Color calibration methods for OLED displays

***Maliha Ashraf*[1], *Alejandro Sztrajman*[1], *Dounia Hammou*[1], *Rafał K. Mantiuk*[1]**

[1]*Department of Computer Science and Technology, University of Cambridge, UK*

Abstract

Accurate color reproduction on a display requires an inverse display model, mapping colorimetric values (e.g. CIE XYZ) into RGB values driving the display. To create such a model, we collected a large dataset of display color measurements for a high refresh-rate 4-primary OLED display. We demonstrated that, unlike traditional LCD displays, multi-primary OLED display color responses are non-additive and non-linear for some colors. We tested the performances of different regression methods: polynomial regression, look-up tables, multi-layer perceptrons, and others. The best-performing models were additionally validated on newly measured (unseen) test colors. We found that the performances of several variations of 4th-degree polynomial models were comparable to the look-up table and machine-learning-based models while being less resource-intensive.

Introduction

OLED (Organic Light Emitting Diodes) displays have gained significant popularity due to their ability to achieve better contrast than LCD displays and wider color gamuts. OLED displays utilize self-emitting sub-pixels, rendering the need for a backlight and filters obsolete. This inherent design leads to some significant advantages. One key advantage is the narrower spectral emission bandwidth response of their primaries when compared to traditional LCDs, resulting in larger color gamuts. Moreover, the self-emissive nature of OLED sub-pixels allows for the attainment of true black by completely turning off pixels. This capability dramatically enhances the dynamic range and contrast of OLED displays. To further extend the luminance range and minimize power consumption, many OLED display manufacturers introduce an additional white primary alongside the traditional red, green, and blue channels. While this enhances the display dynamic range, it also introduces difficulties in color calibration as the relationship between input signals and emitted light becomes more non-linear. Since, the additional white primary lies within the chromaticity gamut spanned by the red, green, and blue primaries, it does not expand the chromaticity gamut but enables the display to produce high luminance, low chroma additional colors. Bergquist (2008) refers to these additional colors as "degenerate colors" [1]. In addition to these challenges, OLED display manufacturers contend with various engineering constraints. These include constraints on power consumption, limitations on current density, and the potential for pixel burn-in. These factors introduce additional layers of complexity to display control and, consequently, the calibration process. To characterize the color of these OLEDs with high fidelity, more complex models are required to circumvent the non-linearities inherent to this technology.

We are specifically interested in evaluating the performance of inverse display models, mapping device-independent CIE XYZ color coordinates to the native RGB values of the display. This is in contrast to most previous works, which evaluated forward display models. We focus on the inverse models because, in most use cases, we want to accurately reproduce colors on a display rather than accurately predict colors emitted by a display. While simple display models, such as gain-gamma-offset, are bijective and thus invertible, most of the models we tested in this work are non-invertible, and, therefore, the performance of a forward display model is not indicative of the performance of an inverse display model. Inverse display models are also more difficult to validate as their validation requires collecting new measurements.

Related work

OLEDs are a fairly nascent technology and thus there has not been a lot of research on accurate color reproduction for these displays yet. There have been some studies characterizing and recommending the use of OLED displays for vision research [2], for medical imaging [3] and for improved image quality [4]. The fast temporal response of the OLED displays is shown to be especially advantageous for a number of applications, but the displays used in these studies are more than 10-years old and different from the newer displays with additional sub-pixels and more sophisticated energy-saving, gamut mapping algorithms, etc. There is a wealth of literature on calibration recommendations and colorimetric characterization standards for older display technologies such as CRT and LCDs [5, 6, 7, 8], but most of these techniques cannot be applied for OLED color calibration directly. This has elicited the development of multiple RGB-to-RGBW conversion algorithms [9, 10, 11]. Wang et al. [11] point out that many conventional approaches aim to increase displayed brightness while preserving hue and saturation, often disregarding color distortion and power consumption. Sun and Luo (2013) [12] and Tian et al. (2019) [13] proposed a set of useful models for calibrating OLEDs. Sun and Luo (2013) tested traditional color calibration models; GOG (Gain-Offset Gamma) [7], S-Curve [8], and PLCC (Piece-wise Linear interpolation assuming Constant Chromaticity) [14] on OLED measurements and demonstrated that they resulted in very large color errors. They proposed two novel models in their work, which compensate for the non-linearities introduced by the fourth white primary. They also measured stimuli with different screen coverage areas to model the effect of lowered brightness with higher screen coverage. This is a limitation of OLED technology caused by the limitation on peak current that the display can draw. Sun and Luo (2013) proposed a parameter that would scale the luminance of the stimulus relative to the average luminance of the full screen. Tian et al. (2019) [13] have pro-

posed a PLCC-based model followed by a 3D LUT to transform native RGB values to XYZ tristimulus values, which is demonstrated to perform better than the PLCC and 3D LUT models applied separately. Bodner et al. (2019) [15] proposed splitting the main gamut into three sub-gamuts for more accurate color measurement when using four-primary displays. Their model corrects the XYZ values from a test colorimeter device with respect to a reference spectroradiometer, to compensate for luminance measurement errors in devices that do not measure the full spectrum. The proposed model does not compute the transformations between the device RGB and the measured XYZ values. However, a similar approach can be used to develop forward and inverse display color characterization models for OLED displays. In contrast to those previous works, we focus on the inverse display models, mapping colorimetric CIE XYZ values to the native RGB values driving the display.

