

# Automated optimization of non-imaging optics for luminaires

Sergey Kudaev<sup>\*</sup>, Peter Schreiber

Fraunhofer Institute for Applied Optics and Precision Mechanics  
Albert-Einstein-Str. 7, D-07745 Jena, Germany

## ABSTRACT

Specifics of non-imaging optical systems require special algorithms for automated optimization. We have implemented two methods into commercially available optical design software, which are robust and numerically effective. The first one is a modification of the edge-ray principle. In this case the optimization criterion should be expressed in geometrical terms (like, for example, collimation of an extended light source). This gives us the possibility to design not only CPC-like collimators, but also rather complex refractive-reflective (RXI-like) devices. For the second (more general) case the optimization criterion is expressed in energetic terms. In this case stochastic behavior of the merit function due to Monte-Carlo ray-tracing procedure limits the applicability of standard optimization routines available in optical design software. We have realized a direct optimization algorithm, which does not calculate the derivatives of the merit function leading to reduced sensitivity with respect to local statistical deviations. The proposed algorithm is deterministic and does not suffer from redundant trials of random search. As a parametric description for the objects to be optimized we propose the use of piecewise Bezier splines. This allows relative strong shape bending but requires control for intersections. A “red-blue intersection reporting” algorithm is realized as a constraint for optimization.

Keywords: LED light source, optimization algorithms, non-imaging optics, edge-ray principle, non-sequential ray tracing

## 1. INTRODUCTION

The use of LED as a light source in optical devices and general illumination offers significant advantages concerning power consumption, lifetime and color management. However, LEDs are still relatively expensive. Therefore, the first task of the designer is to deliver all the light into the system. Analysis of available solutions<sup>1</sup> shows, that non-imaging optics outperforms its imaging counterpart in collimation of extended sources. However, in contrast to well-known design approaches of imaging optics, design algorithms of non-imaging optics are often very sophisticated and some of them are even patented<sup>2</sup>.

The key technology in imaging optical design is automated optimization. The designer has to provide a reasonable initial system prescription and a merit (or error) function for estimation of system performance. Different nonlinear optimization algorithms (damped least square (DLS) for local search; simulated annealing or genetic algorithms for global exploring) are employed to find the - hopefully - best possible solution<sup>3</sup>.

Throughout the long history of design of image-forming optics correspondence between image quality/energetic performances and simple geometrical parameters like spot radius as a deviation of the ray incidence coordinates from the chief ray coordinates or wavefront deformations as an optical path difference was used. It's sufficient to trace a small number of rays to determine the system performances. Moreover, such merit functions are continuous with respect to almost all the system parameters.

The edge ray principle used to be the workhorse in design of non-imaging optics. Most of the developed algorithms use it for sequential tailoring of concentrator profiles<sup>4</sup>, but direct implementation into commercially available optical design software is comparatively uncommon.

We have developed design tools and algorithms for optimization of non-imaging concentrators within the optical design program ZEMAX®. These include optimization of concentrators according to the edge-ray principle as well as direct optimization of concentrator shape to obtain prescribed light distributions. Developed examples are specific for OSRAM OSTAR® modules with following parameters<sup>5</sup>:

- 4 LEDs (1×1 mm each) in square arrangement with pitch 1.1 mm (RGGB or monochrome configuration);
- Typical luminous flux per LED: R – 30 lm; G – 40 lm; B – 9 lm;
- Lambertian radiation pattern;
- Without lens.

---

<sup>\*</sup> sergey.kudaev@iof.fraunhofer.de; phone +49 3641 807-422; www.microoptics.org

## 2. EDGE RAY PRINCIPLE IMPLEMENTATION

The edge-ray design principle applied to light sources can be formulated as following: rays, originated at the edge point of the extended light source should propagate under the maximum target angle (aimed remaining divergence of the output beam) after leaving the collimator. For ray-tracing software this means in particular:

- Launch one-dimensional ray fans from 2 opposed edge points of the source surface;
  - Optimize the system in such a way, that rays in one fan propagate collinear and under the maximum target angle.
- Choice of correspondence between edge point and target angle (positive or negative) should be made by the designer in advance (this depends on the type of the system). Such optimization is similar to the one of imaging optics and can be performed by standard DLS algorithms taking into account following features:
- Merit function can be expressed by RMS deviation of ray angle within each fan;
  - Ray-tracing is non-sequential. Even for CPC-like solid devices ray path is not predefined: rays either refract through the exit aperture directly (so-called direct rays), or undergo reflections from the side surfaces before refraction at the exit aperture;
  - Concentrator shapes should allow strong shape bending during optimization and easy parametrical transformation between straight, concave and convex shapes;
  - On the other hand, the shape should allow for “fine tuning” to approximate conics or Cartesian oval curves;
  - 2D modeling and optimization is advantageous because of tracing only one-dimensional ray fans, which saves computation time.

