Statistical Signal Processing (SSP)

Lecture 7a

ch2: Linear Time-Frequency Representations

- short-term Fourier transform
- filter banks
- discrete Wavelet transform

Linear Time-Frequency Representations

- non-stationary processes $\Rightarrow$ no ergodicity $\Rightarrow$ cannot obtain statistical averages as limits of time averages $\Rightarrow$ no time-averaging

- spectral estimation $\rightarrow$ spectral representation

- fundamental spectral representation: Fourier transform

$$Y(f) = \int_{-\infty}^{\infty} y(t)\, e^{-j2\pi ft}\, dt$$

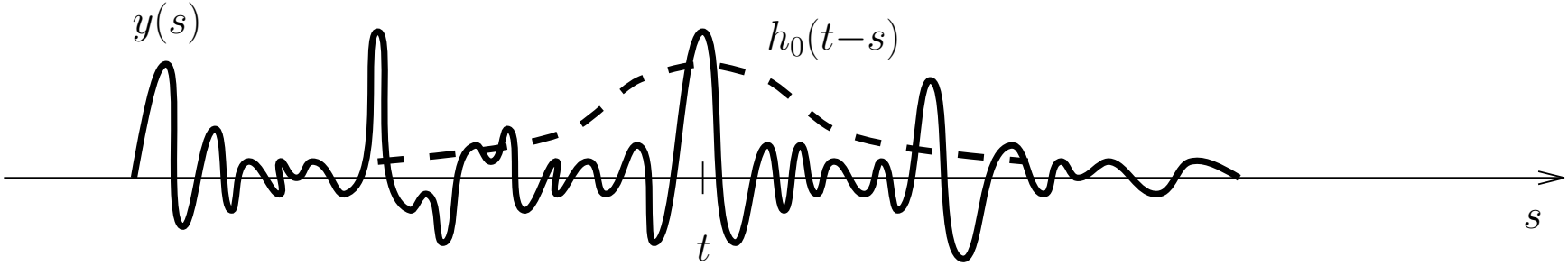
  $y(t)$: we know precisely at what time something happens but we don't know at which frequencies
  $Y(f)$: we know precisely the different spectral components of the signal, but we don't know when they occur

- non-stationary signals: would like a joint time-frequency representation
  e.g.: piano piece: pitch (fundamental frequency) is a piecewise constant function of time (notes being played), $y(t)$ does not tell us which notes are being played, $Y(f)$ shows all the notes but does not tell us when they occur.

Short-Time Fourier Transform (STFT)

- to capture the variation in time of the frequency contents we introduce a sliding window $h_0(t-s)$ centered at t

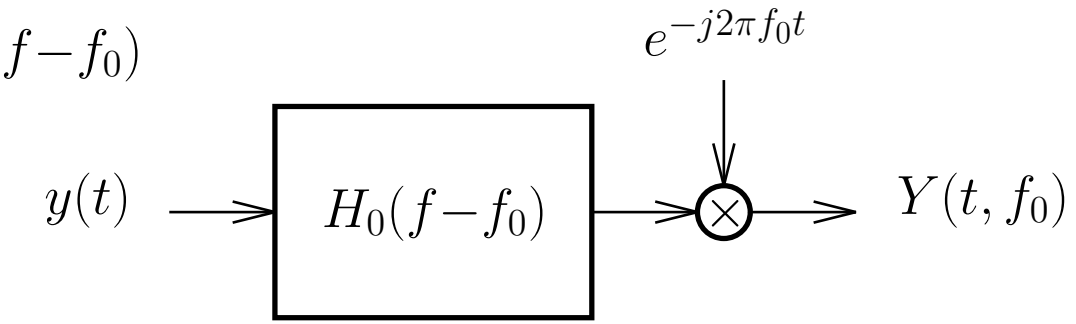
$$Y(t,f) \;=\; \int_{-\infty}^{\infty} y(s)\, h_0(t-s)\, e^{-j2\pi fs}\, ds$$



- filterbank interpretation of the STFT

$$Y(t,f_0) \;=\; e^{-j2\pi f_0 t} \int_{-\infty}^{\infty} y(s)\, h_0(t-s)\, e^{j2\pi f_0(t-s)} ds$$

$h_0(t)\, e^{j2\pi f_0 t}$ = impulse response of $H_0(f-f_0)$



$y(t)$ is bandpass filtered by $H_0(f-f_0)$ ($\Rightarrow$ centered at frequency f_0), and then down modulated by f_0 to give $Y(t,f_0)$ which is centered around dc

Short-Time Fourier Transform (STFT) (2)

- $\begin{cases} h_0(t) \text{ centered around } t = 0: & \int_{-\infty}^{\infty} t\,|h_0(t)|^2 dt \;=\; 0 \\ H_0(f) \text{ centered around } f = 0: & \int_{-\infty}^{\infty} f\,|H_0(f)|^2 df \;=\; 0 \end{cases}$ // unbiased

- $\begin{cases} \text{precise time location} & \Rightarrow\; h_0(t) \;\text{ narrow} \\ \text{precise frequency location} & \Rightarrow\; H_0(f) \;\text{ narrow} \end{cases}$

  time multiplication $\Rightarrow$ frequency-domain convolution $\Rightarrow$ smearing

- $\begin{cases} \text{effective duration:} & D_{eff} = \sqrt{\dfrac{\int_{-\infty}^{\infty} t^2\,|h_0(t)|^2 dt}{\int_{-\infty}^{\infty} |h_0(t)|^2 dt}} \\ \text{effective bandwidth:} & B_{eff} = \sqrt{\dfrac{\int_{-\infty}^{\infty} f^2\,|H_0(f)|^2 df}{\int_{-\infty}^{\infty} |H_0(f)|^2 df}} \end{cases}$ // standard deviation