Display Measurements

We used the LG 27" UltraGear™ (Model no: 27GR95QE) OLED Gaming Monitor with QHD (2560 x 1440) pixel resolution and 240 Hz display refresh rate. A Windows 10 workstation equipped with an NVIDIA GeForce RTX 3080 GPU was driving the display via a 4k HDMI cable. To utilize the full display color gamut, Windows HDR settings were enabled. The Psychtoolbox software [16] running in MATLAB was used for presenting the measurement patches with `HDR10` and 16-bit floating point precision per color channel enabled. The measurements were done in a dark room. We selected the FPS (first-person shooter) HDR mode in the display settings as it provided the best balance of linear response along with high peak brightness. To avoid any unwanted alteration of the display outputs, all auto adjustment and enhancement settings, such as the auto dynamic contrast, dynamic tone mapping, super-resolution, OLED care, AI services, and others, were turned off.

The linear RGB input to the display was passed via `PsychHDR` mode in the Psychtoolbox in units of nits or cd/m^2 which was then PQ-encoded with the ST-2084 PQ Perceptual Quantizer EOTF (Electro-Optica Transfer Function) [17]. We collected 3 kinds of measurements for our training dataset: i) ramps, ii) grid, and iii) local neighborhood grid. The measurements were done by displaying the color in a rectangle in the middle of the screen. The test patches covered 5% of the total screen area, and the rest of the screen was black. The spectrum corresponding to each colour patch as well as the XYZ tristimulus values were measured using the JETI Specbos 1211 broadband spectroradiometer. The wavelenegth range of the spectroradiometer is 350 - 1000 nm with a 4.5 nm optical resolution. The measurement angle was 1.8°. The luminance measurement range is 0.2 - 150,000 cd/m^2 and the manufacturer reported measurement accuracy at 10 cd/m^2 is ±2% for luminance and ±0.002 for chromaticity.

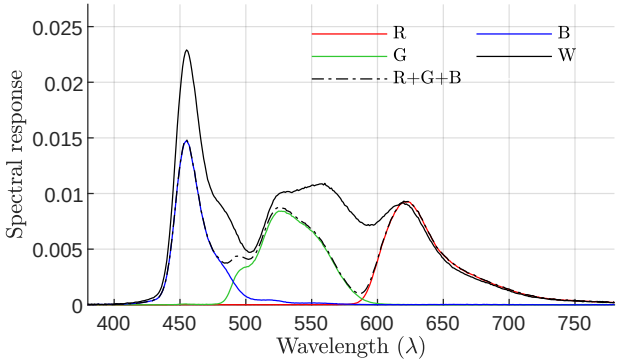


Figure 1. Emission spectra of the display. The white sub-pixel response (solid black line) is not equal to the sum of the red, green, and blue responses (dashed black line) demonstrating the non-additivity in the color response of the display.

Ramps The individual responses of each of the red, green, blue, and white channels were measured using 240 linear pixel values (per color) between 0.01 and 700 uniformly sampled in the logarithmic space. The maximum intensity responses of the four channels at their maximum pixel drive were obtained at $R = (a,0,0)$, $G = (0,a,0)$, $B = (0,0,a)$, and $W = (a,a,a)$, where $a = 700$ cd/m^2. The spectral responses of these individual sub-pixels are shown in Figure 1. Note that the spectral emission curve of the white sub-pixel was not the same as the R+G+B spectral response curve which demonstrates the non-additive characteristic of OLED color emission. Unlike traditional three-primary displays, where white is produced by adding the responses of the red, green, and blue primaries, OLED displays, like the one used in this study, usually have an additional white primary, which helps achieve higher luminance levels. The chromaticity gamut of the display is compared with Rec.2020, Rec.709, and P3-D65 colorspaces in Figure 2. The chromaticity gamut of our display is very similar to the P3-D65 gamut but slightly different in the chromaticity of the green primary. The color intensity responses of the secondary hues (cyan, magenta, and yellow), as well as 6 tertiary hues (orange, chartreuse green, spring green, azure, violet, and rose), were also measured using the same logarithmically increasing pixel values as the primary color ramps. The ramp measurements of the primary and secondary hues are shown in Figure 3 (top-left) along with the corresponding chromaticity values (top-right). The relationship between the pixel values and the measured luminance responses was linear (both in linear and log-log scale) for all 7 colors. In the chromaticity gamut, the chromaticities of these primary and secondary colors at higher intensities were close to the boundaries of the gamut, as expected. At lower pixel intensities, the color gamut of the display shrank, and the colors did not appear as saturated.

