Minimal-Energy Control Sequences for Linear Multi-Primary Displays

Paul Centore*

Abstract

A control sequence gives the intensities of the primaries for a pixel in a display. A multi-primary display has four or more primaries, so that multiple control sequences can sometimes produce the same colour. Different primaries likely consume different amounts of energy; furthermore, the energy consumption can be a complicated function. A minimal-energy control sequence for a target colour produces that colour with as little energy as possible. This paper shows that such minimal-energy sequences take a simple geometric form when each primary's energy function is linear. The display gamut, in CIE XYZ space, can be dissected into parallelepipeds. The originating vertex of each parallelepiped is the sum of a set of primaries at full intensity. Each edge of a parallelepiped is the translation of one primary. A colour with XYZ coordinates in a certain parallelepiped is a unique linear combination of the primaries in the originating vertex, and the three edge primaries. This paper proves that there exists a dissection such that these linear combinations are minimal-energy control sequences. In the generic case, this dissection is unique. An algorithm for a minimal-energy dissection is presented, along with an example.

Keywords — multi-primary, display, metamerism, energy, control sequence, gamut

1 Introduction

A display device such as a computer monitor combines a limited set of primaries to produce a wide gamut of colours. Many display devices use only three primaries (typically a red, green, and blue), but multi-primary devices use four or more. While additional primaries increase the gamut, they also introduce metamerism, in which two different combinations of primaries produce the same colour output. A multi-primary designer must therefore choose one combination from many. Various schemes[1,2,3] with various goals have been proposed for these choices. This paper was motivated by the goal of determining minimal-energy combinations for multi-primary displays with limited battery life or power availability.

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

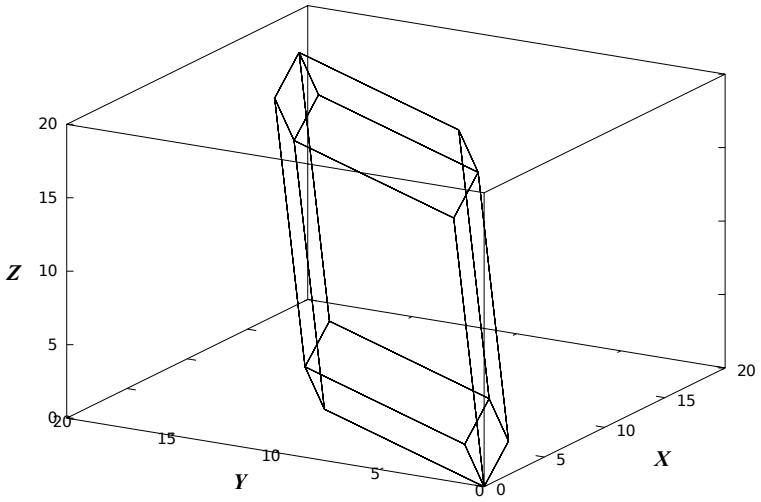

Figure 1: Four-Primary Gamut in XYZ Space

The general problem is difficult because the energy consumption function varies widely over different displays. Activating the red primary at full intensity, for instance, might require more than twice as much power as activating it at half intensity—or require no more power at all. In some displays, a primary draws power even when producing nothing visible. Since the general problem is likely too variable to be tractable, this paper addresses the special case in which each primary's intensity is a linear function of power. Furthermore, additivity is assumed: the power consumed by a combination of primaries is the sum of the powers consumed by the individual primaries.

In the linear case, activating a primary at full intensity would always require twice the power of activating it at half intensity. Different primaries, however, could require different amounts of power for full intensity, and their luminous efficiencies at full intensity could be different. These complications make even the linear case a sufficiently challenging problem. It is hoped that this paper's linear results will be a stepping stone to results for other, non-linear, displays.

The linear case, fortunately, has a simple geometric solution. The minimal-energy results for linear multi-primary displays can best be envisioned in three-dimensional colour space, indexed by XYZ coordinates. The colour gamut G of a display is the set of all colours which can be produced by that display. In the displays of interest in this paper, linear power response and additivity make the XYZ gamut into a zonohedron.[3] Figure 1 shows an example of a gamut when there are four primaries. The gamut is a convex polytope, and its faces (in the generic case) are all parallelograms. A *node* is the colour in XYZ space that is produced when a subset of the primaries is fully activated, and all other primaries are fully deactivated. Every vertex of the zonohedron is a node, but many nodes are also located inside the zonohedron.

Any XYZ point in the zonohedron can be produced by activating each of the primaries with a certain power. Scale the power for each primary so that its maximum value is 1. The set of scaled powers, one for each primary in some fixed order, is called a *control sequence*. A multi-primary display (i.e. one with four or more primaries) ex-

hibits metamerism: many different control sequences can produce the same XYZ colour. Though they produce the same colour, the control sequences can use more or less energy, depending on the primaries' power requirements. For energy-sensitive applications, an engineer would like to assign a control sequence that not only produces a specified colour, but also uses minimal energy.

The central result of this paper is that the minimal-energy control sequences for a linear display correspond to a parallelepiped dissection of the gamut in XYZ space. This dissection consists of disjoint parallelepipeds whose union is the entire gamut. Figure 2 is a dissection of Figure 1; it is an exploded view, in which the parallelepipeds are shrunken and spread out for clarity. Furthermore, the vertices of each parallelepiped are nodes in the gamut.

Each parallelepiped has an initial vertex and a final vertex, with corresponding initial and final nodes. The initial node is a sum of fully-activated primaries; its control sequence consists entirely of 0's and 1's, where the 1's correspond to primaries at full intensity. The control sequence for the final node similarly consists of 0's and 1's. Every 1 in the initial control sequence is also a 1 in the final control sequence, but the final sequence has three additional 1's that do not appear in the initial sequence. Every 0 in the final sequence is also a 0 in the initial sequence. The final and initial sequences therefore differ in three coefficients. The three edges of the parallelepiped that originate at the initial vertex are translations of the primaries, at full intensity, that correspond to those coefficients. These three edges form a vector space basis for the parallelepiped. As the three coefficients vary independently between 0 and 1, they can be used naturally as vector coordinates for the three-dimensional parallelepiped, with respect to that basis.

The dissection provides a map from the colour gamut to the set of control sequences. To produce a certain XYZ colour, find the parallelepiped in the dissection that contains those XYZ coordinates. The minimal-energy control sequence for that colour will have 1's where the parallelepiped's initial sequence has 1's, and 0's where the final sequence has 0's. The remaining three coefficients will be the coordinates of the XYZ point within the parallelepiped.

