Advanced analysis of internal quantum efficiency measurements of GaAs solar cells using machine learning

Zubair Abdullah-Vetter, Brendan Wright, Thorsten Trupke, Ziv Hameiri

University of New South Wales, Sydney, Australia

*z.abdullahvetter@unsw.edu.au

1. Abstract

Internal quantum efficiency measurements are an important characterisation technique that provides critical information on the performance of solar cells. Common methods to analyse these measurements include qualitative comparisons to the ideal quantum efficiency, identifying current losses at specific regions of the cell, and extracting key electrical and optical parameters. Extracting those parameters often requires time-consuming (and often manual) fitting approaches. This study proposes using machine learning algorithms to automatically extract these parameters for gallium arsenide solar cells with a very high level of accuracy. The proposed method will assist in unlocking the full potential of quantum efficiency measurements as a powerful characterisation tool for a wide range of solar cells.

2. Introduction

The spectral response is the current output of a device under different wavelengths of incident light; it reveals crucial information about the device's performance [1]. The internal quantum efficiency (IQE) is obtained by measuring the incident monochromatic illumination intensity, the output current, and the reflectance, all as a function of wavelength [1]. Commonly, IQE is defined as the ratio of the number of carriers collected to the number of photons that have penetrated the device. IQE measurements are often conducted to obtain information about the current losses associated with specific solar cell regions such as parasitic absorption at the front or rear surfaces. IQE measurements are also used to calculate the device’s expected short circuit current [2].

When analysing IQE measurements of solar cells, users often attempt to fit their data by implementing analytical or numerical methods, or simulating the solar cells in software such as PC1D [3], SunSolve [4], Griddler [5], or others [6], [7]. These methods can be time-consuming and frequently require deep knowledge about the solar cell and reasonable assumptions regarding its electrical and optical parameters, making the process quite challenging. Furthermore, different analytical approaches often need to be used simultaneously as each method may only be suitable for a specific wavelength range [8]. These limitations often lead to the utilisation of only part of the full potential of IQE measurements of gallium arsenide (GaAs) solar cells, focusing only on qualitative analysis of the expected losses.

An additional, although rarely discussed, challenge is associated with the fitting of non-unique solutions. These cases result from multiple combinations of parameters that provide near-identical fits to the IQE measurement, although some of these combinations are far from the actual values.

Therefore, to overcome these limitations this study proposes using a machine learning (ML) pipeline to accurately extract electrical parameters from IQE measurements of GaAs solar cells. We also show that the correct utilisation of powerful ML models can help to address cases where non-unique solutions exist. This automated approach is shown to unlock the full potential of IQE measurements and will strongly support the research and development activities of GaAs solar cells.

3. Methodology

A. Dataset

The Python-based program SolCore [7] was used to generate a large training dataset of 40,000 GaAs solar cells following the structure suggested by Tobin et. al [9]. This structure consists of a

front $Al_{0.83}Ga_{0.17}As$ window (29.3 nm), p-type emitter (660 nm), n-type base (3,600 nm), and $Al_{0.3}Ga_{0.7}As$ rear buffer layer (1,000 nm) [9]. A dual layer of ZnS/MgF_2 (50 nm / 100 nm) antireflection coating (ARC) was applied on the front side.

The parameters varied in the simulation were the diffusion length of the minority carriers of the emitter (L_e) and bulk (L_h), and the surface recombination velocity of those regions (S_p and S_n, respectively). The resulting IQE curves were obtained at 10 nm steps in the wavelength range of 300 – 1,000 nm. The effect of each parameter on the IQE curve is shown in Fig. 1. As expected, the IQE in the short and long wavelength regions increases with longer L_e and L_h, respectively. On the other hand, increasing the value of S_p and S_n reduces the IQE in the short and long-wavelength regions, respectively. Notice that the IQE is much less sensitive to the bulk terms (L_h and S_n), so it was expected that the trained models would have higher errors predicting these terms.