We also measured the ramps from primary and secondary hues (red, green, blue, cyan, magenta, and yellow) toward white. For example, for the red-to-white ramp, the pixel values were $RW = (700,a,a)$ where the value of a increased from 0.01 to 700 incrementally in 100 steps logarithmically. These kinds of ramps were measured to detect whether the response of the display was non-linear when the color changes from high-chroma (red, for example), to lower chroma (white/grey) values. The luminance and chromaticity responses of these ramps are shown in Figure 3 (bottom row). The change in luminance was highly non-linear with respect to the change in the corresponding pixel values. For all the tested colors, the luminance remained constant up to a certain pixel value (depending on the color) and then started increasing, signifying the point where the white sub-pixel was activated. In the chromaticity gamut, we can also see some gamut clipping as the high-chroma colors moved towards low-chroma values.

Grids We measured a 13x13x13 grid with pixel values from 0.01 to 700 spaced logarithmically, covering the whole display's gamut. In addition to these 2,197 data points, we measured the local neighbor grids of each point in a coarser 6x6x6 grid. These local grids were the 26 points around each of the central grid points (arranged like a Rubik's cube). The R, G, B coordinates of each point in this local neighbor grid are either the same or $\pm 10\%$ of the RGB values of the central grid point. The local neighborhood grid was measured to ensure that gradients (Jacobians) of the color space transformations were correctly modeled.

Training dataset All the ramp, grid, and local neighbor grid measurements were combined to form the training dataset for the color calibration models. Any repeated data points between the different types of measurements were removed. A few data points had high measurement errors and were identified by converting the input RGB and the measured XYZ values to L*a*b* color space and removing the data points with color differences greater than 10. The RGB to L*a*b* transformation assumed that the input values were in sRGB space, which is incorrect but enough to identify data points with high measurement errors. After all the pre-processing, the training dataset had 11,296 RGB and XYZ measurement pairs.

Models

The goal of this work was to create an inverse display model that can transform the tristimulus coordinates represented in the device-independent CIE 1931 XYZ colorspace to the native RGB coordinates of the display. As discussed in the previous sections, OLED displays use several strategies to maximize their dynamic ranges and color gamut and to optimize energy consumption, which make accurate color calibration of these displays non-trivial. This also means that the forward and inverse display models for such displays are non-invertible. We aim to use these displays in psychophysical vision experiments, and so rather than the more commonly used forward models (mapping native RGB coordinates to a device-independent color space), we are interested in the inverse display model that can accurately map our required XYZ value to the corresponding RGB input to the display. In the following sections, we describe the models used in this study.

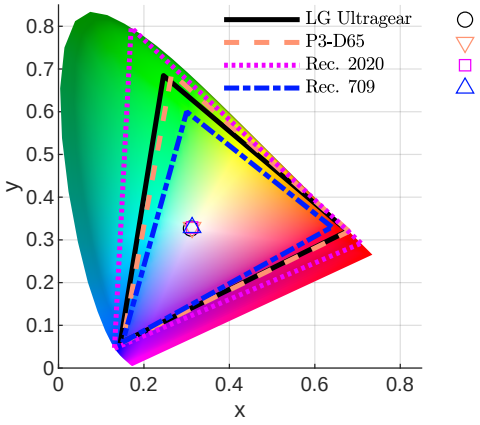


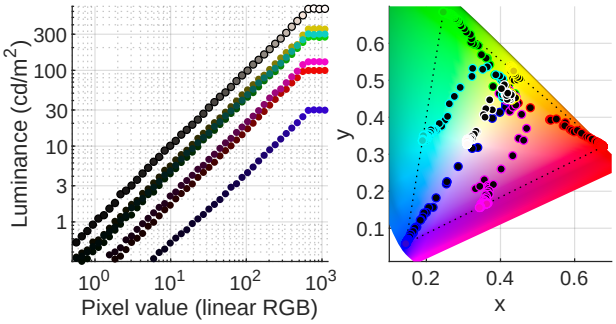
Figure 2. Gamut of our OLED displays compared with Rec. 2020, Rec. 709, and P3-D65 color spaces. The lines show the RGB primaries, and the markers show the un-calibrated white points of the displays and those of the standard color spaces.

Lookup tables

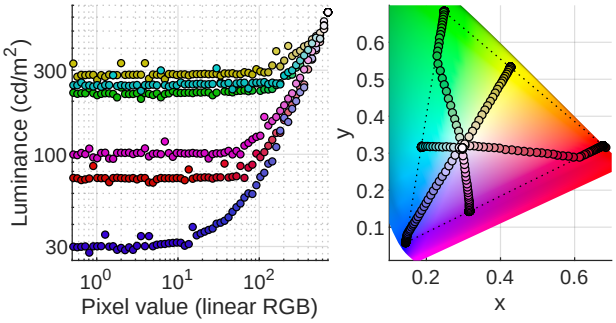
Lookup tables (LUT) are commonly used for device color calibration [18], but require a very large number of measurements and are computationally expensive. In our work, the LUT method served as a benchmark to compare against the results of the other machine learning-based and analytical models tested in this work. Since our training dataset was not in the form of a uniform grid, we interpolated the points using 3D Delaunay Triangulation using the MATLAB function `delaunayTriangulation`. The points were queried using the MATLAB function `pointLocation`.

