Wavelet Packet Transforms and Best Bases: I

- discrete wavelet transforms (DWTs)
  - yields time/scale analysis of $\mathbf{X}$ of sample size N
  - need N to be a multiple of 2^{J_0} for partial DWT of level J_0
  - one partial DWT for each level $j = 1, \ldots, J_0$
  - scale τ_j related to frequencies in $(1/2^{j+1}, 1/2^j]$
  - scale λ_j related to frequencies in $(0, 1/2^{j+1}]$
  - splits $(0, 1/2]$ into octave bands
  - computed via pyramid algorithm
  - maximal overlap DWT also of interest

Wavelet Packet Transforms and Best Bases: II

- discrete wavelet packet transforms (DWPTs)
  - yields time/frequency analysis of $\mathbf{X}$
  - need N to be a multiple of 2^{J_0} for DWPT of level J_0
  - one DWPT for each level $j = 1, \ldots, J_0$
  - splits $(0, 1/2]$ into 2^j equal intervals
  - splitting resembles DFT (or ‘short time’ DFT)
  - computed via modification of pyramid algorithm
  - can ‘mix’ parts of DWPTs of different levels j, leading to many more orthonormal transforms and to the notion of a ‘best basis’ for a particular $\mathbf{X}$
  - maximal overlap DWPT (MODWPT) also of interest

Wavelet Packets – Basic Concepts: I

- recall that DWT pyramid algorithm can be expressed in terms of matrices $\mathcal{A}_j$ and $\mathcal{B}_j$ as $\mathbf{V}_j = \mathcal{A}_j \mathbf{V}_{j-1}$ and $\mathbf{W}_j = \mathcal{B}_j \mathbf{V}_{j-1}$, where, when, e.g., $L = 4$ and $N/2^{j-1} = 16$, we have

$$\mathcal{A}_j = \begin{bmatrix} g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 \\ g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g_3 & g_2 & g_1 & g_0 \end{bmatrix}$$

(there is a similar formulation for $\mathcal{B}_j$ in terms of $\{h_l\}$)

WMTSA: 207–208

Wavelet Packets – Basic Concepts: II

- 1st stage of DWT pyramid algorithm:

$$\mathcal{P}_1\mathbf{X} = \begin{bmatrix}\mathcal{B}_1\\ \mathcal{A}_1\end{bmatrix}\mathbf{X} = \begin{bmatrix}\mathbf{W}_1\\ \mathbf{V}_1\end{bmatrix} \equiv \begin{bmatrix}\mathbf{W}_{1,1}\\ \mathbf{W}_{1,0}\end{bmatrix}$$

  - $\mathbf{W}_{1,1} \equiv \mathbf{W}_1$ associated with $f \in (\frac{1}{4}, \frac{1}{2}]$
  - $\mathbf{W}_{1,0} \equiv \mathbf{V}_1$ associated with $f \in [0, \frac{1}{4}]$

- $\mathcal{P}_1$ is orthonormal:

$$\begin{aligned}\mathcal{P}_1\mathcal{P}_1^T = \begin{bmatrix}\mathcal{B}_1\\ \mathcal{A}_1\end{bmatrix}\begin{bmatrix}\mathcal{B}_1^T & \mathcal{A}_1^T\end{bmatrix} &= \begin{bmatrix}\mathcal{B}_1\mathcal{B}_1^T & \mathcal{B}_1\mathcal{A}_1^T\\ \mathcal{A}_1\mathcal{B}_1^T & \mathcal{A}_1\mathcal{A}_1^T\end{bmatrix}\\ &= \begin{bmatrix}I_{\frac{N}{2}} & 0_{\frac{N}{2}}\\ 0_{\frac{N}{2}} & I_{\frac{N}{2}}\end{bmatrix} = I_N\end{aligned}$$

- transform is $J_0 = 1$ partial DWT

Wavelet Packets – Basic Concepts: III

- likewise, 2nd stage defines $J_0 = 2$ partial DWT:

$$\begin{bmatrix} \mathcal{B}_1 \\ \mathcal{B}_2\mathcal{A}_1 \\ \mathcal{A}_2\mathcal{A}_1 \end{bmatrix} \mathbf{X} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \mathbf{V}_2 \end{bmatrix} \equiv \begin{bmatrix} \mathbf{W}_{1,1} \\ \mathbf{W}_{2,1} \\ \mathbf{W}_{2,0} \end{bmatrix}$$

  - $\mathbf{W}_{2,1} \equiv \mathbf{W}_2$ associated with $f \in (\frac{1}{8}, \frac{1}{4}]$
  - $\mathbf{W}_{2,0} \equiv \mathbf{V}_2$ associated with $f \in [0, \frac{1}{8}]$

- interpretation: we left $\mathcal{B}_1$ alone and rotated $\mathcal{A}_1$

- if we were to leave $\mathcal{A}_1$ alone and rotate $\mathcal{B}_1$ instead, we get a different transform, but one that is still orthonormal:

$$\begin{bmatrix} \mathcal{A}_2\mathcal{B}_1 \\ \mathcal{B}_2\mathcal{B}_1 \\ \mathcal{A}_1 \end{bmatrix} \mathbf{X} \equiv \begin{bmatrix} \mathbf{W}_{2,3} \\ \mathbf{W}_{2,2} \\ \mathbf{W}_{1,0} \end{bmatrix}$$

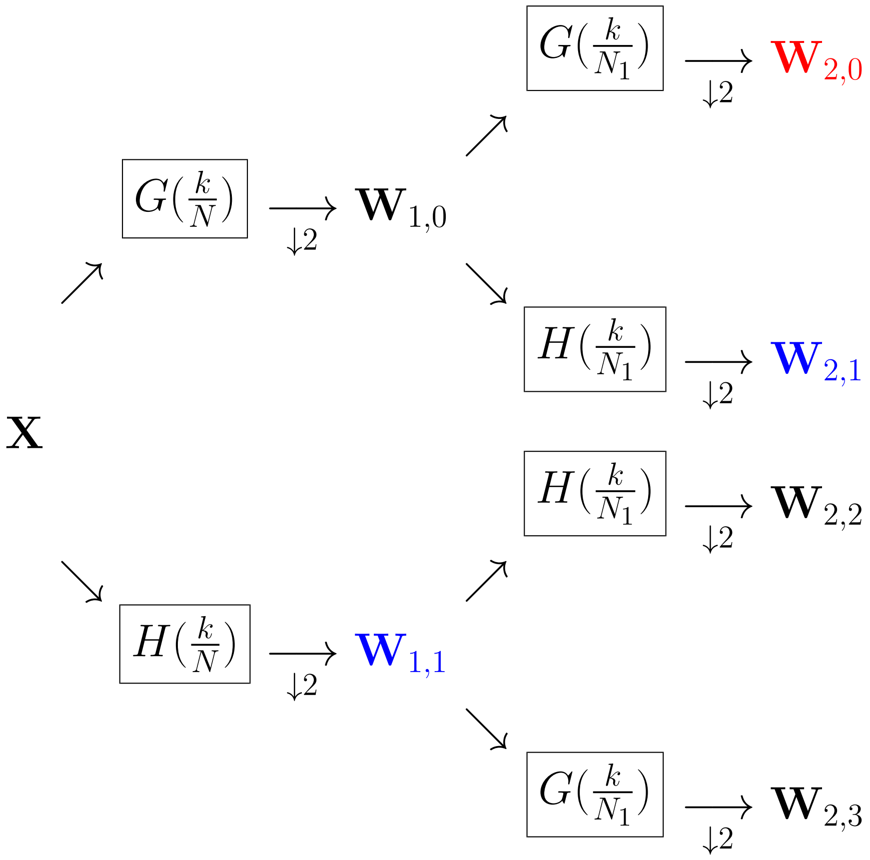
Wavelet Packets – Basic Concepts: IV

- to get yet another orthonormal transform, we can rotate both $\mathcal{B}_1$ and $\mathcal{A}_1$:

$$\begin{bmatrix} \mathcal{A}_2\mathcal{B}_1 \\ \mathcal{B}_2\mathcal{B}_1 \\ \mathcal{B}_2\mathcal{A}_1 \\ \mathcal{A}_2\mathcal{A}_1 \end{bmatrix} \mathbf{X} \equiv \begin{bmatrix} \mathbf{W}_{2,3} \\ \mathbf{W}_{2,2} \\ \mathbf{W}_{2,1} \\ \mathbf{W}_{2,0} \end{bmatrix}$$

Wavelet Packets – Basic Concepts: V

- flow diagram for transform from $\mathbf{X}$ to $\mathbf{W}_{2,0}$, $\mathbf{W}_{2,1}$, $\mathbf{W}_{2,2}$ and $\mathbf{W}_{2,3}$:

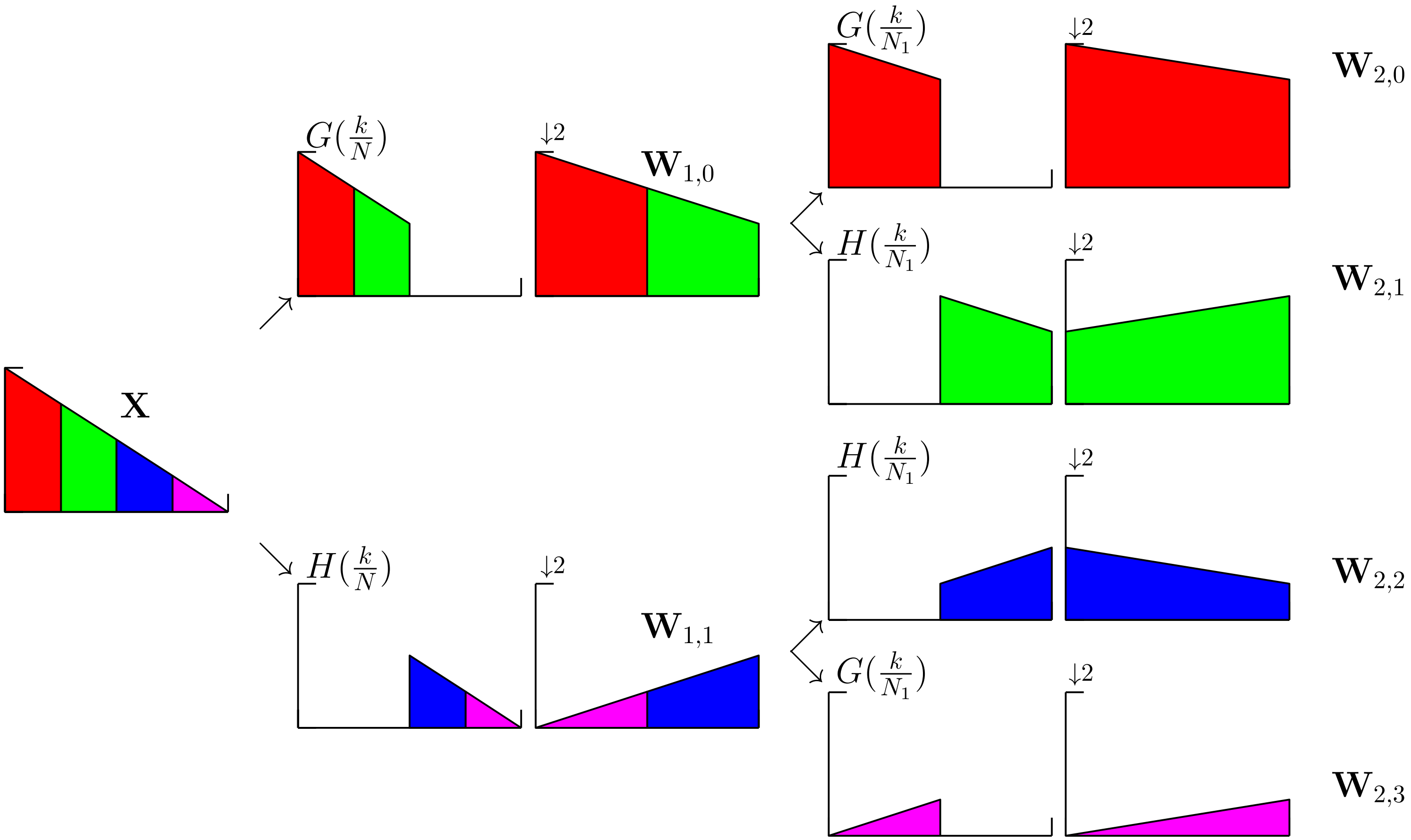


Wavelet Packets – Basic Concepts: VI

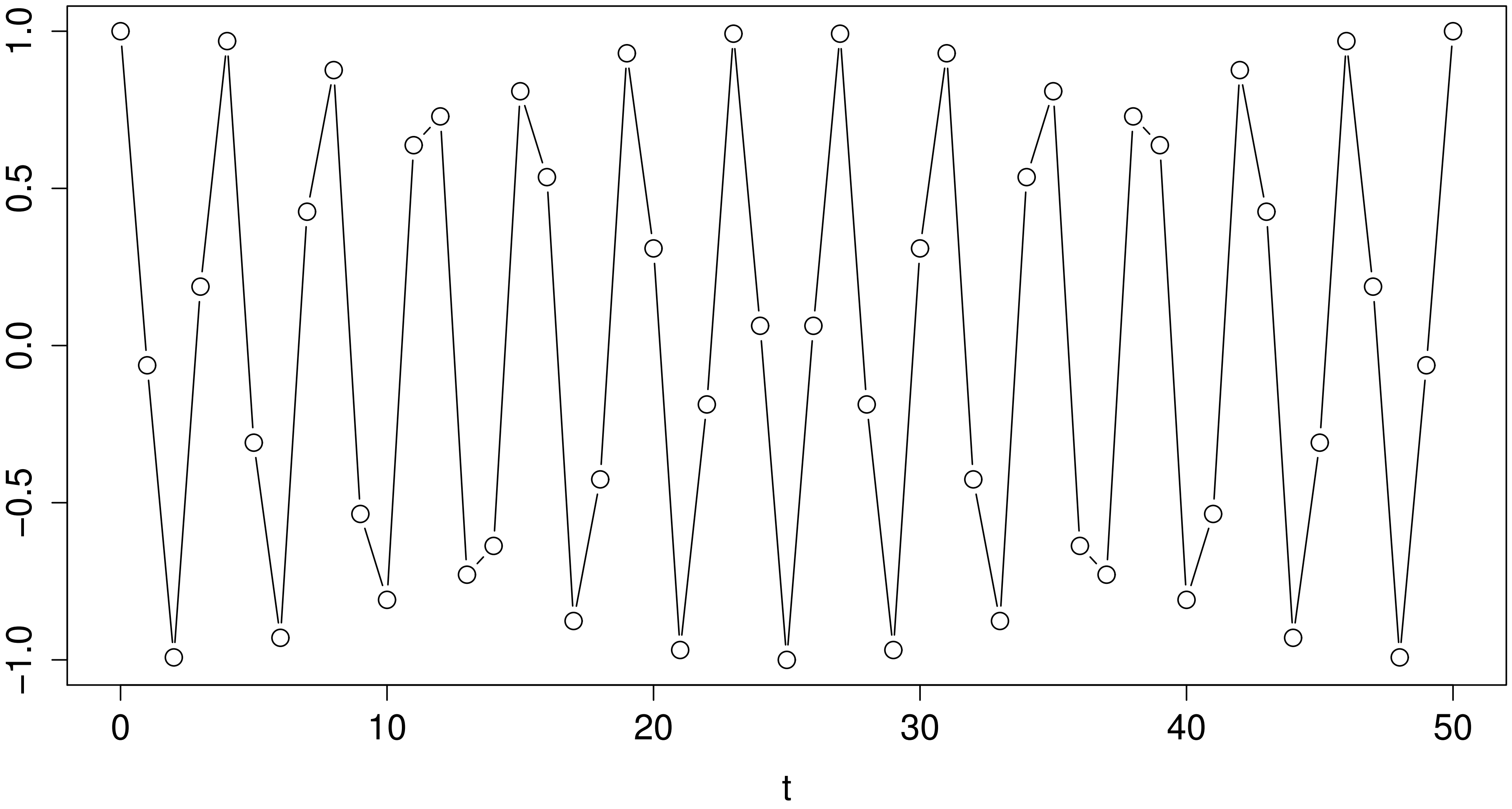
- can argue $\mathbf{W}_{2,0}$, $\mathbf{W}_{2,1}$, $\mathbf{W}_{2,2}$ and $\mathbf{W}_{2,3}$ are associated with $f \in [0, \frac{1}{8}]$, $(\frac{1}{8}, \frac{1}{4}]$, $(\frac{1}{4}, \frac{3}{8}]$ and $(\frac{3}{8}, \frac{1}{2}]$
- scheme sometimes called a 'regular' DWT because it splits $[0, \frac{1}{2}]$ split into 4 'regular' subintervals, each of width $1/8$
- basis for argument given in Section 4.4:
  - $\mathbf{V}_1$ related to $f \in [0, \frac{1}{4}]$ portion of $\mathbf{X}$
  - $\mathbf{W}_1$ related to $f \in (\frac{1}{4}, \frac{1}{2}]$ portion of $\mathbf{X}$ *but with reversal of order of frequencies*

