Extended Solar Cell Parameters - General Purpose Descriptive I/V Parameters for Solar Cells

B.E. Pieters
IEK-5 Photovoltaik, Forschungszentrum Jülich, Germany
(Dated: September 12, 2023)

For the automated analysis of I/V-characteristics of solar cells and modules, descriptive parameters are essential. In particular with the rise in machine-learning techniques and the related increase data volumes, there is a need for good, general purpose, descriptive parameters. The most commonly used descriptive parameters for I/V are the standard solar cells parameters, consisting of V_{oc}, I_{sc}, V_{mpp}, and I_{mpp}. Also other representations may be considered, such as one diode model parameters corresponding to a particular I/V. However, these representations are very coarse and cannot distinguish or represent many common (non-ideal) features of an I/V (e.g. an S-shape). In this work we propose an extended set of solar cell parameters, which are well defined, and easy to determine. We evaluate the effectiveness of the extended solar cell parameters by reconstructing the I/V from the extracted parameters. This allows one to "measure" information loss. We compare the accuracy of our parameters with other commonly used curve models for I/V, namely the one diode model, and the Karmalkar-Haneefa model. The models are applied to a large set of I/V (about 2.2 million curves), covering a wide range of technologies and conditions. We demonstrate our extended solar cell parameters consistently provide an accurate description of nearly all I/V in these datasets. Furthermore, we present our I/V analysis tool which we use to process these datasets. This tool is fast and capable of extracting the extended solar cell parameters, as well as parameters for the one diode model and the Karmalka-Haneefa model. Finally, we exemplary show how the extended solar cell parameters may be used to detect partial shading in outdoor data, by training a simple random-forest classifier based on extended solar cell parameters.

I. INTRODUCTION

Current-Voltage (I/V) characteristics are commonly used to evaluate Photo-Voltaic (PV) performance. With rapid development of Machine Learning (ML) techniques, and the ever increasing volumes of data available for analysis, the need arises to efficiently analyze I/V-curves. The most widely used descriptive parameters for I/V-characteristics are the Standard Solar Cell Parameters (SSP), which consist of the open-circuit voltage (V_{oc}), the short-circuit current (I_{sc}), the maximum-power-point voltage (V_{mpp}), and the maximum-power-point current (I_{mpp}). The SSP are simple to determine, standardized, and convey direct information on the solar cell performance. Such simple and standardized parameters allow meaningful comparisons of different I/V-characteristics, and thus serve many purposes. However, reducing an I/V to its SSP is a (strong) reduction in dimensionality, and therefore leads to information loss. For example, it is not possible to distinguish an I/V with an S-shape from one with high series resistance using only the SSP. In this work we aim to extend the SSP to obtain solar cell parameters which preserve more detailed shape information, i.e. allow a better I/V feature extraction than the SSP.

Many authors have worked on feature extraction from I/V curves. Besides using the SSP, fitting a one diode model is a common approach [1–5]. The one diode model parametrization provides additional information on the curve beyond the information provided by the SSP (5 parameters instead of 4). Furthermore, because the diode model is based on physics, its parameterisation provides directly physically meaningful information. The parametrization of the diode model is the topic of a vast number of papers. A quick overview of the proposed methods can be obtained by looking at some review papers on the topic, see for example [2–5]. The large volume of literature on this topic is certainly an expression of great interest in the parametrization of this equation. However, it is also an expression of just how hard it is to obtain a robust and efficient parametrization method. This hampers the application of diode model parameters as I/V curve features. An additional downside of directly using the diode parameters is that the correlated nature of the diode parameters make the direct comparison of extracted values difficult (e.g. caution is required when comparing the saturation current density of solar cells with different ideality factors). Furthermore, the correlated nature of the parameters, in combination with the complicated non-linear parametrization methods, can easily lead to ill conditioned parametrization procedures, where small differences in the I/V-characteristics lead to large differences in obtained parameters.

The work by Ma et al. should be discussed in more detail as the aims are fairly similar to our work and this work also extends on the SSP [6]. In this paper the SSP are extended with the slopes at short-circuit and open circuit [6]. Furthermore, the I/V-characteristics are classified into partially shaded (with steps) and non partially shaded, and for the partially shaded curves additional features are extracted [6]. In contrast to the work by Ma et al., we propose to use a single, more extended, set of standardized parameters for all curves, which simplifies both the parameter extraction and the subsequent analysis.

The here proposed Extended Solar cell Parameters (ESP) aim to be general purpose I/V parameters. As

such, we aim for a compact, standardized description of the I/V-characteristics with meaningful and easy to understand parameters. Furthermore, the extraction of the ESP should be straight forward and efficient, allowing the processing of large data volumes. Naturally, one would like the parameters to be always determinable in an unambiguous manner. For this reason we restrict ourselves for the ESP to parameters which, like the SSP, describe voltages, currents, and slopes at several key points in the curve. Here we try to strike a balance between the number of parameters and the level of detail we can describe, i.e. the level of detail must suffice to accurately describe the vast majority of I/V-characteristics whilst keeping the number of parameters small.

The 4 parameters of the SSPs describe three key coordinates (voltage,current) in the IV characteristics, namely, the short circuit, open-circuit and maximum power points. In addition, due to the nature of a maximum power point, the SSP also describe the slope in the maximum power point ($\partial I(V)/\partial V|_{V=V_{\mathrm{mpp}}} = -I_{\mathrm{mpp}}/V_{\mathrm{mpp}}$). Our proposed ESP extends this definition to entail 5 key points and include currents, voltages, and slopes at all these key points. In total, this results in 10 parameters, including the 4 SSP.

In order to asses and compare the SSP and ESP we first determine solar cell parameters of an I/V-characteristics and, subsequently, parameterize various curve models to reconstruct the I/V, using *only* the determined solar cell parameters. The difference between the reconstructed I/V and the original measurement is a direct measure for the amount of information which was lost for a given set of parameters and the used curve model. Note that the used curve model implicitly adds information. As some curve models may be more suitable to describe solar cell I/V for specific types of solar cell technologies, we, in this paper, apply the I/V reconstruction models to a wide variety of solar cell technologies and I/V under various conditions.

For the curve models to reconstruct the solar cell IV, a wide variety of models are available. The most widely used model is certainly the one diode model (1D), which has 5 parameters, namely the saturation current, I_0, thermal voltage, kT/q, ideality factor n, photo-current, I_{ph}, series resistance, R_{s}, and shunt resistance, R_{sh} (note that we do not describe temperature dependencies, i.e., we consider only a single I/V at a time). Another widely used model is the Karmalkar-Haneefa (KH) formalism [7, 8]. This model has 4 parameters, V_{oc}, I_{sc}, and two "shape parameters" γ and m.

In addition we developed two curve models which are both based on a smooth and monotonous interpolation between the key points described by the ESP. The use of interpolation models has the additional advantage that interpolation techniques are gerenrally easy to apply, and typically do not require a parametrization step to determine the model parameters. Unlike, for example, the diode model, which requires the use of non-linear least squares or other iterative optimization methods, to extract the model parameters. The here introduced interpolation methods are the "PV-Spline" and "PV-Bézier" models. The models apply different interpolation methods, namely a cubic spline and a cubic Bézier, respectively. Both interpolation methods are adapted such that in case the ESP are monotonous, the resulting interpolation curve is also monotonous.

We analyze a database of outdoor measured I/V with 9 different modules types in 5 different locations. The module technologies include crystalline silicon (c-Si), Cadmium Telluride (CdTe), Copper-Indium-Gallium di-Selenice (CIGS), and hydrogenated amorphous/microcrystalline silicon (a-Si:H/μc-Si:H) tandem. This database entails about 2.2 million I/V. We determine the ESP for the entire dataset. Subsequently we parameterize 4 different curve models, namely the KH, 1D, PV-Spline, and the PV-Bézier models. We compare the fit errors for the different models to determine how accurate the curve models and corresponding parameters describe the I/V for the various module types. Furthermore, we demonstrate the use of the ESP as a feature vector for machine learning applications by developing a simple random-forest classifier to detect I/V affected by partial shading. For many data-analysis methods it is vital to be able to process a large volume of data. To this end we developed an open source library, the PhotoVoltaic CuRve AnalyZEr (PV-CRAZE), which provides code to extract the ESP and provides a framework to fit various curve models using the ESP. Our focus was on robust and efficient methods, suitable for large datasets.

The paper is organized as follows. We first discuss the ESP in Section II. Next, in Section III, we discuss the curve models used in this paper, including the here developed monotonous interpolation methods. We breifly introduce PV-CRAZE in Section IV. In Section V we first demonstrate the curve models on several selected I/V, and subsequently analyze the large outdoor datasets to asses the ESP and the performance of the various curve models. Furthermore, we demonstrate the use of the ESP as a feature vector to detect partial shading conditions from the shape of an I/V using a random-forest classifier. We conclude with a short note on the performance of the PV-CRAZE library to extract the ESP and for the curve model parametrization for the different curve models.

II. EXTENDED SOLAR CELL PARAMETERS

A. Definition of the ESP

In this section we discuss the here used descriptive parameters for an I/V. The ESP key points are optimally chosen to provide information on the area under the I/V. The parameters are illustrated in Fig. 1. The SSP describe 3 key points in the I/V, namely the short-circuit, maximum-power, and open-circuit points. The short circuit point is described by I_{sc}, open-circuit by the V_{oc} and the maximum-power point by both I_{mpp} and V_{mpp}. For

a maximum power point we have that the derivative of power to voltage is 0, i.e.

$$\left.\frac{\partial P(V)}{\partial V}\right|_{V=V_{\mathrm{mpp}}} = 0. \tag{1}$$

From this it follows

$$\begin{aligned} \left.\frac{\partial (I(V)V)}{\partial V}\right|_{V=V_{\mathrm{mpp}}} &= 0 \\ V_{\mathrm{mpp}} \left.\frac{\partial I(V)}{\partial V}\right|_{V=V_{\mathrm{mpp}}} + I_{\mathrm{mpp}} &= 0 \\ \left.\frac{\partial I(V)}{\partial V}\right|_{V=V_{\mathrm{mpp}}} &= -\frac{I_{\mathrm{mpp}}}{V_{\mathrm{mpp}}}. \end{aligned} \tag{2}$$

Thus from the SSP we know both the current, voltage, and, the slope of the I/V at the maximum power point.

