

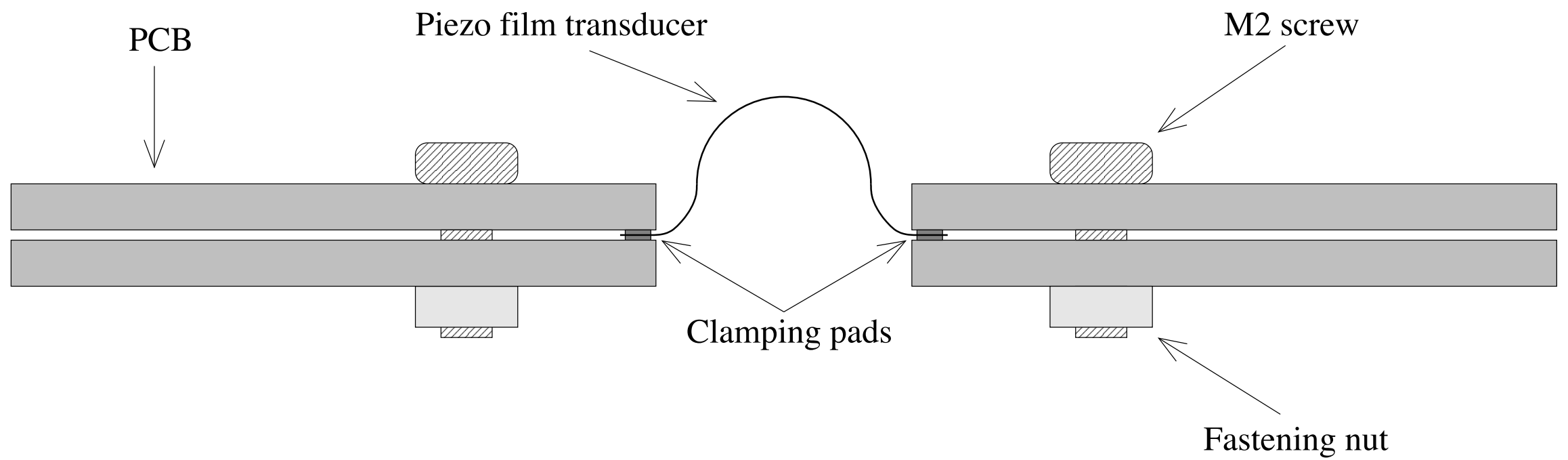
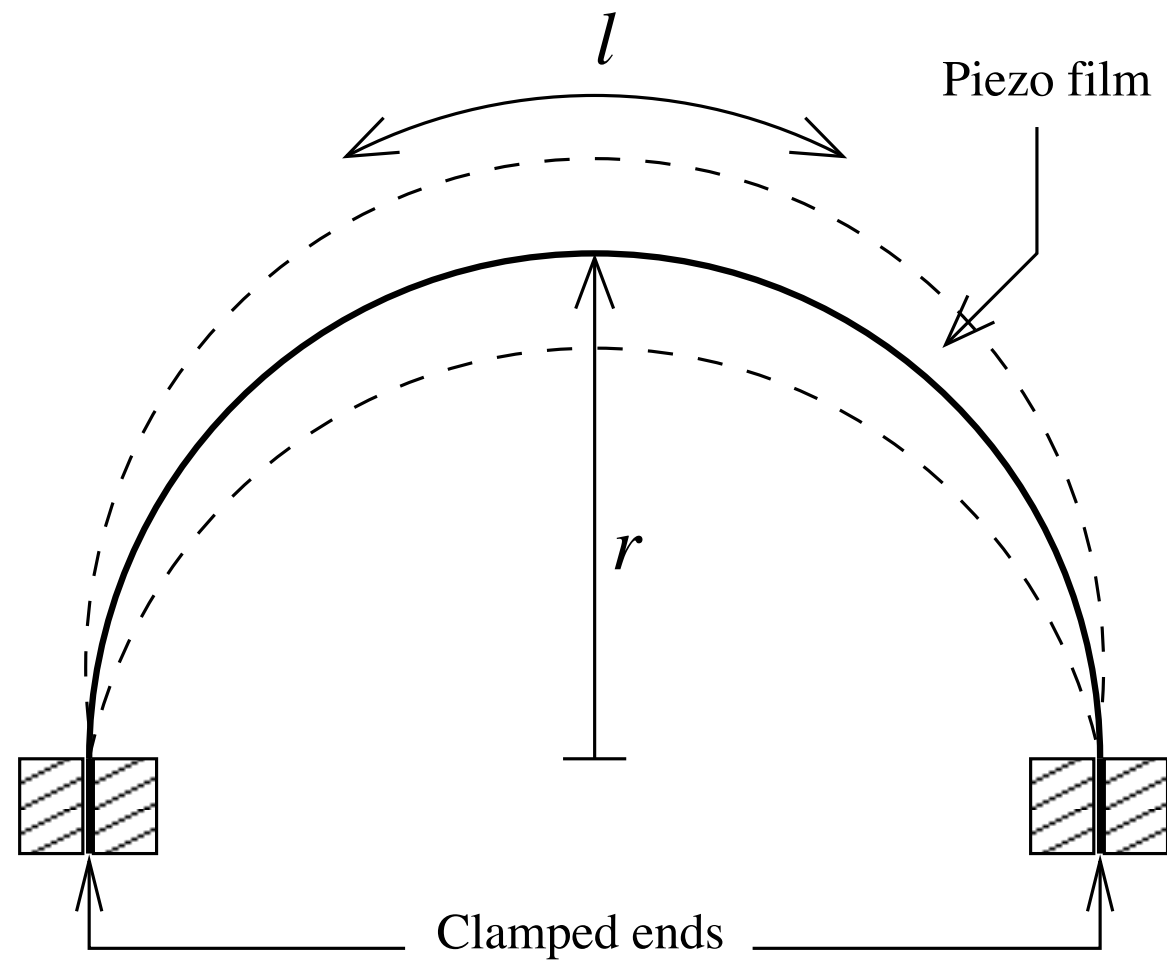
# Acoustic Localisation: New Developments

Mike Hazas  
Lancaster University, UK

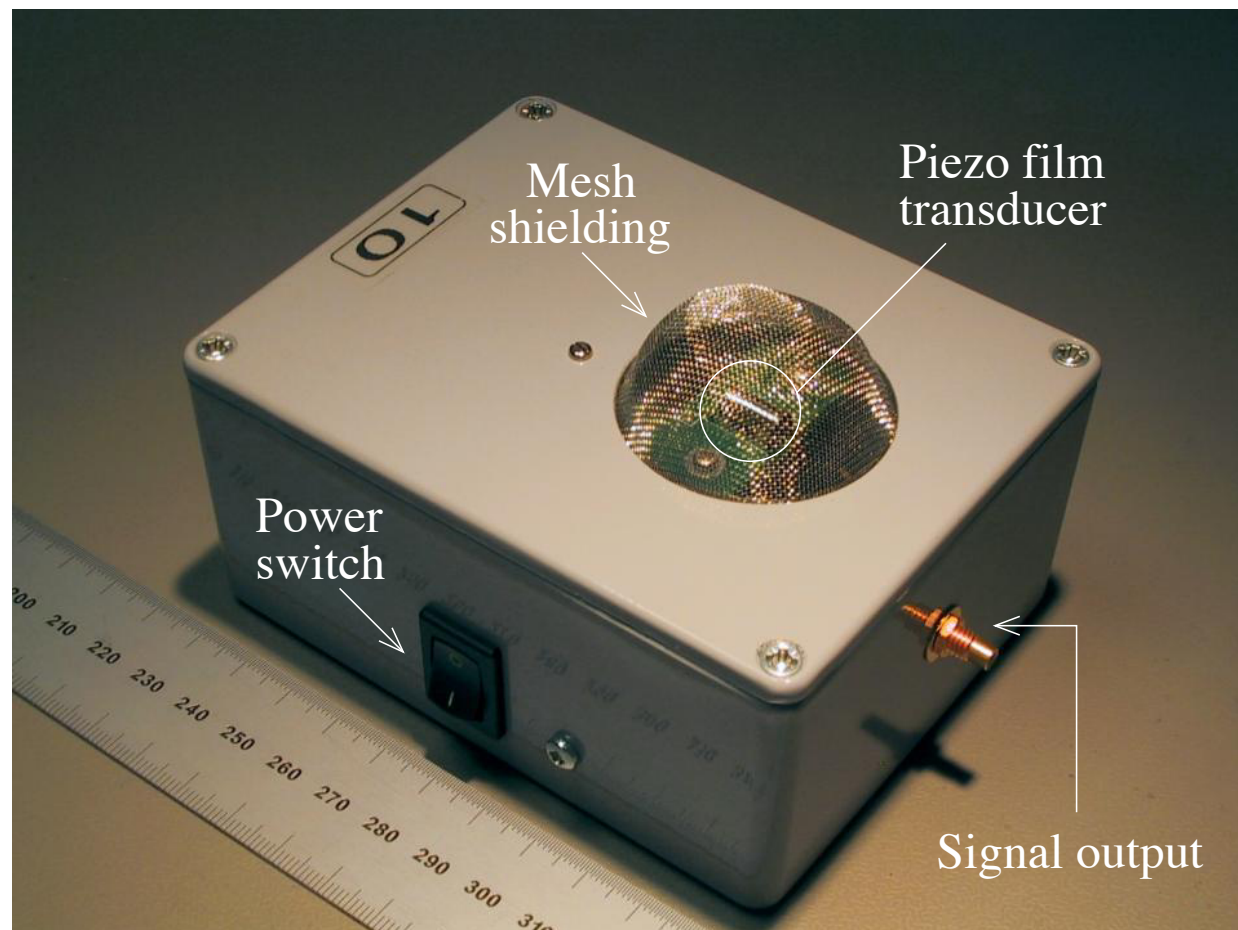
IPIN 2011, Fri 23 Sept

# Introduction

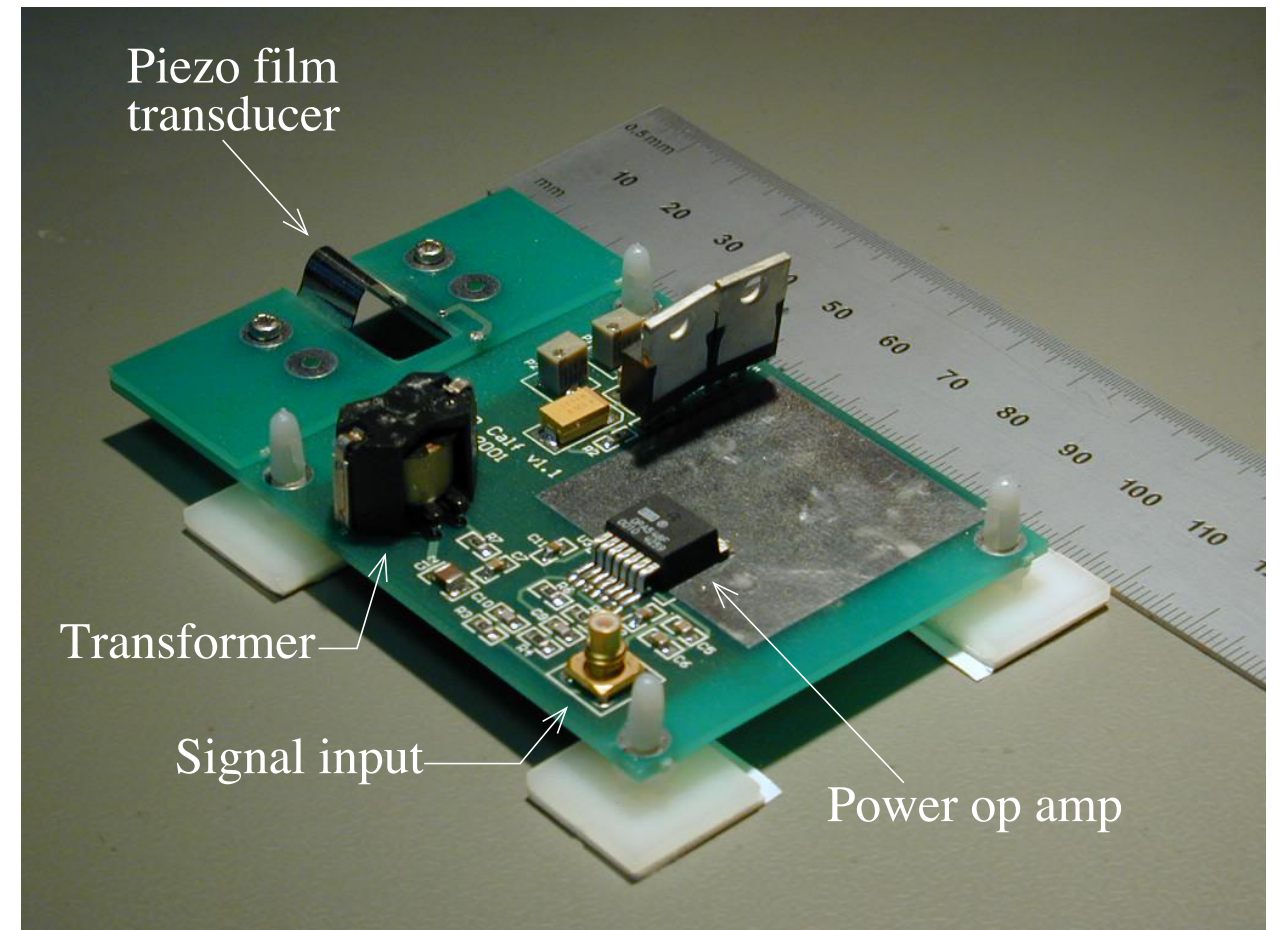
- material in this presentation is that of two PhD students:
  - Mohammed Alloulah
  - Yukang Guo



