Tutorial

Julius Muschaweck* and Henning Rehn

Illumination design patterns for homogenization and color mixing

https://doi.org/10.1515/aot-2018-0051
Received October 1, 2018; accepted December 7, 2018; previously published online January 11, 2019

Abstract: In illumination optics, color mixing is a key design task, but the realization can be a challenge. While tunable light sources based on multiple LEDs are commonplace, color homogenization is just as important for white LEDs, due to their spatial and angular color variation. In this tutorial, we first look at color mixing from an abstract, phase space-based viewpoint. From there, we derive a taxonomy of color mixing problems: How is the multi-color light source composed? What kind of homogeneity is required in the target? How is the homogenization influenced by source and target étendue? We categorize these problems and we present a toolbox: A selection of optical design elements, e.g. mixing rods and fly's eye arrays, and we show for each design pattern how it fits into the taxonomy.

Keywords: color mixing; illumination; LEDs; optical design.

PACS: 42 (Optics).

1 Introduction

Many lighting systems utilize light sources which do not emit homogeneous color and/or 'brightness' by themselves. For tunable LED lamps using individual red, green, and blue LED chips, this is obvious. However, the filaments of incandescent lamps are of helical structure, high intensity discharge (HID) lamps are volume sources with a high luminance core surrounded by a lower luminance halo. So-called white LEDs, which use a yellow phosphor on a blue emitting LED in many different shapes and ways, are far from homogeneous: Thin film, surface emitting LEDs with phosphor platelets show color variation over angle, and multi chip white light engines look like shiny blue pearls in yellow soup. Inhomogeneous light sources are everywhere, and the often used idealized Lambertian source is just that: an idealization.

What is an acceptably homogeneous light output? The answer varies with application. What may be acceptable in a cheap consumer flash light is unthinkable in a professional TV studio lamp head. However, we need to be more precise in defining what is meant by homogeneity: From an optical design standpoint, the most important distinction is dimensional. A video projector, for example, is expected to homogeneously illuminate the rectangular screen. All that matters here is color and brightness as a function of the x-y location on the screen. This is what we might call two-dimensional (2D) homogeneity. For a TV studio lamp head, however, 2D homogeneity is necessary, but not sufficient: Such lamp heads are also expected to create soft, homogeneous background shadows when an object is in the foreground, like a person in front of a white wall. This adds two angular dimensions to the homogeneity requirements, creating what we might call four-dimensional (4D) homogeneity.

As it turns out, it is this dimensional analysis which will allow us to think about homogenization and color mixing from a general, fundamental viewpoint. From there, we can derive deep and general insights into the nature of the optical design problem. To apply these insights, an optical designer also needs a well-stocked toolbox of what we would like to call *design patterns*, a term borrowed from software engineering [1]: A design pattern is a general, reusable solution approach to a commonly occurring problem within a given context.

When an experienced optical designer thinks 'mixing rod', for example, then he or she thinks of much more that just a stick of transparent material: There is a deep understanding of what a mixing rod does to incoming light, depending on length, cross-sectional shape, making it straight or tapered, and how this action on the incoming light may be used in the larger context of the optical

***Corresponding author: Julius Muschaweck,** JMO GmbH, Zugspitzstr. 66, 82131 Gauting, Germany, e-mail: mail@jmoptics.de. https://orcid.org/0000-0002-0596-3979
Henning Rehn: FISBA AG, Rorschacher Str. 268, 9016 St. Gallen, Switzerland

system as a whole. It is this knowledge that makes 'mixing rod' a design pattern.

In this tutorial, we start with explaining the propagation of inhomogeneous light in terms of *phase space*. There, the dimensional analysis mentioned will get quantitative meaning. We then apply the phase space concepts to light sources and to various applications, deriving a general taxonomy of color mixing and homogenization design problems. To illustrate the homogenization methods, we use the phase space diagram [2].

On the basis of this phase space introduction we present a series of design patterns, a more or less abstract view on methods to influence light distributions, based on the earlier insights of phase space. After rising to this 'high altitude' viewpoint, we descend back to practice, filling the tool box with a set of design elements, explaining what each is doing in terms of the introduced patterns.

2 Light in phase space

Just for the record, we are going to restrict ourselves to geometrical ray optics in 'HIL' materials (homogeneous, isotropic, linear), neglecting diffraction and coherence as well as polarization, gradient index materials and birefringence, all of which is very appropriate for most illumination problems. For more general phase space optics, see e.g. [3]. We also will not discuss deeper physics, i.e. the connection to thermodynamics (see [4, 5]), and we will simplify the mathematics by analyzing phase space on planar surfaces, mostly.

2.1 Rays and reference surfaces

The notion of phase space rests on *reference surfaces*. A reference surface, preferably planar, is a mathematical entity, not a real surface. We may insert reference surfaces into the optical system wherever we want to analyze the flow of light, with one important, seemingly counterintuitive, restriction: The reference surface must not coincide with, or intersect with a real optical surface. The reason is that we want to look at ray directions at the reference surface, and optical surfaces are precisely where ray directions change, and thus are not well defined. Often, it is desirable to analyze rays directly before or after an interaction with an optical surface. Then, we put the reference surface an infinitesimal distance away towards one side. When we do that, we know that the reference surface is embedded in a material of known refractive index n.

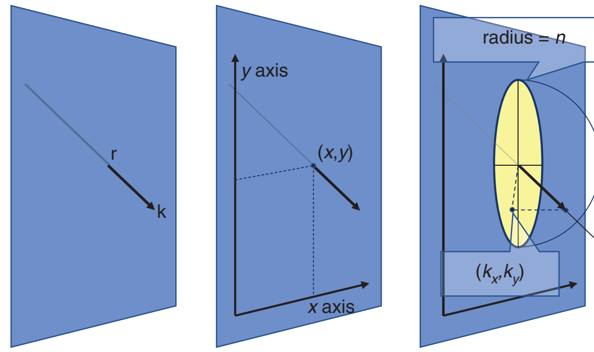


Figure 1: Coordinates (x, y, k_x, k_y) of a ray in phase space.

Let us consider at a ray, intersecting a reference plane from left to right, as shown in Figure 1 on the left. Let us call the intersection point $\vec{r}$ and the ray direction $\vec{k}$, with $\vec{k}$ normalized to $|\vec{k}|=n$. (The reason for normalizing $\vec{k}$ this way will become apparent momentarily.) We define a Cartesian coordinate system on the plane, as shown in the center, such that $\vec{r}=(x, y, 0)$. For ray direction, we attach a hemisphere, whose radius is equal to the refractive index n, to the point (x, y) on the plane. We use a second coordinate system to describe the point where the ray intersects this hemisphere, $\vec{k}=(k_x, k_y, k_z)$, where the k_x and k_y axes are aligned with the x and y axes, respectively. Then, k_z is redundant (given by the normalization of the radius to n), at least when we restrict our attention to rays from left to right, and (k_x, k_y) suffice. Given a ray that intersects the reference surface, this is how we arrive at four numbers that describe the location and direction of this ray in a unique, non-redundant way.

Vice versa: Given four numbers (x, y, k_x, k_y), given a (bounded) planar reference surface, embedded in a medium with refractive index n and equipped with a Cartesian coordinate system, and given on which side of the surface the ray is incident (i.e. $k_z>0$ or $k_z<0$), we can reconstruct the ray this way: First, (x, y) must be a point within the surface boundaries. Second, $k_x^2+k_y^2\le n^2$ must hold, such that $k_z=\pm\sqrt{n^2-k_x^2-k_y^2}$ is well defined. Then we select $\vec{r}=(x, y, 0)$ as the starting point of the ray on the surface, and select $\vec{k}=(k_x, k_y, k_z)$ as the ray direction, its length being normalized to n. Geometrically, this means to attach a tangent disc of radius n (yellow in Figure 1) to the surface at (x, y), mark the point (k_x, k_y) on the disc, and lift it up to the hemisphere over the disc. The point on the hemisphere is where the ray goes. [We should also mention that $(k_x, k_y, k_z)/n$ is the ray direction unit vector, whose components are known as *direction cosines*.]

So why is this a useful description? A first hint comes from looking at refraction. To study refraction at e.g. a

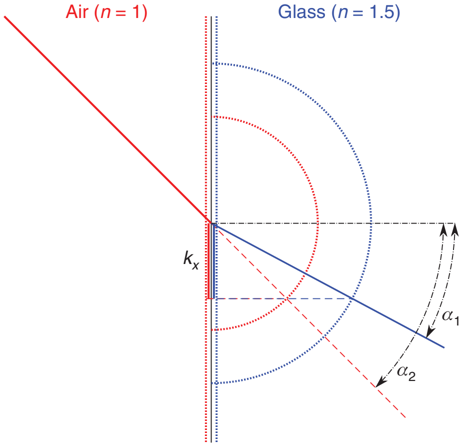


Figure 2: Refraction leaves k_x and k_y constant.

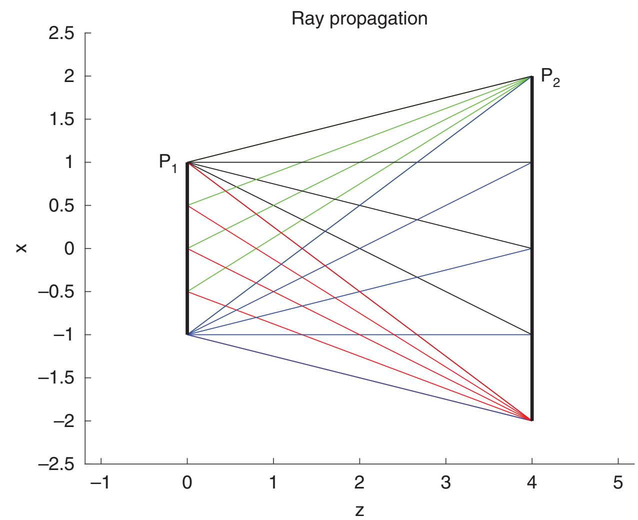


Figure 3: A ray bundle between two reference surfaces (P_1 on the left, P_2 on the right). Shown are the edge rays only: Black and blue rays come from the spatial edge of P_1, green and red rays proceed towards the spatial edge of P_2.

planar glass surface (n=1.5), we place two reference surfaces, one to each side of a refractive surface, at an infinitesimally small distance. One reference surface will be in air, the other in glass (the dashed lines in Figure 2). There, we show a ray being refracted, from air on the left into the glass on the right. The n=1 hemisphere is shown in red, the n=1.5 hemisphere in blue. The geometrical relations show that the tangential components of $\vec{k}$ (the solid red and blue vertical lines) remain unchanged under refraction.

Algebraically, we know from Snell's law that

$$n_1 \sin(\alpha_1) = n_2 \sin(\alpha_2) = \text{const} \tag{1}$$

and from trigonometry that

$$n \sin(\alpha) = \sqrt{k_x^2 + k_y^2} \tag{2}$$

where α denotes a ray's angle with the surface normal.

So when we align the coordinate systems of the two reference surfaces and ignore their infinitesimal distance, we find that the ray in air (before refraction) and in glass (after refraction) are described by the very same numbers (x, y, k_x, k_y). Thus, by defining a ray's direction by the tangential components (k_x, k_y) of its $\vec{k}$ with $|\vec{k}|=n$, we obtain a quantity which is conserved under refraction – and conserved quantities are useful analysis tools. (Physically, this prescription is based on $\vec{k}$ being the wave vector when we use the wavelength as the unit of length, and Maxwell's equation dictating that the tangential components of $\vec{k}$ remain unchanged at an optical interface.)

