NEEA Method for Measuring TV Screen-Average Dynamic Luminance with a Camera Photometer

Revision Control

V15 – Gregg Hardy, 11/2	Updated camera distance range to 1.53-1.55. Added Appendix E: Justification of camera distance.
V16 – Gregg Hardy, 11/16	Updated camera distance to 1.77x ± 0.5'' the screen width and updated Appendix E for same reason.
v17 – Gregg Hardy, 11/19	Updated cam distance and lens model.
V18 – Trevor Swope, 12/18	Updated with independent test lab results, additional camera stability analysis, other small edits.
V19 – Gregg Hardy, 12/28	Updated appendices.
V20 – Carl Bruder 01/08	Updated Labels for tables and figures
V21 – Trevor Swope 02/24	Finished Labels for all tables and figures, incorporated tech-writer feedback.

Pacific Crest Labs on behalf of Northwest Energy Efficiency Alliance (NEEA) with technical support from Dynamic Motion Control (DMC) and VideoQ and funding from the ENERGY STAR® program.

Contents

Table of Figures

Table of Tables

Introduction

This document describes a camera-based method of measuring the luminance of a TV that is playing dynamic video. It covers the technical capabilities of the hardware, the image processing techniques employed, the calibration procedures used, and an accuracy assessment.

Camera photometers have advantages over the spot photometers currently used to measure TV luminance. Camera photometers are capable of viewing and measuring light output across the entirety of the screen, measuring light output during dynamic video play, and recording the TV image during the test.

Our goal is for the sum of all possible sources of error in our camera photometer approach to reach, at worst, <5% expected accuracy to the real luminance value as measured by a hypothetical “perfect camera photometer.” We have met our goal, and we believe our camera approach to be significantly more accurate than current measurement methods, as explained in detail in Error Analysis.

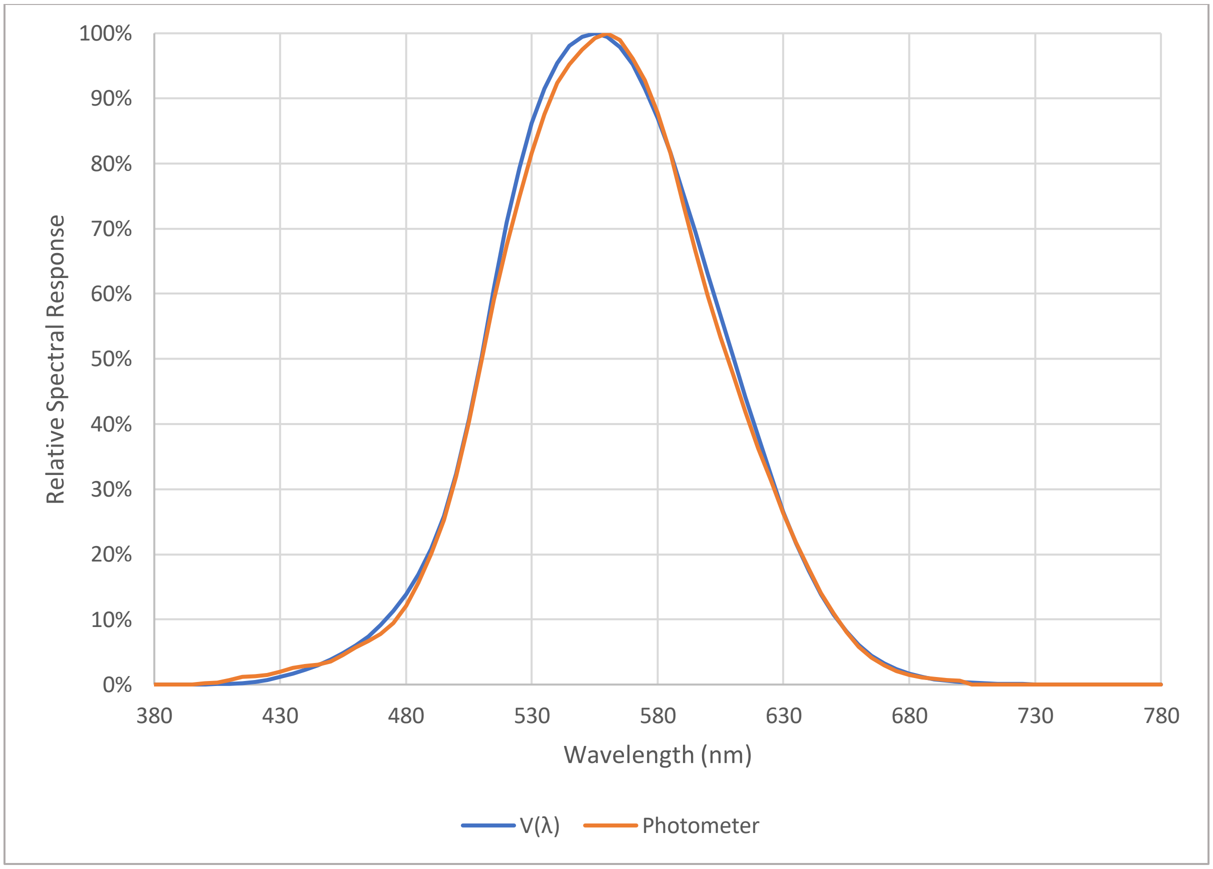
Background

CIE 1931 Luminosity Function

A camera photometer is intended to accurately measure luminance of a light source as observed by a human. To do so, it needs to have a response curve equivalent to the CIE 1931 luminosity function $V(\lambda)$. While a less-used alternative (CIE 1964 Standard Observer) exists, we use CIE 1931 given that Konica-Minolta and other major photometer manufacturers benchmark against this function.

We calculate the expected spectral response of our camera system from the spectral response of the sensor in the camera, and the transmissivity of the lens and the photopic and neutral density filters.

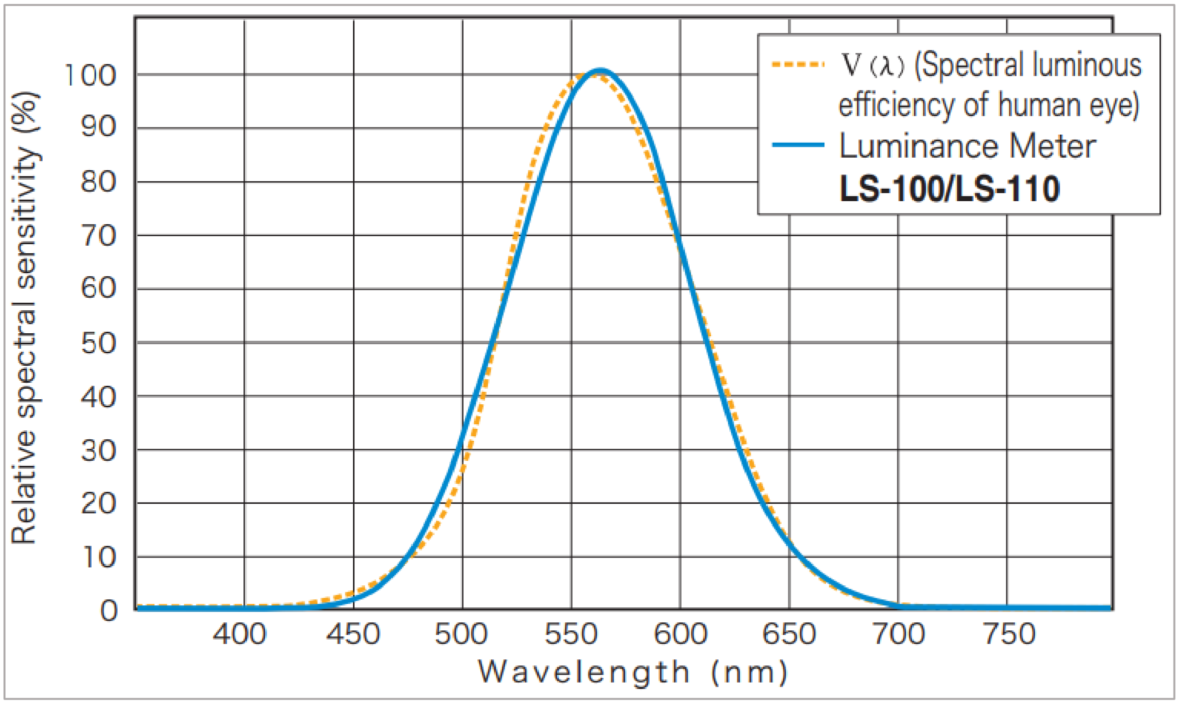
Figure 1. Camera Photometer System Spectral Response vs. V(λ)



Our filter match appears visually similar to the published curve of a Konica Minolta LS-100/LS-110:[1]

[1]Graph from LS 150/160 catalog, see page 2: https://sensing.konicaminolta.us/wp-content/uploads/ls-150_160_catalog-8z1qyj292u.pdf

Figure 2: LS-100/LS-110 Filter Match



f_1' against Illuminant A

A metric for a system's fit to the CIE 1931 curve is defined in CIE 19476, 3.2.2, as an index "describing the deviation of the relative spectral responsivity of the photometer from the $V(\lambda)$ function." Specifications for most luminance meters include the spectral mismatch against standard illuminant A. Although this does not quite fit our use case of measuring LED lights, it helps characterize the quality of our camera photometer against other systems. We more specifically characterize our theoretical accuracy for TV LED, QLED, and OLED light sources in Appendix A: Specific Mismatch to TV backlights. In the absence of an LED standard illuminant, we similarly calculate our expected spectral mismatch index f_1' for our camera response $s_{rel}(\lambda)$ against the photopic curve $V(\lambda)$ for standard illuminant A $S_A(\lambda)$ as:[2]

$$s_{rel}^*(\lambda) = s_{rel}(\lambda) * \frac{\int_{380nm}^{780nm} S_A(\lambda) * V(\lambda) \mathrm{d}\lambda}{\int_{380nm}^{780nm} S_A(\lambda) * s_{rel}(\lambda) \mathrm{d}\lambda}$$

$$f_1' = \frac{\int_{380nm}^{780nm} |s_{rel}^*(\lambda) - V(\lambda)| \mathrm{d}\lambda}{\int_{380nm}^{780nm} V(\lambda) \mathrm{d}\lambda}$$

[2] Equations lifted directly from ISO/CIE 19476: "Characterization of the performance of illuminance meters and luminance meters"

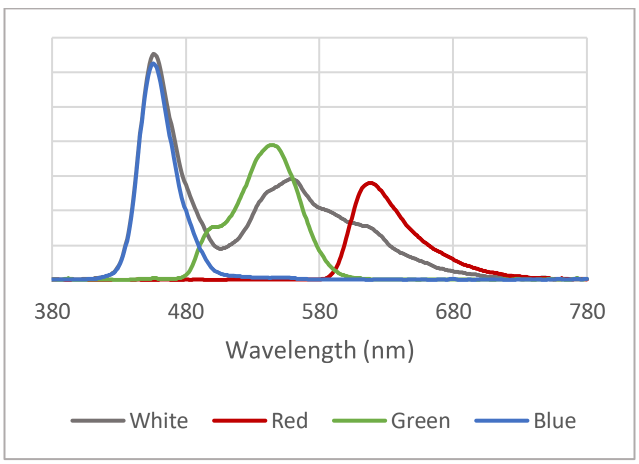
With spectral response estimates for our camera at 5nm,[3] we estimate our spectral mismatch index to be <3% to illuminant A, mostly driven by the filter match of the photopic filter we chose. This value is useful as a general comparison to the quality of the filter match of other photometers, such as the Konica Minolta LS-150, but is not specific to our measurement case: TV LED, QLED, and OLED luminance.

Spectral Mismatch against TV Light

Examining the spectral response of our photometer indicates that our system is likely to under-report luminance in the 580-640nm range, which is where the red channel of most TV displays is going to peak; in other words, our camera system is likely to be under-sensitive to reds. We observe the same difference at lower wavelengths, in blues; however, since blues tend to peak around 440-450 nanometers and have overall less relative luminance than reds for the same spectral power, they are likely to contribute less significantly to error.

Due to differences in spectral profiles among different TV technology types, we cannot simply use a single color correction calibration for the camera system; given different peaks of reds, greens, and blues, and different overall curves, the differences among the spectral profiles of the light of each TV is something to which our camera photometer, like other camera photometers, is sensitive. We measured spectral profiles for the TVs in our lab with a SpectraScan PR650 spectroradiometer to confirm this, noting the difference in spectral profiles among OLEDs, QLEDs, and LCDs.

Figure 3. OLED Spectral Curve



[3] Greater precision in this estimate could be achieved by directly measuring the camera's response with an integrating sphere, rather than assuming the data from the sensor and lens datasheets are perfectly accurate.

Figure 4. QLED Spectral Curve

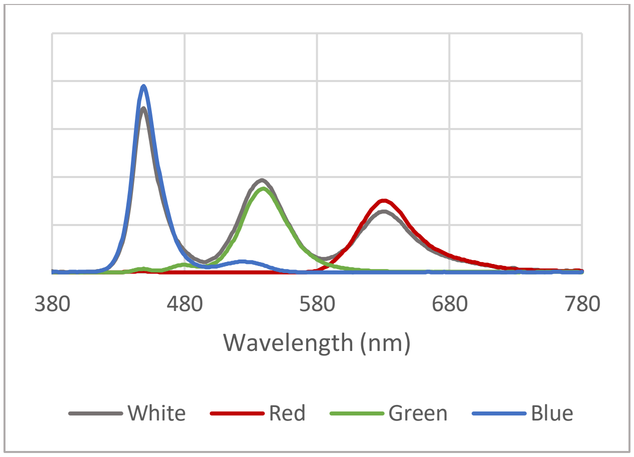


Figure 5. LCD 1 Spectral Curve

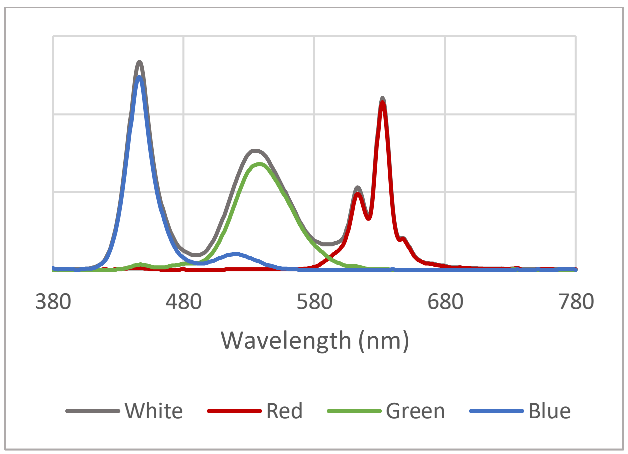
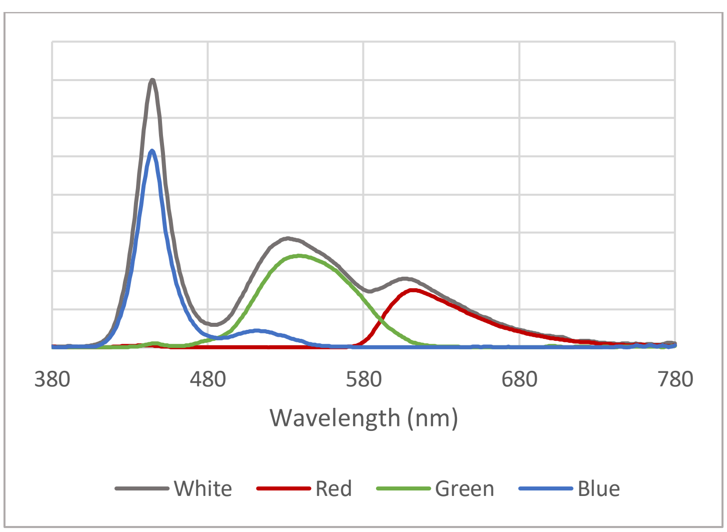


Figure 6. LCD 2 Spectral Curve



Spectral profiles can vary significantly across different TV types, and even from one LCD panel to the next, with different peaks and shapes. Since our camera photometer system is more or less sensitive to different areas of the visible spectrum, we risk TV-to-TV inconsistency. For example, we multiply our camera sensitivity by the spectral curves for the two LCD TVs, to get an idea of the relative perceived luminance of each color for the camera, vs. that of the human eye.

Observing the differences on the red and blue component curves for various TV profiles, it becomes apparent that the amount of error contributed by imperfect filter match can depend on the spectral profile of the TV, particularly the peak wavelengths of the red and blue channels.

