Metal Halide Lighting Systems and Optics for High Efficiency Compact LCD Projectors

Benjamin A. Jacobson[a], Robert D. Gengelbach[a], Charles N. Stewart[b], and Douglas M. Rutan[b]

[a]Illumitech Inc., 5319 S. Dorchester Ave., Chicago, IL 60615-5307
[b]Welch Allyn, Inc., Lighting Products Div., 4619 Jordan Rd., Skaneateles Falls, NY 13153-0187

ABSTRACT

Compact LCD projectors require a high efficiency light source that has the smallest possible spatial extent. Further, they require optical systems that preserve the étendue. Current projector illumination systems have aberrations that produce a light beam whose étendue far exceeds the intrinsic étendue of the light source itself. As a result both efficiency and uniformity fall short of what is theoretically possible. We provide a theoretical framework for understanding these aberrations and the magnitude of their effect. We also present results showing the efficiency, uniformity, and other performance gains which are possible when these aberrations are corrected. This work also describes the performance of long-life, short-arc metal halide lighting systems that are able to increase screen brightness of compact projectors several fold without any increase in system power or heat. With these systems it has been possible to design and validate lamps operating at 50 Watts, producing >3,000 lumens and having excellent lumen maintenance throughout their 4,000 hour life. The benefits of the combination of an improved étendue-preserving optical system and a short-arc metal halide lamp will be demonstrated.

Keywords: projection, LCD, lamp, light source, illumination, non-imaging optics

1. INTRODUCTION

The design of compact, low-cost projection display devices presents complex trade-offs. At the heart of these trade-offs is the illumination system which couples the light source to the light valve and projection lens. We will describe a few key system parameters that often constrain the design team and point out options in the selection of the lamp and collection optics that meet these constraints. We want to focus particularly on the importance of matching lamp and collection optics to the rest of the projector system. We will explore these issues using a theoretical model that describes the performance limits of standard classical, non-faceted ellipsoidal reflectors assuming a lamp with fixed arc length. We will then show how additional performance gains are achieved using non-classical non-imaging optical design techniques to improve the basic efficiency vs. étendue, and also by optimizing arc length for a given system.

Several authors have recently introduced the concept of light-source étendue into discussions of projector illumination efficiency, and this concept is now well accepted. Brennesholtz, in particular, has given an excellent description of the way étendue limits efficiency as luminous power is transferred through an optical system[1].

Étendue is defined as area x solid angle. For circular beams output by an ellipsoidal reflector,

$$\begin{aligned} Area &= \pi r_{\max}{}^2 \\ Solid\ Angle &= \pi \sin^2 \eta_{\max} \\ H &= \pi^2 r_{\max}{}^2 \sin^2 \eta_{\max} \end{aligned} \tag{1}$$

In the simplest projector configuration, no homogenizer or lens array is used to correct the shape mismatch between the circular illumination beam and the rectangular light-valve. In this case the circular beam must fill the light-valve diagonal and the étendue is defined by the light-valve diagonal d_{LV} and the projection lens f/#:

$$H = \pi^2 \left(\frac{d_{LV}}{4(f/\#)} \right)^2 \quad (2)$$

Typical étendue values given by this equation for compact projectors range from 6.9 mm^2sr for a 10 mm reflective display at f/3, to 168 mm^2sr for a 1.3" polysilicon LCD at f/2, but most systems will fall in the range of 25-75 mm^2sr.

Actual étendue included by the rectangular beam is lower by roughly a factor of 0.6-0.7. If a light-pipe homogenizer is used to correct the shape mismatch, then the illumination beam diameter is typically smaller by a factor of 0.7 to 0.8 but there is little or no shape-mismatch loss. In this paper we will consider only circular beams, and any shape mismatch factors must be applied separately.

