Delta}{=} \overline{H\left(\overline{z}\right)}$$

  in which the conjugation of z serves to cancel the outer conjugation.

  We refrain from conjugating z in the definition of the paraconjugate becase $\overline{z}$ is not analytic in the complex-variables sense. Instead, we *invert* z, which *is* analytic, and which reduces to complex conjugation on the unit circle.

  The paraconjugate may be used to characterize allpass filters as follows:

- A causal, stable, filter $H(z)$ is allpass if and only if

$$\tilde{H}(z)H(z) = 1$$

Note that this is equivalent to the previous result on the unit circle since

$$\tilde{H}(e^{j\omega})H(e^{j\omega}) = \overline{H}(1/e^{j\omega})H(e^{j\omega}) = \overline{H(e^{j\omega})}H(e^{j\omega})$$

- To generalize lossless filters to the multi-input, multi-output (MIMO) case, we must generalize conjugation to MIMO transfer function *matrices*:
  - A $p \times q$ transfer function matrix $\mathbf{H}(z)$ is *lossless* if it is stable and its frequency-response matrix $\mathbf{H}(e^{j\omega})$ is *unitary*. That is,

    $$\mathbf{H}^*(e^{j\omega})\mathbf{H}(e^{j\omega}) = \mathbf{I}_q$$

    for all ω, where $\mathbf{I}_q$ denotes the $q \times q$ identity matrix, and $\mathbf{H}^*(e^{j\omega})$ denotes the *Hermitian transpose* (complex-conjugate transpose) of $\mathbf{H}(e^{j\omega})$:

    $$\mathbf{H}^*(e^{j\omega}) \triangleq \overline{\mathbf{H}^T(e^{j\omega})}$$

  - Note that $\mathbf{H}^*(e^{j\omega})\mathbf{H}(e^{j\omega})$ is a $q \times q$ matrix product of a $q \times p$ times a $p \times q$ matrix. If $q > p$, then the rank must be deficient. Therefore, we must have $p \geq q$. (There must be at least as many outputs as there are inputs, but it's ok to have extra outputs.)

– A lossless $p \times q$ transfer function matrix $\mathbf{H}(z)$ is paraunitary, i.e.,

$$\tilde{\mathbf{H}}(z)\mathbf{H}(z) = \mathbf{I}_q$$

Thus, every paraunitary matrix transfer function is *unitary* on the unit circle for all ω. Away from the unit circle, paraunitary $\mathbf{H}(z)$ is the unique analytic continuation of unitary $\mathbf{H}(e^{j\omega})$.

Lossless Filter Examples

- The simplest lossless filter is a unit-modulus gain

$$H(z) = e^{j\phi}$$

where ϕ can be any phase value. In the real case ϕ can only be 0 or π, hence $H(z) = \pm 1$.

- A lossless FIR filter can only consist of a single nonzero tap:

$$H(z) = e^{j\phi} z^{-K}$$

for some fixed integer K, where ϕ is again some constant phase, constrained to be 0 or π in the real-filter case. We consider only causal filters here, so $K \geq 0$.

- Every finite-order, single-input, single-output (SISO), lossless IIR filter (recursive allpass filter) can be written as

$$H(z) = e^{j\phi} z^{-K} \frac{z^{-N}\tilde{A}(z)}{A(z)}$$

where $K \geq 0$,
$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_N z^{-N}$, and
$\tilde{A}(z) \stackrel{\Delta}{=} \overline{A}(z^{-1})$. The polynomial $\tilde{A}(z)$ can be obtained by reversing the order of the coefficients in $A(z)$, conjugating them, and multiplying by z^N. (The factor z^{-N} above serves to restore negative powers of z and hence causality.) Such filters are generally called *allpass filters*.

- The normalized DFT matrix is an $N \times N$ order zero paraunitary transformation. This is because the normalized DFT matrix,
$\mathbf{W} = [W_N^{nk}]/\sqrt{N}$, $n, k = 0, \ldots, N-1$, where
$W_N \stackrel{\Delta}{=} e^{-j2\pi/N}$, is a *unitary* matrix:

$$\frac{\mathbf{W}^*}{\sqrt{N}} \frac{\mathbf{W}}{\sqrt{N}} = \mathbf{I}_N$$

Properties of Paraunitary Systems

Paraunitary systems are essentially multi-input, multi-output (MIMO) allpass filters. Let $\mathbf{H}(z)$ denote the $p \times q$ matrix transfer function of a paraunitary system. Some of its properties include

- In the square case ($p = q$), the matrix determinant, $\det[\mathbf{H}(z)]$, is an *allpass filter*.
- Therefore, if a square $\mathbf{H}(z)$ contains FIR elements, its determinant is a simple delay: $\det[\mathbf{H}(z)] = z^{-K}$ for some integer K.

Properties of Paraunitary Filter Banks

An N-channel analysis filter bank can be viewed as an $N \times 1$ MIMO filter

$$\mathbf{H}(z) = \begin{bmatrix} H_1(z) \\ H_2(z) \\ \vdots \\ H_N(z) \end{bmatrix}$$

A *paraunitary filter bank* must therefore obey

$$\tilde{\mathbf{H}}(z)\mathbf{H}(z) = 1$$

More generally, we allow paraunitary filter banks to scale and/or delay the input signal:

$$\tilde{\mathbf{H}}(z)\mathbf{H}(z) = c_K z^{-K}$$

where K is some nonnegative integer and $c_K \neq 0$.

We can note the following properties of paraunitary filter banks:

- The synthesis filter bank is simply the paraconjugate of the analysis filter bank:

  $$\mathbf{F}(z) = \tilde{\mathbf{H}}(z)$$

  That is, since the paraconjugate is the inverse of a paraunitary filter matrix, it is exactly what we need for perfect reconstruction.

- The channel filters $H_k(z)$ are *power complementary*:

  $$\left|H_1(e^{j\omega})\right|^2 + \left|H_2(e^{j\omega})\right|^2 + \cdots + \left|H_N(e^{j\omega})\right|^2 = 1$$

  This follows immediately from looking at the paraunitary property on the unit circle.

