Principal Component Analysis and Sparse Principal Component Analysis of Plasma CVD Process Data of a-Si:H Films

R. Iwamoto[1], K. Kamataki[1], H. Hara[1], K. Tanaka[1], D. Yamashita[1], N. Itagaki[1], D. Ikeda[2], K. Koga[1] and M. Shiratani[1]

[1]*Department of Electrical and Electronic Engineering, Kyushu University, Fukuoka, Japan*
[2]*Department of Informatics, Kyushu University, Fukuoka, Japan*

Abstract: Plasma processes demand many external parameters to be tuned. For the better tuning, process data of plasma CVD of hydrogenated amorphous silicon films is analysed by two methods: principal component analysis and sparse principal component analysis. These analyses are useful not only for identifying key external parameters and suggesting better experimental conditions, but also for giving predictive insights into experimental results.

Keywords: PCA, SPCA, process data, plasma CVD, a-Si:H

1. Introduction

Hydrogenated amorphous silicon (a-Si:H) films are mostly produced by a plasma enhanced chemical vapor deposition (PECVD) method and used for fabricating crystalline Si solar cells as well as Si thin film solar cells. The current highest efficiency of a-Si:H single junction solar cells in the world is 10.2% [1, 2]. The advantages of a-Si:H solar cells include flexibility and low production cost, while they have one major drawback, light-induced degradation. Upon light exposure, a-Si:H solar cells experience 20-30% drop in conversion efficiency [3]. In our laboratory, over 500 a-Si:H cells have been produced and their conversion efficiency at the initial state and stabilized state after light exposure have been measured. In the previous studies [4], we have found correlation between SiH_2 bond density in films and light-induced degradation, and have successfully reduced the efficiency drop down to 2.4% by using a cluster elimination filter and achieved 9.1% stabilized efficiency.

In general, plasma processes demand many external parameters to be tuned via trial-and-error, and the number of tuning parameters can be enormous in some practical cases. Physical and chemical parameters of reactive plasma in a reactor and products also have many characteristics. In such cases, data-based statistical or machine learning approach offers a novel way for tuning plasma processes in a short period.

Here, our plasma CVD process data of a-Si:H films is analysed by principal component analysis (PCA) [5, 6] and a new version of PCA called sparse principal component analysis (SPCA) [6].

2. Methods

2.1 Film Deposition Process

First, we explain briefly the experimental setup used for the film deposition process [4,7]. Figure 1 shows our experimental setup. Our plasma CVD reactor employs a multi-hollow electrode, frequently together with a cluster eliminating filter. The external parameters are the SiH_4 flow rate, H_2 flow rate, substrate temperature, gas pressure, film thickness, RF power, RF frequency, distance between electrode and substrate, presence/absence and type of the cluster eliminating filter, presence/absence of a 100 mesh, and type of multi-hollow electrode. The last three parameters are discrete, and the rest are continuous. Under each experiment conditions one film was deposited, and then the film was employed to fabricate16 or 7 solar cells. The rows of Table 1 are features of cell characteristics used for PCA and SPCA. The dataset was pre-processed and the data of failed cells was removed.

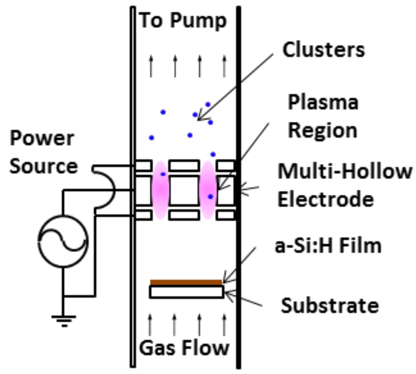

Fig. 1. Plasma CVD reactor of a-Si:H film deposition.

2.2 PCA

PCA is a method of linear transformation of high dimensional data. It transforms axes of data to new axes called principal components (PC1, PC2, …). PCs remain orthogonal and numbered in descending order of data variance along the axes. For example, if a shape of data is bread-like in 3D space, the variance along PC1 is very high, while ones along PC2 and PC3 are small. In such cases, we may ignore PC2 and PC3, and regard the data as one-dimensional. In this way, PCA can be used for dimensional reduction.

2.3 SPCA

SPCA is a family of PCA. The word “sparse” means “mostly zero”. For example, consider a linear combination $a_0x_0 + a_1x_1 + \cdots + a_nx_n$ and if the coefficients $a_0, a_1 \ldots a_n$ are sparse, that means only a small number of them are non-zero. SPCA gives sparse estimations of PCs. This is helpful when you are trying to interpret principal components. Sparsity of loadings can be controlled by changing a regularization parameter.

For PCA and SPCA, an open-source programming package of Python called scikit-learn is used [8].