Machine learning models

In order to model the transformation between the XYZ and the native RGB color spaces, we employed a multi-layer perceptron (MLP) with SIREN layers [19]. This choice was motivated by its improved capability in handling complex function approx-



(a) Ramps from black



(b) Ramps towards white

Figure 3. (a) Ramp measurements from black to primaries of the display. Left: red, green, blue, white, cyan, yellow, and magenta ramp measurements. The luminance response changed linearly with pixel values and saturated for pixel values approximately greater than 700; these values were not included in the training dataset. Right: chromaticity coordinates of the primary and secondary color ramps. At higher intensities, the chromaticity coordinates remained close to the gamut boundary but moved inwards as the gamut shrank for lower luminances. (b) Ramp measurements for high (display primaries) to low chroma colors (display white point white point). Left: The luminance response was non-linear with respect to the pixel values. Right: Chromaticity change from high to low chroma colors was mostly linear. For red and green colors, some gamut remapping occurred, introducing non-linearity in color transformation between the input pixel values and the emitted photometric values.

imations. Our MLP consisted of three hidden layers, each with 60 neurons. The input to our model was the XYZ color space data, from the training dataset described in *Methodology*, while the output was the corresponding RGB values. We utilized the Adam optimizer for 1000 epochs with a learning rate of $1e^{-5}$. We explored various configurations to identify the most accurate mapping between XYZ and RGB spaces. Our experiments included:

- Multiple loss functions, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and both MAE and MSE with logarithmic inputs. The explicit form of the losses can be seen in subsection *Optimization and loss functions*.
- Applying a logarithmic transformation to the input (XYZ), output (RGB), or both.
- Normalizing the range of the input (XYZ), output (RGB), or both to either [0, 1] or [-1, 1].
- Multiple values of the ω_0 hyper-parameter in SIREN networks, which controls the order of magnitude of the frequencies of the data. Lower value assumes a smoother transformation between the colorspaces.

We performed hyper-parameter tuning, which helped us identify the optimal configurations. In Table 3, we show results for the 3 best performing models.

Analytical models

In analytical methods, we optimize the coefficients of the transformation matrix M between the input and output color spaces. These models can be generally expressed as:

$$\rho_{n_\rho \times n_s} = M_{n_\rho \times n_r} \times r_{n_r \times n_s}, \tag{1}$$

where ρ represents the vector with RGB pixel values (the output of the models) and r is the vector of radiometric color values (function of CIE XYZ coordinates and the input of the models). The subscripts denote the sizes of the matrices; n_ρ is the number of channels of the pixel values which is 3 in most models and 4 in the case of the *RGBW* model, n_r is the number of dimensions of the model input. In the simplest linear model, this is 3 when the X,Y,Z values are used as the model input, in other cases it depends on the degree of the polynomial or root polynomial method used. n_s is the number of color coordinates to be transformed.

Polynomial regression

The linear 3×3 model is the simplest and most commonly used type of polynomial regression for color calibration of imaging devices [20, 21]. In our case, the input and output of this linear model were the matrices with XYZ (r) and the RGB values (ρ) respectively, both of size $3\times n_s$. M was the 3×3 transformation matrix. The use of the polynomial regression technique has been shown to improve color accuracy [22, 23]. Since our OLED display likely employs non-linear pixel drive functions, we tested many different variations of polynomial models, as listed in Table 1.

The first row corresponds to the simple linear model, and each of the following rows denotes a higher-degree model with new polynomial terms in addition to the terms of the preceding models. The output of all the polynomial models is the $3\times n_s$ matrix of RGB values, while the size of the M matrix is $3\times n_r$.

Table 1: Coefficients of polynomial regression models

Polynomial degree	n_r	Input terms
$r_{1,1}$	3	X,Y,Z
$r_{1,2}$	4	$r_{1,1}$, 1
$r_{2,1}$	7	$r_{1,2}, XY, XZ, YZ$
$r_{2,2}$	10	$r_{2,1}, X^2, Y^2, Z^2$
$r_{3,1}$	11	$r_{2,2}, XYZ$
$r_{3,2}$	17	$r_{3,1}, X^2Y, X^2Z, Y^2X, Y^2Z, Z^2X, Z^2Y$
$r_{3,3}$	20	$r_{3,2}, X^3, Y^3, Z^3$
$r_{4,1}$	26	$r_{3,3}, X^2Y^2, X^2Z^2, X^2YZ, Y^2Z^2, Y^2XZ, Z^2XY$
$r_{4,2}$	32	$r_{4,1}, X^3Y, X^3Z, Y^3X, Y^3Z, Z^3X, Z^3Y$
$r_{4,3}$	35	$r_{4,2}, X^4, Y^4, Z^4$

Root polynomial regression

The root polynomial regression method was introduced in Finlayson et al. (2015) as a lower-cost exposure-invariant alternative to polynomial regression color calibration [24]. We tested the same root polynomial models up to 4th degree for OLED color calibration. The input terms and the number of dimensions of the input XYZ matrix are listed in Table 2.

