RESEARCH STATEMENT

Gamma-ray bursts (GRBs) are the death cries of massive stars and the birth cries of black holes, and the biggest bangs since the Big Bang itself. Due to ultra-relativistic beaming, they are often bright enough to see across the vastness of space and time, even with small telescopes. However, they fade away quickly, sometimes lasting only minutes.

Since arriving at UNC-Chapel Hill in 2002, my research group has been building and developing fully automated, or robotic, telescopes in the Chilean Andes (PROMPT), and now across the world (Skynet), to observe these distant explosions before they fade away (see §A, and for more detail Appendix Sections §B, §C, and §D). And we have been successful: In 2005, we discovered the most distant explosion in the universe then known, 12.8 billion light years away. GRBs are exciting because they are natural laboratories for ultra-relativistic physics, and natural probes of a part of the universe that is so far away and so old that we currently have few other ways of observing it.

When not observing GRBs, we are using the unique observing resources that we have developed in partnership with others to study blazars, rotating and binary asteroids, a wide variety of variable, pulsating, and eclipsing binary stars, as well as to carry out supernova and exo-planet searches. Between GRB and non-GRB science, we are now publishing approximately one journal article per month.

Since GRB afterglows were first discovered in 1997, a wealth of data has been collected. However, these data have never been modeled in a self-consistent or, frankly, statistically valid way, rendering comparative and population studies meaningless. To begin this work now would be an enormous undertaking. But for the past few years, my research group has been laying the groundwork – one-third in the field of statistics (see §E, and for more detail Appendix Section §E.1), one-third in the field of astrophysics (see §E, and for more detail Appendix Section §E.2), and one-third in the field of computer science (see §F, and for more detail Appendix Sections §F.1 – §F.7) – to do just this.

A. PROMPT and Skynet Overview



Figure 1: Three of the six 16-inch diameter PROMPT telescopes.

Funded primarily by NSF, UNC-Chapel Hill has built "PROMPT" - six 16-inch diameter fully automated, or robotic, optical telescopes at Cerro Tololo Inter-American Observatory (CTIO) in Chile - and "Skynet" - telescope control and webbased, dynamic queue scheduling software capable of controlling many telescopes simultaneously and most types of commercially available small telescope hardware. In

partnership with other institutions, many of them in North Carolina, Skynet has enabled us to grow PROMPT into a network of small, robotic optical telescopes. The Skynet Robotic

Telescope Network currently numbers 14 telescopes between 14 and 40 inches in diameter and currently spans South America, North America, and Europe.



Figure 2: The 20-meter diameter radio telescope at NRAO-Green Bank.

We have recently been funded by the American Recovery and Reinvestment Act through NSF, as well as by the Mt. Cuba Astronomical Foundation and NASA, to expand Skynet's geographic and wavelength footprints to include: (1) a new, 32inch diameter robotic telescope at CTIO, with simultaneous near-infrared (NIR), wide-field optical, and lucky optical imaging capabilities; (2) four new, 16-inch diameter robotic telescopes at

Siding Spring Observatory (SSO) in Australia, also with simultaneous NIR and optical imaging capabilities, enabling near-continuous, simultaneous multi-wavelength observing of southern hemisphere targets, as well as live observing for education and public outreach (EPO) in North Carolina; and (3) a 20-meter diameter radio telescope at the National Radio Astronomy Observatory (NRAO) in Green Bank, West Virginia, including the development of a radio version of our telescope control and web-based, dynamic queue scheduling software.

A.1. PROMPT's Intellectual Merit

PROMPT's primary mission is to observe gamma-ray bursts (GRBs) – deaths of massive stars and births of black holes – simultaneously at multiple wavelengths when they are only tens of seconds old. With bulk Lorentz factors of $\Gamma \sim 100$ and isotropic-equivalent luminosities of L ~ 10^{54} erg/sec, they are both probes of ultra-relativistic physics and backlights with which we can probe star-forming regions and the early universe (e.g., ¹⁻³).

GRBs are first detected and localized by spacecraft, currently NASA's Swift and Fermi. To date, GRB localizations have reached PROMPT within 12 - 78 seconds (90% range) of detection. If observable at that time, PROMPT has responded within 14 - 59 seconds (90% range) of notification, with our fastest response being 12 seconds. To date, PROMPT has observed 43 GRBs on such rapid timescales, detecting 24 optical afterglows.

Our most significant discovery to date occurred on September 4th, 2005, when thenundergraduate student Joshua Haislip and PI Reichart discovered and identified the most distant explosion in the universe then known, GRB 050904 at redshift z = 6.3, using both PROMPT and the 4.1-meter diameter SOAR telescope.⁴⁻⁶ For the WMAP cosmology, this redshift corresponds to 12.8 billion years ago, when the universe was only 6% of its current age. Over the past three years, GRBs have also been discovered at z = 6.7, 8.3, and possibly 9.4.⁷⁻¹⁰



Figure 3: Left panel: Near-infrared discovery image of the bright afterglow of GRB 050904 from SOAR atop Cerro Pachon in Chile. Middle panel: Near-simultaneous non-detection of the afterglow at optical wavelengths, implying z > 6, from one of the six PROMPT telescopes atop Cerro Tololo, only 10 km away. Right panel: Color composite image of the very red afterglow 3.2 days after the burst from Gemini South, also atop Cerro Pachon. From Haislip et al. 2006, Nature, 440, 181.

A.2. PROMPT's Broader Impacts

When no sufficiently bright GRBs are observable, which is approximately 85% of the time, PROMPT is used by professional astronomers, students of all ages – graduate through elementary – and members of the general public across North Carolina, the US, and the world for a wide array of research, research training, and EPO efforts.

PROMPT Collaboration institutions include (1) UNC-Chapel Hill, (2) 12 regional undergraduate institutions, including three minority-serving institutions (Appalachian State University, Elon University, Fayetteville State University, Guilford College, Guilford Technical Community College, Hampden-Sydney College, North Carolina Agricultural and Technical State University, UNC-Asheville, UNC-Charlotte, UNC-Greensboro, UNC-Pembroke, and Western Carolina University), (3) UNC-CH's Morehead Planetarium and Science Center (MPSC), and (4) the US and Chilean astronomical communities. PROMPT Collaboration access began on February 1, 2006, only a year and a half after receiving funding, and to date these four groups have used 7,657, 7,081, 1,759, and 15,888 hours of observing time, respectively.

PROMPT, often in campaigns with other optical and radio telescopes around the world and also with space telescopes, is being used to study blazars, rotating and binary asteroids, a wide variety of variable, pulsating, and eclipsing binary stars, as well as to carry out supernova (SN) and exoplanet searches.¹¹⁻²⁶ The largest of these efforts has been the CHilean Automated Supernova sEarch (CHASE),²⁷ which to date has resulted in the discovery of 135 SNe, including at least 38 Type Ia SNe, which are used to measure Hubble's constant and to calibrate cosmic acceleration. PROMPT is now the most successful discoverer of SNe in the southern hemisphere.

PROMPT's most successful EPO efforts have been carried out in partnership with MPSC. Over the past four years, we have trained approximately 75 high school teachers to use Skynet's professional interface (<u>http://skynet.unc.edu</u>; see §A.3) and these teachers have gone on to train thousands of North Carolina high school students using a 127-page curriculum that we

developed (<u>http://skynet.unc.edu/observe.pdf</u>). This curriculum satisfies North Carolina Earth and environmental science graduation requirements.



Figure 4: Skynet's web interfaces have had 59,000 visits per year, most of which have come from these locations, many of them rural, in North Carolina. The average user spends 7 minutes viewing 10 pages per visit.

Also in partnership with MPSC, we

have developed an introductory version of Skynet's interface and have incorporated it into MPSC's "Zoom In!" exhibit. Over the past two years, approximately 18,000 elementary and middle school students, as well as members of the general public, have used it to request observations on PROMPT. PROMPT takes a unique image for each user, emails them a link to it, and then Skynet allows them to request nine more observations from home or school before having to return to MPSC. **Try it yourself:** <u>http://skynet.unc.edu/morehead/authorize.php</u>, password = "reichart".

A.3. Skynet Robotic Telescope Network

Funded primarily by NSF, we have been developing Skynet, which is telescope control and webbased, dynamic queue scheduling software that we originally developed for PROMPT, but is now capable of controlling many more telescopes and most types of commercially available small telescope hardware. This was a proof of concept effort to see if we could expand our geographic footprint beyond CTIO without having to pay for additional telescopes. This experiment has greatly exceeded expectations: The Skynet Robotic Telescope Network now spans three, and soon four, continents. To date, we have integrated ten non-PROMPT telescopes (California, Colorado, Illinois, Italy, three in North Carolina, Virginia, and a 40-inch diameter telescope in Wisconsin), and are currently scheduled to integrate eight more non-PROMPT telescopes (California, Illinois, New Mexico, five in North Carolina, including a 32-inch diameter telescope, and the 20-meter diameter radio telescope in West Virginia) over the next two years:

Site	Telescope	Owner	Online
Cerro Tololo Inter-American	16″	UNC-Chapel Hill	6/05
Observatory, Chile	16″	UNC-Chapel Hill	8/05
	16″	UNC-Chapel Hill	12/05
	16″	UNC-Chapel Hill	12/05
	16″	UNC-Chapel Hill	12/05
	16″	UNC-Chapel Hill	1,3
	24″	National Astronomical Research	1
		Institute of Thailand & UNC-Chapel	
		Hill	
	32"	Astro Optik & UNC-Chapel Hill	1

Yerkes Observatory, WI	40″	University of Chicago	4/11
Dark Sky Observatory, NC ²	14″	Appalachian State University	6/08
	17″	Appalachian State University 11	
	32″	Appalachian State University	1
Astronomical Research	24″	Astronomical Research Institute	1
Observatory, IL^2	30″	Astronomical Research Institute,	4/11
		Eastern Illinois University & Hands-	
		On Universe	
Siding Spring Observatory,	16″	UNC-Chapel Hill	1
Australia ²	16″	UNC-Chapel Hill	1
	16″	UNC-Chapel Hill	1
	16″	UNC-Chapel Hill	1
Morehead Observatory, NC	24″	UNC-Chapel Hill	9/08
Stone Edge Observatory, CA	20″	Stone Edge Farm Vinyards	1
Cline Observatory, NC	16″	Guilford College	1
Dolomiti Astronomical	16″	Carlo Magno Zeledria Hotel	2/09
Observatory, Italy			
Hampden-Sydney College	16″	Hampden-Sydney College	6/09
Observatory, VA			
Pisgah Astronomical Research	16″	Pisgah Astronomical Research	3/08
Institute, NC		Institute	
Rankin Science Observatory, NC	16″	Appalachian State University	1
Coyote Rim Ranch, CO	14.5″	Jack Harvey	11/05
Hume Observatory, CA	14″	Sonoma State University	5/06
McNair Observatory, NC	14″	NC A & T	1
Smithies-White-Edgell	14″	Oliver Smithies ⁴ & Marshall Edgell	1
Observatory, NM			
National Radio Astronomy	20-m	National Radio Astronomy	1
Observatory, WV		Observatory	

¹Will be integrated into Skynet over the next two years.

² Siding Spring Observatory, Dark Sky Observatory, and Astronomical Research Observatory will have simultaneous multi-wavelength imaging capability like Cerro Tololo Inter-American Observatory.

³ PROMPT-6 enclosure in use to complete AAVSO Photometric All-Sky Survey (APASS).

⁴ Nobel laureate

Skynet has proven to be an attractive option for non-PROMPT telescope owners because (1) they no longer need to staff their telescopes at night, or in the case of campus telescopes they no longer need to keep their students awake night after night if they want to do observational astronomy curricula or research/research training; and (2) Skynet allows telescope owners to queue observations on the other telescopes on the network when they are not otherwise in use, giving them free access to different and often better telescopes, instrumentation, parts of the sky, and site and weather conditions. Skynet has now taken over 3.9 million exposures, currently at a rate of about 80,000 per month and this rate is increasing by about 1,000 per month.



Figure 4: GRB and non-*GRB Skynet journal* articles (published or in press).

Over the past six years, but mostly over the past two years, GRB and non-GRB research has resulted in 23 journal articles^{5,11-26,28-34} (including two in Nature,^{5,17} with another approximately half dozen in preparation across the collaboration), two conference proceedings,^{27,35} over 280 observing reports (GCN, CBET, IAUC, MPB, MPC, ATel), two doctoral dissertations,^{36,37} at least five masters theses, and at least three undergraduate honors theses.

See Appendix B: Targeted Expansions of Skynet

See Appendix C: A Broader Discussion of Skynet

See Appendix D: Foundation for Growing Skynet

E. Afterglow Modeling Project (AMP)

The primary research initiative of Skynet's GRB group is the observation and modeling of GRB afterglows. PROMPT was constructed specifically for the purpose of obtaining simultaneous multi-band photometry of GRB afterglows, beginning tens of seconds after initial detection and localization by spacecraft. With the development of Skynet's image reduction and analysis pipeline, which considerably simplifies and speeds up these often tedious tasks, we are now in a position to focus our efforts not only on modeling data collected with PROMPT and other Skynet telescopes, but also on data mined from all published observations, photometric and spectroscopic, from radio to X-ray wavelengths, and to begin compiling a catalog of GRB afterglow properties. This is the Afterglow Modeling Project (AMP).

AMP will model, in a statistically sound and self-consistent way, the time- and frequencydependent emission and absorption of every GRB afterglow observed since the first detection in 1997, using all published or otherwise available observational data. The result will be a catalog of fitted empirical model parameters describing the intrinsic afterglow emission, and extinction due to dust and absorption due to gas along the line of sight to the GRB. This ever-growing catalog of fitted model parameters will allow us to infer the astrophysical properties of GRBs and their environments, and to explore their variety and evolution over the history of the student universe. Mv graduate Adam Trotter's recent Ph.D. dissertation (http://www.physics.unc.edu/~atrotter/thesis/trotter_thesis.pdf) presents a new-and-improved statistical methodology for the construction of afterglow models, as well as new-and-improved versions of the models themselves, including: intrinsic afterglow emission due to synchrotron radiation from shocks in ultra-relativistic jets; extinction due to dust in the source frame of the GRB (which may change with time as the burst evolves), and in the Milky Way; and absorption due to neutral hydrogen in the host galaxy and the intergalactic medium.³⁷ Presented briefly in §E.1 and §E.2, respectively, this work will constitute Papers I and II of the AMP series, currently in preparation for submission to MNRAS. AMP III will present "Galapagos", which is sophisticated, highly parallel, genetic algorithm-based model-fitting software that we have developed to be used in AMP, as well as in other applications (see §F). AMP IV will present the results obtained from all available afterglow observations from the BeppoSAX/IPN era (1997-2000), AMP V from the HETE/Integral era (2001-2004), and AMP VI and onward from annual divisions of the faster-paced and ongoing Swift/Fermi era. My graduate student Justin Moore's Ph.D. dissertation work includes the development of a database and user-friendly, web-based interfaces to streamline the processes of collating the data from various sources, of customizing the emission and absorption models from modular components, and of organizing the fitted outputs for presentation and publication, for each GRB.

