Extensible Multi-Primary Control Sequences

Paul Centore* Michael H. Brill†

Abstract

In a display with more than three primaries (called a multi-primary display), a color can be expressed as multiple combinations (called control sequences) of primaries. This paper presents an algorithm for assigning control sequences, that preserves current assignments when further primaries are added. We call these control sequences extensible. It is shown that the gamut of any number of primaries is a zonohedron, which can be dissected into parallelepipeds. Control sequences are assigned within each parallelepiped. The current parallelepipeds remain when more primaries are added, so the current assignments are preserved. Multi-primary displays can also cause unwanted metamerism, and make continuous color scales appear discontinuous. The algorithm avoids these problems. When viewed through natural filters, such as yellowed ocular lenses, multi-primary displays can sometimes make two different colors appear identical. If the primaries satisfy the Binet-Cauchy criterion, which is always the case when all primaries are monochromatic, then these spurious matches are avoided.

Keywords — Multi-primary display, metamerism, color, extensible, zonohedron, Binet-Cauchy

1 Introduction

A pixel of a computer monitor, television, or other display device, can usually produce three primaries: red, green, and blue. A primary is a colored illuminant whose

*Paul Centore (corresponding author) lives at 62 Chicago Ave., Groton, CT 06340; e-mail: centore@99main.com

†Michael H. Brill is with Datacolor, 5 Princess Road, Lawrenceville, NJ 08648; e-mail: mbrill@datacolor.com

intensity varies, but whose relative spectral power density does not. Combining the primaries at differing intensities produces a wide gamut of colors. Recently, multi-primary displays, involving more than three primaries, have been used to produce an even wider gamut.

Since the human retina has three kinds of cones, human color space is inherently three-dimensional. Algebraically, the set of linear combinations of three independent primaries should span human color space. Physically, however, the coefficients in the linear combination must be between 0 (no activation) and 1 (full activation). Because of this restriction, any set of three primaries only spans a limited gamut. Adding more primaries expands the device's color gamut, but introduces *metamerism*: a color in the gamut can result from many different combinations of the primaries, instead of from a unique combination.

In 1931, the Commission Internationale de l'Éclairage (CIE) defined a standard observer,[1] who converts a color stimulus to three perceptual coordinates: X, Y, and Z. Two color stimuli, such as two combinations of primaries, are perceptually identical, to the standard observer, if and only if their XYZ coordinates are equal. If viewers' visual systems conformed perfectly to the standard observer, then metamerism would not cause any difficulties—but visual systems usually do not quite conform. As a result, color matches may be spuriously made or broken, for a particular viewer.

When there are three primaries, each color results from a unique combination of primaries, the specification of which is called a control sequence. When extending the display gamut by adding more primaries, it is desirable to assign control sequences to the entire gamut, such that they are consistent with the control sequences of the original gamut. This consistency simplifies the process of converting between the two gamuts. The algorithm presented in this paper possesses such extensibility, and is believed to be the only assignment algorithm that does so.

Besides extensibility, an assignment algorithm should also avoid, or at least mitigate, three other problems that multi-primary displays can cause for viewers: non-matches, non-smoothness, and spurious matches. In a non-match, two metameric colors, produced by two different combinations of primaries, have equal XYZ coordinates. Though the colors appear the same to the standard observer, they differ slightly to many people. Non-smoothness occurs when a viewer perceives a color scale as changing abruptly, even though it is a continuous curve through the gamut, because of the combinations of primaries used to produce it. Spurious matches occur when ambient lighting conditions, or filters, cause two colors, with different XYZ coordinates, to appear identical.

The central insight of this paper is the zonohedral structure of the color gamut. A zonohedron is the Minkowski sum of a set of vectors in $\mathbf{R}^3$. Each primary in a display is a vector in XYZ space. The gamut is the Minkowski sum of the primaries, in three-dimensional XYZ space, and is thus a zonohedron. The paper gives a natural dissection of the zonohedron into disjoint parallelepipeds, each of whose sides is a translated primary vector. A practical contribution of the paper is a simple, easily implemented dissection algorithm. The dissection requires an ordering of the primaries, and different orderings can result in different dissections. The gamut should be dissected at the start of a multi-primary design, and the dissection should be used consistently throughout the design process.

For each parallelepiped in the dissection, assign an originating vertex. The three primary vectors originating from that vertex are a basis, in the vector space sense: each point in the parallelepiped is a unique linear combination of those primary vectors. The dissection algorithm insures that the originating vertex itself is a sum of a subset of primary vectors that does not include any of the three primary vectors making up the edges. A point in the parallelepiped is therefore uniquely expressed as a sum of the subset of primaries that add up to the originating vertex, and a unique linear combination of the primary vectors making up the parallelepiped's edges. This sum defines a unique control sequence.

If more primaries are added, the gamut is expanded. This paper's method of expanding the gamut insures that the current parallelepipeds are also in the new dissection. As a result, the points in the current gamut retain their current control sequences. The assignment algorithm is therefore extensible: new colors can be added without changing the representations of the current colors.

Apart from extensibility, the dissection algorithm has other desirable features. Displaying any set of XYZ coordinates with a unique combination of primaries avoids non-matches. Since every instance of an XYZ color is displayed with the same combination of primaries, every instance will look identical to any viewer, no matter how much he differs from the standard observer. In addition, it will be shown that this parallelepiped dissection avoids discontinuities. Spurious matches can be avoided if the primary spectra obey a criterion discussed in Section **??** below. Since a set of monochromatic primaries (in the appropriate order) always satisfies this criterion, spurious matches will not occur when all primaries are monochromatic; this paper will focus on the monochromatic, or nearly monochromatic, case.

2 Color Gamuts

2.1 Geometry of Display Gamuts

A display device produces a wide gamut of colors by "mixing" a limited set of primaries. A primary is an illuminant of fixed relative spectral density, whose intensity can vary, from 0 to some maximum, which can be denoted 1. The CIE's XYZ coordinates can be used to standardize human perception of both primaries, and combinations of primaries. Geometrically, X, Y, and Z, are non-negative real numbers, that can be plotted in the positive octant of $\mathbf{R}^3$. Points in this octant can be equally well viewed as vectors from the origin; the two viewpoints will be used interchangeably. A primary at full intensity is a point in the octant. A combination of primaries, at varying intensities, is also a point in the octant. The set of all possible combinations of primaries forms a solid in the octant, which is called the display's gamut. This section elucidates the display gamut's zonohedral structure, which later sections will use to assign control sequences extensibly.

