Theoretical Analysis of Coronagraphs

Olivier Guyon
Subaru Telescope

ABSTRACT

Many high contrast coronagraph concepts have recently been proposed, and many more will likely be developed in the near future. In this paper, I look at the intrinsic performance each concept can theoretically offer (with perfect wavefront, perfect optics). I also show that simple physics sets fundamental limits on the performance of coronagraphs—these limits are quantified and compared to existing designs. This theoretical work provides valuable insight into how to design high performance coronagraphs and explains many of the characteristics of particular designs. This new perspective reveals the importance of stellar angular diameter on coronagraph performance.

The simulated observation a sample of target stars reveals very large performance differences between coronagraph designs considered for exoplanet detection. These performance differences are equivalent to changing the telescope diameter by a factor 2 to 3. This analysis shows that by adopting a high performance coronagraph design, the TPF-C science goals which were previously thought to require an ~8m telescope, could be achieved with a 4m telescope.

TABLE 1
CORONAGRAPHS ABLE TO ACHIEVE 10^{10} PSF CONTRAST WITHIN 5 λ/d

Coronagraph	abrev.	reference	Design(s) adopted
"Interferometric" Coronagraphs			
Achromatic Interferometric Coronagraph	AIC	Baudoz et al. (2000)	
Common-Path Achromatic Interferometer-Coronagraph	CPAIC	Tavrov et al. (2005)	(=AIC)
Visible Nulling Coronagraph, X-Y shear (4th order null)[a]	VNC	Mennesson et al. (2003)	Shear distance = ± 0.3 pupil radius
Pupil Swapping Coronagraph	PSC	Guyon & Shao (2006)	Shear distance = 0.4 pupil diameter
Pupil apodization			
Conventional Pupil Apodization and Shaped-Pupil[b]	CPA	Kasdin et al. (2003)	Prolate[c]($r = 4.2\lambda/d$, 8% throughput)
Achromatic Pupil Phase Apodization	PPA	Yang & Kostinski (2004)	$\phi = \phi_2(x) + \phi_2(y)$: $a = 2$: $\epsilon = 0.01$
Phase Induced Amplitude Apodization Coronagraph	PIAAC	Guyon (2003)	Prolate apodization
Phase Induced Zonal Zernike Apodization	PIZZA	Martinache (2004)	Not simulated
Improvement on the Lyot concept with amplitude masks			
Apodized Pupil Lyot Coronagraph	APLC	Soummer et al. (2003)	$r = 1.8\lambda/d$
Apodized Pupil Lyot Coronagraph, N steps	$APLC_N$	Aime & Soummer (2004)	$(N, r) = (2, 1.4)$: $(3, 1.2)$: $(4, 1.0)$
Band limited, 4th order[a]	BL4	Kuchner & Traub (2002)	$\sin^4$ intensity mask, $\epsilon = 0.21$
Band limited, 8th order	BL8	Kuchner et al. (2005)	$m = 1, l = 3, \epsilon = 0.6$
Improvement on the Lyot concept with phase masks			
Phase Mask	PM	Roddier & Roddier (1997)	with mild prolate pupil apod.
4 quadrant	4QPM	Rouan et al. (2000)	
Achromatic Phase Knife Coronagraph	APKC	Abe et al. (2001)	(=4QPM)
Optical Vortex Coronagraph, topological charge m	OVC_m	Palacios (2005)	$m = 2, 4, 6, 8$
Angular Groove Phase Mask Coronagraph	AGPMC	Mawet et al. (2005)	(=OVC)
Optical Differentiation	ODC	Oti et al. (2005)	mask: $x \times \exp^{-(x/10)^2}$ d

[a]The Visible Nulling Coronagraph (VNC) and Band limited 4th order (BL4) coronagraphs belong to the same class of pupil-shearing 4th order coronagraphs, and are simply 2 ways of achieving the same result. They can be designed to have exactly the same performance. In this Table, the VNC is chosen with a small IWA and 2 orthogonal shear directions, while the BL4 is designed with a larger IWA and 2 shears in the same direction. To reflect this similarity, they are referred to as VNC/BL4(1) for the small IWA option (listed as VNC in this Table) and VNC/BL4(2) for the large IWA option (listed as BL4 in this Table).

[b]The CPA design adopted here is a continuous apodization (rather than binary apodization/shaped pupil) which maximizes the radially averaged performance at $\approx 4\lambda/d$. More optimal designs exist in other conditions: CPA with high contrast at specific position angles for observations at $\approx 3\lambda/d$ or high throughput CPA for observations at $> 4\lambda/d$.