2. EFFICIENCY VS. ETENDUE USING CONVENTIONAL OPTICAL SYSTEMS

It is critical to note that the practical source étendue is that of the output beam of the collection optics system, not of the source itself. It is true, as some workers have emphasized, that the final "thermodynamic limit" for collection efficiency is the étendue ratio of output to the arc source; that is,

$$E \le \frac{H_{out}}{\pi^2 Z d_{arc}} \quad (3)$$

where Z is the length and d_{arc} is the diameter of a cylindrical arc volume. It has not always been appreciated, however, that in practice Equation 3 is hardly useful for quantitative work, because practical collection optics systems do not nearly approach this limit.

A much more sophisticated treatment was pioneered by Brennesholtz, who explicitly included collection optics effects via a semi-empirical quadratic dependence of output étendue vs. arc size for a particular reflector.[1] However, this semi-empirical approach offered no guidance in applying the results to different reflectors, and provided no insight as to the aberration mechanism. We want to derive a more general approximate expression for collection efficiency vs. output étendue for a wide range of reflector designs, and especially for design parameters common in compact projectors. Finally, we want to derive the performance limits for an improved collection system based on non-imaging optics.

For these relative efficiency comparisons, in order to simplify the theory and provide greater insight we will use a highly idealized arc modeled as a pure cylinder. *Therefore the theoretical efficiencies that appear in the Figures in Sections 2 and 3 are relative efficiencies and should be used only to compare different optical designs. Real metal halide lamps have a significant fraction of their energy in a surrounding halo, and therefore have significantly lower efficiency for small output étendue.* Work in our labs has shown that, while the real lamp trade-offs are more complex, the trends vs. reflector design are quite similar, albeit at lower overall efficiencies. Section 4 will show the actual efficiencies obtained with the Welch Allyn 50 W Sõlarc™ metal halide lamp.

2.1 Sources of aberrations in conventional collection optics

We will see that even the best ellipsoid increases the étendue by roughly a factor of 4-6 over the limit of Equation 3. In fact, all conventional systems (not just ellipsoidal reflectors) fail to approach the limit of Equation 3 unless they sacrifice significant efficiency. This failure occurs for two reasons. First, the transition from a cylindrical source to a planar output entails a significant increase in étendue in order to achieve high efficiency with any rotationally symmetric optical system. This effect arises from "skew conservation" (comparable to conservation of angular momentum) and has been described in detail by Ries et al.[2] The end result, in the absence of significant losses, is to increase the étendue by a factor of Z/d_{arc}, or approximately 2 for most short-arc lamps.

The remaining factor of 2-3 increase in étendue results from the aberrations intrinsic to any bowl-shaped reflector which surrounds an arc. We can see this effect at work using an ellipsoid. A conventional ellipsoid-based optical system is shown in Figure 1. We want to calculate efficiency vs. étendue for this system. To simplify the calculation we will assume reflectivity = 1, and use $\phi_1 = 0$. This assumption does not affect relative comparisons, since the various approaches are affected similarly by reflectivity and back-hole losses. Assuming the entire beam fits on the target, the efficiency is just the well-known expression for solid angle weighted by the angular distribution of a Lambertian cylinder:

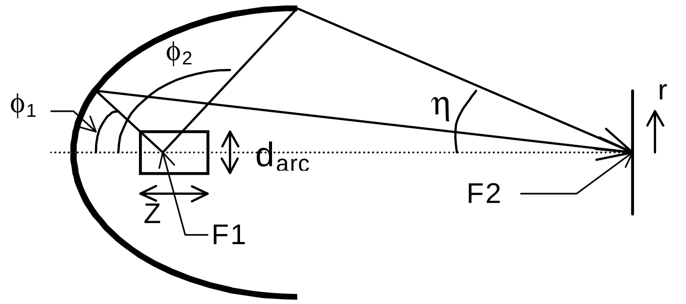

Figure 1. Sketch of reflector (first focus, F1, and second focus, F2) containing the radiating arc (diameter, d, and length, h) at the first focus.