Denote the number of available primaries by N. Display devices usually have three primaries, that appear red, green, and blue. Devices with four or more primaries are called multi-primary devices. Denote each primary by p_i, where $i = 1, \ldots, N$. A primary at maximum intensity has a spectral density, $p_i(\lambda)$, which is a non-negative function on the visible spectrum, from 400 nm to 700 nm. The CIE color-matching functions[1] can be used to transform p_i into a vector $\mathbf{v_i} = (v_{iX}, v_{iY}, v_{iZ})$, in XYZ coordinates:

$$v_{iX} = \int_{400}^{700} \bar{x}(\lambda)p_i(\lambda)d\lambda, \tag{1}$$

$$v_{iY} = \int_{400}^{700} \bar{y}(\lambda)p_i(\lambda)d\lambda, \tag{2}$$

$$v_{iZ} = \int_{400}^{700} \bar{z}(\lambda)p_i(\lambda)d\lambda. \tag{3}$$

The display gamut, G, consists of all possible combinations of primaries, at all intensities from 0 to 1:

$$G = \left\{ \sum_{i=1}^{N} \alpha_i \mathbf{v_i} \middle| \alpha_i \in [0,1]\,\forall i \right\}. \tag{4}$$

A sequence of N coefficients $\alpha_i, i = 1 \ldots N$, all of which are between 0 and 1, will be called a control sequence. The set of all control sequences is written formally as

$$S = \{(\alpha_1, \alpha_2, \ldots, \alpha_N)|\alpha_i \in [0,1]\forall i = 1, \ldots, N\}. \tag{5}$$

In practical terms, a control sequence gives instructions to display some combination of primaries, resulting in a particular color. Equations (1) through (3) are linear in $p_i(\lambda)$. Therefore, the XYZ coordinate vector of a combination of primaries, such as appear in Equation (4), is the sum, in the vector space sense, of the XYZ coordinates of the individual primaries, multiplied by the appropriate α_i. We can define a linear transformation, T, from control sequences to XYZ space:

$$T\left(\alpha_1, \alpha_2, \ldots, \alpha_N\right) = \left(\sum_{i=1}^{N} \alpha_i v_{iX}, \sum_{i=1}^{N} \alpha_i v_{iY}, \sum_{i=1}^{N} \alpha_i v_{iZ}\right). \tag{6}$$

Geometrically, the primaries are seen as the vectors $\mathbf{v_1}$ through $\mathbf{v_N}$ in the positive octant. Equation (4) shows that the gamut G is the zonohedron generated by the N primary vectors. A zonohedron is the Minkowski sum of a set of vectors. The Minkowski sum (also called the vector sum) of two sets, $\mathbf{A}$ and $\mathbf{B}$, in $\mathbf{R}^n$, is defined as

$$\mathbf{A} \oplus \mathbf{B} = \{a + b | a \in \mathbf{A}, b \in \mathbf{B}\}. \tag{7}$$

More concretely, the Minkowski sum of $\mathbf{A}$ and $\mathbf{B}$ is the volume (or area) that $\mathbf{B}$ sweeps out when its tail can be located at any point in $\mathbf{A}$. The Minkowski sum is both commutative and associative.

The Minkowski sum of two vectors is the parallelogram swept out by letting either vector slide along the other. The Minkowski sum of a parallelogram and another vector, not in the same plane, is the parallelepiped swept out by sliding the parallelogram along that vector. Figure 1 shows the Minkoswki sum of a parallelepiped and another vector in $\mathbf{R}^3$. The parallelepiped generated by $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$ appears on the left. On the right is the zonohedron of $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}, \mathbf{v_4}\}$. To make the figure on the right, place a copy of $\mathbf{v_4}$ at each vertex of the figure on the left, and then take the convex hull; alternately, slide the parallelepiped on the left along the vector $\mathbf{v_4}$.

Zonohedra are convex and two-fold rotationally symmetric. They begin at the origin, which corresponds to black, because no primary illuminants are being emitted. Their vertices take the form

$$\sum_{i=1}^{N} \epsilon_i \mathbf{v_i} \Big| \epsilon_i = 0 \text{ or } 1, \tag{8}$$

though not every sum of that form is necessarily a vertex.[2] By the construction in Figure 1, every edge of a zonohedron is a translation of a generating vector; conversely, the construction shows that every generating vector occurs as an edge.

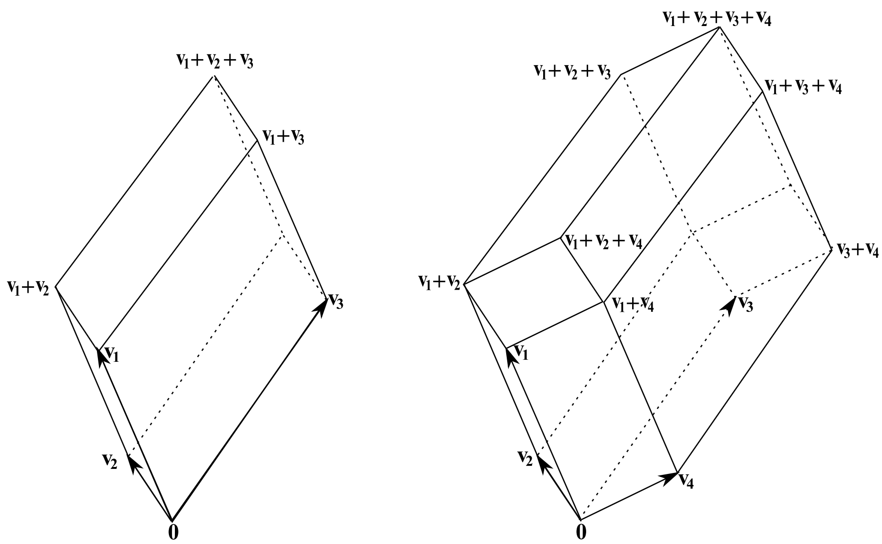

Figure 1: Construction of Zonohedron

A zonohedron's faces are all parallelograms, provided that no three generating vectors are coplanar. This condition is technical and has little practical import, since coplanarity is an unstable condition: adjusting any vector by the slightest amount can destroy coplanarity. In the context of displays, there is always some slight measurement uncertainty about the primaries' tristimulus vectors, so an arbitrarily small adjustment can be made if needed, to ensure this condition. Though they are not labeled as such, Fig. 2 of a paper of Ajito[3] and Fig. 1 of a paper of Kanazawa[4] show zonohedral gamuts, for 4 and 6 primaries, respectively.

More details and examples of Minkowski sums and zonohedra appear in a recent paper.[5] In that paper, the generating vectors were tristimulus vectors for monochromatic spectral densities, and the zonohedron was the object-color solid. The generating vectors in that construction were in cyclic position: their chromaticities traced out a convex curve on the chromaticity diagram. The cyclic property allows an easy algorithm (Algorithm I) for the zonohedron:

1. From Equation (8), the vertices of the N-primary gamut must have all-or-none activation of each primary, yielding 2^N vertex candidates to be culled.

