The Perceptual Quantizer
Design Considerations and Applications

TIMO KUNKEL, PHD

DOLBY LABORATORIES, INC.

SAN FRANCISCO, CA

Introduction

- The Perceptual Quantizer or 'PQ' is an efficient signal non-linearity for HDR applications
- It is modeled after elements of human vision to efficiently and effectively encode a large absolute luminance range for content consumption
- PQ has developed into a foundation of HDR imaging

Overview

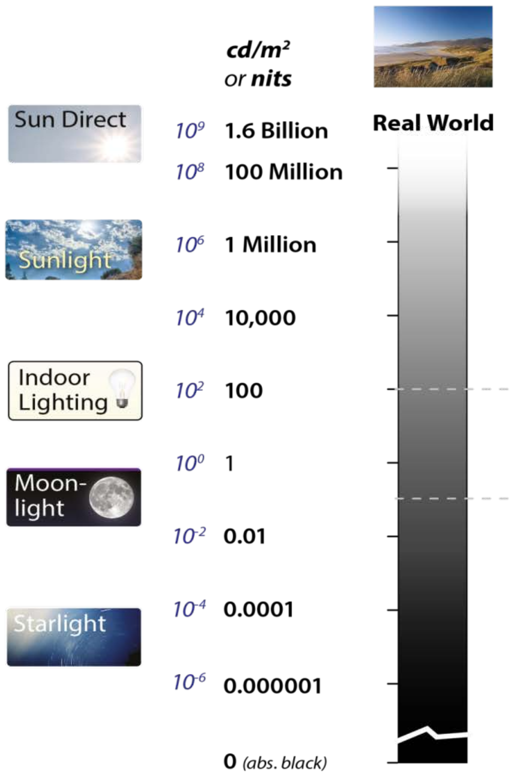
- Discuss the design considerations when developing PQ
- Outline the large field of applications

Example Application

Photo: Timo Kunkel, Dolby   Note: This image is a tone-mapped example

Development of PQ

Encoding HDR for Content Delivery and Image Display



- About 15 years ago, HDR displays used custom drivers and mapping algorithms had to be developed & applied which was not very efficient from a compatibility and scalability point of view.
- Fully log or gamma functions encounter problems with contouring in certain regions when considering HDR luminance ranges
  - ***Efficient*** *but* ***not Effective*** *for our intended Application*
- Linear light formats existed offering extreme luminance ranges (e.g., OpenEXR and Radiance HDR)
  - Good for post processing and CG applications (e.g., Raytracing, SFX & Image Based Lighting - IBL)
  - ***Effective*** *but* ***not Efficient*** *for our intended Application*

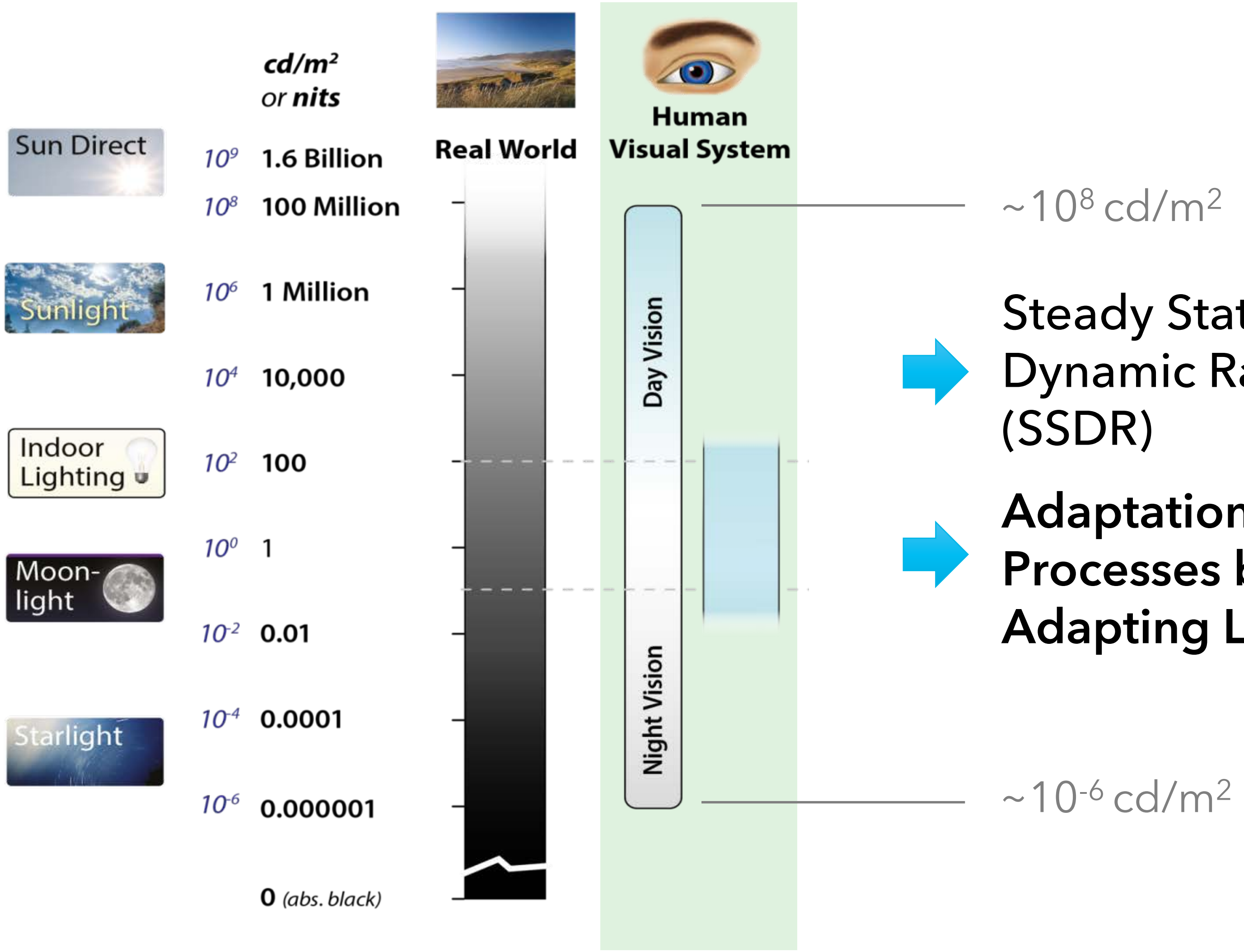
Design considerations:

- Don't design for today's display capabilities (limited performance)
- Nor for ranges offered by float linear luminance (limited efficiency).
- Avoid contouring artifacts (limited effectivity).
- **Base it on properties of human vision!**

Adaptive Processes

Apparently Simple Option:

Reduce Luminance Range to capability of Human Visual System



Steady State Dynamic Range (SSDR)

