Energy & Environmental Science

ANALYSIS

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Energy Environ. Sci.*, 2018, **11**, 730

Critical analysis on the quality of stability studies of perovskite and dye solar cells†

Armi Tiihonen, *[a] Kati Miettunen, [ab] Janne Halme, [a] Sakari Lepikko, [a] Aapo Poskela[a] and Peter D. Lund [a]

Received 17th September 2017,
Accepted 21st December 2017

DOI: 10.1039/c7ee02670f

rsc.li/ees

The success of perovskite and dye-sensitized solar cells will depend on their stability over the whole life-time. Aging tests are of utmost importance to identify deficiencies and to suggest cell improvements. Here we analyzed the quality of 261 recent aging tests and found serious shortcomings in current practices. For example, in about 50% of the studies only one sample was considered, meaning that the sample size was too small for statistical significance. We propose a new procedure for aging tests based on careful planning and scientific reporting. This includes estimating the required sample size for an aging test and avoiding so-called nuisance factors, *i.e.* unintended variations always present in real world testing. The improved procedure can provide more reliable information on stability and lifetime, which could contribute to better understanding of degradation mechanisms important for improving these photovoltaic technologies.

Broader context

Perovskite and dye-sensitized solar cells are promising third generation photovoltaic technologies. In less than a decade, the conversion efficiency of perovskite solar cells has increased almost ten fold, reaching up to 20 percent. Dye solar cells can be manufactured from a variety of materials and in different colors showing wide application areas with commercial potential. Though the lifetime of both PV technologies has constantly improved, they still suffer from major degradation problems, which hamper their market breakthrough. Research on cell stability is therefore of utmost importance, but the quality of related research has shown major shortcomings. Here we comprehensively analyzed the methods used in stability and aging tests, which revealed severe weaknesses in current practices, in particular insufficient reporting and inadequate sample sizes for statistical relevance. As these deficiencies may hamper the progress of perovskite and dye-sensitized solar cells, a new procedure for aging tests is proposed including detailed instructions on how to achieve high quality in such experiments.

Open Access Article. Published on 05 February 2018. Downloaded on 2/21/2025 12:06:10 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

1 Introduction

Dye-sensitized solar cells (DSCs) and perovskite solar cells (PSCs) are among the most promising emerging photovoltaic (PV) technologies. DSCs are close to market breakthrough[1] and can be produced energy efficiently in a variety of colors[2] from low-cost materials.[3] PSCs complement their parent DSCs by reaching high efficiencies using solid-state structures and are thus an exciting, emerging area of research. Researchers have reported efficiencies as high as 22.1 ± 0.7% for small-sized PSCs.[4] The stability of both cell types has improved constantly with promising recent results such as DSCs passing a three-month outdoor aging test without external weather protection[5] and PSCs passing a 1000 hours aging test under UV illumination.[6] Even so, the quest for sufficiently long lifetime for market breakthrough continues. The concept of adequate lifetime depends on how we intend to use the solar cells: from indoor applications under intermediate conditions that require years-long stability, to building-integrated PVs under extreme stress that require decades of stability. However, it is clear that producing more reliable cells requires better understanding of their aging mechanisms.

The number of aging studies is small compared to the number of studies focused on improving the already high efficiencies, as this simplistic comparison reveals: 18% of the articles related to DSCs or PSCs listed in the Web of Science mention stability in their topic (details in ESI,† Section S5), whereas efficiency is mentioned in 79% of the articles. Recently the need for more stability research for both DSCs and PSCs has been recognized.[1,7] Additionally increased attention is directed at the standards of

[a] New Energy Technologies, Department of Applied Physics, Aalto University, P.O. Box 15100, 00076 Aalto, Finland. E-mail: armi.tiihonen@gmail.com
[b] Biobased Colloids and Materials, Department of Bioproducts and Biosystems, Aalto University, P.O. Box 16300, 00076 Aalto, Finland

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7ee02670f

This journal is © The Royal Society of Chemistry 2018

Open Access Article. Published on 05 February 2018. Downloaded on 2/21/2025 12:06:10 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

stability testing – researchers[1,7,8] and publishers[9] have called for better reporting and more uniform methods.

The main motivation in aging tests is (1) to determine the real lifetime of the cells, and (2) to compare the durability of different types of cells under certain stress factors. The first objective is challenging, because the desired lifetimes of the cells are often very long. In practice, the lifetime of the cells is investigated by accelerated aging tests. The challenge is in determining the accelerating factor of the aging test accurately enough, which is a topic still under research even for commercial silicon solar cells.[10]

The latter objective of comparing cell types is principally simple: if everybody performs the same test under the same conditions, the results should be comparable. Currently uniformity and repeatability are poor in the stability testing of both DSCs and PSCs, which is demonstrated in Section 2. This situation leads to seemingly contradictory research results – a recognized problem in the literature.[1] This in turn hampers the progress of the whole field of research as important phenomena may remain unnoticed for too long because of a lack of information on the circumstances of the aging tests.

Similar problems have been found in other fields, which have been targeted by standardizing the aging tests. This approach has also been suggested at least for PSCs.[7] Commercialized technologies – silicon and thin film solar cells – have standardized tests for estimating the durability of PV modules against prolonged exposure in climates that are specified in the standard (IEC 61215-1:2016). These tests are for initial durability testing, not for estimating the long-term stability or lifetime of the modules, although they are sometimes used as the basis for such estimations.[11]

IEC 61215-1:2016 tests are not adaptable for emerging PV technologies as such. To begin with, these third-generation technologies are too young to meet many of the established evaluations for outdoor testing; also entering the market will most likely happen in milder indoor conditions. For instance, a damp heat test designed for terrestrial thin-film solar cells at a temperature of 85 °C and 85% air humidity (IEC 61215-1:2016) would degrade most DSCs and PSCs quickly. Overly harsh tests leading to a rapid failure are unsuitable for research purposes: something more than a binary pass/fail resolution is needed to detect if there is progress in stability. Not to mention that many laboratories researching third-generation solar cells have insufficient equipment to perform the detailed and laborious tests that are designed for commercial large-scale manufacturers. Another example of standardization is the International Summit on Organic and Hybrid Photovoltaic Stability (ISOS) protocol of organic PVs[12] that has been agreed by a wide international consortium of organic PV researchers. ISOS is designed for research purposes, and it is divided into three levels from the highest levels to very basic, encouraging more groups to include stability studies in their research. The ISOS protocol also serves as a practical starting point for designing stability tests for DSCs and PSCs.