- When $\mathbf{H}(z)$ is FIR, the corresponding synthesis filter matrix $\tilde{\mathbf{H}}(z)$ is also FIR.

- When $\mathbf{H}(z)$ is FIR, each synthesis filter, $F_k(z) = \tilde{\mathbf{H}}_k(z)$, $k = 1, \ldots, N$, is simply the $\textsc{Flip}$ of its corresponding analysis filter $H_k(z) = \mathbf{H}_k(z)$:

$$f_k(n) = h_k(L - n)$$

where L is the filter length. (When the filter coefficients are complex, $\textsc{Flip}$ includes a complex conjugation as well.)

This follows from the fact that paraconjugating an FIR filter amounts to simply flipping (and conjugating) its coefficients.

Note that only trivial FIR filters can be paraunitary in the single-input, single-output (SISO) case. In the MIMO case, on the other hand, paraunitary systems can be composed of FIR filters of any order.

- FIR analysis and synthesis filters in paraunitary filter banks have the *same amplitude response*.

  This follows from the fact that $\textsc{Flip}(h) \leftrightarrow \overline{H}$, i.e., flipping an FIR filter impulse response $h(n)$ conjugates the frequency response, which does not affect its amplitude response $|H(e^{j\omega})|$.

- The polyphase matrix $\mathbf{E}(z)$ for any FIR paraunitary perfect reconstruction filter bank can be written as the product of a paraunitary and a *unimodular* matrix.

- A *unimodular* polynomial matrix $\mathbf{U}(z)$ is any square polynomial matrix having a *constant nonzero* *determinant*. For example,

$$\mathbf{U}(z) = \begin{bmatrix} 1+z^3 & z^2 \\ z & 1 \end{bmatrix}$$

is unimodular. See Vaidyanathan for further details (p. 663).

Paraunitary Examples

Consider the Haar filter bank discussed previously, for which

$$\mathbf{H}(z) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1+z^{-1} \\ 1-z^{-1} \end{bmatrix}$$

The paraconjugate of $\mathbf{H}(z)$ is

$$\tilde{\mathbf{H}}(z) = \frac{1}{\sqrt{2}} \begin{bmatrix} 1+z & 1-z \end{bmatrix}$$

so that

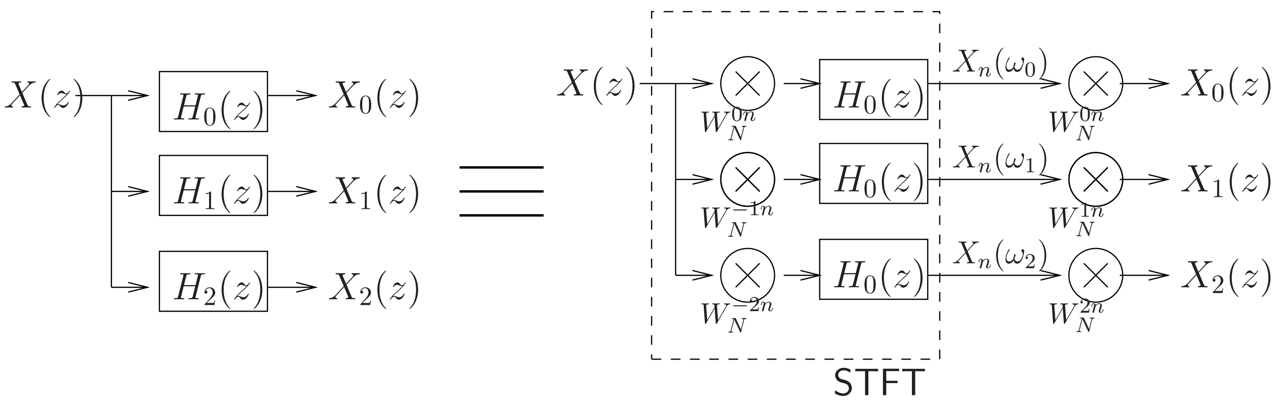
$$\tilde{\mathbf{H}}(z)\mathbf{H}(z) = \begin{bmatrix} 1+z & 1-z \end{bmatrix} \begin{bmatrix} 1+z^{-1} \\ 1-z^{-1} \end{bmatrix} = 1$$

Thus, the Haar filter bank is paraunitary. This is true for any power-complementary filter bank, since when $\tilde{\mathbf{H}}(z)$ is $N \times 1$, power-complementary and paraunitary are the same property.

For more about paraunitary filter banks, see Chapter 6 of Vaidyanathan.

Modulated STFT Filter Bank

- Recall the complex Portnoff analysis bank, where $H_k(z)$, $k = 1 \ldots N$ are Nth-band bandpass filters related to lowpass prototype $H_0(z)$ by modulation (e.g., $H_k(z) = H_0(zW_N^k),\ W_N \stackrel{\Delta}{=} e^{-j\frac{2\pi}{N}}$):



- Convolution gives

$$X_n(\omega_k) = \sum_{m=-\infty}^{\infty} x[m]h_0[n-m]W_N^{-km}.$$

This is the sliding-window STFT implementation, where $h_0[n-m]$ is the sliding window "centered" at time n, and $X_n(\omega_k)$ is the kth DTFT bin at time n.

- After remodulating the DTFT channel outputs and summing, we obtain perfect reconstruction of $X(n)$ provided $H_0(z)$ is Nyquist(N).

- For $H_0(z)$ to be a good anti-aliasing lowpass filter, its length must exceed the number of bins in the DTFT. (Otherwise, the best we have is the rectangular window, which gives only -13 dB stopband rejection.) This means we must use a Portnoff window of some length larger than the DFT length.
- Let the window length be $L > N$. In Lecture 7, it is mentioned that we can still use a length N FFT provided h_0 is replaced by $\text{ALIAS}_N h_0$. I.e., it is *time-aliased* down to length N.