Using a similar calculation of spectral mismatch as for illuminant A, replacing $S_A(\lambda)$ with $S_{TV}(\lambda)$, we can calculate a general spectral mismatch index for the camera photometer for saturated red, green, and blue screens, as well as white screens, on a few of the TVs we have in the test lab. Note that general spectral mismatch index is not an "absolute" error calculation, but a weighted average of how far in general the curve deviates from the target curve, weighted by the profile of the light being measured. Absolute expected error is calculated in Appendix A: Specific Mismatch to TV LEDs.

Table 1. Calculated General Mismatch Index Against Light Sources for Various TV Types

	Pure Red	Pure Green	Pure Blue	**Pure White**
LCD LED#1	2.9%	2.7%	2.7%	**2.7%**
LCD LED#2	2.8%	2.7%	2.7%	**2.7%**
QLED	2.8%	2.7%	3.0%	**2.8%**
OLED	3.0%	2.7%	4.6%	**2.7%**

Camera System

Camera and Lens Selection

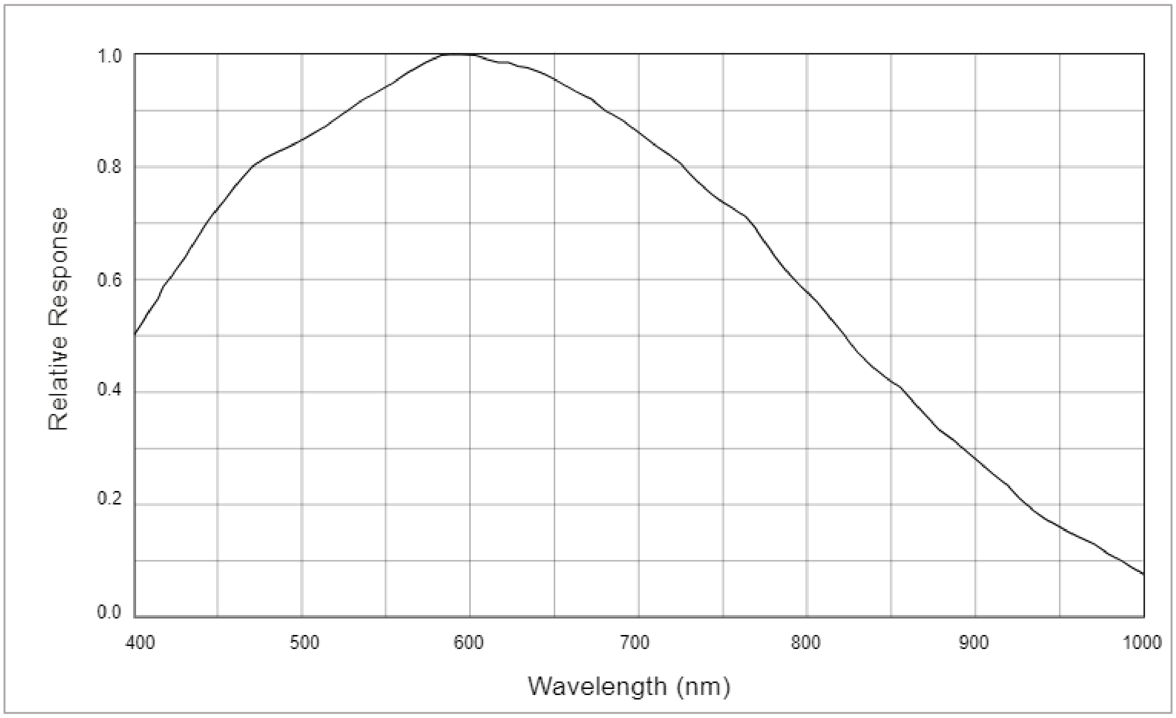
We chose our camera and lens based on the geometry of our test setup and on the need to achieve continuous luminance measurement at approximately 6 frames per second with a data stream that can be processed real-time by a conventional laptop. Other measurement devices designed to have a close match to the luminosity function exist but have limitations making them prohibitive for this application. Spot photometers are incapable of measuring the full screen during dynamic video play, and other camera photometers on the market today are not capable of taking multiple frames a second with continuous exposure. Our camera system solves both of those problems.

Basler acA720-290gm Camera

The spectral response of our camera (without lens), which is largely dictated by the micro-lenses on the Sony IMX287LLR-C CMOS sensors pixels, is:[4]

[4] https://www.baslerweb.com/en/products/cameras/area-scan-cameras/ace/aca720-290gm/

Figure 7: IMX287 CMOS sensor spectral response



Basler C23-0816-2M-S F1.6 f8.6mm Lens

The camera is placed a distance of 1.76–1.78 times the screen width of the TV per the rationale in Appendix E: Justification for Camera Placement Distance from TV. We chose this distance to best represent typical viewing distance across different screen resolutions (e.g., HD, UHD, 8K).

The camera line of sight is positioned normal to the plane of the TV screen and aimed at the center of the TV screen. This positioning simulates the perceived amount of light that a normal viewer would observe when watching TV.

We selected the Basler C23-0816-2M F1.6 f8.6mm lens. An 8mm lens paired with the Basler camera allows for the full width of the TV to be in the image field of view when the camera is placed at a specified distance from the screen. The minimum working distance of the lens is 100mm, or less than 4 inches, well below the minimum TV screen width size.

For an observer centered in front of the TV, pixels at the edges of the TV screen will generally appear dimmer than pixels in the center of the image due to the beam angle of the pixels. The test software does not compensate for this viewing angle effect for the camera because the effect is also observed by a viewer, and the test is intended to emulate the perceived intensity of a viewer.

Basler Lens C23-0816-2M-S F1.6 f8.6mm lens transmittance was unavailable and is assumed to be relatively flat.

Camera Linearity

According to a paper by Hiscocks, 2014, the luminance of the light hitting the camera sensor pixel is directly proportional to the sensor reading value of that pixel in the image,

$$N_d = K_c \left(\frac{t\,S}{{f_s}^2} \right) L_s$$

where the quantities are

N_d	Digital number (value) of the pixel in the image
K_c	Calibration constant for the camera
t	Exposure time, seconds
f_s	Aperture number (f-stop)
S	ISO Sensitivity of the film
L_s	Luminance of the scene, candela/meter^2

In this system, the exposure time, and aperture, of the camera are set at constant values. The ISO sensitivity of the Basler camera is hard-coded into the hardware for optimum linearity. A calibration (described later in this document) is conducted to determine the calibration coefficient to convert pixel brightness to luminance. Our goal is to measure a TV's screen-average luminance during dynamic video play as perceived by the human eye from a typical viewing distance. Since digital camera response is linear to luminance with all other variables fixed, we expect our camera's response to be linear.

Camera Exposure

The frame rate of the camera is set as the reciprocal of the exposure time: for example, if the exposure time is 0.1 seconds, the frame rate is set as 10 fps.

Figure 8: Camera Exposure/Readout Clock Diagram

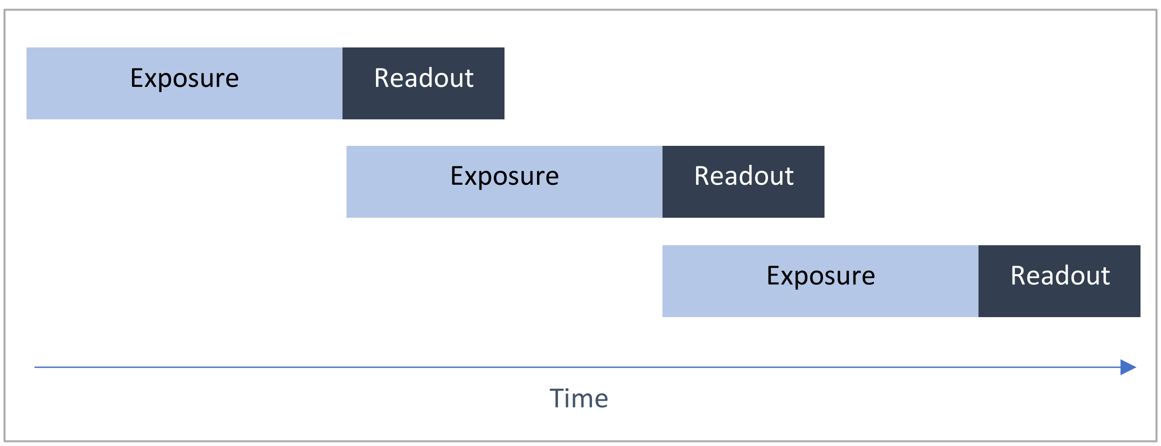


Image acquisition is overlapping, such that during sensor readout, the next frame exposure begins. There is no delay between one frame ending and the next frame beginning: all light passing through the camera lens is recorded in an image. See [Basler Product Documentation] for additional information.

The camera is set at a constant exposure time, aperture, and ISO for all tests with all TVs. We set the camera aperture so that signal level is approximately equal to twice the luminance level (cd/m2) at 6 frames per second. That enables our 12-bit camera to read a maximum of 2048 cd/m^2 (2008 after master black level adjustment discussed later). We chose 6 fps because it limited the image data rate to a level that an affordable laptop could process real-time.

Note: The Basler camera photometer warms up to an equilibrium temperature of about 50°C (122°F),[5] and this can cause the screw used to fix the aperture to loosen; therefore, we secure the lens aperture position using nail polish to prevent it from shifting over time and putting the camera out of calibration. If the nail polish seal is broken, the tester can visually determine that the camera photometer needs recalibration.

Camera Filter Selection

B+W 43/47mm XS-Pro MRC-Nano 806 ND 1.8 Filter (6-Stop)

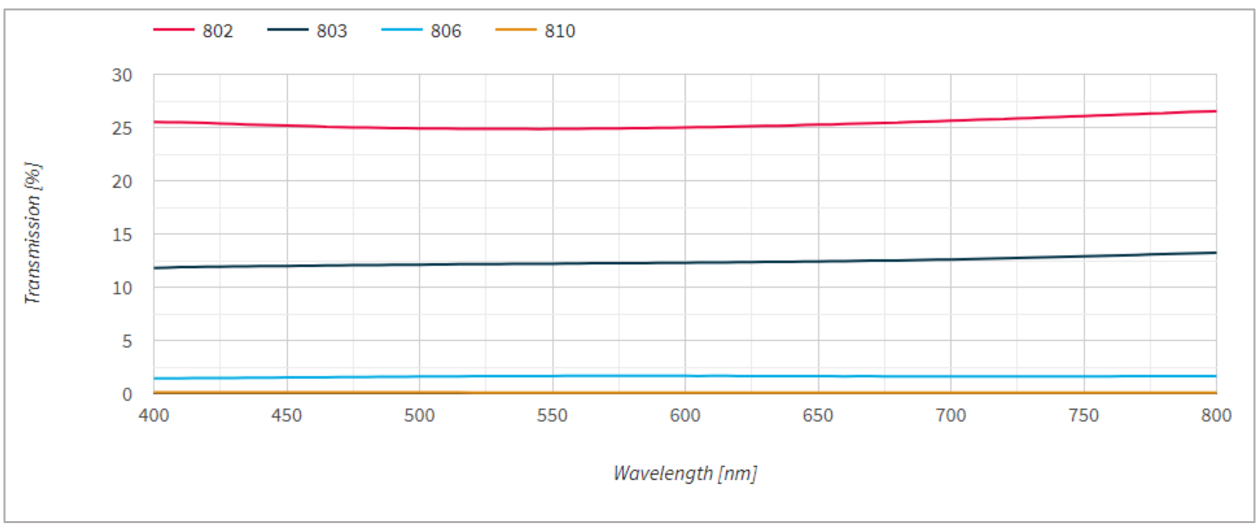
The addition of a neutral density filter allows us to measure brighter objects without over-exposing the pixels; reducing the incoming light by a factor of 64 with a 6-stop neutral density filter allows us to measure up to 2048 cd/m^2, as discussed above, with the aperture set approximately to the middle of its range.

The B+W neutral density filter was chosen for the following reasons:

1. It has a flat spectral response curve. In the following figure, the blue line (806) gives the transmission of this filter line across the spectrum.
2. It has a relatively uniform effect over the surface of the filter.

Any aberrations that do exist across the surface of the filter will be corrected for during vignette correction; see Vignette Effect Correction (Flat Field Correction).

Figure 9: Nano 806 Neutral Density Filter Spectral Transmissivity [6]



Omega 558BP100 38mm photopic filter

To strengthen our camera system's fit to $V(\lambda)$, we choose an off-the-shelf photopic filter with a spectral mismatch index $f_1{}'$ of <3%. We aim to achieve a close fit to luminosity function with the integrated camera system, the sum of the response and transmittance curves for camera, lens, and filters. The final filter match of the camera photometer is a combination of that of the photopic filter, the neutral density filter, the camera sensor response, and lens transmittance. We find that the combined response curve

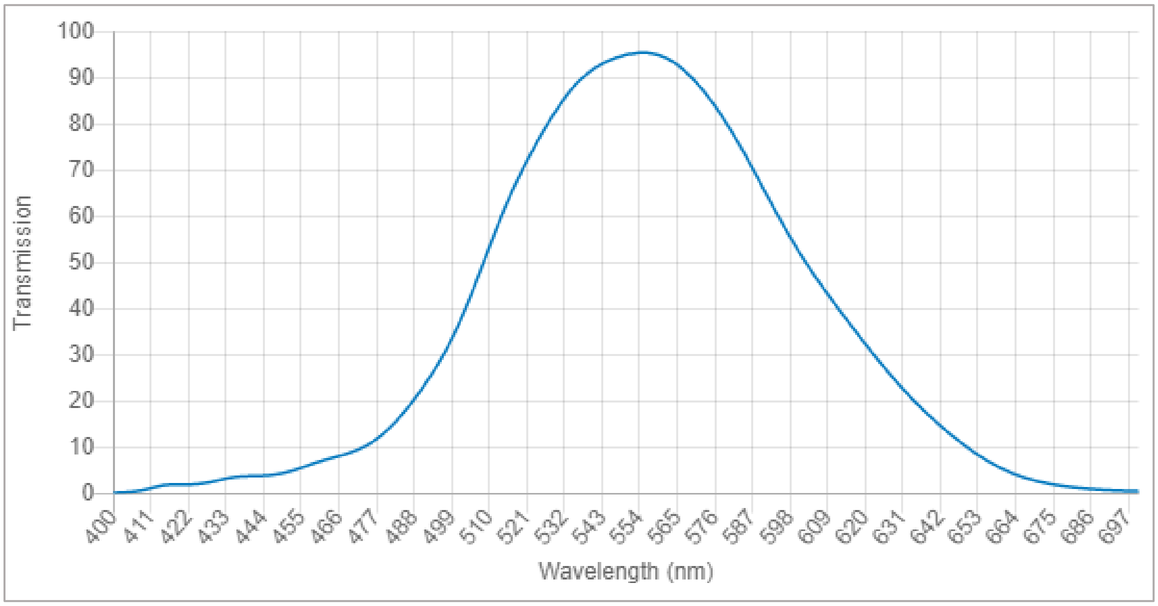
---

[5] PCL Test System Software logs the camera case temperature at 1-second intervals.

[6] https://schneiderkreuznach.com/en/photo-optics/b-w-filters/filtertypes/uv-clear/nd-800-series

of all these components is a closer fit to $V(\lambda)$ than the photopic filter curve by itself, primarily because the camera response mismatch modestly offsets some of the photopic filter mismatch.

Figure 10. Relative transmission for the chosen photopic filter, Omega 558BP100



Camera Calibration

Our camera photometer requires several calibration, configuration, and image processing steps to achieve accurate, repeatable measurements. We perform initial calibrations, to be updated periodically (e.g., annually), in our lab. Other steps must be performed for each TV tested.

