University of Nottingham
UK | CHINA | MALAYSIA

Transmissive Solar Sailing

Thesis submitted to the University of Nottingham for the degree of **Doctor of Philosophy, May 2024.**

Samuel Michael Thompson

14288390

Supervised by

Chantal Cappelletti
Chris Tuck
Nishanth Pushparaj

Signature _____

Date _____ / _____ / _____

Abstract

Uniquely, solar sails are a means of accelerating a spacecraft via ambient space radiation rather than onboard propellant. In principle, this allows for continuous acceleration, and for the attaining of higher velocities than a reaction engine may achieve. These sails are also amongst the most sustainable orbit control systems, being associated with reduced satellite weight, mission cost and space debris footprint. However, they also have weaknesses that constrain how and where they can be used — most notably, their poor performance in low Earth orbit (LEO), where the majority of satellite missions occur. *Transmissive* solar sails may mitigate these weaknesses and thereby promote the broader adoption of solar sails. Advocating for this future, this thesis seeks to collate and expand the existing literature pertaining to these sails across three core domains: orbital dynamics, optical design, and manufacturing processes.

A flight model was developed to characterise the behaviour of differently designed and differently steered solar sails in LEO, with consideration given to atmospheric drag, eclipse and orbital precession during orbit-raising manoeuvres. These designs included transmissive sail proposals of both *refractive* and *diffractive* varieties and a contemporary *reflective* solar sail for comparison. Simplified 'zero-α' steering was found to constrain the flight envelope of transmissive sails. When optimally steered, these sails unanimously demonstrated lower performance sensitivity to altitude but greater sensitivity to orbital inclination than contemporaries. Greater synergy with certain Sun-synchronous orbits (SSO) was

found, but also greater sensitivity to the orbital plane changes that arise in non-SSOs. Sails were sorted according to performance: some refractive and diffractive *metasails* surpassed reflectors in every flight regime (*Type A* transmissive sails); simpler refractive and diffractive sails surpassed reflectors only in low altitude, high inclination orbits, particularly SSO (*Type B*); other diffractive sails surpassed reflectors in low altitude orbits more generally, and were otherwise comparable to reflectors (*Type C*). A case study with a 36 m^2 Type A sail demonstrated transit times from LEO that were comparable to mid-range electric thrusters.

The optical design of refractive sail patterns using single-index materials was explored in-depth via numerical optimisation: a ray tracing simulation was developed to calculate the solar radiation pressure (SRP) and torque per unit area generated by refractive objects illuminated in vacuum. A model-free reinforcement learning optimiser was developed to iterate upon geometries according to a user-defined fitness function. These tools were integrated to generate numerically optimised patterns ranging from high acceleration micro-prisms to passively stable freeform *lightfoils*. The optimiser was shown to substantially improve the performance of optical elements, usually by harnessing total internal reflection. In one notable example, a simple pattern of cylinders evolved into a pattern of freeform elements that achieved 74% of the theoretical maximum tangential SRP — compared to the 42% efficiency reported for single-index refractive designs in prior literature. In another case, a single lightfoil was optimised for stability, and had its maximum corrective torque increased by 50% over that of the initial proposal. However, the optimisation process was shown to be highly sensitive to the configuration of both the simulation and the optimiser AI. For example, identical fitness functions applied to different initial geometries often resulted in very different solutions. Furthermore, geometries were seen to lose resolution during optimisation as a result of the optimiser removing 'low reward' vertices from convex hull calculation. These findings demonstrate the validity — but also the complexity — of using model-free reinforcement learning for the design of

non-imaging optical devices.

The manufacturing of transmissive solar sails was investigated with a focus on feasibility, cost, and accessibility from the perspective of a small satellite developer. The viability and cost of executing processes in-house was compared with the cost of commercial outsourcing, and several in-house trials were conducted. For fabricating the unpatterned sail sections, thin film fabrication trials were carried out using doctor blading and a water-float method. For fabricating patterned moulds, greyscale lithography and dry plasma etching comprised the conventional manufacturing methods trialled, while digital light processing and two-photon polymerisation comprised the additive methods trialled. Finally, pattern transfer trials included silicone moulding, linear thermal nanoimprint lithography, and roll-to-plate UV nanoimprint lithography. The evaluation favoured a commercial solution for the thin film fabrication stage. For mould manufacture, several viable conventional and additive methods were discussed, but a clear winner was not identified. For the pattern transfer stage, procuring a rolling nanoimprinter for in-house use was favoured above a certain volume or size of sail. A prototype solar sail payload was also designed for the University of Nottingham CubeSat *JamSail* in preparation for a future in-orbit demonstration mission.

Acknowledgements

This research project was made possible by the exertions of many patient souls. I first wish to express my deepest gratitude towards my academic supervisory team, comprising Dr. Chantal Cappelletti, Dr. Chris Tuck and Dr. Nishanth Pushparaj. Your mentorship has been invaluable to my development as an engineer, and I thank you for this opportunity. I also wish to thank my peers within the NottSpace team and commemorate our wayward adventures, which ranged from preparing mission proposals at the airport, to hastily storm-proofing ground stations on the laboratory roof. The many hours that we have spent interrogating a misbehaving CubeSat will not soon be forgotten. Special thanks to Daniel Robson for being my academic big brother and taking me underwing when it was needed most. I also wish to acknowledge the contribution of the Engineering and Physical Sciences Research Council (EPSRC) for generously funding my research project (grant numbers: EP/T517902/1, EP/R513283/1). Furthermore, my project took me across the University, and several wayward heroes from different departments stepped up to help me realise it. I wish to thank Dr. James Sharp for lending me his laboratory, time, and considerable expertise with thin film fabrication at the School of Physics and Astronomy; I wish to thank Dr. Jason Hutchinson for generously conducting a two-photon polymerisation trial at the Biodiscovery Institute on my behalf; I wish to thank Dr. Dominic Mosses for spending many hours training me at the Advanced Manufacturing Building, and for the substantial commute that this entailed; I wish to thank Dr. Richard Cousins on the Nanoscale and Microscale Research Centre, without whom much of this project would have been impossible; as well as performing conventional manufacturing trials and microscopy on my behalf, Richard took an active interest in my work and went truly above and beyond.

Finally, I wish to thank my friends and family for their continued support. The unwavering enthusiasm of my parents was a singular blessing. Above all, I wish to thank Sophie Pickford for being with me for the entire journey. I couldn't have done it without her. A heartfelt shoutout to our dog, Arthur Pug, who kept me lucid during lockdown and proved an able research assistant.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Conventional 'reflective' solar sails are a relatively new yet well understood technology that are suited to a narrow band of niche applications, such as space exploration. Despite providing lower rates of acceleration than their contemporaries, they may reach higher velocities (achieve higher 'ΔV') because they utilise solar radiation pressure (SRP) rather than propellant; within reasonable proximity to a star, a solar sail may accelerate for far longer than a reaction engine. Theoretically, a solar sail that is not subject to optical degradation could continue to accelerate indefinitely. Additionally, the performance of a solar sail scales directly with its size. It has been theorised that large solar sails may be manufactured from orbit using in-space additive manufacturing (ISAM) methods [1], allowing for rates of acceleration comparable to chemical engines. By contrast, conventional propulsion systems see depreciating gains in acceleration and ΔV when they are scaled up, as described by the *Tsiolkovsky rocket equation*. Solar sails can also provide continuous torque for attitude control [2, 3]; they are not reliant on propellant as thrusters are, do not become 'saturated' as reaction wheels do, and can operate beyond low Earth orbit (LEO) where magnetorquers would cease to operate [4]. Furthermore, if there is an avenue by which solar sails can be made suitable for wide-scale adoption by the space sector, there is a

strong argument to be made for doing so. This is because solar sails are arguably the most sustainable option for manoeuvring and orienting a satellite: they do not contaminate other systems with exhaust emissions, reduce launch weight and mission costs, and would serve to passively mitigate space debris by their very adoption (see section 4.2.5).

In terms of functional, validated technologies, the state-of-the-art of solar sail design is defined by square quadrants of highly reflective, metallised polymer films (specifically, *aluminised*), deployed via some combination of rotational inertia, elastic energy or motor torque applied to four extendable booms. This describes every solar sail that has managed to deploy in space to date [5–8]. Therefore, in holistic terms, the design of functional sailcraft has changed relatively little in the last 15 years. The state-of-the-art for the SRP-generating membranes that compose these sails has also changed little, as aluminium is already highly efficient in its role and has few competitors. A rare advancement in membrane functionality was demonstrated by NASA's ACS3, which featured a metallised (*chromised*) emission layer on the underside for thermal stability [8]. This indirectly improves SRP over the duration of a mission by mitigating optical degradation [9]. SRP has otherwise been improved by minimising losses and sources of performance uncertainty, such as from sail wrinkling. However, while direct increases in the SRP capacity of reflective sails have been modest, improvements in sail acceleration have continued steadily. This has been achieved by reducing sail thickness (IKAROS, NanoSail: 7.5 μm, LightSail: 4.5 μm, ACS3: 2.1 μm [5–8]) and, with the notable outlier of the spin-deployed IKAROS, increasing deployed sail size (IKAROS: 196 m^2, NanoSail: 10 m^2, LightSail: 32 m^2, ACS3: 80 m^2 [5–8]). Overall, this has effected a trend of greater net force and greater mass efficiency (see section 2.2.5 for definitions). In terms of functionally validated attitude control, the state-of-the-art is the reflectivity control device (RCD), which may induce torques by affecting the distribution of SRP about the sail. This was demonstrated by IKAROS [5], and remains the only example of sail-based

attitude control being demonstrated in space.

Because two of these sailcraft successfully demonstrated functional solar sailing (JAXA's IKAROS [5]; the Planetary Society's LightSail-2 [7]), solar sails of this kind have a technological readiness level (TRL) of 6–7. It it noteworthy that, of the three that failed to perform solar sailing because of issues with the sail itself, two failed to accelerate due to atmospheric drag [6, 7]. The third is currently experiencing attitude control issues arising from boom deformation induced during deployment [8].

Despite their promise, it is clear that solar sails have matured slowly. This is true even when compared to other high ΔV propulsion systems, such as the electric thruster, which reached commercial operation (TRL 9) within 16 years of its debut [10]. This lack of demand is a symptom of the perception that solar sails are unsuitable for most modern satellites, for which LEO and geosynchronous orbit (GEO) are the dominant flight regimes. This perception is largely justified due to the following characteristics of reflective solar sails:

1. Constrained acceleration:

   The acceleration of a solar sail at 1 au (*astronomical unit*) is constrained by the available ambient solar irradiance, and by the size of sail that is able to be stowed, launched and deployed to convert this irradiance into SRP. Technologies such as beamed energy or ISAM may pose a viable means of relaxing either constraint in the future, but for the modern solar sail, only low, *non-impulsive* rates of acceleration are feasible. This is a problem because satellite missions have a time limit dictated by the shelf-life of their other systems; most LEO missions are expected to last for only 4-5 years [11]. Therefore, a rapidly accelerating *impulsive* system may be more attractive — provided that the ΔV requirements of the mission are not prohibitive.

2. Performance sensitivity to altitude and solar cycle:

   In LEO, a solar sail will induce significant atmospheric drag if its surface is presented to the incoming airflow. This is unavoidable for a reflective sail; presenting its reflective surface to the Sun in a conducive manner will *always* involve presenting its opposing surface to the incoming airflow (see section 2.1.3). The sail membrane may also be deformed by this airflow, degrading its ability to generate velocity-wise SRP. The result is that the ability of a solar sail to accelerate will fall as altitude falls (from around $1,000$ km) until orbital decay becomes unavoidable (around 600 km under mean solar activity) [12, 13]. This is a significant disadvantage because the majority of satellites operate entirely within LEO [14]. Moreover, the sensitivity of a solar sail to altitude is itself sensitive to solar activity. This is because heightened solar activity correlates with increased extreme-ultraviolet radiation, which heats the thermosphere and causes atmospheric density to rise. Atmospheric density, and therefore atmospheric drag, may change by an order of magnitude above the expected median as a result, albeit in a predictable manner [15].

3. Performance sensitivity to orbital plane orientation:

   A solar sail that is orbiting a body other than the Sun will exhibit sensitivity to the angle drawn between its orbital plane and the Sun (see section 2.1.4). Notably, orbits with low inclination with respect to the ecliptic will erode sail performance, amplify altitude sensitivity, and frequently encounter eclipses; so will orbits for which one of the two orbital nodes lies near perihelion. This is an issue because LEO missions often constitute a rideshare or 'piggyback' launch, whereby multiple satellites with conflicting orbital requirements are launched together. The injection orbit parameters of individual satellites may therefore be compromised. As a solar sail will be incompatible with certain orbits, its inclusion may restrict the launch opportunities of a satellite developer [16].

4. Performance uncertainties:

   Sail performance is difficult to model accurately as non-linearities may be introduced by billowing, wrinkling, material deformation, satellite self-shadowing and optical degradation [2, 17–20]. Additionally, uncertainties are presented by both the optical properties of a sail [21] and the expected solar irradiance [21, 22].

5. Complexity of steering:

   Solar sails operating from within a heliocentric orbit will have relatively simple manoeuvres to perform, and ample time to perform them. Conversely, solar sails operating from LEO must adjust their attitudes constantly in order to perform even the simplest manoeuvre. Furthermore, unless the sail can be retracted, steering must be performed at all times. In contrast, a reaction engine may perform its manoeuvre quickly, and then be turned off.

6. Impingement on other systems:

   The presence of a solar sail can affect several unrelated systems. Self-shadowing affects solar panels, antennas and optical payloads directly. Self-shadowing (and sail emission) will also cool a satellite, indirectly affecting temperature-sensitive systems such as batteries and biological payloads [23]. Furthermore, the performance of a solar sail depends on its attitude (orientation) relative to the Sun [2], and doing so incorrectly may lengthen the already lengthy manoeuvre times [24, 25]. As discussed above, steering must be maintained continuously; this may conflict with the operation of other systems that require pointing.

7. Dependence on deployment systems:

   Any system that involves moving parts is viewed with suspicion by satellite developers as they introduce single points of failure [26] and deployment

systems are amongst the most complex. Additionally, solar sails have been shown to promote vibration under insufficient tension [27]. Sufficient tension during deployment is particularly difficult to achieve for large sails and, left unchecked, may result in membrane drift that may jam a deployment mechanism [28].

The operational weaknesses described by bullets (1-6.) make solar sails an uncompetitive solution for typical, near-Earth satellite control, and outright infeasible for satellites operating in low LEO. In order for the space sector to embrace solar sails and the many advantages that they entail, these weaknesses must be adequately mitigated. Conventional, reflective solar sails may mitigate some of these issues in time by way of continuing to mature, and by the advent of certain key enabling technologies:

Beamed energy technologies may unconstrain irradiance while on-orbit manufacturing/ISAM technologies may unconstrain sail size; either approach may effectively unconstrain the (1.) acceleration of future solar sails. Meanwhile, the issue of (2.) altitude sensitivity cannot be mitigated by reflective solar sails, nor can the issue of (3.) sensitivity to orbital plane orientation. These three issues are the performance limiters; the remaining issues pertain to reliability and complexity. (4.) Performance uncertainty will be negated in time by more sophisticated models of the space environment, predictive models for the optical degradation of materials in space, and by the maturing of the reflective solar sail class — particularly the maturing of sail structures that can ensure that sail membranes in orbit conform more closely with the 'flat mirror' ideal. On the other hand, the issue of (5.) steering complexity is an innate feature of reflective solar sails that cannot be mitigated. The issue of (6.) impingement on other systems is also not easily mitigated, but it is a lesser issue; indeed, reaction engines impede other systems in different ways (e.g. contamination via exhaust emissions). Finally, the issue of (7.) deployment issues may be negated slowly via increasingly

robust sail deployment systems of ever-growing flight heritage; or more suddenly in the future, by making sail stowing unnecessary with the advent of on-orbit manufacturing and ISAM technologies.

While several of these issues are clearly mitigable without considering alternative classes of solar sail, many are not. In particular, the issues of (2.) altitude sensitivity, (5.) steering complexity and (6.) impingement on other systems are bound up with the reflective solar sail class.

This is where an alternative, 'transmissive' solar sail may be of assistance. Like a conventional solar sail, a transmissive one will generate SRP by redirecting sunlight. Unlike a conventional solar sail, a transmissive one will transmit sunlight through an optically transmissive and typically highly refractive membrane, rather than reflecting it with a metallic one. If perfectly flat — as for an idealised reflector — a transmissive sail will generate zero net SRP because refraction events on entry and egress of sunlight from the membrane will cancel out. To be useful, a transmissive solar sail must incorporate a surface pattern that can effectively harness either *refraction* or *diffraction* to redirect sunlight through the membrane, and for which a plethora of designs have been proposed [29–36]. With the exception of those designs that utilise metamaterials (so-called *metasails*), the functional elements of these sails are based on mature technologies, such as diffraction gratings.

As explored later in this report, the manner in which these sails generate SRP results in certain behaviours that are highly conducive to solar sailing. Advantages include a significant reduction in (2.) altitude sensitivity [29], as well as milder mitigations for (5.) steering complexity [34, 35] and (6.) impingement on other systems. Notably, these issues are not likely to be improved by maturing the reflective sail class. In the case of an enormous and idealised reflective sail of the future, atmospheric drag, rotational inertia, and self-shadowing will be made more prominent, exaggerating the issues of (2.), (5.) and (6.), respec-

tively. Furthermore, some transmissive sail designs have been proposed that may generate higher rates of acceleration than reflective sails even in interplanetary space, where they are not advantaged by their lower altitude sensitivity [30, 31, 33, 35, 37]. As such, these sails are a means of relaxing the constraints of (1.) sail acceleration prior to the advent of ISAM or beamed energy. On the other hand, it should be noted that beamed energy may provide some of the same benefits as switching to a transmissive sail (albeit perhaps at the expense of ceasing to be a low cost system; requiring expensive and precise laser infrastructure): increasing sail irradiance with beamed energy is a means of increasing the capacity of a sail to generate SRP without increasing its size, and therefore, without increasing its capacity for atmospheric drag. This may mitigate the issues of (1.) acceleration, (2.) altitude sensitivity, and may re-frame the issue of (3.) orbital plane orientation sensitivity (in the presence of several lasers that are able to fire from different locations, perhaps nullifying the issue entirely). However, beamed energy could also be applied to a transmissive sail [38] to enjoy the advantages of both innovations.

In summary, a transmissive sail may mitigate issues that even highly advanced reflective sails cannot. Its advantages have overlap and synergy with beam-propulsion, and it would use technologies that have greater readiness than either ISAM or beamed energy. As such, in the near-term, transmissive sail design could facilitate the adoption of solar sails by the terrestrial satellites that dominate the modern space sector, with all the benefits for cost, sustainability and ΔV that this would entail.

This is a glamorous future, but several questions remain to be answered: how substantial are these performance advantages for different sail types in different flight regimes, and are there disadvantages? What satellite applications would they be compatible with? Can the performance of higher readiness designs made of simple materials be brought closer to the performance of lower readiness but

higher performance *metasail* designs (proposals citing *metamaterial* optical properties)? And how easily are these simpler designs to realise for a small satellite developer? These are codified into three research questions:

> Question 1:
>
> In both the present and future space sectors, what roles could different transmissive solar sails fulfil according to their readiness and suitability for various orbital flight regimes, particularly within LEO?
>
> Question 2:
>
> Can the performance of the simplest, highest readiness transmissive solar sails be brought closer to that of the more advanced, lowest readiness transmissive metasails using model-free reinforcement learning?
>
> Question 3:
>
> Can transmissive solar sails be made economically and scalably, and if so, how?

In other words: how strong an argument can one make for transmissive solar sailing, how can the performance of these sails be maximised, and how soon can we bring them into operation? Answering these three research questions entails three bodies of work which can be loosely categorised as: orbital dynamics, optical design, and manufacturing processes.

By answering these questions and formulating an argument for transmissive sails, this thesis aims to hasten the development and adoption of this sail class. The other aim is to ease the process of orienting oneself within this vast body of research, particularly for the new researcher, by providing a resource that collates and expands the literature within these three domains. With respect to these domains and their corresponding research question, the objectives of this thesis are detailed below:

1. Objectives for question 1; orbital dynamics:

    i. Compare the idealised performance of existing transmissive sail proposals by transcribing their performance data into scalar performance parameters; categorise them according to performance to simplify the discussion.

    ii. Extrapolate these scalar performance parameters into approximate performance profiles (SRP versus solar incidence) by imposing a generic model for SRP sensitivity to solar incidence.

    iii. Create a flight model (FM) that realistically depicts solar sailing flight in LEO, accounting for atmospheric drag, eclipse and orbital precession.

    iv. Generate locally optimal steering laws for these sails by importing them into the FM and performing a numerical search over a range of altitudes and inclinations.

    v. Generate tangible flight characteristics for these steered sails, such as operational altitude, by importing them into the FM. Compare the sensitivity of these sails to different flight regimes when steered under either a locally optimal or simplified steering law. Compare with contemporary reflective solar sails, and include a case study that compares them with electric thrusters.

    vi. In light of the results, and by discussing the operational advantages and disadvantages, discuss the suitability and readiness of these sails for near-Earth satellite applications (including space debris mitigation) in the short to long term.

2. Objectives for question 2; optical design:

    i. Develop a ray tracing simulation to calculate the SRP and torque per unit area generated by 3D refractive objects illuminated in vacuum.

ii. Develop a numerical optimisation tool using model-free reinforcement learning to iterate upon geometries according to any user-defined fitness function.

iii. Explore the solution space for refractive surface patterns and demonstrate the optimisation method. To do this, generate numerically optimised surface patterns according to various fitness functions, under various constraints and using different optimiser configurations. Include optimisation runs for both triangular prisms and freeform optical elements; optimise for either SRP (acceleration) or corrective torque per unit area (stability). In particular, compare results with analytically optimised triangular prisms and existing *lightfoil* designs.

3. Objectives for question 3; manufacturing processes:

   i. After identifying and mapping the manufacturing processes that could be used for transmissive solar sail manufacture, trial and assess the feasibility of executing the relevant thin film fabrication processes, additive and conventional mould fabrication processes, and pattern transfer processes.

   ii. Of the processes that are infeasible to execute in-house, identify the cost of commercial alternatives. Of the processes that are feasible to execute in-house, compare the cost of executing in-house versus the cost of outsourcing. From the perspective of a small satellite developer, identify the optimal sequence of processes for transmissive sail manufacture.

   iii. Design a prototype transmissive solar sail for an academic CubeSat in preparation for an in-orbit demonstration.

Chapters 2 and 3 detail the background literature and the employed simulation methodologies, respectively. The former introduces all three areas of interest,

while the latter only covers the first two (manufacturing does not feature a simulation element). Thereafter, each of the technical chapters is dedicated to a single area of interest: chapter 4 covers orbital dynamics and seeks to answer research question one; chapter 5 covers optical design and seeks to answer research question two, chapter 6 covers manufacturing processes and seeks to answer research question three. Finally, chapter 7 comprises the conclusions and identifies future avenues for development.

1.1 Publications

1. S. M. Thompson, C. Tuck, N. Pushparaj, C. Cappelletti, Modelling and nu- merical optimisation of refractive surface patterns for transmissive solar sails, Acta Astronautica 233 (2025) doi:99–112. 10.1016/j.actaastro.2025.03.039. [39]

2. S. M. Thompson, C. Cappelletti, N. Pushparaj, Reflective and transmissive solar sails: Dynamics, flight regimes and applications, Acta Astronautica (2024). doi:10.1016/j.actaastro.2024.04.032 [40]

3. L. Cormier, T. Yousif, S. Thompson, A. Arcia Gil, N. Pushparaj, P. Blunt, C. Cappelletti, Preliminary Design of a GNSS Interference Mapping CubeSat Mission: JamSail, Aerospace 11 (11) (2024). doi:10.3390/ aerospace11110901. [41]

4. L. Cormier, T. Yousif, S. Thompson, A. Arcia, Jamsail: A CubeSat Demonstration Mission for GNSS Interference Mapping and a Refractive Solar Sail, in: 52nd IAF Student Conference, Vol. 233, International Astronautical Fed- eration, 2024, pp. 605–612. doi:10.52202/078379-0053. [42]

5. D. Robson, Y. Ferreira, H. Cope, P. Da Cás, L. Cormier, G. Lionço, S. Thompson, R. da Silva, M. Ghelfi, A. A. Gil, et al., The WormSail Cube-

Sat: An International Educational Project To Elevate Space Science And Education, in: 72nd International Astronautical Congress, 2021. [43]

1.2 Data Access Statement

All data used and produced in the making of this report is referenced in the text or by figures.

Chapter 2

Literature Review

2.1 Solar Sailing Principles

This section provides a mathematical framework by which different kinds of solar sail may be modelled, assessed and compared.

2.1.1 Solar Radiation Pressure

Photons have momentum p, which is expressed in terms of the Planck constant h and photon wavelength λ_k. In vector form $\mathbf{p}$, this is expressed by the reduced Planck constant $\hbar = \frac{h}{2\pi}$ and the wave vector $\mathbf{k}$ wherein $|\mathbf{k}| = \frac{2\pi}{\lambda_k}$ and $\hat{\mathbf{k}}$ acts perpendicular to the wavefront, i.e. perpendicular to the direction of propagation in vacuum (Eq. 2.1-2.2) [44].

$$p = \frac{h}{\lambda_k} \tag{2.1}$$

$$\mathbf{p} = \hbar\mathbf{k} = \frac{h}{\lambda_k}\hat{\mathbf{k}} \tag{2.2}$$

SRP arises because this momentum may be transferred to a particle during a photon-particle interaction. This exchange is described by the mean impulse per

interaction $\mathbf{\Delta\bar{p}}$. Typically, many millions of photons will contribute to SRP at a given instance, as described by the mean interaction rate $\bar{N}$. Divided over the illuminated area of the particle A (the region being bombarded by photons), their product yields the SRP vector $\mathbf{F_S}$ as described by Eq. 2.3. Omitting the division of A would instead yield $\mathbf{F_S}A$, representing the optical force.

$$\mathbf{F_S} = \frac{\bar{N}\mathbf{\Delta\bar{p}}}{A} \tag{2.3}$$

A more convenient approach is to express incident photons as *irradiance* (colloquially, 'illumination'): the incident radiated power upon an illuminated surface ($\mathrm{W/m^2}$). In the case of a solar sail at 1 au, this is described by the solar constant $G_{\mathrm{SC}} = 1,370\ \mathrm{W/m^2}$. Here the term G_{SC}/c is analogous to the rate of momentum transferred from incident photons at any given instant, where $c = 299,792,458$ m/s is the speed of light. For distances other than 1 au, radiation energy dilution due to the *inverse-square law* must be accounted for by dividing by the square of the magnitude of Sun vector $\mathbf{S}$ (au). Furthermore, in the case of a flat particle — such as a sail — the illuminated area will be effectively reduced according to the *cosine* of the solar incidence angle α (Eq. 2.4) [45–47]:

$$\mathbf{F_S} \propto \frac{G_{\mathrm{SC}}}{c\mathrm{S}^2}\cos(\alpha) \tag{2.4}$$

To complete the expression, the *type* of photon interaction must be known. Broadly speaking, these interaction types are *absorption*, *reflection* and *transmission*. In the case of absorption, photons are destroyed and all of their momentum is transferred linearly to the particle such that $\mathbf{F_{S_{absorb}}}$ acts opposite to the light source unit vector $\hat{\mathbf{S}}$ (Eq. 2.5) [2]. During reflection or transmission, photons are not destroyed, and a reaction force is generated that opposes their change in momentum during the interaction [46]. In the reflection case, a distinction must be drawn between *specular* and *diffuse*: in the specular case, this change in

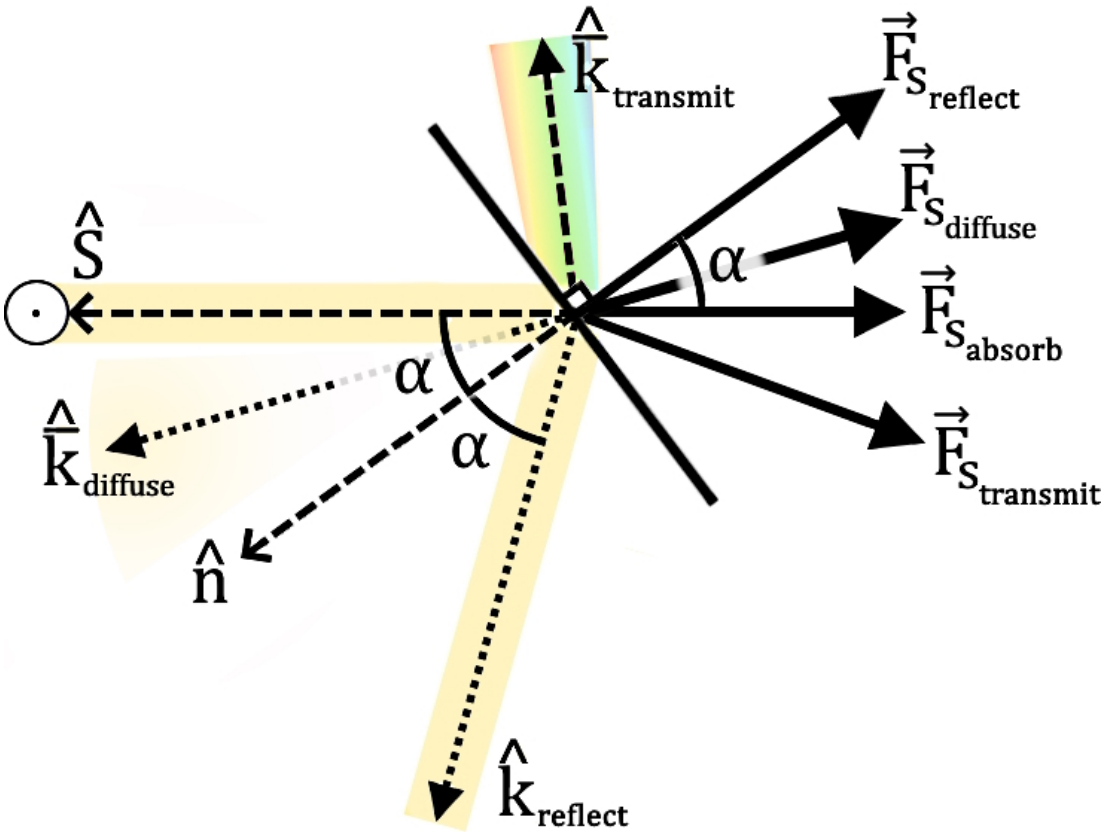

Figure 2.1: SRP Components Arising from Absorption, (Specular) Reflection, Diffuse (Reflection) and Transmission (Refraction and/or Diffraction)

momentum acts along the particle *normal* unit vector $\hat{\mathbf{n}}$ in accordance with the *law of reflection*, and so the pressure $\mathbf{F}_{\mathbf{S}_{\text{reflect}}}$ acts opposite to $\hat{\mathbf{n}}$; the component of photon momentum parallel to the sail is unchanged while the component perpendicular to it is reversed, changing by twice its original value (Eq. 2.6) [48]. In the presence of surface roughness, *diffuse* reflection occurs and the pressure $\mathbf{F}_{\mathbf{S}_{\text{diffuse}}}$ acts in both $-\hat{\mathbf{S}}$ and $-\hat{\mathbf{n}}$ in a 3:2 ratio (Eq. 2.7) [49].

Finally, for transmission — which encompasses both *refraction* and *diffraction* should they occur — the SRP $\mathbf{F}_{\mathbf{S}_{\text{transmit}}}$ is expressed in terms of the incoming and outgoing photons (Eq. 2.8) [46]. The distinction between each pressure and the corresponding direction of outgoing photons ($\hat{\mathbf{k}}_{\text{reflect}}$, $\hat{\mathbf{k}}_{\text{diffuse}}$, $\hat{\mathbf{k}}_{\text{transmit}}$) is

illustrated by Figure 2.1.

$$\mathbf{F}_{\mathbf{S}_{\text{absorb}}} = -\frac{G_{\text{SC}}}{c\text{S}^2}\cos(\alpha)\hat{\mathbf{S}} \tag{2.5}$$

$$\mathbf{F}_{\mathbf{S}_{\text{reflect}}} = -\frac{2G_{\text{SC}}}{c\text{S}^2}\cos^2(\alpha)\hat{\mathbf{n}} \tag{2.6}$$

$$\mathbf{F}_{\mathbf{S}_{\text{diffuse}}} = -\frac{G_{\text{SC}}}{c\text{S}^2}\cos(\alpha)\frac{3\hat{\mathbf{S}}+2\hat{\mathbf{n}}}{\|3\hat{\mathbf{S}}+2\hat{\mathbf{n}}\|} \tag{2.7}$$

$$\begin{aligned}\mathbf{F}_{\mathbf{S}_{\text{transmit}}} &= -\frac{G_{\text{SC}}}{c\text{S}^2}\|\hat{\bar{\mathbf{k}}}_{\text{transmit}}-\hat{\mathbf{S}}\|\cos(\alpha)\frac{\hat{\bar{\mathbf{k}}}_{\text{transmit}}-\hat{\mathbf{S}}}{\|\hat{\bar{\mathbf{k}}}_{\text{transmit}}-\hat{\mathbf{S}}\|} \\ &= -\frac{G_{\text{SC}}}{c\text{S}^2}\cos(\alpha)(\hat{\bar{\mathbf{k}}}_{\text{transmit}}-\hat{\mathbf{S}})\end{aligned} \tag{2.8}$$

The latter expression is unique in that it requires a consideration of the outgoing ray unit vector $\hat{\bar{\mathbf{k}}}_{\text{transmit}}$. This will depend on the geometry and properties of the particle, and can be derived by sequential applications of *Snell's law* (for refraction) [46] or the *grating equation* (for diffraction) [33] according to the number of boundaries passed by the incident photons.

In the event that each of these mechanisms occurs simultaneously, the total pressure is given by Eq. 2.9 wherein Φ_{a}, Φ_{r}, Φ_{d} and Φ_{t} are the proportions of power distributed to absorption, specular reflection, diffuse reflection and transmission, respectively ($\Phi_{\text{a}}+\Phi_{\text{r}}+\Phi_{\text{d}}+\Phi_{\text{t}}=1$).

$$\mathbf{F}_{\mathbf{S}} = \Phi_{\text{a}}\mathbf{F}_{\mathbf{S}_{\text{absorb}}} + \Phi_{\text{r}}\mathbf{F}_{\mathbf{S}_{\text{reflect}}} + \Phi_{\text{d}}\mathbf{F}_{\mathbf{S}_{\text{diffuse}}} + \Phi_{\text{t}}\mathbf{F}_{\mathbf{S}_{\text{transmit}}} \tag{2.9}$$

2.1.2 Reference Frames

How SRP affects a sail in orbit will depend on its attitude, position and velocity, and so the relevant reference frames must be established. Let $\mathcal{T}_B(O;\hat{\mathbf{x}}_{\mathbf{B}},\hat{\mathbf{y}}_{\mathbf{B}},\hat{\mathbf{z}}_{\mathbf{B}})$ be the 3D body-fixed reference frame originating from a solar sail centroid O, whereby the sail normal $\hat{\mathbf{n}}$ emerges from the illuminated side, is related to $\hat{\mathbf{x}}_{\mathbf{B}}$ by

$\hat{\mathbf{x}}_\mathbf{B} = -\hat{\mathbf{n}}$, and where the yz_B plane represents an idealised sail surface (blue, Fig. 2.2). For reflective sails, the direction of $\hat{\mathbf{y}}_\mathbf{B}$ is arbitrary. For transmissive sails, $\hat{\mathbf{y}}_\mathbf{B}$ is chosen to be the direction of the component of SRP that is tangential to the sail (see section 2.1.5). In both cases, $\hat{\mathbf{z}}_\mathbf{B}$ is defined by mutual orthogonality with $\hat{\mathbf{x}}_\mathbf{B}$ and $\hat{\mathbf{y}}_\mathbf{B}$ forming a right-handed frame. Additionally, let $\hat{\mathbf{S}}$ represent the Sun

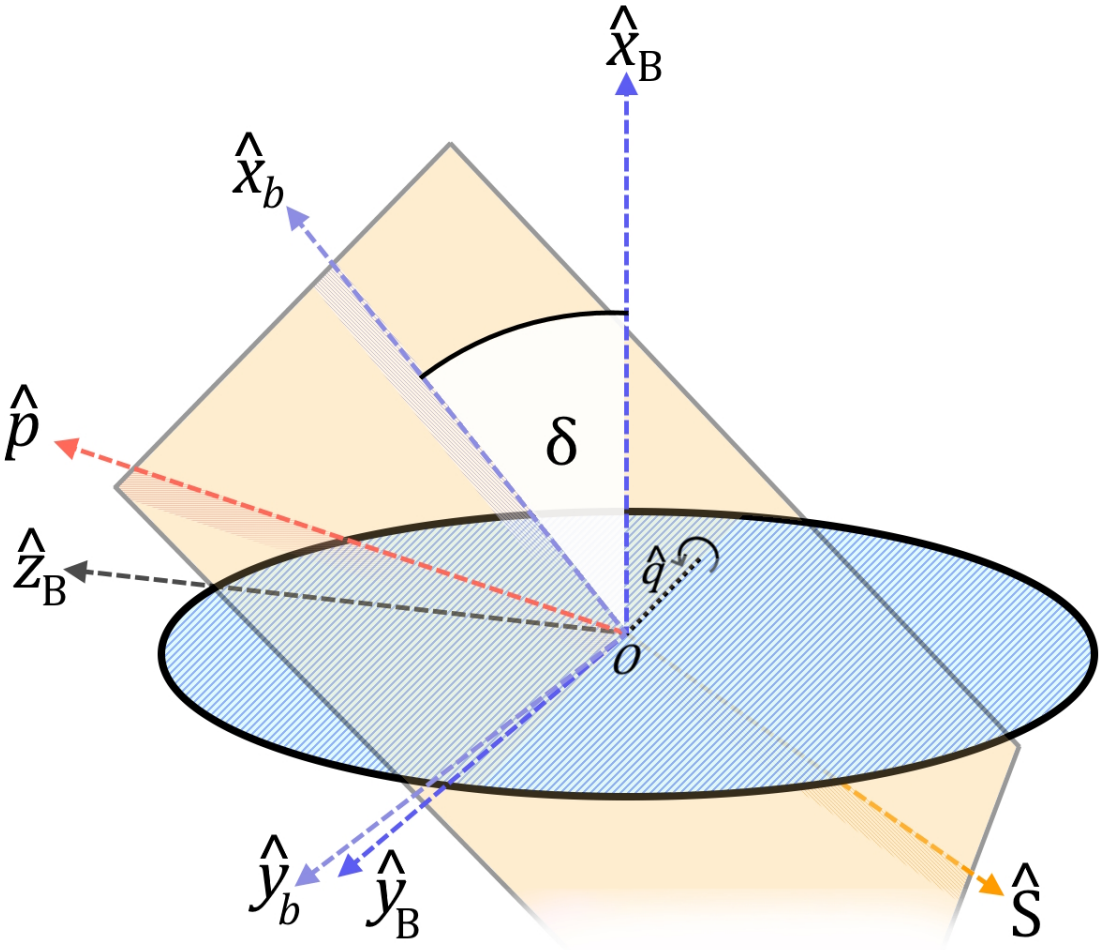

Figure 2.2: 3D Body-fixed Reference Frames Rotated onto the Solar-primer $\mathbf{S\mathfrak{p}}$

unit vector, the direction of the Sun with respect to the satellite position, and $\hat{\mathfrak{p}}$ the primer unit vector, the optimal direction of impulse for a given manoeuvre; in general, $\hat{\mathfrak{p}}$ describes the optimal direction of impulse for a manoeuvre to minimise transit times [50, 51], or to maximise the rate change of a certain orbital parameter [13, 45], depending on the steering law used (see section 2.1.3). During an ideal manoeuvre, $\hat{\mathbf{x}}_\mathbf{B}$ and $\hat{\mathbf{y}}_\mathbf{B}$ will always remain within a plane bound by $\hat{\mathbf{S}}$ and $\hat{\mathfrak{p}}$. It is therefore convenient to confine the analysis to this 2D Solar-primer plane $\mathbf{S\mathfrak{p}}$ (orange, Fig. 2.2), and define the 2D body-fixed reference frame $\mathcal{T}_b(O; \hat{\mathbf{x}}_\mathbf{b}, \hat{\mathbf{y}}_\mathbf{b})$ and

the 2D primer or 'manoeuvring' reference frame $\mathcal{T}_{\mathfrak{p}}(O; \hat{\mathbf{x}}_{\mathfrak{p}}, \hat{\mathbf{y}}_{\mathfrak{p}})$ - for which $\hat{\mathfrak{p}} = \hat{\mathbf{x}}_{\mathfrak{p}}$. For the non-ideal case wherein the attitude control fails to confine $\hat{\mathbf{n}} = -\hat{\mathbf{x}}_{\mathbf{B}}$ and $\hat{\mathbf{y}}_{\mathbf{B}}$ within the solar-primer, forces in $\mathcal{T}_B$ can be projected onto the solar-primer for use within $\mathcal{T}_b$ and the out-of-plane components can be neglected. These two reference frames can be transformed between via a rotation by angle δ about $\hat{\mathbf{q}} = \hat{\mathbf{x}}_{\mathbf{B}} \times \hat{\mathfrak{p}}$ (Eq. 2.10) as depicted by Figure 2.2.

$$\delta = 90 - \arccos\left(\frac{\hat{\mathbf{x}}_b \cdot (\hat{\mathfrak{p}} \times \hat{\mathbf{S}})}{|\hat{\mathbf{x}}_b||(\hat{\mathfrak{p}} \times \hat{\mathbf{S}})|}\right) \tag{2.10}$$

$$\begin{bmatrix} \hat{\mathbf{x}}_b \\ \\ \hat{\mathbf{y}}_b \end{bmatrix} = R_{\hat{q}}(\delta) \begin{bmatrix} \hat{\mathbf{x}}_B \\ \\ \hat{\mathbf{y}}_B \end{bmatrix}$$

Three more angles can be drawn on $\mathbf{S}\mathfrak{p}$ in addition to $\alpha = \angle\, \hat{\mathbf{S}}\hat{\mathbf{n}}$ as depicted in Fig. 2.3: the angle *sail deviation* $\beta = \angle\, \hat{\mathbf{x}}_{\mathbf{b}}\hat{\mathbf{x}}_{\mathfrak{p}}$, which relates the satellite attitude in $\mathcal{T}_b$ to the desired direction of impulse $\hat{\mathfrak{p}}$ in $\mathcal{T}_{\mathfrak{p}}$; *effective sail deviation* $\beta^* = \angle\, \mathbf{F}_{\mathbf{S}}\hat{\mathbf{x}}_{\mathfrak{p}}$, which determines the direction of SRP relative to $\hat{\mathfrak{p}}$; and the *solar misalignment* $\gamma = \angle\, \hat{\mathbf{x}}_{\mathfrak{p}}\hat{\mathbf{S}}$, which describes how suitable the position of the Sun is for the desired manoeuvre at a given instant. These angles do not exist independently of one another, and it may be observed from Fig. 2.3 that these comply with Eq. 2.11. In the $\gamma = 90°$ scenario, Eq. 2.11 simplifies to Eq. 2.12. This is a notable special case that is archetypal of solar sailing flight, and is therefore frequently referenced by this thesis:

$$\forall \gamma \in \mathbb{R} \rightarrow \qquad \alpha + \beta + \gamma = 180° \tag{2.11}$$

$$\gamma = 90° \rightarrow \qquad \beta = 90 - \alpha \rightarrow \qquad \cos(\beta) = \sin(\alpha) \tag{2.12}$$

However, these expressions do not feature β^* at all. As described in Eq. 2.13, the relationship between this angle and the other three will vary depending on the

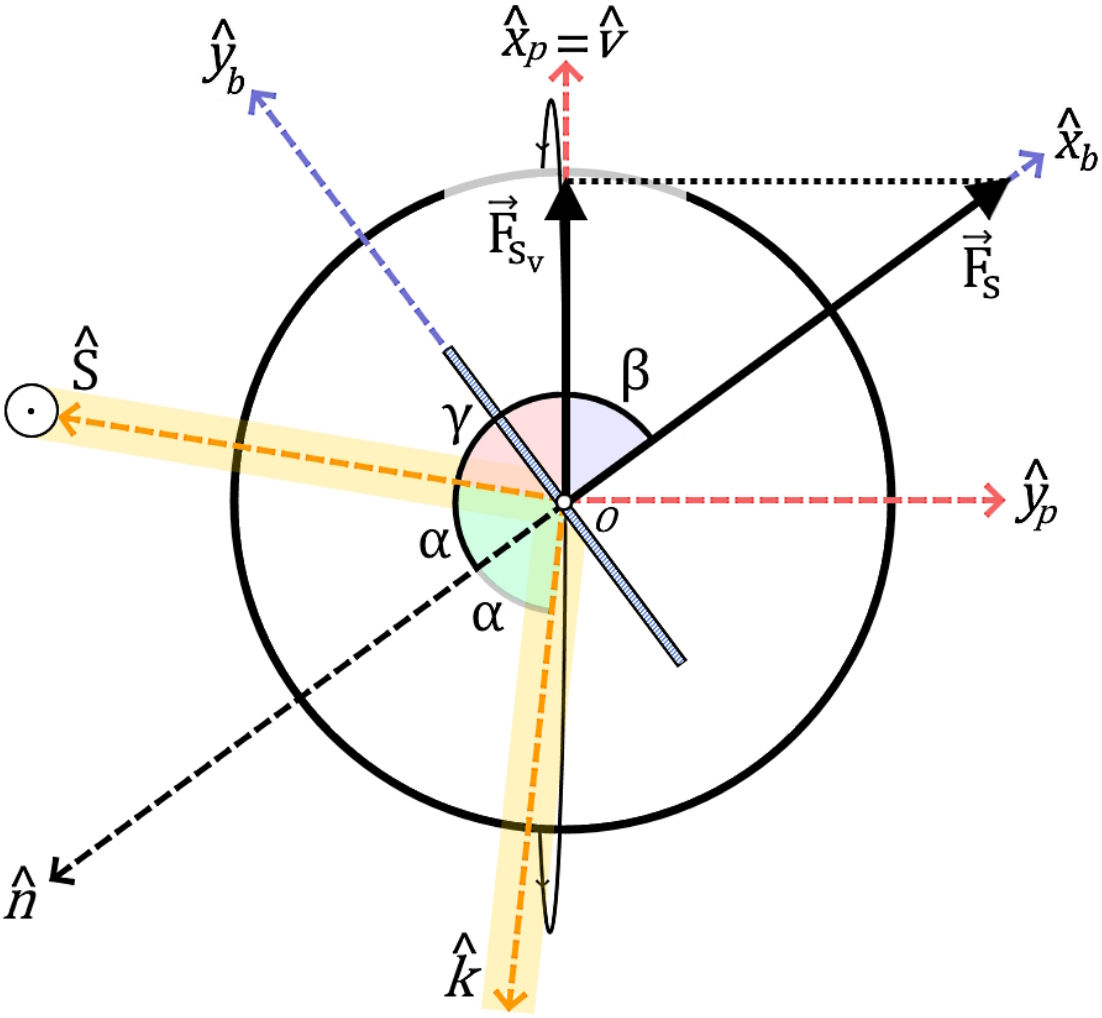

Figure 2.3: 2D Body-fixed Reference Frame in the Solar-primer Plane, applied to a (Specular) Reflective Solar Sail wherein $\mathbf{F_{S_{x_p}}} = \mathbf{F_{S_v}}$ (Orbit-raising Manoeuvre)

mechanism of SRP generation, making it unique. These relationships are derived from Eq. 2.5-2.8 (see section 2.1.1) and Eq. 2.11:

$$\beta^* = \begin{cases} 180^\circ - \gamma & & = f(\gamma) & \text{if absorptive} \\ \beta & = 180^\circ - \gamma - \alpha & = f(\alpha, \gamma) & \text{if reflective} \\ \beta + f(\alpha) & = 180^\circ - \gamma - \alpha + f(\alpha) & = f(\alpha, \gamma) & \text{if transmissive} \end{cases} \tag{2.13}$$

Some noteworthy features of β^* are highlighted by the above. Firstly, in each case, β^* may be expressed as a function of α, γ or both. In the absorption case, β^* is independent of sail attitude. In the reflective case, β^* has a linear

relationship with sail attitude ($\beta^* = \beta$). In the transmissive case, an undefined function $f(\alpha)$ influences β^* that is specific to the sail, due to the material and pattern dependence of transmissive SRP that is described in section 2.1.1, and that is explored extensively in chapter 5.

With the relationship between $\mathcal{T}_b$, $\mathcal{T}_\mathfrak{p}$ and $\hat{\mathbf{S}}$ established, it is convenient to split the SRP vector $\mathbf{F_S}$ into orthogonal components. In the body-fixed frame $\mathcal{T}_b$, these components are the sail normal $\mathbf{F}_{\mathbf{S}_{\mathbf{x_b}}}$ and sail tangential $\mathbf{F}_{\mathbf{S}_{\mathbf{y_b}}}$; in the manoeuvring frame $\mathcal{T}_\mathfrak{p}$, these components are transverse $\mathbf{F}_{\mathbf{S}_{\mathbf{x_\mathfrak{p}}}}$ and longitudinal $\mathbf{F}_{\mathbf{S}_{\mathbf{y_\mathfrak{p}}}}$.

Typically, raw sail performance data expresses $\mathbf{F_S}$ components in terms of $\mathcal{T}_b$. However, it is more convenient to express $\mathbf{F_S}$ in terms of $\mathcal{T}_\mathfrak{p}$ because this pertains directly to performance within a manoeuvre; in this frame, transverse $\mathbf{F}_{\mathbf{S}_{\mathbf{x_\mathfrak{p}}}}$ and longitudinal $\mathbf{F}_{\mathbf{S}_{\mathbf{y_\mathfrak{p}}}}$ may be called the 'useful' and 'wasteful' components. The transverse component $\mathbf{F}_{\mathbf{S}_{\mathbf{x_\mathfrak{p}}}}$ is useful by definition, because it acts in the direction of the primer $\hat{\mathbf{x}}_\mathfrak{p}$, which is the desired direction of impulse. It follows that the longitudinal component $\mathbf{F}_{\mathbf{S}_{\mathbf{y_\mathfrak{p}}}}$ is wasteful because it is perpendicular to the desired direction of impulse. This wasteful component can be neglected, while the useful component may be derived from a raw SRP vector or $\mathcal{T}_b$ performance data using the expressions in Eq. 2.14:

$$\mathbf{F}_{\mathbf{S}_{\mathbf{x_\mathfrak{p}}}} = F_S \cos(\beta^*)\hat{\mathbf{x}}_\mathbf{p} = F_S \cos[\arccos(\frac{F_{S_{y_b}}}{F_S}) + (90^\circ - \gamma)]\hat{\mathbf{x}}_\mathbf{p} \tag{2.14}$$

Eq. 2.5-2.8 can be transcribed to the $\mathcal{T}_\mathfrak{p}$ reference frame to express $\mathbf{F}_{\mathbf{S}_{\mathbf{x_\mathfrak{p}}}}$ for each mechanism (Eq. 2.15-2.18).

$$\mathbf{F}_{\mathbf{S}_{\mathbf{x_p}\text{absorb}}} = \frac{G_{\text{SC}}}{c\text{S}^2}\cos(\alpha)\cos(\beta^*)\hat{\mathbf{x}}_\mathbf{p} \tag{2.15}$$

$$\mathbf{F}_{\mathbf{S}_{\mathbf{x_p}\text{reflect}}} = \frac{2G_{\text{SC}}}{c\text{S}^2}\cos^2(\alpha)\cos(\beta^*)\hat{\mathbf{x}}_\mathbf{p} \tag{2.16}$$

$$\mathbf{F}_{\mathbf{S}_{\mathbf{x_p}\text{diffuse}}} = \frac{G_{\text{SC}}}{5c\text{S}^2}\cos(\alpha)[3\cos(\alpha+\beta^*) + 2\cos(\beta^*)]\hat{\mathbf{x}}_\mathbf{p} \tag{2.17}$$

$$\mathbf{F}_{\mathbf{S}_{\mathbf{x_p}\text{transmit}}} = \frac{G_{\text{SC}}}{c\text{S}^2}\|\hat{\bar{\mathbf{k}}}_{\text{transmit}} - \hat{\mathbf{S}}\|\cos(\alpha)\cos(\beta^*)\hat{\mathbf{x}}_\mathbf{p} \tag{2.18}$$

Wherein specular and diffuse reflection share the β^* definition for reflection of Eq. 2.13. The question then becomes the nature of $\hat{\mathbf{x}}_\mathbf{p}$, which depends upon the manoeuvre being carried out and the steering law being used. Because orbit-raising manoeuvres are both simple and ubiquitous, the analysis will focus on these hereafter.

2.1.3 Steering Law and Optimal Attitude

A steering law dictates how a satellite will carry out a manoeuvre; an optimal steering law aims to carry out the manoeuvre in the shortest time. One such law is the *globally optimal* steering law, which is rigorous and suitable for complex manoeuvres [34, 52]. However, it also poses two-point boundary problems that require computationally expensive solving techniques, such as the shooting method [53]. The Q-law algorithm is an advanced variant of *locally optimal steering law* that has similar capabilities, is suitable for non-impulsive applications, and that can be solved analytically [54, 55]. However, by focusing on the orbit-raising of a solar sail within an approximately circular orbit, a simpler locally optimal steering law will suffice. This law only aims to maximise the rate of change of a specific orbital parameter, but has nonetheless been shown to yield near-optimal transit times for many manoeuvres [13]. Applied to an orbit-raising manoeuvre, this law will aim to maximise the *rate change* of the orbital *semi-major axis* $\dot{\mathtt{a}}$ (half the diameter of an orbital ellipse). This has been shown to be equivalent to maximising the *rate change* of *orbital specific energy* $\dot{\epsilon}$ (Eq. 2.19-2.20) [13, 45], which may be used to form the locally optimal steering law axiom (Eq. 2.21). Through application of Eq. 2.13, α is presented as the sole control variable, making these expressions convenient for a numerical search (note that γ and altitude

h are not control variables):

$$\dot{\epsilon} = v\dot{v} = \frac{v}{m}A[F_{S_v} - D] \tag{2.19}$$

$$\max_\alpha \dot{\epsilon}(\alpha, \gamma, h) = \frac{v(h)}{m}A[F_{S_v}(\alpha_{\text{opt}}, \gamma) - D(\alpha_{\text{opt}}, \gamma, h)] \tag{2.20}$$

$$\text{argmax}_\alpha \dot{\epsilon}(\alpha, \gamma, h) \equiv \text{argmax}_\alpha \dot{\mathsf{a}}(\alpha, \gamma, h) = \alpha_{\text{opt}} \tag{2.21}$$

Where v is velocity, $\dot{v}$ is acceleration, D is the atmospheric drag pressure and α_{opt} is the solar incidence that maximises $\dot{\epsilon}$; h and γ can be said to compose an instance of a flight regime by confining the analysis to circular, 1 au orbits. By adjusting α, the locally optimal steering law will seek to maximise the component of net force that is acting in $\hat{\mathbf{v}}$. It can therefore be said that the primer unit vector is $\hat{\mathbf{x}}_{\mathfrak{p}} = \hat{\mathbf{v}}$ for this manoeuvre.

Eq. 2.19-2.21 presents α as the sole control variable through application of Eq. 2.13. These expressions are convenient, but they occlude the nature of the problem. A more natural depiction is provided by Eq. 2.22, which highlights that the problem is actually a balancing of the three *controllable* solar sailing angles α, β and β^*:

$$\max_\alpha \dot{\epsilon}(\alpha, \beta, \beta^*, h) = \frac{v(h)}{m}A[F_{S_v}(\alpha_{\text{opt}}, \beta^*) - D(\beta, h)] \tag{2.22}$$

This form is useful for understanding the nature of α_{opt} by highlighting the individual contributions of each control angle to $\dot{\epsilon}$. For example, α dictates sail illumination and is an argument of F_S, while β^* defines the alignment of $\mathbf{F_S}$ with $\hat{\mathbf{v}}$ and is an argument of the *transverse SRP ratio* $\lambda = F_{S_v}/F_S$. When atmospheric drag is present, β represents the angle relative to the incoming airflow and $|\beta| - 90°$ determines the magnitude of drag pressure D (drag is minimised when $\beta = \pm 90°$). At a given instance wherein h is considered constant, these contributions are made

clearer yet by the form Eq. 2.23:

$$\dot{\epsilon} \propto F_{S_v}(\alpha, \beta^*) - D(\beta) = F_S(\alpha)\lambda(\beta^*) - D(\beta) \tag{2.23}$$

2.1.4 Flight Regimes

$\{h, \gamma\}$ can only describe a single instance in time. During orbit-raising by solar sail, an orbit that begins as circular will generally remain circular because such sails are non-impulsive. It is therefore convenient to confine the subsequent analyses to circular orbits for our purposes such that h can be said to be constant over the duration of a single orbit. On the other hand, γ generally cycles between some minimum value and some maximum value as a function of true anomaly ν (orbital angular position), albeit always with a mean of $\bar{\gamma} = 90°$ (see section 3.1.2). The variable γ is said to belong to the continuous set S_γ (Eq. 2.24). Additionally, we introduce the Sun-*orbit* angle Γ as the angle between the *specific relative angular momentum unit vector* $\hat{\mathbf{h}}$ (the orbital plane *normal vector*) and $\hat{\mathbf{S}}$. This is the sole argument of S_γ:

$$\gamma(\nu, \Gamma) \qquad \in \mathrm{S}_\gamma(\Gamma) \tag{2.24}$$

$$\mathrm{S}_\gamma(\Gamma) \qquad = [\gamma^{\mathrm{MIN}}, \gamma^{\mathrm{MAX}}] = \{\gamma(\nu, \Gamma) \in \mathbb{R} | \gamma^{\mathrm{MIN}} \leq \gamma \leq \gamma^{\mathrm{MAX}}\} \tag{2.25}$$

$\{h, S_\gamma\}$ can be said to compose a flight regime, and naturally, some flight regimes are more optimal than others. Consider two reflective sails in orbits of arbitrary h wherein one is an ecliptic plane LEO of $\Gamma = 90°$ (Eq. 2.26), and the other is a 'sunrise-sunset' polar LEO of $\Gamma = 0°$ (Eq. 2.27):

$$\mathrm{S}_\gamma(90°) \qquad = \{\gamma(\nu, 90°) \in \mathbb{R} | \qquad 0° \leq \gamma \leq 180°\} \tag{2.26}$$

$$\mathrm{S}_\gamma(0°) \qquad = \{\gamma(\nu, 0°) \in \mathbb{R} | \qquad 90° \leq \gamma \leq 90°\} \tag{2.27}$$

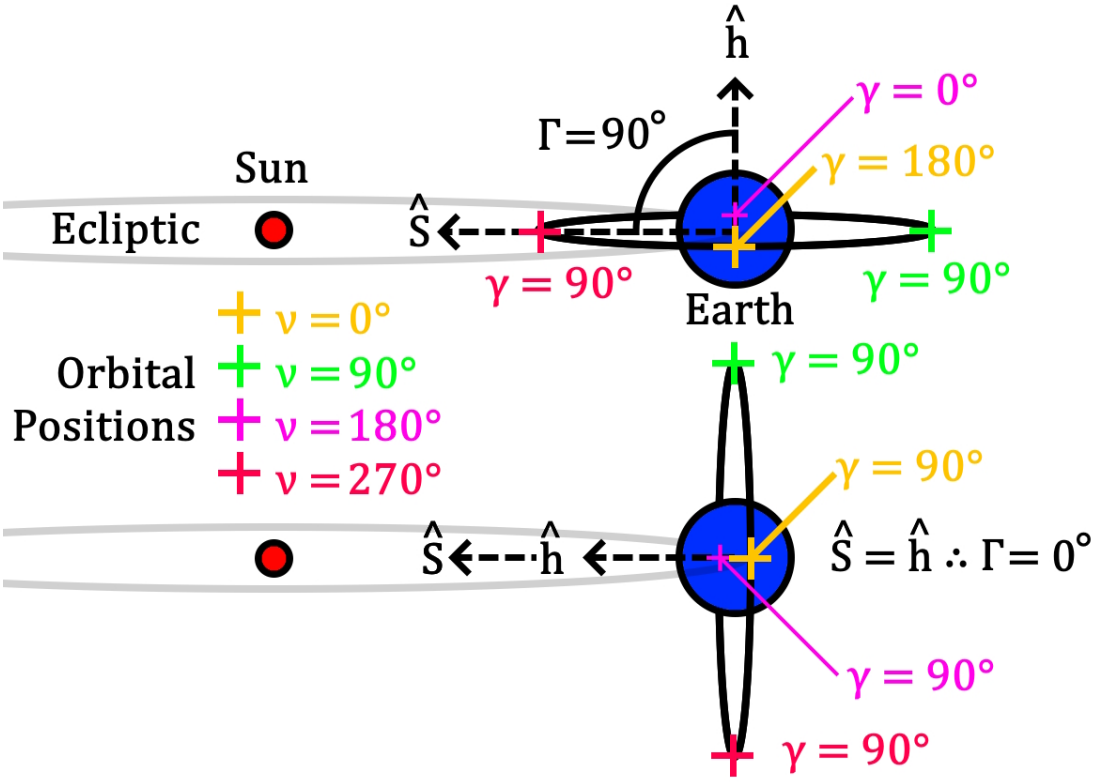

Figure 2.4: $\Gamma = 90°$ and $\Gamma = 0°$ Orbits Split into Four Orbital Position Samples

A low-resolution assessment of these flight regimes can be made by making S_γ discrete and attributing it a sample size of four (each sample representing a quarter of an orbit). In this and all subsequent cases, we define ν according to Fig. 2.4. That is to say that orbits progress anticlockwise about the axis $\hat{\mathbf{h}}$ and $\nu = 0°$ exists at the ascending node of the sailcraft relative to the ecliptic plane (i.e. the point of intersection between the satellite-Earth orbital plane and the ecliptic plane for which the sailcraft is travelling above the ecliptic). When $\Gamma = 90°$, this ascending node is undefined, so instead $\nu = 0°$ occurs at a 90° anticlockwise offset from 'aphelion' (closest point to the Sun). We assign corresponding sets for position $\nu \in \mathrm{S}_\nu$ and optimal solar incidence $\alpha_{\mathrm{opt}} \in \mathrm{S}_{\alpha_{\mathrm{opt}}}$. Note that atmospheric drag is neglected for this example (Eq. 2.29):

$$S_\nu(90°) = \{0, 90, 180, 270\} \qquad S_\nu(0°) = \{0, 90, 180, 270\} \tag{2.28}$$

$$S_\gamma(90°) = \{180, 90, 0, 90\} \qquad S_\gamma(0°) = \{90, 90, 90, 90\} \tag{2.29}$$

$$S_{\alpha_{\text{opt}}}(90°) = \{0, \text{NaN}, 90, 35.26\} \qquad S_{\alpha_{\text{opt}}}(0°) = \{35.26, 35.26, 35.26, 35.26\} \tag{2.30}$$

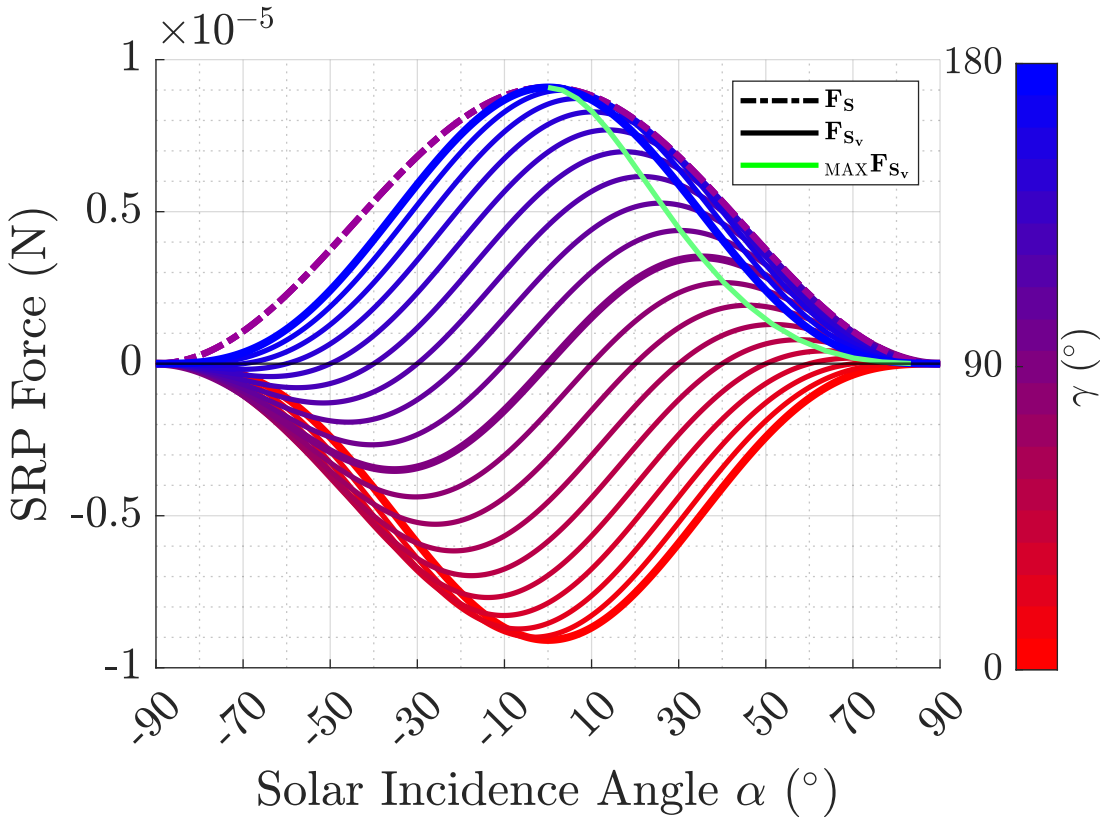

Figure 2.5: 1×1 m^2 Reflective Sail Optical Force-α Profiles for $0° \leq \gamma \leq 180°$ ($\gamma = 0,\ 90,\ 180°$ are Thickened)

Each subsequent set follows from the previous one. How γ follows from ν over the duration of an orbit is simple to derive with reference to Fig. 2.4, and is explored further in section 3.1.2. Conversely, how α_{opt} follows from γ is less intuitive. In each instance, α_{opt} is the solar incidence that achieves maximum F_{S_v}. For a reflective sail in a perfect vacuum, this means maximising the product of $\cos(\alpha)^2$ — the modifier of F_S — and $\cos(\beta^*)$ — the modifier of ratio $\lambda = F_{S_v}/F_S$ (see Eq.

2.15, 2.23). To demonstrate the relationship between SRP, α and α_{opt} for different values of γ, SRP profiles for $0° < \gamma < 180°$ (at 10° intervals) are presented in Fig. 2.5. This figure demonstrates what α_{opt} actually represents at a given γ; the green curve representing $F_{S_v}^{\max}$ traces along all of the possible α_{opt} values for a reflective sail. For example, it can be seen that the $\gamma = 90°$ curve intersects with the $F_{S_v}^{\max}$ curve at $\alpha = 35.26°$, and therefore $\alpha_{\text{opt}} = 35.26°$ here. Of course, an α_{opt} may not truly follow these curves if the sail is orbiting close to a planet due to eclipse and atmospheric drag. For example, the $\Gamma = 90°$ case encounters an eclipse at $\nu = 90°$, and because drag is assumed to be negligible for this example it is true that $\alpha_{\text{opt}} = \text{NaN}$; here, both SRP and drag are zero regardless of attitude.

Another notable characteristic of Fig. 2.5 is that the dotted F_S curves are all superimposed, and so appear as a single curve. This is because F_S is sensitive to α but not γ; the F_S profile for each γ is the same. Conversely, F_{S_v} is sensitive to both. Three F_{S_v} curves in particular are highlighted: the curves for $\gamma = 0°,\ 180°$ represent the 'edge of the envelope' wherein the Sun is directly *prograde* or *retrograde* to the desired direction of impulse, respectively, while $\gamma = 90°$ represents the nominal case wherein the Sun is perpendicular to the desired direction of impulse. Red profiles leading up to $\gamma = 0°$ depict the Sun becoming increasingly prograde, which can be seen to become increasingly problematic for generating SRP. In the perfectly prograde case, it can be seen that positive SRP is impossible to achieve for any sail attitude. Conversely, blue profiles leading up to $\gamma = 180°$ depict the Sun becoming increasingly retrograde, and produce increasingly favourable SRP until the Sun is directly retrograde.

Finally, the efficiency set $S_{\eta_{Sv}}$ can be defined, wherein η_{S_v} represents the F_{S_v} as a percentage of the maximum possible F_{S_v} (at 1 au) through specular reflection. These are defined according to Eq. 2.31-2.32:

$$\eta_{S_v} = 100 \times \frac{F_{S_v}}{(\frac{2G_{\mathrm{SC}}}{c})} \tag{2.31}$$

$$S_{\eta_{Sv}}(90°) = \{100, 0, 0, 38.51\} \quad S_{\eta_{Sv}}(0°) = \{38.51, 38.51, 38.51, 38.51\} \tag{2.32}$$

Returning to a comparison of these two orbits: in the first, $\Gamma = 90°$ case, the flight regime is defined by an oscillation between optimal and sub-optimal conditions due to the shifting γ; at $\nu = 0°$ the Sun is retrograde to sail motion ($\gamma = 180°$) and both trigonometric terms are maximised perfectly by $\alpha_{\mathrm{opt}} = 0°$, which is the maximum SRP configuration depicted by Fig. 2.6. The flight conditions at this attitude yield 100% of the theoretical maximum F_{S_v}. At $\nu = 180°$ (the other side of the orbit), the Sun is prograde to motion ($\gamma = 0°$) and $\alpha_{\mathrm{opt}} = 90°$; this is the zero SRP configuration, and it is optimal because any SRP generated would cause the sail to decelerate. Conversely, at the two intermediary points ($\nu = 90, 270°$) the Sun is perpendicular to motion ($\gamma = 90, 90°$). The optimal solution is arbitrary at the first intermediary point because the sail is eclipsed by the Earth (see section 3.1.3) and drag is neglected. At the second intermediary point, the product of the trigonometric terms is maximised by $\alpha_{\mathrm{opt}} = 35.26°$, which is a *tacked* configuration.

In the second, $\Gamma = 0°$ case, the flight regime is defined by a constant $\gamma = 90°$, yielding the same tacked $\alpha_{\mathrm{opt}} = 35.26°$ configuration as before, but for the entire orbit. This places the average efficiency of the orbit (for reflective solar sailing) at 38.51%. Conversely, averaged across all four samples, the efficiency of the first case flight regime is 34.63%. Clearly, the constant γ second case is more optimal. Incidentally, this unchanging $\gamma = 90°$ flight regime is qualitatively identical to that of a circular heliocentric orbit.

If atmospheric drag is not neglected, the efficiency discrepancy will grow as h is reduced, favouring the second case. This is in part because, in the first case, the

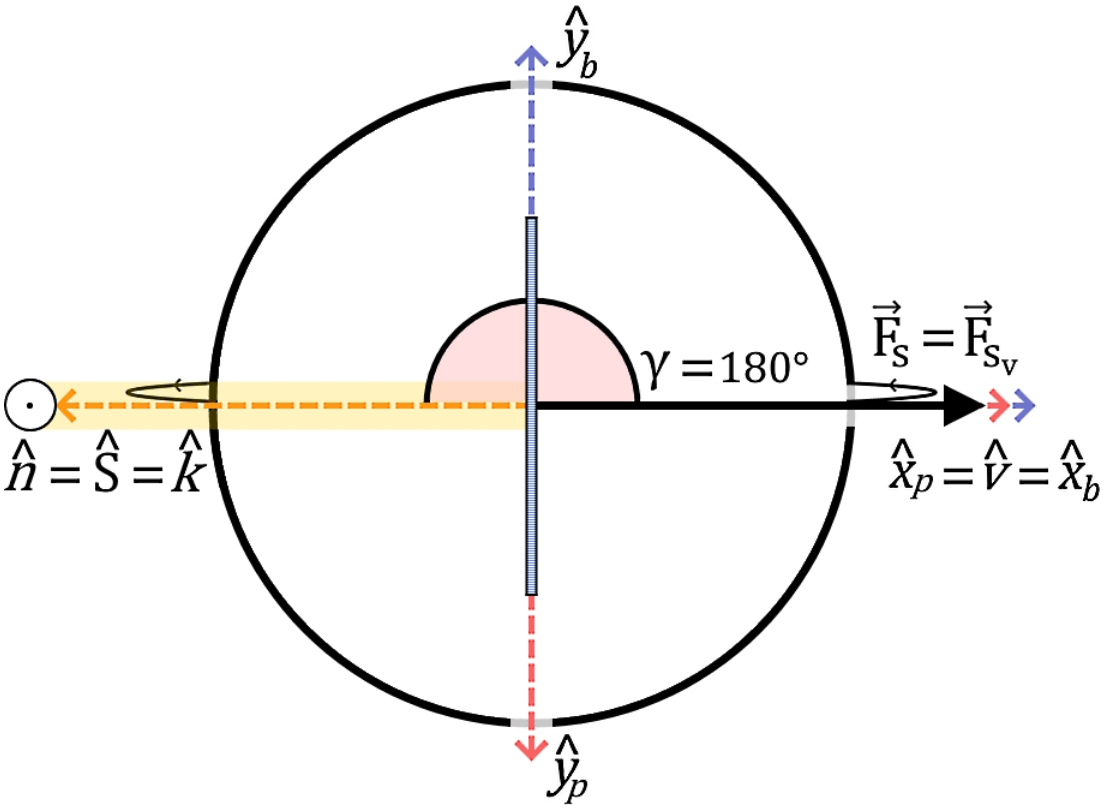

Figure 2.6: Ecliptic LEO (at $\gamma = 0°$) for a Reflective Solar Sail with Atmospheric Drag Neglected

attitude that a sail must adopt to maximise SRP when the Sun is retrograde is also the maximum drag configuration; there are also segments of the orbit (e.g. the prograde Sun region) for which acceleration is impossible. This is exacerbated by eclipse, which in the first case, becomes more prominent as h is reduced. Conversely, the second case maintains a lower drag, tacked configuration throughout, can accelerate constantly, and experiences zero eclipse (at least initially — see section 3.1.3).

2.1.5 Introduction to Transmissive Solar Sails

Transmissive solar sails differ from reflective ones in terms of their mechanisms (see sections 2.1.1) for generating SRP and their behaviour within the various flight regimes. Transmissive solar sails transmit and redirect incident sunlight through a transparent, often highly refractive membrane to generate SRP. This

SRP can be generated even at $\alpha = 0°$ by harnessing surface patterns that act as waveguides or gratings [29–33, 35–37, 56–59]. Whether the primary SRP mechanism of the membrane is refraction or diffraction depends on the dominant spectrum of light that the membrane is designed to transmit (usually the visible spectrum), and the scale of the pattern that this light is transmitted through. If the smallest element of these patterns is an order of magnitude greater than the longest wavelength of said spectrum, refraction will be the primary mechanism; if the patterns are similarly sized to the wavelengths of said spectrum, diffraction must be considered [60]. Crucially, the SRP vector $\mathbf{F_S}$ is not aligned with $\hat{\mathbf{x}}_\mathbf{b}$ (which aligns with the sail normal $\hat{\mathbf{n}}$) for a transmissive sail. This is in contrast to specular reflective sails, for which $\mathbf{F_S} = \mathbf{F_{S_{x_b}}}$ if the sail is idealised. Instead, it will have a significant $\mathbf{F_{S_{y_b}}}$ component that acts tangential to the sail [29, 31, 32, 35]. As depicted by Fig. 2.7, the tangential component $\mathbf{F_{S_{y_b}}}$ will tend to align with the useful transverse component $\mathbf{F_{S_{x_p}}}$ while in a nearly Sun-pointing, $\alpha \approx 0°$ attitude under near-optimal, $\gamma \approx 90°$ conditions (for $\gamma = 90°$: $\mathbf{F_{S_{x_p}}} = \mathbf{F_{S_{y_b}}}$).

