PEA-PODs: Perceptual Evaluation of Algorithms for Power Optimization in XR Displays

KENNETH CHEN, New York University, USA and Meta, USA
THOMAS WAN, Meta, USA
NATHAN MATSUDA, Meta, USA
AJIT NINAN, Meta, USA
ALEXANDRE CHAPIRO*, Meta, USA
QI SUN*, New York University, USA

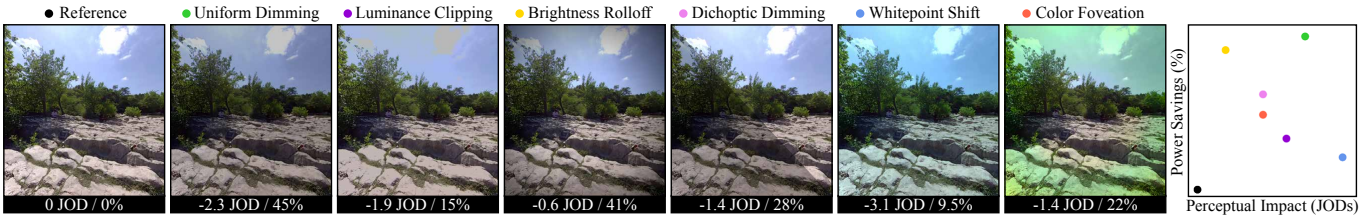


Fig. 1. *Display mapping techniques and study results.* We studied six power-saving display mapping algorithms. These techniques are used in traditional display power reduction as well as recently-proposed methods for VR displays. Just objectionable difference (JOD) scores provide a unified measure of the magnitude of perceptual impact, and percentage values represent relative display power savings (OLED display shown here). Colors ● ● ● ● ● ● ● correspond to plots throughout the manuscript. Optimal techniques have low perceptual impact (perceived close to reference ● in JODs), but provide big power savings.

Display power consumption is an emerging concern for untethered devices. This goes double for augmented and virtual extended reality (XR) displays, which target high refresh rates and high resolutions while conforming to an ergonomically light form factor. A number of image mapping techniques have been proposed to extend battery usage. However, there is currently no comprehensive quantitative understanding of how the power savings provided by these methods compare to their impact on visual quality. We set out to answer this question.

To this end, we present a **p**erceptual **e**valuation of **a**lgorithms (PEA) for **p**ower **o**ptimization in XR **d**isplay**s** (PODs). Consolidating a portfolio of six power-saving display mapping approaches, we begin by performing a large-scale perceptual study to understand the impact of each method on perceived quality in the wild. This results in a unified quality score for each technique, scaled in just-objectionable-difference (JOD) units. In parallel, each technique is analyzed using hardware-accurate power models.

The resulting JOD-to-Milliwatt transfer function provides a first-of-its-kind look into tradeoffs offered by display mapping techniques, and can be directly employed to make architectural decisions for power budgets on XR displays. Finally, we leverage our study data and power models to address important display power applications like the choice of display primary, power implications of eye tracking, and more[1].

CCS Concepts: • **Hardware → Power estimation and optimization**; • **Computing methodologies → Perception**; **Virtual reality**; **Augmented reality**.

Additional Key Words and Phrases: Display Power, Visual Perception, VR/AR

ACM Reference Format:
Kenneth Chen, Thomas Wan, Nathan Matsuda, Ajit Ninan, Alexandre Chapiro, and Qi Sun. 2024. **PEA-PODs**: **P**erceptual **E**valuation of **A**lgorithms for **P**ower **O**ptimization in XR **D**isplay**s**. *ACM Trans. Graph.* 43, 4, Article 67 (July 2024), 17 pages. https://doi.org/10.1145/3658126

1 INTRODUCTION

Modern augmented and virtual extended reality (XR) devices require high resolution, fast refresh rates, wide fields of view, and high dynamic range displays toward the ultimate goal of passing the "visual Turing test" [Matsuda et al. 2022; Zhong et al. 2021]. However, these requirements lead to higher power consumption not only for computation costs, but also for running the display itself. This is problematic for stand-alone XR devices with compact and ergonomic form factors that limit battery size. Balancing the conflicting requirements of high image quality and display power efficiency is a critical problem and requires careful optimization in both hardware and software.

The display alone can consume up to 40% of the power budget of consumer electronics devices [Anand et al. 2011; Tan et al. 2013; Wee et al. 2018]. To remedy this, a number of power-saving algorithms have been proposed which aim to maintain the image's intended appearance while reducing display power use. Throughout this

*Co-corresponding authors.

Authors' addresses: Kenneth Chen, New York University, New York, USA and Meta, Sunnyvale, USA, kennychen@nyu.edu; Thomas Wan, Meta, Sunnyvale, USA, thomaswan@meta.com; Nathan Matsuda, Meta, Remond, USA, nathan.matsuda@meta.com; Ajit Ninan, Meta, Sunnyvale, USA, ajitninan@meta.com; Alexandre Chapiro, Meta, Sunnyvale, USA, alex@chapiro.net; Qi Sun, New York University, New York, USA, qisun@nyu.edu.

0730-0301/2024/7-ART67
https://doi.org/10.1145/3658126

[1]Source code and data is available at github.com/NYU-ICL/pea-pods.

work, we refer to these methods as *display mapping* techniques. These techniques range from simple solutions, like dimming the display [Choi et al. 2002; Gatti et al. 2002], to complex perceptually-inspired color mappings [Duinkharjav et al. 2022b].

When applying these techniques, the resulting display mapped content may contain visible distortions in terms of contrast, color accuracy, or brightness, but no unified analysis of the perceptual impact of this family of techniques has ever been attempted. Predicting and quantifying the trade-offs between the power savings and perceivable image quality loss in the wild across different display types will provide practical guidance for display characteristics, rendering algorithms, and overall XR system design.

In this work, we measure - for the first time under a unified architecture - the perceptual impact of different power-saving methods across common display types, free-form viewing, and employing natural imagery. Interpreting the acquired power-perception data provides quantitative guidance on optimizing display mapping techniques that adapt to display architecture and target battery life spans. To this end, we first characterized the power consumption profiles for common display technologies used in commercial XR devices, including organic light emitting diode (OLED), local and global backlit liquid crystal (LC) displays. We collected subjective judgments of distortion strength for six power-saving methods with a large-scale subjective user study. Our results are scaled in unified perceptual units of just-objectionable-difference (JOD). The scaled data is then used to establish a transfer function between JODs and milliwatts. Finally, we demonstrate learnings and power-aware XR applications with adaptive rendering optimization, multi-primary display, and power-aware eye-tracking control. Source code and data is available at github.com/NYU-ICL/pea-pods.

2 BACKGROUND AND RELATED WORK

2.1 Display Properties and Power Consumption

The power consumption of a screen when displaying the same content varies significantly depending on the underlying architecture. Due to this, some display mapping techniques that are optimal for one display type may be useless for another. In particular, for emissive displays, such as organic light-emitting diode (OLED), individual pixels produce their own light, and each pixel's power profile is proportional to its intensity [Dong and Zhong 2011b]. In contrast, transmissive displays rely on a separate light source, such as a backlight unit (BLU), to illuminate a subtractive filter array. In this case, the power consumption can be dominated by BLU intensity, and have little to no dependence on individual pixel values [Cheng and Pedram 2004]. In Section 3 we examine in detail one XR consumer device in each of these two categories: the emissive OLED HTC Vive Pro Eye [HTC 2020], and the transmissive liquid-crystal (LC) Meta Quest Pro [Rao et al. 2023].

2.2 Display Power Optimization Solutions

The display is one of the most power hungry elements of standalone electronics, such as XR headsets, mobile phones, laptops, etc. Consequently, a number of techniques aiming to reduce their power usage exist. Arguably the most common "power-saving mode" for mobile devices is uniform dimming, scaling down display brightness linearly [Cheng and Pedram 2004; Choi et al. 2002; Gatti et al. 2002]. To avoid globally dimming content, clipping high-luminance regions can preserve the brightness in most of the image at the expense of detail and brightness in highlight regions [Kerofsky and Daly 2006].

In self-emissive LED-based displays, color remapping can save power by shifting pixel colors based on differences in primary efficiencies [Dong and Zhong 2011a]. Different luminous efficiencies of each primary LED, computed from its spectral power distribution, can contribute to imbalances in power consumption by different colors. Wide field displays can take advantage of the limited perceptual acuity of the human visual system through eye-tracked methods such as peripheral dimming [Kim and Lee 2020] or foveated color remapping [Duinkharjav et al. 2022b].

These display mapping proposals in prior art, however, measure perceptual impact in many different ways, making it difficult to determine which techniques are better in terms of both power savings and subjective quality. For example, some works measure image quality by using quantitative metrics [Kerofsky and Daly 2006], or psychophysical thresholds [Kim and Lee 2020], by conducting qualitative surveys [Dong and Zhong 2011a], or by comparing against a baseline [Duinkharjav et al. 2022b]. Wee et al. [2018] study several display mapping techniques (color shift, peripheral dimming, and a "Tron" mode in which images are filtered to contain only salient edges) across several display types. However, the authors conducted a *task completion* evaluation, rather than measuring the *visibility* of introduced distortions, concluding that all methods performed equally well in terms of usability and task completion time. Our goal is to fill in this gap in the literature by studying the perceptual impact of common display mapping techniques with a unified framework and different display types.

2.3 Perceptual Quality Measurement and Optimization

Measuring and quantifying perceptual quality of distorted content is a central topic in visual computing. Existing models may target low-level features, such as contrast sensitivity [Mantiuk et al. 2022], or high-level semantics [Fu et al. 2023]. With different display and content types, some methods extend beyond static images [Yu and Grauman 2015; Zhang et al. 2018], to dynamic [Seshadrinathan and Bovik 2007], or foveation-aware [Mantiuk et al. 2021] videos. Most models of this sort are based on simple stimuli and controllable viewing conditions. However, the highest standard of quality evaluation involves direct subjective testing for the targeted application.

Perceptual models have been leveraged to optimize the quality of XR displays, including improved rendering speed [Krajancich et al. 2023], reduced data volumes [Chen et al. 2022; Kaplanyan et al. 2019], or reduced target-searching times [Duinkharjav et al. 2022a]. We examine this type of optimization in the context of saving display power. Producing optimized solutions is important, because systems benefit from targeted goals, such as more aggressive savings in low battery modes: sacrificing the least possible amount of perceptual quality for the greatest power gain. To do this, we need to be able to perform a quantitative analysis. We present the first study that progressively evaluates and interprets the connection between

power consumption and perceptual quality at various scales, taking into account key considerations like display architecture.

3 CHARACTERIZATION OF XR DISPLAY POWER

Display power consumption is dependent on the image content being displayed. The design of optimal power-saving methods for each display type requires accurate measurement and characterization of how a display's power consumption varies with pixel intensity distribution. As representatives of the two display types, we chose two commercially available VR headsets: Meta Quest Pro and the HTC VIVE Pro Eye for liquid-crystal (LC) and OLED, respectively. Below, we describe our setup to establish accurate electrical power measurements and display-dependent power consumption models.

