Carnegie Mellon

15-826: Multimedia (Databases) and Data Mining

Lecture #22: DSP tools – Wavelets

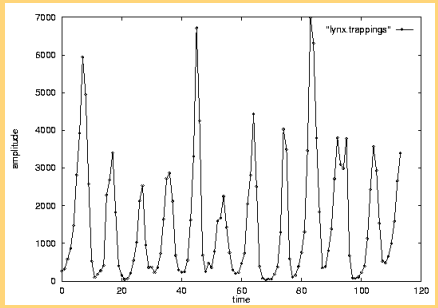
C. Faloutsos

Problem

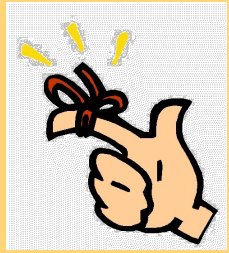
Goal: given a signal (eg., sales over time and/or space)

Q: Find patterns and/or compress

count



year

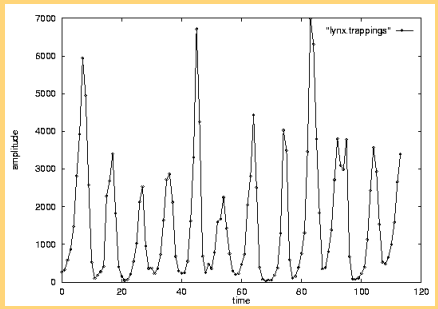


Solutions:

Goal: given a signal (eg., sales over time and/or space)

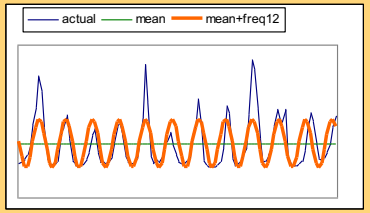
Q: Find patterns and/or compress

count

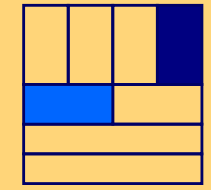


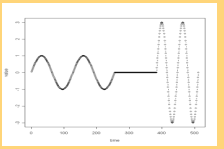
year

A1: Fourier (DFT)



A2: Wavelets (DWT)





Must-read Material

- Wavelets: In PTVF ch. 13.10; in MM Textbook Appendix C

Outline

Goal: 'Find similar / interesting things'

- Intro to DB
- Indexing - similarity search
- Data Mining

Indexing - Detailed outline

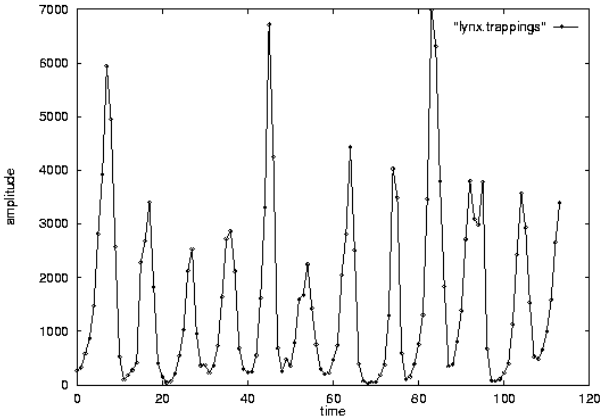
- primary key indexing
- ..
- Multimedia –
  - Digital Signal Processing (DSP) tools
    - Discrete Fourier Transform (DFT)
    - Discrete Wavelet Transform (DWT)

Reminder: Problem:

Goal: given a signal (eg., #packets over time)

Find: patterns, periodicities, and/or **compress**

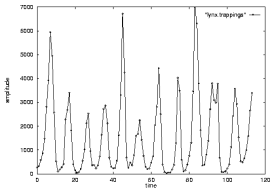
count



lynx caught per year
(packets per day;
virus infections per month)

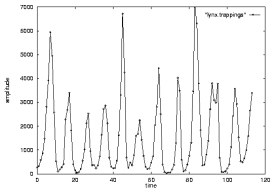
year

Important observations



Patterns, rules, forecasting and similarity indexing are closely related:

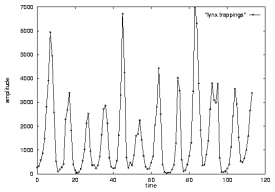
Important observations



Patterns, rules, forecasting and similarity indexing are closely related:

- To do **forecasting**, we need
  - to find **patterns**/rules
  - compress
  - to find **similar** settings in the past
- to find outliers, we need to have forecasts
  - (outlier = too far away from our forecast)

Important observations

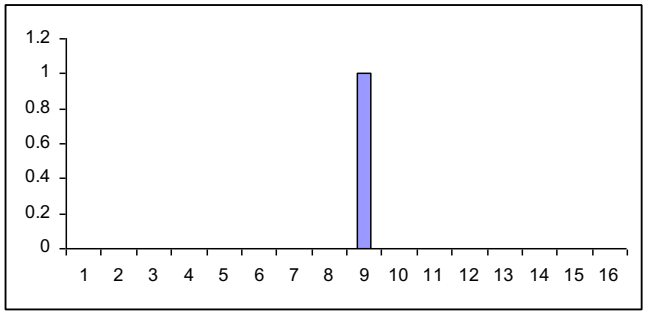


Patterns, rules, forecasting and similarity indexing are closely related:

Wavelets - DWT

- DFT is great - but, how about compressing a spike?

value

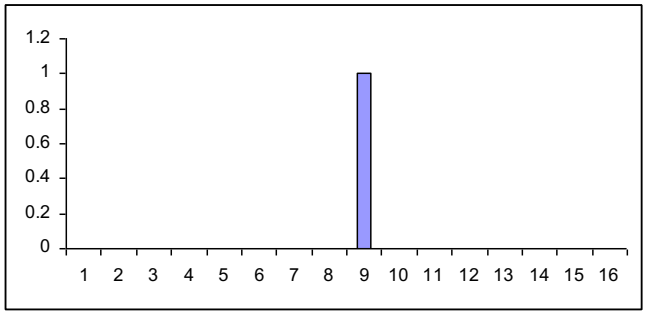


time

Wavelets - DWT

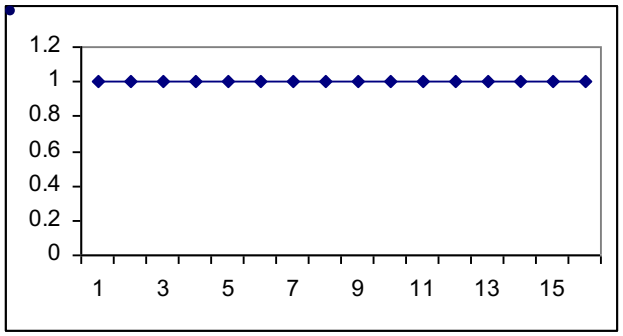
- DFT is great - but, how about compressing a spike?
- A: Terrible - all DFT coefficients needed!

value



time

Ampl.

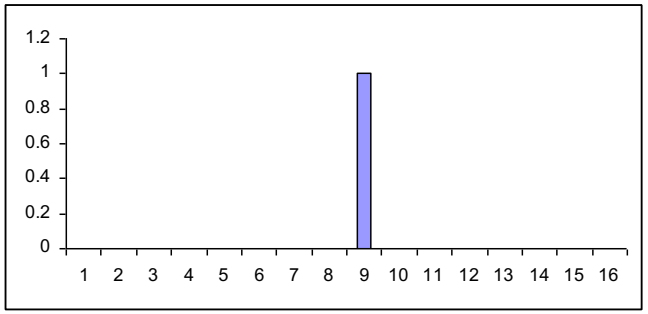


Freq.

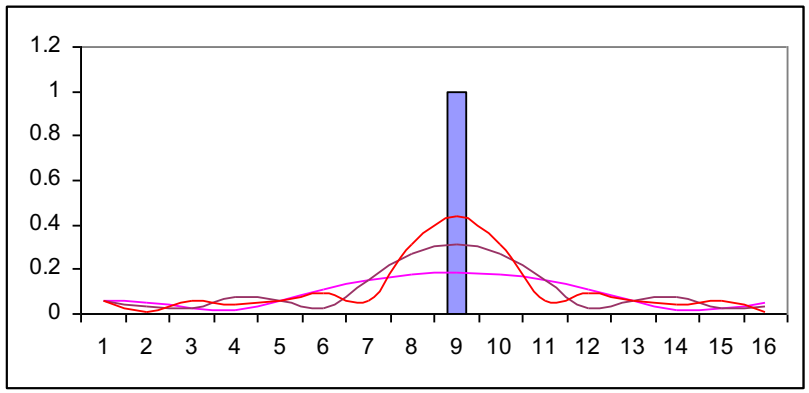
Wavelets - DWT

- DFT is great - but, how about compressing a spike?
- A: Terrible - all DFT coefficients needed!

value



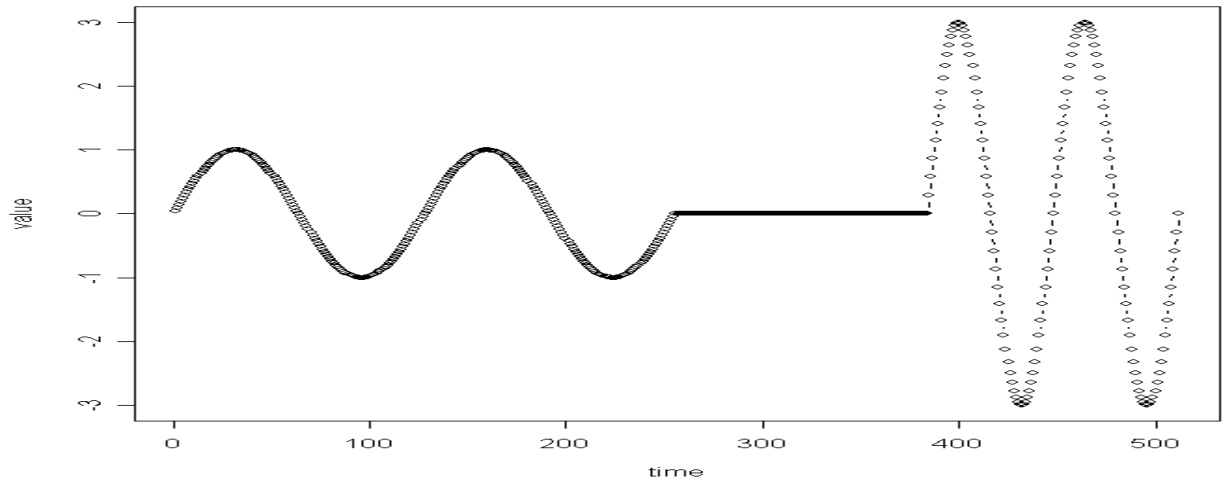
time



Wavelets - DWT

- Similarly, DFT suffers on short-duration waves (eg., baritone, silence, soprano)

value

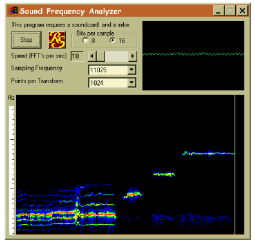


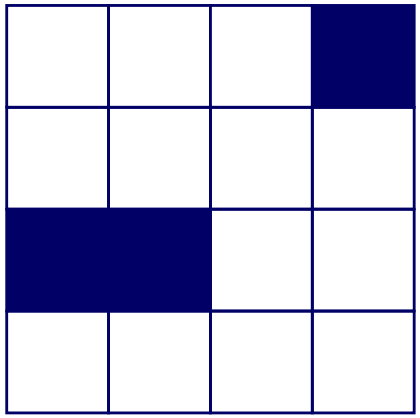
time

Wavelets - DWT

- Solution#1: Short window Fourier transform (SWFT)
- But: how short should be the window?

