What's the depth of field (DOF) for a diffraction-limited lens?

Rik Littlefield --- rj.littlefield@computer.org

Working draft 5/27/2014
typo corrected 6/22/2014

Introduction

This question is usually posed in one of these other forms:

1. What's the depth of field of a microscope objective?
2. How does diffraction affect macro depth of field?
3. When focus stacking, how does image quality depend on step size?

Unfortunately none of these questions is simple to answer. A lens that is operating in its diffraction-limited regime behaves in ways that are counterintuitive to most people (myself included). The intent of this long posting is to look at how image quality varies as the lens is operated at different distances from perfect focus.

Summary

Depth of field for a diffraction-limited lens depends on the magnification, the effective aperture, how much contrast loss you're willing to accept, and what wavelengths you're working with. In a macro/micro setup with a lens operating in air, depth of field can be calculated using either of the following two formulas, which have their roots in wave optics:

1. TDOF_qlwe = lambda / (NA*NA)
2. TDOF_qlwe = lambda * 4 * (f_eff * f_eff) / (m*m)

where

- TDOF_qlwe is the total DOF, front to back, measured at the subject, for "quarter lambda wavefront error". (We'll study later what that phrase means. For now, just take "lambda wavefront error" as a distance unit that describes how badly defocused the lens is. The defocus distance for 1/4 lambda wavefront error is twice that for 1/8 lambda wavefront error, and so on.)
- lambda is the wavelength, typically chosen as 0.00055 mm corresponding to green light.
- NA is the subject-side numerical aperture (such as stamped on a microscope objective).
- f_eff is the camera-side effective aperture, often estimated as f_eff = f_lens * (m+1) .
- m is the magnification, from subject to camera sensor.

Images shot at maximum defocus (1/2 of TDOF_qlwe away from perfect focus) will capture detail at the same size limit as for perfect focus, but the images will have visibly reduced sharpness and contrast for fine detail. You may prefer to use a smaller DOF, which can be very effective because the loss of contrast varies as the square of the defocus distance.

Examples

1. NA 0.25 microscope objective,
   TDOF_qlwe = 0.00055 mm / (0.25*0.25) = 0.0088 mm.

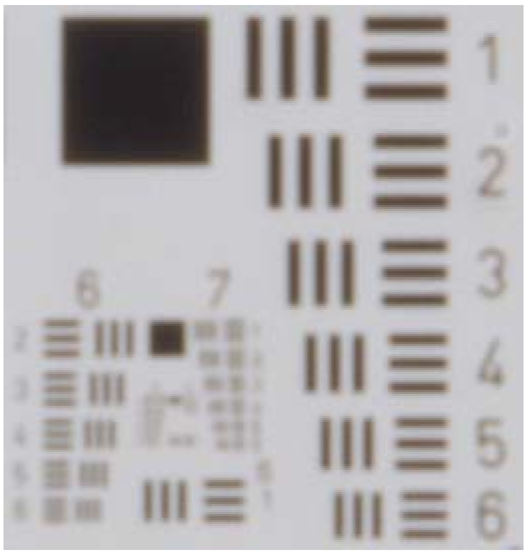
2. Lens set at f/8 and extended to give magnification 0.5 (half life size),
   f_eff = 8 * (0.5+1) = 12
   TDOF_qlwe = 0.00055 mm * 4 * (12*12)/(0.5*0.5) = 1.26 mm.

Tables based on these formulas can be found at
http://zerenesystems.com/cms/stacker/docs/tables/macromicrodof and
http://zerenesystems.com/cms/stacker/docs/tables/landscapes.

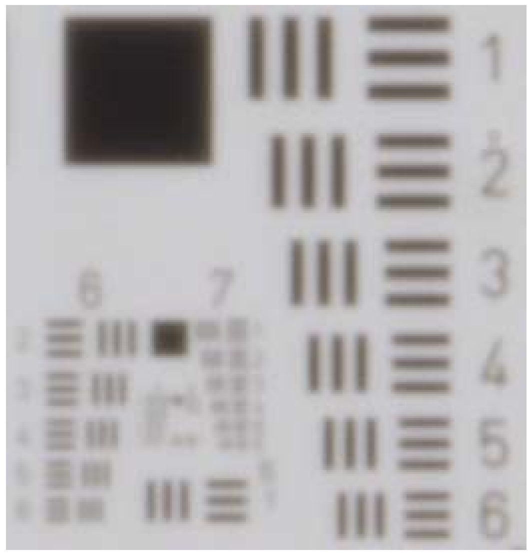
Image Comparisons

For visual comparison, here are images shot at perfect focus and defocused by a specific amount. These are actual captures, shot using a method that removes sensor resolution as a significant factor.