2.2 Ray bundles in phase space

Let us now consider extended *ray bundles* intersecting a reference surface. Each ray is identified by its (x, y, k_x, k_y) coordinates, which naturally span a 4D space, the phase space of geometrical optics. In Figure 3, we show in a 2D cross-section how a ray bundle propagates from one aperture to the next.

To visualize how ray bundles behave and look like in phase space, we use *2D phase space diagrams*. Each ray is plotted as a point in a coordinate system spanned by its x and k_x coordinates [2].

In Figure 4, the left picture shows the phase space diagram for the rays as they leave the left reference surface, P_1. Note that the colors of the points are matched with the colors of the rays in Figure 3. For example, the blue rays come from the lowest point of P_1, all having the same x coordinate but varying direction, k_x. The rays span a region which looks like a parallelogram with two slightly bent sides – a very typical look in this kind of diagram. In the right picture the same rays are shown as they intersect P_2, after some propagation. Note that, now, the green and the red rays have constant x, which is entirely correct, as they are all aimed at the uppermost and lowermost point on P_2, respectively. We see: propagation in free space amounts to shear in phase space.

The reader is asked to take away three key notions from these two figures. (i) There is not just one phase space. On the contrary, there are as many phase spaces as there are reference surfaces. It is important to choose the right ones for fruitful analysis. (ii) These infinitely many phase spaces are not independent. On the contrary, the laws of physical ray propagation *induce a mapping* (see Figure 4) between any two phase spaces. You can pick any point (=ray) in a phase space, and physics will dictate

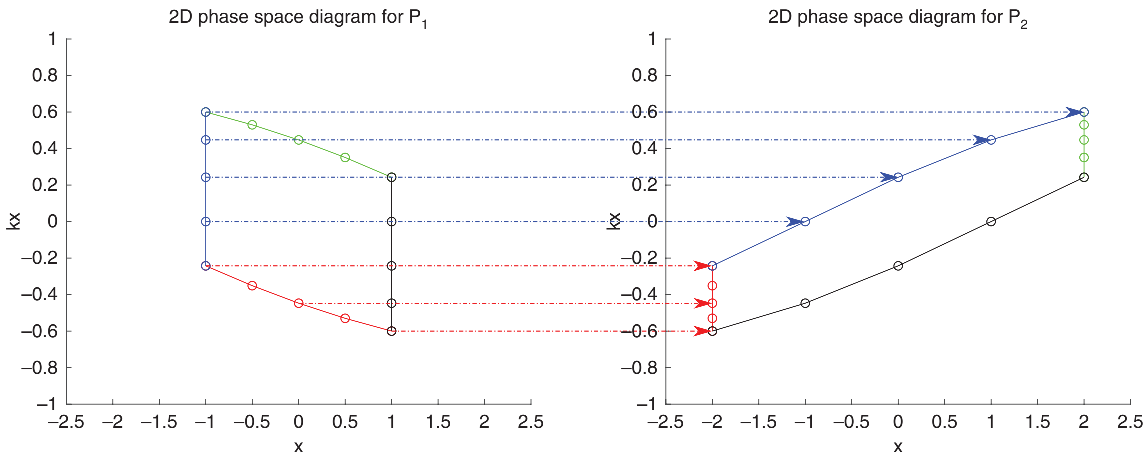


Figure 4: In phase space diagrams corresponding to the two reference surfaces, P_1 and P_2 in Figure 3, rays are shown as points. Colors are matched between this figure and Figure 3: the blue rays, e.g. proceed from the lower edge of P_1 all across P_2. The dashed lines are *not* rays: they visualize the mapping between the two phase spaces which is induced by the laws of optics.

which points (=rays) in other phase spaces of the same system correspond to the one you picked.

2.3 Étendue

The third notion is *volume*. The points (a.k.a. rays) in Figure 4 enclose a 2D volume (a.k.a. area) each. It is hard to visualize, but the process is entirely analogous in full three-dimensional (3D) space, where the phase space has four dimensions, not two. The 4D volume U in phase space, spanned by an extended ray bundle, is known as *étendue*. Just like the area A in 2D can be measured by $A=\int \mathrm{d}x\mathrm{d}y$, and the volume V in 3D by $V=\int \mathrm{d}x\mathrm{d}y\mathrm{d}z$, the étendue U can be measured by

$$U=\int \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}k_x\,\mathrm{d}k_y \tag{3}$$

$$=\int \mathrm{d}An^2\cos(\theta)\mathrm{d}\Omega \tag{4}$$

Étendue is a French word, meaning ‘extent’, and is a quantitative measure of the combined angular and spatial size of a ray bundle. Étendue is intimately related to the Lagrange invariant in imaging optics. However, étendue makes no assumptions about symmetry, or about rays being paraxial, or about an optical axis at all. (In fact, the coordinate free definition in Eq. (4) allows for curved, non-planar reference surfaces, but we will restrict ourselves to planar reference surfaces in this tutorial.)

What makes étendue such an immensely useful quantity is the *law of conservation of étendue* – often mentioned, often not entirely understood. It is important to be very clear what that means, exactly. So let us be precise. We consider a ray bundle, i.e. a set of rays with finite spatial and angular extent, which proceeds through an optical system. We also define a number of reference surfaces, such that each and every ray from the ray bundle intersects each and every reference surface exactly once. Figure 3 is a very simple example, with free space propagation and two reference surfaces. We compute the étendue of the ray bundle according to Eq. (3) as it intersects each reference surface. We assume the system meets the following requirements: (i) The rays undergo only free propagation, refraction and reflection on piecewise smooth surfaces. (ii) There is no scattering, no ray splitting at partially reflecting and/or birefringent surfaces, no wavelength conversion at, e.g. luminescent dyes. (iii) Partial absorption at surfaces and in material is allowed: étendue is about geometry, not flux.

Then, it turns out that the étendue of this ray bundle is precisely the same on each and every such reference surface: the mapping between phase spaces given by the physics of ray propagation conserves étendue. (Proofs are found, e.g. in [4, 6].)

To clarify, let us consider three situations where étendue is *not* conserved, because requirements are violated. (i) A beam from a video projector hits a screen, from where it is diffusely reflected. The reference surface is immediately in front of the screen, in air. We consider two phase spaces on this surface: one for incoming, one for outgoing light. The cross-section area is the same for incoming and reflected ray bundles. However, the incoming beam has a small étendue, because it subtends only a small angular range: the lens of the video projector as seen from the screen. The reflected beam, however, is scattered into the complete hemisphere, increasing the étendue by a large factor. (ii) In a beam splitter, we can define two ray paths: transmitted and reflected. Étendue is conserved in each ray path separately, but if one considers both paths simulteaneously, étendue is precisely doubled. (iii) In an optically pumped single mode laser, e.g. a doped ruby crystal pumped by a flash lamp, the incoming étendue from the flash lamp is large, but the outgoing étendue of the collimated laser beam would be zero, if diffraction at the aperture would not create a small angular spread (interestingly, the étendue of a single mode is always λ^2 [7]).

2.4 Radiometry in phase space

Now, it is time to fill phase space – this purely mathematical, geometrical structure – with light. On some reference surface, let us pick a ray $r=(x, y, k_x, k_y)$, and let us look at its immediate neighbors, i.e. the rays coming from an infinitesimal area element $\mathrm{d}A=\mathrm{d}x\mathrm{d}y$ around (x, y), and going into an infinitesimal angular range $\mathrm{d}k_x\mathrm{d}k_y$ around (k_x, k_y). This defines an infinitesimal étendue element $\mathrm{d}U$,

$$\mathrm{d}U=\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}k_x\,\mathrm{d}k_y \tag{5}$$

There will be an infinitesimal amount of radiant flux $\mathrm{d}\Phi$ flowing in $\mathrm{d}U$. This allows us to define the density of flux in phase space, in an entirely analogous way to the definition of mass density $\rho=\mathrm{d}m/\mathrm{d}V$, or electric current density $J=\mathrm{d}I/\mathrm{d}A$, or any other density which is defined by dividing the infinitesimal amount of something by the infinitesimal n-dimensional volume which that something occupies.

The density of radiant flux in phase space is defined by

$$L=\frac{\mathrm{d}\Phi}{\mathrm{d}U}=\frac{\mathrm{d}\Phi}{\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}k_x\,\mathrm{d}k_y} \tag{6}$$

and is called *radiance*. When radiant flux Φ is weighted with the $V(\lambda)$ function to take the human eye's sensitivity into account (and then called luminous flux Φ_v), the density of luminous flux in phase space is called *luminance* and is denoted by L_v. When we look at *spectral radiant flux*, by restricting our attention to light in an infinitesimally small wavelength interval $\mathrm{d}\lambda$, we obtain *spectral radiance* L_λ,

$$L_\lambda=\frac{\mathrm{d}L}{\mathrm{d}\lambda}=\frac{\mathrm{d}^2\Phi}{\mathrm{d}U\,\mathrm{d}\lambda} \tag{7}$$

Radiance and luminance are functions of four variables, two each for location and direction: They measure how bright the light is, when we look back into a ray, $L=L(x, y, k_x, k_y)$.

When the requirements for étendue conservation hold (see above), and in addition there is no absorption, then radiance L, luminance L_v and spectral radiance L_λ are also conserved along a ray, as that ray passes through the optical system: $\mathrm{d}\Phi$ is conserved by energy conservation, $\mathrm{d}U$ is conserved by étendue conservation, and therefore their quotient is conserved as well. This holds when the ray propagates freely, when it is reflected or when it is refracted into or out of some optical material. As spectrally weighted radiance is what the human eye sees (light coming from a spot as small as the eye can resolve, and going from there into the tiny solid angle of the eye's iris), the reader can confirm the conservation of radiance by looking at a diffuse, self-radiating surface like a white cloud, blue sky, or a computer screen, through an optical system, e.g. an AR-coated singlet lens, or a clear glass of water. The reader will notice that the object looks no brighter or darker through the optics than compared to looking at it directly. (For historic reasons, radiance is officially defined by dividing our L by n^2, and our L is called 'basic radiance'. We feel that it is about time to reserve the short name for the quantity for which we do have a conservation law, and we take the liberty to omit 'basic'. We also mention in passing that L_λ is the quantity that Planck's famous blackbody spectrum formula gives us. Accordingly, the conservation of spectral radiance is intimately connected to the second law of thermodynamics).

There are some more common quantities in illumination engineering. Irradiance E and illuminance E_v denote the density of flux over area,

$$E=\frac{\mathrm{d}\Phi}{\mathrm{d}A}=\frac{\mathrm{d}\Phi}{\mathrm{d}x\,\mathrm{d}y} \tag{8}$$

By comparing Eq. (8) with the radiance definition in Eq. (6), it is apparent that we can obtain the irradiance E at a point (x, y) on some reference surface also by integrating radiance $L(x, y, k_x, k_y)$ over the two angular dimensions of phase space:

$$E(x, y)=\int L\,\mathrm{d}k_x\,\mathrm{d}k_y \tag{9}$$

$$=\int Ln^2\cos\theta\,\mathrm{d}\Omega \tag{10}$$

The second Eq. (10) shows how the integration must be performed over a solid angle Ω instead of k_x, k_y. Analogous equations apply to spectral irradiance E_λ obtained from spectral radiance L_λ by integrating over k_x, k_y. This is the mathematical expression of the definition of irradiance/illuminance: How much light is incident per surface area, no matter from which direction it comes.