Initial/Annual Calibration

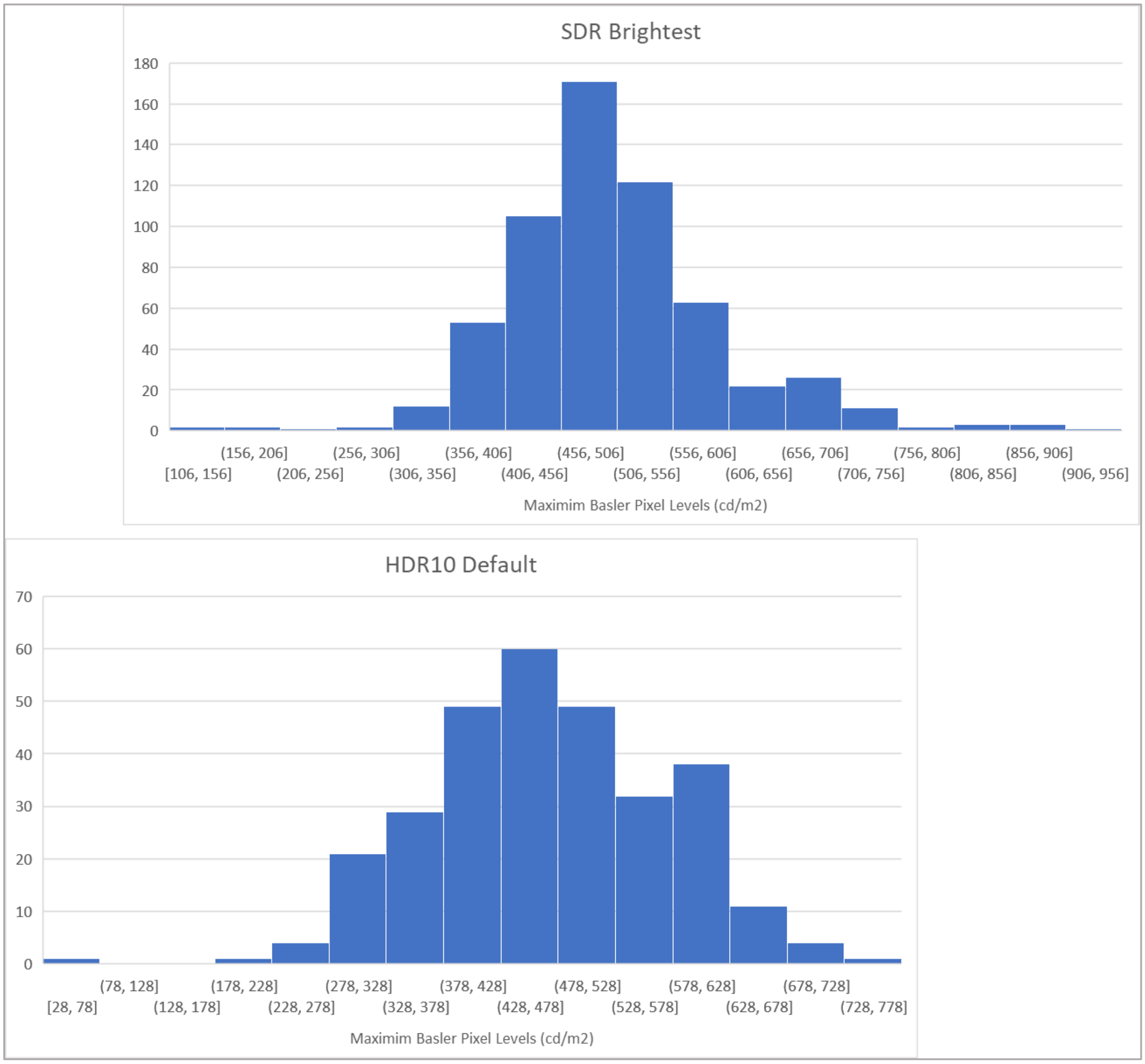
Aperture Setting

As mentioned above, we set the lens aperture so that we achieve an approximately 2:1 ratio between Basler signal level and luminance (cd/m^2) as measured by our PR650. We then fix the lens in place by tightening a small thumb screw in the lens and by applying nail polish to help fix the position and to make it obvious if the aperture setting has shifted. It is impossible to set the aperture to achieve exactly a 2:1 ratio; for example, achieving a 2:1 ratio setting the aperture using one TV might yield a 1.94:1 ratio for a TV with a significantly different spectral power distribution. This step is intended to set the approximate ratio between signal level and luminance. As discussed below, we perform a more precise TV light level calibration for each TV to achieve the needed accuracy level.

Setting the aperture this way maximizes accuracy at low luminance levels while avoiding sensor pixel saturation for even today's brightest TVs. We achieved significantly better accuracy at low light levels with the aperture set to 2:1 signal to cd/m^2 ratio than we did with our initial setting of 1:1 because a 2:1 ratio provides twice the measurement granularity, which is important at low luminance levels in particular.

The 2:1 setting allows us to effectively measure the full range of luminance values we expect to see with today's brightest TVs while allowing flexibility for even brighter TVs in the future. Because our camera supports a 12-bit dynamic range and is set to a master black level adjustment of 80, a 2:1 ratio allows us to measure up to $(2^{12}-80)/2 = 2{,}008\ cd/m^2$, more than twice the maximum camera pixel level we saw for the brightest TV we have tested (advertised peak brightness of 4,000 nits). For this TV, we measured a maximum camera pixel brightness of 945 cd/m^2 when playing either the SDR dynamic test clip in the brightest setting or the HDR10 test clip in the TV's default setting.[7] The histograms below evidence the fact that our camera settings provide sufficient headroom for the foreseeable future. The two histograms are offset to approximately align the bins.

Figure 11: Distribution of Maximum Pixel Brightness Across Multiple Picture Modes



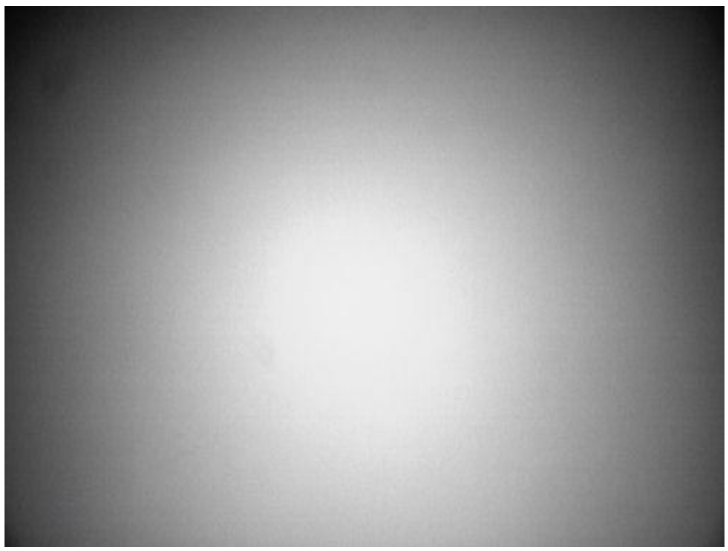
[7] PCL Test System Software logs both average and maximum camera pixel luminance values at one-second intervals.

Vignette Effect Correction (Flat Field Correction)

The purpose of the calibration is to correct for the decreasing brightness of pixels farther away from the center of the image due to optical effects of the lens and camera. This calibration is conducted once per camera per calibration period. The vignette effect is separate from the “viewing angle” effect described in a previous section. This system does not compensate for the “viewing angle effect” for the camera because the effect is also observed by a human viewer. The vignette effect is corrected because it is unique to the camera optics.

An example of the vignette effect is shown here:

Figure 12: Typical Vignette Effect

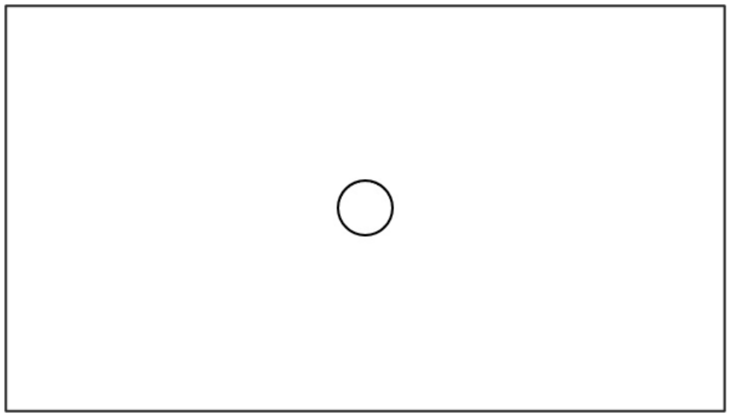


The calibration procedure is as follows:

1. Assemble the camera, lens, and neutral density and photopic filters. Set the camera to the exposure time and aperture settings we are using for our tests.
2. Display the vignette calibration image on the TV (a uniform white image with a black circular outline in the center).[8]

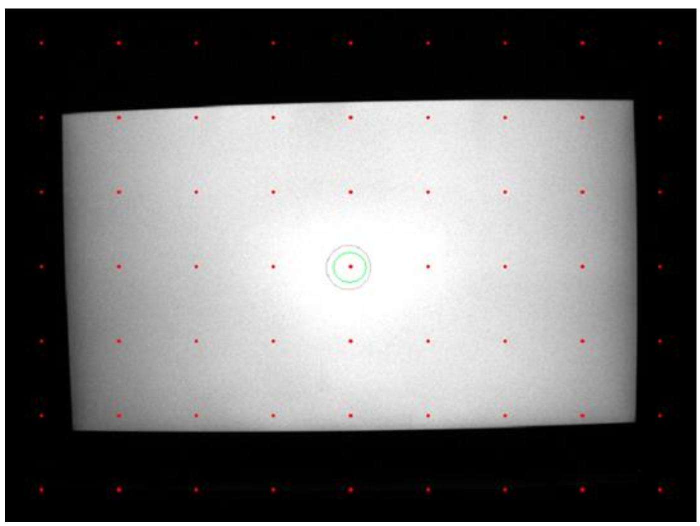
[8] We later moved to a black background outside the grey circle to reduce the risk that glare affects our calibration.

Figure 13: Circle.mp4 Pattern



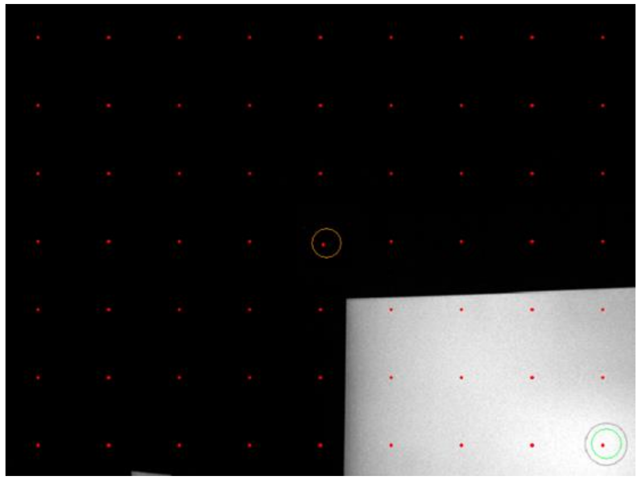
3. Using a special luminance calibration tool, take a luminance measurement of the marked circular area, with the circle lined up in approximately the center of the image. The image below reflects the view of the TV from the calibration tool.

Figure 14: Vignette Calibration Tool, Centered



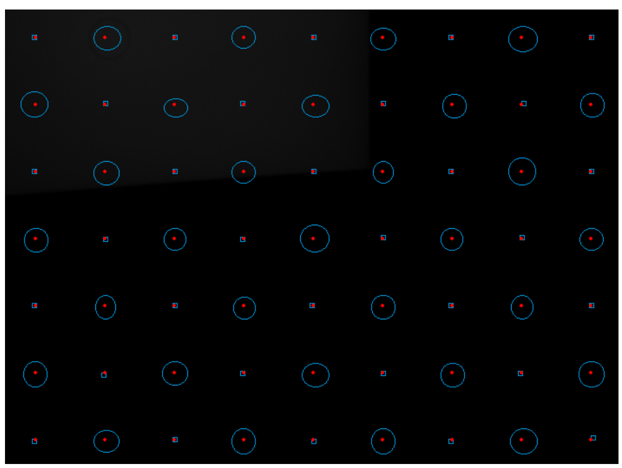
4. Rotate the camera about its center of focus so that the green circle, identifying the location of the next measurement, is located in a different part of the image.

Figure 15: Vignette Calibration Tool, Cornered



5. Take another brightness measurement.
6. Repeat Steps 4-5 until measurements across the entire image area are taken. A grid of small dots in the calibration software shows the intended array of measurements to take. The following image shows a completed vignette with only half the points measured; in practice, we measure all the available grid points when completing a vignette calibration.

Figure 16: Vignette Calibration Tool: Half of Measurements Taken



The calibration software will interpolate between all the measurements and create a vignette correction image that is applied to images during the test. It makes a three-dimensional quartic fit using the least-squares method, with the x and y coordinates of the measurement as the independent variables, and the measured brightness as the dependent variable. A flat field image is created with this quartic regression.

Figure 17: Vignette Tool Output

The test system uses this image to correct for the vignette effect by applying the following equation to the pixel values of acquired images:

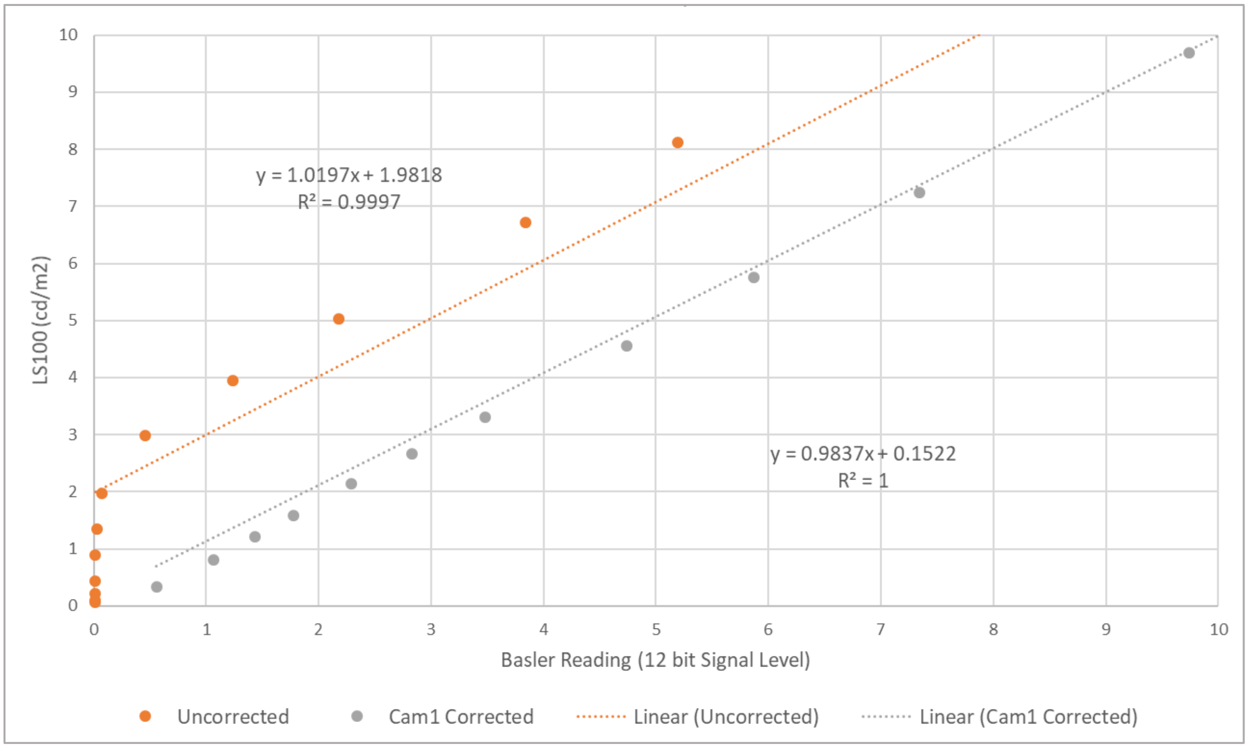
$$Corrected\ Image = Original\ Image \times \frac{Maximum\ Value\ of\ Calibration\ Image}{Calibration\ Image}$$

We observed with the current lens that the vignette effect differs in measured luminance by no more than 0.2% across all possible levels of focus, and as such decided to do a single vignette calibration for each camera. This is possible because with a wider lens, just the center (the least curved area) of the lens transmits light to the camera sensor, and the more curved parts of the lens where vignetting would be more significant are discarded.

Master Black Level

At the low end of the camera dynamic range (pixel brightness values below 5), there is a nonlinearity observed for the relationship between the camera signal level and the actual measured luminance from the reference light meter. This is due to “black crush” (the camera compresses brightness values at the low end of the range). To compensate for this, the camera’s black level setting (which is 0 by default) is set to a positive number, in this case 80. The camera sensor increases the signal of all pixels by the black level setting. This results in more headroom on the low end of brightness, which mitigates the “black crush” effect and increases linearity of image brightness and luminance.

Figure 18. Corrected with Black Level, vs. Uncorrected



Dark Field Correction

Dark field calibration is used to correct for image noise (dark current and fixed-pattern noise). The dark field image is obtained by taking a picture with the camera lens completely obscured. The camera is set to the same settings that are used for testing, including the black level. The brightness level of the dark field image is offset by the black level setting by the same amount for every image acquired during a test.