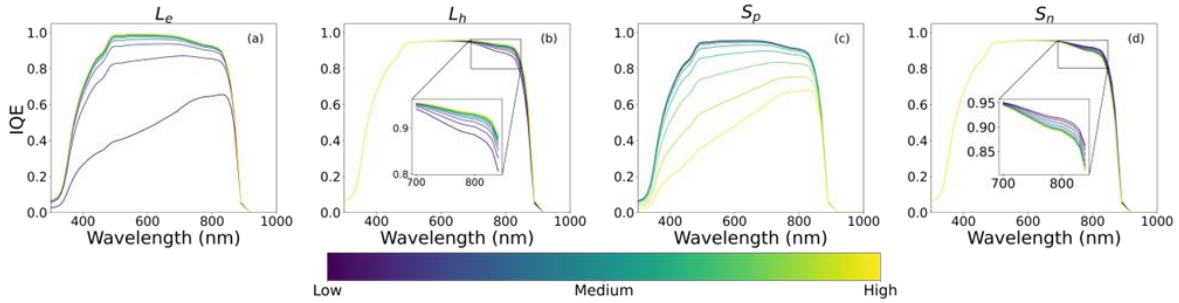

Figure 1. The effect of (a) L_e, (b) L_h, (c) S_p, and (d) S_n on the IQE measurements

B. Machine learning

The ML pipeline involves an array of random forest (RF) models [10]. The RF model is a well-suited solution for this regression problem due to the large collection of uncorrelated estimators (trees) that make up the model (forest). The proposed approach also uses a 'chain regression' technique [11] where the predictions of the previous model are passed to the training data of the next model. This has the added advantage of utilising any correlation between the four parameters being predicted by the array of RF models. The result is a package of four RF models that predict the four electrical parameters in sequential order. The dataset is split into an 80:20 training and test split. The predictions of the test set and the resulting IQE fits are evaluated using the root mean square error (RMSE) [11] formula.

Interestingly it was found that ~14% of the measurements can be fitted with minimal error using more than one set of parameters, hence, there are many non-unique solutions. In this study, non-unique solutions are defined as combinations of parameters (L_e, L_h, S_p, and S_n) that provide great IQE fit (RMSE < 0.002) while still having a relatively large (> 10%) error in the prediction of one of the parameters. To further investigate the non-unique solutions, the outputs of all the RF trees for each parameter prediction were extracted (contrary to the standard approach that only uses the average value across all the trees). As the bulk parameters were found to be more sensitive to this issue, RMSE heatmaps were generated for all the combinations of the bulk parameters predicted by the individual RF trees. Using these heatmaps, all the solutions that provide a good IQE fit can be identified (see Section 4).

To overcome the challenge of non-unique solutions and further improve the prediction results, two feature engineering steps were applied before the training of the RF models. The first step was calculating the IQE difference array (the differences between two IQE measurement points taken 10 nm apart). The second step was based on principal component analysis (PCA) [12] and it was only applied for training the RF models predicting the bulk terms, L_h and S_n. It was found that this mixed feature engineering approach between the emitter and bulk significantly improved the accuracy of the extracted parameters compared to when only the IQE measurements were used to train the four models.

4. Results

An example of a non-unique solution and the corresponding RMSE heatmap is presented in Fig. 2. As shown in Fig. 2(a), despite the significant error in L_h, the measured and predicted IQE are almost identical (RMSE of 0.0006). The RMSE calculated for each of the combinations of the predicted bulk parameters is plotted in the heatmap in Fig. 2(b). The red star is the true (input) combination, whereas the blue diamond is the combination of bulk parameters predicted by the standard RF chain model (average of all the trees). As can be seen, a diverse range of bulk parameter combinations results in RMSE < 0.002.