[c]CPA, APLC, $APLC_N$: r is the radius, in λ/d, of the mask within which the circular prolate function is invariant to a Hankel transform. This parameter is half of the mask diameter a defined in Soummer et al. (2003).

[d]ODC: x is in λ/d. Maximum mask transmission at $7\lambda/d$. Lyot pupil mask radius = 0.85 times pupil radius.

An inventory of current coronagraph designs

The pace at which new coronagraph designs appear seem to be ever accelerating, but I have attempted to provide a snapshot of current coronagraph design in Table 1. Fortunately, they can be grouped in 3 broad categories:

- Inteferometric coronagraphs: The coronagraph design is inspired from interferometry; the single telescope pupil is used to extract several beams which are coherently mixed to produce a null
- Pupil Apodization: These design create an apodized pupil without sharp edges, which is then used to form a high contrast image
- Improvement on the Lyot design: This very rich family of coronagraphs uses a combination of focal and pupil plane masks to selectively reject starlight.

The "external occulter" concept, where a physical mask is placed far in front the telescope to mask the star, is not considered in this study, and is not listed in Table 1.

Fundamental limits of coronagraph performance & comparison to existing coronagraph designs

Before attempting to find fundamental limits and compare coronagraph performances, a performance metric needs to be defined. I propose here to use the **useful throughput,** which is maximum amount of planet light which can be collected (most often by summing up pixels in a focal plane) without collecting at the same time a greater quantity of starlight. This definition is somewhat arbitrary (maybe we can still use planet light which is superimposed on much brighter starlight?) and overly simple (it does not describe well what happens in the presence of exodozi for example). It is however very helpful because it can be applied to any coronagraph design, and, although not perfect, is still a good representation of coronagraph performance.

What is the theoretical limit of coronagraph performance? Clearly, there must be such a limit: it must be impossible to have an optical system remove most of the starlight while transmitting most of the planet light if the planet-star separation is much less than 1 λ/d. This limit can be accurately computed by recognizing **optical systems are linear in complex amplitude**. The complex amplitude in any point of the system (for example a pixel in the focal plane) is a linear function of the complex amplitude in the entrance pupil of the telescope. Fourier transforms (commonly used in coronagraph modeling) and diffraction propagation are linear in complex amplitude. The linear relationship between input (pupil) and output (final focal plane) is entirely fixed by the coronagraph design. We can now think of a coronagraph as a linear system which has to filter out a flat non-tilted wavefront in the pupil plane (= on axis point source). The wavefront of an off-axis source can then be decomposed as the sum of an on-axis wavefront (which the coronagraph will remove) and a residual, which gets smaller as the point source is moved closer to the optical axis. This residual contains all of the energy that can be transmitted by the coronagraph. This simple framework allows us to now quantify a fundamental limit on coronagraph performance. Using the same framework, we can even prove that one could build such a coronagraph with a large number of beamsplitters, so this limit can be reached. It is unfortunately unknown if a simple practical way of building this coronagraph exists.

The stellar angular size problem

A quick glance a Figure 1 paints a very rosy picture of the coronagraph world: the theoretical limit shows that we should be able to get ~50% throughput at 0.5 λ/d separation. We even happen to have several coronagraphs at hand which are getting fairly close to it. Incidentally, one of these (the AIC) is achromatic by design and optically simple to build. It looks as if TPF-C must be possible with a ~2m telescope! Checking (see Figure 2) what happens when taking into account the fact that stars are not points but small disks reveals a true problem in the coronagraph world: none of the coronagraphs which appeared to be "high performance" can cope with the small stellar angular diameter. The star diameter (which is 1% of the planet-star separation for the Sun-Earth system) simply prevents high contrast: the coronagraphic leak for a point at the edge of the star far exceeds 1e-10 for these "high performance" coronagraphs. Worse: even the theoretical limit obtained with our linearity framework is hugely degraded by making the star a fraction of λ/d. It would seem odd that making the star 0.01 λ/d in radius moved the theoretical inner working angle limit from ~0.5 λ/d to more than 1 λ/d. This can however be explained by our framework: when the star gets a little bigger, the coronagraph needs to reject more than just the flat on-axis wavefront. The number of "wavefront modes" which need to be rejected grows, and the wavefront from an off-axis source at ~1 λ/d starts to "look like" these additional modes.