2. If the chromaticities of the primaries form a convex polygon in the chromaticity diagram, continue. Otherwise (as with $RGBW$ displays) turn off any of the primaries (e.g., W) that compromise the convexity.

3. Label the N primaries 1 through N clockwise around the convex polygon de-

fined above. The result for four primaries would look like Fig. 7 in Ref. 5 (but without 0 because black isn't a point in chromaticity space).

4. Now enumerate vertices according to Equation (7) in Ref. 5 (using the example of 4 primaries):

   (a) The level-0 vertex is black: $(0, 0, 0)$.

   (b) The level-1 vertices are the primaries: $\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}, \mathbf{v_4}$.

   (c) The level-2 vertices are adjacent sums of pairs of primaries: $\mathbf{v_1} + \mathbf{v_2}, \mathbf{v_2} + \mathbf{v_3}, \mathbf{v_3} + \mathbf{v_4}, \mathbf{v_4} + \mathbf{v_1}$. Note that adjacency is mod(4).

   (d) The level-3 vertices are adjacent sums of single vectors with the level-2 pairs: $\mathbf{v_1} + \mathbf{v_2} + \mathbf{v_3}, \mathbf{v_2} + \mathbf{v_3} + \mathbf{v_4}, \mathbf{v_3} + \mathbf{v_4} + \mathbf{v_1}, \mathbf{v_4} + \mathbf{v_1} + \mathbf{v_2}$.

   (e) The level-4 vertex is the sum of all primaries: $\mathbf{v_1} + \mathbf{v_2} + \mathbf{v_3} + \mathbf{v_4}$. This is the brightest color, usually white, and it is unique.

This completes the enumeration of the gamut's vertices. There are $2 + N(N-1)$ of them—considerably fewer than the 2^N candidates.

Expression (8) from Ref. 5 shows graphically how to construct edges and parallelograms, producing Figure 2. The entries in Figure 2 are the zonohedron's vertices. Two entries are joined in Figure 2 if and only if the corresponding vertices are joined by an edge. The parallelograms making up the zonohedron's faces are evident. Some vertices and edges appear twice in Figure 2, though they only appear once in the gamut.

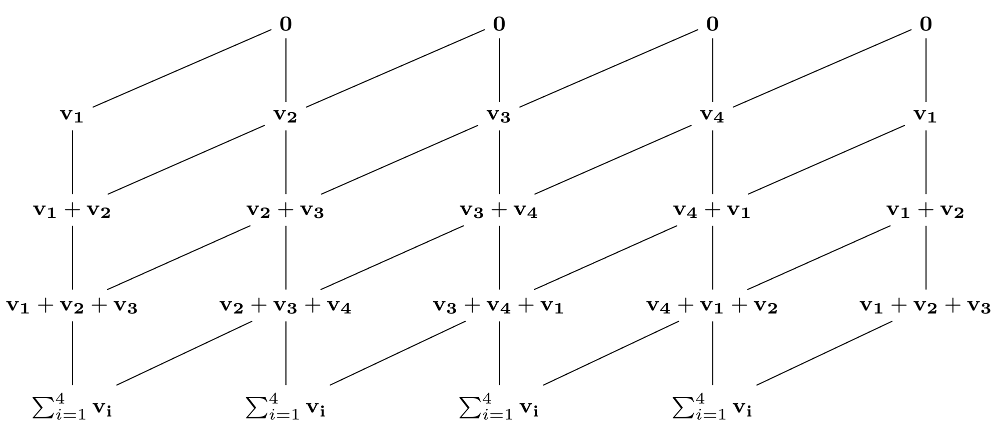

Figure 2: Construction of Edges on a Zonohedron

The cyclic property holds in a multi-primary display when no primary can be written as a positive linear combination of any other primaries. From a practical point of view, this condition means that every primary adds a new chromaticity: no primary appears as a convex combination of the other primaries, in the chromaticity diagram. This condition is often satisfied, but not always. In an $RGBW$ system, for instance, the chromaticity of the white primary can be duplicated by combining the red, green, and blue primaries. If all primaries are monochromatic, on the other hand, then their chromaticities are on the boundary of the chromaticity diagram, so they are automatically cyclic. When the cyclic property holds, the algorithm above conveniently represents the gamut geometrically.

The gamut can still be found for an arbitrary, possibly non-cyclic set of primaries. As König *et al*[6] describe, the gamut is the convex hull of all 2^N points of the form given in Expression (8). One can calculate all such points, then find their convex hull, and then, either manually or with a computer algorithm, determine which of them are vertices, which pairs of vertices are joined by edges, and which sets of edges form faces.

2.2 Metamerism

Equation (4) shows that there can be some redundancy, or *metamerism*, in the display gamut. Metamerism in the context of display devices occurs when two control sequences produce the same XYZ tristimulus values. Figure 3 gives a simple example of metamerism. Suppose there are four primaries, $\mathbf{v_1}$ through $\mathbf{v_4}$, as shown in the figure. They form a square pyramid, whose apex is at the origin. Then one can see geometrically that

$$\mathbf{v_1} + \mathbf{v_3} = \mathbf{v_2} + \mathbf{v_4}, \tag{9}$$

so the two sides of Equation (9) are metameric representations of the same point in the gamut of $\mathbf{v_1}$ through $\mathbf{v_4}$. One could find many other metameric combinations for this point, such as $\frac{1}{3}\mathbf{v_1} + \frac{2}{3}\mathbf{v_2} + \frac{1}{3}\mathbf{v_3} + \frac{2}{3}\mathbf{v_4}$.

Metamerism cannot occur when there are only three (non-coplanar) primaries. XYZ space can be seen as a vector space, in which Equation (4) gives linear combinations of the primaries. When there are only three primaries, they form a basis for the vector space, so any point can be written as only one linear combination of the primaries. Equation (4) requires that the coefficients of the combination be between 0 and 1, so the gamut of three primaries is a parallelepiped.

