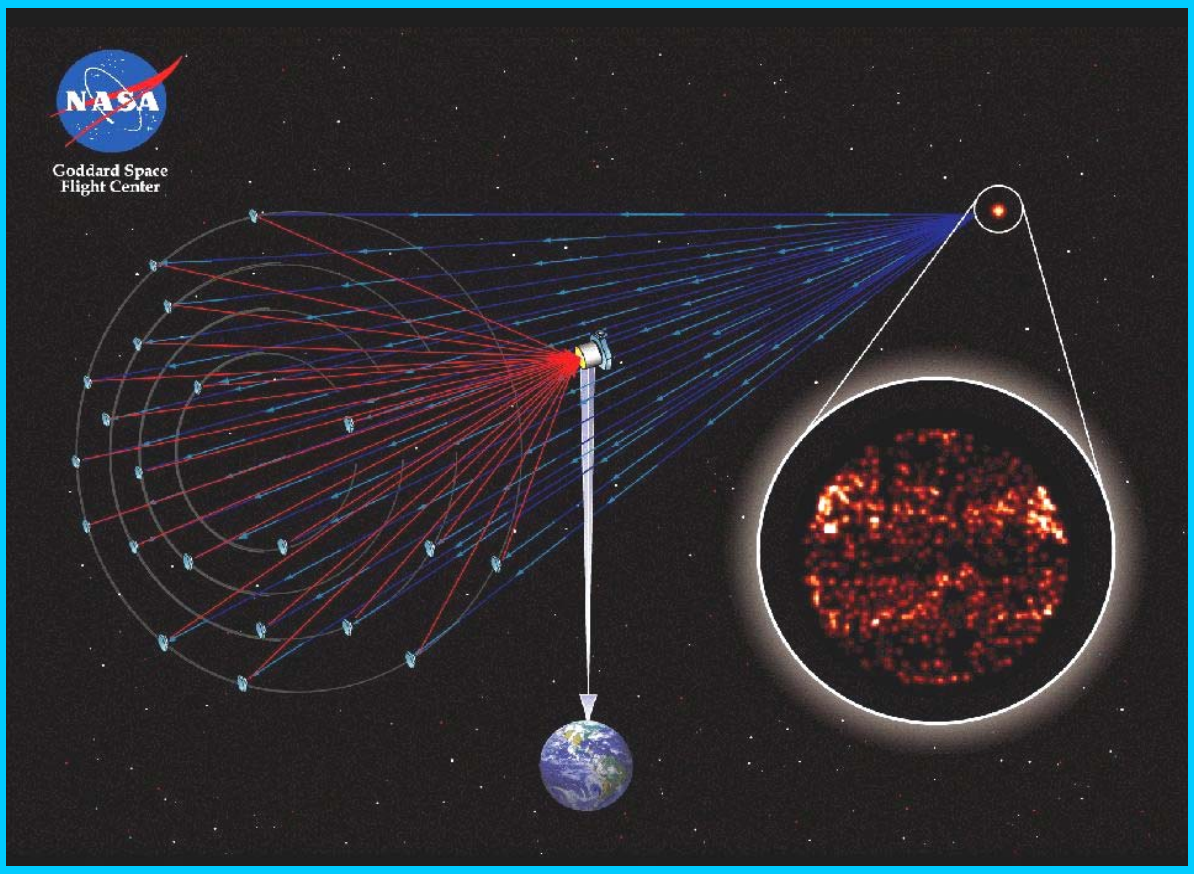
Vision Mission Study Report (15 September 2005)

SI – The Stellar Imager

A UV/Optical deep-space telescope to image stars and observe the Universe with 0.1 milli-arcsec angular resolution



Its mission:

To enable an understanding of solar/stellar magnetic activity and its impact on the:

- origin and continued existence of life in the Universe
- structure and evolution of stars
- habitability of planets

and to study magnetic processes and their roles in the origin and evolution of structure and the transport of matter throughout the Universe.

http://hires.gsfc.nasa.gov/~si/

Front Cover Illustration

The illustration on the front cover shows an artist's concept of the baseline SI design, a Fizeau Interferometer with 20-30 one-meter primary mirrors, which are mounted on formation-flying "mirrorsats" distributed over a parabolic virtual surface whose diameter can be varied from 100 m up to as much as 1000 m, depending on the angular size of the target to be observed. The individual mirrors are fabricated as ultra-smooth, UV-quality flats and are actuated to produce the extremely gentle curvature needed to focus light on the beam-combining hub that is located from 1 – 10 km distant. The focal length scales linearly with the diameter of the primary array: a 100 m diameter array corresponds to a focal length of 1 km and a 1000 m array with a focal length of 10 km.

This Stellar Imager Vision Mission Report presents the results of a study carried out by a broad collaboration led by the Goddard Space Flight Center as shown on the next page. The principal authors of this Report and their primary areas of expertise include:

Kenneth G. Carpenter (GSFC) – Principal Investigator and Lead Author
Carolus J. Schrijver (LMATC) – Solar/Stellar Magnetic Activity Science Lead
Margarita Karovska (SAO) – "Universe" Science Lead
Ron Allen (STScI) – Phase Closure, Interferometer Design/Performance/Simulator
Jim Breckinridge (JPL) – Alternative Architectures
Alex Brown (CU/Boulder) – Target Selection
Peter Chen (CUA) – Lightweight Mirrors
David Folta (GSFC) – Deployment and Orbit and Formation Maintenance
Graham Harper (CU/Boulder) – Effects of ISM/Extragalactic Background
Kate Hartman & Joe Dolan (GSFC) – Editorial Assistance
Steve Kilston & Rich Reinert (BATC) – Operations Assurance/Validation, Mirrorsat design
Barry Kirkham (NGST) - Operations
Jesse Leitner (GSFC) – Precision Formation Flying
Alice Liu (GSFC) – Staged-Control Systems
Richard Lyon (GSFC) – Optical Modeling, Wavefront Sensing & Control, Image Reconstruction
Rud Moe (GSFC) – Role of Humans/Robots
David Mozurkewich (Seabrook Eng.) – Interferometer Architectures/Beam Combiner Designs
Charley Noecker (BATC) – Interferometer Architectures/Beam Combiner Designs
Sten Odenwald (QSS) & Carol Grady (CUA) – Education/Public Outreach
James Phillips (SAO) – Precision Metrology
Eric Stoneking (GSFC) – Staged-Control Systems, Target Acquisition
Fred Walter (SUNY/Stonybrook) – Design Reference Mission

Additional contributions are gratefully acknowledged from the wide range of science and technology investigators and collaborators on the Stellar Imager Vision Mission Team, as shown on the next page.

This work was supported, in part, by Vision Mission Study grants from NASA HQ to NASA-GSFC and from GSFC to Smithsonian Astrophysical Observatory, Seabrook Engineering, SUNY/Stonybrook, U. Colorado/Boulder, and STScI. Substantial complementary internal institutional support is gratefully acknowledged from all of the participating institutions.

Stellar Imager Vision Mission Team

The Stellar Imager Vision Mission concept is under development by NASA's Goddard Space Flight Center (K. Carpenter, PI), in collaboration with a broad variety of industrial, academic, and astronomical science institute partners, as well as an international group of science and technical advisors:

- Mission concept under development by NASA/GSFC in collaboration with experts from industry, universities, & astronomical institutes:

Ball Aerospace & Technologies Corp.	Lockheed Martin Adv. Tech. Center
NASA's Jet Propulsion Laboratory	Naval Research Laboratory/NPOI
Northrop-Grumman Space Tech.	Seabrook Engineering
Sigma Space Corporation	Smithsonian Astrophysical Observatory
Space Telescope Science Institute	State Univ. of New York/Stonybrook
Stanford University	University of Colorado at Boulder
University of Maryland	University of Texas/Arlington
European Space Agency	Kiepenheuer Institute
Potsdam Astronomical Institute	University of Aarhus

- Institutional and topical leads from these institutions include:
  K. Carpenter, C. Schrijver, R. Allen, A. Brown, D. Chenette, D. Mozurkewich, K. Hartman, M. Karovska, S. Kilston, J. Leitner, A. Liu, R. Lyon, J. Marzouk R. Moe, N. Murphy, J. Phillips, F. Walter

- Additional science and technical collaborators include:
  T. Armstrong, T. Ayres, S. Baliunas, C. Bowers, G. Blackwood, J. Breckinridge, F. Bruhweiler, S. Cranmer, M. Cuntz, W. Danchi, A. Dupree, M. Elvis, N. Evans, C. Grady, F. Hadaegh, G. Harper, L. Hartman, R. Kimble, S. Korzennik, P. Liewer, R. Linfield, M. Lieber, J. Leitch, J. Linsky, M. Marengo, L. Mazzuca, J. Morse, L. Mundy, S. Neff, C. Noecker, R. Reinert, R. Reasenberg, D. Sasselov, S. Saar, J. Schou, P. Scherrer, M. Shao, W. Soon, G. Sonneborn, R. Stencel, B. Woodgate

- International Collaborators include:
  J. Christensen-Dalsgaard, F. Favata, K. Strassmeier, O. Von der Luehe

- Student Participants include:
  Linda Watson (undergrad-Univ. Florida/CfA), Darin Ragozzine (undergrad-Harvard, grad-CalTech), Mikhail Dhruv (high school), Fonda Day (undergrad/CU)

Quick Facts: The Stellar Imager (SI) Vision Mission

SI is a UV-Optical, Space-Based Interferometer for 0.1 milli-arcsecond (mas) spectral imaging of stellar surfaces and stellar interiors, via asteroseismology, and of the Universe in general.

Science Goals

To understand:
- Solar and Stellar Magnetic Activity
  and their impact on Space Weather, Planetary Climates, and Life
- Magnetic Processes and their roles in the Origin and Evolution of Structure
  and in the Transport of Matter throughout the Universe

Mission and Performance Parameters

Parameter	Value	Notes
Maximum Baseline (B)	100 – 1000 m (500 m typical)	Outer array diameter
Effective Focal Length	1 – 10 km (5 km typical)	Scales linearly with B
Diameter of Mirrors	1 - 2 m (1 m currently)	Up to 30 mirrors total
λ-Coverage	UV: 1200 – 3200 Å Optical: 3200 – 5000 Å	Wavefront Sensing in optical only
Spectral Resolution	UV: 10 Å (emission lines) UV/Opt: 100 Å (continuum)	
Operational Orbit	Sun-Earth L2 Lissajous, 180 d	200,000x800,000 km
Operational Lifetime	5 yrs (req.) – 10 yrs (goal)	
Accessible Sky	Sun angle: $70^o \leq \beta \leq 110^o$	Entire sky in 180 d
Hub Dry Mass	1455 kg	For each of 2
Mirrorsat Dry Mass	65 kg (BATC) - 120 kg (IMDC)	For each of 30
Ref. Platform Mass	200 kg	
Total Propellant Mass	750 kg	For operational phase
Angular Resolution	50 μas – 208 μas (@1200–5000Å)	Scales linearly ~ λ/B
Typical total time to image stellar surface	< 5 hours for solar type < 1 day for supergiant	
Imaging time resolution	10 – 30 min (10 min typical)	Surface imaging
Seismology time res.	1 min cadence	Internal structure
# res. pixels on star	~1000 total over disk	Solar type at 4 pc
Minimum FOV	> 4 mas	
Minimum flux detectable at 1550 Å	5.0×10^{-14} ergs/cm^2/s integrated over C IV lines	10 Å bandpass
Precision Formation Fly.	s/c control to mm-cm level	
Optical Surfaces Control	Actuated mirrors to μm-nm level	
Phase Corrections	to λ/10 Optical Path Difference	
Aspect Control/Correct.	3 μas for up to 1000 sec	Line of sight mainten.

Executive Summary

The ultra-sharp images of the Stellar Imager (SI) will revolutionize our view of many dynamic astrophysical processes: The 0.1 milliarcsec resolution of this deep-space telescope will transform point sources into extended sources, and snapshots into evolving views. SI's science focuses on the role of magnetism in the universe, particularly on magnetic activity on the surfaces of stars like the Sun. SI's prime goal is to enable long-term forecasting of solar activity and the space weather that it drives in support of the Living With a Star program in the Exploration Era. SI will also revolutionize our understanding of the formation of planetary systems, of the habitability and climatology of distant planets, and of many magneto-hydrodynamically controlled processes in the Universe.

Primary Science Goals for the Stellar Imager

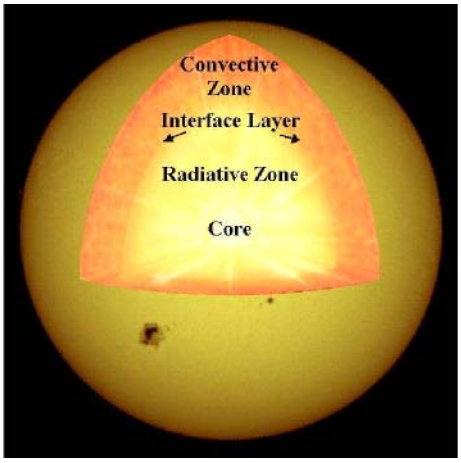


Figure ES-1: Internal structure of the Sun: an essentially static shell surrounds the nuclear furnace in the core, enveloped by a layer in which convective motions transport the solar luminosity. In doing so, these motions drive the solar dynamo.

Most of us rarely give the Sun a second thought. We do not question its presence or its apparent stability as we see it traverse the sky every day. The Sun is, however, a variable star. Its variability affects society by modulating Earth's climate. It also affects our technology, upon which we are becoming ever more reliant: eruptions on the Sun disrupt communications; affect navigation systems; cause radiation harmful to astronauts exploring beyond the Earth's atmosphere and to airline passengers traveling through it; and occasionally push power grids to fail. The cause of this variability is the Sun's magnetic field. This intangible and unfamiliar fundamental force of nature is created in the convective envelope of the Sun (see Fig. ES-1, showing the Sun's internal structure) by a process that we call the dynamo. The Stellar Imager will help us understand the dynamo process in the Sun by observing magnetic processes in Sun-like stars and in other astrophysical systems. That insight into solar activity will help us mitigate the effects of space weather, both on Earth and beyond.

Historical records show that the Sun can change its activity significantly; both upward and downward (see Fig. ES-2). Activity decreased, for example, for multiple decades during the 17th Century when Earth experienced the Little Ice Age. A sustained increase in activity – such as happened during the medieval Grand Maximum – may cause a warm spell, and will be associated with an increase in the frequency of space storms, and in the ultraviolet radiation that is harmful to life on Earth.

The dynamo is one of the truly large mysteries in astrophysics. There is at present no model for a stellar dynamo that can be used to forecast the Sun's activity on the time scale of months to decades. We know that the solar dynamo operates throughout the outermost 200,000 km of the solar interior, in and just below the convective envelope. The vastness of this volume relative to the smallest relevant scales precludes a complete numerical model. There is not even a generally accepted approximate dynamo model. In fact, the experts do not agree where most of the dynamo action occurs within the stellar interior, or which are the key processes that are involved.

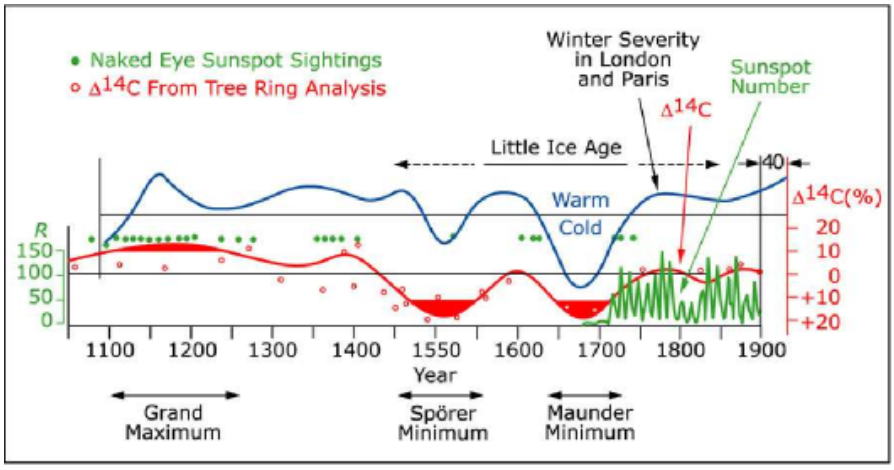


Figure ES-2: Solar activity and Earth's climate:

the nature of the link is yet to be understood, but the correlation suggests that solar activity somehow couples into the Earth's climate system. (From the Living With A Star initiative)

What makes understanding the solar dynamo so difficult? That answer involves two of the major developments of science in the 20th century: a stellar dynamo involves both *non-linear* and *non-local* effects. Such a dynamo can exhibit fundamentally different properties even for relatively small changes in the processes involved. In other words: if a dynamo model does not incorporate all relevant physics in sufficient detail, it will not enable us to predict solar activity on time scales of years or more, or to understand its gross characteristics in the distant past and future. In order to develop a dynamo model with predictive value, we must establish which processes are involved, and which approximations are allowed.

It would take hundreds of years to validate a dynamo model for the Sun using only observations of the Sun, given its irregular 11-year magnetic heartbeat and the significant overlying long-term modulations. The more efficient alternative is to test and validate dynamo models using Stellar Imager observations of the variable magnetic activity of a broad sample of stars. Indeed, surface magnetic activity records of stars on or near the lower main sequence (e.g. from the Mount Wilson Observatory Ca II H&K survey, Soon & Yaskell 2004) show variability similar to the Solar variability, including Maunder minimum-like phases, on time scales of many decades. For example, Figure ES-3 shows two stars with cycles like that of the Sun, with one (the K0 V star HD 3651) showing a rapid decline in its chromospheric activity, possibly reflecting entry into a Maunder-minimum state.

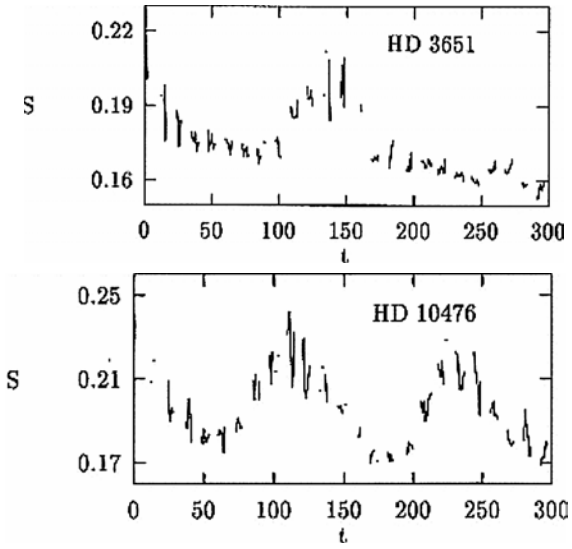


Figure ES-3: Going into a Maunder minimum?

The chromospheric activity of HD 3651 (top panel), as measured by CaII emission strength, is fading over the years, suggesting that this star is rapidly going into a Maunder-minimum state, while HD 10476 (lower panel) continues to cycle like the current Sun. The time "t" on the horizontal axis is measured in months. (Frick et al. 1997).

Key to successfully navigating the route to a workable, predictive dynamo model is the realization that *in order to understand the solar dynamo, we need a population study: we need to study the dynamo-driven activity in a sample of stars like the Sun, and compare it to observations of young stars, old stars, binary stars, etc*. The potential for a breakthrough in our understanding and our prediction ability lies in spatially-resolved imaging of the dynamo-driven activity patterns on a variety of stars. These patterns, and how they depend on stellar properties (including convection, differential rotation and meridional circulation, evolutionary stage/age), are crucial for dynamo theorists to explore the sensitive dependences on many poorly known parameters, to investigate bifurcations in a nonlinear 3-dimensional dynamo theory, and to validate the ultimate model.

"What then is a magnetic field and how does it operate in the astronomical universe to cause all the 'trouble' that we have attributed to it? What is this fascinating entity that like a biological form is able to reproduce itself and carry on an active life in the general outflow of starlight and from there alter the behavior of stars and Galaxies?" E.. Parker, 1979

Direct, interferometric imaging – the goal of the Stellar Imager - is the only way to obtain adequate information on the dynamo patterns for stars of Sun-like activity. Alternative methods that may offer limited information on spatial patterns on much more active stars fail for a Sun-like star:

- rotationally-induced Doppler shifts in such stars are too small compared to the line width to allow Zeeman-Doppler imaging (see section 1.3.2.1)

- the activity level is insufficient to lead to significant spectral changes associated with magnetic line splitting

- rotational modulation measurements are inherently subject to deconvolution limitations that leave substantial ambiguities in the latitude distributions, locations and sizes of spots, and cannot be used to understand the facular contributions in quiet regions that are governed by field dispersal and differential rotation.

The direct imaging by SI of stellar activity will sidestep these problems. Equally importantly, the asteroseismic observations planned with SI will determine the internal properties of stellar structure and rotation, thus directly providing crucial information relevant to the physical operation of the dynamo mechanism.

Imaging magnetically active stars and their surroundings will also provide us with an indirect view of the Sun through time, from its formation in a molecular cloud, through its phase of decaying activity, to its ultimate death beyond the red-giant phase during which the Sun will swell to about the size of the Earth's orbit.

Table ES-1 summarizes the observational requirements that must be met to achieve SI's prime science goals, which are discussed in more detail in Chapter 1.

Table ES-1: Requirements to achieve the prime science goals of the Stellar Imager

Imaging stellar activity using emission from the outer atmosphere:

Image nearby main-sequence and giant stars with at least 1,000 resolution elements on their surface, in UV emission lines originating in the outer atmosphere; *requires a baseline of 500 m for a star at 4 parsec*.

Construct images within ~1% of the stellar rotation period, i.e. 6 h for a star like the Sun; *requires efficient reconfiguration and/or a large number of interferometer components, and an increasing number of interferometer components for increasing rotation rate.*

Compile at least ~ 30 images within one stellar rotation; requires optimized target lists and efficient repointing.

Revisit stars during 3-6 month intervals, spanning > 5 yrs; requires a long operational life, and preferably replaceable component spacecraft.

***Imaging stellar interiors with asteroseismic techniques**:*

Achieve 30 resolution elements on stellar disks with 1 min. cadence, in a broad passband in the optical; *requires at least 9 optical elements, with meter-class collecting areas.*

Continuous observations for ~one rotation, with a duty cycle better than ~ 90%; requires stable environment.

General Astrophysics with the Stellar Imager

Of all the stars in the Universe, only one has been seen as it truly is...highly complex and ever changing. Yet, the Sun is only one of many types of stars. Our comprehension of stars forms the foundation of our understanding of the Universe. Magnetic fields affect the evolution of stars and planetary systems in all phases, from the formation of the star and its planets, to the habitability of these planets through the billions of years during which they live with their stars. *SI will enable detailed study of magnetic processes and their roles in the Origin and Evolution of Structure and in the Transport of Matter throughout the Universe*.

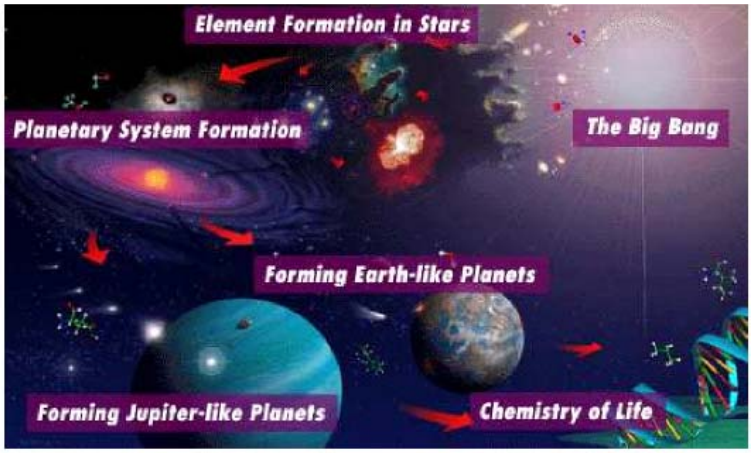


Figure ES-4: The history of the Universe, from the Big Bang, through the formation of stars and planets, to life.

The Stellar Imager focuses on stellar magnetic activity and its role in planetary system formation, the origins of life, space weather, and the habitability of Earth. (Image by P. Rawlings, JPL)

A long-baseline interferometer in space will benefit many fields of astrophysics. Imagine, for example, unprecedented images of active galactic nuclei, quasi-stellar objects, supernovae, interacting binary stars, supergiant stars, hot main-sequence stars, star-forming regions, and protoplanetary disks. Figure ES-5 shows simulated SI results, computed using SISIM (Rajagopal et al. 2003), assuming 30 mirror elements distributed in a non-redundant pattern with the indicated maximum baselines.

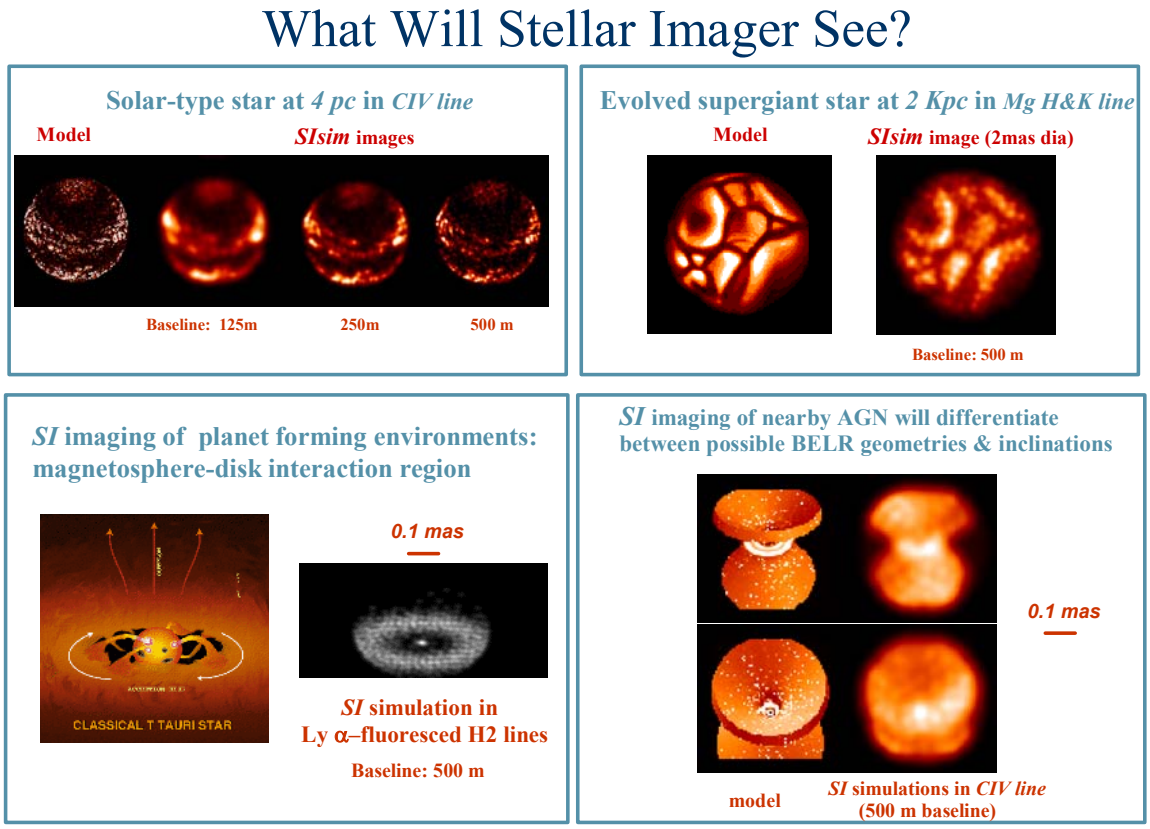


Figure ES-5: Simulations of some of SI's capabilities for UV imaging.

SI will produce images with hundreds of times more detail than Hubble, which in turn will bring the study of dynamical evolution of many astrophysical objects into reach: hours to weeks between successive images will detect dramatic changes in many objects, e.g., mass transfer in binaries, pulsation-driven surface brightness variation and convective cell structure in giants and supergiants, jet formation and propagation in young planetary systems, reverberating active galactic nuclei, and many others (see Fig. ES-6).

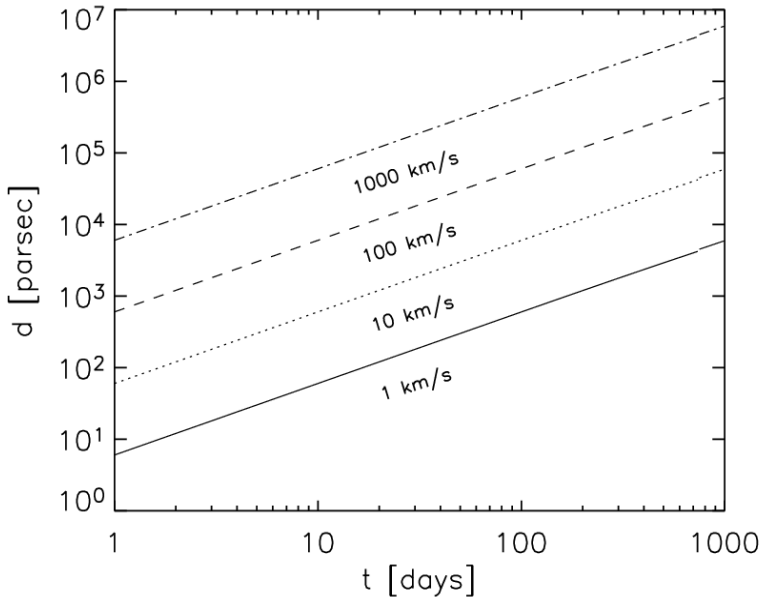


Figure ES-6: Minimum time interval between successive SI images required to resolve the motion of a feature moving at different speeds (line labels) as a function of the object's distance.

The SI Mission Concept

The Stellar Imager (SI) is a mission to understand the various effects of magnetic fields of stars, the dynamos that generate them, and the internal structure and dynamics of the stars in which these dynamos operate. *The ultimate goal of the mission is to achieve the best -possible forecasting of solar activity as a driver of climate and space weather on times scales ranging from months up to decades, and an understanding of the impact of stellar magnetic activity on life in the Universe. The road to that goal will revolutionize our understanding of stars and stellar systems, the building blocks of the Universe.*

The Stellar Imager is a UV/optical interferometer designed to provide images with some 1,000 picture elements of a sample of dozens of stars over a period of up to a decade. This will reveal the surface patterns of dynamos (e.g., Fig. ES-7) in widely different stars, allowing us to differentiate between the various dynamo models.

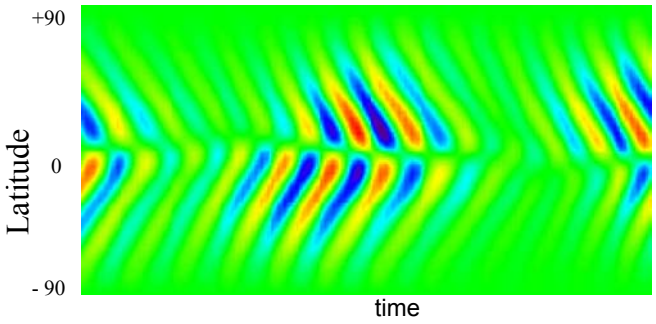


Figure ES-7: The evolution of a latitude-dependent activity pattern for a model dynamo.

Observing the shape of such a pattern and of the large-scale surface flows that help shape it for at least one cycle for a sample of stars will help us discriminate between dynamo models. (Figure from S. Tobias)

SI, with a characteristic angular resolution of 0.1 milli-arcseconds at 2000 Å, represents an advance in image detail of several hundred times over that provided by the Hubble Space Telescope. The Stellar Imager will zoom in on what today - with few exceptions - we only know as point sources, revealing processes never before seen, thus providing a tool as fundamental to astrophysics as the microscope is to the study of life on Earth. The potential of SI for imaging solar-type activity in a star at 4pc is illustrated in Fig. ES-8.

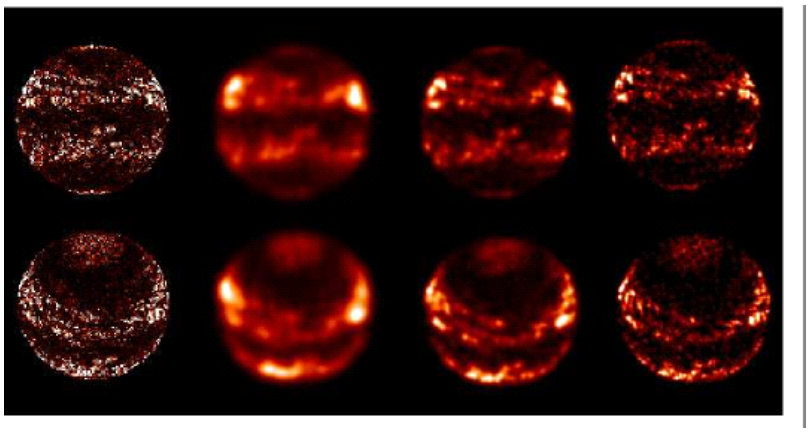


Figure ES-8: The potential of the Stellar Imager:

Model CIV 1550 Å images of a star like the Sun (left) and simulated images for maximum baselines of 125 m, 250 m, and 500 m (2nd-4th columns). The simulated reconstructions assume observations of a star at 4 pc with 870 baseline pairs (e.g. 2 configurations of a 30-element array in a minimum ambiguity, low-redundancy layout or 20 configurations (rotations) of an array of 10 elements in a Y-formation), with 800 CLEAN iterations. The top and bottom rows show views of a Sun-like star with a rotation axis in the plane of the sky and with that axis tilted by 40°, respectively. (Simulations computed with SISIM/Rajagopal et al. 2003)

The full-scale SI is an ultraviolet/optical aperture synthesis imager composed of at least 9, up to perhaps as many as 30, array elements on what we call mirrorsats and a central hub with focal-plane instrumentation that allows spectral energy resolution in pass bands from a few up to hundreds of Angstroms, throughout the UV/optical region from 1216-5000 Å.

The SI mission will allow us not only to image the surfaces of stars, but also to sound their interiors using asteroseismology in order to image internal structure, differential rotation, and large-scale circulations; this will provide accurate knowledge of stellar structure and evolution and complex transport processes, and will impact numerous branches of (astro)physics. *For arrays of 9 or more optical elements, asteroseismic imaging of structure and rotation is possible with a depth resolution of 20,000 km for a star like the Sun.*

The full SI mission may be built up by starting with a small number of optical (array) elements, perhaps utilizing both interferometry and high-resolution spectroscopy. Added optical elements will increase image quality and time resolution. Table ES-2 summarizes the primary science goals and instrument requirements.

Table ES-2: *Overview of the SI science, design, and instrument requirements*

Science requirement	**Design requirement**	**Instrument requirement**
Allow imaging in UV and optical of astrophysically interesting targets with 0.1 mas (milli-arcsec) resolution.	Optical system to be optimized for observing from 1200 Å to at least 5000 Å, in multiple UV pass bands of 2-10 Å width.	Variable effective aperture or interferometer baselines from 100 - 1000 m.
Enable imaging of stars and extended complex sources such as star- and planet-forming regions, accretion disks and jet-forming regions, interacting binaries, super massive black hole environments, etc.	Image frequency components to be high enough for complex sources, and point spread function with well-defined core regions.	20-30 apertures in non-redundant pattern to provide sufficient Fourier (u,v) coverage for ultimate image reconstruction
Image the chromospheric or transition-region emission of a star like the Sun with sufficient resolution to locate large active regions and to map the large-scale surface field.	UV/optical imaging to yield ~700 resolution elements on the disk, or 30 across its equator, for a Sun-like star at 4 pc, equivalent to a resolution of ~0.1 milli-arcseconds.	Effective aperture or interferometer baselines of at least 500 m.
Time to complete one full image should be short enough that rotational smearing does not compromise the required resolution of stellar images.	Image integration time to be less than $P/30\pi$ for a stellar rotation period P (e.g., 6 h for a Sun-like star, or 2.5 h for a star with P = 10 d.)	Individual primary mirrors at least 1 m in diameter; # of interferometer elements ~30, unless fast reconfiguration
Observe at least 25 magnetically-active (cool) single and binary stars over five years, each at least twice per year, to study field pattern evolution and properties of cycles.	Baseline mission to exceed 5 yr; baseline target list to include at least 25 core program stars.	Slew speeds > 10 deg/hour and accessible band on the sky (solar beta angle from 70 to 110 degrees)
Observe at least 25 cool single and binary stars with 30 images within a rotation period, each at least once per year, to measure the field source properties, differential rotation, and other large-scale flows.	Re-targeting must be completed within 2-3 h to enable observing of at least 3 Sun-like targets within a 24 h period. SI pointing to allow imaging of stars for at least 30 days continuously.	Design to allow imaging at least in a 20-30°range centered 90°from the Sun-SI direction
Enable astero-seismology in near-UV or optical to measure internal differential rotation and effects of magnetic fields on internal stellar structure.	Asteroseismological resolution of 30 elements on stellar disks, at a cadence of 1 min. for at least a stellar rotation, at a duty cycle of better than ~90%, in up to three visible passbands of up to 100 Å wide.	Effective aperture to collect 10^{12} photons/band per star per rotation period. Instantaneous number of independent baselines to exceed ~60, and thus # of optical elements to exceed ~8.

The Baseline SI Mission Design

The current baseline architecture concept (see front cover) for the full Stellar Imager (SI) mission is a space-based, UV-Optical Fizeau Interferometer with 20-30 one-meter primary mirrors, mounted on formation-flying "mirrorsats" distributed over a parabolic virtual surface whose diameter can be varied from 100 m up to as much as 1000 m, depending on the angular size of the target to be observed. The individual mirrors are fabricated as ultra-smooth, UV-quality flats and are actuated to produce the extremely gentle curvature needed to focus light on the beam-combining hub that is located at the prime focus from 1 – 10 km distant. The focal length scales linearly with the diameter of the primary array: a 100 m diameter array corresponds to a focal length of 1 km and a 1000 m array with a focal length of 10 km. The hub and all of the mirrorsats are free-flyers in a tightly-controlled formation in a Lissajous orbit around the Sun-Earth L2 point. A second hub is strongly recommended to provide critical-path redundancy and major observing efficiency enhancements. The observatory may also include a "reference craft" to perform metrology on the formation, depending on which metrology design option is chosen. The details of the baseline design are presented in Chapter 2. See also the "Quick Facts" sheet on page iv for additional information and specifications.

Figure ES-9 shows two launch options for SI: a single launch suffices if only one hub is deployed initially (along with a reference spacecraft and 30 mirrorsats), while two launches are needed for designs that include a second beam-combining hub.

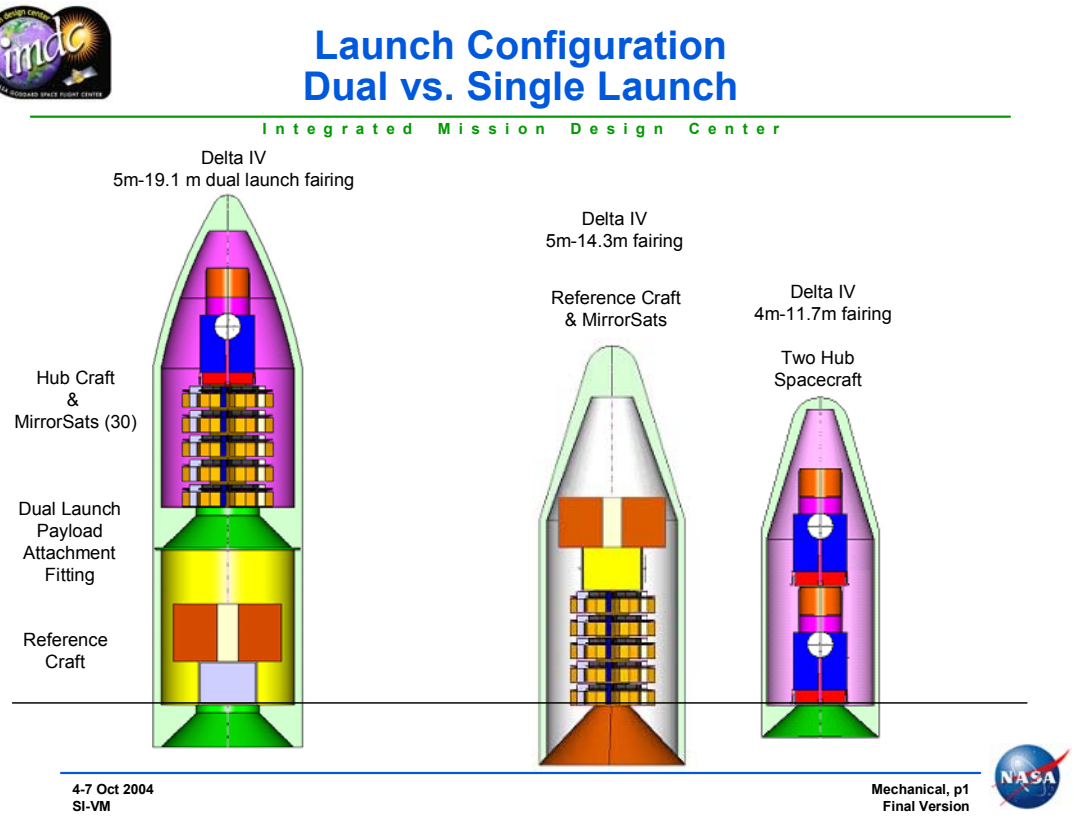


Figure ES-9: Two launch options for SI: a single Delta IV heavy vs. two Delta IV launches.

Figure ES-10 provides an overview of the selected architecture: the upper panel shows a cross-sectional schematic of the entire observatory, while the lower panel shows a close-up of the hub and its major components.

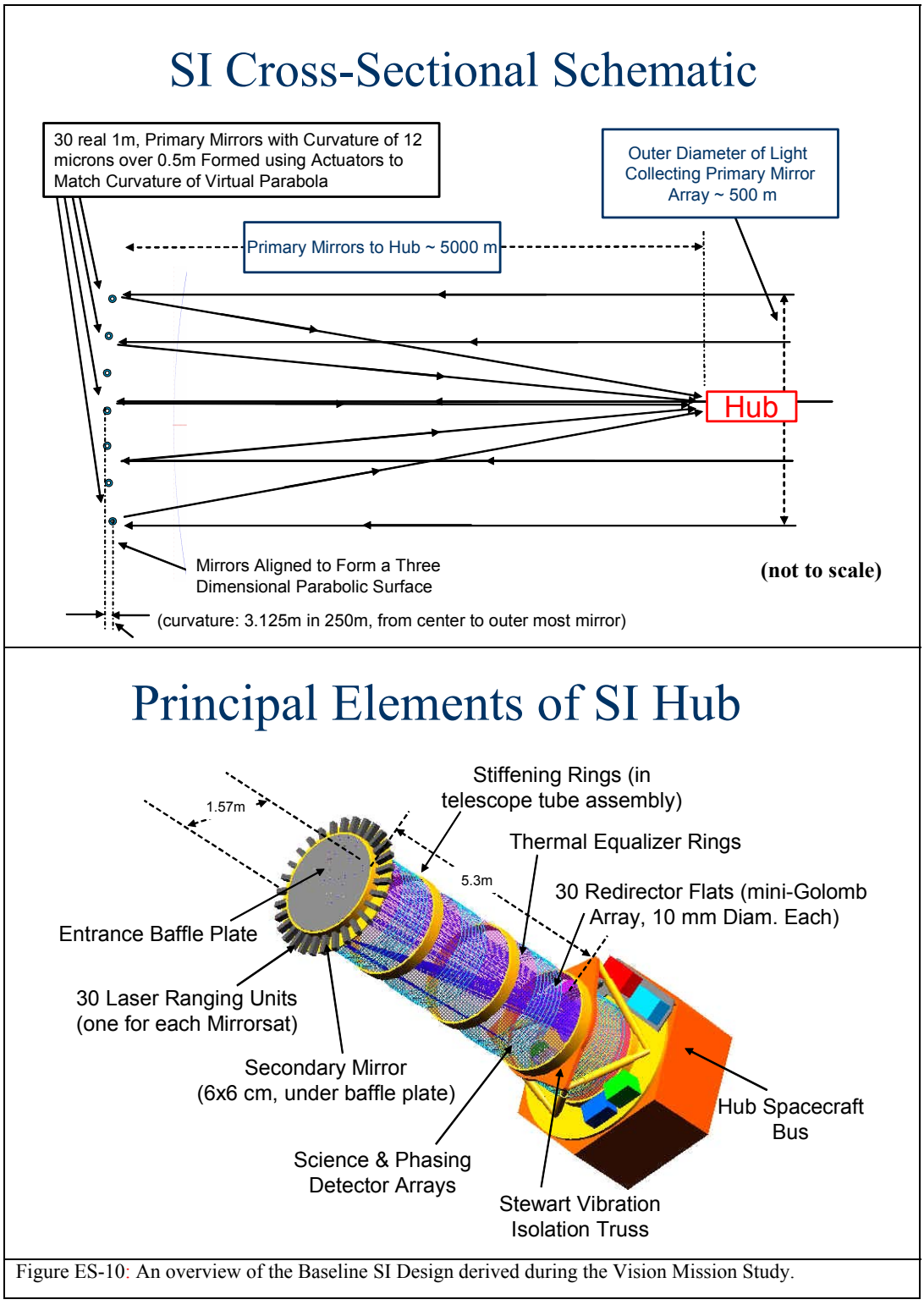


Figure ES-10: An overview of the Baseline SI Design derived during the Vision Mission Study.

Technology Developments Needed to Enable SI

The major enabling technologies needed for SI are:

- ***formation-flying of ~30 spacecraft***
  - *deployment and initial positioning of elements in large formations*
  - *real-time correction and control of formation elements*
    - *staged-control system (km → cm → nm)*
  - *aspect sensing and control to 10's of micro-arcsec*
  - *positioning mirror surfaces to 5 nm*
  - *variable, non-condensing, continuous micro-Newton thrusters*
- ***precision metrology over multi-km baselines***
  - *2nm if used alone for pathlength control (no wavefront sensing)*
  - *0.5 microns if hand-off to wavefront sensing & control for nm-level positioning*
  - *multiple modes to cover wide dynamic range*
- ***wavefront sensing and real-time, autonomous analysis***
- ***methodologies for ground-based validation of distributed systems***
- ***additional challenges***
  - *mass-production of "mirrorsat" spacecraft: cost-effective, high-volume fabrication, integration, & test*
  - *long mission lifetime requirement*
  - *light-weight UV quality mirrors with km-long radii of curvature (likely through active deformation of flats)*
  - *larger format (6 K x 6 K) energy resolving detectors with finer energy resolution (R=100)*

Precision metrology, precision formation-flying, and the methodologies for pre-launch, ground-based testing and integration of such a large, distributed system are identified as the tallest poles among numerous technical challenges. The long mission lifetime requirement is also a concern among the designers: the hub will have redundant components, but it may be necessary to seriously consider building a backup hub for launch on-need or original deployment nonetheless, and additional backup mirrorsats will likely need to be flown so they can be put into the operating array as the original set suffers expected failures (the mirrorsats were designed as inexpensive, low-redundancy, mass-produced craft in these studies). Other technical developments of major importance and/or difficulty include the development of aspect control to 10's of μarcsecs, wavefront sensing and control of large, sparse aperture systems, larger format energy resolving detectors with higher energy resolution, and lightweight, UV-quality mirrors. Further details are given in Chapter 3.

The Stellar Imager, Science Mission Directorate Goals, and Other Projects

Fitting naturally within the NASA long-term time line, SI complements defined and proposed missions (Terrestrial Planet Finder – I, Life Finder, and Planet Imager), and with them will show us entire other solar systems, from the central star to their orbiting planets. It moreover fits on the technology roadmap that leads from interferometers like Keck and SIM to TPF-I/Darwin, MAXIM/Black Hole Imager, Life Finder, and the Planet Imager.

Stellar Imager was included in the 2000 and 2003 SEC Roadmaps and is now identified as a "*Flagship and Landmark-Discovery Mission*" in the draft 2005 Sun Solar System Connection (SSSC) Roadmap. SI is also a candidate for a "*Pathways to Life Observatory*" in the Exploration of the Universe Division (EUD) Roadmap (May, 2005). SI will provide an angular resolution over 200x that of the Hubble Space Telescope (HST), which currently offers the best angular resolution imaging in the UV, and will resolve for the first time the surfaces of Sun-like stars and the details of many astrophysical objects and processes.

The Stellar Imager is a natural culmination of science addressed with ongoing ground-based observatories and a series of space missions (Table ES-3). These efforts will provide information on long-term disk-integrated variability, large-scale internal structure and evolutionary status, distances and other fundamental stellar properties, binary properties, and low-resolution surface imaging for a subset of target classes. Other missions, such as SIM and TPF-I are space-based interferometers in the technology roadmap for the Stellar Imager.

Table ES-3*: The Stellar Imager is part of an array of space and ground-based instrumentation that contribute to our understanding of stellar activity and internal structure.*

Some of these, and their potential role in the study of stellar activity, are summarized below:

Project	Role in activity studies	Observational Technique and/or Technology
Stellar Imager	Dynamo patterns, (internal) dif. rotation binary interaction	UV/Optical interferometry <0.1 mas (milli-arcsecs)
MAXIM	Coronal structure	X-ray interferometry
Terrestrial Planet Finder	Binary properties	SI Technology precursor, IR, free -flying, nulling interferometer, 0.75 mas
Space Interferometry Mission	Binary properties	SI Technology precursor, boom interferometer
James Webb Space Telescope	Stellar mass loss, giant chromospheres	IR imaging, 100 mas
Ground-based interferometry: Keck, Large Binocular Telescope, Very Large Telescope Interferometer	Giant-star imaging, binary properties	Technology precursors
GAIA	Determination of stellar properties	High -precision parallaxes
MOST, COROT, KEPLER	Internal stellar structure	Asteroseismology
Ground-based spectroscopy	Activity monitoring, limited imaging	Automatic telescopes,(Zeeman) Doppler imaging

SI complements and builds on observations made by ground-based interferometers, by asteroseismology missions, JWST, and other missions. It complements TPF-C/I by providing a view of the space-weather environment of the planetary systems studied in those missions, and thus provides critical data needed to understand fully which of the detected planets are indeed habitable.

Table ES-4 summarizes SI's fit into the national science and technology development priorities.

Table ES-4: The Stellar Imager fits in the national science priorities, the NASA strategic plan, the Living With A Star initiative, and the technology roadmap.

SI meets scientific priorities identified by the National Academy of Sciences Astronomy and Astrophysics Survey Committee (2001). With SI we can "survey the universe and its constituents," "use the universe as a unique laboratory," "study the formation of stars and their planetary systems, and the birth and evolution of giant and terrestrial planets," and, by focusing on the driver of space weather in past, present, and future, "understand how the astronomical environment affects Earth."
SI is responsive to a key national priority: imaging of magnetically active stars provides the only means to test any theory of solar magnetic activity as the driver of space weather and climate that can be achieved within a decade after launch.
SI fits in the NASA/OSS strategic plan: it complements the Living With A Star initiative, and shares much of the scientific and technological road that leads to other interferometers such as the Terrestrial Planet Finder, Planet Imager, and the MicroArcsecond X-ray Imaging Mission.

A Timeline for Stellar Imager

A rough flow for the development process for the SI mission or an equivalent long-baseline, UV/Optical, space-based interferometer is outlined below:

2005: Complete Vision Mission Study

2005-08: Continue studies of multi-element fine optical control with Fizeau Interferometer Testbed (FIT)

2005->: Continue other technology development efforts, including precision formation flying, micro-newton level thrusters, wavefront sensing and control, methodologies for integration and test of large distributed system, energy resolving UV-Optical detectors

2006: Develop Pathfinder Concept suitable for future Probe/Discovery-type opportunities and work with other NASA (e.g., ST-9) and ESA projects (e.g., EMMA, SMART-2/LISA-PF) to collaboratively develop relevant technologies

~2015: Fly pathfinder mission(s)

~2025: Fly full mission

Additional information on the Stellar Imager can be found at http://hires.gsfc.nasa.gov/~si/

Contents

Chapter 1

Science Rationale

The Stellar Imager (SI) is a UV/Optical deep-space telescope for 0.1 mas (milli-arcsecond) spectral imaging designed to enable an understanding of:

- **Solar and Stellar Magnetic Activity and Its Impact on Space Weather, Planetary Climates, and Life**

- **Magnetic Processes, the Origin and Evolution of Structure, and the Transport of Matter Throughout the Universe**

The Key science goals of the SI mission are to:

- **Study the evolution of stellar magnetic dynamos by resolving patterns of surface activity & internal structures & flows in a diverse sample of stars**

  - to improve long-term forecasting of solar & stellar activity and understand the impact of stellar magnetic activity on planet formation, planetary climates and the origin and maintenance of life
  - to understand the variable Sun-Earth system

- **Complete the assessment of external solar systems begun with the planet-finding and imaging missions**

  - by imaging the central stars of those systems to determine the impact of their activity on the habitability of the surrounding planets

- **Study the Universe at ultra-high angular resolution to understand**

  - the origin of stars, planetary systems, and life
  - the structure of stars and the life cycle of stars and their planetary systems
  - internal transport processes in stars at different ages, their impact on stellar evolution, and their consequences for the chemical evolution of galaxies
  - dynamo and accretion processes, mass-exchange, and mass flows in, e.g., active galactic nuclei, black hole environments, supernovae, binary stars, and highly evolved stars

SI will provide an angular resolution several hundred times that of the Hubble Space Telescope (HST) and will resolve for the first time the surfaces of Sun-like stars and the details of many astrophysical objects and processes. Examples of scientific areas of study for the Stellar Imager include:

- **Magnetic Processes in Stars**
  - *activity and its impact on planetary climates and on the origin and maintenance of life; stellar structure and evolution*
- **Stellar interiors**
  - *in stars outside solar parameters*
- **Infant Stars/Disk systems**
  - *accretion foot-points, magnetic field structure & star/disk interaction*
- **Hot Stars**
  - *hot polar winds, non-radial pulsations, rotation, structure; envelopes and shells of Be-stars*
- **Cool, Evolved Giant & Supergiant Stars**
  - *spatiotemporal structure of extended atmospheres, pulsation, winds, shocks*
- **Supernovae & Planetary Nebulae**
  - *close-in spatial structure*
- **Interacting Binary Systems**
  - *resolve mass-exchange, dynamical evolution/accretion, study dynamos*
- **Active Galactic Nuclei, Quasars, Black-Hole Environments**
  - *transition zone between Broad and Narrow Emission Line Regions; origin and orientation of jets; distances*

1.1 Key Objectives

1.1.1 Understand Magnetic Activity of the Sun and Stars

In recent decades, we have rapidly increased our knowledge of the weather in the space surrounding Earth (the focus of NASA's Living With A Star initiative); all of this space weather is driven by solar activity. Yet our quantitative understanding of this activity is still rather poor, and our ability to forecast it is very limited, particularly on time scales of years or more. We cannot quantify solar activity accurately enough for the distant past, nor do we understand how it then affected the formation and early evolution of the planetary system, the atmospheres of its planets, or developing life on at least one of these. This is all the more important because we know from stellar studies that the magnetically-driven changes in the Sun have been rather benign in recent decades compared to what we see happen in stars that are otherwise very much like the Sun. Our lack of understanding of the intricate links between the many processes that are involved in the driving of the solar dynamo is our primary obstacle to quantitative forecasting.

We know that magnetic fields and the associated electrical currents played an important, perhaps even crucial, role in the formation of the Sun and its solar system, and that they continue to do so today. These electromagnetic forces – not immediately obvious in our current planetary biosphere – are important in stellar evolution, because most of the matter in and around stars is electrically charged (i.e., forms a plasma), which allows electromagnetic forces to couple distant regions through a mixture

of attractive and repulsive forces. Such coupling is often associated with efficient transport and dissipation of energy and (angular) momentum, resulting in either gradual or explosive energy release. This causes, for example, the outer atmosphere of the Sun to be a thousand times hotter than its surface, so that it radiates in ultraviolet and X-ray light. Furthermore, the electromagnetic field can accelerate particles to velocities that are a significant fraction of the speed of light. We have only in recent decades started to realize just how important the emission of energetic photons and particles is in the coupling of the Sun and the Earth, possibly extending to the evolution of life on Earth.

The combination of solar and stellar observations allows us to appreciate the role of magnetic fields in the initial phases of formation of the solar system, estimate the evolution of the solar brightness in X-rays and in the (extreme) ultraviolet, and assess the kinds of behavior that we can expect from the present-day active Sun. *Combined solar and stellar observation provide the complementary knowledge needed to understand the Sun's dynamo.* Multiple high-quality observatories have imaged the Sun for many years, but the imaging of stars is presently achieved only with cumbersome and highly ambiguous methods, none of which work for a star of solar-like activity. *The Stellar Imager's primary science goal is to image a sample of magnetically active stars with enough resolution to map their evolving dynamo patterns and thus facilitate long-term space-weather forecasts and understanding of the role of solar activity from the very formation of the solar system onward.*

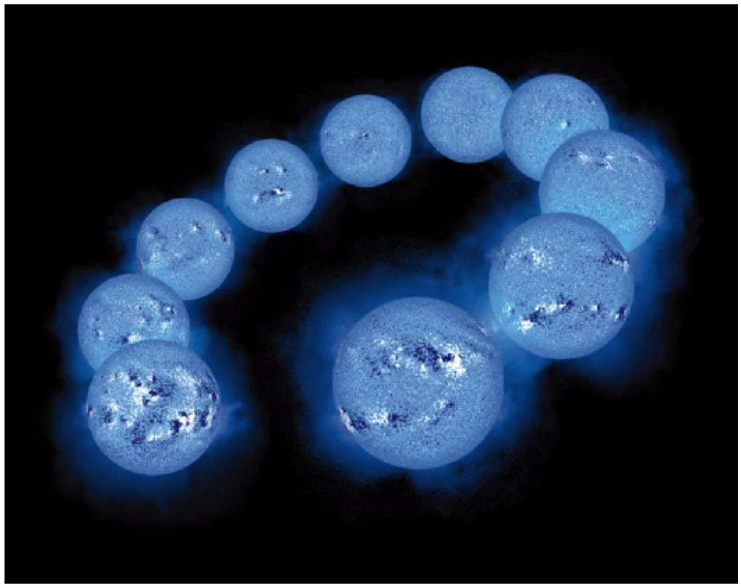


Figure 1.1: The magnetic field on the surface of the Sun, spanning a period 7.5 yr.

The two polarities are shown as dark blue to black and as light blue to white, respectively. The earliest image, taken in 1992, is at the lower left. Starting from this image, follow the arc of images and notice the variation in pattern and strength of the field from one image to the next, taken one half to one year later than the previous. (LMSAL/Schrijver)

1.1.1.1 What is stellar magnetic activity?

Classical stellar models approximate stars like the Sun as dull objects, evolving on a time scale of millions of years or longer, and doing little else. If such classical models described all that stars are, there would be little space weather to speak of, and the solar atmosphere would have been a rather cool, tenuous, and uneventful environment.

In reality, the solar outer atmosphere is highly structured and dynamic, so hot that it shines at extreme-UV and X-ray wavelengths (Figs. 1.1 and 1.2) and that it evaporates into a gusty solar wind, while often displaying immense explosions. These phenomena are caused by the interaction of matter and the complex and dynamic electromagnetic field that it carries. That magnetic field is also the cause of sunspots and the multitude of smaller field concentrations that cluster around them. These surface

fields modulate the Sun's overall brightness: over the past two decades, the bolometric solar irradiance has gone up and down by ~ 0.2%, in step with the 11-year modulation of solar activity. At EUV and X-ray wavelengths, in contrast, the variation reaches a factor of 10 to 100, respectively.

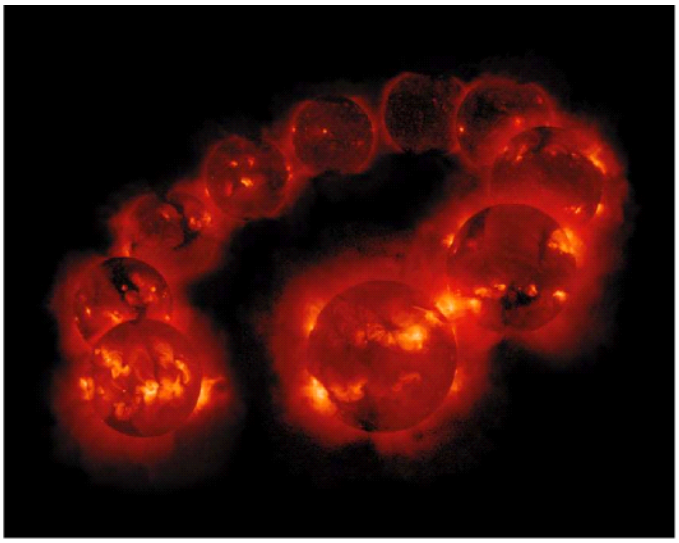
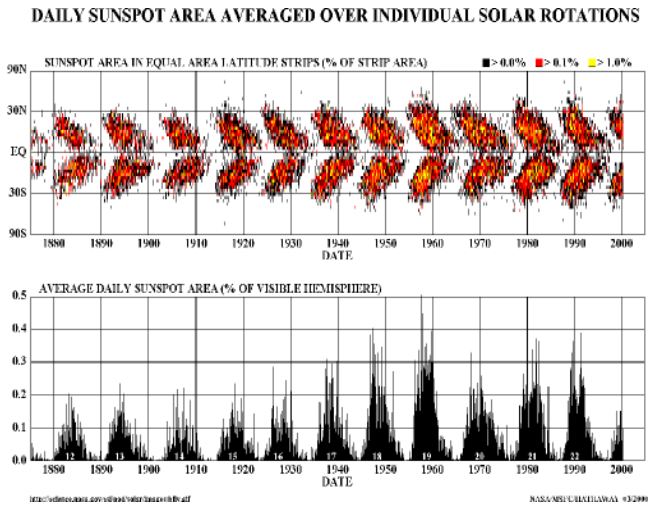


Figure 1.2: The solar corona going from solar-cycle maximum to minimum and back to maximum again, as observed by YOHKOH's Solar X-ray telescope.

The magnetic field of the Sun emerges onto the surface in bipolar regions. In these "active regions," the two magnetic polarities penetrate the surface side by side, connected by field lines through the outer atmosphere. The emergence frequency of active regions onto the surface of the Sun is modulated by a quasi-periodic sunspot cycle, with a duration of approximately 11 years. Two successive sunspot cycles add up to a full magnetic cycle (Figs. 1.1 and 1.3).

Figure 1.3: The solar "butterfly diagram" shows the positions of the spots for each rotation of the sun since May 1874.

Sunspots do not appear at random over the surface of the sun but are concentrated in two latitude bands on either side of the equator. The active-latitude bands first form at mid-latitudes, then widen, and move toward the equator as each cycle progresses. (image by MSFC)



The first active regions of a cycle start at mid-latitudes. Later regions emerge ever closer to the equator, albeit with a substantial range about the mean latitude. As one cycle peters out near the equator, the next, of opposite polarity, has already started at higher latitudes: successive cycles overlap by several years.

The solar magnetic field has a dramatic impact on the atmosphere above the solar surface. Electrical storms and wave phenomena deposit energy, as a result of which the temperature is kept at several million Kelvin, depending on the strength of the field. This high temperature lets plasma rise up into the outer atmosphere against gravity, from where it emits at EUV to X-ray wavelengths (Fig. 1.2). Even in the quietest conditions, parts of the hottest domains within the outer atmosphere, the corona, are forced open into the heliosphere as hot plasma "evaporates" in the solar wind that is further accelerated by some as yet unknown process. The wind speed and density depend on the details of the many sources of this heliospheric field; these sources, in turn, depend on the magnetic configuration of the entire sphere.

Large-scale circulations, smaller-scale convective motions, and the continual emergence of new field tangle the atmospheric magnetic field, stressing it so much that large explosions and eruptions are triggered in which the stresses can relax. These are called flares or mass ejections, depending on whether they primarily emit high–energy radiation (both particles and photons) or eject material and magnetic field into interplanetary space.

Other cool stars exhibit similar phenomena, which are collectively referred to as stellar magnetic activity. Whereas effects of solar magnetic activity have been studied ever since the discovery of sunspots (early in the 17th Century, when the telescope was invented; the fact that they were magnetic phenomena was discovered only 300 years later, in 1908), the study of the stellar equivalent did not seriously start until the second half of the 20th Century. Complementary studies of solar and stellar activity taught us that the Sun is a rather inactive star, typical of its class and age. We now know that many Sun-like stars show activity cycles like that of the Sun, although these actually form a minority: only one in three stars of comparable type and activity exhibits clear cyclic activity.

The levels of activity in cool stars span an enormous range. The most active stars have more than 50% of a visible hemisphere covered by dark starspots, compared to at most 2% for the Sun. With half of their surface covered by spots, much of the rest is likely also covered by a strong magnetic field, but not strong enough that it forms dark spots. In inactive stars, the coverage by spots is insufficient to be detectable by current instruments, but there the associated emissions from the hot outer atmosphere can be used as a measure of activity. At the extreme lower end of activity, inactive giant stars have been observed for which the entire stellar coronae, with temperatures exceeding 1 million Kelvin, are dimmer than the darkest coronal regions on the Sun (the so-called coronal holes).

We have also learned that tidal interaction between companion stars in compact binary systems enhances activity, that stellar rotation is slowed over time by stellar winds flowing away from cool stars, that heating of the outer atmosphere occurs primarily because of the action of the magnetic field and less because of the damping of purely acoustic waves, and that a globally functioning dynamo that generates and maintains significant stellar activity requires convection immediately below the surface as well as some rotation of the star.

The activity of young Sun-like stars is generally much higher than on the Sun: X-ray flares can be 1,000 times larger than those on the Sun, half a stellar hemisphere can be covered by starspots, and the stellar wind may extract so much angular momentum from active, young stars that they decelerate significantly on a time scale of tens of millions of years instead of a billion years as is the case for the present-day Sun. The perpetual spin down by this magnetic brake causes stellar activity to continue to decrease with time.

An interesting side effect of the loss of stellar angular momentum through the magnetized stellar wind occurs in tidally interacting binaries. As each active star in the binary loses angular momentum, the tidal coupling tries to keep the stars rotating at the orbital period. But that means that part of the orbital momentum is transferred into the rotational momentum. As a result the orbit shrinks. This leads

to a shorter period, which induces a higher rate of stellar rotation, and thus higher activity and a more effective loss of angular momentum through the wind. The result is pure catastrophe: the stars spiral in towards each other, until they ultimately collide and merge. A similar process may play a role in the creation of hot Jupiters now being seen in many exoplanetary systems.

The magnetic field that leads to solar and stellar activity is generated in a process that we refer to as the dynamo. The dynamo feeds on convective motions, from which a tangled, rapidly evolving magnetic field is generated. If the star rotates, the convective motions have a preferred sense of motion owing to the Coriolis forces. This is thought to order the field into larger scales, both spatially and temporally. Activity like that on the Sun occurs on all stars that rotate, and that have a convective envelope immediately below their surface. As a rule, the more rapidly the star rotates, the stronger the large-scale dynamo is. There are significant differences from star to star, however, even for stars that are of nearly solar mass, age, and rotation rate.

1.1.1.2 Impact of Solar Activity on Humanity

The Sun's activity plays an important part in our lives. Despite the fact that it does this in ways that overtly affect our everyday routines but little, both the cumulative and the extraordinary effects can be upsetting. The recognition of the importance of these events has led to the development of a large National Space Weather Architecture. Within that Architecture, NASA, and in particular the Sun-Solar-System Division is devoting substantial resources to understanding why and exactly how Earth and humanity are affected by solar activity. This is the focus of NASA's *Living With A Star* program, which aims to develop understanding and mitigation strategies for our society that relies increasingly on stable food supplies and habitats and on resources placed in space, as well as for astronauts traveling beyond Earth's protective magnetosphere. At the core of that problem is the Sun's activity itself: what causes the Sun to be magnetically active, and how can we begin to develop reliable forecasting tools for its activity and the associated space weather and climate changes on Earth? We are beginning to understand the very complicated chain of processes that links the deep-seated solar dynamo to Earth (Fig. 1.4).

Earth Climate

Global warming is real and the magnitude larger than feared only a decade ago (a finding of the WMO/UNEP Intergovernmental Panel on Climate Change in its 2001 report; see http://www.ipcc.ch/). Exactly what the link is between solar variability and climate change continues to be debated. This debate centers on one of the most complicated systems that humans are striving to understand at present. The conclusions of the studies are often ambiguous, compounded by the fact that many of the processes involved are known with inadequate accuracy, if they are not simply ignored.

Over 2,000 climate studies find a significant fingerprint of the long-term solar variability in climate records over the last century. Understanding the nature of this Sun-climate coupling is important, because apparently the solar variations induce a natural background of climate change against which anthropogenic change must be calibrated.

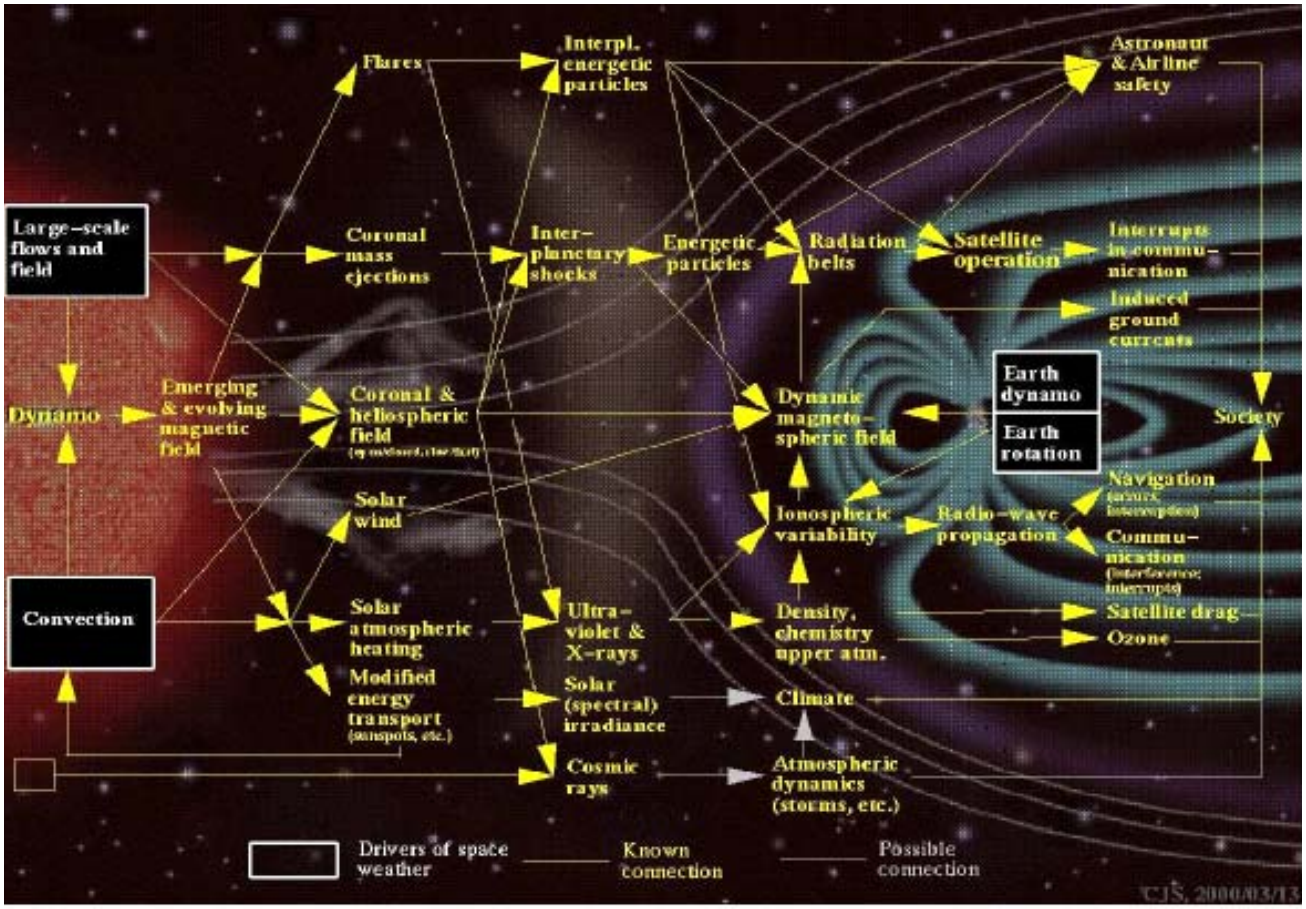


Figure 1.4: Space weather: the solar magnetic field, and coronal mass ejections traveling through it, impact the Earth's magnetic field. This induces changes that propagate throughout geospace. Radiation travels other pathways into the Earth's ionosphere and troposphere.

Detailed climate modeling suggests that the Sun's cyclic irradiance variations even of the order of a few tenths of a percent are strong enough to be a significant driver behind the Earth's climate changes. In addition, there are other ways in which the Sun's variations can couple to the Earth's tropospheric state. For example, the solar cycle modulates the interplanetary magnetic field. The cosmic-ray flux that reaches Earth through this magnetic shield changes as a consequence of this. It has been suggested that cosmic rays induce cloud formation. Changes in cloud cover, and the related atmospheric processes and interaction with radiation, are among the most effective processes driving climatic change. Clouds are so effective in this because they affect the Earth's albedo, which determines how much solar energy can reach into the atmosphere. Moreover, they are storehouses for water vapor, one of the most important greenhouse gases next to CO_2 and CH_4, which helps determine how much energy Earth can radiate away. Unfortunately, the effects of clouds are among the most difficult to model accurately.

Another channel by which the Sun can reach into the Earth's climate system is through the modulation of its EUV and UV brightness. These modulations range from a factor of a few to a hundred, depending on wavelength. These fluxes affect the chemistry of the upper Earth atmosphere, including the ozone balance; ozone is another important greenhouse gas, as well as a shield against

solar UV damage to humans and animals. How the consequences of that chemical balance propagate through the Earth's climate system also remains uncertain. The IPCC report points out that "inclusion of spectrally resolved solar irradiance variations and solar-induced stratospheric ozone changes may improve the realism of model simulations of the impact of solar variability on climate."

Modeling aside, the strongest evidence to date that the solar cycle is coupled with Earth climate on the long term comes from the undeniable empirical finding of the correspondence of the Little Ice Age with the Maunder (and Spörer) Minimum, and of a medieval warm period with the solar Grand Maximum (Fig. ES-2). Another interesting finding is that the Earth's albedo, as measured by the brightness of the dark side of the Moon, may have increased somewhat in a five-year period leading up to the most recent solar maximum (Goode *et al.,* 2001). It is remarkable how little is known about the Earth's albedo and its variations, given that it is a key measure for the atmospheric energy budget. There is an interesting measurement for another planet, however: it has been reported that the albedo of Neptune varies by 4% in antiphase with the ~ 0.2% change in solar irradiance throughout the sunspot cycle (Lockwood *et al.,* 1991), suggesting that there are mechanisms that amplify or exceed the gentle solar irradiance variations.

Even though we do not at present understand the details of these couplings, the correspondence between significant changes in solar activity and Earth climate warrant further study. These studies need to include a better understanding of the solar dynamo mechanism.

Space Weather

The solar wind induces aurorae at high latitudes (Fig. 1.5). These pretty spectacles are caused by the same phenomena that frequently lead to costly or dangerous damage.

Astronauts, for example, who would leave the Earth's magnetosphere, may be exposed to radiation doses that can cause significant damage to their organs, if not kill them en-route through interplanetary space if no protective measures are taken.

Many satellites, particularly those at geostationary heights, must function in environments where energetic particle densities as a result of solar activity frequently show a sudden increase by many orders of magnitude. The most energetic of these particles penetrate deep into spacecraft, where they lead to dielectric charging. Subsequent discharges can damage spacecraft components, or even cause entire spacecraft to fail. Less energetic particles induce detrimental surface charging, and cause solar cells to decay by as much as 3% per solar event.

When these energetic particles penetrate closer to Earth, they can cause dangerous exposure levels for astronauts in low-Earth orbit, and even dangerous life-time dosages for airline personnel. Avoiding strong dosages can require postponing or interrupting extravehicular activities of astronauts, or changing routes for airlines.

Figure 1.5: Most of the consequences of space weather are invisible from Earth.

The aurorae are the exception. They are formed by electrical currents that flow above the magnetic poles, induced by the solar magnetic field that is carried through the heliosphere in the solar wind or during coronal mass ejections.

High-frequency radio communication is restricted to much narrower bandwidths during space-weather events triggered by solar activity. During a Maunder-like minimum in solar activity, there would probably not be enough of an ionosphere around the Earth to enable any HF communications over the horizon. At the frequencies used in satellite communications, solar-induced electron-density variations modulate signal paths and strengths, leading to errors in global positioning and image distortion with synthetic aperture radars.

Changes in the solar EUV and UV brightness causes the Earth's atmosphere to expand or contract, modulating satellite drag (Fig. 1.6). This leads to orbital decay, and limits the lifetime of low-flying satellites. Sometimes, a rapid increase in orbital drag during particularly sensitive times can cause a satellite to tumble and control to be lost permanently (such as happened to the ASCA satellite in July 2000, following a major solar event on the 14th of that month).

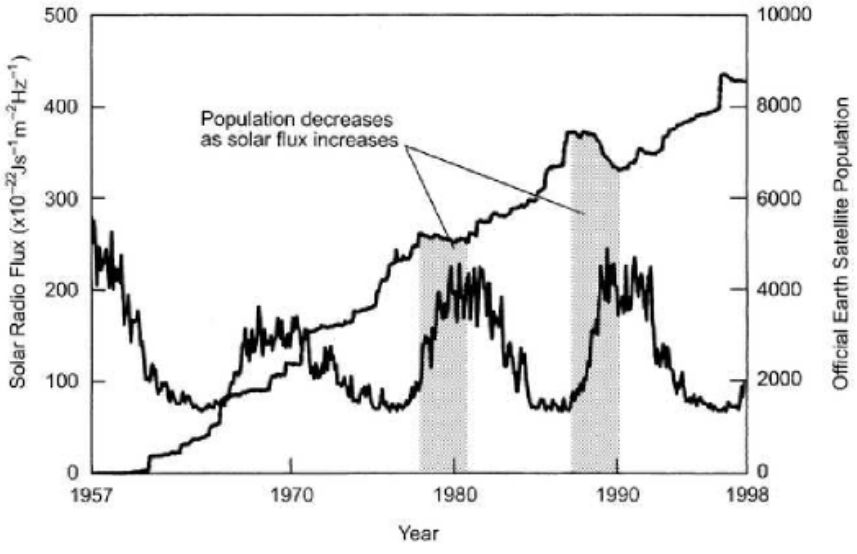


Figure 1.6: The 10.7 cm solar radio flux compared to the total number of spacecraft or spacecraft debris larger than 1 meter in diameter.

Notice that every rise in solar activity corresponds to a decrease in the number of satellites as a result of increased orbital drag through an expanded Earth atmosphere.

Geomagnetic storms triggered by solar activity also induce substantial voltages on long conductors on Earth, such as pipelines and power grids, and--in days gone by--in telegraph networks (which, by the way, contributed substantially to the discovery of space-weather effects on the Earth's magnetosphere). Pipelines may corrode faster because of this. Power grids may fail if the solar event is large enough or if the coupling into the Earth's geomagnetic field is particularly efficient. The best-documented case was a blackout of the entire province of Quebec in 1989, with an estimated economic cost of approximately 5 billion dollars. These and other consequences of space weather are summarized in Table 1.1. Better, more reliable forecasts of solar variability are envisioned to increase safety and reduce cost in all of these areas. *Study of stellar analogs with SI is key to understanding this solar magnetic variability*.

Table 1.1: Space weather affects society in many ways

Weather and climate change
Errors in, and outages of, navigation systems
Astronaut and airline passenger radiation hazard
Satellite operations (upsets, charging, failures)
Satellite orbits and launch trajectories
Interference on monitoring systems
Induced voltages on, and outages of, power systems
Corrosion of long-distance pipelines
Interruption, or failure, of satellite communications
Interruption of HF air and ground communications
UV radiation and ozone barrier
Systematic uncertainties in carbon dating

1.1.1.3 Solar-Stellar Magnetic Activity and the Evolution of Life

Some 4.5 billion years ago, our star was formed in the center of a slowly spinning disk of glowing gas out of which the planets also condensed. The first tens of millions of years, the Sun was magnetically extremely active, with sometimes gigantic starspots on its surface, likely even some covering its polar regions. During the first tens of millions of years, sunspots may have covered up to half of a visible hemisphere, subjecting the young planets to significant modulations in the solar irradiance with a periodicity of only a few days.

Following that early period of activity, the Sun settled into its long stable life as an ordinary, yellow star, with its magnetic activity gradually subsiding. Eventually, the average X-ray brightness declined to its current value of only ~0.1% of the value for the young Sun. In step with this decline, the frequency and magnitude of flares and mass ejections decreased.

In the coming ~5 billion years, solar activity is forecast to continue to subside, while the Sun slightly increases in size and total brightness. Then it will run out of the hydrogen that fuels the fusion reactor in its core. For a while, it will find new ways of nuclear fusion, as it swells up to an orange giant. Within the following few million years, the Sun will grow into a huge red giant, one hundred times its present size. In that phase, only a handful of huge convection cells will cover its surface, which will be radially undulating by millions of miles as the convective flows evolve; the Sun will generally look oddly distorted from the average spherical state. This relatively brief stage heralds the end of the Earth, which will be cooked in the intense radiation, and most likely be engulfed by its star. After that, the Sun will shrink hundredfold, relative to its original size, to become a white dwarf star, being only somewhat larger than Earth. Its brightness will continue to fade as energy, no longer replenished by nuclear reactions, leaks from the star. Activity vs. age for stars observed in clusters is shown in Fig. 1.7.

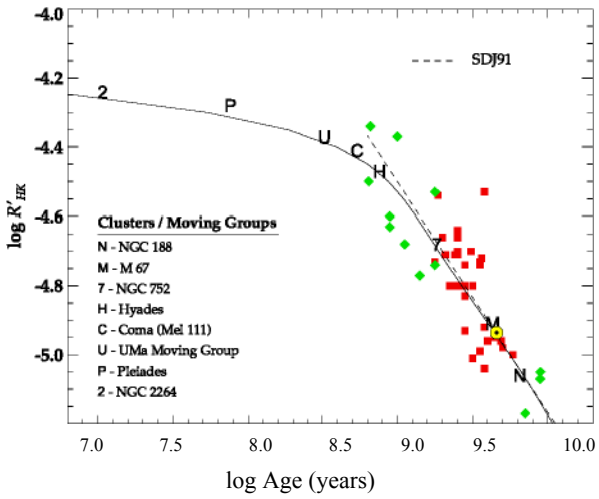


Figure 1.7: Activity versus age for stars in clusters:

The vertical axis is a logarithmic measure for the brightness of stellar chromospheres, the horizontal axis is the logarithm of the estimated age of the star (in years). The Sun is marked as ⊙. (From Baliunas *et al.* (1998).)

How does the stellar activity affect the planets during the star's mature phase? The warming light of the star initially aids the chemical purification of the atmosphere: hydrogen and helium, the most common and lightest elements in the universe, evaporate, leaving the heavier components behind. The residual atmosphere is more and more dominated by molecules such as nitrogen, methane, ammonia, etc. The energetic X-ray and ultraviolet light generated by stellar magnetic activity penetrates into the planetary atmosphere because there is no protective ozone layer yet. This causes dissociation of the molecules there; if that happens relatively high in the atmosphere, the recoil of the constituent atoms contributes to the ablation of the atmosphere. For any planet that does not have a strong magnetic field of its own, like Mars, the solar wind and the electromagnetic phenomena embedded within it, strip away the atmosphere much more readily than in the case of the Earth which is relatively protected with its magnetic cocoon.

Deeper in the atmosphere, the radiation caused by the star's magnetic activity induces reactions in the mixture of nitrogen, methane, water, ammonia, and residual hydrogen. Thus form amino acids, sugars, purines and pyrimidines. The latter two may combine with sugar and naturally occurring phosphate to form nucleotides. Amino acids are the building block of proteins, while nucleotides make up DNA and RNA. These complex molecules are the foundation of all life known thus far. The reactions driven by the solar energetic radiation may have significantly outweighed those induced by Earth's own processes like lightning. It remains to be established how these processes compare to, or combine with, those that may have been energized by other means, say by the high temperatures around deep hydrothermal vents near the deep rifts on the ocean floors.

The geological records show that there was liquid water on Earth even in the first billion years of its existence. Yet, the young Sun should have been some 25% fainter than the current mature Sun according to stellar-evolution models. By any current understanding, this should have resulted in a much cooler Earth, covered in ice. Why this did not occur remains a mystery. When seen in combination with the much more recent occurrences of Sun-induced climatic change like the Little Ice Age that appear to be driven by changes of the solar brightness of only a small fraction of a percent, we must conclude that apparently the Earth's tropospheric system can either dampen or amplify the variations of the Sun that occur on a range of time scales.

In addition to the slow evolutionary change of the solar luminosity, the young planetary atmosphere was subjected to large irradiance variations caused by the magnetic activity of the young star for a few hundred million years as starspots came and went on the rotating Sun. If the mild activity of the Sun

today can already induce climate changes, we wonder what effects short-term variations of ten or twenty times the current magnitude would have.

As the Sun celebrated its billionth birthday, its activity had subsided significantly. The spot coverage decreased, and the relative importance of smaller magnetic elements increased. Such small elements are actually somewhat brighter than the non-magnetic stellar surface (which earned them the name of faculae when they were discovered in the 19th Century). At this age, dark spots and bright faculae compensated almost exactly. But although there were only very weak irradiance variations for a long time, the variable magnetic activity continued with its associated energetic radiation and space weather.

Eventually, as activity subsided even more, the irregular activity variations made way for regular cyclic variations on the time scale of a decade, now known as the sunspot cycle. Ever fewer spots dotted the solar surface, and the balance shifted even more towards the many faculae. Irradiance and activity started to vary in phase causing the Sun to be brightest when most spots occur, as is still the case.

1.1.1.4 Magnetic Activity and the Formation of Planetary Systems

In the initial phases of the contraction of the gaseous cloud out of which the Sun and the planets formed (Fig. 1.8), magnetic fields were instrumental in transporting most of the angular momentum out of the core regions into the outer domains of the cloud. Without the expulsion of the bulk of that initial angular momentum, no star could have formed because centrifugal forces would have exceeded the pull of the gravitational field.

Once a proto-star and a surrounding extended gaseous disk form, stellar magnetic activity truly begins to manifest its significance. In the first 10 million years, the star's strong magnetic field couples to the inner domains of the gaseous disk that surrounds a star before any planets form. The star continues to accumulate mass from the disk for some time, but the disk cannot extend to the stellar surface: the magnetic field of the rapidly rotating star sweeps through the inner disk region, allowing matter to accrete only along the field, as no gas can move perpendicular to it. Accretion thus occurs in evolving columns connecting the disk to patches on the stellar surface.

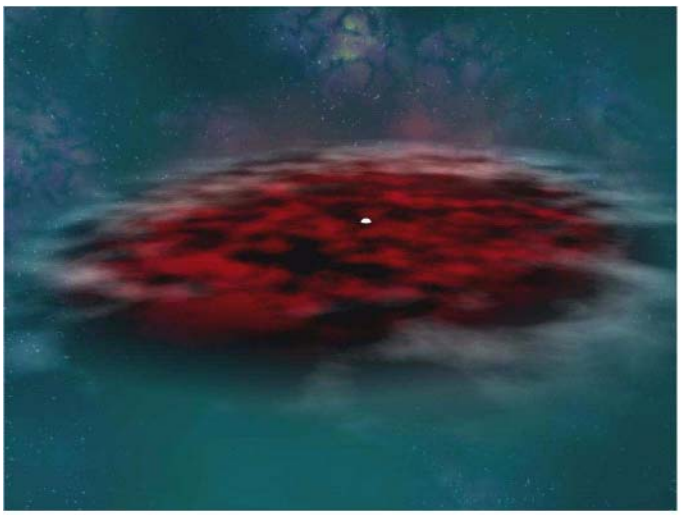


Figure 1.8: All stars form from clouds of gas and dust that collapse under their own gravity; the rotation of the cloud forces an accretion disk to form.

Even as the star continues to grow, lumps form in the disk which will ultimately become planets. As matter accumulates onto the star and its planets, a growing hole forms near the star, until ultimately a fully formed solar system like our own is revealed. (artist's concept, from MIRLIN/JPL)

Where the material cannot reach the surface, the gas pressure aided by the magnetic field channels the material into jets of gas, shooting away from the poles of the star, perpendicular to the disk (Fig.

1.9). More material (perhaps as much as a tenth of the gas reaching the star through the disk) is blown away in a magnetically-powered stellar wind (Fig. 1.10). At the same time, the field acts on the gas in the inner part of the disk, accelerating it to above the orbital velocity, and thus driving it away from the star. This precarious balance, in which matter moves toward and away from the star at different locations, affects the disk's stability, mass content and evolution. That, in turn, likely affects how many planets form, where they form, and what their masses and atmospheres are. The details of this largely hypothetical scenario, and the precise role of the magnetic field in it, remain shrouded, because present-day telescopes cannot image the inner part of such disks around very young stars.

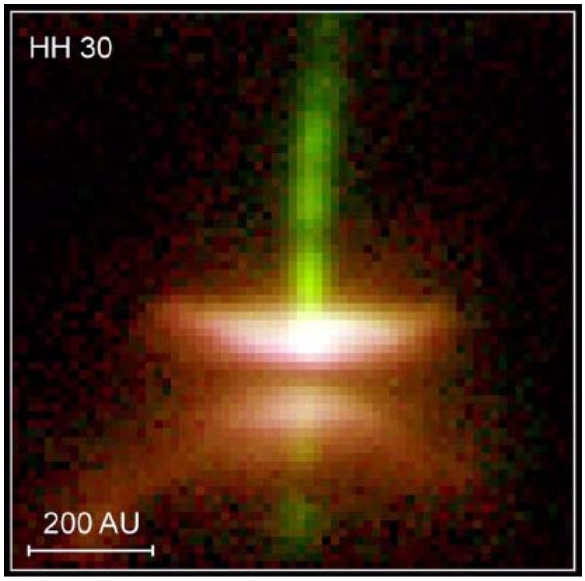


Figure 1.9: This Hubble Space Telescope image shows Herbig-Haro 30, the prototype of a young star surrounded by a thin, dark disk and emitting powerful gaseous jets.

The light of the hidden central star reflects off the upper and lower surfaces of the disk to produce the pair of reddish nebulae. The gas jets are shown in green. The horizontal bar indicates the size of our solar system. (From NASA's Observatorium)

In order to understand the formation of planetary systems, we need to study the balance of angular momentum loss by magnetic braking, storage of angular momentum in stellar binary systems, and the distribution of angular momentum over the central star and any planets.

Figure 1.10: Model of a Young Stellar System.

For single stars, it remains unclear what processes lead to the observed strong decrease in angular momentum in the initial star-forming processes that is inferred from the observations. It is of particular importance to learn what role planetary systems play in this, because these are vast potential reservoirs of angular momentum; Jupiter's orbital angular momentum, for example, is approximately 100,000 times the rotational angular momentum of the present-day Sun. It is interesting that similar stars with and without known pre-planetary disks or planetary systems do not appear to differ significantly in their coronal activity or rotation rates: does the existence of a pre-planetary disk or planetary system not matter to the rotational history of stars despite the disk-star coupling that is invoked to understand how a star could form in the first place? Answering that question requires that we image the dynamics of the innermost regions of your stars that still have pre-planetary disks. Also, asteroseismic determination of the internal rotation for a range of different stars will provide important constraints on the so far poorly understood

establishment of the slow rotation of the deep solar interior, resulting from the spindown of the solar envelope, and the consequences of the associated processes for solar structure and evolution.

1.1.1.5 A Primer on Dynamo Theory

The variable magnetic field that causes stellar activity is generated by a process which we call the dynamo. A dynamo involves the conversion of kinetic energy in convection and large-scale circulations into magnetic energy. This dynamo action eventually leads to the injection of magnetic flux into stellar photospheres. There, the field evolves on times scales that are very short compared to any of the time scales for stellar evolution or for large-scale resistive dissipation of magnetic fields; stellar activity introduces a new domain of time scales for stars, ranging from fractions of a second during flares to decades for cyclic changes.

There is at present no quantitative model for stellar dynamos that is useful to forecast solar activity based on some observed state of the Sun, or even to establish the mean activity level of a star based on, say, its mass, age, and rotation rate. The nonlinear differential equations for the coupling of the vectors of turbulent convection and magnetic field cannot be solved analytically. Nor can the cycle dynamo be simulated numerically in its entirety because of the high magnetic Reynolds number R_m in the solar convection zone (Note: the magnetic Reynolds number is the ratio of the time it takes for the magnetic field to diffuse away from a concentration of flux relative to the time it would take large-scale flows to concentrate it there, and thus measures how fine field structures can be relative to the flows). At such high R_m, the magnetic field is highly intermittent, with length scales down to ~1 km. Since the volume of the solar convection zone is $\sim 10^{18}$ km^3, full numerical coverage would require 10^{18} grid points, which is a factor of order a billion beyond present computational means. Hence, both analytical and numerical studies necessarily make approximations that simplify or ignore much of the physics. Interestingly, even the approximating models are of a richness and diversity that there is no consensus on the model properties, or even on the set of processes that are important in driving the dynamo.

All dynamo models rely on three fundamental operations that plasma flows perform on the embedded magnetic field: (1) stretch, (2) turn, and (3) transport. If these processes occur locally, as must invariably happen in any turbulent convection, a so-called turbulent or small-scale dynamo action occurs. In order to generate the larger scales required for a global dynamo, each of the above three processes needs to be associated with a comparably large scale.

The scientific community concurs that it is the differential rotation within or just below a convective envelope (Fig. 1.11) that takes care of the large-scale *stretching* (commonly referred to as the Ω effect). It remains unclear, however, whether it is the radial shear within the overshoot layer, or the latitudinal shear there and within the envelope itself that is most effective in this. Resolving this requires observational constraints derived for stars in which the differential rotation profile differs from that of the Sun. The internal differential rotation is accessible only through asteroseismology by observing modes of relatively short surface wavelength, inaccessible to unresolved stellar observations. The surface component of the differential rotation can be measured by repeated imaging of the activity patterns on a star.

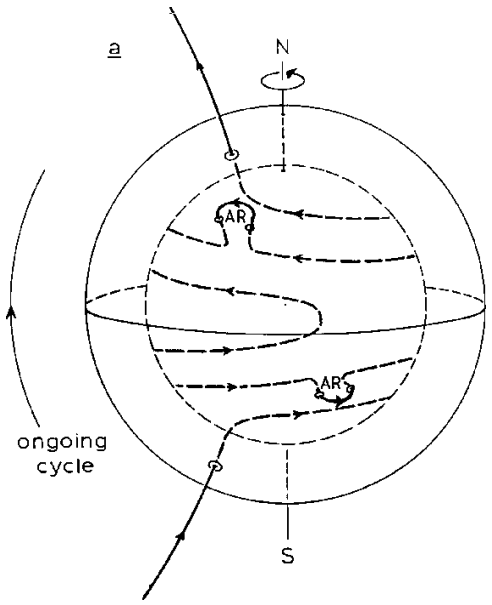


Figure 1.11: Sketch of the nearly toroidal magnetic flux system thought to exist near the bottom of the convective envelope.

Differential rotation stretches field in the E–W direction. When a loop of flux tubes rises to the surface, an-shaped loop breaks through the photosphere, forming a bipolar active region, tilted somewhat relative to the equator because of the Coriolis force acting in the rotating Sun. (From Zwaan, 1996)

The large-scale *turning* may be caused by one or more processes that operate in distinct domains within the stellar interior. The most common assumption is that the turning is caused by the rotationally-induced Coriolis force. Other means to redirect field include the flux emergence and dispersal processes, helical waves within the overshoot layer, and others still have been proposed. The first of these alternative α effects is accessible to observational scrutiny by the Stellar Imager, while others will require detailed numerical studies to be guided by knowledge of the large-scale flows in the stellar interior and at the surface (Fig. 1.12).

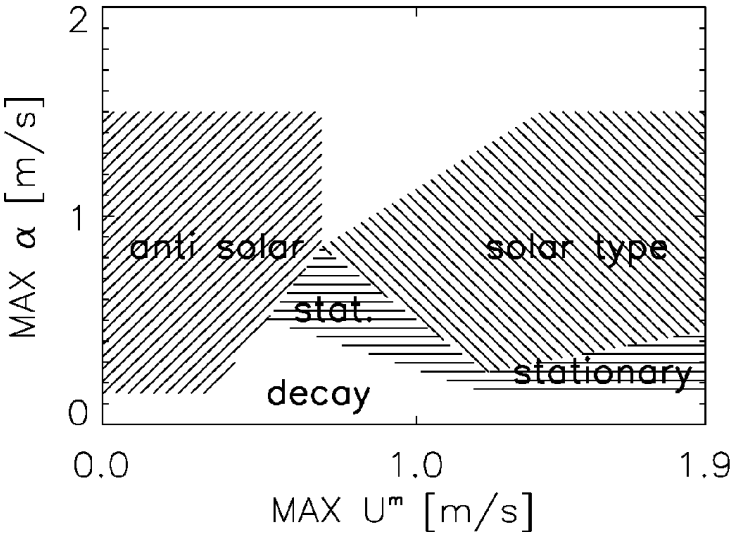


Figure 1.12: Butterfly pattern, meridional circulation, and α effect:

the diagram shows, for one model of an envelope dynamo, where the sunspot migration patterns are equatorward like that of the Sun, poleward, or stationary, as function of the magnitude U^m of the meridional flow at the base of the convective envelope, and the rotation-induced α effect. (From Küker et al., 2001)

The most generally recognized and studied process for the large-scale *transport* of flux is that of turbulent diffusion, or random-walk transport. The large-scale convection can transport fields throughout the convective envelope as well as the layer just below it into which this convection overshoots. Other large-scale flows, up to the scale of the meridional circulation, have only been included in some models in the last decade. These flows can cause fundamentally different dynamo

patterns depending on magnitude, direction, and extent within the convective envelope (e.g., Dikpati and Gilman, 2001; Küker *et al.,* 2001). In addition to this, buoyant rise and downward convective pumping within the rotating star may also play a role that is being explored only in very recent work (e.g., Dorch and Nordlund, 2001). Imaging of surface patterns allows observational determination of the surface meridional flow and supergranular flux dispersal, both key inputs to any dynamo model to be developed.