3.1 Liquid-Crystal (LC) Displays

Measurement. Isolated power measurements of the Meta Quest Pro's LC display module were collected by measuring current and voltage across power supply rails on the display board. Voltage drop across these rails is measured using a Keysight InfiniiVision DSOX3034T oscilloscope with passive probes placed across shunt resistors, and current is computed using Ohm's Law. All measurements were time-averaged over 5 seconds.

To measure BLU power consumption, we uniformly sampled 21 test relative luminance values, y_i. For each sample, all BLU LEDs were set to y_i during measurement. For LC power measurement, we sampled 20 test colors uniformly for each RGB sub-channel (60 colors in total + 1 for black). We set every pixel to the test color when making power measurements. An image of our measurement setup is included in Figure 2a.

Content-Dependent Power Model. We regressed analytical power models that map pixel color to physical units of LC power consumption in milliwatts, mW, derived from the measured data described in the previous paragraph. From these measurements, we found a strong positive relation between BLU power consumption and the relative luminance of LEDs, which we model as a linear function of relative luminance, y,

$$\mathcal{M}_{\mathcal{B}}(y) = \alpha y + \delta \tag{1}$$

with free parameters α and δ. The correlation coefficient of the BLU model regression is $r^2 = .99$. Figure 3 (right) includes a plot of the measured BLU power consumption (y-axis) as a function of relative luminance (x-axis), and the regressed model (dashed line). The LC data can be described as a per-channel summation of 2nd-order polynomial functions of pixel intensity,

$$\mathcal{M}_{\mathrm{LC}}(\mathbf{c}) = \sum_{p=0}^{3} \alpha_p \mathbf{c}_p^2 + \delta_p, \tag{2}$$

where α_p, δ_p are free parameters, p is the index of an RGB primary, and $\mathbf{c}$ is a linear RGB pixel color. The root mean square error and mean absolute percentage error of the LC model regression are 0.31 and 0.92% respectively. Figure 3 (left) visualizes the LC power model for each RGB sub-channel.

The total display power consumption is modeled as the sum of contributions from the BLU and LC panel. Notably, the variation in BLU power consumption is more than an order of magnitude greater than the LC power (>570mW variation for BLU compared to <20mW for LC). Because of its negligible contribution, in following computations of power consumption we ignore the LC panel, similar to prior work [Cheng and Pedram 2004].

Backlight Dimming Scheme. The spatial resolution of the BLU is much lower than the displayed image. Individually-controllable BLU LEDs can thus take different intensities depending on the image processing algorithm employed. For instance, considering an LC display with a *global dimming* backlight, all BLU LEDs take the same driving value, which can be set to the maximum pixel intensity of the displayed image. The power consumption in the global dimming setting is therefore a function of the maximum pixel intensity,

$$\mathcal{P}(\mathcal{I}) = \mathcal{M}_{\mathcal{B}}(\max\{\mathbf{c} : \mathbf{c} \in \mathcal{I}\}) + \sum_{\mathbf{c}\in\mathcal{I}} \mathcal{M}_{\mathrm{LC}}(\mathbf{c}), \tag{3}$$

where $\mathcal{I}$ is an image with linear RGB pixel colors $\mathbf{c}$. In the *local dimming* setting, BLU LEDs are modulated individually. Power consumption is represented as a function of LED driving values, d_i,

$$\mathcal{P}(\mathcal{I}) = \frac{1}{N}\sum_{i=1}^{N} \mathcal{M}_{\mathcal{B}}(d_i) + \sum_{\mathbf{c}\in\mathcal{I}} \mathcal{M}_{\mathrm{LC}}(\mathbf{c}), \tag{4}$$

where N is the total number of backlight LEDs. We determine LED driving values using a heuristic optimization procedure similar to Trentacoste et al. [2007a,b]. Accurate computation of driving values depends on the spatial location of BLU LEDs and the optical blur due to diffusers as well as other physical components of the LC display. This effect is measured by determining the light spread, or point spread function (PSF), of a single illuminated LED. The PSF was measured for the Meta Quest Pro display, and a Lorentzian was fit to the measurements (see supplementary Section A.1). Images of the backlight LED spatial arrangement and the PSF are displayed in Figure 2b and Figure 2c, respectively. More details of this local dimming algorithm are described in supplementary Section A.2.

3.2 Organic Light-Emitting Diode (OLED) Displays

OLED power is often modeled as a linear function of display primaries [Dong and Zhong 2011b; Duinkharjav et al. 2022b],

$$\mathcal{P}(\mathcal{I}) = \sum_{\mathbf{c}\in\mathcal{I}} \mathbf{p}^\intercal \mathbf{c} + \delta. \tag{5}$$

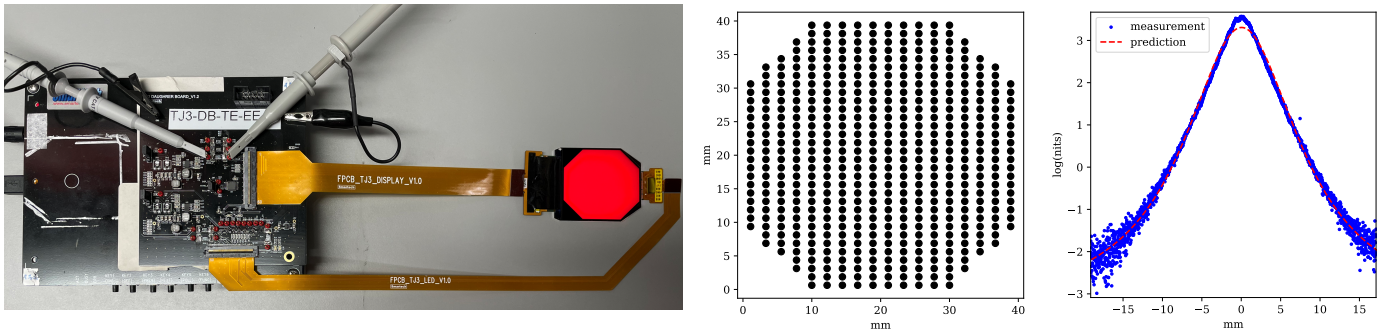
We adopt this function to model the HTC Vive Pro Eye display power consumption by leveraging the free parameters $\mathbf{p}, \delta$ from Duinkharjav et al. [2022b] for the same display.

4 DISPLAY MAPPING TECHNIQUES

We consolidated six display mapping techniques which are either commonly used in traditional mobile displays or have been recently proposed for XR displays. For each display mapping, we introduce a modulation factor, α, that increases the magnitude of each algorithm.

Uniform Dimming. (●) The first technique is uniform dimming, which linearly scales down image pixel values,

$$\mathbf{c}' = (1-\alpha)\mathbf{c}. \tag{6}$$



(a) display measurement setup (b) backlight LED arrangement (c) point spread function (PSF)

Fig. 2. *Meta Quest Pro display.* (a) We measure the power consumption of the Quest Pro LC display module by measuring voltage drop across shunt resistors. This figure includes the display module (displaying an all-red image), driver board, and oscilloscope probes. We visualize the arrangement, (b), of BLU LEDs (black circle) for the Quest Pro display. The point spread function (PSF) and physical measurements are displayed in (c). The x-axis is distance from a single illuminated LED in mm, and the y-axis is in units of log-nits.

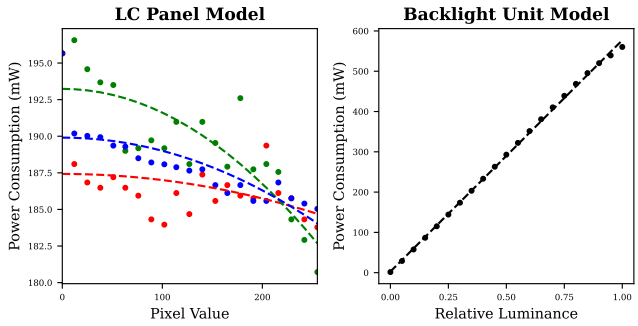


Fig. 3. *LC display power models.* A visualization of the LC (left) and backlight unit (BLU, right) power models. Points show physical measurements, and dashed curves are regressed model fits.

This approach is commonly used in low-battery modes for smart phones and other mobile display technologies, e.g., Windows adaptive brightness control [Microsoft 2022].

Luminance Clipping. (•) The next technique is luminance clipping, which clips the highest-luminance image features,

$$\mathbf{c}' = (1-\alpha)\frac{\mathbf{c}}{\mathcal{Y}(\mathbf{c})}, \text{ if } \mathcal{Y}(\mathbf{c}) > (1-\alpha). \tag{7}$$

The function $\mathcal{Y}$ maps pixel color $\mathbf{c}$ to relative luminance [BT 2002]. Fig. 4a visualizes the difference between luminance clipping and uniform dimming. While uniform dimming decreases the luminance of all pixels in an image, clipping only decreases luminance of highlights. Bright regions (e.g. the cloud) in Fig. 4a are dimmed in both cases, while the tree and rock remain at similar brightness compared to the reference for clipping. However, details are lost in clipped regions compared to uniform dimming.

Brightness Rolloff. (•) Brightness rolloff is an eye-tracked method which applies peripheral dimming using a Gaussian profile,

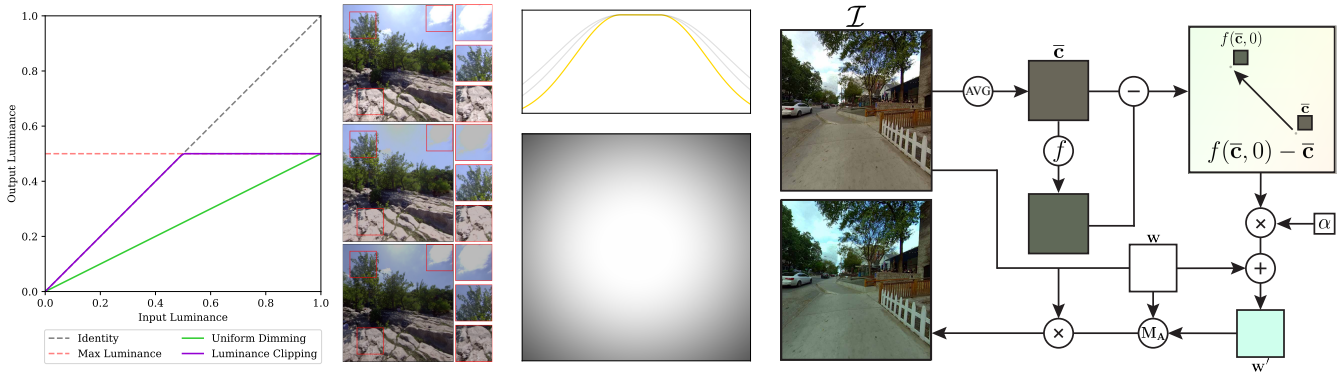
$$\mathbf{c}' = \exp\left(\frac{4\ln(1-\alpha)}{(\text{FOV}-\theta)^2}\phi^2\right)\mathbf{c}, \tag{8}$$

where FOV is the maximum field of view of the display (relevant device specifications in Supplement C). The retinal eccentricity of a pixel located at image coordinates (x, y) is computed as $\phi = \frac{1}{\text{ppd}}\sqrt{(x-\mathbf{g}_x)^2+(y-\mathbf{g}_y)^2}$, where $\mathbf{g}$ is the image-space gaze location and ppd is peak pixel density of the display [Mantiuk et al. 2021]. The foveal region, within $\theta = 10°$ eccentricity, is unmodified. A derivation of Equation (8) is in supplementary Section B.1. Prior work use a linear rolloff [Kim and Lee 2020], which can create visual artifacts due to the C^1 discontinuity [Moulden et al. 1988], while our Gaussian profile ensures C^1 continuity.