The paper's outline is as follows. First, linear algebra is used to rigorously formulate the minimal-energy problem for multi-primary displays. Then nodes and gamut dissections are also formulated rigorously. A technical concept that arises is genericity, in which multiple minimal-energy dissections exist but are unstable; a negligible adjustment of the energy function, smaller than measurement error, will produce a unique dissection. After the formulation, two propositions are proved, from which the existence of a dissection can be mathematically derived. The derivation leads to a dissection algorithm, from which minimal-energy sequences can be found. Finally, an example is given, using four primaries and a simple, hypothetical energy function.

2 Mathematical Formulation

Considerable mathematical machinery is needed to rigorously discuss multi-primary displays and linear energy consumption functions. Linear algebra provides a natural setting,

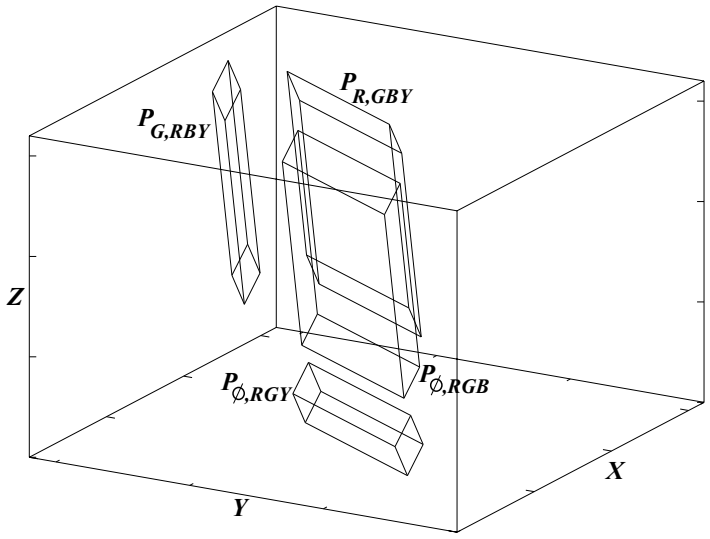

Figure 2: Parallelepipeds in Minimizing Dissection of Gamut

so this section first uses vector spaces and linear transformations to define the relevant terms. Next, we discuss geometric dissections of the display gamut in three-dimensional colour space. A final technical concept that is discussed is genericity.

2.1 The Colour Transformation C

In a linear display, a primary is a light source of fixed relative spectral density, whose intensity can vary from 0 to some maximum, which can be scaled to value 1. The colours produced by primaries and their combinations can be expressed in the XYZ coordinates developed by the Commission Internationale de l'Éclairage (CIE).[4] Colours in this system are vectors in the positive octant of $\mathbf{R}^3$. The set of all possible combinations of primaries forms a solid, called the display gamut, in the positive octant. Formally, suppose that N primaries are available. At full intensity, the colour T_i produced by the i^{th} primary is

$$T_i = [X_i, Y_i, Z_i]\,. \tag{1}$$

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

$$G = \left\{ \sum_{i=1}^{N} \alpha_i T_i \middle| \alpha_i \in [0,1]\,\forall i \right\}. \tag{2}$$

G is a subset of XYZ space. Geometrically, G is a zonohedron, that can be dissected into a set of parallelepipeds.[3,5]

A sequence of N coefficients α_i, all of which are between 0 and 1, is called a *control sequence*. The set of all control sequences can be seen as the unit hypercube, $U = [0,1]^N$, in the vector space $\mathbf{R}^N$. The colour generated by a control sequence can be expressed by

the linear transformation, C, from U to XYZ space:

$$C([\alpha_1, \alpha_2, \ldots, \alpha_N]) = \sum_{i=1}^{N} \alpha_i T_i. \tag{3}$$

The image of U under C is the entire colour gamut G of the display device.

The kernel[6] of C, denoted $\ker C$, is the set of all vectors $\boldsymbol{\alpha} \in \mathbf{R}^N$, such that

$$C(\boldsymbol{\alpha}) = [0, 0, 0]. \tag{4}$$

Wyszecki[7] introduced the term *metameric blacks* as a synonym for $\ker C$. The vector $\boldsymbol{\alpha} = 0$ is in $\ker C$. The image of a non-zero vector for which every α_i is non-negative, however, must consist of a positive X, Y, or Z, so such a vector can never be in $\ker C$. Since all the components of any vector in U are non-negative, U and $\ker C$ only intersect in the zero vector. A non-zero vector that is in $\ker C$ must therefore contain at least one negative component. Since a negative component has no physical meaning, the metameric blacks are imaginary. Like the CIE primaries,[4] however, they are still meaningful for mathematical operations.

It is known from linear algebra[6] that $\ker C$ is a vector subspace of $\mathbf{R}^N$. On practical grounds, the gamut G for any non-monochromatic display device would be expected to be three-dimensional. Therefore the rank of the transformation C is 3. The dimension of the kernel, or the *nullity*, of C is therefore $N-3$. If there were four primaries, for example, then $\ker C$ would be a 1-dimensional subspace of $\mathbf{R}^4$, so it would be a line through the origin.

In a multi-primary device, where $N > 3$, the transformation C is not injective: a given colour might be produced by many different metameric control sequences. Suppose there is a target colour T, given by some XYZ triple in G. Define the *pre-image* of T, denoted $C^{-1}(T)$, to be the set of all sequences that satisfy the first condition:

$$C^{-1}(T) = \{\boldsymbol{\beta} \in \mathbf{R}^N | C(\boldsymbol{\beta}) = T\}. \tag{5}$$

An important result from linear algebra[6] is that

$$C^{-1}(T) = \boldsymbol{\alpha} + \ker C, \tag{6}$$

when $\boldsymbol{\alpha}$ is any vector such that $C(\boldsymbol{\alpha}) = T$. Geometrically, $C^{-1}(T)$ is an affine subspace, that is, a vector subspace which has been shifted by adding the same vector to every vector in the subspace. Since $C^{-1}(T)$ is a parallel translation of $\ker C$, it maintains the same topological dimension of $N-3$. As long as T is not $[0, 0, 0]$, $C^{-1}(T)$ will not contain the origin. Furthermore, $C^{-1}(T)$ will intersect the unit hypercube U in points besides the origin.