We omit radiant or luminous intensity here, as far field angular distribution can easily be related to irradiance or illuminance on a distant screen.

2.5 Color in phase space

Let us now analyze an inhomogeneous source, like two adjacent LED chips of different color. We use the same simple setup as in Figure 3, with two planar, parallel reference surfaces and free space propagation in between. Now, we place two LED chips just behind P_1, one red and one cyan (an admittedly unusual, but instructive and easy to draw method to create white light). The new setup is shown in Figure 5.

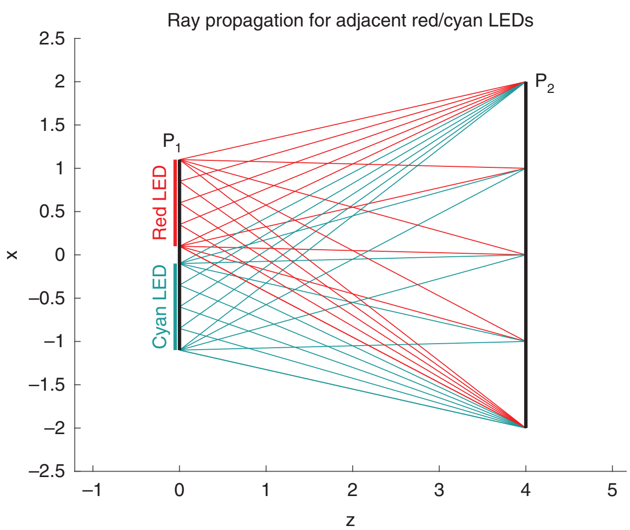


Figure 5: Two ray bundles from two sources intersect two reference surfaces (P_1 on the left, P_2 on the right). Red/cyan rays come from the red/cyan LED, respectively.

In Figure 6, we show how color superposition works from a phase space point of view, restricted to the x, k_x dimensions for visualization. To obtain the spectral irradiance E_λ at a point x, we integrate the spectral radiance L_λ over the angular dimension k_x according to Eq. (9) (Figure 6, left). At P_2, the result is light gray in the center, and darker at the edges, from where the LEDs subtend a smaller angle (Figure 6, right). In addition, the result is slightly cyanish at the edge of a plane between P_1 and P_2 at $x=-2$. As seen from there, the cyan LED is closer and contributes more light, relative to the red LED (vice versa for the $x=+2$ edge).

We will revisit the phase space view when we analyze the action of various optical elements for color mixing.

3 Illumination design for homogenization

3.1 Where is light mixing needed?

Color mixing and homogenization are required and widely applied in diverse fields, such as

- General lighting: Office, architectural, museum
- Transport: Automotive, airplane, train lighting
- Consumer products: data projection, VR glasses, head up displays

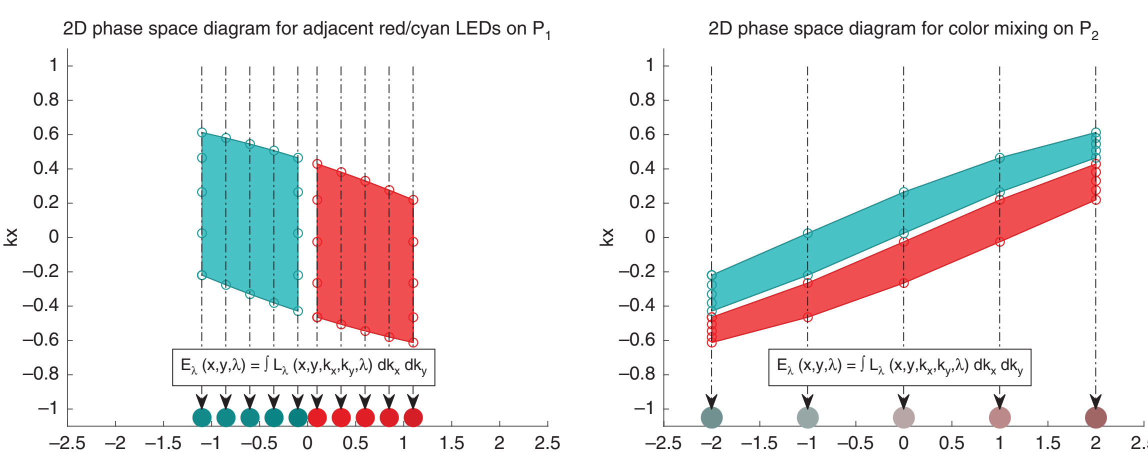


Figure 6: In the 2D phase space diagrams, we can visualize how spectral irradiance is obtained by integrating spectral radiance over the angular dimensions of phase space. Colors are matched between this figure and Figure 5. The dotted vertical arrows show the integration direction. Color and brightness of the dots at the bottom visualize the integration result.

– Stage and studio lighting
– Medical lighting: Surgical microscopes, endoscopes and luminaires
– Pump light shaping for laser activated remote phosphor (LARP) [8, 9]
– Solar collection [10, 11]
– Solar simulators [12–14]
– Sensors and measurement: distribute light equally to a variety of detectors

In the first five fields, the purpose is to create a 'good' visual impression, both for consumers and professionals. In the last four fields, the purpose is of a technical nature: homogeneity is required here to avoid destruction, to increase efficiency or to ensure measurement accuracy. However, we often find similar optical elements and architectures across these fields. Obviously, application is not what differentiates between approaches so much. We will have to find other questions which can we ask about a system to help us arrive at good solutions.

3.2 Design goals

Let us begin with the target, i.e. the required properties of the final light distribution. In most consumer electronics applications like projection, VR/AR displays, in architectural lighting (wall washing), but also in technical systems like concentration of sunlight on a PV cell, or illumination of a sensor, homogeneity of *irradiance/illuminance and color* is required. In phase space terms, this is a *2D problem*. As long as the integration over angular dimensions (see Figure 6) yields spatially homogeneous results at the target, it is not important how strongly radiance, luminance and color vary over angle. For such systems, different color subsystems may even be fully separated at the luminaire side, as long as they all deliver nearly the same illuminance distribution on the target (see Figure 7). Far field homogeneity of intensity or color falls into the same category, as it is equivalent to illuminance or color homogeneity on a very distant screen.

Sometimes, however, more mixing is needed, e.g. for spot light lamp heads for movie set lighting, studio and stage lighting, or for lighting sculptures in museums. In addition to the common requirement that the light should look 'good' on a planar screen, it is also desirable for these systems that they create homogeneous, soft shadows, which are cast onto the background by foreground objects in more complex scenes. Such soft shadows can only be obtained if the lamp's exit pupil looks homogenous, as seen from the target.

Figure 7: A small cell phone flash LED with a cold white (top right) and a warm white (bottom left) LED chip, with identical Fresnel lenses for each LED. Perfect color mixing of these two sources is achieved, except at very close distances which are not relevant for this application. (Image courtesy OSRAM Opto Semiconductors GmbH.).

Thus, homogeneity is required here both in spatial and angular coordinates at the target. To achieve such *4D homogeneity*, the light must be fully mixed in phase space, which requires devices like mixing rods, fly's eye condensers (FECs) and/or scatterers (all of which will be discussed below) as necessary prerequisites.

Other key considerations relate to flux, étendue and luminance. Obviously, the available flux from the source must exceed the required target flux, and the source étendue must not be too large, compared with the available target étendue. (For details on the interplay between sufficient flux and sufficient étendue, especially for strongly inhomogeneous sources, see [15].) Finally, the source luminance must be sufficient to achieve the required illuminance at the target, which may be a limiting factor for automotive headlamps or for wall washing systems. These considerations, treated in more detail, e.g. in [6], are beyond the scope of this paper, however. What is important here is: Do we have an étendue limited system, where target étendue is just large enough to squeeze the light in (e.g. projection), or do we have a system where the target étendue is large (e.g. office lighting)? The difference is that in the latter case, we may use more or less diffuse scattering, while in the former case, scattering may be used only as a final touch-up, if at all.

Finally, it is important what 'sufficiently homogeneous' means. For technical illumination, e.g. concentrated photovoltaics, a typical requirement is $E_{min}/E_{max} > 1 - \varepsilon$, which is fairly easy to use as an optimization target, even if the requirement is strict (small ε). Other fields may use different uniformity [16, 17] and color quality definitions [18]. For many lighting applications, a typical requirement is having homogeneous color while illuminance gently falls off towards the edge. Thus, different problems often require different solutions, but not always: 4D homogeneity implies 2D homogeneity, and good color mixing generally implies good illuminance homogeneity for inhomogeneous single color sources, too.

With perfect sources, little work would be necessary. But in the real world, we have real world sources with imperfect distributions. Let us now look at these.

3.3 The problem – inhomogeneous sources

Widely used light sources, examples of which are shown in Figure 8, comprise white LEDs (partial phosphor conversion of blue light), multi color LEDs (cold white/warm white, RGB, RGB+), and traditional lamps (incandescent, high intensity discharge (HID) lamps).

Neglecting polarization and diffraction, a complete description of an inhomogeneous source would be given by spectral radiance, $L_\lambda(x, y, k_x, k_y, \lambda)$ in 4D phase space. This is what near field goniometry is about, effectively. Visualization, however, usually happens in 2D, showing either spatial (see Figure 8) or angular distributions (intensity plots, given in most LED data sheets).

For illumination design purposes, it is more instructive to look at the distribution in a *mixed 2D plot* – the phase space diagram introduced in the preceding chapter. Here, structured sources would appear as shown in Figure 6 (left) for the case of a red LED chip next to a cyan LED chip. Phase space diagrams of some other sources are shown schematically in Figure 9.

Traditional light sources show similar problems. Incandescent lamps with coil shaped filaments are homogeneous in the angular domain, but very inhomogeneous spatially, giving rise to August Köhler's illumination principle [20].

Compound sources (the bare source combined with some primary optical element) may show inhomogeneities that are due to the optics. Lamps with parabolic or elliptic reflectors, such as HID burners for projection, or incandescent spot lamps, effectively have a ring shaped aperture with a central hole, and angularly dependent magnification [21]. An example (LED with reflector) is shown in Figure 10.

From a phase space point of view, this is the taxonomy of source inhomogeneities:

– Which kind of inhomogeneity has the source? Spatial, angular or both?
– Is there inhomogeneity in color, luminance or both?
– How finely grained is the inhomogeneity? Does the source have many small features like in large white multi-sapphire-chip LED light engines or incandescent filaments? Or does it consist of very few homogeneous sub-blocks like the four chip LED package shown in Figure 8, bottom?