To apply the correction, this dark field image is subtracted from every acquired image from the camera. The brightness offset from the black level setting is also compensated for with this correction.

$$Corrected\ Image = Acquired\ Image - Dark\ Field\ Image$$

Upon request, we can provide detailed pixel-level maps that show the outcomes/consequences of each of the corrections discussed above.

As previously noted in Aperture Setting, the master black level of 80 brings the maximum signal down from 4096 to 4016; with the aperture set to read a 2:1 signal to luminance ratio, this translates to an expected cap of 2008 cd/m^2.

Per-TV Image Processing and Correction Factors

The steps below are performed as an integral part of each TV test.

Screen Detection

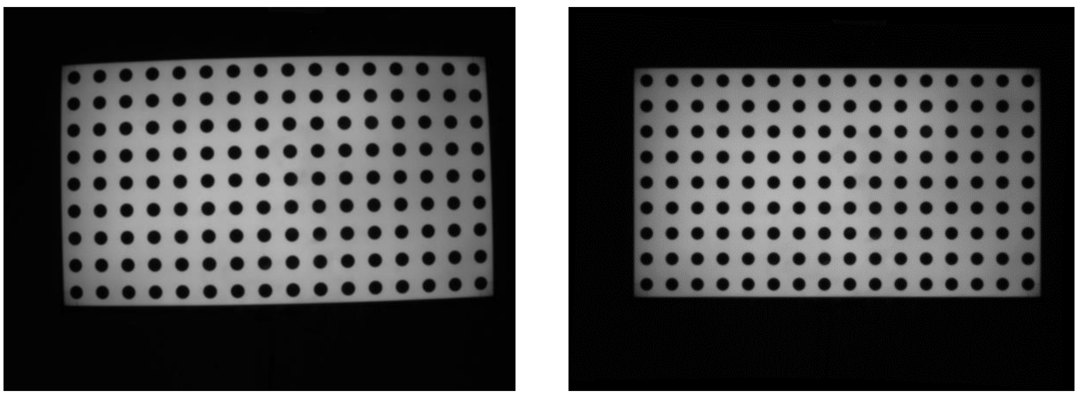
At the beginning of a TV test, immediately following the distortion and perspective correction, the test software detects the border of the TV screen by using a particle detection algorithm that detects the bright image against the dark surrounding environment. The pixels on the edge of this border sometimes overlap both the TV image and the dark environment. The edge pixels could have a small effect on the readings, so the rectangle region of the screen is reduced by one pixel in each direction. The luminance readings that are calculated for the rest of the test will now only use the portion of the camera image that is within the screen border.

Distortion and Perspective Calibration

The purpose of this calibration is to correct for spatial distortion of the wide-angle lens. This calibration is conducted once per test, at the beginning of the test.

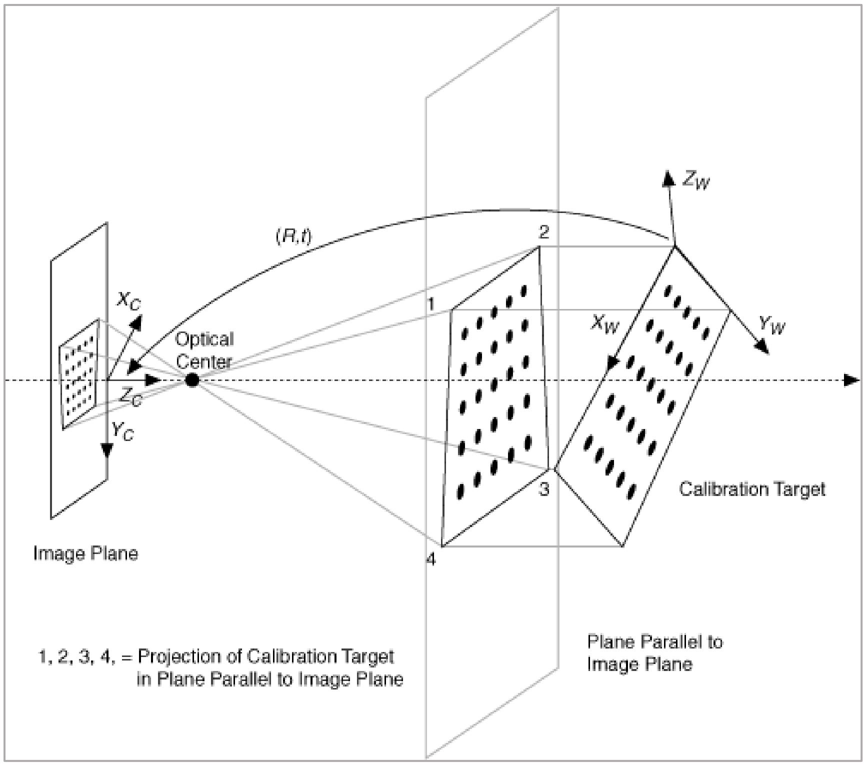
An example of the before-and-after correction for distortion and perspective:

Figure 19: Before-and-After Distortion Correction and Edge Detection



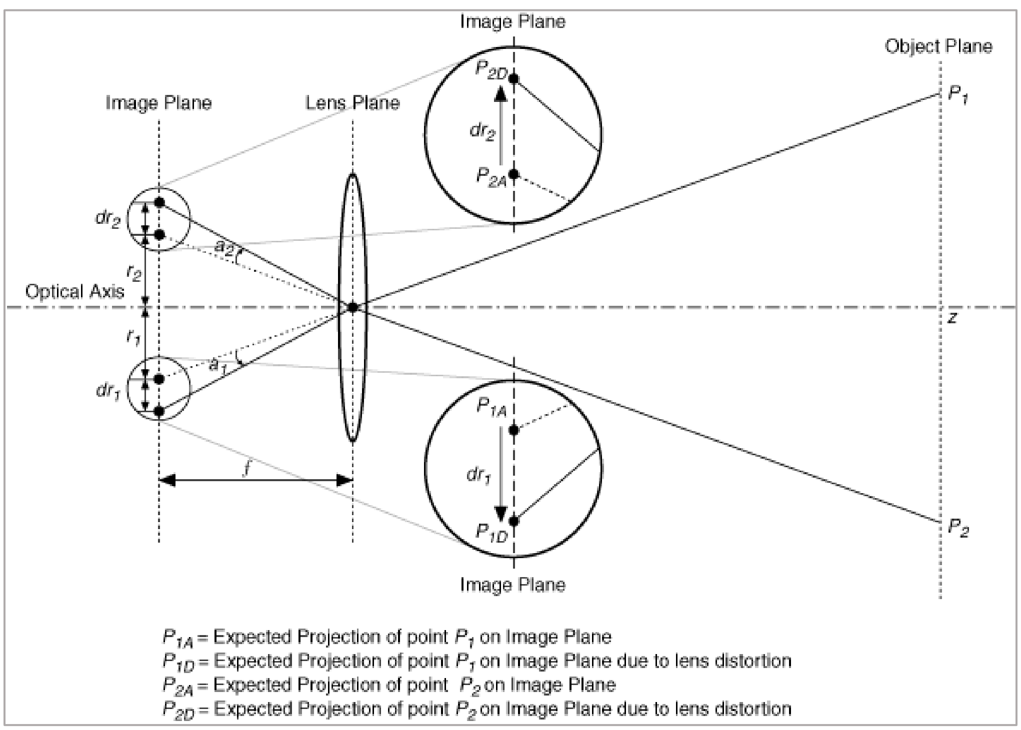
At the beginning of the test, a rectangular grid of dots is displayed. The test system software identifies the position of the dots on the screen and develops a distortion model based on the positions. To correct for perspective, a transformation matrix is created that projects the tilted plane of the original image to a plane parallel to the image sensor.

Figure 20: Perspective Calibration



To correct for lens and camera distortion, the software estimates a distortion model based on the following geometery:

Figure 21: Lens and Camera Distortion Correction



The model equation is: $x_{corrected} = \frac{2x}{1+\sqrt{1-4K(x^2+y^2)}}$

The coefficient K is estimated based on the position of the dots. The inverse equation is used with the estimated K to correct subsequent images. A correction based on these models is applied to images from the rest of the test. Further information about the vision software package used can be found on the National Instruments website.

Luminance Calibration

Rather than calibrate our cameras against a standard illuminant (e.g., illuminant A), we calibrate the luminance of our cameras for each TV, a practice that ensures accurate results across the range of spectral power distributions seen in today's LED and OLED TVs.

Color Correction Factors (CCF)

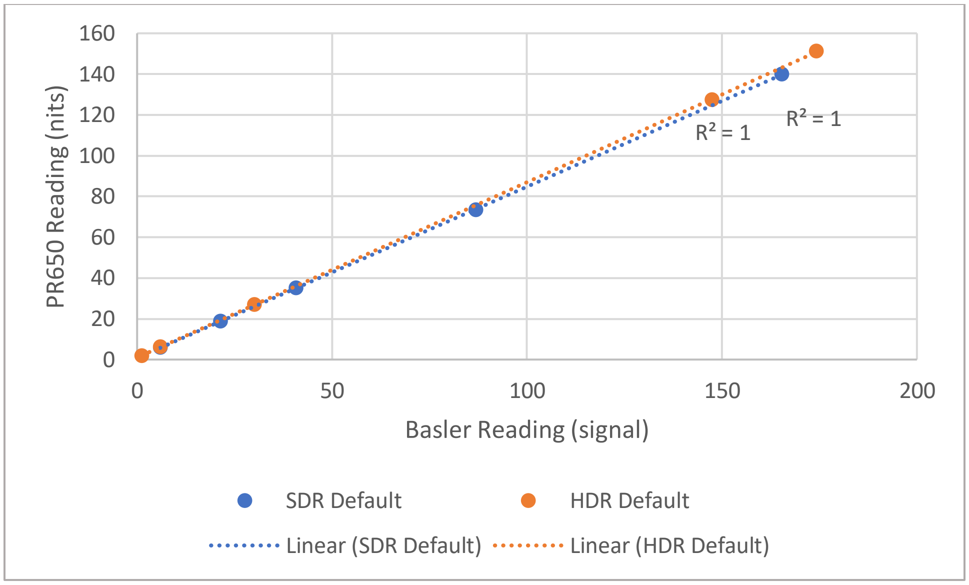
We calibrate the camera photometer against an individual TV's spectral power distribution by taking a range of greyscale values with the camera photometer and the calibrated reference luminance device (in our case, a PR650), and performing a linear fit minimizing relative difference error (Mean Square Percentage Error) between the two. For SDR, we use signal levels of 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, and 90%. We play these nine frames continuously for 35 seconds each with a 5-second black frame in between, taking a measurement at 15s, after which most TVs are observed to have stabilized, based on our research.

This enables us to make luminance measurements specific to each TV tested that are NIST-traceable to a calibrated spot photometer, in our case a PR650 spectroradiometer.[9] The 20s of measurement time, after the intended measurement trigger time at 15s into the color portion of the frame, allows for the long exposure time on the PR650, which uses adaptive exposure time to accommodate for low-light measurements. The following figure shows the linearity of one calculated adjustment; high linearity here suggests that the timing on our calibration process ensures that the PR650 and Basler are measuring the same light output during the calibration clips.[10]

[9] This type of light-source-specific calibration is often called a color correction factor (CCF). For more background on color correction factors, see this presentation: https://www.slideshare.net/theilp/pls-2014-is-measuring-led-illuminance-with-a-lux-meter-accurate

[10] Note that we made these measurements before opening the aperture to achieve a 2:1 signal to cd/m^2 ratio per the earlier section on Aperture Setting.

Figure 22. PR650: Basler Linear Fit (Calibration)



Simultaneous Measurements

While refining the camera photometer system and test process, we found that it is not accurate enough to take the readings with the camera photometer and the reference luminance device asynchronously. Repeatability testing across several TVs revealed variance of up to 10%, run-to-run, independent of the measuring device (see Appendix B: Characterizing TV Stability). Intuitively, if there is a difference in actual light output between two runs, the possible calibrations can vary by as much as the variance of the TV, leading to lower accuracy in the system. As an example, data from several runs on a particular TV is compiled in Appendix C: Effect of TV Instability on Calibration. The expected accuracy in the calibration at a given luminance level is proportional to the error displayed “run-to-run” during the calibration process.

Unfortunately, the reference luminance device cannot sit directly behind the camera photometer and simultaneously take measurements of the exact same light—the field of view of the reference device is obstructed by the camera photometer. Instead, we perform the following process:

Set the reference device directly behind the camera photometer, with its measurement area centered on the following video signal (the same signal that we use for vignette calibration):

Figure 23: Circle.mp4 Pattern

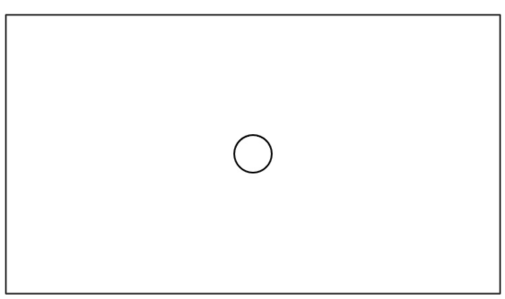
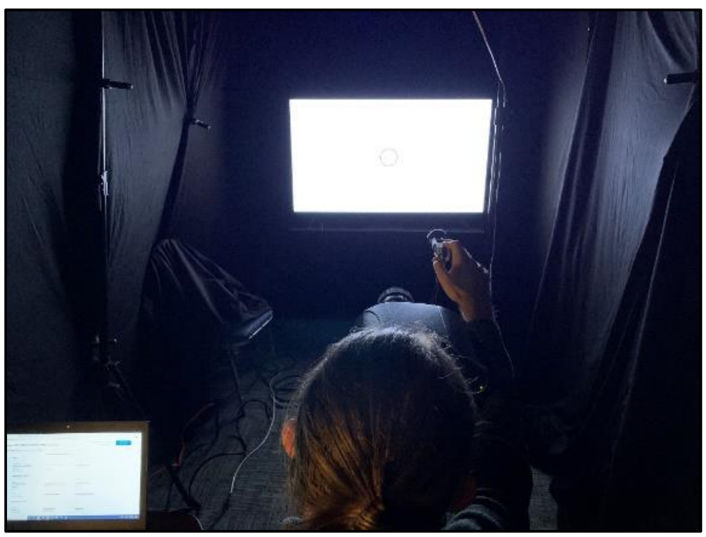


Figure 24. Spot Photometer in Line with Camera

1. Tilt the camera photometer out of the way, then take a reading with the reference luminance device at least 10 seconds after the clip starts.

Figure 25. Luminance Measurement Taken Head-on with Camera Shifted



2. Without changing the clip, and within one minute of the clip starting, adjust the position of the reference device to be as close to parallel with the camera photometer as possible, but at a slight angle to measure the same area of the screen. Note: both the head-on and off-angle measurements should be taken between 10 seconds and a minute after the clip starts, to avoid the effects of initial luminance spikes as discussed later in Temporal Error. This also avoids any Automatic Brightness Limiting features present on OLEDs and other TVs sensitive to static pattern burn-in.
3. Take a second measurement.

Figure 26. Spot Photometer at an Angle Offset from the Camera

4. Use the ratio between the two measurements as a correction for the angular attenuation of the light coming off the TV, then take measurements on the color correction factor clips simultaneously with the camera photometer and the reference device.

Using a Single Preset Picture Setting for Luminance Calibration

We have observed for the small subset of TVs we have tested so far that the calibrations calculated for the camera do not vary significantly (slope typically varies <1%) across picture settings. This makes sense intuitively; unless the actual spectral light profile of the TV changes between picture settings or between SDR and HDR, the needed calibration should not change.

Using a CCF Lookup Table for Known TV Panel Types.

It is possible that a larger data set of TVs may reveal that we can do a single color correction factor for a given technology type (LED vs. OLED vs. QLED). If this proves to be the case in the future, PCL will include calibrations for each TV panel type with the other camera calibration documents, and testers will select from that lookup table instead of performing the color correction factor process themselves. This will have the advantage of reducing possible user error during the color correction step, and removing the requirement for the test lab to have an accurate photometer on hand.

Error Analysis

Below we first present a breakdown of the error of our camera photometer system as used for TV testing. This error includes test process errors as well as camera-specific errors as evident in the overview table below. We then compare the accuracy of our proposed approach (< +/-5.0%) to that of the current approach to measuring TV luminance.

Error Breakdown (Camera Photometer)

The table below shows the error breakdown of our proposed camera photometer approach.