Here we present the current state of stability research of perovskite and dye solar cells, investigated with a focus on the methods and practices of performing aging tests instead of the more commonly investigated findings from aging tests. Based on our literature survey, we present practical procedures for improving the effectiveness and quality of aging testing. Our recommended procedures can be applied to all standards of testing – the focus is to maximize accumulated knowledge and accelerate aging research, regardless of how extensive or humble your facilities may be.

Our results show that the state of aging testing of PSCs and DSCs is alarming. Efforts of the whole community are required for swift corrections. Thus, we propose a series of international summits for agreeing with the principles of stability testing of these cells. The improved methods could greatly enhance the progress in stability research in future.

2 Current state of stability research

For this contribution, a group of recently published stability related articles was selected for literature survey. The selected group was formed of all the articles mentioning stability-related terms in their titles, listed in the Web of Science, and published during 2016 for PSCs and 2015–2016 for DSCs (details in ESI,† Section S5). The 157 articles included a total of 261 individual aging tests.

2.1 Group sizes in aging tests

We begin our analysis of aging methodology by looking at the group sizes used for aging tests. The test cell groups are alarmingly small in studies, as shown in Fig. 1. In half of the tests, the aging data are presented for one cell only in each cell group. It is clear that groups of one cell cannot be statistically reliable. One cell might be enough for stating that in certain circumstances the cell type can be stable for a certain period of time, but a single cell is never enough for stating that the cell type is an improvement when compared with another cell type.

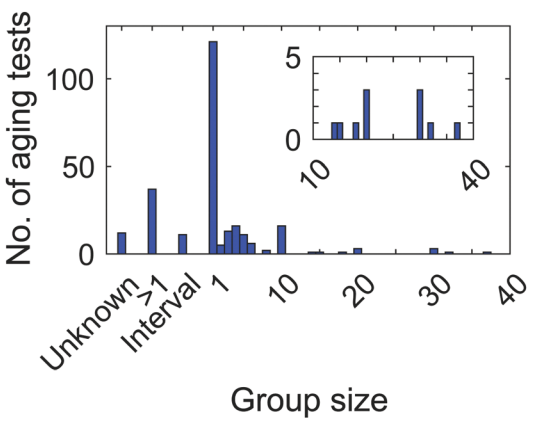

Fig. 1 The number of cells in a cell group in the investigated aging tests. The amount of tests with unknown group size, tests in which the group size is mentioned only by referring to the samples in plural form (>1), and tests in which the sample sizes are mentioned as an interval (*e.g.* 3–6 cells) are also listed. 65% of the investigated aging tests failed to meet the basic criterion of evaluating repeatability (more than one reported cell in each cell group and a clearly defined number of cells in cell groups).

Open Access Article. Published on 05 February 2018. Downloaded on 2/21/2025 12:06:10 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

In some studies, the aging tests have possibly been performed for more cells, but the data are presented only for one cell. Unfortunately the scientific audience has no means of recognizing if the study had more cells or not. Additionally presenting the data only for one cell unnecessarily dissipates the information about the repeatability of the results. Fig. 1 shows that the second most common option for group sizes is to refer to the samples in plural, but then the study neglects to report the exact group size. For the scientific audience, including the number of cells is more informative, as not doing so unnecessarily complicates the interpretation of the results.

Small groups of, *e.g.*, less than five cells are also typically insufficient for statistically significant conclusions, although they result in more reliable conclusions than comparison of single cells. Small cell groups can be used for acquiring tentative data about the differences of the cell types, but the quantitative data are unreliable because the information about the variations of the results is inadequate. As a result, the impact of the resulting article decreases. Therefore it is worthwhile to target statistically acceptable group sizes (see Section 3.2 for determining sufficient group sizes). In only 10% of the investigated tests, ten or more cells have been prepared for each cell group (Fig. 1), demonstrating that increasing group sizes to statistically acceptable levels is possible.

2.2 Aging test conditions

The vast majority of the 261 investigated tests are indoor tests for cells: only 15 tests are outdoor tests and 3 tests are on modules. The statistics might be justified for PSCs that are in the early stage of their development cycle. However, DSCs are a more mature technology and would greatly benefit from an increased number of outdoor and module studies as already recognized in the literature.[1]

38% of the investigated tests are performed for encapsulated or sealed cells, 61% for open devices. Most open devices are PSCs, probably because many DSC types contain liquid electrolytes that would soon leak out or evaporate from a cell left without proper sealing.

More than 60% of the investigated aging tests are performed at open circuit voltage, *i.e.*, operation regime that corresponds to the storage of the cells, and those are mainly done in dark conditions (Fig. 2a). Reverse bias and short circuit conditions are applied on the cells rarely, possibly because realizing these conditions in an aging test setup requires an effort (*cf.* open circuit). They are also seemingly atypical operation states of a cell. However reverse bias conditions exist in panels that are partly shadowed, for example. Also short circuit conditions might appear in a damaged cell or panel. A continuous current–voltage curve measurement (*IV*), applied as a condition in 8% of the aging tests, directly corresponds to none of the operational states of the cell. The benefit of the repeated *IV* test stress is the continuous variation of the electric state that happens in actual cell operation, although typically at a significantly slower pace in daily cycles. Only roughly one-eighth of the aging tests are performed under load, which is the main operating state of the cell (Fig. 2a). The different operating conditions of the cell should be represented in stability research for the sake of completeness. Specifically, operation under load should be utilized more in aging tests because the stability at open circuit or other electric states does not necessarily correlate with stability in real-life use.[13,14]