Fig. 4b illustrates the rolloff mapping applied to a white image at the maximum magnitude used in the user study ($\alpha = 0.88$), as well as a plot of the relative luminance as a function of eccentricity for the three stimulus magnitudes used in the study. The orange curve corresponds to the profile applied to the white image.

Dichoptic Dimming. (•) Techniques which use different rendering modalities for each eye in a binocular display have recently been proposed for tasks like tone mapping [Zhong et al. 2019]. In our implementation, rather than dimming the display equally for both eyes as done for uniform dimming, we only dim the display corresponding to one eye. Because the majority of the population is right-eye dominant (approx. 70% [Ehrenstein et al. 2005; Reiss and Reiss 1997]), we choose to dim the left eye only.

Color Foveation. (•) Human color perceptual acuity is highest in the fovea, and decreases with retinal eccentricity [Cohen et al. 2020; Hansen et al. 2009]. This characteristic can be taken advantage of in displays with non-uniform color efficiencies, such as OLED displays, by modulating pixel chromaticity in a power-aware fashion. We model the power-optimal color shift of $\mathbf{c}$ located at eccentricity ϕ



(a) uniform dimming vs. luminance clipping (b) brightness rolloff (c) whitepoint shift processing diagram

Fig. 4. *Display mapping techniques.* The x- and y-axes of (a) represent input and output luminance respectively, and the horizontal red dashed line is set at $\alpha = 0.5$. A visual comparison of reference, luminance clipping, and uniform dimming (top to bottom) is included with insets in red. (b) We visualize the brightness profile ((b) top; x-, y-axes are eccentricity, brightness) and rolloff applied to a uniform white image ((b) bottom). Whitepoint shift, (c), computes average color $\bar{\mathbf{c}}$ of an input image $\mathcal{I}$ which is passed to a perceptual model that computes a new power-optimal color, $f(\bar{\mathbf{c}}, 0)$. The whitepoint of $\mathcal{I}$ is shifted in the direction of a vector pointing from $\bar{\mathbf{c}}$ to $f(\bar{\mathbf{c}}, 0)$. Given source and destination whitepoints, a transformation matrix $\mathbf{M_A}$ is computed and applied to $\mathcal{I}$.

using the model described by Duinkharjav et al. [2022b],

$$f : (\mathbf{c}, \phi) \mapsto \underset{x \in \mathcal{M}_\Theta(\mathbf{c},\phi)}{\arg\min} \; \mathcal{P}(x), \tag{9}$$

where $\mathcal{M}_\Theta$ is a radial basis network with parameters Θ which returns a set of colors, in the DKL color space [Derrington et al. 1984], perceptually indistinguishable from $\mathbf{c}$ when viewed at eccentricity ϕ. The function f computes the power-minimal color within this set, where $\mathcal{P}$ is a display power model. Final color is computed as

$$\mathbf{c}' = \mathbf{c} + \alpha \left(f(\mathbf{c}, \phi) - \mathbf{c} \right). \tag{10}$$

Whitepoint Shift. (●) Similar to prior color remapping techniques [Dong and Zhong 2011a; Wee et al. 2018], the next technique leverages chromatic adaptation [Luo et al. 2000] to shift the display whitepoint to a more power-optimal whitepoint. Our algorithm applies Equation (9), assuming viewing at $\phi = 0°$ eccentricity. The original whitepoint is then transformed in the computed direction,

$$\mathbf{w}' = \mathbf{w} + \alpha \left(f(\bar{\mathbf{c}}, 0) - \bar{\mathbf{c}} \right), \tag{11}$$

where $\bar{\mathbf{c}} = \text{AVG}(\mathcal{I})$ is the average color in $\mathcal{I}$, $\mathbf{w}$ is the whitepoint under D65 illuminant, and $\mathbf{w}'$ is the whitepoint under a new illuminant. Given the source and destination whitepoints, we can compute a chromatic adaptation matrix by following a linear simplification of the Bradford chromatic adaptation transform [Lindbloom 2017],

$$\begin{aligned} \mathbf{M_A} &= \mathbf{B}^{-1}\mathbf{diag}(\boldsymbol{\gamma}' \oslash \boldsymbol{\gamma})\mathbf{B}, \\ \mathbf{c}' &= \mathbf{M_{xyz2rgb}}\mathbf{M_A}\mathbf{M_{rgb2xyz}}\mathbf{c}, \end{aligned} \tag{12}$$

where $\boldsymbol{\gamma} = \mathbf{Bw}$, $\boldsymbol{\gamma}' = \mathbf{Bw}'$, $\mathbf{B}$ is the Bradford matrix [Luo and Hunt 1998], taking colors from the XYZ to the LMS (cone response) space, and $\mathbf{M_A}$ is the chromatic adaptation transform. The symbol $\oslash$ represents element-wise division. Figure 4c visualizes this procedure.

5 PERCEPTUAL STUDIES

We aim to measure the subjective quality of different power-saving techniques described in Section 4. To do this, we conducted a series of perceptual studies with natural stimuli and free-form viewing. We determined appropriate stimulus magnitudes via pilot studies (Section 5.1), which were then used in our large-scale main study (Section 5.2).

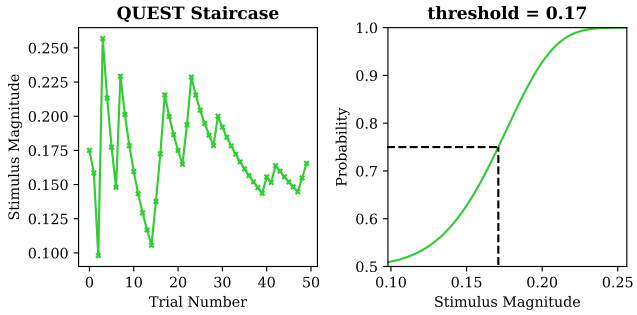


Fig. 5. *Pilot Study.* We conducted a pre-pilot to set initial magnitude values for each display mapping. The left plot shows the stimuli magnitudes as scheduled by QUEST, and the right plot shows the psychometric fit to the user responses (P1). The dashed line represents 75% detection (1 JND).

5.1 Pilot Studies: Setting Stimuli Magnitudes

Introducing invisible or obvious distortions would lead to trivial results. To avoid this, rather than choosing arbitrary technique parameters, we conducted several pilot experiments to find meaningful magnitudes for each of three increasing levels of intensity.

5.1.1 Pre-Pilot. We conducted a small-scale pre-pilot on expert participants ($N = 2$ authors) to set initial magnitudes of each power saving technique. A two-interval forced choice (2IFC) procedure using a QUEST adaptive staircase [Watson and Pelli 1983] was used. Participants were shown two videos, the reference and a display mapped test, and were tasked with selecting the one with better

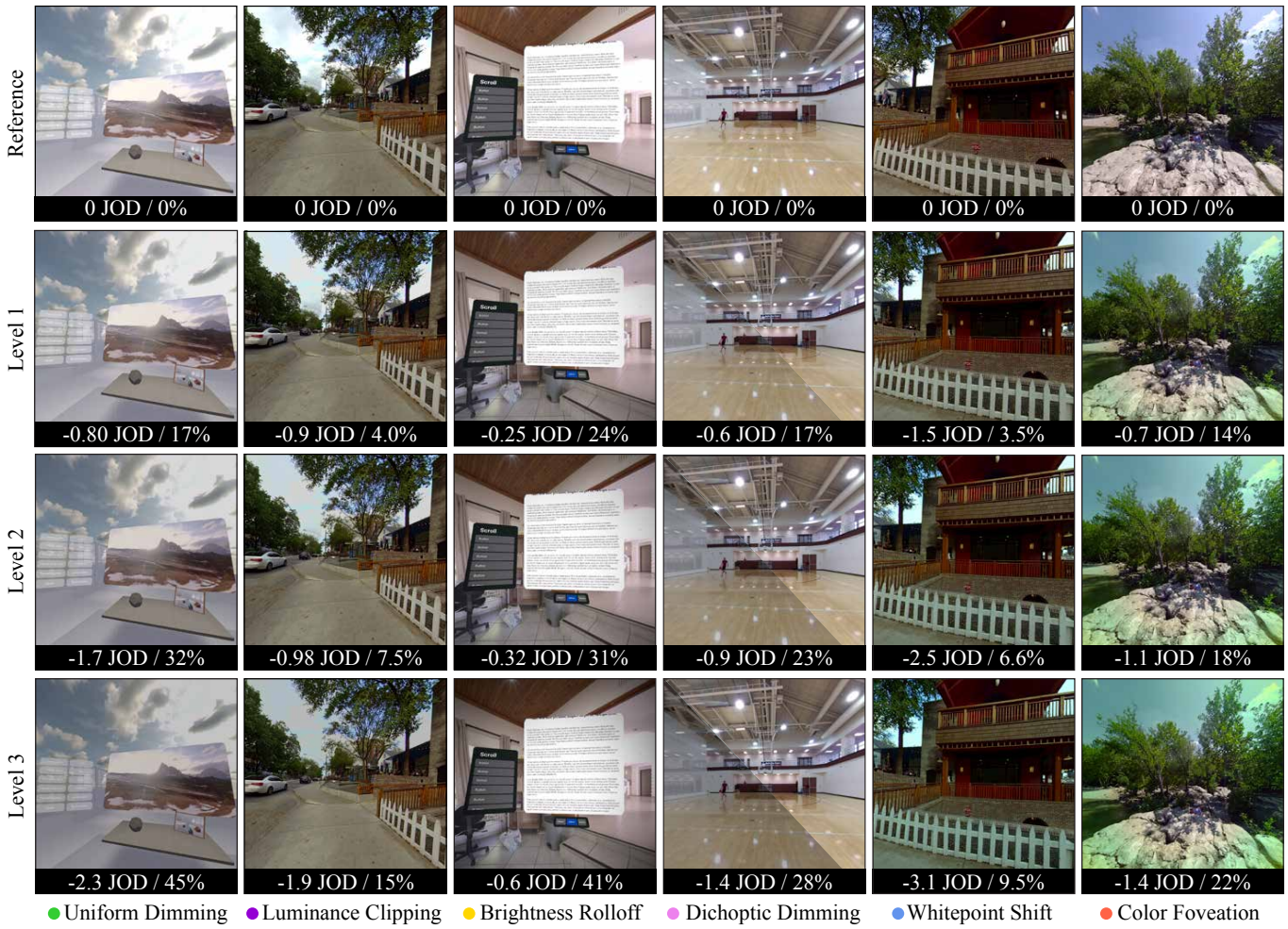


Fig. 6. *Study stimuli.* All display mappings are visualized at the three stimuli magnitudes used in the main study, with rows representing from top to bottom Reference, Level 1, Level 2, and Level 3 magnitudes. From left to right the scenes are *VR Scene 1*, *Street*, *VR Scene 2*, *Dunk*, *Street*, *Sculpture*. JOD / power savings (%) are displayed at the bottom of each image.