In order to extend the SSP, we add the slopes at short-circuit and open-circuit. To this end we define the parameters G_{sc} and R_{oc}

$$G_{\mathrm{sc}} = \left.\frac{\partial I(V)}{\partial V}\right|_{V=0} \tag{3}$$

$$R_{\mathrm{oc}} = \left.\frac{\partial V(I)}{\partial I}\right|_{I=0}. \tag{4}$$

In addition we propose two additional key points, the "lower quasi maximum power point" (qmp-) and the "upper quasi maximum power point" (qmp+). These points are defined using the power curve corresponding to the I/V, see Fig. 1.b. The three SSP in the power curve make a triangle from 0 power at short circuit to P_{mpp} at the maximum power point, and back to 0 at open circuit. For any strictly monotonous I/V the the actual power curve is above this triangle. We define the qmp- point as the point between short circuit and V_{mpp} with the maximum difference between the power curve and the triangle. Likewise we define the qmp+ point between V_{mpp} and V_{oc} such that the difference between the power curve and the power triangle is maximal. In Fig. 1.b the differences between the power curve and the triangle at the qmp's, are indicated with red arrows. The qmp's thus provide additional information on the area under the I/V.

For the qmp- point we derive

$$\begin{aligned} \left.\frac{\partial \left(I(V)V - I_{\mathrm{mpp}}V\right)}{\partial V}\right|_{V=V_{\mathrm{qmp\text{-}}}} &= 0 \\ \left.\frac{\partial I}{\partial V}\right|_{V=V_{\mathrm{qmp\text{-}}}} &= -\frac{I_{\mathrm{qmp\text{-}}} - I_{\mathrm{mpp}}}{V_{\mathrm{qmp\text{-}}}}. \end{aligned} \tag{5}$$

Here we use that $I(V)V - I_{\mathrm{mpp}}V$ is the difference between the power of the I/V minus the power triangle from short-circuit up to the maximum power point. Note that the qmp- point can also be regarded as the maximum power point of the I/V after the current is shifted by -I_{mpp}.

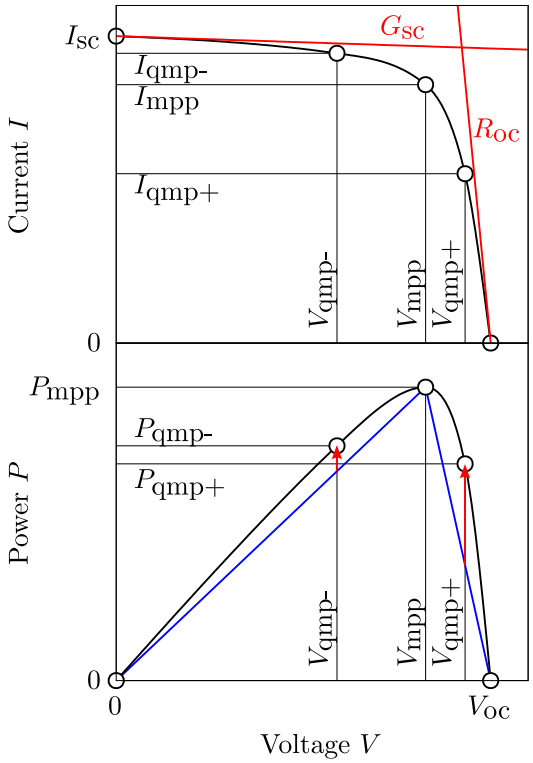

FIG. 1: Extended Solar Cell Parameters

Similarly, for the qmp+ we find

$$\left.\frac{\partial I}{\partial V}\right|_{V=V_{\mathrm{qmp+}}} = -\frac{I_{\mathrm{qmp+}}}{V_{\mathrm{qmp+}} - V_{\mathrm{mpp}}}. \tag{6}$$

The qmp+ point can also be regarded as the mpp of the I/V after the voltage is shifted by -V_{mpp}.

Note that the definition of the qmp's is somewhat recursive. Thus, the same methods which are used to determine the mpp can be used for the qmp's with only minor adaptions (i.e. shifting the I/V). Especially for more elaborate, noise-robust methods to determine the mpp and qmp's, this allows to an optimal reuse of implemented methods. Furthermore, this recursive nature of the qmp's makes that it is generally easy to adapt software to provide the extra points of the ESP in addition to the SSP.

In Appendix A we discuss several algorithms for the noise robust extraction of the ESP including the mpp and qmp's.

III. CURVE MODELS

In this section we present several models to describe an I/V and their parametrization. Before we start with the actual curve models, let us list several properties that we expect to hold for most I/V.

1. In general we expect I/V to be continuous and differentiable. The physics based one diode model is in fact infinitely differentiable [9].

2. Most I/V are monotonously decreasing (with the I_{sc} being positive). In our experience, most non-monotonous I/V are the result of transients, e.g. varying irradiation, capacitive effects, or (metastable) changes in the solar cell. It is also possible for passive semiconductor devices to exhibit non-monotonous behavior under steady state conditions, for example Gunn and tunnel diodes exhibit negative differential resistances. However, we do not expect such effects to play a role in practical photovoltaic devices under normal operating conditions.

3. Normally I/V are concave. There are, however, many cases where an I/V may not be concave everywhere. Most notably due to S-shapes and partial shading of series connected cells (when bypass diodes are active of single cells are driven into junction breakdown).

In this work we will consider various models. However, in light of the expected properties of an I/V we will constrain the models and their parametrization. In particular, all considered curve models are continuous and differentiable. Furthermore, *if* the key points of the ESP are monotonously decreasing, *then* we require the modeled curve to also produce a monotonously decreasing curve (either the model does this by itself or we perform a constrained optimization such that we only consider monotonous solutions).

A. One diode model

The 1D model is one of the most widely used models. It is based on the Shockley diode equation which is rooted in device physics [10]. The one diode model is given by

$$I(V) = I_{\mathrm{ph}} - I_0 \left[\exp\left(\frac{V + I(V)R_{\mathrm{s}}}{a}\right) - 1\right] - \frac{V + I(V)R_{\mathrm{s}}}{R_{\mathrm{sh}}}, \tag{7}$$

where I_{ph} is the photo-current, I_0 the saturation current, a is constant, R_{s} the series resistance, and R_{sh} the shunt resistance. The constant a has the unit volt and can be written as $a = N_{\mathrm{s}} n k T/q$, where N_{s} is the number of series connected cells in a module, n is the ideality factor, k is Bolzmann's constant, T the temperature, and q the elementary charge. As one can see, the equation is transcendental and implicit with the current appearing on both sides of the equation. Thus, to solve the equation, iterative root finding algorithms are typically used, e.g. Newton's or Halley's method. It is possible to write the equation in an explicit form using the Lambert-W function [11]. In terms of solving the equation this does not give a very big advantage as the numerical evaluation of the Lambert-W function itself, relies on the same iterative root finding algorithms [12]. However, you may be able to avoid implementing your own iterative solver if someone implemented the Lambert-W function for you.

As discussed before, the parametrization of the diode model is the topic of a vast number of papers. An impression of the multitude of methods can be obtained from various review papers [1–5]. In this work, we use a hybrid approach utilizing both local and global optimization methods. For the local optimizer we use Levenberg-Marquard and as a global optimizer we implemented the Differential Evolution (DE) algorithm [13]. To obtain a good initial guess we find the Oblique-Asymptote method is quite effective [9].

The overall fitting procedure is to first use the Oblique-Asymptote method and subsequently use the local optimizer. Only if the local optimizer does not obtain a good optimum, we resort to the DE algorithm. Following the DE algorithm we again run the local optimizer with the DE optimum as an intitial guess. A more detailed description of the fitting procedure and used methods can be found in Appendix B.

The diode model has 5 parameters, whereas the ESP provide 10. However, we consider it most important the model accurately reproduces the SSP, which consist only of 4 parameters. For this reason we opt to use the SSP in combination with the G_{sc} value. This choice is somewhat arbitrary but does provide a practical advantage as the Oblique-Asymptote method used for the initial parametrization uses the G_{sc}. We would also note that in case of an I/V with S-Shape, the choice for G_{sc} rather than an ESP beyond the maximum power point, leads diode parametrizations with a high series resistance to account for the low V_{mpp} value w.r.t. V_{oc}. This does not make the 1D model reproducing S-shapes but at least the curve may be somewhat reasonably reproduced up to the mpp (see also Section V A and Fig. 2 therein).

B. Karmalkar-Haneefa

A phenomenological equation for describing I/V was introduced by Karmalkar and Haneefa [7]

$$i(v) = 1 - (1 - \gamma)v - \gamma v^m, \tag{8}$$

where γ and m are the "shape parameters", and

$$i = I/I_{\mathrm{sc}} \tag{9}$$

$$v = V/V_{\mathrm{oc}}. \tag{10}$$

As V_{oc} and I_{sc} are direct parameters the only remaining unknowns are γ and m. The KH model is commonly parameterized using the slopes at short- and open-circuit. This method is indeed convenient as simple analytical expressions can be derived to solve γ and m from these slopes. However, using these slopes will not make the model most accurate in the maximum power point. We

prefer to fit the equation to the SSP, such that the mpp is accurate. To this end we need to solve

$$I_{\mathrm{KH}}(V_{\mathrm{mpp}}) = I_{\mathrm{mpp}} \tag{11}$$

$$\left.\frac{\partial I_{\mathrm{KH}}}{\partial V}\right|_{V_{\mathrm{mpp}}} = -\frac{I_{\mathrm{mpp}}}{V_{\mathrm{mpp}}} \tag{12}$$

Both approximate and accurate expressions exist, see for example [14]. Compared to parameterizing the diode equation, the parametrization of the KH model is generally fast and simple. However, in some cases the solution is non-monotonous.

In order to establish whether a KH parametrization is monotonous, we determine the derivative of Eq. (8)

$$\frac{\partial i(v)}{\partial v} = \gamma(1 - m v^{m-1}) - 1. \tag{13}$$

For a finite derivative at $v = 0$, we must have $m \geq 1$. Furthermore, for a negative or zero derivative at $v = 0$, we must have $\gamma \leq 1$. For a monotonous solution, Eq. (13) must be negative or 0 on the interval $0 \leq v \leq 1$. Thus, we solve $\partial i(v)/\partial v = 0$

$$\frac{\gamma - 1}{\gamma} = m v^{m-1}. \tag{14}$$

Thus, a monotonous solution with finite derivatives ($m \geq 1$), is obtained when $\gamma < 0$ and $1 \leq m \leq (\gamma - 1)/\gamma$, or, for $0 \leq \gamma \leq 1$ and $m \geq 1$. A look at the second derivative

$$\frac{\partial^2 i(v)}{\partial v^2} = -\gamma(m-1) m v^{m-2}, \tag{15}$$

reveals that we obtain concave solutions for $\gamma \geq 0$ and $m \geq 1$, $\gamma \geq 0$ and $m \leq 0$, and, $\gamma \leq 0$ and $0 \leq m \leq 1$. Thus, for normal I/V-characteristics, which are monotonic decreasing and concave, it follows that $0 \leq \gamma \leq 1$ and $m \geq 1$.