# Piezo film prototypes



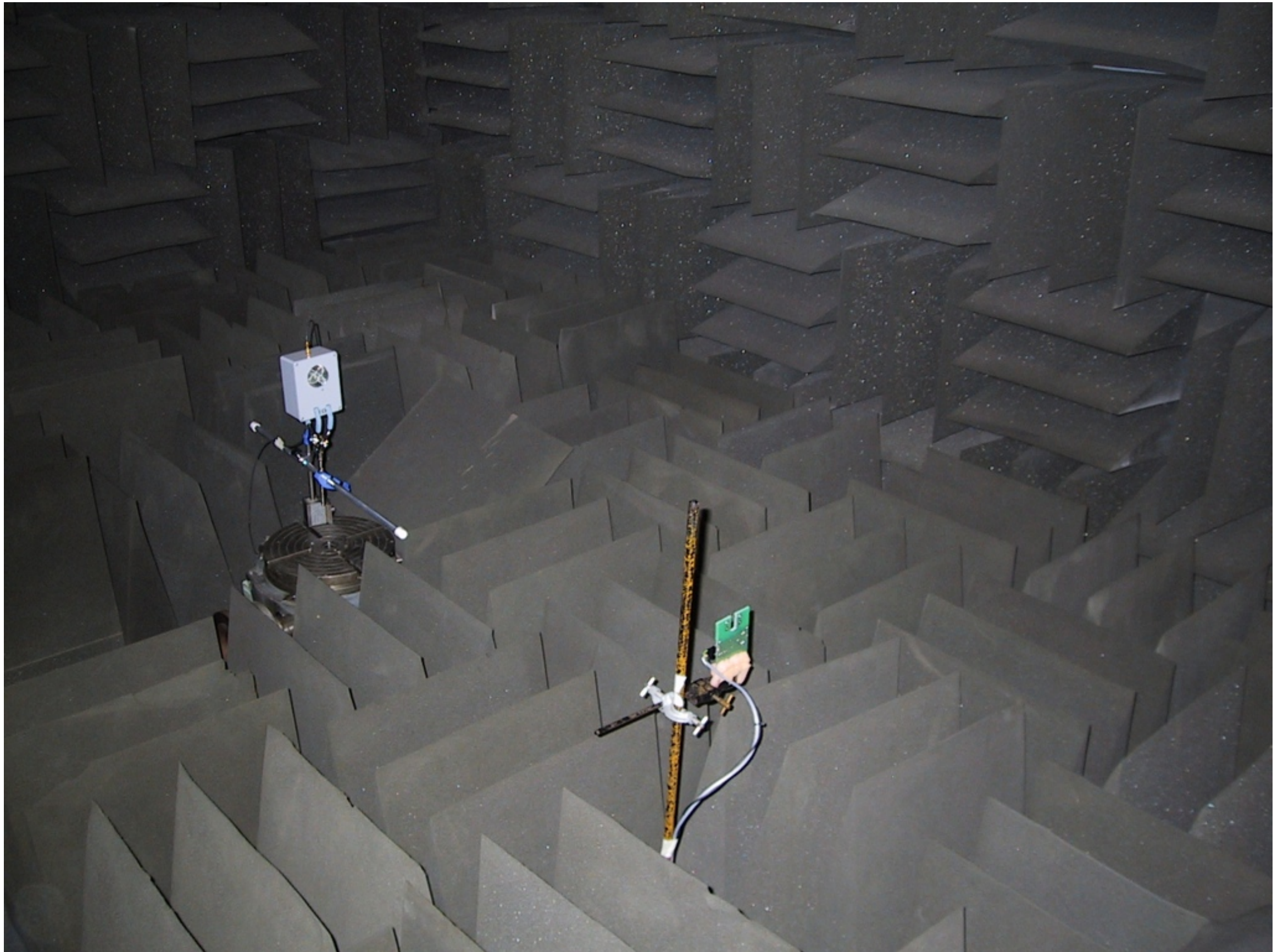
Receiver



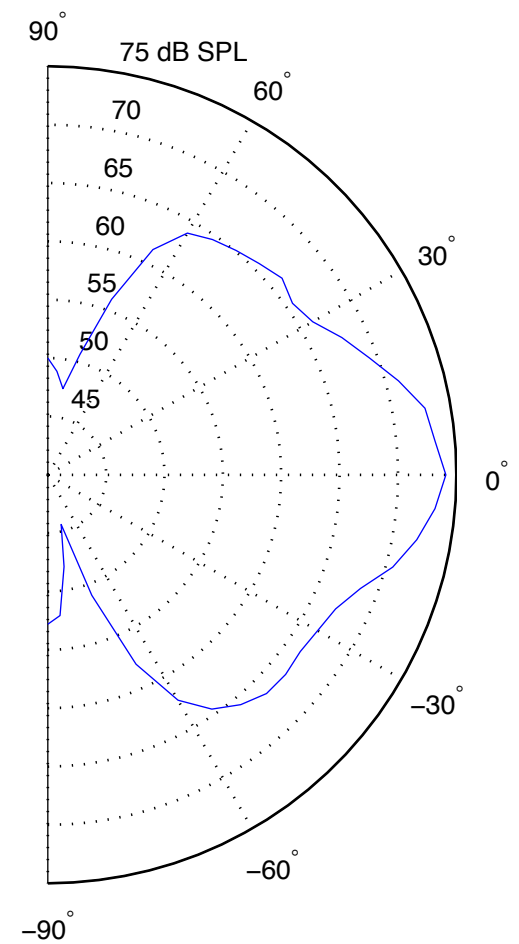
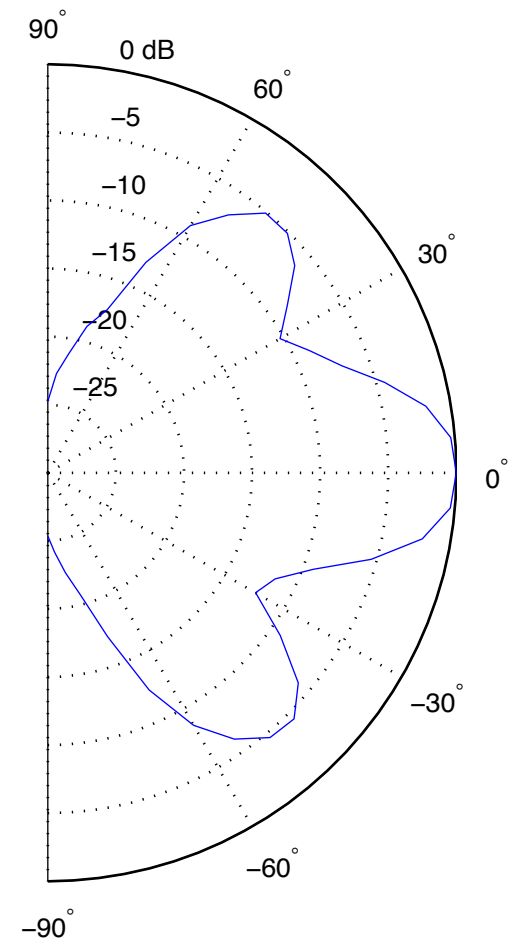
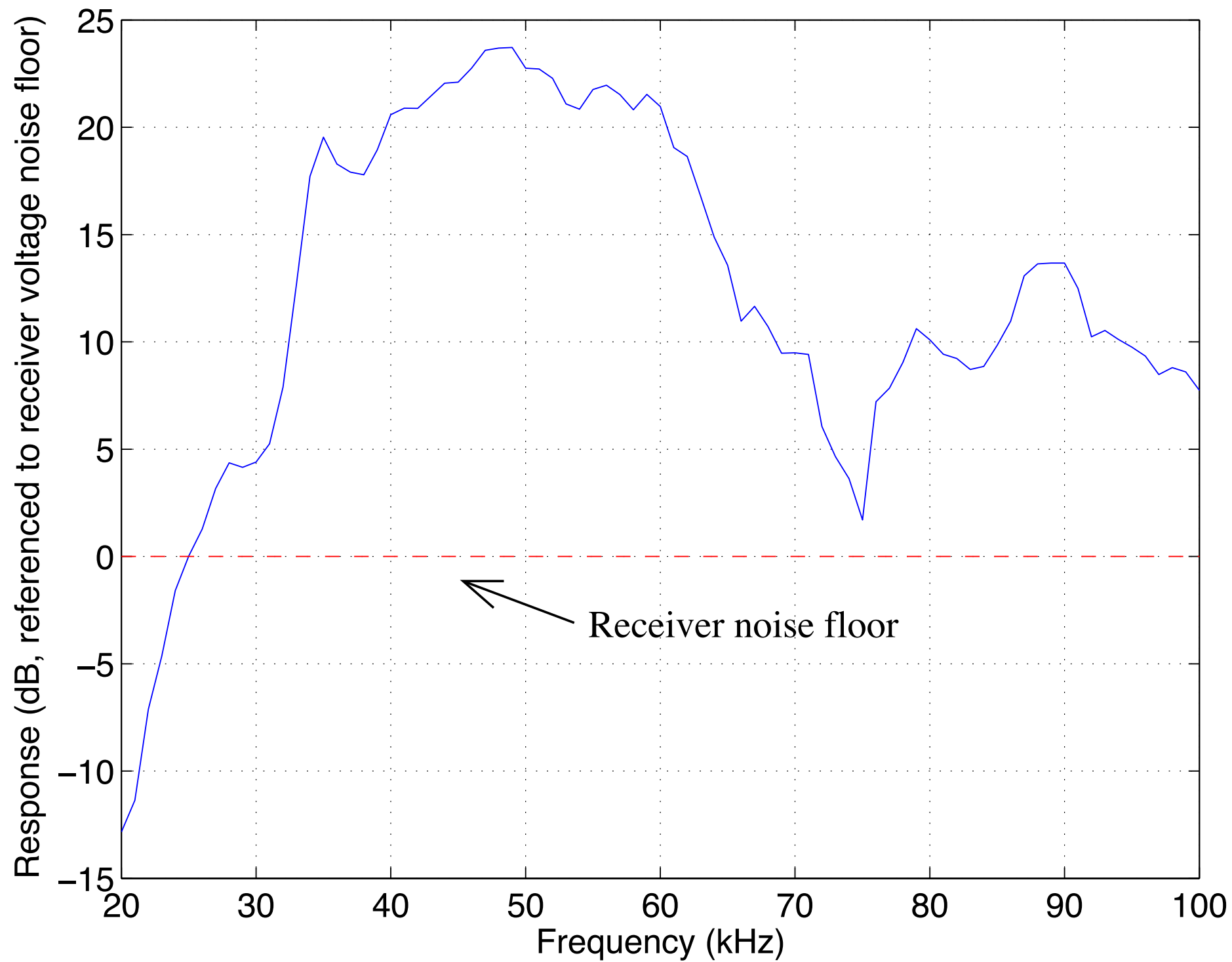
Transmitter

*Hazas and Hopper,  
2006*





# Channel spectrum





# Signaling

- direct sequence CDMA
- 511 length Gold code
- 50 kHz centre frequency
- 20 kHz chip rate



# Location Performance

|                | Success rate | 95% conf. |
|----------------|--------------|-----------|
| 1 transmitter  | 93%          | 2.2 cm    |
| 5 transmitters | 75%          | 2.1 cm    |
| SIC            | 85%          | 2.3 cm    |
| jingling keys  | 68%          | 2.3 cm    |

# Advantages

- noise robustness
- increased update rates
- low-latency positioning
- enhanced identification encoding

# Challenges

1. number crunching required workstation class PC with floating point
2. Doppler distortions due to typical human movement kill the DS/CDMA correlations