3. Results and Discussion

Figure 2 shows the Eigenvalues of the normal PCs. Each of the first three components has an eigenvalues more than 1 and their sum is 91% of all variation of the data. The right side of Table 1 shows the loadings of the SPCA's components. PC1 has high correlation with FF_{initial}, $\eta_{\mathrm{stabilized}}$, and "Degradation Ratio". Thus, PC1 is interpreted as the film performance axis as a solar cell. PC2 has high correlation with η_{initial}, currents, and small correlation with *"Degradation Ratio"*, therefore PC2 is the initial performance axis. PC3 is the voltage related axis. Figure 3 shows PC1 distributions against four external parameters. These results show that, the cluster eliminating filter, SiH_4 flow rate, reactor pressure, and distance between substrate and electrode play key roles in determining the conversion efficiency of a-Si:H solar cells. From the viewpoint of physical mechanisms, these four parameters are reasonable as follows. First, clusters contribute mainly to SiH_2 bond formation in films and tend to enhance the light induced-degradation. The cluster eliminating filter reduces this contribution and hence PC1 depends on the filter type as in Fig. 3(a). Note that "filter type 0" is "w/o filter" and the rest numbers correspond to fineness of filter. Secondly, the SiH_4 flow rate is inversely proportional to the gas residence time in the plasma region, thus the cluster growth rate decreases with increasing the flow rate and hence PC1 depends on the SiH_4 flow rate as in Fig. 3(b). Thirdly, the distance between substrate and electrode affects cluster growth rate, leading to the results in Fig. 3(c). Finally, the pressure play at least two major roles: 1) the higher pressure gives the longer gas residence time leading to the faster growth of clusters, and 2) the higher pressures brings about the less diffusive transport of clusters to the substrate set in the upstream region. According to the balance between these two reverse effects, there is an appropriate pressure to obtain the highest PC1.

Based on these results, we have designed new experiments focusing on the SiH_4 flow rate dependence [9]. SiH_2 bond density in a-Si:H films nonlinearly decreases with increasing the SiH_4 flow rate. Eventually, the experiments realized a lower SiH_2 bond density than previous ones. It should be noted that a-Si:H films with the lower SiH_2 bond density tends to show lower light-induced degradation.

PCA and SPCA analyses are useful not only for identifying key external parameters and suggesting better experimental conditions, but also for giving new insights into experimental results.

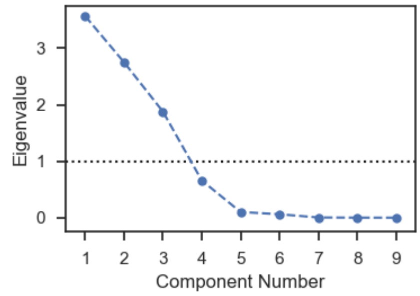

Fig. 2. Eigenvalues of normal PCA's components.

Table 1. Features of cell data used for PCA/SPCA and loadings of PCs of SPCA components.

	Feature	PC1	PC2	PC3
Y0	$J_{\mathrm{sc\ initial}}$	0	0.616	0
Y1	$V_{\mathrm{oc\ initial}}$	0	0	-0.704
Y2	FF_{initial}	0	0.128	0
Y3	η_{initial}	0	0.626	0
Y4	$J_{\mathrm{sc\ stabilized}}$	0	0.427	0.083
Y5	$V_{\mathrm{oc\ stabilized}}$	0	0	-0.705
Y6	$FF_{\mathrm{stabilized}}$	0.654	0	0
Y7	$\eta_{\mathrm{stabilized}}$	0.591	0	0
Y8	*Degradation Ratio*	-0.471	0.175	0

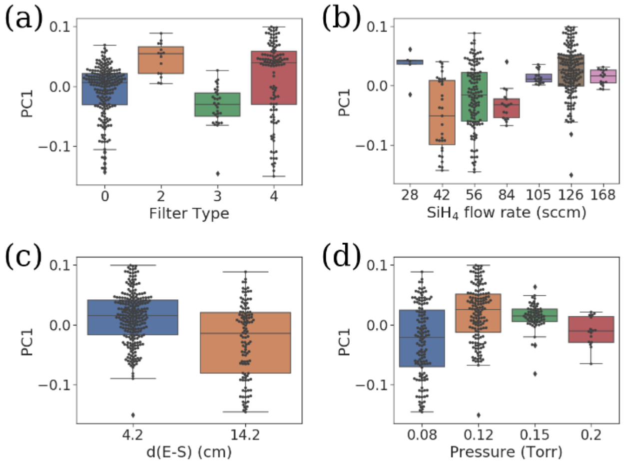

Fig. 3. Box plots showing changes of distributions regards to varying specific external parameters along PC1 obtained with SPCA. Dots represent data points. Each x-axis is (a) filter type, (b) silane flow rate, (c) distance between electrode-substrate, (d) pressure of reactor respectively.

4. References

[1] M. A. Green, *et al*. Prog. Photovoltaics Res. Appl. 27, 3 (2019).
[2] T. Matsui, *et al*., Jpn. J. Appl. Phys. 54, 6 (2015).
[3] Y. Kim, J. Vac. Sci. Technol. A 36, 050601 (2018).
[4] K. Keya, *et. al*., Jpn. J. Appl. Phys. 55, 07LE03 (2016).
[5] I. T. Jolliffe, Springer Ser. Stat. 98, 487 (2002).
[6] S. Yi, *et al*., Pattern Recognit. 61, 524 (2017).
[7] S. Toko, *et al*., Jpn. J. Appl. Phys. 55, 237 (2016).
[8] F. Pedregosa, *et al*., J. Mach. Learn. Res., 12, 2825 (2011).
[9] T. Kojima, *et al*., Plasma Fusion Res., 13, 1406082 (2018).