This means that a transmissive sail may not need to be tacked at all, allowing it to maximise illumination (maximise $\mathbf{F_S}$) without sacrificing β^* and the ratio $\lambda = F_{S_v}/F_S$. Furthermore, within an optimal orbit, this Sun-pointing attitude will coincide with the minimum drag configuration ($\beta = 90°$). In short, the priorities discussed in section 2.1.3 do not necessarily conflict for transmissive sails as they do for reflective ones; to a degree that is afforded by the flight regime, they can often be satisfied simultaneously. The reduced drag under optimal conditions also enables transmissive sails to operate at lower altitudes [29], which is explored in greater detail in section 4.2.2.

Although the magnitude of $\mathbf{F_{S_{y_b}}}$ changes with α, its body-fixed direction is constant. As depicted by Fig. 2.8, a transmissive sail may alternate between an orbit-raising, orbit-lowering or out-of-plane manoeuvre simply by changing its

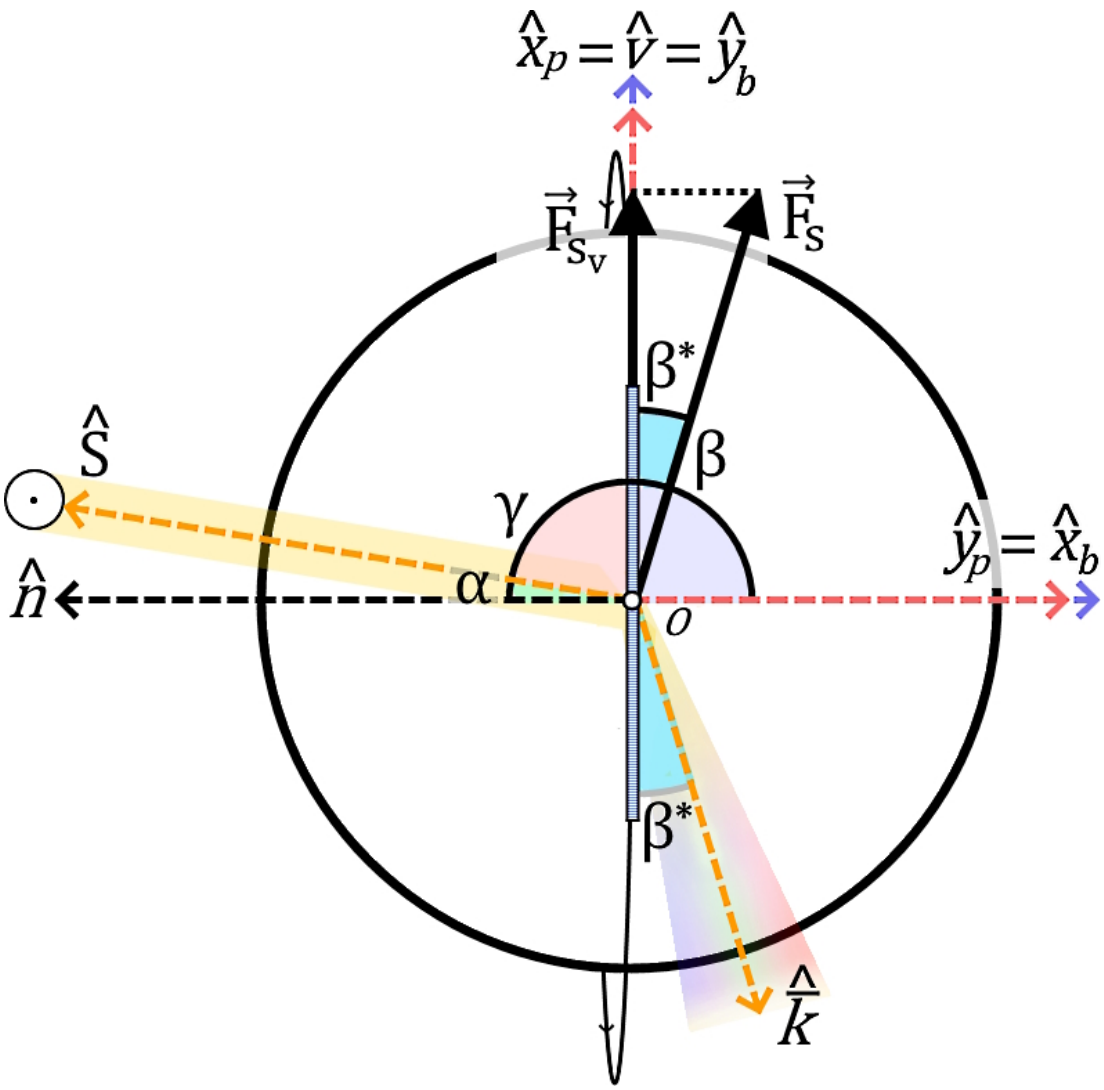

Figure 2.7: Transmissive Sail at $\gamma \approx 90°$

roll angle ϕ while remaining in a Sun-pointing, $\alpha = 0°$ attitude. Some transmissive sail designs have also been proposed to be capable of providing both passive or active Sun-pointing stability [29, 30, 36–38]. Under optimal conditions, these phenomena may allow for greatly simplified steering when compared with reflective sails. For certain interplanetary transits, the cumulative effects of improved F_{S_v} and simplified steering have been suggested to yield reductions in transit times of 30-44% [34].

However, the behaviour of different transmissive sails will also vary significantly according to their pattern and mechanism for generating SRP. It is therefore difficult to make assertions about transmissive sails as a breed, without neglecting

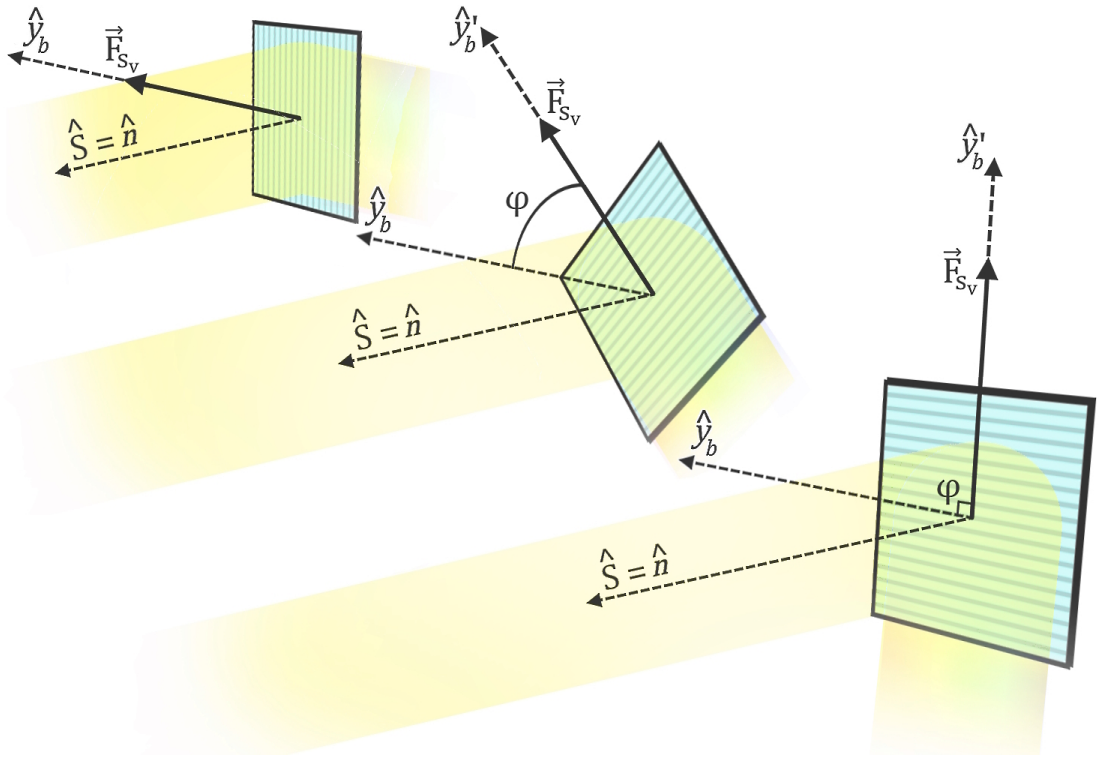

Figure 2.8: Transmissive Sail with $\mathbf{F_{S_v}}$ Locked to $\hat{\mathbf{y}}_\mathbf{b}$ (Idealised so that $\beta^* = 0°$)

the many exceptions. To this end, transmissive sail designs from various proposals are evaluated during this report.

2.2 Solar Sail Design and Manufacture

2.2.1 Evolution of Designs

The state-of-the-art for functional, validated solar sails was described in chapter 1 with notable brevity, as the sails that have been tested in space are few and quite alike. Conversely, the literature contains myriad designs that have not been tested in space (the reader is directed towards source [2] for a deeper dive into conventional solar sail design). The evolution of reflective sails remains pertinent to a discussion of transmissive sails because the holistic design of one kind of sail will generally be adaptable — if not outright applicable — to a sail of the other kind. As detailed at the end of this chapter, a transmissive sail may differently

emphasise one or more solar sail design stressors, but rarely is a reflective sail structure, mechanism or configuration outright incompatible with transmissive sails (one exception is detailed by section 2.2.8). The only compulsory difference between the two lies in the design of the sail membrane itself.

Section 2.1.5 details how these membranes differ by function. Naturally, they must also differ by design in order to reconcile this difference in function. In simple terms, a reflective sail membrane is a nominally two-layer film comprising a reflective metal layer and a thicker, structural polymer layer. The majority of flown solar sail membranes abide by this description [5–7]. Three-layer reflective films also feature a high emissivity metal layer on the underside to promote heat dissipation, with one flown example [8]. Conversely, monolayer reflective films have been proposed that would eschew the structural polymer layer [61, 62]. This is not typically done because the density of metallic films is disproportionately high, such that a mechanically stable metal film is far heavier than a metallised polymer one. Furthermore, thin metallic films are prone to brittle fracture, which is a problem for stowing. Proposed solutions to these problems include tensile meshes [61] or on-orbit manufacturing to circumvent the need for stowing [62].

On the other hand, proposed transmissive sail membranes typically comprise a monolayer polymer, wherein the patterned surface is functional and the unpatterned core is structural. As mentioned in section 2.1.5, the scale of the pattern determines whether the key mechanism for generating SRP is refraction or diffraction. This patterned monolayer polymer is the simplest and most common form that transmissive membrane proposals take, but there are exceptions. Some proposals feature metasails [31–33, 35] with membranes composed not of polymer, but of metamaterials of varying readiness. Others do not feature a static pattern to speak of [36], or even lack a surface pattern entirely [35]. To highlight these variations and the influential works that preceded this report, the evolution of transmissive sail membranes (2011-2024) is described below. The relative effi-

ciencies of these designs are discussed in chapter 4. Note that $\gamma = 90°$ is assumed whenever the manoeuvring frame $\mathcal{T}_\mathfrak{p}$ is referenced.

The transmissive and refractive solar sail concept began with Swartzlander et al. [37] and an experimental demonstration of the 'optical lift' that is generated by refractive cambered geometries. For one, this was the first demonstration of the transverse SRP $\mathbf{F}_{\mathbf{S}_{\mathbf{x}_\mathfrak{p}}}$ that the transmissive sail class would come to be associated with. Secondly, this demonstrated the self-stabilisation capabilities of cambered *lightfoils* with respect to the Sun; these serve as the basis for several stability optimisation runs later in this report (see section 5.5). The concept of lightfoil sails was advanced further by Artusio-Glimpse [30], with several sail configurations being proposed and modelled, and with a micro-scale 'flying carpet' prototype being fabricated.

The first diffractive sails were proposed by Swartzlander [31, 32, 56] based on Littrow reflection, Littrow transmission, and Sun-facing metamaterial gratings. The two Littrow membranes comprise blazed ('sawtooth') gratings configured so that light at the relevant wavelengths are diffracted in accordance with the direction of the incident sunlight. In the reflection case, this means that sunlight is redirected back the way it came. In the transmission case, sunlight is redirected through the membrane such that the incident and the diffracted sunlight are symmetric about the sail plane ($\hat{\mathbf{y}}_\mathbf{B}\hat{\mathbf{z}}_\mathbf{B}$). Interestingly, the Littrow reflection membrane [31] generates SRP in a manner akin to an absorption sail — but with the greater efficiency of a reflective sail. As well as not being truly transmissive, it also flips the transmissive sail paradigm by being incapable of transverse $\mathbf{F}_{\mathbf{S}_{\mathbf{x}_\mathfrak{p}}}$ while efficiently generating longitudinal $\mathbf{F}_{\mathbf{S}_{\mathbf{y}_\mathfrak{p}}}$. In section 2.1.2, longitudinal SRP was called the 'wasteful' SRP component, and indeed, it is not useful for orbit raising in the ideal $\gamma = 90°$ scenario. However, such a behaviour could be advantageous in certain unusual scenarios, e.g. for being used to 'hover' in a fixed point above the Sun.

The proposed Littrow transmission membrane [31] was also atypical for a transmissive sail, with behaviour more akin to a conventional reflective solar sail. Like a reflector, it was incapable of generating useful $\mathbf{F}_{\mathbf{S}_{\mathbf{x_p}}}$ at $\alpha = 0°$, and instead maximised $\mathbf{F}_{\mathbf{S}_{\mathbf{x_p}}}$ by tacking to $\alpha_{\text{opt}} = 35.26°$. The distinction between this and a reflective sail was that α_{opt} yielded equivalent $\mathbf{F}_{\mathbf{S}_{\mathbf{x_p}}}$ but lower $\mathbf{F}_{\mathbf{S}_{\mathbf{y_p}}}$ to a similarly steered reflective sail. This means that its net SRP was lower but its transverse SRP ratio λ was higher. Because both Littrow sails are incompatible with the sensitivity model used in section 4.2, they are mostly excluded from the FM simulations.

The Sun-facing transmission membrane was proposed alongside its peers [31], but was also developed further [32, 56]. This was the first diffractive sail to behave according to the 'transmissive sail archetype' — characterised by high transverse $\mathbf{F}_{\mathbf{S}_{\mathbf{x_p}}}$ from a Sun-pointing $\alpha_{\text{opt}} \approx 0°$ — but was also the first to be inextricably linked to metamaterials [32]. This paper also suggested that performance could be improved further by dynamically controlling the period of the grating via electro-optical means. Later, membranes comprising right-prism gratings were also described [33] and shown to demonstrate very high, archetypal performance — particularly its metasail variants.

A high readiness refractive solar sail composed of simple triangular micro-prisms was proposed by Firuzi et al. [29]. The design of this patterned membrane evokes the blazed gratings of earlier diffractive designs, but at a larger scale so as to prevent diffraction from occurring at the relevant wavelengths. The higher readiness of this design is a result of the lower resolution pattern transfer methods that it entails, as well as its reliance on simple materials (albeit not without its issues; see section 2.2.2). These designs also adhered to the transmissive sail archetype, but were modest in terms of performance. This paper was also notable for its proposal of utilising micro-prisms for active attitude (roll) control by combining alternating micro-prisms with an opacity control device, and for proposing to

utilise incrementally different sized prisms to provide passive attitude (pitch and yaw) control. An alternative configuration was also proposed that would act as a solar collector for the realisation of a solar photon thruster. Importantly, this paper is apparently the first to recognise the potential for reduced sensitivity to altitude of transmissive sails, and was the first to demonstrate that they may operate at lower altitudes than reflective sails by adopting a Sun-pointing $\alpha_{\text{opt}} = 0°$ attitude. This idea was particularly influential to this report. Furthermore, the triangular prism design is subject to optimisation later in this report (see section 5.3).

A lower readiness but higher performance refractive metasail was also proposed by Firuzi et al. [35] composed of gradient-index metamaterials. This membrane design is notable for not featuring a surface pattern at all; instead, the refractive index of the membrane varies according to a pattern, allowing the path of incident photons to be precisely controlled. Using conformal mapping techniques [63], this gradient-index was fine-tuned to achieve a highly archetypal design with $\alpha_{\text{opt}} = 0°$.

Srivastava et al. [57] proposed a 'bi-grating' diffractive beam rider that would be paired with a beamed energy system. The proposed membrane would overcome the issue of beam-propelled solar sails displacing free from the path of a beam by passively generating corrective transverse SRP; its attitude about this beam was also shown to be passively stable.

Chu et al. [36] proposed a liquid crystal diffractive sail that could feasibly perform electro-optical control of its grating without the use of metamaterials, thereby allowing for tunable SRP mid-flight. The membrane of this sail would be composed of liquid crystal optical phase arrays (LC-OPAs) — which are a higher readiness technology. These liquid crystals are ellipsoidal and will align themselves according to an applied electric field. In this case, the field is applied by an array of electrodes spanning the surface of the membrane. Light passing

through two differently oriented crystals will be refracted differently due to their disparate angles of incidence on entry and egress from the crystal as well as the variable birefringence of these ellipsoidal crystals. Effectually, this can be used to change the refractive index of the medium. A voltage applied uniformly about the surface of the LC-OPA may effect a normal gradient-index in a similar albeit lesser manner to the gradient-index refractive membrane [35]. Importantly, a gradient voltage is applied across the surface so as to effect a tangential gradient-index. This shortens or elongates the path that light must take in the tangential axis. This *optical path difference* in the tangential axis effects a corresponding *phase* gradient for the transmitted light, creating a periodic phase grating with electro-optically tunable phase and diffraction effects. Effectively, this results in electro-optically tunable SRP that, being spatially variable, may be distributed to effect a torque and achieve active attitude control (in a similar manner to that of the opacity-controlled alternating refractive prisms [29]).

2.2.2 Proposed Materials

In diffractive and refractive solar sail literature, sail performance has typically been the focus. Often, desired optical properties are assumed and the intended material is left ambiguous. On the other hand, the structural and thermal requirements are typically neglected when a real material is cited (in practice, this is often necessary for an initial proposal to remain in-scope). Therefore, the true suitability of a proposed sail material for spaceflight is not often addressed.

For example, one of the highest readiness transmissive membrane designs features polystyrene (PS) due to its exceptional optical properties [29, 34]; furthermore, PS is helpfully resistant to space radiation [64, 65]. However, it is also brittle and fractures under slight elongation [66]. The suitability of PS for operating in space as a microns-thick membrane is therefore dubious without certain measures being taken. For one, PS could be strategically fractured and adhesed to a suitably

transparent structural layer, thereby negating the risk of structural failure in a manner similar to that of reflective sails. PS derivatives and nanocomposites (including shape memory polymers) have also been proposed with much improved flexibility [67, 68]. On the other hand, PS-based designs could be optimised for one of a number of alternative candidate materials without a significant performance deficit; as explored in section 6.2.2, certain variants of polyethylene terephalate (PET) and polyimide (PI) exhibit similar optical properties to PS, but with far superior mechanical stability.

On the other hand, lower readiness designs typically feature materials that are problematic due to a lack of literature; many high performance transmissive sails cite metamaterials as the source of their highly desirable optical properties [31, 33, 35, 59, 69]. However, it is not yet known whether these materials can be fabricated in a scalable and economically viable manner with the cited properties, nor is it known whether or not these materials are suitable for spaceflight.

2.2.3 Manufacturing Processes

In manufacturing terms, reflective and transmissive solar sails broadly follow the same sequence of processes. Solar sail manufacture typically comprises four stages: *polymerisation*, *shaping*, *functionalising* and *assembly*. Notably, the first two stages are sometimes executed simultaneously, and furthermore, involve the same processes regardless of whether the sail is reflective or transmissive. Conversely, the functionalising stage is distinct and significantly more complex for transmissive sails, while the assembly stage is *mostly* the same as that of a reflective sail.

The *polymerisation* stage involves combining mutually-reactive monomers that form complex polymer chains when they are activated. These monomers come in a variety of forms, but are typically solids or liquids that are combined to form a

resin. Whatever the form, activation of this precursor often involves the addition of a chemical catalyst, heat, UV radiation or a combination of stimuli [70, 71]. An additional post-polymerisation process may also be required to achieve the desired properties [72].

As previously mentioned, the *shaping* stage is related to polymerisation, and the two are not always distinct processes. If these stages are distinct, the output of the polymerisation stage will typically be the final polymer in bead, powder, or a crude sheet form. In this case, the shaping stage involves reforming the polymer into finished sheets by combining extrusion methods (such as doctor blading [73]) with either UV curing (of a resin) or rapid evaporation (of a solution in which polymer beads or powder has been dissolved). If these stages are not distinct, this polymerisation process is carried out simultaneously with an extrusion process to form a thin sheet. For example, biaxially-oriented PET (BoPET) has been used within certain solar sail designs under the trade-name Mylar$^{\text{TM}}$ [74], and is typically produced in a single stage: molten PET is extruded and cooled to form a PET sheet, which is simultaneously stretched or *drawn* to form BoPET with more desirable mechanical and optical properties [75]. Whether simultaneous or sequential, the first two stages fall under the umbrella term of *thin film fabrication*, which is explored further in section 6.2.

Up to this point, reflective and transmissive sail manufacture appears identical. However, the *functionalising* stage for each is notably distinct. For a reflective sail, functionalising the membrane involves *metallising* the polymer film using a *physical vapour deposition* (PVD) process in which a metal (often aluminium) is vaporised or sputtered within a vacuum in the presence of the polymer film, to which the metallic vapours or sputtered atoms bond uniformly [76]. The metallisation of the membrane bestows the sail with the high reflectivity upon which its functionality is predicated. On the other hand, functionalising a transmissive sail membrane involves transferring a pattern to it from a mould. A transmissive

sail membrane that is composed of an appropriate material is inherently transmissive, but without a pattern, transmission will engender mutually destructive, sequential refraction events on the entry and egress of sunlight to zero net effect. Likewise, diffraction will not occur without a suitably sized, grating-like pattern. *Pattern transfer* may involve imprinting directly into the polymer at temperatures beyond the *glass transition temperature* T_g [77, 78] (sometimes the *heat deflection temperature* HDT [79]) of the polymer, or imprinting the polymer in a semi-cured form (with the cure being finished in post) [80]. Alternatively, the pattern may be transferred to a layer of resin that has been coated onto the surface of the base polymer, and that is cured at the moment of imprinting [77, 81]. In any case, pattern transfer implies the existence of a *mould fabrication* stage, which itself is non-trivial. Mould fabrication and pattern transfer processes are explored in sections 6.1 and 6.3, respectively.

During the final assembly stage, the metallised or patterned films are segmented and adhesed together in a manner that is conducive to stowing. The only additional consideration for the assembly of transmissive sails is membrane alignment: each membrane must be aligned according to the grain of their pattern in accordance with the function of the membrane. For example, all sections of membrane that are responsible for acceleration must align their patterns. Conversely, sections of membrane that are responsible for torque may have mutually opposing patterns, depending on their position relative to the centre of the sail and the axis for which they are responsible.

2.2.4 Scalability and Resolution

The functionalising of a transmissive sail is influenced by two competing stressors: *scalability* and *resolution*. The issue of scalability arises from the need to apply a micron-scale pattern to a metre-scale sail. As a rule, a reduction in the size of the features of a fabricated part corresponds with an increase in the required

resolution and a decrease in the throughput of the fabrication process. In this case, this pertains to mould fabrication. The obvious solution would be to reduce the size of the mould, but this would have a knock-on effect for the throughput of the pattern transfer process, which may not be viable. Taken to the extreme, too small a mould will render sail patterning either practically infeasible or financially untenable (although some methods can increase the throughput of small moulds substantially, see section 6.3). This is the primary reason that resolution and scalability must be balanced.

Conversely, resolution is a competing stressor that cannot be ignored, because the scale of pattern features corresponds with sail thickness L_x; sail thickness must be kept low so that *mass efficiency* η_m remains high, as demonstrated by Eq. 2.33 (note that $\eta_m \propto L_x$ is only approximate for transmissive sails due to the presence of negative space in their patterns). This is necessary because the primary advantage of a transmissive sail — namely, higher *transverse SRP efficiency* η_{S_v} — is predicated upon η_m being comparable to that of a reflective sail. As demonstrated by Eq. 2.34 and described by Eq. 2.35, $\dot{v}$ is directly proportional to the product of η_{S_v} and η_m in the absence of atmospheric drag D.

$$\begin{aligned} \eta_m &= \frac{A_{yz}}{m} \\ &= \frac{A_{yz}}{\rho V} \\ &= \frac{A_{yz}}{\rho A_{xz} L_y} \\ &= \frac{L_y L_z}{\rho L_x L_z L_y} \\ &= \frac{1}{\rho L_x} \end{aligned} \tag{2.33}$$

$$\begin{aligned} \dot{v}(\alpha, \beta, \mathrm{S}, h) &= \frac{A_{yz} \sum F_v(\alpha, \beta, \mathrm{S}, h)}{m} \\ &= \frac{A_{yz}}{m}[F_{S_v}(\alpha, \mathrm{S}) - D(\beta, h)] \\ &= \eta_m[F_{S_v}(\alpha, \mathrm{S}) - D(\beta, h)] \\ &= \eta_m[\frac{2G_{\mathrm{SC}}\ \eta_{S_v}(\alpha)}{c\mathrm{S}^2} - D(\beta, h)] \end{aligned} \tag{2.34}$$

$$\dot{v}(\alpha) \propto \eta_m \eta_{S_v}(\alpha) \quad \text{for } D = 0 \tag{2.35}$$

Wherein A_{yz} is sail area, A_{xz} is sail cross-section, L of various subscript are the dimensions of a square sail, ρ is the sail density, and the product $\eta_m \eta_{S_v}$ is the *overall efficiency*.

2.2.5 True Mass Efficiency

Although the importance of η_m has been demonstrated, it could be argued that its importance has been overstated in conventional solar sail literature. Contrary to Eq. 2.34, a 1% decrease in η_m does not correspond with a 1% decrease in $\dot{v}$ for a real sailcraft because η_m does not account for satellite weight, nor the weight of sail support structures. Proportionally, the mass of a solar sail membrane is low relative to a satellite body, both historically (as for IKAROS, accounting for only 19.7% [5]) and within the proposed designs of future missions (as for the NEA Scout, accounting for roughly 2.6% [82]). Distributing the weight of the entire satellite across the sail yields the *true mass efficiency* η_M that is described by Eq. 2.36. This leads to an amendment of Eq. 2.34-2.35 as expressed by Eq. 2.37-2.38.

$$\eta_M = \frac{1}{\rho L_x + \frac{M}{A_{yz}}} \tag{2.36}$$

$$\dot{v}(\alpha, \beta, \mathrm{S}, h) = \eta_M[\frac{2G_{\mathrm{SC}}\ \eta_{S_v}(\alpha)}{c\mathrm{S}^2} - D(\beta, h)] \tag{2.37}$$

$$\dot{v}(\alpha) \propto \eta_m \eta_{S_v}(\alpha) \quad \text{for } D = 0 \tag{2.38}$$

Wherein M is the mass of the satellite bus and sail support structures. The difference between the two efficiencies can be substantial. For example, for IKAROS we find that $\eta_m = 90.2$ m^2/kg while $\eta_M = 3.31$ m^2/kg [5]. This implies that the true mass efficiency is not nearly as sensitive to sail thickness as one might assume. If so, a transmissive sail that increases η_{S_v} but in proportion decreases η_m may yield a net positive *true* overall efficiency $\eta_M \eta_{S_v}$. To highlight this, consider the efficiencies $\eta_m \eta_{S_v}$ and $\eta_M \eta_{S_v}$ of two sails. Using IKAROS as the first example, we find that $\eta_m \eta_{S_v} = 35.4$ m^2/kg and $\eta_M \eta_{S_v} = 1.28$ m^2/kg for a typical tacked manoeuvre ($\alpha = 35.26°$; $\eta_{S_v} = 38.5\%$). This is reached via Eq. 2.39-2.40. Note that these equations feature two expressions for ρL_x because IKAROS is two-layer; one expression is for the polymer (PI) core for which $\rho_1 L_{x1} = 1420 \times 7.5e{-}6$, and one is for the metallic (aluminium) outer layer for which $\rho_2 L_{x2} = 2710 \times 80e{-}9$:

$$\eta_m \eta_{S_v} = \frac{1}{\rho_1 L_{x1} + \rho_2 L_{x2}} \times \eta_{S_v} \tag{2.39}$$

$$\eta_M \eta_{S_v} = \frac{1}{(\rho_1 L_{x1} + \rho_2 L_{x2}) + \frac{M}{A_{yz}}} \times \eta_{S_v} \tag{2.40}$$

Consider next a hypothetical PI transmissive sail that is 50% thicker than IKAROS, but that also has 50% more η_{S_v} (for simplicity, surface features are ignored and the sail is only made thicker). In this example, we find that $\eta_m \eta_{S_v} = 36.1$ m^2/kg and $\eta_M \eta_{S_v} = 1.88$ m^2/kg. This is reached by Eq. 2.41-2.42 using $\rho L_x = 1420 \times 1.5(7.5e{-}6)$ and $\eta_{S_v} = 1.5(0.385)$:

$$\eta_m \eta_{S_v} = \frac{1}{\rho L_x} \times \eta_{S_v} \tag{2.41}$$

$$\eta_M \eta_{S_v} = \frac{1}{\rho L_x + \frac{M}{A_{yz}}} \times \eta_{S_v} \tag{2.42}$$

When these sails are viewed in isolation from the satellite, their overall efficiencies are similar. This complies with the prevailing narrative. In this case, $\eta_m \eta_{S_v}$ is 2% higher for the hypothetical PI sail, simply by merit of the unmetallised PI sail being less dense on average. However, when the satellites are considered as a whole, we find that the true overall efficiency is 47% higher for the hypothetical PI sail. This comparison is drawn to highlight the difference between the usual efficiency narrative and reality: under realistic conditions, an increase in η_{S_v} is typically more valuable than an equivalent increase in η_m. As such, a transmissive sail with higher transverse SRP efficiency than its reflective peers may retain a significant performance advantage even if its mass efficiency is moderately reduced.

2.2.6 Optical Degradation

Optical degradation effects are an asterisk on the proverbial 'indefinite acceleration' of solar sails. Furthermore, the rate at which optical degradation will occur in space is difficult to reliably predict, as it occurs for a multitude of reasons including outgassing, thermal cycling [83, 84], interactions with electromagnetic (EM) radiation, interactions with corpuscle radiation [65, 85, 86] and interactions with atomic oxygen [65]. Every material exhibits different degrees of sensitivity to these phenomena, and each of these sensitivities may also vary with temperature. Furthermore, the expected degree of exposure to these phenomena is not constant; most vary by orbit, solar activity or both. As a result, it is also difficult to predict whether a transmissive sail would suffer less from optical degradation in space by utilising single-layer polymer films instead of multi-layered, metallised polymer films.

However, the available literature is sufficient to at least formulate the problem and make tentative speculations. On the one hand, reflective sail optical degradation due to high-energy EM (e.g. gamma rays) or corpuscle radiation (e.g. protons) often involves the ionisation of the metallic layer as a prominent mech-

anism [87, 88]; polymers are not usually ionised by radiation [89], and would not be degraded through this mechanism. Furthermore, many polymers are less susceptible to damage from high-energy radiation in general because these forms of radiation target their bonds at random — rather than selectively, as for low-energy forms of radiation, such as ultraviolet (UV). The effects of high-energy radiation are therefore distributed and, in some cases, may be harmlessly dissipated [89]. Variants of polyimide and PET have been shown to tolerate high doses of this high-energy radiation before experiencing significant damage [90]. These phenomena may suggest that transmissive sails would experience a reduced rate of degradation from high-energy sources.

On the other hand, no polymer is immune to these phenomena. With sufficiently high exposure, optical degradation will occur due to the formation and accumulation of free radicals, from the forming of conjugate double-bonds, or from polymeric unsaturation [65, 89]. Also, while the metallic layer of a reflective sail may be prone to ionisation from high-energy radiation, it serves to protect the polymer membrane from low-energy radiation: these metallic films are generally quite reflective to UV, and highly reflective to visible and infrared (IR) EM radiation. Furthermore, while an optical polymer will be practically transparent to visible EM, it will be absorptive to UV and IR. In this respect, the ramifications of having no metallic layer will likely vary from polymer to polymer: some may become rapidly degraded due to the scission of certain chemical bonds arising from their selective absorption of UV; others may have mechanisms to dissipate this energy as heat or fluorescence [89], thereby delaying optical degradation.

Of particular interest to the mechanism of thermal cycling is the absorption and emission of visible and IR radiation, which contribute significantly to the heating of a solar sail [88]. Metals and polymers have comparable absorptivity and emissivity in the visible spectrum, but absorptivity and emissivity are both higher for optical polymers within the IR spectrum [91]. As for the aforementioned phe-

nomena, the effect this has on rates of optical degradation is difficult to surmise and is beyond the scope of this thesis.

2.2.7 Non-idealised Sail Phenomena

Besides the optical properties of sail materials and their eventual degradation, there are other factors that contribute to in-flight losses of efficiency and model uncertainties. These pertain to the geometry of the membrane itself relative to the incident sunlight, and includes self-shadowing, wrinkling and billowing [2, 18, 20]. Self-shadowing occurs when a region of a solar sail partially obstructs sunlight to another region; wrinkling occurs primarily due to a localised lack of tension; and billowing occurs due to pressures acting out-of-plane to the support structures of the sail (i.e. acting along the sail normal vector $\hat{\mathbf{x}}_\mathbf{b}$).

Reflective and transmissive sails are not necessarily affected by these phenomena to the same degree. Self-shadowing would be partially mitigated by a transmissive sail by merit of it being transparent (although the scattering caused would be no-less difficult to model), and would be negated if a Sun-pointing attitude was adopted. Billowing may also be mitigated by transmissive sails because a significant component of SRP would be generated in-plane with the sail support structures. Conversely, wrinkling may be helped or hindered by in-plane SRP acting with or against the sail supports in different regions; for a simple transmissive sail, the issue may be slightly exacerbated. However, tangential SRP may also be harnessed to maintain sail tension through the strategic placement and orientation of certain refractive or diffractive patterns.

2.2.8 Additional Operational Considerations

Being transparent, transmissive sails would be less prone to obstructing other systems such as solar panels and antennas; with the exception of IR, EM waves will be scattered rather than blocked, corresponding with a reduction in power or gain rather than an outright loss of function. Conversely, strategically placed refractive or diffractive membranes could be used as a solar collector [29], increasing the power to solar cells or even providing diffuse heating. Deployment of a sail or another functional membrane may also be aided by transmissive design in some circumstances; the deployment of large membranes requires in-plane stresses to unfold, for which refractive and diffractive membranes are well-suited. On the other hand, an often cited method for deploying large sails involves spin deployment (through rotation about $\hat{\mathbf{x}}_\mathbf{b}$) and the harnessing of centrifugal pseudo-forces [2]. Transmissive sails are not inherently incompatible with this, but the intent of such an approach is often to maintain tension by maintaining a high rate of spin after deployment (as for the 'heliogyro' configuration) and to thereby reduce sailcraft mass by removing redundant tensile elements. This, transmissive sails *are* incompatible with, as their SRP vector is always confined to the xy_B plane; the SRP vector of a spinning transmissive sail would rotate as the sail does, cancelling out any tangential component of SRP and nullifying its primary mechanism for acceleration.

As well as being generally lighter than contemporary propulsion systems (corresponding with a reduced cost of launch), these LEO-capable sails may lend operators greater freedom in the choosing of their orbit due to their ability to expedite orbital decay when a satellite is disabled. At end-of-life, active attitude control cannot be assumed, and a disabled satellite will either tumble randomly or precess about $\hat{\mathbf{v}}$. Nonetheless, 'drag sails' have been shown to be effective at promoting orbital decay in the absence of active steering [92, 93]. Functionally, a solar sail in LEO that is not actively steered will behave as a drag sail because

F_{S_v} will tend to oscillate between positive and negative in both the tumbling case and the precession case, nullifying its effect. Conversely, the magnitude of D will librate, but its sign will not change.

Traditionally, satellites in LEO have been required to deorbit within 25 years. Recently, a five year rule has been adopted by the FCC [94], and by proxy, all operators who use US launch services. A sailcraft in LEO can orbit at a greater altitude than other satellites without fear of breaching these constraints; certain satellites may otherwise need to operate a drag sail to conform with these requirements. Furthermore, a sail has the additional benefit of providing a large and uniform target for any debris-capturing device. In the case of harpoon-based systems [95], a solar sail is not only easier to hit, but also comes with less risk of creating additional debris during recapture (i.e. tearing at failure rather than splintering). For laser-actuated systems [96], solar sails may be pushed into a lower or higher orbit with less precision required on the part of the laser, and without ablation needing to occur. Overall, the wide-scale adoption of transmissive sails in LEO may be instrumental in curbing the issue of space debris passively (as well as actively, see section 4.2.5).

Chapter 3

Methodology

This chapter details the models behind the simulations and design tools that were developed for this project. These models pertain to the generation of reflective and transmissive solar sail steering models and flight characteristics in LEO and beyond (section 3.1), the characterisation of refractive sail patterns (section 3.2), and the numerical optimisation of said patterns (section 3.3). The overarching methodology behind the subsequent manufacturing trials is also touched on (section 3.4), albeit briefly, as the specifics of each trial are recounted in detail within chapter 6.

3.1 Flight Model

First we define the metrics used to evaluate solar sails within the FM:

1. 'Local performance' or rate change of specific orbital energy $\dot{\epsilon}(h, \gamma)$: Eq. 2.19 ($\mathrm{J/kg \cdot s}$)
2. Optimal solar incidence angle $\alpha_{\mathrm{opt}}(h, \gamma)$: Eq. 2.21 ($^\circ$)
3. Performance 'breakpoint' altitude $h_{\mathrm{break}}(h, \gamma)$: the altitude (km) at which

the $\dot{\epsilon}$ of a transmissive sail and the reflective baseline sail are equal (if any).

4. Minimum operational altitudes, instantaneous $h_{\min}(\gamma)$, transient $h_{\min^*}(\Gamma)$, stable $h_{\min^{**}}(i)$: the minimum altitudes (km) at which a sail achieves $\dot{\epsilon} > 0$ J/kg $\cdot$ s over different timescales.

5. Orbit-raising time $t_{h_1 \to h_2}(h, i)$: the time taken (months) to accelerate from a circular orbit of altitude h_1 to one of altitude h_2.

3.1.1 Modelling LEO Flight

The atmosphere is modelled as a free-molecule airflow, and all perturbative forces other than velocity-wise SRP F_{S_v} and atmospheric drag D are assumed to be negligible (aerodynamic lift is neglected). This yields Eq. 3.1 and a simplified form given by Eq. 3.2.

$$\begin{aligned} \dot{\epsilon} &= \frac{v}{m} A(F_{S_v} - D) \\ &= \frac{v}{m} A(\frac{2G_{\text{SC}}}{cS^2} \cos^2(\alpha) \cos(\beta^*) - \frac{1}{2}\rho v^2 C_D) \qquad (3.1) \\ &= \frac{v}{m} A(\eta^*_{S_v} F^{\text{MAX}}_{S_v} - \frac{1}{2}\rho v^2 C_D) \qquad (3.2) \end{aligned}$$

The remaining unknowns — the air density ρ and the coefficient of drag C_D — are calculated using models provided by sources [29, 97] (Eq. 3.3-3.5), wherein h (km) is the altitude. It is assumed that $\rho = \rho_0$ in this report, but to highlight that this is a simplification, we include some *solar activity coefficient* $C_\odot$ that accounts for variations in ρ due to the solar cycle. In fact, this coefficient would itself a function of h due to the fact that solar activity affects different layers of the atmosphere differently (for example, it is well known that solar activity disproportionately affects the density of the thermosphere [15]):

$$\begin{aligned}
\rho_0 &= 3.91 \times 10^{-9} \exp(-0.01841h) \\
&+ 1.304 \times 10^{-11} \exp(-0.009264h) \qquad (3.3) \\
\rho &= C_{\odot}(h)\rho_0 \qquad (3.4) \\
C_D &= 2\sigma_n \frac{v_w}{v} \cos^2(\beta) + \frac{2}{\sqrt{\pi}} \frac{v_a}{v_s} [(2 - \sigma_n) \cos^2(\beta) \qquad (3.5) \\
&+ \sigma_t \sin^2(\beta)] \exp(-(\frac{v_s}{v_a})^2 \cos^2(\beta)) \\
&+ 2(2 - \sigma_n)[\cos^2(\beta) + \frac{1}{2}(\frac{v_a}{v_s})^2 + \sigma_t \sin^2(\beta)] \\
&\times \cos(\beta) \texttt{erf}(\frac{v_s}{v_a} \cos(\beta))
\end{aligned}$$

Additionally, the calculation of C_D requires an approximation of the ambient and sail surface molecule speeds v_a and v_w (Eq. 3.6-3.7), as well as the normal and tangential momentum accommodation coefficients σ_n and σ_t (Eq. 3.8) [24, 97]:

$$v_a = 1089 \exp(-0.000604h) + 22.72 \exp(0.004959h) \qquad (3.6)$$

$$\frac{v_w}{v} \approx 0.05 \qquad (3.7)$$

$$\sigma_n \approx \sigma_t \approx 0.8 \qquad (3.8)$$

3.1.2 Generating Sun-velocity Angles for Dynamic Orbits

As explored in section 2.1.4, a satellite in a circular Earth-centered orbit that is described by Γ will encounter a range of γ denoted by set S_γ. The sole exception to this is the $\Gamma \in \{0, 180°\}$ sunrise-sunset polar orbit, for which a constant $\gamma = 90°$ is initially experienced (depicted earlier by Fig. 2.4). In order to provide the FM with S_γ, orbits are generated in 3D space and propagated to ascertain each $\gamma(\nu, \Gamma)$ numerically. Initial orbits are generated with the ascending node offset by 90° from 'aphelion' (closest point to the Sun); when varying the *initial* orbits, changes

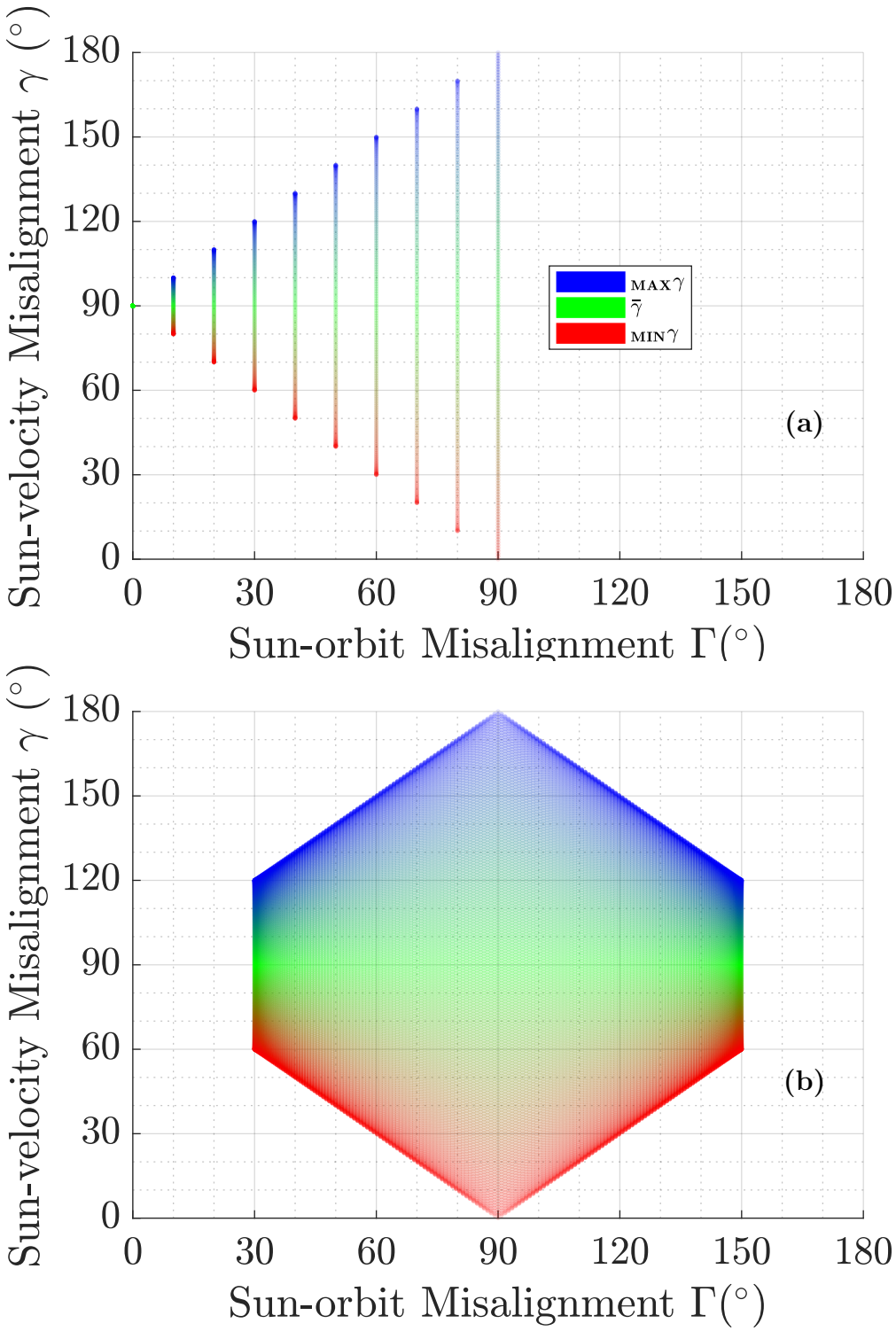

Figure 3.1: (a) S_γ versus Γ for 'initial' orbits of $\Gamma \in \{0, 10, 20... 90°\}$ (sample size lowered for clarity) wherein $i = 90° - \Gamma$, (b) Time-variant S_γ versus Γ for an initial $\Gamma = 30°$, $i = 60°$ orbit

in Γ are achieved by changing orbital inclination i exclusively (never rotating the orbital nodes), such that $\Gamma = 90° - i$. The range of γ experienced by several of these initial orbits is demonstrated by Fig. 3.1a. It can be seen that for each

orbit, the mean is always $\bar{\gamma} = 90°$, while the spread between γ^{MIN} and γ^{MAX} varies more substantially by orbit (generally, a greater spread implies a less optimal orbit for solar sailing).

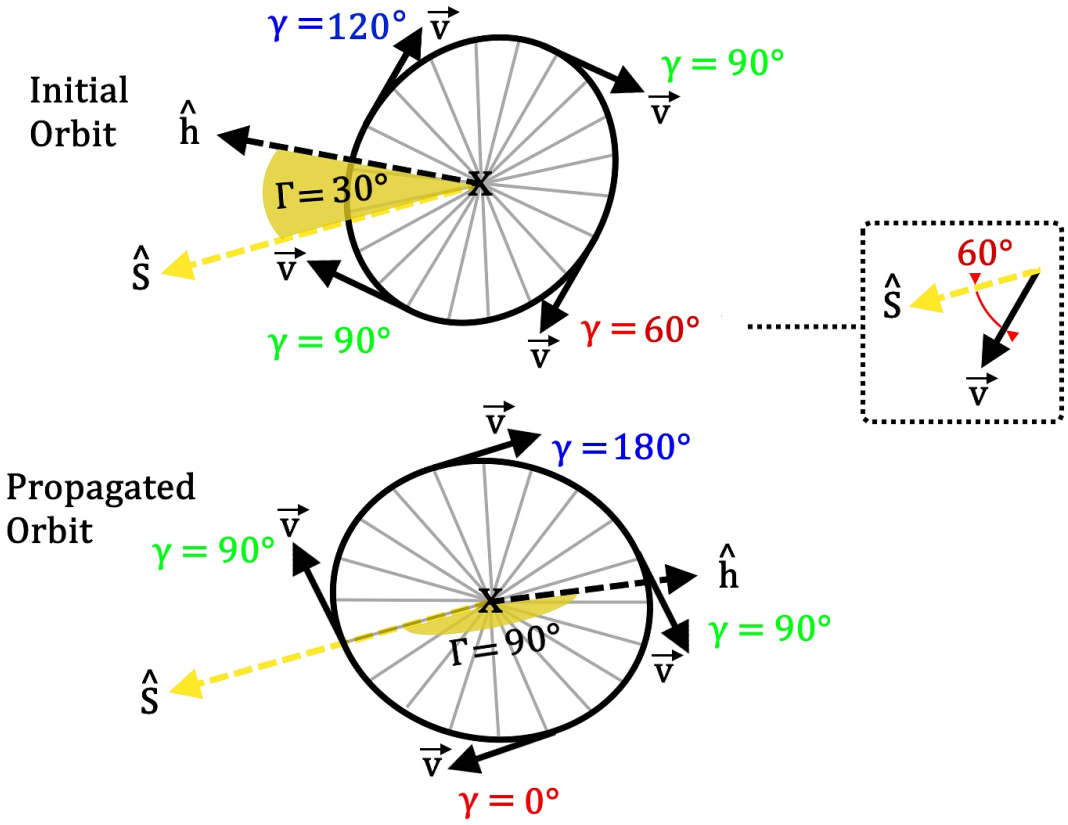

Figure 3.2: Physical Interpretation of Fig. 3.1b from a Sun-fixed Perspective: Four γ Samples for an Initially $\Gamma = 30°$ Orbit, Before (Smallest S_γ Range) and After Propagation (to Largest S_γ Range); γ Colours Correspond with Fig. 3.1

It is also necessary to acknowledge that the Γ of a typical non-heliocentric orbit is not constant; Γ will change because the Sun vector $\hat{\mathbf{S}}$ rotates relative to an Earth satellite's orbital plane as a symptom of Earth's own heliocentric orbit, and because the orbital plane itself will rotate due to the J2 perturbation [98]. A dynamic Γ implies a dynamic range S_γ, and because solar sailing manoeuvres will generally be executed over many months, the S_γ of the orbit may change many times over. This may cause an orbit to oscillate between being optimal and sub-optimal for solar sailing. As an example, Fig. 3.1b demonstrates the substantial variation of γ and Γ that is seen for an initially $\Gamma = 30°$ orbit, and the physical interpretation of this (from a Sun-fixed perspective) is demonstrated

by Fig. 3.2; the initial orbit scenario of Fig. 3.2 is analogous to the $\Gamma = 30°$ spectrum seen in Fig. 3.1b while the propagated orbit scenario is analogous to the taller $\Gamma = 90°$ spectrum.

These time-variant S_γ sets are generated for each orbit by rotating the orbital plane about an Earth-centered ecliptic normal axis in intervals (i = constant). To reduce the number of profiles that need to be generated, axial tilt is neglected; heliocentric Earth and J2 are said to cause rotation about the same axis (in reality, J2 acts about the Earth's axis of rotation [98], which has 23.4° obliquity with the ecliptic normal). These profiles are cycled through by the FM at a rate that is dictated by the current $\dot{\Gamma}(i,h)$ and the time elapsed (see Eq. 3.9-3.11):

$$\dot{\Gamma}(i,h) = \dot{\Gamma}_S + \dot{\Gamma}_{J2}(i,h) \tag{3.9}$$

$$\dot{\Gamma}_S = \frac{360}{T_{\text{Earth}}} = \frac{360}{(365 \times 24 \times 60^2)} = 3.1536 \times 10^7 \ °/\text{s} \tag{3.10}$$

$$\dot{\Gamma}_{J2}(i,h) = -\frac{3}{2}\text{J2}(\frac{r_E}{h+r_E})^2\sqrt{\frac{\mu}{(h+r_E)^3}}\cos(i)\frac{180}{\pi} \tag{3.11}$$

Wherein J2 = $1.08262668 \times 10^{-3}$. It is noteworthy that $\dot{\Gamma}_S$ is independent of orbital parameters, while $\dot{\Gamma}_{J2}(i,h)$ may be selected by controlling them. This brings about a special case known as a Sun-synchronous orbit (SSO), wherein both arguments of $\dot{\Gamma}$ are equal and opposite and $\dot{\Gamma} = 0°/\text{s}$ (see Eq. 3.12-3.13):

$$\dot{\Gamma}_{J2}(i_{\text{SSO}}(h),h) = -\dot{\Gamma}_S \tag{3.12}$$

$$i_{\text{SSO}} = \arccos(-\frac{2}{3} \times \frac{360}{365 \times 24 \times 60^2} \times \frac{\pi}{180} \times \frac{1}{\text{J2}}(\frac{h+r_E}{r_E})^2\sqrt{\frac{(h+r_E)^3}{\mu}}) \tag{3.13}$$

Because this orbit negates dynamic-Γ effects, it is in reality often more optimal

than a sunrise-sunset orbit, particularly for manoeuvres that span many months.

3.1.3 Eclipse

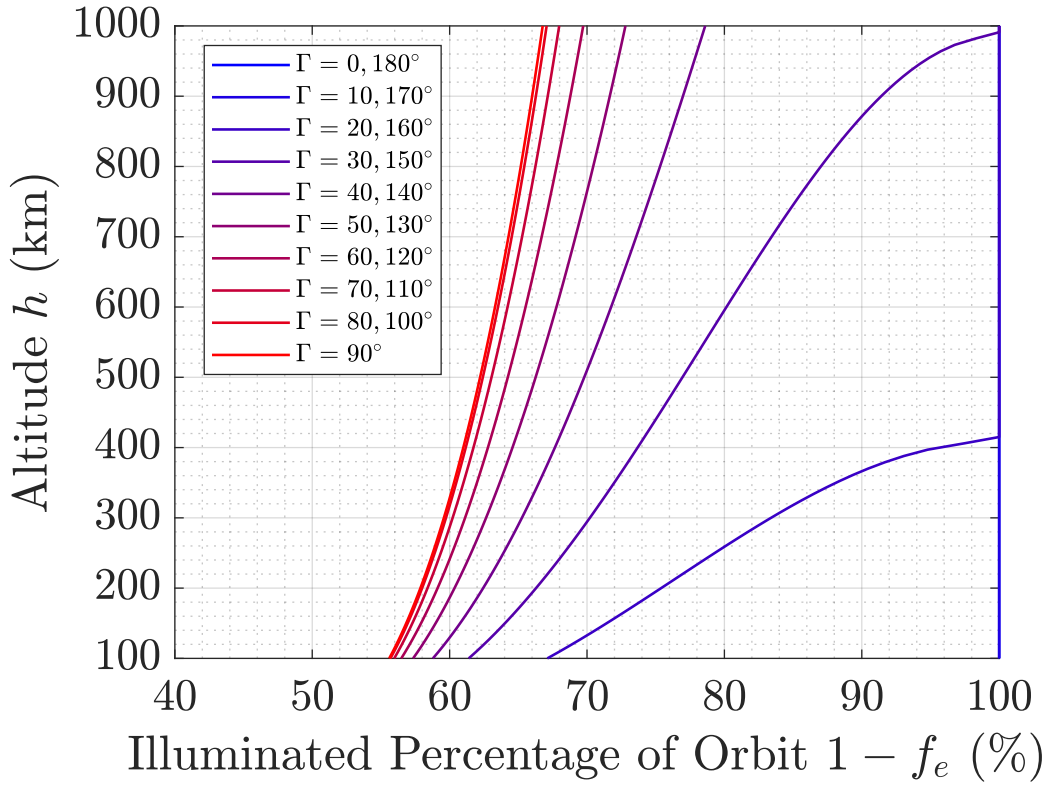

Figure 3.3: Illuminated Orbit Fraction $1 - f_e$ versus Altitude h for a Range of Γ (note that $\Gamma \leq 10°$, $\Gamma \geq 170°$ curves are superimposed at $1 - f_e = 100\%$)

Solar sails in most permutations of LEO are affected by eclipses (see Fig. 3.3). These produce a periodic cessation of any positive $\dot{\epsilon}$, which in turn may erode the propagated performance of the sail in terms of $t_{h_1 \to h_2}$ and minimum operational altitudes $h_{\min^*}$ and $h_{\min^{**}}$. For a given altitude h, a circular orbit will only encounter an eclipse if the Sun-orbit angle Γ is smaller than a certain *eclipse angle* Γ_e (Eq. 3.14) [99]:

$$\Gamma_e = \arcsin\left(\frac{r_E}{r_E + h}\right) + 90° \tag{3.14}$$

If an eclipse will occur, the percentage of an orbit for which a satellite is eclipsed

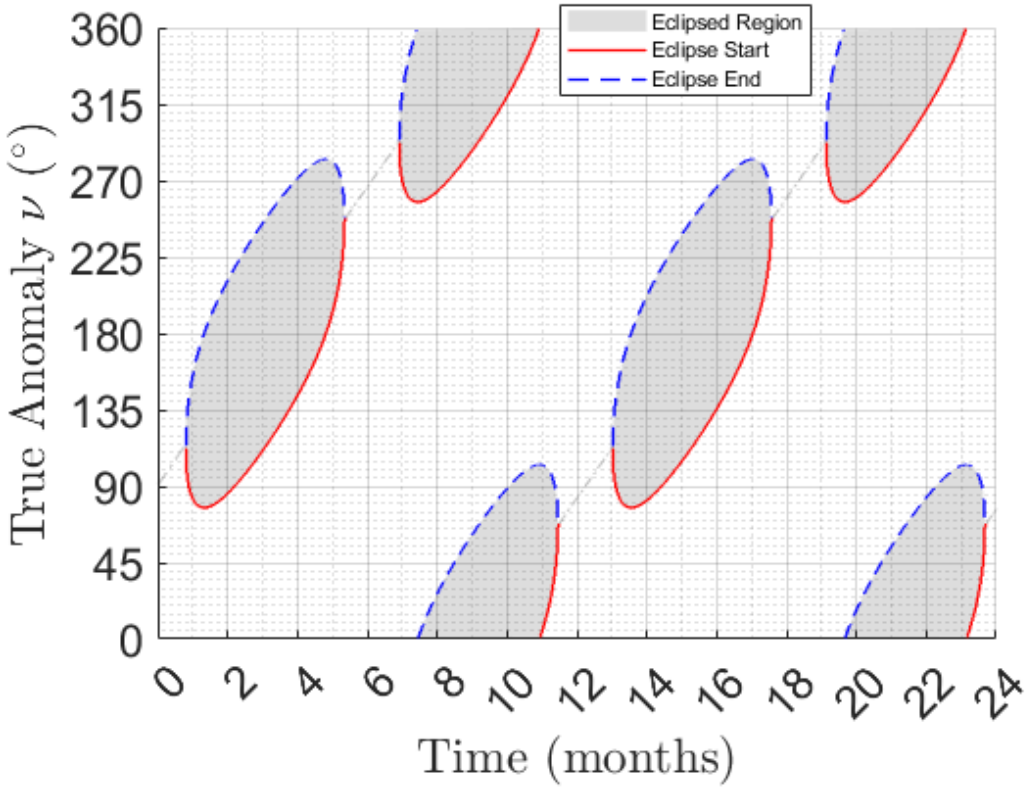

Figure 3.4: Changing Eclipse Region due to a Shifting Γ arising from Relative Sun Position and the J2 Perturbation (550 km Polar Orbiter)

is described by the eclipse factor f_e (Eq. 3.15) [99]:

$$f_e = \begin{cases} \frac{1}{180} \arccos \frac{\sqrt{h^2+2r_E h}}{r_E+h\cos(\Gamma-90)} & \text{if } |\Gamma| < \Gamma_e \\ 0 & \text{if } |\Gamma| \geq \Gamma_e \end{cases} \tag{3.15}$$

Naturally, there is a range of h and Γ for which eclipse will *never* occur (for example, all altitudes at $\Gamma \leq 10°$; see Fig. 3.3). Furthermore, $\dot{\Gamma}$ has a significant effect on eclipse fraction f_e due to the relationship highlighted in Eq. 3.15. For example, a sunrise-sunset polar orbit will eventually precess into a noon-midnight orbit, and then back again. This effect is highlighted by Fig. 3.4 (sunrise-sunset is depicted by the zero-eclipse region; noon-midnight is depicted by the points at which the eclipse region is tallest).

Furthermore, it should be noted that perfectly flat sails are assumed by this model and, because $|\alpha| \leq 90°$ is ensured by attributing these sailcraft with absolute

control authority, the issue of self-shadowing is negated entirely. However, real solar sails are subject to imperfect steering, membrane deformations and occlusion from both the satellite bus and the sail support structures, resulting in self-shadowing effects. This may be particularly severe in low LEO where membrane deformation is elevated by air flux, and in orbits with demanding steering profiles. Lastly, while a binary 'on-off' eclipse is a convenient approximation that does not significantly effect simulation results, in reality the effect is more gradual.

3.1.4 Generating Locally Optimal Steering Laws

The final hurdle for evaluating solar sails is ensuring that each is controlled in-flight optimally. For the reflective solar sail, the locally optimal steering law is determined numerically by solving for α_{opt} via Eq. 3.1 and the definition given earlier by Eq. 2.20-2.21, yielding the steering law shown by Fig. 3.5a. The steering laws of transmissive sails are not so easily derived. A simple Sun-facing $\alpha_{\text{opt}} = 0°$ steering law may suffice for optimal $\gamma = 90°$ orbits, but transmissive solar sails are greatly disadvantaged when this law is applied to other orbits. To maintain equity, true locally optimal $\alpha_{\text{opt}} = f(h, \gamma)$ steering laws are required.

To generate them, α-F_{S_v} profiles are needed for each transmissive sail. Because these are not formulaic as for reflective sails, and because these data sets are not readily available, these profiles are generated through the custom optical simulation that is detailed in chapter 5. The developed simulation focused on refraction, dispersion and reflection by a single-index material; absorption and diffuse reflection are ignored on the assumption that these sail materials are smooth, highly transmissive, and that these mechanisms therefore contribute little to SRP. In particular, neglecting diffuse reflection substantially speeds up simulation times [100]. This simulation is applied to generate α-F_S and α-β^* profiles for a micro-prism array sail of dimensions akin to its original proposal [29].

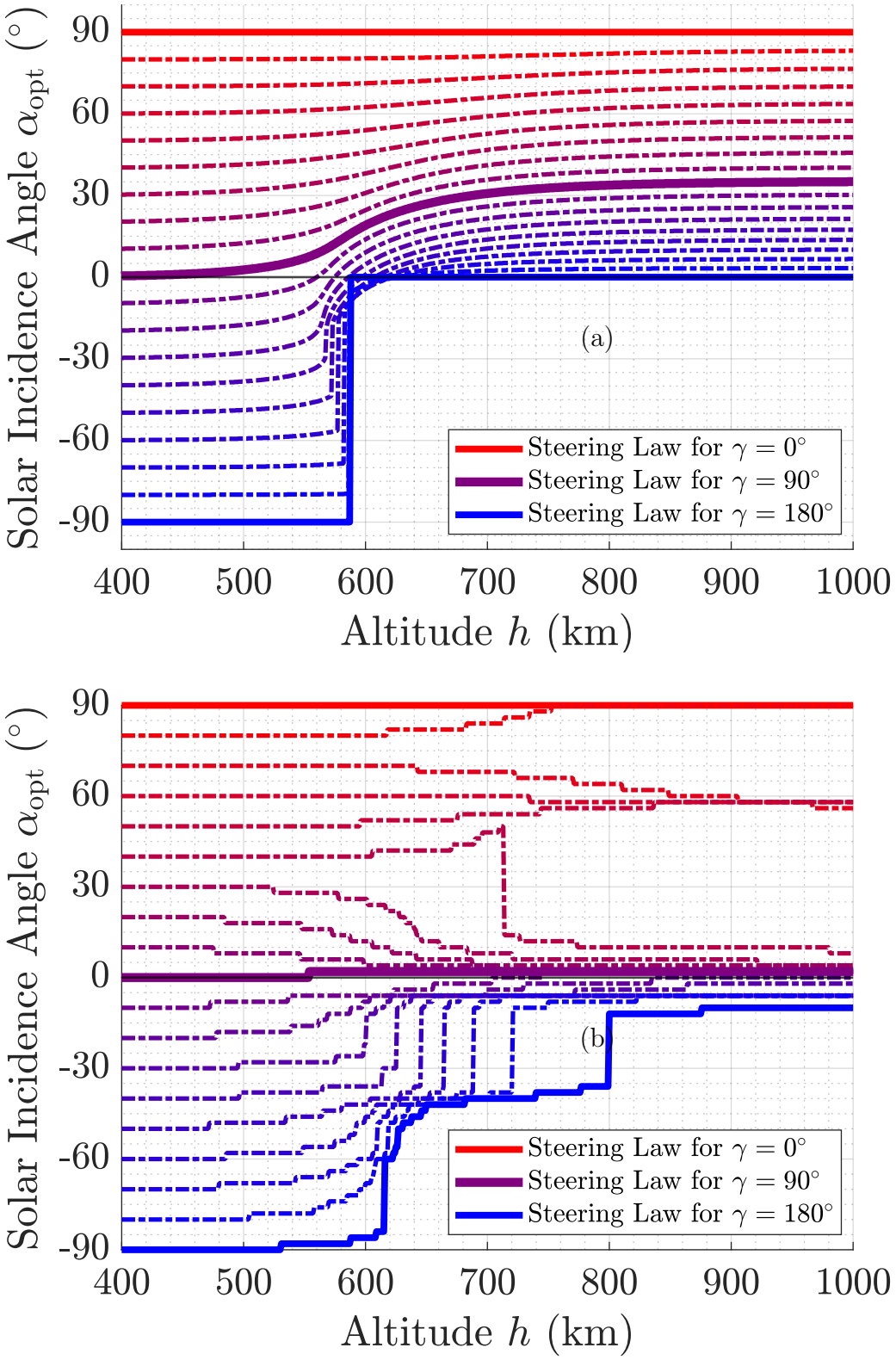

Figure 3.5: LEO α_{opt} for (a) a Perfect Reflector and (b) a (Sensitivity Analog) Gradient-Index Waveguide Solar Sail, using a Locally Optimal Steering Law

This data was used as the generic model for transmissive sail sensitivity to α: the profile was normalised as a percentage of its $\alpha = 0°$ value and multiplied by

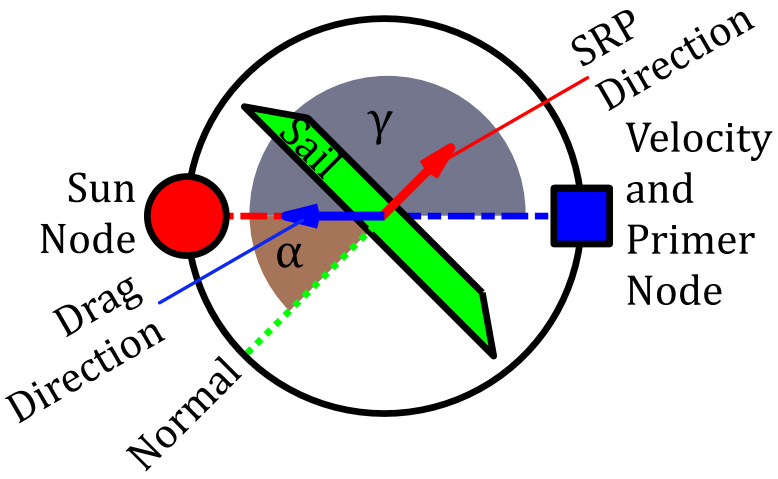

Figure 3.6: Attitude Indication Widget for a Reflective Sail with Normal SRP (for Transmissive Sail Widgets, Tangential and Normal SRP are Assumed to be Equal)

the scalar performance parameters of each sail proposal to generate analogous α-F_S and α-β^* profiles, which in turn were used to generate analogous α-F_{S_v} performance profiles. Therefore, the performance at $\alpha = 0°$ was preserved by these analogs, but their idiosyncratic behaviours across the full γ range were only approximated by the generic model. In particular, the active λ control proposed by the liquid crystal sail [36] may attribute it a flatter sensitivity profile than this model suggests. Nevertheless, most transmissive sail designs demonstrate very similar behaviours within their proposal, such as a slightly off-zero α_{opt}. These transmissive analog profiles were used to generate a locally optimal steering law numerically via Eq. 3.2. An example steering law is shown by Fig. 3.5b. Sails that demonstrate vastly different α sensitivities to this model (e.g. Littrow reflection sail) are omitted from the FM.

It can be seen that an optimal reflective solar sail at low altitude will exhibit an α_{opt} that tends to bring about $\beta = 180 - \gamma - \alpha = 90°$ in order to minimise drag D, and at higher altitudes, converges upon an α_{opt} that favours maximum F_{S_v} as D becomes less significant. Naturally, at $\gamma = 90°$, the reflective sail converges upon $\alpha = 35.26°$. The transmissive sails produced similar steering laws for $400 \leq h \leq 500$ km where the tendency is to minimise drag. Otherwise,

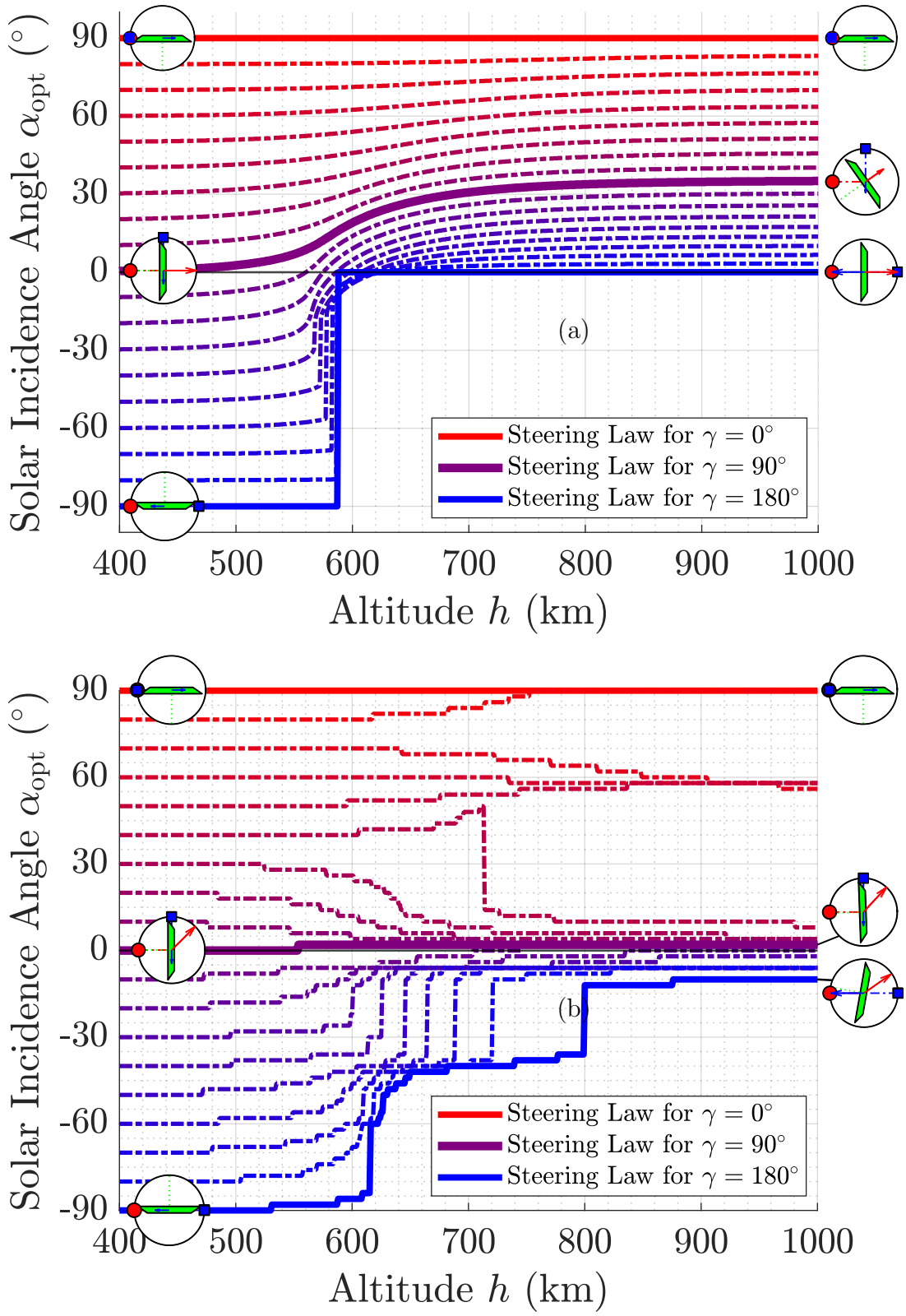

Figure 3.7: Duplicate of Figure 3.5 with Widgets Demonstrating the Low Altitude and High Altitude Attitude for the $\gamma = 0,\ 90\ 180°$ Lines. LEO α_{opt} for (a) a Perfect Reflector and (b) a (Sensitivity Analog) Gradient-Index Waveguide Solar Sail, using a Locally Optimal Steering Law

the steering law curves collate into one of several divergent 'streams' arising due to non-linearities in their SRP-α sensitivity profile. These streams tend to form around peaks in the profile, which become optimal at different γ for different sails. It should be reiterated that these steering laws seek to maximise the rate change of specific orbital energy $\dot{\epsilon}$, which means minimising its magnitude when negative (at lower altitudes) and maximising it otherwise. For example, the steering laws of both sails are identical along the $\gamma = 0°$ curve. Here, F_{S_v} can only be negative, and would act alongside D to reduce ϵ. For the $\gamma = 0°$ curve, maximising $\dot{\epsilon}$ therefore means adopting an attitude for which $F_{S_v} = 0$ Pa and D is minimised, which is the same attitude for both sails.

To aid in the interpretation of these steering laws, duplicates of Fig. 3.5a-b are included that feature attitude widgets. These widgets symbolically represent α and γ through three nodes — the sail, the Sun, and the velocity and primer (desired impulse) node. These widgets are described by Fig. 3.6 and are implemented within duplicates Fig. 3.7a-b. For clarity, these widgets are only added to the $\gamma = 0°,\ 90°,\ 180°$ curves (representing the prograde Sun, nominal, and retrograde Sun steering curves, respectively) and only at the start and end points of these curves (i.e. they represent sail attitude at maximum atmospheric drag, and zero atmospheric drag). Note that an orbit raising manoeuvre is assumed; the velocity node and the primer node are assumed to be one node for a locally optimal steering law.

These duplicates are useful for depicting another feature of these profiles. Each curve is consistent with the positive anticlockwise convention, and the angle definitions given by section 2.1.2 ($\alpha = \angle\ \hat{\mathbf{S}}\hat{\mathbf{n}}$, $\gamma = \angle\ \hat{\mathbf{x}}_\mathbf{p}\hat{\mathbf{S}}$). However, at very low altitudes where F_{S_v} must be zero, the optimal solar incidence is somewhat arbitrary because $\dot{\epsilon}(\alpha_{\text{opt}}) = \dot{\epsilon}(\alpha_{\text{opt}}+180°)$. That is to say that there is only one α_{opt} at altitudes for which SRP is relevant to optimal steering, but there are two possible values of α_{opt} at altitudes for which SRP is irrelevant to optimal steering. For

the sake of clarity, these profiles always favour the α_{opt} that is most congruent with the rest of the $\alpha_{\text{opt}} - h$ curve.