A related concept is an *offset parallelepiped*. Such a parallelepiped is translated from the origin (which is the black point) by some vector $\mathbf{w}$, which extends from the

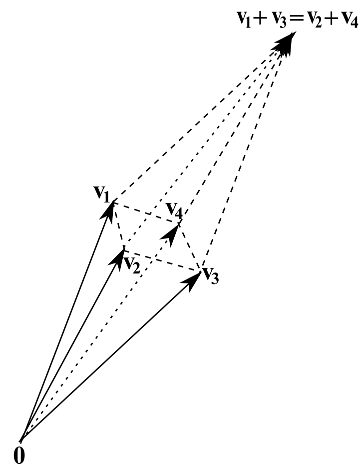

Figure 3: An Example of Metamerism

origin to one vertex of the parallelepiped. The three edges meeting at that vertex are given by three vectors, $\mathbf{u_1}, \mathbf{u_2}$, and $\mathbf{u_3}$. These three vectors can be seen as having been translated by $\mathbf{w}$ from the origin to the parallelepiped vertex. Every point within the parallelepiped has a unique representation as $\mathbf{w}$ plus a linear combination of $\mathbf{u_1}, \mathbf{u_2}$, and $\mathbf{u_3}$. Figure 1 shows that such offset parallelepipeds occur naturally when constructing the zonohedral gamut.

In the sequel, $\mathbf{u_1}, \mathbf{u_2}$, and $\mathbf{u_3}$ will be primary vectors, and $\mathbf{w}$ will be a combination of primary vectors, of the form given in Expression (8). Formally, define the parallelepiped $P_{w,abc}$ to be

$$P_{w,abc} = \left\{ \mathbf{w} + \sum_{i=a,b,c} \alpha_i \mathbf{v_i} \middle| \alpha_i \in [0,1] \, \forall i \right\}. \tag{10}$$

Geometrically, the eight vertices of $P_{w,abc}$ occur when each α_i is 0 or 1 in Equation (10). The parallelepiped $P_{w,abc}$ is the convex hull of those vertices:

$$\begin{aligned} P_{w,abc} \;=\; \mathrm{conv}(&\mathbf{w}, \\ &\mathbf{w} + \mathbf{v_a}, \mathbf{w} + \mathbf{v_b}, \mathbf{w} + \mathbf{v_c}, \\ &\mathbf{w} + \mathbf{v_a} + \mathbf{v_b}, \mathbf{w} + \mathbf{v_a} + \mathbf{v_c}, \mathbf{w} + \mathbf{v_b} + \mathbf{v_c}, \\ &\mathbf{w} + \mathbf{v_a} + \mathbf{v_b} + \mathbf{v_c}). \end{aligned} \tag{11}$$

As long as the coefficients of $\mathbf{v_a}, \mathbf{v_b}$, and $\mathbf{v_c}$, in the expression for $\mathbf{w}$, are all 0, there is no duplication of primaries, and $P_{w,abc}$ is contained in G.

We can define a metamerism function M, from XYZ space to the set of all control sequences that result in a particular set (X, Y, Z). Formally,

$$M(X, Y, Z) = \{s \in S | T(s) = (X, Y, Z)\}, \tag{12}$$

where T is given in Equation (6). M depends on the choice of primaries, and on their ordering. M decomposes the set S into equivalence classes of metamers. T maps each equivalence class to the same (X, Y, Z) values.

3 Zonohedral Control Sequences

The gamut's zonohedral structure suggests a natural dissection into parallelepipeds. Once this dissection is performed, a zonohedral control sequence can be defined for any point in the gamut, by using the parallelepiped containing this point. This section will present an algorithm for the dissection, and the next section will show that zonohedral control sequences have desirable properties. In particular, they are easily extended when more primaries are added.

3.1 Dissecting the Gamut into Parallelepipeds

The following algorithm (Algorithm II) dissects the zonohedral gamut into parallelepipeds :

1. Choose an ordering for the primaries. This ordering is arbitrary, but must be used consistently. Different orderings can lead to different dissections.

2. Construct a parallelepiped from the primary vectors $\{\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3}\}$. This parallelepiped is actually a zonohedron, which can be denoted Z_3. In general, let Z_j denote the zonohedron generated by $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_j}\}$. The parallelepiped Z_3 will be part of the dissection.

3. Loop over Z_j, for $j = 3, \ldots, (N - 1)$. This step will construct the zonohedron Z_{j+1}, from Z_j and the vector $\mathbf{v_{j+1}}$, as follows:

   (a) For each boundary parallelogram B of Z_j :

      i. By construction, two opposite edges of B are copies of some generator $\mathbf{v_a}$. The other two edges are copies of another generator, $\mathbf{v_b}$. The vertices of B are of the form $\mathbf{w}$, $\mathbf{w} + \mathbf{v_a}$, $\mathbf{w} + \mathbf{v_b}$, and $\mathbf{w} + \mathbf{v_a} + \mathbf{v_b}$. Since $\mathbf{w}$ is a vertex of Z_j, it follows from Expression (8) that $\mathbf{w}$ is a sum of a subset of the primaries, at full intensity, that generate Z_j. B has two sides; one is outside Z_j, and the other is inside.

ii. If the vector $\mathbf{v_{j+1}}$ is on the side of B that faces inside the zonohedron, do nothing. Otherwise (when $\mathbf{v_{j+1}}$ is facing outward), let $\mathbf{v_{j+1}}$ sweep out a parallelepiped over B. Add this new parallelepiped to the dissection. Because it was built on an outside face, the new parallelepiped does not intersect any of the current parallelepipeds. The new parallelepiped is $P_{w,abc}$, where $c = j + 1$. The sum of full primaries, $\mathbf{w}$, is the offset point for this parallelepiped.

(b) After looping through all the boundary parallelograms of Z_j, the union of all the previous and new parallelepipeds will be a new zonohedron, Z_{j+1}. Execute the loop again, using Z_{j+1} in place of Z_j.

4. The zonohedron Z_N is the entire color gamut, and the set of all the constructed parallelepipeds defines the dissection.

While the zonohedral construction in Algorithm I depended on cyclic ordering, Algorithm II does not. Even an *RGBW* gamut could be dissected by Algorithm II.

As an example, the dissection algorithm can be applied to the zonohedron in Figure 1. The left side of that figure shows the parallelepiped Z_3, constructed from the first three primaries. Figure 4 applies Algorithm II to build up the zonohedron on the right of Figure 1, parallelepiped by parallelepiped. In Step 2 of Figure 4, $\mathbf{v_4}$ is added, in the Minkowski sense, to the parallelogram originating at $\mathbf{0}$, and bounded by generators $\mathbf{v_1}$ and $\mathbf{v_3}$. Steps 3 and 4 show two other parallelepipeds, constructed similarly, that contribute to the dissection. Only parallelepipeds that are located on the outer sides of bounding parallelograms are included. For example, the parallelepiped $P_{0,124}$ would be built on the inner side of the parallelogram given by 0, $\mathbf{v_1}$, and $\mathbf{v_2}$, so it is not included. In all, four parallelepipeds go to make up Z_4, shown in an exploded view in Step 5. Since there are only 4 primaries, Z_4 is the entire gamut G, and we get