2.2 The Energy Functional E

The ith primary $\mathbf{v_i}$ requires a certain energy, $E(\mathbf{v_i})$, to remain at full intensity for a set time period. For reduced intensity, presumably less energy is needed. The general form

of energy as a function of intensity, though presumably monotonic, can be complicated. This paper will deal only with the case of a linear energy function, i.e., one in which the energy required for each primary is proportional to its intensity. Only half of $E(\mathbf{v_i})$, for example, would be required to maintain $\mathbf{v_i}$ at half its full intensity. Like C, the energy function E is then a linear transformation (in fact, a linear functional) on the vector space $\mathbf{R}^N$ of control sequences:

$$E([\alpha_1, \alpha_2, \ldots, \alpha_N]) = \sum_{i=1}^{N} \alpha_i E(\mathbf{v_i}). \tag{7}$$

This definition also implies that the primaries' power consumptions are independent and additive: the total power required by a combination of primaries is the sum of the powers required by the individual primaries.

Although the colours produced by metameric sequences appear identical to a standard observer, their spectral distributions are different, and likely require different amounts of energy. A minimal-energy control sequence, $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_N]$, for a target colour T, satisfies

$$E(\alpha) \text{ is minimized,} \tag{8}$$

subject to the conditions

$$C(\alpha) = T, \tag{9}$$

$$0 \leq \alpha_i \leq 1, \forall i. \tag{10}$$

Standard linear programming[8] can find such sequences.

The programming problem also has a geometric interpretation that will be useful later. Any control sequence that produces the target colour T must be contained in $C^{-1}(T)$. All control sequences are contained in the unit hypercube U. Therefore, any control sequence that produces T must be in $C^{-1}(T) \cap U$. Both $C^{-1}(T)$ and U are convex sets, so their intersection, which will be denoted I, is also a convex set. In fact, since U is a convex polytope, and $C^{-1}(T)$ is an affine subspace, the set I is also a convex polytope. Any minimal-energy control sequence for T must be in I. When restricted to I, the linear functional E gives the energy consumed by any control sequence that produces T. The problem of finding a minimal-energy sequence is therefore the problem of minimizing a linear functional defined on a convex polytope.

A further useful map can be defined when minimal-energy control sequences are unique: the function σ from G to U, where $\sigma(T)$ is the unique minimal-energy control sequence that produces the target colour T. Geometrically, $\sigma(T)$ is the point on I at which E takes its minimum value. Since the map from target colours such as T to convex polytopes such as I is the pre-image of a linear function (and an intersection), it is continuous in the Hausdorff metric on convex sets. The assignment of the minimum point for the continuous function E to polytopes such as I is itself continuous. σ can be viewed as the composition of these two continuous assignments, so σ must also be a continuous function. Use Σ to denote the image of σ; in notation, $\Sigma = \sigma(G)$. Since

$C(\sigma(T)) = T$ for any T, therefore the inverse of σ is $C|_{\Sigma}$, which is also continuous. σ is therefore a homeomorphism, so G and Σ must have the same dimension, implying that Σ is a three-dimensional subset of U. This fact will be useful later on, when studying coordinate forms for minimal-energy sequences.

While conclusions about dimension are helpful, this paper will show that minimal-energy sequences have considerably more structure. In fact, they correspond to a simple geometric dissection of the three-dimensional gamut.

2.3 Parallelepiped Dissection

This paper shows that the minimal-energy control sequences that satisfy the programming problem in Equations (8) through (10) have a simple geometric form for linear multi-primary displays. The form involves dissecting the display gamut G, which is a polyhedron in three-dimensional colour space, into parallelepipeds. In such a dissection, the gamut is expressed as the union of parallelepipeds; any two parallelepipeds are disjoint, excepting a possible shared boundary when the parallelepipeds are adjacent.

Each vertex of each parallelepiped is produced by a *node*, that is, the sum of a set S of primaries: each primary in S is at full intensity; each primary outside S is at zero intensity. S can equivalently be thought of as a subset of the indices from 1 to N. The image of the node is the colour T given by

$$T = \sum_{i \in S} C(\mathbf{v_i}). \tag{11}$$

The control sequence for any node consists entirely of 0's and 1's. Every vertex of the gamut is a node,[9] but nodes can also be inside the gamut.

Each parallelepiped has an originating vertex ($\mathbf{v_O}$) and three generating vectors ($\mathbf{v_a}$, $\mathbf{v_b}$, and $\mathbf{v_c}$). All three generating vectors are primary vectors; the originating vertex is a sum of a set S of full-intensity primaries. The parallelepiped consists of all colours produced by control sequences of the form

$$\mathbf{v_O} + \sum_{i \in \{a,b,c\}} \alpha_i \mathbf{v_i}, \tag{12}$$

where $0 \leq \alpha_i \leq 1$, for all i. Formally, denote the parallelepiped by the notation

$$P_{S,abc} = \left\{ \sum_{i \in S} \mathbf{v_i} + \left(\sum_{i \in \{a,b,c\}} \alpha_i \mathbf{v_i} \middle| \alpha_i \in [0,1] \, \forall i \right) \right\}. \tag{13}$$

The final vertex, $\mathbf{v_F}$, of the parallelepiped is given by

$$\mathbf{v_F} = \mathbf{v_O} + \mathbf{v_a} + \mathbf{v_b} + \mathbf{v_c}. \tag{14}$$

If a parallelepiped's initial vertex and three generating vectors are specified, then each point in that parallelepiped has a unique control sequence in the form given by Equation

(12). Algebraically, the three generator vectors can be viewed as a vector space basis for the parallelepiped, and the α_i's are the unique coefficients with respect to that basis. The eight vertices of $P_{S,abc}$ occur when each α_i is 0 or 1 in Equation (13). A parallelepiped will be called a *minimal parallelepiped* if all the control sequences it defines are minimal-energy. A minimal-energy parallelepiped dissection is a dissection consisting entirely of minimal-energy parallelepipeds.

In a parallelepiped dissection, each gamut point is either in just one parallelepiped, or on the shared boundary of adjacent parallelepipeds. A gamut point in just one parallelepiped has a unique control sequence in terms of the vertex and generating vectors of that parallelepiped. A gamut point on multiple parallelepiped boundaries has the same control sequence, regardless of which parallelepiped is used to represent it. Rather than having three entries strictly between 0 and 1, gamut points on parallelepiped boundaries have two or fewer entries strictly between 0 and 1. Thus they can be reached from either parallelepiped. Whether inside a parallelepiped or on its boundary, then, a dissection assigns a unique control sequence to each colour in the gamut.

2.4 Genericity

It will be shown later that a minimal-energy dissection exists, but there is still the possibility that that dissection is not unique. The concept of genericity allows us to adjust the parameters, by an amount that is negligible in practice, so that there is a unique minimal-energy dissection.