We require that for monotonous decreasing SSP, the curve must also be monotonous decreasing. In cases where this requirement is not met we thus need to perform a constrained optimization. In our implementation Eqs. (11) and (12) are solved initially using Levenberg-Marquard. In case the solution is not monotonous decreasing and the SSP are, we perform a constrained optimization using DE.

C. PV-Spline

The diode and the KH models are functions we can fit to a set of parameters. However, as all solar cell parameters describe points on a curve we can also use interpolation methods to approximate the curve. The advantage of interpolation is that it is generally easy to parameterize an interpolation function. Furthermore, an interpolation will always reproduce the points on the curve exactly, whereas the fitting models only approximate a set of parameters.

Probably the most obvious interpolation model would be to use a cubic spline. A cubic spline is simply a piece wise cubic polynomial. As with the ESP we have a set of 5 points on the curve with derivatives, we can describe the whole curve using 4 sections between the 5 points. In the terminology of splines the 5 points on the I/V (i.e., sc, qmp- mpp, qmp+, oc) are referred to as knots. The resulting spline is smooth and continuous up to the first derivative.

A downside of a simple cubic spline, however, is that also for a monotonous I/V, the resulting spline may be non-monotonous. Thus, we need a monotone spline interpolation method. Commonly, monotonous splines are obtained by modifying the derivatives at the knots as described by Fritsch and Carlson [15]. However, in our case the knot derivatives are defined by the ESP, and thus we do not want to modify them. However, alternatively one can insert additional knots to provide the needed degrees of freedom without changing the derivatives at the key points of the I/V. There are, however, many ways to insert additional knots to obtain a monotonous curve. Thus, we developed a heuristic method to insert knots such that the curve is guaranteed to be monotonous and remains smooth (e.g. avoid sharp corners at the inserted knots). We refer to the resulting ESP spline interpolation method as a PV-Spline. A full description of PV-Splines, including the monotone knot insertion heuristic used here, can be found in Appendix C.

The procedure 2 from the IEC60891 norm provides a method to translate an I/V to different irradiances and temperatures [16]. A nice feature of the PV-Spline model (and I/V interpolation models in general), is that it can be combined with procedure 2, and thus a complete I/V model as a function of temperature and irradiance may be obtained. In Appendix D we detail how one may use the IEC60891 norm to translate a PV-Spline to different irradiances and temperatures. The PV-CRAZE library also provides fitting and translation methods to work with the IEC 60891 norm and PV-Splines.

D. PV-Bézier

Another interpolation method used here are Bézier curves. We previously found cubic Béziers particularly accurate in describing I/V [17]. A cubic Bézier curve is a parametric curve commonly used in computer graphics. Similar to splines, Bézier curves are used to create smooth curves. Our PV-Bézier interpolation scheme uses a cubic Bézier interpolation between the key points of the ESP, i.e. from sc to qmp-, from qmp- to mpp, from mpp to qmp+, and from qmp+ to oc. The total number of parameters for this Bézier curve is 18 (requiring a smooth curve without corners), i.e. much more than the 10 parameters from the ESP. However, we developed a parametrization procedure which parameterizes these 18 parameters using the 10 ESP, whilst ensuring the resulting curve is monotonous, if the ESP are monotonous.

A complete description of the PV-Bézier interpolation is given in Appendix E.

IV. PV-CRAZE

In this section we describe the previously mentioned PV-CRAZE library. The source code for PV-CRAZE is freely available from [18]. The main functionality of PV-CRAZE is the analysis of I/V and extracting ESP. For this we implemented noise-robust methods to extract the key points form an I/V. Furthermore, in this step we do some rudimentary checks on the I/V data and flag data where no reliable ESP could be extracted. These flags may be used as a first step in filtering large datasets. The PV-CRAZE library is explicitly developed with large datasets in mind. For this reason the library is implemented in C, to achieve a fast analysis tool suitable to process millions of I/V. We briefly discuss PV-CRAZE performance in Section V D.

Optionally PV-CRAZE can also fit several I/V curve models, namely:

- Karmalkar-Haneefa
- Pindado-Cubas IV model [19][**?**]
- One diode model
- PV-Spline
- PV-Bézier
- Piece-Wise Linear

In addition the PV-CRAZE library provides routines to compute an I/V from a model fit, translate a PV-Spline to different temperatures and irradiances using procedure 2 from the IEC 60891 norm, and fit procedure 2 parameters given a set of IV at different irradiances and temperatures.

V. RESULTS

A. Selected IVs

As a first step in analyzing the various models we consider three examplary I/V-characteristics, as shown in Fig. 2. The I/V are selected specifically to demonstrate some differences in the curve models. The the I/V-characteristics consist of

- a crystalline silicon solar cell I/V. This I/V is chosen as it represents a fairly close approximation to the standard diode equation.
- a thin-film hydrogenated amorphous silicon (a-Si:H) I/V. This I/V is chosen as the I/V of a-Si:H solar cells deviates somewhat from the standard diode equation.

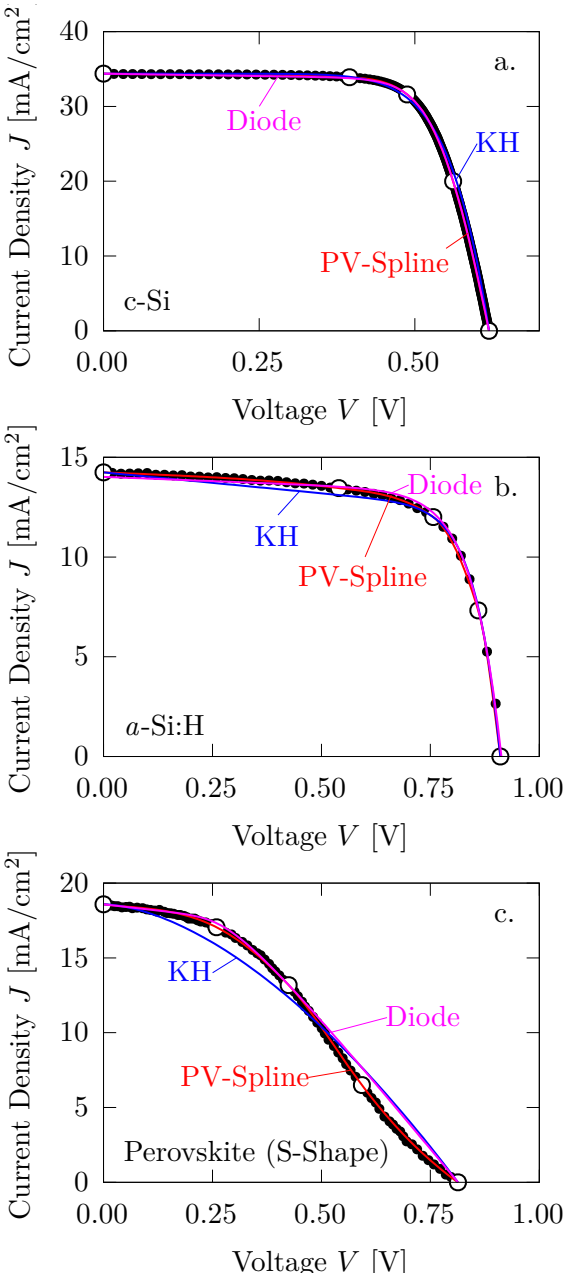

FIG. 2: Examplary ESP and curve model results for three different I/V. The measured I/V is indicated with black dots, the ESP with large white dots, and three lines for the Diode, KH, and PV-Spline models. a. A crystalline silicon wafer based solar cell. b. an a-Si:H solar cell. A perovskite solar cell specifically selected for its severe S-shape (I/V data extracted from Fig. 1b in [20]).

- a perovskite solar cell I/V. This specific perovskite I/V is chosen as it exhibits a rather severe S-Shape. The I/V data shown here was extracted from [20] (Fig. 1b therein).

The results of the curve fits are shown in Fig. 2. The

black dots represent the measured I/V curve under a class A AM1.5g spectrum. The large white dots are the extracted ESP. Furthermore, for each I/V we show three curve fits, the diode model, the KH model and the PV-Spline. In this figure we omit the PV-Bézier model as in this figure its results are indistinguishable from the PV-Spline. As a consequence, in the following discussion of Fig. 2, all statements that apply to the PV-Spline model, also apply to the PV-Bézier model. In Fig. 2a the results are shown for a crystalline silicon wafer solar cell. For this I/V all curve models fit well with the measured curve. Figure 2b shows an a-Si:H solar cell I/V. Here we see that in particular the KH model, but also the diode model clearly deviate from the measured I/V. The PV-Spline still matches perfectly with the measured I/V. In Fig. 2c we show the I/V with S-shape. For this curve both the diode model and the KH model show strong deviations. The KH model maximum power point does not agree with the measured mpp. Note that one can parameterize the KH model to match the mpp exactly, however, this results in a non monotonous curve. The diode model matches well up to the mpp. Note that the selection of ESP to use for fitting the diode equation (see Section III A) was chosen such that the impact of the S-shape is accounted for in a large series resistance value. The PV-Spline reproduces the S-shape.

The results in Fig. 2 exemplify how well the different models can describe various I/V. However, the results are, with only 3 I/V, rather anecdotal. To test the robustness of the various models and the accurate extraction of the ESP, it is important to test things on large datasets. In the following section we compare results for large outdoor datasets.