$$\varepsilon = \frac{\phi_2(r) - \dfrac{\sin 2\phi_2(r)}{2}}{\pi} \tag{4}$$

Next we need to calculate the étendue of the ellipse output. The calculation is simplified by using polar coordinates R,ϕ, with the origin at F1. In these coordinates the equation for an ellipse is

$$R(\phi) = \frac{f(1+e)}{1+e\cos\phi} \tag{5}$$

where f is the focal length (distance from F1 to the vertex) and e is the eccentricity. Note that for an ellipsoid $0<e<1$, while $e=0$ for a hemisphere and $e=1$ for a paraboloid. The quantities f and e are related to the more familiar semi-major and semi-minor axes a and b by:

$$e = \sqrt{1-\frac{b^2}{a^2}} \qquad \text{and} \qquad f = a(1-e) \tag{6}$$

To calculate the étendue we need to find r_{max} and η_{max}. We can find expressions for these quantities by considering Figure 2. For small δ we have

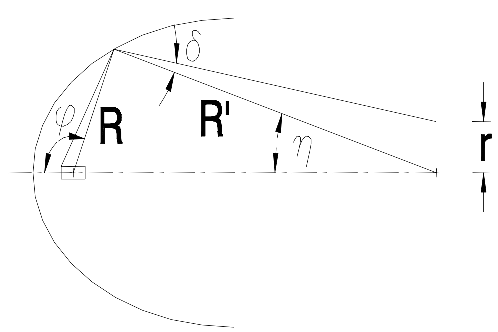

Figure 2. Relevant distances and angles

$$r(\phi) \cong \frac{\sqrt{Z^2 + d_{arc}^{\ 2}}}{2}\,\frac{R'(\phi)}{R(\phi)}\,\frac{\sin\left(\phi + \arctan\left(\dfrac{d_{arc}}{Z}\right)\right)}{\cos\eta(\phi)} \tag{7}$$

and

$$\sin\eta(\phi) = \frac{R(\phi)}{R'(\phi)}\sin\phi \tag{8}$$

(Note that Equation 7 does not hold as η approaches 90°, where $r(\phi)$ becomes infinite, but typical projection ellipsoids do not extend to this depth anyway.) Using the fact that $R + R'$ is constant, we can show that

$$\frac{R(\phi)}{R'(\phi)} = \frac{1-e^2}{1+e^2+2e\cos\phi} \tag{9}$$

Note that the beam étendue is defined by the *maximum* value of r and the *maximum* value of $\sin(\eta)$ as ϕ is varied from ϕ_1 to ϕ_2. The critical point is that these two quantities maximize at quite different values of ϕ. Therefore the magnification factor $R'(\phi)/R(\phi)$ does not cancel when we take the product of r_{max} and $\sin \eta_{max}$. The variation in $R'(\phi)/R(\phi)$ is small for small values of e, but then the $\cos \eta(\phi)$ in Equation 7 becomes a prominent factor. *This is the fundamental source of the étendue distortion caused by an ellipsoid.* Note further that it is not a property of an ellipsoid alone, but of any similar reflector where R'/R varies strongly with ϕ and which focuses light from the source onto a planar aperture. Note also that the output étendue is not reduced by increasing the reflector size so that the source resembles a point source by comparison—in fact, the output beam étendue is completely independent of reflector focal length (F1 to vertex distance) to first order.

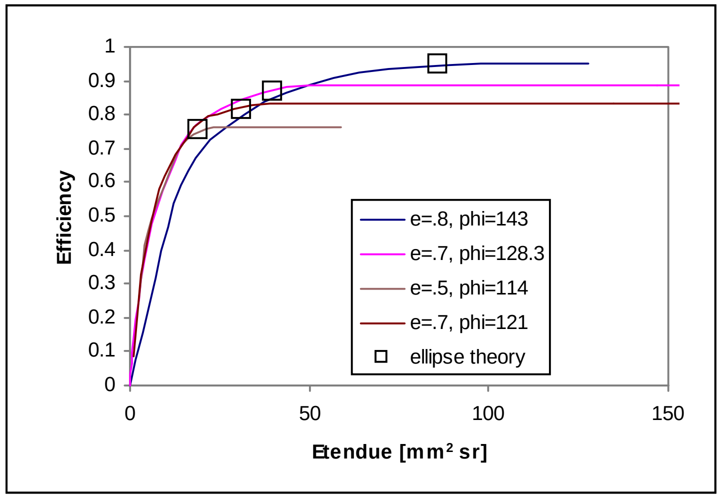

Figure 3. Comparison between analytic theory and ray-traces for specific reflector geometries.