- *uncertainty principle*: $D_{eff} B_{eff} \;\geq\; \dfrac{1}{4\pi}$

  with = for $h_0(t) = e^{-\frac{1}{2}t^2}$: Gaussian window ($\Rightarrow$ $H_0(f)$ Gaussian)

  STFT with Gaussian $h_0(t)$: "Gabor representation" [1946]

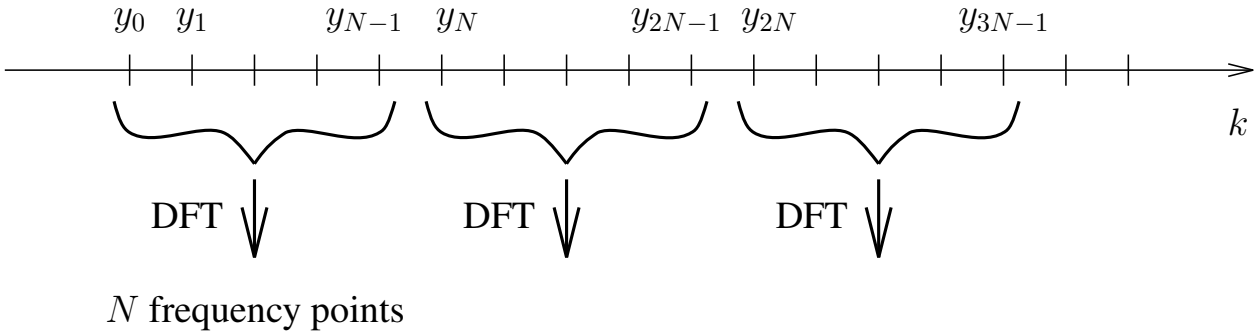
Discrete-Time STFT

- sampling $\Rightarrow$ time axis discretized
  $f \in [-\frac{1}{2}, \frac{1}{2}]$ $\Rightarrow$ frequency more precise, but precision of Δt limited to the sampling period = 1, hence time is less precise

- uncertainty principle: cannot have high precision in both time and frequency
  $\Rightarrow$ $\begin{cases} \text{sample frequency axis} \\ \text{further down-sample time axis} \end{cases}$

- we shall consider *non-redundant* time-frequency representations

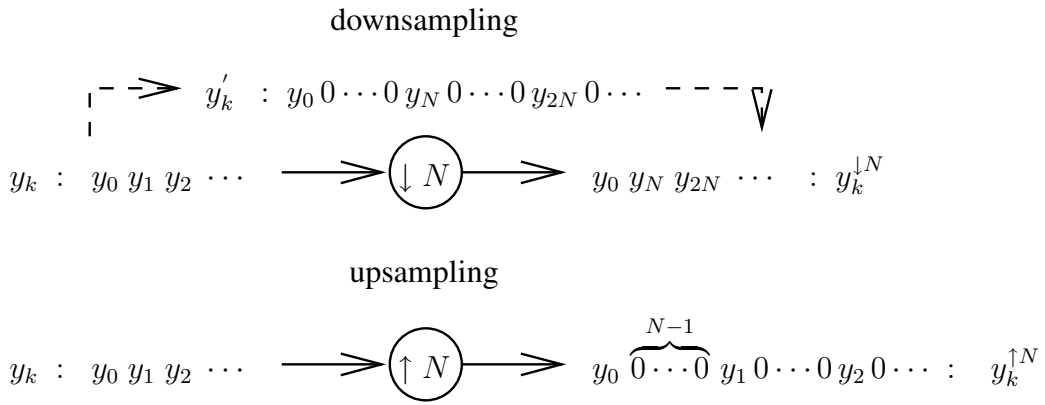
- example: Discrete Fourier Transform (DFT):



  get N frequency points but time axis down-sampled by N
  $\Rightarrow$ conservation of number of degrees of freedom (non-redundant)

Filterbanks

- *downsampling* (decimation) and *upsampling* (interpolation) operations:



- upsampling transformation:

$$\mathbf{Y}^{\uparrow N}(z) = \sum_k y_k^{\uparrow N} z^{-k} = \sum_k y_k z^{-Nk} = \mathbf{Y}(z^N)$$

- downsampling transformation:

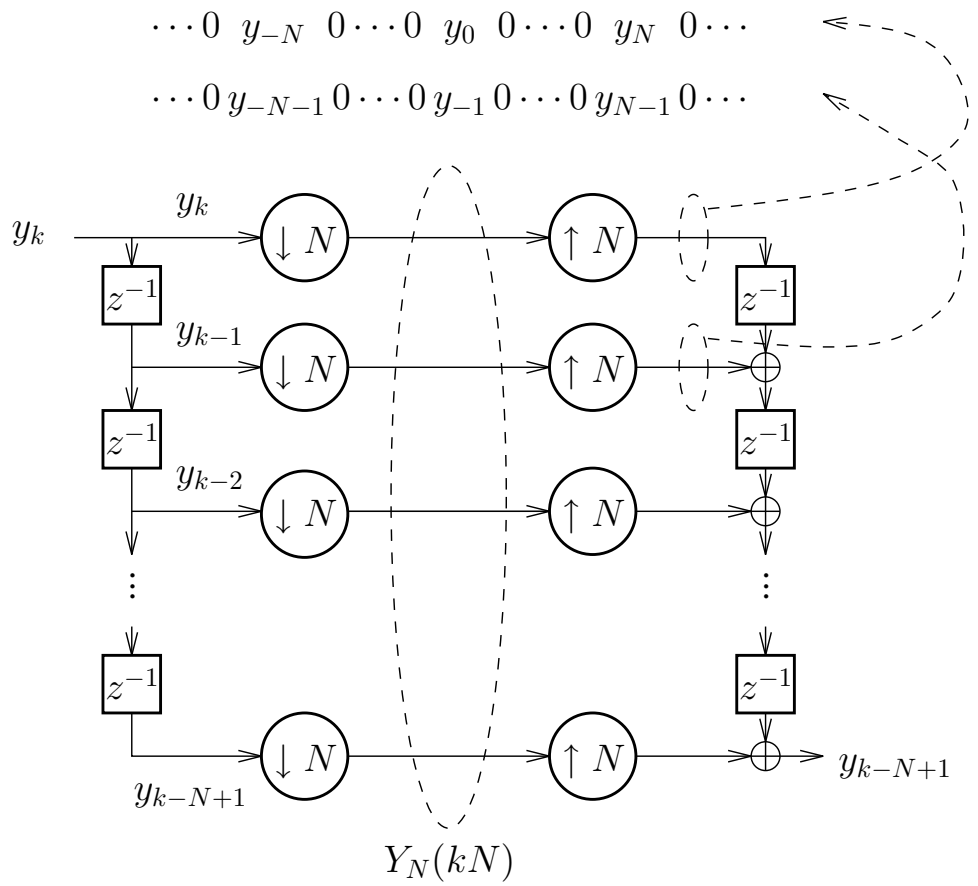
$\diamond\ y_k' = \begin{cases} y_k & , \frac{k}{N} \in \mathcal{Z} \\ 0 & , \text{otherwise} \end{cases} \Rightarrow\ \mathbf{Y}^{\downarrow N}(z) = \sum_k y_{Nk} z^{-k} = \sum_k y_k' z^{-k/N} = \mathbf{Y}'(z^{1/N})$

$\diamond\ y_k' = c_k\, y_k\,,\ c_k = \begin{cases} 1 & , \frac{k}{N} \in \mathcal{Z} \\ 0 & , \text{otherwise} \end{cases}\ ,\ c_k = \frac{1}{N}\sum_{n=0}^{N-1} w_N^{-nk}\ ,\ w_N = e^{-j2\pi/N}$

$\diamond\ \mathbf{Y}'(z) = \frac{1}{N}\sum_{n=0}^{N-1}\sum_k y_k w_N^{-nk} z^{-k} = \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{Y}(z\,w_N^n) \Rightarrow \mathbf{Y}^{\downarrow N}(z) = \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{Y}(z^{1/N} w_N^n)$

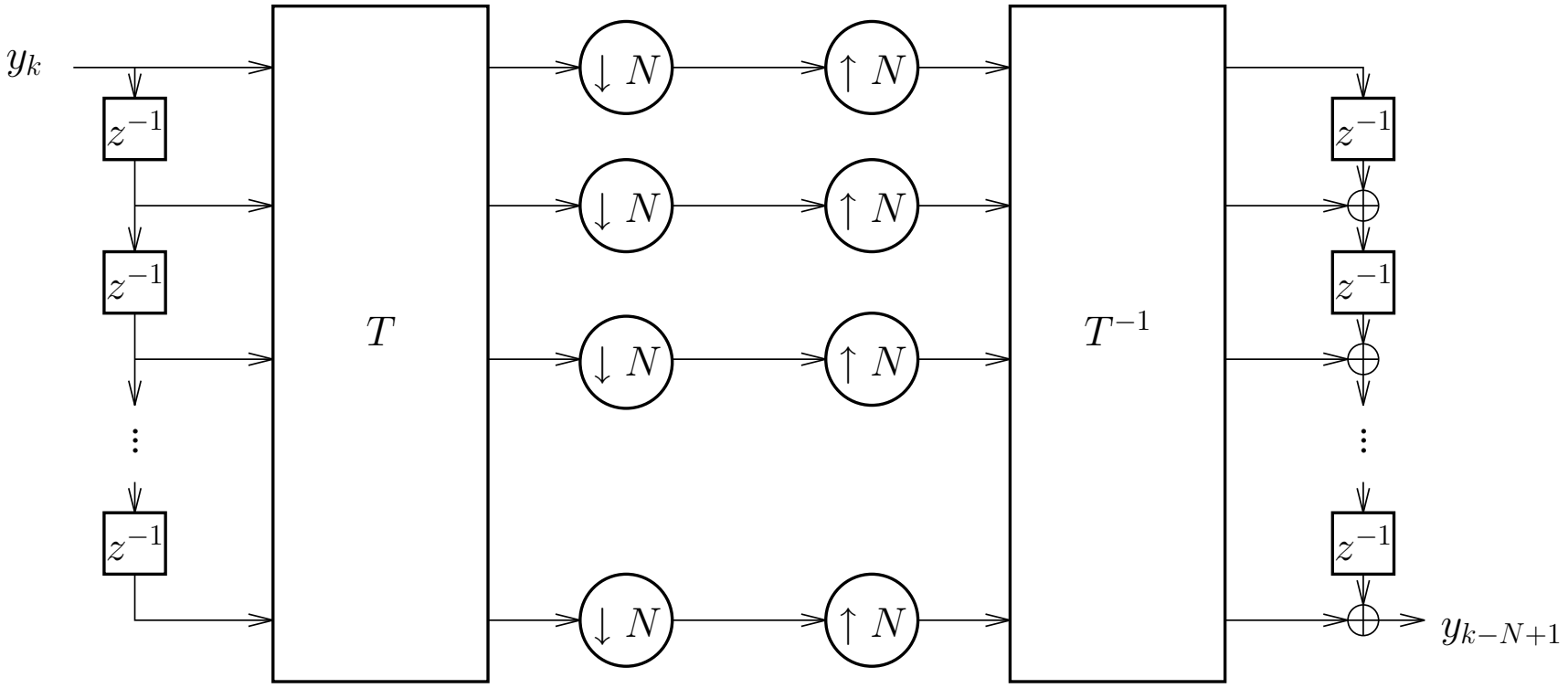
Filterbanks (2)

