A Quantitative Framework for Evaluating Sequential Color Palette Quality in Design Systems

Huu Khanh Nguyen
Independent Researcher
Email: chromametry@gmail.com

***Abstract*—The design of accessible color systems for digital interfaces currently lacks unified quantitative standards, leading to subjective and manual workflows. This document introduces a quantitative framework for evaluating sequential monochromatic color palettes by defining five principal evaluation dimensions: Contrast Efficiency, lightness linearity, chroma smoothness, hue stability, and spacing uniformity. The framework adopts the CIELAB color space for its interpretability and its role as the reference color space for computing perceptual differences using CIEDE2000. By benchmarking eleven industry-leading design systems, the framework objectively differentiates palette quality, with composite scores ranging from 57.16 to 88.59, identifying Adobe Spectrum and IBM Carbon as high-performing benchmarks. This work establishes a quantitative foundation for systematic palette evaluation and comparative analysis.**

Keywords: color ramp; colormap; monochromatic; color accessibility; color palette; design systems

This is the first of two papers from the Domphy project. The second paper addresses the spatial axis (unified sizing model). Together they form a quantitative foundation for UI design systems.
Open-source implementation and benchmarks: https://github.com/chromametry/chromametry
Production integration: https://github.com/domphy/domphy

I. Introduction

Designing color systems for digital interfaces requires balancing uniformity with accessibility constraints. In practice, the absence of a unified quantitative evaluation framework leads to several key limitations:

- **High subjectivity:** Color palette quality assessment remains largely perceptual, lacking objective measurement indicators for comparing different systems.
- **Automation barriers:** Without standardized metrics, the generation of systematic color scales (M hues $\times$ N steps) remains predominantly manual. Existing interpolation methods struggle to preserve linearity and accessibility in the absence of quantitative validation criteria.
- **Combinatorial contrast testing:** Verifying WCAG-compliant contrast across all color pairs is operationally inefficient, often requiring exhaustive inspection during both design and deployment stages.
- **Poor maintainability and scalability:** Color systems without quantitative foundations face difficulties when extending palette ranges or performing synchronized updates while preserving brand-specific characteristics.

This work aims to shift color system design from a heuristic-driven process toward a metrics-based methodology, establishing a quantitative reference for systematic palette evaluation and comparison.

II. Related Work

A. Color Palette Generation

Early work on systematic palette construction focused on perceptual uniformity in color space selection. Zeileis et al. [4] proposed HCL-based palettes for statistical graphics, emphasizing perceptual linearity over aesthetics. Gramazio et al. [11] introduced *Colorgorical*, a palette generation model that jointly optimizes perceptual discriminability and aesthetic preference, establishing discriminability and preference as dual axes of palette quality. Stone et al. [15] examined how step count and lightness range interact in sequential palettes, an observation that motivates our analysis of even-step advantages in contrast span.

B. Color in Design Systems

Norman et al. [13] conducted a qualitative study of professional designer workflows, identifying three interlinked design spaces: purpose, palette, and prototype. Their findings motivate computational scoring that targets purpose-driven constraints (accessibility) rather than aesthetic preference alone. Color Builder [14] provided a direct-manipulation interface for theme construction but offered no quantitative quality criteria for the resulting palettes.

C. Color Theme Evaluation

Yang et al. [12] proposed a computational model for scoring color themes through user preference modelling (ACM TAP, 2024). While the most closely related work in spirit, their approach is *subjective*—trained on perceptual ratings—and evaluates multi-color palettes as sets. By contrast, our framework is *objective*, grounded in geometric properties of the continuous parametric curve underlying a single-hue ramp, and targets design-system scalability rather than global aesthetic appeal.

D. Accessibility Metrics

Existing accessibility evaluation follows WCAG 2.1 [6] contrast ratio thresholds, which are applied pairwise and do not characterize palette-level accessibility structure. Prior tools check compliance post-hoc rather than guiding generation.

Our contrast efficiency metric η reformulates accessibility as a structural property of the ramp, quantifying how economically the required contrast separation is achieved across the full step range.

E. Positioning

We are not aware of prior work that addresses the problem of scoring a *single base color's capacity* to generate an accessible, perceptually coherent sequential scale across industry-standard step counts—a requirement specific to design token systems. The proposed framework aims to fill this gap by combining perceptual geometry, accessibility constraints, and continuous-curve evaluation into a unified, steps-independent score.

III. Definitions and Standards

A. Color Palette Structure

A palette P comprises n monochromatic families:

$$P = \{S_1, S_2, \ldots, S_n\}.$$

Each color family S_j is a *monochromatic ramp* consisting of N lightness steps, generated from a single base color. Formally,

$$S_j = \{c_{j,0}, c_{j,1}, \ldots, c_{j,N-1}\}.$$

Colors within each family are arranged in a strictly monotonic order of perceived lightness, forming a one-dimensional lightness ramp. Within S_j, there exists a unique base color $C_{\text{base},j} \in S_j$ satisfying

$$C_{\text{base},j} = \arg \max_{c_{j,i} \in S_j} C^*(c_{j,i}).$$

This color possesses the maximum chroma in the range and serves as the representative hue for the entire family. By definition, S_j is assumed to be the result of a fixed-hue interpolation where only lightness and its dependent chroma vary. This assumption ensures consistent reference quantities $(L_{\text{base}}, h_{\text{base}}, C_{\text{base}})$ for all subsequent metrics.

B. WCAG Contrast Ratio

According to WCAG 2.1 [6], the contrast ratio is defined based on relative luminance $Y \in [0, 1]$:

$$CR(c_1, c_2) = \frac{Y_1 + 0.05}{Y_2 + 0.05}, \quad Y_1 \geq Y_2.$$

Relative luminance is calculated as

$$Y = 0.2126\, R_{\text{linear}} + 0.7152\, G_{\text{linear}} + 0.0722\, B_{\text{linear}}.$$

Standard contrast ratio thresholds include:

- **3:1** for UI components and large text.
- **4.5:1** for body text (AA standard).
- **7:1** for enhanced readability (AAA).

This framework identifies the **4.5:1 threshold** as the primary benchmark for comparative evaluation.

TABLE I
Illustration using $N = 12$ with contrast span $K = 6$.