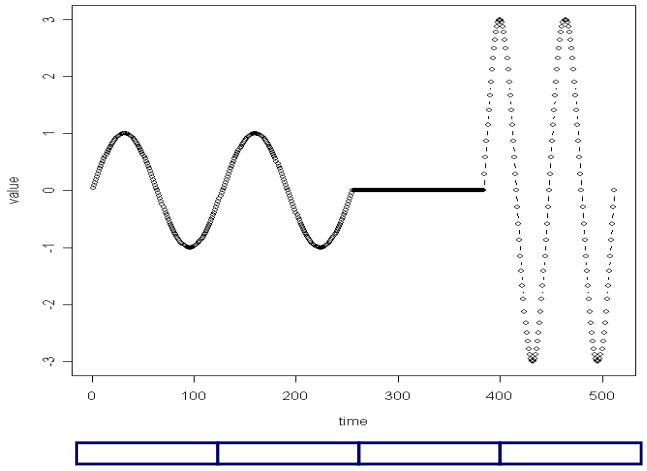
freq





time

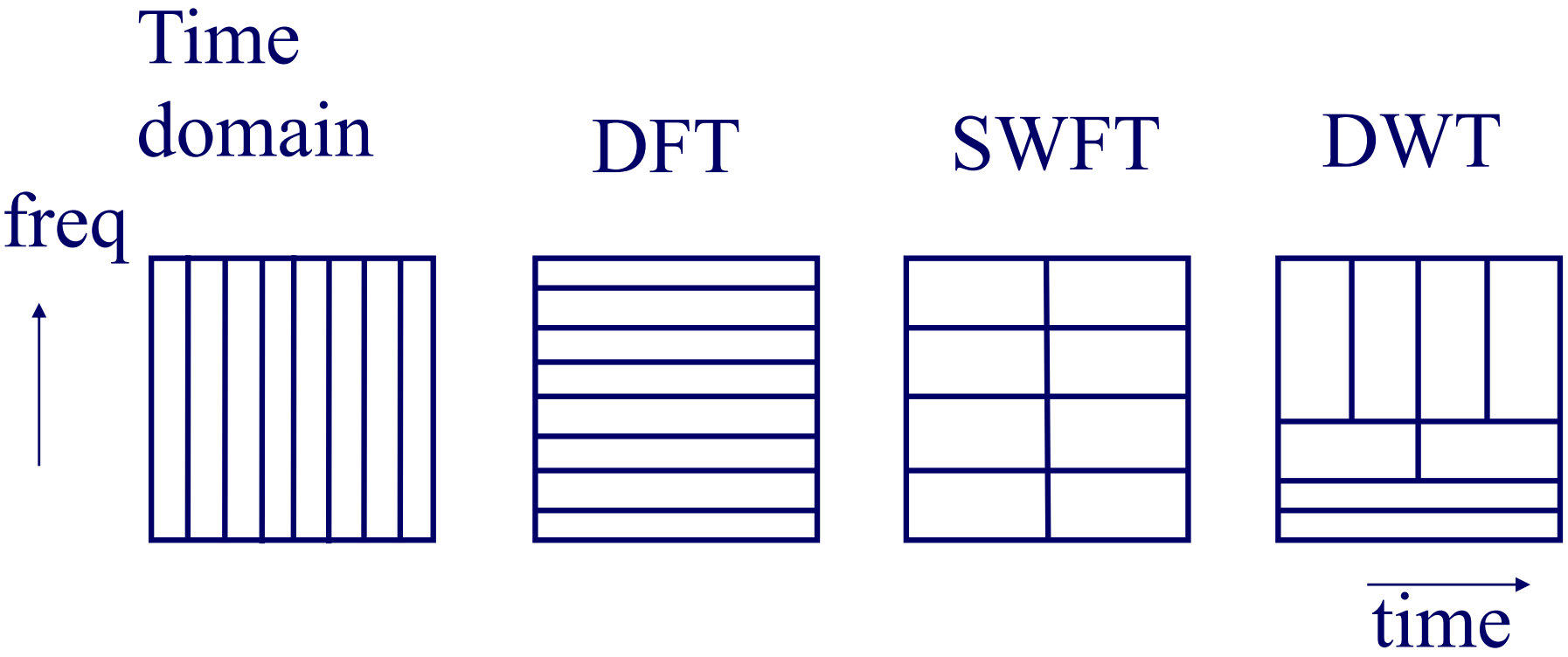
value



time

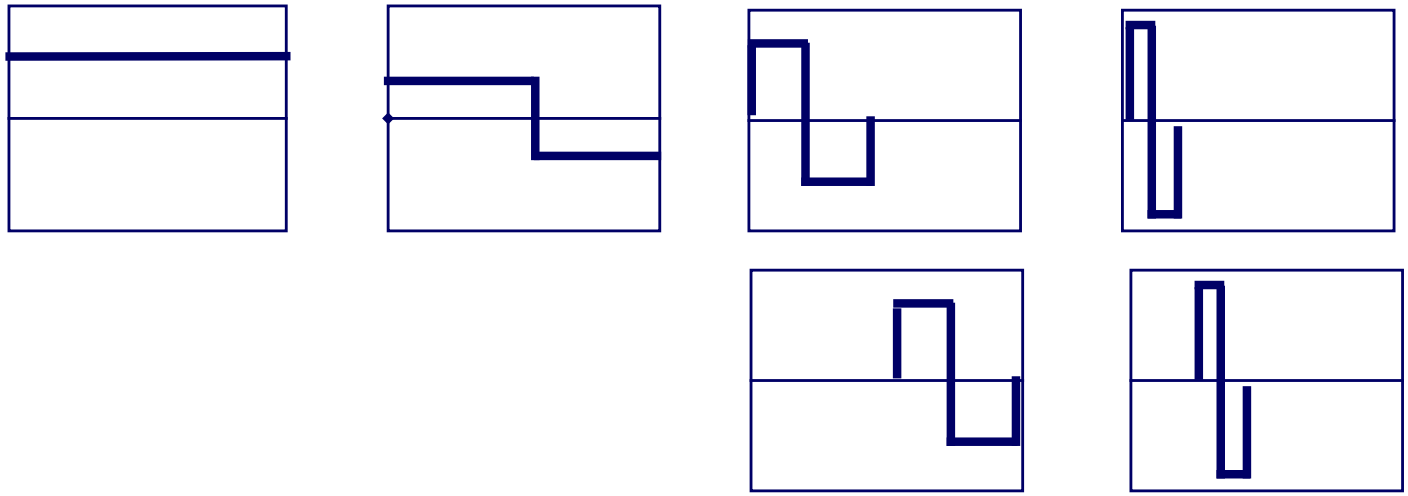
Wavelets - DWT

- Answer: **multiple** window sizes! -> DWT



Haar Wavelets

- subtract sum of left half from right half
- repeat recursively for quarters, eight-ths, ...