Perfect focus



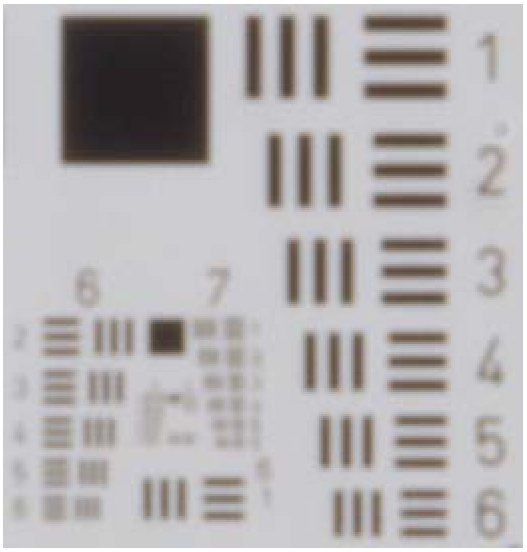
Defocus to 1/4 lambda wavefront error



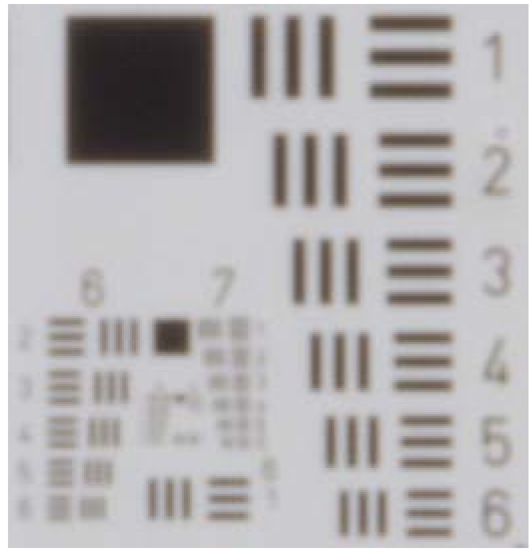
Quantitatively, 1/4 lambda wavefront error corresponds to about 26% loss of contrast at the worst spatial frequency. There is no change in the nominal cutoff frequency, but due to contrast loss some detail that is visible at perfect focus can be effectively lost at maximum defocus.

For more critical work, you may prefer to use smaller DOF. The loss of contrast varies roughly as the square of the defocus distance, so dropping to 1/2 the distance cuts the contrast loss to only 1/4 as much, about 7% at the worst spatial frequency. Here is what those images look like:

Perfect focus

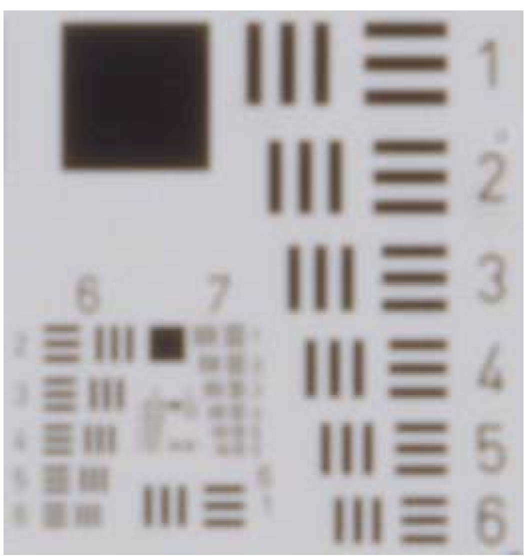


Defocus to 1/8 lambda wavefront error

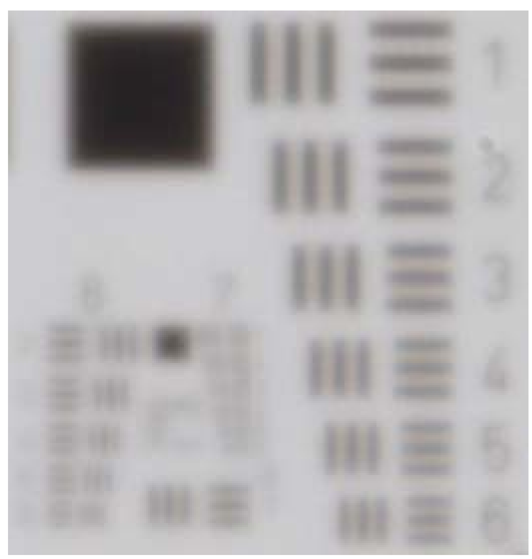


Going in the opposite direction, doubling the defocus distance increases the contrast loss by roughly a factor of 4X. At the resulting 1/2 lambda wavefront error, there is still no change in the nominal cutoff frequency, but the overall loss of contrast and sharpness is so severe that the result is not a pretty picture. Here is the comparison:

Perfect focus

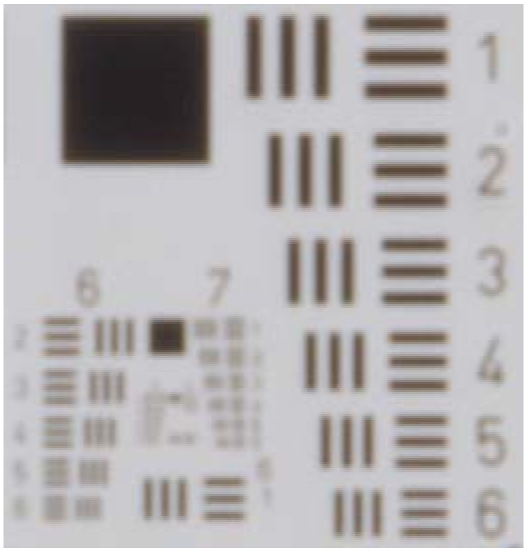


Defocus to 1/2 lambda wavefront error



Going just a little farther than that, the image goes to complete junk. Here's the picture at 3/4 lambda wavefront error, just 50% more defocus from the previous image:

Perfect focus



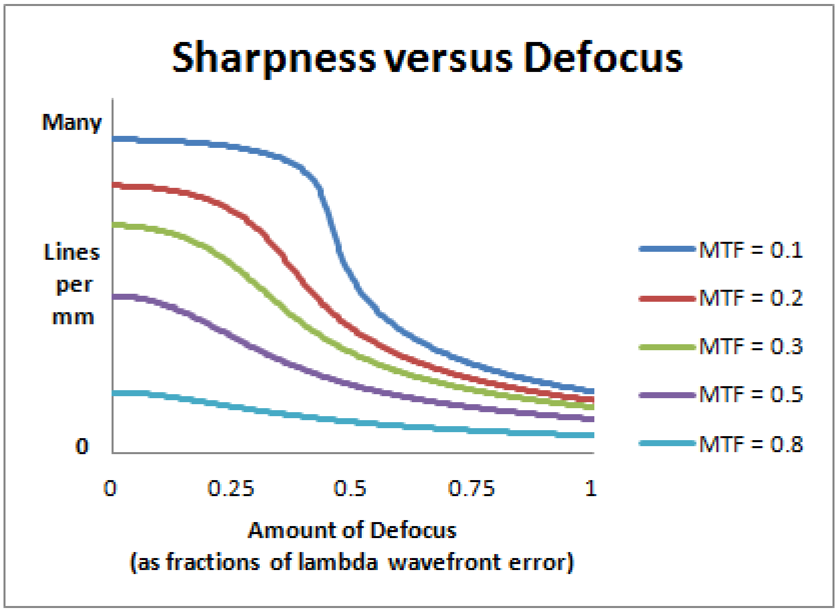
Defocus to 3/4 lambda wavefront error

By this point (3/4 lambda wavefront error), the cutoff frequency of the lens has been reduced by a factor of over 3X. None of the bars in this illustration are correctly resolved, and several of the larger elements along the right side show contrast inversion ("spurious resolution"), with two dark bars between light ones in positions where there should be two light bars between dark ones.

Other representations

The progressive loss of contrast with increasing defocus can be shown in several other representations, each of which may be more useful than the others in some circumstances.

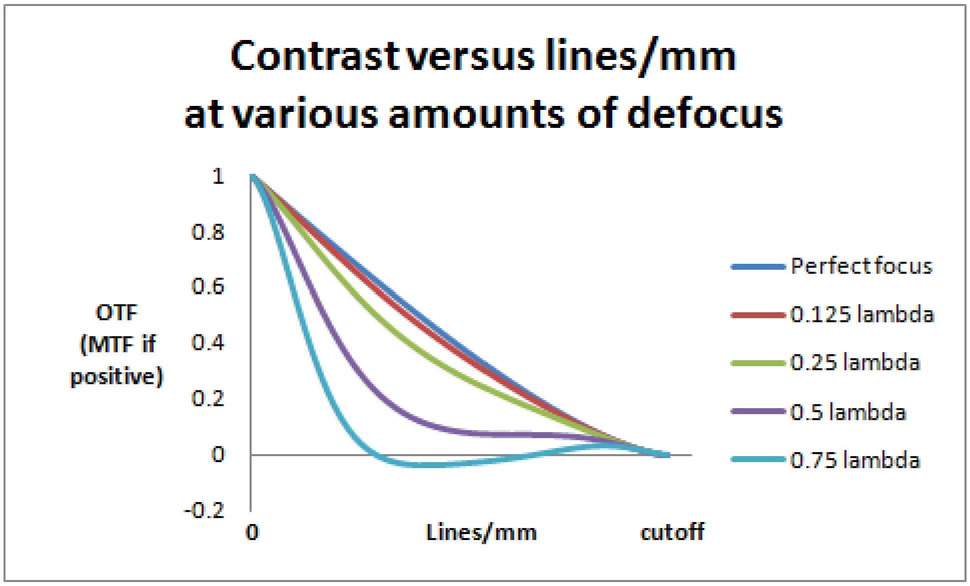
Here is a summary graph that may be helpful:



In reading this graph, the sense is "higher is better". Finer detail is always resolved with lower MTF, and if you can live with lower MTF then you can resolve more lines per mm. That's why smaller values of MTF are on top. As you move away from perfect focus, the lines go downward, meaning that the resolution at which MTF reaches some particular value gets lower and lower. That's bad because it means you're losing sharpness.

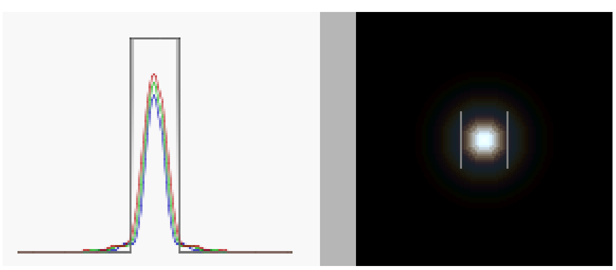
The key point about this graph is that for small amounts of defocus, there is relatively little effect on image sharpness -- the lines are pretty flat near the left side. As you defocus farther, sharpness is lost increasingly quickly -- the curves not only <u>go</u> downward, they <u>bend</u> downward. At some point around 0.4 lambda wavefront error, there is a really painful loss of contrast, especially for fine detail. It's best not to operate there or any farther out.

Another common and useful representation is to look at MTF (Modulation Transfer Function) as it varies depending on spatial frequency (lines per mm). Here is a graph that plots several different curves, one curve for each specific amount of defocus ("lambda wavefront error") as shown in the earlier image comparisons.

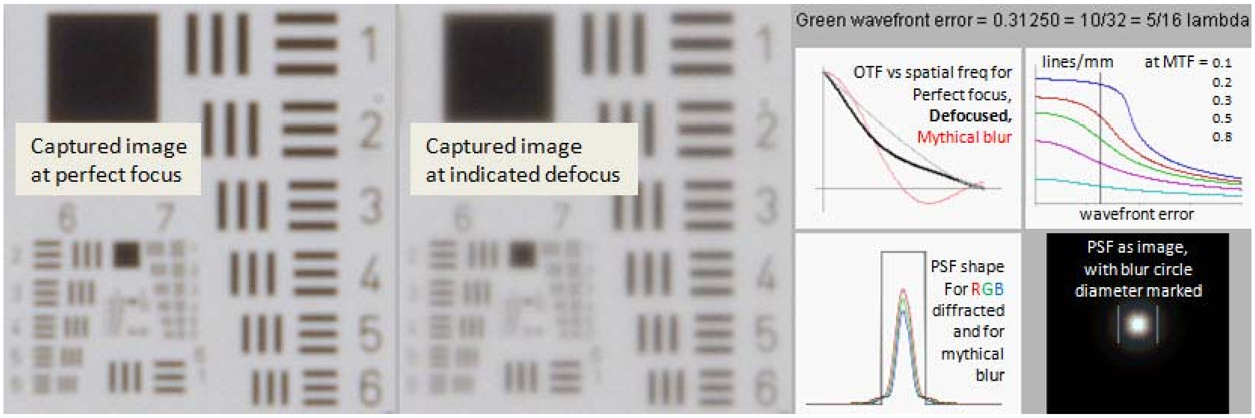


In this graph, the highest of the curves is the one at perfect focus, a lens that is losing sharpness from diffraction but nothing else. Very slightly below that is the curve for 1/8 lambda wavefront error, then 1/4 lambda, then the "not a pretty picture" at 0.5 lambda, and finally the "complete junk" at 3/4 lambda. You can see that the line for 0.75 lambda actually drops below zero on the vertical axis, reflecting that area of contrast inversion and spurious resolution that is shown in the comparison images.

