# A Zonohedral Approach to Optimal Colours

Paul Centore



**Abstract**

*This paper demonstrates that the CIE $XYZ$ colour solid is a zonoid. An approximating zonohedral colour solid is constructed explicitly from a set of generating vectors, which are integrals of colour-matching functions over narrow intervals of the visible spectrum. The zonohedral approach yields an intuitive, constructive proof of the Optimal Colour Theorem: the reflectance function of an optimal colour takes on only the values 0 or 1, with at most two transition wavelengths. In addition, zonohedral techniques can simplify computations: for example, optimal colours can be found without calculating transition wavelengths. Finally, zonohedra provide a simple, unified approach to colour space, and eliminate much of the confusion arising from chromaticity diagrams.*

**Keywords:** optimal colour, zonohedron, MacAdam limit, colour solid

## 1 Introduction

An optimal colour was originally defined as one that could be realized in a material form such as pigment, and that has a greater luminosity than any other material colour of the same chromaticity, when both are viewed under the same illuminant. The material requirement excludes coloured light sources, and limits considerations to materials that reflect light without generating any light of their own. A colour of a material is also called a local colour, because it depends only on the local material, rather than on the viewer or the illuminating light source. Local colours are defined, up to metamerism, by their reflectance functions, which give the percentage of light they reflect, of each wavelength in the visible spectrum. Perceptually, local colours can be expressed in $XYZ$ coordinates, which can be calculated from a reflectance function. The $XYZ$ colour solid consists of the $XYZ$ coordinates for all local colours. In a more modern formulation, optimal colours are those which are on the boundary

of the colour solid. The Optimal Colour Theorem states that a colour is optimal if and only its reflectance function takes on only the values 0 or 1 (which equals 100%), with at most two transition wavelengths.

David MacAdam[1] gave the first complete proof of the Optimal Colour Theorem, in 1935. The proof is unintuitive because it relies on an analogy with center of gravity calculations, that implicitly introduces an extraneous, non-intrinsic, metric concept. Furthermore, the proof works in the chromaticity diagram, which is difficult to interpret, being a transformed two-dimensional section of three-dimensional colour space. The current paper presents a more intuitive proof. The proof is set in three dimensions, and uses an explicit construction for the colour solid, avoiding the difficulties of chromaticity diagrams. Furthermore, the geometric arguments have immediate analogues in colour phenomena. As a byproduct of the proof, computational demands for optimal colour algorithms can likely be reduced. In particular, it should be possible to avoid calculating transition wavelengths, which is currently a typical approach.

Overall, zonohedra unify and simplify concepts of colour space. Zonohedra are three-dimensional projections of Koenderink's hyperparallelepipeds[2], which are constructed in an infinite-dimensional "beam" space. While Koenderink works more algebraically in the spectral domain, the current approach is more geometric, in that every step of the construction can be visualized in three-dimensional space.

The main innovation in this paper is the use of zonohedra. A zonohedron is a convex polyhedron that is generated by Minkowski addition from a set of vectors. It will be shown that the $XYZ$ colour solid can be constructed naturally as a zonohedron. The generating vectors are the colours whose reflectance function is 1 on a narrow interval of the visible spectrum, and 0 elsewhere. In the limit, a generating vector would be a monochromatic colour, whose reflectance function is 1 at a particular wavelength, and 0 elsewhere. In this ideal case, the colour solid is a zonoid, which is the limit of a sequence of discretely constructed zonohedra. Since computations are discrete in practice, we will focus on zonohedra.

The paper is organized as follows. Section 2 describes colour solids and optimal colours. Section 3 presents a mathematical treatment of zonohedra. Section 4 contains the main ideas of the paper, showing in detail how the colour solid can be constructed as a zonohedron, and providing an intuitive proof of the Optimal Colour Theorem. Section 5 describes how zonohedral techniques can simplify optimal colour calculations. Section 6 relates this paper's results to previous work, highlighting some cases where the new constructions clarify concepts. Finally, Section 7 summarizes the paper's conclusions.

## 2 The CIE XYZ Colour Solid

Colour depends both on light sources, and the colour properties of any materials that reflect the light sources. *Local colour* is an object's colour, considered independently of any light source. A local colour is determined entirely by the object's reflectance function. Only local colours, and not light sources, can be optimal. Denote a reflectance function by $r(\lambda)$, where $\lambda$ is a wavelength varying over the visible spectrum, from about 400 to 700 nm. $r(\lambda)$ is the percentage of light the object would reflect if were illuminated solely with light of wavelength $\lambda$. $r(\lambda)$ takes on a value between 0 and 1 (which equals 100%).

Colour perception also requires a light source, which we will assume has a fixed spectral power function, $s(\lambda)$. According to the trichromatic theory, human daylight colour perception derives from three kinds of light receptors, each of which responds differently to stimuli in different parts of the visible spectrum. In 1931, the Commission Internationale de l'Éclairage (CIE) standardized the trichromatic theory[3] by experimentally determining three colour-matching functions, called $\bar{x}(\lambda), \bar{y}(\lambda)$, and $\bar{z}(\lambda)$. Figure 1 plots these functions; they have been scaled so that the area under $\bar{y}(\lambda)$ integrates to 1. Though the receptors' actual response curves were unknown, they were determined to be a linear transformation of the CIE functions, which was sufficient for further analysis.

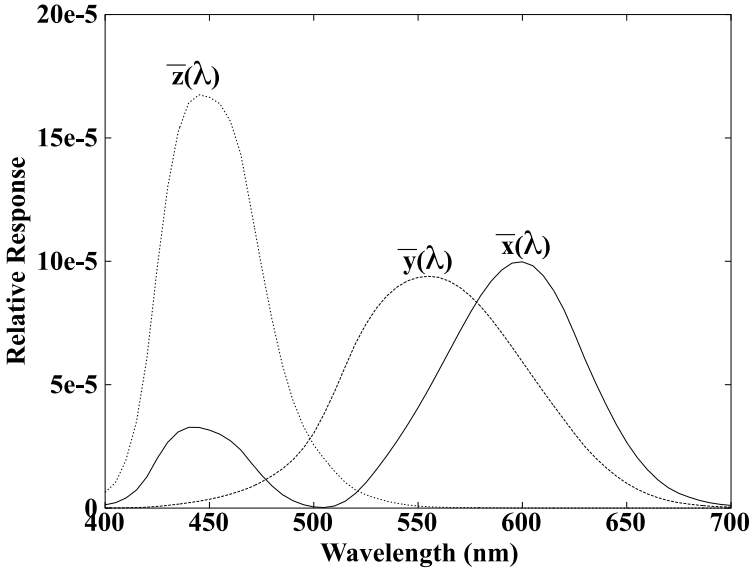



Figure 1: CIE Colour Matching Functions

Perceptually, local colours can be specified by their relative CIE coordinates:

$$X = \frac{\int_{400}^{700} \bar{x}(\lambda) r(\lambda) s(\lambda) d\lambda}{\int_{400}^{700} \bar{y}(\lambda) s(\lambda) d\lambda}, \tag{1}$$

$$Y = \frac{\int_{400}^{700} \bar{y}(\lambda) r(\lambda) s(\lambda) d\lambda}{\int_{400}^{700} \bar{y}(\lambda) s(\lambda) d\lambda}, \tag{2}$$

$$Z = \frac{\int_{400}^{700} \bar{z}(\lambda) r(\lambda) s(\lambda) d\lambda}{\int_{400}^{700} \bar{y}(\lambda) s(\lambda) d\lambda}. \tag{3}$$