Table 2: Error Breakdown (Camera Photometer)

Error Source	Proposed Camera Photometer
Spot luminance device accuracy	+/-2% (Specified accuracy of a PR650, our reference spot photometer)
Spatial luminance measurement	+/-1% (Basler-based camera photometer method spatial accuracy relative to a Radiant ProMetric Y29, our reference camera photometer)
Observed spot measurement accuracy (24 color pattern)	+/-2% (observed accuracy with method improvements: combination of filter match, TV stability, test timing)
Worst Case Total Accuracy	< +/-5%
Expected Repeatability: Observed device-to-device camera photometer variation for dynamic test clips on stable TVs (variations across lots of Basler camera sensors and lenses, filters)	+/-1%

Below, we explain the method used to determine the accuracy figures shown in the above table:

- Assessing reference spot photometer accuracy.
- Determining spatial accuracy by comparing whole-sensor readings from our camera photometer against those of a Radiant ProMetric Y29 camera photometer using a white pattern.
- Determining experimental error due to TV stability, test timing, and filter match with a spot reading test, using static clips with colors representative of those found in the dynamic test clips.

Spot Luminance Device

The PR650 has a specified accuracy within 2%, calibrated against illuminant A. Based on a key difference in how a spectroradiometer measures luminance vs. how a filter-based device such as our camera photometer or an LS-150 measures luminance, the experts we consulted concluded that for the purpose of accurately measuring LED light sources with differen spectral power distributions, a calibrated spectroradiometer is the appropriate reference photometer. A spectroradiometer refracts incoming light across an array of sensors, giving granular data at small wavelength increments (3.5nm for the PR650, 3.12nm for the newer PR655); luminance is then calculated directly from that data.

Spatial Error

For our camera photometer system, we measure the luminance of the entire screen and take the average luminance over all frames in a dynamic video. Clearly, this accounts for spatial luminance distribution across the TV screen. We determined our spatial error by comparing our camera to a calibrated Radiant ProMetric Y29, which has highly accurate vignette calibration and spatial corrections. We took Basler luminance profiles (recorded all Basler luminance values for pixels focused on the TV screen) and compared them to similar readings taken by the Y29 on the same TV.

To simplify the calculation (the two devices take measurements at different resolutions), we down-sampled the luminance measurements to a 7x9 grid, the same resolution at which we perform our vignette calibration. This table represents the ratio of Basler:Y29 measured screen average for that region of the screen, normalized at the center, where we take spot measurements.

Table 3: Deviation from Known Accurate Spatial Measuring Tool

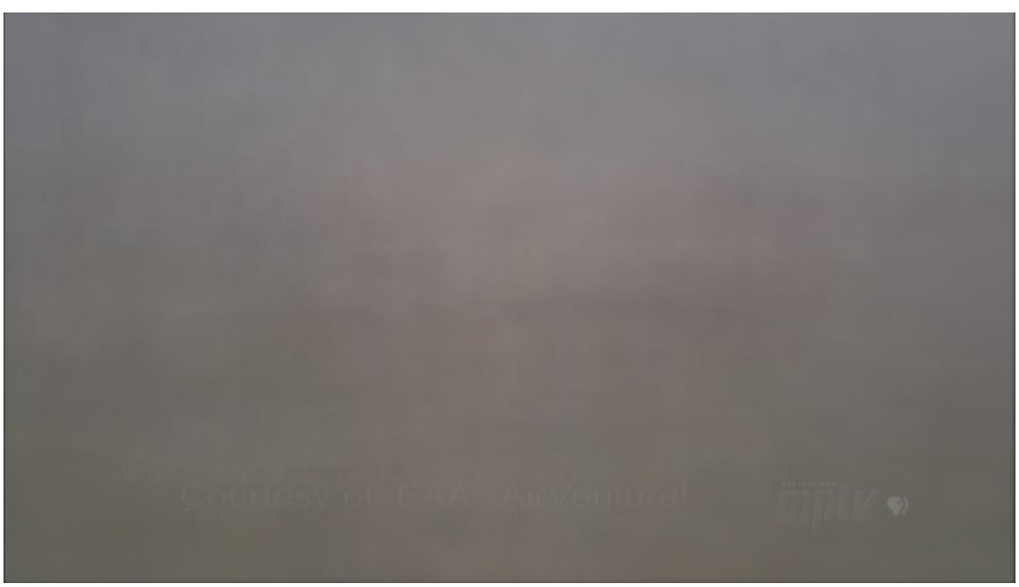
(X,Y)	-4	-3	-2	-1	0	1	2	3	4
-3	1.005	0.985	0.976	0.975	0.976	0.981	0.990	0.993	1.001
-2	1.009	1.002	0.994	0.995	0.991	0.995	0.995	1.002	1.001
-1	1.006	1.000	0.991	0.992	0.992	0.994	0.996	1.001	0.999
0	1.010	0.999	0.993	0.995	1.000	0.998	1.000	1.000	1.000
-1	1.006	0.996	0.987	0.990	0.997	0.996	0.998	1.000	0.998
-2	**1.011**	0.996	0.982	0.990	0.991	0.995	0.990	0.990	0.992
-3	**1.011**	0.988	0.967	**0.963**	**0.958**	0.964	0.967	0.972	0.989

Note the bolded values as the largest deviation from 1.0.

The farther the Basler measurement deviates from that of the Y29, the greater the potential spatial error. Though the relative magnitude of the difference is likely amplified due to small actual readings on the edges of the screen (where it reaches up to 5% difference), we further characterize the extent to which this contributes to error when measuring the dynamic test clip.

Fortunately, we can generate an overall color distribution for the test clip, to see whether, given our knowledge of the spectral response of our camera, spatial differences in color distribution of the clip would be enough to contribute significant error. Calculation shows that this could only account for <1% potential error in the measured screen average luminance, for the dynamic test clips; the overall distribution of the test clip color is relatively uniform, and spatial error, usually due to imprecise placement of the camera, is negligible. We still include it, however, for our worst-case calculations. For reference, below is the averaged picture of the IEC SDR clip used in testing; it is uniformly grey:

Figure 27: IEC SDR clip, signal averaged



It is noteworthy that we performed the above spatial error calculations before switching to a new larger-diameter Basler lens, which exhibits less vignette effect than the originally selected lens (i.e., the one associated with a test distance of 1 x screen diameter). Therefore, we expect that the spatial error of our current camera photometer system is less than (i.e., bounded by) the above calculations.

Experimental Error

Ideally, we would measure the error of our camera photometer against a known-accurate reference camera capable of measuring screen-average dynamic luminance. However, calibrated camera photometers available in the market today cannot measure luminance fast enough for use measuring dynamic video, and we observed low filter match to the luminosity function. So, we experimentally verify the camera system luminance error (minus spatial error) by calibrating against static color patterns. We randomly selected 24 solid colors from the SDR and HDR clips to represent the range of hues, saturation levels, and brightness in the clips themselves.

Figure 28. SDR Frame Colors

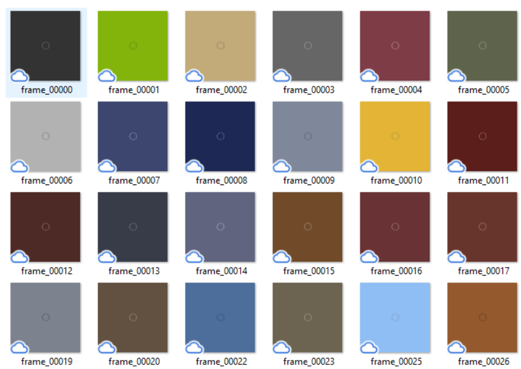
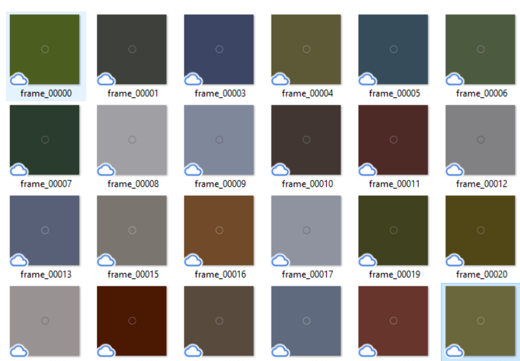


Figure 29. HDR Frame Colors



We measure each frame with both the Basler and the PR650, inside the circle (visible in the center of each frame); the PR650 is only capable of measuring spot luminance, so this simulates measuring the exact same area. We take an average of all 24 readings to simulate averaging the readings across an entire test clip. In our final check, we reduced the cases to just the default SDR and HDR preset picture settings to save time; we did not initially see a remarkable difference in error in the brightest mode vs. the default mode in SDR. We ended up with the following adjusted measurements, representing an experimentally bound accuracy of < +/-2% to the PR650. Since we know the PR650 has a near-perfect fit to the photopic curve, and claims 2% accuracy to luminance measurements, this gives us confidence that experimentally, we are at least 4% accurate in the *worst* case, and strong confidence in our method of calibrating the camera photometer to a white screen, per TV.

Table 4: Camera Accuracy to Measured Color Frame Luminance

	Default SDR average luminance ($\frac{cd}{m^2}$)			Default HDR average luminance ($\frac{cd}{m^2}$)		
TV	PR650	Basler	Accuracy	PR650	Basler	Accuracy
LCD 1	73.86	73.19	-0.9%	70.08	70.03	-0.1%
LCD 2	58.55	58.54	0.0%	43.36	43.45	0.2%
LCD 3	83.77	83.53	-0.3%	75.20	75.39	0.3%
LCD 4	61.20	61.38	0.3%	55.53	55.62	0.2%
LCD 5	46.04	46.10	0.1%	28.56	28.73	0.6%
QLED 1	83.67	83.03	-0.8%	46.93	47.06	0.3%
OLED 1	51.01	51.95	1.8%	50.02	50.73	1.4%
OLED 2	33.84	33.65	-0.6%	28.24	28.15	-0.3%

Though most of the TVs in this set tested under 1.0% error when compared to the PR650, we did observe up to 1.8% error on one of the OLED TVs. OLEDs have the most theoretical error from the filter match; when calibrating against white content on an OLED, we expect the calibration to potentially be off by up to 0.5%: see Appendix A: Specific Mismatch to TV backlights. OLED TVs also have Automatic Brightness Limiting, and we have observed less-predictable stability characteristics on OLED TVs than typical LED TVs, so it is unsurprising that one of our OLEDs shows the largest experimental error, 1.4–1.8%. We attribute most of this observed error to factors such as TV stability and test timing.

Process Improvements

Reducing the measured error to this level was non-trivial and required several considerations in the test setup. These were:

- Observing TV stability across our test sample, we noticed that many TVs take up to 10 seconds to plateau in luminance and power after a static pattern is displayed. To ensure that the measurements of the camera photometer and the reference luminance measurement device (LMD) are taken under the same conditions, we added a timer to our color correction clip, and the exposure for the LMD begins at the same period of the clip that the Basler measures. Failure to follow this procedure can lead to non-linearity in the relationship between the Basler readings and those of the LMD, which is attributed to TV stability, rather than to any characteristics (including filter match, black crush, etc.) of the two devices.
- We observed that the OLED screen burn prevention feature, Automatic Screen Brightness Limiting, kicked in when playing the first version of the 24-color test clip, as the ring around the measurement area was detected as a static pattern. To prevent this from affecting test results, we put a 5 second black frame between each color frame to refresh the internal timer of the TV.
- Camera photometer positioning errors relative to the TV can cause significant variation across test runs, particularly for TVs whose luminance varies more widely across angles. We ensured that our camera sensor and reference LMD sensors were positioned within 1cm tolerance of one another, and that that position was centered directly perpendicular to the center of the TV.

Failure to follow this procedure can lead to non-linearity in the relationship between the Basler readings and those of the LMD. We carefully adhered to the procedure by

  - ensuring the TV was parallel to the back wall of the test lab within 1cm, and
  - measuring from the center of the TV to the side wall and floor and ensuring that the Basler camera was centered exactly.

- We ensured that temperature did not vary by more than 2 degrees Celsius from the start of a test to the end of the test, by controlling air ventilation, and performing all tests sequentially (i.e., not performing the first part of the test with the Basler one day, and the second part with the reference luminance meter the second day). We do not have data to characterize the importance of this precaution, but were careful about it nonetheless.

Error Breakdown (Current Test Method)

Above, we show that our proposed camera photometer approach can achieve worst-case error of < +/-5.0%. Below, we put this value in context with the much larger error associated with today's luminance measurement approach.[11]

Table 5: Worst Case Error

Error Source	Spot Photometer	Camera Photometer
Spot luminance reference device error	**+/- 9%** (observed with LS-100 calibrated against Illuminant A)	**+/-2%** (PR650 calibrated against Illuminant A, but with better filter match)
Spatial error	**+/- 100%** (spot reading can misrepresent screen average luminance)	**+/-1%** (Basler-based camera photometer method spatial accuracy relative to a Radiant ProMetric Y29, our reference camera photometer)
Temporal error	**> +/- 100%** (measurement within first 5s of bright screen can vary widely)	**0%**
Other error	**+/-1 or 2%** (*TV instability*)	**+/-2%** (filter match and/or *TV instability*[12])
Worst case total error	**>> +/- 100%**	**< +/-5.0%**
Expected error	**High**	**+/- 1.0%**

We observed the spot luminance reference device error of +/- 9% when measuring luminance of a grey screen with a Konica-Minolta (KM) LS-100 spot photometer compared to our PR650 spectroradiometer.

[11] See Appendix D: Policy Context on Current Luminance Measurement Approach for discussion of the error associated with current spot luminance policy limits.

[12] *TV instability* is not error, but it's difficult to distinguish between photometer repeatability error and TV instability in the absence of an LED standard illuminant.

One would expect a similar error with the IEC three-bar pattern. Konica Minolta calibrates the spot photometer to standard illuminant A, and TV test labs commonly test LED TVs. Because the LS-100 has an imperfect fit to the CIE 1931 luminosity function (relative to the PR650 or the newer KM LS-150), and LED and OLED TVs have SPDs that differ significantly from illuminant A, we see significant error in the readings. We recommend that test labs perform spot measurements with a more accurate device, which would result in error values in the range of our camera photometer. We also observe that TV stability error (with U.S. federal test method) could be reduced by ensuring that the measurement is taken while the TV is stable (e.g., after initial ramp-up, if applicable, and before automatic brightness limiting, if applicable). We propose this method when determining color correction factors in the new approach, and the current federal test method could be modified by relevant working groups to adopt this approach. However, even with these timing improvements, the current method of using a spot photometer to measure screen-center luminance while playing a static three-bar pattern has the potential to introduce large error in TV luminance measurements due to what we call spatial error.

When the spot photometer method was developed, TVs used CCFL backlights that were relatively constant compared to today's TVs with local dimming. In addition, the spot measurements were used to ensure that the default preset picture setting was at least 65% as bright as the brightest setting to avoid gaming (i.e., a ratio, not an absolute reading). It is our belief that the spot measurement may have been more appropriate for that policy and technology scenario. However, local dimming involves dynamic backlight adjustment, and new policies (e.g., ENERGY STAR v8) set minimum brightness levels which effectively mean that TVs with high spot readings relative to screen-average luminance have an advantage since all certified TVs are required to achieve a minimum spot luminance level. So, we define light delivered to the viewer while playing dynamic video content as the desired metric, and assess up to 100% error to the current static-pattern, spot-measurement method.

Spot Luminance Reference Device Error

The current test method is non-specific about the source of the accuracy calibration for the luminance measuring device; DOE refers to IEC 62087-1 5.1.7, which states "The LMD shall have an accuracy of $\pm$2% $\pm$2 digits of the digitally displayed value or better." However, the method does not dictate to which standard illuminant that accuracy should be measured. Most devices we checked that claim certain accuracy to the photopic curve measure to standard illuminant A, an incandescent, rather than to an LED. It goes without saying that standard illuminant A has a significantly different spectral profile than modern TV light sources. We expect that even the small differences between different LED and OLED TV sources may require individual calibration; clearly, calibrating against an incandescent is inappropriate for all meters except those with the best filter match (e.g., KM LS-150, or KM CA-410 with CA-P427 probe).

A calibrated LS-100 device in our lab read within 2% of the PR650 on an incandescent white light source with spectral power distribution close to illuminant A. Meanwhile, on a QLED displaying a white screen, the instrument read 330 nits, while the PR650 read 361 nits. This amounts to a potential difference in measured luminance of 9% if the device is not calibrated against an LED. This is just a single data point on one of the sharper-peaked spectral power distributions and while it could be worst-case, a TV with sharper peaks in its spectral power distribution could be even worse.