In terms of conventional optics edge-ray principle means perfect projection of near-field of an extended source into the far-field of the system. Therefore, such implementation requires etendue conserving behavior of the concentrator. If this is not possible, we propose to use a special weighting of merit function operands: for each ray the deviation from the target angle is weighted by cosine of this ray’s angle in the initial fan. Thus, direct rays have priority for optimization algorithm (this corresponds to Lambertian source distribution).

### 2.1. RXI-like concentrator design

Analysis of different collimator designs<sup>1</sup> shows, that etendue conserving behavior is fulfilled by concentrators with folded multiple reflections (RXI-like). Moreover, compared with the CPC this type of devices has attractive length/diameter ratio (approximately 1/3). In contrast to the classic RXI scheme we used a concentrator with central lens, because of simplified manufacturing without reflective coating in the center.

The model is implemented as a set of user-defined objects into ZEMAX® and represented by piecewise 2<sup>nd</sup> order Bezier curves<sup>6</sup> (Fig. 1) with overall 9 segments, which are connected by pick-ups:

- Entrance surface (curve 1 on Fig. 1 a) – 4 Bezier splines;
- Reflective surface (curve 2 on Fig. 1 a) – 2 Bezier splines;
- Exit surface (curve 3 on Fig. 1 a) – 2 Bezier splines;
- Central lens (curve 4 on Fig. 1 a) – 1 Bezier spline;
- Areas A and B on Fig. 1, (a) are optically inactive and used for mounting the concentrator.

Because ZEMAX® in non-sequential mode cannot operate with 2D objects, we have developed a 2.5D representation (see Fig. 1 b). Depending on the specially defined parameter the concentrator profile of Fig. 1 (a) can be extruded to obtain some depth (for optimization, see Fig. 1 b) or revolved about the z-axis (for analysis).

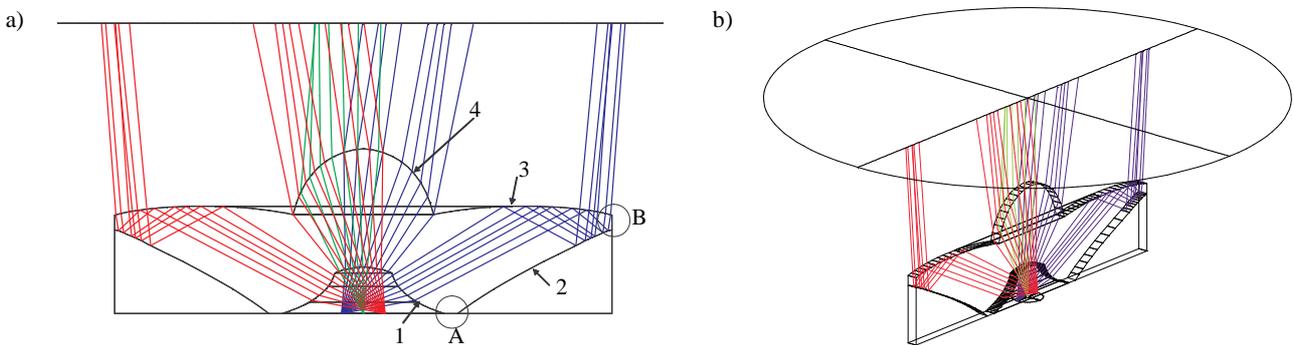


Fig. 1. Bezier-spline based model of combined (RXI-like) concentrator: a) 2D profile and ray fans; b) 2.5D model for optimization, implemented into ZEMAX®.

For optimization we used 3 ray fans: two at  $\pm 1.49$  mm (source diagonal according to OSTAR® specification) and one in the center. This third source point is necessary for central lens optimization (like on-axis field in imaging optics). The starting angle for each fan is selected in such a way, that the nearly perpendicular emanated marginal ray exits from the concentrator approximately at the center (because of axial symmetry of device it is not necessary to trace full  $\pm 90^\circ$  fan). The overall number of degrees of freedom for optimization in this system is 34. Diameter of the concentrator was fixed according to etendue conservation with residual divergence  $\pm 4^\circ$ :

$$etendue = n^2 \cdot area \cdot \pi \cdot \sin^2 \Theta_{max}, \Rightarrow d_{conc} = \frac{2 \cdot size_{source}}{\sin 4^\circ \cdot \sqrt{\pi}} \approx 34 \text{ mm}.$$

## 2.2. Constraints for optimization

Pure numerical optimization with piecewise shape description and a limited number of rays requires additional constraints to prevent the shape of the concentrator to be physically unrealizable. There are two possible classes of such violations:

- Intersection of the fragments near the ends (Fig. 2 a);
- Internal intersection of the fragments (Fig. 2 b).

One feature of a Bezier curve is to pass only through its endpoints<sup>6</sup>. Therefore, it is relatively easy to detect and to control the edge intersection (Fig. 2 a). Internal intersection detection of two Bezier segments (Fig. 2 b) is connected with additional calculations. From the point of view of computational geometry this can be formulated as a “red-blue intersection detection” task<sup>7</sup>: given two separate sets of points; determine, whether the lines, connecting each set are intersected or not.