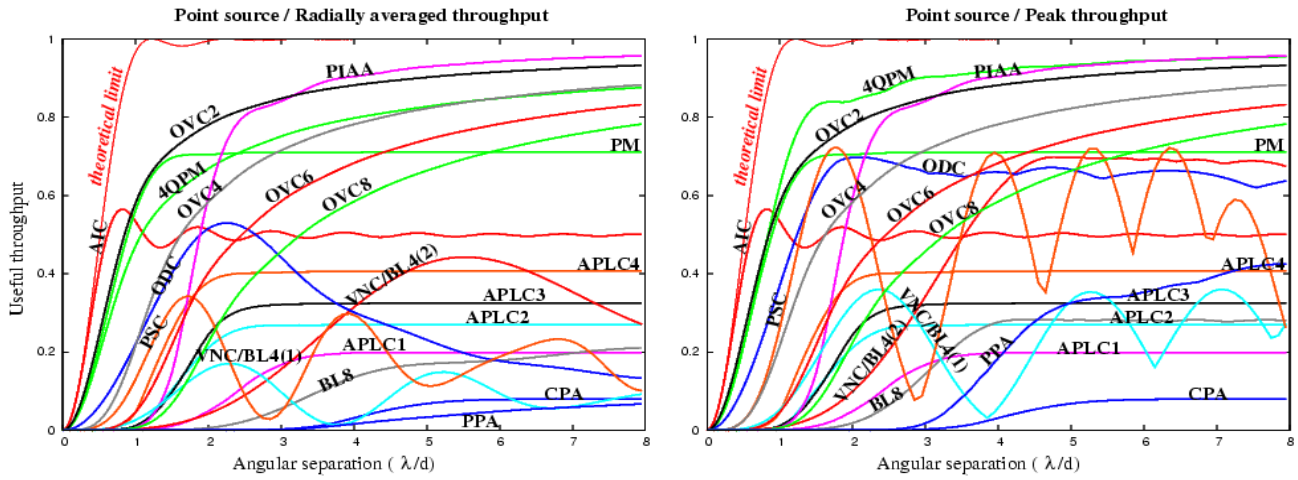

Figure 1: *Useful throughput, at the 1e-10 contrast level, of several coronagraph designs. The theoretical limit of useful throughput is also shown.*