With this in mind, we do not expect every device claiming 2% accuracy to illuminant A to achieve the same accuracy to TV screens. At the very least, we recommend requiring LED-specific calibration of any luminance measuring device, even those with a strong filter match using a similar process to the one described in Luminance Calibration, possibly in addition to some requirement for general spectral mismatch index f_1'. We propose the following test spec language to relevant standards bodies:

> The spot photometer shall have an accuracy of ± 2 % ± 2 digits of the digitally displayed value or better. If the luminance measuring device is neither a spectroradiometer nor calibrated against an illuminant replicating the spectral emission of LEDs, and relies on filters to match the CIE 1931 luminosity function, it shall additionally have a <2% spectral mismatch index f1' against its calibration illuminant as defined in ISO/CIE 19476:2014: "Characterization of the performance of illuminance meters and luminance meters."

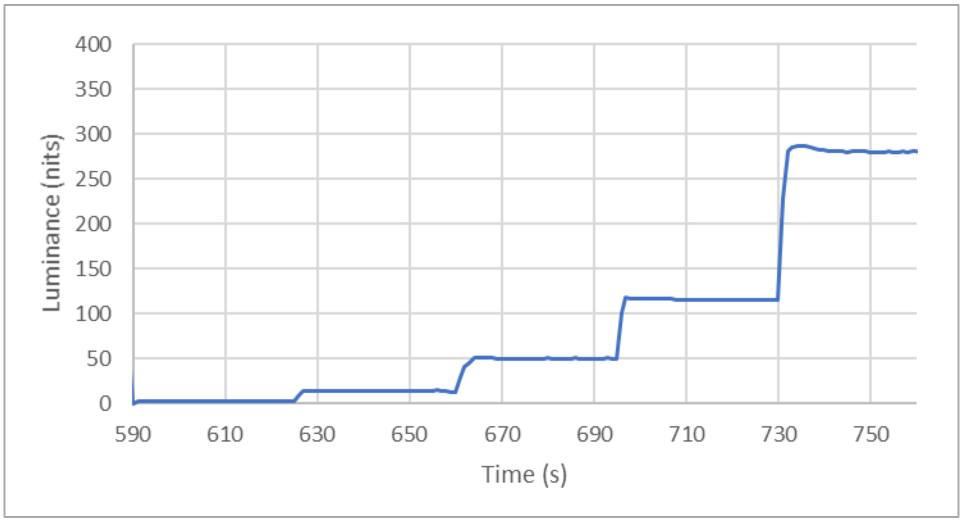
Spatial Error

The current test method measures only at the center of the screen, which does not account for spatial differences in light output of TVs; some TVs, due to the geometry of their backlight or other design reasons, end up with significantly lower luminance measured from the viewer's perspective at the edges or corners. Using a single measurement at the center of the screen to characterize average screen luminance necessarily means extrapolating that measurement to the edges of the screen. However, from our testing in fall 2020, we observed that for some TVs, in the brightest picture settings, center screen brightness can be up to 2x (100% error) brighter than the average screen brightness perceived by the viewer. This rewards uneven lighting as perceived by the viewer in the current test method. We expect this error would be reduced if we backed the camera further away from the TV, as is currently being discussed.

Temporal Error

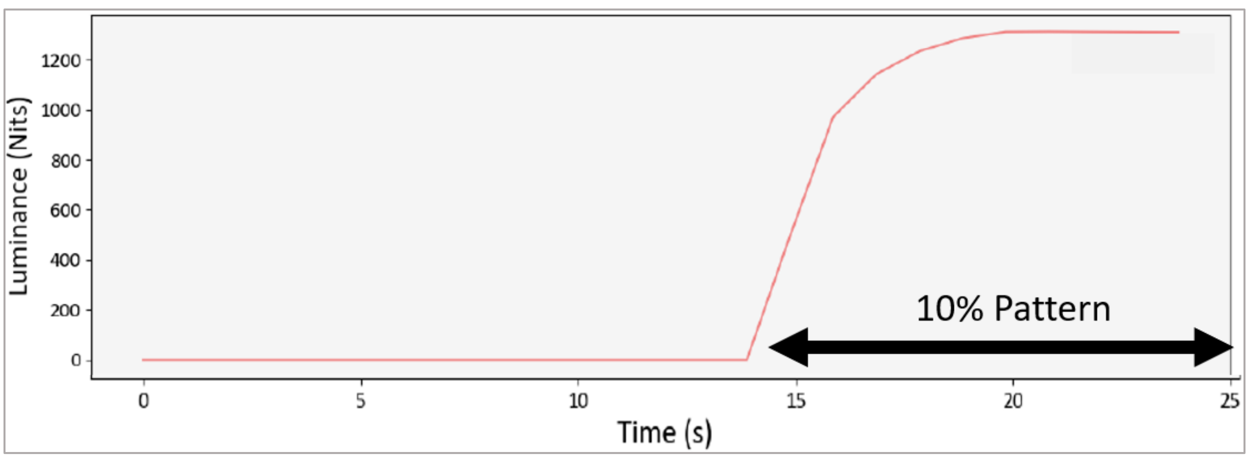
The current method requires the measurement of a single static pattern (the three-bar signal described in IEC 62087-3) within 5 seconds of providing the signal to the TV. Merits of measuring luminance dynamically vs. a single measurement notwithstanding, this method can produce highly variable results, depending on the TV. As part of our investigation into possible sources of error for our method, we observed that some TVs, particularly when the static pattern is bright, spike in brightness within the first 10 seconds of a static pattern appearing, before stabilizing at a normal level. The following figure highlights this behavior, sampled from our testing, representing (in the worst case) a drop from 287 to 280 nits, a difference of 2.5%:

Figure 30. Observed Luminance Over Time on Increasing Signal Level for one TV



In other cases, starting a bright static clip from dark (especially for bright TVs), the TV takes time to ramp up the backlight—for the following TV, a tester measuring even between seconds 1 and 5 of the clip starting could read anywhere from 600 to 1200 nits; at the very least, a test method needs to account for these stability readings by delaying a static reading.

Figure 31: "Backlight Ramp-Up" on a Bright TV



Stability notwithstanding, TVs will dim their backlight locally to save power when dynamic content is playing; this is not reflected by current luminance measurements, as a single static measurement does not reflect the overall brightness of content typically displayed by the TV.

Camera Stability

To verify that the camera system is reliable, we set up two cameras pointing at the same unit, looped a dynamic test clip continuously, and recorded both power and luminance data over a weekend. As the weekend progressed, the test lab cooled off a bit, which for this unit, led to the TV running around a half nit brighter. However, both cameras picked up on this increase proportionally. To determine the expected deviation for the cameras, we can take the ratio of the two luminance readings for the two cameras—one is lower than the other since it was at an angle taking measurements, while the other one was head-on.

Figure 32. Two-Camera Weekend Comparison Run

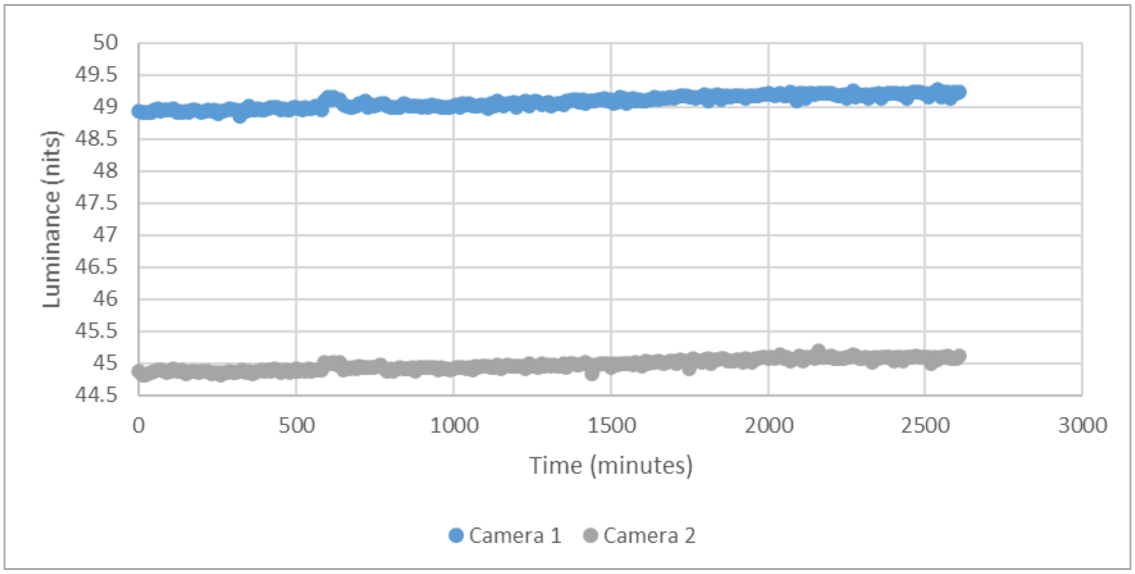
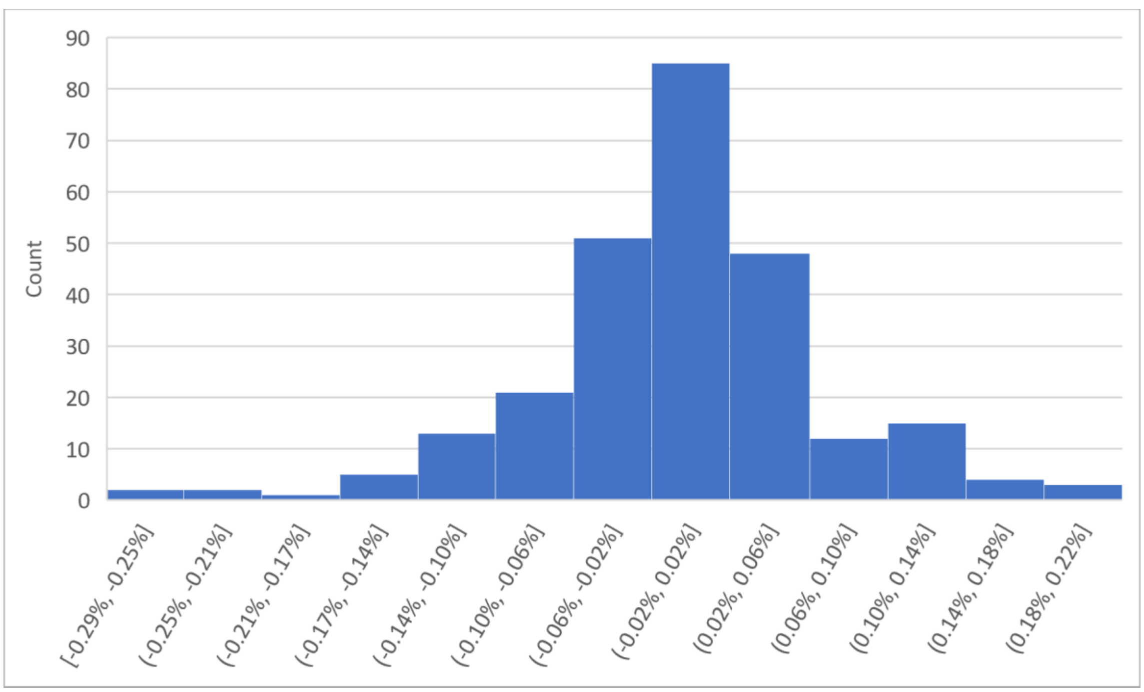


Figure 33. Deviation Between Two Cameras



We calculate the standard deviation of the ratio between the two cameras' measured luminance of the test clip at 0.07%—in other words, the cameras should be considered highly stable, and accurate once a correct calibration is completed.

Further Improvements

Further improvements to the system's fit to the photopic curve $V(\lambda)$ could be made through use of a custom filter that takes into account the spectral response of the camera sensor and transmittance of the lens when correcting for the filter match, improving our camera system's overall fit to the photopic curve. At this time, however, we do not recommend that approach for short-term testing; there is a long lead time on the filters, NRE of about $20,000, and the increase in precision might be incidental. We would like to learn more about the statistical variation in camera/CMOS spectral response curves and to get feedback from TV manufacturers before investing in the development of a custom filter.

Theoretical error could be bound more confidently by running the cameras through an integrating sphere to characterize the spectral response of the camera photometer at higher granularity, as well as giving us an understanding of how widely the spectral response can vary among the individual parts that make up the system. To measure each camera's spectral response curve, we would need an integrating sphere fitted with a monochromator; these systems are in the $80,000 range, which is expensive and not necessary. It may be useful to purchase a much less expensive integrating sphere fitted with an LED standard illuminant in a year or so when they become available.

Appendices

Appendix A: Specific Mismatch to TV backlights

With knowledge of the spectral response of our camera system, and the spectral power distribution of the TV units under test, we can come up with a theoretical spectral mismatch factor based on the equations in ISO/CIE 19476: "Characterization of the performance of illuminance meters and luminance meters" [13]. We can do this calculation for the test clip for a given TV, for the theoretical expected deviation in measured luminance during a dynamic test clip from calibrating against a white screen, if all other sources of error are eliminated (see Experimental Error).

$$s_{rel}(\lambda) = \text{spectral response of camera}$$

$$V(\lambda) = \text{photopic curve}$$

$$S_C(\lambda) = \text{spectral power distribution of calibration source}$$

$$S_M(\lambda) = \text{spectral power distribution of light being measured}$$

$$F^*\big(S_M(\lambda)\big) = \frac{s_C}{s_M} = \left. \frac{\int_{380nm}^{780nm} S_C(\lambda) * s_{rel}(\lambda)\mathrm{d}\lambda}{\int_{380nm}^{780nm} S_C(\lambda) * V(\lambda)\mathrm{d}\lambda} \middle/ \frac{\int_{380nm}^{780nm} S_M(\lambda) * s_{rel}(\lambda)\mathrm{d}\lambda}{\int_{380nm}^{780nm} S_M(\lambda) * V(\lambda)\mathrm{d}\lambda} \right.$$

Table 6. Calculated Spectral Mismatch Correction Factor Against Pure White Calibration

	Pure Red	Pure Green	Pure Blue	**Pure White**	SDR Clip Average	HDR Clip Average
LCD LED#1	1.012	1.000	1.001	**1.0**	1.0026	1.0025
LCD LED#2	1.003	0.999	0.998	**1.0**	1.0000	1.0000
QLED	1.002	0.998	1.011	**1.0**	1.0002	1.0005
OLED	1.013	0.998	1.040	**1.0**	1.0042	1.0048

The sharper wavelengths of the QLED and OLED (see the spectral power distributions in Luminance Calibration) lead to greater expected error, but in general, when we calibrate to white screens, the amount of error theoretically attributed to the filter match is negligible (at most 0.5%) compared to that which can come from simple experimental sources like camera placement, TV stability, etc. We do not calibrate against a calculated 'off-white screen' representative of the clip average, as factors like TV stability, electro-optical transfer functions, and differences in TV technology make the actual average light output hard to predict; using a white screen simplifies the process.

[13] Equations modified from ISO/CIE 19476: "Characterization of the performance of illuminance meters and luminance meters" 5.2.4: "Relative Luminous Responsivity and Spectral Mismatch Correction Factor"

Appendix B: Characterizing TV Stability

In order to isolate error from the camera system and test process, it is necessary to measure accuracy against a TV we consider to be a steady source of light. Unfortunately, not all TVs on the market output the same amount of light across different tests, even if factors like temperature, humidity, and test timing are controlled. In order to assess TV stability and to identify a few stable TVs for use in camera accuracy assessment, we conducted 4 rounds of stability testing in the default picture setting on 13 TVs, models A1-G1 in the table below. For the “original run” we tested all 13 TVs in the default PPS, with ABC and MDD off, first with the 10 min IEC SDR test clip, then with the 5 min IEC test clip. On subsequent days we conducted similar test runs, called runs 2-4 below.