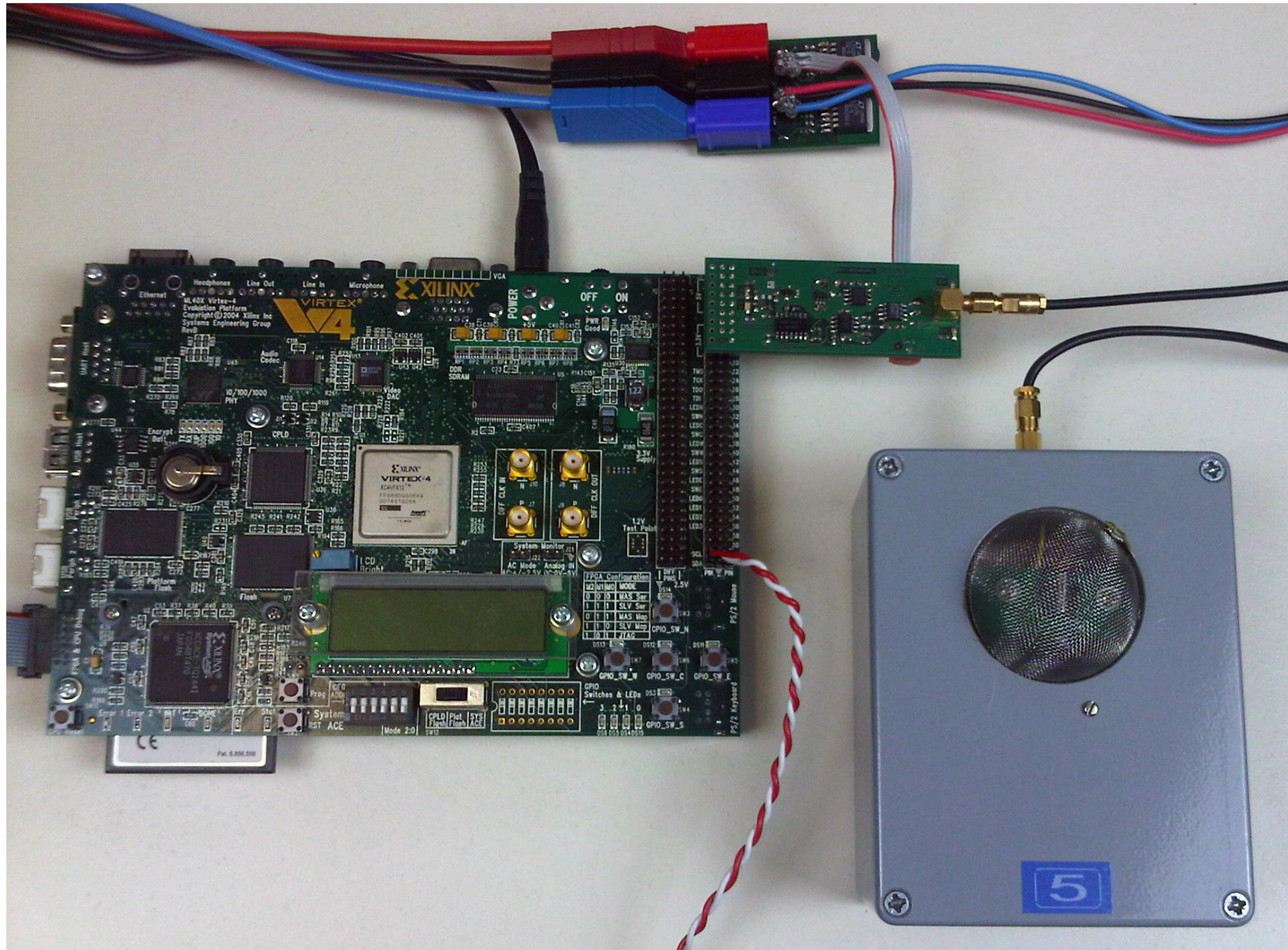
(1) put it on reconfigurable fabric (FPGA)

(2) create an adaptive, Doppler-tolerant tracking algorithm

(1+2) put an adaptive, Doppler-tolerant tracker on reconfigurable fabric

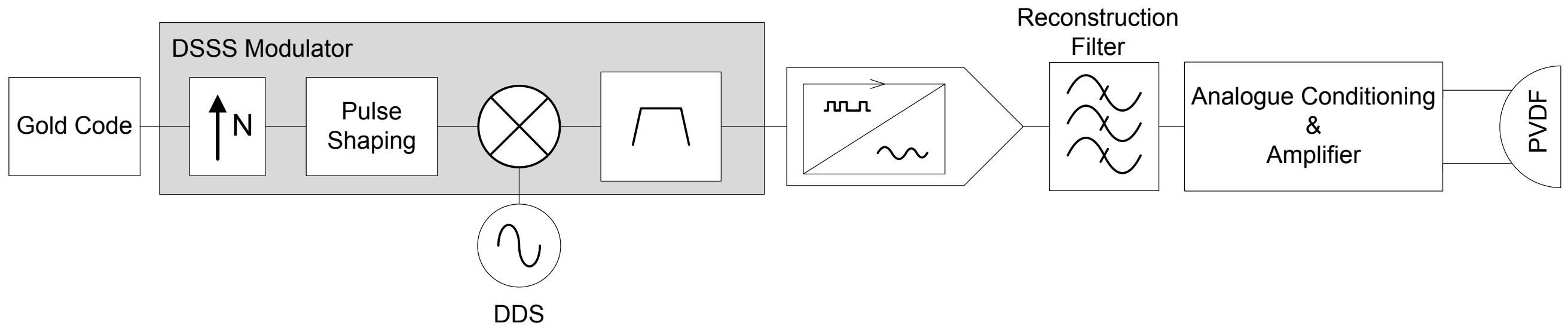


(I) put it on reconfigurable fabric





# FPGA transmitter



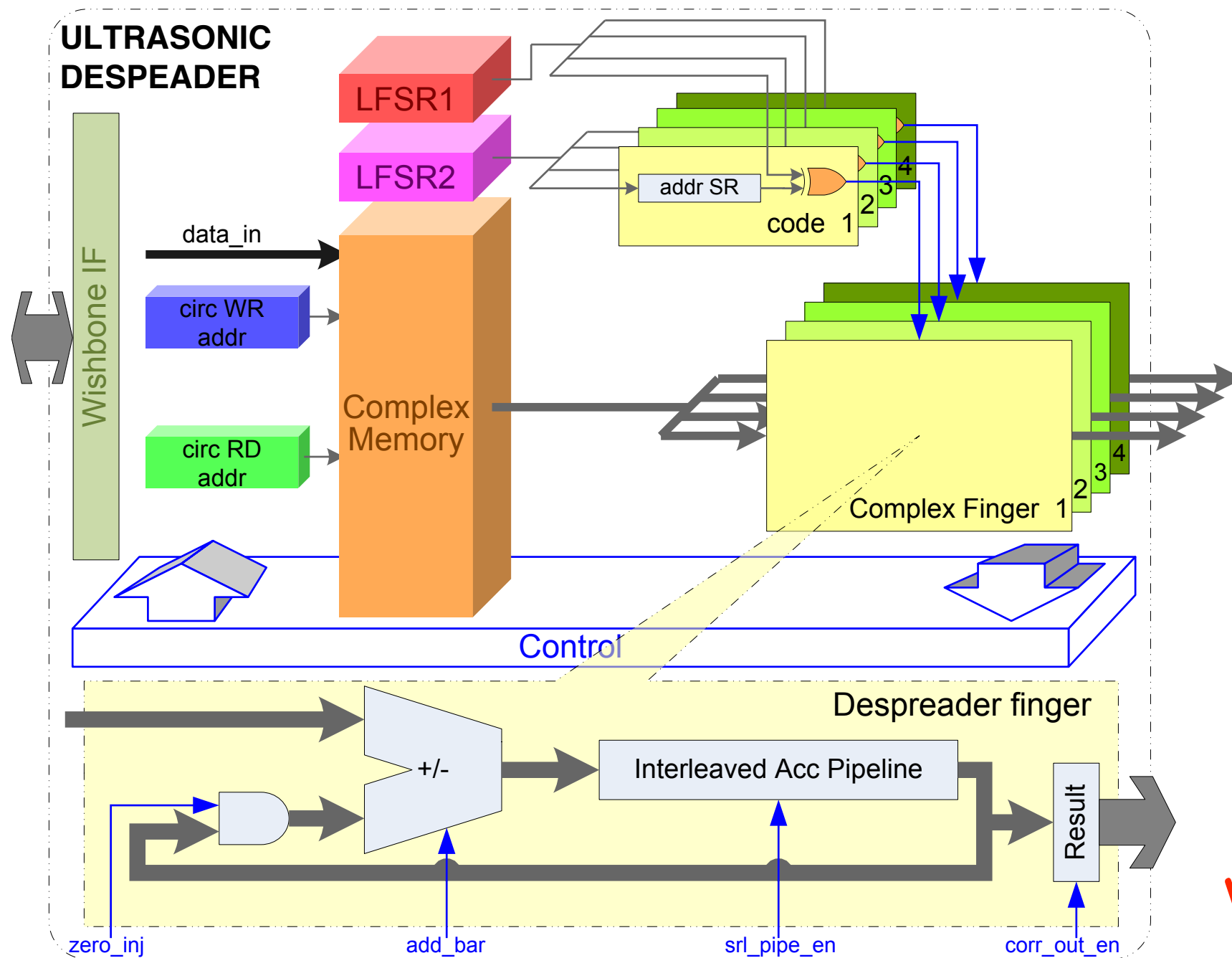
# FPGA transmitter area

| Resource | Units used | Ratio |
|----------|------------|-------|
| DSP48s   | 1          | 3%    |
| BRAMs    | 5          | 13%   |
| Slices   | 308        | 5%    |

FPGA device: Virtex-4 XC4FX12

# FPGA receiver

(or, “multicode embedded despreader core”)



Alloulah and  
Hazas, Proc. of  
IPIN 2010

# FPGA receiver area

| Resource  | Units used | Ratio |
|-----------|------------|-------|
| Slices    | 632        | 11%   |
| LUTs      | 699        | 6%    |
| BRAMs     | 4          | 11%   |
| Max clock | 103.67 MHz |       |

can despread eight users

FPGA device: Virtex-4 XC4FX12

**(2) create an adaptive, Doppler-tolerant tracking algorithm**



# Underwater acoustics

M. Stojanovic and L. Freitag. “Multiuser code acquisition in multipath channels.” In *Proc. of OCEANS 2005 - Europe*, volume 1, pages 74–79.

M. Stojanovic. “Efficient processing of acoustic signals for high-rate information transmission over sparse underwater channels.” *Physical Communication*, 1(2):146 – 161, 2008.

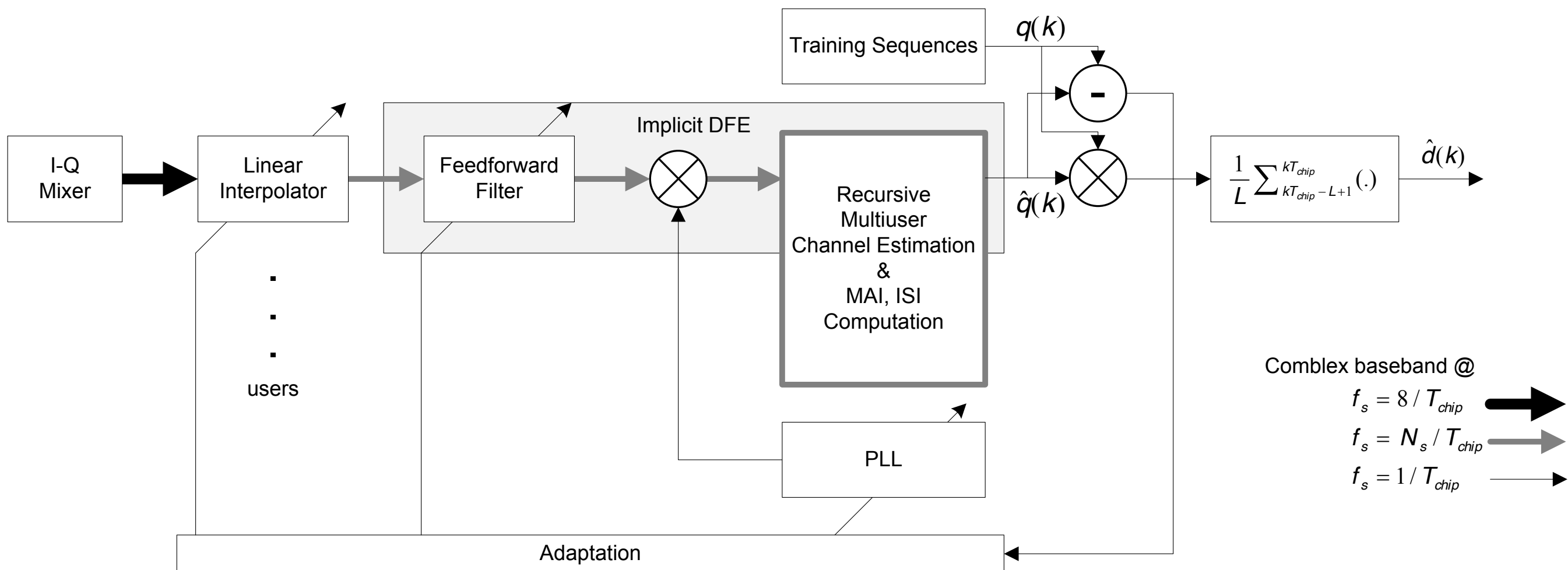
B.S. Sharif, J. Neasham, O.R. Hinton, and A.E. Adams. “A computationally efficient doppler compensation system for underwater acoustic communications.” *IEEE Journal of Oceanic Engineering*, 25(1):52–61, Jan 2000.