B. Outdoor data

In this work we will use a subset of data previously published in [21]. For this data set several industrial modules were set up at 5 test locations in different climate zones around the world. The chosen test sites are located in Cologne (Germany), Ancona (Italy), Phoenix (United States of America), and Chennai (India). In this work we also include data from a fifth test site, constructed and operated in Thuwal, Saudi Arabia (SA). [**?**]. Some data on the test sites are summarized in Table I. Each location was equipped with a set of modules of the same types. Periodic measurements were performed of both weather data and current/voltage sweeps of each module. At all test locations the weather station and test equipment were identical. From these data sets we use a subset of 9 module types, one c-Si, one a-Si:H/μc-Si:H, three CdTe, and four CIGS modules. More information on the test sites and methodology can be found in [21].

The complete dataset used in the work consists of 2.24 million I/V. Some of these I/V, however, exhibit problems such as multiple maximum power points, insufficient current or voltage ranges to determine V_{oc} or I_{sc}, etc. Of the 2.24 million, we could obtain reliable ESP for 2.19 million I/V, i.e. 2.2% of the data points were discarded.

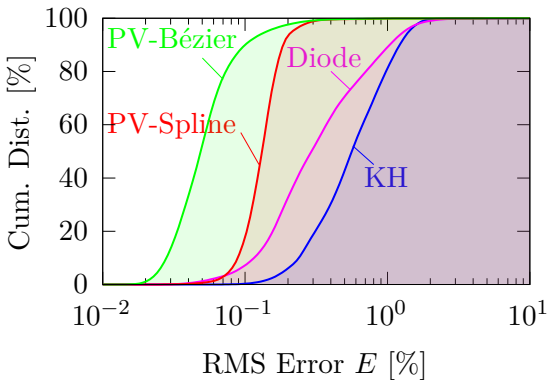

FIG. 3: Cumulative error distributions. For the different curve models, this plot shows the percentage of curves that have an RMS fit-error of less than the error value on the x-axis, e.g. we can see that the fit error of the PV-Bézier model is less than 0.1 % for about 88 % of the I/V's in the database.

In Fig. 3 cumulative fit error distributions are shown for the four I/V models. The curves depict the percentage of curves for which the RMS fit error is less than the fit error in the x-axis for all four models. The fit error of the PV-Bézier model is generally smallest, followed by the PV-Spline model. The PV-Spline model fit errors are distributed in a remarkably narrow range w.r.t. the other models. The Diode and KH models perform worst, where the Diode model outperforms the KH model.

The cumulative plots in Fig. 3 give a good overall indicator for the performance of the different models, however, it does not reveal whether there are sub classes of I/V for which a particular model outperforms another. In Table II we compare how often one model performs better than another for the various curves in the dataset. In the first row we see that the PV-Bézier model is the best model in 93% of the cases, whereas the KH model is the best only in 0.3% of the cases. Likewise, the second row shows the frequency by which each model is the worst model. The remaining rows in Table II each compare the models in the columns with one of the models. For example the last row shows how often the PV-Bézier model is better then the model for each column. From this table we can see that not only is the PV-Béziermodel is the best model in the vast majority of cases, it is also very rarely the worst model.

In a next step we examine whether the fit errors for the different curve models depend on the PV technology. In Fig. 4 we show a box plot for the fit error per model and module technology. In Fig. 4.a the results for the KH model are shown for all 9 module types in the datasets. Overall the model performs best on the CdTe-1 dataset, with a median fit error of 0.4%, and worst on

TABLE I: Test site specifications. All values are in degrees.

	Cologne, Germany	Ancona, Italy	Phoenix, USA	Chennai, India	Thuwal, SA
Latitue	50.922 813	43.474 195	33.424 04	12.984 217	22.312 602
Longitude	6.991 705	13.074 653	−111.910 036	79.987 987	39.110 47
Tilt angle	35.0	35.0	35.5	15.0	25.0
Azimuth	0.0	0.0	0.0	0.0	0.0

TABLE II: Comparison of the various models. The first two rows indicate how often one particular model is best or worst compared to the other 3 models. In the remaining rows we show pairwise comparisons between the models, where each row depicts the rate at which one particular model outperforms the other models.

	KH	Diode	PV-Spline	PV-Bézier
Best	0.3	1.3	5.4	93.0
Worst	87.8	10.7	1.3	0.2
KH <	-	11.3	2.6	0.7
Diode <	88.7	-	15.5	1.9
PV-Spline <	97.4	84.5	-	5.8
PV-Bézier <	99.3	98.1	94.2	-

the CIGS-4 data set with an error of 1%. Although the KH model systematically fits better with some modules than with others, there is no clear trend with technology and the differences are not very big (0.4 % versus 1 %). For the diode model in Fig. 4.b the fit errors are generally smaller than for the KH model. The fit errors show a somewhat similar trend, e.g. like the KH model the smallest and largest fit errors are obtained on the CdTe-1, and the CIGS-4 data sets, respectively. The PV-Spline model in Fig. 4.c poses a significant improvement for all module technologies as compared to both the KH and Diode models, with a fit error typically below 0.2 % for all data sets. Furthermore, the fit errors for all data sets are very similar. Finally, in Fig. 4.d, we show the results for the PV-Bézier model. With the PV-Bézier we obtain fit errors typically well below 0.1 %, outperforming all other tested models.

Inspecting the curves where the interpolation models exhibit a relatively large error, we find that most these curves suffer from specific problems. In particular, we find signs these curves are affected by partial shading, unstable irradiance conditions, or various other issues of unknown origin. In Fig. 5 we show several examples of curves with such problems. Figure 5.a depicts an I/V affacted by partial shading in a crystalline silicon module. In general partial shading can be recognized by kinks and partly convex curves. Figure 5.b is an example where we suspect the irradiance conditions changed during the measurement. Also unstable irradiance conditions may lead to kinks and convex parts in the curve. Finally, an example where the measurement equipment appears to have malfunctioned is shown in Figure Fig. 5.b. Such curves affected by some malfunction are fairly rare and very heterogeneous in appearance. Figure 5 also shows the various curve models for these particular I/V.

It may be desirable for an analysis to specifically remove or select data points such as those shown in Fig. 5. In the following section we will demonstrate the ability of the ESP to capture the shape of an I/V by developing a filter for partial shading in silicon modules as shown in Fig. 5.a.

C. An I/V shape classifier

In this section we develop a random forest classifier to filter the I/V by some shape attributes. For the classifier we define "class A" as a healthy I/V, and "class B" as an I/V with problems (primarily due to partial shading). In this work we limit the scope of this filter to the crystalline silicon modules. In order to train a random forest classifier we created a labeled dataset by manually inspecting I/V-characteristics. Here it should be noted that there exists no ground truth for the labels A and B. The manual labels are based on a somewhat subjective opinion of the author as to whether an I/V is abnormal.

We decided to train the classifier with normalized solar cell parameters, i.e. all currents divided by I_{sc} and all voltages by V_{oc}. This way the classifier is provided only with unit less information on the shape of the I/V, and cannot implicitly learn other attributes in the data such as possible correlations with irradiance level and temperature.

Overall, class B I/V are underrepresented in the dataset, i.e. the majority of I/V are of class A. We try to obtain a somewhat balanced labeled dataset with about equal numbers of class A and B I/V. To this end we used an iterative labeling process where we first labeled only a small dataset to train the classifier. Subsequently, we use the trained model to randomly select I/V for manual labeling such that we obtain a well balanced labeled dataset. The resulting labeled dataset consists of 936 I/V of class A, and 917 of class B. The vast majority of the class B I/V is affected by partial shading.

The final labeled dataset was split into a training and testing set of 1499 and 354 data points, respectively. We trained two random forest classifiers. The first was trained using only the normalized I_{mpp} and V_{mpp} data. This classifier was trained as a reference of the accuracy that may be achieved using the SSP alone. The second classifier was trained with the full set of normalized ESP. The confusion matrices for both random forest classifiers are shown in Table III. The precision in selecting a class A I/V of the SSP classifier is 83 %, and the recall is

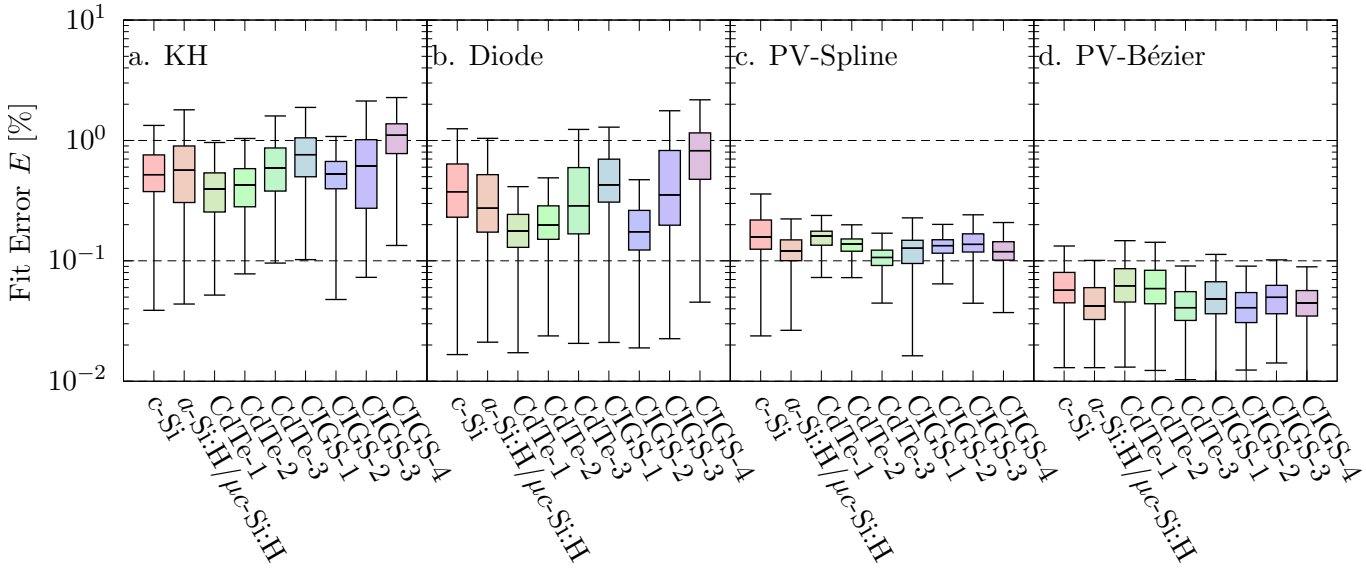

FIG. 4: Fit error box plot for various module technologies and the four models. a. The Karmalka-Haneefa model. b. The one diode model. c. The PV-Spline model. d. The PV-Bézier model

TABLE III: Confusion matrix for the random forest I/V classifiers. (a) Classifier trained using only the normalized V_{mpp} and I_{mpp} values. (b) Classifier trained on the normalized ESPs.

		predicted A	predicted B
label	A	154	40
	B	31	129

(a)

		predicted A	predicted B
label	A	185	8
	B	9	169

(b)

TABLE IV: Confusion matrix to test the accuracy of the manual labeling process. Both the label and the predicted label are the result of manual inspection of the I/V by the author.

		predicted A	predicted B
label	A	160	5
	B	6	146

79 %. The ESP classifier obtains a significantly higher accuracy with a precision of 96% and a recall of 95%.