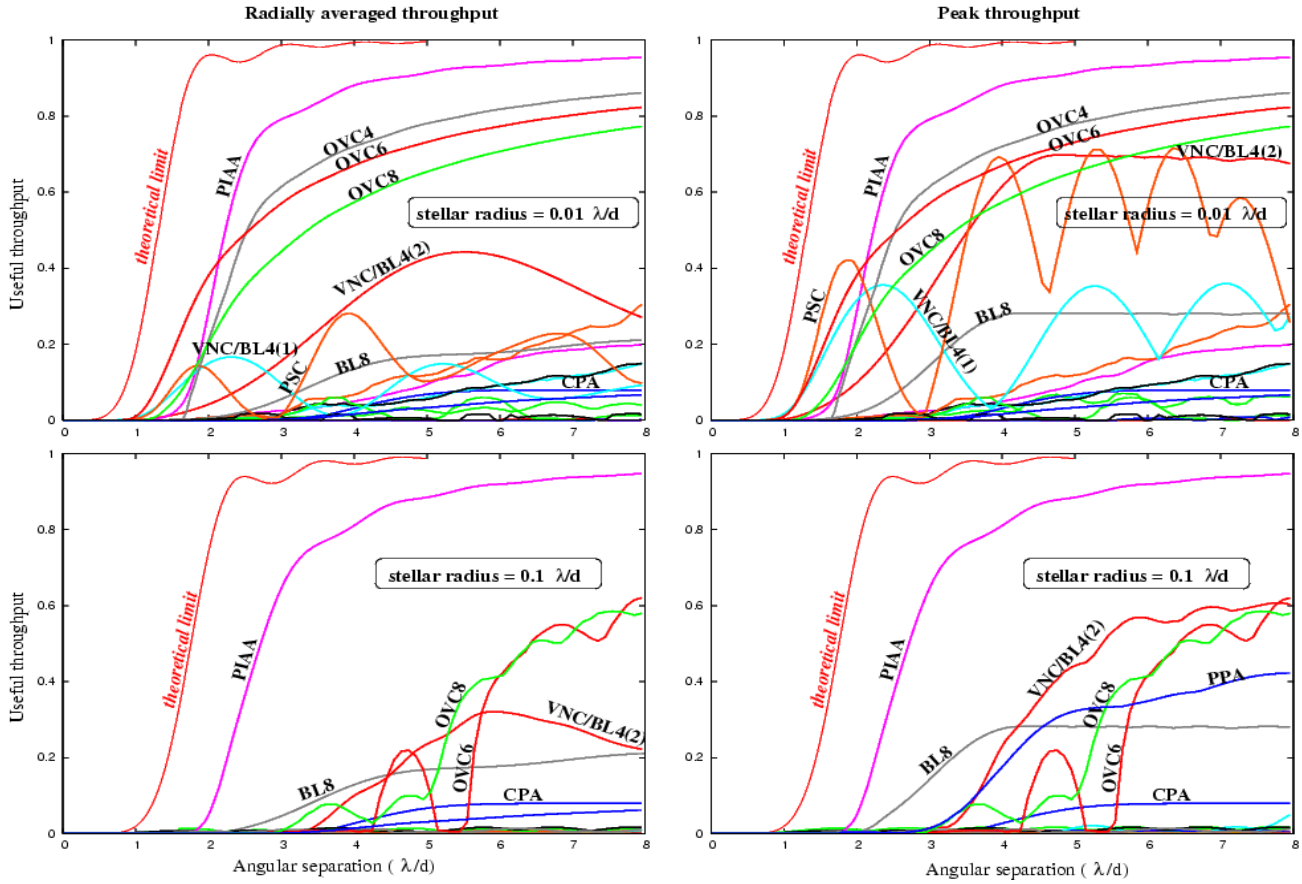

Figure 2: *Same as figure 1, but with non-zero angular stellar diameter. Note that the PPA coronagraph was omitted in the upper plots, but included in the lower plots (it is not affected by stellar angular diameter).*

Now that stellar diameter is taken into account, the most promising coronagraphs appear to be the PIAA and OVC. Even though they can function cope with stellar angular diameter, their performance degrades as the star gets bigger (from 0.01 λ/d to 0.1 λ/d).The BL8 and CPA coronagraphs, even though they offer significantly lower performance than the PIAA or OVC, appear to be unaffected by stellar angular size.

Zodiacal and Exozodiacal light & Coronagraph design

So far, we have developed an accurate model of coronagraph performance for looking at a single planet around a partially resolved star. The useful throughput tells us how many photons we can hope to detect from the planet in a given exposure time with a given telescope diameter. We have however not included zodiacal or exozodiacal backgrounds, which can easily overwhelm the planet signal.

Unlike starlight (which can be removed by the coronagraph), not much can be done about background light. For distant targets, the planet is instrisically fainter, but both the zodi and exozodi have the same surface brightness: the effect of zodi and exozodi is especially strong for distant targets. Although the level of background light is independent of coronagraph design, the amount of background light mixed with the planet image is not. Ideally, this amount is the equal the the background level integrated over a ~1 λ/d disk (the size of the plane image in a conventional telescope). Unfortunately, many coronagraphs do not preserve the full angular resolution of the telescope, or produce non-ideal off-axis PSFs (for example, PSFs in which a large amount of flux is scattered in large "wings"). This effect can be quantified by the diffraction efficiency factor (DEF), which tells us how much more background light is mixed with the planet image compared to a normal diffraction limited imaging system. DEFs for the coronagraphs studied in this paper are listed in Table 2.

Wavefront control

From the coronagraph point of view, wavefront errors appear to be a fatality: if the telescope upstream of the coronagraph has a "wiggly" primary mirror, the coronagraph contrast is going to be degraded. It would also seem that "high performance" coronagraphs (PIAA, OVC) with smaller IWA, since they are more sensitive to low order aberrations, have a more serious problem with low order wavefront errors. Mid-spatial frequencies which create "speckles" at the planet's position affect all coronagraphs equally, since the speckle contrast is simply a function of the magnitude of the wavefront aberration.

In fact, the coronagraph has a central role in the wavefront control scheme for the telescope, as it provides the precious photons for wavefront sensing:

(1) Low IWAs enables active closed-loop control of low order aberrations. Low order aberrations are therefore much smaller in a "high performance" coronagraph than in a large-IWA coronagraph
(2) High throughput maximizes the number of photons available for wavefront sensing, increasing the temporal bandwidth of the wavefront control loop. With a higher throughput coronagraph, the requirements on the stability of the telescope optics can be relaxed.

Coronagraph designs characteristics: Summary

Table 2 lists, for the coronagraphs of Table 1, the values obtained for the performance metrics used in this study: IWAs and throughput derived from the useful throughput vs. angular separation curves (and how stellar angular diameter affects the IWA), angular resolution and DEFs. The last line of table shows the performance of the "ideal" coronagraph constructed with the theoretical framework used in this paper.

This table shows that the most promising coronagraph designs are the PIAA and OVC. The BL8, CPA, and VNC/BL4 also appear to be suitable for exoplanet detection, but have a lower efficiency (lower throughput, higher DEF, larger IWA and poor angular resolution).