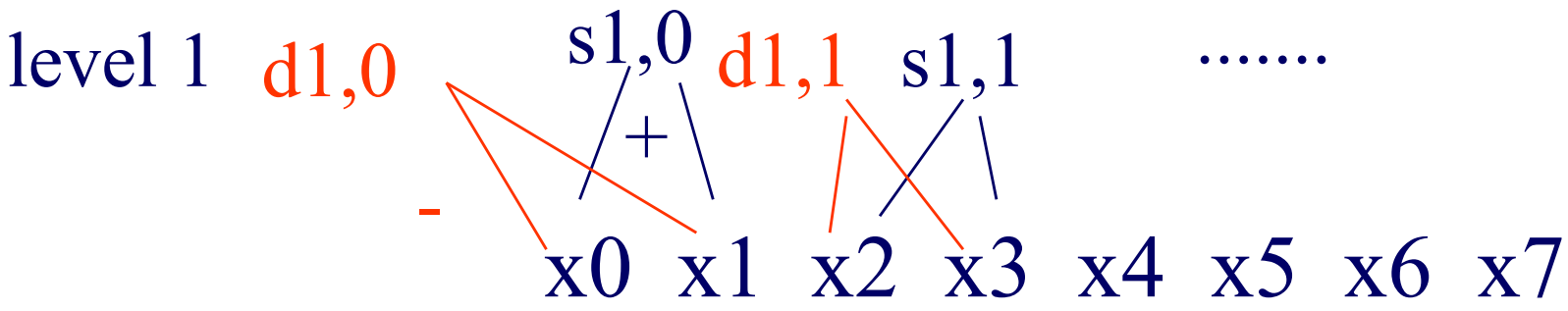


Wavelets - construction

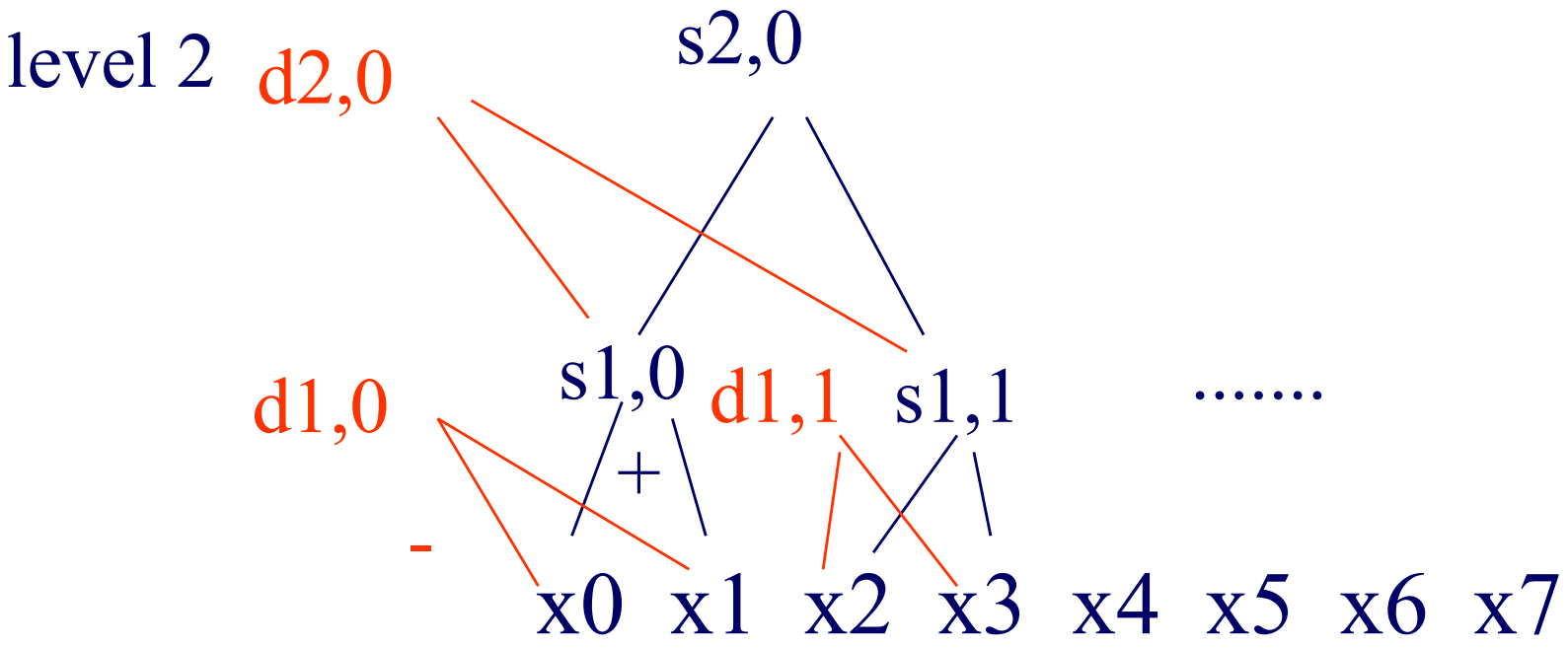
x0 x1 x2 x3 x4 x5 x6 x7

Carnegie Mellon

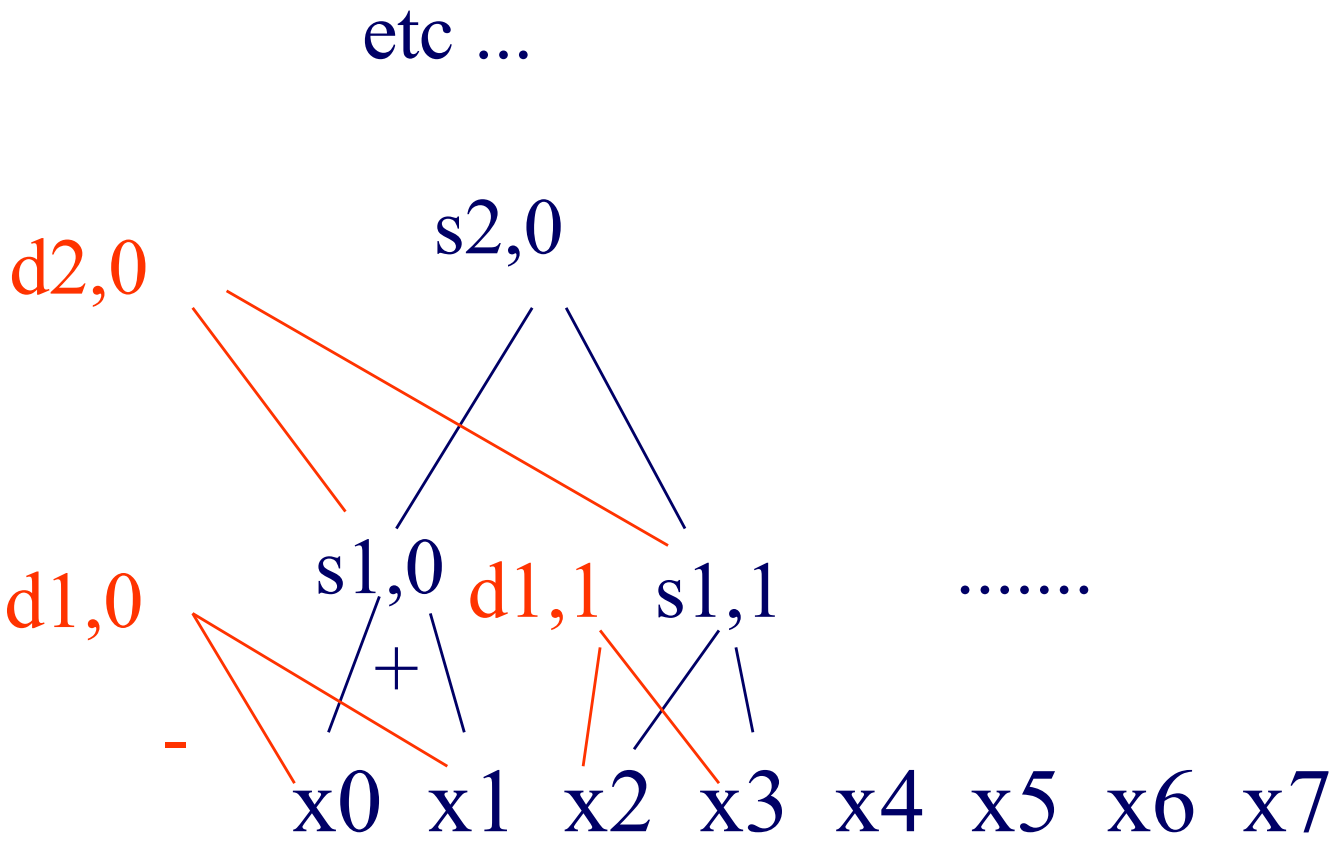
Wavelets - construction



Wavelets - construction



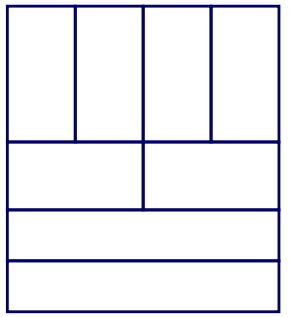
Wavelets - construction

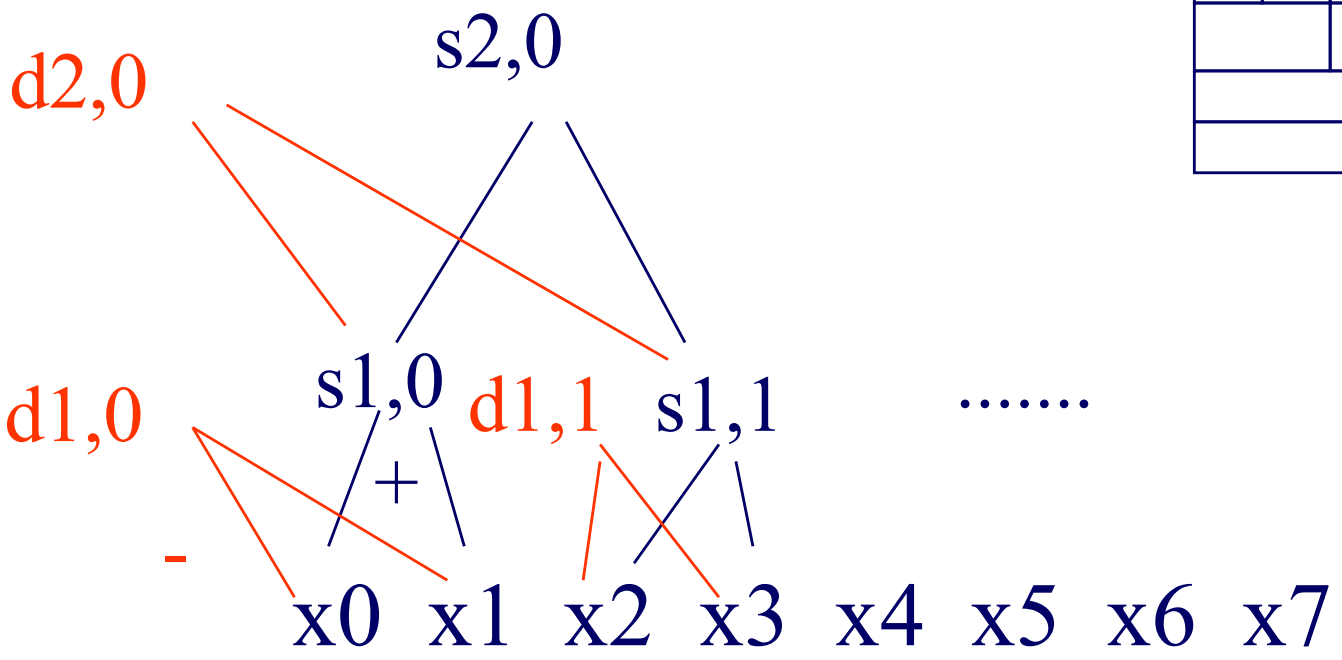


Wavelets - construction

Q: map each coefficient

on the time-freq. plane

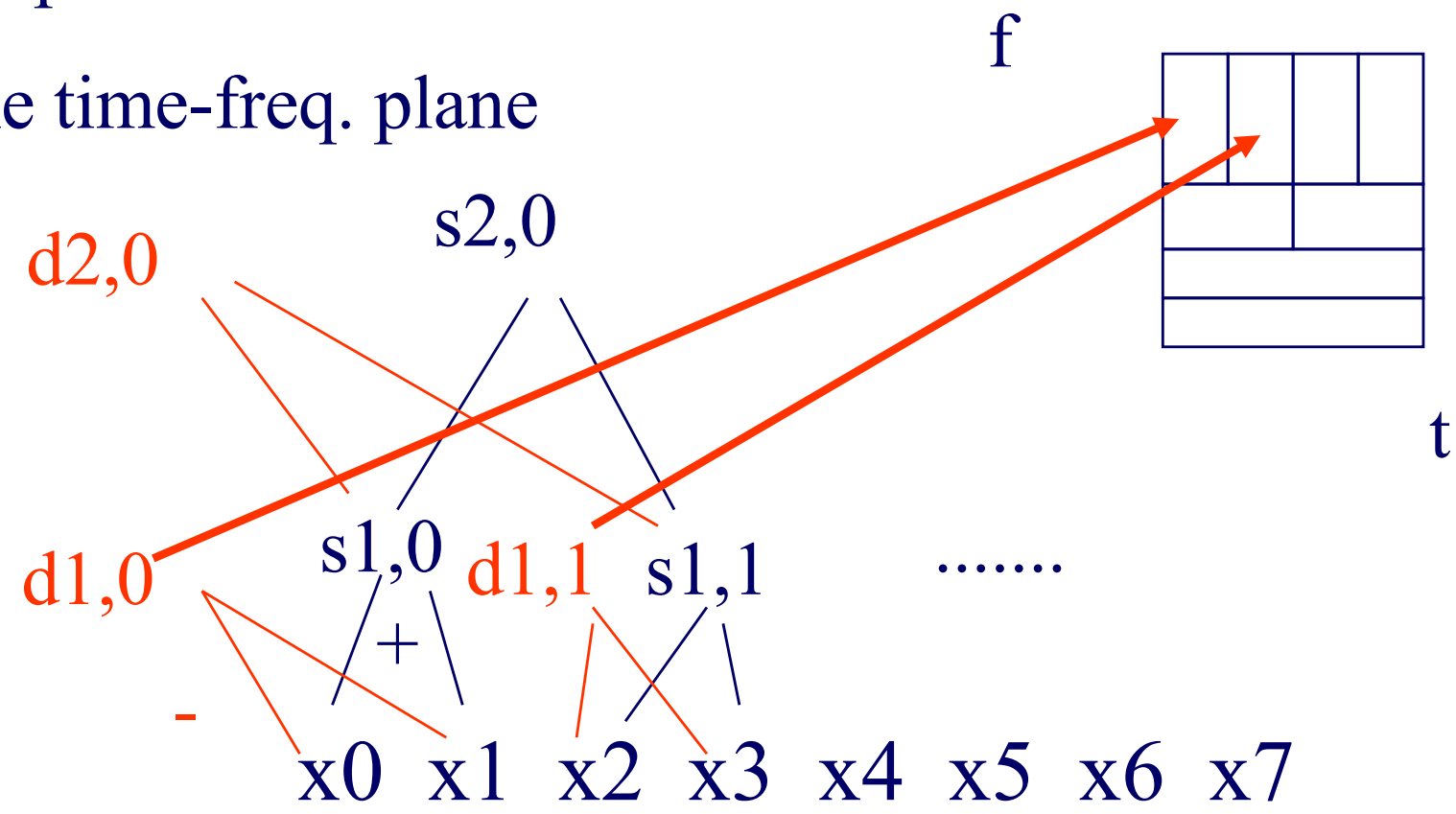




Wavelets - construction

Q: map each coefficient

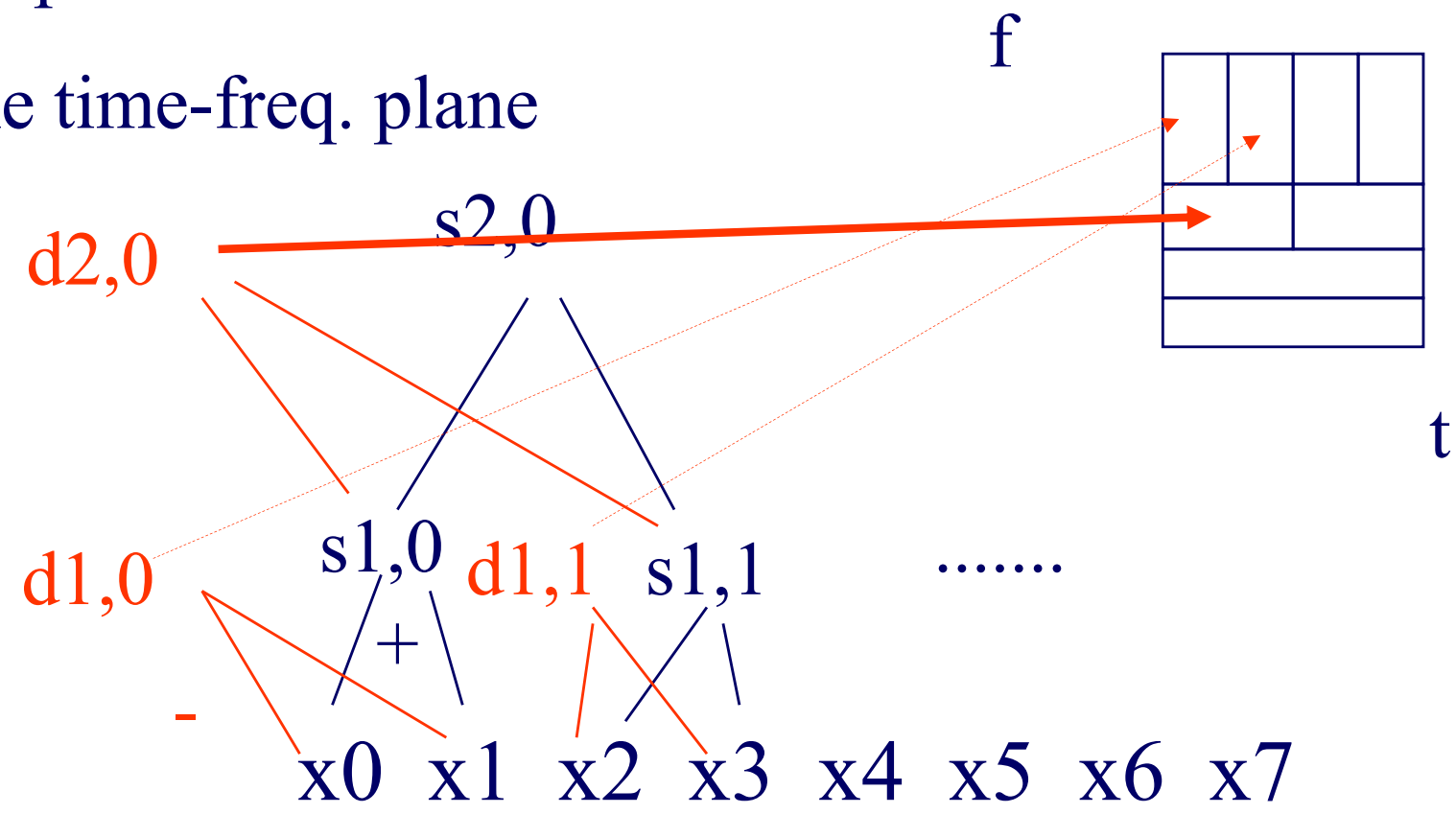
on the time-freq. plane



Wavelets - construction

Q: map each coefficient

on the time-freq. plane



Carnegie Mellon

Haar wavelets - code

```
#!/usr/bin/perl5
# expects a file with numbers
# and prints the dwt transform
# The number of time-ticks should be a power of 2
# USAGE
#    haar.pl <fname>

my @vals=();
my @smooth; # the smooth component of the signal
my @diff;   # the high-freq. component

# collect the values into the array @val
while(<>){
        @vals = ( @vals ,  split );
}
```

```
my $len = scalar(@vals);
my $half = int($len/2);
while($half >= 1 ){
  for(my $i=0; $i< $half; $i++){
        $diff [$i] = ($vals[2*$i] - $vals[2*$i + 1] )/ sqrt(2);
        print "\t", $diff[$i];
        $smooth [$i] = ($vals[2*$i] + $vals[2*$i + 1] )/ sqrt(2);
  }
  print "\n";
  @vals = @smooth;
  $half = int($half/2);
}
print "\t", $vals[0], "\n" ;      # the final, smooth component
```

Also at: www.cs.cmu.edu/~christos/SRC/DWT-Haar-all.tar

Wavelets - construction

Observation1:

'+' can be some weighted addition

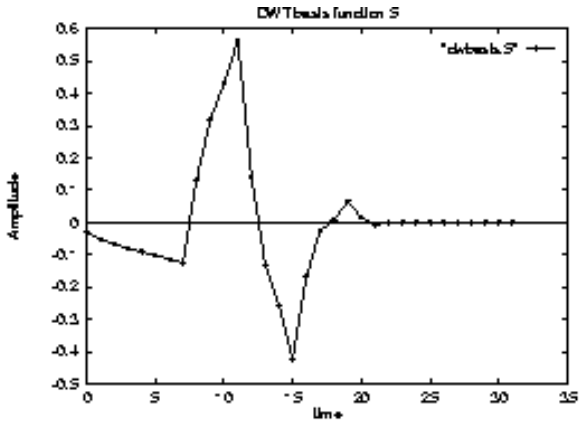
'-' is the corresponding weighted difference ('Quadrature mirror filters')

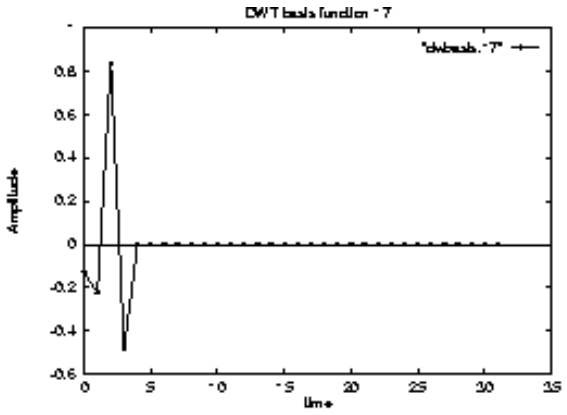
Observation2: unlike DFT/DCT,

there are *many* wavelet bases: Haar, Daubechies-4, Daubechies-6, Coifman, Morlet, Gabor, ...

Carnegie Mellon

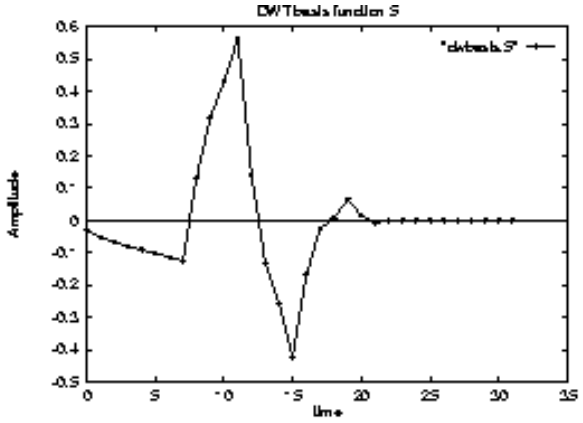
Wavelets - how do they look like?

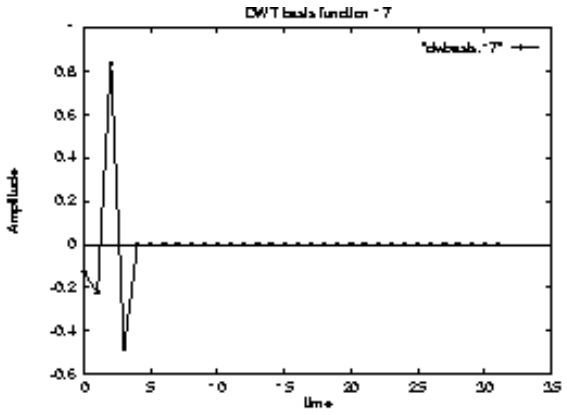




- E.g., Daubechies-4

Wavelets - how do they look like?

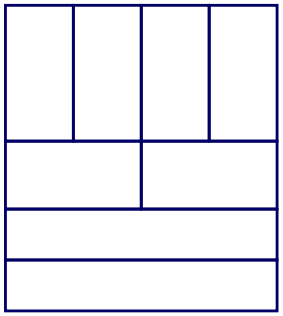




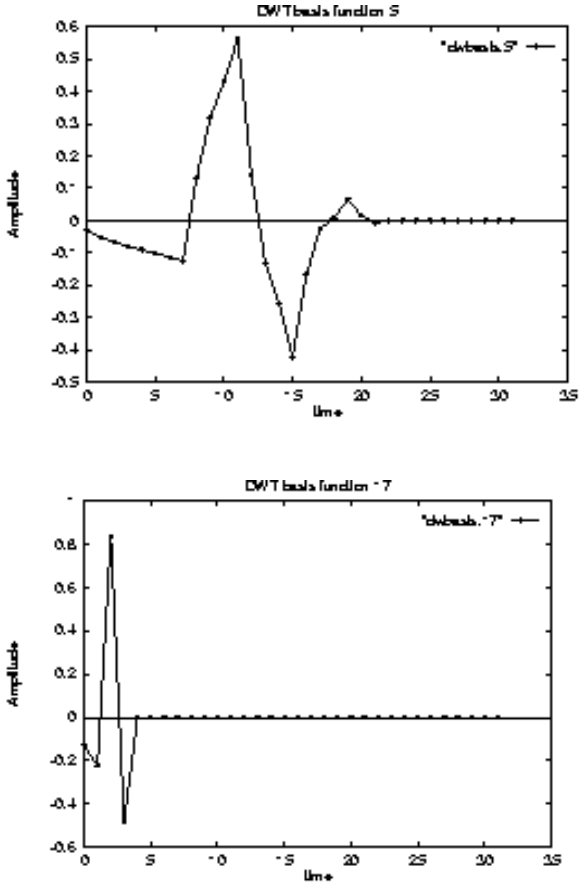
- E.g., Daubechies-4

?