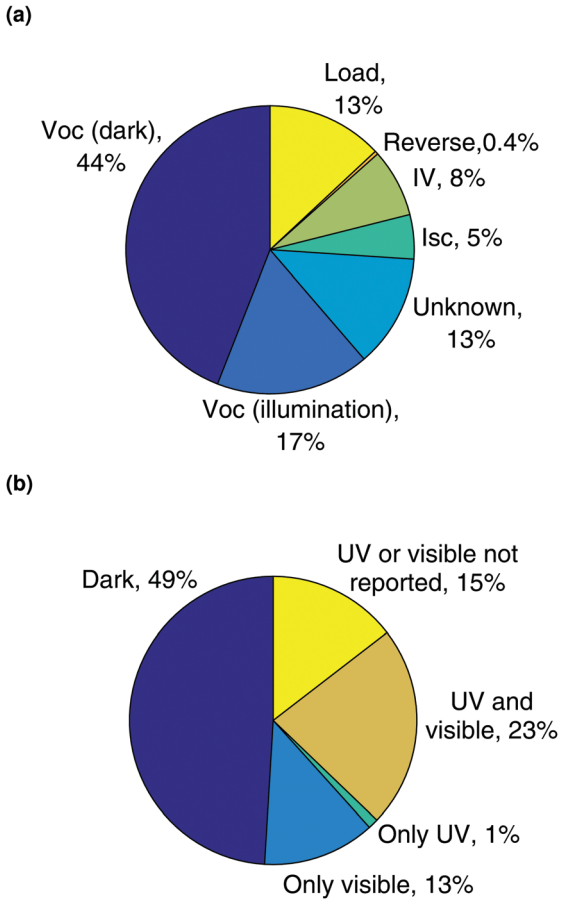

Fig. 2 (a) The electric condition of the cells in the investigated aging tests. The cells are aged at open circuit (V_{oc}) either under illumination or in the dark, under load, under reverse bias, by cycling *IV* repeatedly (*IV*), at short circuit (I_{sc}), or the electric state remains unknown (unknown). Only a minority of aging tests are performed under operational conditions (*i.e.* under load). (b) The investigated aging tests divided into dark tests, tests illuminated with visible and/or ultraviolet light, and tests that did not mention if the illumination contained ultraviolet and/or visible light. Cells are dominantly aged in the dark. See ESI,† Section S5 for more detailed information on the classification.

2.3 Environmental conditions in aging tests

Aging tests are typically performed in dark conditions, as illustrated in Fig. 2b. It is important to investigate the behavior of the cells in the dark, but the number of dark tests in comparison to the illuminated tests is currently out of proportion. Increasing the share of illuminated aging tests would be beneficial for forming a full picture of the aging behavior of PSCs and DSCs. A quarter of aging tests use both visible light and ultraviolet (UV) light. The proportion of tests that do not report the type of illumination the cells were exposed to (visible and/or UV light) is large, 15%. UV light is a recognized stress factor for both DSCs and PSCs, in many cases resulting in significantly shorter lifetimes for

 This journal is © The Royal Society of Chemistry 2018

Open Access Article. Published on 05 February 2018. Downloaded on 2/21/2025 12:06:10 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

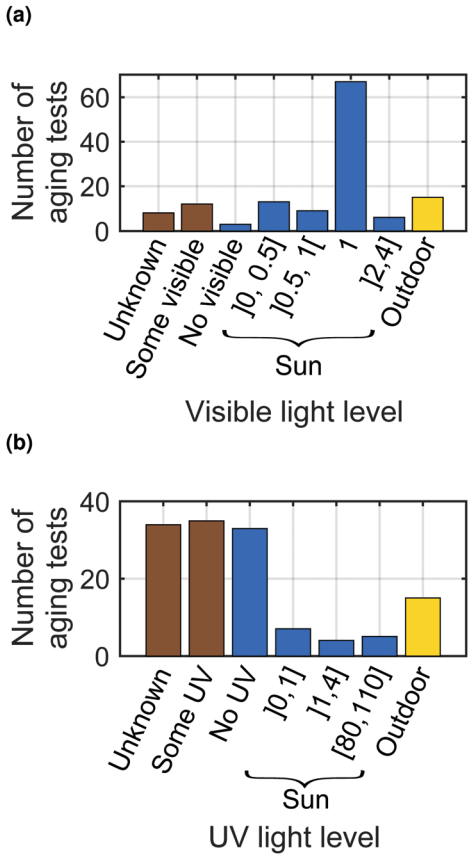

Fig. 3 The amount of (a) visible and (b) UV light in all but dark aging tests. The visible light illumination is divided into completely unknown illumination level (unknown), visible illumination but with unknown intensity (some visible), quantitatively stated illumination level (either no visible, 0–0.5 Sun, 0.5–1 Sun, 1 Sun, or 2–4 Sun), and outdoor tests. The UV light illumination is classified in a similar way. In the case of UV light, "unknown" means that the existence of UV irradiation could not be determined, and "some UV" means that UV illumination exists based on lamp type or solar simulator brand but the actual intensity of UV light is not reported by the authors or simulator manufacturer. Quantitative values are presented in Suns (0–1, 1–4, or 80–110 times the intensity of the UV part of the AM 1.5G spectrum).

the cells than pure visible light.[3,15] Therefore it is important to state what kind of illumination the cells are exposed to.

The aging of the cells is also affected by the intensity of light. For the majority of illuminated aging tests, the visible light intensity is reported quantitatively as Fig. 3a indicates; typically tests are performed at 1 Sun intensity. The state of the reporting of visible illumination seems good. In contrast, only a minority of illuminated tests are performed with quantitatively stated UV intensity (Fig. 3b). A quarter of illuminated tests apply only visible light using LED lamps for example, and another quarter provide no information about the spectrum of the light. Commonly the presence of UV light is deduced from the reported lamp type or solar simulator model, but the intensity of UV light remains unknown ("some UV" in Fig. 3b, more details in ESI,† Section S5).

It seems that generally having a commercial solar simulator with a high-quality calibration cell is regarded as sufficient to describe the aging spectrum in detail. However many standards for the solar simulator spectral accuracy specify only the spectrum above 400 nm (*e.g.* SFS-EN 60904-9, JIS C 8912, and ASTM E927 for AM 1.5G). Thus commercial solar simulators are not necessarily accurate in the UV part of the spectrum, in fact they could actually emit no UV at all, even if they would be very accurate in the visible part of the spectrum.