#	Indices	0	1	2	3	4	5	6	7	8	9	10	11
1	span 0–6	[0	1	2	3	4	5	6]	7	8	9	10	11
2	span 1–7	0	[1	2	3	4	5	6	7]	8	9	10	11
3	span 2–8	0	1	[2	3	4	5	6	7	8]	9	10	11
4	span 3–9	0	1	2	[3	4	5	6	7	8	9]	10	11
5	span 4–10	0	1	2	3	[4	5	6	7	8	9	10]	11
6	span 5–11	0	1	2	3	4	[5	6	7	8	9	10	11]

C. Contrast Span

The *Contrast Span* K of a hue family is the minimum index distance that guarantees a contrast ratio ≥ 4.5:1 for all color pairs:

$$K = \min \{k \in \mathbb{N} : CR(c_i, c_{i+k}) \geq 4.5,\ \forall i \in [0, N-k-1]\}.$$

For multi-family palettes, the global span is defined as

$$K_{\text{palette}} = \max_{j=1}^{n} K_j.$$

This metric enables predictable accessibility: designers can ensure compliant contrast by selecting colors separated by at least K steps without requiring runtime calculations. As illustrated in Table I, a fixed span guarantees WCAG-compliant contrast across all valid index pairs.

IV. Evaluation Metrics

This section defines five quantitative metrics for evaluating sequential monochromatic color palettes. Three metrics assess the internal behavior of individual CIELAB components: lightness linearity, chroma smoothness, and hue stability, capturing how each perceptual dimension evolves along a color ramp. A fourth metric evaluates perceptual spacing uniformity using the CIEDE2000 color difference, quantifying the consistency of stepwise transitions. The final metric measures contrast efficiency by analyzing the accessibility span required to satisfy the WCAG 4.5:1 contrast ratio. Together, these five metrics characterize both the perceptual structure and accessibility performance of a color palette.

A. Contrast Efficiency (η)

Contrast Efficiency evaluates the economy of index separation required to satisfy accessibility constraints. To establish a **reference baseline** for accessibility, this framework derives the minimum lightness fraction needed for WCAG compliance from a **neutral gray ramp** ($a^* = 0, b^* = 0$).

The choice of an achromatic scale as reference is motivated by its analytical tractability: neutral colors provide the most direct mapping between relative luminance Y and perceived lightness L^*, as they are unaffected by the Helmholtz–Kohlrausch effect—where high chroma increases perceived brightness despite constant luminance [3]. Furthermore, neutral scales occupy the achromatic axis of the color gamut, minimizing non-linearities induced by gamut mapping constraints common in highly saturated hues. Because the WCAG contrast ratio depends solely on relative luminance (not on chrominance), the achromatic derivation yields a λ value that applies to any hue family whose lightness-to-luminance mapping is monotonic—a condition satisfied by all well-formed sequential ramps within the sRGB gamut.

In this ideal achromatic case, L^* is related to relative luminance $Y \in [0,1]$ by the CIE 1976 transform [1]:

$$L^*(Y) = \begin{cases} 116\,Y^{1/3} - 16, & Y > \left(\frac{6}{29}\right)^3 \\ \frac{116}{3}\left(\frac{29}{6}\right)^2 Y, & Y \le \left(\frac{6}{29}\right)^3 \end{cases}$$

In the linear branch, the additive constants cancel: $116 \cdot \frac{4}{29} - 16 = 0$, yielding the simplified form $L^* \approx 903.3\,Y$.

Applying the WCAG 4.5:1 boundary conditions:

- For $Y_{\min} = 0$ (black background), the required boundary luminance is $Y_{\text{black}} = 0.175 \Rightarrow L^*_{\text{black}} \approx 48.9$, giving $\lambda_1 \approx 0.489$.
- For $Y_{\max} = 1$ (white background), the required boundary luminance is $Y_{\text{white}} \approx 0.183 \Rightarrow L^*_{\text{white}} \approx 49.9$, whose distance from $L^* = 100$ gives $\lambda_2 \approx 0.501$.

Since accessibility must hold under both black-on-white and white-on-black conditions, this work adopts the conservative estimate

$$\lambda = \max(\lambda_1, \lambda_2) \approx 0.501.$$

For a palette with N discrete steps, let K denote the contrast span, i.e. the minimum index gap such that all pairs (c_i, c_{i+K}) satisfy the WCAG threshold.

The corresponding **ideal contrast span** is defined as:

$$K_{\text{ideal}} = \lceil \lambda \cdot (N-1) \rceil$$

The **observed density** is defined as:

$$D = \frac{K}{N-1}$$

a) Continuity and discretization.: Color ramps are generated from an underlying continuous lightness trajectory, while N represents a discrete sampling of that trajectory. The span K is therefore integer-valued and changes stepwise as N varies, producing quantization artifacts in direct score calculation. By normalizing to $D = K/(N-1)$, the metric approximates the continuous position of the accessibility boundary along the ramp.

The ideal accessibility boundary is defined in the continuous domain by λ. As $N \to \infty$, the discrete span satisfies

$$\frac{K_{\text{ideal}}}{N-1} \to \lambda,$$

ensuring consistency between discrete construction and continuous theory.

For example, with $N = 12$:

- $K_{\text{ideal}} = \lceil 0.501 \times 11 \rceil = 6$
- $D = \frac{K}{11}$ (e.g., $K = 6 \Rightarrow D \approx 0.545 > \lambda$)

Contrast Efficiency η is defined by comparing the observed density D against the continuous reference λ:

$$\eta = \begin{cases} 1, & D \le \lambda \\ 0, & D \ge 1 \\ 1 - \dfrac{D-\lambda}{1-\lambda}, & \text{otherwise} \end{cases}$$

TABLE II
REFERENCE VALUES: K_{IDEAL} (DESIGN TARGET) FOR $\lambda \approx 0.501$

Steps (N)	Formula	K_{ideal}
10	$\lceil \lambda \cdot 9 \rceil$	5
11	$\lceil \lambda \cdot 10 \rceil$	5
12	$\lceil \lambda \cdot 11 \rceil$	6
13	$\lceil \lambda \cdot 12 \rceil$	6
14	$\lceil \lambda \cdot 13 \rceil$	7
15	$\lceil \lambda \cdot 14 \rceil$	7
16	$\lceil \lambda \cdot 15 \rceil$	8
17	$\lceil \lambda \cdot 16 \rceil$	8
18	$\lceil \lambda \cdot 17 \rceil$	9

TABLE III
DERIVED λ VALUES AND K_{IDEAL} FOR STANDARD WCAG CONTRAST THRESHOLDS ($N = 18$). BOUNDARY LUMINANCES Y_{BLACK} AND Y_{WHITE} ARE THE EXTREMAL TEXT LUMINANCES SATISFYING CR $\ge R$ AGAINST A PURE BLACK OR WHITE BACKGROUND, RESPECTIVELY.