- *polyphase* representation of the pure delay $z^{-(N-1)}$



Filterbanks (3)

- Can insert $I_N = T^{-1}T$ in the middle. The instantaneous operations multiplication by T or by T^{-1} commute with the down/upsampling, so we get:



Filterbanks (4)

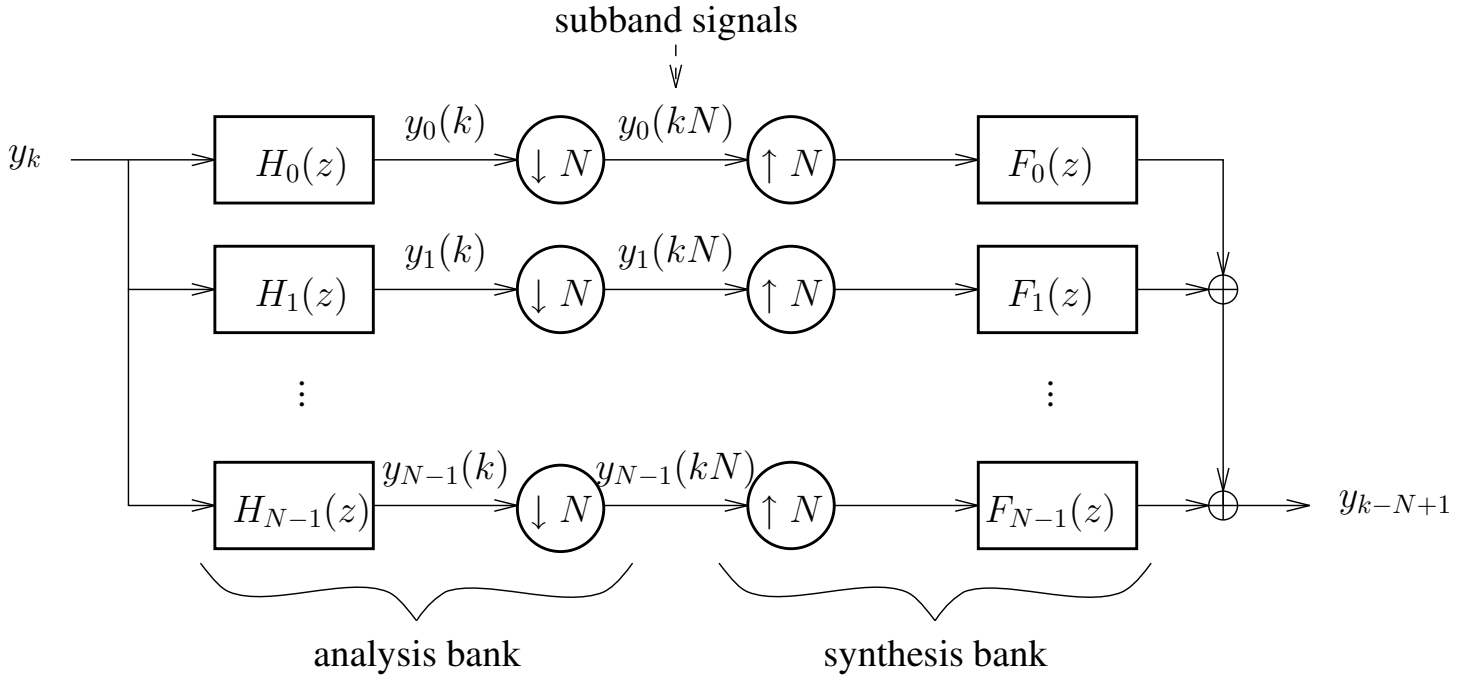
- can interpret the rows of T and the columns of T^{-1} as filters:

$$T = \begin{bmatrix} h_0(0) & h_0(1) & \cdots & h_0(N-1) \\ h_1(0) & h_1(1) & \cdots & h_1(N-1) \\ \vdots & \vdots & & \vdots \\ h_{N-1}(0) & h_{N-1}(1) & \cdots & h_{N-1}(N-1) \end{bmatrix}, \quad T^{-1} = \begin{bmatrix} f_0(N-1) & f_1(N-1) & \cdots & f_{N-1}(N-1) \\ \vdots & \vdots & & \vdots \\ f_0(1) & f_1(1) & \cdots & f_{N-1}(1) \\ f_0(0) & f_1(0) & \cdots & f_{N-1}(0) \end{bmatrix}$$

- can introduce analysis and synthesis filters

$$\begin{cases} \mathrm{H}(z) = \begin{bmatrix} \mathrm{H}_0(z) \\ \mathrm{H}_1(z) \\ \vdots \\ \mathrm{H}_{N-1}(z) \end{bmatrix} = T \begin{bmatrix} 1 \\ z^{-1} \\ \vdots \\ z^{-(N-1)} \end{bmatrix} \\ \\ \mathrm{F}(z) = [\mathrm{F}_0(z) \cdots \mathrm{F}_{N-1}(z)] = \left[z^{-(N-1)} \cdots z^{-1}\ 1\right]\ T^{-1} \end{cases}$$