Table 2: Coefficients of root polynomial regression models.

Root polynomial degree	n_r	Input terms
$\bar{r}_2$	7	$r_{1,2}, \sqrt{XY}, \sqrt{XZ}, \sqrt{YZ}$
$\bar{r}_3$	14	$\bar{r}_2, \sqrt[3]{X^2Y}, \sqrt[3]{X^2Z}, \sqrt[3]{Y^2X}, \sqrt[3]{Y^2Z}, \sqrt[3]{Z^2X}, \sqrt[3]{Z^2Y}, \sqrt[3]{XYZ}$
$\bar{r}_4$	23	$\bar{r}_3, \sqrt[4]{X^3Y}, \sqrt[4]{X^3Z}, \sqrt[4]{Y^3X}, \sqrt[4]{Y^3Z}, \sqrt[4]{Z^3X}, \sqrt[4]{Z^3Y}, \sqrt[4]{X^2YZ}, \sqrt[4]{Y^2XZ}, \sqrt[4]{Z^2XY}$

RGBW gamut

We tested another model that assumes that the red, green, blue, and white (RGBW) sub-pixels in OLED displays are independent and that their combined color response can be derived through linear combinations of their individual pixel responses. However, it's important to note that in our OLED display, all four sub-pixels are not activated simultaneously. Instead, the white (W) sub-pixel can replace either the red (R), green (G), or blue (B) sub-pixel based on which one has the lowest drive value for any given color. In the case of purely achromatic colors, the white sub-pixel alone drives the response of the display. The assumed conversion from *RGB* to $\widetilde{R}\widetilde{G}\widetilde{B}\widetilde{W}$ colorspace was as follows:

$$\begin{aligned} \widetilde{R} &= R - \min(R,G,B), & \widetilde{G} &= G - \min(R,G,B), \\ \widetilde{B} &= B - \min(R,G,B), & \widetilde{W} &= \min(R,G,B) \end{aligned} \tag{2}$$

The size of the output RGB matrix for training this model was $4\times n_s$. Different polynomial and root polynomial configurations for the XYZ input matrix were tested as outlined in Table 1-2 and the size of the XYZ input and the transformation matrix changed accordingly. The $\widetilde{R}\widetilde{G}\widetilde{B}\widetilde{W}$ predictions from this model were then converted to the native RGB values as follows:

$$R = \widetilde{R} + \widetilde{W}, \quad G = \widetilde{G} + \widetilde{W}, \quad B = \widetilde{B} + \widetilde{W} \tag{3}$$

3-gamut

This model is based on the methodology proposed by Bodner et al. (2019) [15], in which they segmented the four-primary RGBW color gamut into three distinct sub-gamuts: RGW, RBW, and GBW for color correction. This segmentation is based on

Table 3: Comparison of color calibration models. The LUT model is the benchmark against which other models are tested. The analytical and machine learning-based models are grouped separately. N: number of parameters. C: complexity of the model, computed as the total number of multiplication and addition operations required to transform a single color coordinate. The best (lime green), second-best (light green), and third-best (tea green) models in each column are highlighted. We include the 3 best performing MLP models, indicating the value for ω_0 and the training loss. The tag normXYZ indicates that minmax scaling was only applied to the input.

Model names	N	C	Mean error at each luminance level (cd/m^2)										Mean ΔE_{00}	Median ΔE_{00}	$\Delta E_{00} >$ 5(%)
			0.15	1.7	4.8	20	25	63	100	250	320	400			
LUT	✗		0.0658	0.5349	1.632	1.803	2.867	2.98	3.931	3.955	5.026	5.174	3.179	2.535	21.67
Poly 4,2 log MAE	96	189	0.1254	1.675	2.709	3.156	3.254	3.722	4.335	4.564	4.581	4.524	3.586	3.509	20.42
Poly 4,2 log MSE	96	189	0.1311	1.694	2.703	3.187	3.267	3.766	4.324	4.597	4.645	4.621	3.589	3.527	20
Poly 4,3 log MSE	105	207	0.125	1.569	2.759	3.677	3.552	3.67	4.39	5.092	5.119	4.7	3.875	3.715	21.25
Poly 4,3 MAE	105	207	0.1262	1.555	2.786	3.646	3.486	3.709	4.39	5.102	5.148	4.55	3.711	3.687	21.25
RGBW Poly 4,3 MAE	140	276	0.1277	1.562	2.756	3.632	3.599	3.707	4.296	5.132	5.088	4.482	3.69	3.716	21.25
3-gamut Poly 3,1 MSE	99	63	0.1307	1.302	2.796	4.126	4.099	4.397	4.586	4.049	4.027	4.094	3.82	3.589	26.67
MLP $\omega_0 = 10$ MAE	$1.14e^4$	$2.23e^4$	0.167	0.9394	2.312	3.696	3.771	3.96	4.188	4.582	4.862	5.671	3.509	3.343	23.75
MLP $\omega_0 = 10$ MAE normXYZ	$1.14e^4$	$2.23e^4$	0.1238	1.423	2.304	4.47	4.262	4.105	4.598	4.718	5.028	4.69	3.759	3.553	26.67
MLP $\omega_0 = 30$ MAE	$1.14e^4$	$2.23e^4$	0.6154	1.411	2.142	2.785	3.155	3.275	3.626	4.125	5.743	6.079	4.043	3.004	21.25