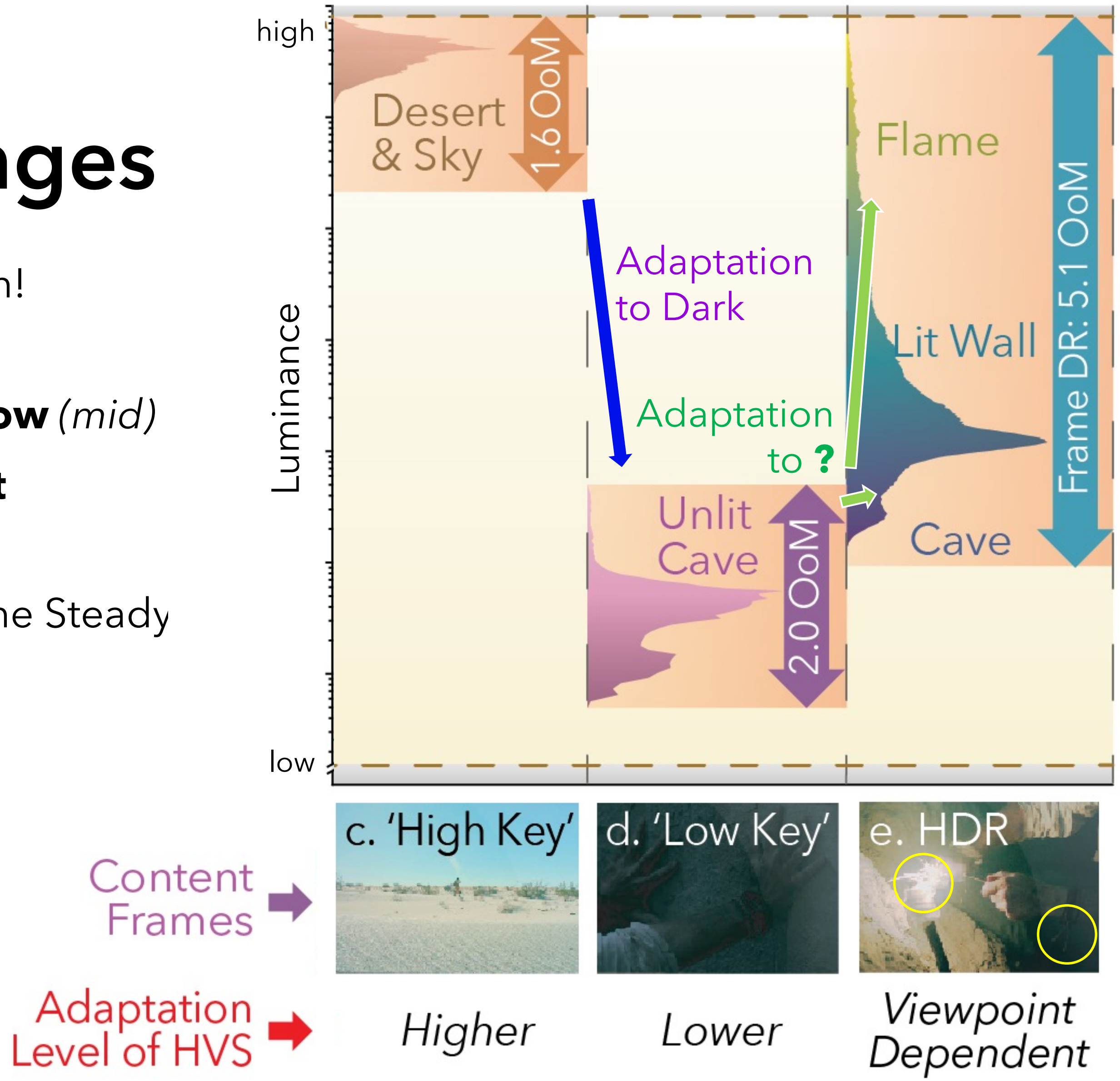
Adaptation Processes based on Adapting Luminance L_A

Note: This is for illustration only!

The adaptive behavior of the HVS is more complex than shown here.

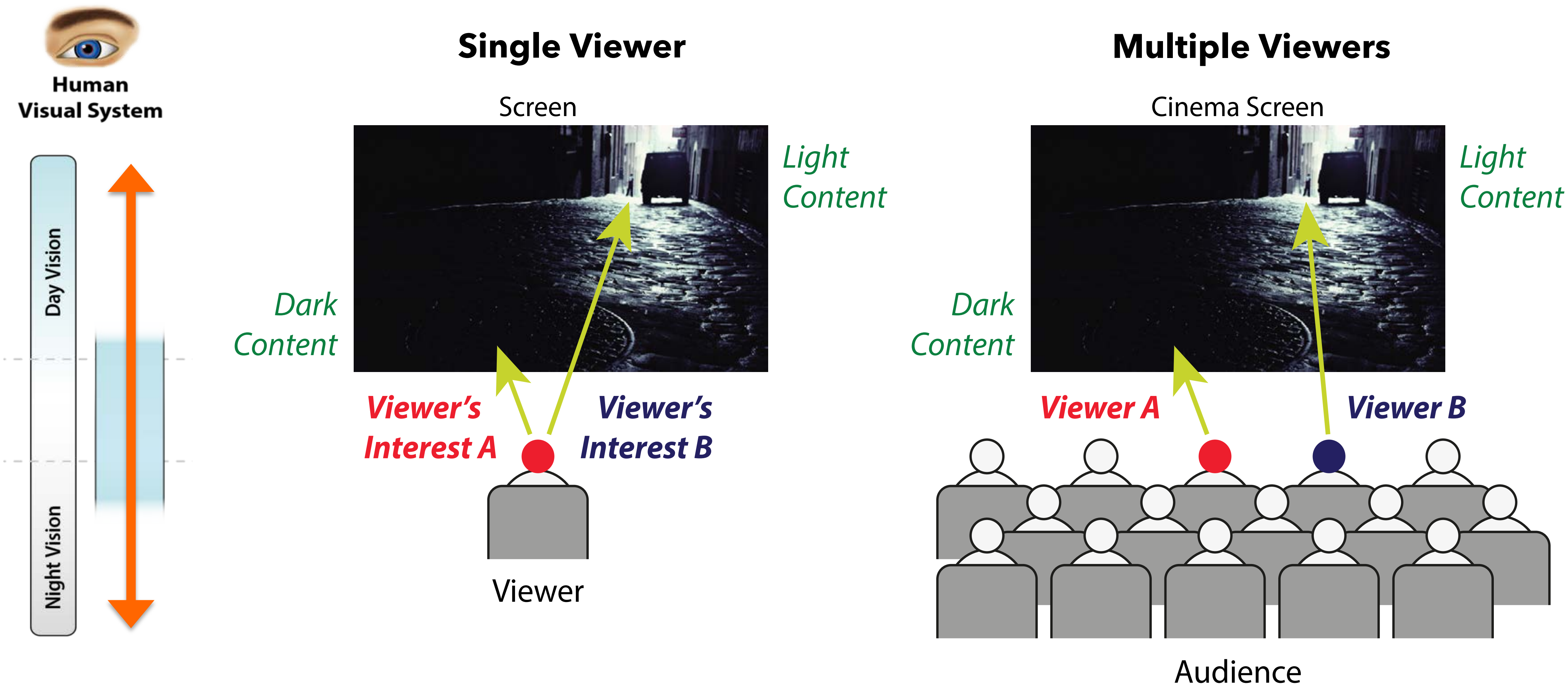
Content Luminance Ranges

- Content (Frame) Range does not have to be high!
  - Steady State typically sufficient
- But average Luminance might be **high** *(left)* **or low** *(mid)*
- Adaptation* adjusts sensitivity of HVS **to highest sensitivity at Adapting Luminance L_A**
- Option for HDR format: Encode and Transport the Steady State Range & Adapting Luminance?
- **Challenge:** Content contrast **can** be high *(right)*
- Now, adaptation is viewpoint and viewing environment dependent
- **Imaging encoding must facilitate all cases!**



** This is a simplification. Adaptive processes are more complex.*

Viewing Conditions and Preferences



Summary (so far)

- Fully log or gamma functions only approximate perception in certain regions when considering HDR luminance ranges
- They can work, but then, the bit depth requirements are too high
- A smaller, floating range (steady state) with adaptation metadata (such as L_A) follows perceptual properties very well but is challenging to predict and robustly implement for the intended application

Are there other options that:

- follow perceptual principles
- don't require knowledge of L_A?
- Are both effective AND efficient?