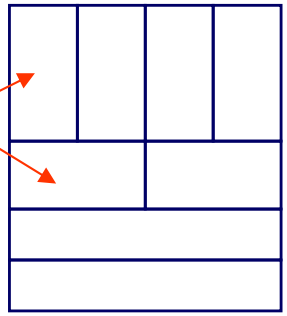
?



Wavelets - how do they look like?

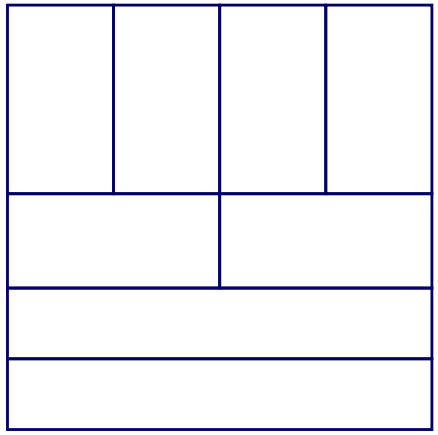


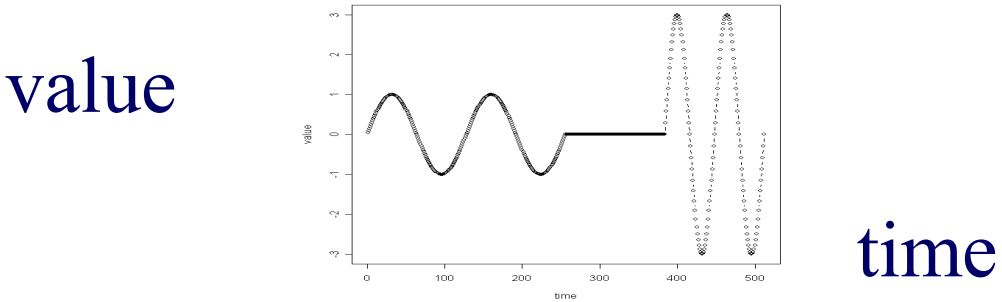
- E.g., Daubechies-4



Wavelets - Drill#1:

- Q: baritone/silence/soprano - DWT?

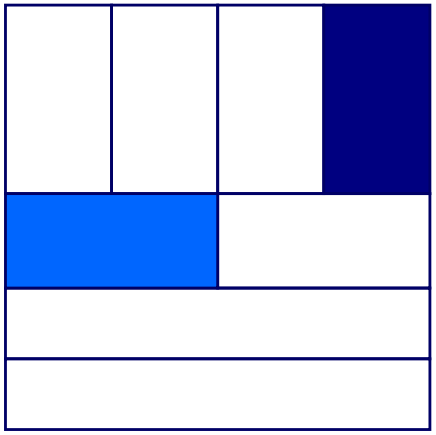




Wavelets - Drill#1:

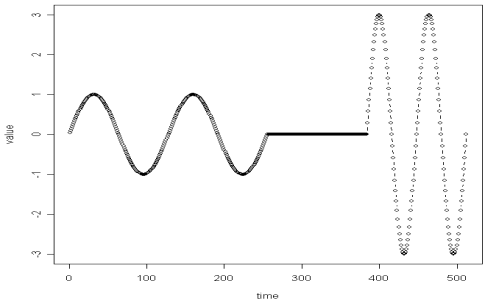
- Q: baritone/silence/soprano - DWT?

f



t

value

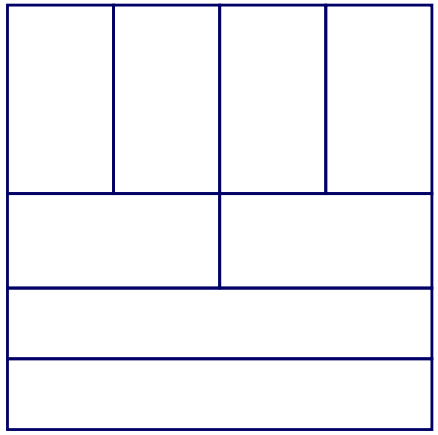


time

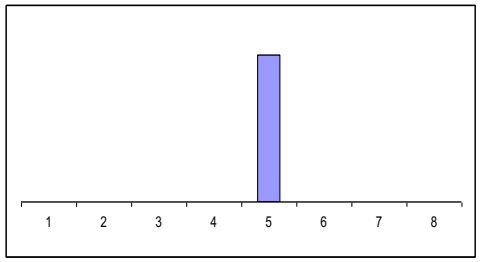
Wavelets - Drill#2:

- Q: spike - DWT?

f



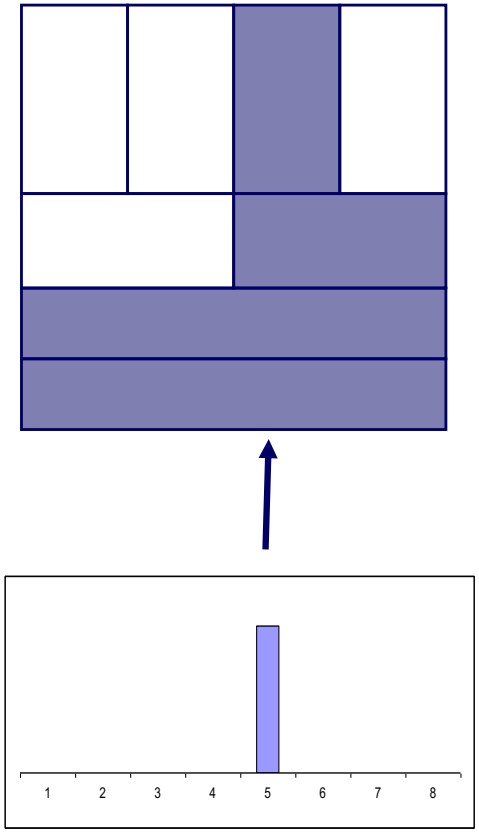
t



Wavelets - Drill#2:

- Q: spike - DWT?

f



t

0.00 0.00 **0.71** 0.00

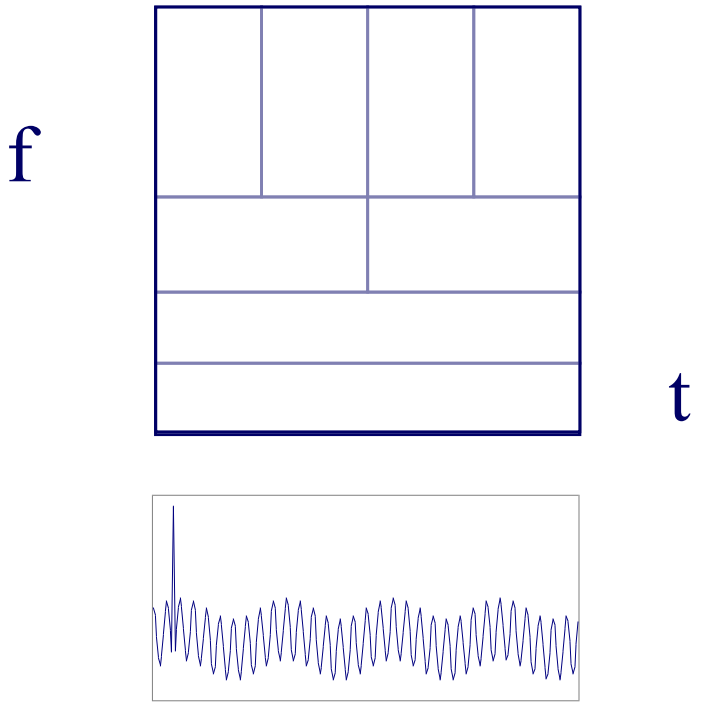
0.00 **0.50**

-0.35

0.35

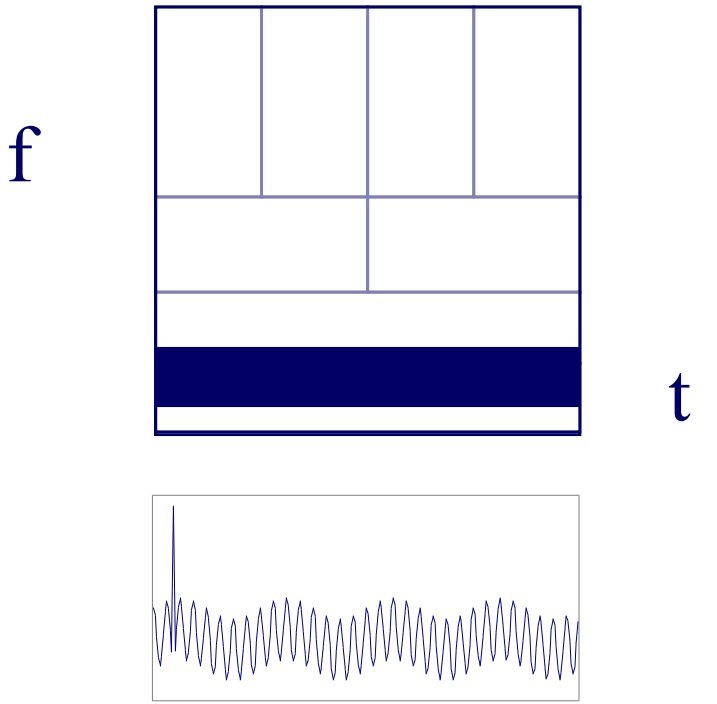
Wavelets - Drill#3:

- Q: weekly + daily periodicity, + spike - DWT?



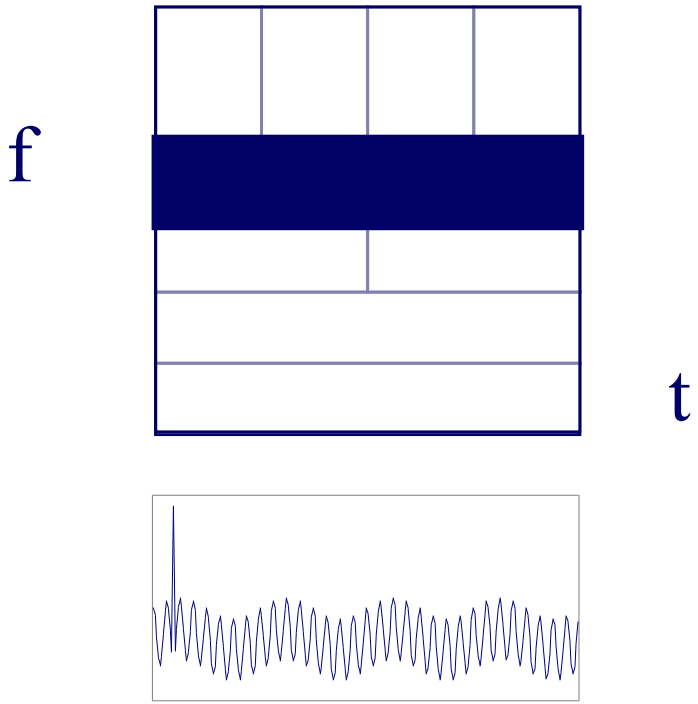
Wavelets - Drill#3:

- Q: **weekly** + daily periodicity, + spike - DWT?



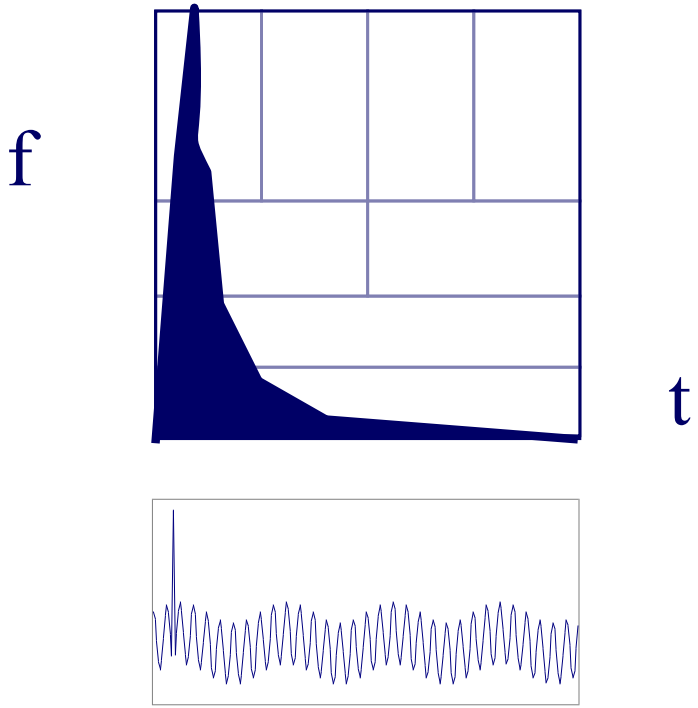
Wavelets - Drill#3:

- Q: weekly + **daily** periodicity, + spike - DWT?



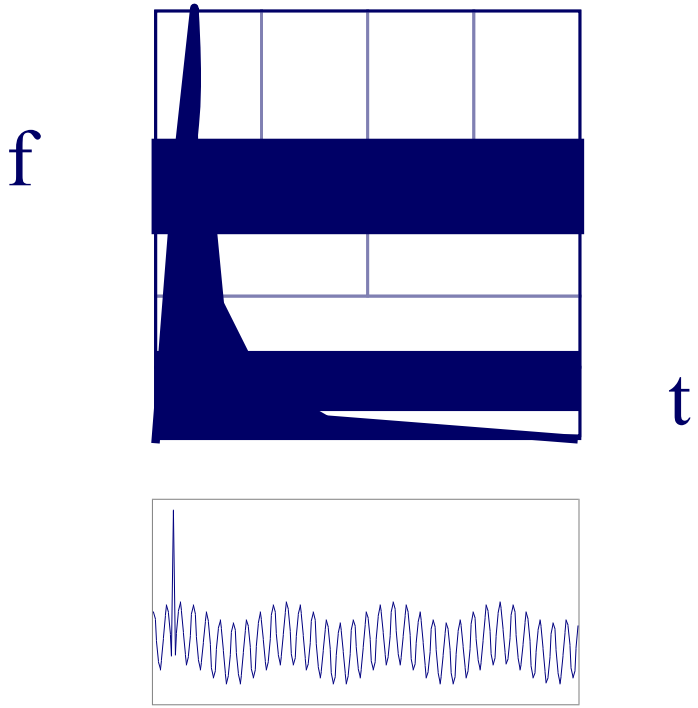
Wavelets - Drill#3:

- Q: weekly + daily periodicity, + **spike** - DWT?