The time scale of the activity is also an important clue for the processes involved in driving the dynamo. For example, in the traditional dynamo of the so-called $\alpha\Omega$ type, time scales for large-scale convective flux dispersal and field-amplification by differential rotation together set the cycle period. In other dynamo models the time scale for meridional circulation determines the cycle period. In a star like the Sun, with a well-defined cycle, the cycle period can, of course, be determined even by ground-based, non-imaging observations in appropriate diagnostics. But such a pronounced cycle is relatively rare, perhaps in part because multiple subsequent cycles are present at any given time, as is the case for the Sun for some two years in each cycle, or because of the nonlinear chaos intrinsic to the dynamo problem. Tests of dynamo theories are in need not only of period measurements, but also particularly of the data with which these can be compared, namely properties of meridional flow, large-scale flux dispersal, and differential rotation. These can be measured only by some form of stellar imaging for stars of Sun-like activity levels (Section 1.3).

During the first decades of research into stellar dynamos, one of the key difficulties was to find any model dynamo that would have a flux-emergence (or butterfly) pattern that migrated towards the equator, as on the Sun (Fig. 1.3). More recent studies show that such solutions can be found for particular combinations of parameters (e.g., Fig. 1.13). By exploring parameter space, these early studies already demonstrated that a dynamo can exhibit widely different spatio-temporal patterns depending on type and magnitude of the processes that are included (e.g., Markiel and Thomas, 1999, Dikpati and Gilman, 2001, Tobias *et al.,* 2001, and references therein). Perturbations on these parameters can cause pattern changes with time; such perturbations include excursions about the mean meridional flow, torsional oscillations in the differential rotation, or fluctuations about the average α effect (e.g., Ossendrijver and Hoyng, 1997).

Modeling efforts suggest that dynamos may have either a dominant dipolar or quadrupolar component, reflected in patterns that are symmetric or antisymmetric about the equator. Other modes, that show up with particular choices of, e.g. the latitude dependence of differential rotation and the α effect, may show a triple butterfly pattern moving towards the equator, or two pronounced wings migrating in opposite directions away from mid-latitudes, or even a polar flux-emergence pattern at a significantly different period than the low-latitude emergence pattern, or phase differences as large as a few years between opposite hemispheres (e.g., Markiel and Thomas, 1999, Dikpati and Gilman, 2001; see Fig. 1.13).

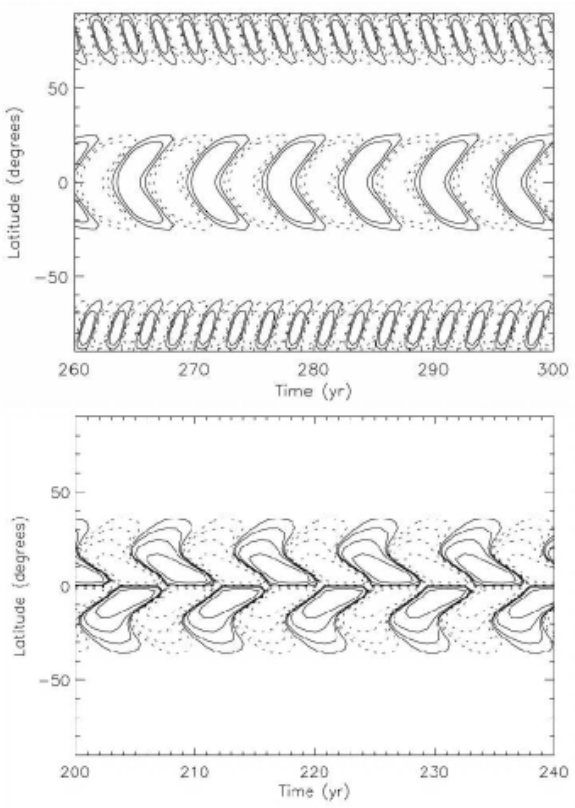


Figure 1.13: Contour maps of model magnetic fields at the bottom of the convective envelope in a Sun-like star, showing a poleward migrating pattern at low latitudes and a higher-frequency polar cycle (top) and a Sun-like butterfly diagram (bottom). These models differ only in the latitude dependence of the field-turning α effect. (From Markiel and Thomas, 1999)

Most dynamo models that were developed until now are kinematic. That is to say that the presence of the field is assumed not to affect the plasma flows. Such models are linear, and cannot be used to draw conclusions about the magnitude of the field. A self-consistent model would need to include the back-reaction of the field on the flows. Whereas some such models have been tested with ad-hoc parameterizations of the interaction of the flow and the field, none can at present be developed from first principles, and observational constraints are thus essential: comparisons of dynamo patterns as observed on stars for different activity levels with those resulting from modeling will guide dynamo research and help validate candidate dynamo models.

Numerical research will undoubtedly make significant advances in the coming years, but only the comparative analysis of many Sun-like stars with a range of activity levels, masses, and evolutionary stages will allow adequate tests of complex dynamo models, validation of any detailed dynamo model, and exploration of the possible spatio-temporal patterns of the nonlinear dynamo.

Our knowledge of the stellar internal stratification stands in stark contrast to that of the large-scale convection, differential rotation and other large-scale flows. There is at present no model that reproduces the internal differential rotation profile of the convective envelope in detail. Our understanding of the meridional circulation is particularly limited: it may be that this circulation is driven by the rotationally induced asymmetry in convective motions (e.g., Kitchatinov and Rüdiger, 1995), but there is no empirical test for that as the Sun, the only currently available test case, shows it only near the surface. And even at the surface it is hard to characterize: the meridional flow may be equatorward at low latitudes and even at the poles; it may vary with time; it may not be axially symmetric. As to the third dimension, we have no constraints on the depth extent of the circulation, or whether it is more than a single large eddy with depth (e.g., Hathaway, 1996, Schou and Bogart, 1998,

Braun and Fan, 1998, Basu *et al.,* 1999, Gonzalez Hernandez *et al.,* 1999). Although far from impressive in its magnitude, this weak meridional flow likely has profound consequences for the dynamo (as discussed below). Given the state of our knowledge, the question whether the dynamo field affects the magnitude or pattern of this weak flow is unanswerable until we have a sample of stars on which the flow can be inferred, if only at their surfaces (Fig. 1.14).

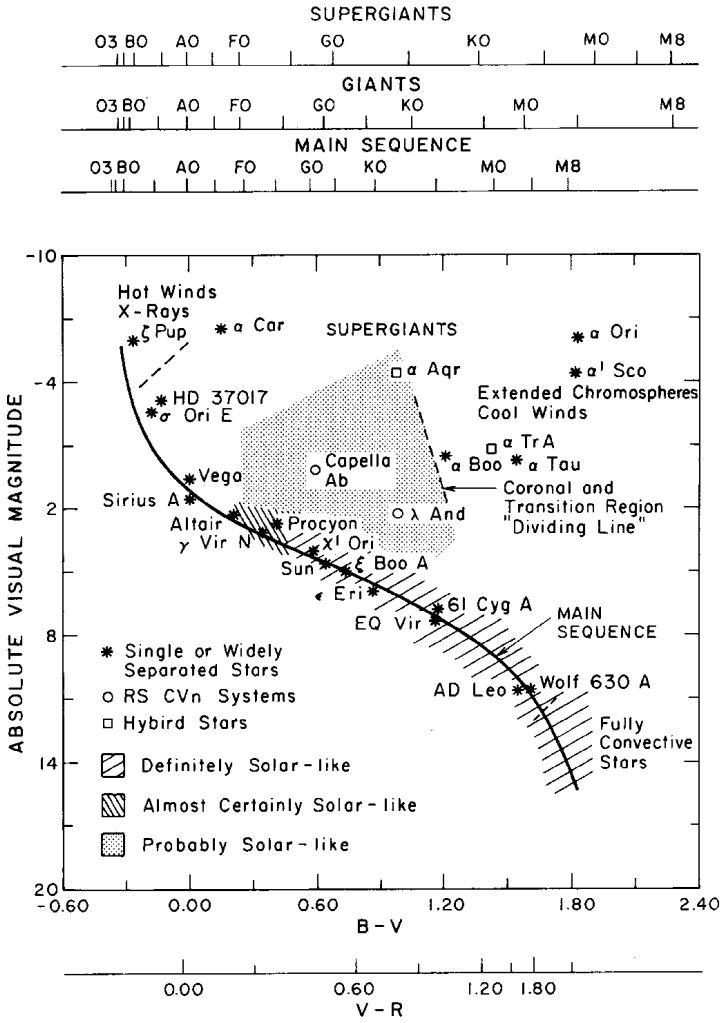


Figure 1.14: An H–R diagram showing stars with magnetic activity, which in the original paper (from Linsky, 1986) are distinguished in groups of solar-likeness. Also indicated is the region where massive winds occur and where magnetically-confined hot coronal plasma is likely absent. Some frequently studied stars (both magnetically active and nonactive) are identified by name.

The situation is a little better for the differential rotation, because here we do have access to at least the magnitudes of that rotation for a limited category of stars even now. This is possible, because the analysis of the disk-integrated signal of the Sun-as-a-star confirms that the rotation period that is inferred depends on the latitude at which active regions preferentially emerge. Thus, as the sunspot cycle progresses, the rotation period changes with the mean active latitude. Stellar astronomers have assumed that the spread in observed rotation periods for a given star over the years is a measure of at least the magnitude of the surface differential rotation of a star, although with the caveat that there is no clear matching of measured rotation period with the rotation period of the instantaneous corresponding mean emergence latitude (e.g., Schrijver, 1996). In principle, for stars for which frequent Doppler images are available in a series of seasons over the years, an even better constraint on differential rotation may be achieved, but that can be done for only very few, very active stars.

The surprising result of efforts to estimate stellar differential rotation is that the differential rotation, when expressed as the time it takes for pole and equator to differ by one full rotation, is about the same for most of the main-sequence stars, whether single or in close binary systems. There is some evidence that this profile may be different for the somewhat heavier F-type stars with shallow convective envelopes, and for some of the evolved binary systems (see references in Schrijver and Zwaan, 2000).

1.1.1.6 Empirical constraints on the magnetic dynamo

As current stellar observations do not allow us to measure differential rotation or meridional circulation in stars, it will come as no surprise that we know essentially nothing about stellar butterfly diagrams or - on even smaller scales - about the spectrum of emerging bipoles, the statistical properties of their orientations and latitudes, and their clustering and nesting properties for any star other than the Sun. Only imaging of stellar surfaces to a scale of better than ~50,000 km can provide the necessary information to begin to explore these properties.

The studies of average activity levels of stars have at least helped us piece together what some of the essential ingredients to dynamo action are on the largest scales. For example, we know that a dynamo associated with stellar activity operates in all rotating stars with a convection zone directly beneath the photosphere. In single stars, the dynamo strength varies smoothly, and mostly monotonically, with rotation rate, at least down to the intrinsic scatter associated with stellar variability. It also depends on some other stellar property or properties. For main sequence stars, for example, the primary factor in determining activity resembles the convective turnover time scale at the bottom of the convective envelope (Fig. 1.15). But no such dependence holds if we test the relationship on evolved stars (e.g., Basri, 1987), so that apparently other, as yet unidentified parameters, play a role, such as surface gravity (Jordan and Montesino 1991).

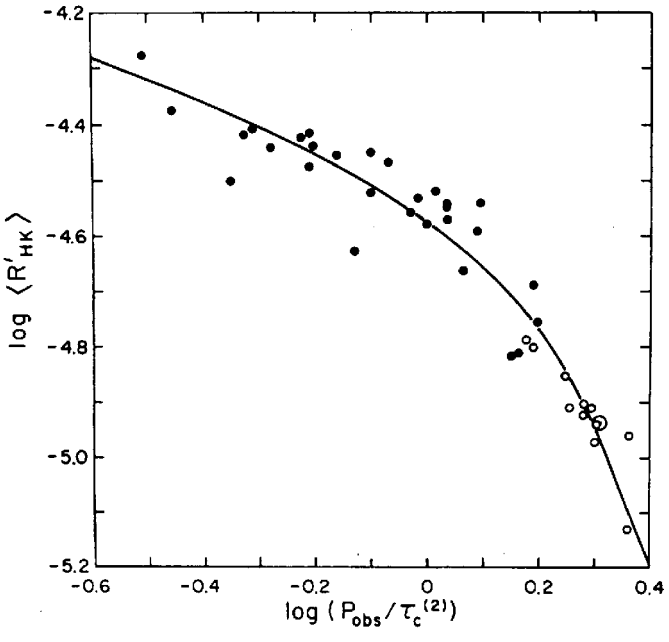


Figure 1.15: Ca II emission strength vs. rotation period & convective turnover time. Ratio R'_{HK} of the chromospheric Ca II H+K emission over a nonmagnetic background to the bolometric emission of main–sequence stars versus the ratio of the stellar rotation period to essentially the convective turnover time $\tau_c^{(2)}$at the bottom of the convective envelope. Open circles indicate stars of relatively low activity; these less active stars predominantly show regular cycles, whereas more active stars often have irregular cycles or noncyclic behavior. (After Noyes et al., 1984).

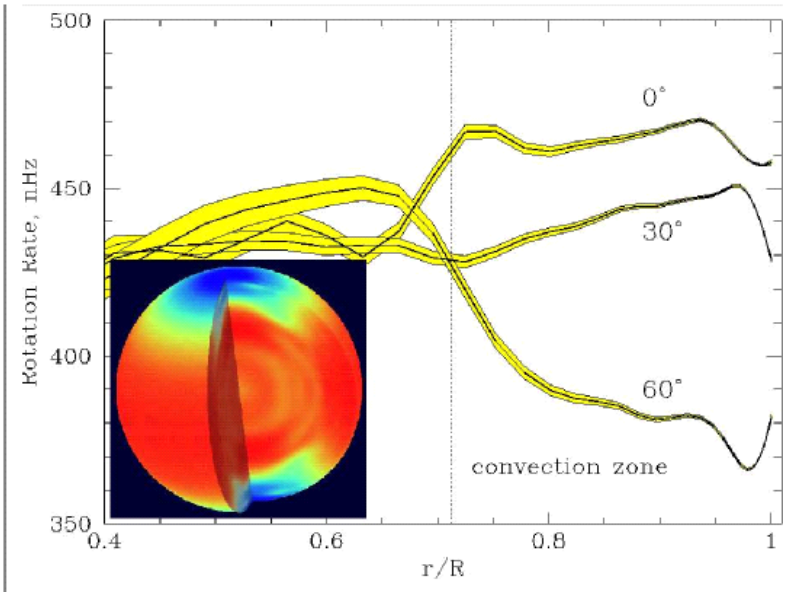


Figure 1.16: Internal rotation of the Sun; the image shows the rotation rate coded from blue (slow) to red (fast). The solar core rotates nearly as a rigid body, but the convective envelope rotates faster at the equator than at the poles. (courtesy SOHO/MDI team)

Whereas almost all models for the solar dynamo emphasize the important role of the overshoot layer in the dynamo process we must also acknowledge that stars that are deemed to be fully convective also are magnetically active, without any clear change at the mass where stars become fully convective. This means that an overshoot layer is clearly not essential for an effective dynamo to function. A dynamo in a fully convective star may not be like the solar dynamo in its patterns, however; at least the quasi-regular variability of the solar cycle seems to have no counterpart on these stars. The comparison of empirically determined flux patterns on partially and fully convective stars will help us understand the role of the overshoot layer in the solar dynamo.

The Sun shows a relatively regular heartbeat with its 11-year sunspot cycle, even though cycle strength and duration are modulated. Such a pattern is not the rule among the cool main–sequence stars; we find a variety of patterns of variability in their activity, while only a minority of these stars show cyclic variations that resemble that of the Sun (Figs. 1.17 and 1.18). For main–sequence stars with moderate to low rotation rates, activity tends to be cyclic, but no clear trend of cycle period with stellar parameters has been found, although there are hints of relationships between cycle period, rotation period, and the time scale for deep convection (e.g., Saar and Brandenburg, 1999). For truly active stars, various variability patterns exist, but generally no unambiguous activity cycle is seen. In this context, it is interesting to add that the moderately and very active stars never reach low states of activity, in contrast to the Sun during its cycle minima. Perhaps there is no butterfly diagram in these active stars, or perhaps multiple sunspot cycles are present on the surface (i.e., a more extreme overlap of successive cycles than seen for the solar cycle).

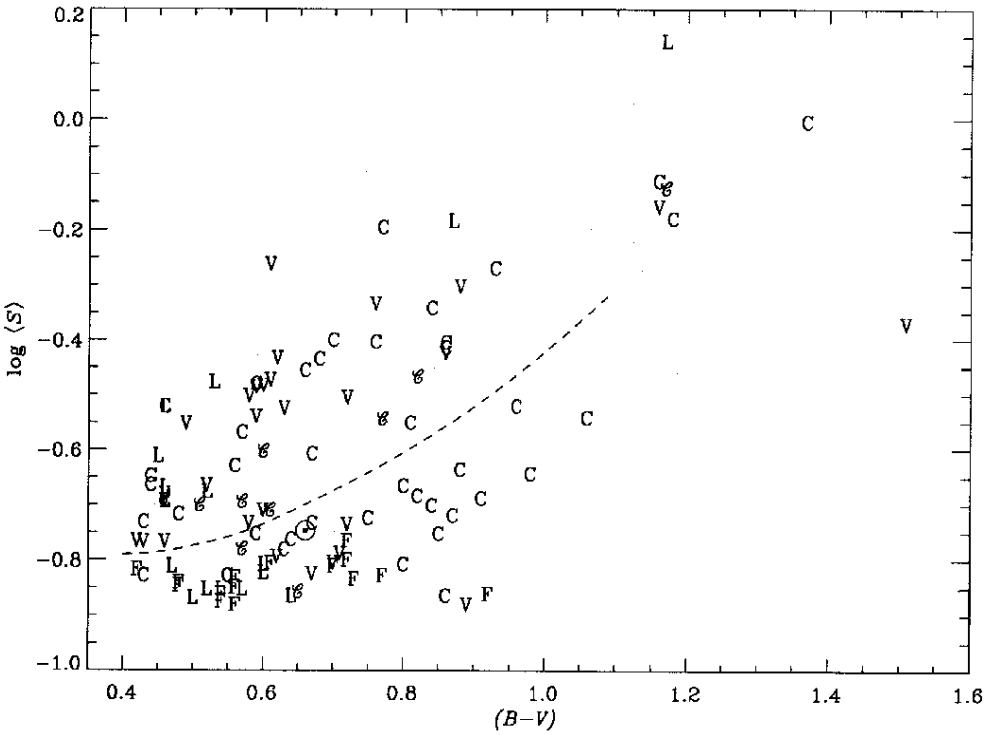


Figure 1.17: A diagram of the mean Ca II H+K emission index ⟨S⟩ against spectral color B–V shows that stars like the Sun (⊙) mostly do not show an 11-year heartbeat. The symbols show the long-term variability (determined by measurement of Ca II fluxes vs. time): C, cycle period determined; *C*, secondary period; L, long-term trend; V, variable without clear period; F, flat; ⊙, Sun. The dashed curve separates the very active from less active stars (from Baliunas et al., 1995).

Activity does not continue to increase as one observes stars with increasingly high rotation rates. The atmospheric radiative losses appear to "saturate" for stars with rotation periods below about one or two days, but it remains unclear whether this reflects a saturation *(i)* in the dynamo action through the backreaction of the field on the flows, *(ii)* in the heating mechanism because no more energy can be channeled into the atmosphere from the (sub)photospheric convection, or *(iii)* in the photospheric flux content through increased annihilation of opposite polarities in the overfilled photosphere. Establishing which of these actually occurs will help us parameterize the couplings of field and flow in the dynamo.

Not only the internal properties of a star determine its activity: the presence of a companion star enhances the dynamo action in cool components of tidally interacting binaries not only by enforcing rapid rotation through tidal interaction, but presumably also by affecting the pattern of differential rotation. There lies another interesting field of study for the Stellar Imager. How does a dynamo function when the differential rotation is subjected by strong tidal interaction? Why is the differential rotation not obviously different in tidally interacting binaries compared to single stars? What is the cause of the persistent preferred longitudes for activity (often at the quadrature points) in these stellar systems?

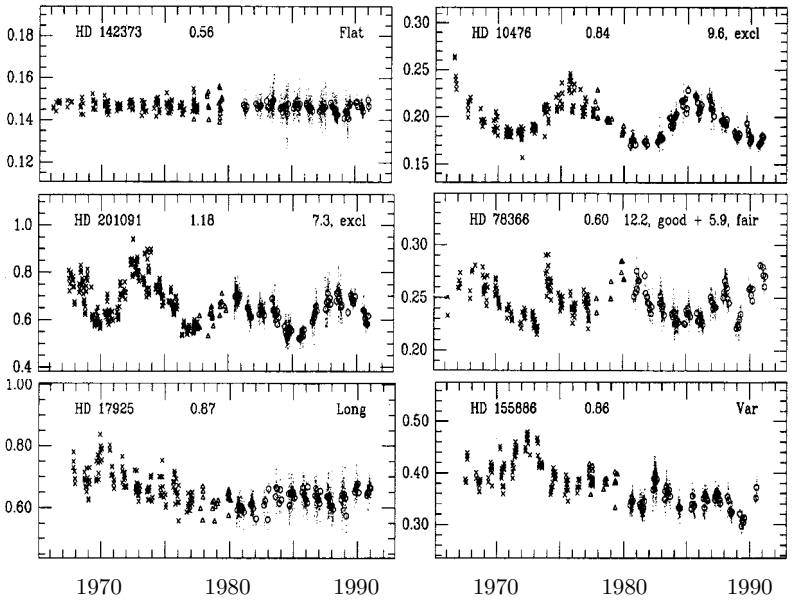


Figure 1.18: Records of the relative Ca II H+K fluxes of main-sequence stars:

x, Wilson's records (1966–1977); triangles and dots, Ca II HK survey (1977–1992); open circles are 30-day averages. The top of each panel shows the stellar identification, color index B –V , and a classification of the long-term variability or period(s) in case of cyclic activity (figure from Baliunas et al., 1995)

Most rapidly rotating stars exhibit polar starspots (e.g., Fig 1.20). Never in recorded history has the Sun had a starspot above 60 degrees latitude (compare Fig. 1.3). How do such unusual concentrations of field form at stellar rotational poles? There are now two very different models for this process: (*i*) they may form because field emerges at high latitudes if not at the poles, or (ii) flux may emerge at mid and low latitudes to be advected towards the poles, and there concentrated in starspots. The first explanation relies on the strong Coriolis force that acts on flux bundles that rise from deep within a star. This force is expected to deflect rising flux bundles towards latitudes of up to 60° (Fig. 1.19). Truly polar spots could result either from flux eruption originating very deep in stars with relatively small radiative interiors, or by a poleward slip of the previously anchored deep segments of flux rings following an earlier eruption of flux at mid latitudes elsewhere on that ring. If this is the correct explanation, the butterfly diagrams on these stars should look very different, with not only much flux emerging at high latitudes, but also a zone of avoidance near the stellar equator (Fig. 1.20).

A recent model for the dispersal of flux within the photosphere has shown that there is a viable alternative explanation for the formation of polar starspots: polar caps would form by advection of preferentially the trailing flux in active regions, as happens in the formation of the dipole component of the solar field, but increased activity leads to an increase in flux as well as closer packing of that flux because of the interaction between flux and convective dispersal. Are polar spots a signature of modified flux emergence, or a consequence of simple surface advection of flux emerging elsewhere? Only stellar imaging can distinguish between these two very different explanations.

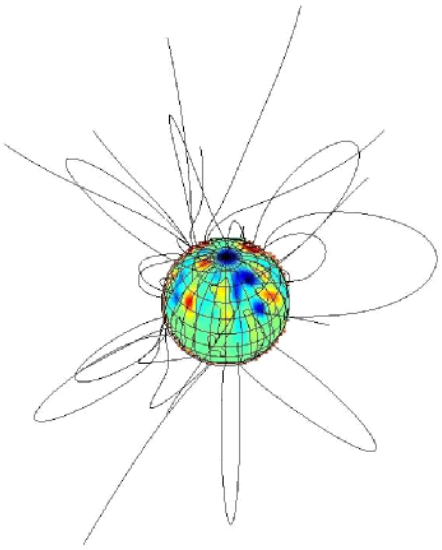


Figure 1.19: Magnetic fields on stars are very difficult to measure, but on some very rapidly rotating stars they are so strong it can be done at least coarsely using Zeeman–Doppler imaging.

This image shows the magnetic field for AB Doradus. The surface field has been extrapolated assuming it to be a magnetic potential field, shown here by the lines, to obtain an impression of the star's coronal field configuration. (Courtesy J.-F. Donati)

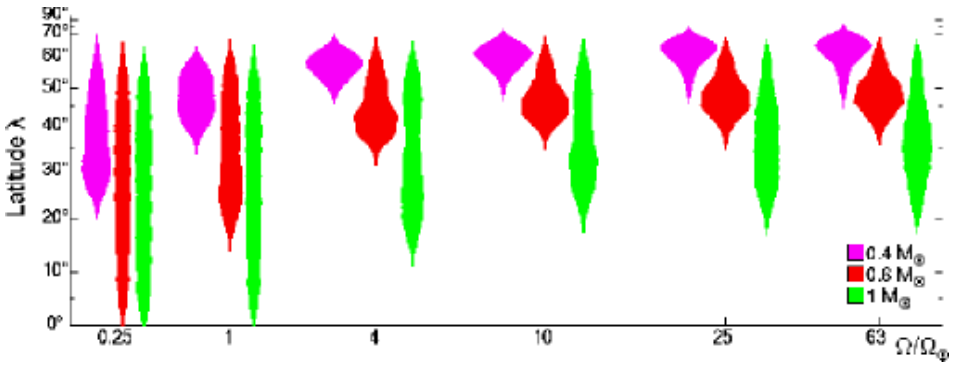


Figure 1.20: Model for latitude distribution of emerging field for different stars, as a function of rotation rate.

Increased rotation and deeper convective envelopes are expected to result in higher latitudes for emergence. (by Th. Granzer)

1.1.1.7 Seismic Studies of Stellar Interiors

Helioseismology has given us an extremely detailed view of the solar interior, thanks to the analysis of frequencies of a very large number of different acoustic modes, spanning degrees between 0 and several thousand. The large range in degree allows resolution of all layers of the Sun, from the core to the near-surface region, and provides fairly detailed resolution in latitude at least in the outer 40 % of the solar radius. These results are of great importance to our understanding of the structure and evolution of stars, and of the physical properties and processes that control this evolution. At the time of the launch of the SI, seismic investigations of other stars will have been undertaken by several space missions, including MOST and COROT. However, these will only observe low-degree modes, through intensity variations in light integrated over the stellar disks. Such point-source observations will provide information about the global properties of solar-like stars, which allows the study of global structure, including, e.g., gravitational settling of helium and large-scale mixing processes. SI observations, however, allow us to go far beyond that: modes of degree as high as 60 should be reachable with an array of N=10 elements, increasing as N^2 for larger arrays. By analogy with the Sun, in solar-like stars this will allow inferences with good radial and reasonable latitude resolution to be made in the radiative interior and the lower part of the convective envelope, for both structure and the patterns and magnitudes of the differential rotation with depth and latitude. With a careful choice of

target stars SI observations will allow us to obtain such detailed information about the interiors of stars over a broad range of stellar parameters, in terms of mass, age and composition.

Studies of the internal rotation as a function of mass and age, and related to the activity studies carried out with SI, will provide unique information about the evolution of stellar internal rotation with age, in response to the activity-driven angular-momentum loss in stellar winds. This will provide stringent constraints on models of the rotational evolution, elucidating the processes responsible for transport of angular momentum in stellar interiors; these studies are also fundamental to the understanding of the dynamo processes likely responsible for stellar activity. By correlating the rotation profile with the profile of the helium abundance, as reflected in the seismically inferred sound speed, an understanding can be achieved of the rotationally-driven mixing processes in stellar interiors. This is of great importance for calibrating the primordial abundances in the universe as well as to the improvement and validation of stellar evolution models. For example, the data will provide constraints on the convective overshoot at the base of the convective envelope which also contributes to the mixing. The resulting understanding can then be applied to the mixing and destruction of lithium, finally providing the means to relate the observed lithium abundance in old halo stars to the primordial lithium content of the Universe. For stars slightly more massive than the Sun the data, combined with the more extensive data on low-degree modes likely available at the time from earlier missions, will allow detailed investigations of the properties of convective cores and related internal mixing; an understanding of these processes is essential to the modeling of the evolution of massive stars, leading to the formation of supernovae.

1.1.2 Discover the Universe at High Angular Resolution

Observing the Universe at ultra-high angular resolutions will enable a fundamental understanding of magnetic processes, the origin and maintenance of structure, and the transport of matter throughout the Universe. We highlight here with only a few examples the vast discovery potential of the Stellar Imager.

1.1.2.1 Stellar Magnetism and Dynamo Extremes

A stellar dynamo involves many processes whose interactions are currently poorly understood. These include internal rotation profiles, coupling of magnetic fields and plasma flows, pulsations, and even the basic nature of stellar convection. Testing of dynamo theories will greatly benefit from the study of a sample of stars with stellar parameters that are much like those of the Sun, as we have argued in the preceding sections. However, to understand dynamo evolution and the exceptional environments in which dynamos operate, or fail to operate, it is also necessary to study a sample of astrophysical sources that are quite different from the Sun. Magnetic interaction in close binary systems plays a key role in any population study, especially since close interacting binaries often contain rapidly rotating stars and thus provide a testbed for studies of dynamo-driven stellar activity and stellar surface structure.

The SI will observe a sample of stars with a range of stellar masses, effective temperatures, abundances, and depths of the convective envelopes, including main-sequence dwarf stars, giant stars, stars in the process of forming, and stars approaching their death. This will provide a database to test the evolution of stellar magnetism from star forming regions to dying stars.

Star Formation, Protoplanetary Disks, and Jets

Protoplanetary disks are where planets form, migrate to their final locations, and where the materials that can ultimately produce life-bearing worlds are assembled. For our own Solar System, the first 30 Myr spans the formation and evolution of the proto-Solar nebula, the assembly of the meteorite parent bodies, the formation of the proto-Earth and proto-Mars, and the early phases of the Era of Heavy Bombardment. If we are to understand not only the history of our Solar System, but also how planetary systems develop in general, we need to understand the disks, how long they last, how they interact with their central stars, and how they evolve.

Young stellar objects (YSOs), e.g. T Tauri stars, represent the parent stars of planetary systems presumed to form from the remnant circumstellar disks that encircle them. The inner boundaries of such disks are expected to be at the corotation radius from the star, typically 3-5 stellar radii (the inner disk boundaries might explain the observed pile-up of Jupiter-like planets at orbital periods of 3-4 days). The environment within that distance is controlled by the strong magnetic field of the rapidly spinning star. The temperatures of the accreting plasma increase from several thousand to a few million degrees in this region. Due to the high temperatures and relatively low densities, UV emission provides an efficient and direct means to image the regions close to YSOs. Spatially resolved fluorescent H_2 emission has been detected for the Herbig Ae star HD 100546 with HST (Grady et al. 2005). This indicates that SI should be able to resolve, for example, at least the inner 0.05 AU of disks in the Taurus-Auriga young stellar association region, which is about 160 pc away. Figure 1.21 shows a simulation of an SI observation of the Lyα-fluoresced H_2 emission originating in the magnetosphere-disk interaction region of the star TW Hya.

Determining the size of the field-dominated region is of great importance to understanding several fundamental questions about stellar rotational breaking, accretion rates as a function of global disk parameters, and orbits for (giant) planets. But SI can provide much more than merely the size of the region: The past decade has seen the identification of mature planetary systems, and the first routine imaging of disks around young stars, as well as the recognition that a number of 8-50 Myr old stellar associations lie within 100 pc of the Sun. At d<150 pc protoplanetary disks have been detected around 0.7-2.5 solar mass stars. This gives us a pool of well-studied targets for detailed imaging by Stellar Imager. For objects with the larger cavities, repeated observations by SI may reveal rotation of resonances and indirectly point to the location of planets and direct detection of planets may be possible via their UV auroral emissions.

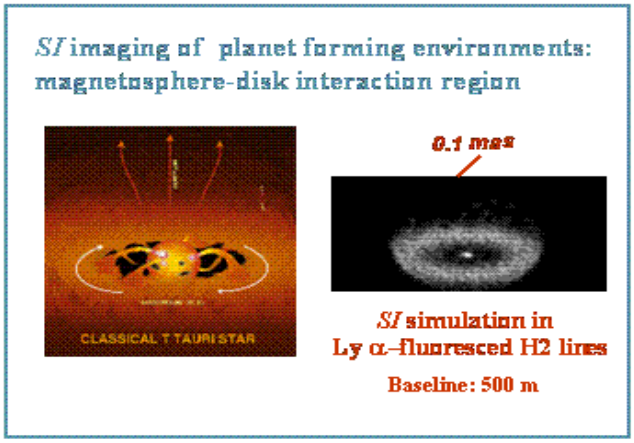


Figure 1.21: A simulation of an SI observation of the Lyα-fluoresced H2 emission originating in the magnetosphere-disk interaction region of TW Hya.

Common red-shifted absorption features have been interpreted (Hartigan et al. 1995) as the signature of accretion onto high latitudes. Magnetic fields are thought to be critical in this process, with the protoplanetary disk extending into the corotation radius with the star (Hartmann 2004), and with the accreting material funneled along magnetic field lines to high latitudes on the photosphere. Stellar Imager will have the necessary spatial resolution to verify this inferred geometry, and to directly image the accretion hot spot(s). With the discovery of FUV excesses (Grady et al. 2004, 2005) and magnetic field detections for Herbig Ae stars (Hubrig et al. 2004), this process can be extended to at least 2.5 Solar masses. With the detection of disks and Hα emission suggestive of activity similar to T Tauri stars, jets may also be present in brown dwarfs (Maschiadri & Raga 2004), suggesting that similar accretion activity may extend well below the sub-stellar limit. The Stellar Imager would have the capability to map the accretion flow from the co-rotation radius of the disk onto the accretion footprint of the star, using emission lines spanning a wide ionization range.

Young stars frequently drive bipolar outflows that can be traced, in some cases, on parsec scales (McGroarty et al. 2004a,b). For nearby stars, these outflows can be resolved from the central star in H Lyα and optical forbidden lines, and in some cases in Si III 1206.5 Å (Devine et al. 2000; Grady et al. 2004, 2005). Comparison of the disk inclination data obtained with HST and the outflow data from the same images suggests that the bipolar outflows are seen orthogonal to the disk. This is unexpected for outflows that are solely following the stellar magnetic field geometry, since the stellar-rotation and global-magnetic axes are not perfectly aligned in the Sun. One possibility is that the outflow launch and collimation region extends over the inner few tens of AU of the disk (Woitas et al. 2005). Stellar Imager can easily resolve this region, at the distances of the nearest star formation regions, and will not only provide data on the wind geometry in the outflow, but will also allow us to resolve any uncollimated wind component. Such a wind component has been proposed as a means of transporting annealed silicates and processed organics from the inner parts of the protoplanetary disk into more distant icy planetesimals, thus accounting for the compositional diversity of comet nuclei (Shu et al. 2001).

Very Cool Stars, Giant Stars, and Supergiant Stars

Stars that are at least 1.5 times heavier than the Sun are not magnetically active during their mature life on the main sequence because they lack envelope convection. Consequently, they begin their transformation to red giant stars with essentially the same rotational energy they had after their initial formative epochs. As they begin to expand, a dynamo is activated once the star cools enough to develop envelope convection. That may lead to significant, sudden magnetic braking, which possibly results in a substantial difference between the rotation rates of the deep interior and the magnetically-active convective envelope (e.g., Schrijver and Pols, 1993). Observations indicate that such a difference may last for up to some tens of millions of years. Detailed understanding of the onset of dynamos in evolving stars with such shear layers between envelope and interior, and of the possible consequences for the internal dynamics, will greatly benefit from imaging and seismic observations of stars in such evolutionary phases.

Continuing their evolution as red giants, the stars then reach a point where the coronal activity disappears again, to be replaced by substantial mass loss at much lower temperatures. In a Hertzsprung-Russell (HR) diagram this behavior occurs on either side of a dividing line. Even though there is an absence of magnetically heated transition-region and coronal plasma in the late-K and M-type giant stars, their winds are thought to be driven by magneto-hydrodynamic waves. Rosner et al. (1995) have proposed that the coronal dividing line is a consequence of a dynamo transition: large-scale structures with closed field lines and coronal heating, and small-scale structures with open field lines and increased mass loss. The hybrid stars that display both phenomena are the key to understanding the dividing line and the associated change in the dynamo mode from global to local. Imaging both transition-region and chromospheric emission, the Stellar Imager will reveal the magnetic field topology on stars on both sides of the dividing line, and the hybrid-atmosphere stars.

As stars continue to expand to supergiant stages, the scale of the surface convection changes to the point that we expect only a few convective 'granules' to cover the entire star. Figure 1.22 shows a model and simulated SI observation (assuming a 500 m baseline and a distance of the star of 2 Kpc) of this convection. Does this really happen? Some doubt it because the spectral lines of these stars show little sign of such large-scale turbulence. And if it does, then a turbulent local dynamo may again create magnetic fields on a near-global scale. The Stellar Imager can image both the large-scale convection (and its evolution) and possible chromospheric patterns driven by this process.

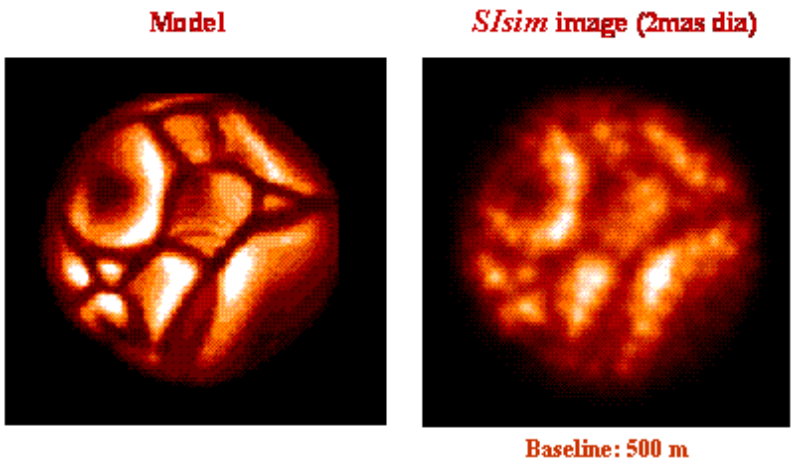


Figure 1.22: Model and simulated SI observation of the convection on a supergiant like Betelgeuse at 2 kpc.

These "granules" transport the energy from the interior to the surface, evolving on a timescale of years, with only a dozen of so granules filling the entire surface (model courtesy B. Freytag).

Hot stars: Rotation, Structure, Winds, and Disks

There are many competitive processes on stars that produce structures on the stellar surface or in the circumstellar environment. These processes include radiative winds, rapid rotation, pulsations, mass motions, and magnetic fields, many of which may operate simultaneously within the envelope of the same star.

Understanding how massive stars rotate is important for the accurate modeling of stellar evolution and computing the final chemical yields of stars (see, e.g., Maeder & Meynet 2000a). Hot (O, B, Wolf-Rayet) stars tend to be the most rapidly rotating types of stars (excluding the degenerate white-dwarf and neutron stars), and many are rotating so fast that their shapes are centrifugally distorted into oblate spheroids. Although rapid rotation in the very rare eclipsing binaries is measurable using light curves and radial velocity profiles, it is extremely difficult to pin down the detailed properties of single-star rapid rotation. With 10 to 1000 interferometric resolution elements across a stellar disk, though, several new methods of measuring hot-star rotation become possible:

Obviously, the rotation rate can be measured directly (along with any differential rotation) by tracking features that move across the star at different latitudes. There is ample evidence for rotationally modulated spots in several cases (e.g., Balona & Engelbrecht 1986; Smith et al. 1994; Reiners et al. 2000). However, stellar oblateness is sensitive to the rotation of the entire stellar interior (thus possibly giving us a better measure of the star's total angular momentum than feature-tracking alone could provide).

An important but seldom directly measured aspect of hot-star rotation is the phenomenon of *gravity darkening* (see Figure 1.23). A distorted gaseous star in radiative and hydrostatic equilibrium exhibits a change in its net radiative flux which is proportional to the local "effective gravity" over its surface (i.e., gravity + centrifugal acceleration; von Zeipel 1924). Thus, the equators of rapidly rotating stars become dimmer and cooler than their poles. For eclipsing binaries, gravity darkening has been determined by its inclusion in the list of optimization parameters that are varied in order to produce agreement with multi-color light curves. For single stars, the phenomenon has been modeled extensively, but only measured indirectly by, e.g., the comparison of absorption lines formed at different latitudes (e.g., Stoeckley & Buscombe 1987). High-resolution imaging in the UV and optical would constrain how much gravity darkening actually exists for different types of stars far better than, e.g., lower-resolution, ground-based optical interferometric measurements. We could then assess how it gradually disappears as subsurface convection eventually sets in later than the early/mid-F spectral range. Current models are still evolving (e.g., Claret 2004), and observational constraints from eclipsing binary light curves sometimes yield types of gravity darkening that are outside the bounds of present theoretical understanding (Djurasevic et al. 2003).

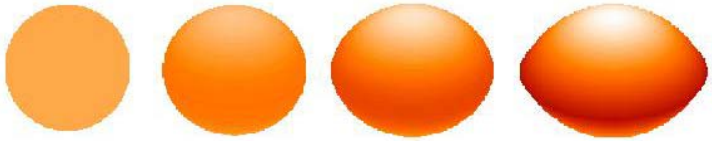


Fig. 1.23: Gravity Darkening in Rapidly Rotating Stars:
Color representation of the bolometric flux emergent from a B2 main-sequence star rotating at 0, 60%, 80%, and 99% of its critical "breakup" rotation speed (left to right) and obeying von Zeipel (1924) gravity darkening.

Broadly speaking, continuum-bandpass filters can be used efficiently to search for thermal and diffusive inhomogeneities on a hot star's disk. Most other processes are best studied by imaging in spectral lines. The best single line available from the ground is Hα, which can be used to map matter both nearby and at moderate distances from the star. In the UV, the CIV doublet can be employed to study inner winds and co-orbiting structures of hot stars. The MgII doublet can be used to trace the discrete ejections of mass and the extent of disks out to several stellar radii.

For O stars and early B supergiants, radiative winds generally dominate over other mass-loss processes. These winds can be optically thick and thus resolvable in high mass loss stars such as Wolf-Rayet and interacting massive binaries. In principle, the structure of these winds provides a means to document the past ejections of shells in stars with histories of discrete mass loss episodes. Imaging winds would help us understand the density distribution and, from the continuity equation, outflow velocities in the inner wind. Imaging lines in sufficient detail could tell us whether wind X-ray exciting clumps form, and if the winds are anisotropic. Anisotropies are important because they hint at a partial confinement of the wind by rotation or magnetic fields. Anisotropic outflows (polar ejections and Shu's X-wind) have been imaged by HST in continuum wavelengths. However, the SI will be able to image the flows with much finer resolution, closer to the star and to the protoplanetary accretion disk.

Proceeding toward the B stars, the efficacy of winds decreases rapidly, but they still play a key role in the development of circumstellar structures in the ejection of matter into extended equatorial disks and co-rotating magnetospheres. Classical Be stars are ostensibly-single, post-Zero Age Main Sequence (ZAMS) stars which episodically eject mass. The ejecta may collapse quickly onto an equatorial disk.

Be stars are rapidly rotating, non-supergiant B-type stars that exhibit emission in the hydrogen Balmer lines. The observed properties of Be stars and their circumstellar gas are consistent with the coexistence of a dense equatorial disk and a variable stellar wind (for a recent review, see Porter & Rivinius 2003; see also Figure 1.24). The gas in the so-called `decretion disk' is generally believed to be ejected from the star and not accreted from an external source, and the rapid rotation of Be stars has been associated with the presence of the disk since at least the 1930s. One of the longest-standing puzzles in hot-star astrophysics is the physical origin of this disk, both from the standpoint of mass supply (the winds may be too tenuous) and from the standpoint of angular momentum supply (the disks are Keplerian but the stellar surfaces are not). Also, there are many examples of stars that have exhibited alternating Be and "B-normal" phases of activity (the latter implying disappearance of the disk), with time scales of various kinds of variability ranging from days to possibly centuries.

Fig. 1.24. Artist's conception of the circumstellar environment of the Be star in the φ Per binary system [FOV 6 x 3 mas] (Image credit: Bill Pounds, STScI-PRC1997-39).

Direct imaging of Be stars with the Stellar Imager will provide answers regarding the physical distribution of matter, structures within the disks and winds (spiral density waves or clumpy structures), wind/disk interaction regions, and ionization structure. The sizes of disks as a function of underlying stellar parameters could be studied. The C IV and Mg II imaging would provide views of regions of the circumstellar envelopes which may be anisotropic but ordered, a finding which has important ramifications for magnetic ejection mechanisms of the matter. Temporal monitoring can shed light on questions of the stability and variability of disks. Comparisons between the disks of the classical (main sequence) Be stars and the Herbig (pre-main sequence) Ae/Be (HAeBe) stars could be carried out directly, and characteristics of HAeBe disks could be compared to those of lower mass pre-main-sequence stars.

Optical interferometry has begun to probe the broad-brush properties of a few nearby Be-star disks (e.g., Quirrenbach et al. 1997), but the low resolution only barely provides the ellipticity of the inclined disk and any large-scale inhomogeneity. Milliarcsecond resolution would allow excellent characterization of the mean disk properties (e.g., inclination, radial density structure, and thickness) as a function of spectral type and stellar rotation rate, which would provide stringent empirical constraints on the currently bewildering number of proposed theories (see below).

Also, the high-resolution imaging of non-axisymmetric structures in Be-star disks would allow a conclusive determination of what gives rise to the well-known V/R (violet/red) variability in double-peaked Hα emission lines. These variability patterns are especially puzzling because they do not rotate with the disk material, but seem to take decades to precess around the star (for possible interpretations, see Okazaki 1991; Savonije 1998). For the bright Be star 7 Cas, a combination of ground-based interferometry and kinematic data from the Hα line have been used to map the precession of a supposed "one-armed" density perturbation over the last several years (Berio et al. 1999), but the features are far from well-resolved.

With SI, the much-improved set of empirically derived star and disk plasma parameters would allow us to choose handily between the major proposed scenarios of disk formation (see Bjorkman 2000 for a summary of the various ideas).

Wolf-Rayet (WR) stars are believed to be the central, heavy cores of evolved O-type stars that have lost most of their hydrogen-rich outer layers as a stellar wind. WR stars have observed mass loss rates at least an order of magnitude higher than other O stars (i.e., of order 10^{-4} M_{sun}/yr), and the origin of these extremely dense and optically thick outflows is still not well understood. The only way that line-driven wind theory can account for such large mass loss rates is if the opacity in the lines is utilized many times (i.e., if photons multiply scatter through the optically thick outer atmosphere before they give up all of their radiative momentum to the gas); see, e.g., Gayley et al. (1995). However, other ideas exist, including fast magnetic rotation (Ignace et al. 1998) and "strange-mode" pulsations in the chemically enriched interiors (Glatzel et al. 1993). The direct imaging of the innermost emitting surface in the wind would lead to stringent constraints on these ideas.

WR winds are observed to contain dense "clumps" with a range of scales suggesting turbulence (e.g., Lepine & Moffat 1999). The existence of these clumps has recently been found to greatly complicate the study of line-driven mass loss (Brown et al. 2004a), with the general conclusion that the presence of optically thick clumps makes it much more difficult to understand how WR winds are accelerated by radiation forces. Clearly higher resolution observations are needed in order to identify the origin and nature of these clumps, and, by inference, help clarify the physical processes responsible for the dense winds as a whole.

Some WR winds also exhibit larger-scale variability similar to the "photospheric connection" variations. Even some of the most basic properties of some stars are not understood. For EZ CMa (St.-

Louis et al. 1995; Georgiev et al. 1999), it is still not known if the reasonably regular variability is driven by rotation, a binary companion, or magnetic fields.

Luminous Blue Variables (LBVs) are evolved, extremely luminous, unstable supergiant stars that undergo large-amplitude wind variability. The prototypical LBV is η Car, which experienced remarkable luminosity variations between 1830 and 1860 that we now interpret as an episodic mass loss of 1-2 solar masses (see Davidson & Humphreys 1997). LBVs pose some of the key questions in massive-star astrophysics: What causes the outbursts? What is the specific internal disturbance, and what sets its time scale? Is the observed bipolarity a result of rapid rotation, a binary phenomenon, or both? (Questions paraphrased from Barbossa & Figer 2004).

LBV stars are known to be very near the so-called *Eddington limit* in the H-R Diagram, where the outward radiative acceleration associated with free-electron scattering equals the inward acceleration of gravity. Although the potential for strong outbursts in such situations is clear, it is still not known precisely how line-driven winds (Owocki et al. 2004) and rapid rotation (Maeder & Meynet 2000b) contribute specifically to LBV mass loss. Figure 1.25 illustrates recent observations of η Car that indicate the gas in the inner wind is elongated along the supposed rotation axis (van Boekel et al. 2003).

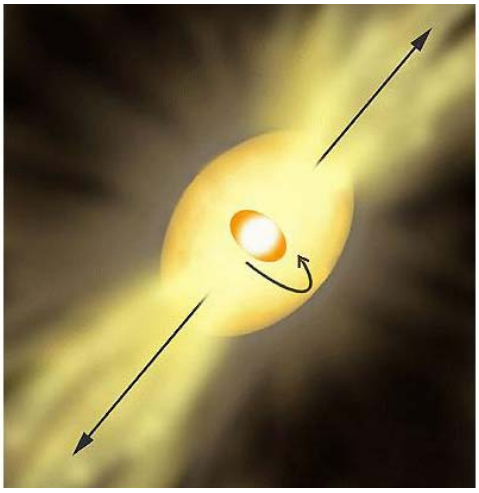


Fig. 1.25. Artist's conception of the prolate inner wind surface of η Car as imaged by the VINCI instrument on the VLT. The central star is ~5x8mas in size and well-suited to SI resolution.
(Image credit: van Boekel et al. 2003; see also ESO Press Release 31/03, PR Photo 32b/03).

1.1.2.2 Pulsation: Stellar Structure and Mass Loss

Atmospheres of Pulsating Variable Stars

Pulsations are found in many different types of stars, ranging from very hot main-sequence stars to dying cool giants and stellar relics. The signatures of pulsation are very prominent in the UV (e.g. Mg H&K lines) and hence of great interest for SI observations. In many cases stellar pulsations, radial or non-radial, significantly affect the extent, composition, and structure of stellar atmospheres. The SI will have a unique capability of direct imaging of pulsation effects including surface structures and shock fronts as they propagate through the dynamical atmospheres. Imaging the effects of the pulsation will provide key inputs to hydrodynamical models for a range of diverse pulsators, including radial pulsators such as Miras and Cepheids, as well as non-radial pulsators such as β Cephei stars.

Nonradial pulsations (NRP's), which are observed spectroscopically on B stars, produce evenly-spaced temperature modulations that can be imaged as bright and dark zones around the star. Surface thermal modulations amplify wind flows into clumps which are apt to collide and produce X-rays at great distances. Diffusive processes begin beneath the surfaces of B stars, but they do not show spectrophotometric features until one encounters the Ap Si stars, the spectra of which can have strong absorption edges of metallic species. The "quantum" description of the bright/dark zone modes can be estimated from line profiles and pulsation theory, but only with considerable error. The ultimate tests of both interior pulsation theory and line profile models will be the counting of the hot/cool zone pairs on the star and the determination of whether they only are concentrated on a star's equator. The temperature variations in early B stars can modify the velocities and densities of the radiative wind from the star, producing features that in some cases could be inferred from a very few dedicated IUE campaigns before this spectrographic satellite was terminated.

An important example of pulsating B stars are β Cephei stars, variables which generally possess radial or low-degree nonradial modes and therefore show a coherent variation over the surface at any given phase. The pulsations in some of these stars are the largest (in terms of velocity amplitude) of any known variable class. The pulsations emerge as waves at the star's surface which eventually produce shocks in some of these stars. These shocks are thought to modulate the star's wind density and its X-ray emission. In some pulsators these shocks form transient shells which could be imaged in lines of CIV or H-alpha as they emerge from the surface, coast to a standstill, and subsequently return to the star in near free fall. Imaging of a "residue" of this process is a practical way of determining whether mass loss can be driven effectively by this piston-type interaction. Direct observation of the spatial structure of these pulsations will allow immediate identification of the pulsation modes and hence very valuable information about the internal properties of these stars which are progenitors of supernovae. In particular, overshoot and other instabilities from their massive convective cores are likely to have a substantial effect on their later evolution, including the properties of the final supernova explosion and the related synthesis of elements.

For early-type Be stars with strong pulsational amplitudes, there is evidence that the disks are strengthened at the times of constructive interference between multiple oscillation modes (Rivinius et al. 1998, 2001). It has been suggested that the isotropic ejection of gas from a "node" on the star could lead to some material being propelled forward into orbit and some propelled backward to fall back onto the star (Kroll & Hanuschik 1997; Owocki & Cranmer 2002). High-resolution observations would certainly be able to locate the sites of these impulsive explosions and directly correlate them with any possible pulsational events in the star's atmosphere. High time resolution could also track the outward drift of the propelled gas and its smearing into a Keplerian disk.

Theories of NRPs in very rapidly rotating stars are still evolving, and the imaging of how rotation affects the latitudinal profile of pulsation amplitudes would verify or falsify certain modeling assumptions and directly diagnose the angular momentum profiles of these stars (e.g., Lee and Saio 1990; Townsend 2003). In addition, the possible "leakage" of pulsational power into the circumstellar gas could be responsible for several types of observed variations. For example, the direct imaging of a cause-and-effect relationship between stellar and circumstellar features could provide the long-sought explanation for the Be phenomenon. It should be possible to make "movies" of the shock-propagation in extended atmospheres and winds of these stars and thereby determine the time evolution and spatial symmetries of these shocks to constrain and improve theoretical shock models of massive stars. The SI observations will answer a large number of questions about stellar interiors, core convection, chemical mixing, and magnetic fields.

Magnetic B Stars include the He-peculiar (Bp) and β Cephei pulsating stars, which lie in the region of the H-R diagram where winds become weak and are susceptible to competition from other processes such as diffusion and magnetic acceleration. Winds from these stars are preferentially guided from the magnetic poles along the field lines where they collide with particles emanating from the other pole and come to rest, forming a torus-shaped magnetosphere which co-rotates with the star. Current models do not strongly constrain the thickness or extent of these toruses, and some suggest that they could be warped. The determination of these geometrical properties from imaging would permit a better understanding of the interaction of winds and magnetic fields, which can be applied to other stars with potential magnetospheres (possibly some Be stars at some phases, and cool, active, rapidly rotating stars such as AB Dor).

The "Photospheric Connection" between Stellar Variability and Wind Variability

Hot stars exhibit strong stellar winds that contribute significantly to the mass and energy balance of the interstellar medium. Observations of P-Cygni type UV line profiles have been made for decades, and the *quantitative spectroscopy* of hot-star winds has evolved into a reasonably accurate means of deriving fundamental stellar parameters and distances (see review article by Kudritzki & Puls 2000).

The atmospheres and winds of hot stars are intrinsically variable, and it is now accepted that in many cases time-dependent phenomena (e.g., pulsations or magnetic field evolution) in the photosphere provide "shape and structure" to the larger-scale wind (Fullerton & Kaper 1995). The direct observational confirmation of a causal connection between specific stellar variations and specific wind variations, though, has proved elusive. For many O and B stars, it is not clear whether large-scale wind inhomogeneities are rotationally modulated (i.e., due to spots) or if pulsations are responsible, or if the variability occurs spontaneously in the wind.

High-resolution observations with SI would shed light on the stellar origins of wind variability. Simply seeing correlations between individual spots (no matter their physical origin) and modulations in the wind outflow would be key to understanding how hot stars affect their local environments. One suggested paradigm that can be tested is the idea that UV discrete absorption components (DACs) are caused by corotating interaction regions (CIRs) in the winds (Mullan 1984; Cranmer & Owocki 1996; Dessart 2004). Figure 1.26 shows the results of multidimensional hydrodynamic simulations of how bright or dark spots can lead to corotating wind variability patterns that give rise to spectral features resembling the observed DACs.

SI may also resolve the intrinsic small-scale variability of hot-star winds that arises from a nonlinear line-driven instability. Simulations have shown that tiny perturbations can rapidly grow into large-amplitude shocks that fill the surrounding volume, affect the shapes of P-Cygni profiles, and emit X-rays (for recent work see, Owocki & Puls 1999; Oskinova et al. 2004; and references therein). *Chandra* observations of hot-star X-ray emission lines have supported this paradigm for some stars, but not others (e.g., Miller et al. 2002), leaving open the question of how wind instabilities originate and manifest themselves for hot stars of various types. An improved observational understanding of the wind instability would solidify our interpretation of UV P-Cygni profiles and thus lead to improved accuracy in the determination of fundamental stellar parameters and distances.

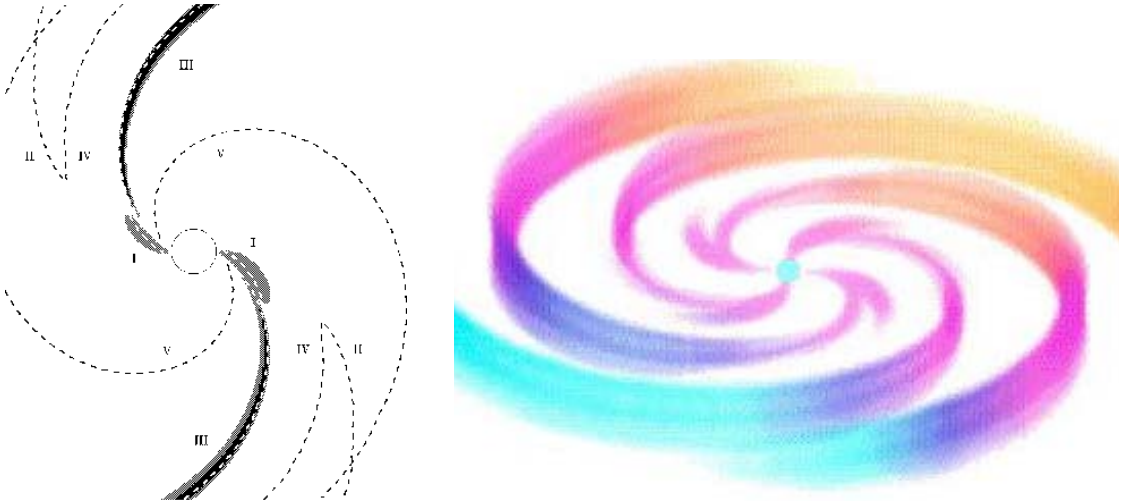


Fig. 1.26: Hyrodynamical simulations of wind variability patterns due to spots.

(**Left**) Grayscale representation of density perturbations in the wind of a rotating O supergiant caused by two bright spots on the stellar surface (Cranmer & Owocki 1996). Overplotted are dashed lines that trace specific dynamical features in the wind: (I) direct spot mass-loss enhancement, (II) prograde precursor, (III) Corating Interaction Region (CIR) compression front, (IV) CIR rarefaction front, and (V) Abbott-mode velocity-gradient kink. (**Right**) Three-dimensional model of equatorial CIR density perturbations viewed from intermediate inclination (Dessart 2004).

1.1.2.3 Interacting Binary Systems as Astrophysical Laboratories

Almost all high-energy sources in the Universe are powered through the potential energy released via accretion. Understanding accretion driven flows in binaries will directly affect our understanding of similar flows around YSOs, including the formation of planets in the circumstellar disk as well as the much larger scale accretion flows in active galactic nuclei (AGNs). Compact, mass transferring binaries provide us with laboratories for testing energetic processes such as magnetically driven accretion and accretion geometries, various evolutionary scenarios, and conditions for induced stellar activity.

Accretion

In close binary stars the flow of material from one component into the potential well of the other is a key in determining the future evolutionary histories of each component and the system itself, and particularly the production of degenerate companions and supernovae. Our cosmological standard candles, the Type Ia supernovae, for example, may be a consequence of accretion onto a white dwarf in a close binary.

Currently, most of our accretion paradigms are based on time-resolved spectroscopic observations. For example, in Cataclysmic Variables (CVs) the picture of accretion onto compact objects via an extended accretion disc is solidly based on spectral and timing information. However, several objects challenge our standard picture and there are significant gaps in our understanding of their formation and evolution.

Large uncertainties exist in our quantitative understanding of accreting processes in many interacting systems. The interaction between the components in close binaries is believed to occur via Roche lobe overflow and/or wind accretion. 3-D hydrodynamic simulations show that the accretion processes in tidally interacting systems are very complex. Wind accretion is even more complicated. The amount of the accreted material depends on the characteristics of both components including

stellar activity and wind properties (e.g. density and velocity), the binary parameters (e.g. orbital period and separation), and the dynamics of the flow. In the case of Roche lobe overflow, the accretion may form an extended accretion disk whose turbulent magnetic dynamo drives the flow through it. Stellar activity of the rapidly rotating donors and their impact on the binary remains poorly constrained despite being crucial in regulating the mass transfer rate and setting the long-term evolution.

The key to further advance in accretion studies is resolving a wide range of interacting binaries and studying their components and mass flows. The SI sub-milliarsecond resolution in the UV will lead to unprecedented opportunities for detailed studies accretion phenomena in many interacting systems including symbiotics, Algol type binaries, Cataclysmic Variables (CVs) and their progenitors. The SI will be able to resolve the components of numerous interacting systems and will therefore provide a unique laboratory for studying accretion processes. The binary components can be studied individually at many wavelengths including Lyα, CIV, and MgH&K lines, and the geometry of accretion can be imaged directly, giving us the first direct constraints on the accretion geometries for a range of systems. This in turn will allow us to benchmark crucial accretion paradigms that affect any stellar population and even the structure evolution of galaxies whose central black-holes are steadily accreting, shaping the long term evolution.

For example, the SI will be able to separate the components of many currently unresolved symbiotic systems. Symbiotics are interacting binaries showing a composite spectrum with signatures of a late-type giant and a high-temperature component, often a compact object in the form of a white dwarf. They are some of the most fascinating interacting binary systems because of their dramatic transformations, and extremely complex circumbinary environment. Symbiotics are very important systems because they are likely progenitors of bipolar planetary nebulae. They have been also invoked as potential progenitors of at least a fraction of Supernovae type Ia, a key cosmological distance indicators. So far, the individual components have been resolved clearly in only one nearby symbiotic system - Mira AB (Karovska et al. 2005 and references therein). The SI will have the capability of resolving a significant fraction of currently known symbiotic systems, and will be able to image directly the individual components and the dynamical accretion flows in dozens of nearby systems.

SI will be able to study the direct impact accretion that occurs in Algols, where the mass-losing donor star feeds the primary with a stream of gas that hits the surface of the primary. The physical properties of the impact site and the associated transfer of angular momentum are crucial in setting the accretion history. For example, Algol itself (β Per) is a nearby 2 mas eclipsing binary system (at 30 pc) and its accretion environment and geometry will be easily imaged by SI. A simulation of the mass transfer in Algol is shown in Figure 1.27.

Cataclysmic variables form the most common source of compact accreting binaries, harboring a primary white dwarf accreting in a short-period (1-10 hrs) system. The SI will be able to resolve AE Aqr, one of the brightest and closest CVs. This 9.8 hour system contains a K5 donor star feeding a rapidly spinning and magnetized white dwarf. The 33s spin period provides a rapidly rotating magnetosphere around the dwarf that prevents normal disc accretion and instead propels most of the transferred gas out of the binary system, producing a large outflow in the orbital plane, as shown in Figure 1.28. This outflow can easily be resolved with SI, and would be the confirmation that is needed to support this propeller mechanism. The result would be a key to our understanding of other systems containing rapidly spinning and magnetized compact objects including accreting pulsars.

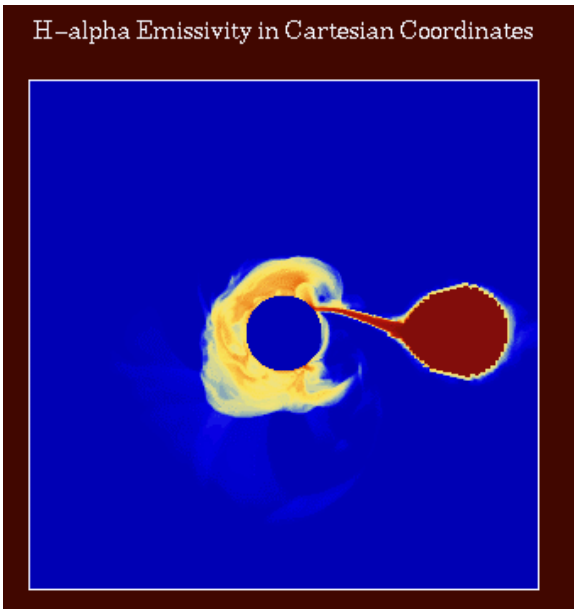


Figure 1.27: Hydrodynamic simulations of the mass transfer in the Algol prototype β Per.

The gas stream impact onto the surface of the primary forming a local hotspot as well as an extended flow around the accretor. Colorscale represents the emission strength from such a flow as derived from the simulations in Richards & Ratliff 1998. The separation between the components is 2 mas.

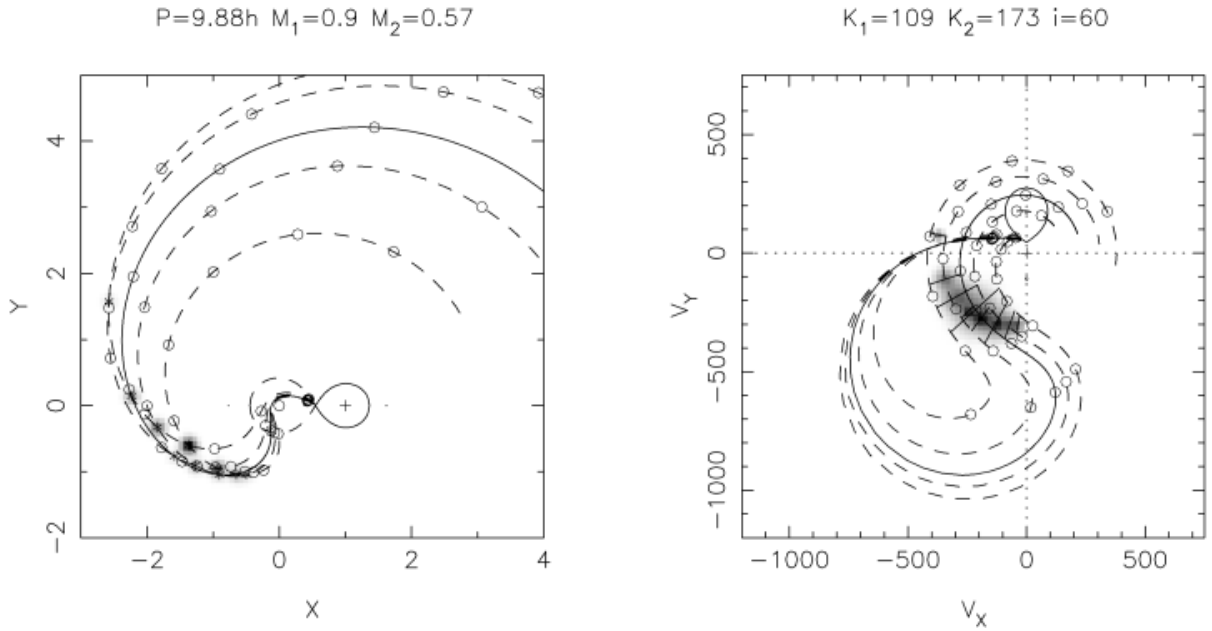


Figure 1.28: Solid curves are model particle trajectories of gas that is transferred from the donor star toward the magnetic white dwarf in the accreting binary AE Aqr.

The rapidly spinning white dwarf propels the infalling gas and launches a large outflow away from the binary system. Colliding trajectories downstream lead to local line and continuum emission as indicated by the grayscale. Right panel plots the expected velocities of this emission, corresponding in detail to the line emission observed from this binary (from Pearson, Horne & Skidmore 2003, MNRAS, 338, 1067). [separation between stars in the center is 0.1 mas].

The SI will detect colliding winds in the photoionization wakes in high-mass X-ray binaries. Phase-correlated emissions from UV lines strongly hint at the existence of neutron star companions to B supergiants and Be stars as the latter plow through a wind or disk along their orbits. For example, in the X-ray source Vela X-1 the neutron star completely ionizes the wind from the Be star flowing

toward it, but at the edges of this Stromgren sphere the wind encounters a "photoionization shock," a structure which may be resolvable. The size and orientation of this shock are sensitive to the wind velocity and the effective temperature of the neutron star. The latter is difficult to infer by present observational means other than by heavy reliance on accretion theory with uncertain parameters.

Stellar Activity Enhancements

There is substantial indirect evidence that magnetic activity plays an important role in driving the formation and secular evolution of a wide range of binary systems. In some cases, tomographic imaging has identified signatures of star spots on the rapidly rotating donor stars in binaries. SI can provide a breakthrough by resolving the surface structure of such stars directly. Furthermore, with the SI, we will be able to study the presence of active coronae and stellar prominences for the first time.

For example, RS CVn binaries are detached systems containing a very active sub-giant. Large star spots are inferred through light-curve modeling and the presence of emission lines and UV/X-ray emission indicates active chromospheres and coronae. Direct imaging of the stellar activity structures with SI would be invaluable in linking these active sub-giants to solar-type activity. [An example SI imaging target is HR1099, a bright RS CVn system located 29 pc a way containing a K1 sub-giant and G5 dwarf in a 2.8 day orbit.]

Activity Enhancements by Planets – appearance of hot spots on the stellar surface

Nearby giant planets can also enhance stellar activity as seems to be the case with the close-in giant planets around HD 179949 and Upsilon And (Cuntz, Saar, & Musielak 2000, Shkolnik, Walker, & Bohlender 2003 Shkolnik et al. 2005). But rather than the tidal forces in close stellar binaries, the enhanced stellar emission in these cases appears to be due to the interaction of the stellar and planetary magnetic fields The additional heating contributes ~10% of the emission in Ca II H&K and should be greater in higher temperature UV and EUV lines. The increase of the line emission depends both on the stellar and the planetary magnetic field strengths. Also, details of the emission, such as the phase shift between the emission peak and the sub-planet point on the star, yield information on the properties of the close-in stellar wind and stellar magnetic geometry.

Only 10 planetary systems have been studied for this phenomenon, with the two with planets within 0.1 AU exhibiting it. We expect for that ratio at least 6 systems for SI to observe. We estimate the possibility of detecting the signal, based on the current detection in Ca II H+K of Upsilon And as an example. The star has a 1.1 mas diameter, and thus the SI image will contain about 100 resolution elements on the surface in the Mg II. If the bright spot is concentrated in one resolution element (reasonable, given the phase duration of the Ca II HK enhancement), we expect a local Mg II contrast of a factor of 6.4. This bright spot moves around the star in $P_{orb} \approx 4.62$ days, considerably faster than the stellar rotation at ≈ 14 days. In the case of HD 179949, the diameter is 0.41 mas, or about 10 resolution elements across in the Mg II line for SI. The estimated enhancement of about 12% of the stellar Mg II emission should result in a local enhancement of a factor of about 2.1, moving around with a period of $P_{orb} = 3.09$ d, while the star has $P_{rot} \sim 9$ days.

1.1.2.4 Extragalactic and 3-D Universe

Supernovae

With the exception of the relatively nearby SN1987A (in the LMC), which could be well-studied by HST, it has not been possible to obtain much information about the close-in spatial structure of supernovae (typical sizes remain below about 1 mas, which is not reached by current ground-based optical telescopes). Radio VLBI observations have resolved a few supernovae, but are more a probe of the interaction of the SN shock front with the circumstellar material than of the supernova. Direct imaging by SI of early stages of expansion would be possible of supernovae at a distance of few Mpc. These images would provide essential information on the nature of the explosion, especially in regard to its symmetry or asymmetry, and of the early evolution of its structure with time. Spectral radial velocity observations could be combined with SI's interferometric measurements of the geometric expansion to derive important estimates of the distances to the supernovae.

Active Galactic Nuclei

Images of Active Galactic Nuclei (AGN's) could resolve the transition zone between the broad and narrow emission line regions and help resolve the origin and orientation of jets. Sub-milliarcsec resolution could enable study of broad and narrow-line emission regions (but probably not the central engine) at 0.5 milli-arcsec resolution (0.02pc at 10 Mpc). Images of the transition zone between broad-line and narrow-line regions would answer the question: "is material being stripped from the broad-line clouds, which are in close to the nucleus, and driven out to the narrow-line region?" It is best studied in the UV/optical emission lines within a fraction of a parsec of the nucleus. The problems of the origin and orientation of the jets could be addressed with detailed broad-band images of the inner regions, especially if some spectral resolution is possible with, e.g., energy-resolving detectors. It would also enable a search for electron scattering by outflowing plasma associated with jets or with massive winds driven off accretion disks. Such images could also provide an answer to the question: "do type 1 Seyferts have molecular tori?". Broad-band imaging at sub-parsec scales could tell us if tori are obscuring starlight.

AGN winds

The structure of AGN is unknown, and may be complex, random, or perhaps orderly, with that order masked under the profusion of observables that AGN present to us. One major new area of research is AGN winds. This area of study has only become established since about 1995, and has come into its own with the advent of high resolution spectra from Chandra and XMM-Newton, and from their UV counterparts, STIS and FUSE.

These AGN winds carry a substantial mass loss rate compared with the m(dot) needed to power the AGN continuum itself. They are therefore important to understanding the dynamics and structure of AGN. Since these winds enrich the surrounding intergalactic medium they have larger implications for cosmology.

Though least ambiguously seen in absorption, AGN winds also produce emission lines. The location of the gas producing these emission lines is much debated. Suggestions range from the size of the optical/UV "broad emission line region" (a few 10s of light-days), to the size of the "obscuring torus" at a few parsecs. Optical observations of similarly high ionization emission lines (e.g. FeX) suggest both a small and a large scale region, differentiated by their velocity structure.

For AGN out to the Virgo Cluster or a little beyond (D=20Mpc) 1.0 mas is 0.1pc. Hence the 'obscuring torus' scale is readily resolved and should yield telling images of the AGN wind. If the CIV remains point-like at this level, the more radical BELR-scale hypothesis will be greatly strengthened. Figure 1.29 shows a simulation of SI capabilities for differentiating between different AGN broad emission line region morphologies and inclinations. The hourglass shape model can be discerned in the simulated observation and its tilt is clearly visible as well. The simulation was produced using the SISIM code (Rajagopal et al. 2003) assuming 30 mirror elements distributed in a non-redundant pattern, within a 500 meter circle and using a 10 Å bandpass around the C IV doublet near 1550 Å.

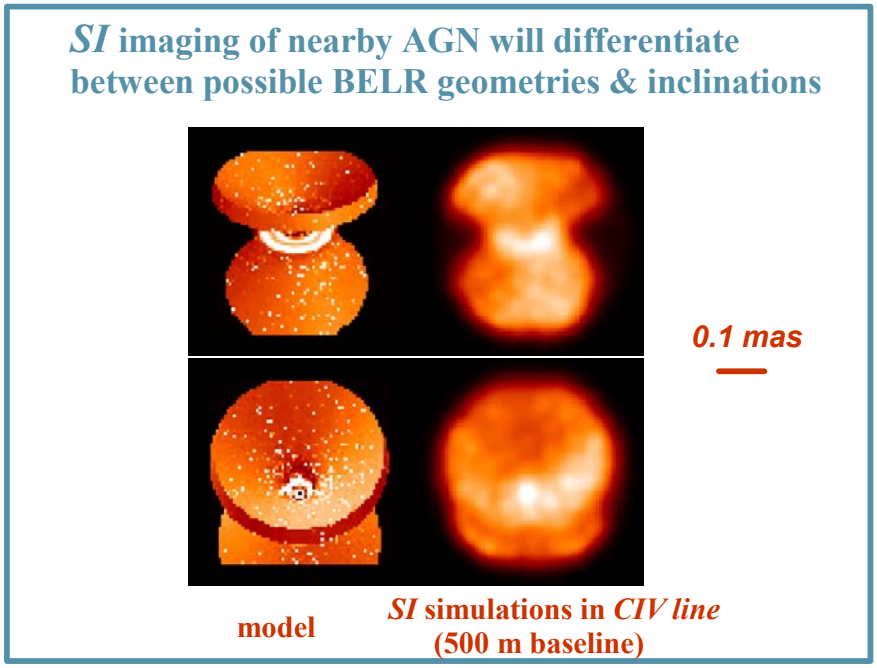


Figure 1.29: Simulations of SI's capabilities for UV imaging of AGN morphologies.

Distance Measurements with SI

Cosmic distance scale determination methods include relative and absolute distance estimators. Absolute distance estimators include various applications of the: Baade-Wesselink method (e.g. Sasselov and Karovska 1994), and distance estimators using time delays (e.g., SN1987A; Panagia 1991). Relative distance estimating methods include period-luminosity relation for Cepheid star variability (used by the *Hubble* Key Project ; Freedman et al. 2001), and supernovae of type SN 1a as 'standard candles'. Relative distance estimators often involve assumptions and correlations, and have inevitable model dependencies. Absolute methods on the other hand have the advantage of having lesser dependence on physical models and provide an independent way to determine the distance scale. SI will provide a new avenue for determining distances to various astronomical sources including many nearby pulsating stars and high redshift sources such as distant supernovae and quasars.

Mapping the 3-D geometry of the Universe involves measurement of the large "cosmic" scale distances of high redshift sources such as distant supernovae and quasars. Distance measurements using a relative distance estimation, e.g., the brightness of supernovae of type SN1a at z~1.5 as "standard candles" (Perlmutter et al. 1999), suggest that that the expansion of the Universe is currently

accelerating. Recently Elvis and Karovska (2002) proposed an absolute method for estimating distances to quasars at different redshifts using long-baseline interferometry of quasars Broad Emission Line Regions (BELRs). This geometric method uses the size of the quasar BELR from reverberation mapping (Peterson et al. 1993, 2001) combined with interferometric measurements of the angular diameter of the emitting region to derive the distance to the quasar. When compared to relative distance estimators this method is much less dependent on physical models and of changes in the fundamental constants (other than c, the speed of light) because it uses a standard-length measurement approach rather than a standard-candle approximation.

The sub-milliarcsecond resolution of SI provides a unique opportunity to measure the angular sizes of BELRs of quasars at $z < 1$ in several UV lines including C IV (1550Å) and Mg II h&k (2800Å) (Peterson 2001). The quasar broad emission lines (v~ 5000 – 10,000 km s^{-1}) originating in the BELR gas clouds respond to changes in the continuum source in the center by changing their intensity (~20% in the UV) with a time-lag of a few days to years. This time-lag is induced by the light travel time from the continuum source. For low redshift quasars the size of the BELRs is ~10 light days corresponding to an angular size of a fraction of a milliarcsecond (Figure 1.30). Reverberation mapping suggests that shorter wavelength emission lines (especially in the UV) are more favorable for BELR size measurements.

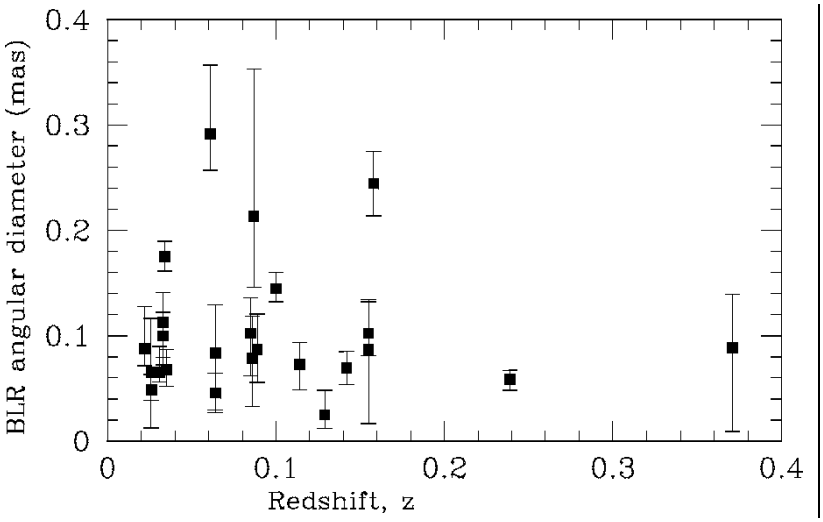


Figure 1.30: Angular diameters for the Hα and Hβ BELRs of nearby active galaxies, assuming H_0=65 km $s^{-1}$$Mpc^{-1}$ (Peterson et al. 1997, Kaspi et al. 2000)

By monitoring a number of quasars for continuum changes, one could identify the onset of a significant continuum increase in a particular quasar, which could then trigger SI observations. The SI would observe the target quasar in both low and high continuum states, with appropriate lags included, so that the difference map will be a measure only of the "responsive fraction" of the line, i.e. that part which responds to continuum changes.

1.2 Relation to NASA/Science Mission Directorate (SMD) Strategic Plans

Stellar Imager was included in the 2000 and 2003 SEC Roadmaps and is now identified as a "*Flagship and Landmark Discovery Mission*" in the draft 2005 Sun Solar System Connection (SSSC) Roadmap (see Figure 1.31). SI is also a candidate for a "*Pathways to Life Observatory*" in the Exploration of the Universe Division (EUD) Roadmap (May, 2005). SI will provide an angular resolution several hundred times that of the Hubble Space Telescope (HST) and will resolve for the first time the surfaces of Sun-like stars and the details of many astrophysical objects and processes. SI address several of the major objectives in these 2005 Strategic Plans and Roadmaps, including:

- Explore the origins & evolution of structure & life in the Universe (Strategic Objective #5)
- Understand the Sun and its interaction with and impact on the Earth and environmental conditions in solar system space that will be experienced by future human explorers (Strategic Objective # 15)
- SI supports Strategic Objective # 4 (Search for Earth-like planets and habitable environments around other stars using advanced telescopes) by assessing the impact of stellar activity on the habitability of planets found around other stars
- Inspire and motivate the nation's students and teachers (Strategic Objective #13) – just imagine what it will be like to see the Universe with >200x the resolution of the Hubble Space Telescope!

SI fits into the President's Exploration Initiative explicitly in two distinct areas:

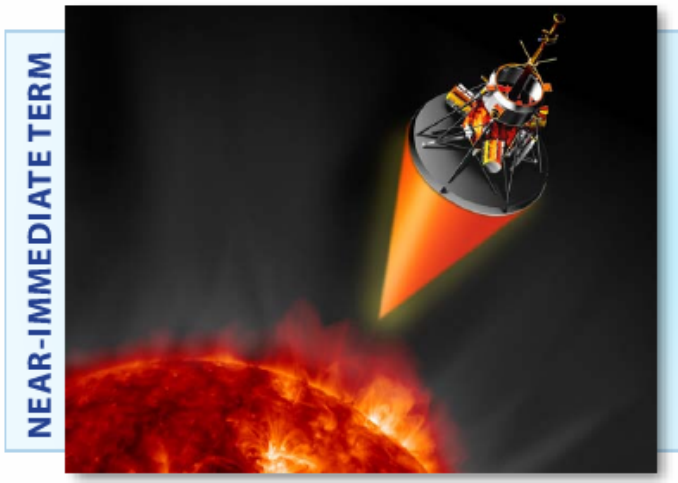
1) as one of the "deep-space observatories" which will be a part of the search for and study of habitable planets around other stars:

Stellar Imager (SI) is an essential part of this mandate since it enables the assessment of the impact of stellar magnetic activity on the habitability of planets found by the planet search and imaging missions (e.g., TPF-I and Planet Imager (PI)).

2) as a means to improve our ability to forecast space weather within our own solar system:

Exploration requires that we know space weather throughout much of the heliosphere, and that means we need long-term forecasts of solar activity, which in turn requires a fundamental understanding of the solar dynamo and of all related transport processes. The Living With a Star initiative addresses that on the fairly short term, while the Stellar Imager is to provide the knowledge (constraints from a broad population of stars of differing activity level and stellar parameters/properties) critically needed to test and validate models developed under LWS.

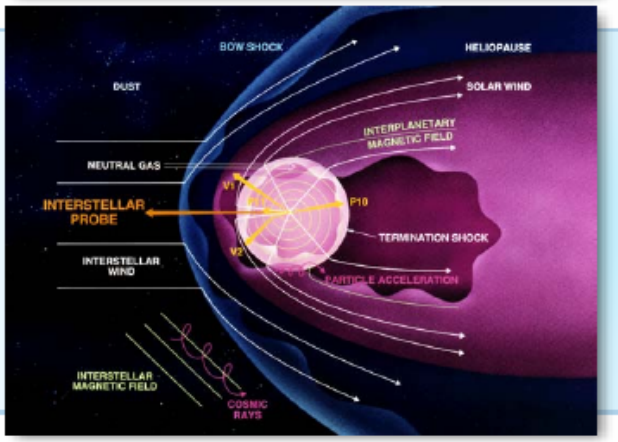
SSSC Landmark Discovery Missions



NEAR-IMMEDIATE TERM

Solar Probe

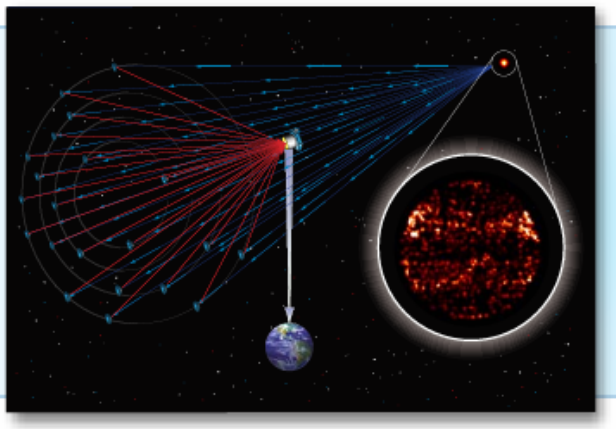
- Measure magnetic reconnection at the Sun
- Thermal shielding protection for in situ solar wind measurement at 4Rs



LONG-TERM

Interstellar Probe

- Analyze the first direct sample of the interstellar medium
- Advanced propulsion for 200Au in 15 years



FAR-TERM

Stellar Imager

- Image activity in other stellar systems
- UV interferometry in space with precision formation flying autonomous constellation

Figure 1.31: The Stellar Imager is a "Flagship Landmark Discovery Mission" in the Sun Solar System Connection 2005 Roadmap.

With the Stellar Imager we aim to understand the inner workings of stars leading the dynamo action that drives stellar magnetic activity. In obtaining the observations and knowledge to develop and test a dynamo model that can be used to forecast solar activity on time scales of years to decades, this primary science objective will also fundamentally advance our understanding of

- the development of structure/flows of magnetic fields
- the formation of stars and planetary systems,
- the origin of life near stars,
- the evolution of stars, and
- a star's impacts on the habitability of a planetary biosphere.

The exciting overarching research themes discussed earlier in this Chapter require the imaging of stars and of their environments, and the measurement of their internal structure and dynamics. In parallel to the primary science goals, the high-resolution imaging of the Stellar Imager will also enable us to explore and study the universe in general in unprecedented detail. The combined results from the primary and supplemental science programs of SI constitute a voyage of exploration that will

- test theories of stellar interiors,
- deepen our understanding of a broad range of physical processes, and
- help forecast the activity of the Sun for our society that is Living With A Star.

These research themes of the Stellar Imager mission align with the nation's science priorities for astronomy and astrophysics as identified in NAS Decadal Surveys as well as NASA strategic plans. For example, Stellar Imager contributes significantly to each of the mission areas and corresponding research goals as identified in the 2003 Strategy of the NASA Office of Space Sciences (now Science Mission Directorate, SMD), as illustrated in Table 1.2. The Stellar Imager also contributes to all but one of the research focus areas of the SMD Sun-Earth Connections theme (Table 1.3), which is why SI has been included in its three most recent strategic plans as formulated in the SEC Road Maps. The SI mission also contributes to top-level elements identified in the 2001 Decadal Survey of Astronomy and Astrophysics:

- "Survey the universe and its constituents, including galaxies as they evolve through cosmic time, stars and planets as they form out of collapsing interstellar clouds in our galaxy, [. ..]",
- "Use the universe as a unique laboratory for probing the laws of physics in regimes not accessible on Earth [. ..]"

SI also addresses two of the key Challenges that were formulated in the 2002 "Decadal Research Strategy in Solar and Space Physics:"

- 1) "Understanding the structure and dynamics of the Sun's interior, the generation of magnetic fields, the origin of the solar cycle, the causes of solar activity, and the structure and dynamics of the corona."
- 4) "Understanding the basic physical principles manifest in processes observed in solar and space plasmas."

Table 1.2: *Stellar Imager within the primary NASA/SMD Areas and Goals as formulated in the 2003 Strategy. The last column identifies how SI contributes in a primary (P) or supporting (S) role.*

OSS Mission Area	Goal	Role of SI
Understand and protect our home planet	Understand the Earth system and apply Earth system science to improve prediction of climate, weather, and natural hazards	P: Understand solar activity, and its consequences on space weather and Earth climate, to enable reliable forecasting of the climates that make up the Sun-Earth system
	Enable a safer, more secure, efficient, and environmentally friendly air transportation system	S: Enable long-term forecasting of radiation hazards to air travelers
	Create a more secure world and improve the quality of life by investing in technologies and collaborating with other agencies, industry, and academia	S: Stimulate research at the forefront of our understanding of astrophysical and plasma-physics processes at NASA, NSF, universities, and industry
Explore the universe and search for life	Explore the fundamental principles of physics, chemistry, and biology through research in the unique laboratory of space	P: Observe stellar surface activity and interior structure and dynamics to test our understanding of fundamental physics, and of complex, nonlinear systems
	Explore the solar system and the Universe beyond, understand the origin and evolution of life, and search for evidence for life elsewhere	P: Contribute to our understanding of planetary system formation and of the habitability and radiation environments of planets on which life originates and evolves
Inspire the next generation of explorers	Inspire and motivate students to pursue careers in science technology, engineering, and mathematics	P: Reveal, for the very first time, the true nature of stars like the Sun that until then are only known as point sources in the sky
	Engage the public in shaping and sharing the experience of exploration and discovery	P: Share the experience of discovery as images of a wide variety of stars and other astrophysical objects change with time

Table 1.3: *SI within the primary Sun Earth Connection science objectives and research focus areas (2003 Strategic Plan) within which SI plays a primary (P) or supporting (S) role.*

SEC science objective	SEC research focus area	Role of SI
Understand the changing flow of energy and matter throughout the Sun, heliosphere, & planetary environments	Understand the transport of energy and matter within the Sun, the solar atmosphere, and into the solar wind	P: Enable the development and testing of a solar dynamo model, including the transport of field throughout the interior and atmosphere
	Determine the evolution of the heliosphere and its interaction with the galaxy	S: Measure stellar surface fields to interpret observations of stellar asterospheres, thus simulating an outside view of our heliosphere
	Understand the response of magnetospheres and atmosph. To external and internal drivers	-
Explore the fundamental physical processes of plasma systems in the solar system	Discover how magnetic fields are created and how charged particles are accelerated	P: Image stellar dynamo patterns and develop and validate a comprehensive dynamo model
	Understand coupling across multiple scale lengths and its generality in plasma systems	P: Observe and model scale couplings for stellar dynamos, large-scale circulations, and convection
Define the origins and societal impacts of variability in the Sun-Earth connection	Develop the capability to predict solar activity and the evolution of solar disturb. As they propagate in the heliosph. and affect Earth	P: Develop and test a model for solar activity on scales of years to decades
	Develop the capability to specify and predict changes to the Earth's radiation environment, ionosphere, and upper atmosphere	P: Develop and test a model for solar activity on scales of years to decades, which determines solar atmospheric activity
	Understand the role of solar variability in driving global change in the Earth's atmosphere and in controlling long-term space climate	P: Understand the solar dynamo on time scales from years to billions of years, and evaluate its effects on Earth's habitability

The Stellar Imager complements ongoing and planned research programs within NASA. For example, the recognition of the coherence of many of the problems facing astrophysicists, the importance of understanding the impact of stellar activity on humanity and the role it plays in the origin of life, has led the Office of Space Sciences to promote an initiative, called "Living With a Star." This initiative specifically focuses on the solar aspects of stellar dynamos through future mission such as the Solar Dynamics Observatory, which itself builds on successful current projects that include several instruments on the Solar and Heliospheric Observatory (SOHO), the Solar X-ray telescope on YOHKOH, and the Transition Region and Coronal Explorer (TRACE).

The aspects of the origin of life in the Universe to which SI contributes are addressed with a series of NASA missions along a time line that extends well into the future, and include the Space-Interferometer Mission (SIM), the James Webb Space Telescope (JWST), the Terrestrial Planet Finder Interferometer (TPF-I), the Life Finder (LF) and ultimately the Planet Imager (PI). Together with these missions, SI allows the complete viewing of other solar systems from the central stars to the orbiting giant gaseous planets and smaller terrestrial, possibly life-bearing planets.

SI provides a stepping stone towards crucial technology needed for the major future interferometry missions. SI is comparable in complexity to the Terrestrial Planet Finder - I (compare Table 1.4), and serves as a useful technological and operational pathfinder for the Planet Imager: the SI resolution is approximately 40 times less demanding than that ultimate NASA goal in terms of its required angular resolution. Note that the baseline requirements for synthesis imaging are ~ 6 times more relaxed than on TPF-I, if SI tolerates a mere 1% light loss due to phase uncertainties compared to the strict nulling requirement of the light of the central star for TPF-I. Consequently, the roadmap for the baseline metrology for TPF-I will yield the technology needed for SI as an intermediate product. Other technological synergies are found in formation flying, pointing stabilization and vibration control, determination of optimal array configuration, beam combining systems, optical path-length stabilization, metrology technology, lightweight mirrors, and energy-resolving detectors.

The science of the TPF-I/LF/PI triplet is entirely complementary to that of SI: whereas these missions require the successful removal of the light from the central star, SI requires that very light. Note that the TPF-I/LF/PI missions cannot be used to study stellar magnetic activity in adequate detail: (a) the optimal wavelength for activity studies is the UV rather than the infrared that is to be used for molecular spectroscopy and imaging of the planets, and (b) the use of IR light instead of UV light to reach the targeted angular resolution for SI would require a 40x longer baseline than envisaged by TPF-I, i.e. 20km versus 500m (not impossible in principle, perhaps, but resulting in a very much sparser aperture and slower aperture filling, both vastly increasing integration times, compromising activity studies and ruling out asteroseismology). Hence, the TPF-I and SI science goals can only be met with separate missions.

Table 1.4: *SI and TPF-I free-flying, multi-telescope, spectroscopic interferometers.*

Together they provide a complete view of other solar systems.

SI	TPF-I
Focus on central star	Focus on planet atmospheres
1-meter class UV/optical	3.5-meter class near-IR, cryogenically cooled
λ/50 baseline stability for 1% light loss: 3 nm precision	nulling requires phase stability with 0.5-nm precision at 7 micron
9-30 telescopes	4-6 telescopes

The science and technology interdependencies of SI and other missions and observatories are shown in Figure 1.32.

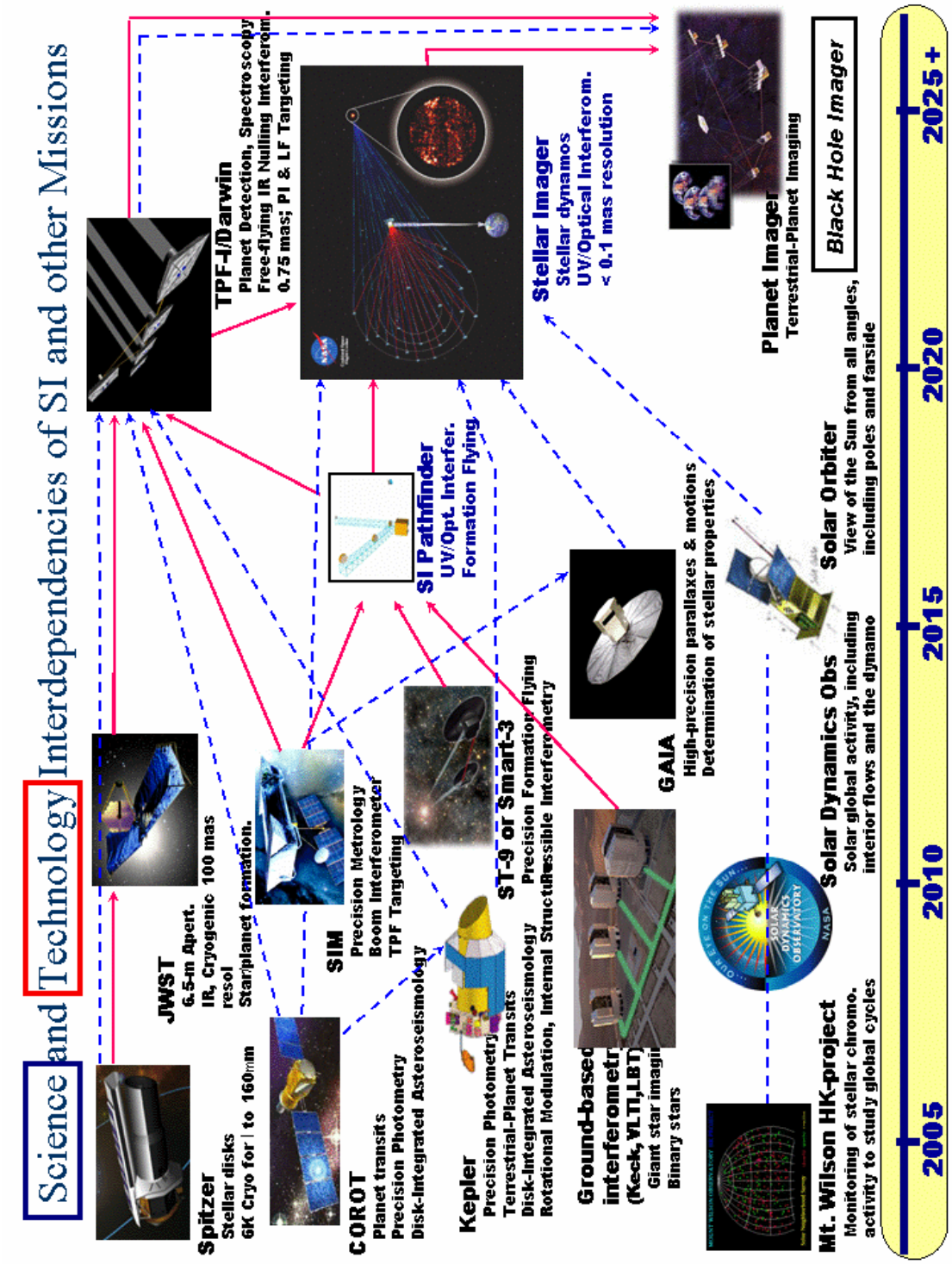


Figure 1.32: Interdependencies of SI and other Missions and Observatories. The Stellar Imager fits naturally on the timeline for missions designed to study the origin of life in the Universe.

