The Discrete Fourier Transform and the Short-time Fourier Transform

Xavier Serra
Music Technology Group
Universitat Pompeu Fabra, Barcelona

Discrete Fourier Transform

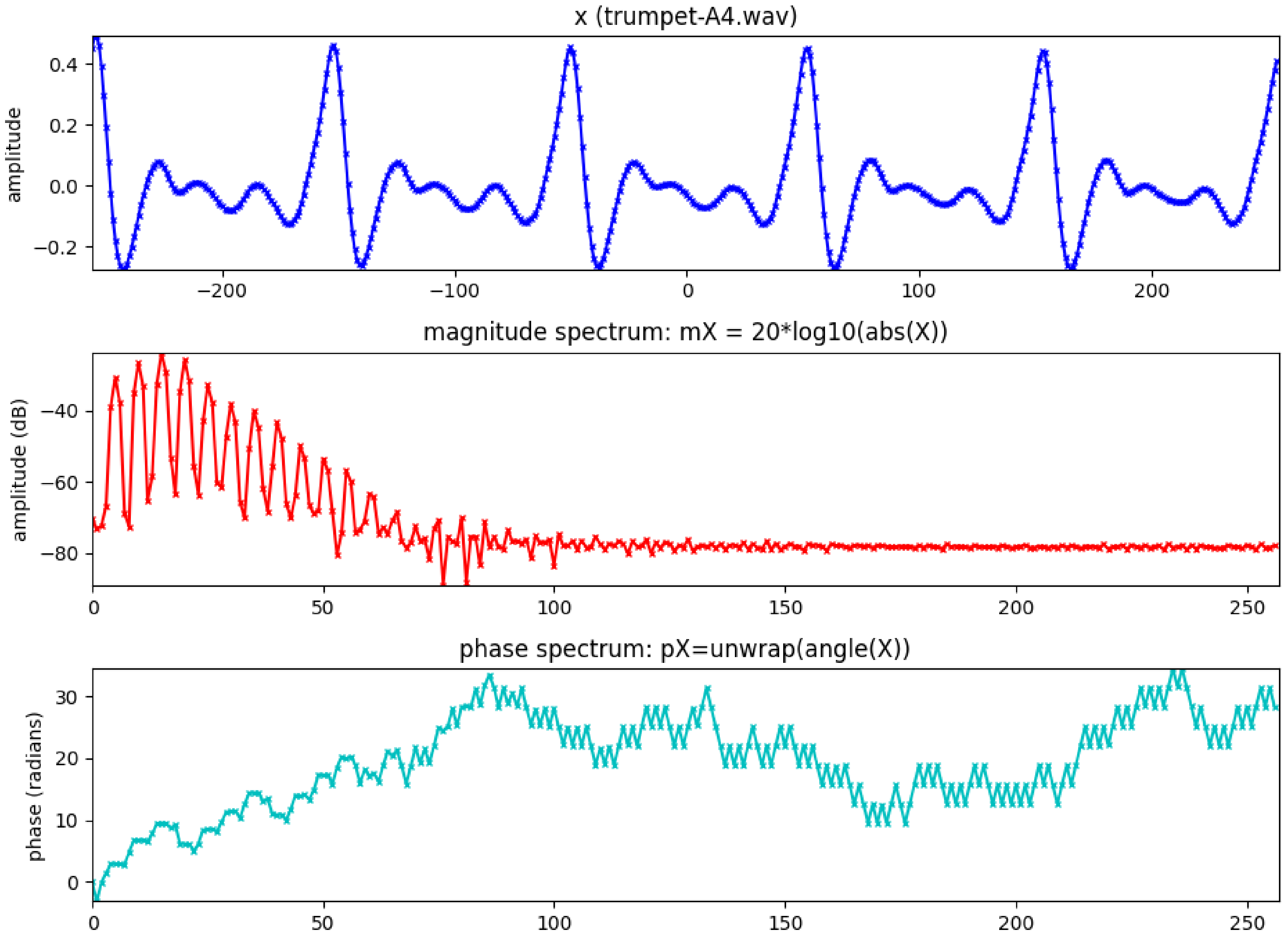
$$X[k]=\sum_{n=0}^{N-1} x[n]e^{-j2\pi kn/N} \quad k=0,\ldots,N-1$$

n : discrete time index (normalized time, $T=1$)

k : discrete frequency index

$\omega_k = 2\pi k/N$: frequency in radians

$f_k = f_s k/N$: frequency in Hz (f_s : sampling rate)



x (trumpet-A4.wav)
amplitude
0.4
0.2
0.0
−0.2
−200
−100
0
100
200
magnitude spectrum: mX = 20*log10(abs(X))
amplitude (dB)
−40
−60
−80
0
50
100
150
200
250
phase spectrum: pX=unwrap(angle(X))
phase (radians)
30
20
10
0
0
50
100
150
200
250