R	Y_{black}	L^*_{black}	λ_1	Y_{white}	L^*_{white}	λ_2	$\lambda = \max(\lambda_1, \lambda_2)$
2:1	0.050	26.7	0.267	0.475	74.5	0.255	0.267
3:1	0.100	37.8	0.378	0.300	61.7	0.383	0.383
4.5:1	0.175	48.9	0.489	0.183	49.9	0.501	0.501
7:1	0.300	61.7	0.617	0.100	37.8	0.622	0.622

R	λ	$K_{\text{ideal}} = \lceil \lambda \cdot (N-1) \rceil$	K_{ideal} ($N = 18$)
2:1	0.267	$\lceil \lambda \cdot (N-1) \rceil$	5
3:1	0.383	$\lceil \lambda \cdot (N-1) \rceil$	7
4.5:1	0.501	$\lceil \lambda \cdot (N-1) \rceil$	9
7:1	0.622	$\lceil \lambda \cdot (N-1) \rceil$	11

Extension to Other WCAG Thresholds: The same derivation can be applied to any WCAG contrast ratio R. The boundary conditions yield:

- **Black background** ($Y_{\min} = 0$): the lightest text satisfying CR $\ge R$ has $Y_{\text{text}} = 0.05(R-1)$, located at $L^*_{\text{black}} = L^*(0.05(R-1))$, whose distance from $L^* = 0$ gives $\lambda_1 = L^*_{\text{black}}/100$.
- **White background** ($Y_{\max} = 1$): the darkest text satisfying CR $\ge R$ has $Y_{\text{text}} = 1.05/R - 0.05$, located at $L^*_{\text{white}} = L^*(1.05/R - 0.05)$, whose distance from $L^* = 100$ gives $\lambda_2 = (100 - L^*_{\text{white}})/100$.

Taking the conservative estimate

$$\lambda = \max(\lambda_1, \lambda_2),$$

the same formulation applies:

$$K_{\text{ideal}} = \lceil \lambda \cdot (N-1) \rceil$$

A notable structural property emerges: the λ values for 3:1 and 7:1 are approximately complementary ($\lambda_{3:1} + \lambda_{7:1} \approx 1.00$), as are those for 2:1 and a ratio near 10:1. This symmetry arises because the two boundary conditions for ratio R and its reciprocal role (light-on-dark vs. dark-on-light) swap the same luminance values $Y = 0.05(R-1)$ and $Y = 1.05/R - 0.05$, producing mirror-image distances along the L^* axis.

B. Lightness Linearity ($\mathcal{L}$)

The **Lightness Linearity** metric evaluates the degree to which *perceptual* lightness progression in a palette follows a linear trend with step index.

1) Motivation: A linear lightness trajectory is desirable for two reasons. First, it ensures that the perceptual distance between any two steps separated by the same index gap k remains approximately constant, which is a prerequisite for predictable contrast pairing (cf. the contrast span K in Section IV-A). Second, non-linear progressions—such as lightness plateaus or abrupt jumps—produce uneven visual weight when palette steps are mapped to sequential data or UI surfaces, undermining the ordinal semantics that sequential palettes are designed to convey [4].

Unlike pure L^*, lightness is affected by the Helmholtz–Kohlrausch effect [3], where high-chroma colors are perceived as lighter than achromatic colors with the same L^*.

2) Equivalent Achromatic Lightness: To compensate for this effect, we use **Equivalent Achromatic Lightness (EAL)** according to High et al. [3]:

$$L_{\mathrm{EAL}} = L^* + \big(f_{BY}(h) + f_R(h)\big)\, C^*$$

where chroma and hue in CIELAB space are determined by

$$C = \sqrt{a^{*2} + b^{*2}}$$

$$h = \mathrm{atan2}(b^*, a^*)$$

The hue-dependent correction functions are given by

$$f_{BY}(h) = 0.1644 \left| \sin\left(\frac{h - 90^\circ}{2}\right) \right| + 0.0603$$

$$f_R(h) = \begin{cases} 0.1307\,|\cos(h)| + 0.0060, & h \in [0^\circ, 90^\circ] \cup [270^\circ, 360^\circ] \\ 0, & \text{otherwise.} \end{cases}$$

3) Linear Regression by Step Index: For a hue family of N colors ordered by **monotonic lightness progression** (either strictly increasing or strictly decreasing), let $L_{\mathrm{EAL},i}$ denote the EAL value at step i. A linear model is fitted using **ordinary least squares**:

$$\hat{L}_{\mathrm{EAL}}(i) = \alpha i + \beta, \qquad i = 0, \ldots, N-1.$$

Here, the sign of α implicitly captures the direction of progression: $\alpha > 0$ corresponds to increasing lightness, while $\alpha < 0$ corresponds to decreasing lightness.

The root mean square error is computed as

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} \big(L_{\mathrm{EAL},i} - \hat{L}_{\mathrm{EAL}}(i)\big)^2}.$$

4) Normalization by Fitted Range: To normalize error independently of absolute lightness magnitude and progression direction, error is normalized by the **extremal values of the fitted line**:

$$L^{\mathrm{fit}}_{\mathrm{min}} = \min\big(\hat{L}_{\mathrm{EAL}}(0),\ \hat{L}_{\mathrm{EAL}}(N-1)\big).$$

$$L^{\mathrm{fit}}_{\mathrm{max}} = \max\big(\hat{L}_{\mathrm{EAL}}(0),\ \hat{L}_{\mathrm{EAL}}(N-1)\big).$$

At each step i, the maximum allowable deviation within this fitted range is

$$\epsilon_i^{\max} = \max\left(\left| \hat{L}_{\mathrm{EAL}}(i) - L^{\mathrm{fit}}_{\mathrm{min}} \right|, \left| L^{\mathrm{fit}}_{\mathrm{max}} - \hat{L}_{\mathrm{EAL}}(i) \right| \right).$$

From this, the maximum normalized error is defined as

$$RMSE_{\max} = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} \big(\epsilon_i^{\max}\big)^2}.$$

This normalization constrains the metric to the envelope of the fitted linear model and is invariant to whether the palette progresses from dark-to-light or light-to-dark, ensuring score stability independent of absolute position or direction on the L^* axis.