Currently both UV and visible light intensities are reported numerically only in 31% of the investigated illuminated indoor tests. Reporting of the intensity separately for the visible and UV parts of the spectrum would be valuable because UV illumination strongly degrades some cell types.

Fig. 4a and b illustrate the temperature and humidity applied in dark and illuminated aging tests. For most aging tests, the humidity and/or temperature are reported only qualitatively, such as being "ambient", or are not reported at all. Illuminated aging tests are commonly performed in a dry atmosphere and at low temperature. Dark storage tests are performed in a narrower temperature range than illuminated tests, but the humidity range is wider. The scarcity of published illuminated aging tests applying higher than 60 °C temperature or more than 10% humidity suggests that these conditions remain a severe stress factor for DSCs and PSCs.

This trend of missing reporting of environmental conditions during the aging tests spreads across all basic environmental variables. Numerical values are stated for visible and UV light intensities, humidity, and temperature in only one-third of aging tests. The absence of reporting suggests a lack of monitoring of the environmental variables, a situation that is problematic because other environmental factors than the intended ones (*e.g.*, humidity in a light soaking test) could greatly affect the test result. Even if the monitoring would have been appropriate, the environmental details missing from the article complicate the comparison of results with other tests.

2.4 Measurements during aging tests

The current–voltage curve (*IV*) is measured in 91% of the aging tests, which makes *IV* the most typical measurement. Other additional measurements are performed in 41% of the aging tests. These tests include a variety of measurements from the short circuit current (instead of *IV*) or maximum power point tracking and X-ray diffraction spectroscopy to electrochemical impedance spectroscopy. Selection principles for measurements applied in aging tests are further discussed in Section 3.4.

In 89% of the tests, the performance of the cells has been monitored with measurements during the aging test, in addition to the beginning and end of the test. The regular monitoring of performance provides information about the progress of degradation mechanisms that could be, for example, linear or step-wise. Typically the measurements seem to be performed manually on a regular basis during the test. Manual measurements are laborious, so working hours would be saved if the commonly repeated measurements were automated.

2.5 Stability in aging tests

The length and final efficiency in aging tests are properties with strong reverse correlation, yet the combination is a question of

Open Access Article. Published on 05 February 2018. Downloaded on 2/21/2025 12:06:10 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

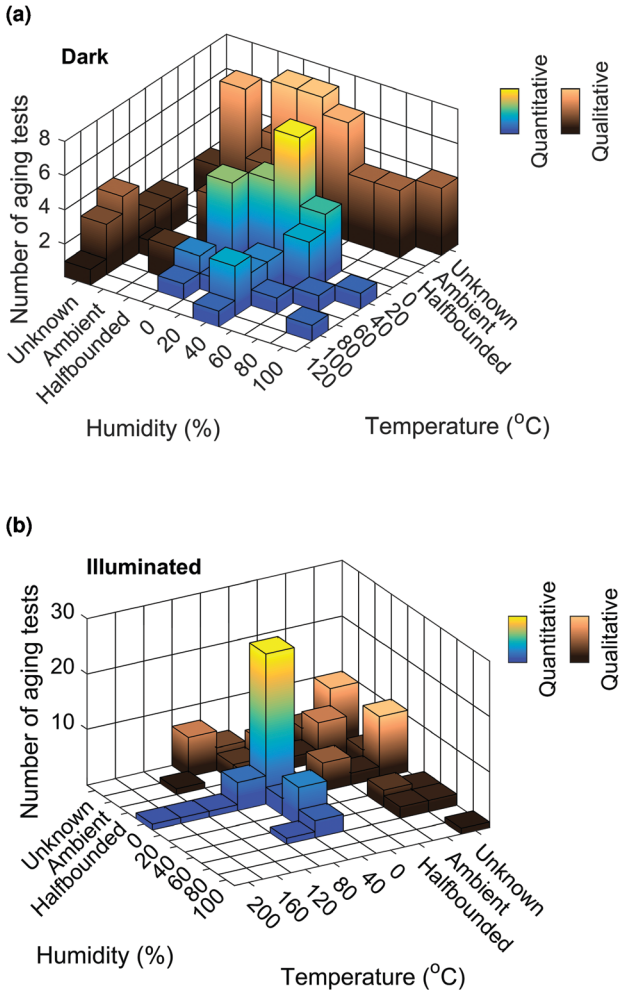

Fig. 4 The aging tests performed (a) in the dark and (b) under illumination classified based on the temperature and humidity level during the test. The tests that do not specify the value (''unknown''), define the value as ambient (''ambient''), or declare only the upper or lower boundary value (''halfbounded'') are also listed. Qualitatively reported or unknown temperature and humidity are common in aging tests which complicates the comparison of test results. See ESI,† Section S5 for more information on the classification details.

test design just like any test condition: depending on the aims of the study, one selects a longer test (potentially with lower final efficiency) or a shorter test (potentially with higher final efficiency). Fig. 5 illustrates the distribution of the final efficiency and the duration of the aging tests under illumination and in dark conditions for DSCs and PSCs. All the values represent the most stable cell group in the aging test. In the majority of published aging tests of both cell types, the final efficiency remains at 80–100% of the original efficiency (Fig. 5a and b), which could be regarded as stable. To get more information out of these studies, be it making certain aging mechanisms visible or comparing the stability of the cells with another cell type, exceeding the test period of stable performance to the degradation phase would be valuable. A change of attitudes is needed in the planning and reporting of aging tests. For example, measuring until a certain decreased efficiency level has been reached instead of a fixed test duration, and declaring more clearly the aims of the aging test (*e.g.*, the comparison of cell types, demonstration of stability, investigation of degradation mechanisms) in the resulting article would be fruitful. Presenting aging data beyond good stability should also be valued in the peer-review process and not considered a demerit.

Whereas the final efficiency distribution is rather similar for both DSCs and PSCs, the distribution of aging test durations shown in 12 and 13 varies. PSCs are typically aged for less than 250 hours, whereas DSCs are aged for more than 1000 hours. A shorter test is perhaps regarded as sufficient for PSCs that are at the early stages of development with shorter lifetime expectations, whereas for DSCs with a longer history, the 1000 hours test has become a standard.

