The Short-Time Fourier Transform (2 of 2)

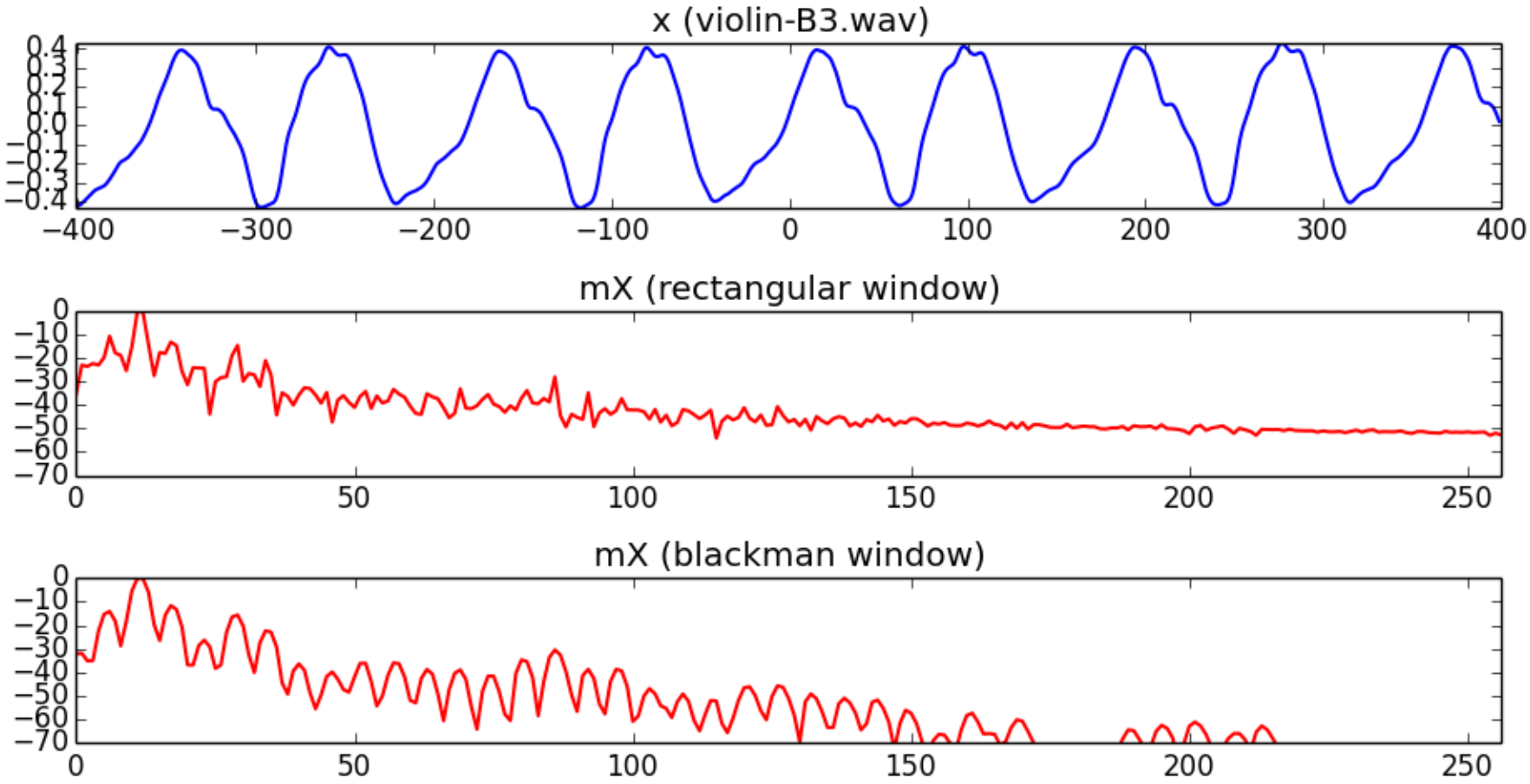
Xavier Serra
Universitat Pompeu Fabra, Barcelona
&
Stanford University

Index

- STFT and analysis window
- Window size
- FFT size
- Hop size
- Time-frequency compromise
- Inverse STFT
- STFT system