Several UNC-Chapel Hill undergraduate and graduate students have already worked on modeling various GRB afterglows with earlier versions of the models, and of Galapagos. The learning curve has at times been steep; however, with Trotter's final models in place and my recent purchase of a 48-core machine, we are now poised to begin modeling afterglows in earnest, including preliminary modeling of afterglows – in real time – as data from PROMPT and other Skynet telescopes come through our now developed image reduction and analysis pipeline. We will need the assistance of all the talent that we can recruit, to keep this project moving forward in a timely fashion. AMP will provide a wealth of research opportunities for both undergraduate and graduate students for years to come, and in a range of sub-disciplines – from applied mathematics and computer science to theoretical astrophysics and cosmology. More than half of the excitement is that we do not yet know what we will discover, or what new tools we will have to invent, when we begin compiling and comparing models of past and future GRBs; this is, literally, unexplored territory.

In §E.1, we introduce a new statistic for fitting data with both statistical and systematic uncertainty in two dimensions. It is both invertible, unlike the statistic of D'Agostini (2005), and reduces to the 1D case, unlike the statistic of Reichart (2001).^{62,63} As a general solution to the problem of fitting data in two dimensions, this work is broadly applicable, not only across astrophysics but across all of science.

In §E.2, we use this statistic to significantly upgrade the dust-extinction and Ly α -forest models of Reichart (2001), which are also broadly applicable, in this case not only to GRB afterglows but to any extragalactic point source.

See Appendix E.1: A New Statistic

See Appendix E.2: AMP Extinction and Absorption Models

F. Galapagos

Galapagos is sophisticated genetic algorithm-based model-fitting software that we have been developing, and recently have begun to use, for many applications, including model-fitting GRB afterglow data obtained by Skynet, as well as data collected with other telescopes and archival data. However, Galapagos is broadly applicable, not only across astrophysics but across all of

science. We describe Galapagos in F.1 - F.5. In F.6 and F.7, we present plans to continue to develop and optimize Galapagos and release it to the scientific community in ways that are both easy to deploy and easy to use. In particular, we will initially grow Galapagos as an ad hoc grid on which users can benefit from each other's unused clock cycles (similar to how Skynet users benefit from each other's unused observing time) and to which the general public can contribute clock cycles through what we will call Science@home (similar to SETI@home and Folding@home). The Afterglow Modeling Project (AMP), described in F. is a large key project that will significantly exercise and test Galapagos.

See Appendix F.1: Motivation

See Appendix F.2: Genetic Algorithm

See Appendix F.3: Fitness Function

- See Appendix F.4: Data Structures
- **See Appendix F.5: Implementation**
- See Appendix F.6: Next Steps

See Appendix F.7: Broader Impacts: Broad Availability and Ease of Use

APPENDICES

B. Targeted Expansions of Skynet

B.1. Research Activities Overview

Funded by the American Recovery and Reinvestment Act through NSF, as well as by the Mt. Cuba Astronomical Foundation and NASA, we are pursuing three targeted expansions of the Skynet Robotic Telescope Network's wavelength, geographic, and user-community footprints. These efforts will not only significantly impact UNC-Chapel Hill's ability to study GRBs, they will significantly impact non-GRB research and research training that is being carried out by Skynet's broader user community, primarily on transient and time-variable phenomena. Furthermore, these efforts will significantly grow this community, both in opening Skynet to the Whole Earth Telescope's (WET's) and NRAO's user communities.

B.1.a. PROMPT-7: First, we are pursuing a targeted expansion of Skynet's wavelength footprint into the NIR with the addition of a 32-inch diameter, simultaneous optical/NIR telescope at the PROMPT site.

PROMPT's combination of rapid slewing (9°/sec) and simultaneous multi-wavelength imaging (ugriz or UBVRI) is unique among GRB follow-up telescopes. However, the most distant GRBs, as well as the most dust-obscured GRBs, are only detectable in the NIR. Originally, we planned to outfit one of the 16-inch diameter PROMPT telescopes with an LN2-cooled NIR camera, a Rockwell Scientific MicroCam capable of YJH imaging, which we purchased. However, we found that PROMPT's German equatorial mounts are too lightweight for reliable pointing and tracking when carrying this camera, given its dewar's changing weight distribution and heavy filling cables and its small, 5' field of view. Consequently, we have had to rely on other telescopes, such as SOAR, for NIR observations. However, such telescopes are human controlled and often take tens of minutes to interrupt, during which time GRB afterglows can fade considerably and are often no longer detectable in the NIR. The addition of a NIR-capable, robotic telescope to PROMPT's already rapid and simultaneous multi-optical wavelength imaging capability will allow us to detect and identify high-redshift and highly extinguished GRBs within tens of seconds of spacecraft notification and will significantly better inform us and, through our rapid release of GCN observing reports, the broader GRB community as to when to interrupt larger telescopes.

To help facilitate this and to significantly reduce the cost of this part of the project, UNC-Chapel Hill has partnered with telescope builder and owner of Astro Optik Phillip Keller and professional astrophotographer Johannes Schedler to build a 32-inch diameter, simultaneous optical/NIR telescope at the PROMPT site. In addition to our LN2-cooled NIR camera, this telescope will be outfitted with a wide-field optical camera, in part to better observe Fermi/LAT-localized GRBs, and a LuckyCam (also already owned by UNC-Chapel Hill), primarily for planetary research (e.g., ³⁸).

The division of observing time will be as follows: We will have priority whenever a GRB occurs. Keller and Schedler will have priority for the rest of the dark time. We will have

priority for the rest of the bright time, which is not a significant drawback in the NIR or for the bright objects that the LuckyCam will target. The non-GRB bright time, which is most of this time, will be available to the PROMPT Collaboration, to other Skynet telescope owners and their collaborators, and to WET during their large campaigns (about 1 - 2 weeks of bright time per year; see §B.1.b).

All three of PROMPT-7's instruments will add new capabilities to Skynet and will facilitate a diversity of research and research training efforts within Skynet's growing user community. A sampling of the research and research training that will be enabled across the collaboration includes:



Figure 5: UVRI Skynet/PROMPT observations of naked-eye GRB 080319B beginning 32 seconds after detection (15 seconds after notification) and best-fit three-component (internal, reverse, forward) shock model.³⁹ The chromatic inversion at early times occurs too quickly to be explained by the synchrotron peak frequency passing through these bands.

• In addition to allowing us to detect and identify high-redshift and highly extinguished GRBs within tens of seconds of spacecraft notification, PROMPT-7 will allow us to better probe GRB physics on these timescales, when internal and/or reverse shocks can dominate the emission (e.g., ⁴⁰). For example, on March 19th,

2008, PROMPT autonomously observed the most (isotropic-equivalent) luminous object in the universe, GRB 080319B, within 32 seconds of detection and 15 seconds of notification by Swift.^{41,42} Known as the "Naked-Eye" GRB, it reached 5th magnitude before fading away, despite being 7.5 billion light years away. PROMPT's unique rapid, simultaneous multi-wavelength design made possible the discovery of its chromatic inversion – the afterglow quickly changed in color from blue to red.³⁹ This occurred too quickly to be due to the synchrotron peak frequency passing through the optical bands, but could be due to a curvature effect associated with an internal shock, so far only seen in x rays, or due to a rapid fragmentation of dust in the circumburst medium (e.g., ⁴³⁻⁴⁶). Simultaneous NIR observations with PROMPT-7 should allow us to disentangle such effects in the future.

• The combination of NIR and optical observations is crucial in calibrating SNe as standard candles. Moreover, with such multi-wavelength data, it is possible to study the extinction law in the SN host galaxy. Differences between the extinction law in our own galaxy and the host galaxy of a given SN is one of the most important sources of systematic error in the use of SNe as a probe of the expansion history of the universe. This study will be undertaken by members of the CHilean Automatic Supernova sEarch (CHASE) project, the most successful SN search in

the southern hemisphere and prolific users of the current PROMPT telescopes. Their work will also benefit from the proposed expansion to SSO (see §B.1.b), allowing them near-continuous sky coverage to facilitate the discovery of ever younger SNe.

• Studies of high-mass x-ray binaries (HMXBs) benefit from NIR imaging, as they contain much dust, which can result in considerable variation in the IR. Identification of HMXBs with interesting IR variations will facilitate follow up with other IR instruments such as Spitzer. The proposed telescopes at SSO (see §B.1.b) will also help to eliminate large temporal gaps in the observations of southern hemisphere HMXBs. In addition, observations from the 20-meter radio telescope (see §B.1.c) are important because HMXBs are known to vary in the radio due to mass-transfer effects.

• An ongoing project studying possible phase lags between optical bandpasses on semi-regular variable (SRV) stars can be expanded into the NIR. This study requires coordinated simultaneous observations in both the NIR and optical bandpasses and is perfectly suited to the expanded PROMPT site. Such observations would enable color measurements for these poorly studied objects, and enable a review and improvement of published periods. Distances will then be derivable based on calibrations obtained from Hipparcos photometry and parallaxes for nearby SRVs, ultimately leading to possible improvements in the classification system for such objects. The addition of telescopes at the SSO site (see §B.1.b) would also benefit the study of these SRVs. While quoted periods of SRVs are never less than 20 to 40 days, there is some evidence of night to night variations in the current data. Near-continuous observations of selected objects would help to search for these shorter variations and could provide for important diagnostics for the nature of the pulsations producing the variability.

• NIR imaging will significantly enhance asteroid research, as the nature and texture of asteroidal surfaces and the resulting light scattering effects are best determined in the infrared. The simultaneous NIR and optical observations that will be possible at PROMPT will greatly enhance the accuracy of the surface parameters being measured and perfectly compliment an ongoing cooperative project which utilizes radar imaging of near-Earth objects from radio telescopes at Goldstone and Arecibo. The wider field of view on the optical camera on PROMPT-7, as well as its larger collecting area, will allow for more detailed study of asteroid "families" – groups of smaller asteroids formed by the collisional disruption of larger bodies. These families provide insights in the interior structures, strengths, and compositions of these primitive bodies by presenting samples that have the same age and dynamical history.

• The LuckyCam will be used to image lunar transient sites through mineralogy diagnostic filters from 0.6 to 0.9 microns. The aim of these observations is to correlate discrete gas release events detected by upcoming particle spectrometers in close lunar orbit with surface albedo/color changes around suspected vent sites.

B.1.b. Skynet at SSO: Next, we are pursuing targeted expansions of Skynet's geographic and user-community footprints with the addition of four 16-inch diameter telescopes at Siding Spring Observatory (SSO) in Australia.

Although telescopes at locations other than CTIO are now joining Skynet (§A.3), so far they are campus and privately owned telescopes in the US, Europe, and soon the Middle East. A PROMPT-like array of telescopes at the best site on the other side of the southern hemisphere will not only double PROMPT's scientific output (simultaneous multi-wavelength observations of twice as many GRBs on the rapid timescale, twice as many SNe discovered, etc.), it will add a significant new capability to Skynet: near-continuous, simultaneous multi-wavelength observing of southern sky objects. Not only will we be able to monitor bright GRB afterglows without significant interruption as Earth rotates, and then with larger PROMPT-7 and SOAR as they grow fainter, near-continuous and near-continuous, simultaneous multi-wavelength coverage enables a great deal of science for variable and eclipsing binary star, pulsating white dwarf, and blazar observers, many of whom are already using PROMPT (e.g., ^{11-15,18-20}; also, see §B.1.a).



Figure 6: Fourier transform of a single pulsation as observed from CTIO (left) and CTIO + SSO (right).

To more fully exploit this new capability, we have partnered with the Whole Earth Telescope (WET), a worldwide collaboration of

astronomers that organizes global observing campaigns targeting compact pulsating stars in order to obtain near-continuous coverage. Figure 6 illustrates how near-continuous coverage will remove large alias peaks in the Fourier transform which would otherwise make it nearly impossible to properly identify the correct frequency of pulsation, particularly in the presence of noise or closely spaced modes.

The WET community will make full use of the unique advantages of the Skynet telescopes: near-continuous coverage, year-round accessibility, and simultaneous multi-wavelength observations. In order to obtain continuous coverage, WET currently employs a large number of telescopes (>8) for a 2 - 3 week period, each manned by a local astronomer. (PROMPT has already participated in two WET campaigns: November 2008 and May 2009.) The overhead required to coordinate such campaigns is large, limiting WET to two such campaigns per year. With year-round availability of robotic telescopes at both CTIO and SSO, WET will be able to increase its scientific impact with minimal increase in manpower. Furthermore, future WET campaigns can easily improve their continuous coverage by employing all of the telescopes on the Skynet system, at guest or higher priority. Finally, the near-continuous, simultaneous multi-wavelength coverage which PROMPT + Skynet-SSO offer will enable unique observations of pulsators that are otherwise prohibitively difficult to obtain, coordinate, and calibrate.¹

The division of observing time will be as follows: UNC-Chapel Hill will have priority whenever a GRB occurs. Once MPSC completes its live observing center (see §B.3.b.2), it will have next

¹ A possible future expansion of this effort would be the addition of a telescope or an array of telescopes at the South African Astronomical Observatory (SAAO) where UNC-Chapel Hill is a partner in the 9.2-meter effective diameter Southern African Large Telescope (SALT). This would make possible completely continuous observing from three nearly identical southern latitudes spaced relatively uniformly in longitude around Earth.

priority on three of the telescopes and WET will have next priority on one of the telescopes and on one of the 16-inch diameter PROMPT telescopes at CTIO for near-continuous observing, except when near-continuous, simultaneous multi-wavelength observing is required, in which case WET will have next priority on additional telescopes at both sites. WET targets are selected from community proposals through a TAC process and will be input into Skynet. We will make the remaining time available to the PROMPT Collaboration, and to other Skynet telescope owners and their collaborators.