Notably in comparison with dark tests, the majority of illuminated perovskite aging tests are very short (Fig. 5c) and end up with more scattered final efficiency for the best cell group (Fig. 5a). This suggests that illumination remains a significant stress factor for the PSCs, whereas dark tests are more easily endured (although simultaneous analysis of all the test conditions would be needed to confirm this hypothesis – which is out of the scope of this work). It is, however, clear that our understanding of the aging of the PSCs would increase if longer aging tests would be conducted more commonly, even if they would result in a lower final efficiency.

3 Principles of stability testing

In this section, we go through practical methods for achieving improved quality of aging tests. We have also gathered a practical checklist for quality aging testing compiling the findings discussed in detail in the next sections (ESI,† Section S1).

3.1 Cell design according to the aims of the aging test

The aims of an aging test can be divided into being related to either intrinsic or extrinsic stability of the cells. Intrinsic stability includes the chemical and structural stability of the cells with impurities consisting of those that are presented in the cell during the cell preparation, and extrinsic stability is the stability of the cells when external impurities can access the cell.[3]

Thus, intrinsic and extrinsic stability are closely linked to the selection of the possible encapsulation for the cells. Encapsulation isolates the active components of the cell from the environment whereas open devices remain susceptible to the incoming impurities, the loss of active cell materials from the cell, and mechanical stress, such as bending or scratching of the cell. From the industrial viewpoint, encapsulated cells resemble cells in a well-sealed solar panel, whereas open devices bring information about the stability in case of a failed sealing. The perfectness of the encapsulation varies with the method[16–20] and should be confirmed to be sufficient for the planned conditions.

3.2 How to determine the sufficient amount of test cells

Section 2 demonstrates that the group sizes in aging tests are generally small. In the extreme and unfortunately also the most common case, the data are presented only for one cell of each

 This journal is © The Royal Society of Chemistry 2018

Open Access Article. Published on 05 February 2018. Downloaded on 2/21/2025 12:06:10 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

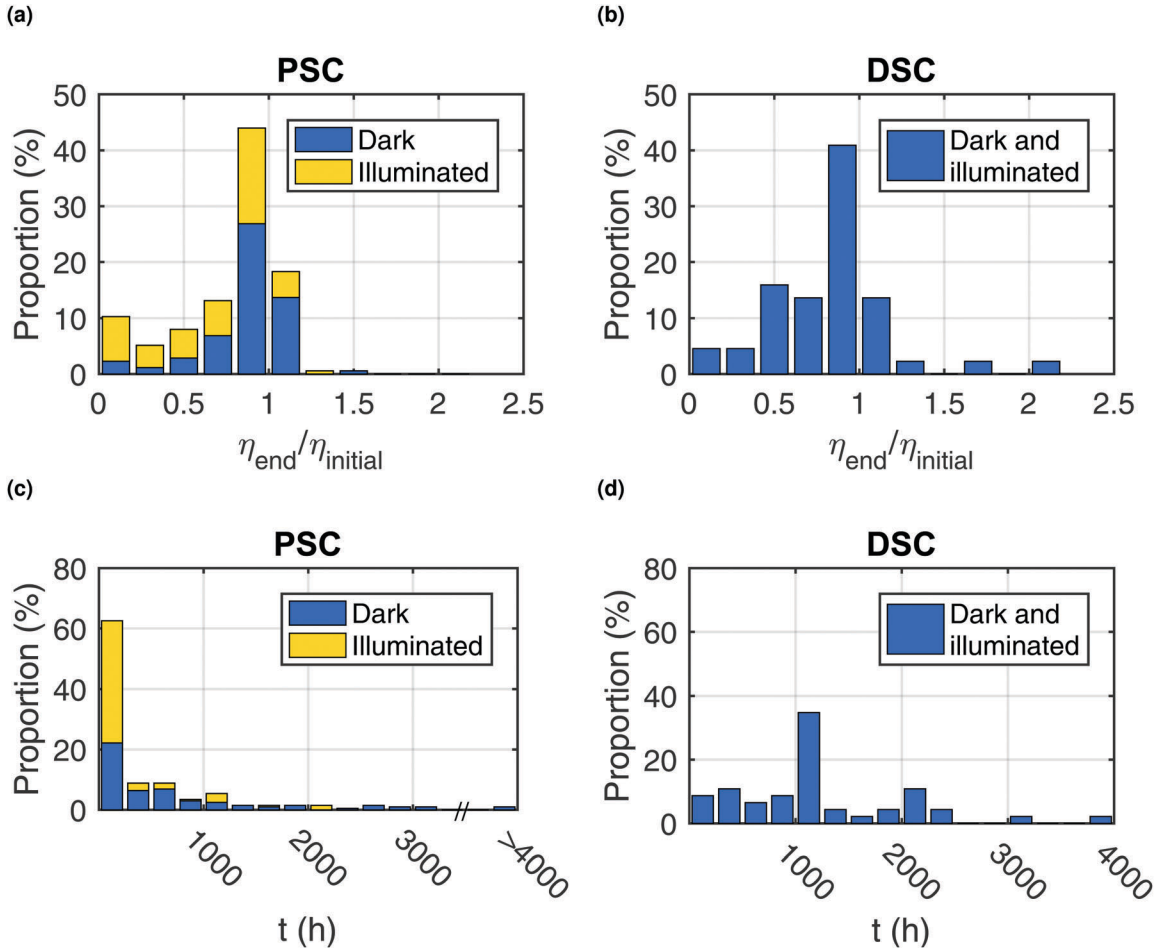

Fig. 5 The distributions of final efficiency in proportion to the initial efficiency in aging tests of (a) perovskite and (b) dye solar cells, and duration of the aging tests of (c) perovskite and (d) dye solar cells. The illuminated and dark tests are shown separately for perovskite solar cells. In the case of dye solar cells, they are combined together. All the final efficiencies are the efficiencies of the most stable cell group in the aging test. The distribution of the reported final efficiencies is similar for both cell types: cells are not aged beyond high stability in most tests. The distributions of aging test duration differ. In the case of dye solar cells, specifically 1000–1250 hours tests seem to be in favor whereas most published aging tests of perovskite solar cells are very short.

cell type. The optimal group size depends on three factors: (1) the expected effect (*e.g.*, expected difference in the final efficiency between the cell groups), (2) the variations (*e.g.*, standard deviation of the efficiency of the cells), and (3) the desired reliability of the test (*e.g.*, confidence level of 95% in a *t*-test). The larger the effect you expect to see, the less cells you need. On the other hand, the more variations you expect in the results or the more confident you want to be in the result, the more cells you need.