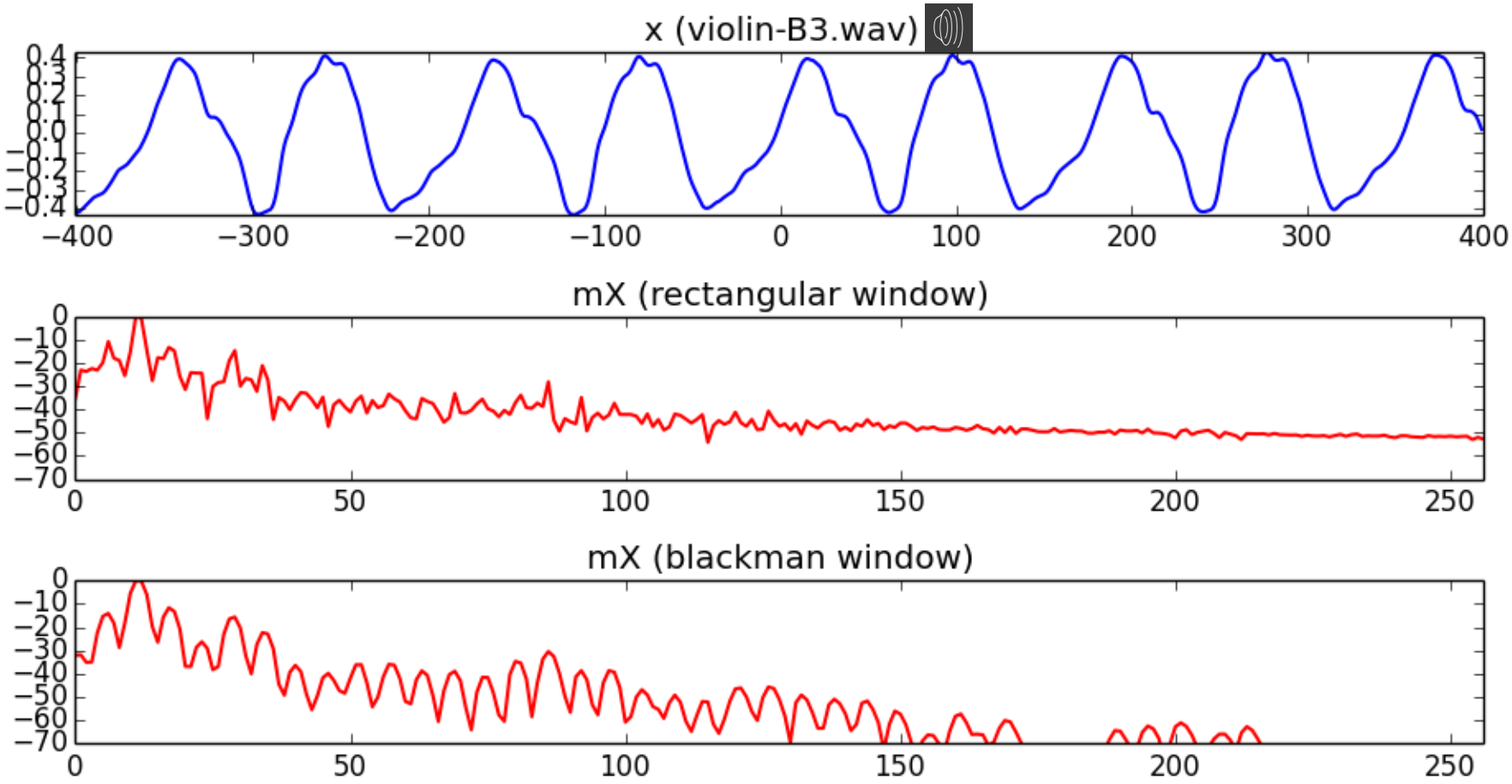
STFT and analysis window

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

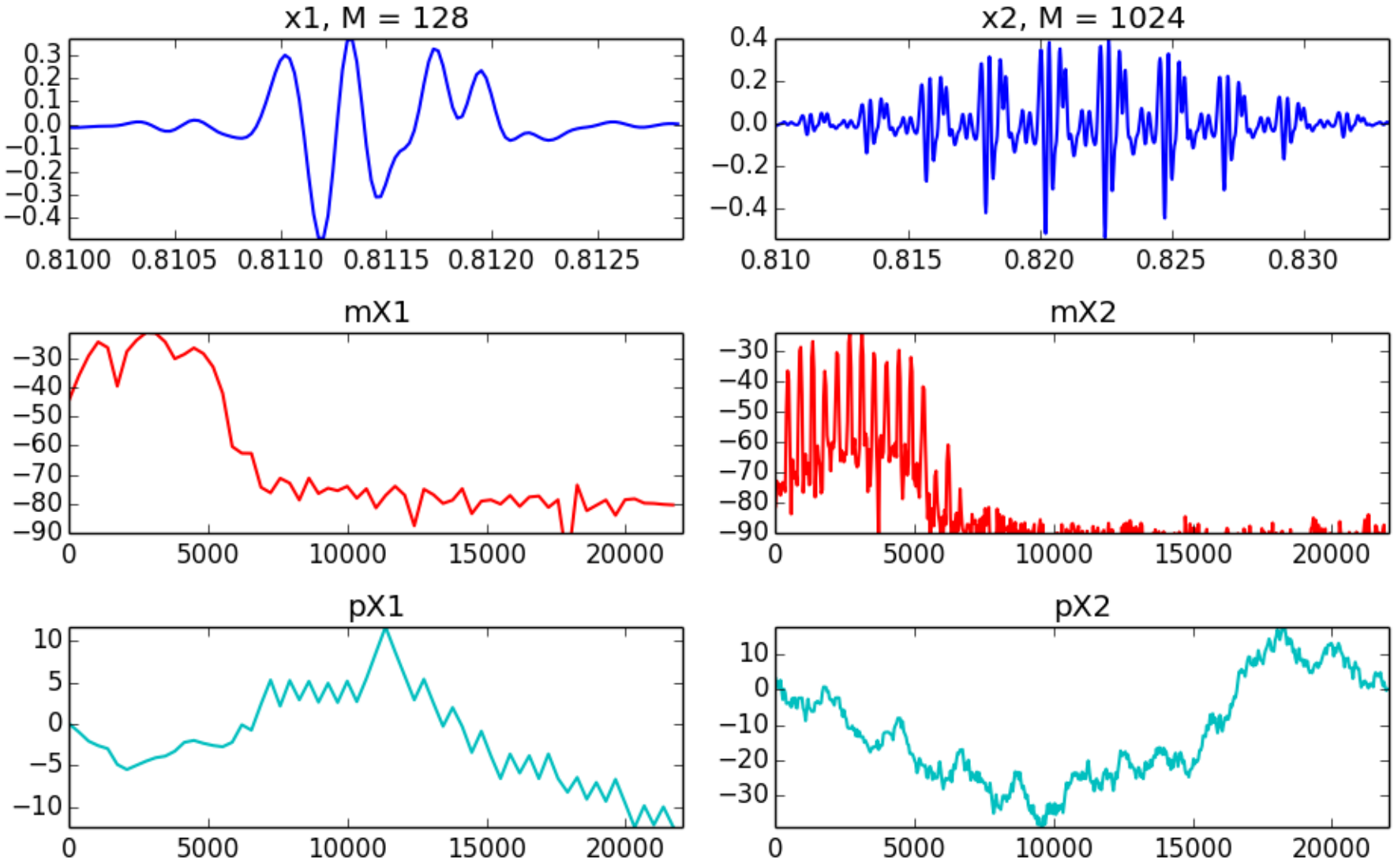


STFT and analysis window

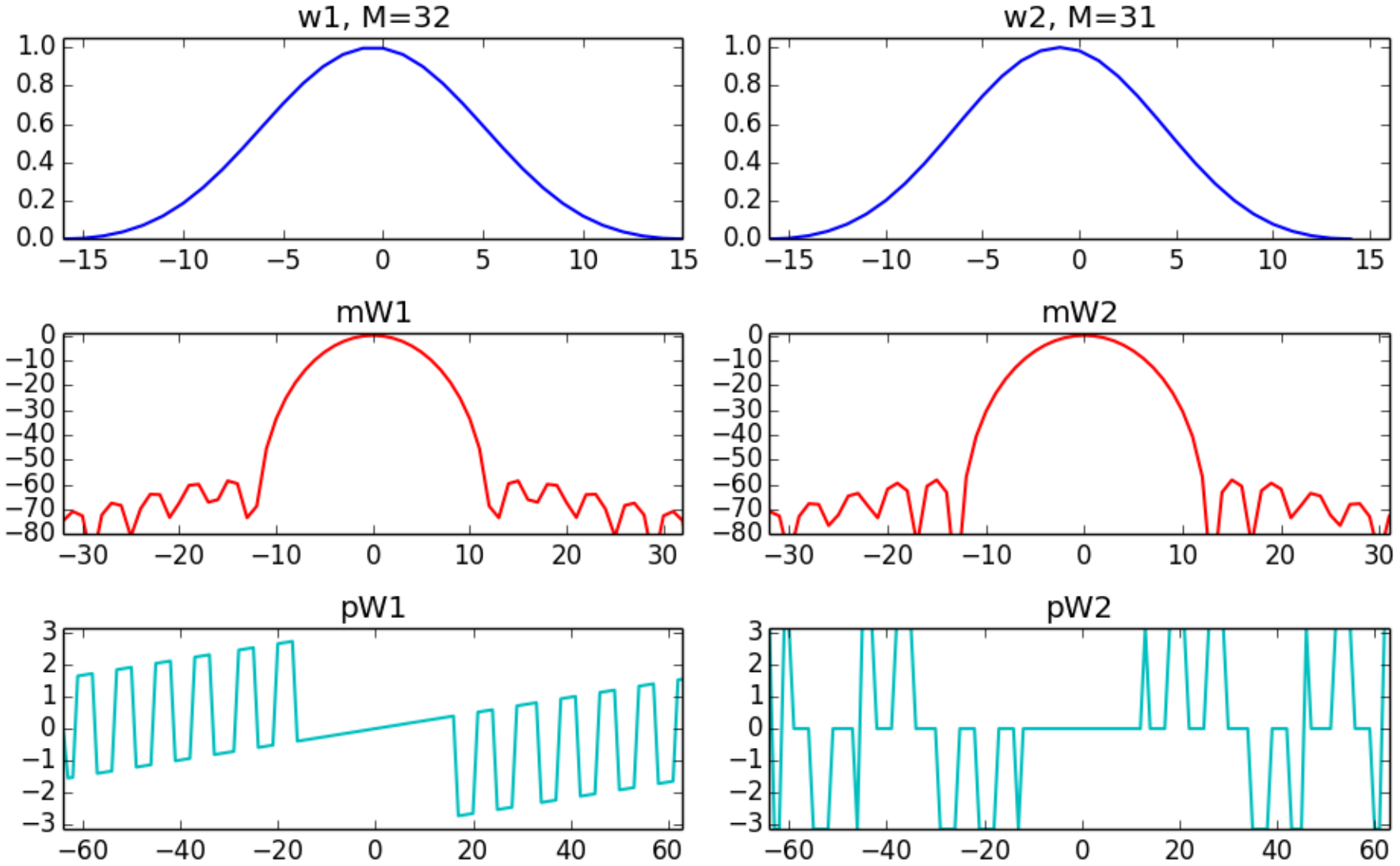
$$X_l[k] = \sum_{n=-N/2}^{N/2-1} w[n]\,x[n+lH]\,e^{-j2\pi kn/N} \qquad l = 0,1,\ldots,$$



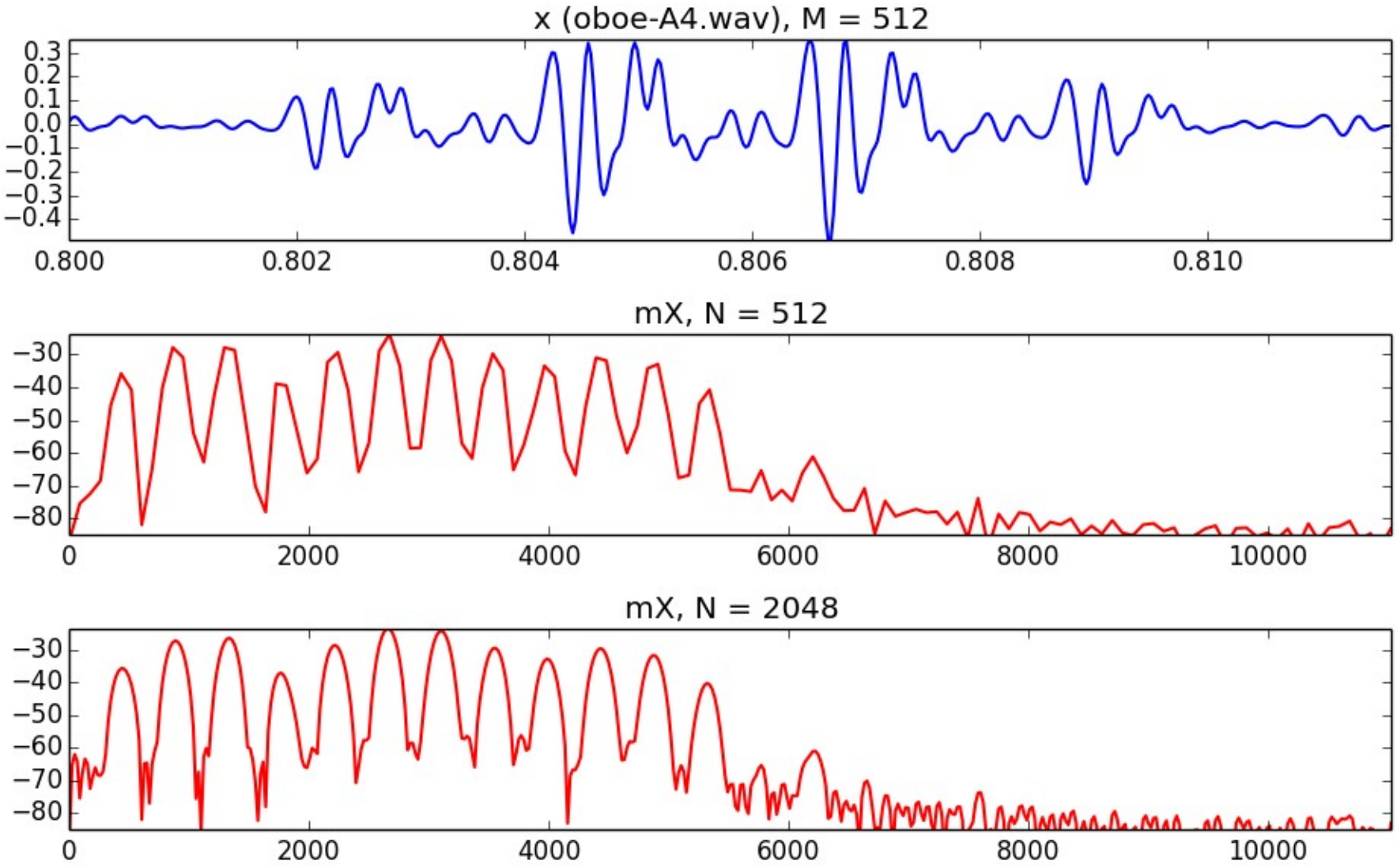
Window size


x1, M = 128
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
0.8100 0.8105 0.8110 0.8115 0.8120 0.8125
x2, M = 1024
0.4
0.2
0.0
−0.2
−0.4
0.810 0.815 0.820 0.825 0.830
mX1
−30
−40
−50
−60
−70
−80
−90
0 5000 10000 15000 20000
mX2
−30
−40
−50
−60
−70
−80
−90
0 5000 10000 15000 20000
pX1
10
5
0
−5
−10
0 5000 10000 15000 20000
pX2
10
0
−10
−20
−30
0 5000 10000 15000 20000