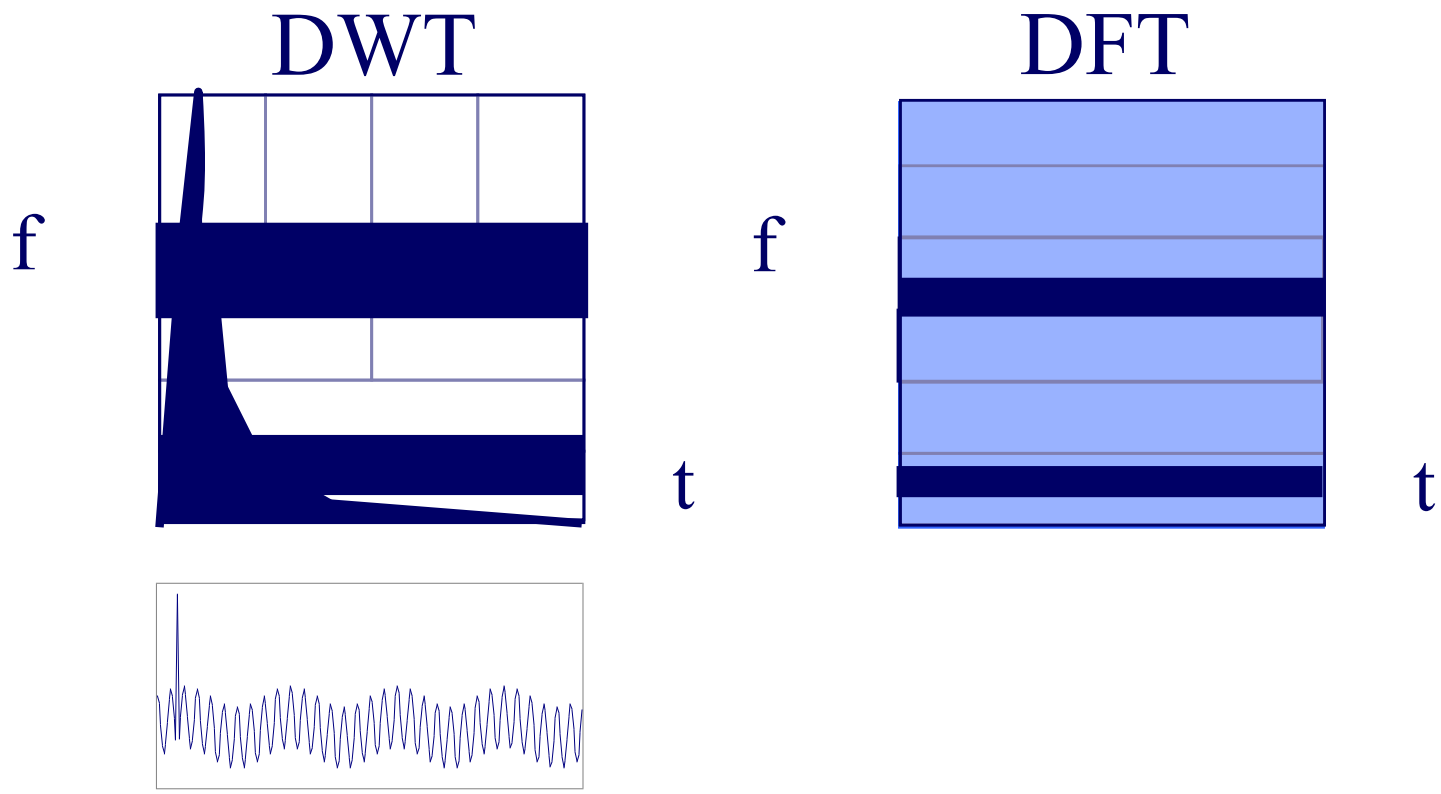
Wavelets - Drill#3:

- Q: weekly + daily periodicity, + spike - DWT?



Wavelets - Drill#3:

- Q: DFT?



Wavelets - Drill:

Let's see it live:

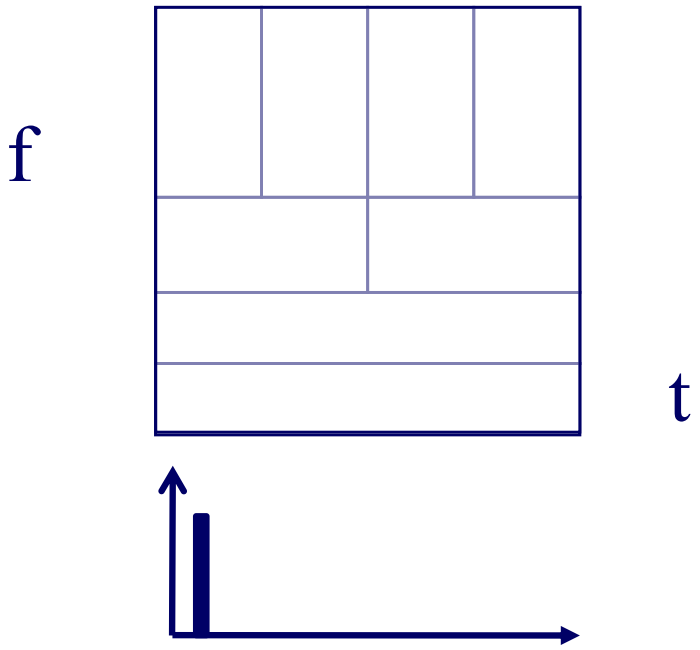
http://dsp.rice.edu/software/dsp-teaching-tools

delta; cosine; cosine2; chirp

- Haar vs Daubechies-4, -6, etc

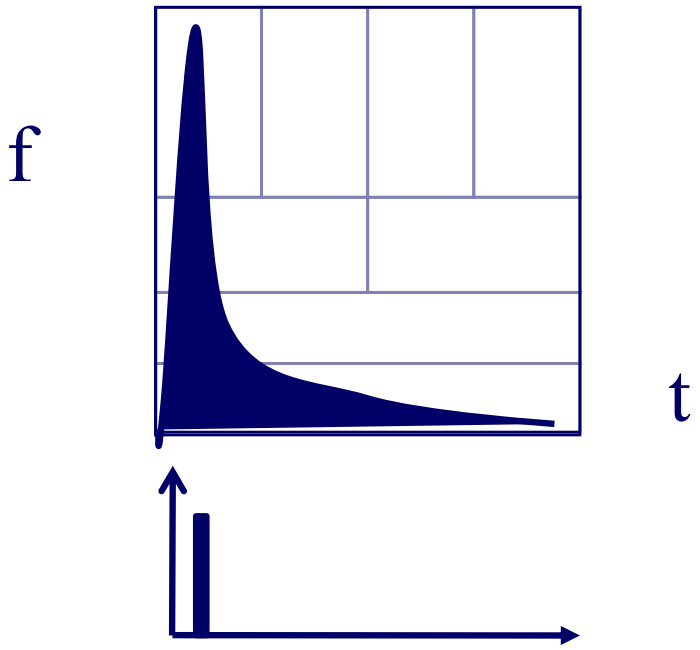
Delta?

x(0)=1; x(t)= 0 elsewhere



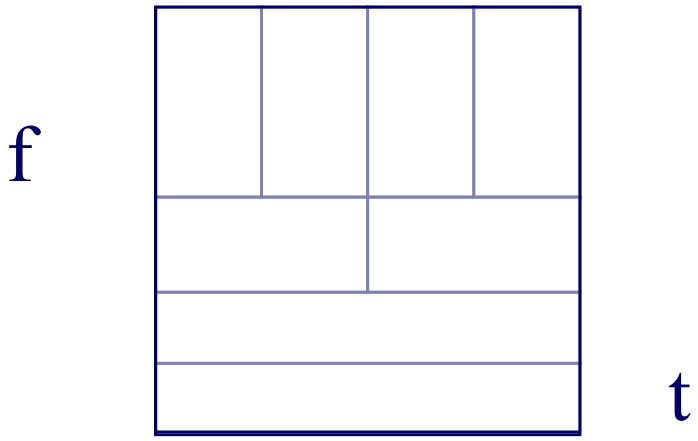
Delta?

x(0)=1; x(t)= 0 elsewhere



2 cosines?

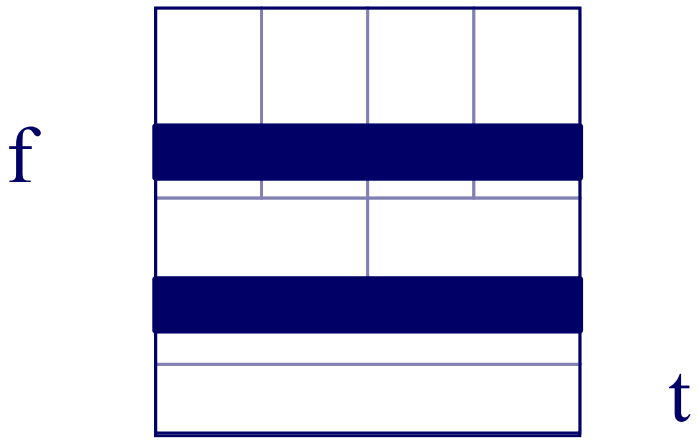
$$x(t) = \cos(2 * pi * 4 * t / 1024) + 5 * \cos(2 * pi * 8 * t / 1024)$$



Carnegie Mellon

2 cosines?

$$x(t) = \cos(2 * pi * 4 * t / 1024) + 5 * \cos(2 * pi * 8 * t / 1024)$$



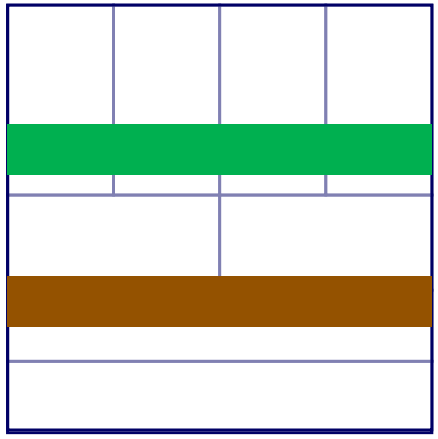
Which one
is for freq.=4?

2 cosines?

$$x(t) = \cos(2 * pi * 4 * t / 1024) + 5 * \cos(2 * pi * 8 * t / 1024)$$

f~8 ⟶

f~4 ⟶

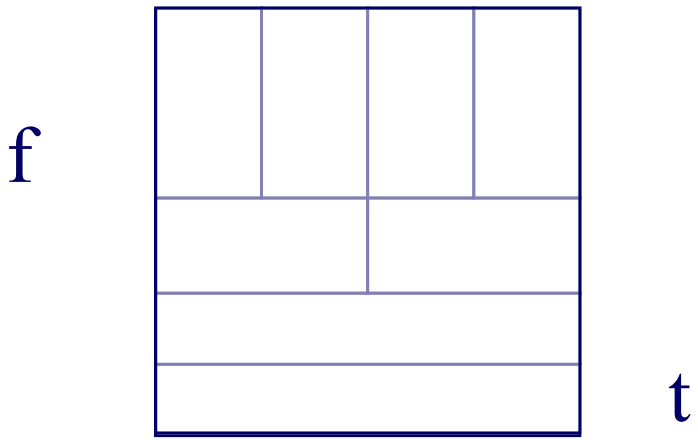


Which one is for freq.=4?

Carnegie Mellon

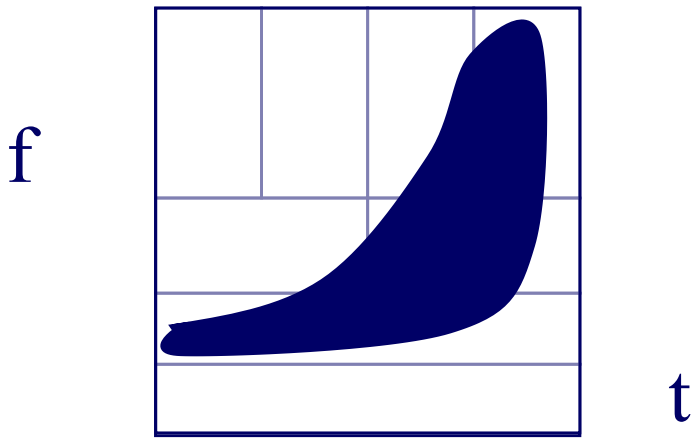
Chirp?

$$x(t) = \cos(2 * pi * t * t / 1024)$$



Chirp?

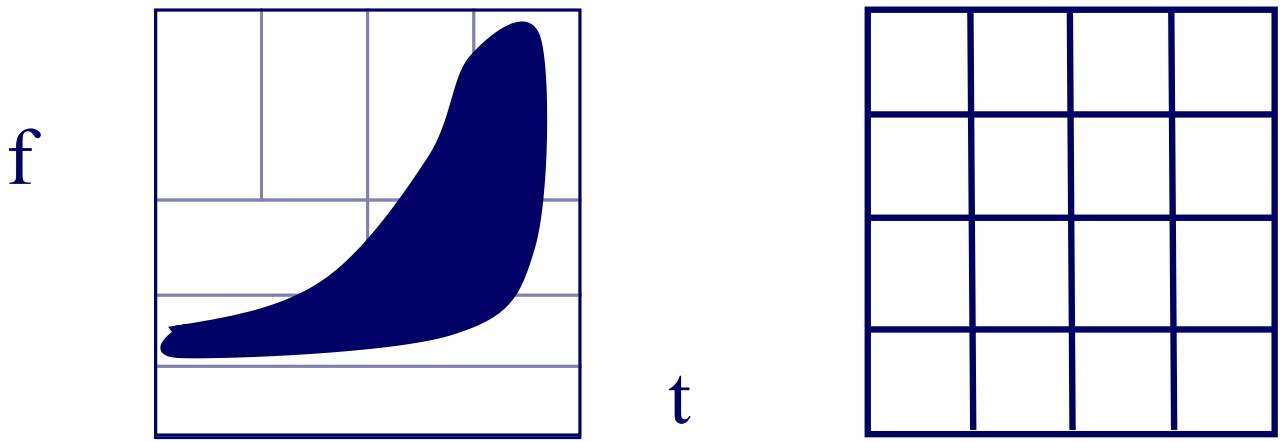
$$x(t) = \cos(2 * pi * t * t / 1024)$$



Chirp?

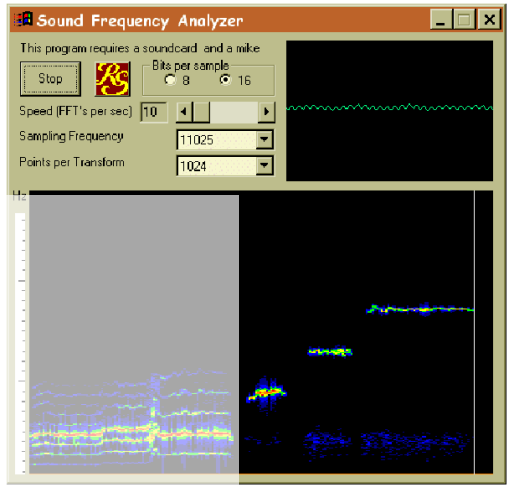
$$x(t) = \cos(2 * pi * t * t / 1024)$$

SWFT?