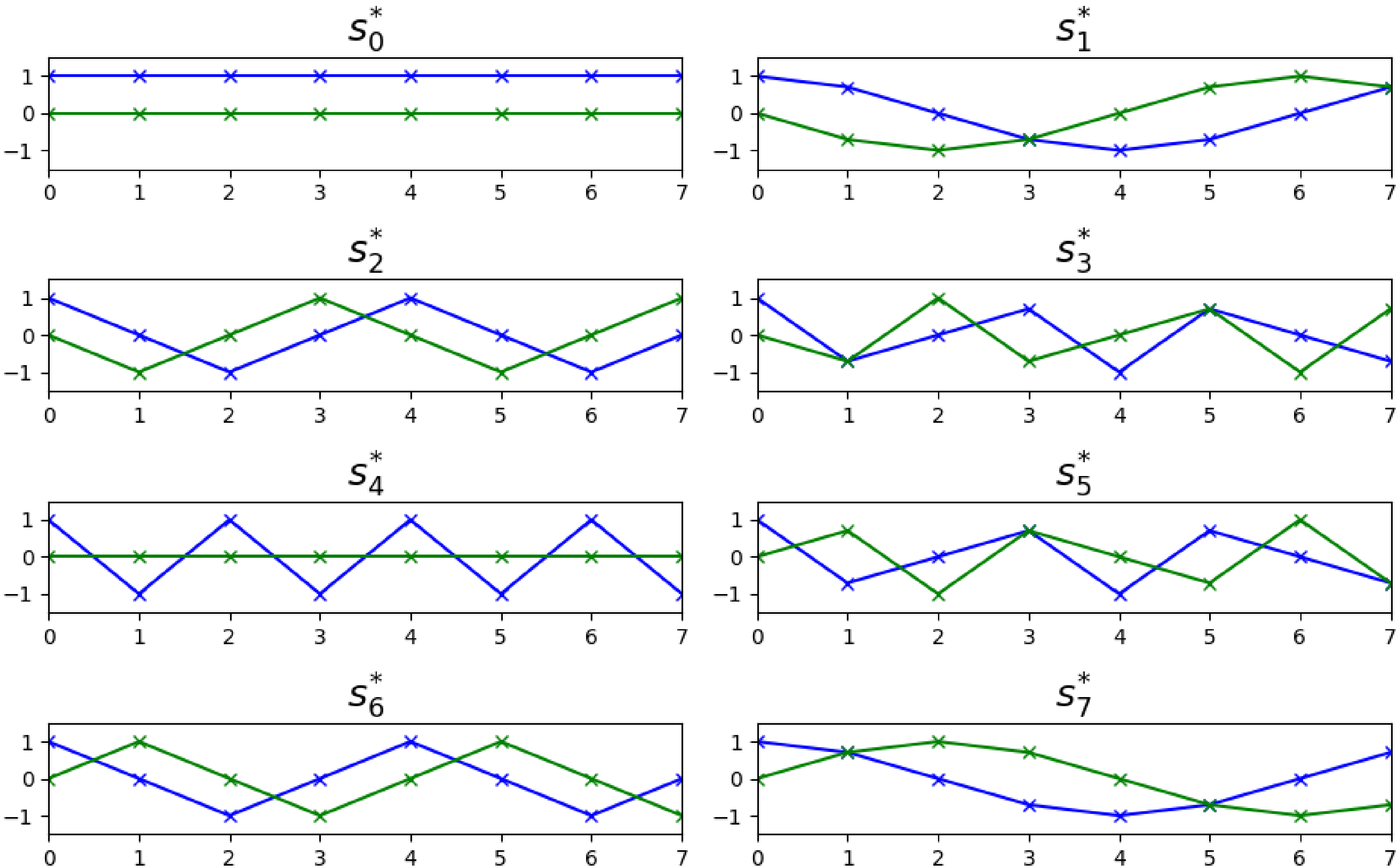
DFT: complex exponentials

$$s_k^* = e^{-j2\pi kn/N} = \cos(2\pi kn/N) - j\sin(2\pi kn/N)$$

for $N = 4$, thus for $n = 0,1,2,3;\ \ k = 0,1,2,3$

$$s_0^* = \cos(2\pi \times 0 \times n/4) - j\sin(2\pi \times 0 \times n/4) = [1,1,1,1]$$
$$s_1^* = \cos(2\pi \times 1 \times n/4) - j\sin(2\pi \times 1 \times n/4) = [1,-j,-1,j]$$
$$s_2^* = \cos(2\pi \times 2 \times n/4) - j\sin(2\pi \times 2 \times n/4) = [1,-1,1,-1]$$
$$s_3^* = \cos(2\pi \times 3 \times n/4) - j\sin(2\pi \times 3 \times n/4) = [1,j,-1,-j]$$

DFT: complex exponentials



DFT: scalar product

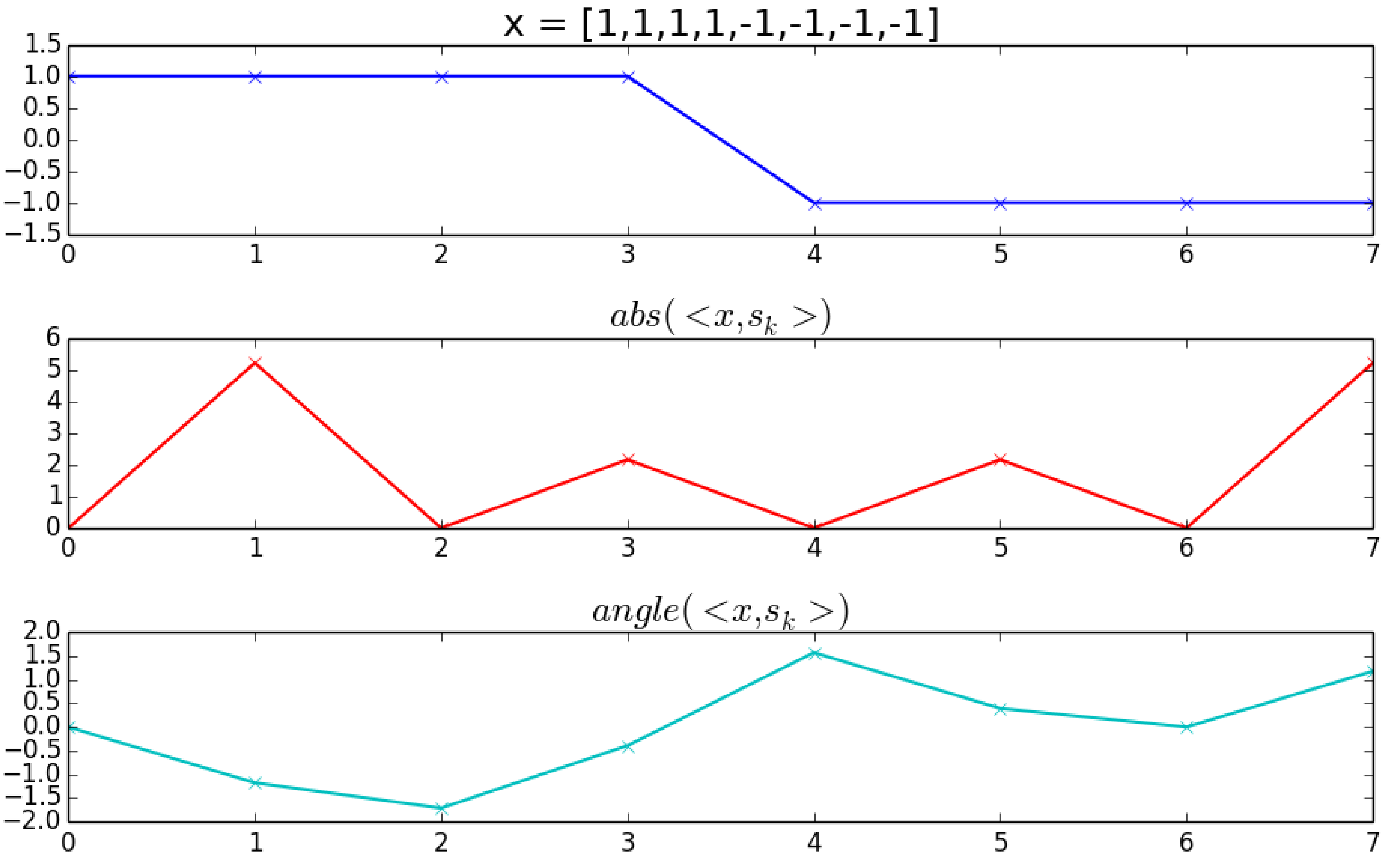
$$\langle x, s_k \rangle = \sum_{n=0}^{N-1} x[n] s_k^*[n] = \sum_{n=0}^{N-1} x[n] e^{-j2\pi kn/N}$$

Example:

$$x[n] = [1, -1, 1, -1];\ N = 4$$

$$\langle x, s_0 \rangle = 1 \times 1 + (-1) \times 1 + 1 \times 1 + (-1) \times 1 = 0$$
$$\langle x, s_1 \rangle = 1 \times 1 + (-1) \times (-j) + 1 \times (-1) + (-1) \times j = 0$$
$$\langle x, s_2 \rangle = 1 \times 1 + (-1) \times (-1) + 1 \times 1 + (-1) \times (-1) = 4$$
$$\langle x, s_3 \rangle = 1 \times 1 + (-1) \times j + 1 \times (-1) + (-1) \times (-j) = 0$$

DFT: scalar product



DFT of real sinusoids

$$x_3[n] = A_0 \cos(2\pi k_0 n/N) = \frac{A_0}{2} e^{j2\pi k_0 n/N} + \frac{A_0}{2} e^{-j2\pi k_0 n/N}$$