the assumption that, depending on a color's location within these sub-gamuts, three out of the four primary colors drive the display to produce that particular color. They then derived correction matrices to minimize errors in XYZ color measurements when comparing results from a reference spectroradiometer with those from a test colorimeter. In our modified inverse display model, we adopt the three-sub-gamut approach to determine the transformation matrices that map the CIE XYZ tristimulus values to the corresponding $\widetilde{RGW}$, $\widetilde{RBW}$, or $\widetilde{GBW}$ pixel values. The transformation to and from the *RGB* and $\widetilde{RGBW}$ colorspace follow Eqs. 2-3. The chromaticity coordinates from the XYZ values are triangulated with the chromaticity coordinates of the red, green, blue, and white primaries to identify the sub-gamut each color belongs to. The training dataset was thus classified into three sub-gamuts and 3 different transformation matrices were estimated. We also tested the polynomial and root polynomial versions of the model where the XYZ input to the model was transformed accordingly.

Optimization and loss functions

The coefficients of the transformation matrices in all the proposed analytical models were optimized using MATLAB `GlobalSearch` with 4 different error functions to be minimized:

$$
\begin{aligned}
E_{MAE} &= \frac{1}{n_s}\sum_{n=1}^{n_s} |\Delta R_n| + |\Delta G_n| + |\Delta B_n|\,, \\
E_{MSE} &= \sqrt{\frac{1}{n_s}\sum_{n=1}^{n_s} \Delta R_n{}^2 + \Delta G_n{}^2 + \Delta B_n{}^2}\,, \\
E_{logMAE} &= \frac{1}{n_s}\sum_{n=1}^{n_s} |\Delta \log R_n| + |\Delta \log G_n| + |\Delta \log B_n|\,, \\
E_{logMSE} &= \sqrt{\frac{1}{n_s}\sum_{n=1}^{n_s} (\Delta \log R_n)^2 + (\Delta \log G_n)^2 + (\Delta \log B_n)^2}\,,
\end{aligned}
\tag{4}
$$

where $[\Delta R, \Delta G, \Delta B]' = [R,G,B]' - M[X,Y,Z]'$ is the difference between the measured and the predicted RGB values. The predicted RGB values are the product of the matrix multiplication between the optimized transformation matrix M and the XYZ values of the training dataset. The combination of all the discussed models with the above four loss functions resulted in 156 different variations of the analytical models.

Validation

We evaluated the models' performances using a set of XYZ test values that were not included in the training dataset. These XYZ values were the color measurements of the X-Rite ColorChecker captured under Illuminant C from [25], scaled to 10 different luminance levels. The values of these luminance levels were optimized such that the histogram of the test dataset luminance values matched that of the training dataset. The optimized luminance levels were: 0.15, 1.7, 4.8, 20, 25, 63, 100, 250, 320, and 400 cd/m^2. The test dataset consisted of 240 different XYZ color coordinates. To test the performance of the model, the predicted RGB values were input to the display, and the XYZ values were measured. The XYZ values of the test dataset and the measured XYZ values were converted to CIE L*a*b* colorspace with a D65 white point at 650 cd/m^2 (highest luminance in the tested mode and screen coverage), and the CIE ΔE_{00} color difference values were used as the metric for the models' performances. However, it was impractical to measure the predictions for such a large amount of models. Thus, we shortlisted the best models using the forward LUT display model as a first approximation. The predicted RGB values from each of the models were passed through the forward LUT to predict the corresponding XYZ values. The color difference values between the LUT-predicted XYZ and the XYZ values of the test datasets were calculated. The models were ranked according to the lowest percentage of test color patches with $\Delta E_{00} > 5$. We identified the resulting top 10 models and then measured their predicted RGB color values.