Coronagraph performance for exoplanet imaging

The theoretical model and performance metrics used so far should provide a fairly complete and accurate picture of coronagraph performances. To verify the conclusion obtained, and to derive an accurate estimate of the coronagraph performance for exoplanet imaging, a Monte-Carlo imaging simulation of coronagraphic observations was developed.

A sample of target was first established, assuming each sample star has an Earth-like planet. Coronagraphic images were computed for each target and possible planet position along its orbit. Sample images can be seen for star HIP56997 in Figure 3.

TABLE 2

CORONAGRAPH CHARACTERISTICS FOR 10^{10} CONTRAST

Coronagraph name	Throughput average	Throughput peak	Angular resolution[a]	diffractive eff. fact.[b]	IWA[c]	IWA(5%) with stellar radius[d] $0\ \lambda/d$	$0.001\lambda/d$	$0.01\lambda/d$	$0.1\lambda/d$
AIC	0.50	0.57	1.0	0.5	0.38	0.15	6.14	> 8.0	> 8.0
VNC/BL4(1)	0.09	0.36	1.65[e]	0.37[e]	1.49	1.27	1.27	1.43	> 8.0
PSC	0.18	0.73	1.0[f]	0.5[f]	1.13	0.79	0.79	3.24	> 8.0
CPA	0.08	0.08	1.81	0.31[g]	4.20	4.42	4.42	4.42	4.47
PPA	0.08[h]	0.33[h]	1.35	0.18	3.91	3.29	3.29	3.30	4.02
APLC	0.20	0.20	1.22	0.67	2.41	2.01	2.17	4.02	> 8.0
$APLC_2$	0.27	0.27	1.11	0.81	1.64	1.26	1.78	4.71	> 8.0
VNC/BL4(2)	0.35	0.70	1.19	0.71	3.02	2.17	2.20	2.21	3.77
BL8	0.26	0.28	1.86	0.29	2.96	3.04	3.04	3.04	3.05
PM	0.71	0.71	1.0	1.0	0.69	0.23	2.94	3.59	> 8.0
4QPM	1.0	1.0	1.0	1.0	0.84	0.28	2.39	> 8.0	> 8.0
OVC_2	1.0	1.0	1.0	1.0	0.84	0.21	2.21	> 8.0	> 8.0
OVC_4	1.0	1.0	1.0	1.0	1.62	0.62	0.66	1.74	> 8.0
OVC_6	1.0	1.0	1.0	1.0	2.45	1.02	1.02	1.19	4.32
OVC_8	1.0	1.0	1.0	1.0	3.26	1.41	1.42	1.42	3.34
ODC	-[i]	0.70	1.18[j]	0.72[j]	0.97	0.50	4.59	> 8.0	> 8.0
PIAA	1.0	1.0	1.0	1.0	1.88	1.28	1.38	1.66	2.0
ICC_6	**1.0**	**1.0**	**1.0**	**1.0**	**1.32**	**0.76**	**0.76**	**0.76**	**1.76**

[a] Square root of the smallest focal plane area containing half of the planet's light. Normalized to 1.0 for the Airy pattern of an non-coronagraphic telescope. In the case of the AIC, the angular resolution is measured on one of the 2 images.

[b] Diffractive efficiency factor measured at $20\lambda/d$ unless otherwise specified.

[c] Angular separation at which the planet's useful throughput first reaches half of the peak throughput. This is the standard definition of inner working angle (IWA): it assumes that the planet is favorably placed (no radial average) and that the star is point-like.

[d] Angular separation at which the planet's radially averaged useful throughput first reaches 5% for the stellar radius shown.

[e] Measured at $2.36\lambda/d$.

[f] Measured at $1.77\lambda/d$.

[g] Higher if the apodization is performed by a binary mask

[h] Only the main PSF diffraction peak is considered in the throughput

[i] The outer part of the field is not transmitted in the ODC

[j] Measured at $2\lambda/d$.

Figure 3 nicely illustrates the coronagraphic characteristics quantified in the previous sections:

- **Throughput**. The CPA, BL8, and to a lesser extent VNC/BL4(2) suffer from low coronagraphic throughput. As a result, the planet's image, even if well outside the coronagraph mask's influence, appears noisy (few photons detected). The PIAAC, OVC6 and ICC6, on the other hand, enjoy nearly 100% throughput: the planet image is brighter and less noisy.
- **Angular** resolution. The CPA, BL8, and to a lesser extent VNC/BL4(2) have poorer angular resolution: the planet image is larger and more zodi/exozodi light is mixed with it.
- **Ability to work at small angular separation**. None of the coronagraph tested can detect the planet on a 2m telescope. Detection appears feasible on a 4m telescope with the VNC/BL4(2), PIAAC, OVC6 and ICC6, but requires a 6m telescope with the BL8. Finally, an 8m telescope is needed for detection with a CPA.