With polyphase analysis we obtain this result, along with an efficient FFT implementation:

Polyphase Analysis of Portnoff STFT

- Recall that the analysis filter in an STFT can be expressed as a frequency-shift of a prototype lowpass filter:

  $$H_k(z) = H_0(zW_N^k),\ k = 1, \ldots, N-1$$

  In principle, the lowpass filter's impulse-response can be any length L.
- Denote the N-channel polyphase components of $H_0(z)$ by $E_l(z)$, $l = 0, 1, \ldots, N-1$.

- Then by the polyphase decomposition,

$$
\begin{aligned}
H_0(z) &= \sum_{l=0}^{N-1} z^{-l} E_l(z^N) \\
H_k(z) &= \sum_{l=0}^{N-1} (zW_N^k)^{-l} E_l((zW_N^k)^N) \\
&= \sum_{l=0}^{N-1} z^{-l} E_l(z^N) W_N^{-kl}
\end{aligned}
$$

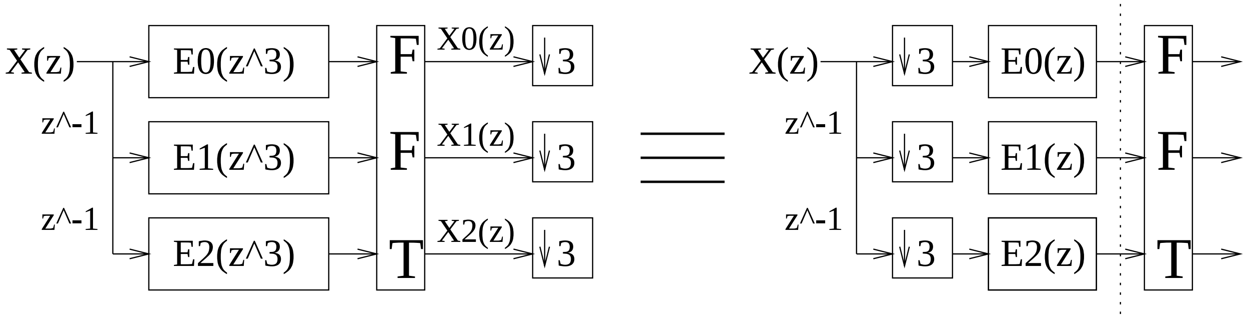
- Consequently,

$$
H_k(z)X(z) = \sum_{l=0}^{N-1} z^{-l} E_l(z^N) X(z) W_N^{-kl}
$$

$$
\begin{bmatrix} H_0(z) \\ \dots \\ H_{N-1}(z) \end{bmatrix} = \begin{bmatrix} \\ W_N^{-kl} \\ \\ \end{bmatrix} \begin{bmatrix} E_0(z^N) z^{-0} X(z) \\ \dots \\ E_{N-1}(z^N) z^{-(N-1)} X(z) \end{bmatrix}
$$

- If $H_0(z)$ is a good Nth-band lowpass, the subband signals $x_k[n]$ are bandlimited to a region of width $2\pi/N$. As a result, there is negligible aliasing when we downsample each of the subbands by N.

- Commuting the downsamplers to get an efficient implementation is straightforward:



- We see that the polyphase filters compute the appropriate time-aliases of the *flipped* window $H_0(z^{-1})$.
- The window is hopped by N samples, but recall that it is operating on input date time aliased by the factor L/N, where L is the filter length. Thus, the hop size is only a fraction of the anti-aliasing filter impulse-response length.

Question: What familiar case do we get when $E_k(z) = 1$ for all k?

Critically Sampled STFT Filter Bank

We will now analyze the filter bank interpretation of the STFT with hop size set to $R = M = N$.

The kth subband signal in the DFT filter bank can be written as:

$$X_k(m) = \sum_{n=-\infty}^{\infty} h(mN - n)x(n)W_N^{kn}$$

with $W_N \equiv e^{-j2\pi/N}$. The signal $X_k(m)$ is regarded as a complex sequence formed from the kth DFT bin over time m (in frames).

Making the change of variable $n = lN - r$, the above equation becomes:

$$X_k(m) = \sum_{r=0}^{N-1} u_r(m)W_N^{kn} \tag{11}$$

with

$$u_r(m) \equiv \sum_{l=-\infty}^{\infty} h(mN - lN + r)x(lN - r) \tag{12}$$

Therefore, (11) can be interpreted as computing a length-N FFT applied to the input block $[u_0(m) \cdots u_{N-1}(m)]^T$.

Now, we'll form a simpler representation of the sequence $u_r(m)$. First, define the polyphase decomposition of $h(n)$ and $x(m)$:

$$p_i(m) = h(mN + i), \quad i = 0, 1, ..., N - 1$$

If the filter $h(n)$ has length $M = LN$, then each of its polyphase components $p_r(m)$ has length L. Now define the sequence $x_r(m)$

$$x_r(m) \equiv x(mN - r), \quad r = 0, 1, ..., N - 1$$

Finally, (12) can be expressed as:

$$u_r(m) = \sum_{l=-\infty}^{\infty} p_r(m - l)x_r(m) \tag{13}$$

Thus, every input bin to the DFT is actually a convolution between $x_r(m)$ and $p_i(m)$, the ith polyphase filter of the lowpass prototype filter, $h(n)$.

- Complexity is now the cost of one full-rate convolution with $h(n)$ and the FFT cost.
- If $M \gg N$ ($L \gg 1$), we can keep aliasing error within tolerable limits.
- If $M = 10N$ ($L = 10$), it is possible to keep the reconstruction error below 0.1%.
- Similar development for the synthesis DFT bank
- Notice that when the polyphase filters are scalars (1-tap FIR filters) of unit gain, then this is simply a DFT block transform, with a rectangular window. The frequency response is a sinc, with poor frequency characteristics.
- By extending the length of the polyphase filters, $L > 1$, then the freqeuency characteristics of the window can become much better.
- The window length is no longer restricted to be of the same length as the transform.