$$G = P_{0,123} \cup P_{0,134} \cup P_{1,234} \cup P_{3,124}. \tag{13}$$

This dissection algorithm could be easily extended to an arbirtrary number of primaries. Suppose, for example, that an engineer wanted to add a fifth primary. He would start with the dissection of Z_4 shown in Figure 4, and a fifth primary, $\mathbf{v_5}$. Six additional parallelepipeds can be created by sweeping $\mathbf{v_5}$ over different outward-facing surfaces of Z_4. Figure 5 shows one of the new parallelepipeds. It is built on the bounding parallelogram whose edges are copies of $\mathbf{v_1}$ and $\mathbf{v_2}$, originating at $\mathbf{v_3} + \mathbf{v_4}$, which is a vertex of Z_4. The new parallelepiped is $P_{3+4,125}$, where $3+4$ is shorthand for $\mathbf{v_3} + \mathbf{v_4}$. Other parallelepipeds are constructed similarly on other faces of Z_4. Any point that is already in Z_4 remains in its previous parallelepiped. The new parallelepipeds are just added to the list in Equation 13, extending it to the entire Z_5.

3.2 An Algebraic Interpretation

Apart from its geometric interpretation, the assignment of coefficients α_i has an elegant algebraic interpretation. The parallelepiped dissection gives more weight to earlier primaries than later primaries. The primary $\mathbf{v_i}$, for example, will not be used to define a color $C \in G$, if C is in Z_{i-1}, the zonohedron generated by the first $i-1$ primaries. Even if $\mathbf{v_i}$ is necessary, its coefficient will be as small as possible, because the zonohedron spanned by the first $i-1$ primaries is as large as possible, and there is a unique nearest point to C on Z_{N-1} along the line generated by $\mathbf{v_i}$. If $i = N$, then, the coefficient α_N is given by

$$\alpha_N = \min\left\{\alpha | C = \alpha\mathbf{v_N} + \sum_{i=1}^{N-1} \alpha_i\mathbf{v_i},\ \text{for some } \alpha, \alpha_1, \alpha_2, \ldots, \alpha_{N-1}\right\}. \tag{14}$$

Once α_N is fixed as a component of C, use the same minimizing procedure to define α_{N-1} :

$$\alpha_{N-1} = \min\left\{\alpha | C = \alpha_N\mathbf{v_N} + \alpha\mathbf{v_{N-1}} + \sum_{i=1}^{N-2} \alpha_i\mathbf{v_i},\ \text{for some } \alpha, \alpha_1, \alpha_2, \ldots, \alpha_{N-2}\right\}. \tag{15}$$

Continue this procedure, counting downward until 1, by which time all coefficients will have been found. A converse approach, which gives the same result, is to maximize the early primaries. Define the coefficients by

$$\alpha_1 = \max\left\{\alpha | C = \alpha\mathbf{v_1} + \sum_{i=2}^{N} \alpha_i\mathbf{v_i} \text{ for some } \alpha, \alpha_2, \alpha_3, \ldots, \alpha_N\right\}, \tag{16}$$

$$\alpha_2 = \max\left\{\alpha | C = \alpha_1\mathbf{v_1} + \alpha\mathbf{v_2} + \sum_{i=3}^{N} \alpha_i\mathbf{v_i} \text{ for some } \alpha, \alpha_3, \alpha_4, \ldots, \alpha_N\right\}, \tag{17}$$

and so on, until α_N.

These two other algorithms give identical results to Algorithm II, and are conceptually simpler and more elegant. They make no reference to zonohedra or geometric

