Using SEM Data to Detect Wafer Abnormality:
Prevent TEM Data Usage for Cost Savings and Improving Production Output

Jia-Wei Jessie Liang

Abstract

This research proposal aims to use image classification on SEM data to detect wafer abnormality and prevent the usage of TEM data, which can reduce tremendous cost in the long run—at least 36.5 million USD (for a nominal two-year R&D process). Current methods of detection largely entail process engineers sorting through every piece of abnormal wafer—not only is this time-consuming, but also this manual process generates too much "wastage" stemming from false-positives and false-negatives. This research aims to answer the question of how much better ML (machine learning) can correctly spot and declare abnormalities on wafer than process engineers can.

This research proposes a new method for abnormal wafer detection that does not simply do partitioning and correlation: This study uses an improved process flow: detect => categorize => diagnose. Detection means looking into every SEM data to catch the ones that are considered pattern abnormal; categorization entails classifying *open* or *short* type of pattern abnormal; and through learning from training and test dataset, diagnosis requires that ML has the capability to determine whether future data is considered pattern abnormal.

In this research paper, machine learning, fitted with different types of statistical methods, such as logistic regression, ADALINE, SVM, and kernel method, is implemented to demonstrate a new structure on pattern recognition: *image classification on SEM data (ICSD)*. Of these, logistic regression is the best classifier for detecting wafer abnormality. Given the limited dataset, the logistic regression has arrived at approximately 60% accuracy.

The overall benefits this research will afford the semicon industry are savings on manpower, time, and cost. Additionally, it allows detecting wafer abnormalities with improved efficiency that certainly improves output.

Introduction

The challenges associated with the design and manufacturing of leading edge integrated circuits (IC) have increased with the complexity of chip functionalities. In order to improve the product yield in the IC design and manufacturing cycle, it is important to identify the underlying factors that contribute most to yield loss.

Current practice entails that process engineers guard the wafer manufacturing process yield outcome. If the yield performance is under 80%, which is considered an unsuccessful process, then the process engineers will need to look back at the circuit diagrams in order to ascertain whether a specific section is faulty. The entire inspection process usually takes about two weeks.

Process engineers use SEM[1] and TEM[2] (two types of electron microscope) to look for abnormal spots in wafer. SEM focuses on a sample wafer's surface and its composition, while TEM seeks to see what is inside or beyond the surface. However, in preparation for TEM imaging, not only does the process requires much more manpower since part of the inspection process requires that the wafer be sliced, but also costs much more because the wafer itself is damaged. SEM, by contrast, is easy to obtain since it is only a picture taken of the wafer surface.

Based on their experience and knowledge, the process engineers know where the "weak spots" are on a wafer. Weak spots are those that have high frequency of pattern abnormal[3]. For example, in the MOL (middle of line), an "open contact issue"[4] is likely to occur; while at the BEOL (back end of line), Cu line short[5] or via open issue are the likely culprits. Nevertheless, an entire wafer has at least 10,000 spots that should be inspected. Current industry practice dictates that when pattern abnormal is present, engineers could only look into approximately two to three weak spots (such as MOL and BEOL). This process is categorized as problem-identification. Most often than not, problems are not resolved by simply looking into these weak spots. What is required in reality, however, is that the process engineers widen their inspection more extensively and comprehensively to capture other potentially faulty spots on the wafer. In the name of efficiency, the shortchanging of the in point of fact required process will create false conclusions due to (1) only recognizing a failing pattern on the wafer as belonging to a class of known issues, and (2) miss abnormal patterns that potentially reflects a yield issue.

---

[1] SEM (scanning electron microscope): a type of electron microscope that produces images of a sample by scanning the surface with a focused beam of electrons. The electrons interact with atoms in the sample, producing various signals that contain information about the surface topography and composition of the sample.

[2] TEM (transmission electron microscope): a microscopy technique in which a beam of electrons is transmitted through a specimen to form an image. The specimen is most often an ultrathin section less than 100 nm thick or a suspension on a grid. An image is formed from the interaction of the electrons with the sample as the beam is transmitted through the specimen.