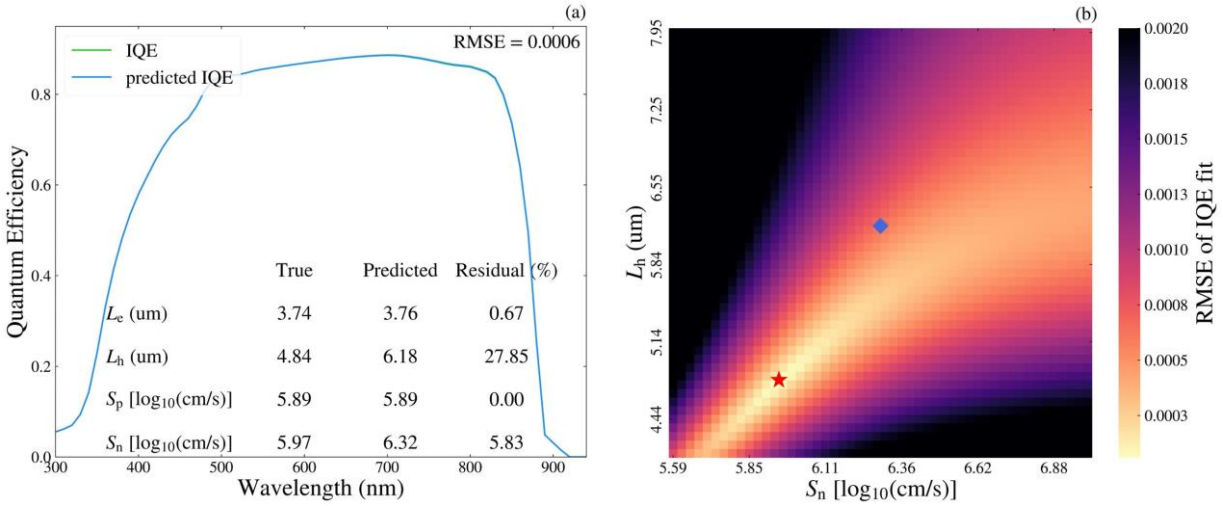

Figure 2. An example of a non-unique solution case from the test set. The true and predicted IQE are plotted in (a) with an inserted table of the four parameters and their residuals. A heatmap of the bulk terms in (b) shows the range of values that provide a good IQE fit. The blue diamond is the ML prediction while the red star is the true value.

Surprisingly, the feature engineering steps can minimise this challenge and obtain accurate parameters despite the non-unique solution problem. The predicted vs true plots of each of the four parameters after the feature engineering steps are shown in Fig. 3. The RMSE values show that the ML-based method achieves very accurate predictions for all four parameters. Furthermore, most errors can be described by a decreased sensitivity of the IQE at certain ranges for those parameters. In Fig. 3(a), the higher errors in L_e are caused by the combination of S_p dominating the emitter at larger values and diffusion lengths that are much greater than the width of the emitter. The IQE is less sensitive in both cases and so the ML model struggles to obtain the correct values of L_e. Similarly, in Fig. 3(b), the higher errors in S_p are caused by the low values of L_e dominating the region. The higher errors found in Fig. 3(c) are due to the lower sensitivity of the IQE at the extreme ranges of S_n, with a similar issue in Fig. 3(d) for larger values of L_h. The emitter parameters also impact the long wavelength region of the IQE measurements [see Fig. 1(a) and 1(b)], especially at low values of L_e. In such cases, it was expected to cause further challenges for the bulk parameter predictions. This effect is represented in Fig. 4, where the cases with higher prediction errors in L_h and S_n also had low values of L_e. Nevertheless, the ML pipeline still achieves very accurate predictions for most of the test set. Investigating the IQE fits of the test set indicated that only 0.7% of the 8,000 samples had an RMSE above the threshold of 0.002 (bad fit). Of the good fits, only 0.4% still encountered the non-unique solution issue (RMSE < 0.002 while still having a relative error above 10% for at least one of the parameters).

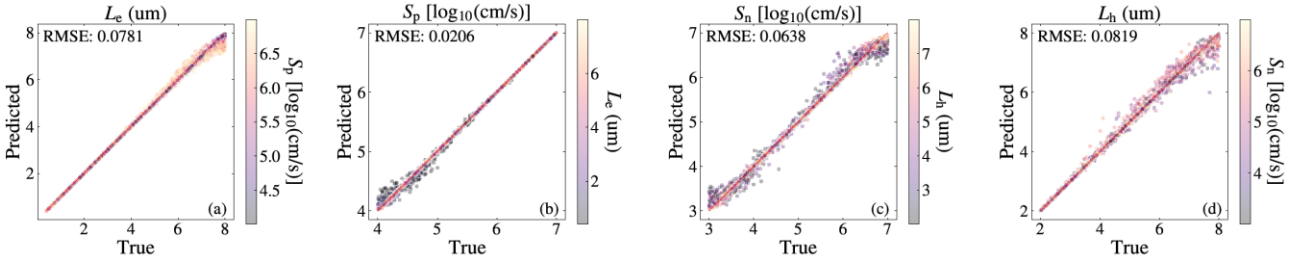

Figure 3. Predicted vs true plots are presented in the chain order: a) L_e, b) S_p, c) S_n, and d) L_h. The colour bars represent the true values of the corresponding parameter in the same region of the solar cell.