As mentioned before, the manual labeling is somewhat subjective. Furthermore, during the manual labeling we found is sometimes difficult to decide whether an I/V should be considered class A or B. For this reason we decided to determine baseline precision and recall values by re-labeling some of the I/V in the labeled set. The re-labeling was again performed by the author, but several days after the labeled dataset was generated. For the re-labeling we selected 300 data points from the labeled set. The confusion matrix resulting from the re-labeling is shown in Table IV. The precision in selecting a class A data points is 96 % and the recall is 97 %. Note that the precision and recall of the ESP classifier is nearly identical, i.e., the author is only slightly less confused than the ESP classifier.

Finally, we applied the classifier to all c-Si datasets. In Fig. 6 we show for each location the percentage of curves that are of class B. We find the number of detected class B data points varies strongly for various locations. In particular the c-Si module in India is strongly affected, as almost 72% of the curves is labeled B. Inspection reveals that indeed this dataset exhibits many curves which look partially shaded. Furthermore, from the timestamps for these class B I/V's, we observe longer periods where nearly all curves are labeled B (not shown). Thus, we suspect these partial shading effects originate from soiling.

D. PV-CRAZE Performance

The PV-CRAZE package is developed to aid the analysis of large datasets, and thus performance is very important. In this work the curve analysis with PV-CRAZE consisted of several tasks:

- Curve Analysis (extended solar cell parameter extraction)
- Evaluate the PV-Spline model
- Fit and evaluate the KH model
- Fit and evaluate the Diode model

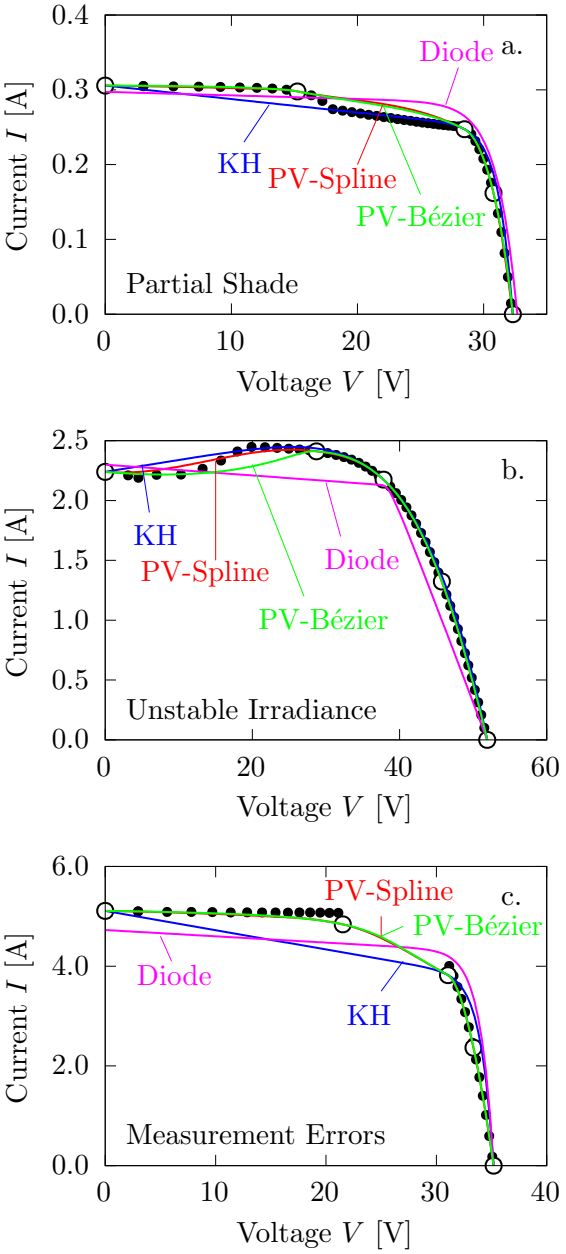

FIG. 5: Several examples of curves labeled as bad. Black dots are the measured I/V, white dots the extracted ESP, and the lines the various curve models. a. Typical example of a partially shaded c-Si module. b. I/V with maximum currents above I_{sc}, presumably the result of unstable irradiation conditions during the measurement. c. Data acquisition problem of unknown origin.

- Evaluate the PV-Bézier model

For the complete dataset we list the amount of time PV-CRAZE spent on the various tasks in Table V. These times were measured for a single thread on a consumer grade laptop (CPU: dual core Intel i5-7200U at 2.50GHz,

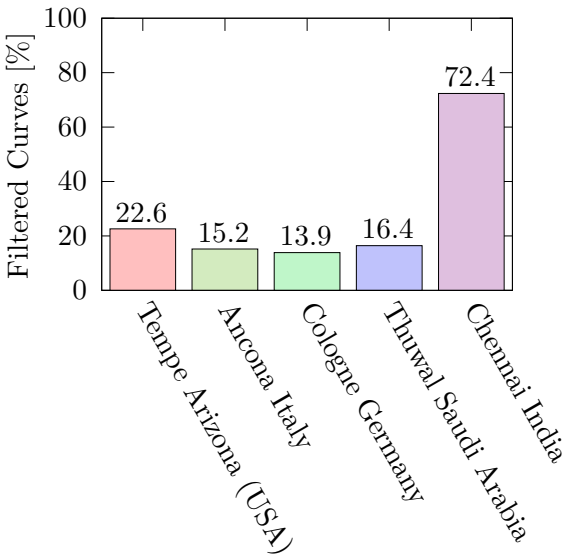

FIG. 6: Bar plot with the percentage of bad curves per location.

TABLE V: PV-CRAZE performance on 2241539 curves on a single core

task	time [s]	time/datum [μs]
Curve Analysis	79.2	35.3
PV-Spline	85.2	38.0
KH	152.7	68.1
Diode	2849.0	1271.0
PV-Bézier	72.3	32.2
Rest	13.5	6.0
total	3251.8	1450.7

and 20 GB DDR4 RAM). The times in Table V were obtained with the octave extension of PV-CRAZE. The total time used was 54:12 (min.:sec.). From Table V we observe almost 87.6 % of the time is spent on the fitting of the diode model, which translates to about 1.3 ms per data point. Most time for fitting the diode model is spent on a relatively small subset of I/V-characteristics (approx 10 %) for which the initial Levenberg-Marquard optimizer does not directly find a good optimum. The KH model is comparatively efficient and uses only 64 μs per data point. The curve analysis, and the evaluation of both interpolation models are fastest and each take between 30-40 μs per data point. If we only extract the ESP, the complete dataset is processed in less than 100 s. It is trivial to parallel process a large dataset by dividing the dataset in chunks, thus, PV-CRAZE provides a fast and salable means to analyze large I/V datasets.

VI. CONCLUSIONS

In this paper we propose a set of Extended Solar Cell Parameters. The ESP provide a general purpose parametrization of I/V-characteristics, allowing a detailed and accurate description of an I/Vcurve. We have developed two methods to reconstruct an I/V-characteristics from the ESP, namely the PV-Bézier and PV-Spline models. By reconstructing the I/V-characteristics we can obtain a measure of information loss by computing the distance between the original and reconstructed I/V. This way we demonstrate on a large dataset of 2.2 million I/V, covering a wide range of PV technologies, that the ESP provide an accurate description of the I/V. For the PV-Bézier model we find that for 90% of the I/V curves we obtain an RMS error below 0.1%. For comparison, with the PV-Spline model this is 0.2%, for the diode model about 1% and for the KH model 1.2%. This comparatively low fit error demonstrates that the ESP capture much more detail than the commonly used I/V parametrizations.

We argue that these general purpose, descriptive I/V parameters are useful for analyzing large datasets. To demonstrate this we developed a random forest classifier to detect partial shading in the crystalline silicon modules in our outdoor dataset. The random forest classifier obtains high precision and recall values (96%, and 95%, respectively). These precision and recall values are more or less identical to the precision and recall values obtained by manually labeling the labeled dataset a second time, i.e. the classifier is about as accurate as the labeled dataset is. This random forest classifier can thus be used to rapidly filter the complete dataset for partial shading conditions.

Finally, we presented our open source analysis tool for I/V-characteristics, PV-CRAZE. The tool implements robust I/V curve analysis methods and was developed with large datasets in mind. PV-CRAZE can extract the ESP, and fit various I/V curve models (among others the PV-Bézier, PV-Spline, One diode, and KH models). We demonstrate PV-CRAZE can extract the ESP for a large dataset of 2.2 million I/V in under 100 s, on a single thread using a consumer grade laptop.

VII. ACKNOWLEDGMENTS

This work was partly funded by the Living Lab Energy Campus (LLEC) project, and PV Klima, BMWK (Federal Ministry for Economic Affairs and Climate Action), under contract 0325517C. Furthermore, I am grateful to Andreas Gerber, Timon Vaas, and Evgenii Sovetkin for discussions and comments.

