

## A DISPLAY MODELLING

### A.1 Modelling the Point Spread Function

We model the point spread function (PSF) as a Lorentzian,

$$\text{PSF}(x) = \alpha \left[ 1 + \left( \frac{x - x_0}{.5\gamma} \right)^2 \right]^{-1} + \epsilon,$$

where the input variable  $x$  is the distance from an illuminated pixel, and  $\alpha, \gamma, \epsilon, x_0$  are learned parameters. Because the Quest Pro BLU has significantly lower spatial resolution than the displayed image (as is the case for most LC displays), LED light can pass through the subtractive filter layer even when pixel intensities are set to 0. Perceptually, this leads to artifacts such as flare or decreased contrast in LC display systems [Reinhard et al. 2010]. Methods have been proposed to reduce these artifacts through deconvolution using hardware-accurate measurements of the PSF. In this work, we model the PSF in order to estimate plausible backlight LED driving values, but do not otherwise account for these additional complex artifacts.

### A.2 Local Dimming Simulation

The local dimming algorithm of the Meta Quest Pro display is not publicly available. As such, we implement a proxy local dimming algorithm simply to explore the power savings of such a display.

The image formation model of an LC display can be described by LC pixel intensities and backlight driving values,

$$\mathcal{I} = \mathcal{I}_{LC}\mathcal{B}. \quad (14)$$

where  $\mathcal{I}_{LC}$  are the LC panel pixel intensities and  $\mathcal{B}$  is the result of blurring of BLU LEDs due to LC display optical components, which can be physically approximated by convolving the display PSF with the BLU LED driving values,  $\mathcal{B} = \mathcal{W}\mathcal{d}$ . The matrix  $\mathcal{W}$  has shape  $n \times m$ , where  $n = w \times h$  ( $w, h$  are width and height of the displayed image) and  $m$  is the number of LEDs in  $d$ , and describes the PSF at the corresponding LED positions. Eq. (14) implies an inverse relationship between BLU driving values and LC pixel values, which means that decreasing backlight LED luminance can be compensated by an increase in LC pixel transmissivity.

To determine LED driving values for a local dimming backlight, we compute an approximate deconvolution of the blur due to diffusers and other optical components in the LC by solving a constrained least squares optimization problem,

$$\min_d \|\mathcal{B}^* - \mathcal{W}\mathcal{d}\|, 0 \leq d_i \leq 1.$$

$\mathcal{B}^*$  is the target BLU that is being approximated by solving for the LED driving values,  $d$ . The target BLU is computed by downsampling the target image to the resolution of the BLU, setting each pixel value to the maximum pixel intensity of the downsampled patch, and then scaling to photometric units [Trentacoste et al. 2007a]. In practice, heuristics are used to solve this optimization for real-time computation of  $d$ . In this work, we use the simplification described by Trentacoste et al. [2007a,b],

$$d_j = \frac{\mathcal{B}_j^* - \sum_{i \in N} \mathcal{W}_{ji} \mathcal{B}_i^*}{\mathcal{W}_{jj}},$$

where  $N$  is a neighborhood of BLU LEDs.

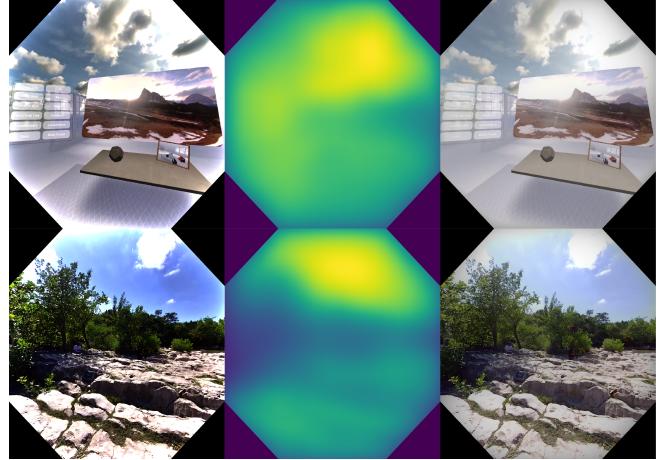


Fig. 10. *Local dimming simulation*. LC pixel intensities  $\mathcal{I}_{LC}$  (left) are used to filter light from the BLU  $\mathcal{B}$  (middle) to produce the displayed image (right).