A sampling of the research and research training that will be enabled across the collaboration includes:

• PROMPT + Skynet-SSO will be an integral part in the discovery and follow-up of new white dwarf (WD) and subdwarf B (sdB) pulsators. PROMPT has already been used to confirm the variability of two new WD pulsators and to discover one of the largest amplitude sdB pulsators, CS 1246.¹³ We intend to observe more than 200 sdB stars using Skynet. As approximately 10% of the stars that fall in the instability strip are variables, we expect to find a considerable number of new pulsators from this survey.



Figure 7: Folded u'g'r' light curves of sdB CS 1246 from Skynet/PROMPT.¹³ The different pulsation amplitudes constrain l = 1 or 0.

• A key feature of PROMPT + Skynet-SSO is the ability to do simultaneous multi-wavelength observations. This is one of the only ways to directly measure

the geometry of a pulsation mode. The non-radial pulsations on compact pulsators appear as spherical harmonics (denoted with the letters l and m). Multi-wavelength observations allow us to distinguish between these different pulsational geometries. Once we know the values of these spherical harmonics we can use the pulsations to measure the internal properties of the star, including mass, atmospheric depth, and core composition. The PROMPT telescopes have already shown their ability to identify spherical harmonic indices of pulsational modes by constraining the pulsation geometry of modes on sdB CS 1246 (see Figure 7).^{18,20}

• With dedicated telescope use at CTIO and SSO, we will be able to measure month-to-month variations of pulsation frequencies and amplitudes on WDs, a parameter space rarely explored. This mode of operation, while precluded by the relative brevity of previous WET runs, will allow us to explore the underlying physics of compact stars. By mapping the slow changes in mode frequencies and amplitudes over several months we will be able to explore questions of energy transfer between pulsations and the variability of weak magnetic fields on WDs.⁴⁷ Additionally, the nonlinear light curves of WDs can be used to measure the turbulent energy transport (convection) at specific temperatures and pressures.⁴⁸ Convection is very poorly understood and by participating in full WET runs, the Skynet telescopes will help obtain measurements to constrain new models of convection physics.

• Pulsating WDs are accurate clocks with a slow period drift due to the cooling of the star: $dP/P \sim 10^{-15} \text{ sec/sec.}^{49}$ One possible contributing factor to the cooling rate is the production of axions in the core of WDs.⁵⁰ The axions, though hypothetical, are a possible dark matter candidate. By measuring the periods of the bright southern hemisphere WD pulsators with PROMPT + Skynet-SSO over the next 5 – 10 years, we will provide stringent constraints on the mass of the hypothetical axion.

• The Skynet telescopes will also participate in the search for planets around evolved stars. As the stable WD and sdB pulsators orbit the center of mass of the star-planet system, the pulsations will periodically arrive earlier and later.^{51,52} This method is especially sensitive to more distant planets, a parameter space not explored by the radial velocity or transiting surveys. The large temporal coverage afforded by PROMPT + Skynet-SSO will maximize the chances for seeing these variations.

• The expanded temporal coverage enhances observational capabilities for a number of variable star projects. Long-period eclipsing binaries can have eclipses lasting many hours, sometimes even longer than a single site can observe. Obtaining a time of minimum can be extremely difficult, and as such they are virtually unstudied for period changes, eccentric orbits, apsidal motion, and possible additional components. The increased temporal coverage of additional telescopes at SSO will enable such studies. Additionally, objects such as RR Lyrae stars, which have pulsation periods of roughly half a day to a day, are difficult to study for second order pulsations as well as variations in the primary period. The ability to observe these types of variables almost uninterrupted through complete pulsation cycles is important for measuring these variations.

• Blazars are well known to vary substantially over periods of months, years, and even decades. Shorter variations, over periods of days, hours, and even minutes, are well established, but less well studied. While these shorter variations are seen and studied in high-energy x rays and gamma rays, they are virtually unstudied in the optical and NIR. SSO, along with PROMPT (and its new NIR capability, see §B.1.a) is well suited to studying a selection of blazars for such variations.

B.1.c. Radio Skynet at NRAO-Green Bank: Finally, we are pursuing targeted expansions of Skynet's wavelength and user-community footprints into the radio with the refurbishment and addition of NRAO-Green Bank's 20-meter diameter telescope.

The 20-meter will be used as a development platform for a radio version of Skynet, with the goals of (1) providing radio astronomy research and research training opportunities to a larger community, (2) expanding the reach of NRAO-Green Bank's excellent EPO programs from regional to national and international scales, and (3) enabling collaborative research and development in radio astronomy instrumentation.

The 20-meter telescope, originally constructed for the US Naval Observatory, was part of the National Earth Orientation Service network from 1995-2000 and participated in a global program of Earth Orientation VLBI measurements in cooperation with the International Earth Rotation Service. Although the 20-meter is of modest size, it has a precision surface (0.8 mm RMS) and

extremely fast slew rates (2° /sec, each axis). Science that can best be done with the 20-meter takes advantage of these attributes.

A sampling of the research and research training that will be enabled across the collaboration includes:

Investigating the transient sky with the 20-meter

• Because of the telescope's fast slew rates and large beam size, one can map the sky visible to Green Bank in 8 hrs at 21 cm and 24 hrs at 3 cm to the confusion limit. By making multiple maps with varying time spacing, one can explore source variability with time scales (8 hrs) that are shorter than usually explored and up to time scales as long as months.^{53,54} The high time resolution of these maps allows one to capture possible transient events of modest (few Jy) intensity and at distances with interesting extragalactic (>~200) dispersion measures which can then be followed up by directed observations with the 20-meter or other telescopes.⁵⁵

• The recent discovery of a bright, 5-ms burst of radio waves of cosmological origin suggests that we have been missing some important and unknown astrophysical phenomena. The energy release from such a source at 1 Gpc (3.3 billion light years) would be $\sim 4 \times 10^{33}$ J, or the total energy released by the Sun in four months!⁵⁶ Light-travel-time arguments strongly suggest an origin on or around compact objects. Possible mechanisms are coalescing neutron stars, magnetar hyperflares, or cosmic strings. It is impossible to discriminate between scenarios from the properties of only one event. However, a key implication is that hundreds of similar bursts should be detectable each day, but the small field of view and limited temporal and spectral resolution of most radio telescopes has so far rendered them largely invisible. The 20-meter is well suited for performing a whole-sky survey for similar transients. As long as the sky is sampled sufficiently quickly $(50 - 80 \,\mu\text{s})$ these observations can piggyback on any other science being conducted at 20 cm. West Virginia University has is supplying the necessary second backend (IBOB + Xilinx FPGAs) for this project and other tandem observing projects that might occur. As was the case for GRBs, it is likely that these radio bursts will have detectable signals elsewhere in the electromagnetic or gravitational wave spectra. Any detected burst information will be placed rapidly into the public domain, creating opportunities for detection at other wavebands. This will undoubtedly lead to accelerating new research opportunities in understanding the sources.

• 20-meter observations of easily-detected giant radio pulses from the Crab pulsar could be used to correlate giant pulses with gamma rays detected by Fermi. This work would help to constrain the giant-pulse emission mechanism and probe the physics of relativistic particle acceleration in high magnetic fields.

Probing the cosmic web: HI science with the 20-meter

• An exciting area of radio-optical synergy for the 20-meter will be UNC-Chapel Hill's RESOLVE Survey.⁵⁷ RESOLVE will survey 53,000 cubic Mpc of the z = 0 cosmic web, constructing the first ever volume-limited mass census of the gas, stars, and dark matter for all galaxies in a cosmologically significant volume, down to a dwarf mass scale and spanning multiple clusters, filaments, and voids. The data include existing public UV/optical/IR surveys, new SOAR/SALT high-resolution spectroscopy to measure dynamical masses, and new Arecibo

21-cm data to measure HI gas masses from the ALFALFA survey. The 20-meter will facilitate true completeness for RESOLVE, with deep 21-cm follow up of galaxies too faint for ALFALFA's blind and necessarily shallow wide-area drift scan. Going deep on this ~20% of RESOLVE galaxies will enable an unprecedented cosmic mass census, answering basic questions concerning the nature and substructure of dark matter, the location of "missing baryons" predicted by current cosmology, and how gas in the cosmic web reshapes galaxies destroyed by mergers to recreate galaxies like our own.

• Recent research has revealed large, low surface-brightness HI clouds around galaxies and galaxy groups, possibly associated with the warm-hot intergalactic medium (e.g., ^{58,59}). Their large angular size makes them excellent targets for a dedicated 20-meter telescope capable of making integrations of many tens of hours. The 20-meter can explore the global trends in a large sample of nearby groups and clusters by measuring their integrated HI masses above 10¹⁰ solar masses. Detections can then be confirmed using the GBT.

Other science with the 20-meter

• A major goal of 21st century astrophysics is identifying the nature of dark matter, thought to be an as-yet unidentified, weakly interacting massive particle (WIMP).⁶⁰ These particles can decay resulting in both gamma-ray emission – being searched for with Fermi, although with very low expected count rates – and radio synchrotron emission from electron-positron pairs.⁶¹ Searches for these decay products are underway with the GBT at 1.4 GHz. The 20-meter telescope will provide a valuable high-frequency channel to complement the GBT data, approximately matched in angular resolution, and allowing spectral discrimination between dark matter decay signals and (steeper-spectrum) foreground Galactic synchrotron.

• The bandwidth of the 21-cm receiver covers the frequencies emitted by the hydroxyl radical as well as HI. Astronomers may want to use the 20-meter to map the extent of OH production from comets, or monitor the motion and evolution of OH-rich envelopes of evolved, preplanetary nebulae stars. The 3-cm receiver could be used to map the galactic plane for radio recombination lines from HII regions that cannot be executed economically with other telescopes. Such an experiment might have sufficient sensitivity to inventory all of the highmass star formation in our galaxy and provide an entrée to radio science for optical astronomers involved in using the Skynet telescopes.

Research and development in radio astronomy instrumentation with the 20-meter

• In 2007, the 20-meter was briefly re-activated as part of a collaboration with engineering faculty and students at Brigham Young University to conduct a series of array feed experiments. То date. 15 papers have resulted from their experiments (see http://www.ece.byu.edu/faculty/bjeffs/publications.phtml). As a test platform for other instrumentation collaborations, the 20-meter could be invaluable. We view this refurbishment as an opportunity for university groups to partner with NRAO in developing new technologies for radio astronomy, such as focal plane arrays, and RFI cancellation techniques.

The division of observing time will be as follows: NRAO will have priority about 70% of the time. NRAO projects will be selected from community proposals through a TAC process and

will be input into Skynet. We will make the remaining time available to the PROMPT Collaboration, and to other Skynet telescope owners and their collaborators.



B.2. Description of Research Instrumentation

Figure 8: PROMPT-7 and enclosure/dome design.

B.2.a. PROMPT-7: The hardware design of PROMPT-7 is driven primarily by (1) the need to host multiple simultaneous instruments for rapid, unattended switching between observing programs, (2) its suitability for hosting our existing NIR camera; and (3) its ability to aid in the localization of GRB detections with large error circles in position, such as those reported by Fermi/LAT. PROMPT-7 will consist of a 32-inch diameter, f/7 Astro Optik optical system with LOMO Sital mirrors and a 100-mm diffraction-limited corrected field at Cassegrain focus. The large corrected field permits simultaneous mounting of multiple instruments, which will include (1) a Kodak KAF-16803based optical camera provided by Keller

and Schedler (see §B.1.a), (2) an Andor Technology Luca EMCCD "Lucky" camera already owned by UNC-Chapel Hill, and (3) an LN2-cooled Rockwell Scientific MicroCam NIR system, also already owned by UNC-Chapel Hill.

The optical camera will have a $4,096 \times 4,096$ pixel imaging array with a 22.2' FOV (0.32" pixels), 60% peak QE at 550 nm, and 10e⁻ RMS readout noise. Combined with the increased light gathering power of the 32-inch mirror, PROMPT-7 will not only operate as a capable general-purpose optical scope, but will also be able to quickly scan the sky around a Fermi/LAT GRB detection and localize the optical afterglow. Once localized, all of the PROMPT telescopes can pursue simultaneous multi-wavelength observations of the target field, and participants in the GRB Coordinate Network may make use of the localized coordinates to perform their own follow-up observations. Under ideal conditions, a 1-minute exposure is expected to have a limiting magnitude of V \approx 20 (S/N = 3).

The Luca camera uses a 658×496 pixel imaging array with a peak QE of 52% at 500 nm. The system will have a field of view of 4' × 3' (0.36" pixels). Full-frame images can be read out at up to 37.2 frames per second with 15e⁻ RMS readout noise. Subframe exposures can be transferred at proportionally higher rates, and readout noise can be reduced below 1e⁻ by slowing the readout speed. This camera enables the use of lucky imaging techniques, which can drastically reduce the effect of atmospheric seeing. The fast readout rate, combined with longer exposure lengths,

will also be useful for observations of rapidly varying phenomena (<2 minute period), where a duty cycle near 100% is highly beneficial.

The NIR camera features a 256×256 pixel imaging array and will have a field of view of 5' (1.2" pixels). This configuration balances the competing requirements of acquiring comparison stars within the field of view without under-sampling individual stars. The camera will be LN2 cooled with an autofill system, and includes Y, J, and H-band filters. A 1-minute exposure has an expected limiting magnitude of J \approx 16.5 (S/N = 3).

The optical system will be installed on a rapidly slewing (10°/sec) fork mount with 5" RMS pointing accuracy and 0.35" RMS/10 min unguided tracking accuracy. The torque characteristics of the mount compensate for the changing LN2 weight distribution of the NIR camera, and the fork design simplifies the problem of running cooling hoses to the equipment. Moreover, the pointing and tracking performance aids in observing with the relatively small FOV imposed by the NIR and Luca camera hardware.