The integrands in the denominator model the incoming light, $s(\lambda)$, of which only a reflected percentage, $r(\lambda)$, reaches the viewer, where it is perceptually weighted by a colour-matching function. The denominators are all the same constant, and account for human adaptation to the ambient light level. Since the colour-matching functions are all bounded, the illuminant level is fixed, and $r(\lambda)$ is between 0 and 1, it follows that $X$, $Y$, and $Z$ are also bounded. The $XYZ$ colour solid is the set of all triples, $(X(r), Y(r), Z(r))$, that result from a local colour, $r(\lambda)$. This solid can be plotted in $\mathbf{R}^3$.

It is easily seen that the $XYZ$ colour solid is convex: Let $(X(r_1), Y(r_1), Z(r_1))$ and $(X(r_2), Y(r_2), Z(r_2))$ be the $XYZ$ triples corresponding to the local colours $r_1(\lambda)$ and $r_2(\lambda)$. The $XYZ$ solid is convex if the triple $\alpha\,(X(r_1), Y(r_1), Z(r_1)) + (1-\alpha)\,(X(r_2), Y(r_2), Z(r_2))$ is also in the solid, for any $\alpha$ between 0 and 1. Equations 1 through 3 show that the local colour $r_3(\lambda) = \alpha r_1(\lambda) + (1-\alpha) r_2(\lambda)$ produces this new $XYZ$ triple. Since $r_1(\lambda)$ and $r_2(\lambda)$ are both restricted to the range $[0, 1]$, $r_3(\lambda)$ is also in that range, and is therefore a local colour.

In fact, the $XYZ$ colour solid shows considerably more structure than just convexity. This paper will show that it is a zonoid, and can be well approximated by zonohedra, for which an explicit construction will be given. The $XYZ$ colour solid depends on the illuminant, $s(\lambda)$. If the illuminant is changed, then the $XYZ$ coordinates in Equations 1 through 3 change, even when $r(\lambda)$ remains constant. Changing the illuminant, however, does not affect the proof of convexity, nor the later derivations involving zonohedra.

Chromaticity coordinates, $(x, y, Y)$, are an alternative set of colour coordinates from $XYZ$. The colour-matching functions were chosen so that a colour's $Y$-coordinate would correspond to its intensity: the total percentage of light that it reflected, weighted by human visual sensitivity. These new coordinates separate a local colour into an intensity component and a chromatic component, which is given by $x$ and $y$ :

$$x = \frac{X}{X + Y + Z}, \tag{4}$$

$$y = \frac{Y}{X+Y+Z}. \tag{5}$$

The set of all $(x, y, Y)$ corresponding to local colours, when plotted in $\mathbf{R}^3$, is called the *Rösch colour solid*.[4] Unlike the $XYZ$ solid, the Rösch solid is not convex. In the Rösch colour solid, any colour in the plane $Y = 0$ is an ideal black (that is, a local colour that reflects absolutely no light), regardless of the values of $x$ and $y$.

## 2.1 Optimal Colours

An optimal colour is a colour that is on the boundary of the $XYZ$ colour solid. This definition generalizes the original definition: an optimal colour is a local colour, $(x, y, Y)$, for which $Y$ could not be increased without changing $x$ or $y$. When seen in shadow, a local colour, $(x, y, Y)$, has chromaticity coordinates $(x, y, \tilde{Y})$, where $\tilde{Y} < Y$, but $x$ and $y$ are unchanged. An optimal colour could therefore be interpreted as a colour that could not appear as a shadow of another colour.

A shadow series consists of an optimal colour, and all the possible colours it could take on in shadow. A shadow series goes from an optimal colour $(x, y, Y_0)$, to the colour $(x, y, 0)$, which is an ideal black. In the Rösch colour solid, shadow series are lines parallel to the $Y$-axis, which is usually vertical, so an optimal colour is the maximum height a fixed chromaticity can reach. In the $XYZ$ solid, a shadow series is a straight line that originates at ideal black, whose $XYZ$ coordinates are $(0, 0, 0)$. The shadow series stays inside the colour solid until it intersects the solid's boundary, at an optimal colour.

Optimal colours also have a natural interpretation in terms of the Munsell system. The Munsell system classifies local colours by hue (red, green, yellow, etc.), value (how light or dark the colour is), and chroma (how saturated the colour is; the less dull a colour is, the more saturated it is). For a fixed hue and value, the chromas extend from 0, corresponding to the grey of that value, to a maximum, called the MacAdam limit. A colour at a MacAdam limit is an optimal colour. The MacAdam limits are theoretical bounds; in practice, the Munsell system extends only as far as actual colorants allow, stopping well short of the MacAdam limits.

An important result characterizes optimal colours in terms of their reflectance functions:

**Optimal Colour Theorem.** *A local colour is optimal if and only if the values of its reflectance function are either 0 or 1, with at most two transitions between those two values.*

Wilhelm Ostwald[5] first noted this result for some special cases, in 1916. In 1920, Erwin Schrödinger[6] proved that the reflectance values of an optimal colour are either

0 or 1, and stated without proof that there were at most two transitions. Finally, David MacAdam[1] proved the complete result in 1935. The zonohedral constructions to be presented in this paper will give an alternate proof of the Optimal Colour Theorem, that is simpler and more direct.