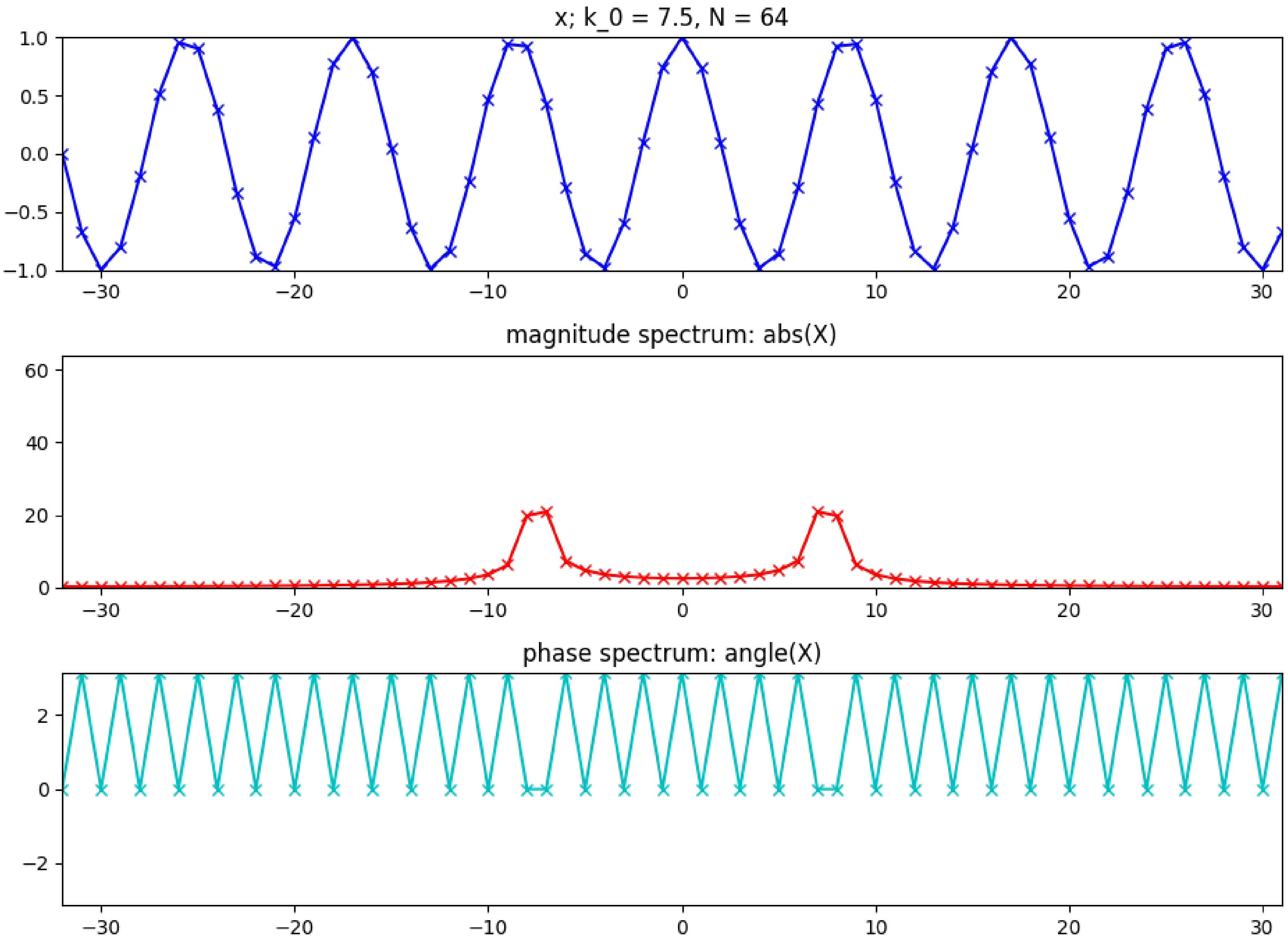
$$X_3[k] = \sum_{n=-N/2}^{N/2-1} x_3[n] e^{-j2\pi kn/N}$$

$$= \sum_{n=-N/2}^{N/2-1} \left(\frac{A_0}{2} e^{j2\pi k_0 n/N} + \frac{A_0}{2} e^{-j2\pi k_0 n/N}\right) e^{-j2\pi kn/N}$$

$$= \sum_{n=-N/2}^{N/2-1} \frac{A_0}{2} e^{j2\pi k_0 n/N} e^{-j2\pi kn/N} + \sum_{n=-N/2}^{N/2-1} \frac{A_0}{2} e^{-j2\pi k_0 n/N} e^{-j2\pi kn/N}$$

$$= \sum_{n=-N/2}^{N/2-1} \frac{A_0}{2} e^{-j2\pi (k-k_0) n/N} + \sum_{n=-N/2}^{N/2-1} \frac{A_0}{2} e^{-j2\pi (k+k_0) n/N}$$

$$= N\frac{A_0}{2} \text{ for } k = k_{0,} -k_0 \text{; } 0 \text{ for rest of } k$$



x; k_0 = 7.5, N = 64
1.0
0.5
0.0
−0.5
−1.0
−30
−20
−10
0
10
20
30
magnitude spectrum: abs(X)
60
40
20
0
phase spectrum: angle(X)
2
0
−2

Inverse DFT

$$x[n]=\frac{1}{N}\sum_{k=0}^{N-1} X[k]s_k[n]=\frac{1}{N}\sum_{k=0}^{N-1} X[k]e^{j2\pi kn/N} \quad n=0,1,\ldots,N-1$$

Example:

$X[k]=[0,4,0,0];N=4$

$$x[0]=\frac{1}{4}(X*s)[n=0]=\frac{1}{4}(0*1+4*1+0*1+0*1)=1$$

$$x[1]=\frac{1}{4}(X*s)[n=1]=\frac{1}{4}(0*1+4*j+0*(-1)+0*(-j))=j$$

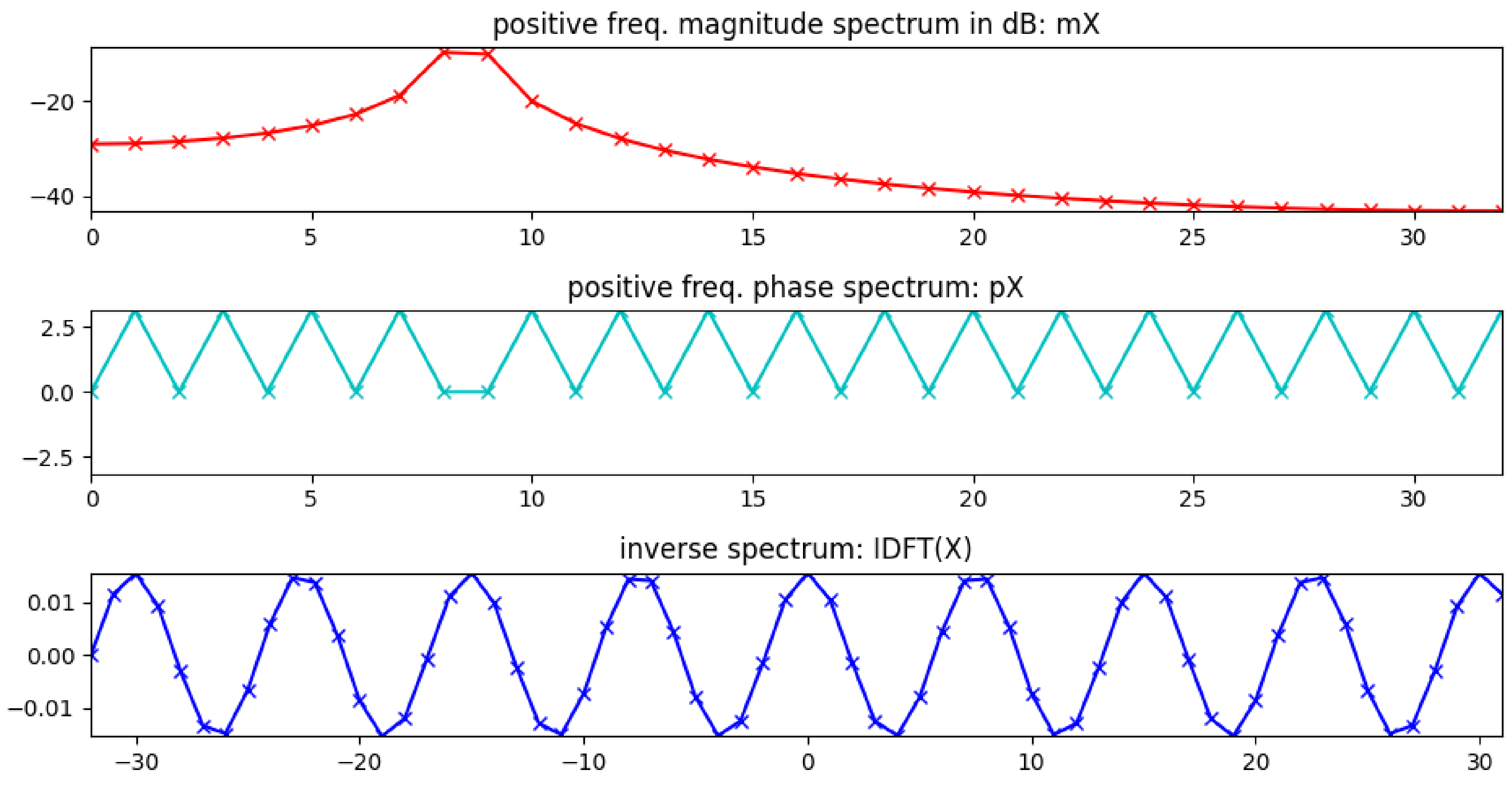
$$x[2]=\frac{1}{4}(X*s)[n=2]=\frac{1}{4}(0*1+4*(-1)+0*1+0*(-1))=-1$$

$$x[3]=\frac{1}{4}(X*s)[n=3]=\frac{1}{4}(0*1+4*(-j)+0*(-1)+0*j)=-j$$

Inverse DFT for real signals

$$X[k]=|X[k]|e^{j<X[k]} \quad \text{and} \quad X[-k]=|X[k]|e^{-j<X[k]}$$

for $k=0,1,\ldots,N/2$

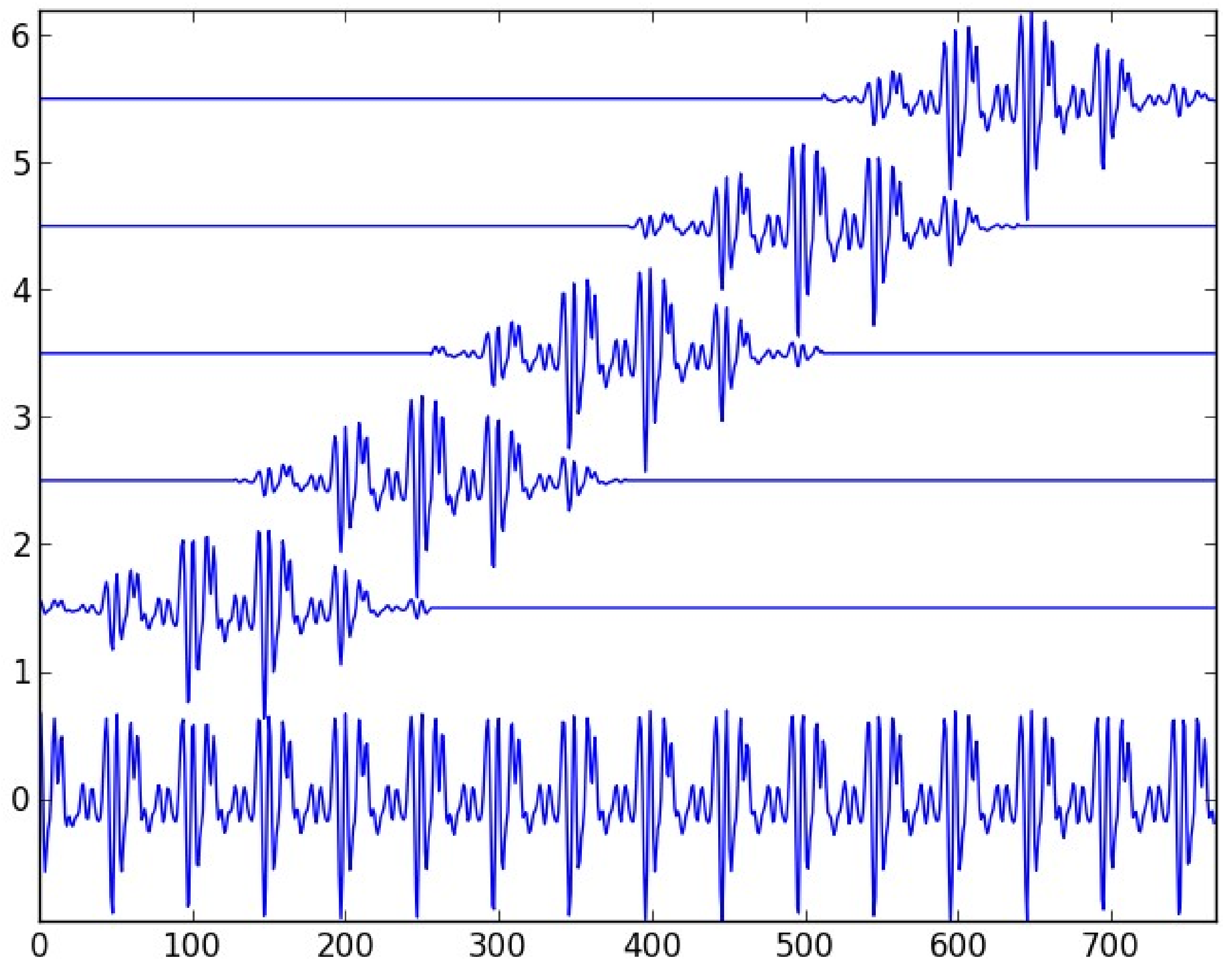


Short-time Fourier Transform

$$X_l[k] = \sum_{n=-N/2}^{N/2-1} w[n]\,x[n+lH]\,e^{-j2\pi kn/N} \qquad l = 0, 1, \ldots,$$

w: analysis window
l: frame number
H: hop-size

$$xw_l[n] = w[n]\,x[n + lH] \qquad l = 0, 1, \ldots,$$



Transform of a windowed sinewave

$$x[n] = A_0 \cos(2\pi k_0 n/N) = \frac{A_0}{2} e^{j2\pi k_0 n/N} + \frac{A_0}{2} e^{-j2\pi k_0 n/N}$$