Finally, we might be interested in the profile of the point spread function (PSF), either alone or as compared to the diameter of the blur circle that is predicted by simple ray optics. Here's the situation at 1/4 lambda wavefront error. The straight lines represent the diameter of the ray optics blur circle; the colored curves are the diffracted PSF for red, green, and blue.



For purposes of presentation, I have bundled all this information together into a collection of image panels that look like this:



The collection can be downloaded as a single .zip file containing 33 PNG files, one each for wavefront error from 0 to 1 by 1/32. A simple way of viewing these images is to load the whole .zip into Zerene Stacker (no need to unpack it first), then use press-and-drag to quickly scroll through the list as if you were running a focus slider. (It helps to maximize the source window size so that ZS can show the images at 100%.)

Alternatively, you can download a .psd file containing 33 layers, one for each of the above images. This file can be loaded into Photoshop or GIMP, and then the little "eye" icons that control layer visibility can be toggled on and off to flash between layers.

Or, you can download the images as a .mov file, and use the timeline pointer in QuickTime player to

What is "wavefront error" and why so much emphasis on it?

The term "wavefront error" comes from the wave optics model of imaging. To understand the term, we first need to understand a little about wave optics.

Most photographers are familiar with the ray optics model of light. This model assumes that (1) light propagates in straight lines called "rays" and (2) you can just add up the intensities for all arriving rays to figure out what the image looks like.

Wave optics retains the first assumption that light travels in rays, but it replaces the second assumption by recognizing that no, you can't just add up intensities, you also have to consider phase. So in wave optics, we calculate the optical path length along each ray, then use the differences in optical path length to compute differences in phase, and finally add up the contributions along various paths as sines and cosines depending on the phase differences..

With a lens that has no aberrations and is perfectly focused, all of the ray paths have exactly the same optical path length and thus all of the contributions have the same phase at the focus point. As the lens move away from perfect focus, paths through various parts of the lens change in length by slightly different amounts. These differences in optical path length mean that various contributions no longer arrive exactly in phase. It is these phase errors that cause contrast to drop.

In this model, the optical path lengths do not depend on wavelength, but the phase error does. As a result, it is traditional to express the phase error in terms of the wavelength lambda. When two waves arrive with phases that are 90 degrees apart, this is because the corresponding path lengths differ by 1/4 of the wavelength, so that particular error is called "1/4 lambda".

For overall image quality, it turns out that what matters is the maximum phase error between any two paths through the lens to the image point. This is called the "wavefront error".

At this point, we have what we need to define the term: "1/4 lambda wavefront error" means that there's a maximum phase shift of 90 degrees between contributions along any two ray paths through the lens to the image point.

The reason we emphasize wavefront error, as opposed to other measures of defocus, is that wavefront error bundles together most of the interactions between lens focal length, aperture size, focus distance, and so on. Assuming that all the wavefront distortions are due to focus errors, **any two optical setups that have the same wavefront error will have MTF curves with exactly the same shape**. The MTF curves will vary only in scale to match the cutoff frequency, which is determined by aperture size as follows:

subject side cutoff frequency = (2*NA)/lambda
camera side cutoff frequency = 1/(lambda*f_eff)

(As above, NA and f_eff refer to subject-side numerical aperture and camera-side effective aperture.)

At the very beginning of this document, we gave one formula for a defocus distance that is tied directly to 1/4 lambda wavefront error:

- TDOF_qlwe = lambda / (NA*NA)

This formula gives the total depth of field, front limit to back limit, within which the wavefront error is no greater than 1/4 lambda. In the absence of other aberrations the wavefront error is symmetric around focus, so we also have that the one-sided DZ corresponding to 1/4 lambda wavefront error is just half as large as the total depth of field:

- DZ_qlwe = lambda / (2*NA*NA)

It is important to know that different manufacturers use different conventions when quoting depth of field. The formulas given by Nikon refer to total depth of field, front limit to back limit, and our TDOF_qlwe can be recognized as the first term in Nikon's formula at http://www.microscopyu.com/articles/formulas/formulasfielddepth.html. In contrast, Mitutoyo quotes the depth of field for their microscope objectives as one-sided, from perfect focus to either the front or the back limit, so their formula agrees with our DZ_qlwe.

When working with wavefront error, there are two useful relationships to keep in mind:

- If aperture is fixed while the mm of defocus distance is changed, then wavefront error scales linearly in direct proportion to the mm of defocus. This is why I said earlier that wavefront error could be treated as a distance unit indicating how badly the lens was defocused. However many mm of defocus gives 1/4 lambda wavefront error, half that much gives 1/8 lambda and twice that much gives 1/2 lambda.
- If the mm of defocus distance is fixed while the aperture is changed, then wavefront error scales as the square of the aperture width. This is why the TDOF and DZ formulas include NA squared in order to produce a constant amount of wavefront error.

Comparison to the classic "Circle Of Confusion" model

Now, let's get back to talking about depth of field.