5) Metric Definition: The **Lightness Linearity** metric is defined by

$$\mathcal{L} = \max\left(0,\ 1 - \frac{RMSE}{RMSE_{\max}}\right).$$

The value $\mathcal{L} \in [0, 1]$ approaches 1 when lightness progression closely follows a linear trend; strong local fluctuations or non-linearity will decrease the score.

In degenerate cases where the fitted line magnitude is negligible (palette nearly flat in lightness), the metric is conventionally set to $\mathcal{L} = 1$.

6) Interpretation: This metric evaluates **consistency of lightness progression** rather than purely geometric L^*. Using EAL allows $\mathcal{L}$ to more accurately reflect users' visual perception, especially for high-chroma palettes where the Helmholtz–Kohlrausch effect plays a significant role.

C. Chroma Smoothness ($\mathcal{S}_C$)

1) Theoretical Background: Zeileis et al. (2009) Power Function Model [4]: Zeileis et al. [4] propose a power-function-based parameterization to control the rate of chroma and lightness variation along a sequential color scale, improving contrast distribution. Using a *normalized continuous position* $t \in [0, 1]$, the model is defined as

$$\begin{cases} H(t) = H_2 - t\,(H_1 - H_2), \\ C(t) = C_{\max} - t^{p_1}\,(C_{\max} - C_{\min}), \\ L(t) = L_{\max} - t^{p_2}\,(L_{\max} - L_{\min}), \end{cases} \tag{1}$$

where H_1, H_2 are hue values at the two ends of the scale; $C_{\max}, C_{\min}$ and $L_{\max}, L_{\min}$ are chroma and lightness bounds; and $p_1, p_2 > 0$ control the curvature of chroma and lightness variation.

TABLE IV
DETERMINATION OF REFERENCE CHROMA C^*_{ref} FROM THE SRGB GAMUT.

Vertex (sRGB)	(R, G, B)	(L^*, a^*, b^*)	C^*
Black	$(0, 0, 0)$	$(0.0,\ 0.0,\ 0.0)$	0.0
Red	$(1, 0, 0)$	$(53.2,\ 80.1,\ 67.2)$	≈ 104.6
Green	$(0, 1, 0)$	$(87.7, -86.2,\ 83.2)$	≈ 119.8
Blue	$(0, 0, 1)$	$(32.3,\ 79.2, -107.8)$	≈ 133.8
Yellow	$(1, 1, 0)$	$(97.1, -21.6,\ 94.5)$	≈ 97.0
Magenta	$(1, 0, 1)$	$(60.3,\ 98.2, -60.8)$	≈ 115.5
Cyan	$(0, 1, 1)$	$(91.1, -48.1, -14.1)$	≈ 50.1
White	$(1, 1, 1)$	$(100.0,\ 0.0,\ 0.0)$	0.0

2) Derivative Discontinuity and Kink Artifacts: Although the single power-function model guarantees monotonicity, applying it to sequential palettes with *maximum chroma near the center* typically requires piecewise definitions (dark-to-peak and peak-to-light). The chroma derivative is

$$\frac{dC}{dt} = -p_1\, t^{p_1 - 1}\,(C_{\max} - C_{\min}). \tag{2}$$

When parameters differ across the two segments, the first derivative becomes discontinuous at the peak position t_{peak}:

$$\lim_{t \to t^-_{\text{peak}}} \frac{dC}{dt} \neq \lim_{t \to t^+_{\text{peak}}} \frac{dC}{dt}. \tag{3}$$

This lack of C^1 continuity produces a sharp chroma cusp, which can induce visible artifacts (Mach bands) due to the human visual system's sensitivity to first-derivative luminance changes [8]. This motivates the use of *monotonic cubic splines* [2], which allow enforcing a zero derivative at the chroma peak while preserving global monotonicity.

3) Chroma in CIELAB Space: Chroma in CIELAB space is computed as

$$C^* = \sqrt{a^{*2} + b^{*2}}. \tag{4}$$

While C^* is not perceptually uniform, it provides a device-independent chroma magnitude suitable for relative comparison when properly normalized.

4) Reference Chroma Standard (C^*_{ref}): To stabilize chroma magnitude across palettes, chroma is first normalized by the maximum chroma within the palette and subsequently rescaled into a fixed reference frame derived from the sRGB gamut extremum in CIELAB space.

Let C^*_i denote the CIELAB chroma of step i, and define

$$C^*_{\max} = \max_i C^*_i. \tag{5}$$

The rescaled chroma used for smoothness evaluation is

$$\tilde{C}_i = \frac{C^*_i}{C^*_{\max}} \cdot C^*_{\text{ref}}. \tag{6}$$

Evaluating primary and secondary vertices of the sRGB cube (Table IV) shows that the maximum chroma occurs at the blue primary, yielding

$$C^*_{\text{ref}} \approx 133.8. \tag{7}$$

The reference is derived from the sRGB gamut because it remains the dominant color space for web content delivery [10]. For wide-gamut displays (e.g., Display P3), C^*_{ref} would increase, but because chroma enters the metric only as a *ratio* $C^*_i / C^*_{\max}$ that is subsequently rescaled, the smoothness score depends on the *shape* of the chroma curve rather than its absolute magnitude. Replacing C^*_{ref} with a P3-derived value would therefore affect only palettes whose peak chroma exceeds the sRGB boundary, leaving the vast majority of current design-system palettes unaffected.