Given the anticlockwise convention and the aforementioned definitions, these widgets demonstrate why negative and positive α_{opt} arise. This is easiest to observe in the reflective sail profile of Fig. 3.7a, for which F_{S_v} changes very predictably. However, it is initially unclear as to why the $\gamma = 90°$ curve of this profile changes so gradually, while the $\gamma = 180°$ apparently flips between one of two states. Indeed, each intermediary γ curve approaching 180° becomes increasingly sharp, and no negative α is seen after 600 km.

The reason for this pertains to section 3.1.1 and the equations Eq. 3.1-3.5. The fact that $\dot{\epsilon} \propto F_{S_v} - D$ has been covered extensively, but in order to explain why these curves differ, the problem must be reframed as the following: F_{S_v} and D are products of *environmental arguments* and *trigonometric arguments*. The environmental arguments of F_{S_v} and D are the available SRP $2G_{\text{SC}}/cS^2$ and the air density $\rho(h)$, respectively. Importantly, the environmental argument of F_{S_v} can be said to be constant because S is effectively a constant in LEO, while the environmental argument of D is a variable, because h is a variable in LEO. On the other hand, the trigonometric arguments of F_{S_v} and D are $\cos^2\alpha\cos\beta$ (based on the assumption that for reflective sails, $\beta^* = \beta$) and $\cos^2\beta$, respectively. In the case of $\cos^2\beta$, this is true because of C_D, which is proportional to several trigonometric terms, but is principally proportional to $\cos^2\beta$. When it is convenient to remove β, this may also be expressed as a balancing of $\cos^2\alpha\cos(180° - \alpha - \gamma)$ and $\cos^2(180° - \alpha - \gamma)$.

The different shape of these steering laws by γ is a result of how these trigonometric terms compare, and how this interacts with the environmental argument of D. The problem may be abstracted as a left-hand-side $\cos^2\alpha\cos\beta$ ('LHS') versus a right-hand-side $C_\rho\cos^2\beta$ ('RHS') for different values of C_ρ, where C_ρ is the environmental argument of D normalised such that the environmental argument

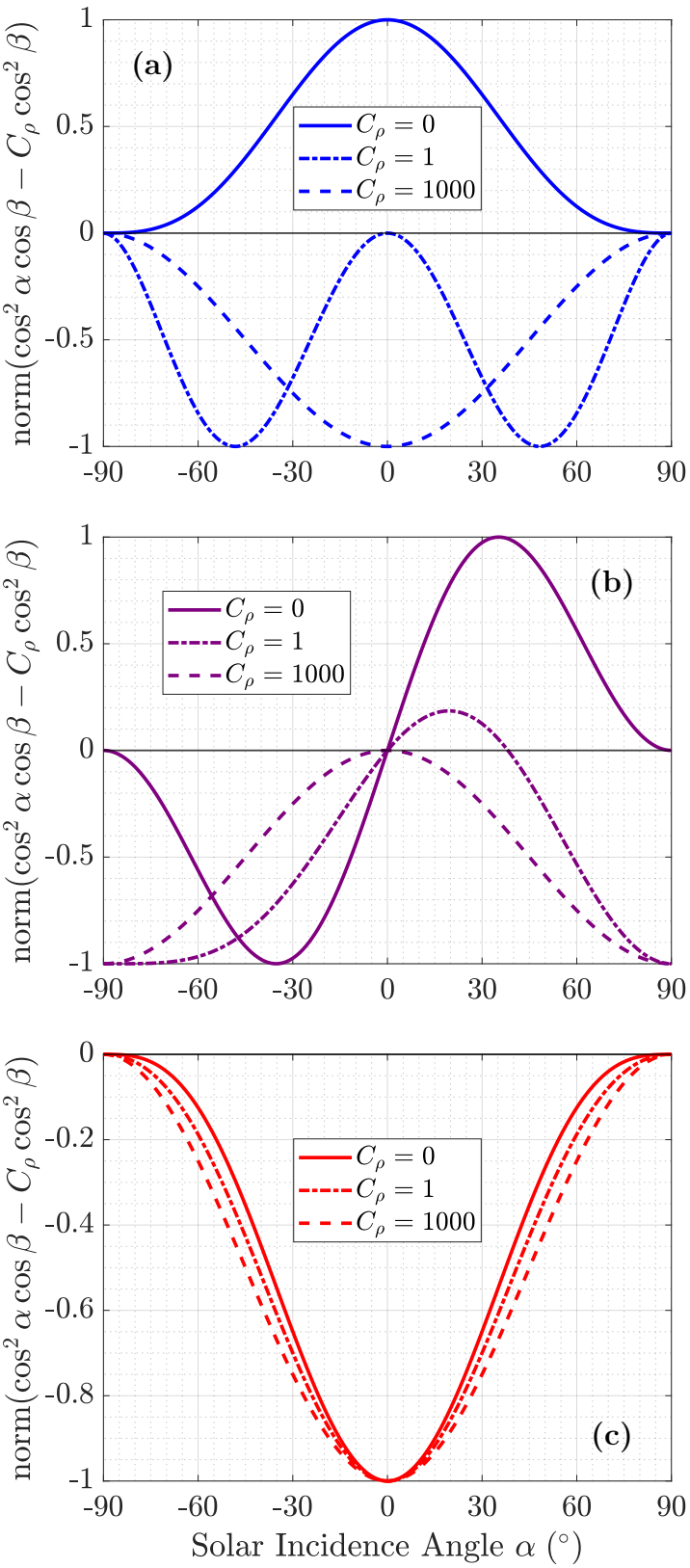

Figure 3.8: Difference between $\cos^2\alpha\cos\beta$ (trigonometric expression governing SRP) and $\cos^2\beta$ (trigonometric term governing drag): for (a) $\gamma = 180°$, (b) $\gamma = 90°$, (c) $\gamma = 0°$

of F_{S_v} equals one. I.e. $C_\rho = 1$ when these arguments are the same size, meaning that SRP and drag are balanced before trigonometric factors. Conversely, $C_\rho = 1,000$ represents a low altitude environment that is dominated by drag, while $C_\rho = 0$ represents a high altitude, zero drag environment. This study is depicted by Fig. 3.8a-c. The important thing to note here is when these curves register positive values, and how these curves progress as C_ρ increases.

The $\gamma = 180°$ (retrograde Sun) case is depicted by Fig. 3.8a, and this demonstrates why the corresponding steering law is so sharp: there is no advantage to be had by changing α from $\alpha_{\text{opt}} = \pm 90°$ while at low altitude (where $C_\rho \geq 1$), as the dashed curves are never positive. Instead, when the altitude is high enough ($0 \leq C_\rho < 1$), the attitude flips to $\alpha_{\text{opt}} = 0°$ because the available SRP exceeds the available drag. From the prior steering laws, we can infer that this takes place just before $h = 600$ km.

Conversely, the $\gamma = 90°$ (perpendicular Sun) case that is depicted by Fig. 3.8b demonstrates why its corresponding steering law curve changes gradually. In the previous case, there were two values of α_{opt}: one for $C_\rho \geq 1$ and another for $C_\rho < 1$. But in this case, α_{opt} can be seen to progress gradually from $\alpha_{\text{opt}} = 0°$ when drag is dominant, to 19.47° when SRP and drag are balanced, and finally to the familiar 35.26° once drag is absent.

Naturally, the steering law curves for $90° < \gamma < 180°$ are influenced more by Fig. 3.8a or Fig. 3.8b depending on there proximity to these γ, explaining why they become increasingly sharp as they near 180°. Conversely, the $\gamma = 0°$ (prograde Sun) case of Fig. 3.8c is included for comprehensiveness, but is hardly worth discussing. It is perhaps noteworthy that Fig. 3.8a and Fig. 3.8c are symmetric because, like the RHS, the LHS is always of the same sign for these figures — either positive (a) or negative (c) and never changing between the two. Conversely, Fig. 3.8b is asymmetric because the sign of the LHS changes based on the sign of α, i.e. SRP may act with or against drag depending on the attitude. It

is also worth noting that this 'balancing of terms' would look different if applied to any of the transmissive sails — as should be apparent from the shape of the example steering law shown in Fig. 3.5b/3.7b — but the fundamental problem would be the same.

3.1.5 Simplified Satellite Bus and Thruster Model

Some simulations require additional parameters pertaining to the satellite itself. Satellite bus drag is assumed to be negligible, and each bus is assigned $m_{bus} = 3.6$ kg. It is assumed that all sails have equivalent momentum accommodation coefficients to that of a conventional solar sail $\sigma_t = \sigma_n = 0.8$ [24], and are attributed identical mass $m = 1$ kg and area ($A = 1$ m^2 unless otherwise stated). For comparison, some of these satellites are modelled with thrusters. Where not specified by a datasheet, 0.5 kg of propellant is assigned. These thrusters [101–104] are attributed equivalent dry masses to that of the sails, and are only made heavier by their wet mass. This ensures that the results are not skewed in favour of the sails by assuming that they are of significantly lower mass (in reality, they almost certainly are, but the magnitude of this discrepancy is unknown). The circular orbit assumption is maintained for all satellites, and so the *Oberth effect* is neglected.

3.2 Optical Simulation

The strategy executed to design and optimise sail membrane patterns was principally to develop both the optical simulation and the optimisation tool in-house, the rationale being that a bespoke simulation and optimiser would afford greater control over which phenomena to explore, which assumptions to make or not make, and over the optimisation criteria. The developed simulation is also mod-

ular, and may be developed further for subsequent research projects [39]. This section details the techniques used by this simulation. However, the optical simulation is more complex than the FM and would be more difficult to replicate as a result. For this reason, the simulation sequence is described in detail at the beginning of chapter 5.

The objective of the optical simulation was to model the SRP generated by a refractive, illuminated object of arbitrary geometry, and thereby test the performance of different patterned sail membranes in a vacuum. Consistent with other literature, objects in the simulation environment are referred to as *particles* [46, 105]. Sails were modelled either per-element (a single particle), or by an array or pattern (multiple particles). The former configuration is useful for calibration and for the exploration of the more orderly optical phenomena, while the latter is useful for the exploration of real sail patterns and the (often chaotic) sequences of ray-particle interactions that dictate their behaviour. Performance data generated by this simulation is expressed as either a pressure F_S or a torque per unit area τ_S in one or more axes, and may be applied to calculate the optical forces and torques exerted upon sails of arbitrary size. Ray tracing is employed for its relative simplicity and suitability for modelling ray-particle interactions in 3D space; this technique may be used to accurately simulate optical phenomena within environments of arbitrary complexity — the trade-off being computation time, which itself may grow arbitrarily long without procedural optimisations.

3.2.1 Ray Tracing

The optical simulation represents light as a grid of parallel rays that are cast from a single plane, as seen in Fig. 3.9. This plane is located at the lightsource position $\mathbf{r_S}$, and lies perpendicular to the lightsource axis $\hat{\mathbf{S}}$. A single ray q is said to have direction unit vector $\hat{\mathbf{c}}_\mathbf{q}$, and this is initially the same for all rays ($\hat{\mathbf{c}}_\mathbf{q}(0) = -\hat{\mathbf{S}}$). Rays are propagated by *ray marching*, whereby the position vector of a ray $\mathbf{r_q}$

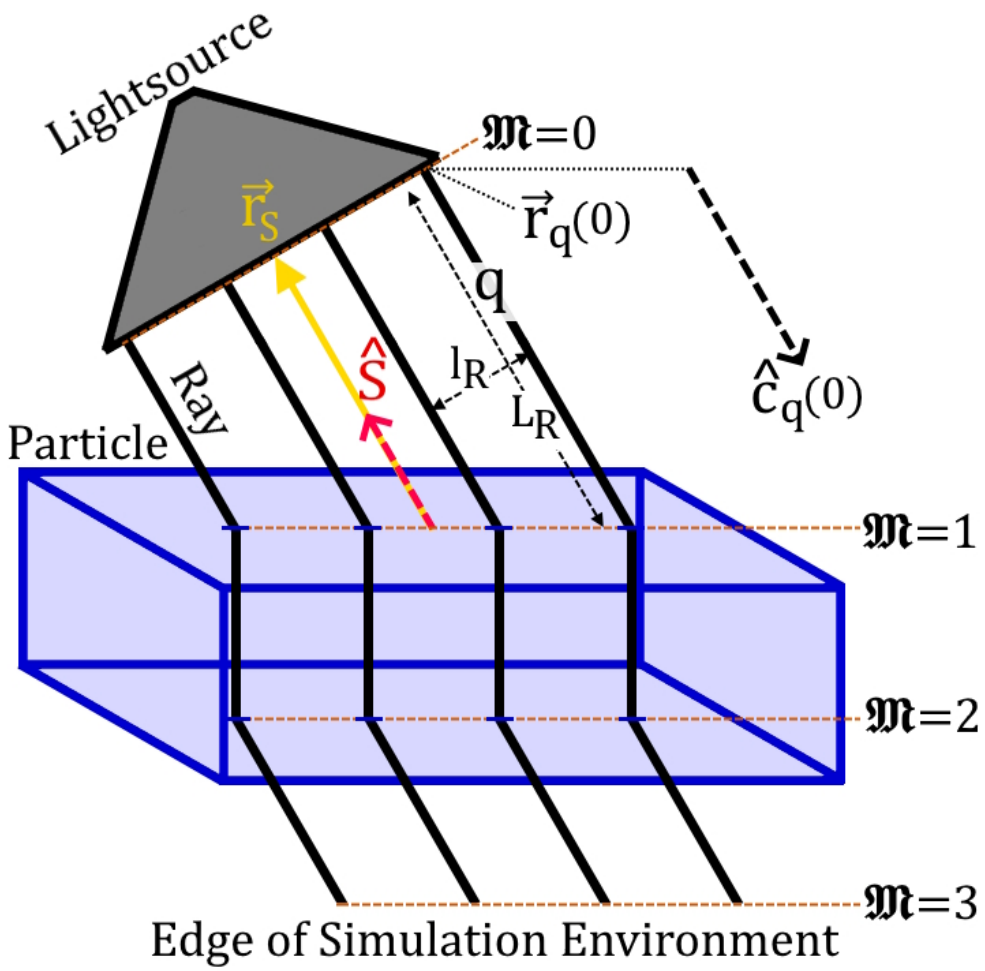

Figure 3.9: Four Rays forming a 2D Grid in 3D Space, Projected onto a Cuboidal Particle (simple case wherein all rays have identical paths and escape the environment at $\mathfrak{m} = 3$)

is advanced along $\hat{\mathbf{c}}_\mathbf{q}$ in small, discrete intervals described by the *march attempt length* ∂L_R. The ray is then checked for collision with any simulated particles. This process repeats until either a collision is detected or the ray leaves the simulation environment. Once either event occurs, a single march is said to have been completed for that ray, and its *march number* $\mathfrak{m}$ is increased by one. $\mathfrak{m}$ is increased by one for every ray that is still within the simulation environment before progressing to the next march. Furthermore, all simulation outputs are time-invariant, and so rays are analysed with respect to $\mathfrak{m}$ in lieu of a true time domain. For example, ray q is cast from the lightsource with position $\mathbf{r}_\mathbf{q}(0)$ and direction $\hat{\mathbf{c}}_\mathbf{q}(0)$; it encounters a particle for the first time at $\mathbf{r}_\mathbf{q}(1)$, at which point its new direction unit vector $\hat{\mathbf{c}}_\mathbf{q}(1)$ is calculated and the process repeats. The

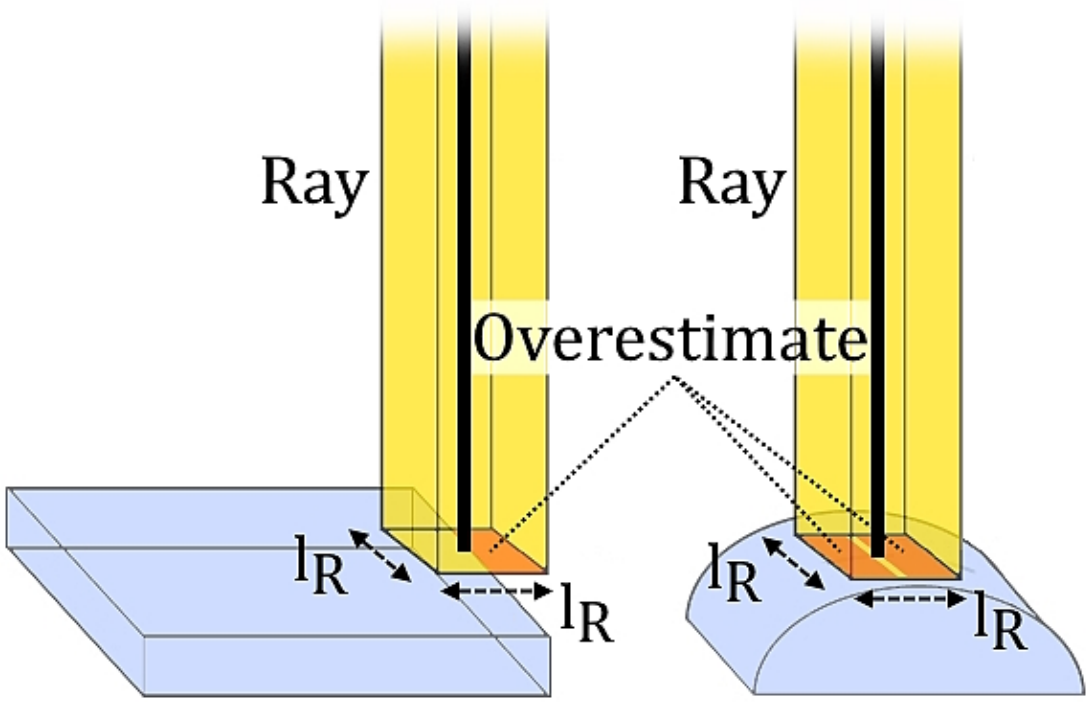

Figure 3.10: Rays Locally Overestimating the Illuminated Area due to Very Large l_R

simulation ends once $\mathfrak{m}$ reaches the user-defined *march limit* $\mathfrak{m} = \mathfrak{m}_{\max}$, or may end prematurely if all rays have escaped the simulation environment. Rays may also be removed from the simulation prematurely according to the ray filter (see section 3.2.7).

In physical terms, a single ray represents the continuum of photons propagating through an $l_R \times l_R \times L_R$ cuboid towards a particle over a single second interval as depicted by Fig. 3.10, where l_R is the ray spacing and L_R is the ray length. This approximation is good when l_R is at least an order of magnitude smaller than the particle. For larger l_R, a ray may significantly overestimate the SRP at the edge of a shape or in regions of sharp curvature as seen in Fig. 3.10. l_R may therefore be decreased for greater accuracy, but higher computation times. Conversely, if l_R is substantially smaller than the resolution of the particle polyhedron, decreasing l_R further (increasing the 'ray density') will yield diminishing returns for accuracy.

Of course, photons are not a purely geometric phenomenon, and so it is for rays. Each ray will have a number of properties attributed to it by its lightsource, some

of which may be altered by subsequent ray-particle interactions. Such properties are said to be *march-variant*, and these must be tracked according to the $\mathfrak{m}$ index. The most important march-variant property is radiant flux $W_q(\mathfrak{m})$, which describes the energy passing through the l_{R}^2 end section of a cuboidal continuum of photons. For example, if a lightsource is to imitate the Sun at 1 au, ray q will be attributed $W_q(0) = G_{\mathrm{SC}} = 1,370\ \mathrm{W/m^2}$ at simulation start. In physical terms, this describes the product of the number of photons per second hitting a surface multiplied by their mean energy, which is a function of Planck constant h and mean photon frequency $\bar{f}_\gamma$ (Eq. 3.16).

$$W_q = \frac{dN_\gamma}{dt} h \bar{f}_\gamma \tag{3.16}$$

Wherein dN_γ/dt is the rate of photon collisions. Other relevant properties are detailed in section 5.1.

3.2.2 Particles and Intersection Detection

Each particle in the simulation environment is represented by a convex hull constructed from a set of vertices using *Delaunay triangulation*. Each convex hull comprises a number of triangles. After a march attempt, the end position of a ray may be cross-referenced with the vertex set of these convex hulls to ascertain whether a particle has been intersected. Once an intersection is found, a numerical search is carried out on the triangles nearest to $\mathbf{r_q}(\mathfrak{m})$: after confirming that a triangle plane is not parallel to the ray in question, the intersection point is cross-referenced with the triangle vertices to ascertain whether or not it lies within the triangle perimeter. In the event that intersection with multiple triangles is found, the nearest to $\mathbf{r_q}(\mathfrak{m}\text{-}1)$ is chosen (this may happen at a cusp, where two edges come very close to one another). To reduce the computational strain of this method, each convex hull is attributed a cuboidal bounding box which

delineates the outer region for which a ray march must check for intersection.

The other reason for determining which triangle was intersected is to query the surface normal vector $\hat{\mathbf{n}}_{\text{tri}}$. This is needed to calculate the angle of intersection θ_i as described by Eq. 3.17. This in turn is used to calculate angles of transmission and reflection (see section 3.2.3).

$$\theta_i = \arccos\left(\frac{\hat{\mathbf{c}}_\mathbf{q} \cdot \hat{\mathbf{n}}_{\text{tri}}}{|\hat{\mathbf{c}}_\mathbf{q}||\hat{\mathbf{n}}_{\text{tri}}|}\right) \tag{3.17}$$

3.2.3 Ray-particle Interactions

When a ray intersects with a particle, one or more ray-particle interactions will occur depending upon the angle of intersection θ_i and the properties of the two mediums (the particle medium and the vacuum medium). As explored in section 2.1.1, a photon-particle interaction will typically comprise absorption, reflection (specular or diffuse) and transmission to varying degrees. In the case of optical membranes that are highly transparent, absorption is typically a minor contributor to SRP. It also occurs irrespective of θ_i, and therefore will vary less between designs. As such, it is less relevant to the design process and is neglected. Furthermore, the geometries explored are assumed to be an order of magnitude larger than that of the dominant wavelengths of the solar spectrum, and so diffraction effects are also neglected in favour of pure refraction. In summary, when a ray encounters a particle within the simulation environment, a portion of the ray will be transmitted (and if $\theta_i \neq 0°$, refracted), and a portion will be reflected.

The exception to this rule is *total internal reflection* (TIR), which results in reflection without a transmitted component. This may occur when a ray attempts to propagate from a more refractive medium to a less refractive one (i.e. $n_1 > n_2$), such as when a ray attempts to travel from a particle to a vacuum. This will only occur if θ_i is larger than a certain *critical angle* θ_c, which itself is determined by

the ratio of refractive indices of the two mediums (Eq. 3.18).

$$\theta_c = \arcsin(\frac{n_2}{n_1}) \tag{3.18}$$

Otherwise, an incident ray will always split into a transmitted ray and a reflected ray. For the reflected portion, the angle of reflection θ_r is dictated by the law of specular reflection (Eq. 3.19). For the transmitted portion, the angle of refraction θ_t is determined by the indices of refraction of the mediums that are being left — n_1 — and entered — n_2 (Eq. 3.20). As explored in section 5.1 and 5.2.1, the ratio of radiant flux attributed to the reflected ray and the transmitted ray is dependent on the material properties (reflectivities $\{R_S, R_P\}$ and transmissivities $\{T_S, T_P\}$), which are wavelength-dependent.

$$\theta_i = \theta_r \tag{3.19}$$

$$n_1 \sin(\theta_i) = n_2 \sin(\theta_t) \tag{3.20}$$

Because refractive index n is also a function of wavelength, transmission may incur another kind of interaction called *dispersion*. Within the simulation, Sellmeier equations (Eq. 3.21) are used to determine n depending on the wavelength λ_k of the specific ray that is being refracted (for example, infrared rays are refracted less than ultraviolet rays).

$$n^2(\lambda_k) = 1 + \frac{B_s \lambda_k^2}{\lambda_k^2 - C_s} \tag{3.21}$$

Wherein B_s and C_s are the material Sellmeier coefficients. When the transmitted ray is monochromatic, dispersion is not incurred and both the reflected component and the transmitted component will share the same colouration. When a ray is polychromatic (e.g. white light), the reflected ray will retain the polychromatic composition of its parent, but the transmitted ray will disperse into several monochromatic rays of different colour. In this case, the simulation carries out

Eq. 3.21 once for each ray that is generated by the dispersion event. Additionally, the radiant flux that would otherwise be attributed to a single transmitted ray must now be subdivided between each dispersion ray according to the power distribution of the electromagnetic spectrum of the lightsource, as described in section 3.2.4.

3.2.4 Black-body Radiation Model

In truth, dispersion rays that are created by the simulation are not really *mono*chromatic as this would require a ray to be generated for every wavelength within a spectrum — an uncountable set. Instead, it is convenient to subdivide a spectrum into several smaller spectra or *bands* and attribute a ray to each of them. To determine how the radiant flux of a polychromatic parent ray should be distributed amongst its pseudo-monochromatic children rays after dispersion, the spectral power distribution (or radiance) of its lightsource must be known. For a lightsource radiating idealised white light, this distribution would be uniform and flat. Conversely, the Sun radiates polychromatic light with a non-uniform power distribution: white by convention, but not perfectly so. Within the visible spectrum alone, more red and green photons are produced than blue ones, producing light with a yellow hue.

Despite its non-uniformity, the power distribution of the solar spectrum is simple to model by approximating it to be black-body radiation. The spectral radiance B of a black-body radiator at wavelength λ_k may be calculated via Eq. 3.22, which is derived from *Planck's law*.

$$B(\lambda_k) = \frac{2\pi hc^2}{\lambda_k^5(e^{\frac{hc}{k_B\lambda_k T_o}} - 1)} \tag{3.22}$$

Wherein h is the Planck constant, c is the speed of light in vacuum, λ_k is the ray wavelength (in μm), k_B is the Boltzmann constant, and T_o is the absolute

temperature. By attributing the Sun a temperature of $T_o = 5778$ K, this model yields the spectral power distribution depicted by Fig. 3.11. Next, the solar

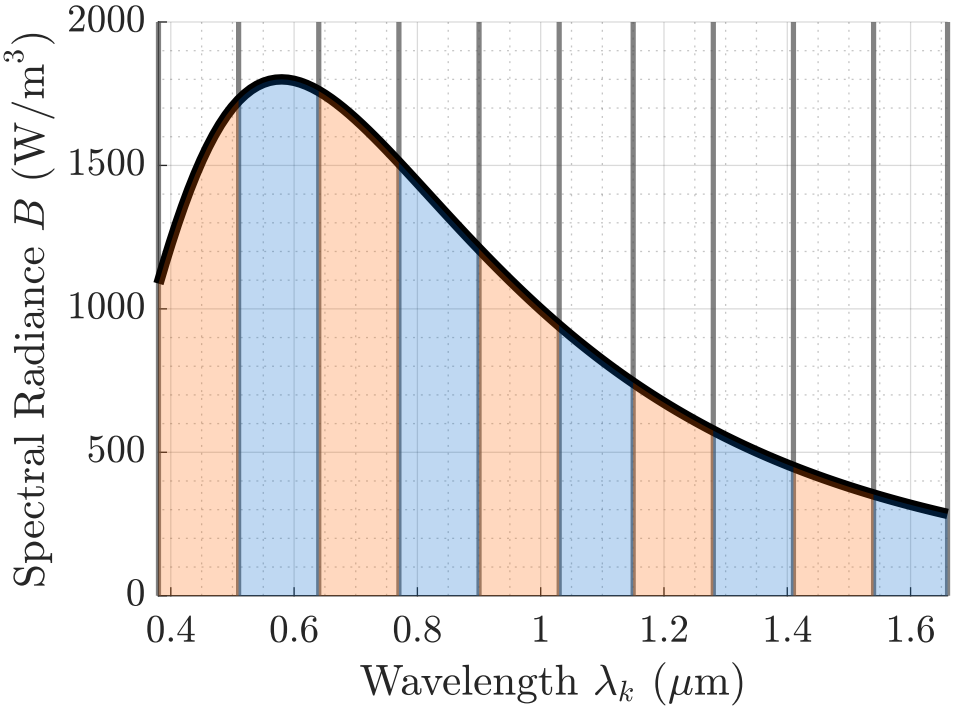

Figure 3.11: Spectral Radiance of Sunlight as Black-Body Radiation, Divided into $D = 10$ Bands, wherein the Area of Each Segment Determines the Ray Flux of Dispersed Rays

spectrum is divided into a user-defined number of bands, D (also present in Fig. 3.11, wherein D = 10). The spectral radiance of each band $B_{\text{band}}(\text{d})$ may be calculated by numerically integrating B across the full range of wavelengths for said band. For our purposes this is approximated by the trapezium rule as shown in Eq. 3.23.

$$\begin{aligned} B_{\text{band}}(\text{d}) &= \int_{\lambda_{d1}}^{\lambda_{dI}} B \, d\lambda_d \\ &\approx \frac{\Delta\lambda_d}{2}[B(\lambda_{d1}) + \sum_{i=2}^{I-1} B(\lambda_{di}) + B(\lambda_{dI})] \end{aligned} \tag{3.23}$$

Wherein λ_d is a wavelength sample within band d and $\Delta\lambda_d$ is the sample interval. Next it is necessary to calculate the spectral radiance of the entire spectrum B_{spec}, for which the procedure is identical as Eq. 3.23: the only difference is the limits,

which are chosen to be $\lambda_{d1} = 380$ nm and $\lambda_{dI} = 1,660$ nm because most optical materials demonstrate good optical transmissivity within this range [106, 107]. When a parent ray q with radiant flux W_q disperses into several children rays, the power that is to be attributed to a dispersed ray belonging to band d can now be determined using Eq. 3.24.

$$W_d = \frac{B_{\text{band}}(d)}{B_{\text{spec}}} W_q \tag{3.24}$$

This band approach allows the user to further select for accuracy or computational speed by adjusting the number of bands D, but should be carefully controlled. This is because although dispersion can only occur once within the 'family tree' of a ray, each ray may still produce children via partial reflection, as may their children, and so on. A small increase in D may therefore correspond with a disproportionate increase in the number of children rays and greatly increased computation times.

3.2.5 Solar Radiation Pressure from Radiant Flux

Because the direction of a ray is changed by ray-particle interactions, its momentum must change as well. According to *Newton's third law*, the particle responsible for this must experience an equal and opposite reaction force to that of the ray. In the case of a large number of rays being projected, the SRP exerted upon a particle $\mathbf{F_S}$ may be calculated by comparing the sum of the flux of every *input* ray with the sum of the flux of every *output* ray as depicted by Eq. 3.25 [46]:

$$\mathbf{F_S} = -\frac{1}{c}\left(\sum_{n=1}^{N_{\text{out}}} W_{\text{out}} \hat{\mathbf{c}}_{\text{out}} - \sum_{n=1}^{N_{\text{in}}} W_{\text{in}} \hat{\mathbf{c}}_{\text{in}}\right) \tag{3.25}$$

Wherein W is the ray flux, $\hat{\mathbf{c}}$ is the ray direction unit vector, and N_{in} and N_{out} are the numbers of rays input and output of the simulation, respectively. This approach is valid, but does not lend itself well to calculating torques, for which

the location of each ray-particle interaction is relevant; it also does not allow for calculating pressures imparted by absorption, if one wished to add that functionality. A more robust method is to calculate the change in ray flux at *each interaction*. This is described by the simulation sequence in section 5.1, and so it is not repeated here. These simulations may be repeated across a range of Sun incidence angles α to create a performance envelope for SRP or optical torque per unit area.

3.2.6 Assumptions and Ray Grid Configuration

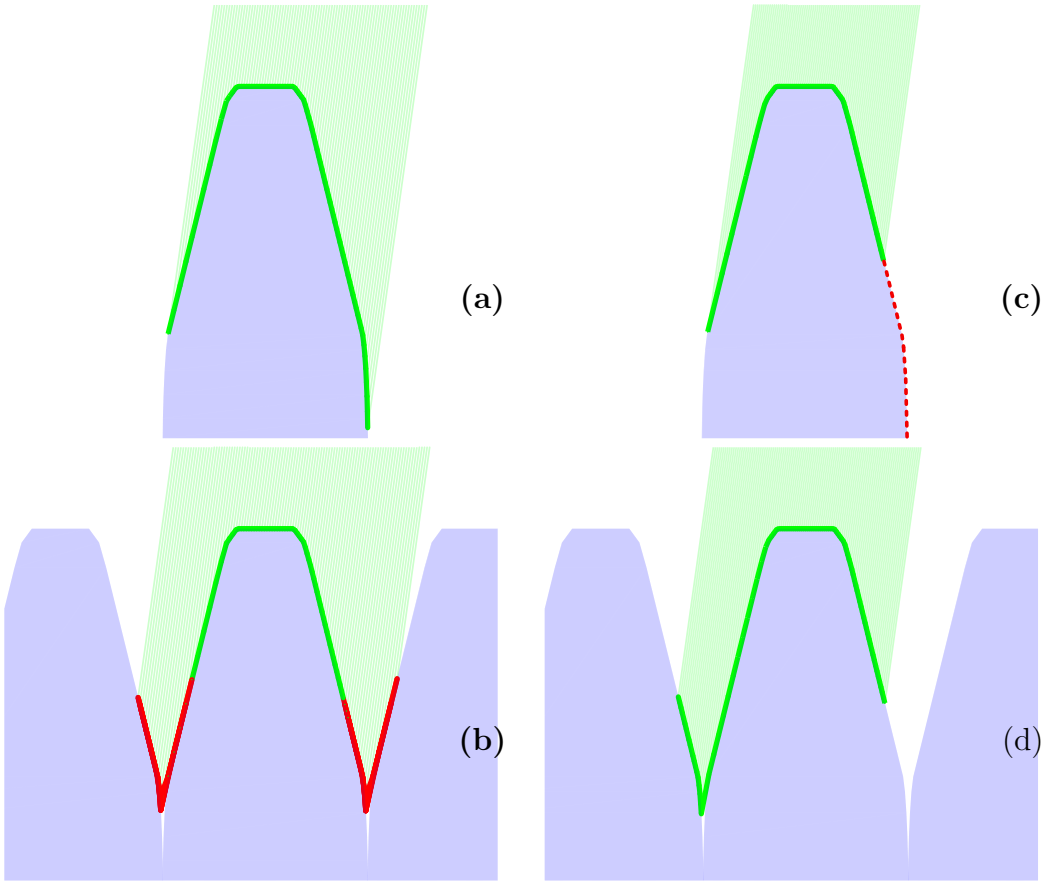

Figure 3.12: Illumination of a (a-b) Particle, (c-d) Pattern, using a (a, c) Particle Ray Grid Configuration, (b, d) Pattern Ray Grid Configuration. Full, Red Edges Show Double-illuminated Regions; Dashed, Red Edges Show Omitted Regions

A *photon*-particle interaction that effects a change in photon momentum would correspond with a point force at a single location on the particle. The aggregate of these momentum changes would naturally be expressed as a net force on the

particle, and their aggregate divided by area as a pressure on the particle surface. Given that rays themselves are often represented as vectors acting upon a point, one might erroneously assume that changes in ray energy would correspond with a force or a pressure when summated, or summated and averaged over an area, respectively. However, rays represent the photon continuum, not individual photons, so this is not true. In lieu of true momentum or energy they posses momentum per unit area, averaged over an area, and energy per unit area, averaged over an area. The latter, which is used by this simulation, is the aforementioned ray radiant flux. When the radiant flux of a ray is changed by a particle, it corresponds with a change in energy per unit area, averaged over the area of the particle that was affected by the ray (which is subject to the ray spacing l_{R}).

Due to the 'per area, averaged over an area' duality, the aggregate of change in radiant flux is indeed dimensionally a force, and as an aggregate divided by area, a pressure. However, unless $l_{\mathsf{R}} = 1$ m (which would be unlikely, given the scale of these optical elements), they will be the wrong force and the wrong pressure. Provided that l_{R} is the same for every ray, and provided that the entirety of an element is illuminated evenly, the pressure acting upon a sail composed of these elements is nonetheless trivial to calculate, being the average change of radiant flux of all illuminating (or 'valid') rays. This is true because, if the entirety of the element is illuminated, the change in energy 'per unit area, averaged over an area' becomes the change in energy 'per unit area, averaged over an element' - which is simply the pressure acting on an element. This in turn is the same as the pressure acting on a sail comprising identical elements. The force acting upon the sail is simply the product of this per-element pressure multiplied by the sail. Assuming diffraction is not considered, such a pressure and a force would be true for a metre-wide sail containing thousands of microscopic elements, or a few, centimetre wide elements; only the dimensions of the sail itself are relevant to the force calculation. Furthermore, while the geometry of the elements is relevant to the calculation of pressure, neither the bounding dimensions of the elements

(their scale) nor the dimensions of the sail will affect this pressure.

As every element in a pattern is identical, the SRP generated by a pattern is the same as the SRP generated by just one of its constituent elements. This may lead one to assume that simulating a single element would suffice, but this would occlude the true behaviour of the pattern as it would prevent ray propagation between adjacent pattern elements, hereafter referred to as *pattern propagation*. Generative designs informed by such a model would be low quality, as the prevention — and sometimes harnessing — of pattern propagation is a major design stressor. At the other extreme, the brute-force approach to accuracy would be to model a small sail in its entirety, attributing a particle to every element of its pattern. Provided that all particles were illuminated evenly, this would be highly accurate. This approach is infeasible though, as a single square metre of sail could imply upwards of a hundred thousand pattern elements and - at a reasonable ray density - over a million rays. The solution is to compromise by modelling just a few pattern elements and to only illuminate the central one. If the central element can be considered typical, the SRP generated by it should be representative of the pattern as a whole, allowing one to predict the performance of a sail, so patterned, of arbitrary size.

Of course, certain assumptions need to be made for this 'per-element' SRP calculation to be truly representative of a real sail. These are described by two axioms. The first axiom is that per-element performance is the same regardless of the position of an optical element within its pattern. Even for an idealised sail, this is obviously false given the prevalence of pattern propagation; an optical element placed at the edge of the sail at either of the $\hat{\mathbf{y}}_\mathbf{B}$ extremities will not be subject to pattern propagation effects when rays are redirected away from the sail edge, and so these would not behave the same as a central element. Yet despite being objectively untrue, this axiom is particularly easy to justify because even a very small sail would comprise thousands of optical elements. Even if a measur-

able deficit could be felt a hundred rows adjacent from the edge, the performance discrepancy between an idealised sail and the per-element estimate would be negligible. Furthermore, the discrepancy would become less pronounced with each adjacent row moving centre-wise and so it would likely not be measurable at a hundred rows' depth. Anecdotally, rays entering a pattern of symmetric optical elements (e.g. lightfoils) were rarely seen propagating through the pattern any more than five elements away from their point of entry, and for asymmetric optical elements (e.g. triangular micro-prisms), rarely more than three. Furthermore, certain patterns do not exhibit pattern propagation at all. Within the simulation, this axiom is kept valid if the number of pattern elements are sufficient to ensure that the ejection of rays from the outer edges of the pattern is an uncommon occurrence. It is difficult to prevent this entirely because this phenomenon – as for all optical phenomena – is sensitive to both α and particle geometry, neither of which are fixed during numerical optimisation. Here the aim is to ensure that first-order rays (see section 5.2.1) are not ejected in this manner, but a few low power rays being ejected so will not significantly affect simulation results. For most geometries, five elements distributed to either side of the central element were more than sufficient to this end.

The second axiom is that the entire particle is illuminated evenly. The validity of this axiom is more complicated to maintain. Naturally, self-occlusion is unavoidable; rotating a lightsource around a sphere will cause one part of its surface to become occluded (where $\hat{\mathbf{n}}_{\text{tri}}$ is perpendicular or obtuse with respect to $\hat{\mathbf{c}}_{\mathbf{q}}$) and another to be illuminated (where $\hat{\mathbf{n}}_{\text{tri}}$ forms an acute angle with $\hat{\mathbf{c}}_{\mathbf{q}}$). Furthermore, for non-spherical particles, protrusions and concavities may occlude parts of a surface that would otherwise be illuminated. Nevertheless, the entire particle can be said to be illuminated if every triangle has the *opportunity* to be illuminated, i.e. if the entire surface *would be* illuminated if rays were allowed to pass through the particle unimpeded. In practice, as long as the rays at the extremities of the ray grid are cast towards the projected extremities of the par-

ticle (namely, projected onto a plane that is perpendicular to $\hat{\mathbf{c}}_\mathbf{q}$), then the entire particle can be said to be illuminated. This also requires that the central ray is cast towards the centre of the particle wherein the centre is the mean average of its *extremities* as described by Eq. 3.26, not the centroid and/or centre of mass. This is the same as the the centre of the bounding box, and indeed, the projected dimensions of the bounding box are also used to size the ray grid for each α for this ray grid configuration as depicted by Eq. 3.27.

$$Y_{\text{grid}-\text{particle}} = X_B|\cos(\alpha)| + Y_B|\sin(\alpha)| \tag{3.26}$$

$$O_{\text{grid}-\text{particle}} = \frac{1}{2}\begin{bmatrix} x^{\text{MAX}} - x^{\text{MIN}} \\ y^{\text{MAX}} - y^{\text{MIN}} \\ z^{\text{MAX}} - z^{\text{MIN}} \end{bmatrix} \tag{3.27}$$

This is called the *particle configuration* hereafter because it is appropriate for single particles as seen in Fig. 3.12a. Conversely, it will not ensure compliance with the aforementioned axiom if the particle is part of a pattern because certain regions will be 'double illuminated' at non-zero α; certain regions that are illuminated on the central particle — which is fully illuminated — are also illuminated on the adjacent particles — which are (erroneously) partially illuminated. The regions for which there is overlap are counted twice by the simulation and so the per-element assumption is broken. Furthermore, at high α the central particle will be partially occluded by an adjacent element. If the region of the adjacent element that is partially illuminated corresponds exactly with the region on the central element that is partially occluded, the per-element assumption holds, but this is not guaranteed by this configuration.

The solution is to have a separate ray grid configuration for patterns. The most appropriate, hereafter called the *pattern configuration*, was found to be one for which the central ray is cast towards the centre of the upper edge of the bounding box as described by Eq. 3.29. Furthermore, the size of the ray grid is only

influenced by the projection of the upper edge of the bounding box as described by Eq. 3.28, meaning that the grid size falls to zero at $\alpha = 90°$ and every ray is superimposed. In practice, this means that SRP and torque are not generated when the sail surface is perpendicular to a lightsource. This faithfully reproduces the assumptions laid out in the FM — namely, that these sails are flat, and so their effective area is zero when the sail normal is perpendicular to the Sun.

$$Y_{\text{grid-pattern}} = X_B |\cos(\alpha)| \tag{3.28}$$

$$O_{\text{grid-pattern}} = \begin{bmatrix} x^{\text{MAX}} - x^{\text{MIN}} \\ y^{\text{MAX}} - y^{\text{MIN}} \\ z^{\text{MAX}} \end{bmatrix} \tag{3.29}$$

Effectively, this configuration treats the particle as a 2D surface with no accounting for depth. It is therefore inappropriate for single particles as seen in Fig. 3.12c, wherein a significant section of the particle that should be illuminated is missed. It is notable that the pattern configuration is also prone to partially illuminating adjacent particles. However, it can be seen in Fig. 3.12d that this does not matter, as the illuminated region of the adjacent particle corresponds exactly with the missed region of the central one. This configuration will only ever cause up to one adjacent particle to be partially illuminated. Nonetheless, this partial illumination means that the centre of mass of the central particle is not always an appropriate argument for the calculation of per-element torque. For example, in Fig. 3.12d, the rays that intersect the left adjacent particle *would* have a ray-particle interaction that occurs to the right of the centre of mass, if they were to have instead intersected with the central element at the corresponding per-element position. Instead, they undergo an interaction to the left of the centre of mass. As such, the local SRP may effect a torque of incorrect sign. Here, the simplest means of maintaining the validity of the second axiom is to calculate torque with respect to the centre of mass of the particle *that was initially intersected* at $\mathfrak{m} = 1$, hereafter called the *reference particle*. For children

rays that did not exist at $\mathfrak{m} = 1$, their reference particle will be the particle that was intersected by their earliest parent ray at $\mathfrak{m} = 1$.

3.2.7 Child Limits and Ray Filtering

With the exception of TIR events, every ray-particle interaction will result in at least one new ray being generated, causing the number of rays being simulated to grow exponentially as $\mathfrak{m}$ is increased. Because it is impossible for a ray to exit a particle without partial reflection occurring, and because absorption is neglected, a ray that has entered a particle is not able to escape without simultaneously sending a new child ray back into the particle. That child ray may itself escape later, but not without sending yet another child ray back into the particle *ad infinitum*. Rarely, an individual ray may also become trapped by cyclical TIR events. For these reasons, it is impossible for the simulation to run to completion with all rays having left the simulation environment. The effect that this has on simulation results is negligible; the ray components that are trapped decrease in flux as $\mathfrak{m}$ increases, and by the end of a simulation their contribution to net SRP is typically less than 0.1%. Nonetheless, they are a hindrance to computation times; only so many interactions can occur within a simulation before the number of rays becomes infeasible to simulate within a reasonable span of time.

Previous works have circumvented this issue by using *sequential* ray tracing [46]. Using this method, any children rays that are created during a ray-splitting event — such as dispersion or partial reflection — would not be subsequently tracked by a sequential ray tracer, and any non-sequential ray-particle interactions would not be considered. Because this may occlude the more nuanced behaviours of a ray propagating through a patterned membrane, this method was not adopted. However, tracking the behaviour of each child ray poses problems for computation speed.

To curb this issue, three control parameters are established. Firstly, the *march limit* $\mathfrak{m}_{\max}$ is introduced to limit the number of successful ray marches that may occur; this is effectively a limit on the number of ray-splitting events that may be permitted. Secondly, the *child limit* $q_{\max}$ is introduced to prevent new rays from being created if the original parent ray already has a certain number of children; if $q_{\max} = 0$, the simulation effectively becomes a sequential ray tracer. Finally, a ray filter is introduced alongside the parameter *filter flux* $W_{\min}$; below this flux, the ray filter will delete a ray if it attempts to create more children. These measures are made more important in the absence of absorption which, as for the sequential ray tracing optical force model [46], is not considered here. This is regrettable because absorption would naturally cause trapped rays to attenuate and more rapidly fall below the filter flux. The author therefore recommends that future models do not neglect absorption.

3.2.8 Validation

The optical simulation was validated via direct comparison with established models through several means. The simplest but least comprehensive of these was to parametrise the compliance of the simulation with simple, fundamental physical laws. For example, one parameter compared the net ray flux of the incident and outgoing valid rays to ensure that the law of the conservation of energy was not breached.

Despite being strictly deterministic, other physical laws are difficult to validate parametrically because of the chaotic and unpredictable manner with which their effects accumulate. In particular, laws governing ray paths (e.g. Snell's law) were validated by confining the simulation to 2D and emulating a given scenario in a well-established and open-source 2D ray tracing model [108]. Validation of these laws was achieved by overlaying the two outputted ray diagrams within a digital photo editor and ensuring that their ray paths were identical.

The next layer of validation was to compare the simulation outputs directly with existing models. This was initially achieved by emulating the simplest scenarios — perfect reflectors and perfect transmitters — and comparing their outputs with the outputs of Eq. 2.15-2.18. For example, it was validated that the SRP profile of a perfect reflector (Fig. 2.5) could be reproduced via the optical simulation. Notably, all prior validation measures have high certainty because the phenomenon that they are validating are simple; in these simple idealised scenarios, the optical simulation and the model being compared feature either no assumptions or the same assumptions.

Conversely, the final stage of validation involves comparing the final outputs of the simulation against the final outputs of prior works that examine transmissive solar sails. This stage of validation is more comprehensive, but owing to the disparate assumptions and methods used by different simulations, is less conclusive. For example, a ray filtered, 'branching' ray tracer may yield different results to that of a sequential ray tracer. Furthermore, simulation outputs in the existing literature are typically not paired with a ray diagram, making an exact reproduction of their scenario impossible. Nevertheless, to validate the optical force model in context, micro-prisms profiles provided by source [29] were compared (these are discussed in section 5.3.2). To validate the optical torque model, lightfoil profiles provided by source [30] were compared. Profiles were seen to agree in terms of shape but disagree in magnitude. For example, the SRP generated at zero α in the micro-prism case deviated by 22% with the prisms in source [29]; conversely, the transverse SRP ratio λ agreed nearly exactly with the optical simulation, registering a deviation of 1% at zero α.

3.3 Pattern Optimisation

The described optical simulation is suitable for assessing the performance and behaviour of sail membranes, but on its own, is not well suited to the *designing* of said membranes. Furthermore, it is preferable to automate pattern optimisation because the solution space — particularly for freeform optical elements — is vast. A wide variety of such optimisation methods exist for optical elements, and these can broadly be categorised according to whether they are analytical, numerical or hybrid in their approach. Broadly speaking, analytical methods tend to produce very exact solutions by framing the optimisation problem in a manner that is solvable. Because it produces solvable problems, analytical optimisation is quick to execute. However, these exact and efficient solutions are for fairly approximate representations of the actual problem. Analytical optimisers operate in the absence of simulation feedback and so must make assumptions as to what constitutes an optimal solution. Furthermore, the state space of the scenario must be simplified or abstracted so as to remove any ambiguous, chaotic or otherwise hard-to-parameterise phenomena.

Conversely, numerical optimisers produce solutions that may be more approximate, but for representations of the problem that are more accurate. A numerical optimiser is always *able* to yield a globally optimal solution — which is not always true for an analytical one — but at the cost of greatly increased computation times. Of course, there is no guarantee of finding a global solution; a numerical optimiser may yield many solutions, but it is difficult to prove that a better one does not exist. Nonetheless, these methodologies offer considerable flexibility to a designer as the parameters that are to be maximised may be changed easily according to various priorities, allowing one to generate various designs for various applications. Lastly, a numerical optimiser does not require the scenario or system to be exhaustively understood; they are generally functional in the presence of ambiguity and may be compatible with certain *black-box problems*. Due

to its synergy with the optical simulation, numerical optimisation is adopted for the pattern optimiser.

The developed pattern optimiser is a simple AI that numerically optimises sail membranes. It interacts with the optical simulation to generate optimised optical membranes through an iterative process wherein membrane geometry is adjusted between simulation runs and membrane efficiency is determined using SRP profiles and a user-defined fitness function f_f. Because this process can be computationally expensive, the pattern optimiser operates according to reinforcement learning principles based on Q-learning [109]. This greatly reduces the number of simulations that are needed to converge upon an optimal design.

3.3.1 Actions, Fitness and Reward

The pattern optimiser uses *actions* A to shape patterns, *fitness functions* f_f to evaluate the success of actions, and *reward functions* f_r to shape optimiser behaviour accordingly. An action is any change that the agent makes to the simulated particle. For the purposes of this thesis, this involves changing its geometry, but it can just as easily involve changing material properties or particle orientation. If an action is successful, it is made permanent. Each time that an action is accepted, a new *iteration* of the design is made, denoted by the *iteration number* $\mathrm{n_I}$. Typically an optimisation run will end after a user-defined number of actions are accepted, called the *iteration limit* $\mathrm{n_{I\,max}}$.

The fitness function f_f is defined by the user to describe what constitutes an optimal design. In the case of sail patterns, f_f is usually a function of SRP or torque per unit area for a certain α or range of α, but could just as easily include parameters that do not pertain to performance, such as thickness. Conversely, the reward function f_r grants rewards at the end of each simulation according to the relative success of a given action, and it is these that the optimiser seeks to

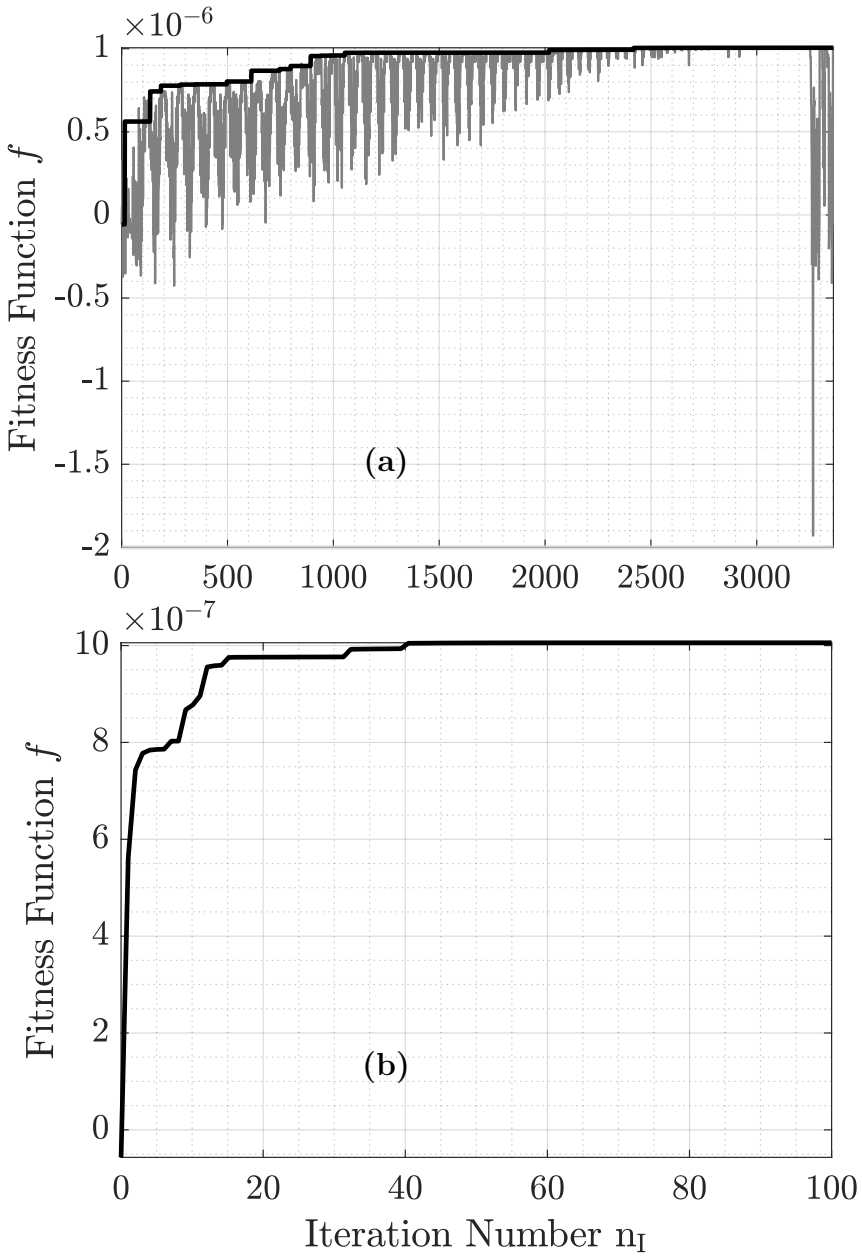

Figure 3.13: Example of Fitness Curves by (a) Simulation Number n_S, (b) Iteration Number n_I

maximise. f_r is simply the fitness after an action relative to the fitness of the previous iteration. By attempting to maximise this reward, the optimiser will naturally converge upon an optimal solution. For action A and iteration n_I, the reward is calculated by Eq. 3.30.

$$f_r(n_I, A) = f_f(n_I, A) - f_f(n_I\text{-}1, A^*(n_I\text{-}1)) \tag{3.30}$$

Wherein A^* denotes a successful action that was made permanent, and that is

defined by Eq 3.31:

$$f_r(\mathrm{n_I}, A^*) = f_r^{\mathrm{MAX}}(\mathrm{n_I}, A) \qquad \text{if } f_r^{\mathrm{MAX}}(\mathrm{n_I}, A) > 0 \tag{3.31}$$

The reward of each action can be positive, null, or negative, but an action will only be accepted if it is positive. f_r is stored according to $\mathrm{n_I}$ and A because it allows the optimiser to make predictions as to which action is most likely to be successful in the future. It does this through a *discounted* variant of the reward function $f_r^*(A)$ which takes the sum of f_r for a given action, but with the rewards discounted according to how many iterations have passed since the action that yielded it was taken (Eq. 3.32).

$$f_f^*(A) = \sum_{i=1}^{\mathrm{n_I}-1} \frac{1}{i} f_f(\mathrm{n_I}\text{-}i, A) \tag{3.32}$$

This is used to influence the behaviour of the optimiser and is the basis of the reinforcement learning process. The probability that it will attempt a given action is elevated if that action was successful *recently*. Similarly, actions that were historically unsuccessful become more likely to be attempted again after a large number of iterations have passed — as it is more likely that the conditions will have become conducive to its success after the state space has changed significantly. Notably, this form of reinforcement learning is *model-free* because the AI learns exclusively from data generated by the simulation, which is itself treated as a black-box. This is appropriate given that certain rays may have chaotic paths through a particle, particularly those that comprise a complex pattern.

Additionally, the optimiser mitigates the issue of 'overshoot' by adjusting the magnitude of actions taken dynamically. As the optimiser converges upon a solution, the $df_f/d\mathrm{n_I}$ will typically decrease, and this is monitored to allow it to preemptively adjust the magnitude of the responsible action accordingly. Of course, $df_f/d\mathrm{n_I}$ is not directly proportional to said magnitude — particularly

for chaotic systems — and so any attempt to prevent overshoot is only a best guess, and one that will ultimately fail. However, once that action has (for the time being) been exhausted, the negative reward that it yields will encourage the optimiser to explore other actions more frequently. In the example depicted by Fig. 3.13a-b, this ebb-and-flow may be observed: the fitness curves are initially steep, as large actions are initially tenable, and yield proportionally large gains in fitness. As these actions begin to yield less reward, the curve levels out. Eventually, an alternative optimisation route is found and the curve becomes steep again. This repeats with dwindling efficacy until only very small actions yield positive rewards. For example, in Fig. 3.13b it can be seen that there is a significant difference in f_f between $\mathrm{n_I} = 0$ and $\mathrm{n_I} = 40$, but practically no difference between $\mathrm{n_I} = 40$ and $\mathrm{n_I} = 100$.

3.3.2 Exploration and Exploitation

Because the optimiser treats the simulation as a black-box, it cannot make meaningful decisions until it has sufficiently built up the discounted reward function f_r^*. As such, optimisation is split into two distinct phases wherein the first phase is dedicated to *exploration* and the second is dedicated to *exploitation*.

During the exploration phase, the optimiser attempts every action successively and records the reward. After every action has been attempted, the action with the highest f_r is made permanent and the particle is saved as a new iteration. The action with the lowest f_r is barred from being attempted for the remainder of the phase. This process repeats until all actions have been taken or until all actions that have not been barred have yielded zero or negative reward. This phase is distinctly visible in Fig. 3.14 as the orderly, curved region on the left side of the diagram. During the exploitation phase, the optimiser is allowed to choose which actions it takes and any action with a positive f_r is made permanent. This phase is depicted in Fig. 3.14 by the more prominent chaotic region, in

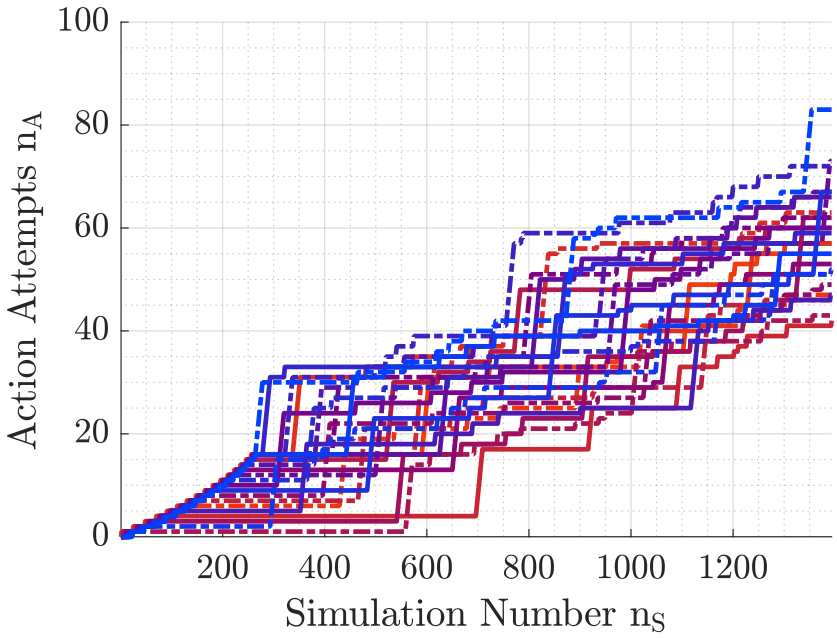

Figure 3.14: Action Space of a Freeform Optical Element Represented by the Number of Times that a Positive (Blue) Action or Negative (Red) Action was Taken n_A with respect to the Number of Simulations n_S, Showing Two Distinct Optimisation Phases

which a variety of actions are attempted. As for a *Markov decision process* [110], the probability that an action will be taken is proportional to the f_r^* that each action has accrued. This phase is more efficient than the first phase, but is less exhaustive. Conversely, if the optimiser consisted entirely of an exploration phase, it would always converge upon the same solution. Given that numerical optimisers are prone to converging upon *local maxima* irrespective of other, better solutions existing, this would negate the possibility of finding them through successive optimisation runs.

3.3.3 Particle Parameterisation

The state space may be very large for these simulations if exhaustively described, but as far as the optimiser is concerned, it may be simplified to α, plus the state space of the particle. The size of the particle state space depends on whether it has a *structured* or *freeform* shape. For example, if the particle is a structured triangular prism, its state space will be just four parameters, of which three are

control parameters that the optimiser may interact with. Namely, the control parameters are the three dimensions needed to describe a triangle and the final non-control parameter is the depth of the 3D shape. Other structured shapes, such as cuboids and cylinders, may be described in a similarly concise manner. Conversely, a freeform optical element that is built from a polynomial may be described by dozens of control parameters, representing dozens of spline control points needed to build a freeform 2D profile. Alternatively, the optimiser may be allowed to edit the output vertices directly (rather than the control points) for even greater control, at the cost of a much larger state space.

During optimisation, the action space is twice the size of the particle state space for structured particles — as both the negative or positive increment of one of these states is attributed its own action. Furthermore, the action space is four times the size of the particle state space for freeform particles — as there may be a negative or positive increment in either axis of the 2D cross-section profile.

3.3.4 Drawbacks

A drawback of an optimiser that has the ability to make decisions is that the reasoning behind those decisions is not always easy to decipher. An increase in performance may be observed, but deductive reasoning is often required to understand why it has converged upon a given solution. This is particularly true of simulations involving many rays and multiple particles.

Furthermore, certain optimisation problems can have multiple near-optimal solutions, for which transmissive sail membranes are no exception. As already mentioned, an AI can become convinced that a near-optimal solution is the true optimal. This occurs because the extent to which an action is successful or unsuccessful is dependent upon the current state space. Typically, when the state space is in a near-optimal state due to it having found a local reward maximum,

the optimiser may not be willing to leave that near-optimal state because all further actions yields zero or negative reward. This could be because a sequence of different actions would be needed to access the positive reward of an alternative maximum, or it could be because the magnitude of individual actions taken are currently too small to reach said alternative maximum. Therefore the outcome of the optimisation process can be highly dependent upon the state of the object at the start of the process – in this case, the user-input geometry of the particle.

Mitigating strategies are employed to give the optimiser the ability to escape a local maximum in case a better solution lies nearby. Firstly, it is important to verify that the maximum — whether global or local — has been accurately reached. To do this, the optimiser will temporarily revert to the exploration phase once all positive actions appear to have been exhausted. Simultaneously, all actions are greatly reduced in magnitude to prevent overshoot. If this fails also, a maximum must have been found to a high degree of accuracy. The optimiser then attempts to search for other maxima by substantially increasing the magnitude of its actions and executing a random walk. This *partially* mitigates the issue of local maxima but does not eliminate it; these mitigating strategies are predicated upon the presence of maxima that are near enough to be found by these exploratory actions, which is not always the case. It is also predicated upon said alternative maxima being reachable by a single large action, rather than a sequence of different actions; allowing sequences of actions to be attempted that yield initially negative reward was deemed too computationally expensive for the existing model, as this would have effectively increased the size of the action space by several orders of magnitude. However, this may be made feasible by the employment of more advanced, model-based AI, ray tracing optimisations to reduce simulation time, or very powerful hardware.

3.4 Experiments and Prototypes

The aim of the manufacturing segment of this project was to map the processes surrounding the development of a novel refractive or diffractive sailcraft. This was achieved by collating the relevant literature, evaluating the suitability of different methods for transmissive sail fabrication, and by performing *in situ* experiments — the latter of which engendered a significant amount of trial-and-error. In chapter 6, this work is represented in a structured manner. In truth, this was an organic process, and dead-ends and new avenues for research were discovered along the way in equal measure. The breadth of this work led to collaboration with a number of UoN faculties and research groups. Within the Faculty of Engineering, this was principally via the *Nottingham Geospatial Institute* (NGI) and the *Centre for Additive Manufacturing* (CFaM). However, this work was also aided by the Faculty of Science via the *School of Physics & Astronomy* and the *School of Chemistry*, as well as interdisciplinary research centres comprising the *Biodiscovery Institute* (BDI) and, in particular, the *Nanoscale and Microscale Research Centre* (nmRC). This is stated to highlight the fact that expertise and infrastructure belonging to a diverse set of academics had to be solicited (and graciously given) in order to gain a holistic understanding of transmissive sail manufacture and to execute many of these processes. Given the breadth of this work, it would have been immensely helpful at project outset to have had access to a piece of literature that detailed the relevant concepts, processes, and the associated common pitfalls. In the absence of this coveted resource, it was decided that chapter 6 should be made into said resource.

As well as mapping and trialling the manufacturing processes involved in transmissive sail development, these processes were evaluated based on how feasible and cost effective they are as an in-house investment (procuring the relevant infrastructure and materials for in-house processing) versus how cost effective they are as a commercial solution (outsourcing the process). During evaluations,

in-house process feasibility was typically assessed according to the resolution, repeatability and scalability of the relevant fabrication methods. These evaluations were supplemented by experience gained from performing these processes *in situ*. However, the wide scope of this project ensured that no single process was able to be explored exhaustively. Conversely, the most cost-effective route was determined by soliciting quotations for the cost of the relevant infrastructure (and where substantial, the cost of raw materials) versus the cost of pre-processed parts and outsourced processes. For example, it was found that thin film fabrication as a process was absolutely feasible to carry out in-house with relatively little in the way of infrastructure, but provided too small a cost advantage over the commercial solution to be optimal. This information is particularly pertinent to academic and small-to-medium sized business (SMB) satellite developers that are seeking to develop a sailcraft, for which time and capital must be very carefully allocated in order to bring a project to fruition.

The secondary aim of the manufacturing segment of this project was to develop a sail prototype that would serve as a precursor to an in-orbit demonstration (IOD) mission in collaboration with the *NottSpace* CubeSat Programme, NGI and UoN Faculty of Engineering. A refractive solar sail and deployment system were designed as a payload for the UoN CubeSat *JamSail* to this end. This serves as a design case study at the conclusion of chapter 6, and is followed by an account of the future work that will be carried out to bring the payload to maturity.

Chapter 4

Orbital Dynamics

4.1 Analysis of Transmissive Solar Sail Proposals

This section details and compares the existing transmissive sail proposals. To facilitate their evaluation, this begins with the defining of the parameters by which they may be initially compared, with a focus on their performance under ideal conditions. Analysis of these sails under non-ideal conditions is subsequently carried out in section 4.2.

4.1.1 Scalar Performance Parameters

Eq. 4.1-4.2 characterises η_S and η_{S_v}, which are the percentage incident solar irradiance that is converted to SRP and *transverse* SRP, respectively. Eq. 4.3 describes λ, which is the the transverse SRP ratio. Both η_{S_v} and λ have been expressed before in some form, but are reiterated here for clarity. Note that all parameters that are sensitive to flight regime assume $\gamma = 90°$ at 1 au ($G_{\mathrm{SC}} =$

$1,370\ \mathrm{W/m^2}$), and that $F_S^{\mathrm{MAX}} = 2G_{\mathrm{SC}}/c$:

$$\eta_{S_v} = \frac{F_{S_v}}{F_S^{\mathrm{MAX}}} \tag{4.1}$$

$$\eta_S = \frac{F_S}{F_S^{\mathrm{MAX}}} \tag{4.2}$$

$$\lambda = \cos(\beta^*) = \frac{F_{S_v}}{F_S} \tag{4.3}$$

Since idealised reflective sails are used as the baseline for comparison, it is convenient to express the efficiencies of transmissive sails not in terms of percentage irradiance converted, but in proportion to the theoretical maximum efficiencies of reflective ones. Efficiencies that are not absolute, but are instead expressed as a percentage relative to this reflective baseline, are denoted by an asterisk. These relative efficiencies η_S^* and $\eta_{S_v}^*$ are expressed by Eq. 4.4-4.5. Note that η_S^* is expressed only to highlight that its relative and absolute forms are indistinguishable, because the F_S^{MAX} that a reflective sail can achieve is theoretically 100% of $2G_{\mathrm{SC}}/c$:

$$\eta_S^* = \frac{F_S}{F_S^{\mathrm{MAX}}} = \frac{F_S}{(\frac{2G_{\mathrm{SC}}}{c})} = \eta_S \tag{4.4}$$

$$\eta_{S_v}^* = \frac{F_{S_v}}{F_{S_v}^{\mathrm{MAX}}} \tag{4.5}$$

4.1.2 Transmissive Sail Proposals

Developments in the field of study of refractive and diffractive solar sails have lead to myriad designs. The performance data of these designs has been collated and converted to use the scalar performance parameters described in section 4.1.1; these are tabulated in Table 4.1.

These designs can be roughly partitioned by performance and conformity with

Table 4.1: Scalar Performance Parameters of Solar Sails at $\gamma = 90°$, Conversion of Approximate Data Transcribed or Extracted from Figures of Tabulated Sources

Category	Mechanism	α (°)	η_{S_v} (%)	$\eta^*_{S_v}$ (%)	η_S (%)	λ
Reflective	Specular Reflection	35.26	38.49	100	81.65	0.471
		0	0	0	100	0
Diffractive	Littrow Reflection [31]	0	0	0	100	0
	Littrow Transmission [31]	35.26	39.00	101.33	47.43	0.822
	Sun-facing Transmission [32]	20	24.73	64.24	27.17	0.910
		0	21.50	55.86	23.12	0.930
	Bi-grating Beam Rider [57]	0	28.00	72.75	100	0.280
	Liquid Crystal [36]	0	25.00	64.95	93.41	0.268
	Prism Grating [33]	0	43.50	113.02	50.42	0.863
Refractive	Lightfoil (50° Rotation)[30, 37]	0	50.00	129.90	64.03	0.781
	Prism Array [29]	0	20.16	52.39	24.70	0.816
	Gradient-index Waveguide [35]	0	45.14	117.29	58.12	0.777

the transmissive sail 'archetype' that was described in section 2.1.5; they may be described as high performance or low performance, and archetypal or non-archetypal. A sail is said to be high performance if it demonstrate greater transverse acceleration than the reflective baseline, i.e. $\eta^*_{S_v} > 100\%$, while archetypal ones are said to demonstrate an SRP ratio of $\lambda > 0.5$. Notably, all high performance sails are archetypal, but not all low performance sails are non-archetypal, and so these sails fall into one of three categories: high performance (archetypes), low performance archetypes, and low performance non-archetypes. Hereafter, these categories are referred to as Type A, B or C, respectively.

High performance Type A designs include the refractive rotated lightfoil [30], refractive gradient-index waveguide [35], diffractive Littrow transmission [31] and diffractive prism grating [33] sails. Sails within this category tend to outperform reflective sails in every flight regime (as demonstrated in section 4.2). Type A sails also tend to be metasails. The rotated lightfoil sail is an exception, being a non-metasail with very high performance, and that could reasonably be fabricated via double-sided nanoimprint lithography processes (non-rotated lightfoils may be fabricated with conventional nanoimprinting, but such a sail would generate

zero net SRP at $\alpha = 0°$).

Conversely, the Sun-facing transmission [32] variant is a metasail design that belongs to Type B: it is lower performance but still archetypal (and in this case, notable for featuring the highest λ of any design). Type B sails — of which the refractive prism array [29] is the only other example — tend to outperform reflective sails only within a narrow band of flight regimes (see section 4.2).

Lower performance, non-archetypal, Type C designs include the diffractive Littrow reflection [31], bi-grating [57] and liquid crystal [36] sails. The former is not truly transmissive, and has superficial behavioural similarities to a reflective sail. However, it diffracts sunlight in such a way as to gain no benefit from tacking; its SRP vector is always locked to $-\hat{\mathbf{S}}$ (which for $\gamma = 90°$ is the longitudinal axis $\hat{\mathbf{y}}_\mathbf{p}$). This behaviour is a hindrance here, but may have application within artificial Lagrange points or certain laser-driven sail configurations. The latter two are notable for generating a significant component of SRP in the sail normal $\hat{\mathbf{x}}_\mathbf{b}$ (indicated by their high η_S but low λ). This sometimes plays to their advantage in later analyses (see section 4.2) as, paired with a suitable steering law, they may exhibit behaviours similar to either the reflective or transmissive sail archetype, depending on which mode is most beneficial at the time. In general, they tend to outperform Type B transmissive sails, match or outperform reflective sails, and be outperformed by Type A transmissive sails when these are compared within the FM.

Finally, it should be noted that each sail proposal applies different assumptions, simplifications and modelling techniques. Generally, sail performance will therefore be skewed in favour of designs with less detailed, more idealised source models. Furthermore, many proposals offer a variety of possible designs. In these cases, the data transcribed represents only one of a number of possibilities. Furthermore, while scalar performance parameters allow for easy 'at a glance' comparisons of sail designs, they do not account for the potential benefits of simplified

steering and, in some cases, active control of λ that may be achieved by certain designs [29, 31, 36]. These initial evaluations also pertain only to a single ($\Gamma = 0°$) flight regime.

4.2 Simulation and Evaluation

4.2.1 Instantaneous Performance Profiles in LEO

Type	Mechanism	α (°)	$\eta^*_{S_v}$ (%)	$\dot{\epsilon}_{500\text{ km}}$ (J/kg)	$\dot{\epsilon}_{1{,}000\text{ km}}$ (J/kg)	h_{MIN} (km)	h_{break} (km)
-	Specular Reflection	$\alpha(h)$	$\eta^*_{S_v}(\alpha)$	-0.004	0.007	568	-
		35.26	100	-0.028	0.007	602	-
		0	0	-0.004	0	-	-
A	Lightfoil (50° Rotation)	0	129.90	0.006	0.009	446	∞
	Gradient-index Waveguide	0	117.29	0.005	0.008	452	∞
	Prism Grating	0	113.02	0.004	0.008	454	∞
B	Sun-facing Transmission	0	55.86	0	0.004	499	635
	Prism Array	0	52.39	0	0.004	503	628
C	Bi-grating Beam Rider	0	72.75	0.001	0.005	482	678
	Liquid Crystal	0	64.95	0.001	0.005	489	656

Table 4.2: LEO Performance Parameters of Idealised Solar Sails in a Typical $\gamma = 90°$ Orbit (Zero-α Steering Law Example)

Fixed-γ, variable-h profiles are generated for $\dot{\epsilon}$-h wherein $\Gamma = 0°, \gamma = S_\gamma = 90°$ (Fig. 4.1a, 4.2a), as well as fixed-h, variable-γ profiles for $\dot{\epsilon}$-γ wherein $h = 550$ km (Fig. 4.1b, 4.2b). Two simulation runs are carried out: in the first, transmissive sails operate with a zero-α steering law (Fig. 4.1, Table 4.2). In the second, these use a locally optimal steering law (Fig. 4.2). During both runs, a reflective solar sail using a locally optimal steering law is used as a baseline.