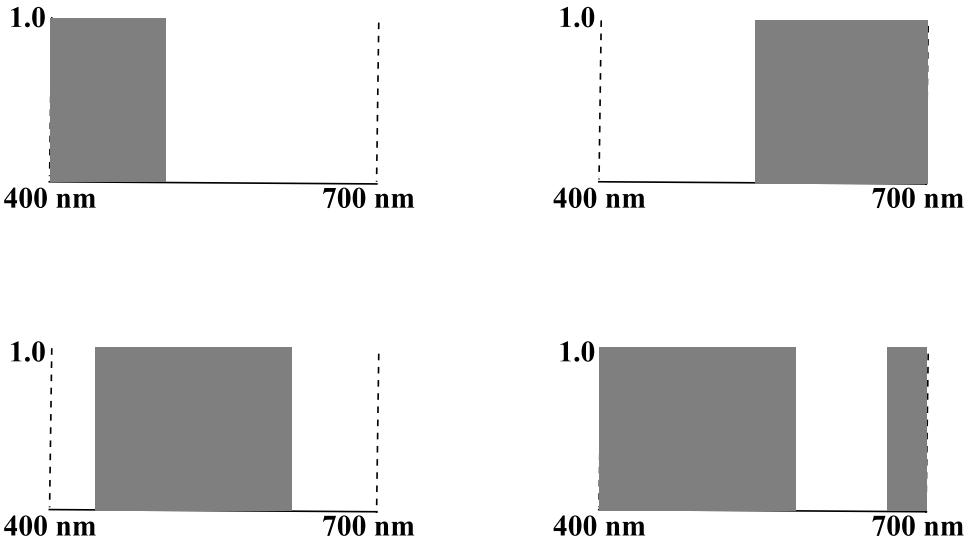



Figure 2: Four Forms for Optimal Colour Functions

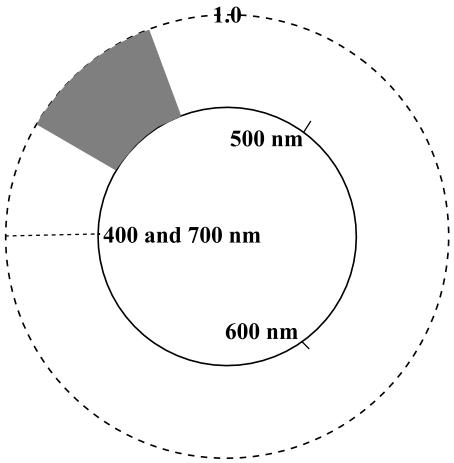



Figure 3: Circular Form for Optimal Colour Functions

Schrödinger[6,7] classified the optimal colour functions into four forms, shown in Figure 2, depending on the number and locations of transition wavelengths. The

first form, in the upper left, has one transition, and takes on the value 1 between 400 nm and that transition. The second form, in the upper right, also has one transition, but takes on the value 1 to the right of that transition. The third form, in the lower left, has two transitions, between which it is 1, and outside which it is 0. The fourth form also has two transitions, but is 0 between them, and 1 outside.

These four forms can be subsumed into one form, if, following Newton, the visible spectrum is wrapped around to make a circle, and the ends, at 400 and 700 nm, are joined. On the circle, the reflectance function for any optimal colour fills out a sector of an annulus, with one transition wavelength where the sector starts, and another where the sector ends. In the first form, the sector starts at 400 nm (which is also 700 nm). In the second form, the sector ends at 400 (or 700) nm. In the third form, the sector does not cross the 400 nm mark. In the fourth form, the sector starts before 700 nm, continues clockwise across 700 nm, which is also 400 nm, and then fills out some of the clockwise side of the 400 nm mark. The zonohedral constructions used later will provide another interpretation for the circular form.

# 3 Zonohedra

The mathematical constructions in this section will be used later to construct the $XYZ$ colour solid in the form of a zonohedron.

## 3.1 Minkowski Sums

The Minkowski sum (also called the vector sum) of two convex sets, $\mathbf{A}$ and $\mathbf{B}$, in $\mathbf{R}^n$, is defined as

$$\mathbf{A} \oplus \mathbf{B} = \{a + b | a \in \mathbf{A}, b \in \mathbf{B}\}. \tag{6}$$

The Minkowski sum of two convex sets is itself convex.[8] The Minkowski sum can be naturally generalized to include any number of convex sets. Up to translation, Minkowski addition is commutative: $\mathbf{A} \oplus \mathbf{B}$ is the same shape as $\mathbf{B} \oplus \mathbf{A}$, although possibly in a different location. It is also associative: $(\mathbf{A} \oplus \mathbf{B}) \oplus \mathbf{C}$ and $\mathbf{A} \oplus (\mathbf{B} \oplus \mathbf{C})$ are the same shape, so that we can unambiguously write $\mathbf{A} \oplus \mathbf{B} \oplus \mathbf{C}$.

Geometrically, the Minkowski sum of $\mathbf{A}$ and $\mathbf{B}$ is the shape covered by all the copies of $\mathbf{A}$ whose centers touch $\mathbf{B}$ at at least one point. To illustrate, Figure 4 shows the Minkowski sum of an ellipse $\mathbf{A}$ and a quadrilateral $\mathbf{B}$, in the same plane. A copy of the ellipse can be centered on any interior point of the quadrilateral, so the Minkowski sum contains all of $\mathbf{B}$. In addition, the ellipse can be placed so that

its center is on a boundary point of the quadrilateral. The dotted ellipses show some copies of **A** that just touch the quadrilateral's corners. The ellipse can also be slid along each side of the quadrilateral, producing a line parallel to that side of the quadrilateral.

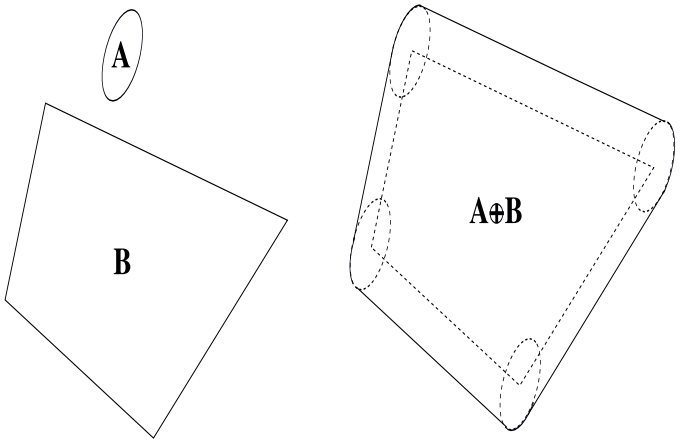



Figure 4: Minkowski Sum of an Ellipse and a Quadrilateral

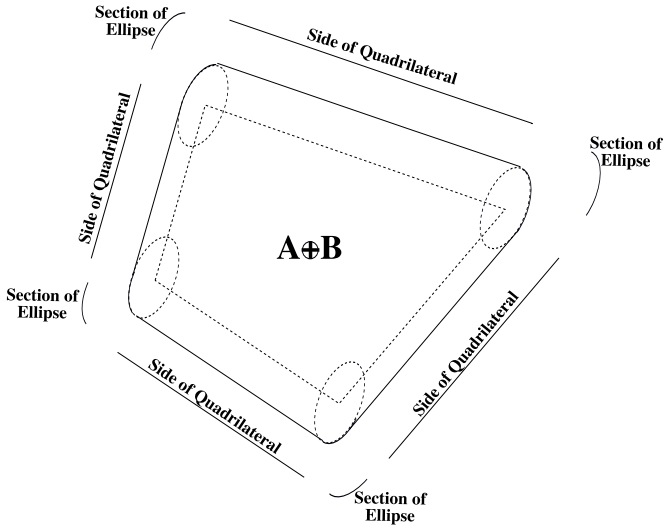



Figure 5: Boundary of Minkowski Sum, and Boundary Segments of Summands

In fact, the Minkowski sum can be visualized as the shape produced by sliding **A** along the boundary of **B**, without letting **A** roll. The figure outlines $\mathbf{A} \oplus \mathbf{B}$

with a thick line. Figure 5 shows how the boundary of $\mathbf{A} \oplus \mathbf{B}$ can be broken into segments that are translations of the sides of the quadrilateral, and translations of four sections of the ellipse. The union of the segments contains the entire perimeters of the quadrilateral and ellipse.