Example images of the BLU and LC images are displayed in Figure 10 for the VR Scene 1 (top row) and Sculpture scenes (bottom row). Brighter colors in the BLU image (middle column) represent higher LED driving value. The right column images are simulations of what will be displayed as a result of applying Eq. (14), by multiplying LC pixel intensities (column 1) by BLU responses (column 2) in Figure 10.

## B DISPLAY MAPPING TECHNIQUES

We provide additional detail and discussion for display mapping techniques described in Section 4.

### B.1 Brightness Rolloff

The general form of the Gaussian rolloff curve in Equation (8) is

$$y = \exp(-\beta\phi^2)$$

where  $y$  is relative luminance,  $\phi$  is retinal eccentricity, and  $\beta$  is a constant controlling the minimum curve value. Solving for  $\beta$  so that  $\alpha$  becomes the minimum value at the edge of the display FOV,

$$\begin{aligned} \exp(-\beta * (\text{FOV}/2)^2) &= 1 - \alpha \\ \beta &= -4 \ln(1 - \alpha) / \text{FOV}^2. \end{aligned}$$

### B.2 Dichoptic Dimming

This display mapping technique could lead to unintended percepts, such as the Pulfrich Effect, which can enhance depth cues due to slower signal processing times for lower-luminance images [Doi et al. 2023] or cause binocular rivalry [Asano and Wang 2024; Wang et al. 2023; Wolfe 1983].

## C DEVICE SPECIFICATIONS

We list relevant device specifications in Table 2. To our knowledge, there are no publicly available specifications related to the Quest

Table 2. Relevant device specifications.

	Meta Quest Pro	HTC VIVE Pro Eye
<b>Resolution</b>	1400 × 1660	1800 × 1920
<b>PPD (horizontal/vertical)</b>	19.17/18.83	14.58/13.36
<b>Field of View</b>	106° × 96°	98° × 98°
<b>Eye Tracker Frequency</b>	–	120 Hz
<b>Eye Tracker Accuracy</b>	1.249° – 1.813°	0.5° – 1.1°

Pro eye tracking frequency. We use reasonable estimates from the literature [Wei et al. 2023] for the eye tracking accuracy.

## D USER STUDY DETAILS

We include additional details of the user study software implementation and data processing.

### D.1 Stimuli

The scenes used in the pilot studies and main study were of resolutions 4096 × 4096 for the hand-crafted scenes, and 5376 × 5376 for the LIVE-FBT-FCVR scenes. Hand-crafted scenes were captured using the Unity Recorder, which has a maximum capture resolution of 4096 × 4096 for 360° stereoscopic video. The LIVE dataset scenes contain natural motion (e.g. humans walking, clouds moving) and the virtual scenes contain UI panels with scrolling text and a dynamic fly-through video. Panoramic images of each scene are displayed in Figure 11.

### D.2 Pilot Studies

*Pre-Pilot.* The pre-pilot stimuli were set using a QUEST adaptive staircase procedure, with thresholds measured using a Weibull-shaped psychometric function,

$$\Psi(x) = \delta\gamma + (1 - \delta) \left[ 1 - (1 - \delta) \exp(-10^{\beta(x-T+\epsilon)}) \right],$$

as implemented in the open source PsychoPy package [Peirce et al. 2019]. To select the stimuli magnitudes to be used in pilot 1, the inverse of the Weibull curve was evaluated at 1, 2, and 3 JND. The psychometric functions for each participant are displayed in Figure 12, with the first row corresponding to the Weibull curves for participant 1 (P1), and the second row for participant 2 (P2). The blank plots indicate that only one of the two users participated in the pre-pilot for the specific display mapping.

Notably, the pre-pilot was conducted on two of the authors. As a result, the 1 JND thresholds are lower than those in the main study, which was conducted on naive users, for all display mappings except for whitepoint shift. This result is understandable, because the expert participants were aware of the types of display mappings used and more sensitive to the techniques.

*Pilots 1 & 2.* After each pilot study, new stimulus magnitudes were selected by fitting a best-fit line to the stimulus magnitude vs. JOD data from the previous pilot. The results of this procedure after pilot 1 and pilot 2 are visualized in Figure 13 (first and second row correspond to pilot 1 and 2, respectively). Because many of the magnitudes at 1 JND in the pre-pilot were close to the 1 JOD value in the first pilot, we kept them the same in the second pilot.