Most photographers are familiar with the classic method for calculating depth of field based on some acceptable COC (Circle Of Confusion). A typical formula for macro applications is:

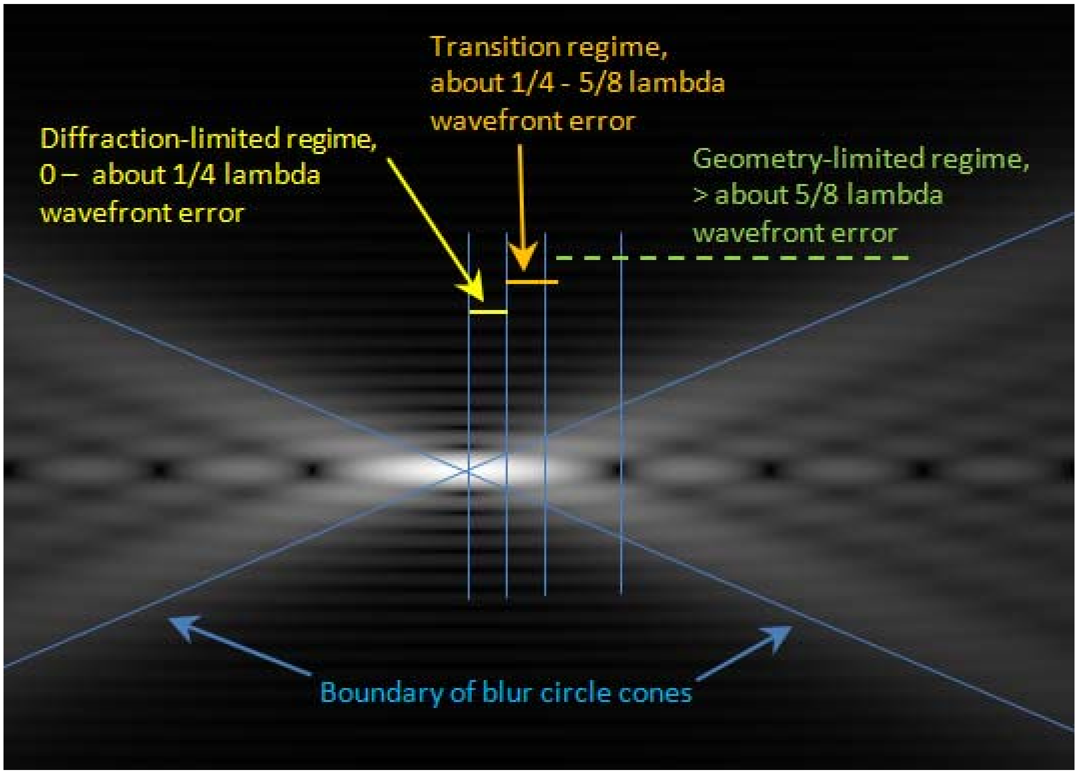
TDOF_coc = 2 * C * f_lens * (m+1) / (m*m)

where

- TDOF_coc is the total DOF (front to back) such that the lens "blur circle" remains smaller than C.
- C is the diameter of the COC, typically taken as 1/1000 to 1/1500 of the sensor width.
- f_lens * (m+1) is the effective f-number of the lens, estimated from the lens's own f-number and the magnification.
- m is the magnification

Formulas like this always ignore diffraction and instead are based on the simplifying assumptions of ray optics (also called geometrical optics). As mentioned above, ray optics assumes that 1) light propagates in straight lines called "rays" and 2) you can just add up the intensities for all arriving rays to figure out what the image looks like. Under these assumptions, a lens with a circular aperture creates uniformly illuminated "blur circles" whose diameter is just proportional to the distance from perfect focus. Given a point source, light coming from the aperture forms a cone of light that converges to a single point at perfect focus, then expands again on the far side.

However, what the light actually does is quite a bit more complicated than that. The diagram that appears next is a (computed) longitudinal section of the 3-D point spread function of a perfect lens imaging a point source. In other words, this is a map of the light intensity, near focus, if you point your camera at a bright pinhole that is emitting light.



(Base image extracted from http://upload.wikimedia.org/wikipedia/commons/1/10/Spherical-aberration-slice.jpg)

I have marked on this diagram the boundary of the cones predicted by ray optics, plus the positions corresponding to perfect focus and to 1/4 lambda, 1/2 lambda , and 1 lambda wavefront error. I have also identified three regions that I find are useful for thinking about the relationship between optics that are limited by diffraction, and optics that are limited by geometry.

In brief, these regimes are:

1. Geometry limited: when the Airy disk is much smaller than the COC, the classic blur circle model is accurate because there's a good match between the envelope of the diffraction pattern and the diameter of the ray optics blur circles.

2. Transition: when the Airy disk is only slightly smaller than the COC, the blur circle model no longer accurately predicts the shape of the MTF curve, but it still does a pretty good job of predicting overall sharpness and thus overall DOF. In this regime, the actual diffracting lens will not resolve coarse detail at the same high contrast predicted by ray

optics, but in exchange it resolves substantially finer detail at low contrast.