When the phase space structure of the source inhomogeneity is understood, the optical designer can choose from

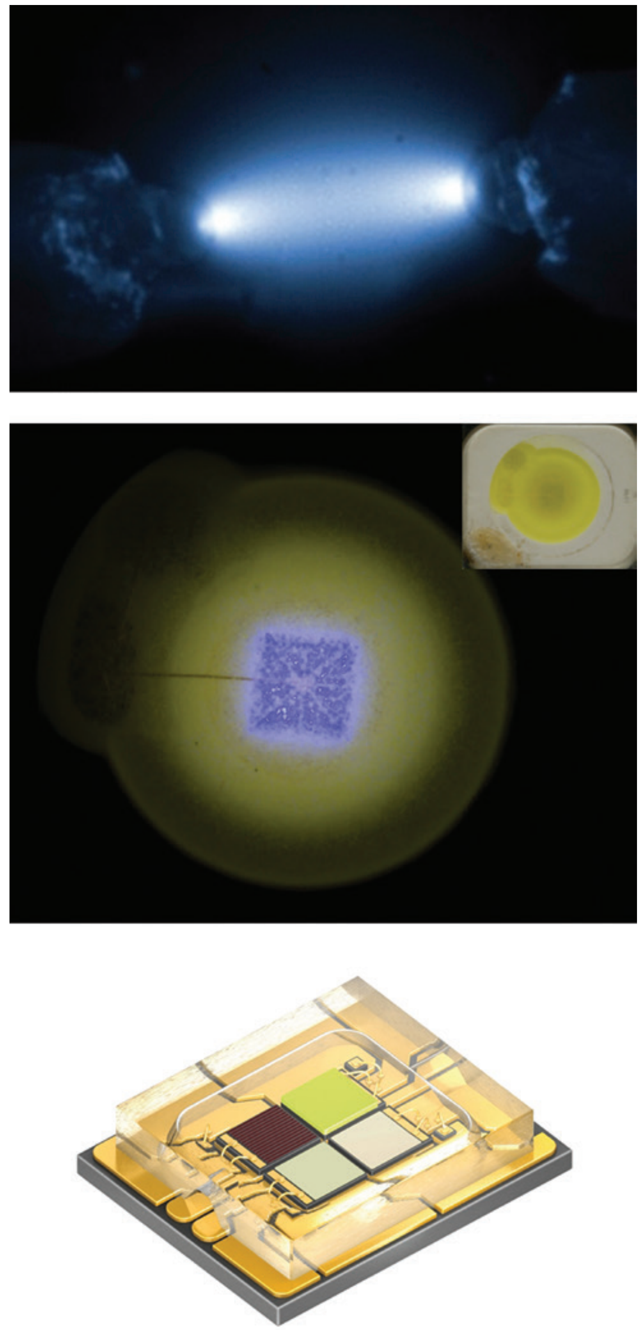


Figure 8: Various inhomogeneous sources. Top: Arc of a HID lamp with hot spots and halo [19]. Center: White LED with blue emitting LED chip embedded in phosphor matrix (insert shows unlit package). Bottom: Multi color LED package, with white, red, green and blue LEDs. Top and center pictures by the authors, bottom picture courtesy OSRAM Opto Semiconductors.

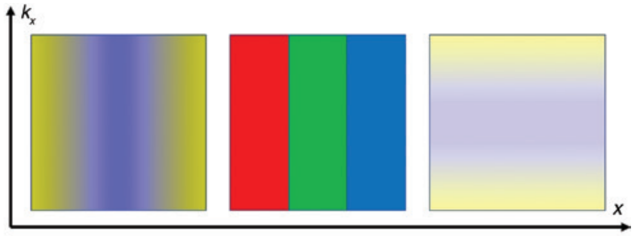


Figure 9: Schematic phase space diagrams of various LEDs. Left: White LED (blue chip embedded in phosphor matrix) as shown in Figure 8, middle, with its spatial color distribution. Center: Multi color LED package with adjacent red, green and blue chips. Right: Thin film white LED (blue thin film, surface emitting chip covered with thin phosphor layer), with its spatial homogeneity and its typical color over angle distribution.

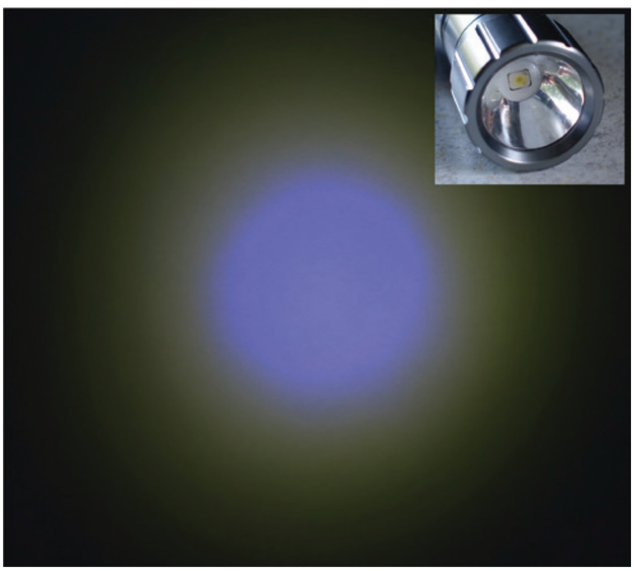


Figure 10: Far field distribution on a distant screen of a white LED in a parabolic reflector (see insert). The LED is the same as shown in Figure 8 (middle). The parabolic reflector images the inhomogeneous spatial source distribution into the far field. As a light source to be used with further optics, this device is both spatially and angularly inhomogeneous. As a cheap consumer flash light (what it actually is), this device is an example of a particularly bad optical design for the given source and the given purpose.

various tools in the toolbox to transform the source light into the desired homogeneous output. However, tools are only useful when it is fully understood what the tools actually do. Therefore, it is necessary to study the *optical function* first, before we look at individual *optical elements*.

4 Design patterns: basic principles

First, we present basic principles which provide light mixing either by themselves, or in combination with other elements. Then, we will describe several design elements which use these basic principles in various ways.

4.1 Propagation

Propagation in free space or in a medium causes a spatial light distribution to mix, as long as an angular extent is present. As seen in Figure 4, propagation causes shear in phase space. Ray intersection points move spatially (sideways), whereas the angular distribution remains unchanged. It may well be sufficient to let a spatially inhomogeneous distribution propagate a certain distance, to obtain some light mixing. A well-known example is to use defocus as a means to improve the homogeneity of a distribution. However, regions of phase space ‘stay together’ under propagation, and they keep their étendue. Only their shape in phase space changes.

4.2 Étendue transformation

Optical systems can provide a transformation of the phase space region where the inhomogeneous distribution is located. The best known example is the area to angle conversion provided by a positive lens imaging a source to infinity: The angular distribution of the source shows up on the lens aperture, and the spatial distribution of the source is imaged into far field directions. This amounts to *rotation* in phase space. So one can often use the most useful source property, i.e. its homogeneous far field instead of an inhomogeneous near field (spatial distribution), and put that onto the target. This is the basic idea of Köhler’s illumination principle [20]. For demonstration, we apply Köhler illumination to the now familiar red/cyan LEDs (see Figure 11).

4.3 Étendue expansion

If the inhomogeneity is associated with a small étendue source to be propagated to a large étendue target, the task is to spread the light into the latter. With propagation and smooth optical surfaces only, the small source étendue will be conserved, and the target étendue will be only partly filled: There will be bright regions in the target phases space, surrounded by dark regions. This provides the opportunity to increase the source étendue, e.g. by scattering, and achieving homogeneity at the same time.

Consider an automotive HUD (head up display), with a reflection from the windshield and with a virtual image

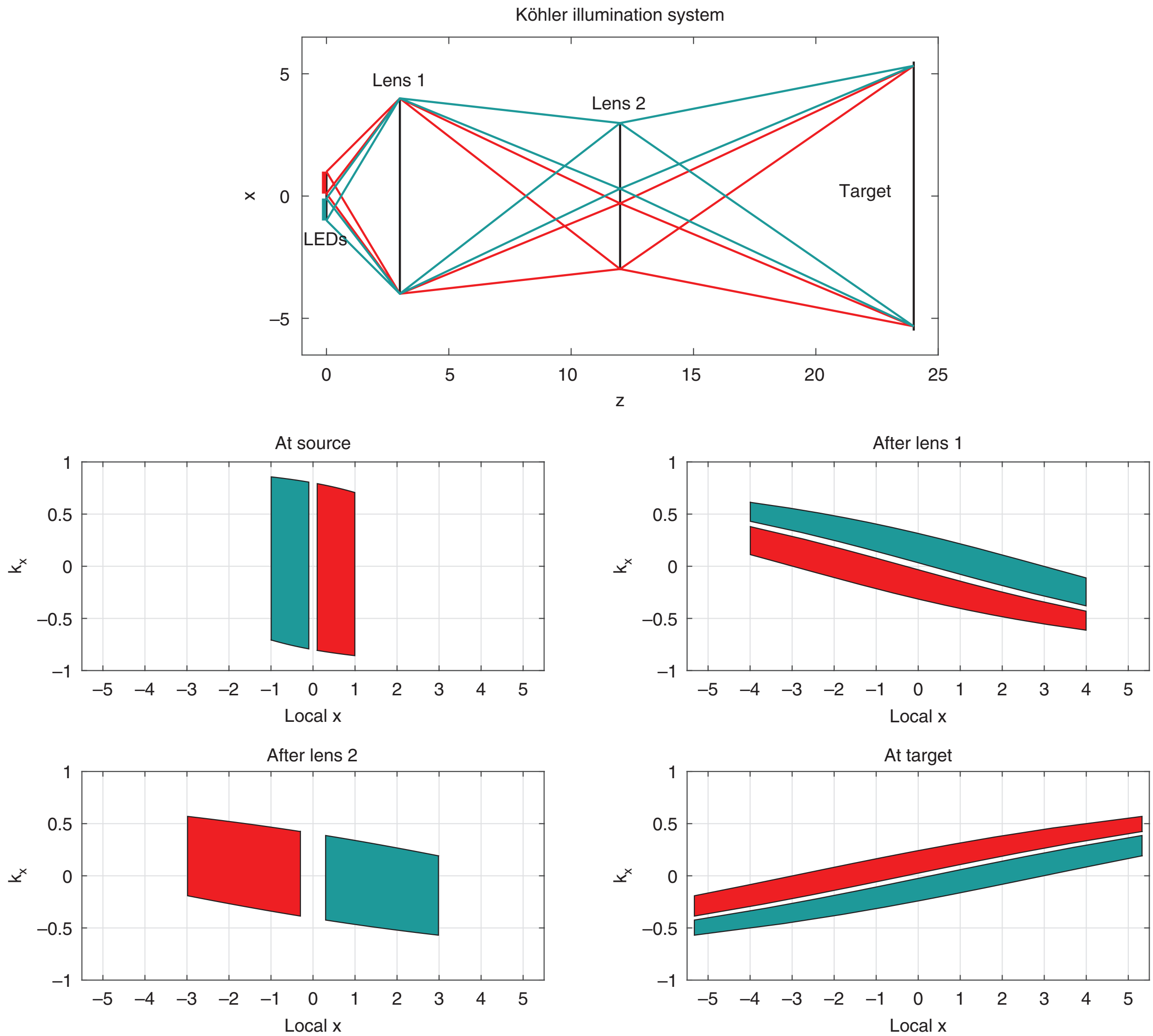


Figure 11: Köhler illumination exploits the homogeneous angular distribution of a source. Top: The system with red/cyan LEDs on the left, two ideal lenses and the target, and rays. Bottom: Four phase space diagrams at various positions. The sources illuminate lens 1 homogeneously, thanks to propagation. Lens 1 images the sources onto lens 2. Lens 2, in turn, images the homogeneous distribution on lens 1 onto the target.

somewhere in front of the car. Commonly, the étendue of the projector is considerably smaller than the target étendue defined by the size of the virtual image, its distance and the size of the eye box from where it should be viewable. Without any further measures, étendue conservation would cause the eye box to be rather small. The solution is to place a weak scatterer into the light path, to adapt the projector étendue to the target étendue and to distribute the light of each pixel homogeneously across the eye box.