Table 3. Variable strengths for each magnitude of the display mapping techniques used in the main study. The reported numbers correspond to the  $\alpha$  values described in Section 4 of the main manuscript.

Display Mapping Technique	Level 1	Level 2	Level 3
Uniform Dimming	0.17	0.32	0.45
Luminance Clipping	0.22	0.35	0.51
Brightness Rolloff	0.64	0.75	0.88
Dichoptic Dimming	0.33	0.45	0.56
Whitepoint Shift	1.76	3.46	5.2
Color Foveation	0.58	0.76	0.93

Dichoptic dimming is missing in the first pilot data because it was added after the first pilot was completed.

### D.3 Main Study

The magnitudes used in the main study are displayed in Table 3. We use a Python implementation of ASAP [Mikhailuk et al. 2021], and communicate with the rendering application in Unity to update stimuli at each trial. The main study data scaled to JODs is displayed in Figure 14.

## E JOD VS. POWER SAVED TRANSFER FUNCTION

### E.1 Fitting the Psychometric Function

Psychometric functions were fit to the main study JOD data. The data was first converted to units of percentage preference, as described by Mantiuk et al. [2021]. This conversion is also visualized in Figure 15. A Weibull function was fit to this data with additional control points added based on the assumptions discussed in Section 7.1.1. Namely, points at ( $\alpha = 0$ , 0 JOD) and ( $\alpha_{\max} << -3$  JOD) were added during the curve-fitting process. The psychometric curves for two of the methods, color foveation and whitepoint shift, were not adjusted using this procedure because the study data already produce this saturation behavior. For example, power savings for color foveation plateau around 23% (for OLED display), because this is the maximum power savings that the method can achieve. The psychometric curves for each method are visualized in Figure 16. Optimization error however, with respect to the original three data points from the main study, may increase slightly after fitting the curve to the two additional control points.

### E.2 Relative Power Computation

We computed power consumption on two salient frustums from the 360° videos used in the main study. These regions were selected based on areas where participants spent the most time fixating, using the collected head- and eye-tracking data from the main study. To find salient regions, we used the software package from Sitzmann et al. [2017] which computes a saliency map given normalized fixation coordinates.

Because the contribution of the LC panel does not vary much with content, we discard it from the power consumption measurements as discussed in Section 3.1. Additionally, the static power consumption ( $\delta$  in Equations (1) and (5)) is excluded.



Fig. 11. Scenes used in the user study. Five scenes were used in the main user study. Left three videos original from the LIVE-FBT-FCVR database, and the right two videos were hand-crafted by the authors.