Results

The results of the final validation for the best 10 models are shown in Table 3. The mean and median errors for the whole test set as well as the per-luminance mean color difference errors were calculated. Another useful measure of testing color is the proportion of color patches with $\Delta E_{00} > 5$. We found that the LUT method was the best-performing model for most of the conditions

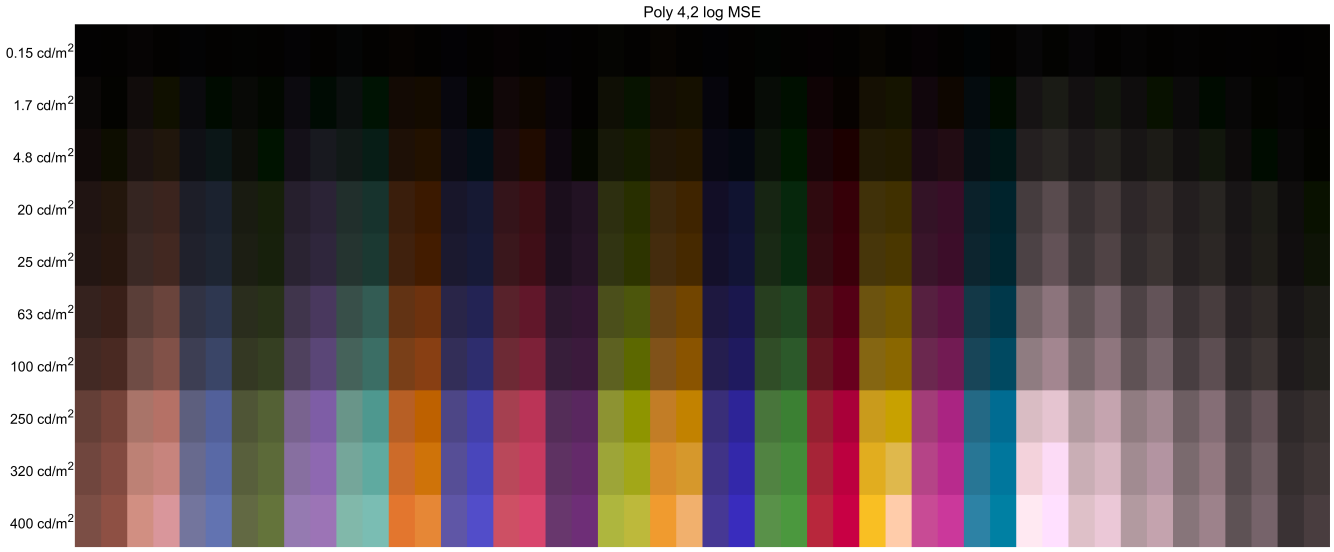


Figure 4. Approximate sRGB representation of the reference (left side in each patch) colors and the measured colors (right side in each patch) from the predicted RGB values from the 4th degree polynomial model ($r_{4,2}$) with E_{logMSE} optimization loss function.

and tested metrics, which is expected for a large training dataset. However, we were interested in the model with lower computational costs and simpler to implement as an OpenGL shader. The MLP-based models performed well, with the *"MLP* $\omega_0 = 30$ *MAE"* model as the second-best-performing model for most of the metrics. However, machine learning models are also more computationally expensive to implement compared to analytical methods. The tested analytical models slightly outperformed the LUT model in terms of lower percentages of color patches with color difference values greater than 5. We can argue that this metric is a better indicator of model performance as it indicates a more consistent performance and a smaller spread of color differences among the test predictions. Out of all our 156 analytical models, variations of 4th-degree polynomial models performed the best on average. Mustafa et al., (2022) have also demonstrated that a 4th degree polynomial basis is very effective in mapping between SDR and HDR colorspaces [26].

The luminance-dependent mean errors show that the performances of all models were best for lower luminances and worsened for higher ones. Our training dataset was coarser in the high-luminance region, which could be the cause of these errors. Another reason could be that in OLED displays specifically, the white sub-pixel plays a larger role in the high luminance region and introduces non-linearities that cannot be adequately captured by the color calibration models. The color difference between the reference test colors and the measured color values from the predictions of the *Poly 4,2 log MSE* model are shown in Figure 4.

Conclusions

This work provides a comprehensive comparison of various color calibration models, assessing their effectiveness for OLED displays. Based on our findings, we recommend employing 4th-degree polynomial analytical models as they produce good color fidelity while being less computationally expensive as compared to look-up tables or machine-learning-based methods. It should be noted that the same models would not necessarily be the most optimal for all types of OLED displays as different manufacturers use their own in-house technologies to enhance the performance of their displays. This work does not consider the effect of different Average Picture Levels (APL). In OLED displays the proportion of the active pixels affects the luminance response of the sub-pixels, and so the same RGB value would result in different color intensities if the mean luminance of the content is different.

Acknowledgements

This research received funding from Meta. We would also like to thank Alexandre Chapiro for his valuable feedback and suggestions.

References

[1] J. Bergquist, "Fast and accurate colour gamut volume measurements of RGB displays with degenerate colours," *VHF7-3, IDW Proc*, pp. 1030–1033, 2018.

[2] E. A. Cooper, H. Jiang, V. Vildavski, J. E. Farrell, and A. M. Norcia, "Assessment of OLED displays for vision research," *Journal of vision*, vol. 13, no. 12, pp. 16–16, 2013.