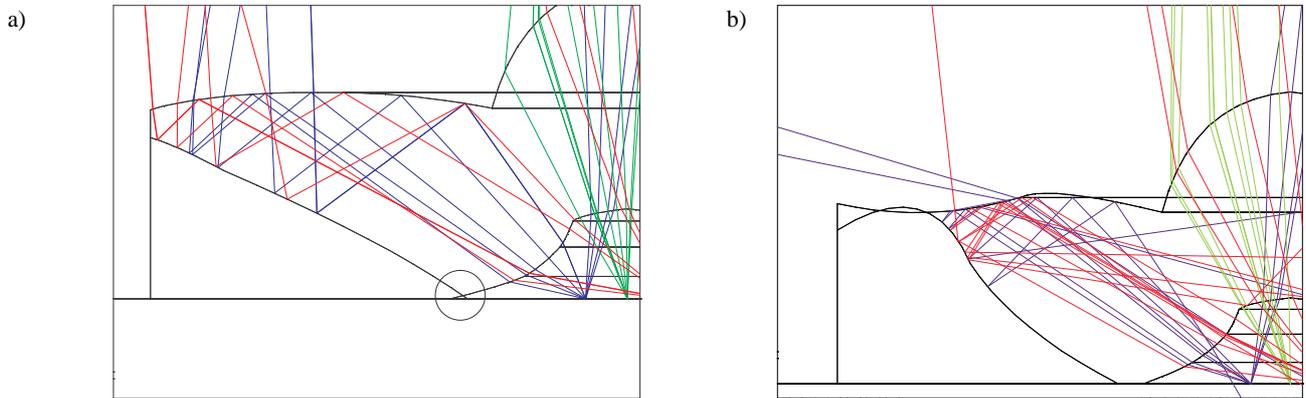


Fig. 2. Artefacts of numerical optimization: physically unrealizable concentrator shapes: a - intersection of the fragments near the ends, b - Internal intersection of the fragments.

Compared to the general problem<sup>7</sup> in our case the computation complexity is reduced because of an ordered sequence of the points in sets. Taking this into account the intersection detection is realised as a binary lookup table:

1. Select one set with smaller number of points as a “base” and the second (with larger number of points) – as a “test” curve;
2. Select first point from the “base” set;
3. Use binary search<sup>8</sup> to find the point from the “test” set, which has nearest coordinate value to this “base” point;
4. Calculate the direction of the vector between these two points;
5. Select next point from the “base” set;
6. Find the point from the “test” set, which has nearest coordinate value to this “base” point;
7. If the direction of the vector between these two points does not correspond to vector, calculated on step 4, intersection is detected; break the algorithm and do not perform ray-tracing for such a system;
8. Repeat steps 5...7 for each point in “base” set.

The predefined geometry of the object also reduces computational efforts by proper selection of the coordinate to test: for intersection detection between the curves 1 and 2 (see Fig. 1 a) we have to operate with the  $y$ -coordinate; for intersection detection between the curves 2 and 3 this will be the  $x$ -coordinate.

This algorithm is fast enough to be implemented into the merit function with one calculation for each system prescription during optimization.

### 2.3. Optimized concentrator shape and performances

The theoretical limit of outcoupling of such a concentrator is approximately 85% because of Fresnel losses (light undergoes refraction at entrance and exit interfaces) and reflection losses (in the model the reflecting surface has ideal aluminum coating with  $R=0.95$ ).

The profile of the concentrator, optimized with the described technique is shown in Fig. 1. The concentrator has length 11.3 mm and diameter 34 mm. The central lens has diameter 9.6 mm and length 4.5 mm. Ray-tracing results (Fig. 3,  $10^6$  rays traced) show, that overall outcoupling from the concentrator is approximately 84% of emitted light and most of the light propagates with residual divergence less than  $\pm 8^\circ$ .