The hardware for PROMPT-7 will be enclosed in a 5-m rotating Ash Dome that has already been supplied by Keller and Schedler, and a lower observatory structure that has already been built by UNC-Chapel Hill. As delivered, the dome rotates 3.2°/sec (56 second worst-case repositioning delay). In order to improve GRB response time, we calculate that the maximum rotation speed can safely be doubled by adding smooth acceleration/deceleration to the rotation mechanism. The dome will be automated with computer control and weather/safety interlocks by UNC-Chapel Hill.

B.2.b. Skynet at SSO: The design criteria of Skynet-SSO are to (1) provide rapid, simultaneous multi-wavelength observations of transient events, such as GRBs; (2) provide near-continuous observing capabilities in conjunction with PROMPT; and (3) allow general-purpose observations to be performed during the day from the U.S.



Figure 9: The Skynet-SSO telescopes will be nearly identical to the 16-inch diameter PROMPT telescopes.

Each of the four Skynet-SSO systems will consist of a 16-inch f/9 Ritchey-Chrétien telescope, mounted on a fast slewing (9°/sec) Paramount ME mount and enclosed in a 12-ft Astro Haven clamshell dome. Each telescope will be equipped with an Apogee Alta 1,024 × 1,024 pixel CCD camera with a 12.5' FOV (0.74" pixels), a peak QE of 92% at 550 nm, 10e⁻ RMS readout noise, and a 1-second full-frame readout delay. An Apogee filter wheel with BVRIg'r'i'z'C filters allows a single telescope to perform multi-color observations, and also permits multiple available telescopes to be combined for simultaneous color observations. The expected limiting magnitude of a 1-minute exposure is approximately V \approx 19 (S/N = 3).

One telescope will simultaneously host an Indigo Systems Corporation Alpha NIR camera, which UNC-Chapel-Hill already owns. The camera is thermoelectrically cooled, which eliminates the weight distribution problems associated with liquid cooling and permits the camera to be installed on a relatively small German equatorial mount. The camera uses a 320×256 pixel array and will observe unfiltered across YJH bands. The expected limiting magnitude of a 1-minute exposure is J \approx 14 (S/N = 3). Although the sensitivity of this thermoelectrically-cooled system is not as high as the LN2-cooled camera on PROMPT-7, this remains a cost effective method for identifying relatively bright high-redshift/highly extinguished GRBs that are not observable from PROMPT on the rapid timescale, and will likely have other non-GRB science applications as well (e.g., §B.1.a).

Normal use of Skynet-SSO and PROMPT-7 will be coordinated through the Skynet software written by UNC-Chapel Hill. Nearly all Skynet-SSO hardware is already compatible with our existing software system. Major components of the control software for the Alpha and Rockwell Scientific NIR cameras have been written, with minor additional work being required for full integration into Skynet. Some minimal software work will also be required to integrate the mount for PROMPT-7.

B.2.c. Radio Skynet at NRAO-Green Bank

L-band receiver: The L-band receiver, originally used on the 140-foot telescope, will be modified for operation on the 20-meter. This receiver is still quite impressive, with a bandwidth spanning 1.3 - 1.8 GHz, and a receiver temperature of 8 K (total system temperature of 18 K). To refurbish this receiver, the receiver box will be outfitted with a water tight panel that contains a fiber optic cable interface. This cable contains the single mode fiber for analog transmission of the received signal to the control room, and multimode fiber for transmission of the digital monitor and control signals. Two analog fiber optic modulators will be installed for signal transmission. A digital modem is also required for remote monitoring and control of the receiver. Input and output electronics will be designed and installed to interface the existing monitoring electronics to the digital buss.

X-band receiver: Refurbishing the S/X band receiver system originally designed for VLBI observations will improve the performance for other science drivers. The original bandwidth is limited to 9% as a consequence of the dual S/X band design. We will eliminate S-band and build a new X-band feed for this system, increasing the bandwidth to 30%. The new feed will also reduce the length of circular wave guide coupling the cryogenic dewar to the feed, thus lowering the system noise temperature by 10 K. Mechanical modifications to the receiver box are also necessary to accommodate the changes. Wideband filters of the appropriate center frequency will be purchased and installed. The left and right polarizations will be modulated at the X band frequencies onto analog fiber optic links. At the output of the fiber optic links, down conversion to the analog-to-digital converter sampling frequency is necessary. A tunable down converter will be constructed of two mixers, a band-pass filter for image rejection, a tunable LO, and a fixed LO. This scheme is identical to the GBT IF system for rejecting the unwanted sideband and providing band selection.

Cryogenics: Since both receivers are cooled to 15 K, the telescope requires a helium compressor and stainless steel pressure lines. Recent improvements in the NRAO compressor design along with the purchase of a TIG welder for installing SS lines warrant an upgrade of the 20-meter cryogenic system. This will reduce cryogenic failures, eliminate potential leaks in the lines, and improve the reliability of the system.

Backend: We are replicating instrumentation designed for pulsar signal processing and spectral analysis for the 20-meter. The backend will be based on hardware and development software from the Center for Astronomical Signal Processing and Electronics Research (CASPER) which uses reconfigurable off-the-shelf hardware platforms and software tools to rapidly design and deploy astronomical signal processing systems. One such system, called the "Green Bank Ultimate Pulsar Processing Instrument", or GUPPI, was designed using Field Programmable Gate Arrays (FPGAs) and was in shared-risk use within a year after construction began. For the 20-meter, off-the-shelf Reconfigurable Open Architecture Computing Hardware (ROACH) FPGA boards and iADC analog to digital converters will be combined with a multi-core computer system to form the backend. The data will be stored on an array of disks.

Radio Skynet: The development of Radio Skynet will require (1) modifying Skynet and the Skynet database to accept radio telescopes and radio observing types; (2) the development of a radio telescope version of our Terminator software for local hardware control; (3) the development of a variety of web interfaces, introductory through advanced, for requesting observations; and (4) the development of a radio version of our reduction pipeline, which will be done in consultation with NRAO.

B.3. Impact on Research and Training Infrastructure

The proposed, targeted expansions of Skynet's wavelength and geographic footprints will not only significantly impact the research, research training, and EPO that is already being carried out by Skynet's existing user community (§A.2, §A.3), it will significantly expand this community by offering unique research, research training, and EPO opportunities to members of WET's and NRAO's user communities (§B.1.b, §B.1.c).

B.3.a. Impact on User Community Infrastructure: PROMPT's user community consists of UNC-Chapel Hill, 12 regional undergraduate institutions, including three minority-serving institutions, MPSC, and, through TAC processes, the US and Chilean optical astronomy communities (§A.2). Skynet's existing user community consists of the PROMPT Collaboration, the growing number of institutions that and individuals who are putting non-PROMPT telescopes on Skynet (§A.3), and their collaborators. Aside from UNC-Chapel Hill, most of Skynet's research and research training users are from smaller institutions without comparable facilities – some do not even have campus telescopes. However, leveraged by PROMPT, our users have forged partnerships with users of other optical and radio telescopes around the world, and also of space telescopes. ^{11,12,14-17,19,21-27}

Given Skynet's growing geographic footprint (§A.3), WET is a very natural community to partner with, particularly given the growing demand within Skynet's existing user community for near-continuous and near-continuous, multi-wavelength coverage (§B.1.b). The partnership

with NRAO is an exciting opportunity to bridge the two ground-based observing communities and offer the US radio astronomy community a parallel facility to Skynet.

Key to the success of Skynet has been and will continue to be the easy-to-use web interfaces that we have developed and continually improve upon in response to the needs of our growing user community (see §C), and the nation's (and world's) cyber-infrastructure, through which the entire system functions. Users no longer need to raise money to travel to telescopes. They no longer even need to stay awake at night. Some no longer need to write proposals to win time. It is not an overstatement to say that Skynet has created a small telescope and small telescope science renaissance in North Carolina and southern Virginia, with seven of the region's telescopes now on or gearing up to go on Skynet (three of which were built for this purpose, including one at HBCU NC A & T), with a renewal of research and research training across the PROMPT Collaboration's undergraduate institutions, and with approximately 20,000 high-school, middle-school, and elementary-school students and members of the general public across North Carolina having used Skynet through MPSC programs (§A.2, §A.3).

B.3.b. Impact on Students



Figure 10: Mosaic of images of near-Earth asteroid 2001 FE90 simultaneously obtained from PROMPT and Dark Sky Observatory in North Carolina. A parallactic shift of about 8' and a rotational period of about 30 minutes can be measured from the images.

B.3.b.1. Undergraduate Students: In addition to facilitating research and research training, Skynet has been working its way into undergraduate curricula across North Carolina and southern Virginia. At UNC-Chapel Hill, both our introductory and advanced astronomy laboratory

courses are now Skynet based, and our now two-semester introductory astronomy lecture course will make regular use of the introductory version of Skynet's interface (§A.2), which we are expanding to include the moon, planets, and dwarf planets. In the introductory laboratory course, students measure the masses of all four Jovian planets using Newton's form of Kepler's Third Law, parallaxes and distances to planets and minor planets using simultaneous imaging from Skynet telescopes in different hemispheres (see Figure 10), etc. (see Teaching Statement). Radio astronomy labs will be developed upon the completion of Radio Skynet and Radio Afterglow (see §B.2.c). Altogether, we are reaching approximately 1,500 students per year on campus and approximately 100 distance-learning students per year online, including members of the US military stationed in Iraq and Afghanistan.

B.3.b.2. MPSC EPO: MPSC is about to be renovated and expanded to make room for what will be one of the best mid-sized science centers in the country. Designed by the same company that recently redesigned Griffith Observatory in LA, the exhibit space will feature a live observing center, from which visitors will be able observe using Skynet-SSO and a small radio

telescope in real time. The interfaces will be based on our introductory Skynet interface (§B.1.b). A similar interface will be set up at NRAO-Green Bank's science center.



Figure 11: Artist conception of MPSC's live observing center.

B.3.b.3. NRAO EPO: NRAO has many educational programs that reach out to broader communities: professional, educational, and lay. Some provide research experiences and training for K-16 teachers and students, while others provide hands-on informal experiences for families and the general public. But. these programs are primarily delivered in the communities and regions co-located with NRAO's

major telescopes. Radio Skynet presents a singular opportunity to extend these efforts beyond the boundaries of their research facilities. NRAO is committed to broadening participation in radio astronomy; to increase participation of underrepresented groups and diverse institutions throughout the US and beyond. Radio Skynet will be an effective way to provide authentic experiences in radio astronomy research activities to a new, national audience. In addition, lessons learned in piloting Radio Skynet with the 20-meter will assist NRAO in developing the outreach program for the ALMA telescope currently under construction.

B.3.b.4. Afterglow: UNC-Chapel Hill is developing web-based, image reduction and analysis software that is easy enough for middle school students to use, but identical to IRAF in its outputs (see Teaching Statement). "Afterglow", which is directly linked to Skynet's database, is already being used in UNC-Chapel Hill's introductory astronomy laboratory course (§B.3.b.a). **Try a beta version:** <u>http://skynet.unc.edu/afterglow</u>, login = "guest", password = "reichart". A radio version, which will build upon my 20 years of experience developing radio astronomy reduction and analysis software for NRAO-Green Bank's 40-foot diameter telescope (see Teaching Statement) is being planned.

C. A Broader Discussion of Skynet

Astronomers have witnessed tremendous progress in the development of robotic telescopes over the past fifteen years. Computer-controlled telescope mounts are able to run night-long observing programs while completely unattended, using sophisticated modeling software originally developed for large professional observatories to achieve remarkable pointing and tracking performance. CCD technology has provided astronomers with efficient, high-quality imaging devices that allow a small telescope to do work that a generation ago would have required a system many times larger. Auxiliary equipment commonly used in optical observatories, including filter wheels, focusing mechanisms, and enclosures, can be easily automated. Perhaps most astonishingly, all of this equipment has become relatively inexpensive and straightforward to operate. Over the past several years, high-quality robotic telescopes have become well within the reach of many universities, small colleges, and dedicated amateur astronomers. With the rapidly growing number of robotic telescopes around the world, and with the advent of widespread broadband Internet access, the prospect of linking many robotic telescopes into a network became increasingly attractive. Development of the Skynet Robotic Telescope Network has effectively continued the trend of expanding availability of high quality telescope equipment to new and formerly underserved audiences. Skynet has been used by professional astronomers, graduate, undergraduate, and high school students, and members of the general public to take nearly four million exposures of the night sky.

Many features of telescope networks in general, and Skynet in particular, are attractive to telescope owners and observers. Perhaps most obvious is that a worldwide telescope network is not constrained to a particular site and telescope. Skynet users have access to a variety of telescope equipment that covers both the northern and southern sky and includes systems at premier observing sites like CTIO. For an observer carrying out time-critical observations, obstacles such as equipment outages and bad weather at the primary observing site can be mitigated. Furthermore, the longitudinal range of Skynet allows uninterrupted observations for extended periods of time. Our recent expansion to a telescope in Italy and upcoming expansion to Australia dramatically improve our total sky coverage and our ability to support continuous observations.

By joining Skynet, telescope owners immediately benefit from the infrastructure we have developed. Our website provides a convenient multi-user interface to the telescope, which effectively solves the problem of giving many students or external collaborators access to a particular telescope. With our queue scheduling system and automated control software, observers and telescope operators do not need to stay awake to manually monitor and control telescope operation. Nearly all night-to-night observing tasks, such as opening and closing the observatory dome, monitoring for bad weather conditions, and reinitializing telescope hardware in case of common problems, can be performed autonomously.

All Skynet images are permanently archived on a disk array at UNC-Chapel Hill, which is currently capable of storing 25 terabytes of data. Skynet members are not faced with the need to provide storage, server equipment, and network bandwidth to distribute telescope data to their users.

Finally, observers on any Skynet telescope benefit from the data analysis pipeline and tools that we have developed. This pipeline, which is now the same system we use to rapidly analyze GRB data and produce publication-quality results, automatically handles the tasks of generating and applying master CCD calibration images, detecting stars, finding astrometric solutions, and generating well calibrated photometry. With further development, the pipeline will evaluate the quality of incoming images in real time according to several metrics, allowing telescope owners and observers to keep a close watch on telescope performance.