visual quality. In this setting, display mapped imagery is perceptually identical to the reference when participants select the modulated video 50% of the time (at chance). We estimated the parameters of a Weibull-shaped psychometric function [Weibull 1951] fit to the 2IFC data to determine the threshold at 75% probability of detection, or 1 Just-Noticeable Difference (JND). An example fit is shown in Figure 5 for uniform dimming, in which the 1 JND threshold was roughly $\alpha = 0.171$. Psychometric curves for each power-saving modality are displayed in supplementary Figure 12. To guide the selection of stimuli magnitude, QUEST uses a prior, which was estimated using a method of adjustment procedure conducted on the first author. Sessions concluded after 50 trials or until the QUEST procedure surpassed a 95% confidence threshold. A 360° video from our set of stimuli (described in Section 5.2) is randomly selected and used for the duration of the trials. We set three magnitudes for each power-saving technique at 1, 2, and 3 JNDs (as extrapolated from the psychometric curve) to be used in the first pilot study.

5.1.2 Pilot Studies. After setting the initial magnitudes in the pre-pilot, we conducted two further pilots using the main study's protocol (see Section 5.2). After each pilot, we adjusted algorithm magnitudes by fitting a best-fit linear model to the magnitude vs. JOD pilot data. New magnitudes were set at intersections with 1, 2, and 3 JODs (see supplementary Section D.2 for details). Both pilots were conducted on expert participants ($N = 8$ each). The final magnitudes used in the main study for each technique are reported in supplementary Section D.3.

5.2 Main Study: Examining Display Mapping Techniques

Hardware Setup. Two commercial head mounted displays were employed – the HTC VIVE Pro Eye and Meta Quest Pro (relevant device specifications can be found in Supplement C). Both devices have integrated eye and head tracking, with data readily obtainable from the respective SDKs and collected during the study.

Participants. 30 naïve paid participants aged 25 to 65 (9M, 19F, 2 other) were recruited, none of whom participated in the pilots. All participants signed informed consent, passed an Ishihara color test and reported normal or corrected-to-normal vision. The study was approved by an Institutional Review Board (IRB). User data was anonymized and sent to a secure server after each session.

Stimuli. We render 360° stereoscopic videos, applying display mappings in real-time via custom post-processing shaders using Unity's ShaderLab language. We chose three real-world videos from the LIVE-FBT-FCVR Database [Jin et al. 2020, 2019, 2021] and hand-crafted two additional videos representative of common VR scenarios: scrolling text and viewing video. All videos were between 10-12 seconds long, and played on loop. Representative frames are shown in Figure 6, with each technique applied at the 3 corresponding magnitudes (additional details in Supplementary D.1).

Experimental Procedure. The experimental procedure consisted of a two-interval forced choice task (2IFC) using the method of paired comparisons. We chose this protocol because it has been shown to result in less noisy data compared to direct rating studies [Perez-Ortiz et al. 2019]. At the start of each trial, users are shown the reference video, and can freely switch between this and two test videos using a 3-button keyboard, with final selections made using a foot pedal. A grey blank screen was displayed for 500ms when stimuli were switched, and was introduced to aid with adaptation and so participants could not make direct comparisons between stimuli by "flipping" between conditions. We allow users to make natural eye and head movements to simulate behavior representative of natural VR/AR use. We included, but did not enforce, a 20-second timer for each trial.

Participants were comfortably seated for the duration of the study. Each participant performed the experiment on one of the two display types (Meta Quest Pro or HTC VIVE Pro Eye). Participants were instructed to select the video with higher quality or fewer distortions, and were required to view both test videos at least once before proceeding. They completed 6 controlled training trials prior to the experiment to familiarize themselves with the protocol, showing each power-saving modality set at the highest magnitude against the reference for different scenes.

In total, our study included (6 display mappers + 1 reference) × 3 magnitudes × 5 scenes = 105 conditions. A full pairwise study design would require an estimated 40 hours (at 20 seconds per trial). To keep duration tractable, we used an active sampling method, ASAP [Mikhailiuk et al. 2021], which determines comparisons that maximize expected information gain given all previous user responses. This allows us to significantly reduce the study length to an average of 49 minutes per participant, including training, 5-point eye tracking calibration, inter-pupillary distance adjustment, and a post-study qualitative survey. Each participant completed a full set of trials as scheduled by ASAP.

6 STUDY RESULTS

The pairwise comparison data from our main study was scaled to units of JODs using Bayesian maximum likelihood estimation under Thurstone's Case V model as implemented using the *pwcmp* software [Perez-Ortiz et al. 2019], which was also used to detect outliers, removing the data of one participant. Scaling our data in JODs allows us to make comparisons between methods on the same perceptual scale, and enables an easy conversion to interpretable units of percentage preference (visualized in Figure 15). For example, if a method A has score 1 JOD greater than method B, this means method A was selected 75% of the time over method B.

The main study results are shown in Figure 7. JOD error bars are computed by simulating 2,000 bootstrap samples using the procedure described by Perez-Ortiz et al. [2019]. The JOD scale is relative, with the reference pegged to 0; scores for each display mapping represent distance from the reference, which is itself a condition of the study and thus is also shown with a confidence interval. Several noteworthy insights can be drawn from this plot. For instance, brightness rolloff has the lowest perceptual impact across all magnitudes compared to other techniques. Two methods – uniform and dichoptic dimming, are device independent and perform similarly across display types. Whitepoint shift had the worst perceptual impact among all other techniques at each magnitude level. These insights are analyzed in the context of power savings in Section 7.1.

An ANOVA analysis was conducted to determine the significance of relevant independent variables on the JOD scores. Significant effects included the display mapping technique type ($F = 23.49, p << .001$) and magnitude ($F = 35.67, p << .001$). Remaining variables (scene, display type, and observer) were not found to be significant.

7 APPLICATIONS

7.1 Display and Battery Level Adaptive Power Mode

Our study highlights the interplay between display type and mapping techniques in terms of image quality and power savings. In real-world applications, the intensity of display mapping may be adjusted following the system status, such as the remaining battery level or expected charging schedule. In addition, a device may or may not support eye-tracking or stereo viewing, determining whether gaze-adaptive methods or dichoptic techniques are applicable.

7.1.1 Fitting Psychometric Functions to the Main Study Data. We combined our study data and display power models to build a transfer function of perceptual impact (in JODs) and relative power savings. To accomplish this, we first fit Weibull-shaped psychometric functions to the main study data for each display mapping (Figure 16) under basic assumptions: all methods rendered with $\alpha = 0$ (equivalent to the reference) must have a JOD score of 0, and a large score (indicating high likelihood of detection) for large modulation factors. Full details for the optimization procedure are described in supplementary Section E.1. These psychometric curves allow us to extrapolate perceptual impact given display mapping intensity along a continuous scale. Display power models are then applied (see supplementary Section E.2) to establish transfer functions relating JOD scores to relative power savings, as shown in Figure 7. This allow us to answer these practical questions for XR applications:

7.1.2 Q1: Which method delivers the best visual quality for a given power target? Techniques can be ranked by evaluating transfer functions in JODs. For example, our results indicate that for a 20% power saving target on an OLED display, ● > ● > ● > ● > ● > ●.

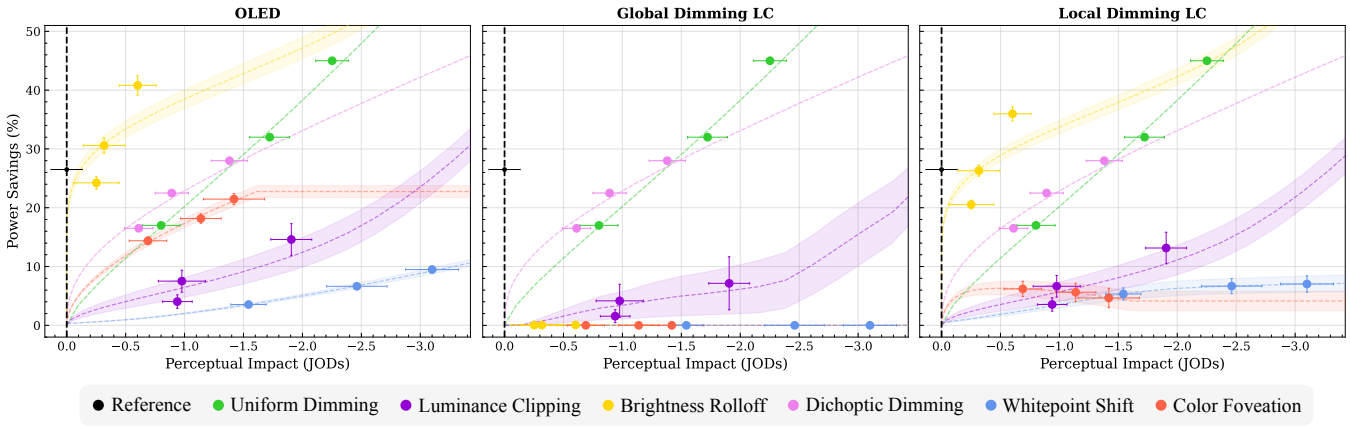


Fig. 7. *Perceptual Impact (JODs) vs. Power Savings (%).* Using the hardware-accurate power models described in Section 3 and the user study data, we fit transfer functions of JODs vs. % power savings for each display mapping for three different display types – OLED, global and local dimming LC. Shaded regions represent 95% confidence intervals of percentage power savings; horizontal and vertical error bars represent 95% confidence intervals of JOD scores and power savings, respectively. Note that methods like uniform and dichoptic dimming are content-independent and thus do not exhibit any vertical error bars.

Table 1. *Power saving rankings at -1 JOD.* The transfer functions in Figure 7 were evaluated for each method at -1 JOD for common XR display types, including non eye-tracked (ET) as well as eye-tracked monocular and binocular displays. Colors represent power saving rankings for a column, blacked out cells are techniques which do not apply to the specific display modality, and grey cells are methods which save near-zero power.