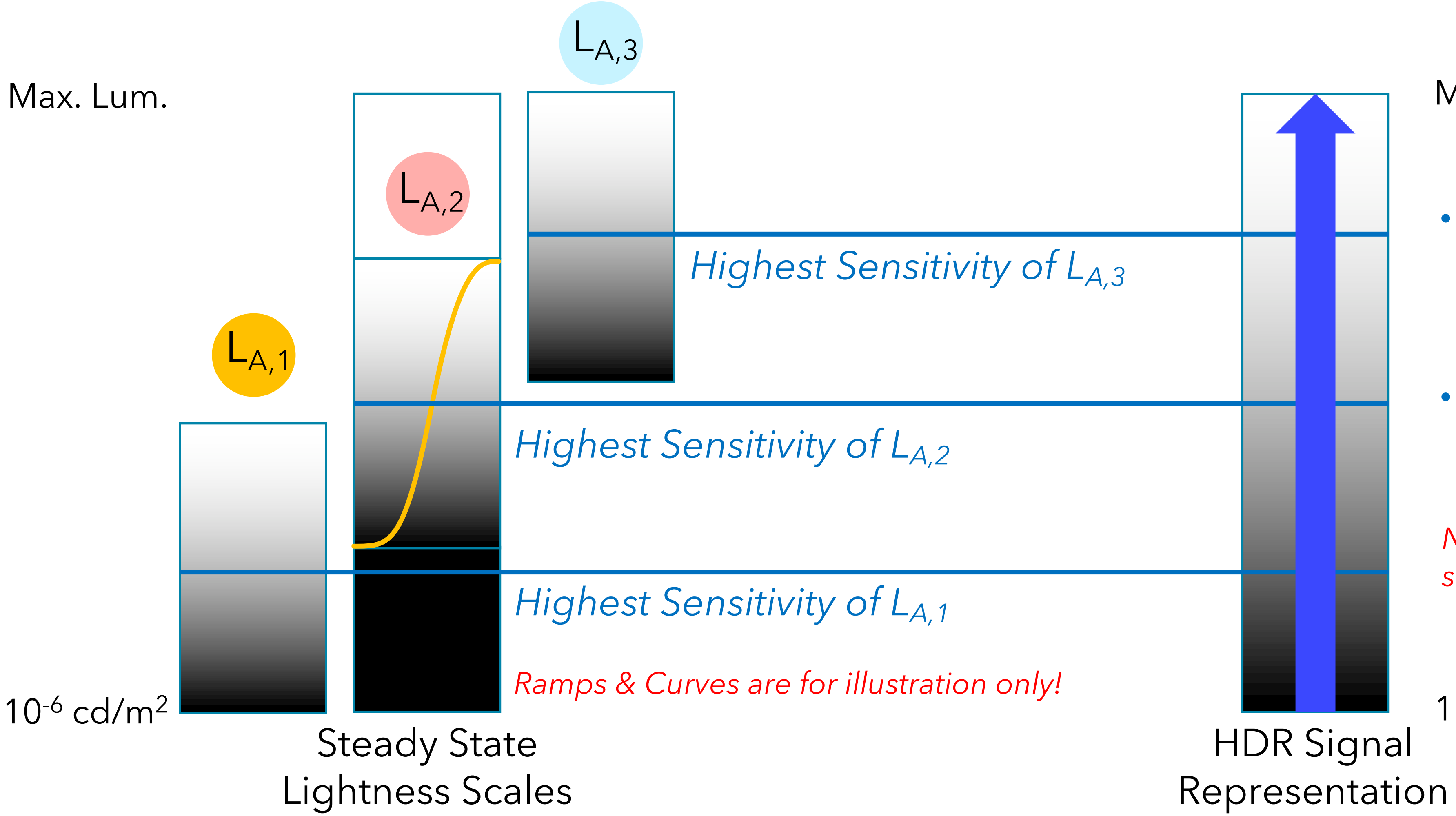
Potential Approach:

- 'Worst Case Engineering' paradigm
- Intervals at Highest HVS Sensitivity
- Model discrimination thresholds through a useful luminance range

Worst Case Engineering

- **Fundamental question:** How to determine if a signal non-linearity retains quantization steps that are all equal or below the threshold of one JND over an absolute luminance range.
  - Approach in terms of sensitivity (the inverse of threshold).
  - Particularly: How does the sensitivity of the visual system vary with luminance level.
- One approach to answer this question is to follow **worst-case engineering design principles**.
- **Fundamental Concept:**
  - Worst-case engineering considers the most severe possible behavior of a system that can reasonably be projected to occur in a given situation.
  - In the context of imaging, the worst case can be defined by either contouring steps to become visible or by unnecessarily wasting bandwidth for a given luminance level.

Worst Case Engineering



- *$L_{A,1...n}$ spaced (quantized) in **Just Noticeable Difference (JND) steps** over intended luminance range*
- ***JNDs based on Highest Sensitivity!***

Note: Not all tone values simultaneously visible to the HVS

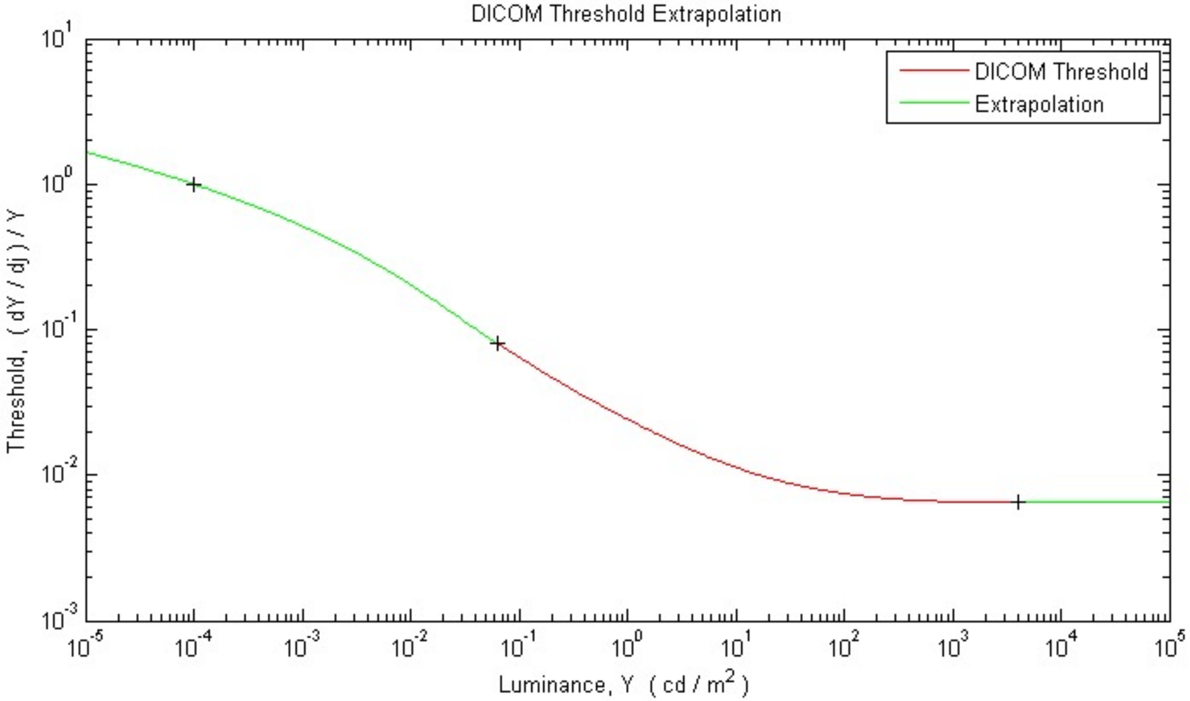
DICOM

Gray Scale Display Function

- Created for medical imaging, adopted by VESA
- Covers 0.05 to 4000 cd/m^2 with 1023 steps
- Could be extend it to a broader range by adding range above and below limits

Challenges

- Could see visible differences between adjacent bars on test HDR display
- Especially at darker levels
- Why?