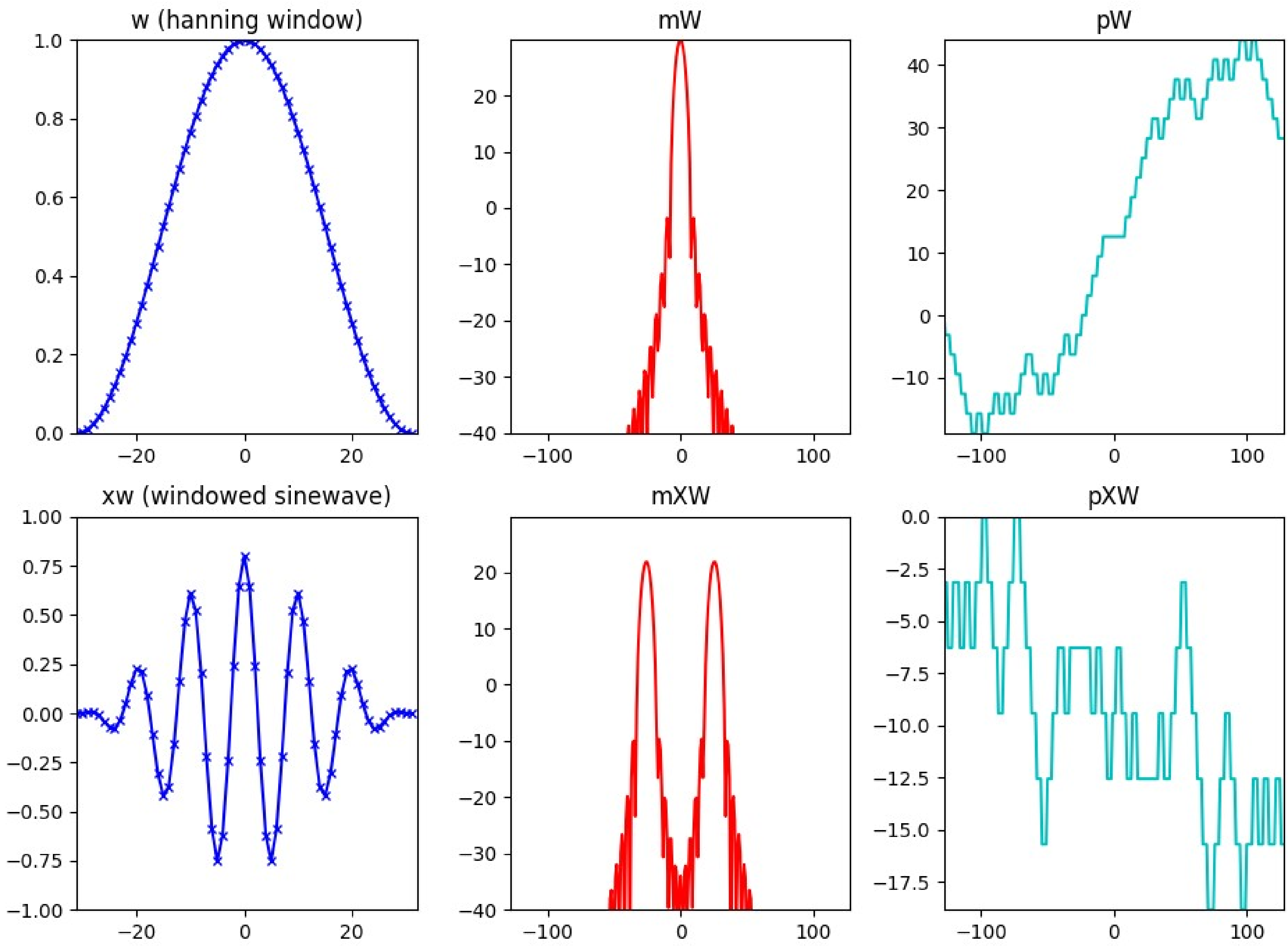
$$X[k] = \sum_{n=-N/2}^{N/2-1} w[n] x[n] e^{-j2\pi kn/N}$$

$$= \sum_{n=-N/2}^{N/2-1} w[n] \left(\frac{A_0}{2} e^{j2\pi k_0 n/N} + \frac{A_0}{2} e^{-j2\pi k_0 n/N} \right) e^{-j2\pi kn/N}$$

$$= \sum_{n=-N/2}^{N/2-1} w[n] \frac{A_0}{2} e^{j2\pi k_0 n/N} e^{-j2\pi kn/N} + \sum_{n=-N/2}^{N/2-1} w[n] \frac{A_0}{2} e^{-j2\pi k_0 n/N} e^{-j2\pi kn/N}$$

$$= \frac{A_0}{2} \sum_{n=-N/2}^{N/2-1} w[n] e^{-j2\pi(k-k_0)n/N} + \frac{A_0}{2} \sum_{n=-N/2}^{N/2-1} w[n] e^{-j2\pi(k+k_0)n/N}$$

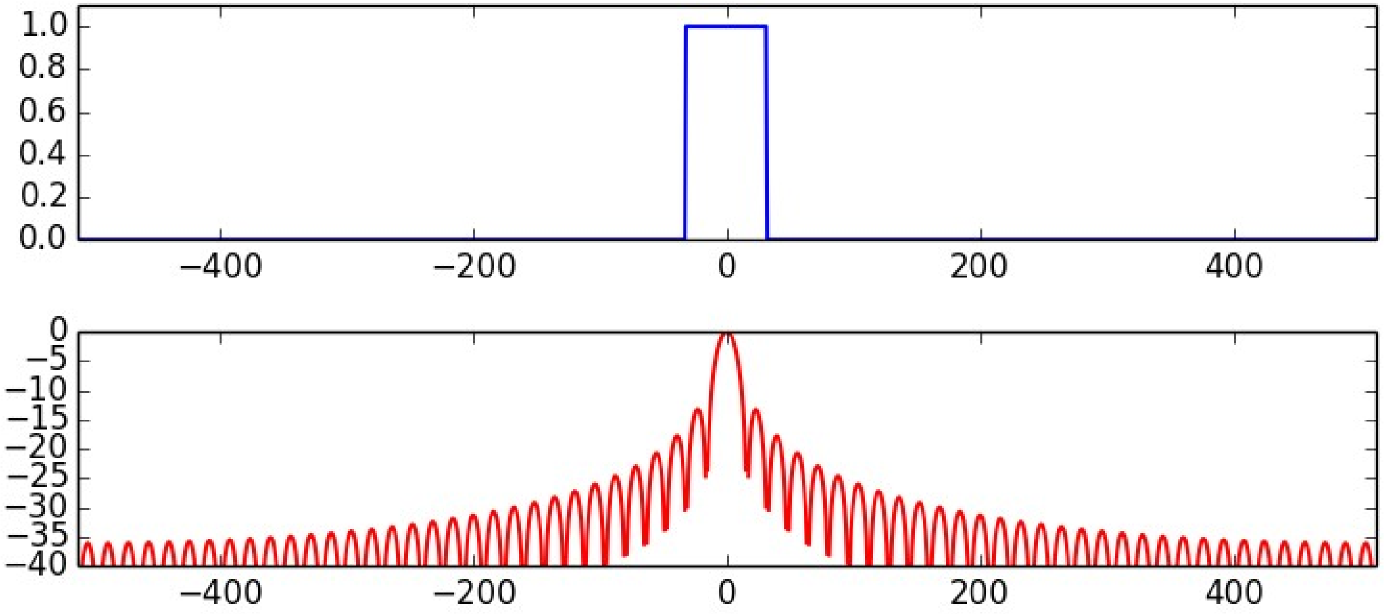
$$= \frac{A_0}{2} W[k-k_0] + \frac{A_0}{2} W[k+k_0]$$


w (hanning window)
mW
pW
xw (windowed sinewave)
mXW
pXW

Rectangular window

$$w[n] = 1, \quad n = -M/2, \ldots, 0, \ldots M/2$$
$$\quad\;\, = 0, \quad n = \text{elsewhere}$$

