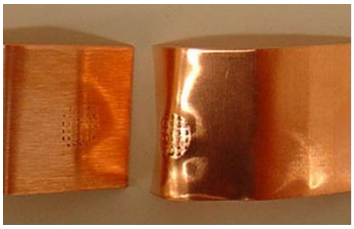
# Lab 3: Feature Generation and Selection

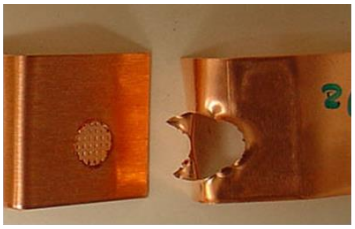
Assigned: October 23, 2019 Due: November 8, 2019

## Description of the Data Set

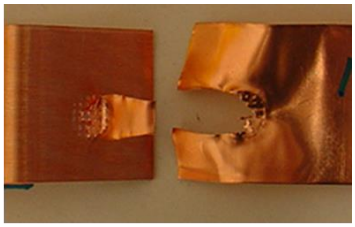
In this lab, 69 welding experiments were conducted under seven pressure levels (psi): 10, 20, 30, 40, 50, 60, and 75. Each condition was repeated by ten times. Two types of signals, power and force, were collected for each sample during the welding experiments. Peel test was performed subsequently to evaluate the welding quality. The samples are categorized into three classes: cold, excessive, and good.



**Class 1: Cold**



**Class 2: Excessive**

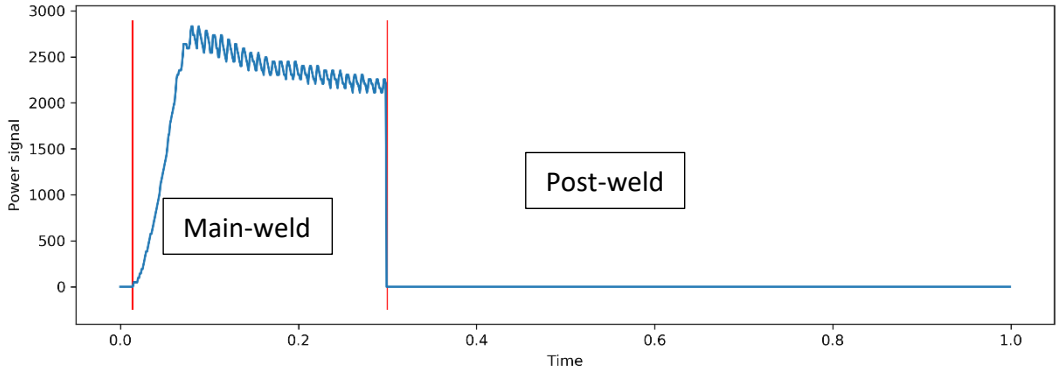


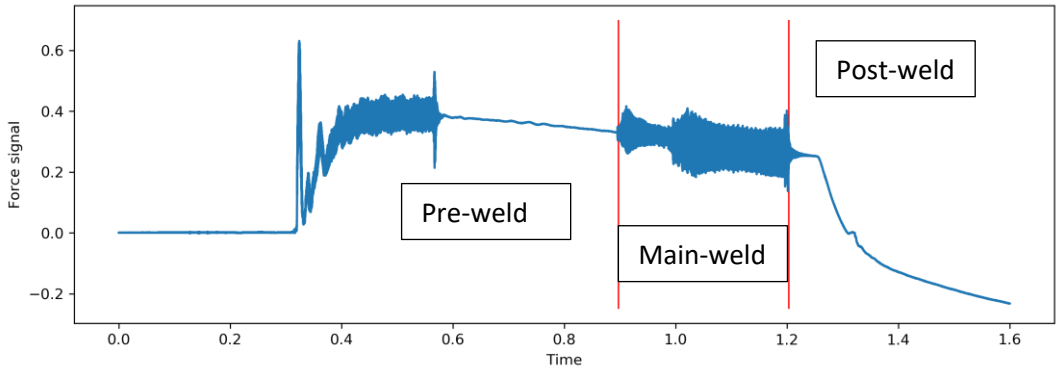
**Class 3: Good**

The experimental table and the data set are available on Box: https://uofi.box.com/s/zuj5nzm49vcwxuinhgtke11835dcts6k.

## Analysis Procedure and Lab Report Requirement

1. Data preprocessing. As shown by the figure below, the power signals include a main-weld segment and a post-weld segment, and the power signals include a pre-weld segment, a main-weld segment, and a post-weld segment. Based on a physical understanding of the welding process, we know that the main-weld segment is the most informative for quality monitoring and our following analysis will focus on this segment only. Develop an algorithm to automatically extract the main weld segment for all welds. In the lab report, describe your algorithm briefly and demonstrate its effectiveness with one example for each type of signals, i.e., plot the raw signal and show the boundaries of the main-weld segment.






2. Feature generation. Create features following the steps below and record all features in a csv file. Each row corresponds to one experiment and each column corresponds to one parameter/feature. The columns should follow this order: experiment number, quality label, four features from group a, four features from group b, and four features from group c. This table should be included in your submission.
   a. Process parameters. Welding pressure, pre-height, post-height, and height change. First three were recorded by the welder. Height change is the difference between pre-height and post-height.
   b. Time-domain features from power signals. Brainstorm and generate four features. In the lab report, provide the definitions of all features and a justification why these features may be helpful.
   c. Frequency-domain features from force signals. Generate the following features: 1$^{st}$ peak frequency and magnitude, 2$^{nd}$ peak frequency and magnitude.
3. Fisher's ratio. Calculate Fisher's ratio for Good-Cold (Fisher's ratio 1) and Good-Excessive (Fisher's ratio 2) for all 12 features.
4. Feature selection. Select three features from the feature pool based on the following criterion: max (Fisher's ratio 1 + Fisher's ratio 2). Visualize the three features in a 3D plot. Use legend to distinguish the data points from different class.
5. Classification. Develop control-chart-type monitoring limits for the selected features as follows.
   a. Plot histograms of each class for each feature.
   b. Determine appropriate monitoring limits for each feature. Explain how the limits are determined.
   c. Find the misdetection rate for each feature.
6. Sensor selection. Based the analysis above, discuss which sensor, power or force, is more useful for quality monitoring.