1.3 Uniqueness or Scientific Advantages of the Proposed Approach

1.3.1 Overview of Possible Approaches

The SI concept will yield a spatial resolution of about 0.1 mas (milli-arcsec) at ultraviolet wavelengths. This corresponds to the scientific requirement of about 50,000 km resolution (3% of a solar diameter) at a distance of 4 pc. There are other techniques that can be used to attain high angular resolution, but none meet our needs as well as the SI concept.

1.3.1.1 Adequate Spatial Resolution

VLBI: The VLBA radio array achieves 1.3 mas spatial resolution at its shortest operational wavelength, 1.3 cm, about 10x lower resolution than that planned for SI in the mid-UV.

Optical: The spatial resolution of interferometers is constrained by the Rayleigh relation, resolution = 1.22 lambda/D. To achieve comparable resolution, an optical interferometer must have 3-4 times the baseline of a UV instrument.

Doppler imaging (DI): This technique relies on the rotational broadening on a rotating star to spread spectrally the image of the star. The line profile, in velocity space, can be inverted to generate a map of the brightness as a function of velocity. Over a rotation, the acceleration of these contrasting features provides information on the latitude (velocity corresponds directly to longitude). The resolution currently attainable for a rapidly rotating star (v sin i ~ 100 km/s) is about 10 degrees in longitude. This is comparable to the SI goal.

The advantage of DI techniques is that the spatial resolution is independent of the distance to the star (given sufficient S/N).

The disadvantage is that only a very small percentage of cool stars can be Doppler imaged, since the technique only works on rapidly rotating stars (V sin i >~25 km/s) and the spatial resolution depends on the number of resolution elements across the line profile. DI techniques are less sensitive to latitudinal structures, and characterization of permanent features (e.g., polar spots) is model-dependent.

DI is inappropriate for study of the activity in solar-like stars, the prime goal of the SI mission, because in these stars the line broadening is dominated by turbulence and not by rotation.

It will be important, in order to obtain the end-truth about the utility of the Doppler Imaging technique to include some stars in the Verification Stage of SI's program that can be imaged with DI as well as SI and thus enable an assessment of the success of DI on some stars.

1.3.1.2 Appropriate Wavelength Coverage

The prime driver for the SI is to image stellar magnetic activity. As argued elsewhere, this is best done in the ultraviolet where the contrast with the underlying stellar photosphere is very high.

Radio emission is emitted from coronal magnetic loops, but detectable radio-emitting electrons are restricted to regions far up in the corona, due to opacity effects. The radio emission is non-thermal and provides a limited view of the thermal plasma; in particular, the structure of magnetic loops as they emerge from the photosphere cannot be studied well.

Optical DI techniques are sensitive to the dark photospheric spots. In the Sun, these do track the regions of magnetic flux emergence.

Optical interferometry can in principle detect the active regions through their plage emission, primarily at H-alpha or the Ca II H&K lines. However, the contrast with the bright photosphere is low, and narrow band filters (<1 Angstrom) will be required. One also needs to obtain off-band images.

X-rays are emitted from the tops of the same active region loops the SI will study. There is no problem with contrast against the underlying star. In principle, an X-ray interferometer can achieve our goals, and can reach our spatial resolution requirement with a baseline about 1% of the SI baseline. However, stars emit many more UV photons than X-ray photons: a solar like star at 10 pc will have an X-ray flux of about 5×10^{-5} ph/cm^2/s, while the same star will emit about 30x that in the C IV lines. An X-ray mission will require a much larger effective collecting area to satisfy our sensitivity requirements.

The SI prime Science goals require

- Imaging stellar surfaces to measure flux emergence patterns (in latitude and longitude) and flux dispersal and advection (by convection, differential rotation, and meridional circulation).

- The use of spatially-resolved asteroseismology to measure large-scale flows on the surface and in the interior.

which only can be met by high angular-resolution UV/optical imaging (UV for surface imaging, broad-band optical for seismology).

Table 1.5 Diagnostics for activity and seismology

Technique:		Because:
Doppler imaging	Fails	Sources evolve well before a rotation is completed on a Sun-like star; latitude ambiguity on fast rotators
Rotational modulation	Fails	Sources evolve too fast; no latitude information; no reference level
X-ray imaging	Fails	No access to asteroseismology; too much confusion by rapid coronal evolution
Optical only imaging	Fails	Works for seismology, but not for surface imaging (Spot coverage too small on Sun-like stars; no access to surface flows as spots dissolve)
UV & optical imaging	**Succeeds**	UV ➔ High contrast to detect active regions and their dispersed patterns; Optical ➔ seismology

1.3.2 Methodologies in Detail

1.3.2.1 Imaging Magnetic Activity at the Stellar Surface

Magnetic activity affects all layers of the stellar outer atmosphere. The nonradiative heating associated with the presence of a magnetic field modulates the emission from the various domains: optical and

UV emission from the surface layers and from the relatively warm chromospheres; EUV and X-ray emission from the hot coronae; and radio emission with wavelength-dependent contributions from multiple regions within the outer atmosphere.

The signals that are currently most valuable to the mapping of stellar magnetic fields originate from the lower domains of the atmosphere, and are observable in the optical and UV with high-quality spectrographs on the ground and in space. On the surface, the starspots and faculae cause modulation of the white-light signal as well as distortions of line profiles associated with Doppler and Zeeman effects. The changes in the broad-band flux observed at Earth can be large for the most active stars, in which more than half of a hemisphere may be covered by spots (such as shown in Figure 1.33 for V410 Tau, XX Tri but also in II Peg). For a star like the Sun, in contrast, the brightness change is limited to a few tenths of one percent.

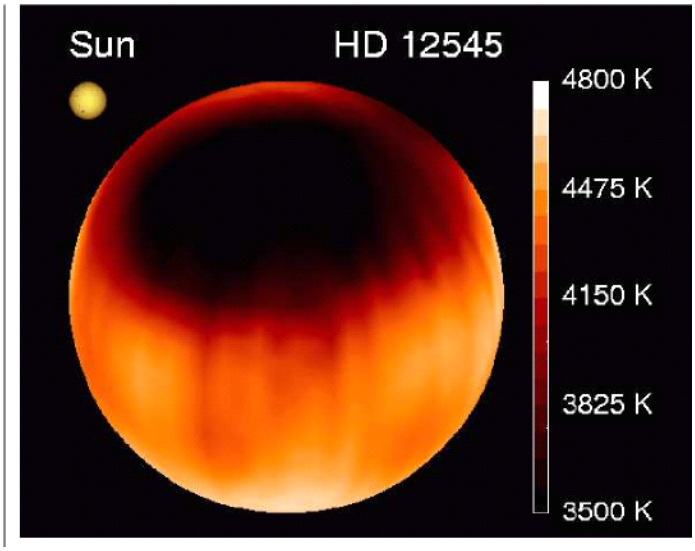


Figure 1.33: Doppler image of the giant star HD 12545 (XX Trian-gulum) and its enormous starspot near its rotational pole (the rotation axis lies in the plane of the paper). The color coding is proportional to temperature. Dark means cooler than the effective temperature and white means hotter. An image of the solar disk is shown to scale as a comparison.

Signals from higher regions in the stellar atmosphere are much harder to interpret than those from near the surface. The thermal line widths are larger and thus more often exceed the rotational Doppler widths. Hence, few useful Doppler signals are observable even if we had the high-quality spectrographic capabilities available in the visible and UV, with the exception of the most rapid rotators. The observed coronal intensities can be used to reconstruct longitude maps of the coronal brightness, but little latitude information can be retrieved (except in strongly tilted cases where high-latitude non-eclipsing regions may be differentiated from lower-latitude eclipsing ones).

Given measurements of (spectral) intensities, there are five classes of methods to explore the spatial structure of the magnetic fields of Sun-like stars. We discuss these here briefly, with their limitations:

- *Rotational-modulation mapping.* Structures that rotate onto and off the visible hemisphere cause brightness variations that can be used to derive longitude maps of activity. A limiting factor is that the time scale for intrinsic evolution of the source regions should be large in comparison to the stellar rotation period, which is not the case for a star like the Sun, where active-region evolution happens on a time scale of one to two weeks, compared to the rotation period of one month.

- *Eclipse mapping.* In binary systems, the crossing of one component in front of the other allows some low-resolution recovery of atmospheric structure on the sides of the stars facing each other during eclipse, and off the limbs in favorable conditions. For the few eclipsing systems that are Sun-like, with orbits wide enough to ignore tidal coupling, this method typically provides longitude information in the eclipsed part of the star, whereas our problem requires latitude information on the entire star.
- *(Zeeman) Doppler imaging.* Inhomogeneities that move across the stellar disk because of stellar rotation result in line-profile distortions caused by the Doppler effect. Stellar surfaces can be imaged by modeling these time-dependent distortions. Latitude information can be recovered by the observed velocity amplitude of the signal, and some height information (as in the case of high prominences) from the rapidity with which structures cross the line profile compared to the surface rotation period. The technique is limited to rapid rotators with $v \sin(i)$ values significantly exceeding the intrinsic line widths (that is, in excess of both the thermal and non-thermal broadening), i.e. stars should rotate at least some five times faster than the Sun; on the other hand, the rotational Doppler broadening of the target star should not be so large that the line becomes only a weak depression upon the overall emission. These restrictions imply that the method is useful for studies of relatively active stars, but not for stars of Sun-like activity. In principle, the method can be combined with an analysis of the photospheric magnetic field through the use of Zeeman-sensitive lines. The method requires knowledge of the unperturbed profile, and is subject to inversion ambiguities that are inherently associated with the Fredholm equations that are to be inverted.
- *Interferometric imaging*. In principle, stellar sources are large enough to allow resolution of details on their surfaces by interferometric means. This method is currently most frequently used at radio wavelengths. Only a few optical studies have been performed on some cool supergiants to date (Baldwin, 1996), but new instrumentation will boost the potential. The ESO Very Large Telescope, for example, promises to resolve several hundred cool stars with at least a dozen pixels on their disks.
- *Direct imaging.* If the telescope and the apparent size of the star are large enough, some information can be obtained on the structure of the stellar surface by direct imaging. The number of stars for which this is feasible (e.g., Gilliland and Dupree, 1996) is very small, and these tend to be giant stars rather than main-sequence stars. The star α Ori is one of the largest in the night sky, and even imaging that star with 30 elements across the photospheric diameter would require a telescope with ten times the diameter of the Hubble Space Telescope. Nearby dwarf stars like the Sun would require mirror diameters of several hundred meters.

For stars with rotation periods of less than approximately a week useful information can be gathered by using (combinations of) these methods. But only very large baseline interferometric imaging provides the required latitude and longitude information on the magnetic fields of stars that rotate as slowly as the Sun.

Figure 1.34 shows the distribution of chromospheric activity (#occurrences vs. Ca II strength), while Figure 1.35 shows the various techniques described above for exploring the spatial structure of magnetic fields of stars.

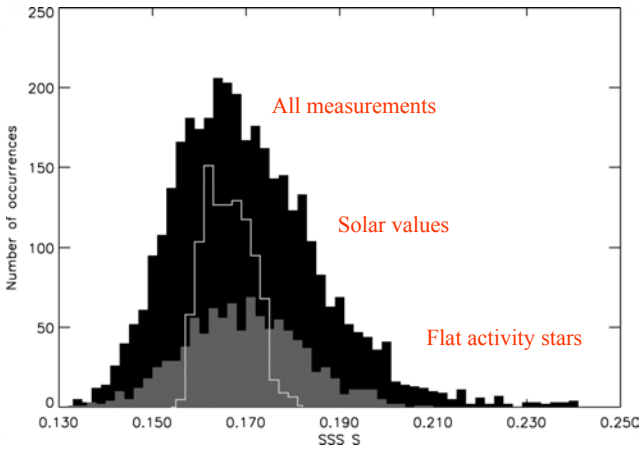


Figure 1.34: Stellar chromospheric activity distribution.

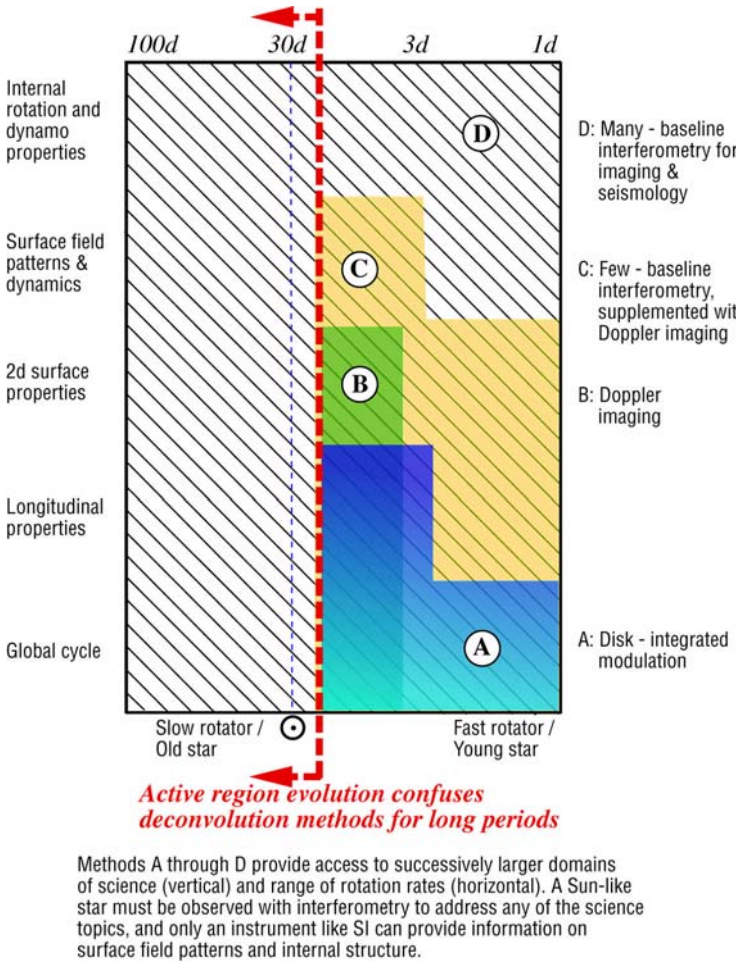


Figure 1.35: Many questions regarding geometrical information on surface flows and surface magnetic fields on cool stars like the Sun require some form of interferometric imaging.

Spectroscopic Doppler imaging can help in this area only by coarse mapping of stars that rotate neither too slowly nor too rapidly. Rotational modulation provides no latitudinal differentiation within the visible hemisphere.

1.3.2.2 Direct Measurement of Stellar Magnetic Fields

For stars of moderate activity, the magnetic field at the stellar surface can be measured directly by spectrometric means, because the broadening of Zeeman-sensitive spectral lines survives even in the

unpolarized disk-integrated intensity spectrum. The quantity that is most accurately determined with this method is the unsigned flux density, which is the product of the area filling factor and the intrinsic magnetic field strength (cf. Fig. 1.36). Field strength and filling factor can be determined separately provided that the non-magnetic line profile is known. That profile can, for example, be derived from a comparison with a similar but inactive star, or by using state-of-the-art model atmospheres.

Even with high-quality reference spectra available for the inactive state, however, magnetic filling factors remain subject to substantial uncertainties owing to the fact that assumptions must be made about the distribution of the field across the stellar surface, and about the differential visibility of field components such as spots, pores, and faculae as a function of wavelength, filling factor, and position on the stellar surface. Consequently, the determination of filling factors is subject to considerable systematic and random uncertainties. Moreover, these estimates can be made with reasonable accuracy only if a star rotates neither too slowly - so that the non-magnetic component dominates the stellar signal while active regions evolve before the star completes a full rotation - nor so fast that the Doppler smearing washes out the magnetic signal. In short, the method allows the mapping of fields on stars with rotation periods between 2 and 10 days.

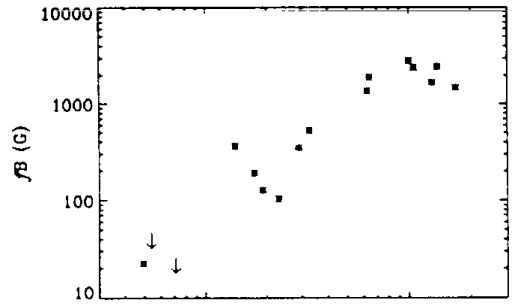


Figure 1.36: The measured average surface flux density of magnetic field versus the Rossby number in a sample of main–sequence stars.

The Rossby number (convective turnover time over rotation period measures the effectiveness of Coriolis forces. From Saar (1997)).

Currently available estimates of field strengths and filling factors confirm our expectations based on the solar example: the field strengths are a direct function of the photospheric properties, e.g., gas pressure, while the area coverage of the field is determined by the strength of the dynamo.

Line-profile studies of moderately active stars of spectral types G and early K have intrinsic field strengths B ranging from 1.0 to 1.9 kG; the very active Ke and Me stars (i.e. stars with chromospheric lines in emission over the nearby continuum) exhibit B up to 4.2 kG. These intrinsic field strengths are compatible with predictions from stellar atmospheric models based on the solar example, which have the field strength dictated by a balance between gas and magnetic pressures. For evolved giant stars, B decreases from B $\simeq$ 650 G for early G-type giants to 400 G at K5 III, in accord with the lower gas pressure at the surface owing to the much lower surface gravity of these stars.

The fractional surface coverage (or surface filling factor) f covers a much larger range than that of the field strengths, with observed values ranging from as low as ~ 1:5% (for the Sun, as the only star for which such a low value could be measured) up to 70% in active Ke and Me dwarf stars.

If a star rotates sufficiently rapidly, the features on the disk are subject to a differential Doppler effect that is large enough to allow some of the polarization associated with the opposite-polarity patches to survive in the disk-integrated line profile. This then allows Zeeman-Doppler imaging of at least the largest-scale magnetic structures (e.g., Donati, 1996). For this method to work, the rotation rate should be of the order of the thermal width of the line or larger. Consequently, the method is limited to rotation periods below about 5 days, i.e. significantly faster than the Sun. When this criterion is met, however, the method appears to be remarkably successful and accurate (Hussain *et al.,* 2000),

with a resolution claimed to be as good as 3 asterocentric degrees at the equator. So far, this has been achieved for only one star, however: AB Doradus, a K0 dwarf star with a rotation period of half a day, i.e. 50 times faster than the Sun.

Mapping results for rapid rotators can probably still be improved somewhat. For example, a combination of magnetically sensitive lines and insensitive lines should provide complementary information on differential contributions from spots and faculae. For stars with rotation periods of order a week and longer, however, the Zeeman signals are too small compared to the background photospheric signal to be measurable at all, setting a hard physical limitation to surface mapping by Doppler methods.

The ambiguities of inversion methods limit our ability for accurate mapping even for the most favorable of conditions, however. To demonstrate some of these limitations, we performed an experiment in which we generate a state-of-the-art model for the surface pattern of magnetic field, compute the expected signal for the star, and apply an inversion method. For this purpose, we apply the most advanced inversion method which uses the polarization signatures of activity rather than only the associated Zeeman broadening. This is possible because the position-dependent Doppler shift that is introduced for regions distributed over a visible hemisphere allows some polarization to survive. The method is very powerful in principle, because it allows us not only to map the position of magnetic flux, but also its polarity.

Our simulation (Fig. 1.37) demonstrates that the large-scale polarity pattern of the simulated star is recovered quite well, particularly the pattern near the rotational poles. But whereas a first impression of the smaller-scale patterns at lower latitudes on the simulated star may suggest a promising agreement, a detailed comparison demonstrates that the patterns show very little if any correlation. Therefore, we conclude that Zeeman Doppler signals do not allow an unambiguous translation of observed polarized spectra into surface maps even under the best of circumstances. Only direct or interferometric imaging can provide the unambiguous information on stars of a wide range of activity levels that is needed to test models of astrophysical dynamos.

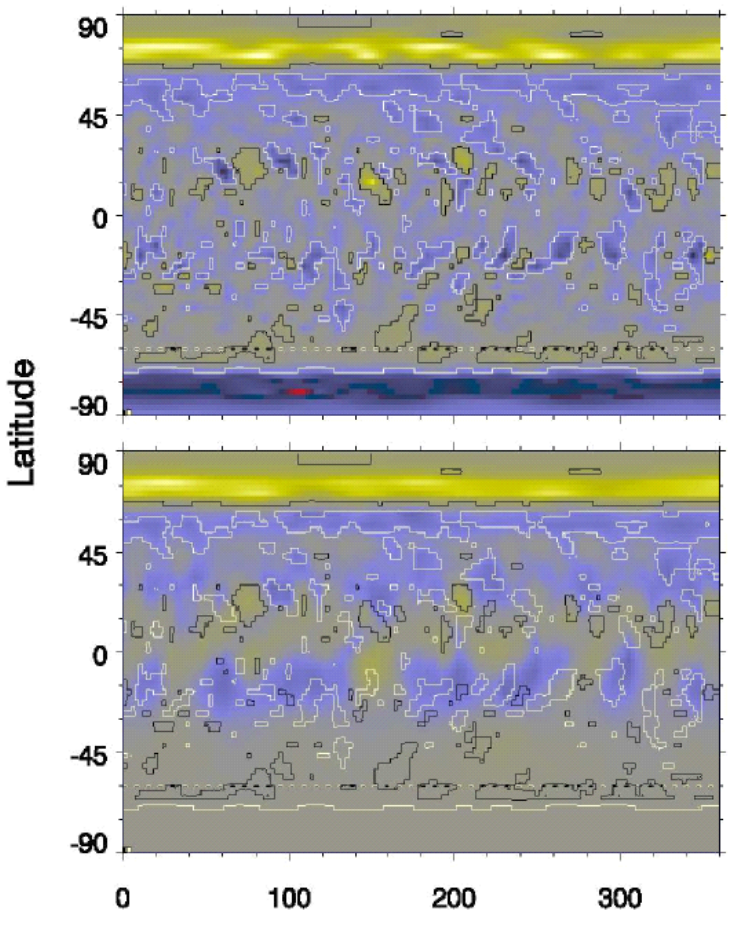


Figure 1.37: Top: Magnetic map of a simulated star. Bottom: Simulated reconstructed map based on Zeeman Doppler imaging (bottom) with I, Q, U, and V Stokes polarization, for a rotation velocity of $v \sin(i) = 10$ km/s and a mean spectral signal to noise ratio of 350.

The contours from the input map are copied onto this map. The polar-cap field, its surrounding ring, and the activity belt pattern are recovered quite well, also quantitatively, but even in this case for a static surface field during one full simulated rotation, the features on the scale of active regions are poorly mapped. Courtesy N. Piskunov

1.3.2.3 How to Achieve the Required Resolution of Activity Maps

Images of the activity in the stellar outer atmosphere can, in principle, be obtained by interferometry at radio wavelengths. The solar example (see sample radio images in, e.g., Dulk and Gary, 1983, and White, 1999) suggests that the most appropriate wavelength range for the imaging of a stellar corona would lie in the range of 30 cm to 3 m. At these wavelengths, the quiescent emission is dominated by free-free emission, i.e. thermal Bremsstrahlung; this has a cutoff frequency that corresponds to electron densities of 10^{10} cm^{-3} to 10^{8} cm^{-3}. This choice leaves the corona over the quiet network optically thin, while active-region coronae are optically thick. This ensures an optimal contrast between the weak-field and strong-field domains in the stellar coronae to be imaged.

If we assume that the required low sensitivity levels can in principle be attained to detect stars of Sun-like activity, ground-based VLBI imaging at these wavelengths is limited to an angular resolution of several milli-arcsec to several tens of milli-arcsec. What is, in fact, needed is a resolution of order a few tenths of a milli-arcsec. At a wavelength of ~15 cm, the required baseline would be of order 500,000 km, i.e. 1.3 times the distance to the Moon. Reconfiguration times of an interferometric array of a reasonable number of apertures would be far too large to effectively image a rotating star covered by an evolving magnetic field.

At shorter radio wavelengths, gyroresonant radio emission can be expected from regions of very strong fields, as is observed on the Sun for wavelengths as short as 2 cm. The associated baseline requirement still does not allow ground-based VLBI imaging. Moreover, the only detectable sources would be starspot fields, which do not offer an adequate proxy for the entire surface field, nor do they provide empirical information on the mechanisms of flux transport by differential rotation and meridional circulation, or on the random-walk dispersal driven by supergranulation. If the optically-thick chromospheric emission at relatively short radio wavelengths would be detectable separate from the gyroresonant emission over starspots, then at best a rather featureless image would be seen because the brightness temperature would be the same nearly everywhere. We conclude that radio observations in general are ill-suited to meet our scientific requirements of high-quality, high-resolution imaging of stellar activity.

The low contrast of faculae in the visible or infrared rules out those wavelength ranges as useful for mapping stellar magnetic fields across the surface, unless they were used for starspot studies in combination with some other diagnostic mostly sensitive for the field surrounding them.

Despite the fact that the contrast between active and non-active areas in visible, (E)UV, and X-rays tends to increase with increasing formation temperature characteristic of a diagnostic of activity, this does not necessarily make the soft X-ray flux the most useful for imaging purposes. The coronal emission often varies strongly, not only as a result of flares, but also because of changing interactions between active regions and because of the internal evolution of active-region coronae. This probably explains in part why rotational X-ray modulation studies have rarely resulted in useful information on the longitudinal distribution of activity on stars. The magnetic interaction between distant regions, and possibly between these regions and the magnetic poles of very active stars, may also play a role in the remarkable fact that stellar cycles in soft X-rays have yet to be observed: either there are no such cycles, or their amplitude is smaller than the scatter associated with rotation and active region evolution and the uncertainties on the relative calibrations of the multiple satellites--each with its unique pass band--that have been used for stellar cycle research thus far.

If a 5-m normal-incidence telescope could be built, tuned to a narrow X-ray passband around 15 Å, with a 1-km focal length to achieve the required angular resolution and plate scale on the detector, then that could image stellar coronae with the desired angular resolution. The emission at such wavelengths would be dominated by the active-region coronae, and therefore would form a suitable diagnostic to explore the input of magnetic field into the outer atmosphere in relatively inactive stars. In active stars, the coronal field would be expected to connect to all neighboring active regions, thus obfuscating the active-region spectrum as well as lowering the contrast between active region proper and surroundings. For stars of Sun-like activity, the corona over quiet regions might not be detectable in a narrow wavelength band because it is dominated by emission from plasma at a significantly lower temperature and typically emitting at longer wavelengths. If the passband could be made to be sensitive to plasma from 1 MK up to at least 4 MK, then the rather unstructured corona over quiet regions would, although now faintly observable, not provide adequate information on the evolution of the underlying field in response to the large-scale surface flows.

We conclude that the diagnostics that are best suited for the study of patterns in the surface magnetic field lie in the (near) ultraviolet. Best would be a combination of imaging of starspots in a photospheric continuum band as well as in either the strong chromospheric Mg II h&k lines near 2800 Å (with high-contrast between active and inactive regions, and with a relatively weak photospheric contribution) or in the transition-region C IV doublet at 1,550 Å (with little background continuum except in the warmest of the cool stars). The latter diagnostic would show the largest contrast between inactive and active regions, particularly if the bandpass is relatively narrow, because the lines lie on a

weak continuum for all stars cooler than spectral type mid F. The short wavelength would also reduce the number of sources in the immediate vicinity of the star to be imaged, which is important to reduce the ambiguities in the image reconstruction process. For the C IV 1550 Å doublet formed in the transition region, a baseline of order 500 m would suffice to image a star like the Sun at 4pc with 1,000 resolution elements on its surface, as detailed in Sect. 7.1.1.4.

1.3.2.4 Assessing the Impact of the Interstellar Medium on Sub-mas UV Imaging

As UV radiation passes through the interstellar medium (ISM), it interacts with any dust and gas that may be present. While scattering out of a sightline may limit which sources are bright enough to be observed, scattering back into the sightline creates image halos. These halos could, in principle, limit the attainable spatial resolution of Stellar Imager (SI). We have examined the limitations imposed by ultraviolet halos that result from refractive scintillation from clouds of neutral and molecular hydrogen, and small-angle single scattering from hydrogen and diffuse interstellar dust.

Refractive scintillation occurs when photons travel through inhomogeneous media where the refractive indices vary with position. Light traversing slightly different paths between the source and SI will pass through different gaseous components of the ISM and accumulate different phase shifts. If there are many such "cloudlets" along the sightline, each with $N_H > 10^{18}$ cm^{-2} for scintillation of C IV, the characteristic scattering angle can be found using the "thin phase screen model" (Scheuer 1968 Nature 218, 920). For a kiloparsec sightline with a typical "cloudlet" density of n_H=3 cm^{-3} which is correlated on scales of 0.1 parsec, the angular broadening of a point source is extremely small, i.e., ~10^{-9} μarcsec compared to the proposed SI 30 μarcsec resolution. Orders of magnitude changes in the adopted parameters still lead to negligible broadening. (A discussion of different scintillation regimes can be found in Moniez (2003).). While the physical properties of the inhomogeneous ISM are not well constrained, especially on the smaller scales, the ISM is unlikely to limit SI's spatial resolution.

Dust can exhibit strong forward scattering which can create image halos. Small angle dust scattering is known to be important for interpreting X-ray images where arcminute halos are observed around bright embedded sources. The halos of UV sources that are not heavily obscured can be examined in the single-scattering approximation (Mathis & Lee 1991). The ratio of the halo intensity to the source flux is essentially independent of distance and source size, but is controlled by the relative angular scattering efficiency. Scattering by diffuse interstellar dust and neutral hydrogen are potential sources of image halos. The diffuse dust angular scattering phase function has recently been derived by Schiminovich et al. (2001 ApJ 563, L161) from 1740 Å NUVIEWS far-UV imaging measurements. [These have an albedo a= 0.45 and a Henyey-Greenstein (1941) scattering asymmetry parameter of g=0.77.]

Figure 1.38 shows the ratio of halo intensity to source flux for uniform distributions of interstellar dust and hydrogen with a total optical depth of unity towards the source. There is assumed to be a small region that is free of dust and neutral hydrogen surrounding the source, e.g., stellar coronae. The source/halo contrast is very large. The scattering intensity for very small angles depends sensitively on the material located near the source where the source illumination is strongest, rather than on the long path-length through the ISM. If the scattering material is concentrated near the source the width of the halo decreases while its brightness increases. One corollary is that for UV bright targets in external galaxies the ISM of the host galaxy will act as a small angle scattering source leading to a smaller and brighter halo. The dust scattering is much more forward-peaked and consequently has a narrower and brighter halo than hydrogen. The source/halo contrast is so large that even if the source were 1/30th of an SI image pixel the adjacent halo image pixel would still be nine orders of magnitude fainter. If the

optical depth towards the source is much greater than unity, where the source/halo contrast declines exponentially, the UV source will be faint and not suited for prime SI science observations.

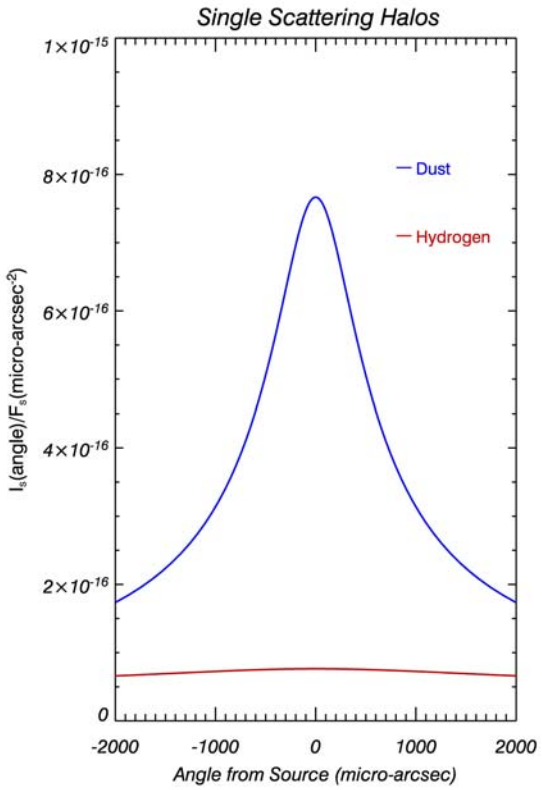


Figure 1.38: Halo intensity to flux ratio vs. angle from source in the ISM

Conclusions:

At the 50 μarcsec resolution (@ Ly α) planned for Stellar Imager, refractive scintillation from cloudlets of neutral or molecular hydrogen and small-angle scattering from hydrogen and dust with optical depths of order unity will not limit the attainable image resolution.

1.3.2.5 Impact of Galactic/Extra-Galactic Background in the UV

SI will be viewing its science targets against a background of scattered Galactic and, at high Galactic latitudes, extragalactic UV photons (Schiminovich et al. 2001). At the same time, SI requires that the background counts per image pixel be negligible compared to that from the science target for successful image reconstruction. The FUV/NUV background from the fields of view surrounding each mirrorsat does, however, provide a non-negligible contribution to the total count rates and this is mitigated by the use of entrance aperture masks.

The FUV background is dominated by dust scattered starlight, and the positional variations depend on the location of the dust and illuminating UV bright stars - the intensity being correlated with the hydrogen column density (Paresce & Jakobsen 1980). Other contributions to the FUV background come from H_2 fluorescence, H I 2-photon emission and line emission from the warm and hot ISM. The

FUV count rates can vary from 300-2,000 photons cm^{-2} sec^{-1} $Å^{-1}$ sr^{-1} (from information provided by the GALEX web site) to as high as ~300,000 at 1550 Å. (from unpublished results from the SPEAR instrument).

A typical science source will have an angular diameter of 1~mas and a total source flux at C~IV 1550 Å of $5x10^{-14}$ erg cm^{-2} sec^{-1} which corresponds to $3.8x10^{-3}$ photons cm^{-2} sec^{-1}. Assuming the upper range of the FUV background and a 10 Angstrom passband this corresponds to $7x10^{-11}$ photons cm^{-2} sec^{-1} for the same solid angle, which is totally negligible. The stellar source will be surrounded by a background which is unlikely to be detectable.

The NUV background can also reach 100,000 photons cm^{-2} sec^{-1} $Å^{-1}$ sr^{-1}. This NUV emission is normally dominated by zodiacal light with a contribution from diffuse Galactic emission. The strength is strongly dependent on the position of the Sun, target, and ecliptic plane (Leinert et al. 1999). Since SI will not observe targets at angles close to the Sun [the nominal SI observing will be at the anti-solar angle range: 70-110 degrees] the NUV background will also pose no limitation for SI.

The very high angular resolution of SI will also resolve out any FUV and NUV background contributions from numerous faint sources which remain unresolved in present day observations. The very small angular scale of the science targets ensure that the Galactic FUV and NUV background will be negligible compared to the science targets brightness.

1.3.2.6 Mapping flows below the stellar surface

The principle unknowns about stellar interiors that need to be established in order to understand stellar magnetic activity are the large-scale transport mechanisms for the magnetic field, i.e., the internal differential rotation and meridional flow. The Sun's internal differential rotation is measured using helioseismic means: the higher-order acoustic modes that resonate within the Sun are slightly altered in frequency as seen from the Earth because of the Doppler effects associated with the Sun's rotation. Note that the modes with degrees $\ell = 0, 1, 2, 3$ can be measured in integrated star light; these provide access to the mean rotation period (as measured by precursor missions such as COROT and KEPLER), but that measurement of differential rotation within the convective envelopes necessarily requires spatial resolution to measure modes well above $\ell \sim 10$.

Using the differential paths probed by modes of different order and degree, the latitude and depth dependence of the internal differential rotation can be accurately determined (Fig. 1.39). The corresponding signature of the meridional flow is minute, but has recently been successfully measured using careful analysis of the many modes observed to resonate within the Sun.

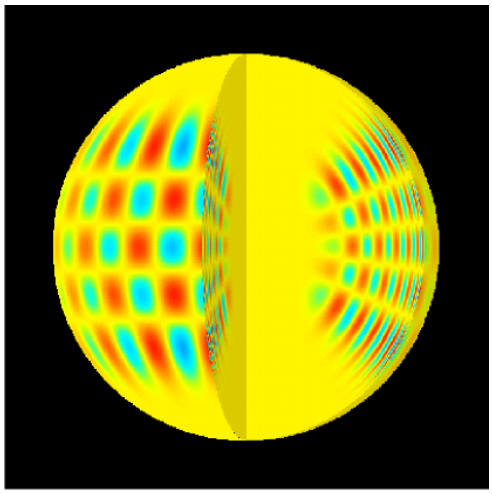


Figure 1.39: Astero-seismology, the analysis of standing sound waves within stars, allows mapping of the internal structure and dynamics.

For stars, the acoustic signature of meridional circulation will remain out of reach even for the Stellar Imager (although it will hopefully be visible in the evolution of the magnetic activity pattern), but the internal differential rotation can be determined by observing the *p*-modes in the intensity signal at stellar surfaces. This signal can be determined from measurements taken well within the characteristic period of the modes; for the Sun, a 1-minute spacing between measurements is required to measure the 3-8 minute period *p* modes. The theory of stellar interiors shows that stellar oscillation periods Π approximately scale as the inverse square root of the mean density of the star, $\Pi \propto 1/(M/R^3)^{-1/2}$, for stellar radius *R* and mass *M*. Thus typical periods of solar-like stellar *p* modes are

$$\Pi \cong 300s\left[\frac{R_*}{R_{sun}}\right]^{3/2}\left[\frac{M_{sun}}{M_*}\right]^{1/2} \qquad (1.1)$$

expressed in solar units R_{sun} and M_{sun}. From this it appears that a minimal cadence of 1 minute is adequate for all cool stars along the main sequence, while a slower cadence suffices for increasingly larger evolved stars.

With n elements in the interferometer, we can observe $n(n\text{-}1)/2$ baselines at a given wavelength. In principle, an interferometer can observe multiple wavelengths, providing additional baselines if the wavelengths are different enough. For a *k*-channel interferometer, there are thus in total $kn(n\text{-}1)/2$ baselines being recorded. An essentially linear array that is positioned perpendicular to a stellar rotation axis can therefore measure sectoral modes up to $m \approx kn(n\text{-}1)/4$, each of which has a splitting that is *m* times the characteristic rotational splitting of $1/P$. In order to resolve differential rotation to a fraction *f* of the stellar rotation period, sampling with a frequency resolution of mf/P Hz, requires an observing interval spanning $\Delta t \approx 4P/(fkn\ (n - 1))$ s. In the Sun, the equatorial differential rotation is ~1/10 of the mean rotation rate. Requiring a resolution for the stellar imager that covers this in, e.g., 5 resolution elements (or $f \sim 1/50$), and with $n = 10$, $k = 3$, and $P \sim$ 25d, yields an integration time of $\Delta t =$ 19d.

As the Stellar Imager will be observing stars in general for at least one full rotation period in order to image the entire surface, the number of mirror elements *n* in the interferometer must be $n \sim 2/\sqrt{fk}$

in order to achieve a resolution of a fraction f of the rotation rate for the differential rotation profile with depth.

The asteroseismic resolution that can be achieved at a given location within a star approximately equals the local wavelength $\lambda = c_s/\nu$, where c_s is the sound speed and ν is the cyclic frequency. Thus the resolution improves from the stellar center to the surface as the sound speed decreases. The best resolution is obtained at the lower turning depth of the most shallowly penetrating modes for given ν, i.e., those with highest m. At this location $mc_s/r = \omega = 2\pi\nu$. Thus the resolution is approximately $\lambda = 2\pi r/m \approx 2\pi R_*/m$, where the last approximation holds in the outermost layers of the star. For a Stellar Imager with $n = 9$ and $k = 3$, we thus find a depth resolution of $2\pi R_*/m \sim 81{,}000$ km, or 40% of the depth of the convective envelope in a star like the Sun. Hence differential rotation with depth within the envelope can just be constrained, but any difference between envelope and interior can be measured quite accurately. For a Stellar Imager with $n = 17$ and $k = 3$, the resolution is 10% of the depth of the convective envelope for a star like the Sun, so that the differential rotation can be mapped accurately with depth even within the envelope.

Within the total observing period, the net fraction of the time spent on the target star must be of the order of 50% or more, although in principle alternating intervals of some 12 h on a pair of stars nearby on the sky would suffice, thus doubling the number of stars that can be studied in this way. Increasing the number of interferometer elements allows a shorter integration period needed to measure internal rotation rate, although it must necessarily remain a substantial fraction of the rotation period in order to be able to separate the frequencies in Fourier space. Hence, it appears that the number of elements should at least be 10 from the viewpoint of non-radial asteroseismology studies.

Future trade-off studies will need to be performed to balance the scientific needs with the observing possibilities, for example to assess how best to position the mirrors of the interferometer to obtain latitudinal as well as depth information on the internal differential rotation.

1.4 Mission Requirements Derived from Science Goals

1.4.1 Stellar surface activity

1.4.1.1 Scientific requirements

To understand the dynamo, we need to know how magnetic fields are generated and how they behave in different circumstances. The Sun is only one example, which provides insufficient constraints on theories of dynamos, turbulence, structure, and internal mixing. We must observe other stars, i.e. perform a *Population Study,* to establish how mass, rotation, brightness and age affect the patterns of activity. The goal of this population study would be to provide strict observational constraints on dynamo behavior, in order to guide dynamo research by answering questions that include:

- what determines cycle strength and duration?
- how common is solar-like activity?
- can multiple cycles exist at the surface?
- what are Maunder-minimum states like?
- how do polar spots form?
- what are extremely active and inactive stars like?

In order to characterize these "patterns of activity" we need sufficient spatial resolution across the surface of the target stars and temporal resolution and sampling to be able to detect and follow an average-sized region of surface magnetic activity across the disk of the star, and asteroseismic measurements to determine the largest-scale flows within stars that power their activity. We address these requirements in the following sections.

The **primary wavelength bands** must enable imaging of both starspots and of surrounding plage and enhanced network. Imaging of starspots embedded within the stellar photosphere requires access to the stellar photospheric continuum, i.e., to visible wavelengths for the stars of interest. Imaging of the magnetic regions across the stellar disk is best achieved at UV wavelengths blueward of the bulk of the photospheric continuum, particularly at the high-contrast C IV doublet around 1550 Å emitted by transition-region plasma, or in the chromospheric Mg II h & k lines around 2800 Å.

The **spatial resolution for activity imaging** must resolve structures down to approximately 50,000 km. At such a resolution, Stellar Imager can image large active regions, map the regions of decaying active regions, as well as resolve some of the largest starspots or starspot clusters that occur on young, active stars. This resolution corresponds to 30 pixels across the equator of a Sun-like star, or of order 700 pixels on the stellar surface. For a star at 4pc, that resolution is achieved with an angular resolution of order 0.1 mas (milli-arcsecond).

The **sample of target stars** must include at least two dozen stars like the Sun, i.e. main-sequence stars of moderate activity. The sample should also include at least several of each of the following categories: active binary stars, Herztsprung-gap giants, evolved (super)giants, and young main-sequence stars.

The **time to complete a full image** must be short enough so as to not introduce rotational smearing at a resolution of 50,000 km on a star like the Sun. For a G2 main-sequence star with a rotation period P(days), the imaging time must therefore be less than 0.3 *P* hours, or 7 hours at *P*= 25 d, and 1.4 hours at *P*= 5d.

The **number of pointings per stellar rotation** must enable the measurement of differential rotation within the time scale for substantial active-region evolution, while providing sufficient separate views to accurately map the entire stellar surface. The requires at least six pairs of images, taken at rotation phases approximately 60 degrees apart, with pairs less than the lower of 20 degrees or three days apart.

The **mission duration** must allow observations that span a significant fraction of stellar activity cycles, which range from 8 to 23 years. An absolute minimum requirement of 5 years, with a goal of 10 years or more, is thus indicated.

The **number of full maps during the mission** must be large enough to evaluate the change in the flux-emergence patterns. For a star like the Sun, at least ten visits in an 11-year sunspot period would be required for that purpose.

1.4.1.2 Required Angular Resolution

For a star like the Sun, a surface resolution is required of approximately $\Delta r \sim 50{,}000$ km, comparable to an active region of average size. For a star with diameter D_{star}, the apparent size of a star is

$$\theta(\mu arc\,\mathrm{sec}) = 9345 \times \frac{D_{star}/D_{sun}}{d_{pc}} \qquad (1.2)$$

Hence, for a Sun-like star at 4 pc, the required instrumental angular resolution is

$$\alpha = \frac{\Delta r}{D_{star}}\theta \qquad (1.3)$$

or 83 microarcsec.

An angular resolution α (microarcsec) requires a baseline of b_{max} (m)

$$b_{max}(\mathrm{m}) = 21\,\frac{\lambda}{\alpha} \quad \text{with } \lambda \text{ in Å.} \qquad (1.4)$$

For $\alpha = 83$, $b_{max} = 390$m at 1550Å, or $b_{max} = 700$m at 2800Å. As to be expected, the required baseline is shorter if one can observe at shorter wavelengths. Observations in the light of the C IV lines at 1550Å would therefore allow shorter baselines as well as provide the best sensitivity to the hot plasma ($\approx 10^5$ K) in active regions. However, observations in the light from the Mg II doublet at 2800Å are still required to sample lower temperature ($\sim 10^4$ K) material in the active regions and longer baselines associated with those observations will need to be available. In practice, we will probably need to compromise slightly, aiming for baselines ~500m, which will provide resolution slightly above

specification for the C IV observations, and slightly below for the Mg II observations, since reconfiguring the array between observations at different wavelengths would be prohibitive in time and in fuel.

1.4.1.3 Required Field of View and Detector Elements

The field of view, f, of the interferometer is dictated by the Nyquist frequency N = λ/*2a*, where *a* is the aperture of the interferometer elements, and the number n_{pix} of pixels across the detector:

$$f = \frac{\lambda}{2a} n_{pix} \text{ radians,} \tag{1.5}$$

or

$$n_{pix} = 0.063a \frac{1550}{\lambda} f, \tag{1.6}$$

for *a* in m, and *f* in milliarcsec. For a solar type star at a typical distance of 4 pc, our requirement for resolving active regions translates to an angular resolution ~0.1 milliarcsec, and provides approximately 36 resolution elements across the diameter of the star (≈1000 total over the full disk). The field of view needed to contain the stellar disk would need to be at least 36x0.100 mas = 3-4 mas and require ~252 pixels across the detector. For substantially larger targets, such as the supergiant α Ori and the Capella binary system (~ 50 mas), a factor of ten increase in field of view and number of pixels would be required. In fact, to observe these larger targets at all, the baselines would have to be reduced, and the resolution decreased, to enable sufficient sampling of the Fourier UV-plane in a reasonable amount of time.

However, the interference pattern produced on the detector by SI's array extends over a much broader area than the stellar disk subtends (see Figure E.2), so that much larger detectors (about 6 K x 6 K pixels) must be used in order to ensure that most of the light from the target is detected and system efficiency kept high.

1.4.1.4 Limitations on Total Observation Times

The time interval within which the image synthesis has to be completed may be limited by the stellar rotation, the proper motions of nearby targets, and intrinsic variability of the stars.

For example, for *n* resolution elements across the equator of the stellar disk, on a star that rotates with a period *P*, the rotational smearing reaches one resolution element at disk center (and increasingly less than that with increasing distance from disk center) for a total integration time t_{obs} of

$$t_{obs} = \frac{P}{\pi n} = \frac{P\Delta r}{\pi D_*} \tag{1.7}$$

For a star with a rotation period of 25d, and a resolution of Δ*r* = 50,000 km, the maximum allowed integration time equals 7 hrs.

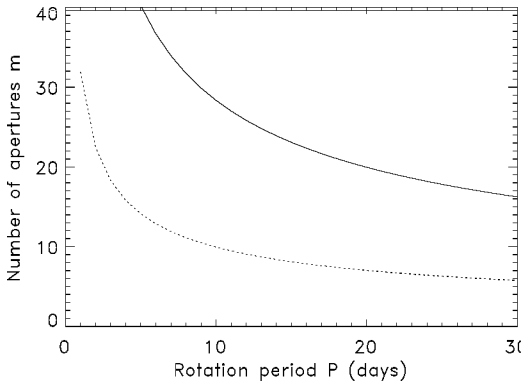


Figure 1.40: Required number of apertures to complete a full image synthesis fast enough to avoid rotation smearing.

The number of apertures must exceed the solid curve for N_{pix} = 1000 and a configuration plus exposure time of 1 hour, and the dotted line for N_{pix} = 400 and a configuration plus exposure time of 0.5 hour.

To collect enough information to produce an image, a sparse aperture interferometer needs to make observations using a wide range of baseline lengths and orientations. The arrangement of mirrorsats either needs enough apertures to give sufficient baseline sampling with a single configuration or will need to be reconfigured to gather the needed measurements. The number of configurations n_{con} required to synthesize an image containing N_{pix} pixels for an interferometer with m apertures is

$$n_{con} = \frac{2N_{pix}}{m(m-1) - 4} \tag{1.8}$$

If we assume that the time to reconfigure is t_{con} and the exposure time per configuration is t_{exp}, then the total time required to complete an image is less than the rotational-smearing time only if

$$\frac{(t_{con} + t_{exp})}{m(m\text{-}1) - 4} \leq \frac{P}{2\pi\,(N)^{3/2}} \tag{1.9}$$

For N_{pix} = 1000, and a total time of one hour per reconfiguration, for example, there should be at least 17 apertures for a star with a rotation period comparable to that of the Sun, and at least 40 apertures for a star with a rotation period of 5 days. That number scales inversely with the square root of $t_{con} + t_{exp}$.

The rotation rate of a star like the Sun limits the total integration time to approximately one day for the full-resolution aperture-synthesis imaging. For more rapidly rotating stars, the total exposure duration should not exceed 4-6 hrs. For supergiant convection studies, exposures of up to several days are in principle allowed (allowing extremely high-resolution studies by continual interferometer reconfigurations).

The rotation dominates the smearing for most objects on this timescale, except perhaps for the stars with the highest proper motions. Intrinsic variability of the active regions is probably not a concern, aside from flare events, on timescales shorter than a day.

It may be possible to loosen this restriction on total integration times if the image reconstruction algorithms are sophisticated enough to incorporate corrections for stellar rotation and other target motions which occur during the integrations.

1.4.1.5 Required Duration of Mission

The study of cycle dynamics such as the migration patterns of the solar activity belts in the so-called butterfly diagram require a mission baseline of at least 5 years, though we set 10 years as an unrequired "goal" for the mission, with repeat visits up to a few times a year for selected targets.

1.4.1.6 Sample Targets

SI needs to be able to image a significant sample of stars similar to the Sun, as well as a sample of other cool stars characteristic of different classes of magnetically active stars. There are 72 main-sequence star systems (single and binary stars) within 6.5 parsecs of the Sun. If we require a resolution comparable to a medium-sized active region on the Sun, for a typical distance of 4 parsecs, then a maximum baseline follows of 500 meters when observing in the mid-UV, as discussed above. This corresponds to a dwarf star with an angular diameter of 2 mas. For imaging smaller stars with diameters of 1 mas, of which there are many, longer baselines up to 1 km will be necessary.

Taking a maximum baseline of 500 meters (more compact and more extended configurations are possible), the potential targets include a variety of magnetically active stars, as shown in Fig. 1.41. The apparently larger objects, such as α Ori, would actually be observed with a condensed (i.e., shorter baseline array), since the number of configurations and total observing time needed to properly sample a star with 600 resolution elements across its diameter (more than 300,000 pixels total!), using even a 30 mirror element array, would be prohibitive.

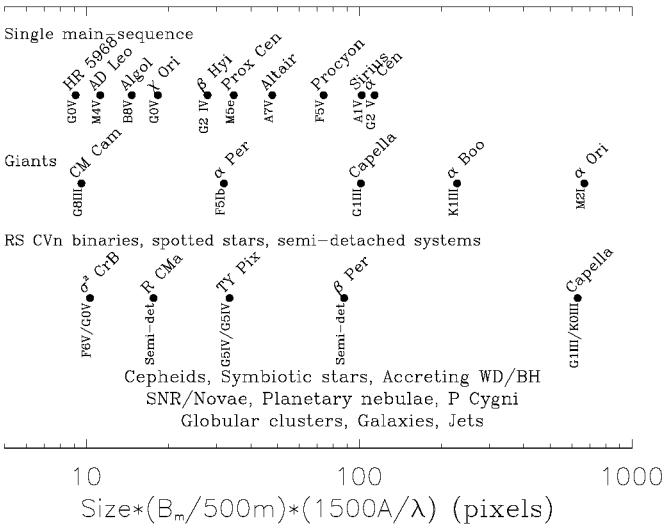


Figure 1.41: Sample targets for SI.

The horizontal axis shows the number of resolution elements across the equator for single stars or across the system at quadrature for binary stars, assuming a baseline of 500 meters observing at 1,500 Å.

A wide variety of potential target stars is available to investigate the science topics described earlier in this report; Table 1.6 contains a set of stars for which SI observations would be suitable to address the questions related to solar-like magnetic activity, while Table 1.7 contains a more diverse set of targets addressing other issues. The four sections of Table 1.6 show how different samples might be constructed with i) a number of wide (few to tens of arcseconds) binaries that would permit very efficient observation of multiple targets with the minimum of slewing, ii) a sample of stars with known planetary systems, iii) a group of single stars spanning the late F–G– K-early M spectral range, and iv) a set of stars that probe the evolution of solar-like activity.

These are interesting solar counterparts that span a range of ages and activity conditions:

- There are numerous solar analogs that explore solar-like activity for a range of metallicities (e.g. α Cen A) or different evolutionary states (e.g. χ Ori and β Hyi) for a solar mass star.
- Similarly there are many targets that would explore the changes in magnetic activity as the convection zone depth increases, all the way to the fully-convective mid-M dwarfs.
- ρ CrB (HR 5968) is the nearest of the solar analogs that is likely to be in a Maunder minimum, as well as hosting a planetary system: its activity level is low, nearly constant, yet its age and rotation rate suggest it should be as active as the Sun.

Other sample targets, shown in Table 1.7, extend the range of stellar parameters:

- Red giant stars, like α Boo, α Tau, and β Gem, gives us a view of the Sun in the distant future: it has expanded to become a giant star, having enveloped any nearby planets it might have had
- Altair, Procyon, and α Per are warm stars, with shallow convective envelopes, and a strongly subdued dynamo activity given their rotation rate. The shallow convective envelope is thought to generate only small active regions, but that conjecture remains to be tested.
- Cepheid variables, such as δ Cep, η Aql, and Polaris, can be both imaged and their pulsational spectrum explored. Imaging the size changes of Cepheids has important implications for the accuracy of the galactic and extragalactic distance scales.
- AD Leo and Proxima Centauri are small, cool stars, that are nearly fully convective. They are important testing grounds for the importance of the overshoot layer just below the convective envelope in the dynamo process
- Capella, HR 1099, and σ^2 CrB are close binaries, with both components active, and magnetically coupled at only a few stellar radii. The separation of the Capella system is so large that it will allow imaging of any chromospheric and transition-region material that exists between the stars. Dynamos in such tidally-interacting binaries are up to two orders of magnitude more active then expected for a single star of similar rotation rate; SI will study how the tidal interaction affects the dynamo action
- TY Pyx is a compact RS-CVn-type binary in which the stars revolve around each other only two stellar radii apart; tidal effects are expected to change differential rotation and meridional circulation, and therefore the dynamo.
- R CMa and β Per are examples of semi-detached eclipsing binaries in which the evolving primary fills its potential envelope: mass extending above it falls towards its companion
- α Ori is a supergiant star. It is not expected to show much activity, but its size makes it an exceptional target to study convective patterns on a star. In this case, its low surface gravity is expected to result on only a handful of convective cells (granules) on its surface, while the Sun is covered by tens of millions of such granules
- Algol is an example of a binary in which mass is transferred from one to the other component
- Sirius is a hot star, not magnetically active, significant as a reference star for cool-star studies

Observations in a variety of bandpasses are needed to record emission from various elements in plasmas of different temperatures. Thus a minimum set of filters/bandpasses would be comprised of ~10 Å-wide UV passbands including: the C IV doublet at 1550 Å (characteristic of 100,000 K plasmas) and Mg II doublet at 2800 Å (10,000 K), plus bands other lines such as Ly a 1216 Å, O I 1304 Å, C II 1335 Å, Si IV 1400 Å, and He II 1640 Å. Observations in broadband, near UV or optical continuum passbands (~100Å wide) representative of plasmas at 3,000 K to 10,000 K would be used for asteroseismology.

All these targets can be studied with close to one hundred up to potentially many thousands of pixels on their surfaces. The range of angular scales that must be explored implies that SI and its Pathfinder mission must be adaptable and capable of reconfiguration to match the baseline lengths employed to the size of the target and the wavelength of observation.

Table 1.6: Characteristics of Potential Sun-like Science Targets for Stellar Imager.

Target star	Target Description	V-Mag	Dist. (pc)	[Fe/H]	Rotation Period (days)	Apparent Angular Size (mas)	# pixels (500 m baseline)
α Cen A	G2 V -supersolar abund.	-0.1	1.33	0.15	25.1	8.5	13,400
α Cen B	K1 V -supersolar abund	1.3	1.33	0.15	42.	6.0	6,700
Prox Cen	M5 Ve - Fully-convective	11.1	1.31	0.15	31.5	1.1	225
70 Oph A	K0 V -- subsolar abund.	4.0	5.1	-0.29	19.9	(1.8)	600
70 Oph B	K4 V -- subsolar abund.	6.0	5.1	-0.29	---	(1.6)	475
36 Oph A	K0 V -- subsolar abund.	5.1	6.0	-0.30	20.7	(1.2)	270
36 Oph A	K1 V -- subsolar abund.	5.1	6.0	-0.30	21.1	(1.2)	270
61 Cyg A	K5 V -- subsolar abund	5.2	3.5	-0.37	34.7	2.0	740
61 Cyg B	K7 V -- subsolar abund	6.0	3.5	(-0.37)	40.9	1.9	670
ε Eri	K2 V – Planet host	3.7	3.2	0.00	11.7	2.2	900
γ Cep	K1 IV – Planet host	3.2	13.8	---	---	3.6	2,400
54 Psc	K0 V – Planet host	5.9	11.1	0.19	48.	1.0	190
70 Vir	G6 V – Planet host	5.0	18.1	-0.09	35.	1.0	190
47 UMa	G1 V – Planet host	5.1	14.1	0.01	---	(0.8)	120
ρ CrB (HR 5968)	G0 V -- Maunder-min. star	5.4	17.4	-0.21	---	0.6	67
υ And	F8 V – Planet host	4.1	13.5	0.00	---	1.2	270
ι Peg	F5 V -- Single	3.8	11.8	-0.08	---	1.1	225
ι Per	G0 V -- Single	4.1	10.5	0.16	14.5	1.3	315
λ Ser	G0 V -- Single	4.4	11.8	0.05	18.0	(1.1)	225
κ^1 Cet	G5 V -- Single	4.8	9.2	0.06	9.2	1.0	190
61 Vir	G6 V -- Single	4.7	8.5	0.04	28.0	(1.1)	225
σ Dra	K0 V -- Single	4.7	5.8	-0.29	27.0	1.3	315
o^2 Eri	K1 V -- Single	4.4	5.0	-0.05	43.0	1.6	475
ε Ind	K5 V -- Single	4.7	3.6	-0.06	22.0	1.9	670
GJ 887	M0.5 V -- Single	7.3	3.3	-0.22	---	1.4	365
GJ 411	M1.5 V -- Single	7.5	2.5	-0.42	48.	1.4	365
GJ 699 (Barnard's *)	M4 V -- Single	9.5	1.8	---	---	1.0	190
GJ 388 (AD Leo)	M4 V -- Flare star	9.4	4.7	---	2.7	0.7	90
β Com	G0 V – 1.6 Gyr "Sun"	4.3	9.2	0.00	12.	1.2	270
χ^1 Ori	G0 V -- Young (0.3 Gyr) Sun	4.4	8.7	-0.06	5.2	1.1	225
π^1 UMa	G1.5 V--Young (0.3 Gyr) Sun	5.6	14.3	-0.20	4.9	(0.7)	90
AB Dor	K0 V - Young (0.01 Gyr) Sun	6.9	14.9	0.18	0.5	(0.7)	90
ϑ Cet	G8 V – "Old (7 Gyr) Sun"	3.5	3.6	-0.42	---	2.1	820
β Hyi	G2 IV -- "Old Sun "	2.8	6.5	0.01	28.	1.7	540
α Aql (Altair)	Onset activity (A7 IV-V)	0.8	5.1	---	---	3.5-3.0	1,650
α CMi (Procyon)	Onset activity (F5 IV-V)	0.3	3.5	-0.05	---	5.5	5,600

Table 1.7: *Potential Non-Sun-like Targets for Stellar Imager mission.*

The final column indicates targets for imaging (I), seismology(S), or both (B)

Target star or system	Target Description	V-Mag	Dist. (pc)	Apparent Angular Size (mas)	S
ζ Oph	O star (O9 V)	2.6	140	1.0	I
ε Ori	B supergiant (B0 Iab)	1.7	400	1.1	I
β Ori	B supergiant (B8 Iab)	0.1	240	2.8	I
β Per (Algol)	Semi-detached binary	2.1	28.5	5.5	I
Algol	Primary star (B8V)	2.1	28.5	1.1	I
γ^2 Vel (WR11)	Orbital separation (WC8+O8 III)	1.8	260	5.4	I
γ^2 Vel (WR11)	Nearest Wolf-Rayet star	1.8	260	0.18	---
α CMa (Sirius)	A dwarf (A1V)	-1.5	2.7	6.0	B
α Cyg (Deneb)	A supergiant (A2 Ia)	1.2	2.7	2.4	B
R CMa	Algol-binary (F2 III+)	5.7	44	1.1	I
α Per	Onset activity (F5 Ib)	1.8	150	2.0	B
α UMi (Polaris)	Cepheid (F7 Ib)	2.0	130	3.2	B
η Aql	Cepheid (F6 Ib)	3.9	360	1.8	B
δ Cep	Cepheid (F5 Iab)	4.1	300	1.6	B
β Cet	K-giant star (K0III)	2.0	29.4	6.7	B
β Gem	K-giant star (K0III)	1.1	10.7	8.0	B
α Boo	"Ancient Sun" (K1III)	-0.0	11.3	21.4	B
α Tau	K-giant star (K5III)	0.9	20.0	21.1	B
α Ori	M-Supergiant star (M2Iab)	0.5	130	45 [optical]	S
α Ori	M-Supergiant star (M2Iab)	0.5	130	125 [UV]	I
o Cet (Mira)	Mira Variable (M7 III)	3.0	130	8.3	B
R Leo	Mira variable (M8 III)	6.0	100	38	B
α TrA	Hybrid star (K2 II)	1.9	130	9.0	B
γ Aql	Hybrid star (K3 II)	2.7	140	7.1	B
α Aur (Capella)	G1III/K0III binary system	0.1	13.0	55.	I
Capella Primary star	G1 III component	0.1	13.0	5.8	B
HR1099	Active binary (K1 IV primary)	5.7	29.0	1.2	B
σ Gem	Active binary (K0 III primary)	4.3	37.5	2.3	B
λ And	Active binary (G8 III primary)	3.8	25.8	3.1	B
σ^2 CrB	Active binary (F6V/G0V)	5.6	21.7	0.6	I
TY Pyx	Active binary (G5IV/G5IV)	6.9	55.	2.1	I

1.4.2 Internal Stellar Structure, Convection, and Rotation

1.4.2.1 Scientific Requirements

In order to meet the scientific goals related to internal structure and dynamics, SI must be capable of asteroseismology, in which low to intermediate degree non-radial modes (detected via high time cadence intensity fluctuations on the surface) are used to measure internal stellar structure and rotation. This requires very short integration times: minutes for dwarf stars to hours for giant stars. In order to obtain sufficient signal/noise on these timescales, it will be necessary to utilize broadband optical wavelengths to get sufficiently high fluxes. The typical SI target is considerably brighter in the optical and the non-radial modes are visible there as easily as at the shorter wavelengths. Low-resolution imaging is sufficient to measure non-radial resonant waves.

1.4.2.2 Required Angular Resolution

The seismological measurements to map the internal dynamics of large-scale flows require intensity measurements of the stellar photospheric emission. A resolution of 40 to 120 elements across the equator (using a linear array of apertures operating in multiple passbands simultaneously) with a time cadence of 1 min allows a similar resolution in the stellar interior to reconstruct the internal structure. A resolution of approximately 40 to 120 elements across the stellar surface (for a Y-shaped configuration) is adequate for imaging the differential rotation of the deep convective envelope. At these resolutions, which far exceed the 3 elements that are attainable from point source imaging, there is sufficient sampling to learn about differential rotation, convective-envelope depth, core structure, and even a coarse assessment of the overshoot layer in which the dynamo is thought to operate.

Note that multiple wavelength bands can be used to mimic different effective baselines: if SI can observe at 1500 Å, 3000 Å, and 6000 Å simultaneously, the number of Fourier components for seismology will be three times that of a single configuration, or conversely the number of array elements can be reduced correspondingly.

1.4.2.3 Requirements on Integrations Times

In order to measure the differential rotation of a star from asteroseismic observations, the duration T of the observing interval must be of the order of the stellar rotation period P. In that time interval, the photometric precision, assuming a similar amplitude of p-modes to that of the Sun, should reach approximately 0.2 ppm per resolution element, which requires a total of $2{:}5 \times 10^{13}$ photons. For 9 mirrors, for example, there are 36 resolution elements and $n_{\mathrm{ph}} \sim 9 \times 10^{14}$ photons are required.

The photon flux F through a system of m apertures with diameter a(m), for a star at d(pc) with luminosity L_* equals

$$F = 9 \times 10^{10} \frac{L_*}{L_{sun}} E\left[\frac{ma^2}{d^2{}_{pc}}\right] (\text{photons/s}) \qquad (1.10)$$

where E is the efficiency of the system, determined by filters, detectors, and mirror reflectivity. For E~ 0.3, the net integration time t_{as} to reach 0.2 ppm per resolution element in signal-to-noise is

$$t_{as} = 140 \frac{(m-1)}{F} \frac{L_{sun}}{L_*} \left[\frac{d_{pc}}{a} \right]^2 \text{ (s)} \qquad (1.11)$$

For m= 8, d_{pc}= 4, and a= 1, t_{as} ~ 16 hr, so after a stellar rotation, the S/N will be significantly higher than required for solar-mode detection.

1.4.2.3 Sample Targets

The targets to be observed in the asteroseismology mode will be selected from the sample which will be observed in imaging mode. The last column in Table 1.6 shows which targets will be observed in both modes (flagged by a"B" in the last column), as well as those which will likely only be observed in imaging mode. The asteroseismology mode will only be used on single stars, since its purpose is to "image" internal structure and rotation, which are not relevant for studies of mass-exchange or binary system dynamics. The maximum science is extracted from the observations when both surface and internal imaging can be applied. However, the total time required for the asteroseismology observations may limit the number of stars which can be observed in this mode.

1.4.3 Target Selection and Observing Strategies

SI will likely operate using a largely predefined target list to minimize time used for repointing and amount of consumables used for that purpose. A mix of targets will be established that balances the needs for quantitative analysis of stellar activity, structure, and dynamics, against the desire to explore the universe at unprecedented angular resolution. Targets of opportunity, such as supernovae within our Galaxy or in sufficiently nearby sister galaxies, will be accommodated. Primary science goals require long studies of targets, ranging from days to a month in selected cases. Exploratory studies and snapshots are likely to take several hours.

SI will observe within an annulus subtending +/- 20 degrees from a sun angle of 90 degrees, to permit the proper shading of the telescope mirrors. Each target will be visible to SI a minimum of twice a year, as this annulus rotates around the sky. Targets will be selected in a sequence which requires only small slews (~ 15 degrees or less) of the full array to minimize propellant usage and overhead time between observations.

It should be noted that there are thousands of stars that are larger than 1.4 milliarcsec, and hence allow imaging by SI with 20 pixels across the diameter, or over 300 pixels on the surface. There are some 750 stars larger than 2.5 milliarcsec, allowing imaging with 1000 pixels on their surface. However, only a few of these stars are the high priority targets that address the science goals of the SI mission optimally and careful target selection will be essential to maximize the science return of the long duration observations that are envisaged. Other science targets, outside of the stellar activity list, can be fit into the schedule so as to optimize the scheduling efficiency and provide maximum flexibility to the schedulers.

1.4.4 Design Reference Mission

In order to estimate the need for system resources and to estimate how many targets a mission like the SI will be able to observe over its mission, we generated some approximate illustrative Design Reference Missions (DRM's). These were only intended to provide a "feel" for the limitations and capabilities of the system and to point out the need, for example, for faster slew capabilities or additional propellant than initially assumed, rather than be definitive in any sense.

1.4.4.1 Targets

The target list for the DRM consists of three categories of targets. The high priority targets (prime targets) are 19 stars identified in the original SI proposal. Nine of these are solar-like stars to be observed daily for a month; the rest are other stars with lower observational cadences. Forty-five secondary targets are non-solar like stars, or AGNs, with 1 to 10 observations within a month. We augmented this with a list of 345 solar-like stars included mainly to ensure that a target would always be visible.

The spatial distribution of the targets is shown in Figure 1.42.

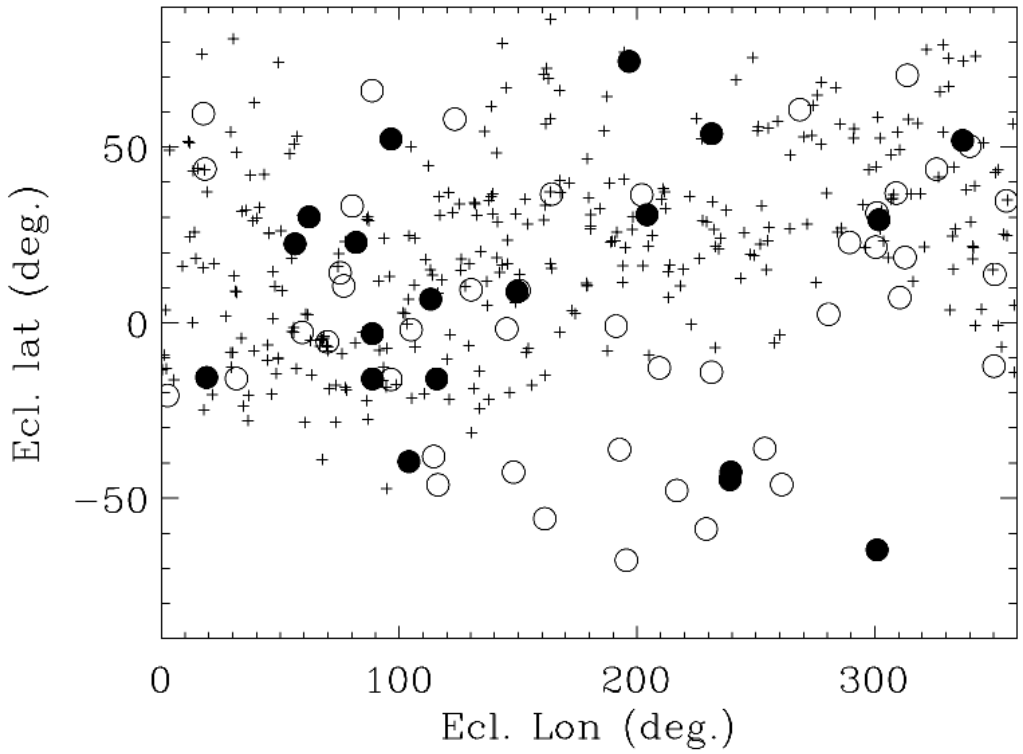


Figure 1.42 : The input target positions, in ecliptic coordinates.

The prime targets are filled circles; secondary targets are open circles. The tertiary targets (plus signs) are confined to the northern hemisphere.

Each target is assigned a priority (1, 2, or 3), as stated above. Each target is designated to be observed N times within a window of length set by the ~1 month time the target is visible to the observatory. The interval between observations, which we call the cadence, is generally 30/N days.

The integration times are set somewhat arbitrarily at this point. Exposure times must be a fraction of the rotational period, so we set the times to 8 hours for the priority one targets. These times will be refined later as the total system efficiency becomes better known with design refinements and better radiometric analysis. The distribution of exposure times is shown in Figure 1.43.

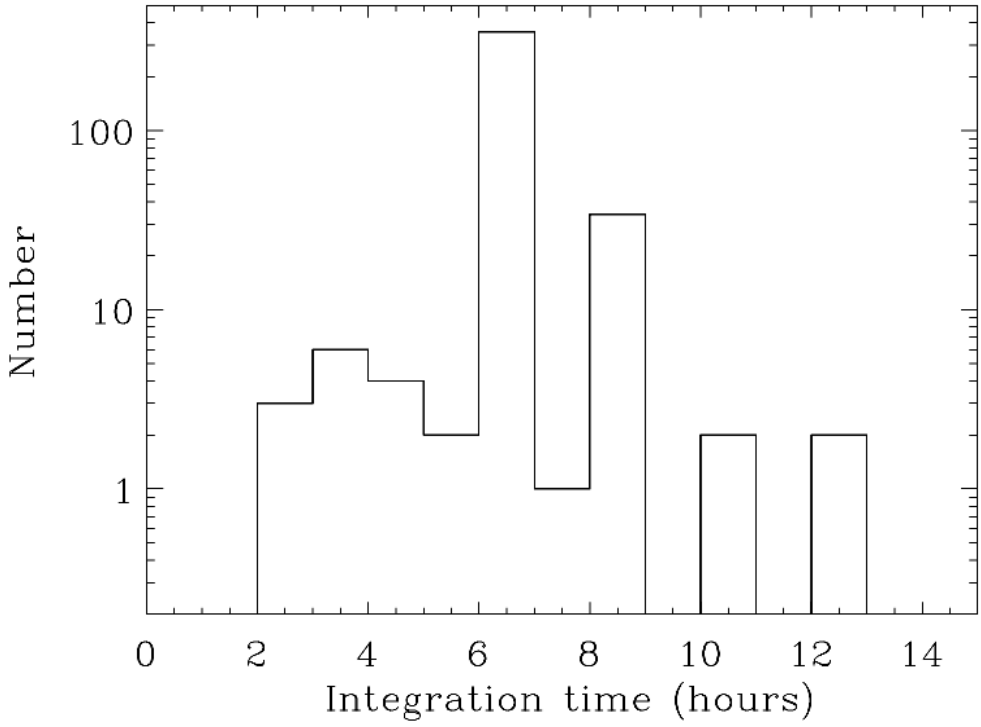


Figure 1.43. The distribution of observation times among the 409 targets in the in input catalog. The default integration time for solar-like stars has been set to 8 hours.

1.4.4.2 Assumptions and Scheduling Constraints

The S/C will point within 20 degrees of the great circle perpendicular to the Sun-S/C line, or 70<β<110 degrees, called the β-limit. With a +/-20 degree β-limit, a target on the ecliptic is observable for 40 days every 6 months.

The S/C will slew with constant acceleration, so that the time for a slew is proportional to the square root of the slew distance. We set the maximum slew rate to 10 deg/hr, which corresponds to an acceleration of 0.11 deg/hr^2, and a time of 18 hours to slew 180 degrees.

We assume that there is a 1 hour overhead required to settle on and acquire a target following each slew.

1.4.4.3 The Scheduling Software

The Design Reference Mission (DRM) software takes the input target list and the constraints and schedules an illustrative 1 year mission. Target priorities are determined by weighting the target priority, the slew time, and the number of days left in the observing window. These weightings are adjustable.

We set the slew time weighting to minimize the mean slew distance, although it is possible to alter this to achieve more efficient (or more complete) target scheduling, if needed, for example, by increasing the weight of the target priority or days left in observing window.

Other selectable parameters are the beta angle constraint, the slew rate, the overhead time, the date the mission starts, and the position of the S/C at the start of the mission.

1.4.4.4 Results

We simulated 36 mission plans, all using an assumed slew rate of 10 degrees/hr, 1 hour overhead time to set up at each target, and a +/-20 degree β-limit. These simulations all used the same input target list, but started at 6 different days of the year (1 + 60n, where n = 0-5), and from 6 different initial S/C pointings (latitude 0, +/- 90; longitude 0 or 180). We set the slew weighting to 100 to lower the mean slew distance.

Looking at the overall set of 36 simulations, we find: There are 661 +/- 46 slews (the error is the standard deviations of the mean), ranging in length from 0 to 179.4 degrees. The mean and median slew distances are 31 +/- 10 and 18 +/- 3 degrees, respectively. Of the 19 prime sources, between 0 and 17 (median=9) are started and 0 to 4 (median=3) are completed. In all, 112 targets are observed and 23 observing sequences completed (the low completion rate is due to the simple scheduling algorithm used). With the default settings, the observing efficiency (time actually observing) is 45%; 55% of the time is spent slewing and settling. A typical distribution of slew distances is shown in Figure 1.44.

From this we conclude that one or more of the assumptions must be changed to enable a higher completion rate of the prime targets: a faster slew rate, a different target weighting, or a lower overhead time at each target. We have assessed the sensitivity of the simulations to slew rates (using 1, 9, and 18 hours for 180 degree slews), overhead times (using 1, 3, and 5 hours) and β-limits (using 15 and 30 degrees vs. the initial 20 degrees). For each set of input parameters we generated 42 DRM's, starting on different days of the year and with the S/C at different initial pointings.

Some of the relevant results are summarized in the Tables 1.7 and 1.8, where we show how the number of targets observable during a 1 year mission depends on the slew rate and the overhead time.

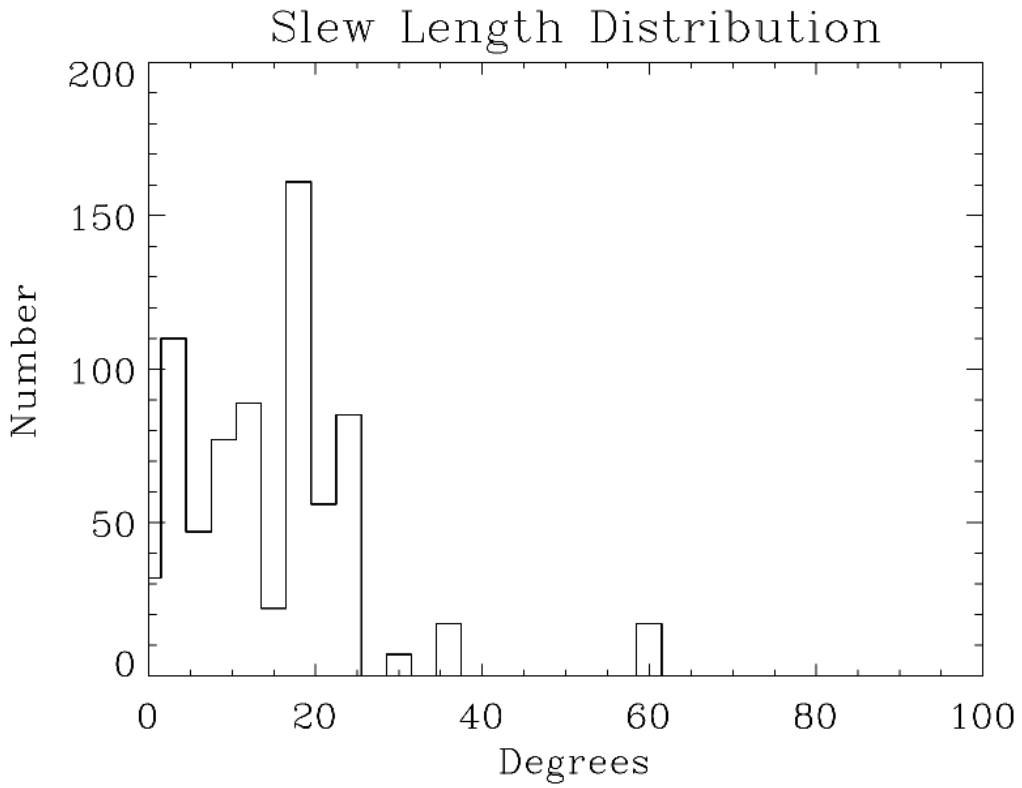


Figure 1.44. The slew distribution for one realization of the DRM.

The slew rate is 10 degrees/hr. Overhead is 1 hour/target. the beta limit is 20 degrees. 721 slews were executed, with lengths from 1.2 to 114.9 degrees. The mean and median slew distances are 16.4 and 15.9 degrees, respectively.

Table 1.7: Sensitivity of #targets (#slews) to slew rate (beta=20deg, ovhd=1hr)

Slew Rate (deg/hr)	# targets	Median slew (deg)
60	1078 +/- 29	16.4
20	806 +/- 74	21.1
10	649 +/- 52	19.7

Table 1.8: Sensitivity of #targets (#slews) to overhead time per target (beta=20deg, slew rate=10 deg/hr)

Overhead (hours)	# slews (deg)	Median slew (deg)
1	649 +/- 52	19.7
3	577 +/- 49	16.4
5	501 +/- 44	15.0

1.4.4.5 Conclusions

We have performed a very preliminary study of a possible 1 year (of 5-10 total) "reference mission schedule" to identify the significant aspects of the mission design relative to observing efficiency. Much work remains to be done in this area. For now, we can say that the mean and median slew distances cannot come down below about 15 degrees because of the spacing of targets on the sky. Similarly, the number of the prime targets observed is limited if we weight so as to minimize slew distances, so a compromise has to be reached to schedule a "complete" program that does not use too much propellant.

Reducing the slew times (increasing the maximum slew rate) clearly increases the number of observable targets. For example, doubling the slew rate to 20 deg/hr enables observations of about 20% more targets (see Table 1.8). Decreasing the overhead time will obviously increase the number of targets observable. In our example case where the maximum slew rate is 10 deg/hr, the spacecraft must carry sufficient maneuvering fuel for of order 700 slews for each year of the mission. With faster slews (and more complete target coverage), more fuel for more slews is required – a 6-fold increase in slew rate adds 66% more targets (and an equivalent number of slews) to the yearly list. The inclusion of an asteroseismology target (30 days continuous pointing at the same target, planned for once/year) will decrease the number of slews considerably (by 8%), while increasing the observing efficiency, as one would expect.

These DRM studies have indicated that a reasonable schedule can be created to perform the SI Science Program, but further more detailed and more robust studies of scheduling scenarios are required to find the optimal program that balances science productivity and mission design demands and technology capabilities (e.g., propellant usage and impulse available from the desired non-contaminating, long-lasting, continuously-variable thrusters). This study also needs to be revisited as future mission architecture trade studies are performed (e.g., the baseline mission described in Chapter 2 vs. alternative designs described in Appendices F and G).

This exercise also assumed that only one Hub spacecraft was flown. As pointed out elsewhere in this Report, the inclusion of a second Hub during initial deployment not only provides a backup to a single-point failure (loss of a single Hub spacecraft), but also provides an opportunity for dramatically increasing the efficiency of the observing program. This occurs because the second Hub can be pre-positioned for the next target while the first Hub is being used on the current target. The "slew" of the observatory from target one to target two then simply involves a tilting of the array to align with the second Hub and removes the wait involved for a single Hub to translate relatively large distances in real time to its new observing position (at target two). An extreme example of this involves positioning the Hubs on opposite sides of the primary array, thus allowing the observatory to "slew" 180 degrees by simply flipping the mirrorsats over to point at the opposite side of the sky – nearly instantaneously compared to the day or so otherwise required.

Chapter 2

Architecture and Implementation

2.1 Space Systems Architecture

2.1.1 Design Requirements

The diagnostic means that are to be employed by SI have been summarized in Chapter 1. Based on these, the following requirements can be placed on the "optimal" design:

- 9-30 one-meter class primary mirrors plus a central beam-combining hub
- Free-flying cluster of spacecraft with micro-thrusters for formation flying and repointing.
- Largest mirror-to-mirror baseline must be variable from 100 to 1000 meters
- A resolution in the reconstructed image of 60 μarcsec at 1,550 Å and 120 μarcsec at 2,800 Å.
- Observations in ~ 10 Å wide ultraviolet passbands, including the C IV doublet at 1550 Å (characteristic of 100,000 K plasmas) and Mg II doublet at 2800 Å (10,000 K), plus other lines such as Ly a 1216 Å, O I 1304 Å, C II 1335 Å, Si IV 1400 Å, and He II 1640 Å sampling a range of temperatures
- observations in broadband, near UV or optical continuum passbands representative of plasmas at 3,000 K to 10,000 K
- image stellar surface at high time resolution (1min cadences) with 30-100 resolution elements covering the stellar disk to perform astereoseimological studies of interior structures and flows for convection and oscillation studies
- reconfigurable telescope formation to enable imaging synthesis: must sample aperture for full resolution in 6-24 hours, depending on target
- 5 year required lifetime, goal of 10 or more years
- individual mirrorsats should be serviceable or replaceable as needed
- the hub ("focal-plane package"), which is far more mission-critical than any one mirrorsat, must contain a higher level of redundancy and be serviceable or there must be a second (backup) hub in place or available for launch on need. Detectors should be upgradeable.
- adequate time resolution for asteroseismology: obtain sufficient Fourier-component coverage fast enough to sample p-modes even for dwarf stars; a 9-telescope array provides 36 baseline pairs or 36 Fourier components (modes: $\ell < 10$ ~15 for Y-formation, or $m < 40$ or linear-array formation); operation in multiple passbands increases that number proportionally
- Orbit: likely distant from Earth to avoid gravitational perturbations and light contamination: Sun-Earth L2 point or Earth-Trailing? Sun-Earth L2 preferred for servicing access unless "Arrested Earth-Trailing" can be utilized – though servicing access and available infrastructure is likely to be less than at L2.
- Consumables may not be a problem, but preserve option of replenishing them as part of a periodic refurbishment and upgrade of the facility.

2.1.2 Optimal Approach to Achieving High Angular Resolution

It is clear that in order to fully address the science goals outlined previously that high angular resolution, on the order of a 100 μarcsecs or better in the mid-UV, is required. This requires mirror diameters or baselines between sparse aperture or interferometric elements on the order of 500 meters. Although a large monolithic mirror might possibly be considered for a precursor mission (where the resolution requirements are ~25x lower and baselines of 20m could suffice), even there the costs and technical challenges are high. Problems with obtaining sufficient rigidity without excessive mass and near-perfect manufacturing are significant for true monoliths. Segmented mirrors require precise surface control and relatively high mass can still be a problem. Both suffer from difficulties with launch because of the likely high mass and size. And it is clear that at 500m, a monolithic mirror is not feasible in the desired timeframe.

It thus appears that some type of sparse aperture mirror system using large booms or distributed spacecraft is needed. A boom arrangement can perhaps suffice for 10 - 50 m baselines, though the control of boom dynamics becomes increasing difficult with the longer booms, and even relatively short ones are extremely challenging, as has been seen in the development of the SIM mission.

As we head out to baselines beyond 50m baselines, we are led to consider either tethered formations for a limited number of optical elements (currently under study for the 3-element SPECS mission concept) or true free-flyers (e.g., the LISA mission), but the dynamics and control issues are difficult and may in the end not turn out to be any easier for a tethered system than for a system of true free-flyers. In the case of SI where a large number of optical elements are required to enable relatively rapid integrations on a given target (to avoid smearing of images due to stellar rotation, proper motion, and intrinsic variations of the stars), tethers seem fraught with dangers and a free-flying architecture is optimal. A free-flyer design does present significant challenges, including high-precision metrology and formation control over scales of hundreds of meters, but it represents the optimal solution in terms of the configuration flexibility needed to meet the science requirements. These options are summarized in Figure 2.1.

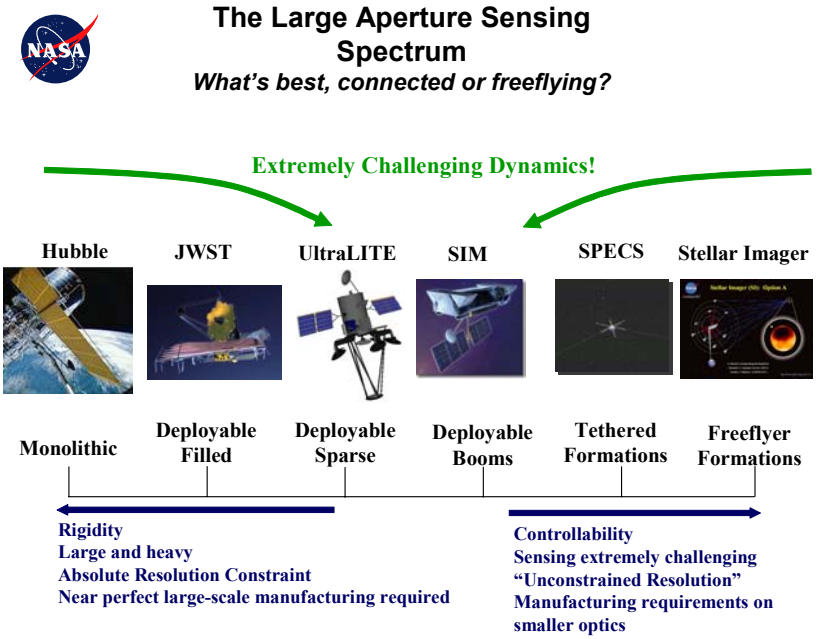


Figure 2.1: Large Aperture Sensing Spectrum

2.1.3 System Flexibility

A driving requirement shared by the various concepts under consideration is one of "flexibility" - the architecture must be able to accommodate changes including:

- The array must be reconfigurable to meet the image sampling requirements, i.e. ensure proper coverage of the Fourier UV-plane with variable numbers of operational "mirrorsats".
- The repair of the observatory through the repair or replacement of faulty "mirrorsats", or even the hub spacecraft.
- The expansion of the array via the addition of "mirrorsats" to improve the imaging capabilities of the array. This means that both the formation flying/mirror phasing control system as well as the hub must be able to accommodate more optical paths than in the configuration originally launched, at least if an incremental approach to building the facility is chosen.
- The system must be robust against the loss of one or more "mirrorsats", i.e. science operations and quality imaging must continue to be supported even if several of the "mirrorsats" fail. This may require longer observations times, due to the need for additional reconfigurations of the array, but not a halt to science operations.
- The system should be designed in a modular fashion to enable routine servicing by robotic and/or human means, to, at a minimum refuel the spacecraft, but perhaps to do much more.

2.1.4 Sparse-Aperture Imaging

The principle of sparse-aperture imaging is that each baseline (pair of apertures) measures the target at one spatial-frequency in one orientation. This is the reason a telescope with a central obscuration works; the entire telescope is not needed, just enough of it so that all spatial frequencies, out to some maximum value, are measured. The maximum measured spatial frequency determines the resolution of the image. Since an annulus of any thickness contains all spacings from zero to its diameter, it can be used to make an image of a target regardless of the size of the central hole.

The central hole has two effects on the measurement. First, and most importantly, the larger the central hole, the less light there is from which to make the image and with less light, longer integration times are needed. If resolution, rather than sensitivity, is the limitation, the longer integration times are not a serious problem.

The second effect of the central obscuration is to change the relative weighting of the different spatial frequencies in the resulting image. In a filled aperture, the spatial frequency with the largest amplitude is sampled only by two points opposite each other at the edge of the aperture. Lower frequencies are sampled by progressively more pairs of points in the aperture. In fact this weighting decreases almost linearly from a maximum at zero spatial frequency to zero at the maximum frequency sampled by the aperture. A central obscuration decreases the relative weight of the lower spatial frequencies. You may think that with this uniform weighting that gives higher weight to the higher spatial frequencies, the resolution of the image would increase -- and it does -- but at the expense of increasing the ringing in the image. The weighting of a filled aperture provides a natural apodization; the resulting image has a central core surrounded by reasonably faint Airy rings. As the central obscuration is increased, the core of the point spread function (PSF) becomes slightly smaller but at the expense of significantly worse ringing.

So far, this discussion has been based on annular apertures but there is no need to keep the aperture connected and symmetric – several disjoint apertures can be combined to form a single telescope as

long as those apertures are arranged so that they sample all spatial frequencies. But these asymmetric apertures do form more complex point spread functions with secondary peaks (sidelobes) instead of rings seen in filled aperture systems. Typically, there will be enough structure in the PSF to completely mask the actual image unless deconvolution techniques are used to remove the effect of the PSF.

Everything discussed so far is completely general. But we can make one additional change to our design requirements if we assume the sources we are going to observe are compact -- that is have finite spatial extent and in particular are smaller than the Airy disk of the sub-apertures. In this case we no longer have to sample all spatial frequencies! This is because of the Fourier transform nature of the relationship between the image and its spatial frequencies. If we know that ALL source structure falls within a region of extent E, then we also know that its Fourier transform can only vary on scales larger than 1/E. Thus we can uniquely determine the source image by sampling spatial frequencies on all points of a grid with spacing 1/2E.

Because the primary target list for the Stellar Imager is stars and because it takes apertures of hundreds of meters to resolve all but the largest stars, it is possible to build the SI primary mirror out of a number of 1.0-2.0 meter diameter apertures. But note that we cannot continue to make the resolution higher indefinitely. An advantage of using free-fliers for SI is that we can adjust the resolution of SI simply by moving the segments closer together or further apart. But the resolution cannot be increased to an arbitrary amount. With N telescopes in a non-redundant configuration, resolutions finer than about 1/N times the diameter of the star will produce holes in our spatial frequency knowledge that are larger than the limit imposed by the compactness of the source and the data will no longer uniquely constrain the image.

At this point, it becomes useful to stop thinking about how SI works in terms of a normal telescope, a point spread function, and deconvolution and to start thinking of SI's function as generating a list of amplitudes and phases at particular spatial frequencies and orientations. There are three advantages to this. First, we do not need to measure all the spatial frequencies at the same time, as long as the source is not varying over the total time spent observing the target. For example, if we want to increase the resolution beyond N resolution elements across the star: a factor of two in resolution requires four times as many spatial frequencies, but we can measure these a few at a time, reconfiguring the separations of the spacecrafts between observations. When we have enough data, we stop reconfiguring and use the list of data to form an image. Note that this reconfiguring is inefficient; it is not feasible to increase the resolution arbitrarily. Since some spatial frequencies will occur in more than one configuration, it will take more than 4 configurations to increase the number of spatial frequencies by a factor of two.

A second advantage to thinking about the data as a list of fringe amplitudes and phases is calibration. The observed amplitudes and phases are perturbed by tip/tilt/piston errors between the apertures. These errors are associated with the apertures so they grow linearly as the number of apertures increases. But the amount of data grows as the number of pairs of apertures, which is proportional to the square of the number of apertures. Consequently, the system is overconstrained -- there are combinations of amplitudes and phases that are not corrupted by static alignment errors. Taking advantage of this additional information could, potentially, improve the precision of the measurements.

A third advantage is that for a non-redundant configuration, the signal to noise is not affected by the piston errors. This converts a control problem into a knowledge problem: as long as we know how large the error is, and as long as it is stable during an integration, we can correct for it in post processing *with no signal to noise penalty*.

2.1.5 Baseline Architectural Concept for Full-Mission

The current baseline architecture concept for the full Stellar Imager (SI) mission is that of a 0.5 km diameter space-based UV-optical Fizeau Interferometer composed of a reconfigurable array of 10 - 30 one-meter-class deformable flat mirror array elements on "mirrorsats". The flats are deformable to the very slight curvatures needed to match the paraboloid surface of the long-focal-length (1 to 10 km) virtual primary and the deformation needed changes as the array diameter and focal length change depending on the size of the target to be observed. Those elements direct light to an image plane beam combination facility in a hub at the prime focus. The hub and all of the mirrorsats are free-flyers in a tightly-controlled formation. This design would provide:

- an angular resolution of 60 and 120 micro-arcsec at 1550 Å and 2800 Å
- ≈ 1000 pixels of resolution over the surface of a nearby (at 4 pc) dwarf star
- observations in 10-Angstrom UV passbands around, e.g., C IV(100,000 K), Mg II h&k (10,000 K) emission lines for activity studies and other lines in the 1200 – 2800 Å regime
- observations in broadband near-UV or optical continuum (formed at 3,000-10,000 K) for asteroseismology
- a long-term (5 - 10 year) mission to study stellar activity/magnetic cycles by ensuring that individual mirrorsats and the central hub can be refurbished or replaced on need

The Vision Study utilized the services of the two components of the GSFC Integrated Design Center (IDC) to devise a baseline design vetted by the various engineering disciplines and responsive to the scientific goals of the mission. A major purpose of these studies were to develop a "Technology Roadmap" to summarize the various paths that might be taken to move from the current state of technology to that needed to fly the mission. This technology roadmap is presented in Chapter 3. In the current chapter we describe the design concept itself.

The Integrated Mission Design Center (IMDC) was utilized to work out the details of the overall system, including the primary "mirrorsat" array, formation flying methodologies, launch, deployment, station-keeping requirements and observational scenarios.