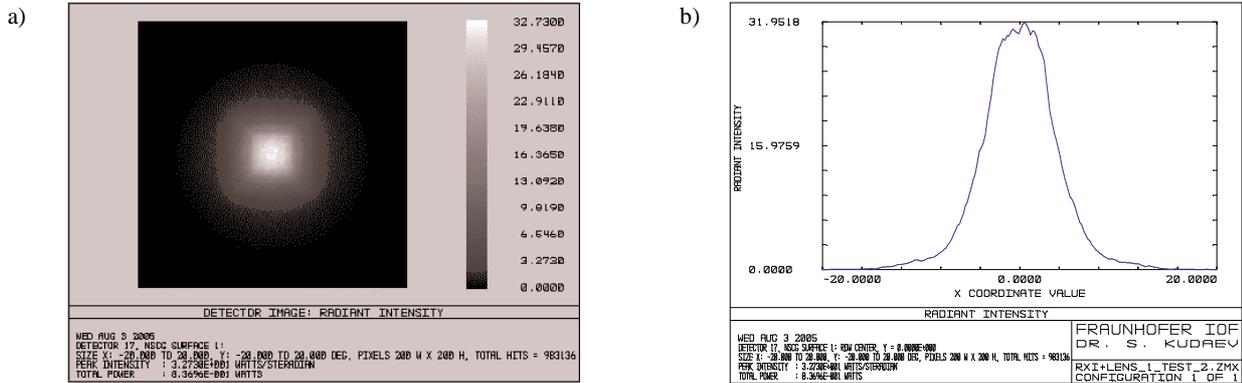


Fig. 3. Ray-tracing results with complete model of concentrator: a) 2D far-field distribution,  $\pm 20^\circ$  area; b) Axial cross-section of distribution.

Analysis of Fig. 3 shows, that the concentrator is not perfectly etendue-conserving with residual divergence  $\pm 4^\circ$  (optimization criterion). We suppose that this difference has following reasons:

- Diameter of the concentrator was selected according to etendue conservation under the assumption that the whole exit area is emitting. From Fig. 1 (a) follows, that some part of exit surface does not emit;
- Second order Bezier curves are not flexible enough to model the required shapes;
- Residual divergence of the light from such large extended source collimated by simple central lens is much larger, than the aimed  $\pm 4^\circ$ .

The overcome these disadvantages requires further improvement of the model and optimization algorithms. However, this concentrator already shows very good performances: for example, a truncated CPC with 80% of energy within  $\pm 8^\circ$  residual divergence and the same source size has a length of at least 100 mm.

It is worth mentioned, that neighbored Bezier splines are almost continuous at the connecting endpoints after optimization (the tangents are almost the same). This means, that the optimization process is steady and convergence rate can be improved by replacing piecewise Bezier spline description by a single higher-order spline curve. But this would additionally require further development of the fast ray-tracing algorithms.

### 2.4. Realization and measurements

For verification of the developed design methods some prototype concentrators were manufactured and characterized (see Fig. 4, Fig. 5). Solid concentrators were produced by direct single-point diamond turning of PMMA blanks. The resulting shape tolerance (deviations less than  $5 \mu\text{m}$ ) and surface quality were sufficient for “proof of concept”.

The light energy, encircled within the residual divergence angle (Fig. 5) is calculated by integration of goniometric measurements of light distribution. Design and measured values are in good agreement, but the overall outcoupling energy is approximately 5...6% smaller than design value. This difference can be explained by diffraction on the turned grooves, scattering on the surface microroughness and reflection losses.

Two-dimensional goniometric scans of the far-field light distribution after the concentrator (Fig. 6) also show good coincidence with the simulated distribution. The measurements were performed with 2 LEDs in diagonal position emitting equal flux (adjustment was done with integrating sphere). One can resolve the positions of LEDs on the graphs with linear scale (Fig. 6 a, b). The bright spot has an angular size of  $\pm 5^\circ$ .

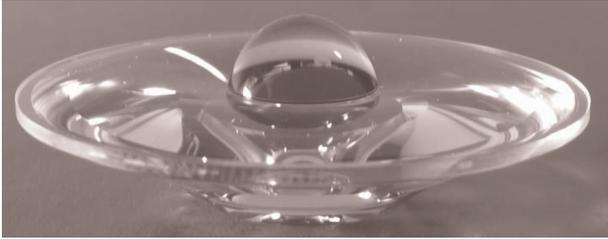


Fig. 4. Diamond-turned concentrator without reflection coating.

Stray light distribution (see the graphs with logarithmic scale, Fig. 6 c, d) also shows good agreement with design prediction.