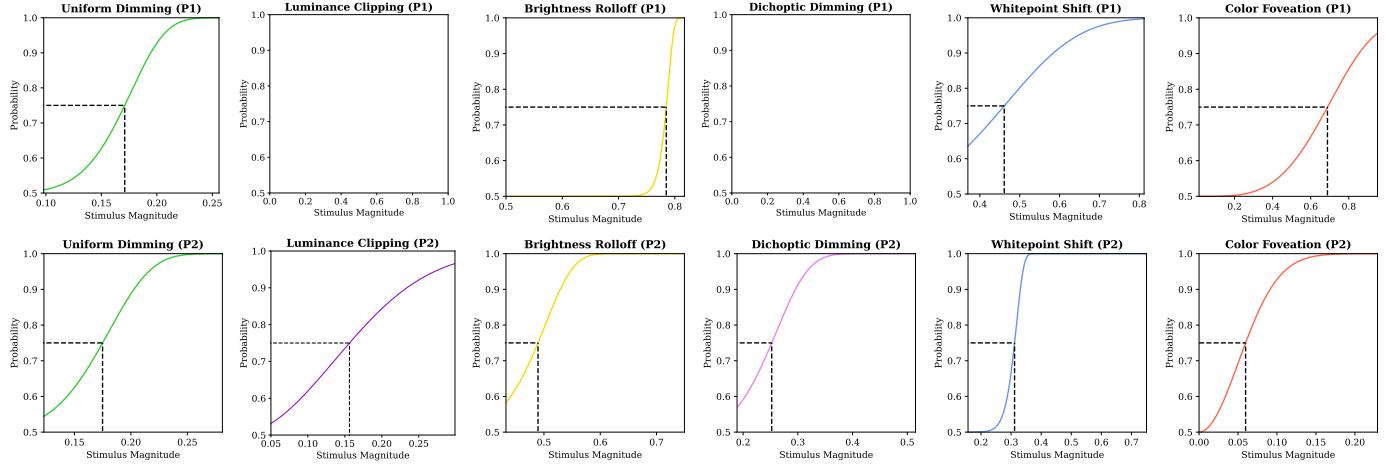


Fig. 12. Pre-pilot staircase results. Psychometric curves fit to the pre-pilot study results for both participants.  $x$ -axis is stimulus magnitude and  $y$ -axis is probability of detection. Empty plots were conditions not done by P1.

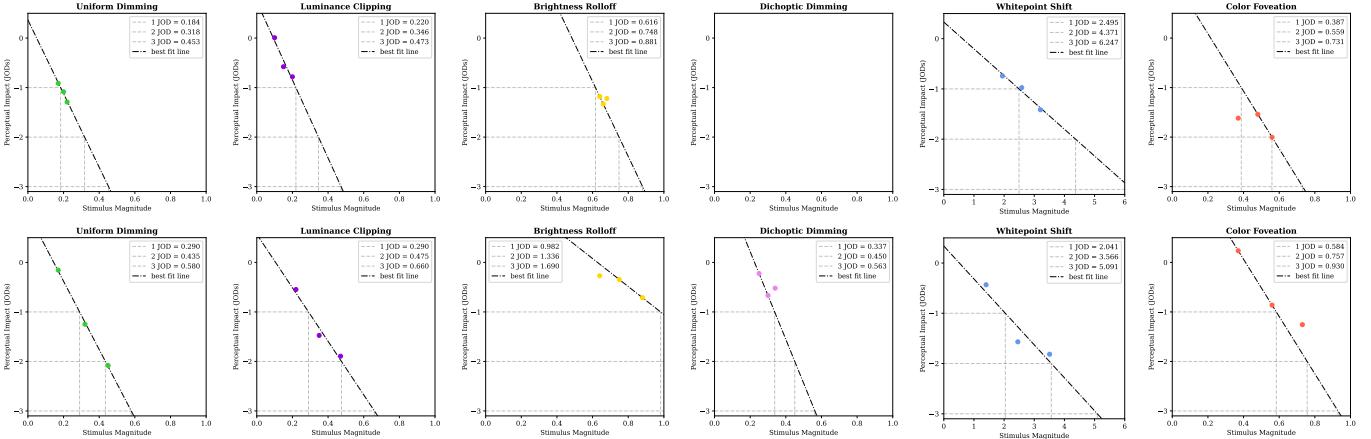


Fig. 13. Extrapolating pilot data. The top row shows results after the first pilot, and bottom row results after the second pilot.  $x$ -axis represents stimulus magnitude,  $\alpha$ , and the  $y$ -axis shows perceptual impact (JODs). Missing plot refers to a condition added after the first pilot.

### E.3 Validation Study

This experiment aims to validate the accuracy of the assumptions made when fitting psychometric curves to the main study data.

**Hardware.** We chose to conduct the study with the HTC VIVE Pro Eye, with stimuli magnitudes corresponding to 20% and 40% power saving targets as computed by the OLED power model. A similar study could have been conducted for the Quest Pro's LC

display model, but we decided that using the OLED model would allow us to test more display mapping techniques, such as color foveation.

**Participants.** We recruited  $N = 5$  naive participants, none of whom participated in the pilot or main studies. The same eye tracking calibration and vision testing were done as in the main study.

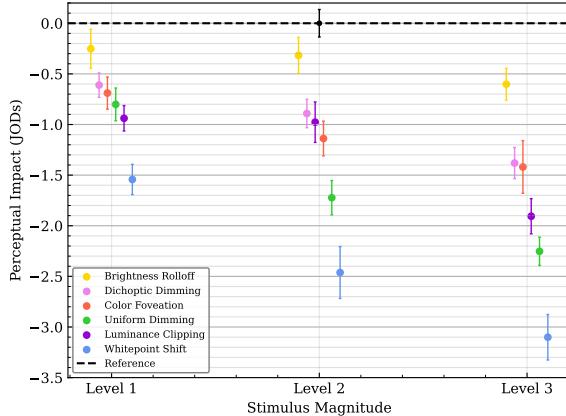


Fig. 14. *Main study results.* The main study data was scaled to units of JODs ( $y$ -axis) for three increasing levels of magnitude ( $x$ -axis). Vertical error bars represent 95% percent confidence intervals.