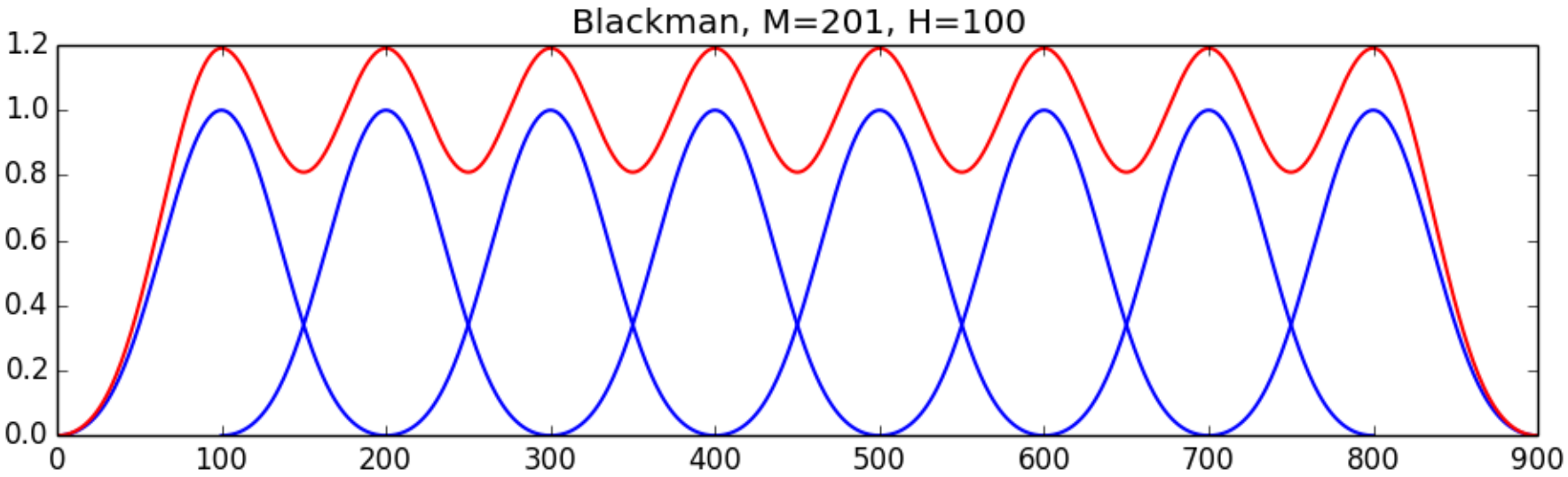
Even-odd size window

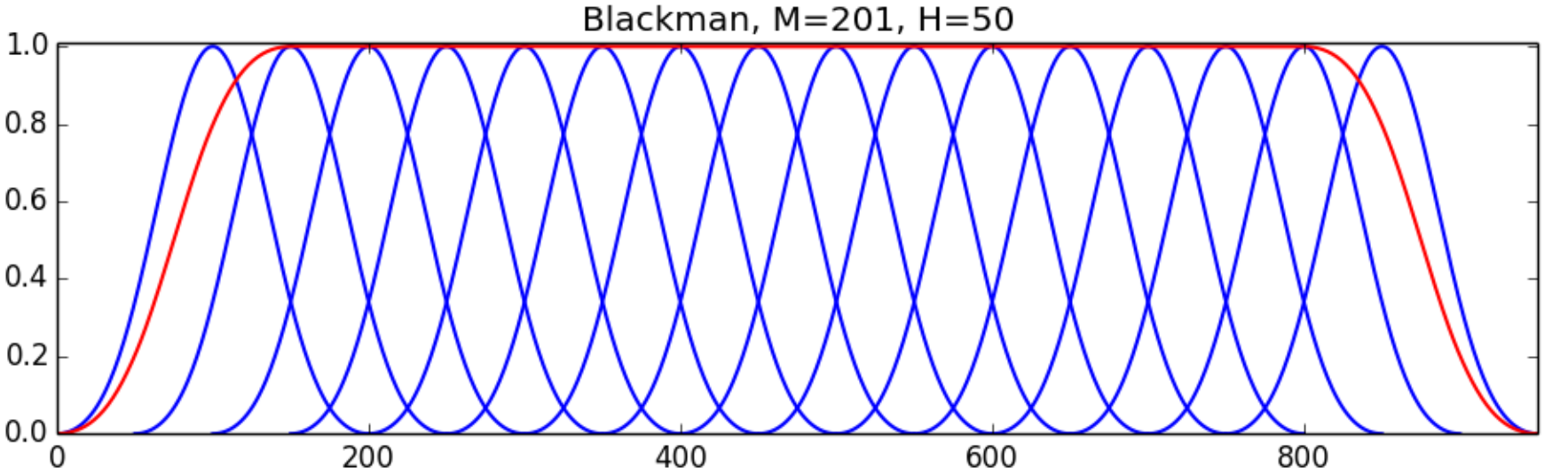


FFT size

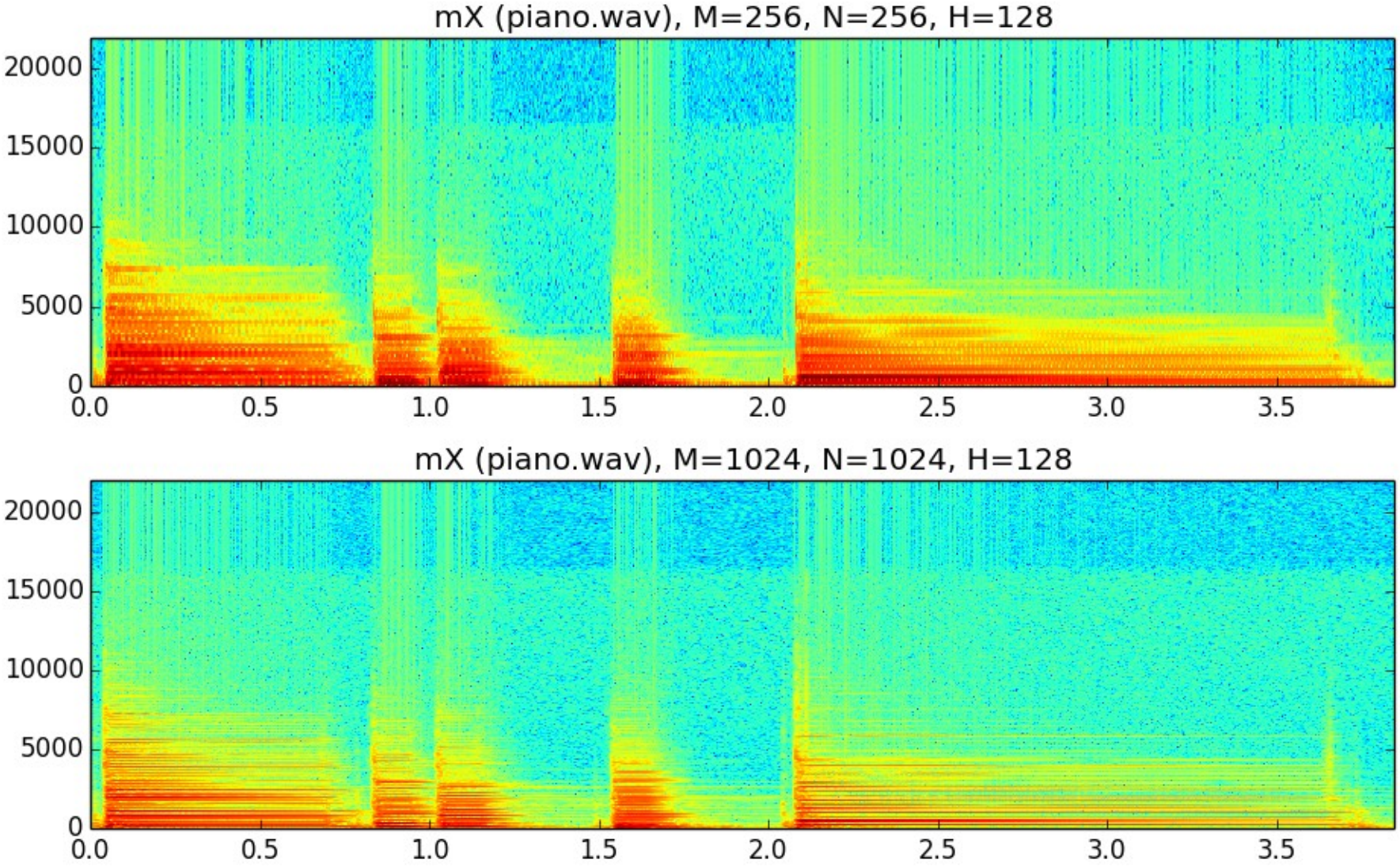

x (oboe-A4.wav), M = 512
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
0.800
0.802
0.804
0.806
0.808
0.810
mX, N = 512
−30
−40
−50
−60
−70
−80
0
2000
4000
6000
8000
10000
mX, N = 2048
−30
−40
−50
−60
−70
−80
0
2000
4000
6000
8000
10000