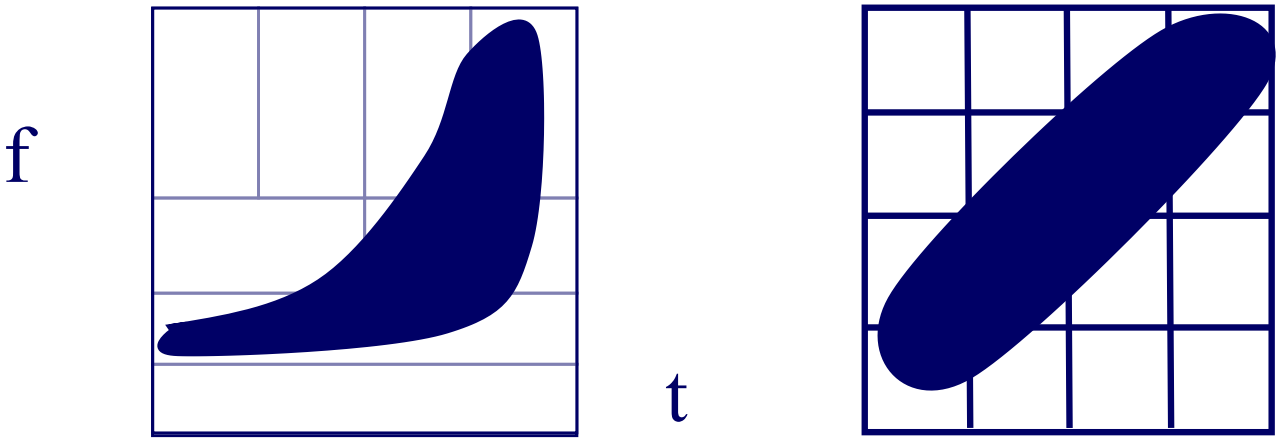


Chirp?

$$x(t) = \cos(2 * pi * t * t / 1024)$$



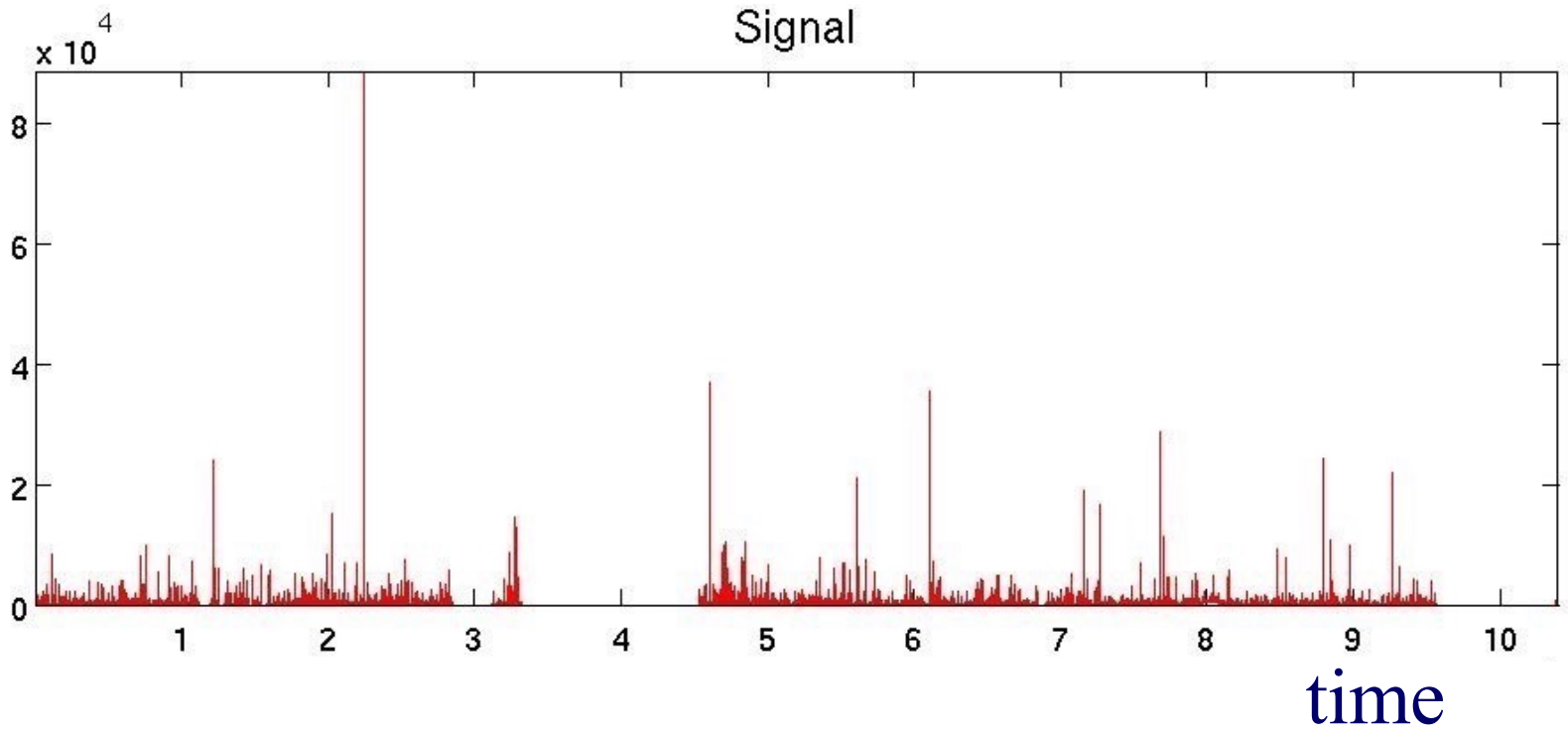
SWFT



Carnegie Mellon

More examples (BGP updates)

#updates

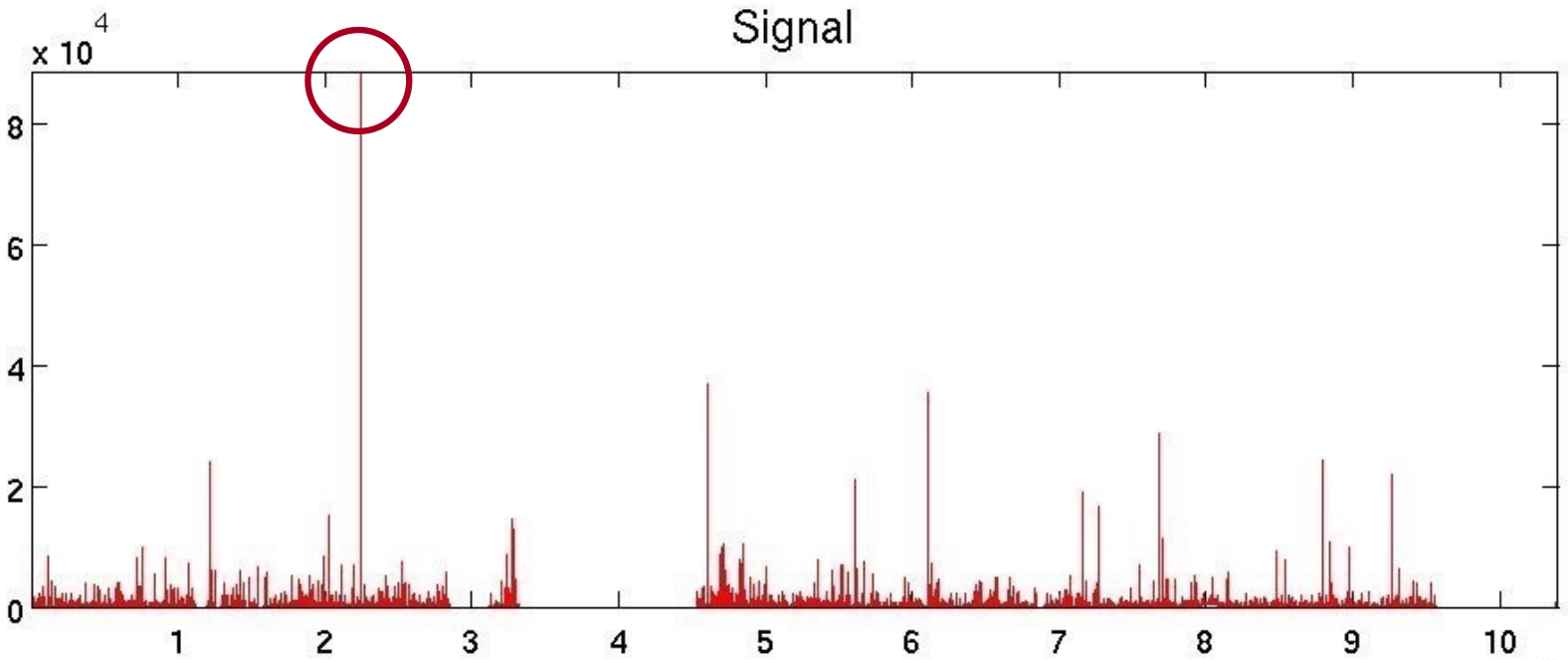


BGP-lens: Patterns and Anomalies in Internet Routing Updates B. Aditya Prakash et al, SIGKDD 2009

Carnegie Mellon

More examples (BGP updates)

#updates

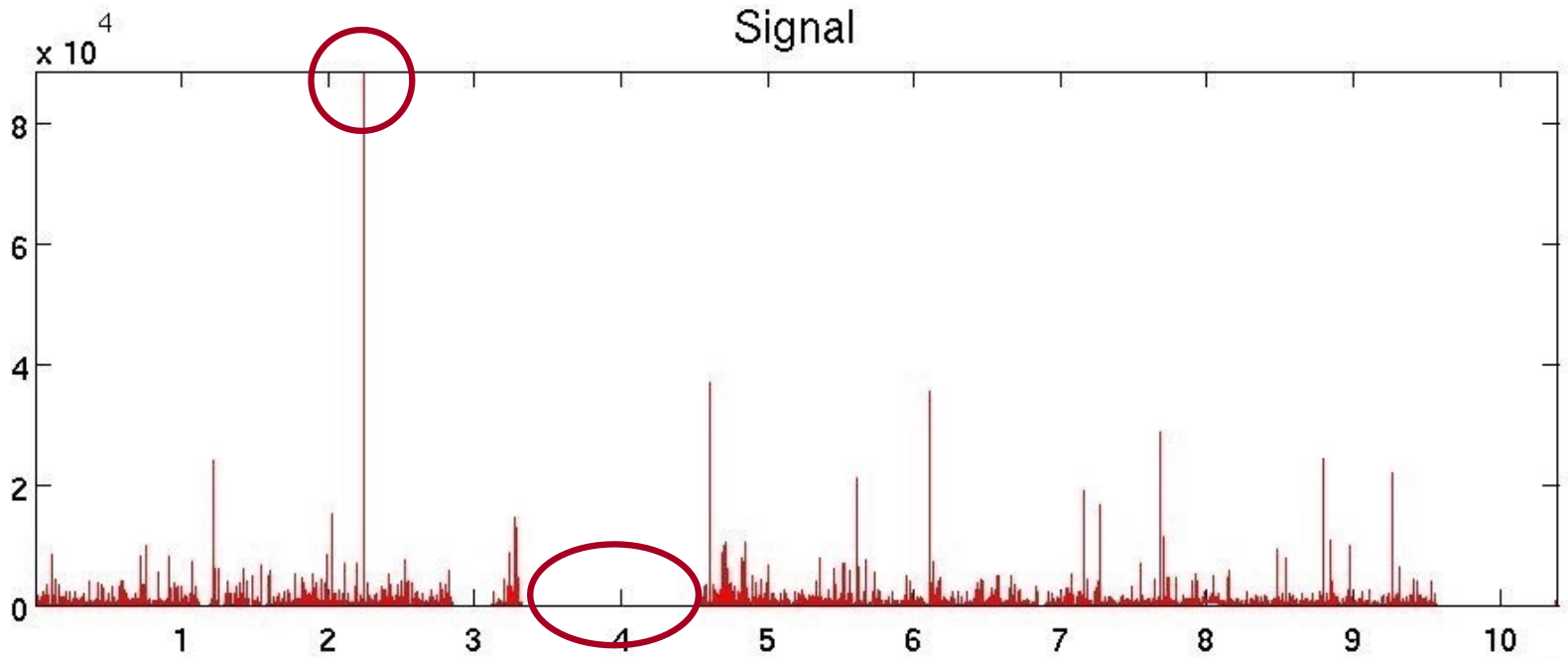


BGP-lens: Patterns and Anomalies in Internet Routing Updates B. Aditya Prakash et al, SIGKDD 2009

Carnegie Mellon

More examples (BGP updates)

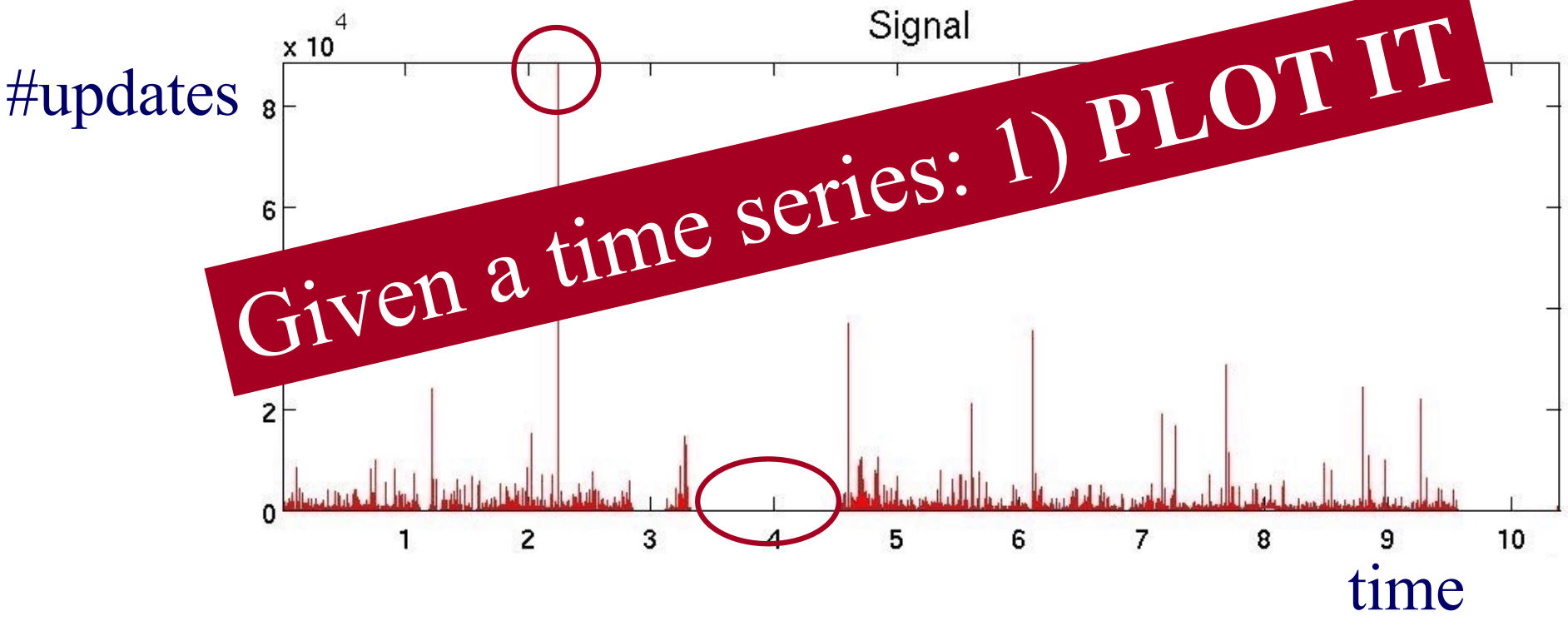
#updates



BGP-lens: Patterns and Anomalies in Internet Routing Updates B. Aditya Prakash et al, SIGKDD 2009