Wavelet Packets – Basic Concepts: VII

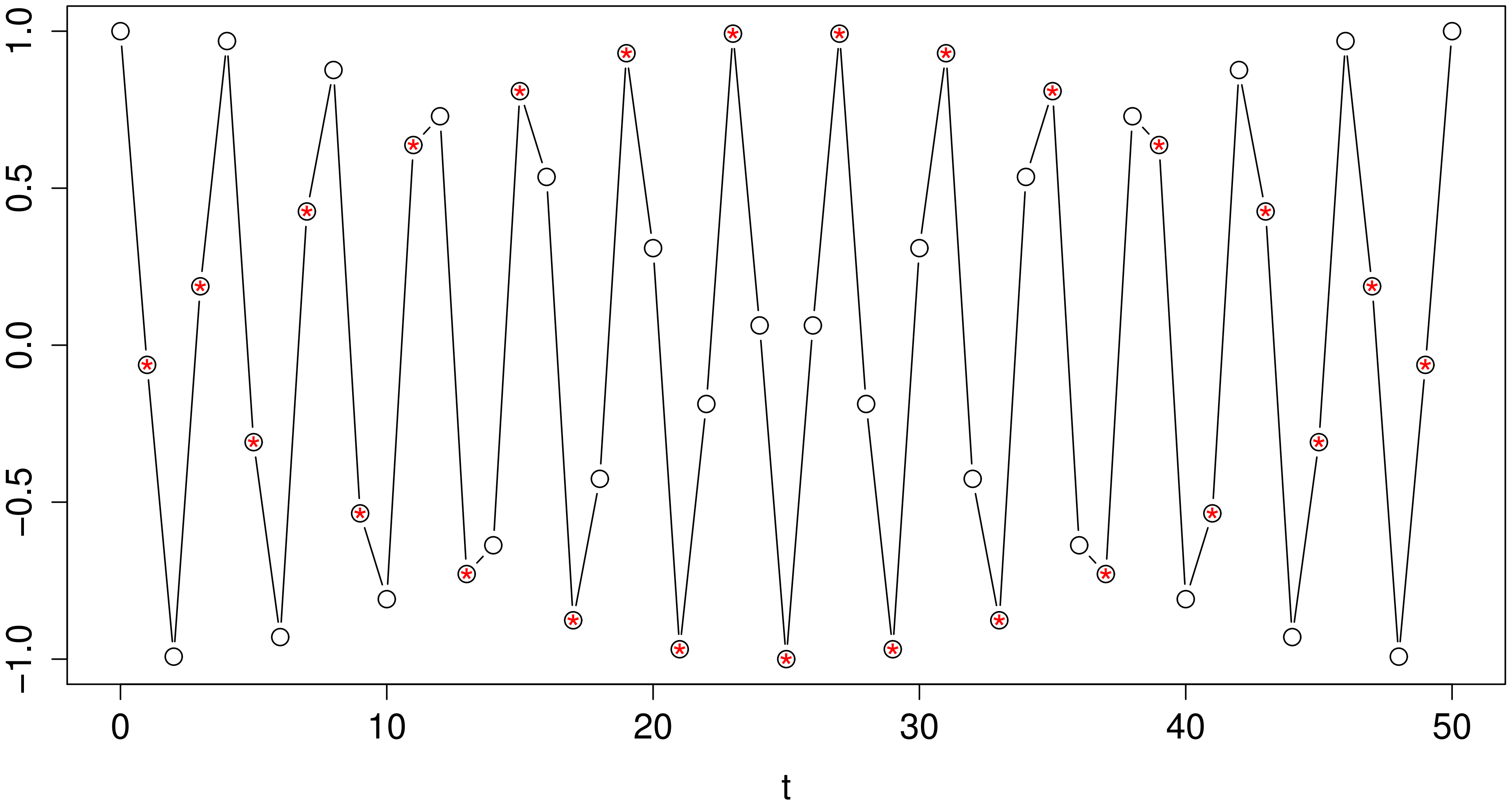
- flow diagram in frequency domain:



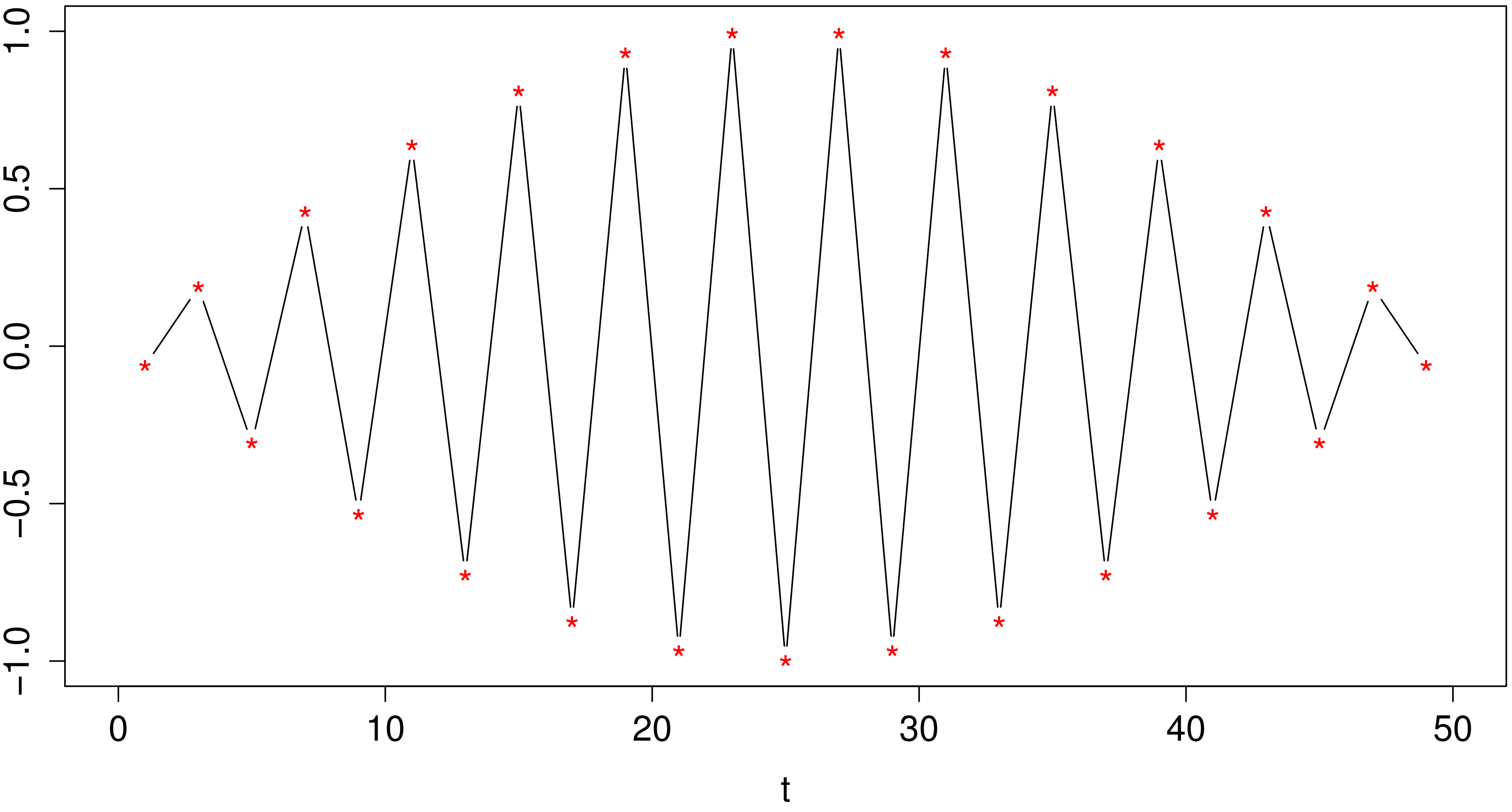
Downsampling $X_t = \cos(2\pi f_0 t)$ with $f_0 = 0.26$: I



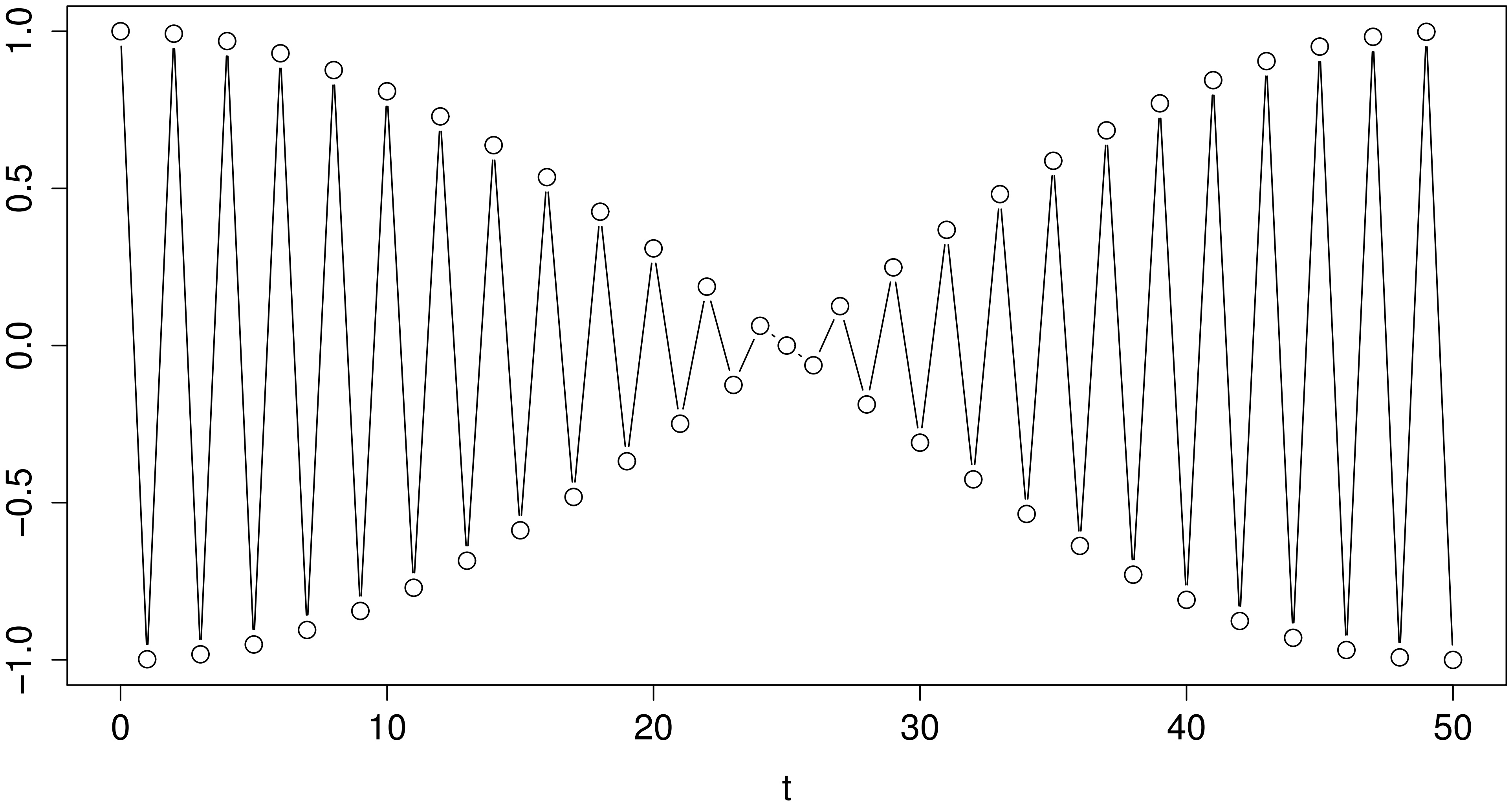
Downsampling $X_t = \cos(2\pi f_0 t)$ with $f_0 = 0.26$: II