Figure 2.2 shows an artist's concept of the baseline SI design, a Fizeau Interferometer with > 20 one-meter primary mirrors (mounted on formation-flying "mirrorsats") distributed over a parabolic virtual surface whose diameter can be varied from 100 m up to as much as 1000 m, depending on the angular size of the target to be observed. The individual mirrors are fabricated as ultra-smooth, UV-quality flats and are actuated to produce the extremely gentle curvature needed to focus light on the beam combining hub that is located from 1 – 10 km distant (the focal length scales linearly with the diameter of the primary array, a 100 m diameter array goes with a focal length of 1 km, a 1000 m array with a focal length of 10 km).

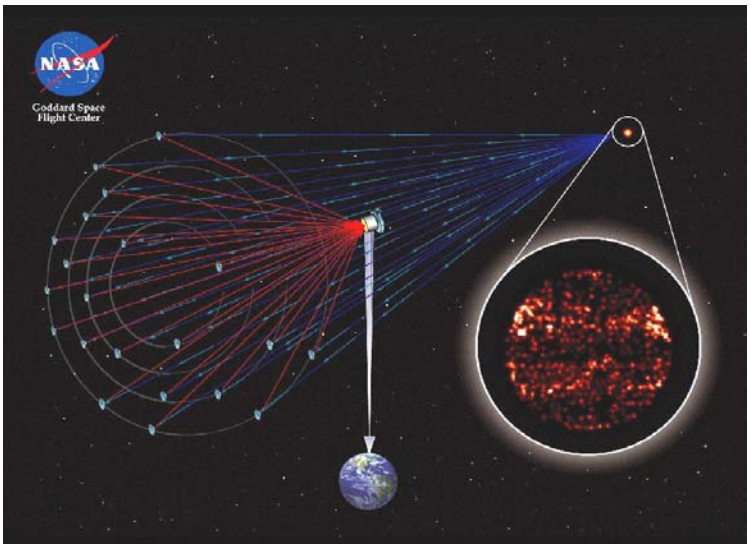


Figure 2.2: An artist's illustration of the "baseline" design adopted for SI during the Vision Mission Study.

Figure 2.3 (below) shows a cross-sectional diagram of the overall layout of the observatory with the primary mirror array on the left side of the diagram and the hub on the right side. Important characteristic sizes and distances are indicated.

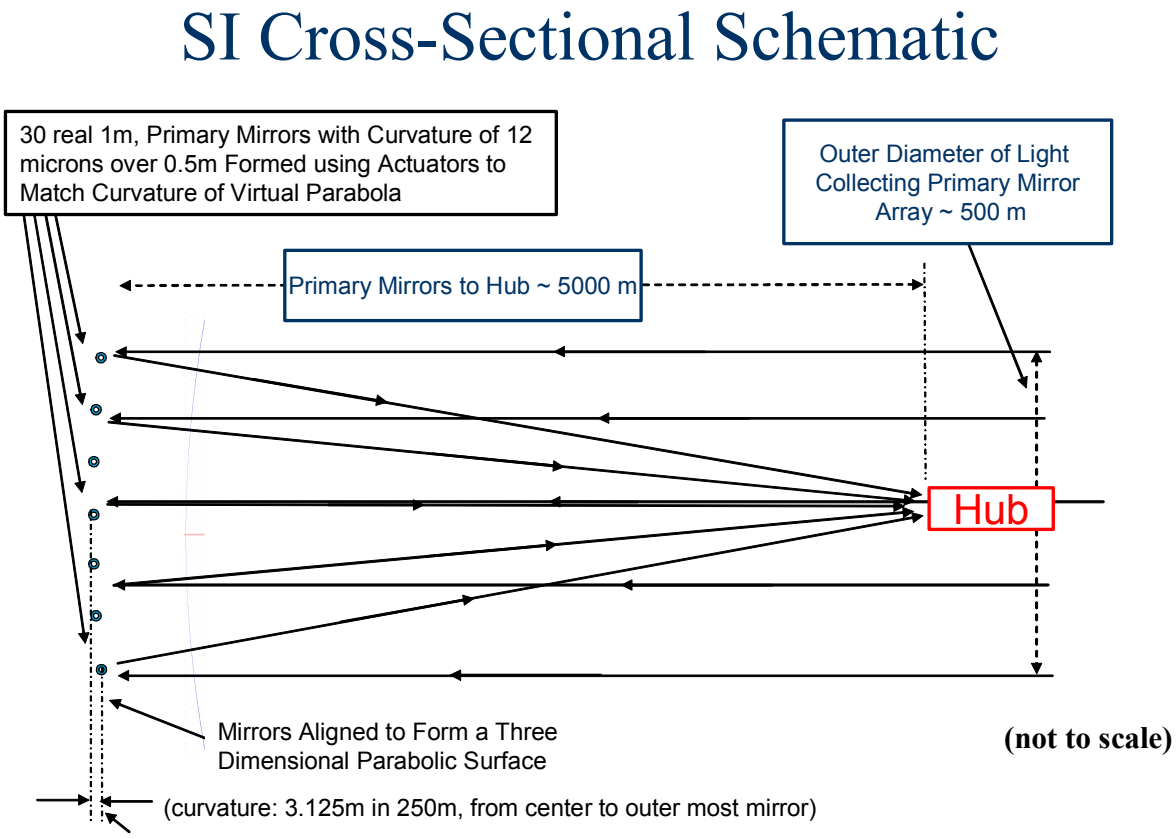


Figure 2.3: A schematic layout of the mission concept illustrated in Figure 2.2.

Figure 2.4 shows a view from the direction of the hub (or the target) toward the primary mirror array. In the left panel, the individual mirrors are shown much larger than they would appear (on this scale they would be dots, 1/500th the full array diameter). The mirrors are arranged in a "Golomb Rectangle" (Golomb 1982) pattern to minimize the redundancies in the Fourier (u,v) baselines sampled by the array. The Golomb pattern consists of 30 separate 1 meter diameter spherical mirrors, each comprising a component of the primary array. The mirrors are arrayed in a 2D non-redundant aperture pattern with the constraint that each of the mirrors falls at a vertex of the 30 x 30 rectilinear grid pattern shown on the left side of Figure 2.4 (not to scale). The centers of the two most distant apertures (longest baseline) are 500 meters from each other and the centers of all the apertures fall within a circle of diameter of 500 meters. The right side of Figure 2.4 shows a similar aperture pattern drawn to scale; note this is a very sparse array, most of the virtual primary mirror is empty. The Golomb pattern is a non-redundant aperture pattern, implying that no two baseline pair vectors are of the same length or rotation angle – this maximizes information content for a fixed given aperture area. The net effect is to synthesize a virtual primary mirror which spans the set of mirrors, thereby allowing ultra-high resolution spatial imaging, but, at the expense of missing and/or attenuated image spatial frequencies and reduced throughput as compared to a filled aperture of the same size.

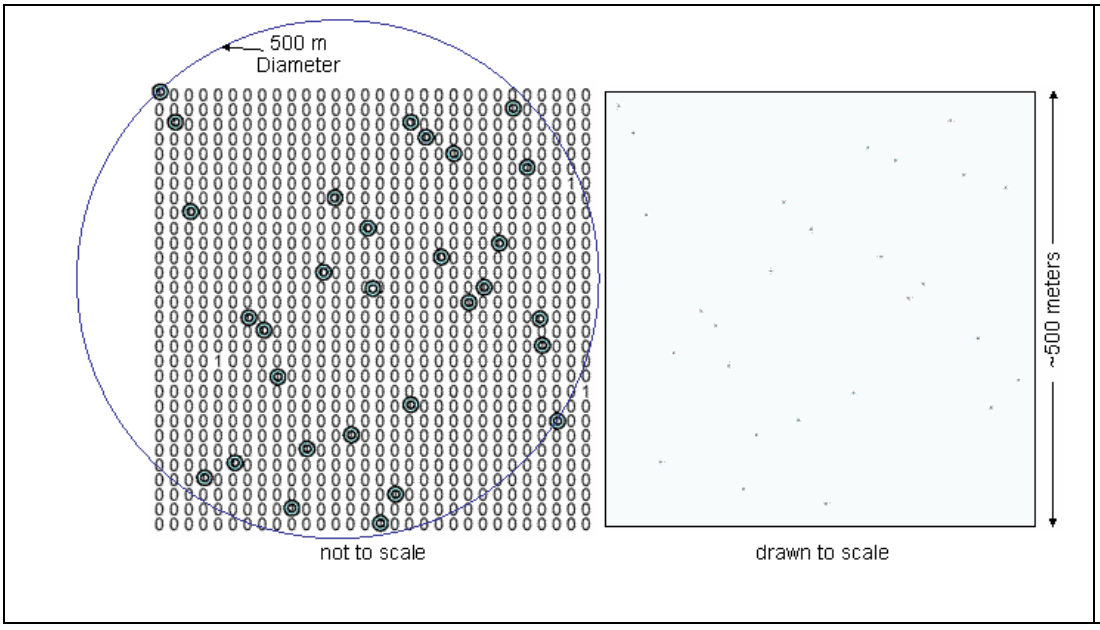


Figure 2.4: Views of the SI primary array from its hub: Left: 30 separate mirrors (enlarged) at vertices of a rectilinear grid, producing non-redundant baseline pairs. Right: a similar pattern, but drawn to scale to show the true sparsity of the array.

The Instrument Synthesis and Analysis Lab (ISAL) was used to work out the details of the beam-combining hub, including the optical design, detector requirements, and metrology needs. Figure 2.5 shows the major components of the hub spacecraft as designed in the ISAL. Further details of the contents of the hub are given in section 2.2 ("Science Instrumentation").

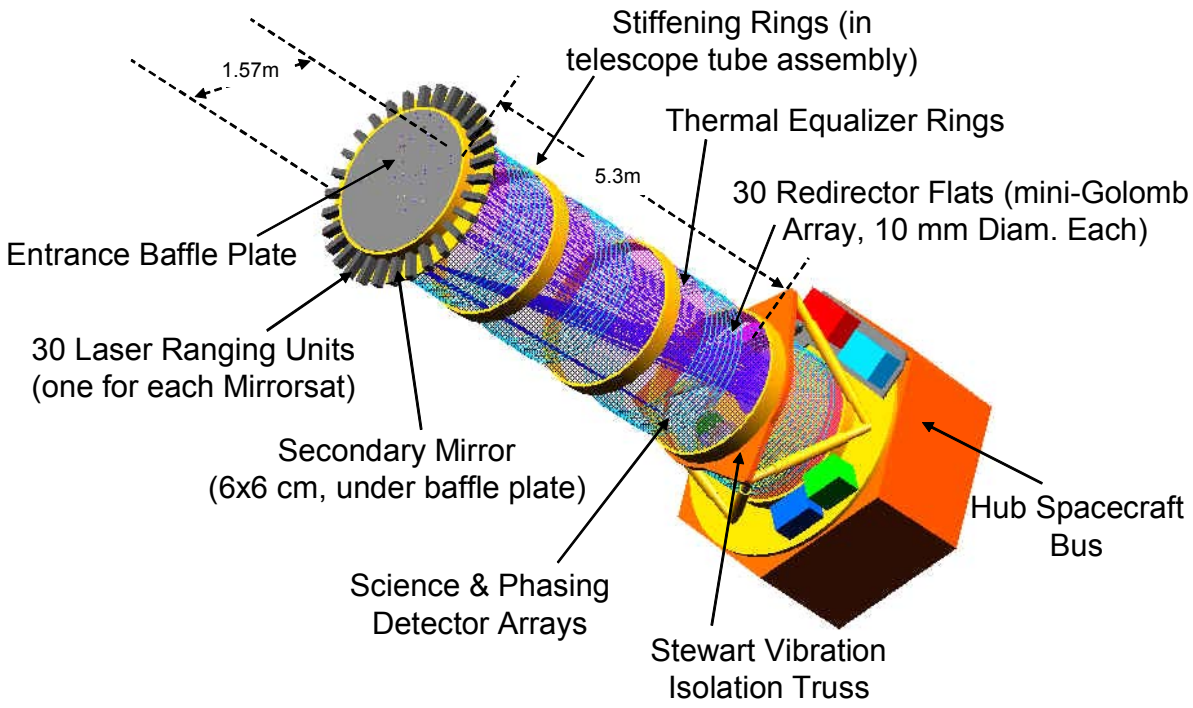


Figure 2.5: An external view of the SI hub with its major elements identified.

2.1.5.1 Control and Communications Concepts in Baseline Design

The overall control approach will be to use a "staged" metrology system utilizing Radio Frequency (RF) ranging at the coarsest level, laser ranging at the intermediate level, and Wavefront Sensing and control at the finest level.

The coarse metrology will use S-Band RF ranging from the Hub to each Mirrorsat and from each Mirrorsat to at least 2 other Mirrorsats. With the known distances between 4 objects there remains only one solution for the geometry ("Truss") - the constellation geometry is thus established by adding these "Trusses" together. The RF ranging mode is used to position the Mirrorsats relative to the Hub**.** The RF ranging system accuracy is determined by the S-band frequency, or between 7.5 and 15 cm and it is assumed that other attitude sensors will orient the Mirrosats to within ~1 arcsec orientation.

The intermediate, laser ranging will be done from the Hub to retro reflectors on each Mirrorsat. The same technique is used to construct the constellation geometry, but with more precise range data. The errors between the desired satellite-satellite distances and the measured distances are used to control the thrusters to fine tune the constellation geometry. The control limit of the laser ranging system and thrusters is approximately 1-2 microns.

The finest level of control will be done via WaveFront Sensing Control (WFC): once the various path lengths are within a few wavelengths of each other, fringes produced by interference of the observed starlight from the separate apertures become visible on the diversity camera (the optical

channel devoted to wavefront sensing, separate from the two science channels) and this information can be used to control the fine position of the piston mechanisms (on both Hub and Mirrorsats). The control limit of the WFC system is approximately 3nm.

A critical issue for proper observatory operation is how and at what scale to perform the handoff from one metrology system to another in this staged control system.

- Handoff between S-Band and Laser Systems
  - Once the Mirrorsat position is known to the limiting accuracy of the S-Band system, ~10cm, and the attitude of the Mirrorsat is correct, then the retro-reflectors on the Mirrorsat will be somewhere within the FOV of the Laser system
  - The Laser scans up to 400 points (discrete points or smooth motion) within a 10 arcminute by 10 arc minute box. Once the laser illuminates the retro-reflector on the Mirrorsat, a return signal is confirmed and the scan stops.
  - The S-Band ranging system now operates in parallel, but the laser system has a more accurate measure of the distances, so it will generate the error signals that control the thrusters.

- Handoff between Laser and WaveFront Control (WFC) Systems
  - Once the laser system reaches its control limit, range within a few microns, the WFC might be able to pick up some fringes on some of the optical paths.
  - Even with the satellites position controlled at the limiting accuracy of the laser system, biases between the focal plane, retro-reflectors, and lasers may be larger than the coherence length of the starlight (~12 microns at 500 nm) , thus preventing the various optical paths from generating fringes.
  - While controlling position at the limiting accuracy of the laser system, the optical path length can be changed with the piston mechanisms (either on the Mirrorsat, Hub, or both). This is done until fringes appear and the bias for that optical path can be established.
  - This is repeated until all of the optical path biases are established and all are within the control range of the WFC system.
  - The laser and S-band ranging systems both still operate in parallel
    - The WFC system controls the fine positioning of the piston mechanisms
    - The laser and s-band systems manage the control of the thrusters to assure that the piston mechanisms remain near the center of their operating range

Tip/Tilt mechanism control is another important concern. Each incoming beam reflected from the primary mirror of each Mirrorsat should enter a 1cm hole on the aperture plate located on the Hub. At 1 arc second pointing accuracy, the beam will lie within 2.5cm of the aperture plate hole. Photo diodes surrounding each aperture plate hole are used to provide feedback to the tip/tilt mechanisms on the Mirrorsats. Once the beam is successfully within the aperture plate hole, it will illuminate one of the corrector flat mirrors. The tip/tilt of the Hub corrector flat and the tip/tilt of the ellipsoidal secondary mirror is modulated until the beam lands on the desired part of the focal plane. This staged control concept in illustrated in the following Figures 2.6 and 2.7.

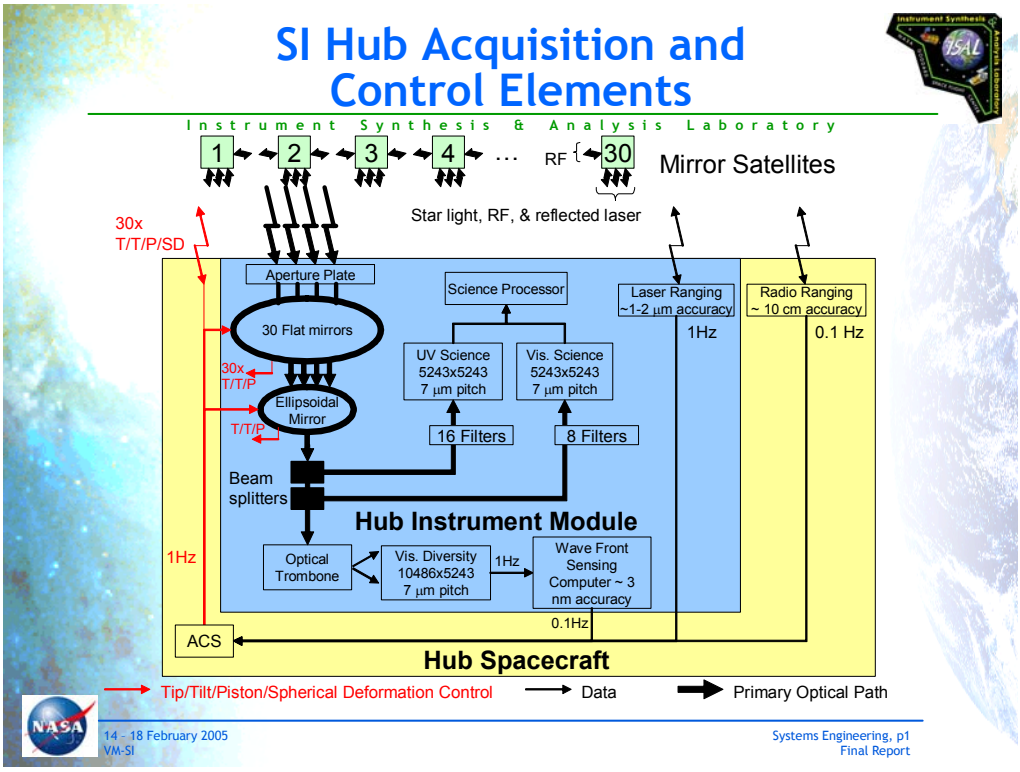


Figure 2.6: Acquisition and Control Elements on Hub Spacecraft.

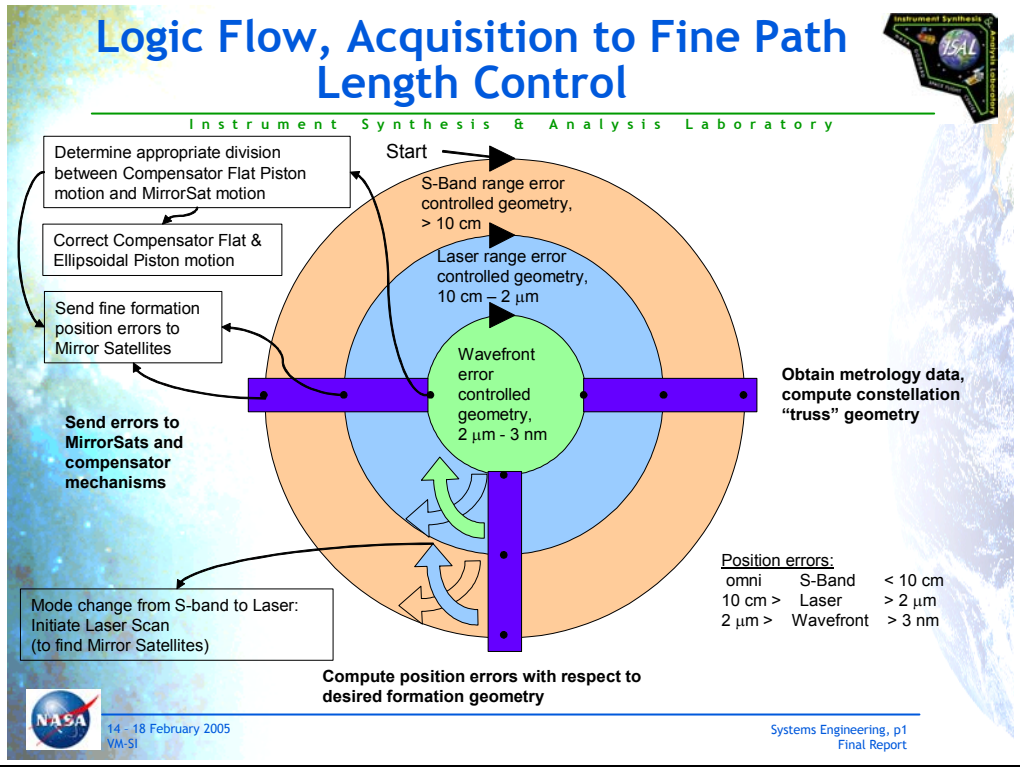


Figure 2.7: Logic Flow from Acquisition to Fine Path-length Control.

An example of a handoff between two levels of the control system is shown in Figure 2.8

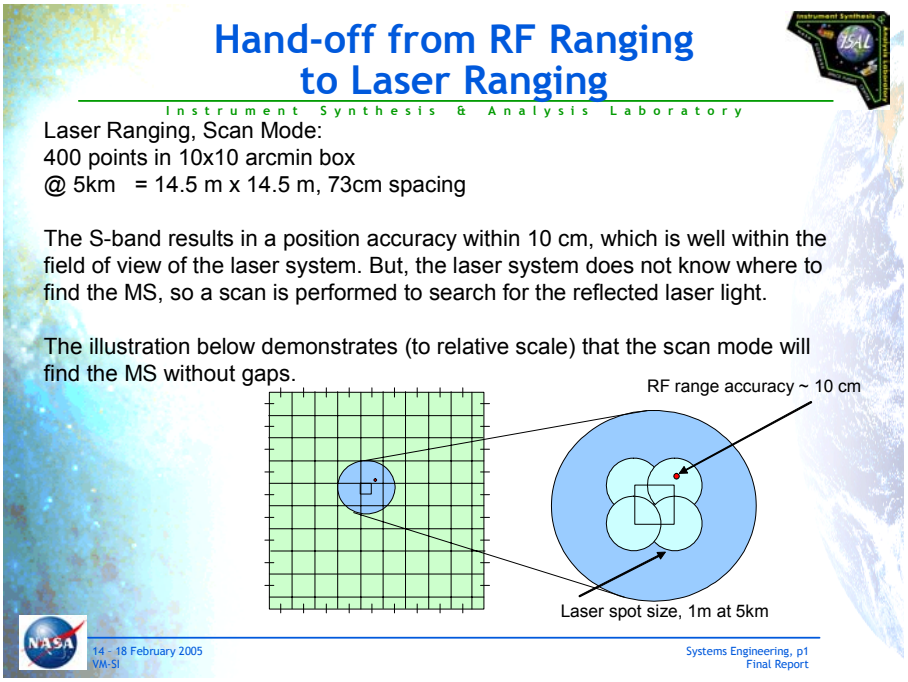


Figure 2.8: Hand-off from RF Ranging to Laser Ranging Control.

and the Path-length Control bands are shown in Figure 2.9.

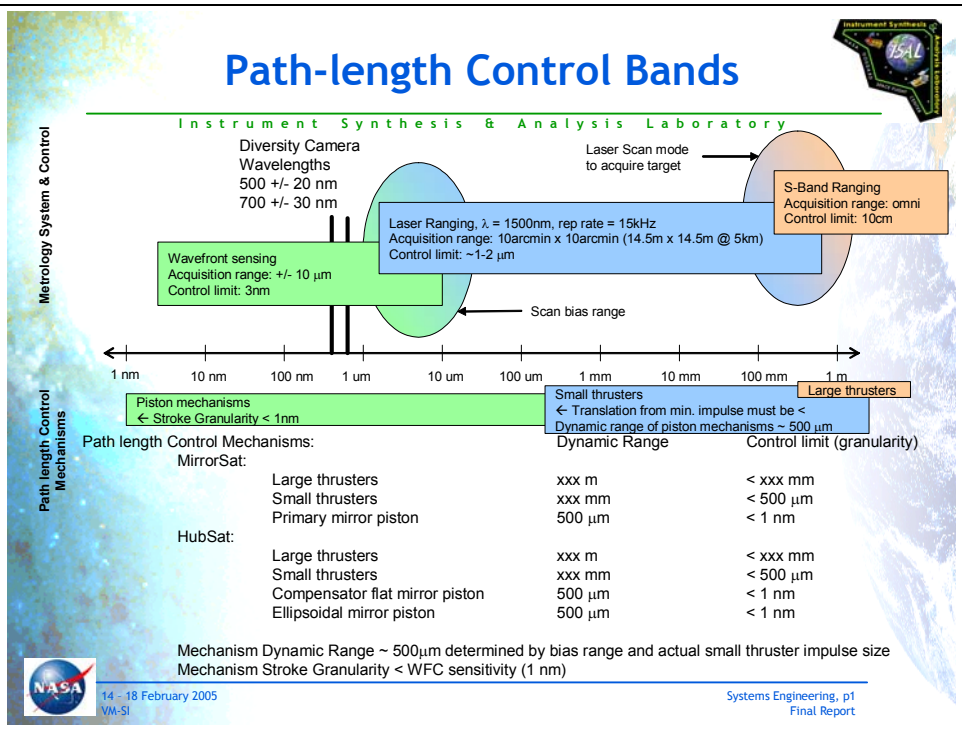


Figure 2.9: Path-length Control Bands.

The primary communication links that enable this control system and the science data collection to work are shown in the Figure 2.10, with the maximum supported data rates in Figure 2.11.

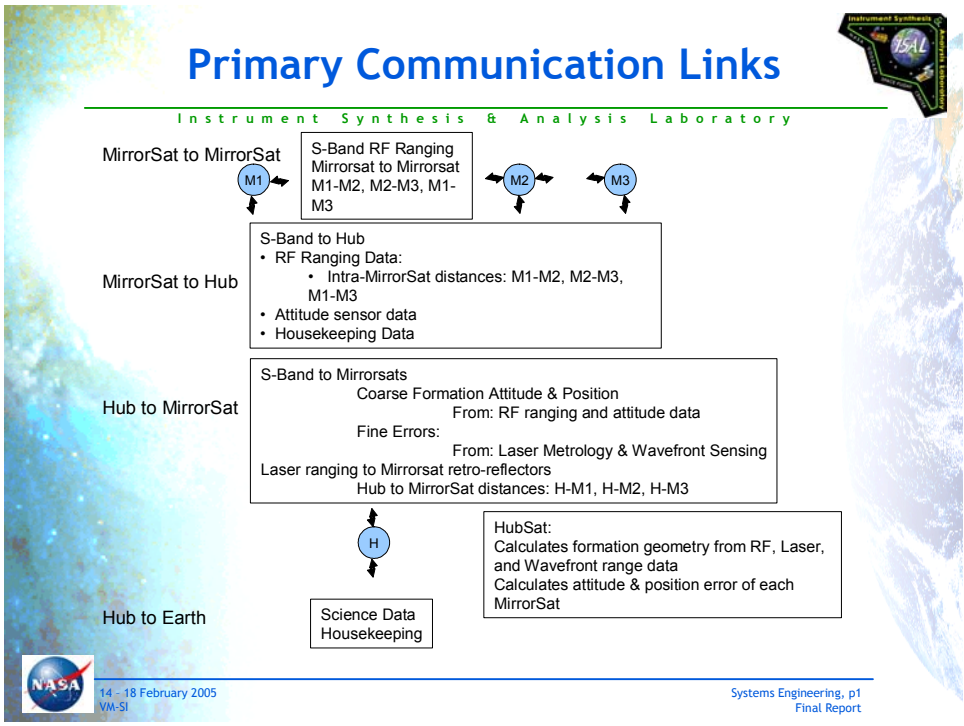


Figure 2.10: Primary Communication Links for Control and Science.

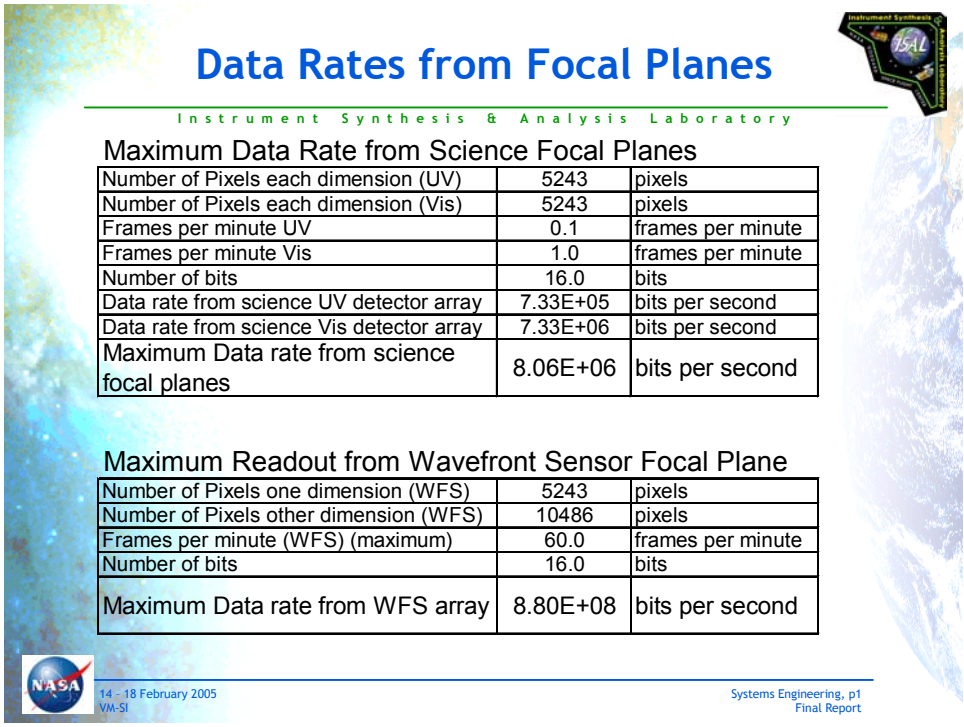


Figure 2.11: Maximum Expected Data Rates from Detectors.

The primary mirror array can be reconfigured to different baselines, with the normal expected range of operations being performed with maximum baselines (array diameters) in the 100 m to 1000 m range, with 500 m being the "typical" configuration. As shown in the following Figure (2.12), the system f/ratio is constant at f/10, so the system focal length changes linearly with the primary array diameter. The individual primary mirrors are actuated to enable their focal lengths to change accordingly.

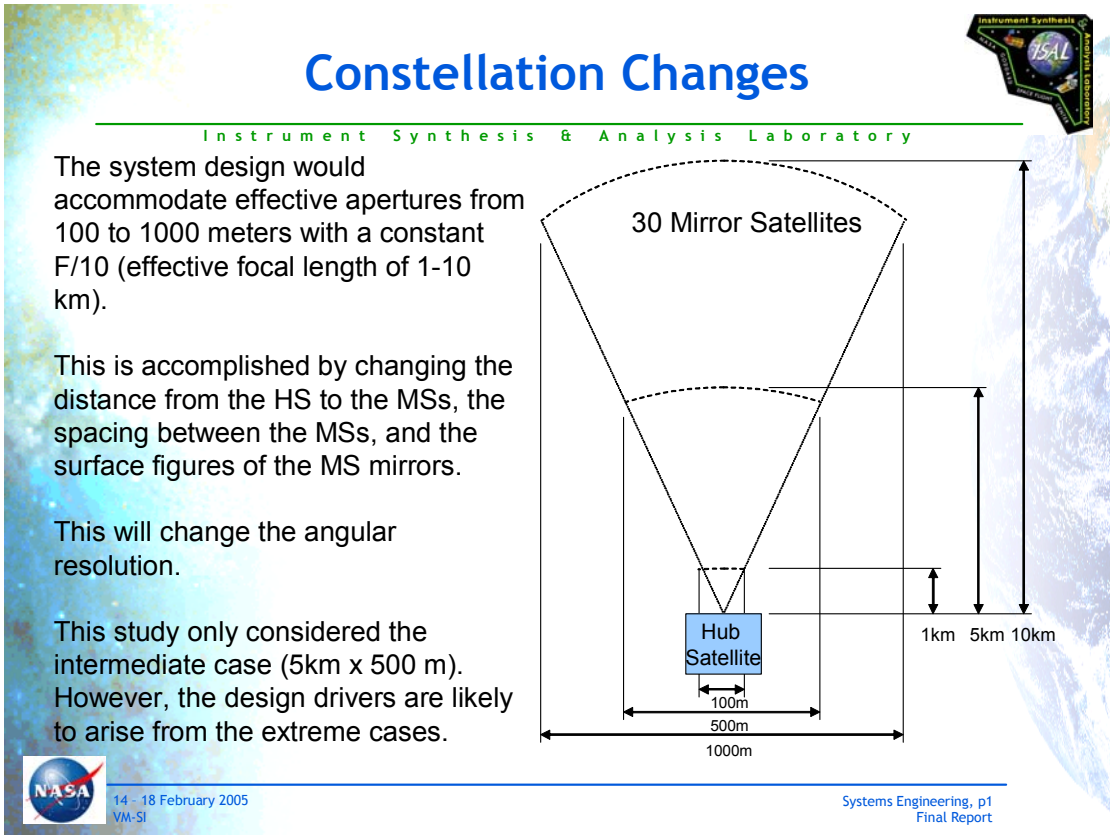


Figure 2.12: The System Focal Length Changes with Primary Array Diameter (angular resolution).

2.1.5.2 Spatial Location of SI

The best location for the SI observatory would be in a Lissajous orbit around the sun-earth L2 point. It cannot be in a low-earth orbit because strong gravity gradient there would not permit precise formation flying (and there would be potential scattered light concerns as well). An un-arrested, earth-trailing orbit is not desirable since replacement of failed array elements and addition of improved (larger) array elements would not be possible. L2 has both a small and very well characterized gravity gradient (permits precise formation flying) and should be accessible in the 2015 and beyond time frame for servicing and upgrade by robotic and/or manned missions. The location of the L2 point in the Sun-Earth system and the transfer orbits available to L2 are shown in Figure 2.13. L2 is likely to have a better general infrastructure available to support missions like SI, e.g., for communications and/or servicing, as well.

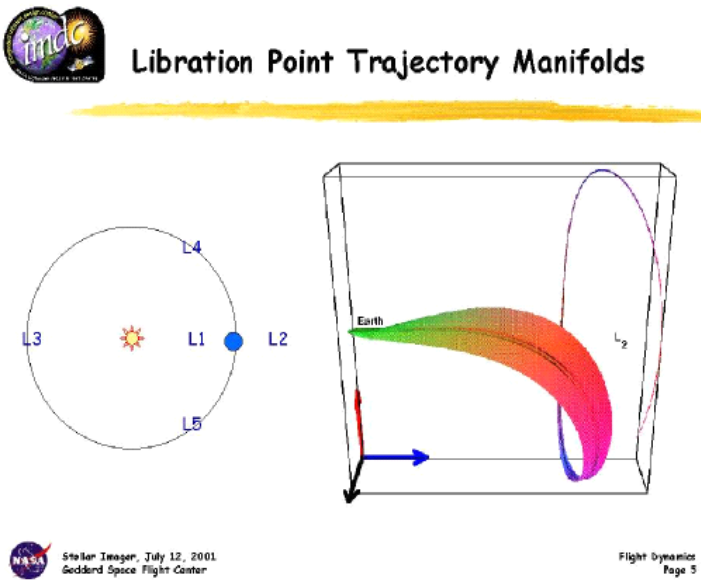


Figure 2.13: SI orbit trajectory and final location around the Sun-Earth L2 point.

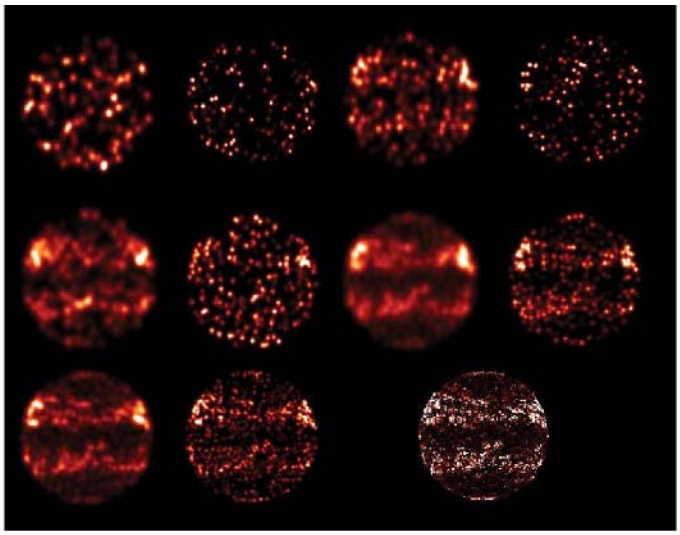
2.1.5.3 How Many Primary Mirror Elements are Needed in Baseline Design?

The images that could be obtained with the strawman mission design are illustrated in Figure 2.14 for various numbers of elements and re-configuration strategies. These simulations were computed with SISIM (Rajagopal et al. 2003) assuming 250 (first and third columns) and 500 (second and fourth columns) meter maximum baseline arrays. The first two rows assume Y-shaped configurations with 6 and 12 elements, respectively. The last two columns of those rows assume that the array is rotated 24 times (15 degree motions) to acquire sufficient Fourier UV-plane sampling. The 1st two images in the last row assume 30 elements arranged in a low-redundancy Golomb rectangle (Golomb, 1982). The first two columns in all cases show snapshots taken without rotating the arrays. The image in the lower right is the input image. This figure shows that 30 static elements appear to be sufficient to adequately synthesize this particular stellar image. The 435 baselines provided by the static 30-element array works well because only about half of the 1000 pixels in the image are truly filled. If all the image pixels were filled (or a large number of the remaining pixels), then a second configuration of the array (e.g., a 90 degree rotation) would be necessary for sufficient sampling. Alternatively, fewer elements can be used with a larger number of rotations (6 elements with 24 rotations or 12 elements with 6 rotations).

Figure 2.15 compares the resolution of the 30-element design for baselines of 100, 250, and 500 meters for the same stellar model seen both equator-on and from 40 degrees north latitude. The activity belts and the larger groupings of active regions are visible at the shortest baseline, smaller groupings of plages are resolved at 250 meters, and the full 500 meter baseline is required to clearly resolve the individual active regions.

Figure 2.14: Simulated interferometric (CIV (1550 Å)) images of a sun-like star at 4 pc, viewed equator-on, based on the model solar image at bottom right.

See Section 2.1.5.3 for details.



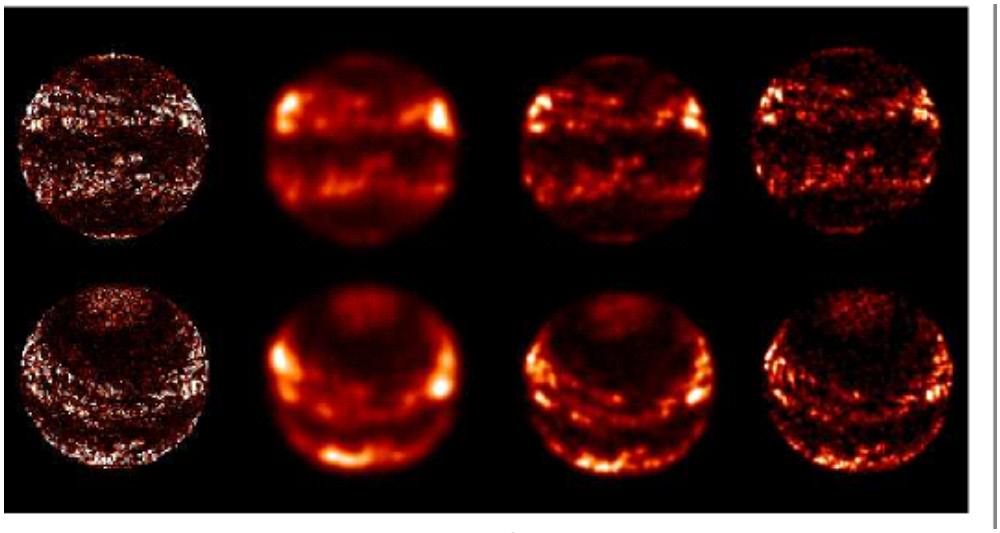


Figure 2.15: Simulated interferometric (CIV (1550 Å)) images of a sun-like star at 4 pc, viewed equator on (top) and from 40 degrees north latitude (bottom).

The input images are in the first column, simulations of observations with 100, 250, and 500 m baselines are shown in columns 2, 3, and 4.

2.1.5.4 Further Considerations for Baseline Design

Several variations on the basic strawman design are under consideration.

A primary concern is in the number of array elements (mirrorsats) to be used. A smaller number are cheaper to build and easier (perhaps) to control and the hub optics are simpler in this case, but more re-configurations of the array are required to properly sample and image the targets, and that has its own cost in dramatically increased overhead times, increased propellant use, and severe limitations on the speed with which targets can be imaged. A larger number of elements solves the latter problems, but of course leads to increased total production costs and more complex formation-flying control solutions. The final decision on this may depend critically on whether or not the use of optical delay lines in each light path can be avoided by precision formation flying of the mirrorsats. If optical delay lines are required, it is difficult to conceive of processing more than perhaps 10-12 simultaneous beams

in a hub full of delay lines. If, on the other hand, the mirrorsats and hub can be formation-flown with sufficient precision to maintain equal optical path lengths for each element, then a hub of reasonable complexity and cost is feasible, even with a large number of primary mirrors.

One possible mission concept is to start with an "n" aperture mission to observe slowly-evolving objects. Then a second set of "n" apertures can be added, and the two optical systems can then be reconfigured alternately, while the hub utilizes the non-moving set to observe the target. That reduces the effective reconfiguration time to the time it takes the spacecraft to adjust (either a maneuver or a repositioning of the hub's internal optics) to match the other set of apertures, plus the time required for baseline determination and stabilization.

A variation on the preceding idea is to start with the full number of apertures, but to launch a second beam-combining hub, either initially or at a later date. The presence of two hubs would allow one hub to be moved into position for observing a new target, while the other hub is being used to observe the current target, thus shortening the re-targeting time substantially. This has an advantage over the multiple sets of apertures in that it only requires one spacecraft (the hub) to be moved large distances – the aperture array only needs to be tilted nearly in-place, so that the large number of spacecraft in the array will not require large amounts of fuel or re-targeting time. This also provides redundancy for a critical path element, the hub, without which the observatory could not function. This approach also provides the full number of array elements needed to obtain "snapshot" observations (without the need to reconfigure the array while at a given target simply to get sufficient sampling of the Fourier (u,v) image plane). A clever use of the two hubs would enable the observatory to look at opposite sides of the sky very efficiently: placing the second hub "behind" the primary array by one focal length would allow the facility to "slew" 180 degrees around the sky simply by flipping the primary mirrorsats over to point at the other side of the sky, using very little fuel or time.

If the hub(s) is/are designed from the beginning to handle the full complement of primary beams, then the mirrorsats could be launched in groups smaller than the entire set, if that was desired from financial or technical reasons. Observations would take longer at the beginning with fewer elements (due to the required reconfiguration of the array during observations on a single target), but they could be done. Efficiency would increase as more elements were added to the array and targets changing more quickly or moving more quickly could be added to the observing list as the numbers of elements were built up. The launch scenarios described in this document would not require this, since sufficient launch capabilities are available to launch the entire suite of spacecraft in one or two launches, but the possibilities exist if other reasons argue for breaking the launch and deployment into smaller sets of mirrorsats.

2.1.6 Alternative Architectures

The "baseline" design for SI described throughout this report is for a Fizeau Interferometer/Sparse Aperture Telescope composed of up to 30 free-flying mirrorsats, carrying 1 m mirrors and distributed over a virtual paraboloidal surface 100 – 1000 m in diameter, which focus light on a common beam-combining hub, located 1 – 10 km distant. This architecture was selected by SI Team as the best option at the current time. The Team did consider several other alternative concepts during the course of the study, including the possible use of a large Fresnel Zone Plate (FZP) or Photon Sieve. In Appendix H, we discuss these alternative options, and discuss the trade-off between filled and sparse apertures and between using a large number of small apertures or a small number of large apertures in the sparse aperture case. Appendix H concludes with our rationale for the choice of the many-element Fizeau Interferometer as the best architecture to meet the science requirements.

2.2 Science Instrumentation

2.2.1 Overview of Hub in Baseline Design

In the case of a Fizeau Interferometer, the distinction between telescope and instrumentation becomes somewhat murky in that, in some sense, the instrument is the telescope. However, since we have described the overall architecture in the preceding section, we will take 'science instrumentation' here to mean the optical and detector systems inside the beam-combining hub.

Figure 2.16 shows a detailed block diagram of the hub and the optics, detectors, and supporting instrumentation contained therein. Light from the source is reflected off the 30 mirrors in the primary array and relayed into the hub spacecraft. The hub spacecraft effectively controls metrology, pointing and wavefront control between each of the mirrorsats and between the mirrorsats and the hub, and ultimately constructs both the UV and visible light science imagery. The baseline hub consists of multiple subsystems which include: spacecraft bus, telescope tube assembly, internal optics, entrance baffle plate, metrology subsystem, wavefront control subsystem (visible light) and science focal planes (visible & UV light).

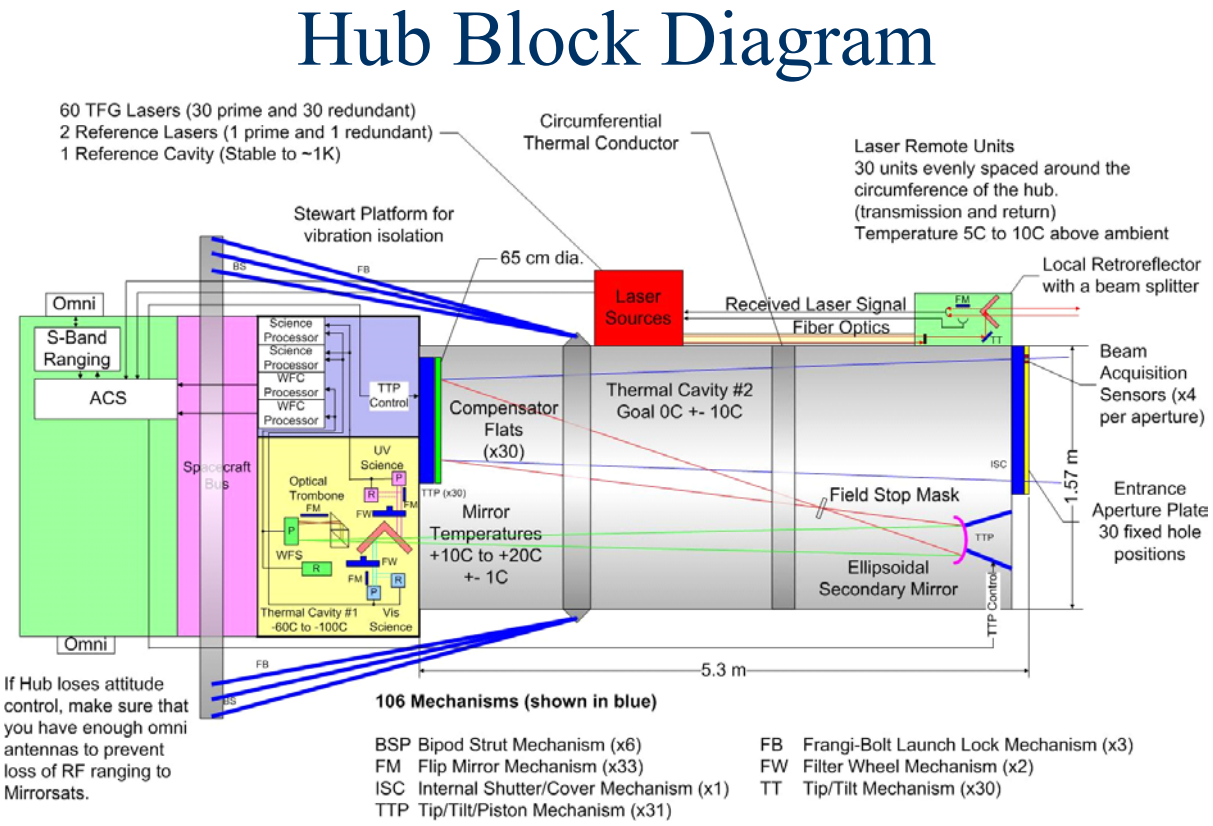


Figure 2.16: A detailed look at the hub design, showing optics, detectors, metrology components, and support systems.

An optical ray-trace is shown in Figure 2.17. Broadband light initially enters the hub from the 30 primary mirrors through the entrance baffle plate. This plate contains 30 holes, one per optical beam and in the same pattern as the primary mirror array. Its purpose is to mitigate against stray light from the sky background as seen from between the 30 formation flying spacecraft. If other (non-subset) patterns were to used, the plates would need to be "active", i.e. in that number and placement of apertures would need to be commandable. After passing through the plate the light travels the length of the hub tube (~5.3 meters) and is incident on 30 redirector flats, each of which is 10 mm in diameter and also arrayed in a scaled version of the Golomb array pattern. These flats move in piston, tip and tilt to facilitate pointing, metrology and wavefront control. After reflection off the flats the light comes to focus at the field stop mask and travels to an ellipsoidal secondary mirror (SM) mounted on tip/tilt control actuators. The SM relays the beams to the focal plane science instrument.

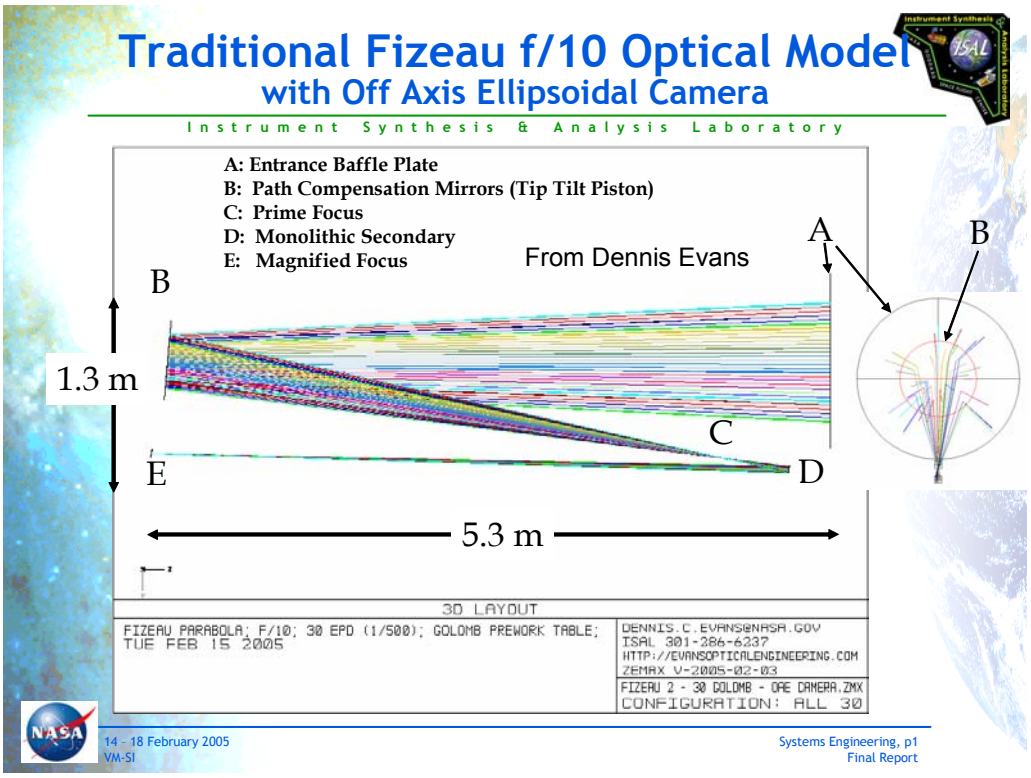


Figure 2.17: A ray trace of the optical design of the baseline architecture.

The Hub Tube Assembly

The primary structural tube serves as optical metering structure and requires a very stiff structure which is stable with respect to time varying thermal gradients.

The tube is subdivided into two thermal cavities by optical benches which are thermally isolated from the exterior tube. The exterior tube (Figure 2.18) uses loop heat pipes to isothermalize its temperature, by transporting heat from the sun-side to the anti-sun side of the tube. There are 4 loop heat pipe circuits with 8 pipes each. Warming the tube to the target temperature relies on using waste heat from the Laser Ranging Assembly/Base Unit Electronics (LRA/BUE), Main Electronics Box (MEB), and Laser Ranging Assy/Base Unit Optics (LRA/BUO). These 3 boxes should be designed to

conduct all their dissipated power into their mounting interface. These boxes are attached to a common LHP (Loop Heat Pipe) thermal conduction plate which is shared with the LHP evaporators and subcooling radiators. The LHP Thermal Conduction Plate is thermally isolated at its mount to the Hub S/C deck.

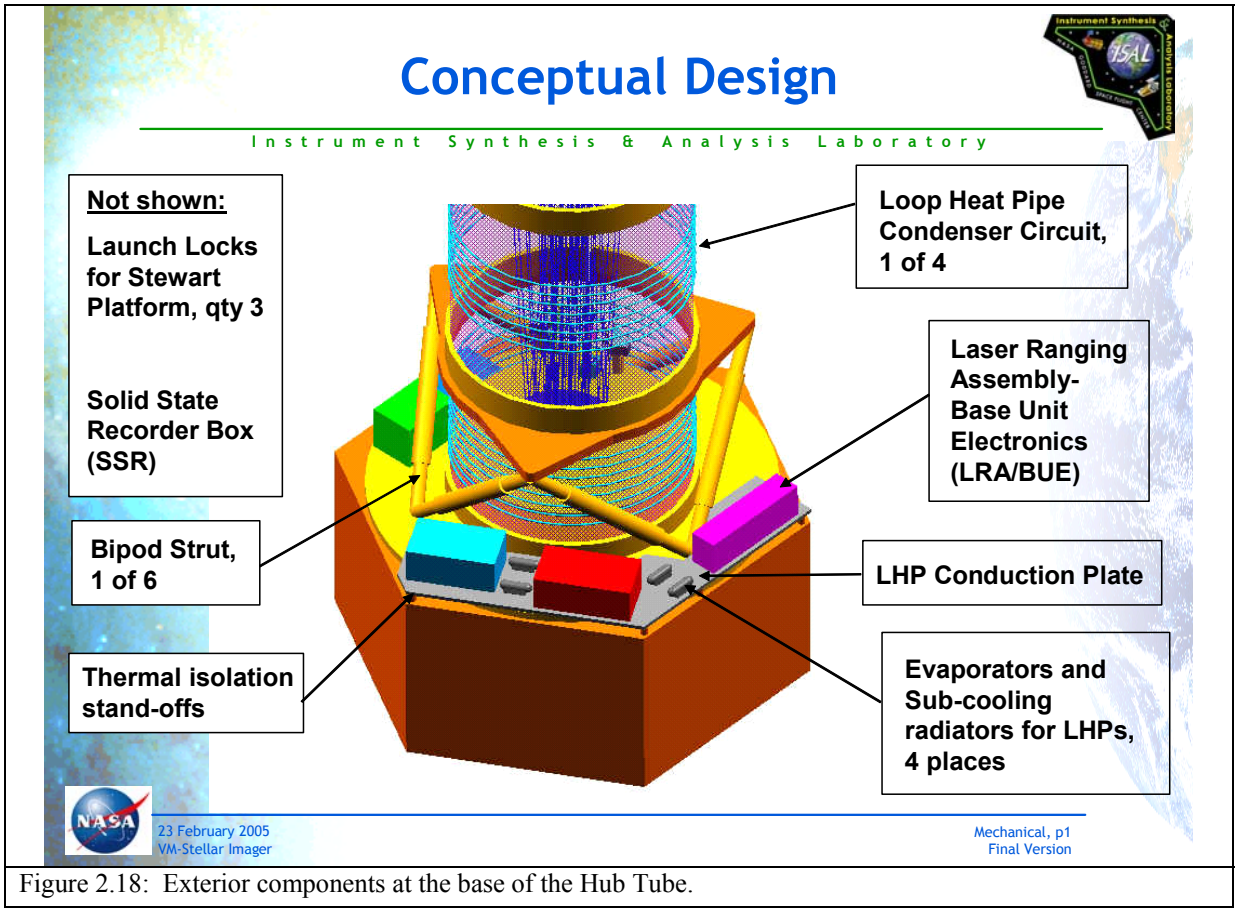


Figure 2.18: Exterior components at the base of the Hub Tube.

Thermal Cavity No. 1 contains the detectors (2 visible science, 2 UV science and 2 wave-front sensing (WFS) channels). Constant conductance heat pipes (CCHP's) are used to transport the detector heat to a radiator on the exterior, anti-sun side of the telescope tube assembly. This radiator will likely have to be at a station elevation to place it above the numerous boxes at the base of the tube assembly, in order to gain an unobstructed clear FOV to cold space. Of the 6 optical channels, 3 are primary and 3 are redundant secondaries. A flip mirror mechanism, just before the detector, is used to select the appropriate primary or secondary detector for each of the UV, visible and WFS channels. Each science channel (UV and visible) has a 16-position filter wheel just upstream of the flip mechanism. Flip mirrors will run cold-biased at 10-20 degrees (exact value TBD) above the cavity temperature, and use heater power to stabilize their temperature to within +/-1 deg. C. All channels have Front End Electronics (FEE) placed near the detectors. Typically, these must be located within 18 inches of the detectors. The signal A/D conversion occurs here.

Thermal Cavity No. 2 contains the beam compensator optics with mechanisms, secondary optic with mechanism, and a cover/shutter mechanism. The optics will have heaters bonded directly to their backs and be thermally isolated from the mechanisms to which they attach. These optics will run cold-

biased at 10-20 degrees (exact value TBD) above the cavity temperature, and use heater power to stabilize their temperature to within +/-1 deg. C. This scheme will drive most contaminants off of the mirrors. (See Figure 2.19).

Each of the optical benches, to which secondary structures attach, will also require a very stiff structure that is stable with respect to time varying thermal gradients. These benches must be thermally isolated with respect to the Hub Tube to which they attach.

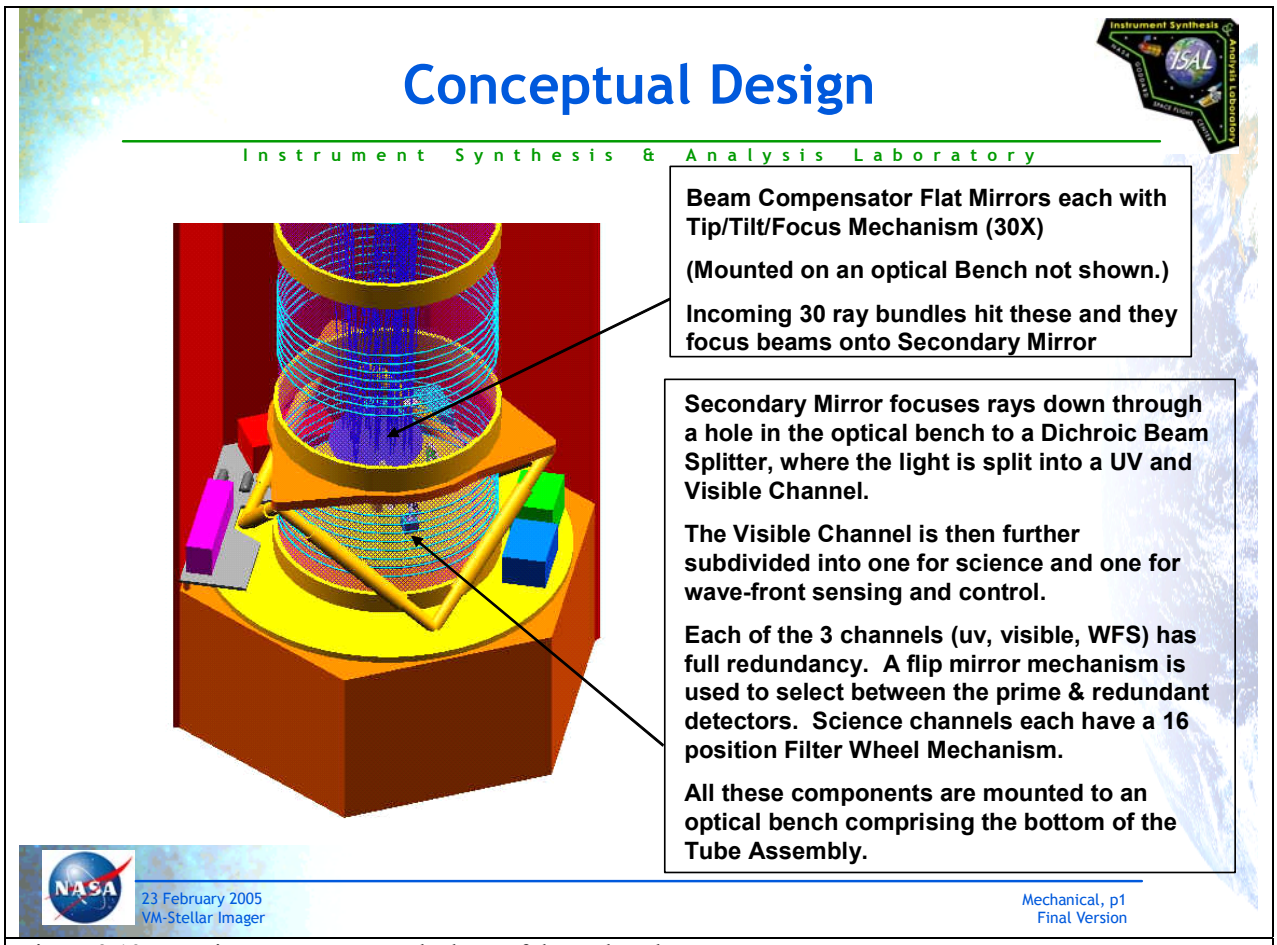


Figure 2.19: Interior components at the base of the Hub Tube.

At the top exterior (see Figure 2.5) of the Tube Assembly is a shelf, supported by gussets, to which each of 30 Laser Ranging Assy Remote Units (LRA/RU) attach. The purpose of these LRA/RUs is to establish range between the 30 MirrorSats and the Hub to within TBD microns. Therefore the mount for these boxes will need to be stable with respect to time-varying, thermal gradient induced distortions. The plan is for a fixed sunshade/solar array, independently attached to the Hub S/C Bus, to shade the entire telescope tube assembly. This puts about one half of the LRA/RU's in the space between the Telescope Tube and the fixed solar array. Future thermal analysis will have to address whether there is enough cold space view for the TEC cooled detectors to radiate off the top of the LRA/RU boxes, when faced with the radiative heat load off the back of the solar arrays. It may require that all 30 boxes be mounted on the anti-sun side of the Telescope Tube. Also, the height of the solar arrays above the telescope tube will be driven by the required shading and straylight baffling when the

Hub S/C is pointing +/-20 degrees off of the orthogonal to the sun. At the Hub Tube Assembly center of gravity (cg) is located the exterior interface ring to the Stewart vibration isolation platform. The entire Hub Tube Assembly is suspended above the Hub S/C deck via this attachment. The tube interface ring provides 3 mounting locations to the upper deck of the Stewart vibration isolation platform. The thermal and vibration isolation considerations are shown in Figure 2.20.

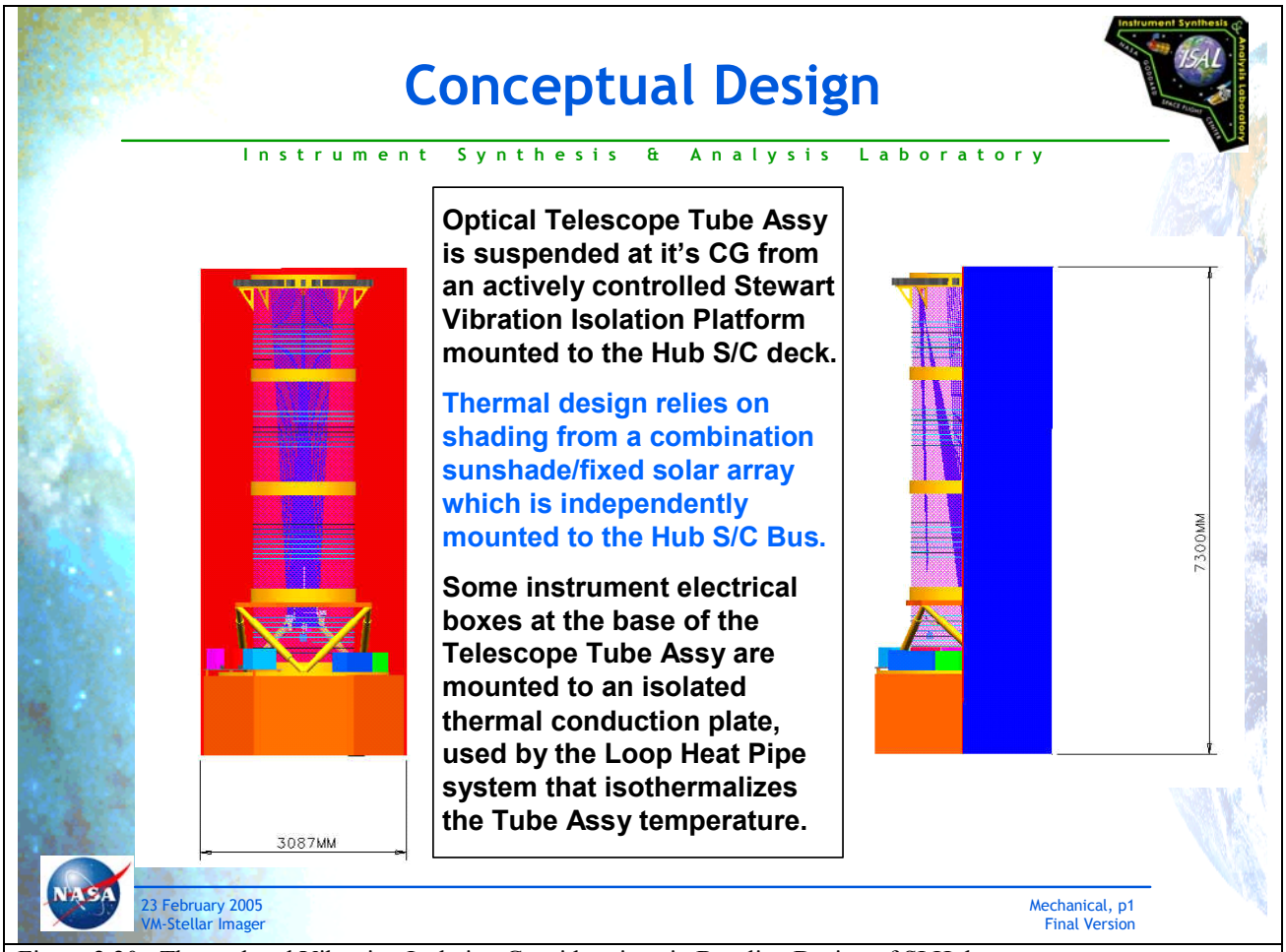


Figure 2.20: Thermal and Vibration Isolation Considerations in Baseline Design of SI Hub.

2.2.2 The SI Focal Plane Instrument Package in Baseline Design

The focal plane science instrument package consists of 3 cameras: (i) UV science camera, (ii) Visible science camera and (iii) wavefront sensing camera (Figures 2.21, 2.22, 2.23). The UV science camera is 5243 X 5243 pixels, with a Nyquist sampling at 1550 Å of λ/2B (where B=max. baseline) of 32 μas and a full science field-of-view (FOV) of 168 mas. The visible science camera has 5243 X 5243 pixels, while the wavefront control camera has 10486 X 5243 pixels with Nyquist sampling at 5000 Å of 103 μas and a FOV of 541 mas. The larger format of the wavefront sensing camera enables the simultaneous recording of two "diversity" images of the source on the same detector. It also could be used for visible light science, as a "wider field camera" than the dedicated (higher resolution, smaller FOV) visible science camera. The wavefront sensing camera and optical control will be described in detail in section 2.2.3 below. Each of the channels has two identical, redundant detectors to ensure long lifetimes. The two science channels have, in this baseline design, filters wheels in front of the detectors to produce the desired bandpasses for the observations. Alternative designs are envisioned which could replace this filter+standard detector set with either energy-resolving detectors, or with a more complex optical system that re-maps the 2D distribution of the beams into a 1D non-redundant array, whose light is then dispersed orthogonally at every point to produce more complete spectral information (Figure 2.24, Sections 2.2.4 & 3.1.8).

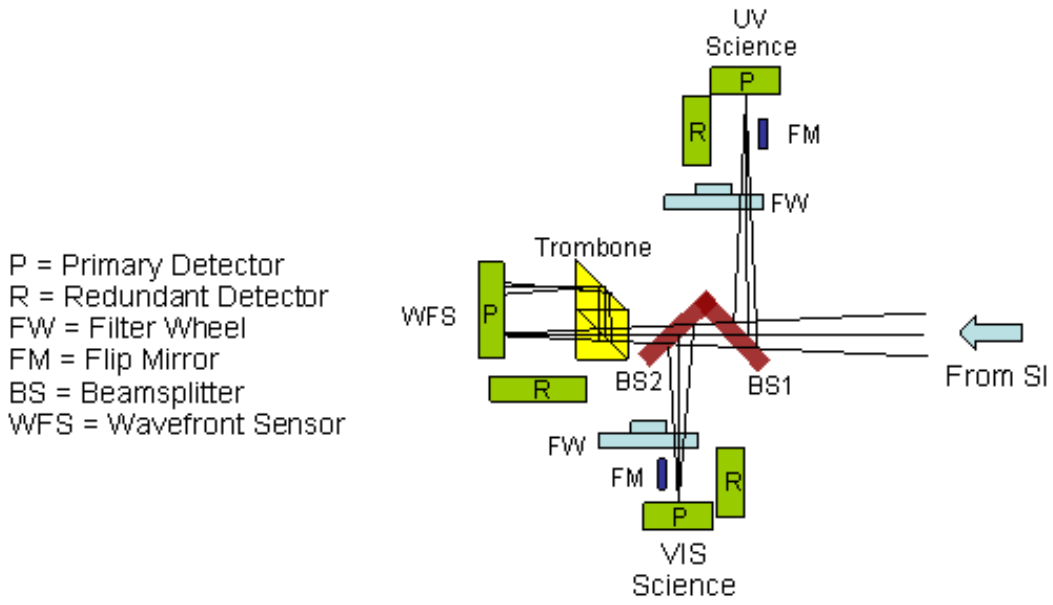


Figure 2.21: The science and wavefront control focal planes of the SI beam-combiner.

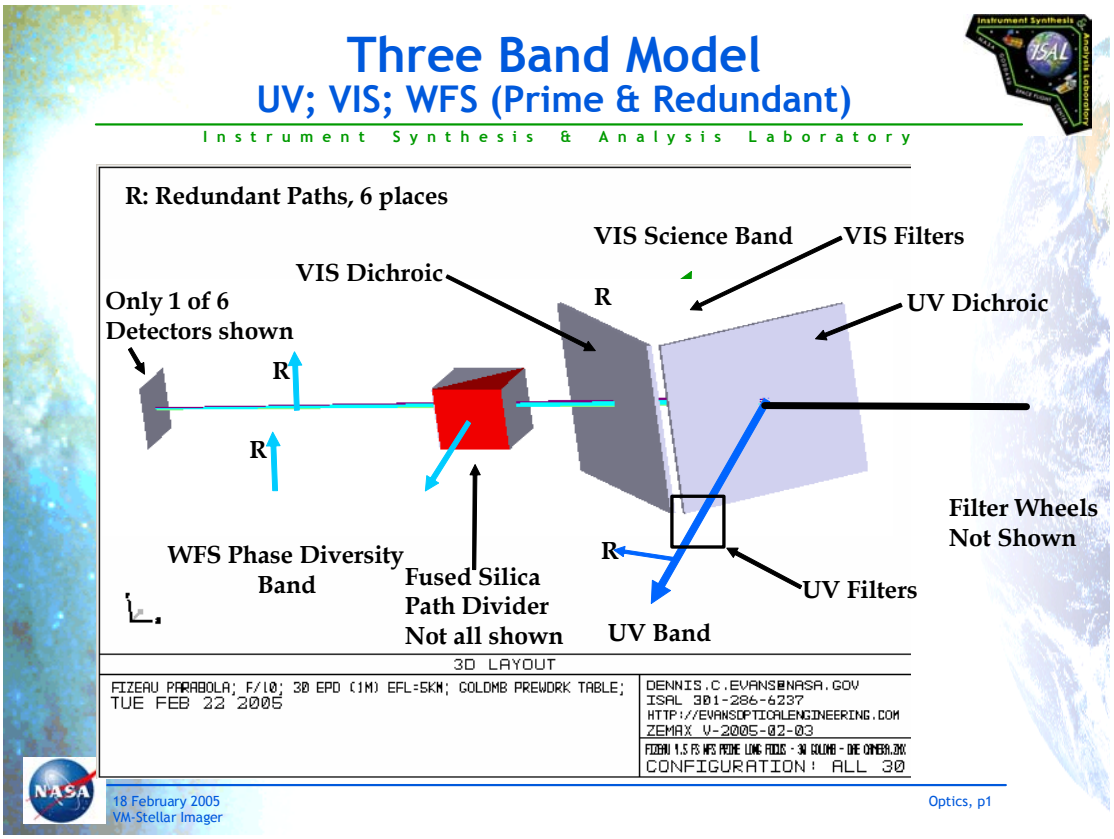


Figure 2.22: The focal plane in 3D.

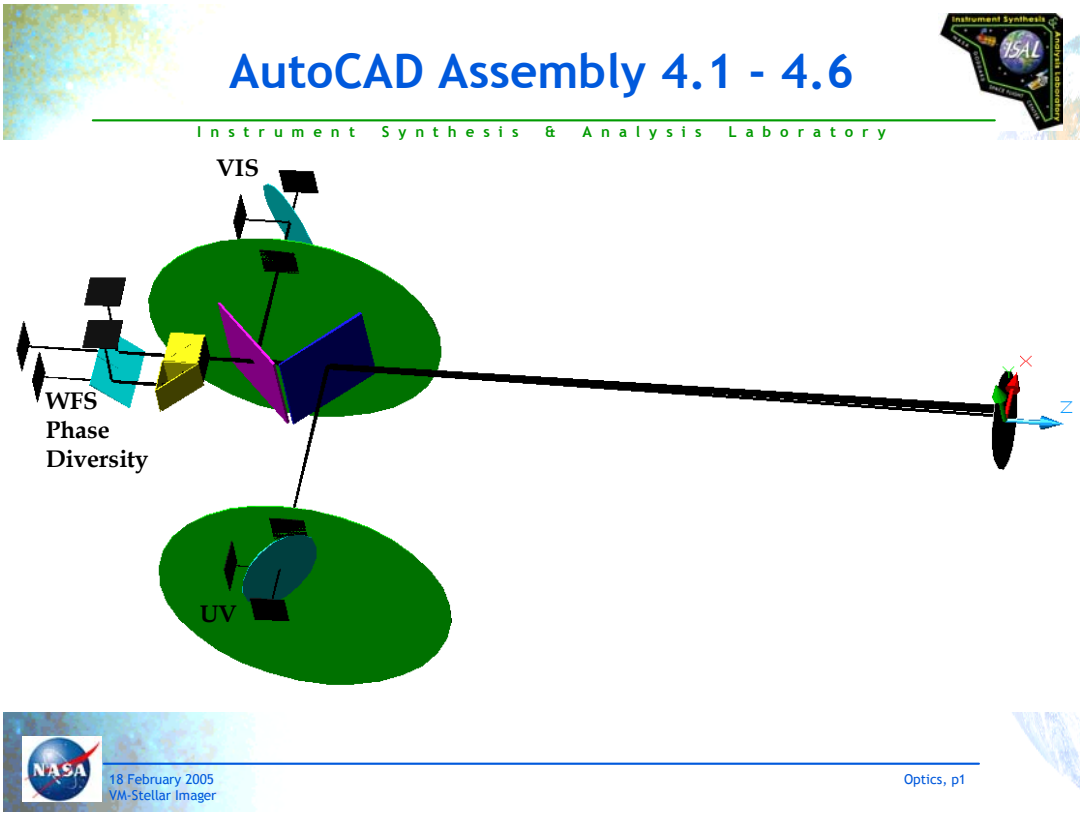


Figure 2.23: An AutoCad drawing of the focal plane.

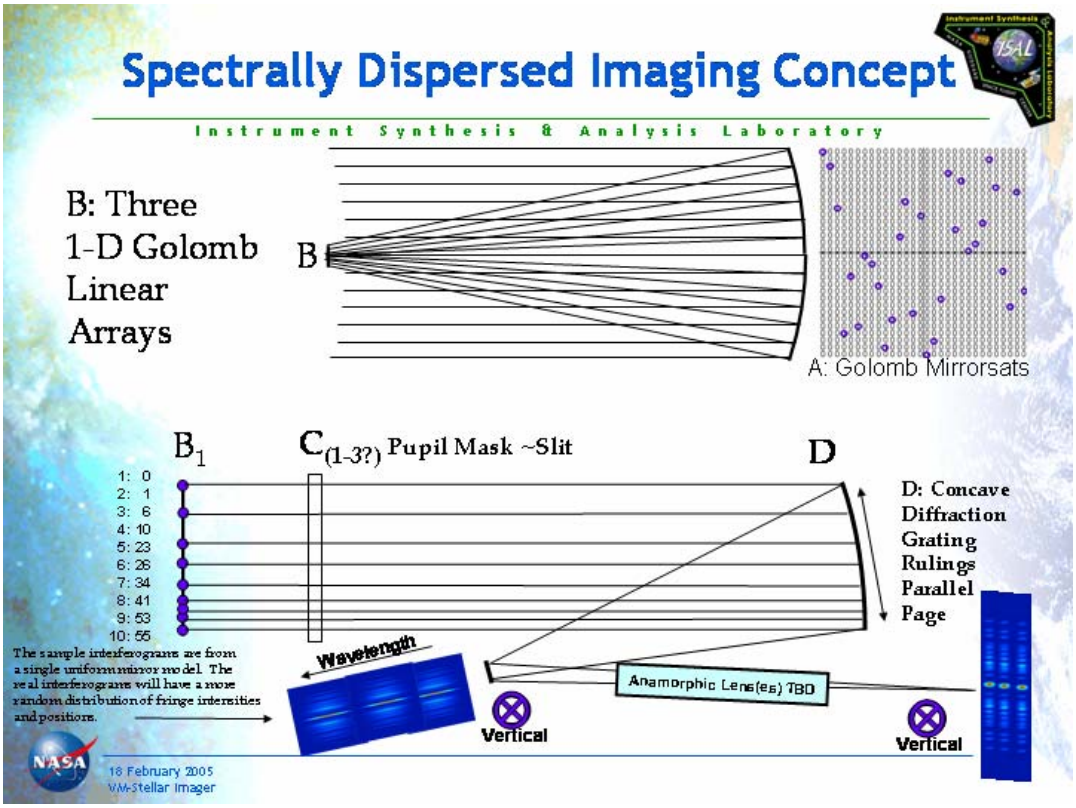


Figure 2.24: An alternate design to enable spectral dispersion of the beams.

The specifications for each of the science cameras are as follows:

- UV science camera
  - Detector pitch = 7 microns
  - Back thinned silicon CCD with special UV coating
  - UV science array size = 5243 x 5243 pixels
    (note final array size may be 4kx4k or 8kx8k)
  - Number of bits = 16
  - Temperature 173K +/- 1 K
  - Readout rate = science requires 1 frame per 10 minutes, but shorter subexposures may be necessary or desired to mitigate cosmic ray hits, background levels, etc.

- Visible science camera
  - Detector pitch = 7 microns
  - Back thinned silicon CCD
  - Visible science array size = 5243 x 5243 pixels
    (note final array size may be 4kx4k or 8kx8k)
  - Number of bits = 16
  - Temperature 173K +/- 1 K
  - Readout rate = 1 frame per minute

If filters are used to obtain spectral information (rather than energy resolving detectors or spectral dispersion on remapped beams mentioned above), then the minimum filter set for the UV channel requires 7 "line" filters + 7 "continuum" filters near those lines, one "open" position" and 1 spare position, as follows:

Table 2.1: Minimum filter set for UV wavelengths

UV Lines & Broadband	**Center λ (Å)**	**Bandwidth (Å)**
Ly α	1216	10
Broadband	Offset from above	100
O I	1303	10
Broadband	Offset from above	100
C II	1335	10
Broadband	Offset from above	100
Si IV	1400	10
Broadband	Offset from above	100
C IV	1550	10
Broadband	Offset from above	100
He II	1640	10
Broadband	Offset from above	100
Mg II	2800	10
Broadband	Offset from above	100

The minimum filter set for the visible wavelengths science camera contain two "line" filters (for Ca II h&k at 3934 and 3968 Å) and five broadband regions distributed over the 3000 to 5000 Å zone, plus one "open" position on the filter wheel, as follows:

Table 2.2: Minimum filter set for optical wavelengths

Visible Lines & Broadband	**Center λ (Å)**	**Bandwidth (Å)**
Ca II h	3934	2.5
Ca II k	3968	2.5
Broadband	3000	100
Broadband	3500	100
Broadband	4000	100
Broadband	4500	100
Broadband	5000	100

It would be highly desirable to have a "tunable" filter (or, better yet, an energy/spectral resolving instrument, see, e.g. Appendix I) for the Ca II lines, since their width varies from target to target and it is very important to be able to isolate the emission from the surrounding, strong continuum. An energy resolving instrument is also better for the UV lines, though not quite as important due to the

weaker continuum at those wavelengths.

The specifications and capabilities for the wavefront sensing and control camera are:

- Separate from visible camera for science
- Control loop using wavefront sensing camera gradually picks up from the laser ranging system. Optical path length control goes from within a few microns for the laser ranging system to within 3 to 5 nm with the WSC. With the WSC, optical path length is sensed to the one nm level. The RF control system, laser ranging system, and WSC system are all operational at the same time.
- Detector pitch = 7 microns
- Back thinned silicon CCD
- Visible fringe array size = 10486 x 5243 pixels (note final array may be 8kx4k or 16kx8k)
- Number of bits = 16
- Temperature 173K +/- 1 K
- Optics separates beam into two beams
- No moving parts for optics or camera except mirror which flips into place for redundant camera
- Focal plane twice the size of science camera focal plane to accommodate the two sets of fringes
- Frame rate and impact on control loop
    - If camera frame rate equals 1 Hz, then
        - Typical control loop at 0.1 times frame rate = 0.1 Hz
        - Conservatively correct disturbances from dc up to about 0.5 times control loop rate = 0.05 Hz
        - Disturbances above 0.05 Hz likely caused by instrument or spacecraft
        - Try to assure no spacecraft or instrument disturbances above 0.05 Hz
        - Increase camera frame rate if necessary

2.2.3 Wavefront Sensing and Control System

Wavefront sensing and control (WFC) is the final method by which SI is kept within specification to realize the unprecedented resolution requirement needed to obtain the science. A top-level schematic of the SI WFC system is shown in Figure 2.25 and will be discussed herein.

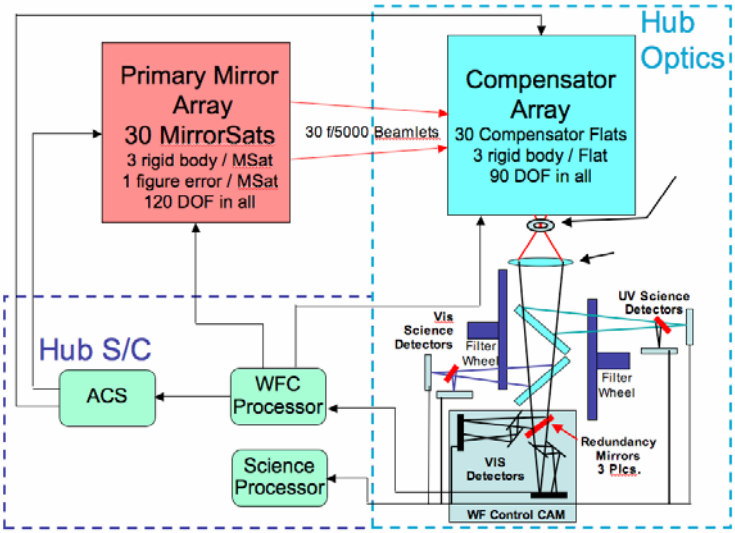


Figure 2.25 – Schematic of WFC System

During its mission life Stellar Imager (SI) will sequentially cycle through a series of science targets, one at a time. For each science target SI must repoint the hub spacecraft, slew and repoint each of primary mirror array spacecraft, acquire the new source on the focal plane, rephase the primary mirror array and hold the pointing and optical phase stable throughout the science observation. This process can be broken into a sequence of steps beginning at the end of the science observation of the current source and continuing through to the end of the science observation of the new source. After finishing a science observation the hub spacecraft is rotated under star tracker control to point its back end, i.e. its end away from its entrance aperture, toward the new science source. Each of the mirrorsats will be separately *repointed* to bring light into the entrance aperture of the hub and through to the science cameras and wavefront sensing and control (WFC) camera (WFCCam). The WFCCam (light blue box, lower right of Figure 2.25) collects images, shown in Figure 2.26, and processes them through the WFC processor to recover the phase errors (wavefront errors) (Figure 2.27) in propagating through SI and also to the science processor to apply software phase corrections to the science imagery as it is being collected. The WFC processor feeds back to the attitude control system for repointing and translations of the

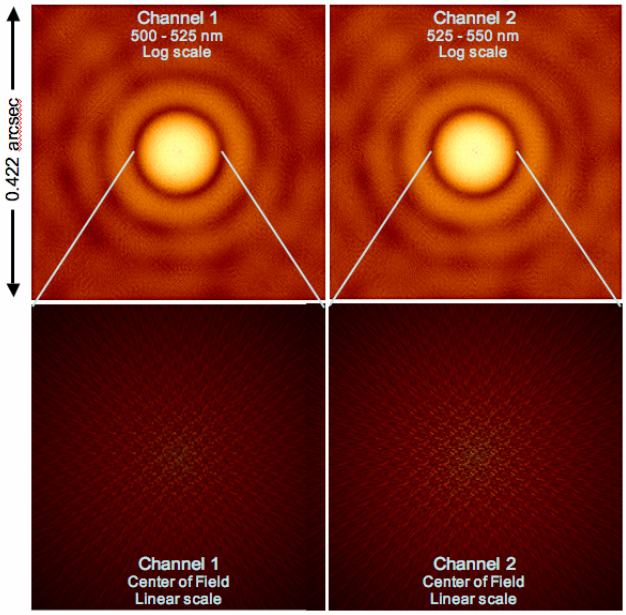


Figure 2.26 – WFCCam Images

Images collected on the WFCCam detectors, upper row is channel 1 and 2 respectively on a log scale. Lower row shows center of images on linear scale. Images differ by 3 waves of defocus and in passband.

individual spacecraft and to a set of 30 compensator flats within the hub spacecraft. During normal operations the system is completely autonomous and operates in a closed-loop fashion.

Light of wavelengths 500 – 550 nm is split off from the science beam and sent into the WFCCam in a converging beam. Within the WFCCam are two redundant optical paths which can be switched via a rotating beamsplitter; only one path operates at a given time. Within each path is an optical "trombone" consisting of a beamsplitter and a flat. The beamsplitter reflects light from 500 – 525 nm and passes light from 525 – 550 nm. The reflected light subsequently reflects off a flat and both the 500 – 525 nm and 525 – 550 nm images are brought to focus on the photon counting detector chip, but at different locations. Since the two paths are of non-equal length the two different images are actually at different foci with approximately 3 waves of focal difference between them and are at slightly different passbands. Each of the images are 5243 x 5243 samples and are sampled at λ/B_{max}= 0.103 milli-arcseconds, i.e. are Nyquist sampled at the shortest wavefront sensing wavelength of 500 nm. The two spatially separated images of the extended stellar object are known as *diversity* images since they have known wavefront differences between them in terms of deliberately induced delta-focus and delta-wavelengths. The two diversity images are periodically passed to the WFC processor which employs an algorithmic approach to separate the unknown object phases from the unknown phase incurred in propagation through SI. The net effect of using two different wavelengths is to use an effective wavelength of $\lambda_1\lambda_2/(\lambda_2-\lambda_1)=11$ microns, thereby countering the problem of 2π ambiguities so long as the system does not stray by more that +/- 5.5 microns from alignment. Using two different foci allows resolution of the problem of separating the object phase from the system phase; nominally an ill-posed problem for a single image. Algorithms of this type are known as *phase diversity* [Gonsalves, 1982, Kendrick et al. 1994] and will be described in detail in section 3.1.4.

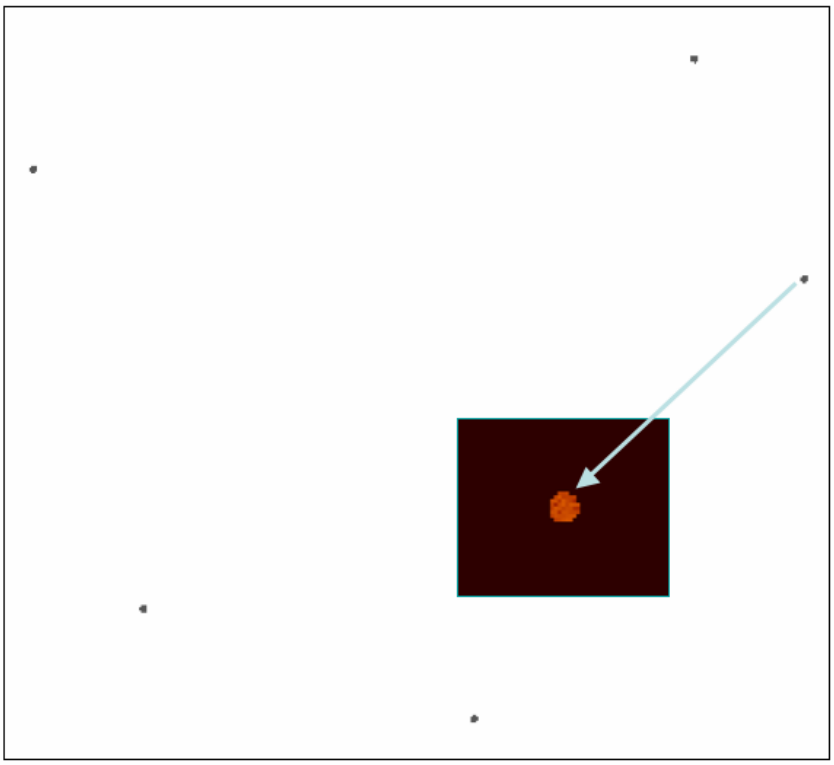


Figure 2.27 – Phase Diversity Recovered Wavefront Error.
Recovered wavefront error over a portion of SI aperture sampled at locations of apertures. Inset shows error recovered over single sub-aperture

Use of a WFCCam very near to the science focal plane allows for mitigation of non-common path errors typically introduced in metrology systems. Since the light propagates the same optical path up to the beamsplitter which separates the light into the science camera and WFCCam only errors with occur after the beam splitter introduces any systematic errors in the system.

Figure 2.28 shows the wavefront sensing and optical control during normal operations. Throughout science observations two WFC images are periodically collected with the WFCCam and processed through modulation transfer function (MTF) fitting and phase diversity. MTF fitting recovers aperture lateral translations along the virtual PM surface, i.e. the baselines; the algorithm will be discussed in more detail in section 3.1.4. Phase diversity recovers the wavefront error. Phase diversity uses inputs from MTF fitting in order to construct relative aperture locations. Estimated wavefronts are decomposed into correctable piston, tip and tilt modes and into the uncorrectable higher order modes.

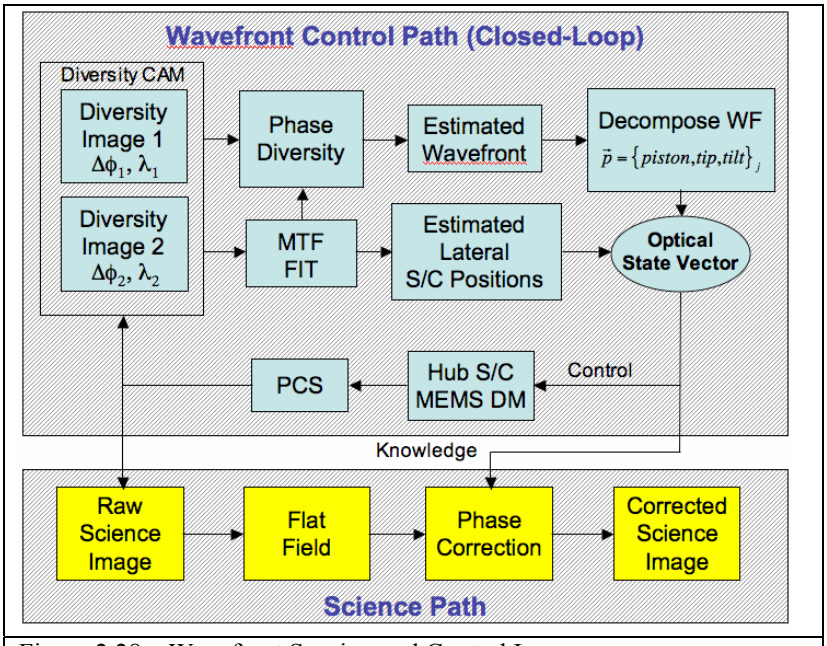


Figure 2.28 – Wavefront Sensing and Control Loop

Small correctable modes are fedback to the hub MEMS compensator mirror array, and the larger lower bandwidth errors are feedback to the PCS pointing. The MEMS mirror array consists of 30 separate flat mirrors, mounted on a curved surface, each of which is individually articulated in piston, tip and tilt. Piston, tip and tilt errors are simultaneously fedback to the science processor to phase-correct the science image in-situ. In-situ refers to the science image being periodically read out, phase corrected and summed with the previous time step. The entire process is cycled in closed loop control throughout science observations to maintain image quality within specification.

Simulations have shown that pointing and phase (wavefront) must be known to ~λ/30 rms wavefront error (5.17 nm WFE) in order to phase correct the science imagery to λ/10 at the shortest science wavelength of 1550 Angstroms. Absolute pointing and phase control needs to be maintained to < ½ $\lambda^2/\Delta\lambda$ = +/-5 microns for a 5% passband. Treating this as a 3-sigma number implies that for any baseline pair the rms pointing and phase difference must be maintained to 1.67 microns rms per baseline pair; if we allocate this uniformly between piston and tip/tilt and use $\sigma_T^2 = \sigma_p^2 + \sigma_{tip/tilt}^2$ then $\sigma_{piston} = 1.18$ um and $\sigma_{tip/tilt} = 1.18$ um and converting the tip/tilt allocation to angle yields 0.244 arcseconds. This is significantly looser than 5.17 nm knowledge requirement since phase diversity will be used to sense the phase corrections that will be applied to the science imagery. Thus the science imagery will be phase corrected, using the estimated phases from phase diversity, to the required 15.5 nm.

Phase diversity recovers the wavefront differences between all baseline pairs of apertures and as such its performance is driven by the number of photons, visibility and stability of the fringes across the image during collection of the diversity images. Mathematical models of this process are developed in section 3.1.4. Figure 2.29 shows the required number of photoelectrons per baseline pair of apertures versus baseline to sense 10 microns of OPD with phase diversity. The number of photoelectrons are

for a 2 milli-arcsecond diameter source, with 5% ($\Delta\lambda$=25nm) passband filter on the WFCCam. Increasing number of photoelectrons are required with increasing baseline separation due to the corresponding decrease in visibility from the extended source. Shown are 4 curves defining the number of photo-electrons needed to achieve a given level of OPD knowledge as a function of baseline length. The curves represent sensing an OPD of 10 μm +/- {λ/10 - black, λ/20 - red, λ/30 - blue, λ/40 – green}. For a 500 meter baseline to sense to 5.17 nm we need ~5 x 10^4 photoelectrons per integration period of the WFCCam. Figure 2.30 plots integration time versus baseline for a 2 milli-arcsecond extended source of visual magnitudes 5 – 15. A 13th-magnitude star will allow collection of enough photons to sense the OPD to 5.17 nm in 10 seconds.

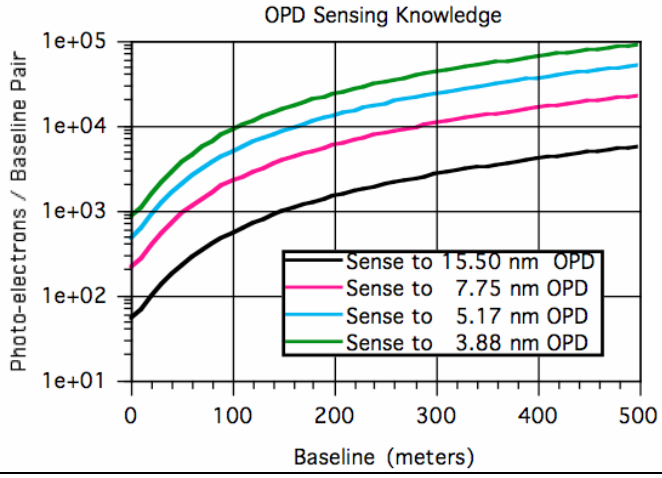


Figure 2.29 – WFC Photo-electrons per baseline pair versus baseline required to sense 10 um of OPD between the apertures in a 5% passband (25 nm).