Analysing the variable-h profiles first (Fig. 4.1a, 4.2a): using both steering laws, every transmissive sail outperforms the baseline at altitudes lower than $h = 630$ km – even those with inferior $\eta^*_{S_v}$. Notably, all transmissive sails could continue to accelerate below 505 km, in contrast to the baseline specular reflector,

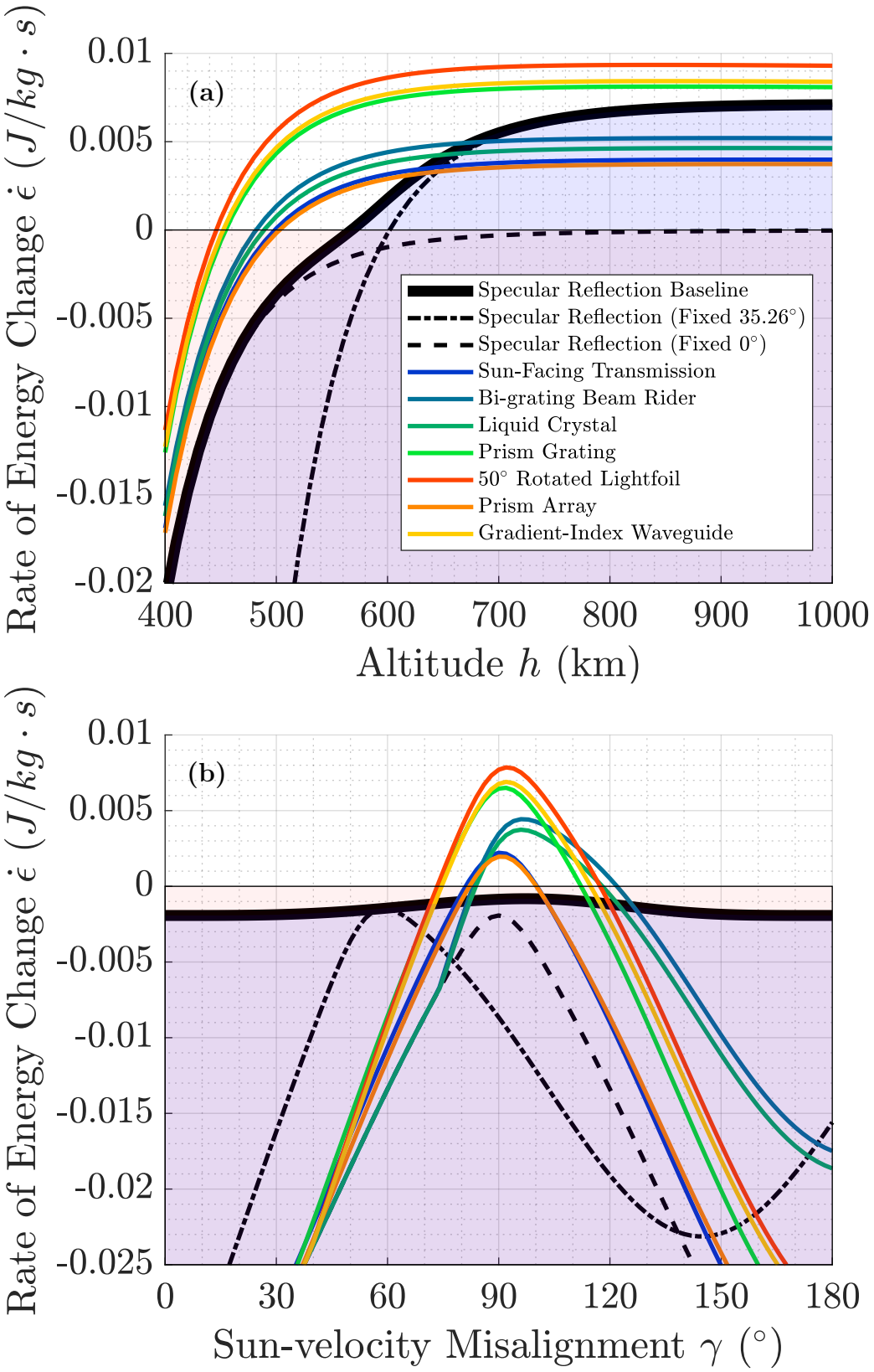

Figure 4.1: Performance Profiles of Transmissive Sails using a zero-α Steering Law, a Reflective Solar Sail using a Locally Optimal Steering Law and Two Fixed Specular Reflectors in Circular LEOs: (a) Fixed $\gamma = 90°$, (b) fixed $h = 550$ km

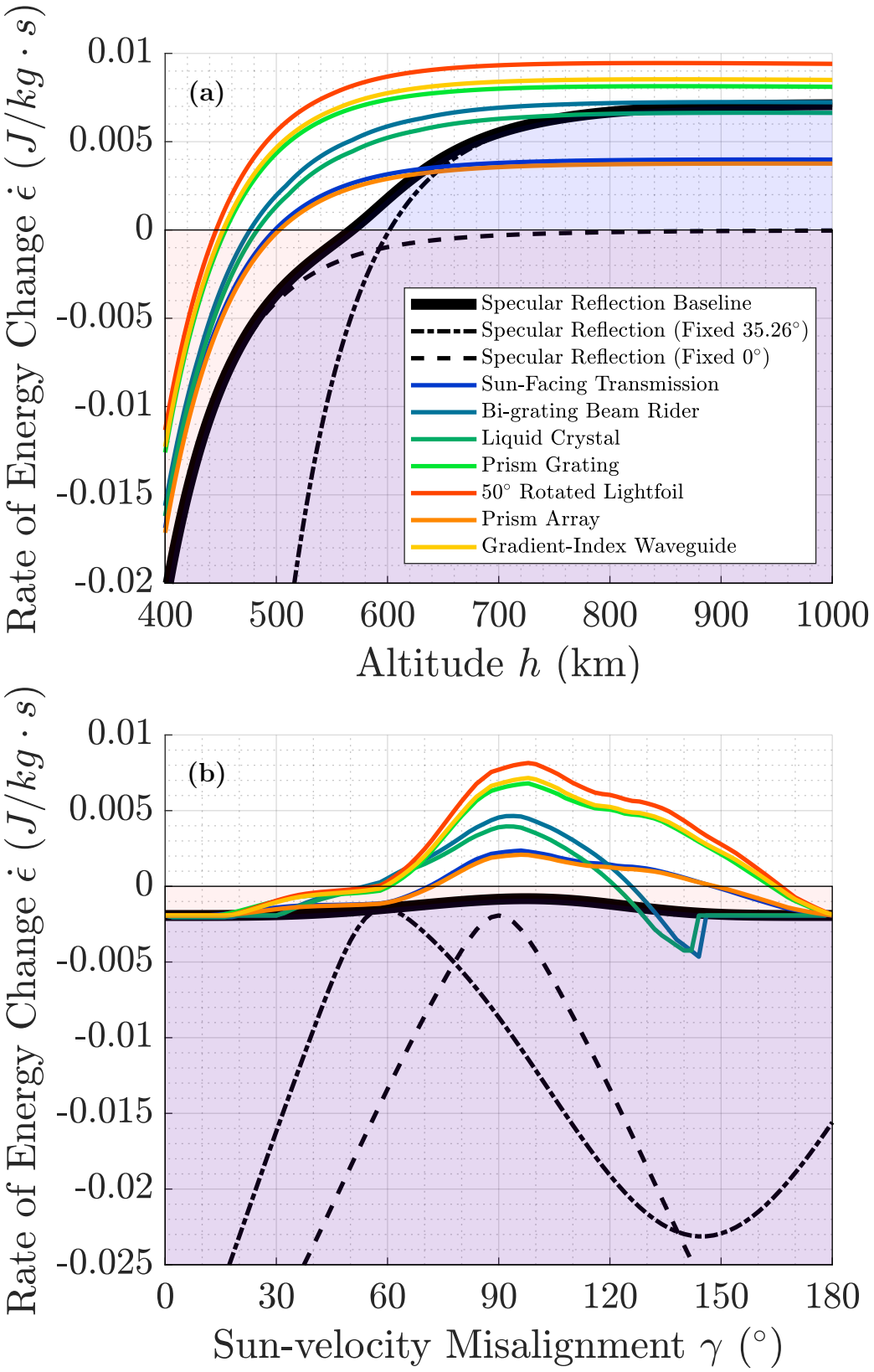

Figure 4.2: Performance Profiles for Solar Sails using a Locally Optimal Steering Law plus Two Fixed Specular Reflectors in Circular LEOs: (a) Fixed $\gamma = 90°$, (b) fixed $h = 550$ km

which could not do so any lower than 570 km. A list of minimum operational altitudes h_{MIN} is tabulated in Table 4.2 for the zero-α law. Notably, even lower performance transmissive sails demonstrate a range of altitudes for which they outperform the baseline reflector. This is described by $0 < h < h_{\text{break}}$. The parameter h_{break} is represented in Fig. 4.2 by the point at which the $\dot{\epsilon}$ curve of a sail intersects with the $\dot{\epsilon}$ curve of the baseline. Transmissive sails that have higher scalar $\eta^*_{S_v}$ outperform the baseline at every altitude, and so their curves never intersect (denoted $h_{\text{break}} = \infty$).

The two variable-h profiles are indistinguishable at lower altitudes (where the two steering laws are practically identical) but differ at high altitude. When a zero-α law is applied (Fig. 4.1a), the relative success of *all* solar sails at $h = 1,000$ km (where drag is negligible) corresponds with the rankings of their scalar $\eta^*_{S_v}$; those with higher $\eta^*_{S_v}$ than the baseline perform better and *vice versa*. The equivalent profile that was generated with a locally optimal steering law (Fig. 4.2a) yields a higher $\dot{\epsilon}$ at $h = 1,000$ km for transmissive sails than the zero-α equivalent. This occurs because the locally optimal steering law places these sails into a slightly off-zero-α attitude which, according to the generic sensitivity model, registers a slightly higher $\eta^*_{S_v}$ than those at zero-α. However, it is only substantial for Type C sails (e.g. bi-grating beam rider). This is because these have a large normal component to SRP and so receive a disproportionate increase in performance at high altitude when adopting a locally optimal steering law, where they benefit greatly from tacking.

A greater discrepancy may be observed between the two variable-γ profiles (Fig. 4.1b, 4.2b): the zero-α transmissive sails experience a significant erosion of performance as their γ deviates from 90°. Conversely, locally optimal transmissive sails demonstrate considerable robustness, and are able to operate at γ far beyond the optimal. Type A sails demonstrated the widest range of tenable γ for both laws. Of the lower performance sails, non-archetypal Type C variants

demonstrated a wider range of tenable γ than their archetypal Type B peers when using a zero-α law. This is likely by merit of their higher $\eta^*_{S_v}$, rather than any nuanced behaviours pertaining to non-archetypes. Indeed, when using a locally optimal steering law, the Type C archetypes were more versatile than Type B non-archetypes, despite having lower $\eta^*_{S_v}$. This demonstrates that archetypal, high-λ behaviour is more conducive at lower altitudes. Conversely, Type B non-archetypes benefited greatly from tacking, and were seen to gain a larger performance increase from a locally optimal steering law than their peers at higher altitudes (e.g. bi-grating beam rider as seen in Fig. 4.2a).

When compared with the baseline reflective sail, transmissive sails are shown to be less sensitive to altitude h by Fig. 4.1a, 4.2a, and are seen to be *more* sensitive to γ by Fig. 4.1b, 4.2b. During the latter two simulations, every transmissive sail nonetheless retained a performance advantage for nearly the entire γ range despite their heightened sensitivity to it. However, this is only by merit of having much higher maximum performance at $h = 550$ km.

4.2.2 Instantaneous, Transient and Stable Minimum Operational Altitudes

To build a more comprehensive picture of the operational flexibility of these sails, minimum operational altitude profiles are generated. Minimum operational altitude is defined as the lowest altitude at which $\dot{\epsilon} > 0$ J/kg $\cdot$ s. Three minimum operational altitudes are defined to represent different timescales over which this may be achieved: *instantaneous* $h_{\min}$ represents the altitude needed to meet this condition over a single instance, depending on local conditions (single γ); *transient* $h_{\min^*}$ pertains to an entire orbital period, depending on the cumulative effect of the range of conditions experienced (variable γ, single S_γ as seen in Fig. 3.1a); *stable* $h_{\min^{**}}$ pertains to an initial orbit propagated to account for

dynamic-Γ effects (variable γ and S_γ as seen in Fig. 3.1b). Hereafter, a locally optimal steering law is employed by all actively steered sails. The *instantaneous*

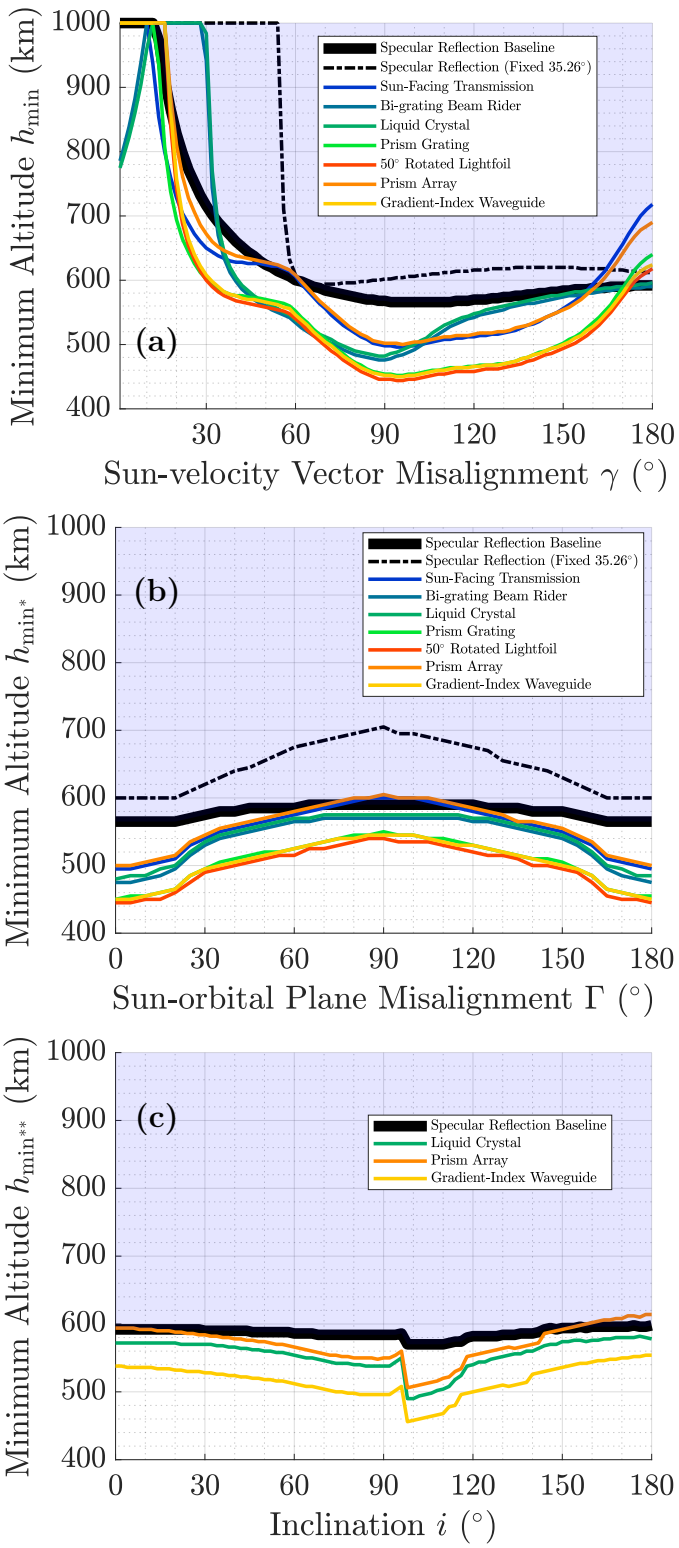

Figure 4.3: Minimum Operational Altitudes: (a) Instantaneous $h_{\min}$ versus γ, (b) Transient $h_{\min^*}$ versus Γ, (c) Stable $h_{\min^{**}}$ versus i

$h_{\min}$-γ profiles (Fig. 4.3a) reveal interesting behaviours through their asymmetry. Firstly, all solar sails struggle to operate effectively when their motion is carrying

them towards a *prograde* Sun (particularly for $0° \leq \gamma \leq 30°$). By a small margin, reflective and Type C transmissive sails are the most disadvantaged by these conditions. Conversely, at its extremity, the *retrograde* Sun regime favours reflective and Type C transmissive sails due to their greater ability to generate velocity-wise SRP under these conditions (enabled by their larger SRP component in the sail normal $\hat{\mathbf{x}}_\mathbf{b}$). The majority of the γ range ($30° \leq \gamma \leq 150°$) is dominated by transmissive sails of all varieties; of these, Type A variants demonstrate the lowest $h_{\min}$ nearly throughout.

The *transient* $h_{\min^*}$-Γ profiles (Fig. 4.3b) reveal how low a solar sail can orbit without said orbit decaying. It is most pertinent to large ISAM sails which, being able to escape Earth's atmosphere rapidly, do not need to consider dynamic-Γ effects. At the optimal extremes ($\Gamma = 0, 180°$), $h_{\min^*}$ is shown to be lowest for transmissive sails (in order of their $\eta^*_{S_v}$ ranking) and highest for reflective sails. At the sub-optimal extreme ($\Gamma = 90°$), transmissive sails retain advantage but by a smaller margin. The exception to this is the Type B transmissive sails; these are shown to struggle within sub-optimal (e.g. near-equatorial) orbits, where they demonstrate a higher $h_{\min^*}$ than even reflective sails. It may also be observed that in general, the $h_{\min^*}$ of these profiles never rise as high as the highest peaks of the $h_{\min}$ profile, but exhibit similar valleys. This is to be expected, as the average $\bar{\gamma}$ is always 90°; constantly optimal orbits (for which $S_\gamma = 90°$) are possible, as depicted by the identical valleys of both profiles. However, it is impossible to have an orbit that is constantly sub-optimal (e.g. no orbits can comprise only the prograde Sun regime). It is also of note that the Γ for which the eclipse fraction f_e becomes non-zero ($\Gamma = 20°$) is identifiable by a sudden increase in $h_{\min^*}$. It is notable that this increase is more significant for transmissive sails than for the reflective baseline, which highlights an interesting behavior: eclipse prevents acceleration from occurring during one of the two segments of an orbit for which $\alpha \approx 0°$ is tenable (the orbital segment near 'perihelion'; furthest from the Sun). These are ideal flight conditions for transmissive sails. Conversely, the

retrograde Sun portion of these orbits marginally favour reflective sails and are never occluded.

The *stable* $h_{\min^{**}}$-i profiles (Fig. 4.3c) reveal how low a solar sail can orbit when said orbit is subject to dynamic-Γ effects. This is most pertinent to modern solar sails which may take weeks or even months to escape Earth's atmosphere. Because propagating hundreds of orbits over several months is computationally expensive, only three transmissive sails were propagated; the liquid crystal, prism array and gradient-index waveguide sails are chosen to represent the Type C, B and A sails, respectively. Considering that $\Gamma = 90 - i$ for *initial* orbits, the $h_{\min^{**}}$ profile largely agrees with $h_{\min^{*}}$, and the relative ranking of these sails is mostly conserved. The asymmetry of this profile and the concentration of valleys around i_{SSO} is predominantly due to the J2 perturbation, the sensitivity of different orbital inclinations to it, and the subsequent effects of a dynamic-Γ (see sections 3.1.2-3.1.3). Furthermore, for these profiles the effect of eclipse is only truly absent at i_{SSO}, and is a significant contributor to the sharp fall in $h_{\min^{**}}$ that may be observed.

4.2.3 Propagated Performance for Typical Manoeuvres

The final simulation explores the performance of these sails as they carry out an orbit-raising manoeuvre, and demonstrates the effects of time-variant S_γ arising from heliocentric Earth and J2 perturbation effects. Sail performance is compared with that of contemporary propulsion systems (model described in section 3.1.5). Sails are assigned areas of $A = 36$ m^2 consistent with a 6 m square sail. Two profiles are generated for manoeuvres spanning up to 24 months, which are discussed below. The first profile (Fig. 4.4) depicts various satellites starting from the same injection orbit and serves to highlight the performance of medium-sized transmissive sails relative to (approximated) contemporary propulsion systems. All satellites begin from a $h = 530$ km, circular sunrise-sunset polar

orbit ($i = 90°, \Gamma = 0°$). The $t_{h_1 \to h_2}$ for each sail and thruster is visible for any $h_1 \geq 530$ km, $h_2 \leq 6,000$ km. As expected, the reflective sails failed to perform

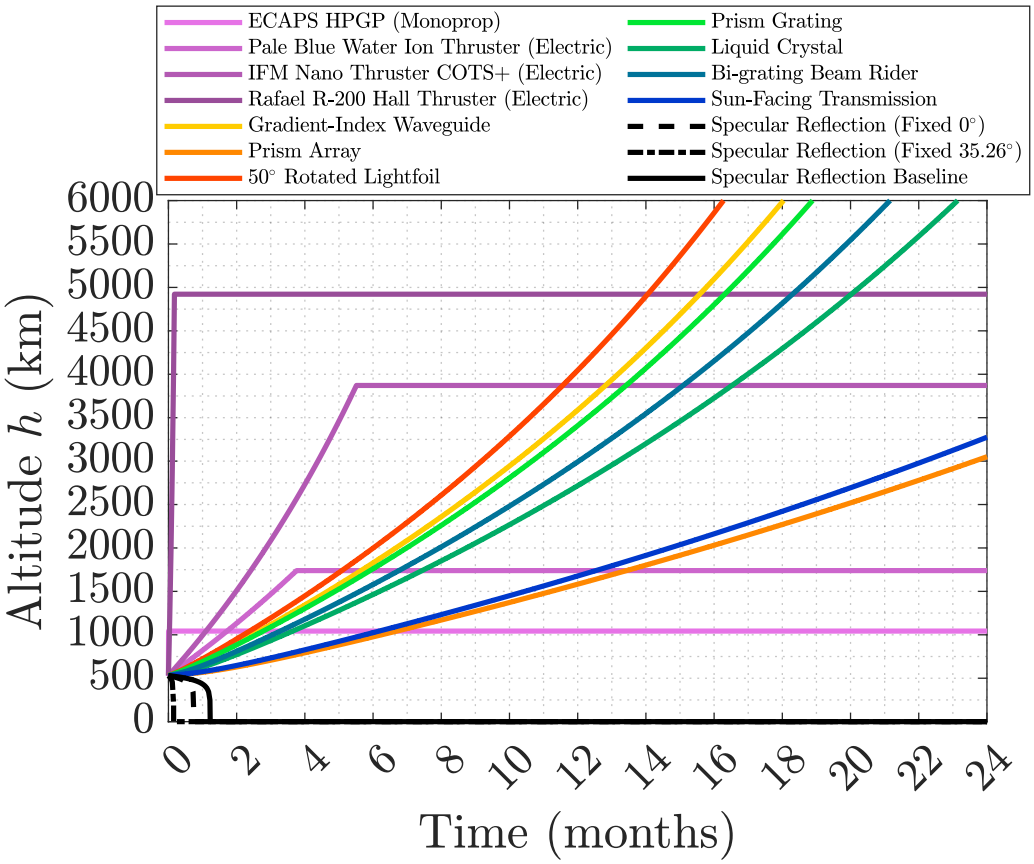

Figure 4.4: Orbit-raising for Thrusters and $A = 36$ m^2 Solar Sails from a $h = 530$ km, Circular Sunrise-sunset Polar Orbit ($\Gamma = 0°$)

orbit-raising from such a low altitude, which is consistent with the results of Fig. 4.3. The $t_{530 \to 6000\text{ km}}$ of the Type A and C transmissive sails only differed by around 20%, while Type A and B differed by nearly 100%. Relative to thrusters, the $t_{530 \to h_2}$ of these sails were greater at any h_2 for which the thrusters had not run out of fuel; as expected, larger sails would be needed to compete in terms of raw acceleration. However, the rates of acceleration for Type A and B sails were comparable to that of the mid-range electric thrusters (Pale Blue Water Ion [104] and IFM Nano Thruster [102]), though significantly outclassed by Hall effect [101] and cold gas thrusters [103] (these happen to be the highest and lowest ΔV thrusters, respectively). By merit of not requiring propellant, Type A and C transmissive sails exceeded the h_2 and ΔV of the cold gas and Hall effect

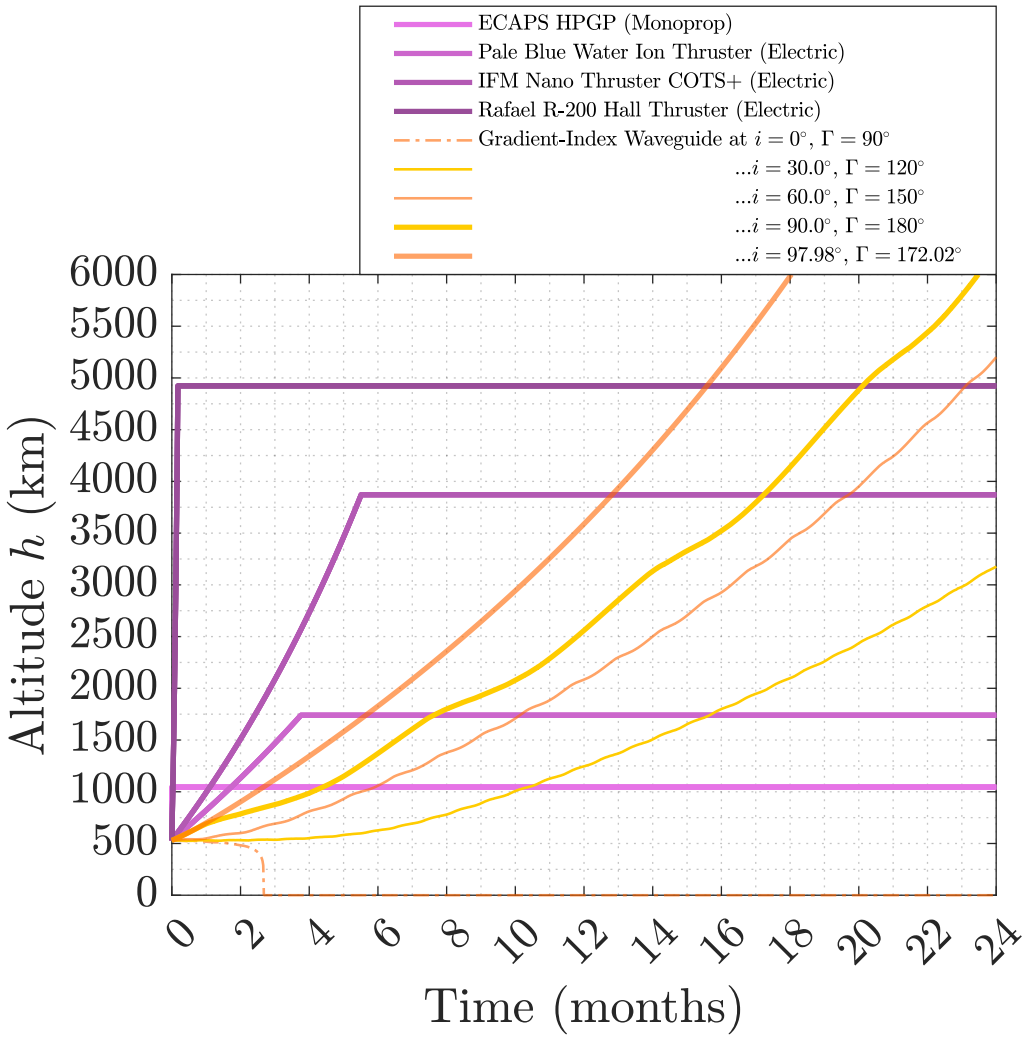

Figure 4.5: Orbit-raising for Thrusters and an $A = 36$ m^2 Transmissive Solar Sail from a $h = 530$ km at Various Inclinations i

thrusters within 2.5-3.5 months and 14-20 months, respectively. If one assumes that these sail may be fabricated in an economically viable manner akin to that of a cold gas thruster, these results are encouraging, and suggest that such a sail may be favourable for many applications in both the short and long term (see section 4.2.5).

The second profile focuses on a single Type A sailcraft injected into a $h = 530$ km circular orbit (Fig. 4.5), but from various inclinations i and Sun-velocity misalignment Γ at injection. This profile serves to explore the operational flexibility of transmissive sails. Naturally, thruster satellite performance is not influenced by i or Γ. As expected, injection into an $i = 90°$ sunrise-sunset polar orbit is not truly the most efficient for solar sailing; the $i = 97.98°$ dusk-dawn SSO yields

substantially higher $\mathrm{d}h/\mathrm{d}t$ because dynamic-Γ effects are mitigated. Initially, the manoeuvre from the sunrise-sunset polar orbit *is* faster than the manoeuvre from a dusk-dawn SSO. However, J2 and heliocentric Earth effects cause the orbit to precess from sunrise-sunset to noon-midnight, which implies a less optimal S_γ and a higher eclipse fraction f_e. Eventually, this orbit returns to sunrise-sunset — and so on — leading to the 'wobbly' $\mathrm{d}h/\mathrm{d}t$ of the manoeuvre. It can be seen that at $i = 0°$, the sailcraft is unable to perform the manoeuvre from this altitude. These results demonstrate that, although the operational flexibility of these sails are considerably improved, they are inherently less flexible than modern thrusters.

4.2.4 Orbital Simulation Overview

Relative to reflective solar sails, transmissive sails demonstrate lower sensitivity of their performance metrics to h, but greater sensitivity to i throughout; all solar sails prefer SSO and polar orbits, but this proclivity is pronounced in transmissive sails (particularly archetypal Type A and B variants). Type A and C variants generally retained advantage even within less favourable flight regimes despite this, by merit of either (A) their much higher peak performance or (C) flexibility of steering. For Type B sails, advantage over reflective sails was typically only found in low h, near-polar i ($i \approx 90°$) flight regimes. Furthermore, the (stable) minimum altitude advantage of transmissive solar sails over reflective ones was shown to be less substantial in non-polar and non-SSO orbits.

Relative to conventional propulsion systems, the 36 m^2 transmissive sails that were used in the case study were shown to be capable of comparable rates of acceleration to mid-range electric thrusters, but were greatly outclassed in this respect by cold gas and Hall effect thrusters. Unlike thrusters, the $t_{h_1 \to h_2}$ of solar sails was also shown to be highly sensitive to i. Naturally, all thrusters were outclassed by all transmissive sails in terms of maximum h_2 and ΔV.

4.2.5 Potential Applications

It has been shown that transmissive sails operating in LEO may compete with contemporary propulsion systems if they are suitably sized, which implies that they may carry out a multitude of different LEO satellite missions. However, economics also plays a significant role in the choosing of satellite systems, and these sails may yet be relegated to niche applications if the assumption that they hold an economic advantage over other systems is proven to be invalid. While this is a safe assumption for some transmissive sail designs, it is more speculative for others (e.g. certain metasails).

Generally, transmissive sails may be the optimal solution for missions that entail high ΔV in LEO and the inner Solar System and, more speculatively, missions that seek to minimise cost. Missions with the greatest compatibility with transmissive sails include long-duration missions, high altitude missions, and low altitude missions in SSO. Missions with the least compatibility include those that require rapid manoeuvring, those that require operation at very low altitude, or those that begin from low inclination orbits.

In the short term, Type B or C transmissive sails composed of conventional materials could provide a low-cost solution to propulsion for small satellite developers operating in LEO and beyond. Amateur developers — who traditionally use cold gas thrusters or none — could gain access to an inexpensive, high-ΔV system, perhaps enabling amateur space probes or similar. Their simplified steering may also endear them for operation in highly atypical orbits, such as those involving 'artificial Lagrange points'. However, lacking the flexibility of the high performance variants, their operation in low LEO may be relegated to SSO and near-polar inclinations.

In the medium term, Type A or C transmissive sails could be used for a variety of contemporary applications in LEO in a similar role to that of the electric thruster.

This would enable greater longevity at presumably reduced cost, with the trade-off being an incompatibility with certain LEO flight regimes. Their pseudo-indefinite acceleration may also be exploited for the performing of debris mitigation missions, as a transmissive sailcraft could complete multiple rendezvous-and-ferry trips between debris clouds and re-entry altitudes (albeit likely paired with a small thruster for fine-control at rendezvous).

In the long term, it is possible that sails of arbitrary size may be fabricated in space, thereby increasing acceleration to an arbitrary degree. Perhaps in an ISAM future, this could even be done economically. A large sail (or multiple small sails) could be used to ferry manned interplanetary spacecraft, or even tow mineral-rich asteroids into near-Earth orbit for exploitation. Advantage may even be found in the outer Solar System or interstellar space through the use of a 'sun-diving manoeuvre', beam-propulsion configurations, or sails of tremendous size.

Chapter 5

Membrane Optics

This section details the exploration of the solution space for refractive sail pattern designs using numerical optimisation in tandem with a ray tracing optical simulation that generates profiles for SRP and torque per unit area with respect to α.

The methods employed in the development of this simulation were explored in section 3.2. However, because this optical simulation is reasonably complex and would be difficult to replicate by referring to the methodology alone, this chapter begins with a detailed description of the simulation sequence.

5.1 Optical Simulation Sequence

In this chapter, axes $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$ refer to the Cartesian reference frame of the simulation; as before, $\{\hat{\mathbf{x_B}}, \hat{\mathbf{y_B}}, \hat{\mathbf{z_B}}\}$ is the body-fixed reference frame of a solar sail and $\{\hat{\mathbf{x_b}}, \hat{\mathbf{y_b}}\}$ is the simplified, 2D body-fixed reference frame that has been rotated onto the solar-primer (see section 2.1.2). In practice, $\hat{\mathbf{x}}_\mathbf{B} = \hat{\mathbf{x}}_\mathbf{b}$ and $\hat{\mathbf{y}}_\mathbf{B} = \hat{\mathbf{y}}_\mathbf{b}$ for this chapter and the two are used interchangeably. Furthermore, because the lightsource is rotated rather than the particle (see section 3.2.6), these axes do

not change, and have a constant relationship with the Cartesian reference frame — in this case, $\hat{\mathbf{x}}_\mathbf{B} = \hat{\mathbf{z}}$, $\hat{\mathbf{y}}_\mathbf{B} = \hat{\mathbf{x}}$ and $\hat{\mathbf{z}}_\mathbf{B} = -\hat{\mathbf{y}}$. Finally, the subscripts P and LS denote a particle or lightsource, respectively, and an `alternative font` is used to denote code. This optical simulation was created in MATLAB, but could be recreated in any Turing complete language. The simulation procedure is depicted graphically by Fig. 5.1 and each step is described below:

1. **Build the lightsource and particles.** These are built as *structs* and are referred to collectively as `entities`. These entities are attributed the following *fields* (user-inputs are denoted by an asterisk*; the rest should be calculated automatically):

   i. `type*` (double) – 0 for lightsource, 1 for a particle

   ii. `position*` (1×3 column vector) – placement of the particle or lightsource in 3D space

   iii. `rotation*` (1×3 column vector) – roll ϕ, pitch θ and yaw ψ

   iv. `axes` (3×3 matrices) – rows for $\hat{\mathbf{x}}_\mathbf{B}$, $\hat{\mathbf{y}}_\mathbf{B}$ and $\hat{\mathbf{z}}_\mathbf{B}$ (if the entity is a lightsource, $\hat{\mathbf{x}}_{\mathbf{B(LS)}} = -\hat{\mathbf{S}}$: the direction of ray casting)

   As mentioned, there are two kinds of entity in this simulation: particles (`entity(i).type = 1`) and lightsources (`entity(i).type = 0`). The following fields are included within the struct, but are only used by particles:

   i. `centre_of_mass` (1×3 column vector) - centre of mass position vector

   ii. `shape*` (string) – the geometry type, e.g. 'prism' for structured prism or 'gen' for generative, freeform shape

   iii. `dimensions*` (1×4 column vector) – dimensions of the shape; the usage of columns 1-3 varies by shape, but column 4 is always depth

   iv. `repetition*` (1×3 column vector) – used to form a pattern or shape array by repeating the base geometry in $\hat{\mathbf{x}}_\mathbf{B}$, $\hat{\mathbf{y}}_\mathbf{B}$ or $\hat{\mathbf{z}}_\mathbf{B}$ (but typically only in $\hat{\mathbf{y}}_\mathbf{B}$)

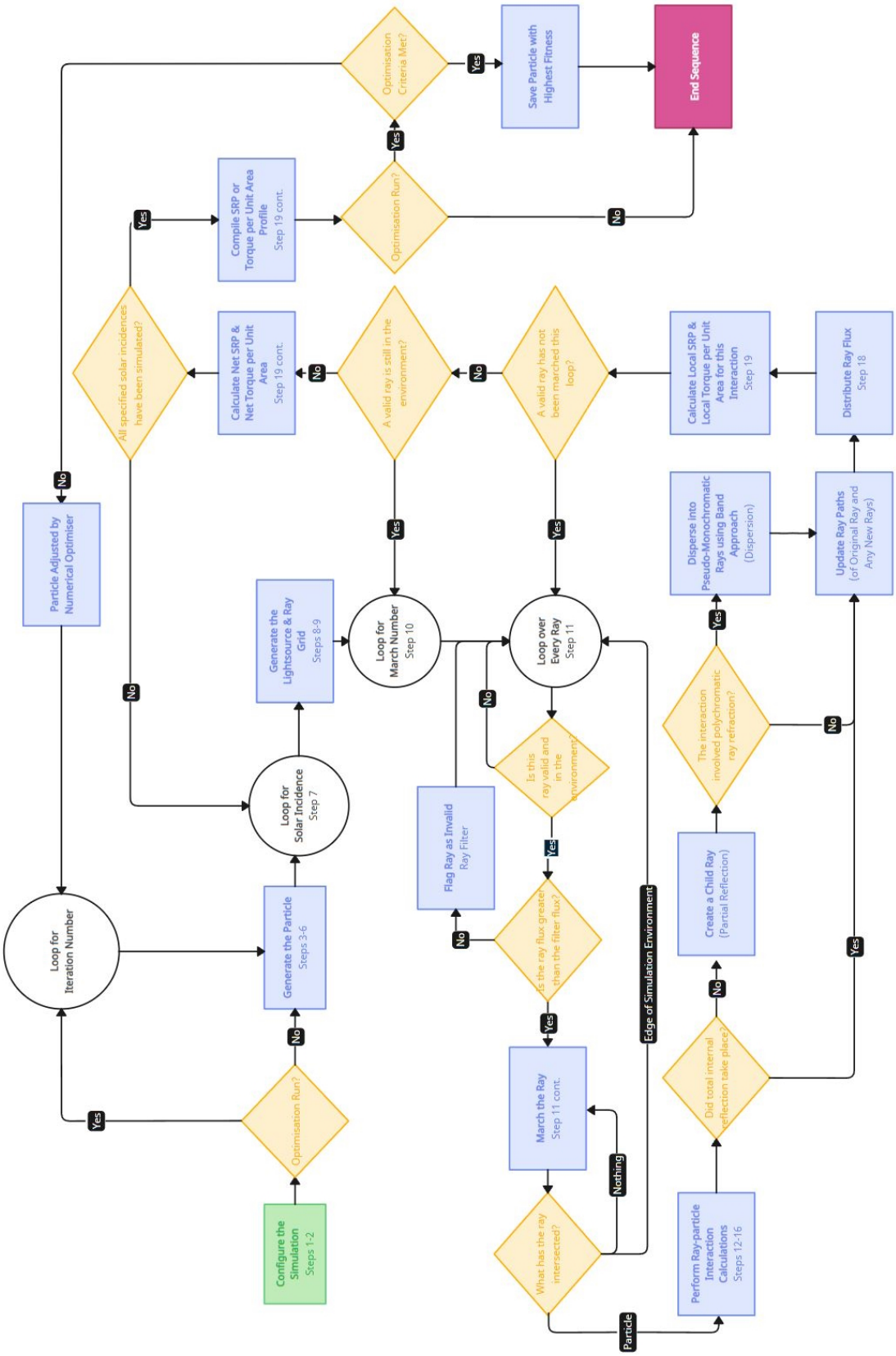

Figure 5.1: Simulation Sequence Flow Diagram

v. `alpha_shape` (alphShape object or array) – the convex hull data

vi. `bounding_box` (3×2 matrix) – the minimum and maximum x-coordinates of the shape compose the first row, then so on for y and z

vii. `Sellmeier_coefficients*` (1×2 column vector) – used in a Sellmeier equation to calculate the effective refractive index of a material according to the wavelength of the ray that has intersected it

Likewise, the following fields are only relevant to the lightsource entity.

i. `power*` (double) – typically set to G_{SC}

ii. `ray_interaction_data` ($5 + \mathrm{E} \times \mathrm{D}$ matrix) – stores information pertaining to power distribution (see bullet 2.)

iii. `grid*` (1×2 column vector) – the dimensions of the ray grid $\{Y_{\mathsf{R}}, Z_{\mathsf{R}}\}$. Setting this to [0 0] will tell the simulation to dynamically fit the ray grid to the size of the particle if the number of rays R is defined

iv. `spacing*` (double) – the distance between projected rays l_{R}. This field is made redundant if the number of rays is defined

v. `number*` (double) – the number of projected rays R. This field is made redundant if the ray spacing l_{R} is defined

2. **Fill the `ray_interaction_data` field.** The ray-interaction data field matrix has the following rows: wavelength (m), dispersed ray flux ratio (as a percentage of polychromatic ray flux - %), red (0-1), green (0-1), blue (0-1), followed by E rows which record the effective refractive index of each simulated particle for this specific ray, according to its wavelength and the Sellmeier coefficients of the particle. The red, green and blue rows are optional; they are only used for figures.

   Each column represents a ray that will be created during transmission. If dispersion effects are disabled or the lightsource is monochromatic, this will become a row vector as D = 1. The wavelength will default to 483 nm if a wavelength is not specified (which is the wavelength for which the black

body radiance of the Sun peaks), and the dispersed ray flux ratio will always be undefined. When the light is polychromatic and dispersion effects are enabled, the matrix will have D > 1 columns and the wavelength will be defined by the average wavelength of the solar spectrum *band* that said ray represents (detailed in section 3.2.4). In this case, the percentage power that each dispersed ray takes from the polychromatic ray will be defined by Eq. 3.23-3.24. It is important to note that this field is used to calculate the power that a dispersed ray takes from a polychromatic ray *after* the polychromatic ray has experienced other losses, such as partial reflection.

3. **Build the particle vertex set in 3D space.** Particles are defined by a set of vertices which are used to build a convex hull, in this case an alpha shape. These vertices are defined within the $\mathcal{T}_B$ reference frame (the axes extracted from the `axes` field). At simulation start, $\hat{\mathbf{x}}_\mathbf{B} = \hat{\mathbf{x}} = [1\ 0\ 0]$ and so on, so these axes must first be rotated according to $[\phi\ \theta\ \psi]$ from the `rotation` field to yield their true vector. This is done using the following sequence:

$$R_z(\psi) = \begin{bmatrix} \cos(\psi) & -\sin(\psi) & 0 \\ \sin(\psi) & \cos(\psi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5.1}$$

$$R_y(\theta) = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \end{bmatrix} \tag{5.2}$$

$$R_x(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\phi) & -\sin(\phi) \\ 0 & \sin(\phi) & \cos(\phi) \end{bmatrix} \tag{5.3}$$

$$R = R_z R_y R_x \tag{5.4}$$

In the case of the particle, these angles are user (or optimiser) inputs. In

the case of a lightsource, $\phi = 0°$, $\theta = \alpha$, $\psi = 0°$. Using $\hat{\mathbf{x}}_\mathbf{B}$ as an example, each axis is then rotated via Eq. 5.5:

$$\hat{\mathbf{x}}'_\mathbf{B} = (R \times \hat{\mathbf{x}}^T_\mathbf{B})^T \tag{5.5}$$

Because these particles have constant xy_B cross-sections (being representative of a patterned surface), the 2D cross-section is built first and its vertices are mirrored and offset in z_B to form the 3D vertex set. For structured shapes such a simple prism, the dimensions of the 2D shape are extracted from the `dimensions` field (in this case the dimensions of a triangle). For freeform optical elements, cubic spline interpolation is performed using a look-up table of control points which is prepared beforehand, and the `dimensions` field provides the indexing to locate the relevant control points in said table. Finally, the `repetition` field is queried to determine how many times the geometry is to be repeated to form a pattern or array; the vertices are cloned and offset by the length of the original shape within the specified axis.

In the former case, the optimiser can edit the `dimensions` field directly. In the latter case, *either* the look-up table of control points is edited or the vertices themselves are moved (editing control points is usually more computationally efficient as there are fewer of them, resulting in a smaller action and state space). If rotation and material changes are allowed, the optimiser can also edit the `rotation` and `Sellmeier_coefficients` fields directly (typically this is not done; rotation confuses the optimiser by changing the position of every vertex, and giving the optimiser access to the Sellmeier coefficients tends to result in geometries that require theoretical materials).

4. **Convert the vertex set into a convex hull, and the convex hull into a triangle set.** This is done using *Delaunay triangulation*. In MATLAB, this is achieved via the '`delaunayTriangulation`', '`freeBoundary`'

and '`trianglulation`' functions, generating both a point matrix (containing the coordinates of three vertices per triangle) and a connectivity matrix (containing the ID of the three vertices that a given triangle is composed of).

Given time, one might consider writing their own convex hull algorithm for better integration with their optimiser. This is because of issues pertaining to *redundant vertices* that arise when using generic convex hull algorithms to define freeform shapes that have been edited by an optimiser: after the creation of a convex hull from a set of vertices, any vertex that lies within the convex hull (rather than on its boundary) is said to be redundant. A vertex that has *only* formed triangles with vertices that have the same z_B coordinate as itself must lie on one of the two 2D faces, but not on their boundary, so these are also said to be redundant. Redundancy is easily avoided when a vertex set is defined manually, but is harder to prevent during numerical optimisation. This is particularly true of a model-free reinforcement learning algorithm (see section 3.3), as the AI has no understanding of what constitutes a shape, and instead interprets the validity of an action wholly according to the discounted reward function f_r^*.

In lieu of a custom convex hull algorithm, the removal of redundant vertices is necessary for torque calculation as they lead to erroneous calculation of the centre of mass (during bullet 6., Eq. 5.8). It is preferable to remove redundant vertices in any case, as these can produce unnecessary triangles (i.e. multiple triangles that are both adjacent and co-planar), adding to computation time for no gain. The simplest way to do this is to build the convex hull, ascertain which vertices form the outer boundary of the two 2D faces by querying the connectivity matrix, and then to rebuild the convex hull using only the boundary vertices.

5. **Fill the `bounding_box` field for the particle.** This is a matrix B_{box}

described by Eq. 5.6:

$$B_{\text{box}} = \begin{bmatrix} x^{\text{MIN}} & x^{\text{MAX}} \\ y^{\text{MIN}} & y^{\text{MAX}} \\ z^{\text{MIN}} & z^{\text{MAX}} \end{bmatrix} \tag{5.6}$$

6. **Fill the `centre_of_mass` field** (if calculating torque). First, the cross-section A_{xy_B} is calculated through the *shoelace formula* presented by Eq. 5.7. Because these particles are symmetric about xy_B and are of uniform density, the centre of mass can then be calculated from Eq. 5.8 [111].

$$A_{xy_B} = \frac{1}{2}\sum_{u=0}^{U-1}(x_u y_{u+1} - x_{u+1}y_u) \tag{5.7}$$

$$x_{\text{CoM}} = \frac{1}{6A_{xy_B}}\sum_{u=0}^{U-1}(x_u + x_{u+1})(x_u y_{u+1} - x_{u+1}y_u)$$

$$y_{\text{CoM}} = \frac{1}{6A_{xy_B}}\sum_{u=0}^{U-1}(y_u + y_{u+1})(x_u y_{u+1} - x_{u+1}y_u)$$

$$z_{\text{CoM}} = \frac{1}{U}\sum_{u=1}^{U} z_u = \bar{z}_u = 0 \tag{5.8}$$

Where u is the vertex index, U is the number of vertices, and $\{x_u, y_u, z_u\}$ are the Cartesian coordinates of a vertex.

7. **Begin the α loop needed for SRP or torque per unit area profiles.** Place the subsequent steps within a loop for solar incidence α wherein $\alpha_1 \leq \alpha \leq \alpha_2$ and $\{\alpha_1, \alpha_2\}$ are user-defined: an SRP profile will be complete once $\alpha = \alpha_2$.

8. **Set up the ray grid.** In the case of patterns with constant A_{xy_B}, a 3D ray grid will yield the same results as a 2D ray grid, so a 2D one will suffice. This is discussed in depth in section 3.2.6. Position the lightsource

above the focus via $O_{\text{grid}} + [0\ 0\ 10]$ and offset each ray in either direction according to $Y_{\text{grid}}/\mathsf{R}$, wherein O_{grid} and Y_{grid} are calculated via Eq. 3.26-3.27 for a particle and Eq. 3.28-3.29 for a pattern.

9. **Create the rays themselves.** These are built as structs with the following format:

   i. `index` (double) – the ray number according to the order in which the rays were created, i.e. `ray(1).index = 1`

   ii. `source` (double) – the entity index of the light source, not to be confused with the ray index of a `parent` ray

   iii. `parent` (double) – the `index` of any parent rays, if said ray was created during dispersion or partial reflection; if the ray is an original ray with no parent, it is said to be its own parent (`ray(q).parent = ray(q).index`)

   iv. `children` (column vector) – the `index` of any children rays created during dispersion or partial reflection

   v. `flux` (column vector) – the power of the ray at each ray march

   vi. `source_flux` (double) – the power of the light source (`ray(q).source_flux = entity(ray(q).source).power`)

   vii. `lambda` (double) – wavelength of the ray

   viii. `history` (cell array containing strings) - history of encounters at each simulation stage, e.g. 'refracted' and 'left environment'; at ray creation, this field will contain one cell that reads 'propagating'

   ix. `intersects` (matrix) – a row for each encounter; column 1 denotes the entity intersected, column 2 denotes the specific shape (in the case of an array), column 3 denotes the triangle that was intersected, and column 4 denotes the angle of intersection

x. `position` ($3 \times \mathfrak{m}_q$ matrix) – history of the position of the ray, where $\mathfrak{m}_q$ is the number of successful marches executed by ray q

xi. `direction` ($3 \times \mathfrak{m}_q$ matrix) – history of the direction of the ray

xii. `colour` (1×3 matrix) – RGB colour of the ray

10. **Begin the $\mathfrak{m}$ loop to begin ray tracing.** Place the subsequent steps within a loop for march number $\mathfrak{m}$ wherein $1 \leq \mathfrak{m} \leq \mathfrak{m}_{\max}$ and $\mathfrak{m}_{\max}$ is user-defined.

11. **March each ray that is within the simulation environment.** If the latest `history` cell of a ray reads 'propagating' or 'repropagating', march it forwards by march length ∂L_{R} along the $\hat{\mathbf{c}}_{\mathbf{q}}$ axis that is stored in the `direction` field. Do this on repeat until it falls between the coordinates of a bounding box or exits the simulation environment. Naturally, it is important that ∂L_{R} is much smaller than the smallest dimension of the boundary box. What happens next depends on whether the ray leaves the environment or encounters a particle.

12. **Update the ray according to the outcome of the march.** If the ray enters a bounding box, intersection can be queried by generating sample points between the last and second-to-last $\mathbf{r}_{\mathbf{q}}$ using the ray `position` field. In MATLAB, whether each of these vertices lies within the `alpha_shape` can be queried using the *inShape* function, generating a row vector containing 0 for vertices that lie outside the `alpha_shape` and 1 for vertices that lie within it. The position of the intersection will be refined in the next step.

    i. If an intersection is found, update the latest `history` cell to read 'intersected'. Also update the first two columns of the `intersects` field with the ID of the entity and shape that was intersected.

    ii. If no intersection is found, check whether the latest ray `position` lies within the simulation boundaries.

a. If so, repeat the march attempt.

b. If no, update the latest `history` cell to read 'left environment'.

In the case of an array, this process is repeated for each shape. In practice, it is a reasonable shortcut to only check the three nearest shapes for intersection.

13. **Determine the nature of the interaction on a per-triangle basis** if the latest ray `history` cell reads 'intersected', using the triangle set that was built during bullet 4.. This can be queried using the two ID's generated for the `intersects` field in the previous step.

    First, test for intersection with each triangle in the set (for better efficiency, first filter for distance from the ray). This is done by attempting to calculate $\mathbf{P_0}$, the point of intersection with the triangle plane (Eq. 5.9).

$$\mathbf{P_0} = \mathbf{r_q}(\mathfrak{m}\text{-}1) + [\mathbf{r_q}(\mathfrak{m}) - \mathbf{r_q}(\mathfrak{m}\text{-}1)]\frac{[\mathbf{P_1} - \mathbf{r_q}(\mathfrak{m}\text{-}1)] \cdot \hat{\mathbf{n}}_{\text{tri}}}{[\mathbf{r_q}(\mathfrak{m}) - \mathbf{r_q}(\mathfrak{m}\text{-}1)] \cdot \hat{\mathbf{n}}_{\text{tri}}} \tag{5.9}$$

    Wherein $\mathbf{r_q}(\mathfrak{m}\text{-}1)$ and $\mathbf{r_q}(\mathfrak{m})$ are the beginning and end points of the ray march, $\{\mathbf{P_1}, \mathbf{P_2}, \mathbf{P_3}\}$ are the vertices of the triangle and $\hat{\mathbf{n}}_{\text{tri}} = (\mathbf{P_2} - \mathbf{P_1}) \times (\mathbf{P_3} - \mathbf{P_1})$ is the triangle normal. To intersect with a triangle, the following conditions must be met:

`Cond. 1:`

$$\mathbf{P_0} \in \mathbb{R} \tag{5.10}$$

`Cond. 2:`

$$\begin{aligned} (\mathbf{P_2} - \mathbf{P_1}) \times (\mathbf{P_0} - \mathbf{P_1}) \cdot \hat{\mathbf{n}}_{\text{tri}} \geq 0 \\ (\mathbf{P_3} - \mathbf{P_2}) \times (\mathbf{P_0} - \mathbf{P_2}) \cdot \hat{\mathbf{n}}_{\text{tri}} \geq 0 \\ (\mathbf{P_1} - \mathbf{P_3}) \times (\mathbf{P_0} - \mathbf{P_3}) \cdot \hat{\mathbf{n}}_{\text{tri}} \geq 0 \end{aligned} \tag{5.11}$$

These state that $\mathbf{P_0}$ must be real and finite to have intersected the triangle plane (Eq. 5.10) and that $\mathbf{P_0}$ must lie between the vertices of the intersected triangle to have intersected the triangle itself (Eq. 5.11). Because each ray march attempt will propagate the ray by only a small distance, generally only a single triangle will be intersected. However, where two edges of a particle converge (e.g. a cusp), a ray may intersect two triangles. A simple test to determine which triangle was intersected first is to determine which $\mathbf{P_0}$ lies closest to the vertex of the previous ray march (Eq. 5.12):

$$|\mathbf{r_q}(\mathfrak{m}\text{-}1) - \mathbf{P_0}| > |\mathbf{r_q}(\mathfrak{m}\text{-}1) - \mathbf{P_{0_{\text{false}}}}| \tag{5.12}$$

14. **Calculate the angle of incidence** θ_i with the intersected triangle (Eq 5.13).

$$\theta_i(\mathfrak{m}) = \tag{5.13}$$
$$\text{atan2}\big(|\hat{\mathbf{n}}_{\text{tri}} \times [\mathbf{r_q}(\mathfrak{m}\text{-}1) - \mathbf{r_q}(\mathfrak{m})]|,$$
$$|\hat{\mathbf{n}}_{\text{tri}} \cdot [\mathbf{r_q}(\mathfrak{m}\text{-}1) - \mathbf{r_q}(\mathfrak{m})]|\big)$$

15. **Calculate the reflected ray path** (because of the potential for TIR, do this even if reflection is disabled). To do this, first calculate the axis of ray rotation for reflection $\hat{\mathbf{o}}_r$ (Eq. 5.14) and the reflection angle θ_r (Eq. 5.15).

$$\hat{\mathbf{o}}_r = \begin{bmatrix} o_x \\ o_y \\ o_z \end{bmatrix} = \hat{\mathbf{c}}_q(\mathfrak{m}\text{-}1) \times \hat{\mathbf{n}}_{\text{tri}} \tag{5.14}$$

$$\theta_r = 180^\circ - 2\theta_i \tag{5.15}$$

Given the definition for the ray direction vector $\hat{\mathbf{c}}_q(\mathfrak{m}\text{-}1)$ provided by Eq.

5.16, calculate the reflected ray direction vector $\hat{\mathbf{c}}_r$ (Eq. 5.17).

$$\begin{bmatrix} c_x \\ c_y \\ c_z \end{bmatrix} = \hat{\mathbf{c}}_q(\mathfrak{m}\text{-}\mathfrak{1}) \tag{5.16}$$

These values are stored for now and not applied to a new ray. This is because it is not yet known whether the ray has undergone partial reflection or TIR.

$$\hat{\mathbf{c}}_r = \begin{bmatrix} ((o_y^2 + o_z^2) - o_x(o_y + o_z - o_x c_x - o_y c_y - o_z c_z))(1 - \cos(\theta_r)) + c_x \cos(\theta_r) + (-o_y + o_z - o_z c_y + o_y c_z) \sin(\theta_r) \\ ((o_x^2 + o_z^2) - o_y(o_x + o_z - o_x c_x - o_y c_y - o_z c_z))(1 - \cos(\theta_r)) + c_y \cos(\theta_r) + (+o_x - o_z + o_z c_x - o_x c_z) \sin(\theta_r) \\ ((o_x^2 + o_y^2) - o_z(o_x + o_y - o_x c_x - o_y c_y - o_z c_z))(1 - \cos(\theta_r)) + c_z \cos(\theta_r) + (-o_x + o_y - o_y c_x + o_x c_y) \sin(\theta_r) \end{bmatrix} \tag{5.17}$$

$$\hat{\mathbf{c}}_t = \begin{bmatrix} ((o_y^2 + o_z^2) - o_x(o_y + o_z - o_x n_x - o_y n_y - o_z n_z))(1 - \cos(\theta_t)) + n_x \cos(\theta_t) + (-o_y + o_z - o_z n_y + o_y n_z) \sin(\theta_t) \\ ((o_x^2 + o_z^2) - o_y(o_x + o_z - o_x n_x - o_y n_y - o_z n_z))(1 - \cos(\theta_t)) + n_y \cos(\theta_t) + (+o_x - o_z + o_z n_x - o_x n_z) \sin(\theta_t) \\ ((o_x^2 + o_y^2) - o_z(o_x + o_y - o_x n_x - o_y n_y - o_z n_z))(1 - \cos(\theta_t)) + n_z \cos(\theta_t) + (-o_x + o_y - o_y n_x + o_x n_y) \sin(\theta_t) \end{bmatrix} \tag{5.18}$$

16. **Calculate dispersed ray paths** if dispersion is enabled ($D > 1$) and if the ray has not already been dispersed. To check if a ray is going to disperse, query `entity.history`. If it contains the term '`refracted`' at any point, the ray has already been split into coloured rays and cannot be split again.

If dispersion is going to take place, several new rays are created by cloning the parent ray ($D-1$ new rays, as the first of these 'diverged' rays will overwrite the parent ray). This will change the ray count to $Q' = Q+D-1$. Next, each ray is updated according to the `entity.ray_interaction_data` field belonging to the lightsource: update `lambda`, `colour` and `flux` according to the data stored here during bullet 2.. Finally, set the effective refractive index for this particle-ray combination n_{particle} as described by the final rows of `entity.ray_interaction_data`.

It is also necessary to determine whether the ray is exiting or entering a particle. This can be done by querying `history` at cell $\mathfrak{m}$-1: if it contains '`reflected`', '`refracted out`' or '`propagating`', the `history` cell for $\mathfrak{m}$ should be set to '`refracted`'. If it does not, it should be set to '`refracted out`'.

In the `refracted` case, $n_1 = 1$ and $n_2 = n_{\text{particle}}$, wherein subscript 1 refers to the exited medium and 2 the entered medium. In the `refracted out` case, $n_1 = n_{\text{particle}}$ and $n_2 = 1$.

Next, calculate the refraction angle θ_t (Eq. 5.19) and the refracted ray path $\hat{\mathbf{c}}_t$ (Eq. 5.18). This is calculated in a similar manner to $\hat{\mathbf{c}}_r$, only this time, the inputs are components of the triangle normal $\hat{\mathbf{n}}_{\text{tri}}$ (Eq. 5.20) rather than of the ray path vector $\hat{\mathbf{c}}_q(\mathfrak{m}\text{-}1)$.

$$\theta_t = 180^\circ + \arcsin\left(\frac{n_1}{n_2}\sin(\theta_i)\right) \tag{5.19}$$

$$\begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} = \hat{\mathbf{n}}_{\text{tri}} \tag{5.20}$$

Finally, check for exceptions before updating `direction`. For the first exception, refraction will not occur at $\theta_i = 0^\circ$. In this case, Eq. 5.18 may yield

an undefined $\hat{\mathbf{c}}_t$, requiring it to be set manually as depicted by Eq. 5.21. For the second exception, Eq. 5.18 will yield a $\hat{\mathbf{c}}_t$ that has an imaginary component, indicating that $\theta_i > \theta_c$ and TIR has occurred. In this case, $\hat{\mathbf{c}}_t$ is actually undefined as no component of the ray is transmitted to the next medium, and so it is set to equal $\hat{\mathbf{c}}_r$ as depicted by Eq. 5.22.

$$\texttt{Excep. 1:} \tag{5.21}$$
$$\hat{\mathbf{c}}_t \ni \text{NaN} \rightarrow \qquad \theta_t = 0°, \qquad \hat{\mathbf{c}}_t = \hat{\mathbf{c}}_q(\mathfrak{m} - \mathfrak{1})$$
$$\texttt{Excep. 2:} \tag{5.22}$$
$$\hat{\mathbf{c}}_t \notin \mathbb{R} \rightarrow \qquad \theta_t = \text{NaN} \qquad \hat{\mathbf{c}}_t = \hat{\mathbf{c}}_r$$

In the first case, the `direction` of this ray for this ray march will be unchanged from that of the previous ray march. In the second case, the new `direction` of the ray is dictated by the reflected ray path and `history` should be updated to '`internally reflected`'. If neither of these exceptions occur, the `direction` $\hat{\mathbf{c}}_t$ has been calculated normally. In this case, and in the case of the first exception, `history` should be updated to '`refracted`' for each dispersed ray.

The ray path of the transmitted and reflected rays have now been determined. If TIR occurred, no further action is required.

17. **Create children rays and calculate the distribution of ray flux** according to the nature of the ray-particle interaction. If transmission occurred, partial reflection will have also occurred. In this case an additional ray is created by cloning the parent and setting its `direction` to $\hat{\mathbf{c}}_r$. It is also helpful to clear the `history` cells of the new ray for all marches preceding this splitting event, thereby making it clear that this is not an original ray.

    In the case of polychromatic ray transmission, D new rays are created by ray q, with D-1 new rays being transmitted and dispersed (in addition to ray

q itself, which does not add to the ray count), plus one additional ray from partial reflection. If the ray is monochromatic, only ray q is transmitted and — due to partial reflection — the ray count increases by just one. Conversely, in the case of TIR, no new rays are created and ray q is both the sole input and the sole output.

For this step, it is necessary to record the indices of each ray involved in the interaction. This is because the index of a ray relates only to the ray number Q at the time of its creation. As such, ray q and its children are not able to be stored as a sequence unless $Q = 1$ at the time of the interaction. For demonstration purposes we define two 'local' ray numbers that are specific to this interaction: Q' is to be the new ray number after dispersion, and Q'' is to be the new ray number after a sequential partial reflection; after an interaction is finished, we set $Q = Q''$ as this will be the most up-to-date ray number. How Q will change depends on the interaction type and the composition of the ray as expressed by Eq. 5.23-5.25:

`Polychromatic Transmission:` (5.23)

$$Q' = Q + D - 1$$

$$Q'' = Q' + 1 = Q + D$$

`Monochromatic Transmission:` (5.24)

$$Q' = Q$$

$$Q'' = Q' + 1 = Q + 1$$

`Total Internal Reflection:` (5.25)

$$Q'' = Q' = Q$$

As each interactions is analysed on a per-ray basis, only one ray — ray q — is present at the start of the interaction, and ray q is also always amongst the outgoing rays.

To surmount the issue of $q{:}Q''$ not being in-sequence when new rays are created, we record three index sets to be queried later (Eq. 5.26-5.30):

$$\mathrm{S_i} = \{q\} \tag{5.26}$$

$$\mathrm{S_t} = \{q, Q{+}1, Q{+}2, ..., Q'\} \tag{5.27}$$

$$\mathrm{S_r} = \{Q''\} \tag{5.28}$$

Or in the case of TIR:

$$\mathrm{S_i} = \mathrm{S_r} = \{q\} \tag{5.29}$$

$$\mathrm{S_t} = \{\} \tag{5.30}$$

18. Finally, **the ray flux W can be divided amongst the rays involved** in this particular interaction according to the S and P-oriented reflectivities $\{R_S, R_P\}$ and transmissivities $\{T_S, T_P\}$, respectively, wherein $R_S + R_P + T_S + T_P = 100\%$ (Eq 5.31-5.32) [46]. The relevant rays can be queried according to the sets recorded during the previous step.

`Transmission & Partial Reflection:` (5.31)

$$R_S = |\frac{n_1 \cos(\theta_i) - n_2 \cos(\theta_t - 180)}{n_1 \cos(\theta_i) + n_2 \cos(\theta_t - 180)}|^2$$

$$R_P = |\frac{n_1 \cos(\theta_t - 180) - n_2 \cos(\theta_i)}{n_1 \cos(\theta_t - 180) + n_2 \cos(\theta_i)}|^2$$

$$T_S = 1 - R_S$$

$$T_P = 1 - R_P$$

`Total Internal Reflection:` (5.32)

$$R_S = R_P = 1$$

$$T_S = T_P = 0$$

This can be applied to determine the power assigned to the reflected ray

and the transmitted ray(s), if any, as depicted by Eq. 5.33.

$$W_r = \frac{1}{2}(R_S + R_P)W_q(\mathfrak{m}\text{-}1) \tag{5.33}$$
$$W_t = \frac{1}{2}(T_S + T_P)W_q(\mathfrak{m}\text{-}1)$$

Wherein $W_r + W_t = W_q(\mathfrak{m}\text{-}1)$ (*conservation of energy law*). Because diffuse reflection is not considered, $\mathrm{S_r}$ always comprises a single ray that receives all of W_r. If ray q is monochromatic and TIR has not occurred, W_t is likewise assigned to the single transmitted ray of $\mathrm{S_t}$. Conversely, if ray q is polychromatic, W_t is divided amongst rays $\mathrm{S_t}$ according to the flux ratio stored in the `ray_interaction_data` field of the lightsource for each ray band.

19. **Determine the SRP F_S and torque per unit area $\tau_{\mathbf{S}}$.** This is done by first calculating the local SRP $\mathbf{F_{Si}}$ and local torque per unit area $\tau_{\mathbf{Si}}$ for this interaction (Eq. 5.34-5.35).

$$\mathbf{F_{Si}} = \frac{1}{c} \sum_{q* \in \{\mathrm{S_t}, \mathrm{S_r}\}} \left[W_{q*}(\mathfrak{m}) \cdot \hat{\mathbf{c}}_{q*}(\mathfrak{m}) \right] \tag{5.34}$$
$$- W_q(\mathfrak{m}\text{-}1) \cdot \hat{\mathbf{c}}_q(\mathfrak{m}\text{-}1) \Big]$$

$$\tau_{\mathbf{Si}} = [\mathbf{r_q}(\mathfrak{m}) - \mathbf{r}_\mathrm{g}] \times \mathbf{F_{Si}} \tag{5.35}$$

Wherein $\mathbf{r}_\mathrm{g}$ is the position of the centre of mass. Assuming that total illumination of the particle was achieved, the net SRP vector $\mathbf{F}_\mathrm{S}$ and torque per unit area $\tau_{\mathbf{S}}$ may be calculated by taking the summation of local pressures and local torques, respectively, and adjusting these for the number of rays that interacted with the particle (Eq. 5.36-5.37).

$$\mathbf{F}_{\mathrm{S}} = \frac{1}{\mathsf{R}^*} \sum_{\mathsf{i}=1}^{\mathsf{I}} \mathbf{F}_{\mathbf{S}\mathsf{i}} \cdot |\cos(\alpha)| \tag{5.36}$$

$$\tau_{\mathbf{S}} = \frac{1}{\mathsf{R}^*} \sum_{\mathsf{i}=1}^{\mathsf{I}} \tau_{\mathbf{S}\mathsf{i}} \cdot |\cos(\alpha)| \tag{5.37}$$

Wherein I is the total number of ray-particle interactions during this simulation run and R^* is the number of *valid* rays: a ray is valid if it encounters a particle during the simulation. Depending upon the ray casting method employed, R^* may vary with α.

20. **Adjust sequence for optimisation.** If optimising rather than simply generating profiles, place steps 2.-19. within a loop for iteration number $\mathrm{n_I}$, wherein the optimiser function is called upon at the start of each loop. Define a fitness function f_f outside of this loop. Using the profiles generated during step 19., the fitness of each geometry can be assessed. Optimisation will be complete once $\mathrm{n_I} = \mathrm{n_{I\,max}}$.

    Supplementary information pertaining to the specific optimisation runs that were carried out during this chapter are provided in their respective sections, as well as section 5.2. Furthermore, extensive literature exists for practically every permutation of optimising algorithm [110, 112, 113] of which many have been adapted to optical systems in particular [114, 115]; any number of these could be adapted to the problem of solar sail pattern optimisation.

5.2 Optimisation Setup

In order to explore the solution space, membrane patterns have been generated according to a number of different constraints, initial geometries and fitness functions f_f in this chapter. During one set of optimisation runs (section 5.3.2), numerically optimised patterns are compared with analytically optimised pat-

terns from source [29]. In this case, the numerically optimised sail is assumed to be composed of PS so as to maintain consistency with the original proposal, and Sellmeier coefficients are cited as $B_S = 1.4435$ and $C_S = 0.020216$ [116]. All other optimisation runs assume the sail material to be PET. In the absence of literature pertaining to the Sellmeier coefficients of PET, B_S and C_S were derived from an empirical dataset [117] containing PET refractive indices n with respect to wavelength λ_k. Eq. 3.21 was used to generate a curve of best fit while solving for n. This curve is depicted by Fig. 5.2.

This yielded $B_S = 1.57$ and $C_S = 0.02505$ with a sum of square error of SSE = 6.105×10^{-5}, coefficient of determination of $R^2 = 0.9995$ and root-mean-square deviation of RMSE = 3.868×10^{-4}. The ascertained Sellmeier coefficients were then used to extrapolate $n(\lambda_k)$ over a wider λ_k range according to Eq. 3.21.

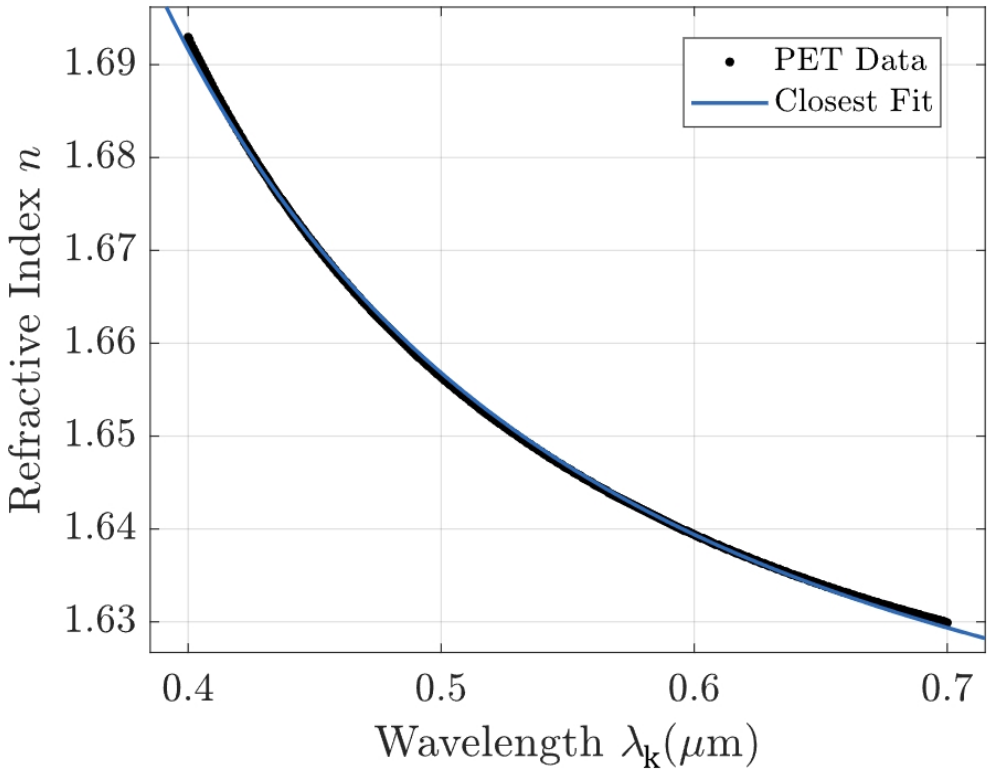

Figure 5.2: Sellmeier Curve for PET via Curve Fitting of Empirical Data, yielding $B_S = 1.57$ and $C_S = 0.02505$

The next consideration is the reference frame that should be used when gener-

ating SRP profiles. In general, $F_{S_{xp}}$ of the manoeuvre reference frame $\mathcal{T}_{\mathfrak{p}}$ is the most widely applicable metric for the assessing of sail performance in any given flight regime. This was defined in section 2.1.2. It correlates with (but is not directly proportional to) rate change of specific orbital energy for LEO and above, minimum operational altitude in LEO, and transit times. However, it is an inconvenient metric for an optimisation problem as it is dependent upon the flight regime, particularly the Sun-velocity misalignment angle γ [40]. Furthermore, $\mathcal{T}_{\mathfrak{p}}$ is in general an unintuitive reference frame for the determining of torques. Instead, it is more convenient to express pressures and torques in the body-fixed frame $\mathcal{T}_b$. These can be directly applied to $\mathcal{T}_{\mathfrak{p}}$ for a given flight regime using basic trigonometry.