3.2 Performance limits of conventional collection optics

We can now extend Equations 6 and 7 to derive an approximate expression for efficiency vs. étendue as a function of the reflector properties e and ϕ_2. The value of ϕ which maximizes $r(\phi)$ is a function of e, but for all relevant values of e it is between 45° and 50°. Therefore we can set $r_{max} = r(\phi=45°)$, and combine Equations 7 and 9 to find

$$r_{\max} \cong \frac{\sqrt{Z^2 + d_{arc}^{\;2}}}{2} \cdot \frac{1+e^2+\sqrt{2}e}{\left(1-e^2\right)} \cdot \frac{\sin\left(45^\circ + \arctan\left(\frac{d_{arc}}{Z}\right)\right)}{\cos\eta(45^\circ)} \tag{10}$$

Eqns 8 and 9 can also be combined to show, with considerable rearranging,

$$\cos\eta(\phi) = \frac{2e + \left(1+e^2\right)\cos\phi}{1+e^2+2e\cos\phi} \tag{11}$$

which can be substituted into Eqn 10 to yield an expression for beam size.

$$r_{\max} \cong \sqrt{Z^2 + d_{arc}^{\;2}}\,\sin\left(45^\circ + \arctan\left(\frac{d_{arc}}{Z}\right)\right)\frac{\left(1+e^2+\sqrt{2}e\right)^2}{\left(1-e^2\right)\left(\sqrt{2}\left(1+e^2\right)+4e\right)} \tag{12}$$

Next we need to find the output angle η, which is clearly maximized at $\phi = \phi_2$.

$$\sin\eta_{\max} = \frac{1-e^2}{1+e^2+2e\cos(\phi_2)}\sin\phi_2 \tag{13}$$

Equation 12 gives the absolute largest radius required to accept all the rays from a cylindrical source. In practice we can accept almost all the rays with a somewhat smaller radius. Comparison with ray-trace results for a pure cylindrical source shows an excellent match with a radius = 90% of the value in Equation 12. We can combine this "fudge factor" with Equations 12 and 13 can to yield a complete expression for étendue vs. arc size, eccentricity, and rim angle.

$$H_{ellipse} = \pi^2 (0.9)^2 \left(Z^2 + d_{arc}{}^2\right) \cdot \sin^2\left(45^\circ + \arctan\left(\frac{d_{arc}}{Z}\right)\right) \cdot \frac{\left(1 + e^2 + \sqrt{2}e\right)^4}{\left(\sqrt{2}\left(1 + e^2\right) + 4e\right)^2} \frac{\sin^2 \phi_2}{\left(1 + e^2 + 2e\cos\phi_2\right)^2} \tag{14}$$

Figure 3 compares four efficiency vs. étendue points described by Equations 4 and 14 with ray-trace results for Lambertian cylinders. The equation does an excellent job of identifying the knee on each curve—that is, the point where these reflector parameters are likely to be optimal.

In principle we could also use Eqns 4 and 14 to plot efficiency vs. étendue for a variety of values of e and ϕ_2, and find the optimal combination of e and ϕ_2 for each value of the eccentricity. In practice, however, ϕ_2 is limited by the difficulty of molding ellipsoids deeper than a hemi-ellipsoid. Moreover, Equation 19 takes a particularly simple form for hemi-ellipsoids. Substituting the hemi-ellipsoid condition $\cos\phi_2 = -e$ into Equations 4 and 14, our expression for optimal efficiency vs. étendue is

$$H_{opt} = \pi^2 (0.9)^2 \left(Z^2 + d_{arc}{}^2\right) \cdot \sin^2\left(45^\circ + \arctan\left(\frac{d_{arc}}{Z}\right)\right) \cdot \frac{\left(1 + e^2 + \sqrt{2}e\right)^4}{\left(\sqrt{2}\left(1 + e^2\right) + 4e\right)^2 \left(1 - e^2\right)}$$

