1

Pathways to Discovery: From Foundations to Frontiers

We live in a time of extraordinary discovery and progress in astronomy and astrophysics. Since the dawn of the millennium, breakthroughs have come at an astounding rate, with highlights that include the first direct detection of gravitational radiation from astronomical sources; the discovery of thousands of extrasolar planets, including potential Earth-like analogs and the first characterizations of the physical properties and atmospheres for gaseous giant planets; mapping of the nascent disks of other solar systems as they are forming; a unified paradigm for the formation and evolution of galaxies, including deep insights gained from the fossil record of the Milky Way Galaxy; precision measurements of the supermassive black hole in the Milky Way's center; the first direct image of the shadow of a supermassive black hole; and precision measurements of the dark contents of the universe itself. Six Nobel Prizes for discoveries made using astronomical data have been awarded over the past decade alone (dark energy, gravitational waves, neutrino oscillations, the discovery of exoplanets, cosmology, supermassive black holes). Many ambitious scientific visions have been fulfilled in the past 10 years, but, if anything, momentum has only grown.

Every decade, the agencies that provide primary federal funding for astronomy and astrophysics—the National Aeronautics and Space Administration (NASA), the National Science Foundation (NSF), and the Department of Energy (DOE) Office of Science—request a decadal survey to assess the status of, and opportunities for the nation's efforts to forward our understanding of the cosmos. The National Academies of Sciences, Engineering, and Medicine responds by convening a body of experts with diverse interests and expertise to undertake this task, with a resulting report that advises the agencies about how to best deploy resources to advance knowledge in these areas. This survey's key objective is to map the national and international scientific landscape and to chart a path for investment, identifying programs with transformational scientific potential and new observational capabilities. Also central to the survey's charge is to assess the health of the profession and the balance of investments in the people and scientific infrastructure crucial to advancing the understanding of the cosmos. This report lays out a strategy for federal investments aimed at paving a pathway from the foundations of the profession to the bold scientific frontiers.

This chapter provides an integrated view of the strategy, analysis, and advice contained in Chapters 2-7. It is not a comprehensive summary of the report, but rather describes the recommended program in the broader context and framework in which this decadal survey was conducted, articulating the approach for building a scientifically broad, balanced, sustainable program that seizes the opportunities before us.

1.1 THE SCIENTIFIC OPPORTUNITIES

We stand on the threshold of new endeavors that will transform not only our understanding of the universe and the processes and physical paradigms that govern it, but also humanity's place in it. The tremendous richness of 21st century astrophysics is evident in the 573 science white papers authored by more than 4,500 individuals that lay out a wide array of questions we are now poised to answer. Six expert science panels formulated these into key science questions and discovery areas ripe for rapid progress in the coming decade.

Three broad themes, described in Chapter 2, encompass these opportunities—*Worlds and Suns in Context*, *New Messengers and New Physics*, and *Cosmic Ecosystems.* The diversity of the science and observational techniques used to advance the associated goals is striking. Because of the balanced and varied programs put forward by prior decadal surveys, small telescopes, inventive experiments, and competed missions operating across the spectrum have harnessed the creativity and technical ingenuity of the community, resulting in an intensely dynamic and rapidly evolving enterprise. As a result, many of the questions at the forefront of the survey's themes could not have been framed even a decade ago. The richness of these three themes demands that a broad and varied suite of capabilities be sustained over the full electromagnetic spectrum and in the new windows of gravitational waves and high-energy neutrinos. Within each overarching theme, with its multiple science objectives, the survey identifies a priority science area that captures the most transformative and far-reaching goal, where, given new, ambitious facilities, we are poised to take giant strides forward.

1.1.1 Worlds and Suns in Context

The science theme of *Worlds and Suns in Context* captures the quest to understand the interconnected systems of stars and the worlds orbiting them, tracing them from the nascent disks of dust and gas from which they form, through the formation and evolution of the vast array of extrasolar planetary systems so wildly different than the one in which Earth resides. This is an area where advances over the past decade have been stunning, and progress in the next decade will be similarly rapid. By 2020, just 25 years after the discovery of the first exoplanet, the inventory of known exoplanets had exceeded 4,000, with more being identified nearly every week, thanks to ground-based radial velocity measurements and surveys of systems where the exoplanet partially eclipses its star (transit surveys), as well as dedicated space missions. The Kepler Discovery-class mission,[1] launched in 2009, revolutionized exoplanet studies by monitoring more than 150,000 stars to detect thousands of transiting planets, enabling astronomers to explore the structure and vast diversity of planetary systems for the first time. Combining Kepler's data with ground-based radial velocity measurements is providing essential information on exoplanet masses and densities. The Transiting Exoplanet Survey Satellite (TESS) Explorer-class mission,[2] launched in 2018, is surveying the entire sky to find nearby exoplanets, thereby providing the best sample for detailed follow-up studies using current and future ground- and space-based facilities. These same missions, along with the European Space Agency (ESA) Gaia astrometric and photometric observatory, launched in 2013, and large ground-based spectroscopic surveys have also enabled great leaps in the understanding of the physics of stars, the stellar populations of stars of the Milky Way, and the Milky Way's formation history.

The astronomical community and the public alike have been galvanized by the extraordinary progress in detecting and studying exoplanets. The 2018 National Academies report *Exoplanet Science Strategy*[3] captures this progress in rich detail. For the coming decade, key goals include applying spectroscopic and photometric observations to characterize exoplanet surfaces and atmospheres, and fully characterizing not only individual planets but also the properties of entire extrasolar planetary systems. The past decade has revealed how diverse and often different these are from our own solar system. But far more is needed to reliably assess the relative numbers of different system architectures. The upcoming Nancy Grace Roman Space Telescope, with launch expected in 2026, will conduct a microlensing survey

---

[1] The Discovery Program is a series of small to medium-sized competed solar system exploration missions funded by NASA Planetary Science Division.

[2] The Astrophysics Explorer Program is a series of small to medium-sized competed missions.

[3] National Academies of Sciences, Engineering, and Medicine, 2018, *Exoplanet Science Strategy*, Washington, DC: The National Academies Press, https://doi.org/10.17226/25187.

of the Milky Way's galactic bulge, filling out the census by finding exoplanets in the outer reaches of planetary systems that are inaccessible by other detection techniques. Ground-based 6–10 m optical and infrared telescopes with custom instrumentation will continue to broaden demographic samples and diagnose their properties. For the study of atmospheres of exoplanets in close-in orbits, as were found in abundance by Kepler and TESS, spectroscopy with the James Webb Space Telescope (JWST), to be launched by the end of 2021, will be transformational. Millimeter, radio, and infrared observations of the gas and dust disks of forming protoplanetary systems are providing complementary clues to the factors shaping the extent and architectures of solar systems, and this is an area of great discovery potential.

A rich agenda of discovery and scientific opportunity also lies ahead for stellar astrophysics. Over the coming decade, attention will focus on the most important unanswered questions, including understanding the effects of stellar multiplicity on the evolution of the stars in the system, the nature of stellar activity and activity cycles, and reconstructing the formation and assembly of the Milky Way as derived from its ancient stars. Precise distances, until recently only available for ~100,000 stars, are now available for hundreds of millions of stars, along with high-precision photometry, thanks to the ESA Gaia mission. With such a large sample, even rare types of stars and short-lived stellar evolutionary stages are well represented. At the same time, precision time-domain measurements of thousands of stars from Kepler, TESS, and the National Centre for Space Studies (CNES)/ESA Convection, Rotation and planetary Transits mission (CoRoT) have provided detailed asteroseismological measurements of their oscillations, which, like seismic measurements on Earth, unveil the internal structures and motions of material. Ground-based spectroscopy of the stars measured by the space missions will be crucial to obtain orbital velocities, chemical compositions, surface gravities, masses, rotation rates, and other fundamental properties. Spectroscopic survey telescopes in the 4-10 m class capable of observing thousands of stars simultaneously promise major advances. Finally, the Daniel K. Inouye Solar Telescope (DKIST) will revolutionize observations of the Sun's atmosphere.

Priority Area: Pathways to Habitable Worlds

Over the past two decades, thousands of extrasolar planets have been discovered, almost all of them extremely different from any world in our own solar system. This decadal survey's science theme of *Worlds and Suns in Context* encompasses the interlinked studies of stars, planetary systems, and the solar system. Within this broader science theme, the survey has identified the priority science area of Pathways to Habitable Worlds with the goal of trying to discover worlds that could resemble Earth and answer the fundamental question: "Are we alone?" Such planets will be found in the "habitable zone" of their parent stars—not too close and hot and not too distant and cold. Measurements indicate that around 30 percent of stars possess such a planet. The task for the next decades will be finding the easiest of such planets to characterize, and then studying them in detail, searching for signatures of life.

Life on Earth has profoundly altered the planet's atmosphere (Figure 1.1). Interpreting such "biosignatures" is not simple, but the interplay of atmospheric components such as water, oxygen, methane, and carbon dioxide can be modeled to search for evidence of life on other planets. Astronomers have already demonstrated the ability to use spectroscopy to study the atmospheres of large, hot worlds; with future facilities, the same techniques will measure the composition of small, habitable planets.

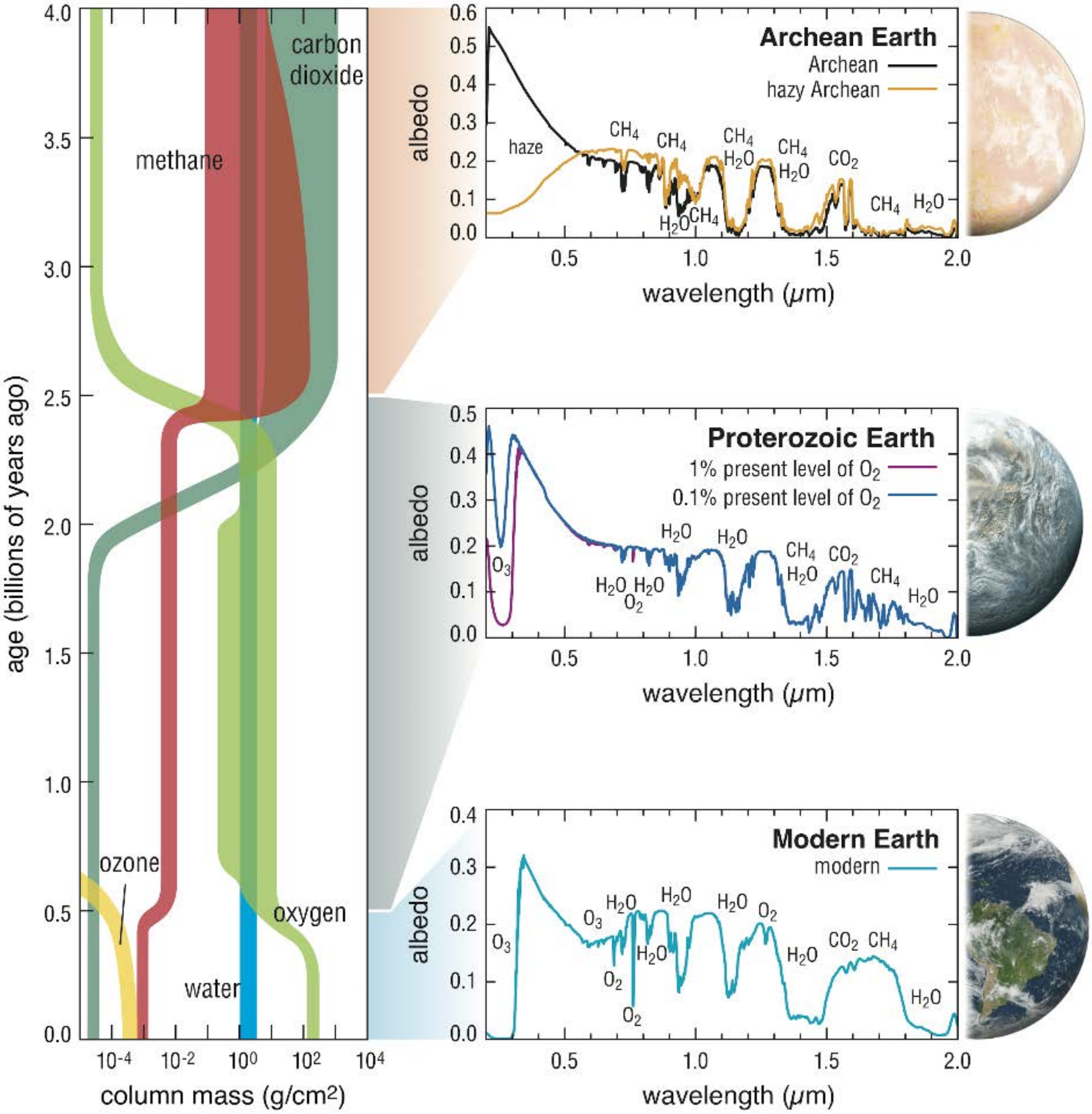

FIGURE 1.1 Evolution of the reflectivity spectrum of Earth. Simulated spectra of Earth before life had significantly altered its atmosphere (*top, Archean era 2.5 to 5 Gyr ago*), before the development of complex life (*middle, Proterozoic era from 0.54 to 2.5 Gyr ago*), and the modern oxygen-bearing Earth (*bottom*). SOURCE: LUVOIR Report; G. Arney, S. Domagal-Goldman, T. B. Griswold (NASA GSFC).

The pathway to searching for biosignatures on habitable worlds depends strongly on the properties of their parent stars. The most common stars in the Milky Way Galaxy are dim, red "M dwarfs." Their habitable zone will be very close to the star, making the systems accessible with the transit technique. JWST will observe a few of the very best target systems. To expand that sample will require the spectroscopic sensitivity of ground-based 25-40 m extremely large telescopes (ELTs).

However, M dwarf stars may not be the best harbor for life—they have massive super-flares and intense, potentially life-destroying energetic emissions. The planets around more placid Sun-like stars are essentially inaccessible to the transit technique and beyond the reach of ELTs, which must observe through Earth's atmosphere. Only an ultra-stable, space-based telescope equipped to block the star's light and directly image the planet can reach this level of sensitivity. The larger the telescope, the larger the number of stars whose planets can be searched for signatures of life.

Properly interpreting these observations will also require a scientific context—understanding the formation and history of these planetary systems to see how life-enabling chemicals flow onto worlds, laboratory studies and simulations of planetary atmospheres, and deeper knowledge of the stars themselves—driving a large part of the overall *Worlds and Suns in Context* theme.

Key capabilities required on the pathway to habitable worlds include the following:

- Ground-based extremely large telescopes equipped with high-resolution spectroscopy, high-performance adaptive optics, and high-contrast imaging;
- A large, stable, space-based infrared/optical/ultraviolet (IR/O/UV) telescope with high-contrast imaging capable of observing planets 10 billion times fainter than their star, and UV, visible, and near-IR exoplanet spectroscopic capabilities;
- A high spatial and spectral resolution X-ray space observatory to probe stellar activity across the entire range of stellar types, including host stars of potentially life-sustaining exoplanets; and
- Laboratory and theoretical studies of planetary formation, evolution, and atmospheres.

Life on Earth may be the result of a common process, or it may require such an unusual set of circumstances that we are the only living beings within our part of the galaxy, or even in the universe. Either answer is profound. If planets like Earth are rare, our own world becomes even more precious. If we do discover the signature of life in another planetary system, it will change our place in the universe in a way not seen since the days of Copernicus—placing Earth among a community and continuum of worlds. The coming decades will set humanity down a path to determine whether we are alone.

1.1.2 New Messengers and New Physics

Our understanding of the universe has been repeatedly transformed by looking at the sky in new ways, from exploiting the full range of electromagnetic phenomena, to making large-scale, high-cadence astronomical movies, to exploring the universe in non-electromagnetic messengers. This has led to remarkable progress in astronomy over the past century, including the ever-growing impact of astronomy on basic physics. The *New Messengers and New Physics* theme captures the key scientific questions associated with a broad range of inquiries, from astronomical constraints on the nature of dark matter and dark energy, to the new astrophysics enabled by combined observations with particles, neutrinos, gravitational waves, and light.

The unknown physical natures of dark matter and dark energy, both discovered through astronomical measurements, remain outstanding grand challenges in both physics and astronomy, and great observational progress will be made in the coming decade. Addressing these profound mysteries were prime motivations for the Roman Space Telescope, with a field of view 100 times that of the Hubble Space Telescope (HST); the NSF/DOE Vera C. Rubin Observatory, a wide-field 8.4 m telescope devoted to a decade-long mapping of the entire southern sky; as well as ESA's Euclid mission, with a planned launch in 2022. These telescopes are all poised to address the nature of dark energy through large optical and infrared surveys aimed at measuring the distribution of galaxies on large scales, and by detecting distant supernovae. These measurements will also provide a lasting astronomical legacy, with data that can be mined to answer a variety of foundational astronomical questions. High-sensitivity and wide-angle mapping of the cosmic microwave background (CMB) has the potential to create virtual 3D tomographic maps of the matter distribution between the young universe—when there were free electrons that could readily scatter CMB photons—and Earth. These measurements can also be used to map the cosmic structure by mass (rather than by light, which is the structure traced by Roman, Rubin, and Euclid through the starlight of galaxies). Comparing the two will reveal vital information about the structure itself, its evolution, and the evolution of differences between the distribution of light and mass.

In the past decade, a new, perplexing inconsistency between the expansion rates of the universe (Hubble constant) measured from nearby stellar distance ladders versus the CMB and other cosmological yardsticks has also emerged. The latter could be an observational issue, but it could also conceivably point to a missing element of physics in the current cosmological model. New measurements of the Hubble constant made by combining gravitational wave signals with associated redshift measurements will be an entirely independent way to resolve (or confirm) this tension.

The power of near-continuous monitoring of large regions of the sky in the X-ray, gamma-ray, optical, infrared, and radio bands has been dramatically demonstrated over the past two decades. Time-domain astronomy is now a mature field central to many astrophysical inquiries, from diagnosing the wide array of stellar explosions, to exoplanet detection, probing stellar structure, and measuring dramatic and unexplained changes in the appearance of active galactic nuclei—the regions closely surrounding supermassive black holes. New phenomena such as fast radio bursts, besides being events of mysterious origin, can provide a means of probing the tenuous gas in and in between galaxies. Progress in this subject will accelerate dramatically with the commissioning of the Rubin Observatory and later in the decade with the launch of the Roman Space Telescope. In particular, Rubin's unique time domain mapping of the southern sky is expected to detect roughly 10 million variable events per night, providing optical color information necessary for rapid characterization and unique scientific inquiries.

The ground-based Laser Interferometer Gravitational Wave Observatory's (LIGO) discovery in 2015 of gravitational waves from a pair of merging 30 solar mass black holes is certainly one of the watershed moments in physics and astronomy of the last decades. Future upgrades of LIGO, the European Virgo interferometer, and the Japanese Kagra, together with the launch of the Laser Interferometer Space Antenna (LISA) low-frequency gravitational wave observatory in the early 2030s have tremendous promise to answer fundamental questions in physics and astronomy and to open vast new discovery space. Upgrades of NSF's IceCube high-energy neutrino detector will enable these nearly massless subatomic particles to be associated with individual astrophysical objects, probing extreme environments where particles are accelerated to near-light speeds. The recent addition of the entirely new messengers—gravitational waves and high-energy neutrinos—to time domain astrophysics provides the motivation for the survey's priority science theme within *New Messengers and New Physics*.

Priority Area: New Windows on the Dynamic Universe

This report's science theme of *New Messengers and New Physics* captures the broad array of science made possible by observing the sky in new ways. Within this theme, the decadal survey has identified the priority science area of "New Windows on the Dynamic Universe"—the study of neutron stars, white dwarfs, collisions of black holes, and stellar explosions using the complementary perspectives provided by the wide range of messengers from light in all its forms from radio to gamma rays, gravitational waves, neutrinos, and high-energy particles. In parallel to remarkable advances in observations with multiple messengers from the LIGO/Virgo/Kagra gravitational wave and the IceCube high-energy neutrino observatories, the combination of large detectors, big data, and software advances for handling that data continues to transform the previously static view of the sky to one with nearly daily movies. Future upgrades of ground-based gravitational wave facilities, together with the launch of LISA make this a high priority for discovering new physics, and making astronomical measurements that will change paradigms.

Just like our everyday experience benefits from combining the information provided by sight, sound, taste, and smell, so too observations with these complementary messengers open new ways of doing astronomy and new ways of testing the laws of physics. This will reshape the understanding of topics as diverse as the origin of the carbon in bones and the metal in phones, the history of the expansion of the universe since the Big Bang, the life and death of stars, and the physics of black hole event horizons.

New, coordinated advances in several areas are required to unlock the workings of the dynamic universe. These include the following:

- A suite of small and medium-scale ground and space-based observational facilities across the electromagnetic spectrum to discover and characterize the brightness and spectra of transient sources as they appear and fade away.
- Ground-based 20-40 m optical-infrared telescopes and an IR/O/UV space telescope significantly larger than HST to see the light coincident with colliding neutron stars detected in gravitational waves—most of these are sufficiently distant to be undetectable with current facilities. These same telescopes will diagnose in exquisite detail the elements produced in stellar explosions.
- A sensitive next-generation radio observatory more powerful than the Very Large Array (VLA) to detect the jets of relativistic gas produced by neutron stars and black holes, including those in mergers observed by ground and space-based gravitational wave facilities.
- Next-generation CMB telescopes to search for the polarization signatures of gravitational waves produced in the infant universe.
- Upgrades to improve the sensitivity of current ground-based gravitational wave detectors, and development of technologies to enable next-generation facilities.
- Improvements in the sensitivity and angular resolution of high energy neutrino observatories.
- Strong software and theoretical foundations to numerically interpret the gravitational wave signals from merging compact objects to extract new physics in the extremes of density and gravity, and ensure easy user access to the wealth of data on the dynamic universe and to model and interpret astronomical sources whose physical conditions cannot be replicated in laboratories on Earth.

1.1.3 Cosmic Ecosystems

The universe is characterized by an enormous range of physical scales and hierarchy in structure, from stars and planetary systems to galaxies and a cosmological web of complex filaments connecting them. A major advance in recent years has been the realization that the physical processes taking place on all scales are intimately interconnected, and that the universe and all its constituent systems are part of a constantly evolving ecosystem. The seeds of galaxies were planted during the first moments of the Big Bang, and modern numerical simulations trace the gravitational growth of cosmic structure from 300,000 years after the Big Bang to the structures and galaxies seen today. The galaxies are ecosystems of their own, with further condensation of matter to form stars and planets balanced by “feedback” from stellar winds, outflows, and supernovae that return mass and energy to the gaseous environment. The supermassive black holes that form and grow within nearly all massive galaxies also play a key role in this feedback process. Unraveling the nature of this connection is one of the key science goals of the decade.

The time is ripe for major breakthroughs. JWST will provide definitive observations of the earliest stages of galaxy formation and evolution, and the histories of star formation, chemical enrichment, and feedback processes over cosmic time. The combination of wide-area observations of distant galaxies by the Rubin Observatory, Roman, and Euclid will provide imaging and spectral energy information for millions of galaxies, complementing the in-depth observations from JWST and HST.

The upcoming observations with JWST, the Rubin Observatory, and Roman will be profound but will not on their own be able to address the central problem of understanding how galaxies grow. Probing the heart of the galactic feedback process requires detecting and measuring the tenuous gases at the boundaries of galaxies and their intergalactic surroundings, the circumgalactic medium (CGM), where the accretion and recycling of gas and metals from feedback processes take place. This goal motivates the priority area within *Cosmic Ecosystems*.

Priority Area: Unveiling the Drivers of Galaxy Growth

Processes on a wide range of time and length scales shape the behavior of most astronomical objects, from the scales of planet formation in disks around young stars to galaxies and clusters of galaxies. The science theme of *Cosmic Ecosystems* captures the interconnectedness of these astronomical systems across cosmic scales. Within this theme, the decadal survey has identified the priority science area of Unveiling the Drivers of Galaxy Growth. The allure of galaxies—to scientists and the public alike—stems from their diversity and complexity. Their rich internal structures and tremendous variety make understanding the origin of galaxies one of the most continuously compelling stories in astronomy. The past decades have seen a growing understanding of the origin of this complexity: gas flows into galaxies, fueling new generations of stars and the buildup of central black holes, but these same stars and black holes send matter back out, potentially shutting down any chance for new material to stream in. These processes must have profound effects on galaxies, but astronomers have only a tenuous grasp on the full coupling between the larger galaxy environment that holds the gas transiting in and out of a galaxy, and the properties of the galaxy itself.

This profoundly multiscale problem requires connecting galaxies from their central black holes, in a region no larger than the solar system, to their outermost reaches millions of light years from the center. Technologically, these demanding requirements drive investments in reaching high resolution—to uncover the parsec-scale astrophysics powering feedback—and towards high sensitivity—to both detect the most tenuous and diffuse emission and to allow spectroscopy against faint background light sources with sufficient density to sample a dozen or more lines of sight in a single galaxy. Furthermore, the range of gas temperatures (from more than a million degrees kelvin down to temperatures approaching absolute zero) and redshifts naturally motivates a multiwavelength approach.

New observational capabilities across the electromagnetic spectrum along with computation and theory are needed to resolve the rich workings of galaxies on all scales. These include the following:

- Large 20–40 m ground-based O/IR telescopes to observe the transition-rich rest-frame UV, in both emission and absorption, for galaxies in the young universe. This will reveal the faint networks of gas that surround galaxies and the gas’s chemistry, temperature, density, and motions.
- A next-generation VLA radio telescope will, for the same early epochs, map emission lines of molecular gas, tracing the cold gas associated with both the extended galactic environment and fueling AGN and star formation within the galaxy itself.
- A next-generation IR/O/UV large space telescope to trace much of the same physics as the ELTs but in the nearby, evolved universe, and in dramatic detail, revealing the full multiphase complexity of the local ecosystem.
- To complement these capabilities a capable far-IR and/or X-ray mission will further transform these views by peering into the dusty hearts of galaxies to reveal enshrouded accreting black holes, or tracing the hottest phases of gas driven outward by this same accretion, with the spatial and spectral resolution needed to isolate critical physical quantities in massive galaxies.
- Investments in theory and in the community of scientific experts exploring these data are essential for synthesizing a new scientific foundation for galaxy evolution from these observational advances.

1.2 BACKGROUND AND CONTEXT

The range and variety of compelling scientific opportunities illustrates the dynamic nature of modern astrophysics, with future directions propelled both by steady evolution and by dramatic revolution, powered by new discoveries, emerging capabilities, and an increasingly diverse set of ideas. The survey's recommended program is driven by the science, but it is also shaped by the global landscape, and the scientific, technical, and human context of the times. Multifaceted considerations led to the balance of science, the emphasis on sustainable investments in projects and people, and the wide-ranging activities on all scales that are prioritized through recommendations in this report.

The major scientific progress in astronomy over the past decade has been mirrored by a continued transformation in the national and international landscape in which this research is being conducted. Astronomy continues to become more global and interconnected, and many of the major space missions in recent decades (HST, JWST, Herschel Space Observatory, and Planck) have been carried out as partnerships between NASA, ESA, and/or the Japan Aerospace Exploration Agency (JAXA). With the XRISM and Athena X-ray observatories, Euclid, and LISA on the horizon, the survey's scientific goals are crucially dependent on such partnerships continuing and even strengthening going forward. On the ground, the Gemini and ALMA ground-based observatories are international collaborations with NSF participation. This trend is likely to continue; a majority of the large ground-based projects presented to this survey have, or plan to have, significant international partners. Data produced by other European-led observatories such as the ESA Gaia mission and the European Southern Observatory have also contributed to major advances by U.S. researchers, either individually or as members of international collaborations. This international context of current and planned facilities has been fully incorporated into the survey's science and strategy planning.

The imminent launch of JWST is a momentous occasion that will shape the course of astronomy and astrophysics in the coming decades. Arguably the most ambitious robotic science mission that NASA has ever undertaken, JWST will influence essentially every area of astronomy, from peering back in time to view nascent galaxies as they begin to form in the early universe, to exploring the atmospheres of exoplanets in exquisite detail. JWST, more than two decades in the making, reminds us of the transformational nature of the ambitious, large strategic missions that NASA is uniquely capable of undertaking.