Although this example is in two dimensions, Minkowski sums are also defined for three dimensions, and indeed for an arbitrary number of dimensions. Also, a Minkowski sum can have arbitrarily many summands, rather than just two.

## 3.2 Construction of Zonohedra

A special case of Minkowski summation occurs when the summands are all line segments, or vectors, starting at the origin. Line segments are always convex sets, so their Minkowski sum is also a convex set, called a *zonotope.* In three dimensions, a zonotope is called a *zonohedron.* Each face of a zonohedron is a centrally symmetric, convex polygon. If no three generating vectors are coplanar, then every face is a parallelogram. A configuration with three coplanar vectors is unstable, because moving any one of them by the slightest amount can destroy the coplanarity. Generically, then, any zonohedron can be approximated arbitrarily closely by a zonohedron with only parallelograms for faces. Alternately, if a zonohedron's face is not a parallelogram, it can be naturally tessellated into parallelograms, and we can think of the vertices of the tessellating parallelograms as vertices of the zonohedron.

Figure 6 shows an example. On the left, three generating vectors of different lengths and directions, labeled $\mathbf{v_1}$, $\mathbf{v_2}$, and $\mathbf{v_3}$, start at the origin, labeled $\mathbf{0}$. The zonohedron they generate is a parallelepiped, all of whose faces are parallelograms. The vertices of the parallelepiped are sums of the generating vectors. The farthest vector from the origin is the sum of all the generators. On the right, a fourth generator, $\mathbf{v_4}$, has been added. The parallelepiped on the right fills the space swept out by moving the parallelepiped on the left, along $\mathbf{v_4}$. Again, all the vertices are sums of generators, and the sum of all the generators is the vertex farthest from the origin.

Not every combination of generators appears as a vertex. The vertices of a zonohedron can be characterized as follows.[9] Draw an arbitrary plane through the origin, and find the generating vectors that lie on one side of that plane: the sum of those vectors is a vertex, and every vertex can be found by this method. An important special case, which will be of interest later, occurs when the generating vectors have no "interior" vectors in the following sense. Construct all the rays that begin at the origin and lie along generating vectors. The convex hull of the rays is called a convex cone. If each ray is an edge of the convex cone, then the generating

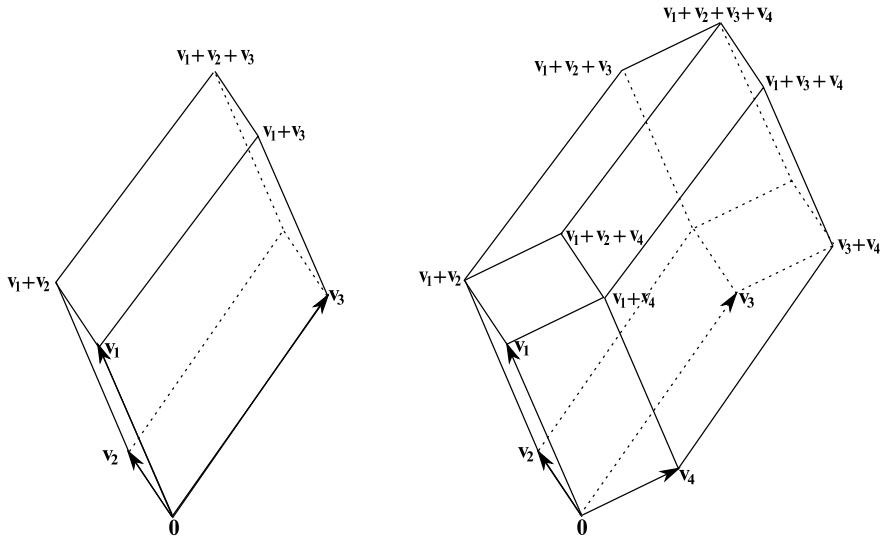



Figure 6: Construction of Zonohedra

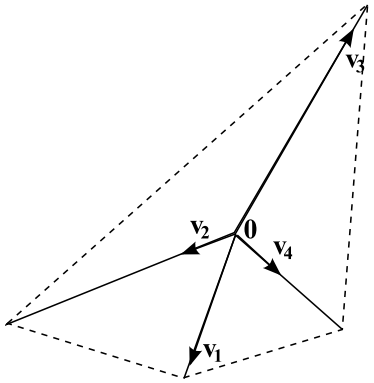



Figure 7: Four Vectors in Cyclic Position

vectors will be said to be in *cyclic position.*

The four vectors in Figure 6 are in cyclic position. Figure 7 shows the vectors seen from above, and extended to rays. A plane perpendicular to the vertical axis has been drawn, and the generators' intersections with the plane are shown, connected by a dashed line. The intersections form a convex polygon: none of the intersection points is inside the triangle generated by the other three intersection points. Further generators could be added; as long as none is inside the convex hull of the other generators, the entire set will be in cyclic position.

Figure 6 suggests a simple means of calculating a zonohedron's vertices, when the

generators are in cyclic position. The vertices adjacent to the origin are simply the generating vectors; think of these as forming the first level of vertices. The second level of vertices is above and adjacent to the first level, so they consist of sums of two generators. These levels continue until the farthest vertex, which is the sum of all the generators.

The following matrix method formalizes this approach to calculating the zonohedron's vertices. In the first row of the matrix, list the zero vector $n$ times. In the second row of the matrix, list the $n$ generators, in sequence around the loop they form. The matrix's third row consists of sums of two vectors. Entry $(3, j)$, for $j = 1 \ldots n$, is the sum of entry $(2, j)$ and the vector immediately adjacent to the vector in $(2, j)$, in the clockwise direction. The adjacent vectors should be selected modularly: the first vector comes immediately after the $n$th vector. The fourth row consists of sums of three vectors. Entry $(4, j)$, for $j = 1 \ldots n$, is the sum of entry $(3, j)$, which is already the sum of two vectors, and the vector adjacent to those two vectors, in the clockwise direction. Continue this process until the $(n+1)^{\text{st}}$ row, which will contain $n$ copies of the sum of all the vectors. Expression (7) shows the matrix when $n$ is 4:

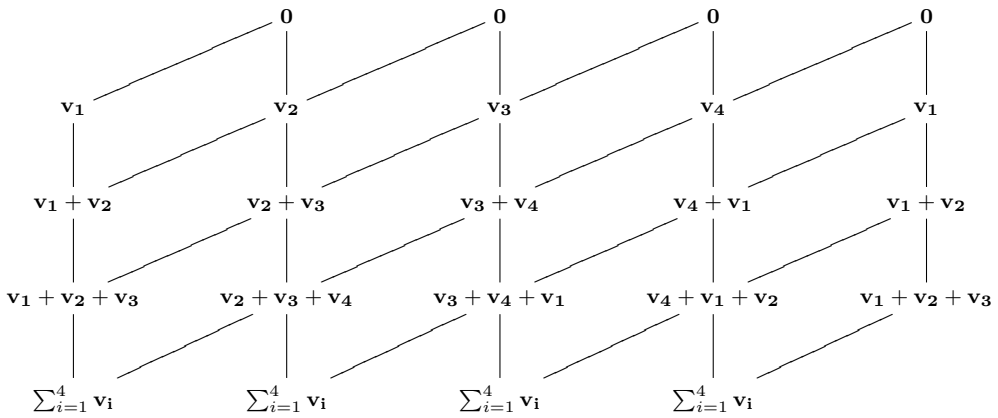
$$\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{v_1} & \mathbf{v_2} & \mathbf{v_3} & \mathbf{v_4} \\ \mathbf{v_1} + \mathbf{v_2} & \mathbf{v_2} + \mathbf{v_3} & \mathbf{v_3} + \mathbf{v_4} & \mathbf{v_4} + \mathbf{v_1} \\ \mathbf{v_1} + \mathbf{v_2} + \mathbf{v_3} & \mathbf{v_2} + \mathbf{v_3} + \mathbf{v_4} & \mathbf{v_3} + \mathbf{v_4} + \mathbf{v_1} & \mathbf{v_4} + \mathbf{v_1} + \mathbf{v_2} \\ \sum_{i=1}^{4} \mathbf{v_i} & \sum_{i=1}^{4} \mathbf{v_i} & \sum_{i=1}^{4} \mathbf{v_i} & \sum_{i=1}^{4} \mathbf{v_i} \end{bmatrix} \tag{7}$$

Every entry in Expression (7) is a vertex of the zonohedron, although with some duplication: the entries in the first row are all the origin, and the entries in the last row are all the farthest vertex. In all, there are $n^2 - n + 2$ vertices. Each vertex consists of adjacent vectors in the cyclic loop. There are two adjacent vectors in the third row, three in the fourth row, and so on. This is just the form predicted earlier.[9] Imagine a plane drawn through the origin, dividing the generators into two groups. In Figure 7, that plane would cut the picture plane in a straight line, which would divide the intersection points into two groups—the two groups of intersection points correspond to the two groups of generators. When the generators are in cyclic position, the intersection points are the vertices of a convex polygon. When cut by a straight line, the vertices of a convex polygon are divided into two consecutive segments: these consecutive segments correspond to the entries of Expression (7).

Not only can the zonohedron's vertex structure be inferred from the matrix constructed, but so can its face structure. Two vertices are adjacent, i.e. joined by an

edge in the zonohedron, if and only if their difference is a generating vector. Expression (8) joins the vertices in Expression (7) that are adjacent. The first column has been repeated at the right side of the matrix, in order to show all the adjacencies. Some edges appear twice in Expression (8). For example $\mathbf{v_2}$, in entry $(2, 2)$ is joined to $\mathbf{0}$ by two lines. Since all the entries in the first row represent the same point, the origin, the two lines are actually one line in the zonohedron. Similar comments apply to the last row. Visually, Expression (8) displays $n(n-1)$ parallelograms. Each parallelogram is a face on the zonohedron, and the parallelogram's vertices and edges on the zonohedron are as shown in Expression (8).




(8)

*Zonoids* are a useful generation of zonotopes. A zonoid is the limit of a convergent sequence of zonotopes. Convergence in this context is with respect to the Hausdorff metric.[10] Since the limit of a convergent sequence of convex sets must itself be convex, therefore zonoids are convex. It is possible, however, for a zonoid to be smooth, even though every zonotope in the convergence sequence has vertices. Daoudi et al.[11] discuss another approach, that will be used later, in which a zonoid is constructed from a curve in space.

# 4 The $XYZ$ Colour Solid as a Zonohedron

Section 2 described the calculations behind the $XYZ$ colour solid, and Section 3 described some of the mathematics of zonohedra. The current section reveals a hidden zonohedral structure in the $XYZ$ colour solid calculations, and uses that structure in an intuitive proof of the Optimal Colour Theorem.

## 4.1 Generating Vectors for the $XYZ$ Colour Solid

Rather than using a continuous reflectance function, $r(\lambda)$, approximate $r(\lambda)$ by dividing the visible spectrum into $n$ channels of equal width, and selecting $n$ point masses, at wavelengths $\lambda_i, i = 1 \ldots n$, each of which is at the center of a channel. For example, $n$ could be 31, each channel could be 10 nm wide, the first channel could be centered on 400 nm, the second on 410 nm, and the 31st on 700 nm. Then the integrals in Equations (1) through (3) are approximated by finite sums:

$$X = k \sum_{i=1}^{n} \bar{x}(\lambda_i) r(\lambda_i) s(\lambda_i), \tag{9}$$

$$Y = k \sum_{i=1}^{n} \bar{y}(\lambda_i) r(\lambda_i) s(\lambda_i), \tag{10}$$

$$Z = k \sum_{i=1}^{n} \bar{z}(\lambda_i) r(\lambda_i) s(\lambda_i). \tag{11}$$

The value of $k$ in these equations is another finite sum:

$$k = \frac{1}{\sum_{i=1}^{n} \bar{y}(\lambda_i) s(\lambda_i)}. \tag{12}$$

Express Equations (9) through (11) more concisely as a vector in $\mathbf{R^3}$, using the notation $\mathbf{S_i} = ks(\lambda_i)\left(\bar{x}(\lambda_i), \bar{y}(\lambda_i), \bar{y}(\lambda_i)\right)$ :

$$(X, Y, Z) = \sum_{i=1}^{n} r(\lambda_i) \mathbf{S_i}. \tag{13}$$

Equation (13) expresses the relative CIE coordinates of $r(\lambda)$ as a sum of $n$ different vectors, where all the coefficients in the sum are between 0 and 1. Think of each vector $\mathbf{S_i}$ as a line segment in $\mathbf{R^3}$. Then, geometrically, Equation (13) is the vector sum of a finite set of line segments. Since the $XYZ$ coordinates for any local colour can be written in the form given in Equation (13), the $XYZ$ colour solid as a whole can be approximated by the Minkowski sum of the vectors $\mathbf{S_i}$, making a zonohedron.

Each choice of $n$ gives a different zonohedron. As $n$ goes to $\infty$, the zonohedra approach the smooth colour solid more and more closely. The Hausdorff limit of a convergent set of zonohedra is a zonoid, so the $XYZ$ colour solid is a zonoid, which can be approximated arbitrarily closely by zonohedra constructed in accordance with Equation (13). Calculations involving the $XYZ$ colour solid can be performed with this zonohedron, with $n$ as large as computational power permits.