In street lighting and in office lighting, the primary task is to spread the light of a high luminance source to a prescribed spatial distribution. But to avoid glare, it is preferred if the light leaves the lamp from a not too small exit pupil.

Scattering surfaces or volumes can be applied for pump light shaping (de-peaking) for LARP where defocus only is just not reliable enough [22, 23].

The étendue expansion itself can be achieved efficiently by *scattering*, both random (e.g. injection molding with eroded mold inserts, ground glass) and deterministic (e.g. single sided micro lens arrays, faceted reflectors). A notable class of plastic materials uses volume scattering by embedded clear particles, which exhibit small scattering angles and nearly no backscattering, making efficient, simple systems possible.

All these measures provide limited phase space transformations: shear, rotate or stretch, all of which can bring up the better side of a light source's distribution, but cannot provide full *4D homogenization* without étendue increase. To realize such a complete homogenization, we need other measures.

4.4 Étendue splitting

This technique is the key to efficient and effective light mixing in étendue limited systems. The design pattern is to:
- split the source étendue into parts
- transform the parts appropriately
- recombine them to the target étendue

The 'treatment' can mean to:
- throw away some parts
- redirect each part, e.g. to different regions of the target. This is a shift in phase space.
- spread each part on the whole target. This is a phase space transformation.

Étendue splitting means that adjacent rays which started out from nearly the same point in phase space (having nearly identical starting points and directions) end up at very different points in another phase space later in their path: The mapping between phase spaces that is dictated by ray propagation physics becomes discontinuous.

The phase space of an incident light distribution may be split by:
- area
- angle
- spectrum
- polarization

In a simple, but not very practical example, we start with a simple relay lens that images a square source onto a target. The goal is to illuminate a 2:1 aspect ratio. This is achieved by cutting the lens into to halves, each of them slightly decentered (Figure 12). This simple example shows that splitting the étendue requires discontinuities in the optical surfaces: Steps, or kinks, serve as knives in phase space, cutting étendue into pieces which then proceed separately through the system.

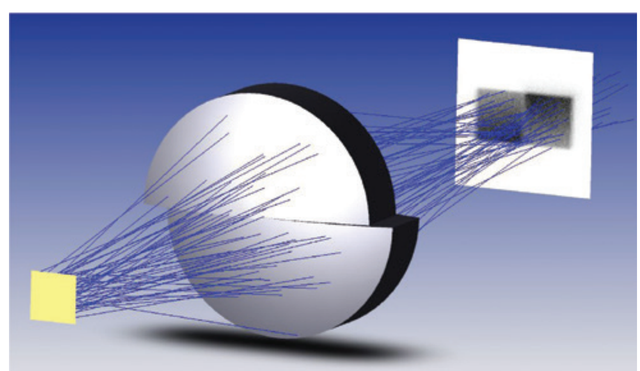
An extreme 'Gedankenexperiment', jokingly called the *spaghetti design*, means to collect every étendue bin of a source étendue by a suitable optical fiber (the single spaghetti) and then combine all those fiber outputs (with additional optical elements) into a beam of suitable area and angle. For example, we could place the fiber inputs densely onto a sphere centered around the source, arrange the fiber outputs into a plane, with a lens to each fiber end, and thus create a perfectly collimated beam without geometrical losses and without étendue increase. This line of thought is useful to see if something can be done in principle: If it is possible with a spaghetti design, then it does not violate physics, and 'all' that remains to be done is to find a more practical solution.

Facets on a reflector surface work by the same principle: the far field of the source is split by the facets and each facet provides a different way to send it to the whole target. As a result, the target illumination is well mixed. However, facets on a single surface increase effective étendue: the outgoing ray bundle consists of many small bright pieces, one from each facet, with dark phase space regions in between, such that the volume of the envelope of this outgoing ray bundle has increased by precisely the volume of these dark phase space regions.

To conclude this section: Most optical elements in illumination systems perform one of these functions. Let us now look at them in more detail.

5 Design patterns: elements

Now, we finally describe the toolbox, reviewing practical homogenization devices and discussing design aspects.



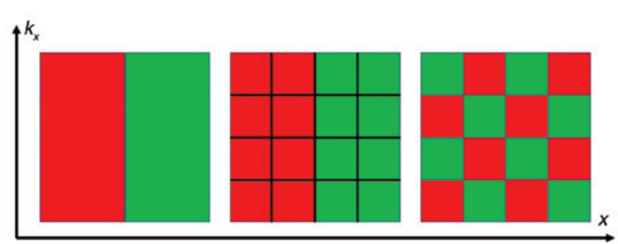


Figure 12: Top: Étendue split as a means to illuminate a target of a different aspect ratio. Bottom: The principle of étendue splitting: The large phase space regions (left) are cut into small pieces (center) and reassembled in a different arrangement (right), which is now homogeneous in 4D: both in angular and in spatial dimensions.

5.1 Rod integrators

A mixing or integrating rod is a lightguide which realizes homogenization of the incident light distribution by multiple reflections at its side faces. Mixing rods, or rods for short, are not complete opticals systems by themselves. They are mostly used as a subsystems, to provide a spatially homogenous secondary source which then is imaged somewhere.

Rods come in two main flavors, both of which have pros and cons: They may be made of a transparent material such as plastic/glass (see Figure 13), or by assembling mirror segments [24–26]. Solid rods are not only relatively easy to make an cheap, they also use lossless total internal reflection, whereas highly reflective coatings of a hollow rod's segments are difficult to realize for a wide range of angles of incidence. On the other hand, it takes some effort to apply an AR coating on the rod's entrance and exit faces, and if the rod exit is imaged directly or indirectly into the far field, special care must be taken to keep the exit face perfect, clean and free of dust. A hollow rod of a given length is a more effective mixer than a solid rod of same dimensions, because refraction at the entrance of a solid rod bends the rays such that they undergo fewer reflections. Finally, there is some effort to hold a solid rod by mechanical elements. Sometimes a collar is added to the body of the rod for easy mounting [27, 28].

Design parameters of a mixing rod are:

- length
- material (hollow or solid)
- cross-section shape (rectangle, hexagon, edges straight or grooved)
- shape along the axis: straight, tapered, curved

Figure 13: Integrator rods made of glass (Photo courtesy of Auer Lighting).

The device acts on the incident light, so we have to take into account its spatial and angular distributions [29].

A mixing rod acts as an étendue splitter: it cuts the phase space of the incident light into pieces, and the knife cuts along constant angle, approximately. The longer the rod, and the wider the source angle, the more and smaller the pieces get. The pieces undergo individual phase space transformations, and then are reassembled at the exit surface.

5.1.1 Straight mixing rods

We start with the basic type of a mixing rod, whose rectangular or hexagonal cross-sectional shape does not change along the axis. Straight mixing rods work well for spatially inhomogenous light at the entrance, but angular inhomogeneities, e.g. the color over angle problems of many white LEDs, or the 'hole' in the center of an arc lamp source (see Figure 14) remain unaffected.

Mixing rods work like caleidoscopes: Imagine yourself as being a tiny observer at the exit face. When you look back into the mixing rod, you see the source, and many mirror images of the source. One way to see how mixing rods homogenize is to imagine yourself now moving across the exit face. What you see while you move is the same scene, from slightly varying angles – and the same scene will cause the same overall color and brightness.

In phase space terms, the edges of the side faces split the incoming ray bundle. Rays with increasing angle towards the rod axis will undergo an increasing number of reflections on the side faces. These pieces are then transformed by free propagation and reassembled at the exit face (see Figure 15).

As a note, the beam leaving a straight mixing rod is always centered around the axis of the rod; it is *telecentric*. Accordingly, a projection lens that provides an image of the rod's exit face should be telecentric as well.

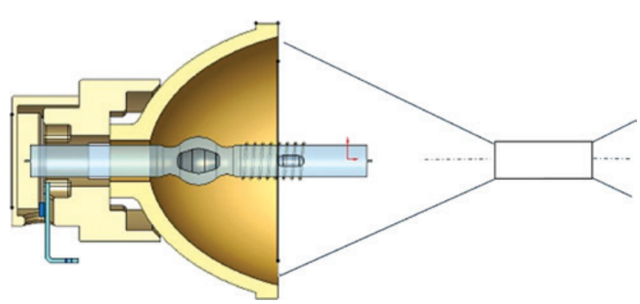


Figure 14: Rod integrator combined with a lamp in a elliptical reflector for use in a projector [30]. See also [31].

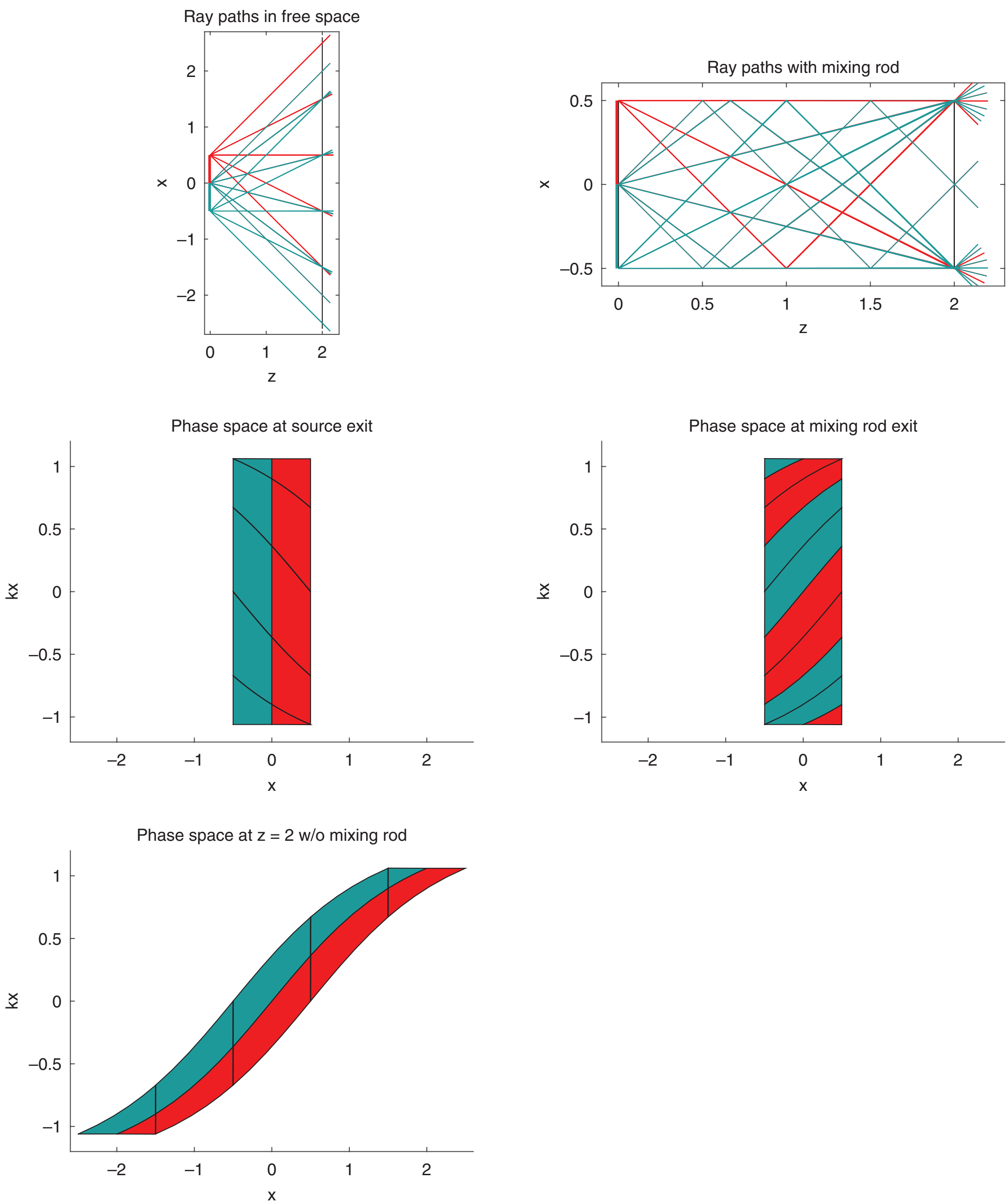


Figure 15: Mixing rod in phase space view. We use the familiar red/cyan source combination at $z = 0$, now with telecentric output up to 45°. On the left, the free propagation to $z = 2$ is shown: The ray paths (top left), and the phase space distributions at source exit (center left) and after propagation (bottom left). On the right, by using a mixing rod with diameter 1 and length 2, all the light from the sources is propagated to the exit surface at $z = 2$ (top right). The phase space distribution at the exit face is shown at center right. The phase space pieces are transformed and reassembled in a way that the angular integration (vertical) yields nearly identical contributions of red and cyan across all values of x: homogenization.