While large strategic missions are transformative, 21st century astrophysics owes much of its richness to NASA's panchromatic suite of Great Observatories that spanned the spectrum from gamma rays to infrared, and which were accomplished with a wide range of scales, from what today is referred to as "Probe scale" up to the very ambitious HST and JWST missions. Diverse missions of all scales, national and international, designed to view the universe in a multiplicity of complementary ways are now essential to progress in modern astrophysics. The broad science laid out in this report requires a wide variety of space-based techniques and capabilities spanning not just the electromagnetic spectrum, but, with the launch of ESA's LISA mission, in which the United States is a significant partner, the gravitational wave spectrum as well. While, as noted above, sustaining broad observational capabilities is crucially dependent on international partnerships and missions, essential capabilities, such as very high-contrast imaging and spectroscopy in the IR/O/UV bands, far-IR imaging and spectroscopy, and high-resolution X-ray imaging and spectroscopy, are not planned in ESA's Voyage 2050 program,[4] or by other international agencies. Because of the significant U.S. leadership in the development of the needed technologies and capabilities, and the high priority these have for this survey, it is essential for NASA to lead their development. However, without a major change in the approach to developing strategic missions, combined with expanding the range of mission scales, it will take many decades to realize the necessary range of observational capabilities.

---

[4] See https://www.esa.int/Science_Exploration/Space_Science/Voyage_2050_sets_sail_ESA_chooses_future_science_mission_themes.

On the ground, the astronomical community eagerly awaits the commissioning of the Rubin Observatory, which will be devoted to a decade-long mapping of the entire southern sky in multiple colors and with multiple time-domain cadences. By harnessing the power of the digital revolution, and building on past large surveys, large public data sets, data science, and computational astrophysics, Rubin will leave a legacy of data that will be mined far into the future by a diversity of astronomers. The challenge going forward is to ensure that the vast range of variable and transient phenomena that Rubin will uncover can be quickly discovered and studied by facilities spanning the wavelength spectrum.

Concerning new ground-based activities, NSF and DOE strongly urged the survey to be ambitious and challenged it to consider bold, transformative initiatives. At the same time, NSF Division of Astronomical Sciences (NSF AST) is faced with an historic underinvestment in smaller scale, foundational activities such as the general investigator grants that ensure high scientific return from projects of all scales. Together with the lack of a sustainable model for operating new facilities, the agency faces structural issues it must address to capitalize on the opportunities. Nationally, attention is growing on the country's urgent need to build its infrastructure, technological base, and scientific foundations, and this movement aligns well with NSF AST's needs. Being a field that captures the imagination of the public, pushes technology, and is a gateway to science, technology, engineering, and mathematics (STEM) education, astronomy is in a good position to argue for addressing these foundational issues through increased basic investments.

The activities and deliberations of this decadal survey took place in a time of tremendous national and international upheaval. The global COVID-19 pandemic has disrupted every aspect of life, from seemingly mundane issues of how to conduct the survey's business, to health, childcare, elder care, and education. The impacts have not been equally felt by women and men, and they also depend on socioeconomic status and race. The careers of many young people, including scientists, have been paused, and this will have a lasting impact on the profession. The pandemic also strongly underscored the important role of science, and scientific reasoning in combatting the epidemic, from the rapid development of mnRNA vaccines, to the factual, analytic presentation of the data necessary to design protective measures. The ultimate economic and social impacts of the pandemic remain unclear, adding to the uncertainty of the future landscape.

As a final, important backdrop, this survey was strongly influenced by the urgent need to advance diversity, equality, and inclusion in all aspects of society. This need came into sharp focus with the Black Lives Matter movement, sexual harassment and the inequalities highlighted by the #MeToo movement, the inequitable impacts of COVID-19, and the shocking increase in hate crimes against Asian Americans. These harsh realities have invigorated the nation into a renewed conviction to tackle systemic issues of race, gender bias, and privilege at a local and global scale. There is momentum to effect change, and the time is overdue to actively focus on these activities. Changing the defaults under which astronomy is practiced will only happen with energetic engagement and a diversity-, equity-, and inclusion-focused lens.

1.3 FRAMEWORK FOR THE SURVEY'S RECOMMENDATIONS

In the context described above, the decadal survey committee weighed many considerations in designing its recommended program (Figure 1.2). Primary among these is that the portfolio must be scientifically balanced, broad, and sustainable. Also, the program must be structured to draw from the widest range of human talent. The first consideration drives the need for a balance of investments among activities that lay the foundations of the science and the profession, and that advance a variety of projects on all scales. The second consideration requires that the profession and the agencies nurture, structure, and expand programs in such a way that they eliminate barriers, create inclusive environments, and actively encourage broad participation.

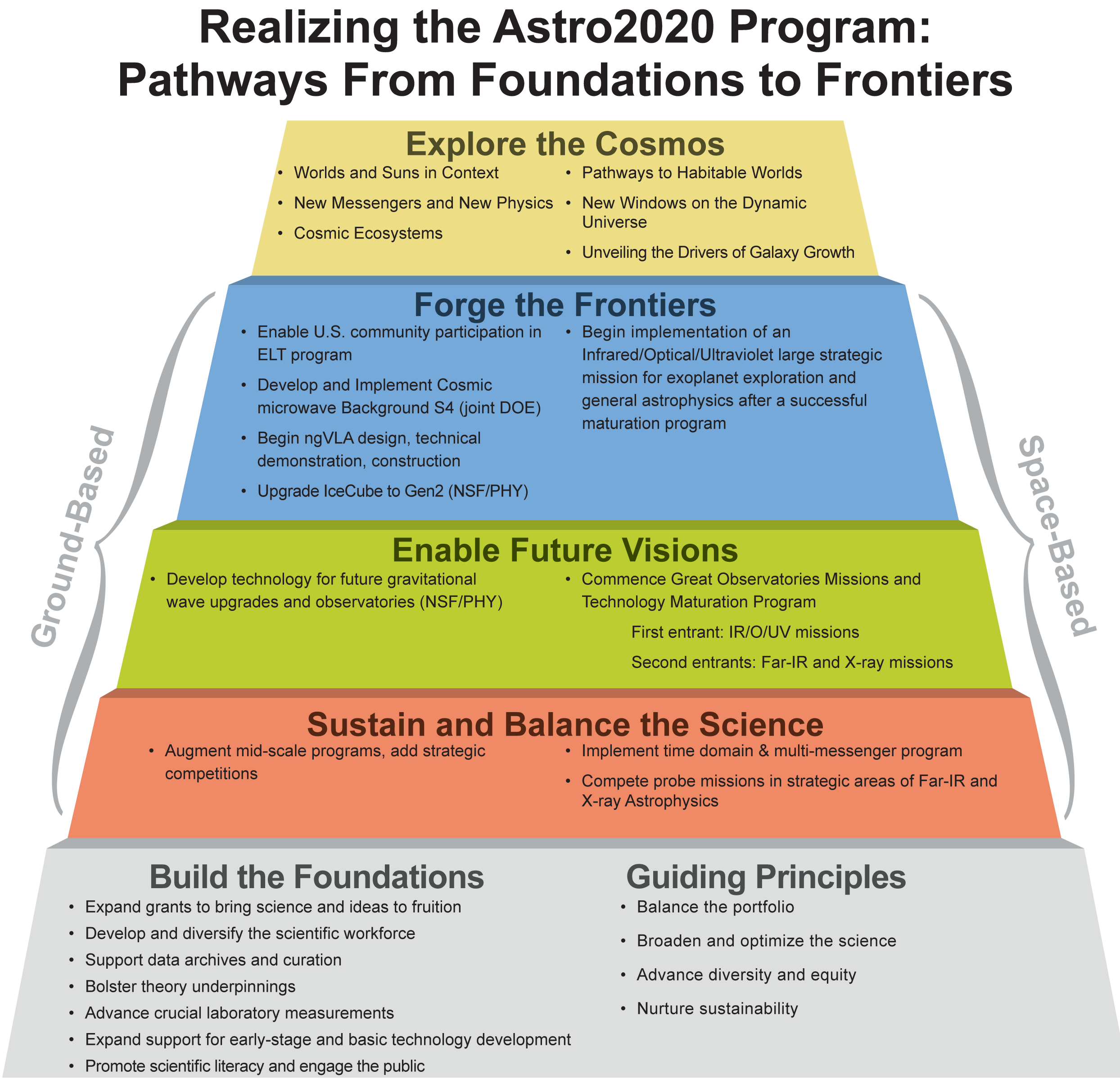

FIGURE 1.2 The recommended program includes elements that pave the way to transformative science by building a strong research and technology foundation, promoting programs on a range of scales that balance and sustain observational capabilities, enabling future large projects, and advancing new frontier observatories.

The survey's organization of projects and activities into categories is a departure from past practice. It emphasizes the function of the activity within the program rather than the cost, although there is a rough equivalence. Prior surveys have divided programs strictly by budgetary requirements (small, medium, and large) and have in general not prioritized projects in one cost category compared to those in another as a means of emphasizing the need for balance. The approach taken by Astro2020 is to adopt functional categories. Projects that build the foundations consist in large part of competed grants to individual investigators and programs that support modest scale activities, and sustaining projects consist of competed mid-scale experiments and missions. Two categories capture the large, ambitious initiatives: programs that enable future visions and those that realize frontier facilities. This survey, like its predecessors, places strong emphasis on balance and the need for projects on a variety of scales and does not prioritize one category over another. In most categories, the survey identifies the highest-priority activity, and for others there is a natural time ordering based on scientific urgency and/or project readiness.

Another consideration is the budget uncertainty associated both with agency guidance, and with the landscape of federal funding discussed above. NASA and NSF strongly urged the survey to develop a program that is aspirational and inspirational, but that also conforms to budgetary norms. Given the uncertain landscape, the survey committee concluded that it is not possible to imagine and plan for the many possible contingencies. Rather, the recommended program forwards the frontiers of science through ambitious projects, and at the same time strongly advocates for balance. With this guidance in hand, the agencies have the flexibility to seize opportunities that arise on all scales, and the strong motivation to do so given the analysis of this report. For major projects that dominate budgetary requirements, the survey establishes decision rules and off ramps that guide agencies in the event of technical issues, or changes in the budgetary landscape. Interim advice from the mid-decadal survey, and from committees such as the Space Studies Board, the Board on Physics and Astronomy, and the Committee on Astronomy and Astrophysics, are an effective means for the agencies to request input on issues resulting from changing circumstances. These corrections would of course be strongly guided by, and be based on, the full analysis contained in this report. Additional prescriptions are unlikely to be helpful to the agencies given the many constraints, fiscal, political, and organizational, that they are faced with.

The greatest challenge faced by the survey committee in developing new recommendations for the nation's space astrophysics program is how to realize large strategic missions, yet at the same time achieve the wavelength balance, and the overlapping operational lifetimes that characterized NASA's Great Observatories, a model that so successfully propelled many, varied fields of astrophysics. While international partnerships are essential, they are not sufficient to accomplish the broad and aspirational science program laid out in Astro2020. Doing so will require a range of missions significantly larger than Explorers, yet with a mix of cost and implementation time scales spanning from a less than a decade to the multiple decades required to realize a mission of the ambition and complexity of JWST.

As evidenced by the four Large Mission Concept Studies prepared for this survey, the community's most ambitious and visionary ideas now require timelines that are pan-decadal, and even multi-generational (Chapter 7, Table 7.4). We are poised to tackle some questions that are so grand that the facilities and instruments needed to address them require vision and commitment beyond our individual horizons. But to do this sustainably, and to realize the broad capabilities demanded by the richness of the science requires a re-imagining of the ways in which large missions are developed and implemented. The ambitious strategic missions demand much more significant early investments in co-maturing mission concepts and technologies prior to adoption, with appropriate decadal input on scope, and with checks and course corrections along the way. In addition, adding a competed probe mission line that spans the large gap between Explorers and ambitious strategic missions, with science foci identified by decadal surveys will be a further move toward a capable, panchromatic mission fleet.

The greatest challenge for NSF going forward is its need to develop the appropriate programmatic balance of projects spanning the needed range of budgetary levels required to optimize the return on federal investments. Seizing the compelling scientific opportunities, and retaining U.S. competitiveness in astronomy requires capabilities that are uniquely provided by large, ambitious facilities. However, it also requires supporting operations of NSF's wide range of productive facilities, including upgrading instrumentation, ensuring a balance of project scales, and most importantly supporting the community of individual investigators to realize the scientific goals set out for the decade. The complex challenge associated with achieving this balance has been an impediment to the field for multiple decades, and it must be addressed if we are to reap the scientific rewards going forward.

For NASA, NSF, and DOE, overruns and delays in major projects have historically been a significant threat to improving and maintaining program balance. The survey addresses this in two ways.

First, the recommendations in this report emphasize more significant project and technology maturation prior to a commitment to, and commencement of, implementation. This will enable requisite budgets to be more firmly established when the project is adopted by the agency. In addition, for major projects, the decision rules are intended to guide the agencies in how to manage changing circumstances, technical or budgetary.

1.4 CRITERIA FOR ADVANCING NEW ACTIVITIES

The program of new activities in this report was conceived in the context of numerous exciting large strategic projects and missions, including international partnerships, that have yet to begin scientific operation (see Table 7.1 for a comprehensive list). This survey assumes that these compelling programs will be all be completed and sustained through their scientifically productive lifetimes. Ambitious and transformative large-scale efforts often take multiple decades to realize, and all of those scheduled for completion in the coming decade will provide essential capabilities upon which the Survey's scientific goals rely. Further, programs resulting from decadal recommendations, such as NASA's expanded Explorer program and NSF's Mid-Scale Innovations Program, play essential roles in sustaining scientific breadth and ensuring timely response to new opportunities. These continued and future capabilities are essential underpinnings upon which new recommendations are predicated.

For NSF, as noted above, the pressure imposed by operations costs of large NSF facilities on grants and other programs has been a systemic issue plaguing astronomy. By the middle of the decade, this will escalate to unsustainable levels unless changes are made to the way that large facilities are supported. The survey's recommendation is that new, large Major Research Equipment and Facilities Construction (MREFC) recommendations described below be predicated on NSF developing a sustainable plan for supporting the operations and maintenance costs of its astronomical facilities, while preserving an appropriate balance with funding essential scientific foundations and the remainder of the NSF AST portfolio.

1.5 RECOMMENDED PROGRAM OF NEW ACTIVITIES

The survey's recommendations for new programs and program augmentations are organized into steps that form the pathway from the foundations of the profession out to the scientific frontiers (Figure 1.2). The full text of the survey committee's analyses and recommendations is found in Chapters 2-7, while this chapter provides a broad overview. These recommendations advance transformative science in the coming decade and set the stage for enabling the bold visions in the future (Figure 1.3).

1.5.1 Guiding Principles

Major investments must advance a bold and broad scientific vision, while at the same time ensuring a balanced program that responds to scientific opportunity. Astronomy and astrophysics advances in a global context, and the survey recognized and responded to the need for synergy with, and complementarity to, activities worldwide. Especially for ground-based observatories, private institutions and philanthropic entities have been, and continue to be central to some of the most ambitious endeavors. The survey committee carefully considered how to best leverage these private-public partnerships in a way that achieves ambitious science and advances the aspirations of the entire community. There is also the challenging issue of balancing scientific ambition with feasibility and timeliness. All of these factors shaped the recommended programs, and their phasing.

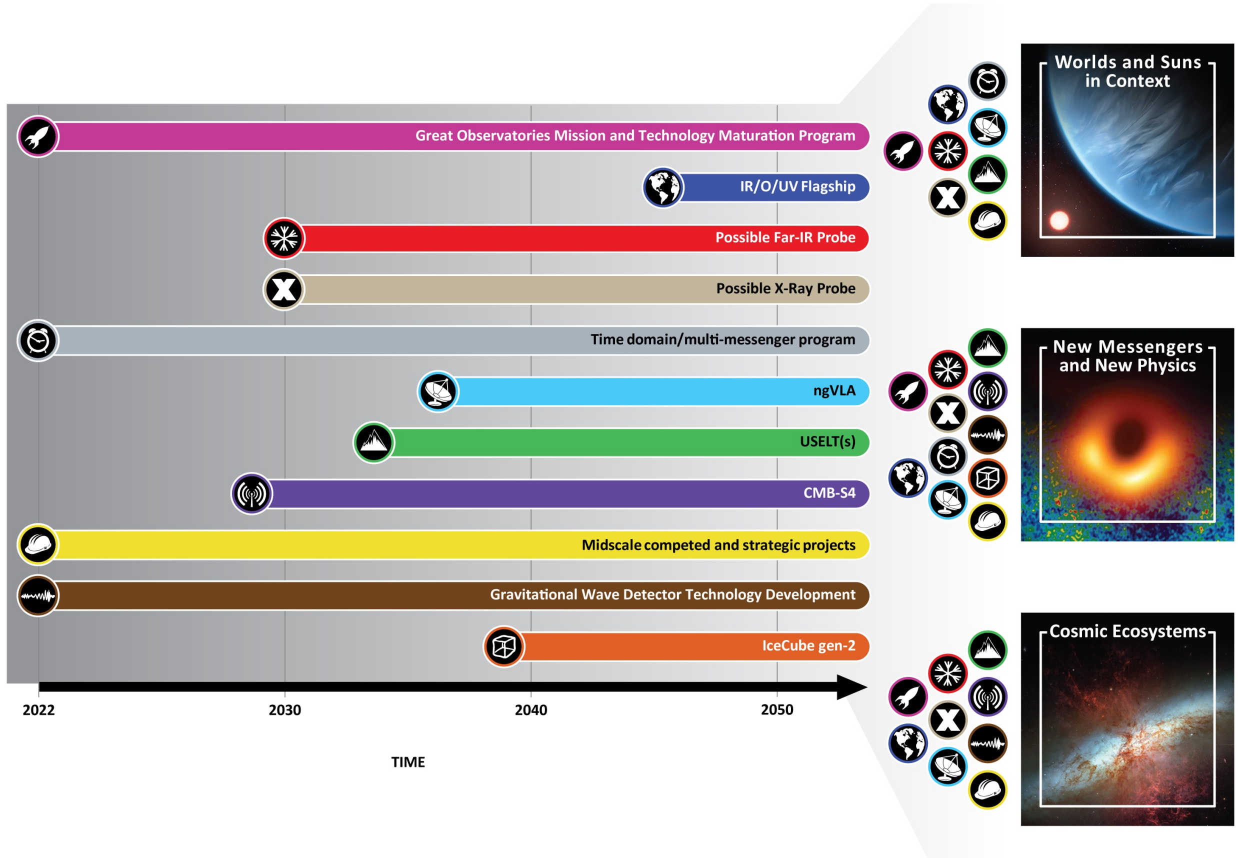

FIGURE 1.3 Timeline for the medium and large programs and projects recommended by this Astro2020 decadal survey. The starting point of each, indicated by the logos, shows the projected start of science operations for missions and observatories, or the start date of the program. The boxes on the right show the survey's three broad science themes, and the placement of the logos to the left of the boxes indicate which activities address the indicated theme. As evidenced in the figure, advancing each of the survey's broad science themes requires a range of facilities and programs.

The recommendations in this report are also guided by the precepts and principles of diversity, equity, benefit to the nation and the world, and sustainability. Diversity is a driver of innovation, and the astronomy and astrophysics enterprise can be at its most innovative only when it maximizes and fully utilizes the diversity of its human talent, ensures equitable access to opportunities, removes barriers to participation, and when it values diverse forms of expertise in its leadership. Equity demands that what is pursued with the nation's resources are done in a manner consistent with the principles of fairness and equal opportunity that are core to society's ideals. Anyone with the ability and the drive to contribute through astronomical discovery should have a fair chance to do so, and be free of fear, harassment, or discrimination. The benefits of astronomy and astrophysics extend beyond its fundamental discoveries. They provide lifelong learning opportunities and science literacy to the public and contribute to the development of the nation's broader, technically trained STEM workforce. In terms of sustainability and accountability, the substantial investments in people and the use of natural resources in astronomy require responsible stewardship, transparency, and accountability for outcomes. This is a core responsibility of the organizations, agencies, and stakeholders that benefit from the human labor and products of the field.

1.5.2 Foundational Activities

The pathway begins with strong support for foundational activities that build the people and the profession, (Chapter 3, The Foundations of the Profession), bolster the core activities necessary for a vibrant research enterprise (Chapter 4, The Research Foundations), and lay the technological foundations for the future (Chapter 6, The Technological Foundations and Small and Medium Programs that Balance the Science). The key new programs with these aims are described below.

- *Develop and diversify the scientific workforce.* The diversity of the astronomy and astrophysics profession remains an area where much improvement is needed. While there have been some notable improvements, especially with regards to the representation of women at the early career ranks of the profession, the overall demographics of the field remain very far from parity with the larger population. Addressing this will require action on many fronts: recommendations in this decadal survey report span the career stages from undergraduate to faculty and beyond, with targeted programs to improve diversity at each level; bridge critical transitions in the pipeline; and work to improve diversity of project teams, participants, and beneficiaries. The ugly realization of continued discrimination in the form of racism, bias, and harassment hampers progress towards building a fully diverse and inclusive workforce, and a recommendation of the report in this area suggests adoption of scientific integrity policies that address discrimination and harassment as forms of research or scientific misconduct. At the core of a diversity-, equity-, and inclusivity-focused approach is the need for data to evaluate equitable outcomes of proposal competitions; such data was sorely lacking in the preparation of this report, and a recommendation to collect, evaluate, and publicly report such data would enable future assessments.
- *Promote scientific literacy and engage the public*. By capturing the public's attention with discoveries, including the participation of citizen scientists in the research process, promoting science literacy, and realizing advanced technologies that can then find real-world applications, astronomy has a clear benefit to the nation. Astronomy education is effective as a broad gateway to STEM careers. Considering the rapidly increasing need for advanced computational skills in both the public and private sector for students to be competitive, embedding computational training in the undergraduate curriculum is even more important to integrate in the coming decade.
- *Promote sustainability and accountability*. The future of the field requires that greater attention be paid to issues of sustainability and accountability, whether it is in the context of the natural resources required for astronomy research activities at observing sites, or the current crisis of a large number of low Earth–orbiting satellites that will impact wide-field imaging at optical wavelengths and radio frequency observations. Adapting to the realities of climate change requires a decrease of the field's impact on carbon emissions. Recognizing the need for active, up-front, and sustained engagement with local and Indigenous communities, the survey committee recommends the implementation of a Community Astronomy model of engagement, similar to community-based approaches in other scientific disciplines. The goal for such an approach is to advance scientific research while also respecting, empowering, and benefitting local communities.
- *Expand the NSF grants program (highest-priority foundational recommendation).* Robust individual investigator grant funding is crucial to achieving the science goals of this decadal survey and to ensure more equitable access to resources. The NSF Astronomy and Astrophysics Grants (AAG) program is a cornerstone of the enabling foundation for research in astronomy and astrophysics in the United States, supporting research projects across nearly all subfields of the astrophysical sciences. This program is not currently at a healthy level, and the recommendation for an augmentation over 5 years is designed to restore success rates

to a healthy competitive environment. This is the highest-priority item amongst the many recommendations for building the foundation of the nation's research enterprise.

- *Bolster theory underpinnings.* Theoretical investigations, crucial as both a mechanism for driving new discoveries and a framework for interpreting essentially all signals received from space, are, like grants at NSF, lacking crucial funding at a level that can sustain the necessary projects. A recommendation to increase the amount of funding for NASA's Astrophysics Theory Program (ATP), and restore its cadence to an annual call, reflects increases to recover from past limited funding.
- *Maximize science from large programs on ground-based facilities.* Another survey recommendation in the foundations category urges NSF to establish a mechanism of research funding and production of high level data products for large principal-investigator programs on MREFC-scale astronomical facilities. This would accelerate scientific output and maximize the timeliness and community impact of large key projects.
- *Support data archives and curation.* Astronomy is evolving rapidly into a profession in which archiving of individual observations can produce scientific impacts that rival the original studies, and large-scale surveys are designed for science-ready archival manipulation from the beginning. As demonstrated by space missions and some ground observatories (e.g., ALMA, the European Southern Observatory [ESO]), readily-accessible archival data can substantially increase the scientific impact of facilities for a relatively modest incremental cost. The situation is less uniform for the large number of ground-based optical/infrared (OIR) facilities managed by universities and other institutions. A survey recommendation to NSF and stakeholders for enabling science-ready data across all general-purpose ground-based observatories is an attempt to ensure that all pipelined observations are archived for eventual public use.
- *Advance crucial laboratory measurements.* Laboratory astrophysics is a critical but often hidden and underappreciated cornerstone of the enabling research foundation. It has been chronically underfunded; concerns were raised in both the 2000 and 2010 decadal surveys, but the problem persists. Research in this area needs to be regarded as a high priority, and the existing approaches are not sufficiently advancing the field. A multi-step recommendation in this area urges the agencies to identify the needs for supporting laboratory data to interpret the results of new astronomical observatories, identify resources, and consider new approaches or programs for building the requisite databases. The recommendation also points out the need to include not only experts in laboratory astrophysics, but also users of the data to identify the highest priority applications.
- *Expand support for early-stage and basic technology development.* Analyses of the needs for basic technology funding to support future innovation, as well as to advance identified goals for, for example, high-contrast imaging, adaptive optics, highly multiplexed detectors, and technologies that will drive the next generation of instruments, observatories and missions, identifies increased investments in basic technology as a priority. Another important factor is that basic technology grants are too small to support infrastructure or significant involvement by industrial partners. To be able to fuel innovative future projects on all scales, it is important for the basic technology development portion of the Astrophysics Research and Analysis Program (APRA) to be significantly increased, and for cuts to NSF's Advanced Technologies and Instrumentation (ATI) program over the last decade be rapidly reversed, and additional funding be added to bring the program to the levels recommended by Astro2010.

1.5.3 Programs that Sustain and Balance the Science

Chapter 7 (The New Medium and Large Investments that Sustain Science and Forge Frontiers) lays out an ambitious roadmap for high-priority, space- and ground-based, large and medium-scale initiatives that are compelling and ready to begin implementation in the decade 2023–2033. This roadmap has at its core recommendations are aimed at capitalizing on the upcoming Roman, Rubin, Athena, and LISA observatories, and balancing scientific progress among the survey's priorities, thereby addressing the extraordinary richness of 21st century astrophysics.

Time Domain Astrophysics Program (Highest Priority Sustaining Activity for Space)

Exploring the cosmos in the multi-messenger and time domains is a key scientific priority for the coming decade, with new capabilities for discovery on the horizon with the Rubin Observatory, Roman, LIGO/Virgo and the Kamioka Gravitational Wave Detector (KAGRA), and IceCube. To advance this science, it is essential to maintain and expand space-based time-domain and follow up facilities in space. Many of the necessary observational capabilities can be realized on Explorer-scale platforms, or possibly somewhat larger. As the international landscape and health of NASA assets change, it will be important for NASA to seek regular advice over the coming decade on needed capabilities and to ensure their development. The open Explorer program calls have reached a healthy funding level, and as noted in Section 6.2.1.1.3, maintaining the current cadence of open calls is a condition for new initiatives. This time-domain program is therefore recommended as an augmentation to those levels, and would be executed through competed calls in broad, identified areas.

Astrophysics Probe Mission Program (Space)

The large gap in cost and capability between medium-class Explorer missions and the large strategic missions presented to the survey is a significant impediment to achieving the broad set of decadal scientific priorities. Institution of Probe-class line of missions with a cost cap of ~$1.5 billion per mission, a cadence of ~one per decade, and competed within selected priority areas identified by this and future decadal surveys, is a crucial addition to NASA's astrophysics portfolio. The two priorities for the first Probe-class mission competition are a far-IR probe or an X-ray probe to complement the Athena mission. Both areas represent important observational needs where advances in technology and focused objectives can yield transformative science on a moderate-sized platform.

Augmentation and Expansion of the NSF Astronomy Mid-Scale Program (Highest Priority Sustaining Activity for Ground)

Mid-scale programs—across the entire range of ~$4 million to 120 million—enable new transformative capabilities by incentivizing creative approaches from the community for cutting-edge instruments and experiments. They also ensure robust capabilities for basic research through continually refreshed instrumentation suites and can respond rapidly to strategic priorities. For these reasons it is essential to expand funding levels for the astronomy funding available through mid-scale programs, MSIP and Mid-scale Research Infrastructure (MSRI). It is also essential to add components to the astronomy mid-scale program to target strategic areas through dedicated calls, and to sustain and advance instrumentation on existing telescopes. For the next 10 years, time-domain astrophysics, highly multiplexed spectroscopy, and radio instrumentation (including radio transient cameras and neutral hydrogen mappers) are the priorities for strategic calls. Dedicated calls are also needed to ensure the regular upgrading of instrumentation on existing facilities, with an emphasis on 4–10 m class optical/infrared telescopes. These two new elements would be added, in addition to entirely open competitions of new ideas, in a balanced way that responds to proposal pressure.