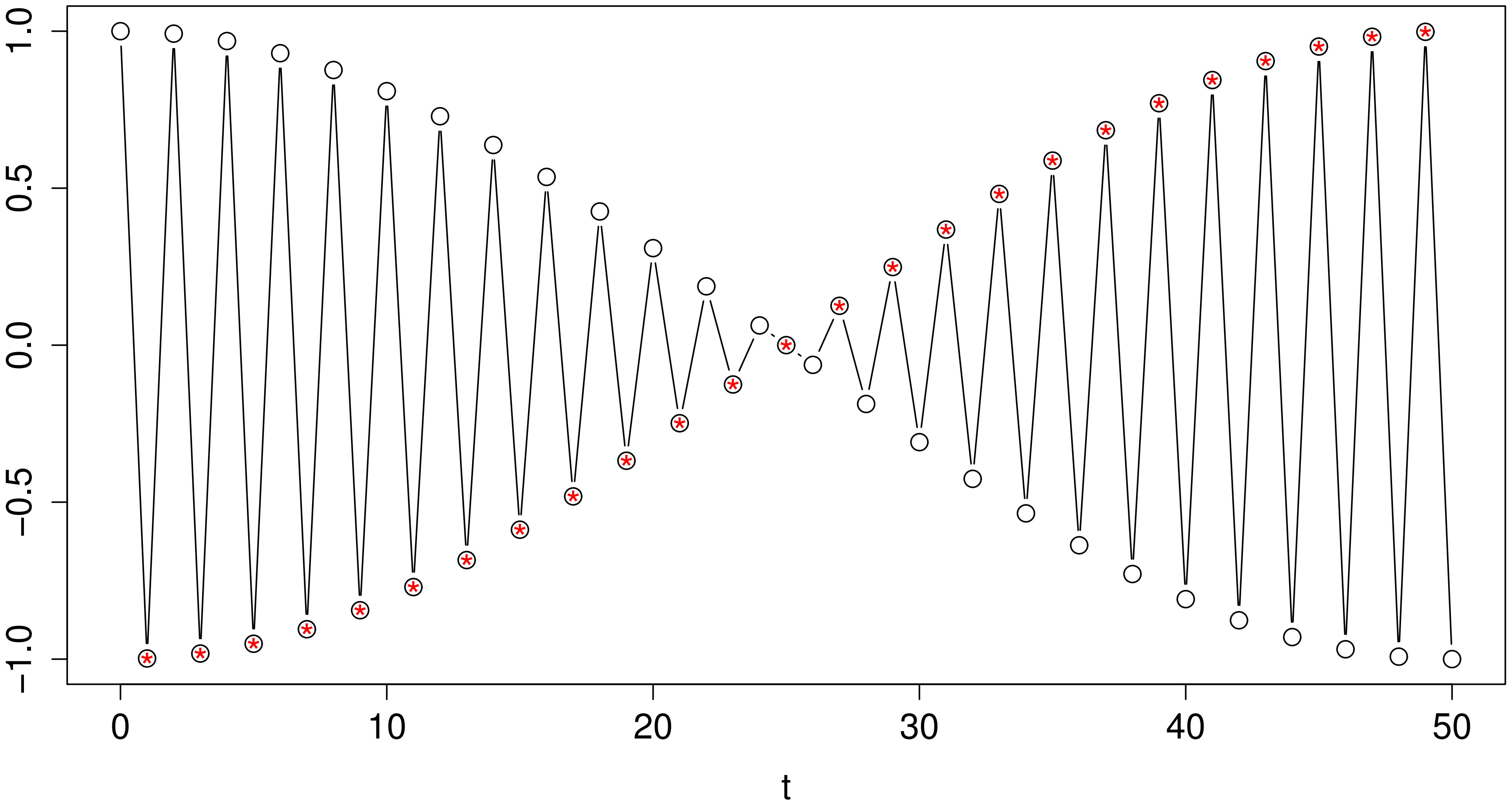
Downsampling $X_t = \cos(2\pi f_0 t)$ **with** $f_0 = 0.26$**: III**



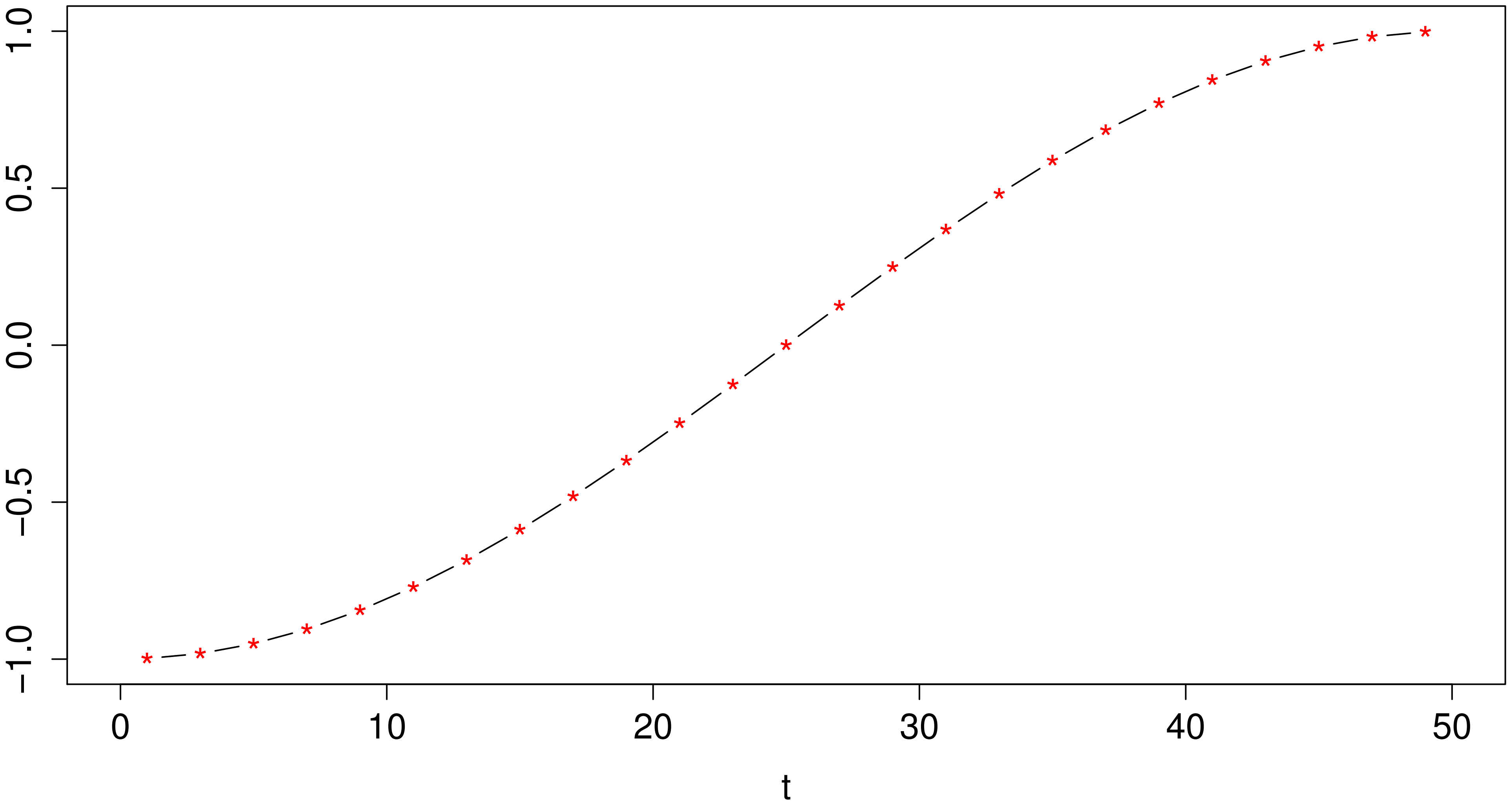
Downsampling $X_t = \cos(2\pi f_0 t)$ **with** $f_0 = 0.49$**: I**



Downsampling $X_t = \cos(2\pi f_0 t)$ with $f_0 = 0.49$: II



Downsampling $X_t = \cos(2\pi f_0 t)$ **with** $f_0 = 0.49$**: III**



Wavelet Packets – Basic Concepts: VIII

- transform from $\mathbf{X}$ to $\mathbf{W}_{2,0}$, $\mathbf{W}_{2,1}$, $\mathbf{W}_{2,2}$ and $\mathbf{W}_{2,3}$ is called a level $j = 2$ discrete wavelet packet transform
  - abbreviated as DWPT
  - splitting of $[0, \frac{1}{2}]$ similar to DFT
  - unlike DFT, DWPT coefficients localized (similar to so-called 'short time' Fourier transform)
  - DWPT is 'time/frequency'; DWT is 'time/scale'
- because level $j = 2$ DWPT is an orthonormal transform, we obtain an energy decomposition:

$$\|\mathbf{X}\|^2 = \sum_{n=0}^{3} \|\mathbf{W}_{2,n}\|^2$$

Wavelet Packets – Basic Concepts: IX

- can use level $j = 2$ DWPT to produce an additive decomposition (similar to an MRA):

$$\mathbf{X} = \left[\mathcal{B}_1^T\mathcal{A}_2^T, \mathcal{B}_1^T\mathcal{B}_2^T, \mathcal{A}_1^T\mathcal{B}_2^T, \mathcal{A}_1^T\mathcal{A}_2^T\right]\begin{bmatrix}\mathbf{W}_{2,3}\\ \mathbf{W}_{2,2}\\ \mathbf{W}_{2,1}\\ \mathbf{W}_{2,0}\end{bmatrix}$$

$$= \mathcal{B}_1^T\mathcal{A}_2^T\mathbf{W}_{2,3} + \mathcal{B}_1^T\mathcal{B}_2^T\mathbf{W}_{2,2} + \mathcal{A}_1^T\mathcal{B}_2^T\mathbf{W}_{2,1} + \mathcal{A}_1^T\mathcal{A}_2^T\mathbf{W}_{2,0}$$

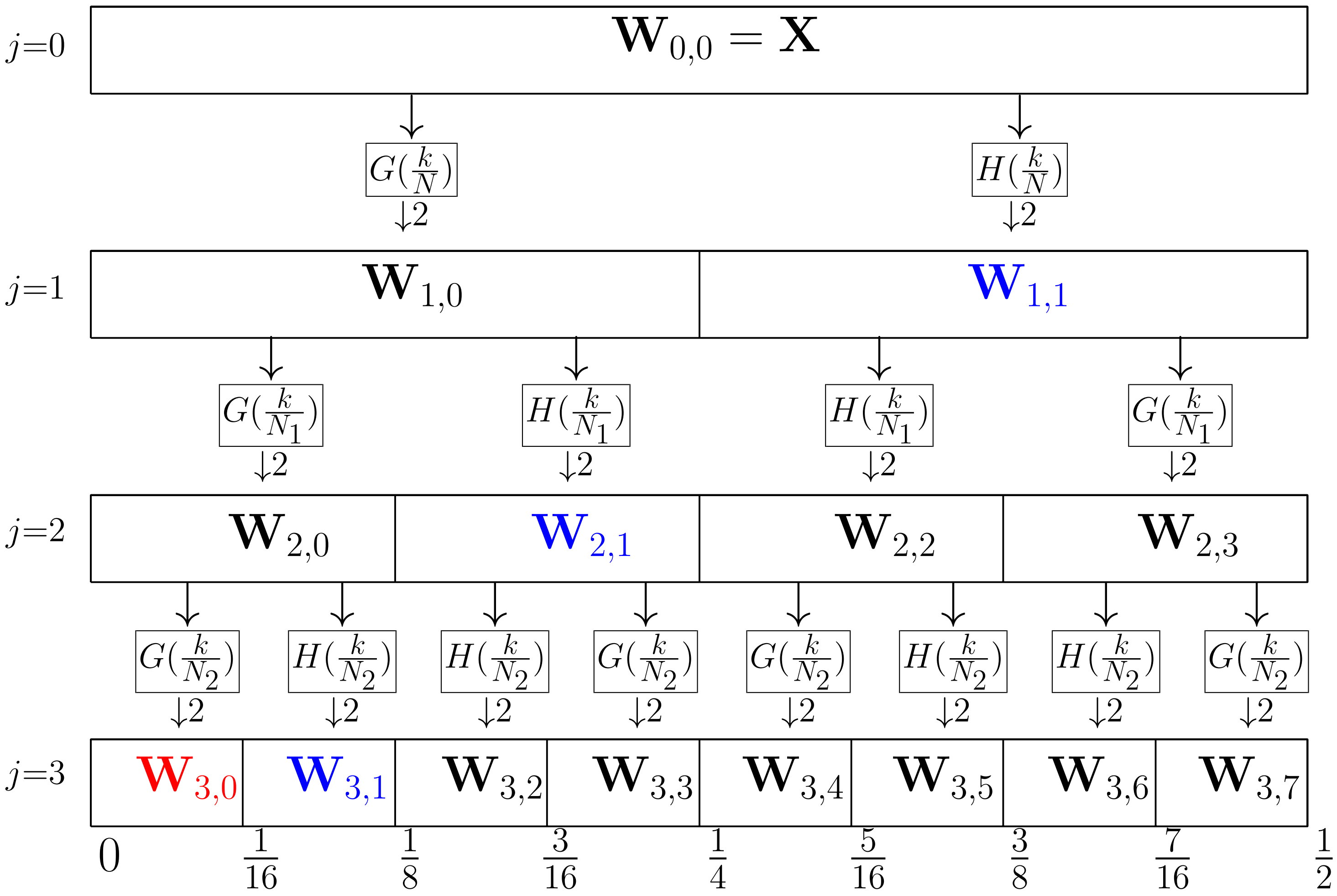
  - $\mathcal{B}_1^T\mathcal{A}_2^T\mathbf{W}_{2,3}$ associated with $f \in (\frac{3}{8}, \frac{1}{2}]$
  - $\mathcal{B}_1^T\mathcal{B}_2^T\mathbf{W}_{2,2}$ associated with $f \in (\frac{1}{4}, \frac{3}{8}]$
  - $\mathcal{A}_1^T\mathcal{B}_2^T\mathbf{W}_{2,1}$ associated with $f \in (\frac{1}{8}, \frac{1}{4}]$
  - $\mathcal{A}_1^T\mathcal{A}_2^T\mathbf{W}_{2,0}$ associated with $f \in [0, \frac{1}{8}]$

DWPTs of General Levels: I

- can generalize scheme to define DWPTs for levels $j = 0, 1, 2, 3, \ldots$ (with $\mathbf{W}_{0,0}$ defined to be $\mathbf{X}$)
- idea behind DWPT is to use $G(\cdot)$ and $H(\cdot)$ to split each of the 2^{j-1} vectors on level $j-1$ into 2 new vectors, ending up with a level j transform with 2^j vectors
- given $\mathbf{W}_{j-1,n}$'s, here is the rule for generating $\mathbf{W}_{j,n}$'s:
  - if n in $\mathbf{W}_{j-1,n}$ is even:
    - to get $\mathbf{W}_{j,2n}$, use $G(\cdot)$ to transform $\mathbf{W}_{j-1,n}$
    - to get $\mathbf{W}_{j,2n+1}$, use $H(\cdot)$ to transform $\mathbf{W}_{j-1,n}$
  - if n in $\mathbf{W}_{j-1,n}$ is odd:
    - to get $\mathbf{W}_{j,2n}$, use $H(\cdot)$ to transform $\mathbf{W}_{j-1,n}$
    - to get $\mathbf{W}_{j,2n+1}$,use $G(\cdot)$ to transform $\mathbf{W}_{j-1,n}$

DWPTs of General Levels: II

- example of rule, yielding level $j = 3$ DWPT in the bottom row



$j{=}0$ $\mathbf{W}_{0,0} = \mathbf{X}$

$G(\frac{k}{N})$ $\downarrow 2$ $H(\frac{k}{N})$ $\downarrow 2$

$j{=}1$ $\mathbf{W}_{1,0}$ $\mathbf{W}_{1,1}$

$G(\frac{k}{N_1})$ $\downarrow 2$ $H(\frac{k}{N_1})$ $\downarrow 2$ $H(\frac{k}{N_1})$ $\downarrow 2$ $G(\frac{k}{N_1})$ $\downarrow 2$

$j{=}2$ $\mathbf{W}_{2,0}$ $\mathbf{W}_{2,1}$ $\mathbf{W}_{2,2}$ $\mathbf{W}_{2,3}$