The metric that is most pertinent to the following optimisation problems is the tangential SRP $F_{S_{yb}}$, as this correlates with greater $F_{S_{xp}}$ for typical solar sailing orbits in a near Sun-pointing attitude, such as circular heliocentric orbits and dusk-dawn SSOs [40]. Indeed, under ideal $\gamma = 0°$, $\alpha = 0°$ conditions, $F_{S_{yb}}$ is equivalent to $F_{S_{xp}}$. In general, $F_{S_{yb}}$ is a good indicator of overall transmissive sail performance, so long as one remembers that this is only an indication, and one that becomes less valid if γ deviates significantly from zero.

For torque generating patterns that are concerned with active control rather than passive stability, $F_{S_{yb}}$ is a suitable indicator of performance. This is because such patterns generate the majority of their torque by merit of tangential SRP that occurs with significant offset from the centre of mass [29], and so for these, the SRP and torque per unit area optimisation problems are essentially the same; a prism that is optimised for $F_{S_{yb}}$ will naturally be optimal for torque. In particular, torque generating prisms should be optimised for $\alpha = 0°$. This may seem somewhat contradictory — and indeed it would be if these patterns were expected to generate pitching torque $\tau_{S_{yb}}$ or yaw torque $\tau_{S_{zb}}$, as they would need to provide torque over a wide range of α. In practice, exclusively active torque generating

patterns are relegated to rolling torque $\tau_{S_{xb}}$, for which near-zero α may be assumed. It would be atypical for a refractive or diffractive membrane to be used for active pitch or yaw torques, because simpler, reflective membranes are superior in this role. This is because, with the exception of certain boom-extended solar sail configurations, the offset in $\hat{\mathbf{y}}_\mathbf{b}$ between the sail perimeter and the axis $\hat{\mathbf{x}}_\mathbf{b}$ will typically be larger than the offset in $\hat{\mathbf{x}}_\mathbf{b}$ between the sail and the plane yz_B (wherein all body-fixed axes converge at the centre of mass). In other words, a reflection at the sail perimeter will effect a larger $\tau_{S_{yb}}$ (or $\tau_{S_{zB}}$) than refraction can at any arbitrary location on the solar sail. Furthermore, the maximum SRP that may be generated by reflection is twice as large as the maximum SRP that may be generated by transmission. For the sake of argument: a possible advantage of a transmissive, active torque pattern is that it may generate torque irrespective of its location upon a sail, while a reflective membrane will generate less torque at the sail centre than at its perimeter. In practice, this means that you could pattern an entire sail with transmissive, active torque patterns to brute force higher torque. However, such a sail would be incapable of effectively accelerating a sailcraft.

Conversely, where passive control is concerned, the stability of the pattern is the primary concern and so the torque per unit area profiles must be made available to the optimiser. Specifically, the problem pertains to stability in $\hat{\mathbf{y}}_\mathbf{b}$ and $\hat{\mathbf{z}}_\mathbf{B}$. The optimisation problems are identical for both axes, and so the choice between optimising for pitch or yaw is arbitrary. For convenience, we choose $\tau_{S_{yb}}$ to be the optimisation metric because y exists in both $\mathcal{T}_b$ and $\mathcal{T}_B$. The *corrective* equivalent of this torque is $\tau^*_{S_{yb}}$, and it is this that is principally used by the optimiser to ascertain fitness for a self-stabilising optical element (see section 5.5 for definition).

It should also be noted that the aim here is to explore sail designs that have high performance in the broadest sense (albeit typically with a bias towards zero-

α performance). However, a benefit of transmissive sails is that they may be optimised on a per-mission basis (through this method, this would be done by simply adjusting the fitness function and α-range).

5.2.1 Nomenclature for Analysing Ray Paths

SRP profiles and normalised pattern geometries adequately describe a pattern, and are the primary outputs of these simulations. However, they are less useful for determining *how* a pattern achieves its SRP profile and *why* the optimiser converged upon a given geometry. These questions can be answered by analysing the ray diagram to ascertain the ray paths that have had the greatest contribution to the SRP profile and reward function. For example, if the majority of the radiant flux that is passed through a particle can be seen to travel along a specific path or group of paths, then the SRP profile can be largely attributed to them. The actions that were taken by the optimiser to generate geometries conducive to these ray paths will have yielded the greatest reward owing to their disproportionately large contribution to fitness, explaining why these geometries came about.

To utilise these diagrams we require some new terminology. Hereafter, rays are described according to their *order*, which relates to the number of *parents* that a ray has (parents = order − 1). First-order rays are those that were initially cast and have no parents. During a ray splitting event, the *child* ray created by partial reflection is always one order higher than its parent, so the children of a first-order ray will always be of the second-order. When a ray is transmitted across a boundary it is considered to be a continuation of the original ray, so a first-order ray will persist in this form, and will always remain a first-order ray. This rule extends to rays created during the transmission and subsequent dispersion of polychromatic rays; each of the dispersed rays will be first-order. Order is a convenient descriptor because the number of parents that a ray has is a good indicator of radiant flux. This in turn is a good indicator of the contribution

of the ray to the SRP profile; first-order rays generally contribute the most by a significant margin.

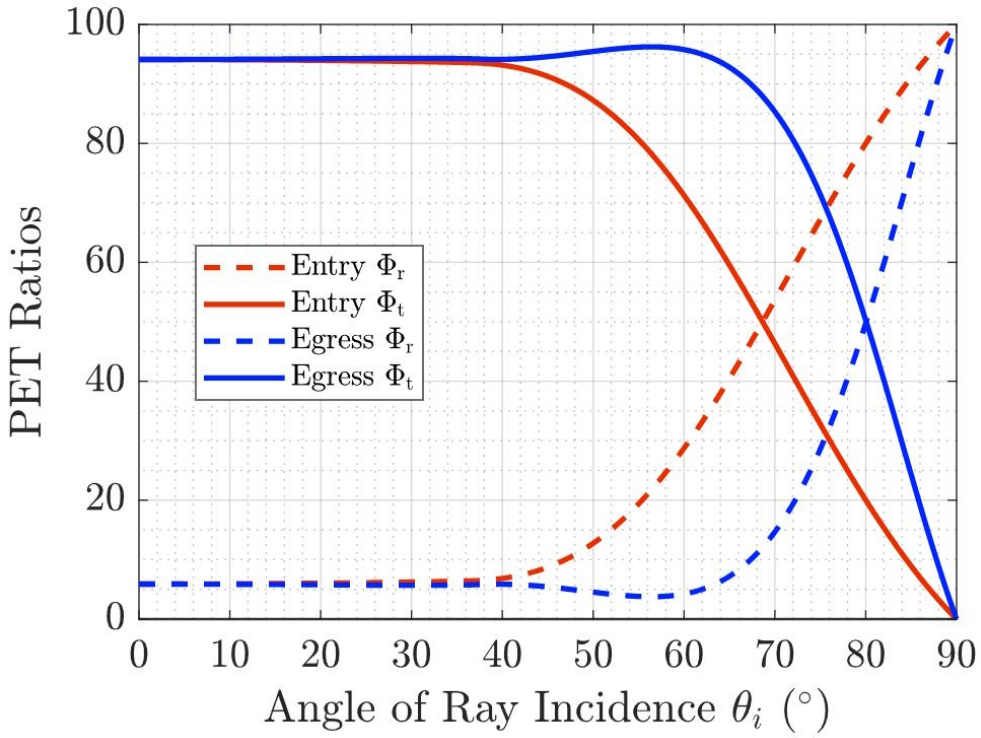

Figure 5.3: Reflectivity Φ_r versus Transmissivity Φ_t for PET in Vacuum on Entry and Egress of a $\lambda_k = 483$ nm Ray

The reason that order correlates with ray flux is that it describes how many times the original parent ray has distributed its power to other rays. When interacting with a particle composed of optical polymer and surrounded by a vacuum, a transmitted parent ray will generally retain far more flux than it confers to a reflected child ray — unless the angle of incidence is very high (see Fig. 5.3 for transmitted versus reflected power ratios of PET). For example, PET will transmit more power than it reflects up to an incidence angle of about 68° on ray entry, or 80° on ray egress, which are exceptional circumstances. Taking 45° to be the average incidence, the typical proportion of flux given to a reflected child ray is 9% on parent ray entry or 6% on parent ray egress. Therefore, a child ray will on average have a ray flux that is an order of magnitude less than its parent. Conversely, while a transmitted parent ray will attenuate each time that this happens, its ray flux will typically not change by an order of magnitude unless

many of these interactions occur — such as when it is propagated through several pattern elements. As such, first-order rays nearly always retain their order-of-magnitude flux advantage over their peers, and so the assumption that order and flux inversely correlate remains generally true.

Order is also a convenient indicator of optimiser agency over a given ray: generally, the paths of first-order rays will have been central to the decision making of the optimiser, whereas any positive or negative impetus from a fifth-order ray will almost certainly have arisen through random chance, rather than by design. This is true for two reasons: for one, lower order rays contribute the most to fitness, and because increasing their effectiveness yields the greatest reward, they will be prioritised over higher order rays. A second-order ray will generally be prioritised over a third-order ray, and so on. Secondly, the higher the order of a ray, the more sensitive it is to small changes made by the optimiser. This is because with each increase in order, an additional interaction has to take place to bring about the given ray path. The more complex the ray path, the more sensitive the ray is to initial conditions. This is an incarnation of the *butterfly effect*: while the paths of these rays are deterministic as for any ray path, they nonetheless become harder for the optimiser to interpret. Combined with the lower motivation for the optimiser to try to interpret them, their contributions to the action and reward functions are akin to noise.

5.2.2 Employing New Nomenclature to Improve Ray Diagrams

One can infer the approximate power distribution and SRP contributions of a ray diagram by observing how many ray splitting events have occurred to produce a given ray, as well as by comparing its $\hat{\mathbf{c}}_q$ on egress (final $\mathfrak{m}$) with the $\hat{\mathbf{c}}_q$ of its original parent on entry ($\mathfrak{m} = 0$). In particular, the *angular deviation* θ_k is the

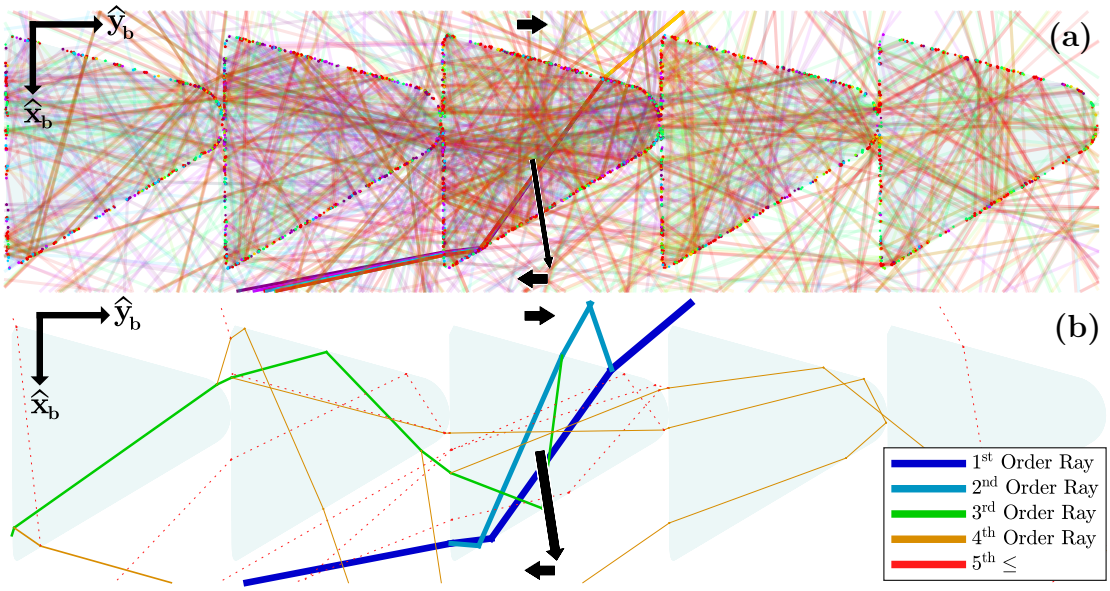

Figure 5.4: Ray Diagram for a Single Ray Entering a Pattern of 11 Elements (Only Five Elements Shown and Ray Count Reduced for Clarity) from $\alpha = -60°$, Direction of Net SRP and Torque Indicated by One Black Arrow by the Centre of Mass and Two Opposing Black Arrows Beyond the Particle, Respectively: (a) Unfiltered, Coloured by Ray Wavelength, Opacity Determined by Ray flux, (b) Filtered, Coloured by Ray Order, Width Determined by Natural Logarithm of Ray flux

angle between $\hat{\mathbf{c}}_q(0)$ and $\hat{\mathbf{c}}_q(\mathfrak{m})$, and is a convenient indicator of SRP contribution: while the positive-anticlockwise convention is generally adhered to in this report, θ_k is said to be positive if clockwise as this will effect a positive contribution to $F_{S_{yb}}$. Furthermore, a ray that is passed *around* the centre of mass with a positive, clockwise θ_k *tends* to effect an anticlockwise, conventionally positive $\tau_{S_{yb}}$, but this is not guaranteed. Generally, $\theta_k = 90°$ is most conducive to $F_{S_{yb}}$ while $\theta_k = -90°$ is most harmful to it; $\theta_k = 0°$ has zero net effect, while $\theta_k = 180°$ effects $F_{S_{xb}}$ exclusively. The outcome of θ_k may nonetheless be negligible for a high order ray, whereas a few first order rays with a conducive θ_k may dominate an SRP or torque profile.

However, reading ray diagrams can be laborious or, in the case of Fig. 5.4a, impossible. This diagram depicts a single ray undergoing $\mathfrak{m} = 20$ marches with zero filtering and $D = 10$ ray bands allowed. This represents the most natural interpretation of a ray diagram wherein colour is determined by wavelength as

for real life, with ultraviolet and the unhelpfully-prominent infrared rays being depicted as violet and red, respectively. Furthermore, to give some semblance of legibility, here ray flux determines opacity (albeit with a lower limit so that low flux rays remain visible), and for this reason the dominant ray path can be discerned. However, the diagram is too busy to determine much else. Fig. 5.4b shows an improved, more legible interpretation of a ray diagram. Firstly, applying a ray filter for only 0.001% of G_{SC} culls hundreds of very low power rays from view. Next, the natural logarithm of ray flux is used to determine the width of each ray. This is done to account for the fact that the power distribution will typically span several orders of magnitude. The lower limit for ray width occurs at $< 0.01\%$ G_{SC} and after this point the ray becomes a dotted line. Finally, the colour of each ray is determined by its order, which is discerned by tallying the number of parents that a ray has. Because of the aforementioned order-flux relationship, ray colour and width tend to correlate in these diagrams, but this is not always the case. For example, a second-order ray created at the start of a simulation may be much higher flux than one born from the same parent near the end of a simulation, because said parent will have attenuated. In subsequent ray diagrams, multiple rays are cast, and naturally this may threaten legibility. To curtail this, the filter is typically set to be more ruthless, such that the appearance of fifth-order rays is rare. Furthermore, up to a dozen rays are cast for the ray diagram, which is far lower than the 50 that are cast during optimisation, and the 50 or 100 that are cast to generate the SRP or torque density profiles, respectively. Furthermore, SRP and torque arrows are normalised relative to the pattern size for clarity, and their length is not indicative of vector magnitude.

In the single ray example depicted by Fig. 5.4b, it can be seen that the majority of ray flux is retained by the first-order ray that enters the particle. It can be seen that this ray undergoes TIR (which one identifies by noting that no children rays were created during that particular interaction), and is then ejected with high $-\hat{\mathbf{y}}_{\mathbf{b}}$ alignment. However, one may also note that the vector of the first-order

ray at egress deviates little from the vector of the first-order ray at entry. For reasons detailed in section 5.2.3, this low angular displacement θ_k should effect an SRP vector that is low in magnitude, but with a component acting perpendicular to the incoming ray that is proportionally high (analogous to high λ when at $\alpha = 0°$). The SRP vector in Fig. 5.4b is consistent with this. Furthermore, the direction of SRP is to be expected if one recalls that SRP pertains to the reaction force experienced by a particle, and the negative of this vector can be thought of as the pressure applied to the ray. Given that c is constant, when this pressure effects an increase of ray velocity in one direction, it must decelerate the ray in another direction. The negative of the SRP vector can be seen to correlate with a change in direction of $\hat{\mathbf{c}}$ that conserves $|\mathbf{c}|$. However, some discrepancy would be expected owing to the effect of other rays.

5.2.3 Theoretical Maximums: SRP Profile and SRP Ratio of a Lossless Ray according to Angular Displacement

To aid evaluation, it is useful to define the theoretical maximum performance for these patterns beforehand. This standard is defined according to a lossless ray (no absorption, partial reflection or dispersion) with angular displacement θ_k in a fixed Sun-pointing attitude. The maximum $F_{S_{yb}}$ is $G_{\text{SC}}/c = 4.5698$ μPa as depicted in Fig. 5.5. It is intuitive that this occurs at $\theta_k = 90°$ where ray alignment with $-\hat{\mathbf{y}}_\mathbf{b}$ is greatest. It is also intuitive that $F_{S_{xb}}$ is greatest when $\theta_k = 180°$ and the direction of the incident ray has been exactly reversed, as for a reflective sail with a retrograde Sun. Under these conditions, $F_{S_{yb}} = 0$ Pa and $\lambda = 0$ because $|F_{S_{xb}}| = F_S = 2G_{\text{SC}}/c = 9.1397$ μPa (i.e. $F_{S_{xb}}$ is at the theoretical maximum for SRP, so $\hat{\mathbf{x}}_\mathbf{b}$ must be the only axis in which SRP is acting). What is more noteworthy is that from zero θ_k, $F_{S_{yb}}$ initially increases in magnitude much faster than $F_{S_{xb}}$. This means that, somewhat less intuitively, the less that

a ray has been redirected towards the axis tangential to the sail $-\hat{\mathbf{y}}_\mathbf{b}$, the greater the proportion of SRP that is acting in $+\hat{\mathbf{y}}_\mathbf{b}$. This is represented in Fig. 5.5 by the SRP ratio λ, which decreases after $\alpha = 0 + 1/\infty$ (at $\alpha = 0°$ it is undefined). Indeed, at the maximum $F_{S_{yb}}$, the ratio λ is 0.7071 which is equivalent to $\sin(45°)$, ergo $F_{S_{yb}} = F_{S_{xb}}$. The conclusion that one can draw from this behaviour is that a transmissive sail cannot maximise $F_{S_{yb}}$ and λ simultaneously; either the absolute magnitude or proportional magnitude of $F_{S_{yb}}$ must be sacrificed. Conversely, there is no theoretical maximum torque because the offset from the centre of mass can always be increased.

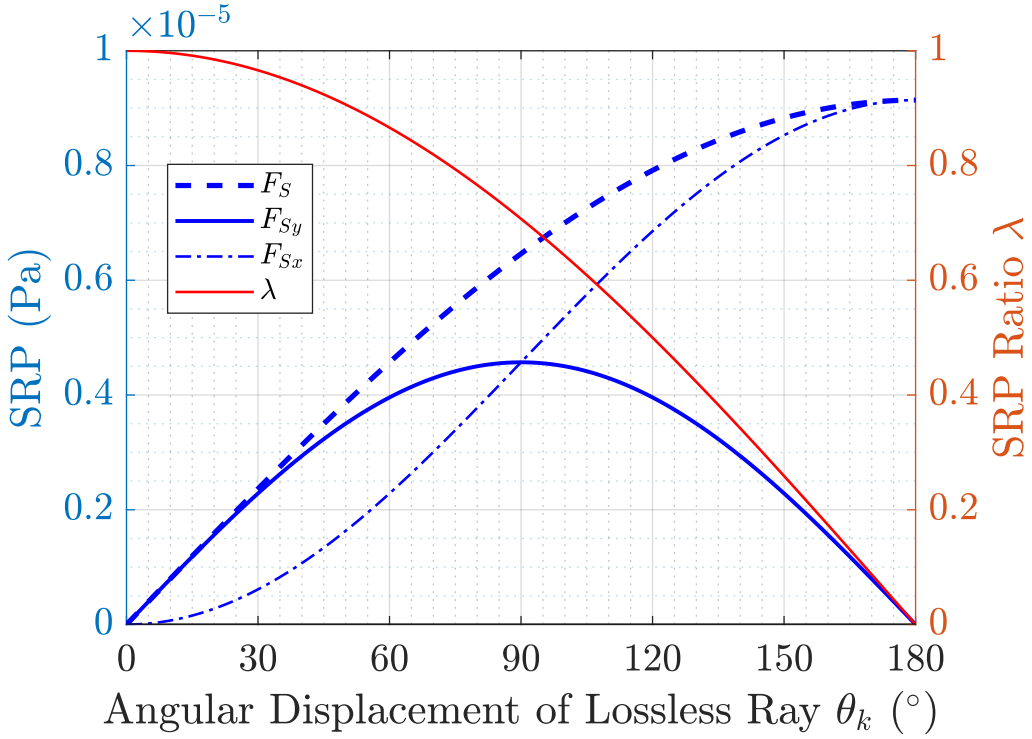

Figure 5.5: SRP arising from Theoretically Lossless Redirection, Body-fixed Reference Frame, $\alpha = 0°$ Assumed

5.3 Generative Micro-prisms for Acceleration

The triangular micro-prism (hereafter, just 'micro-prism') represents the simplest and most scalable interpretation of the transmissive sail. Furthermore, although the focus here is on refractive sails, many diffractive sails harness similar geome-

tries known as *blazed gratings* [31, 33]. Though a refractive sail does not harness diffractive effects, a diffractive one may harness both diffraction and refraction.

The application of micro-prisms within a transmissive sail is the generation of tangential pressure $F_{S_{yb}}$ or active roll torque $\tau_{S_{xb}}$. Micro-prisms are not typically considered for passive control outside of certain *solar collector* configurations comprising prisms of varying size [29]. Active roll torque $\tau_{S_{xb}}$ is achieved by positioning prisms around the circumference of a sail with alternating clockwise or counterclockwise orientations [29]; opacity control devices may be placed above these to achieve active control by preventing the illumination of either the clockwise or counterclockwise prisms. As mentioned in section 5.2, an element that is optimised for $F_{S_{yb}}$ will be optimised for $\tau_{S_{xb}}$ inherently. The optimisation problem therefore simplifies to maximising $F_{S_{yb}}$ over the desired α range.

Each prism begins as a right triangle, but it is inappropriate to describe them using this convention as they rarely stay a right triangle during numerical optimisation. An anatomical convention is instead adopted: the terms *left lateral* edge and *right lateral* edge refer to the edges on either side of the *base* edge, which generally remains horizontal. The term *tilt* refers to the angle between the left lateral and base edges, which is 90° for the initial right triangle. Because diffraction is not considered, and because the following fitness functions do not consider thickness as an argument, the size of the prisms is arbitrary. As such, a shorthand is used for prism dimensions of the following format: *tilt, left lateral to base edge ratio* (e.g. 90°, 1:1 for the initial right prism with equally sized left lateral and base edges). Note that scale is not arbitrary for torque calculation as explored in section 5.5.1. The optimiser was allowed $\mathrm{n_{I\,max}} = 100$ iterations for each of these runs.

5.3.1 Prism Pattern Optimisation for Sun-pointing with Comparison to a Single Prism Solution

These optimisation runs were carried out to explore the effects of pattern propagation. To do this, it is first necessary to see how patterns function in its absence. As such, a single PET prism is illuminated in isolation by 50 rays, and is optimised for a Sun-pointing attitude via the fitness function f_{f1} (Eq. 5.38). This optimises the prism for $F_{S_{yb}}$ at $\alpha = 0°$ and nothing else, which is the broadest interpretation of transmissive sailing fitness (for certain manoeuvres, it may be conducive to optimise for a wider range of α or achieve a specific ratio λ).

$$f_{f1} = F_{S_{yb}}(\alpha = 0°) \tag{5.38}$$

After a brief optimisation, the f_{f1} single prism achieved a maximum tangential SRP of $F_{S_{yb}}(0°) = 4.057$ μPa — which is 88.8% of the theoretical maximum — as shown in Fig. 5.7a. As stated, the profile of the initial prism was a right triangle with a 1:1 ratio between the left lateral and the base. The geometry of the f_{f1} single prism seen in Fig. 5.7b deviated very little from this start point with dimensions 90.3°, 0.965:1. This is on account of the initial right triangle being already near-optimal for a Sun-pointing attitude in the absence of pattern propagation, and so the optimiser had little room for improvement. Why this initial geometry is optimal, and why it changed at all, is discussed below.

A 1:1 right triangle is effective here because all rays are ejected with nearly perfect alignment to $-\hat{\mathbf{y}}_\mathbf{b}$, primarily through TIR on the 45°-inclined right lateral edge; at $\alpha = 0°$, this 45° incline results in $\theta_k = 90°$ after TIR; were it not for losses accrued from partial reflection (which occurs on entry and egress from the prism), this configuration would yield 100% of the maximum tangential SRP and the optimiser would have failed to improve upon it. Generally, it is more efficient to redirect a ray via TIR because a ray undergoing TIR does not have to share

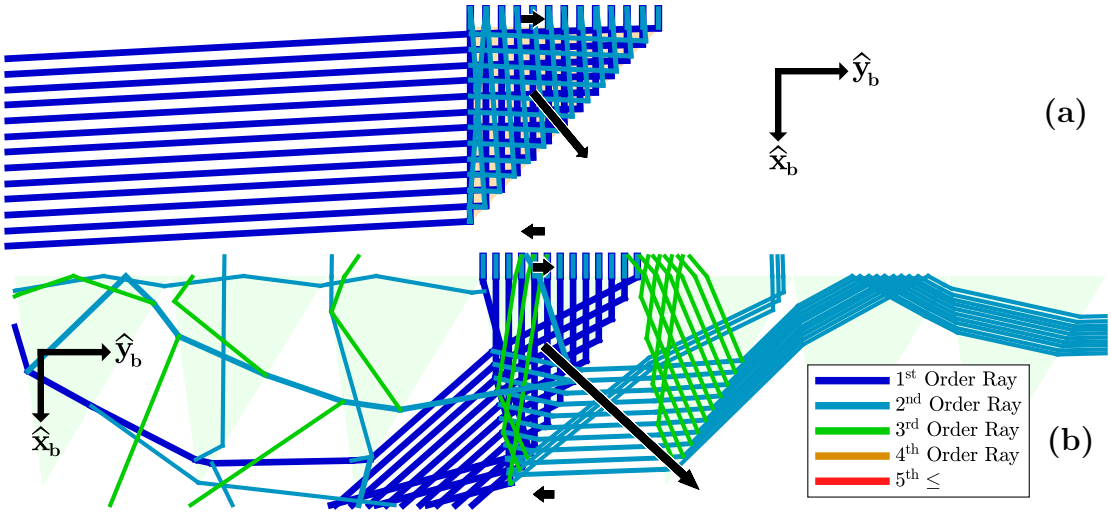

Figure 5.6: Ray Diagram for f_{f1} (a) Single Prism and (b) Prism Pattern, Filter Set to 1%, Reduced Ray Count for Clarity

its flux with any children rays under this model. This may seem like delaying the inevitable — after all, a ray that is made to undergo TIR by a prism will experience the same number of ray-splitting events as one that does not: one on entry, and one on egress. The difference is that a ray that undergoes TIR may exit the particle with very low θ_i with respect to the final boundary, as it does not require refraction on egress to achieve a change in θ_k. Conversely, a ray that undergoes pure refraction will find that the magnitude of θ_k is proportional to θ_i by proxy of Snell's law (Eq. 3.20). As seen in Fig. 5.3, this implies that for the pure refraction case, θ_k is also *inversely* proportional to transmissivity Φ_t. A ray that is refracted very effectively via pure refraction will also incur great loss by conferring a larger portion of its ray flux to the second-order child ray created on egress. Note that if diffuse reflection were considered, additional rays *would* be created by TIR. However, this would be more akin to the dispersion of a polychromatic transmitted ray — which under this model, would be considered to be a continuation of the parent rather than a true child ray. This is because, as for dispersion, diffusely reflected rays do not oppose the motion of a parent that is undergoing TIR — instead the effect is more akin to scattering. Conversely, partial reflection always implies a drastically different ray path to that of its parent (whether specular or otherwise).

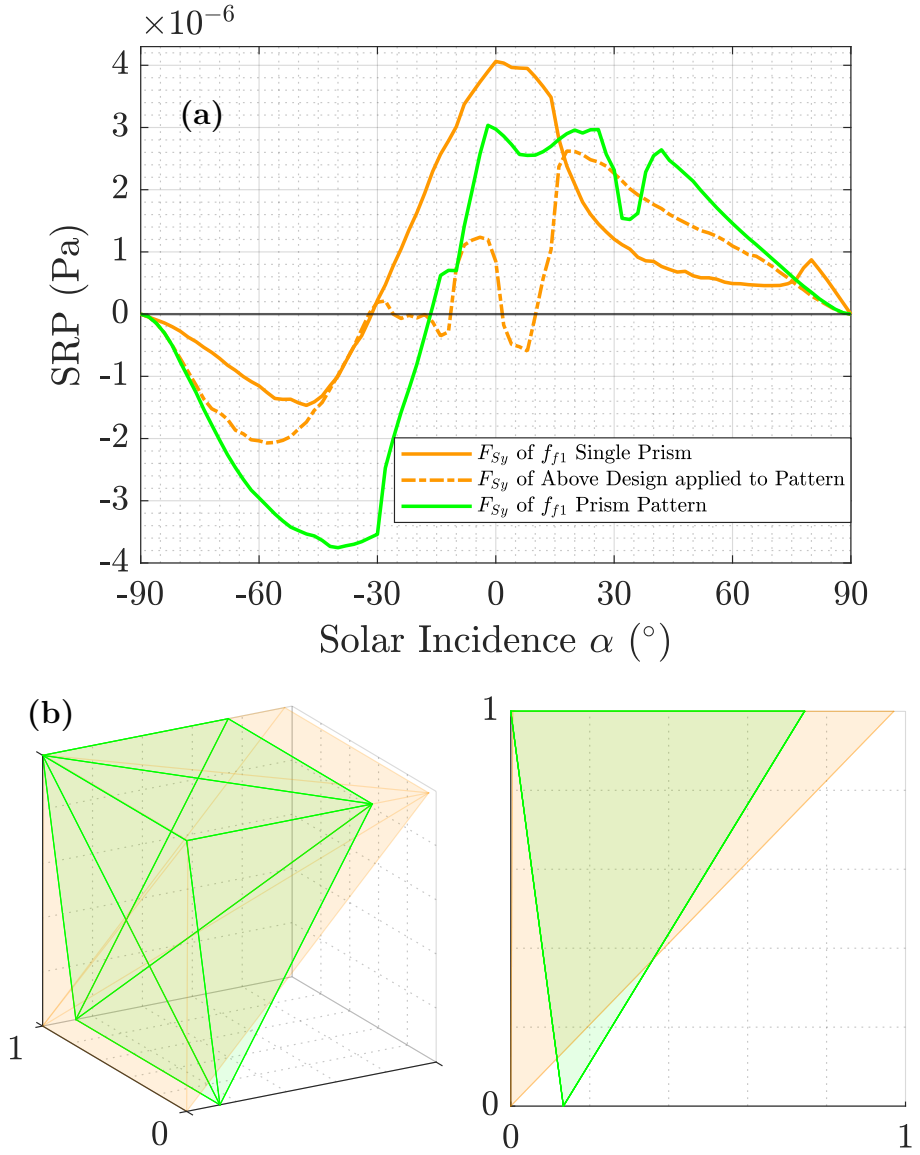

Figure 5.7: Micro-prisms Optimised by f_{f1} in Single and Pattern Form: (a) SRP Profile, (b) Normalised Geometry

Given that the 1:1 right triangle configuration would yield 100% of the maximum tangential SRP under idealised conditions, the fact that the optimiser changed the geometry at all may therefore be surprising. However, the reasoning behind it helpfully demonstrates the nuance of sail pattern optimisation for even very simple scenarios. It is perhaps more surprising that this change occurred to the benefit of second-order rays at the expense of their first-order parents, and yet the overall tangential SRP was increased. Specifically, this change benefited the second-order rays created during the egress of first-order rays from the particle. Because this occurred at low incidence, we can infer that the flux of these chil-

dren was less than $1/10^{\text{th}}$ of G_{SC}/c, making this decision even more puzzling. The first key to this problem is noting that the second-order rays gained slightly more alignment with $-\hat{\mathbf{y}}_\mathbf{b}$ on their egress than the first-order rays lost as a result of this change (see Fig. 5.6a). The second key lies in the $F_{S_{yb}}$ curve of Fig. 5.5: the $|dF_{S_{yb}}/d\theta_k|$ is greatest near $\theta_k = \{0°,\ 180°\}$ and lowest near $\theta_k = 90°$. First-order rays were ejected by the 1:1 prism with $\theta_k = 90°$, where $\pm dF_{S_{yb}}/d\theta_k$ is lowest. Therefore, the smaller $\Delta\theta_k$ that the optimiser effected on the first-order rays resulted in a proportionally even smaller $\Delta F_{S_{yb}}$, despite the flux advantage. Conversely, second-order rays ejected by the 1:1 prism did so from $\theta_k = 180°$ where $-dF_{S_{yb}}/d\theta_k$ is highest, and so the larger $\Delta\theta_k$ effected on these rays engendered a proportionally even larger $\Delta F_{S_{yb}}$, despite their flux disadvantage (note that $\Delta\theta_k$ is negative here).

Next, the same geometry that had been optimised for a single prism was applied to a pattern. It can be seen in Fig. 5.7a that this resulted in a significant performance decrease ($F_{S_{yb}}(0°) = 0.753$ μPa, which is 16.5% of the theoretical maximum). The reason for this discrepancy is that the f_{f1} single prism is optimised to achieve $\theta_k \approx 90°$ for its first-order rays. However, when such a prism is a part of a pattern, these conditions guarantee that every first-order ray will be subject to pattern propagation. By entering the adjacent prism and undergoing TIR on the left lateral, these rays were mostly ejected via the base with very poor $-\hat{\mathbf{y}}_\mathbf{b}$-alignment ($\theta_k < 0°$), resulting in low $F_{S_{yb}}$.

Conversely, Fig. 5.7 also depicts the performance and geometry of a pattern that has been optimised by f_{f1} specifically for pattern form. This pattern achieved ($F_{S_{yb}}(0°) = 3.10$ μPa (or 67.8% of the maximum). In this case, the optimiser favoured a geometry that prevented first-order rays from being propagated through the pattern by ejecting them with a smaller θ_k, as can be seen in Fig. 5.6b. At the same time, it appears to have maximised θ_k to the degree that is possible while still adhering to its no-pattern-propagation strategy. Visually, this

is made apparent by the fact that ejected first-order rays pass about as closely to the adjacent prism as is possible without (with the exception of one ray) ever touching an adjacent prism. This is a common feature of patterns employing this strategy and is referred to as *pattern skimming* hereafter. Maximising the efficacy of pattern skimming is also the goal of the aforementioned analytical optimisation method employed in previous works [29]. What is interesting is that the optimiser deviated from the right triangle structure to good effect: by tilting the left lateral, rays that enter above said left lateral undergo TIR upon it and then, more steeply than its peers, undergo TIR on the right lateral, causing these rays to be ejected with greater θ_k. Because these rays accumulate near the tip of the prism, they can be ejected with greater θ_k than their peers while still avoiding pattern propagation.

Nevertheless, the average θ_k discrepancy between the first-order rays of the single and pattern designs is significant at zero α — so much so that it may appear strange that the performance discrepancy is relatively small. After all, the mean θ_k of first-order rays in the single prism case is nearly double that of the prism pattern case, and yet the performance deficit is only 23.7%. The performance discrepancy being so small is again the outcome of the lesson taught by the Fig. 5.5 curve, namely, that increasing θ_k yields diminishing returns for $F_{S_{yb}}$.

It is also notable that this design does not make effective use of second-order rays in a Sun-pointing attitude, as their propagation through the pattern in the $+\hat{\mathbf{y}}_\mathbf{b}$ direction counteracts the prevailing $F_{S_{yb}}$. However, when reducing α below zero, the right lateral of the right adjacent prism begins to affect TIR on the second-order rays passed to it, preventing further pattern propagation and redirecting the rays with $-\hat{\mathbf{y}}_\mathbf{b}$ alignment. It is for this reason that a performance spike can be seen at $\alpha = -3°$. It seemed reasonable that pattern propagation could be reliably harnessed at a single α as demonstrated here, but whether it could be made reliable for a range of α was less certain according to the seemingly chaotic

nature of pattern propagation.

5.3.2 Prism Pattern Optimisation for Low-α with Comparison to the Analytical Solution

These optimisation runs explore whether micro-prism patterns optimised analytically can be optimised further through numerical means, as speculated by the authors of source [29]. In said source, a prism pattern was optimised for $-10° \leq \alpha \leq 10°$ by employing a strategy that we earlier referred to as pattern skimming. This pattern is depicted at zero α by Fig. 5.8a. In theory, this would imply ejecting rays as close as possible to the tip of the left adjacent prism, thereby maximising $-\hat{\mathbf{y}}_\mathbf{b}$ alignment while preventing pattern propagation. However, this pattern was constrained by two additional requirements: the profile had to be a right triangle, and it had to prevent TIR — as this was said to promote pattern propagation and lead to erratic SRP profiles. As such, θ_k is lower than it may be otherwise. This source prism has a right triangle profile with an edge ratio of 1.64:1. For the sake of consistency, the prisms in this section are made of PS as for the source.

By avoiding both TIR and pattern propagation for the first-order rays over the entire α range, the SRP profile of the source prism is remarkably stable for a larger range of α than it is designed for: for the range $-10° \leq \alpha \leq 90°$, the $F_{S_{yb}}$ curve could be described as exactly that — a curve — with no significant spikes or valleys to speak of. Furthermore, although its $F_{S_{yb}}$ is modest, this does endow it with a much larger λ than other designs (as demonstrated by Fig. 5.5, these things go hand-in-hand). In its own literature, this pattern registered $F_{S_{yb}}(-5°) = 1.60$ μPa maximum and $F_{S_{yb}}(0°) = 1.55$ μPa. In our own model we register a more favourable $F_{S_{yb}}(-2°) = 1.94$ μPa maximum and $F_{S_{yb}}(0°) = 1.90$ μPa. This difference may be due to the source model employing sequential

ray tracing, or otherwise employing different assumptions to those that are used here. Conversely, the shape of the two profiles is nearly identical and the values for λ are largely in agreement — for example, registering $\lambda(0°) = 0.876$ in the source model and $\lambda(0°) = 0.867$ in our own.

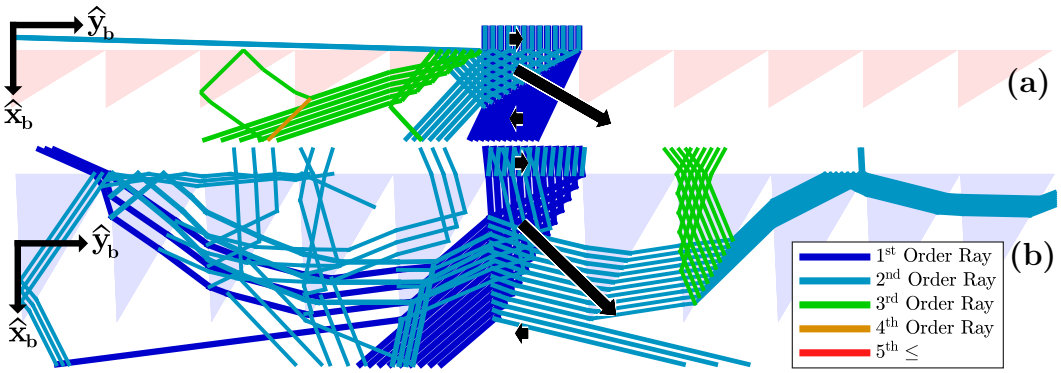

Figure 5.8: Ray Diagram for (a) Source Prism Pattern at $\alpha = 0°$ and (b) f_{f2} Prism Pattern $\alpha = 6°$; Filter Set to 1%, Reduced Ray Count for Clarity

Despite the fact that the analytical method only optimises with first-order rays in mind, this source pattern has very interesting high order ray behaviours. They are notable in that — as if by design — they are highly conducive to tangential SRP, and furthermore, are not restricted to a single α. Said behaviours can be seen to be exhibited by this pattern for a range $-5° \leq \alpha \leq 10°$, thereby answering the question of whether pattern propagation can be reliably harnessed for an α range with an affirmative. The behaviour in question (seen in Fig. 5.8a) is the tendency for second and third-order rays to be ejected with higher θ_k than the first-order rays, through a few different mechanisms. In the zero α case, second-order rays follow one of two prevailing ray paths: the first are the children of first-order rays that enter from the right half of the base. After being created at the right lateral edge, these avoid pattern propagation by undergoing TIR on the base and subsequent ejection from the left lateral (with slightly improved θ_k over the first-order rays). The second group are the children of first-order rays that enter from the left half of the base. After being created, these are ejected before they encounter the base of the original prism. These are then passed to the base of the left adjacent prism, which they encounter with an incidence that is only

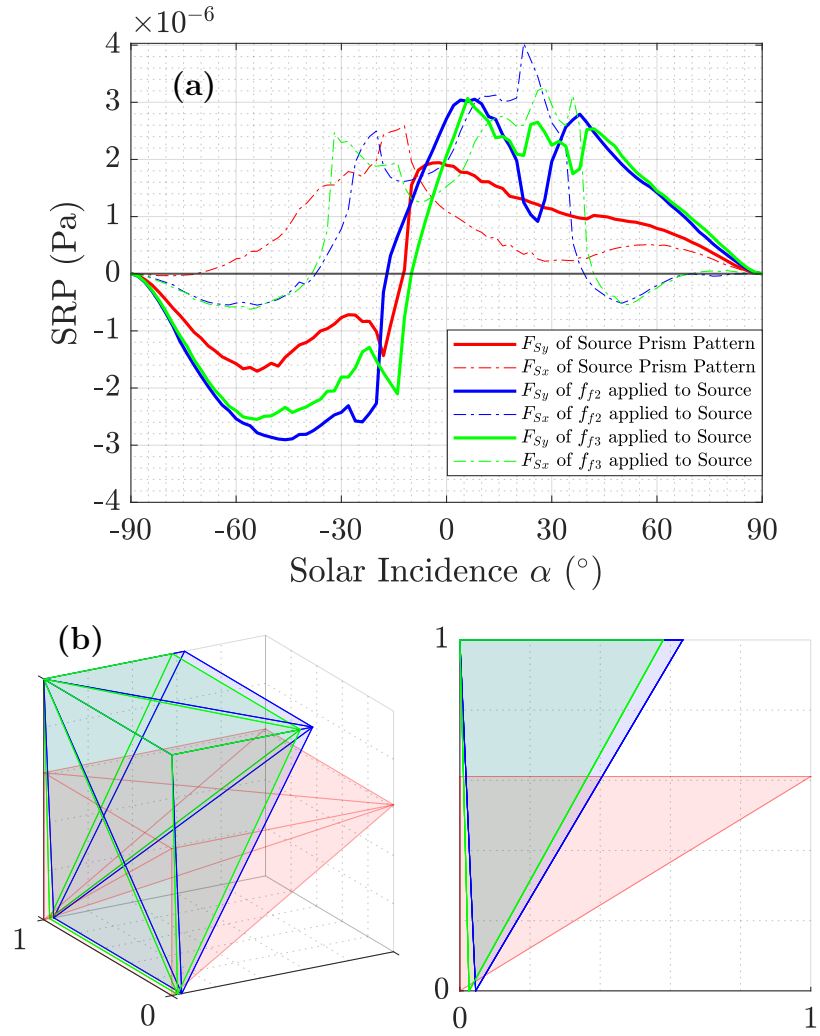

Figure 5.9: Source, f_{f1} and f_{f2} Prism Patterns: (a) SRP Profile, (b) Normalised Geometry

just low enough to avoid TIR; as such, the effect of refraction is amplified and they are ejected with a highly favourable $\theta_k \approx 90°$. However, because the angle of incidence is barely tenable for transmission, a large portion of their power is distributed to third-order rays, which helpfully mimic the behaviour of the first group of second-order rays — and with even higher θ_k than said group.

This system is not perfect, and the occasional third-order ray is ejected with an unfavourable ray path — but these are outliers. Furthermore, if one moves the ray filter to 0.5%, a family of third-order rays can be seen being ejected via the right lateral prism with negative θ_k, but these are of lower power than those that do behave favourably. Overall, this pattern is notable amongst its

peers for having so many divergent ray paths that all ultimately end up to the benefit of performance. It is somewhat paradoxical that the pattern that sought to avoid complex ray paths and TIR ended up with some of the most optimal high order ray paths brought about by TIR. Their existence was only obfuscated by sequential ray tracing, which would not have tracked their progress through the pattern.

Two optimisation runs were carried out using this source prism as the initial geometry (see Fig. 5.9). The first used fitness function f_{f2}, which is simply f_{f1} applied over the range $-10° \leq \alpha \leq 10°$ (Eq. 5.39). This pattern is depicted at zero α by Fig. 5.8b. The second used f_{f3}, which seeks to maximise $F_{S_{yb}}$ while minimising $F_{S_{xb}}$ over this same α range so as to generate designs with more favourable λ (Eq. 5.40).

$$f_{f2}(\mathrm{n_I}) = \int_{-10}^{10} F_{S_{yb}}\, d\alpha \tag{5.39}$$

$$f_{f3}(\mathrm{n_I}) = \int_{-10}^{10} (F_{S_{yb}} - |F_{S_{xb}}|)\, d\alpha \tag{5.40}$$

The f_{f2} and f_{f3} prism patterns were finished with dimensions 87.4°, 0.634:1 and 88.5°, 0.579:1, respectively. These differed only marginally in performance as well as geometry, with the f_{f2} prism registering slightly higher $F_{S_{yb}}$ near zero α, but with the f_{f3} prism demonstrating lower $F_{S_{xb}}$ for the $\alpha \pm 10°$ range. Compared to the source pattern, both of the numerically optimised prisms yielded significantly higher maximum tangential SRP, with $F_{S_{yb}}(8°) = 3.05$ μPa and $F_{S_{yb}}(6°) = 3.07$ μPa, respectively. However, only the f_{f2} pattern significantly improved upon zero α performance with $F_{S_{yb}}(0°) = 2.72$ μPa; the f_{f3} pattern represented only a marginal improvement with $F_{S_{yb}}(0°) = 2.07$ μPa. These values represent an improvement of 57% and 58% in the maximum case and 43% and 9% in the zero α case, respectively.

Conversely, both of the numerically optimised patterns saw a significant decrease

in λ. For example, at zero α the f_{f2} and f_{f3} prisms demonstrated $\lambda(0°) = 0.813$ and $\lambda(0°) = 0.805$, representing a 22% and 23% decrease relative to the source pattern, respectively (the f_{f3} pattern is usually better than the f_{f2} pattern in this respect, but its performance at zero α is greatly hampered). Moreover, while the f_{f3} prism did sacrifice λ less than its peer for the majority of the $\alpha \pm 10°$ range, it improved $F_{S_{yb}}$ far more than it decreased $F_{S_{xb}}$ despite both arguments having equal weight within the fitness function. This suggests that there was more reward to be found in increasing the former than there was in trying to maintain λ. As such, converging upon a high-λ design through numerical means would likely require a fitness function with bias towards λ or against $|F_{S_{yb}}|$.

Fig. 5.8b demonstrates the ray paths for the f_{f2} prism at $\alpha = 6°$; here the f_{f3} pattern behaves nearly identically and so it is not shown. As for before, these patterns improve performance principally by pattern skimming its first-order rays in an optimal manner. Furthermore, these patterns do not harness pattern propagation consistently in a single direction as for the source pattern; at certain α, favourable pattern propagation through the left and the right adjacent prisms may occur simultaneously. Pattern propagation through the right adjacent prisms occurs reliably across the $\alpha \pm 10°$ range, and serves to mitigate the otherwise counterproductive second-order rays that are produced on first-order ray egress. They do this by capturing these on the immediately right adjacent prism and passing them (with ever increasing θ_k) to the next adjacent prism, and so on. By the third adjacent prism, their angle of incidence is too large for transmission (exceeding the critical angle θ_c) and via TIR they are passed to the base, from which they are ejected with their negative θ_k having become slightly positive. As α is increased, this mechanism becomes more effective as more of rays are captured by the immediately right adjacent prism.

Conversely, pattern propagation through the left adjacent prisms occurs opportunistically as α increases, beginning at $\alpha = 1°$ for the f_{f2} pattern and 5° for

f_{f3}. At these thresholds, a single ray is passed through a single adjacent prism, increasing θ_k slightly. At higher α, this same ray will be passed through several adjacent prisms, increasing θ_k significantly and effecting a curved ray path. This ultimately culminates in an ejection from the base edge of an adjacent prism with greatly improved $-\hat{\mathbf{y}}_\mathbf{b}$ alignment. Furthermore, as α increases, a greater proportion of first-order rays are subject to pattern propagation through these means. It is for these reasons that the performance of these patterns is much higher at $\alpha = 6°$ than it is at zero α. For example, at $\alpha = 8°$, nearly half of all first-order rays will be pattern propagated by the f_{f2} pattern.

Furthermore, the performance drop-off seen at $8° \leq \alpha \leq 38°$ can be attributed to this mechanism beginning to fail (this is present for both numerically optimised designs). This occurs because increasing α also increases the 'curvature' of these sequentially diverted ray paths, ultimately causing them to undergo TIR at the base edge of a pattern element where they would otherwise be ejected. This may seem counterintuitive because increasing ray path curvature should imply a smaller angle of incidence with the base edge, preventing TIR rather than enabling it. However, increasing this curvature reduces how far along the $-\hat{\mathbf{y}}_\mathbf{b}$ axis a ray can travel before it encounters a base edge. Eventually, this causes the ray path to encounter the base edge an entire prism 'sooner' than otherwise. This is a problem because refraction on the left lateral of the penultimate prism gives the curved ray path the additional curvature it needs to encounter the base at a low enough incidence angle for transmission. So, somewhat paradoxically, too much *initial* curvature of a ray path can situationally reduce the *overall* curvature of the ray path, leading to shallower incidence with the base edge and promoting an unfavourable TIR interaction.

Ultimately, in terms of θ_k the pattern propagation demonstrated by these numerically optimised prisms is not necessarily an improvement over that which is demonstrated by the source prisms — and when it is, it is typically very slight.

Furthermore, the numerically optimised patterns may incur greater loss by utilising more prisms within their pattern propagated ray paths. This implies more children rays to distribute ray flux to. However, while the source pattern harnesses pattern propagation to better the θ_k of second and third-order rays, the numerically optimised prisms improve the θ_k of higher flux, first-order rays, to a proportionally greater effect. This is the first example of a prism pattern harnessing pattern propagation for its first-order rays, and furthermore, these patterns manage to make use of this mechanism over a range of α.

5.4 Generative Freeform Optical Elements for Acceleration

Geometries optimised for a range of α would be expected to demonstrate a reduced sensitivity to α, but lower zero α performance when compared to those that were simply optimised for a Sun-pointing attitude. However, differences in sensitivity were marginal during the preceding micro-prism optimisation runs, while zero α performance deficits were typically more substantial. Furthermore, there is little incentive to optimise for a larger α range due to the diminishing performance that is found at larger α values, and because single α optimisation is faster, this section focuses on Sun-pointing designs optimised with f_{f1}. Making the fitness function a constant also allows one to explore the efficacy of other optimiser parameters, such as initial geometry and 'greedy' optimisation strategies.

To validate that the numerical optimisation of freeform optical elements would be effective, a simple test was carried out for the first optimisation run. A single PET cylinder is optimised for Sun-pointing attitude via the fitness function f_{f1} while illuminated by only a single ray. This scenario is depicted by Fig. 5.10. This resulted in a strange approximation of a tilted lightfoil (see Fig. 5.11)

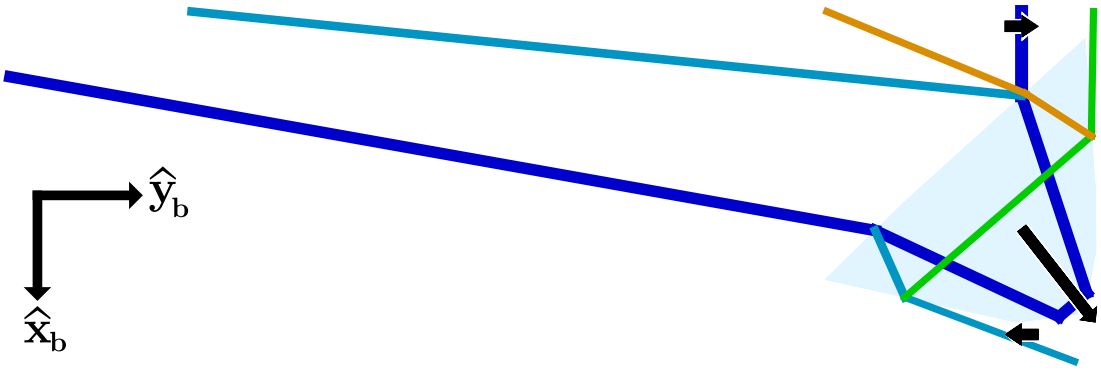

Figure 5.10: Ray Diagram for a Single Freeform Optical Element at $\alpha = 0°$; Filter Set to 1%, Single Ray

that registered an extremely high $F_{S_{yb}}(0°) = 4.17$ μPa or 91% or the theoretical maximum.

It can be seen in Fig. 5.10 that the optical element achieves this by ejecting both its first-order and highest flux second-order rays with a significant $-\hat{\mathbf{y}}_\mathbf{b}$ component, which is allowable given that pattern propagation is not a factor here. It is notable that this is the first geometry to harness external reflection as a means of generating tangential SRP, which it achieves via a straight base edge inclined by roughly 45°. Naturally, it is much easier to optimise for a single ray and this design would likely be ineffective when illuminated in a realistic manner that complied with the axioms of section 3.2.6. Nonetheless, this demonstrated that the optimiser was effective at exploring the much larger solution space of a freeform optical element.

5.4.1 Patterns of Freeform Optical Elements Optimised with f_{f1} to Compare the Effects of Initial Geometry and Optimiser Greed

The larger solution space provided by freeform optical elements would imply a greater number of local maxima that a numerical optimiser could fall into. For this reason, it is worth exploring the effect of varying the starting conditions.

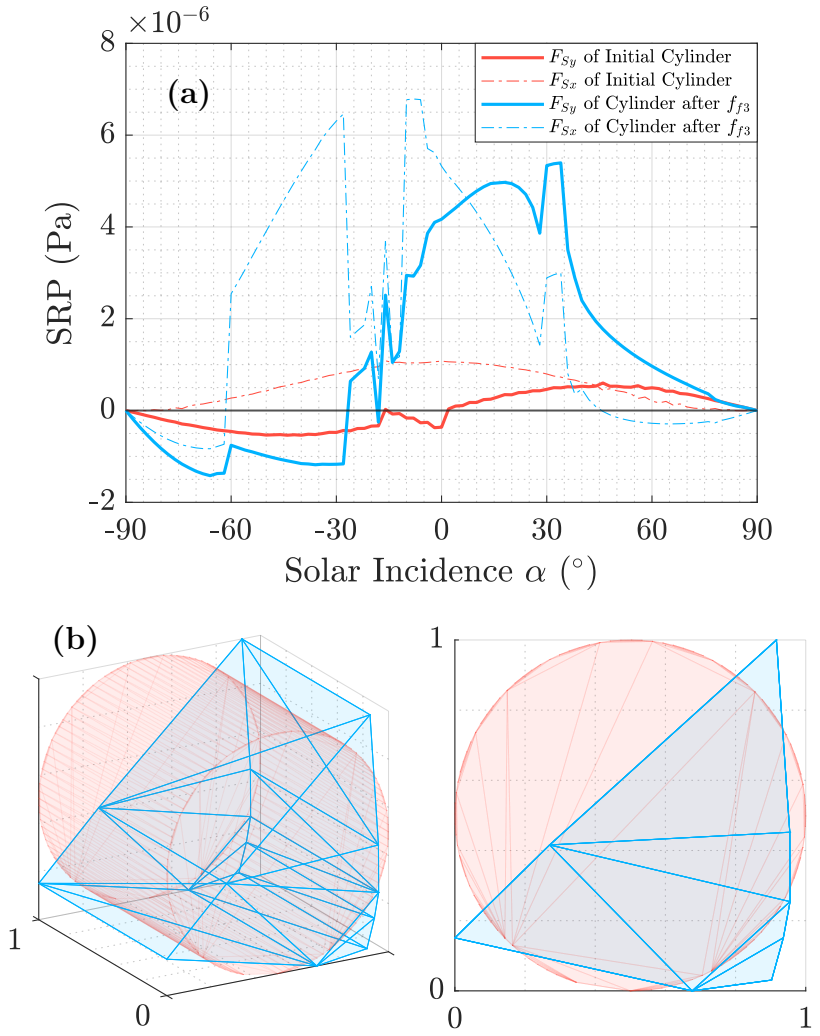

Figure 5.11: Freeform Element Optimised for f_{f1} with a Single Ray, Starting as a Cylinder, Before and After: (a) SRP Profile, (b) Normalised Geometry

In this section, the pattern begins as either a cylinder or a prism-like freeform element and the outputs of the two are compared. Furthermore, the concept of optimiser *greed* was introduced to mitigate against long optimisation times: when enabled, this ensures that during phase two of optimisation (the exploitation phase), the optimiser will repeat any successful action that it finds until it cannot gain any more reward from that action (reducing the size of the action where appropriate to prevent overshoot). This somewhat skews the exploration-exploitation balance of phase two even more heavily towards exploitation. On the one hand, this can greatly speed up optimisation and, if only a small $n_{I\,max}$ is allocated to the optimisation run, may improve the output of the optimiser. On the other hand, the emphasis placed upon exploitation could render the op-

timiser more prone to falling into local maxima. For the sake of comparison, the cylinder-derived and prism-derived freeform elements were optimised twice, once by a greedy optimiser and once by a non-greedy one. The optimiser was also allowed $n_{I\,max} = 100$ iterations as for the structured prisms. This disadvantages the freeform elements in principle, as a larger solution space would imply that a greater number of iterations will be needed for convergence — particularly for the exploration phase. However, to explore the efficacy of the optimiser as an optimiser of freeform elements versus structured elements, the parameter was not changed. Because of the large amount of data depicted, $F_{S_{xb}}$ is neglected from these profiles and the triangulated shapes are replaced by opaque patches (see Fig. 5.13). The most successful of these patterns is depicted at zero α by Fig. 5.12.

If viewed through the lens of the fitness function, it can be seen from Fig. 5.13a that the non-greedy and cylinder-derived patterns were the most successful at zero α (with the non-greedy cylinder-derived pattern being the most successful of all, and the greedy prism-derived pattern being the least successful of all). The non-greedy cylinder-derived pattern is noteworthy for exceeding the maximum and zero-α tangential SRP of all prior patterns that considered pattern propagation, with $F_{S_{yb}}(6°) = 3.56$ μPa maximum and $F_{S_{yb}}(0°) = 3.17$ μPa, which are 78% and 69% of the theoretical maximum, respectively. The performance of the other freeform elements was only middling, and generally inferior to their structured prism counterparts. This seems to confirm the hypothesis that a numerical optimiser can achieve greater performance through freeform elements, but at the cost of longer computation and perhaps repeated optimisation runs due to the heightened chance for an optimiser to fall into a local maximum. The greedy configuration appears to make this more likely. Anecdotally, in subsequent tests it was seen to be rare (but not unheard of) for the greedy optimiser to outperform the normal one with $n_{I\,max} = 100$. Conversely, it was not uncommon for the greedy and non-greedy optimisers to converge upon the same result.

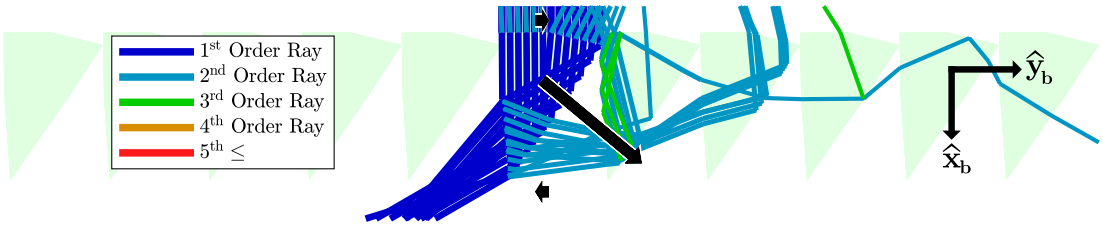

Figure 5.12: Ray Diagram for Freeform Optical Elements at $\alpha = 0°$, Normal/Non-greedy, Cylinder-derived, Control Point Mode; Filter Set to 1%, Reduced Ray Count for Clarity

Focusing on the most successful pattern: it can be seen in Fig. 5.12 that this performance is achieved by pattern skimming the first-order rays and pattern propagating the second and third-order rays. No attempt is made to pattern propagate the first-orders. The reason that this pattern exceeds the performance of prism patterns that utilised similar strategies is a result of it having a cambered base and left lateral edge. These two cambers can be seen to work in tandem.

The base camber causes rays to be concentrated near the tip of the optical element; rays that enter on the right half of the base edge are diverted away from the right lateral so that they encounter and reflect off of it further towards the tip. Meanwhile, rays that enter on the left half of the base edge are unimpeded. This causes variable θ_k depending on point of egress. This behaviour was also demonstrated by non-right triangle prisms, but the effect is more pronounced here. In fact, the ray concentration effect makes the pattern skimming strategy appear less aggressive, as there is still room for higher θ_k. Indeed, this is the reason why the performance continues to improve up to $\alpha = 6°$, whereupon the pattern skimming effect is maximised. The slight camber on the left lateral edge slightly amplifies this effect. Furthermore, the left lateral camber causes partially reflected, second-order rays to converge, such that rays being created near the tip begin their life with a larger $\hat{\mathbf{x}}_\mathbf{b}$ component than those that begin their life higher up within the optical element. This second-order convergence ensures that all rays are captured by the right adjacent optical element, and these are subsequently ejected with a small, positive θ_k.

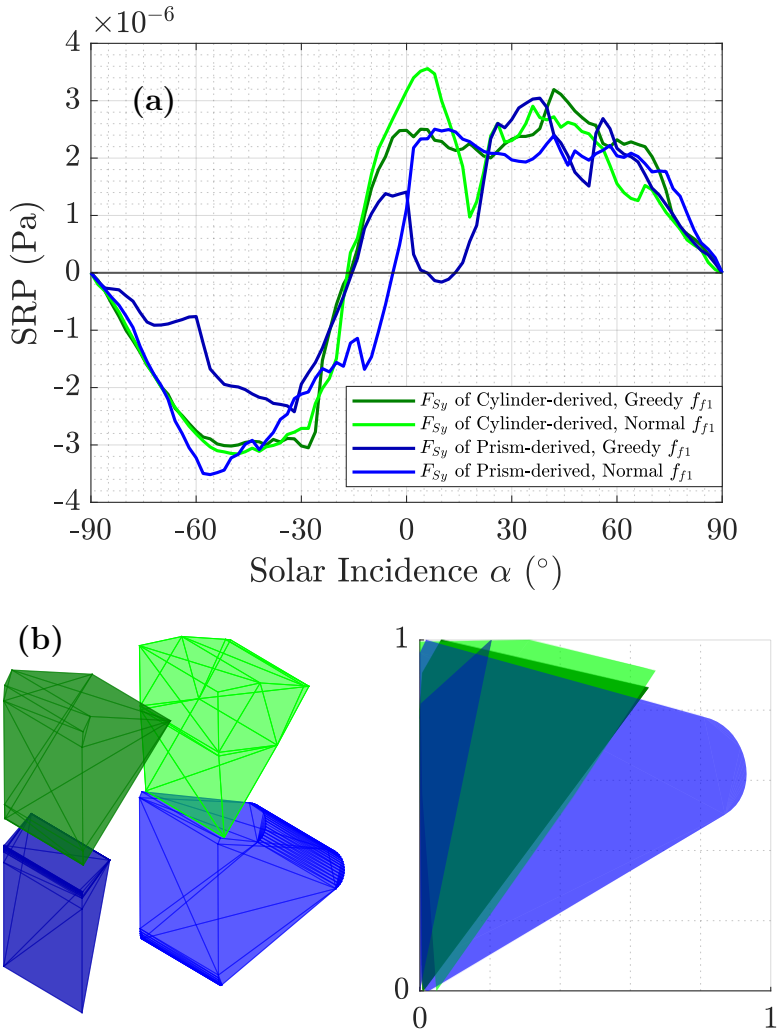

Figure 5.13: Pattern of Freeform Elements Optimised for f_{f1}, Cylinder or Prism-derived, Greedy and Normal: (a) SRP Profile, (b) Normalised Geometry

However, it is noteworthy that the base camber is very low resolution, implying that the optimiser did not effectively distribute its vertices here. Because these shapes typically grow as they are optimised, the vertices that were not utilised will typically be pushed inwards towards the centre of the shape after the geometry has been normalised. This causes the convex hull algorithm to skip them during triangulation and so they cease to influence fitness. Actions taken to effect such vertices will yield zero reward unless said action manages to bring them back into the near-perimeter region, where the convex hull algorithm will consider them. This becomes less likely with each action for which they are not utilised, and so eventually, the optimiser largely stops trying to use them. This could be mitigated by employing an algorithm that redistributes vertices after each iteration while

not changing the shape of the convex hull.

5.4.2 Patterns of Freeform Optical Elements Optimised with f_{f1} Utilising Direct Vertex Editing

Another variable that affects freeform element optimisation is the type of node that the optimiser is allowed to interact with — the vertices, or the spline control points. As previously mentioned, using control points greatly reduces the size of the action space and speeds up optimisation. As such, this was used as default. However, not allowing the optimiser to interact directly with vertices implies that some control may have been relinquished from it. This section explores the effect of allowing the optimiser to 'brute-force' generative shapes by enabling it to interact with every vertex directly. Two of these patterns are depicted by Fig. 5.14a-b at their optimal α.

Fig. 5.15 demonstrates two geometries that were created using vertex mode, wherein one was generated from a cylinder and the other from a prism-analog. Notably, both are very high performance: the cylinder-derived pattern yielded $F_{S_{yb}}(6°) = 3.78$ μPa maximum and $F_{S_{yb}}(0°) = 2.99$ μPa, or 83% and 65% the theoretical maximum, respectively. Meanwhile, the prism-derived pattern yielded $F_{S_{yb}}(0°) = 3.40$ μPa — which was also its maximum — which is 74% of the theoretical maximum. This attributes the highest maximum tangential SRP so far to the cylinder-derived, vertex mode pattern, and the highest zero-α tangential SRP so far to the prism-derived, vertex mode pattern. Given that both of these geometries yielded record breaking performance, whereas only one in four of the control point mode patterns did likewise, this methodology appears to be preferable for performance. Furthermore, while control point optimisation is faster per-optimisation-run, if it takes four optimisation runs to achieve the same performance as a single vertex mode run, then effectively it is slower. Much as for

optimiser greed, streamlining optimisation seems to speed up computation at the cost of optimiser efficacy.

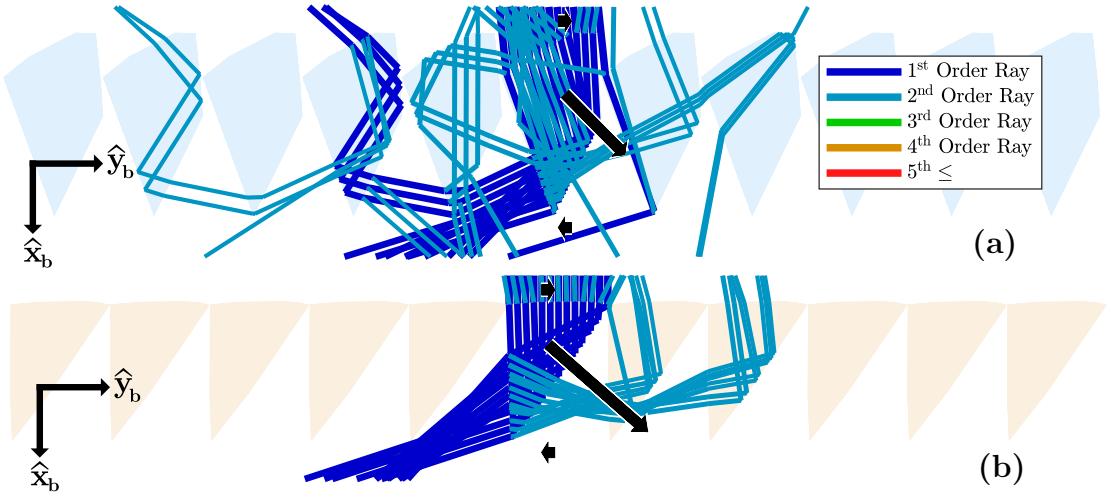

Figure 5.14: Ray Diagram for Freeform Optical Elements using Vertex Mode (and Greedy Mode), (a) Cylinder-derived at $\alpha = 6°$ (b) Prism-derived at $\alpha = 0°$; Filter Set to 1%, Reduced Ray Count for Clarity

The strategy employed by each pattern differs: in Fig. 5.14a, it can be seen that the cylinder-derived optical element — which deviates significantly from the usual prism-analog archetype — favours a variety of different mechanisms to achieve its extremely high, off-zero α tangential SRP. In fact, it utilises most of the aforementioned strategies to some degree: external reflection is harnessed using an inclined base edge in a similar manner to the calibration optical element seen in Fig. 5.10. It can be seen that pattern skimming is the prevailing strategy for first-order rays, as the majority of first-order rays do not undergo pattern propagation. This is achieved with high θ_k by concentrating the first-order rays near the tip of the optical element, before facilitating ejection via a $\Delta\theta_k \approx 45°$ TIR (ray convergence is also present, facilitating some degree of per-ray θ_k fine-tuning, but it is cruder here than for other patterns). However, this strategy is pursued very aggressively, such that at $\alpha = 6°$, several first-order rays do undergo pattern propagation. Unlike the prism pattern seen in section 5.3.2 (Fig. 5.8b), pattern propagation is not used to improve the θ_k of the effected rays; instead, it simply mitigates the potentially negative effects very effectively. It does this so

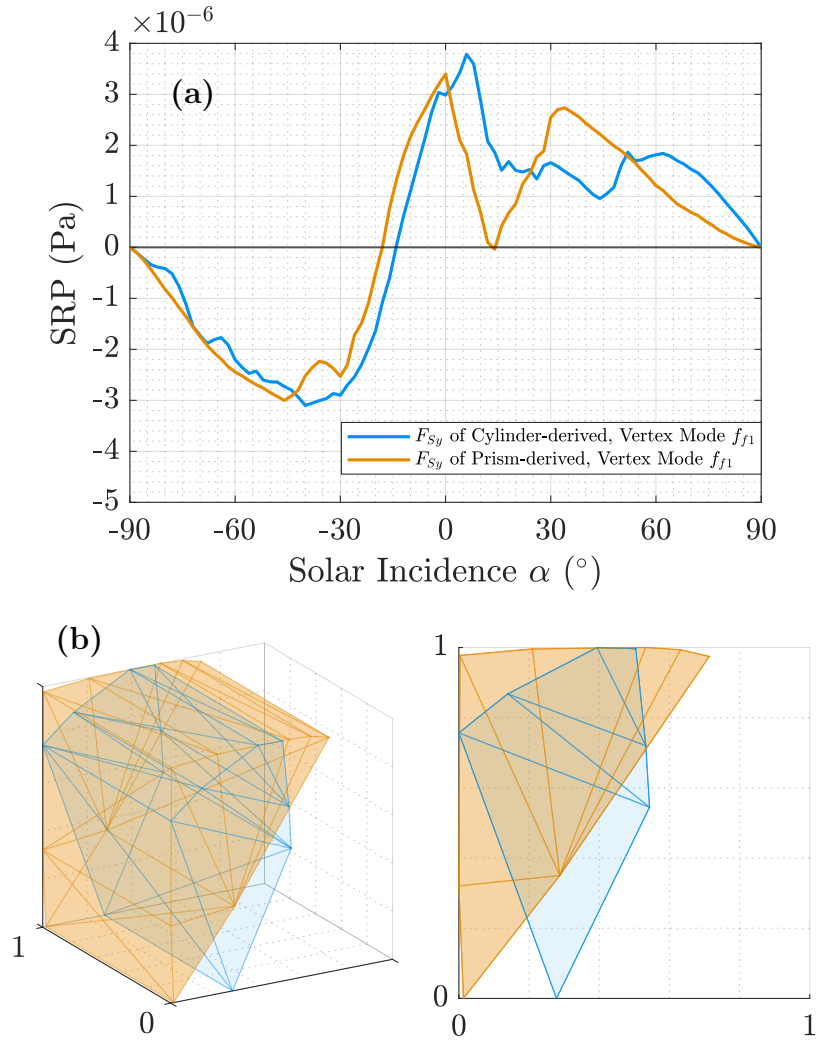

Figure 5.15: Pattern of Freeform Elements Optimised for f_{f1}, Cylinder or Prism-derived in Vertex Mode: (a) SRP Profile, (b) Normalised Geometry

successfully that the small θ_k loss experienced by the affected first-order rays is lower than the gain of θ_k experienced by the unaffected first-order rays. Losses incurred by children rays during pattern propagation may otherwise diminish the effectiveness of this, but it can be seen that these are also predominantly ejected with positive θ_k (externally reflected rays created on adjacent particle entry do have negative θ_k, but as the first order rays enter with nearly zero-θ_i these partially reflected children are very lower power).

Conversely, it can be seen in Fig. 5.14b that the prism-derived optical element largely retained the shape of a prism and exhibits a more conventional pattern skimming behaviour. It achieves this through a cambered base and left lateral

edge to effect first-order ray convergence for optimal θ_k without pattern propagation. This convergence also applies to the second-order rays which exhibit a mitigating pattern propagation behaviour. In fact, this pattern is essentially the same as the most successful control point mode pattern that was converged upon in 5.4.1, namely, the normal (non-greedy), cylinder-derived pattern. The fact that both have converged upon a similar geometry suggests that this configuration is either a particularly attractive local maximum, or perhaps even a globally optimal solution. However, this vertex mode geometry appears to be better optimised for zero-α, as its pattern skimming behaviour is more aggressive (for most rays, a small increase in θ_k would induce pattern propagation). The reason that the vertex version is more effective is that it is better at distributing its vertices, and renders fewer vertices redundant for convex hull calculation during optimisation. As such, the final geometry is higher resolution, and so the fine-tuning of θ_k for each ray can be made finer. This is most visible on the base edge: in Fig. 5.15, it can be seen that six vertices form the cambered base edge. Conversely, in the earlier Fig. 5.13 it can be seen that the equivalent geometry utilised just three vertices (and so its camber would be more accurately described as a chamfer). The other manner in which the vertex version is superior is that its cambered edge is, on average, parallel with the $\hat{\mathbf{y}}_\mathbf{b}$ axis, meaning that partially reflected children rays created on particle entry are reflected with $\theta_k \approx 180°$ when at zero α. Conversely, it can be seen in Fig. 5.12 that the control mode equivalent has a cambered edge that has a clockwise tilt, and so its partially reflected children rays are on average reflected with a negative θ_k, which counteracts the prevailing direction of $F_{S_{yb}}$.