# Phase tracking algorithms

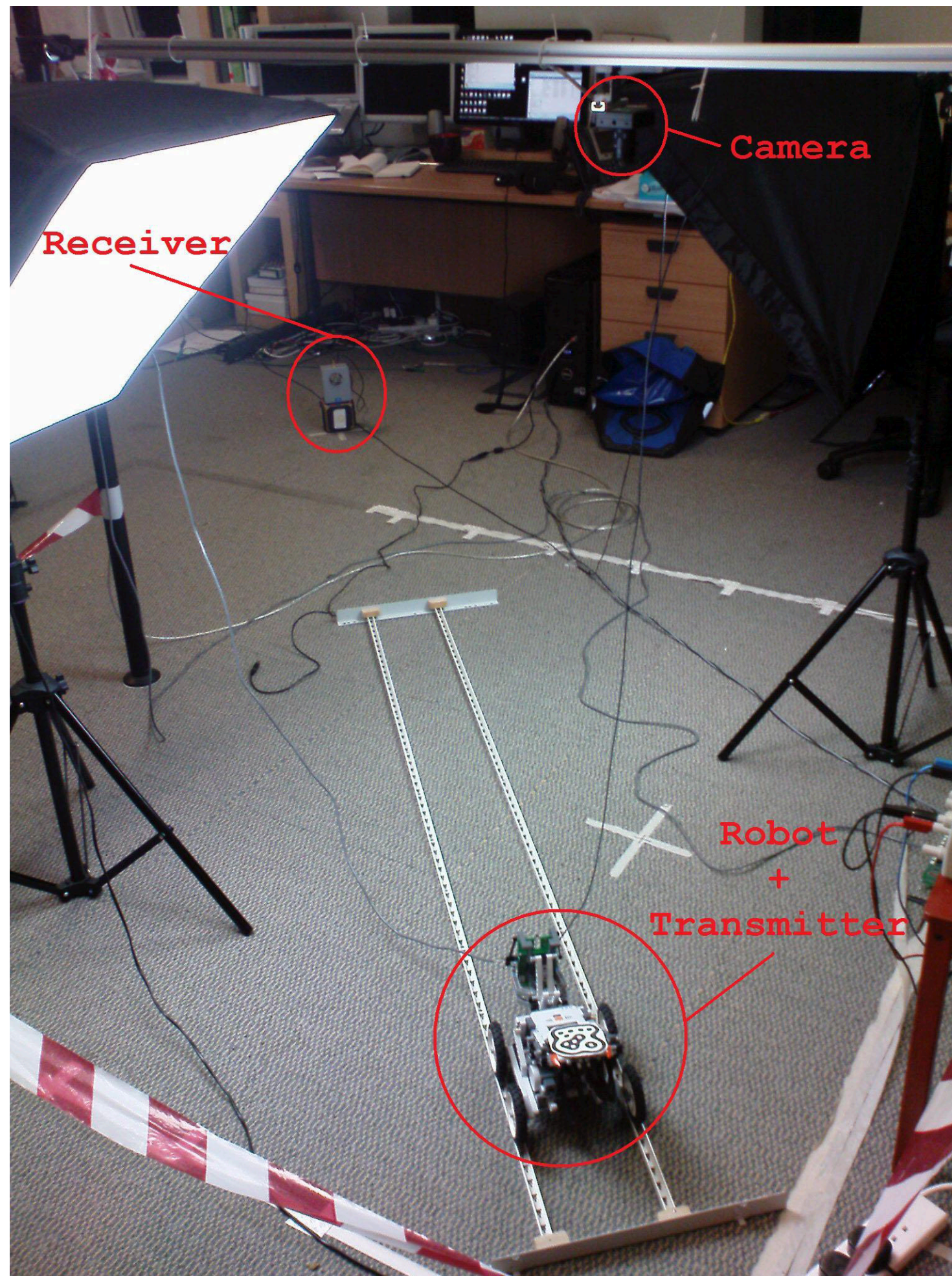
- second-order phase-locked loop (PLL)
  1. sensitive to higher-order motion moments (jerk stress)
  2. unlike RF, both carrier (50 kHz) and code (20 kHz) are affected by Doppler
- implicit decision feedback equaliser (DFE)

interference cancellation, but cannot work alone
- linear interpolation (LI)
  1. tracks total distortion (carrier and code)
  2. agile to the point of being overreactive; can overcompensate or distort

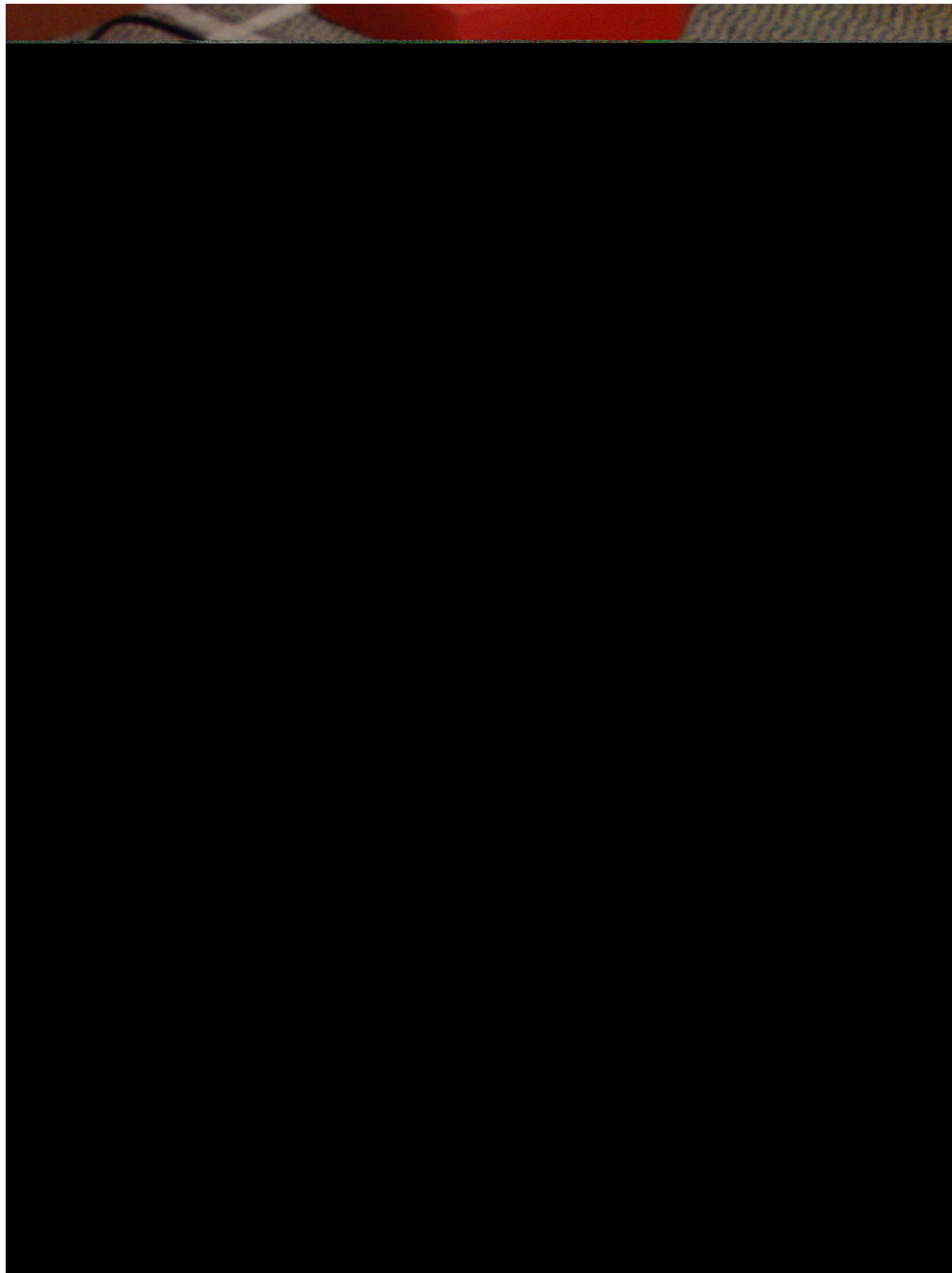
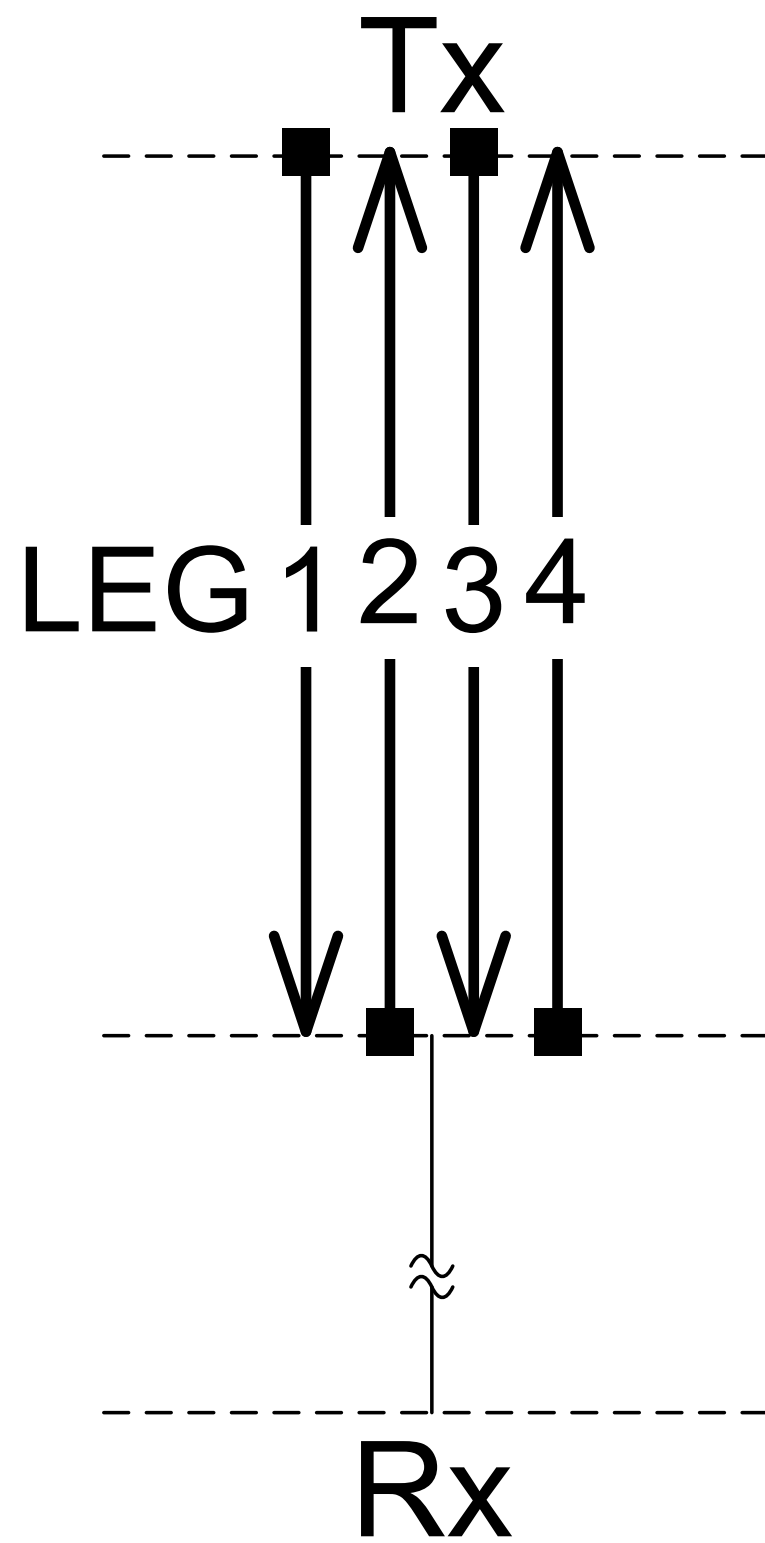
# All together, now.