Table 7: Observed Stability on Several TV Models, SDR

	SDR (Orig)	SDR (Run 2)	SDR (Run 3)	SDR (Run 4)	SDR (Orig)	SDR (Run 2)	SDR (Run 3)	SDR (Run 4)	SDR (Orig)	SDR (Run 2)	SDR (Run 3)	SDR (Run 4)	SDR (Orig)	SDR (Run 2)	SDR (Run 3)	SDR (Run 4)
Model	Power (Watts)				Luminance (Nits)				Absolute Deviation from Average Power				Absolute Deviation from Average Luminance			
A1	124.8	125.4	125.3	125.1	72.5	71.9	72.3	71.9	0.3%	0.2%	0.1%	0.0%	0.5%	0.4%	0.2%	0.4%
B1	96.6	96.3	96.4	96.1	53.3	53.7	54.1	51.8	0.2%	0.1%	0.1%	0.2%	0.2%	0.8%	1.7%	2.7%
C1	93.7	94.8	94.6	94.8	60.2	61.3	61.2	60.6	0.8%	0.3%	0.1%	0.4%	1.0%	0.8%	0.5%	0.3%
C2	92.5	93.0	93.1	93.1	72.5	75.5	76.4	76.5	0.5%	0.1%	0.2%	0.2%	3.6%	0.3%	1.5%	1.7%
C3	140.1	140.3	141.0	140.6	82.4	82.1	83.6	83.4	0.3%	0.1%	0.4%	0.0%	0.6%	0.9%	0.9%	0.7%
C4	122.1	122.2	123.3	122.9	49.7	48.6	49.1	48.3	0.5%	0.3%	0.5%	0.3%	1.6%	0.7%	0.4%	1.3%
D2	107.6	108.0	107.8	108.1	54.3	52.4	52.3	53.0	0.2%	0.1%	0.1%	0.2%	2.4%	1.1%	1.3%	0.0%
D3	208.4	209.6	211.0	210.9	102.5	104.7	105.6	105.7	0.8%	0.2%	0.5%	0.4%	2.1%	0.1%	1.0%	1.1%
D4	329.2	331.3	330.3	330.1	69.6	71.9	72.4	72.7	0.3%	0.3%	0.0%	0.0%	2.9%	0.3%	1.1%	1.5%
E1	111.5	112.0	112.8	112.5	48.0	47.7	48.6	48.6	0.6%	0.2%	0.5%	0.3%	0.5%	1.1%	0.7%	0.8%
E3	178.2	178.9	178.7	177.9	44.9	44.4	45.0	44.3	0.1%	0.2%	0.2%	0.3%	0.4%	0.6%	0.9%	0.8%
F1	84.0	83.1	83.0	83.1	48.1	48.9	49.0	49.0	0.8%	0.2%	0.4%	0.3%	1.3%	0.3%	0.5%	0.5%
G1	106.7	107.0	106.5	106.5	41.1	41.5	41.5	42.5	0.0%	0.3%	0.1%	0.2%	1.3%	0.4%	0.3%	2.0%
									Standard Dev =		0.19%		Standard Dev =		0.75%	

Figure 34. TV Stability Playing SDR 10 Min Test Clip, Clustered

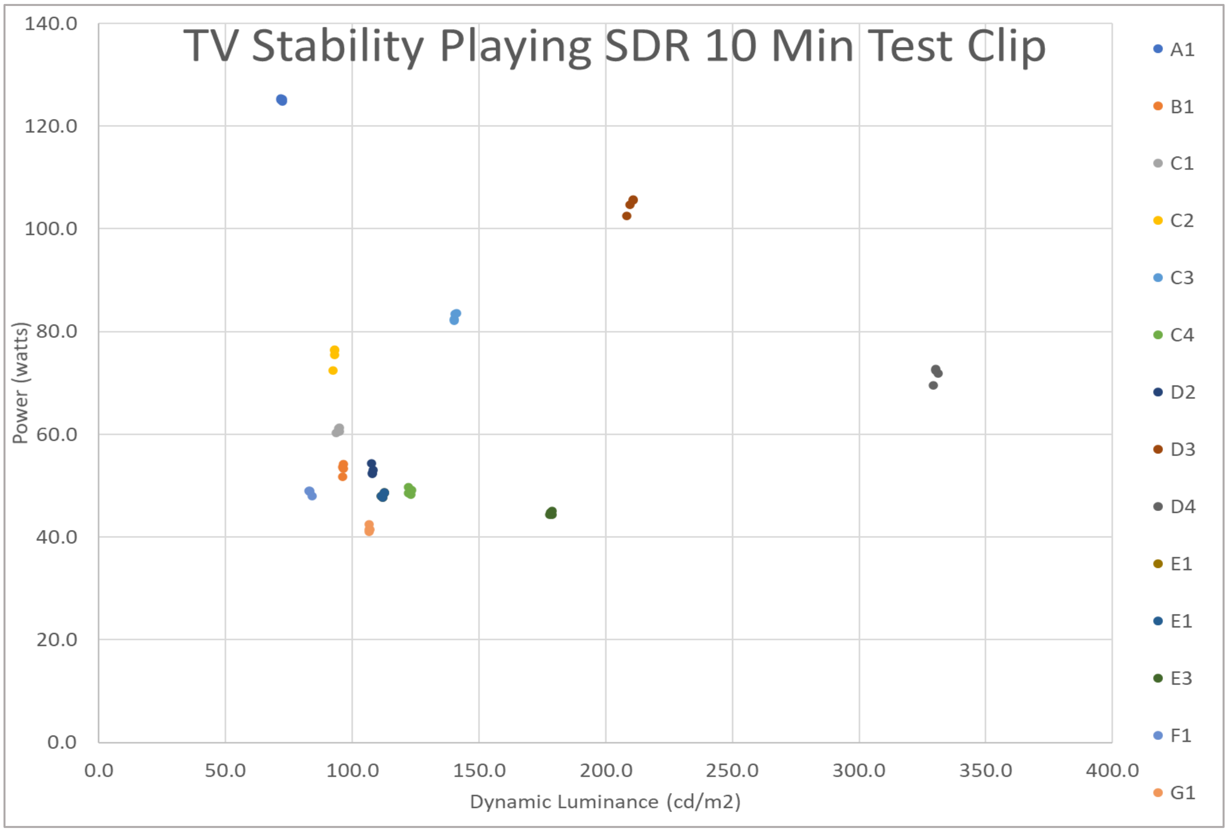
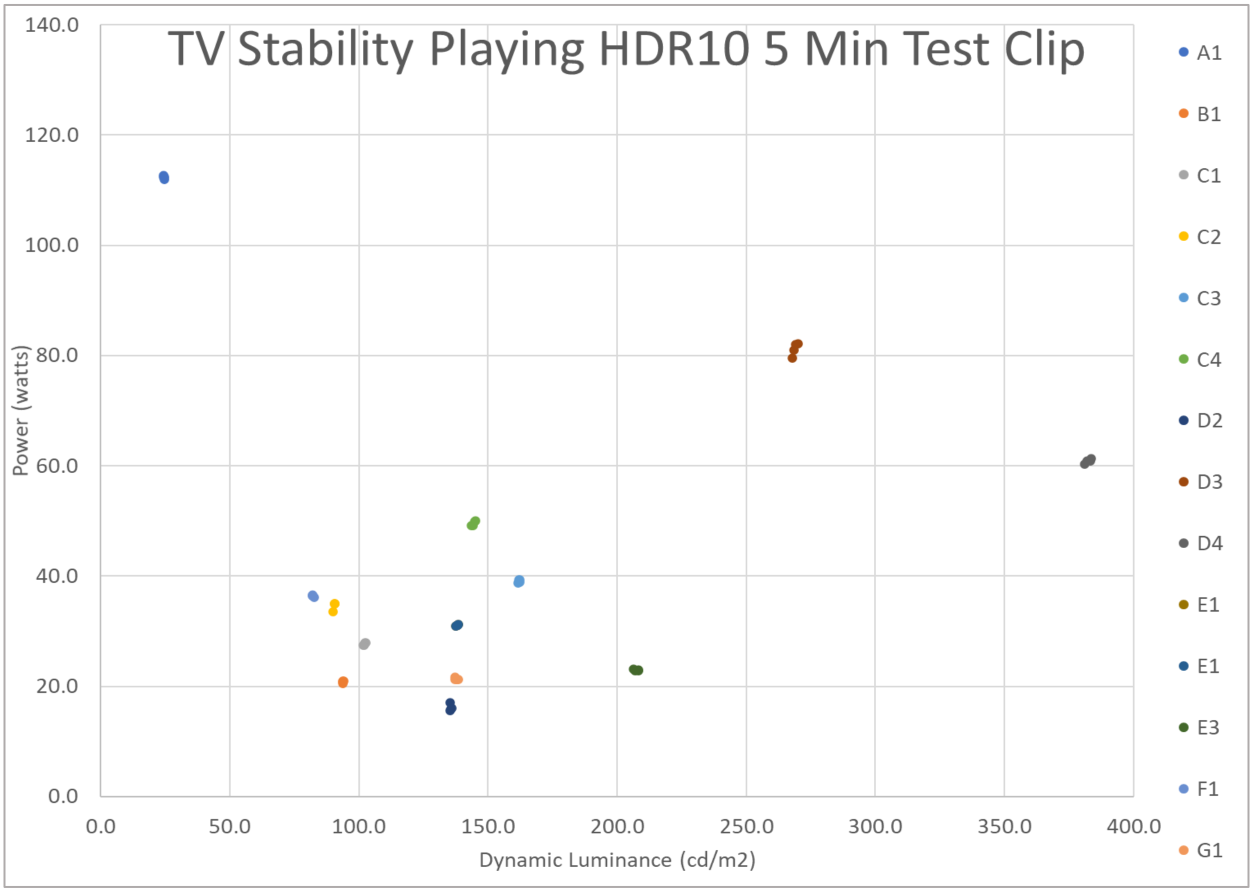


Table 8: Observed Stability on Several TV Models, HDR10

Model	Power (Watts) HDR10 (Orig)	HDR10 (Run 2)	HDR10 (Run 3)	HDR10 (Run4)	Luminance (Nits) HDR10 (Orig)	HDR10 (Run 2)	HDR10 (Run 3)	HDR10 (Run4)	Absolute Deviation from Average Power HDR10 (Orig)	HDR10 (Run 2)	HDR10 (Run 3)	HDR10 (Run4)	Absolute Deviation from Average Luminance HDR10 (Orig)	HDR10 (Run 2)	HDR10 (Run 3)	HDR10 (Run4)
A1	111.9	112.7	112.4	112.4	24.7	24.2	24.5	24.3	0.4%	0.3%	0.0%	0.1%	1.2%	1.1%	0.4%	0.5%
B1	94.2	93.9	93.9	93.8	21.0	20.8	21.0	20.6	0.2%	0.1%	0.1%	0.1%	0.6%	0.0%	0.7%	1.3%
C1	101.6	102.5	102.4	102.1	27.5	27.9	27.9	27.5	0.6%	0.3%	0.3%	0.0%	0.7%	0.7%	0.8%	0.7%
C2	90.1	90.7	90.6	90.6	33.5	35.0	35.0	35.0	0.4%	0.3%	0.1%	0.1%	3.2%	1.1%	1.1%	1.0%
C3	161.5	162.3	162.0	162.1	38.7	38.9	39.4	39.3	0.3%	0.2%	0.0%	0.1%	1.0%	0.4%	0.7%	0.6%
C4	145.3	143.7	144.7	144.1	50.0	49.1	49.9	49.1	0.6%	0.5%	0.2%	0.2%	1.0%	0.9%	0.7%	0.8%
D2	135.2	135.7	135.2	135.9	17.0	15.8	15.6	16.0	0.2%	0.2%	0.2%	0.3%	5.4%	1.7%	3.1%	0.6%
D3	267.9	268.4	269.9	269.2	79.5	81.0	82.1	82.1	0.4%	0.2%	0.4%	0.1%	2.1%	0.3%	1.2%	1.1%
D4	381.1	381.9	383.3	383.6	60.2	60.9	60.9	61.4	0.4%	0.2%	0.2%	0.3%	1.0%	0.0%	0.1%	0.9%
E1	137.7	138.2	138.2	138.4	30.9	31.1	31.0	31.2	0.3%	0.1%	0.0%	0.2%	0.6%	0.0%	0.1%	0.6%
E3	206.4	208.3	208.4	206.9	23.2	22.8	22.9	22.9	0.5%	0.4%	0.4%	0.3%	1.0%	0.7%	0.0%	0.3%
F1	82.6	81.9	81.9	81.9	36.1	36.4	36.4	36.6	0.6%	0.2%	0.2%	0.2%	0.8%	0.2%	0.1%	0.5%
G1	137.8	138.4	137.3	137.4	21.2	21.3	21.2	21.7	0.1%	0.5%	0.3%	0.3%	0.8%	0.2%	0.7%	1.7%
									Standard Dev =	0.16%			Standard Dev =	0.91%		

Figure 35. TV Stability Playing HDR10 5 Min Test Clip, Clustered



The data shows that on average, the deviation is low. However, some TVs are significantly less stable than others, especially when playing HDR, where the worst-case deviation was 5.4% (model D2). We avoid these TVs when conducting camera accuracy checks (e.g., round robin tests) or when performing camera vignette calibrations.

Appendix C: Effect of TV Instability on Calibration

To demonstrate the potential effect of TV instability and inconsistency between even back-to-back runs on the potential calculation of signal to luminance calibration for a given TV, we compare the calibrations done synchronously to those done asynchronously for the same TV and the same camera. This data was taken from the CCF step on a test we repeated across four cameras on a single stable TV.

The following table contains the calculated calibrations for each picture setting using the original camera data for the first run as the x-value in the calibration fit. The y-value is the (synced) reference data for each run, the first being synced with the original run and the other three being from the other, asynchronous runs. Expected CCF error is calculated at a given luminance level

Table 9: Calculated Expected Accuracy if CCF Calibrations are not done Synchronously

	SDR Default		**SDR Brightest**		**HDR Default**	
	Slope	Intercept	Slope	Intercept	Slope	Intercept
Run 1 (synced)	0.5144	0.7848	0.5139	0.7272	0.5122	0.7429
Run 2	0.5204	1.1307	0.5258	0.9313	0.5162	0.8439
Run 3	0.5239	0.5507	0.5268	0.475	0.5122	0.7429
Run 4	0.5155	0.8922	0.5123	0.6972	0.515	0.6906
Expected CCF Accuracy at 10 nits	4.5%		4.5%		1.6%	
Expected CCF Accuracy at 50 nits	1.8%		3.0%		1.0%	
Expected CCF Accuracy at 100 nits	1.6%		2.8%		0.9%	

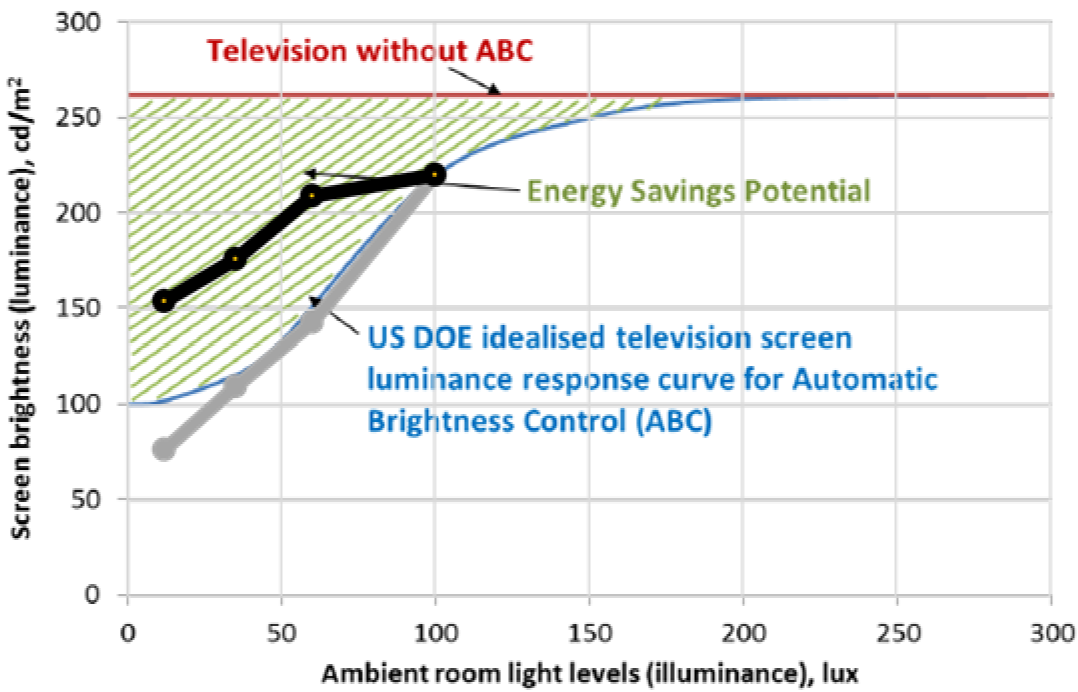
Especially at low light levels, the calibration process is sensitive to differences between runs, and the error introduced will be proportional to the error observed between runs. Full 50 TV testing will give us a full picture of what the range of 'stability' errors can be; however, performing the CCF calibration synchronously prevents this factor from introducing additional error in the test process and camera system.