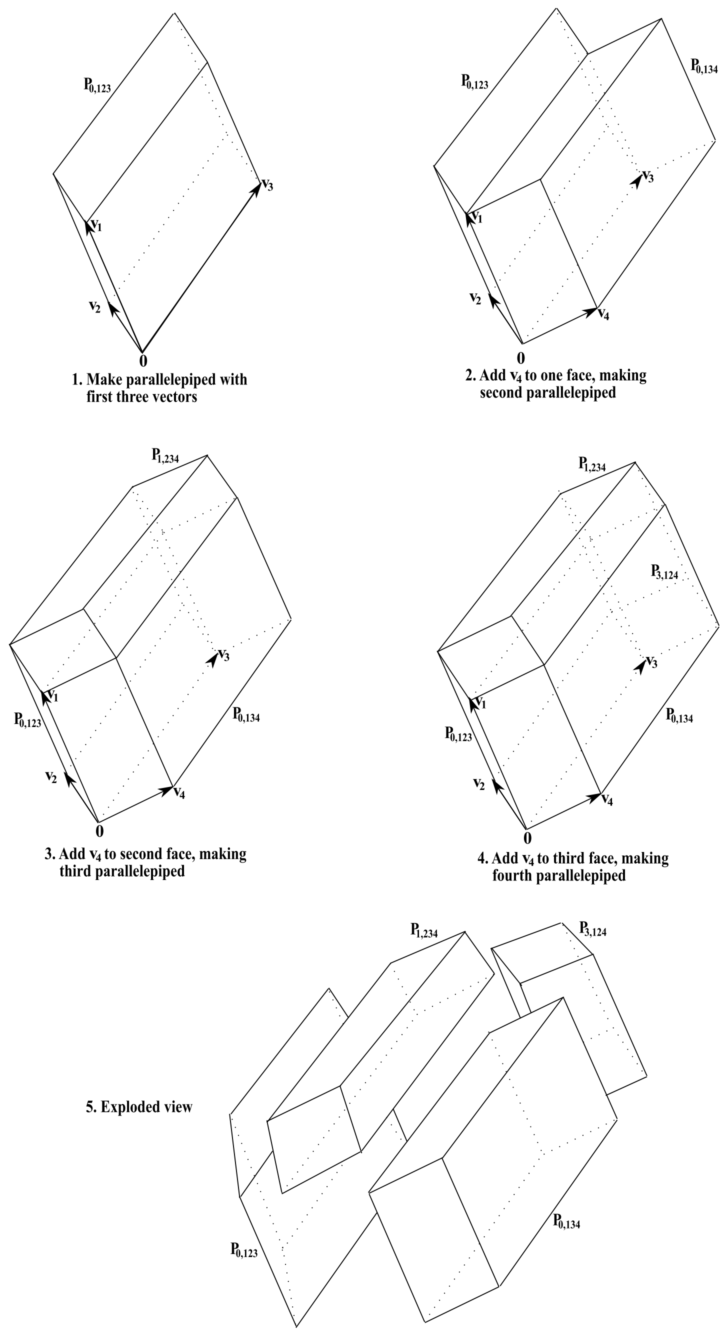

Figure 4: Dissection of Zonohedron into Parallelepipeds

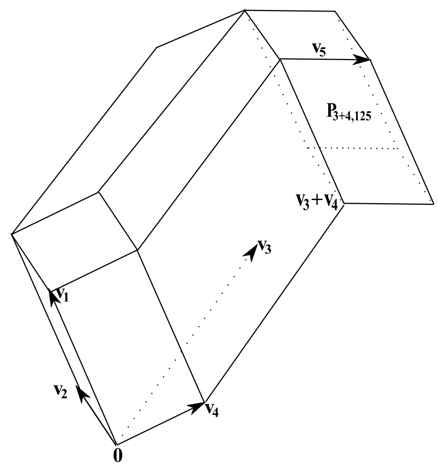

Figure 5: Extending the Dissection to a Fifth Primary Vector

constructions. It is clear from the defining equations that the resulting coefficients are unique, and that they depend on the order chosen from the primaries. As presented, however, it is not at all obvious how the coefficients can be computed efficiently. The parallelepiped dissection provides a geometric approach that is computationally undemanding.

4 Properties of Zonohedral Control Sequences

This section shows that zonohedral control sequences have several desirable properties. They avoid metamerism, are continuous, and, under some mild conditions, avoid spurious matches. Many other control sequence algorithms[3,4,6,7,8] share some or all of these properties, but extensibility is believed to be unique to the zonohedral algorithm. Extensibility in the context of displays means that further primaries can be added, without changing the control sequences of the existing gamut. Each property will be discussed in detail.

4.1 Extensibility

Extensibility in the display context means that additional primaries can be added, without having to recompute the control sequences for points currently in the gamut. For example, Figure 5 shows a fifth primary being added to the four-primary gamut in Figure 4. Using zonohedral control sequences, the current gamut would preserve its current sequences, but any new points would involve a non-zero coefficient α_5.

Because of this property, a manufacturer could offer a low-end, three-primary, display, that is compatible with a high-end, six-primary, display, by making the low-end primaries a subset of the high-end primaries. When using Algorithm II, the primaries would be ordered so the low-end ones come first. Gamut mapping would be simplified, so that a multi-primary image could more easily be displayed on either system.

4.2 Avoiding Non-Matches

A metameric non-match occurs, in the notation of Equation (12), when there are different control sequences, s and t, such that $T(s)$ and $T(t)$ have identical CIE coordinates, but a non-standard observer judges that they do not match. The solution to this problem is conceptually simple: assign a unique control sequence to each (X, Y, Z) triple. Formally, define a function $A : G \to S$. Anytime it is desired to display the color $(X, Y, Z) \in G$, use the control sequence given by $A(X, Y, Z)$ to determine what combination of primaries to use. $A(X, Y, Z) \in M(X, Y, Z)$, so A specifies which of the many metamers should be displayed. With this method, one would not use both s and t, so metamerism would never occur.

If there are only three (linearly independent) primaries, then the construction of A is automatic. Since the three primaries are independent, no combination of any two of them can make the third, and the gamut is a parallelepiped based at the origin in XYZ space. The three primaries form a basis in a three-dimensional vector space, so any vector in the gamut can only be written as a unique linear combination of the primaries. The set G, as can be seen from Equation (4), consists of linear combinations of the primaries. Therefore A is uniquely defined as the components of the primaries needed to sum up to a given (X, Y, Z) triple.

If a color C is in a (possibly offset) parallelepiped, $P_{w,abc}$, define A in the natural way. Subtract the vector $\mathbf{w}$ from both C and $P_{w,abc}$. $P_{w,abc} - \mathbf{w}$ is now a parallelepiped at the origin, that is generated by $\{\mathbf{v_a}, \mathbf{v_b}, \mathbf{v_c}\}$, and contains the vector $C - \mathbf{w}$. Since the parallelepiped is three-dimensional, and its vector space basis is $\{\mathbf{v_a}, \mathbf{v_b}, \mathbf{v_c}\}$, the components of $C - \mathbf{w}$ in this basis are unique. Denote them by α_a, α_b, and α_c. Define

$$A(C) = \mathbf{w} + \alpha_a \mathbf{v_a} + \alpha_b \mathbf{v_b} + \alpha_c \mathbf{v_c}. \tag{18}$$

Since $\mathbf{w}$ is a sum of primaries that are not $\mathbf{v_a}, \mathbf{v_b}$, or $\mathbf{v_c}$, the right side of Equation (18) is within the gamut of the primaries. By construction, $T(A(C)) = C$, so, once we have chosen $P_{w,abc}$, we can use Equation (18) to specify a unique metamer.

The dissection algorithm decomposes the gamut into parallelepipeds. By defining A on each parallelepiped as in the above paragraph, A becomes a map on the entire gamut G. Since A is unique on each parallelepiped, A uniquely specifies the zonohedral control sequence as the choice of metamer. The uniqueness is sufficient to avoid non-matches, as desired

4.3 Avoiding Discontinuity

The parallelepiped dissection not only avoids non-matches, but, as this section will show, it also avoids discontinuity. Suppose that a continuous curve is drawn through the display gamut. Suppose further that a multi-primary display presents this curve to a viewer as a scale, much like the "hot" or "aquatic" color scales used in some graphics programs. Ideally, a viewer should perceive no discontinuity or jumps in such scales. The display maps gamut colors to control sequences, so the form of these mappings determines whether discontinuity occurs. Formally, A is a function from the gamut G, seen as a topological subspace of XYZ space, to the set S of control sequences, which are points in $[0,1]^N$. Each component, α_i, of a control sequence, can also be seen as a function from G to $[0,1]$. A as a whole is continuous if and only if each component α_i is continuous on G, which we will now show.

By construction, the dissection allows the gamut to be written as a disjoint union:

$$G = Z_3 \cup (Z_4 \backslash Z_3) \cup (Z_5 \backslash Z_4) \cup \cdots \cup (Z_N \backslash Z_{N-1}), \tag{19}$$

where $A \backslash B$ is the set of all elements that are in set A, but not in set B. The gamut can be compared to an onion, with each $Z_j \backslash Z_{j-1}$ adding a new layer of skin onto the previous layers. Since $\mathbf{v_i}$ is not used in assigning control sequences to the zonohedron generated by the first $i-1$ vectors, it follows that $\alpha_i = 0$ everywhere in Z_{i-1}, and so is continuous in Z_{i-1}. (If $i \leq 3$, then α_i is already continuous in Z_3, so assume $i \geq 4$.) The construction uses the Minkowski sum of $\mathbf{v_i}$, and some boundary parallelograms of Z_{i-1}, to make $Z_i \backslash Z_{i-1}$. Geometrically, the Minkowski sum sweeps out copies of those parallelograms, along a copy of $\mathbf{v_i}$. The sweeping motion is continuous, and is indexed by $\mathbf{v_i}$'s coefficient α_i, which varies from 0 to 1. The assignment of α_i is therefore continuous on $Z_i \backslash Z_{i-1}$. Furthermore, when the sweeping begins, at $\alpha_i = 0$, the parallelograms are on the boundary of Z_{i-1}, where α_i is always 0. Therefore, α_i moves continuously from Z_{i-1} to Z_i, and so is continuous on all Z_i.