Relationships bewteen OLA,LOT, & ELT

- Let N be the length of the DFT (or the number of frequency bins). N is no longer constrained to be the same as R.
- R is the downsampling factor on each branch
- L is the length of each polyphase filter, $p_r(m)$.
- The length of the lowpass prototype filter $h(n) = M = LN$
- Case 1 : critically sampled, no overlap
  - $R = N$, the filter bank is critically sampled
  - $L = 1, M = N$, the polyphase filters are simply scalar multiplies
  - Equivalent to a block transform
  - Perfect reconstruction only if $h(n) = 1, \quad n = 0, ...N - 1$
- Case 2 : oversampled OLA of 50% overlap
  - $R = \frac{N}{2}$, the filter bank output is oversampled by 2
  - $L = 1, M = N$, the polyphase filters are simply scalar multiplies

- PR requires that the prototype lowpass filter (window) has constant overlap add in time:

$$\sum_{i=-\infty}^{\infty} h(n - iR) = 1$$

- Case 3 : critically sampled OLA of 50% overlap
  - $R = N$, the filter bank is critically sampled
  - $L = 2$, the polyphase filters are two-tap FIR filters
  - $M = 2N$, the lowpass protoype filter $h(n)$ is twice as long as the transform length, N
  - A transform slightly different than the DFT matrix is needed for perfect reconstruction
  - The same form as the Princen-Bradley filter bank, where $h(n) = h(M - 1 - n)$ and $h^2(n + \frac{M}{2}) + h^2(n) = 2$
  - Considered a Lapped Transform
- Case 4 : critically sampled OLA of 8:1 overlap
  - Similar to Case 3, but $L = 8$ and thus $M = 8N$.
  - Transform matrix is close to that of Case 3.
  - The same form as the MPEG layer I,II filter bank
  - Considered an Extended Lapped Transform
- How do we generate the new transform matrices in Cases 3 & 4?

Pseudo-QMF Cosine Modulation Filter Bank

- Now, we want an N-channel transform with real filters and real outputs.
- *Only eliminate aliasing between adjacent bands.* Bandpass filters used in practice attenuate noise about $96dB$, so neglected bands are not much of a concern.
- First, design a lowpass prototype window, $h(n)$, with length $M = LN, \quad L, M, N \in \mathbb{Z}$.
- The lowpass design is constrained to give aliasing cancellation in neighboring subbands:

$$\begin{aligned} |H(e^{j\omega})|^2 + |H(e^{j(\pi/N)-\omega})|^2 &= 2, \quad 0 < |\omega| < \pi/2N \\ |H(e^{j\omega})|^2 &= 0, \quad |w| > \pi/N \end{aligned}$$

- The filter bank analysis filters $h_k(n)$ are cosine modulations of $h(n)$:

$$h_k(n) = h(n)\cos\left[\left(k + \frac{1}{2}\right)\left(n - \frac{M-1}{2}\right)\frac{\pi}{N} + \phi_k\right],$$

$k = 0, \ldots, N-1$, where the phases are restricted according to

$$\phi_{k+1} - \phi_k = (2r+1)\frac{\pi}{2}$$

again for aliasing cancellation.

- Since it is an orthogonal filter bank, the synthesis filters are simply the time reverse of the analysis filters:
$$f_k(n) = h_k(M - 1 - n)$$
- This filter bank is used in MPEG audio, layers I,II.
  - N=32 bands
  - M=512 taps (L=8)

Perfect Reconstruction Cosine Modulated Filter Banks

- By changing the phases ϕ_k, the pseudo-QMF filter bank can yield perfect reconstruction:

$$\phi_k = \left(k + \frac{1}{2}\right)(L+1)\frac{\pi}{2}$$

  where L is the length of the polyphase filter ($M = LN$).

- If $M = 2N$, then this is the *oddly-stacked Princen-Bradley filter bank*, and the analysis filters are related by cosine modulations of the lowpass prototype:

$$f_k(n) = h(n)\cos\left[\left(n + \frac{N+1}{2}\right)\left(k + \frac{1}{2}\right)\frac{\pi}{N}\right],\ k = 0 : N-1$$

- However, the length of the filters M can be any *even* multiple of N:

$$M = LN, \quad (L/2) \in \mathcal{Z}$$

- L is called the $overlapping\ factor$

- These filter banks are also referred to as *Extended Lapped Transforms*, when $K \geq 2$.

MPEG Layer III Filter Bank

- Original MPEG subbands were about 687 Hz wide.
- Finer frequency resolution was needed.
- A Princen-Bradley filter bank with 12 to 36 sub-bands is appended after each subband of the 32-channel PQMF cosine modulated analysis filter bank.
- The number of sub-bands and window shape are signal dependent:
  - Transients use 12 subbands, which gives better time resolution and poorer frequency resolution.
  - Steady-state tones use 36 subbands, which gives better frequency resolution and poorer time resolution.
  - The encoder generates a function called the *perceptual entropy* (PE), which tells the coder when to switch resolutions.

Geometric Signal Theory

We will now approach filter bank derivation from a "Hilbert space" point of view. This is the most natural setting for the study of *wavelet filter banks*.

- Signals can be expanded as a linear combination of *orthonormal basis signals* φ_k:

$$x(n) = \sum_{k=-\infty}^{\infty} \langle \varphi_k, x \rangle \, \varphi_k(n)$$

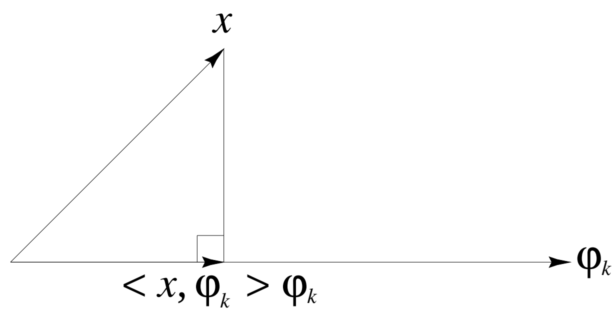
$$n \in (-\infty, \infty), \quad x(n), \varphi_k(n) \in \mathcal{C}$$

  where the coefficient of projection of x onto φ_k is given by

$$\langle \varphi_k, x \rangle \stackrel{\Delta}{=} \sum_{n=-\infty}^{\infty} \varphi_k^*(n) x(n) \qquad \text{(Inner Product)}$$

$$\langle \varphi_k, \varphi_l \rangle = \delta(k-l) = \begin{cases} 1, & k = l \\ 0, & k \neq l \end{cases} \quad \text{(Orthonormal)}$$