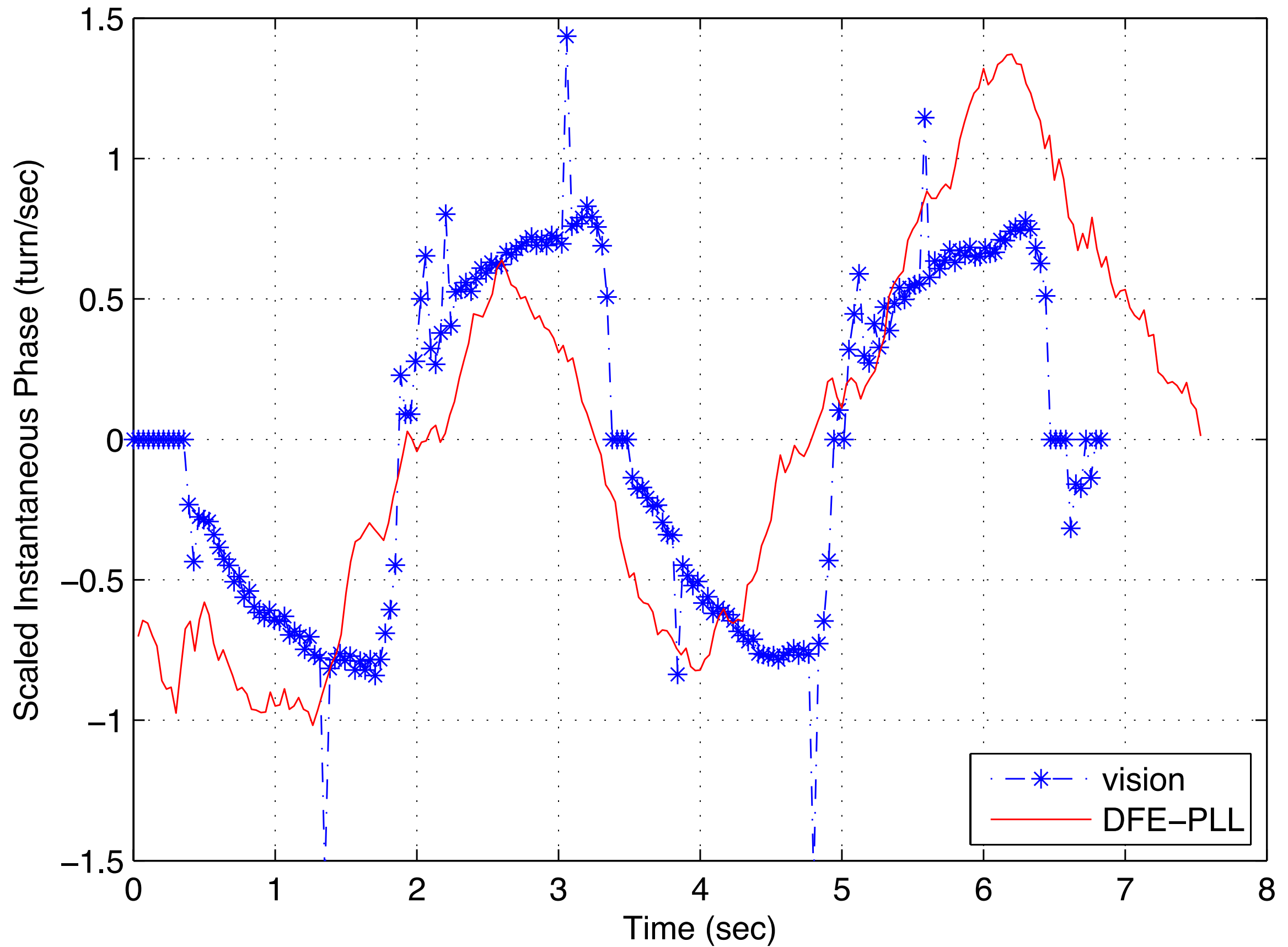






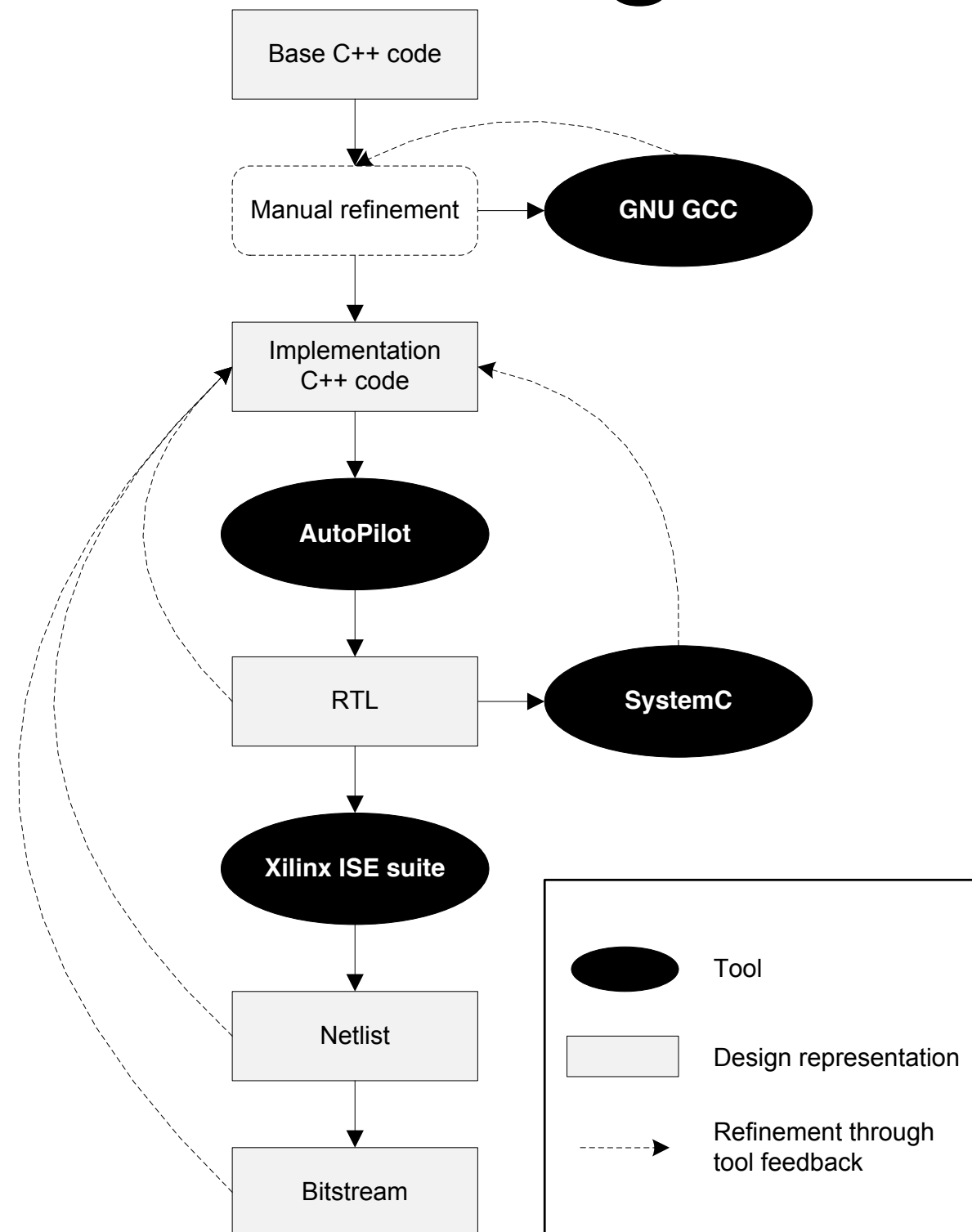


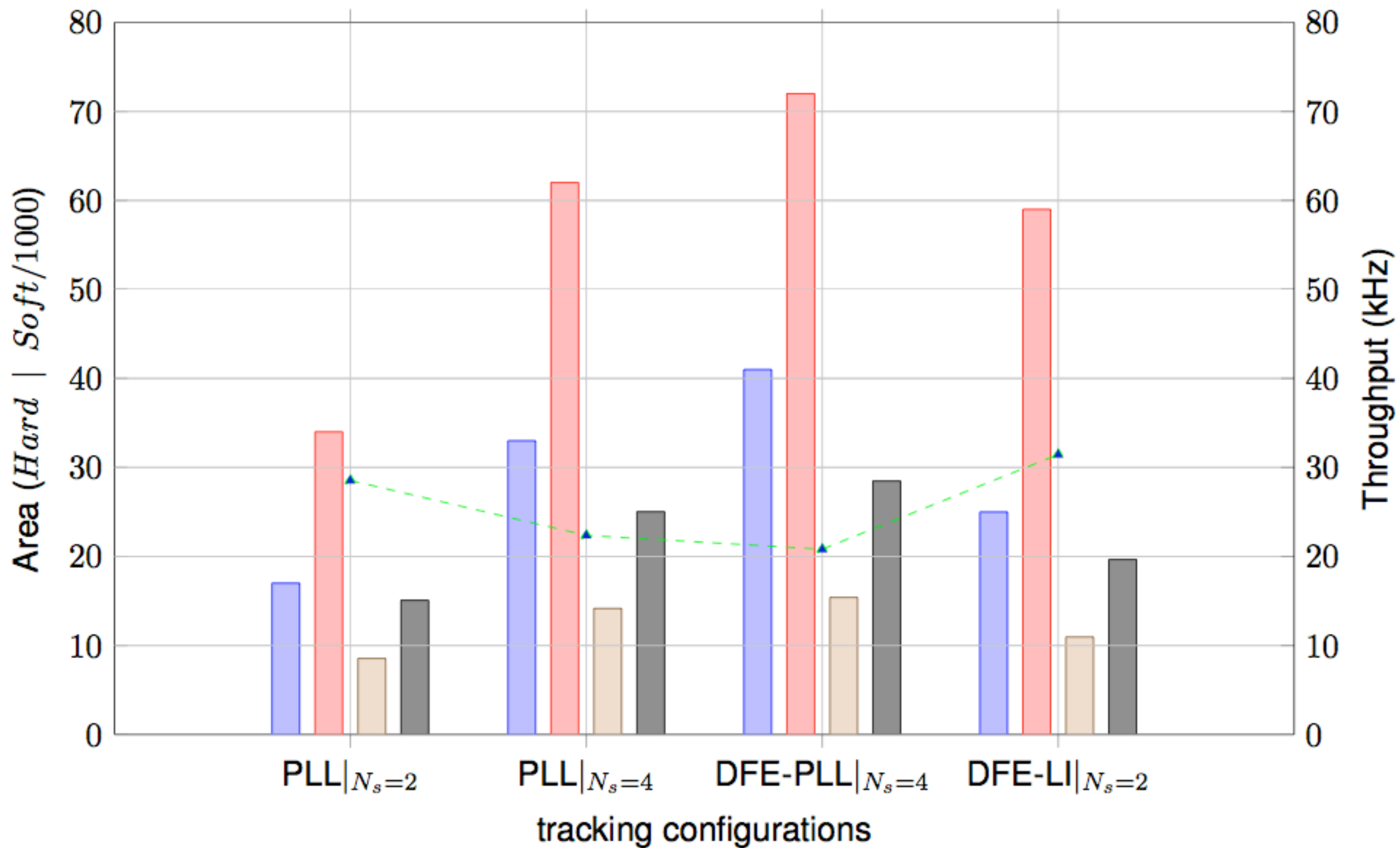
DFE-PLL pattern & Ground Truth





# (1+2) put an adaptive Doppler-tolerant tracker on reconfigurable fabric





BRAM DSP48E FF LUT Throughput

device: Virtex5 XC5VSX94T

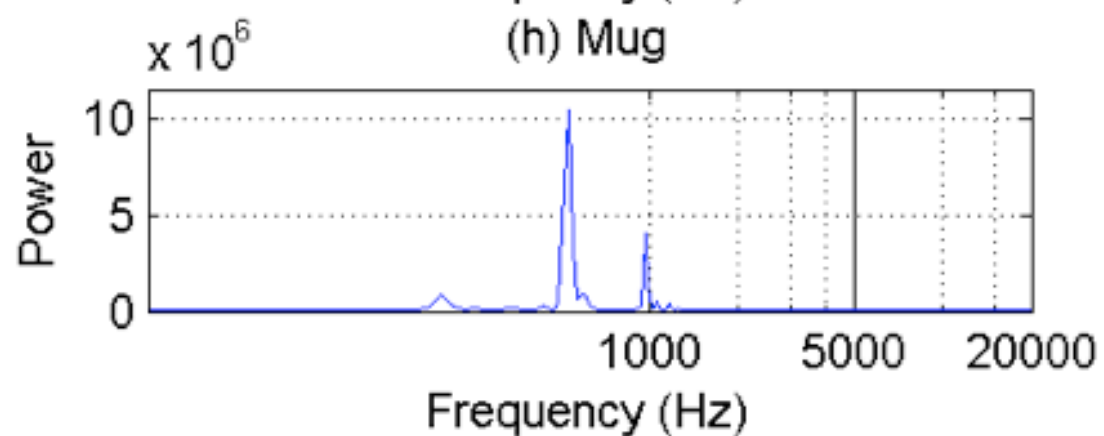
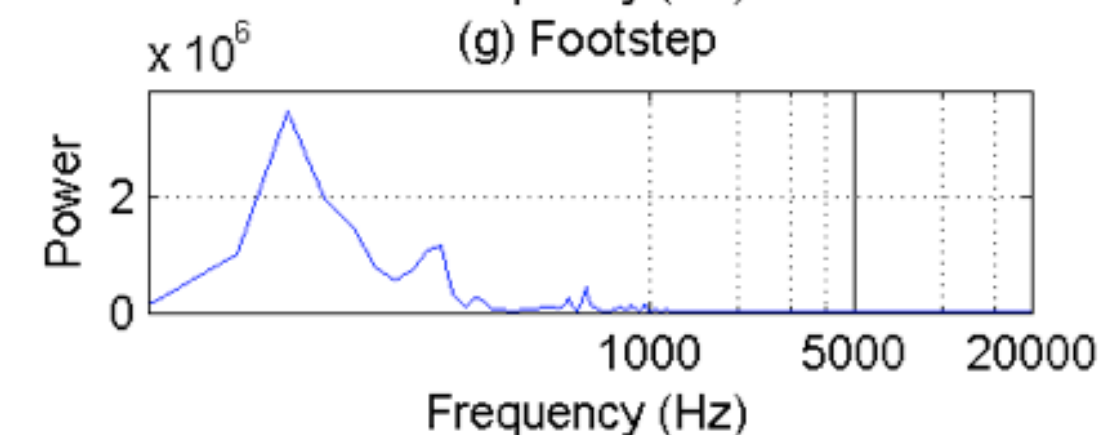
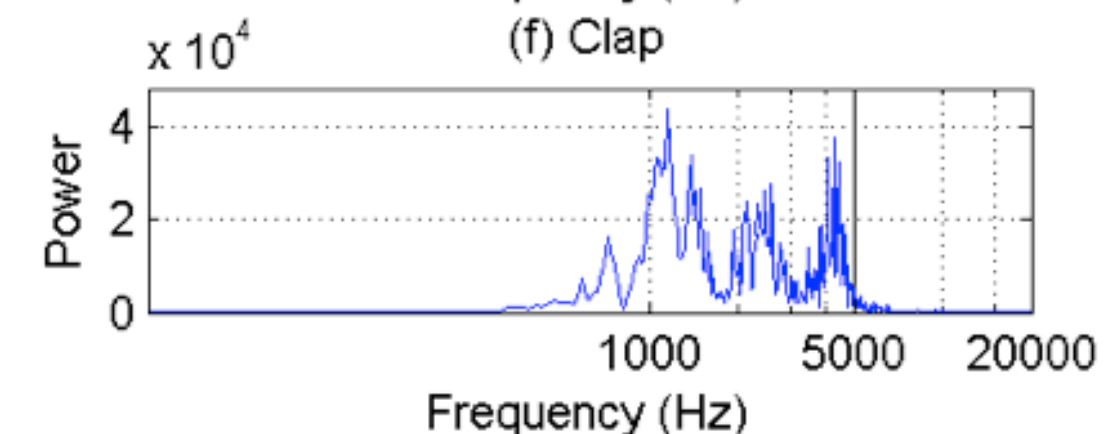
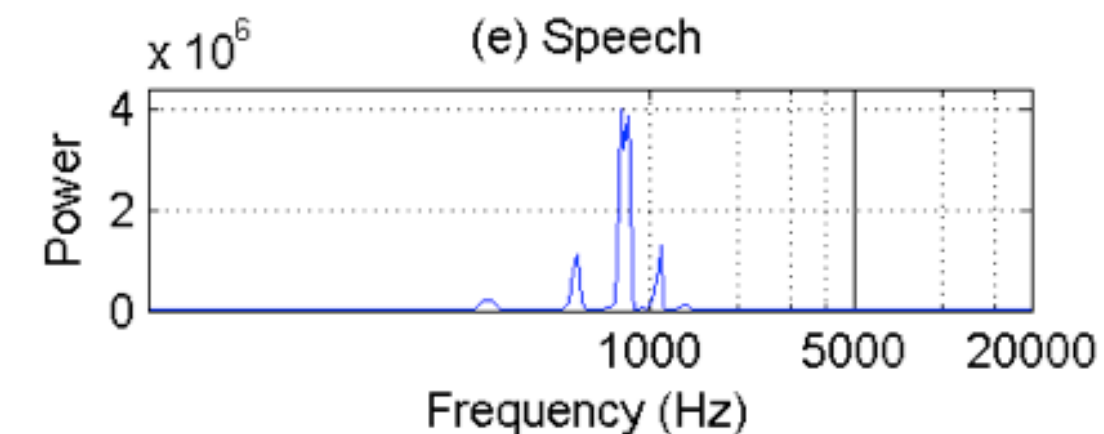
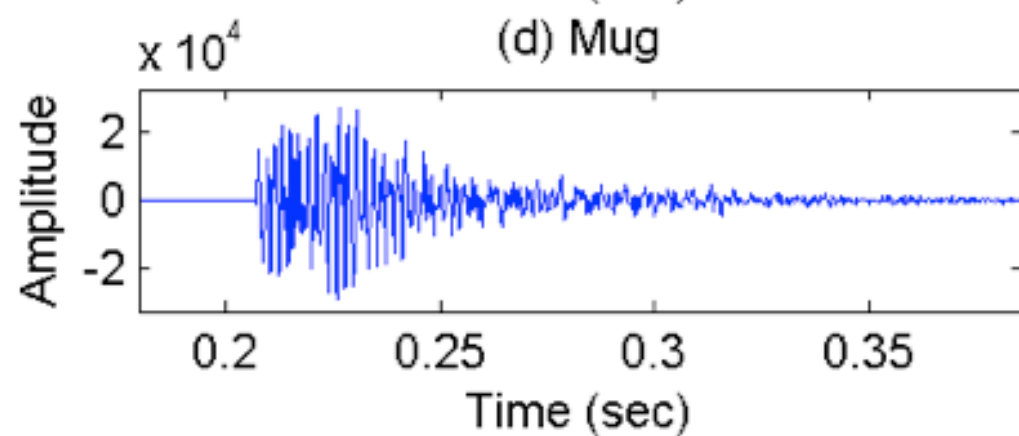
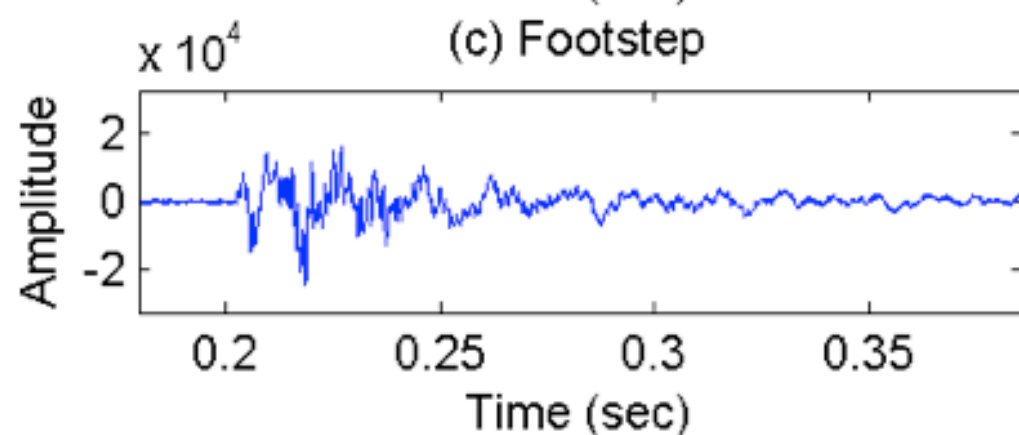
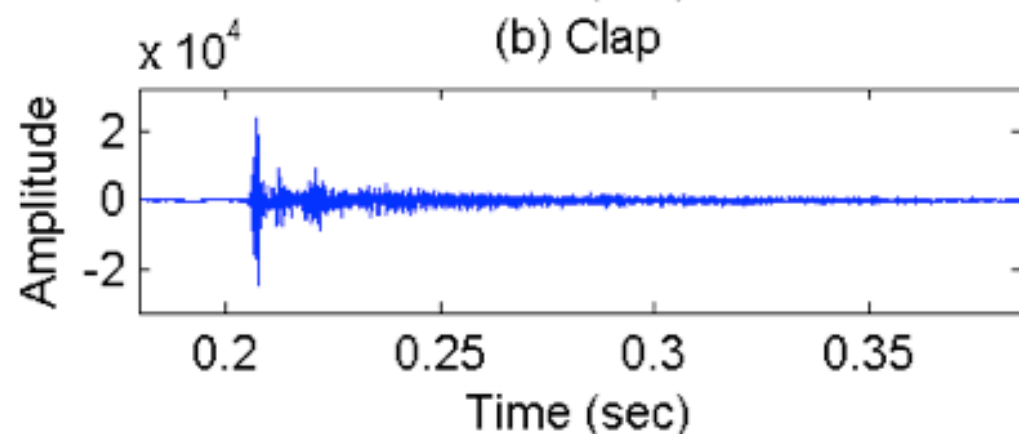
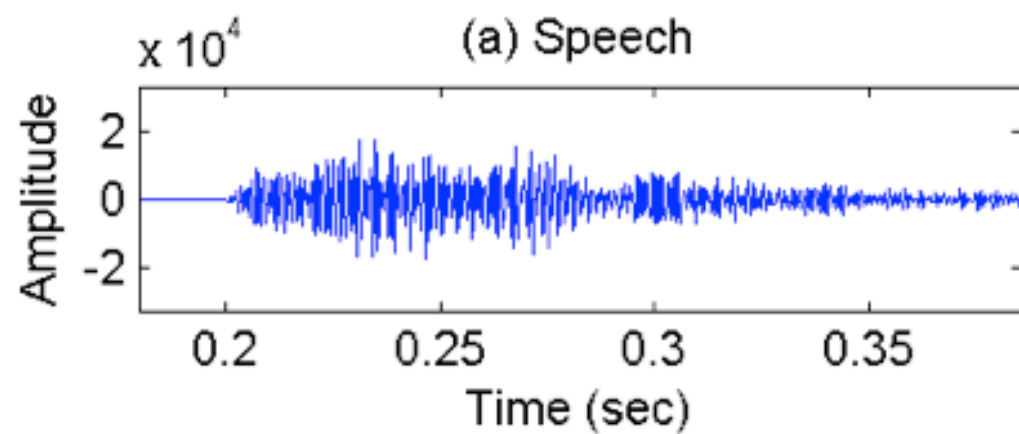
# Audible Sound Localisation

Guo and Hazas,  
proc. of IPSN 2011

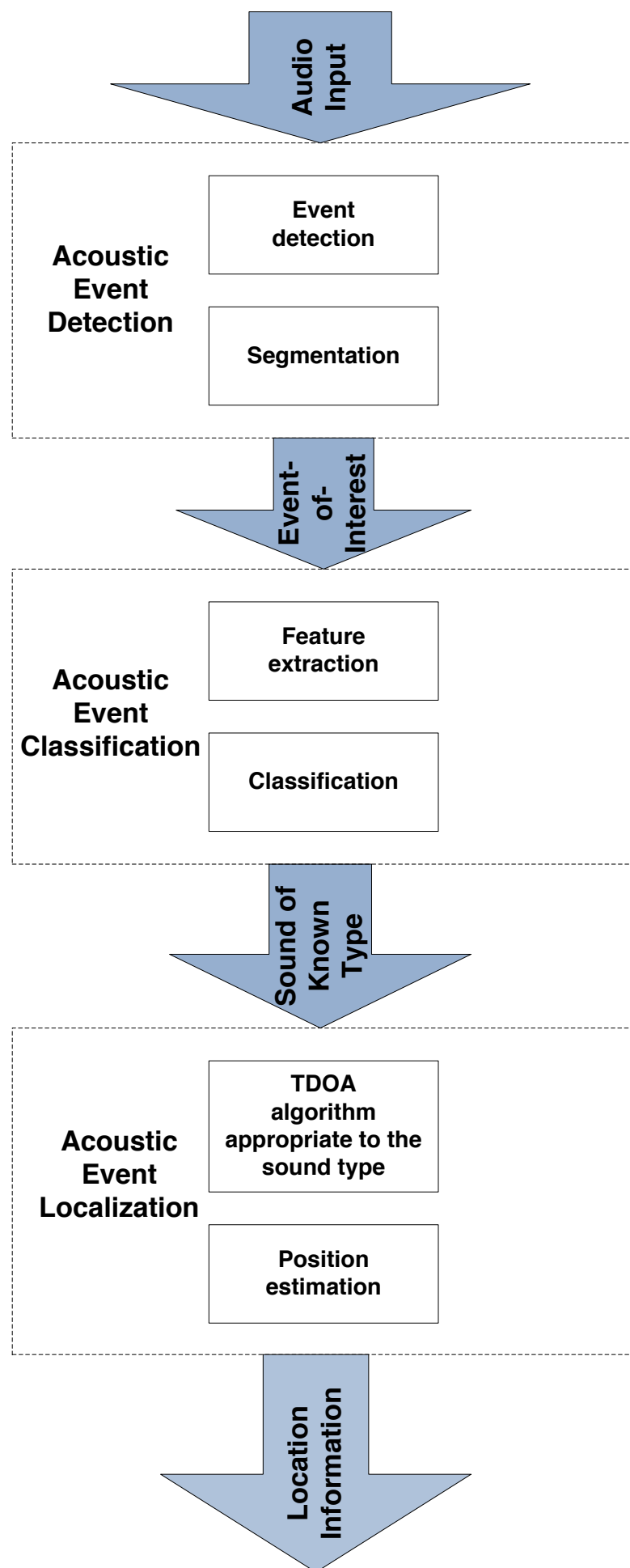
- Two traditional approaches
  1. calibrated microphone arrays
  2. embedded sensor networks

# Goals

- localise everyday indoor sounds generated by people's movement
- with wireless sensor nodes
- talking, walking, placing items on table or shelf surfaces, or handling tools or kitchen utensils