5) Ideal Chroma Trajectory: An ideal chroma trajectory is assumed to increase monotonically from the palette start, reach a single maximum, and then decrease monotonically toward the end. This unimodal assumption follows directly from the structure of sequential ramps anchored at white and black: both endpoints have $C^* = 0$ by definition, so chroma must rise and fall at least once. A single peak is the simplest such trajectory and matches the design intent of monochromatic families, where the "brand color" (maximum saturation) occupies the mid-range while extremes fade toward the achromatic anchors [4]. For a discrete palette of N colors indexed by $i = 0, \ldots, N-1$, the ideal trajectory $C_{\text{ideal}}(i)$ is constructed using a *monotonic cubic spline* [2] passing through three anchors:

- $(0, C_0)$ — start chroma,
- $(i_{\text{peak}}, \tilde{C}_{\max})$ — maximum chroma,
- $(N-1, C_{N-1})$ — end chroma.

This avoids polynomial oscillation (Runge's phenomenon) [9] and enforces smooth, monotonic chroma variation.

6) Smoothness Metric Definition: Deviation from the ideal trajectory is measured using root mean square error:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} \left(\tilde{C}_i - C_{\text{ideal}}(i)\right)^2}. \tag{8}$$

The maximum deviation envelope is defined as

$$\epsilon_{\max,i} = \max\Big(C_{\text{ideal}}(i) - \tilde{C}_{\min}, \tilde{C}_{\max} - C_{\text{ideal}}(i)\Big), \tag{9}$$

where $\tilde{C}_{\min} = \min_i \tilde{C}_i$ and $\tilde{C}_{\max} = \max_i \tilde{C}_i$.

The maximum achievable error is

$$RMSE_{\max} = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} \epsilon^2_{\max,i}}. \tag{10}$$

The *Chroma Smoothness* metric is defined as

$$\mathcal{S}_C = \max\left(0,\ 1 - \frac{RMSE}{RMSE_{\max}}\right). \tag{11}$$

7) Interpretation: $\mathcal{S}_C \approx 1$ indicates smooth chroma progression closely matching the ideal trajectory, while $\mathcal{S}_C \to 0$ reflects strong chroma oscillation or derivative discontinuities. For achromatic or near-achromatic palettes where $C^*_{\max} \approx 0$, chroma smoothness is defined as maximal, reflecting the absence of chroma variation.

D. Hue Stability ($\mathcal{H}$)

The **Hue Stability** metric evaluates the consistency of hue in a monochromatic color scale as lightness varies. A palette is considered stable if hue fluctuates minimally around a reference color rather than drifting toward other hues.

1) Hue Representation and Periodicity Handling: Hue in CIELAB space is determined by

$$h_i = \operatorname{atan2}(b_i^*, a_i^*). \tag{12}$$

Since hue is a periodic angular quantity, the hue sequence is **unwrapped** to remove discontinuities at 360° and ensure continuity:

$$h_i^{\text{unwrap}} = h_{i-1}^{\text{unwrap}} + \Delta h_i, \qquad \Delta h_i \in (-180^\circ, 180^\circ]. \tag{13}$$

This unwrapping ensures that hue deviations reflect actual shifts rather than artifacts from angular periodicity.

2) Endpoint Exclusion: Since the first and last steps of a design-system ramp are typically achromatic anchors (pure white and pure black), their hue angles are numerically undefined or arbitrary ($C^* \approx 0$). Including them would inject noise unrelated to the palette's chromatic identity. Accordingly, the hue stability metric is evaluated only over the *inner* steps $\{c_1, \ldots, c_{N-2}\}$, yielding $N' = N - 2$ observations.

3) Reference Color and Hue Deviation: The reference hue h_{base} is chosen as the hue of the **color with maximum chroma** in the palette, as this typically represents the dominant hue identity.

The angular distance between each step and the reference color is defined as

$$d_i = \min(|h_i - h_{\text{base}}|,\ 360^\circ - |h_i - h_{\text{base}}|)\,, \tag{14}$$

where $d_i \in [0, 180^\circ]$ represents the hue deviation at step i.

4) Normalization Envelope: To normalize deviation magnitudes into a $[0,1]$ score, a fixed reference envelope is constructed representing systematic linear drift from zero deviation at one end of the ramp to the maximum distinguishable angular distance (180°) at the other:

$$d_i^{\max} = 180^\circ \cdot \frac{i}{N'-1}. \tag{15}$$

This envelope serves as a **normalization scale** rather than a strict theoretical maximum: its purpose is to provide a consistent denominator that is independent of palette content, ensuring that scores remain comparable across palettes with different step counts and base-color positions. Because the envelope is monotonically increasing, palettes whose base color lies near the ramp center will have small actual deviations d_i relative to $d_i^{\max}$ on both sides of the base, yielding high stability scores as intended.

5) Hue Stability Metric Definition: The root mean square error of hue deviation is computed over the N' inner steps:

$$RMSE = \sqrt{\frac{1}{N'}\sum_{i=0}^{N'-1} d_i^2}. \tag{16}$$

The corresponding worst-case error is

$$RMSE_{\max} = \sqrt{\frac{1}{N'}\sum_{i=0}^{N'-1} (d_i^{\max})^2}. \tag{17}$$

The **Hue Stability** metric is defined by

$$\mathcal{H} = \max\left(0,\ 1 - \frac{RMSE}{RMSE_{\max}}\right). \tag{18}$$

6) Interpretation:

- $\mathcal{H} \approx 1$: Hue remains stable around the reference color throughout the palette.
- $\mathcal{H} \to 0$: Strong hue drift approaching the worst-case linear scenario.

This metric is independent of lightness and chroma, measuring only the **geometric stability of hue** in color space.

E. Spacing Uniformity ($\mathcal{U}$)

Spacing between adjacent color steps is measured using the **CIEDE2000 color-difference metric** [5]. For a sequential color scale of N ordered steps $\{c_0, c_1, \ldots, c_{N-1}\}$, differences are computed only between consecutive entries:

$$\delta_i = \Delta E_{00}(c_{i-1}, c_i), \qquad i = 1, \ldots, N-1. \tag{19}$$

An ideally spaced color scale exhibits approximately constant perceptual increments, i.e., $\delta_i \approx \delta_j$ for all i, j. To quantify relative dispersion of these increments independently of their absolute magnitude, spacing uniformity is evaluated using the **coefficient of variation (CV)**:

$$CV = \frac{\sigma(\{\delta_i\})}{\mu(\{\delta_i\})}, \tag{20}$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ denote the mean and standard deviation of the set $\{\delta_i\}$, respectively. As a dimensionless quantity, CV provides a scale-invariant measure of non-uniformity.