5.4.3 Patterns of Freeform Optical Elements Optimised with f_{f1} and Constrained by Base Edge Colinearity

Another consideration for a patterned sail is its ease of manufacture, which is particularly important for very large sails. Unlike the structured prisms in section 5.3, the freeform optical elements explored in this section do not provide a flat edge that is colinear to that of its adjacent optical elements in its pattern. A flat edge implies that a pattern is compatible with one-sided pattern transfer processes — such as conventional nanoimprint lithography (NIL) — while an irregular edge would imply that a pattern is only compatible with double-sided pattern transfer processes. This may be problematic because double-sided pattern transfer methods are not as widely available [118]. To ensure designs are generated of a sufficiently high TRL, we therefore constrain the optimiser to the generation of flat patterns. Fig. 5.17 demonstrates two geometries that were constrained in this way, wherein one was generated from a prism-analog and the other from a cylinder. The cylinder-derived variant was not particularly effective, so the focus here is on the prism-derived variant. Fig. 5.16a shows the ray diagram for this pattern at $\alpha = 0°$.

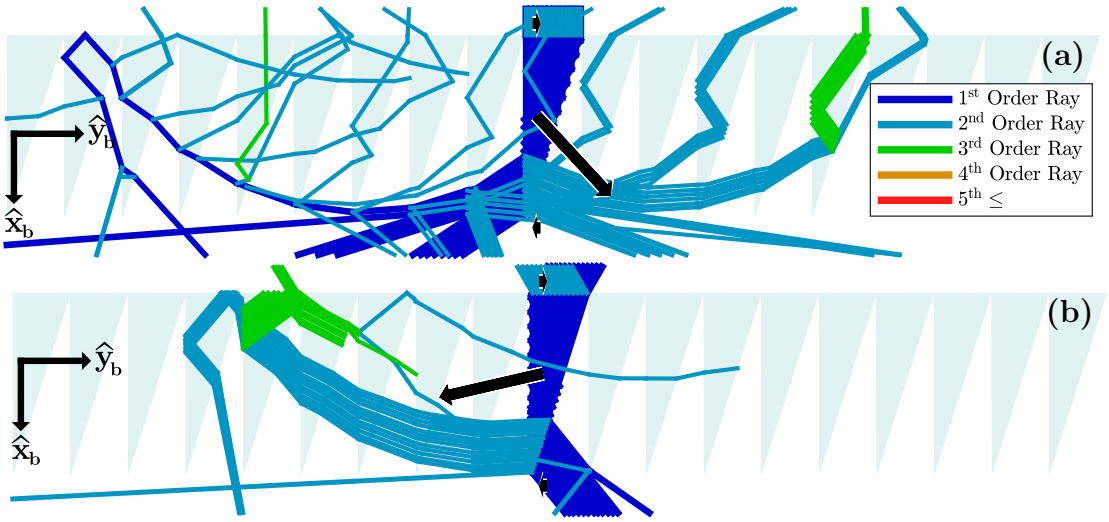

Figure 5.16: Ray Diagram for Prism-derived Optical Element (a) at $\alpha = 0°$, (b) at $\alpha = -30°$; Filter Set to 1%, Reduced Ray Count for Clarity

This prism-derived pattern yielded $F_{S_{yb}}(0°) = 2.88$ μPa — which is 63.0% of

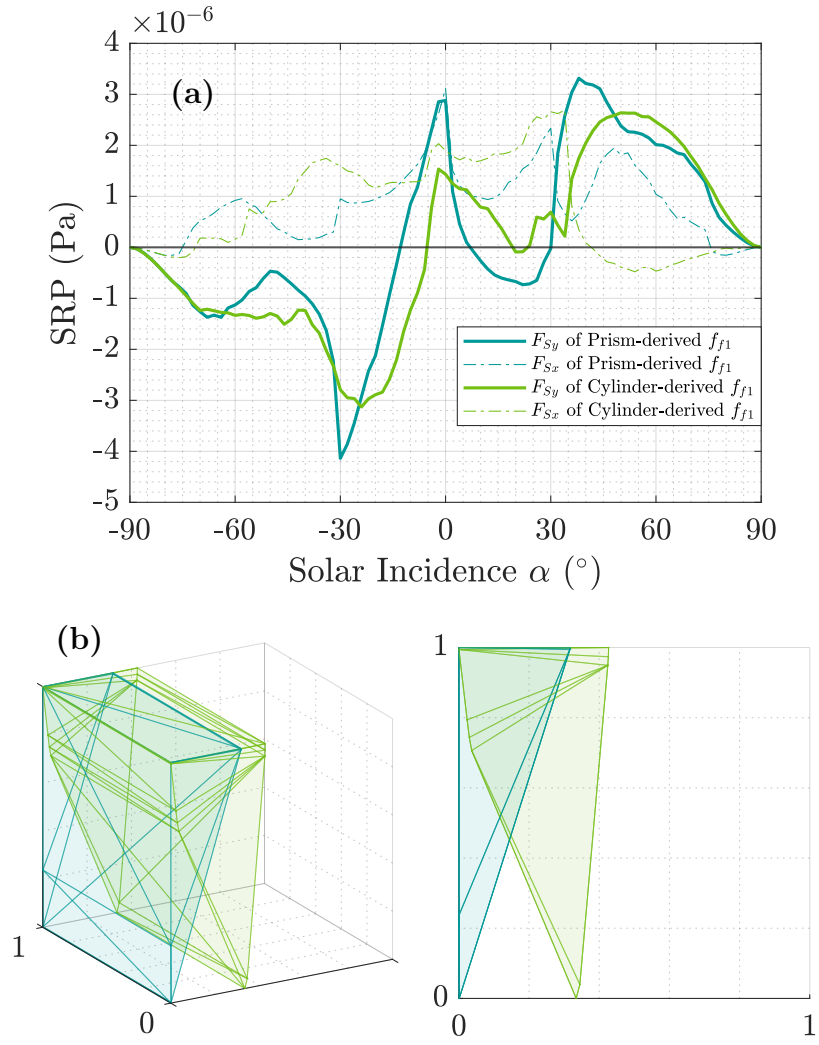

Figure 5.17: Pattern of Constrained Freeform Elements Optimised for f_{f1}, Cylinder or Prism-derived: (a) SRP Profile, (b) Normalised Geometry

the theoretical maximum. This performance is similar to that of the higher performance micro-prism patterns from section 5.3, albeit with much greater sensitivity to α. This performance makes sense because the final geometry can hardly be called freeform: its shape is exactly that of a prism, and it can be seen that the optimiser ended up utilising only four vertices (of which two lie on a colinear edge — so effectively, only three vertices were used). Overall, it appears that the optimiser was not well configured for this constraint, and the issue likely pertains to the manner in which flatness was ensured: vertices above a certain z were removed from convex hull calculation. Effectively, this not only initially reduces the number of vertices available to the optimiser, but it also gives the

optimiser another avenue for rendering them redundant. In the future, it may be better to constrain this pattern in another way. For example, two vertices could be fixed at a certain z to comprise the flat edge. Of course, this would require fine tuning as it is conceivable that the optimiser would attempt to make those fixed vertices redundant, thereby circumventing this constraint and producing irregular edged patterns once again.

Furthermore, the torque profile was unusual in the prism-derived case because the number of marches and the number of elements within had to be increased in order to allow the rays to adequately escape the finished pattern. This implies a greater discrepancy between the true performance, and the performance that was describing fitness to the optimiser near the end of optimisation, which may have contributed to its middling performance. The reason that the marches and elements needed to be increased is that these high aspect ratio prisms promote very long ray paths via a lengthy pattern propagation sequence. None of the pattern propagated ray paths are particularly conducive to $F_{S_{yb}}$, and indeed the primary mechanism for generating SRP here is pattern skimming. However, these pattern propagated rays do at least have a roughly neutral contribution to SRP rather than an outright negative one.

Furthermore, this pattern yielded an interesting behaviour by chance rather than by design: Fig. 5.16b demonstrates the behaviour of this pattern at $\alpha = -30°$, where it can be seen in Fig. 5.17a that this corresponds with $F_{S_{yb}}(-30°) = -4.137$ μPa. If we were to mirror this geometry about z or x_b, this would correspond with $F_{S_{yb}}(30°) = 4.137$ μPa. This is very high efficiency, and achieves 90.5% of the theoretical maximum. On the one hand, efficiency would be expected to be lowered at this α due to effective area attenuation. On the other hand, the theoretical maximum performance profile of Fig. 5.5 was predicated upon zero α and, as discussed, said profile is actually applicable to any α if one renames $\{F_{S_{xb}}, F_{S_{yb}}\}$ to $\{F_{S_{xS}}, F_{S_{yS}}\}$ - i.e. the components of SRP parallel and

perpendicular to the incoming rays. The reason efficiency is able to get so high here is that the component $F_{S_{xS}}$ (which can reach up to $2G_{\mathrm{SC}}/c$ Pa) has a significant component that contributes to $F_{S_{yb}}$. This is in contrast to the $\alpha = 0°$ case, for which $F_{S_{yb}} = F_{S_{yS}}$ (for which the theoretical maximum $F_{S_{yS}}$ is only G_{SC}/c Pa). For interplanetary transits wherein atmospheric drag is not a factor, it may therefore be optimal to optimise using a fitness function that takes the highest $F_{S_{yb}}$ in a range, rather than for a single α. This configuration also achieves an unusually high λ of 0.9745.

5.5 Generative Freeform Optical Elements for Stability

Semi-cylindrical lightfoils have demonstrated the ability to provide corrective pitch and yaw torques to a solar sail in previous literature, both via simulation and in-situ experiments [30, 37]. This behaviour is reproducible in the described optical simulation, as demonstrated by the single lightfoil torque profile in Fig. 5.19a. In this profile, a curve with a negative gradient can be seen passing through $\alpha = 0°, \tau_{S_{yb}} = 0$ Nm/m^2 such that a negative α corresponds with a positive $\tau_{S_{yb}}$ and vice versa. This indicates that the illuminated particle will rotate towards $\alpha_s = 0°$ in this region. This holds true until the curve next passes $\tau_{S_{yb}} = 0$ Nm/m^2, which for the single lightfoil occurs at $\alpha = \pm 45°$. This behaviour makes lightfoils a convenient starting point for optimisation. It is noteworthy that these torque simulations were particularly sensitive to ray density, necessitating a greater number of rays to be cast with a smaller l_R in order for the phenomena of interest to be reliably reproduced. As such, $\tau_{S_{yb}}$-α profiles are formed by casting 100 rays, up from 50.

Next, we define corrective torque. A sail torque is *corrective* if it causes the sail to angularly accelerate towards some stable incidence angle α_s. In the case of

lightfoils, several α_s may exist [30]. This report will focus on $\alpha_s = 0°$ (and the associated corrective torques) because it promotes the Sun-pointing attitude that so many transmissive sails are designed for. We define corrective torque $\tau^*_{S_{yb}}$ to be positive if it effects a rotation towards $\alpha_s = 0°$ according to Eq. 5.41.

$$\tau^*_{S_{yb}} = -\tau_{S_{yb}} \cdot \frac{\alpha}{|\alpha|} \tag{5.41}$$

In other words, $\tau^*_{S_{yb}}$ is positive if the sign of $\tau_{S_{yb}}$ is opposite to the sign of α — which is intuitive, given that the purpose of $\tau^*_{S_{yb}}$ is to 'undo' any non-zero α. By this definition, $\tau_{S_{yb}} = \tau^*_{S_{yb}}$ for negative α as demonstrated by Fig. 5.19a (where the two profiles are superimposed for $\alpha \leq 0°$). This parameter is principally defined because it makes for simpler stability-promoting fitness functions, but also has the benefit of visually delineating the stabilising behaviour of these optical elements via the symmetry of a $\tau^*_{S_{yb}}$-α profile. For convenience in later analyses, we also define the range of α over which the optical element is stable: the *stable range* $\mathbb{A}_s$ (in this case, $\mathbb{A}_0 = 45°$).

It should be noted that in space, a sail providing passive corrective torque will not induce its satellite to neatly rotate from some arbitrary α to a state of rest at the nearest α_s. In the absence of damping torques, such a satellite would instead oscillate about its α_s, to zero net effect on rotational energy. Furthermore, this is the ideal case for which the rotational energy of the satellite is low; if its rotational energy is very high, a corrective torque will only serve to periodically speed-up — and then slow-down — rotation as the satellite passes α_s. Conversely, atmospheric drag torques can provide some damping to an oscillating sail in certain LEOs, but this would only work if they were small relative to the corrective torques, which is highly dependent on the orbit (for example, a high altitude, dusk-dawn Sun-synchronous orbit would roughly place α_s at an aerodynamically stable attitude for an archetypal transmissive sail, and so atmospheric damping would be more effective here).

Despite its limitations, passive corrective torques remain useful for several reasons. For one, corrective torques provide *passive stability*. In the absence of passive stability, even small disturbance torques would cause a satellite that is not being actively controlled to rotate away from its desired attitude *in perpetuum*; if this disturbance torque is applied in an axis other than that of one of the principal axes of inertia of the satellite, the satellite may never return to its initial attitude, even fleetingly. Conversely, when a passively stable satellite is positioned at α_s, any disturbance torques will be counteracted by a corrective torque. The resulting small oscillation about α_s is typically preferable to the complete departure from the intended attitude that a disturbance torque would otherwise induce. Secondly, corrective torque *can* be harnessed to bring about rotation from an arbitrary α to α_s if it is either actively controlled or supplemented.

The most direct method for control is to cover the corrective torque pattern with an opacity or reflectivity control device, such as a PDLC film, allowing the satellite to activate, disable or modulate the corrective torques at will. This transforms the patterned sail from being a passive system to being an active one. By not impeding the pattern when the corrective torques are opposing rotation but occluding the pattern when it would supplement rotation, the rotational energy of the satellite can be reduced, and the satellite may be effectively brought to rest at α_s.

Alternatively, these corrective torques may be supplemented by a secondary, active system to provide damping. Many sailcraft operate auxiliary, active attitude control systems capable of orienting a satellite alone, which may lead one to conclude that these passively stable patterns are at risk of being redundant (in the classical, negative sense; many engineers would view this as a plus). However, a passive attitude control system can make the job of an active attitude control system much easier as the $\hat{\mathbf{y}}_\mathbf{B}$ and $\hat{\mathbf{z}}_\mathbf{B}$ pointing of the satellite can be effectively carried out by the patterned sail. In this case, the active control system only

needs to concern itself with $\hat{\mathbf{x}}_\mathbf{B}$ pointing, and with damping the other axes by providing a small torque that opposes the prevailing direction of rotation. Such a satellite would naturally come to rest at α_s if the active control torques were very small.

The aforementioned approaches can be combined in a lightweight manner if the auxiliary system is a magnetorquer. Other than redundancy, this approach comes with the benefit of having the weaknesses of each system mitigated by the strengths of the other; solar sail control systems are rendered periodically inoperable at low altitude due to the prevalence of eclipse, becoming more reliable as altitude increases. Conversely, magnetorquers eventually become inoperable as altitude is increased, but are highly reliable at low altitude where the Earth's magnetic field is strongest. Conversely, a reaction wheel or thruster has neither synergy with patterned sails, nor a weakness in any flight regime that would make such a synergy necessary. They also retain effectiveness in very high, interplanetary orbits, which neither solar sails nor magnetorquers can boast. However, they are considerably heavier than magnetorquers and come with other limitations, such as momentum saturation and contaminating exhaust fumes, respectively.

5.5.1 Scale Factors

As mentioned in section 5.3, scale is arbitrary for SRP calculation when dealing with purely refractive optical elements. However, this is not the case for calculating torque per unit area $\tau_{S_{yb}}$, also known as *torque surface density*. This is because the relative offset of each ray-particle interaction with respect to the centre of mass is pertinent to torque calculation. For these optimisation runs, we introduce the *scale factor* Σ. Naturally, this scale factor is applied in all three axes, but for convenience we calibrate $\Sigma = 1$ to be indicative of a pattern or element with height $z^{\text{MAX}} - z^{\text{MIN}} = 1$ m.

Because Σ is applied in all three axes multiplicatively, the distance from the centre of mass to any arbitrary point on the particle boundary is also changed multiplicatively by Σ. As such a distance is an argument of torque for an arbitrary ray-particle interaction (which always occurs at a particle boundary), the torque is also changed multiplicatively by Σ. In other words, a $\tau_{S_{yb}}$-α profile does not change shape when Σ is changed, and new simulations do not need to be run when it is changed. Instead, the entire $\tau_{S_{yb}}$ dataset can be multiplied by Σ to make it applicable to a pattern of a different size, provided that it has the same geometry when normalised.

Conventional wisdom would state that in subsequent figures, Σ should be set to one so that a reader may apply their own scale factor in a single operation. However, we instead set this to be $\Sigma = 10^{-5}$, consistent with a 10 μm tall element. This is done because a profile using $\Sigma = 1$ would be calculating $\tau_{S_{yb}}$ for an optical element that is one metre tall, which would provide wildly misleading values at a glance. Accordingly, to convert these figures with profile τ_1-α to a new profile τ_2-α according to a new scale factor, the operation becomes $\tau_2 = \tau_1 \times \Sigma/10^{-5}$.

5.5.2 Self-stabilising Freeform Optical Elements

Most prior literature has focused on single lightfoils in isolation, so the optimisation of a single self-stabilising optical element is a natural start point. The ray diagram for one such lightfoil at $\alpha = 20°$ is depicted by Fig. 5.18a. For this and subsequent stability optimisation runs, it is assumed that geometric symmetry is necessary and — to ensure compatibility with a wide range of pattern transfer methods — that the pattern should have a flat base edge (contrary to the prism convention, this base edge faces away from the Sun). As such, the optimiser is constrained to produce geometries with symmetry about the z (or x_b) axis, and vertices beneath the x (or y_b) axis are removed from convex hull calculation. The ray diagram for the optimised geometry is depicted by Fig. 5.18b.

Fig. 5.19 depicts the torque surface density profiles and normalised geometries of a semi-cylindrical lightfoil and, with the aforementioned constraints, a lightfoil-derived freeform optical element. The latter was optimised using the fitness function f_{f4} (Eq. 5.42) wherein $\alpha_2 = \text{-}\alpha_1 = 20°$, which seeks to maximise the integral of corrective torque over the optimisation range. This relatively narrow optimisation range was chosen to encourage the maximisation of corrective torque within the traditional $\mathbb{A}_0$ of a lightfoil. The ray diagrams of Fig. 5.18 depict these optical elements illuminated at $\alpha = 20$, where the corrective torques for both designs are large. While the number of rays cast for diagrams is reduced as for previous optimisation runs, the number of rays cast here is higher because the behaviour of these pattern is obscured when ray density is low.

$$f_{f4} = \int_{\alpha_1}^{\alpha_2} \tau^*_{S_{yb}} \, d\alpha \tag{5.42}$$

The maximum corrective torque per unit area achieved by the initial lightfoil was $\tau^*_{S_{yb}}(20°) = 6.18$ pNm/m^2 (piconewton-metres per square metre) with a stable range of $\mathbb{A}_0 = 45°$. It should be noted here that there is no theoretical maximum for $\tau^*_{S_{yb}}$ because there is no theoretical maximum displacement from the centre of mass. Furthermore, in the absence of absorption and diffuse reflection, a trapped ray ricocheting around the perimeter of a particle could generate theoretically infinite torque without breaking energy conservation laws (this could skew optimisation runs, but fortunately this never happened). It can be seen in 5.19a that the optimiser improved fitness f_{f4} by greatly increasing the mean and peak $\tau^*_{S_{yb}}$ at the cost of a small $\mathbb{A}_0$ deficit. The maximum corrective torque achieved was $\tau^*_{S_{yb}}(18°) = 9.30$ pNm/m^2 (an increase of 50.5%) with a stable range of $\mathbb{A}_0 = 42°$ (a deficit of 6.7%).

It can be seen in Fig. 5.19b that it achieved this by condensing the lightfoil vertices to form a 'smooth corner', in contrast to the continuous curvature of the lightfoil. The f_{f4} element also converges to a convex point at its extrem-

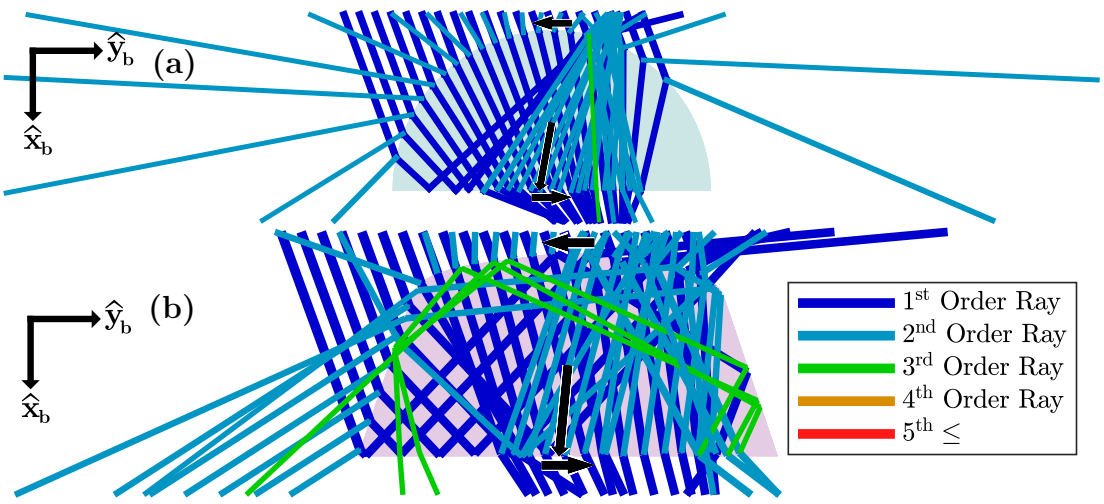

Figure 5.18: Ray Diagram for Single, Self-stabilising Optical Elements at $\alpha = 20°$ (a) Lightfoil (b) Lightfoil-derived, Optimised for f_{f4}; Filter Set to 1%, Accurate Ray Count

ity. Furthermore, relatively little resolution was lost here when compared with previous freeform elements that were optimised for SRP. The optimiser seems actively discouraged from creating redundant vertices for the same reason that torque simulations are highly sensitive to ray density: a lightfoil only functions because a large number of disparate interactions occur that — on average — correspond with a corrective torque $\tau^*_{S_{yb}}$. Reducing the ray density reduces the sample size that is fed into that average, and certain ray-particle interactions that would otherwise have their effects 'averaged out' instead have a disproportionate contribution to the calculation of $\tau^*_{S_{yb}}$. This is simply by merit of the ray-particle interactions that *would* have mitigated its effects not occurring, because the triangles that needed to be intersected for said sequence of interactions to occur was missed. Similarly, an optimiser action that attempts to disrupt this delicate balance with a large vertex change will meet considerable resistance because performance is not dependent on one single edge or vertex; it is dependent on the curvature of the shape as a whole. This is in contrast to pure SRP optimisation, for which curvature is useful but not a functional requirement.

The final geometry resembles a rounded pentagonal prism. To determine why this is optimal, we should first approximate what a desirable ray diagram would looks

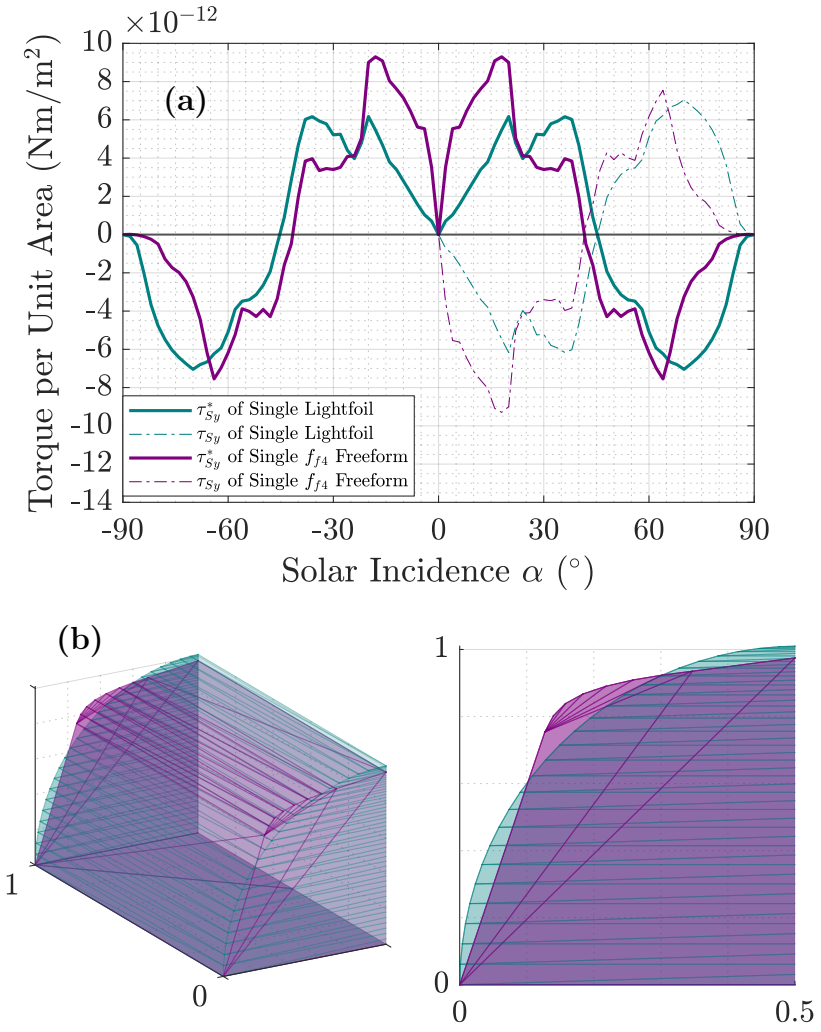

Figure 5.19: Single, Self-stabilising Optical Elements: (a) Torque Profile with $\Sigma = 10^{-5}$, (b) Normalised Geometry Bisected about Symmetry Line

like for a torque-generating optical element. A first-order ray with a high θ_k and large offset relative to the centre of mass for each of its ray-particle interactions is an archetypal, high contribution ray for generating torque. Conversely, a higher order ray with low θ_k that is redirected near or through the centre of mass would be a low or zero contribution ray. The ideal ray diagram would therefore be highly asymmetric, with first-order rays being redirected around the perimeter, and being ejected in a common direction *around* the centre of mass (i.e. all clockwise or all counterclockwise). Generally, rays that follow a clockwise path around the centre of mass will effect a counterclockwise $\tau_{S_{yb}}$, and vice versa.

The reason that the optimiser converged upon this geometry appears to be due

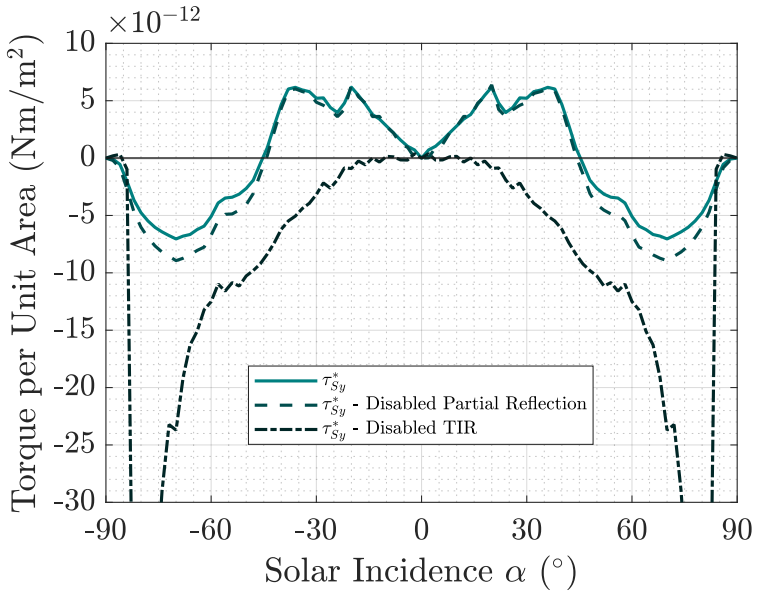

Figure 5.20: Torque Profile for a Lightfoil with $\Sigma = 10^{-5}$ Demonstrating the Effect of Disabling Partial or Total Internal Reflection

to its effects upon the rays that underwent TIR. The sides of the f_{f4}-optimised element are, on average, steeper than the lightfoil, and so rays around this α encounter it with higher incidence θ_i. On transmission, the refraction effect therefore induces greater $\Delta\theta_k$, and this in turn reduces ray θ_i when they encounter the base edge. The effect of this is that many more rays undergo TIR here than for the lightfoil. On its own, TIR here would actually effect a negative $\tau_{S_{yb}}$, but these rays are subsequently either ejected from the top right of the particle with a clockwise alignment relative to the centre of mass, or undergo a second TIR to be ejected from the base, again with clockwise alignment.

Maximising the number and contribution of these TIR rays is a logical means of optimisation because, for both the lightfoil and the f_{f4} element, these rays have by far the greatest contribution to torque. For example, rays born from partial reflection occasionally display a small positive contribution to corrective torque, but far more often it is either neutral or negative. Conversely, the contribution of transmitted rays that do not undergo TIR ranges from near-zero to extremely negative. This is shown in Fig. 5.20 for a lightfoil, wherein disabling partial reflection has practically no effect on torque, whereas disabling rays that undergo

TIR renders the lightfoil highly unstable.

5.5.3 Self-stabilising Freeform Patterns

The next optimisation run deals with the self-stabilisation of an entire pattern, with the fitness function unchanged. Ray diagrams for the initial and the optimised lightfoil patterns at $\alpha = 14°$ are depicted by Fig. 5.21a and Fig. 5.21b, respectively. It can be seen in Fig. 5.22a that lightfoil performance does indeed degrade when it is applied to an entire pattern, as for the SRP generating elements prior. In pattern form, the lightfoils yielded maximum corrective torque $\tau^*_{S_{yb}}(28°) = 5.40$ pNm/m^2 (down from 6.18 pNm/m^2 at 20°) with a stable range of $\mathbb{A}_0 = 41°$ (down from 45°). Furthermore, the regions of instability are aggravated somewhat, and the largest negative $\tau^*_{S_{yb}}$ can be seen to exceed the magnitude of the largest positive $\tau^*_{S_{yb}}$ by a large margin ($\tau^*_{S_{yb}}(-70°) = -17.6$ pNm/m^2). However, the loss of performance is notably less pronounced than it was for SRP-generating prisms at and around a Sun-pointing attitude. This is because the function of a lightfoil is not predicated on the ejection of rays with high $\hat{\mathbf{y}}_\mathbf{b}$ alignment, which makes pattern propagation more likely; generally, rays are ejected from a lightfoil via the base edge or near the top of the camber, where they are less likely to impinge upon adjacent lightfoils. Fig. 5.21a demonstrates the ray diagram for a lightfoil pattern at maximum $\tau^*_{S_{yb}}$, and it can be seen that no first-order rays undergo pattern propagation under this configuration (those that enter an adjacent lightfoil do so due to the ray casting scheme, not pattern propagation – see Fig. 3.12, section 3.2.6).

In Fig. 5.22b, it can be seen that the f_{f4} optimised pattern bears a strong resemblance to that of the pattern seen in section 5.5.2 just prior. The main distinctions are that this pattern version is proportionally taller, retains significant curvature on its cambered edge, and does not come to a convex point at its extremity. The method for optimisation also appears to be the same, namely, the capturing and

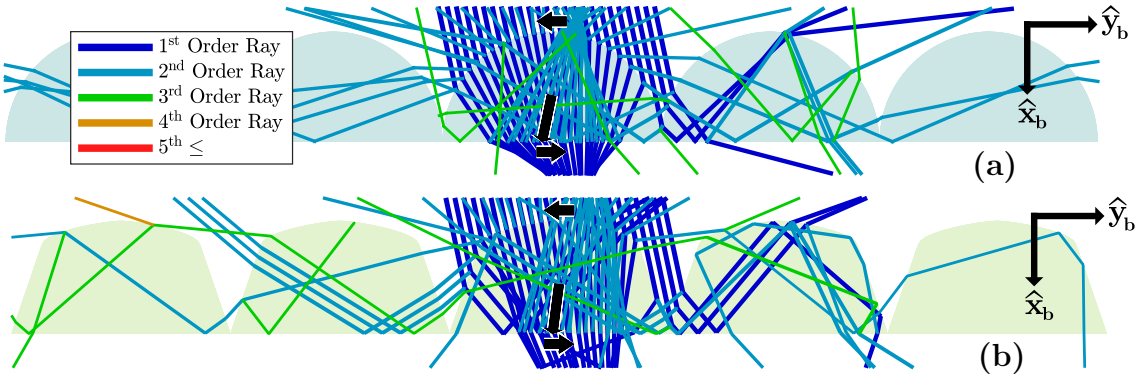

Figure 5.21: Ray Diagram for Self-stabilising Patterns at $\alpha = 14°$ (a) Lightfoil (b) Lightfoil-derived, Optimised for f_{f4}; Filter Set to 1%, Reduced Ray Count

effective redirection of more rays via TIR, to the effect of increased mean and peak $\tau^*_{S_{yb}}$ at the cost of $\mathbb{A}_0$: the final geometry yields peak $\tau^*_{S_{yb}}(12°) = 9.381$ pNm/m^2 with a stable range of $\mathbb{A}_0 = 32°$, corresponding with a 74.0% increase and 22.0% decrease, respectively.

5.5.4 Self-stabilising Freeform Patterns with Offset Centre of Mass

As convention dictates, prior torque profiles were generated for elements and patterns using their own centres of mass. This is the natural approach to take, and may be suitable for arbitrarily large, unmanned sailcraft for which the mass of the sail is dominant. However, this approach is less well suited to the optimisation of sails for modern satellites, for which the mass of the sail is generally low relative to the mass of a satellite bus. Unless such a satellite were to deploy its sail from a very long boom, the centre of mass of the satellite with or without the sail attached would be very similar. For this optimisation run, sail mass is assumed to be negligible. As an analogy for a 3U CubeSat bus, it is assumed that the centre of mass is located 15 cm beneath the sail in $+\hat{\mathbf{x}}_\mathbf{b}$. A ray diagram for this optimised pattern is depicted by Fig. 5.23.

On the scale of a sail pattern, the offset of a ray-particle interaction with respect

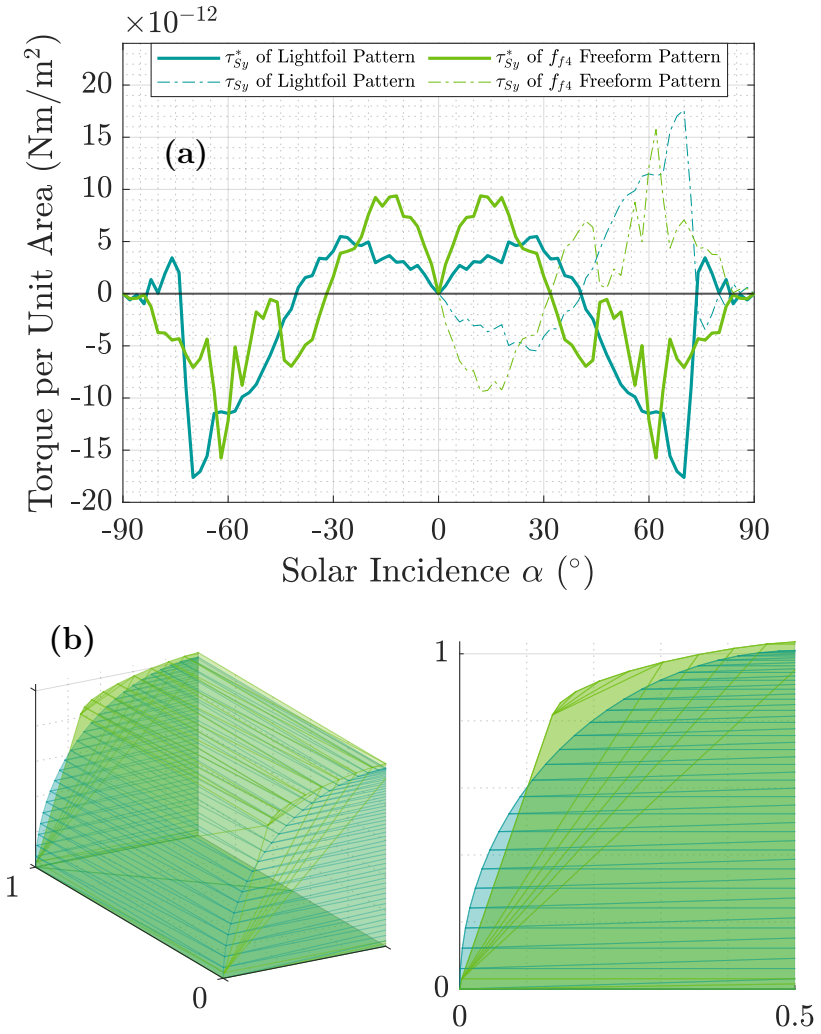

Figure 5.22: Self-stabilising Patterns: (a) Torque Profile with $\Sigma = 10^{-5}$, (b) Normalised Geometry Bisected about Symmetry Line

to the centre of mass is generally measured in microns (with $\Sigma = 10^{-5}$, this offset typically did not exceed 5 μm). An external centre of mass shifts the paradigm because the position of each ray-particle interaction, in effect, becomes redundant: in terms of offset from the centre of mass, a ray-particle interaction occurring at the top of a pattern will only differ from one occurring at the bottom by about 10 μm, which is a negligible difference when the offset is being measured in terms of cm with respect to a satellite bus. Instead, the direction of torque is determined by the sign of $F_{S_{yb}}$.

Nevertheless, it can be seen in Fig. 5.24a that the function of a lightfoil is not compromised by this change. Naturally, corrective torques are greater here due to

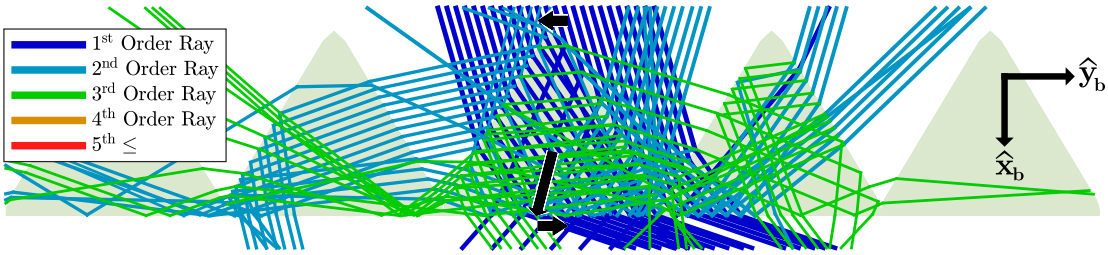

Figure 5.23: Ray Diagram for a Self-stabilising, Lightfoil-derived Pattern with a Centre of Mass Offset of 15 cm at $\alpha = 16°$, Optimised for f_{f4}; Filter Set to 1%, Reduced Ray Count

the larger centre of mass offset, with the lightfoil demonstrating a peak corrective torque of $\tau^*_{S_{yb}}(18°) = 0.053\ \mu\text{Nm/m}^2$. More importantly, the shape of the profile is not adversely affected. Instead, it is smoothed, with the only adverse affect being a reduction in stable range, which has decreased from $\mathbb{A}_0 = 41°$ to $\mathbb{A}_0 = 34°$. This smoothing is a result of ray paths that would have otherwise had opposing contributions to $\tau^*_{S_{yb}}$ being made constructive, as the dominant ray paths tend to have contributions to $F_{S_{yb}}$ of the same sign. For example, it was seen in Fig. 5.23 that almost all first-order rays were ejected with negative θ_k and positive $\hat{\mathbf{y}}_\mathbf{b}$ alignment, suggesting that they each effect a negative contribution to $F_{S_{yb}}$.

In contrast to previous optimisation runs, it can be seen in Fig. 5.24b that the f_{f4} optimised pattern deviates significantly from the lightfoil that it was derived from, with a form that resembles an equilateral prism rather than a cambered foil. This pattern operates on the principle that an equilateral prism will generate an $F_{S_{yb}}$ of opposite sign to that of its α, at least within the stable range $\mathbb{A}_0$. Again, fitness was improved by increasing the mean and peak $\tau^*_{S_{yb}}$ at the cost of $\mathbb{A}_0$: the final pattern yields a peak corrective torque of $\tau^*_{S_{yb}}(16°) = 0.131\ \mu\text{Nm/m}^2$ with a stable range of $\mathbb{A}_0 = 21°$, representing a 147% increase and 38.2% decrease, respectively. Notably, performance drops sharply at $\alpha = \pm 20°$, which is the edge of the optimisation range. It is interesting to note that this pattern promotes pattern skimming, as for certain SRP-optimising patterns. This pattern is also the *only* example of the optimiser not utilising TIR as the primary mechanism for achieving high fitness.

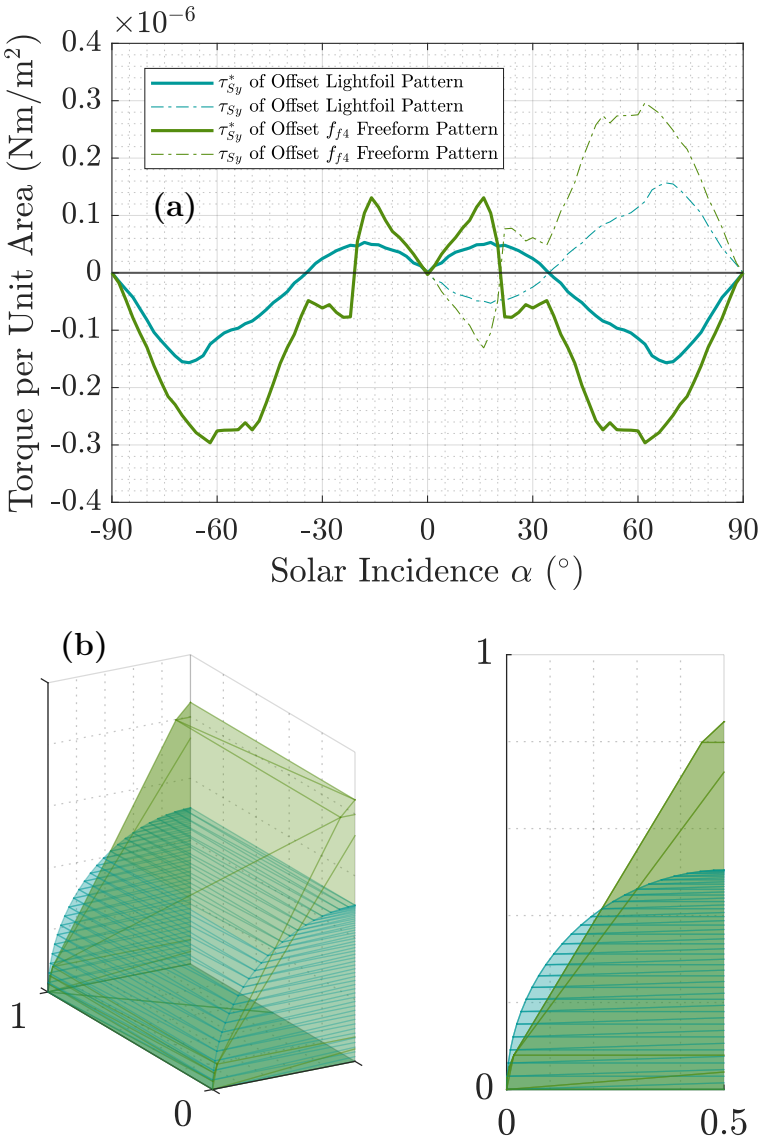

Figure 5.24: Self-stabilising, Offset Patterns: (a) Torque Profile with $\Sigma = 10^{-5}$, (b) Normalised Geometry Bisected about Symmetry Line

5.6 Optimisation Overview

Although the designs varied substantially between optimisation runs according to the input, the behaviours that they emphasised were similar. *Almost* every optimised design utilised TIR as the primary mechanism for generating $F_{S_{yb}}$ or $\tau^*_{S_{yb}}$. Designs that utilised pure refraction demonstrated some desirable traits — including a more uniform SRP profile, with fewer and smaller fluctuations — but were shown to be typically sub-optimal in terms of raw performance.

When optimising an SRP profile for acceleration, the optimiser naturally sought

to maximise ray alignment with $-\hat{\mathbf{y}}_\mathbf{b}$, i.e. to achieve $\theta_k = 90°$. High flux, first-order rays were invariably prioritised, but the best designs also considered the ray paths of lower flux, second-order rays borne of partial reflection (rays of a higher order than the second-order were rarely considered). For single optical elements, achieving nearly absolute $-\hat{\mathbf{y}}_\mathbf{b}$ alignment was straightforward and optimal. However, patterns attempting to do the same would induce pattern propagation. Patterns therefore posed a more difficult optimisation problem for the optimiser. To circumvent this issue, the optimiser converged upon pattern geometries that fell under one of two archetypes: pattern skimming — wherein first-order ray ray alignment was maximised to the greatest extent possible without inducing pattern propagation — or harnessing — wherein the optimised patterns utilised pattern propagation to improve ray alignment. Generally, pattern skimming was the favoured strategy, but many of these patterns also harnessed pattern propagation in a mitigating capacity. Such patterns could carry out pattern skimming more aggressively and effectively because the effects of being 'too aggressive' (i.e. of having a ray shoot too close to the pattern and reenter it erroneously) were mitigated by ensuring that any rays that propagated through the pattern did not counteract the prevailing direction of SRP.

For stability optimisation, the strategy utilised by the optimiser depended on whether the centre of mass was dictated by the sail itself, or by a greater satellite. For the former, performance was increased by promoting first-order ray TIR at the base of the pattern, ensuring ejection at the upper cambered surface with $\hat{\mathbf{y}}_\mathbf{b}$ alignment of opposite sign to that of α. For the latter, performance was increased by promoting pattern skimming behaviour, again with $\hat{\mathbf{y}}_\mathbf{b}$ alignment of opposite sign to α. In this case, the optimisation problem was similar to that of the pure SRP runs: the key difference was that these geometries and SRP profiles had to be symmetric. Higher order rays had very little contribution to fitness during stability optimisation runs and were largely ignored by the optimiser.

Broadly speaking, the optimisation runs that were carried out could be partitioned by: triangular micro-prism optimisation for SRP, freeform element optimisation for SRP, or freeform element optimisation for stability. In one case, numerically optimised triangular prisms were compared with an analytical solution. The numerically optimised patterns demonstrated zero α $F_{S_{yb}}$ and peak $F_{S_{yb}}$ of up to 3.07 μPa and 2.72 μPa, respectively, representing an increase over the analytical solution of up to 43% and 58%. In another case, a pattern of freeform optical elements was generated that registered zero α $F_{S_{yb}}$ of 3.40 μPa, which is 74% of the theoretical maximum. Likewise, during a stability optimisation run, a pattern of freeform optical elements derived from lightfoils (scale factor $\Sigma = 10^{-5}$) demonstrated peak corrective torque $\tau^*_{S_{yb}}$ of 9.38 pNm/m^2, representing an increase of 74% from that of the initial lightfoils.

Nevertheless, these patterns can only be called optimal within the purview of their fitness functions: increasing fitness according to these functions often came at the cost of a performance deficit that lay beyond the scope of said functions. For example, stability optimisation utilised a fitness function that optimised for $-20^\circ \leq \alpha \leq 20^\circ$ and resulted in higher performance within this range, but at the cost of the total range over which stability could be achieved. On the other hand, SRP optimisation typically greatly improved $F_{S_{yb}}$, but at the expense of the pressure ratio λ, which was not considered by most fitness functions. Such designs will be optimal for certain manoeuvres, but not for every manoeuvre, emphasising the importance of tailoring the fitness function to a specific mission, flight regime or manoeuvre when utilising numerical optimisation. Conversely, the fact that performance gains were demonstrated in such a specific manner according to the user-defined fitness functions should commend this method for such applications.

The model-free reinforcement learning pattern optimiser proved to be a reliable method for improving performance according to user-defined fitness. However,

the efficacy of the method was sensitive to optimiser configuration and initial conditions. Furthermore, while two differently-configured optimisation runs with identical optimisation criteria would sometimes converge upon the same solution, they would just as often diverge to output very different solutions. This suggests that repeated runs with a variety of different optimisation parameters would be required in order to carry out an exhaustive pattern optimisation. Furthermore, while very good solutions may be reliably found by this method, a globally optimal solution cannot be guaranteed.

Chapter 6

Manufacturing Processes

6.1 Mould Fabrication Experiments

The geometry of the negative space of a mould pattern dictates the surface geometry of a patterned sail. When forming a mould for a sail, one must contend with the mutually competing stressors of resolution and scalability: the mould must have very small features to comply with the specifications of a refractive (or to an even greater extent, diffractive) solar sail, and yet the mould itself must be large enough to pattern a sail in an economical and timely manner. First, one must consider the pattern transfer method, as this will dictate the form factor of the mould.

6.1.1 Linear versus Rolling Systems

The difficulty of balancing these stressors can range from near-trivial to highly demanding, depending on whether a *rolling* pattern transfer system is available or not. There is little nuance in the decision making process in this respect: if it feasible to use or procure, a rolling system is *always* preferable because both of

the processes that pertain to the mould — namely, mould fabrication and pattern transfer — are more costly and time-consuming in the absence of such a system. The reason that rolling systems are so desirable is that they partially automate the pattern transfer process and dramatically increase throughput. As such, a greater area of sail can be feasibly patterned with a smaller mould and fewer person-hours. The issue of scalability is therefore mitigated from the outset, and the requirements of the mould fabrication stage are dramatically reduced.

In theory, a roll-to-roll (R2R) or roll-to-plate (R2P) pattern transfer system comes with the drawback of greater cost according to the greater sophistication of the technology involved. In practice, this is only initially true when investing in one for use in-house. However, rolling systems typically incur reduced costs for high volume applications [119], which is pertinent when patterning at the scale of a solar sail. At such a scale, this translates to reduced processing costs when using R2R/R2P in-house, and a more general reduction in cost when using a rolling system as a commercial service.

In both cases, the source of the reduced cost is the aforementioned reduction in person-hours (material costs for rolling and linear systems do not substantially differ). To highlight why a linear system may be so much more laborious, it was calculated how many individual pattern transfers would be required to complete a 1.5 m diameter, circular transmissive sail using a linear pattern transfer system. In this case study, we attempt to address the issue of throughput by allowing for the creation of sequentially larger moulds before imprinting the sail. This is referred to as a *parent-child mould* methodology hereafter. Three square inch *parent* moulds are assumed, as this is a standard form factor for patterned silicon wafers, and one to which the moulds used in later trials comply (the fabrication of these moulds is explored in section 6.1.5; their usage for pattern transfer is explored in section 6.3). Furthermore, while this is quite a large form factor for moulds fabricated through micro-scale AM, it is still feasible through the method

explored at the end of section 6.1.3. Under this methodology, it is first assumed that parent moulds may be combined to form a larger parent before imprinting the *child* mould (either through a silicon wafer bonding process [120, 121] or a joining method for AM-fabricated parts [122]). This parent mould may then transfer its pattern to a child mould through an injection moulding process using silicone or similar (explored at the end of section 6.3). Finally, this child mould is used to imprint the sail. The output of this case study is Fig. 6.1, which depicts the relationship between the number of child imprints and the total number of imprints. For example, it can be seen in this figure that the optimal solution engendered 20 child mould imprints for a total of 50 pattern transfers when the parent was constrained to a single wafer. For lower throughput methods, this could imply over a hundred hours of constant imprinting (see section 6.3). Conversely, a small desktop R2P nanoimprinter operating at peak capacity could do the same within as little as four hours [123].

Furthermore, while Fig. 6.1 demonstrates that increasing the size of the parent mould may greatly curtail the issue of linear throughput, this approach has other issues: producing parent moulds in a scalable manner is non-trivial when executing in-house (explored in sections 6.1.2—6.1.5) and is one of the most costly stages of sail manufacture to execute (explored in section 6.4). This is true regardless of whether it is outsourced or executed in-house. Therefore, increasing the output of this stage should be a last resort. Conversely, imprinting into a child mould may be considerably more inexpensive (explored in section 6.3). The principal drawback of this approach is that a child mould implies an additional pattern transfer process through which to propagate any inaccuracies or defects that are incurred by said process.

In any case, both approaches result in very large moulds that require very large pattern transfer systems. For the 1.5 m diameter sail of this case study, this is not necessarily prohibitive: in the aforementioned single wafer case, the optimal

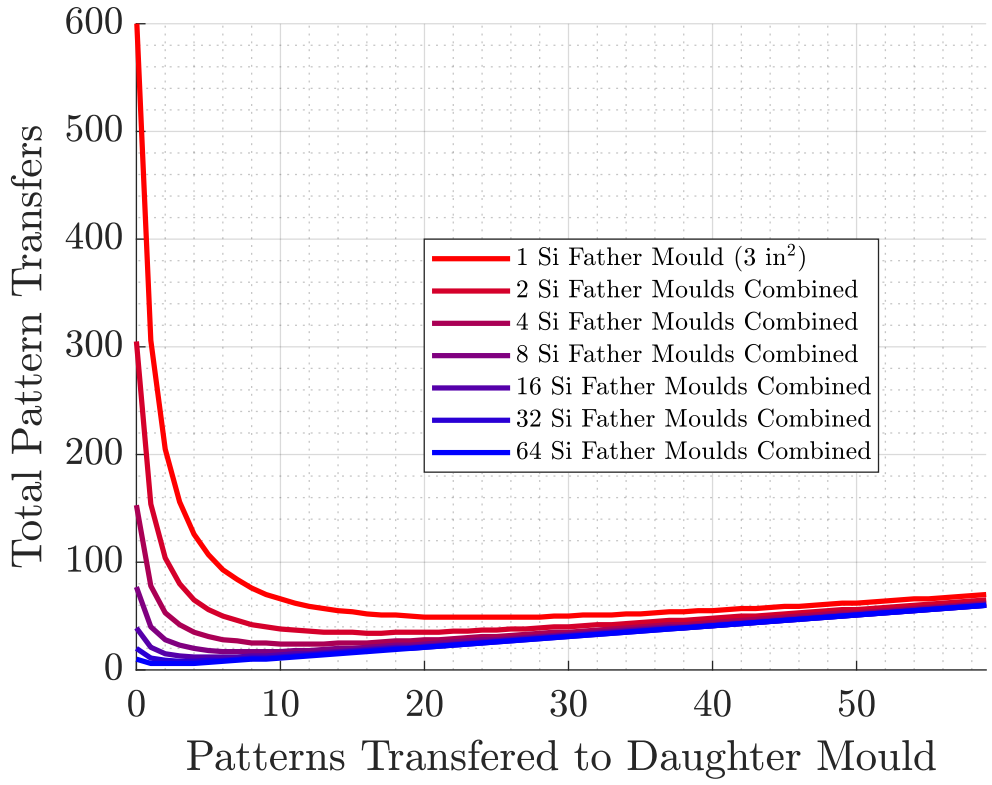

Figure 6.1: Linear System Pattern Transfers Required to Pattern a Circular Sail of 1.5 m Diameter using a Parent-Child Mould Methodology: standard 3 in^2 form factor for parent (father) moulds that may be combined before imprinting the child (daughter) mould, which is used to pattern the sail in turn

pattern transfer process would demand a square imprinter of 1.52 × 1.52 m — which is large but achievable. Furthermore, one may intentionally choose a less optimal parent-child route in order to relax this constraint according to the circumstances of the developer, should such a large printer prove to be infeasible to procure or develop. However, for very large sails, an optimal parent-child route may be necessary for the process to be feasible. Ultimately, increasing the size of this system will narrow the upfront cost advantage that a linear system holds, thereby making the rolling alternative even more desirable.

In short, a parent-child methodology is evidently the most efficient means of performing linear pattern transfer in the absence of a micro-scale *deus ex machina* to produce singular, very large moulds with very fine features (such systems are actually an ongoing area of research [124]). However, even when one is able to implement this methodology, linear pattern transfer is only competitive in the

leanest of circumstances for which available capital and sail requirements are both low.

6.1.2 Additive Methods

Several additive manufacturing (AM) methods are appropriate for mould manufacture. These come with the benefit of offering considerable control over sail geometry, and of being able to rapidly produce new prototypes as a design evolves. Of the seven standardised categories of AM, six are compatible with polymers making them suitable for producing reusable moulds: these are material extrusion, powder bed fusion, vat photopolymerisation, sheet lamination, material jetting and binder jetting [125]. Of course, not all methods are capable of resolutions that are up to the task. For our purposes, geometries as small as 20 μm may be required [29], requiring resolutions of at least 2 μm; for diffractive sails, this is the upper limit, with finer resolutions being desirable. An issue with these requirements is that higher resolution AM methods are typically not the most efficient for high volume production, and might be expected to struggle to meet the mould demands of lower throughput pattern transfer methods that require large moulds.

Of these, four are traditionally considered for small scale optical applications: these are material extrusion via fused deposition modelling (FDM) or direct ink writing (DIW), powder bed fusion via selective laser sintering or melting (SLS/SLM), material jetting via inkjet printing (IJP), and photopolymerisation via stereolithography (SLA) [126]. These are discussed below.

Extrusion methods involve feeding molten polymer through a nozzle, which then rapidly cools to form a discrete parcel or continuous line of material that binds to any material adjacent or below [127]. The most common of these methods is FDM, which has one of the highest throughputs but lowest resolutions. This

is typically in excess of 100 μm [128], owing to the fact that resolution is tied to nozzle diameter [129]. The mechanical properties of its materials are well-documented if typically poor, in part due to its tendency to produce porous structures [128]. On the one hand, these pores are typically internal and do not affect surface features, and furthermore, the mechanical properties are not prohibitively low for every pattern transfer process when appropriate materials are used [130]. Furthermore, FDM has successfully been used for the fabrication of optical components in the past, such as lenses [131] and infrared diffraction gratings [132]. However, these were on the millimetre scale. FDM remains inappropriate for micro-scale mould fabrication owing to its limited resolution, which is particularly low in the z axis [128]. Nonetheless, FDM is useful for prototyping structural parts owing to its high throughput, and was used as such during this project. Furthermore, the structural optimisations that AM allows can outweigh the drawbacks of using materials with lower mechanical properties, particularly for FDM-compatible materials that mitigate these drawbacks and demonstrate radiation tolerance [133]. For this reason, FDM-fabricated structures have been proposed for satellites and are an ongoing field of study [1, 134]. As such, this method certainly has application for the fabrication of elements of a sail other than that of the membrane if suitable materials are chosen. This comes at the cost of incurring greater risk by flying materials with low or zero flight heritage (though being the first to demonstrate these materials in space may be of enough inherent value to the scientific community to justify the additional risk). Another extrusion method of note is DIW, which is sometimes called *robocasting* [135, 136]. DIW is notable for being a lower throughput extrusion method, and for having an extremely wide range of cited resolutions — even within recent literature [136, 137]. In principle, its resolution is limited by its nozzle diameter as for any extrusion method. However, much higher resolutions than FDM are possible because DIW employs inks, pastes and slurries that may be of much lower viscosity than a molten filament. As such, DIW may employ smaller nozzles than

FDM. In its more refined forms, this method has reportedly achieved sub-micron resolutions [137]. Although DIW was not trialled during this project, it would be technically feasible to perform transmissive sail mould manufacture using one of the higher-end expressions of this method.

While extrusion deposits material onto a build plate in an already molten state, powder bed fusion instead deposits and sinters or melts the material sequentially: thin layers of powder are deposited on the build plate and energy is directed onto it as either an electron-beam or (in the case of SLS/SLM) a laser, thereby selectively melting said powder layer to form a pseudo-2D shape [126, 138]. The build plate is then lowered and a new layer of powder is deposited above, and the process repeats layer-by-layer [127]. In the case of SLS, the presence of this powder bed comes with the benefit of supporting overhangs without the need for temporary support structures [139, 140], which is rare amongst AM methods. Conversely, support structures are required for SLM because of the elevated temperatures associated with rendering its powder molten [141]; these support structures are needed in order to dissipate excess heat and to resist the warping that might be induced by thermal stresses [142]. On the other hand, this powder bed comes with the drawback that any powder melted will naturally adhere to the surrounding powder on the same layer, rather than exclusively adhering to the previously-melted material of the layer beneath it. This contributes to an overall poor surface finish that requires significant post-processing to make the printed part suitable for optics [126]. After post-processing, these parts may exhibit excellent surface properties that are of comparable quality to conventionally manufactured optical elements [143]. However, while capable of higher resolution than extrusion, powder bed fusion methods such as SLM are nonetheless ill-suited to micro-scale optics because their resolution is typically in excess of 50 μm in xy [144]. In this case, the key limiting factor for resolution is the size of the powder particles [140]. Furthermore, their resolution in z is disproportionately poor at around 200 μm, which is only slightly better than FDM [128, 144]. This

is because a powder layer is typically several particles thick. This is partially by design and partially incidental: multiple particles in a layer can be desirable to mitigate against single-layer inhomogeneity, which arises due to suboptimal powder flowability, non-uniform particle sizes [145], and gaps due to the geometry of powder particles being suboptimal for packing [139]. Conversely, defects or inaccuracies in the powder deposition process will cause layers of more than one particle deep to arise naturally [145]. Although unsuitable for pattern moulds, SLS/SLM nonetheless have the same structural applications for a sail as FDM; they even have access to many of the same radiation tolerant materials (as powder rather than filament), such as *polyether ether ketone* (PEEK) [133, 146]. Other than having less ubiquity than FDM printers, their only major drawback is the additional post-processing that is required, which makes them less convenient for rapid prototyping.

Material jetting has also been used for optics, particularly IJP [126], which is sometimes called direct inkjet printing (DIP) [147]. This method bears similarity with extrusion in that it revolves around the deposition of material via a nozzle, which then solidifies. The primary difference between the two is that extrusion methods deposit a continuous molten filament onto a build plate, while material jetting deposits discrete droplets onto a build plate or substrate [148]. Furthermore, the deposited droplets are composed of either an aqueous or photopolymeric ink which solidifies via partial evaporation or UV-curing, respectively [149]. IJP is a low throughput method that is mostly used for patterning. Historically, it was also considered a low resolution patterning method [137, 150]. In recent years, sub-micron resolutions have been achieved of around 0.3 μm [151, 152], which would be suitable for transmissive sail moulds. However, parts produced through IJP can suffer from poor mechanical properties arising due to imperfect adhesion between its layers, and can be fragile [148, 153]. This may limit its usefulness for the fabrication of reusable moulds. Nevertheless, low pressure pattern transfer methods do exist [154] for which IJP-fabricated moulds may be suitable.

Finally, vat photopolymerisation methods utilise the selective curing of photopolymeric resins. The most common form of vat photopolymerisation is stereolithography (SLA), wherein UV light is focused over a small area to cure a small volume of monomer resin [126] — which is solid in its cured, polymer form. This polymer will bind to any previously-cured polymer adjacent or below, and this curing process is repeated layer-by-layer. Photopolymerisation methods are generally low throughput, but they also comprise some of the highest resolution methods [148]. These characteristics are influenced by a number of parameters pertaining to the properties of the UV beam. In particular, higher power densities correspond with greater throughput by enabling faster scanning speeds, but higher beam diameter corresponds with greater throughput to the detriment of resolution [138]. Furthermore, while traditional SLA is typically too coarse for micron-scale fabrication, so-called micro-stereolithography (MSL, micro-SLA or μSLA) is a more recent expression of this method that is capable of sub-micron spatial resolutions. Functionally, this method differs from traditional SLA only by way of having a substantially lower beam diameter [127]. Furthermore, a desire to improve throughput led to the development of digital light processing (DLP) *a.k.a.* projection microstereolithography (PμSL), for which the scanning laser was exchanged for a series of small mirrors and a digital mask. Through this method, an entire layer of resin may be cured at a time, and the resolution is limited instead by the resolution of mask pixels and by diffraction effects incurred at the scale of said pixels [127]. Despite its higher throughput, DLP is capable of resolutions as high as 0.6 μm [137]. Considering also the decent mechanical properties of its finished parts [128], DLP is well-suited to transmissive sail mould fabrication. Accordingly, this was chosen to be the first method to be trialled for carrying out mould fabrication. However, photopolymerisation methods do come with some inherent drawbacks. For one, they are restricted to using photopolymers exclusively. Secondly, parts produced via photopolymerisation are particularly susceptible to having their mechanical properties degrade over time

[155], though their initial properties are typically good [128].

Furthermore, a DLP system that was capable of resolutions of around 2 μm was available [156], which matched the requirements of these moulds very precisely. However, this also meant that these moulds existed at the edge-of-the-envelope for the system, with little in the way of margin for error. For this reason, an alternative form of photopolymerisation was also explored that was capable of finer sub-micron resolutions, known as two-photon polymerisation (2PP) or direct laser writing (DLW) [126, 157]. This method induces resin to cure through near-infrared (NIR) light excitation instead of the more traditional UV, wherein NIR light is deployed as two photons via a pulsed laser beam [158]. Furthermore, using NIR allows 2PP to break the layer-by-layer AM paradigm and selectively cure any region of resin in 3D space [159]. This is possible because 2PP photopolymeric resins are generally transparent to NIR, but will nonetheless absorb NIR if two or more such photons excite it simultaneously. In practice, this means that curing only occurs at the focal point of said photons. Through this mechanism, 2PP can achieve resolutions of less than 0.1 μm [160], with resolutions as fine as 0.015 μm (15 nm) having been reported [158]. This is in excess of our requirements, implying a greater margin for error and higher repeatability. Initially, this method was deemed unsuitable due to its very low throughput [158]. This low throughput is symptomatic of its laser scanning architecture [157] — which is similar to that of traditional SLA — combined with an extremely low beam diameter. However, recent advances in 2PP — such as the use of femtosecond fiber lasers — have led to much faster scanning behaviours and a corresponding surge in throughput [161], resulting in build rates that are suitable for transmissive sail mould fabrication. As such, 2PP was chosen to be the second AM method investigated for the fabrication of sail moulds.

Irrespective of the method that is used, there are certain AM phenomena that can influence the performance of fabricated optical parts. The most prominent

of these is the difference between xy and z resolutions which — though certainly related — are in some ways distinct. This distinction arises for a few reasons. The first is gravity, which may influence the mechanism used for controlling the motion of the build plate, nozzle or optical devices in z, but not in xy: traditionally, xy control is achieved by a belt driven rotary actuator — which is fast but low torque — while z control is achieved by a ball screw driven rotary actuator — which is slow but high torque [162–164], and therefore better suited to opposing heavy loads. In recent years, faster ball screw actuators and high torque linear actuators have enabled AM systems that employ one mechanism for all three axes [164], but in any case, the control of z is constrained to high torque actuators. Furthermore, even if a single kind of actuator is used, a load bearing z actuator must be calibrated differently to those in xy. Furthermore, a z actuator may lose accuracy as successively more weight is added to a build plate or powder bed during fabrication. For this reason, methods that have a printing platform that is stationary in z may incur a smaller xy-z resolution discrepancy, particularly for high mass builds [163].

Furthermore, methods that melt or deposit a liquid must contend with the competing effects of gravity, surface tension, cohesion to adjacent droplets, and adhesion to adjacent material; the former two counteract one another to form a uniform droplet with xy and z variations that approximately emulate a spherical cap (assuming a level build plate or bed) [165]; the effect of the latter two is dependent upon the surroundings of the droplet, and so cannot be characterised with much specificity, other than to say that their effect is not necessarily uniform. Furthermore, for certain, very high viscosity droplets, solidification could occur before dynamic equilibrium is reached, which would confer even greater non-uniformity. Similarly, methods that excite powders, inks or resins must contend with the issue of having somewhat limited control over the penetration of the energising beam. In both the droplet case and the beam case (and the combination case, as for the material jetting of photopolymers), the net effect of

these phenomena places an innate limitation on z resolution that is described by the *minimum layer thickness*. Conversely, minimum layer thickness is sometimes imposed by an operator rather than by nature. This is done because layers of higher thickness are sometimes necessary from a repeatability standpoint, as layer thickness has been shown to correlate with dimensional accuracy [166] and resistance to warping [167]. This may indirectly reduce the practicable z resolution, especially for less robust AM materials. Furthermore, dimensions in xy are entirely constrained by the geometry of the final part, as the edges of each xy profile determines the topology of a part. In contrast, the z dimension of an xy profile is only strictly relevant to topology when the lower or upper surfaces are being defined (e.g. in the case of a simple cuboid, z would only be relevant to the first and final layers of the part). xy dimensions are therefore not as readily sacrificed in the name of repeatability, which may be interpreted as a procedural bias against z resolution. The overall effect of these natural and procedural phenomena is that AM resolutions are *usually* lower in z than in xy [128].

Perhaps the most important cause of this discrepancy is the layer-by-layer architecture itself, which incurs certain limitations on the z resolution exclusively (this does not necessarily apply to 2PP, and other 'true 3D' outliers): AM methods that fabricate layer-by-layer often proceed under the assumption that each layer is planar with xy, presenting a flat and uniform surface on which to build the next layer. This means that the thickness of a layer cannot vary, and each layer is effectively discrete. Combined with the aforementioned *generally* lower z resolution, discrete layer thicknesses contribute to the *staircase effect* [168], wherein the inclined surfaces of a part are actually represented by stacked, horizontal shelves or steps. For non-optical parts, the staircase effect does not necessarily hinder part functionality. For optical parts — and refractive or reflective optical elements in particular — the slope of a surface is essential to optical performance, and the staircase effect may render them nonfunctional. For example, a staircase-marred lens would be incapable of focusing light in any meaningful way even if

its macroscopic topology appeared to be accurate.

Fortunately, sail patterns typically have a continuous xy_B cross-section. In other words, their geometry can be treated as a stack of identical 2D shapes. The staircase effect can therefore be neutralised by using an optimal build orientation [169], which in this case is achieved by simply aligning the xy plane of the AM apparatus with the xy_B profile of the mould that is to be printed (i.e. with the mould on its side so that the 2D cross-section is in-plane with the build platform and every layer is identical).

6.1.3 Additive Trials

DLP trials were carried out with a HTL resin [170] via a Boston Micro Fabrication BMF microArch S130, which is capable of xy resolutions of up to 2 μm and z resolutions of 5–10 μm depending on how it is configured [156]. The initial DLP trial focused on discerning the minimum micro-prism and lightfoil dimensions that could be fabricated through this method, using the prism dimensions of source [29] as a benchmark (optical elements with a width of 27 μm). An array of prisms and lightfoils of increasing size were fabricated for this purpose, wherein the central element was sized according to the aforementioned benchmark. As seen in Fig. 6.2, the first attempt was unsuccessful as a result of the staircase effect (see section 6.1.2). This occurred because of sub-optimal print orientation, and subsequent trials were more successful once this was corrected. It can be seen in Fig. 6.3 that the median geometry was replicated to a high degree of accuracy at the final print layer. Generally, prisms of 20 μm width or more were well defined. However, issues with *overcure* could be seen to effect prototype moulds as a whole, as indicated by the dark silhouette surrounding the part. This indicated that layers fabricated near the start of the print continued to absorb UV while later layers were being fabricated. This issue is exacerbated when features are parallel to the print platform [171]. In other words, an 'optimal' print orientation

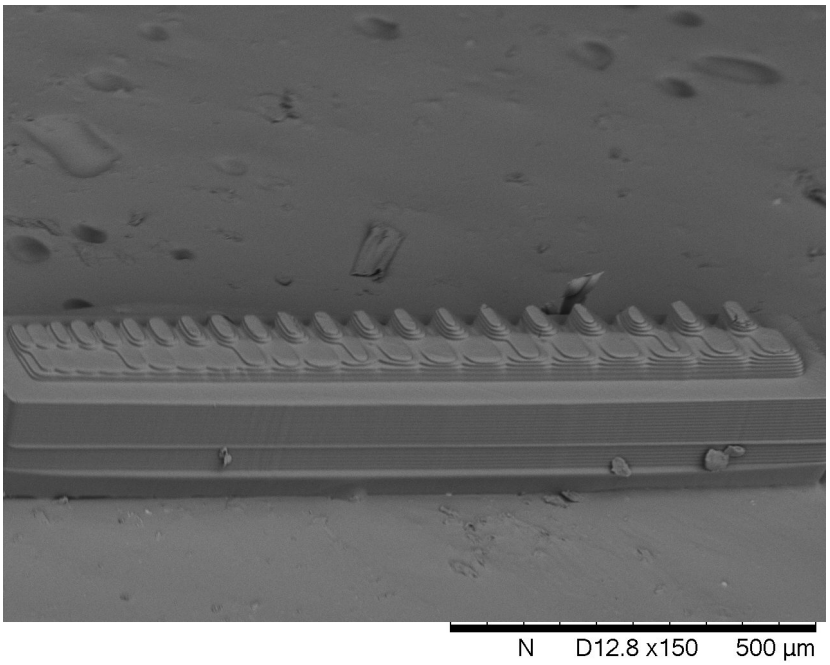

Figure 6.2: Scanning Electron Microscope (SEM) Image of Prototype Prism-Lightfoil Array with Geometries of Varying Size, Fabricated via DLP, Prominent Staircase Effect due to Sub-optimal Build Orientation

exchanges the issue of the staircase effect for the issue of overcure. This issue was mitigated for future prints by reducing the exposure time for earlier layers in anticipation of sustained UV absorption. This was a trial-and-error process because reducing the exposure too much would lead to undercure, corresponding with a reduction in resolution and adhesion to the print platform.

According to some ongoing issues with machine calibration, 27 μm patterns suffered from repeatability issues. However, increasing the size of the prisms beyond

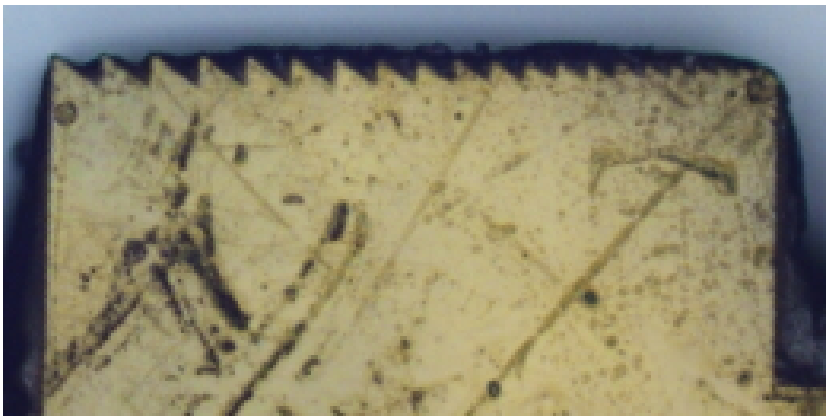

Figure 6.3: Microscope Image of Array of Micro-prism with Varying Size, Fabricated via DLP, Visible Overcure

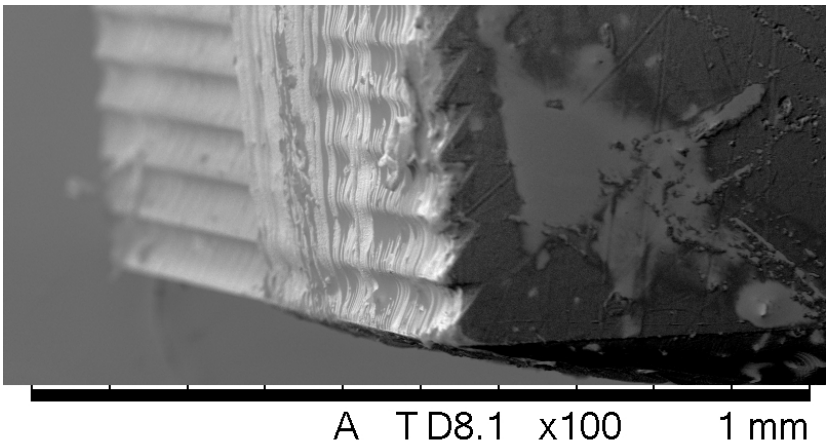

Figure 6.4: SEM Image of Well-formed, Scaled-up Micro-prism Array Fabricated via DLP (with some smoothing visible due to the gold sputtering process that preceded SEM)

that of the benchmark resulted in high repeatability prints; Fig. 6.4 depicts the output of one such trial for 50 μm width prisms. On the one hand, it was demonstrated in section 2.2.5 that larger prisms (i.e. increased sail thickness) does not necessarily correspond with a substantial performance deficit; as such, these larger, higher repeatability prisms pose a valid solution to the problem. On the other hand, since the conclusion of these trials, this same DLP system has demonstrated the highly repeatable fabrication of parts of diameters of around 27 μm for other ongoing projects; the repeatability issues at the original scale are therefore not representative of the method or system as a whole.