In our experience, a great deal of telescope time goes unused at robotic observatories around the world, not because of poor equipment or sky conditions, but because the primary users of the telescope are often, by themselves, simply unable to generate sufficient demand to keep the telescope observing every clear night. For example, a campus telescope used mainly for lab

exercises may only see heavy use for a few weeks each semester. Moreover, many sites lack the resources and user community to effectively manage and share their telescope time. Finally, sites may not have the manpower, tools, or specialized experience to manually monitor a telescope and keep it in top operating condition night after night.

Nevertheless, many of these telescopes could, with minimal effort, join the Skynet Robotic Telescope Network. By integrating with Skynet, telescope owners benefit from our automation tools, community management, and small telescope experience. Groups are able to interact with their own telescopes more efficiently, and any unused time on a particular telescope can be easily shared with the wider astronomical community, making each integrated telescope more productive and useful.

In order for Skynet to expand effectively, we must be able to integrate new hardware quickly and keep telescopes running smoothly once integrated. We are making it possible to manage a large telescope network with a relatively small staff by: (1) modularizing our control software to make integration of new observatory hardware a straightforward task; (2) developing tools to make necessary telescope evaluation and maintenance as automated and efficient as possible; (3) implementing sophisticated and thorough methods for detecting and notifying telescope owners of problems that affect imaging quality or telescope availability; (4) continuing to develop our image analysis pipeline to allow us and our user community to work with an ever-expanding quantity of data; and (5) providing telescope owners with thorough documentation and other resources for anticipating, preventing, and recovering from common problems and failure modes:

Hardware Integration: We have written our control software to operate with a variety of common interfaces to astronomical hardware. We currently support telescope mounts, filter wheels, and cameras that can communicate with standard software such as TheSky6, MaxIm DL, and the ASCOM platform. This represents the vast majority of basic telescope equipment available in the small telescope market today. However, other types of hardware, such as weather stations, cloud sensors, and enclosures, frequently lack robust driver support or do not have well-defined standard interfaces. In many cases, observatories will have custom-designed equipment with an entirely unique software interface (e.g., PROMPT-7's NIR camera; see §B.2.a). To facilitate quick integration of this type of hardware, we will create frameworks and interfaces to handles the common tasks involved in communicating with such equipment. Our Skynet software will be restructured with a more modular design and incorporate this framework. Support for new devices can then be quickly implemented and tested with a relatively small amount of programming effort. Furthermore, the insufficiently robust control software that is delivered with some equipment will be updated and reprogrammed as necessary to support improved error detection and problem recovery capabilities.

Maintenance Tools: When a telescope is integrated with Skynet, a fair amount of time is spent evaluating the quality and performance of the hardware. For example, CCD cameras are run through a set of commissioning tests to evaluate fundamental characteristics of the camera, and telescope mounts undergo a suite of tests we have developed to detect potential operational problems or inconsistent corner case behaviors. In addition, once a telescope has been integrated with Skynet, there are several maintenance tasks that must be performed regularly. For example,

the pointing and tracking model for the mount may become ineffective over time and need to be rebuilt. We are developing tools to allow Skynet staff and telescope owners to efficiently perform telescope integration testing and maintenance to achieve consistent, high-quality performance and results across the network with minimal manual effort.

Detecting Errors: As previously mentioned, many sites lack the resources to manually monitor telescope performance every night. Moreover, as we add systems to Skynet, it becomes impractical for our staff to constantly track the operation of every telescope. It is therefore critical to detect errors and performance degradation in a fast, reliable, and fully automated fashion. Such error detection is especially important where the safety of the telescope is concerned; in particular, staff must be notified immediately in the case of dome failure. We put great emphasis on detecting equipment failures (e.g., the computer is unable to communicate with the mount), unexpected behaviors (e.g., a 5 second exposure only took 3 seconds to complete), and configuration errors (e.g., the system clock is incorrect). The appropriate personnel are notified of such errors via e-mail and cell phone alerts. We will continue to add error detection support for new failure modes as we discover them. We will also extend these capabilities to include real-time, automated evaluation of telescope performance based on recent images. We will be able to detect problems such as pointing and tracking errors, out-of-focus stars, excessive wind shake, stuck camera shutters, and more. Curious image artifacts such as streaks caused by orbiting satellites can also be flagged. In cases where immediate action must be taken to continue producing useful observations, staff will be notified automatically. Other metrics, such as star roundness, telescope drift, and small pointing errors, will be summarized in daily web reports that allow administrators to easily track telescope performance over time and identify potential areas for improvement.

Pipeline Development: We have developed a software pipeline to support quick analysis of Skynet images. The pipeline uses a C++ library we have built on top of cfitsio that implements and improves upon stacking, calibration, and photometry routines from IRAF. Our library also supports automated WCS solutions for arbitrary rotation angles and plate scales, source detection using the SExtractor package, catalog lookups for the major star surveys, and custom algorithms such as "smart stacking" that automatically prevent poor quality single exposures (e.g., those observed when a thin cloud passed overhead) from lowering the SNR of a final stack. This library is callable from the Python programming language, facilitating quick development of analysis scripts. We are also developing a web application called Afterglow, which provides Skynet users an easy-to-use interface for analyzing images from any computer using our pipeline. Approximately 1100 introductory lab students at UNC-Chapel Hill have used a beta version of Afterglow since 2008 (see §B.3.b.4).

We will continue to develop this pipeline to support the automated image quality metrics and error detection mentioned above. We will also continue to add analysis capabilities as they are needed by us and our user community. Finally, we will continue to develop Afterglow as a tool for making image analysis accessible for beginning users, and for giving professional users a new way to quickly analyze Skynet data and produce publication-quality results.

Documentation: Except for infrequent emergency support and our annual two-week maintenance trip, PROMPT is an entirely unmanned facility. Through our work to make

PROMPT and other Skynet telescopes as reliable and autonomous as possible, we have identified many of the most common problems and design considerations associated with running a robotic telescope. We have compiled a list of approximately 150 failure modes related to third party software bugs, hardware errors, configuration errors, and simple wear and tear over time. Many of these problems are preventable, and others have straightforward solutions. We will continue to thoroughly document and share our findings so that this hard-earned knowledge can benefit all Skynet telescope operators and keep troubleshooting and repair time to a minimum. In addition to written documentation, we plan to produce a set of online videos that show telescope owners how to perform a variety of the most important hardware and software tasks related to operating a telescope.

D. Foundation for Growing Skynet

An overarching goal for Skynet development over the past year and a half has been to expand the telescope network to new observatories while preserving a level of maintenance that is sustainable with a relatively small technical staff. In §C, we outlined five primary areas of development that would help us to achieve these goals: (1) modularizing our control software to make integration of new observatory hardware a straightforward task; (2) developing tools to make telescope evaluation and maintenance as automated and efficient as possible; (3) implementing sophisticated and thorough methods for detecting and notifying telescope owners of problems that affect imaging quality or telescope availability; (4) continuing to develop our image analysis pipeline to allow us and our user community to work with an ever-expanding quantity of data, and (5) providing telescope owners with thorough documentation and other resources for anticipating, preventing, and recovering from common problems and failure modes. Our progress in these areas is summarized below.

Hardware integration: Skynet was originally written primarily to support the hardware used on the original PROMPT telescopes, and took advantage of third-party software to do the low-level control. Mount operation was done via Software Bisque's TheSky, camera and filter wheel operation was done via MaxIm DL, and the focuser was operated through an ASCOM-like interface. Although these packages can be configured to work with a variety of available hardware, we found that some important hardware remains unsupported, some driver implementations are too buggy to support autonomous telescope operation, and some critical hardware features are not accessible through the third-party programming interfaces. Moreover, our original dome control and weather monitoring was done through custom-built software that was limited in terms of extensibility beyond systems very similar to PROMPT.

In the past 18 months, we have greatly expanded the level of hardware support offered by Skynet. We now have full ASCOM support for mounts, filter wheels, domes, and a wide variety of focusers. This allows us to communicate with numerous devices that were previously unsupported, and gives us finer-grained control over the operation of these devices. Where we have found it to be necessary, we have added options to our control software to accommodate various subtle differences in behavior between driver implementations.

In several cases, we have found that vendor-supplied drivers, while suitable for casual use, are not robust enough to provide reliable remote/autonomous operation. As the need has arisen, we

have implemented custom drivers to improve reliability, error detection capabilities, and access to device-specific features. Moreover, there are occasionally instances where a device is working poorly and cannot be repaired for an extended period. In these cases, we can customize a driver to work around these limitations so that the telescope can continue to operate in a slightly diminished mode until a repair is possible. Such custom drivers have been written for FLI and ACE filter wheels, AstroHaven domes, RCOS TCC focusers, and custom focuser units in use on the Morehead 24" and Yerkes 41" telescopes.

We have dramatically improved Skynet support for rotating domes. Skynet can now perform dome-telescope alignment using a variety of software packages, including ASCOMDome and Software Bisque's AutomaDome. Unfortunately, these packages are not able to correctly calculate the dome geometry for all telescope types and dome configurations – for example, the cross-axis mount design of the Yerkes 41" telescope is not well handled by either program. In these cases, we have implemented custom dome geometry routines that can be adapted and optimized to particular situations.

Some rotating domes use solar panels mounted on the side of the dome to charge batteries that power a wirelessly-controlled shutter motor. Our control software now rotates the dome periodically during the daytime to keep the solar panels aligned with the sun and maximize the battery charge before nighttime operation.

Skynet also now recognizes custom local horizons that may be enabled at each observatory. This helps to avoid observing targets in parts of the sky that may be blocked by local obstructions. This is also used as a safety feature for telescopes that cannot be slewed to certain orientations without risking cable damage or collisions within the observatory.

Support for weather monitoring has improved dramatically since 2009. We now have a generic weather interface that can communicate with a variety of weather sources, including Boltwood I and II cloud sensors, Davis WeatherLink and Vantage Pro weather stations, devices compatible with the Virtual Weather Station software package, and custom sources of weather information such as the CTIO weather database and weather telemetry from the GONG observatory. All of this data is reported back to the Skynet website in a consistent manner, so that our Skynet user community can see weather conditions at every Skynet telescope in an organized and consolidated manner.

In developing support for all of this new hardware, we have found it to be useful to create simulators that accurately imitate the behavior of each device. By having a device simulator, we can systematically recreate any number of failure modes and ensure that Skynet handles these failures in a sensible way. On multiple occasions, these simulators have also proven to be indispensable in diagnosing problems with remote devices. Our simulators typically work at the level of the serial protocol, although in some cases we work at higher or lower levels. We can now simulate approximately 25 different devices, summarized in the table below. These roughly correspond to the devices that have been tested and integrated with Skynet over the past two years.

Device type	Simulated devices

Mounts	ACL-compatible mounts; COMSoft control systems;				
	Paramount/BisqueTCS control systems				
Focusers	DFM TCS systems; RCOS TCC; Planewave EFA; specialized ASCOM				
	focus simulator				
Filter wheels	ACE; DFM; SBIG; specialized ASCOM filter wheel simulator				
Domes	DFM TCS-integrated systems; AstroHaven; MaxDome; ObservaDome				
	Labs; TI ProDome; AutomaDome simulator driver; specialized ASCOM				
	dome simulator				
Weather stations	Boltwood I; Boltwood II; Davis Vantage Pro; Davis WeatherLink II;				
	GONG weather source; VWS-compatible weather stations				
Cameras	Specialized ASCOM/MaximDL camera simulator				

The integration of the Yerkes 41" and ARO 30" telescopes stand as excellent examples of the new hardware control capabilities of Skynet. The 41" telescope was originally built in the 1960s, and the ARO telescope is a completely custom design. Both systems use control systems that are significantly different from the original PROMPT telescopes. However, both telescopes are now controlled by Skynet in a standardized manner as a result of the hardware work described above. These enhancements also allow previously integrated Skynet telescopes, such as GORT and Dolomiti, to operate in a more optimized manner. Moreover, Skynet is now in a strong position to integrate a number of new telescopes that we have been working with as they become available in early 2012.

Telescope evaluation / maintenance tools: We have a suite of evaluation, diagnosis, and maintenance tools that we continue to improve and expand upon. In the past two years, we have added tools for measuring telescope tracking performance around the sky, using open-shutter CCD readouts to measure vibrations (e.g., due to fans that are failing), and characterizing optical effects such as astigmatism, field curvature, and collimation error. We have also developed DIMM seeing monitor software to determine what the expected telescope performance should be at any given site.

Our software to build pointing models has been optimized and enhanced, adding support for new pointing modeling software such as PointXP (used with Planewave and SiTech mounts), supporting new types of sky-mapping sequences, and optimizing the exposure length necessary to get an accurate mapping at any given point in the sky.

We have also enhanced our auto-focus software, adding support for searching and slewing to appropriate nearby stars and intentionally de-centering the focus star slightly so that images taken immediately after the focus run are not affected by a ghost/RBI afterimage of the star near the target of interest. Our focus software can now map the dependence of telescope orientation on focus position – a relationship that can be quite substantial on some telescopes. We hope to incorporate focus/orientation modeling into Skynet in the near future to reduce the need for frequent auto-focus runs.

A small but crucial enhancement to the Skynet control software was to add a quick way to toggle between Skynet control and manual control. Under manual control, a user can run telescope tests or diagnostics while still benefiting from the dome control and status/error monitoring features, and once finished it is trivial to switch back to automated Skynet mode, minimizing the amount of telescope downtime.

Finally, the Skynet website now includes support for high-quality video and audio streaming to a user's web browser. This is a desirable feature for demonstrations and curious Skynet users, but it is critically important for staff that are operating and diagnosing problems at a remote site. Audio feedback can be used to detect problems such as failing motors or camera shutters. Video feedback can be used to diagnose issues such as dome misalignment and wrapped cables.

Error detection / notification: Error detection has always been an important part of the Skynet control software, and as we have integrated new hardware we have added support to the existing system for detecting new failure modes that we have discovered. We have also been developing a new system, the Skynet EventHub, which will be a far more sophisticated and flexible way of managing and distributing event notifications from all parts of the Skynet system. For example, the EventHub will detect if any of our critical servers (e.g. website, database, FTP) are offline, if operating system updates need to be applied to any Skynet computer, if hard drives are becoming too full, if system backups were not performed recently enough, if any part of the Skynet system is not responding in a timely fashion, and much more. This new system will make it far easier to add support for new types of notifications that might impact Skynet operation, even if they don't specifically come from a telescope. We expect to deploy the first version of EventHub before the end of the year.