[3] T. Elze, C. Taylor, and P. J. Bex, "An evaluation of organic light emitting diode monitors for medical applications: Great timing, but luminance artifacts," *Medical physics*, vol. 40, no. 9, p. 092701, 2013.

[4] D. M. Hoffman, P. V. Johnson, J. S. Kim, A. D. Vargas, and M. S. Banks, "240 Hz OLED technology properties that can enable improved image quality," *Journal of the Society for Information Display*, vol. 22, no. 7, pp. 346–356, 2014.

[5] I. E. Commission *et al.*, "Color measurement and management in multimedia systems and equipment: Part 6. Equipment used for digital image projection," *IEC Committee Draft*, pp. 61966–6, 1998.

[6] L. D. Silverstein and T. G. Fiske, "Colorimetric and photometric modeling of liquid crystal displays," in *Color and Imaging Conference*, pp. 149–156, Society for Imaging Science and Technology, 1993.

[7] R. S. Berns, "Methods for characterizing CRT displays," *Displays*,

vol. 16, no. 4, pp. 173–182, 1996.

[8] Y. Kwak and L. Macdonald, "Method for Characterizing an LCD Projection Display," *Proc. SPIE*, vol. 4294, Mar. 2001.

[9] C. Lee and V. Monga, "Power-Constrained RGB-to-RGBW Conversion for Emissive Displays: Optimization-Based Approaches," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, pp. 1821–1834, Oct. 2016.

[10] M. E. Miller and M. J. Murdoch, "RGB-to-RGBW Conversion with Current Limiting for OLED Displays," *Journal of the Society for Information Display*, vol. 17, no. 3, pp. 195–202, 2009.

[11] L. Wang, Y. Tu, L. Chen, K. Teunissen, and I. Heynderickx, "27.2: Trade-off between Luminance and Color in RGBW Displays for Mobile-phone Usage," *SID Symposium Digest of Technical Papers*, vol. 38, no. 1, pp. 1142–1145, 2007.

[12] P.-L. Sun and R. M. Luo, "Color Characterization Models for OLED Displays," *SID Symposium Digest of Technical Papers*, vol. 44, no. 1, pp. 1453–1456, 2013.

[13] D. Tian, L. Xu, and M. R. Luo, "The Characterization of HDR OLED Display," *Electronic Imaging*, vol. 31, pp. 326–1–326–6, Jan. 2019.

[14] D. L. Post and C. S. Calhoun, "An Evaluation of Methods for Producing Desired Colors on CRT Monitors," *Color Research & Application*, vol. 14, no. 4, pp. 172–186, 1989.

[15] B. Bodner and N. Robinson, "83-3: Calibration of Colorimeters for RGBW Displays," *SID Symposium Digest of Technical Papers*, vol. 50, no. 1, pp. 1206–1209, 2019.

[16] M. Kleiner, D. Brainard, and D. Pelli, "What's new in Psychtoolbox-3?," *Perception*, vol. 36, pp. 1–16, 2007.

[17] S. Standard, "High dynamic range electro-optical transfer function of mastering reference displays," *SMPTE ST*, vol. 2084, no. 2014, p. 11, 2014.

[18] P.-C. Hung, "Colorimetric calibration in electronic imaging devices using a look-up-table model and interpolations," *Journal of Electronic imaging*, vol. 2, no. 1, pp. 53–61, 1993.

[19] V. Sitzmann, J. N. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein, "Implicit Neural Representations with Periodic Activation Functions," in *Proc. NeurIPS*, 2020.

[20] D. H. Brainard, D. G. Pelli, and T. Robson, "Display characterization," *Signal Process*, vol. 80, pp. 2–067, 2002.

[21] M. J. Vrhel and H. J. Trussell, "Color device calibration: A mathematical formulation," *IEEE Transactions on Image Processing*, vol. 8, no. 12, pp. 1796–1806, 1999.

[22] G. D. Finlayson and M. S. Drew, "Constrained least-squares regression in color spaces," *Journal of electronic imaging*, vol. 6, no. 4, pp. 484–493, 1997.

[23] G. Hong, M. R. Luo, and P. A. Rhodes, "A study of digital camera colorimetric characterization based on polynomial modeling," *Color Research & Application*, vol. 26, no. 1, pp. 76–84, 2001.

[24] G. D. Finlayson, M. Mackiewicz, and A. Hurlbert, "Color correction using root-polynomial regression," *IEEE Transactions on Image Processing*, vol. 24, no. 5, pp. 1460–1470, 2015.

[25] "ColorChecker (Macbeth) Chart." https://poynton.ca/notes/color/GretagMacbeth-ColorChecker.html.

[26] A. Mustafa, P. Hanji, and R. Mantiuk, "Distilling style from image pairs for global forward and inverse tone mapping," in *Proceedings of the 19th ACM SIGGRAPH European Conference on Visual Media Production*, pp. 1–10, 2022.