Conversely, an issue that could not be easily circumvented was that of the low throughput of this method. This was surprising given that the chief advantage of DLP when compared to μSLA is throughput. This made machine calibration via trial-and-error extremely time consuming. For example, a $2 \times 2 \times 2$ mm cube with a single patterned surface would complete in 2–4 hr depending on the settings, corresponding with a build rate of 1–2 mm^3/hr. While certainly not prohibitive for a R2R/R2P methodology, this build rate may be untenable for the fabricating of very large, linear pattern transfer moulds.

Subsequent AM trials focused on 2PP via the UpNano NanoOne [161]. This

machine is revolutionary in terms of 2PP throughput, which is achieved in part through a remarkably powerful femtosecond laser, and in part through its modular objectives. These objectives could be swapped out to achieve different magnifications of the pulsed laser, allowing throughput to be increased at the expense of resolution. As such, xy and z resolutions are described by the considerably wide ranges 0.22–1.20 μm and 0.55–23 μm, respectively [161].

Two trials were carried out for the fabrication of arrays of micro-prisms sized according to source [29], which composed the upper surface of a $10 \times 10 \times 0.5$ mm cuboidal mould. In both cases, this was executed without issue. Both trials utilised a two-phase strategy wherein the first 0.45 mm of z depth used a coarser objective or writing mode, and the final 0.05 mm of z depth (the patterned topography) is completed using a finer objective or writing mode. This approach is enabled by aligning pattern x_b with printer z so that the bulk mould structure is printed before the finer patterned region, rather than simultaneously. This is a build orientation that was previously described as 'sub-optimal' due to the influence of the staircase effect. However, this orientation is feasible here because 2PP is a 'true 3D' AM method that does not necessarily need to fabricate its parts layer-by-layer. This makes 2PP inherently resilient to the staircase effect.

In trial one, both phases used a 10x objective, but phase one used a coarse writing mode (resolutions of 4.2 μm in xy, 5.0 μm in z) and phase two used a fine writing mode (resolutions of 0.5 μm in xy, 3.0 μm in z). The output of this trial is depicted by Fig. 6.5. During this trial, phases one and two took 1 hour 28 min and 1 hour 22 min, respectively, for a total print time of 2 hour 50 min. This corresponds with an average build rate of 17.6 mm^3/hr.

In trial two, both phases used a coarse writing mode. Phase one used a 10x objective and phase two used a 20x objective (resolutions of 2.1 μm in xy, 1.5 μm in z). The output of this trial is depicted by Fig. 6.5. During this trial, phases one and two took 1 hour 28 min and 2 hour 27 min, respectively, for a total

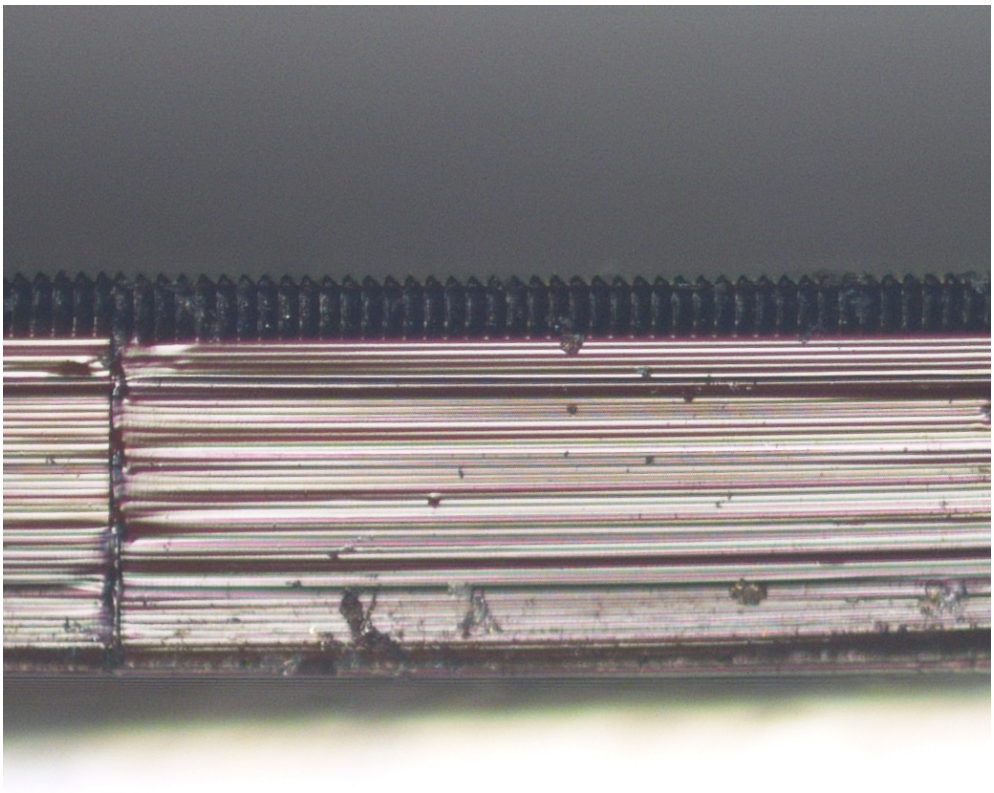

Figure 6.5: Microscope Image of 2PP Pattern Showing Two Distinct Layers, Topography Achieved via 10x Objective, Fine

print time of 3 hour 55 min. This corresponds with an average build rate of 12.8 mm^3/hr.

Fig. 6.5 demonstrates the distinct appearance of the regions fabricated through either phase (in this case for trial one); Fig. 6.6 demonstrates the well-defined geometry of the patterned region (in this case for trial two). Both trials were successful on the first attempt. The two-phase approach greatly sped up print times and effected build rates that were an order of magnitude higher than that of the DLP trials. Of course, DLP parts had to fabricated with an optimal xy-x_by_b build orientation, and so it could not execute a two-phase print as 2PP could. This demonstrates that higher resolution, true 3D methods may enable print strategies that lead to higher average build rates than their lower resolution and theoretically higher throughput contemporaries, in some cases.

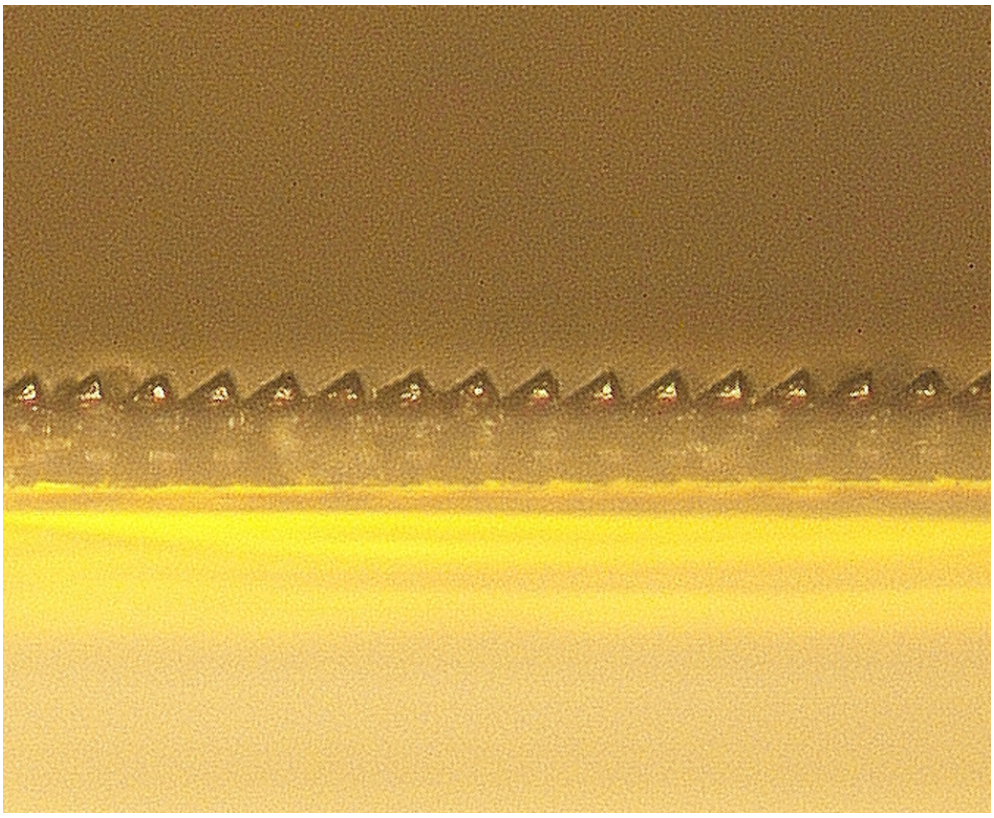

Figure 6.6: Microscope Image of Successful 2PP Layer, Topography Achieved via 20x Objective, Coarse

6.1.4 Conventional Methods

Naturally, micro-scale mould manufacture predates AM. Many conventional manufacturing methods have a long history of creating patterned moulds, and can do so with precision and repeatability that few AM methods can match. These methods also often come with the advantage of greater throughput for large-scale manufacture. Furthermore, the materials involved in conventional manufacturing processes typically have more extensive literature; AM resins and inks in particular are often bespoke or proprietary, which introduces a degree of uncertainty when using them for pattern transfer at elevated pressure or temperature. Conversely, conventional manufacturing methods are typically costlier and less convenient to fabricate with on a smaller scale — as for unique or prototype parts — and generally offer less flexibility than AM methods in terms of the lead time between design and fabrication, and in terms of the geometries that can be

achieved. Furthermore, many conventional methods have a very narrow range of materials that they are compatible with; typically, only semiconducting materials are compatible with the final stage of conventional micro-mould manufacture, and so silicon (Si) is nearly ubiquitous as the material of the finished mould [172].

The most common methods involve two or more phases, wherein a *template fabrication* phase is used to create a patterned resist or mask, and is followed by an *etching* phase that is used to transfer the pattern of that resist or mask to a substrate [173]; once patterned, this substrate may be used as a mould. Many methods require a mask for the executing of the template fabrication phase itself, and so template fabrication may be two-phase.

Common methods for producing the resist include photolithography and electron-beam lithography (EBL) [174]; in both cases, a photosensitive and/or electron-sensitive polymer is exposed to a beam that induces a chemical change. Unlike AM vat photopolymerisation, this chemical change is a change in solubility rather than curing, and the sensitive material is not necessarily a resin (although for photolithography it often is). Other than beam composition, the primary differences between photolithography and EBL is the nature of the mask that is used to make the exposure selective: for photolithography, this is a physical *photomask* that is placed above the resist, and that has opaque regions that selectively occlude a (typically UV) beam of light. As such, template fabrication via photolithography is two-phase, and the entire mould fabrication process including etching is three-phase. For EBL, this mask is a digital file that is used to control the path of the beam, and so EBL is more properly referred to as a *maskless* lithography method [172]. As such, template fabrication via EBL is single phase, and the entire mould fabrication process is two-phase. In terms of output, EBL is typically much higher resolution, but much lower throughput [175].

Because of this higher throughput, and because its lower resolutions are still suitable for micro-mould fabrication, photolithography would be the optimal method

for fabricating the resist. However, for reasons pertaining to the available masks (discussed in section 6.1.5), EBL was used instead. For the etching phase, dry (plasma) etching is most common [173, 176], and this was adopted for the purposes of the manufacturing trial.

6.1.5 Conventional Trials

A lithographic process was used to form a patterned mask, and this mask underwent dry etching to transfer its pattern to a 3 in^2 Si mould. Initially, issues arose regarding how to achieve the slope of a micro-prism; photolithography and EBL are not typically suited to making gradual z variations owing to their binary nature, as regions of a resist are either fully exposed or fully occluded by a conventional photomask or digital mask. To circumvent this issue, it was decided to use a *greyscale* mask and lithographic process. Such processes utilise masks that can control not only the area that is exposed, but also the power that is supplied to each area. In the case of *greyscale photolithography*, this is achieved by controlling the power density at exposed areas through partially transparent, 'grey' regions of the photomask: the transparency of these regions may change gradually in order to effect gradual variations in exposure, which effect gradual z (x_b) variations for the final, patterned resist in turn. In the case of *greyscale EBL*, power is supplied differently to each area of the resist by controlling the exposure time that is dictated by the digital mask. Again, exposure time may change gradually in one or more directions to effect a slope for the final resist.

Both photolithography and EBL systems were available for use in these trials. *Greyscale photolithography* was the method initially chosen for these trials according to its higher throughput, but a greyscale photomask was not available. *Greyscale EBL* was trialled instead because EBL masks can be created on-the-fly, being entirely digital. In this case, the greyscale EBL masks were limited to maximum thicknesses of 1.4 μm, requiring the pattern to be scaled down such

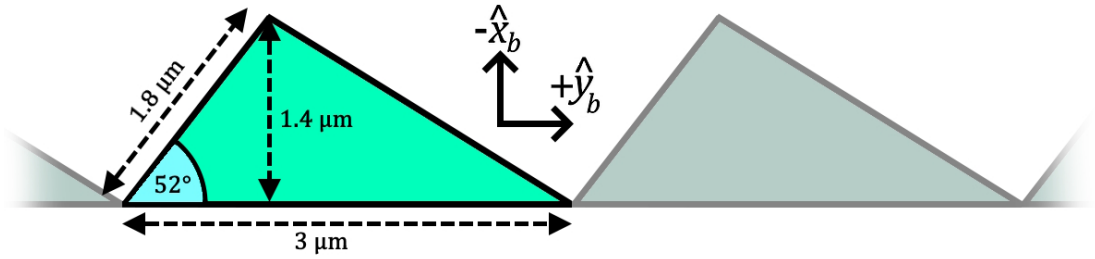

Figure 6.7: Dimensions of Mould Pattern for Conventional Manufacturing Trials: the negative of this pattern comprises the SU-8 resist geometry produced using SU-8 in a greyscale EBL process

that the width of a single prism was only about 3.0 μm. The EBL system used was the NanoBeam nB5 which, surplus to requirements, is capable of resolutions of 10 nm or less [177]. This was used to produce patterned resists composed of SU-8 for subsequent etching into Si.

For these trials, the pattern is described by Fig. 6.7. As can be seem in Fig. 6.8a-b, this geometry was reproduced successfully. A discrepancy does exist between the mask geometry and the final pattern, in that the latter is smoothed; the left lateral edge of the prism in particular (convention replicated by Fig. 6.8a) has a notable curve. However, this could feasibly be mitigated by adjusting the greyscale mask. Overall, this method demonstrated higher accuracy than any AM technique by producing geometries at about $1/10^{\text{th}}$ of the original scale which, if scales were normalised, would be of comparable quality. Overall, this serves to highlight the exceptional resolution that can be achieved through this method, and that would be compatible with even very fine diffractive sails (but are far in excess of the requirements of most transmissive sails). In practice, a thicker resist would be necessary to realise a purely refractive sail pattern, as at this scale, the prism pattern would serve as a blazed grating and would induce diffraction for sunlight within its dominant visible spectrum.

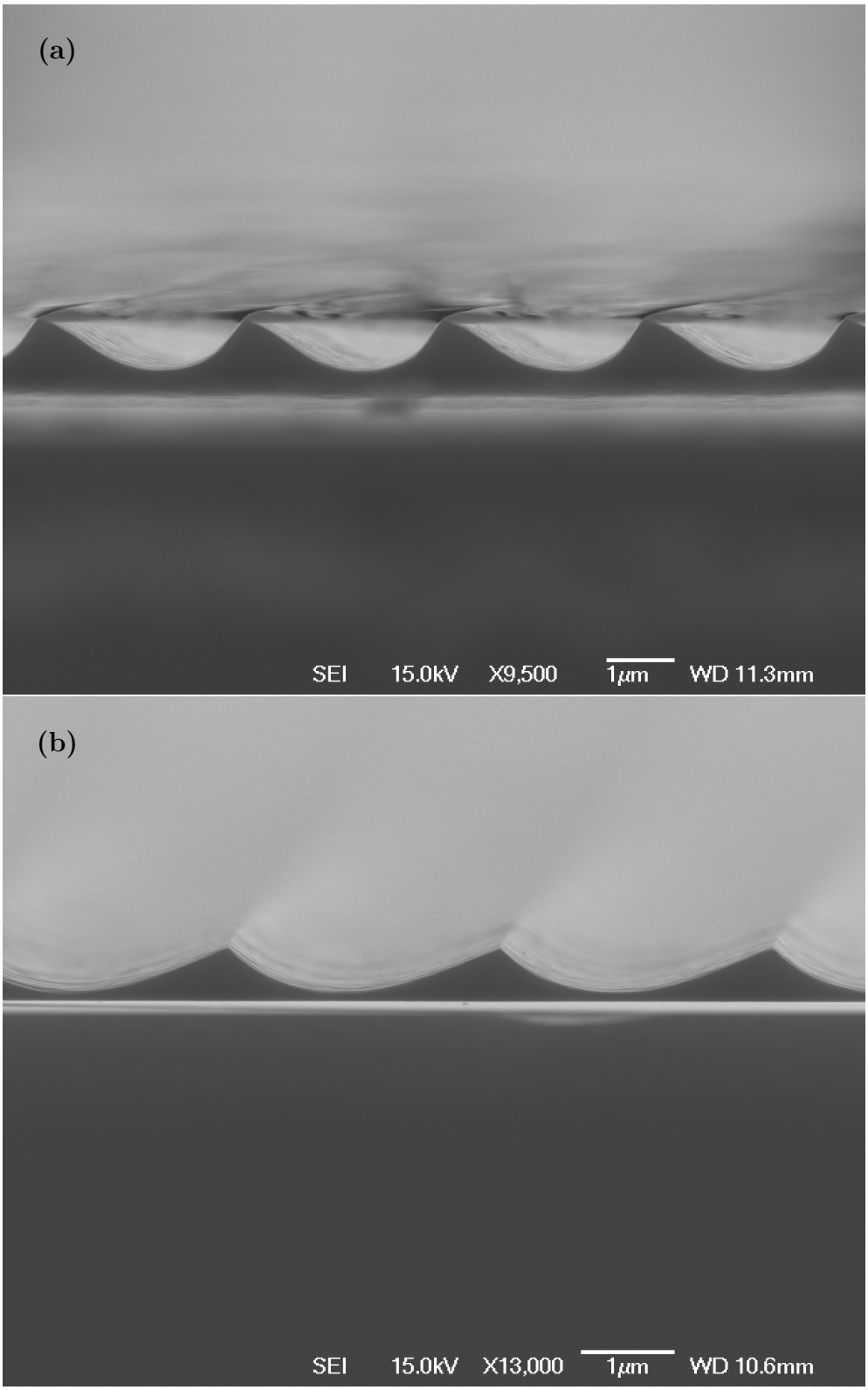

Figure 6.8: 3.0 μm-wide Micro-Prisms Fabricated via Greyscale (Electron Beam) Lithography

6.2 Thin Film Trials and Membrane Material Selection

Thin film fabrication trials were carried out to explore the suitability of fabricating these films in-house versus simply procuring appropriate thin films, with the dual-purpose of creating inexpensive films on which to carry out pattern transfer trials. This process is inextricably linked to both material selection and the pattern transfer method; different materials are compatible with different thin film fabrication methods, and the films that they form will only be compatible with certain pattern transfer methods.

6.2.1 Thin Film Trials using Polystyrene-toluene Solution

Thin film trials focused on forming thin films composed of polystyrene (PS), according to the recommendations of source [29], as this was assumed to be the material that the sail prototype would use — though this assumption would be challenged later. Despite being a common method for fabricating thin films, spin-coating was not explored due to its limited scalability [178]. Two alternative, more scalable methods for thin film fabrication were trialled: doctor blading, and a water-floating method. In both cases, thin films were achieved by dissolving PS within the solvent toluene, and coating a smooth substrate with a thin layer of the resulting PS-toluene solution. Being highly volatile, toluene evaporates rapidly, and the high area-to-volume ratio of the thinly-coated solution expedites this. Trials were carried out beneath a fume hood to mitigate against the hazards of toluene exposure.

Doctor blading is a conventional means of forming large area thin films [179] wherein a wide blade is secured above a smooth plate, with a small and fixed displacement between the edge of the blade and the plate surface. Alongside

the composition of the deposited solution, it is this distance that determines the thickness of the final film; films fabricated through this method may range in thickness from 20 μm to several hundred μm. As such, this technique is applicable to both fine diffractive and larger refractive sail membranes. The solution is deposited upon the plate and in front of the blade. Next, either the plate or the blade is moved so that there is relative motion between the two. The solution must pass through the blade-plate gap, which constrains it to a thin film form factor and facilitates a homogeneous spreading of the solution over the plate.

This process was carried out successfully during the trial with thickness uniformity at the micrometric scale. However, the scale of the films themselves was very small; the size of film that may be fabricated is dependent on the size of the doctor blade apparatus, which in this case was only a few centimetres. Although the apparatus is conceptually simple and could be made in-house without much difficulty, it would be more difficult — albeit not impossible — to execute on the scale of a solar sail membrane.

An alternative and potentially more scalable method was trialled wherein the PS-toluene solution was poured onto water [180]. Because toluene is less dense than water, this solution was buoyant so long as the percentage PS was kept suitably low. The thickness of the film could be controlled by calculating the PS content of the solution by volume, per millilitre of PS-toluene, and cross-referencing this with the dimensions of the tank. This method has the advantage of being extremely low-tech, requiring only a suitably large tank and a fume hood that can accommodate it. It is also quick to setup and requires very little supervision. The drawback is that the evaporation process is much slower through this method owing to the diluting effect of the water. A jig was made to separate different compositions of PS-toluene for these trials as seen in Fig. 6.9.

Films fabricated this way achieved thickness uniformity at the scale of tens of microns, with the film appearing slightly thicker near the centre. In any case,

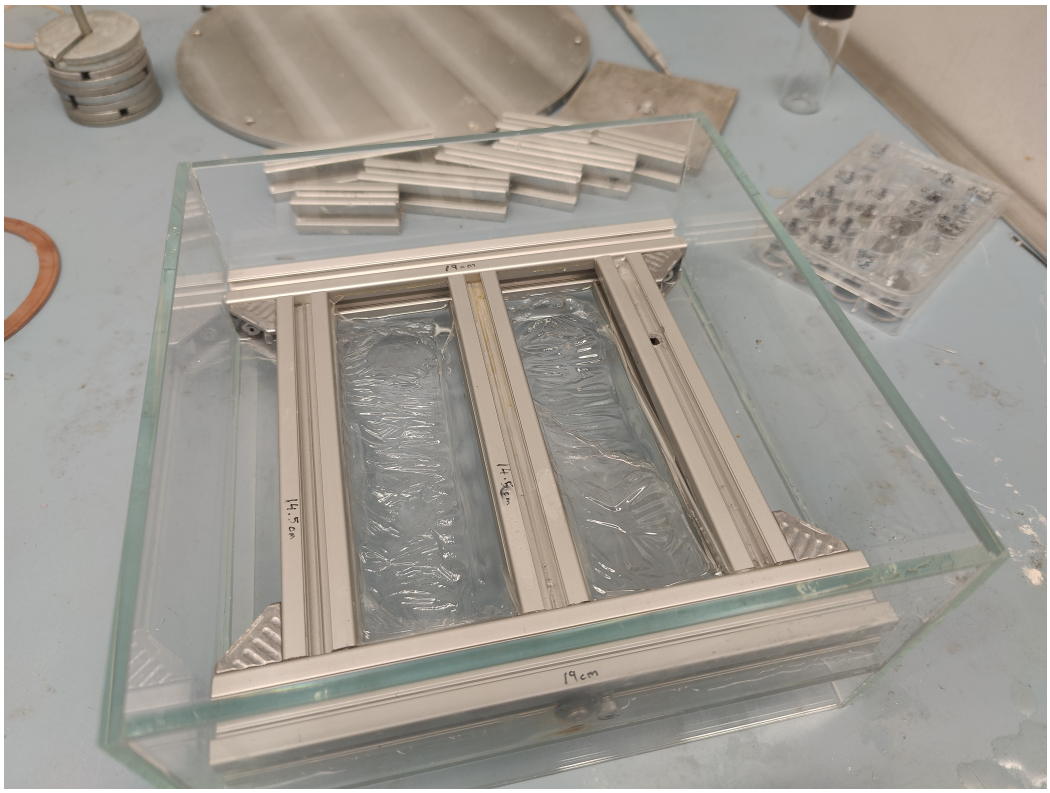

Figure 6.9: Jig Created for Floating and Evaporating Different Concentrations of PS-Toluene Solution above Distilled Water

handling the PS thin films made their unsuitability for spaceflight very apparent. Even films that had formed with too much thickness failed very easily due to how brittle they were, and it seemed unlikely that these would survive deployment under any significant tension.

6.2.2 Membrane Material Evaluation

Ten polymers were evaluated against PS, principally based upon the optical and mechanical properties shown in Table 6.1, where they are each listed. However, tolerance to radiation was also considered where literature was available, and some materials were rejected outright owing to their sensitivity to it. Qualitatively speaking, the criteria for selection were:

1. High optical performance; be comparable to PS, which has been shown to be optimal for these applications [29]:

   i. High transmissivity Φ_t within the visible spectrum

   ii. High dispersion curve; a high visible refractive index n and low Abbe number V_{Abbe}

2. Sufficient mechanical stability to survive launch, deployment, and operating under tension; be comparable to traditional solar sail materials such as Mylar$^{\text{TM}}$ and CP1$^{\text{TM}}$:

   i. High resistance to fracture during tension; high σ_T (UTS), high elongation at break (L).

   ii. Preferably: high elastic modulus E and high flexural strength σ_F; acceptable but less preferable: flexible/low elastic modulus E for low flexural strength σ_F materials.

   Several other properties were considered, but these only influenced the evaluation if they were unusually high or low for this class of material. These include:

3. Density ρ — low density is desirable, but is not a priority for reasons detailed in section 2.2.5.

4. Be simple to work with — this has different meanings for thin film fabrication and pattern transfer. For example, for pattern transfer using thermal nanoimprint lithography (T-NIL), this may mean having a reasonably low glass transition temperature T_g and/or heat deflection temperature HDT, while not being so low as to induce deflection in space.

5. Be tolerant to radiation (particularly UV) and outgassing when in a vacuum.

6. Be inexpensive.

Firstly, polysulfone (PSU) and polyphenylsulfone (PPSU) were rejected because sulfones have been found to absorb an inordinate amount of UV [89], which

Name	Acronym	n	V_{Abbe}	Φ_t	E (GPa)	σ_T (MPa)	σ_F (MPa)	L (%)
Polystyrene	PS	1.603	45	0.92	3.5	55	85	2
Polycarbonate	PC	1.586	29	0.89	2.2	68	110	100
Polyimide	PI	1.67	20	0.9	2.8	120	170	30
Polymethyl methacrylate	PMMA	1.484	53	0.92	3.2	70	90	3
Polyethylene terephthalate	PET	1.58	64	0.89	3.5	50	80	50
Polyvinyl chloride	PVC	1.544	56	0.9	2.8	45	65	40
Polysulfone	PSU	1.65	21	0.88	2.8	80	120	30
Polyphenylsulfone	PPSU	1.585	22	0.9	2.9	95	140	40
Polypropylene	PP	1.49	32.8	0.91	1.5	32	45	500
Polyethylene	PE	1.479	42	0.91	0.9	20	40	800
Cyclic olefin copolymer	COC	1.53	58	0.91	3.1	66	70	3

Table 6.1: Evaluated Materials for Transmissive Sail Membranes [66, 106, 107]

would greatly harm their longevity in space. The remaining rejected materials were rejected due to either poor optical or mechanical properties.

Materials rejected due to poor optical properties include polymethyl methacrylate (PMMA), polypropylene (PP) and polyethylene (PE). Of the remaining materials, cyclic olefin copolymer (COC) was rejected due to having poor mechanical properties that were too similar to fragile PS. Conversely, polyvinyl chloride (PVC) was rejected because its optical and mechanical properties were both only borderline acceptable, which made for an unattractive if technically viable candidate material.

In order of ascending suitability, the following three materials were favoured by the material evaluation: polycarbonate (PC), polyethylene terephalate (PET) and polyimide (PI). Firstly, PC has only middling optical properties, corresponding with a reduced sail performance that would only be slightly offset by the very low density of PC. They are nonetheless acceptable and PC escaped rejection according to its exceptional mechanical properties. For this reason, PC would make for a good backup material, and would be appropriate for sails for which higher stresses than usual would be expected (according to the demands of the orbit and the deployment system). However, it is notably more expensive than PS (around three times the cost per unit weight). In the case of the final, most

suitable material: PI is the most attractive option because its mechanical properties are comparable to PC while its optical properties are comparable to PS. Furthermore, PI has heritage as a solar sail material (metallised CP1^{TM}). Despite being the most attractive candidate, PI is problematic due to its relatively high cost (around five times the cost per unit weight, relative to PS) and the low availability of its more attractive, space-proven variants (such as CP1^{TM}). Furthermore, certain processes would have been more complex with PI because it is a thermoset (sometimes a *pseudothermoplastic*) as detailed in section 6.2.3.

The second most suitable material was ultimately favoured for future prototypes: PET demonstrated very good optical properties (in its biaxially-oriented BoPET form) and acceptable mechanical properties. Specifically, its elongation before failure is significantly higher than PS. The fact that PET is otherwise mechanically similar to PS could be perceived as a problem, but sail tension would be relieved by its generous ability to elongate, which would naturally mitigate against tearing. Furthermore, it is similar to PS and PMMA in that it is very stiff, which would mitigate against the wrinkling to which transmissive sails would be highly sensitive [40].

6.2.3 Process Compatibility of Selected Materials

In terms of performing thin film fabrication in-house, PC is similar to PS in that it is very easy to work with. Like PS, it is available in a wide array of bead and resin forms, and is compatible with the low-density solvent toluene, as well as the more ubiquitous (but slightly heavier) solvent acetone. As both are lower density than water, they would form solutions with PC that are compatible with the water float method, provided that the percentage PC is kept suitably low.

Conversely, PI and PET are both significantly more difficult to form thin films with than PS. In the case of PI, this is due to its status as a thermoset or pseu-

dothermoplastic, necessitating that it is cured during the thin film fabrication process. In the case of the latter, thin film fabrication is hazardous due to PET requiring very toxic solvents (e.g. *o-chlorophenol* or *o-cresol*), and that are incompatible with the float method according to their density — which is higher than that of water.

In terms of performing pattern transfer, the compatibility of a polymer is determined by the type of polymer that it is. Generally speaking, thermoplastics are universally compatible with T-NIL and incompatible with UV NIL. Conversely, thermosets (including most elastomers) are compatible with one or the other method, depending on the mechanism that is used to induce their precursor to cure. Thermally curable thermosets will be compatible with T-NIL, whereas photo-curable thermosets (photopolymers) are generally compatible with UV NIL. Thermosets that are cured chemically — such as via a condensation reaction — are not necessarily compatible with either method. Regardless of the mechanism, the principal difference between performing NIL on a thermoplastic and a thermoset is that imprinting of the former is carried out upon the final polymer, while the imprinting of the latter is carried out during or before the production of the final polymer; thermoset NIL is executed using an uncured resin or partially cured polymer precursor. All three of the chosen materials are compatible with some form of nanoimprinting [78, 80, 181]. Being thermoplastics, both PC and PET are compatible with T-NIL [78, 181]. The minimum temperature at which this is carried out is dependent upon T_g. For PC and PET, these minimums are $T_g = 147°C$ and $T_g = 71°C$, respectively. In practice, T-NIL is generally carried out far above T_g so as to minimise viscosity and promote the filling of the negative NIL mould space. For example, T-NIL has been carried out at 220°C for PC at 5 MPa [78] and 95°C for PET at 2 MPa [181]. However, higher pressures and longer imprinting times may allow for lower imprinting temperatures that are closer to T_g.

PI is comparatively complex to perform NIL with, being a thermoset that sometimes behaves as a thermoplastic (depending on composition). Furthermore, PI is usually cured chemically rather than through heat or radiation alone, to which one may conclude that PI NIL would be infeasible without a bespoke PI nanoimprinter. However, compositions of PI that behave as a pseudothermoplastic are compatible with T-NIL, albeit with difficulty according to their very high minimum imprinting temperatures of $T_g = 260–370°C$. This correlates with elevated thermal stresses that may affect the quality of the pattern. Conversely, thermally curable compositions of PI exist that may allow for T-NIL at lower temperatures, following a thermoset T-NIL methodology [80]. This would involve performing T-NIL with PI resin or a partially cured PI precursor and fully curing it in post. Such an approach induces lower thermal stresses, but PI imprinted this way is reportedly prone to having pattern defects arise due to outgassing [80]. Finally, photosensitive compositions of PI resin also exist, and that may allow for lower temperature cure [182]. More speculatively, some of these compositions may be compatible with certain UV NIL processes. It is possible that UV-curable photopolymers would be less suitable as a membrane material in general, as these may continue to cure in space. As for the conventional effects of polymer exposure to space radiation, overcure is associated with optical degradation.

6.3 Pattern Transfer Trials

6.3.1 Linear Pattern Transfer

Initial trials focused on exploring the feasibility of linear T-NIL performed on 50 μm thick films of PET using readily available lab infrastructure, such as a vacuum oven (for heat and mitigation of outgassing) and CNC mill (for manufacturing custom apparatus). This was carried out using the micro-prism patterned Si moulds fabricated via greyscale EBL and dry etching, described in section

6.1.5.

First, the Si moulds were treated to apply an *anti-stick* layer. This was done to mitigate against adhesion between the imprinted polymer and the imprinting mould, as this would otherwise introduce significant stresses during demoulding that may deform the imprinted pattern [181]. Commercial anti-stick aerosols would have been infeasible, as aerosol particles would disrupt the pattern because they are of a similar size to the micro-prisms. Chemical vapour deposition (CVD) would have been a viable alternative but the necessary apparatus was out-of-action. Instead of applying a layer additively, it was decided to reduce the surface free-energy of the Si surface itself. This was achieved via Si silanisation [183], wherein Si is exposed to a *silanising agent* and forms a monolayer with silanol groups and hydrophobic properties [184, 185]. In this case, the silanising agent used was *trichloro(1H,1H,2H,2H-perfluorooctyl)silane* (PFOCTS). First, the Si mould was cleaned using acetone and pressurised nitrogen. From beneath a fume hood, two drops of PFOCTS were deposited into an aluminium foil cap of lower diameter than the Si mould. The Si mould was placed on top of this foil cap (patterned face down) and the assembly was placed inside a vacuum desiccator for 30 minutes using the setup shown by Fig. 6.10a. Finally, the residual PFOCTS was baked off of the Si mould at 150°C over 10 minutes using a hot plate beneath a fume hood in accordance with the sequence detailed by source [183], resulting in the treated mould shown by Fig. 6.10b.

Next, a mild steel T-NIL press was fabricated via CNC milling that allowed for 28×30 cm imprints (shown in Fig. 6.11); the size of the press was only constrained by the size of the cavity of the available vacuum ovens (Fistreem$^{\text{TM}}$ and Thermo Fisher$^{\text{TM}}$). In practice, the imprints carried out during these trials were much smaller, as they were constrained by the size of the Si moulds that had been fabricated during the small-scale mould fabrication trials that are detailed in section 6.1.5. Principally, the static elements of this press consisted of a large

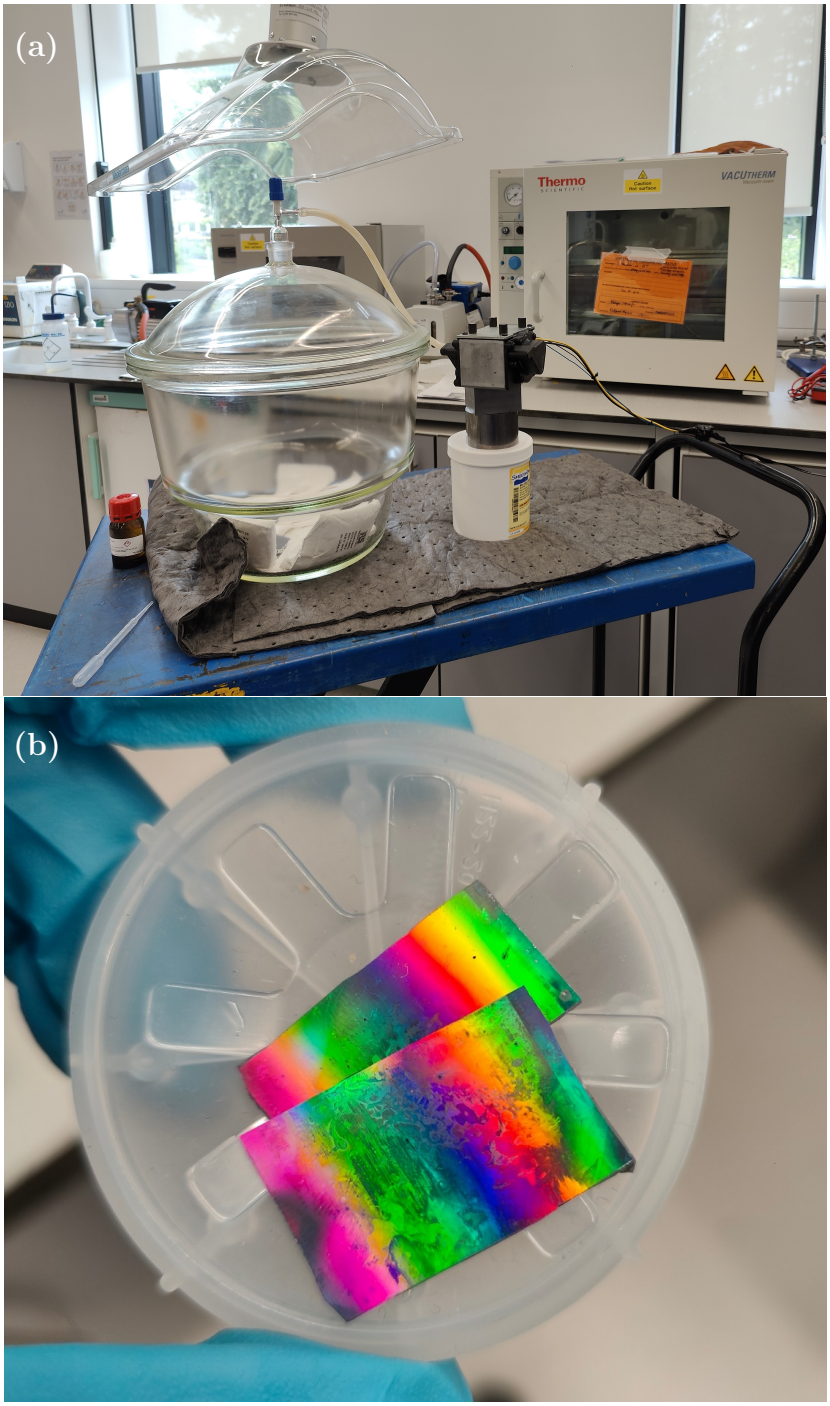

Figure 6.10: Mould Preparation via Silanisation, (a) Setup, (b) a Silanised Si Mould above an Untreated Si Mould

bracket fastened to a plate. Two separate sliding plates would sandwich the Si mould and the membrane that was to be patterned, and this assembly was slid under the bracket. The functional element of this press was a smaller, circular press plate that was secured to the end of a large screw. This screw was threaded through the centre of the bracket, and was suspended above the centre of the

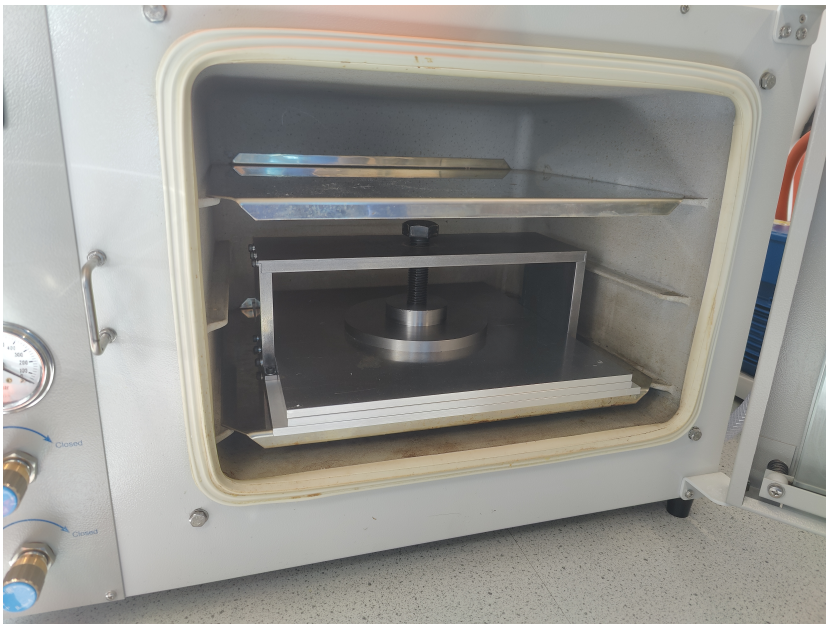

Figure 6.11: Custom Press for Thermal Nanoimprint Lithography within a Vacuum Oven

imprinting area. Pressure could be applied by extending this screw to the uppermost plate by operating the screw nut by hand, before applying a specific torque via a torque wrench. Alongside the high thickness of the two sliding mild steel plates, flexing was mitigated by the raised edges of the static plate and the indented plate-press interface of the uppermost plate, which constrained their motion. This made flexing unlikely for these T-NIL trials, which were carried out with relatively low pressures of 2–15 MPa).

Because this was an attempt to achieve T-NIL in the 'lowest tech' means (using infrastructure that might be available to small teams), it was expected that this approach would be sub-optimal. Indeed, this largely became an exercise of documenting all of the issues with the procedure. The first issue was the throughput, which was even lower than predicted. This was so because the heat capacity of the mild steel press was very high, while supplying heat to a system in a vacuum is very slow (in this case, it was supplied via heated shelves on which the T-NIL press rested). Furthermore, the process involved some guesswork as the thermometer of the system pertained to the temperature of the shelf, and the shelf reaching the desired temperature did not necessarily mean that the mould-

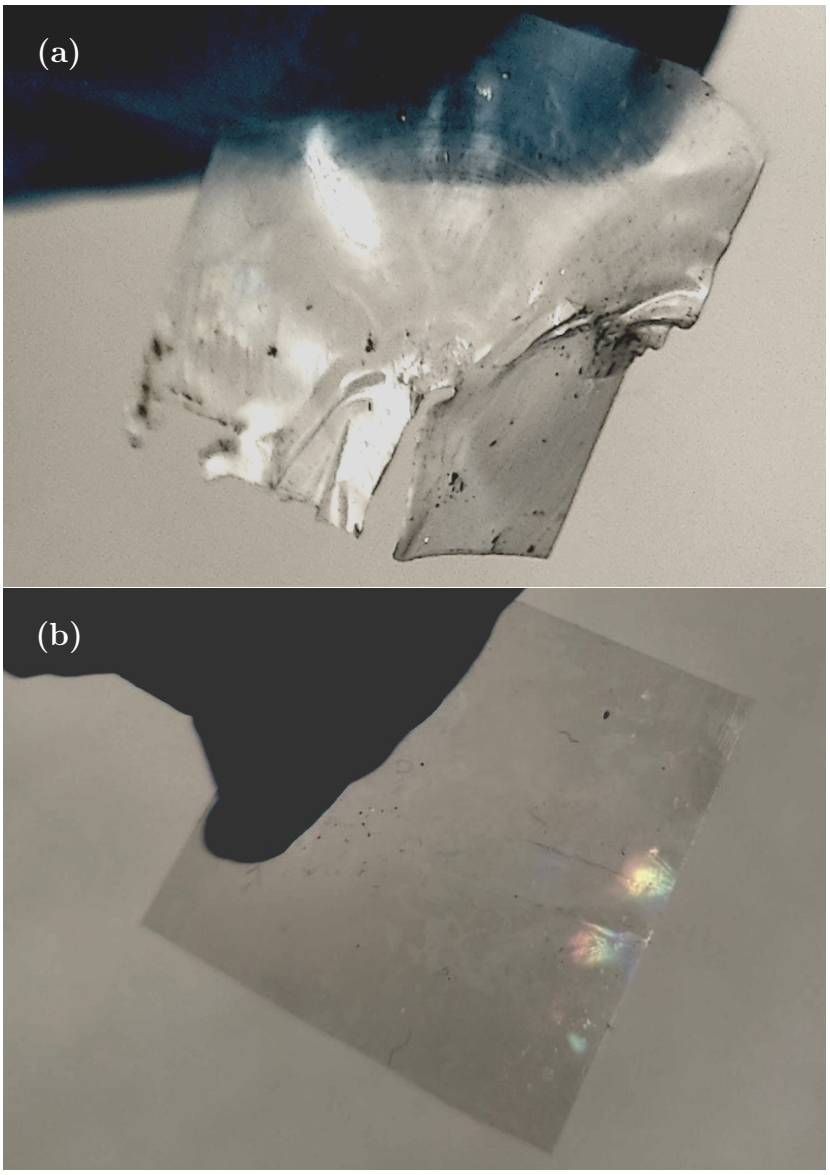

Figure 6.12: T-NIL Failure Modes, (a) Damaged Membrane with Concentric Rings and Debris from Si Ablation (b) Localised Patterning

polymer sandwich had done so. As such, a safety margin of 10 mins was applied. Cooling could be completed at atmospheric pressure, but was nonetheless a slow process as it was entirely passive. Furthermore, demoulding temperatures higher than 65°C have been found to induce significant plastic deformation of nanoimprinted PET (with 55°C being the optimal) [181], so demoulding could not be carried out prematurely.

In totality, the heating and cooling processes would take roughly 40 mins and 60 mins for imprinting at 120°C, respectively, or 90 mins for both when imprinting

at 200°C. With a generous imprint time allowed of 30 mins, a single imprint could take anywhere from 130 mins to 210 mins. If the entire 840 cm^2 print surface was used, this would correspond with a throughput of 4–6.46 cm^2/min. This could be increased if duplicate moulds and sliding plates were made, as these plates could be slid out, replaced, and left to cool while the new mould and plate assembly is heated. Furthermore, the heating of the new assembly would be faster as the surrounding T-NIL press would already be close to the necessary temperature. This could feasibly half the heating time (after the first batch) and completely remove the cooling time from the equation, resulting in an estimated throughput of 11.2–16.8 cm^2/min. At its best, this would still mean that a metre-wide square sail would take at least 10 hours of continuous imprinting; even for larger sails this would not be prohibitive, but it is certainly a significant disadvantage of this low-tech approach. Of course, more or larger vacuum ovens and presses could increase the throughput *ad infinitum*, but this would quickly become more expensive than the roll-to-roll methods that this low-tech approach attempts to circumvent. For the purposes of these trials, only one mould was available at a time, and so the throughput remained low. This had a knock-on effect on these trials, as typically only three imprints could be carried out within a single working day.

The other issue pertained to the quality of the imprints, which itself is a multi-faceted issue. No single trial with the custom T-NIL press was entirely successful: PET films were either damaged during the imprint or were only partially imprinted as seen in Fig. 6.12a-b. In both cases, it is clear that either imprinting is not occurring, or the pattern is being damaged before the end of the process. Damaged patterns were usually the more plausible explanation because most of these issues persisted even when stresses, temperatures and imprinting times were increased beyond what is normally required for successful PET T-NIL.

For the first failure mode seen in Fig. 6.12a, two root causes appear to be responsible for this damage. The first and most obvious was a fractured or ablated

Si mould, which could occur at higher pressures. This happened rarely, but was more frequent during tests that were intended to rule out the possibility of imprinting not occurring by using higher temperatures and pressures. At best, this would contaminate the film, and at worst, lead to puncturing and tearing. It was more common for this failure mode to occur in the absence of Si debris, which implies that plastic deformation is occurring on the PET surface. In the absence of mould damage, the obvious explanation for this plastic deformation would be demoulding forces [181]. However, there was very little evidence of demoulding being a factor, as the silanisation process had been effective: PET and the silanised Si appeared to adhere together very weakly. Two alternative mechanisms for this failure mode were identified: the first was thermal stresses during cooling, and the second was shear stresses during separation of the sliding plates.

The possibility of pattern deformation occurring due to thermal stresses was identified after the trials. This mechanism could explain why increasing the temperature did not increase pattern uniformity, but its existence is difficult to validate in post. Conversely, there is substantial evidence that shear stresses during the separation of sliding plates were responsible for failure mode one: concentric lines could be seen on the surface of many films experiencing this failure mode. This suggested that when relaxing the screw, the topmost sliding plate was not totally constrained, and in fact was able to rotate very slightly. This would cause the PET film to rotate upon the Si mould, which in turn would scratch and stretch the PET, overwriting its pattern with circular indents at the scale of the micro-prism pattern. Furthermore, the two plates would occasionally slide apart during retrieval, which usually coincided with no patterning being visible at all. This issue is essentially a design flaw in the T-NIL press: the two sliding plates were designed to allow for quick switchover of batches by sliding in and out of the press, but their relative motion is not constrained during this process; it would be prudent to have these plates interlock or mesh together for

subsequent designs.

The second failure mode suggests that either the pattern of the Si mould had been damaged, or pressure was not being applied uniformly about the membrane. In the case of mould damage, a likely cause would be the abrasion of the Si moulds over repeated imprints, particularly during the separation of the sliding plates. This further supports the notion that the relative motion of the sliding plates should be constrained during retrieval. In actual fact, SEM imagery on the reused moulds revealed that the patterns were not especially abraded. Non-uniform pressures are therefore the more likely source of this failure mode.

In this case, one possible mechanism would be the flexing of plates. This flexing was assumed to be constrained, and indeed, this mechanism is unlikely. This is because the Si moulds used in these small-scale trials were smaller than the diameter of the press; flexing would be most prominent at the perimeter of the plates, very low near the smaller press plate, and practically zero beneath the smaller press plate (where the Si mould was located). Another explanation may have to do with trapped air, as with this configuration, pressure had to be applied *before* achieving vacuum. It is therefore possible that pockets of air were preventing the PET from flowing into the cavities of the Si mould pattern. This could occur because the high pressure between the PET and Si mould could have made the PET into essentially an airtight wrap, preventing air from escaping while vacuum was being established. In support of this theory, it is notable that partial patterning always occurred near the perimeter of the film as shown in Fig. 6.12b, where it may have been easier for air to escape.

Finally, outgassing of the PET itself was not considered a likely mechanism for imprint failure, because no outgassing bubbles were ever observed. Of course, it is also noteworthy that these moulds were imprinting at the scale of a diffractive sail rather than a refractive one, and small non-uniformities that cause issues on the scale of the former may not have led to failure on the scale of the latter.

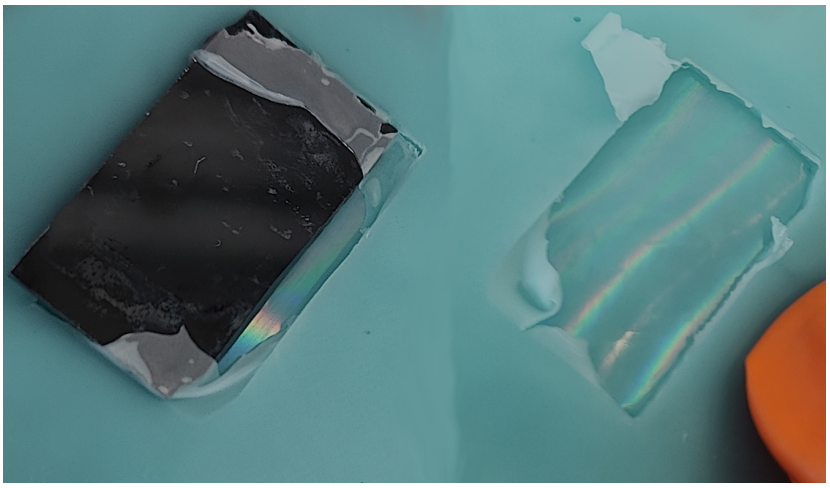

Figure 6.13: Pattern Transferred to Silicone from Si Mould

However, as the name implies, nanoimprinting should be more than capable of producing patterns at a 3 μm scale.

Whatever the reason, it was decided to explore the transferring of the pattern to a flexible silicone mould: such a mould could be perforated to prevent the trapping of air, and its ability to deform would allow it to conform to any non-uniform pressures that arose. These moulds would also be less likely to break, and would delay testing less if they did. The only drawback to such a mould is that it would be limited to lower pressures than Si, as a more malleable silicone pattern could be deformed at higher pressures. Even if this deformation were elastic, it would affect the geometry of the transferred pattern.

In any case, silicone transfer would also make for an effective means of scaling up Si or AM parent moulds, as their pattern could be transferred quickly and repeatedly to a larger silicone child mould. Naturally, anytime a mould transfers its pattern to another, the negative pattern space of the parent becomes positive space for the child, and vice versa. Si or AM moulds that are to undergo silicone pattern transfer should therefore be fabricated with the geometry of the final pattern rather than of the negative of said pattern.

Mold Start$^{\text{TM}}$ *16 FAST* was chosen due to its short cure times and suitability for

T-NIL; this rubber is heat resistant up to 232°C, which is much higher than the temperatures at which PET is nanoimprinted. First, an Si mould was treated with a silicone release agent and was placed inside a container of comparable diameter, with its pattern facing upwards. The silicone was then mixed with a platinum catalyst within a separate container and allowed to degas in a vacuum oven for 3 mins. Finally, the silicone-catalyst mixture was poured over the Si mould so that it was fully immersed within its container, and left to cure for 30 mins. The initial trial was successful as the pattern was transferred to the silicone as seen in Fig. 6.13; this was then trimmed away to form a mould. However, due to the sudden availability of an R2P solution, the linear T-NIL trials pivoted towards R2P (UV) NIL.

6.3.2 Rolling Pattern Transfer

R2P-NIL was carried out using the *Stensborg Desktop R2P Nanoimprinter* shown in Fig. 6.14. This printer has a 10.5 cm roller capable of imprinting at 0.5–500 cm/min depending on the configuration [123]. This corresponds with a throughput of 5.25–5250 cm^2/min, which could produce a metre-wide square sail in 31.7 hours at its slowest or 1.9 mins at its fastest. Once again, the 3 in^2 Si moulds produced in section 6.1.5 were used as a template, and so this imprinting would occur at diffractive scales (according to the limitation imparted by the 1.4 μm greyscale mask). The true rate achieved with these moulds was around 600 cm^2/min corresponding with a hypothetical 16.6 mins of imprinting for the aforementioned sail.

Because this system utilises UV curing, it is incompatible with thermoplastics such as PET. As such, a photopolymeric resin with comparable transmissivity to PET (up to wavelengths of $1,700$ nm) was applied as a 30 μm thick layer onto a PET film. This resin (Stensborg X30 [186]) is proprietary and not fully characterised. As such, it is suitable for process testing but is unverified for

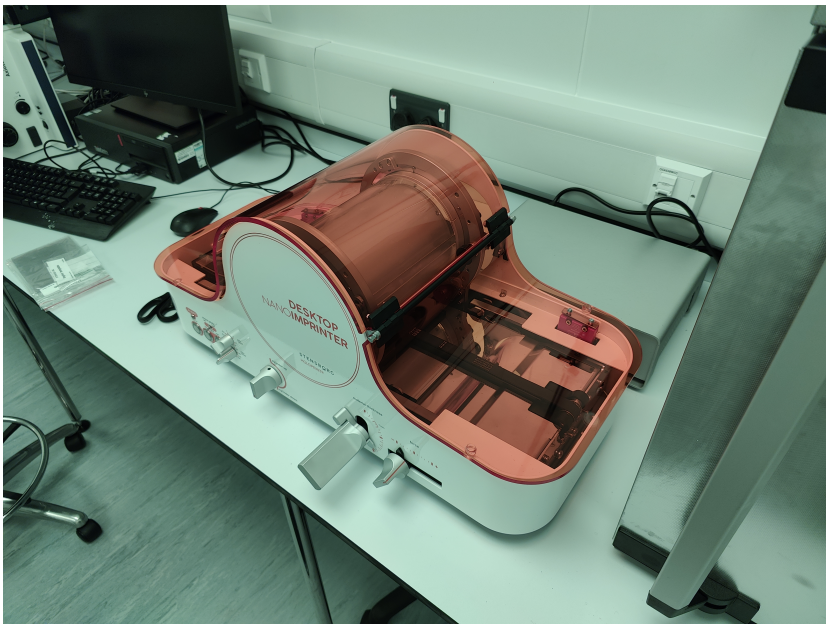

Figure 6.14: Stensborg Roll-to-Plate Nanoimprinter

flight. For a flight-ready sail patterned via this nanoimprinter, the real chosen resin would need to adhere to the criteria of the material review detailed in section 6.2.2, as for the PET itself. As mentioned in section 6.2.3, the suitability of photopolymers for spaceflight requires investigation as it pertains to optical degradation. However, the mechanical requirements may be somewhat relaxed for this resin if the thickness of the unpatterned PET is kept large relative to the thickness of the patterned photopolymer layer, as said layer would not be load bearing to the same degree as if it composed the entire membrane. Finally, the optical behaviour of a coated film differs slightly from that of homogeneous films, and because optical simulations assumed a single material, a design utilising this method would need to be optimised under a slightly revised model [46, 178].

Initial trials using this method were successful. As can be seen in Fig. 6.15, the imprinted film shows good parity with the Si pattern seen in Fig. 6.8. The smoothing seen in the latter is exaggerated by the former, and there is clearly room for both processes to be improved by subsequent trials. This could be rectified by adjusting the greyscale mask to account for such effects. Overall, while greyscale trial-and-error would be required to achieve perfect accuracy, the

resolution and precision of this method is demonstrated to be in excess of what is required of refractive sails, and is appropriate for diffractive sails.

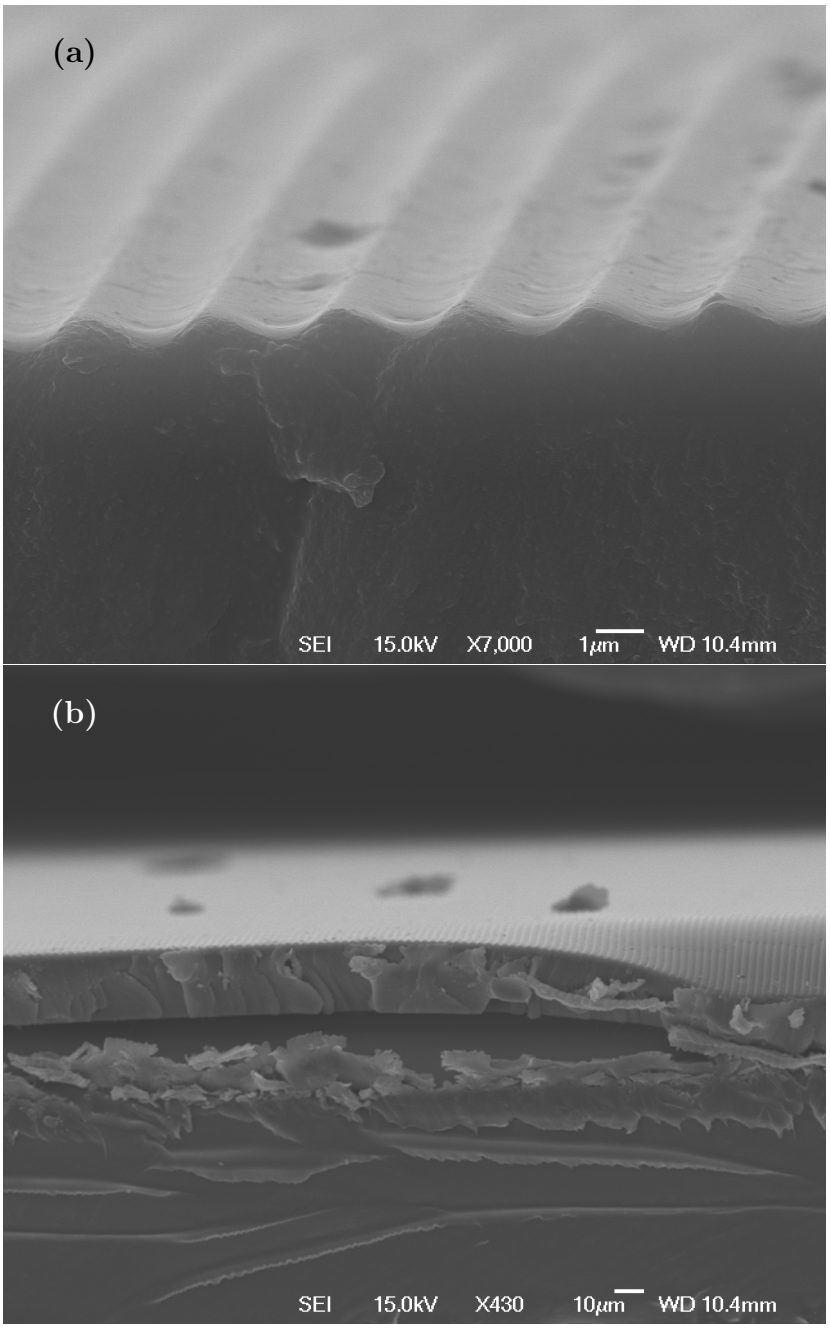

Figure 6.15: 3.0 μm-wide Micro-Prism Pattern Transferred via UV Roll-to-Roll Nanoimprint Lithography

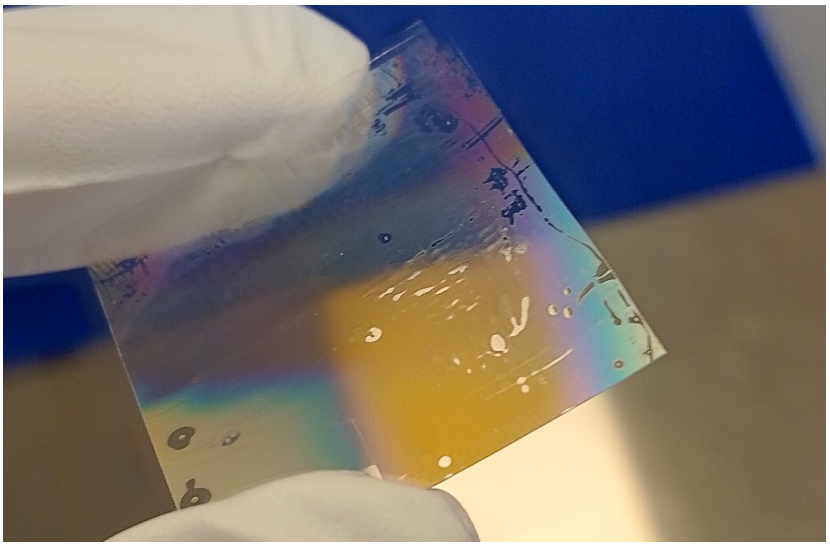

Figure 6.16: R2P Patterned Resin-PET Film

6.4 Evaluation of Processes as In-house or Outsourced Solutions

The sail manufacturing processes explored in this report are now evaluated according to how feasible and advantageous it is to execute them in-house versus outsourcing them as a commercial solution. This pertains to the expertise required to execute them effectively, whether the necessary infrastructure is likely already available to an engineering team, and crucially, the cost of procuring that infrastructure if it is not readily available, versus the cost of paying others to use their own as a service. The articles explored include:

1. Thin film fabrication versus commercial thin films.
2. Mould fabrication in-house versus as a service.
3. Pattern transfer in-house versus as a service.

It is assumed that sail design and assembly would be completed in-house, as this evaluation is from the perspective of a small satellite developer. Furthermore, the above list is in order of chronology, but the following discussion is in order of

certainty and is summarised by the list below:

1. Thin films: commercial solution unambiguously favoured.
2. Pattern transfer: the optimal solution is formulaic (capital and volume dependent) and fairly unambiguous.
3. Mould fabrication: ambiguous owing to the quantity of in-house and commercial solutions.

For thin films, it can be said with confidence that the commercial solution will be preferable for most sail developers, and small teams in particular. Although the processes involved are some of the simplest explored, and require some of the lowest investment in infrastructure, financially there is very little incentive to do this. This is because the cost of the raw materials is comparable to the cost of the pre-processed thin films. For example, a 50 μm thick PET (BoPET) film with dimensions 10 m $\times$ 0.4 m could be acquired for £259, and corresponds with about 0.276 kg of material. From the same source [187], the closest quantity of PET powder (to equally apply economies of scale) is 0.250 kg and is priced at £219. The cost-to-weight ratios of the processed films and raw materials are £938/kg and £875/kg, respectively, meaning that it is only 7% more expensive to buy ready made films. This discrepancy also does not account for the cost of solvents (o-chlorophenol or similar), nor the cost of waste incurred from failed processes or from infrastructure; these may widen the discrepancy considerably. Furthermore, additional processing costs may be incurred. For example, in the case of this PET powder, not only would it need to be formed into a thin film, but also it would need to be bi-axially oriented (made into BoPET) to endow it with the desired optical properties. This would require its own infrastructure and expertise at greater cost.

Conversely, the viability of pattern transfer as an in-house solution is largely a function of the available capital and the size or number of sails that are to

be produced. For a sail area of 1.5 m^2, using a small R2P nanoimprinter as a service was quoted at £2,098 corresponding with 1–2 days of work from a single engineer. Conversely, procuring the nanoimprinter itself for use in-house was quoted at £57,710. These quotes may be used as a case study, and extrapolated to approximate 'how much sail' a team would need to produce in order for the in-house investment to become optimal. In the case of R2P-NIL as a service, it can be assumed that the person-hours on the part of the engineer would incur most of the cost on the part of the provider, according to the low cost of materials that was discussed above. Of course, the provider will benefit from the economies of scale of raw materials, and this will almost certainly incur a discount on the part of the client — but as these materials are already a small contributor to overall cost, the discount will be proportionally small when compared to that of the materials themselves. Therefore, material costs — which benefit from economies of scale — can be said to be negligible, while person-hours — which do not benefit from economies of scale — can be said to be dominant. This simplifies things, as it can be assumed that the cost-to-imprint area is roughly constant as sail area is increased. Under this assumption, the imprinting of a sail as a service would be set to roughly £1,399/m^2 for this case study, requiring 41.25 m^2 of sail for the in-house investment of an R2P-NIL imprinter to be optimal. Of course, this is merely a benchmark, and this breakpoint will vary from provider to provider.

In any case, if the desired sail area is too low to justify the investment, the commercial solution becomes optimal *unless* the cost-to-imprint area is prohibitively high for a given team. If so, building a simple but low throughput nanoimprinter in-house becomes 'optimal' in that it is the only remaining option. A CNC machine is likely available to a satellite development team, but in its absence, can be used as a service by many providers for a low hourly rate of around £200 [188]. Using the previous trials as a case study, the cost of materials for the T-NIL press only came to £64.30. Vacuum ovens (which could also be used for silanisation in lieu of a vacuum desiccator) can be found from a number of sources for less than

£500. In the absence of any other infrastructure, a setup for pattern transfer could therefore be established for around £800 before materials. However, as established in section 6.3, carrying out this process 'from scratch' is non-trivial and would likely incur significant cost in terms of time if not capital.

Conversely, mould fabrication is much more convoluted to price according to the number of potential solutions. Furthermore, it may be viable to fabricate a small mould via one method and scale it up using sequential imprints into silicone as discussed in section 6.3. On the additive side, the most successful trials were carried out via 2PP. However, typical 2PP machines have far too low a throughput, and these trials were only feasible using a specific machine — the UpNano NanoOne 1000 — and which also happens to be amongst the most expensive 3D printers on the market. Its exact price varies by quote but is certainly greater than £375,000 [189]. Naturally, this will be prohibitive for many teams. μSLA and DLP machines capable of the required resolutions are themselves uncommon, and come with a lower but nonetheless substantial price tag. For example, the DLP machine used in our trials is typically priced at around £187,730 [190]. This is in contrast to traditional SLA printers, which may be acquired for less than £200. As a service, AM may be more affordable; online quotations place parts produced via μSLA at around £23/cm^2 for a 3 mm thick mould [191]; extrapolated to achieve the dimensions of the mould used for the aforementioned R2P-NIL trial, this would cost £445. Of course, smaller moulds may be feasible and the true cost may vary. In terms of conventional fabrication, some form of greyscale lithography (photolithography or EBL) combined with etching is invariably the optimal solution. As an in-house investment, industrial photolithography machines may cost over £7 million [192]; based on a previous quote and adjusted for inflation, the EBL machine used in our trials would cost approximately £1 million today. As a service, the University of Nottingham provides a useful benchmark via its commercial services: photolithography and subsequent dry etching are provided from £150/hour and £105/hour, respectively [193]. While the etch-

ing step typically can be completed within an hour, photolithography may take several hours; at the scale of the Si mould used in trials, this would generally cost in excess of £1,000. On a surface level, this would appear to favour AM as a service. However, the number of available services and the opaque pricing practices of conventional and AM mould fabrication services makes this far from conclusive. Furthermore, in support of photolithography is the fact that the only AM method that can compete with its resolution and consistency is 2PP — which is not available as a service.

Despite the high cost of many of these processes, it is worth remembering that the upfront cost of the greatest solar sail competitor — the electric thruster — typically ranges from £32,000-150,000 [194, 195]. Several manufacturing processes have been demonstrated that would enable the fabrication of a transmissive sail at a fraction of the cost. Furthermore, with access to high throughput systems such as R2R/R2P-NIL, the cost-per-unit area of a sail will not necessarily increase as it is scaled up. In fact, it may decrease as material acquisitions will benefit from economies of scale. This is in contrast to electric thrusters, which incur exponentially greater cost in terms of both hardware and onboard propellant as the requirements of a mission grows, and regularly incurs costs in excess of £1 million for larger satellites [196].

6.5 Solar Sail Payload

Another aspect of this project was the development of a solar sail payload for the *JamSail* CubeSat in collaboration with the *NottSpace* CubeSat Programme at the University of Nottingham. The in-depth design of the payload began with the origami folding pattern. Alongside the dimensional constraints of the hub, this would determine the size of the sail that could be launched. Early on, the design was constrained to a *Miura-ori* crease pattern applied to a *flasher* configuration.

This crease pattern describes a membrane tessellated by parallelograms, which the flasher configuration harnesses to deploy in a spiral motion. As such, both the folded form and deployed form of a flasher has rotational symmetry. This configuration was chosen according to the existence of literature that advocated for it as an optimal solution for deployment systems of this size [197–199].

Miura-ori flasher designs were generated using *Mathematica* via a modified version of the workbook *Tessellatica* [200]. Miura-ori flashers can be described according to four parameters, and which serve as the input for the generation of designs:

1. r_{ori} — the number of rings.
2. h_{ori} — the height order; each ring is h_{ori} panels high.
3. m_{ori} — is the rotational order/symmetry. $m_{\text{ori}} = 4$ for square, $m_{\text{ori}} = 6$ for hexagonal, and so forth.
4. dr_{ori} — is the radial increment; vertices at the same height in the folded form are spaced radially by this amount (relative to the radial distance of the corresponding inner polygon vertex). This is the value that results in the skewed geometry of the crease pattern; the larger the value of dr_{ori}, the faster that outer rings grow in height and the more curved the diagonal folds become. dr_{ori} is used to accommodate for thickness.

The designs that are generated by Tessellatica are normalised relative to the height of the folded form of the flasher H_{ori} such that $H_{\text{ori}} = 1$ in some undefined unit. This was reconfigured so that they would be normalised according to the folded form diameter W_{ori}, which was fixed according to the available space within the satellite (minus the estimated wall thickness of the hub). Initially, the design focused on a 'TunaCan' configuration wherein the sail would be stowed externally. According to CubeSat Design Specification Rev. 14 [201], this constrained the maximum dimensions of the folded form flasher to $H_{\text{ori}} \leq 3.6$ cm, $W_{\text{ori}} \leq 6.4$ cm.

The latter was constrained further to $W_{\text{ori}} \leq 5.9$ cm to account for hub wall thickness. The initial design sought only to maximise sail area, and as a result, was very complex. This design had parameters $\{r_{\text{ori}} = 5,\ h_{\text{ori}} = 7,\ m_{\text{ori}} = 8,\ dr_{\text{ori}} = 0.015\}$ and is depicted by Fig. 6.17. With folded form diameter fixed to $W_{\text{ori}} = 5.9$ cm this design demonstrated $H_{\text{ori}} = 3.43$ cm, which is compatible with a TunaCan. However, these constraints had a significant effect on the deployed sail, which registered a circumcircle diameter of only 0.93 m and an approximate area of only 0.68 m^2.

It was decided to revert to an internally stowed payload for a number of reasons. The first reason was that sail area showed greater sensitivity to folded form diameter than height. Stowing the sail internally would relax the W_{ori} constraint to $W_{\text{ori}} \leq 10$ cm (9 cm with a margin for wall thickness) and therefore allow for a more reasonable crease patter and — so long as the H_{ori} constraint were not made disproportionately small by this change — larger deployed sails. The second reason was that additional space had been freed up within the satellite bus due to the adoption of a '3U' form factor ($30 \times 10 \times 10$ cm bus). This meant that the H_{ori} constraint could be relaxed by this change as well. Although the exact available space is in flux according to revisions occurring for other systems, a preliminary constraint of $H_{\text{ori}} = 7$ cm was prescribed.

Although the layout of functional surfaces was largely carried over from the TunaCan design, the crease pattern itself was simplified with parameters $\{r_{\text{ori}} = 8,\ h_{\text{ori}} = 1,\ m_{\text{ori}} = 4,\ dr_{\text{ori}} = 0.005\}$. The most notable outcome of this simplification is that the sail became approximately square, with only four flat springs needed, and a much simpler to realise crease pattern. With folded form diameter fixed to $W_{\text{ori}} = 9$ cm, this design (shown in Fig. 6.18) remained reasonably compact, demonstrating $H_{\text{ori}} = 6.83$ cm. When deployed, the sail has a circumcircle diameter of 1.67 m and an approximate area of 1.68 m^2.