As the construction proceeds to Z_{i+1}, additional points will be added to Z_i. $Z_{i+1} \backslash Z_i$ is the Cartesian product of $\alpha_{i+1}\mathbf{v_{i+1}}$, for $0 < \alpha_{i+1} \leq 1$, and a set, σ, of some (but not all) bounding parallelograms of Z_i. Since σ is a subset of Z_i, and α_i is continuous on Z_i, it follows that α_i is continuous on σ. By the construction in Figure 4, σ is swept along $\mathbf{v_{i+1}}$, to construct some new parallelepipeds, in $Z_{i+1} \backslash Z_i$. Since the translated vector $\mathbf{v_{i+1}}$ is a continuous set, and the new parallelepipeds are the Cartesian product of $\mathbf{v_{i+1}}$ and σ, it follows that α_i is continuous on all the new parallelepipeds. The union of the parallelepipeds constitutes $Z_{i+1}\backslash Z_i$, so α_i is continuous on all $Z_{i+1}\backslash Z_i$. By the construction, the limit of α_i as it approaches Z_i on a path through $Z_{i+1}\backslash Z_i$, equals its value on the boundary of Z_i, so α_i is continuous on all Z_{i+1}. Continue this argument inductively, to show that α_i is continuous on Z_{i+2}, Z_{i+3}, and so on until Z_N, which equals G.

The preceding argument has shown that the parallelepiped dissection guarantees color continuity, in a mathematical sense, on a multi-primary display. Unfortunately, it is not always clear how to go from a mathematical property, like continuity, to a perceived quality, like color smoothness. Other methods, such as Kanazawa's spherical average,[4] give coefficient functions that are not only continuous, but also have some differentiability. Using experiments involving human subjects, Murakumi *et al*[9] compare different control sequence assignment methods. Though they found that functions with some differentiability appear smoother than functions that are just continuous, it is still an open question as to how gradually coefficient functions should change, to produce acceptably smooth color scales for observers.

4.4 Avoiding Spurious Matches

Spurious matching, in which two colors that are different to the standard observer appear identical to an actual observer, is also possible. For example, a yellowed ocular lens might cause matching of two colors with different XYZ coordinates. In the context of our tiled parallelepipeds, a filter can cause spurious matches by making a vertex of one parallelepiped traverse one of the faces of the same parallelepiped and land at another point in which there is already a control sequence assigned. Brill and Larimer[7,10] give an example of such traversal, and show that the following two conditions are sufficient to avoid such spurious matches:

1. Using the CIE color-matching functions, the chromaticity curve $(x(\lambda), y(\lambda))$ traced out in λ (i.e., the spectrum locus) is convex and well-ordered in wavelength.

2. The spectra of any triplet of primaries, $\{p_1(\lambda), p_2(\lambda), p_3(\lambda)\}$, traces out in λ a curve $(p_1(\lambda), p_2(\lambda))/(p_1(\lambda)+p_2(\lambda)+p_3(\lambda))$ that is also convex and well-ordered in wavelength.

Many authors, e.g. West and Brill,[11] have noted that the first point is very nearly true. Brill and Larimer[7] refer to the criterion in the second point as the Binet-Cauchy criterion.

Any triad of primary spectra either satisfies the Binet-Cauchy criterion or not, regardless of how control sequences are assigned. However, the criterion significantly affects whether a given assignment algorithm avoids spurious matches. Consider the matrix-switching algorithm of Ajito *et al*,[3] which dissects the XYZ gamut into quadrangular pyramids. Each pyramid's vertices comprise the black point $\mathbf{b}$ and four vertices $\mathbf{q_j}, j = 1 \ldots 4$, of a parallelogram on the boundary of the display gamut. Each of the pyramids is spanned by three control variables: $\mathbf{q_1} - \mathbf{b}, \mathbf{q_2} - \mathbf{q_1}$, and $\mathbf{q_3} - \mathbf{q_1}$. Note that these spanning vectors are composite primaries, each being the sum of fully-activated single-primary vectors.

To see the effect of using simple versus composite primaries to address a color, we now pose an example. Suppose that all the simple primaries had spectra that were Gaussian in wavelength with the same width (given by a standard deviation, σ), but different peak wavelengths, denoted μ_i. Consider any triad of the Gaussian curves:

$$p_i(\lambda) = A_i \exp\left[-\frac{1}{2}\left(\frac{\lambda - \mu_i}{\sigma}\right)^2\right], i = 1, 2, 3. \tag{20}$$

As λ varies, the vector $(p_1(\lambda), p_2(\lambda), p_3(\lambda))$ traces out a curve in three-dimensional space. Construct a chromaticity analogue of this curve by centrally projecting the coordinates to the plane $p_3 = 1$. The chromaticity coordinates will be (P_1, P_2) , where

$$P_1(\lambda) = \frac{p_1(\lambda)}{p_3(\lambda)}, \tag{21}$$

$$P_2(\lambda) = \frac{p_2(\lambda)}{p_3(\lambda)}. \tag{22}$$

Substituting Equation (20) into Equations (21) and (22) gives

$$P_i(\lambda) = \frac{A_i}{A_3} \exp\left[\frac{1}{2\sigma^2}\left[\mu_3^2 - \mu_i^2 + 2\lambda(\mu_i - \mu_3)\right]\right], i = 1, 2. \tag{23}$$

Eliminating λ from the two equations in (23) reveals P_2 to be a power function of P_1, so the resulting curve of P_2 versus P_1 is convex. Map the coordinate system (P_1, P_2) into the coordinates used in point 2, as follows:

$$\frac{(p_1, p_2)}{p_1 + p_2 + p_3} = \frac{(P_1, P_2)}{P_1 + P_2 + 1}. \tag{24}$$

This mapping is a projective transformation, which sends straight lines into straight lines and hence preserves the convexity of convex curves. The spectrum locus of the triad in Expression (20) is therefore convex and well-ordered in wavelength, as required by point 2.