Carnegie Mellon

More examples (BGP updates)



BGP-lens: Patterns and Anomalies in Internet Routing Updates B. Aditya Prakash et al, SIGKDD 2009

More examples (BGP updates)

Given a time series: 1) PLOT IT

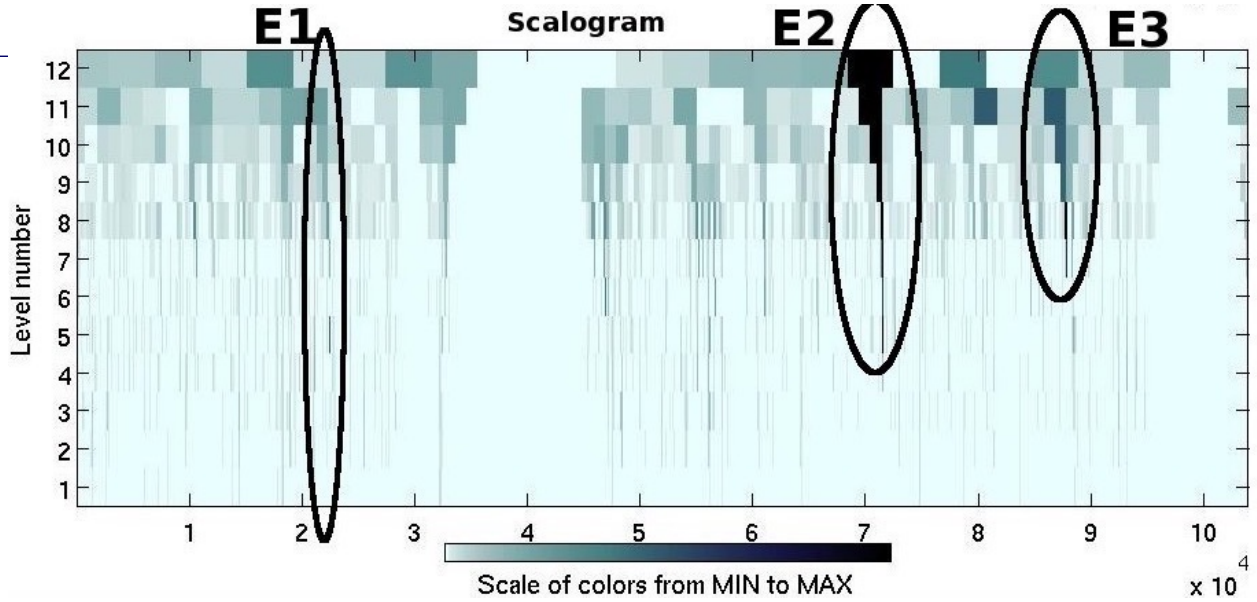
2) DFT / DWT

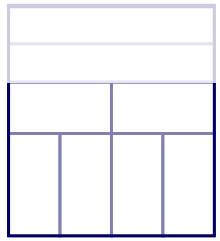
Carnegie Mellon

More examples (BGP updates)

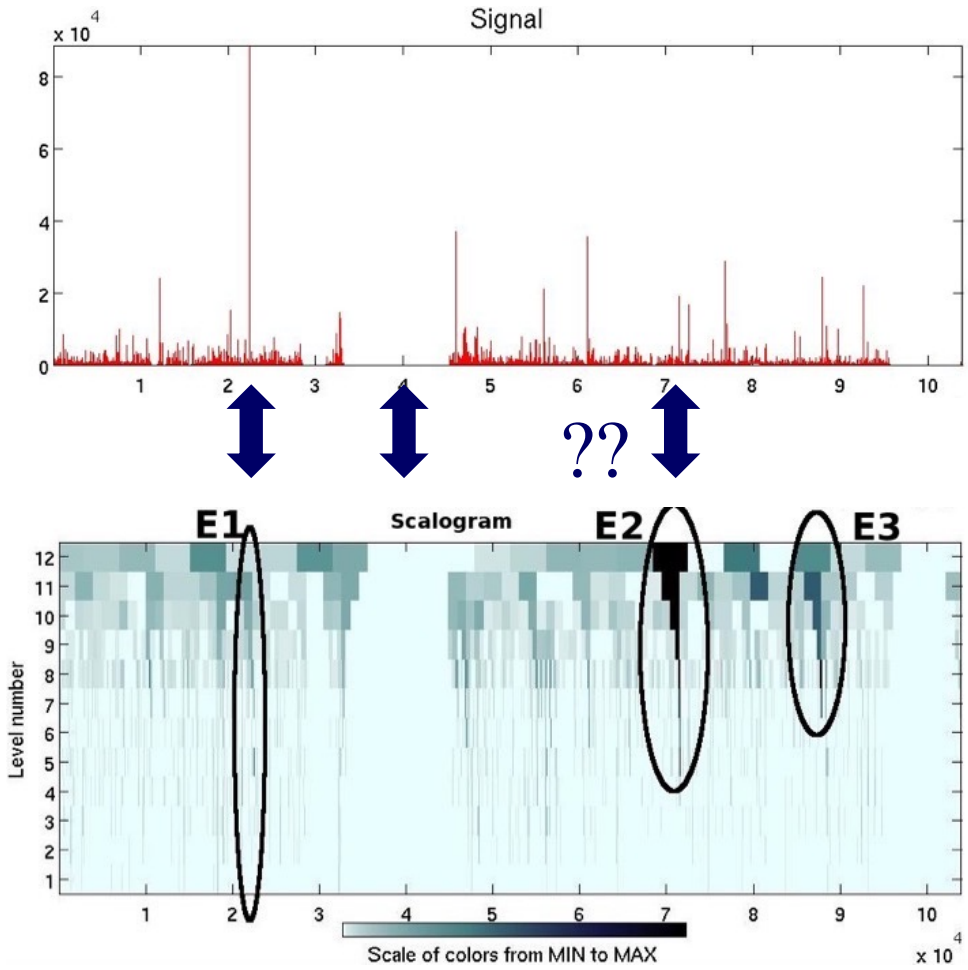
Low freq.:
omitted

freq.





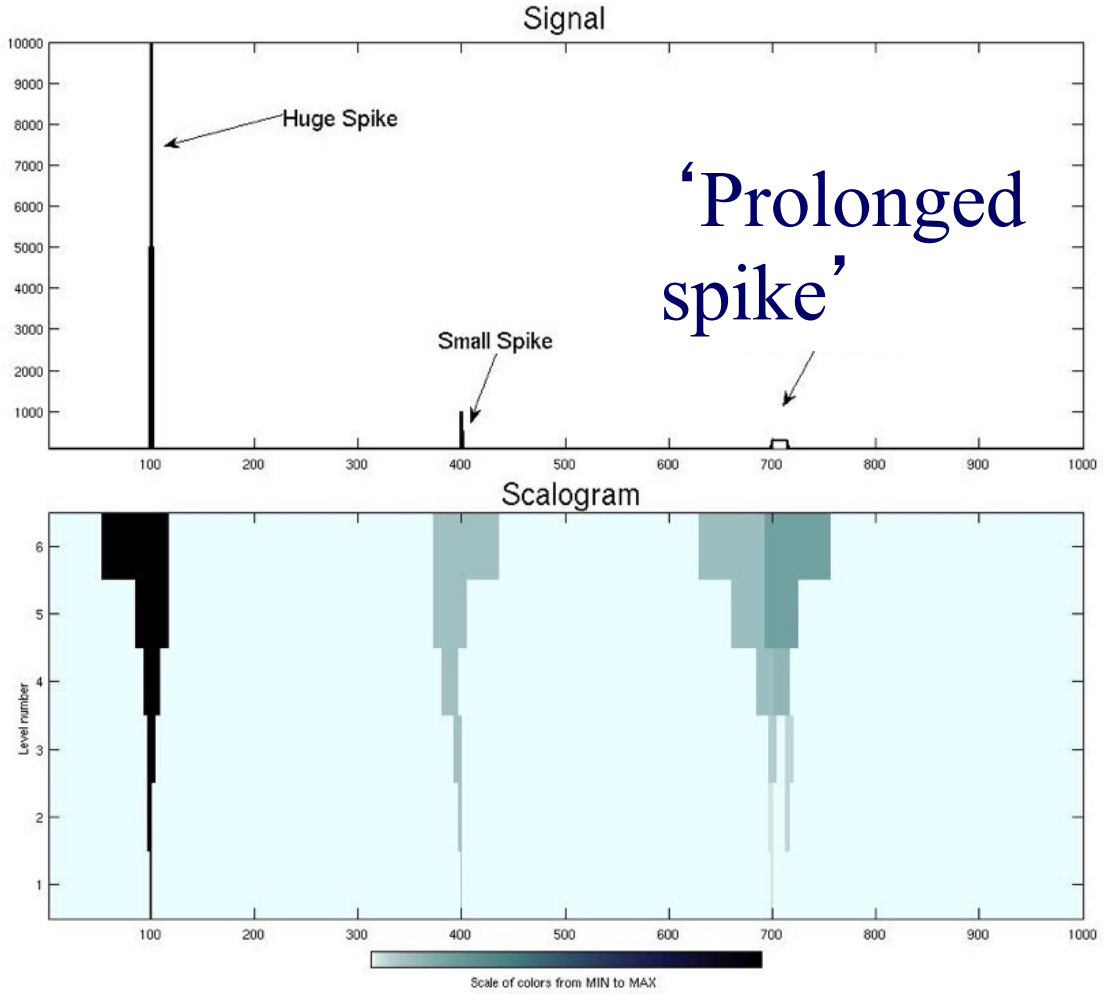
More examples (BGP updates)



freq.

Carnegie Mellon

More examples (BGP updates)



freq.

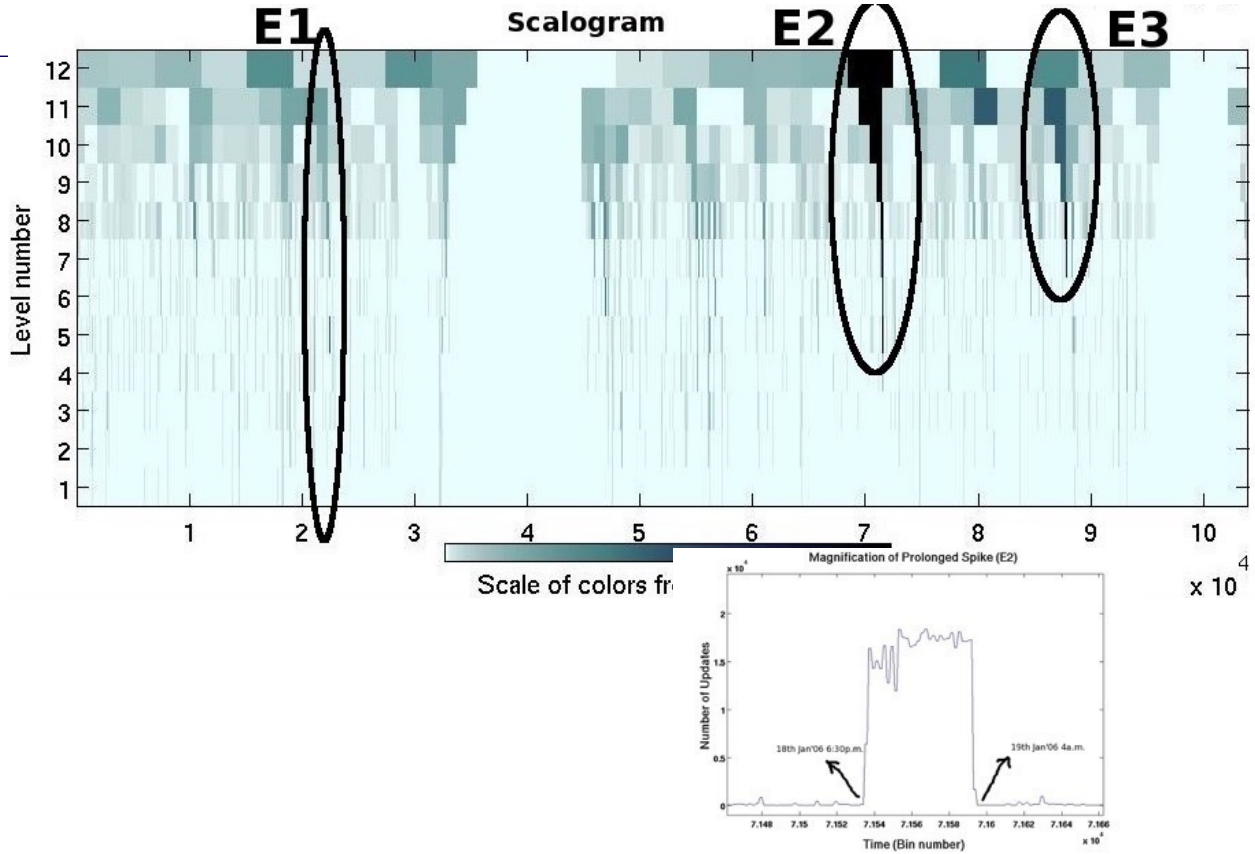
time

Carnegie Mellon

More examples (BGP updates)

15K msgs, for several hours: 6pm-4am

freq.

