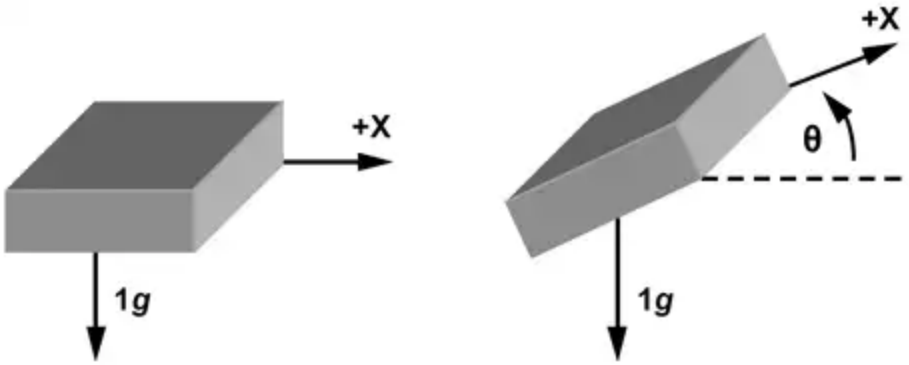
# SmartWheel: Accelerometer Data & Analysis

Yaroslav Burdin

Department of Electrical and Computer Engineering
October 21, 2016

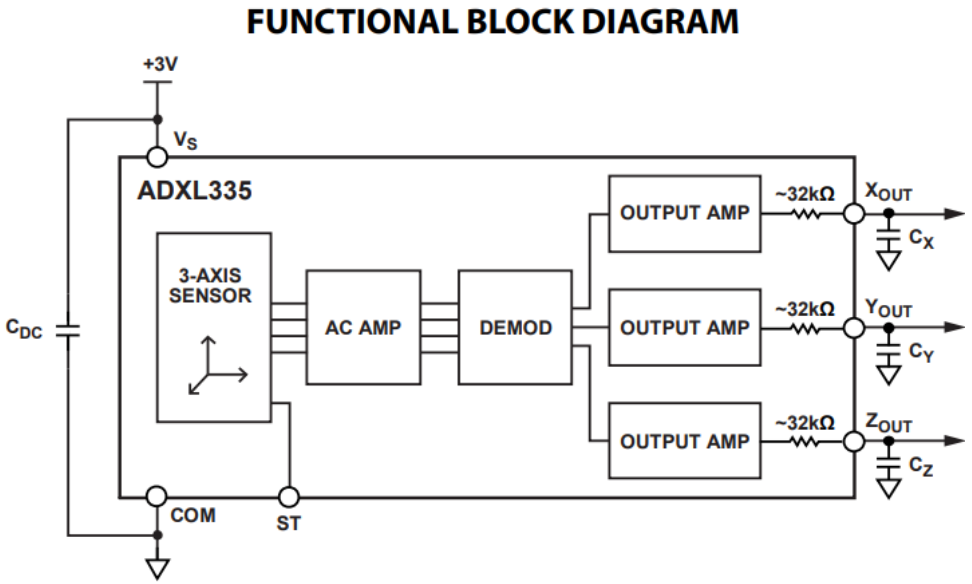
# Change of Acceleration




E.g. acceleration on x-axis:

$$A_{X,OUT}\,[g] = 1\,g \times \sin(\theta)$$

# ADXL335 Accelerometer




Detecting dynamic acceleration with signal conditioned voltage output

# Accelerometer Data

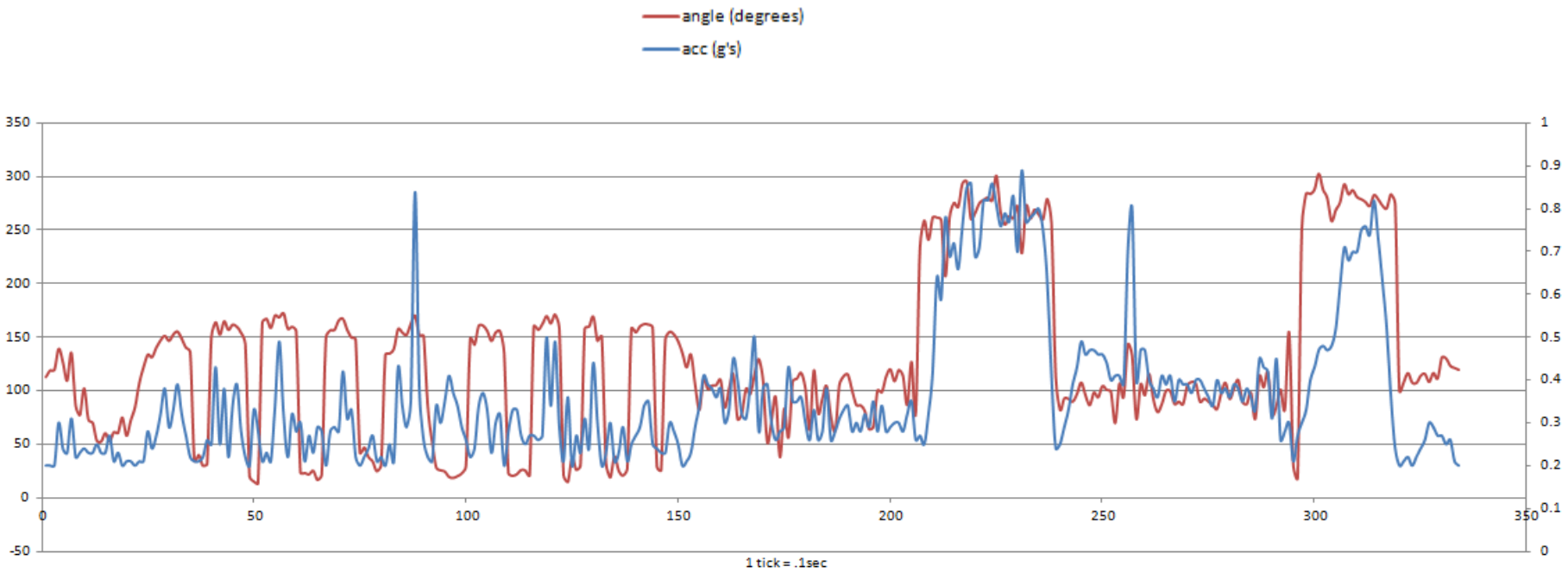