Note: The extrapolation of the DICOM curve was carried out by Dolby for exploration of options

References:

- Bradley M. Hemminger, Richard Eugene Johnston, Jannick P. Rolland, and Keith E. Muller "Perceptual linearization of video-display monitors for medical image presentation", Proc. SPIE 2164, Medical Imaging 1994: Image Capture, Formatting, and Display, (1 May 1994); https://doi.org/10.1117/12.174005
- NEMA Standards Publication PS 3.14-2008, Digital Imaging and Communications in Medicine (DICOM), Part 14: Grayscale Standard Display Function, National Electrical Manufacturers Association, 2008.
- P. G. J. Barten, "*Contrast Sensitivity of the Human Eye and its Effects on Image Quality*", SPIE Optical Engineering Press, 1999.
- P. G. J. Barten, "Formula for the contrast sensitivity of the human eye", *Proc. SPIE-IS&T Vol. 5294*:231-238, Jan. 2004.

The Barten Model

$$CSF = \frac{1}{m_t} = \frac{M_{opt}(u)/k}{\sqrt{\frac{2}{T}\left(\frac{1}{X_0^2}+\frac{1}{X_{\max}^2}+\frac{u^2}{N_{\max}^2}\right)\left(\frac{1}{\eta pE}+\frac{\Phi_0}{1-e^{-(u/u_0)^2}}\right)}}$$

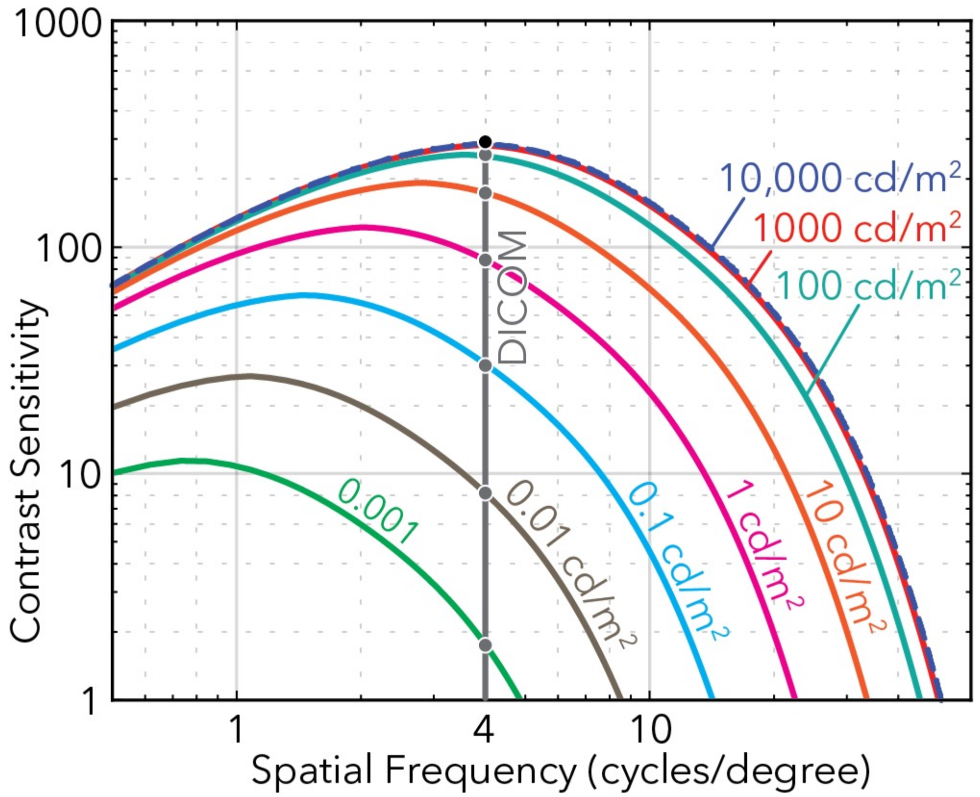
$$M_{opt}(u) = e^{-2\pi^2\sigma^2u^2}$$

$$\sigma = \sqrt{\sigma_0^2 + (C_{ab}d)^2}$$

$$d = 5 - 3\tanh\left(0.4\log\left(LX_0^2/40^2\right)\right)$$

$$E = \frac{\pi d^2}{4}L\left(1-(d/9.7)^2+(d/12.4)^4\right)$$

- DICOM is based on the Barten Model
- The Barten model is not a fixed curve, but a model with many parameter values
- DICOM's choice of parameters are not ideal for our use case



Reference:

- M. Cowan, G. Kennel, T. Maier, and B. Walker, "Contrast Sensitivity Experiment to Determine the Bit Depth for Digital Cinema", *SMPTE Mot. Imag. J.*, 113:281-292, Sept. 2004.
- Mantiuk, R., Daly, S., Myszkowski, K., and Seidel, H. 2005. "Predicting visible differences in high dynamic range images: model and its calibration". In Proc. SPIE, vol. 5666, 204–214.

Optimize Barten Model Parameters

$$CSF = \frac{1}{m_t} = \frac{M_{opt}(u)/k}{\sqrt{\frac{2}{T}\left(\frac{1}{X_0^2}+\frac{1}{X_{\max}^2}+\frac{u^2}{N_{\max}^2}\right)\left(\frac{1}{\eta pE}+\frac{\Phi_0}{1-e^{-(u/u_0)^2}}\right)}}$$