5.1.2 Tapered mixing rods and compound parabolic concentrators (CPCs)

Both combine two functions: Homogenization and collimation for large (up to hemispherical) source angular ranges. However, CPCs, explained in detail in e.g. [6]) work less well: Homogeneity is not nearly as good for sources smaller than the entrance area, they are more difficult to make, and there is less freedom to adjust the CPC's length independent of angular ranges. Their only advantage is that output is telecentric. Accordingly, not many CPCs are actually used in illumination systems.

Tapered mixing rods, on the other hand, provide a flexible, easy to make solution when both homogenization and collimation are desired. Figure 16 shows a view through a tapered rod whose entrance area is illuminated from the side. The virtual images are concatenated at an angle, which is caused by the tilted side faces. In effect, a tapered mixing rod converts the source into a mosaic of sources on a spherical surface, which send their light through the exit face. As the virtual source sphere is at a finite distance, the resulting ray bundle is homocentric and may require an additional field lens at the output.

When used with a field lens, their degree of collimation (i.e. angular exit beam width) depends only on the relative size of the entrance and exit faces, while the length can be chosen according to homogeneity and other requirements, which adds significant flexibility.

In phase space, tapered mixing rods work similar to straight mixing rods (Figure 15). The main difference is collimation: the exit phase space region becomes wider in x and narrower in k_x.

5.1.3 Shape and other design aspects

For a proper mixing function, it is helpful when the cross-section of a rod allows for the tesselation of the plane (rectangles, hexagons) [32, 33]. Cylindrical rods usually mix in the azimuthal domain only, and tend to concentrate the light in the center. However, other cross sections may work surprisingly well. For example, longitudinal *ripples* on the outer faces ([34], Figure 17) improve the homogenization function and may be even effective in the angular domain. In addition, we mention some recent research on chaotic ray propagation in specially shaped mixers [35].

The longer the rod, the better is the homogeneity at the exit face. In phase space (see Figure 15), we simply get more reflected portions overlaid, and even in the case of an inhomogeneous incoming angular distribution it is likely to get a spatially homogeneous output.

The incident angular distribution of light is basically conserved, perhaps rotated or mixed in the azimuthal domain. A really weird angular distribution may even disable the homogenizer function [29]. On the other hand, harmless incident light distributions such as that of an arc lamp in an elliptical reflector for DLP projection [36] require very few reflections for a homogeneous output (an average of slightly less than one reflection per ray may be sufficient).

In practical illumination designs, an angular distribution of around ±30° within the rod is best suited for the mixing rod's function. Narrower beams may require extended rod lengths and are often better handled with a FEC. On the other hand, focusing an incident beam to higher angles (higher numerical aperture) implies a smaller rod area and may cause tolerance problems.

Entrance and exit faces do not need to have the same shape. A *shape changer* (Figure 18) may be a smart way to adapt between different shapes of source and target area [37, 38] as long as some modifications of the angular distribution are not that important. The exit area may even form a skew quadrangle for a better adaptation to a asymmetric optical system [39].

A *field lens* may be provided in the exit face of an integrating rod. It will not affect the light distribution in that

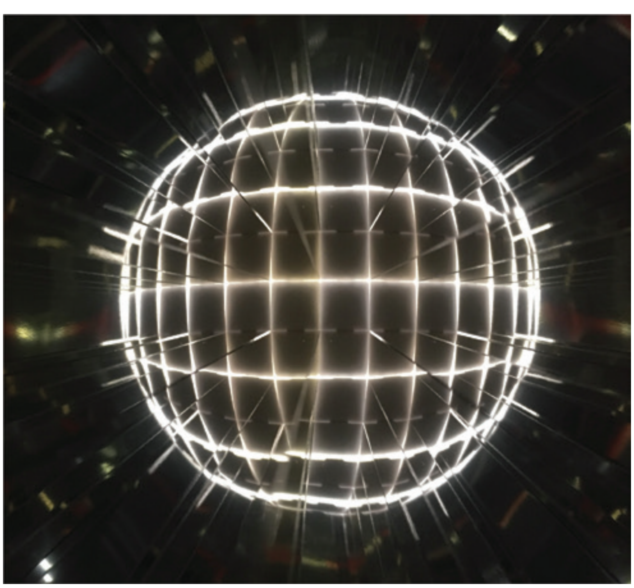


Figure 16: Look into a hollow tapered rod of square cross section. The rod was standing on a table with the rim of the smaller bottom face illuminated from the side (Photo courtesy by Markus Stange, OSRAM).

Figure 17: Rippled mixing rod.

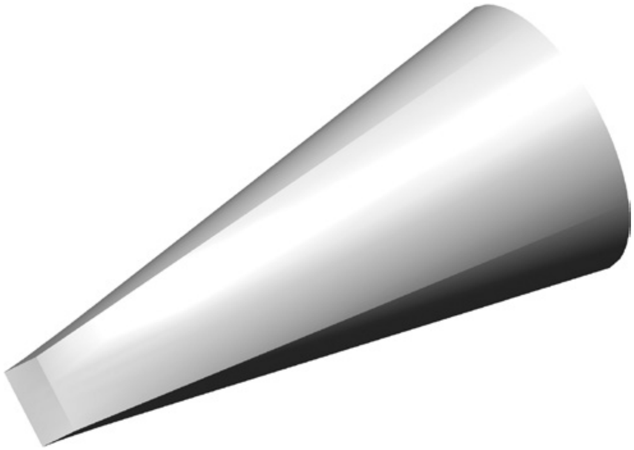


Figure 18: Shape changer.

plane, but may help to direct the light into the next optical aperture, while producing a virtual image of the entrance face far away (even at infinity) that is not imaged by a subsequent projection lens.

5.2 Fly's eye condensers

A FEC is a specific microlens array arrangement, where the lenslet surfaces are arranged in entrance/exit pairs, forming rather thick lenslets. Commonly, the lenslet pairs of FECs are laterally arranged in a regular, hexagonal or rectangular grid, and the lenslets all have equal shape. However, the principle can be realized by a reflective design as well [40], the entrance/exit surfaces in each lenslet do not need to be of equal size or shape [41], and there is considerable and useful freedom in irregular arrangements [42]. The oldest FEC reference we are aware of is [43].

Whereas mixing rods make distributions spatially homogeneous, not much affecting angular distribution, FECs make distributions angularly homogeneous, leaving the spatial distribution largely unchanged – just the other way around. In contrast to a long, narrow mixing rod, a FEC is a short and wide plate. It does work well for spatially inhomogeneous, but not too finely grained distributions, and its function is rather independent of the incident angular distribution as long as the incoming angular distribution is less wide than the *design angle* of the device, which can be up to ±15°.

The principle of operation of a FEC plate is to split the area of the inhomogeneous incident distribution into pieces or channels, to transform them individually and to recombine them in the far field. Homogeneity is achieved by the superposition of a sufficient number of different individual distributions. This implies a first design rule: the number of lenslets should be not too small.

To achieve this, the exit of each channel is formed as a lens, designed to provide an image of the channel entrance at infinity. Consequently, the shape of the channel (the shape of the lenslet), or more specifically its illuminated portion, determines the shape of the intensity distribution of the considered channel.

Any light from an adjacent channel that hits the exit lens under consideration would cause additional light in the far field *outside* the image of the entrance lenslet [44]. To avoid this crosstalk, all light that enters a channel must be *confined* in the channel, which is ensured by the refracting entrance surface which acts as a field lens.

The working principle of FECs is explained in phase space as follows (see Figure 19): Each lens performs a *separate* angle-to-area transformation. The angular distributions of all channels are added in the far field and provide a homogeneous intensity distribution. This works fine as long as a sufficient number of completely illuminated channels is involved or partially illuminated channels compensate each other.

The entrance (field) lens provides an image of the source in the area (the pupil) of the exit (projection) lens. In this sense, the operation of a channel is related to Köhler's illumination principle [20] and the term 'Köhler integrator' is occasionally used for FECs. Thus, a FEC can be viewed as a plurality of Köhler systems.

5.3 Facets on optical surfaces

Imaging a source to a target plane with a lens or a reflector is simple and efficient, but any source inhomogeneities are imaged to the target as well. A particularly bad example is shown above in Figure 10. So we need a way to disturb the optical image. Defocusing is an easy way to realize that, but it is not always reliable because of the interplay with other aberrations, and its single degree of freedom is often insufficient. A safer and more flexible method is to implement the étendue split approach introduced above by using *facets*. To avoid introducing an extra optical element, we put the facets on an existing optical surface and assign an optical function to each facet. The simplest facets are just flat, but they may have any convex or concave shape (see Figure 21).

Facets are often used on elliptical or parabolic reflectors. They can simply be applied to the originally smooth reflector surface, and each of them will then produce a different image of the source. Their superposition delivers the desired homogeneous distribution. For example, faceted

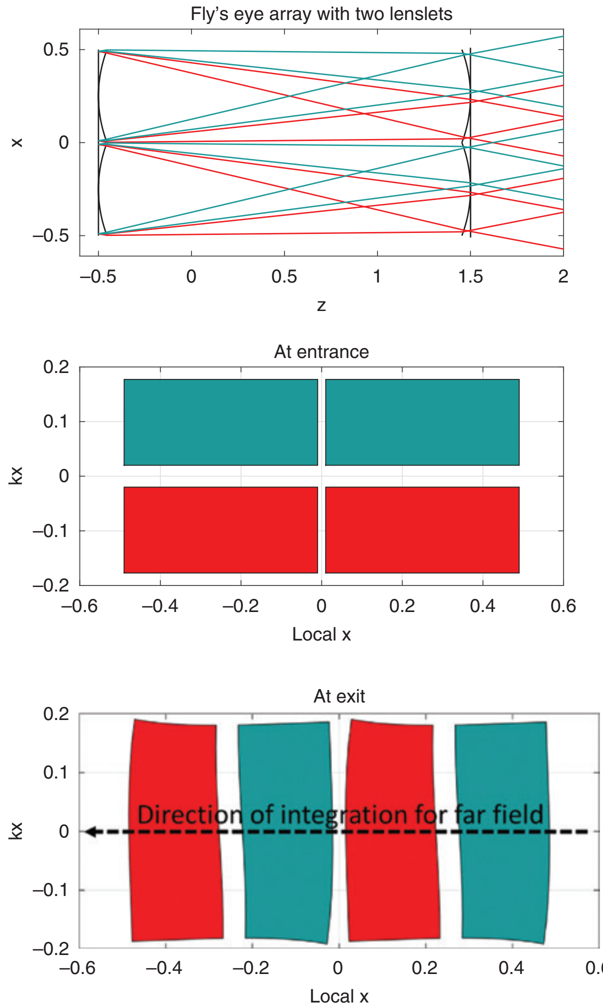


Figure 19: A FEC splits the incident distribution into pieces, rotates them in phase space and transports each one into the far field. Here, we use red rays going into lower angles and cyan rays going into higher angles, a typical use case when light coming from adjacent sources is first imaged into infinity to make it telecentric. Note that the angular intensity distribution at the exit aperture is now obtained by integration of spectral radiance over spatial dimensions, as shown by the dashed arrow. In this integration direction, the light is nearly perfectly homogeneous, except at the angular edges, due to aberrations – an effect that has been seen in Figure 20 already.

reflectors are used in the still ubiquitious MR16 halogen spot lamps, in lamp-based stage lights to provide a smooth illuminance distribution on the gate or gobo. Such reflectors are also available for the use with LEDs (Figure 22).