Typically the aging test is performed with two or more cell groups in order to have a reference group to which the stability of the investigated group is compared, *e.g.*, an illuminated aging test with a reference group stored in the dark as in ref. 21. The groups can be compared with a statistical test, a *t*-test for two cell groups, analysis of variance for multiple groups, or analysis of covariance for compensation of nuisance factors (see Section 3.3), for example. The sufficient group size can be estimated by calculating the statistical power of the statistical test that you plan to use (see a detailed example for *t*-test in ESI,† Section S4). Often long-term aging tests require larger group sizes than tests without aging. This is caused by nuisance factors (see Section 3.3) that increase variation in the final performance of the cells and accumulate during the whole aging test. *E.g.*, light intensity variations during the aging test between the aged cells could act as a nuisance factor.

The reality of solar cell research is that the cell preparation often requires manual work. The space in aging test devices is limited, as well. This holds true especially for light soaking tests where the cells cannot be stacked, unlike in dark tests. These factors create pressure to decrease the group size from the optimum value. In such cases, one could preferably decrease the amount of cell groups. Additionally, if one cell preparation method is well-established with low variation, but the other results in more variation, more cells could be in the latter group.

In practice, some cells could fail during the cell preparation or the aging test because of factors unrelated to the research question of the aging study (*e.g.*, sealing failures or a breakdown of the electrical contact). The likelihood of such cell failures should be taken into account by increasing the group sizes correspondingly.

Open Access Article. Published on 05 February 2018. Downloaded on 2/21/2025 12:06:10 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

3.3 Nuisance factors in aging tests

Nuisance factors affect the results of the aging test, although they are not studied intentionally in the aging test. They create additional variation – noise – in the results. As an example, Fig. 6 illustrates different illumination levels acting as a nuisance factor of the post-aging cell efficiency. The increased noise is undesirable because the differences between the cell groups in aging tests could, in the worst case, even be completely covered by the noise. There are two strategies for handling the nuisance factors. They can be minimized or compensated for in the result analysis.

In the primary strategy, minimization, the cell preparation procedure and the environmental conditions are ensured to remain constant between the cells and not to contain unexpected factors. A healthy dose of paranoia may be very helpful in this. If exercised, it does not significantly complicate or slow down the work.

For example, the order in which the cells are prepared can affect their performance: the components might adsorb contamination during the cell preparation, and thus the cells prepared first might have better stability than the later ones. Thus preparing and measuring different cell types in an alternating fashion is a very easy way to prevent false positive and negative results with practically zero extra work. Regarding the nuisance factors related to cell materials, different material batches might result in different stability, and even the cell assembly date might affect the results if ambient conditions (especially humidity) vary greatly from day to day. Therefore it is advisable to prepare all the cells during a short time period and when applicable from the same material batches. If that is not feasible, it would be worthwhile to have reference cells for each assembling session to verify equal quality.

Aspects related to illumination are likely nuisance factors. The spectrum could affect the results of the aging test greatly, in the case of many cell configurations.[1,22,23] Therefore, one should be aware of the spectrum of the light (especially if it contains UV light) and the effects of possible filters on it. The spatial variations of illumination intensity across the aging platform could be tens of percent and still remain unnoticed by the human eye because of the eye's good intensity adaptivity. The most simple option to measure the spatial intensity distribution is to use a photodiode that is sensitive to the applied illumination spectrum. In our light soaking system, we record the light intensity for each cell separately, for example on a weekly basis, in addition to constant tracking of a few spots. Other significant environmental factors than intensity should also be followed during the aging test. Just stating "ambient" is not enough. To give an example, ESI,† Fig. S7 illustrates the indoor air humidity varying greatly in both the short-term (between days) and long-term (between seasons). These variations certainly affect the aging of moisture-sensitive unsealed devices.

Nuisance factors could remain significant even after they are minimized, their importance could be detected only after the aging test, or they might be impossible to control during the test. In these cases, the alternative strategy is to compensate for the most significant nuisance factor(s) after the aging test with regression analysis or analysis of covariance (ANCOVA), for example. ANCOVA is used for determining if the groups are different regardless of a covariate, that is, the nuisance factor (application example in ref. 22). The compensation is naturally practicable only for nuisance factors that are measurable.

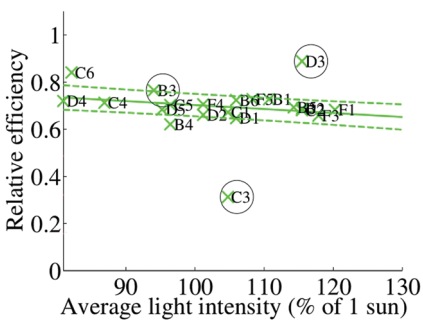

Fig. 6 Illumination level is acting as a nuisance factor of the efficiency of dye solar cells after 700 hours of aging under illumination[22] because the post-aging efficiencies form a line with negative slope with respect to the illumination intensity (measured separately for each cell). A reprinted derivative from ref. 22 under CC-BY license.

3.4 Measurements during aging tests

The measurements performed in the aging test should be capable of revealing the effects that are anticipated in the hypothesis. The *IV* curve measurement/characterization is used for tracking the stability of the cells, but its use in the analysis of degradation phenomena is very limited. Therefore, all but the most simple aging tests require additional measurements.

The measurements should be selected and performed so that they do not add unnecessary nuisance factors to the aging test. Some tests, such as electrochemical impedance spectroscopy (EIS) or *IV* cycling, could affect the electrochemical state of the cells and consequently the results of the following measurements of the same sample. Thus, the measurement sequence should be kept unchanged during the aging test.