[1] C. Hansen, Tech. Rep., Sandia National Lab.(SNL-NM), Albuquerque, NM (United States) (2015).
[2] A. Ortiz-Conde, F. J. García-Sánchez, J. Muci, and A. Sucre-González, Facta universitatis-series: Electronics and Energetics **27**, 57 (2014).
[3] D. Cotfas, P. Cotfas, and S. Kaplanis, Renewable and Sustainable Energy Reviews **28**, 588 (2013).
[4] A. R. Jordehi, Renewable and Sustainable Energy Reviews **61**, 354 (2016), ISSN 1364-0321, URL https://www.sciencedirect.com/science/article/pii/S1364032116300016.
[5] R. Venkateswari and N. Rajasekar, International Transactions on Electrical Energy Systems **31**, e13113 (2021), https://onlinelibrary.wiley.com/doi/pdf/10.1002/2050-7038.13113, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/2050-7038.13113.
[6] X. Ma, W.-H. Huang, E. Schnabel, M. Köhl, J. Brynjarsdóttir, J. L. Braid, and R. H. French, IEEE Journal of Photovoltaics **9**, 1405 (2019).
[7] S. Karmalkar and S. Haneefa, IEEE electron device letters **29**, 449 (2008).
[8] C. Ulbrich, D. C. Jordan, S. R. Kurtz, A. Gerber, and U. Rau, Solar energy **113**, 88 (2015).
[9] F. J. Toledo and J. M. Blanes, Renewable Energy **72**, 125 (2014).
[10] W. Shockley, Bell System Technical Journal **28**, 435 (1949).
[11] T. Banwell and A. Jayakumar, Electronics Letters **36**, 291 (2000), ISSN 0013-5194.
[12] R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey, and D. E. Knuth, Advances in Computational mathematics **5**, 329 (1996).
[13] R. Storn and K. Price, Journal of Global Optimization **11**, 341 (1997).
[14] S. Pindado, J. Cubas, E. Roibás-Millán, F. Bugallo-Siegel, and F. Sorribes-Palmer, Energies **11**, 1353 (2018).
[15] F. N. Fritsch and R. E. Carlson, SIAM Journal on Numerical Analysis **17**, 238 (1980).
[16] ISO Central Secretary, Standard IEC 60891:2021, International Organization for Standardization, Geneva, CH (2021), URL https://webstore.iec.ch/publication/61766.
[17] T. Vaas and B. Pieters, in *2021 IEEE 48th Photovoltaic Specialists Conference (PVSC)* (IEEE, 2021), pp. 1560–1564.
[18] B. E. Pieters, *PV-Craze (placeholder citation, will be online by the time of publication)*, https://github.com/IEK-5/PV-CRAZE (2022).
[19] S. Pindado and J. Cubas, Renewable Energy **103**, 729 (2017), ISSN 0960-1481, URL https://www.sciencedirect.com/science/article/pii/S0960148116309697.
[20] R. A. Z. Razera, D. A. Jacobs, F. Fu, P. Fiala, M. Dussouillez, F. Sahli, T. C. J. Yang, L. Ding, A. Walter, A. F. Feil, et al., J. Mater. Chem. A **8**, 242 (2020), URL http://dx.doi.org/10.1039/C9TA12032G.
[21] M. Schweiger, J. Bonilla, W. Herrmann, A. Gerber, and U. Rau, Progress in Photovoltaics: Research and Applications **25**, 968 (2017), URL https://onlinelibrary.wiley.com/doi/abs/10.1002/pip.2904.

[22] A. Savitzky and M. J. Golay, Analytical chemistry **36**, 1627 (1964).
[23] C. Runge, *Zeitschrift für Mathematik und Physik* (B. G. Teubner., 1901), pp. 224–243, 46.
[24] M. Galassi and B. Gough, *GNU Scientific Library: Reference Manual*, GNU manual (Network Theory, 2009), ISBN 9780954612078, URL http://www.gnu.org/software/gsl/.
[25] ISO Central Secretary, Standard IEC 60891:2009, International Organization for Standardization, Geneva, CH (2009), URL https://webstore.iec.ch/publication/3821.
[26] P. de Faget de Casteljau, in *Mathematical Methods in Computer Aided Geometric Design II*, edited by T. Lyche and L. L. Schumaker (Academic Press, 1992), pp. 57–68, ISBN 978-0-12-460510-7, URL https://www.sciencedirect.com/science/article/pii/B9780124605107500100.
[27] This model provided no advantage over KH and uses the same number of parameters, we therefore omit the results for this model.
[28] Due to a delayed start of measurements at the Thuwal test site, its results were not included in reference [21]

Appendix A: Noise Robust Extraction of ESP

In this section we discuss the methods implemented in PV-CRAZE to extract the ESP. To extract the ESP from an I/V we first determine the mpp, then the two qmp's and finally the sc and oc points.

1. Determine the (q)mpp's

The simplest method to obtain a mpp estimate is simply multiplying all voltage and current pairs and select current-voltage pair with the highest power. This estimate is, however, not always accurate, especially if there is noise or the I/V is coarsely sampled. We refer to this mpp estimate with $\widetilde{\text{mpp}}$. To determine a more accurate mpp estimate we generally fit a polynomial to the data and compute the maximum from the polynomial. For the fitting of a polynomial we apply Ordinary Least Squares (OLS). For accurate results we only fit a certain range of the curve. The range is specified with a start voltage, V_{start}, and an end current, I_{end}. The range is selected with respect to the coarse mpp estimate $\widetilde{\text{mpp}}$

$$V_{\text{start}} = (1 - f_{\text{mpp},V}) V_{\widetilde{\text{mpp}}} \tag{A1}$$
$$I_{\text{start}} = (1 - f_{\text{mpp},I}) I_{\widetilde{\text{mpp}}}, \tag{A2}$$

where $f_{\text{mpp},V}$, and $F_{\text{mpp},I}$ are both constants between 0 and 1.

For the polynomial fitting we have several approaches implemented in PV-CRAZE:

- Fit a 3rd order polynomial to the power curve ($P(V)$) *and* to the current curve ($I(V)$). Determine the V_{mpp} from the maximum of power fit and substitute the found V_{mpp} in the current fit to obtain I_{mpp}
- Same as above but now considering both the voltage and the current as functions of "time", where we assume voltages and currents are sampled at equidistant time. In this case we fit three 3rd order polynomials to the power curve ($P(t)$), the current curve ($I(t)$), and the voltage curve ($V(t)$). From the fit to the power, we obtain a time for the mpp and we substitute this time value in the voltage and current fits, thus obtaining V_{mpp} and I_{mpp}.
- A faster algorithm may be obtained using Savitzky-Golay filters [22]. Savitzky-Golay filters provide a method to obtain a polynomial fit, with the requirement that the data is sampled equidistantly. We cannot generally assume this to be the case for the curve as a function of voltage. However, for the parametric approach we do assume equidistant sampling rates in time. In this work we use a symmetric 3rd order Savitzky-Golay filter with a kernel length of 7. Here we still try to use the complete I/V range between V_{start}, and I_{end}. To this end we place 3 points before $\widetilde{\text{mpp}}$ at approximately equidistant voltages, and 3 points after $\widetilde{\text{mpp}}$ at approximately equidistant currents, i.e. this method always considers only 7 points within the selected range.

On the data we worked with we find that the Savitzky-Golay method is most robust. In addition the Savitzky-Golay method is always faster than the explicit application of OLS, especially for I/V with many samples around the mpp.

For the quasi-mpp's we simply use the same algorithms as above where we merely shift the I/V down by I_{mpp}, to obtain the qmp- point, or left by V_{mpp} to obtain the qmp+ point.

2. Extrapolation to short- and open-circuit

Laboratory cell I/V are typically measured with 4 quadrant measurement equipment, which allows the measurement of the I/V from voltages below short circuit and above open circuit. However, PV modules are often measured with cheaper programmable loads. With such equipment one can measure only the active range of the solar cell and thus the short- and open-circuit points need to be extrapolated.

In general I/V tend to be fairly flat near the short- and open-circuit points. However this is not always the case. Furthermore, we again need to make sure the extraction procedure is noise robust. With a plain linear extrapolation one quickly ends up with a a problem that one needs to define a "linear" range of the measured I/V, where the linear function may be fitted to. If this linear range is

chosen too large, the slope of the linear fit will not accurately represent the slope one is after, and if the range is too small noise may affect the fit. To mitigate this effect we can again, like for the maximum power point, use can us a higher order polynomial fit, which can adapt to the curvature of the data whilst rejecting noise. However, a polynomial extrapolation can easily become ill conditioned, see for example Runge's phenomenon [23].

In our approach we use a 3rd polynomial OLS fit to obtain an approximation of the curve which is robust to noise. In the second step we, however, avoid using the polynomial directly for the extrapolation. Instead we extrapolate with a line through two points on the polynomial approximation. The procedure is exemplified using the extrapolation of the short-circuit current.

We start with a 3rd order polynomial OLS fit to $I(V)$, where we select points in the curve with voltages up to $V_{\text{qmp-}}$. Then we consider two poinst in the fit. The first point is the minimum voltage larger than 0, present in the data, V_1. The second point is defined as

$$V_2 = V_1 + f_{\text{sc},V} V_{\text{qmp-}}, \tag{A3}$$

where $f_{\text{sc},V}$ is a constant between 0 and 1. We use the 3rd order polynomial fit to determine the corresponding currents I_1 and I_2. Subsequently, we parameterize a straight line through these two points and determine the intersection with the current axis. Thus, we assume we have a linear range between V_1, and V_2. If the value of $f_{\text{sc},V}$ is chosen too small, the extrapolation may become ill conditioned (Runge's phenomenon [23]). If, on the other hand, $f_{\text{sc},V}$ is chosen too large, the range between V_1, and V_2 is not well approximated with a straight line. Per default PV-CRAZE uses $f_{\text{sc},V} = 0.12$.

For the open circuit we use an analogous approach. However, for the open circuit point we swap the voltages and currents, i.e. we select data with currents up to $I_{\text{qpm+}}$, fit a polynomial to $V(I)$, and make a linear extrapolation through two points at I_1 (minimum current larger than 0 in the data) and $I_2 = I_1 + f_{\text{oc},I} I_{\text{qmp+}}$. The default value in PV-CRAZE is again $f_{\text{sc},I} = 0.12$.

Appendix B: Diode Model Parametrization

Here we give a short description of the used parametrization methods for the one diode model. However, we start with the disclaimer that, in the face of the vast amount of literature of which we have barely scratched the surface, we do not claim a particular expertise on fitting the diode equation. The here presented method has not been bench-marked against any other procedure, and thus, we make no claims about how this procedure performs w.r.t. other methods. Our only claim is that "it works for us" on large sets of I/V. We feel the latter is an important aspect as in our experience it is easy to develop a method that works well on any small set of I/V. However, it is very hard to find a robust method that works well on virtually all I/V.

In general, local optimizers, such as non-linear least squares and Levenberg-Marquard, frequently get "stuck" in local minima, making a good initial guess imperative. In literature many methods exist for this purpose. We found the oblique asymptote method to be of great value [9]. In the following we first describe the oblique asymptote method.