1.5.4 Programs that Enable Future Visions

Great Observatories Mission and Technology Maturation Program (Highest Priority for Enabling Programs for Space)

NASA's flagship missions are driven by transformative scientific visions, and they advance a broad range of scientific objectives. Overlapping or near-simultaneous wavelength coverage is particularly impactful, as evidenced by the success of NASA's Great Observatories. Given the large costs and development timescales associated with the large missions presented to this survey, achieving this will only be possible if a new approach is taken to mission maturation, and in particular phasing it with decadal survey advice. The Great Observatories Mission and Technology Maturation Program is aimed at increasing the cadence of large missions by designating appropriate scope at an early stage and making significant investments in maturing missions to the appropriate level prior to ultimate recommendation and implementation. This motivates the recommendation that a large IR/O/UV mission first enter the maturation program, and only when that has been successful as defined by a review, would it proceed to formulation. It is also important that additional missions enter the maturation program in the next 10 years to ensure the needed cadence for panchromatic capabilities, and the priorities for this are a far-IR flagship with some of the capabilities of the proposed Origins, and a high-resolution X-ray mission with some of the capabilities of the proposed Lynx. An important aspect is that both the X-ray and far-IR missions are to be matured with cost targets of $3 billion to $5 billion. Determining the range of capabilities for these missions will be part of this maturation program, and will be informed by the first Probe mission selection.

Technology Development for Future Gravitational Wave Observatories (Ground)

Gravitational wave astrophysics is one of the most exciting frontiers in science. One of the survey's key priorities is the opening of new windows on the dynamic universe, with gravitational wave detection at the forefront. The continued growth in sensitivity of current-generation facilities, such as LIGO, through phased upgrades and planning the next-generation observatory, such as Cosmic Explorer, is essential. This will require investment in technology development now. The survey committee strongly endorses gravitational wave observations as central to many crucial science objectives. Because the technology development for future upgrades and observatories is funded by NSF Physics, it is beyond the survey's charge to formally recommend this investment.

1.5.5 Large Programs that Forge the Frontiers

A Future Large Infrared/Optical/Ultraviolet Telescope Optimized for Observing Habitable Exoplanets and General Astrophysics (Highest Priority for Space Frontier Missions)

Inspired by the vision of searching for signatures of life on planets outside of our solar system, and by the transformative capability such a telescope would have for a wide range of astrophysics, the priority recommendation in the frontier category for space is a large (~6 m diameter) IR/O/UV telescope with high-contrast (10^{-10}) imaging and spectroscopy. This is an ambitious mission, of a scale comparable to the HST and JWST space telescopes. It is also one that will be revolutionary, and that worldwide only NASA is positioned to lead. A period of mission and technology maturation is required, however with sufficient investment this could be completed before the end of the decade, and the mission could commence formulation prior to 2030. (Section 7.5.2)

Decision Rules: Prior to commencing mission formulation, a successful Great Observatories Mission and Technology Maturation program must be completed, and a review held to assess plans in light of mission budgetary needs and fiscal realities.

The U.S. Extremely Large Telescope Program (Highest Priority in the Ground-Based Frontier Category)

Because of the transformative potential that large (20–40 m) telescopes with diffraction-limited adaptive optics have for astronomy, and because of the readiness of the projects, the survey committee's top recommendation for frontier ground-based observatories is investment in the U.S. ELT program. The U.S. ELT program is made up of three elements: the Giant Magellan Telescope (GMT), the Thirty Meter Telescope (TMT), and NSF's National Optical-Infrared Astronomy Research Laboratory (NOIRLab). The primary mirror of the GMT has a total diameter of 24.5 m and the telescope has a 25 arcmin field-of-view (FOV). The GMT will be located at the Las Campanas Observatory in Chile. The TMT primary mirror has a diameter of 30 m, and the telescope has a 20 arcmin FOV. The TMT will either be sited on Maunakea in Hawaii, or at Roque de los Muchachos Observatory on La Palma in the Canary Islands. These observatories will create enormous opportunities for scientific progress over the coming decades and well beyond, and they will address nearly every important science question across all three priority science themes. Both projects are essential for keeping the U.S. community's global scientific leadership, providing important synergistic capabilities that complement those planned for the European ELT. However, both projects have significant remaining risks primarily associated with the need to raise additional private or international contributions. The success of at least one of these projects is absolutely essential if the United States is to maintain a position as a leader in ground-based astronomy. The objective is to achieve a time share that is equivalent to 25 percent in each telescope. If only one project is viable, then a larger fraction on that telescope is required to meet the survey's scientific goals, with the aim of achieving an NSF share up to 50 percent time in that project. (Section 7.6.1.1)

Decision Rules: Successful completion of an external review that will determine the financial viability of both projects, final site selection (in the case of TMT), development of an appropriate management plan and governance structure, and appropriate plans for public access and data archiving.

The Cosmic Microwave Background Stage 4 Observatory (CMB-S4)

Given technical and scientific progress over the last decades, ground-based studies of the CMB are poised to take a major step forward in the coming decade. The Cosmic Microwave Background Stage 4 (CMB-S4) observatory will leverage this progress and will have broad impact on both cosmology and astrophysics. Realizing the ultimate scientific potential of ground based CMB observations will take an effort far beyond what can be achieved by independently scaling up existing experiments. CMB-S4 observatory, a joint effort of NSF and DOE, is the compelling and timely next leap for ground-based observations. It will conduct a 7-year ultra-deep survey of a few percent of the sky from the South Pole with a combination of large and multiple small aperture telescopes observing from 30-270 GHz. This will be done in parallel with a 7-year deep/wide survey of roughly half the sky with additional telescopes sited in the Atacama desert in Chile. The Survey is also excited by the breadth of science, including time-domain and transient studies, and the potential engagement of a community well beyond traditional CMB cosmologists. To maximize the science, transient alerts and well calibrated maps from all surveys will need to be made available to the entire community in a timely fashion, even if it requires some extra resources to do so. (Section 7.6.1.3)

The Next Generation Very Large Array (ngVLA)

For the past four decades, the Karl Jansky Very Large Array (JVLA) and the Very Long Baseline Array (VLBA) have been the premiere observatories worldwide for accessing the sky at centimeter wavelengths. It is of essential importance to astronomy that the JVLA and VLBA be replaced by an observatory that can achieve roughly an order of magnitude improvement in sensitivity compared to these facilities, with the ability to image radio sources at centimeter to millimeter wavelengths on scales of arcminutes to fractions of a milliarcsecond. The ngVLA is such a facility; however, it is immature in its development, and considerable effort must be put into studies to understand and reduce the cost relative to current estimates, secure international partnerships, and prototype the antennae. With such an effort commencing soon, the ngVLA would be ready to commence construction by about 2030. It will be important to begin implementation as soon as it is technically and fiscally possible. (Section 7.6.1.4)

Decision Rules: Implementation is contingent on a successful design, development and prototyping program, cost studies, and commitments from any foreign partners. A review will determine the project's readiness and consistency with budgetary constraints prior to commencement of construction.

The IceCube-Generation 2 (IceCube-Gen2) Neutrino Observatory

Observations of high-energy neutrinos enable astrophysical advances in the study of some of the most energetic phenomena in the universe. The IceCube-Gen2 would greatly enhance the capabilities relative to IceCube, would be able to resolve the bright, hard-spectrum TeV-PeV diffuse neutrino background into discrete sources, and would make the first detections at higher neutrino energies. Multi-messenger astrophysics is a major theme of this report, and the survey endorses the IceCube-Gen2 observatory as important to many key survey scientific objectives. Because it is funded by NSF Physics, it is beyond the survey's charge to recommend this investment. (Section 7.6.2.3)

1.6 ADDITIONAL ADVICE

In addition to the vision for new, recommended future endeavors, this decadal survey report offers advice on aspects of the agencies' programs aimed at optimizing returns for their existing programs.

Data Archives. An important component of creating effective archives is coordinating with cross-agency and international archiving services to develop best practices and interoperability. As a step toward this, it is important for NASA and NSF to explore mechanisms to improve coordination among U.S. archive centers and to create a centralized nexus for interacting with the international archive communities. The goals of this effort are best defined by the broad scientific needs of the astronomical community.

Solar Physics. Solar physics is directly relevant to astronomy, as it is the study of our nearest star, and interacts with stellar astrophysics; is input to studying the Earth-Sun connection and expanding to stellar-planetary interactions; and is vital to understanding Earth's climate and space weather. The survey committee concluded that an appropriate role for astronomy and astrophysics decadal surveys is to comment on the value of ground-based solar physics projects for astronomy and astrophysics priorities, with the solar and space physics decadal survey being the more appropriate body to prioritize and rank ground-based solar physics projects within the context of the full range of multi-agency activities in solar physics.

NSF Portfolio Reviews. Regular reviews of more mature observatories are essential to determine how to optimize their scientific return and cost effectiveness, and to determine when a facility is at the end of its operational life. While some aspects of ground-based facility reviews are considered as part of the review of operating agreements for observatories, these are not an appropriate substitute for a review that considers the entire portfolio on a self-consistent, holistic basis. It is essential that NSF AST establish a regular cadence of reviews of its operational portfolio, at a frequency sufficient to respond to changes in scientific and strategic priorities in the field. An appropriate target is two reviews per decade.

SOFIA. The survey committee has significant concerns about SOFIA, given its high cost and modest scientific productivity. The NASA portion of SOFIA's operating budget is out of balance with its scientific output, which is a fraction of that of comparable cost missions (e.g., HST, Chandra) and often less than those of Explorer missions. The survey committee finds no evidence that SOFIA could transition to a significantly more productive future and notes the minimal mention of SOFIA science by the science panels. The committee found no path by which SOFIA can significantly increase its scientific output to a degree that is commensurate with its cost and endorses NASA's current plan to discontinue operations in 2023.

NASA's Balloon Program. NASA's balloon program plays an important role in offers access to a near-space environment with a wide variety of options for duration and sky coverage, for developing technologies, and training future generations of technologies and mission leaders. It is, however, clear that the balloon program is not yet achieving the potential promised by the advent of ultra-long duration balloon (ULDB) flight capabilities. It is important that the balloon program be critically reviewed to evaluate how to optimally support innovative payload development and to increase the cadence and reliability of LDB and ULDB flights.

NASA's Program of Record. NASA's upcoming Roman Space Telescope, and ESA's Athena X-ray Observatory and LISA mission, in which NASA is a significant partner, are essential to the survey's science program. Advice on how to optimize the science return includes: holding a non-advocate review of Roman Space Telescope's science program to set the appropriate mix of survey time to guest investigator-led observing programs; and at the appropriate time, establishing funding for LISA science at a level that ensures U.S. scientists can fully participate in LISA analysis, interpretation, and theory.

1.7 CONCLUSION

The integrated program forwarded in this report advances a vision for discovery and progress for the coming decades. The content of the remaining chapters, together with the panel reports, represent an enormous effort that took years of preparation on the part of a large fraction of the astronomical community, and more than 2 years for the survey and its committees to complete. The full context of the recommendations and advice summarized in this chapter can only be appreciated by reading the report in its entirety. Realizing the opportunities presented in these pages will only be possible with the continued dedication and energy of the community, the agencies, and the excitement of the nation to explore the cosmos and answer some of humanity's most profound questions.

2

A New Cosmic Perspective

The past decade has been one of extraordinary discoveries in astronomy and astrophysics, and a realization of the scientific vision of the 2010 decadal survey, *New Worlds, New Horizons in Astronomy and Astrophysics.*[1] Scientific advances range from the detection of gravitational waves from merging black holes, to a direct image of a black hole in a nearby galaxy, to the production of heavy elements in the merger of two neutron stars, long-hypothesized but observed in exquisite detail for the first time. Increasingly sensitive instrumentation and powerful simulations are uncovering connections between the complex gaseous surroundings of galaxies and the forces that shape them. An explosion in the number of known exoplanetary systems has been accompanied by the detailed characterization of a subset of these other worlds, with insights into their formation arising from imaging of the disks where young planets are assembling. Newly discovered fossil structures from the formation of the Milky Way Galaxy open a window on the Milky Way's distant past, and observations take several steps closer to diagnosing the conditions present shortly after the Big Bang.

The investments of previous decades bore fruit in this decade in the awarding of Nobel Prizes in Physics for six discoveries derived from astronomical measurements: dark energy, neutrino oscillations, gravitational waves, exoplanets, physical cosmology, and black holes (Figure 2.1). At the same time, international collaborations greatly expanded. A salient example of this is the seminal paper detailing the discovery of two merging neutron stars and their gravitational and electromagnetic signatures: it encompassed nearly 4,000 authors from 900 institutions and 70 observatories, spanning all seven continents and space-based facilities, or roughly one third of the professional astronomical community as well as most of the gravitational wave community world-wide.

[1] National Research Council, 2010, *New Worlds, New Horizons in Astronomy and Astrophysics*, The National Academies Press, Washington, D.C., https://doi.org/10.17226/12951.

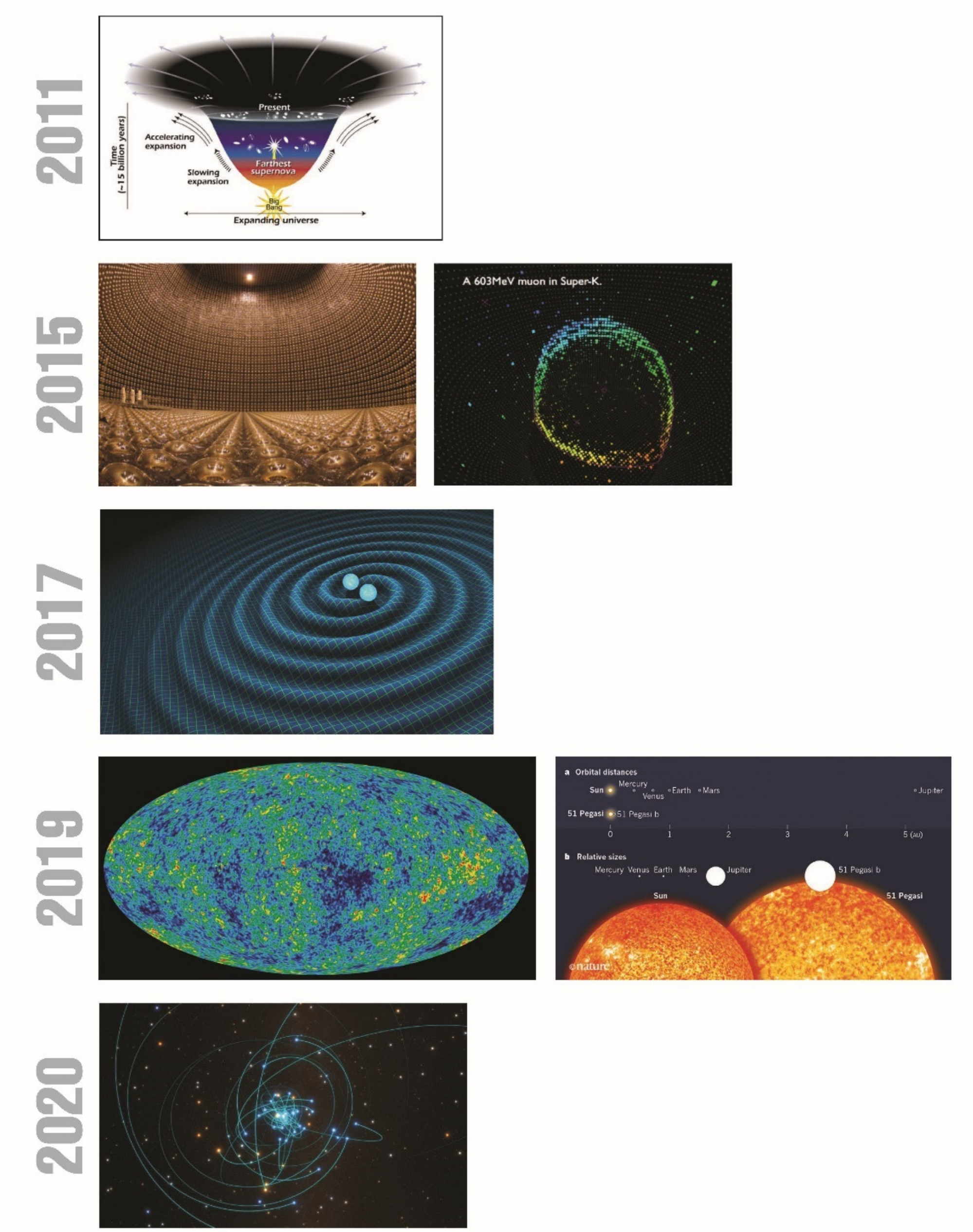

FIGURE 2.1 Physics Nobel Prizes derived from astrophysical measurements. The decade of 2011-2020 witness the awarding of Nobel Prizes for six different astronomical discoveries. In 2011, Saul Perlmutter, Adam Riess, and Brian Schmidt received the prize "for the discovery of the accelerating expansion of the universe through observations of distant supernovae." The citation to Takaaki Kajita and Arthur McDonald in 2015 was "for the discovery of neutrino oscillations, which shows that neutrinos have mass." In 2017, Kip Thorne, Rainer Weiss and Barry Barish were awarded the prize "for decisive contributions to the LIGO detector and the observations of gravitational waves." The year 2019 saw the awarding of the Nobel Prize in Physics to James Peebles "for theoretical discoveries in physical cosmology" and to Dider Queloz and Michael Mayor "for the discovery of an exoplanet orbiting a solar-type star." Most recently, in 2020, the topic of black holes received Nobel attention, with recognition to Roger Penrose "for the discovery that black hole formation is a robust theory of general relativity" and to Andrea Ghez and Reinhard Genzel "for the discovery of a supermassive compact object at the center of our galaxy." SOURCE: *2011:* NASA/STScI/Ann Field; *2015:* Kamioka Observatory, ICRR (Institute for Cosmic Ray Research), The University of Tokyo; *2017:* R. Hurt/Caltech-JPL; *2019-left:* NASA/WMAP Science Team; *2019-right: pending*; *2020:* ESO/L. Calcada/spaceengine.org.

The *New Worlds, New Horizons* decadal survey identified three main science objectives for their decade, while also enabling a wider discovery potential. The resulting scientific program is still bearing fruit in this decade and will in the next. On the topic of *Cosmic Dawn: Searching for the First Stars, Galaxies, Black Holes*, the James Webb Space Telescope (JWST) (launch expected in 2021) will directly examine the youngest observable galaxies, the Vera Rubin Observatory (science first light expected in 2021) and Nancy Grace Roman Observatory (launch expected in 2025) will transform views of dwarf galaxies at the extremes of galaxy formation and the record of ancient stars they left behind, and the European Space Agency's (ESA) Laser Interferometer Space Antenna (LISA) (launch expected in mid-2030s) should identify mergers of black holes all the way back to their earliest formation. In the arena of *New Worlds: Seeking Nearby Habitable Planets*, the Transiting Exoplanet Survey Satellite (TESS) (launched 2018), JWST, and soon Roman will explore a widening array of exoplanet types. For the *Physics of the Universe: Understanding Scientific Principles* objective, myriad ground- and space-based telescopes will use the universe as a laboratory to probe the nature of dark matter and dark energy. Roman, Rubin, ESA's Euclid (launch expected in the latter half of 2022), the most recent Laser Interferometer Gravitational-Wave Observatory (LIGO) upgrade to Advanced LIGO Plus (operations expected to begin 2024), the Event Horizon Telescope, and ground cosmic microwave background (CMB) experiments will advance understanding of the conditions present in the infant universe, the properties of dark energy, and the fundamental physics associated with black holes.

The development of scientific priorities for this survey began with the receipt of 573 science white papers (written by 4516 unique authors and/or endorsers) in early 2019; 573 such papers were received. These papers formed the starting point for deliberations by six expert science panels, organized by subfield: Compact Objects and Energetic Phenomena; Cosmology; Galaxies; Exoplanets, Astrobiology and the Solar System; The Interstellar Medium and Star and Planet Formation; and Stars, the Sun, and Stellar Populations. Each panel reviewed recent progress in their fields and identified key challenges and priorities for the coming decade and beyond. Their reports are included as an Appendix to this report. From these wide array of opportunities each panel identified four key science questions regarded as being especially ripe for investigation in the coming decade, along with a "discovery area" where emerging capabilities or techniques offer great promise for major advances. Tables 2.1 and 2.2 at the end of this chapter provide a summary list of these questions and discovery areas. These are not intended to encompass all of the important science needed, but rather to highlight particularly important questions and opportunities.

The panel reports were then integrated by the steering committee into the summary science case which forms the remainder of this chapter. It soon became clear that most of the science questions and discovery areas could be organized into three broad thematic areas: *Worlds and Suns in Context* highlights the extraordinary advances over the past decade in the study of exoplanets, stars, and their associated planetary systems, and the opportunities for transformational advances in these areas, including the ultimate search and characterization of habitable planets, in the decades ahead. *Cosmic Ecosystems* represents an integration and culmination of understanding the origins of galaxies, stars, planets, and massive black holes, and the realization that the life cycles of the universe over this billionfold range of scales are intimately connected, through feedback processes propagating through the gas within, surrounding, and between galaxies. The *New Messengers and New Physics* theme embodies the dual revolutions brought about by the marriage of observations of light with those from gravitational waves and elementary particles (multi-messenger astrophysics) along with the expansion of measurements of the sky over time (time-domain), as well as the opportunities for major advances answering some of the most fundamental questions in astrophysics and physics, such as the nature of dark matter and dark energy. Each of these three themes is summarized below, and Tables 2.1 and 2.2 list the thematic areas associated with each of the science panel questions and discovery areas. All are represented in the themes and many cross multiple themes.

Although the three overarching themes effectively distill the multitude of science goals and priorities contained within the six panel reports, their encompassing nature do not provide clear scientific goals which can be easily grasps by the non-professional readers of this report. With that in mind we have identified a key priority science question for each of the themes, which help to motivate the strategic program later in this report: Pathways to Habitable Planets; New Windows on the Dynamic Universe; and Unveiling the Drivers of Galaxy Growth. The priority areas for each theme are described at the end of each respective section.

2.1 WORLDS AND SUNS IN CONTEXT

The planets in our solar system, and the Sun at the center of it, provide the most direct connection to the myriad other stars and planets in our galaxy and the universe. With the flowering of capabilities expected in the next decade, progress in both stellar astrophysics and planetary science will expand and provide a broader context with which to understand and appreciate our cosmic perspective.

Within the last decade, progress in planetary demographics has achieved a reversal of sorts; new knowledge that planets are common, along with multi-planet systems, translates to a new cosmic understanding that there are likely more planets than stars in the universe. The Copernican revolution continues, in the realization that amongst the incredible variety of exoplanetary systems, our own solar system may prove to be an outlier rather than an exemplar. Along with pondering the cosmic order to understand stars, their formation, evolution, and ultimate fate, there is a parallel track for planets—how do they form and evolve?—and a merging of these tracks when considering questions related to habitability. It is an exciting time in which to practice the astronomical craft, as humanity edges ever closer to being able to answer the age-old question “Are we alone?” It is humbling and exciting to contemplate that the question of whether life exists elsewhere could be answered with the technology humanity now possesses.

The past decade has also witnessed a renaissance in stellar astronomy. Gaia, an ESA mission to deliver fundamental stellar parameters such as distances and three-dimensional motions on the sky, has proved revolutionary even with its initial data releases, to articulate the connections between and among groups of stars. Other time-domain observatories primarily designed to detect exoplanets have returned a wealth of asteroseismological observations allowing us to probe the interior structures of stars. Whereas in the past only a few fundamental parameters of a star could be determined accurately, and most of these in a relative sense, now mass, size, distance, age and chemical makeup for a wide swath of stars are measurable. The coming decade will continue this trend: where Gaia enabled precision stellar parameters of roughly a billion stars in the Milky Way, the Rubin Observatory promises to expand by a factor of 10 the number of main sequence stars for which distances can be determined.

2.1.1 Stellar Demographics

We are in the midst of a stellar renaissance, as astronomers come to know the individual journey of each star, separating them from anonymous points of light in the heavens to having the equivalent of detailed dossiers of physical characteristics and histories.

FIGURE 2.2 Mosaic image of part of our neighboring galaxy, Andromeda, as viewed by the Hubble Space Telescope as part of the Panchromatic Hubble Andromeda Treasury Project (PHAT). Roughly 100 million stars in this galaxy's pancake-shaped disk are resolved with this data. Precision characterization of the individual stellar properties enables determination of factors affecting galactic structure and evolution. SOURCE: https://hubblesite.org/image/3476/gallery/73-phat; NASA, ESA, J. Dalcanton, B.F. Williams, and L.C. Johnson (University of Washington), the PHAT team, and R. Gendler

This increasing complexity of stellar astrophysics extends to exposing the internal states of stars and the insights that come along with that revelation. Asteroseismology was a once-bespoke method of sounding the internal state of a star via detections of oscillation modes. The technique is now an established tool to determine precise stellar ages in large data sets, as well as returning masses and radii which can be used to calibrate models. The emergence of ultra-precise, long-duration, and continuous light curves from space (first with the European mission Convection Rotation and planetary Transits [CoRoT], then with NASA's Kepler and TESS as well as ESA's Characterizing Exoplanets Satellite [CHEOPS] and soon, ESA's Planetary Transits and Oscillations of stars [PLATO] mission) served both the exoplanet community and the stellar astrophysics community by enabling planet detections via the transit method and vastly expanding the number of stars for which asteroseismic oscillations (and subsequent stellar astrophysics studies; see below) could be detected. Probes of the internal states of stars via this method now return constraints on stellar structure previously only theorized. The cores of red giant stars appear to rotate faster than the surface, and oscillation frequencies differentiate red giant stars in which core helium burning is occurring, versus those only burning hydrogen in a shell. Latitudinal differential rotation in the convection zones of Sun-like stars, revealed through asteroseismic observations, indicate a shear much larger than predicted from numerical simulations. The mass distribution of red giant stars probed by asteroseismology does not match predictions from stellar population models. Asteroseismology detects strong magnetic fields in the cores of red giant stars. Indeed, high-precision, high-cadence light curve observations now join traditional methods of photometry and spectroscopy as essential tools of observational stellar astrophysics.

The next decade will continue to provide precision tests of stellar evolutionary models, and ultimately advance our understanding of the structure, dynamics, and evolution of galaxies as a whole (Figure 2.2). The elements present in a star's atmosphere reveals its origins and refines the understanding not just of that star, but when applied to large samples of stars, reveals how the assemblage of stars in our galaxy came to have its form, structure and content. Identifying and studying particular stellar subsamples such as cool subdwarfs provide the fossil record of the early history of star formation in the Milky Way Galaxy, as their elemental compositions are unchanged from that at their birth. Necessary links in this chain of chemical tagging are improvement in the laboratory and numerical calculations of atomic and molecular transitions and opacities, together with inclusion of more realistic geometries beyond assuming stars are spherical, and non-equilibrium effects in stellar model atmospheres. In the coming decades, high-resolution spectroscopy with extremely large telescopes will expand these abundance measurements throughout the Local Group. Connecting this chemical record of the galaxy with the dynamical record from Gaia and spectroscopy should produce a definitive fossil record of the assembly and life history of our galaxy. Progress in the next decade will require advances in industrial-scale spectroscopy (e.g., SDSS V and future even larger surveys) along with computational methods to harness the large data sets of photometric, spectroscopic and astrometric surveys.

Whether a star has one or more partners, and the nature of those partners, influences its life history because of mass exchange, particularly for massive stars. This is especially true for the end state of stellar evolution, as the formation of compact objects provides connection points to probes of extreme gravity and extreme particle acceleration. The detection of gravitational waves from astrophysical objects this decade leads to invigorated research into the endpoints of stellar evolution for the next decade. Stars in multiple systems can have very different evolutionary pathways compared to their single counterparts; detailed photometric and spectroscopic electromagnetic observations spanning infrared through X-ray wavelengths, coupled with gravitational wave measurements and attention from theory, will elucidate these multiple routes and their consequences. Theoretical modelling of binary co-evolution is necessary to understand the fate of close binaries that do interact. White dwarfs caught in the act of merging will be studied by the Rubin Observatory, and can be linked to resolved gravitational wave signals detected by LISA once it launches, as these systems will also likely contribute a major part of the stochastic background expected from the Milky Way Galaxy. Further observational and theoretical research will sharpen constraints on the mass threshold that separates an end state of a white dwarf from core-collapse supernova formation. The effects of rotation, binary, and magnetic fields can be implemented in population synthesis models, with wide applicability ranging from understanding the ionizing output of massive stars to interpreting the spectra of distant galaxies. More generally, mapping out the myriad evolutionary pathways by which stars come to populate every part of the Hertzsprung-Russell diagram will require a robust observational and theoretical understanding of the formation, evolution, and especially interaction of stars in multi-star systems.