Sometimes, a measurement can even accelerate or activate the degradation of the cells. For example, measuring the *IV* curve of the cells with metal substrates far to the reverse load conditions can trigger corrosion. The corrosion reactions have a certain polarization after which corrosion occurs. Consequently, polarization can be used even as corrosion prevention but also to trigger it. Degradation reactions related to device polarization have also been observed in cells without metal electrodes.[24] As another example, the cells could be so sensitive to UV light that the prolonged and repeated *IV* measurements under full spectrum illumination could trigger cell degradation even though they might otherwise pass the aging test. The *IV* curve of the cell could be measured before and after the whole measurement sequence to confirm that the cells remain stable during the measurements.

The measurements can be performed either only before and after the aging test, or repeatedly during the aging test. Measuring the cells only before and after the aging test saves working hours but gives limited information on the degradation phenomena.

 This journal is © The Royal Society of Chemistry 2018

Open Access Article. Published on 05 February 2018. Downloaded on 2/21/2025 12:06:10 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

For ambitious aging studies, the frequent monitoring of all the cells and the environmental parameters, like in ref. 25, is recommended. Tracking the environmental parameters permits the compensation for nuisance factors when necessary, and monitoring the cells allows investigating nonlinear degradation phenomena. Additionally, laborious measurements can be timed optimally so that the aging signs are visible in the cells but they still operate well enough for the measurement. For example, electrochemical impedance spectroscopy could provide detailed information about the degrading components of the cell but the results become difficult to analyze if the cell is too degraded (see a demonstration in ESI,† Section S3).

The parameters could be tracked automatically or manually. Automatic measurements can save working hours, and in some cases, they increase the accuracy of the data collection. For example, the variations in humidity in ESI,† Fig. S7, partly connected to the different level of the air-conditioning outside office hours, could remain unnoticed without automated measurements.

3.5 Reporting of aging data

The literature review in Section 2 revealed that insufficient reporting of the measurement data and conditions is one of the main problems in aging tests at the moment. This leads to difficulties in comparing the results. The situation can be improved simply by reporting the measurement data of all the cells that have been aged in ESI,† as in ref. 26. Similarly it should be reported how the environmental parameters have been followed, like in ref. 15 and 27.

Reporting outliers, that is cells dropped from the final analysis, increases the reliability of the study. Cells with major scratches or leaking electrolytes are typical justified outliers that could lead to false conclusions if they would be kept in the analysis. There are also statistical methods, such as Peirce's criterion,[28] for objectively detecting outliers from the measurement data.

Using statistical methods in the analysis of results increases the weight of the article because they provide a collectively agreed basis for the conclusions. Statistical methods can be utilized if the assumptions of the selected statistical test are met and the cell groups are large enough. The assumptions could include having no outliers, equal variances in the groups, normally distributed data, or equal group sizes. The sufficient group size for statistical testing varies from test to test. For example, a *t*-test can technically be performed even for groups of two cells.

However the probability of getting false negative results increases if the groups are small. In practice, this means that no difference between the two groups is detected even if there is one, unless the difference is very notable. To give an example, if the group size is three cells and the difference between the distributions is twice the standard deviation of them, the probability of not detecting the difference between the two groups is more than 50%.[29] Increasing the group size to five cells already decreases the probability to approximately 20%.[29] Additionally, numerical simulations show that small groups combined with the violation of the assumptions of the *t*-test lead to increased probability of both false positive and negative results.[29]

Last but not least, the significance of the results should be expressed whether or not statistical methods are applied in the analysis. In principle, even the smallest differences between two distributions could be detected and be statistically significant if enough samples are investigated. But a remarkable difference is not only statistically significant but also practically noteworthy. Since the definition of noteworthiness varies from person to person, every research team should ask themselves if the acquired differences are large enough to matter. For example, would a statistically significant 5% higher mean short circuit current density in a test group compared to a reference group be a noteworthy result, or should the results be regarded practically equal? How about 5% higher mean open circuit voltage?

4 Conclusions

The two main deficiencies in the recent aging studies are (1) inadequate group sizes for statistical analysis and (2) insufficient reporting of the measurement conditions. The aging data are typically reported only for one cell, and the basic environmental parameters, temperature, humidity, and intensities of both ultraviolet and visible illumination, were given quantitatively only in one-third of the 261 reviewed aging studies. A numerical value for UV light intensity is glaringly absent, and commonly, only "ambient" humidity is reported which could in reality be almost any value between 0 and 100%. The fixing of these shortcomings is straightforward – increasing the number of cells in the reported aging groups, and monitoring and reporting quantitatively at least the basic environmental parameters. This would be a great start in reducing the amount of seemingly contradictory results, and should be required by the publishers. The checklist provided in the ESI† should be helpful for this purpose. The improved reporting could open new opportunities, for instance the accumulated aging knowledge could be used as input data for machine learning algorithms to investigate degradation mechanisms and the rate of degradation.

We could as a community be bolder in aging testing since the typical applied environmental conditions are still limited to storage conditions, for example dark tests under ambient temperature and moisture, and the open circuit state of the cells. More tests under load and illumination are needed for mapping the stability and degradation of the cells under all the conditions in which the cells actually are used. One issue that is often left unspoken is nuisance factors creating unintended variance in the test results. The best defense against nuisance factors is to overcome them by elimination but is all hope lost if the nuisance factors persist anyway? Comforting news is that their effect can be analyzed with statistical methods, in the best case giving additional insight into degradation (*e.g.*, degradation as a function of light intensity) – provided that the research team is proactive in monitoring potential nuisance factors.

The state of aging testing of perovskite and dye solar cells investigated thoroughly in this work requires swift actions for

Open Access Article. Published on 05 February 2018. Downloaded on 2/21/2025 12:06:10 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

improvements. The whole community should collaborate in the process. Thus, we suggest a series of international summits for determining the definitive standards of stability testing of these cells. This move could greatly enhance the progress in stability research in future.

Authors contributions

A. T. is the corresponding author who made the literature review, compiled the principles of improving stability testing, and wrote the article. All authors discussed the results and implications, and edited the article. A. T. and S. L. performed the air humidity measurement presented in ESI.† K. M. instructed the work and P. L. supervised it.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

The authors thank Tiina and Antti Herlin Foundation, Kone Foundation, and the Academy of Finland (project SOLID, 271081) for the financial support. The authors thank Heidi Henrickson for valuable comments on the article.