1. Initial Guess: The Oblique-Asymptote Method

For the Oblique Asymptote method the diode equation is rewritten as

$$I = A - B(C^V D^I - 1) - EV \tag{B1}$$

where

$$A := \frac{I_{\text{ph}} R_{\text{sh}}}{R_{\text{sh}} + R_{\text{s}}} \tag{B2}$$

$$B := \frac{I_0 R_{\text{sh}}}{R_{\text{sh}} + R_{\text{s}}} \tag{B3}$$

$$C := \exp\left(\frac{1}{a}\right) \tag{B4}$$

$$D := \exp\left(\frac{R_{\text{s}}}{a}\right) \tag{B5}$$

$$E := \frac{1}{R_{\text{sh}} + R_{\text{s}}}. \tag{B6}$$

The line

$$I' = A + B - EV \tag{B7}$$

is an oblique asymptote of Eq. (B1) in the limit for $V \to -\infty$. If we now approximate this asymptote with $I' = I_{\text{sc}} - VG_{sc}$, we can reduce Eq. (B1) to

$$I - I' = BC^V D^I. \tag{B8}$$

Thus it follows

$$\ln(I - I') = \ln(B) + V\ln(C) + I\ln(D). \tag{B9}$$

It is trivial to solve $\ln(B)$, $\ln(C)$, and $\ln(D)$ using three points on the I/V, after which we can substitute A and B in Eq. (B7) to obtain E, and we have a full parametrization of the diode equation.

For the three points on the I/V, we cannot use the short circuit point as there $I - I' = 0$ and the logarithm is not finite. However, as we always determine the ESP, we have the mpp, the oc and two qmp's at our disposal to obtain an initial parametrization. In some cases $I - I' \leq 0$ also for other points on the curve apart form short circuit. To deal with this situation we modify the assumption for the oblique asymptote to $I' = I'_{\text{sc}} - VG_{sc}$, with $I'_{\text{sc}} > I_{\text{sc}}$.

2. Diode Parameter Optimization

After using the oblique-asymptote method for the initial guess we found Levenberg-Marquard with geodesic

acceleration, as implemented in the GNU Scientific Library [24] to be quite effective. In most cases this leads to satisfactory results. However, in a small fraction, the resulting fits are still not accurate. In these cases we resort to Differential-Evolution (DE) [13].

The effectiveness of the DE optimizer depends strongly on the initial population. In the diode equation, the various parameters are strongly correlated. This needs to be considered when initializing agents for DE. When choosing parameters randomly without considering the correlated nature of the diode parameters, many agents are not even close to the actual curve. For this reason we use the oblique asymptote method with random perturbations. This way one can effectively limit the used parameter space to a region with reasonable solutions.

Our final diode model parametrization method is as follows. We start with the initial guess using the oblique asymptote method. Subsequently, we use Levenberg-Marquard with geodesic acceleration. In case the fit error is below a user defined threshold we are finished. Otherwise, we switch to the DE algorithm to obtain a better fit. This final fit from DE is used as the initial guess in one final Levenberg-Marquard optimization.

Appendix C: A derivative preserving monotone spline interpolation

Fritsch and Carlson developed a commonly used monotone cubic spline interpolation method [15]. However, in this method the derivatives at the given knots are modified to ensure the curve is monotonous. In our case we do not want this as the derivatives at the given knots should be fixed. For this reason, we resort to inserting additional knots, for which we choose both the coordinate and slope in such a way that the resulting curve is monotonous. Before we go into the knot insertion algorithm, we first discuss under which conditions a cubic spline is monotonous. Fritsch and Carlson formulated several criteria for a cubic spline to be monotonous [15].

$$m_k = \frac{I_{k+1} - I_k}{V_{k+1} - V_k} \tag{C1}$$

$$\alpha_k = \frac{\partial I_k / \partial V_k}{m_k} \tag{C2}$$

$$\beta_k = \frac{\partial I_{k+1} / \partial V_{k+1}}{m_k}, \tag{C3}$$

where k refers to the knot index, i.e. index of the key point on the I/V. We define k such that k increases with voltage and $k = 0$ for short circuit and $k = 4$ for open circuit. The first criterion is

$$\operatorname{sgn}(\partial I_k / \partial V_k) = \operatorname{sgn}(\partial I_{k+1} / \partial V_{k+1}) = \operatorname{sgn}(I_{k+1} - I_k), \tag{C4}$$

which must always be fulfilled for a monotonous curve. In case we in addition have

$$\alpha_k + \beta_k - 2 \leq 0, \tag{C5}$$

the curve is monotonous. For the case that Eq. (C5) is not met, one of the following criteria must be met

$$\alpha_k - \frac{(2\alpha_k + \beta_k - 3)^2}{3(\alpha_k + \beta_k - 2)} \geq 0 \tag{C6}$$

$$\alpha_k + 2\beta_k - 3 \leq 0 \tag{C7}$$

$$2\alpha_k + \beta_k - 3 \leq 0. \tag{C8}$$

In the following we describe our knot insertion heuristics. As a first step we test whether a section is monotonous. If this is not the case, we will insert one or two knots. We define $\partial I'_k$

$$\partial I'_k = 3m_k - \frac{\partial I_k + \partial I_{k+1}}{2} \tag{C9}$$

We now consider two cases. In case both the following criteria are true

$$\partial I'_k > 0 \tag{C10}$$

$$(\partial I_k - m_k)(\partial I_{k+1} - m_k) > 0, \tag{C11}$$

we insert two additional knots at

$$I_{k+1/3} = \frac{\partial I_{k+1} I_k + \partial I_k I_{k+1}}{\partial I_k + \partial I_{k+1}} \tag{C12}$$

$$V_{k+1/3} = V_k + 3\frac{I_{k+1/3} - I_k}{\partial I_k} \tag{C13}$$

$$\partial I_{k+1/3} = 0, \tag{C14}$$

and

$$I_{k+2/3} = \frac{\partial I_{k+1} I_k + \partial I_k I_{k+1}}{\partial I_k + \partial I_{k+1}} \tag{C15}$$

$$V_{k+2/3} = V_{k+1} + 3\frac{I_{k+2/3} - I_{k+1}}{\partial I_{k+1}}, \tag{C16}$$

$$\partial I_{k+2/3} = 0. \tag{C17}$$

Otherwise we insert one knot. If $(\partial I_k - m_k)(\partial I_{k+1} - m_k) < 0$, we insert a knot at

$$I_{k+1/2} = \frac{2[(m+\partial I_{k+1})I_k - (m+\partial I_k)I_{k+1}] + (m+\partial I_k)(m+\partial I_{k+1})(V_{k+1}-V_k)}{2(\partial I_{k+1}-\partial I_k)} \tag{C18}$$

$$V_{k+1/2} = \frac{2(I_k - I_{k+1}) + (m+\partial I_{k+1})V_{k+1} - (m+\partial I_k)V_k}{\partial I_{k+1}-\partial I_k}, \tag{C19}$$

$$\partial I_{k+1/2} = m. \tag{C20}$$

For $(\partial I_k - m_k)(\partial I_{k+1} - m_k) >= 0$, we insert a knot at

$$I_{k+1/2} = I_k + \frac{(6m+\partial I_k - \partial I_{k+1})(V_{k+1}-V_k)}{12} \tag{C21}$$

$$V_{k+1/2} = (V_k + V_{k+1})/2, \tag{C22}$$

$$\partial I_{k+1/2} = \partial I_k. \tag{C23}$$

Appendix D: Using the IEC 60891 with PV-Splines

We can use the IEC 60891 in combination with PV-Splines. To this end we need to translate the knots and the derivatives. Various versions of the IEC 60891 exist. The last two versions are version 2 and 3, which can be found in [25] and [16], respectively. For simplicity we use version 2 of the IEC 60891, as the translation equations in version 3 are more complicated. However, a similar derivation can be made for version 3 of the norm, and PV-CRAZE implements PV-Spline translation routines for both version 2 and version 3 of the IEC 60891.

The translation of knot coordinates is performed using the two IEC 60891 v2 translation equations [25]

$$I_2 = I_1\left(1+\alpha\left(T_2-T_1\right)\right)G_2/G_1 \tag{D1}$$

$$V_2 = V_1 + V_{oc1}\left(\beta\left(T_2-T_1\right) + a\ln\left(G_2/G_1\right)\right) - R'_s\left(I_2-I_1\right) - \kappa I_2\left(T_2-T_1\right), \tag{D2}$$

where a subscript 1 refers to values before- and 2 to after translation, e.g. (V_1, I_1) is a knot coordinate before translation. The IEC 60891 v2 norm has 5 coefficients describing the irradiance (G) and temperature dependencies of the voltages and currents, namely α, β, a, R'_s, κ. From Eqs. (D1) and (D2) we can also derive how the derivatives change. The derivatives of Eqs. (D1) and (D2) to I_1 are

$$\frac{\partial I_2}{\partial I_1} = \left(1+\alpha\left(T_2-T_1\right)\right)G_2/G_1 \tag{D3}$$

$$\frac{\partial V_2}{\partial I_1} = \frac{\partial V_1}{\partial I_1} - R'_s\frac{\partial I_2}{\partial I_1} + R'_s - \kappa\frac{\partial I_2}{\partial I_1}\left(T_2-T_1\right). \tag{D4}$$

By dividing Eq. (D4) by Eq. (D3) we obtain

$$\frac{\partial V_2}{\partial I_2} = \frac{G_1}{G_2}\frac{\frac{\partial V_1}{\partial I_1} + R'_s}{\left(1+\alpha\left(T_2-T_1\right)\right)} - R'_s - \kappa\left(T_2-T_1\right). \tag{D5}$$

Equation (D5) allows us to translate the derivatives.

By translating all knots and derivatives we obtain a new spline. Note that the translated points do not correspond exactly to the ESP of the translated curve. However, we can then simply determine the ESP from the newly computed spline. Note that one may need to extrapolate the new V_{oc} or I_{sc} values, for which we use a linear extrapolation from the translated oc and sc points, respectively.