[3] Pattern abnormal can be catagorized in three classes:

(1) Failing frequency based statistics indicating abnormal yield fluctuations
(2) Failing location based statistics revealing abnormal concentration of failures (clustering)
(3) Spatial failing patterns that can be correlated to some special causes, such as scratches from material handling, non-uniformities in film thickness, edge-die effects

[4] "Open" is defined as an open or incomplete circuitry in which metallic lines that should supposedly touch or continue are disrupted (contact open, via open). (1) "via" is defined as an open or hole where metals on an upper layer fail to touch or connect to metals on a lower layer; (2) "contact" is defined as an open or hole where metals on an upper layer fail to touch or connect to Si on a lower layer

[5] "Short" is defined as two metallic lines (parts of the embedded circuitry) touching (due to defect; metal short) where they should not be touching.

Current industry practice and research papers in this domain stop at the problem-identification step, which are pattern abnormal detection and pattern abnormal classification. The problem-identification step usually entails monitoring some failing frequency statistics based on counting the number of failing dies[6]. Beyond simply identifying failures, the pattern-recognition step this study propounds, however, is more complicated, the implementation of which requires training ML algorithm to recognize patterns, accelerate learning, and automate its learning process through running various statistical models to ascertain the best method to achieve optimum accuracy.

Literature Review

Analysis of wafer abnormalities has been a common practice in the semiconductor industry for many years, and there are so many related papers and research studies. Almost all of their ultimate goal is to apply ideas in creating a high-quality production line for higher yield.

A Pattern Mining Framework for Inter-Wafer Abnormality Analysis (2013 IEEE) presents three pattern mining methodologies for wafer abnormality analysis: Abnormality Detection[7], Perspective Search[8], and Similarity Search[9]. *Identifying Systematic Spatial Failure through Wafer Clustering (2016 IEEE)* proposes another method that consists of SVD (singular value decomposition), hierarchical clustering, and dictionary learning to take the testing results (pass or fail) of a number of dies over different wafers, then cluster all these wafers according to their failures, which in the end identify the underlying spatial failure patterns. *Process Monitoring through Wafer-level Spatial Variation Decomposition (IEEE, 2016)* introduces a spatial decomposition method for breaking down the variation of a wafer to its spatial constituents, based on a small number of measurements samples across the wafer.

To summarize these various studies, there are two main methods on detecting wafer abnormalities: the first is using partitioning and the second is using correlation statistical methods.

1. partitioning the wafers into groups with similar spatial signatures, and classify wafers through clustering analysis in order to detect abnormal wafers and plan future production, which will help process engineers to focus on the failure causes associated with the significant yield loss

---

[6] Wafer die: a die is a small block of semiconducting material on which a given functional circuit is fabricated (Wikipedia)

[7] Abnormality detection: identify wafers with patterns that are abnormal as compared to other wafers. Given a large population of wafers, the methodology identifies wafer with abnormal patterns based on a test or a group of test.

[8] Perspective search: given a wafer of interest, identify a test perspective (i.e. a test or a group of test) that exposes a pattern on the wafer, where the pattern is novel as compared to other wafers. Given a wafer of interest, the methodology searches for a test perspective that reveals the abnormality of the wafer.

[9] Similarity search: given a known abnormal pattern, detect wafers containing similar patterns. Given a particular pattern of interest, the third methodology implements a monitor to detect wafers containing similar pattern.

2. identify the most prominent spatial variation component and the main contributor to yield variation by analyzing the correlation between the estimated weight vector and the yield. This could further predict yield using correlation functions which map the estimated weight vector to the actual yield

Improving on existing studies and methodologies, this research proposes a new method for pattern abnormal detection not only by just doing partitioning and correlation, but also by improving and expanding on the process by appending diagnosis to detection and categorization, in order to provide the capability to reveal previously undetected pattern abnormal that affects production yield.

Methodology

In practice, many measurements are taken to verify whether chips on a wafer meet a required design specification. Hence, the result of such testing is binary. The binary nature of the measurements (pass or fail) justifies the use of classification statistics methods. In order to improve on the current method of relying on process engineers to classify pattern abnormal, the methodology chosen as the better alternative is to train ML algorithm to achieve optimum accuracy. Image classifying technique is applied to identify significant abnormal patterns on wafers. Then a clustering method is applied to group these failing dies, followed by a classification method to put each group into a known category of issue—open or short. This research differentiates itself from others in *not assuming* a set of known pattern abnormal (problematic pattern) *as known in advance*.