References

1 A. Hinsch, W. Veurman, H. Brandt, K. Flarup Jensen and S. Mastroianni, *ChemPhysChem*, 2014, **15**, 1076–1087.
2 S. G. Hashmi, M. Özkan, J. Halme, S. M. Zakeeruddin, J. Paltakari, M. Grätzel and P. D. Lund, *Energy Environ. Sci.*, 2016, **9**, 2453–2462.
3 M. Asghar, J. Zhang, H. Wang and P. Lund, *Renewable Sustainable Energy Rev.*, 2017, **77**, 131–146.
4 M. A. Green, K. Emery, Y. Hishikawa, W. Warta, E. D. Dunlop, D. H. Levi and A. W. Y. Ho-Baillie, *Prog. Photovoltaics Res. Appl.*, 2017, **25**, 3–13.
5 G. Griffini, F. Bella, F. Nisic, C. Dragonetti, D. Roberto, M. Levi, R. Bongiovanni and S. Turri, *Adv. Energy Mater.*, 2015, **5**, 1401312.
6 S. G. Hashmi, A. Tiihonen, D. Martineau, M. Ozkan, P. Vivo, K. Kaunisto, V. Ulla, S. M. Zakeeruddin and M. Grätzel, *J. Mater. Chem. A*, 2017, **5**, 4797–4802.
7 Y. Yang and J. You, *Nature*, 2017, **544**, 155–156.
8 T. Leijtens, G. E. Eperon, S. Pathak, A. Abate, M. M. Lee and H. J. Snaith, *Nat. Commun.*, 2013, **4**, 2885.
9 Editorial, *Nat. Mater.*, 2015, **14**, 1073.
10 V. Sharma and S. Chandel, *Renewable Sustainable Energy Rev.*, 2013, **27**, 753–767.
11 K. Branker, M. Pathak and J. Pearce, *Renewable Sustainable Energy Rev.*, 2011, **15**, 4470–4482.
12 M. O. Reese, S. A. Gevorgyan, M. Jørgensen, E. Bundgaard, S. R. Kurtz, D. S. Ginley, D. C. Olson, M. T. Lloyd, P. Morvillo, E. A. Katz, A. Elschner, O. Haillant, T. R. Currier, V. Shrotriya, M. Hermenau, M. Riede, K. R. Kirov, G. Trimmel, T. Rath, O. Inganäs, F. Zhang, M. Andersson, K. Tvingstedt, M. Lira-Cantu, D. Laird, C. McGuiness, S. J. Gowrisanker, M. Pannone, M. Xiao, J. Hauch, R. Steim, D. M. DeLongchamp, R. Rösch, H. Hoppe, N. Espinosa, A. Urbina, G. Yaman-Uzunoglu, J.-B. Bonekamp, A. J. van Breemen, C. Girotto, E. Voroshazi and F. C. Krebs, *Sol. Energy Mater. Sol. Cells*, 2011, **95**, 1253–1267.
13 M. Berginc, U. O. Krašovec and M. Topič, *Sol. Energy Mater. Sol. Cells*, 2014, **120**, 491–499.
14 S. Mastroianni, A. Lanuti, S. Penna, A. Reale, T. M. Brown, A. Di Carlo and F. Decker, *ChemPhysChem*, 2012, **13**, 2925–2936.
15 F. Bella, G. Griffini, J.-P. Correa-Baena, G. Saracco, M. Grätzel, A. Hagfeldt, S. Turri and C. Gerbaldi, *Science*, 2016, **354**, 203–206.
16 J. Macaira, L. Andrade and A. Mendes, *Sol. Energy Mater. Sol. Cells*, 2016, **157**, 134–138.
17 D. Ivanou, R. Santos, J. Macaira, L. Andrade and A. Mendes, *Sol. Energy*, 2016, **135**, 674–681.
18 T. J. Wilderspin, F. D. Rossi and T. M. Watson, *Sol. Energy*, 2016, **139**, 426–432.
19 Q. Dong, F. Liu, M. K. Wong, H. W. Tam, A. B. Djurisic, A. Ng, C. Surya, W. K. Chan and A. M. C. Ng, *ChemSusChem*, 2016, **9**, 2597–2603.
20 F. Matteocci, L. Cina, E. Lamanna, S. Cacovich, G. Divitini, P. A. Midgley, C. Ducati and A. D. Carlo, *Nano Energy*, 2016, **30**, 162–172.
21 S.-W. Lee, S. Kim, S. Bae, K. Cho, T. Chung, L. E. Mundt, S. Lee, S. Park, H. Park and M. C. Schubert, *et al.*, *Sci. Rep.*, 2016, **6**, 38150.
22 A. Tiihonen, K. Miettunen, S. Rendon, D. Mavrynsky, J. Halme, R. Leino and P. Lund, *J. Electrochem. Soc.*, 2015, **162**, H661–H670.
23 W. Li, W. Zhang, S. Van Reenen, R. J. Sutton, J. Fan, A. A. Haghighirad, M. B. Johnston, L. Wang and H. J. Snaith, *Energy Environ. Sci.*, 2016, **9**, 490–498.
24 T. Handa, D. M. Tex, A. Shimazaki, T. Aharen, A. Wakamiya and Y. Kanemitsu, *Opt. Express*, 2016, **24**, A917–A924.
25 Y. Reyna, M. Salado, S. Kazim, A. Pérez-Tomas, S. Ahmad and M. Lira-Cantu, *Nano Energy*, 2016, **30**, 570–579.
26 M. Salado, J. Idigoras, L. Calio, S. Kazim, M. K. Nazeeruddin, J. A. Anta and S. Ahmad, *ACS Appl. Mater. Interfaces*, 2016, **8**, 34414–34421.
27 Y. Dkhissi, S. Meyer, D. Chen, H. C. Weerasinghe, L. Spiccia, Y.-B. Cheng and R. A. Caruso, *ChemSusChem*, 2016, **9**, 687–695.
28 S. M. Ross, *J. Eng. Technol.*, 2003, **20**, 38–41.
29 J. C. De Winter, *Pract. Assess., Res. Eval.*, 2013, **18**, 10.

 This journal is © The Royal Society of Chemistry 2018