	OLED				Global Dimming LC				Local Dimming LC			
	Monocular	Binocular	Mono + ET	Bino + ET	Monocular	Binocular	Mono + ET	Bino + ET	Monocular	Binocular	Mono + ET	Bino + ET
Uniform Dimming	20.2%	20.2%	20.2%	20.2%	20.2%	20.2%	20.2%	20.2%	20.2%	20.2%	20.2%	20.2%
Luminance Clipping	6.45%	6.45%	6.45%	6.45%	3.43%	3.43%	3.43%	3.43%	5.69%	5.69%	5.69%	5.69%
Brightness Rolloff			38.5%	38.5%			.001%	.001%			33.7%	33.7%
Dichoptic Dimming		22.9%		22.9%		22.9%		22.9%		22.9%		22.9%
Whitepoint Shift	2.01%	2.01%	2.01%	2.01%	0.00%	0.00%	0.00%	0.00%	3.54%	3.54%	3.54%	3.54%
Color Foveation			17.4%	17.4%			0.00%	0.00%			5.79%	5.79%

We validated that the extrapolation provided by the transfer functions aligns with user preference via a validation study ($N = 5$). A full within-condition 2IFC design was used for 20% and 40% savings targets (see Fig. 17 for conditions) on the OLED VIVE Pro Eye using the task described in Section 5.2. 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). Please see supplementary Section E.3 for detailed study protocol and full results.

7.1.3 Q2: Which method saves the most power for a given visual quality target? Table 1 summarizes a sample ranking of each display mapping technique in terms of power saved at -1 JOD for different XR device modalities, including monocular, binocular, and eye-tracked versus non eye-tracked devices.

7.2 Display Primary Selection

Commercial display manufacturers typically choose display primaries which provide good coverage of industry standard color gamuts (e.g., sRGB, DCI-P3). In the literature, the design of display primaries often targets increasing the color gamut [Ajito et al. 2000; Xie et al. 2017], improving power efficiency [Li et al. 2020; Takaya et al. 2005], and reducing color metamerism [Hu et al. 2020; Konig et al. 2002]. Several works in the graphics community explored these ideas for other applications, such as boosting brightness in holographic displays [Kavaklı et al. 2023], reproducing colors using fewer primaries [Huang et al. 2017], and expanding the gamut for projector systems [Kauvar et al. 2015]. Our goal is to determine a possible set of laser primaries that minimizes power consumption and maximizes color accuracy when displaying sRGB images.

We formulated a proxy power model based on the primary luminous efficacy similar to prior work [Li et al. 2020; Xie et al. 2017],

$$\mathcal{K}(\lambda) = k \cdot \eta(\lambda) V(\lambda), \tag{13}$$

where η is wall-plug efficiency and V is photopic luminous efficiency, both a function of wavelength λ. The constant $k = 683 \frac{\text{lm}}{\text{W}}$ is the photopic luminous efficacy of an ideal monochromatic source (555 nm). While the wall-plug efficiency data (see Supplement F) used in our analysis was measured from LEDs with different peak wavelengths, we simplify the computation by assuming that it is applicable to laser diodes. We assume that luminous efficacy, $\mathcal{K}$, is inversely proportional to display power consumption, i.e. an increase in luminous efficacy corresponds to a decrease in power consumption. We uniformly sampled a mix of 1,000 laser diode primaries along the spectral locus, and solved a constrained optimization problem to jointly minimize color difference and power consumption, as defined by our proxy model. Significant power savings can be

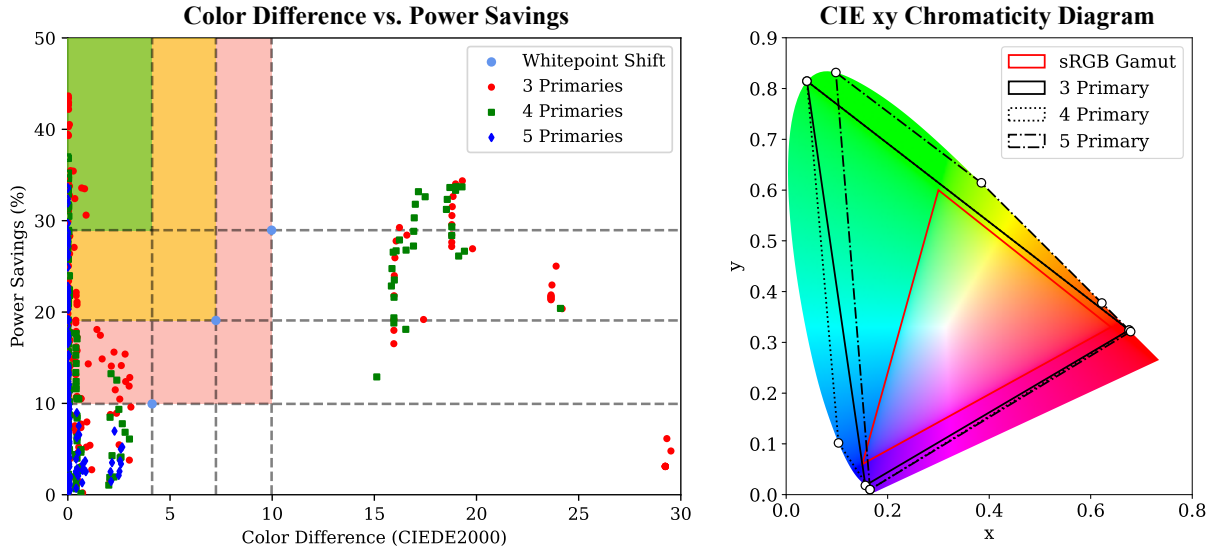


Fig. 8. *Display primary optimization.* (Left) Display primaries which introduce non-zero power savings are plotted, with average color difference (CIEDE2000, ΔE^*) on the x-axis and mean power savings (%) on the y-axis, computed on a fixed set of 500 images from the ImageNet dataset. Whitepoint shift with the magnitudes in the main study are plotted as light-blue points for reference. The green-shaded region (top-left) has higher power savings and lower ΔE^* score than all examined whitepoint shift magnitudes. (Right) The primaries which save the most power with the lowest mean ΔE^* score are plotted on the CIE xy chromaticity diagram.

achieved with minimal loss in color accuracy (Figure 8) when displaying a fixed subset ($N = 500$ images) of the ImageNet dataset [Deng et al. 2009]. Power-optimal primaries for 3-, 4-, and 5-primary displays with minimal ΔE^* [Sharma et al. 2005] are plotted on the CIE xy chromaticity diagram (Figure 8).

Our optimization attempts to minimize color error (dE) when displaying a fixed set of images in a standard sRGB color space rather than maximizing gamut. As such, primary combinations which provide good coverage of sRGB, and in particular of the 500 images in the dataset, minimize our objective. This explains the choice of the optimal 4-primary display, for example, which includes an additional blue primary that covers sRGB but only extends the gamut marginally. In practical implementation, considerations such as physical space for additional primaries and manufacturing costs would also need to be addressed. Furthermore, narrow-primary displays face many practical limitations, including color metamerism, speckle noise, and hardware challenges, which each have associated cost and efficiency tradeoffs to address [Kunkel and Ninan 2023].

7.3 Eye Tracker Power Consumption Analysis

In Section 7.1 we found brightness rolloff as the optimal display mapping for OLED and local dimming LC displays. However, this discussion and Figure 7 do not account for power consumed by the eye tracker itself, which can be significant [Singh et al. 2023]. A reference power estimate for a modern device, TobiiPro Glasses 3, is 189.69mW [Meyer 2022], while the ideal power consumption of an eye tracking system should not exceed 100mW, according to Hong et al. [2018].

As illustrated in Figure 9, brightness rolloff saves less power than non-eye-tracked uniform or dichoptic dimming if factoring eye tracking costs. Thus, using brightness rolloff is not justified unless eye tracking is already a system requirement, in which case brightness rolloff should be employed and can offset the ideal eye tracker power or more than half of the Tobii device at -1 JOD for the HTC VIVE Pro Eye. Similarly, nearly half of the ideal eye tracker power can be offset at -1 JOD for the Meta Quest Pro.

8 LIMITATIONS & FUTURE WORK

Hardware. Conducting perceptual studies on commercial VR devices allows us to develop a realistic study framework with head and eye tracking. However, these displays have their own limitations including optical distortions, eye tracking latency/error, inaccuracy in color reproduction, limited contrast, etc., which influence image quality along with our display power saving methods.

Display Mapping Algorithms. We focused on a fixed set of techniques to maintain study feasibility, but alternative implementations or approaches to display mapping, such as different brightness rolloff profiles, or combinations of multiple techniques presented here are always possible. Further discussion of the limitations of the discussed methods is provided in Supplement B.

Compute vs. Display Power. This work focused on reducing display power, but computational power can also be optimized, for example, via foveated rendering [Patney et al. 2016] or adaptive shading approaches [Jindal et al. 2021]. We leave the analysis of these techniques to future work.

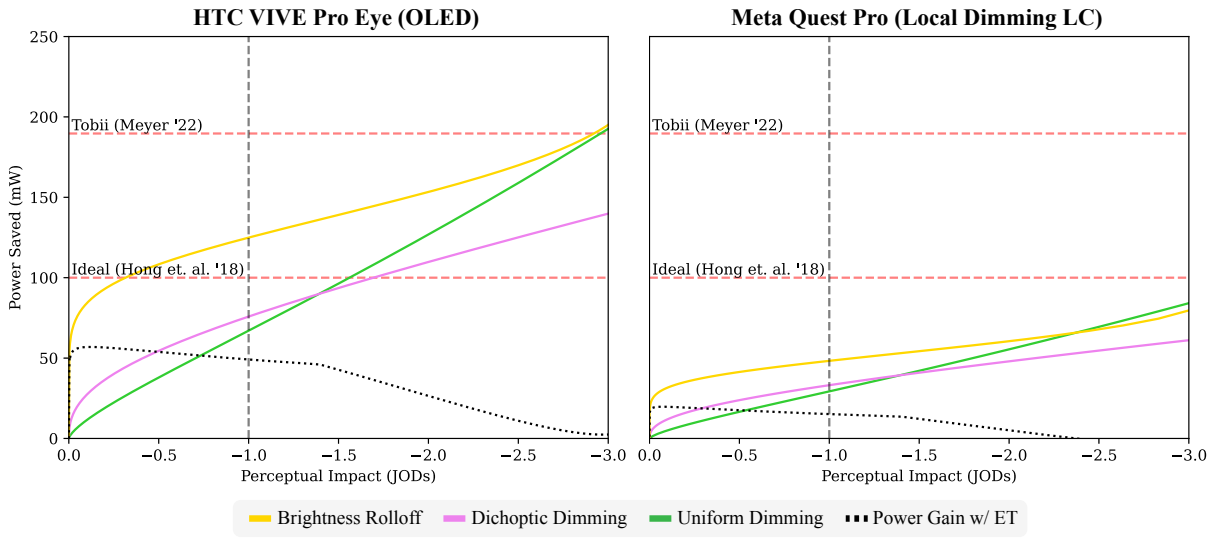


Fig. 9. *Eye tracker power consumption.* We conduct an analysis of gaze-contingent power-saving techniques, while considering estimations of eye tracker power consumption. Solid lines represent power *saved* for three methods (brightness rolloff, uniform dimming, and dichoptic dimming), and the dotted line represents the difference in power saved between brightness rolloff and the next best method with the same perceptual impact.