The angular momentum of a star, typically measured by its surface rotation period, is a key parameter to unlocking its current state and is a fundamental parameter in its own right. Determining rotation periods for a few tens of stars would in years past have been the subject of a Ph.D. dissertation. Now, with automated light curve analysis of stellar targets to search for transiting planets, a click of a button is practically all that is needed to compute rotation periods of tens of thousands of stars within a single research paper. This has proceeded in recent years largely as the stellar byproduct of exoplanet studies (Figure 2.3), a fortuitous result but one biased by the particular target selection criteria used for exoplanet searches. The next decade will see rotation rates of stars (and other time-domain stellar astrophysics measures such as flaring) determined over nearly the entire galaxy, likely with the aid of modern machine learning practices. Application of these rotation periods as clocks for stellar age dating converts stellar time-domain astronomy into the industrial-scale extraction of stellar birth dates. This furthers the personalization of stellar histories and adds precision to testing stellar models, as well as providing ground-truth age dating to constrain models of the Milky Way’s formation and evolution history.

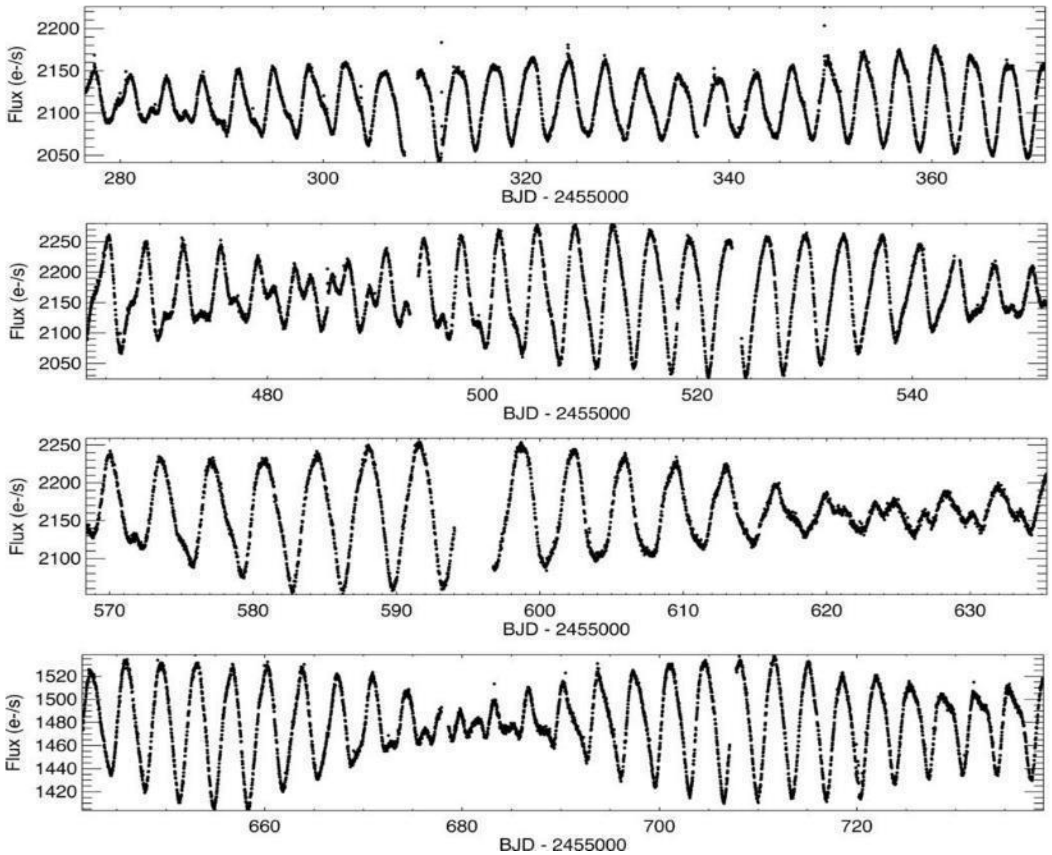

FIGURE 2.3 Evidence of stellar rotation appears from periodic surface features in the Kepler light curve of this K-type dwarf studied by Roettenbacher et al. (2013). Long-term evolution in the number and distribution of starspots is indicated by the changing patterns in the light curve, as well as short-term brightenings that indicate flares.
SOURCE: Adapted from "Imaging starspot evolution on Kepler target KIC 5110407 using light curve inversion," Rachael M. Roettenbacher et al 2013 *The Astrophysical Journal* 767 60, © AAS. Reproduced with permission. doi:10.1088/0004-637X/767/1/60.

Amongst all the detailed stellar knowledge, astronomy still discovers entirely new types of stars and stellar phenomena. Progress in the next decade is expected to reveal more and different types of stellar exotica, as testaments to the incompleteness of the knowledge about the many forms and varieties of stellar behavior. The unprecedented dimming of Betelgeuse**,** one of the brightest stars in the sky and the shoulder of the Orion constellation, from December 2019 through May 2020, proved that even well-studied stars can throw an unexpected astronomical curveball. Multi-wavelength observational capabilities which include ultraviolet and even X-ray wavelengths provide key diagnostics of temperature, density, abundances, and kinematics, essential for the post-mortem understanding of these events. The rare phenomena, such as the unusual dimmings of Boyajian's star, sparsely erupting pre-main sequence stars, or the elusive luminous blue variables, will become commonplace, and understanding will hopefully follow discovery of stars and their environments in a virtuous cycle.

The Sun is a singular star amongst all the others, primarily for its proximity but also because of our dependence on its behavior for our existence. Observations of the Sun are necessarily a touchstone for virtually all of stellar theory, with ripples throughout the understanding of all of stellar astrophysics. The anticipated first science to be done with the revolutionary Daniel K. Inouye Solar Telescope (DKIST) facility promises more answers to fundamental questions of how magnetism controls our star. This four-meter, solar-dedicated telescope on the ground (a "Hubble for the Sun"; see Figure 2.4) represents a large leap in solar observing capabilities. The interplay between magnetic flux and mass flows is of universal astrophysical importance. This decade's detailed observations will have the requisite spatial, temporal, and spectral resolution as well as the dynamic range to provide a ground-truth for models of basic magnetic structures. Advances originating from this zoomed-in view of solar magnetic structures require complementary global measurements—particularly of the solar corona—to understand magnetic energy storage and release. Complementary radio observations could generate three dimensional mapping of the magnetic field in sunspots, and magnetic field measurements in the global corona provide context for the more detailed, restricted field of view measurements.

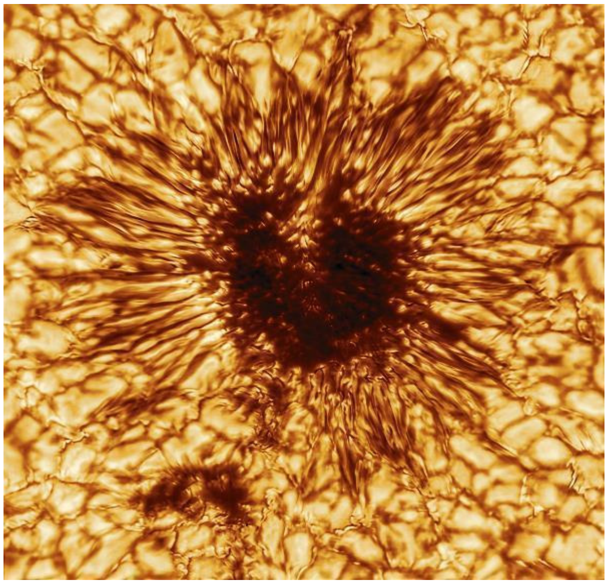

FIGURE 2.4 This is the first sunspot image taken on January 28, 2020, by the National Science Foundation's Daniel K. Inouye Solar Telescope's (DKIST) Wave Front Correction Context Viewer. The image reveals striking details of the sunspot's structure as seen at the Sun's surface. The sunspot is sculpted by a convergence of intense magnetic fields and hot gas roiling up from below. This image uses a warm palette of red and orange, but the context viewer took this sunspot image at the wavelength of 530 nanometers - in the greenish-yellow part of the visible spectrum. SOURCE: NSO/AURA/NSF (see https://nso.edu/press-release/inouye-solar-telescope-releases-first-image-of-a-sunspot/).

The singular nature of the Sun is both its blessing and its curse. Its ability to provide constraints on astrophysical questions with unrivalled accuracy is in tension with its fitness as a prototype for all things stellar. While it is a spectacularly well-studied star, it is a unique case study observed at one point in its 4.6 billion year evolutionary history. Observations in the last decade provided tantalizing hints that the Sun's cycle of magnetic activity does not behave the same as other solar-like stars. Work in the next decade will broaden the range to include a larger sample and provide the perspective of stars of different ages and spectral types. While hints exist now of a disconnect in the relationship between fundamental stellar parameters and magnetic properties—magnetically active stars can have larger radii than predicted based on their temperatures; stellar twins identical in all other properties can have very differing levels and amounts of magnetic field distribution and magnetic activity signatures—observations in the next decade will further this discovery space and fill important gaps in the foundation of stellar structure theory.

On a broader scale, it is clear that stars are not uniform discs of light. Time-domain astronomy enables the characterization of surface structure using the changes in light seen either in broad wavelength bands (Figure 2.3) or in narrow spectral features that reveal inhomogeneities. Optical and infrared interferometric observations have begun to localize these regions. Observations spanning multiple regions of the electromagnetic spectrum provide independent constraints on the properties of these features, characterizing the magnetic, chemical, or cloud-like nature of the features. Indeed, the ability to measure rotation periods arises from the periodic brightening and dimming caused by blemishes passing across the face of each star. On stars like the Sun those spots reveal the telltale existence of subsurface magnetism; the multitude of long-duration high precision photometric light curves available from space missions including CoRoT, Kepler, and TESS, reveal the ubiquity and variety of surface inhomogeneities of cool stars. Future progress in interferometers, from optical to infrared to radio wavelengths, will advance knowledge of the underlying physics causing spatial inhomogeneities on stellar surfaces.

The importance of magnetic fields in a stellar context was expected but is now unequivocal. Magnetism in massive stars has been confirmed (but not entirely explained) in a minority of cases, and stronger magnetic fields than initially expected have been found at the end of the main sequence, in the ultracool dwarfs spanning stellar and substellar origins. A variety of observational tools enable these discoveries: high-energy observations detail the magnetic shaping of stellar winds and unveil the existence of heated plasma in an above-the-surface corona; optical and infrared spectra reveal the tell-tale signature of magnetic splitting of atomic and molecular lines; and radio observations expose the action of accelerated electrons thereby revealing the presence of magnetic fields in their atmospheres. Stellar coronae are common in cool main-sequence stars, but the origin of this heated plasma is still unknown, even on the Sun. Future sensitive high energy and radio measurements will probe the coronae of stars in more detail and expand the number of objects amenable to study. Ultraviolet spectra probe the chromosphere and above, regions in a cool star's atmosphere where magnetic forces begin to dominate. These observational advances also require associated advances in modelling the intricacies of stellar structure for full understanding of the impact of convection, rotation, and magnetic field generation on the complex and dynamic nature of stars.

2.1.2 Exploring Alien Worlds

Over the past two decades an incredible diversity of worlds and systems has been discovered, from giant planets over 10 times the mass of Jupiter to the Trappist-1 system with seven Earth-sized planets packed in an area smaller than the orbit of Mercury. This revolution is ongoing. New capabilities led to an explosion of discoveries in the last decade (Figure 2.5), filling in demographics of planets, expanding the ability to characterize the composition and nature of individual members, and advances in new facilities adding to the depth of characterization possibilities.

These discoveries are helping us understand fundamental questions about ourselves. How did the solar system form? Are systems like our own common or rare? Are planets like Earth common or rare? And, ultimately, do any of those Earth-like planets harbor life? With current technology, the planets in solar systems like our own are nearly undetectable; with new facilities over the next two decades, this will change, and the picture of other worlds will start to become complete.

Finding exoplanets is challenging. Many exoplanet detection methods have been developed, and each gives some important, but limited, information. The radial velocity (RV) technique measures orbits and constrains the mass of a planet using the motion of its parent star. This technique has been used for a number of exoplanet "firsts" over 25 years (including the Nobel Prize winning discovery of 51 Peg b), and astronomers are pushing the technology to be able to detect Earth-mass planets on periods of months, to perhaps a year, in search of objects similar to Earth. In the last decade, another technique, exoplanet transits, has transformed the view of exoplanetary systems, in particular when employed via high precision space missions like NASA's Kepler and TESS missions, and before them the European CoRoT mission. A transit, when a planet passes directly in front of its parent star, allows for a measurement of a planet's radius and orbital period. Kepler showed that planets on close-in orbits (within 100 days) are extremely common (Figure 2.5), which has revolutionized the understanding of planet formation, and is one of many examples that show that architecture of planetary systems is exceedingly diverse.

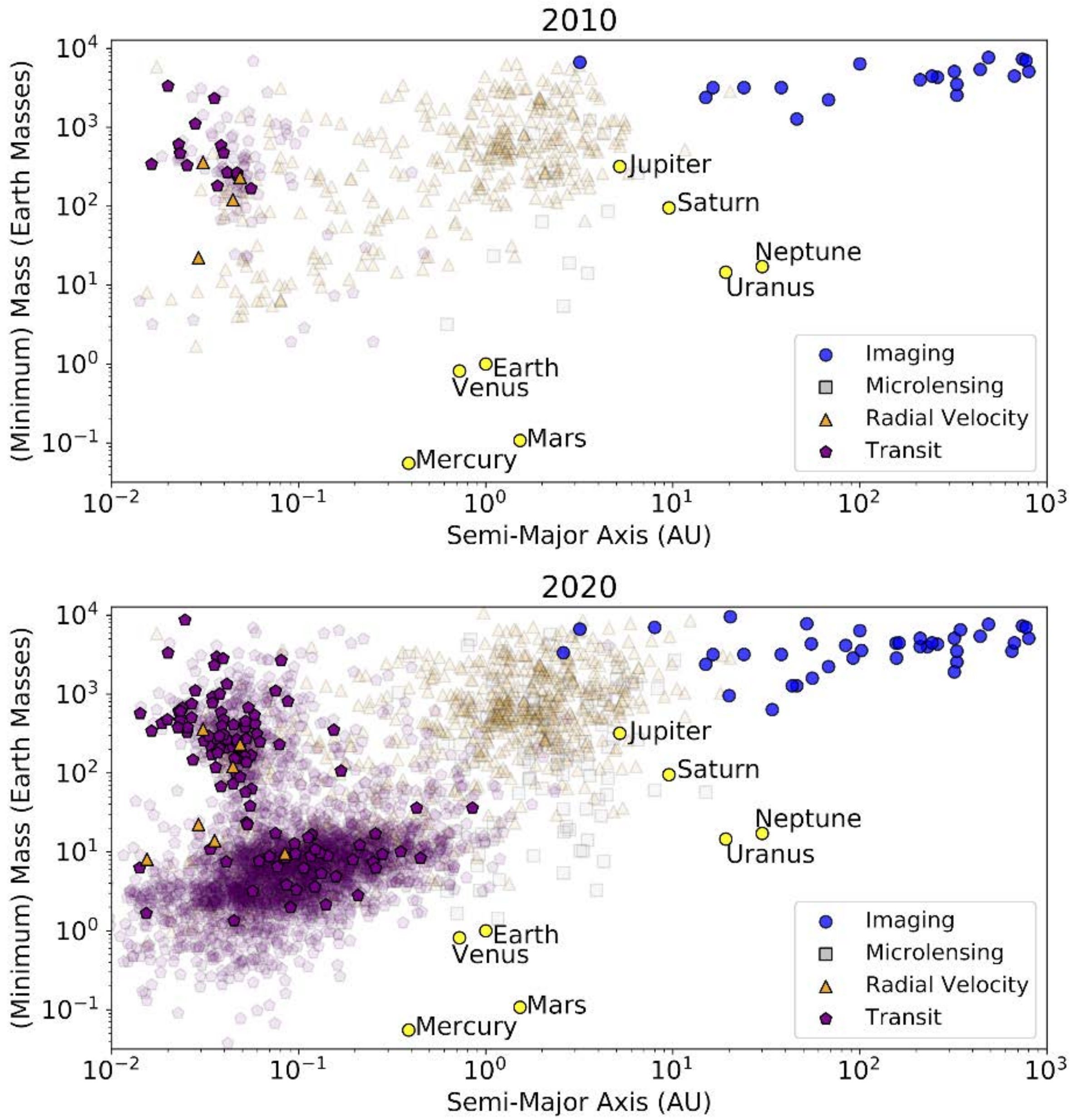

FIGURE 2.5 The population of known exoplanets in 2010 (*top*) and 2020 (*bottom*). Each symbol represents a known extrasolar planet, colored by initial discovery method. Hollow symbols are planets that have been discovered. Filled symbols are planets whose atmospheric composition have been *characterized* by measurements of its spectrum or brightness. Over the past decade astronomers have begun to move from the era of planetary census-taking to detailed characterization, and the next decades will both complete the missing parts of the census—planets like our own solar system—and see an explosion in characterization. SOURCE: D. Savransky and B. Macintosh, with data from the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program.

The radial velocity and transit methods are best suited for finding planets on relatively close-in orbits. A method prioritized by *New Worlds, New Horizons,* gravitational microlensing, will be used by the Roman Space Telescope in this decade to complete the planetary census by finding planets from 1 to 100 AU, and even free-floating planets. Microlensing exploits the bending of light by the gravity of the star and planet to detect distant systems. This technique will be powerful for measuring how common such planets are on these wider orbits. However, these planets cannot be followed up for future in-depth observations, because this chance star-planet alignment does not repeat.

Another major technique of long-term importance to the field is direct imaging, where the light of the parent star is blocked to make a faint planet directly visible. Currently this is only practical for young giant planets in the outer parts of solar systems. However, with suitable technology development there is now a clear path forward to use direct imaging techniques on ground-based extremely large telescopes (ELTs) and optical interferometry with 8 m-class telescopes to study the atmospheres of temperate rocky planets around low-mass stars, and to use future space missions to study potentially Earth-like planets around Sun-like stars, as well as the enormous diversity of non-Earthlike planets.

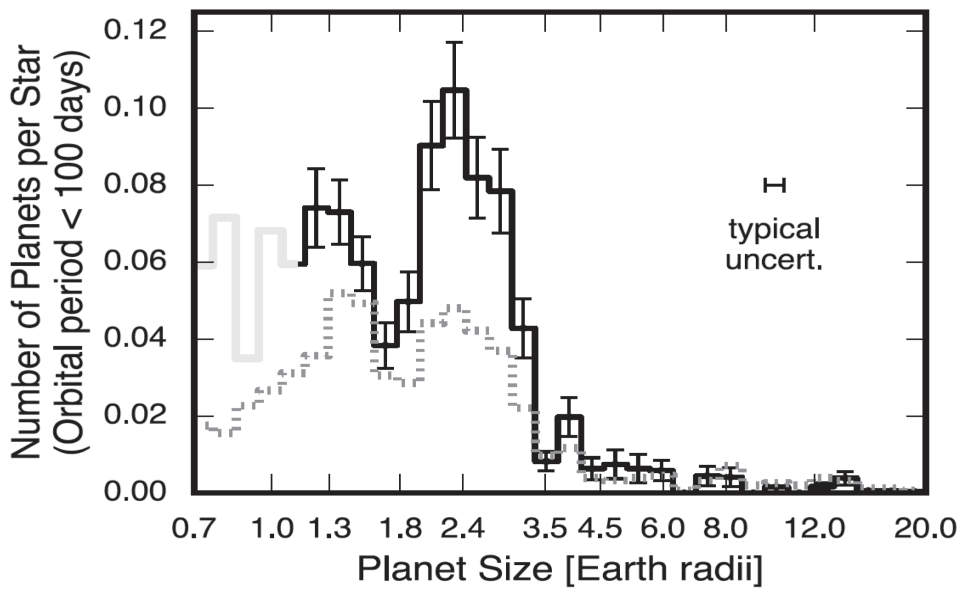

FIGURE 2.6 Plot of number of planets per star vs. planetary radius, for orbital periods less than 100 days, shown in solid black. The dashed grey line is from planet detections before including the Kepler Mission completeness corrections, which becomes more pronounced at smaller planet sizes. The analysis of data from the Kepler Mission showed a pronounced gap between smaller, denser worlds (1-1.7 Earth radii) and larger more Neptune-like planets (1.7-4 Earth radii) with thick hydrogen atmospheres. Detailed analysis of the gap as a function of orbital period and system age suggests that the smaller planets are likely the "stripped cores" of formerly more Neptune-like worlds. SOURCE: Adapted from B.J. Fulton and E.A. Petigura, 2018, "The California-Kepler Survey. VII. Precise Planet Radii Leveraging Gaia DR2 Reveal the Stellar Mass Dependence of the Planet Radius Gap," *The Astronomical Journal* 156 264, © AAS. Reproduced with permission. doi:10.3847/1538-3881/aae828.

With these diverse techniques astronomers have produced a partial census of exoplanets. The results have been extraordinary, showing that on the average there are at least two planets per star in the Milky Way, and that many planetary systems are very different than our own, with crowded systems of planets intermediate between Earth and Neptune in size. So far it appears that lower-mass M-dwarf stars have more planets than Sun-like stars. Careful characterization of the stars themselves has been crucial to interpreting Kepler results, and broader studies will reveal how these distinctions persist across different regimes. The census remains incomplete; all current planet detection surveys would be essentially blind to every planet in the solar system at their known orbital separations, except Jupiter.

The Kepler Mission has also shown that the changing nature of planets—their time evolution—is also fundamental to understanding the planetary population. Figure 2.6 shows a recent discovery using Kepler data, of a fundamental divide between rocky planets below about 1.7 times the radius of Earth, and larger "sub-Neptunes" that have thick hydrogen-dominated envelopes that increase the planetary radius.

The "gap" between these two populations is thought to be due to the gradual loss of hydrogen-envelopes possibly due to a stellar wind that drives these atmospheres off the planet and into space. The detection of this gap was made possible due to exquisite characterization of the host stars and shows the important connection between stellar and planetary physics.

Approximately twenty Earth-sized (<1.7 R_Earth) planets have been discovered within the habitable zone of low-mass stars—neither too close nor too far, but receiving enough energy to allow liquid water on their surface. A key question that Kepler and other projects have tried to address is the frequency of potentially habitable planets - the average number of Earth-sized planets within the habitable zone of their star, particularly around Sun-like stars. A decade ago, this quantity—known as η_{Earth}—was almost completely uncertain. These planets are very difficult to detect directly, so their exact occurrence rate must be extrapolated from planets higher in mass, closer to their star, and/or from the frequency of habitable zone planets orbiting lower-mass stars. Current constraints indicate that 18-28 percent of Sun-like stars have an Earth-sized (80 to 140 percent of Earth's radius) planet in their habitable zone, enough to make their study with a future large mission practical. (The number is likely a factor of two higher for cooler lower-mass stars.) The nature of these planets is of course almost completely unknown, whether they have followed a path of planetary evolution amenable to life or something completely different than Earth; but that very uncertainty is a compelling scientific reason to try and answer this most profound question.

2.1.2.1 Planet Formation

The detailed process of planet formation is one of the great unknown frontiers in astrophysics. The core concept of gas and tiny dust particles in a circumstellar disk assembling into planets is well known. But the details of how particles assemble, how the process overcomes barriers to operate quickly enough before the disk dissipates, how this leads to the incredible diversity of known planets, and whether this process can frequently produce planets with Earth's key characteristics, remain unknown. Understanding this is crucial to placing potentially habitable planets in context, including questions as to how water and atmospheric volatiles are delivered to Earth, whether giant planets affect the evolution of terrestrial worlds, and whether "mini-Neptune" planets can evolve towards habitability. More sensitive observations of planet-forming disks to understand the astrochemistry, dynamics, and role of water in the formation of habitable planets through radio, mm, and far IR spectroscopy will help advance this field.

Planetary demographics and composition only reveal the endpoint of this process. Stunning observational advances from ALMA show complex structures in planet-forming disks (see Figure 2.20), and recent advances with large ground-based telescopes have even caught a protoplanet with its own accretion disk (Figure 2.7). These structures, however, still correspond to large planets in the outer parts of solar systems; in the habitable zone or the regime of the Kepler super-Earths the process itself remains almost inaccessible. Current observations of continuum dust emission cannot probe the innermost regions of the planet-forming disks where Earth-like planets may be located; sensitive radio interferometry at longer wavelengths than those probed by ALMA is needed to see the optically thin emission originating in these locations. Spectra at far infrared wavelengths would provide a unique and revolutionary census of water within these disks, which is a key to understanding giant planet formation and the distribution of water among terrestrial planets. New radio and far infrared surveys of planet forming disks would enable leaps in understanding that would surpass this era's ALMA-driven revolution.

There have been tremendous advances in the theory that underlies planet formation, as new ideas such as the "streaming instability" and "pebble accretion" use gas and particle interplay and physical changes in the disk to trap planet-forming material and rapidly form giant planets. The models incorporate uncertainties such as turbulence and feedback between the forming planets and disks. Better observations of gas and dust in disks (especially on smaller scales and across a range of ages) and of forming planets (especially of lower masses) will advance this field, combined with larger-scale computational models and laboratory experiments.

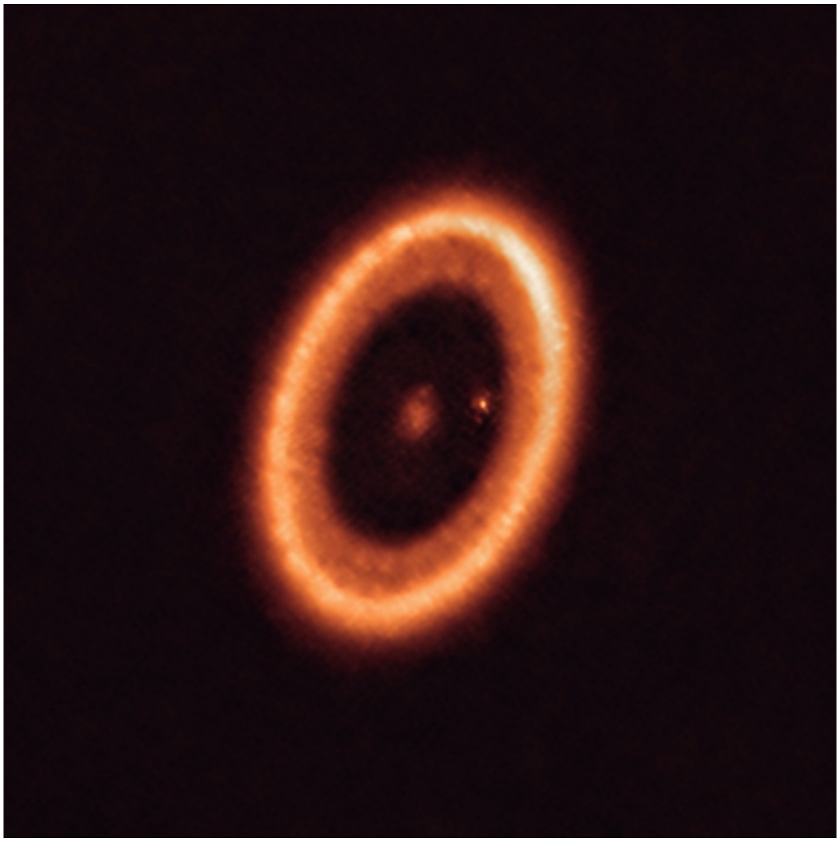

FIGURE 2.7 ALMA image of the young planet-forming star PDS70 showing a outer ring of leftover debris, an inner disk of planetesimals, and a potentially moon-forming disk orbiting a young Jupiter-like planet (at 3:00). The outer ring is about 75 AU in radius, about twice the size of our own solar system. SOURCE: ALMA (ESO/NAOJ/NRAO)/Benisty et al. See https://www.eso.org/public/images/eso2111b/.