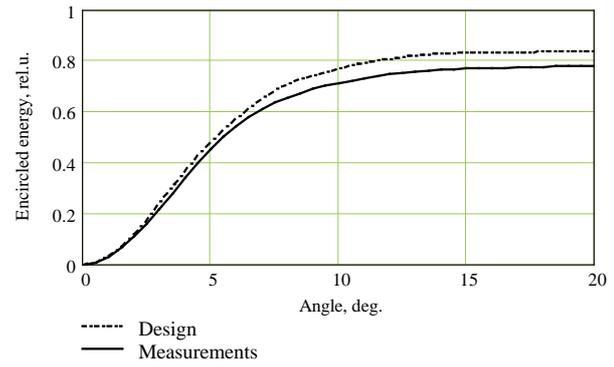


Fig. 5. Encircled energy in the far-field

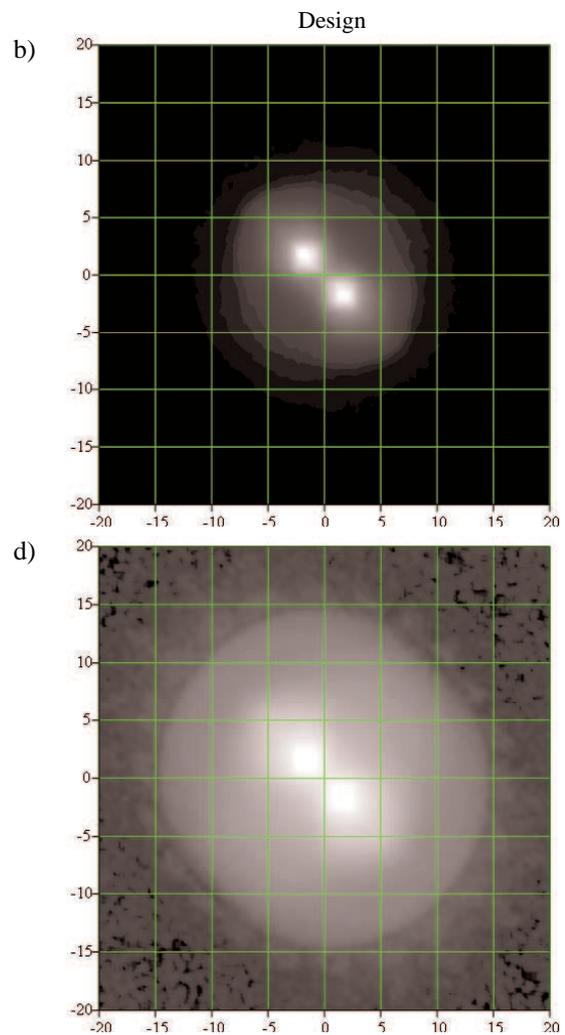
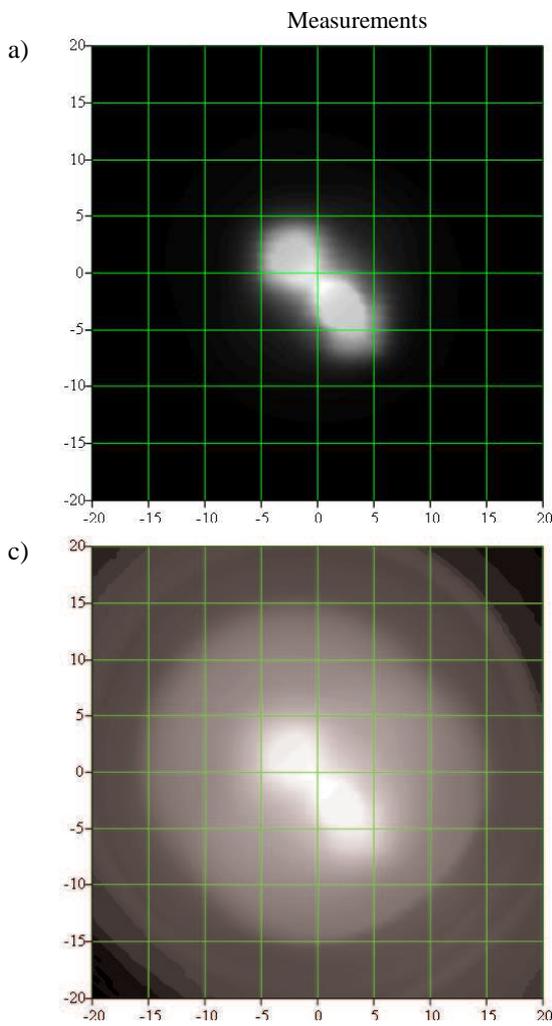


Fig. 6. Far-field distribution for combined concentrator. Design data vs. measurements: a), b) – linear brightness scale; c), d) – logarithmic scale.

### 3. DIRECT OPTIMIZATION IMPLEMENTATION

Despite of good collimation properties, combined multiple-reflective/refractive concentrators are not the universal device for primary optics for LED. A couple of applications for both general illumination and image projection require homogeneous near-field light distribution.

Homogeneous illumination can be achieved only if the concentrator operates with multiple reflections, does not conserve the skewness of the source and has one of the homogenizing cross-sections: triangle, rectangle or hexagon<sup>9</sup>. Thus, etendue conserving behavior cannot be maintained anymore and optimization should be performed by estimating the energetic terms: overall luminous flux and its distribution. The merit function in this case is a weighted sum of deviations of efficiency and estimation of homogeneity of the distribution:

$$MF^2 = \frac{w_1 \cdot \sigma^2 + w_2 (1 - k_{angle})^2}{w_1 + w_2}, \quad (1)$$

where:  $\sigma$  - either RMS or PV deviation of irradiance;  $k_{angle}$  - outcoupling efficiency (flux from concentrator within target angle over the flux from LED);  $w_1, w_2$  - weights of operands.