$$W[k] = \frac{\sin(\pi k)}{\sin(\pi k/M)}$$

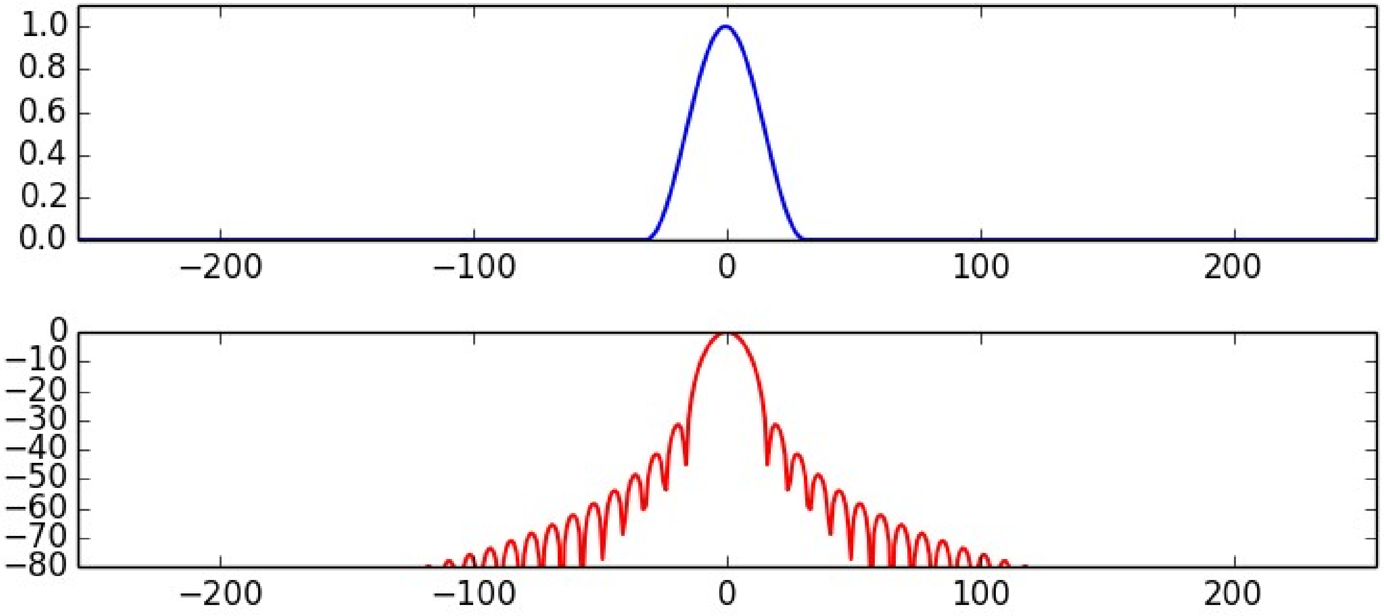


main-lobe width: 2 bins
side-lobe level: -13.3 dB

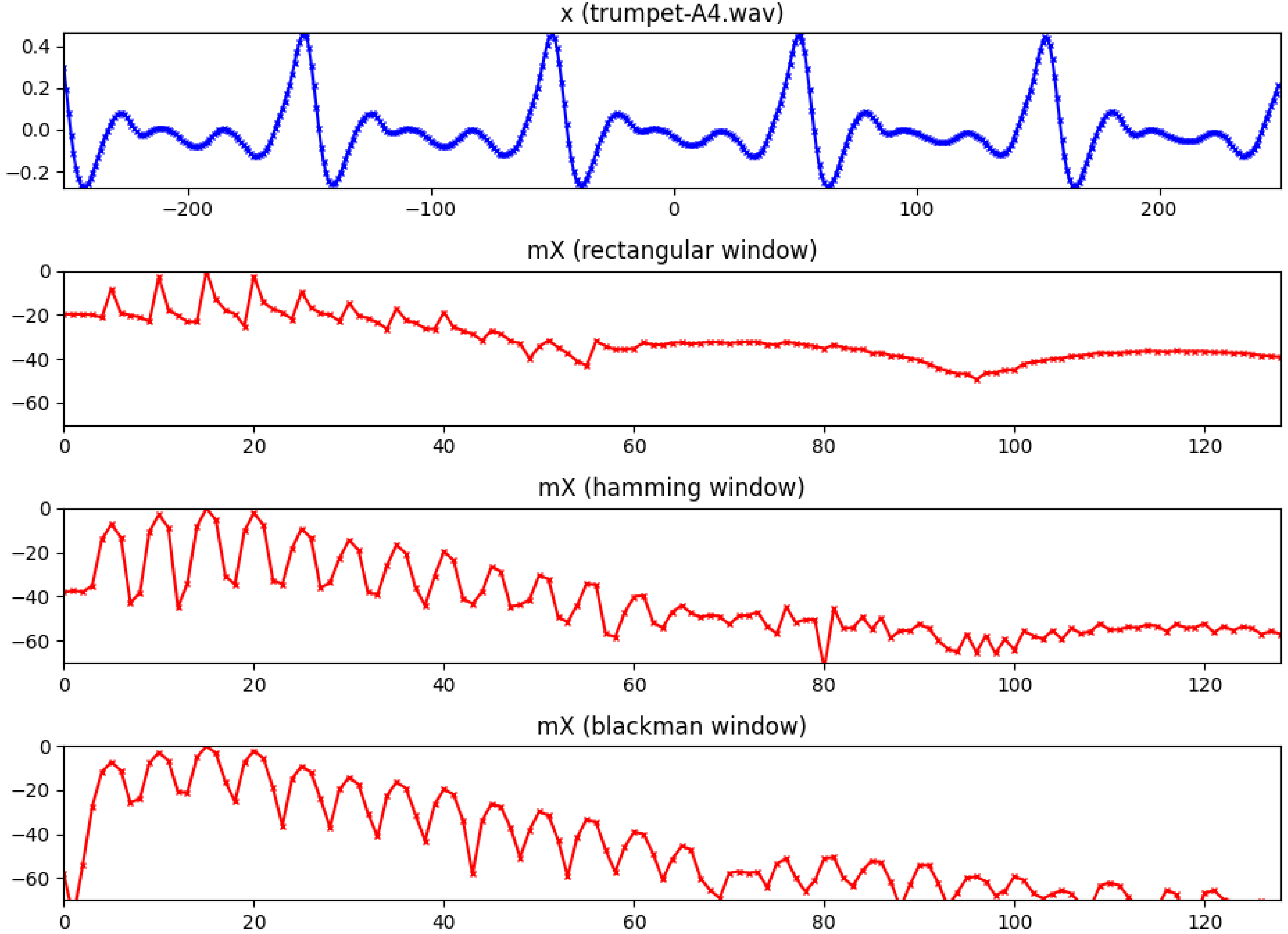
Hanning window

$$w[n] = .5 + .5\cos(2\pi n/M), \quad n = -M/2, \ldots, 0, \ldots M/2$$

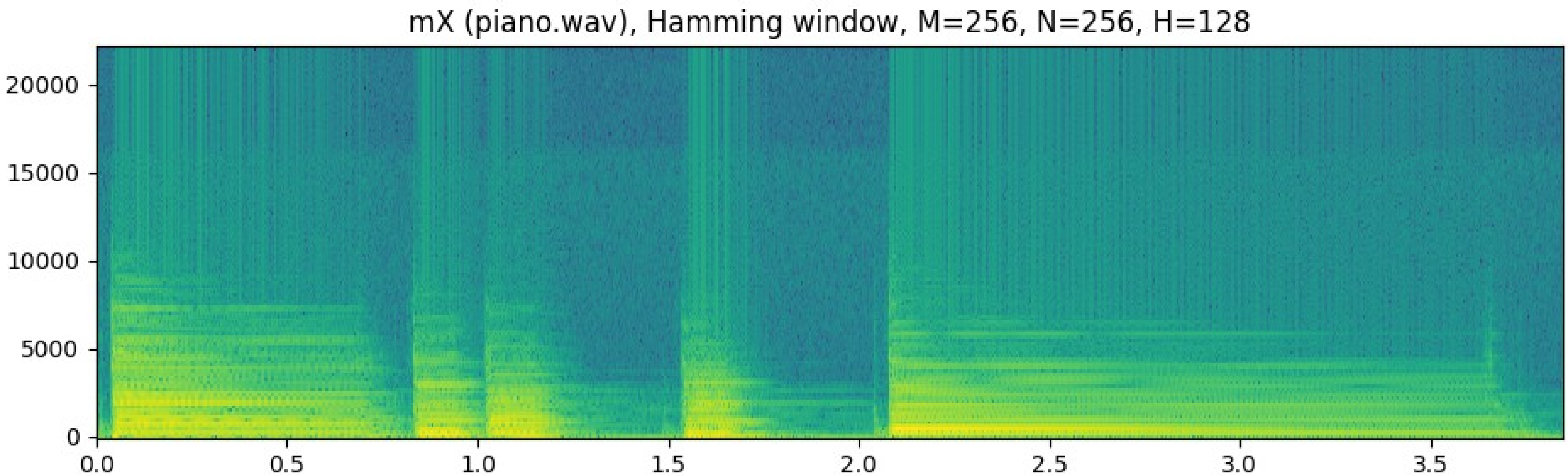
$$W[k] = .5\,D[k] + .25\left(D[k-1] + D[k+1]\right) \quad \text{where } D[k] = \frac{\sin(\pi k)}{\sin(\pi k/M)}$$