$G(\frac{k}{N_2})$ $\downarrow 2$ $H(\frac{k}{N_2})$ $\downarrow 2$ $H(\frac{k}{N_2})$ $\downarrow 2$ $G(\frac{k}{N_2})$ $\downarrow 2$ $G(\frac{k}{N_2})$ $\downarrow 2$ $H(\frac{k}{N_2})$ $\downarrow 2$ $H(\frac{k}{N_2})$ $\downarrow 2$ $G(\frac{k}{N_2})$ $\downarrow 2$

$j{=}3$ $\mathbf{W}_{3,0}$ $\mathbf{W}_{3,1}$ $\mathbf{W}_{3,2}$ $\mathbf{W}_{3,3}$ $\mathbf{W}_{3,4}$ $\mathbf{W}_{3,5}$ $\mathbf{W}_{3,6}$ $\mathbf{W}_{3,7}$

0 $\frac{1}{16}$ $\frac{1}{8}$ $\frac{3}{16}$ $\frac{1}{4}$ $\frac{5}{16}$ $\frac{3}{8}$ $\frac{7}{16}$ $\frac{1}{2}$

WMTSA: 212

DWPTs of General Levels: III

- note: $\mathbf{W}_{j,0}$ and $\mathbf{W}_{j,1}$ correspond to vectors $\mathbf{V}_j$ and $\mathbf{W}_j$ in a jth level partial DWT
- $\mathbf{W}_{j,n}$, $n = 0, \ldots, 2^j - 1$, is associated with $f \in (\frac{n}{2^{j+1}}, \frac{n+1}{2^{j+1}}]$
- n is called the 'sequency' index
- in terms of circular filtering, we can write

$$W_{j,n,t} = \sum_{l=0}^{L-1} u_{n,l} W_{j-1,\lfloor \frac{n}{2} \rfloor, 2t+1-l \bmod N/2^j}, \quad t = 0, \ldots, \frac{N}{2^j} - 1,$$

where $W_{j,n,t}$ is the tth element of $\mathbf{W}_{j,n}$ and

$$u_{n,l} \equiv \begin{cases} g_l, & \text{if } n \bmod 4 = 0 \text{ or } 3; \\ h_l, & \text{if } n \bmod 4 = 1 \text{ or } 2. \end{cases}$$

DWPTs of General Levels: IV

- can also get $\mathbf{W}_{j,n}$ by filtering $\mathbf{X}$ and downsampling:

$$W_{j,n,t} = \sum_{l=0}^{L_j-1} u_{j,n,l} X_{2^j[t+1]-1-l \bmod N}, \quad t = 0, 1, \ldots, \frac{N}{2^j}-1,$$

  where $\{u_{j,n,l}\}$ is the equivalent filter associated with $\mathbf{W}_{j,n}$

- let $\{u_{j,n,l}\} \longleftrightarrow U_{j,n}(\cdot), n = 0, \ldots, 2^j - 1$
- to construct $U_{j,n}(\cdot)$, define $M_0(f) = G(f)$ & $M_1(f) = H(f)$
- let $\mathbf{c}_{1,0} \equiv [0]$ & $\mathbf{c}_{1,1} \equiv [1]$ &, for $j > 1$, create $\mathbf{c}_{j,n}$ recursively
  - by appending 0 to $\mathbf{c}_{j-1,\lfloor\frac{n}{2}\rfloor}$ if $n \bmod 4 = 0$ or 3 or
  - by appending 1 to $\mathbf{c}_{j-1,\lfloor\frac{n}{2}\rfloor}$ if $n \bmod 4 = 1$ or 2

DWPTs of General Levels: V

- letting $c_{j,n,m}$ be mth element of $\mathbf{c}_{j,n}$, then

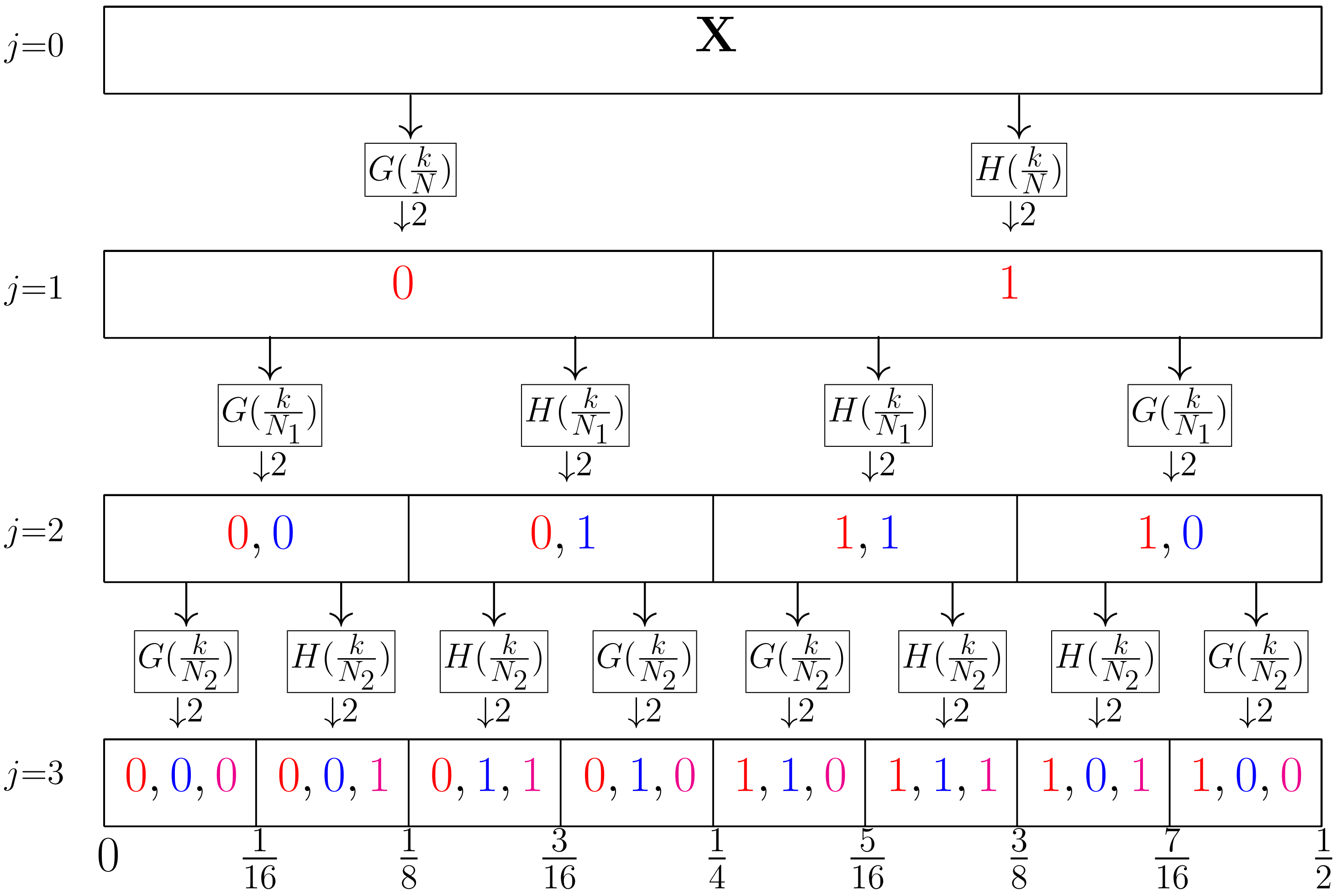
$$U_{j,n}(f) = \prod_{m=0}^{j-1} M_{c_{j,n,m}}(2^m f)$$

- example: $\mathbf{c}_{3,3} = [0, 1, 0]^T$ says

$$U_{3,3}(f) = M_0(f)M_1(2f)M_0(4f) = G(f)H(2f)G(4f)$$

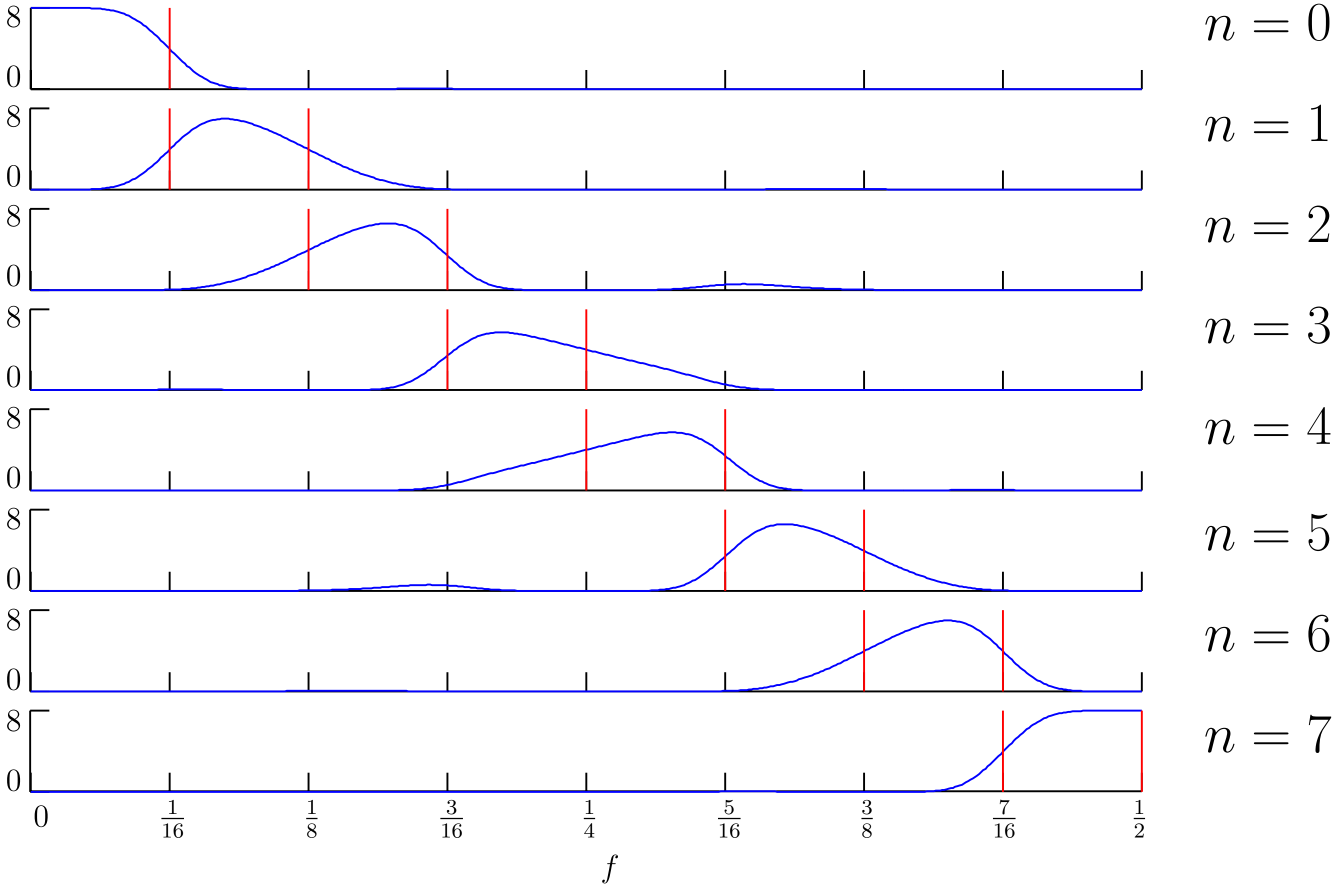
DWPTs of General Levels: VI

- contents of $\mathbf{c}_{j,n}$ for $j = 1, 2$ & 3 and $n = 0, \ldots, 2^j - 1$



DWPTs of General Levels: VII

- squared gain functions $|U_{3,n}(\cdot)|^2$ using LA(8) $\{g_l\}$ & $\{h_l\}$



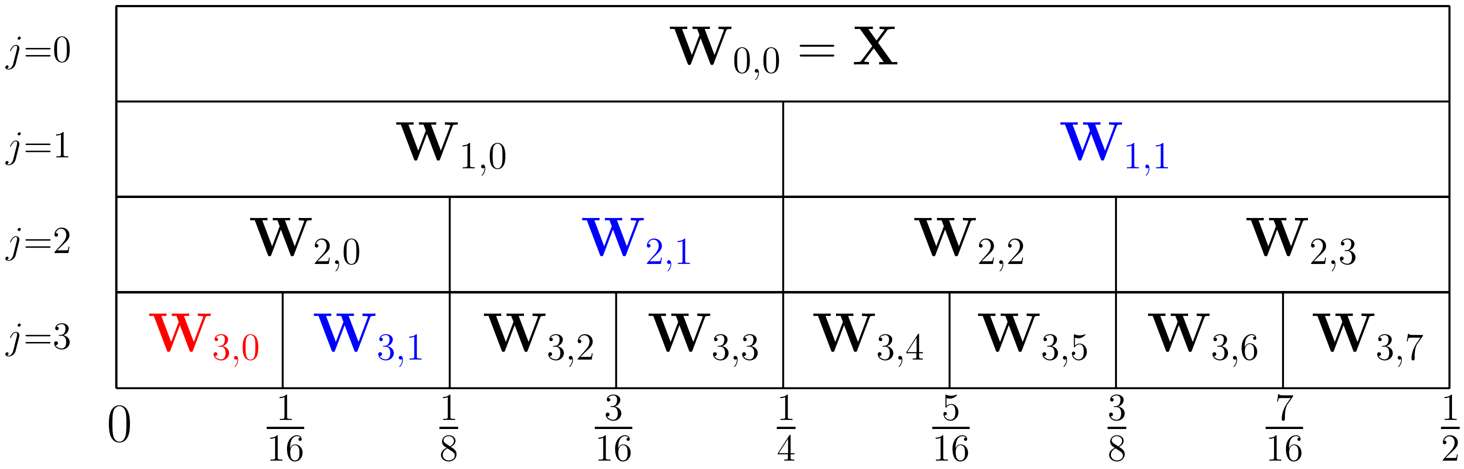
- note overlap in $n = 3$ and 4 bands – not well separated

DWPTs of General Levels: VIII

- $\mathbf{W}_{j,n}$ nominally associated with bandwidth $1/2^{j+1}$ (corresponding frequency interval is $\mathcal{I}_{j,n} \equiv (\frac{n}{2^{j+1}}, \frac{n+1}{2^{j+1}}])$
- $\mathbf{W}_{j,0}$ same as $\mathbf{V}_j$ in level j partial DWT
- since $\mathbf{V}_j$ has scale $\lambda_j = 2^j$, can say $\mathbf{W}_{j,0}$ has 'time width' λ_j
- each $\{u_{j,n,l}\}$ has width L_j, so each $\mathbf{W}_{j,n}$ has time width λ_j
- $j = 0$: time width is unity and bandwidth is $1/2$
- $j = J$: time width is $N = 2^J$ and bandwidth is $1/2N$
- note that time width × bandwidth is constant, which is an example of 'reciprocity relationship'
- aside: level J Haar DWPT same as Walsh transform (see Fig. 218 – 'global,' as is DFT)

Wavelet Packet Tables/Trees: I

- collection of DWPTs called a wavelet packet table (or tree), with the tree nodes being labeled by the doublets (j, n):