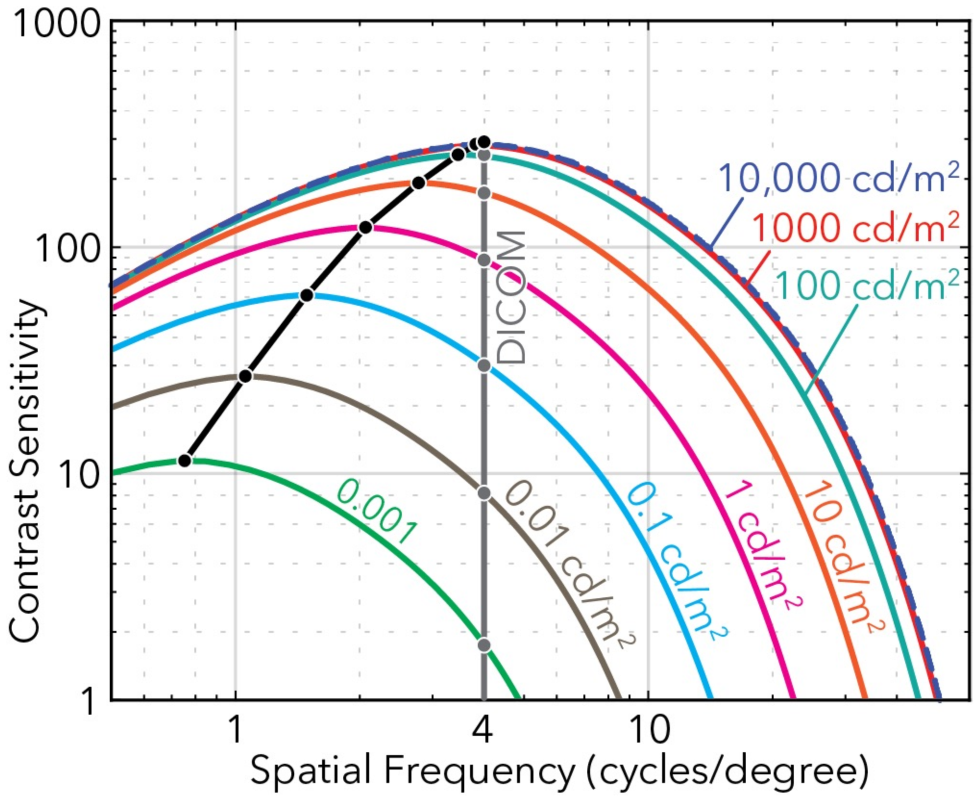
$$M_{opt}(u) = e^{-2\pi^2\sigma^2u^2}$$

$$\sigma = \sqrt{\sigma_0^2 + (C_{ab}d)^2}$$

$$d = 5 - 3\tanh\left(0.4\log\left(LX_0^2/40^2\right)\right)$$

$$E = \frac{\pi d^2}{4}L\left(1-(d/9.7)^2+(d/12.4)^4\right)$$

- Highest Sensitivity at any light adaptation level, regardless of frequency
- Must work with zero noise imagery
- Angular size of 40° instead of DICOM's 2°
- Recomputing the photon conversion factor for a D65 white point
- Tracking the peaks of contrast sensitivity instead of DICOM's fixed 4 cycles/degree



Reference: P. Whittle, "Increments and decrements: luminance discrimination", Vis Res. V 26, #10, 1677-1691, 1986.

Using Modulation to Set JND Steps

- Once we know the modulation threshold (m_t) at a given luminance level, **we can calculate the next JND step up or down** from that level

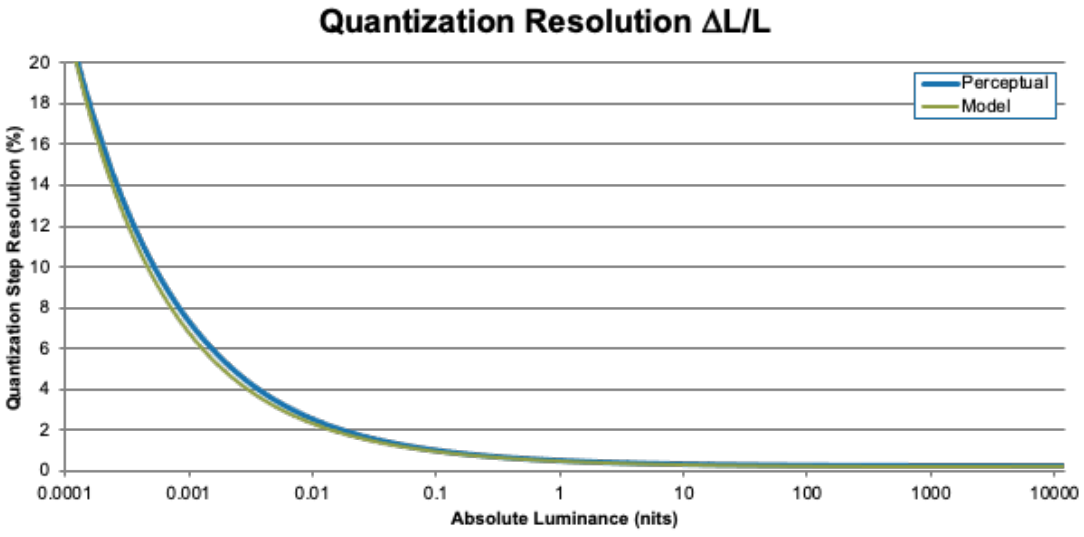
$$m_t = \frac{L_{\max} - L_{\min}}{L_{\max} + L_{\min}} \quad so: \quad L_{j+1} = L_j \frac{1+m_t}{1-m_t} \quad and \quad L_{j-1} = L_j \frac{1-m_t}{1+m_t}$$

- We can also **use fractions of a JND step to cover a desired range with a desired bit depth**

References:

1. S. Miller, M. Nezamabadi and S. Daly, "Perceptual Signal Coding for More Efficient Usage of Bit Codes," *The 2012 Annual Technical Conference & Exhibition*, 2012, SMPTE, pp. 1-9, doi: 10.5594/M001446.

Finding a Functional Representation



The perceptual model is a table built by iteration of fractional JNDs

- Awkward for standardization & specification

Identify functional form:

- Continuous function
- With no transitions required
- Invertible

Encode Equation

$$L = \left(\frac{D^{1/m} - c_1}{c_2 - c_3 D^{1/m}}\right)^{1/n}$$

Decode Equation

$$D = \left(\frac{c_1 + c_2 L^n}{1 + c_3 L^n}\right)^m$$

Absolute Boundaries:

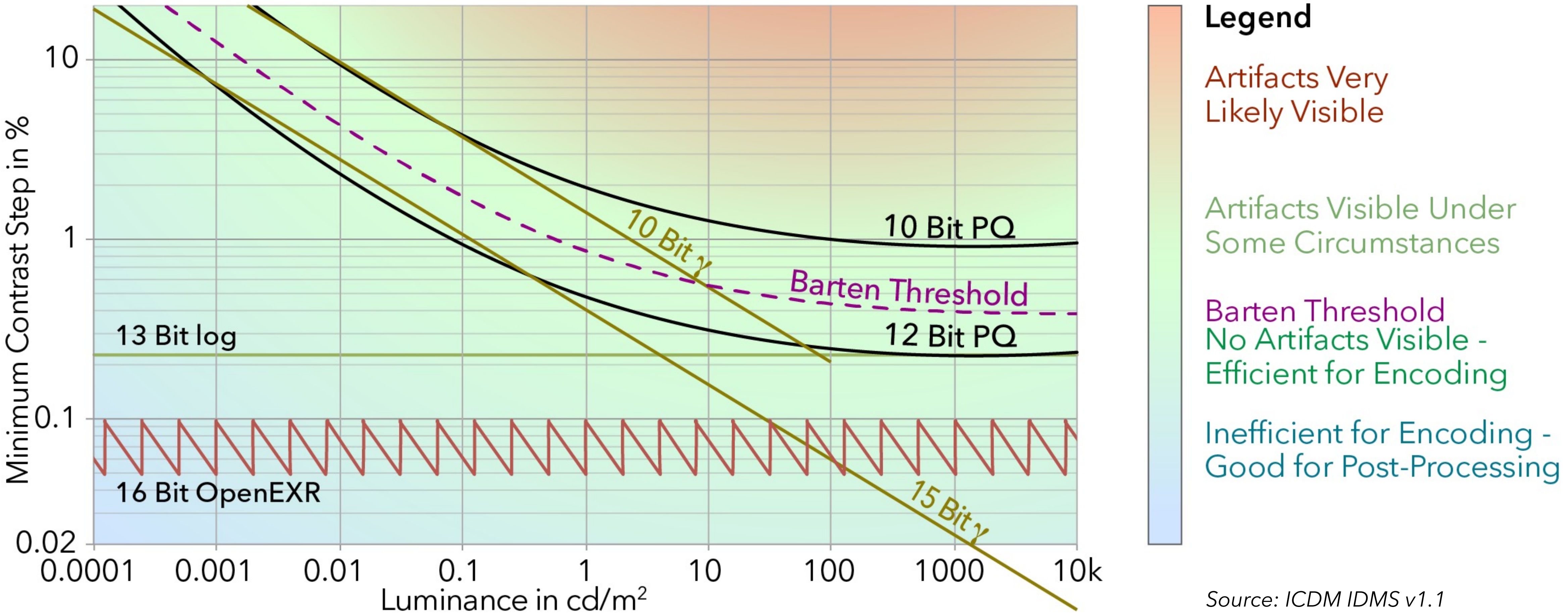
PQ$_{Max}$

- Accepted for PQ max: **10,000 cd/m²**
  - To fit to 12 Bits as usable by SDI
  - Also preferred by Hollywood studios & MovieLabs