Accuracy vs. Storage

Accelerometer Bandwidth Range: 0.5Hz to 550Hz

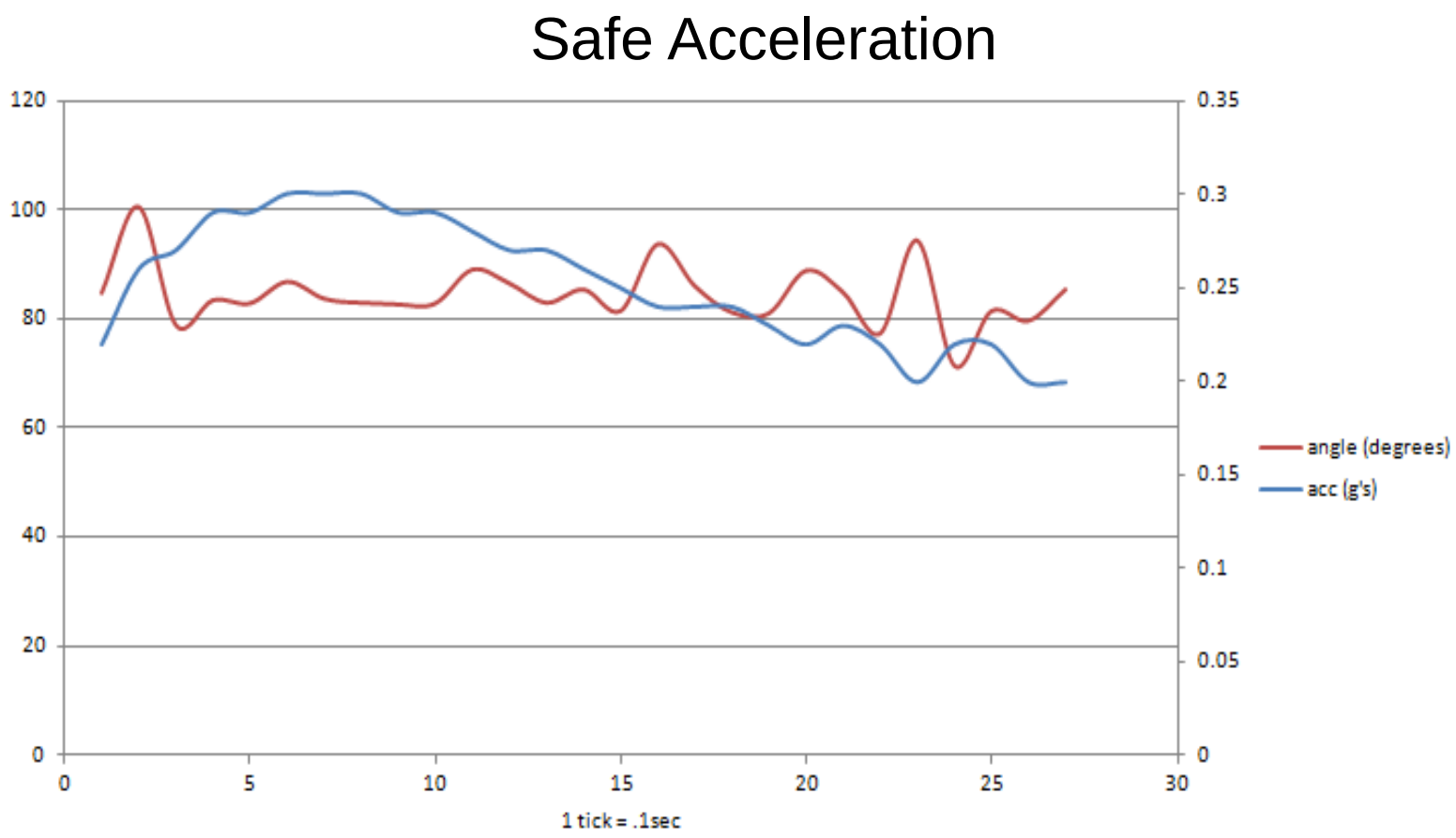
Chosen Bandwidth: 5Hz
1 Hour of driving = 954KB with data compression