- Signal expansion can be viewed as sum of geometric projections of x onto $\{\varphi_k\}$



Example Basis Signals

- "Natural" Basis:

$$\varphi_k \stackrel{\Delta}{=} [\ldots, 0, \underbrace{1}_{k^{\text{th}}}, 0, \ldots], \quad \text{i.e.,}$$

$$\varphi_k(n) \stackrel{\Delta}{=} \delta(n-k)$$

$$\Longrightarrow \qquad \langle \varphi_k, x \rangle = x(k)$$

$$x(n) = \sum_{k=-\infty}^{\infty} x(k)\varphi_k(n)$$

$$= \sum_{k=-\infty}^{\infty} x(k)\delta(n-k) \stackrel{\Delta}{=} (x * \delta)(n)$$

- Normalized DFT Basis for $\mathbb{C}^N$:

$$\varphi_k(n) \stackrel{\Delta}{=} e^{j\omega_k n}/\sqrt{N}, \quad \omega_k \stackrel{\Delta}{=} 2\pi k/N, \quad n, k \in [0, N-1]$$

$$
\begin{aligned}
\Longrightarrow \qquad \langle \varphi_k, x \rangle &\isdef \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x(n) e^{-j\omega_k n} \\
&\isdef \mathsf{DFT}_{N,k}(x)/\sqrt{N} \isdef X(\omega_k)/\sqrt{N}
\end{aligned}
$$

$$
\begin{aligned}
x(n) &= \sum_{k=0}^{N-1} \langle \varphi_k, x \rangle \, \varphi_k(n) \\
&= \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j\omega_k n} \\
&\isdef \mathsf{DFT}_{N,n}^{-1}(X)
\end{aligned}
$$

- Normalized Fourier Transform Basis:

$$\varphi_\omega(t) \stackrel{\Delta}{=} e^{j\omega t}/\sqrt{2\pi}, \quad \omega, t \in (-\infty, \infty)$$

$$\begin{aligned}
\Longrightarrow \qquad \langle \varphi_\omega, x \rangle &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \\
&\stackrel{\Delta}{=} \mathsf{FT}_\omega(x)/\sqrt{2\pi} \stackrel{\Delta}{=} X(\omega)/\sqrt{2\pi}
\end{aligned}$$

$$\begin{aligned}
x(t) &= \int_{-\infty}^{\infty} \langle \varphi_\omega, x \rangle \, \varphi_\omega(t) d\omega \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{j\omega t} d\omega \\
&\stackrel{\Delta}{=} \mathsf{FT}_t^{-1}(X)
\end{aligned}$$

- Normalized DTFT Basis:

$$\varphi_\omega(n) \stackrel{\Delta}{=} e^{j\omega n}/\sqrt{2\pi}, \quad \omega \in (-\pi, \pi], \quad n \in (-\infty, \infty)$$

(Find inner product $\langle \varphi_\omega, x \rangle$ and reconstruction of $x(n)$ in terms of $\{\varphi_\omega\}$ as an exercise.)

- Normalized STFT Basis:

$$\varphi_{mk}(n) \triangleq \frac{w(n-mR)e^{j\omega_k n}}{\|\,\textsc{Shift}_{mR}(w)e^{j\omega_k(\cdot)}\,\|} = \frac{w(n-mR)e^{j\omega_k n}}{\sqrt{\sum_n w^2(n)}},$$

$\omega_k = 2\pi k/N,\ k \in [0, N-1],\ n \in (-\infty, \infty),\ w(n) \in \mathcal{R}$

  - *Overcomplete* in general
  - *Orthonormal* when

    $R = M = N$

    (Hop size = Window length = DFT length)

    $w =$ Rectangular Window w_R

    $\Longrightarrow\ \varphi_{mk} = \textsc{Shift}_{mN}\left[\textsc{ZeroPad}_{\infty}\left(\varphi_k^{\text{DFT}}\right)\right]$,
    i.e., $\varphi_{mk}(n) = e^{j\omega_k n}/\sqrt{N}$ for
    $mN \le n \le (m+1)N - 1$, and 0 otherwise

  In this case,

$$\begin{aligned}\langle \varphi_{mk}, x\rangle &= \frac{1}{\sqrt{N}} \sum_{n=-\infty}^{\infty} x(n) w_R(n-mN) e^{-j\omega_k n} \\ &\triangleq \text{STFT}_{N,m,k}(x)/\sqrt{N} \triangleq X_m(\omega_k)/\sqrt{N}\end{aligned}$$

$$
\begin{aligned}
x(n) &= \sum_{m=-\infty}^{\infty} \sum_{k=0}^{N-1} \langle \varphi_{mk}, x \rangle \, \varphi_{mk}(n) \\
&= \sum_{m=-\infty}^{\infty} w_R(n - mN) \frac{1}{N} \sum_{k=0}^{N-1} X_m(\omega_k) e^{j\omega_k n} \\
&= \sum_{m=-\infty}^{\infty} \textsc{Shift}_{mN,n} \left\{ \textsc{ZeroPad}_{\infty} \left[\mathsf{DFT}_N^{-1}(X_m) \right] \right\} \\
&\stackrel{\Delta}{=} \mathsf{STFT}_{N,n}^{-1}(X)
\end{aligned}
$$

- Continuous Wavelet Transform Basis:

$$
\varphi_{s\tau}(t) \stackrel{\Delta}{=} \frac{1}{\sqrt{|s|}} f^* \left(\frac{\tau - t}{s} \right),
$$