$$E_{opt} = \frac{\arccos(-e) + e\sqrt{1 - e^2}}{\pi} \tag{15}$$

Figure 4 shows a plot of ray-trace results for a wide variety of ellipsoids, plotted in terms of efficiency vs. étendue. The solid line is the locus of points traced by Equations 15 for e ranging from 0.1 to 0.9. This curve clearly also traces the envelope of highest efficiency found in the ray-trace for each étendue value, and therefore is the ellipsoid performance limit within the restrictions of our idealized model.

3. IMPROVED COLLECTION EFFICIENCY VIA NON-IMAGING OPTICS

3.1 Improved optical efficiency with fixed lamp

Now we want to calculate the maximum collection efficiency vs. étendue if we remove the constraint of a conventional reflector. Even with non-imaging optics, we cannot approach the thermodynamic limit of Equation 3. Typical projection systems target collection efficiencies of 90% or more. Consider the system in a plane containing the optical axis. In order to collect all the reflected rays we must collect all the rays in this plane. But for these in-plane rays, the 2-dimensional étendue of the target cross-section must at least equal the 2-dimensional étendue of the source cross-section[3]. This limit is found by imagining that, instead of rotationally symmetric sources and reflectors, we are using a slab source and trough reflector having the cross-sections shown in Figures 1 and 2. If our design works

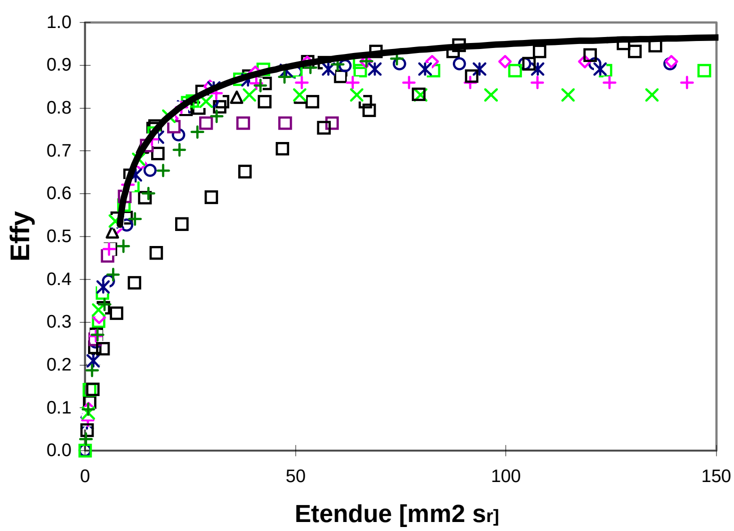

Figure 4. Optimal efficiency vs. etendue. Solid line is locus of points given by Equation 15. Symbols are ray-trace data for a large variety of reflector parameters

for all in-plane rays, then it is subject to the étendue limits for the slab-trough system:

$$r \sin \eta_{\max} \geq h \tag{16}$$

where h is the arc length. We further assume that, although we are not using a conventional elliptical or parabolic reflector, our collection angle is still bounded by angles ϕ_1 and ϕ_2. If we are concentrating the rays within that 2-D collection angle with maximum theoretical effectiveness, then the 2-D target étendue ($r \sin\eta$) is just the 2-D source étendue (h) multiplied by the 2-D solid angle efficiency:

$$r \sin \eta_{\max} \geq h \frac{\cos \phi_1 - \cos \phi_2}{2} \tag{17}$$

This limit is known as the "edge-ray limit" because the 2-dimensional, in-plane collection problem can often be solved by non-imaging optical methods known as "edge-ray" methods.

We can calculate the edge-ray limit efficiency vs. étendue curve by combining Equations 1, 4, and 17. The resulting limit is plotted in Figure 5, along with the ellipsoid performance curve from Figure 4. Both curves reach their maximum efficiency at 95% because we have limited ϕ_2 to 145° in order to limit the size of the reflector. The dashed vertical lines in Figure 5 show the range of étendue values for a circular beam with diameter equal to the diagonal of typical new-generation small light-valves, with angles corresponding typical projection lens f/#'s.