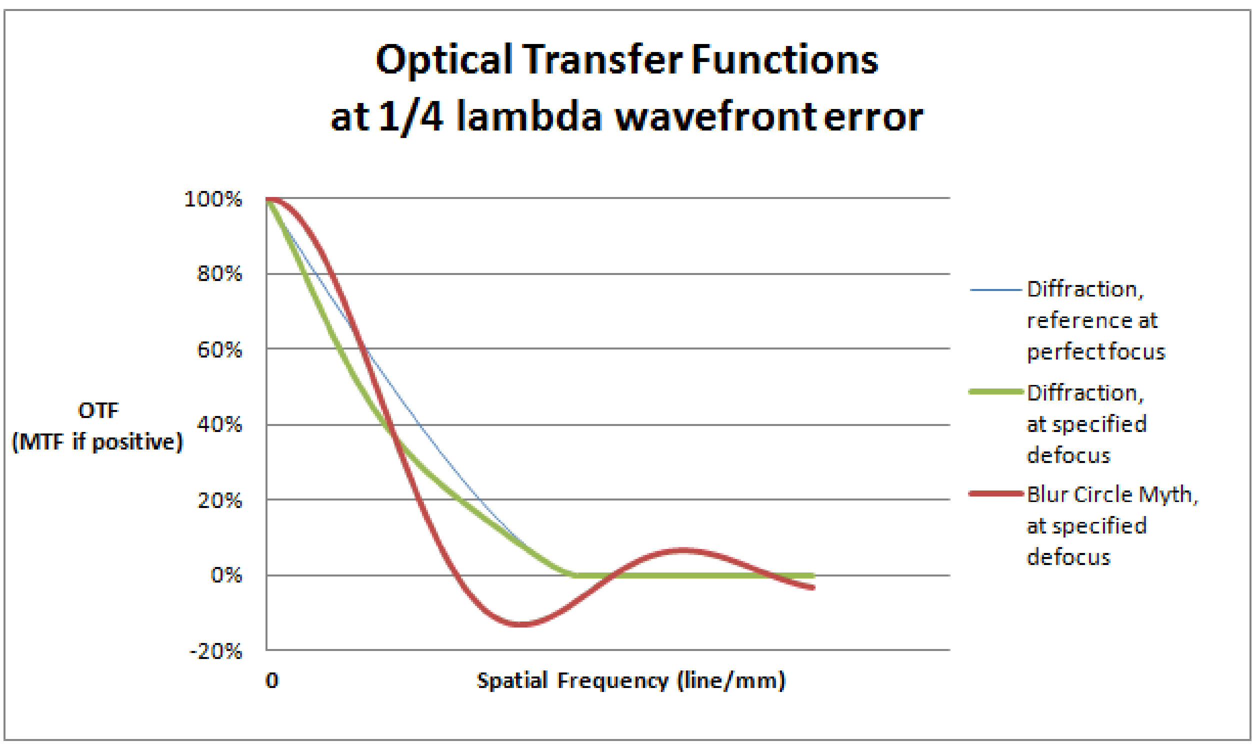
3. Diffraction limited: when the Airy disk is larger than the COC, the blur circle model becomes unusably inaccurate. In this regime the resulting images will not meet the sharpness criterion implied by the selected COC, but in exchange the DOF will be significantly greater if you're willing to accept what sharpness there is. Another reasonable way of thinking about this regime is that if you're working here, then you've just implicitly agreed to accept a larger COC, at which point you could repeat the classic calculation with the larger COC and get a decent prediction that way.

Looking in more detail at both the math and the MTF curves, we can get the following.

Slogging through the algebra, it's simple enough to figure out that both models give the same value for TDOF when C = 2 * lambda * f_eff.

This is significant because that value (2 * lambda * f_eff) also happens to be a fixed fraction of the wave optics Airy disk diameter, which is (2.44 * lambda * f_eff) to the first black ring. That fraction -- the area covered by the COC -- contains over 98% of the total power in the Airy disk's central peak. Further, because the Airy disk is strongly center-weighted, it is able to resolve significantly finer detail than a uniformly illuminated blur circle could.

Here's the picture for that situation, expressed as two curves of transfer function versus spatial frequency:



There are three things I want you to notice about this graph.

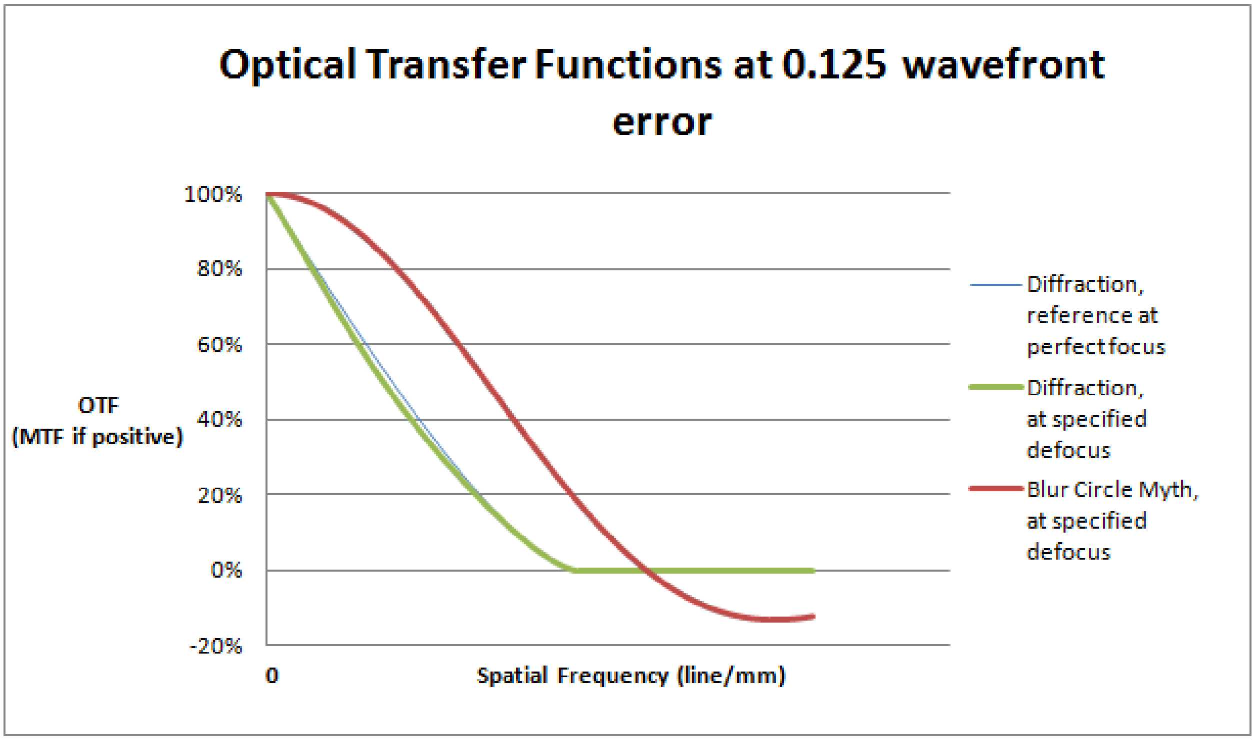
First, the vertical axis is labeled OTF (Optical Transfer Function). OTF is just like MTF except for one critical difference: MTF is always positive but OTF can go negative. Where OTF goes negative, it represents contrast inversion, the phenomenon in which dark and light bars trade places, as shown earlier in the comparison image for 3/4 lambda wavefront error. Lenses that are operating in the diffraction-limited regime do not have negative OTF for any frequency.