Faceted lenses can be used to spread (laser) pump light on the phosphor of a LARP source [22]. This technique is somewhat related to MTF shaping in a low beam projection headlamp by facets on the lens surface (an image of a knife edge by an aspheric lens is too sharp) [45]).

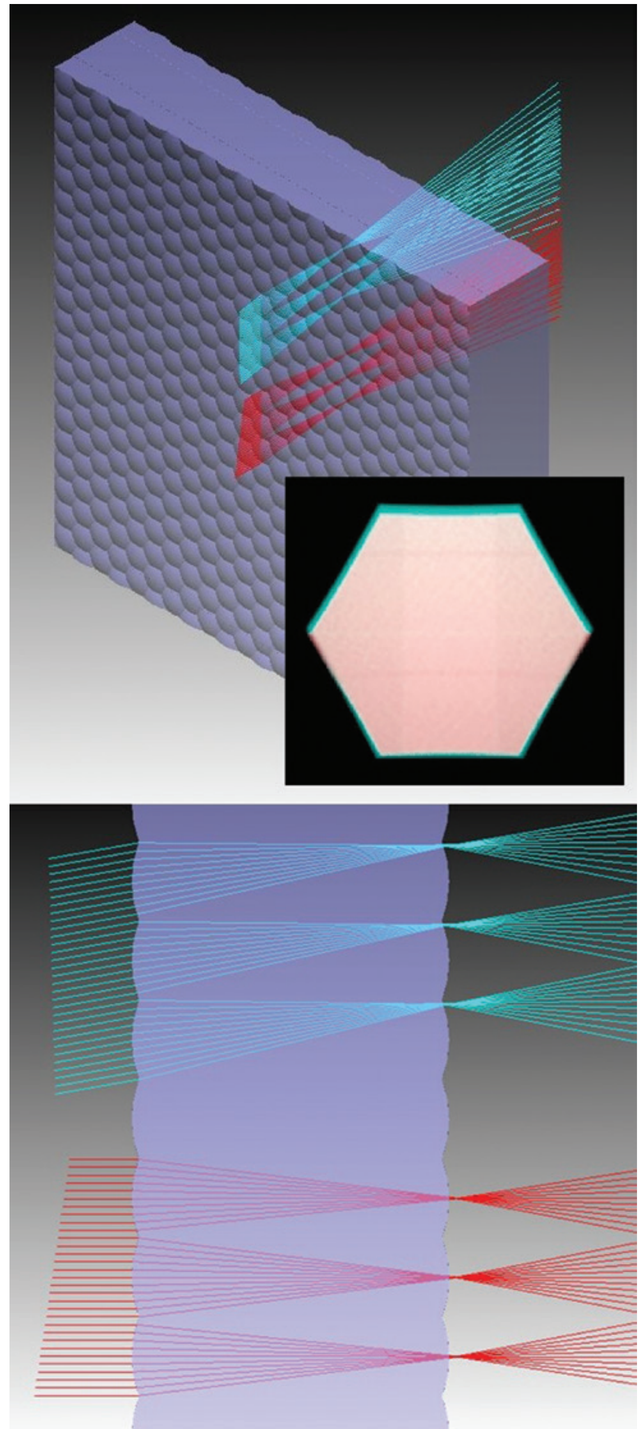


Figure 20: A FEC transforms any incident angular distribution within its design angle into an angularly homogeneous output beam. Note that the output beams (right) cover nearly the same telecentric angular range, irrespective of the incoming direction, red (0°) and cyan (10°). The bottom image shows the far field distribution obtained from the two collimated input beams. Each exit lenslet surface images the entrance hexagon into the far field. The color fringes are a result of the aberrations associated with this image. (Images created with LightTools.).

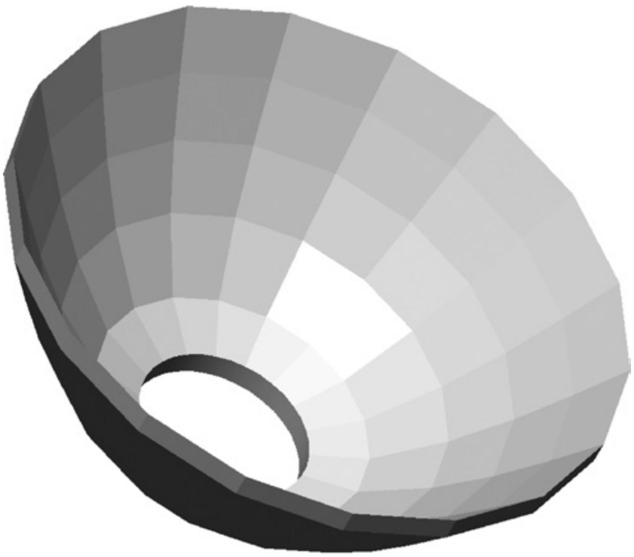


Figure 21: Simulation model of a simple reflector with planar facets. Each facet serves as an aperture, through which the virtual mirror image of the source shines its light. While there will be remaining inhomogeneities in each facet's contribution, they tend to average out.

Figure 22: Faceted reflector for LEDs. (Photo courtesy by Auer Lighting.).

Illumination design software is often assisting the user for the generation of facets on a reflector surface. The mixing effect can be tuned by the number of facets, their shape (plane or curved) and their arrangement [46]. They can also be individually tailored freeform reflectors [47], an example is shown in Figure 21.

5.4 4D homogenization

For some applications, one needs to homogenize the light distribution for a whole phase space region and not for just a cross section. For example, a stage light should have a 'nice beam' everywhere between the exit pupil and the target (image). A source which is homogeneous just in 2D may produce colored edges if an object in the beam casts a shadow. In the condenser optics within a video projector, angular separation of colors as seen from the homogeneously lit digital micromirror device (DMD) imager may lead to strange effects due to the complex interplay with the color aberrations of the projection lens; we have seen similar problems in gobo projectors.

4D homogenization is not easy to realize, especially when the system is nearly étendue limited. Except if the source is already homogeneous in two dimensions, no single optical element will suffice. A two-stage approach works well in practice: First, we find a way to illuminate an intermediate plane with spatial homogeneity, by using a rod integrator or by just aiming several collimated beams at the same target. The angular distribution will be rather inhomogeneous at this point. Then, we use a FEC to thoroughly mix the angular distribution without much increase of étendue. The incoming light at the intermediate plane must then be telecentric and limited in angular extent. Another approach would be to operate two integrators in a sequence [29], but is still difficult to get good results. The right dose of scattering can help to wash out residual artefacts.

5.5 Scattering

In illumination design, weak and strong scatterers are widely used as a tool for homogenization. If a source (with an inhomogeneous exitance distribution) is simply imaged to a target area ('critical illumination'), the source structure will appear on the target. A moderate scatterer, e.g. positioned in the exit pupil, may mitigate the problem, but may incur an increased target étendue. In principle, scattering works for any kind of sources, for all homogeneity requirements, even extreme ones. However, scattering tends to be inefficient, to increase étendue, and tends to make systems large.

Scattering causes an expansion of the *effective étendue*: it is a means to introduce an angular spread to a specific phase space distribution in a random way, i.e. without introducing obvious patterns. In phase space, the scattering of a ray (a single point in phase space) is represented by the point being spread out over the angular dimensions; in the phase space diagram, points become vertical lines (see Figure 23). A sufficient amount of angular spreading causes homogenization in the angular domain, and, after propagating some distance, in the

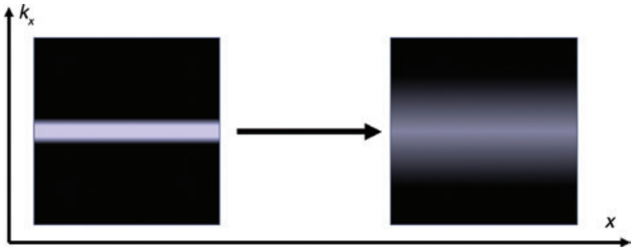


Figure 23: Scattering increases étendue and decreases luminance, but also smoothes out inhomogeneous structure (schematic image).

spatial domain as well. However, if scattering brings light into previously empty regions, it causes an increase of the source étendue.

While faceted reflectors or lenses may be viewed as étendue splitting systems (see previous section) which – in contrast to fly's eye condensers – also increase effective étendue, it is equally appropriate to view them as scatterers, if the facets are small and their number is large. However, some important points are worth noting: (i) Faceted reflectors perform what could be called deterministic scattering: The induced angular spread is given by the deterministic facet surface shape, not by random surface roughness. (ii) The phase space structures induced by faceted reflectors are small, but finite in size: When looking closely enough, the full source structure is apparent. Consequently, they work well only if any relevant quantity at the target is obtained by integration over sufficiently large areas of phase space.

5.5.1 Integrating spheres

The most extreme scattering configuration is the *integrating sphere*, a hollow sphere, whose inside is covered with a diffuse white reflector. (In German, the 'Ulbrichtkugel' is named after Richard Ulbricht, a railway engineer dealing with the illumination of train stations around 1900). Integrating spheres are used to homogenize an incident light distribution for use with sensors, but also to generate a spatially and angularly homogeneous light beam for whatever purpose. For example, an integrating sphere illuminates the film gate in a professional movie film scanner with red, green, blue and infrared light [48]. At the heart of integrating sphere theory is the fact that a small patch of a fully diffuse scatterer on the inner surface of a hollow sphere acts as a Lambertian source, and will illuminate the hollow sphere with perfect homogeneity: incidence angles, the cosine intensity distribution of Lambertian sources and source-target distances all compensate each other. Thus, after the second surface interaction, the incoming light has completely 'forgotten' where it came from.

However, designing good integrating spheres is not trivial. Sphere size, location and size of entrance and exit ports, internal baffles, reflectivity and diffuseness of the white coating all play important roles. The key disadvantage of an integrating sphere used for illumination is the fact that efficiency, homogeneity and small phase space expansion are conflicting targets. For more information on design, theory, history and application of integrating spheres, see [49–53].