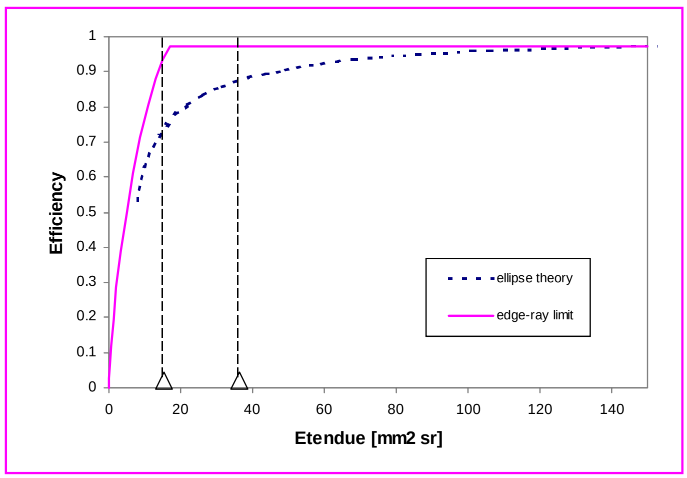

Figure 5. Theoretical limits of non-imaging vs. conventional collection optics. Vertical lines show etendue range bounded by 10 mm diagonal light-valve with f/2 lens and a 23-mm light-valve with f/3 lens.

3.2 Improved lamp efficiency enabled by optical improvements

Figure 5 shows the efficiency improvements possible by decreasing the étendue alone. Because the efficiency curve is so shallow at larger étendue values, however, large reductions in étendue yield fairly small efficiency increases. In effect, in this étendue range the target area and angles are large enough to accommodate the ellipsoid aberrations and little is gained by better optical design alone.

One alternative is to redesign the lamp to match the target, given the improved étendue. This approach yields significant improvements because the luminous efficacy of the lamp increases with the arc gap. The impact of gap size has been discussed before[4]. The overall efficiency of the lamp, η_L, is a function of arc gap, Z:

$$\eta_L = \eta_{pos} \left(\frac{1}{1+\rho} \right) \tag{18}$$

where η_{pos} is the efficacy of the arc in the positive column between the electrodes.

$$\rho = \frac{V_C}{ZE_z + V_C} = \frac{1}{1 + \frac{ZE_z}{V_C}} \tag{19}$$

where V_C is a constant voltage associated with the physical transition of the arc to the electrodes and E_Z is the constant electric field of the arc plasma. The current passing through this constant voltage, V_C, results in a non-radiative power loss. The effect of this loss of power at the ends of the arc is more detrimental to short arcs than long arcs. For very long arcs, the relative end power loss is small and the lamp efficacy, η_L, approaches that of the positive column efficacy, η_{pos}. Typical values for η_{pos}, V_C, and E_z for our 50W miniature metal halide lamp are 85lm/W, 15 V, and 33 V/mm, respectively. The effective lamp efficacy is now seen to be functionally dependent on arc gap as shown in Figure 6. Therefore, if the constraint on arc gap can be relieved early in the optical design cycle as a result of overall system étendue, the impact could be improvement in lamp design. The shift in the trade-off curves using this approach is illustrated in Figure 7. This graph shows the same comparison as Figure 5, but now the designer has chosen to exploit the more efficient optics by increasing the arc gap. Figure 7 shows total system efficacy (efficiency x luminous efficacy) instead of just the geometric efficiency shown in Figure 5. With the larger arc gap, the non-imaging system underperforms at small étendue values, but outperforms the conventional system at larger values.