Additional Display Modalities. While our current analysis is limited to OLED and local/global backlit LC displays, it could be applied to other display modalities, such as mini/micro-LED or liquid crystal on silicon (LCOS) displays. We recommend that practitioners leverage the template laid out in this work to provide a unified perception–to-power-savings analysis via subjective experimentation and modeling when tackling this problem for a specific display architecture not already included in this work.

9 CONCLUSIONS

In this work, we aimed to characterize the tradeoffs of display mapping techniques in terms of power savings and image quality impact. Previous measurements of their perceptual impact were not standardized, making comparisons impossible. Motivated by this gap in the literature, we conducted a large-scale user study to determine the subjective quality of 6 different display mapping techniques, obtaining results scaled in a unified perceptual JOD scale. Our results allowed us to provide insights into practical tasks like selecting a display mapper for power savings, designing display primaries, or making informed choices on power tradeoffs of using eye tracking.

ACKNOWLEDGMENTS

The authors would like to thank Cole Williams for conducting the user study, and John Hill, Cameron Wood, and Helen Ayele for coordinating it. We thank Alex Klement for providing the display module and relevant power data, Henry Milani for EE support and for providing the oscilloscope, Zhao Dong for providing the HTC VIVE Pro Eye, and Yongmin Park for providing PSF measurements. Thanks go to Takahiro Doi and Saeideh Ghahghaeinezamabadi for preliminary data related to dichoptic dimming, Yuta Asano for deriving the brightness rolloff curve, and Christopher Reidy for discussions about compute power. We are grateful to Ken Koh and Yu-Jen Lin for Unity support, Eric Fest for power model discussions, Michael Vaganov, Karsten Behrendt, and Ross Ning for eye tracking support, and Minjung Kim for help with logistics. Finally, we thank all the user study participants for their time. This project is partially supported by the National Science Foundation grants #2225861 and #2232817.

REFERENCES

Takeyuki Ajito, Takashi Obi, Masahiro Yamaguchi, and Nagaaki Ohyama. 2000. Expanded color gamut reproduced by six-primary projection display. In *Projection Displays 2000: Sixth in a Series*, Vol. 3954. SPIE, 130–137.

Bhojan Anand, Karthik Thirugnanam, Jeena Sebastian, Pravein G Kannan, Akhihebbal L Ananda, Mun Choon Chan, and Rajesh Krishna Balan. 2011. Adaptive display power management for mobile games. In *Proceedings of the 9th international conference on Mobile systems, applications, and services*. 57–70.

Yuta Asano and Minqi Wang. 2024. An investigation of color difference for binocular rivalry and a preliminary rivalry metric, Δ E* bino. *Color Research & Application* 49, 1 (2024), 51–64.

ITU Recommendation BT. 2002. Parameter values for the hdtv standards for production and international programme exchange. *International Telecommunication Union, Recommendation, May* (2002).

Shaoyu Chen, Budmonde Duinkharjav, Xin Sun, Li-Yi Wei, Stefano Petrangeli, Jose Echevarria, Claudio Silva, and Qi Sun. 2022. Instant reality: Gaze-contingent perceptual optimization for 3d virtual reality streaming. *IEEE Transactions on Visualization and Computer Graphics* 28, 5 (2022), 2157–2167.

Wei-Chung Cheng and Massoud Pedram. 2004. Power minimization in a backlit TFT-LCD display by concurrent brightness and contrast scaling. *IEEE Transactions on Consumer Electronics* 50, 1 (2004), 25–32.

Inseok Choi, Hojun Shim, and Naehyuck Chang. 2002. Low-power color TFT LCD display for hand-held embedded systems. In *Proceedings of the 2002 international symposium on Low power electronics and design*. 112–117.

Michael A. Cohen, Thomas L. Botch, and Caroline E. Robertson. 2020. The limits of color awareness during active, real-world vision. *Proceedings of the National Academy of Sciences* 117, 24 (2020), 13821–13827. https://doi.org/10.1073/pnas.1922294117 arXiv:https://www.pnas.org/content/117/24/13821.full.pdf

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.

Andrew M Derrington, John Krauskopf, and Peter Lennie. 1984. Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of physiology* 357, 1 (1984), 241–265.

Takahiro Doi, Laurie Wilcox, and T Scott Murdison. 2023. Stereopsis from interocular temporal delay: disentangling the effects of target versus background luminance. *Journal of Vision* 23, 9 (2023), 5159–5159.

Mian Dong and Lin Zhong. 2011a. Chameleon: A color-adaptive web browser for mobile OLED displays. In *Proceedings of the 9th international conference on Mobile systems, applications, and services*. 85–98.

Mian Dong and Lin Zhong. 2011b. Power modeling and optimization for OLED displays. *IEEE Transactions on Mobile Computing* 11, 9 (2011), 1587–1599.

Budmonde Duinkharjav, Praneeth Chakravarthula, Rachel Brown, Anjul Patney, and Qi Sun. 2022a. Image features influence reaction time: A learned probabilistic perceptual model for saccade latency. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–15.

Budmonde Duinkharjav, Kenneth Chen, Abhishek Tyagi, Jiayi He, Yuhao Zhu, and Qi Sun. 2022b. Color-Perception-Guided Display Power Reduction for Virtual Reality. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 41, 6 (2022), 144:1–144:16.

Walter H Ehrenstein, Birgit E Arnold-Schulz-Gahmen, and Wolfgang Jaschinski. 2005. Eye preference within the context of binocular functions. *Graefe's Archive for Clinical and Experimental Ophthalmology* 243 (2005), 926–932.

Stephanie Fu, Netanel Tamir, Shobhita Sundaram, Lucy Chai, Richard Zhang, Tali Dekel, and Phillip Isola. 2023. DreamSim: Learning New Dimensions of Human Visual Similarity using Synthetic Data. *arXiv preprint arXiv:2306.09344* (2023).

Franco Gatti, Andrea Acquaviva, Luca Benini, and Bruno Ricco'. 2002. Low power control techniques for TFT LCD displays. In *Proceedings of the 2002 international conference on Compilers, architecture, and synthesis for embedded systems*. 218–224.

B Hahn. 2016. Closing the green efficiency gap, status and recent approaches. In *Proc. DOE Workshop Raleigh*.

B Hahn, A Weimar, M Peter, and J Baur. 2008. High-power InGaN LEDs: Present status and future prospects. *Light-Emitting Diodes: Research, Manufacturing, and Applications XII* 6910 (2008), 9–16.

Thorsten Hansen, Lars Pracejus, and Karl R Gegenfurtner. 2009. Color perception in the intermediate periphery of the visual field. *Journal of vision* 9, 4 (2009), 26–26.

Injoon Hong, Kyeongryeol Bong, and Hoi-Jun Yoo. 2018. Challenges of eye tracking systems for mobile XR glasses. In *Applications of Digital Image Processing XLI*, Vol. 10752. SPIE, 391–397.

HTC. 2020. VIVE Pro Eye Specs & User Guide. https://developer.vive.com/resources/hardware-guides/vive-pro-eye-specs-user-guide/.

Yu Hu, Minchen Wei, and Ming Ronnier Luo. 2020. Observer metamerism to display white point using different primary sets. *Optics Express* 28, 14 (2020), 20305–20323.

Fu-Chung Huang, Dawid Pająk, Jonghyun Kim, Jan Kautz, and David Luebke. 2017. Mixed-primary factorization for dual-frame computational displays. *ACM Trans. Graph.* 36, 4, Article 149 (jul 2017), 13 pages. https://doi.org/10.1145/3072959.3073654

Yize Jin, Meixu Chen, Todd Goodall Bell, Zhaolin Wan, and Alan Bovik. 2020. Study of 2D foveated video quality in virtual reality. In *Applications of digital image processing XLIII*, Vol. 11510. SPIE, 18–26.

Yize Jin, Meixu Chen, Todd Goodall, Anjul Patney, and Alan Bovik. 2019. LIVE-Facebook Technologies-Compressed Virtual Reality (LIVE-FBT-FCVR) Databases. *Online: http://live. ece. utexas. edu/research/LIVEFBTFCVR/index. html* (2019).

Yize Jin, Meixu Chen, Todd Goodall, Anjul Patney, and Alan C Bovik. 2021. Subjective and objective quality assessment of 2D and 3D foveated video compression in virtual reality. *IEEE Transactions on Image Processing* 30 (2021), 5905–5919.

Akshay Jindal, Krzysztof Wolski, Karol Myszkowski, and Rafał K Mantiuk. 2021. Perceptual model for adaptive local shading and refresh rate. *ACM Transactions on Graphics (TOG)* 40, 6 (2021), 1–18.

Anton S Kaplanyan, Anton Sochenov, Thomas Leimkühler, Mikhail Okunev, Todd Goodall, and Gizem Rufo. 2019. DeepFovea: Neural reconstruction for foveated rendering and video compression using learned statistics of natural videos. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 1–13.

Isaac Kauvar, Samuel J. Yang, Liang Shi, Ian McDowall, and Gordon Wetzstein. 2015. Adaptive color display via perceptually-driven factored spectral projection. *ACM Trans. Graph.* 34, 6, Article 165 (nov 2015), 10 pages. https://doi.org/10.1145/2816795.2818070

Koray Kavaklı, Liang Shi, Hakan Urey, Wojciech Matusik, and Kaan Akşit. 2023. Multicolor Holograms Improve Brightness in Holographic Displays. In *SIGGRAPH Asia 2023 Conference Papers* (Sydney, NSW, Australia) *(SA '23)*. Association for Computing Machinery, New York, NY, USA, Article 20, 11 pages. https://doi.org/10.1145/3610548.3618135

Louis Kerofsky and Scott Daly. 2006. Brightness preservation for LCD backlight dimming. *Journal of the Society for Information Display* 14, 12 (2006), 1111–1118.

Jeong-Sik Kim and Seung-Woo Lee. 2020. Peripheral dimming: A new low-power technology for OLED display based on gaze tracking. *IEEE Access* 8 (2020), 209064–209073.

F Konig, Kenro Ohsawa, Masahiro Yamaguchia, Nagaaki Ohyama, and Bernhard Hill. 2002. A multiprimary display: Optimized control values for displaying tristimulus values. In *IS AND TS PICS CONFERENCE*. Society for Imaging Science & Technology, 215–220.

Brooke Krajancich, Petr Kellnhofer, and Gordon Wetzstein. 2023. Towards Attention–aware Foveated Rendering. *ACM Trans. Graph.* 42, 4, Article 77 (jul 2023), 10 pages. https://doi.org/10.1145/3592406

Timo Kunkel and Ajit Ninan. 2023. HDR and Wide Color Gamut Display Technologies and Considerations. *Fundamentals and Applications of Colour Engineering* (2023), 311–334.

Hongru Li, Han Gao, Gökhan Kirca, and Zhichun Lei. 2020. Energy-saving display by color pixel re-representation. *IEEE Transactions on Circuits and Systems for Video Technology* 30, 12 (2020), 4728–4738.