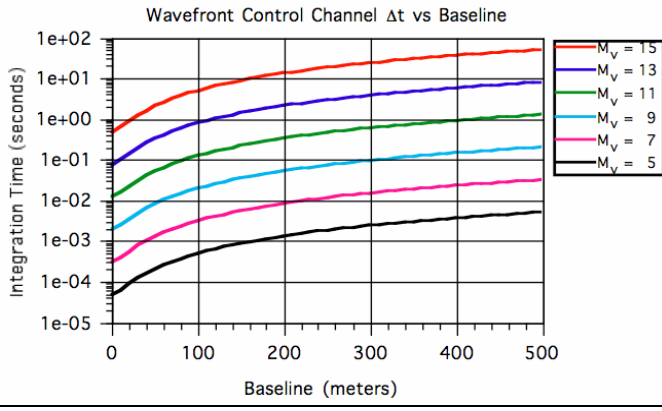


Figure 2.30 – Integration time to sense 10 microns of OPD to 5.17 nm versus baseline and visual magnitude with a 5% passband and 50% throughput.

Knowledge of the relative lateral locations (baselines) of the individual spacecrafts along the virtual primary surface must be known to better than 10 cm (1/10th of an aperture width) to determine the baselines and hence the pupil function used for wavefront control and for accurate phase correction. This is accomplished using the MTF fitting algorithm operating on the WFCCam images provided all the images from the 30 apertures are within 0.244 arcseconds of the center of the WFCCam detector.

Both repointing and phasing are initially accomplished by the laser metrology system which brings each of the beams to relative pointing within the science and WFC focal planes to approximately ¼ of an Airy disk width per primary beam and the relative piston differences to < +/- 5 microns per baseline pair. This effectively removes tip/tilt errors to approximately ¼ arcsecond and allows calibration of systematic errors to this same level between the laser metrology system and WFC system and science focal plane due to the non-common paths of these three systems. Following this each of the beams must be phased in piston. The WFC focal plane, operating in the visible, will be used for the initial hand-off from the laser metrology systems and during fine phasing of SI and for monitoring and maintenance of the wavefront quality during a science observation. The WFC detector is actually two CCD chips, placed side-by-side (or one long chip), each of size 5243 x 5243 per channel and is Nyquist sampled, at λ = 500 nm, for the longest baseline (B = 500m) giving 103 micro-arcseconds per pixel. The diameter of the core of the Airy disk (primary beam point spread function) is 2.44 λ/D = 252 milli-arcseconds. Thus the full field of view (FOV) of a single WFC CCD chip is 5243 x 103 micro-arcseconds = 540 milli-arcseconds or 2.14 Airy

diameters. In order to centroid and accurately place and hold each primary beam onto the WFC detector we desire that a primary beam Airy disk core be centered on the WFC detector chip, thus we desire at this stage of the control that each mirrorlet spacecraft be absolutely pointed to within ½ of Airy disk or ¼ of the field of view with a precision of ½ an Airy ring, i.e. to ~270 milli-arcseconds +/- 52 milli-arcseconds.

Each of the mirrorlet spacecraft contains a curved portion of the primary mirror each of which should ideally be separately focused onto the WFC detector and ultimately the science detector. In principle this could be accomplished by tip/tilting all but one of the mirrorlet beams outside the FOV of the WFC detector and pistoning the one mirrorlet still within the FOV and looking for the minimal primary beam size. However this is impractical due to slow F/# = 5000 giving a large depth of focus of $\lambda(F/\#)^2$ = +/- 12.5 meters. While not a practical method for SI it could be used as contingency option should a mirrorlet spacecraft become significantly lost within its piston range.

The baseline method for estimating the phase differences is known as *Phase Diversity* [Gonsalves 1982, Kendrick et al. 1994]. Phase diversity requires as input: two (or more) collected images from the wavefront control camera with known phase differences (non-common path wavefront error), an optical model of the system, and statistical model of the noise. With these inputs a Gonsalves metric is constructed of the form:

$$E(\vec{a}) = \sum_{u,v} M(u,v) \frac{\left| OTF_1(u,v;\vec{a})\tilde{I}_2(u,v) - OTF_2(u,v;\vec{a})\tilde{I}_1(u,v) \right|^2}{\left| OTF_1(u,v;\vec{a}) \right|^2 + \left| OTF_2(u,v;\vec{a}) \right|^2} \tag{1}$$

where $OTF_1(u,v;\vec{a})$ and $OTF_2(u,v;\vec{a})$ are the u,v-plane optical transfer functions for channels 1 and 2 respectively, $\tilde{I}_1(u,v)$ and $\tilde{I}_2(u,v)$ are the uv-plane spatial spectrums (2D Fourier transforms) of the observed images, and $M(u,v)$ is a binary mask that is unity at the valid spatial frequency points within the uv-plane and zero otherwise. $\vec{a}$ represents the vector of unknown piston, tip and tilts of each of the mirrorlet beam paths; note that higher order terms can be included. The solution vector is the minima of equation (1) with respect to $\vec{a}$ and is found by nonlinear optimization methods. Numerous other metrics can also be used to solve the problem, see [Kendrick, 1994] for additional choices. Phase diversity algorithms will be developed in more detail in section 3.1.4.

A top-level error budget for the wavefront sensing and optical control is shown in Table 3.1. The handoff from the laser metrology to WFC can occur at +/- 5 microns (3σ) of optical path difference error per baseline pair of apertures and at +/-0.244 arcseconds (3σ) of tip/tilt per aperture. The laser metrology system can continue operation throughout wavefront control. The WFC system senses at 0.1 Hz (once per 10 sec) and controls at 0.01 Hz (once per 100 sec) down to 13th visual magnitude source and generates an update to send to the attitude control system and an update to the phase correction for the science image. Much faster control is possible on brighter stars. The residual phase diversity sensing errors are < 3.66 nm (3σ) for OPD per baseline pair for the longest baseline and < 0.76 milli-arcseconds (3σ) for tip/tilt per aperture. During the 10 second interval the OPD cannot drift by more than 8.41 nm (3σ) per baseline pair and by 2.45 milli-arcseconds per aperture in tip/tilt. The phase correction algorithm has errors similar to phase diversity of 3.66 nm (3σ) for OPD correction and 0.76 milli-arcseconds for tip/tilt correction. The aggregate errors in closed-loop are 1/10 wave for the shortest science wavelength of 155 nm.

Table 2.3: Wavefront Control Error Budget

Laser Metrology Handoff		
Piston	+/- 5 microns	(3σ) per baseline pair
Tip/Tilt	+/- 0.244 arcsec	(3σ) per aperture
Lateral Translation		
Δx, Δy	10 cm	apertures knowledge wrt reference aperture
Phase Diversity Sensing Residual		
Piston	3.66 nm	(3σ) per baseline pair
Tip/Tilt	0.76 mas	(3σ) per aperture
Stability		
Piston	8.41 nm	(3σ) per 10 seconds
Tip/Tilt	2.45 mas	(3σ) per aperture
Phase Correction Residual		
Piston	3.66 nm	(3σ) per baseline pair
Tip/Tilt	0.76 mas	(3σ) per aperture
Total Closed-Loop Science Channel Error	**15.5 nm**	**(3σ)**

[1]All units are in wavefront OPD
[2]mas = milli-arcseconds

2.2.3 Alternative Hub Designs

Alternative, more sophisticated hub designs than assumed in the baseline design are also well worth considering. These would be based on the alternative beam-combination designs described in Appendix I, which include not only the straightforward "Fizeau" (pupil-plane beam combination) design, but three other possible architectures. These alternatives include: a hypertelescope architecture (Labeyrie, 2002)**;** a two-dimensional, spatial-frequency remapping architecture (a variation on the hypertelescope using "partial densification" of the pupils); and a one-dimensional, spatial-frequency remapping architecture. Although these alternatives may be more difficult to construct in the near-term, they are quite plausible in the mid- (2015) and longer-term (2024) timescales envisioned for the SI Pathfinder and full SI mission, respectively, and they hold great promise for increasing the efficiency of the observing program and the amount of spectral information that can be gathered by the observatory along with the high angular resolution imaging. Appendix I also discusses the possible use of "Closure Phase" to reduce pathlength control requirements.

2.3 Infrastructure and Constraints Assumed in Place for the Time of Implementation

The design and implementation plan presented in this document for the Stellar Imager does not require major improvements in infrastructure for a 2024 launch. Heavy lift vehicles in the Delta IV Heavy (or the future Atlas V heavy) are assumed available to launch the entire constellation in one or two launches – which is the most efficient ways to launch and deploy the observatory, though more numerous launches on smaller ELV's could be utilized if needed. Capabilities for supporting significant science and operations telecom data rates to/from Sun-Earth L2 are assumed (Rough assumptions for SI data collection rates include 900 kbps daily average for 11 months/year and 5 Mbps average for 1 month/year). The most important capability not currently available would be the ability to reach and service facilities in Lissajous orbits around the L2 point. The long lifetime goal for SI suggests that it could benefit greatly from a human and/or robotic capability to refuel at a minimum, and, optimally, service the various components of the mirrorsats and hub – and the design of all the spacecraft is envisioned as modular to enable servicing/exchange of the various important components.

2.4 Role of humans or robots for in-space servicing

Although the SI baseline design does not require that humans and/or robots be able to access and work on SI at the Sun-Earth L2 site, the mission could benefit greatly from such a capability. In particular, the long lifetime requirement for SI (5-10 years or more) is most easily met if the design can be made modular so that humans and/or robots can readily service and replace key components of the mirrorsats and hub. An obvious and simple capability that would help enable SI would be the ability to refuel the spacecraft to ensure it will be able to perform station-keeping/orbit maintenance and target-to-target maneuvering over the desired long lifetime. Servicing of the critical hub spacecraft would also be of great utility, since it, unlike the mirrorsats, is a single-point failure, unless more than one hub is launched (or is available for launch-on-need).

In the timeframe of interest for the 10-year mission life of SI, NASA systems developments for Moon, Mars, and in-space sites may be expected to produce major advances in robotic capabilities. These developments and mission deployments will make a variety of robots increasingly available with increasing performance and versatility for assembly, servicing, and related operations in space. The Sun-Earth L2 site is relatively unfavorable for human visits, due to radiation hazards and the vehicle provisions and flight durations required for return. Robotic servicing at L2 will be relatively advantaged early in this timeframe, although the operations will require extensive autonomy, with limited supervision by ground-based controllers, due to command-response latencies of several seconds at that distance. Modular systems with simple interfaces are particularly well suited for such operations. As the capabilities for human activities are extended to Mars and other destinations later in the timeframe, visits to L2 by humans may become a more readily supported approach. Although the uniquely human capabilities for handling unstructured situations could conceivably relieve the need for simplifying operations through modular design, modularity will continue to support increased efficiency and risk reduction.

In-space servicing of the SI hubs or mirrorsats will require provisions for access, capture, and handling by the servicing system visiting vehicles and robots or EVA astronauts. Standard features and modular designs greatly reduce the mission risks, costs, and operations impacts associated with servicing compared to high-risk handling of non-modular, unaccommodating systems and components.

The baseline design from the IMDC and ISAL Vision Studies does include basic modularity to enable such servicing; i.e., system partitioning and clear access are features of the design. Additional features will be needed in a detailed design, e.g., module handling provisions (markings, grasping points, positioning guides, kinematic mounts, blind mating connectors, position locks or latches), module self-protection (covers, blankets, shields, keep-alive power), and verification provisions (built-in test equipment, configuration indicators, interfaces for diagnostic tools) as needed for each type of module. Some of the passive features can be integrated into the structure with slight impact to system mass and complexity by adding surface markings, reshaping of hardpoints, rearrangement of elements, or replacements for non-separable mounting interfaces and connections. The additional degrees of freedom needed for operations of separable interfaces generally lead to some increment of mass and complexity.

Refueling provisions will involve more complexity. Approaches to refueling include propulsion module exchange, fuel canister exchange, or retanking of fluids. Module exchange may have simple interfaces as used in other modular servicing. If the thrusters, controls, and integrated module systems are inexpensive, the simplicities may outweigh mass considerations. Fuel canister exchange and retanking would require leakless fluid joints and verification, and retanking would also need zero-gravity fluid transfer pumping and control systems, a significant complexity for the servicer system. Refueling systems may be in operational service for a variety of missions during the SI timeframe, so accommodating the prevailing support system approach may be advantageous. However, for refueling a series of mirrorsats, a system that imposes minimal accommodation burden on the mirrorsats may be more strongly advantaged.

Such modular designs may be increasingly used with standardized interfaces as industry providers move to supply systems that are compatible with the future NASA architecture. Earlier experience has demonstrated cost savings through operational efficiencies in the integration and test of modular systems, with additional savings available to ongoing programs using a family of standardized modular components; these benefits are independent of assembly and servicing.

Design for in-space servicing of the hubs and mirrorsats imposes more than modularity, i.e., markings or targets; grasping points for capture; docking ports for the temporary structural coupling of the visiting servicer; power ports and data ports for support to joint operations; integrated structural hardpoints for supporting local manipulation, mobility loads and torque reactions. Operational considerations and contamination control are also important.

Chapter 3

Technology

The technology requirements for SI are derived from a flowdown of the science requirements, as shown in Figure 3.1:

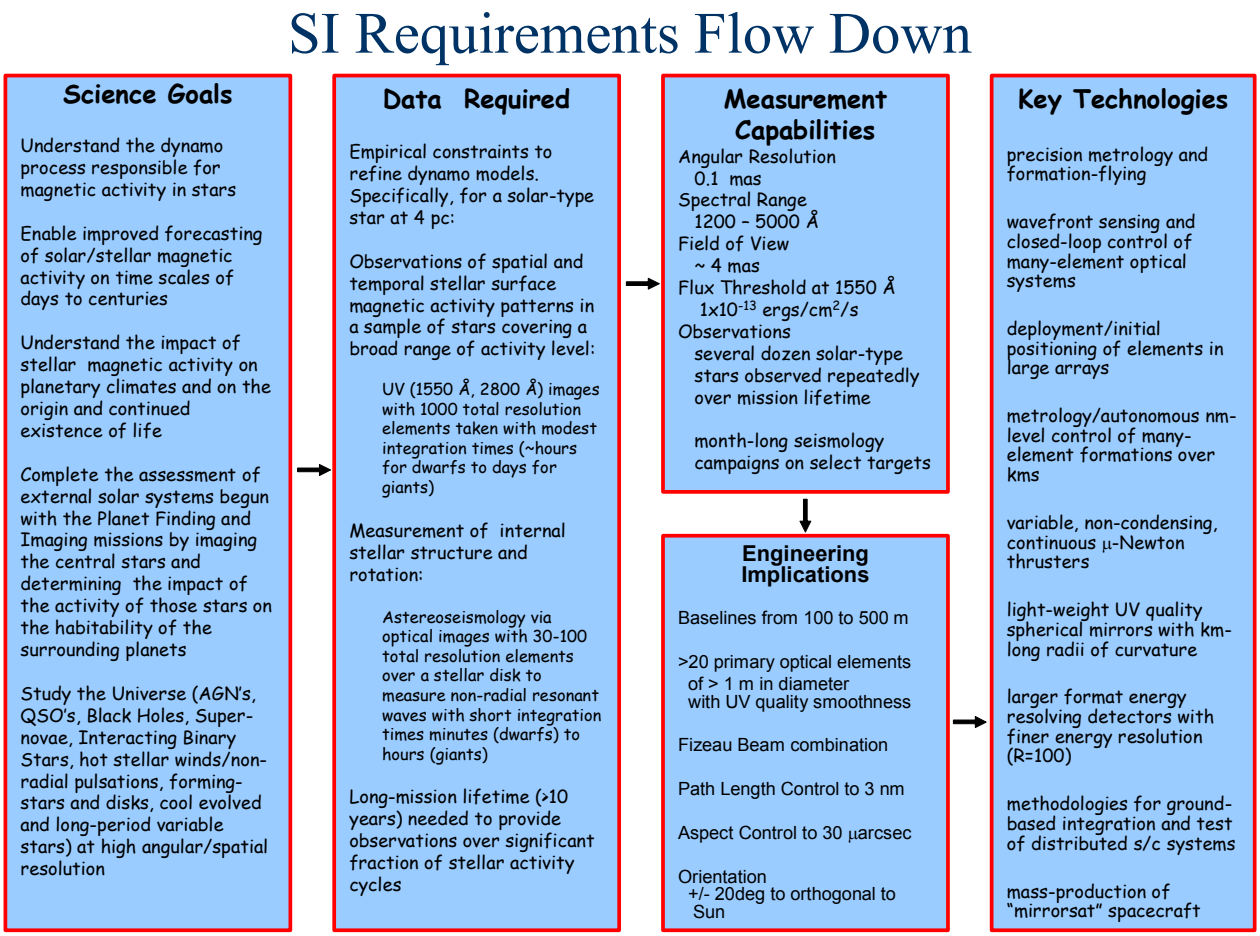


Figure 3.1: The flow down of science requirements through data requirements, measurement capabilities, and engineering implications to key technologies needed for SI.

3.1 Unique Requirements, Their Priority, and Sensitivity of Design to Each

The baseline full-mission concept for SI was studied by the GSFC Integrated Mission Design Center (IMDC) and Instrument Synthesis and Analysis Lab (ISAL) to explore the technical feasibility of the mission and to identify the technology developments needed to enable the mission in the 2024 timeframe. The IMDC examined the strawman concept described in the preceding chapter for a space-based Fizeau interferometer, located in a Lissajous orbit around the sun-earth L2 point, with a maximum baseline adjustable from 100 m to 1000 m and focal length that simultaneously adjusts from 1 km to 10 km, with the typical configuration having a 500 m diameter and 5 km focal length. The 1 m primary mirror size was chosen to ensure that the primary stellar activity targets can be well observed with good signal/noise. Sizes up to 2 m may be considered in the future, depending on the breadth of science targets that SI is required to observe – e.g., some fainter extragalactic objects may need larger mirrors, but those will come at a cost to the packaging for launch, the number of launches needed, and total mission cost. The design considered included 30 mirrorsats formation flying with a beam-combining hub, where the satellites are controlled to mm radial precision and the mirror surfaces to 5 nm precision, rather than using optical delay lines inside the hub for fine tuning the optical path lengths. A variety of disciplines considered the implications of this general design, including power, guidance & navigation, flight dynamics, operations, communications, quality assurance, system engineering, etc. The ISAL concentrated its efforts on the design of the beam-combining hub, again from a multiple-discipline viewpoint, including accommodation of the IMDC results. The results of these IMDC and ISAL studies and of related work carried out throughout the course of the Vision Study by Team members are presented here as well as in Chapter 2.

Many spacecraft engineering requirements exist which are a natural consequence of the defined science goals of the SI mission. The following represent the most significant issues that arise when considering the science objectives of the mission:

- *Telescope pointing:* In order to center the disk of a star that is approximately 3 milliarcsec across, the spacecraft configuration needs to point to the center of the disk within a fraction of a pixel (less than 40 μarcsec pointing); this places the further constraint of the jitter to be no more than 20 μarcsec to avoid possible smearing of the image
- *Formation flying:* Refer to Sections 3.1.1 and 3.3.3 for details
- *Hub focal plane / mirrorsat mirror position*: All mirrors must be kept in phase while in science mode. This requires the following control and knowledge:
    - Mirrorsat piston position (relative to virtual parabola) controlled to < 1 mm
    - Mirror piston position controlled to < 5 nm via closed-loop-controlled piezoelectric mounts
    - Lateral position knowledge to < 10 cm
    - Tip / tilt < 4 milliarcsec
- *Mission Lifetime:* The 5-10^{+} year mission duration raisings several concerns in several areas, including the power system (batteries), long term reliability of components, total propellant needs, and level of redundancy at the component and/or spacecraft level

Secondary items include:

- *Target exposure time:* Observations of targets must occur on a small enough timescale (~ 4 to 6

hours) so that the star's rotation, intrinsic variations, and proper motion do not smear the image
- *Spacecraft Pointing:* It is crucial to keep the mirrors and detector in the shade with a modest size sunshade. Therefore the spacecraft must point to within +/- 20 degrees of the perpendicular to the sunline; it is also a requirement to have continuous full sun on the solar arrays
- Lightweight, UV quality mirrors

As part of the end-to-end mission design approach, the IMDC examined several subsystems. Each subsystem was analyzed according to the given SI concept design. In contrast to the major issues discussed above, the following are considered more moderate challenges that should be readily addressable, although requiring significant work and investment in the desired frame:

- *Launch Requirements:* The launch requirements can be handled with current technology. There are several options that exist for placing all of the component parts of SI in orbit about the Sun-Earth L2 point. If the selected design includes a single Hub and no Reference Craft (an optional metrology spacecraft), then the options (defined in IMDC studies prior to the Vision Mission runs) are: 3 Delta III launches, 1 Atlas V launch, or 2 Delta (III/IV) launches. If a Reference Craft is included in the selected design, then the options (defined in the Vision Mission IMDC & ISAL runs) are: a single Delta IV launch using a 5mx19.1m dual launch fairing or a dual launch using two Delta IV's, one with a 5mx14.3m fairing and one with a 4mx11.7m fairing. The single Delta IV launch is preferred for a design which includes a single Hub spacecraft plus a Reference Craft. If two Hubs plus a Reference Craft are to be launched initially then the dual Delta IV launches are needed.
- *Power Requirements:* Although power requirements can be handled by existing solar cells, they must, on the mirrorsats at least, be body-mounted to avoid unacceptable impact on precision formation-flying and station-keeping. Battery life and storage are also a concern for a mission which is intended to last for perhaps a decade.
- *Propellant Requirements:* Propellant requirements at L2 are modest in the current design (requirements could go up if faster slews are needed): Field Emission Electric Propulsion (FEEP) thrusters should be capable in the 2024 timeframe of generating continuous, variable μ-Newton thrust for required 10 year lifetime on approximately 3.0 kg (per mirrorsat) and 643 kg (per hub) of solid fuel. The most recent IMDC study suggests using Hall Thrusters on the (larger, more massive) Hub spacecraft to obtain the higher thrusts needed to move its mass around (relative to the less massive mirrorsats) for the hub slews, and FEEP's for Hub fine thrust. The fuel estimates above include both FEEP and Hall Thrusters.
- *Operations Concept:* The operations concept is straightforward and assumes autonomous control of array station-keeping, reconfiguration, and slewing, with ground interaction only for command uploads and anomaly resolution.
- *Thermal Design:* The main concern of the thermal engineers is keeping the mirrors isothermal and protected from the Sun. A protective coating can be added to reduce the chance of damage in case of accidental sun exposure.
- *Communications Requirements:* Communications requirements are not excessive. In normal operations the mirrorsats talk to the hub and each other, and the hub talks to earth. In contingency operations: mirrorsats can be commanded directly from earth. A desired enhancement in this area would be a central communications hub at L2 for all missions flying in that locale. Sample data rates are provided in Chapter 2 (section 2.1.1.5).

Precision metrology and formation-flying are identified as the tallest poles among numerous technical challenges. The long mission lifetime requirement was the next biggest concern among the designers: the hub will have redundant components, but it may very well be necessary to seriously consider building a backup hub for launch on-need or original deployment nonetheless, and additional backup mirrorsats will likely need to be flown so they can be put into the operating array as the original set suffers expected failures (the mirrorsats were designed as inexpensive, low-redundancy, mass-produced craft in these studies). Other technical developments of major importance and/or difficulty include the development of aspect control to 10's of μarcsecs, wavefront sensing and control of large, sparse aperture systems, the functional and performance verification of the entire (distributed) system prior to launch, and lightweight, UV-quality mirrors. The major enabling technologies needed for SI are given in Table 3.1

Table 3.1*: Summary of the most important enabling technologies needing further study and development for Stellar Imager.*

- ***formation-flying of ~ 30 spacecraft***
  - *deployment and initial positioning of elements in large formations*
  - *real-time correction and control of formation elements*
    - *staged-control system (km → cm → nm)*
  - *aspect control to 10's of micro-arcsec*
  - *positioning mirror surfaces to 5 nm*
  - *variable, non-condensing, continuous micro-Newton thrusters*
- ***precision metrology over multi-km baselines***
  - *2nm if used alone for pathlength control (no wavefront sensing)*
  - *0.5 microns if hand-off to wavefront sensing & control for nm-level positioning*
  - *multiple modes to cover wide dynamic range*
- ***wavefront sensing and real-time, autonomous analysis***
- ***methodologies for ground-based validation of distributed systems***
- ***additional challenges***
  - *mass-production of "mirrorsat" spacecraft: cost-effective, high-volume fabrication, integration, & test*
  - *long mission lifetime requirement*
  - *light-weight UV quality mirrors with km-long radii of curvature (perhaps using deformable flats)*
  - *larger format (6 K x 6 K) energy resolving detectors with finer energy resolution (R=100)*

We discuss each of the most significant technology areas in detail in the follow sections.

3.1.1 Precision Formation Flying

In an engineering sense, formation flying is about understanding, harnessing, and exploiting the dynamics of relative motion. To the science community, formation flying is the collective use of multiple space vehicles to act virtually as a single large sensor. For the most part, formation flying for the Stellar Imager mission lies in the overlapping region of these engineering and science definitions as shown in Figure 3.2. Formation flying missions are those missions in which the relative separations and/or relative orientations of vehicles must be controlled and not simply measured. Precision formation flying is the subset of formation flying where the relative states of the component spacecraft are controlled continuously or quasi-continuously (i.e., not just once or twice per orbit). Such a task requires direct communication among the spacecraft. Very specifically, the "precision" qualifier in precision formation flying carves out a somewhat well-defined niche in the formation flying field with the following characteristics:

- Continuous and robust, possibly high bandwidth intersatellite communications
- On-board relative navigation/bearing at high data rate with high-precision through the communication links
- Continuous formation control at high bandwidth and high-precision through the communication links
- Highly-optimized formation/mission design and analysis
- Integrated hardware-in-the-loop, high-fidelity simulations
- Autonomous and robust closed-loop on-board control during science gathering

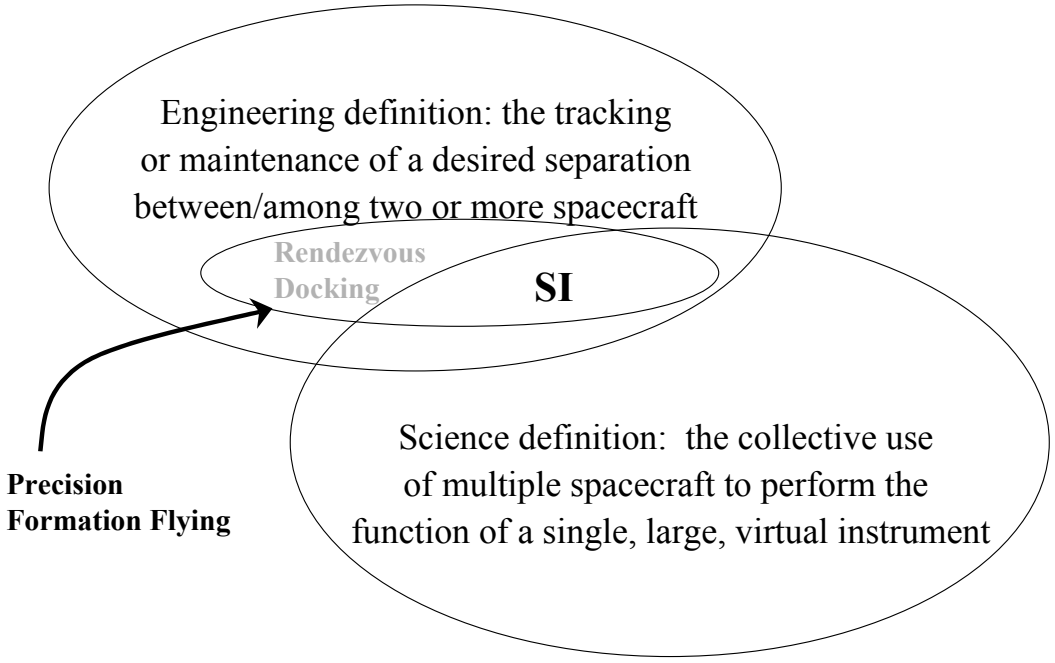


Figure 3.2: Formation flying science versus engineering

3.1.1.1 The Elements of Formation Flying

There are four elements that are unique to the formation flying problem: formation design, relative navigation/metrology, formation control, and intersatellite communication. These elements and specific considerations in the SI mission are described as follows.

3.1.1.2 Formation Design

Formation design is the guidance problem for the desired geometry as a function of time as dictated by the science needs of the mission. The dynamics of relative motion as applied to formation design continues to be a major research area. The problem is not only the specification of where the spacecraft needs to be as a function of time but also how to do this in the most fuel-efficient manner, since differential effects between spacecraft and small errors in initialization can be very costly in fuel. Concisely, formation design is the science and art of designing the desired relative motion of the vehicles to best meet science requirements without prohibitive fuel consumption.

For the Stellar Imager mission, formation design involves (1) modeling of the relative motion dynamics between multiple vehicles near the Earth-Sun L2 point, determining the appropriate level of fidelity equations to use so as not compromise the accuracy at the required performance levels, (2) determining natural relative motion between vehicles as a function of formation geometry to understand the frequency requirements in the controls, (3) assessing the sensitivity of fuel consumption on the orientation of the formation based on the geometric configuration, and finally (4) the definition of the formation geometry that meets scientific constraints while minimizing fuel usage requirements.

3.1.1.3 Relative Navigation and Metrology

Relative navigation is the estimation of relative positions based on the measurements between adjacent spacecraft. It includes the sensors, metrology systems, and wavefront error sensing systems and algorithms needed to determine relative position and attitude either to process science data or to feedback for control. The overall measurements in this area represent combined performance of loose ranging systems and precision metrology so as to meet overall science requirements on knowledge of relative positions. This can be driven by science requirements or possibly indirectly through other engineering requirements. In particular, formation control requirements may drive the relative navigation requirements more stringently than direct science requirements will. Ground performance simulation must be performed with the sensor in the loop with a high fidelity channel simulator. Relative navigation is constrained by technology, but component level demonstrations of ranging systems have shown more than sufficient performance, but not quite in relevant environments to this date.

Specifically, the relative navigation system will provide the intravehicle measurements taking the cluster of vehicles from a “lost-in-space” situation at insertion to finest level of “instrument-level” phasing of mirrors for science mode. Since it is unlikely to find a single system to reliably cover the entire dynamic range of the sensing problem (hundreds of meters at insertion, down to nanometers during science operations), the system will have several modes with carefully designed handoff and/or overlap such that the resolution of coarse mode is within the dynamic range of the next finer mode. More details on relative navigation are provided in the sections on wavefront sensing, pointing and metrology, and staged-control. Although the baseline design includes a multi-stage control system including wavefront sensing, an interesting future trade study will be on whether wavefront sensing will be necessary or if the finest level of measurement can be performed by an independent metrology system.

3.1.1.4 Formation Control

Formation control is responsible for rejecting disturbances, maintaining formation stability and reconfiguring or repointing the formation. Specifically, this involves the application of forces and moments required to regulate and/or track desired formation geometry and orientation. Formation control includes the actuators, other components and algorithms, together with autonomy and higher-level command and control. This is heavily dependent on new technology. This is truly a system-level problem, depending critically on performance of the intersatellite communications, the relative navigation, and the formation design. Formation control is the principle driver for concepts such as 6-Degrees-of-Freedom spacecraft control and closed-loop orbit control. Analogous to the relative navigation problem, multiple modes will likely be required, with overlapping dynamic range and resolution between coarse and fine modes. Some key trades will be between control of spacecraft, movement of mirrors or other devices on the spacecraft, and control of pathlength of science light, such as by path delay lines. The level of precision in control will necessitate control algorithms that are very insensitive to small variations and uncertainties in the dynamics as well as simulation systems that can verify performance at microscopic levels even over large distances.

3.1.1.5 Intersatellite Communications

The intersatellite communication system is the data bus of the formation. More so than in other data buses robustness and continuity are essential. The primary areas of development are mass, power and cost reduction, integrating communications and ranging functions. This area includes hardware (transceivers or transponders), algorithms and network architectures, and software. Substantial work is still needed in developing requirements for communication bandwidth and time synchronization and transfer for precision formation control performance. Since the formation control laws are implemented through this system, a lack of integrity in the system will be a showstopper. For the most part, the technology is available today, however, there is much analysis to be done to ensure that the communication is robust enough to support very tight control loops enabling sub-mm control performance for long-duration science operations. Significant analytical efforts are required to assess the effects of variable communication delays within a formation and the results will drive the technology development process to support SI and multiple other similar mission concepts.

Using such general expressions, we can divide and conquer the critical subsystem-level challenges unique to the formation flying problem while, at the same time, we acknowledge that there may be multi-stage sensing, actuation, and communication which cannot be easily divided among their individual stages. For example, a key element of precision metrology will be the handoff between a coarse and fine measurement stage. The challenges of this handoff will not be adequately addressed if the metrology were not an element of an overall relative navigation process.

3.1.2 Metrology

Stellar Imager can make images with pixels as fine as $\lambda/D = 30$ μarcsec, for λ=155 nm and D=1000 m. The system requires control of the Optical Path Difference (OPD) through any pair of apertures to $\lambda/10$ (equivalent at 1500 Å to 3 μas) for up to 1000 sec, the time it takes to resolve the 2π ambiguity in the phase of the complex visibility. Therefore the metrology and pointing system must provide position information for each aperture for $\lambda/30$ for each mirror surface, equivalent to 1 μas. While the ambiguity can in many cases be resolved in much less time, the long-time portion of the requirement (the 100-1000 sec regime) is the easiest to meet, so the requirement has been set conservatively long to ensure that it is possible to work with faint targets.

The metrology and pointing technology for SI is an enabling technology for several future NASA missions, for example Black Hole Imager and Terrestrial Planet Finder.

This section describes two approaches for meeting these requirements:

Case (A) All targets can be observed, including those too faint or diffuse to serve themselves as a pointing reference. A reference platform uses interferometers to sense alignment with primary and secondary guide stars.

Case (B) Only targets sufficiently bright and pointlike to serve as a pointing reference can be observed. For this case the metrology can be from a satellite at the center of curvature (if the primary array is spherical), or point to point metrology between pairs of mirrorsats and between mirrorsats and the detector, with no extra satellite (needed when using parabolic primary array).

In both cases, laser gauges determine mirror positions with respect to the pointing reference, and detector position is monitored by laser gauges or observations of an artificial star. Table 3.2 summarizes some metrology and pointing trades. There is a possibility for blending cases A and B, if observations are made under case B for bright and compact targets. For faint or diffuse targets, several mirrorsats would be used to form a reference platform, and in those observations would not process science light. The impact on science throughput would need to be evaluated.

Table 3.2: Metrology and Pointing Trades

Pointing reference	Target star (Case B)	Target must have V<~7.5 and ang. dia. < 10 milliarcsec
	Guide star (Case A)	No restriction on target star.
Metrology architecture	Pointing platform (Case A)	Reference platform required, near mirrorsats. Number of required laser gauges = 2x number of mirrorsats.
	Center of curvature (Case B)	Metrology spacecraft required, at center of curvature. Distance from mirrorsats = 2× focal length = 2-20 km. A single laser gauge (of a different design) measures positions of all mirrorsats. (can only be done with spherical primary)
	Image-based (Case B)	Processing of images yields information driving pointing. If the science derives only closure phase, not visibility amplitude, from data, it may be possible to relax the pointing requirement as much as 100-fold. Target requirements and impact on science data have not yet been determined. It may even be possible to derive sufficient information from the science data to control individual mirrorsat positions. For image-based pointing and metrology, the time scale for a disturbance to violate the OPD requirement must be longer than the time to acquire enough photons from the source to determine the phase of the complex visibility.

3.1.2.1. Metrology and Pointing Requirements

The metrology and pointing systems provide OPD information to the control system to allow it to keep the image sharp. In this section we discuss Case (A). It is necessary to keep the Strehl ratio near the diffraction limit (for the dilute aperture, possibly with pupil densification), and to keep pointing stable to a small fraction of the instrument resolution to satisfy the science requirements. Specifically, Pointing and OPD variations must meet the requirements set forth above.

Three time scales are important to this requirement. First, t_1 is the time it takes for disturbances to violate the requirement. The control system should have a step response time $\leq t_1/10$, which requires that it have a unity gain frequency $>\sim 10/(2\pi t_1)$, and the metrology system must provide updates at a rate$>\sim 100/(2\pi t_1)$.

Disturbances to the spacecraft positions are likely to be dominated by solar radiation pressure. A 50 kg spacecraft with a face of area 2 m^2 and 100% specular reflectivity, oriented perpendicular to sunlight, accelerates at 0.3 $\mu m/sec^2$, which would violate the requirement in 0.2 sec. However, the spacecraft reflectivity and orientation vary very slowly with time and predictably with orientation, and the control system can maintain a model for reflectivity of each spacecraft. It can correct the mirror position either by extremely fine adjustments to spacecraft position, or by having actuators of limited range on the mirror. Only unmodeled changes in reflectivity provide a disturbance which the control system must counteract dynamically. Therefore, we take t_1 = 20 sec, and the metrology and pointing system must provide OPD updates once per sec.

Second, t_2 is the required integration time for a frame or binning of the data. This might be the integration time of light on a CCD for signal to overcome read noise. If a photon-counting detector is used, it might be the time required for the 1-σ error of the phase of the complex visibility on a baseline to diminish to $\pi/3$, so that a 2π error is improbable. The metrology and pointing system and the control system must be stable to at least the above accuracy for a time of at least t_2. Above, we assume (conservatively, for faint targets) that t_2 is 1000 sec.

Third, t_3 is the time required for the entire integration. While pointing may be permitted to drift for times longer than t_2, i.e., the control system may drift, the pointing direction must be known to the required accuracy, to permit *a posteriori* image reconstruction. Thus, the metrology and pointing system must also remain stable for a time of at least t_3. The control system needs to hold OPD and pointing constant over t_2 . While it does not need to hold it constant over t_3, doing so is probably the easiest part of its job, and would reduce the work to be done in data reduction. Therefore, the control system will probably hold OPD and pointing constant over t_3 as well.

3.1.2.2. Metrology and Pointing Approach

The pointing system establishes a reference platform of the required stability, using observations of guide stars or an inertial reference. The metrology system measures the positions of the apertures and detector with respect to that reference, and estimates OPD's from the target through each subaperture.

A schematic diagram of the optics is shown in Fig. 3.3. Each mirrorsat containing a subaperture A_i brings light from the science target to a focus at the detector, D. The path to be measured and held constant, for subaperture i, is from a reference wavefront W to D *via* A_i. W is perpendicular to the line to the center of the science target, and passes through D.

3.1.2.3. Pointing: Guide Star Interferometer

The reference spacecraft, R, is located a distance h above the vertex of the primary, O. Let z be the science instrument's optical axis, and x and y complete an orthogonal right-handed coordinate system. The reference spacecraft has a pair of interferometers that observe a guide star, with baseline and aperture set by the required accuracy and pointing update interval, as shown in Table 3.3, which shows baseline and aperture combinations meeting various pointing precision and update rate requirements, for a magnitude 7.5 guide star, which can be found within 2° of the science target. The longest baselines would require five free flying spacecraft: four apertures, plus a combining hub and are thus undesirable since they add complexity to the reference spacecraft. More desirable are the combination of parameters that allow the shorter (< 10 m) baselines. Assume that the guide star is of solar type, with V=7.5, and that the detection bandwidth is 4000 Å, centered at 5500 Å. The rate of detection of photons, per aperture area, is $N_o = 4\times10^6$/sec/m^2, assuming 10% overall efficiency (obscuration, reflection losses, and detector efficiency). An interferometer of baseline B detecting N photons at wavelength λ has a precision (Reasenberg 1988)

$$\sigma = \frac{1}{2\pi}\frac{1}{\sqrt{N}}\frac{\lambda}{B} \tag{3.1}$$

With apertures of diameter

$$d = \frac{B}{\alpha} \tag{3.2}$$

observing for a time τ, a precision σ is obtained when

$$B = \sqrt{\alpha\frac{1}{\pi\sqrt{2\pi}}\frac{\lambda}{\sigma}\frac{1}{\sqrt{N_o\tau}}} \tag{3.3}$$

Setting α=6 (corresponding to a compact reference interferometer), we obtain the values in Table 3.3 (except for the first line, where we have limited the aperture diameter to 1 m). The highest accuracy requirements imply baselines that would be achieved with free flying spacecraft. Herein, we refer to the spacecraft used for pointing as a "pointing platform," even if there are actually several spacecraft operating in concert.

Table 3.3: Baseline and Aperture Combinations vs. Requirements

σ(θ) μas	Update interval sec	Baseline m	Aperture m
0.1	1	71	1
0.1	1	21	3.4
0.1	100	6.5	1.1
1	1	6.5	1.1
1	100	2.1	0.3
10	1	2.1	0.3
10	100	0.7	0.1

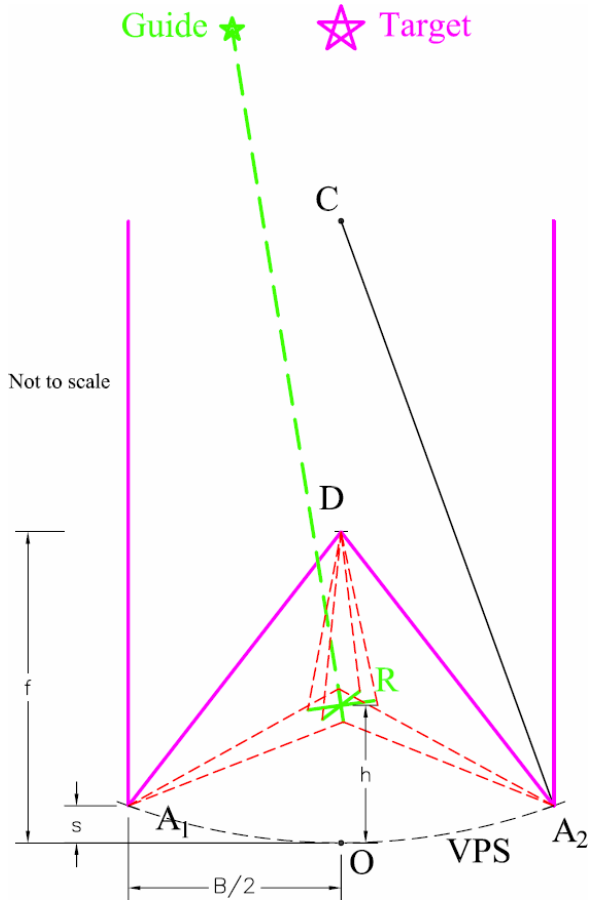


Figure 3.3 - Schematic diagram of Stellar Imager metrology and pointing system.
VPS = virtual primary surface. f = focal length = 1-10 km. B = baseline = 100-1000 m. s = sagitta = 1.25-12.5 m. h = height of reference platform above VPS, nominally 50-100 m.

A guide star of V<7.5 can almost always be found within 2° of the science target, even at the Galactic pole. The guide star interferometer's measurements are fed to the reference platform's attitude control system, which keeps it aligned with the direction to the guide star to within <1 arcsec.

Each guide star interferometer will have a "pseudobaseline," formed by a pair of optical fiducial blocks, one near the entrance to each telescope. The pseudobaseline is approximately parallel to the guide star interferometer baseline, with the offset measured to high precision. The pseudobaseline's defining fiducial blocks provide optical endpoints for accurate laser gauge determinations of the direction to the science apertures and detector. The guide interferometer OPD is stabilized with respect to the pseudobaseline using laser gauge measurements, and adjustments of the primary beamsplitter position. The offset of the star from the pseudobaseline is <1 arcsec, and is measured to within 1 μas. The offset is used to correct the laser gauge measurements of aperture and detector positions before they are sent to the control system.

Since the fiducial block can be compact and enclosed in thermal shielding, with negligible view factors outside the instrument, it is thermally very stable. With the optical elements made of ultra-low expansion material, thermal changes of distances among them can be held to 1 pm (10^{-12} m) or below.

Since the guide star is up to 2° away from the science target, rotation of the reference platform about z must be known with an uncertainty only ~30 times as great as that about x and y. This requires a third guide star interferometer. As can be seen from Eqs. 3.2 and 3.3, if an equally bright guide star is used, the product of B and d (guide star baseline and subaperture diameter) for the interferometer that senses rotation about z need only be 1/30 that for the other two guide star interferometers, so the latter interferometer may be comparatively small and inexpensive.

3.1.2.4. Metrology of Mirror Positions

To measure the position of a primary mirror element, laser gauges measure the distance to it from the fiducial blocks at each end of the z-direction pseudobaseline. For the detector, the measurements are from both x- and y-direction pseudobaselines. The distance measurements in each case are used to estimate an angle. It is also necessary to measure the distance to both mirrors and detector, which can be done with the same measurements used for angle.

The directions in which the laser gauges must aim their beams to follow the mirrors and detector will vary for different target star-guide star offsets, and this requires that the laser gauge beams interrogate different parts of their endpoint optics. This is termed beamwalk, and it causes measurement error due to manufacturing errors in the surfaces.

The required tolerance of the angle measurement for the mirrors is the permissible position error divided by the distance from metrology platform to mirror, ~2.5 nm/500 m. The tolerance for the detector is a fraction of the permissible pointing error. Both tolerances are of the order of 1 μas. On a 2 m baseline, the entire error corresponds to a distance of 10 pm; however because the error of several laser gauges contributes, the tolerance for an individual laser gauge needs to be about 1 pm.

3.1.2.5. Detector Position

There are several ways to measure the detector position with respect to the reference spacecraft. For example, the detector could contain an artificial star that was observed by an interferometer on the guide star platform. Or, laser gauges could be used.

While intercepting the entire metrology beam at each endpoint might be required at the highest accuracy levels (well under 1 picometer), it is probably not necessary for the pm and greater tolerances of SI. Intercepting the entire beam would require endpoint optics of ~20 cm diameter (for a 10 km path at 1550 nm metrology wavelength), but in fact, they can be considerably smaller.

3.1.2.6. Guide Star Acquisition

For the reference platform to acquire a guide star, the ordinary star tracker provides initial attitude information, allowing the platform to slew to within ~1 arcsec of the guide star direction. This may be close enough to allow the guide star interferometers to acquire fringes. If not, a "super star tracker" is used to refine the reference platform attitude sufficiently for interferometric acquisition. The super star tracker would be similar to the guide star telescope used in GP-B, which observes a single star, bore-sighted, with ~1 mas precision.

Once the guide star interferometer has acquired fringes, its attitude is servo-controlled so that its pseudobaselines remain within about 1 arcsec of the guide star. The offset is measured with a dispersed fringe ("channeled spectrum") approach, to an accuracy of 1 μas. The guide stars will be on the axes of the guide star interferometers' beam-compressing telescopes to within about 1 arcsec. This

will reduce cost by eliminating delay lines and articulating siderostats. Also, the telescopes require only a 1 arcsec field of view, and aberrations will be negligible.

3.1.2.7. Alternative Approaches

There are several alternative pointing schemes at an early stage of research. With all pointing schemes, metrology is needed to measure mirror and detector positions.

Science Star as Guide Star

If it is acceptable to require the science target to be moderately bright and compact (V<~7.5 and diameter <~10 milliarcsec), then the guide star interferometers may observe the target itself. In this case there is no need to determine the angle of the reference platform about the direction to the star, and the reasons for employing a separate metrology platform (discussed above) would not apply. There may also be an option to obtain pointing information from the science target via the main apertures instead of via separate guide star interferometers. This might be most appropriate for an SI Pathfinder or for the early Phases of the full mission, in a scenario where the brightest, most compact targets are observed early-on and upgrades to the metrology system are made later to enable observations of brighter targets in later Phases of the mission (this would require advance planning in the design of the hardware flown originally on the s/c, to accommodate this later upgrade.).

Superfluid Gyro

Gyros based on superfluid ^{4}He (at temperatures below the lambda point, 2.2°K) have been constructed, with angular sensitivity of 3 arcsec/s/sqrt{Hz} (Schwab 1997). Work is going on at GSFC and the University of Maryland on a version employing microfabrication on a Si wafer and a Single Electron Transistor (SET) intended to improve this to 4 μas/s/sqrt{Hz}.

Kilometric Optical Gyro

In the presence of rotation, the Sagnac effect introduces a phase shift between beams traveling in opposite directions around a closed path. This is the basis for various rotation sensors. A ring laser gyroscope of area 1 m^2 has been operated at the shot noise limit (Stedman 1997).

For Stellar Imager, a km scale gyro would be required. This instrument will be complicated by several factors. The work cited above was performed in a cavity bored out of a solid block of Zerodur. Mirrors with reflectivity 0.999999 were used, resulting in a finesse approaching 10^6. The Stellar Imager gyro would be based on mirrors mounted on free-flying spacecraft whose orientation would be under servo control. Residual vibrations will tend to couple power into higher order cavity modes, limiting the finesse. Also, keeping mirrors in space sufficiently clean to maintain this reflectivity would be difficult.

Telescope

A telescope could in principle observe guide stars and a beacon transmitted by the detector spacecraft. However, the sensitivity of a telescope is inferior to that of an interferometer. In order to compare instruments requiring comparable structure, we set the telescope focal length equal to the

interferometer baseline. The ratio of the uncertainty of the astrometric telescope to that of the interferometer is (Reasenberg 1988)

$$\frac{\sigma_T}{\sigma_I} = \frac{3\pi^2 F}{8\sqrt{2}} \tag{3.4}$$

where F is the focal ratio of the telescope's primary. The cited work has a factor of 2 that pertains to a Ronchi ruling, which is omitted here, since this telescope may be bore-sighted and employ a roof prism. Also, the uncertainty for an interferometer has been increased by a factor $\sqrt{2}$ because with the telescope and an appropriate arrangement of roof prisms (or a CCD), all photons can contribute information in both orthogonal directions, but this is not so for the interferometer. For an f/15 astrometric telescope, the ratio in Eq. 3.4 is 40. Since the cost of a spacecraft is closely related to its size and weight, this predicts that an interferometer has a significant cost advantage over a telescope.

Note also that, by comparison with the interferometer example above, the telescope would need to observe guide stars *at least* as bright as V=7.5, and therefore must look as far as 2° from the science star to find a guide star. Since the detector is near the science star, as seen from the reference platform, the telescope must have a field of view of 2° in order to observe both detector and guide star.

Another advantage of an interferometer over a telescope is in the metering of the instrument in the presence of thermal changes. For the same astrometric precision, the interferometer's apertures will be smaller than the telescope aperture. Thus, baffling will be more effective, and monitoring the position of one or a limited number of points on the mirror yields a more faithful indication of the position of the whole surface. The GP-B telescope does achieve stability of ~0.1 mas/yr, but it does so by rolling about the line of sight every 1-3 min., and by operating at liquid helium temperature, which reduces thermal expansion.

Mechanical Gyroscope

A mechanical gyroscope, even one as refined as that developed for GP-B, is not a viable option, because its readout noise is too high. The GP-B gyro uses an electrostatic suspension, and drag-free control of the satellite to reduce spurious torques due to the suspension. Operation at liquid helium temperature conveys several advantages: reduced thermal expansion and a readout based on the London moment and Superconducting Quantum Interference Devices (SQUID's). This readout minimizes spurious torques. However, the short-term noise of these instruments makes them inadequate for the Stellar Imager task. The limit to gyro sensitivity imposed by SQUID noise is 0.1 arcsec $Hz^{-1/2}$, presumably for signals at the spacecraft roll rate. This is consistent with the statement that the gyro sensitivity is limited by readout noise to 0.1 mas/yr after 14 months of integration (M. Ratner, priv. comm. 2004). To point Stellar Imager to 1 μas in 1 sec requires 10^5 times less noise for a signal frequency of 1 Hz. The GP-B gyro falls far short of meeting the Stellar Imager requirement.

3.1.2.8 Mirror Position: Metrology from the Center of Curvature

Another concept for measuring the positions of A_i is to transmit a spherical wavefront from an additional spacecraft placed at C (Fig. 3.3). The portion of the beam falling on each aperture returns to a focus at C, and it interferes with a reference beam. An image of the primary is formed on a detector array. Each aperture corresponds to a distinct area of the image. The several fringe patterns are analyzed separately to obtain the tilt and piston of each aperture. (Higher order distortions could be

monitored as well.) Absolute distance can be measured by shifting the wavelength.

This scheme measures very nearly the set of quantities desired, and measures directly to the surfaces of the primary mirrors. The resulting simplicity may compensate for the expense of a separate spacecraft, and the much larger fuel consumption required for repointing, since the extra spacecraft is located 2f from the system center of mass, while the metrology reference platform would be near the CM. With this scheme for internal metrology, a pointing reference is still needed, and the best approach is likely to be the guide star interferometer outlined above. The position of the satellite at C would need to be monitored, similarly to the way the detector craft, D, is monitored.

3.1.2.9 Tracking Frequency Gauge

The Tracking Frequency laser distance Gauge (TFG) differs from the traditional high-precision (heterodyne) laser gauge (Noecker 1993, Reasenberg 1995, Phillips 2004, Phillips 2005). The TFG employs a single beam, not two distinct beams as with the heterodyne gauge. This frees the TFG of the nm-scale cyclic bias of heterodyne gauges and simplifies alignment.

The classic version of the TFG has demonstrated 2 picometer (pm) incremental distance accuracy in 1 min on a stabilized optical path, and 10 pm in 0.1 sec on a path with only passive stabilization. While the requirement for Stellar Imager is ~1 nm, other work (Reasenberg 2001, Reasenberg 2005) has a goal of 0.1 pm. (Incremental distance is the change of distance from an initial offset. Absolute distance includes the estimation of the offset.) Absolute distance determination has been subjected to a preliminary test, to an accuracy of 0.1 mm. The TFG's precision for both incremental and absolute distance measurements can be independent of the distance measured. A semiconductor-laser version, the SL-TFG, now undergoing laboratory testing, will achieve similar or improved incremental distance precision, and improved absolute distance. The selected laser is a distributed feedback (DFB) unit operating at 1550 nm. These are space-qualifiable, and all components will be in fiber-connected packages, reducing sensitivity to air turbulence and thermal expansion, and simplifying setup, testing, reconfiguration, and repair.

It is possible to build a simplified version of the SL-TFG, comprising only a tunable DFB laser, with no separate phase modulator. As with all laser gauges, the simplified version would require beam-launching optics, the interferometer whose length is to be measured, and a photodetector, plus electronics. Besides those required components, the simplified TFG would require only the DFB laser. The precision of the simplified version is expected to be of the order of 1 pm in 1 sec.

3.1.3 Staged-Control Systems

To achieve an ultra-high angular resolution, SI must maintain relative position and attitude stability of all collector spacecraft to nanometer and microarcsecond level over baselines of approximately 500 meters. Dynamic disturbances, introduced by variations in gravitation and by on-board mechanisms, perturb optic stabilities and degrade the optical performance of SI. It is unlikely that the magnitude of the perturbations will be below the specified level identified by the SI requirements. Therefore, closed-loop control systems must be used to reject the optic perturbations to within the specified tolerances. Meeting these stringent stability requirements in the presence of disturbances poses great challenges to the control system design.

Suppose a single actuator were available to control the spacecraft position with nanometer accuracy, and to maintain this accuracy in the face of high frequency perturbations. Then this single actuator would be sufficient to meet the stabilization requirements. Unfortunately, no single actuator

simultaneously possesses large control authority and high bandwidth. Therefore, a suite of actuators with overlapping strokes and bandwidths is employed; such a construction is termed a "staged control system." There are four main components associated with the design and analysis of the staged control system. The first step involves identifying and characterizing the disturbance sources and actuator capabilities. After collecting such information, possible staged control system designs can be formulated. Given a staged control system, the challenge is then to determine how to use the individual actuators intelligently in order to satisfy the system requirements.

In this section of the report, preliminary studies of the perturbation sources acting on SI are described in order to understand SI's disturbance environment. The frequency content and amplitude of the disturbances can be used to estimate actuator bandwidth and stroke requirements. Common actuator and sensor constraints are then presented to illustrate the capabilities of current actuation systems. By considering disturbances and actuator imperfections, candidate staged control systems required for SI are identified. As the SI design matures, detailed analyses that evaluate the performance of each design will be conducted in the future to down select the most appropriate control architecture for SI.

3.1.3.1 Characterization of the Disturbance Environment

The disturbances acting on the spacecraft can be classified into two categories: external and internal disturbances. At the L2 orbit, the external disturbances are dominated by solar radiation pressure and non-uniformity of gravitational field effects. Due to the offset between center of mass and center of pressure, any external force acting on the system will also introduce torques on the system. These types of disturbance forces and torques are slow varying. The internal disturbances, on the other hand, are induced from mechanisms onboard the spacecraft and can introduce much higher frequency disturbances.

External disturbances

The dynamical disturbance environment at L2 is much more benign than the near-Earth environment, which makes it especially attractive to high-accuracy missions like the SI. Given SI's very stringent attitude and translation control requirements, though, environmental disturbances may not be entirely neglected. Solar radiation pressure and differential gravitational acceleration affect observatory dynamics, necessitating continuous or near-continuous formation control, and driving the sizing of momentum unloading resources.

Solar radiation pressure arises from solar photons transferring linear momentum to a surface as they are absorbed or reflected. At L2, the solar radiation pressure is essentially constant, to within 1% of the near-Earth value of $P = 4.64E\text{-}6\ N/m^2$. Since the hub spacecraft is substantially larger than the mirrorsats (15 m^2 solar panels rather than 1 m^2), the hub spacecraft will experience an anti-sunward force relative to the remainder of the formation. Assuming an area difference $A = 14\ m^2$ and a momentum transfer coefficient of $\rho = 1.5$ (1.0 for total absorption, 2.0 for specular reflection), the relative force is simply

$$F = \rho PA = 1.5 \times 4.6E\text{-}6 \times 14 = 9.7E\text{-}5\ N$$

Note that specular reflection may also exert a force normal to the sun line. The simple calculation here is sufficient for preliminary actuator sizing.

Two bodies near L2, separated by a small distance, experience very slightly different gravitational

accelerations; the body closer to the Earth and Sun is attracted more strongly by them than the farther body is. In the vicinity of L2, this differential gravitational acceleration (a) is proportional to the radial separation (d) between the two bodies:

$$a = Kd$$

The constant of proportionality may be shown to be K = 3.5E-13 $(m/s^2)/m$. So two spacecraft separated by 1000 m would experience differential gravitational acceleration of 3.5E-10 m/s^2. To compensate for this acceleration, a 500-kg spacecraft would have to exert a steady 1.8E-7 N. This value is small compared to the relative solar pressure force computed above, but is still comparable to the minimum thrust level (0.5 μN) of the baseline micro-thrusters.

Internal disturbances

Any moving mechanisms onboard the spacecraft can introduce disturbances on the optical instruments due to structure interactions. Potential mechanisms that can generate large disturbances on SI include thrusters and reaction wheels. Other actuators such as voice coil or PZT for mirror articulations are expected to generate much less disturbance than thrusters and wheels. General descriptions of the dominant actuator noises are given below.

Thrusters

Traditional on-off cycled thrusters produce impulsive disturbances on the spacecraft which excite dynamics in all frequency ranges. More recently developed Colloid and Field Emission Electric Propulsion (FEEP) thrusters are promising candidates that can provide the continuous, gentle thrusting required to control the position of the spacecraft while avoiding the injection of impulsive disturbances on the system. However, these electric thrusters may still generate small, high frequency disturbances due to electronic noises. Representative FEEP thruster noise developed for the Laser Interferometer Space Antenna (LISA) mission is shown in Figure 3.4 (Tajmar et al., 2004). The thruster noise power spectral density (PSD) shown in the figure illustrates that the noise content can span a large frequency range.

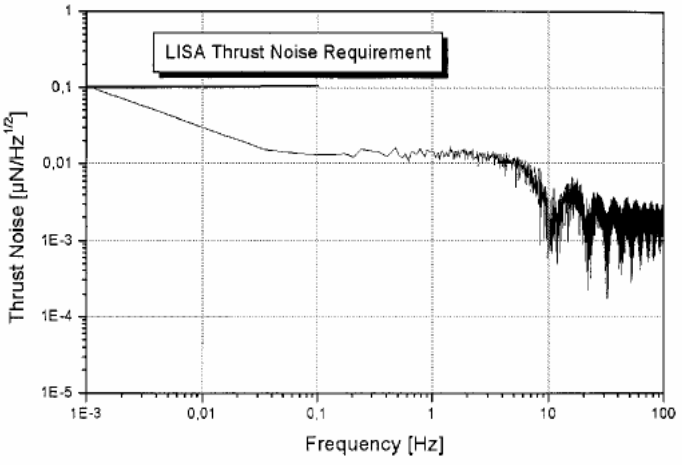


Figure 3.4 Representative thruster noise power spectrum

Reaction Wheels

Most spacecraft require reaction wheels for attitude control. SI mirrorsats will employ only thrusters for both orbit and attitude control, but the hub spacecraft may use wheels for large slews to reduce propellant consumption. However, reaction wheels are inherently noisy systems and can generate large disturbances on the spacecraft due to flywheel imbalances and bearing irregularities. A representative wheel disturbance waterfall plot is shown in Figure 3.5, where for each wheel speed, the disturbance spectrum is plotted. As the wheel speed increases, the disturbance frequency and amplitude increase proportionally to the wheel speed and speed squared, respectively. As a result, wheel disturbances can be detrimental at high frequencies due to large disturbance magnitudes at high wheel speeds.

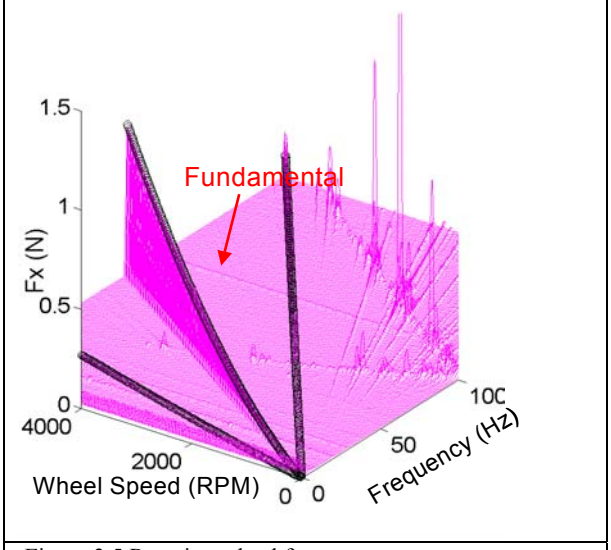


Figure 3.5 Reaction wheel force spectrum

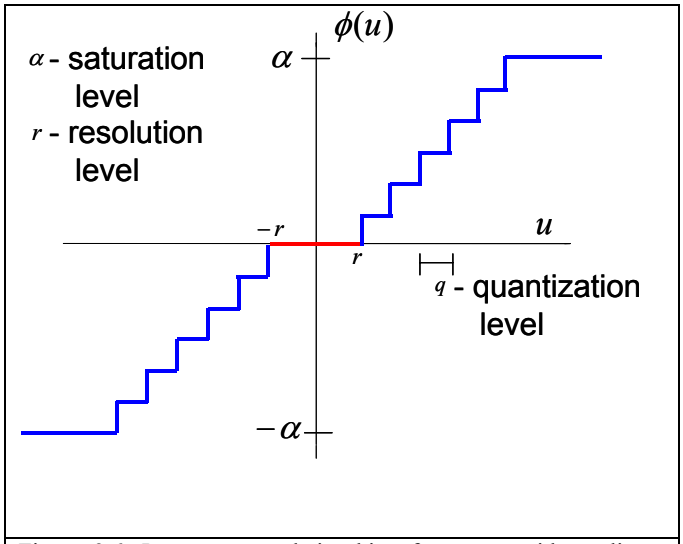


Figure 3.6: Input-output relationship of actuator with nonlinear constraints

3.1.3.2 Actuator constraints

All actuators have non-idealities which result in an imperfect match between the requested force/torque and the actual force/torque acting on the system. Three of the actuator non-idealities play a significant role in the staged control system design: saturation (stroke), resolution (quantization), and bandwidth (Liu, 2003). Saturation is one of the most common actuator constraints, defined as the maximum possible output that can be obtained from the actuation system. For example, there is an upper limit to the amount of force that a given thruster can apply, regardless of the magnitude of the commanded signal. Resolution is another common nonlinear constraint that defines the minimum physical output of the actuator. It can also be used to characterize the behavior of the minimum on-time of electrical or chemical thrusters. In digital systems, quantization effects are unavoidable and introduce round-off errors during the digital to analog (D/A) or analog to digital (A/D) conversion process. Furthermore, quantization also creates a deadband of one quantization level around zero input and thus contributes to the actuator resolution effect. A plot of the composite nonlinear effects of saturation and resolution/quantization is shown in Figure 3.6.

In addition to the nonlinear effects, most actuators also have bandwidth limitations. The bandwidth is defined as the frequency range where the actuator can follow the input without much error. The actuator bandwidth can be limited by the rise time of the amplifier current or the stiffness of the mechanical system. The bandwidth limitation of the actuator can be modeled as a low pass filter, and a representative first order low pass filter with corner frequency at 1 Hz is shown in Figure 3.7.

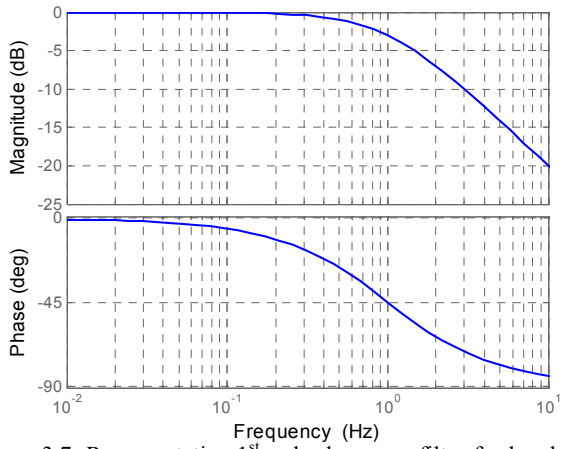


Figure 3.7: Representative 1st order low pass filter for bandwidth characterization

These linear and nonlinear actuator characteristics alter the desired operation executed on the system and thereby affecting the behavior of the closed-loop system. Since SI is a highly accurate interferometer, it is important to take the actuator non-idealities into account in order to predict realistic system performance. The relevant constraints of the SI actuators are summarized in Table 3.4.

Either an approximate or a range of values for saturation/stroke, resolution, and bandwidth of each potential SI actuator are given in the table to illustrate the current technology capabilities. The actuator information is obtained from various published product specs, and the full reference can be found in the Bibliography section. Notice that the operational range and resolution of the thruster are described in terms of force. Since the thrusters can push the spacecraft around with no theoretical position limitations, there is no stroke limit for this actuator type, only a force limit. Similarly the reaction wheel itself does not have an angular rotation limitation, but it does have a maximum torque limit. The minimum resolution of the wheel is the D/A quantization level assuming a 12-bit converter.

Table 3.4 SI actuator resolution, stroke, and bandwidth

Actuators	Resolution	Force/Torque or Stroke Limit	Bandwidth
Position			
Thruster (Tajmar et al)	~0.1 μN	0.5-100 μN	~ 10 Hz
VC actuator (Bei Tech)	< 0.1 μm	~1 μm to 1 cm	< 1 kHz
PZT actuator (Physik)	~0.1 nm	~10 μm	2-4 kHz
Pointing			
Thrusters*	~0.1 μN-m	0.5-100 μN-m	~ 10 Hz
Reaction Wheels (Tajmar et al. 2004)	50-70 μN-m	0.2-0.3 N-m	~ 10 Hz
Tip/tilt PZT actuator (Physik Instr.)	~0.05 mrad	2-6 mrad	2-4 kHz

*Assuming 1m moment arm between thruster location and center of mass

3.1.3.3 Staged control architectures

Each degree of freedom of the observatory (mirrorsats+hub) may be perturbed by undesirable disturbances from either the environment or from the actions of other mechanisms that interact with, or

support, operation of the interferometer. For example, there may be solar pressure or gravity gradients acting on the hub, and onboard mechanisms such as reaction wheels or thrusters may introduce unwanted vibrations. Additional actuator such as voice coil or PZT actuator may be required to suppress the effects of these disturbances to a specified level.

Traditionally, spacecraft control using thrusters and/or reaction wheels has been considered separately from instrument control using piezo actuators and steering mirrors. However, the rigid body and the instrument controllers are naturally coupled through the dynamics of the entire system. As a result, spacecraft and optical controllers must collaborate closely and can be designed simultaneously to satisfy the demanding stabilization requirements. A controller design that combines the spacecraft and instrument controllers is an example of a two-stage control system (see Figure 3.8).

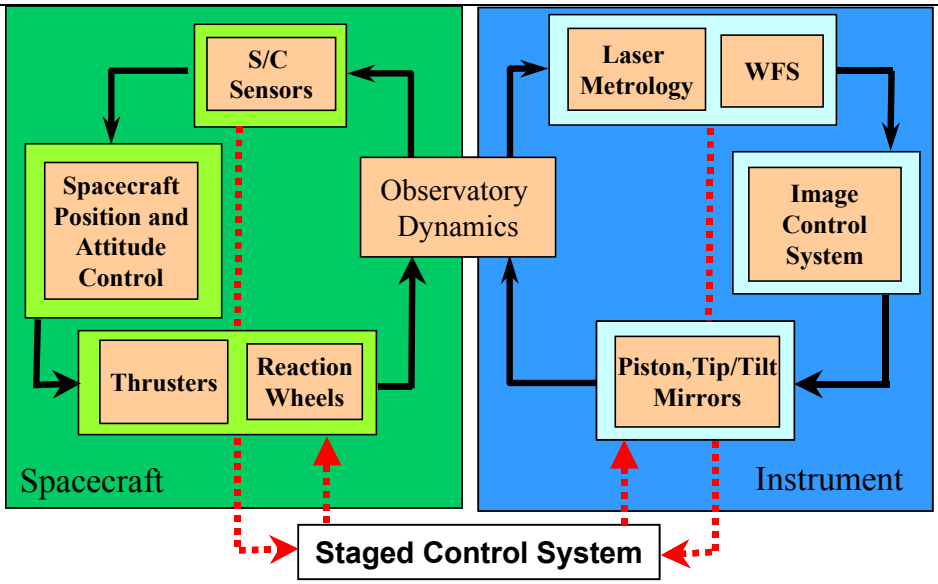


Figure 3.8 Two-stage control system design

Several control architectures are considered for the SI observing mode where laser metrology and wavefront sensing data are available:

	Case 1	Case 2	Case 3
Hub actuators	thrusters/wheels	Two-staged control	Two-staged control
Mirrorsat actuators	thrusters	thrusters	Two-staged control

Case 1. Controller uses thrusters and wheels on the hub and thrusters on the mirrorsats to accomplish station keeping and maintaining relative attitude among all spacecraft. The primary control task is to reject the slow varying environmental disturbances. The high frequency perturbations generated by thrusters and wheels may be ignored, so no staged control system is required to meet the stability requirements.

Case 2. Control scheme features a two-stage system (thrusters/wheels + mirror actuators) for the hub and uses just thrusters for the mirrorsats. In this case, the wheels on the hub may be generating large disturbances and would require the mirror controllers (piston, tip and tilt) to reject some of the perturbations. Comparing Figures 3.4 and 3.5, the expected thruster

disturbances are much smaller than that of the wheels. It is possible that the mirrorsats do not need an additional stage of control to reject thruster disturbances.

Case 3. Two-stage control systems are employed on both the hub and mirrorsats. If the previous two cases cannot achieve the desired performance, this strategy is implemented, utilizing overlapping stroke and bandwidth actuators, to reject as much disturbance as possible given the available actuators.

In order to select the most appropriate control architecture for SI, it is necessary to develop accurate dynamic models of the disturbances, actuator imperfections, and hub and mirrorsats. After integrating all the dynamic models, control algorithms are applied to the integrated model, and an estimate of the closed-loop performance (motion stability) of the system is obtained.

3.1.3.4 Control algorithm development and simulation

Given a staged control design, the challenge is to determine how to use the individual actuators appropriately in order to satisfy the system requirements. As shown in Table 3.5, actuators with larger maximum output tend to be lower bandwidth and coarser resolution, while the high bandwidth actuators tend to have smaller maximum output and finer resolution. The objective of the staging control algorithm is to specify how to piece together the individual actuators with different stroke and bandwidth limitations into the best possible control system for the observatory.

The synthesis of staged-control algorithms is still an active research field. Some progress has been made in the area of multi-stage optical delay line control (Liu 2003, Grogan et al. 1998, Hench et al. 2000) which can be extended to the two-staged control systems proposed for SI. However, these algorithms currently do not couple translation and attitude control schemes. A unified staged-control theory that properly integrates both translation and attitude control may achieve better performance than strategies that treat them separately.

Developments of the staged-control algorithms and high fidelity simulations of the SI system have begun at NASA Goddard. As the SI program matures, better structural characterizations of the hub and mirrorsat will become available. These structure models along with environmental, actuator, and optics models will form the integrated dynamic system of SI. The newly developed staged-control algorithms will be implemented to assess the closed-loop performance of the integrated system. The dynamic simulation will be used to demonstrate the end-to-end disturbance to optical performance of SI and select the most appropriate control architecture for SI.

3.1.4 Wavefront Sensing and Control

The wavefront sensing and control (WFC) system consists of the WFC camera (WFCCam), the WFC processor, a MEMS mirror array consisting of rigid body actuators, and rigid body actuation of each spacecraft, and possibly rigid body actuation of each mirrorlet on each spacecraft. The WFCCam, as described in section 2.2.3 collects diversity images and periodically sends them to the WFC processor. The WFC processors uses a variant of phase diversity that directly solves for phase differences between the baseline pairs. This phase diversity approach is a direct non-iterative solve for the phase differences which works for any non-redundant aperture configuration. This algorithm was conceived, designed and developed as part of this Stellar Imager study. Once the phase differences (OPDs) are known they are decomposed, by the WFC processor, into the control modes of the system and are also used in the phase correction algorithm on the science image (Figure 2.25 – Section 2.2.3).

The technologies required in the WFCCam are high quality low scatter dichroic beamsplitters, reflective flats and large format non-destructive read photon counting 16 bit detectors or higher. The WFC processor must be capable of running the MTF fitting algorithm, the phase diversity algorithm and the decomposition of the recovered wavefront into control modes of the MEMS mirror array and the spacecraft all within 10 seconds including data collection, bussing of the data from the detectors to processor and command generation and must run autonomously in closed-loop control. In addition as science images are collected and sent to memory within the science processor input is needed, from the WFC processor, in terms of the phase corrections output from the phase diversity algorithm to phase correct the science imagery in-situ. The actuation subsystem consists of a MEMS mirror array consisting of 3 degree of freedom articulation of piston, tip and tilt on 30 separate flat mirrors mounted on a curved surface. In addition each of the 30 mirrorsats comprising the primary mirror array need rigid body motions in terms of vehicle attitude and translation through the use of thrusters. In addition each spacecraft will need its 1-meter primary mirror mounted on piston, tip and tilt actuators and possibly need radius of curvature correction. The combination of the spacecraft thrusters and actuators will allow relative placement of individual spacecrafts mirrors on the virtual primary surface. The thrusters will continuously perform coarse updates to attitude and location while the onboard actuators will provide fine actuation.

Visibility and Wavefront Sensing

In imaging interferometry there is a dichotomy of terminology in that one can think of an image of an extended source as a set of fringes, each fringe with different amplitudes, directions, phase and contrast (visibility). Alternatively one can think of the image as the optical point spread function spatially convolved with the extended object. Each of these models gives different physical insight into the problem (Appendix E). If the source has the same spectral characteristics at each point on the source and a filter is used such $\Delta\lambda << \lambda_0$ then in this quasi-monochromatic approximation each fringe is formed as the interference of each non-redundant baseline pair. The focal plane distribution of photo-electrons can be spatially represented (Equation E.27, Appendix E) as:

$$n\left(\vec{\theta}\right) = n_e \frac{\pi D^2}{4\lambda_0^2} PSF_{PB}\left(\vec{\theta};\lambda_0\right)\left\{1 + \frac{2}{N}\sum_{k=1}^{N-1}\sum_{j>k}^{N} \mathrm{V_t}\left(\vec{B}_{jk},\vec{\theta},L_c\right)\mathrm{V_s}\left(\vec{B}_{jk},W_0,\lambda_0\right)\cos\left(\frac{2\pi}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta} + \phi\left(\vec{B}_{jk}\right)\right)\right\} \qquad (3.5)$$

where $\vec{\theta}$ is the location projected on the sky and n_e is the total number of photons collected by the two apertures comprising a baseline pair. $PSF_{PB}\left(\vec{\theta}\right)$ is the primary beam point spread function normalized to unity at the origin, $\vec{B}$ is the baseline spacing and direction, $\mathrm{V_t}\left(\vec{B}_{jk},\vec{\theta},L_c\right)$ is the visibility due to the finite passband and $\mathrm{V_s}\left(\vec{B}_{jk},W_0,\lambda_0\right)$ is the visibility due to the finite source size and $\phi\left(\vec{B}\right)$ is the phase due to both the object phase and any phase delays incurred in propagation through the system.

The total focal plane intensity can be found by summing over each of the fringe packets, one per baseline pair. It is evident from equation 3.5 that each of the fringe packets has a different unique spatial carrier frequency due to the non-redundant aperture configuration. Thus, in the uv-plane (Fourier domain) each of the fringe packets is given by the convolution of 3 delta functions with the Fourier transform of the primary beam point spread function. In the quasi-monochromatic approximation the UV-plane photon distribution is given by the 2D spatial Fourier transform of equation 3.5 as:

$$\tilde{n}(u,v) = OTF_{PB}(u,v) ** \left\{ n_e \delta(0) + V_T \frac{n_e}{2} e^{ik_0 \Delta z} \delta\left(u - \frac{B}{\lambda_0}\right) + V_T \frac{n_e}{2} e^{-ik_0 \Delta z} \delta\left(u + \frac{B}{\lambda_0}\right) \right\} \quad (3.6)$$

Where $OTF_{PB}(u,v)$ denotes the optical transfer function of the primary beam and "**" denotes 2D spatial convolution.

The uv-plane distribution of photons, for a given baseline pair, is given by three terms consisting of a u,v = 0 (zero frequency) term and two terms at $u = \pm B/\lambda_0$. The photons useable for sensing the phase difference between this baseline pair is given only by sum of the photons in the two terms at $\pm B/\lambda_0$ or by $V_T n_e$. The photons in the uv = 0 component contribute only to a reduction in the contrast of the fringes. For photon counting detection the theoretical lower bound in sensing the phase difference between the two apertures is given by $\sigma_\phi = 1/\sqrt{V_T n_e}$ [Loudon, 1985] or in units of optical path length difference as $\sigma_{OPD} = \frac{\lambda_0}{2\pi} \frac{1}{\sqrt{V_T n_e}}$. Thus the total number of photons per baseline pair required to sense the piston difference (OPD) to knowledge of σ_{piston} is given by:

$$n_e = \left(\frac{\lambda_0}{2\pi}\right)^2 \frac{1}{V_T \sigma_{OPD}^2} \quad (3.7)$$

If we desire knowledge of the piston difference between any baseline pairs to σ_{piston} we require that n_e photo-electrons are collected by this pair of apertures. Note that the visibility in the denominator is a function of OPD difference, passband, center wavelength, baseline and angular diameter of the source. These numbers represent only lower bounds on the number of photo-electrons required to achieve a given sensing knowledge and that due to algorithmic errors and other unknowns the numbers are expected to be larger.

The visibility function for any baseline pair of apertures is separable into a temporal and a spatial visibility function, $V_T = V_{temporal} V_{spatial}$, if the extended source has the same spectral characteristics at each point in the source. For a circular extended source, with piston differences between the apertures, the total visibility function can be shown to be approximately given by:

$$V_{T,jk} \approx \left[e^{-\frac{\pi^2}{4 L_c^2 \ln 2}\left[\vec{B}_{jk}\cdot\vec{\theta}+2\Delta z_{jk}\right]^2} \right] \left[\frac{2J_1\left(\pi B_{jk} W_0/\lambda_0\right)}{\pi B_{jk} W_0/\lambda_0} \right]$$

where the coherence length is given by $L_c = \lambda_0^2/\Delta\lambda$ and λ_0 is the center wavelength, $\Delta\lambda$ is the full-width ½ max passband of the source and spectral filters, B_{jk} is the baseline spacing of the j-k th aperture and W_0 is the angular diameter of the source on the sky and Δz_{jk} is the piston difference between the j-k th apertures. The first term in brackets (Figure 3.9: Top) is the temporal visibility function with various piston differences between the j-k th aperture with $\theta = 0$. The middle term represents a decay in visibility with increasing baseline and angle on the sky, i.e. this component is spatially non-stationary in the focal plane for a Fizeau interferometer. Note the temporal visibility is a rapidly decaying function with increasing passband, baseline, piston errors and angle on the sky. The 2nd term is due to the spatial extent of the source (Figure 3.9: Middle) and is rapidly decaying function of baseline and source diameter. The total visibility for a nominal source size of 2 milli-arcseconds versus baseline and for OPD differences from 0 to 20 microns is plotted at the bottom of Figure 3.9.

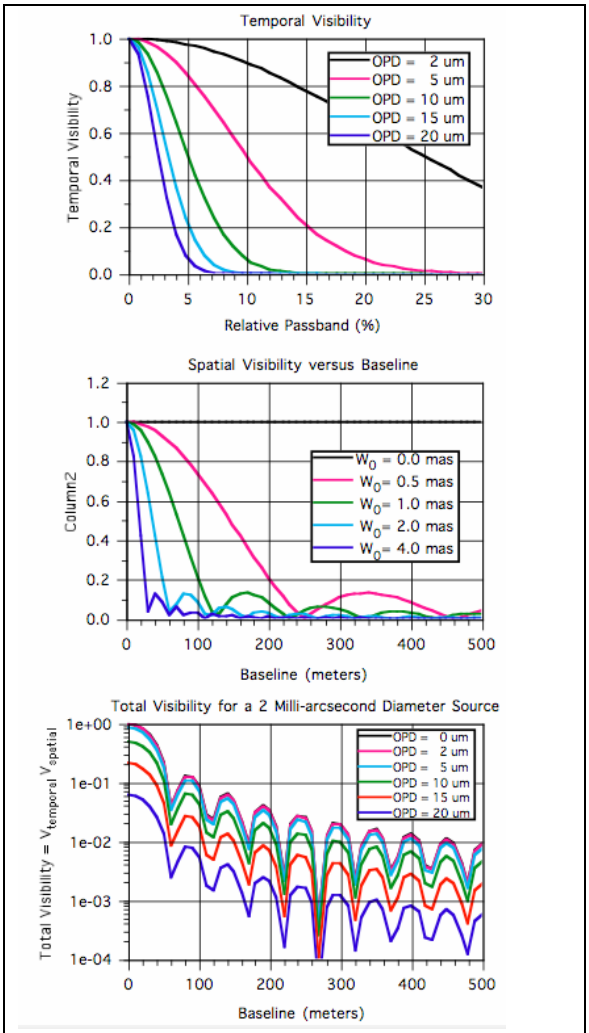


Figure 3.9 – Temporal, Spatial and Total Visibility

The total number of photo-electrons collected by any given baseline pair is:

$$n_e = M_0 \cdot 10^{-M_v/2.5} \cdot 2\Delta A \cdot \Delta\lambda \cdot T(\lambda_0) \cdot q.e.(\lambda_0) \cdot \Delta t \qquad (3.8)$$

where $M_0 = 1 \times 10^8$ photons/meter^2/second/nanometer is the source photon rate for a zero visual magnitude source and M_v is the apparent visual magnitude, ΔA is the area of a single aperture, $T(\lambda_0)$ is the transmission of the optics, $q.e.(\lambda_0)$ is the quantum efficiency of the detector, and Δt is the integration time in seconds. Using D as the diameter of a single aperture and noting that only two apertures are used per baseline pair yields:

$$n_e = M_0 \cdot 10^{-M_v/2.5} \cdot \frac{1}{2} \pi D^2 \cdot \Delta\lambda \cdot T(\lambda_0) \cdot q.e.(\lambda_0) \cdot \Delta t$$

Plotted at the top of Figure 3.10 are the photo-electrons per baseline pair required to recover 10 microns of OPD at 5% passband, for a source diameter of 2 milli-arcseconds, to knowledge 3.88 – 15.50 nm versus baseline. Recovery to 3.88 nm at a 500 meter baseline requires ~10^5 photo-electrons collected per integration period for the two apertures.

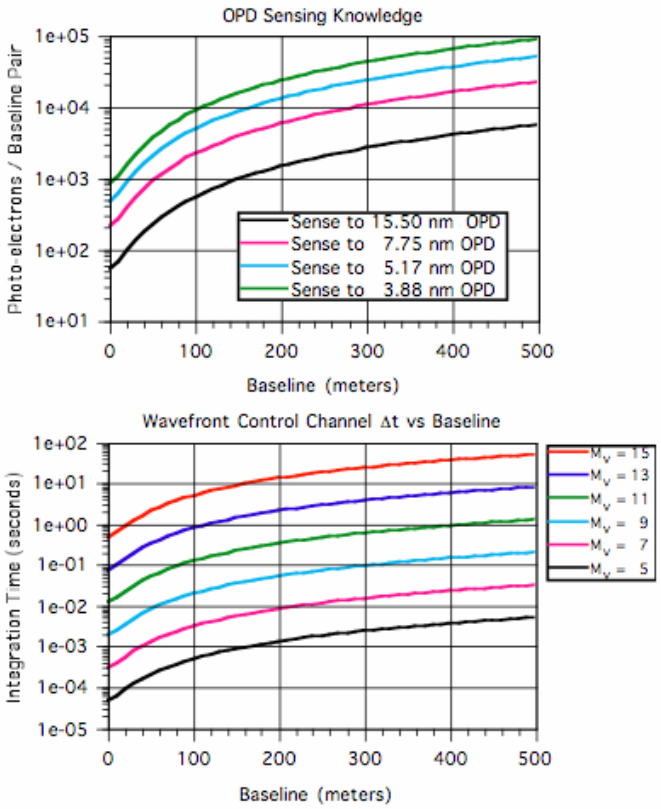


Figure 3.10 – WFS Knowledge & WFC Integration Time

Using the above and solving for the integration time to meet a given requirement for knowledge of OPD between any baseline pair gives the integration time as:

$$\Delta t = \frac{\lambda_0^2}{2\pi^3 V_T \sigma_{OPD}^2} \frac{1}{M_0 \cdot 10^{-M_v/2.5} \cdot D^2 \cdot \Delta\lambda \cdot T(\lambda_0) \cdot q.e.(\lambda_0)} \tag{3.9}$$

The WFC channel integration time versus baseline to sense 10 microns of OPD to 5.167 nm knowledge per baseline pair for a 5% passband with 50% SI throughput is plotted at the bottom of Figure 3.10 versus baseline for 2 milli-arcsecond diameter stellar sources of visual magnitudes from 5 to 15 magnitudes.

Noting that $\lambda_0 = 500$ nm is the center wavelength of the wavefront sensing system and that we desire knowledge of the OPD to $\lambda_s/10 = 15.5$ nm ($\lambda_s/20 = 7.75$ nm for spacecraft motion) where $\lambda_s = 155$ nm is the shortest science wavelength. Thus we require the wavefront sensing channel to sense to 7.75 nm. Operating the WFC channel at once per 10 sec (0.1 Hz) down to a M_v = 13 magnitude source appears feasible and would allow closed loop feedback to phase correct the science image to better λ/10 frame every 10 seconds for all baselines and to provide an error signal to the actuators and pointing control system to maintain alignment of sub-apertures to less than 10 microns.

Phase Diversity Wavefront Sensing

The baseline method for estimating the phase differences between baseline pairs, from imagery collected by the WFCCam, is known as *Phase Diversity* [Gonsalves 1982, Kendrick et al. 1994]. Phase Diversity was briefly described in section 2.2.3 and is described in more detail and simulations of its performance are shown herein.

Phase diversity requires as input: two (or more) collected images from the WFCCam with known phase differences (non-common path wavefront error) deliberately added in, an optical model of the system, and statistical model of the noise. The data model is given by:

$$\begin{cases} I_1\left(\vec{\theta},\lambda_1\right) = PSF_1\left(\vec{\theta},\lambda_1;\vec{a},\Delta\phi_1\right) ** O\left(\vec{\theta}\right) + \eta_1\left(\vec{\theta}\right) \\ I_2\left(\vec{\theta},\lambda_2\right) = PSF_2\left(\vec{\theta},\lambda_2;\vec{a},\Delta\phi_2\right) ** O\left(\vec{\theta}\right) + \eta_2\left(\vec{\theta}\right) \end{cases} \tag{3.10}$$

where $I_j\left(\vec{\theta},\lambda_j\right)$ represents the two, $j=1...2$, collected sampled diversity images, collected at the same time, each at a different wavelength λ. $O\left(\vec{\theta}\right)$ represents the unknown object and $\eta_j\left(\vec{\theta}\right)$ noise, not necessarily additive, $\vec{a}$ represents the vector of piston, tip and tilt coefficients to be solved for. Each of the two channels has a different PSF due to the difference in wavelength and the deliberately introduced phase differences $\Delta\phi_j$. There are two sets of unknowns in this problem: (i) the object and (ii) the wavefront parameterized in terms of piston, tip and tilt. The equations are linear in the object and non-linear in the phase. If a metric is constructed with both these as unknowns, and we assume that the PSFs are known then using simple inverse filtering the object can be estimated and inserted back into the metric, thereby removing the object from the equations. Without belaboring the math a Gonsalves metric can be constructed of the form:

$$E\left(\vec{a}\right) = \sum_{u,v} M\left(u,v\right) \frac{\left|OTF_1\left(u,v;\vec{a},\lambda_1\right)\tilde{I}_2\left(u,v\right) - OTF_2\left(u,v;\vec{a},\lambda_2\right)\tilde{I}_1\left(u,v\right)\right|^2}{\left|OTF_1\left(u,v;\vec{a},\lambda_1\right)\right|^2 + \left|OTF_2\left(u,v;\vec{a},\lambda_2\right)\right|^2} \tag{3.11}$$

where $OTF_1\left(u,v;\vec{a},\lambda_1\right)$ and $OTF_2\left(u,v;\vec{a},\lambda_2\right)$ are the u,v-plane optical transfer functions for channels 1 and 2 respectively, $\tilde{I}_1\left(u,v\right)$ and $\tilde{I}_2\left(u,v\right)$ are the uv-plane spatial spectrums (2D Fourier transforms) of the observed images, and $M\left(u,v\right)$ is a binary mask that is unity at the valid spatial frequency points within the uv-plane and zero otherwise and is used to mitigate against zeros in the denominator. In practice the denominator is regularized via the noise spectrum (Wiener or Wiener-Helstrom filter). $\vec{a}$ represents the vector of unknown piston, tip and tilts of each of the mirrorlet beam paths; note that higher order terms can be included. The constructed metric is now only a function of the unknown vector of piston, tip and tilts. Multiple other metrics can be constructed see Kendrick, 1994 of other metrics. In principle the solution vector can be found by minimization of equation 3.11 with respect to $\vec{a}$ and can be found by convergent nonlinear optimization methods. However for known wavelength differences and known diversity functions for non-redundant aperture systems the solution can be found by a more direct method (Lyon, 2005 – In progress). This approach relies on the separability of the baseline pairs in the uv-plane both with wavelength and with spatial frequency. If no tip or tilt is present and only one wavelength and passband are used for both images the piston differences can be directly solved for, however, with 2 (or more wavelengths) tip and tilt can also be directly solved for.

The net effect of using two wavelengths is to mitigate against the phase *wrapping* problem. Using only a single wavelength nominally recovers the phase modulo 2π, or 1 wavelength of light. For a contiguous wavefront this ambiguity can resolved by exploiting the fact that any closed contour of the phase must integrate to zero, however, for a discontinuous wavefront such as in a sparse aperture system, this cannot be used. Using two wavelengths gives and effective wavelength of $\lambda_1\lambda_2/(\lambda_2-\lambda_1)=11$ microns, making the range over which phase diversity can recover the phase larger. Also the for a single wavelength the spatial visibility has zero points in the Bessel function, yielding theoretically no visibility, or no fringe contrast and a uniform fringe, at these points. In practice this is seldom the case since the Bessel function comes from a uniform disk and it is unlikely that any real source would have this characteristic; more likely the disk has structure and soft edges which soften the zeros of the Bessel function. However there would still likely be regions of low visibility versus baseline and using two wavelengths helps mitigate this effect as the regions will be shifted with wavelengths. The drawback to using multiple wavelengths is that if the source itself changes with wavelength then the phase diversity results will loose accuracy.

Figure 3.11 shows a simulation of the phase diversity accuracy in sensing 10 microns of piston and 2.5 microns of tip versus baseline. Two 5% passband images were used, one from 500 – 525 nm and the other from 525 – 550 nm. Each image has a peak SNR of 10, with photon noise only, and a source size of 2 milli-arcseconds diameter. For baselines longer than 50 meters the piston sensing was better than 7 nm of absolute accuracy and the tip sensing was better 14 nm of absolute accuracy; shortward of 50 meter baselines the results are even better.

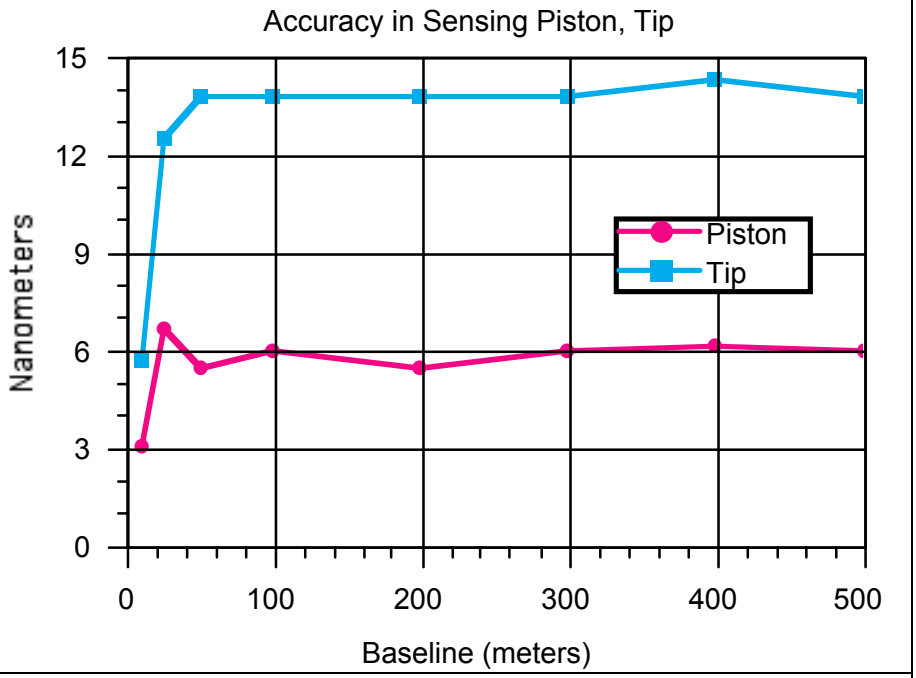


Figure 3.11 – Accuracy in Sensing Piston & Tip

Sensing of piston and tip with a 2 milli-arcsecond source and 5% passband at a peak SNR of 10.

3.1.5 Detectors

Some candidate detectors that could satisfy the design requirements given in section 2.2 are as follows:

- Back-thinned silicon CCD arrays
  - Mature technology
  - Multiple readouts
  - UV through visible
  - +/- 1 degree C stability
  - -20 to + 20 degrees C
- Active pixel photo-diode arrays
  - Silicon based
  - Coming to maturity
  - Requires coating for UV
  - Multiple readouts
  - UV through visible
  - +/- 1 degree C stability
  - -20 to + 20 degrees C
- Silicon carbide arrays
  - GSFC and JPL are pursuing funding of SiC detector development in the UV from NASA HQ
  - Maturing fast
  - Radiation hard
  - Good in UV because of band gap, but cuts off at ~400 nm
  - Can operate up to 800 degrees C; also can operate cold
  - +/- 10 degree C stability

However, the best approach for the UV detectors might be use of AlGa-Nitride photocathodes with photon-counting detectors, such as EBCCD's (electron-bombarded CCD's) or micro-channel plate arrays. Although we call out the possible utility of energy-resolving detectors in several parts of this report – the need for them can be eliminated by adoption of the one the alternative beam-combiner designs discussed in Appendix F for remapping the two-dimensional array pattern into a one-dimensional non-redundant pattern that can be crossed-dispersed to obtain spectral (i.e., energy-resolution) information. Radiation considerations for any detectors included in the design are summarized in the following two figures:

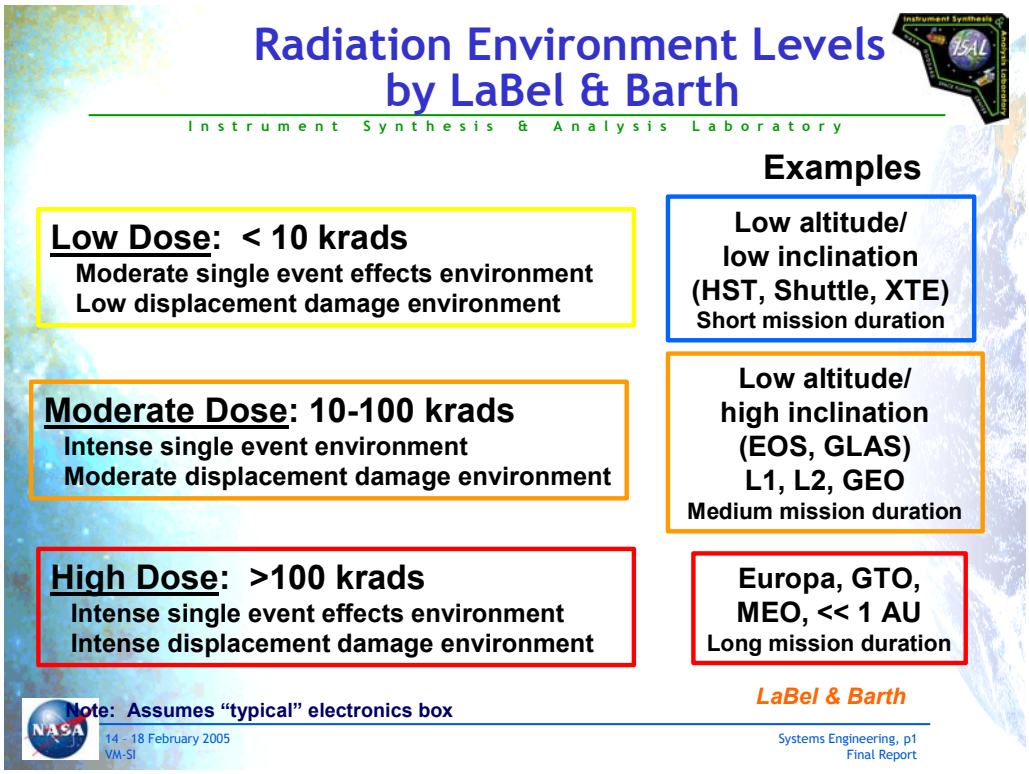


Figure 3.12: Radiation Environment at various locales, including L2.