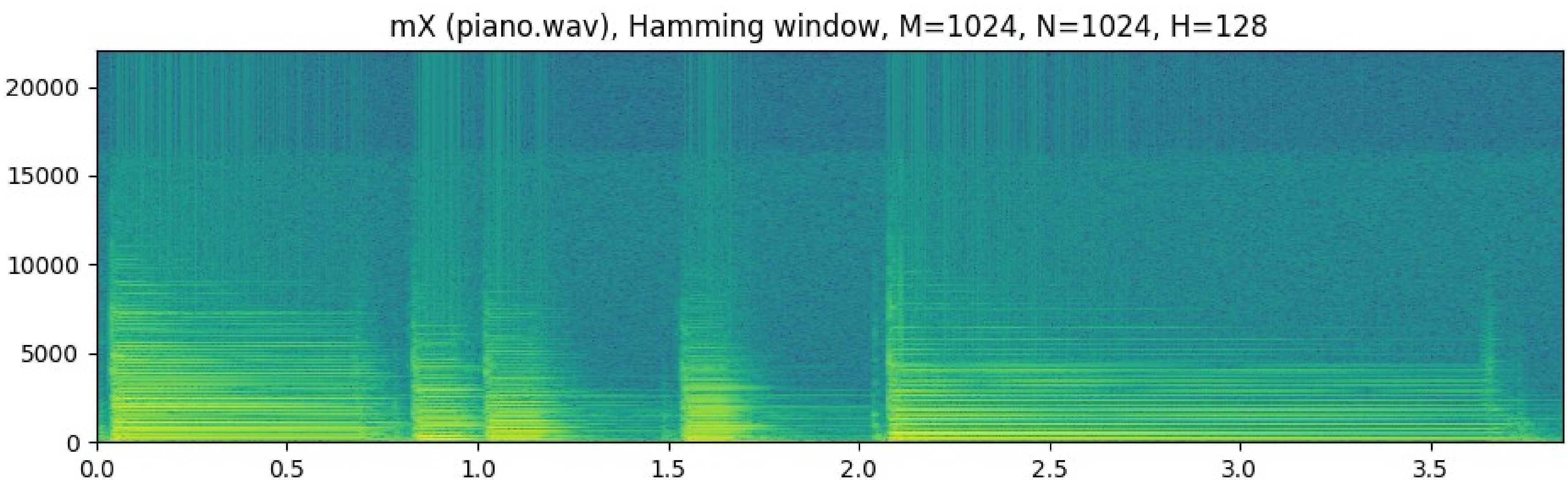


main-lobe width: 4 bins
side-lobe level: -31.5 dB

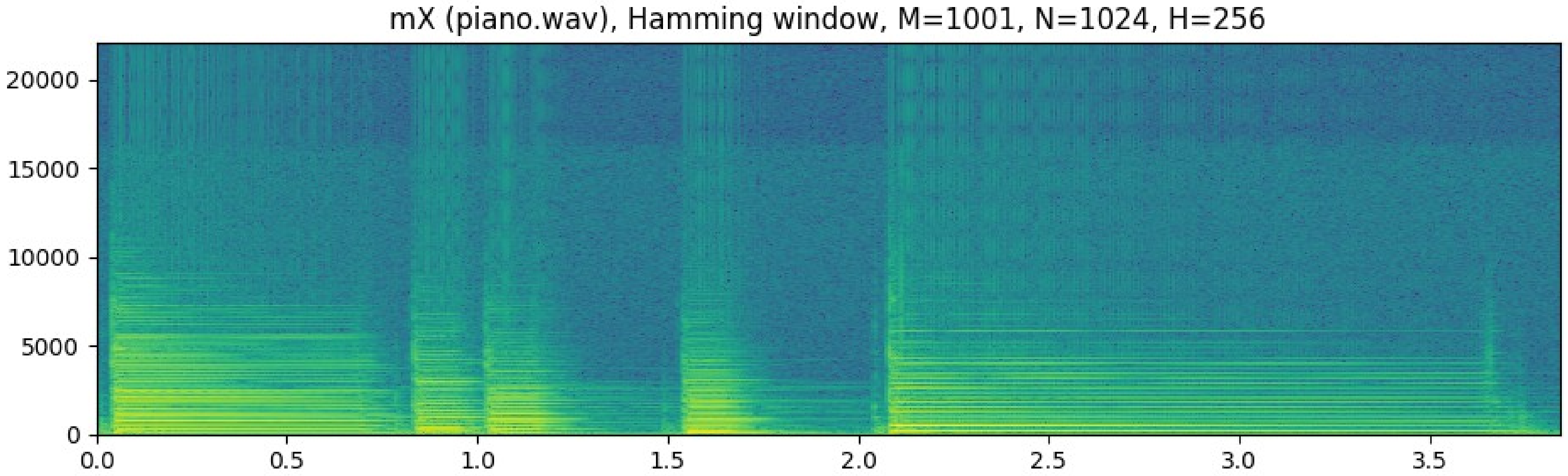

x (trumpet-A4.wav)
0.4
0.2
0.0
−0.2
−200
−100
0
100
200
mX (rectangular window)
0
−20
−40
−60
0
20
40
60
80
100
120
mX (hamming window)
0
−20
−40
−60
0
20
40
60
80
100
120
mX (blackman window)
0
−20
−40
−60
0
20
40
60
80
100
120