Since CV is unbounded above and inversely related to quality, it is mapped to a bounded score with intuitive directionality using the following monotonic transform:

$$\mathcal{U} = \frac{1}{1 + CV}. \tag{21}$$

This particular transform is chosen over alternatives (e.g., e^{-CV} or $1 - \tanh(CV)$) for three reasons: (i) it maps $[0, \infty) \to (0, 1]$ with a closed-form inverse ($CV = 1/\mathcal{U} - 1$), making the score directly interpretable; (ii) it is parameter-free, avoiding arbitrary tuning constants; and (iii) its gradient $-1/(1+CV)^2$ provides moderate sensitivity across the empirically observed range $CV \in [0.3, 1.5]$ without saturating prematurely.

This definition ensures $\mathcal{U} \in (0, 1]$, with $\mathcal{U} = 1$ corresponding to perfectly uniform spacing ($CV = 0$), and $\mathcal{U}$ decreasing monotonically as dispersion increases.

A palette with uniform sampling along the color ramp therefore achieves $\mathcal{U}$ values close to unity.

V. Composite Quality Score

To aggregate multiple quality metrics into a single scalar score without allowing strong dimensions to compensate for weak ones, the **geometric mean** is used instead of the arithmetic mean. The composite score is defined as

$$\text{SCORE} = 100 \cdot \left(\prod_{k=1}^{5} (M_k + \varepsilon) \right)^{1/5}, \tag{22}$$

where

$$M_k \in \{\eta, \mathcal{L}, \mathcal{S}_C, \mathcal{H}, \mathcal{U}\}$$

denote the five normalized component metrics, each bounded in $[0, 1]$, and $\varepsilon = 10^{-6}$ is a small constant introduced solely to ensure numerical stability when any component metric approaches zero.

The use of the geometric mean has several desirable properties:

- **Strong penalty for imbalance**: a single poor metric significantly reduces the overall score, preventing compensation by unrelated dimensions.
- **Encouragement of uniform quality** across accessibility, perceptual uniformity, and structural consistency criteria.
- **Scale invariance** with respect to the component metrics, preserving their relative contributions.

Since all component metrics are bounded in $[0, 1]$, the resulting SCORE is guaranteed to lie in the interval $[0, 100]$, enabling direct and intuitive comparison across different palette generation methods.

a) Equal weighting.: The five metrics are weighted equally in the current formulation. This choice reflects two considerations: (i) each metric captures a structurally independent dimension of palette quality—contrast, lightness, chroma, hue, and spacing—with no a priori reason to privilege one over another in a general-purpose evaluation; and (ii) the geometric mean already imposes a strong implicit penalty on weak dimensions, making explicit weighting less critical than in an arithmetic aggregation. Domain-specific applications (e.g., accessibility-first tools, data visualization) may benefit from weighted variants; however, determining principled weights would require empirical user studies that are beyond the scope of this work.

Accordingly, SCORE reflects overall palette completeness rather than excellence in any single isolated attribute.

VI. Experimental Evaluation

This section reports quantitative evaluation results for **11 widely used design systems**, serving as a benchmark for the proposed metric framework when applied to real-world monochromatic color ramps.

An interactive online report containing all benchmark data and visualizations is available at:

> https://chromametry.github.io/chromametry/benchmark

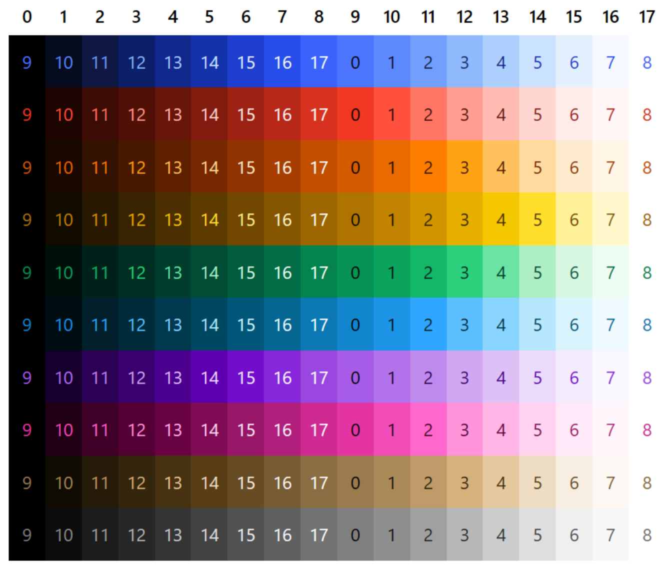

Fig. 1. Adobe Spectrum Color Palette. Cell value = Background Index ±9 steps (WCAG 4.5:1).

A. Quantitative Analysis Results

Table V summarizes the measured quality metrics across all evaluated design systems. Clear differentiation is observed across both individual metric dimensions and composite scores, indicating substantial variation in how existing design systems balance accessibility, uniformity, and structural consistency.

Several trends emerge from these results. Design systems that tightly control contrast span and lightness progression (e.g., IBM Carbon and Adobe Spectrum) achieve consistently high scores across most dimensions. In contrast, systems with large span values or uneven spacing exhibit reduced contrast efficiency and spacing uniformity, which significantly impacts the composite score due to the use of geometric aggregation.

> **Note.** Design systems such as Bootstrap, Google Material 3, Apple Human Interface Guidelines, and Fluent UI are excluded from this evaluation, as they primarily define discrete semantic color tokens rather than algorithmically constructed sequential color ramps.

B. Example: A Typical Report

The contrast span value K is taken directly from measured *Span*. Observed density D is determined by

$$D = \frac{K}{N-1}, \tag{23}$$

where N is the number of palette steps.

This table shows that design systems achieving high efficiency all have low density, while palettes with excessively wide span lead to wasted lightness space and are strongly penalized by metric η.

Thus, contrast span K is observed consistently across popular industry palettes, though previously not systematically

TABLE V
BENCHMARK RESULTS OF POPULAR DESIGN SYSTEMS EVALUATED USING THE PROPOSED METRIC FRAMEWORK. CONTRAST EFFICIENCY (η) IS COMPUTED AS THE ROOT-MEAN-SQUARE (RMS) ACROSS RAMPS.