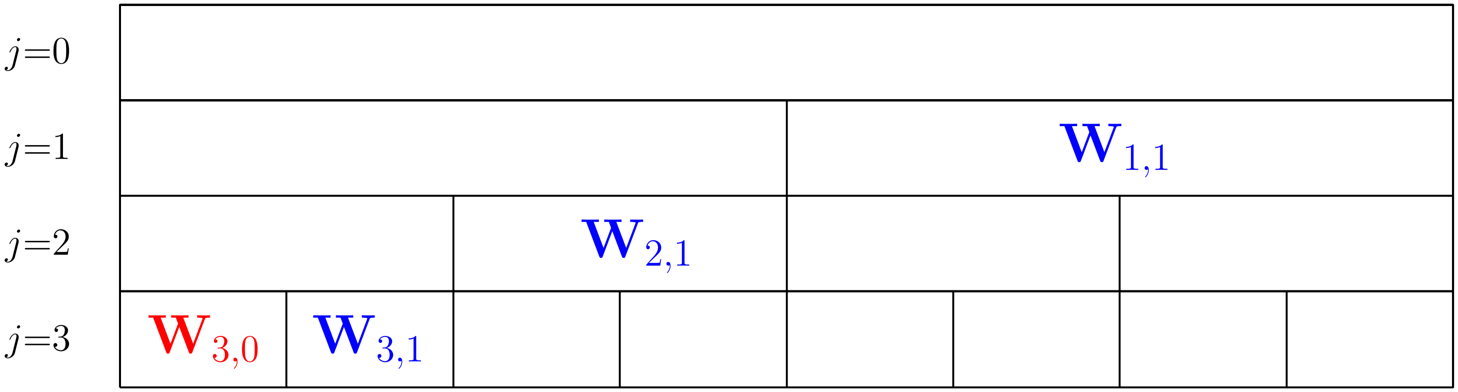


<table>
<tr><td>j=0</td><td colspan="8">$\mathbf{W}_{0,0} = \mathbf{X}$</td></tr>
<tr><td>j=1</td><td colspan="4">$\mathbf{W}_{1,0}$</td><td colspan="4">$\mathbf{W}_{1,1}$</td></tr>
<tr><td>j=2</td><td colspan="2">$\mathbf{W}_{2,0}$</td><td colspan="2">$\mathbf{W}_{2,1}$</td><td colspan="2">$\mathbf{W}_{2,2}$</td><td colspan="2">$\mathbf{W}_{2,3}$</td></tr>
<tr><td>j=3</td><td>$\mathbf{W}_{3,0}$</td><td>$\mathbf{W}_{3,1}$</td><td>$\mathbf{W}_{3,2}$</td><td>$\mathbf{W}_{3,3}$</td><td>$\mathbf{W}_{3,4}$</td><td>$\mathbf{W}_{3,5}$</td><td>$\mathbf{W}_{3,6}$</td><td>$\mathbf{W}_{3,7}$</td></tr>
</table>

- nodes $\mathcal{C} \equiv \{(j, n) : n = 0, \ldots, 2^j - 1\}$ for row j form a DWPT
- nonoverlapping complete covering of $[0, \frac{1}{2}]$ yields coefficients for an orthonormal transform $\mathbf{O}$ ('disjoint dyadic decomposition')
- let's consider 2 sets of doublets yielding such a decomposition

Wavelet Packet Tables/Trees: II

- $\mathcal{C} = \{(3,0),(3,1),(2,1),(1,1)\}$ yields the DWT:



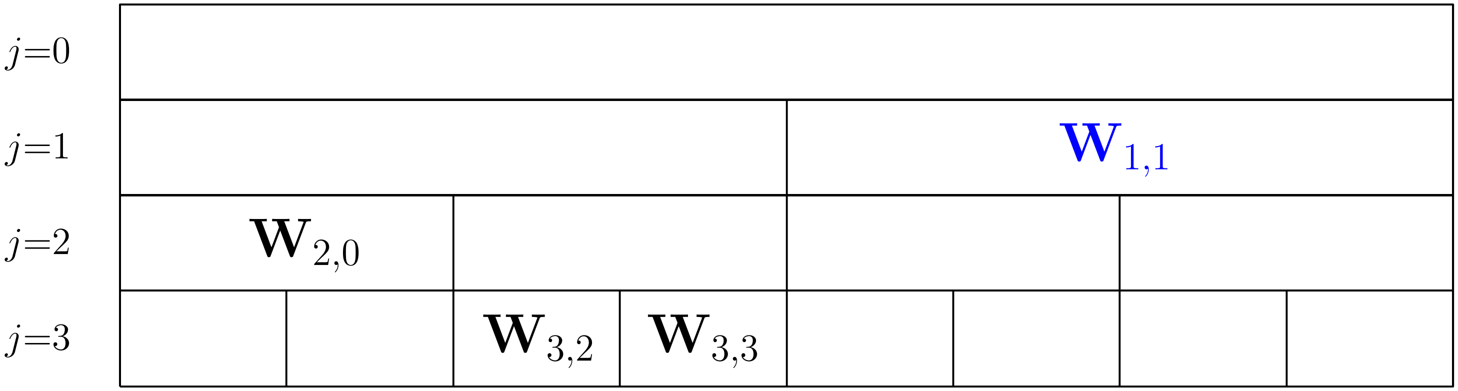
j=0

j=1 $\mathbf{W}_{1,1}$

j=2 $\mathbf{W}_{2,1}$

j=3 $\mathbf{W}_{3,0}$ $\mathbf{W}_{3,1}$

- $\mathcal{C} = \{(2,0),(3,2),(3,3),(1,1)\}$ yields another $\mathbf{O}$:



j=0

j=1 $\mathbf{W}_{1,1}$

j=2 $\mathbf{W}_{2,0}$

j=3 $\mathbf{W}_{3,2}$ $\mathbf{W}_{3,3}$

Optimal Orthonormal Transform: I

- WP table yields *many* **O**'s: does one match **X** 'optimally'?
- Coifman & Wickerhauser (1992) proposed notion of 'best basis'
- form WP table out to level J, and assign 'cost' to $\mathbf{W}_{j,n}$ via

$$M(\mathbf{W}_{j,n}) \equiv \sum_{t=0}^{N_j-1} m(|W_{j,n,t}|)$$

  where $m(\cdot)$ is real-valued cost function (require $m(0) = 0$)
- let $\mathcal{C}$ be any collection of indices in the set $\mathcal{N}$ of all possible indices forming an orthonormal transform
- 'optimal' such transform satisfies

$$\min_{\mathcal{C}\in\mathcal{N}} \sum_{(j,n)\in\mathcal{C}} M(\mathbf{W}_{j,n})$$

Optimal Orthonormal Transform: II

- consider following 2 unit norm vectors:

$$\mathbf{W}_{j,n}^{(1)} = \left[\tfrac{1}{2}, -\tfrac{1}{2}, \tfrac{1}{2}, -\tfrac{1}{2}\right]^T \text{ and } \mathbf{W}_{j,n}^{(2)} = [1, 0, 0, 0]^T$$

- example: 'entropy-based' cost function

$$m(|\overline{W}_{j,n,t}|) = -\overline{W}^2_{j,n,t} \log(\overline{W}^2_{j,n,t}),$$

  where $\overline{W}_{j,n,t} \equiv W_{j,n,t}/\|\mathbf{X}\|$ (assume $\|\mathbf{X}\| = 1$ in what follows)

  - note: since $|x| \log(|x|) \to 0$ as $x \to 0$, will interpret $0 \cdot \log(0)$ as 0

- here $M(\mathbf{W}_{j,n}^{(1)}) = 4 \cdot (-\frac{1}{4} \log \frac{1}{4}) > 0$ and $M(\mathbf{W}_{j,n}^{(2)}) = 0$ (lower cost if energy is concentrated in a few $|W_{j,n,t}|$'s)

Optimal Orthonormal Transform: III

- continue looking at $\mathbf{W}_{j,n}^{(1)} = \left[\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right]^T$ & $\mathbf{W}_{j,n}^{(2)} = [1, 0, 0, 0]^T$
- 2nd example: threshold cost function

$$m(|W_{j,n,t}|) = \begin{cases} 1, & \text{if } |W_{j,n,t}| > \delta; \\ 0, & \text{otherwise.} \end{cases}$$

if $\delta = 1/4$, $M(\mathbf{W}_{j,n}^{(1)}) = 4$ and $M(\mathbf{W}_{j,n}^{(2)}) = 1$
(lower cost if there are only a few large $|W_{j,n,t}|$'s)

- 3rd example: ℓ_p cost function $m(|W_{j,n,t}|) = |W_{j,n,t}|^p$

if $p = 1$, $M(\mathbf{W}_{j,n}^{(1)}) = 2$ and $M(\mathbf{W}_{j,n}^{(2)}) = 1$
(same pattern as before)

- once costs assigned, need to find optimal transform

Optimal Orthonormal Transform: IV

- example: consider Haar DWPTs out to level $j = 3$:

$$\begin{bmatrix} \mathbf{W}_{1,1} \\ \mathbf{W}_{1,0} \end{bmatrix} = \begin{bmatrix} \mathcal{B}_1 \\ \mathcal{A}_1 \end{bmatrix} \mathbf{X}, \quad \begin{bmatrix} \mathbf{W}_{2,3} \\ \mathbf{W}_{2,2} \\ \mathbf{W}_{2,1} \\ \mathbf{W}_{2,0} \end{bmatrix} = \begin{bmatrix} \mathcal{A}_2\mathcal{B}_1 \\ \mathcal{B}_2\mathcal{B}_1 \\ \mathcal{B}_2\mathcal{A}_1 \\ \mathcal{A}_2\mathcal{A}_1 \end{bmatrix} \mathbf{X},$$

$$\begin{bmatrix} \mathbf{W}_{3,7} \\ \mathbf{W}_{3,6} \\ \mathbf{W}_{3,5} \\ \mathbf{W}_{3,4} \\ \mathbf{W}_{3,3} \\ \mathbf{W}_{3,2} \\ \mathbf{W}_{3,1} \\ \mathbf{W}_{3,0} \end{bmatrix} = \begin{bmatrix} \mathcal{A}_3\mathcal{A}_2\mathcal{B}_1 \\ \mathcal{B}_3\mathcal{A}_2\mathcal{B}_1 \\ \mathcal{B}_3\mathcal{B}_2\mathcal{B}_1 \\ \mathcal{A}_3\mathcal{B}_2\mathcal{B}_1 \\ \mathcal{A}_3\mathcal{B}_2\mathcal{A}_1 \\ \mathcal{B}_3\mathcal{B}_2\mathcal{A}_1 \\ \mathcal{B}_3\mathcal{A}_2\mathcal{A}_1 \\ \mathcal{A}_3\mathcal{A}_2\mathcal{A}_1 \end{bmatrix} \mathbf{X}$$

Optimal Orthonormal Transform: V

- let $\mathbf{X}$ be following series of length $N = 8$:

$$\mathbf{X} = \sqrt{2}\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + 2\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \end{bmatrix} + \sqrt{8}\begin{bmatrix} \frac{1}{\sqrt{8}} \\ -\frac{1}{\sqrt{8}} \\ -\frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ -\frac{1}{\sqrt{8}} \\ -\frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \\ 0 \\ 0 \\ -2 \\ 2 \end{bmatrix}$$

- note that $\mathbf{X}$ is a linear combination of transposes of 1st row of $\mathcal{A}_1$, 2nd row of $\mathcal{A}_2\mathcal{B}_1$ and single row of $\mathcal{A}_3\mathcal{B}_2\mathcal{B}_1$

Optimal Orthonormal Transform: VI

- Haar DWPT coefficients, levels $j = 1, 2$ and 3 (three underlined coefficents correspond to basis vectors used in forming $\mathbf{X}$):

<table>
<tr><td>j=0</td><td colspan="8">$\mathbf{X} = [2, 0, -1, 1, 0, 0, -2, 2]^T$</td></tr>
<tr><td>j=1</td><td colspan="4">$[\underline{\sqrt{2}}, 0, 0, 0]$</td><td colspan="4">$[-\sqrt{2}, \sqrt{2}, 0, \sqrt{8}]$</td></tr>
<tr><td>j=2</td><td colspan="2">$[1, 0]$</td><td colspan="2">$[-1, 0]$</td><td colspan="2">$[2, 2]$</td><td colspan="2">$[0, \underline{2}]$</td></tr>
<tr><td>j=3</td><td>$[\frac{1}{\sqrt{2}}]$</td><td>$[-\frac{1}{\sqrt{2}}]$</td><td>$[\frac{1}{\sqrt{2}}]$</td><td>$[-\frac{1}{\sqrt{2}}]$</td><td>$[\underline{\sqrt{8}}]$</td><td>$[0]$</td><td>$[\sqrt{2}]$</td><td>$[\sqrt{2}]$</td></tr>
</table>

Optimal Orthonormal Transform: VII

- cost table using $-W_{j,n,t}^2 \log(W_{j,n,t}^2)$ cost function:

j=0	1.45							
j=1	0.28				0.88			
j=2	0.19		0.19		0.72		0.36	
j=3	**0.12**	**0.12**	**0.12**	**0.12**	**0.32**	**0.00**	**0.28**	**0.28**

- algorithm to find ‘best’ basis
  - mark all costs of ‘childern’ nodes at bottom
  - compare cost of children with their ‘parent’
    - if parent cheaper, mark parent node
    - if children cheaper, replace cost of parent
  - repeat for each level; when done, look for top-marked nodes

Optimal Orthonormal Transform: VIII

- compare 0.12 + 0.12 = 0.24 to 0.19: parent cheaper!

<table>
<tr><td>j=0</td><td colspan="8">1.45</td></tr>
<tr><td>j=1</td><td colspan="4">0.28</td><td colspan="4">0.88</td></tr>
<tr><td>j=2</td><td colspan="2">0.19</td><td colspan="2">0.19</td><td colspan="2">0.72</td><td colspan="2">0.36</td></tr>
<tr><td>j=3</td><td>0.12</td><td>0.12</td><td>0.12</td><td>0.12</td><td>0.32</td><td>0.00</td><td>0.28</td><td>0.28</td></tr>
</table>

Optimal Orthonormal Transform: VIII

- mark parent node

<table>
<tr><td>$j{=}0$</td><td colspan="8">1.45</td></tr>
<tr><td>$j{=}1$</td><td colspan="4">0.28</td><td colspan="4">0.88</td></tr>
<tr><td>$j{=}2$</td><td colspan="2">0.19</td><td colspan="2">0.19</td><td colspan="2">0.72</td><td colspan="2">0.36</td></tr>
<tr><td>$j{=}3$</td><td>0.12</td><td>0.12</td><td>0.12</td><td>0.12</td><td>0.32</td><td>0.00</td><td>0.28</td><td>0.28</td></tr>
</table>

Optimal Orthonormal Transform: VIII

- next comparison is the same, so again mark parent node

<table>
<tr><td>j=0</td><td colspan="8">1.45</td></tr>
<tr><td>j=1</td><td colspan="4">0.28</td><td colspan="4">0.88</td></tr>
<tr><td>j=2</td><td colspan="2">**0.19**</td><td colspan="2">**0.19**</td><td colspan="2">0.72</td><td colspan="2">0.36</td></tr>
<tr><td>j=3</td><td>**0.12**</td><td>**0.12**</td><td>**0.12**</td><td>**0.12**</td><td>**0.32**</td><td>**0.00**</td><td>**0.28**</td><td>**0.28**</td></tr>
</table>

Optimal Orthonormal Transform: VIII

- compare 0.32 + 0.00 = 0.32 to 0.72: children cheaper!