Second, the curve corresponding to the blur circle model is labeled "Blur Circle Myth". The word "Myth" is in there to emphasize that no real lens actually does what the blur circle model pretends that it does. In many situations the blur circle model is plenty good enough to make useful predictions. But as you approach perfect focus, the blur circle model becomes progressively less accurate and it becomes important to pay attention to what the more accurate wave optics model says.

Finally, notice that here at 1/4 lambda wavefront error, the real lens which is suffering from diffraction still has a cutoff frequency (lines/mm) that is over 50% higher than the blur circle model predicts. In exchange, there is some loss of contrast for low frequency detail, reaching its worst situation where the blur circle model predicts about 75% MTF while the more accurate diffraction model says 60%. Crossover between the two models occurs at a point where the MTF is about 40%.

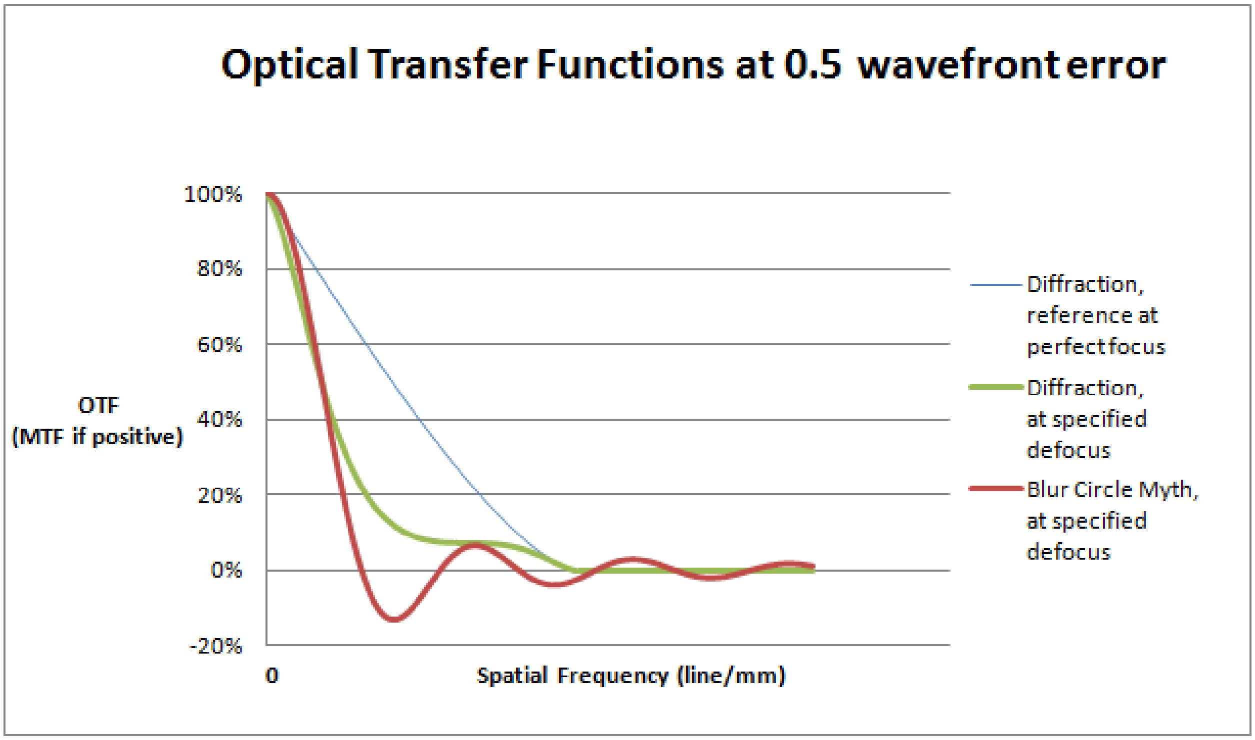
Although the curves have different shapes, it's not a huge stretch to say that here at 1/4 lambda wavefront error both curves represent similar overall performance.