Hop size $A_w[n] = \sum_{l=0}^{L-1} w[n - lH] = c$

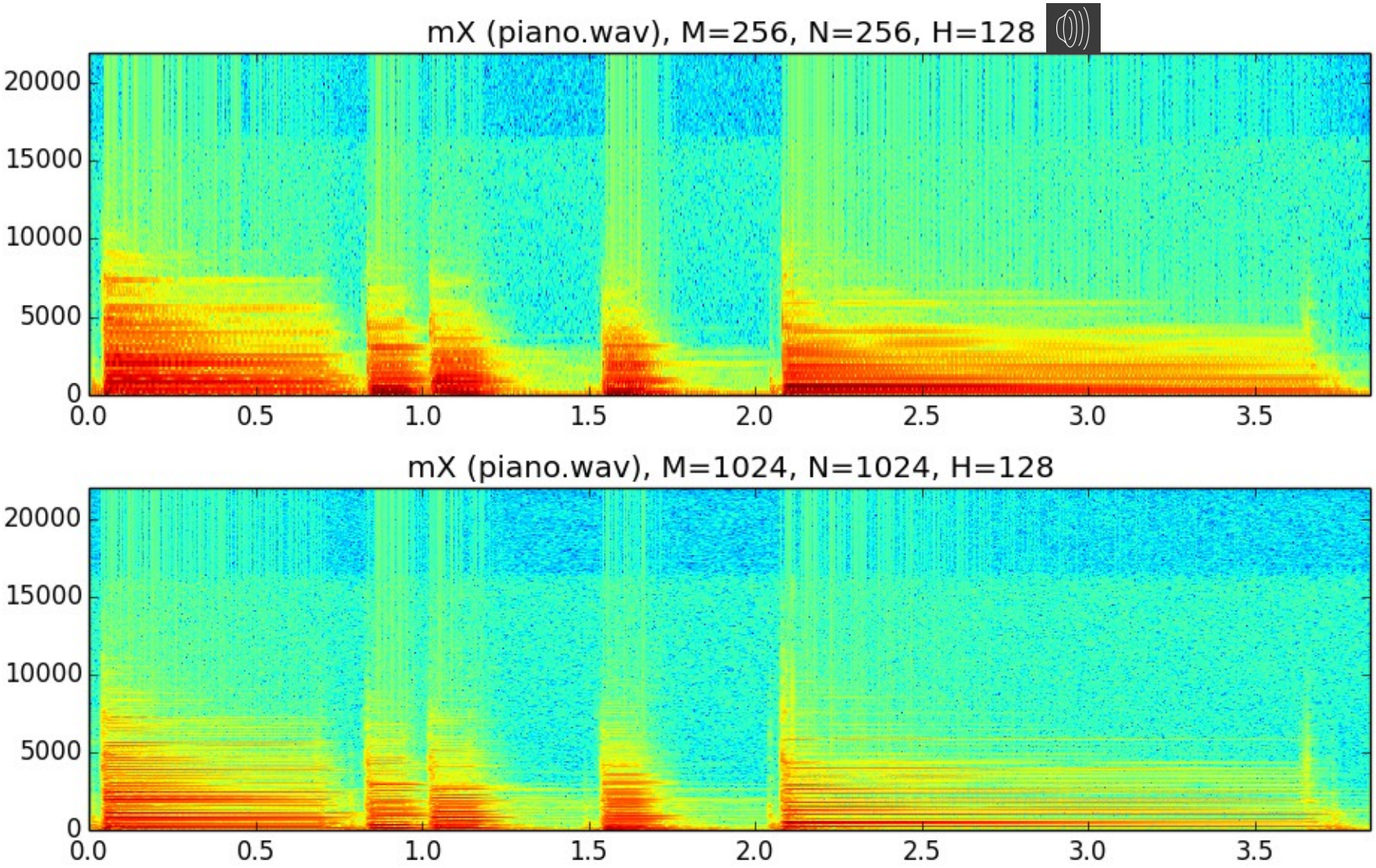




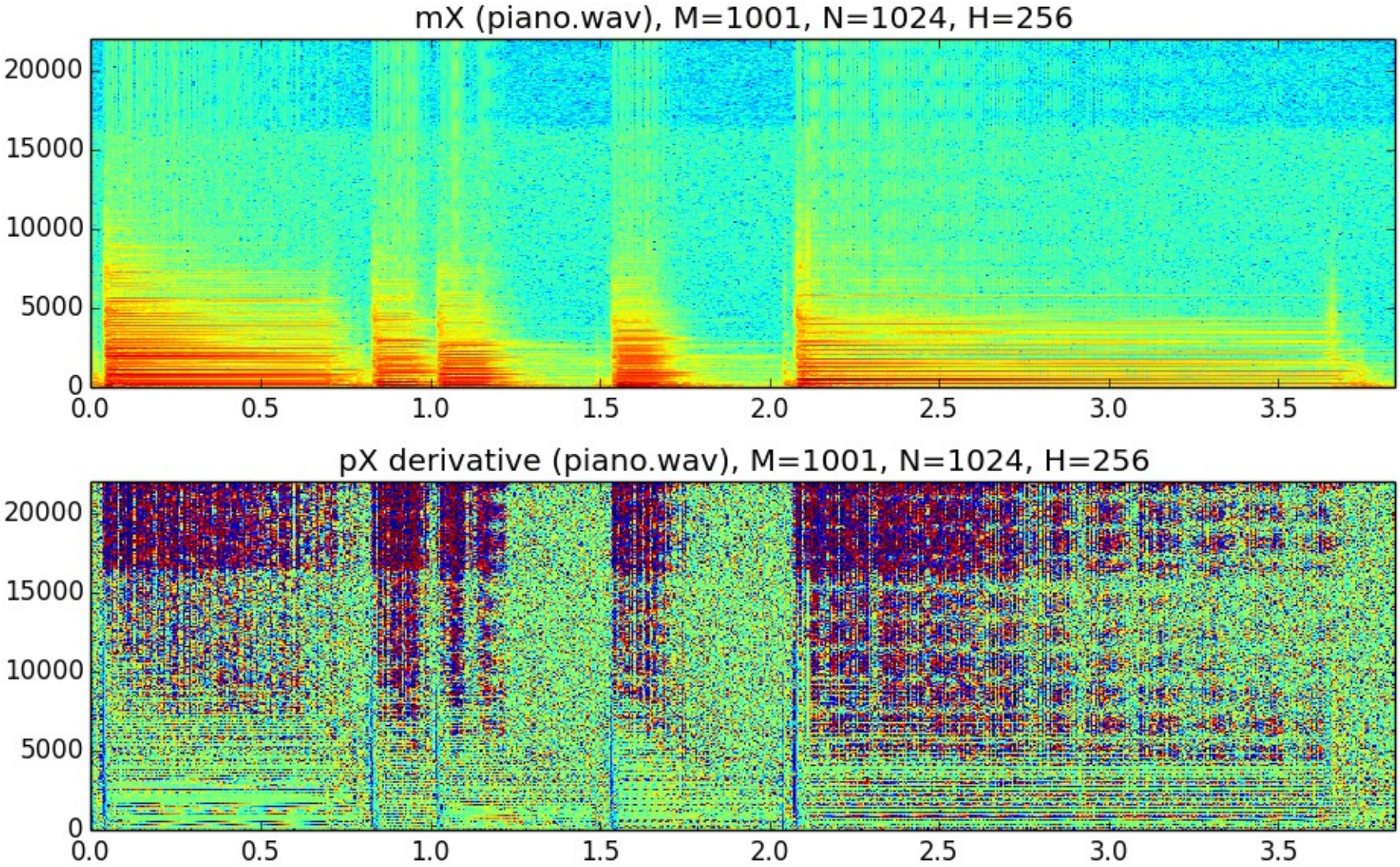
Time-frequency compromise



Time-frequency compromise