A property is *generic* for a set if a dense, open subset possesses that property. As a consequence, any non-generic case can be modified, by arbitrarily small parameter adjustments, to produce a generic case. The property of being in *general position*, for example, is generic in the set of N-tuples of vectors. An N-tuple of vectors is in general position if any three of its N vectors are linearly independent. This property is generic because any set of three linearly dependent vectors can be made independent, by adjusting one vector coordinate by an amount less than any small ϵ. From a practical point of view, the vectors' coordinates are only known to within some ϵ, so the adjustment is too small to be measured. As a result, there is no loss of generality in assuming that the primaries are in general position. If needed, we will apply such adjustments to the energy functional E, to insure that the minimal-energy dissection is unique.

3 Mathematical Derivations

The definitions in the previous section will be used for the derivations in the current section. Equations (8) through (10) formalize the minimal-energy control sequence problem in linear programming terms. Although linear programming can readily find a minimal sequence for any particular T, we will be more interested in deriving conclusions about the general form of minimal-energy sequences. In particular, we will prove the mathematical theorem that a minimal-energy parallelepiped dissection always exists. The uniqueness of a minimal dissection is a generic property. An example will be given of a special case

where the dissection is not unique. but an arbitrarily small adjustment to the energy functional makes the dissection unique.

Two propositions will be proven, from which the dissection theorem follows. First, all but at most three coefficients of a minimal-energy sequence can be set to either 0 or 1. Second, any other sequence, with the same set of 0's and 1's as a minimal-energy sequence, is also minimal-energy. The three other coefficients can vary freely, filling out a parallelepiped, and all control sequences created in this way will be minimal-energy. The initial vertex of the parallelepiped will be the sum of a set S of primaries, all at full intensity, so it will be a node. The final vertex, and in fact all the vertices, will be nodes, too. The derivations imply that the gamut can be dissected into parallelepipeds, such that all the colours in one parallelepiped have minimal-energy control sequences with 0's and 1's for the all but the same three coefficients.

3.1 First Proposition

Lemma. *Suppose that F is a linear functional on a vector space V, and that L is an affine line in V (i.e. L is a translation of a one-dimensional subspace). Then either F is constant when restricted to L, or F is always increasing in one direction along L, and decreasing in the opposite direction.*

Proof. F can be viewed as a differentiable function from $\mathbf{R}^m$ to $\mathbf{R}$, where m is the dimension of V. L can be viewed as generated by a unit vector $\mathbf{v}$. Since F is linear, the directional derivative of F in the direction $\mathbf{v}$ is the same at any point on L. If the directional derivative is 0, then F is constant on L, which is the first alternative in the lemma's statement. If the directional derivative is not zero, then it is either negative or positive. If negative, then F is always decreasing along L in the direction given by $\mathbf{v}$, and increasing in the opposite direction$-\mathbf{v}$. If positive, then F is always increasing in the direction $\mathbf{v}$, and decreasing in the opposite direction. Either result satisfies the second alternative in the lemma's statement. ■

Proposition 1. *Suppose there is some colour T, in the gamut G, for which we have found a minimal-energy control sequence, $\boldsymbol{\alpha}$. In other words, $C(\boldsymbol{\alpha}) = T$, and $E(\boldsymbol{\alpha}) \leq E(\boldsymbol{\beta})$, for any other sequence $\boldsymbol{\beta}$, such that $C(\boldsymbol{\beta}) = T$. Then all but at most three of the coefficients of $\boldsymbol{\alpha}$ can be chosen to be either 0 or 1.*

Proof. Assume, by way of contradiction, that four or more coefficients, α_i through α_n, are strictly between 0 and 1. If needed, reorder the primaries so that

$$\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n, 1, 1, \ldots, 1, 0, 0, 0, \ldots, 0], \tag{15}$$

where the number of 1's and 0's are chosen appropriately, and $n > 3$.

We have seen that $\boldsymbol{\alpha}$ is in the set I, which is the intersection of $C^{-1}(T)$ and U. The energy functional E takes on various values over I, but the lemma restricts its behavior along affine lines in I. For each i from 1 through n, construct the line L_i that contains $\boldsymbol{\alpha}$, and extends indefinitely along both $-\mathbf{v}_i$ and $+\mathbf{v}_i$. Each point on L_i has the same form as

Equation (15), except that the coordinate α_i will be replaced by some β_i. The span of the n lines L_i is an affine subspace of dimension n, that contains $\boldsymbol{\alpha}$. $C^{-1}(T)$ is also an affine subspace, of dimension $N-3$, that contains $\boldsymbol{\alpha}$. Since $n > 3$, the intersection J of these two subspaces must have dimension at least 1, so J contains at least one line L, which itself contains $\boldsymbol{\alpha}$. Since L is a subset of J, which is a subset of $C^{-1}(T)$, it follow that C maps every point in L to T.

From the lemma, it follows that E decreases (or at least stays at the same value), as one moves away from $\boldsymbol{\alpha}$, in one of the directions along L. To reduce E as much as possible, keep moving in that direction until the boundary of U is reached, at some point $\boldsymbol{\gamma}$, which occurs when one of the β_i's becomes 0 or 1. Now $E(\boldsymbol{\gamma}) \leq E(\boldsymbol{\alpha})$, and $\boldsymbol{\gamma}$ has at least one more 0 or 1 than the expression for α in Equation (15). Furthermore, $C(\boldsymbol{\gamma}) = T$, so we have found a control sequence that has one less interior coordinate than $\boldsymbol{\alpha}$, and also requires no more energy.

Repeat the above argument with $\boldsymbol{\gamma}$ instead of $\boldsymbol{\alpha}$. We can then produce a $\boldsymbol{\gamma}$ with $n-2$ coefficients strictly inside $[0,1]$, a $\boldsymbol{\gamma}$ with $n-3$ such coefficients, and so on. The process will end when the dimension of the span of the L_i's is 3 or less, at which time there is no longer guaranteed to be a non-empty intersection with $C^{-1}(T)$, which has dimension $N-3$. When the dimension of the span is 3, however, n is also 3. In that case Equation (15) has no more than three entries which are not 0 or 1, proving the proposition. ■

3.2 Second Proposition

Proposition 2. *Suppose that a control sequence $\boldsymbol{\alpha}$ is a minimal-energy control sequence for some target colour T_α in the gamut G, and that $\boldsymbol{\alpha}$ has the form*

$$\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \alpha_3, 1, 1, \ldots, 1, 0, 0, 0, \ldots, 0], \tag{16}$$

where α_1, α_2, and α_3 are all strictly between 0 and 1. (Proposition 1 implies that this form is the rule rather than the exception.) Then any *control sequence $\boldsymbol{\beta}$ of the form*

$$\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3, 1, 1, \ldots, 1, 0, 0, 0, \ldots, 0], \tag{17}$$

where β_1, β_2, and β_3 are all strictly between 0 and 1, is also a minimal-energy control sequence.