Design System	Ramps	Steps	K	η (RMS)	$\mathcal{L}$	$\mathcal{S}_C$	$\mathcal{H}$	$\mathcal{U}$	SCORE
Adobe Spectrum	10	18	9	0.943	0.9333	0.8786	0.9138	0.7722	88.59
IBM Carbon	12	12	6	0.911	0.9303	0.8688	0.9288	0.7919	88.46
U.S. Web Design System	25	12	6	0.911	0.9359	0.8096	0.9380	0.7997	87.67
Salesforce Lightning 2	13	14	7	0.925	0.9187	0.8464	0.9372	0.7107	86.31
GitHub Primer Brand	13	12	6	0.911	0.9243	0.8405	0.9408	0.6841	85.45
Atlassian	9	14	8	0.771	0.8964	0.9094	0.9465	0.7129	84.23
Tailwind CSS	18	13	8	0.756	0.8705	0.8565	0.9147	0.6780	81.04
Ant Design	12	12	9	0.665	0.8586	0.8734	0.9276	0.6550	78.76
Material UI	19	12	11	0.507	0.7967	0.7861	0.9239	0.5500	69.43
Radix UI	16	13	10	0.474	0.7979	0.7679	0.9468	0.5207	67.80
Shopify Polaris	12	17	15	0.282	0.7281	0.6892	0.9223	0.4667	57.16

Base Color	Span (3.0)	Span (4.5)	Span (7.0)	Contrast Efficiency	Lightness Linearity	Chroma Smoothness	Hue Stability	Spacing Uniformity
	7	**9**	11	0.9431	0.8876	0.8669	0.8221	0.7902
	7	**9**	11	0.9431	0.9175	0.8210	0.9428	0.7757
	7	**9**	11	0.9431	0.9336	0.8971	0.8658	0.7854
	7	**9**	11	0.9431	0.9605	0.9317	0.8871	0.7343
	7	**9**	11	0.9431	0.9554	0.8617	0.9005	0.8153
	7	**9**	11	0.9431	0.9518	0.8394	0.7848	0.8481
	7	**9**	11	0.9431	0.8693	0.8888	0.9868	0.6695
	7	**9**	11	0.9431	0.8979	0.8866	0.9544	0.7232
	7	**9**	11	0.9431	0.9741	0.7736	0.9683	0.8215
	7	**9**	11	0.9431	0.9781	1.0000	1.0000	0.7425

Fig. 2. Adobe Spectrum Palette Metrics. Each radar axis corresponds to one of the five normalized metrics (η, $\mathcal{L}$, $\mathcal{S}_C$, $\mathcal{H}$, $\mathcal{U}$).

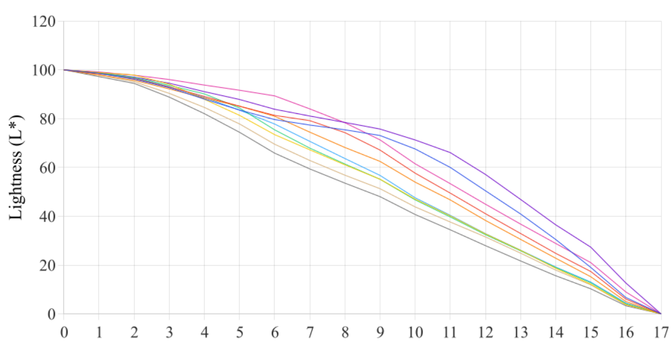

Fig. 3. Adobe Spectrum: Helmholtz–Kohlrausch corrected lightness (L_{EAL}) vs. step index, with least-squares linear fit overlay.

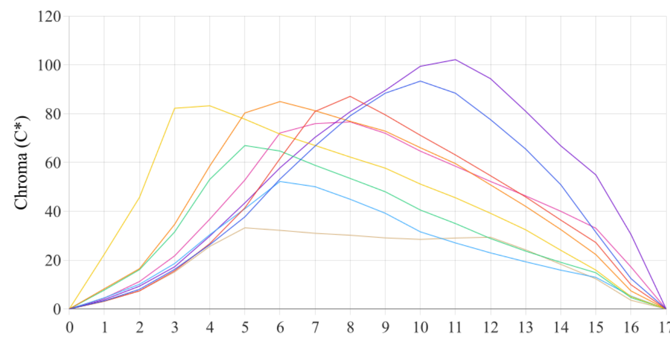

Fig. 4. Adobe Spectrum: Normalized chroma trajectory vs. monotone cubic spline reference.

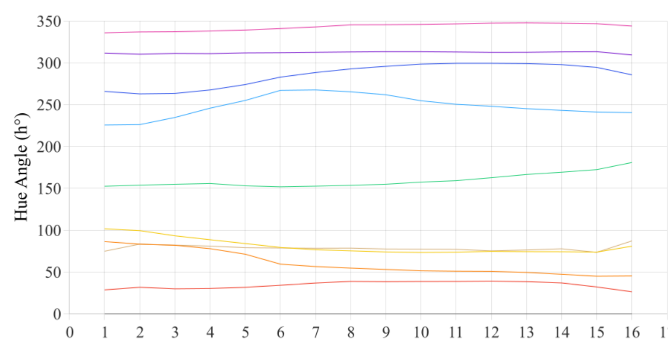

Fig. 5. Adobe Spectrum: Unwrapped hue angle across inner ramp steps, showing deviation from the base hue.

identified or exploited. With these results, users can select color pairs for background and text with distance K without requiring runtime verification.

C. Discussion

1) High Score Group (Score > 85): IBM Carbon, Adobe Spectrum, and USWDS maintain consistently high indicators across all aspects.

- **Observation:** High Contrast Efficiency ($\eta > 0.9$) and Lightness Linearity ($\mathcal{L} > 0.93$).
- **Analysis:** Span distance K maintains optimal ratio to total steps N, ensuring high density of WCAG-compliant color pairs.

2) Medium Score Group (75–85): Tailwind CSS and Ant Design show disparities among component indicators.

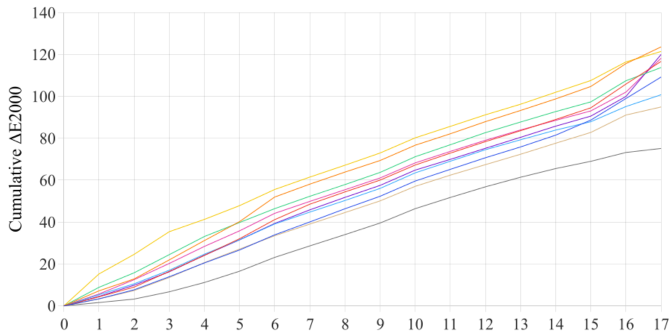

Fig. 6. Adobe Spectrum: Cumulative ΔE_{00} curve. A linear trajectory indicates uniform perceptual spacing.