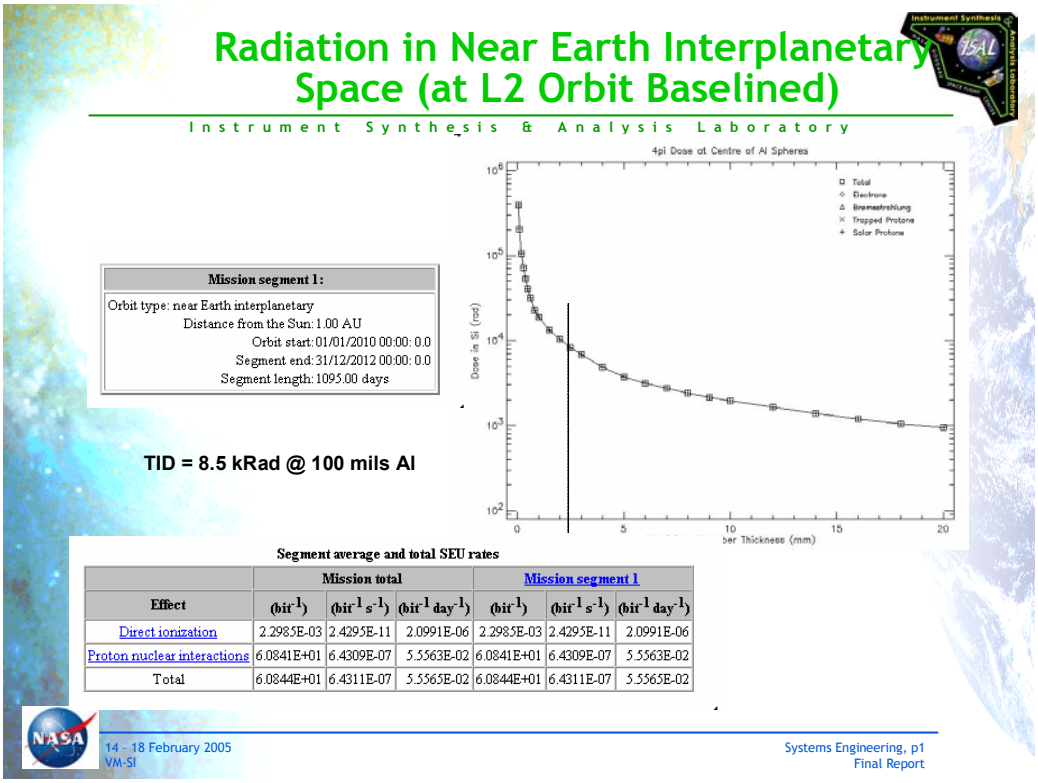


Figure 3.13: Radiation Environment details near Sun-Earth L2.

3.1.6 Lightweight UV Quality Mirrors

P. C. Chen (CUA/Code 667 GSFC) has been working on the development of lightweight telescope mirrors that meet SI requirements, namely meter class aperture size, very low areal density, ultra smooth surfaces, long focal lengths or flats, ability to be produced in quantity with excellent uniformity, and low cost. The work is an extension of studies previously carried out at GSFC to demonstrate the feasibility of making precision lightweight mirrors (with areal density in the 5-10 kg/sq.m range) using graphite composite laminates and an overcoat of polymer resin.

There has been considerable success in the development of special polymer resins that give very smooth surfaces without the need for super high quality mandrels. A non-contact process has been found that can further smooth the polymer resin, as well as to shape the surface contour to obtain high optical figure accuracy. The process is similar in principle to ion milling (IM) which is a well known, established industrial practice to perform ultra-high accuracy (1/100 wave) on super fine optics. Unlike IM, however, the new process both etches and smooths surfaces, and does so at a much higher rate.

The following images and their captions give a brief summary of the work to date. All measurements were made using a Zygo Maxim 5700 Interferometric Microscope Profiler with a 20X Mirau objective, multiple phase integrations, and system error subtract.

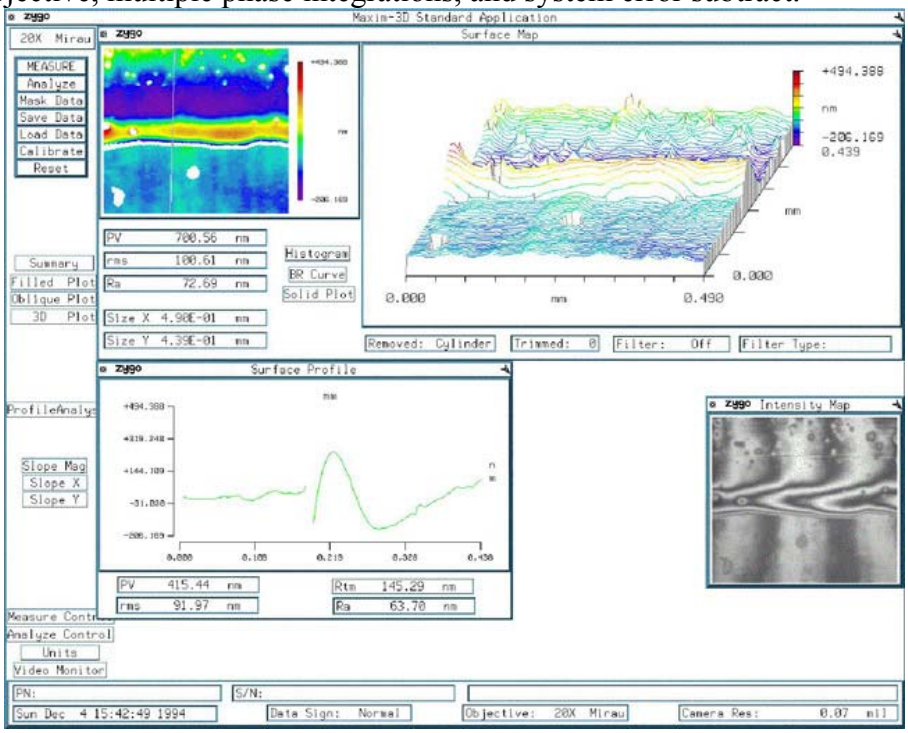


Figure 3.14: Plasma Treated Composite Mirror Surface

A scan using a chemically enhanced AC plasma across the surface of a composite mirror having a resin-rich optical surface. The upper left corner image shows the surface profile across the plasma treated (front) and the covered (back) regions. It can be seen that the front side is noticeably smoother than the back. Additional measurements indicate that: rms microroughness (front) ~ 25 nm; rms microroughness (back) ~ 37 nm; etch rate ~400 nm/16s ~1.6 microns/min. The plasma therefore both etches (figure correct) the surface and makes it smoother at the same time.

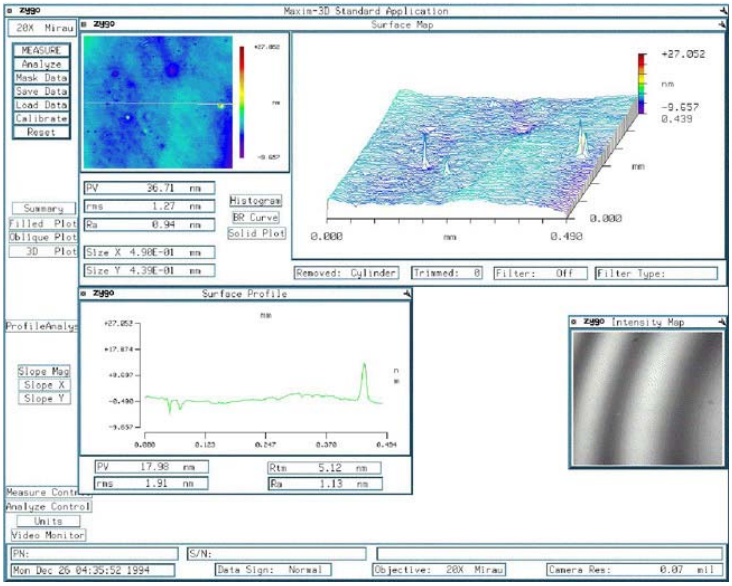


Figure 3.15: New Resin System:

A specially formulated, proprietary polymer resin was utilized as the basis for the mirror. The surface was cured (hardened) while spinning on a turntable (spincasting) to generate a parabolic figure. RMS microroughness is 12.7 A after subtracting the surface curvature term ('cylinder' in box at bottom of upper left figure). The result shows that very smooth surfaces can be obtained by curing this polymer resin in air, without the use of super-smooth mandrels or the usual labor-intensive polishing methods.

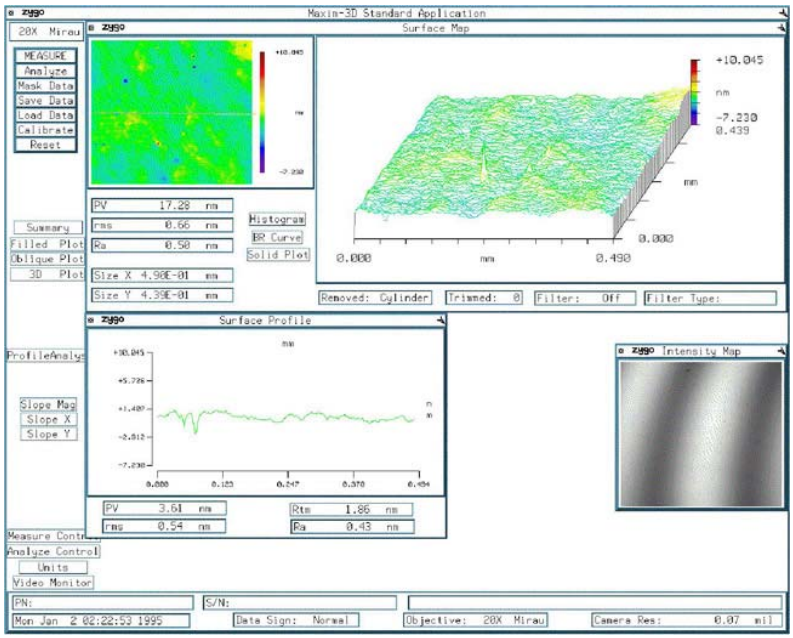


Figure 3.16: Plasma Treated Polymer Resin Surface.

The Zygo plot shows the texture of the polymer resin after plasma treatment. By using a suitable combination of polymer, plasma type, plasma density, and feed gas, the surface has been smoothed down to 6.6 A rms. The data indicates that a mechanism has been found which is capable of generating large area, extremely smooth surfaces. The process is very rapid (seconds) and low in cost. In combination with lightweight graphite fiber composite mirror structures, the new surfacing technology provides a promising venue for the development of high performance space mirrors such as that required for the SI and similar projects.

Fig. 3.17: An ultra-smooth 20 cm diameter flat mirror.

The optical surface is formed by pouring a liquid polymer onto a prefabricated graphite fiber composite laminate and letting it cure in place. Surface microroughness is 3.6 Angstroms rms. The ability to form large area optics with extremely smooth surfaces makes this new technology a very promising avenue for SI.

The baseline SI design requires very shallow mirrors (~ f/1000 – f/10000) whose curvature can be adjusted to provide different (long) focal lengths, varying from ~ 1 km to 10 km. Such long, variable focal length mirrors can be produced most easily by manufacturing ultra-smooth flats and actively forming the desired surface. The mirrors described above can easily be adjusted to the desired curvature by bending or slightly heating the flat mirror (Chen & Carpenter 2003; also, see Figure 3.18.). This actuation would also enable a fine-tuning of the mirror surface shape to match the shape of the virtual parabola describing the main array at any particular point in that parabola.

For the SI, the fabrication procedure would be:

1. Make l m flat composite plates
2. Apply resin and cure
3. Use Reactive Ion Etching (RIE) for fine figuring, then aluminize
4. Attach ~6 actuators for shape control

In summary, very significant progress has been made in a new mirror technology that combines low areal density, low cost, extremely smooth surfaces, rapid fabrication, and the ability to be fine figured by a deterministic non-contact method. Pending further developments and refinements, the new process holds promise to enhance the science capability of the SI mission while reducing the project cost.

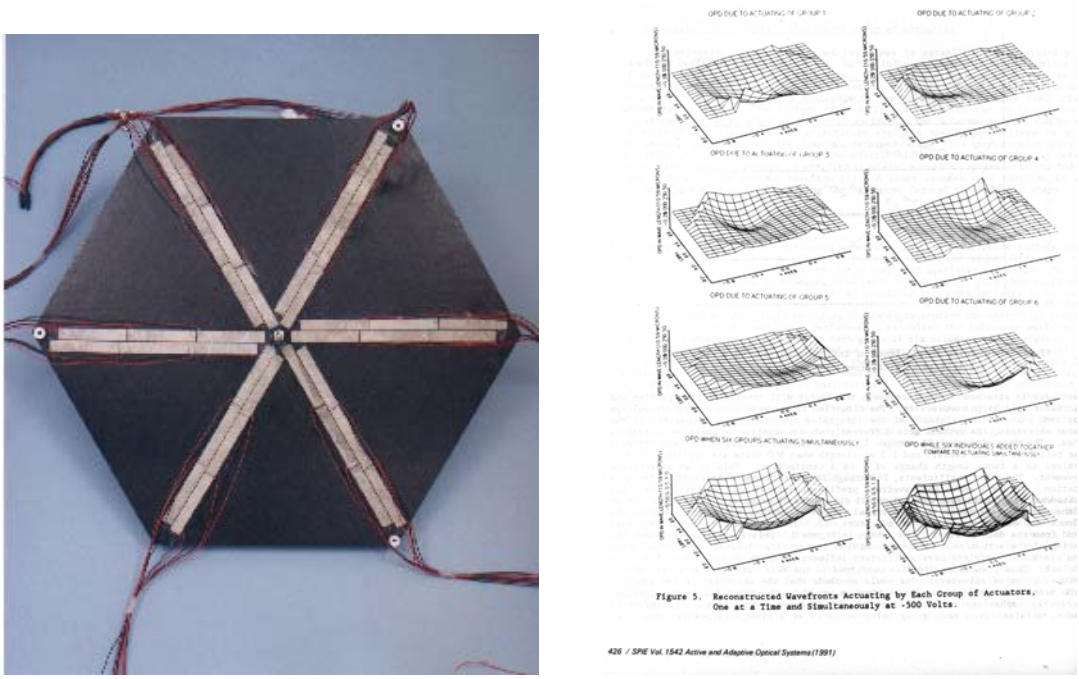


Fig. 3.18. The SI can use a simple proven scheme to adjust shape, focal length, and correct for thermal gradients. The image above shows an earlier generation composite mirror from the JPL Precision Segmented Reflector program. Six strips of PZT are bonded to the back of this 0.5m spherical mirror. By actuating the PZT strips individually or in unison, it is possible to generate (or correct) >10μ of the Seidel terms including focus, tip-tilt, coma, astigmatism, etc. (Source C.P. Kuo, SPIE 1542, p420, 1991. Photo courtesy of C.P. Kuo, NAS JPL).

3.1.7 Mirrorsat Spacecraft

The 30 Mirrorsats and their dispenser spacecraft (S/C) as an ensemble are the most massive element to be deployed for SI – it is thus important that and mass (and volume) savings in a single Mirrorsat be an important consideration in their design, since any savings can be multiplied by a factor of 30. Ball Aerospace Technology Corporation therefore considers in this section ways of achieving such savings relative to the original mirrorsat design from the October, 2004 IMDC results.

The resulting notional SI Mirrorsat spacecraft design is optimized for interferometry mission support and uses an innovative configuration and functional architecture plus a suite of advanced subsystem technologies to minimize both mass and volume. Compared to October 2004's reference Mirrorsat design, this approach offers potential reductions of a factor of two in both mass and volume. Much of the S/C hardware used and between 60 and 80% of the flight software loads are flight-proven or flight-qualified, adding credibility to our notional Mirrorsat approach.

The alternative mirrorsat design presented in this section has not been fully integrated with the overall "baseline" SI design, but it should be possible to do so with modest future effort, if it is deemed feasible and worthwhile.

3.1.7.1 Mirrorsat S/C Configuration

Figure 3.19 shows forward (towards the Hub S/C) and aft views of the Mirrorsat S/C, which is optimized for mirror carriage and support. Key features visible include the 1m diameter mirror, equipment shelf, solar array and the two deployed sunshades. Subsystem units on the aft side of the

equipment shelf are identified. The star cameras' locations on the aft side of the equipment shelf provides them with clear Fields of View (FOV's). Most other units mount to the forward side of the shelf. This single-shelf approach minimizes mass and volume and provides easy access to avionics, facilitating Assembly, Test, and Launch Operations (ATLO).

The mirror and its triangular mounting frame are supported from the forward side of the shelf using 6 voice-coil actuators, each in series with a piezo actuator. The six cascaded actuators are arranged as 3 bipods which together form a hexapod mount to allow the mirror six degrees of freedom with respect to the shelf. This mirror support approach combines mm-scale articulation range with nm level resolution.

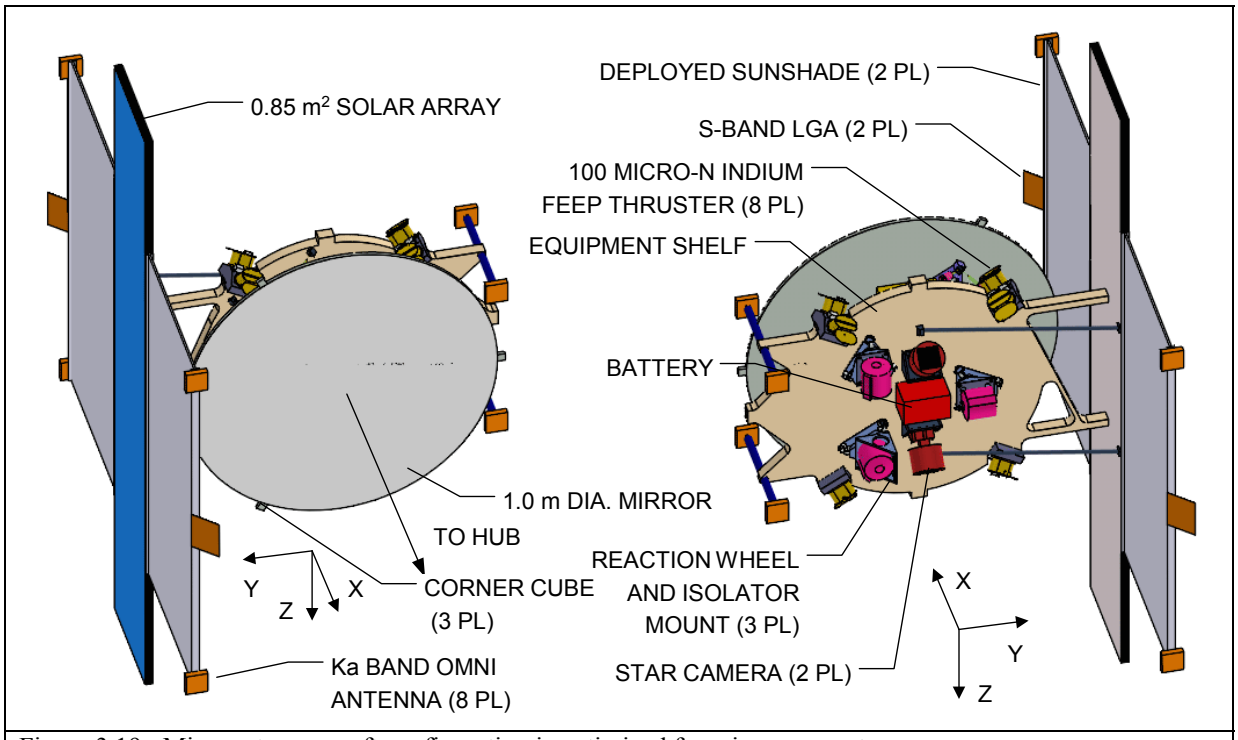


Figure 3.19: Mirrorsat spacecraft configuration is optimized for mirror support.

The 0.85-m^2 fixed rectangular solar array attaches directly to the equipment shelf and is braced by two struts. Inflatably deployed sunshields mount to the top and bottom of the array. In this location the array and sunshields continuously shadow the Mirrorsat S/C and the mirror during the entire mission to simplify thermal control and maximize mirror stability. Three laser corner cubes equally spaced on the mirror periphery are used by the Hub S/C to aid nm level mirror positioning.

Eight Field Emission Electric Propulsion (FEEP) thrusters are mounted to the periphery of the equipment shelf. Their location and orientation allows them to provide 3-axes of rotation using full couples for momentum management, and three axes of translation without rotation for station keeping. Three reaction wheels and their Jitter Isolation Mounts (JIM) mount to the bottom of the equipment shelf.

A ten-element antenna farm supports Mirrorsat communications and navigation functions. Two S-band LGA's are mounted to the centers of the fore and aft sunshades to provide each with a hemispherical FOV. Eight Ka-Band omni antennas, two transmit and six receive, together provide close to spherical coverage for the pseudo-GPS-based coarse navigation system.

3.1.7.2 LV Interface

Figure 3.20 shows the SI Mirrorsat launch configuration. 30 Mirrorsats are arranged in six columns of five each spaced around the hexagonal Dispenser S/C. Minimum separation between mirrorsats is 25.4-mm (1.0"). Once the Dispenser is properly aligned and spaced with respect to the Hub S/C at L_2, Mirrorsats are individually released from the Dispenser with each proceeding independently to its proper station.

The Dispenser provides the LV interface for the entire stack. It is attached by a 1676-mm (66") diameter marmon clamp to the Delta model 1666-5 Payload Attach Fitting (PAF). CAD layouts show a minimum clearance of 165 mm (6.5") between the periphery of the Mirrorsat solar arrays and Delta-IV 5-m fairing static envelope. Clearances elsewhere are much larger.

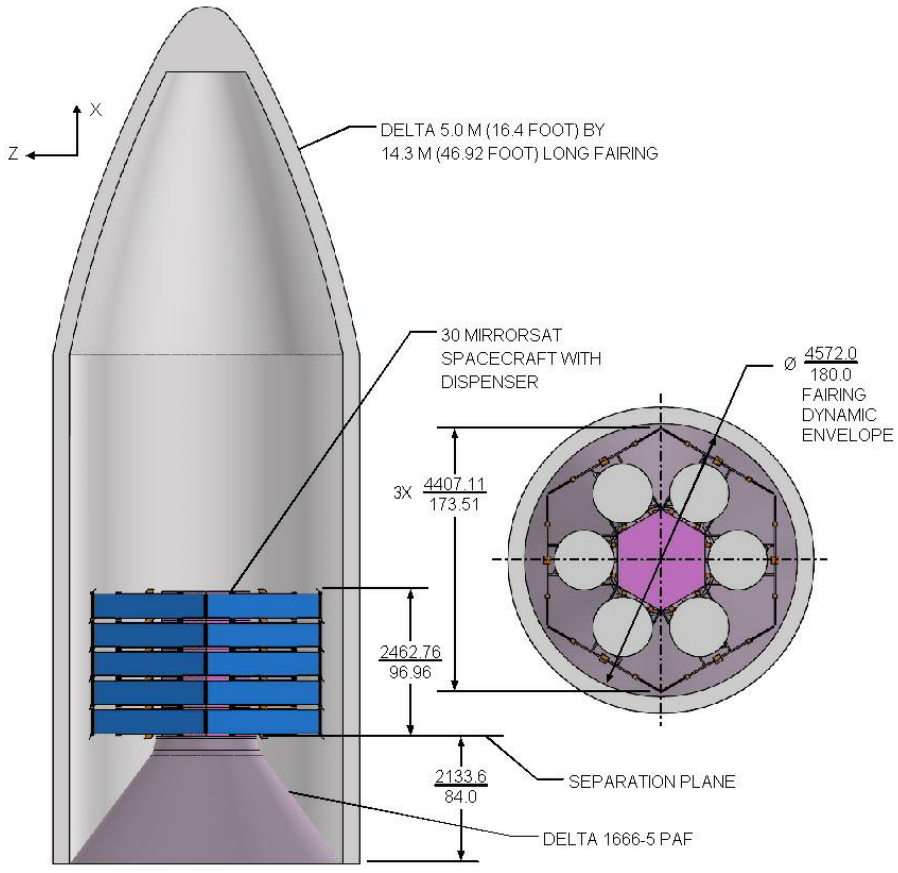


Figure 3.20: One possible Mirrorsat launch configuration.

3.1.7.3 Mirrorsat Deployment and Operations Overview.

In operation, the Mirrorsat spends 99.99% of its ten-year life in observation mode, with its mirror located and mirror normal oriented with respect to the Hub with nanometer and nanoradian accuracy respectively, and with its array normal pointed within 20 deg of the perpendicular to the sun.

Prior to beginning observation support, the Mirrorsat is delivered to its station orbiting the earth-sun L_2 Lagrange point by the Dispenser S/C. Mirrorsat deployment begins after the Dispenser and its

cargo of 30 Mirrorsats are orbiting the L_2 point and are correctly located and oriented with respect to the Hub S/C. The deployment process begins as the Dispenser S/C rolls until one of the six columns of Mirrorsats is oriented so that all their arrays are normal to the sunline. This allows each Mirrorsat to achieve power balance prior to separation. Separation of an individual Mirrorsat is accomplished by three shape-memory separation bolts. Actuation of the bolts frees the S/C, which is then separated radially at about 30-mm/s by a separation spring.

As it leaves the dispenser, the Mirrorsat is powered up with its reaction wheels cycling around their nominal rpm to keep the S/C momentum at zero. Immediately after separation the reaction wheels are used to zero any residual S/C rates and to point the solar array normal to the sun, and the mirror normal to the nominal location of the Hub S/C.

Following initial orientation, the mirror hexapod actuators and reaction wheel Jitter Isolation Mechanisms are uncaged by shape-memory-alloy separation devices. Uncaging is followed by checkout of the 8 FEEP thrusters, deployment of the two sunshades, and actuation and checkout of the pseudo-GPS coarse navigation system.

After checkout, the pseudo-GPS now provides the S/C with its location with respect to the Hub S/C. The Mirrorsat knows where it should be. The current and goal destination are differenced onboard to generate a trajectory to the goal. FEEP thrusters are used to implement the transfer, which concludes with the Mirrorsat on station to an accuracy of about 50-mm (2.1")

When located to this level of accuracy, the S/C mirror can be acquired by the Hub S/C's metrology/attitude control system and the Mirrorsat can be guided into its final position and orientation with the required level of accuracy. At this point the Mirrorsat transitions to its Observation Support (OS) mode.

In OS mode the Mirrorsat bus maneuvers using its reaction wheels and its FEEP thrusters such as to keep the mirror hexapod voice-coil actuators in the center of their articulation range. The voice-coil actuators are controlled to keep the piezo actuators in the middle of their range. The piezo actuators in turn are controlled by high speed control loops driven by the Hub S/C's metrology system to keep the mirror in position and aligned.

At the end of its life, or if afflicted with an incurable fault, the Mirrorsat will use its FEEP thrusters to accelerate outward along the radius vector back to the Hub S/C until it reaches "escape velocity" with respect to the L_2 Lagrange point, eliminating any possibility of interference with the rest of the SI constellation.

3.1.7.4 Structure & Mechanisms.

The Mirrorsat structure uses conventional GFRP construction to minimize mass**.** The primary structure consists of only two elements, the equipment shelf and the solar array substrate. Both are fabricated from aluminum honeycomb using GFRP facesheets. Secondary structure elements such as thruster brackets and the array brace tubes are machined from Al Alloy bar stock or fabricated from GFRP tube stock.

Mechanism elements include the reaction wheel Jitter Isolation Mounts (JIM), the sunshades, and the mirror support voice-coil/piezo actuator cascade.

Each reaction wheel is supported by three JIMs. Each JIM consists of two elliptical constrained-layer flexures at right angles to each other that provide high-frequency isolation in six degrees of freedom. Each ellipse consists of a 10-mil thick viscoelastic layer constrained between two 9-mil stainless steel sheets. To withstand launch loads, each reaction wheel is caged by shape-memory separation devices to unload the JIMs during launch and ascent. This JIM approach was developed and extensively proven during a two-year BATC IR&D program.

The two sunshades consist of parallel inflatable tubes that deploy an aluminized Mylar curtain when pressurized. Self-rigidization is provided by choosing an inflation pressure sufficient to slightly yield the thin aluminum walls of the deployment tubes, providing rigid sunshade support after removal of pressurization. Each sunshade incorporates an independent and redundant inflation system.

The series voice-coil and piezo actuator struts used by the mirror support hexapod are designed to provide a total stroke of 5 mm with resolution measured in nm. Together six struts provide the mirror with 6 degrees of freedom, allowing use of a single FEEP thrust level and reducing the FEEP duty cycle.

3.1.7.5 Functional Block Diagram.

The Mirrorsat's processor and data-bus based functional architecture is summarized in the top-level system block diagram shown in Figure 3.21. The diagram is color coded to show hardware heritage. The figure shows the Mirrorsat's division into six functionally distinct subsystems plus the S/C's structure and mechanisms. The system architecture is single-string for minimum mass and cost.

3.1.7.6 Electrical Power (EP).

The Mirrorsat EP subsystem must maintain a positive S/C power balance for all mission modes and S/C orientations. Table 3.5 provides a summary spacecraft power budget by operational mode. The large margins shown further reduce risk. A fixed 0.8-m^2 (net) solar array using cascade multi-junction solar cells with 30% efficiency provides power generation. Energy for potential temporary off-sun operations during deployment is stored by a 6 A-H Li ion battery. Battery charging is controlled by solar array switching implemented by the charge control card in the SCU, which also provides under-voltage and over-current protection.

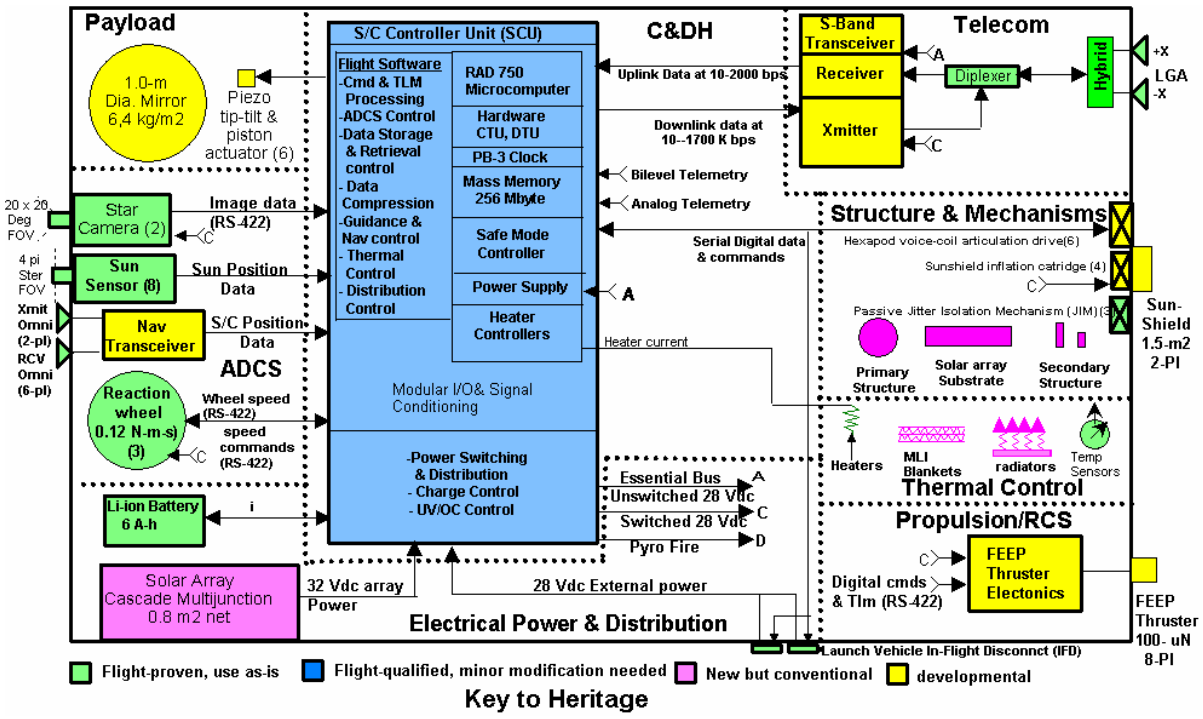


Figure 3.21 Mirrorsat spacecraft functional block diagram shows processor based " single-box" architecture for minimum mass.

Table 3.5: Mirrorsat S/C Summary power budget shows 26% available margin

Subsystem/Item	CBE Power (W)	Cont. (%)	Cruise Electrical Power (W)	Heritage/Comments
Electrical Power & Distribution	1.0	30%	1.3	Cabling losses per RS-300 7/01 MEL
Command, Control & Data Handling	35.0	15%	40.3	Per Broadreach Tech XSS-11 SCU
Telecommunications	4.0	30%	5.2	Per GSFC Estimate
Thermal Control	5.0	45%	7.3	Worst -case
Attitude Determination & Control	45.9	2%	46.8	Star cameras, wheels, Nav transceiver
Propulsion/RCS	40.0	30%	52.0	Allocated based on 2 X 100 μN Indium FEEP thrusters
Subtotal, SI Mirrorsat S/C Bus	130.9	17%	152.8	
Science Payload (Total)	10.00	0%	10.0	For piezoactuators and voice coil actuators
Total, Power draw, Sailcraft	140.9	16%	162.8	
Array Power, EOL (Worst-Case)	205.6		205.6	0.80 m2 Cascade Multijunction array 30 deg from sunline
Margin on S/C Electrical Power (W)	64.7		42.8	
Margin on S/C Electrical Power EOL (%)	46%		26%	

3.1.7.7 Command, Control and Data Handling (CC&DH).

The Mirrorsat CC&DH must provide all S/C C&DH functionality and support computation requirements for the ADCS, EP and mirror actuator subsystems. Our approach uses a RAD-750 based 8-card Spacecraft Control Unit (SCU) originally designed for the DOD TechSat-21 mission; it integrates all these functions into a single unit by extensive use of Application Specific Integrated Circuits (ASIC's).

With over 100 MIPS of computational capability, the SCU provides a test-proven 284% margin on a notional DI-derived FSW load's throughput requirements. The 41 MB FSW code image operates out of SCU RAM (256 Mbytes) providing over 500% margin. This generous program memory and the high throughput capabilities will be used to minimize CC&DH development risk. Risk is further reduced by inheriting DI's FSW-based autonomous fault management capability.

The SCU supports selectable downlink telemetry rates from 16 bps to 330 kbps and payload data transfer rates (storage and playback) of up to 20 Mbps. Housekeeping data rates are typically < 500 bps. The SCU RAM provides > 200% margin for estimated housekeeping data storage.

3.1.7.8 Telecommunications.

The Mirrorsat telecom subsystem must support high-rate communication with the Hub S/C and low-rate communications direct to earth. Our approach is based on that used by the IMDC Mirrorsat approach. It uses a single S-band transceiver combined with two low gain antennas (LGA) fed through a diplexer and a 6-dB hybrid. Together, the LGA's provide essentially 4π steradian coverage allowing communication with either the Hub S/C or earth, regardless of S/C orientation. IMDC studies indicate that this approach can provide a Hub/Mirrorsat two-way data rate of 100 kbps with a maximum range link margin < 23 dB. A 400 bps direct to ground link can be supported with a link margin of 0.3 dB

3.1.7.9 Thermal Control (TCS).

The Mirrorsat TCS must keep all S/C equipment within flight-required temperature limits for the nominal sun range of 1.0 AU and for the full range of S/C attitudes. The baseline Mirrorsat TCS approach is primarily passive and takes advantage of the benign environment at L2 and of having the S/C bus always shadowed. With the exception of portions its aft face, the entire equipment shelf is covered by multi-layer insulation. Radiator surfaces on the aft face of the shelf are sized to provide passive thermal balance over all spacecraft modes. Make-up heaters maintain equipment power dissipation effectively constant to simplify obtaining thermal balance.

3.1.7.10 ADCS.

The Mirrorsat ADCS must autonomously maintain the S/C in its operational attitude (+ X axis to Hub and + Y axis to sun) from Dispenser separation until mirror lockup, when control passes to the Hub S/C.

Table 3.6 summarizes ADCS characteristics and estimated performance. The S/C operates in two primary modes. As a free flyer after separation from the Dispenser the two star cameras serve as the attitude reference; attitude control is provided by the reaction wheels with accumulated wheel momentum periodically dumped by the FEEP thrusters. Adcole coarse sun sensors augment the star cameras and provide and alternate attitude reference to cope with anomalies.

Following "Lock up" of the mirror's position and orientation by the Hub S/C, the mirror itself becomes the primary attitude reference with the reaction wheels and the FEEP thrusters acting together in a low-bandwith control loop to keep the voice-coil actuators in the center of their range and the wheels close to their nominal rpm.

Table 3.6: Mirrorsat spacecraft ADCS Autonomous Mode Requirements and Characteristics

Characteristic/Parameter	Approach/Requirement
Control Method	3-axis zero-net-momentum with RCS wheel dump
Attitude Control Reference	Stellar-Inertial plus mirror
Pointing Control (3- Axes)	0.1 deg.
Pointing Knowledge (all axes)	20 arcseconds
Pointing Stability	N/A
Pointing Jitter	N/A
Attitude Knowledge Processing	Onboard
Slew Rate, all axes, max	0.02 deg/s
Agility	Array normal up to 20 deg. From sunline.
Postion Accuracy WRT Hub S/C	±54-mm, 3 -axes
On-orbit Calibrations	None required
Deployment/articulation	None required

3.1.7.12 Propulsion/RCS.

The Mirrorsat Prop/RCS must provide 3-axes of reaction control torques for reaction wheel momentum management and up to 164.4-m/s of delta-V for deployment and station-keeping. Our approach uses eight 100 μN FEEP thrusters located and oriented to provide three axes of torque-free delta-V for maneuvers and station-keeping and to provide 3-axes of rotation using full couples.

FEEP thrusters were chosen for their combination of high Isp which enables the 10 year mission life, and of linear thrust modulation capability, which simplifies S/C control.

The Mirrorsat notional propellant budget is calculated using a FEEP Isp of 5000 seconds and a 50% efficiency factor to account for thruster cant. A 30% contingency has been added to the calculated RCS and delta-V propellant loads.

3.1.7.13 Flight Software.

The existing DI FSW currently in the final stages of testing prior to launch performs most of the functions needed for a Mirrorsat, including ADCS, command processing (including delayed-command storage and software command processing), telemetry processing (including software data collection, formatting, and compression), autonomous S/C fault management, and general CPU management. It is compatible with CCSDS commanding and telemetry standards. Changes to the DI FSW package are limited to those required to adapt it to the Mirrorsat's set of operational modes, and to addition of the mirror articulation control functions.

3.1.7.14 Resources and Margins.

Mirrorsat S/C mass properties are summarized in Table 3.7 and show a wet mass at separation of 63.6 kg as well as an average 11% contingency for mass growth allocated at the bus level to reduce risk.

Table 3.7: Mirrorsat S/C Summary Mass List

Subsystem/Item	CBE Mass Kg	Contin-gency (%)	Mature Mass (kg)
S/C element peculiar Structure	6.46	30.0	8.4
Mechanisms	2.71	30.0	3.5
Electrical Power & Distribution	4.2	19.5	5.0
Command, Control & Data Handling	4.8	15.0	5.5
Communications	2.2	17.9	2.6
Thermal Control	0.4	30.0	0.5
NAV/ADCS	7.6	7.1	8.1
Propulsion/RCS	21.0	2.0	21.4
Subtotal, Mirrorsat S/C bus	49.3	12%	55.0
Science Payload (Total)	10.0	0%	10.0
Subtotal, Mirrorsat S/C, dry	59.3	10%	65.0
Propellant	0.4	10%	0.4
Total, Mirrorsat S/C, wet	59.7	10%	65.5

3.1.8 Ground Integration and Test of Distributed S/C Systems

The pre-launch integration and testing of large, distributed spacecraft systems like SI is a challenging problem, since the entire observatory cannot be "deployed and utilized" on the ground when the sparse primary array is designed to have a diameter ranging from 100 to 1000 m in diameter and its beam-combining hub be 1 – 10 km distant in normal operations. Since it is not possible to reproduce the precision formation flying and operation of such a system on the ground, the pre-launch testing will have to consist of robust component-level testing, coupled with very sophisticated computer modeling of system performance to validate them for flight. This is a problem common to all future distributed spacecraft systems – a good solution to it will benefit a large cross-section of missions. Further discussion of these problems and some solutions are given in Chapter 6 (Operations Assurance).

3.2 Key Technology Risks and Uncertainties

The Key technology needs and risks were highlighted at the beginning of the chapter as part of Table 3.1. The need for these technologies and how they are used in our baseline design for SI is described in Chapter 2. The technologies themselves are described in section 3.1, along with a number of other easier, though still challenging, technical development needs. We will discuss in the following section (3.3) a development roadmap. In this section, we first briefly highlight the risks and uncertainties in the development of these technologies.

Probably the tallest pole among all these technologies is the precision formation flying of as many as 33 distinct spacecraft: 30 mirrorsats, 1-2 beam-combining hubs, and possibly a reference spacecraft for metrology and aspect control. This is a complicated, multi-stage controls problem. However, similar control systems will be needed for many future missions (e.g. at some level, all missions composed of distributed spacecraft flying in a formation with tight constraints), so there is a great deal of motivation for such development. The biggest risk at the moment is the lack of a well-defined sequence of intermediate demonstration missions – with the cancellation of STARLIGHT, only SMART-3 and, possibly, ST9, are currently under consideration for flight prior to attempts at flying the large strategic missions like TPF-I, SI, LF, etc. We propose to develop an SI Pathfinder mission to both fill-in this development "hole", as well as to prove other technologies such as UV beam-combination and pursue intermediate science goals as well – but even more could and should be done.

Precision metrology over the long baselines required in interferometric missions like SI needs further development. Efforts are underway at JPL and SAO, but there is no assurance they will be continued as long as needed and to the fine levels required in the current long-term plan.

Wavefront sensing and control, based on feedback from the science data stream, especially in the context of a very sparse aperture imaging system needs continued long-term work. The Fizeau Interferometer Testbed (FIT) is exploring this technology now with 7 elements and has plans to expand to as many as 20 elements, but it is a small effort that needs to be expanded to fully develop the needed algorithms and control laws. And it needs eventually to be integrated with a formation flying testbed, such as the FFTB (GSFC) or the SPHERES (MIT) experiment to develop and prove the staged-control laws needed to cover the full dynamic range from km's to m's to cm's to nm's.

Finally, one of the most challenging technology needs for SI and all large, distributed spacecraft missions: how does one test and validate on the ground, prior to flight a system whose components are numerous (~ 30) and whose separations are order of 100's of meters to many kilometers? A critical need for, e.g., TPF-I, MAXIM (BHI), LF, and PI, in addition to SI.

In the following section we discuss a Roadmap for getting from today's state of the art to that needed to support these technologies.

3.3 Development Roadmap, with Alternative Approaches

3.3.1 Overview

The successful design and construction of the Stellar Imager will rely on the development and validation of a number of critical technologies highlighted in the preceding sections. These include, for example, precision formation flying, coarse ranging and array alignment, high-precision metrology, on-board autonomous computing and control systems, and closed-loop optical control to maintain array alignment based on the science data, along with a host of additional, somewhat easier challenges. A high-level technology roadmap for these items is given in Table 3.8.

Table 3.8: Technology Roadmap for the Stellar Imager

Technology Needed by SI	Development Plan and/or Candidate Technologies	Readiness Date
Wavefront Sensing and Control	Phase Diverse Testbed (PDT), Fizeau Interferometry Testbed (FIT), Wavefront Control Testbed (WCT)	2004 2007 2007
Closed-loop optical path control	Phase Diverse Testbed (PDT), Fizeau Interferometry Testbed (FIT)	2004 2006
Mass-production of spacecraft (SI "mirrorsats")	TBD (but see BATC approach in section 3.18)	2007?
Lightweight, UV-quality mirrors with km-long radii of curvature	Chen (2002), etc.	2007
Large format energy-resolving UV detectors with resolution >100	TBD – but driven by many missions	2008?
Methodologies for combining 20-30 simultaneous beams	Ground-based interferometers, FIT	2006?
Variable, non-condensing micro-newton thrusters	FEEPs, etc.	2007?
Precision Formation Flying	GSFC Distributed Space Systems Roadmap (Figure 3.20)	2009 SI PF, 2013 full-SI
Aspect Control to 10's of micro-arcsecs	Trade external metrology vs. wavefront sen.	2013
Precision Metrology over long baselines	JPL & SAO metrology labs	2010
Methodologies/control processes for deployment and initial positioning of elements in large formations	GSFC Distributed Space Systems Roadmap (Figure 3.20)	2013

Study of these technologies is ongoing at NASA/GSFC, JPL, SAO, various universities, and in industry, and significant leveraging and cross-fertilization will occur across projects, e.g. with JWST, TPF, and LISA. A series of testbeds are in operation or are under development at GSFC, including the: Wavefront Control Testbed (WCT) to study image-based optical control methods for JWST, Phase Diverse Testbed (PDT) to study extended scene phase diversity optical control with moving array elements, Wide-Field Imaging Interferometry Testbed (WIIT) to study extending the field of Michelson imaging interferometers, and the Fizeau Interferometry Testbed (FIT) to study closed-loop control of an array of elements, as well as assess and refine technical requirements on hardware, control, and imaging algorithms. Studies of the full *SI* mission as well as pathfinder concepts continue in

GSFC's Integrated Design Center and Metrology Testbeds are under development at SAO and at JPL (Lay, 2003). We present in Figure 3.22 a graphical representation of flow of technology development and mission capabilities for space-based interferometric facilities, from ground-based testbeds and operational interferometers to space missions that will logically precede and follow SI.

One of the more interesting technology options that is being pursued is an investigation of how much of the measurement and control job (of the various spacecraft and mirror surfaces in the distributed system) can be done purely by "external" (to the science data stream) metrology using, for example, lasers and at what point, and if, it will be necessary to handoff the measurement and control job to a system based on feedback from analysis of the science data stream. Our "baseline" mission concept in fact assumes that the external metrology system has measurement and command authority down to ~ the millimeter or micron level and that a "closed-loop" optical control system, based on phase diversity analysis of the science data stream, takes over at smaller scales to obtain control down to the nanometer level. The exact point at which that handoff occurs in the multi-stage control system is one of the interesting points still to be resolved. Our technology development plan is based on pushing both technologies to their limits, i.e., driving the external metrology to the smallest attainable scales (effectively testing in the process if we can do the "entire job" this way) and driving the development of the wavefront sensing & control to the largest possible scales, in the hope that the two systems will in the end at least have a significant region of overlap in their control authority.

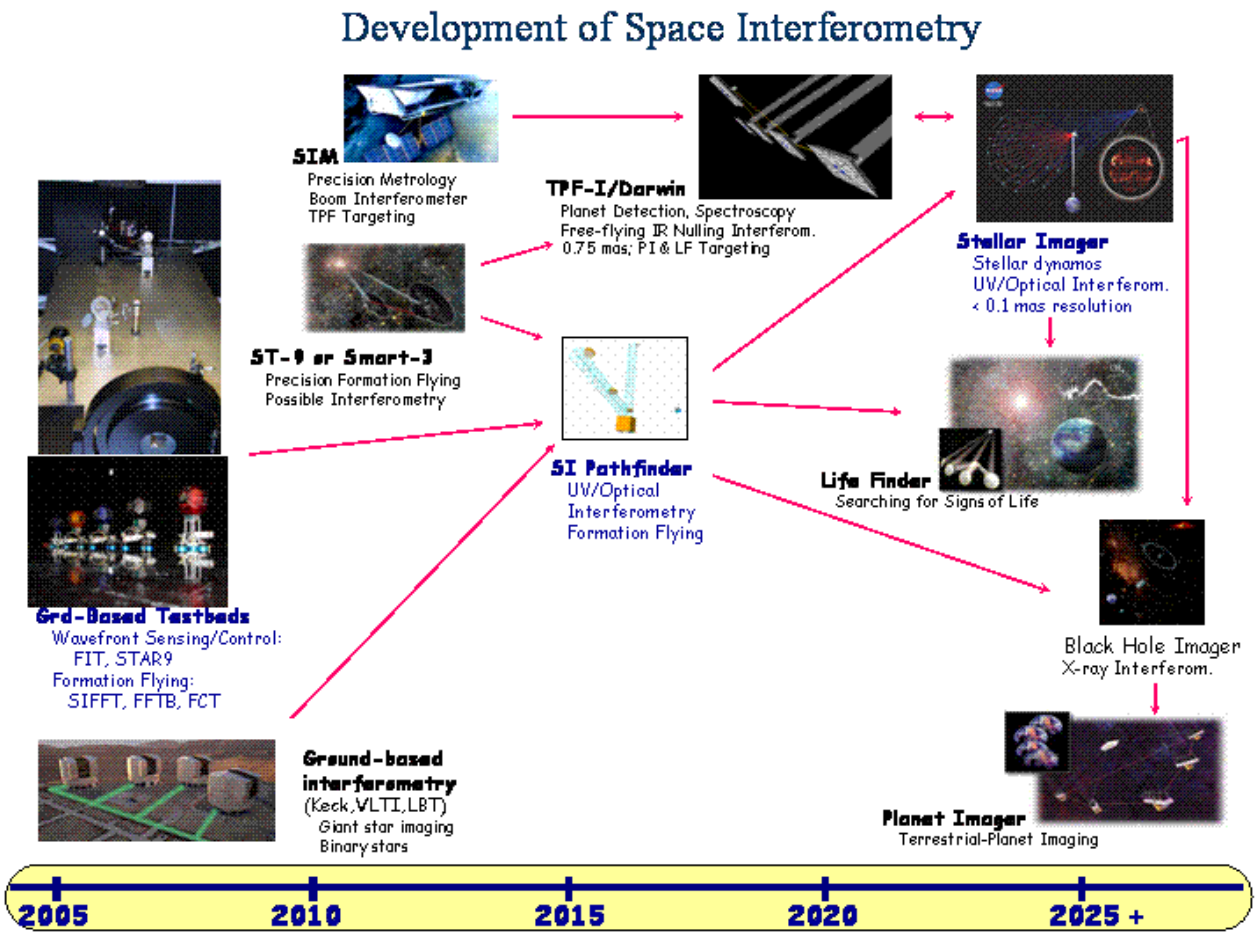


Figure 3.22: A Roadmap for the development of Space Interferometry

In the next several sub-sections we discuss in further detail existing plans and progress in the most critical and difficult areas mentioned above, including precision formation flying, closed-loop control of the complex multi-component optical systems based on wavefront sensing, and precision metrology. Finally, we discuss important synergies with other missions and facilities that will help to develop the needed technologies.

3.3.2 Precision Formation Flying

Precision Formation Flying system technology is critical for a broad range of future NASA Space Science missions, including the Terrestrial Planet Finder-Interferometer (TPF-I), the Micro-Arcsecond X-ray Imaging Mission (MAXIM)/Black Hole Imager, the Submillimeter Probe of the Evolution of the Cosmic Structure (SPECS), Life Finder (LF), and Planet Imager (PI), to name a few. Furthermore, the technology development occurring for other formation flying and distributed spacecraft missions, including Magnetospheric Multi-Scale, Constellation-X, and the Laser Interferometer Space Antenna, feeds into the SI technology roadmap as well. A high-level view of how the some of the technologies from these missions and science areas feed into the SI development process, as well as a collection of requirements is shown in Figures 3.20 - 3.22.

3.3.3.1 Technology needs for formation flying

Recently, under the context of the New Millennium Program ST-9, the formation flying community was brought together to identify the critical capability requirements for precision formation flying in the Space Science Enterprise. The following roadmap (Figure 3.23) was developed by a cross-cutting team consisting of GSFC, JPL, AFRL, NRL, industry, and academia, to represent the progression of capabilities needed vs. rolled up formation flying capability (expressed in terms of formation control precision requirements). "Relative navigation" requirements in this table include the "end-to-end" requirements all the way down to the measurement of the mirror surfaces, i.e., not just of the gross spacecraft positions. Some of the associated relative control requirements may therefore be accommodated through the motion of actuated mirrors and not solely by spacecraft maneuvers.

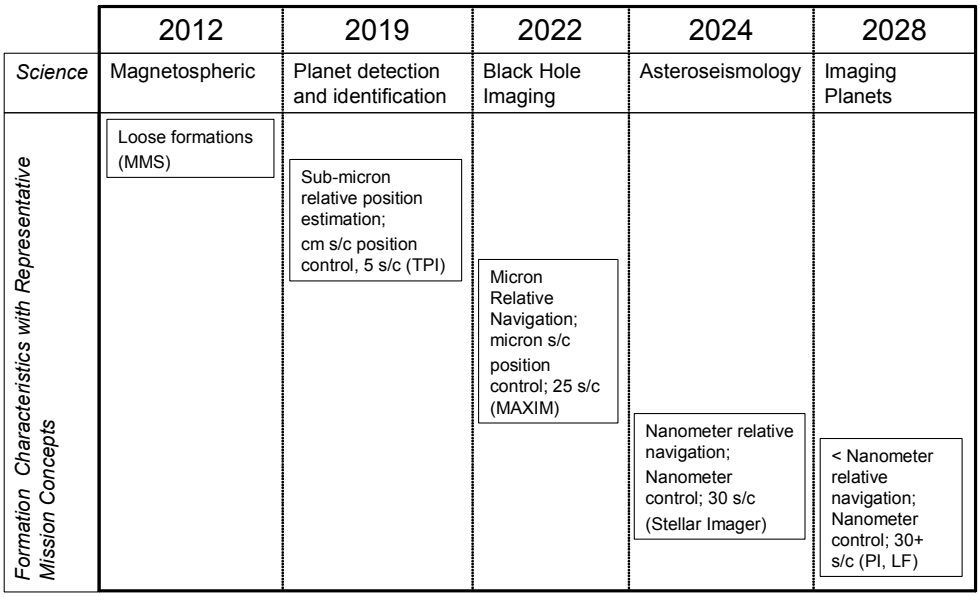


Figure 3.23: Science Capability/Formation Flying Capability Progression

3.3.3.2 Key Technology Challenges

The following represents some key technology challenges for formation flying and, in general, distributed spacecraft, future capabilities.

- Centimeter to nanometer control over S/C separations ranging from meters to 1000s of kilometers
- Affordable sensing at high altitudes (above GPS constellation, libration points)
- Affordable spacecraft components
- Control of formation line-of-sight to sub micro-arcsecond
- Coordinated Orbit/Attitude control of multiple spacecraft
- Initialization of multi-spacecraft fleets: collision avoidance and accurate placement
- Autonomous operations for formations and constellations; extreme challenge is a mission consisting of 100's to 1000's of satellites
- Multiple spacecraft deployment systems : deployerships and release mechanisms
- Affordable and robust interspacecraft communications
- Modeling, simulation and testbed infrastructure

Table 3.9 summarizes the current state-of-the-art, a near-term desirable demonstration level, proposed to be demonstrated in ST9, the Stellar Imager requirement, and the definition of what would constitute TRL (Technology Readiness Level)-5, i.e., readiness for flight-demonstration.

Figure 3.9: Formation flying requirements for Stellar Imager

	Figure of Merit				
Required Capability	Now	ST9	Stellar Imager	Current TRL	TRL 5 Test Requirement
Number of Satellites	2 S/C, non-collaborative (LS-7/EO-1)	4 desired 2 minimum	30	For constellations, 9 For formations, 6	Distributed simulation environment
Measure relative position	2 cm postproc (over 20,000 km measurement to GPS transmitter)	< 2 cm on-board, real-time	< 1 nm on-board	2 cm: 6 < cm: 4	RF or optical channel simulator with high fidelity dynamic simulator and real-time estimation
Measure S/C-S/C bearing angles (combination of relative attitude & 3 axis position)	N/A	1 am	~1 mas	4	HW prototype integrated into high fidelity simulation, with real-time estimation
Control relative position through comm. link	Rendez/Docking, < 1m short range	10 cm	1 nm (combined optics/S/C)	4	RF or optical channel simulator with high fidelity dynamic simulator and real-time estimation and control loops wrapped around.
Control S/C-S/C bearing angle	N/A	5 am	~10 mas	2	HW prototype integrated into high fidelity simulation, with real-time estimation and control loops wrapped around
Formation line-of-sight Control	N/A	N/A	50 μas	1	Interferometric verification
Inter-S/C Communication Rate	300 Mbps TDRSS	10-1,000 Kbps < 20 W, 20 kg	~1-10 Mbps	6	Testing of low power lightweight device through RF or optical channel simulator
Constellation Operating Range	1 km	100m - 1km	1-500 km	N/A	Channel simulator
Formation Commanding	On-board, one spacecraft relative to other	On-Board, collaborative	On-Board, collaborative	4	Distributed simulation
Autonomous collision avoidance	N	Y	Y	4	High-fidelity simulation
Precision of time synchronization	3 ns GPS, on-board real-time	< 1 μs	~1 ps	9	Time transfer simulator with GPS or other accurate clock

An overall Formation Flying Technology Development Roadmap is shown in Figure 3.24.

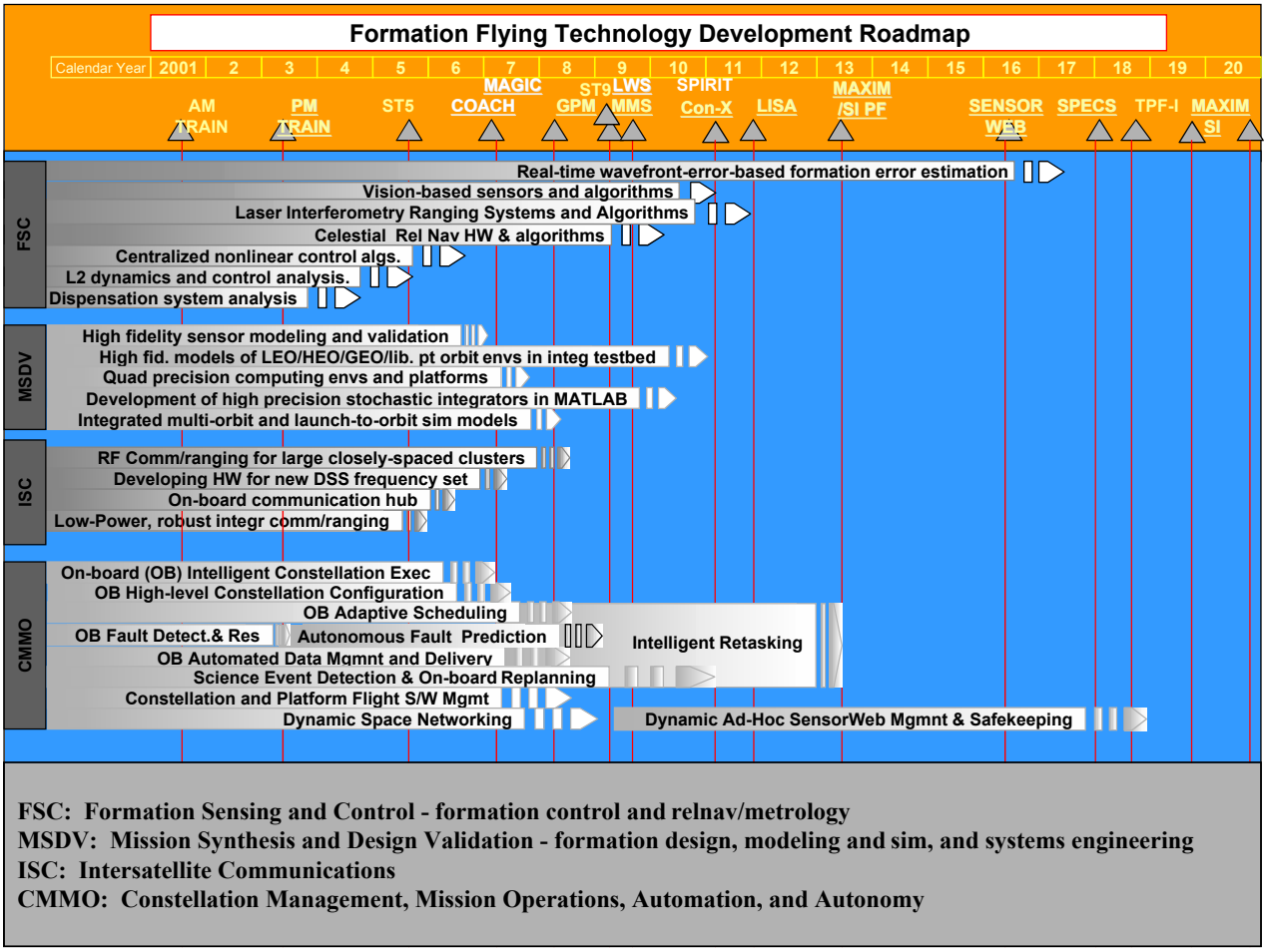


Figure 3.24: Formation flying technology roadmap

3.3.3 Closed-Loop nm-Level Optical Pathlength Control via Wavefront Sensing

In any many-element sparse-aperture telescope or Fizeau interferometer, the most critical concern is how to get the light beams from the various mirror elements into phase and how to maintain that alignment over time. At the highest level, this requires that the optical path lengths from the celestial target to each mirror and onward (perhaps, but not necessarily, via numerous intervening reflections) to the final detector be identical to within about 1/10 of a wavelength. In the case of SI, the most stringent requirement is from the shortest wavelength of the planned observational capability, i.e., the wavelength of the CIV doublet at 150 nm, which produces a requirement that the optical paths be held identical to within about 5 nm. It is unclear at this time whether external (to the science data path) metrology and control systems will be able, in the time frame under consideration, to attain this precision by themselves. It is therefore necessary and prudent to pursue alternatives for the finest level of control (down to the nm level). The best candidate for such an alternative requires the use of the actual science data itself (or light that is somehow sent through the same optical paths). Preliminary

studies of SI concepts have envisioned a hand-off from the external (perhaps based on laser ranging) metrology systems to systems based on analysis of the actual science data stream to get from the cm/mm to the nm level of control.

Optical image-analysis methods, such as phase diversity (Lyon et al. 2004a, 2004b), exist that theoretically are capable of determining the errors in the locations and attitude of the mirror array elements from numerical analysis of the distorted image created by the combined beams. The output from such an analysis can then be used to correct the positions (tip, tilt, and piston) of the individual mirrors to improve and maintain the image quality. It is important to demonstrate that these theoretical capabilities will work in the real world since they are critical to the eventual success of sparse-aperture systems.

Goddard Space Flight Center (GSFC), the University of Maryland, the Naval Research Laboratory, and Sigma Space Corporation have collaboratively designed and constructed the Fizeau Interferometer Testbed (FIT) at NASA/GSFC. It will be used to explore the principles of and requirements for the *Stellar Imager* mission concept and other Fizeau Interferometers/Sparse Aperture Telescope missions. FIT utilizes a large number of truly separate, articulated apertures (each with 5 degrees of freedom: tip, tilt, piston, and 2D translation of array elements) in a sparse distribution. It has the long-term goal of demonstrating closed-loop control of articulated mirrors and the overall system to keep beams in phase and optimize imaging. FIT also enables critical assessment of various image reconstruction algorithms (phase diversity, clean, MEM, etc.) for utility and accuracy by application to real data. FIT Phase I (7 primary mirror elements) is now in operation and has been successful in demonstrating that such a system can be fabricated, aligned and stably controlled in a laboratory environment. The main components of the current FIT are schematically illustrated in Figure 3.25.

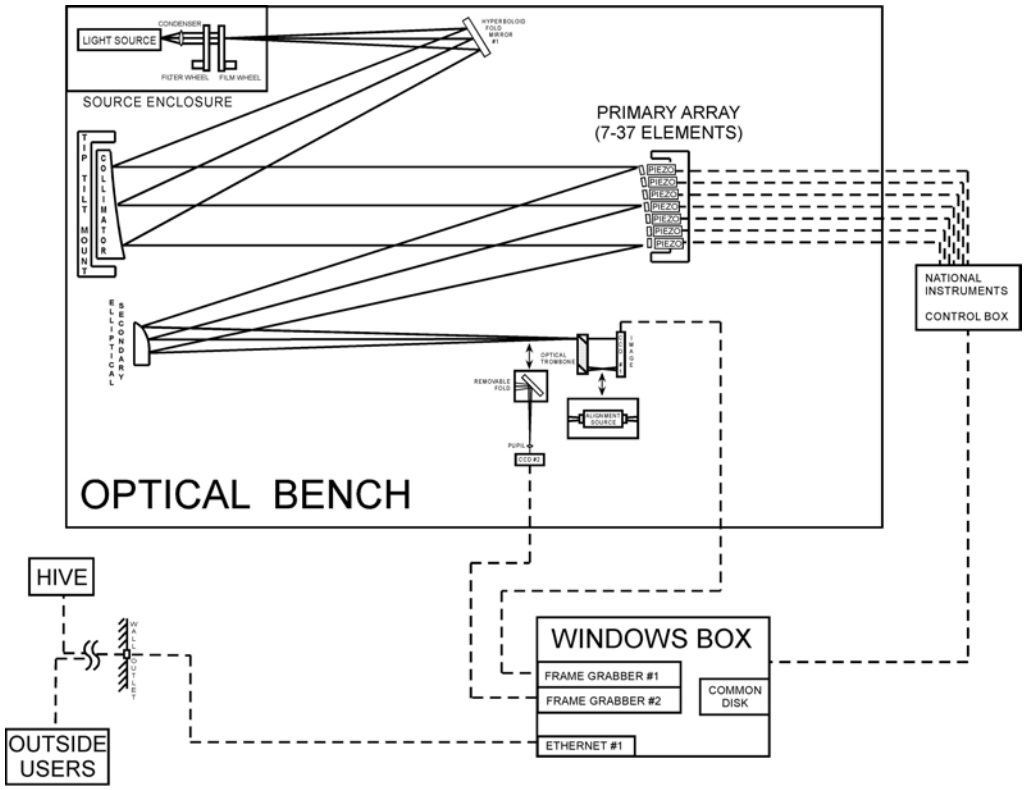


Figure 3.25: A sketch of the Fizeau testbed layout.

FIT consists of source module with both laser and thermal sources, and with narrow band and medium band filters and extended scene images. These extended scenes are imaged through the system with off-axis conic optics through to the focal planes. One of the mirrors in the relay is the aperture plate consisting currently of 7 (to be expanded to ≥20 in Phase II) piston, tip, tilt articulated mirror elements. From two images of the extended scene, at different focii and wavelengths, we have demonstrated that the wavefront, and hence the misalignments, can be sensed and the actuators moved to compensate. Figure 3.26 shows some of the FIT hardware (baffles have been removed to improve visibility of the main elements). Further information on FIT is presented, along with early results, in 3 papers from the 2004 SPIE meeting (Mazzuca et al 2004, Lyon et al. 2004a, Petrone et al. 2004) and in two IEEE papers (Zhang et al. 2003, Lyon et al. 2004b). Phase II of the FIT is expected to go into operation in 2006 and will enable test of realistically-sized arrays (in terms of # of elements) for the first time.

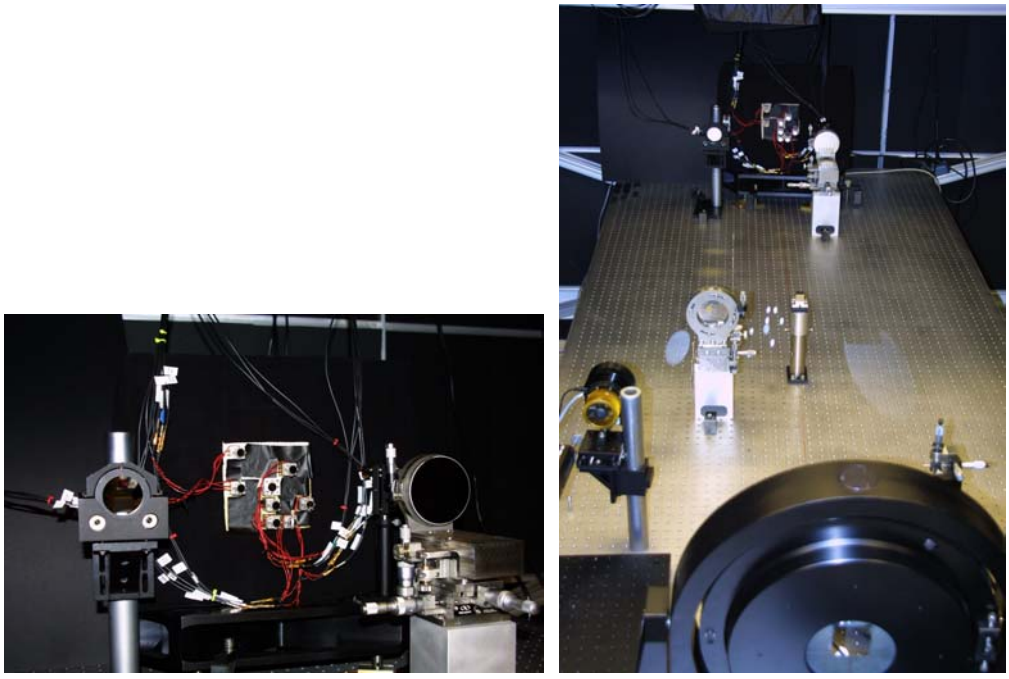


Figure 3.26: Left: Photo of the Phase I primary array and 2 secondary mirrors.
Right: An overview of the Phase I FIT, with baffles removed to show the optical elements clearly.

3.3.4 Precision Metrology Over Long Baselines

As discussed in Sec. 3.3.3, advances in astronomy and cosmology in the coming decades will require dramatic improvements in imaging resolution at a wide variety of wavelengths. These will be realized in a number of ambitious missions exemplified by those in Fig. 3.22. Interferometric range sensing with picometer accuracy will be required, between spacecraft separated by km distances. Missions employing dozens or hundreds of distance gauges are envisioned.

Current capabilities must be enhanced in accuracy, range, speed of operation, and slew rate. The technology must be made ready for use in space, employing suitable components and demonstrating reliability. In addition, particularly for missions considering many distance gauges, a substantial reduction of the cost per gauge and improvements in modularity and simplification of setup will enable some missions, and for others in the conceptual design stage, will provide options.

The SAO Metrology Testbed is based on the successful development of the Tracking Frequency Gauge (TFG) [Fig. 3.27]. It will demonstrate metrology over long paths, 3-D metrology, a compact multi-beam launcher, and metrology of sub-nm precision at low cost.

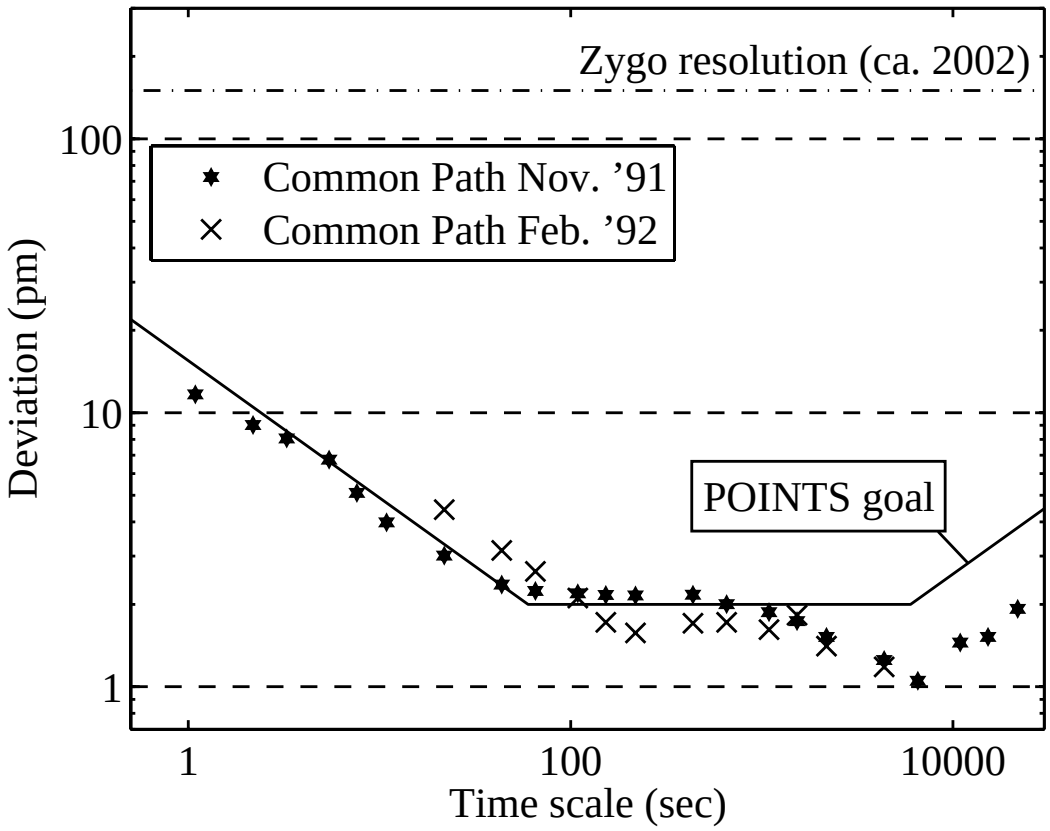


Figure 3.27: TFG incremental distance accuracy vs. averaging time.

The TFG was developed following an investigation of available laser gauges for POINTS [Reasenberg 1988]. The investigation found that none of the available approaches could meet the picometer requirements of POINTS. High-resolution imaging missions in many cases have nm OPD requirements, but metrology requirements similar to POINTS. This is because substantial structural mass, and therefore cost, can be saved by employing a pair of relatively closely-spaced pm laser gauges to determine the transverse position of an object to nm accuracy, and because of the division of error allocations among several sources.

The classical precision laser gauges employ two beams and a heterodyne principle. Imperfect separation of the beams leads to nm-scale cyclic bias. The bias has been reduced by dithering and averaging over the cycle, but this limits the readout rate. The bias has also been reduced without averaging, replacing polarization with the spatial separation of beams, and with a combination of polarization and spatial separation. These methods involve trades of performance against complexity.

All laser gauges employ a measurement interferometer tied to the distance to be measured. The TFG is based on the technique of locking a laser to the measurement interferometer, and measuring the optical frequency by heterodyne against a stable laser. For changes too great for the available range (which is often no more than several wavelengths), the laser hops from one fringe to another. We have demonstrated a TFG that can hop as often as every 20 μs.

The TFG improves upon the classical heterodyne gauge in several ways:

- it is intrinsically free of cyclic bias.
- the TFG employs only a single optical beam, simplifying the optics of the measurement interferometer, and making possible a compact multi-beam launcher.
- the new version of the TFG can be built with only a single optical component in addition to the measurement interferometer. This component is a semiconductor distributed feedback (DFB) laser, an item produced in quantity for telecommunications. For telecommunications, DFB lasers must meet performance requirements over wide variations of environmental conditions similar to those required of space electronics. Semiconductor DFB lasers have been used for atomic spectroscopy for more than a decade [Wieman 1991]. This TFG version is termed the SL-TFG.
- the TFG can be built with all components connected by optical fiber, and no free-space path (except in the measurement interferometer). Current implementations of the heterodyne gauge employ acousto-optic frequency shifters, which require a free-space path.
- since the TFG design involves fringe hopping, absolute distance can be measured with little or no additional hardware.
- the TFG's hops are agile (the time scale of the DFB laser is ns), improving the accuracy of absolute distance measurements.
- the TFG solves a problem due to polarization rotation in retroreflectors (distinct from the cyclic bias) that is also of nm scale.
- since the TFG can operate in a resonant optical cavity, alignment sensing is facilitated and alignment errors suppressed.

3.3.5 Synergies with Other Missions and Facilities

There will be considerable benefit from technology development for both ground-based facilities and space-based missions planned prior to the launch of SI. Table 3.10 summarizes various space-based missions and/or studies that have the potential to contribute significantly to development of technologies relevant to SI and/or its Pathfinder mission.

Table 3.10*:* Technologies *of Space-based Precursor Missions*

Mission	Description	Enabled Technologies
StarLight mission studies	formation flying out to 1 km with 2-spacecraft interferometry out to 200 meter baselines	precision formation flying, separated spacecraft interferometry
SIM	optical interferometer with baseline 10 meters with 2 telescopes; imaging resolution: 10 milliarcsec; launch 2011	precision picometer-level metrology, precision actuators, control laws
TPF-I	IR nulling interferometer option ; 45-135 meter baseline; 4 collector spacecraft; one combiner spacecraft	precision formation flying and metrology; launch 2020
LISA-Pathfinder (formerly SMART-2)	2-spacecraft laser interferometer flying in formation; launch 2007	formation flying sensors, thrusters, formation flying control, optical components (achromatic phase shifters, delay lines, spatial filters, amplitude/polarization matching units), and integrated optics
LISA	3-spacecraft laser interferometer flying 5 million km apart in orbits around the Sun; each satellite will be controlled to the nanometer level; launch 2014	laser-based metrology and control; based on SMART-2 development
Darwin	8-spacecraft IR interferometer at L2; launch 2015; ESA equivalent to NASA interferometric TPF concept	radio-frequency ranging system to determine relative positions of the spacecraft to within 1 cm, laser-based metrology and control to within 10's of nanometers; based on SMART-2 development
ST-9 – Formation Flying Option	4 satellites; less than 2 cm relative position measurement real-time; measure spacecraft to spacecraft bearing angles to 1 arcmin; control relative position through communications link to 10 cm; control spacecraft to spacecraft bearing angles to 5 arcmin; constellation operating range: 100m - 1km; autonomous collision avoidance	precision formation flying

Similarly there are a number of existing and future ground-based interferometric facilities that will be developing relevant technologies, as described below in Table 3.11. MRO in particular should be useful, since it is the closest in design to that of SI and actually uses Fizeau beam combination, unlike most of the others.

Table 3.11: *Technologies of Ground-based Precursor Missions*

Mission	Description	Enabled Technologies
COAST	Optical Michelson interferometry; baselines up to 100m to give images with a resolution down to 1 milliarcsecond	high resolution imaging, multiple-beam combination
NPOI	Optical interferometer with imaging subarray baseline lengths from 2.0 to 437 m; laser metrology system from 19 m to 38 m; best angular resolution of the imaging subarray is 200 microarcsec	fast delay lines; close phase demonstration; visible light beam combination; synthesis imaging
CHARA	Optical/IR interferometer; 200 micro-arcsecond resolution; maximum baseline of 330 meters	visible light synthesis imaging; optical path length equalization
KI	Optical/IR interferometer, expanding from 2 to 6 elements; narrow-angle differential astrometry with a precision of 30 microarcseconds	metrology systems, control systems; path length equalization
MRO	Optical/IR 10-element interferometer; completion in 2008	aperture synthesis imaging with many elements; closest to SI in concept of any ground facility
VLTI	IR 7-element interferometer; maximum baseline of large unit telescopes is 130m, and 200m for auxiliary telescopes; best resolution for large telescopes is 1.5 milliarcsec and 1 milliarcsec for aux telescopes	high resolution imaging
LBT	IR 2-element interferometer; 2 8-m primaries, each F/1.142; 22.8 m baseline	full coverage of Fourier UV-plane

3.4 Validation and/or Demonstration Approach

The main special challenges in deploying and operating a complex formation flying interferometer successfully are likely in two areas: formation control and beam control. Both of these technologies can be demonstrated in obvious ways on a simpler SI Precursor mission with 2 or more spacecraft, interferometric beam combination, and laser metrology. But the ground validation of both the Precursor and the full-up SI system requires even greater creativity.

3.4.1 Formation Control Validation

The key requirement on SI formation control is to maintain absolute positional knowledge adequate to reduce OPD (Optical Path Difference) estimation error sufficiently to acquire stellar fringes in a reasonable time. The white light fringe condition is a perfect match of absolute travel distance to a detector via each pair of separate optical paths. Since the white light fringe signal is ~1 micron wide in OPD, the Hub spacecraft ideally needs absolute positional knowledge of the ballistically-drifting Mirrorsats to this precision; this is not available initially. Therefore the Mirrorsats must be navigated as close as possible to their initial desired positions using the range and angle measurements from the formation control and metrology systems. At the beginning of the fringe acquisition, metrology has measured the angle between the starlight direction and the inter-satellite vectors with ~1 arcsec

accuracy, as well as measuring absolute range with an uncertainty of 1 cm. Improving the absolute range accuracy at short range provides considerable benefit, and is likely realistic given the improved signal strength at close range. A maximum OPD scan rate of 10 μm/sec is limited by reliable fringe detection from dim stars. Thus, the estimated OPD range can be scanned within the desired 1000 sec search time. Although absolute positional knowledge sufficiently reduces OPD uncertainty so as to acquire fringes in a reasonable time, the system requires 1) careful calibrations to achieve this narrow OPD range, and 2) a methodical means of searching through this range.

Testing and characterization of the SI formation control system should validate the performance of the system in its flight configuration and will prove proper operation of the GNC (guidance, navigation, and control) system using the formation control sensor inputs. The proposed tests address SI formation control performance early in the program, reducing schedule risk associated with system improvements identified by the tests. The tests require early formation control sensor system delivery to validate GNC performance using flight mockups before integration of the flight hardware into the actual S/C. These tests verify the formation control system, including inputs from the course guidance sensors such as star trackers, gyros, and formation control sensors. The GNC system will select which guidance sensors it responds to at which phases of acquisition and operation. The test program for the formation control system is broken down into several stages. The formation control transceiver test with S/C characterization characterizes the formation control antenna patterns for use by a formation control sensor simulator. This RF Simulator will be developed in the formation control/GNC Pre-Acceptance Test for use in later integrated system tests. Verification of this simulator is accomplished by the formation control/GNC Acceptance Test, an end-to-end test of the formation control and GNC controls and interfaces using the Hub and at least one Mirrorsat at a time. Following this early testing of the formation control system on S/C mockups, the flight formation control hardware will be integrated into the actual S/C. Testing of the formation control systems on the flight units will consist of an antenna characterization measurement and formation control RF simulator aliveness and functional tests.

3.4.2 Beam Control Validation

We can produce a beam-control test set-up with a long effective optical pathlength relevant to the SI mission in a few different ways. Doing the whole job inside a large test chamber such as is available at LM Sunnyvale, MSFC Huntsville, or Plum Brook near Cleveland is possible if we use a parallel-mirror multi-reflection pathlength extender. To do that with a modest size extender, the transmitted beam(s) must not be too broad – thus the optical magnification should be at least a factor of 10 (compressing a 1 meter beam diameter down to 10 cm or less). This will probably be what's needed for space operation as well. The consequences of such magnification include a considerable increase in the effects of mirror tip and tilt, and of beamwalk. The higher the magnification, the greater the decrease in beam quality from given levels of optical surface imperfections, alignment errors, and pointing instabilities. The net conclusion is that if we can make a high-magnification demonstration achieve required system performance levels in a test chamber, we can have considerable confidence in the actual performance of the real space system operating at more modest beam compression levels.

Some chambers will be large enough to test a few Mirrorsats and the Hub simultaneously in a rather complex set-up, with a test source feeding all of these with highly parallel star simulator beams. Another possible test setup is to connect two vacuum chambers with a long evacuated tube, to give a long, straight optical path between two spacecraft at each end.

A Metrology Acceptance Test will demonstrate acquisition and fringe tracking under the expected operating conditions. The metrology system will be verified over the operating range of 50 to 1000 m, despite the effects of Gaussian beam propagation, attenuation, and pointing errors. A vacuum test will also measure fringe intensity, which is indicative of optical alignments and wavefront. The final interferometer subsystem test is the End-to-End Optical Acceptance Test. This test verifies the optics and associated control systems that will be used in the Formation Observation Mode. This layout uses the Hub's internal optics and the metrology system between the S/C.

3.4.2.1 System Validation

Integrated formation control, beam control, and interferometry tests will be performed after the flight formation control and optical systems are integrated. The control systems are tested in circumstances similar to flight operation, in that the tests parallel the dominant control moving from the guidance system to the metrology and interferometer. The Metrology Acceptance Test demonstrates that the beam control system autonomously acquires the metrology system fringe lock, given the positional information provided by the guidance system. Once metrology lock is acquired, the control system is ready to transition its dominant inputs from the guidance systems to the metrology and interferometer. At this stage, the system (or a representative part) is moved into a thermal vacuum chamber for the final integrated test. The End-to-End Performance Test verifies that the interferometer can acquire white light fringes given a locked metrology system, completing the transition of dominant control input from the formation control system to the interferometer. Our sequence of subsystem and integrated tests verifies each system at an early stage of integration. The formation control system is tested in advance of the interferometer and before final S/C completion. The interferometer tests verify autonomous acquisition and maintenance of optical beams across the broad range of separations expected. The integrated tests demonstrate appropriate hand-off points in the control system, culminating in a full end-to-end vacuum performance test.

3.4.2.2 Basic Test Principles

For the SI mission, Test and Evaluation (T&E) is a broad set of activities that includes the physical and analytic testing of components, subsystems, or systems during Formulation and Implementation, including testing during hardware and software integration and verification. Effective T&E should balance the elements performed early in the design process to identify and correct problems with those elements later in the design process to verify that all system requirements have been met.

The prime contractor will develop a matrix of all tests required to assure that the mission requirements are being met. From this matrix, the contractor selects only those tests that make economic and final product verification sense. Early in Formulation, NASA in-house testing may be necessary to identify whether the system concept is verifiable. Later in Formulation, NASA may test to develop acceptance criteria to be used in evaluating the contracted system. In either case, test data usually will have more fidelity than predictive data, consequently the results obtained from any of this testing can be used to improve the systems design as well as update preliminary information with data that more accurately reflects operational use. For example, failure rates used in Life-Cycle Cost (LCC) models could be updated to reflect operational data obtained during test.

A summary of test types appropriate to the main program phases is listed here:

Formulation

- Concept Verification Testing
- Testing for Acceptance Criteria

Implementation

- Alternative Design Testing
- Functional Testing
- Environmental Testing
- Preliminary Operational Testing
- Turnover & Acceptance Testing
- System Wearout & Replacement Testing
- Operational Testing
- Maintainability Demonstration

3.4.3 Pathfinder Mission Concepts

The path to a large-baseline UV-optical interferometer in space such as *SI* is long and difficult. Hence, one or more pathfinder missions that take smaller technological steps and produce science results within a reasonable time-span are desirable. Existing useful precursor missions are limited: TPF-I will be a nulling, cryogenic interferometer operating in the infrared; SIM does not use the free-flying formations that will be needed for truly long-baseline facilities, and it will operate only at longer (optical) wavelengths and it will be used primarily as an astrometer, rather than an imager.

It would therefore be desirable to have a nearer-term mission (perhaps using booms, but free-flying if possible) with modest baselines (~ 20-50 m), a small number of primary elements (~ 3-5), decent size mirrors (~1 m) and the ability to perform ultraviolet beam combination and produce images in ultraviolet light. The small number of spacecraft/mirrors in this pathfinder mission would require frequent reconfigurations and therefore limit observations to targets whose variability does not preclude long integrations. However, such a mission would both test most of the technologies needed for the full mission, as well as be capable of producing a significant scientific return. A pathfinder with 20-50 m baselines could, for example, image the surfaces of the apparently larger stars, such as the red supergiant Betelgeuse and several long-period variables (e.g. Mira), as well as cataclysmic variables exhibiting mass-exchange between the components. The addition of high-resolution spectroscopy to such a mission could increase the science return even further at modest additional cost.

The Integrated Design Center (IDC) at GSFC has examined several possibilities in its Instrument Synthesis and Analysis Lab (ISAL) and Integrated Mission Design Center (IMDC). One of the more interesting designs is shown in Figure 3.28. It uses three one-meter mirrors (sections of a larger virtual parabola) on an 8.4 meter boom, a secondary mirror above them on an 18 meter mast, and a completely separate free-flying spacecraft with a single, deformable mirror, whose surface is adjustable to match the figure needed at a given location on the large virtual mirror surface, shown in Figure 3.29.

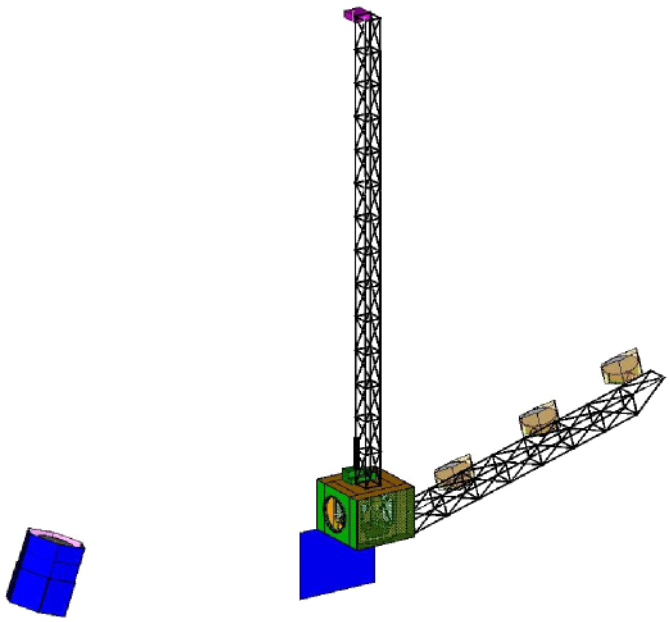


Figure 3.28: A concept for a UV/optical imaging interferometer pathfinder mission, utilizing 3 primary mirrors on a boom and a fourth on a free-flying spacecraft, and a secondary mirror on a mast.

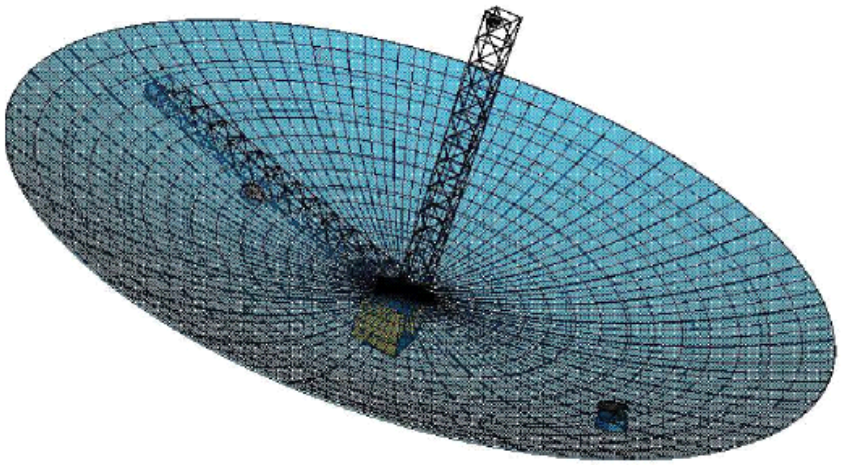


Figure 3.29: A view of the Pathfinder mission concept with the "virtual array" drawn in to show the surface along which the free-flyer moves and on which all the primary mirrors lie.

Additional details on this version of the pathfinder design:

This version of the pathfinder mission is a 4-element UV interferometer at L2, which comprises two rigid hinged booms and one free-flyer. Three one-meter mirrors are secured on an 8.4 meter boom, with a secondary mirror above them on an 18 meter mast. The free-flyer is a spacecraft with a single deformable mirror, whose surface is adjustable to match the figure needed at a given location on the large virtual mirror surface. The parent virtual primary mirror is f/0.5 in order for the mast and boom to be short and rigid. Therefore, the primary mirror actuators on the free-flyer spacecraft will have to adapt the mirror surface to new subaperture figures as the mirror is moved to different radial positions on the parent.

These high-level requirements include the following:

- At least a 20 meter baseline separation of the outer elements 1-m class UV/optical-quality mirrors
- Provide metrology to determine mirror locations to within about 20 nanometers of proper positions, and use science data on engineering focal plane in final control of the optics to within 5 nanometers for the optical path lengths of the beams
- Capability to observe in ~ 10 Å bandpasses in UV and in broad-band near-UV or optical
- Array reconfigurable for synthesis imaging and asteroseismology
- 2D detectors (512x512), either energy-resolving or CCD-like with filters
- Maintain thermal control precision of the boom/free flyer structure to within 0.1 C, and the optical benches to within 0.01 C.
- Lifetime is negotiable

A few significant issues need to be addressed in order to achieve a successful pathfinder mission as presented. Similar to the full SI mission, metrology and autonomous nanometer control of the mirror positions, formation flying, and Earth-based functional/performance verification are crucial. Metrology and autonomous nanometer-level mirror control span many of the subsystem components, including the following:

- *Boom Dynamics & Damping:* We are anticipating that missions like SIM, which will utilize booms, will have resolved most of the basic problems associated with stabilization of the boom.
- *Mirror Wavefront Control:* Wavefront control methodologies are currently being developed in the Fizeau Interferometer Testbed (FIT) at GSFC and in other testbeds such as the Wavefront Control Testbed (WCT).
- *Pointing:* Since the Pathfinder mission will operate with 25 times lower resolution, the constraints on aspect control become more relaxed, although they remain a significant challenge.

Formation flying capabilities are significantly simplified in this pathfinder mission since there will exist only one flyer. The advantage to this scenario, though, is the ability to still provide science if the free-flyer component fails. Targets will be constrained to slower moving ones since the boom-only scenario will require longer exposure times to fill in the UV-plane. Formation flying issues are also being addressed by SMART2 and perhaps ST-9.

Other Pathfinder concepts of interest include designs that are pure formation flyers, i.e. no booms involved. This approach would, of course, be more directly applicable to the full-up SI missions, which has no booms, and may in the end be no more difficult than a system utilizing booms (SIM is demonstrating that booms can create almost as many problems as they solve, relative to a pure formation-flying system – the "control problem" is of approximately the same level of complexity either way). A design similar to that shown above for the boom plus formation flyer option, but lacking the booms, could be used for a pure formation-flyer Pathfinder option.

The derivation of an optimal SI Pathfinder design will be the next step (post Vision Mission Study) in our overall SI development process.

Chapter 4: Deployment and Orbit/Formation Maintenance

Several launch concepts were examined by the IMDC and ISAL at GSFC.

The IMDC recommended consideration of two options, depending on whether one or two hubs were to be included in the initial deployment. Both options include a reference spacecraft for controlling pointing of the observatory without use of the target light, i.e. by using independent guide stars tracking by the reference spacecraft (mini-interferometer). These are shown in Figure 4.1.

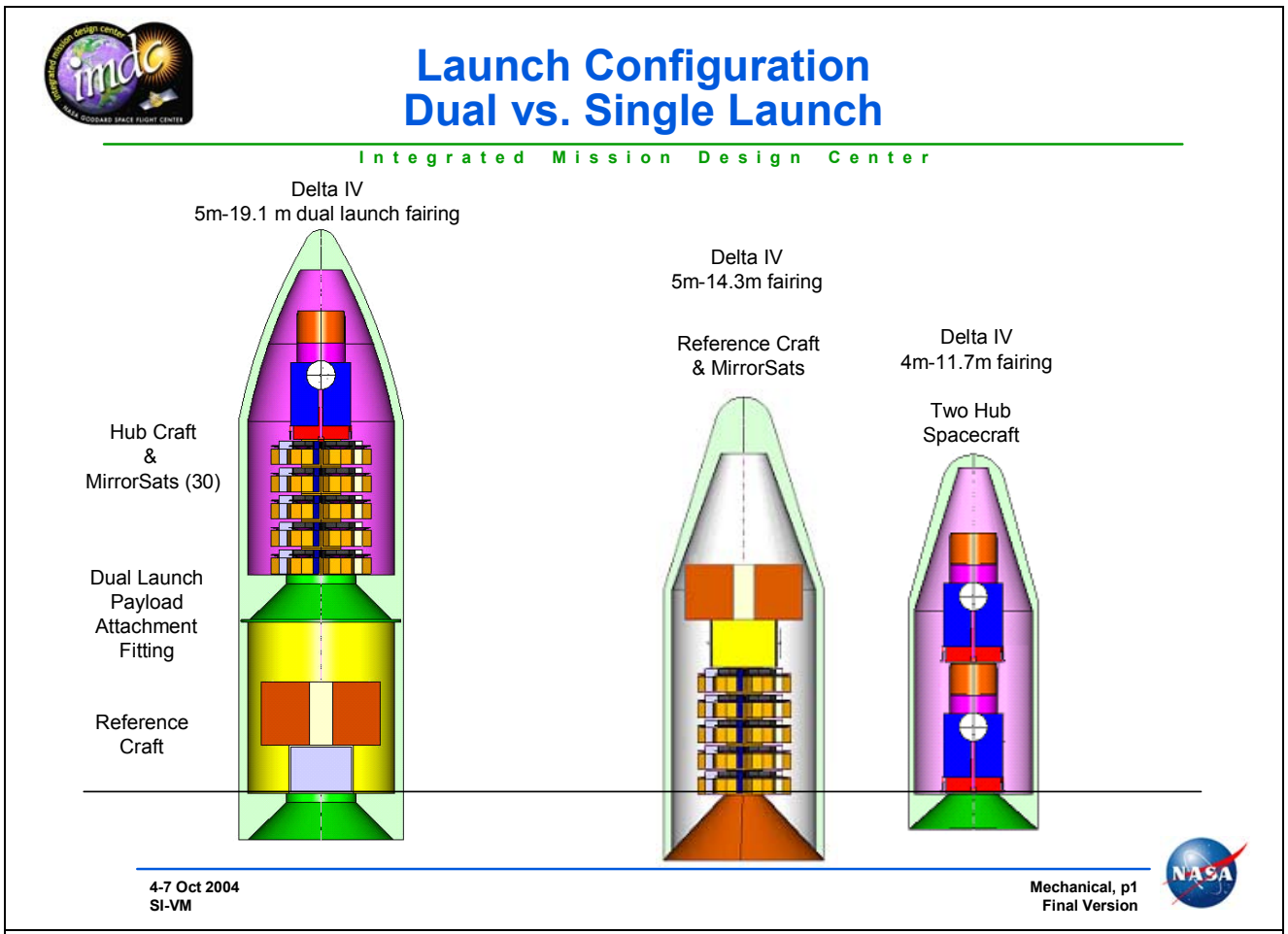


Figure 4.1: Two launch options for SI, a single Delta IV heavy vs. two Delta IV launches.