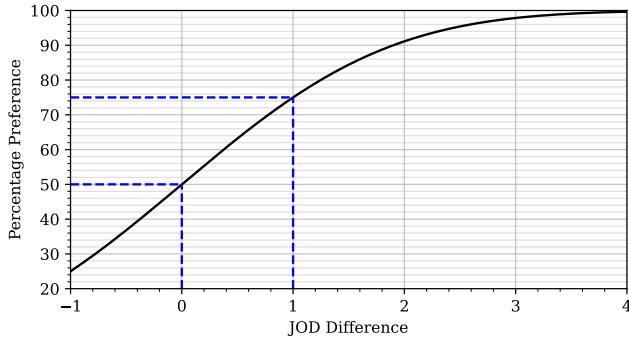


Fig. 15. *JODs vs. % preference.* We replicate the plot from Mantiuk et al. [2021], which maps JODs to interpretable units of percentage preference. The difference in 1 JOD between two techniques, A and B, indicates 75% percent selection of A over B.

*Stimuli.* Two previously unused scenes from the LIVE-FBT-FCVR Database were presented in this study (Bar and Bridge scenes).

*Experimental Procedure.* Participants completed a 2IFC study with the same task as the main study. Unlike the main study, we conducted this experiment with a full pairwise design within the power-saving conditions. That is, for stimuli magnitudes within each power-saving level (20% and 40%), all combinations of scenes and display mappings were directly compared against each other. For 20% power savings, this resulted in  $2 \text{ scenes} \times (5 \text{ display mappings} + 1 \text{ reference}) = 12$  conditions and for 40% power savings,  $2 \text{ scenes} \times (4 \text{ display mappings} + 1 \text{ reference}) = 10$  conditions. The full pairwise study contained  $C(12, 2) + C(10, 2) = 111$  conditions. We opted to discard whitepoint shift from this study because its perceptual impact (JODs) at 20% and 40% savings is very large, making it redundant. In total, the study took approximately 30 minutes to complete. See Figure 17 for visualization of validation study conditions.

Table 4. Validation study vote counts. Two columns under each participant represent vote counts for the 20% power saving condition and 40% conditions, respectively. The second column for color foveation is left blank as it was not studied in the 40% condition trials.

	P1	P2	P3	P4	P5	SUM						
• Reference	14	13	19	16	17	13	16	14	18	13	84	69
• Uniform Dimming	17	9	11	8	12	5	7	3	13	5	60	30
• Luminance Clipping	0	0	2	3	2	0	4	1	5	3	13	7
• Brightness Rolloff	12	14	16	12	17	14	15	11	13	14	73	65
• Dichoptic Dimming	13	4	10	1	10	8	9	11	8	5	50	29
• Color Foveation	4	—	2	—	2	—	9	—	3	—	20	—

*Results.* We visualize the results as a plot, in Figure 17, of the JOD values evaluated by the psychometric curves vs. the validation study data scaled to JODs. The identity line represents a hypothetical perfect match between the psychometric fit and validation. Spearman’s rank-order correlation analysis showed a strong positive correlation between JOD scores predicted by the transfer functions and those collected from the validation study, ( $r = 0.943, p < .005$  and  $r = 0.999, p << .001$ , for 20% and 40% savings, respectively). The tabulated study results in Table 4 show the vote counts for each display mapping technique.

## F WALL-PLUG EFFICIENCY DATASET

We use data from five separate datasets aggregated by Hahn [2016] – Narukawa et al. [2010], Hahn et al. [2008], Peter et al. [2008], Schiavon et al. [2013], and Steigerwald et al. [1997] – to determine wall-plug efficiency (WPE) in Equation (13), displayed in Figure 18. In order to sample from this data at continuous wavelengths  $\lambda$ , we fit a linear function to the data, relating  $\lambda$  to WPE.