Filterbanks (5)



filterbank characteristics:

- *critical subsampling*, subsampling factor = number of subbands, non-redundant
- *perfect reconstruction*: can do an exact inverse transform, the original signal can be recovered exactly from the subsampled subband signals
- *losslessness*: $T^{-1} = T^H = T^{*T}$ unitary T
  conservation of energy: $\|TY_N(k)\|^2 = Y_N^T(k)\,\underbrace{T^H T}_{=I}\,Y_N(k) = \|Y_N(k)\|^2$
- $T^{-1} = T^H \;\Rightarrow\; f_n(k) = h_n^*(N-1-k)\,,\; k,n = 0,1,\ldots,N-1$

Example 1: DFT

- DFT matrix W: $W_{km} = w_N^{(k-1)(m-1)}$, $w_N = e^{-j2\pi/N}$ $\quad T = W^*$, $T^{-1} = \frac{1}{N}W$
- let $Y_{N,k}(f)$ = FT of $\{y_{k-N+1}, \ldots, y_k\}$ of finite duration,
  hence $\{Y_{N,k}(n/N),\ n = 0, 1, \ldots, N-1\}$ = its DFT
- subband signals interpretation:

$$y_n(k) = \sum_{i=0}^{N-1} h_n(i) y_{k-i} = \sum_{i=0}^{N-1} w_N^{-ni} y_{k-i} = \sum_{i=0}^{N-1} w_N^{-ni} y_{k-N+1+\underbrace{N-1-i}_{=m}}$$
$$= \sum_{m=0}^{N-1} w_N^{mn} \underbrace{w_N^{-nN}}_{=1} w_N^n\, y_{k-N+1+m} = w_N^n \sum_{m=0}^{N-1} w_N^{mn}\, y_{k-N+1+m} = w_N^n\, Y_{N,k}(n/N)$$

  $\Rightarrow$ filterbank = running spectrum analyzer,
  trade-off: frequency resolution vs. time resolution

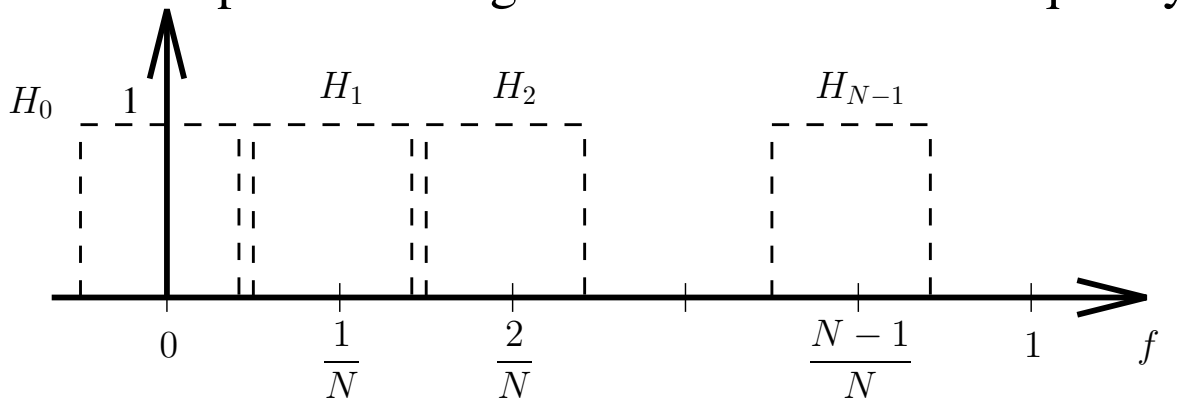
- $\mathrm{H}_0(z) = 1 + z^{-1} + \cdots + z^{-N+1}$ or $|H_0(f)| = \left|\dfrac{\sin N\pi f}{\sin \pi f}\right|$:
  low-pass filter corresponding to a rectangular window $h_0(k)$
- $\mathrm{H}_n(z) = \mathrm{H}_0(z\, w_N^n)$: *modulated filterbank*,
  $H_n(f) = H_0(f - \frac{n}{N})$: frequency translation

More General Filterbanks

- so far: analysis and synthesis filters FIR of length N = number of subbands = subsampling factor
- more general: T and T^{-1} get replaced by $\mathrm{E}(z)$ and $\mathrm{R}(z)$ resp. $(N \times N)$
- perfect reconstruction: $\mathrm{R}(z)\,\mathrm{E}(z) = I_N \quad \Rightarrow \quad \mathrm{R}(z) = \mathrm{E}^{-1}(z)$
- after inserting $\mathrm{R}(z)\,\mathrm{E}(z) = I_N$ into the polyphase decomposition of $z^{-(N-1)}$, $\mathrm{R}(z)$ and $\mathrm{E}(z)$ becomes $\mathrm{R}(z^N)$ and $\mathrm{E}(z^N)$ resp. after moving them across the downsampling and upsampling operations
- losslessness: $\mathrm{R}(z) = \mathrm{E}^{\dagger}(z) = \mathrm{E}^{H}(1/z^*)$: $\mathrm{E}(z)$ paraunitary: $\mathrm{E}^{\dagger}(z)\,\mathrm{E}(z) = I$
- analysis filter bank: $\mathrm{H}(z) = \begin{bmatrix} \mathrm{H}_0(z) \\ \vdots \\ \mathrm{H}_{N-1}(z) \end{bmatrix} = \mathrm{E}(z^N) \begin{bmatrix} 1 \\ \vdots \\ z^{-(N-1)} \end{bmatrix}$
- synthesis filter bank: $\mathrm{F}(z) = \left[\mathrm{F}_0(z) \cdots \mathrm{F}_{N-1}(z)\right] = \left[z^{-(N-1)} \cdots z^{-1} \;\; 1\right] \mathrm{R}(z^N)$
- lossless case: $\mathrm{H}^{\dagger}(z)\mathrm{H}(z) = N \quad \Rightarrow$ sum of the variances in the N subbands is $N\sigma_y^2$ per N sample periods, hence σ_y^2 per sample period $\Rightarrow$ power conservation
- further generalizations: non-uniform filter banks, not critically subsampled, not lossless (biorthogonal), linear phase, IIR, ...