The specific methodology employed is that of image recognition and classification. The mathematical logic behind the algorithm can be summarized as:
Given a wafer w and a set C of known problematic pattern categories {c1, c2, ….cn}, decide if there is a failing pattern on w that closely resembles a ci belonging to C. If it belongs to C, then will further identify it is open or short.

The data utilized is the SEM data gathered from the demos that the researcher uses for customers. The data is divided into a training (80%)[10], and a test[11] (20%) set. For analysis and classification, this research proposal employs different statistical classification models, which can be seen in the codes below, to train a classifier for sorting these images into two classes: *open* and *short*. In addition, the test dataset, which is generally well curated to sample data that spans the various classes that the model would face, when used in the real world will identify which model is the best classifier.

[10] Training dataset: the sample of data used to fit the model, which is the actual dataset that we use to train the model (weights and biases in the case of Neural Network). The model sees and learns from this data.

[11] Test dataset: The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset.

```python
from PIL import Image
from glob import glob
import pandas as pd
import numpy as np
from sklearn import svm
from sklearn.decomposition import PCA
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import LogisticRegression

def imToArray (path):
    # First, read and convert the picture into black and white (reduce the d
    pic = Image.open(path).convert('L')
    # Resize the picture to the same size (reduce dimension as well)
    pic = pic.resize((125,75))
    # Convert picture into nparray
    pic = np.array(pic)
    # Convert numpy.ndarray into vector
    x1 = np.array(pic.flatten())
    return x1;


topic = np.array(["Open", "Short"])

filetrain = np.array(['/Users/jessieliang/Desktop/SEM Data/Open/Train/image*.jpg',
                      '/Users/jessieliang/Desktop/SEM Data/Short/Train/image*.jpg'])

filetest = np.array(['/Users/jessieliang/Desktop/SEM Data/Open/Test/image*.jpg',
                     '/Users/jessieliang/Desktop/SEM Data/Short/Test/image*.jpg'])

data = np.array(['/Users/jessieliang/Desktop/SEM Data/Open_data.csv',
                 '/Users/jessieliang/Desktop/SEM Data/Short_data.csv'])

for j in range(0,len(filetrain)):
    # Read all the image names from the train folder
    train_filelist = glob(filetrain[j])

    # Make the train_filelist into a pandas dataframe
    name_train = pd.Series(train_filelist)
    for i in range(0,len(train_filelist)):
        name_train[i] = name_train[i].split('/')[-1]
    name_train = pd.DataFrame(name_train)
    name_train.columns = ["name"]

    # Read all the train images
    x = list()
    for i in range(0,len(train_filelist)):
        item = imToArray(train_filelist[i])
        x.append(item)
    x = np.array(x)

    # Read the label of each picture
    y = pd.read_csv(data[j])

    # Convert y into pandas dataframe
    # Match the name of the file with train_filelist
    for i in range(0,len(y)):
        y.ix[i,0] = y.ix[i,0].split('_',1)[-1]

    # Left join these two tables
    ynew = pd.merge(name_train, y, on='name')
    label = np.ravel(ynew['final'])

    # Read the test dataset
    test_filelist = glob(filetest[j])

    # Make the test_filelist into a pandas dataframe
    name_test = pd.Series(test_filelist)
    for i in range(0,len(test_filelist)):
        name_test[i] = name_test[i].split('/')[-1]
    name_test = pd.DataFrame(name_test)
    name_test.columns = ["name"]
```

```python
# Read all the test images
xtest = list()
for i in range(0,len(test_filelist)):
    item = imToArray(test_filelist[i])
    xtest.append(item)
xtest = np.array(xtest)

# Get the labels for test dataset
ytest = pd.merge(name_test, y, on='name')
labeltest = np.ravel(ytest['final'])

# Use PCA to reduce dimension
n = np.array([100,50])
pca = PCA(n_components = n[j], svd_solver='full')
x1 = pca.fit(x)
x2 = x1.transform(x)
xtest1 = x1.transform(xtest)

# Fit them different models
#1. Logistic regression
lgmodel = LogisticRegression()
lgmodel = lgmodel.fit(x2,label)

#2. Ada line
lnmodel = LinearRegression()
lnmodel = lnmodel.fit(x2,label)
lnpredict = lnmodel.predict(x2)
count = 0
for i in range(0,len(lnpredict)):
    if lnpredict[i] > 0:
        lnpredict[i] = 1
    if lnpredict[i] <= 0:
        lnpredict[i] = 0
for i in range(0,len(lnpredict)):
    if lnpredict[i] == label[i]:
        count = count + 1
ada_train_accuracy = float(count)/float(len(lnpredict))

lntestpredict = lnmodel.predict(xtest1)
count = 0
for i in range(0,len(lntestpredict)):
    if lntestpredict[i] > 0:
        lntestpredict[i] = 1
    if lntestpredict[i] <= 0:
        lntestpredict[i] = 0
for i in range(0,len(lntestpredict)):
    if lntestpredict[i] == labeltest[i]:
        count = count + 1
ada_test_accuracy = float(count)/float(len(lntestpredict))


#3.SVM
svmmodel = svm.LinearSVC()
svmmodel = svmmodel.fit(x2,label)

#4. Kernal SVM
ksvmmodel = svm.SVC(kernel = 'rbf')
ksvmmodel = svmmodel.fit(x2,label)


# Print the result
print(topic[j]+":")
print("Logistic Accuracy for train dataset:"+str(lgmodel.score(x2,label)))
print("Logistic Accuracy for test dataset:"+str(lgmodel.score(xtest1,labeltest)))
print("adaline Accuracy for train dataset:"+str(ada_train_accuracy))
print("adaline Accuracy for test dataset:"+str(ada_test_accuracy))
print("SVM Accuracy for train dataset:"+str(svmmodel.score(x2,label)))
print("SVM Accuracy for test dataset:"+str(svmmodel.score(xtest1,labeltest)))
print("Kernal SVM Accuracy for train dataset:"+str(ksvmmodel.score(x2,label)))
print("Kernal SVM Accuracy for test dataset:"+str(ksvmmodel.score(xtest1,labeltest)))
print("\n")
```