Proof. A linear functional such as E can be characterized by its behavior in all possible directions in its vector space domain. Let $k\boldsymbol{\delta}$ denote a vector $\boldsymbol{\delta}$, multiplied by a positive constant k. Then $k\boldsymbol{\delta}$ defines a direction in the vector space: as long as k is positive, even if it is arbitrarily large or small, every $k\boldsymbol{\delta}$ is in the same direction. The direction $k\boldsymbol{\delta}$ can originate at any point in the vector space. As k increases, according to the lemma, the linear functional E, evaluated along $k\boldsymbol{\delta}$, either always increases, always decreases, or always maintains the same value. Thus it is only necessary to test E with an arbitrarily small k to conclude that E increases along $k\boldsymbol{\delta}$.

The points in the vector space $\mathbf{R}^N$ represent different control sequences. The sequence $\boldsymbol{\alpha}$ is a minimal-energy sequence, so E takes on a minimum value at $\boldsymbol{\alpha}$, when restricted to

the convex set $I_\alpha = C^{-1}(T_\alpha) \cap U$. A direction $k\boldsymbol{\delta}_\alpha$ that originates at $\boldsymbol{\alpha}$ and extends into I_α will be called an interior direction. Since $\boldsymbol{\alpha}$ minimizes E, it follows that E must increase (or at least maintain the same value) along every interior direction. In fact, this criterion is reversible: if E increases (or maintains the same value) along every interior direction originating at $\boldsymbol{\alpha}$, then $\boldsymbol{\alpha}$ must be a minimal-energy sequence. This characterization of minima also applies to $\boldsymbol{\beta}$. $\boldsymbol{\beta}$ is a minimal sequence if E increases along every $k\boldsymbol{\delta}_\beta$ that extends into $I_\beta = C^{-1}(T_\beta) \cap U$, where T_β is the colour produced by the control sequence $\boldsymbol{\beta}$.

To show that $\boldsymbol{\beta}$ is a minimal-energy sequence, we will identify every interior direction originating at $\boldsymbol{\beta}$, and show that E increases in each of those directions. In fact, we will show that the interior directions that originate at $\boldsymbol{\beta}$ are identical to the interior directions that originate at $\boldsymbol{\alpha}$. Since $\boldsymbol{\alpha}$ is minimal, E must increase in each of those directions, and the result follows. The argument will use Proposition 1, which says that a minimal-energy sequence contains many 0's and 1's. (A loose geometric interpretation is that a minimal-energy sequence is on the boundary of its corresponding I.) If a component is 0, then every interior direction must have a positive entry for that component; if a component is 1, a negative entry is necessary. The large number of 0's and 1's in $\boldsymbol{\alpha}$ therefore limits the set of interior directions.

A direction $k\boldsymbol{\delta}_\beta$ is an interior direction at $\boldsymbol{\beta}$ if $\boldsymbol{\beta} + k\boldsymbol{\delta}_\beta$ is inside I_β for a sufficiently small k. Since U and $C^{-1}(T_\beta)$ both contain I_β, it follows that each component $\beta_i + k\delta_{\beta i}$ must satisfy

$$0 \leq \beta_i + k\delta_{\beta i} \leq 1, \forall i, \tag{18}$$

and that

$$C(k\boldsymbol{\delta}_\beta) = 0. \tag{19}$$

Thus $\boldsymbol{\delta}_\beta$ is in the kernel of C. The direction $\boldsymbol{\delta}_\alpha$ (which is just $\boldsymbol{\delta}$ translated to α) is still in the kernel of C, so $C^{-1}(T_\alpha)$ and $C^{-1}(T_\beta)$ consist of the same directions.

We will similarly show that there is a one-to-one correspondence between the interior directions from $\boldsymbol{\beta}$ into U, and the interior directions from $\boldsymbol{\alpha}$ into U. Add and subtract the coefficients of α in Equation (18):

$$0 \leq (\beta_i - \alpha_i) + \alpha_i + k\delta_{\beta i} \leq 1, \forall i. \tag{20}$$

The direction $k\delta$ can be shifted to any originating point, so Equation (20) is equivalent to

$$0 \leq (\beta_i - \alpha_i) + \alpha_i + k\delta_{\alpha i} \leq 1, \forall i. \tag{21}$$

From the statement of the proposition, it follows that

$$\boldsymbol{\beta} - \boldsymbol{\alpha} = (\beta_1 - \alpha_1, \beta_2 - \alpha_2, \beta_3 - \alpha_3, 0, 0, \ldots, 0)\,, \tag{22}$$

so Equation (21) reduces to

$$0 \leq \alpha_i + k\delta_{\alpha i} \leq 1, \text{for } i = 4 \ldots N. \tag{23}$$

Since $\boldsymbol{\alpha}$ has been assumed to be minimal, there is a small enough k so that Equation (23) holds. Therefore, the last $N-3$ components satisfy Equation (18).

We will now show that the first three components also satisfy Equation (18). The first three components of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are strictly between 0 and 1, so, regardless of $\boldsymbol{\delta}$, there exists a small enough k to simultaneously insure

$$0 \leq \alpha_i + k\delta_{\alpha i} \leq 1, \quad \text{for} \quad i = 1, 2, 3, \text{ and} \tag{24}$$

$$0 \leq \beta_i + k\delta_{\beta i} \leq 1, \quad \text{for} \quad i = 1, 2, 3. \tag{25}$$

Thus every interior direction from $\boldsymbol{\alpha}$ into U is also an interior direction from $\boldsymbol{\beta}$ into U.