Amplitude and phase spectrogram



Inverse STFT

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

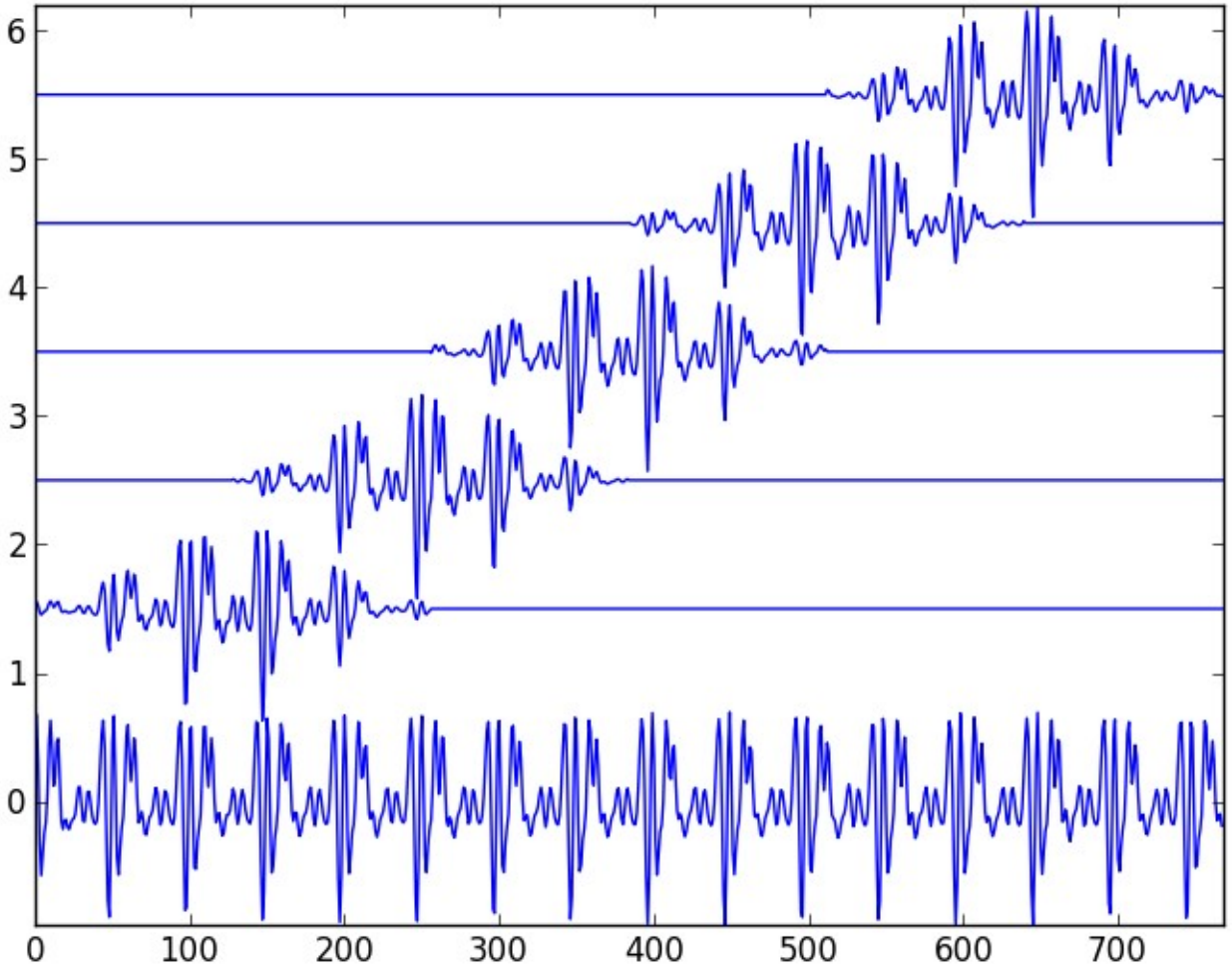
each output frame is:

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