PQ$_{Min}$

- Lowest Level of Visibility at 10^{-6} cd/m^2
- To simplify math, extended down to **0.0 cd/m²** (added negligible ‘cost’ to the accuracy)

Quantization Performance



Source: ICDM IDMS v1.1

Verification

Psychophysical Experiment: Detectability

- Test detectability thresholds using JND cross pattern & real images

Psychophysical Experiment: Preference Luminance Range

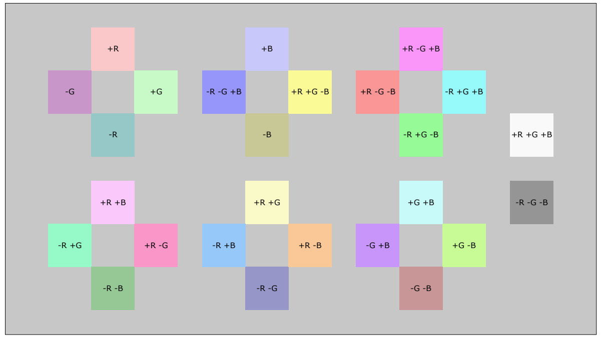
- In parallel to the mathematical modelling, a psychophysical experiment was carried out
- Objective: Identify the preference HDR luminance range
- Result: Reflected similar ranges for entertainment content

Vetting by Studios & Standardization Efforts

SMPTE standardization efforts involved extremely rigorous vetting of the technology using

- Professional, stable equipment, with careful calibration,
- Critical viewing, with all sorts of content and corner cases

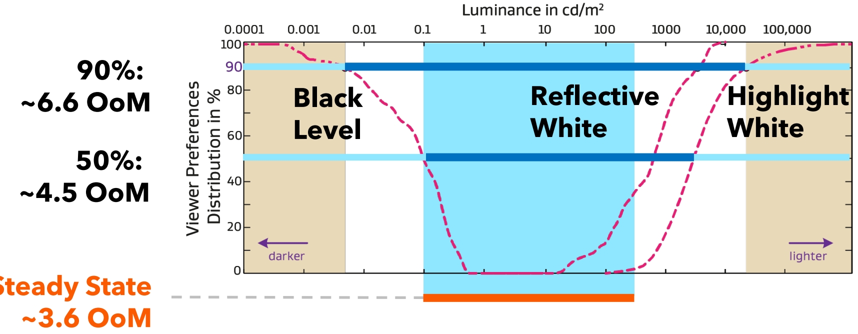
JND Cross Test Pattern (nearest neighbor test)[1]



Note: Here, patch color contrast to background is amplified for visualization.

Real Images[1]

Viewer Preferences for Shadow, Diffuse, Specular, and Emissive Luminance Limits[2]



References:
1. S. Miller, M. Nezamabadi and S. Daly, "Perceptual Signal Coding for More Efficient Usage of Bit Codes," *The 2012 Annual Technical Conference & Exhibition*, 2012, SMPTE, pp. 1-9, doi: 10.5594/M001446.
2. S. Daly, T. Kunkel, X. Sun, S. Farrell & P. Crum (2013): '41.1: Distinguished Paper: Viewer Preferences for Shadow, Diffuse, Specular, and Emissive Luminance Limits of High Dynamic Range Displays', SID Symposium Digest of Technical Papers, 44: 563-566. DOI: 10.1002/j.2168-0159.2013.tb06271.x

Final Form of the Perceptual Quantizer

In the end, 10 years ago, in January 2012, PQ took on its finished form with a range of 0 to 10,000 cd/m^2 and the following formula:

$$L = 10{,}000\left(\frac{\max\left[\left(V^{1/m_2} - c_1\right),0\right]}{c_2 - c_3 V^{1/m_2}}\right)^{1/m_1}$$

$$m_1 = \frac{2610}{4096}\times\frac{1}{4} = 0.1593017578125$$

$$m_2 = \frac{2523}{4096}\times 128 = 78.84375$$

$$c_1 = \frac{3424}{4096} = 0.8359375$$

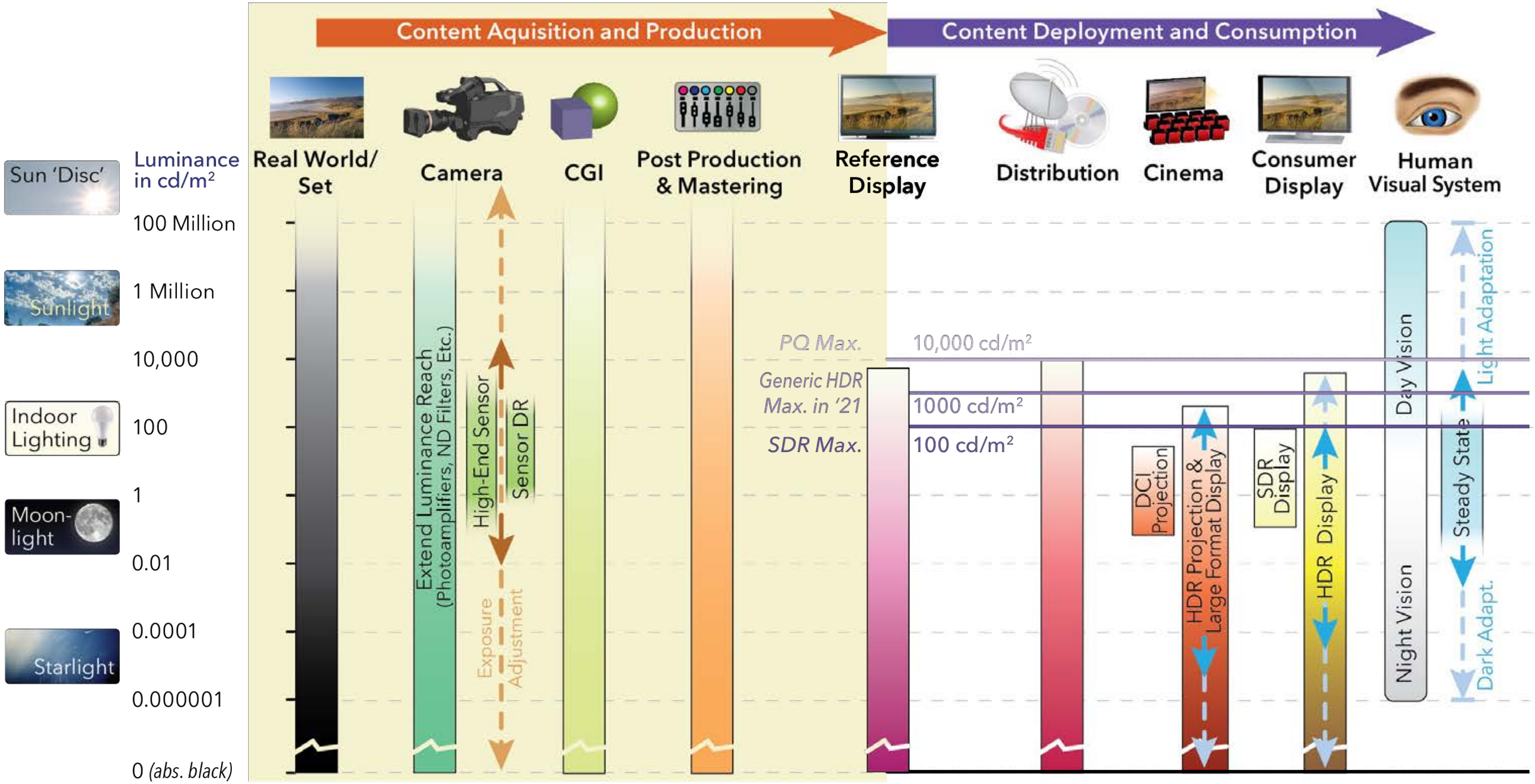
$$c_2 = \frac{2413}{4096}\times 32 = 18.8515625$$