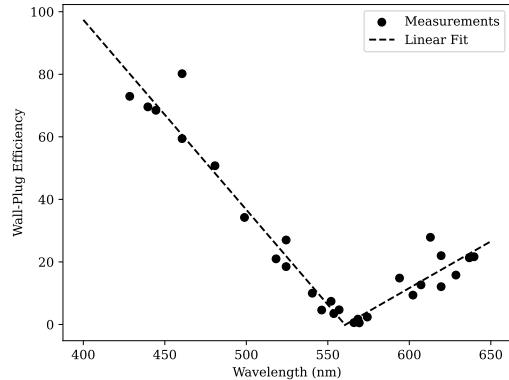


Fig. 18. *Wavelength versus wall-plug efficiency data.* Wall-plug efficiency data for LEDs is aggregated across five datasets, and lines are fit to the measurements.

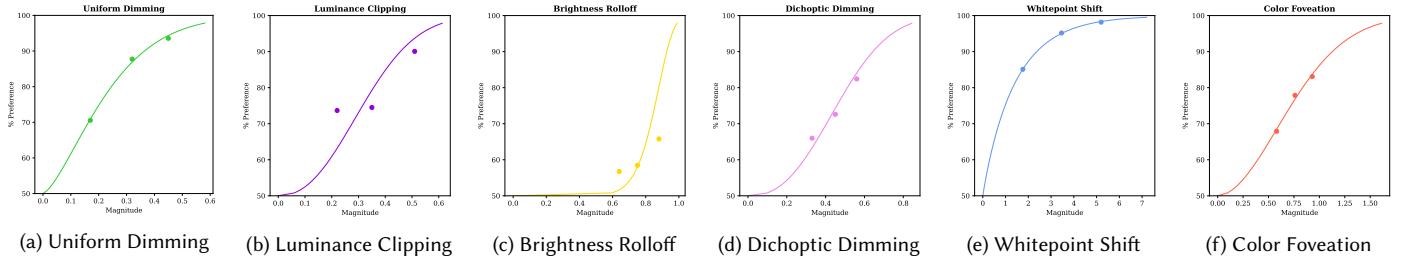


Fig. 16. *Psychometric fitting.* We fit a psychometric function to the main user study data after perceptual scaling and conversion to percentage preference.

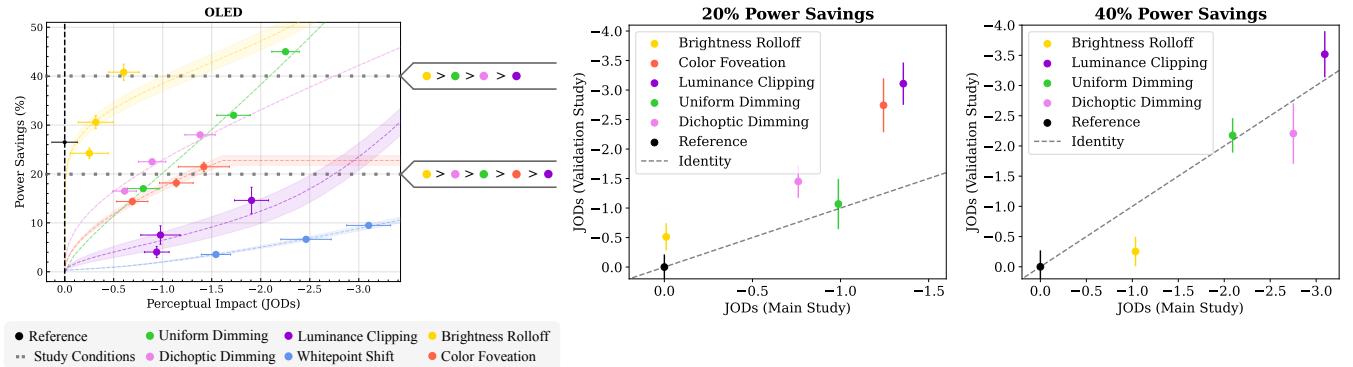


Fig. 17. *Validation study conditions and results.* The two conditions used in the validation study, 20% and 40%, are visualized as gray dotted lines (left plot). The ranking of display mappings is displayed as circle markers to the right of the first figure. JOD scores predicted by the psychometric curve fit to the main study data ( $x$ -axis) against JOD scores from the validation study ( $y$ -axis) are displayed (right two figures). Vertical error bars show 95% confidence intervals.