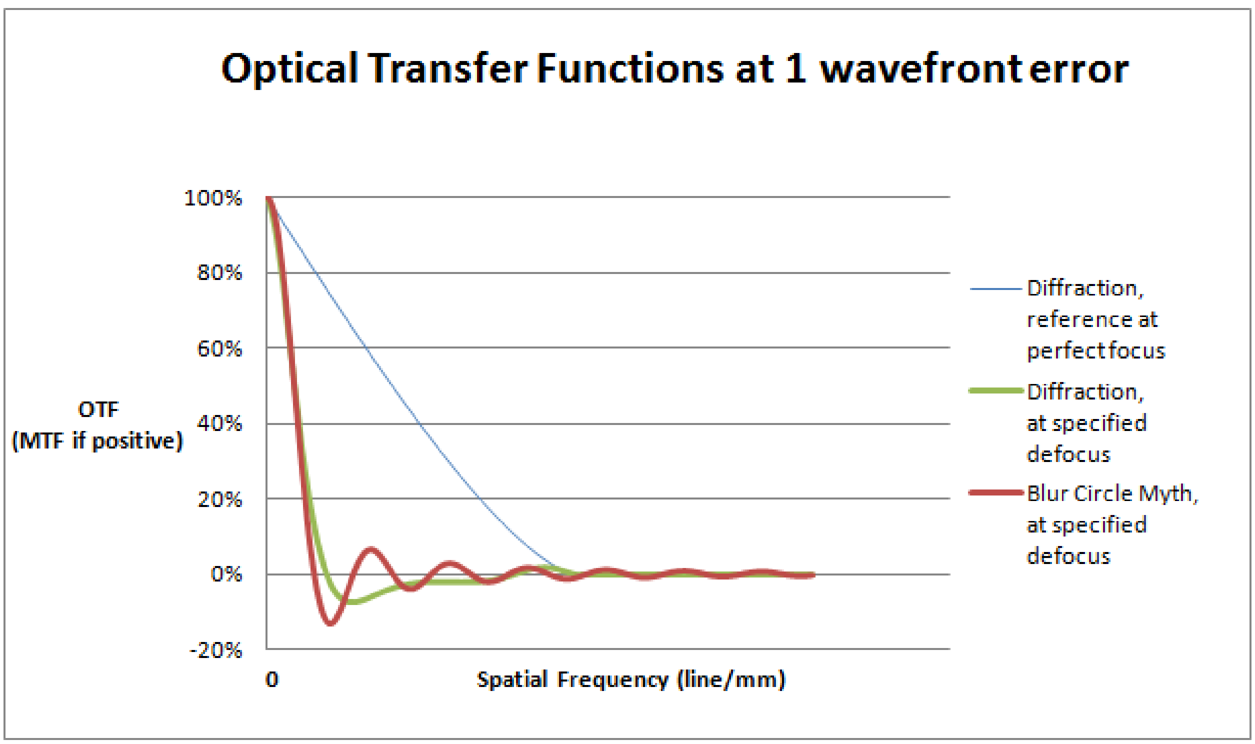
However, that situation changes dramatically at points closer to perfect focus. Here is the picture at 1/8 lambda wavefront error:



At this point, the Diffraction and Blur Circle curves have clearly diverged from each other. Here the lens is running in the diffraction-limited regime.

On the other hand, for larger amount of defocus, the Diffraction and Blur Circle curves match each other even better. Here are the pictures at 0.5 lambda and 1.0 lambda wavefront errors:





Personally, I find that it's difficult and probably pointless to keep track of all the details of this behavior. That's why I think in terms of those three regimes that I mentioned earlier: geometry-limited, transition, and diffraction-limited.

As a matter of practice, it's even simpler to ignore the transition zone and just split the behavior into only two regimes. If the COC is larger than the Airy disk, then compute DOF using the classic formula; if the Airy disk is larger than the COC, then compute DOF using the quarter-lambda formulas, or almost equivalently, set COC = Airy disk diameter and use the classic formula.

Conclusions

I confess, this section is here only because the article comes to an awkward end if it's omitted. Really, you've seen the conclusions already. Go read the summary and look at the pictures again if you're not sure.

Appendix: Technical details and references

The actual captured images shown here were shot with a Mitutoyo M Plan Apo 5X NA 0.14 objective on a Mitutoyo MT-1 tube lens, with the tube lens stopped down to f/80 using a 14-bladed iris closed around a 2.5 mm diameter drill. This gives NA 0.031 on the front side. Images were shot using a StackShot rail with nominal step size 0.005 mm. Running the numbers, this results in a front-side cutoff frequency of 114 line pairs/mm, almost exactly at Group 6, Element 6 in the high resolution chrome-on-glass 1951 USAF chart being used as a target. The 0.005 mm physical step size gives about 7 physical steps per each 1/32 lambda wavefront error selected to be shown in the datasets.

Images as shown here were captured using a Canon T1i camera, then reduced to 50% and color-balanced before being montaged by a Java program that also computed all the curves and PSFs.

OTF curves were computed using the formulas described at http://photo.net/learn/optics/lensTutorial#part5 , which trace back to H.H. Hopkins, "The frequency response of a defocused optical system," Proceedings of the Royal Society A, v. 231, London (1955), pp 91-103. Reprinted in Lionel Baker (ed), *Optical Transfer Function: Foundation and Theory*, SPIE Optical Engineering Press, 1992, pp 143-153. Diffracted PSF's were computed using numerical integration based on the same theory, averaging over 91 wavelengths from 400 to 700 nm as seen by a sensor with nominal sensitivity as shown at http://www.photomacrography.net/forum/viewtopic.php?p=145697#145697. Calculation of the PSFs was confirmed by physical experiments in the range of 1-7 lambda wavefront error (presented as a "puzzle piece" at http://www.photomacrography.net/forum/viewtopic.php?t=23448), and subsequently by comparison with the irradiance profiles computed by http://wyant.optics.arizona.edu/fresnelZones/fresnelZones.htm, as described by http://wyant.optics.arizona.edu/webMathematica/myprograms/fresnelZones/FresnelDiffraction.pdf.

Other useful info at http://www.bobatkins.com/photography/technical/resolution.html, http://photo.net/learn/optics/mtf/, http://www.lonestardigital.com/aperture_diffraction_limits.htm,

http://www.cambridgeincolour.com/tutorials/diffraction-photography.htm,
http://voi.opt.uh.edu/VOI/WavefrontCongress/2006/presentations/1ROORDAprincip.pdf,

There's an interesting tie-in with the question of "What's the sharpest aperture for a given DOF?", as discussed by http://www.largeformatphotography.info/articles/DoFinDepth.pdf, http://www.kenrockwell.com/tech/focus.htm, and http://www.cs.cmu.edu/~ILIM/courses/vision-sensors/readings/ViewCam.pdf, but that's a topic for another day.

--- the end ---