$$c_3 = \frac{2392}{4096}\times 32 = 18.6875$$

Today, PQ is standardized in SMPTE ST.2084 and ITU-R Rec. BT.2100

Applications

Luminance Capabilities of Today's HDR Imaging Pipeline

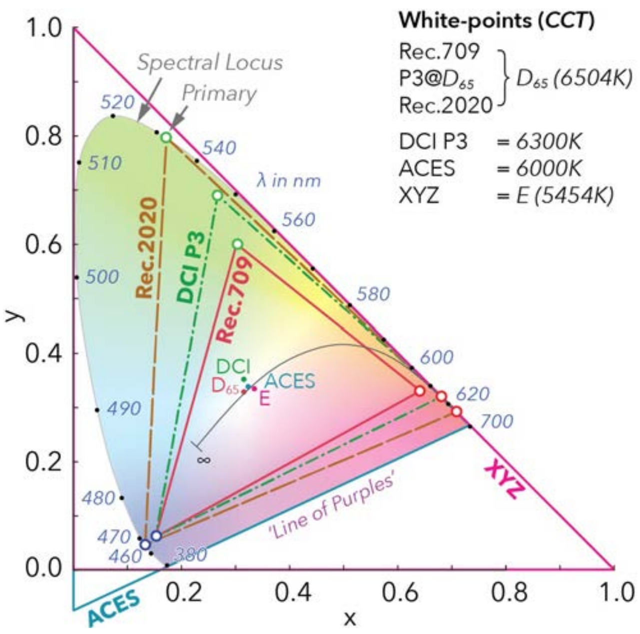


Capture, Postproduction & Deployment

- **Tools for capture, process and distribute HDR in PQ are deployed at scale**
- Both **professional and consumer level options** are available to create, process and distribute HDR PQ content
  - Most major video editing tools support PQ
  - Availability of higher end HDR-capable content editing and grading displays with high luminance capability of ~700 to 4000 cd/m^2
  - Support on computers, mobile devices, or tablets for UGC content

Display

- HDR capable displays are widely available in the market at different price points, and all support the PQ non-linearity
- Luminance capabilities
  - Deep blacks (<0.1 cd/m^2) down to no light emission
  - Maximum luminance levels reaching 700-1000cd/m^2 (some offer up to 2000 cd/m^2).
  - Color gamut has extended from ITU-R rec. BT.709 to P3 or towards ITU-R Rec. BT.2020
- HDR TVs & mobile phones have been around for several years,
- Computer & tablet displays are also catching up.



Key Standards for Color Gamut:

ITU-R Rec. BT.2020-2	Parameter values for ultra-high definition television systems for production and international programme exchange.
ITU-R Rec. BT.2100-2	Image parameter values for high dynamic range television for use in production and international programme exchange.

Availability

Displays

- Of **225 million TVs** sold globally in 2020,
- **58% included HDR** functionality
- **10% even provided 500 cd/m²** peak luminance or higher *(Source: OMDIA)*
- Many also support **dynamic metadata**-enabled HDR formats.

HDR supported on **PCs, mobile phones, tablets**, as well as **gaming platforms**

HDR content is readily available

- Cinema, Blu-ray, broadcast, gaming & OTT

Summary

Summary

1

Designing PQ

Complex considerations went into designing PQ

This included modelling as well as several psychophysical studies

2

PQ is established!

After its standardization in SMPTE ST.2084 and later ITU-R Rec. BT.2100, PQ has found wide-spread adoption with many HDR applications.

3

'Backbone' of Consumer HDR

PQ, together with HLG, form the backbone of today's consumer HDR.

Continue enjoying the quality and fidelity improvements, HDR continues to provide!

THANK YOU!