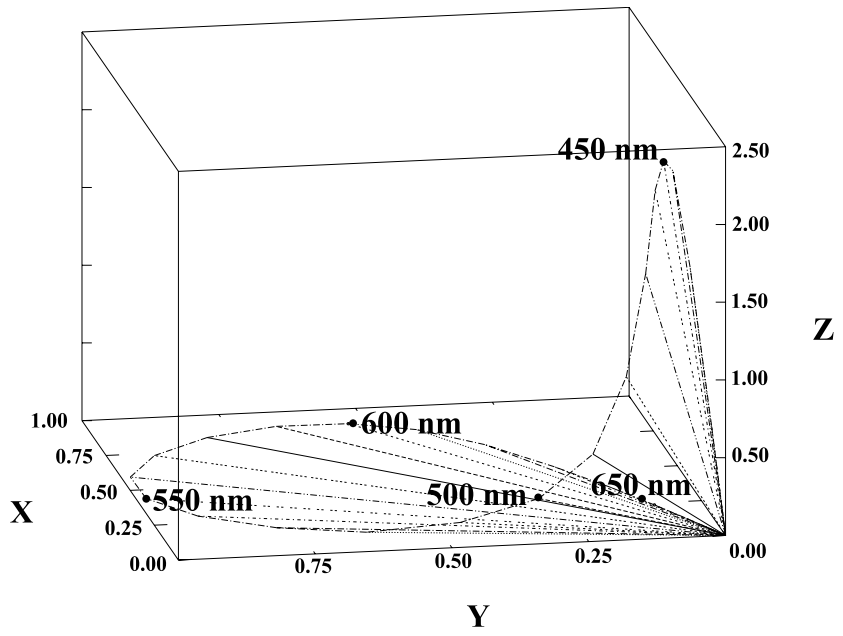



Figure 8: Generating Vectors for 10 nm Channels, Illuminant C

Geometrically, for each $n$, the vectors $\mathbf{S_i}$ are a finite set of generating vectors for the zonohedron. Figure 8 shows the generating vectors when $n$ is 31, and the light source is illuminant C. At 400 and 700 nm, both generators are nearly zero, because the three colour-matching functions shown in Figure 1 all nearly vanish at these wavelengths. When the wavelength is greater than 550 nm, the function $\bar{z}(\lambda)$ is practically zero. As a result, generating vectors for wavelengths greater than 550 nm lie in the horizontal plane, $Z = 0$. This effect can be seen in Figure 8, where the vectors near 400 nm are high above the base plane, and then descend towards it, and are restricted to it after 550 nm.

As $n$ increases to $\infty$, the generating vectors shrink to zero length. Suppose, for example, that the channels were 5 nm wide, giving 61 vectors instead of 31. Each channel, being only half as wide, would reflect approximately half as much of the illuminant, so that $X$, $Y$, and $Z$ would shrink to half their values. The scale of the colour solid would not change, however, because the vertices in Expression (7) could be the sums of twice as many generating vectors. The relative values of $X$, $Y$, and $Z$, are therefore more important than the absolute values indicated in Figure 8.

Figure 9 shows the $XYZ$ colour solid, as the zonohedron generated by the 31 vectors in Figure 8. As expected, each face of Figure 9 is a parallelogram, or is tessellated into parallelograms. The top and bottom of the solid are flat planes.

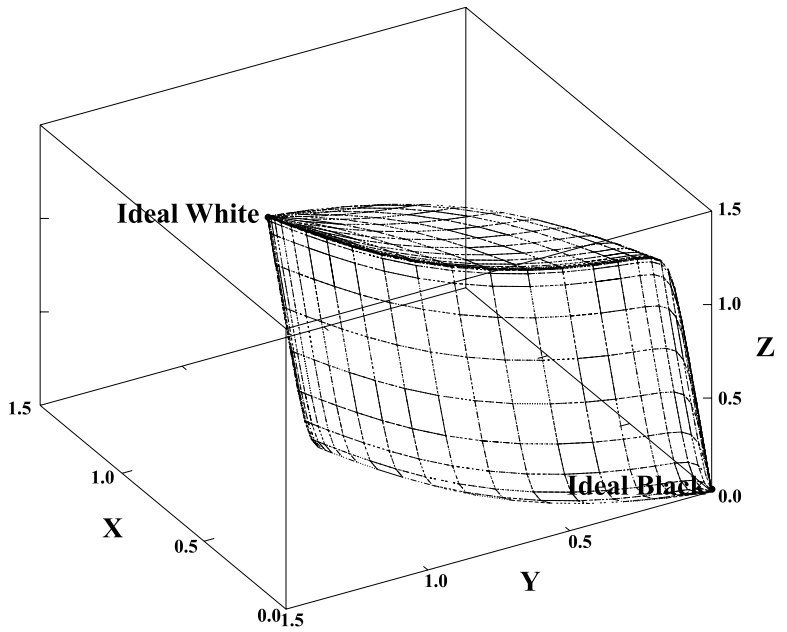



Figure 9: $XYZ$ Colour Solid, as a Zonohedron (10 nm Channels, Illuminant C)

These correspond to the wavelengths greater than 550 nm, where the $Z$-component is negligible. The vertex at the origin represents ideal black, a local colour whose reflectance function is identically 0. The farthest vertex represents ideal white, a local colour whose reflectance function is identically 1. The colours on the boundary of the solid are optimal colours. The non-optimal colours fill the interior of the solid. The maximum $(X, Y, Z)$ values are (0.98, 1.00, 1.18). The maximum $Y$-value is exactly 1 because of the normalization in Equation 2.

## 4.2 A Constructive Proof of the Optimal Colour Theorem

Zonohedral constructions can be used to prove the Optimal Colour Theorem, more simply and directly than the Schrödinger-MacAdam proof. There are two major components in the proof, one mathematical and the other empirical.

The mathematical component comes from the characterization of the vertices in a zonohedron. Optimal colours are colours on the boundary of the colour solid, so the vertices, which are all on the boundary, are taken to be good approximations to optimal colours. Expression (7) lists the solid's vertices, for $n = 4$. They are all sums of a consecutive set of vectors: $\mathbf{v_i} + \mathbf{v_{i+1}} + \mathbf{v_{i+2}} + \ldots \mathbf{v_j}$, allowing for wrapping around to the first vector once the $n$th vector is reached. If the vectors $\mathbf{v_i}$ are the

generating vectors $\mathbf{S_i}$, then the vector sums in Equation (13) would all have $r(\lambda_i)$ set to 1 for a consecutive set of vectors (allowing for wraparound), and set to 0 for any vectors not in that consecutive set. Since the generating vectors $\mathbf{S_i}$ are the $(X, Y, Z)$ triples for channels, $r(\lambda)$ is 1 on a consecutive set of channels, and 0 elsewhere. When wraparound is accounted for, $r(\lambda)$ for an optimal colour must be of the form shown in Figure 3, which can be reduced to one of the four forms in Figure 2. This conclusion is just the substance of the Optimal Colour Theorem.