Image analysis pipeline: We have rewritten the system that is used to calculate WCS (World Coordinate System) values for telescope images, and this system can now process every image that comes through Skynet. WCS establishes a relationship between image pixel positions and celestial coordinates, and we use this for many purposes, including determining telescope pointing errors, aligning images to produce stacks, and performing automated photometric calibrations against catalog sources. Our new system uses multiple WCS-solving programs when searching for a solution. Each program has different strengths, and one may succeed where another fails. For images that are difficult to solve automatically (e.g., images with only 2-3 stars, or near bright stars where catalogs tend to be sparse or polluted with false detections), we have created a web interface where users can interactively build and verify WCS solutions. In addition, our software can now produce WCS FITS headers in multiple formats to support the variety of image analysis software used by our Skynet community.

We have developed an enhanced process for generating master calibration images from the biases, darks, and flats that are automatically taken by Skynet each night. This process employs a number of techniques to reject bad component images, and to reject master calibration images if they are determined to be unsuitable. For example, a telescope may be having problems with the camera shutter, and a set of dark images might be taken with the camera shutter stuck partially open. The master calibration script will detect and report such anomalies, and revert to the most recent suitable master dark image.

A new system for distributing Skynet image data will be deployed by the end of this year. This system adds a number of new features. For example, users will have the ability to select between downloading raw images/calibration frames (for those who prefer to do their own calibrations),

or images that are calibrated using Skynet-generated master images. This system can also manipulate images on demand when a download is requested – for example, it can apply a 180 degree flip for images that were taken west of the meridian on a German Equatorial mount, it can update and inject new FITS headers that may be of use but were not written to the original FITS file, and it can rename or organize image files according to user preferences.

Documentation: We place a high level of importance on documenting potential failure modes and problems that Skynet operators may encounter. We also produce documentation describing procedures for telescope maintenance and operation. As of 2009, we had documented approximately 100 such problem/solution scenarios and procedures. In the past two years, we have added documentation for approximately 50 new problems or procedures. Many of these relate to the new hardware that we have

been integrating into Skynet.



Figure 5: 3D CAD model of PROMPT-Australia.

For telescopes that are more technically involved, we produce detailed startup and shutdown procedures so that telescope operators can set up their telescope in a way that is most likely to result in reliable operation. We have also created a set of recommendations for operating system and software installation to increase system stability, reliability, and security.

To aid in the design and construction of a number of upcoming Skynet systems, we have created accurate 3D CAD models for a variety of telescope, mount, and enclosure types. These models have been critical in making decisions related to current and future telescope and enclosure design.

E. Afterglow Modeling Project (AMP)

E.1. A New Statistic

The primary motivation for the development of Galapagos was the need to fit time- and frequency-dependent models to photometric observations of GRB afterglows. The complete model for the afterglow emission and all sources of absorption and extinction along the line of sight contains dozens of parameters. However, most of these parameters are not truly "free", but are constrained by prior probability distributions, either on the parameter values themselves, or on sets of parameters that describe the correlation of one parameter with another. These correlation parameter priors are obtained by using Galapagos to fit models to various two-dimensional data sets (see §E.2).

We have derived a new statistical technique to approach the very general problem of fitting a model curve to a two-dimensional data set, where there is intrinsic uncertainty in the measured quantities in both dimensions, as well as additional scatter in the data that is greater than can be

accounted for by the intrinsic uncertainties alone. In all that follows, we assume that the intrinsic measurement uncertainties and the extrinsic scatter, or sample variance, are normally distributed and independent in both dimensions.



Figure 12: Left: Intrinsic error ellipse of a data point and extrinsic sample variance convolved with model curve. The joint probability of the data point and the model is the integral over the x-y plane of the product of these two distributions. **Right:** This joint probability is equivalent to a path integral through a convolved error ellipse, where intrinsic errors and extrinsic sample variance are added in quadrature. The model curve is approximated by the red line tangent to the convolved error ellipse. TRF projects the differential element of path integration onto the blue line. Shaded areas indicate 1, 2, and 3σ confidence regions.

Consider a set of *N* points in the *x*-*y* plane with intrinsic uncertainties ("error bars") in both dimensions $\{x_n, y_n; \sigma_{xn}, \sigma_{yn}\}$ and a model distribution, described by a curve defined by *M* parameters $y_c(x; \theta_m)$ and by extrinsic sample variances σ_x and σ_y (see Figure 12a). The intrinsic two-dimensional probability distribution function of a measured data point is:

$$p_{\text{int}}(x, y \mid x_n, y_n, \sigma_{xn}, \sigma_{yn}) = G(x, x_n, \sigma_{xn})G(y, y_n, \sigma_{yn}),$$

where *G* is the Gaussian function:

$$G(x, x_n, \sigma_{xn}) \equiv \frac{1}{\sqrt{2\pi}\sigma_{xn}} \exp\left[-\frac{1}{2}\left(\frac{x - x_n}{\sigma_{xn}}\right)^2\right]$$

The model probability distribution is given by the convolution of the model curve with a twodimensional Gaussian:

$$p_{\text{mod}}(x, y \mid \theta_m, \sigma_x, \sigma_y) = \iint_{x', y'} \delta(y' - y_c(x'; \theta_m)) G(x, x', \sigma_x) G(y, y', \sigma_y) dx' dy' .$$

Bayes' theorem allows us to compute the probability of a given model distribution, given a set of measurements and any prior constraints on the values of the model parameters. Assuming the prior distributions of the parameters are flat, the best-fit model is found by maximizing the likelihood:

$$p(\theta_m, \sigma_x, \sigma_y \mid x_n, y_n, \sigma_{xn}, \sigma_{yn}) \propto \prod_{n=1}^N \iint_{x, y} p_{\text{int}}(x, y \mid x_n, y_n, \sigma_{xn}, \sigma_{yn}) \cdot p_{\text{mod}}(x, y \mid \theta_m, \sigma_x, \sigma_y) dx dy \equiv \prod_{n=1}^N p_n dx dy$$

For independent Gaussian intrinsic uncertainties and extrinsic sample distributions, the joint probability p_n is equivalent to a path integral through a convolved two-dimensional Gaussian probability distribution:

$$p_n(\theta_m, \sigma_x, \sigma_y \mid x_n, y_n, \sigma_{xn}, \sigma_{yn}) = \int_s G(x, x_n, \Sigma_{xn}) G(y_c(x; \theta_m), y_n, \Sigma_{yn}) ds,$$

where $\Sigma_{xn} \equiv \sqrt{\sigma_x^2 + \sigma_{xn}^2}$ and $\Sigma_{yn} \equiv \sqrt{\sigma_y^2 + \sigma_{yn}^2}$ (see Figure 12b).

For a given data point, the path integral along the curve y_c is approximated by finding the point at which the curve is tangent to an error ellipse centered on (x_n, y_n) with axes proportional to (\sum_{x_n}, y_n) Σ_{vn}) and integrating the convolved probability along a line through that point with slope $m_t =$ $\tan \vartheta_t$ (see Figure 12b). The surprising fact is that the result of this 1D linear path integral depends upon which axis the differential path element ds is projected onto (or, equivalently, whether you rotate the x-y coordinate system before performing the integral). D'Agostini (2005) rotation). Reichart [D05] chose ds = dx(no (2001)[**R01**] chose $ds = \sqrt{dx^2 + dy^2} = dx\sqrt{1 + m_t^2} = 1/\cos\theta_t$ (rotation by the angle θ_t).

We present a new statistic **[TRF]** where the differential element is projected onto a line perpendicular to the segment connecting the data point centroid and the tangent point of the curve to the convolved error ellipse; the angle between this line and the tangent line is φ_t , and its angle with respect to the x-axis is ψ_t (see Figure 12b). This is equivalent to setting the differential path element to:

$$ds = \frac{\cos\varphi_t}{\cos\varphi_t} dx = \frac{\sum_{yn}^2 + m_t^2 \sum_{xn}^2}{\sqrt{\sum_{yn}^4 + m_t^2 \sum_{xn}^4}} dx,$$

or to rotating the coordinate system by an angle ψ_t before performing the integration. A more intuitive interpretation of TRF is that it is mathematically equivalent to a modified 1D χ^2 statistic:

$$p_n^{\text{TRF}} = \frac{1}{\sqrt{2\pi}\Sigma_m} \exp\left[-\frac{1}{2}\left(\frac{\delta_m}{\Sigma_m}\right)^2\right],$$

where δ_m is the radial distance between the data point centroid and the tangent point (the point at which the model curve is closest to the data point, in terms of Σ), and Σ_m is the 1-sigma radius of the convolved error ellipse along that axis.

Three criteria motivate our choice: (1) The statistic should be invertible, *i.e.*, if $x_n \leftrightarrow y_n$, $\sigma_{xn} \leftrightarrow \sigma_{yn}$, and $\sigma_x \leftrightarrow \sigma_y$, the best-fit model parameters should describe the curve $x_c(y;\theta_m) = y_c^{-1}(x;\theta_m)$; (2) The statistic should reduce to the traditional 1D χ^2 statistic in y or x when $\Sigma_{xn} = 0$ or $\Sigma_{yn} = 0$, respectively; and (3) When applied to a data set consisting of only two points, the statistic should produce a best-fit line that intersects the centroids of both points' error ellipses. The following table summarizes the properties of the D05, R01, and TRF statistics:

	D05	R01	TRF
Differential Element <i>ds</i> =	dx	$1/\cos \theta_t dx$	$\cos\varphi_t/\cos\vartheta_tdx$
Invertible?	No	Yes	Yes
	D05	R01	TRF
Reduces to 1D χ^2 ?	Only if $\Sigma_{xn} = 0$	No	Yes
Fits 2 Data Points?	Only if $\Sigma_{xn} = 0$	Only if $\Sigma_{xn} = \Sigma_{yn}$	Yes

Linear fits to circularly symmetric Gaussian random clouds of points illustrate the bias inherent in the D05 statistic. R01 and TRF fits are invertible $(m_{xy}=1/m_{yx})$, while D05 is biased towards m = 0 whether fitting to y vs. x or x vs. y: The probability distributions of $\mathcal{G} = \tan^{-1}(m)$ for ensembles of fits to Gaussian random clouds of N points are $p_{R01}(\mathcal{G}) = p_{TRF}(\mathcal{G}) = \text{constant}$, while $p_{D05}(\mathcal{G}) \propto \cos^N \mathcal{G}$. Linear fits to sets of data points generated by adding Gaussian random noise to a linear model show that D05 consistently underestimates the generating slope when there are error bars in two dimensions. The only case for which D05 correctly recovers the generating parameters is when the intrinsic uncertainty and sample variance are in the y-direction only, where D05 is equivalent to the 1D χ^2 statistic. The R01 statistic does not reduce to χ^2 in the 1D case, and the best-fit slopes are biased in cases where the axial ratios of the error ellipses are large (R01 is equivalent to TRF when $\Sigma_{xn} = \Sigma_{yn}$).

E.2. AMP Extinction and Absorption Models

While the intrinsic emission of an afterglow can typically be described by one or more powerlaw curves (sometimes smoothly broken) in time and frequency, accounting for the absorption due to gas and extinction due to dust in the host galaxy, the Milky Way, and the intergalactic medium is rather more complicated. The full extinction/absorption model contains no fewer than 50 parameters. Though most of these parameters are constrained by priors, finding the best-fit afterglow model in this complicated parameter space would be effectively impossible without a tool as flexible and efficient as Galapagos. Furthermore, using Galapagos's parameter-linking capabilities, it is possible to explore changes in the circumburst environment over time: Those parameters that describe source-frame extinction and absorption can be allowed to vary independently for different subsets of the data, while others, such as the burst redshift and the parameters that describe intergalactic absorption and extinction due to dust in the Milky Way along the line of sight, can be linked across the entire data set.



Figure 13: The combined CCM/FM dust- extinction model.

We model extinction due to dust in the source frame of the host galaxy and in the Milky Way using the near-UV through infrared extinction model of Cardelli, Clayton & Mathis (1989) [CCM] and the UV extinction model of Fitzpatrick & Massa (1988, 1990) [FM].⁶⁴⁻⁶⁶ The extinction at a given wavelength λ can be expressed as:

The parameter A_V normalizes the extinction curve in the V band, and $R_V = A_V / E(B-V)$ is a measure of the extinction in the B band relative to that in the V band.

Extinction due to dust in the Milky Way is described by the CCM model, with an asymmetric Gaussian prior, $\log R_V^{MW} = 0.423_{-0.010}^{+0.082}$, and a fixed value of $E(B-V)_{MW}$ from all-sky IR dust-emission maps.⁶⁷

Extinction in the host galaxy is described by a combination of the CCM and FM models (see Figure 13); the only free parameters are A_V and the CCM model parameters c_2 and c_4 . The other parameters in the model, including R_V , are constrained by priors obtained by fitting empirical functional relationships, using Galapagos and the TRF statistic, to measured extinction parameters for 417 stars in the Milky Way and 23 stars in the Large and Small Magellanic Clouds (see Figure 14).^{68,69}

For source-frame $x = (\lambda / 1 \mu m)^{-1} < 1.82$, we use the CCM extinction model:

$$\frac{A_{\lambda}}{A_V} = a(x) + \frac{b(x)}{R_V},$$

where a(x) and b(x) are empirical functions fitted by CCM. For 3.3 < x < 10.96, we use the FM extinction model:

$$\frac{E(\lambda - V)}{E(B - V)} = c_1 + c_2 x + \frac{c_3}{\gamma^2} \frac{x^2}{\left((x^2 - x_0^2) / \gamma \right)^2 + x^2} + c_4 F(x) ,$$

where F(x) is an empirical function fitted by FM that describes the shape of the far-UV excess. For 1.82 < x < 3.3, we use a weighted average of the two models (see Figure 13). In the AMP dust-extinction model: c_2 , c_4 and A_V are free parameters; c_1 , UV bump height c_3/γ^2 , and R_V are constrained by priors on the correlation model parameters (see Figure 14); while γ and x_0 are constrained by Gaussian priors: $\gamma = 0.895 \pm 0.141$ and $x_0 = 4.584 \pm 0.019$.