Fig. 7. Composite score ranking of eleven evaluated design systems. Higher scores indicate stronger overall balance across all five metrics.

- **Observation:** High Chroma Smoothness ($\mathcal{S}_C > 0.84$) but lower Contrast Efficiency.
- **Analysis:** Large span relative to N reduces simultaneous usability of color pairs.

3) Low Score Group (< 75): Material UI (v4) and Shopify Polaris record the lowest scores.

- **Observation:** Low Spacing Uniformity ($\mathcal{U} < 0.6$) and Lightness Linearity.
- **Analysis:** Non-uniform step spacing and non-linear lightness progression.

D. Experimental Conclusions

Benchmark data confirms the framework's ability to classify palettes based on physical and mathematical characteristics. Results suggest that using an even number of steps with contrast span $\approx (N-1)/2$ yields favorable usability properties, where every color can serve as background and a corresponding text color always exists satisfying WCAG 4.5:1.

VII. Conclusion

This work introduces a unified quantitative framework for evaluating the quality of sequential monochromatic color palettes, grounded in the CIELAB color space and motivated by both perceptual uniformity and accessibility requirements. Five complementary metrics are proposed, covering contrast efficiency under WCAG 4.5:1 constraints, lightness linearity, chroma smoothness, hue stability, and perceptual spacing uniformity. Each metric captures a distinct structural or perceptual property and is normalized to enable consistent aggregation.

Through large-scale benchmarking of widely used industry design systems, the framework reveals systematic differences in how palettes allocate lightness range, distribute perceptual steps, and preserve hue identity. In particular, the analysis identifies contrast span K as a previously under-formalized but critical parameter, with empirical results indicating that spans close to $(N-1)/2$ maximize accessibility efficiency while preserving usable color density. These findings demonstrate that many high-quality palettes converge toward similar structural ratios, despite differing design origins.

The proposed composite score, based on geometric aggregation, enables objective comparison without allowing compensation between weak and strong dimensions, thereby reflecting overall palette completeness rather than isolated excellence. As a result, the framework supports reproducible benchmarking, automated palette validation, and optimization-driven color system design.

A. Limitations

Several limitations should be acknowledged. First, while CIELAB is widely adopted as a standard perceptual color space, it exhibits known non-uniformities—particularly in the blue region [5]—that may affect chroma and hue measurements for certain hue families. More recent color appearance models such as CAM16 or perceptually uniform spaces like Oklab [7] may offer improved uniformity, though at the cost of reduced compatibility with the CIEDE2000 metric. Second, the framework evaluates only *sequential monochromatic* ramps and does not address diverging, categorical, or multi-hue palettes, which require fundamentally different structural assumptions. Third, the five proposed metrics carry equal weight in the geometric mean composite score; domain-specific applications may benefit from weighted aggregation, though determining appropriate weights would require empirical user studies beyond the scope of this work. Finally, the benchmark evaluation is limited to design systems that publish algorithmically constructed sequential ramps; systems relying on hand-picked semantic tokens were excluded, which may limit generalizability.

B. Future Work

Future work will extend the framework to diverging and multivariate palettes, and investigate how analogous metric formulations may be adapted to alternative perceptually uniform color spaces (e.g., CAM02-UCS or CAM16) while preserving cross-space comparability. Empirical validation through user studies comparing metric rankings with human preference judgments would further strengthen the framework's applicability.

References

[1] ISO/CIE. *Colorimetry — Part 4: CIE 1976 L*a*b* Colour Space*. ISO/CIE 11664-4:2019, 2019.

[2] F. N. Fritsch and R. E. Carlson. Monotone piecewise cubic interpolation. *SIAM Journal on Numerical Analysis*, 17(2):238–246, 1980.

[3] R. High, P. Green, and P. Nussbaum. The Helmholtz–Kohlrausch effect on display-based light colors and simulated substrate colors. *Color Research & Application*, 48(3):341–354, 2023.

[4] A. Zeileis, K. Hornik, and P. Murrell. Escaping RGBland: Selecting colors for statistical graphics. *Computational Statistics & Data Analysis*, 53(9):3259–3270, 2009.

[5] M. R. Luo, G. Cui, and B. Rigg. The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Research & Application*, 26(5):340–350, 2001.

[6] W3C. *Web Content Accessibility Guidelines (WCAG) 2.1*. 2018. Available at https://www.w3.org/TR/WCAG21/.

[7] M. Safdar, G. Cui, Y. J. Kim, and M. R. Luo. Perceptually uniform color space for image signals including high dynamic range and wide gamut. *Optics Express*, 25(13):15131–15151, 2017.

[8] E. Mach. Über die Wirkung der räumlichen Verteilung des Lichtreizes auf die Netzhaut. *Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften, Wien*, 52:303–322, 1865.

[9] J. F. Epperson. On the Runge example. *The American Mathematical Monthly*, 94(4):329–341, 1987.

[10] IEC. *Multimedia systems and equipment — Colour measurement and management — Part 2-1: Colour management — Default RGB colour space — sRGB*. IEC 61966-2-1:1999, 1999.

[11] C. C. Gramazio, D. H. Laidlaw, and K. B. Schloss. Colorgorical: Creating discriminable and preferable color palettes for information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):521–530, 2017.

[12] F. Yang, W. Zhang, and X. Chen. Computational color theme scoring via user preference modelling. *ACM Transactions on Applied Perception*, 21(2), 2024.

[13] J. Norman, A. Klokmose, and B. Victor. Purpose, palette, prototype: Understanding professional color design workflows. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, 2024.

[14] S. Lin and J. Fortuna and C. Kulkarni and M. Stone and J. Heer. Modeling Color Difference for Visualization Design. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, 2019.

[15] M. Stone, D. Albers Szafir, and V. Setlur. An engineering model for color difference as a function of size. In *Proceedings of the 22nd Color and Imaging Conference*, pages 253–258, 2014.