Bruce Lindbloom. 2017. Chromatic Adaptation. http://www.brucelindbloom.com/index.html?Eqn_ChromAdapt.html.

MR Luo and RWG Hunt. 1998. The structure of the CIE 1997 colour appearance model (CIECAM97s). *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur* 23, 3 (1998), 138–146.

M Ronnier Luo et al. 2000. A review of chromatic adaptation transforms. *Review of Progress in Coloration and Related Topics* 30 (2000), 77–92.

Rafał K. Mantiuk, Maliha Ashraf, and Alexandre Chapiro. 2022. stelaCSF: a unified model of contrast sensitivity as the function of spatio-temporal frequency, eccentricity, luminance and area. *ACM Trans. Graph.* 41, 4, Article 145 (jul 2022), 16 pages. https://doi.org/10.1145/3528223.3530115

Rafał K. Mantiuk, Gyorgy Denes, Alexandre Chapiro, Anton Kaplanyan, Gizem Rufo, Romain Bachy, Trisha Lian, and Anjul Patney. 2021. FovVideoVDP: a visible difference predictor for wide field-of-view video. *ACM Trans. Graph.* 40, 4, Article 49 (jul 2021), 19 pages. https://doi.org/10.1145/3450626.3459831

Nathan Matsuda, Alex Chapiro, Yang Zhao, Clinton Smith, Romain Bachy, and Douglas Lanman. 2022. Realistic Luminance in VR. In *SIGGRAPH Asia 2022 Conference Papers* (Daegu, Republic of Korea) *(SA '22)*. Association for Computing Machinery, New York, NY, USA, Article 21, 8 pages. https://doi.org/10.1145/3550469.3555427

Johannes Meyer. 2022. Towards energy efficient mobile eye tracking for AR glasses through optical sensor technology. *arXiv preprint arXiv:2212.03189* (2022).

Microsoft. 2022. Adaptive brightness. https://support.microsoft.com/en-us/windows/content-adaptive-brightness-control-in-windows-292d1f7f-9e02-4b37-a9c8-dab3e1727e78.

Aliaksei Mikhailiuk, Clifford Wilmot, Maria Perez-Ortiz, Dingcheng Yue, and Rafal Mantiuk. 2021. Active Sampling for Pairwise Comparisons via Approximate Message Passing and Information Gain Maximization. In *2020 IEEE International Conference on Pattern Recognition (ICPR)*.

Bernard Moulden et al. 1988. Border effects on brightness: a review of findings, models and issues. *Spatial vision* 3, 4 (1988), 225–262.

Yukio Narukawa, Masatsugu Ichikawa, Daisuke Sanga, Masahiko Sano, and Takashi Mukai. 2010. White light emitting diodes with super-high luminous efficacy. *Journal of physics D: Applied physics* 43, 35 (2010), 354002.

Anjul Patney, Marco Salvi, Joohwan Kim, Anton Kaplanyan, Chris Wyman, Nir Benty, David Luebke, and Aaron Lefohn. 2016. Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 1–12.

Jonathan Peirce, Jeremy R Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Kristoffer Lindeløv. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior research methods* 51 (2019), 195–203.

Maria Perez-Ortiz, Aliaksei Mikhailiuk, Emin Zerman, Vedad Hulusic, Giuseppe Valenzise, and Rafał K Mantiuk. 2019. From pairwise comparisons and rating to a unified quality scale. *IEEE Transactions on Image Processing* 29 (2019), 1139–1151.

M Peter, A Laubsch, P Stauss, A Walter, J Baur, and B Hahn. 2008. Green ThinGaN power-LED demonstrates 100 lm. *physica status solidi c* 5, 6 (2008), 2050–2052.

Linghui Rao, Yongmin Park, Alex Klement, Chris Kang, Eric Park, Jim Zhuang, Cheonhong Kim, Hsin-Ying Chiu, Ross Ning, Daozhi Wang, et al. 2023. 5-1: Invited Paper: Infinite Display for Meta Quest Pro. In *SID Symposium Digest of Technical Papers*, Vol. 54. Wiley Online Library, 32–35.

Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. 2010. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann. 152–159 pages.

Michael Reiss and Gilfe Reiss. 1997. Ocular dominance: some family data. *Laterality: Asymmetries of Body, Brain and Cognition* 2, 1 (1997), 7–16.

Dario Schiavon, Michael Binder, Andreas Loeffler, and Matthias Peter. 2013. Optically pumped GaInN/GaN multiple quantum wells for the realization of efficient green light-emitting devices. *Applied Physics Letters* 102, 11 (2013).

Kalpana Seshadrinathan and Alan C Bovik. 2007. A structural similarity metric for video based on motion models. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, Vol. 1. IEEE, I–869.

Gaurav Sharma, Wencheng Wu, and Edul N Dalal. 2005. The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur* 30, 1 (2005), 21–30.

Rahul Singh, Muhammad Huzaifa, Jeffrey Liu, Anjul Patney, Hashim Sharif, Yifan Zhao, and Sarita Adve. 2023. Power, performance, and image quality tradeoffs in foveated rendering. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 205–214.

Vincent Sitzmann, Ana Serrano, Amy Pavel, Maneesh Agrawala, Diego Gutierrez, Belen Masia, and Gordon Wetzstein. 2017. How do people explore virtual environments? *IEEE Transactions on Visualization and Computer Graphics* (2017).

Daniel Steigerwald, Serge Rudaz, Heng Liu, R Scott Kern, Werner Götz, and Robert Fletcher. 1997. III–V nitride semiconductors for high-performance blue and green light-emitting devices. *Jom* 49 (1997), 18–23.

Masanori Takaya, Ken Ito, Gosuke Ohashi, and Yoshifumi Shimodaira. 2005. Color-conversion method for a multi-primary display to reduce power consumption and conversion time. *Journal of the Society for Information Display* 13, 8 (2005), 685–690.

Kiat Wee Tan, Tadashi Okoshi, Archan Misra, and Rajesh Krishna Balan. 2013. FOCUS: a usable & effective approach to OLED display power management. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. 573–582.

Matthew Trentacoste, Wolfgang Heidrich, Lorne Whitehead, Helge Seetzen, and Greg Ward. 2007a. Photometric image processing for high dynamic range displays. *Journal of Visual Communication and Image Representation* 18, 5 (2007), 439–451.

Matthew Trentacoste, Wolfgang Heidrich, Lorne Whitehead, Helge Seetzen, and Greg Ward. 2007b. Photometric image processing for high dynamic range displays. *Journal of Visual Communication and Image Representation* 18 (10 2007), 439–451. https://doi.org/10.1016/j.jvcir.2007.06.006

Minqi Wang, Jian Ding, Dennis M Levi, and Emily A Cooper. 2023. The effect of interocular contrast differences on the appearance of augmented reality imagery. *ACM Transactions on Applied Perception* (2023).

Andrew B Watson and Denis G Pelli. 1983. QUEST: A Bayesian adaptive psychometric method. *Perception & psychophysics* 33, 2 (1983), 113–120.

Tan Kiat Wee, Eduardo Cuervo, and Rajesh Balan. 2018. FocusVR: Effective 8 usable VR display power management. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–25.

Shu Wei, Desmond Bloemers, and Aitor Rovira. 2023. A preliminary study of the eye tracker in the meta quest pro. In *Proceedings of the 2023 ACM International Conference on Interactive Media Experiences*. 216–221.

Waloddi Weibull. 1951. A statistical distribution function of wide applicability. *Journal of applied mechanics* (1951).

Jeremy M Wolfe. 1983. Influence of spatial frequency, luminance, and duration on binocular rivalry and abnormal fusion of briefly presented dichoptic stimuli. *Perception* 12, 4 (1983), 447–456.

Hao Xie, Carlos Eduardo Rodríguez-Pardo, and Gaurav Sharma. 2017. Multiobjective optimization for color display primary designs. *Journal of Electronic Imaging* 26, 6 (2017), 063013–063013.

Aron Yu and Kristen Grauman. 2015. Just noticeable differences in visual attributes. In *Proceedings of the IEEE International Conference on Computer Vision*. 2416–2424.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 586–595.

Fangcheng Zhong, Akshay Jindal, Ali Özgür Yöntem, Param Hanji, Simon J. Watt, and Rafał K. Mantiuk. 2021. Reproducing reality with a high-dynamic-range multi-focal stereo display. *ACM Trans. Graph.* 40, 6, Article 241 (dec 2021), 14 pages. https://doi.org/10.1145/3478513.3480513

Fangcheng Zhong, George Alex Koulieris, George Drettakis, Martin S Banks, Mathieu Chambe, Frédo Durand, and Rafał K Mantiuk. 2019. DiCE: Dichoptic contrast enhancement for VR and stereo displays. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 1–13.

A DISPLAY MODELLING

A.1 Modelling the Point Spread Function

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

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

where the input variable x is the distance from an illuminated pixel, and α, γ, ϵ, 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,

$$\mathcal{I} = \mathcal{I}_{\mathrm{LC}} \mathcal{B}. \tag{14}$$

where $\mathcal{I}_{\mathrm{LC}}$ 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, $\mathcal{B} = \mathcal{W}d$. 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,

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

$\mathcal{B}^*$ 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],

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

where $\mathcal{N}$ is a neighborhood of BLU LEDs.

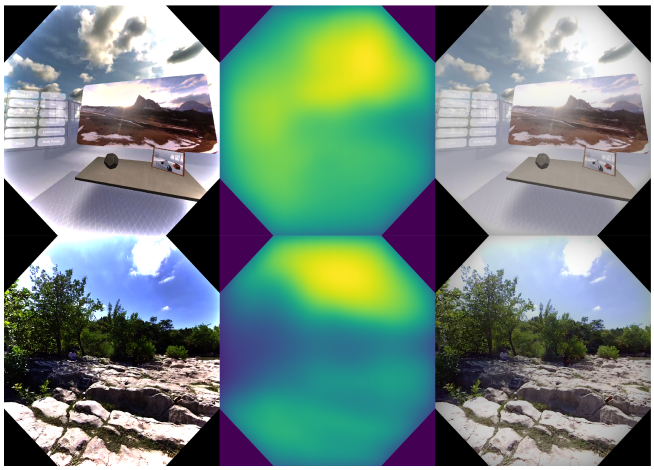


Fig. 10. *Local dimming simulation.* LC pixel intensities $\mathcal{I}_{\mathrm{LC}}$ (left) are used to filter light from the BLU $\mathcal{B}$ (middle) to produce the displayed image (right).

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, ϕ is retinal eccentricity, and β is a constant controlling the minimum curve value. Solving for β so that α becomes the minimum value at the edge of the display FOV,

$$\exp(-\beta * (\mathrm{FOV}/2)^2) = 1 - \alpha$$
$$\beta = -4 \ln(1 - \alpha)/\mathrm{FOV}^2.$$

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
Resolution	1400×1660	1800×1920
PPD (horizontal/vertical)	19.17/18.83	14.58/13.36
Field of View	$106^\circ \times 96^\circ$	$98^\circ \times 98^\circ$
Eye Tracker Frequency	–	120 Hz
Eye Tracker Accuracy	$1.249^\circ - 1.813^\circ$	$0.5^\circ - 1.1^\circ$

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,

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

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 α 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 [Mikhailiuk 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 (δ 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.