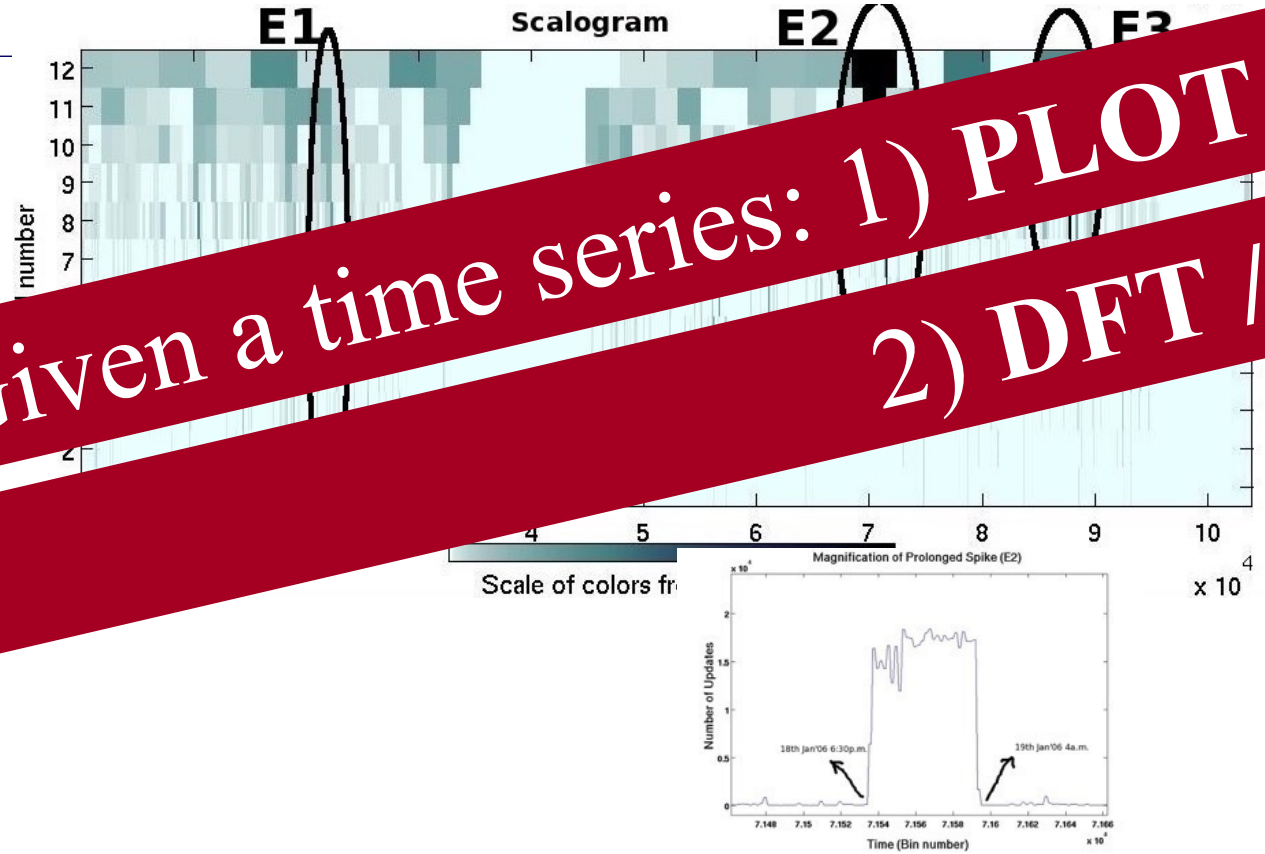


Carnegie Mellon

More examples (BGP updates)

15K msgs, for several hours: 6pm-4am

freq.



Given a time series: 1) PLOT IT

2) DFT / DWT

Wavelets - Drill

- Or use 'R', 'octave' or 'matlab' – R:

```
install.packages("wavelets")
library("wavelets")
X1<-c(1,2,3,4,5,6,7,8)
dwt(X1, n.levels=3, filter="d4")
mra(X1, n.levels=3, filter="d4")
```

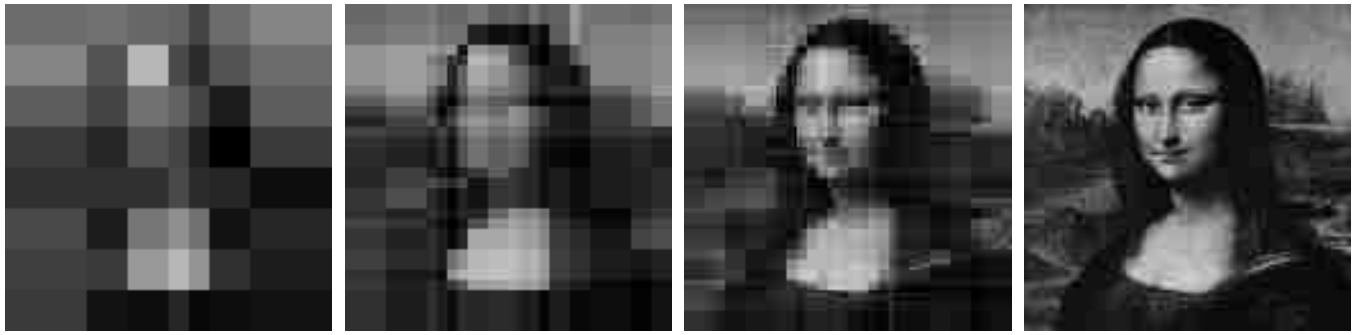
Wavelets - k-dimensions?

- easily defined for any dimensionality (like DFT, DCT)

Wavelets - example

http://grail.cs.washington.edu/projects/query/

Wavelets achieve *great* compression:

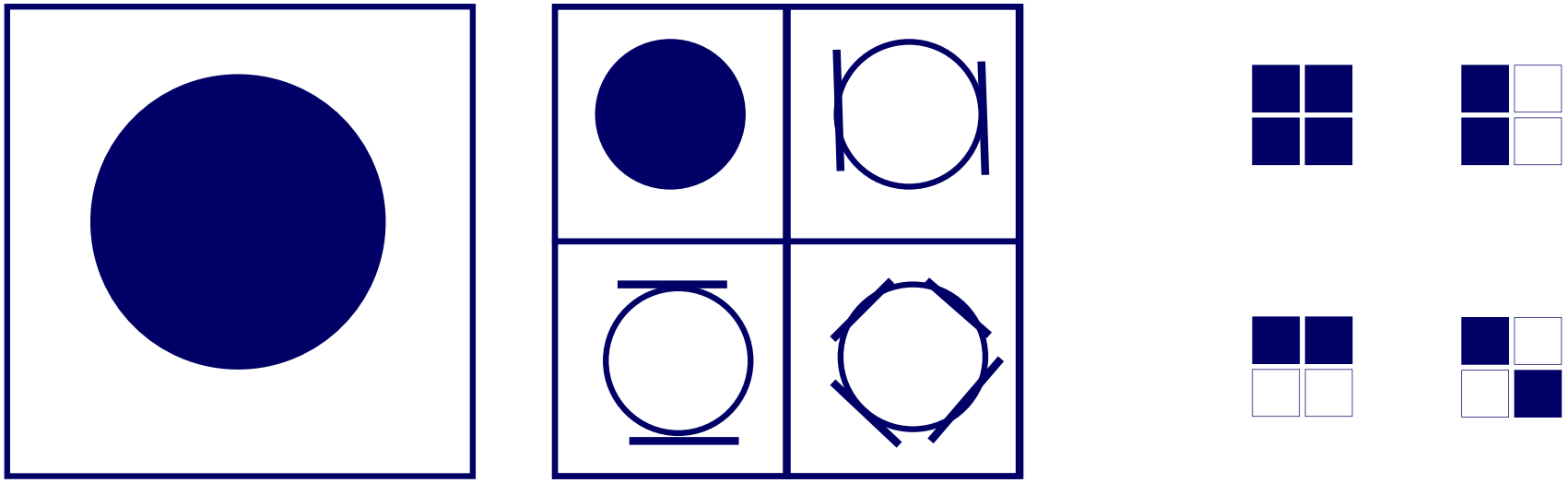


20 100 400 16,000

coefficients

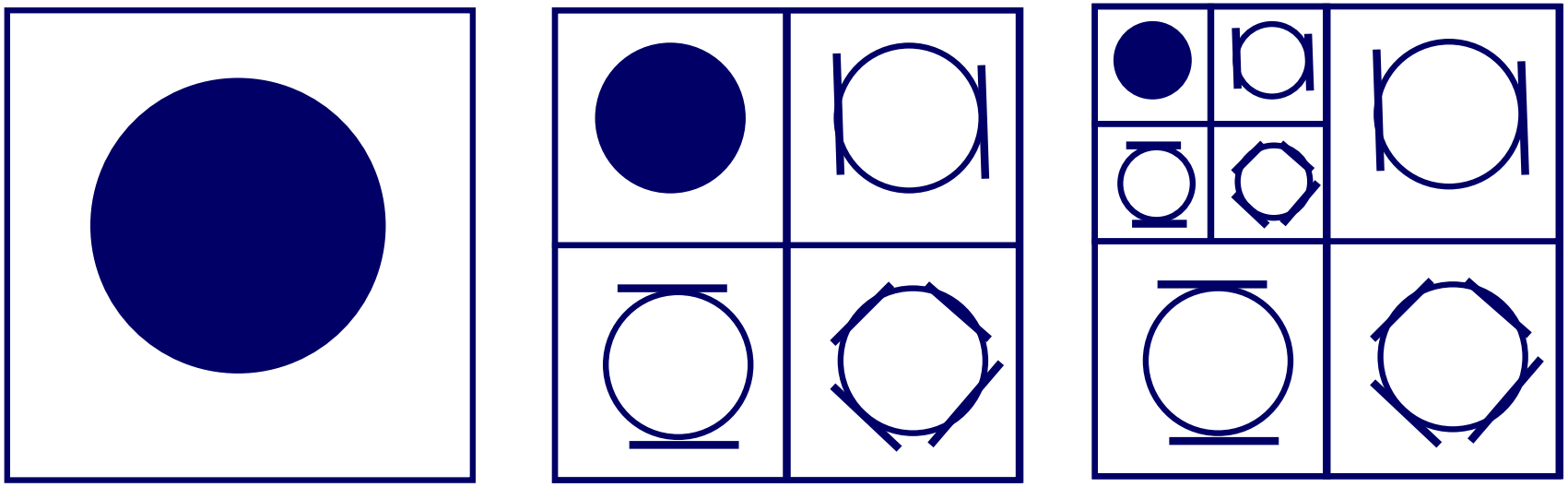
Wavelets - intuition

- Edges (horizontal; vertical; diagonal)



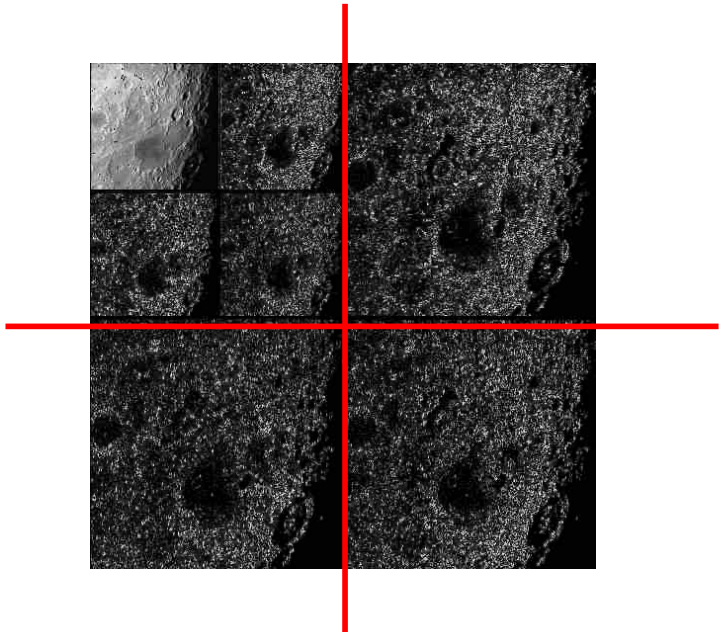
Wavelets - intuition

- Edges (horizontal; vertical; diagonal)
- recurse



Wavelets - intuition

- Edges (horizontal; vertical; diagonal)
- http://www331.jpl.nasa.gov/public/wave.html

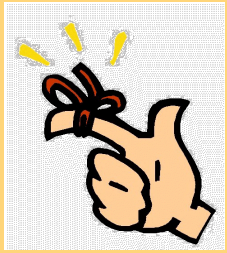


Advantages of Wavelets

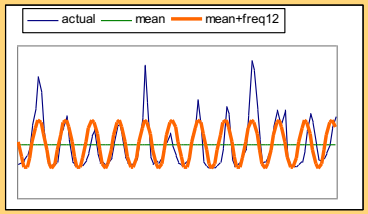
- Better compression (better RMSE with same number of coefficients)
- closely related to the processing of the mammalian eye and ear

- Good for progressive transmission

- handle spikes well
- usually, fast to compute (O(n)!)

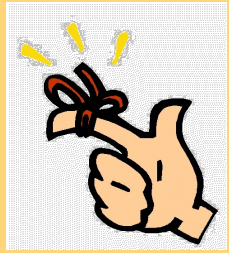


Overall Conclusions



- DFT spots periodicities
- **DWT** : multi-resolution - matches processing of mammalian ear/eye better

- Both: powerful tools for **compression**, **pattern detection** in real signals
- Both: included in math packages (matlab, R, python)

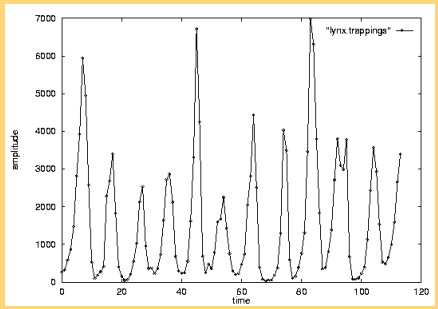


Solutions:

Goal: given a signal (eg., sales over time and/or space)

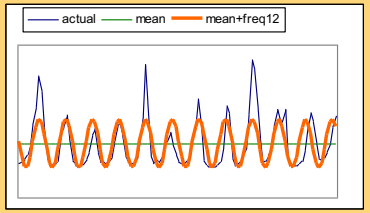
Q: Find patterns and/or compress

count

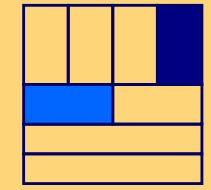


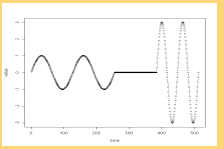
year

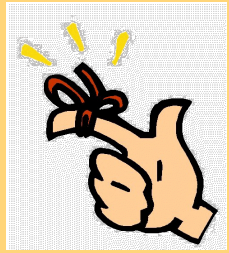
✔ A1: Fourier (DFT)



✔ A2: Wavelets (DWT)





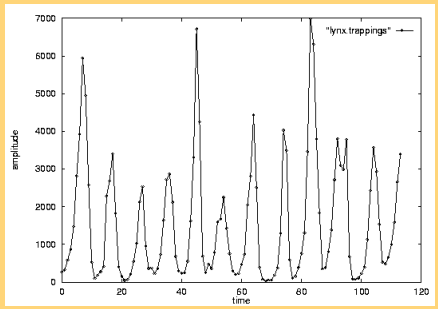


Solutions:

Goal: given a signal (eg., sales over time and/or space)

Q: Find patterns and/or compress

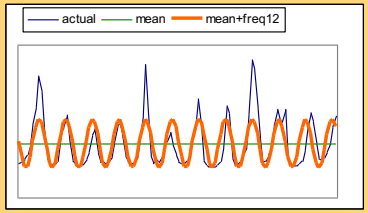
count



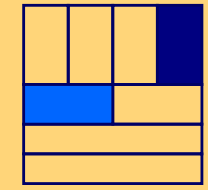
year

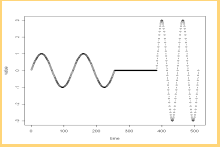
A0: PLOT IT

✔ A1: Fourier (DFT)



✔ A2: Wavelets (DWT)





Resources

- Numerical Recipes in C: great description, intuition and code for all three tools
- *xwpl:* open source wavelet package from Yale, with excellent GUI.

Resources (cont'd)

- www-dsp.rice.edu/software/EDU/mra.shtml (wavelets and other demos)
- R ('install.packages("wavelets"))