Appendix E: PV-Bézier interpolation

A Bézier curve is defined by a set of control points. The number of control points depend on the order of the Bézier curve. In this work we will focus on cubic Bézier curves, for which we need 4 points (coordinates), $\boldsymbol{P_0}$-$\boldsymbol{P_3}$. The first and last points are the beginning and end of the curve. The other two control points generally do not lie on the curve. A cubic Bézier curve can be written as:

$$\boldsymbol{B}(t) = (1-t)^3\boldsymbol{P_0} + 3(1-t)^2t\boldsymbol{P_1} + 3(1-t)t^2\boldsymbol{P_2} + t^3\boldsymbol{P_3}, \tag{E1}$$

where the time t ranges from 0 to 1, $\boldsymbol{B}(t)$ is the coordinate at time t. At $t=0$ we have $\boldsymbol{B}(t)|_{t=0} = \boldsymbol{P_0}$ and at $t=1$ we have $\boldsymbol{B}(t)|_{t=1} = \boldsymbol{P_3}$. The curve can be understood as a trajectory, where at each point the curve has a derivative and a velocity w.r.t. the time variable. The derivative of the Bézier curve with respect to time is given by:

$$\begin{aligned}\boldsymbol{B'}(t) =& 3(1-t)^2(\boldsymbol{P_1}-\boldsymbol{P_0}) + 6(1-t)t(\boldsymbol{P_2}-\boldsymbol{P_1}) \\ &+ 3t^2(\boldsymbol{P_3}-\boldsymbol{P_2}).\end{aligned} \tag{E2}$$

From Eq. (E2) we see that in the point $\boldsymbol{P_0}$, the vector $\boldsymbol{P_0}$-$\boldsymbol{P_1}$ is a tangent vector. Furthermore, the velocity (derivative w.r.t. time) is proportional to the length of

this vector ($|\boldsymbol{P_0}$-$\boldsymbol{P_1}|$). Likewise the direction and velocity in the point $\boldsymbol{P_3}$ is given by the vector $\boldsymbol{P_2}$-$\boldsymbol{P_3}$.

To parameterize an I/V we proceed along several steps. In the first step we split the I/V in two sections, one from short circuit to the maximum power point, and one from the maximum power point to open-circuit, see Fig. 7a. We then determine the control points for each section such that the curve approximates the quasi-mpp's as well as possible, where we consider both the qmp coordinate as well as slope. For a cubic Bézier section with given slopes at the beginning and end knots, we have that the two control points lie on the two respective tangential lines at the beginning and end knots. Thus, we write

$$\begin{aligned}\boldsymbol{P_1} &= \boldsymbol{P_0} + a\boldsymbol{P_0'} \\ \boldsymbol{P_2} &= \boldsymbol{P_3} - b\boldsymbol{P_3'},\end{aligned} \tag{E3}$$

where, a and b are constants to be determined, and $\boldsymbol{P_0'}$ and $\boldsymbol{P_3'}$ are the tangent vectors at the beginning and end knot. If we now take any arbitrary point ($\boldsymbol{P_x}$) on the desired Bézier section, *and* a specific time $0 < t_x < 1$, for which the curve passes through this point, we can solve a and b as:

$$a = \frac{x_3'((2t_x^3 - 3t_x^2)(y_3 - y_0) + (y_x - y_0)) - y_3'((2t_x^3 - 3t_x^2)(x_3 - x_0) + (x_x - x_0))}{3D(t_x^3 - 2t_x^2 + t)} \tag{E4}$$

$$b = \frac{x_0'((2t_x^3 - 3t_x^2)(y_0 - y_3) + (y_0 - y_x)) + y_0'((2t_x^3 - 3t_x^2)(x_3 - x_0) + (x_x - x_0))}{3D(t_x^3 - t_x^2)}, \tag{E5}$$

where, x_i and y_i are the coordinates of $\boldsymbol{P_i}$, and x_i' and y_i' the elements of $\boldsymbol{P_i'}$ (i equal to 0 or 3). The factor D equals $x_3'y_0' - x_0'y_3'$, and is thus the magnitude of the cross product of the tangential vectors ($\boldsymbol{P_0'} \times \boldsymbol{P_3'}$). Note that if $D = 0$ we cannot solve a and b as the Bézier section is a straight line (i.e. the tangential vectors have the same direction and thus they do not span the 2D plane).

Using Eqs. (E4) and (E5) we can reduce the optimization problem to find the two control points to a linear search. To this end we substitute the qmp for $\boldsymbol{P_x}$, and search for the time $0 < t_{\text{qmp}} < 1$ for which the resulting Bézier curve best matches the slope in the qmp. It may happen that the solution for a or b is negative. This would lead to a multi-valued curve and thus we do not consider such solutions valid. If the optimization does not produce a valid solution (i.e. all a and b are negative) we set $t_{\text{qmp}} = 0.5$ and take the absolute values of a and b.

In the second step we split the Bézier curve at the approximate quasi-mpp's using de Casteljau subdivision [26], as illustrated in Fig. 7b. The qmp point itself and the adjacent control points are then adapted to match the exact qmp coordinate and derivative.

It may happen that the Bézier curve is not monotonous but the ESP are. However, as our final Bézier representation, as shown in Fig. 7b, has more degrees of freedom than the ESP provide, we do not need to insert additional knots as with the PV-Splines. It suffices to change the control points such that the derivatives in the knots are unaffected and the curve becomes monotonous. A first step is, however, checking whether the Bézier curve is monotonous.

A sufficient condition to guarantee a monotonic cubic Bézier curve in 2 dimensions is if both $x(t)$ and $y(t)$ are monotonic, where x and y are the elements of the vector $\boldsymbol{B(t)}$ in Eq. (E1). Note that we only need to consider the time interval $0 \leq t \leq 1$. In the following we discuss only $x(t)$, however, the same arguments and derivations hold for $y(t)$. For a curve to be non-monotonic, there must be at least one extrema on the the time interval $0 \leq t \leq 1$. As we have a cubic equation we can have at most two extrema. We can consider various cases.

- If for the derivative of x at $t = 0$ and $t = 1$, we can write

$$\left.\frac{\partial x(t)}{\partial t}\right|_{t=0} \cdot \left.\frac{\partial x(t)}{\partial t}\right|_{t=1} < 0, \tag{E6}$$

  we must have exactly one extrema on the time interval $0 \leq t \leq 1$, and the curve is non-monotonic.

- If the condition in Eq. (E6) is not true, any curve which is non-monotonic on $0 \leq t \leq 1$ must have an inflection point on $0 \leq t \leq 1$. We rewrite Eq. (E1) in terms of x

$$x(t) = (1-t)^3x_0 + 3(1-t)^2tx_1 + 3(1-t)t^2x_2 + t^3x_3. \tag{E7}$$

  We can solve the inflection point for Eq. (E7) and find

$$t_{\text{infl.}} = \frac{x_0 - 2x_1 + x_2}{x_0 - 3x_1 + 3x_2 - x_3} \tag{E8}$$

  Having the inflection point we can immediately conclude that if $t_{\text{infl.}} < 0$ or $t_{\text{infl.}} > 1$, the curve must be monotonic on $0 \leq t \leq 1$.

- If $0 \leq t_{\text{infl.}} \leq 1$ the curve *may* still be monotonic. In this case we can consider the derivative in the inflection point. If the following two conditions hold

the curve is monotonic:

$$\left.\frac{\partial x(t)}{\partial t}\right|_{t=0} \cdot \left.\frac{\partial x(t)}{\partial t}\right|_{t=t_{\text{infl.}}} \geq 0 \tag{E9}$$

$$\left.\frac{\partial x(t)}{\partial t}\right|_{t=1} \cdot \left.\frac{\partial x(t)}{\partial t}\right|_{t=t_{\text{infl.}}} \geq 0. \tag{E10}$$

As it is sufficient if both $x(t)$ and $y(t)$ are monotonous, we can consider the x and y coordinates separately. If the x (y) coordinates of $\boldsymbol{P_0}$, $\boldsymbol{P_1}$, $\boldsymbol{P_2}$, and $\boldsymbol{P_3}$, are monotonous, the resulting $x(t)$ ($y(t)$) function is also monotonous. Thus in order to make a section of our curve monotonous, we compute two factors a_x and a_y:

$$a_x = \frac{x_3 - x_0}{x_3 - x_0 + x_1 - x_2} \tag{E11}$$

$$a_y = \frac{y_3 - y_0}{y_3 - y_0 + y_1 - y_2}. \tag{E12}$$

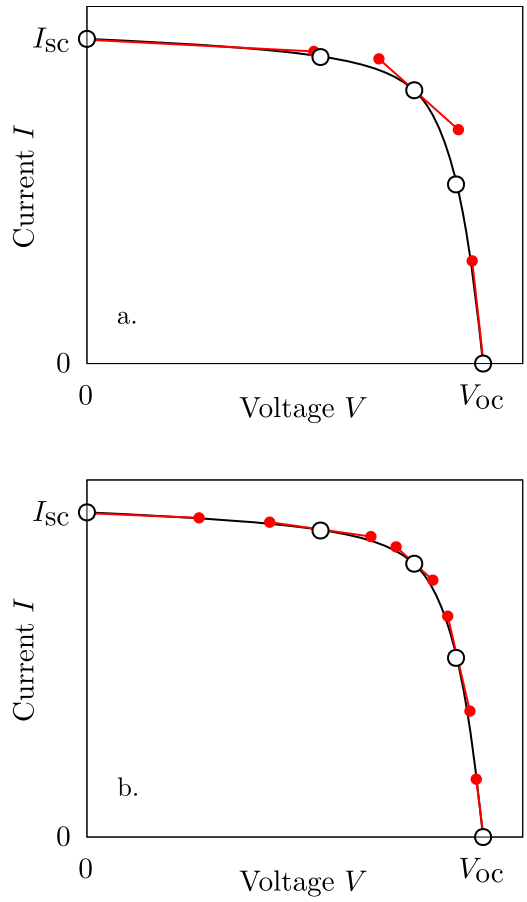

FIG. 7: Illustration of the PV-Bézier parametrization steps. a. In the first step a reduced Bézier curve is constructed with knots at sc, mpp, and oc. The control points are fitted such that the qmp's are best approximated (both in coordinate as well as slope). b. In the second step we split the two Bézier sections in two at the approximate qmp's points.

We take the minimum value of a_x and a_y, i.e. $a = \min(a_x, a_y)$, and then compute new control points $\boldsymbol{P_1'}$ and $\boldsymbol{P_2'}$

$$\boldsymbol{P_1'} = \boldsymbol{P_0} + a(\boldsymbol{P_1} - \boldsymbol{P_0}) \tag{E13}$$

$$\boldsymbol{P_2'} = \boldsymbol{P_3} + a(\boldsymbol{P_2} - \boldsymbol{P_3}). \tag{E14}$$

Note that we only change controll point coordinates, and both the coordinates and derivatives in the main knots (e.g. the ESPs), are unaffected.