The ISAL launch scenario assumed the simpler case of a single Hub spacecraft plus 30 mirrorsats, but no reference spacecraft. In this design all tracking and guiding of SI is done based on target light detect within the Hub. This launch scenario is shown in Figure 4.2.

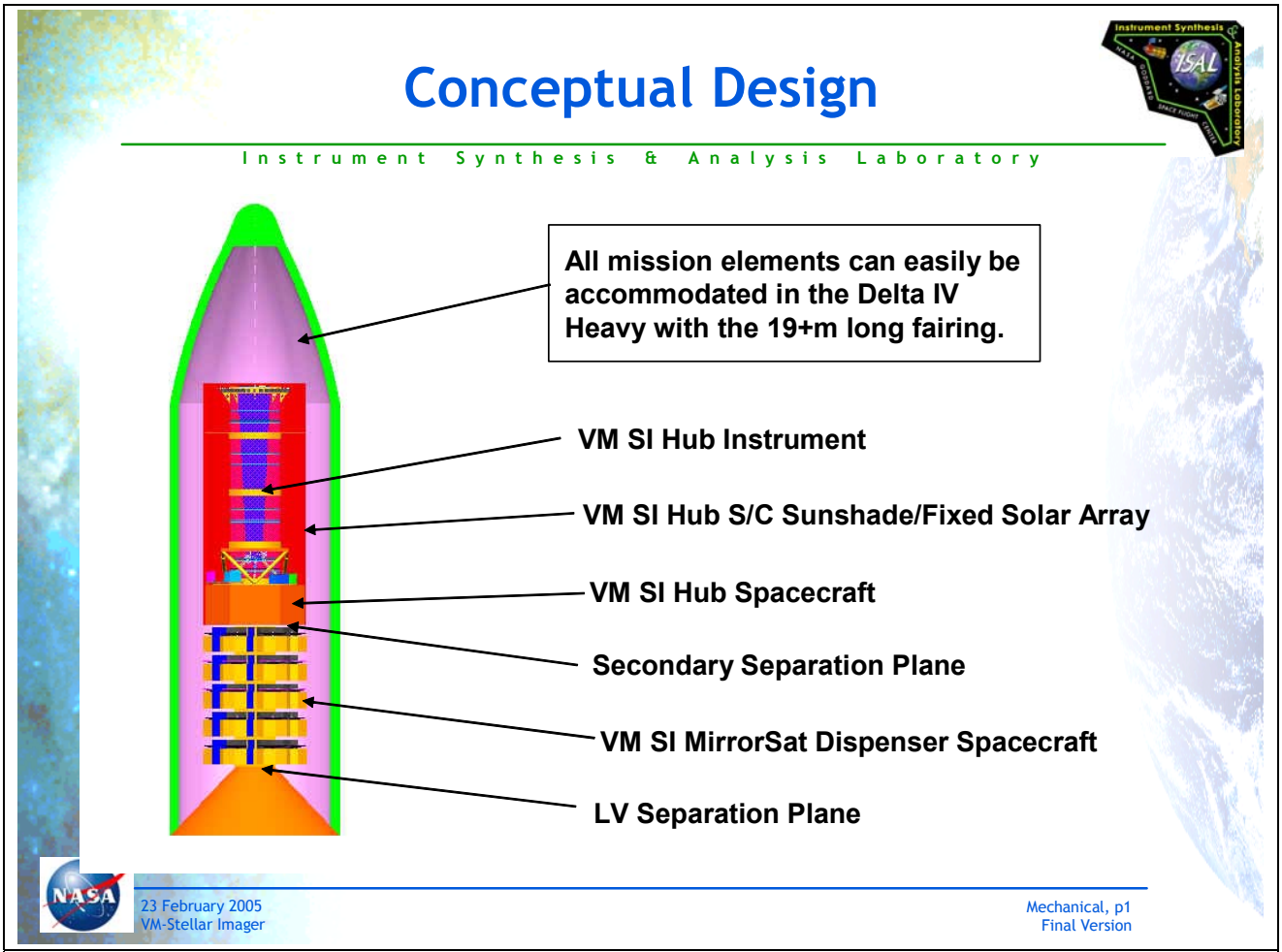


Figure 4.2: Launch concept for the "simpler" SI system with 1 Hub and no reference spacecraft.

4.1 Transportation to Operational Location

SI will be transferred to a Sun-Earth L2 libration orbit using a direct transfer trajectory shown in Figure 4.3. This type of transfer can be designed using a formulation of invariant manifolds (4.3a) that describes all the possible trajectories from the Earth parking orbit to that of the mission orbit. Using a large Lissajous or halo orbit (4-3b) as the mission orbit will either minimize or eliminate the need for any large insertion maneuver. The parking orbit is a generic low Earth orbit with orbit parameters of 185-km in altitude, eccentricity near zero, and an orbit inclination near 28.5 degrees if launched from the Eastern Test Range at Cape Canaveral, Florida. The parking orbit is normally restricted to less than one orbit period due to battery and power constraints. This orbit permits both long and short coast durations before the insertion from the parking orbit onto the transfer trajectory. The insertion maneuver, performed by the upper stage of the launch vehicle, is on the order of 3.14 km/s and represents an energy of approximately –0.7 km^2/s^2. This energy level is important in that it is used by the launch vehicle manufacture to determine the payload capacity into the transfer orbit. An estimate of the maximum payload mass for the launch vehicle can be found on the KSC launch vehicle web site.

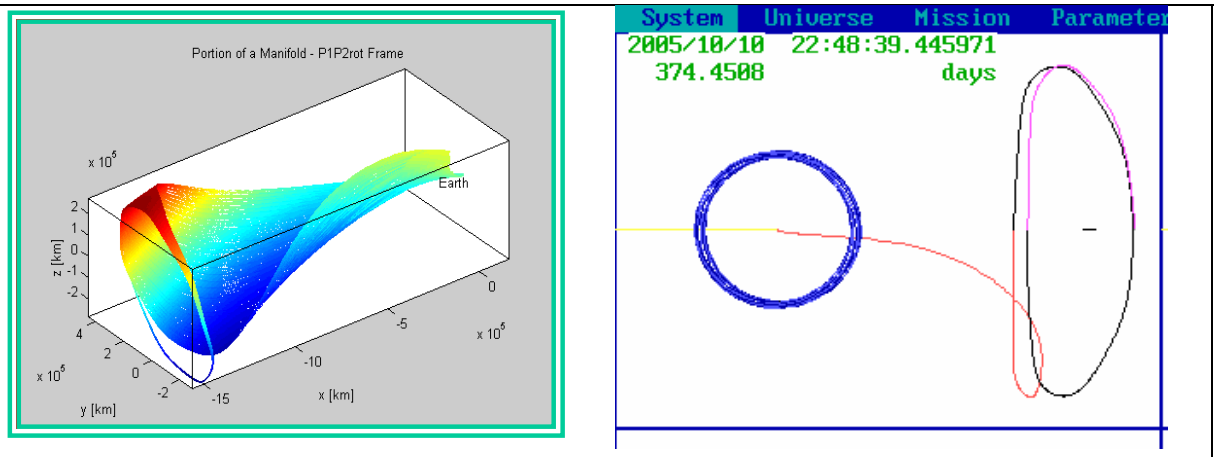


Figure 4.3: Sample Manifold Design (a) and Simulation of transfer trajectory (b)

Once launched, the spacecraft is placed into the proper orbit and the correct attitude is established for the insertion maneuver. After the insertion maneuver, the transfer trajectory enters a coast phase that takes approximately 120 days. During this coast phase, from approximately 12 hours after insertion onward, midcourse correction maneuvers will be performed to correct any insertion energy errors and misalignments in the insertion orbit parameters. These maneuvers are segmented to take out the majority of the insertion error and to target the mission orbit goals as un-modeled accelerations due to environmental perturbations and attitude re-orientation effects on the estimated area to mass ratio will need to be corrected. Some of these maneuvers may be designed to allow a multiple day launch window. Upon arrival at the mission orbit, an insertion maneuver will be performed to balance the energy, allowing the spacecraft to be placed on the reference libration orbit. The size and orientation of the mission orbit for SI is not critical, therefore the maneuver (Delta-V) budget can be minimized for the mission lifetime. During this coast phase, routine orbit determination (navigation) will begin. The orbit determination accuracy is dependent upon the number of and duration of the tracking passes. These tracking passes use S, K, or X Band Doppler and range measurements as input into the orbit determination process. Convergence to a solution will take days to weeks and is dependent on the position and velocity with respect to the ground station in the orbit. For example, a two-week tracking arc is typical for the mission orbit while only 12 hours are needed during the early coast phase when the velocity is directed radially away from the Earth.

4.2 Deployment

SI will be transferred to the mission orbit as one entity. Upon arrival and insertion into the mission orbit, a deployment of the components will begin. This is a critical event as not only are the components maneuvered into their proper location, but also collision avoidance must be performed. This means that the relative navigation system and individual propulsion systems must be operating. The relative drift of the components will be in predictable directions, as the components will follow their own orbits and drift in patterns that are determined by the natural dynamics of the Sun-Earth libration region. Figure 4.4 shows one prediction of the drift relative to a central point on a reference orbit when the components are relatively close. See spacecraft bus section (3.1.4.3) for additional discussion on deployment of the mirrorsats.

4.3 The Mission Orbit and Formation Control

A common approximation in research of this type models the dynamics of a satellite in the vicinity of the sun-Earth L2 point using the assumption of circular restricted three-body motion. This assumption accounts only for gravitational forces from the sun and Earth. The moon is also included, but not as an independent body. The masses of the Earth and moon are combined and assumed to be at the Earth-moon barycenter. The motion of the sun and the earth-moon barycenter is also assumed to be circular around the system barycenter. GSFC analysis uses high fidelity dynamics to create more realistic Lissajous orbits than those derived from the circular restricted three-body problem, taking into account the accelerations due to lunar and Earth mass and eccentricity, the other planets of the solar system, and solar radiation pressure. The resulting orbit can then be used as an accurate reference orbit. In addition to providing the reference positions and velocities, non-linear dynamics information is generated for each SI component at each epoch. This data can be used onboard for autonomous control by simple uploads or onboard computation as a background task of 36 matrix elements and the state vector. A sample reference orbit in a rotating coordinate system is shown in Figure 4.4. The *X*-axis coordinate connects the two primary bodies (sun and Earth), the *Z*-axis coordinate is parallel to the angular velocity of the system (Ecliptic Pole), and the *Y*-axis coordinate completes a right-handed system. The reference orbit period is 180 days and the scale is in kilometers. The amplitudes in each axis vary for each particular design but can be as large as 800,000km in Y, with the other axes scale accordingly.

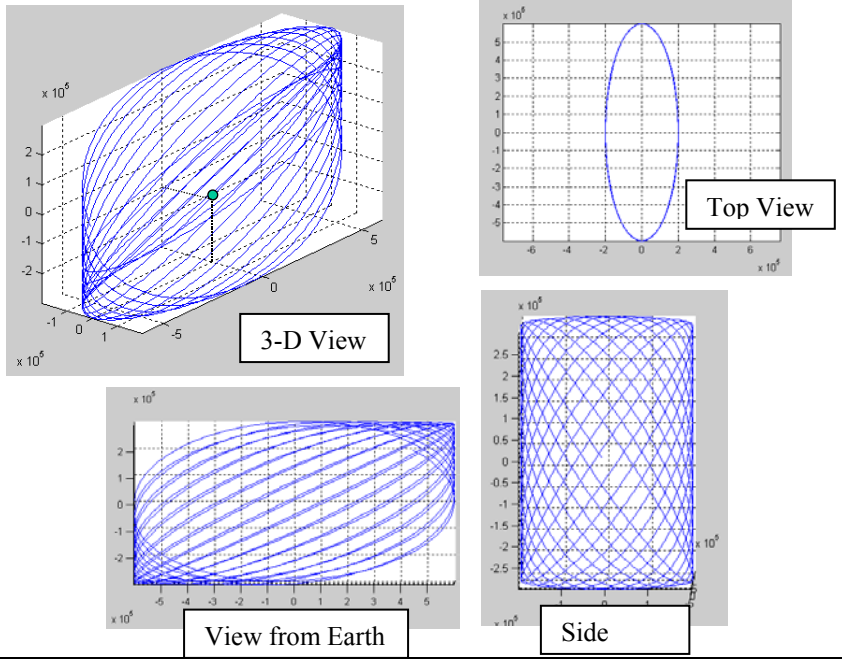


Figure 4.4: A Sample Lissajous Orbit at Sun-Earth L2

A subset of the Lissajous orbit pattern is the halo orbit shown in Figure 4.5. In the halo orbit, the frequency in the Z direction is equal to the frequency in the in-plane (X-Y) direction. This permits the generation of a unique set of orbits of either class I or class II, where class defines the tilt of the orbit either away from or towards the Earth. In all these cases, any orbit can be achieved, but the impact of selecting a particular orbit will be to limit launch window opportunities or increase the required maneuver control and fuel allocations.

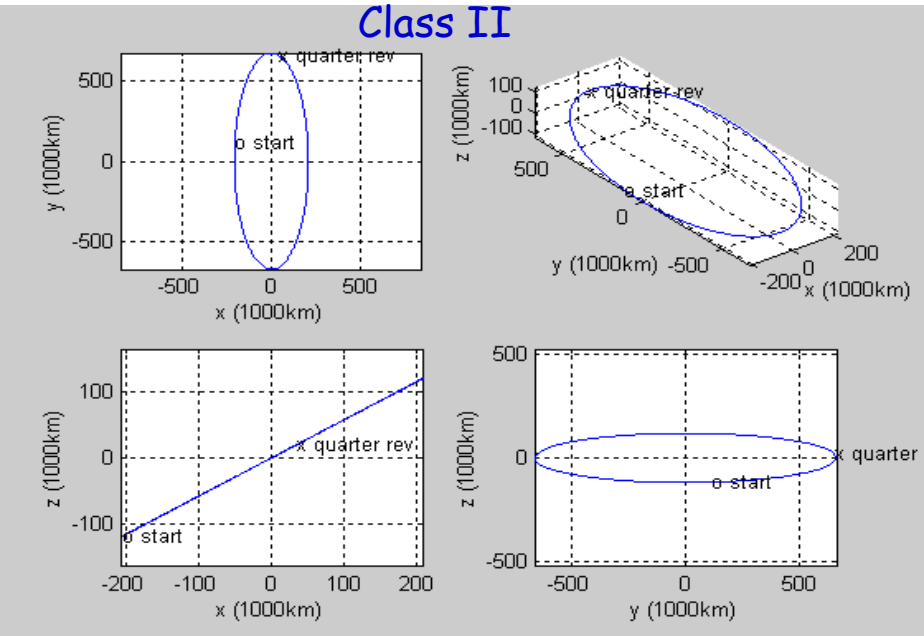


Figure 4.5: Halo Orbit

4.3.1 Formation Control

Three different scenarios make up the position control problem; maintaining the Lissajous orbit, slewing the formation, and reconfiguring the formation. These three scenarios can be treated independently. To determine the amount of fuel, the velocity change, or ΔV, is needed. The ΔV in each direction is found by numerical integration:

$$\Delta V_x = \sum_{k=1}^{359} |u_x| T \,,\ \Delta V_y = \sum_{k=1}^{359} |u_y| T \,, \Delta V_z = \sum_{k=1}^{359} |u_z| T$$

where T is the maneuver interval, the upper limit of the sum is the simulation length. The absolute value of the control is taken because the direction of the maneuver has no bearing on the fuel used. The total ΔV for one simulation is calculated by

$$\Delta V = \Delta V_x + \Delta V_y + \Delta V_z$$

Depending on the mass and area of the mirrors the drifting between the mirrors and the detector can range from mm/day to cm/day. The key here is that the drifting is exponential and is dependent upon the unstable dynamics of the Sun-Earth region. A reference orbit and a sample formation are shown in Figure 4.6 and Figure 4.7, respectively.

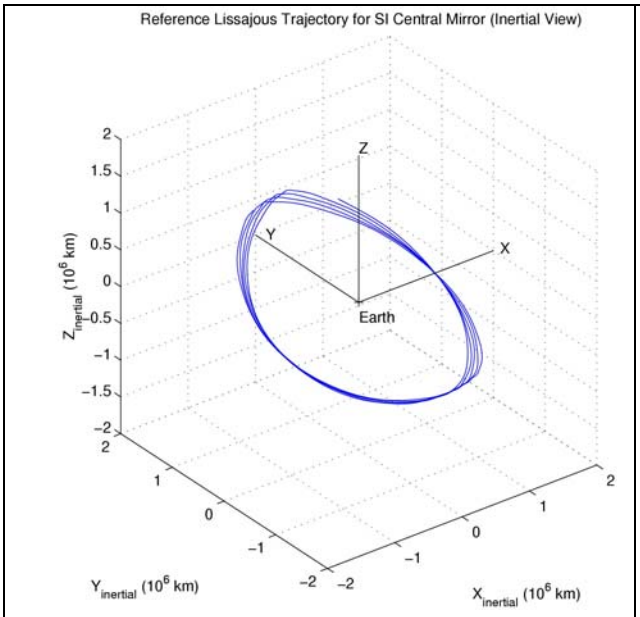


Figure 4.6: SI Reference orbit

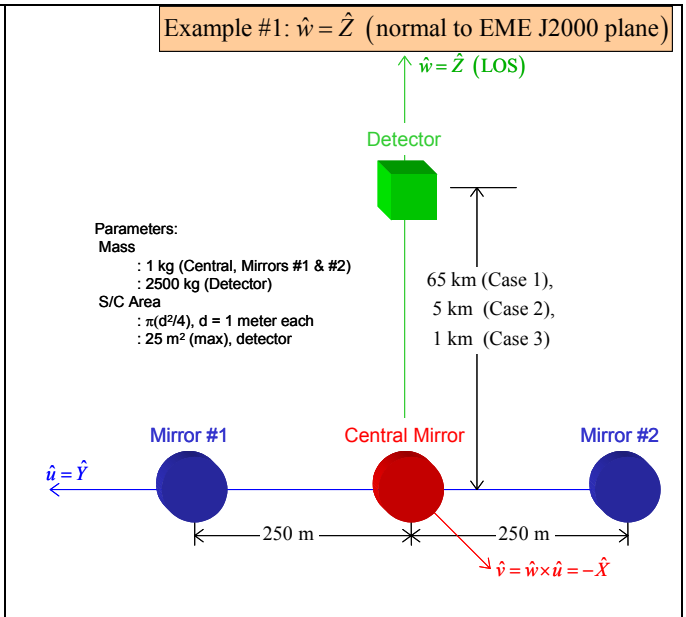


Figure 4.7: SI Reference Formation

4.3.2 An Example of Formation Control

A key goal of the SI mission is to image many stars. Following a Lissajous orbit, SI could view the entire sky approximately every half-year by slewing about the radial-to-the-sun (x) axis. A formation slewing simulation follows a similar algorithm as the Lissajous orbit simulation. As part of the formation maintenance and slewing, estimation of the formation geometry must be performed. Assuming that relative navigation will provide the input to an autonomous system for control, a 'simple' controller can be developed to maintain a reference location with respect to the objective or other mirrors. These controllers can take the formulation of Linear Quadratic Regulator (LQR), Input Feedback Linearization (IFL), or a Floquet controllers. The tracking errors of all formation components should be completely understood for accurate control. Figure 4.4 shows a reference orbit for the central spacecraft.

Using an IFL controller, Figure 4.8 (top) shows the control effort for a sample drift with a 1-km separation. Note the drift is exponential in nature. After a time of 3200 seconds a controller is turned on to effect a re-alignment of the formation. The ΔV is shown in Figure 4.9 (bottom) and is in mm/sec. Note that Figure 4.8 is for the hub (detector s/c) control.

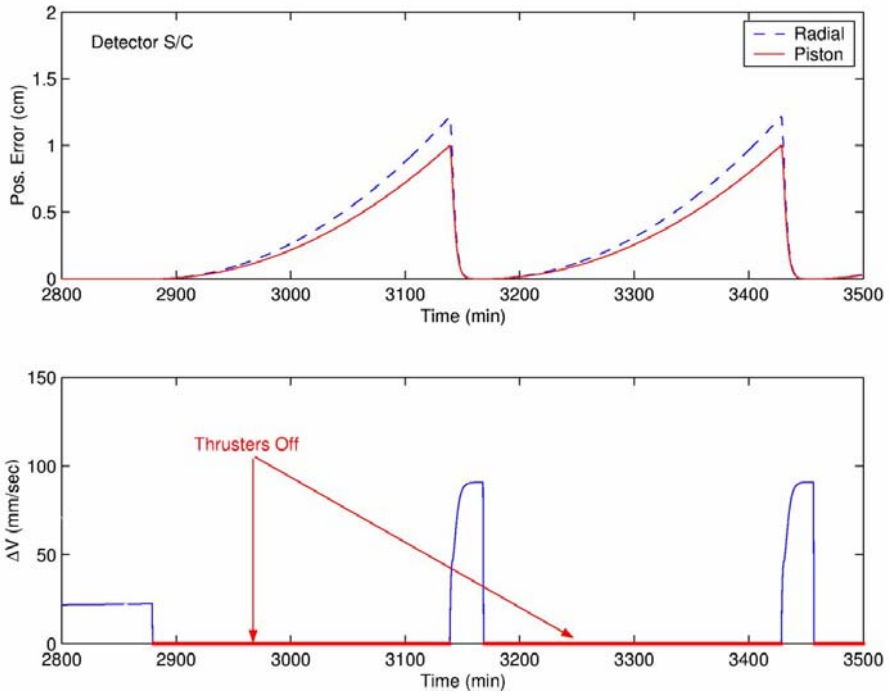


Figure 4.8: Detector control

For the mirror control a similar control effort is required and is shown in Figure 4.9.

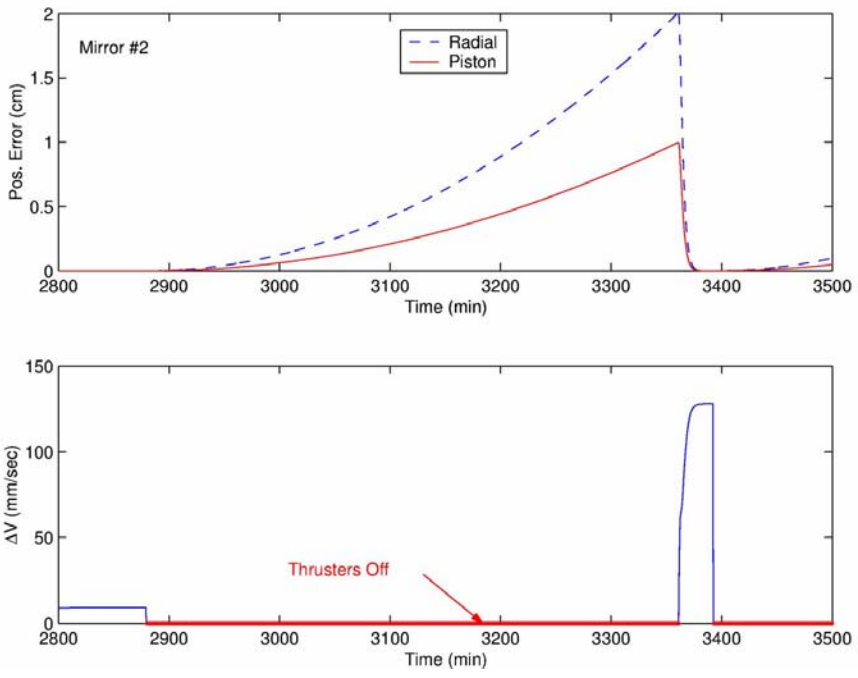


Figure 4.9: Mirror Control Effort for large separation (up to 65km)

The ΔV requirement is proportional to the formation slew angle and the focal length; however for perfect control and no noise on the system the ΔV for each mirror per maneuver can be significantly lower as shown in Table 4.1

Table 4.1 Time To Drift And Control Effort To Return

	Distance (km)	1-cm violation (Minutes)	ΔV to return (mm/s)
Detector	65	33	110
Detector	5	100	100
Detector	1	230	100
Mirror	0.25	460	130

4.3.3 End of Mission

The individual spacecraft will eventually leave their unstable L2 halo orbits after loss of station keeping ability caused by depletion of on-board propellant. The individual spacecraft will then drift off into separate solar orbits that do not intersect the Earth. A statistical analysis of the departure orbit will need to be performed to provide a timeframe but a plausible number is on the order of several thousand years.

Chapter 5: Operations

This section describes the operations considerations for Stellar Imager.

5.1 Space Segment

After initial check-out and commissioning, Stellar Imager will be an autonomously controlled constellation using onboard software to maintain the optical configuration of the system. Commands to re-point the system to a new target will come from a stored command area onboard and be combined automatically onboard to re-point the observatory and re-establish the optical configuration at the end-point of the maneuver to a new target. SI flight software will constantly monitor and constrain various parameters onboard to stay within defined limits (Sun angles, solar arrays, etc). The frequency of re-pointing of the SI will vary greatly, between once per hour and once per month. The sequence of re-pointings and associated science instrument commands for each pointing will reside in stored command memory onboard. SI flight software will check each sequence of commands for health and safety rules before executing them. At each pointing of the SI, the onboard systems will automatically acquire guide stars, verify attitude, acquire the science target, and initialize the observing sequence and the optical configurations required. Confirmational data from each of the steps will be stored in onboard Solid State Recorders (SSRs) for later transmission to the ground.

The SI design will include autonomous capability for re-configuring the component spacecraft as necessary for the science observing programs requiring this function. A collision avoidance backup system will ensure that this capability operates successfully or interrupts attempts to make an unsafe maneuver and alerts the ground to the interruption.

Stellar Imager will store commands sent periodically from the operations center and periodically send data to the ground from its onboard data storage area. The Hub design will contain the communications equipment for space-ground contact. The Hub will be designed for optimal lifetime by including various redundant features for all essential functions. In addition, it is desirable for two Hubs to be in operation in the optimal design. If one fails, this "critical path" component has an immediately available backup. In addition, the availability of two Hubs greatly increases the efficiency of the observatory – the second Hub can be pre-positioned while the first one is in use observing a target and the observatory can be re-pointing simply by tilting the primary array to align with the second Hub, without any large slews for the numerous (~ 30) mirrorsats. An extreme example of this could be obtained by positioning one Hub on each "side" of the primary array – thus enabling the observatory to be repointed halfway around the sky with simple "flips" of the primary mirrorsats, taking perhaps minutes instead of the days to do so with only one Hub. The SI design will include alternate communication capability for the unlikely event of a loss of primary Hub space-ground capabilities. SI will include onboard capability for recognizing failures in any given primary mirror unit and ability to avoid collision with the other members of the constellation.

Stellar Imager will broadcast a "distress" beacon signal within TBD seconds of sensing an onboard emergency situation as defined in the onboard software and safing subsystems. The safing subsystem will be hosted in an internally redundant computer independent of the main operations computers onboard. If the safing subsystem finds that critical events have been triggered as defined in its database, it will autonomously put the SI into a safe state as defined in its database. The distress signal

will contain critical information about the event and certain engineering data onboard at the time of the event. If the safing system is triggered, the alert signal will be sent in parallel with the safing actions.

5.1.1 Orbit

The Stellar Imager constellation will orbit the Sun-Earth Lagrange Point known as L2. The entire SI constellation will have to be commanded to correct its orbit periodically to maintain long-term orbit around L2 and to optimize fuel usage in such corrections (see Chapter 4). The Mission Operations Center will provide the necessary commands to SI when tracking data indicate that orbit correction is necessary. In this orbit, SI will have uninterrupted view of the Sun and hence uninterrupted electrical power from its solar arrays. High gain antenna pointing towards Earth for optimal communications gain will be controlled onboard by the flight software, using the onboard SI and Earth ephemerides and the onboard inertial navigation data.

5.1.2 Formation and Science Target Acquisition

Due to SI's distributed architecture and exceedingly stringent control requirements, a multi-step process is required to acquire a science target. This sequence includes several handoffs from "coarse" sensors to "fine" sensors, with more accuracy but limited dynamic range. The acquisition sequence is described here in narrative form, and the requirements are summarized in Table 5.1.

1) Deployment: The launch vehicle ejects hub, reference, and mirrorsats one at a time. As each spacecraft is ejected, it is activated and basic attitude and positional control is established. After deployment, the spacecraft will be dispersed over a range of a few kilometers at most, with relative drift of a few kilometers per hour, and arbitrary (but known and controlled) attitude. This is the beginning state of the acquisition sequence.
2) Formation Acquisition: Using RF ranging sensors, the hub, reference spacecraft, and mirrorsats move to form the nominal relative configuration, pointing (to star tracker accuracy) at a pre-selected acquisition target.
3) Laser metrology acquisition: Laser sources on the hub are directionally scanned to find reflectors on the mirrorsats. Laser ranging is then used to trim the hub-mirrorsat distance to within micrometers of its desired value.
4) Coarse spot acquisition: Recall that the hub is pointing at the acquisition target, using star tracker knowledge. Now one mirrorsat's attitude is adjusted to put its light beam onto the Optical Wavefront Sensor, using a search scan. This helps refine the pointing of the hub to the target. The other mirrorsats are added sequentially, adjusting each to put its light beam onto the Optical Wavefront Sensor.
5) Refine Pointing: With target light falling onto the Optical Wavefront Sensor, each spacecraft's relative position and attitude are adjusted to better point the synthetic telescope at the target.
[**Note:** For clarity, steps 3-5 have been described sequentially. In practice, however, they may overlap to reduce the acquisition time, e.g. some mirrorsats performing coarse spot acquisition while others are still in laser metrology acquisition.]
6) Fine spot acquisition: At this point, each mirrorsat has a spot of target starlight falling somewhere on the detector of the Optical Wavefront Sensor. Now each mirrorsat is adjusted to place its spot on the central region of the Optical Wavefront Sensor.
7) Refine pointing to target: Again, the improvement in pointing information allows a refinement in pointing of the overall formation. This refinement nulls any residual angular drift of the system.

8) Fringe acquisition in Wavefront Sensor: Optical Wavefront Sensor image feedback is used to trim up pointing and position to achieve science-quality Optical Path Length control. With the pointing accuracy achieved in step 7, fringes will appear on smaller baselines. Fringes can be acquired in a few seconds using the wavefront sensor. Fringes can also be acquired on the shorter transverse baselines. Using fringe feedback, the mirrorsats adjust piston to null Optical Path Differences (OPD) to the nanometer level. The fringes on short baselines are then analyzed to measure OPD rate more accurately, which enables acquisition of fringes on longer baselines. This "bootstrap" process continues until fringes are acquired on all baselines.

9) Fringe acquisition in Science Sensor (UV or optical): Finally, fringes may be acquired in the Science Sensor, enabling the collection of science data. This is the goal and end of the acquisition sequence.

Table 5.1 Acquisition Modes

Mode	Formation Sensor	Exit Criterion	Linear Rate Required	Angular Rate Required
Deployment	None (maybe RF ranging)	S/C deployed from launcher.	5 km/hr	10 deg/sec
Formation Acquisition	RF Ranging	Form nominal relative formation, pointed at pre-selected acquisition target.	N/A	N/A
Laser Metrology Acquisition	RF, Laser Metrology (LM)	Laser metrology acquired.	N/A	N/A
Coarse Spot Acquisition	LM, Optical Wavefront Sensor (OWS)	All spots on OWS.	N/A	N/A
Fine Spot Acquisition	LM, OWS	All spots centered in OWS.	N/A	N/A
Fringe Acquisition in Wavefront Sensor	LM, OWS	Wavefront sensed and controlled.	N/A	N/A
Fringe Acquisition in Science Sensor	LM, OWS, Science Sensor (SS)	Science-quality OPD control.	N/A	N/A
Science	LM, OWS, SS	Exposure complete.	N/A	N/A

Table 5.2 Acquisition Handoff Requirements

Mode	Position Requirement for Exit			Rationale	Attitude Requirement for Exit		Rationale
	OPL/Piston	Radial	Tangential		Tip/Tilt	Roll	
Deployment	10 km	10 km	10 km	Assume limit of RF comm is about 50 km. Allow room to damp any deployment rate.	Any	Any	Assumes mirror covered for safety.
Formation Acquisition	10 cm	10 cm	10 cm	Metrology sensors on hub can acquire cornercubes on mirrorsats within 1 cm with 10x10 pixel raster scan. Limit of RF ranging capability.	1 deg	5 deg	Retro-reflectors insensitive to attitude. Laser sources on hub are steerable. 5 deg reasonable limits for sun shield, solar array offpointing.
Laser Metrology Acquisition	10 um	10 um	10 cm	Intermediate in preparation for Coarse Spot Acquisition. Well within capability of laser metrology system.	1 arcsec	5 deg	Intermediate in preparation for Coarse Spot Acquisition. Within capability of ST/gyro.
Coarse Spot Acquisition	10 um	10 um	10 cm	Maintain accuracy from previous mode.	0.4 arcsec	5 deg	Place a 1-cm beam on a detector from 5 km away.
Fine Spot Acquisition	5 nm	10 um	10 cm	Control OPL to fraction of UV wavelength to facilitate fringe acquisition in OWS. Radial position well controlled to maintain mirror on virtual primary surface. Tangential direction lies in VPS, so tangential control much less critical.	0.04 arcsec	5 deg	Superpose 1-cm beams on detector to within 1/10 of beam width.
Fringe Acquisition in Wavefront Sensor	5 nm	10 um	10 cm	Same as above.	0.04 arcsec	5 deg	Same as above.
Fringe Acquisition in Science Sensor	5 nm	10 um	10 cm	Same as above.	0.04 arcsec	5 deg	Same as above.
Science	5 nm	10 um	10 cm	Same as above.	0.04 arcsec	5 deg	Same as above.

5.2 Communications

Communications services through Deep Space Network will be used to update onboard command memory, allow daily reproduction of science and engineering data from the SSR(s) to the ground, collect tracking and ranging data for use in calculating orbital elements of SI, and send any re-configuration commanding deemed necessary for maintaining and enhancing the SI system. On an occasional basis, the contents of onboard computer memories will be dumped to ground for analysis and occasionally new software and database content will be sent to onboard memories from the ground.

5.2.1 Uplink

All communications to SI are planned assuming use of the Deep Space Network (DSN). Primary uplink communications, to the Hubs, will be at 2kbps using X band, with 2kbps S band backup. The primary link will include automatic communications from the Hubs to each of the Mirrorsats using SI internal communications subsystem. Backup link to the Mirrorsats will be via S band at 2kbps from the ground.

5.2.2 Downlink

All communications from SI are planned assuming use of the DSN. The nominal aggregate data rate from SI to ground is about 125 Gb/day for about 11 months of each year. This requires approximately one 30 minute Ka-band downlink per day. For about one month per year continuous data rate aggregate of about 250 Gb/day is expected assuming a 2:1 lossless compression of the science data, which will require approximately one 60 minute Ka-band downlink per day. These figures include about 15% overhead for CCSDS formatting. Primary downlink of stored data from the Hubs will be 75 Mbps on Ka band. SI will automatically control data from the Mirrorsats to the Hubs for storage. Real-time data downlink from the Hubs will be via X band at 10 kbps with a backup of 6 kbps on S band. Backup real-time telemetry from the Mirrorsats directly to ground will be via S band at 3 kbps.

5.3 Ground Segment

The Stellar Imager Mission Operations Center (MOC) will be staffed one 8-hour shift per day for 5 days per week by the Flight Operations Team (FOT). This FOT will operate the prime and backup control center systems to maintain normal operations of the constellation. The control center system will automatically detect anomalous conditions, warn operators and switch to backup systems if operators do not respond. The application software systems in the control center will be based on heritage software from the SI development and I&T phases, together with IP (internet protocol) communications software between the MOC and SI. IP/COTS (Commercial-Off-The-Shelf) applications developed for control centers are assumed to be mature by the time SI requires them. The MOC will communicate externally via T3 lines. These products are assumed to provide data delivery assurance technology built-in. The command and telemetry databases used in the ground system will be inherited from the SI development and Integration & Test phases. Normal operations will include routine generation of science observing schedules and associated command loads, and transmission of these command loads at times scheduled for uplink contacts. These schedules will be generated based on the science plan residing in the operations control center system, periodically transmitted or updated from the Science Operations Center (SOC).

The FOT will also schedule contacts for replaying data from the onboard data storage system, command the SSR playbacks and receive and confirm the data at the control center. They will use control center software systems to receive, analyze and confirm engineering data from all SI subsystems and verify health and safety of the subsystems. The FOT will be able to process real-time telemetry and in parallel process dumps from the SSR(s). The science data received will be forwarded (level 0) to the data distribution system for processing and distribution to the SOC, normally within 48 hours of collection onboard. On an occasional basis, when the science plan warrants it, science data latency can be reduced to 6 hours by FOT selective control of the SSR pointer. Level 0 data will have duplicates removed, and quality flags attached for all the data in a downlink in chronological order.

The ground system will include a data archive facility, with a shadow backup repository for restoration in the event of a catastrophic loss of data in the prime archive. The archive will store all science and engineering data from SI in raw and processed forms as well as all versions of the SI ground system databases and software, and calibration databases. The long term average accumulation rate for the archive is expected to be approximately 400Gbits/day or about 145Tbits/year. This data will be online for access via web connections for general research use once the initial proprietary data rights period has elapsed for each particular dataset. The archive science data will be available for incorporation into other astronomical datasets for coordinated research purposes.

The ground system will include a software suite for monitoring the optical performance of SI. This software will be inherited from the design and development phase of SI and will represent a very powerful tool for use by the FOT. The operations form of this software will be a more standardized version of the development software, to enable routine use by the FOT.

On an occasional basis, the FOT will send re-configuration commands to SI (e.g. orbit maintenance, flight software updates, etc.) and confirm successful completion of these activities. Tracking and ranging data for SI will routinely be sent to the Flight Dynamics Facility (FDF) at NASA/GSFC for routine analysis. The FDF will send orbit element reports to the MOC. These reports will be used to determine orbit maintenance activities and commands, and associated critical communications schedules.

The SI ground system will include simulators at appropriate locations. There will be a training simulator in or near the MOC for training of all operations personnel. There will be a high-fidelity

simulator (including engineering model hardware and flight software) inherited from pre-launch design and test activities. Post-launch, this hi-fi simulator will be used to aid in maintenance of flight software and for trouble-shooting unexpected anomalies on SI. The location of this hi-fi simulator is TBD and location might change based on the phase of the mission.

5.3.1 Safing System Operation

The essence of keeping SI safe in normal operations will be based on the onboard fully-autonomous safing system. When a distress beacon alert signal is broadcast by SI via the "always on" channel used for this, the ground system will automatically detect the signal, prioritize the receipt and analysis of the content of the signal and immediately relay the appropriate alerts (paging, email, etc) to the FOT. This system will operate at all times, whether the FOT is on-station or not. In the event of a delta-V control system failure the FOT will have the ability to command appropriate responses but, in general, collision avoidance requires immediate action so this will rely on onboard autonomy to be followed by ground intervention. At least one FOT member will be on-call to receive the alert and respond by either remotely accessing the control center systems or going physically to the MOC to respond. The FOT response will include Health and Safety checks, other pre-defined immediate action procedures, and establishing communications as soon as possible through the available communications systems. The FOT will have a suite of contingency operations to initiate onboard "macro" command constructs for rapid and safe intervention. After suitable engineering analysis and understanding of the causes of the safing, the FOT will restore SI to normal operations and recover to a normal observing schedule. In anomalous events such as safing, the FOT will have rapid access to expert engineering assistance for in-depth analysis, understanding and response.

5.3.2 Staffing

The normal long-term FOT staff will consist of a shift supervisor/system engineer plus 3 Mirrorsat/Hub engineers (4 people total) working 8 hours/day x 5 days/week. There will be a staff of two (one hardware plus one software engineer) 8 hours x 5 days to maintain the MOC systems used by the FOT. With two FOT teams available, it will be possible to staff on-call and vacations, etc. The FOT teams will be rotated regularly to maintain skill levels and the various staffing requirements. Ground system software will include a suite of analysis and graphics displays etc. to enable the FOT to monitor all SI subsystems and recognize and understand anomalous conditions. The FOT will also schedule and maintain calibration databases and onboard configurations. At least one FOT member will be on-call at all off-hours. When an anomalous condition is recognized, the FOT will have the resources to call designer-level support from all engineering disciplines at short notice to work non-routine situations. It will be possible for the FOT to share any data they have with the designers within 5 minutes.

Initial normal operations will be supported 24 hours/day x 7 days/week for a settling period (~6 months) until normal operations support will be adequate. The initial operations staff (6 per 12-hour shift) will augment the skill mix as well as the round-the-clock schedule. This initial augmented staff will consist primarily of all of the systems engineers associated with the development and test phases, and they in turn will have access to the design engineers on short notice (within ~1 hour).

Throughout the operational phase of SI, all the simulators will be available at short notice to aid in trouble-shooting and developing "fixes" to problems onboard as necessary. A flight software update capability and staff (3) will be available as necessary.

5.3.3 Training

The FOT and support staff will be trained over a period of ~9 months prior to launch. Each FOT member will take a training curriculum prepared by the SI designers and systems engineers, followed by an examination to confirm proficiency. Training will consist of classroom sessions and hands-on activities at flight consoles supported by simulators. The simulators will simulate all normal operations activities and all contingency operations documented in the SI operations documentation with reasonable fidelity to actual SI behavior, such that the FOT will be ready to respond to all prepared contingency situations. This training might be updated and/or re-enforced periodically during the SI flight mission lifetime.

Chapter 6: Operations Assurance

Operations Assurance is a term generally defined to be a combination of Mission Assurance and Systems Engineering, with activities conducted throughout all phases of the mission for the purpose of raising the probability of complete mission success. NASA defines Mission Assurance to include Reliability, Maintainability, and Quality Assurance. But some other organizations, such as JPL, have defined Mission Assurance to include EEE Parts, Environmental Design & Test Requirements, Reliability, Quality Assurance (H/W & S/W), Materials & Processes Control, Systems Safety (includes personnel and H/W), and Operations Assurance! Although there is considerable circularity in these definitions, we shall assume the Operations Assurance effort for SI will consist of adequate attention and activities pertinent to all these elements throughout all phases of the mission. The main goal, always, is to provide the highest affordable probability of the mission being successful.

In the beginning, we must provide realistic Reliability and Maintainability (R&M) requirements for system development specifications and requirements documents. Those requirements will promote operations assurance if accompanied by a commitment to:

- Allow for early and continuing attention to R&M principles during system design.
- Achieve system R&M as defined by the SI mission objectives.
- Control the SI system life-cycle cost by addressing operations and maintenance support cost drivers during system design.
- Measure, assess, and report R&M performance throughout the SI life cycle.
- Maintain a comprehensive and readily accessible database of success and failure data for use in prediction, problem trending, and assessment of progress toward system success goals throughout the SI life cycle (as well as to establish R&M performance requirements for follow-on or programs).

We usually measure R&M operational effectiveness in terms of the parameter, Mean Time Between Failures (MTBF). To achieve the longest MTBFs for SI by applying R&M principles, the following tasks should be applied in the sequence listed:

1. Identify desired outcomes, emphasizing objective and measurable outcomes based on SI mission requirements.
2. Select measures and indicators for evaluations. These may change as program perturbations occur.
3. Set performance and surveillance standards based on SI mission requirements and operating environments. The program Statements of Work (SOW) ensures that surveillance plans are established and implemented early in the program.
4. Report results. Results of the surveillance effort are reported to support assessment of the progress made toward meeting SI mission requirements.
5. Use results for planning, managing and budgeting. Assessing progress toward meeting requirements provides the feedback needed to adjust planning, managing and budgeting of the SI program.

Desired Outcomes: Risk Management Plan, and Flowdown of R&M to subsystems

The following steps are essential to successful operations assurance via completion of the risk management plan and the flowdown of R&M objectives to the SI subsystems:

- Completion of the detailed system design. It can include products such as simulated and physical mock-ups and test articles of critical systems and subsystems; complete detailed system and component specifications, systems baseline description, and comprehensive requirements traceability of all derived requirements to SI mission requirements.
- Development of the system hardware and software and delivery of an operational SI system acceptable to the ultimate user. This comprises the fabrication, integration, testing, and certification of all system hardware/software required for system initiation and subsequent operations. Contractor activities must be monitored early and coupled with analyses and tests to measure and evaluate achievement of the required R&M performance prior to acceptance of specified hardware and software.
- Satisfaction of SI mission needs throughout operations. R&M activities continue during Implementation, for logistics support, sustaining engineering, and to document "lessons learned." Four main factors control R&M operational effectiveness and must receive our attention. These factors are:

    Active Corrective Maintenance
    Preventive Maintenance
    Logistics Downtime
    Administrative Downtime

Evaluation

The objective of Evaluation is to provide timely assessment of the continuing ability of the SI program to meet its technical and programmatic commitments. It also provides value-added assistance to the SI program manager as required. A major portion of this process is the continuing assessment of work initiated during Formulation. Specifically this includes the use of program metrics, life-cycle cost models and risk management analyses, all of which can include elements of R&M. Evaluation continues concurrently with Implementation, in the form of reviews to measure program performance against program plans. These reviews are expected to include the Preliminary Design Review (PDR) and Critical Design Review (CDR), and would address progress toward meeting R&M and other related requirements (e.g., for safety, logistics, and maintenance).

R&M Integration With Other Organizational Elements

Integration of R&M with other organizational elements ensures that requirements are met expeditiously and economically. Integration includes coordinating efforts, establishing/ controlling/verifying interfaces, eliminating redundant activities, facilitating the flow of information, and applying R&M engineering techniques and practices appropriately during hardware and software system development. Integration/verification must assure the achievement of all R&M performance requirements. During Formulation, the SI program manager will develop the program structure and establish the foundation for integration of the associated disciplines. The following paragraphs highlight the relationship between R&M and the organizational elements of the SI program.

. 1. Manufacturing and Quality Assurance

A primary R&M concern during manufacturing is to prevent degradation of the inherent R&M designed into the product during Formulation and Implementation. The Quality and Product Assurance activities should work closely with the R&M development team to ensure a full understanding of the impact of the manufacturing processes on end item R&M and to assure the integrity of the product by using capable and controlled program-critical manufacturing and operational processes. Involvement of R&M Engineering in the review/approval loop for the selection of parts and materials, manufacturing processes and procedures, and assembly procedures is critical.

2. Diagnostics and Maintenance

The various R&M analyses, such as Failure Mode and Effects Analysis (FMEA), reliability predictions, and maintainability predictions are a significant source of information in designing a diagnostics system and maintenance plan. An important component of maintainability is testability, defined as a design characteristic that allows the status (operable, inoperable, or degraded) of an item to be determined and the isolation of faults within the item to be performed in a timely and efficient manner, especially by means of built-in-test capabilities. Decisions arising from development of diagnostic systems and the corresponding maintenance procedures may require updating the R&M analyses and should be fed back to R&M engineering on a timely basis.

3. Software

The computer software life-cycle covers the period from its conception until the time it is no longer available for use. Software is seldom static, and long after a system has become operational and is in use, upgrades to its software are likely, to increase performance or take advantage of new technology. Software R&M is concerned with the initial release and subsequent upgrades of software to ensure that minimal downtime is attributed to the software. R&M performance requirements should be specified for software as well as hardware. The resulting tradeoffs among hardware, software and operations become an important aspect of R&M and System Engineering. Close coordination of all design, development, operational, and support decisions among Hardware, Software, and R&M groups is essential for achievement of software R&M performance requirements.

4. Program Engineering

R&M engineers and program engineers have a common interest in identifying risks and hardware and software failure modes, along with potential effect(s). An interface between these two disciplines will ensure timely communication of all potential risks, provide early identification, and aid in development/application of methods to mitigate the risk from the possible failure modes in the proposed design. R&M engineering should review the risk mitigation methods to ensure that reliability and/or maintainability are not degraded by the resulting risk mitigation method.

Developing a Fault Detection, Isolation and Recovery (FDIR) Capability

The main goal of fault detection, isolation, and recovery (FDIR) is to effectively detect faults and accurately isolate them to a failed component in the shortest time possible. Development and use of such a capability leads to reduction in diagnostic time or downtime in general and, therefore, increased system availability. A good inherent diagnostic of a system also enhances confidence in operating the system, the main driver of mission success. Effective FDIR can keep a difficult to maintain system up and running where normal methods would lead to system downtime, especially beneficial for an on-orbit system where maintenance is limited.

Risk Management Activities Related to R&M

A fundamental element of Formulation is the initiation of program risk management activities. The following activities are central to a risk management process that promotes successful R&M for the SI mission:

Establish Mission Success Criteria

Environmental stress screening (such as burn-in, temperature cycling and vibration testing)
- · FMEA
- · Reliability Centered

Maintenance
- · Reliability Requirements
- · Redundancy
- · Parts selection criteria and control
- · Derating of parts
- · Reliability Plan
- · Components service
- · Conformal coating
- · Simplicity of design
- · Product survival in the intended mission environment (radiation, plasma, micro-meteor/orbital debris, humidity, temperature, mechanical shock, vibration, electromagnetic compatibility)
- · Predictions
- · Control of the physical environment
- · Expected operating and storage times (for limited life items)
- · Failure propagation
- · Protection of cables, wires, receptacles, plug ends, connectors
- · Probabilistic Structural Analysis
- · Failure/fault tolerance
- · Use of preferred parts and materials
- · Verification of operational status for redundant paths
- · Redundancy management
- · Burn-in to eliminate infant mortality parts

Reliability Prediction
- · R&M allocation
- · Failure modes and effects analysis
- · Criticality analyses
- · Fault tree analysis
- · Worst case circuit analysis
- · Maintainability assessment.

R&M Performance Requirements Verification

Verification establishes that R&M performance requirements have been met. The contractor can use various verification techniques (e.g., test, analysis, and inspection) to ensure that the system or product items being developed meet the R&M performance requirements and will perform effectively in the intended operational environment**.** The contractor or supplier must prepare R&M test and evaluation plans that provide the details of the R&M performance and demonstration tests. During R&M testing activities, adequate budgeting must be provided for both NASA and the contractor to perform all the test programs necessary to verify the R&M performance requirements.

Correct selection of a verification method, i.e., Test and Evaluation (T&E), helps ensure that the SI system Architecture complies with the R&M performance requirements as the program evolves. Verifying the R&M hardware performance requirements early in the program will provide timely logistics planning and spares projections; however, due to the early application it may not ensure that all R&M performance functions have been exercised.

R&M T&E is conducted to evaluate how well the system meets the specified R&M performance requirements for successful operations assurance. Once the system R&M performance requirements have been established and documented, T&E can be planned**.** Selection of the most appropriate verification method or methods for a given requirement should be based on the following:

- the test or evaluation method that can be applied by the contractor as early as possible in the system or product life-cycle to demonstrate compliance with the R&M performance requirements.
- the evaluation method or combination of methods or tests that are most effective for demonstrating R&M compliance with the system requirements, and
- the evaluation or test methods that are most efficient when considering the safety and cost risks involved.

6.2 Reliability

Reliability analyses such as reliability block diagrams analysis are used to verify the fulfillment of quantitative requirements. The attribute of reliability, by definition, lies in the probabilistic realm while most performance attributes or parameters such as temperature, thrust, voltage, or material strength contain more measurable characteristics. No such measuring device exists for reliability. It is usually estimated through comparison with similar components or systems through inference, analysis, and the use of statistics. The following are the main tools to estimate system reliability.

- Reliability Analysis (Block Diagram Assessments, Availability Simulation)
- Probabilistic Risk Assessment (Fault Tree Analysis, Event Tree Analysis)
- Reliability Qualification or Acceptance Testing

A reliability requirement specified without a probability value, such as “the system shall perform the SI mission on-orbit without failure for 5 years,” is impossible to verify during qualification or

acceptance testing. The statistical likelihood, or probability, that the requirement will be met is assessable, and this activity is inherently equivalent to assessing the reliability.

During Operations the system truly demonstrates its capability to meet the requirements developed during program Formulation. Operations also provides a unique opportunity to continue the evaluation and upgrading of the system(s) R&M performance with the dual benefit of ensuring that the R&M performance meets and maintains intended capabilities and incurs lower lifetime costs. The corrective action system developed during Formulation and Implementation should continue to be used in operations to support upgrading R&M performance.

Validation of hardware after operations have begun may be necessary to evaluate how the hardware is performing under actual operating conditions. History has shown typical reasons for such validation include: 1) flight hardware can be slightly different from qualification units; 2) environments may differ from what was expected and qualified to, and 3) changes in subsystems can induce some surprises in performance characteristics. Use of a structured and controlled data acquisition process provides the necessary information to perform trend analysis on the behavior of the SI system and to support root cause analyses of failure situations.

6.3 Reviews

For a strategic program of this size, it is clear that there would be a very long and exhaustive series of reviews, including System Requirements Reviews, Preliminary Design Reviews, Critical Design Reviews, etc. And it would be best to have them for various components of the system separately - surely for the mirrorsats and hub, but perhaps also for the optical control and formation flying, before the overall system reviews are held, due to the complexity and variety of the components and overall system.

6.4 System Resilience

The SI design is resilient in two major ways.

The most important perhaps is that the observatory is robust against the failure of one or more individual mirrorsats. If the full complement of 30 mirrorsats is put into place during the initial launch and deployment, then science observations can continue even if individual mirrorsats are lost due to hardware or other failures. The number of elements in the array has been chosen to enable efficient synthesis imaging, i.e., 30 elements covers the necessary number and diversity of baselines to adequately sample the Fourier (u,v) plane with few or no reconfigurations of the primary array during the observation of a given target. Many of the targets can thus be observed in a “snapshot” mode – where the array is pointed at a target and all the necessary data are acquired with the array elements in same pattern for the entire time. A few, more complex, extended targets may require that the array be rotated or reconfigured once or twice to get the necessary sampling, but most will not. As mirrorsats fail, the baselines covered by the remaining elements will decrease and the quality of the imaging synthesis will degrade, unless the remaining operating elements are moved around (reconfigured into new patterns and/or rotated as a whole) to fill-in the missing Fourier frequencies. Thus the “snapshot” observing mode will not be available and the observations will require more and more movement of the mirrorsats to maintain image quality – and the length of observation at each target will increase until, eventually, the efficiency of the observatory becomes so low that the “nominal” observing program would have to be halted and replaced perhaps by a different one that monitored a small number of targets for extended periods of time. And targets would be restricted to those whose variability timescale was longer than the required observation times. Our basic science requirements

for the primary "typical" science targets require that a stellar surface imaging observation be complete in a period of ~5 hours to avoid smearing of the images due to stellar rotation, proper motion, and intrinsic variability of the active regions. Once the observation time exceeds that, then prime science begins to be lost and the number of suitable targets begins to decrease. There is no "magic number" at which it becomes impossible to observe, but below 20 elements the impact is very significant – so the goal should be to maintain the number of elements in the 20-30 range for the duration of the mission.

The beam combining hub is obviously a single-point failure whose failure could lead to a loss of mission scenario. The design concept addresses this in two ways. First, the hub is designed to be highly redundant at the component level, at least for all parts with plausible and significant failure scenarios over the mission lifetime. The design for the most critical components is modular, thus enabling in-situ servicing by robotic or human means. Second, it is highly desirable from both a redundancy viewpoint and an operational efficiency viewpoint to actually launch and use in normal operations two identical hubs. With two hubs, one can be in motion while the other is being used for an observation, and thus "pre-positioned" for the next target. When observation of the first target is finished, then a small change in the orientation of the array to line-up with the pre-positioned second hub is all that is needed to get setup for observation of the next target. An extreme case that well-illustrates the utility of a second hub is one in which it is actually positioned on the "back" side of the primary array, at the same distance as the first hub is positioned on the "front" side of the array (i.e., one system focal length distant, typically about 5 km, though sometimes as close as 1 km, sometimes as far as 10 km). With this setup, the observatory could actually switch which "half" of the sky is being observed, simply by flipping over the mirrorsats in-place, thus accomplishing a repointing half-way around the sky in minutes instead of what otherwise would take hours if not days (normally re-targeting will only move 10 degrees or less on the sky from target to target to avoid excessive propellant and time usage). The availability of a second hub would thus immensely increase the efficiency of the observatory at the same time as providing insurance against catastrophic loss-of-mission due to a failure of a single hub. An alternative, of course, is to have available a second (or third) hub on the ground ready for a launch-on-need should a failure in the primary hub(s) occur. This can enable a recovery from a hub failure, but at the cost of some down-time while the backup hub is launched and deployed at L2.

6.5 Maintenance or servicing

The SI baseline design does not require servicing at the Sun-Earth L2 site to achieve the mission goals and objectives. Nevertheless, the overall mission reliability and operations lifetime could benefit greatly from servicing. Servicing can replace key components of the mirrorsats and hub and refuel the spacecraft for station-keeping/orbit maintenance and target-to-target maneuvering over the desired long lifetime.

The level of modularity and serviceability of the numerous mirrorsats is something to be determined in a future study that would trade the ease and cost of producing extra mirrorsats to hold in reserve vs. the cost of making the minimum-set mirrorsats serviceable (or with redundant components).

The critical hub spacecraft is a single-point failure, unless more than one hub is launched (or is available for launch-on-need). A future systems optimization study would also be needed to determine the appropriate level of modularity on the critical hub spacecraft, considering the comparative value of the hub as a whole, the critical parts, the modularity impacts, and the user cost of a servicer vehicle visit.

Chapter 7: Safety

Safety engineers will identify hardware and software failure modes and hazards, along with their potential effects. The interface between safety engineering, design engineering, and R&M engineering should ensure timely communication of information on all potential hazards, and provide for the early identification and correction of problems inherent in the proposed design. Safety engineers will use R&M data to help develop hazard analyses that identify and address all hazards resulting from the failure modes. R&M engineering should review the hazard analyses to ensure that reliability and/or maintainability are not degraded by the resulting design recommendations.

For the SI mission the primary special failure modes involve possible spacecraft collisions, spacecraft drift and loss from the formation, and problems with system (including beam and position) control maintained by the Hub spacecraft.

7.1 Launch and near-Earth operations

The SI launch(es) will quickly get beyond Low Earth Orbit (LEO). The usual range safety considerations apply prior to leaving near-Earth space. No radioactive power sources or calibration sources are onboard the spacecraft in the baseline design.

7.2 Planetary protection

The SI will deployed into a Lissajous orbit around the Sun-Earth L2 point. There will be no non-Earth planetary encounters and thus no "planetary protection" issues.

7.3 End of mission safety issues

The individual spacecraft will eventually leave their unstable L2 halo orbits after loss of station keeping ability caused by depletion of on-board propellant. The individual spacecraft will drift off into separate solar orbits that do not intersect the Earth for many years. A statistical analysis of the departure orbit will need to be performed to provide a timeframe but a number on the order of several thousand years is plausible.

References

Allen, R. J., 2005, *Station-Keeping Requirements for Constellations of Free-Flying Collectors Used in Astronomical Imaging*, in AIAA Journal of Guidance, Control and Dynamics (submitted)

Anderson, G. 2005, *A Large Optical Photon Sieve*, Optics Letters, in press

Baldwin, J. E.: 1996, in K. G. Strassmeier and J. L. Linsky (Eds.), *Stellar Surface Structure,* IAU Symp. 176, Kluwer Academic Publishers, Dordrecht, The Netherlands, p. 139

Baliunas, S. L., Donahue, R. A., Soon, W., and Henry, G. W.: 1998, in R. A. Donahue and J. A. Bookbinder (Eds.), *Cool Stars, Stellar Systems, and the Sun,* ASP Conf. Ser. 154, Astron. Soc. Pacific, San Francisco, p. 153

Balona, L. A., and Engelbrecht, C. A.: 1986. *MNRAS* **219**, 131

Barbossa, C. L., and Figer, D.: 2004, arXiv.org/abs/astro-ph/0408491

Barrett, H. H. and Myers, K. J.: 2004, Foundations of Image Science, (NJ: Wiley Interscience)

Barton, I.M., et. al. 2001, *Fabrication of large-aperture light-weight diffractive lenses for use in space*, Applied Optics 40, 447-451

Basri, G.: 1987, *ApJ* **316,** 377

Basu, S., Antia, H. M., and Tripathy, S. C.: 1999, *ApJ* **512,** 458

Bei Technologies, Inc., http://www.beikimco.com/products/actuator/guide/BEIKimco , VCA-AppGuide2002.pdf.

Berio, P., Stee, P., Vakili, F., et al.: 1999, *A&A* **345**, 203

Bjorkman, J. E.: 2000, in *The Be Phenomenon in Early-type Stars*, ed. M. A. Smith, H. F. Henrichs, J. Fabregat, IAU Colloq. 175, ASP Conf. Ser. **214**, 435

Braun, D. C. and Fan, Y.: 1998, *ApJL* **508,** 10

Breckinridge, J., Lorre, J., and Bryant, 1999, *Image Contrast Constraints Upon Fill Factor, CCD Fullwell, and Exposure Time*, Internal JPL report.

Brown, J. C., Telfer, D., Li, Q., Hanuschik, R., Cassinelli, J. P., and Kholtygin, A.: 2004, *MNRAS* **352**, 1061

Cao, Q. and Jahns, J. 2002, *Focusing analysis of the pinhole photon sieve: individual far field model*, J. Opt. Soc. Am A 19, 2387-2393

Cao, Q. and Jahns, J. 2003a, *Non-paraxial model for the focusing of high-numerical aperture photon sieves*, J. Opt. Soc. Am A. 20 1005-1012

Cao, Q. and Jahns, J. 2003b, *Modified Fresnel zone plates that produce sharp Gaussian focal points*, J. Opt. Soc. Am A 20, 1576-1581

Cao, Q. and Jahns, J. 2004, *Comprehensive focusing analysis of Various Fresnel Zone plates*, J. Opt. Soc. Amer. A 21, 561-571

Chen, P., Carpenter, K.G., Romeo, R.C., 2003. in "Future EUV/UV and Visible Space Astrophysics Missions and Instrumentation". Edited by J. Chris Blades and Oswald H. W. Siegmund, Proceedings of the SPIE, Volume 4854, pp. 21-28

Claret, A.: 2004, Astron. & Ap., 424, 919

Cranmer, S. R., and Owocki, S. P.: 1996, *ApJ* **462**,469

Cuntz, M. Saar, S. H., and Musielak, Z. E.: 2000, *ApJ* **533**, L151

Davidson, K., and Humphreys, R. M.: 1997, *Ann. Rev. Astron. Astrophys.* **35**, 1

Dessart, L.: 2004, *A&A* **423**, 693

Devine, D. 2000, *ApJ* **579**, 800

Dikpati, M. and Gilman, P.: 2001, *ApJ* **559,** 428

Djurasevic, G., Rovithis-Livaniou, H., Rovithis, P., Georgiades, N. Erla[oc. S/. and Pavlovic, R.: 2003, *A&A* **402**, 667

Donati, J.-F.: 1996, in J.-F. Donati (Ed.), *Solar and stellar dynamos; constraints from observations,* Observatoire Midi-Pyréne's, Toulouse, France, p. 83

Dorch, S. B. F. and Nordlund, Å.: 2001, in P. Brekke, B. Fleck, and J. B. Gurman (Eds.), Recent insights into the physics of the Sun and the heliosphere - Highlights from SOHO and other space missions, Conference series Vol. 200, American Soc. Pac., San Francisco, 186

Dorch, S. B. F. and Nordlund, Å.: 2001, A&A **365,** 562

Dulk, G. A. and Gary, D. E.: 1983, A&A **124,** 103

Elvis, M. and Karovska, M.: 2002, ApJ, 581, L67

Fienup, J. R., Griffith, D., Harrington, L., Kowalczyk, A.M., Miller, J. J., and Mooney, J. A. 2002, *Comparison of Reconstruction Algorithms for Images from Sparse-Aperture Systems*, Proc. SPIE 4792-0 Image Reconstruction from Incomplete Data Conference, Seattle WA.

Fienup, J. R. 2000, *MTF and Integration Time versus Fill Factor for Sparse Aperture Imaging Systems*, Proc. SPIE 4091, Imaging Technology and Telescopes, J. W. Bilbro editor, pp. 43 – 47.

Fiete, R.D., Tantalo, T.A., Calus, J.R., Mooney, J.A. 2002, *Image quality of sparse-aperture designs for remote sensing*, Optical Engineering, Vol. 41 No. 8, pp. 1957–1969.

Freedman, W. et al.: 2001, ApJ, 553, 47

Frick, P., Baliunas, S.L., Galyagin, D., Sokoloff, D., and Soon, W.: 1997, ApJ, 483, 426

Fullerton, A., and Kaper, L.: 1995, Hot/Massive Star Newsletter, no. 15, http://www.astroscu.unam.mx/massive_stars/

Gatewood, G., Kiewiet de Jonge, J., and Stephenson, B.: 1993, *PASP* **105**, 1101

Gayley, K. G., Owocki, S. P., and Cranmer, S. R.: 1995, *ApJ* **442**, 296

Georgiev, L. N., Koenigsberger, G., Ivanov, M. M., St.-Louis, N., and Cardona, O.: 1999, *A&A* **347**, 583

Gilliland, R. L. and Dupree, A. K.: 1996, ApJL **463,** 29

Glatzel, W., Kiriakidis, M., and Fricke, K. J.: 1993, *MNRAS* **262**, L7

Golomb, S. 1982 . IEEE Trans. On Information Theory, vol. IT-28, No. 4, 600

Gonsalves, R.A. *Phase Retrieval and Diversity in Adaptive Optics*, Optical Engineering, Vol. 21, p 829, 1982

Gonzalez Hernandez, I., Patr´on, J., Bogart, R. S., and The SOI Ring Diagram Team: 1999, ApJL **510,** L53

Goode, Ph. R., Qin, J., Yurcheshyn, V., Hicky, J., Chu, M.-C., e. Kolbe, Brown, C. T., and Koonin, S. E.: 2001, Geophys. Res. Lett. **28,** 1671

Grady, C., Woodgate, B., Torres, C. A. O., Henning, Th., Apal, D., Rodman, J., Wang, H., Stecklum, B., Linz, H., Williger, G. M., Brown, A., Wilinson, E., Harper, G. M., Herczeg, G. J., Danks, A., Viera, G. L., Malamuth, E., Collins, N. R., and Hill, R. S.: 2004, *ApJ* **608**, 809

Grady, C., et al.: 2005, *ApJ* **630**, 958

Grogan, R. L., Blackwood, G. H., and Calvet, R. J. 1998, Optical delay line nanometer level pathlength control law design for space-based interferometry. SPIE Int. Symposium on Astronomical Interferometry, April 1998.

Hadaegh, F. 2005, private communication

Hartigan, P., Edwards, S., and Ghandour, L.: 1995, *ApJ* **452**, 736

Hartmann, L. 2004, in *Star Formation at High Angular Resolution,* **IAU Symp. 221** (ed. M. Burton, R. Jayawardhana, T Bourke), 201

Hathaway, D. H.: 1996, ApJ **460,** 1027

Hench, J.J., Lurie, B, Grogan, R. and Johnson, R. 2000, Implementation of nonlinear control laws for an optical delay line. IEEE Aerospace Conference, March 2000.

Henyey, L. and Greenstein, J. 1941, ApJ **93**, 70

Hill, H. G. M., Grady, C. A., Nuth, J. A. III, Hallenbeck, S. L., and Sitko, M. L.: 2001, *Proc. NAS* **98**, 2182

Hubrig, S., Scholler, M., and Yudin, R. V.: 2004, *A&A* **428**, L1

Hyde, R., Dixit, S., Weisenberg, A., Rushford, M. 2002. *Eyeglass: a very large aperture diffractive space telescope*, in Proceedings of the SPIE vol.4849: 28-39

Hyde, R.A., 1999. "Eyeglass.1. Very large aperture diffractive telescopes", Applied Optics **38**, 4198-4212

Hussain, G. A. J., Donati, J.-F., Collier Cameron, A., and Barnes, J. R.: 2000, *MNRAS* **318,** 961

Ignace, R., Cassinelli, J. P., and Bjorkman, J. E.: 1998, *ApJ* **505**, 910

Jennison, R.C.: 1958, *MNRAS*, 118, 276

Jennison, R.C.: 1961, *Proc. Phys. Soc*., 78, 596

Jennison, R.C. and Latham, V.: 1959, *MNRAS*, 119, 174

Jordan, C., and Montesino, B.: 1991, *MNRAS* **252**, 21

Karovska, M., Schlegel, E., Hack, W., Raymond, J.C., Wood, B.E., 2005, ApJ, 623, L137

Kaspi, S., Smith, P. S., Netzer, H., Maoz, D., Jannuzi, B. T., & Giveon, U.: 2000, *ApJ*, 533, 631

Kendrick, R.L., Acton, D.S., and Duncan, A.L. 1994, *Phase Diversity Wavefront Sensor for Imaging Systems*, Applied Optics, Vol. 33, No 27, 6533-6546

Kipp, L., Skibowski, M., Johnson, R.L., Berndt, R., Adelug, R., Harm, S. and Seemann, R. 2001, *Sharper Images by Focusing Soft X-Rays With Photon Sieves*, Nature 414, 184-188

Kitchatinov, L. L. and R¨udiger, G.: 1995, *A&A* **299,** 446

Krishnamurthi, A., Pinsonneault, M. H., Barnes, S., and Sofia, S.: 1997, *ApJ* **480,** 303

Kroll, P., and Hanuschik, R. W.: 1997, in *Accretion Phenomena and Related Outflows*, ed. Wickramasinghe D. T., Bicknell, G. V., Ferrario, L., Proc. IAU Colloq. 163, ASP Conf. Ser. 121, 494

Kudritzki, R.-P., and Puls, J.: 2000, "Winds from hot stars", *Ann. Rev. Astron. Astrophys.* **38**, 613

Küker, M., Rüdiger, G., and Schultz, M.: 2001, *A&A* **374,** 301

Kuo, C.P.: 1991, SPIE **1542**, 420

Labeyrie, A. 2002. in *Beyond conventional adaptive optics : a conference devoted to the development of adaptive optics for extremely large telescopes*. Proceedings of the Topical Meeting held May 7-10, 2001, Venice, Italy. Edited by E. Vernet, R. Ragazzoni, S. Esposito, and N. Hubin. Garching, Germany: European Southern Observatory, ESO Conference and Workshop Proceedings, Vol. 58, ISBN 3923524617, p. 473

Lay, O.P., Dubovitsky, S., Peters, R. D., Burger, J. P., Ahn, S.-W., Steier, W. H., Fetterman, H. R., Chang, Y. "MSTAR: a submicrometer absolute metrology system", 2003, Opt. Lett. 28, 890

Leinert, Ch., et al. 1999, A&AS, **127**, 1

Lee, U. and Saio, H. 1990, *ApJ* [**349**, 570][**360**, 590]

Levashov, V. E. and Vinogradoz, A. V. 1994, *Analytical theory of zone plate efficiency*, Phys. Rev. E 49, 5797-5803

Linsky, J.L.: 1986, NASA STI/Recon Technical Report.

Liu, K. 2003, *Stochastic performance analysis and staged control system designs for space based interferometers*. PhD dissertation, MIT, Department of Aeronautics and Astronautics.

Lockwood, G. W., Thompson, D. T., Lutz, B. L., and Howell, E. S.: 1991, *ApJ* **368,** 287

Loudon, R., *The Quantum Theory of Light*, Oxford Science Publications, 1985

Lyon, R.G., Carpenter, K.G., Petrone, P., Dogoda, P., Liiva, P., Marzouk, J., Solyar, G., Mazzuca, L., 2004a, "Fizeau Interferometry Testbed Wavefront Control", in Optical, Infrared, and Millimeter Space Telescopes, SPIE Proc., Vol 5487.

Lyon, R.G., Carpenter, K.G., Huet, H., Cottle, P., Petrone, P., Dogoda, P., Liiva, P., Marzouk, J., Solyar, G., Mazzuca, L., Zhang, X. 2004b, ``Building the Fizeau Interferometer Testbed", Proc. IEEE Aerospace Conference

Maeder, A. and Meynet, G. 2000a, ``The evolution of rotating stars," Ann. Rev. Astron. Ap., 38, 143

Maeder, A. and Meynet, G. 2000b, A&A, 361, 159

Markiel, J. A. and Thomas, J. H. 1999, *ApJ* **523,** 827

Maschiadri, E., and Raga, A. C.: 2004, *ApJ* **615**, 850

Mathis, J.S. and Lee, C.W. 1991, *ApJ* **376**, 490

Mazzuca, L., Carpenter, K.G., Lyon, R.G., Petrone, P., Huet, H., Cottle, P., Dogoda, P., Liiva, P., Marzouk, J., Solyar, G., Mozurkewich, D., Armstrong, J. T., Zhang, X., Mundy, L., 2004, "The Fizeau Interferometer Testbed(FIT): Developing and Testing the Technologies Needed for Space-Based Interferometric Imaging Systems", in New Frontiers in Stellar Interferometry, SPIE proceedings, Vol 5491.

McGroarty, F., and Ray, T. P.: 2004a, *A&A* **420**, 975

McGroarty, F, Ray, T. P., and Bally, J.: 2004b, *A&A* **415**, 189

Meinel, A. B., and Meinel, M. P. 2002, *Parametric dependencies of high diffraction order achromatized aplanatic configuration that employ circular of crossed linear diffractive optical elements.* Applied Optics, 41, 7155-7166

Meinel, A. B., and Meinel, M. P. 2003, *Large membrane space optics: imagery and aberrations of diffractive and holographic achromatized optical elements of high diffraction order*, Opt. Engr. 41, 1995-2007

Miller, N. A., Cassinelli, J.P., Waldron, W.L., MacFarlane, J.J., and Cohen, D.H.: 2002, *ApJ*, 577, 951

Monet, D., Dahn, C. C., Vrba, F. J., Harris, H. C., Pier, J. R., Luginbuhl, C. B., and Ables, H. D. 1992, *AJ* **103**, 638

Mullan, D. J. 1984, *ApJ* **282**, 603

Noecker, M.C., Phillips, J.D., Babcock, R.W., and Reasenberg, R.D. "Internal laser metrology for POINTS," in *The Proceedings of the SPIE Conference # 1947 on Spaceborne Interferometry*, (Orlando, FL, April 14-16, 1993), Vol. 1947, p. 174, 1993.

Noyes, R.W., [Weiss, N. O., and Vaughn, A. H.] [Hartman, L. W., Baliunas, S. L., Duncan, D. K., and Vaughn, A. H.] 1984, *ApJ* [**287**, 769] [**279**, 763]

Okazaki, A. T.: 1991, *PASJ* **43**, 75

Oskinova, L. M., Feldmeier, A., and Hamann, W.-R.: 2004, *A&A* **422**, 675

Ossendrijver, A. J. H. and Hoyng, P.: 1997, *A&A* **324,** 3290

Owocki, S. P., and Cranmer, S. R.: 2002, in *Radial and Nonradial Pulsations as Probes of Stellar Physics*, ed. C. Aerts, T. Bedding, and J. Christensen-Dalsgaard, ASP Conf. Ser. 259, IAU Colloq. 185, 512

Owocki, S. P., Gayley, K. G., and Shaviv, N. J.; 2004, *ApJ* **616**, 525

Owocki, S. P., and Puls, J.: 1999, *ApJ* **510**, 355

Panagia, N., Gilmozzi, R.; Macchetto, F.; Adorf, H.-M.; Kirshner, R. P..: 1991, ApJ Letters, 380, 23

Paresce, F., and Jakobsen, P., 1980 Nature, 288, 119

Parker, E. N.: 1955, *ApJ* **122,** 293

Perlmutter, S., Turner, M. S. and White, M.: 1999, *PhysRevLett* **83**, 670

Pearson, K.J., Horne, K. and Skidmore, W. 2003, *MNRAS* **338**, 1067

Peterson, B. M. 1993, PASP, 105, 247

Peterson B. M. 1997, An Introduction to Active Galactic Nuclei (Cambridge: Cambridge Univ. Press)

Peterson, B.: 2001, "Variability of AGN", in Starburst-AGN Connection", Singapore: World Scientific (also: arXiv:astro-ph/0109495)

Petrone, P., Clark, D., Lyon, R.G., Carpenter, K.G., Huet, H., Cottle, P., Dogoda, P., Liiva, P., Marzouk, J., Solyar, G., Mazzuca, L., Mozurkewich, D., 2004, "Fabrication and Characterization of the Fizeau Interferometer Testbed", in New Frontiers in Stellar Interferometry, SPIE proceedings, Vol 5491.

Phillips, J.D and Reasenberg, R.D.: 2005, "Tracking Frequency Laser Distance Gauge," *Review of Scientific Instruments*, **76**, 064501

Phillips, J.D. and Reasenberg, R.D.: 2004, "Toward a spaceworthy picometer laser gauge," in the *Proceedings of the SPIE Conference #5495 on Astronomical Telescopes and Instrumentation,* (Glasgow, 21-25 June, 2004), 320-327

Physik Instrumente. http://www.physikinstrumente.de/products/section4/content.php.

Porter, J. M., and Rivinius, T.: 2003, *PASP* **115**, 1153

Quirrenbach, A. et al.: 1997, *ApJ* **479**, 477

Rajagopal, J., Bőoker, T., Allen, Ronald J., and Carpenter, K.G., 2003, "Simulating Dilute-Aperture Imaging: The Stellar Imager", in *Astronomical Telescopes and Instrumentation: Interferometry in Space*, ed. M. Shao (Proc. SPIE, Vol. 4852), 652 - 656

Reasenberg, R.D., Babcock, R.W., Chandler, J.F., Gorenstein, M.V., Huchra, J.P., Pearlman, M.R., Shapiro, I.I., Taylor, R.S., Bender, P., Buffington, A., Carney, B., Hughes, J.A., Johnston, K.J., Jones, B.F., and Matson, L.E.: 1988, "Microarcsecond Optical Astrometry: An Instrument and Its Astrophysical Applications," *Astron. J.*, **32**, 1731-1745

Reasenberg, R.D., Phillips, J.D., Noecker, M.C.: 1995, "High Precision Interferometric Distance Gauge," U.S. Patent 5,412,474, issued 2 May

Reasenberg, R.D. and Phillips, J.D.: 2001, "Testing the equivalence principle on a trampoline," *Classical and Quantum Gravity* **18**, 2435-2445

Reasenberg, R.D. and Phillips, J.D.: 2005, "A Laboratory Free-Fall Test of the Equivalence Principle – POEM," in *Proceedings of the Third Meeting on CPT and Lorentz Symmetry*, (Indiana University, Bloomington, 4-7 August 2004), V. Alan Kostelecky, ed., World Scientific, New Jersey

Reiners, A., Stahl, O. Wolf, B., Kaufer, A., and Tivinius, T.: 2000, *A&A* **333**, 125

Ribak, E., Roddier, C., Roddier, F., and Breckinridge, J. 1988, *Signal-to-noise limitations in white-light holography*, Applied Optics, 27, 1183-1186

Richards, M. T. and Ratliff, M. A.: 1998, *ApJ* **493**, 326

Rivinius, T., Baade, D., Stefl, S., Stahl, O., Wolf, B. and Kaufer, A.: 1998, *A&A* **333**, 125

Rivinius, T., Baade, D., Stefl, S., Townsend, R.H.D., Stahl, O., Wolf, B., Kaufer, A: 2001, A.&Ap., 369, 1058

Rogers, A. E. E.; Hinteregger, H. F.; Whitney, A. R.; Counselman, C. C.; Shapiro, I. I.; Wittels, J. J.; Klemperer, W. K.; Warnock, W. W.; Clark, T. A.; Hutton, L. K.: 1974, *ApJ*, 193, 293

Rosner, R., Musielak, Z. E., Cattaneo, F., Moore, R. L., and Suess, S. T.: 1995, *ApJ* **442**, L25

Saar, S. H. And Osten, R. A.: 1997, *MNRAS* **284**, 874

Saar, S. H. and Brandenburg, A.: 1999, *ApJ* **524**, 295

Savonije, G. J.: 1998, in *Cyclical Variability in Stellar Winds*, ed. L. Kaper and A. Fullerton (Berlin: Springer), 337

Sasselov, D. D. and Karovska, M.: 1994, *ApJ* **432**, 367

Scheuer, P.A.G. 1968, *Nature* **218**, 920

Schou, J. and Bogart, R. S.: 1998, *ApJL* **504,** 131

Schiminovich, D., Friedman, P.G., Martin, C., and Morrissey, P. F., 2001, ApJ, 563, L161

Schrijver, C. J.: 1996, in K. G. Strassmeier and J. L. Linsky (Eds.), *Stellar Surface Structure,* IAU Symp. 176, Kluwer Academic Publishers, Dordrecht, The Netherlands, p. 1

Schrijver, C. J. and Pols, O. R.: 1993, *A&A* **278,** 51, (Erratum in A& A 293, 640 (1995))

Schrijver, C. J. and Zwaan, C.: 2000, *Solar and Stellar Magnetic Activity,* Cambridge University Press, Cambridge, U.K.

Schwab, K., Bruckner, N., and Packard, R. E.: 1997, "Detection of the Earth's Rotation Using Superfluid Phase Coherence," Nature **386**, pp. 585-587

Shkolnik, E., Walker, G. A. H., and Bohlender, D. A.: 2003, *ApJ* **597**, 1092

Shkolnik, E., Walker, G. A. H., Bohlender, D. A., Gu, P.-G., & Kűrster, M. 2005, ApJ, 622, 1075

Shu, F., Shang, H., Gounelle, M., Glassgold, A. E., and Lee, T.: 2001, *ApJ* **548**, 1029

Smith, M. A., Hubeny, I, Lanz, T., and Meylan, T.: 1994, *ApJ* **432**, 392

Soon, W.-H. and Yaskell S. H., 2004, in "Maunder Minimum and the Variable Sun-Earth Connection", World Scientific Publishing, 296pp.

Stedman, G. E.: 1997, "Ring-laser tests of fundamental physics and geophysics," *Reports on Progress in Physics,* **60**, 615-688

St.-Louis, N., Dalton, M. J., Marchenko, S. V., Moffat, A. F. J., and Willis, A. J.: 1995, *ApJ* **452**, L57

Stoeckley, T. R., and Buscomb, W.: 1987, *MNRAS* **227**, 801

Tajmar, M., Genovese, A. and Steiger, W. 2004, *Indium Field Emission Electric Propulsion Microthruster Experimental Characterization*. Journal of Propulsion and Power, Vol. 20, No. 2, March–April 2004.

Tobias, S. M., Brummell, N. H., Clune, Th. L., and Toomre, J.: 2001, *ApJ* **549,** 1183

Townsend, R. H. D. 2003, MNRAS, 340, 1020

Townsend, R. H. D., Owocki, S. P., and Howarth, I. D.: 2004, *MNRAS* **350**, 189

van Boekel, R., Kervella, P., Schöller, M. et al.: 2003, *A&A* **410**, L37

von Zeipel, H.: 1924, *MNRAS* **84**, 665

Waldman, G. S. 1966, *Variations on the Fresnel Zone Plate*, JOSA, 56, 215-218

West, J. 1998, *Attitude Control for the Geostorm Warning Mission*, internal JPL engineering design memorandum, 12 October 1998

Wieman, C. E. and Hollberg, L. 1991, in Review of Scientific Instruments, vol. 62, Jan. 1991, p. 1-20.

Zhang, X., Carpenter, K.G., Lyon, R.G., Huet, H., Marzouk, J., Solyar, G., 2003, ``The Fizeau Interferometer Testbed", Proc. IEEE Aerospace Conference.

Zwaan, C.: 1996, *SPh* **169,** 265

Appendix A: Education and Public Outreach

A.1 Introduction

Stars illuminate our night skies, fill our literature, and populate our artwork, music, and popular speech. They are topics of both public fascination and a key component in astronomy curricula, beginning at the elementary level. Yet, understanding stars as more than points of light has been very challenging, both for the general public and for students, in large part due to the lack of spatially resolved imagery.

The technical advances represented by the Stellar Imager will provide us with the opportunity to view stars closer than 4 pc much as we have historically viewed the Sun, and, for the first time to resolve main sequence stars out to 44pc and young stars out to 88-100 pc, providing an opportunity to see how the stellar atmospheric structure changes dynamically with rotation, vibration, stellar activity cycles, and ultimately through the entire course of stellar evolution. E/PO efforts in astronomy are critical in conveying to students and the public this process of discovery and the excitement of understanding new phenomena.

A.2 Innovation for SI: The SI "Virtual Mall"

We are sufficiently early in the planning for the Stellar Imager that it is premature to specify the technology and tools needed to implement an E/PO plan. However, we can organize material and develop an information outreach paradigm which is useful both in identifying the capabilities that will be needed, and in ensuring that it truly provides a "science for all" experience. Humans learn new material most efficiently when it is presented in a familiar context, and when it is anchored in what they already know. Therefore, the paradigm should incorporate familiar functions or activities, should be flexible, with information presented in multiple ways, and should incorporate the best practices of universal design to make it possible. Many of the capabilities needed here already exist: we only seek to merge them seamlessly.

One paradigm which might meet these needs would be to organize access to the SI datasets as an information/technology virtual mall. Such a mall could include the following:

- A newsstand providing a picture of the day, press releases, topical podcasts (or then-current equivalent) and any breaking news regarding the mission.
- A picture gallery which could include posters, still images using SI data, and the press release images.
- A sculpture gallery that provides images in formats which can be printed with braille printers, and 3D file formats suitable for CAD/CAM machines (or equivalent) to produce small 3D models of the stars.
- A movie theater can provide previously produced movies. This approach can provide stars as a function of time, stars as a function of wavelength, stellar activity cycles, stellar evolution, stars as a function of mass, binary star orbital motions, eclipses by hot Jupiters, eruptive phenomena, flybys and animations.
- A multimedia production studio, which could enable users to pick a still image of a particular star, or from a date with significance to the user, generate movies incorporating data of interest

to the user using state-of-the-art visualization tools. This area could also enable the user to generate fly-bys of a particular object and combine the SI data with data resources for circumstellar disks and planets associated with that program star, thus facilitating study of the star-exoplanet connection.

- A classroom/university, which can provide teachers with lesson plans and slide sets with good captions, for example. All of these will be sorted by grade level to make them most useful. We would propose developing these plans with the aide of teacher interns.
- A library containing reference materials on the scientific background, illustrative examples and links to the extensive materials on the Sun. This is an area where hypermedia presentations and e-books can be derived, providing state- of-the-art materials to interested readers world-wide.
- An information kiosk for stellar imager could provide information on where the science targets are in the sky, which of them are known to have planetary systems (and what they are like, what the space weather forecast for one of the planets in that system might be, etc.), as well as the more conventional site map information and internal search facility.

The launch of the SI Mission ~2024 will mark an important milestone in our views of the local universe. Heretofore, we have dealt with solar system exploration and sun-earth connection issues in isolation from an interstellar context. With SI and the continued discovery of extra-solar planets, that will be the era of thinking of distant suns and planetary systems as a familiar extension of our own circumstances. This will allow us to develop exciting new types of classroom and museum resources that cross the boundary between contemporary science fiction, and the reality of living in that wondrous future.

The SI Mall will be designed to serve a variety of communities including formal and informal education and the news media. The education communities will be able to convey the idea that our sun is a star, and that other stars have similar activity cycles and recognizable phenomena. This will serve as a thematic bridge to the Sun-Earth Connection community and their emphasis on space weather issues.

We anticipate that current trends will continue in the manner in which information is delivered to the public using the World Wide Web or its next incarnation. Current online resources will benefit from years of continued innovation in web technology, so that we will truly be working with a ‘paperless’ communication environment by the time SI is launched and students and the public begin to explore its data and results.

A.3 Conventional Outreach Activities for SI

Once the SI Mission Virtual Mall EPO resource is in place, we will work closely with teachers and museum staff to create specific products tailored to their unique needs. Current missions (ca 2005) have adopted the universally accepted concept that teachers will not use your resources unless you train them. To that end, we will provide teacher workshops at national conventions (NSTA, NCTM, SACNAS, etc). We will also host workshops for local teachers at each of the participating mission centers of the Co-Is. We will offer accredited mini-courses to teachers to upgrade their background in astronomy and solar/stellar physics. Teachers are strongly motivated to participate in a science workshop if they receive something in return, and credit towards a pay raise or advanced certificate in teaching is a powerful inducement to learn the material and to use it later. But to be truly successful in impacting the education of 26 million students, we will need to do much more that train teachers to use our curriculum guides.

A major formal education objective will be to work directly with teacher education at the state and national level. We will also work with textbook companies to enhance their materials. Teachers can be motivated to use SI materials only if they see a pressing need to change the way in which they are currently teaching. A powerful inducement for change is when a Superintendent or a state curriculum standard actually requires a teacher to incorporate a new 'SI-friendly' concept. Moreover, pre-service teacher education is an area that offers tremendous potential for impacting how a teacher teaches, and which type of content they choose to include.

Not to be forgotten are a number of already-existing NASA resources that will also be available by the time of SI launch. NASA/CONNECT designs 10 TV programs each year that are broadcast to 8 million students and 300,000 teachers annually. We will collaborate with CONNECT to create one new program each year that will feature SI science in an exciting way. We will also work with programs such as the *Why File*s and *Destination:Tomorrow* which also have sizable audiences. These TV programs are 'pricey' at a cost of $90,000 for CONNECT, but their impact is typically in the millions of viewers. This kind of leverage for a mission is crucial in getting SI EPO resources into the hands of teachers. We have already worked with these programs as part of the outreach activities of the Sun-Earth Connection Education Forum. We are well known to the Producers, and have participated as co-hosts in many of these programs. A similar collaboration will be set up for SI. In addition, NASA's new digital education channel will be debuting by 2007, so we will have the opportunity to create short 20-minute TV programs to further distribute our science results to the formal and informal community.

We will work very closely with the TV, radio and newsprint media to bring our science to the attention of the general public. We will facilitate this by creating an archive of high-definition video and still images of appropriate topics, which we will produce in consultation with the NASA/GSFC Public Affairs Office, with whom we have worked in missions such as SoHO, IMAGE, Polar, COBE etc. We will create drafts of press releases that will be available to the PAO office to evaluate and edit as needed. We will create a timely archive of 'Recent Discoveries', and a library of Information Briefs that define and describe specific issues, technologies, science concepts that are central to SI and its mission. We will also work with radio programs such as Earth and Sky Radio, NPR 'Science Friday' or the future programs that replace them to educate the general public about stellar astronomy and 'the sun as a star'.

All of the aforementioned ideas sound exotic and remote, but in fact they are all based on programs which the current EPO leads are already implementing during their tenure with existing NASA Missions. This, however, does not mean that what we are proposing is merely an 'incremental' and perhaps uninspired extension of existing EPO strategies. SI will be a unique instrument, and its mission will be an historic exploration of distant stars, rendering them into resolvable cousins to our own sun. To meet the grandeur of this opportunity, the EPO effort we are envisioning for SI will be at least as dramatic as what was employed so successfully with the Hubble Space Telescope. Just as it would have been impossible to anticipate the central role played by the Internet for Hubble in 1978, so too must we be a bit patient and cautious in anticipating what new resources will be available to SI in EPO some 10 years hence. We cannot know what directions teaching will take, or which technologies will be best suited to meet these goals. Nevertheless, no matter what the new resources may be, the EPO activities of SI will adapt to use them.

Appendix B

Resources/suggested reading

Web pages with information pertinent to the Stellar Imager (precede URLs below with "http://")

www.nas.edu/ssb/cover.html	CSSP/CSTR report entitled "Space Weather; A Research Perspective
www.lmsal.com/sec/Roadmap	Sun-Earth Connection: Roadmap 2003-2028
sec.gsfc.nasa.gov/lws.htm	NASA/GSFC home pages for the "Living with a star" initiative
www.ispe.arizona.edu/-conferences/sunmeet/	Proceedings of a NASA workshop on Sun-Climate Connections

Web pages with information pertinent to other space-based interferometric missions

SIM	Planetquest.jpl.nasa.gov/SIM/sim_index.html
TPF	Planetquest.jpl.nasa.gov/TPF/tpf_index.html
Darwin	Sci.esa.int/darwin/

Web pages with information pertinent to ground-based interferometric missions

CHARA	www.chara.gsu.edu/CHARA/
COAST	www.mrao.cam.ac.uk/telescopes/coast/index.html
KI	planetquest.jpl.nasa.gov/Keck/kick_index.html
LBT	medusa.as.arizona.edu/lbtwww/lbt.html
MRO	mro.nmt.edu/
NPOI	ftp.nofs.navy.mil/projects/npoi/
VLTI	www.eso.org/projects/vlti

Appendix C

Acronyms

ADCS	Attitude Determination and Control System
AGN	Active Galactic Nuclei
ASIC	Application Specific Integrated Circuit
AU	Astronomical Unit
BATC	Ball Aerospace Technology Company
BELR	Broad Emission Line Regions
BLR	Broad Line Region
CCD	Charge Coupled Device
CC&DH	Command, Control and Data Handling
CCSDS	Consultative Committee for Space Data Systems
CDG	Concept Development Group
CPU	Central Processing Unit
CRD	Critical Design Review
CfA	Center for Astrophysics
CHARA	Center for High Angular Resolution Astronomy
COAST	Cambridge Optical Aperture Synthesis Telescope
COROT	Convection Rotation and Planetary Transits
COTS	Commercial Off-The-Shelf
DAC	Discrete Absorption Components
DI	Doppler Imaging
DNA	Deoxyribonucleic acid
DRM	Design Reference Mission
DSN	Deep Space Network
EP	Electrical Power
E/PO	Education and Public Outreach
EUV	Extreme Ultraviolet
FDF	Flight Dynamics Facility
FEEP	Field Emission Electric Propulsion
FDIR	Fault Detection, Isolation and Recovery
FIT	Fizeau Interferometer Testbed
FMEA	Failure Mode and Effects Analysis
FOT	Flight Operations Team
FSW	Flight Software

FUSE	Far Ultraviolet Spectroscopic Explorer
GNC	Guidance, Navigation and Control
GPS	Global Positioning System
GSFC	Goddard Space Flight Center
H/W	Hardware
IDC	Integrated Design Center
IFL	Input Feedback Linearization
IMDC	Integrated Mission Design Center
IP	Internet Protocol
IR&D	In-house Research and Development
ISAL	Instrument Synthesis and Analysis Lab
ISM	Interstellar Medium
I&T	Integration and Testing
JIM	Jitter Isolation Mount
JPL	Jet Propulsion Laboratory
JWST	James Webb Space Telescope
KI	Keck Interferometer
L2	2nd Lagrangian point
LBT	Large Binocular Telescope
LF	Life Finder
LGA	Low Gain Antenna
LISA	Laser Interferometer Space Antenna
LMC	Large Magellanic Cloud
LMATC	Lockheed Martin Advanced Technologies Center
LMMS	Lockheed Martin Missiles and Space
LQR	Linear Quadratic Regulators
MAXIM	Micro-Arcsecond X-ray Imaging Mission
MIPS	Millions of Instructions per Second
MOC	Mission Operations Center
MOST	Microvariability and Oscillations of Stars
MRO	Magdalena Ridge Observatory
MSFC	Marshall Space Flight Center
MTBF	Mean Time Between Failures
NASA	National Aeronautics and Space Administration
NRL	Naval Research Laboratory
NPOI	Naval Prototype Optical Interferometer
OPD	Optical Path Difference
OS	Observation Support
PDR	Preliminary Design Review
PDT	Phase Diverse Testbed

PI	Planet Imager
PSF	Point Spread Function
QSO	Quasi-stellar Object
R&M	Reliability and Maintainability
RAM	Random Access Memory
RCS	Reaction Control System
RF	Radio Frequency
RNA	Ribonucleic acid
S/C	Spacecraft
SCU	Spacecraft Control Unit
SEC	Sun-Earth Connection
SEU	Structure and Evolution of the Universe
SI	Stellar Imager
SIM	Space Interferometry Mission
SOC	Science Operations Center
SPECS	Sub-millimeter Probe of the Evolution of Cosmic Structure
SOHO	Solar and Heliospheric Observatory
SOW	Statement of Work
SRR	Science Requirements Review
SSR	Solid State Recorder
STScI	Space Telescope Science Institute
STIS	Space Telescope Imaging Spectrograph
SUNBEAMS	Students Becoming Excited About Math and Science
S/W	Software
T&E	Test and Evaluation
TCS	Thermal Control System
TFG	Tracking Frequency Gauge
TPF	Terrestrial Planet Finder
TRACE	Transition Region and Coronal Explorer
UMD	University of Maryland
UV	Ultraviolet
VLBI	the Very Long Baseline Interferometer
VLTI	the Very Large Telescope Interferometer
WCT	Wavefront Control Testbed
WITT	Wide-Field Imaging Interferometry Testbed
XMM	X-ray Multi-Mirrors
YSO	Young Stellar Objects
ZAMS	Zero-Age Main Sequence

Appendix D: Derivation of the Point Spread Function

Let each of the set of N sub-apertures be represented by a circular aperture of diameter D with the set of sub-aperture centers located at $\{u_j, v_j\}_{j=1\ldots N}$ where (u,v) are orthogonal coordinates in a plane orthogonal to the system line of sight, and tangent to the virtual vertex of the primary mirror array. This plane is the entrance pupil. Mathematically we can represent this aperture pattern, and hence the *entrance pupil*, by:

$$P(u,v) = \sum_{j=1}^{N} E(u-u_j, v-v_j) \tag{D.1}$$

where $E(u-u_j, v-v_j)$ represents the scalar electric field amplitude at the j-th aperture in the entrance pupil plane. For circular sub-apertures:

$$E(u,v) = E_0 circ\left(\frac{\sqrt{u^2+v^2}}{D}\right) = \begin{cases} E_0 & \text{for } \sqrt{u^2+v^2} \le D/2 \\ 0 & \text{otherwise} \end{cases} \tag{D.2}$$

where E_0 is the constant electric field across each of the sub-apertures for a point source at infinity, i.e. a plane wave in the entrance pupil. At this point we assume no amplitude, phase or polarization errors are introduced.