Vector Magnitude: $a = \sqrt{x^2 + y^2 + z^2}$

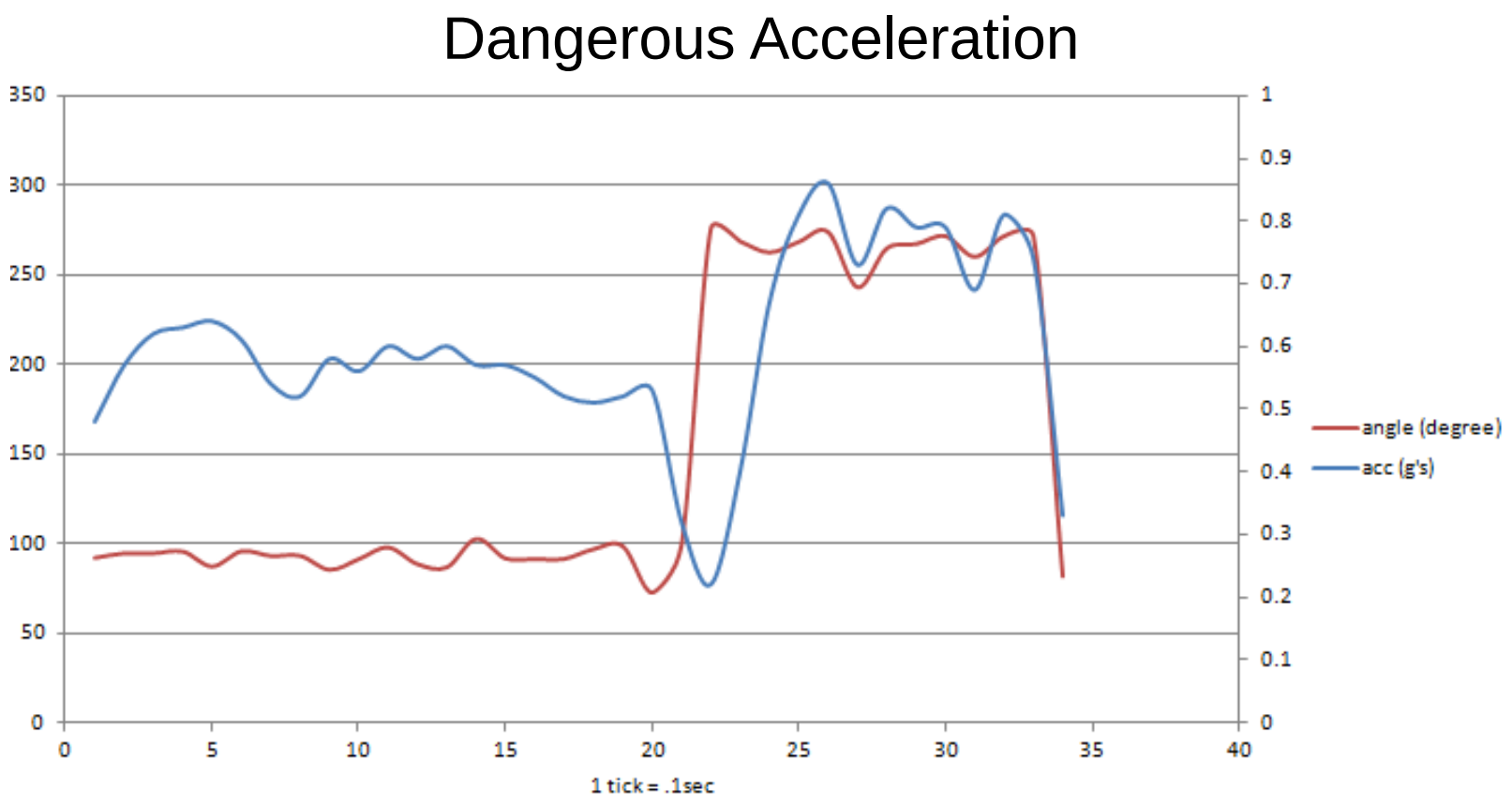
# Acceleration Data Preliminary Visualization



# Acceleration Data Preliminary Visualization



# Acceleration Data Preliminary Visualization



# Data Analysis

Approach: Cluster Analysis

1. Dimensionality Reduction
   - PCA vs t-SNE
2. Clustering Algorithm
   - K-Means vs Hierarchical
3. Performance Metrics
   - Calinski-Harabasz Index
   - Davies-Bouldin Index
   - Silhouette Coefficient

# Available Features

Time Domain Features:

- Mean – average acceleration
- Median – intermediate acceleration
- Standard deviation – dispersion of acceleration from its mean
- Root mean square – square root of the mean square of acceleration signal
- Variance – spread in the acceleration signal

# Available Features

Frequency Domain:

- Repetitive nature of the signal can be extracted from data
- Less susceptible to signal quality variations

Frequency Domain Features:

- Energy – characterizes the frequency components of each activity
- Entropy – measure of consistency
- Mean Frequency – average frequency

# Feature Selection

Principal Component Analysis is used for dimensionality reduction

Three features selected:

1. Mean
2. Median
3. Root Mean Square (RMS)

# Clustering Algorithm

K-Means Algorithm

Pros: simple iterative method to partition n observations in k clusters

Cons: have to specify number of clusters
Approach: the silhouette value

Technology used: Apache Spark

# Evaluation and Results

Linking clusters to true labels

Clustering Accuracy: 78%

Why?

- Data Amount (i.e. difficult to collect prolonged dangerous stopping)
- Window Size

# Thank you!