2.1.2.2 Atmospheric Spectroscopy to Characterize Worlds

Only a small fraction of the thousands of known planets have been characterized beyond their basic mass or radius, along with orbital properties, but those available observations further illustrate the diversity of worlds. In addition to completing the planetary census, the other major frontier of the 2020s is the spectroscopic observation of planetary atmospheres. A spectrum can yield the abundances of atoms and molecules as well as the temperature of a planetary atmosphere (Figure 2.8). Already several dozen planetary atmospheres have been characterized by the Spitzer and Hubble space telescopes and by large ground-based telescopes. Observations of the transiting and directly imaged planets target the extreme inner and outer reaches of planetary systems, respectively. So far this work has mostly focused on the easier to observe giant planets, but already the diversity of planets beyond what is seen in the solar system is on full display. Detections have been made of clouds of rock dust in the hottest planets, metal vapor such as sodium, potassium, iron, and calcium, escaping hydrogen, and molecules including water vapor, carbon monoxide, and methane. Abundance determinations show the current state of these atmospheres, and—as in the solar system—the history of planetary formation and evolution is embedded in them.

NASA's TESS mission, along with ground-based surveys, are finding planets around the nearby bright stars, to further fuel this revolution in atmospheric spectroscopy. The premier tools for obtaining exoplanet spectra in the 2020s will be JWST and extremely large ground-based telescopes. These telescopes will revolutionize the understanding of the composition of exoplanets. Atmospheres are the window into many physical, chemical, and formation processes, and these platforms will deliver spectra for a continuum of worlds, for Jupiter-size gas giants, to the Neptune-size planets, to the mysterious mini-Neptunes that dominate the exoplanet census, and down to Earth-size rocky worlds. The lessons of atmospheric physics and chemistry learned from one planetary class can readily inform the understanding of other classes. As astronomers will be far outside of the realm of solar system expertise, and while tentative predictions of what to expect have been made, the joy of discovery will be seeing the diversity of these new worlds. In particular for terrestrial planets, significant progress in the 2020s will occur for systems around low-mass M-dwarf stars. Nearly all of this science will focus on planets within a few tenths of an astronomical unit (AU) of their parent stars, orbits tucked in far closer than Mercury is to the Sun. Moving beyond this to characterization of systems that look more like the solar system will require new capabilities. To understand potentially Earth-like planets around stars like the Sun, which is the only example, so far, for life, will drive the field towards even loftier goals beyond the 2020s (Box 2.1). Are such planets habitable? Do they show signs of life?

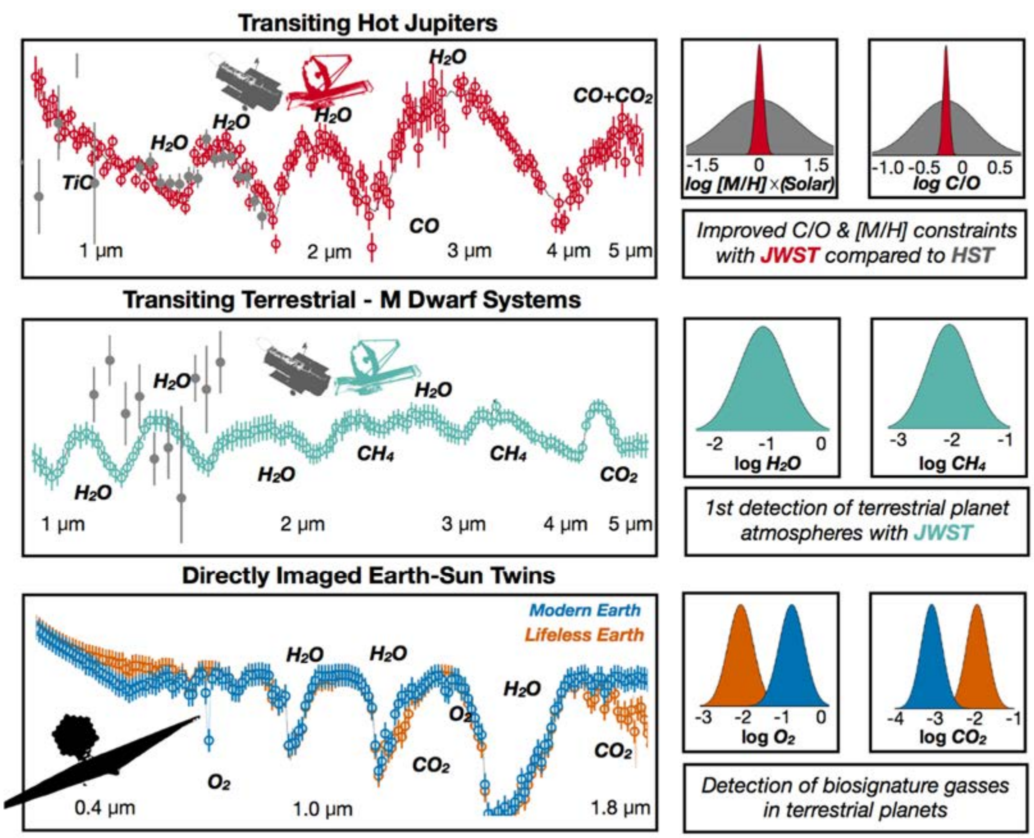

FIGURE 2.8 The 2020s and beyond will be an era of spectroscopy of exoplanet atmospheres. For giant planets such as transiting "hot Jupiters" (*top panel*) the limited wavelength coverage and precision of Hubble will give way to high-precision spectroscopy across JWST's wide wavelength range, yielding the detection of many molecules, and comprehensive atmospheric characterization including metallicities ([M/H]) and carbon-to-oxygen ratios. For terrestrial exoplanets that transit small M dwarf stars (*middle panel*), where Hubble is only able to yield marginal constraints, a significant allotment of JWST observing time will allow for the first reconnaissance of these atmospheres, including the ability to determine the mixing ratios for a range of molecules important for life, like water vapor and methane. Looking to the future (*bottom panel*), to examine the atmospheres of potentially Earth-like planets around Sun-like stars will require further development of a specialized space telescope for high-contrast imaging to measure a reflection spectrum that could show oxygen, methane, water vapor, and carbon dioxide. SOURCE: Courtesy of Natasha Batalha and the PICASO project, http://natashabatalha.github.io/picaso.

BOX 2.1 Detecting Life on Exoplanets

Earth's current atmosphere is shaped in many ways by the presence of life—from the abundance of oxygen, to the atmospheric and climate changes over the past century as humans burn fossil fuels. Such biosignatures could be detected spectroscopically in exoplanets, if life is as prevalent as it is on Earth. Astrobiologists and exoplanetary scientists have assembled lists of proposed biosignatures. Large-scale atmospheric oxygen is one of the most powerful, particularly in combination with the detection of other compounds such as methane. On Earth, atmospheric oxygen would be short-lived if not replenished by photosynthesis. An Earth-like spectrum would be a strong indicator of life-like processes. However, interpretation of such a detection must be extremely careful; abiotic processes can also produce complex chemical signatures. Predicting all the possible chemical pathways in a remote planetary environment, perhaps orbiting a star very different than the Sun, will be challenging. Interpreting any observations as a sign of life must involve high-quality data detecting multiple atmospheric constituents, an integrated view of the planet's evolution and its interaction with its star and solar system, and development of a comprehensive framework allowing a probability analysis. One important component of this will be the study of exoplanets over a wide range of masses, ages, stellar insolations, and compositions; even uninhabitable planets can provide clues to how atmospheres are formed and evolve, as the atmosphere of Venus helped the understanding of the history of Earth.

The search for biosignatures in exoplanet systems is focused on the most Earth-like worlds, particularly those that could have liquid water on their surface. This in turn concentrates attention on planets in the habitable zone. The exact borders of this zone depend on the details of the planet's atmosphere, orbital and rotational motions, and the interactions with the star, but in the solar system at the Sun's current age it extends from just inside the orbit of Earth to somewhere around the orbit of Mars. Since stars with lower mass than the Sun are much less luminous, the habitable zone moves much closer to the star, well inside the equivalent of the orbit of Mercury for the lowest-mass stars.

This range of star/planet separations leads to a range of different pathways for searching for life-bearing planets (Figure 2.1.1). For low-mass stars, the close-in habitable zone means those planets are much more likely to transit their parent star, and the small star yields a larger relative transit signature. These worlds are being identified by ground-based surveys and TESS, and will be studied in transit spectroscopy by JWST and potentially by the ground-based ELTs. Exactly how many Earth-sized habitable zone planets can be probed this way is unclear. It is likely a small number, and not all key biosignatures (including oxygen) may be detectable, but the first glimpses of planets in the habitable zone will come from transiting worlds orbiting the very lowest-mass stars. Directly imaging non-transiting, potentially habitable planets of the nearest low-mass stars requires extreme angular resolution but is only moderately demanding in terms of relative brightness, and hence is feasible from the ground with high-performance adaptive optics. Achieving this is a key science goal of the proposed ground-based extremely large telescopes.

Although the planets orbiting low-mass stars are the easiest to study, those stars are also very different from the Sun, and yield habitable zone environments that may be quite different from the solar system. Low mass stars' high initial luminosities and their prolonged output of high-energy radiation may make atmospheres difficult to keep for habitable zone planets. Potentially Earth-like planets around Sun-like stars orbit at larger distances, and hence are easier to spatially separate from their parent star even with moderately-sized (4 m+) telescopes equipped with coronagraphs or starshades. However, the bright star still makes them hard to see; Earth seen from beyond our solar system is more than 10 billion times fainter than the Sun at visible wavelengths. Reaching this level of planet to star contrast requires extraordinarily precise control of the wavefronts of light and is only practical with dedicated space-based telescopes. At longer wavelengths, the contrast ratio is more favorable, but the thermal background and scattered light would still swamp the planet signal for any but the few very closest sunlike stars.

It will be necessary to study sufficiently large samples of planets both inside and outside the habitable zone to find potentially rare Earth-like planets, so that connections between planetary properties and environment can be explored. Comparative planetology between systems influenced by very different stars and evolutionary pathways and examining multiple planets in a given system will help make the interpretation of atmospheric signatures more certain.

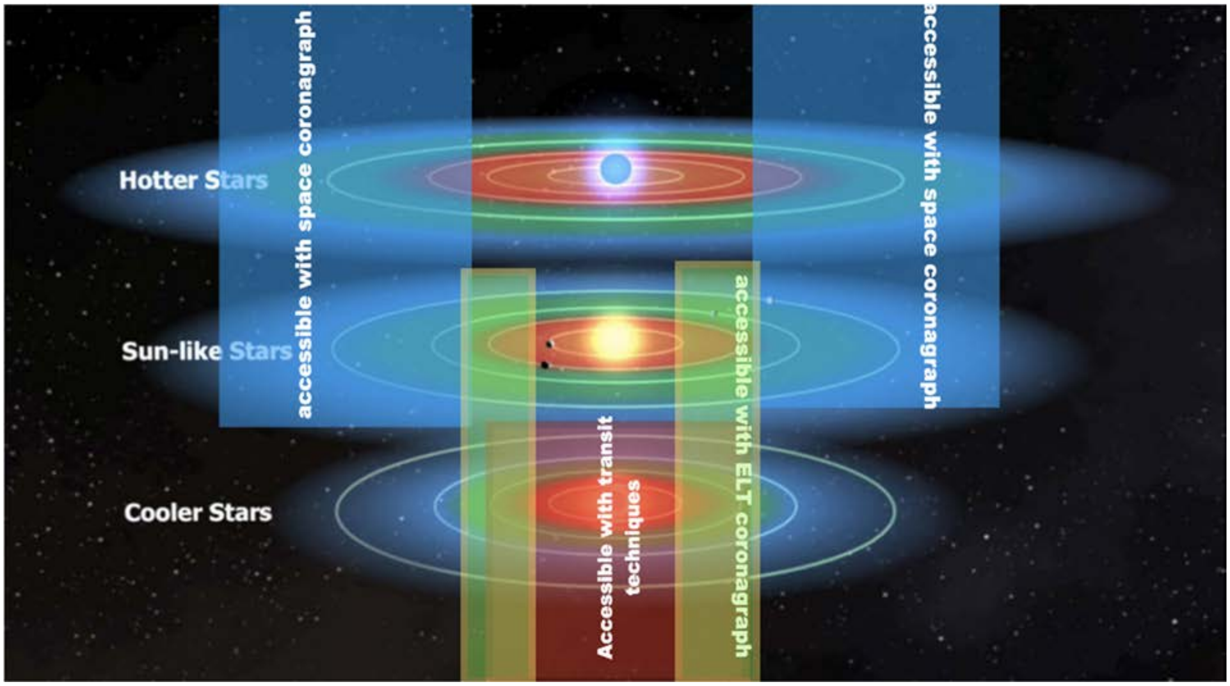

FIGURE 2.1.1 The location around a star where liquid water can exist—the habitable zone—changes with stellar temperature. Different types of telescopes are needed to probe these locations, from space- and ground-based coronagraphs which can return imaging observations of planets, to close-in planets only accessible through transit measurements. SOURCE: NASA/Kepler Mission/Dana Berry, adapted from https://www.nasa.gov/ames/kepler/habitable-zones-of-different-stars.

2.1.2.3 Connections to the Solar System

Studies of exoplanets and of the solar system are tightly intertwined and have enjoyed a profitable give and take in contributing to the understanding of planet formation and atmospheres. Many aspects of exoplanet atmosphere modeling use numerical techniques originally developed for the solar system's own giant and terrestrial planets. The solar system provides "ground truth" to validate spectroscopic observations against in-situ measurements such as elemental abundances on Jupiter or Earth or hazes in satellite atmospheres (Figure 2.9). On the other hand, the vast diversity of exoplanet types—many with no analog in the solar system—and the opportunity to study systems over a range of ages provides insights into formation mechanisms and evolutionary processes that improve the understanding of the solar system's own history.

The solar system contains asteroids and comets as remnants of its original formation. Analogues around other stars take the form of debris disks, massive dust belts produced by collisions among these small bodies. Measurements of the composition, orbital dynamics, and size distributions of small bodies in the solar system provide crucial benchmarks for understanding solar system formation in one spectacularly detailed instance. Time-domain surveys such as the Vera Rubin Observatory's Legacy

Survey of Space and Time will greatly expand the number of known small bodies in the outer solar system, and provide information about its early evolution. Results from recent studies analyzing dynamics of small bodies in the Kuiper Belt provide tantalizing hints, to be confirmed, about the possibility of additional planets.

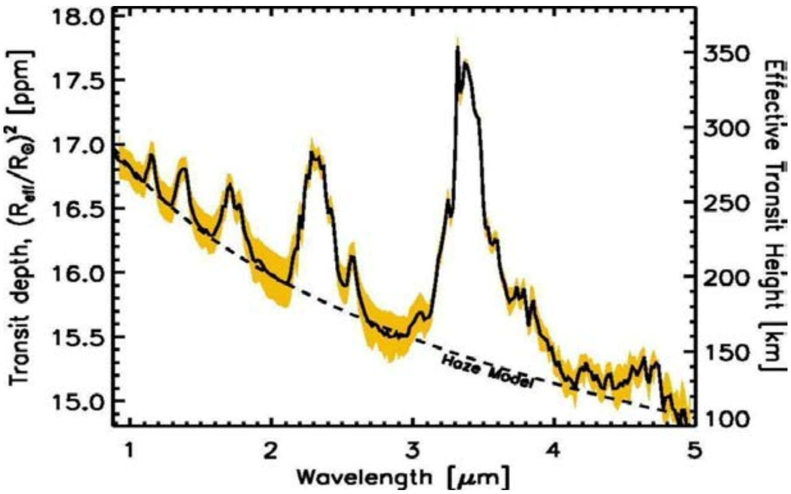

FIGURE 2.9 Spectrum of the hazy satellite Titan during an occultation of the Sun as viewed by the Cassini mission to the Saturn system. Recognizing the analogous geometries between such occultations and exoplanet transits of their host star, Robinson et al. (2014) used these observations to inform how high altitude hazes influence transit spectra of exoplanets. This cross-disciplinary approach will be key to interpreting exoplanet transit spectra taken by the James Webb Space Telescope. SOURCE: Courtesy of T.D. Robinson, L. Maltagliati, M.S. Marley, and J.J. Fortney, *Proceedings of the National Academy of Sciences* 2014, 111 (25) 9042-9047, doi: 10.1073/pnas.1403473111. Copyright 2014 National Academy of Sciences. Reproduced with permission.

While the detailed information available from solar system observations informs theories of planetary system formation in general, such studies are also synergistic. Observations of young planet-forming disks provide a window into the early conditions that led to the formation of the giant planets in this solar system. New generations of radio interferometers will probe inner solar system scales of planet formation, approaching scales of Earth’s orbit and super-Earth planet masses. High angular resolution near- and mid-infrared observations of the innermost regions of planet formation around young stars directly image thermal continuum emission of forming planets and provide kinematic constraints. Increased capabilities to track circumplanetary disks and substructure add the possibility to measure orbital motions of Earth-like planets as they are forming, and provide context for factors potentially affecting our early Earth. Detection and study of complex organic molecules provide the chemical initial conditions of forming solar systems. With these high angular resolution and high sensitivity techniques, young planetary systems can be observed—analogs to the solar system—at their moment of formation, validating models that explain the origin of the life-bearing conditions on Earth.

As the only place for which in situ and other detailed studies are possible, advances in knowledge of the solar system are a crucial part of the astrobiological endeavor, and astronomical observations in turn inform the understanding of the planets in direct reach. Future space coronagraphic observations of young Venus or Mars analogs, for example, could help confirm whether those planets had more Earth-like atmospheres in the past. Continued solar system space missions, complemented by long-term monitoring from telescopes at Earth, provide essential details to broadly understand these planets. Comprehensive theories of planet formation must be able to explain both the solar system and the diverse array of known exoplanetary systems.

One of the most exciting astronomical discoveries of the past few years are two interstellar interlopers—the asteroid ‘Oumuamua and the comet 2I/Borisov—that originated around another star and passed through our solar system. Large-scale surveys such as the Rubin Observatory Legacy Survey of Space and Time will discover many more such objects and will provide an increased understanding of context and impact for these interstellar visitors.

2.1.2.4 The Star-Planet Connection

Stars and the planets that orbit them are inexorably connected to each other by their evolutionary history. The star’s properties and evolution influence the evolution and habitability of the planets, particularly of terrestrial planets. The end stages of star formation provide the initial conditions for planet formation. The coming decade will see the implementation of a systems-level approach to understanding the many factors that influence a planet’s habitability.

Most workhorse planet detection methods are a relative measurement, made with reference to the properties of the host star. Stellar surface inhomogeneities pollute the planet measurements due to the combined light of the system and the breakdown of simplistic assumptions about the stellar surface characteristics. Starspots and other variations produce spurious Doppler shifts that can mask Earth-sized planets. Precision planet characterization in the coming decade will motivate the need for better knowledge about the star, particularly its magnetic properties. Conversely, close-in planets can be used as test particles to reveal small-scale stellar inhomogeneities (Figure 2.10), and increase understanding of stars.

A star can influence its near environment by a combination of its radiation, gravity, and particles. The difference between “super-Earth” and “mini-Neptune” planets can potentially be explained by processes driven by the star. A star’s magnetic field is responsible for the heating producing stellar emissions above the photosphere, which manifests as enhanced ultraviolet through X-ray radiation. High energy stellar radiation is a risk to, and potentially a catalyst for, life. The UV emission of a star influences the planetary atmosphere photochemistry, and can create false positives for biosignatures. The extent and amount of high energy radiation from the star determines how much of the atmosphere a close-in planet keeps over evolutionary time. For planet-hosting stars which are cooler than the Sun in temperature, the magnetic field is also complicit in the star losing mass, through a steady stellar wind and potentially transient coronal mass ejections. Eruptive events, characterized as some combination of stellar flaring, coronal mass ejections, and energetic particle events, produce space weather and in the extreme events, influence habitability. Magnetic fields of low-mass stars may prevent some eruptive events from ejecting mass, lending complexity to a blind application of the solar analog, and there are currently few observational constraints on stellar winds or stellar coronal mass ejections. The next decade will see progress in characterizing the habitability prospects, particularly of M dwarf planets. Due to their proximity to the star, they are the most vulnerable to atmospheric loss, coronal mass ejections, tidal heating, and orbital evolution.

Priority Science Area: Pathways to Habitable Worlds

Among the many exciting discoveries and opportunities ahead within the *Worlds and Suns in Context* theme, one question stands out: “Are there habitable planets harboring life elsewhere in the universe?” The answer to this question, and the questions which immediately follow—“Is the Earth unique?” “Are humans alone?”—would have impacts reaching far beyond astronomy, within science and humankind overall. Advances in the understanding of exoplanets and in astronomical instrumentation now allow the planning of major steps towards identifying and studying candidate habitable planets, and making the first tests for the signatures of habitability and life. Laying out the Pathways to Habitable Worlds is the priority science area for this theme.

Permission Pending

FIGURE 2.10 Surface inhomogeneities on a star can be revealed by departures from the expected depth of transiting planets, as in the case of a dark patch and a different light patch on the star as revealed in this example from the transiting planet WASP-19b. SOURCE: Espinoza et al. (2019).

Meeting this ambitious goal will require progress on a variety of fronts, observationally, theoretically, and in the laboratory. Potentially habitable planets are the exception rather than the rule (roughly 20 out of the ~2,400 planets discovered by Kepler, Sec. 2.3.2), and those close enough to be studied in detail by future facilities is smaller yet, so current and planned exoplanet surveys will play a critical role in enlarging the candidate list. Ongoing efforts to characterize the surfaces and atmospheric compositions of known exoplanets, from space and the ground, will be important for quantifying the demographics of the overall planet population and refining the diagnostic tools which will eventually be brought to bear on the candidate habitable worlds.

Observing and characterizing candidate habitable planets themselves, however, whether by direct imaging or indirect observations in transiting systems, requires new telescopes on the ground and in space. The first target systems are likely to be planets orbiting close to the lowest mass and faintest stars, M-dwarfs. Their low luminosities (thousands of times less than the Sun) make it easier to detect an orbiting planet against the stellar glare, but the habitable planet zones will lie close to the stars themselves. JWST may be able to measure a few such systems in transit, and spectroscopic observations of larger samples form a keystone part of the science cases for the 24-40m ELTs (E-ELT, GMT, TMT).

A longstanding goal ever since the discovery of the first exoplanets is to image and characterize planets orbiting Sun-like stars, including those in the habitable zone. Such stars have long been suspected as offering the most likely sites not only for liquid water but also life, and space-based telescopes with sufficient stability and high-contrast capabilities to image and spectroscopically measure planets around the nearest such stars are now within our reach. A space telescope similar in wavelength coverage to the Hubble Space Telescope, and with an aperture of at least 6m and coronagraphic imaging capability should be capable of observing approximately 100 nearby stars, and successfully detect potentially habitable planets around at least a quarter of the systems. Such an observatory would also provide valuable information on other extrasolar planets, and be versatile enough to carry out groundbreaking observations of stars, galaxies, black holes, and the gases and baryons within and between galaxies, with a scientific impact rivaling that of previous "great observatories" such as HST.

The potentially habitable worlds ripe for discovery in this decade and the next represent only the tip of a deep iceberg in exoplanetary exploration. Studies of Super-Earths and mini-Neptunes as well as larger Jupiter-sized planets fill in critical areas; these measurements can be done with a range of facility sizes and round out a comprehensive approach to the subject. Considering solar systems as a whole and understanding how the individual components interact with each other is a crucial component of understanding the processes at work to arrive at the observed state. While the ultimate quest is to find potentially habitable planets, the perspective of a wide variety of planetary demographics, characteristics, and evolutionary paths (including those that lead away from habitability) is needed to understand where the multiple highway exits lead along the route to answering the question "Are humans alone?"

The path to habitability starts at the beginning of a planet's journey and requires investigation of the chemical and dynamical processes at work to determine conditions early on in a planet's life that lead both towards and away from habitability. Improvements in imaging and spectroscopy with future large facilities at mm and sub-mm wavelengths will probe the rings and gaps caused by forming planets in the disk of gas and dust around a young star, and create a census of the properties of these forming planets. Spectroscopic probes of planet-forming gas reveal the chemical initial conditions that solar systems and individual planets in formation experience; the study of complex prebiotic species paints the picture of chemical evolution pathways needed at the beginning of this path. Measuring the water reservoirs in star- and planet-forming disks is crucial for understanding the mechanisms by which terrestrial planets gain their water.

Although this mapping of the Pathways to Habitable Worlds has emphasized the central roles to be played by major existing and planned observatories, success will also require a wide array of enabling and foundational projects and studies. These include observational and theoretical studies of the linking between star and planet formation, the chemical processes—both inorganic and organic—critical for planet formation, the evolution of planetary atmospheres, and the origins of life, laboratory measurements of spectroscopic tracers and astrochemical processes, and detailed study of stellar activities, variability, and surface structures, all of which can mimic observational signatures of transiting

2.2 NEW MESSENGERS AND NEW PHYSICS

Our understanding of the universe has been repeatedly transformed by observing across the entirety of the electromagnetic spectrum, from the radio to the gamma-rays. These observations can test or reveal physics in ways that are not possible on Earth. Now, nearly daily movies of the sky, as well as the new messengers of neutrinos, particles, and gravitational waves, provide new ways of doing astronomy and new methods for uncovering and testing new physics.

Observing the universe in new ways has historically involved expanding our observations to cover the full electromagnetic spectrum, not just the part accessible with our eyes. In doing so, X-ray observations revealed the dynamic coronae of the Sun and other stars, accreting neutron stars and black holes, and the hot plasma that pervades interstellar and circumgalactic environments. Radio observations revealed the existence of neutron stars, whose remarkably stable rotation rates have since been used to discover planets and confirm the theory of general relativity's prediction of orbital decay via the emission of gravitational waves. And telescopes observing in the infrared can peer into the enshrouded stellar nurseries where stars and planets form.

New views of the universe also come from observing in new ways—for example, by making observations with much higher sensitivity, angular resolution, or time resolution, or by obtaining higher dynamic range views of previously hidden phenomena. The movies of stars orbiting the 4 million solar mass black hole in the center of the Milky Way Galaxy would not have been possible without adaptive optics and near-infrared interferometry to overcome the blurring effects of Earth's atmosphere. More recently, the Event Horizon Telescope's unprecedented angular resolution at millimeter wavelengths (via interferometry) provided the first direct image of the near-horizon environment of a black hole (see Figure 2.11), captivating scientists and the public alike.