The empirical component comes from the experimentally determined shape of the curve in Figure 8. The mathematical component of the proof depends on the zonohedral constructions in Expression (7), which are only valid when the generating vectors are in cyclic position. Figure 8 shows that consecutive channels produce a sequence of generating vectors with no "interior" vectors, and that this sequence maps bijectively into an order-producing loop formation: as one goes left to right, from channel to channel in the visible spectrum, the corresponding generating vectors always move clockwise around the cyclic loop. To borrow a phrase from West and Brill,[12] the curve in Figure 8 is "well-ordered in wavelength." The empirical properties on which the Optimal Colour Theorem rests cannot be demonstrated mathematically. Rather, they depend on the colour-matching functions, which in turn depend on the biochemistry of human retinal receptors, so their properties must be found by experiment.

# 5 Zonohedra in Calculations

Not only do zonohedra simplify colour space, but they can also reduce computational complexity. An example is the original optimal colour problem: given chromaticity coordinates $x$ and $y$, find the maximum $Y_0$ for which $(x, y, Y_0)$ can exist as a local colour. To determine $Y_0$, divide the visible spectrum into $n$ channels, and then use Expression (7) to calculate zonohedral vertices in $XYZ$ space. Each vertex is itself an optimal colour, but probably none has the right $xy$ chromaticity coordinates. The maximum $Y_0$ will occur where the shadow series of $(x, y)$ intersects the boundary of the zonohedron. The shadow series is a straight line, starting at the black point $(x, y, 0)$, and containing all points of the form $(x, yY)$. A straight line, such as a shadow series, intersects the boundary of a convex solid, such as the $XYZ$ zonohedron, in at most two points. One intersection is the black point, so we are looking for the second point, which occurs for some positive $Y$.

Use the bisection method to determine where the shadow series is inside the colour solid, and where it is outside; the dividing point is the optimal colour. Start the bisection with the points $Y = 0$ and $Y = 1.5$. The point $(x, y, 0)$ is inside the solid

(or, rather, on its boundary), while the point $(x, y, 1.5)$ must be outside, because the maximum $Y$-value is 1. The midpoint of the two initial $Y$-values is $Y = 0.75$. Convert $(x, y, 0.75)$ to $XYZ$ coordinates, and determine whether it is inside or outside the solid. At this step we use the fact that a zonohedron is the convex hull of its vertices, which we have already found. Graphics algorithms such as QuickHull can rapidly determine whether a point is inside the convex hull of a given set of vertices. If $(x, y, 0.75)$ is inside, then the optimal $Y_0$ must be between 0.75 and 1.5. If $(x, y, 0.75)$ is outside, then the optimal $Y_0$ must be between 0 and 0.75. Either way, we choose the midpoint of the appropriate interval and continue bisecting. Since every bisection halves the interval containing $Y_0$, and since $2^{10} \doteq 10^3$, it takes ten bisection steps, so ten calls to QuickHull, to determine $Y_0$ to three significant digits.

Bisection works not just for straight lines, but for any parametrized curve that intersects the colour solid. For example, one can determine the maximum Munsell chroma for a given hue and value, such as 5P7/. The colours 5P7/0, 5P7/1, 5P7/2, etc., form a parametrized curve, when converted into $XYZ$ space by the Munsell renotation. The curve starts inside the colour solid, at a Munsell grey. It extends outward until it crosses the colour solid boundary at its maximum chroma. If extrapolated renotation values are used, then one could start with a chroma, such as 40, for which 5P7/40 is definitely outside the solid. As before, perform a bisection on the curve, converging on the maximum chroma.

Calculating optimal colours customarily involves calculating $XYZ$ coordinates for reflectance functions corresponding to different pairs of transition wavelengths. Typically, optimal colour tables, such as Table I(3.7) & Table II(3.7) on pp. 776-779 of Wyszecki and Stiles,[4] list transition wavelengths along with $Y_0$. The zonohedral approach, on the other hand, does not require finding transition wavelengths.

# 6 Relation to Previous Work

## 6.1 The Schrödinger-MacAdam Proof

In 1920, Erwin Schrödinger[6] introduced the methods used to prove the Optimal Colour Theorem; in 1935, David MacAdam[1] used those methods to produce a complete, rigorous proof. Both researchers worked in terms of chromaticity diagrams. To construct the chromaticity diagram, extend the vectors in the convex hull of the spectrum locus in Figure 8, until they intersect the plane $X + Y + Z = 1$, and then calculate $xyY$ chromaticity coordinates for the intersection points. Plot the intersection points by treating $x$ and $y$ as Cartesian coordinates. The visible spectrum wraps around of the boundary of a chromaticity diagram, excepting one non-spectral

*purple line* that joins the 400 nm point to the 700 nm point.

Schrödinger and MacAdam both used the convexity of the chromaticity diagram, and the fact that a "mixture" of two colours in the chromaticity diagram lies on the line joining those two colours. A mixture in this sense corresponds to adding an $r_1(\lambda)$ and an $r_2(\lambda)$ to get a new reflectance function, which is valid as long as the sum is nowhere greater than 1. MacAdam took a further step, expressing the chromaticity of a reflectance function as the center of gravity of that function's image in the chromaticity diagram.

The current proof improves on the Schrödinger-MacAdam proof in three ways. First, by using the three-dimensional colour solid directly, it avoids many of the difficulties in interpreting chromaticity diagrams. Second, it eliminates metric notions, such as the center of gravity, that are extraneous to colour perception. Third, it exposes the linear structure of local colours, so that one can see how optimal colours are constructed.

## 6.2 Zonoids Associated with Parametric Curves

The zonoidal colour solid is an example of a general construction, discussed in a 1994 paper of Daoudi et al.,[11] of a zonoid from a parametric curve. To construct the curve, let $n$ go to $\infty$ in Figure 8. With an appropriate rescaling, the heads of the generating vectors converge to a parametric curve, $F(\lambda)$, in $\mathbf{R}^3$ :

$$F(\lambda) = \frac{s(\lambda)}{\int_{400}^{700} \bar{y}(\lambda)s(\lambda)d\lambda} \left(\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)\right). \tag{14}$$

Following Def. 3 of the 1994 paper, the set $Z(F)$ defined by

$$\left\{ v \in \mathbf{R}^3 \mid v = \int_{400}^{700} r(\lambda)F(\lambda)d\lambda \text{ for some function } r : [400, 700] \to [0, 1] \right\}, \tag{15}$$

is a zonoid. In the current context, the set of all functions $r$ from [400, 700] to [0,1] is the set of reflectance functions, so $Z(F)$ is the $XYZ$ colour solid.