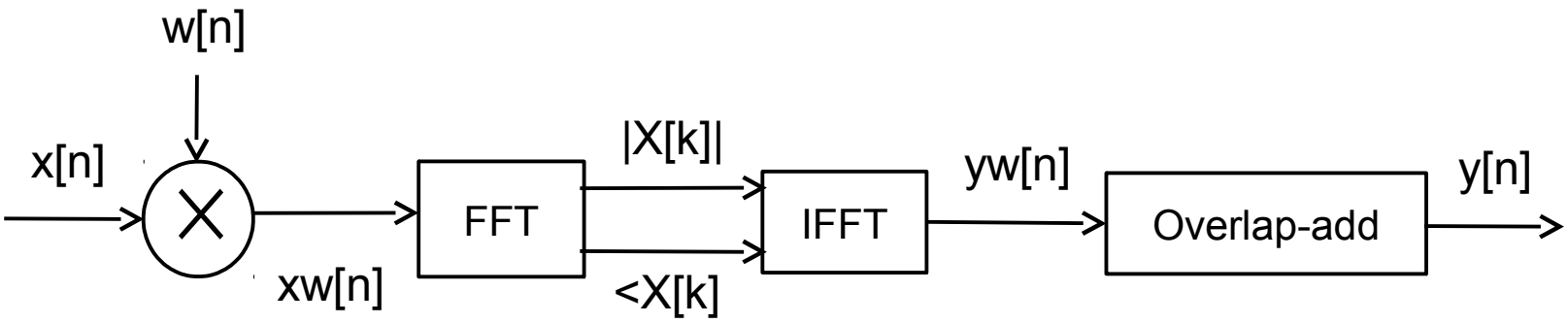
and the output sound is:

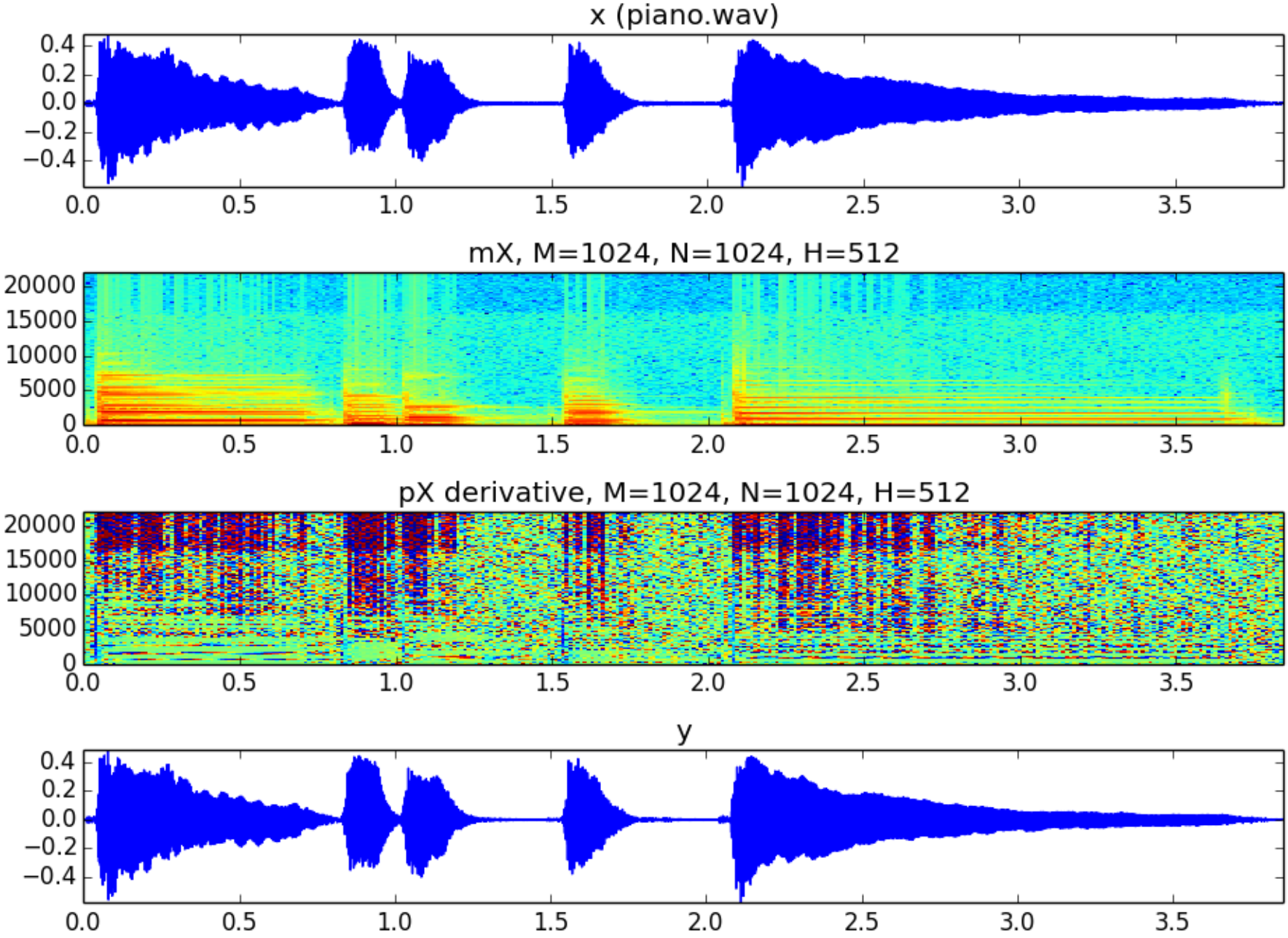
$$y[n] = \sum_{l=0}^{L-1} yw_l[n] = x[n] \sum_{l=0}^{L-1} w[n - lH]$$