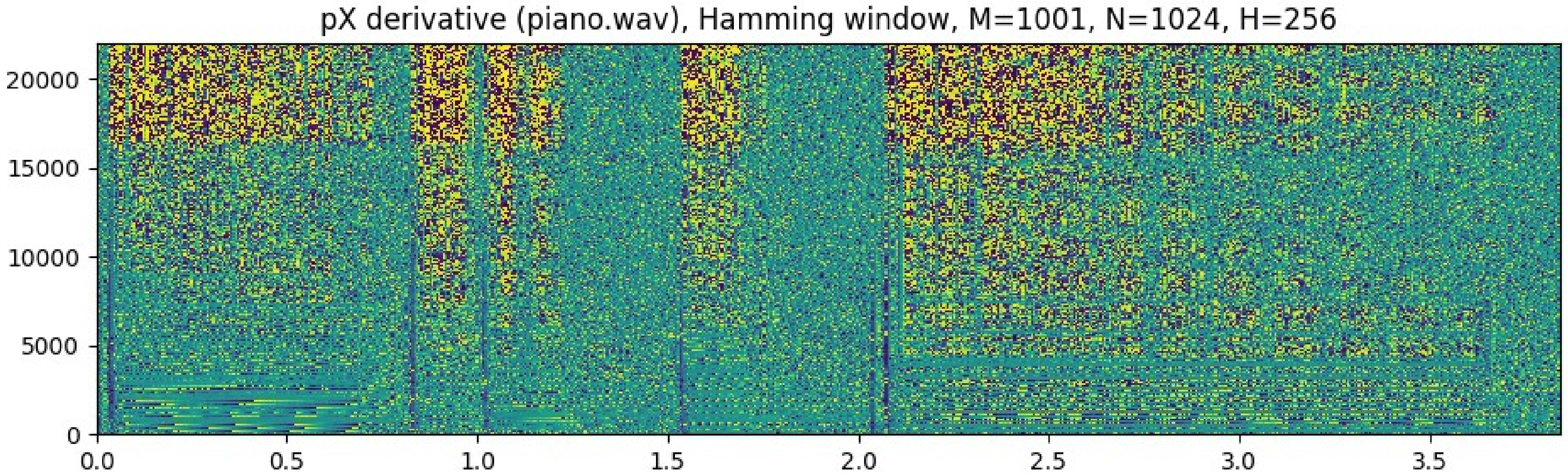
Time-frequency compromise





Amplitude and phase spectrogram





Inverse STFT

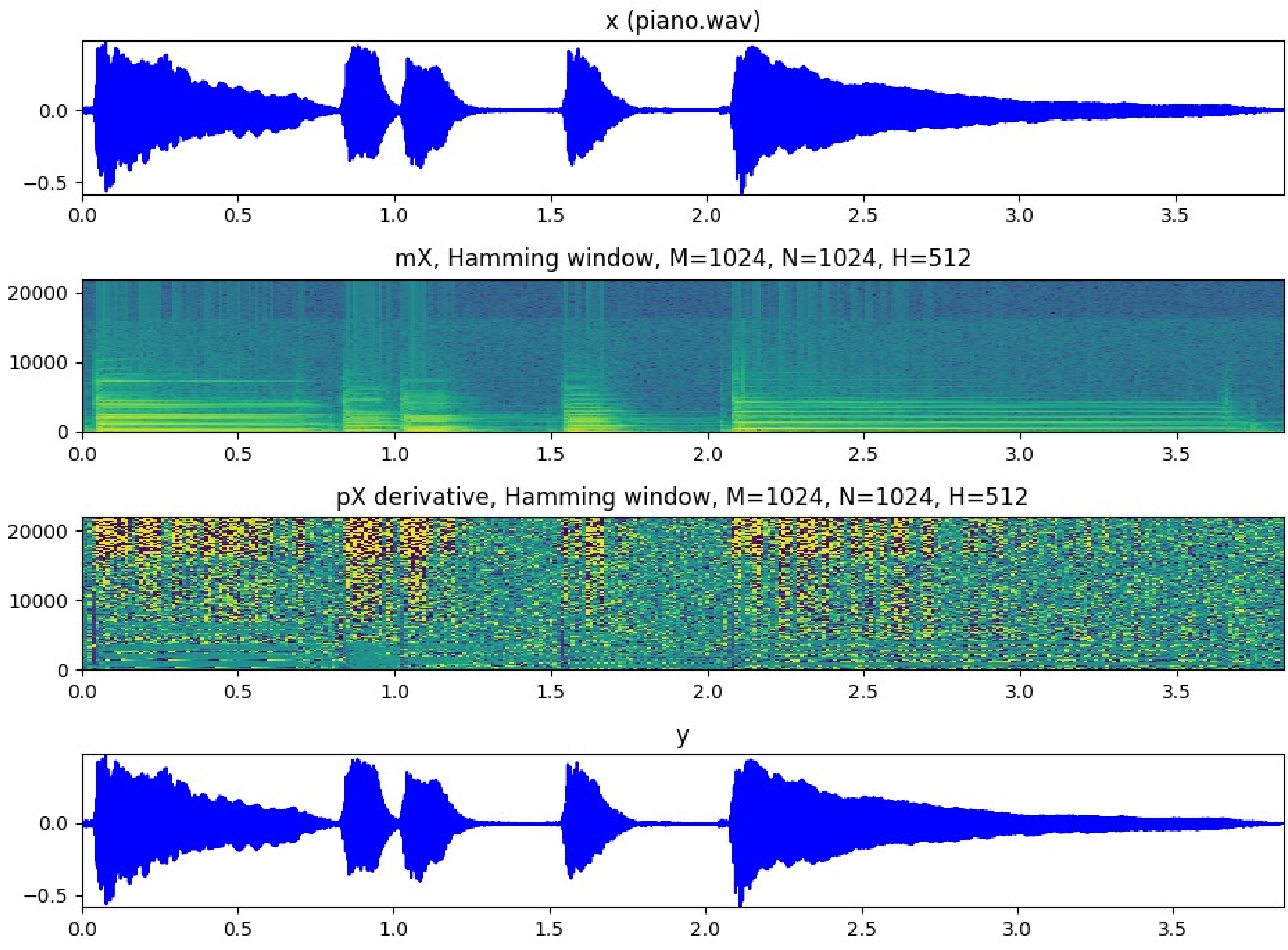
$$y[n] = \sum_{l=0}^{L-1} Shift_{lH,n}\left[\frac{1}{N}\sum_{k=-N/2}^{N/2-1} X_l[k]e^{j2\pi kn/N}\right]$$

each output frame is:

$$yw_l[n] = x(n+lH)w[n]$$

and the output sound is:

$$y[n] = \sum_{l=0}^{L-1} yw_l[n] = x[n]\sum_{l=0}^{L-1} w[n-lH]$$


x (piano.wav)
0.0
−0.5
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
mX, Hamming window, M=1024, N=1024, H=512
20000
10000
0
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
pX derivative, Hamming window, M=1024, N=1024, H=512
20000
10000
0
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
y
0.0
−0.5
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5

References and credits

- More information in:
  https://en.wikipedia.org/wiki/Discrete_Fourier_transform
  https://en.wikipedia.org/wiki/STFT
  https://en.wikipedia.org/wiki/Window_function
- References by Julius O. Smith:
  https://ccrma.stanford.edu/~jos/mdft/
  https://ccrma.stanford.edu/~jos/sasp/
-