When used for sensing, the integrating sphere is a means to measure just the flux, independent of any specific angular/spatial distribution of the incident light. Instead of sending light from the outside, a light source may be placed in or near the center of the sphere, and the sensor on the inner surface.

For particular measurement purposes of collimated sources, such as for large discharge lamps in parabolic reflectors, one can avoid to use of a large and expensive integrating sphere, but use just one or two scatterers in a sequence in front of a receiver. With the right calibration, such a setup may be good enough (see Figure 24).

5.5.2 Other scattering elements

An efficient way is to combine a little bit of scattering with other approaches: Often, other ways of cutting phase space into pieces are just not really good enough, or just too sensitive to tolerances. Then, one would look for a location (i.e. a surface, whether real or in free space) in the system, from where the pieces of the inhomogeneous source look small, i.e. subtend a small angle, and are close to each other. In this situation, just a little bit of scattering, with small scattering angles, will make the output homogeneous (see Figure 25), and make the system less sensitive to tolerances.

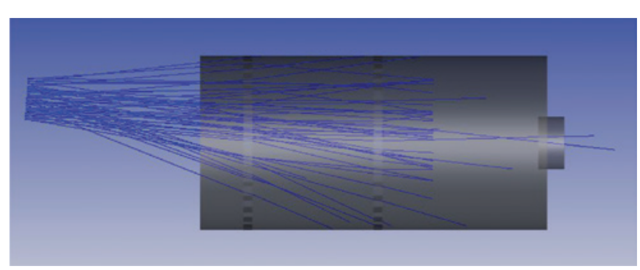


Figure 24: The 'jar': two scatterers homogeneously distribute the incident light, and a detector (right) collects a representative portion of incident light. With the right calibration, such a cheap device works nearly as good as an integrating sphere.

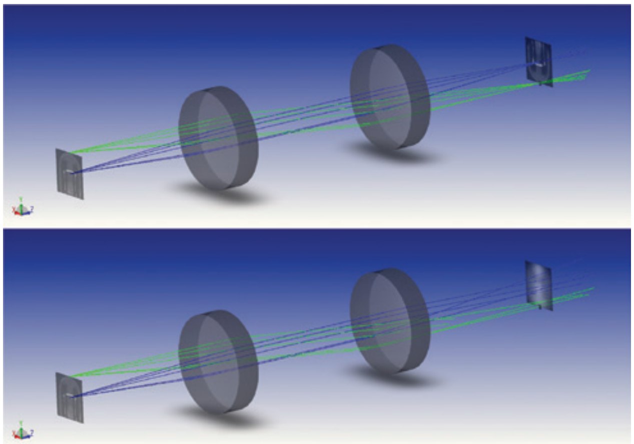


Figure 25: Scattering implemented (on the second lens) to smoothen a source image.

In yet another approach, scattering light back onto the source itself can be used to increase the luminance of the source beyond what it would emit by itself. Several examples of this approach are described in [54]. Although not intended primarily for color mixing, such systems tend emit their light more homogeneously than the sources alone.

6 Summary

Color mixing and homogeneity are major challenges in many illumination designs. However, it is easy to get lost in the labyrinth of often not well-defined requirements, optical principles, optical elements, and light sources. We have presented the phase space view, with its core notion of volume (étendue), as a common, high level viewpoint from where similarities and differences become clear, and from where it is possible to define *design patterns*: general, reusable solution approaches to a commonly occurring problem within a given context. In phase space, optical elements used for color mixing all perform one of the core functions: (i) propagation induces shear in phase space, (ii) smooth optical surfaces transform the shape of phase space regions in many ways, (iii) scattering surfaces, including single surfaces with deterministic micro structures, expand étendue, and (iv) surfaces with kinks or other discontinuities split phase space regions into independent parts. Equipped with these insights, individual optical elements like rod integrators, FECs, faceted and scattering can all be understood in terms of their optical function in phase space, filling the toolbox of the optical designer with a full set of useful and well-understood tools. Color mixing then is a problem of first understanding the given phase space structure of the source, and the desired phase space structure of the target, and then choosing appropriate elements to perform the necessary functions of transforming the former into the latter.

Acknowledgment: The authors would like to thank OSRAM Opto Semiconductors and Auer Lighting for images provided.

References

[1] E. Gamma, R. Helm, R. Johnson and J. Vlissides, in 'Design Patterns: Elements of Reusable Object-Oriented Software', (Addison-Wesley, Boston, MA, USA, 1994).
[2] A. Herkommer, Opt. Eng. 53, 031304 (2013).
[3] M. Testorf, B. Hennelly and J. Ojeda-Castaneda, in 'Phase-Space Optics: Fundamentals and Applications', (McGraw Hill, New York, NY, USA, 2010).
[4] R. Winston, J. C. Miñano and P. Benitez, in 'Nonimaging Optics', (Elsevier Academic Press, London, UK, 2005).
[5] R. Winston, L. Jiang and M. Ricketts, Adv. Opt. Photonics 10, 484–511 (2018).
[6] J. Chaves, in 'Introduction to Nonimaging Optics', (CRC Press, Boca Raton, FL, USA, 2016).
[7] G. Brooker, in 'Modern Classical Optics', (Oxford University Press, Oxford, UK, 2003).
[8] A. Krasnoshchoka, A. Thorseth, C. Dam-Hansen, D. Corell, P. Petersen, et al., Materials 10, 1407 (2017).
[9] O. Mehl and M. Daniels, European Patent Application EP3045950A1 (2016).
[10] P. Benitez, J. C. Minano, P. Zamora, M. Hernandez and A. Cvetkovic, 'Kohler Concentrator', United States Patent Application US20100123954A1 (2010).
[11] O. Dross, R. Mohedano, M. Hernández, A. Cvetkovic, J. C. Miñano, et al., Proc. SPIE 71030G (2008).
[12] R. Bader, S. Haussener and W. Lipinski, J. Sol. Energy Eng. 137, 021012 (2014).
[13] N. Morgenbrod, 'Solar Simulator and Method for Operating a Solar Simulator', United States Patent Application US 2013 0294.045A1 (2013).
[14] H. Rehn and U. Hartwig, Proc. SPIE 77850F (2010).
[15] J. Muschaweck and H. Ries, Proc. SPIE 594206 (2005).
[16] L. Shapiro, Proc. SPIE 2650, 209–216 (1996).
[17] 'Information Display Measurements Standard', Society for Information Displays, https://www.icdm-sid.org/downloads/idms1.html (2012).
[18] T. Q. Khanh, P. Bodrogi and T. Q. Vinh, in 'Color Quality of Semiconductor and Conventional Light Sources', (Wiley VCH, Weinheim, Germany, 2017).
[19] H. Rehn, Proc. SPIE 4775 (2002).
[20] A. Köhler, Z. Wiss. Mikrosk. 10, 433 (1893).
[21] H. Rehn, Opt. Eng. 43, 1480 (2004).
[22] M. Daniels and I. Schmidt, 'Focus when Laser Light Sources', German Patent Application DE102016218245A1 (2018).

[23] K. Franz and U. Hartwig, 'Lighting Apparatus with Phosphor Element', United States Patent Application US 2013 0329448A1 (2013).
[24] S. Magarill, 'Apparatus for Uniformly Illuminating a Light Valve', United States Patent USOO5625738A (1997).
[25] M. E. Levis and H. Gourley, 'Projector System with Hollow Light Pipe Optics', United States Patent USOO5829858A (1998).
[26] B. Wagner, 'Hohlintegrator', United States Patent US6625380B2 (2003).
[27] M. Angelini, C. Bigliati and E. Grossi, 'Light Mixing Lenses and Systems', PCT Patent Application WO2013098387A2 (2013).
[28] T. Vinther, 'Illumination System Comprising an Optical Light Mixing Rod and a Pressure Mechanism to Apply a Pressure Force to the Optical Light Mixing Rod and to Press an Entrance Surface and Emitting Window Together', United States Patent US9612380B2 (2017).
[29] H. Rehn, Proc. SPIE 105901L (2017).
[30] H. Rehn, Proc. SPIE 5529, 157 (2004).
[31] J.-W. Pan, C.-M. Wang, W.-S. Sun and J.-Ya. Chang, Appl. Optics 46, 5097–5102 (2007).
[32] W. J. Cassarly, Proc. SPIE 710307 (2008).
[33] C. F. Dimas, S. Read and J. J. Kuta, Proc. SPIE 4768, 82–92 (2002).
[34] W. J. Cassarly, 'Rippled Mixers for Uniformity and Color Mixing', United States Patent US7777955B2 (2010).
[35] T. S. Bonenberger, J. Baumgart and C. Neumann, Adv. Opt. Technol. 5, 157 (2016).
[36] M. Kuwata, H. Takeuchi, T. Sasagawa, S. Yagyu and H. Sugiura, J. Soc. Inf. Display 15, 829 (2007).
[37] F. Potekev, 'Morphing Light Guide', United States Patent Application US 2006004.4531A1 (2006).
[38] T. L. R. Davenport, W. J. Cassarly and T. A. Hough, Proc. SPIE 594204 (2005).
[39] D. Schmidt, C. Symanowski and G. Schroeter, 'Device for Generating a Quadrangular Illuminating Field and use of such Device in an Optical Device Comprising a Surface to be Illuminated having a Predetermined Shape', United States Patent USOO6578999B2 (2003).
[40] B. G. Crowther, D. G. Koch, J. M. Kunick, J. P. McGuire, R. Harned, et al., Proc. SPIE 4832, 302–310 (2002).
[41] B. A. Jacobson, R. D. Gengelbach and J. M. Ferri, Proc. SPIE 3139, 141–150 (1997).
[42] J. Muschaweck, Proc. SPIE 79540A (2011).
[43] J. Mihalyi, 'Projection System for Color Pictures', United States Patent 1,762,932 (1930).
[44] P. Schreiber, S. Kudaev, P. Dannberg and U. D. Zeitner, Proc. SPIE 59420K (2005).
[45] L. Wang, J. M. Sasián, P. Su and R. John Koshel, Proc. SPIE 74230Y (2009).
[46] W. J. Cassarly, S. R. David, D. G. Jenkins, A. P. Riser and T. L. R. Davenport, Opt. Eng. 39, 1830–1840 (2000).
[47] W. Pohl, C. Anselm, C. Knoflach, A. Timinger, J. Muschaweck, et al., Proc. SPIE 5186 (2003).
[48] ARRI Group, 'ARRISCAN XT', https://www.arri.com/archive_technologies/arriscan, (2018).
[49] R. Ulbricht, in 'Das Kugelphotometer: (Ulbricht'sche Kugel)', (R. Oldenbourg, Germany, 1920).
[50] D. G. Goebel, Appl. Optics 6, 125–128 (1967).
[51] Labsphere, 'Technical guide integrating sphere uniform light source applications', http://labsphere.com/site/assets/files/2552/a-guide-to-integrating-sphere-uniform-light-source-applications.pdf (2018).
[52] Labsphere, 'TECHNICAL GUIDE integrating sphere theory and applications', http://labsphere.com/site/assets/files/2551/integrating_sphere_theory_apps_tech_guide.pdf (2018).
[53] Labsphere, 'A guide to integrating sphere radiometry and photometry', http://labsphere.com/site/assets/files/2550/radiometry_and_photometry_tech_guide-1.pdf (2018).
[54] J. Chaves, W. Falicoff, P. Benitez, J. C. Miñano, W. A. Parkyn, et al., Proc. SPIE 63380L (2006).