Such an estimation leads to a general trade-off of optimization with energetic terms: to obtain high signal-to-noise ratio of irradiance distribution one has to trace a lot of rays, but this will slow down the optimization process significantly. As a rule-of-thumb we are using the estimate, that at each detector pixel should be incident at least 100 rays. In this case statistical deviations are smaller than human eye contrast sensitivity (10...20% for bright light source<sup>3</sup>). Nevertheless, standard deterministic algorithms fail to optimize the system:

- The flux is determined by tracing a limited number of rays. It is possible, that a small change in the value of a variable may not change the flux estimate; if no rays are close enough to an aperture to change from being vignetted to unvignetted or vice-a-versa. Damped least square algorithms cannot compute the finite difference derivative in this case;

- The merit function has long valleys, where changes of optimizing parameters do not affect the MF value;

The problem is illustrated by Fig. 7: for square Bezier concentrator with fixed entrance/exit apertures and length there are 2 degrees of freedom:  $x$  and  $y$  coordinates of the middle control point - Fig. 7 (a). Although phase space of the system shows clear tendency Fig. 7 (b), DLS optimizer exhibits problems in finding the minimum, because of the long valley and non-smooth topology.

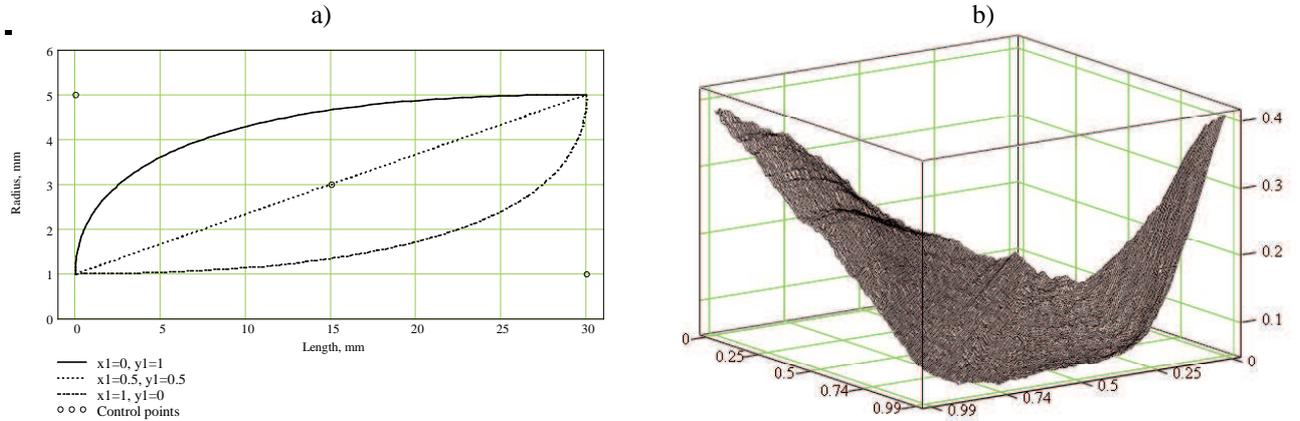


Fig. 7. Allowed range for two-parameter optimization of square Bezier concentrator (a):  $0 \leq x_1 \leq 1$ ,  $0 \leq y_1 \leq 1$ ; optimization phase-space (b): merit function value acc. (1) vs. parameter values.

Among the developed methods of non-linear optimization without derivatives, the Powell method is stable in the most of cases and, especially, in the tasks of statistical estimation of merit function<sup>10</sup>. Detailed description of this method can be found in<sup>11</sup>. We have realized Powell optimization with Brent's one-dimensional search as an add-on for ZEMAX® (see Fig. 8).

- Algorithm performs independent (i.e. for single variable) optimization with either iterative parabolic fit or golden section search: see transfer from points 0→1→2 on Fig. 8.

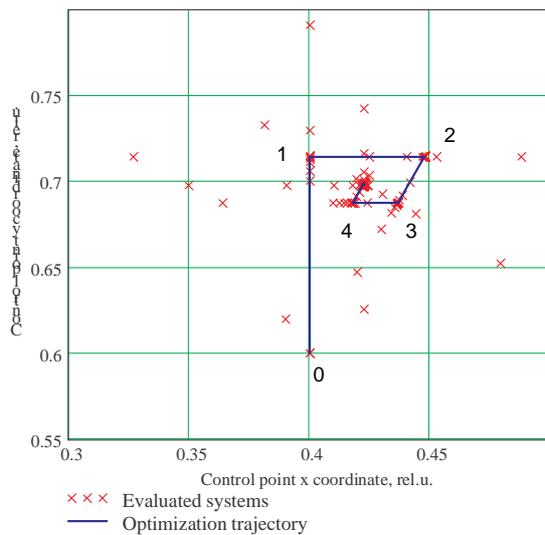


Fig. 8. Trajectory of the optimization with Powell method.

requires 150000 estimations of merit function for transferring the system to approximately the same state. However, implemented algorithm ensures finding only a local optimum. The combination of this algorithm with some global optimization algorithm would be very useful.

Powell algorithm was compared with standard DLS optimizer for some imaging systems. In spite of longer calculation time, the optimized systems in both cases have the same parameter values. Therefore, we assume numerical stability and efficiency of the developed algorithm.