Appendix D: Policy Context on Current Luminance Measurement Approach

Current US and EU policy use minimum luminance requirements to discourage manufacturers from offering luminance levels in their default setting which are not fit for purpose to score better on the energy efficiency test. We have evaluated the basis for these limits below. We find that there are flaws and uncertainties associated with the basis for these levels. We present this data in support of the proposed camera photometer approach, which can be used to accurately measure how efficiently a TV generates light from the viewer's perspective. In the Error Analysis section of this document, we explain how inaccurate the current spot luminance approach can be; here, we show that the policy basis for spot luminance limits is error prone as well.

A 2012 DOE study[14] documents the rationale behind the idealized luminance curve that Europe used as the basis of the ABC qualification criteria (see Figure below).

Figure 36. Illustration of ABC Limits for Europe Compared to Overall Savings Potential



The DOE study references Matsumoto and others, who used a static pattern to measure centre-of-screen luminance with a photometer. As discussed in prior memos, centre-of-screen measurements are not an accurate representation of how bright a TV is from the viewer's perspective. Screen-average is a better metric that we have not had the tools in the past to measure. And Matsumoto's data is based on a 40% peak pattern similar to the 50% peak pattern shown below (In Matsumoto's case, 40% of the screen area was white).

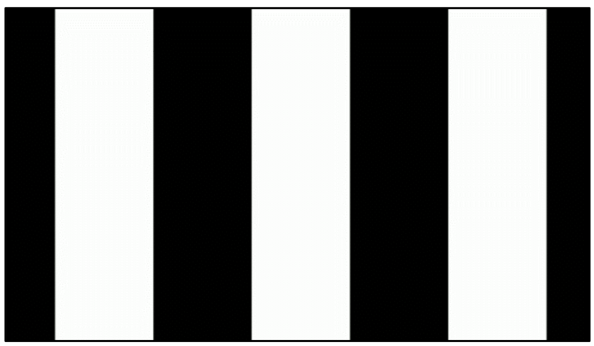
[14] https://www1.eere.energy.gov/buildings/appliance_standards/pdfs/tv_tpnopr_room_illuminance_abc_031912.pdf

Figure 37: 40% Area Peak Window

While Matsumoto used a 40% peak window, DOE and EU use the following patterns:

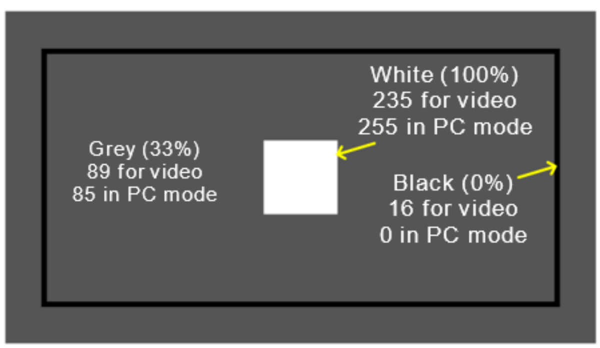
- DOE: three bar video signal (IEC 62087-2:2015, 4.2.2.1)

Figure 38: DOE Three Bar Video Signal



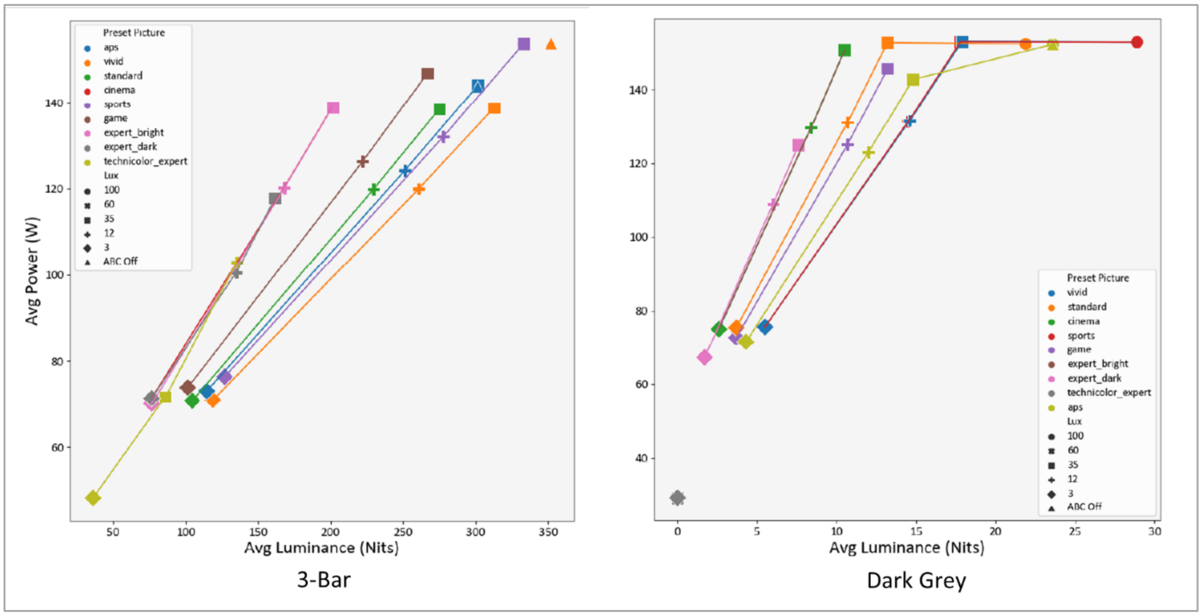
- EC: box and outline video signal (IEC 62087-2:2015, 4.2.2.2).

Figure 39: Box and Outline Video Signal



So, it is likely that the results are not comparable because TVs respond differently to the different patterns. For example, in 2019, NEEA demonstrated that a monochrome 33% grey pattern produced a non-linear relationship between power and luminance for some TVs; whereas the three-bar pattern did not.

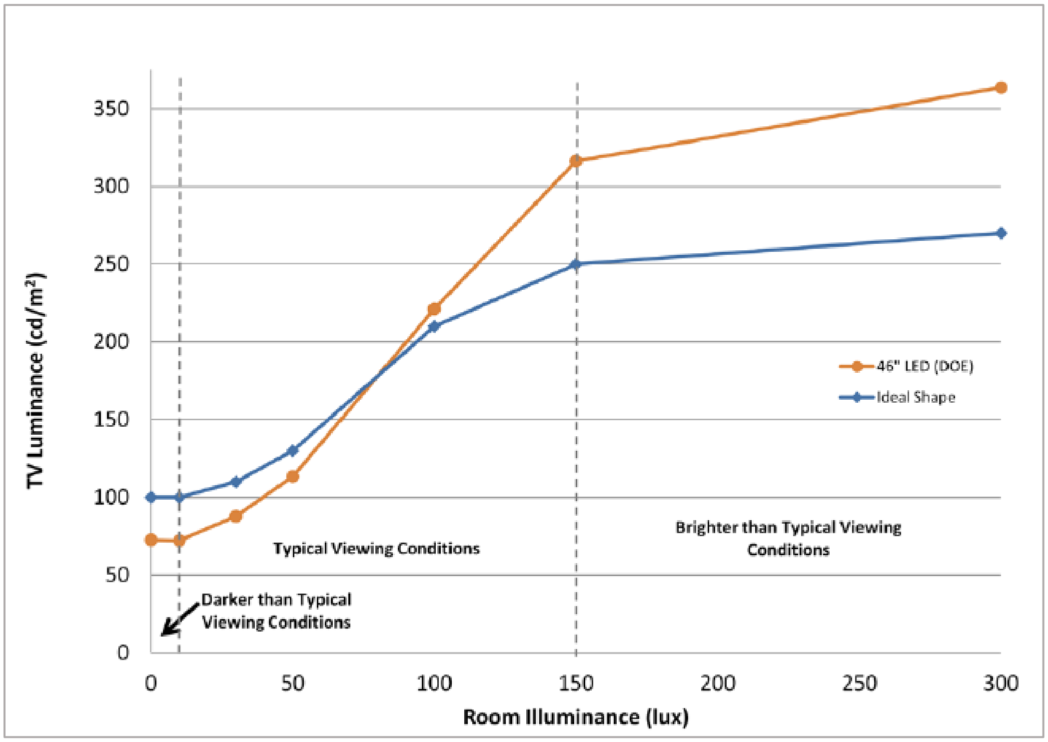
Figure 40. Three-bar vs. Dark Grey: LED TV



And yet, both ENERGY STAR v8 and EU luminance limits are based on the DOE ideal luminance curve[15], which is based in part on Matsumoto. The 2012 DOE illuminance study states, "The ideal TV luminance levels for dark room conditions in Figure 1.3.1 are based on Imaging Science Foundation's (ISF) recommended brightness level for TVs in a dark room setting, while the luminance levels for brighter conditions are based on a 2010 study on appropriate luminance levels, which found that at 100 lux, subjects preferred a TV brightness range from 160 to 248 cd/m^2." And this is based on undocumented video content with an Average Light Level (ALL), presumably the same thing as APL', of 25% vs. 34% for the IEC dynamic test clip and based on an angular screen size of 20 degrees, which is a function of screen diagonal and viewing distance from the TV, which have changed since this data was collected. Also note that the ideal appears to be based on the average of the preferred luminance level for young (160 cd/m^2) and old (248 cd/m^2) people. Because the range between the two is so large, neither group is likely to be satisfied with a TV set to the "ideal" luminance value.

[15] Section 1.3.1 of https://www1.eere.energy.gov/buildings/appliance_standards/pdfs/tv_tpnopr_room_illuminance_abc_031912.pdf

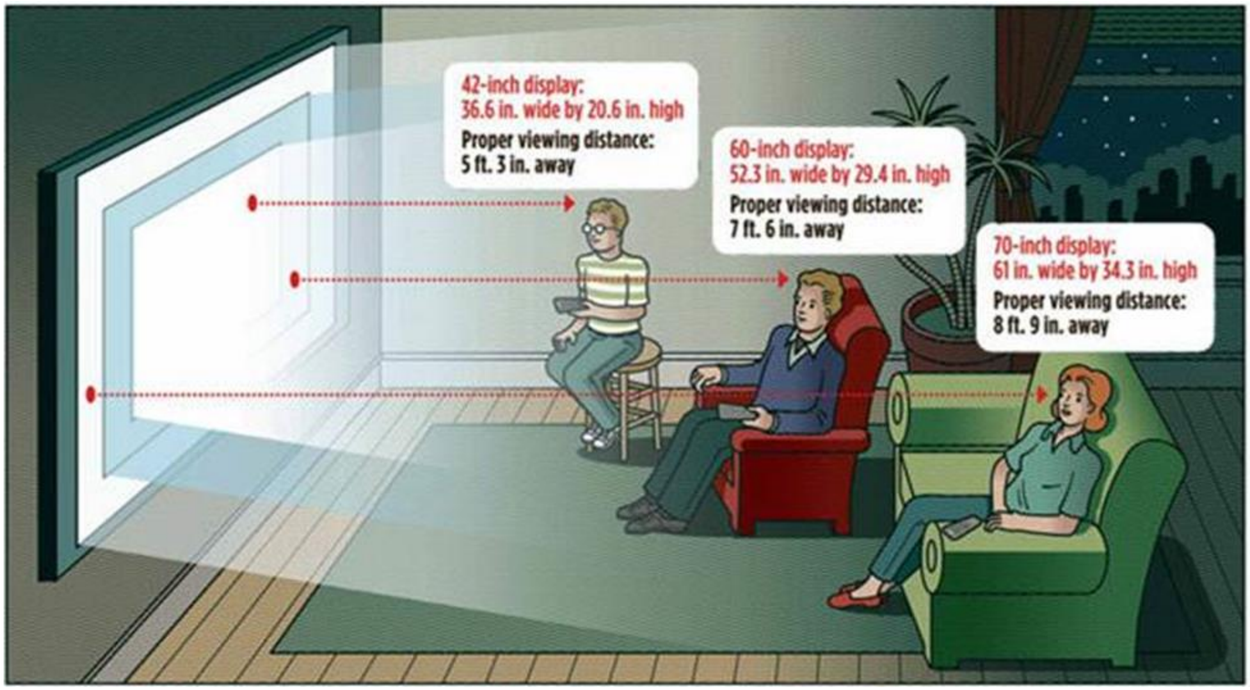
Figure 41. Ideal DOE ABC Luminance Curve



Appendix E: Justification for Camera Placement Distance from TV

A TV manufacturer presented data that suggests a recommended viewing distance of 2H-2D. 2 x height is approximately 1 x diagonal for a 16x9 TV. This range can be expressed as approximately 1-2D with an average of 1.5D, which is what is shown in the figure below:

Figure 42: Typical Viewing Distances



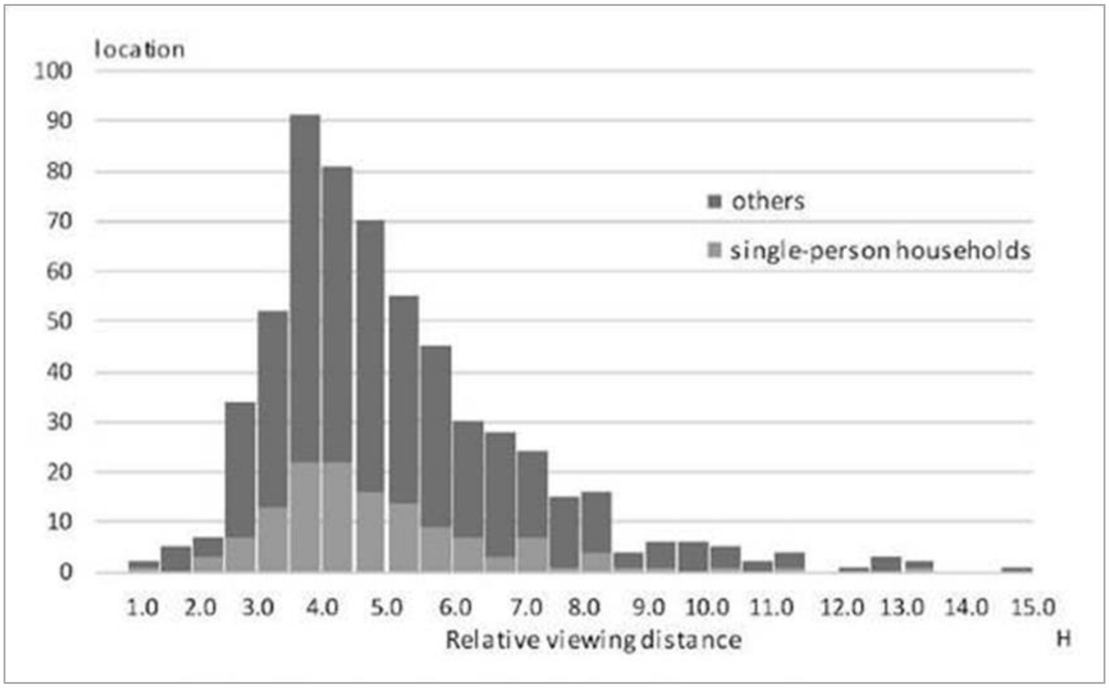
ITU-R Rec. BT.20221 specifies relative viewing distances of 3.2 H (H is the screen height) for 2K television system, 1.6 H for 4K, and 0.8 H for 8K. This translates approximately to the following for 16:9 TVs, the most common aspect ratio:

- HD: 1.5D
- 4K: 0.8D
- 8K: 0.4D

Yagi et. al.[16] show that the average installed TV in Japan in 2019 had a viewing distance of 5H (or ~2.5D for 16:9 TVs), but installed TVs skew towards smaller HD TVs. We are focused on new TVs sold, which are bigger and higher resolution. Yagi et. al. point out that size and resolution are driving the H to distance ratio down over time. And we are designing a test method for use going years into the future.

[16] A Survey of Television Viewing Conditions at Home in Japan, 2019

Figure 43. Distribution of Relative Viewing Distances



Basler lens options are shown below. The TV must fill the lens field of view, and there is no zoom capability. So each lens supports a specific distance where the TV width fills the FOV (for 16:9 TVs).

1. 1.06D
2. 1.53D
3. 2.12D

We have determined that there is a 4% difference in expected luminance readings between the first and second options and even 2% difference between the latter two based on testing.

So we recommend 1.53 x diagonal (1.53D) because we expect it to be representative of real world viewing distance over the timeframe in which this test method is in use, because it aligns with the above stakeholder data on viewing distance, and because it requires less lab space than 2D.

Note: We now represent camera distance as a multiple of screen width in our test kit manual (1.76-1.78 times the screen width, which is equivalent to 1.5 x diagonal). We do this because it simplifies the calculation for screens that have an aspect ratio other than 16x9.

1.5 x diagonal is consistent with rtings.com recommendation of 1.6 x diagonal for mixed usage and 1.2 x diagonal for cinema. See: https://www.rtings.com/tv/reviews/by-size/size-to-distance-relationship