Figure 14: 1, 2, and 3σ fitted model distributions to correlations of dust-extinction parameters measured for 441 stars in the Milky Way and Large and Small Magellanic Clouds. values of the UV extinction parameter $c_2 \sim 0$ correspond to "gray dust", typical of young star-forming regions (SFRs) with strong stellar winds, as in the Orion Nebula; high values correspond to higher ratios of small grains to large grains, typical of older SFRs with supernova shocks, as in the SMC. Left: c1 vs. c2. Middle: R_V vs. c2. While $R_V \sim 3.1$ is typical of stars in the Milky Way and older SFRs, higher values, i.e. smaller E(B-V), are found in young SFRs with low c2 values. **Right:** UV bump height vs. c2. UV bumps are thought to be due to resonances in the lattice structure of graphitic dust grains, and are typical of Milky Way dust-extinction spectra. Dustextinction spectra in SFRs, with both high and low values of c2, exhibit less prominent UV bumps.



Figure 15: IGM Lya transmission vs. sourceframe absorber redshift. The solid line is the best-fit empirical model curve, transformed into lnT space; shaded regions are the transformed 1, 2, and 3σ sample variance intervals for a typical bin width $\Delta z = 0.07$. Note the onset of the Gunn-Peterson trough near redshift $z \sim 6$.

To model Ly α absorption in the intergalactic medium, we use Galapagos and the TRF statistic to fit an empirical model to transmission *T* vs. absorber redshift *z* based on observed flux deficits for 64 QSOs, measured in binned regions of width Δz of their spectra blueward of Ly α in the source frame (see Figure 15).^{70,71} The empirical model has the form:

 $\ln(-\ln T) = \ln\{\exp[b_1 + \tan \theta_1(z - z_1)] + \exp[b_2 + \tan \theta_2(z - z_2)]\},$

with sample variance in ln(-ln*T*) given by $\sigma_{\ln(-\ln T)} \propto \sigma_0 (1+z)^{\alpha} (\Delta z)^{-\frac{1}{2}}$. The redshifts z_1 and z_2 are chosen to minimize correlations among fitted parameters. We model Ly α -forest absorption by fitting to parameters b_1 , ϑ_1 , b_2 , ϑ_2 , σ_0 , and α , which are constrained by priors, and to an offset in each photometric filter in ln(-ln*T*) with a zero-mean Gaussian prior that scales as $(1+\overline{z}_F)^{\alpha} \Delta \overline{z}_F^{-\frac{1}{2}}$, where \overline{z}_F and $\Delta \overline{z}_F$ are the effective weighted mean absorber redshift and bin width, respectively, of the overlap of filter *F* with the Ly α forest. The host-galaxy redshift z_{GRB} is typically held fixed, based on spectral or other observations, but can be allowed to vary if we wish to independently determine a photometric redshift.

The AMP model also includes a damped Ly α absorber profile at the source redshift, whose shape is parameterized by the neutral-hydrogen column density $N_{\rm H}$, and total absorption at wavelengths shorter than the Lyman limit $\lambda < 912$ Å in the source frame. $N_{\rm H}$ may be fit as a free parameter, or with prior constraints from X-ray or preferably optical spectroscopic observations, when available. We have also developed an empirical model of absorption due to roto-vibrationally excited molecular hydrogen, based on fits to theoretical spectra.



Figure 16: Mid-run screen of Galapagos fitting capture Skynet/PROMPT and Skynet/DAO (Italy) observations of the slowrising afterglow of GRB 090313 at z = 3.375, which exercises both AMP's extinction and absorption models (the Ly α forest can be seen in the foreground of the image). AMP measures source-frame $A_V =$ 0.55 - 0.59 mag and an SMC-like

 $c_2 = 2.1 - 2.5$ depending on the emission model.

F. Galapagos

F.1. Motivation

Effectively modeling GRB afterglows, like all data-driven astroinformatics tasks, requires attacking difficult problems in information integration and analysis. Combining previously determined information as well as data in multiple formats and from disparate sources is essential to properly determining parameterizations for detailed models with rigorous statistical methodologies. The complexity of these physical effects and analysis techniques, in turn, induce a vast space of possible solutions that is almost impossible to adequately explore via traditional methods.

Reliable scientific analysis involves properly leveraging a priori information. For instance, a solution obtained by modeling photometric light-curve data can be confirmed in part by comparing the predicted redshift value to one established by more precise means, such as spectroscopy.⁷² While this approach has merit, a more complete treatment actually incorporates

this prior knowledge into the modeling process from the outset by building independently determined information into the measure of overall likelihood for a given solution candidate using a Bayesian inference scheme.⁶²

While PROMPT, and the whole Skynet Robotic Telescope Network, provides a wealth of GRB follow-up images, we are often required to augment these observations with data from other sources. Therefore, a truly definitive system for modeling GRBs will facilitate collaborative research efforts by adequately addressing data format and source considerations.⁷³ While GRB models and spectral data are described over the frequency continuum, most rapid-time follow-up data (often from small telescopes) is reported in terms of discrete photometric magnitudes. Before candidate solutions can be compared to the photometric data, we need to map the modeldictated spectrum to the data-space via filter integrations.⁷² While conceptually undemanding, the multifarious composition of GRB data sets with photometric points from across the spectrum and the sensitivity of these mappings to changes in the underlying model necessitate a flexible, yet robust, and modular framework for applying these translations. The modeling task is further complicated by the relationships among the various observations. For instance, the overall data set may consist of points collected and calibrated by different means and at different times. As a result one subset of the data may possess common calibration errors while a different, but nondisjoint, subset may share specific temporal characteristics. Thus, fluidly enforcing and accounting for these commonalities between various overlapping data subsets is absolutely essential for effective data integration.

Additionally, drawing accurate conclusions from observations requires uniform, rigorous, and complete statistical methodologies and physical models. For example, the proper treatment of data sets whose scatter often overwhelms the measurement error, as is typically the case with GRBs. Rather than applying a simple predictor model, the data must be described by a distribution that accounts for both systematic and statistical errors at the cost of additional model parameters and more complex evaluations of candidate solutions.⁶² Statistical techniques aside, simply applying the physical models entails a whole slew of caveats. Many of these stem from the fact that any comprehensive system for modeling GRBs must be able to describe a series of distinct effects. However, any particular GRB may only present with a subset of these effects at play; as a result, modularity is crucial. Consider the emission and extinction mechanisms, which can be further broken down, respectively, into discrete components, including the internal, reverse, and forward shock emission and the source-frame and Galactic extinction. To effectively describe the data, these (and other) model components must be able to vary independently while behaving in a coherent fashion when considered together.

With this in mind, we can view modeling GRBs as a problem of maximum likelihood estimation. In essence, we are searching for the hypothesis, H, of maximum probability, P(H/E), given the evidence, E. Our hypothesis can be expressed as the combination of a parameterization, θ , and a model, $M(\vec{x} | \theta)$, which, given a vector of independent variables, \vec{x} , produces a prediction. The evidence, in this case, consists of any observed data and may include independently determined information. Once the tools of physical theory dictate the structure of the model, our next charge is to find the corresponding parameterization that maximizes the likelihood of our hypothesis given a set of evidence. With a complex model, this quickly becomes a difficult problem in non-linear function optimization over a vast, multi-dimensional solution space. While deterministic optimization systems offer the comfort of a well-studied theoretical foundation, the efficacy of these methods tends to break down with increasing problem complexity, especially in high-dimensional and deceptive landscapes.⁷⁴⁻⁷⁶

F.2. Genetic Algorithm

There are a number of stochastic heuristics well-suited to this type of global optimization problem; however, none offer the flexibility, scalability, and efficiency of a genetic algorithm (GA). GAs work by simulating the mechanism by which heredity and natural selection guide the evolutionary process in creating organisms able to thrive under a given set of conditions. First, we generate a series of "organisms" or model parameterizations by randomly selecting "genes" or values from the range of all valid parameters for each dimension of the solution. Each of these organisms, the set of which now form a "population", are then assigned a "fitness" by evaluating the probability of the corresponding hypothesis it (along with their model) represents. Speaking loosely, this objectively measures how well the candidate solution matches the data. The organisms are then ranked according to fitness, whence the selection phase occurs; the less "fit" solutions are discarded, while those with greater fitness are combined to replenish the ranks. This "mating" is accomplished by selecting a pair of parents, and then randomly choosing genes from among the pair to build an "offspring". After these offspring are evaluated for fitness, they are added to the population to replace the least-fit members. Iterating this process of continually replacing less-fit solutions with those better suited to the data, the weaker organisms wither away as a particular genotype of high fitness eventually comes to dominate the population. At which point, we arrive at a solution of maximum likelihood and terminate the search.

This approach offers a number of key advantages over competing stochastic optimization schemes in terms of flexibility, scalability, efficacy, and efficiency. First, the "fitness function" evaluating each organism is wildly flexible; in fact, it can be any function which maps a vector model parameterization to a scalar fitness value. Unlike Markov-Chain Monte-Carlo (MCMC) type heuristics (including simulated annealing), rank-based GA's probe the solution space by comparative measures of fitness and are, therefore, independent of the scaling of the fitness function.^{77,75} This quality affords us the flexibility to add and reconfigure model components at will without significant impact to the rest of the algorithm. Second, this method will scale with increasing problem size; even after adding parameters to the model, the algorithm naturally spreads its search over each dimension of the problem more effectively than MCMC-type methods.⁷⁸ GA's often prove to be more effective than MCMC optimization engines by better avoiding local-minimae and producing higher-quality solutions.⁷⁹⁻⁸¹ Finally, since we can measure the fitness of each organism independently, we are able to leverage all available simultaneous computing resources for reduced time-to-solution.⁸²⁻⁸⁴ This kind of parallelism is especially important in light of the fact that the most pervasive trend in terms of maximizing computational throughput in commodity hardware comes in the form of increased multi-tasking over raw clock speed.⁸⁵

To gain the maximum advantage from these, and other, aspects of GAs in building Galapagos, we have deviated from the canonical implementation described above; specifically, we have further sub-tasked the various functions to a truly multi-threaded environment, eliminated the notion of discrete generations, and introduced an additional level of iteration. Galapagos works as follows: Upon initializing the engine, a number of threads are spawned, namely a manager and a series of workers. The workers begin by randomly generating and evaluating organisms

during an initial phase. After an organism is evaluated, it is handed off to the manager thread, which organizes the incoming suggestions into a coherent population. When this population has reached a predetermined size, the workers switch into a main phase and begin generating organisms by mating existing parents. As these new suggested solutions come in, the manager compares them against the current weakest population member, and if appropriate, discards that weakest member in favor of the suggested offspring, essentially maintaining a running list of the best organisms the search has thus encountered. Once the fitness values of the best and worst organisms in the population differ by less than some predefined small constant, rather than terminate the search as above, we re-center the initial search space on the most-fit population member and run the algorithm again. Since our method for modifying the search space includes a damping factor, successive solutions that are sufficiently proximal will cause the search space to shrink in each dimension around some localized area, thus providing a more thorough and complete probing of the region of maximum likelihood. When a sequence of these solutions has the quality that all of their respective fitness values fall within some tolerance of each other, we conclude the search and select the best member of this sequence as the overall solution (see Figure 17). Finally, we deviate from the error estimation methodologies described in Mokiem and Curran in favor of a more rigorous analysis.^{86,72} Parameter uncertainties are obtained by perturbing the optimum values and measuring the effect on the overall solution quality. Concretely, this means nominally incrementing a parameter and then freezing it, intentionally handicapping the fitting procedure. We then use the same genetic algorithm to search this solution space hyper-plane for a "best fit" given the handicap. By slowly marching away from the optimum value and tracking the solution fitness, we establish an empirical relationship between the parameter value and solution quality, from which we extract N-sigma confidence intervals.



Figure 17: Demonstrates the iterative refinement by serial application of GAs. Once the difference between the best and population members worst become negligible (indicating a static population of clones), we re-size the search space and re-run the GA. Once these successive applications begin to produce identical results, we terminate the search. This plot was generated by sampling the best and worst organisms of the population 100 times per second while simultaneously fitting three fourth-order polynomials to a set of test data.

The specifics of this implementation improve both the time-to-solution and the overall convergence stability. First, exploiting this producer-consumer threading strategy results in minimal synchronization, requires no active load-balancing, and maximizes data-locality.^{87,88} The de-centralized nature of the managing strategy also makes it agnostic to the quantity and quality of available workers, meaning the implementation scales effectively from low-power notebooks to large, multi-processor workstations to heterogeneous clusters of networked machines. Secondly, by updating the population as better suggestions become available or

removing the "generation gap", we decrease the latency between the discovery of helpful solution components and their subsequent availability for recombination, ultimately being more aggressive with more promising solutions.^{77,89} Finally, to balance this increased emphasis on exploitation of good-solution "building blocks" versus the exploration of the parameter space, the iterative application of the GA reduces susceptibility to sub-optimal solutions. Consider a problem of fitting an N parameter model; initially generating a population of size M, means M values are in the gene pool for each of the N problem dimensions. Applying the mating and recombination operators allows for a directed exploration of the N^{M} possible points in the solution space. The probability of landing within some specified neighborhood of the optimal solution with one of these points, while initially small, increases as the overall volume of the parameter space shrinks around this region of maximum likelihood. On the contrary, if the mostfit organism from one run lies far from the center of the search space, the resizing scheme can actually expand the box to more adequately search alternative maxima. In fact, this quality allows Galapagos to find optimum solutions which might lie outside the initial search interval. While no stochastic heuristic can guarantee success in global optimization problems, our implementation provides a series of tunable parameters to help the search compensate for difficult fitness landscapes.

F.3. Fitness Function

A fitness function gives the GA context by encapsulating the specific details of the problem at hand. As a result, its organization is important. While Galapagos can be applied to any general problem of function optimization, the following organizational hierarchy applies chiefly to data regression and is discussed in the context of GRB modeling.