### 3.1. Rectangular concentrator: design, realization and measurements

The developed optimization algorithm was applied to design a square free-shape concentrator with following requirements:

- Source type: OSRAM OSTAR® RGGB module;
- Minimum residual divergence;
- Fixed output aperture size;
- Homogeneous distribution of the near-field (distance from concentrator output 0.5...2 mm). This distribution should be estimated within the residual divergence angle.

Because of the relative large exit aperture the solid concentrator was optimized to operate with immersion coupling to LED (this can increase overall outcoupled flux on 30...100%). The shape of the concentrator consists of 3 Bezier curves (two for side surfaces and one for the lens at the top) and has a square cross-section. The overall number of variables is 12.

Optimization takes 10 min on dual processor computer (ZEMAX® is very effectively multithreaded for ray-tracing operations) and requires approximately 800 evaluations of the merit function. The optimized concentrator has length 39 mm, transmission of 72% of LED light and irradiance inhomogeneity less than 15%.

A prototype was manufactured by diamond milling of PMMA and characterized by a telecentric 4F setup with 1:1 imaging and obscuration of the light outside the target  $\pm 10^\circ$  residual divergence.

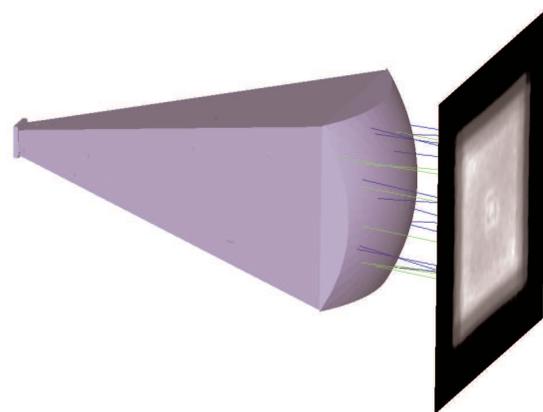


Fig. 9. Square concentrator with homogeneous near-field distribution.

- After all the independent optimization of one iteration, the algorithm performs additional optimization in resulting direction: transfer 2→3 along imaginary line between points 0 and 2.
- This new direction is included into the variable matrix instead of the direction of maximum changes of merit function (there is no transfer in vertical direction after the point 3 on Fig. 8, but there is the transfer in direction, parallel to line 2-3).
- After  $n$  successful iterations, where  $n$  is the number of independent variables, the algorithm performs resetting the variable matrix to its initial state (this is necessary for conservation of quadratic convergence<sup>11</sup>);
- Termination criteria are either negligible relative changes of merit function or user-defined number of iterations.

This algorithm does not suffer from redundant trials of random search. For this simple example only 4 iterations were necessary with less than 100 evaluation of merit function (ray-tracing). For comparison, standard random search, available in ZEMAX® (Hammer optimization),

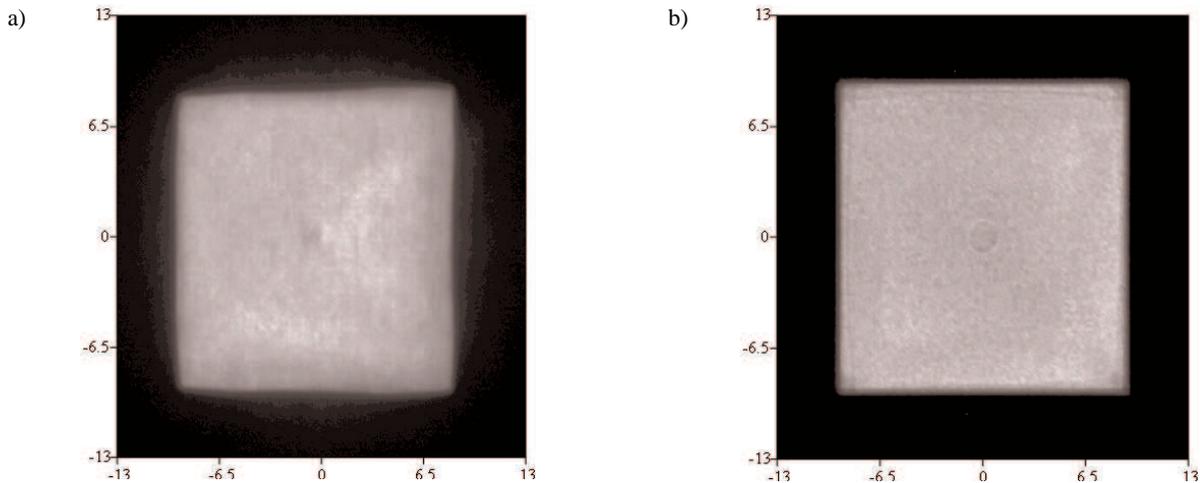


Fig. 10. Near-field distribution for rectangular concentrator with lens: a) – measurement, b) – design. Axes extent on both graphs is  $\pm 13$  mm.