<table>
<tr><td>j=0</td><td colspan="8">1.45</td></tr>
<tr><td>j=1</td><td colspan="4">0.28</td><td colspan="4">0.88</td></tr>
<tr><td>j=2</td><td colspan="2">0.19</td><td colspan="2">0.19</td><td colspan="2">0.72</td><td colspan="2">0.36</td></tr>
<tr><td>j=3</td><td>0.12</td><td>0.12</td><td>0.12</td><td>0.12</td><td>0.32</td><td>0.00</td><td>0.28</td><td>0.28</td></tr>
</table>

Optimal Orthonormal Transform: VIII

- replace cost of parent (0.72) with that of children (0.32)

<table>
<tr><td>$j=0$</td><td colspan="8">1.45</td></tr>
<tr><td>$j=1$</td><td colspan="4">0.28</td><td colspan="4">0.88</td></tr>
<tr><td>$j=2$</td><td colspan="2">0.19</td><td colspan="2">0.19</td><td colspan="2">0.32</td><td colspan="2">0.36</td></tr>
<tr><td>$j=3$</td><td>0.12</td><td>0.12</td><td>0.12</td><td>0.12</td><td>0.32</td><td>0.00</td><td>0.28</td><td>0.28</td></tr>
</table>

Optimal Orthonormal Transform: VIII

- compare $0.28 + 0.28 = 0.56$ to 0.36: parent cheaper!

<table>
<tr><td>j=0</td><td colspan="8">1.45</td></tr>
<tr><td>j=1</td><td colspan="4">0.28</td><td colspan="4">0.88</td></tr>
<tr><td>j=2</td><td colspan="2">**0.19**</td><td colspan="2">**0.19**</td><td colspan="2">0.32</td><td colspan="2">0.36</td></tr>
<tr><td>j=3</td><td>**0.12**</td><td>**0.12**</td><td>**0.12**</td><td>**0.12**</td><td>**0.32**</td><td>**0.00**</td><td>**0.28**</td><td>**0.28**</td></tr>
</table>

Optimal Orthonormal Transform: VIII

- mark parent node

<table>
<tr><td>j=0</td><td colspan="8">1.45</td></tr>
<tr><td>j=1</td><td colspan="4">0.28</td><td colspan="4">0.88</td></tr>
<tr><td>j=2</td><td colspan="2">0.19</td><td colspan="2">0.19</td><td colspan="2">0.32</td><td colspan="2">0.36</td></tr>
<tr><td>j=3</td><td>0.12</td><td>0.12</td><td>0.12</td><td>0.12</td><td>0.32</td><td>0.00</td><td>0.28</td><td>0.28</td></tr>
</table>

Optimal Orthonormal Transform: VIII

- compare $0.19 + 0.19 = 0.38$ to 0.28: parent cheaper!

<table>
<tr><td>j=0</td><td colspan="8">1.45</td></tr>
<tr><td>j=1</td><td colspan="4">0.28</td><td colspan="4">0.88</td></tr>
<tr><td>j=2</td><td colspan="2">**0.19**</td><td colspan="2">**0.19**</td><td colspan="2">0.32</td><td colspan="2">**0.36**</td></tr>
<tr><td>j=3</td><td>**0.12**</td><td>**0.12**</td><td>**0.12**</td><td>**0.12**</td><td>**0.32**</td><td>**0.00**</td><td>**0.28**</td><td>**0.28**</td></tr>
</table>

Optimal Orthonormal Transform: VIII

- mark parent node

<table>
<tr><td>j=0</td><td colspan="8">1.45</td></tr>
<tr><td>j=1</td><td colspan="4">**0.28**</td><td colspan="4">0.88</td></tr>
<tr><td>j=2</td><td colspan="2">0.19</td><td colspan="2">0.19</td><td colspan="2">0.32</td><td colspan="2">**0.36**</td></tr>
<tr><td>j=3</td><td>**0.12**</td><td>**0.12**</td><td>**0.12**</td><td>**0.12**</td><td>**0.32**</td><td>**0.00**</td><td>**0.28**</td><td>**0.28**</td></tr>
</table>

Optimal Orthonormal Transform: VIII

- compare $0.32 + 0.36 = 0.68$ to 0.88: children cheaper!

j=0	1.45							
j=1	**0.28**				0.88			
j=2	0.19		0.19		0.32		**0.36**	
j=3	**0.12**	**0.12**	**0.12**	**0.12**	**0.32**	**0.00**	**0.28**	**0.28**

Optimal Orthonormal Transform: VIII

- replace cost of parent (0.88) with that of children (0.68)

<table>
<tr><td>j=0</td><td colspan="8">1.45</td></tr>
<tr><td>j=1</td><td colspan="4">0.28</td><td colspan="4">0.68</td></tr>
<tr><td>j=2</td><td colspan="2">0.19</td><td colspan="2">0.19</td><td colspan="2">0.32</td><td colspan="2">0.36</td></tr>
<tr><td>j=3</td><td>0.12</td><td>0.12</td><td>0.12</td><td>0.12</td><td>0.32</td><td>0.00</td><td>0.28</td><td>0.28</td></tr>
</table>

Optimal Orthonormal Transform: VIII

- compare 0.28 + 0.68 = 0.96 to 1.45: children cheaper!

j=0	1.45							
j=1	**0.28**				0.68			
j=2	0.19		0.19		0.32		**0.36**	
j=3	**0.12**	**0.12**	**0.12**	**0.12**	**0.32**	**0.00**	**0.28**	**0.28**

Optimal Orthonormal Transform: VIII

- replace cost of parent (1.45) with that of children (0.96)

<table>
<tr><td>$j=0$</td><td colspan="8">0.96</td></tr>
<tr><td>$j=1$</td><td colspan="4">**0.28**</td><td colspan="4">0.68</td></tr>
<tr><td>$j=2$</td><td colspan="2">0.19</td><td colspan="2">0.19</td><td colspan="2">0.32</td><td colspan="2">**0.36**</td></tr>
<tr><td>$j=3$</td><td>**0.12**</td><td>**0.12**</td><td>**0.12**</td><td>**0.12**</td><td>**0.32**</td><td>**0.00**</td><td>**0.28**</td><td>**0.28**</td></tr>
</table>

Optimal Orthonormal Transform: VIII

- final step (best basis includes 3 vectors forming **X**):

<table>
<tr><td>j=0</td><td colspan="8">0.96</td></tr>
<tr><td>j=1</td><td colspan="4">0.28</td><td colspan="4">0.68</td></tr>
<tr><td>j=2</td><td colspan="2">0.19</td><td colspan="2">0.19</td><td colspan="2">0.32</td><td colspan="2">0.36</td></tr>
<tr><td>j=3</td><td>0.12</td><td>0.12</td><td>0.12</td><td>0.12</td><td>0.32</td><td>0.00</td><td>0.28</td><td>0.28</td></tr>
</table>

Time Shifts for Wavelet Packet Filters: I

- use of LA or coiflet filters allows time alignment
- equivalent filter yielding $\mathbf{W}_{j,n}$ is $\{u_{j,n,l}\}$
- can obtain shifts $\nu_{j,n}$ via study of phase function
- transfer function for $\{u_{j,n,l}\}$ given by

$$U_{j,n}(f) = \prod_{m=0}^{j-1} M_{c_{j,n,m}}(2^m f) = \prod_{m=0}^{j-1} |M_{c_{j,n,m}}(2^m f)| e^{i\theta_{c_{j,n,m}}(2^m f)}$$

so phase function is

$$\sum_{m=0}^{j-1} \theta_{c_{j,n,m}}(2^m f)$$

Time Shifts for Wavelet Packet Filters: II

- using

$$\theta^{(G)}(f) \approx 2\pi f \nu \text{ and } \theta^{(H)}(f) \approx -2\pi f(L - 1 + \nu),$$

can get

$$\nu_{j,n} \equiv \nu(2^j - 1) - S_{j,n,1}(2\nu + L - 1),$$

where

$$S_{j,n,1} \equiv \sum_{m=0}^{j-1} c_{j,n,m} 2^m$$

- Table 230 gives example of computing $\nu_{j,n}$ for LA(8)

Maximal Overlap DWPT: I

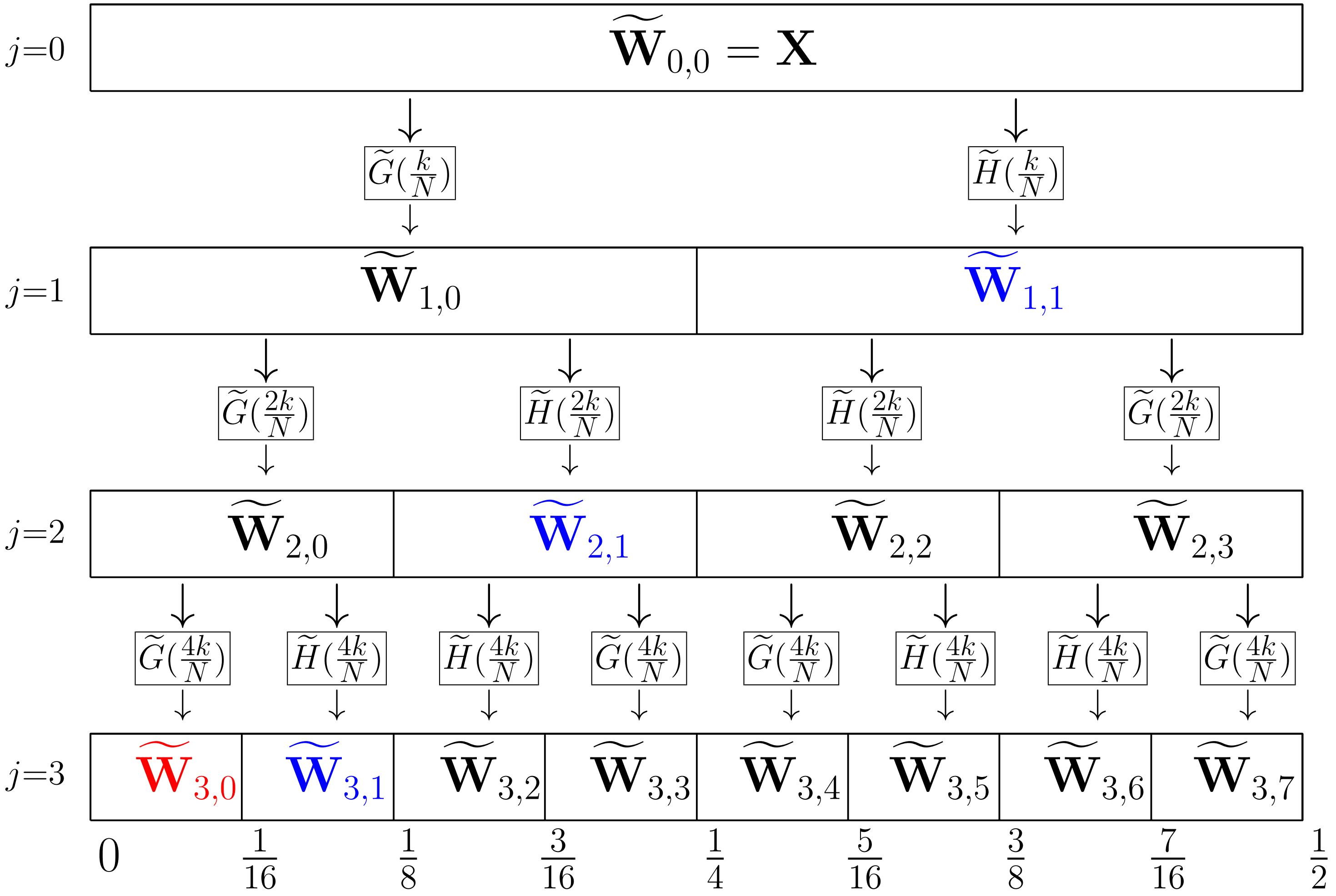
- recall relationship between DWT and MODWT
- MODWT: no downsampling and hence 'shift invariant'
- uses MODWT filters: $\tilde{h}_l \equiv h_l/\sqrt{2}$ and $\tilde{g}_l \equiv g_l/\sqrt{2}$
- level J_0 MODWT maps $\mathbf{X}$ to $J_0 + 1$ vectors $\widetilde{\mathbf{W}}_1$, $\widetilde{\mathbf{W}}_2$, $\ldots$, $\widetilde{\mathbf{W}}_{J_0}$, $\widetilde{\mathbf{V}}_{J_0}$, all of length N (arbitrary)
- with LA wavelet, can align (time shift) using $\mathcal{T}^{\nu_j}\widetilde{\mathbf{W}}_j$
- MODWT multiresolution analysis and analysis of variance:

$$\mathbf{X} = \sum_{j=1}^{J_0} \widetilde{\mathcal{D}}_j + \widetilde{\mathcal{S}}_{J_0} \text{ and } \|\mathbf{X}\|^2 = \sum_{j=1}^{J_0} \|\widetilde{\mathbf{W}}_j\|^2 + \|\widetilde{\mathbf{V}}_{J_0}\|^2$$

- $\widetilde{\mathcal{D}}_j$ is output from zero phase filter

Maximal Overlap DWPT: II

- similarly, can generalize DWPT to MODWPT



$j=0$ $\widetilde{\mathbf{W}}_{0,0} = \mathbf{X}$

$\widetilde{G}(\frac{k}{N})$ $\widetilde{H}(\frac{k}{N})$

$j=1$ $\widetilde{\mathbf{W}}_{1,0}$ $\widetilde{\mathbf{W}}_{1,1}$

$\widetilde{G}(\frac{2k}{N})$ $\widetilde{H}(\frac{2k}{N})$ $\widetilde{H}(\frac{2k}{N})$ $\widetilde{G}(\frac{2k}{N})$

$j=2$ $\widetilde{\mathbf{W}}_{2,0}$ $\widetilde{\mathbf{W}}_{2,1}$ $\widetilde{\mathbf{W}}_{2,2}$ $\widetilde{\mathbf{W}}_{2,3}$

$\widetilde{G}(\frac{4k}{N})$ $\widetilde{H}(\frac{4k}{N})$ $\widetilde{H}(\frac{4k}{N})$ $\widetilde{G}(\frac{4k}{N})$ $\widetilde{G}(\frac{4k}{N})$ $\widetilde{H}(\frac{4k}{N})$ $\widetilde{H}(\frac{4k}{N})$ $\widetilde{G}(\frac{4k}{N})$