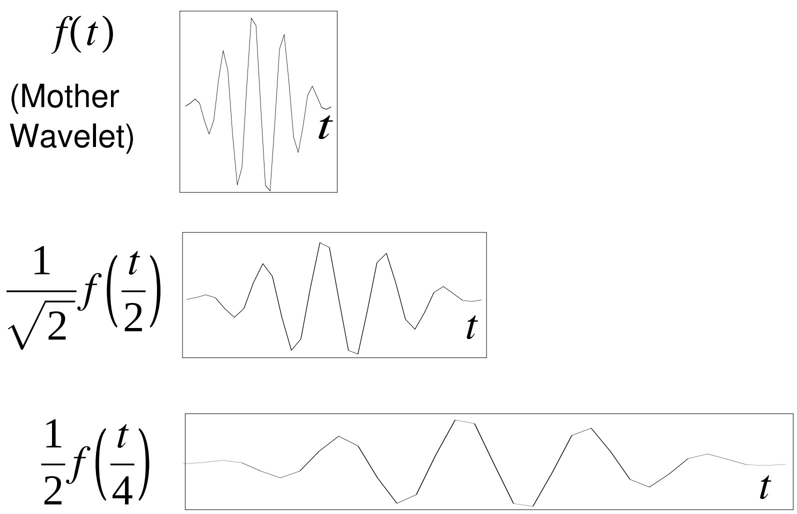
$$
\tau, s, t \in (-\infty, \infty)
$$

$$
X(s, \tau) \stackrel{\Delta}{=} \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} x(t) f \left(\frac{t - \tau}{s} \right) dt
$$

  - $s =$ *scale* parameter
  - $1/\sqrt{|s|}$ maintains energy invariance
  - $X(s, \tau) =$ *wavelet coefficients*
  - $f(t) =$ *mother wavelet*

– $f(t)$ typically *bandpass*

– Qualitative example:



Wavelets

- Admissibility condition for mother wavelet $\psi(t)$:

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty$$

- Given sufficient decay, this reduces to $\Psi(0) = 0$ (mother wavelet must be zero-mean)
- Morlet Wavelet:

$$\begin{aligned} \psi(t) &\stackrel{\Delta}{=} \frac{1}{\sqrt{2\pi}} e^{-j\omega_0 t} e^{-t^2/2} \\ &\longleftrightarrow \\ \Psi(\omega) &= e^{-(\omega-\omega_0)^2/2} \end{aligned}$$

  - Gaussian-windowed complex sinusoid
  - Scaled so that $\| \psi \| = 1$
  - Center frequency ω_0 typically chosen so that second peak is half of first $\Longrightarrow \ \omega_0 = \pi\sqrt{2/\ln 2} \approx 5.336$
  - $\Psi(0) \approx 7 \times 10^{-7} \approx 0$ (close enough)
- Wavelet-based analysis called a *scalogram* (analogous to STFT "spectrogram")

Discrete Wavelet Transform

Transform (filterbank form):

$$\begin{aligned} X(k,n) &= s^{-k/2} \int_{-\infty}^{\infty} x(t) h\left(nT - a^{-k} t\right) dt \\ &= \int_{-\infty}^{\infty} x(t) h\left(n a^{k} T - t\right) dt \end{aligned}$$

n, k integers

Inverse:

$$x(t) = \sum_{k} \sum_{n} X(k,n) \underbrace{\varphi_{kn}(t)}_{\text{basis}}$$

- Wavelet filter banks always *Constant*-Q:
  - $Q \stackrel{\Delta}{=}$ center-frequency / bandwidth
  - Center-frequency $\stackrel{\Delta}{=}$ geometric mean of bandlimits ω_1 and ω_2:

$$\omega_c \stackrel{\Delta}{=} \sqrt{\omega_1\omega_2}$$

$$\omega_c(k) \stackrel{\Delta}{=} \sqrt{\omega_1(k)\omega_2(k)} = \sqrt{a^k\omega_1(0)a^k\omega_2(0)} = a^k\omega_c(0)$$

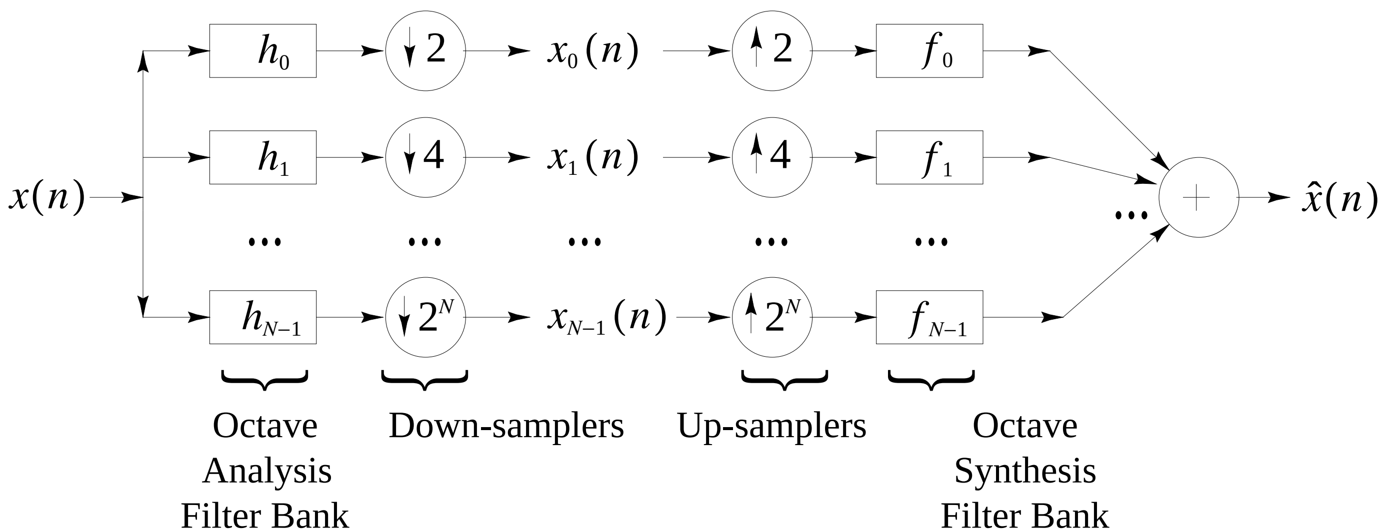
$$Q(k) \stackrel{\Delta}{=} \frac{\omega_c(k)}{\omega_2(k)-\omega_1(k)} = \frac{a^k\omega_c(0)}{a^k\omega_2(0)-a^k\omega_1(0)} = Q(0)$$