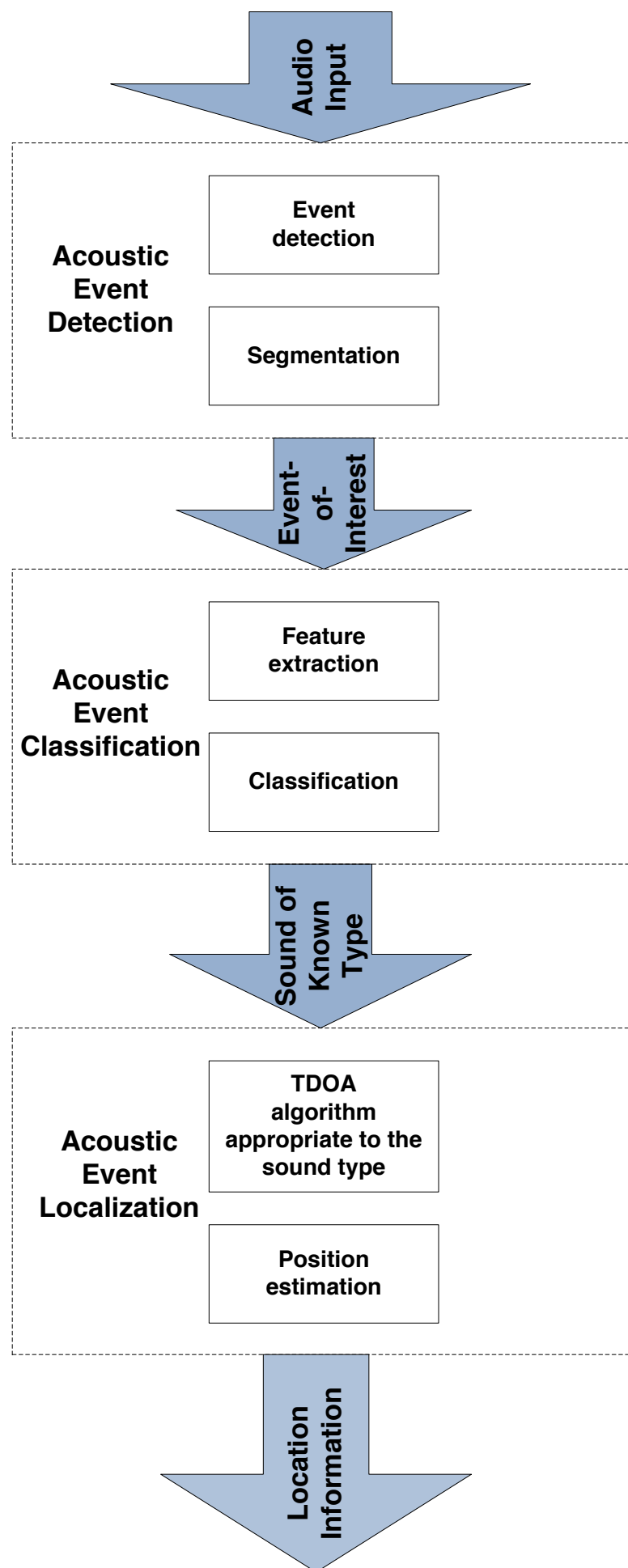






detection by RMS amplitude thresholding

| Subset                          | Features                                   |
|---------------------------------|--------------------------------------------|
| Time domain                     | Root-mean-square (RMS)                     |
|                                 | Low short-time energy ratio (LSTER)        |
|                                 | Zero-crossing rate (ZCR)                   |
| Polynomial                      | Linear predication coefficients (LPC)      |
|                                 | Linear spectral frequencies (LSF)          |
| Human auditory perception model | Mel-frequency cepstral coefficients (MFCC) |
| Spectral                        | Spectral centroid                          |
|                                 | Spectral flux                              |
|                                 | Frequency band energy                      |