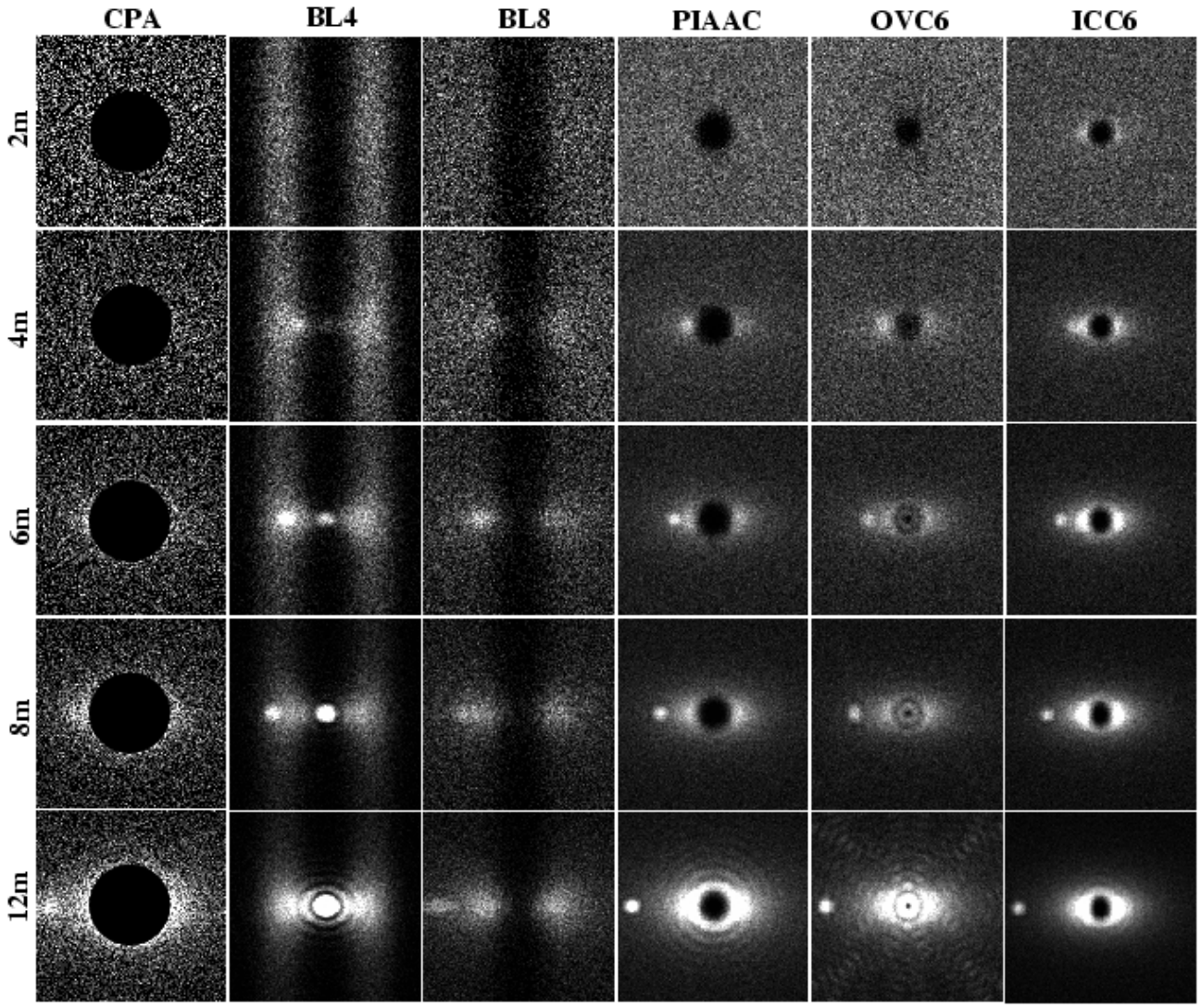

Figure 3: *Simulated 4 hours exposures of HIP56997 and a hypothetical Earth-type planet at maximum elongation for telescopes sizes ranging from 2m to 12m. HIP56997 is a G8 tye main sequence star at 9.54pc. Each image assumes a perfect detector, mV=22.95/sqarcsec zodiacal background, a 1 zodi exozodi cloud, a 25% telescope+camera throughpu, and a 0.1 micron bandpass centered at 0.55 micron. The system inclination for this particular simulation was arbitrarily set at i=59 deg. Each image is 20 × 20 λ/d, and the planet-star separation is 80 mas.*

– **Sensitivity to stellar angular size**. The CPA and BL8 are extremely robust to stellar angular size: starlight leaks are virtually nonexistent even on the 12m telescope. With the VNC/BL4(2), PIAAC, OVC6 and ICC6, starlight is visible on the 12m telescope (equivalent to a 6m telescope observing the same system at 5pc), although it is still fainter that the exozodiacal contribution.