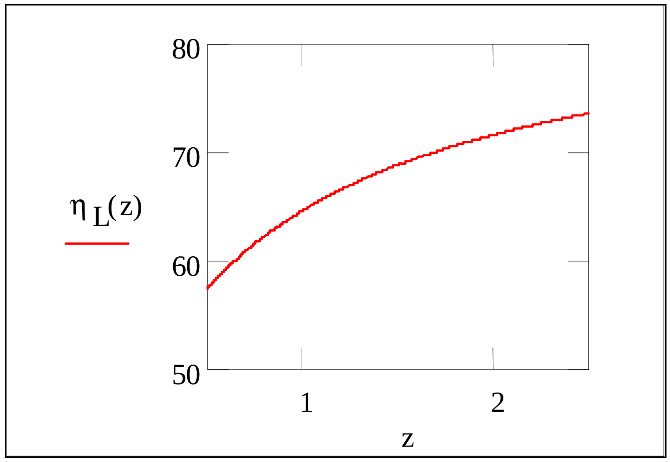

Figure 6. Lamp Efficacy vs. Arc Gap

Longer arc gaps can not only be traded off for improved lamp efficacies, they can also allow the lamp designer flexibility in rated lamp life. The positive column efficacy, η_{pos}, is generally proportional to thermal stresses on the lamp and therefore inversely proportional to the life performance of the lamp. If light output exceeds requirements for a specific arc gap, for example, the lamp designer could leave the arc gap fixed, reduce the efficacy thereby reducing the thermal stresses and increase the life.

Given the selection of panel size and numerical aperture on source size requirements, the system still has a luminosity requirement. The luminous efficacy, Equation 4, is a general formulation for lamp performance and all that remains is selection of the lamp wattage to achieve the desired lumens.

$$W_L = \frac{L}{\eta_L} \qquad (20)$$

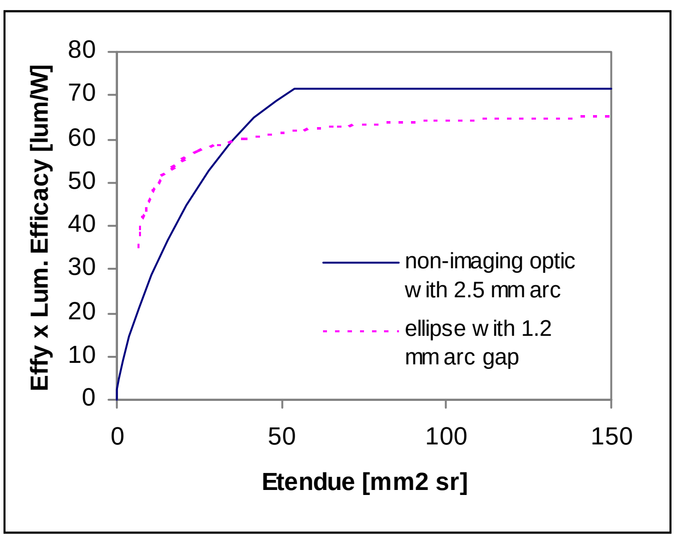

Figure 7. Total system efficacy comparison using larger arc gap with non-imaging collection

4. EFFICIENCY VS. ETENDUE FOR WELCH ALLYN 50 W SÕLARC™ LAMP

The previous discussion has focused on the benefits of a "bottom-up" design of a lamp, including exact selection of arc gap, lamp life, and lamp wattage given the optical constraints of reflector, panel diagonal, projection NA, and lumen output. While this approach produces the best performance, design and validation of a new arc tube can be costly with respect to both time and money. Particularly in the development phase, it is sometimes more expedient to optimize the reflector design and selection for a given lamp. Sections 2 and 3 have discussed relative efficiency effects of different design approaches and parameter choice, but did not discuss absolute efficiencies.

We have characterized the Welch Allyn 50 W Sõlarc arc tube with a number of reflectors shown in the table below. Note that the Welch Allyn tube design allows small reflector diameters, which minimize the overall size of the resulting system. The table below indicates a variety of custom reflectors and designs for device illumination. These are variants of smooth elliptical designs with variations in the resultant beam divergence (NA), working distance, and source magnification.