As stated above, one of the primary design goals for Galapagos was flexibility, which, in a system such as this, means pervasive component modularity. We generally break the fitness function down into a "statistic" that measures how well a hypothesis describes the data and a "model" that maps a range of independent inputs to a hypothesis's prediction. The statistic generally implements the probability function discussed above and takes a model, a corresponding parameterization, and a set of evidence as arguments. This allows the optimization engine to be applied to a whole slew of problems, given the right metrics for classifying candidate solutions. In terms of data regression, one such metric could be a simple χ^2 measure that sums the squares of the differences between the model prediction and the observation at each available data point. Returning the inverse of this summed quantity as a particular parameterization's fitness provides a way of objectively ranking these sets of parameters such that hypotheses matching the data more closely have better fitness values. Therefore, given a data set and an appropriate model, simply pointing the optimization engine at this statistic will produce a χ^2 regression.

In modeling GRBs, however, we employ a significantly more involved full Bayesian treatment. Bayes' theorem states that the posterior probability of a solution is proportional to the product of the probabilities of the hypothesis data given the hypothesis (which we call the likelihood function) and the hypothesis given prior knowledge (which is often expressed as the product of a series of independent "priors"). Given that the evidence, E, is broken down into objective data observations. and prior information. I. illustrate this or D. we can as

 $P(H | E) \propto L(D | HI) \prod p_i(H | I)$. These subcomponents are naturally implemented in a modular fashion. The structure of the Bayesian inference statistic in Galapagos allows the user to add, and interchange priors at will. In fact, since priors operate on the full parameterization of the model being evaluated, they are allowed to be arbitrarily complex, from measuring the likelihood of a single parameter with a simple Gaussian distribution to implementing user-defined functions that operate on the whole candidate genome. In the same manner, the likelihood function is also flexible; in fact, it can be a whole additional nested statistic. For example, we could use a simple χ^2 statistic as the likelihood function to perform a regression as above but bias the parameters using independently determined information in the form of priors to arrive at a more complete solution. Thus implemented, we are even free to extend this treatment using the statistics of Reichart (2001), D'Agostini (2005), or Trotter, Reichart & Foster (§F).^{62,63}



Figure 18: AMP's modular fitness function (§*F*).

The other important component of the fitness function is the model, which is primarily responsible for contextualizing the parameterization. It serves as a function. which. given this parameterization, maps a vector of independent variables to a prediction. Data modeling is often an organic process of synthesizing various sundry

components whose respective functional forms are both well understood and determined by theory. Therefore, like the statistic, the model can range in complexity from a simple function to a whole hierarchy of interdependent modules. For instance, our typical GRB model is separated into emission and absorption models, each of which has a number of sub-components which can be added, removed, and configured by the user at will (see Figure 18).

F.4. Data Structures

Quality and quantity are important concerns when gathering data for any kind of modeling task. Often times the investigator must strike a delicate balance between these two in the interest of drawing useful and reliable conclusions. One technique to increase the quantity of information available without necessarily sacrificing quality is to incorporate related observations into the overall data set. This supplemental data, however, must be handled properly so as to reinforce relevant characteristics of, while adding minimal scatter to, the original data set. This sets the stage for the general problem of defining and accounting for non-trivial relationships between different data subsets.

Consider the following demonstrative problem: We wish to determine the mass of a planet by recording the locations of *N* of its moons as a function of time and modeling the orbital dynamics. For an edge-on projection, the trajectories can be suitably described by the function $x(t) \approx a \sin(\sqrt{GM/a^3}t + \varphi)$, where *a* is the semi-major axis of the orbit, *G* is the

gravitational constant, M is the mass of the central body, and φ is a phase constant. Now, one approach is to model each trajectory independently, and thus arrive at N distinct values for M, which can then be averaged to provide a reasonably accurate estimation. The alternative is to recognize that the underlying phenomena producing the data, while largely independent, actually share some characteristics, namely the parameter M. A more complete treatment would be to include this information from the outset.

Galapagos allows us to do this by defining the notion of data groups, semi-independent subsets of the overall data set. When we perform a regression using multiple data groups, Galapagos actually performs multiple regressions simultaneously by expanding the genome of each organism to include N "chromosomes" or semi-independent parameterizations for the model. In the example above, this means the organisms we generate will have not 3 parameters (a, M, and φ), but 3N parameters, one set for each moon trajectory. During the fitness evaluation, the statistic iterates over the entire data set and compares how well the overall parameterization predicts each moon's position. The statistic decides which 3-parameter set to use when calculating the model prediction based on which data group it is being evaluated at that instant. At the end of the run, the overall solution is N different parameterizations (of 3 values each) that quantify the orbits of the N distinct moons in the context of the functional form mentioned above.

As before, this produces N distinct values for the mass of the central object; however, we are able to enforce a relationship between the data groups by "linking" the parameter M across the different chromosomes. This amounts to setting the N different mass parameters in each organism to be identical after its inception. What was a fit with 3N free parameters becomes one with 2N+1 free parameters, and we gain the advantage that the optimization engine now determines which single value of M best describes the data as a whole, all other factors being unrelated.

Data groups and linking become powerful concepts when combined with GAs. Even though adding groups, in this manner, multiplies the dimensionality of the overall fit by N, linking parameters across groups actually subtracts from the total number of free parameters and thus reduces the size of the solution space. Besides this, GAs are extremely well suited to searches over a large number of dimensions, as mentioned before, because they scale quite naturally in this respect. When modeling GRB data, we apply data groups to capture changing temporal characteristics of burst events. For instance, we define a handful of data groups corresponding to natural breaks in the behavior of the GRB light curve. During the modeling step we hold a number of parameters constant across the entire interval of the GRB; for instance, the redshift parameter, z, does not change measurably over the duration of the burst event. We also allow other parameters to vary, like A_V , which describes the dust-extinction characteristics in the source frame of the burst. By modeling all temporal components simultaneously and sharing or freeing key values, we ensure the integrity of the overall fit while securing the ability to measure small, but telling changes in certain parameters. For example, trends in the value of A_V over time indicate a changing dust profile in the circumburst environment as time progresses, probably a result of jet spreading illuminating different dust populations, both those affected and unaffected by high-energy radiation from the GRB itself.^{90,91}

Decomposing data into semi-independent subsets and modeling the relationships between the groups becomes exceptionally important when integrating data from different sources. For instance, networks of small robotic telescopes provide a wealth of useful photometric data, but non-uniformities in the data acquisition, calibration, and reporting process can skew results if not properly handled. Even highly controlled surveys can exhibit significant photometric zero-point uncertainties that propagate to all measurements based on those standards.⁹²⁻⁹⁴ Assigning data from various sources to specific data groups, we can account for and measure these uncertainties by allowing the optimization engine to perturb entire subsets in a uniform way. Concretely, this is accomplished by supplementing the model's parameterization with additional entries defining data offset values. We then constrain these parameters with zero-centered Gaussian priors with widths suggested by the data calibration uncertainty. The end result is that the optimizing engine can shift the data group, as a unit, by some small but reasonable amount if it benefits the overall fit and provide a measure of the systematic calibration error.

F.5. Implementation

Galapagos has been designed and built with special emphasis on speed, reliability, extensibility, scalability, and portability. The core optimization engine is written entirely in C++, and has been extensively optimized with profiling tools. We took special care in developing our master-slave model for work distribution, which uses separate threads of execution for organism generation and population maintenance. This kind of parallelism requires fewer shared resources and less overall processor communication and thus allows for monumental improvements over turnkey solutions for multi-processing capability. In concrete terms, this means fewer context switches, nominal processor synchronization, and more efficient data caching, all of which reduce CPU overhead and idle-time during costly memory operations and direct every available clock cycle toward generating and evaluating candidate solutions.



Figure 19: Demonstrates the performance scalability of our application for various population sizes. The superlinear speedup seen in the 25 organism population is attributed to processor caching effects. This plot was generated by measuring the overall time-to-solution for a series of three simultaneous fourthorder polynomials to a set of test data and averaging over 30 runs on a workstation with 8 processing cores.

As a system that solves generalized optimization problems, Galapagos is also highly extensible. As explained above, a GA is meaningless without a proper fitness function, which can be any metric of quality that maps a vector candidate solution to a scalar fitness. Therefore the fitness function is implemented using dynamic libraries (Dynamic Link Libraries under Windows and Shared Objects under Linux, written in C or C++). In this way, a user can apply one or a whole chain of established statistics and models to suit his or her needs by specifying the structure of a fitness function at run time. Alternatively, if no pre-built modules prove adequate, a user can write his or her own plug-in, which, for simple models, typically requires less than five lines of code, using an open API.

In designing and implementing Galapagos, we underscored the importance of scalability, both in terms of problem complexity and performance. The "building block" nature of the fitness function, through recursive composition, means it can be viewed as an arbitrarily complex tree of evaluations. For instance, some GRB data sets require over 80 different components, like priors, filter integrations, and absorption and emission models. The limiting factor in solving complex optimization problems becomes solely computational throughput. Therefore, we have also focused on performance scalability and efficient use of all available computational resources. GAs are well suited to parallel computing architectures, and our Galapagos implementation, with minimal resource locking, achieves near-linear speed-up for a range of problem sizes on shared-memory systems. We have also spent time adapting the algorithm to a distributed resource scheme, using clusters of machines to act as hierarchically organized "islands", evolving separate sub-populations, and exchanging organisms over a network through a process of "migration". With this functionality, we can attack larger and more complex problems by scaling computational resources with widely available and even existing, under-utilized commodity hardware.

Finally, we have also built Galapagos to ensure portability and usability. We use Boost libraries for platform-independent threading and networking as well as the commonly used open-source plotting utility, Gnuplot, for visualization. The components of the fitness function, the overall GA behavior, and initial search space are supplied at run time to the Galapagos engine via a plain-text input file. The project is supported for Windows and Linux and has been ported to Mac OSX.

F.6. Next Steps

Next steps include the following efforts in refining and extending Galapagos's abilities. First, we will further develop the algorithm to bring robustness and a greater reduction in time-to-solution. We have carried out preliminary studies probing the interplay between the search-space resizing scheme (how aggressively we pursue particular solutions) and the population size (how densely we sample the solution space) and have found loose relations that guarantee an optimal balance between maximizing algorithmic reliability and minimizing time-to-solution. These results will have an important impact on and supply a general empirical framework for all kinds of stochastic optimization methods. Secondly, we would like to hone the actual software implementation by improving the efficiency of data caching and parallel resource access by restructuring the API.

Most importantly, however, we will enhance Galapagos's utilization of available resources by further improving performance in a distributed environment. GAs differ from other types of parallelizable problems (like N-body simulations or massive data crunching tasks) that simply distribute a fixed amount of work over a set of processors. Instead, they require a truly dynamic communication strategy that allows computing nodes to share information about constantly evolving state of the search. While the speedup for shared-memory, multi-processor workstations scales almost linearly, when working with distributed memory systems, like

computer clusters, the performance scaling is less impressive.⁸⁷ Cantù-Paz proposed an analysis that blamed the diminishing returns on the increased communication cost associated with managing larger fleets of workers and declared that "hybrid" algorithms, combining course and fine-grained parallelization strategies, were the most promising way to break this performance barrier.⁹⁵ We propose extending the hybrid GA concept to a hierarchical mode where distributed managers amalgamate candidate solutions from subsets of the total worker pool and coordinate results among themselves, thus requiring each manager to maintain fewer active connections and ameliorating the relative cost of communication overhead. Through investigating various migration policies and network topologies, we will develop an improved inter-process communication infrastructure that allows Galapagos's near-linear performance scalability to continue into the realm of distributed computing. Enhancing reliability and scalability will genuinely transform how we, as scientists, approach complex optimization problems by lifting the performance ceiling to a new level, limited only by resource availability.

The de-centralized nature of the worker-manager model permits dynamic resource accessibility as well because the system is structured in a manner that is agnostic to the number and quality of collaborating worker threads. This means computers can join and leave long-running searches at will, with no appreciable detriment to the overall progress of the algorithm. Therefore, with the communication infrastructure discussed above, Galapagos will be able to leverage not just the capabilities of individual workstations, or local computing clusters, but even dynamic grids formed by temporarily co-operating clusters from different institutions.⁹⁶ This will make the overall system even more amenable to complex, large-scale problems as greater performance can be gained by simply supplying additional commodity hardware or even exploiting idle-time on existing resources.

F.7. Broader Impacts: Broad Availability and Ease of Use

Along these lines, we will continue to improve the availability and usability of Galapagos with a tool-chain that facilitates working with the optimization engine. Not only has the system been ported to a number of operating systems, but once deployed, it is easy to take advantage of the range of options available to the user, from building custom fitness functions to using available tools to refine modeling techniques. Even setting up simple computing clusters is as easy as pointing slave nodes at a central server's IP address. With the further refinements detailed above, instances of Galapagos running at any number of locations, with the operator's permission, will be able to form a self-organized meta-cluster, or computing grid, for the coordinated application of hundreds or thousands of processors to a single problem. This kind of cooperation with take the current state of optimization studies to a whole new league, not simply because we will be able to find solutions more quickly, but because we will be able to approach problems in a more uniformly rigorous and definitive manner than ever before.

Since this resource pooling can be entirely ad hoc, it is possible to configure local computing resources to only attach themselves to larger optimization processes during large portions of otherwise idle time. Thus, underutilized resources can be easily redirected and amalgamated where necessary. To facilitate this coordination and exchange of resources, we will build a central web hub for tracking the availability and division of computing time, allowing users to develop collaborative computing efforts, as well as supplying useful tools to scientists. For

example, in its current form Galapagos requires a carefully formatted input file, but we will develop a graphical web interface to streamline the process of building fitness functions. Most importantly, since Galapagos has applications to almost any kind of optimization study across a range of disciplines, this online community will promote sharing fitness function components as well as expertise in their implementation and utilization. We will also develop tutorials so scientists can easily develop their own model and statistic modules and publish them for general use, if willing. Along with Galapagos's widespread availability and ease of deployment, a central information exchange such as this will advance the fields of function optimization and general data modeling in providing an educational forum and a proving-ground for both advanced and novel techniques.

Finally, the web presence will not only target scientists but the public at large, much like SETI@home and Folding@home. This component, called Science@home, will allow members of the general public to run Galapagos in the background on their own machines and contribute computing time to larger optimization or modeling efforts. In exchange, users will have information and statistics on the science they are supporting at any given time as well as being able to track their respective contribution. This dimension allows Galapagos users to capitalize on underutilized private computing resources as well as disseminate information about their particular topics.