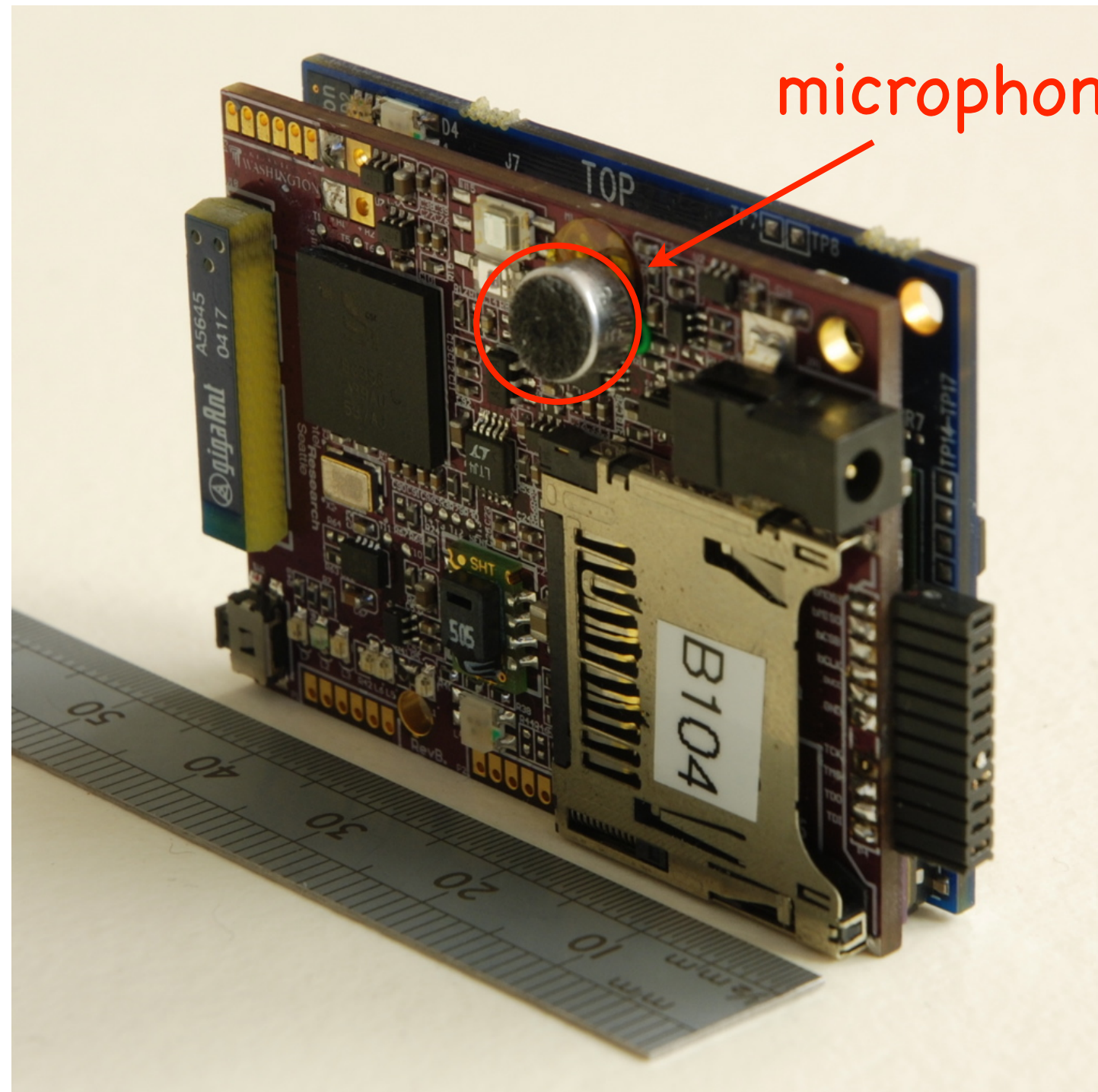
detection by RMS amplitude thresholding

classifier: quadratic discriminant analysis (QDA)

all pairs: dynamic amplitude thresholding (cheap)  
cross-correlation (expensive)

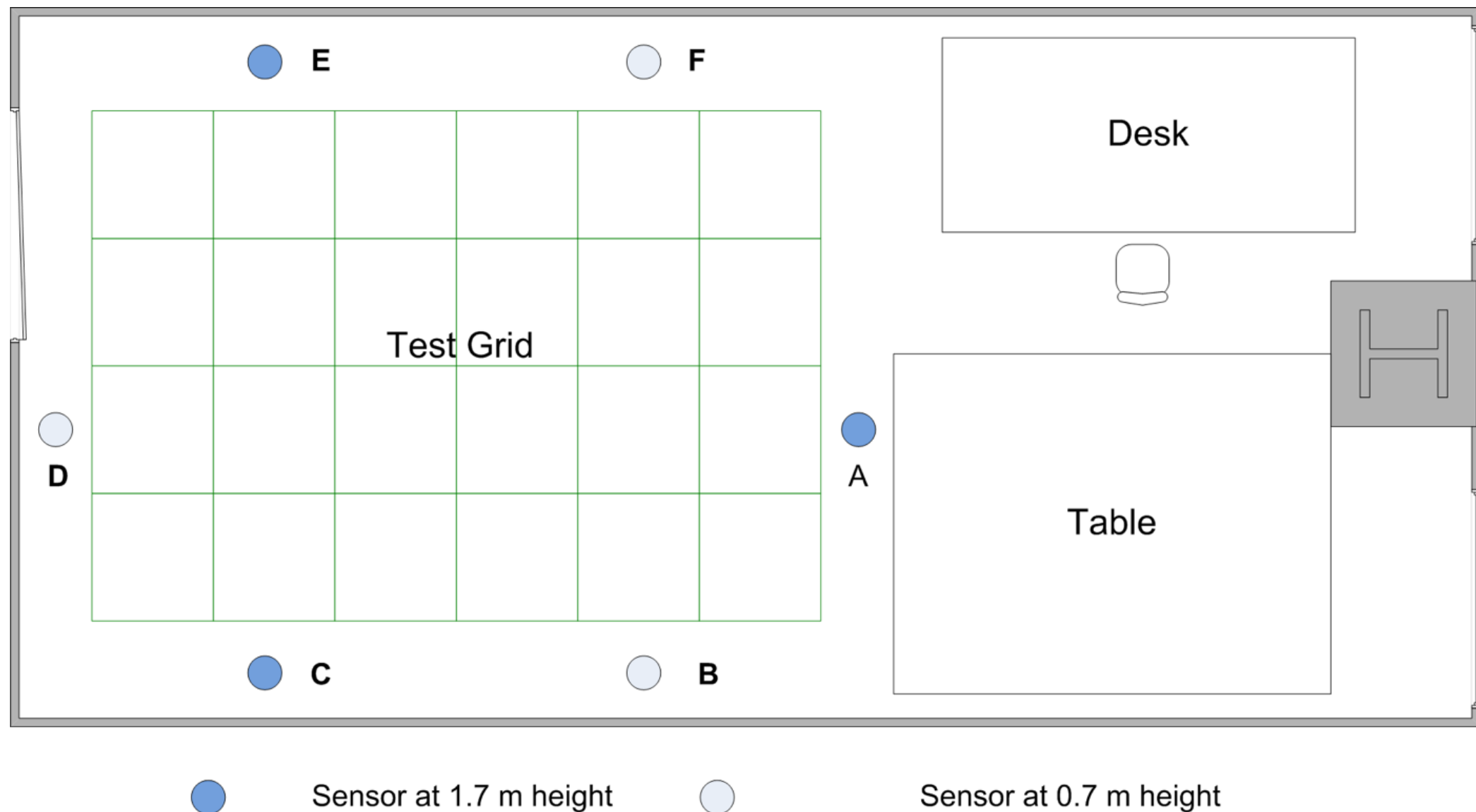
model-based non-linear regression  
iterative residual-based outlier rejection

# iMote2 and sensor board



# Experiment setup

- 3 x 2 x 1.8 m area; six microphones at edges
- 35 sound test points (5x7 grid)
- speech, clap, footstep, mug-on-table (20 times/point)



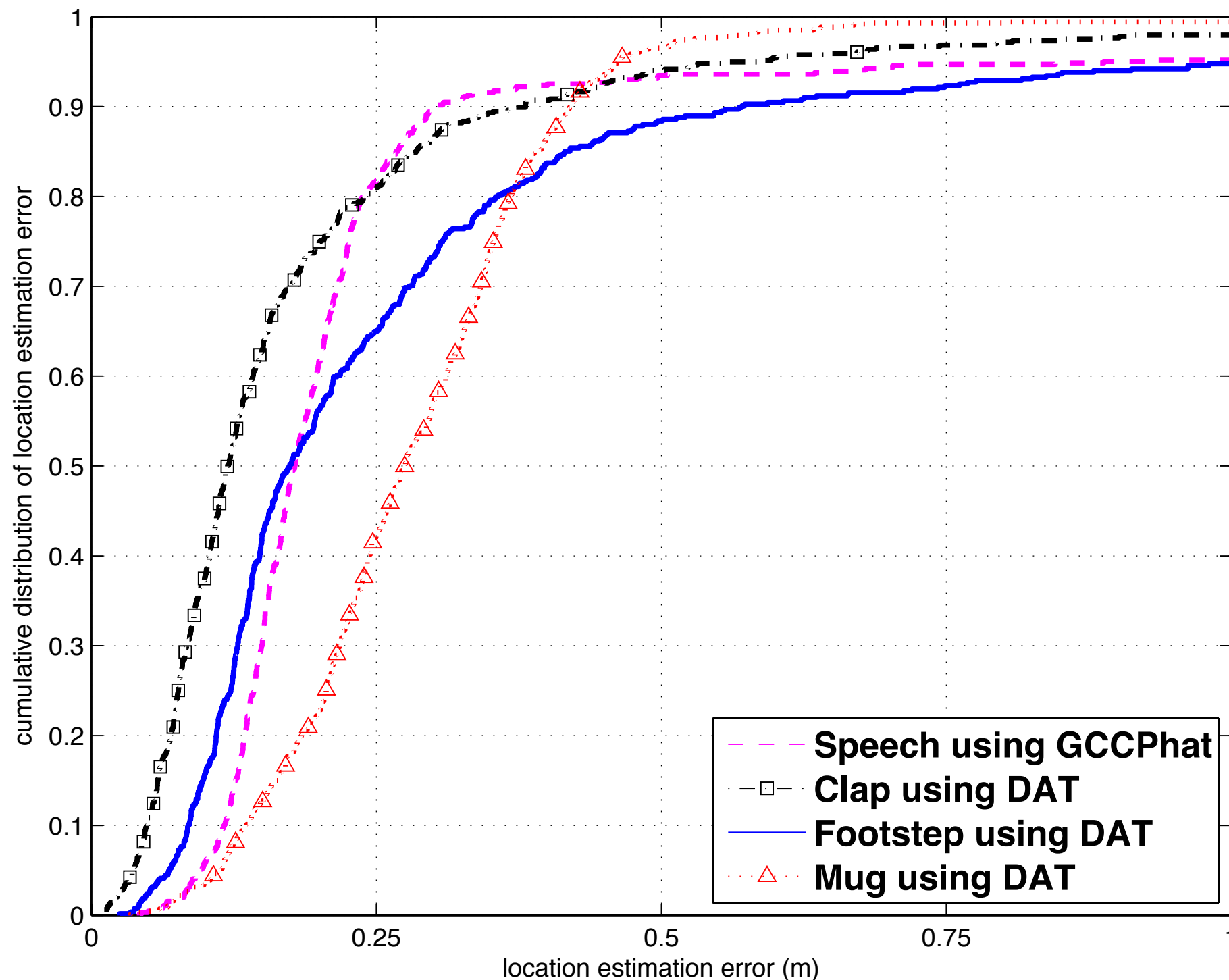
# Classification performance

- 10-fold cross-validation (i.e. 90% for training)
- overall, better than 93%

|          | Speech | Clap   | Footstep | Mug    |
|----------|--------|--------|----------|--------|
| Speech   | 93.46% | 0.02%  | 4.58%    | 1.94%  |
| Clap     | 0.60%  | 99.01% | 0.00%    | 0.40%  |
| Footstep | 1.00%  | 0.27%  | 98.06%   | 0.68%  |
| Mug      | 1.30%  | 0.01%  | 0.26%    | 98.44% |