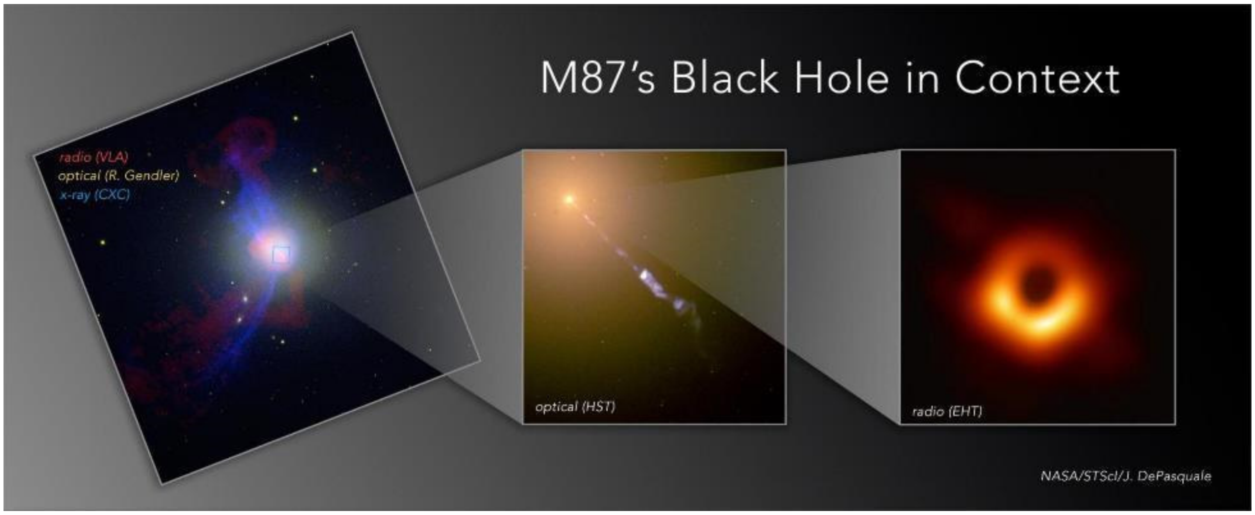

FIGURE 2.11 The galaxy M87 (*left*) harbors a 6.5 billion solar mass black hole at its center that produces a spectacular jet (*middle*). The unprecedented angular resolution at millimeter wavelengths of the Event Horizon Telescope produced the image of electromagnetic radiation from plasma near the horizon of the black hole (*right*). The image shape is interpreted as being due to motion of the radiating plasma at near the speed of light (producing the brighter lower half) and strong gravitational lensing by the black hole (producing the 'shadow of the black hole,' the deficit of light at the center). SOURCE: *Left:* Adapted from Chandra X-Ray Observatory, "M87: A Nearby Galaxy Metropolis," https://chandra.harvard.edu/photo/2008/m87; X-ray: NASA/CXC/CfA/W. Forman et al.; Radio: NRAO/AUI/NSF/W. Cotton; Optical: NASA/ESA/Hubble Heritage Team (STScI/AURA), and R. Gendler; Middle: Adapted from Hubblesite, "Black Hole-Powered Jet of Electrons and Sub-Atomic Particles Streams from Center of Galaxy M87," https://hubblesite.org/contents/media/images/2000/20/968-Image.html, NASA and The Hubble Heritage Team (STScI/AURA). Right: Adapted from The Event Horizon Telescope Collaboration, "First-ever Image of a Black Hole Published by the Event Horizon Telescope Collaboration," https://eventhorizontelescope.org/blog/first-ever-image-black-hole-published-event-horizon-telescope-collaboration.

By observing ever fainter sources across the electromagnetic spectrum, it is also possible to peer back into the distant past when the universe and its constituents were young, and unravel the history of the universe. Perhaps most spectacularly, observations of the cosmic microwave background radiation measured the geometry and mass-energy content of the universe and helped transform cosmology into a precision science.

The most radically different views of the universe are provided by messengers other than electromagnetic radiation in the form of photons or waves. Neutrino observations confirmed the theoretical prediction that hydrogen fusion powers the luminosity of the Sun, and demonstrated that neutrinos have mass, a key insight into physics beyond the standard model of particle physics. Cosmic-ray measurements have found particles whose energies are enormous compared to those that can be produced in the Large Hadron Collider (LHC) at CERN, but whose astrophysical origin remains a mystery. In 2013, the south pole IceCube observatory detected a diffuse high energy neutrino flux of astrophysical, but unknown, origin. Starting in 2015, LIGO opened up the gravitational wave view of the universe by detecting merging binary black holes. The simultaneous detection of gravitational waves and electromagnetic radiation from a binary neutron star merger in 2017 showed the power and complementarity of multi-messenger observations (see Box 2.2).

As views of the universe have expanded so has astronomy's impact on basic physics. Many frontier science questions identified by the Survey's science panels center on the intertwined themes of using new techniques to see the universe in new ways (new messengers) and uncovering new physics with advanced astronomical observations. In what follows the presentation of this science theme is organized by first discussing cosmology and the dark sector and then turning to the new astronomy enabled by observations with particles, neutrinos, gravitational waves, and light.

2.2.1 Cosmology and the Dark Sector

The fundamental paradigm of modern cosmology is the Hot Big Bang, in which an initially hot, dense, and nearly smooth universe rapidly expands and cools. All evidence suggests an initially simple universe, made of a nearly uniform collection of light nuclei and electrons, a sea of radiation, a similar sea of cosmic neutrinos, and dark matter of an unknown nature. As time passed, small initial differences in density grew under the action of gravity to form the rich structure described in the *Cosmic Ecosystems* science theme of this report. The transition from a smooth universe to one with stars and galaxies occurred less than 500 million years after the Big Bang. Finding innovative ways to probe cosmology in the "dark ages" prior to any significant star formation is one of the discovery areas identified here. The development of "LambdaCDM," our current standard cosmological model is one of the major intellectual triumphs of the past few years; the nature and origin of its key ingredients—dark matter, dark energy, and a nearly scale-invariant spectrum of primeval mass fluctuations—remain, however, some of the biggest mysteries in science.

Observations of the motions within galaxies and clusters to the large-scale distribution of intergalactic gas and the CMB require more matter than can be explained by the observed baryons. In the common interpretation, this is cold dark matter, an unseen gravitating material—nearly 10 times more abundant than baryonic matter—that moves non-relativistically in the recent universe. Over the many decades since dark matter was posited by Fritz Zwicky to explain galaxy motions in the Coma cluster and by Vera Rubin to understand galaxy rotation, astronomers have learned primarily what it is not: not like anything that has been seen on Earth. It could be a particle with the mass of a proton that does not significantly interact with normal matter or, at another extreme, it could be a "particle" with a quantum-mechanical wavelength the size of a small galaxy. While physicists attempt to identify dark matter through ambitious and excruciatingly careful laboratory experiments, astronomers in the coming decade will wield the threefold tools of theory, simulation, and observations to search in parallel. Dark matter could leave detectable traces of its potentially more complex interactions through deviations from the simplest version of the cold dark matter paradigm or through emission of unexpected particles (gamma-rays, positrons, narrow radio frequency lines) produced by dark matter interactions. The signatures of complexity in the properties of dark matter could come from the most distant sources (e.g., the CMB) or some of the nearest (e.g., nearby wide stellar binaries or the internal kinematics of dwarf galaxies). Nearly all astronomical facilities—current, imminent, and aspirational—contribute to the study of dark matter. New large ground-based optical-infrared telescopes would be particularly impactful, e.g., by studying the internal motions of stars in dwarf galaxies and testing the nature of the dark matter that holds those galaxies together.

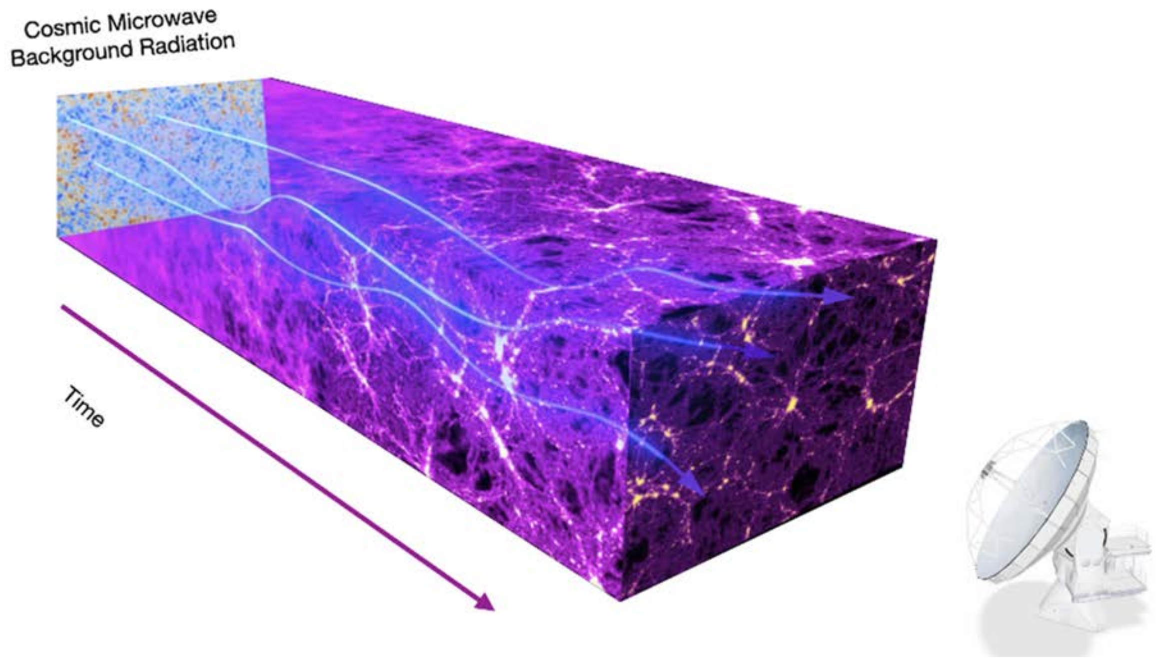

FIGURE 2.12 The cosmic microwave background (CMB) can be thought of as a backlight at the edge of the observable universe as depicted above. Light from it traverses the universe on its way to us on the right. Concentrations of mass, shown here as the brighter areas in the web of dark matter, deflect the light through the process of gravitational lensing. The image of the CMB that finally reaches Earth is then a distorted view of the true CMB, like the view through an imperfect piece of glass. Measuring these distortions determines the three-dimensional mass distribution that the light has passed through, which is shaped by the distribution of dark matter throughout the universe. The light of distant galaxies can be distorted in a similar manner. SOURCE: *Left:* Copyright ESA and Planck Collaboration. *Right:* ALMA antenna, https://public.nrao.edu/telescopes/alma/, National Radio Astronomy Observatory, Associated Universities, Inc., and the National Science Foundation.

Even with the current poor understanding of what dark matter *is*, there is broad consensus around how it behaves gravitationally on large scales. This allows theoretical models to robustly connect the final distribution of dark matter to the conditions in the very early universe. Astronomers are on the brink of using this connection to make 3D maps of the dark matter distribution and use those maps as probes of fundamental physics. The properties of the dark matter distribution can be traced by the ways it subtly shifts light rays as they propagate through the universe, through the effect of gravitational lensing (Figure 2.12). Analyzing the fine details of maps of the CMB or of the shapes of distant galaxies, reveals the 2D matter distribution, which can then be deprojected into the full 3-dimensional view. This is one of the primary goals of Roman, Rubin, Euclid, and new more powerful CMB instruments. The statistical properties of these maps will eventually be able to measure the sum of the masses of the neutrinos, which will be yet another major contribution of astronomical observations to basic physics. In parallel with lensing studies on large scales, more detailed, focused studies of individual lensed galaxies, supernovae, and quasars (accreting black holes), and even lensed stars at cosmological distances, will test the predictions of the cold dark matter paradigm on the smallest scales, where we expect deviations to be most evident.

Along with dark matter, the second great cosmological unknown is the nature of the "dark energy." Observations of the recent universe have shown that its expansion is accelerating. This remarkable discovery is not explained by a model containing only matter (even dark matter), but instead indicates a new feature, dubbed dark energy. Arguably the simplest explanation for dark energy is Einstein's cosmological constant—an energy and pressure characterizing empty space whose gravitational effect drives the acceleration—but the incredibly small value, compared to what is expected from quantum theories, leads physicists to think that dark energy may be more complicated. The cosmological constant predicts a specific form for the acceleration in the expansion of the universe as a function of time. Observations in the coming decade using Rubin, Roman, Euclid and higher resolution, higher sensitivity CMB facilities will test whether observations are, or are not, consistent with a cosmological constant. These results will be a major legacy of the *New Worlds, New Horizons* decadal survey. Any deviations from a cosmological constant model could touch off a second revolution as powerful as the initial discovery of the accelerating universe itself.

While the next generation tests of dark energy are underway, astronomers are already finding hints that the current cosmological model may be incomplete. An unexpected tension has developed between two different ways of determining the present expansion rate of the universe. In one, based on measuring distances and velocities in the local universe, the current expansion rate is roughly 5 percent higher than inferred from the second method, which uses the standard cosmological model to extrapolate from early universe CMB measurements to today. Continued measurements from satellites and the ground, including those using gravitational waves, will be able to distinguish between a systematic effect in one of the methods or the need for a previously unknown component of the universe.

The fluctuations seen in the CMB are believed to have been imprinted in the earliest phases of the Big Bang during a period of cosmological inflation in which extraordinarily rapid expansion established the large-scale homogeneity and flatness of the universe while also causing quantum fluctuations which subsequently grew into the fluctuations we observe. One of the most exciting opportunities in the coming decade is that CMB measurements may reveal remnant gravitational waves from this early epoch, as depicted in Figure 2.16 below. The presence of gravitational waves would manifest as a distinctive polarization pattern in the CMB, called "primordial B-modes," at angular scales of 3 degrees and larger. These scales are accessible from the ground and space. If primordial B-modes are found they will provide critical constraints on the physics of inflation and give new insights into physical processes at energy scales orders of magnitude larger that can be attained at the LHC. Efforts are already underway to detect them, but higher angular resolution and sensitivity CMB observations from the ground and in space will be needed given the difficulty of detecting the small B-mode signal amidst the polarized galactic foreground.

In summary, the standard cosmological model is both a remarkable triumph and an astonishing puzzle. With a relatively modest number of parameters, it continues to match observational results despite orders of magnitude improvement in cosmological measurements over the past twenty years. However, there is a mystery novel's worth of hints that the same model is incomplete, as the most important components are not yet understood. This represents one of the great contributions of astronomical observations to basic physics, and a continued opportunity for breakthroughs and surprises in the coming decades. Unraveling the many cosmological mysteries will require a particularly close interplay between theory, simulation, observations, and laboratory experiments.

2.2.2 New Messenger

Astronomy has long been a science rooted in the observation of light (photons). The past decade, however, has overturned this understanding of what astronomy is and could be, thanks to observations with new messengers that carry new information about the workings of the universe. Gravitational waves, neutrinos, and cosmic rays (Figure 2.13)—long viewed as largely the province of physics—have all now passed into the realm of astronomy. This is due to multiple breakthrough discoveries in the last decade, using facilities such as Auger, IceCube, and LIGO. At the same time, astronomy's traditional pursuit of photons is being transformed by new observational facilities that probe time variable and transient phenomena. Characterizing the time-variable electromagnetic universe has become increasingly sophisticated, thanks to pathfinding optical telescopes such as the All-Sky Automated Survey for Supernovae (ASAS-SN) and the Zwicky Transient Facility (ZTF) that are setting the stage for the Rubin in the coming decade. Outside the optical, dedicated radio surveys with the Karl Jansky Very Large Array (JVLA) and the Canadian Hydrogen Intensity Mapping Experiment (CHIME), and other international facilities are uncovering new and unexpected phenomena, such as fast radio bursts, while high energy space telescopes sensitive to explosive events like gamma ray bursts become ever more central to interpreting signals from new cosmic messengers. It is not an exaggeration to say that nearly daily movies of the sky made across the electromagnetic spectrum are their own form of "new messenger," with information that is fundamentally distinct from a static view of the universe. This led to the identification of time domain astronomy as a key discovery area in *New Worlds, New Horizons*.

These seemingly separate advances in observational techniques are in fact intimately related: most of the known and anticipated sources of gravitational waves, neutrinos, and cosmic rays are also time variable or transient electromagnetic sources (e.g., neutron star mergers, gamma-ray bursts, black hole jets, and stellar explosions). Combining information from all messengers can unravel the physics at the heart of these objects, as was demonstrated so spectacularly in the case of the binary neutron star merger GW170817 (Box 2.2).

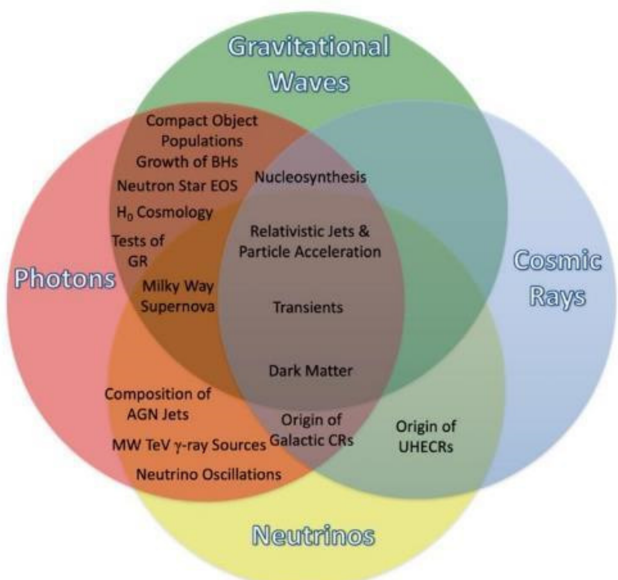

FIGURE 2.13 The combination of multiple messengers provides unique insights into astrophysical sources, particularly those involving strong gravity or relativistic motion (see Box 2.2). SOURCE: Report of the Panel on Particle Astrophysics and Gravitation.

BOX 2.2 Multi-Messenger Astronomy

The binary neutron star merger GW170817, detected both in electromagnetic and gravitational waves, was a watershed event that confirmed the long-anticipated promise of multi-messenger astronomy. The confluence of decades of work in theory, numerical relativity, nuclear astrophysics, gravitational wave detectors and analysis methods, combined with measurements across the electromagnetic spectrum from space and ground, building on the international network for rapid follow up originally developed for the study of gamma ray bursts, produced what has become the archetype for multi-messenger astronomy, a field destined to blossom in the coming decade.

In GW170817, gravitational wave measurements determined the mass of the merging neutron stars and an initial sky localization, while electromagnetic observations determined the host galaxy of the merger and the mass, speed, energy, and composition of matter ejected from the system during the merger (Figure 2.2.1). This ejecta consisted of both a jet of relativistic material that powered non-thermal radiation from the radio to the gamma rays and slower more spherical ejecta that powered the thermal optical and infrared light. The data indicated that the latter was a “kilonova,” optical/infrared light powered by nuclear decays involved in the production of many of the heaviest elements in nature. Indeed, the optical-infrared light provided strong evidence that neutron-star mergers are a significant astrophysical site for the production of rapid neutron capture elements (including the rare Earth metals, platinum, and gold), a long-standing mystery in our understanding of the origin of the elements traced in the spectra of stars.

Permission Pending

FIGURE 2.2.1 Schematic of the binary neutron star merger GW170817 observed in gravitational waves and light (radio to gamma-rays). SOURCE: Margutti and Chornock (2021), ARAA.

The combination of a gravitational wave distance to the merger and a redshift in the spectrum of the host galaxy also allowed a fully independent measurement of the Hubble constant, the value of which is a source of uncomfortable cosmological tension and in need of new measurements (See Section 2.2.1). Although the single measurement with GW170817 is not as precise as other techniques, multi-messenger cosmology will increase in importance in the coming decade as we detect ever more binary neutron star and black hole mergers.

Observations of SN 1987A and the Sun in light and neutrinos and GW170817 in light and gravitational waves revealed the power of multi-messenger astronomy. This is but a taste of the feast that is to come. Higher sensitivity high-energy neutrino experiments will detect individual astrophysical sources. A supernova in the Milky Way Galaxy would be a multi-messenger, multi-wavelength goldmine, detectable in light, neutrinos and possibly gravitational waves. A next generation ground-based gravitational wave network could detect and localize every solar-mass binary black hole merger in the universe, transforming astronomy and cosmology. Pulsar timing and the space-based interferometer LISA will open up other parts of the gravitational wave spectrum, revealing new sources and new surprises, much as the first X-ray and radio telescopes did. Capitalizing on these opportunities will also require a new generation of theoretical and computational models that combine general relativity, nuclear astrophysics and plasma physics. Likewise, new electromagnetic facilities are needed, including those with transient capabilities, larger ground-based optical-infrared telescopes for detailed spectroscopic follow-up, new cm-wavelength radio arrays for observing the non-thermal radiation from jets, and new space-based satellites to provide critical pieces of the picture missing from the ground (e.g., the ultraviolet and gamma rays).

2.2.3 Neutrinos and Cosmic Rays

Trillions of neutrinos stream through us from the Sun every second, with a similar number from the cosmic neutrino background. This is an indication of just how difficult neutrinos are to detect relative to how common they are, but in this difficulty lies their promise. Precisely because neutrinos interact so little with matter, neutrinos carry information about the inner workings of some of nature’s most important energy sources: the nuclear furnaces inside stars, the formation of neutron stars in stellar explosions, and the conditions in the jets of relativistic particles that originate near the event horizons of black holes in galaxy nuclei. Neutrinos that are detected also point directly to the celestial position of their source. In the next few decades, astronomical observations will begin to routinely measure the cosmos with these most elusive of particles. Here it is important to distinguish between cosmological neutrinos (which, like the CMB, have redshifted to millielectronvolts), neutrinos emitted by fusion processes in stars, which are in the megaelectronvolt range, and neutrinos in the gigaelectronvolt range and above, which are produced by hadronic collisions between cosmic rays and ambient matter. We focus on the lattermost here, as they are an unambiguous signature of ion acceleration, follow straight line orbits from their source to our detectors (unlike cosmic rays themselves, which are scrambled by magnetic fields), and unlike their photon counterparts, are immune from optical depth effects.

Theoretical models of the early phases of the Big Bang predict that there should be a nearly uniform cosmic neutrino background similar to the CMB. There are laboratory experiments aimed at detecting the cosmic neutrino background but it is an extremely challenging measurement. In contrast to the sea of low energy neutrinos produced in the Big Bang, higher energy neutrinos are messengers from some of Nature’s most dramatic events. When the relatively nearby supernova 1987A exploded, some 25 neutrinos were measured in the Kamiokande, IMB, and Baksan neutrino detectors, which were state of the art for their times. Thirty-five years later, modern neutrino detectors would measure tens of thousands of neutrinos from a supernova in the Milky Way. As similar supernovae explode throughout the universe, they collectively produce a diffuse background of megaelectronvolt (MeV) neutrinos, one that should be within reach of forthcoming experiments if current theories are correct. Detection of either a galactic supernova or the diffuse background would test models of the formation of neutron stars and black holes and the core-collapse explosion mechanism. Only neutrinos and gravitational waves can peer into the inner regions of these events where densities reach nuclear scales.

The skies are full of even more energetic sources that likely produce neutrinos, some continuous, some episodic, and some transient. For example, relativistic jets—collimated beams of ejected material moving at nearly light-speed—are known to emanate from supermassive spinning black holes in active galactic nuclei (see Figure 2.11). These jets span up to millions of light years in extent. Similar relativistic jets are also associated with gamma ray bursts (GRBs), extremely energetic events that emit in just seconds the same amount of energy the Sun will emit over its lifetime. This rich population is likely to be a significant source of neutrinos detectable with future facilities.

The prospects for future neutrino astrophysics are promising. Over this past decade, the IceCube experiment at the South Pole detected an unresolved extragalactic background of 60 neutrinos with energies in the teraelectronvolt to petaelectronvolt (TeV-PeV)[2] range. Their distribution on the sky indicates that they are produced by distant sources well outside our galaxy, and thus are likely to be by products of energetic events throughout the universe (much like diffuse X-ray and gamma ray backgrounds). To date, only the Sun (Figure 2.14), SN1987A, and potentially one blazar have been imaged in neutrinos of any energy, but with new facilities the excitement of identifying specific individual sources is likely just beginning.

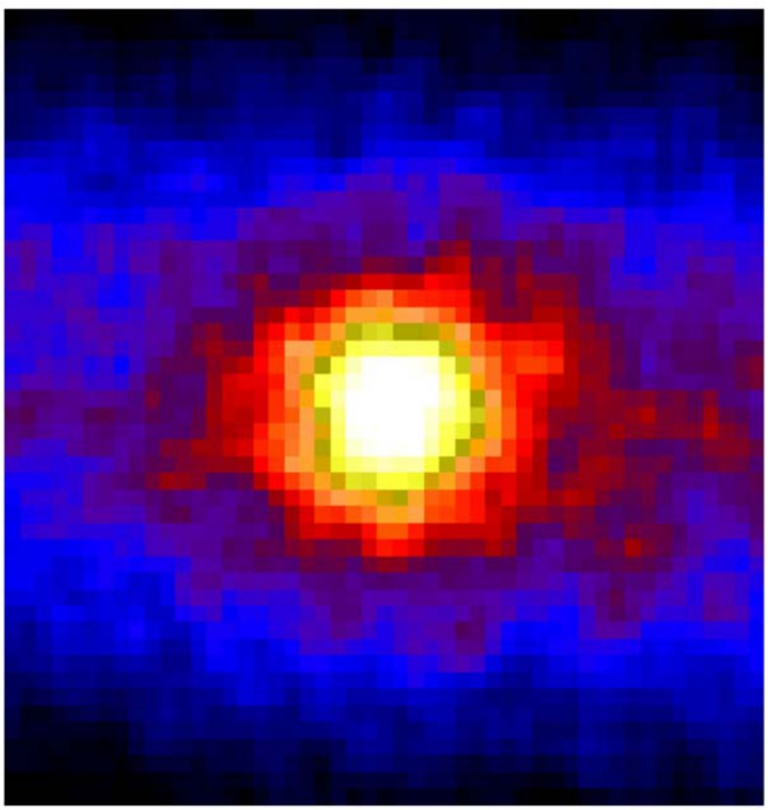

FIGURE 2.14 An image of the Sun taken with the Super-Kamiokande neutrino detector in Japan. These data were collected from neutrinos emitted from the core of the Sun, which after traversing the Earth-Sun separation, travelled through Earth to reach the detector. SOURCE: http://strangepaths.com/the-sun-seen-through-the-earth-in-neutrino-light/2007/01/06/en/, Kamioka Observatory, ICRR (Institute for Cosmic Ray Research),The University of Tokyo.

[2] A teraelectronvolt (TeV) is 10^{12} eV, a petaelectronvolt (PeV) is 10^{15} eV, and an exaelectronvolt (EeV) is 10^{18} eV. The protons in the Large Hadron Collider in CERN have an energy of about 10 TeV, orders of magnitude below those of the highest energy cosmic rays.

Neutrinos are not the only messengers of relativistic and energetic astrophysical phenomena. Quite independently, other experiments have detected ultra-high-energy cosmic rays (UHECRs, in the ~10^{18} eV [EeV] range). What produces these remarkably energetic particles? Are the TeV-PeV neutrinos and ultra-high-energy cosmic rays produced in the same sources? They are widely surmised to be accelerated in the relativistic jets of accreting supermassive black holes or gamma-rays bursts, but this has yet to be tested observationally. Unfortunately, direct identification of a cosmic-ray source is difficult, since UHECRs are charged particles and are thus deflected as they travel through magnetic fields that permeate the universe. However, a clear directional signature would be provided by the high-energy neutrinos that the UHECRs produce in the regions where they are accelerated. Higher sensitivity neutrino observations with better sky localization are critical for unraveling how nature's most extreme particle accelerators work.

2.2.4 Gravitational Waves

Gravitational waves are the newest detectable messenger traveling through the astronomical landscape. They provide a unique probe of regions with large amounts of mass moving at near the speed of light: black hole and neutron star collisions, neutron star and black hole formation in stellar explosions, and the first fractions of a second of the Big Bang (Figure 2.15). The importance of gravitational waves lies in part with the central role that black holes and neutron stars play in many areas of astronomy, from stellar evolution to galaxy formation. In just the five years since LIGO's first detections were announced, gravitational wave measurements have already left astronomers in awe, with insights into the origin of neutron-rich elements (Box 2.2), the detection of stellar-mass black holes much more massive than previously known, and new tests of gravity in the strong field regime.