Results

The results collated here show the level of accuracy from training and test using various statistical models. From the data in Table 1.1 to Table 1.4 below, logistic regression is considered the best classifier.

1. Logistic Regression

Category	Accuracy in Training	Accuracy in Test
Open	0.8334	0.5968
Short	0.8696	0.6099

Table 1.1 Logistic regression on training and test accuracy

2. ADALINE

Category	Accuracy in Training	Accuracy in Test
Open	0.5852	0.4516
Short	0.6014	0.4905

Table 1.2 ADALINE on training and test accuracy

3. SVM

Category	Accuracy in Training	Accuracy in Test
Open	0.6881	0.5000
Short	0.6739	0.5157

Table 1.3 SVM on training and test accuracy

4. Kernel Method

Category	Accuracy in Training	Accuracy in Test
Open	0.7043	0.5887
Short	0.9111	0.5157

Table 1.4 Kernel Method on training and test accuracy

These tables show that using the logistic regression model yields the highest accuracy given test dataset.

Discussion

Most of the accuracy rates are around 50%, which is not considered high, as the dataset is not large enough. In order to improve the accuracy rates, build a SEM picture library and implement of more advanced techniques have to be done.

- Next Step: Improvement on this research proposal

Build a SEM picture library. Import every SEM data generated.

Once there is a bigger data library, I would use TensorFlow to do more in-dept learnings, and further build a model to teach machine how to detect pattern abnormal. If a "pattern abnormal" is detected during the running of the first wafer, then the parameters could be revised. The entire production batch will be deemed "successful" once the BKM (Best Known Method) is established. Putting more of these datasets on training will ensure a more accurate model for validating and testing.