# Localisation performance



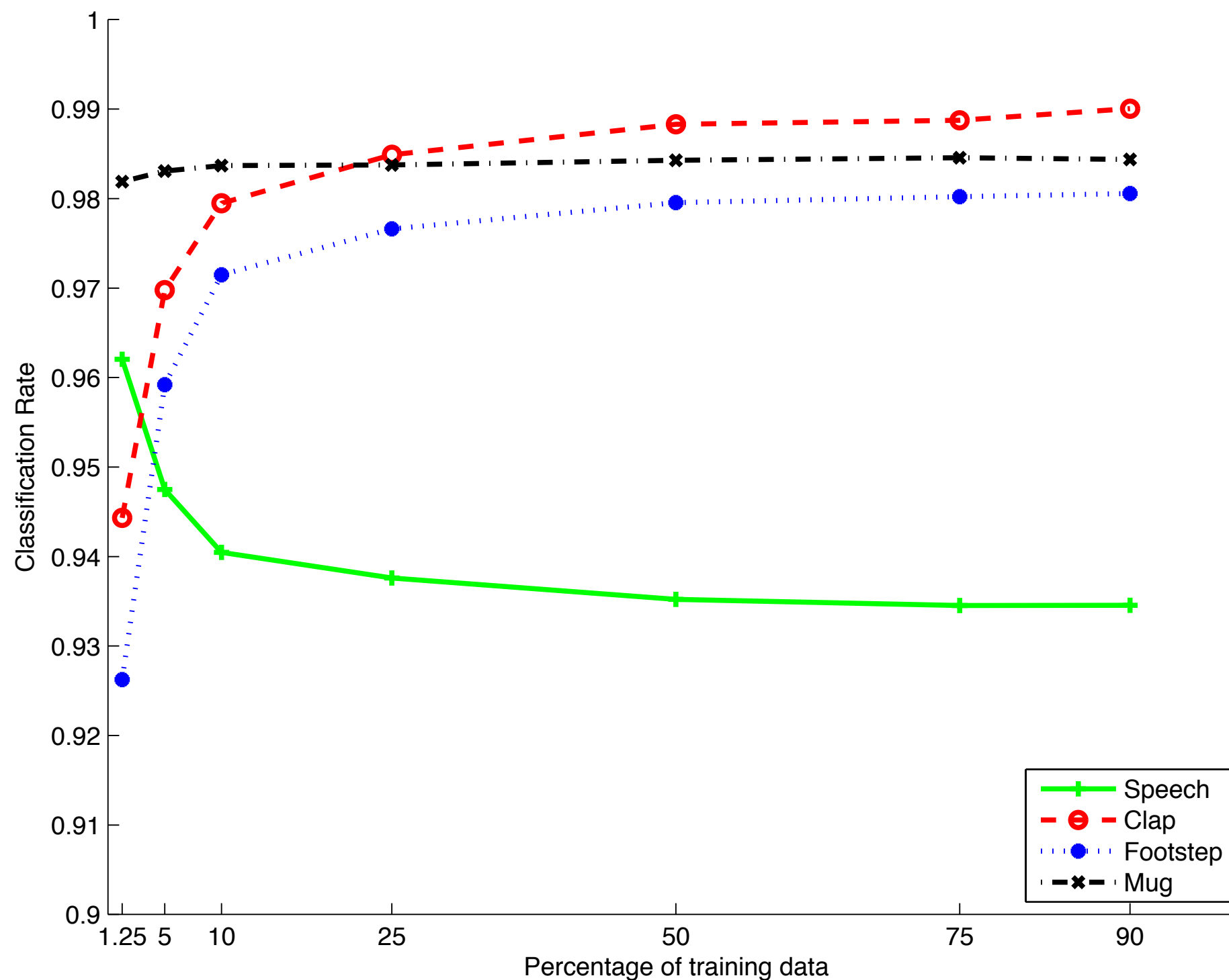
## Rejection rate

|          |      |
|----------|------|
| Speech   | 20%  |
| Clap     | 4.5% |
| Footstep | 40%  |
| Mug      | 2%   |

# Calibration considerations

- What about fewer audio passages from each test point as training?
- What about taking training data from fewer locations?
- What about different speaking subjects?

# Fewer training passages



# Fewer training locations

| Training Points | Speech | Clap   | Footstep | Mug    |
|-----------------|--------|--------|----------|--------|
| 35              | 93.32% | 98.68% | 97.62%   | 98.17% |
| 25              | 94.80% | 98.83% | 97.38%   | 98.52% |
| 15              | 94.03% | 98.60% | 97.83%   | 98.33% |
| 5               | 97.85% | 87.04% | 89.01%   | 96.67% |

# Different speakers

|          | Speech | Clap   | Footstep | Mug    |
|----------|--------|--------|----------|--------|
| Speech   | 91.30% | 0.00%  | 2.42%    | 6.28%  |
| Clap     | 1.00%  | 99.00% | 0.00%    | 0.00%  |
| Footstep | 13.61% | 2.25%  | 83.55%   | 0.59%  |
| Mug      | 2.23%  | 0.00%  | 0.14%    | 97.63% |

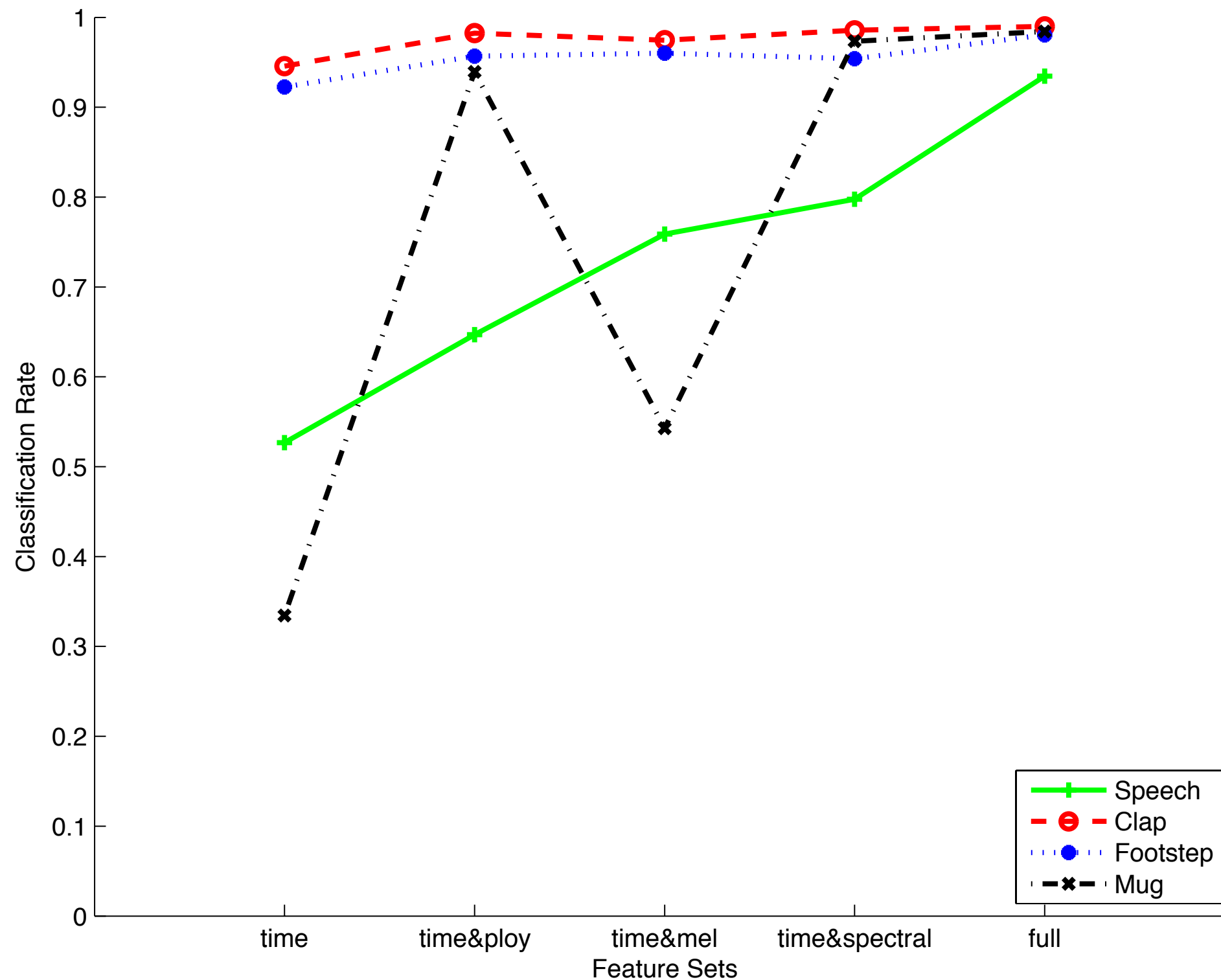
# Performance-resource trade-offs

- constrain the feature set
- lower the sampling rate



| Subset                          | Features                                   |
|---------------------------------|--------------------------------------------|
| Time domain                     | Root-mean-square (RMS)                     |
|                                 | Low short-time energy ratio (LSTER)        |
|                                 | Zero-crossing rate (ZCR)                   |
| Polynomial                      | Linear predication coefficients (LPC)      |
|                                 | Linear spectral frequencies (LSF)          |
| Human auditory perception model | Mel-frequency cepstral coefficients (MFCC) |
| Spectral                        | Spectral centroid                          |
|                                 | Spectral flux                              |
|                                 | Frequency band energy                      |

# Constraining the features



# Lowering the sampling rate (classification)

| Rate  | Speech | Clap   | Footstep | Mug    | N    |
|-------|--------|--------|----------|--------|------|
| 40kHz | 93.46% | 99.06% | 98.04%   | 98.44% | 8192 |
| 20kHz | 92.53% | 99.00% | 97.41%   | 98.45% | 4096 |
| 10kHz | 97.46% | 98.64% | 97.54%   | 97.57% | 2048 |
| 5kHz  | 82.28% | 97.06% | 94.19%   | 97.32% | 1024 |

# Lowering the sampling rate (localisation)

|       |           | Speech | Clap  | Footstep | Mug  |
|-------|-----------|--------|-------|----------|------|
| 40kHz | 75th %ile | 23cm   | 20cm  | 31cm     | 35cm |
|       | 90th %ile | 30cm   | 38cm  | 57cm     | 42cm |
|       | Rejection | 20%    | 4.5%  | 40%      | 2%   |
| 20kHz | 75th %ile | 24cm   | 24cm  | 32cm     | 35cm |
|       | 90th %ile | 54cm   | 47cm  | 66cm     | 41cm |
|       | Rejection | 26%    | 3%    | 25%      | 0.7% |
| 10kHz | 75th %ile | 38cm   | 35cm  | 33cm     | 34cm |
|       | 90th %ile | 148cm  | 66cm  | 74cm     | 40cm |
|       | Rejection | 55%    | 4%    | 20%      | 0.5% |
| 5kHz  | 75th %ile | 139cm  | 53cm  | 36cm     | 35cm |
|       | 90th %ile | 232cm  | 108cm | 94cm     | 42cm |
|       | Rejection | 80%    | 3%    | 10%      | 0.5% |

# Complexity

| Component             | Operations                       | Notes                                                                                                          | Requires |
|-----------------------|----------------------------------|----------------------------------------------------------------------------------------------------------------|----------|
| FFT                   | $N \log N$                       | $N$ : the number of samples per frame, which depends on the frame duration and the sampling rate (see table 9) |          |
| ZCR                   | $3N$                             |                                                                                                                |          |
| LPC [28]              | $3Np - \frac{1}{2}(3D - 1)p^2$   |                                                                                                                |          |
|                       |                                  | autoregressive modelling                                                                                       |          |
|                       |                                  | $p$ : order of autoregressive model                                                                            |          |
|                       |                                  | $D$ : dimension of autoregressive model                                                                        |          |
| MFCC [36]             | $2LM + MN - 1$                   | $M$ : number of mel window filter banks ( $M = 7$ )                                                            | FFT      |
|                       |                                  | $L$ : coefficient order ( $L = 13$ )                                                                           |          |
| RMS                   | $2N$                             |                                                                                                                |          |
| LSTER                 | $2N$                             |                                                                                                                |          |
| Spectral Centroid     | $N$                              |                                                                                                                | FFT      |
| Spectral Flux         | $N$                              |                                                                                                                | FFT      |
| Frequency Band Energy | $N$                              |                                                                                                                | FFT      |
| LDA/QDA               | $CF$                             | $C$ : number of categories ( $C = 4$ )                                                                         |          |
|                       |                                  | $F$ : no. of feature parameters ( $F = 23$ for full feature set; see tab. 1)                                   |          |
| DAT                   | $< 2N$                           |                                                                                                                |          |
| GCC-PHAT [9]          | $3N \log N - \frac{11}{2}N + 20$ |                                                                                                                |          |
| NLR                   | $135I(I - 1) + 1016$             | $I$ : number of microphones which detect the event                                                             |          |