Since $C^{-1}(T_\alpha)$ and $C^{-1}(T_\beta)$ also consist of the same directions, the interior directions from $\boldsymbol{\alpha}$ into I_α, and from $\boldsymbol{\beta}$ into $I_{\boldsymbol{\beta}}$, are therefore identical. Since E increases or decreases along a direction, regardless of where that direction originates, and since E increases along every interior direction from $\boldsymbol{\alpha}$ into I_α, it follows that E increases along every interior direction from $\boldsymbol{\beta}$ into I_β. This statement is equivalent to saying that $\boldsymbol{\beta}$ is a minimal-energy control sequence, as was to be shown. ■

3.3 Parallelepiped Dissection Theorem

Parallelepiped Dissection Theorem. *Assume that there is a unique minimal-energy control sequence for each colour in the gamut. Then there exists a unique minimal-energy parallelepiped dissection for the gamut, and all the parallelepiped vertices are nodes.*

Proof. To begin with, we will prove the technical condition that there exists a target colour T whose minimal-energy sequence has the form given in Equation (16), i.e. T contains exactly three entries that are not 0 or 1. We need such a T so that we can apply Proposition 2. To see that such a T exists, use the previously shown result that Σ, the subset of U consisting of minimal-energy sequences, has dimension three when such sequences are unique. Within Σ, then, there are at three degrees of freedom, so one could adjust three entries of a minimal-energy sequence (i.e. an element of Σ) by an arbitrarily small amount, and still remain in Σ. Thus, by starting with any minimal-energy sequence, we can move to a nearby sequence τ that has at least three entries that are not 0 or 1. Let T be $C(\tau)$.

Next, construct a minimal-energy parallelepiped around such a target colour T. By the theorem's assumptions, the minimal-energy control sequence $\boldsymbol{\alpha}$ such that $C(\boldsymbol{\alpha}) = T$ is unique. By Proposition 2, any control sequences $\boldsymbol{\beta}$ that satisfy Equation (17) are also minimal-energy, and, by the theorem's assumptions, are unique. Geometrically, the control sequences $\boldsymbol{\beta}$ form a parallelepiped in XYZ space. The originating vertex ($\mathbf{v_O}$) of the parallelepiped is the node given by summing up all the primaries represented by a 1 in Equation (17). Equivalently, those primaries are all displayed at full intensity, as shown in Equation (11). Use S to denote the set consisting of indices of these primaries. The first three primaries in Equation (17) will be denoted $\mathbf{v_a}$, $\mathbf{v_b}$, and $\mathbf{v_c}$, and can be displayed at any three intensities between 0 and 1. The remaining primaries are not used at all; their intensities are the 0's in Equation (17). In accordance with Equation (13), the parallelepiped as a whole is denoted $P_{S,abc}$. By Proposition 2, any control sequence

generated by Equation (13) is of minimal energy. $P_{S,abc}$ is therefore a minimal-energy parallelepiped that contains T.

Now choose another colour outside that parallelepiped, insuring that its minimal-energy sequence is of the form in Equation (16), and construct a second minimal-energy parallelepiped. Continue with a third point, outside both parallelepipeds, construct a third parallelepiped, and so on. Since the vertices of each parallelepiped are nodes, and since there are only 2^N nodes, there can be only finitely many parallelepipeds. The linear programming problem and the two propositions imply that a unique minimal parallelepiped can be found around any colour in the gamut, so the finiteness implies that the entire gamut will eventually be filled with non-overlapping minimal parallelepipeds.

It is possible that a target colour T is simultaneously on the boundaries of multiple adjacent parallelepipeds. In that case, its minimal-energy sequence has two or fewer entries that are not 0 or 1. Since the map σ is continuous, the coordinates for any parallelepiped can be extended to the boundary of the parallelepiped, and any boundary control sequences (in parallelepiped coordinates) will still be minimal, since they are the limit points of minimal sequences in the interior. By the continuity of σ, it does not matter which of the adjacent parallelepipeds is used: the boundary control sequence will be identical. ■

Figure 2 shows an example of a parallelepiped dissection of the colour gamut in Figure 1. In that case, there were four primaries, and the gamut was dissected into four parallelepipeds. Details of this example will be given after a dissection algorithm is presented formally.

The existence of a set of minimal parallelepipeds that fills the gamut is a major result of this paper. Leaving aside questions of uniqueness, which the next section will deal with, we have assigned every colour in the gamut to its minimal-energy sequence. Furthermore, the assignment can be expressed in a geometrically tangible form, as a dissection of three-dimensional colour space. Designers can therefore start from the user's world (visible colour stimuli), and arrive naturally at the engineer's world (a set of power outputs for primaries).

3.3.1 Possible Non-Uniqueness of Minimal Sequences

Although a minimal dissection always exists, a simple example will show that it need not be unique: multiple minimal dissections can occur. Cases with multiple minimal dissections are non-generic, however, and can easily be adjusted to produce a unique dissection. Figure 3 gives an example. 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

$$\mathbf{v_1} + \mathbf{v_3} = \mathbf{v_2} + \mathbf{v_4}. \tag{26}$$

Now suppose that each primary consumes one unit of energy when at full intensity. Then either side of Equation (26) requires two units of energy. Furthermore, there are many other metameric combinations for this point, such as $\frac{1}{3}\mathbf{v_1} + \frac{2}{3}\mathbf{v_2} + \frac{1}{3}\mathbf{v_3} + \frac{2}{3}\mathbf{v_4}$. All the

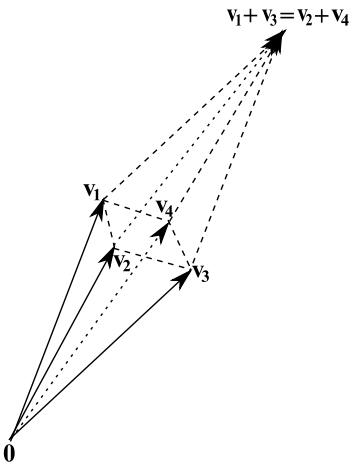

Figure 3: Multiple Minimal-Energy Sequences for the Same Colour

metamers also require two units of energy. If there are only these four primaries, then *all* these metamers are minimal-energy.

Two possible minimal dissections for the gamut G produced by the four primaries are

$$G = P_{0,123} \cup P_{0,134} \cup P_{1,234} \cup P_{3,124}, \text{and} \tag{27}$$

$$G = P_{0,124} \cup P_{0,234} \cup P_{2,134} \cup P_{4,123}. \tag{28}$$

The existence of two minimal dissections, however, is unstable: adjusting some parameters by a negligible amount can eliminate one of the possibilities. A simple approach to this particular problem is to adjust the energy for any one of the primaries, say $\mathbf{v_1}$. On physical grounds, a small adjustment is reasonable, because the measurement of the energy is only good to a certain number of significant digits. Rather then using $E(\mathbf{v_1}) = 1$, for example, assume that $E(\mathbf{v_1}) = 0.999$. Then the left side of Equation (26) uses less energy than the right side. It will then follow that the dissection given in Equation (27) is not minimal, but the dissection in Equation (28) is.

3.3.2 Other Dissections

Gamut dissections for multi-primary displays have occurred previously.[2,3,5] Rodriguez-Pardo et al.,[5] as well as Centore and Brill,[3] each presented parallelepiped dissections. Ajito et al.[2] presented "matrix switching," which dissected the gamut into pyramids with apices at the origin.