Example 2: Brickwall Filterbank

- dual to the DFT example: rectangular filters in the frequency domain:



- subband signal $y_n(k)$ perfectly bandlimited with bandwidth $\frac{1}{N}$, hence the Nyquist theorem allows for downsampling with a factor N without loss of information, can still perfectly reconstruct

- again modulated filterbank: $H_n(f) = H_0(f - \frac{n}{N})$

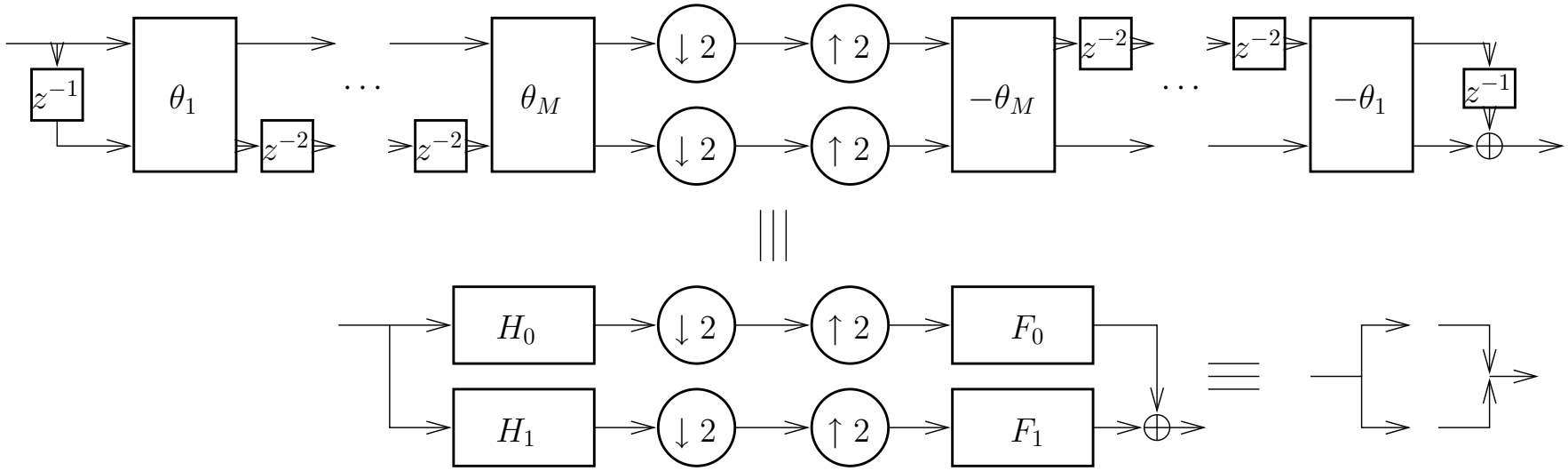
  $h_0(k) = \dfrac{\sin \frac{\pi k}{N}}{\pi k}$ non-causal and of infinite duration

$N = 2$: Quadrature Mirror Filterbanks (QMF)

- general form of lossless FIR two-band filterbank:

$$\mathrm{E}(z) = \Theta(\theta_M) \begin{bmatrix} 1 & 0 \\ 0 & z^{-1} \end{bmatrix} \cdots \Theta(\theta_2) \begin{bmatrix} 1 & 0 \\ 0 & z^{-1} \end{bmatrix} \Theta(\theta_1) \ , \quad \Theta(\theta) = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

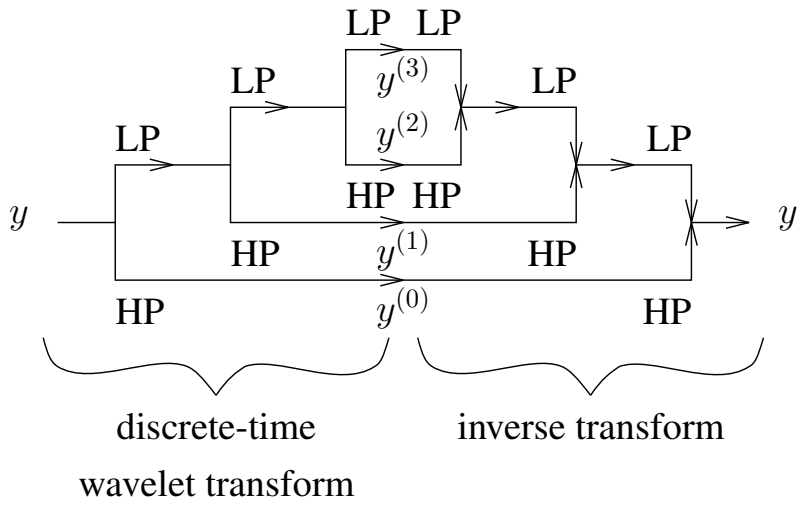
cascade of elementary paraunitary factors