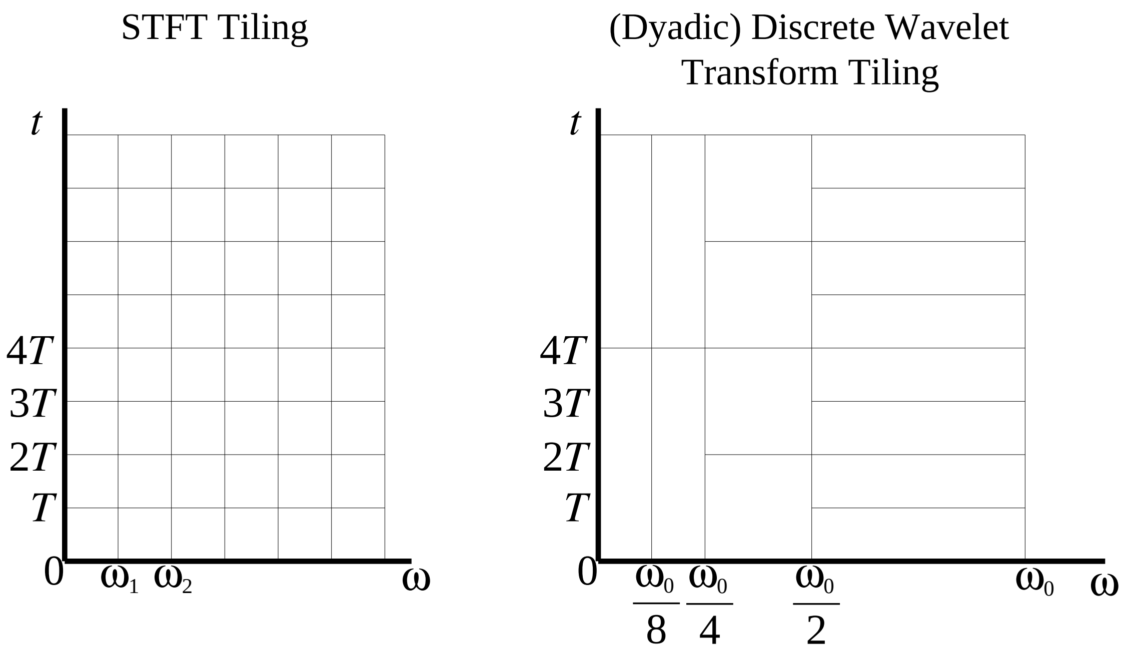
- In dyadic filter bank, $Q = \sqrt{2}$:

$$\text{center-frequency} \stackrel{\Delta}{=} \sqrt{\omega_0(\omega_0 + \text{bandwidth})} = \sqrt{2}\omega_0$$

$$\Longrightarrow \; Q = \frac{\sqrt{2}\omega_0}{2\omega_0-\omega_0} = \sqrt{2}$$

Discrete Wavelet (Dyadic) Filterbank

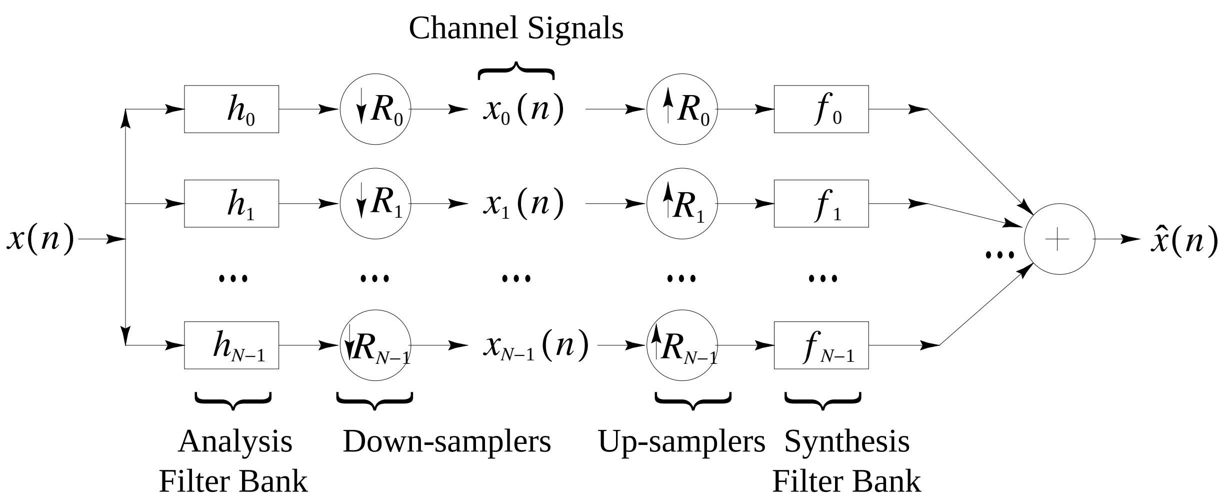



STFT Tiling
t
4T
3T
2T
T
0
ω1
ω2
ω
(Dyadic) Discrete Wavelet
Transform Tiling
t
4T
3T
2T
T
0
ω0/8
ω0/4
ω0/2
ω0
ω

Generalized STFT

$$x_k(n) \;=\; (x * h_k)(nR_k) = \sum_{m=-\infty}^{\infty} x(m)\, \underbrace{h_k(nR_k - m)}_{\text{analysis filter}}$$

$$x(n) \;=\; \sum_k (x_k * f_k)(n) = \sum_{k=0}^{N-1} \sum_{m=-\infty}^{\infty} x_k(m)\, \underbrace{f_k(n - mR_k)}_{\text{synthesis filter}}$$



- Analysis filter h_k typically complex bandpass
- R_k downsamples filter output:
  Critical sampling: $R_k = \pi/\text{Width}(H_k)$
- Impulse response of synth. filter $f_k = k$th *basis signal*
- If $\{f_k\}$ are orthonormal, $f_k(n) = h_k^*(-n)$
- More generally, $\{h_k, f_k\}$ form a *biorthogonal basis*.