While it is interesting to note that dissections, and particularly parallelepiped dissections, have occurred in other contexts, comparisons are not really apropos, because the authors' aims are different. Rodriguez-Pardo and Ajito were both interested in computational efficiency, the first author for calculating the gamut's volume, and the second author for assigning control sequences. Ajito's method reduced a three-dimensional look-up table (LUT) to a two-dimensional LUT. All the pyramids originated at the origin, so each pyramid was restricted to a convex set in the chromaticity diagram; those convex sets were the basis of the two-dimensional LUT. Centore and Brill were looking for a method

that allowed easy extension to further primaries. Other motivations might also be reasonable when assigning control sequences, such as producing spectral matches,[1] reducing the effects of observer metamerism, or catering to colour-deficient viewers. The current paper, of course, aims to minimize energy consumption. While the needed calculations are likely reasonably efficient, no attempt was made to maximize efficiency, nor to achieve other goals. The minimal-energy dissection should therefore stand on its own, rather than be compared with other dissections.

4 The Dissection and Assignment Algorithms

Once the primaries' XYZ colours and energy consumptions are known, the following dissection algorithm, which is based on the proof of the theorem, will generate a minimal-energy dissection:

1. Construct all the parallelepipeds in the gamut:
   (a) List the 2^N nodes in the gamut. A node is the sum of all primaries whose indices are in a subset S of $\{1, 2, 3, \ldots, N\}$.
   (b) The set S^c is the complement of S. S^c is the indices of the remaining primaries, that are not in the node. For each node, construct S^c.
   (c) For each node, list the combinations (if any) of three primaries in S^c. The i^{th} combination comb(i) is a set of three distinct indices, $\{a_i, b_i, c_i\}$, all of which are in S^c.
   (d) For each S and comb(i), there is a parallelepiped $P_{S,\text{comb}(i)}$. Every parallelepiped is of this form, so we have constructed a complete list of parallelepipeds.
2. For each parallelepiped $P_{S,\text{comb}(i)}$:
   (a) Choose an arbitrary control sequence α in the interior of $P_{S,\text{comb}(i)}$. This sequence will have 1's for every primary in S, entries between 0 and 1 for the three primaries in comb(i), and 0's elsewhere.
   (b) Calculate the colour $T = C(\alpha)$, and solve the linear programming problem defined by Equations (8) through (10). The open-source Octave/MATLAB routine glpk() is sufficient for this problem.
   (c) If the solution to the linear programming problem is again α, then Proposition 2 implies that every control sequence defined by $P_{S,\text{comb}(i)}$ is minimal-energy. Declare the parallelepiped as a whole to be minimal-energy; otherwise, declare the parallelepiped is not minimal-energy.
3. The result of Step 2 is a list of minimal-energy parallelepipeds. In the generic case, this list will dissect the gamut completely. Check for dissection by making sure that none of the minimal-energy parallelepipeds overlap; if there are no overlaps, then we have found a dissection, and the algorithm terminates.

Primary	X	Y	Z	Energy
R	6.3	2.5	0.0	1
G	1.9	8.7	2.5	3
B	2.4	2.8	16.7	2
Y	4.9	3.8	0.0	4

Table 1: XYZ Coordinates and Energy Consumption for Primaries

4. If there are overlaps, then identify unstable equal-energy relationships like the one in Equation (26). Adjust one or more energy levels very slightly, to invalidate those relationships, and re-run the algorithm.

Once the gamut has been dissected into parallelepipeds, the following assignment algorithm determines the minimal-energy sequence for a target colour T, without solving a linear programming problem:

1. Determine which parallelepiped T belongs to.

2. Subtract the originating vertex of that parallelepiped from both T and the parallelepiped. The result will be a shifted parallelepiped that starts at the origin, and a T' that is contained in the shifted parallelepiped.

3. Construct a change-of-basis matrix M from XYZ coordinates to the generating vectors of the shifted parallelepiped.

4. Multiply T' by M^{-1} to find the coefficients for the generating vectors. All the primaries that make up the originating vector will have coefficients of 1, and all remaining primaries will have coefficients of 0.

Once the primaries and their energy consumptions are known, the dissection algorithm needs to be run only one time, to determine a set of parallelepipeds. Similarly, a matrix M can be precomputed for each parallelepiped, to make later computations faster.

5 Dissection Example

This section will give a simple hypothetical example involving four primaries, shown in Figure 4. The primaries are the red, green, blue, and yellow (R,G,B,Y) primaries, taken from Figure 2 of Reference 1. Table 1 gives the CIE XYZ coordinates for these primaries, and Figure 5 displays them in XYZ space. Table 1 also gives the primaries' energy consumption; these consumptions are hypothetical values that provide a simple illustration . The entries in this table will reappear as entries in the matrix representations for the linear transformations C and E. The complete gamut in Figure 1 is the polyhedron (in fact, the zonohedron) generated from these four primaries, using Equation (2).

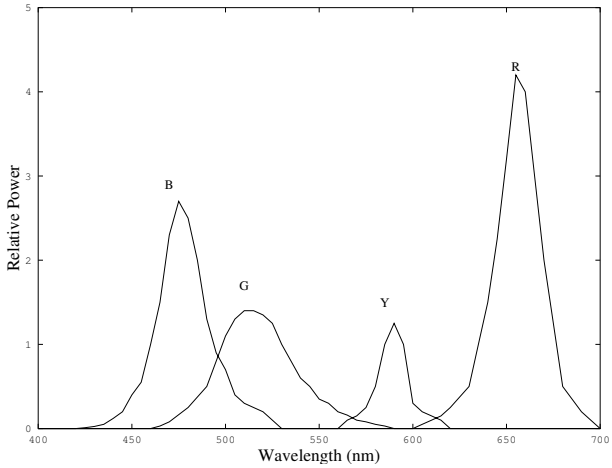

Figure 4: Relative Spectral Power Densities of Four Primaries

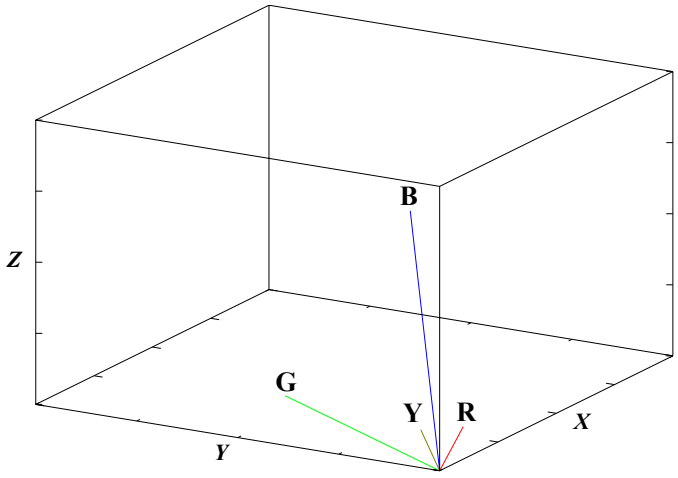

Figure 5: Four Primaries in XYZ Space

The first step in the dissection algorithm is to construct all the parallelepipeds in the gamut, by an exhaustive search. The eight possible parallelepipeds are:

$$P_{\phi,RGB}, \tag{29}$$
$$P_{\phi,RGY}, \tag{30}$$
$$P_{\phi,RBY}, \tag{31}$$
$$P_{\phi,GBY}, \tag{32}$$
$$P_{R,GBY}, \tag{33}$$
$$P_{G,RBY}, \tag{34}$$
$$P_{B,RGY}, \tag{35}$$
$$P_{Y,RGB}, \tag{36}$$

where ϕ denotes the empty set: parallelepipeds with ϕ for an originating vertex start at the origin.

In accordance with Step 2 of the dissection algorithm, each parallelepiped was investigated individually, to see whether the control sequences it generated were in fact minimal-energy. $P_{G,RBY}$ provides an example. The center of this parallelepiped was chosen as an arbitrary control sequence in the parallelepiped's interior. The center is given by $G + (1/2)(R + B + Y)$, which is approximately [8.7, 13.3, 10.8] in XYZ coordinates. The linear programming problem given by Equations (8) through (10) was then solved, with

$$E(R,G,B,Y) = R + 3G + 2B + 4Y, \tag{37}$$
$$C(R,G,B,Y) = \begin{bmatrix} 6.3 & 1.9 & 2.4 & 4.9 \\ 2.5 & 8.7 & 2.8 & 3.8 \\ 0.0 & 2.5 & 16.7 & 0.0 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \\ Y \end{bmatrix}, \tag{38}$$
$$T = [8.7, 13.3, 10.8]. \tag{39}$$

The solution to this programming problem was a minimal-energy control sequence:

$$\alpha = \left[\frac{1}{2}, 1, \frac{1}{2}, \frac{1}{2}\right], \tag{40}$$

which is just the coordinate expression for the center of the parallelepiped. Since the coordinates given by the parallelepiped for the center are in fact a minimal-energy control sequence, Proposition 2 insures that the coordinates for *every* point in that parallelepiped are also minimal-energy sequences. Therefore, the parallelepiped $P_{G,RBY}$ was included in the minimal dissection.

Similar tests were performed on the remaining seven parallelepipeds, using their centers as a test sequence for convenience. Sometimes the control sequence at the center was not minimal-energy, in which case that parallelepiped was rejected. Otherwise, it was

accepted. In all, four parallelepipeds were accepted:

$$P_{\phi,RGB}, \tag{41}$$
$$P_{\phi,RGY}, \tag{42}$$
$$P_{R,GBY}, \tag{43}$$
$$P_{G,RBY}. \tag{44}$$

These four parallelepipeds are shown in Figure 2, in an exploded view. When re-assembled, they perfectly dissect the gamut shown in Figure 1. This dissection does not use the Y primary if it can be avoided. The R and G primaries are preferred instead, as origins of the dissecting parallelepipeds. This preference makes sense: Y consumes more energy than the other primaries, so it should only be used when substitutions are impossible. As is almost always the case, this example is generic, so there is no overlap between any of the accepted parallelepipeds. Had some of the parallelepipeds overlapped, some of the energy levels could have been adjusted very slightly, to move to a generic problem.

6 Summary and Further Investigations

This paper has proven that minimal-energy control sequences take a simple geometric form for linear multi-primary displays. The display gamut in CIE XYZ space can be dissected into parallelepipeds. The originating vertex of each parallelepiped is a sum of primaries at full intensity. The sides of each parallelepiped are translations of three other primary vectors. The minimal-energy control sequence for a colour in a parallelepiped is the sum of the originating vertex, plus a unique linear combination of the three vectors that generate the sides. In the generic case, each colour has only one minimal-energy sequence, so there is only one such dissection. The existence of a dissection was proven mathematically, a simple algorithm was given to determine the dissection, and an example was presented.

This paper has dealt with only the special case in which energy functions are linear. Of course, energy functions for different display technologies can take many forms, often considerably more complicated. As a result, their minimal-energy sequences will also likely be more involved than a simple dissection. The mathematical approach used here is already somewhat elaborate, and even more elaborate approaches would likely be needed for other energy functions. It is hoped, however, that this rigorous analysis of the linear case will serve as a stepping stone for more general cases.

1. Yuri Murakumi, Jun-ichiro Ishii, Takashi Obi, Masahiro Yamaguchi, & Nagasaki Ohyama, “Color conversion method for multi-primary display for spectral color reproduction,” Journal of Electronic Imaging 13(4), pp. 701-708, October 2004.
2. Takeyuki Ajito, Kenro Ohsawa, Takashi Obi, Masahiro Yamaguchi, & Nagasaki Ohyama, “Color Conversion Method for Multiprimary Display Using Matrix Switching,” Optical Review, Vol. 8, N0. 3, pp. 191-197, 2001.
3. Paul Centore & Michael H. Brill, “Extensible Multi-Primary Control Sequences,” Journal of the Society for Information Display, Vol. 20, Issue 1, pp. 12-21, Jan 2012.

4. CIE, *Colorimetry, 3rd ed.*, CIE Publication No. 15:2004, Vienna, 2004.
5. Carlos Eduardo Rodriguez-Pardo, Gaurav Sharma, Jon Speigle, Xiao-Fan Feng, & Ibrahim Sezan, "Efficient Computation of Display Gamut Volumes in Perceptual Spaces," 19th Color and Imaging Conference Final Program and Proceedings, pp. 132-138, November 2011.
6. Kenneth Hoffman & Ray Kunze, *Linear Algebra, 2nd ed.,*Prentice-Hall, Inc., 1971.
7. G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae, 2nd ed.* (New York: John Wiley & Sons, 1982).
8. S. R. Lay, *Convex Sets and Their Applications*, Dover Publications, 2007.
9. P. McMullen, "On Zonotopes," *Transactions of the American Mathematical Society*, Vol. 159, pp. 91-109, September 1971.