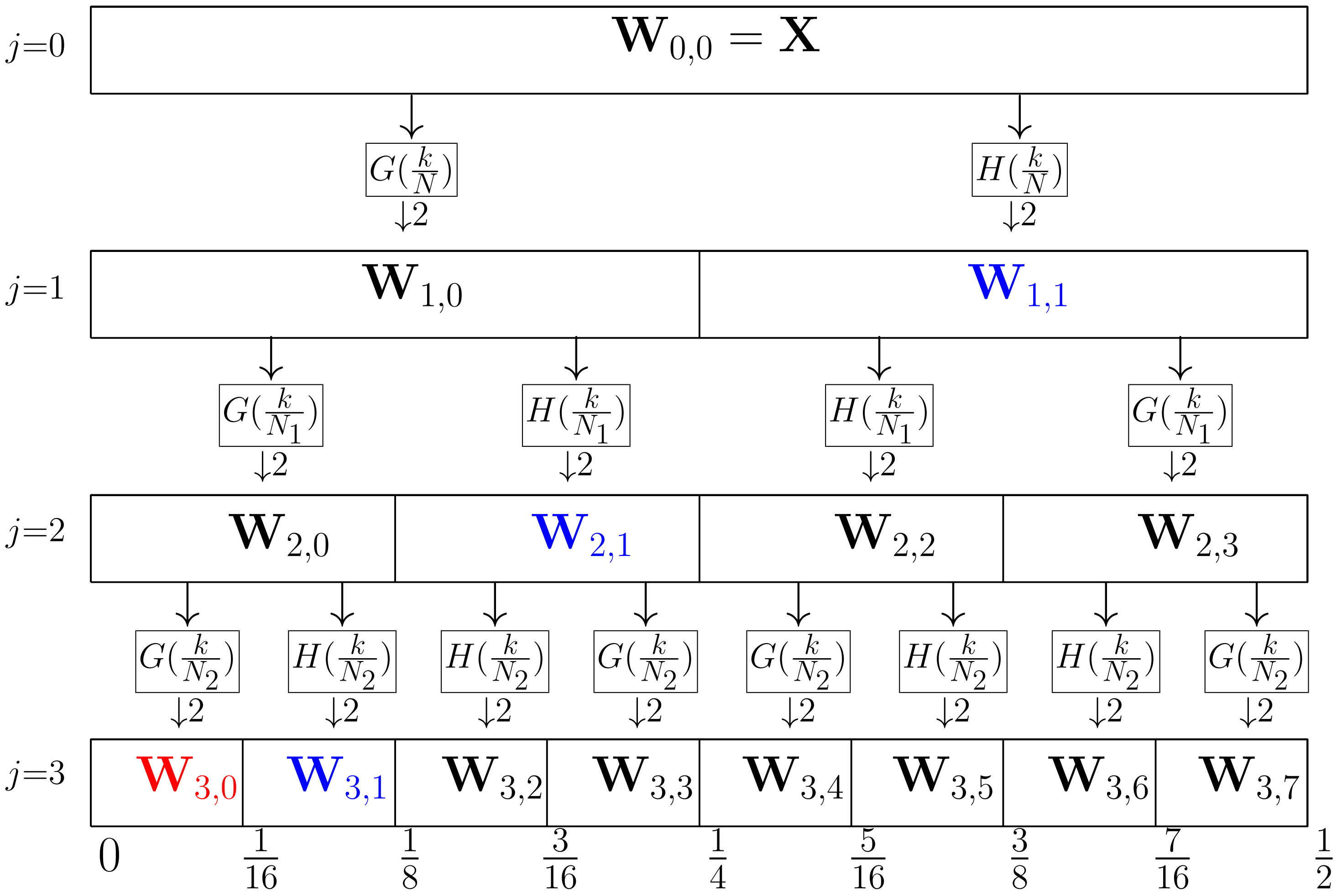
$j=3$ $\widetilde{\mathbf{W}}_{3,0}$ $\widetilde{\mathbf{W}}_{3,1}$ $\widetilde{\mathbf{W}}_{3,2}$ $\widetilde{\mathbf{W}}_{3,3}$ $\widetilde{\mathbf{W}}_{3,4}$ $\widetilde{\mathbf{W}}_{3,5}$ $\widetilde{\mathbf{W}}_{3,6}$ $\widetilde{\mathbf{W}}_{3,7}$

0 $\frac{1}{16}$ $\frac{1}{8}$ $\frac{3}{16}$ $\frac{1}{4}$ $\frac{5}{16}$ $\frac{3}{8}$ $\frac{7}{16}$ $\frac{1}{2}$

WMTSA: 231–234

DWPTs of General Levels: II

- example of rule, yielding level $j = 3$ DWPT in the bottom row



Maximal Overlap DWPT: III

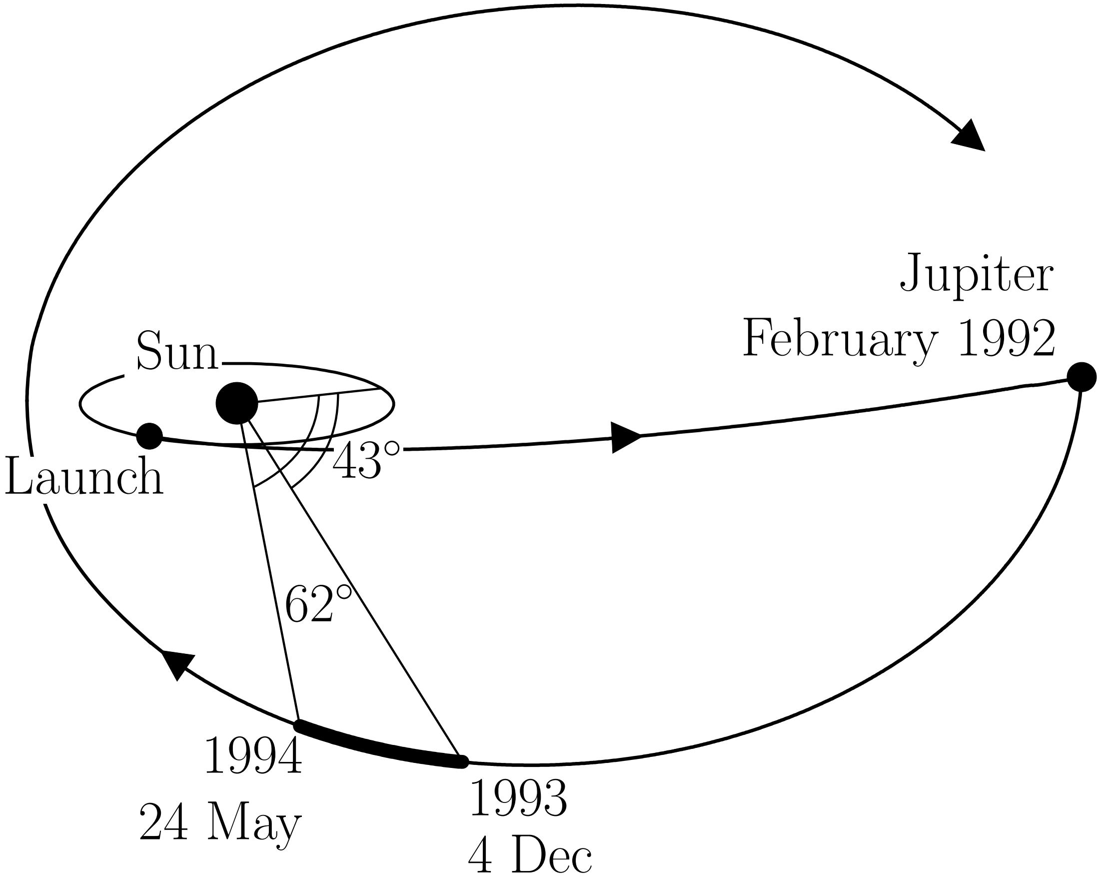
- uses renormalized DWPT filters
- every $\widetilde{\mathbf{W}}_{j,n}$ is now a vector of length N
- with LA wavelet, can align using $\mathcal{T}^{\nu_{j,n}}\widetilde{\mathbf{W}}_{j,n}$
- let $\mathcal{C}$ be indices for disjoint dyadic decomposition
- MODWPT additive decomposition and analysis of variance:

$$\mathbf{X} = \sum_{(j,n)\in\mathcal{C}} \widetilde{\mathcal{D}}_{j,n} \text{ and } \|\mathbf{X}\|^2 = \sum_{(j,n)\in\mathcal{C}} \|\widetilde{\mathbf{W}}_{j,n}\|^2$$

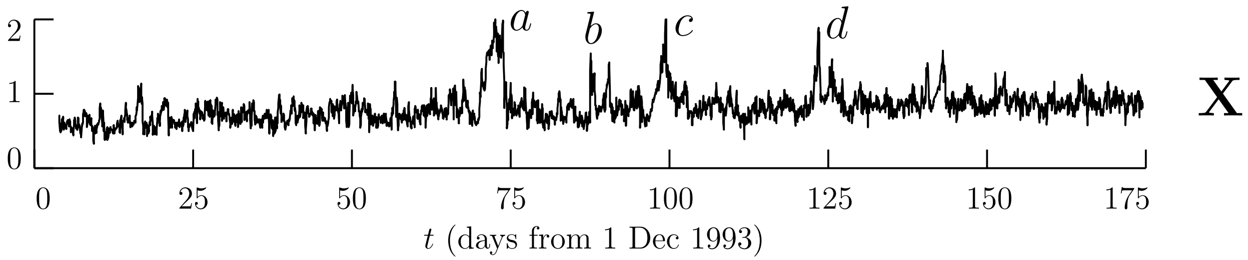
- $\widetilde{\mathcal{D}}_{j,n}$ is analogous to MODWT detail (and is created by applying inverse MODWPT to $\widetilde{\mathbf{W}}_{j,n}$ and vectors of zeros)
- $\widetilde{\mathcal{D}}_{j,n}$ is output from zero phase filter

Example – Analysis of Solar Physics Data: I

- path of Ulysses spacecraft (records magnetic field of heliosphere)

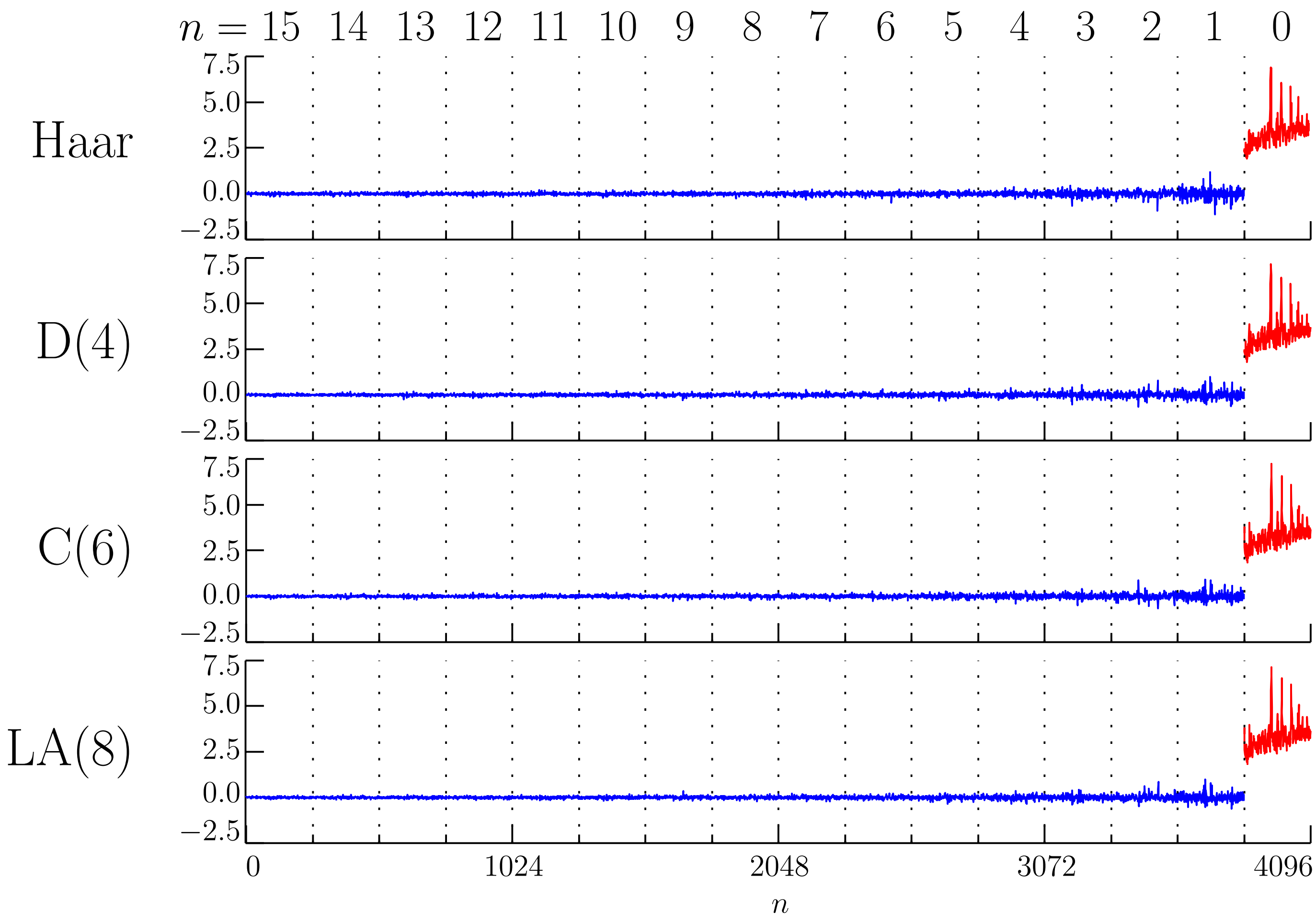


Example – Analysis of Solar Physics Data: II



- magnetic field measurements of polar region of sun recorded hourly from 4 Dec 1993 to 24 May 1994 ($\Delta t = 1/24$ day)
- Ulysses moved from 4 AU to 3 AU (explains upward trend)
- a, b, c, d are fast solar wind streams from polar coronal holes
- two classifications for these 'shocks'
  - corotating interaction regions (CIRs) – recur every solar rotation (about 25 days)
  - fast coronal mass ejections (CMEs) – transient in nature