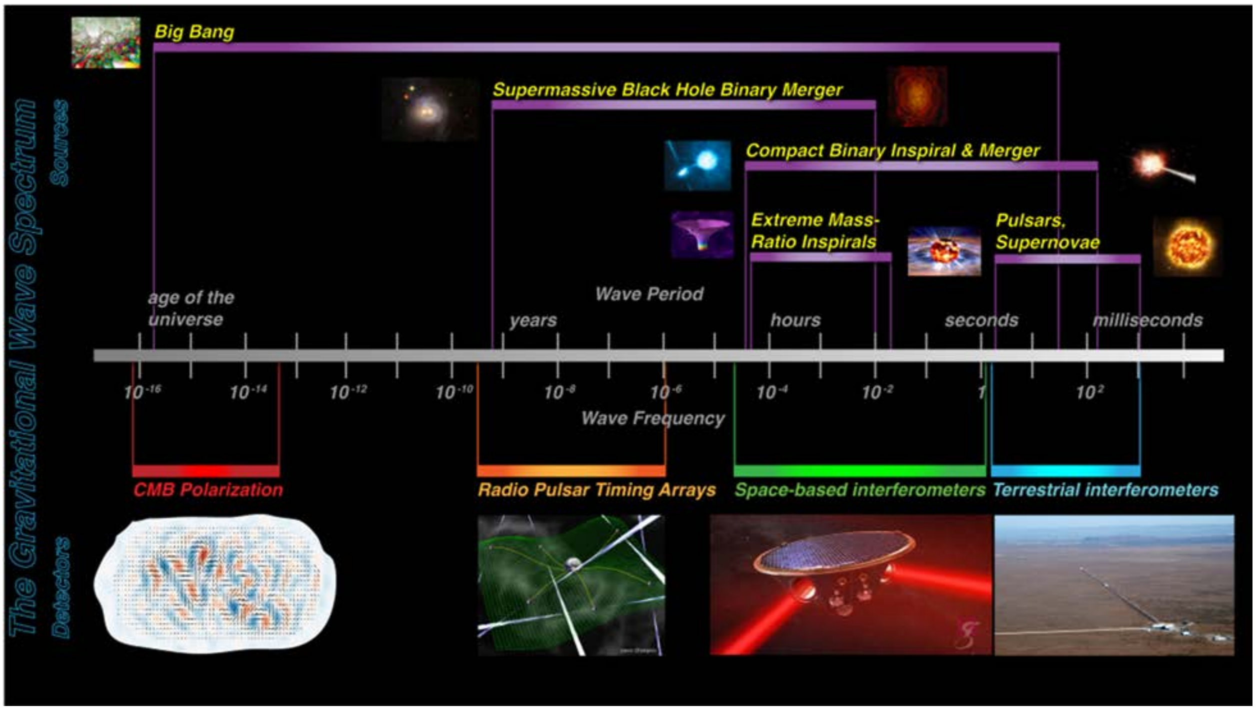

FIGURE 2.15 The full spectrum of gravitational-wave emission. The diversity of objects and events at the top of the image produce gravitational waves of different frequencies that need different detection instruments shown at the bottom. SOURCE: NASA/J. I. Thorpe.

Similar to the electromagnetic spectrum, a variety of astronomical phenomena produce a broad frequency range of gravitational waves (see Figure 2.15). The kilometers-long ground-based detectors LIGO and Virgo are sensitive to signals from relatively small systems: neutron stars and stellar mass black holes with radii from 10 to a few hundred kilometers, and with masses up to several hundreds of solar masses. Colliding massive black holes with millions of solar masses at the centers of galaxies and inspiraling white dwarfs are thousands to millions of kilometers in extent, and need space-based detectors, such as the LISA mission led by ESA with NASA contributions, scheduled to launch in the mid 2030's. To detect the gravitational waves from yet larger objects, like black holes weighing billions of solar masses, requires galaxy-sized baselines. This is done with high precision timing of the arrival of radio pulses from pulsars distributed throughout our galaxy. The loss of Arecibo is a setback for pulsar timing experiments, and underscores the need for radio facilities that can continue this critical science.

The larger and more sensitive ground-based gravitational wave interferometer network anticipated in the coming decade will detect a large number of stellar-mass black hole mergers and determine their masses and spins. Careful comparison of the properties of these mergers with theoretical models will inform which arise from binary or triple stellar evolution, which arise in dense stellar systems like globular clusters, and which might have entirely different origins. The population of neutron star and black hole mergers with masses of ~2-5 M_{sun} in the 'mass gap' between neutron stars and black holes will provide key constraints on our understanding of massive stellar evolution, the maximum mass of neutron stars, and core-collapse explosion physics.

Combined gravitational wave and electromagnetic observations have the potential to finally crack the longstanding puzzle of the origin and growth of massive black holes, one that lies at the intersection of the understanding of stars, galaxies, accretion disks, and cosmology. LISA and future ground-based gravitational wave detectors will detect black hole mergers in the earliest phases of galaxy formation less than a billion years after the Big Bang, while LISA, ground-based detectors, and pulsar timing arrays will constrain the merger rate over cosmic time for a range of black hole masses. Combining this gravitational wave data with deeper electromagnetic observations of accretion onto black holes, particularly in the infrared (e.g., JWST) and X-ray, will inform whether massive black holes originate from the remnants of massive stars. Are they built up by stellar and black hole collisions in dense star clusters, or perhaps they formed from the direct collapse of gas clouds in some of the first galaxies? Gravitational wave measurements will also be critical for disentangling the role of mergers and gas accretion in growing black holes over cosmic time.

The simultaneous detection of gravitational waves and light from the binary neutron star merger GW170817 was a transformative event (Box 2.2). More sensitive ground-based gravitational wave interferometers will provide a much larger sample of binary neutron star and neutron star-black hole mergers. The light from GW170817 was powered by material ejected during the merger and from the accretion disk left behind after the merger. Theoretical models predict that there will be considerable diversity in the electromagnetic counterparts to such events depending on the total mass of the binary neutron star system, the mass ratio, and the equation of state of dense nuclear matter, which sets the maximum mass of a neutron star beyond which it collapses to a black hole. Characterizing this diversity using observations of light and gravitational waves will be critical for unraveling the physics of accretion and jet production in binary mergers and their role in the origin of the heavy elements. Given the larger distance to gravitational wave sources as the facilities become more sensitive, this will require new observational capabilities for electromagnetic follow up. Large ground-based optical-infrared telescopes for spectroscopy to characterize heavy element production and sensitive ground based cm wavelength interferometers and space-based high energy telescopes to characterize relativistic jets will be particularly important. The science produced by joint gravitational wave and electromagnetic observations is likely to extend to at least a subset of LISA sources, namely binary black hole mergers in gas-rich galaxies. This will enable detailed studies of host galaxies, reveal the role of gas in facilitating massive black hole mergers in galactic nuclei, and provide a sample of black hole mergers at high redshift suitable for cosmology.

Part of the power and promise of gravitational wave measurements is their ability to simultaneously enable “new astronomy” and “new physics.” Tests of General Relativity using gravitational wave measurements are in their infancy. In the coming decades these tests will become far more stringent, with increasingly more precise measurements either cementing the quantitative applicability of General Relativity in the strong field regime, or perhaps revealing new physics. Sensitive ground-based gravitational wave detectors at higher frequencies and louder signals will constrain the radii of neutron stars through their tidal deformation. This in turn will constrain the equation of state of nuclear matter better than with current measurements, and in a way that is not possible in laboratories on Earth. Gravitational wave constraints on the neutron star equation of state will complement ongoing electromagnetic efforts using radio and X-ray timing and X-ray spectroscopy, likely leading to high precision measurements of neutron star radii in the coming decade. Larger samples of binary black hole and neutron star mergers will also significantly increase their utility for cosmology. Gravitational wave detections determine the distances to sources. Simultaneous electromagnetic observations of host galaxies to determine redshift can thus provide a measurement of the Hubble constant that is completely independent of current techniques.

Gravitational wave astronomy started with a bang in 2015, opening a new window to the universe with the detection of merging black holes by LIGO/Virgo. Over the few years since then, gravitational wave observations have become an indispensable astronomical tool. The coming decade, with the potential of detections in other parts of the gravitational wave spectrum, signals from new sources, and large numbers of black hole and neutron star detections, will be the start of a new era of precision and multi-wavelength gravitational wave astronomy.

2.2.5 Astronomical Transient Events

Although the night sky looks placid, millennia of observations have shown that it in fact varies systematically on many timescales. There are secular changes in the positions and speeds of objects due to their motion—be it interstellar interlopers in the solar system (see Section 2.1) or stars orbiting the 4 million solar mass black hole in the center of the Milky Way Galaxy. On top of this, there is a dizzying variety of dynamic astrophysical events that emit large amounts of energy in anywhere from the blink of an eye to timescales longer than human lifetimes. Some of these are cataclysmic events that herald the formation of neutron stars and black holes in stellar core-collapse, the thermonuclear explosions of white dwarfs, or the mergers of stars or compact objects. Others are repeating phenomena, such as stellar flares or the explosions of the surface layers of white dwarfs. All of these phenomena involve astronomical sources that produce light at many wavelengths, as well as in some cases gravitational waves and high energy neutrinos and cosmic rays. Astronomical transients impact nearly every area of astronomy. Supernovae and neutron star mergers disperse heavy elements into the interstellar medium, seeding gas with the elements necessary for life. Thermonuclear explosions of white dwarfs are valuable “standard candles” used to trace the acceleration of the universe. Fast radio bursts, millisecond bursts of radio emission of uncertain origin, have the potential to become a powerful new probe of the distribution of baryons throughout cosmic ecosystems (Section 2.3). Dedicated archives of brightness measurements now spanning more than 100 years enable the study of transient phenomena that recur on decade timescales or longer, such as stellar occultations by circumstellar material, the slow but dramatic brightening of newly unveiled young stars and of old stars like Betelgeuse in their death throes (see Section 2.1).

Rapid advances in detector technology and computing power have led to a revolution in astronomical time-domain surveys, which in turn have led to the discovery of new classes of transients (e.g., fast radio bursts and stellar mergers; Figure 2.16 depicts various classes of optical transients). The Rubin Observatory will take this effort to the next stage. Just one night of observation is expected to detect 10 million transient events with sub-arcsecond positional accuracy, sending triggers to telescopes around the globe and in space, to follow up with observations at other wavelengths and to correlate with observations using other messengers. This revolution is extending to a broader range of wavelengths outside the traditional optical and gamma ray bands. The Roman satellite will carry out a near-infrared supernova survey that will also likely discover new classes of infrared transients. Roman's microlensing survey will allow characterization for the first time of the mass function of the majority of neutron stars and black holes in the galaxy. The extended Roentgen Survey with an Imaging Telescope Array (eROSITA) is carrying out the first all-sky X-ray survey since the Roentgen Satellite (ROSAT) in the early 1990s. Additionally, there is a tremendous increase in the number of international radio facilities searching for radio transients (e.g., CHIME, the Australian Square Kilometre Array Pathfinder (ASKAP), MeerKat).

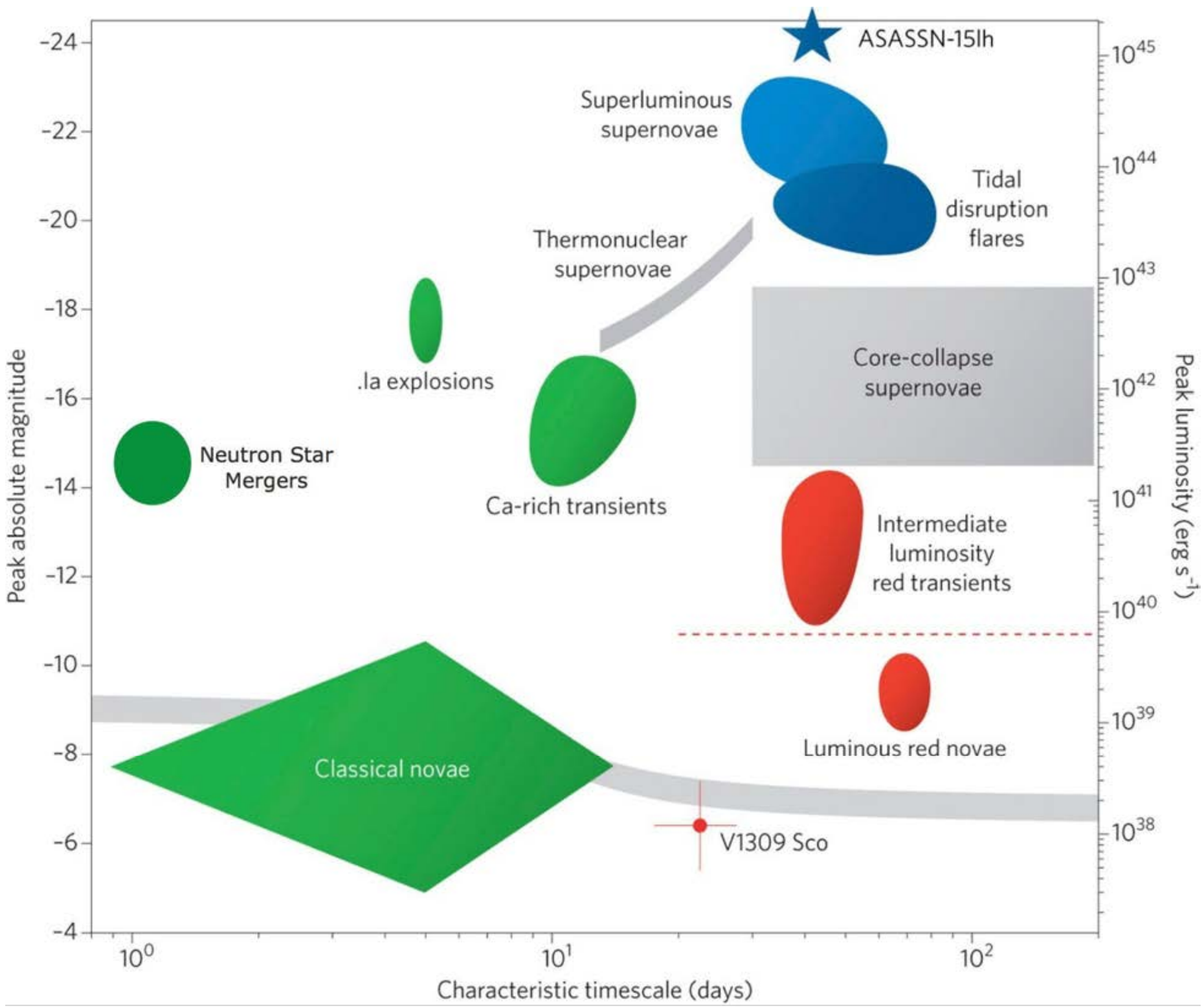

FIGURE 2.16 Types of optical transients, distributed in peak brightness and characteristic timescale, adapted from Cenko (2017). New wide-field optical surveys expected to come online this decade will surely populate this diagram with other transient exotica. SOURCE: Adapted by E. Quataert, with permission from Springer Nature: S.B. Cenko, 20187, Astrophysics: The true nature of transients, *Nature Astronomy* 1:0008, https://doi.org/10.1038/s41550-016-0008, copyright 2017.

The next decade has the potential to realize the full capabilities of transient and multi-messenger science, and very likely discover more surprises along the way. This work will ultimately provide a mapping between transients, the energy sources and central objects (e.g., black holes, neutron stars, stars) that power them, and their broader astronomical consequences (e.g., feedback and nucleosynthesis). Fully realizing this vision will, however, require a wide range of observational, theoretical, and software capabilities. The data required are obtained with simultaneous and coordinated operation of different instruments on Earth and in space. Ensuring easy user access to this wealth of data and the ability to cross-correlate multiple sources of data will maximize the science enabled by existing and planned facilities (see Sections 4.4 and 7.4.1). In addition, NASA's workhorse hard X-ray and gamma ray transient facilities (Swift and Fermi, respectively) are aging and their longevity is uncertain. Higher sensitivity all-sky monitoring of the high-energy sky, complemented by capabilities in the optical such as from Kepler and TESS, is a critical part of our vision for the next decade in transient and multi-messenger astronomy. Likewise, there are tremendous scientific opportunities for dedicated transient facilities at other wavelengths (e.g., the ultraviolet from space and radio on the ground) and dedicated spectroscopic follow up facilities, to complement the major U.S. investment in optical and near-infrared imaging surveys.

Priority Science Area: New Windows on the Dynamic Universe

The combination of new multi-messenger probes of astronomical phenomena with the maturation of time-domain observations opens up tremendous discovery spaces across nearly all areas of astrophysics. Within this discovery landscape, driven by improvements in gravitational wave and neutrino detection, and upcoming facilities such as the Rubin Observatory, one priority area stands out: the application of these new tools to the formation, evolution, and nature of compact stellar remnants such as white dwarfs, neutron stars, and black holes, as probed by the gravitational wave signatures of their mergers, together with rare explosive events that can be explored by the unique cadence and multi-color sensitivity of the Rubin Observatory. Sensitive observations of high-energy neutrinos and charged particles add new elements of discovery space, which will probe the universe's most extreme particle accelerators—New Windows on the Dynamic Universe.

The formation and evolution of compact stellar remnants, signaled by their accompanying multi-messenger transient phenomena are now serving to probe the range of neutron star and black hole masses in entirely new ways. These measurements provide information about the nature of matter under the extreme conditions that cannot be replicated in the laboratory, and about how the most extreme compact stellar remnants are formed and evolve. The mergers of neutron stars, uniquely observable at very early times through their gravitational wave signatures, can inform how elements such as gold and platinum are produced, which has been a mystery for many decades. Physical conditions near the surfaces and event horizons of neutron stars and black holes also represent extremes of matter, density, and gravitation, and serve as unique probes of fundamental physics. The coming revolution in temporal observations of the sky non-electromagnetic messengers, and through new observational cadences and spectroscopic follow-up across the electromagnetic spectrum is at the frontier of modern astrophysics.

As with the other priority science areas, progress will require coordinated advances in observations, experiments, and theory. The Vera Rubin Observatory and future large-area high cadence radio facilities will increase the numbers of variable and transient objects by orders of magnitude, with regular sampling over time. The Rubin Observatory will be unique for an optical time domain facility, as it will provide multiple optical colors at uniform cadence with unprecedented sensitivity. Advanced algorithms utilizing artificial intelligence algorithms to sift through Rubin Observatory's massive amounts of data will find the interesting outliers which will be heart of future progress and discovery. Current facilities such as the Chandra, SWIFT, and Fermi space observatories and ground-based radio and OIR observatories will play vital roles in such follow-up work. The full exploitation of this potential for multi-wavelength time-domain observations will require maintaining and expanding these observatories. Such facilities need not be large or expensive but need to be optimized for the task, and this survey recommends the establishment of such facilities both for space (including needed replacement of capabilities currently handled by aging facilities) and on the ground. Since this science tends to be global by its very nature, international cooperation both in complementary facilities and data sharing would greatly enhance the scientific outputs from these investments.

The major new facilities recommended by this survey will all play important roles in extending the power of these capabilities in the future. Many of the visible counterparts of these sources (often originating at cosmological distances) are extremely faint, beyond the limits of present-day telescopes, but should be within reach of the next generation of ELTs. The relativistic outflows produced by these events often can be readily detected in the radio, and will be prime targets for next-generation facilities such as the ngVLA (recommended for design studies by this survey). Design studies for a next generation ground-based gravitational wave observatory will set the pathway towards a revolutionary new facility in future decades. X-ray observations—critical to fully understanding the physics of these phenomena—motivate the design and construction of new facilities ranging from the scale of Explorers to a future large mission.

Foundational research is also essential for maximizing progress in this field. Chief among these will be support for theoretical modeling and simulations of these highly relativistic and energetic phenomena, including numerical relativity to determine the nature of the gravitational signatures, plasma physics to understand the particle acceleration, and dynamical modeling to determine the populations that could lead to compact object mergers. Efficiently assimilating the large flows of data from the surveys, extracting the measurements, and interpreting the observations pose major challenges, but with benefits that will expand our astrophysical knowledge in entirely new ways.

2.3 COSMIC ECOSYSTEMS

Processes on a wide range of time and length scales together drive the formation, evolution, and interaction of the remarkable diversity of objects we observe, from exoplanets and stars to black holes and galaxies. A confluence of advances in theory, computational modeling, and observational capabilities expected in the next decade will transform our understanding by identifying the key mechanisms shaping this web of interconnected systems.

2.3.1 Overview

Arguably the single most important lesson in the last ~30 years of understanding the origin of structure in the universe is that it is not a one-way street, dictated solely by gravity from large scales to small. The formation of some of the smallest and densest objects in the universe, stars and massive black holes, dramatically alters how most other astronomical objects form, from planets and galaxies to stars and black holes themselves. Stars and black holes impact their surroundings through a broad set of energetic processes collectively known as feedback. These span an enormous range of time and length-scales, from gas as close as the planet-forming disk around a young star to as distant as in another galaxy millions of light years away.

Many aspects of star and galaxy formation can be viewed as a cosmic tug-of-war between feedback and gravitational collapse, as illustrated schematically in Figure 2.17. It is now known that the luminous bodies of galaxies, far from being disconnected from their surroundings, are part of a vast system that includes their surrounding circumgalactic medium out to intergalactic scales. Theory predicts that giant rivers of gas flow into galaxies, but most of the gas in galaxies is subsequently ejected back out into the circumgalactic medium by powerful galaxy-scale outflows. The flow of matter and energy throughout the entire system is likely responsible for the commonalities and differences among galaxies, but the details of how have been elusive. Likewise, the flow of matter and energy within a galaxy—again due to the combined effects of gravity and feedback—determines the distribution of gas in the interstellar medium and where and how stellar and planetary systems form. The same flows depicted in Figure 2.17 also disburse the heavy elements produced by stellar processes, from the carbon in our bones to the rare-Earth metals in phones.

Understanding the interplay of gravitational and feedback-driven processes is challenging in part because it involves such a wide range of length and time-scales. In addition, much like understanding human health requires understanding how cells function, myriad small-scale physical processes regulate the flow of mass and energy illustrated in Figure 2.17, because they determine how gas cools, sheds its angular momentum, and mixes with other gas.

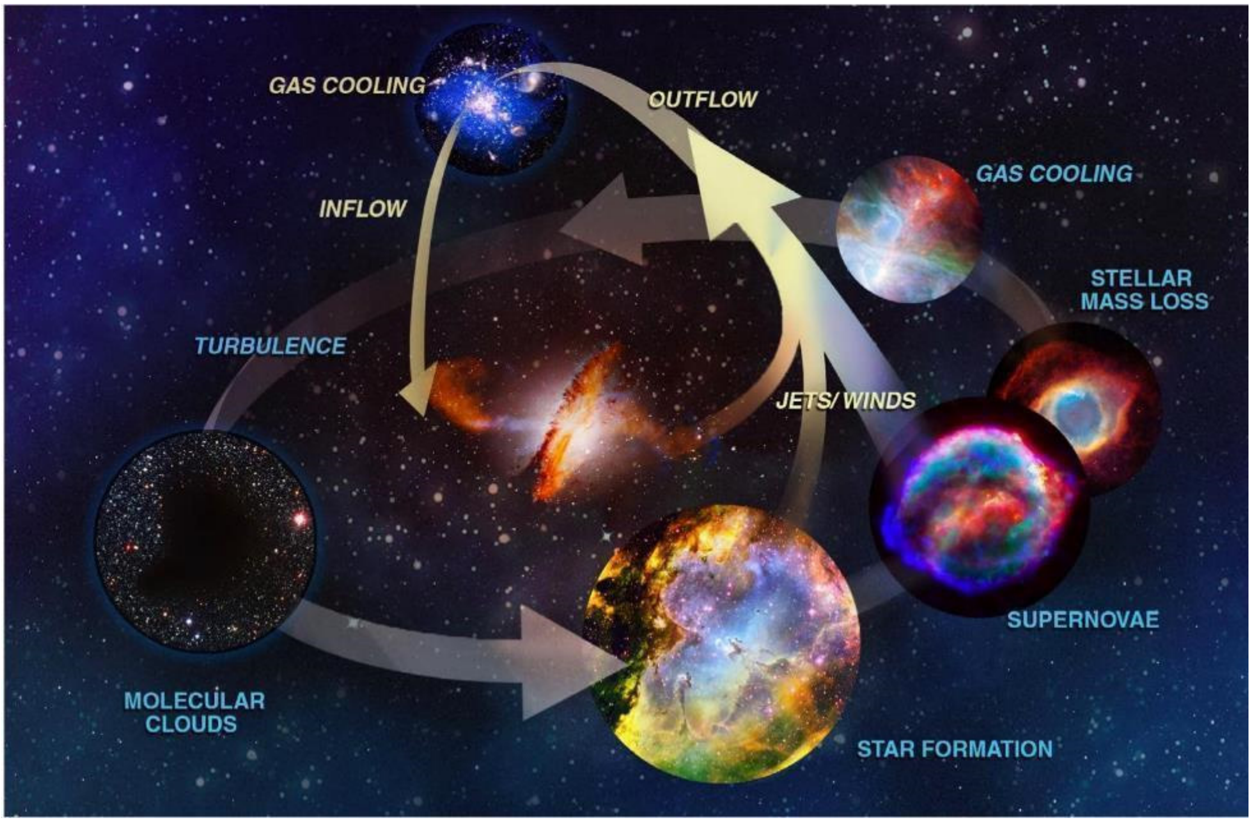

FIGURE 2.17 Illustration of the flow of gas into and out of the interstellar medium and galaxies through the combined effects of gravity and feedback. Heavy elements formed by fusion in massive stars are dispersed into the interstellar medium by stellar winds and supernovae. Much of this gas is in turn ejected from galaxies into the circumgalactic medium by galactic winds. Pristine gas accretes from the intergalactic medium to the circumgalactic medium, and subsequently accretes into galaxies, replenishing the fuel for star formation and subsequent generations of stars and supernovae. SOURCE: HABEX Report, The Habitable Exoplanet Observatory Study Team.

Small-scale physical processes at work in the cosmic ecosystem can thus have a surprisingly large impact on the large-scale behavior of astrophysical systems. An example is ionizing radiation, which is produced by massive stars and black hole accretion disks, and can regulate star formation and black hole accretion on sub-parsec scales. Some of this radiation can escape from the dense gaseous environments in which it was produced, and propagate out of galaxies into the intergalactic medium. In this way, the first stars and black holes were able to cause a global phase transition over scales of hundreds of Megaparsecs, in which most of the hydrogen in the universe was converted from a neutral to an ionized state, during what is referred to as the “Epoch of Reionization”. Identifying the sources of cosmic reionization, and better understanding how photons escape from their gas-rich and dust-enshrouded sources, will be an important science area over the coming decade. Numerical simulations of the propagation of radiation through galaxies (radiative transport) are highly computationally intensive, but critical for gaining a full picture of the role of this key process. Observational studies of gas kinematics, luminosity functions, and chemical compositions of galaxies spanning the Epoch of Reionization are needed, as well as studies of local galaxies that "leak" ionizing radiation, which may help shed light on the process of photon escape.

The symbiosis among cosmological phenomena on such different scales has been recognized for decades. Now, however, a confluence of advances in theory, computational modeling, and new observational capabilities will enable us to identify and understand the actual mechanisms at work in regulating this cosmic ecosystem. New observational probes of these interconnected flows at two widely separated spatial scales are highlighted in two of the discovery areas identified by the science panels: mapping the circumgalactic and intergalactic medium in emission and detecting and characterizing planets as they form.

2.3.2 Stellar and Black Hole Feedback

Stars have densities exceeding the mean density of the universe by more than 30 orders of magnitude; for black holes, the conditions are even more extreme. Despite their small sizes, stars and black holes are the most efficient sources of energy production yet discovered. This efficiency is a consequence of nuclear fusion in stellar interiors and the deep gravitational potential wells produced in stellar core-collapse and tapped by accretion onto black holes. Additional energy can be extracted at the end of a star's life. There is approximately one core-collapse supernova for every 100 M_{sun} of stars formed, and this single10^{51} erg supernova explosion can in principle accelerate 1000 M_{sun} of gas to speeds sufficient to unbind the gas from most galaxies. For accreting black holes the energy per unit mass is even higher. Thus, objects occupying a small fraction of the total mass and volume of the universe are critical to the evolution of much of the structure that is observed.

Many aspects of how stars live and die are currently uncertain enough that it limits the ability to model and interpret the effects of stellar feedback, be it in the form of radiation, stellar winds, or supernovae. Even in the local universe, for example, there are significant gaps in the understanding of stellar winds. This leads to large uncertainties in which massive stars become neutron stars and which become black holes. Stellar winds are also believed to play an important role in dispersing star-forming molecular clouds and driving turbulence in the interstellar medium, but the efficiency of this feedback again depends on the uncertain strength of winds from massive stars.

Even less is known about how the properties of stellar feedback vary with quantities like stellar metallicity, which strongly affects the efficiency of driving stellar winds, but varies from galaxy to galaxy and throughout cosmic time. Much of this uncertainty can be traced to the need for better theoretical models and observational diagnostics of stellar mass loss. More sensitive space-based ultraviolet (UV) spectroscopy is necessary to better characterize the spectra and winds of massive stars. Also, deep visible-wavelength and near-infrared spectroscopy of large numbers of stars in the Milky Way and other nearby galaxies from the ground—"industrial-scale" spectroscopy, one of the discovery areas identified by this Survey—would significantly sharpen constraints on stellar models.

Similarly large unknowns about feedback stem from uncertainties in binary stellar evolution. High mass binary stars evolve differently than single stars, which affects both their radiative output in life and the supernova explosions in which they die. Understanding binary stellar evolution is thus critical for understanding the global energetics of stellar feedback in galaxies. It is likewise important for understanding the spectra of galaxies, in particular the UV radiation that photoionizes the interstellar medium and likely reionized the universe during the initial epoch of galaxy and star formation. Separately, the dramatic advances in directly seeing the outcomes of binary stellar evolution with transient detection and gravitational wave facilities—for example, stellar mergers and compact object mergers—will continue to provide critical new insights into the life cycle of binary stars (see Section 2.2).