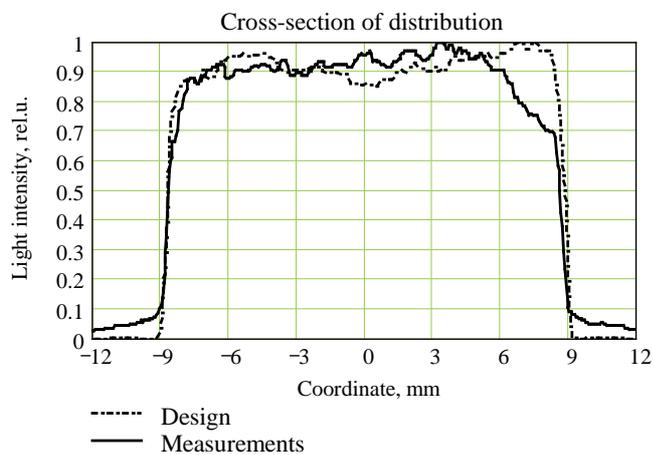


Fig. 11. Cross-section of the distribution.

The measurements were performed with 1 LED only switched on. The two-dimensional irradiance distribution (Fig. 10) shows good agreement with design predictions.

There is a dark spot in the center of the aperture, which is visible in design as well as in the measurements. The reason for it is the shape of the output lens: it was realized as rotation symmetric free-shape Bezier profile. The sag has a discontinuous derivative in the center of rotation, i.e. a small cusp. However, if this free-shape lens is replaced by a best-fit conic asphere, general transmission of the system drops down on 5% (simulation results).

The cross-section of the distribution (Fig. 11) shows, that the distribution is not perfectly homogeneous, but irradiance drop at the edges is about 20% of maximum value (this value is near the limit of

contrast sensitivity of human eye). The graph also shows that the distribution is slightly non-symmetric: right side has approximately 5...10% higher irradiance (this is valid for both the design and the measurements).

Nevertheless, visual perception of the module is that light distributed almost homogeneously over the emitting area.

#### 4. CONCLUSION

Main result of this work is that design algorithms for non-imaging optics can be implemented into universal optical design software. The combination of universal features and such add-ons provides great flexibility to cover quite different applications and design goals:

- DLS optimization can be used for almost etendue-limited systems by means of edge-ray principle implementation;
- Combined refractive/folded-multiple-reflective concentrators, optimized according to edge-ray principle are very effective (transmission more, than 80%) and size-saving solutions for collimating LED light with small residual divergence (less, than  $\pm 10^\circ$ );
- Piecewise Bezier description of concentrator profile is well suited for direct optimization according to different criteria, but should be constrained against physically unrealizable configurations;

- Design of concentrators for prescribed (in particular – homogeneous) irradiance distributions requires deterministic optimization algorithms without derivatives;
- Faceted concentrators with homogeneous irradiance distribution are not perfectly étendue conserving, but can collimate more than 70% of LED light within moderate (approximately  $\pm 10^\circ$  or more) target residual divergence angle;
- Index matching ensures higher outcoupling of light from LED, but has to be considered during the optimization process;
- Concentrator prototypes, optimized according to the developed algorithms, show good agreement with design goals and simulation predictions.

### ACKNOWLEDGEMENTS

The authors would like to thank Stefan Groetsch and Dr. Georg Bogner from OSRAM OS for supplying the LED sources. This work was funded by German Federal Ministry of Education and Research (BMBF) through contract 13N8271.

### REFERENCES

1. R. Winston, J.C. Minano, P.G. Benitez, *Nonimaging Optics*, Academic Press, San Diego, 2004
2. J. C. Minano, P. Benitez, J. C. Gonzalez, W. Falicoff, and H. J. Caufield, *High efficiency non-imaging optics*, U.S. Patent 6,639,733, Oct. 28, 2003
3. W.J. Smith, *Modern optical engineering*, 3rd ed., McGraw-Hill, New York, 2000
4. H. Ries and R. Winston, *Tailored edge-ray reflectors for illumination*, J. Opt. Soc. Am. A, **11**, 1260 (1994).
5. <http://www.osram-os.com/ostar/index.php?init=1>
6. S. Kudaev, P. Schreiber, Optimization of symmetrical free-shape non-imaging concentrators for LED light source applications, Proceedings of SPIE, Vol. 5942, to be published
7. P.J. Schneider, D.H. Eberly, *Geometric Tools for Computer Graphics*, Morgan Kaufmann Publishers, San Francisco, 2003
8. D. Knuth, *The Art of Computer Programming, Volume 3*, 1st edition, Addison-Wesley Professional, Stanford, 1998
9. W. Cassarly, *Nonimaging optics: Concentration and Illumination*, in Handbook of Optics, Vol. 3, McGraw-Hill, New-York, 2001
10. D.M. Himmelblau, *Applied nonlinear programming*, McGraw-Hill, New York, 1972
11. R.P Brent, *Algorithms for minimization without derivatives*, Prentice-Hall, New Jersey, 1973