Images similar to the ones showed in Figure 3 have been computed for all possible positions of the planet along its orbit. With photon noise the only source of noise, the exposure time required to detect the planet (SNR=7 detection threshold) is computed for each planet position. For a given exposure time, the **planet detection probability** can then be measured as the fraction of the planet orbit along which it can be detected. These results are shown for HIP56997 in the table below for a selection of coronagraph designs, including the theoretically ideal ICC6 coronagraph.

TABLE 3

THEORETICAL EXPOSURE TIMES REQUIRED TO REACH 20% AND 50% DETECTION (SNR=7) PROBABILITY ON HIP 56997 IN THE 0.5–0.6 μM BAND.

Tel. Diam.	CPA	VNC/BL4(2)	BL8	PIAAC	OVC6	ICC6
			20% detection probability			
2m	... [a]	...	...	...	...	12.4 hour
4m	...	2.1 hour	11.4 hour	34 min	48 min	14.3 min
6m	6.9 hour	14 min	44.8 min	5.7 min	8.6 min	3.78 min
8m	63 min	4.8 min	17 min	2.1 min	3.1 min	110 sec
12m	12 min	2.7 min	4.3 min	45 sec	56 sec	42 sec
			50% detection probability			
2m	...	...	...	...	...	...
4m	...	12.7 hour	...	2.3 hour	2.2 hour	44.1 min
6m	...	59 min	5.0 hour	14 min	19.8 min	12.23 min
8m	4.3 hour	18 min	1.8 hour	5.8 min	7.6 min	5.4 min
12m	31 min	55 min	19 min	2.0 min	2.4 min	113 sec

[a]Exposure times above 1 day are not shown in this Table, as they are practically unrealistic.

This table shows the large performance differences between coronagraph designs. To reach a 50% detection probability for HIP56997's planet, the CPA requires 4.3 hours of exposure time, while the PIAAC requires 5.8 min. Comparison of exposure times in this table suggests that a PIAAC or OVC6 is equivalent to a CPA or BL8 on a telescope 2 to 3 times larger.

This exposure time vs. detection probability analysis was repeated for each star of the sample, for which an "exposure time required to reach 50% planet detection probability" was assigned. For a given telescope/coronagraph combination, the sample stars can then be ranked from the easiest (shorter exposure time) to the hardest (longer exposure time). A given total "open shutter" observation time can then be optimally assigned to maximize the number of targets sampled at the 50% detection probability level. Results of this analysis are shown in Table 4.

Table 4 confirms the results obtained so far:

- The PIAAC and OVC appear to be the most promising coronagraphic designs.
- The theoretically ideal coronagraph ICC6, thanks to its smaller IWA, would enable a larger sample size than allowed by either PIAAC or OVC. For a small sample size however, most the targets do not require this small IWA, and the PIAAC and OVC performance is not largely different from the theoretical optimal coronagraph
- The CPA and BL8 coronagraphs would require a telescope about twice as large to reach the same exoplanet detection productivity than the PIAAC or OVC.

The analysis presented in this paper ignores important challenges which must be overcome to reach the performance levels described in this paper:

- wavefront control
- chromaticity

More fundamentally, the assumption that detection is photon noise limited is optimistic: ETPs are embedded in an exozodiacal light unlikely to be perfectly smooth and symmetrical. Since good angular resolution may help to distinguish between a planet and exozodiacal light features (such as arcs or rings), it is important to avoid using coronagraphs that clip or strongly attenuate the edge of the telescope pupil if at all possible.