The optical point spread function (PSF) is proportional to the modulus squared of the spatial Fourier transform of the *exit pupil* function. If the optical system is perfect, i.e. no wavefront errors, misalignments, deformations, coating errors etc. then the exit pupil is ideally an image of the entrance pupil and the monochromatic PSF, normalized to unity at the origin, is given by:

$$PSF(\theta_x, \theta_y; \lambda) = \frac{1}{\left|\iint P(u,v)dudv\right|^2} \left| \iint P(u,v) e^{-i\frac{2\pi}{\lambda}(\theta_x u + \theta_y v)} dudv \right|^2 \tag{D.3}$$

where (θ_x, θ_y) are angular sky coordinates and λ is the radiation wavelength. If optical errors are present these are represented as amplitude and phase errors in propagation between the entrance and exit pupils of the system.

The integral over the aperture in the denominator of (3) is given by the aperture intensity times the aperture area, thus $\iint P(u,v)dudv = E_0 N \frac{\pi}{4} D^2$ yielding:

$$PSF(\theta_x,\theta_y;\lambda)=\frac{16}{I_0N^2\pi^2D^4}\left|\iint P(u,v)e^{-i\frac{2\pi}{\lambda}(\theta_x u+\theta_y v)}dudv\right|^2 \tag{D.4}$$

where $I_0=|E_0|^2$ is the field intensity at the entrance pupil. The integral term of equation (A4) can also be rearranged to yield:

$$\iint P(u,v)e^{-i\frac{2\pi}{\lambda}(\theta_x u+\theta_y v)}dudv=\iint E(x,y)e^{-i\frac{2\pi}{\lambda}(\theta_x x+\theta_y y)}dxdy\sum_{j=1}^{N}e^{-i\frac{2\pi}{\lambda}(\theta_x u_j+\theta_y v_j)} \tag{D.5}$$

and for a single circular sub-aperture:

$$\iint E(x,y)e^{-i\frac{2\pi}{\lambda}(\theta_x x+\theta_y y)}dxdy\propto\frac{2J_1\left(\frac{\pi D}{\lambda}\sqrt{\theta_x^2+\theta_y^2}\right)}{\frac{\pi D}{\lambda}\sqrt{\theta_x^2+\theta_y^2}} \tag{D.6}$$

where J_1 is a Bessel function or order 1. This yields the idealized system PSF, normalized to unity at the origin $PSF(0,0;\lambda)=1$, of:

$$PSF(\theta_x,\theta_y)=\frac{1}{N^2}\left|\frac{2J_1\left(\frac{\pi D}{\lambda}\sqrt{\theta_x^2+\theta_y^2}\right)}{\frac{\pi D}{\lambda}\sqrt{\theta_x^2+\theta_y^2}}\right|^2\left|\sum_{j=1}^{N}e^{-i\frac{2\pi}{\lambda}(\theta_x u_j+\theta_y v_j)}\right|^2 \tag{D.7}$$

Denoting the normalized primary beam PSF, i.e. the PSF due to a single subaperture as:

$$PSF_{PB}(\theta_x,\theta_y;\lambda)=\left|2J_1\left(\frac{\pi D}{\lambda}\sqrt{\theta_x^2+\theta_y^2}\right)\Big/\frac{\pi D}{\lambda}\sqrt{\theta_x^2+\theta_y^2}\right|^2 \tag{D.8}$$

allows the system PSF to be succinctly expressed as:

$$PSF(\theta_x,\theta_y;\lambda)=PSF_{PB}(\theta_x,\theta_y;\lambda)\frac{1}{N^2}\left|\sum_{j=1}^{N}e^{-i\frac{2\pi}{\lambda}(\theta_x u_j+\theta_y v_j)}\right|^2 \tag{D.9}$$

Equation (D.9) expresses the system PSF as the product of the primary beam PSF and the modulus squared of a finite phasor sum. The phasor sum consists of a finite set of Fourier terms which is unto itself periodic. This terms accounts for the fringing in the PSF. The phasor sum term can be re-expressed as:

$$\left|\sum_{j=1}^{N}e^{-i\frac{2\pi}{\lambda}(\theta_x u_j+\theta_y v_j)}\right|^2=\sum_{j=1}^{N}\sum_{k=1}^{N}e^{-i\frac{2\pi}{\lambda}(\theta_x u_j+\theta_y v_j)}e^{-i\frac{2\pi}{\lambda}(\theta_x u_k+\theta_y v_k)}=\sum_{j=1}^{N}\sum_{k=1}^{N}e^{-i\frac{2\pi}{\lambda}(\theta_x(u_j-u_k)+\theta_y(v_j-v_k))} \tag{D.10}$$

and defining the baselines along the U and V directions as $B_{u,jk}\equiv u_j-u_k$ and $B_{v,jk}\equiv v_j-v_k$ allows equation (D.10) to be re-expressed as:

$$\sum_{j=1}^{N}\sum_{k=1}^{N} e^{-i\frac{2\pi}{\lambda}\left(\theta_x B_{u,jk}+\theta_y B_{v,jk}\right)} = \sum_{j=1}^{N}\sum_{k=1}^{N} e^{-i\frac{2\pi}{\lambda}\vec{\theta}\cdot\vec{B}_{jk}} = N + \sum_{k=1}^{N-1}\sum_{j>k}^{N}\left[e^{-i\frac{2\pi}{\lambda}\vec{\theta}\cdot\vec{B}_{jk}} + e^{i\frac{2\pi}{\lambda}\vec{\theta}\cdot\vec{B}_{jk}} \right] = N + 2\sum_{k=1}^{N-1}\sum_{j>k}^{N}\cos\left[\frac{2\pi}{\lambda}\vec{\theta}\cdot\vec{B}_{jk}\right] \quad \text{(D.11)}$$

Thus the PSF, normalized to unity at the origin, becomes:

$$\boxed{PSF\left(\theta_x,\theta_y;\lambda\right) = PSF_{PB}\left(\theta_x,\theta_y;\lambda\right)\frac{1}{N^2}\left\{N + 2\sum_{k=1}^{N-1}\sum_{j>k}^{N}\cos\left[\frac{2\pi}{\lambda}\vec{\theta}\cdot\vec{B}_{jk}\right]\right\}} \quad \text{(D.12)}$$

The term in brackets shows that there are only $N(N-1)/2$ independent set of fringes in a PSF and since the fringe patterns are real, their Fourier transforms are symmetric with respect to the origin in the Fourier plane. The direction of a given fringe, i.e. along the direction of minimal spacing between successive maxima, is in the direction of the vector adjoining that baseline pair, i.e. along $\vec{B}_{jk}$ and the period is given by $\lambda/\left|B_{jk}\right|$ in angular units on the sky.

Unit Integral Normalization

The PSF can be renormalized to unit integral – this form is more suitable for the photometric calculations. The unit integral PSF is denoted as:

$$PSF'\left(\theta_x,\theta_y;\lambda\right) = \frac{PSF\left(\theta_x,\theta_y;\lambda\right)}{\iint PSF\left(\theta_x,\theta_y;\lambda\right)d\theta_x d\theta_y} \quad \text{(D.13)}$$

Labeling the denominator as:

$$P = \iint\left\{\frac{1}{N}PSF_{PB}\left(\theta_x,\theta_y;\lambda\right) + \frac{2}{N^2}\sum_{k=1}^{N-1}\sum_{j>k}^{N} PSF_{PB}\left(\theta_x,\theta_y;\lambda\right)\cos\left[\frac{2\pi}{\lambda}\vec{\theta}\cdot\vec{B}_{jk}\right]\right\}d\theta_x d\theta_y \quad \text{(D.14)}$$

dividing into two terms:

$$\begin{aligned} P_1 &= \frac{1}{N}\iint PSF_{PB}\left(\theta_x,\theta_y;\lambda\right)d\theta_x d\theta_y \\ P_2 &= \frac{2}{N^2}\sum_{k=1}^{N-1}\sum_{j>k}^{N}\iint PSF_{PB}\left(\theta_x,\theta_y;\lambda\right)\cos\left[\frac{2\pi}{\lambda}\vec{\theta}\cdot\vec{B}_{jk}\right]d\theta_x d\theta_y \end{aligned} \quad \text{(D.15)}$$

and expressing the integral term within the 2nd term in D.15 as:

$$P_{2,jk} = \iint PSF_{PB}\left(\theta_x,\theta_y;\lambda\right)\cos\left[\frac{2\pi}{\lambda}\vec{\theta}\cdot\vec{B}_{jk}\right]d\theta_x d\theta_y = \mathrm{Re}\left\{\iint PSF_{PB}\left(\theta_x,\theta_y;\lambda\right)e^{-i\frac{2\pi}{\lambda}\vec{\theta}\cdot\vec{B}_{jk}}d\theta_x d\theta_y\right\} \quad \text{(D.16)}$$

The primary beam PSF can be expressed as:

$$PSF_{PB}\left(\theta_x,\theta_y;\lambda\right) = \left|\iint A(u,v)e^{-i\frac{2\pi}{\lambda}\left(u\theta_x+v\theta_y\right)}dudv\right|^2 = \iint\iint A(u,v)A(u',v')e^{-i\frac{2\pi}{\lambda}\left[(u-u')\theta_x+(v-v')\theta_y\right]}dudvdu'dv' \quad \text{(D.17)}$$

yielding:

$$P'_{2,jk} = \iint \iint \iint A(u,v)A(u',v')e^{-i\frac{2\pi}{\lambda}\left[(u-u'+B_{x,jk})\theta_x+(v-v'+B_{y,jk})\theta_y\right]}dudvdu'dv'd\theta_x d\theta_y \quad \text{(D.18)}$$

interchanging the order of integration and evaluating the integrals over $d\theta_x d\theta_y$ yields:

$$P'_{2,jk} = \iint A(u,v)A(u+B_{x,jk},v+B_{y,jk})dudv = \begin{cases} \frac{2}{\pi}\left\{\cos^{-1}\left(\frac{\left|\vec{B}_{jk}\right|}{D}\right)-\frac{\left|\vec{B}_{jk}\right|}{D}\sqrt{1-\frac{\left|\vec{B}_{jk}\right|^2}{D^2}}\right\}; & \left|\vec{B}_{jk}\right| < D \\ 0; & \left|\vec{B}_{jk}\right| \geq D \end{cases} \quad \text{(D.19)}$$

For any sparse system of interest the baselines are always greater than the sub-aperture diameter, hence $P'_{2,jk} = 0$ and therefore $P_2 = 0$, thus the denominator of D.13 only depends on the number of sub-apertures and the integral of the primary beam, i.e.

$$PSF'(\theta_x,\theta_y;\lambda) = \frac{PSF(\theta_x,\theta_y;\lambda)}{\frac{1}{N}\iint PSF_{PB}(\theta_x,\theta_y;\lambda)d\theta_x d\theta_y} \quad \text{(D.20)}$$

For a circular sub-aperture the primary beam integral in D.20 can be evaluated to give:

$$\iint PSF_{PB}(\theta_x,\theta_y;\lambda)d\theta_x d\theta_y = 2\pi\int_0^\infty \left|\frac{2J_1\left(\pi\frac{D}{\lambda}\theta\right)}{\pi\frac{D}{\lambda}\theta}\right|^2 \theta d\theta = 2\pi\frac{2}{\pi^2}\left(\frac{\lambda}{D}\right)^2 = \frac{\lambda^2}{\Delta A} \quad \text{(D.21)}$$

where $\Delta A = \pi(D/2)^2$ is the area of a single sub-aperture. Thus the PSF normalized to unit integral can be expressed as:

$$PSF'(\theta_x,\theta_y;\lambda) = \frac{\Delta A}{\lambda^2}PSF_{PB}(\theta_x,\theta_y;\lambda)\left\{1+\frac{2}{N}\sum_{k=1}^{N-1}\sum_{j>k}^{N}\cos\left[\frac{2\pi}{\lambda}\vec{\theta}\cdot\vec{B}_{jk}\right]\right\} \quad \text{(D.22)}$$

or folding in the Airy disk as the primary beam PSF to yield:

$$\boxed{PSF'(\theta_x,\theta_y;\lambda) = \frac{\Delta A}{\lambda^2}\left\{\frac{2J_1\left(\pi\frac{D}{\lambda}\theta\right)}{\pi\frac{D}{\lambda}\theta}\right\}^2\left\{1+\frac{2}{N}\sum_{k=1}^{N-1}\sum_{j>k}^{N}\cos\left[\frac{2\pi}{\lambda}\vec{\theta}\cdot\vec{B}_{jk}\right]\right\}} \quad \text{(D.23)}$$

PSF

Photometry

Multiplying D.23 by the total number of photo-electrons collected by the aperture

$$n_e = M_0 \cdot 10^{-M_v/2.5} \cdot N \cdot \Delta A \cdot \Delta\lambda \cdot T(\lambda_0) \cdot q.e.(\lambda_0) \cdot \Delta t \tag{D.24}$$

where $M_0 = 1 \times 10^8$ photons/meter^2/second/nanometer is the source photon rate for a zero visual magnitude source and M_v is the visual magnitude, ΔA is the area of a single aperture, $T(\lambda_0)$ is the transmission of the optics, $q.e.(\lambda_0)$ is the quantum efficiency of the detector, Δt is the integration time in seconds and N is the total number of sub-apertures. For small pixels ($< \lambda/B$) the number of photoelectrons collected per pixel, versus angle on the sky, is given by:

$$n(\vec{\theta};\lambda) \approx n_e \Delta\Omega_{pixel} PSF'(\vec{\theta};\lambda) \tag{D.25}$$

where $\Delta\Omega_{pixel}$ is the solid angle subtended by a pixel projected on the sky.

Combining D.23 and D.25 yields the photo-electrons per pixel for an unresolved point source as:

$$n(\vec{\theta};\lambda) \approx n_e \Delta\Omega_{pixel} \frac{\Delta A}{\lambda^2} \left\{ \frac{2J_1\left(\pi \frac{D}{\lambda}\theta\right)}{\pi \frac{D}{\lambda}\theta} \right\}^2 \left\{ 1 + \frac{2}{N} \sum_{k=1}^{N-1} \sum_{j>k}^{N} \cos\left[\frac{2\pi}{\lambda} \vec{\theta} \cdot \vec{B}_{jk} \right] \right\} \tag{D.26}$$

Equation D.26 expresses the number of photo-electrons per pixel in the focal plane for an unresolved monochromatic point source.

Appendix E: Characterization and Modeling of System Performance

An array of formation flying spacecraft, each comprising a single piece of a primary mirror, can be combined interferometrically to form ultra-high resolution imagery thereby allowing direct resolution of the structure of nearby stellar disks (~30 micro-arcseconds) and enabling the Stellar Imager (SI) mission. Furthermore, formation flying interferometers significantly reduce mass, launch volume and cost over a filled aperture system of comparable size, thereby allowing systems of this type to become achievable for flight within 20 years. This, however, comes at the expense of a complex hierarchical control system. Each of the mirrorlet spacecraft must have its own attitude control system working in harmony with a system wide attitude and wavefront control system with post-processing to achieve and maintain the required level of image quality during extended science observations. In order to approach development of a flight control system of this magnitude we first develop models of imaging through such interferometers. These models must include beam combination, wavefront sensing and optical control, as well as pointing and attitude and their respective control systems. These combined models are coupled to perform sensitivity analysis. The science requirements are flowed to system level and sub-system requirements and combined in the form of an error budget for SI.

E.1 Point Spread Function

A Fizeau interferometer's spatial response to an unresolved point source is known as the point spread function (PSF) and an extended scene image is given as the 2 dimensional spatial convolution of the PSF with the object radiance distribution with detector effects folded in. The monochromatic optical point spread function (derived in Appendix D), without errors, is succinctly stated as:

$$PSF\left(\theta_x,\theta_y;\lambda\right)= PSF_{PB}\left(\theta_x,\theta_y;\lambda\right)\frac{1}{N^2}\left\{N+2\sum_{k=1}^{N-1}\sum_{j>k}^{N}\cos\left[\frac{2\pi}{\lambda}\vec{\theta}\cdot\vec{B}_{jk}\right]\right\} \tag{E.1}$$

where $\vec{\theta}=\left(\theta_x,\theta_y\right)$ are angular sky coordinates, λ is the wavelength, $N=30$ is the number of apertures and $\left\{\vec{B}_{jk}\right\}=\left\{B_{u,jk},B_{v,jk}\right\}=\left\{\left(u_j-u_k\right)\left(v_j-v_k\right)\right\}_{j,k=1\ldots N}$ represents the set of baseline vectors in the UV plane. The double summation extends from $k=1\ldots N-1$ and $j>k\ldots N$ and expresses that the Golomb aperture pattern is non-redundant; each of the Fourier terms in the double summation is unique of which there are $N\left(N-1\right)/2=435$ spatial frequencies and their complex conjugates, not including the DC (0 spatial frequency) term which occurs $N=30$ times. It is this double summation term which is responsible for the fringing in the point spread function. The direction of a given fringe, i.e. along the direction of minimal spacing between successive maxima, is in the direction of the vector adjoining that baseline pair, i.e. along $\vec{B}_{jk}$ and the period is given by $\lambda/\left|B_{jk}\right|$ in angular units on the sky. The term PSF_{PB} in equation (E.1) represents the primary beam PSF, i.e. the PSF of a single circular sub-aperture of diameter $D=1$ meter, and is given by:

$$PSF_{PB}\left(\theta_x,\theta_y;\lambda\right)=\left|2J_1\left(\frac{\pi D}{\lambda}\sqrt{\theta_x^2+\theta_y^2}\right)\Big/\frac{\pi D}{\lambda}\sqrt{\theta_x^2+\theta_y^2}\right|^2 \tag{E.2}$$

where J_1 is a Bessel function of order 1. The primary beam PSF and the system PSF are normalized to

unity at the origin, i.e. $PSF(0,0;\lambda) = PSF_{PB}(0,0;\lambda) = 1.0$

Figure E.1 shows idealized SI PSFs at wavelengths of $\lambda = 0.155$ microns and $\lambda = 0.280$ microns for SI. The primary beam PSF (1st term in equation 1) is seen as the concentric Airy rings in the two leftmost images of Figure E.1. The pattern due to the interference (2nd term in equation 1) of the different baselines is seen in the two rightmost images of Figure E.1.

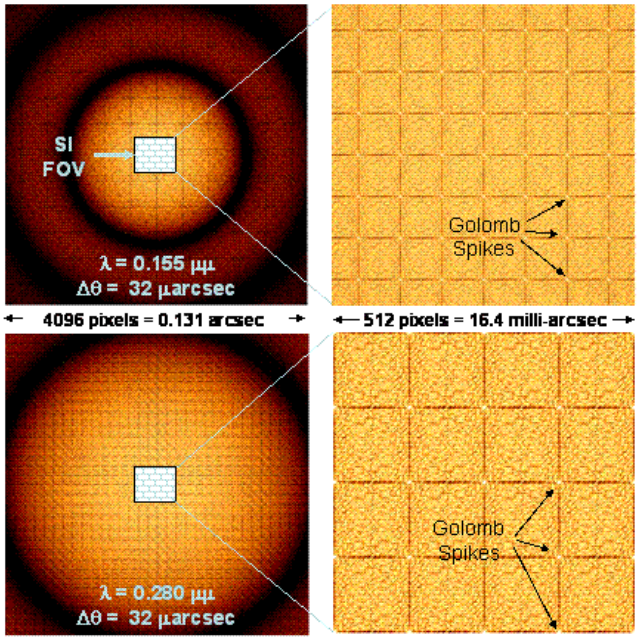


Figure E.1: PSFs and FOV
Upper Left: PSF at λ = 0.155 microns

Upper Right: PSF Zoomed to SI Field of View

Lower Left: PSF at λ = 0.280 microns

Lower Right: PSF Zoomed to SI Field of View

All Images on Log Scale (R. Lyon)

E.2 Resolution, Contrast and Sampling

The *resolution* of SI is defined to be the full width half max (FWHM) of the central PSF Golomb sidelobe. Figure E.2 shows plots along the θ_x direction for the two primary science wavelengths. The FWHM is ~ 50 micro-arcseconds for a wavelength of 0.155 microns and ~100 micro-arcseconds for a wavelength of 0.280 microns. The periodic pattern of grating like sidelobes are spaced by $\Delta\alpha = (N_{ugrid} - 1)\lambda / B_{u\max}$ and $\Delta\beta = (N_{vgrid} - 1)\lambda / B_{v\max}$ on the sky. λ is the wavelength, and $(N_{ugrid}, N_{vgrid}) = (30,30)$ and $(B_{u\max}, B_{v\max}) = (411.6064, 411.6064)$ meters are the maximum baseline components in the u and v directions respectively. Note that $(\Delta\alpha, \Delta\beta) = (2.255094, 2.255094)$ milli-arcseconds for λ = 0.155 microns and $(\Delta\alpha, \Delta\beta) = (4.073718, 4.073718)$ milli-arcseconds for λ = 0.280 microns. These Golomb sidelobes generally decrease as one moves radially outward from the core of the PSF. The central Golomb spike has a peak contrast of ~ $1/N^2$ = 1/900 where N=30 is the number of sub-apertures. This property can be seen in the lower plot in Figure E.2. Note that this contrast is the theoretical upper bound for a perfectly phased system without any other sources of errors and for a point source. An extended source will lower the contrast and will be discussed in more detail in later sections.

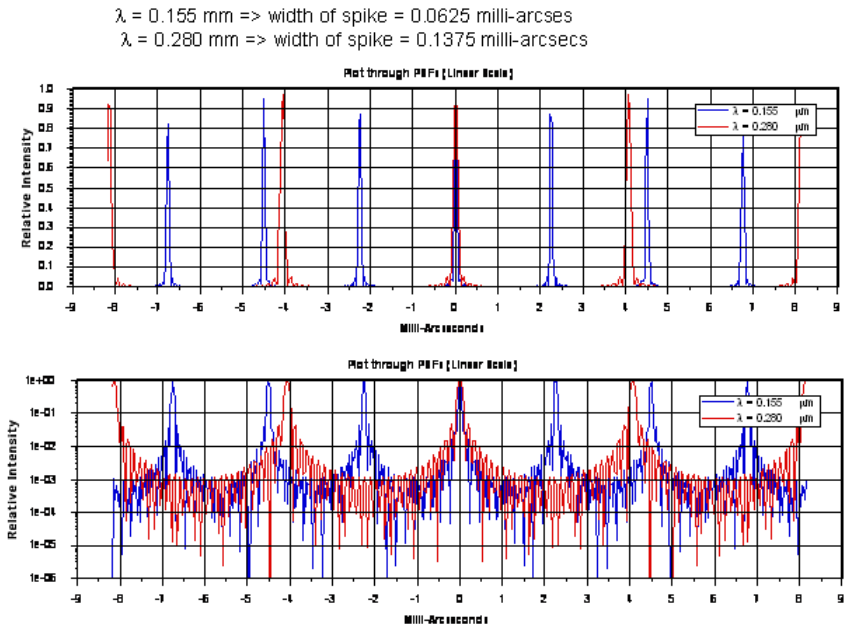


Figure E.2: Plots through center of PSFs along θ_x (R. Lyon).

Nyquist sampling of the focal plane is driven by the shortest wavelength and the longest baseline and is required to be $\Delta\theta_s \le \frac{\lambda}{2B_{\max}} = 32$ micro-arcseconds or sampling in the focal plane of $\Delta x \le \frac{\lambda}{2B_{\max}} F = \frac{\lambda(f/\#)}{2}$. If we assume the detector pixel sizes are 9 microns to yield an effective focal length of $F \ge \frac{2B_{\max}}{\lambda}\Delta x = 58.065$ kilometers, or a final effective beam $f/\# \ge \frac{2\Delta x}{\lambda} = 116.13$, i.e. an optically slow system. Sampling at, or better than, the Nyquist limit will insure optimal use of post-processing to construct the image.

E.3 Encircled Energy

The simultaneous mixing of all 30 beamlets results in a point spread function which is widely distributed in the focal plane. The encircled energy is a measure of this energy distribution. Figure E.3 plots the encircled energy, i.e. the energy within a circle of radius "R" on the sky for a point source. For the baseline design of SI the encircled energies are plotted in Figure E.3 for the two primary science wavelengths of λ = 0.155 microns and λ = 0.280 microns. The top plot shows the encircled energy on a linear scale and the lower plot of Figure F.3 shows the same on a log-log scale. Within twice the width of the central Golomb spike, 100 micro-arcsec for 0.155 um and 200 micro-arcsecs for 0.280 microns, the encircled energy is ~ $1x10^{-4}$, a very small encircled energy. Thus only ~1 in 10,000 photons is contained within the central Golomb spike, the rest are diffracted away from the core and into the surrounding structure. This might appear to be a problem for SI, however, since SI is a Fizeau interferometer the post-processing to recover the spatial frequencies will significantly alter this.

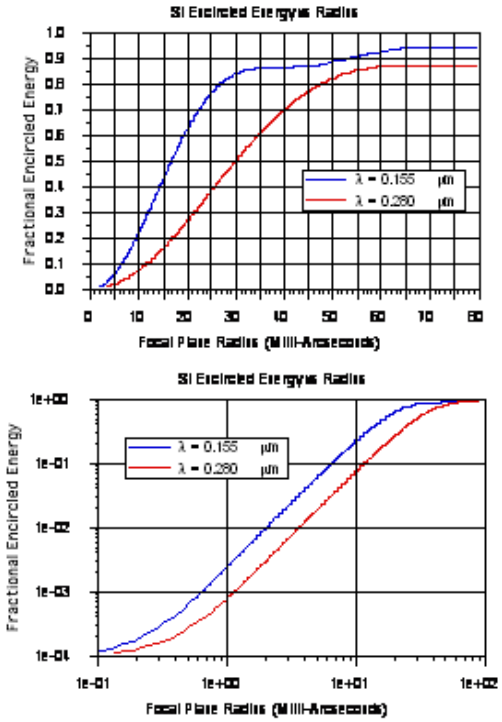


Figure E.3: PSF Encircled Energy

(R. Lyon).

E.4 Point Response Function and Detector Effects

The point spread function represents the focal plane relative radiance distribution due to an unresolved point source. However with real detectors the point spread function integrated over the solid angle subtended by a pixel, sampled onto a finite detector size and integrated over the spectral radiance of the source and transmissivity of the optics is what is actually measured. This function is known as the point response function (PRF). The PRF incorporates the effects of source spectral flux $S(\lambda)$, the spectral transmission of the optics and spectral filters $T(\lambda)$, the finite area of a single pixel of the detector $rect\left[\frac{\theta_x}{\Delta\theta_x},\frac{\theta_y}{\Delta\theta_y}\right]$, and sampling and the finite extent of the detector array. The PRF is given by:

$$PRF(n\Delta\theta_x,m\Delta\theta_y)=\left\{\int_{\Delta\lambda} S(\lambda)T(\lambda)PSF(\theta_x,\theta_y;\lambda)**rect\left[\frac{\theta_x}{\Delta\theta_x},\frac{\theta_y}{\Delta\theta_y}\right]d\lambda\right\}\sum_{m=-\frac{N_{px}}{2}}^{N_{px}/2}\sum_{n=-\frac{N_{py}}{2}}^{N_{py}/2}\delta(\theta_x-n\Delta\theta_x,\theta_y-m\Delta\theta_y)+\eta_{nm} \quad \text{(E.3)}$$

where (n,m) are the pixel indices in the focal plane, $(\Delta\theta_x,\Delta\theta_y)$ are the finite size of the detector pixels projected on the sky (angular units), and (N_{px},N_{py}) are the number of focal plane pixels in the θ_x and θ_y directions respectively, and "$**$" represents 2D spatial convolution, η_{nm} represents additive noise. The PRF is actually the PSF convolved with the detector footprint and integrated over the spectral passband of the source and optics and multiplied by the sampling function. The PRF contains the

combined effects of optical diffraction, aberrations, spectral response and sampling and is the best representation to assess the performance of the Stellar Imager performance.

E.5 Piston, Tip and Tilt

In order to assess the sensitivities with respect to rigid body misalignments over the set of 30 mirrorlet spacecraft a Monte-Carlo type simulation was conducted using the Optical Systems Characterization and Analysis Research (OSCAR) package. The Strehl ratio was calculated for each of a 100 realizations of random piston, tip and tilt errors on each of the mirrorlets. The mean piston, tip and tilt over the ensemble of 30 mirrors was set to zero and the standard deviation of the set was set to σ. Figure E.4 shows the results of this simulation for piston, tip and tilt. The top two plots in Figure E.4 show the effect on the Strehl ratio for random piston (left) and random tip/tilts (right) at $\lambda = 0.155$ microns and the bottom two plots for piston (left) and random tip/tilt (right) at $\lambda = 0.280$ microns. Plotted is the mean Strehl ratio versus rms piston or tip/tilt over an ensemble of 100 realizations as the solid black lines and the +/- 1 standard deviation over the ensemble as dashed lines. Thus for example in the upper left plot for a random piston error of 10 nm rms over all 30 mirrorlets we would expect the Strehl to fall to approximately 50% of its nominal value. In this simulation a 10 nm rms piston value refers to the 1 sigma standard deviation of a Gaussian distributed set of pistons over all the 30 mirrorlets, thus some mirrorlets have more and some have less but such that the mean is zero and similarly for tip and tilt. These plots show that a direct image from the Golomb array is very sensitive to random uncorrected misalignments, however during wavefront sensing piston, tip and tilt will be accurately measured and will be compensated for via phase correction during post-processing.

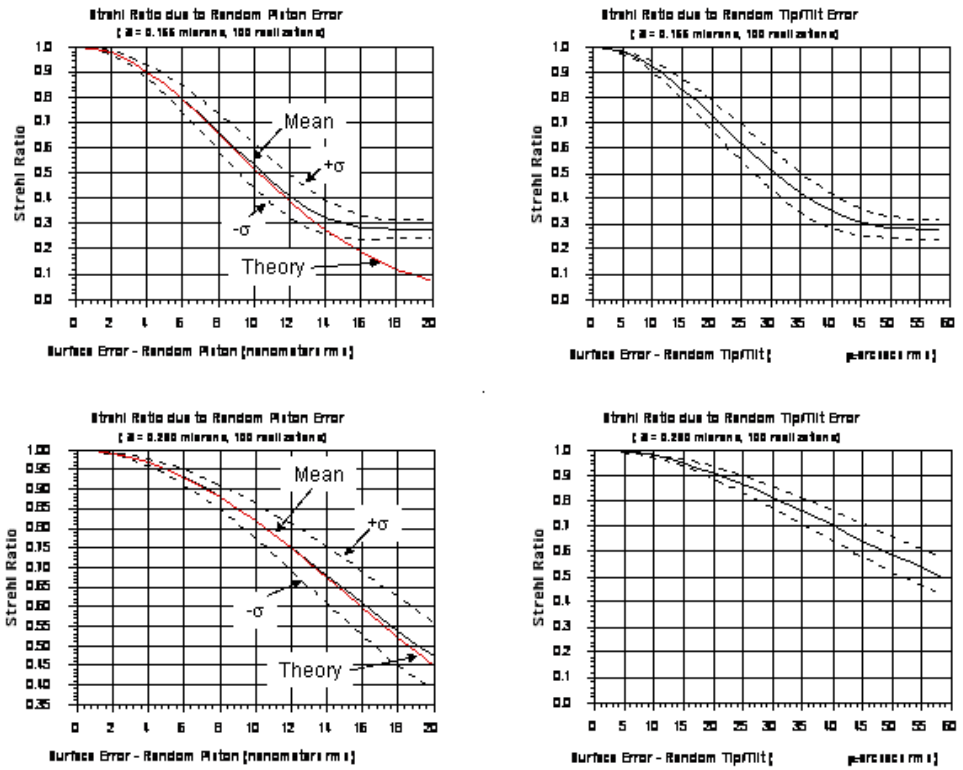


Figure E.4: Effect of random piston, tip, and tilt errors (R. Lyon).

E.6 Optical Transfer Function and UV Coverage

The optical transfer function (OTF) is given as the 2D spatial Fourier transform of the point spread function. Its significance is that it conveys how well the spatial frequencies, for incoherent imaging, of an object are transferred to the image plane. For an imaging interferometer it also gives the UV coverage.

The OTF is given as:

$$OTF\left(f_x,f_y;\lambda\right)=\frac{FT_{f_x,f_y}\left\{PSF\left(\theta_x,\theta_y;\lambda\right)\right\}}{FT_{f_x,f_y}\left\{PSF\left(\theta_x,\theta_y;\lambda\right)\right\}\Big|_{f_x=f_y=0}} \tag{E.4}$$

where FT denotes the 2D spatial Fourier transform. Inserting equation E.1 into E.4 yields the OTF as:

$$OTF\left(\vec{f};\lambda\right)=\frac{2}{\pi}\left\{\cos^{-1}\left(\frac{\lambda}{D}f\right)-\frac{\lambda}{D}f\sqrt{1-\left(\frac{\lambda}{D}f\right)^2}\right\}**\left\{\delta\left(\vec{f}\right)+\frac{1}{N}\sum_{k=1}^{N-1}\sum_{j>k}^{N}\delta\left(\vec{f}+\frac{1}{\lambda}\vec{B}_{jk}\right)+\delta\left(\vec{f}-\frac{1}{\lambda}\vec{B}_{jk}\right)\right\} \tag{E.5}$$

where $\vec{f}\equiv\left(f_x,f_y\right)$ are the spatial frequencies in units of cycles per radians on the sky and $f=\left|\vec{f}\right|=\sqrt{f_x^2+f_y^2}$. 2D spatial convolution is denoted by "$**$". The OTF is given by the primary beam OTF convolved with an array of delta functions, each of which is located at the baseline spacing divided by the wavelength.

$$OTF\left(\vec{f};\lambda\right)=OTF_{PB}\left(\vec{f};\lambda\right)**\left\{\delta\left(\vec{f}\right)+\frac{1}{N}\sum_{k=1}^{N-1}\sum_{j>k}^{N}\delta\left(\vec{f}+\frac{1}{\lambda}\vec{B}_{jk}\right)+\delta\left(\vec{f}-\frac{1}{\lambda}\vec{B}_{jk}\right)\right\} \tag{E.6}$$

The OTF is normalized such $OTF\left(\vec{f}=0;\lambda\right)=1$ and the peak of each of the cones is given by the inverse of the number of mirrorlets $1/N$. The primary beam OTF is a nearly conical shaped function and is independent of the baseline spacings, conversely the location of each of the conical regions is independent of the diameter of each sub-aperture and they fall at $\vec{f}_{0,jk}=\pm\frac{1}{\lambda}\vec{B}_{jk}$. Figure E.5 shows the optical transfer function at λ = 0.155 microns. Each sparse white point represents a spatial frequency cone due to the primary beam, while the minimum spacing is quantized in units of $B_{\min}/\lambda$ where $B_{\min}$ represents the minimum baseline spacing of the Golomb array. The zero baseline frequency is at the center of the figure. Note that with the 30 element Golomb configuration the UV coverage is incomplete, i.e. there are holes in the UV coverage and the density of the holes grows as one moves radially outward from the center.

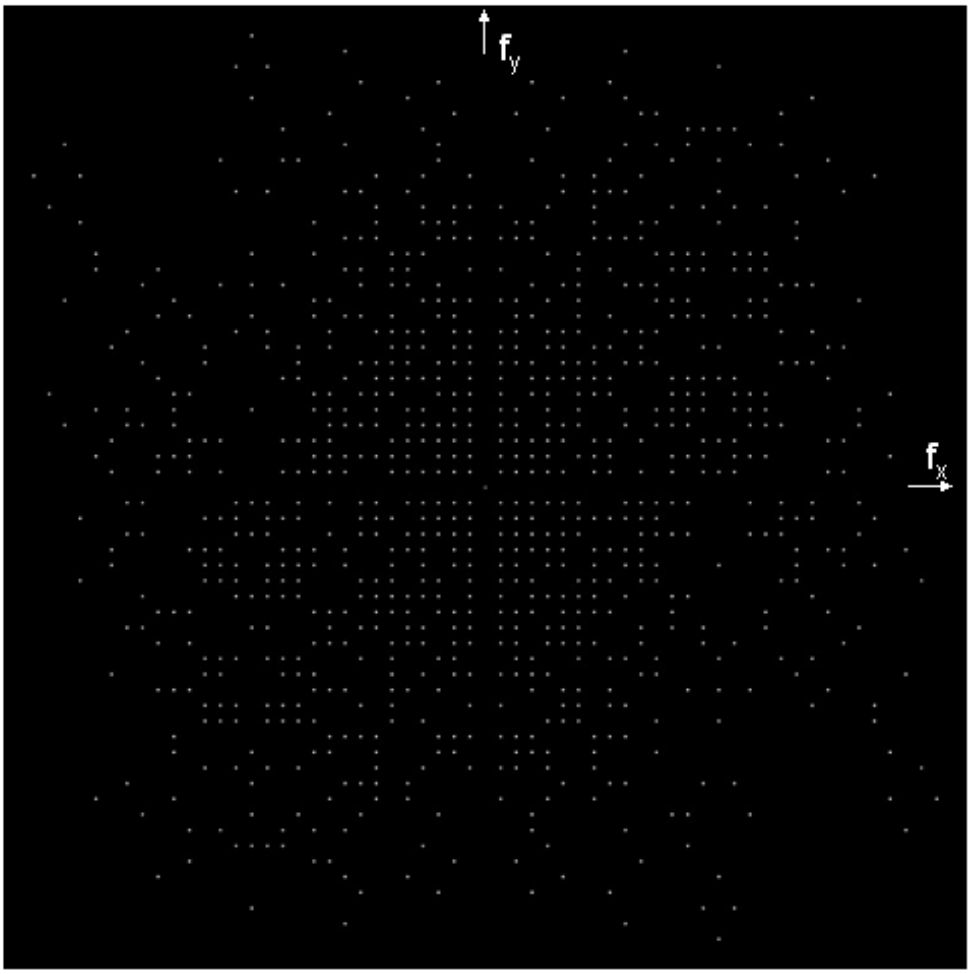


Figure E.5: Optical Transfer Function and UV Coverage. OTF@ λ= 0.155 um
(R. Lyon).

E.7 Quasi-Monochromatic Compact Extended Sources

Generally an optical systems spatial response is described in terms of the optical point spread function (PSF). An image is given as the 2D spatial convolution of the point spread function with the object if the PSF is space invariant, i.e. it does not change over the field of view. Imaging interferometry has historically described the systems response in terms of spatial fringes and spatial and temporal visibility functions (fringe contrast). Both approaches are insightful into the image forming properties of SI and both approaches are equivalent for a compactly supported source imaged through spectral filter such that its passband is much less than its center wavelength. A compactly supported source is unresolved with respect to the PSF of a single sub-aperture, i.e. the primary beam, yet resolved with respect to the PSF of the entire system . Developed first will be a point source in the quasi-monochromatic approximation followed by the compact source and the results will be combined for an analytic expression of the spatial response of a Fizeau imaging interferometer. The results are specialized to the case of Stellar Imager in sections 2.2.3 and 3.1.4.

Quasi-Monochromatic Point Spread Function

For an unresolved source at each baseline the optical point spread function is given by (Appendix D, equation D12):

$$PSF\left(\vec{\theta};\lambda\right)= PSF_{PB}\left(\vec{\theta};\lambda\right)\frac{1}{N^2}\left\{N+2\sum_{k=1}^{N-1}\sum_{j>k}^{N}\cos\left[\frac{2\pi}{\lambda}\vec{B}_{jk}\cdot\vec{\theta}\right]\right\} \quad \text{(E.7)}$$

where N is the number of sub-apertures comprising SI, λ is the wavelength, $\vec{B}_{jk}$ is the baseline spacing between the j,k-th apertures and $\vec{\theta}$ is the angle on the sky.

The PSF integrated across a spectral filter function, $g(\lambda-\lambda_0,\Delta\lambda)$, centered on wavelength λ_0 and of width $\Delta\lambda$ is given by:

$$PSF\left(\vec{\theta}\right)= \int g(\lambda-\lambda_0,\Delta\lambda)PSF_{PB}\left(\vec{\theta};\lambda\right)\frac{1}{N^2}\left\{N+2\sum_{k=1}^{N-1}\sum_{j>k}^{N}\cos\left[\frac{2\pi}{\lambda}\vec{B}_{jk}\cdot\vec{\theta}\right]\right\}d\lambda \quad \text{(E.8)}$$

In the quasi-monochromatic approximation the spectral passband is narrow with respect to the center wavelength, i.e. $\Delta\lambda << \lambda_0$ then the primary beam PSF can be approximately evaluated at only the center wavelength and removed from the integrand to yield:

$$PSF\left(\vec{\theta}\right)\approx \frac{1}{N^2}PSF_{PB}\left(\vec{\theta};\lambda_0\right)\int g(\lambda-\lambda_0,\Delta\lambda)\left\{N+2\sum_{k=1}^{N-1}\sum_{j>k}^{N}\cos\left[\frac{2\pi}{\lambda}\vec{B}_{jk}\cdot\vec{\theta}\right]\right\}d\lambda \quad \text{(E.9)}$$

Interchanging the order of integration with the summation and rearranging terms yields:

$$PSF\left(\vec{\theta}\right)\approx \frac{1}{N^2}PSF_{PB}\left(\vec{\theta};\lambda_0\right)\left\{N\int g(\lambda-\lambda_0,\Delta\lambda)d\lambda+2\sum_{k=1}^{N-1}\sum_{j>k}^{N}\int g(\lambda-\lambda_0,\Delta\lambda)\cos\left[\frac{2\pi}{\lambda}\vec{B}_{jk}\cdot\vec{\theta}\right]d\lambda\right\} \quad \text{(E.10)}$$

Changing variables such that $\lambda'=\lambda-\lambda_0$ yields:

$$PSF\left(\vec{\theta}\right)\approx \frac{1}{N^2}PSF_{PB}\left(\vec{\theta};\lambda_0\right)\left\{N\int g(\lambda',\Delta\lambda)d\lambda+2\sum_{k=1}^{N-1}\sum_{j>k}^{N}\int g(\lambda',\Delta\lambda)\cos\left[\frac{2\pi}{\lambda'+\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}\right]d\lambda\right\} \quad \text{(E.11)}$$

and expanding $\frac{1}{\lambda'+\lambda_0}$ in powers of λ'/λ_0 and keeping only terms to first order yields $\frac{1}{\lambda'+\lambda_0}\approx\frac{1}{\lambda_0}(1-\lambda'/\lambda_0)$. This allows the integral within the summation in E.11 to be expressed to 1st order as:

$$\int g(\lambda',\Delta\lambda)\cos\left[\frac{2\pi}{\lambda_0}\left(1-\frac{\lambda'}{\lambda_0}\right)\vec{B}_{jk}\cdot\vec{\theta}\right]d\lambda' =$$
$$\int g(\lambda',\Delta\lambda)\left\{\cos\left(\frac{2\pi}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}\right)\cos\left(\frac{2\pi}{\lambda_0}\frac{\lambda'}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}\right)+\sin\left(\frac{2\pi}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}\right)\sin\left(\frac{2\pi}{\lambda_0}\frac{\lambda'}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}\right)\right\}d\lambda' \quad \text{(E.12)}$$

For any filter function symmetric with respect to its central wavelength, i.e. $g(\lambda',\Delta\lambda)=g(-\lambda',\Delta\lambda)$, the 2nd term is zero due to odd-even symmetry of the integrand and thus E.12 becomes:

$$\cos\left(\frac{2\pi}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}\right)\int g(\lambda',\Delta\lambda)\left\{\cos\left(\frac{2\pi}{\lambda_0}\frac{\lambda'}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}\right)\right\}d\lambda' \quad \text{(E.13)}$$

Noting that the filter function is real allows E.13 to be succinctly expressed as:

$$\cos\left(\frac{2\pi}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}\right)\mathrm{Re}\left\{\int g(\lambda',\Delta\lambda)e^{-i2\pi\left[\frac{\vec{B}_{jk}\cdot\vec{\theta}}{\lambda_0^2}\right]\lambda'}d\lambda'\right\} \quad \text{(E.14)}$$

Thus the term in brackets is the Fourier transform of the filter function, shifted to its center wavelength, with the frequency variable identified as $\vec{B}_{jk}\cdot\vec{\theta}/\lambda_0^2$. If we renormalized E.11 by dividing through by $\int g(\lambda',\Delta\lambda)d\lambda'$ and use E.14 to yield for the point spread function in the quasi-monochromatic approximation as:

$$PSF(\vec{\theta})\approx\frac{1}{N^2}PSF_{PB}(\vec{\theta};\lambda_0)\left\{N+2\sum_{k=1}^{N-1}\sum_{j>k}^{N}\frac{\int g(\lambda',\Delta\lambda)e^{-i2\pi\left[\frac{\vec{B}_{jk}\cdot\vec{\theta}}{\lambda_0^2}\right]\lambda'}d\lambda'}{\int g(\lambda',\Delta\lambda)d\lambda'}\cos\left(\frac{2\pi}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}\right)\right\} \quad \text{(E.15)}$$

This somewhat unwieldy expression can be simplified by recognizing the temporal visibility function as:

$$V_t(\vec{B}_{jk},\vec{\theta},\lambda_0,\Delta\lambda)\equiv\frac{\int g(\lambda',\Delta\lambda)e^{-i2\pi\left[\frac{\vec{B}_{jk}\cdot\vec{\theta}}{\lambda_0^2}\right]\lambda'}d\lambda'}{\int g(\lambda',\Delta\lambda)d\lambda'} \quad \text{(E.16)}$$

The temporal visibility is normalized to be unity at $\vec{B}_{jk}=\vec{\theta}=0$ and represents the local fringe contrast for a given baseline and sky location.

Thus the quasi-monochromatic point spread function becomes:

$$PSF\left(\vec{\theta}\right) \approx \frac{1}{N^2} PSF_{PB}\left(\vec{\theta};\lambda_0\right)\left\{N + 2\sum_{k=1}^{N-1}\sum_{j>k}^{N} V_t\left(\vec{B}_{jk},\vec{\theta},\lambda_0,\Delta\lambda\right)\cos\left(\frac{2\pi}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}\right)\right\} \qquad \text{(E.17)}$$

The normalization of the PSF is such that $PSF\left(\vec{\theta}\right)_{\vec{\theta}=0} = PSF_{PB}\left(\vec{\theta};\lambda_0\right)_{\vec{\theta}=0} = 1.$

For a Gaussian filter function, with FWHM of $\Delta\lambda$, of the form $g\left(\lambda-\lambda_0,\Delta\lambda\right) = e^{-\frac{4\ln 2}{\Delta\lambda^2}(\lambda-\lambda_0)^2}$ the visibility function (E.16) evaluates to:

$$V_t\left(\vec{B}_{jk},\vec{\theta},\lambda_0,\Delta\lambda\right) = e^{-\frac{\pi^2}{4L_c^2\ln 2}\left[\vec{B}_{jk}\cdot\vec{\theta}\right]^2} \qquad \text{(E.18)}$$

where $L_c \equiv \lambda_0^2/\Delta\lambda$ is the temporal coherence length. The FWHM of the visibility function is given by $\vec{B}_{jk}\cdot\vec{\theta} = \frac{4(\ln 2)}{\pi}L_c \approx 0.88L_c$.

The temporal visibility is a function of: baseline, angle on the sky and coherence length. For a given baseline spacing the fringe visibility changes spatially on the sky, with the greatest visibility occurring at the origin and generally decreasing as one moves away from the PSF core in the in the direction of the baseline.

For a Fizeau interferometer it is improper to think of a given fringe as a cosine term with the same contrast across the entire focal plane; the fringe contrast generally decreases away from the core along the baseline direction. The temporal visibility function is actually given by the Fourier transform of the spectral transmission.

The point on the sky, along the baseline direction, at which the visibility function falls by ½ occurs at $\theta \approx 0.44\,L_c/B_{jk}$ with the visibility, or fringe contrast, decaying more rapidly for increasing baseline. The first fringe maxima (and fringe spacing) occurs at λ_0/B_{jk} and thus the number of spatial fringes at which the visibility falls by ½ is $\frac{0.44\,L_c/B_{jk}}{\lambda_0/B_{jk}} = 0.44\frac{L_c}{\lambda_0} = 0.44\frac{\lambda_0}{\Delta\lambda} = 0.44R$ where "R" is the spectral resolution – this is *independent of baseline.* Thus it is of interest to note that for any given baseline and spectral resolution the fringes begin to lose contrast at the same number of fringes from the PSF center but not the same distance.

<u>Compact Circular Source</u>

Assume a compact circular source of angular diameter W_0 of the form:

$$O\left(\vec{\theta}\right) = O_0 circ\left\{\frac{\vec{\theta}}{W_0}\right\} = \begin{cases} 1; & \sqrt{\theta_x^2+\theta_y^2} \le W_0/2 \\ 0; & \text{otherwise} \end{cases} \qquad \text{(E.19)}$$

and that the source is spectrally invariant over the passband of interest, i.e. that each source point has

the same spectrum. The image is given by the 2D convolution of the source with the point spread function: $I(\vec{\theta}) = PSF(\vec{\theta}) ** O(\vec{\theta})$. Using the equation for the monochromatic PSF gives the image as:

$$I(\vec{\theta}) = O(\vec{\theta}) ** \frac{1}{N^2} PSF_{PB}(\vec{\theta};\lambda_0)\left\{N + 2\sum_{k=1}^{N-1}\sum_{j>k}^{N}\cos\left(\frac{2\pi}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}\right)\right\} \quad \text{(E.20)}$$

The source diameter is much smaller in extent that the primary beam point spread function for all sources of interest for SI, i.e. the source is unresolved with respect to a primary beam PSF. This allows the primary beam PSF to be approximately extracted from the convolution to yield:

$$I(\vec{\theta}) \approx \frac{1}{N} PSF_{PB}(\vec{\theta};\lambda_0) + \frac{2}{N^2} PSF_{PB}(\vec{\theta};\lambda_0)\sum_{k=1}^{N-1}\sum_{j>k}^{N} O(\vec{\theta}) ** \cos\left(\frac{2\pi}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}\right) \quad \text{(E.21)}$$

The convolution term, since the object is real, can be expressed as:

$$O(\vec{\theta}) ** \cos\left(\frac{2\pi}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}\right) = \iint O(\vec{\theta}')\cos\left(\frac{2\pi}{\lambda_0}\vec{B}_{jk}\cdot(\vec{\theta}-\vec{\theta}')\right)d\vec{\theta}' = \mathrm{Re}\left\{e^{i\frac{2\pi}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}}\iint O(\vec{\theta}')e^{-i\frac{2\pi}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}'}d\vec{\theta}'\right\} \quad \text{(E.22)}$$

For a circular object E.22 evaluates to:

$$\mathrm{Re}\left\{e^{i\frac{2\pi}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}}\frac{2J_1\left(\pi\frac{B}{\lambda_0}W_0\right)}{\pi\frac{B}{\lambda_0}W_0}\right\} = \frac{2J_1\left(\pi\frac{B_{jk}}{\lambda_0}W_0\right)}{\pi\frac{B_{jk}}{\lambda_0}W_0}\cos\left(\frac{2\pi}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}\right) \quad \text{(E.23)}$$

Thus the image of a circular source can be expressed at a single wavelength as:

$$I(\vec{\theta}) \approx \frac{1}{N} PSF_{PB}(\vec{\theta};\lambda_0) + \frac{2}{N^2} PSF_{PB}(\vec{\theta};\lambda_0)\sum_{k=1}^{N-1}\sum_{j>k}^{N}\frac{2J_1\left(\pi\frac{B_{jk}}{\lambda_0}W_0\right)}{\pi\frac{B_{jk}}{\lambda_0}W_0}\cos\left(\frac{2\pi}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}\right) \quad \text{(E.24)}$$

or for a more general source shape and quasi-monochromatically by including the temporal visibility function as and identifying E.23 as the spatial visibility function:

$$I(\vec{\theta}) \approx \frac{1}{N^2} PSF_{PB}(\vec{\theta};\lambda_0)\left\{N + 2\sum_{k=1}^{N-1}\sum_{j>k}^{N} \mathrm{V_t}(\vec{B}_{jk},\vec{\theta},L_c)\mathrm{V_s}(\vec{B}_{jk},W_0,\lambda_0)\cos\left(\frac{2\pi}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}\right)\right\} \quad \text{(E.25)}$$

Equation E.25 is normalized to be unity at the origin. E.25 can be renormalized to unit integral by

dividing through by $\int I(\vec{\theta})d\vec{\theta} = \frac{1}{N}\frac{4\lambda_0^2}{\pi D^2}$ to yield:

$$I'(\vec{\theta}) = \frac{\pi D^2}{4\lambda_0^2} PSF_{PB}(\vec{\theta};\lambda_0)\left\{1 + \frac{2}{N}\sum_{k=1}^{N-1}\sum_{j>k}^{N} \mathrm{V_t}(\vec{B}_{jk},\vec{\theta},L_c)\mathrm{V_s}(\vec{B}_{jk},W_0,\lambda_0)\cos\left(\frac{2\pi}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}\right)\right\} \quad \text{(E.26)}$$

where the primary beam PSF is normalized to unity at the origin ($\vec{\theta} = 0$). Thus the focal plane distribution of photo-electrons is $n_e I'(\vec{\theta})$ where n_e are the total number of photo-electrons. Thus we have:

$$\boxed{n(\vec{\theta}) = n_e \frac{\pi D^2}{4\lambda_0^2} PSF_{PB}(\vec{\theta};\lambda_0)\left\{1 + \frac{2}{N}\sum_{k=1}^{N-1}\sum_{j>k}^{N} \mathrm{V_t}(\vec{B}_{jk},\vec{\theta},L_c)\mathrm{V_s}(\vec{B}_{jk},W_0,\lambda_0)\cos\left(\frac{2\pi}{\lambda_0}\vec{B}_{jk}\cdot\vec{\theta}\right)\right\}} \quad \text{(E.27)}$$

E.27 expresses the photo-electron distribution in the focal plane (projected on the sky) for a quasi-monochromatic compactly supported extended source.

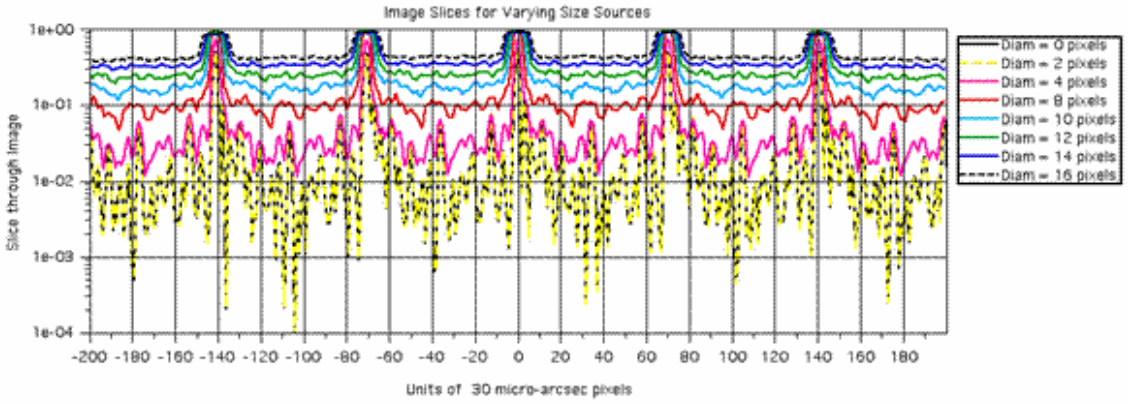


Figure E.6: Extended Scene Image Slices

(R. Lyon).

Appendix F: Alternative Concepts for "Primary Array"

The challenge for this mission is to create an ultraviolet-optical telescope with the resolution of at least a 500-meter-diameter aperture, with a portion of the program requiring 1000 m baselines. The separation of the aperture from the detectors will be in the range of 0.5 to 10 kilometers, depending on how fast an optical system the particular design will allow. Over this range of distances, we need to sense and control the structure to adjust the optical wavefront at the focal plane to the order of a tenth of the wavelength of the radiation in which we wish to image.

This is a challenging mission and it is important to choose the design that is most feasible. Since any design concept for this mission requires the invention of new technology, our challenge is to predict which inventions will be the least difficult. Fortunately, based on our present understanding of optics and opto-mechanical systems, there are only two main design trades that need to be made at this point in the study. These are 1) full-aperture versus sparse aperture and 2) within the sparse aperture regime: a large number of small telescopes vs. a small number of large telescopes. These decisions affect the fundamental technology that needs to be developed. We chose a sparse aperture design with a large number of small apertures as the baseline design for the Vision Mission Study. In this Appendix we discuss the rationale for that choice, which only gets more compelling as the desired "aperture" (i.e., baseline) gets larger. Within the chosen design, there remain a large number of lesser decisions. For example, the arrangement of apertures and the details of how the beam combination should be achieved. These choices do not have a major affect on the fundamental technology required for the mission and are discussed in Appendix G. Figure F.1 shows a basic sparse aperture telescope design.

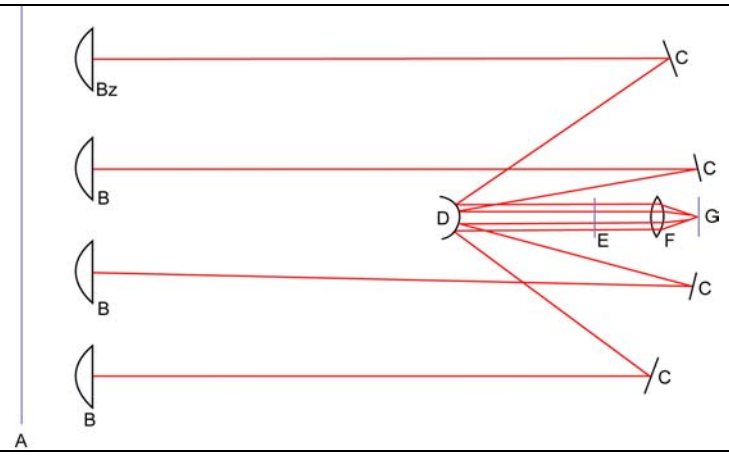


Figure F.1: Schematic drawing of a sparse aperture telescopes.

Surface A is the aperture plane. It is usually, but not always coincident with the collectors B which are drawn as lenses for simplicity but will be mirrors in the case if SI. Optics C and D represent the transfer optics that deliver light collected at A to the central hub. Plane E is an image of plane A formed by the intermediate optics. This is the entrance aperture to the beam combiner. F represents the beam combiner optics. It job is to provide the Fourier transform of the E-field in plane F on the Image plane G. In a Fizeau interferometer, plane E is a to-scale image of plane A. This gives a field of view limited only by aberrations in the optics. In other configurations, the transfer optics image their own segment of the aperture plane at E but their relative separations are not maintained.

F.1 Filled-Aperture Designs

A normal filled-aperture telescope is prohibitive in cost at the desired sizes. A mirror that is lightweight enough to make a 500 meter diameter telescope feasible is likely not to exist until the distant future. People who dream about large, lightweight optics talk about tens of meters, not hundreds. To make a 500 meter diameter, filled-aperture telescope will require an extreme design. The most promising of those designs proposed to date are either a Fresnel Zone Plate (FZP) or the closely related Photon Sieve.

These are diffractive optical elements. The idea behind these designs is to make the powered element from a thin sheet. In a normal lens, the thickness of a high index of refraction material is adjusted so that the optical path to the focus is the same from any point on a plane in front of the lens. For a FZP, regions where the optical path to the focus differ by an integral number of wavelengths are allowed to pass while those a half-wave out of phase are blocked. Efficiencies as high as about 20% are feasible for narrow band light. This design is considered interesting because of the claim that a transmissive optic does not have to be held with the same precision as a reflective optic. The lens would be a thin sheet consisting of alternating transmissive and reflective (or absorptive) rings. A picture of a Fresnel Zone Plate (abbreviated FZP) is shown in Figure F.2. Further information on the use of FZP's as space telescopes can be found in Hyde (1999), and Hyde et al. (2002) Barton (2001), and Meinel and Meinel (2002, 2003).

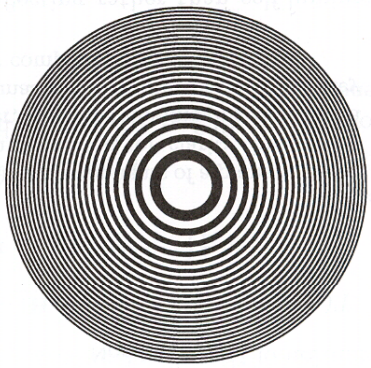


Figure F.2: Image of a Fresnel Zone plate, from H. Barrett and Myers (2004)

A Photon Sieve is a sparse-aperture version of a FZP. Compared to a FZP, a Photon Sieve will have poorer throughput but it is still high throughput compared to the baseline design. It is deserving of its own name because proper choice of what portions of the transmissive zones should be filled in can actually improve the point spread function. For Stellar Imager, the advantage of a photon sieve is mechanical. The parts of the transmissive bands of a FZP that are filled in to form a photon sieve provide support for holding the blocking portions of the screen in place. This is necessary since for the ultraviolet wavelengths, the transmissive zones have to be vacuum. Figure F.3 shows a picture of a photon Sieve from Anderson (2005). Additional information on photon sieves, in particular for soft x-ray imaging, can be found in: Kipp (2001), Waldman (1966), Levashov and Vinogradov (1994) and Q. Cao and Jurgen Jahns in 4 seminal papers (2002, 2003a, 2003b, 2004).

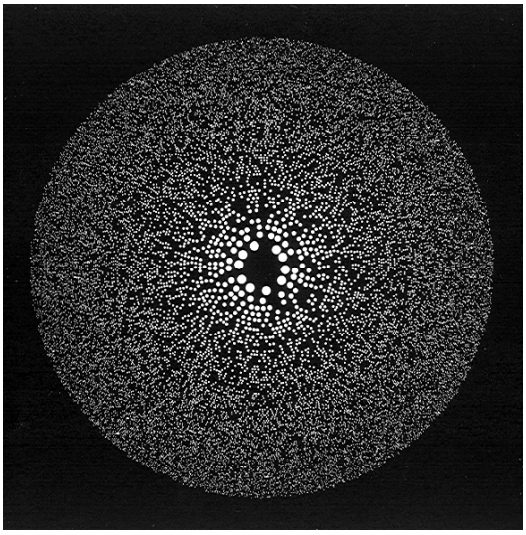


Figure F.3: An image of a photon sieve for optical applications (Anderson 2005)

The estimated mass of a 500 meter diameter aperture assuming a density of 200 grams per square meter is about 39,000 kg. This is 1.7 times heavier than the payload a Delta IV Heavy can lift to low Earth orbit. It does not include the mass of the structure and equipment that will be needed to control the film after it is deployed. The FZP option for SI will require new materials (for a lighter film), new technologies (for stowing and deploying the film) and a new launch vehicle to lift that much mass into a high orbit. Next, we consider control of the sheet after it is deployed in orbit.

The phase plate for a FZP consists of concentric rings alternating between transmissive and reflective. Each consecutive transmitting ring is one wavelength further from the focal point. For a focal length, f, the radius, r, of the k^{th} ring is therefore given by

$$(f + k\,\lambda)^2 = f^2 + r^2$$

or

$$r^2 = 2\,f\,k\,\lambda$$

If the $(k+1)^{st}$ ring has a diameter of $r+\delta$ we have

$$2\,f\,(k+1)\,\lambda = (r+\delta)^2.$$

By combining the last two equations

$$\delta = f\,\lambda / r = 2\,F\,\lambda.$$

F is the f-number of the system at that radius r and λ is the wavelength. The reflective gap between successive transmitting regions is $\delta/2$ and the tolerance on positioning each ring is on the order of $\delta/5$ to $\delta/10$ depending on the desired quality of the resulting image. Assuming the looser requirement and

assuming we can maintain positions to 1 μm, we need to build an F/16 system (8 km focal length) to make observations at 1550 Å. One can argue that the outermost ring, which is the most critical, is also the easiest to control since actuators can be attached to the edge of the lens. But it is difficult or impossible to place actuators in the interior of the lens without blocking transmissive regions. Rings that are at r/2 have only a factor of two looser tolerances but can be hundreds of meters from the nearest actuator.

This is a more difficult control problem than the baseline design despite the fact that micrometer tolerances are quite a bit looser than the ten nanometer tolerances needed for the baseline design. This is because the baseline design is a number of small, rigid bodies whose positions and orientations can be sensed with a metrology system and independently controlled in a straightforward fashion by applying forces and torques. The FZP is a single structure which needs to be maintained to a fractional accuracy of $\delta L/L$ = 1 μm/500 m = 2×10^{-9}. This type of mechanical stability is possible in small, thermally controlled structures, but borders on impossible for a large structure that is not built from low CTE materials and is subjected to full solar illumination at varying angles of incidence. A *uniform* size change yields a focal length change, which could be compensated by focus optics in the combiner or by moving the combiner axially, but a *non-uniform* expansion cannot be compensated for in this fashion.

In addition to the seemingly insurmountable static problems, there is also the issue of dynamics. The normal control scheme for the membrane is to rotate it for stability and then use thrusters located along the outer edge of the membrane for control. But these thrusters will be hundreds of meters from some places in the membrane that need to be controlled and the mechanical connection is through the membrane which is both floppy and has a complicated response function due to the holes which pass the star light. Mounting additional thrusters directly to the membrane destroys reliability due to the high probability of ripping with anything but perfect control. Adding ribs and connecting both the membrane and thrusters to them makes the mass budget even more problematic. Although sensing and control in the baseline design is far from trivial, it more closely related to problems with known solutions. The dynamics and control of large membranes is space is a subject with only theoretical results.

F.2 Signal-to-noise: do we need a filled aperture?

Fortunately, the target stars for Stellar Imager are bright - the large aperture is needed for resolution, not sensitivity. Characteristic fluxes for sample target stars are ~ 1.6×10^{-13} ergs/cm^2/sec at 1550A and 8×10^{-13} ergs/cm^2/sec at 2800A. A detection with a signal to noise of 100 with a filled aperture can be achieved in 20 milliseconds at 1550 Å and 2.25 ms at 2800 Å. These integration times assume 20% efficiency.

Clearly the integration times for a sparse aperture will be longer. Fienup (2002, 2000) and Ribak et al. (1988) showed that to obtain an image using a sparse aperture telescope with fill fraction a that is comparable to what would have been obtained with a filled aperture with equivalent resolution, the integration time needs to be increased by a^{-3}. These calculations assume that the image is of an extended target. In the regime where the target is unresolved by the individual apertures, we do not need to sample all spatial frequencies in the image and the ratio of integration times is increased to $1/[a/N^2]$, where N is the number of apertures. The integration times needed for the baseline design is on the order of 4.25 hours for the 1550 Å image and just less than ½ hour for the 2880 Å image. While it would be nice to have shorter integration times, these are sufficient for meeting the mission's science requirements.

F.3 Sparse-apertures -- How many apertures?

Having settled on a sparse-aperture design, the most important consideration is the number of apertures. In order to image an object with k by k resolution elements across its field of diameter θ, we need to measure k^2 baselines, roughly evenly distributed in a rectangle of diameter kλ/θ. In the baseline design, all of these measurements are made at the same time. With N apertures, we can measure up to N(N-1)/2 baselines at one time. In practice the number of useful baselines will be a little smaller since either the same baseline will be measured by different pairs of telescopes or some other baseline will fall outside the required region. With a smaller number of telescopes we measure fewer baselines at a time then reconfigure the spacing and orientation of the telescopes to fill in the rest of the required baselines. With a smaller number of apertures, efficiency is increased because fewer unnecessary baselines are measured and fewer necessary baselines are missed. On the other hand, the time it takes to reconfigure takes away from observing time and hurts efficiency. An exception to this may be with a very small number of apertures. For example, two apertures rotating around each other in an increasingly widening spiral can provide uniform coverage of the MTF with no loss of integration time to reconfiguring as long as observations can be made while the spacecraft are moving. Efficiency may be a net plus for a small number of apertures, but this is a small advantage.

Given a constant total collecting area, integration time is roughly independent of how that collecting area is divided between apertures. To see this for the photon-noise limited case applicable to SI, we will use the sensitivity calculation for a Michelson beam combiner since the calculation is independent of whether the combiner is a Michelson or Fizeau design and for the case involving multiple reconfigurations, the calculation is easier for a Michelson design. Assuming a Michelson beam combiner with N apertures, each of diameter D, the signal to noise on a single baseline is given by

$$SNR^2 = (F\ \ NT\pi D^2/4)(2/N)^2 = \pi F(D^2/N)T.$$

The first term, with F representing the detected number of photons per unit area and time, is the total number of collected photons. The second term is the reduction of fringe contrast due to combining multiple signals on the same detector. With N apertures, N(N-1)/2 baselines are measured simultaneously and the integration time per baseline is given by

$$T = (SNR)^2/[\pi F(N-1)D^2]$$

$$ND^2T = [N/(N-1)](SNR)^2/\pi F$$

The left hand side is the product of the total collecting area times the integration time. The leading coefficient on the right-hand side is 2 for 2 apertures and slowly decreases to 1 for a very large number of apertures. Now, while a factor of two in integration time is well worth pursuing, this is also a small factor.

Most designers of telescopes will admit that both the cost and the mass of a telescope increases more rapidly than the collecting area; and the cost, at least for ground-based systems grows at a rate between $D^{2.3}$ and $D^{2.7}$. In space, the cost couples strongly with mass and may be closer to D^3. For a

constant cost mission, the increase in collecting area achieved by using 30 small telescopes rather than 2 large telescopes should be a factor of 2 to 2.5. Thus the integration times for a constant-cost mission should be roughly independent of how many apertures are deployed.

There are additional advantages for a mission with a large number of apertures. Having a large number of telescopes builds redundancy naturally into the system. The reliability of the telescopes in a mission with two telescopes where both are essential needs to be substantially higher than in a mission with 30 telescopes where the loss of one or even two telescopes would have a negligible effect on the mission. There is also an advantage for observing variable sources. When all spatial frequencies are measured simultaneously, the resulting image is an average of what the star looked like during the observation. If the spatial frequencies are measured sequentially, the resulting image can contain artifacts since different spatial frequencies are taken from the image of the star at different times – this is especially important for the prime science targets which are intrinsically variable and rotating, as well as, in many cases, moving across the line of sight, at rates significant to the timescale of observation.

We therefore chose a system with a large number of apertures to ensure timely and unambiguous observations of moving and temporally varying targets and to ease the redundancy and reliability issues.

Appendix G: Alternative Beam Control and Combiner Designs

In Appendix F, we described the reasoning that shaped the overall design of SI. This appendix deals with the next level of the design -- the placement of the apertures and the way the beams are combined in the central hub. These decisions are important for obtaining the best possible performance from SI but they are secondary considerations in that they have little effect on the appearance of the mission or the technology that needs to be developed. In the body of this report we presented one design that appears to be viable. Because they have received a lot of publicity over the last several years, we spent a part of this study examining the possibility of using a hyper-telescope. That concept fails miserably but while studying it we proceeded down a line of modifications that eventually lead to an alternative design which we labeled "One-dimensional spatial frequency remapping." This concept not only appears to be viable but seems to have some properties that may make it the preferred choice. However, this concept has not yet been studied in as much detail as the baseline design. It mostly uses the same technology as the baseline design and if anything has somewhat looser tolerances. We describe our current understanding of it and its potential benefits in this Appendix. It is our intention to pursue this option in further detail beyond the end of the current Vision Mission study.

We start in Section G.1 with a description of a Fizeau interferometer, point out its shortcomings and present the hyper-telescope as an alternative and explain why it will not work. A discussion of the hyper-telescope's shortcomings leads us to a discussion of redundant versus non-redundant configurations in Section G.2 and a discussion of what conditions need to be met in order to guarantee a sparse aperture telescope can image a particular source. In Section G.3, we make a short diversion and discuss path-length control and how the choice of a beam combination design can convert a requirement on path length control into one of knowledge and stability. Finally, in Section G.4 the hyper-telescope design is modified to produce a viable alternative that not only combines most of the advantages of the other options but also *provides a natural way to build-in spectral resolution*.

G.1 Hyper-Telescope

To understand the concepts presented in this section, we need to understand the basic operation of a sparse-aperture telescope. These telescopes can be understood by characterizing three optical planes (Figure G.1). At the front end we have all the individual apertures (surface A). This surface, the observatory entrance pupil, is characterized by the relative position and diameter of each aperture. We will call A the aperture plane. The light from all the apertures must be transferred to the entrance of the central hub. The relevant plane here is B, the entrance pupil of the beam combiner. At the back end there is an image plane C.

In a Fizeau interferometer, the image plane is a convolution of the image of the source and a point spread function (which is independent of field angle). This light transfer can be accomplished by making B a scaled image of A. By scaled image we mean that there is a single scale factor *m* applied to both the aperture diameters and their locations in the aperture plane to give the beam diameters and locations in the beam combiner entrance pupil B. Thus if a photograph of the telescopes looks like Figure G.2(a), then at plane B the beams also look exactly like G.2(a) except with a much smaller lateral scale.

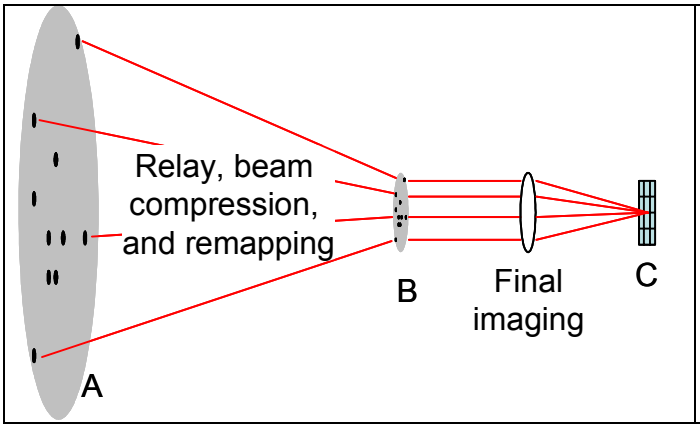


Figure G.1: Significant planes in a sparse aperture telescope system.
The collecting telescopes are in plane A, which has a typical lateral scale of tens to thousands of meters. Their beams are transferred to the combiner entrance pupil B, which typically would be meter-sized. For a focal-plane or Fizeau interferometer, the final imaging combiner can be as simple as an ideal lens. Then finally at the focal plane C, the beams overlap and produce fringes.

Following that, the simplest beam combiner (going from B to C) is a lens. The image at C is then the squared modulus of the Fourier transform of the E-field at B, which is in turn a scaled copy of the E-field at A. It follows that the image plane is the image of the target convolved with a Point Spread Function (PSF) and the PSF is the Fourier transform of the autocorrelation of the aperture plane. This is true over an almost arbitrarily large angle limited only by our ability to design and fabricate the optics.

A limitation of a Fizeau interferometer is that for very small fill fractions, the PSF is very spread out with many side lobes and a low Strehl ratio, equal to the fill fraction. Labeyrie pointed out that if we do not need the wide field of view we do not have to maintain the rigid scaling between A and B. He suggested close-packing the beams at the beam combiner as in Figure G.2(d). Now, for a source exactly on axis, at B the stellar field approximates a uniformly illuminated, filled aperture with a flat wavefront, and the point spread function approaches an Airy disk with a single peak and unit Strehl. Labeyrie's innovation was to point out that with a very large number of apertures it is possible to obtain a non-zero field of view. It is still necessary to 1) scale all aperture diameters by the same amount, m_A, and 2) scale the separations by the same amount, m_S, between the two planes. But m_A and m_S can be different. Thus, if we are to achieve a close-packed configuration in plane B, the apertures in plane A must also be in a highly redundant configuration, as in Figure G.2(c) and therefore a very large number of apertures are needed to achieve a usable field of view. This configuration is called a hyper-telescope and the ratio of m_A to m_S is called the densification.

Forming a nearly perfect image of the target in the focal plane is a very attractive option. Unfortunately it takes a very large number of apertures to generate a usable field of view. To achieve the goals of SI would take nearly 1000 separate apertures. This is truly a formation-flying nightmare compared to the order of 30 apertures in the baseline design. This should not be surprising since another way to look at this problem is to realize that in order to fill the beam combiner entrance pupil, the apertures need to be in a highly redundant formation. For circular or hexagonal apertures they need to fall on a hexagonal grid; for square apertures they need to fall on a square grid and in both cases they need to fill all adjacent grid points.

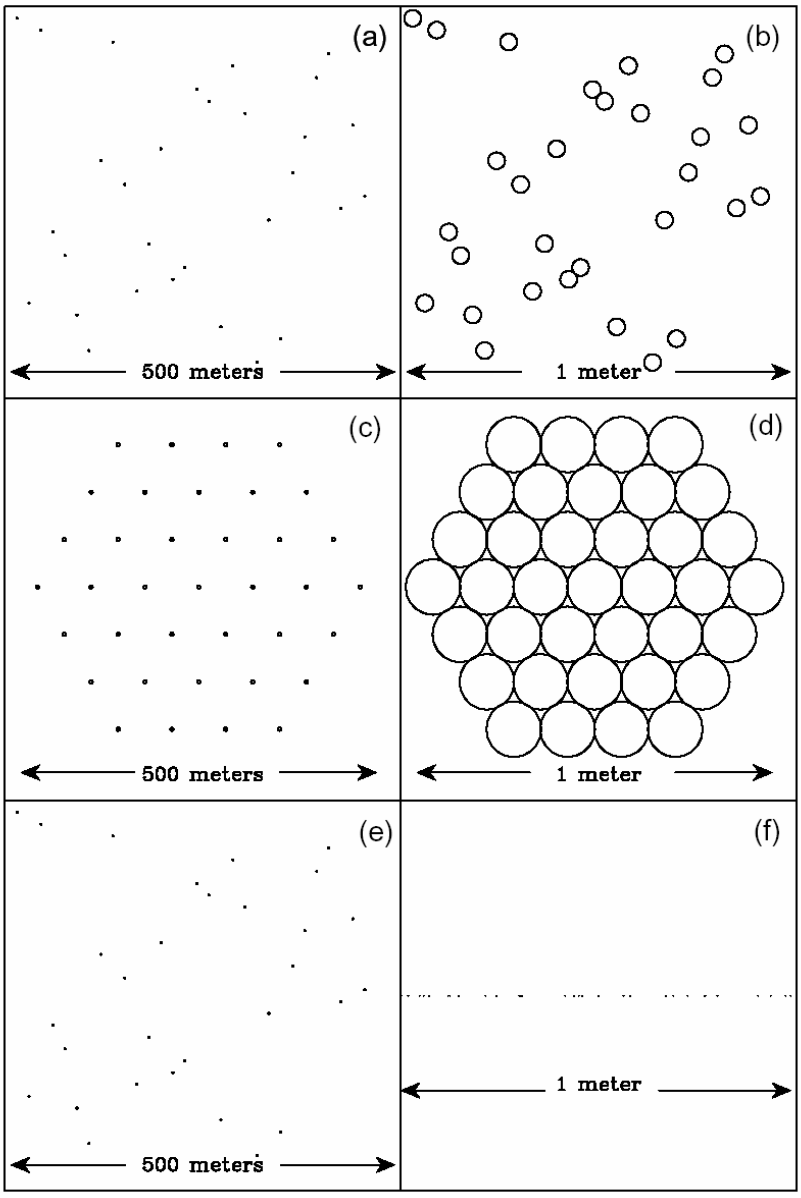


Figure G.2: Aperture layouts for aperture plane (left column) and combiner entrance pupil (right column) for different beam-combination architectures.

Panel (a) - the standard Fizeau interferometer: shows the mirror locations and diameters in the aperture plane. The combiner entrance pupil is not shown for this case, since it looks identical to (a) but with a much smaller scale. This is the requirement for a wide field of view.

The first row, panels (a) and (b) - a modified hyper-telescope (using partial densification): shows that the locations and beam diameters are scaled differently, so that the nearest beams touch in the pupil plane. Well-defined PSF with small FOV, but one that is larger than for hyper-telescope. Requires significant post-processing to form image.

The second row, panels (c) and (d) - a true hyper-telescope: shows the pupil is fully-densified (all the beams touch). This requires a redundant layout and many more mirrors than other options. FOV is small, but non-zero.

The third row, panels (e) and (f) - Spatial-frequency remapping interferometer in which the beams are remapped into a non-redundant linear arrangement: The combiner entrance pupil is mapped into one dimension, so the second dimension of the detector can be used for spectral dispersion, eliminating the need for an energy-resolving detector and its associated cyrogenics.

G.2 Redundant versus Non-Redundant Configurations.

The autocorrelation of the E-field sampled by the aperture plane is called the Optical Transfer Function. Usually, we are only interested in its magnitude, the Modulation Transfer Function, MTF. The MTF is the Fourier Transform of the PSF and is important because its non-zero values indicate the spatial frequencies which the optical system can "see" well in the scene. For a telescope with a very sparse aperture, such as SI, the MTF is approximately a summation of delta functions; one for each pair of apertures. The gaps in between represent features of the scene to which the system is blind. Here is the real advantage of using non-redundant aperture spacings: it gives the best possible sampling of the scene's features. The number of spatial frequencies sampled by an array of N elements grows as N(N-1)/2 for a completely non-redundant configuration but a little slower than 2N for a redundant array on a square grid. This is also the heart of the reason why a hyper-telescope needs so many apertures to form an image: to get 100 peaks in the MTF, the hypertelescope needs 50 apertures while the non-redundant array needs only 15.

While on the topic of spatial frequency sampling, it is time to address the issue of why a sparse aperture telescope works. Clearly the largest aperture spacings determine the maximum resolution in the image. But what determines the fidelity of the image and what determines whether the image after deconvolution of the PSF is unique? Although it is difficult to answer these questions in general, it is possible to prove uniqueness and therefore fidelity in one special case. If the spatial frequencies are sampled on a regular grid with spacings $k=B/\lambda$, then the image is unique provided the object has an extent of no greater than λ/B. Thus, while SI with ~30 elements will be able to make an image with roughly 500 resolution elements on the stellar surface, it will not be able to increase the number of resolution elements simply by looking at a larger star. The larger star requires shorter minimum baselines if the image is to be unique. Adjusting the grid spacing to match the star being observed results in the same number of resolution element across the star regardless of the diameter of the star (as long as the maximum baseline available to SI is not exceeded).

The adopted configuration for the baseline design is a minimum-redundancy Golomb rectangle because configurations for N=30 exist (Golomb 1982). These are not ideal configurations because some required spatial frequencies are missing. The authors know of no configuration for more than 18 elements that has been proven to be optimum; but if better configurations are discovered, undoubtedly the configuration will be adjusted before launch.

G.3 The Use of Closure Phase to Reduce Path length Control Requirements

There is a more subtle reason why non-redundant array configurations are important -- redundant arrays suffer from speckle noise, while non-redundant arrays do not. Redundancy means we have two different pairs of telescopes with the same baseline vector, and thus sampling the same spatial frequency in the scene. This means the errors for two pairs of telescopes are added before detection. Thus the effects of these errors are entwined and cannot be isolated.

So far we have only discussed a perfect sparse aperture telescope. We now discuss the effect of phase errors between the apertures. We start by assuming an observation of a star consists of a large number of short exposures. We assume the phase errors vary from exposure to exposure but are constant during each exposure. If we know or can determine the phase errors after the fact, then the phase errors can be corrected in post-processing, with absolutely no loss of signal to noise. This is because a Fourier transform of the image separates the spatial frequencies. Since each spatial frequency comes from a unique baseline, the phase of that Fourier component is the source phase plus the phase error. We can correct the phase error and do an inverse Fourier transform to form a corrected image. This

process replaces the requirement for phase control to one of knowledge and stability; perhaps a critical advantage. If an array is redundant with multiple baselines contributing to the same spatial frequency, we can correct that spatial frequency for the mean phase error but there is a reduction of the Fourier amplitude that increases with the variance of the phase errors on the contributing baselines. With knowledge of the phase variance we can boost the amplitude by the appropriate amount but that also boosts the noise. This is why aperture masking on large, ground-based telescopes works better than speckle interferometry.

Even if the phase errors are not known perfectly, non-redundant baselines allow complete isolation of signals from each baseline, and thus can still provide useful information through "closure phase" techniques. Closure phase reconstruction, pioneered at radio wavelengths, is becoming increasingly common in ground-based optical imaging as a way of reducing the deleterious effects of fringe phase instability caused by the earth's turbulent atmosphere. In the space-based systems considered here, station-keeping position drift plays a similar role to atmospheric instability in ground-based interferometry, and computer-based image reconstruction techniques using closure phase and related concepts are therefore likely to be widely used in future space astrophysics imaging systems as well.

Jennison (1958} presented a technique for measuring relative fringe phase which used three radio-linked collectors coupled as three interferometers operating at a wavelength of 2.4 meters over baselines up to ~10 km. Owing to a variety of instrumental problems related to the amplifiers and local oscillator electronics available at the time, the fringe phase between any two collectors was unstable and could normally not be measured; the source structure information had to be derived from the (squared) fringe amplitudes alone (a familiar situation in present-day ground-based optical interferometry). Jennison showed that, if the three observed fringe phases were summed, the resultant combined phase was insensitive to equipment instabilities. With this approach, Jennison & Latham (1959) showed that the brightness distribution of the radio source Cygnus A, which until then was only known to be elongated, actually consisted of two separated sources of nearly equal brightness straddling a peculiar optical object tentatively identified at the time with two galaxies in collision. This was the first observation to reveal the double-lobed structure of a powerful radio galaxy.

Applications of this method to circumvent atmospheric phase instabilities in optical interferometry were described by Jennison (1961) and, apparently independently, by Rogstad (1968). The first use of the term ``closure phase'' seems to be in the paper by Rogers et al. (1974) describing an application at radio wavelengths using very stable and accurate, but independent, reference oscillators at the three stations in a so-called "very-long-baseline" interferometer array. Since that time, closure phase has been used extensively at radio, IR, and optical wavelengths, and there are many papers describing the subject, its virtues, and its limitations. For newcomers, the lectures presented at the "Michelson summer schools" by John Monnier[1] and David Buscher[2] in 2001, and by Peter Tuthill[3] in 2003 are good sources.

One of the major problems affecting the design of future systems under study for high-resolution astronomical imaging using constellations of free-flying interferometers (examples of which include not only the "Stellar Imager" (SI), but the "Terrestrial Planet Finder - Interferometer" (TPF-I), the SPECS sub-millimeter space interferometer, and Black Hole Imager, an X-ray interferometer) is the necessity to maintain the "figure" of the equivalent aperture by keeping the individual elements of the constellation at their designated stations to a high degree of accuracy for extended periods of time. The

[1] http://olbin.jpl.nasa.gov/iss2001/cdrom1/monnier.htm
[2] http://olbin.jpl.nasa.gov/iss2001/cdrom1/buscher.htm
[3] http://msc.caltech.edu/school/2003/2003_MSS/08_Tuesday/peter_tuthill2003.pdf

current view is that the required precision transverse to the optical axis of the constellation is only of the order of a modest fraction of the diameter of the individual collector elements, but in the longitudinal direction parallel to the optical axis (the "boresight", or "piston" direction) the required precision is generally expected to be a small fraction (~1/10) of a wavelength. Sophisticated radio + laser ranging systems may be adequate for station-keeping in the transverse direction, but more elaborate measures will be required in the longitudinal direction. Current thoughts for achieving the required level of accuracy in this direction include fine control using photons from the target itself; unfortunately this means that observations of faint targets may in general be impossible.

R. Allen (2005) has described a simple model of the imaging process for such free-flying constellations and points out that "indirect" imaging systems, where the images are formed by post-processing of interferometer data in a computer, can be made less sensitive to station-keeping errors in the longitudinal direction than are the more classical "direct" imaging systems. This model leads to succinct expressions (formulated as two "theorems") for the precision with which the positions of the collectors must be known in the directions transverse to, and parallel with, the line of sight to the target. In the transverse direction, the required knowledge precision depends on the wavelength and the angular size of the target field of view; for a representative example the required precision turns out to be a few meters. In the longitudinal direction (piston), the relevant scale length is the correlation length of the signal, which is related to the signal wavelength and the bandwidth. For a representative example the required "piston" precision turns out to be a few wavelengths.

Although it clearly makes no sense to build a Stellar Imager that cannot hold its phases, there are potentially three advantages for considering closure phases in the data reduction chain. Firstly, this could be a useful technique during the commissioning or debugging phases of the array. Secondly, it may allow observations of fainter targets where the phase-stabilization techniques do not work as well. And thirdly, closure phases are potentially free from systematic errors and may be the technique of choice when extremely high precision measurements are required, such as for the asteroseismology observations.

G.4 Spatial-Frequency Remapping.

Armed with the knowledge gained from the previous sections, we can improve on the hyper-telescope design. We start with a non-redundant configuration and increase the densification until the beams from the shortest baseline just touch, as in the example in Figure G.2(b). Now, the fraction of the densified pupil that is filled is not unity, but on the order of 1/N, where N is the number of apertures. The fill fraction is smaller than the unity of a hyper-telescope but much larger than in a very sparse Fizeau design. Because the PSF is not a good Airy disk, image processing will be required to construct the image. The configuration of the combiner entrance pupil does not matter as long as it is non-redundant since a Fourier transform of the image plane cleanly separates the signals from each baseline. Those signals can be moved to their "correct" positions in the pupil plane and a Fourier transform then restores the image that would have been detected with the Fizeau configuration. Compared to the standard Fizeau design, one main advantage of this “partial densification” is that it eliminates the need for a variable magnification to match the image scale to the pixels in the detector.

It is possible to improve further on this, since there is no need to maintain any relationship between the spatial frequencies in the aperture plane and the spatial frequencies in the beam combiner entrance pupil. As long as the frequencies in the beam combiner entrance pupil are non-redundant, we can separate them after they are detected, do any phase corrections that are necessary, rearrange their positions, and then Fourier transform back to form an image. This processing can be performed with

no loss of signal to noise. This gives us a useful degree of freedom -- if we choose a one-dimensional configuration for the beam combiner entrance pupil, we can get by with a one-dimensional detector or *we can add a diffraction grating and use the second dimension of the detector for spectral information and thus avoid the need for an extensive set of difficult-to-make UV filters and/or energy-resolving detectors with their associated cryogenics*. For this case, the aperture plane mirror layout is shown Figure G.2(e) and the combiner entrance pupil in Figure G.2(f).

G.4.1 Conceptual Optics for Frequency Re-Mapping

To use one dimension of the detector array for fringe detection and the other dimension for spectral information, the magnifications for the two transverse axes of the optical system must be different. This is because we want to sample the primary beam (single telescope PSF) with more than 1000 pixels along one axis, to capture the fringes, but only two pixels along the other axis for adequate spectral sampling. An example of an optical system that will accomplish this is shown in Figure G.3.
By itself, the first lens forms an image between the second and third lenses. The second lens is a negative focal length cylindrical lens which re-images one axis of the image to the final image plane. The final lens is also a cylindrical lens but rotated 90 degrees so that it re-images the other axis of the image at A to B. The red and green lines correspond to rays in orthogonal views of the system. The ratio of the angles these rays make with the optical axis at the image is the ratio of magnifications. Because this system need only work on-axis, it is possible to achieve very good performance with simple optics. We developed a design which produces nanometer-level wavefronts that uses off-axis mirrors and have achieved a magnification ratio in excess of 1000.

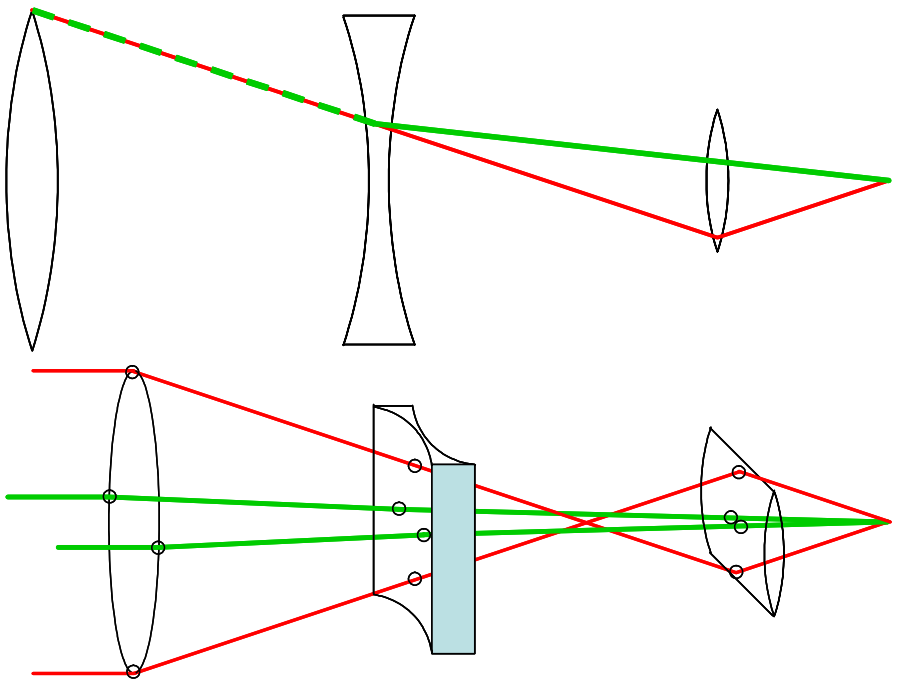


Figure G.3: Two views of a conceptual optical design to remap beams to 1D.

G.4.2 Choice of 1-D Configurations

The final beam combiner design decision is whether all the beams should be combined onto a single detector or if they should be split into two or three groups of beams, each directed to its own beam combiner. The multiple-combiners design does not observe all the baselines simultaneously -- the input beams need to be redistributed between the combiners part way through the observation -- but the signal to noise per baseline per unit time is increased by using multiple combiners in such a way that the total integration time remains roughly constant. The multiple combiner design adds flexibility if we need longer integration times on some baselines and it can shorten total observing time if we do not need some baselines for a particular observing campaign. We leave this decision open but provide one- dimensional configurations for the three cases of 30, 15 and 10 beam combiners:

10 beam combiner positions -- 0 1 6 10 23 26 34 41 53 55

15 beam combiner positions -- 0 6 7 15 28 40 51 75 89 92 94 121 131 147 151

30 beam combiner positions -- 0 12 32 39 49 82 85 100 147 166 206 207 211 286 302 310 316 344
388 399 462 475 500 529 531 552 623 645 671 680

Figure G.4 shows one concept for such a design, from the SI ISAL study. Figure G.2(d) shows the beam layout in plane B for a 10-beam linear non-redundant system.

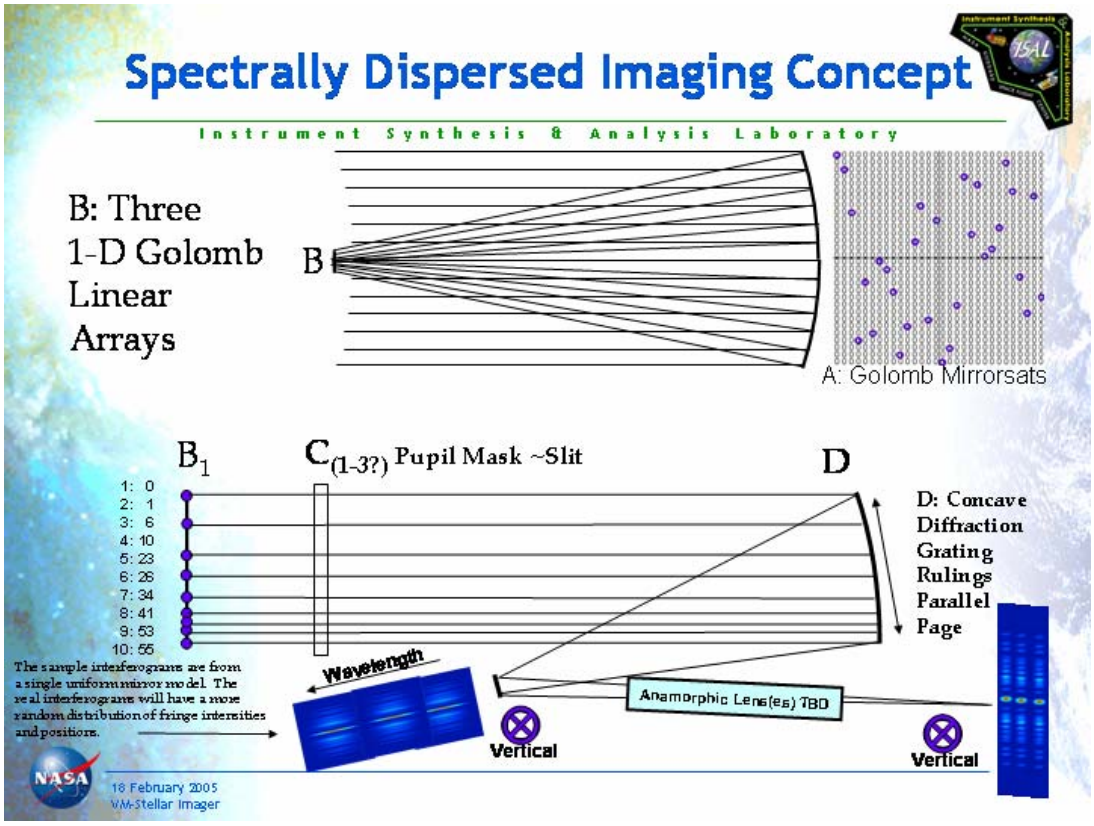


Figure G.4: An alternate design to enable spectral dispersion of the beams.

Potentially, this beam combiner has advantages over the baseline design. The most important is that it provides a natural method for obtaining images in multiple spectral lines simultaneously. It also uses all the light and spreads it over a minimum number of pixels. A fourth advantage is that the array configuration need not be completely non-redundant. Two pairs of apertures with the same spacing contribute to the same spatial frequency in the image but they can be given different frequencies in the combiner. In this case phase corrections can be applied in post processing to all baselines even for partially redundant arrays. This is important since a non-redundant array of the size we need cannot be ideal since it must either miss spatial frequencies we need to observe or measure baselines that are longer than we need. A fifth advantage is that for special purpose applications the beam combination and control does not place special constraints on the arrangement of the apertures. For some observations, such as asteroseismology, a redundant array may provide better signal to noise than a non-redundant array.

List of Figures

List of Tables