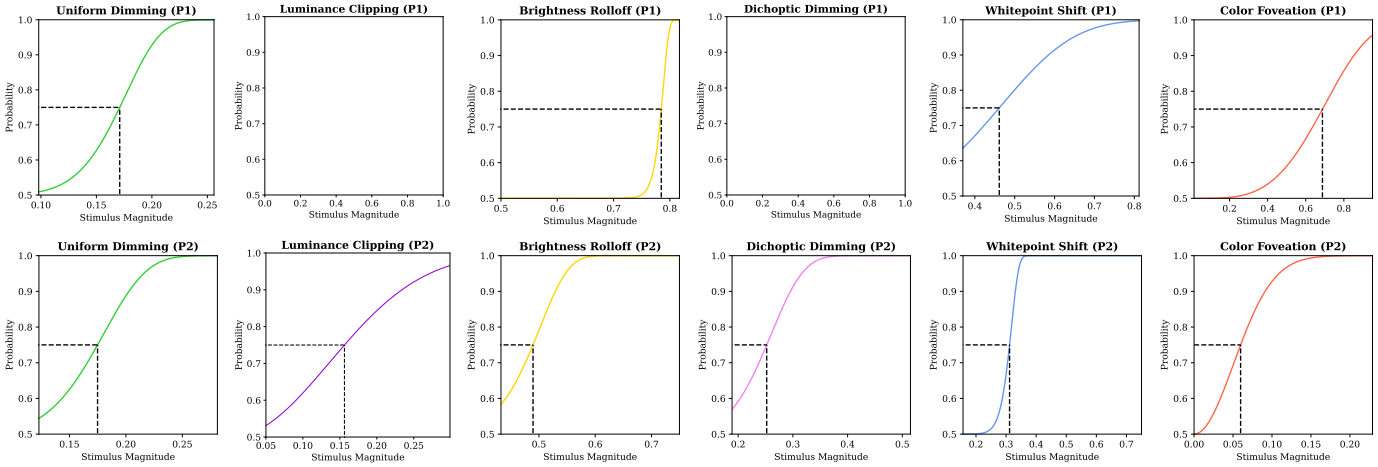


Fig. 12. *Pre-pilot staircase results.* Psychometric curves fit to the pre-pilot study results for both participants. x-axis is stimulus magnitude and y-axis is probability of detection. Empty plots were conditions not done by P1.

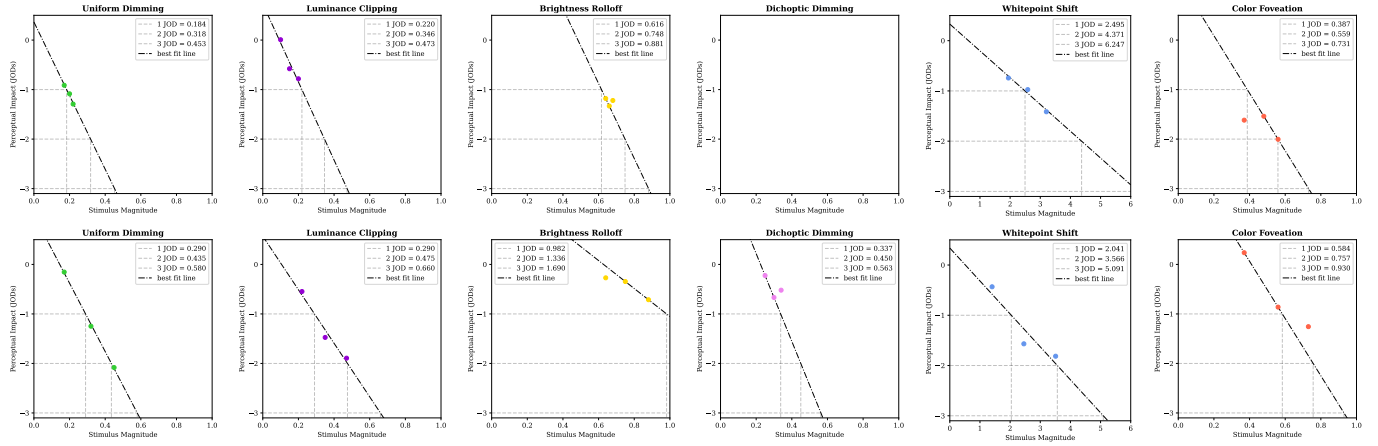


Fig. 13. *Extrapolating pilot data.* The top row shows results after the first pilot, and bottom row results after the second pilot. x-axis represents stimulus magnitude, α, and the y-axis shows perceptual impact (JODs). Missing plot refers to a condition added after the first pilot.

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.

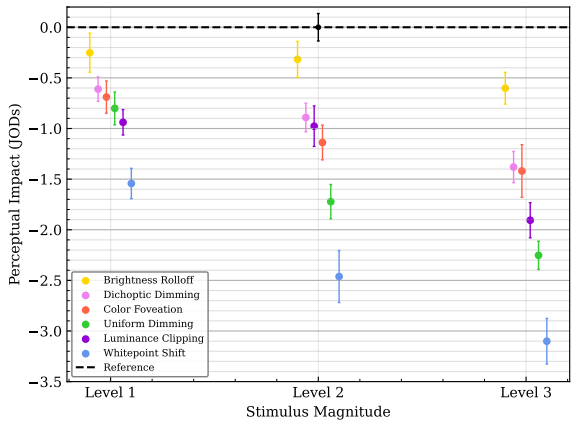


Fig. 14. *Main study results.* The main study data was scaled to units of JODs (y-axis) for three increasing levels of magnitude (x-axis). Vertical error bars represent 95% percent confidence intervals.

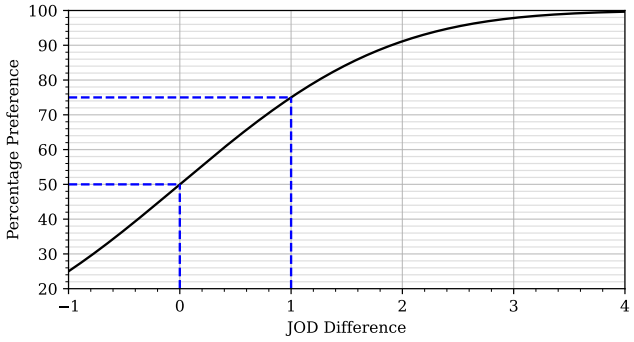


Fig. 15. *JODs vs. % preference.* We replicate the plot from Mantiuk et al. [2021], which maps JODs to interpretable units of percentage preference. The difference in 1 JOD between two techniques, A and B, indicates 75% percent selection of A over B.

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 scenes × (5 display mappings + 1 reference) = 12 conditions and for 40% power savings, 2 scenes × (4 display mappings + 1 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 λ, we fit a linear function to the data, relating λ to WPE.

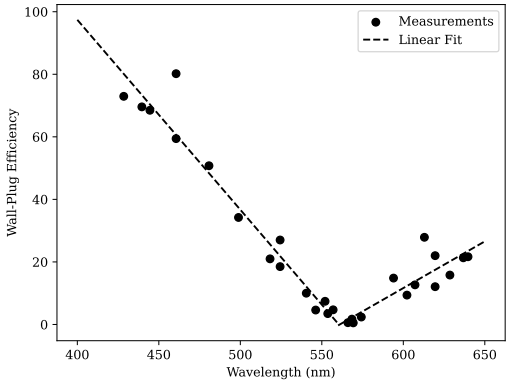
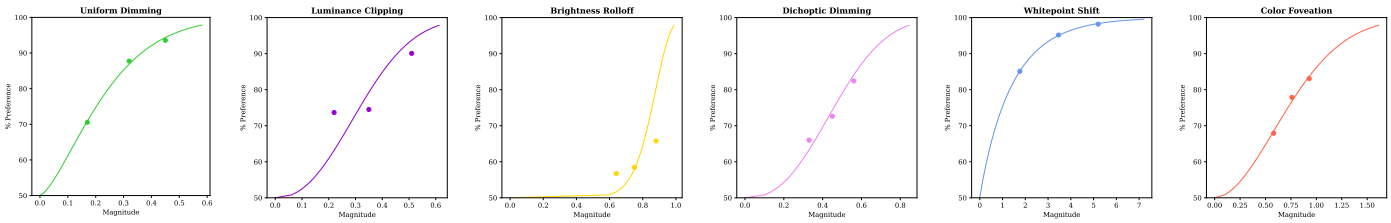


Fig. 18. *Wavelength versus wall-plug efficiency data.* Wall-plug efficiency data for LEDs is aggregated across five datasets, and lines are fit to the measurements.



(a) Uniform Dimming (b) Luminance Clipping (c) Brightness Rolloff (d) Dichoptic Dimming (e) Whitepoint Shift (f) Color Foveation

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

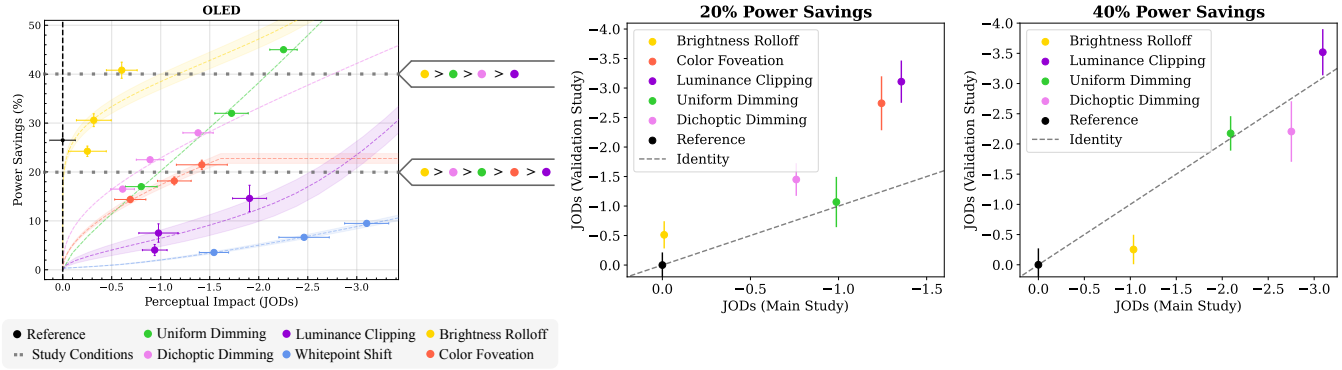


Fig. 17. *Validation study conditions and results.* The two conditions used in the validation study, 20% and 40%, are visualized as gray dotted lines (left plot). The ranking of display mappings is displayed as circle markers to the right of the first figure. JOD scores predicted by the psychometric 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.