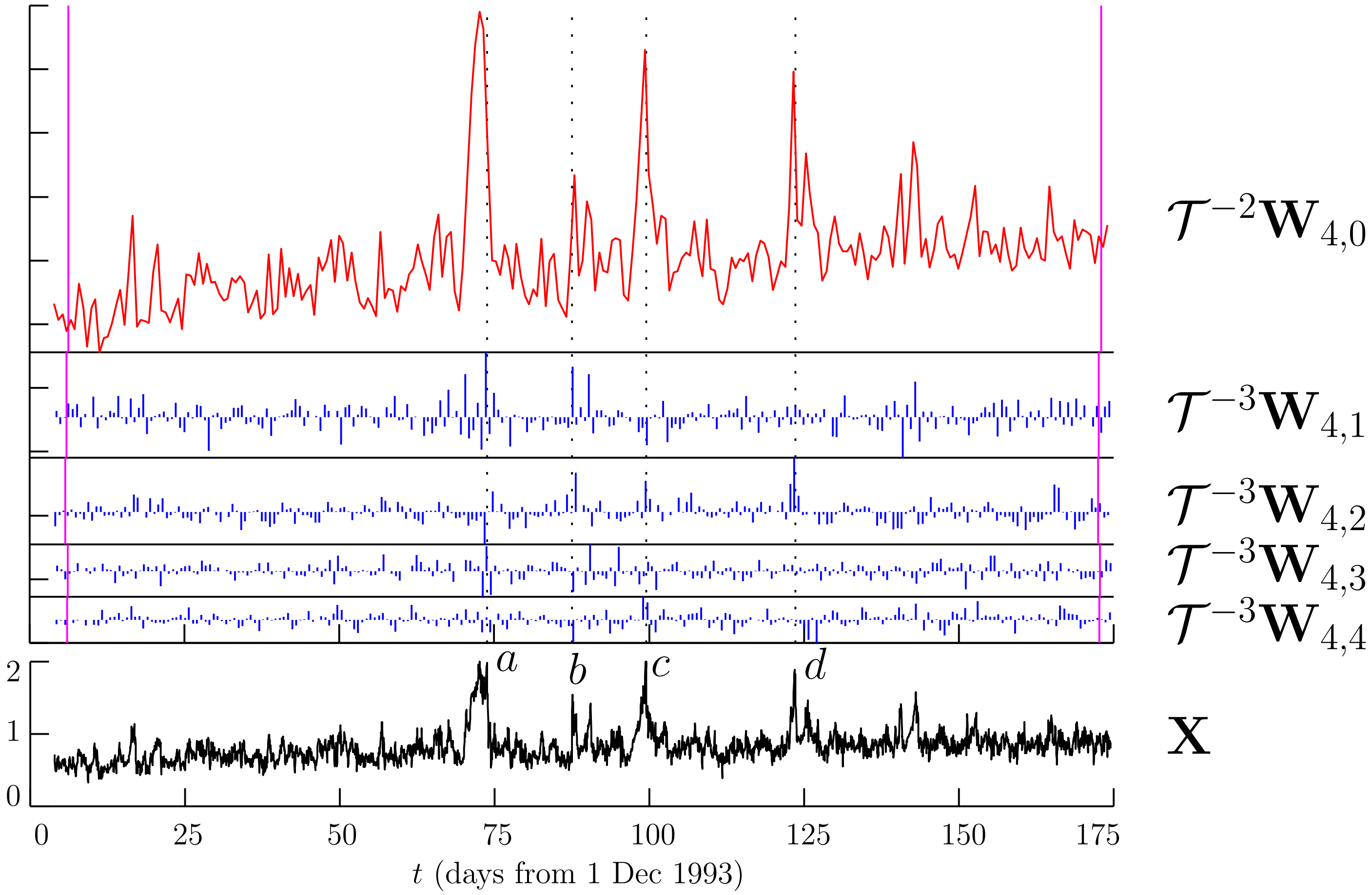
Example – Analysis of Solar Physics Data: III



- 4 different level $j = 4$ DWPTs, each partitioning $(0, 1/2\,\Delta t]$ into 16 intervals

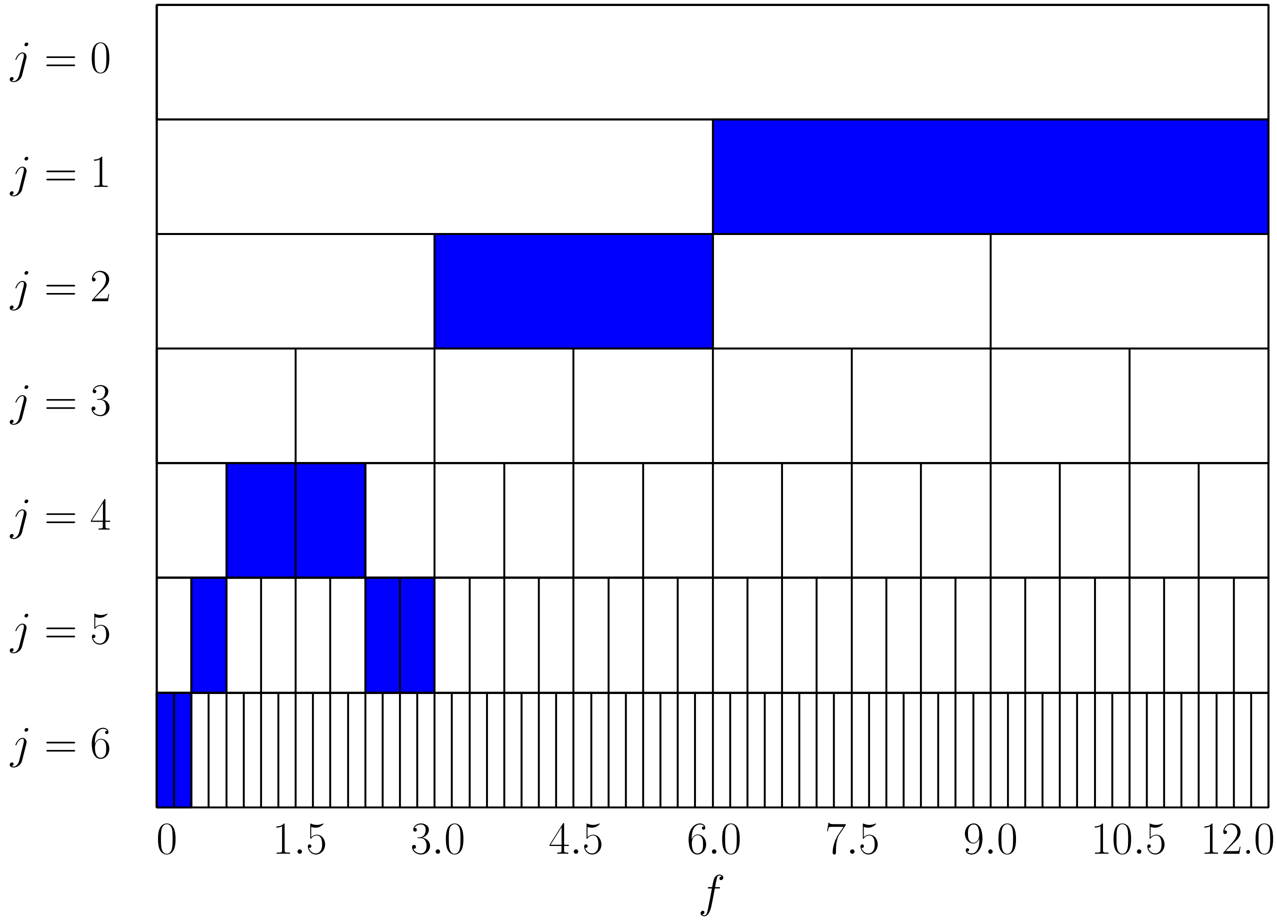
WMTSA: 218–221

Example – Analysis of Solar Physics Data: IV



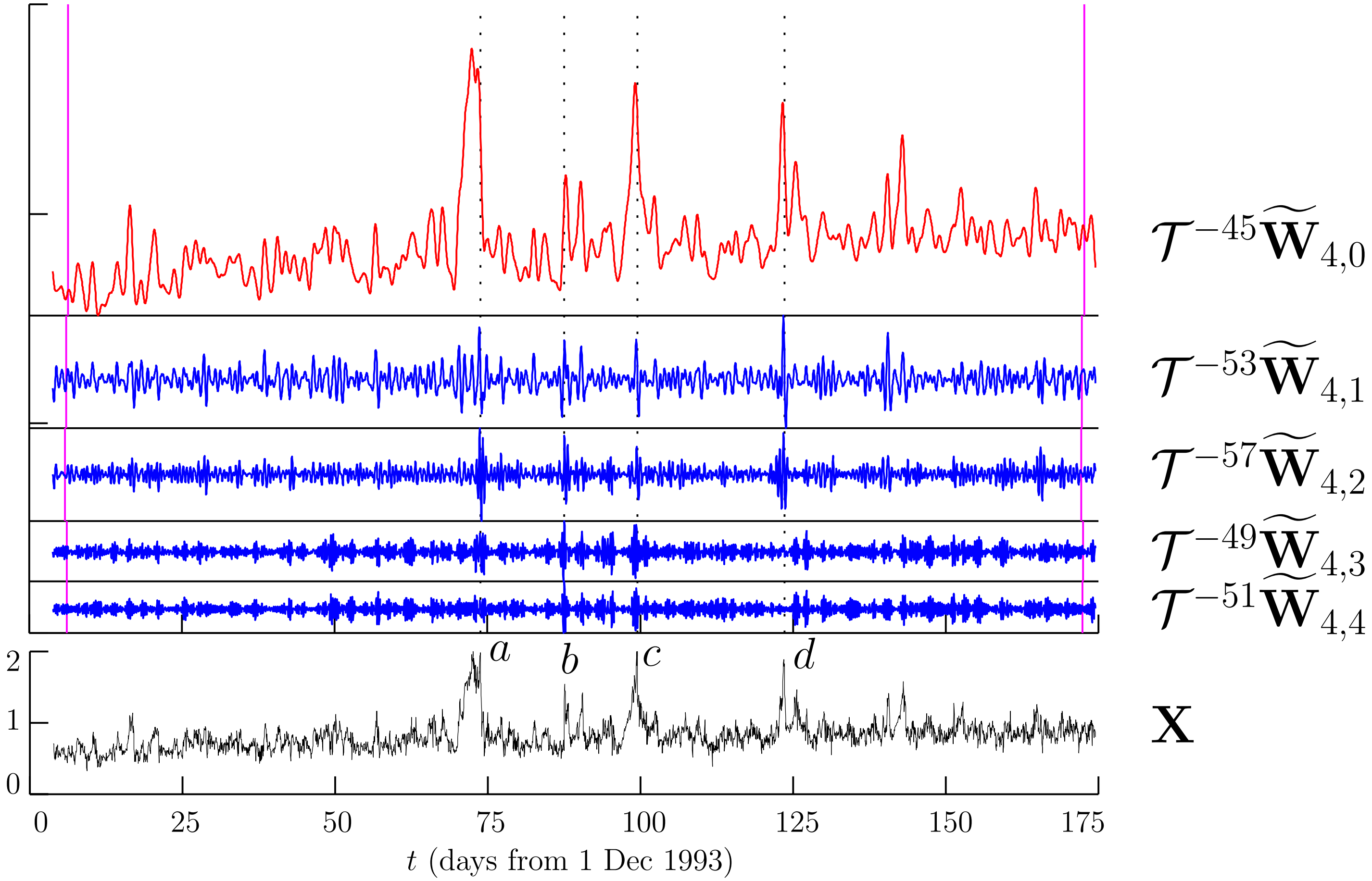
- level $j = 4$ LA(8) DWPT coefficients $\mathbf{W}_{4,n}$, $n = 0, \ldots, 4$, after time alignments

Example – Analysis of Solar Physics Data: V



- best basis transform using LA(8) filter and $-W_{j,n,t}^2 \log(W_{j,n,t}^2)$ cost function

Example – Analysis of Solar Physics Data: VI

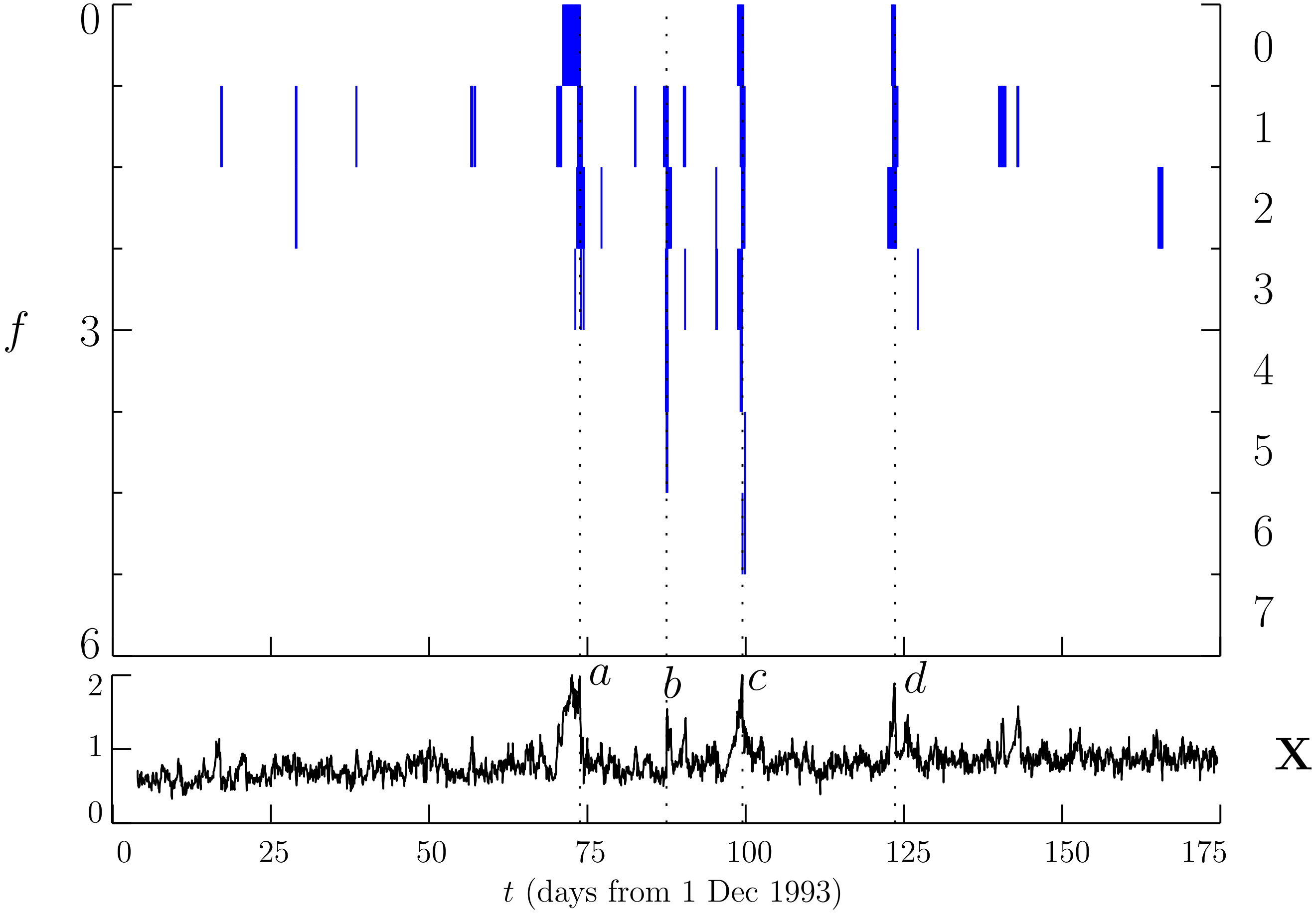


- level $j = 4$ LA(8) MODWPT coefficients $\mathbf{W}_{4,n}$, $n = 0, \ldots, 4$, after time alignments

Example – Analysis of Solar Physics Data: VII

- will summarize using a modified time/frequency plot, which indicates locations of
  - 100 largest values in $\mathcal{T}^{-|\nu_{4,0}|}\widetilde{\mathbf{W}}_{4,0}$
  - 100 largest values in $\mathcal{T}^{-|\nu_{4,1}|}\widetilde{\mathbf{W}}_{4,1}$
  - 100 largest values in $\mathcal{T}^{-|\nu_{4,n}|}\widetilde{\mathbf{W}}_{4,n}$, $n = 2, \ldots, 15$ (in fact these all occur in $n = 2, \ldots, 6$)

Example – Analysis of Solar Physics Data: VIII



- 4 events coherently broad-band; events a, c, d are recurrent; b is transient; a might be two events (recurrent & transient)