Prop. 3 of Daoudi's paper expresses the zonoid's boundary points, which in our case are optimal colours, in terms of reflectance functions: $\mathbf{u}$ is the outward normal to the zonoid at the boundary point given by

$$\int_{\omega} F(\lambda)d\lambda, \tag{16}$$

where

$$\omega = \{\lambda \in [400, 700] \mid \langle F(\lambda), \mathbf{u}\rangle \geq 0\}. \tag{17}$$

A generating vector $F(\lambda)$ satisfies the inequality in Equation (17) if and only if it is on the positive side of the plane through the origin that is perpendicular to $\mathbf{u}$. The integral in Expression (16) is the limit of the sum of all generating vectors on one side of a plane, so it is the limit of a sequence of vertices on zonohedra. Rather than being a discrete set of vectors, however, the set $\omega$ is composed of one or more intervals of $F(\lambda)$. From the form of the curve in Figure 8, each $\omega$ is composed of exactly one interval, if wraparound is accounted for. These remarks extend the zonohedral characterization of optimal colours (as the sum of discrete, consecutive channels), showing that reflectance functions that are 1 on an annular interval, and 0 elsewhere, give optimal colours for the $XYZ$ solid when viewed as a zonoid.

Historically, Goethe's edge colours led to similar constructions. Along these lines, Equation (5) in a recent paper of Koenderink[13] is the integral of the parametric curve $F(\lambda)$, defined in Equation 14. As he remarks, this integral, which is itself a curve parametrized by $\lambda$, forms a spiral in $\mathbf{R}^3$, going from ideal black to ideal white. From the development in this paper, it can be seen that this spiral contains only optimal colours, so it must be restricted to the surface of the colour solid. Both $F(\lambda)$ and its integral in Koenderink's Equation (5) contain enough information to construct the complete colour solid.

## 6.3 West and Brill's Conditions for Optimal Colours

In 1983, West and Brill[12] investigated mathematical conditions under which reflectance functions of the forms shown in Figure 2 are not optimal. For human vision, of course, such forms always give optimal colours, but Fig. 3 of their 1983 paper shows data for bees, which implies that those forms do not give optimal colours for all species. The authors follow MacAdam's approach, to find a necessary condition for the forms in Figure 2 always to give optimal colours: the spectrum locus must be a convex curve in the chromaticity diagram. The current paper generalizes West and Brill's condition to a necessary and sufficient condition in three dimensions. The spectrum locus is now a curve in space, such as appears in Figure 8. If the locus is in cyclic position, i.e. no vector from the origin to the curve is contained in the interior of the convex cone generated by the curve, then the boundary points of the associated zonohedron are the sums over connected subsets of the locus. The connected subsets are of the forms in Figure 2, so the Optimal Colour Theorem holds.

West and Brill also presented a construction for optimality for arbitrary colour-matching functions: optimal reflectance functions are those whose value is 1 for wavelengths on one side of a line dividing the spectrum locus in the chromaticity

diagram, and 0 on the other side. The current paper gives a similar construction, where the spectrum locus is a set of generating vectors in $\mathbf{R}^3$. Let any plane through the origin divide the spectrum locus into two sets: optimal reflectance functions are those whose value is 1 for wavelengths on one side of the plane, and 0 on the other side.

# 7 Summary

The main insight of this paper is the zonohedral structure of the $XYZ$ colour solid (Figure 9). The zonohedron for a particular illuminant is generated by the vectors in the spectrum locus for that illuminant (Figure 8). The head of each vector is the $XYZ$ coordinates for a reflectance function which is 1 on a narrow interval of the visible spectrum, and 0 elsewhere. Expression (8) gives a simple, easily implemented computation for the faces and vertices of the zonohedron.

Optimal colours are those that appear on the boundary of the colour solid. From the solid's zonohedral structure, and the fact that the vectors of the spectrum locus are in cyclic position (for standard illuminants), it follows that the reflectance functions of optimal colours take on only the values 0 and 1, with at most two transitions. This is the content of the already-known Optimal Colour Theorem, for which the zonohedral approach provides a constructive, geometric approach. An equivalent statement of the theorem is that, if the ends of the visible spectrum are joined together to make a circle, then an optimal reflectance function is 1 on an annular interval over that circle, and 0 elsewhere.

The zonohedral structure also leads to simple algorithms for optimal colour calculations, which do not involve finding transition wavelengths. In general, the zonohedral approach to the colour solid complements the spectral domain approach. Both approaches eliminate the misleading features of chromaticity diagrams. By working with colour space explicitly in three dimensions, the zonohedral interpretation clarifies ideas of convexity, and colour "mixtures." It is hoped that zonohedra can be a useful tool in further problems involving colour combinations and constructions.

1. David L. MacAdam. "The Theory of the Maximum Visual Efficiency of Colored Materials," JOSA, Vol. 25, 1935, pp. 249-252.
2. Jan J. Koenderink. *Color for the Sciences,* MIT Press, 2010.
3. Mark D. Fairchild. *Color Appearance Models,* $2^{nd}$ *ed.,* John Wiley & Sons, Ltd., 2005, Sect. 3.6.
4. Günter Wyszecki & W. S. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulae,* $2^{nd}$ *ed.*, John Wiley & Sons, 1982.

5. Wilhelm Ostwald. "Neue Forschungen zur Farbenlehre," Physikalische Zeitschrift XVII, 1916, pp. 322-332. In 2010, Rolf G. Kuehni prepared an English translation, "New researches in color science," which includes a technical introduction by Michael H. Brill and Rolf G. Kuehni.
6. Erwin Schrödinger. "Theorie der Pigmente von grösster Leuchtkraft," Annalen der Physik 4, 62, 1920, pp. 603-622. In 2010, Rolf G. Kuehni prepared an English translation, "Theorie [sic] of pigments of greatest lightness," which includes a technical introduction by Michael H. Brill.
7. David L. MacAdam. *Color Measurement: Theme and Variations,* Springer Series in Optical Science, Vol. 27, Springer-Verlag, 1981, Sect. 7.5.
8. Russell V. Benson. *Euclidean Geometry and Convexity,* McGraw-Hill, Inc., 1966, Theorem 14.4.
9. H. S. M. Coxeter. *Regular Polytopes,* $3^{rd}$ *ed.,* Dover Publications, 1963, p. 28.
10. Steven R. Lay. *Convex Sets and Their Applications,* Dover Publications, 2007, Sect. 14.
11. O. Daoudi, B. Lacolle, N. Szafran, & P. Valentin. "Zonoidal Surfaces," pp. 113-120 in *Curves and Surfaces in Geometric Design,* eds. Pierre-Jean Laurent, Alain Mehauté, & Larry L. Schumaker, A. K. Peters, 1994.
12. Gerhard West & Michael H. Brill. "Conditions under which Schrödinger object colors are optimal," JOSA, Vol. 73, No. 9, September 1983, pp. 1223-1225.
13. Jan J. Koenderink. "The prior statistics of object colors," JOSA, Vol. 27, No. 2, February 2010, pp. 206-217.