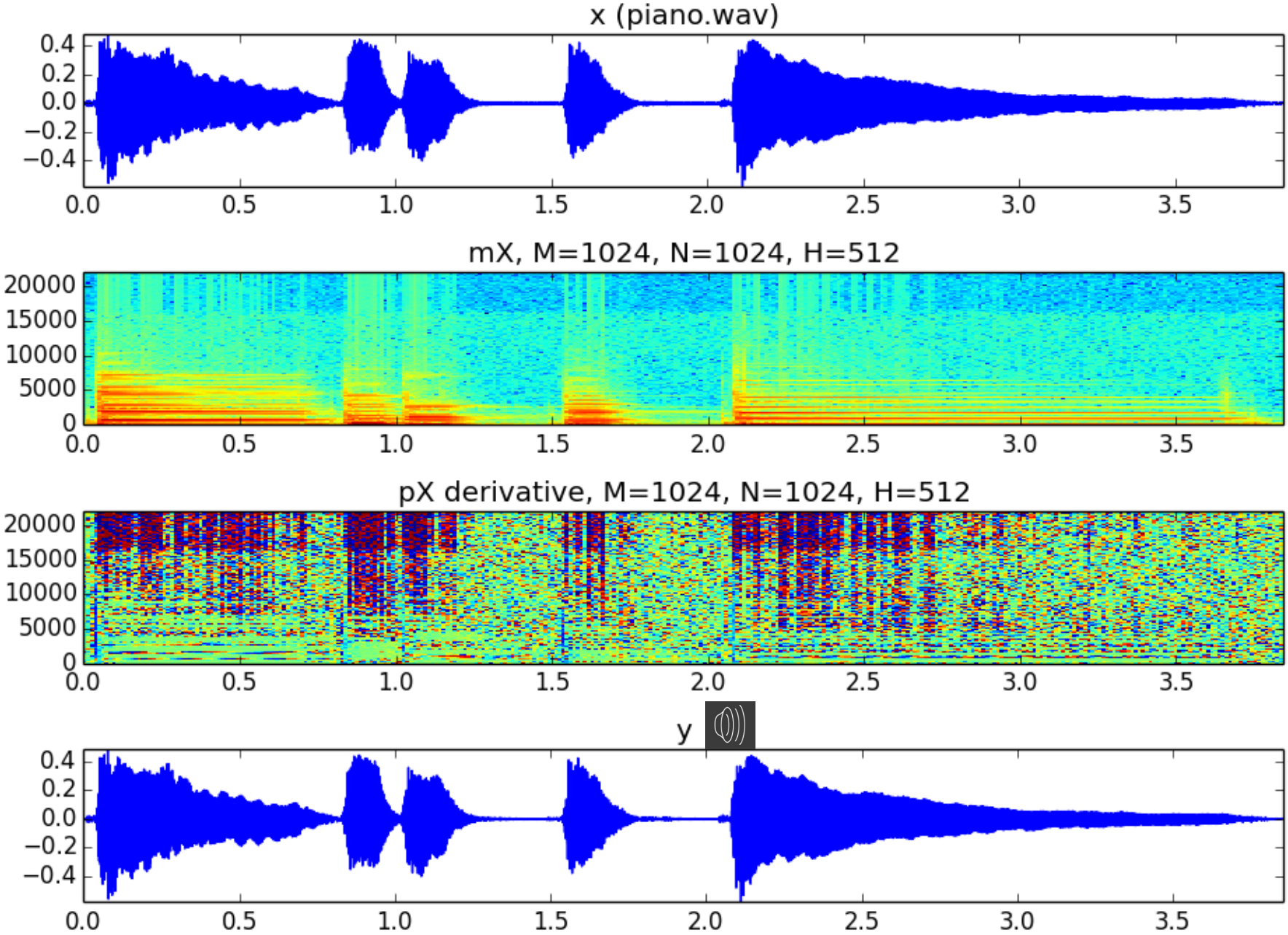
$$yw_l[n] = w[n]\,y[n + lH] \qquad l = 0,1,\ldots,$$



STFT system




x (piano.wav)
0.4
0.2
0.0
−0.2
−0.4
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
mX, M=1024, N=1024, H=512
20000
15000
10000
5000
0
pX derivative, M=1024, N=1024, H=512
y


x (piano.wav)
0.4
0.2
0.0
−0.2
−0.4
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
mX, M=1024, N=1024, H=512
20000
15000
10000
5000
0
pX derivative, M=1024, N=1024, H=512
y

References and credits

- More information in:
  - https://en.wikipedia.org/wiki/STFT
  - https://en.wikipedia.org/wiki/Window_function
  - http://en.wikipedia.org/wiki/Spectrogram
- Reference on the STFT by Julius O. Smith: https://ccrma.stanford.edu/~jos/sasp/
- Sounds from: http://www.freesound.org/people/xserra/packs/13038/
- Slides and code released using the CC Attribution-Noncommercial-Share Alike license or the Affero GPL license and available from https://github.com/MTG/sms-tools

The Short-Time Fourier Transform (2 of 2)

Xavier Serra
Universitat Pompeu Fabra, Barcelona
&
Stanford University