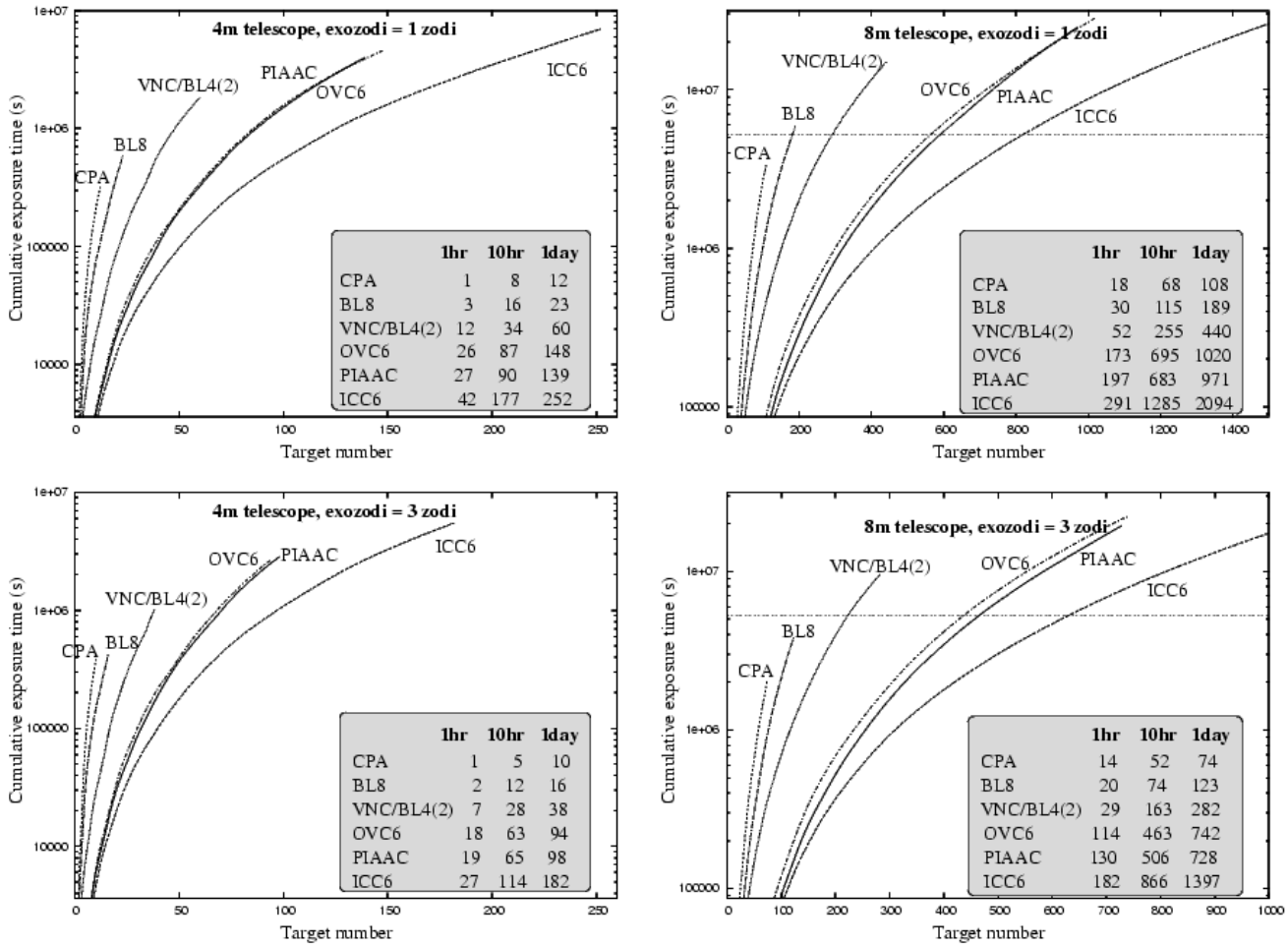

Figure 4: *Total cumulative exposure times required to reach a 50% planet detection (SNR=7) probability for a single observation as a function of number of targets. For all simulations, a 25% throughput in the 0.5–0.6 micron band is adopted, and targets are ordered with increasing exposure time. Each curve terminates when the required exposure time per target reaches 1 day. Results are shown for a 4m telescope (left) and an 8m telescope (right) with exozodi levels of 1 zodi (top) and 3 zodi (bottom). The number of accessible targets for which the required exposure times are less than 1 hour, 10 hour and 1 day are listed for each case in the Grey boxes. In the 8m telescope plots, the horizontal line corresponds to a 2 months ``shutter open'' cumulative exposure time, and may be considered as a practical limit on the number of targets that can be visited.*

The results obtained in this work should therefore be cautiously considered as upper limits on what may be achieved with a particular coronagraph/space telescope design for the specific goal of imaging ETPs around nearby stars.

ACKNOWLEDGEMENTS:

This work was carried out with the support of the National Astronomical Observatory of Japan and JPL contract numbers 1254445 and 1257767 for Development of Technologies for the Terrestrial Planet Finder Mission.

A more detailed report of this study, including the full list of references, can be found in:
"Theoretical Limits on Extrasolar Terrestrial Planet Detection with Coronagraphs" Guyon, O., Pluzhnik, E.A., Kuchner, M.J., Collins, B., Ridgway, S.T. ApJS, 167, 81 (2006) astro-ph/0608506