The solar sail payload can be partitioned into three elements: the sail mem-

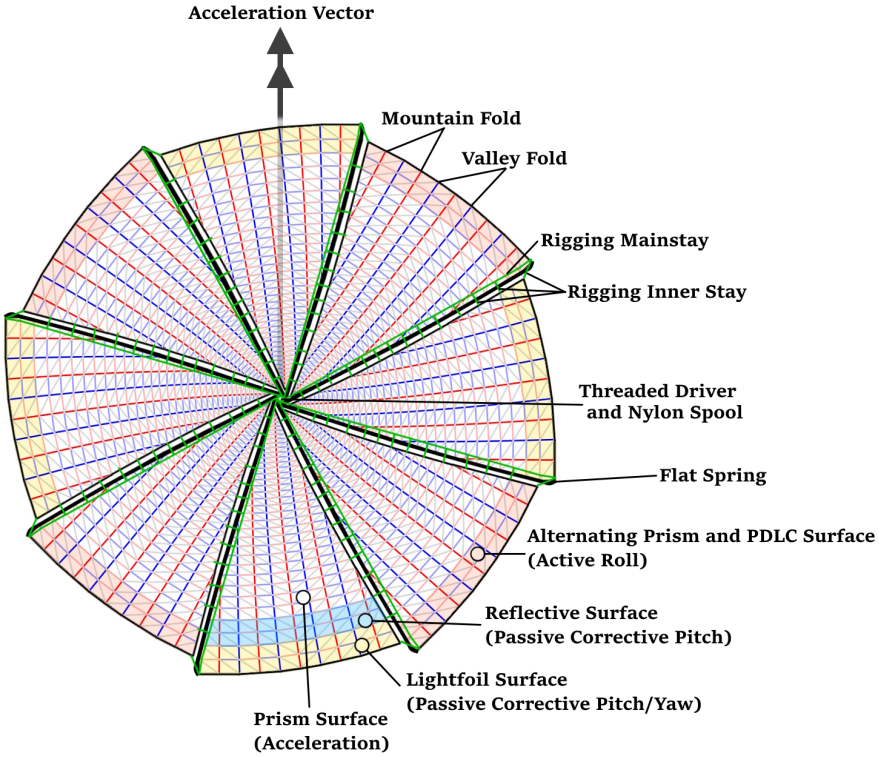

Figure 6.17: Deployed Sail Original Concept

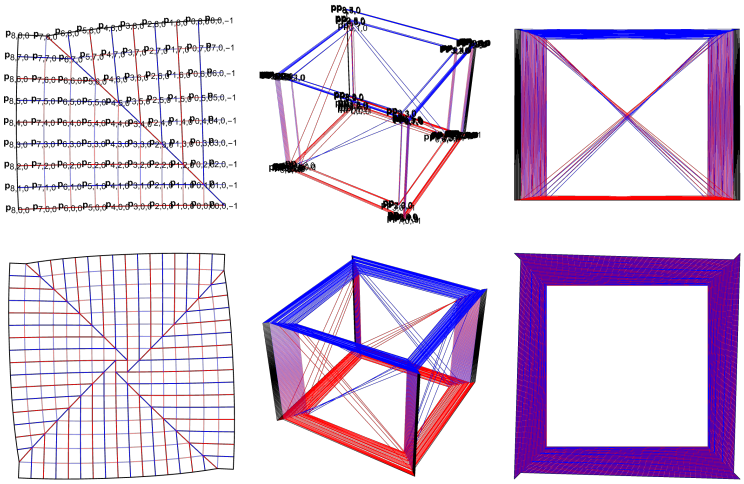

Figure 6.18: Simplified Design, Miura-ori Flasher for CubeSat-deployed Transmissive Sail

brane, the static assembly and the moving assembly. The sail membrane is a 50 μm thick optical film composed of BoPET. The majority of this film is unmetallised and embossed with a micro-prism pattern. An onboard magnetorquer is utilised by JamSail for attitude control, and is supplemented by the solar sail

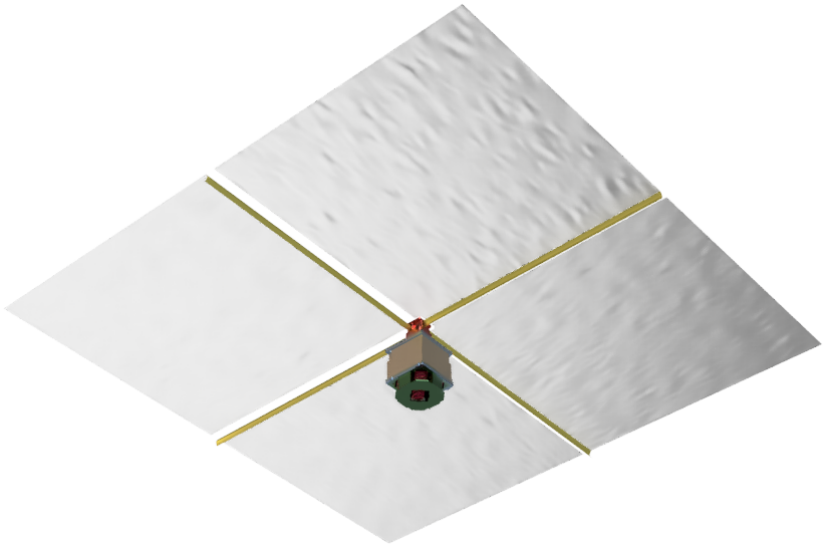

Figure 6.19: Render of Deployed Sail Payload (No Satellite Bus)

via an alternatively patterned region around the sail perimeter, wherein the pattern comprises an array of self-stabilising optical elements derived from *lightfoils*. These geometries will cause the sail to naturally oscillate about the Sun-pointing attitude in the absence of active control, and expedite the process of converging upon said attitude in its presence. Furthermore, a small section of sail at the anterior edge is metallised; here, SRP is generated normal to the sail through conventional reflective means in order to counteract the torque generated by the micro-prism regions; this is necessary because SRP that acts parallel to the sail also acts perpendicular to any radial vector drawn from the satellite centre of mass. This membrane is attached to the rest of the payload by way of four flat springs. These are fastened to the moving assembly in turn, which is constrained by the static assembly.

This static assembly serves as the structural interface between the payload and the JamSail bus, houses the membrane, and prevents the moving assembly from being ejected during and after deployment. It is principally composed of a static hub (beige, Fig. 6.20) that is fastened to the satellite bus via screws located at the sail plate above, and a reinforcement ring below (blue, Fig. 6.20). This hub also features a membrane cavity, within which the membrane and flat springs are

stowed under tension. Furthermore, beneath the static hub lies the compression spring housing (green, Fig. 6.20). This constrains the moving assembly prior to deployment and serves to 'sandwich' the four compression springs, which are compressed between the spring dock of the static assembly and the spring moving dock of the moving assembly.

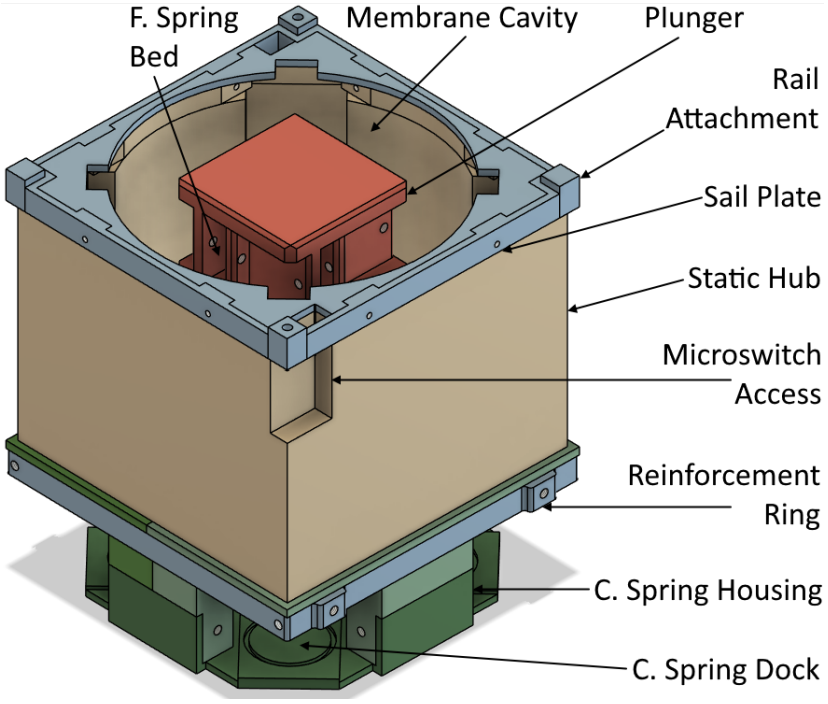

Figure 6.20: Stowed Sail Payload Annotated without Flat Springs, Compression Springs or Motor Assembly

The core component of the moving assembly is the plunger (red, Fig. 6.20), within which the four coiled flat springs are fastened, and upon which the membrane rests prior to deployment. This assembly is initially kept static by the aforementioned compression spring housing of the static assembly and by a temporary wire fastening that moors it to the static assembly. At the start of deployment, these fastenings are severed by way of a thermocutter, causing the moving assembly to rise under the influence of the four compression springs. These act upon the plunger via the four moving spring docks (see Fig. 6.21a). After the flat springs atop the plunger are free of the static housing, they are able to spring outwards and thereby unfurl the membrane. The moving assembly is prevented from being

ejected from the satellite by the sail plate, which partially overlaps with the (now extended) plunger at four locations that lie towards the rail attachment points (see Fig. 6.21b).

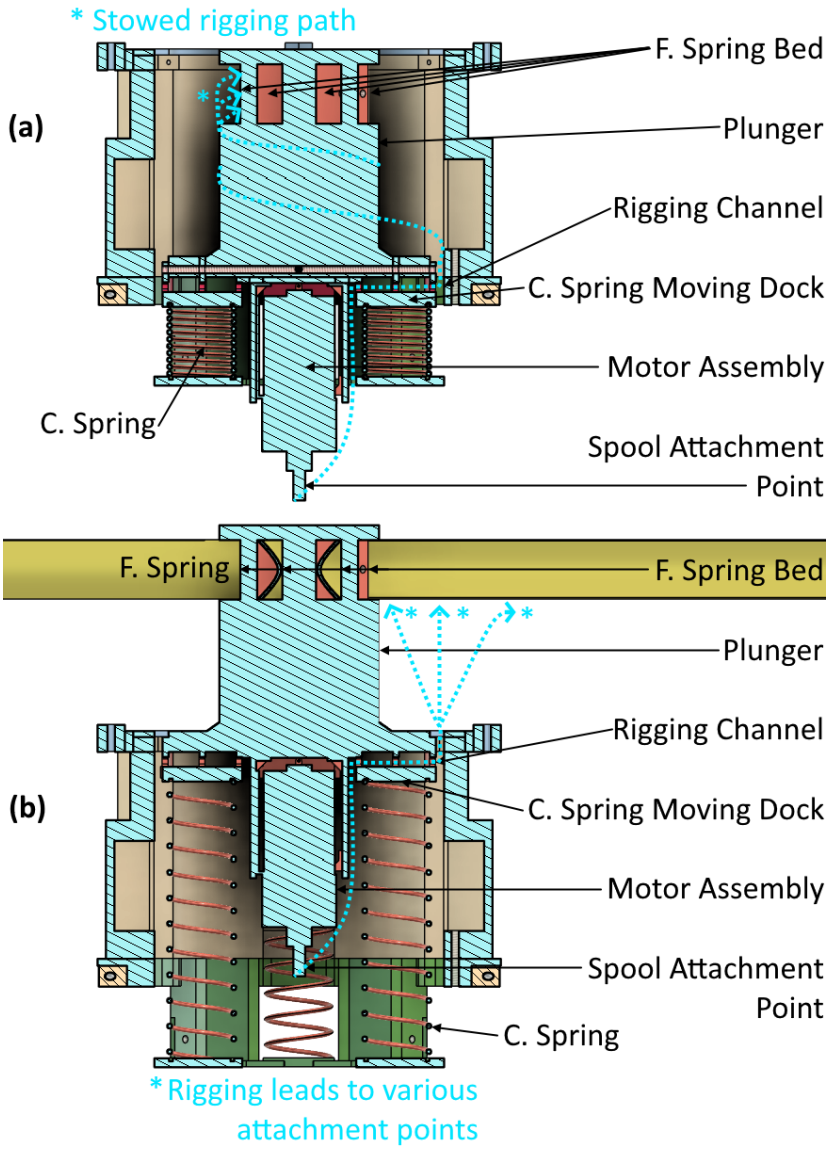

Figure 6.21: Sail Payload Diagonal Section, Moving Assembly Annotated (a) Stowed (b) Deployed

The moving assembly would be functional with these elements alone. However, several design stressors are placed upon it by the refractive sail membrane. The first is the need to maintain a high degree of sail flatness. A refractive sail has greater sensitivity to solar incidence angle than its reflective contemporaries (as demonstrated in chapter 4), and so an exaggerated sensitivity to wrinkling and

billowing may be inferred. Furthermore, unmetallised, patterned polymer membranes are atypical of spaceflight, and their tensile performance — particularly under the effects of radiation and vacuum — are not well documented. Greater control over membrane tension is therefore desirable, particularly during deployment and the later stages of the JamSail mission. Both issues can be mitigated through the active control of sail tension. To achieve this, first the sail membrane is segmented into four membranes; during deployment, each membrane segment is only constrained by the single flat spring to which it is attached, allowing it to deploy with theoretically zero tension. In practice, some tension is needed to enable the folded sail to unfurl. The degree of tension that is supplied to the membrane is controlled via nylon rigging that is sewn through the unconstrained sail edge nearest to the flat spring of the neighbouring membrane. It should be noted that the ratio of tension supplied to the sail near the spring tip relative to the spring base cannot be controlled actively, and must be determined prior to launch. This is because, despite having several attachment points spanning the flat spring and the neighbouring free membrane edge, each rig is principally coiled into a single coil of wire. It is only by merit of having different lengths that they are able to 'fray' into distinct wires with different attachment points. Conversely, the net tension can be controlled actively during a mission by way of a motorised spool. One such wire coil is attributed to each flat spring, and all four share this common spool. Said spool is attached to a single motor, which in turn is attached to the underside of the plunger and rises with it during deployment (see Fig. 6.21a-6.21b). This is so that the rigging has zero motion relative to the moving assembly during deployment; otherwise, sail tension would spike at a critical moment. This motorised spool ensures that tension can be controlled at every stage of the mission, and may be used to bring the membrane to failure once a timely deorbit has been guaranteed by the sail, thereby generating useful data regarding unmetallised sail membrane performance in space. However, this configuration comes at the cost of complexity and mass, and may be unnecessary

for future missions once the relevant literature has been expanded.

Chapter 7

Conclusions

The mechanisms for transmissive solar sailing and their various designs have been explored. Performance has been modelled for these sails with approximated α-sensitivity (informed by a custom ray-tracing optical force simulation) and a γ-adaptive locally optimal steering law for a variety of flight regimes, with a particular focus on LEO flight and the effects of atmospheric drag, eclipse and orbital precession. The results have been compared with those of similarly-modelled reflective solar sails, as well as cold gas and electric thrusters. As a breed, transmissive sails were shown to have lower sensitivity to altitude h but greater sensitivity to inclination i; they could operate at lower altitudes than reflective sails, but saw their performance disproportionately eroded at lower inclination orbits. Nonetheless, these sails often demonstrated higher performance even within low inclination orbits by merit of their peak performance being higher. The transmissive sails were then partitioned into three groups based on their performance and behaviour: (A) high performance archetypal, (B) lower performance archetypal, and (C) lower performance non-archetypal sails. Type A demonstrated the highest rate change of orbital energy $\dot{\epsilon}$, the shortest transit times $t_{h_1 \to h_2}$, and outperformed reflective sails in every flight regime. Type B was the lowest performing, and only found advantage over reflective sails at low altitudes and at inclinations

near i_{SSO}. Type C were amongst the most robust, and demonstrated some of the least sensitivity to inclination i or Sun-velocity angle γ due to their large normal component to SRP, which allowed them to situationally mimic the behaviour of reflective solar sails; these also tended to outperform reflective sails in every flight regime. When compared with thrusters, it was shown that medium-sized 36 m^2 transmissive sails performing an orbit-raising manoeuvre from a 530 km SSO generated comparable rates of acceleration to mid-range electric thrusters, but were outclassed in this respect by Hall effect and cold gas thrusters. The sensitivity of the $t_{530\rightarrow h_2}$ of a Type A transmissive sail to i was shown to be substantial — for example, differing by up to 440% between $i = 30°$ and $i = i_{\mathrm{SSO}}$, with failure to escape the atmosphere at $i = 0°$. Naturally, thruster transit times were independent of inclination. Based on the theoretical performance alone, transmissive sails are suggested to be valid solutions for modern satellites completing a variety of missions in LEO and beyond; these demonstrate greater operational flexibility than their reflective peers, and are suitable for a wider range of flight regimes. However, there still exist flight regimes for which only thrusters are suitable — particularly low inclination, very low altitude orbits. In the short term, Type C and B solar sails may provide an inexpensive, high ΔV system for escaping Earth's atmosphere; in the medium term, Type A or C sails (perhaps metasails) may be used for contemporary satellite applications, and may enable edge-of-the-envelope missions such as reusable space debris mitigation. The adoption of these sails over thrusters for contemporary LEO applications may be encouraged due to their ability to passively deorbit defunct satellites, the growing issue of space debris, and the narrowing window of time allowed for satellites to deorbit [94]. In the long term, ISAM may enable solar sails of arbitrary size to be fabricated in space [1]; these sails may be able to out-accelerate chemical thrusters (which see depreciating gains when increasing their size, as described by the *Tsiolkovsky rocket equation*); if realised via transmissive designs, these sails may retain these properties even within LEO. Very large sails or swarms of

smaller sails could be used for applications ranging from manned interplanetary flight to asteroid towing. The practical advantages and limiting factors for the introduction of transmissive solar sails were also discussed, with several requiring further research — particularly the suitability of metamaterials and the optical degradation of non-metallised optical films in space.

An optical simulation was developed that used ray tracing to model the solar radiation pressure and torque per unit area generated by illuminated, refractive solar sail patterns. This was developed in tandem with a numerical optimiser that used model-free reinforcement learning, allowing for designs to be optimised according to a user-defined fitness function. Generative designs focused on sails composed of optical polymer and were optimised for either tangential-to-sail solar radiation pressure $F_{S_{yb}}$ or self-stabilising corrective torque $\tau^*_{S_{yb}}$ at a single solar incidence α or range of α, with an emphasis on near-zero α (Sun-pointing) performance. Designs varied substantially between optimisation runs. However, *almost* every optimised design utilised total internal reflection (TIR) as the primary mechanism for generating $F_{S_{yb}}$ or $\tau^*_{S_{yb}}$, suggesting that pure refraction is typically sub-optimal. A variety of novel designs were generated by this process. For SRP optimisation, the optimiser naturally sought to maximise ray alignment with $-\hat{\mathbf{y}}_\mathbf{b}$. High flux, first-order rays were invariably prioritised, but the best designs also considered the ray paths of lower flux, second-order rays borne of partial reflection (rays of a higher order than the second-order were rarely considered). For single optical elements, achieving nearly absolute $-\hat{\mathbf{y}}_\mathbf{b}$ alignment was straightforward and optimal. However, patterns attempting to do the same would induce ray *pattern propagation*, which was more complex to design for. The behaviours that the optimiser encouraged fell under one of two archetypes: *pattern skimming* — wherein first-order ray ray alignment was maximised to the greatest extent possible without inducing pattern propagation — and *harnessing* — wherein the optimised patterns utilised pattern propagation to improve ray alignment. Generally, pattern skimming was the favoured strat-

egy, but many patterns harnessed pattern propagation in a mitigating capacity. For example, pattern skimming was carried out more aggressively and effectively by patterns that could effectively redirect rays that failed in the former endeavour via pattern propagation. For stability optimisation, the strategy utilised by the optimiser depended on whether the centre of mass was dictated by the sail itself, or by a greater satellite. For the former, performance was increased by promoting first-order ray TIR at the base of the pattern, ensuring ejection at the upper cambered surface with $\hat{\mathbf{y}}_\mathbf{b}$ alignment of opposite sign to that of α. For the latter, performance was increased by promoting pattern skimming behaviour, again with $\hat{\mathbf{y}}_\mathbf{b}$ alignment of opposite sign to α. Higher order rays had very little contribution to fitness in both cases and were largely ignored by the optimiser. The latter could imply either the improving of alignment for rays with already favourable alignment, or mitigating the poor alignment of rays with poor alignment. Broadly speaking, the optimisation runs that were carried out could be partitioned by: micro-prism optimisation for SRP, freeform element optimisation for SRP, or freeform element optimisation for stability. In one case, numerically optimised micro-prisms were compared with an analytical solution. The numerically optimised patterns demonstrated zero α $F_{S_{yb}}$ and peak $F_{S_{yb}}$ of up to 3.07 μPa and 2.72 μPa, respectively, representing an increase over the analytical solution of up to 43% and 58%. In another case, a pattern of freeform optical elements was generated that registered zero α $F_{S_{yb}}$ of 3.40 μPa, which is 74% of the theoretical maximum. Likewise, during a stability optimisation run, a pattern of freeform optical elements derived from lightfoils (scale factor $\Sigma = 10^{-5}$) demonstrated peak corrective torque $\tau^*_{S_{yb}}$ of 9.38 pNm/m^2, representing an increase of 74% from that of the initial lightfoils. It was also demonstrated that there exist geometries that can generate similar and higher magnitude self-stabilising torques than lightfoils, particularly when satellite centre of mass is considered instead of merely that of the sail. Nevertheless, these patterns can only be called optimal within the purview of their fitness functions: increasing fitness according

to these functions often came at the cost of a performance deficit that lay beyond the scope of said functions. For example, stability optimisation utilised a fitness function that optimised for $-20^\circ \leq \alpha \leq 20^\circ$ and resulted in higher performance within this range, but at the cost of the total range over which stability could be achieved. On the other hand, SRP optimisation typically greatly increased $F_{S_{yb}}$, but at the expense of the pressure ratio λ, which was not considered by most fitness functions. Such designs will be optimal for certain manoeuvres, but not for every manoeuvre, emphasising the importance of tailoring the fitness function to a specific mission, flight regime or manoeuvre when utilising numerical optimisation. Conversely, the fact that performance gains were demonstrated in such a specific manner according to the user-defined fitness functions should commend this method for such applications. The model-free reinforcement learning optimiser proved to a reliable method for improving performance according to user-defined fitness. However, the efficacy of the method was sensitive to optimiser configuration and initial conditions. Furthermore, while two differently configured optimisation runs with identical optimisation criteria would sometimes converge upon the same solution, they would just as often diverge to output very different solutions. This suggests that repeated runs with a variety of different optimisation parameters would be required in order to carry out an exhaustive pattern optimisation. Furthermore, while very good solutions may be reliably found by this method, a globally optimal solution cannot be guaranteed.

Each aspect of sail manufacture was also explored, trialled and evaluated. The feasibility and cost of executing each process in-house was explored against the cost of outsourcing. Low-volume thin film fabrication trials were carried out using doctor blading and a water-float method and were compared with the commercial solution. Conventional and additive manufacturing (AM) methods were compared for the fabricating of sail patterning moulds. Greyscale (electron beam) lithography and dry plasma etching comprised the conventional manufacturing methods trialled, while digital light processing (DLP) and two-photon polymeri-

sation (2PP) comprised the AM methods. Finally, pattern transfer processes were explored, with trials comprising silicone injection moulding and nanoimprint lithography. Nanoimprinting trials included the testing of a bespoke, low-tech, thermal nanoimprinter for usage within a vacuum oven, as well as the trialling of a commercial roll-to-plate UV nanoimprinter. Finally, a prototype solar sail payload was designed for the University of Nottingham CubeSat *JamSail* in preparation for a future in-orbit demonstration mission.

7.1 Recommendations for Future Work

During chapter 4, flight regimes were examined with respect to either altitude or one of the misalignment angles (γ, Γ or i). If altitude was fixed, the relevant angle would be varied, and *vice versa*. Doing so yields a 2D slice of the true flight envelope; future works may wish to consider producing a 3D flight envelope, for example, comprising mutually orthogonal axes for altitude h, velocity misalignment γ and rate change of orbital specific energy $\dot{\epsilon}$.

The sensitivity of various transmissive solar sails with respect to solar incidence α was also only approximated by a generic sensitivity model. This model was informed by an early version of the optical force model and also did not account for the more unique behaviours of certain designs — such as those that were capable of electro-optical tuning of their SRP profiles. Future works may improve upon the characterisation of these sails for LEO flight by directly importing performance data from an optical simulation into a flight model. Other assumptions that were applied for the sake of scope could also be removed — for example, examining the LEO behaviour of sails during different stages of the solar cycle, accounting for aerodynamic lift of flat or billowed sails exposed to an airflow, or accounting for the fact that the J2 perturbation is a function of equatorial inclination rather than ecliptic inclination. Future case studies should focus on verifying

the suitability of these sails for specific missions — for example, demonstrating the rate at which debris objects could be de-orbited via cyclically performing orbit-raising, rendezvous and orbit-lowering manoeuvres.

Furthermore, the solution space for transmissive sail patterns remains vast and several phenomena have not been explored in this report. Based on the outputs of chapter 5, the author makes the following recommendations:

1. Future models should account for absorption owing to the utility of this phenomenon as a natural means of attenuating and culling rays. This may reduce computation times. Furthermore, certain patterns encouraged long ray paths, thereby challenging the assumption that absorption has a negligible contribution to SRP for thin films. A model that accounts for absorption, emission and wavelength-dependent absorptivity, reflectivity and transmissivity would be both highly accurate and well-poised to answer questions pertaining to the suitability of these membranes for spaceflight. The infrared absorption and thermal stability of these membranes is particularly interesting in the context of mitigating optical degradation.

2. The current model can only be used for the designing of refractive solar sails, which is not the whole solution space. Future work could involve the development of a ray tracer that is capable of emulating diffraction (for which there is some precedence [202]) in order to explore transmissive sails more generally. Additionally, future models may wish to account for polarity, birefringence, and the different behaviour of coated films [46]. The latter in particular may be important for modelling the behaviour of sails that are patterned via UV nanoimprint lithography. Also note that to be truly 'general', an optical force and torque model would need to be compatible with metamaterials.

3. Future works would benefit from a bespoke convex hull calculation algo-

rithm. MATLAB's in-built convex hull algorithms behaved sub-optimally when paired with the numerical optimiser, particularly because of the 'redundant vertex' issue that degraded the resolution of optimised geometries. In particular, this reduced the effectiveness of the optimisation process for geometries that were constrained with base edge collinearity (ensuring that they could be fabricated via conventional patter transfer methods).

4. Many configurations remain unexplored. For one, multiple refractive (or diffractive) films may be layered. Non-uniform patterns or patterns with smaller, interstitial optical elements may also exhibit unique behaviours. Furthermore, a reflective film may be placed underneath a transmissive one to maximise the net SRP generated (though this would not necessarily improve transverse SRP). Finally, solar collector configurations could also be explored. Focussing sunlight onto a small, actuated mirror could allow a hybrid transmissive sail to effectively 'tack without tacking'; maintaining a Sun-pointing attitude for the transmissive solar collector while tacking the small mirror would allow such a sail to access the low altitude performance benefits of a transmissive sail while retaining the high efficiency of a reflective one.

As it pertains to manufacturability, the greatest unknowns pertain to metasails. Certain designs proposed in previous literature are predicated upon the existence of scalable, high-performance metamaterials that can survive spaceflight. The suitability for space of higher readiness designs is also not perfectly understood, particularly as it pertains to photopolymers used in certain UV-sensitive imprinting processes. Furthermore, though the design of a payload has been completed and the relevant processes have been detailed, more work is needed to actualise a real prototype. For example, an origami jig must be designed so as to enable the accurate folding of a large sail, and deployment tests must be carried out; this is non-trivial, as the additional thickness that a transmissive sail entails is an

additional stressor placed upon the stowing and deployment systems. Similarly, the pattern transfer process must be applied to a to-scale membrane.

In the case of the UV-R2R nanoimprint lithography trials, the proprietary resin being used would need to be characterised. If its performance proved to be significantly lower than that of PET, differently-configured optimisation runs would be needed, alternative resins may need to be chosen, or a thermal R2P/R2R imprinter that is compatible with PET would need to be secured.

References

[1] E. Sacco, S. K. Moon, Additive manufacturing for space: status and promises, Int. J. Adv. Manuf. Technol. 105 (10) (2019) 4123–4146. doi:10.1007/s00170-019-03786-z.

[2] B. Fu, E. Sperber, F. Eke, Solar sail technology—A state of the art review, Prog. Aerosp. Sci. 86 (2016) 1–19. doi:10.1016/j.paerosci.2016.07.001.

[3] S. Hassanpour, C. J. Damaren, Collocated attitude and vibrations control for square solar sails with tip vanes, Acta Astronaut. 166 (October 2018) (2020) 482–492. doi:10.1016/j.actaastro.2019.07.038.

[4] A. J. Westfall, Design of an Attitude Control System for Spin-Axis Control of a 3U CubeSat (2015). doi:10.31979/etd.u6fr-6t4m.

[5] Earth Observation Portal, JAXA, IKAROS (Interplanetary Kite-craft Accelerated by Radiation of the Sun) (2010).
URL https://www.eoportal.org/satellite-missions/ikaros#mission-status

[6] Earth Observation Portal, NASA, NanoSail-D2 (2010).
URL https://directory.eoportal.org/satellite-missions/nanosail-d2

[7] Earth Observation Portal, The Planetary Society, LightSail Missions (2019).
URL https://directory.eoportal.org/satellite-missions/nanosail-d2

[8] K. Wilkie, J. Fernandez, Advanced composite solar sail system (ACS3) mission update, in: The 6th International Symposium on Space Sailing, 2023.
URL https://ntrs.nasa.gov/citations/20230008378

[9] R. Y. Kezerashvili, Solar Sail: Materials and Space Environmental Effects, Adv. Sol. Sail. (2014) 573–592.

[10] D. Lev, R. M. Myers, K. M. Lemmer, J. Kolbeck, H. Koizumi, K. Polzin, The technological and commercial expansion of electric propulsion, Acta Astronaut. 159 (February) (2019) 213–227. doi:10.1016/j.actaastro.2019.03.058.

[11] A. D. Pathak, S. Saha, V. K. Bharti, M. M. Gaikwad, C. S. Sharma, A review on battery technology for space application, J. Energy Storage 61 (2023) 106792. doi:10.1016/j.est.2023.106792.

[12] P. D. Fieseler, A method for solar sailing in a low earth orbit, Acta Astronaut. 43 (9-10) (1998) 531–541. doi:10.1016/S0094-5765(98)00175-1.

[13] G. Mengali, A. A. Quarta, Near-optimal solar-sail orbit-raising from low earth orbit, J. Spacecr. Rockets 42 (5) (2005) 954–958. doi:10.2514/1.14184.

[14] J. Behrens, Global Trends in Small Satellites, Tech. rep. (2017).
URL https://www.researchgate.net/publication/335429689

[15] R. L. Walterscheid, Solar cycle effects on the upper atmosphere: Implications for satellite drag, J. Spacecr. Rockets 26 (6) (1989) 439–444. doi:10.2514/3.26089.

[16] M. N. Sweeting, Modern small satellites-changing the economics of space, Proceedings of the IEEE 106 (3) (2018) 343–361. doi:10.1109/JPROC.2018.2806218.

[17] O. L. Starinova, M. A. Rozhkov, B. Alipova, I. V. Chernyakina, Modeling the process of optical characteristics variation for a solar sail surface during heliocentric flights, J. Phys. Conf. Ser. 1368 (2) (2019). doi:10.1088/1742-6596/1368/2/022036.

[18] G. Vulpetti, D. Apponi, X. Zeng, C. Circi, Wrinkling analysis of solar-photon sails, Advances in Space Research 67 (9) (2021) 2669–2687, solar Sailing: Concepts, Technology, and Missions II. doi:10.1016/j.asr.2020.07.016.

[19] L. Niccolai, A. A. Quarta, G. Mengali, Trajectory approximation of a solar sail with constant pitch angle and optical degradation, IEEE Transactions on Aerospace and Electronic Systems 58 (4) (2022) 3643–3649. doi:10.1109/TAES.2021.3124867.

[20] L. Boni, M. Bassetto, L. Niccolai, G. Mengali, A. A. Quarta, C. Circi, R. C. Pellegrini, E. Cavallini, Structural response of helianthus solar sail during attitude maneuvers, Aerospace Science and Technology 133 (2023) 108152. doi:10.1016/j.ast.2023.108152.

[21] L. Niccolai, A. Anderlini, G. Mengali, A. A. Quarta, Effects of optical parameter measurement uncertainties and solar irradiance fluctuations on solar sailing, Advances in Space Research 67 (9) (2021) 2784–2794, solar Sailing: Concepts, Technology, and Missions II. doi:10.1016/j.asr.2019.11.037.

[22] A. Caruso, L. Niccolai, A. A. Quarta, G. Mengali, Role of solar irradiance fluctuations on optimal solar sail trajectories, Journal of Spacecraft and Rockets 57 (5) (2020) 1098–1102. doi:10.2514/1.A34709.

[23] J. Mahaney, E. A. Thornton, Self-shadowing effects on the thermal-structural response of orbiting trusses, J. Spacecr. Rockets 24 (4) (1987) 342–348. doi:10.2514/3.25922.

[24] L. Carzana, P. Visser, M. J. Heiligers, Solar-sail control laws for perturbed Earth-bound trajectories, 72nd Int. Astronaut. Conf. (2021).
URL https://pure.tudelft.nl/ws/portalfiles/portal/104389460/IAC_21C135x65648.pdf

[25] C. Bianchi, L. Niccolai, G. Mengali, M. Ceriotti, Preliminary design of a space debris removal mission in LEO using a solar sail, Adv. Sp. Res. 73 (8) (2024) 4254–4268. doi:10.1016/j.asr.2024.01.024.

[26] M. Swartwout, The first one hundred CubeSats : A statistical look, J. Small Satell. 2 (2) (2013) 213–233.
URL https://www.jossonline.com/storage/2014/12/0202-The-First-One-Hundred-Cubesats.pdf

[27] B. Vatankhahghadim, C. J. Damaren, Solar sail deployment dynamics, Advances in Space Research 67 (9) (2021) 2746–2756, solar Sailing: Concepts, Technology, and Missions II. doi:10.1016/j.asr.2020.03.029.

[28] P. Seefeldt, A stowing and deployment strategy for large membrane space systems on the example of gossamer-1, Advances in Space Research 60 (6) (2017) 1345–1362. doi:10.1016/j.asr.2017.06.006.

[29] S. Firuzi, S. Gong, Refractive sail and its applications in solar sailing, Aerosp. Sci. Technol. 77 (2018) 362–372. doi:10.1016/j.ast.2018.03.016.

[30] A. B. Artusio-Glimpse, The Realization and Study of Optical Wings (2016).
URL https://repository.rit.edu/cgi/viewcontent.cgi?article=10377&context=theses

[31] G. A. Swartzlander, Radiation Pressure on a Diffractive Sailcraft, arXiv 34 (6) (2017) 25–30. doi:10.1364/josab.34.000c25.

[32] G. A. Swartzlander, Flying on a rainbow: A solar-driven diffractive sailcraft, JBIS - J. Br. Interplanet. Soc. 71 (4) (2018) 130–132. doi:10.48550/arXiv.1805.05864.

[33] G. A. Swartzlander, Theory of radiation pressure on a diffractive solar sail, J. Opt. Soc. Am. B 39 (9) (2022) 2556. arXiv:2206.10052, doi:10.1364/josab.468588.

[34] M. Bassetto, A. Caruso, A. A. Quarta, G. Mengali, Optimal steering law of refractive sail, Adv. Sp. Res. 67 (9) (2019) 2855–2864. doi:10.1016/j.asr.2019.10.033.

[35] S. Firuzi, Y. Song, S. Gong, Gradient-index solar sail and its optimal orbital control, arXiv (2019) 1–11arXiv:1905.10977, doi:10.1016/j.ast.2021.107103.

[36] Y. Chu, S. Firuzi, S. Gong, Controllable liquid crystal diffractive sail and its potential applications, Acta Astronaut. 182 (February) (2021) 37–45. doi:10.1016/j.actaastro.2021.02.003.

[37] G. A. Swartzlander, T. J. Peterson, A. B. Artusio-Glimpse, A. D. Raisanen, Stable optical lift, Nat. Photonics 5 (1) (2011) 48–51. doi:10.1038/nphoton.2010.266.

[38] P. Zhang, S. Firuzi, C. Yuan, X. Gong, S. Gong, General passive stability criteria for a Sun-pointing attitude using the metasurface sail, Aerosp. Sci. Technol. 122 (2022) 107380. doi:10.1016/j.ast.2022.107380.

[39] S. M. Thompson, C. Tuck, N. Pushparaj, C. Cappelletti, Modelling and numerical optimisation of refractive surface patterns for transmissive solar sails, Acta Astronautica 233 (2025) 99–112. doi:10.1016/j.actaastro.2025.03.039.

[40] S. M. Thompson, C. Cappelletti, N. Pushparaj, Reflective and transmissive

solar sails: Dynamics, flight regimes and applications, Acta Astronautica (2024). doi:10.1016/j.actaastro.2024.04.032.

[41] L. Cormier, T. Yousif, S. Thompson, A. Arcia Gil, N. Pushparaj, P. Blunt, C. Cappelletti, Preliminary Design of a GNSS Interference Mapping CubeSat Mission: JamSail, Aerospace 11 (11) (2024). doi:10.3390/aerospace11110901.

[42] L. Cormier, T. Yousif, S. Thompson, A. Arcia, JamSail: A CubeSat Demonstration Mission for GNSS Interference Mapping and a Refractive Solar Sail, in: 52nd IAF Student Conference, Vol. 233, International Astronautical Federation, 2024, pp. 605–612. doi:10.52202/078379-0053.

[43] D. Robson, Y. Ferreira, H. Cope, P. Da Cás, L. Cormier, G. Lionço, S. Thompson, R. da Silva, M. Ghelfi, A. A. Gil, et al., The WormSail CubeSat: An International Educational Project To Elevate Space Science And Education, in: 72nd International Astronautical Congress, 2021.

[44] M. Alonso, E. Finn, Fundamental University Physics.—Vol. III: Quantum and Statistical Physics (1970). doi:10.1063/1.3035704.

[45] C. McInnes, Solar Sailing: Technology, Dynamics and Mission Applications, Astronomy and Planetary Sciences, Springer, 2004. doi:10.1007/978-1-4471-3992-8.

[46] E. Aspnes, T. D. Milster, K. Visscher, Optical force model based on sequential ray tracing, Appl. Opt. 48 (9) (2009) 1642–1650. doi:10.1364/AO.48.001642.

[47] D. Zola, C. Circi, G. Vulpetti, S. Scaglione, Photon momentum change of quasi-smooth solar sails, J. Opt. Soc. Am. A. Opt. Image Sci. Vis. 35 (8) (2018) 1261–1271. doi:10.1364/JOSAA.35.001261.

[48] R. Georgevic, Mathematical Model of the Solar Radiation Force and

Torques Acting on the Components of a Spacecraft (1971) 33–494.
URL https://ntrs.nasa.gov/citations/19720004068

[49] O. Montenbruck, P. Steigenberger, U. Hugentobler, Enhanced solar radiation pressure modeling for Galileo satellites, J. Geod. 89 (3) (2015) 283–297. doi:10.1007/s00190-014-0774-0.

[50] L. Niccolai, A. A. Quarta, G. Mengali, Analytical solution of the optimal steering law for non-ideal solar sail, Aerospace Science and Technology 62 (2017) 11–18. doi:10.1016/j.ast.2016.11.031.

[51] K. Oguri, G. Lantoine, Indirect trajectory optimization via solar sailing primer vector theory: Minimum solar-angle transfers, Acta Astronautica 211 (2023) 405–415. doi:10.1016/j.actaastro.2023.06.032.

[52] L. Carzana, P. Visser, J. Heiligers, Locally optimal control laws for Earth-bound solar sailing with atmospheric drag, Aerosp. Sci. Technol. 127 (2022). doi:10.1016/j.ast.2022.107666.

[53] Z. J. Zhou, N. N. Yan, A survey of numerical methods for convection-diffusion optimal control problems, J. Numer. Math. 22 (1) (2014) 61–85. doi:10.1515/jnum-2014-0003.

[54] S. Narayanaswamy, C. J. Damaren, Equinoctial lyapunov control law for low-thrust rendezvous, Journal of Guidance, Control, and Dynamics 46 (4) (2023) 781–795. doi:10.2514/1.G006662.

[55] L. Niccolai, A. A. Quarta, G. Mengali, Solar sail heliocentric transfers with a Q-law, Acta Astronautica 188 (2021) 352–361. doi:10.1016/j.actaastro.2021.07.037.

[56] G. Swartzlander, Diffractive solar sails could outperform reflective-metal-coated sails (October 2017) (2018) 35–38.
URL https://www.laserfocusworld.com/optics/article/16548235/space-

and-astronomy-diffractive-solar-sails-could-outperform-reflective-metal-coated-sails

[57] P. R. Srivastava, Y.-J. L. Chu, G. A. Swartzlander, Stable diffractive beam rider, Opt. Lett. 44 (12) (2019) 3082. doi:10.1364/ol.44.003082.

[58] Y.-j. L. Chu, Measurements of Radiation Pressure on Diffractive Films, Rochester Institute of Technology (2021).
URL https://repository.rit.edu/theses/10938/

[59] K. Achouri, O. V. Cespedes, C. Caloz, Solar 'Meta-Sails' for Agile Optical Force Control, IEEE Trans. Antennas Propag. 67 (11) (2019) 6924–6934. doi:10.1109/TAP.2019.2925279.

[60] A. Ashkin, Forces of a single-beam gradient laser trap on a dielectric sphere in the ray optics regime, Biophys. J. 61 (2) (1992) 569–582. doi:10.1016/S0006-3495(92)81860-X.

[61] K. E. Drexler, Design of a high performance solar sail system., Ph.D. thesis, Massachusetts Institute of Technology (1979).

[62] G. L. Matloff, R. Y. Kezerashvili, Interstellar solar sailing: A figure of merit for monolayer sail, JBIS - J. Br. Interplanet. Soc. 61 (8) (2008) 330–333.

[63] S. Firuzi, S. Gong, All-dielectric tailored-index index-matched arbitrary waveguides based on transformation optics, arXiv (2019). arXiv:1903.04398.

[64] R. Wilken, A. Holländer, J. Behnisch, Vacuum Ultraviolet Photolysis of Polyethylene, Polypropylene, and Polystyrene, Plasmas Polym. 7 (2) (2002) 185–205. doi:10.1023/A:1016247524912.

[65] C. P. Ennis, R. I. Kaiser, Mechanistical studies on the electron-induced degradation of polymers: Polyethylene, polytetrafluoroethylene, and

polystyrene, Phys. Chem. Chem. Phys. 12 (45) (2010) 14884–14901. doi:10.1039/c0cp00493f.

[66] D. W. Van Krevelen, Properties of Polymers: Their Correlation with Chemical Structure; their Numerical Estimation and Prediction from Additive Group Contributions: Fourth Edition, 2009. doi:10.1016/C2009-0-05459-2.

[67] Y. Liu, J. K. Boyles, J. Genzer, M. D. Dickey, Self-folding of polymer sheets using local light absorption, Soft Matter 8 (2012) 1764–1769. doi:10.1039/C1SM06564E.

[68] G. S. Martins, I. M. Pereira, R. L. Oréfice, Toughening brittle polymers with shape memory polymers, Polymer 135 (2018) 30–38. doi:10.1016/j.polymer.2017.12.017.

[69] D. C. Ullery, S. Soleymani, A. Heaton, J. Orphee, L. Johnson, R. Sood, P. Kung, S. M. Kim, Strong Solar Radiation Forces from Anomalously Reflecting Metasurfaces for Solar Sail Attitude Control, Sci. Rep. 8 (1) (2018) 1–11. doi:10.1038/s41598-018-28133-2.

[70] V. P. M., High Performance Polymers and Their Nanocomposites., 1st Edition, John Wiley & Sons, Incorporated, 2019.

[71] K. Pang, R. Kotek, A. Tonelli, Review of conventional and novel polymerization processes for polyesters, Progress in Polymer Science 31 (11) (2006) 1009–1037. doi:https://doi.org/10.1016/j.progpolymsci.2006.08.008.

[72] Y. Ma, U. Agarwal, Solvent assisted post-polymerization of PET, Polymer (Guilford) 46 (15) (2005) 5447–5455.

[73] J. Hanisch, T. Wahl, C. D. Wessendorf, E. Ahlswede, Efficient polymer tandem modules and solar cells by doctor blading, J. Mater. Chem. A 4

(2016) 4771–4775. doi:10.1039/C5TA10257J.
URL http://dx.doi.org/10.1039/C5TA10257J

[74] L. Johnson, R. Young, E. Montgomery, D. Alhorn, Status of solar sail technology within NASA, Adv. Sp. Res. 48 (11) (2011) 1687–1694. doi:10.1016/j.asr.2010.12.011.

[75] M. T. DeMeuse, 5.2 polyethylene terephthalate (pet) (2011).
URL https://app.knovel.com/hotlink/khtml/id:kt0094H484/biaxial-stretching-film/polyethylene-terephthalate

[76] R. Melentiev, A. Yudhanto, R. Tao, T. Vuchkov, G. Lubineau, Metallization of polymers and composites: State-of-the-art approaches, Materials & Design 221 (2022) 110958. doi:https://doi.org/10.1016/j.matdes.2022.110958.

[77] M.-W. Wang, C.-C. Tseng, Analysis and fabrication of a prism film with roll-to-roll fabrication process, Opt. Express 17 (6) (2009) 4718. doi:10.1364/oe.17.004718.

[78] Y. Jiang, B. Luo, X. Cheng, Enhanced thermal stability of thermoplastic polymer nanostructures for nanoimprint lithography, Materials (Basel). 12 (3) (2019). doi:10.3390/ma12030545.

[79] D. Wan, H. L. Chen, Y. T. Lai, C. C. Yu, K. F. Lin, Use of reversal nanoimprinting of nanoparticles to prepare flexible waveguide sensors exhibiting enhanced scattering of the surface plasmon resonance, Adv. Funct. Mater. 20 (11) (2010) 1742–1749. doi:10.1002/adfm.201000057.

[80] B. Cui, Y. Cortot, T. Veres, Polyimide nanostructures fabricated by nanoimprint lithography and its applications, Microelectron. Eng. 83 (4-9 SPEC. ISS.) (2006) 906–909. doi:10.1016/j.mee.2006.01.014.

[81] J. Kim, H. Yoon, Transfer tiling of Nanostructures for large-area fabrication, Micromachines 9 (11) (2018). doi:10.3390/mi9110569.

[82] A. R. Sobey, T. R. Lockett, Design and Development of NEA Scout Solar Sail Deployer Mechanism, Proc. 43rd Aerosp. Mech. Symp. (September 2018) (2016) 315–328.
URL https://ntrs.nasa.gov/citations/20160008143

[83] B. Putz, S. Wurster, T. Edwards, B. Völker, G. Milassin, D. Többens, C. Semprimoschnig, M. Cordill, Mechanical and optical degradation of flexible optical solar reflectors during simulated low earth orbit thermal cycling, Acta Astronautica 175 (2020) 277–289. doi:10.1016/j.actaastro.2020.05.032.

[84] G. F. Abdelal, N. Abuelfoutouh, A. Hamdy, A. Atef, Thermal fatigue analysis of small-satellite structure, International Journal of Mechanics and Materials in Design 3 (2006) 145–159. doi:10.1007/s10999-007-9019-1.

[85] W. Saylor, Space Radiation Guide, Vol. 62, 1962.
URL https://www.osti.gov/biblio/4728733

[86] N. Meyer-Vernet, Basics of the Solar Wind, Cambridge University Press, 2007. doi:10.1017/CBO9780511535765.

[87] R. Y. Kezerashvili, G. L. Matloff, Solar radiation and the beryllium hollow-body sail-1 - the ionization and disintegration effects, Journal of the British Interplanetary Society 60 (2007) 169–179.
URL https://ui.adsabs.harvard.edu/abs/2007JBIS...60..169K

[88] R. Y. Kezerashvili, Solar sail: Materials and space environmental effects, arXiv.org (Jul 28 2013). doi:10.1007/978-3-642-34907-2_36.

[89] C. Giori, T. Yamauchi, Space Radiation Resistant Transparent Polymeric Materials, NASA - Natl. Aeronaut. Sp. Adm. (1977).
URL https://ntrs.nasa.gov/citations/19780009276

[90] A. El-Hameed, Radiation effects on composite materials used in space systems: a review, NRIAG Journal of Astronomy and Geophysics 11 (2022) 313–324. doi:10.1080/20909977.2022.2079902.

[91] M. Cutroneo, L. Torrisi, L. Calcagno, A. Torrisi, Characterization of thin films for TNSA laser irradiation, J. Phys. Conf. Ser. 508 (1) (2014). doi:10.1088/1742-6596/508/1/012012.

[92] V. Tarantini, G. Bonin, R. Zee, De-orbit analysis of a tumbling sailcraft, in: Proceedings of the Small Satellites Systems and Services Symposium, Porto Perto, 2014.

[93] Z. Serfontein, J. Kingston, S. Hobbs, I. E. Holbrough, J. C. Beck, Drag augmentation systems for space debris mitigation, Acta Astronautica 188 (2021) 278–288. doi:10.1016/j.actaastro.2021.05.038.

[94] W. Wiquist, FCC Adopts New '5-Year Rule' for Deorbiting Satellites (2022).
URL https://www.fcc.gov/document/fcc-adopts-new-5-year-rule-deorbiting-satellites

[95] J. Hsu, Fishing for space junk, IEEE Spectr. 55 (6) (2018) 7–9. doi:10.1109/MSPEC.2018.8362211.

[96] C. R. Phipps, LADROIT - A spaceborne ultraviolet laser system for space debris clearing, Acta Astronaut. 104 (1) (2014) 243–255. doi:10.1016/j.actaastro.2014.08.007.

[97] J. A. Storch, Aerodynamic Disturbances on Rapidly Rotating Spacecraft in Free-Molecular Flow, Eng. Constr. Oper. Challenging Environ. Earth Sp. 2004 Proc. Ninth Bienn. ASCE Aerosp. Div. Int. Conf. 2003 (3397) (2004) 429–436. doi:10.1061/40722(153)60.

[98] L. Casalino, A. Forestieri, Approximate optimal LEO transfers with J2

perturbation and dragsail, Acta Astronaut. 192 (2022) 379–389. doi:10.1016/j.actaastro.2021.12.006.

[99] R. Sumanth, Computation of eclipse time for low-earth orbiting small satellites, International Journal of Aviation, Aeronautics, and Aerospace (01 2019). doi:10.15394/ijaaa.2019.1412.

[100] G. J. Ward, F. M. Rubinstein, R. D. Clear, A ray tracing solution for diffuse interreflection, Proc. 15th Annu. Conf. Comput. Graph. Interact. Tech. SIGGRAPH 1988 22 (4) (1988) 85–92. doi:10.1145/54852.378490.

[101] R-800 Hall Thruster Datasheet, URL https://satsearch.co/products/rafael-r-800-hall-thruster (Accessed on 01/18/2023).

[102] IFM Nano Thruster Se / Cots + Datasheet (Accessed on 01/18/2023).
URL https://enpulsion.com/wp-content/uploads/ENP2019-086.B-IFM-Nano-Thruster-COTS-Product-Overview-1.pdf

[103] 100 mN HPGP Thruster Datasheet (Accessed on 01/18/2023).
URL https://www.satcatalog.com/component/100mn-hpgp/

[104] Pale blue water ion thruster 1u datasheet, https://satcatalog.s3.amazonaws.com/components/1220/SatCatalog_-_Pale_Blue_-_Water_Ion_Thruster_-_Datasheet.pdf?lastmod=20230531084516 (Accessed on 01/18/2023).

[105] T. J. Peterson, Ray-Traced Simulation of Radiation Pressure for Optical Lift (2011).
URL http://scholarworks.rit.edu/theses/266/

[106] N. Sultanova, S. Kasarova, I. Nikolov, Dispersion properties of optical polymers, Acta Phys. Pol. A 116 (4) (2009) 585–587. doi:10.12693/APhysPolA.116.585.

[107] N. G. Sultanova, S. N. Kasarova, I. D. Nikolov, Characterization of optical properties of optical polymers, Opt. Quantum Electron. 45 (3) (2013) 221–232. doi:10.1007/s11082-012-9616-6.

[108] Y.-T. Tu, et al., Ray Optics Simulation (2016). doi:10.5281/zenodo.6386611.
URL https://phydemo.app/ray-optics/

[109] T. N. Huynh, D. T. Do, J. Lee, Q-learning-based parameter control in differential evolution for structural optimization, Applied Soft Computing 107 (2021) 107464.

[110] C. J. Watkins, P. Dayan, Q-learning, Machine learning 8 (1992) 279–292.

[111] P. Bourke, Polygons and meshes (1997).
URL https://paulbourke.net/geometry/polygonmesh/

[112] J. Nocedal, S. J. Wright, Numerical optimization, Springer, 1999.

[113] S. Jaimungal, Reinforcement learning and stochastic optimisation, Finance and Stochastics 26 (1) (2022) 103–129.

[114] P.-I. Schneider, X. Garcia Santiago, V. Soltwisch, M. Hammerschmidt, S. Burger, C. Rockstuhl, Benchmarking five global optimization approaches for nano-optical shape optimization and parameter reconstruction, ACS Photonics 6 (11) (2019) 2726–2733.

[115] M. M. Elsawy, S. Lanteri, R. Duvigneau, J. A. Fan, P. Genevet, Numerical optimization methods for metasurfaces, Laser & Photonics Reviews 14 (10) (2020) 1900445.

[116] M. Polyanskiy, Optical Constants of Plastics, PS (2024) 9–10.
URL https://refractiveindex.info/?shelf=3d{&}book=plastics{&}page=ps

[117] Filmetrics, Refractive Index of PET, Estar, Melinex, Mylar 9300 (2024) 1–2.

URL https://www.filmetrics.com/refractive-index-database/PET/Estar-Melinex-Mylar

[118] M. K. Kwak, C. W. Park, Continuous fabrication process for double sided micro/nanopatterns, Mater. Res. Innov. 18 (2014) S5–21–S5–24. doi:10.1179/1432891714Z.000000000903.

[119] H. Lan, Y. Ding, Nanoimprint lithography, in: M. Wang (Ed.), Lithography, IntechOpen, Rijeka, 2010, Ch. 23. doi:10.5772/8189.

[120] R. Stengl, T. Tan, U. Gösele, A model for the silicon wafer bonding process, Jpn. J. Appl. Phys. 28 (10 R) (1989) 1735–1741. doi:10.1143/JJAP.28.1735.

[121] H. Sui, N. Huda, Z. Shen, J. Z. Wen, Al–NiO energetic composites as heat source for joining silicon wafer, J. Mater. Process. Technol. 279 (November 2019) (2020) 116572. doi:10.1016/j.jmatprotec.2019.116572.

[122] N. Abbasnezhad, A Joining Strategy for the Stereolithography-Printed Parts: Thermal and Mechanical Characterization, J. Mater. Eng. Perform. (Ref 19) (2024). doi:10.1007/s11665-024-09816-6.

[123] Stensborg, Desktop Nanoimprinter 4 (1) (2014) 2014.
URL https://stensborg.com/desktopr2pnanoimprinter/

[124] D. F. Buscher, Manufacturing metre-sized gratings with nanometre precision, lookup.phy (2024).
URL https://people.phy.cam.ac.uk/db106/grating-manufacturing/

[125] Y. W. Adugna, A. D. Akessa, H. G. Lemu, Overview study on challenges of additive manufacturing for a healthcare application, IOP Conf. Ser. Mater. Sci. Eng. 1201 (1) (2021) 012041. doi:10.1088/1757-899x/1201/1/012041.

[126] A. Zolfaghari, T. Chen, A. Y. Yi, Additive manufacturing of precision optics at micro and nanoscale, Int. J. Extrem. Manuf. 1 (1) (2019). doi:10.1088/2631-7990/ab0fa5.

[127] F. Calignano, D. Manfredi, E. P. Ambrosio, S. Biamino, M. Lombardi, E. Atzeni, A. Salmi, P. Minetola, L. Iuliano, P. Fino, Overview on additive manufacturing technologies, Proc. IEEE 105 (4) (2017) 593–612. doi:10.1109/JPROC.2016.2625098.

[128] T. Han, S. Kundu, A. Nag, Y. Xu, 3D printed sensors for biomedical applications: A review, Sensors (Switzerland) 19 (7) (2019). doi:10.3390/s19071706.

[129] G. Berglund, A. Wisniowiecki, J. Gawedzinski, B. Applegate, T. S. Tkaczyk, Additive manufacturing for the development of optical/photonic systems and components, Optica 9 (6) (2022) 623. doi:10.1364/optica.451642.

[130] L. Rodríguez-Parada, P. F. Mayuet, A. J. Gámez, Industrial product design: Study of FDM technology for the manufacture of thermoformed prototypes, Procedia Manuf. 41 (2019) 587–593. doi:10.1016/j.promfg.2019.09.046.

[131] S. Zhang, R. K. Arya, S. Pandey, Y. Vardaxoglou, W. Whittow, R. Mittra, 3d-printed planar graded index lenses, IET Microwaves, Antennas & Propagation 10 (13) (2016) 1411–1419. arXiv:https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-map.2016.0013, doi:https://doi.org/10.1049/iet-map.2016.0013.

[132] A. D. Squires, E. Constable, R. A. Lewis, 3D Printed Terahertz Diffraction Gratings And Lenses, J. Infrared, Millimeter, Terahertz Waves 36 (1) (2015) 72–80. doi:10.1007/s10762-014-0122-8.

[133] M. Rinaldi, T. Ghidini, F. Cecchini, A. Brandao, F. Nanni, Additive layer manufacturing of poly (ether ether ketone) via FDM, Compos. Part B

Eng. 145 (December 2017) (2018) 162–172. doi:10.1016/j.compositesb.2018.03.029.

[134] F. Abdullah, K. I. Okuyama, I. Fajardo, N. Urakami, In situ measurement of carbon fibre/polyether ether ketone thermal expansion in low earth orbit, Aerospace 7 (4) (2020). doi:10.3390/aerospace7040035.

[135] N. W. S. Pinargote, A. Smirnov, N. Peretyagin, A. Seleznev, P. Peretyagin, Direct ink writing technology (3d printing) of graphene-based ceramic nanocomposites: A review, Nanomaterials 10 (7) (2020) 1–48. doi:10.3390/nano10071300.

[136] S. C. Altıparmak, V. A. Yardley, Z. Shi, J. Lin, Extrusion-based additive manufacturing technologies: State of the art and future perspectives, J. Manuf. Process. 83 (July) (2022) 607–636. doi:10.1016/j.jmapro.2022.09.032.

[137] J. Yang, B. Zhou, D. Han, N. Cui, B. Li, J. Shen, Z. Zhang, A. Du, High-Precision Three-Dimensional Printing in a Flexible, Low-Cost and Versatile Way: A Review, ES Mater. Manuf. 15 (2022) 1–13. doi:10.30919/esmm5f526.

[138] N. Tepylo, X. Huang, P. C. Patnaik, Laser-Based Additive Manufacturing Technologies for Aerospace Applications, Adv. Eng. Mater. 21 (11) (2019) 1–35. doi:10.1002/adem.201900617.

[139] S. M. Hur, K. H. Choi, S. H. Lee, P. K. Chang, Determination of fabricating orientation and packing in SLS process, J. Mater. Process. Technol. 112 (2-3) (2001) 236–243. doi:10.1016/S0924-0136(01)00581-7.

[140] Y. Song, Y. Ghafari, A. Asefnejad, D. Toghraie, An overview of selective laser sintering 3D printing technology for biomedical and sports device applications: Processes, materials, and applications, Opt. Laser Technol. 171 (April 2023) (2024) 110459. doi:10.1016/j.optlastec.2023.110459.

[141] K. Zhang, G. Fu, P. Zhang, Z. Ma, Z. Mao, D. Z. Zhang, Study on the geometric design of supports for overhanging structures fabricated by selective laser melting, Materials (Basel). 12 (1) (2018). doi:10.3390/ma12010027.

[142] F. Calignano, Design optimization of supports for overhanging structures in aluminum and titanium alloys by selective laser melting, Mater. Des. 64 (2014) 203–213. doi:10.1016/j.matdes.2014.07.043.

[143] M. Sweeney, M. Acreman, T. Vettese, R. Myatt, M. Thompson, Application and testing of additive manufacturing for mirrors and precision structures, Mater. Technol. Appl. to Opt. Struct. Components, Sub-Systems II 9574 (Mmc) (2015) 957406. doi:10.1117/12.2189202.

[144] T. Laumer, T. Stichel, K. Nagulin, M. Schmidt, Optical analysis of polymer powder materials for Selective Laser Sintering, Polym. Test. 56 (2016) 207–213. doi:10.1016/j.polymertesting.2016.10.010.

[145] M. Lupo, S. Z. Ajabshir, D. Sofia, D. Barletta, M. Poletto, Experimental metrics of the powder layer quality in the selective laser sintering process, Powder Technol. 419 (February) (2023) 118346. doi:10.1016/j.powtec.2023.118346.

[146] M. Schmidt, D. Pohle, T. Rechtenwald, Selective laser sintering of PEEK, CIRP Ann. - Manuf. Technol. 56 (1) (2007) 205–208. doi:10.1016/j.cirp.2007.05.097.

[147] Y. Lakhdar, C. Tuck, J. Binner, A. Terry, R. Goodridge, Additive manufacturing of advanced ceramic materials, Prog. Mater. Sci. 116 (September 2020) (2021) 100736. doi:10.1016/j.pmatsci.2020.100736.

[148] R. Kumar, M. Kumar, J. S. Chohan, Material-specific properties and applications of additive manufacturing techniques: a comprehensive review, Bull. Mater. Sci. 44 (3) (2021). doi:10.1007/s12034-021-02364-y.

[149] M. Jiménez, L. Romero, I. A. Domínguez, M. D. M. Espinosa, M. Domínguez, Additive Manufacturing Technologies: An Overview about 3D Printing Methods and Future Prospects, Complexity 2019 (2019). doi:10.1155/2019/9656938.

[150] T. Kawase, T. Shimoda, C. Newsome, H. Sirringhaus, R. H. Friend, Inkjet printing of polymer thin film transistors, Thin Solid Films 438-439 (03) (2003) 279–287. doi:10.1016/S0040-6090(03)00801-0.

[151] L. Wu, Z. Dong, F. Li, H. Zhou, Y. Song, Emerging Progress of Inkjet Technology in Printing Optical Materials, Adv. Opt. Mater. 4 (12) (2016) 1915–1932. doi:10.1002/adom.201600466.

[152] Q. Jin, Q. Zhang, C. Rainer, H. Hu, J. Chen, T. Gehring, J. Dycke, R. Singh, U. W. Paetzold, G. Hernández-Sosa, R. Kling, U. Lemmer, Inkjet-printed optical interference filters, Nat. Commun. 15 (1) (2024) 1–7. doi:10.1038/s41467-024-47086-x.

[153] M. Thymianidis, C. Achillas, D. Tzetzis, Modern Additive Manufacturing Technologies : An Up- to-Date Synthesis and Impact on Supply Chain Design, Int. Conf. supply Chain. (2013) 21.

[154] S. H. Ahn, L. J. Guo, Large-area roll-to-roll and roll-to-plate Nanoimprint Lithography: A step toward high-throughput application of continuous nanoimprinting, ACS Nano 3 (8) (2009) 2304–2310. doi:10.1021/nn9003633.

[155] C. A. Costa, P. R. Linzmaier, F. M. Pasquali, Rapid prototyping material degradation: A study of mechanical properties, IFAC Proc. Vol. 6 (PART 1) (2013) 350–355. doi:10.3182/20130911-3-BR-3021.00118.

[156] Boston Micro Fabrication, Ultra-High Resolution 3D Printers (2022). URL https://bmf3d.com/ultra-high-resolution-3d-printers/

[157] F. Cantoni, D. Maher, E. Bosler, S. Kühne, L. Barbe, D. Oberschmidt, C. Marquette, R. Taboryski, M. Tenje, A.-I. Bunea, Round-robin testing of commercial two-photon polymerization 3d printers, Additive Manufacturing 76 (2023) 103761. doi:https://doi.org/10.1016/j.addma.2023.103761.

[158] A. Jaiswal, C. K. Rastogi, S. Rani, G. P. Singh, S. Saxena, S. Shukla, Two decades of two-photon lithography: Materials science perspective for additive manufacturing of 2D/3D nano-microstructures, iScience 26 (4) (2023) 106374. doi:10.1016/j.isci.2023.106374.

[159] S. C. Ligon, R. Liska, J. Stampfl, M. Gurr, R. Mülhaupt, Polymers for 3D Printing and Customized Additive Manufacturing, Chem. Rev. 117 (15) (2017) 10212–10290. doi:10.1021/acs.chemrev.7b00074.

[160] A. K. Nguyen, R. J. Narayan, Two-photon polymerization for biological applications, Mater. Today 20 (6) (2017) 314–322. doi:10.1016/j.mattod.2017.06.004.

[161] UpNano, NanoOne 1000 (2024) 1–7.
URL https://www.upnano.com/nanoone-1000/

[162] M. Lemire, L. Boni, B. Furnish, The straight story on linear actuators, Motion Syst. Des. 49 (11) (2007) 24–31.
URL https://www.santecindia.com/pdfs/white-papers3.pdf

[163] J. Shah, B. Snider, T. Clarke, S. Kozutsky, M. Lacki, A. Hosseini, Large-scale 3D printers for additive manufacturing: design considerations and challenges, Int. J. Adv. Manuf. Technol. 104 (9-12) (2019) 3679–3693. doi:10.1007/s00170-019-04074-6.

[164] A. Kampker, J. Triebs, S. Kawollek, P. Ayvaz, S. Hohenstein, Review on machine designs of material extrusion based additive manufacturing (AM)

systems - Status-Quo and potential analysis for future AM systems, Procedia CIRP 81 (March) (2019) 815–819. doi:10.1016/j.procir.2019.03.205.

[165] B. Derby, Additive Manufacture of Ceramics Components by Inkjet Printing, Engineering 1 (1) (2015) 113–123. doi:10.15302/J-ENG-2015014.

[166] A. Pilipović, G. Baršić, M. Katić, M. R. Havstad, Repeatability and reproducibility assessment of a polyjet technology using x-ray computed tomography, Appl. Sci. 10 (20) (2020) 1–14. doi:10.3390/app10207040.

[167] B. N. Panda, K. Shankhwar, A. Garg, Z. Jian, Performance evaluation of warping characteristic of fused deposition modelling process, Int. J. Adv. Manuf. Technol. 88 (5-8) (2017) 1799–1811. doi:10.1007/s00170-016-8914-8.

[168] L. C. B. Carolo, R. E. Cooper O., A review on the influence of process variables on the surface roughness of Ti-6Al-4V by electron beam powder bed fusion, Addit. Manuf. 59 (PA) (2022) 103103. doi:10.1016/j.addma.2022.103103.

[169] P. Delfs, M. T̈ows, H. J. Schmid, Optimized build orientation of additive manufactured parts for improved surface quality and build time, Addit. Manuf. 12 (2016) 314–320. doi:10.1016/j.addma.2016.06.003.

[170] BMF, HTL Resin.
URL https://bmf3d.com/micro-3d-printing-materials/

[171] P. F. O'Neill, N. Kent, D. Brabazon, Mitigation and control of the over-curing effect in mask projection micro-stereolithography, AIP Conf. Proc. 1896 (2017). doi:10.1063/1.5008249.

[172] N. L. Chiromawa, K. Ibrahim, Applications of Electron Beam Lithography (EBL) in Optoelectronics Device Fabrication., AASCIT J. Physics. 4 (May) (2018) 53–58. arXiv:40192771.

[173] Y. Chen, Applications of nanoimprint lithography/hot embossing: a review, Applied Physics A: Materials Science & Processing 121 (2) (2015) 451–465. doi:10.1007/s00339-015-9071-x.

[174] M. Koç, T. Özel, Fundamentals of Micro-Manufacturing, Micro-Manufacturing Des. Manuf. Micro-Products (May) (2011) 1–23. doi:10.1002/9781118010570.ch1.

[175] B. J. Wiley, D. Qin, Y. Xia, Nanofabrication at high throughput and low cost, ACS Nano 4 (7) (2010) 3554–3559. doi:10.1021/nn101472p.

[176] M. Choi, J. W. Leem, J. S. Yu, Antireflective gradient-refractive-index material-distributed microstructures with high haze and superhydrophilicity for silicon-based optoelectronic applications, RSC Adv. 5 (32) (2015) 25616–25624. doi:10.1039/c4ra15686b.

[177] NanoBeam Ltd, Advanced Electron Beam Lithography, NanoBeam nB5 (2011).
URL https://www.tesscorn-nanoscience.com/wp-content/uploads/2017/02/nB5.pdf

[178] J. Bailey, Thin Film Polymer Photonics : Spin Cast Distributed Bragg Reflectors Thesis submitted to The University of Nottingham for the degree of Doctor of Philosophy Abstract (July) (2014).

[179] A. Berni, M. Mennig, H. Schmidt, 2.2. 8 Doctor Blade, Sol-Gel Technol. Glas. ... (2004) 8–11.
URL http://books.google.com/books?hl=en{&}lr={&}id=ONBNBdPyuzAC{&}oi=fnd{&}pg=PA89{&}dq=2.2.8+doctor+blade{&}ots=UjVyk837zV{&}sig=kAURdhSUolRt{_}nbcgF-lby89Jik

[180] E. Bormashenko, A. Musin, Y. Bormashenko, G. Whyman, R. Pogreb, O. Gendelman, Formation of films on water droplets floating on a polymer

solution surface, Macromol. Chem. Phys. 208 (7) (2007) 702–709. doi:10.1002/macp.200600485.

[181] Z. Yin, L. Sun, H. Zou, Numerical analysis on PET demolding stage in thermal nanoimprinting lithography, Microsyst. Technol. 23 (4) (2017) 899–905. doi:10.1007/s00542-015-2761-7.

[182] B. Z. Karim, Y. E. Systems, Low-temperature polyimide processing for next-gen backend applications (2021) 1–5.

[183] H. Microfluidics/Microfabrication Facility, PROCEDURE FOR SILANIZATION OF SU-8/SILICON MASTER Microfabrication Core Facility, Harvard Medical School (2015) 3–4.
URL https://hms.harvard.edu/sites/default/files/Departments/Microfluidics and Microfabrication Facility/files/Silanization of Photoresist Master Protocol.pdf

[184] S. Schoell, A. Oliveros, M. Steenackers, S. Saddow, I. Sharp, Chapter 3 - multifunctional sic surfaces: From passivation to biofunctionalization, in: S. E. Saddow (Ed.), Silicon Carbide Biotechnology, Elsevier, Oxford, 2012, pp. 63–117. doi:https://doi.org/10.1016/B978-0-12-385906-8.00003-9.

[185] K. Subramani, Chapter 11 - fabrication of hydrogel micropatterns by soft photolithography, in: W. Ahmed, M. J. Jackson (Eds.), Emerging Nanotechnologies for Manufacturing (Second Edition), second edition Edition, Micro and Nano Technologies, William Andrew Publishing, Boston, 2015, pp. 279–293. doi:https://doi.org/10.1016/B978-0-323-28990-0.00011-7.

[186] Stensborg, X30 Technical Data Sheet.
URL https://stensborg.com//wp-content/uploads/2023/01/Tech-specs-for-X30-resin-11072022.pdf

[187] Goodfellow Advanced Materials (Accessed on 06/09/2024).
URL https://www.goodfellow.com/uk/

[188] R. Ye, How Much Does CNC Milling Cost: Factors and Tips to Consider (2024) 1–13.
URL https://www.3erp.com/blog/cnc-milling-cost/

[189] All3DP, The World's Most Expensive 3D Printers of 2024 (2024) 1–22.
URL https://all3dp.com/1/biggest-large-3d-printer-world-most-expensive/

[190] T. Magazine, Interview: Boston Micro Fabrication announces global launch of PμSL 3D printing technology - TCT Magazine (2024) 1–7.
URL https://www.tctmagazine.com/additive-manufacturing-3d-printing-news/boston-micro-fabrication-pusl-3d-printing-technology/

[191] Malcolm Nicholls Limited, 3DP Online Quotation (2024) 1–7.
URL https://mnl.co.uk/3d-printing-quote/

[192] M. Jalali, J. J. Marti, A. L. Kirchhoff, F. Lawrenz, S. A. Campbell, A low-cost hands-on laboratory to introduce lithography concepts, IEEE Trans. Educ. 55 (4) (2012) 517–524. doi:10.1109/TE.2012.2192442.

[193] University of Nottingham, Nanofabrication Nottingham Facilities, Nanofabrication Nottingham Services and price list 770.
URL https://www.nottingham.ac.uk/nmrc/documents/services-and-charges-external.pdf

[194] CubeSatShop, IFM Nano Thruster for CubeSats (2024) 1–4.
URL https://www.cubesatshop.com/product/ifm-nano-thruster/

[195] F. Leverone, A. Cervone, E. Gill, Cost analysis of solar thermal propulsion systems for microsatellite applications, Acta Astronaut. 155 (December) (2019) 90–110. doi:10.1016/j.actaastro.2018.11.025.

[196] National Research Council and Division on Engineering and Physical Sciences and Aeronautics and Space Engineering Board and Committee on Thermionic Research, Thermionics Quo Vadis?: An Assessment of the DTRA's Advanced Thermionics Research and Development Program, The National Academies Press, Washington, DC, 2001, see Appendix C: Electric Propulsion Considerations.

[197] S. A. Zirbel, R. J. Lang, M. W. Thomson, D. A. Sigel, P. E. Walkemeyer, B. P. Trease, S. P. Magleby, L. L. Howell, Accommodating thickness in origami-based deployable arrays, J. Mech. Des. Trans. ASME 135 (11) (2013) 1–11. doi:10.1115/1.4025372.

[198] R. J. Lang, S. Magleby, L. Howell, Single degree-of-freedom rigidly foldable cut origami flashers, J. Mech. Robot. 8 (3) (2016) 1–15. doi:10.1115/1.4032102.

[199] K. Ikeya, H. Sakamoto, H. Nakanishi, H. Furuya, T. Tomura, R. Ide, R. Iijima, Y. Iwasaki, K. Ohno, K. Omoto, T. Furuya, T. Hayashi, M. Kato, S. Koide, M. Kurosaki, Y. Nakatsuka, S. Okuyama, R. Kashiyama, J. Nakamura, W. Nio, T. Tsunemitsu, Y. Yamazaki, K. Taga, B. Hohmann, T. Amamoto, T. Chubachi, S. Tamura, H. Okada, A. Watanabe, N. Kawabata, T. Hori, H. Ito, T. Kuratomi, Y. Shimoda, N. Hidaka, K. Watanabe, A. Torisaka, M. Yamazaki, Significance of 3U CubeSat OrigamiSat-1 for space demonstration of multifunctional deployable membrane, Acta Astronaut. 173 (2020) 363–377. doi:10.1016/j.actaastro.2020.04.016.

[200] R. J. Lang, Tessellatica.
URL https://langorigami.com/article/tessellatica/

[201] California Polytechnic State University, Cubesat Design Specification Rev 14.1, CubeSat Program, Cal Poly SLO 8651 (February) (2022) 34.
URL https://www.cubesat.org/cubesatinfo

[202] E. Agu, F. S. Hill, A simple method for ray tracing diffraction, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 2669 (2003) 336–345. doi:10.1007/3-540-44842-x_35.