Reflector #	**EXP0950**	**EXP1090**	**EXP1091**
Size	MR16	MR16	MR16
Reflector Diameter (mm)	51	51	51
Rim to F2 (mm)	42.7	23.7	28.8
Numerical Aperture (NA = sin η_{max})	0.46	0.69	0.60
Magnification	6.9	5.7	4.6

Figure 8 shows the resulting lumens for each reflector design as a function of aperture diameter. Figure 9 shows the same data as a function of output étendue. Note that absolute efficiencies are significantly below the geometric efficiencies in Sections 2 and 3. This shortfall results from the fact that the main cylindrical arc region produces only about 50% of the total arc lumens. The remaining lumens are emitted by the surrounding halo. This explanation has been confirmed by direct imaging of the arc, and also by ray-traces which reproduce the data in Figure 8 using a more realistic arc model.

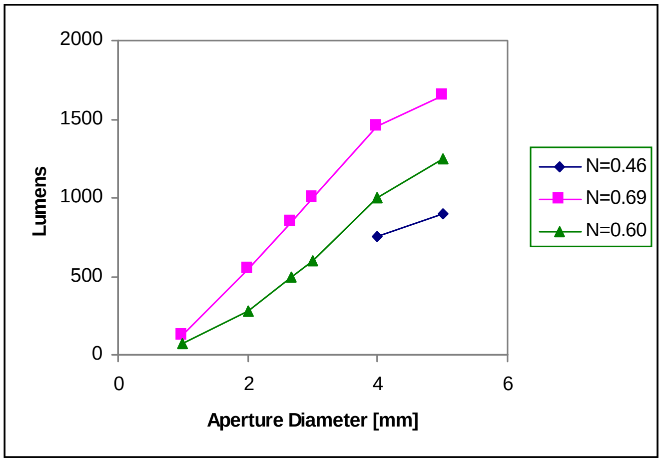

Figure 8. Lumen output vs. aperture diameter for various reflectors

5. CONCLUSIONS

Designers of optical systems for projection display products are faced with a difficult job. A basic understanding of how the lamp design interacts with the optical components can improve component selection at the outset of a project. Start with a lamp that demonstrates small source size, long life, and high efficiency. Then design an optimum reflector to maximum the amount of light into a numerically constrained panel of fixed size.

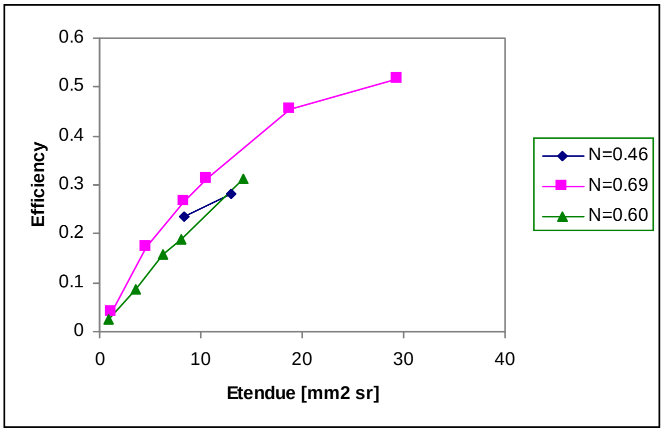

Figure 9. Same data as Figure 8, but expressed as efficiency vs. etendue

6. REFERENCES

1. Matthew S. Brennesholtz, “Light collection efficiency for light valve projection systems”, in *Projection Displays II*, Ming H. Wu, Editor, Proc. SPIE 2650, pp.71-79 (1996)
2. Harald Ries, Narkis Shatz, John Bortz, and Wolfgang Sprikl, “Performance limitations of rotationally symmetric nonimaging devices”, to be published in *J. Opt. Soc. Am. A.*
3. W. T. Welford and R. Winston, *High Collection Nonimaging Optics*. Academic Press, San Diego, California, 1989
4. Douglas M. Rutan, Charles N. Stewart, and Daniel J. Savage, Poster, Eurodisplay ’96, Birmingham.

[1] Brennesholtz paper
[2] Harald's JOSA article.
[3] Roland's book
[4] Rutan, Stewart, Savage - Eurodisplay'96 Poster, Birmingham