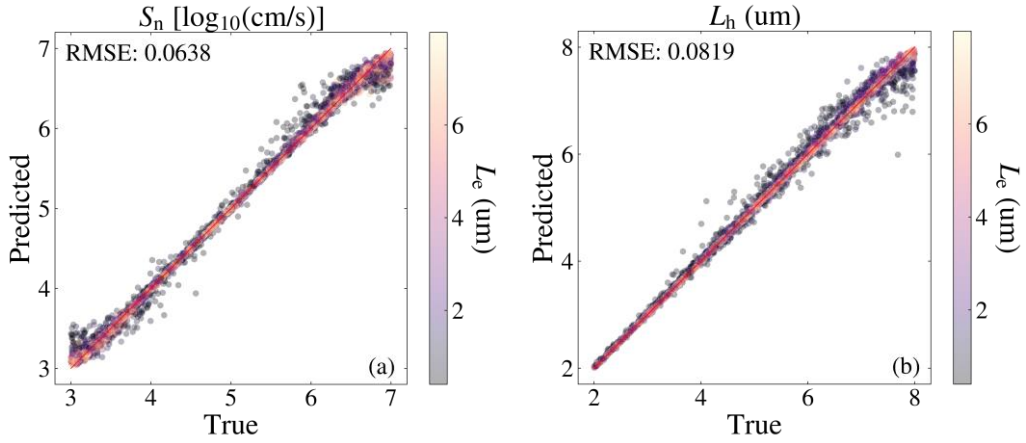

Figure 4. The same predicted vs true plots of a) S_n and b) L_h from Fig. 3(c) and 3(d), but the applied colour bars are the true values of L_e in each sample.

These results show that the proposed ML pipeline can obtain very accurate predictions for the four electrical parameters of GaAs solar cells. However, the training data used has been simulation data, which is absent of the expected noise in typical IQE measurements. Therefore, in the final conference proceedings, the resilience of this approach to noise will be investigated with both simulated noise and real GaAs solar cell IQE measurements.

5. Conclusion

This study proposes an ML-based method to predict four critical performance parameters from IQE measurements of GaAs solar cells. The method employs a mixed feature engineering approach for the emitter and bulk terms while also applying a chain regression to utilise any correlation between the parameters. Remarkably, training the RF models in this manner made them very resilient to the challenge of non-unique solutions where a large range of L_h and S_n would result in very good fits of the measured IQE. This training method can be used for different solar cell structures, unlocking the full potential of IQE measurements for the research and development of GaAs and other types of cells.

6. References

[1] J. S. Hartman and M. A. Lind, “Spectral response measurements for solar cells,” *Sol. Cells*, vol. 7, no. 1, pp. 147–157, 1982,

[2] B. Fischer, “Loss Analysis of Crystalline Silicon Solar Cells using Photoconductance and Quantum Efficiency Measurements,” PhD Thesis, Konstanz University, 2003.

[3] D. A. Clugston and P. A. Basore, “PC1D version 5: 32-bit solar cell modelling on personal computers,” in *26th IEEE Photovoltaic Specialists Conference*, 1997, pp. 207–210.

[4] “SunSolve™.” https://www.pvlighthouse.com.au/sunsolve (accessed 2022).

[5] J. Wong, “Griddler: Intelligent computer aided design of complex solar cell metallization patterns,” in *39th IEEE Photovoltaic Specialists Conference*, 2013, pp. 0933–0938.

[6] A. Fell, “A free and fast three-dimensional/two-dimensional solar cell simulator featuring conductive boundary and quasi-neutrality approximations,” *IEEE Trans. Electron Devices*, vol. 60, no. 2, pp. 733–738, 2013,

[7] D. Alonso-Álvarez, T. Wilson, P. Pearce, M. Führer, D. Farrell, and N. Ekins-Daukes, “Solcore: a multi-scale, Python-based library for modelling solar cells and semiconductor materials,” *J. Comput. Electron.*, vol. 17, no. 3, pp. 1099–1123, 2018,

[8] P. A. Basore, “Extended spectral analysis of internal quantum efficiency,” in *23rd IEEE Photovoltaic Specialists Conference*, 1993, pp. 147–152.

[9] S. P. Tobin *et al.*, “Assessment of MOCVD- and MBE-growth GaAs for high-efficiency solar cell applications,” *IEEE Trans. Electron Devices*, vol. 37, no. 2, pp. 469–477, 1990,

[10] L. Breiman, “Random Forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001,

[11] F. Pedregosa *et al.*, “Scikit-learn: machine learning in python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[12] I. Jolliffe, “Principal Component Analysis,” in *International Encyclopedia of Statistical Science*, M. Lovric, Ed., Berlin, Heidelberg: Springer, 2011, pp. 1094–1096.