Biorthogonal Signal Expansions

A set of signals $\{h_k, f_k\}_{k=1}^{N}$ is said to be a *biorthogonal basis set* if any signal x can be represented as

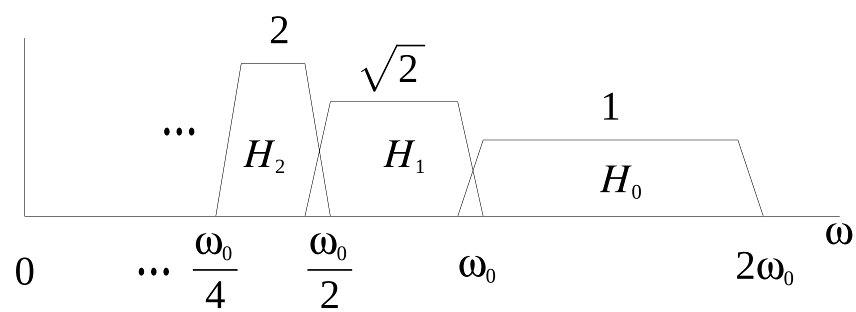
$$x = \sum_{k=1}^{N} \alpha_k \langle x, h_k \rangle f_k$$

where α_k is some normalizing scalar dependent only on h_k and/or f_k. Thus, in a biorthogonal system, we project onto the signals h_k and resynthesize in terms of f_k.

Discrete Wavelet Filterbank

$$
\begin{aligned}
h_k(t) &= \frac{1}{\sqrt{a^k}} h\left(\frac{t}{a^k}\right), \quad a > 1 \\
\longleftrightarrow\ H_k(\omega) &= \sqrt{a^k} H(a^k \omega)
\end{aligned}
$$

- Wavelet channel-filter $H_k(\omega)$ is a *scaling* of channel-filter $H_0(\omega)$ (scaling in time domain also)
- In STFT, channel filter $H_k(\omega)$ is a *shift* of channel-filter $H_0(\omega)$ (modulation in time domain)
- As k increases, h_k lengthens, H_k narrows
- Dyadic filter bank ($a = 2$):



- $H_0(\omega) =$ top-octave bandpass (BP) filter
- $H_1(w) = \sqrt{2} H_0(2\omega) =$ BP for next octave down
- $H_2(w) = 2H_0(4\omega) =$ octave bandpass below that, etc.

Review of STFT Filterbanks

Let's take a look at some of the STFT processors we've seen before, now viewed as a *polyphase filter bank*.

Since they are all based on the FFT, they are all efficient, but most are oversampled as "filter banks" go. Some oversampling is usually preferred outside of a compression context.

The STFT also computes a uniform filter bank, but it can be used as the basis for a variety of non-uniform filter banks giving frequency resolution closer to that of hearing.

STFT, rectangular window, no overlap

- Perfect reconstruction
- Critically sampled (aliasing cancellation)
- Poor channel isolation (13dB)
- Not robust to filter-bank modifications

STFT, rectangular window, 50% overlap

- Perfect reconstruction
- Oversampled by 2 (aliasing cancellation)
- Poor channel isolation (13dB)
- Not robust to filter-bank modifications, but better

STFT, triangular window, 50% overlap

- Perfect reconstruction
- Oversampled by 2
- Better channel isolation (26dB)
- Moderately robust to filter-bank modifications

STFT, Hamming window, 75% overlap

- Perfect reconstruction
- Oversampled by 4
- Aliasing from sidelobes only
- Good channel isolation (42dB)
- Moderately robust to filter-bank modifications

STFT, Kaiser window, Beta=10, 90 percent overlap

- Approximate perfect reconstruction (sidelobes controlled by β)
- Oversampled by 10
- Excellent channel isolation (80 dB)
- Very robust to filter-bank modifications
- Aliasing from sidelobes only

Sliding FFT, any window, maximum overlap, zero-padded by 5

- Perfect reconstruction
  (always true when hop-size $= 1$)
- Oversampled by $5M$:
  - $M =$ window length [time-domain oversampling factor]
  - 5 $=$ zero-padding factor [frequency-domain oversampling factor]
- Excellent channel isolation (set by window sidelobes)
- Extremely robust to filter-bank modifications
- No aliasing to cancel

References

Books

- P.P.Vaidyanthan, *Multirate Systems and Filter Banks*, Prentice Hall, 1993. (Comprehensive reference on perfect reconstruction filter banks. Little or no treatment of the oversampled case.)
- M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, Prentice Hall, 1995. (Good depth on both perfect reconstruction filter banks and wavelets, and their connections.)
- H.S.Malvar, *Signal Processing with Lapped Transforms*, Artech House, Boston, 1992.
- M. Bosi and R.E. Goldberg, *Introduction to Digital Audio Coding and Standards*, Kluwer Academic Publishers, Boston, 2003.

Papers

- J.P. Princen, A.W.Johnson, and A.B. Bradley, *Subband/Transform Coding Using Filter Bank Desings Based on Time Domain Aliasing Cancellation*, ICASSP 1987, pp 2161-2164.
- K. Brandenburg and G. Stoll, *ISO-MPEG-I Audio: A Generic Standard for Coding of High-Quality Digital Audio*, Journal of the AES, Vol 42, no 10, October 1994.