Need more advanced techniques to dig deeper into every images

Dig deeper into images in order to find the range in the SEM data that shows a high probability of "pattern abnormal" but in reality will not cause problems. For example, in the SEM data, if it shows two lines coming too close to each other, tradition models will categorize this condition as pattern abnormal. In reality, however, there is a certain range that this condition will not cause problems. The aim is to better ascertain a more precise range... in order to lessen false-positives and to correctly classify false-negatives as abnormal

- Challenges faced by this research proposal

1. Similarity can be rotation invariant such that a pattern rotated by a certain degree is recognized as the same pattern
2. The presence of random defects and variations that can mask the underlying systemic failure pattern
3. Few wafers can have abnormal signatures that may be due to equipment malfunction. It is important to detect these special wafers as outliers and avoid including them in the desired clusters

Conclusion

The accuracy results show that even with a limited dataset, the algorithm is still able to establish a viable level of accuracy. This is an indication that given a large enough dataset, the different statistical models will provide further valuable insights, which in the end gian higher yield with less time and money.

By rough estimation, when a wafer is destroyed in the process of obtaining TEM imaging data, the cost associated with one wafer is $50,000 USD. The R&D usually takes two years. In a minimum scenario of just slicing one wafer per day for TEM analysis, the annual cost is 36.5 million USD (=50,000*365*2). Above and beyond cost savings is the time saved when SEM data is used instead of TEM data. Process engineers will be able to identify problems on the production line immediately, preventing the production line to continue making the same mistakes. Thus, using ML to automate detection, classification, and diagnosis prevents further wastage and improves production output.

Appendix

- Detailed Differentiation on SEM and TEM Data:

The main difference between SEM and TEM is that SEM creates an image by detecting reflected or knocked-off electrons while TEM uses transmitted electrons (electrons which are passing through the sample) to create an image. As a result, TEM offers valuable information on the inner structure of the sample, such as crystal structure, morphology and stress state information, while SEM provides information on the sample's surface and its composition.

	Scanning Electron Microscope (SEM)	Transmission Electron Microscope (TEM)
Method	Both instruments use electrons or electron beams	
Result	Both images produced are highly magnified and offer high resolution (TEM comparatively has high resolution and magnifying power)	
Technique	Scans the surface of the sample by releasing electrons and making the electrons bounce or scatter upon impact; The machine collects the scattered electrons and produces an image	TEM processes the sample by directing an electron beam through the sample; The result is seen using a fluorescent screen
Type of Electron	Scattered electrons; The scattered electrons in SEM are classified as backscattered or secondary electrons	Transmitted electrons; There is no other classification of electrons in TEM
Sample	An SEM sample is stained by an element that captures the scattered electrons	The sample in TEM is cut thinner in contrast to a SEM sample
Presentation	Three-dimensional and are accurate representations SEM is that the area where the sample is placed can be rotated in different angles	Two-dimensional and might require a bit of interpretations
Field of View	Large	Limited
Preparation technique	Easy	Skilled, very thin sample is required

- Two High Probability Cases of Wafer Abnormal:

1. Line-end shortening

(Solid line: target; Dotted line: image printed)

2. Corner rounding

Reference

Difference Between TEM and SEM
http://www.differencebetween.net/science/difference-between-tem-and-sem/

About Train, Validation and Test Sets in Machine Learning
https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7

A Pattern Mining Framework for Inter-Wafer Abnormality Analysis (2013 IEEE)
http://mtv.ece.ucsb.edu/licwang/PDF/2013-ITC.pdf

Identifying Systematic Spatial Failure through Wafer Clustering (2016 IEEE)
https://users.ece.cmu.edu/~xinli/papers/2016_ISCAS_wafer.pdf

Process Monitoring through Wafer-level Spatial Variation Decomposition (IEEE, 2016)
https://www.utdallas.edu/~gxm112130/papers/itc13b.pdf

S. Cunningham and S. MacKinnon. Statistical methods for visual defect metrology. IEEE Trans. Semi. Manuf., vol. 11, no. 1, pp. 48-53, 1998

Tao Yuan, Way Kuo, and Suk Joo Bae. Detection of Spatial Defect Patterns Generated in Semiconductor Fabrication Processes IEEE Tran. on Semi Manuf., Vol 24, No 3, 2011