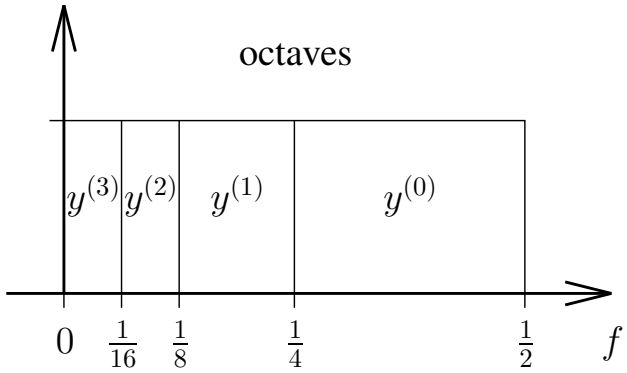


- a delay has been introduced in the synthesis bank to make it causal

Discrete-Time Wavelet Transform (DTWT)

- so far: uniform filterbanks (e.g. discrete STFTF): frequency resolution constant as a function of frequency
- discrete-time wavelet transform: tree-structured two-band filterbank, leads to frequency analysis in octaves

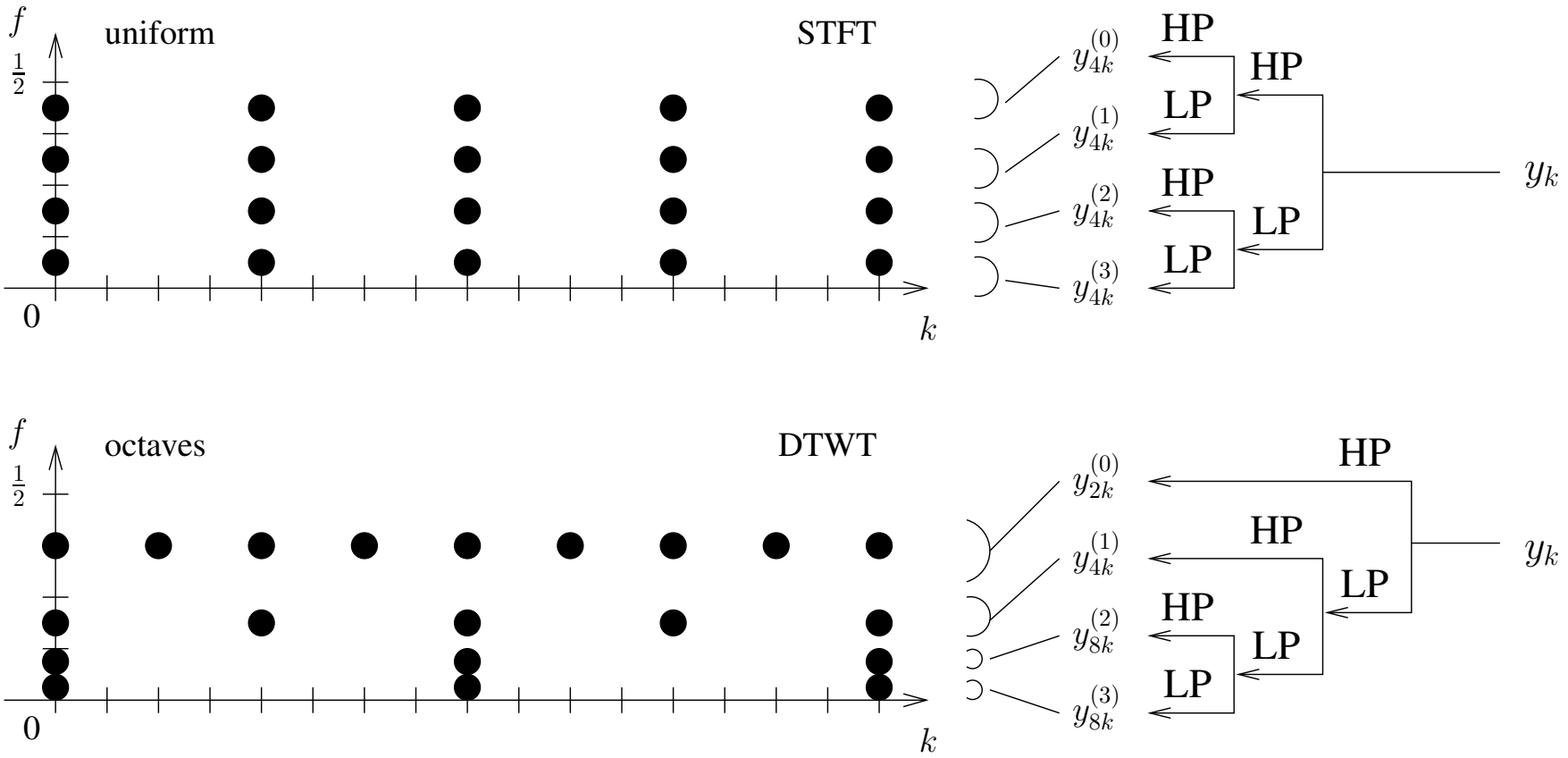




- H_1: high-pass filter, H_0: low-pass filter

Discrete-Time Wavelet Transform (2)

- time-frequency sampling patterns of STFT and DTWT:



Discrete-Time Wavelet Transform (3)

- wavelet transform:
  - $\rightarrow$ high frequency resolution at low frequencies
  - $\rightarrow$ high temporal resolution at high frequencies
- very well adapted to
  - $\rightarrow$ image coding (see source coding chapter)
  - $\rightarrow$ image analysis (// eye)
  - $\rightarrow$ sound analysis (// ear)
- logarithmic frequency = *scale*
-
  - high frequency components = *detail signals*
  - low frequency components $\rightarrow$ approximation of signal at a lower scale (as if standing further away $\Rightarrow$ looks more blurred)