Although the spectrum locus of an arbitrary triad of primary curves is convex, sums of such curves, which will have multiple maxima in wavelength, can give triads whose spectrum locus will be far from convex. Hence the composite primaries used by Ajito *et al*[3] to span their quadrangular pyramids will be more susceptible to spurious matches than if the spanning primaries were chosen to be simple. The same can be said for the composite primaries in the algorithm of Brill and Larimer.[7,10] That algorithm tiles the available gamut with tetrahedra, many of whose spanning vectors comprise sums of fully-activated primaries. The algorithm works for monochromatic primaries because the convexity criterion applies only to the wavelength support of the primaries. However, when the primary spectra are Gaussians of equal width, the composite primaries will generate non-convex spectrum loci. Furthermore, the algorithm of Brill and Larimer does not exhaust the three-dimensional gamut of the display in XYZ space, and this is an additional problem.

In contrast to these methods, the current algorithm uses only one primary for each of the three spanning vectors of any offset parallelepiped. Therefore, if the simple primaries are the above-mentioned Gaussians, the spanning primaries of each offset parallelepiped will have a convex spectrum locus, the Binet-Cauchy criterion will be satisfied, and spurious color matches will be avoided. The above discussion used equal-spread Gaussians as primary spectra only to suggest plausibility. A more rigorous discussion (one that makes precise the likelihood of satisfying Binet-Cauchy and gives a best dissection for a given set of primaries) would be a fitting subject of a future investigation.

5 Comparison with Previous Work

Several algorithms have already been proposed, for addressing multi-primary displays. Apart from extensibility, most desirable properties of zonohedral control sequences occur in previous algorithms.

Like the current algorithm, Ajito *et al*'s matrix switching technique[3] dissects the color gamut into a disjoint union of discrete sets. Rather than parallelepipeds, their dissection consists of pyramids whose apex is at the origin, and whose base is one bounding parallelogram. A different color assignment matrix is used on each pyramid, assuring uniqueness so as to avoid non-matches.

Matrix switching and zonohedral sequencing are both mathematically continuous, but other methods were designed to maximize perceptual smoothness. König *et al*,[6] for example, noted that matrix switching caused visible artifacts in images when their colors crossed from pyramid to pyramid. To remedy this lack of smoothness, several averaging methods were suggested. Motomura[8] produced the LIQUID method, which interpolates over three colors of the same luminance as the color of interest. König[6] used a center of gravity construction, combined with a local choice of three primaries. Kanazawa[4] used a spherical average around the color of interest: each ray starting at the color of interest intersects the boundary of the gamut. Averaging methods insure uniqueness and improve smoothness, but have some drawbacks. First, they can be computationally demanding. Kanazawa's spherical average, for examples, requires calculating many intersections between lines and the display gamut, finding the control sequences at the intersections, and then evaluating a surface integral. Second, they can be difficult to visualize and therefore unintuitive to work with.

Brill and Larimer[7,10] approached the problem of metamerism in multi-primary displays in the context of the chromaticity diagram. To dissect the gamut into local XYZ regions and primary triplets, they used a set of parallelepipeds (with spanning vectors being either primaries or composite primaries) that shared a vertex at the origin (0, 0, 0). These parallelepipeds, which appear as triangles in (x, y) chromaticity space, did not cover the full three-dimensional gamut of the N-primary system (generated by Algorithm I). The reason was that each triad of composite primaries was constrained by maximum activation of any of the possible common ingredients (R, Y, G, etc.).

The current paper abandons composite primaries (with common simple-primary ingredients). Instead, only simple primaries can be used in a region of XYZ space. In the present paper, the elementary regions in XYZ space are parallelepipeds, each spanned by three primaries from a common black-point offset, that represents the sum of a number of other fully-activated primaries. Given a candidate triplet (X, Y, Z), we first find the elementary region to which it belongs, and then find the primary settings in that region that will render the triplet. This construction completes Brill and Larimer's, by working with the zonohedral color gamut directly, rather than with the chromaticity diagram.

6 Conclusion

This paper has presented an algorithm for efficiently finding the bounding vertices, edges, and faces of the tristimulus color gamut of a multi-primary display. A second, zonohedral, algorithm has also been presented, for dissecting that gamut into parallelepipeds. The dissection algorithm provides a unique address for every gamut point, with continuous control sequence transitions between points. In addition, it insures that current control sequences can be preserved when the gamut is extended by adding more primaries. The zonohedral algorithm is believed to be the first to possess this extensibility property. In addition, the algorithm's unique addressing avoids non-matches, in which two metameric control sequences appear different to a viewer. The assignment of control sequences is mathematically continuous in each primary's coefficient. If all the primaries are monochromatic, or narrow functions with a single peak, then an interposed filter or observer change will not result in spurious matches. More generally, any primaries that satisfy the Binet-Cauchy criterion will avoid spurious matches. Less strict rules than the Binet-Cauchy criterion might offer more practical alternatives that existing displays can meet. Finding such good-enough rules will be a future effort.

1 D B Judd, "The 1931 I. C. I. Standard Observer and Coordinate System for Colorimetry," *J Opt Soc Am* **23**, 359-374 (1933).
2 P McMullen, "On Zonotopes," *Trans Am Math Soc* **159**, September, 91-109 (1971).
3 T Ajito, K Ohsawa, T Obi, M Yamaguchi, and N Ohyama, "Color Conversion Method for Multiprimary Display Using Matrix Switching," *Opt Rev* **8**, No. 3, 191-197 (2001).
4 H Kanazawa, M Mitsui, M Yamaguchi, H Haneishi, and N Ohyama, "Color conversion for multi-primary displays using a spherical average method," *Proc CIC*, 65-69 (2004).
5 P Centore, "A Zonohedral Approach to Optimal Colours," accepted for publication in *Color Research and Application*, 2012.
6 F König, K Ohsawa, M Yamaguchi, N Ohyama, and B Hill, "A Multiprimary Display: Optimized Control Values For Displaying Tristimulus Values," *Proc PICS*, 215-220 (2002).
7 M H Brill and J Larimer, "Avoiding on-screen metamerism in N-primary displays," *J Soc Info Display* **13**, 509-516 (2005).
8 H Motomura, "Color conversion for a multi-primary display using linear interpolation on equi-luminance plane method (LIQUID)," *J Soc Info Display* **11**, 371-378 (2003).
9 Y Murakumi, N Hatano, J Takine, M Yamaguchi, and N Ohyama, "Evaluation of smooth tonal change reproduction on multi-primary display: comparison of color conversion algorithms," *Proc SPIE IS&T Electronic Imaging* **5289**, 275-283 (2004).
10 M H Brill and J Larimer, "Metamerism and multi-primary displays," *Info Display* **23**, No. 7 July, 16-21 (2007).
11 G West and M H Brill, "Conditions under which Schrödinger object colors are optimal," *J Opt Soc Am* **73**, 1223-1225 (1983).