Contact: timo.kunkel@dolby.com

Acknowledgement: Scott Daly & Scott Miller

APPENDIX

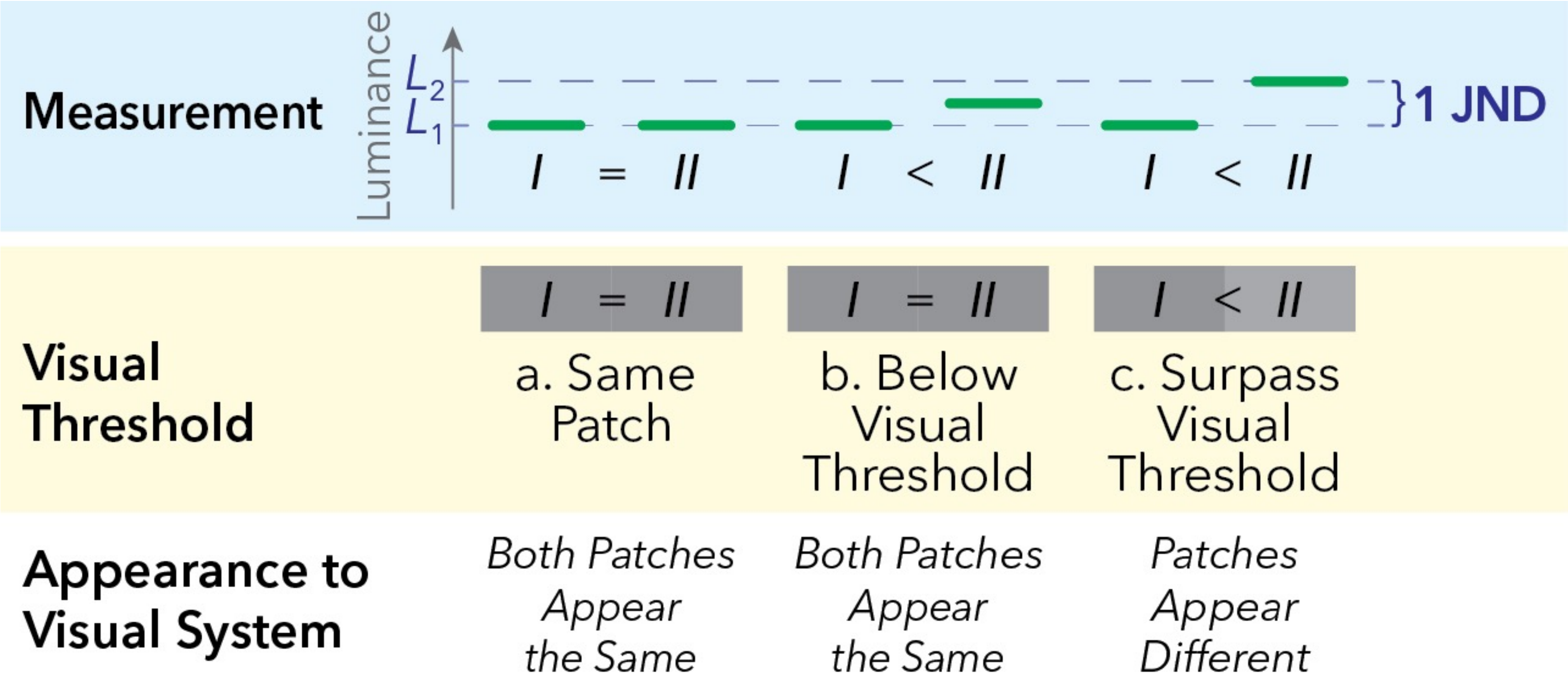
REFERENCES

Standards Related to HDR

• **ARIB STD-B67**	Essential Parameter Values for the Extended Image Dynamic Range Television (EIDRTV) System for Programme Production
• **CTA 861.3-A-2016**	HDR Static Metadata Extensions
• **EBU Tech 3320**	User Requirements for Video Monitors in Television Production. Version 4.1 (9/2019)
• **ETSI TS 103 433-1**	High-Performance Single Layer High Dynamic Range (HDR) System for use in Consumer Electronics devices
• **SID ICDM IDMS 1.1**	SID International Committee for Display Metrology (ICDM) Information Display Measurements Standard v1.1
• **ITU-R Rec. BT.601-7**	Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios. ITU-R. March 2011
• **ITU-R Rec. BT.709-6**	Parameter values for the HDTV standards for production and international programme exchange. ITU-R. June 2015
• **ITU-R Rec. BT.1886**	Reference electro-optical transfer function for flat panel displays used in HDTV studio production. ITU-R. March 2011
• **ITU-R Rec. BT.2020-2**	Parameter values for ultra-high definition television systems for production and international programme exchange. Oct. 2015
• **ITU-R Rec. BT.2035**	A reference viewing environment for evaluation of HDTV program material or completed programmes
• **ITU-R Rec. BT.2100-2**	Image parameter values for high dynamic range television for use in production and international programme exchange. July 2018
• **ITU-R Rec. BT.2124**	Objective metric for the assessment of the potential visibility of colour differences in television. ITU-R. Jan. 2019.
• **ITU-R Rec. BT.2408-0**	Operational practices in HDR television production. ITU-R. Oct. 2017.
• **ITU-T Rec. H.273**	Coding-independent code points for video signal type identification, 2016
• **OpenEXR**	High-dynamic-range scene-linear image data and associated metadata format. www.openexr.com
• **SMPTE ST 0196-2003**	Motion-Picture Film - Indoor Theater and Review Room Projection - Screen Luminance & Viewing Conditions
• **SMPTE ST 0431-1-2006**	D-Cinema Quality - Screen Luminance Level, Chromaticity and Uniformity
• **SMPTE RP 0431-2-2007**	D-Cinema Quality - Reference Projector and Environment
• **SMPTE ST 2084:2014**	High Dynamic Range Electro-Optical Transfer Function of Mastering Reference Displays. SMPTE 2014
• **SMPTE ST 2086:2018**	Mastering Display Color Volume Metadata Supporting High Luminance and Wide Color Gamut Images
• **SMPTE ST 2094-0:2017**	Overview Document - Dynamic Metadata for Color Volume Transformation. DOI: 10.5594/SMPTE.OV2094-0.2017.
• **VESA DisplayHDR:**	VESA High-performance Monitor and Display Compliance Test Specification (DisplayHDR CTS). Rev. 1.1 (2019). www.displayhdr.org

Appendix

What is a JND (Just Noticeable Difference)



Source: ICDM IDMS v1.1

Crispening Effect

- The visual system is more **sensitive to lightness variations around the background** luminance
- This is known as the Crispening Effect.

Light Background | **Dark Background** | **Background Color Close to Patches**

Patch 1

Patch 2

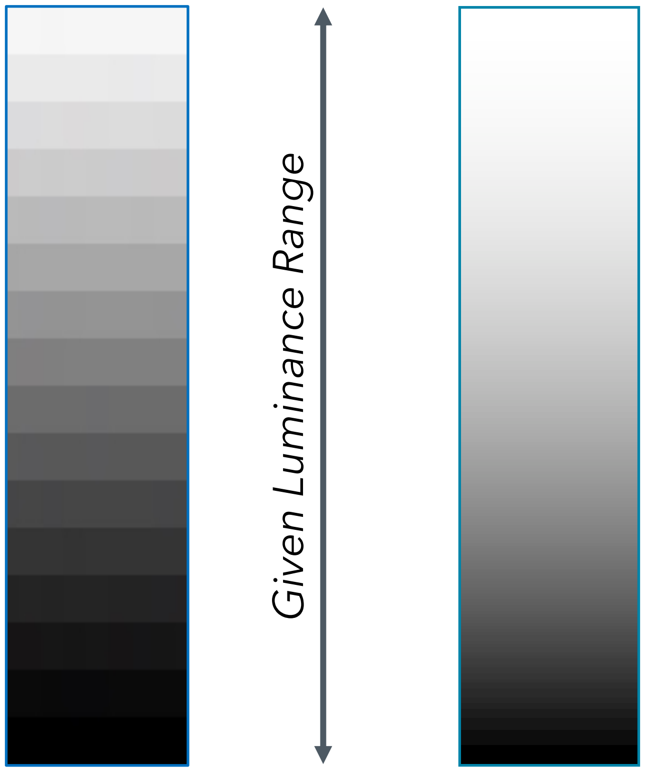
Color Difference appears to be **small** | Color Difference appears to be **small** | Color Difference appears to be **larger**

References:
- Takasaki, H. (1966). Lightness change of gray induced by change in reflectance of gray background. Journal of the Optical Society of America, 56(4), 504-509. https://doi.org/10.1364/JOSA.56.000504
- Takasaki, H. (1967). Chromatic Changes Induced by Changes in Chromaticity of Background of Constant Lightness. Journal of the Optical Society of America, 57(1), 93-96. https://doi.org/10.1364/JOSA.57.000093
- Semmelroth, C. C. (1970). Prediction of Lightness and Brightness on Different Backgrounds. Journal of the Optical Society of America, 60(12), 1685-1689.
- P. Whittle, "Increments and decrements: luminance discrimination", Vis Res. V 26, #10, 1677-1691, 1986.

Contouring

10k cd/m^2

Insufficient Intervals:
Contouring visible



Sufficiently Small Intervals:
'Smooth' Ramp

10^{-6} cd/m^2

The Barten Model

The Barten model parameters as used for PQ

$$CSF = \frac{1}{m_t} = \frac{M_{opt}(u)/k}{\sqrt{\frac{2}{T}\left(\frac{1}{X_0^2} + \frac{1}{X_{\max}^2} + \frac{u^2}{N_{\max}^2}\right)\left(\frac{1}{\eta p E} + \frac{\Phi_0}{1 - e^{-(u/u_0)^2}}\right)}}$$

$$M_{opt}(u) = e^{-2\pi^2\sigma^2u^2}$$

$$\sigma = \sqrt{\sigma_0^2 + (C_{ab}d)^2}$$

$$d = 5 - 3\tanh\left(0.4\log\left(L\,X_0^2/40^2\right)\right)$$

$$E = \frac{\pi d^2}{4}L\left(1 - (d/9.7)^2 + (d/12.4)^4\right)$$

$k = 3.0$

$\sigma_0 = 0.5$ arcmin

$C_{ab} = 0.08$ arcmin/mm

$T = 0.1$ sec

$X_0 = 40°$

$X_{\max} = 12°$

$N_{\max} = 15$ cycles

$\eta = 0.03$

$\Phi_0 = 3\times10^{-8}$ sec deg^2

$u_0 = 7$ cycles/deg

$p = 1.25\times10^6$ photons/sec/deg^2/Td