  $\Rightarrow$ *multiresolution analysis*:
  different scales $\leftrightarrow$ different levels of resolution
- hierarchical signal reconstruction by adding detail signals at higher and higher levels of resolution

Decorrelating Transformations (8)

choice of transformation for various source coding applications

- linear prediction
  - speech coding: because of the connection between linear prediction and autoregressive modeling and the fact that speech signals can be approximated well with autoregressive models
  - low complexity solutions for image and video coding
- DCT-based transform coding
  - image and video coding
    artefacts appearing at low bit rate: blocking (discontinuities at the borders between consecutive image blocks that get transformed)
- subband coding
  - image and video coding
    artefacts appearing at low bit rate: ringing (large quantization errors at edges in the image get convolved with the synthesis bank impulse response)
  - audio coding (large bandwidth)
    limitations of hearing that can be expressed in the frequency domain can be incorporated in the coding process

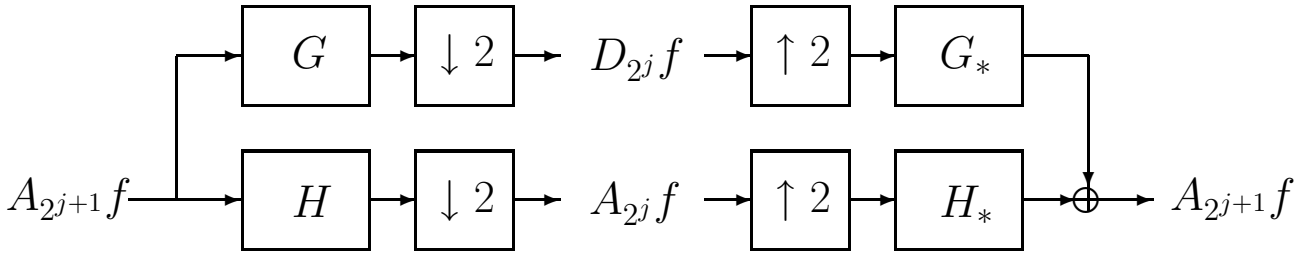
Decorrelating Transformations (9)

advantages of the Wavelet Transform (WT) for image coding :

- *perfect reconstruction* is possible
- the WT is a *multiresolution* description
  $\Rightarrow$ progressive decoding from lowest to highest resolution
- The WT is closer to the human visual system than the DCT transform
  $\Rightarrow$ artefacts introduced by WT coding are less annoying than those introduced by DCT coding (esp. at a high compression ratio)
- wavelet transform $\Rightarrow$ *scale-space representation*:
  - high frequency components precisely located in the pixel domain
    spatial resolution of the WT increases linearly with frequency
    $\Rightarrow$ allows the WT to *zoom* into the strong high frequency components of sharp edges and locate them accurately
  - low frequency components precisely located in frequency domain
    overall spectrum of most images very much of a low-pass type
    the frequency resolution is inversely proportional to frequency in the WT
    $\Rightarrow$ high decorrelation

Wavelet Transforms and Scale-Space Analysis

- the approximation $A_{2^j} f$ of a signal f at a resolution 2^j (estimate of f derived from 2^j measurements per unit length): computed by uniformly sampling at the rate 2^j the signal f smoothed by a low-pass filter whose bandwidth is proportional to 2^j. The approximation at resolution 2^j can be computed from the approximation at resolution 2^{j+1} by filtering it with a low-pass digital filter H and subsampling the output by a factor two.
- to extract the details of f which appear in $A_{2^{j+1}} f$ but not in $A_{2^j} f$: the *discrete detail* signal $D_{2^j} f$ at the resolution 2^j can be obtained by passing $A_{2^{j+1}} f$ through a high-pass filter G and subsampling the output by a factor two.
  $D_{2^j} f$ can be obtained by uniformly sampling at the rate 2^j the continuous-time signal f passed through an analog band-pass filter whose bandwidth is proportional to 2^j. This filter is usually referred to as a *wavelet*, and the mapping from the signal f to its discrete details is called the *discrete wavelet transform*
- Conversely, the approximation signal at the resolution 2^{j+1} can be recovered from the approximation and detail signals at the resolution 2^j by means of the "dual filtering system".



2D Wavelet Transform and Image Coding

- Consider the *separable* extension from 1D to 2D: the 2D impulse responses are obtained as the convolution of two 1D impulse responses, one for each of the horizontal and vertical directions. We first apply the 1D approach to the rows of the image $A_1 f$, yielding two subimages. Then we run the 1D QMF filters on the columns of these two subimages, yielding four subimages, each being one fourth of the original image in size: $A_{2^{-1}} f$ is the low-pass approximation in both directions, on which the decomposition process can be repeated, $D^1_{2^{-1}} f$ gives the vertical higher frequencies (horizontal edges), $D^2_{2^{-1}} f$ gives the horizontal higher frequencies (vertical edges), $D^3_{2^{-1}} f$ gives the higher frequencies in both directions (corners).

<table>
<tr><td>$A_{2^{-2}} f$</td><td>$D^1_{2^{-2}} f$</td><td rowspan="2">$D^1_{2^{-1}} f$</td></tr>
<tr><td>$D^2_{2^{-2}} f$</td><td>$D^3_{2^{-2}} f$</td></tr>
<tr><td colspan="2">$D^2_{2^{-1}} f$</td><td>$D^3_{2^{-1}} f$</td></tr>
</table>