In addition to understanding the energy, mass, and momentum that stars supply to their surroundings, determining how these winds, radiation, and supernovae interact with the surrounding gas on different scales in the interstellar medium is equally important to untangling their interplay. Newly forming low-mass stars produce winds and jets that modify the structure of the clouds in which they form. They also produce radiation that heats and evaporates their surrounding protostellar disks, influencing the conditions for planet formation. Evaporation of planetary atmospheres by the same stellar radiation can also explain a bimodal distribution of planetary radii seen in transit observations (see Section 2.1). The higher energy radiation (UV and X ray) from low mass stars can also drastically alter the chemistry of planetary atmospheres, and thus the habitability of planets, but more precise determinations are needed of how the relevant radiation changes as a function of stellar mass or age.

In regions of high mass star formation, the radiation and stellar winds produced by massive stars can dominate the dispersal of molecular clouds (Figure 2.18), but observationally diagnosing which processes are the most important in different environments has proved challenging. There are tantalizing observational and theoretical clues suggesting that star-forming clouds with sufficiently high densities are difficult to disrupt by stellar feedback and may form super-star clusters and perhaps globular clusters at high redshift. Studies of reionization era galaxies (e.g., with JWST) and local analogues in the coming decade may finally resolve this long-standing puzzle. Regions of high mass star formation are often buried behind huge layers of dusty gas so improved long wavelength observations (far infrared, sub-mm, radio) are required to peer through the dust.

FIGURE 2.18 Multiwavelength image of the star-forming region 30 Doradus in the Large Magellanic Cloud illustrating the complex physical processes responsible for the disruption of star-forming giant molecular clouds and the production of hot, multiphase gas on the scale of 10’s of parsecs. Hot gas from stellar winds and supernovae (blue, Chandra), radiation from massive stars (green, Hubble), and re-radiated infrared emission from dust (red, Spitzer) all trace feedback into the interstellar medium on sub-kiloparsec scales. SOURCE: NASA, https://chandra.si.edu/photo/2012/30dor/, X-ray: NASA/CXC/PSU/L.Townsley et al.; Optical: NASA/STScI; Infrared: NASA/JPL/PSU/L.Townsley et al.

Supernovae generally explode too late after stars form to dominate the dynamics within most molecular clouds, but they are a critical source of feedback on galactic scales. The Chandra X-ray observatory led to major progress in the last two decades on understanding supernova feedback, but higher resolution and higher sensitivity X-ray imaging and spectroscopy would enable much more quantitative probes of supernova feedback and its role in powering galactic winds. Supernovae are important for a second, less direct, reason, in that they produce the gigaelectronvolt (GeV) cosmic rays that dominate the energy of relativistic particles in many galaxies. The impact of cosmic rays is one of the largest uncertainties in understanding feedback in galaxy formation. The primary uncertainty is how cosmic rays are scattered by small-scale fluctuations in the magnetic field, which sets whether cosmic-rays can escape a region or whether their pressure builds up to the point where it can drive an outflow. On smaller scales, these cosmic rays can affect the thermal balance and chemistry of molecular clouds and their ability to form stars. It is remarkable that tiny solar-system scale fluctuations in the galactic magnetic field are a key ingredient in understanding how galaxies drive winds on scales of tens of kiloparsecs, or that the large scale magnetic field properties or distant supernovae can affect the formation of pre-stellar cores. This is an area where additional theoretical advances are particularly needed, including advances in plasma simulation techniques.

FIGURE 2.19 Multiwavelength image of the Perseus galaxy cluster illustrating the impact of black hole feedback in clusters and the coexistence of cool gas in the hot intracluster medium that provides fuel for ongoing star formation and black hole accretion. Radio emission from jets (pinkish lobes, Karl Jansky Very Large Array (JVLA)) fill cavities in the X-ray emission (violet, Chandra). Optical emission shows cooler photoionized gas (red filaments, HST). SOURCE: NASA and STScI, https://hubblesite.org/image/2376/gallery/135-multiwavelength, X-ray: NASA/CXC/IoA/A.Fabian et al.; Radio: NRAO/VLA/G. Taylor; Optical: NASA, ESA, the Hubble Heritage (STScI/AURA)-ESA/Hubble Collaboration, and A. Fabian (Institute of Astronomy, University of Cambridge, UK).

Feedback from supermassive black holes during galaxy formation and evolution is one of the most dramatic examples of the small scales in the cosmic ecosystem impacting the large (Figure 2.19). Accreting black holes can influence their surroundings through UV and X-ray radiation, collimated relativistic jets, and wider-angle winds. However, the understanding of how active galactic nuclei (AGN) spectra, jets and winds vary with luminosity, black hole mass, black hole spin, and perhaps other properties is rudimentary. This currently is a major bottleneck in understanding the evolution of galaxies. Theoretical progress in these areas is likely to continue to be rapid, with advances in general relativistic simulations of black hole accretion predicting increasingly realistic jets and winds for comparison to observations. These predictions push the frontiers of radiation theory, plasma theory (to determine how and where the plasma is heated), and computational astrophysics. Observationally, a combination of high-resolution millimeter (mm) imaging and multiwavelength, multi-messenger observations can reveal the launching mechanism and particle content of relativistic jets (electron-proton vs. electron-positron) and thus the energetics of jet feedback. More detailed studies of molecular gas emission will reveal how the densest star-forming components of the interstellar medium are impacted by AGN jets. And higher sensitivity and spectral resolution optical-UV and X-ray spectroscopy of broad emission and absorption line outflows from AGN are needed to better diagnose the physical properties of accretion disk winds and determine how much mass and energy they actually carry. In addition to measuring the jet and wind properties, better constraints on the masses and spins of the supermassive black holes themselves will play in important role in understanding how these objects formed and how they grow, as well as their role in the feedback processes described above. X-ray telescopes and next generation gravitational wave experiments can constrain the black hole spins, and their masses will be better constrained by measurements with next generation optical and radio telescopes.

Observations of galaxy clusters have revealed the critical role of black hole feedback by jets in the intracluster medium (Figure 2.19), though exactly how the energy from the black hole couples to the surrounding gas is still uncertain. This is a prime example of the need for multiwavelength observations: the combination of radio, X-ray, and optical data reveals the interplay between the centimeter-wave-emitting relativistic jets, mm-emitting molecular gas, X-ray emitting thermal intracluster plasma, and the optical-emitting photoionized gas. Higher spectral resolution X-ray spectroscopy of clusters would sharpen the understanding of AGN feedback in this critical environment that serves as a laboratory for understanding AGN feedback more broadly. Another observational frontier lies in extending studies of the hot intracluster medium to galaxies, probing the transition from galaxies that have largely ceased star formation to actively star forming galaxies. This is likely to transform the understanding of the role of feedback across a wide range of environments by directly observing the impact of feedback on the gaseous halos that contain the fuel for galaxy growth (see Box 2.3). Multiwavelength studies will again be key. UV and optical spectroscopy, and mm dust continuum observations probe the cooler multiphase gas. Combining UV absorption and UV emission studies of the CGM would be particularly valuable, as would deeper X-ray imaging and spectroscopy. The Sunyaev-Zel’dovich (SZ) effect is a direct probe of the thermal pressure of ionized gas in galactic halos. Of all the observational diagnostics at our disposal, the SZ effect most directly constrains the energy content of gas in galactic halos produced by the combined effects of gravitational collapse and stellar and black hole feedback. Higher sensitivity and higher resolution CMB observations motivated to a significant extent by cosmology (Section 2.2) will have a large impact on the understanding of galaxy formation as well.

2.3.3 Multi-Scale Cosmic Flows of Gas

Feedback and gravity are the key ingredients that determine how gas flows across cosmic scales. Directly observing these gas flows is challenging because of the diffuse nature of the gas in galaxy halos and the high spatial resolution required to peer into regions of ongoing star, planet, and massive black hole formation. New theoretical and observational capabilities are, however, allowing this critical aspect of the tug-of-war between gravity and feedback to be tackled.

Within galaxies, star-forming clouds form and disperse on timescales of millions of years. The structures are thus constantly being reshaped by cosmic flows of gas driven by the interplay between gravity and stellar feedback. Within those clouds, a major puzzle is how turbulence and magnetic fields determine the gas flows down to young stars and the planet-forming disks that surround them. Theoretically, this subtle problem requires understanding the degree to which the mostly neutral gas is coupled to the magnetic field—small-scale physics which dramatically impacts the large-scale problem of how disks around young stars form. Observationally, higher resolution radio and infrared imaging of protostars and their surrounding gaseous environments with ALMA and other instruments are required for progress in this area (Figure 2.20). Images of dust emission from ESA’s Herschel Space Observatory revealed that gaseous filaments are responsible for fueling star formation on the scale of star clusters, but the role of filaments in determining cluster structure and stellar fragmentation is not yet clear. The accretion disks around young stars fed by these gas flows in star-forming cores set the conditions under which planets form.

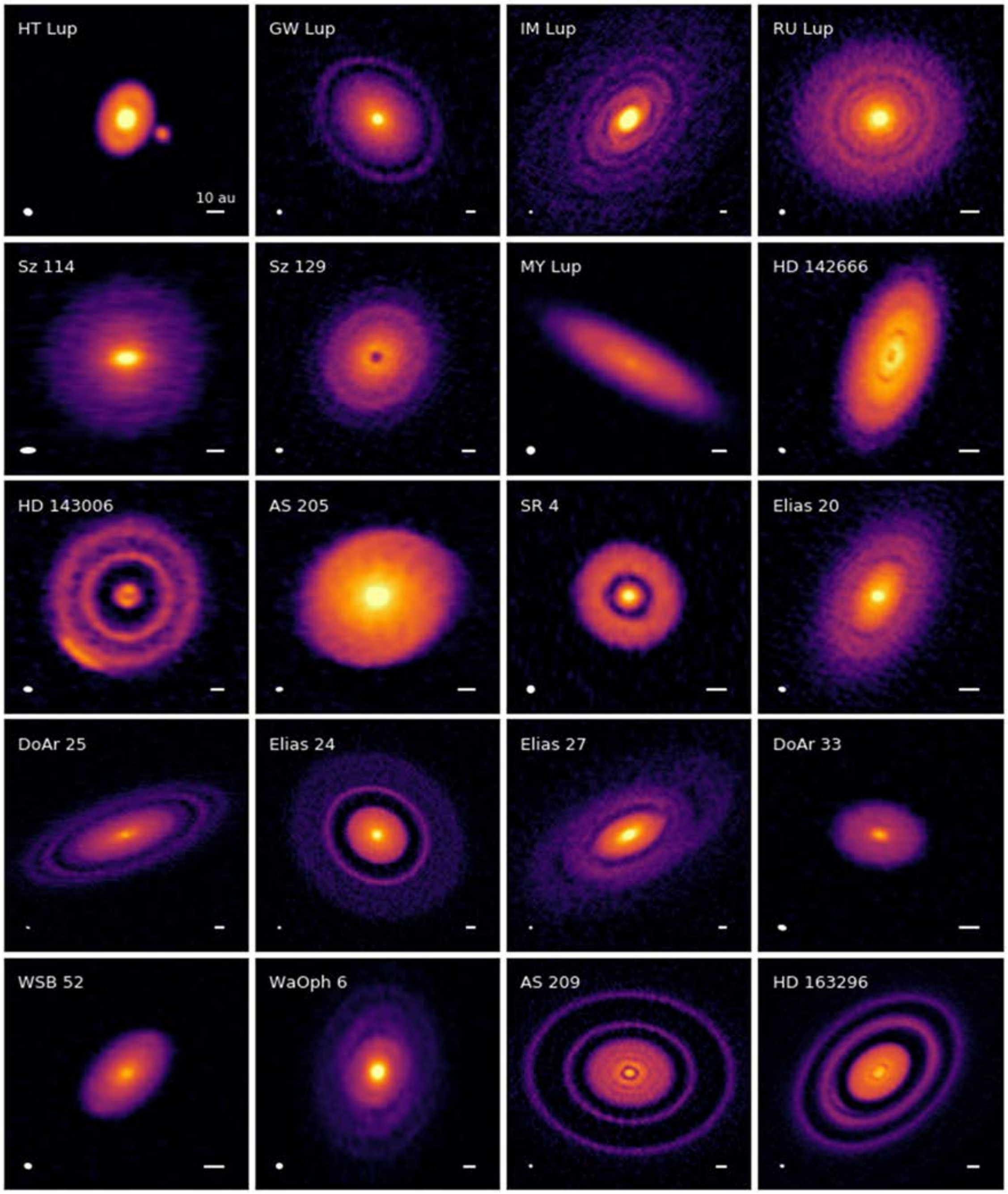

FIGURE 2.20 Montage of planet-forming disks around young stars as revealed by millimeter observations from the Atacama Large Millimeter/submillimeter Array (ALMA) observatory. ALMA's sensitivity and angular resolution enable the discovery of substructure in the disks as revealed by continuum dust emission. There is a near-ubiquitous geometry of gaps and rings which may point to the existence of forming planets, or possibly a signpost of magnetohydrodynamical processes occurring in the disks. Spectral-line observations indicate velocity structures in the gas and reveal flows similar to what has been predicted for the early stages of planet formation. In each image, the lower left icon indicates the beam size and the lower right icon is a 10 au scalebar. SOURCE: From of S.M. Andrews et al 2018, "The Disk Substructures at High Angular Resolution Project (DSHARP). I. Motivation, Sample, Calibration, and Overview," *The Astrophysical Journal Letters*, 869 L41. © AAS. Reproduced with permission. doi:10.3847/2041-8213/aaf741.

Studies with Spitzer, Herschel, and the Atacama Large Millimeter/submillimeter Array (ALMA) of local star-forming regions have identified protostars down to low masses and have brought the complex interactions with their surroundings to light. An area of particular focus is the relation between protostars and the dynamics, chemical evolution abundances, and physical conditions in protostellar disks. High resolution ALMA images of gas and dust have shown that many disks have small scale substructures such as gaps and rings (Figure 2.20), which are either created by already formed planets or by gas instabilities that could shape future planet formation. An unexpected result of these studies is that the central regions of the disks are often dust-obscured even at submillimeter wavelengths. Imaging of disks at high spatial resolution by ALMA or JWST can map the distribution within disks of a wide variety of molecules, with lines sampling an extensive range of density and temperature. These molecules, as interpreted by chemical models, are beginning to sketch out the early chemical evolution of planetary systems as shaped by the young star, but many key molecules, such as water, remain to be observed. A question for the coming decade is to understand the coupling between these small scale, solar system-forming regions, the larger cloud environment, and the diffuse ISM. The challenge is that these different scales harbor gas with a wide range of temperatures, phases, and chemical species, which require panchromatic observations from the millimeter and far-infrared through UV. Ground-based spectroscopy in the radio, optical, and near-infrared will be complemented by UV and far-IR spectroscopy from space to map the full cascade of star formation from the diffuse ISM down to individual protostars.

Our understanding of the flow of gas on the much larger distance scales of galaxies is also evolving rapidly. Despite the vast distances between observed galaxies, theoretical models predict that they are connected via the cosmic web of dark matter that forms the backbone for gas flows on cosmological scales (Figure 2.21). Theory predicts that gas flows into galaxies from the circumgalactic medium that fills the dark matter halos surrounding galaxies (Box 2.3). The circumgalactic medium is in turn fed by flows of gas from the more diffuse intergalactic medium and from gas ejected into the circumgalactic medium in galaxy-scale outflows. Characterizing the structure, metallicity, and dynamics of the circumgalactic and intergalactic medium with optical-UV and X-ray spectroscopy is a major frontier highlighted by the science panels.

During the formation of galaxies a small fraction ($\sim 10^{-3}$) of the gas somehow loses essentially all of its angular momentum to end up in a massive black hole at the center of the galaxy. Exactly how this happens remains a mystery, though theoretical models are beginning to connect the growth of black holes to the properties of gas in the host galaxy on much larger scales (Figure 2.21). Given the energetic importance of black hole feedback, we need to understand how and when black holes grow, to assess their role in galaxy formation. High spatial and spectral resolution molecular gas observations are the key to peering into galactic nuclei, and can reveal how gas accretes and the properties of the black hole itself. One key problem highlighted by two of the science panels is whether massive black holes first form and grow by accretion from seed black holes formed in stellar core-collapse, or whether under rare circumstances it is possible for gas clouds to collapse directly to $\sim 10^{4\text{-}5}$ M_{sun} black holes, perhaps via a supermassive star intermediary. Or does an entirely different set of processes seed galactic nuclei with massive black holes at high redshift? A combination of gravitational wave measurements of black hole mergers across cosmic time (see Section 2.2), deep near-infrared (JWST), far-infrared, and X-ray imaging and spectroscopy of high redshift accreting black holes would help piece together the origins story for massive black holes.

The collective result of feedback from star formation and black hole accretion powers galactic winds, galaxy-scale outflows of mass, energy, and heavy elements that have a major impact on the star formation histories and chemical evolution of galaxies and on the gaseous medium surrounding galaxies. The morphological and spectroscopic evidence for galactic winds is overwhelming (e.g., Figure 2.224), as is the heavy element enrichment they produce hundreds of kiloparsecs from star forming galaxies and even out into the diffuse intergalactic medium.

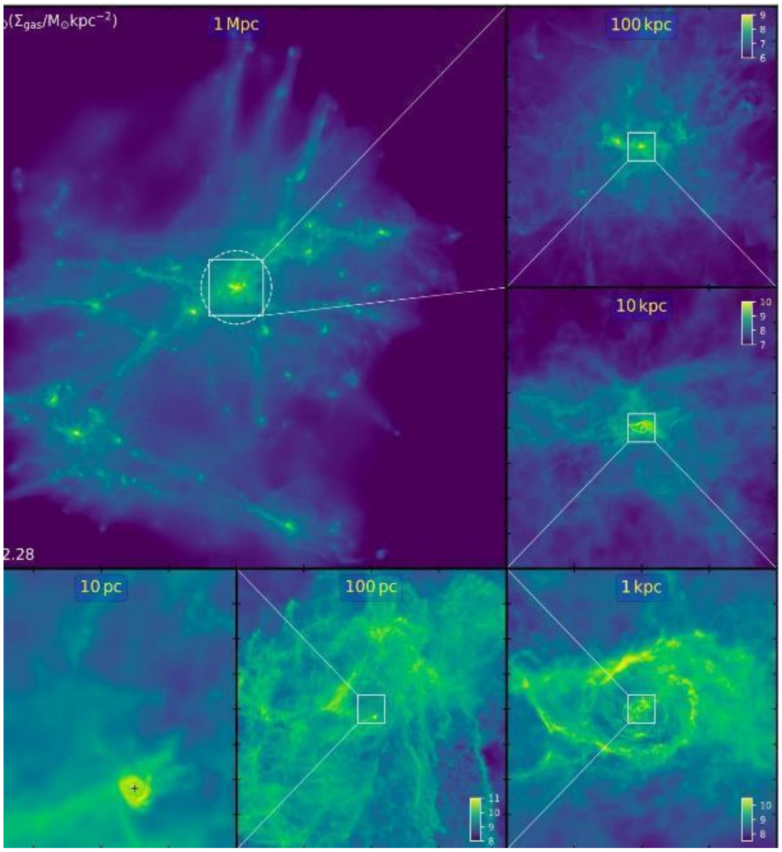

FIGURE 2.21 Flows of gas in a cosmological simulation from scales of the cosmic web on millions of parsecs to the central parsec of a galaxy, in order to study inflow of gas that fuels star formation and accretion onto the central massive black hole (the + in the lower left image). Color shows gas mass surface density. SOURCE: From D. Anglés-Alcázar et al 2021, "Cosmological Simulations of Quasar Fueling to Subparsec Scales Using Lagrangian Hyper-refinement," *The Astrophysical Journal*, 917 53. © AAS. Reproduced with permission. doi:10.3847/1538-4357/ac09e8.

Massive stars and their supernovae are sources of mass, energy and momentum, which emerge in the form of fast-moving shock-heated gas, and relativistic cosmic ray particles as well as photons. Radiation, winds, and jets from accreting supermassive black holes are also likely important in powering galaxy-scale outflows in galaxies. It is, however, an open question whether AGN are only important for high mass galaxies or have a significant impact in low mass galaxies as well. There are major uncertainties in how the ingredients of stellar and AGN feedback combine to produce the outflows that we observe. Perhaps the biggest puzzle lies in understanding the subtle problem of the co-existence of gas at very different temperatures and densities in the outflows. Galactic winds are observed to be multiphase, with molecular, atomic, and warm and hot ionized gas apparently coexisting in the same outflow, but at different velocities. This is revealed by multiwavelength images and spectra ranging from the mm to the X-ray: the wide range of physical conditions requires a panchromatic observational approach. It is unclear how much of the colder gas is launched from the galaxy and how much forms in situ in the outflow. Recent far infrared dust polarization maps showing smooth vertical magnetic fields over large portions of galaxy disks argue that the cold gas in which this field is embedded is influenced by the wind; future infrared data will test these ideas and provide direct information on feedback processes in and around the cold ISM. Theoretically, it is likely that small-scale physical processes such as instabilities and thermal conduction regulate the transfer of gas between different phases. Although they are small in scale, these processes are in fact critical because they determine how the wind material cools and thus how much mass and energy can actually be ejected from a galaxy. A better understanding of how mass moves between phases is also necessary to quantitatively interpret the wealth of multi-wavelength data on galactic winds, and connect those observations to the physical quantities of most interest such as the mass and energy winds carry. Some of the most pressing questions, and greatest opportunities, are on the largest scales, as outlined in Box 2.3.

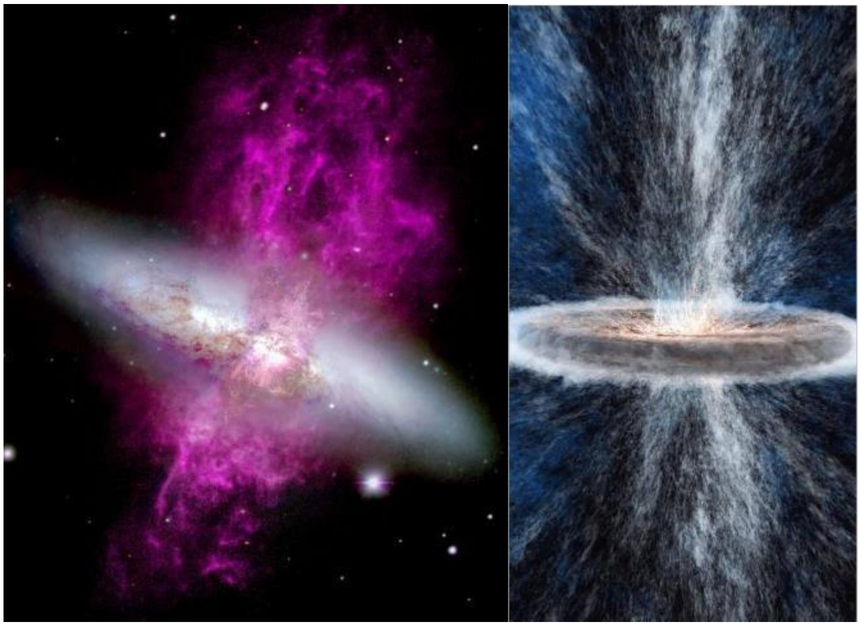

FIGURE 2.22 (*left*) Multiwavelength image of the starburst M82 illustrating the H-alpha emission from the galactic wind (pink) and the optical continuum from the galactic disk (BVI colors). (*right*) Numerical simulation of a bi-conical galaxy-scale outflow driven by multiple supernovae in a disk galaxy. SOURCE: Left: International Gemini Observatory/NOIRLab/NSF/AURA/M. Westmoquette (UCL)/J. Gallagher (Wisconsin-Madison)/ L. Smith (STScI/UCL). *Right:* From E.E. Schneider et al 2020, "The Physical Nature of Starburst-driven Galactic Outflows," The Astrophysical Journal, 895 43. © AAS. Reproduced with permission. doi:10.3847/1538-4357/ab8ae8.

BOX 2.3 Connecting Galaxies to the Cosmic Web

Observations of star-forming galaxies show that entire gaseous galactic disks would typically be converted into stars over less than a billion years, much less than the age of the universe at most redshifts. A continuous supply of gas into galaxies is thus required to maintain the observed ongoing star formation. This is believed to come from the circumgalactic medium (CGM) that fills the dark matter halos surrounding galaxies. The circumgalactic medium is in turn fed by flows of gas accreted from the more diffuse intergalactic medium and by material ejected from galaxies by galactic winds (Figure 2.3.1): the CGM is thus where all the elements of the galactic ecosystem connect. Based on constraints from current observations, about 10% of the baryons in the universe reside in stars and in the cold, dense interstellar medium, about 10% resides in the CGM within virialized galactic halos, and about 80% is in the diffuse medium. Within a virialized halo, the CGM may comprise 80% or more of the baryons.

Box 2.3 continued

All processes that drive galaxy evolution, from the evolution of individual stars and accretion onto black holes to the galactic-scale winds that are powered by them, take place against the backdrop of the large scale structure of the universe which feeds and sustains galaxies. This structure is gravitationally-dominated by dark matter, and simulations have long predicted that the dark matter is arranged into walls and filaments that are organized into the foam-like cosmic web. Gravitationally channeled by the dark matter filaments, gas is predicted to follow broadly the same large scale structure. Galaxies and the stars within them assemble at the nodes of the dark matter filaments, where gas accumulates, cools, and ultimately reaches high enough densities to form stars (Figure 2.21).

The interface between galaxies and the cosmic web is the circumgalactic medium (CGM), the complex and crucial immediate environment of galaxies. Giant rivers of pristine gas move inwards, from multiple directions, delivering the fuel for future generations of stars. These inflows are impeded as they encounter equally impressive galaxy-scale outflows, driven by star formation and black hole accretion deep inside the galaxies. This encounter creates shocks and turbulence, and it is now thought that the fate of galaxies—whether they continuously form stars at a low rate, like the Milky Way, undergo a period of intense star formation like M82 (Figure 2.22), or become quiescent—rests on the outcome of this interaction of competing gas flows. These interactions happen throughout cosmic history: the most intense winds likely occurred at “cosmic noon” ($z\sim2$), with the CGM of today’s galaxies providing a record of the past history of feedback. Furthermore, the flows are predicted to be strongly dependent on galaxy mass: theoretical models predict a transition from gas that is primarily hot and pressure-supported in massive galaxies to gas that is primarily cold in lower mass galaxies.

So far these spectacular interactions, which are among the largest causally-connected events in the universe, have almost exclusively been studied in computer simulations. Direct observations are needed to answer some of the most fundamental questions about galaxy formation, such as whether gas actually accretes from the intergalactic medium, what the rate of accretion is, whether gas in outflows is “recycled” back into the galaxies, and what the spatial distribution and physical conditions are of gas on the largest scales. Such observations also constitute one of the most direct tests of the long-standing idea that galaxies are connected with each other through a diffuse cosmic web of gas and dark matter.

The multiphase nature of the diffuse CGM/intracluster medium (ICM) and its complex dynamics, where many physical processes are superposed, necessitate a multi-wavelength and multi-scale approach. Improved theoretical studies of the key physical processes are needed, as are detailed computational models of global CGM dynamics faithful to this physics. On the observational side, major improvements are needed in imaging and spectroscopic studies of individual galaxy halos and clusters over the full spectral range from X ray to radio, and spatial scales that span from 100 kpc scale haloes down to 100 pc scale star-forming clouds. Very large optical telescopes with diameters of 20-30 m can probe the diffuse ionized gas directly with sensitive integral field units (IFUs) and other components through absorption line studies of background sources. A large telescope in space would be sensitive enough to allow ultraviolet absorption spectroscopy towards a dense network of faint background quasars behind a single foreground galaxy, revealing the composition, temperature, velocities, and density structure of the diffuse hot gas that is thought to contain most of the baryons. Large samples of dispersion measures to localized fast radio bursts (see Section 2.2.5) offer a powerful new probe of the ionized gas in galaxy halos. Together, these complementary observations would be transformational for probing and understanding the CGM. These are all major frontiers highlighted by the science panels.

Box 2.3 continued

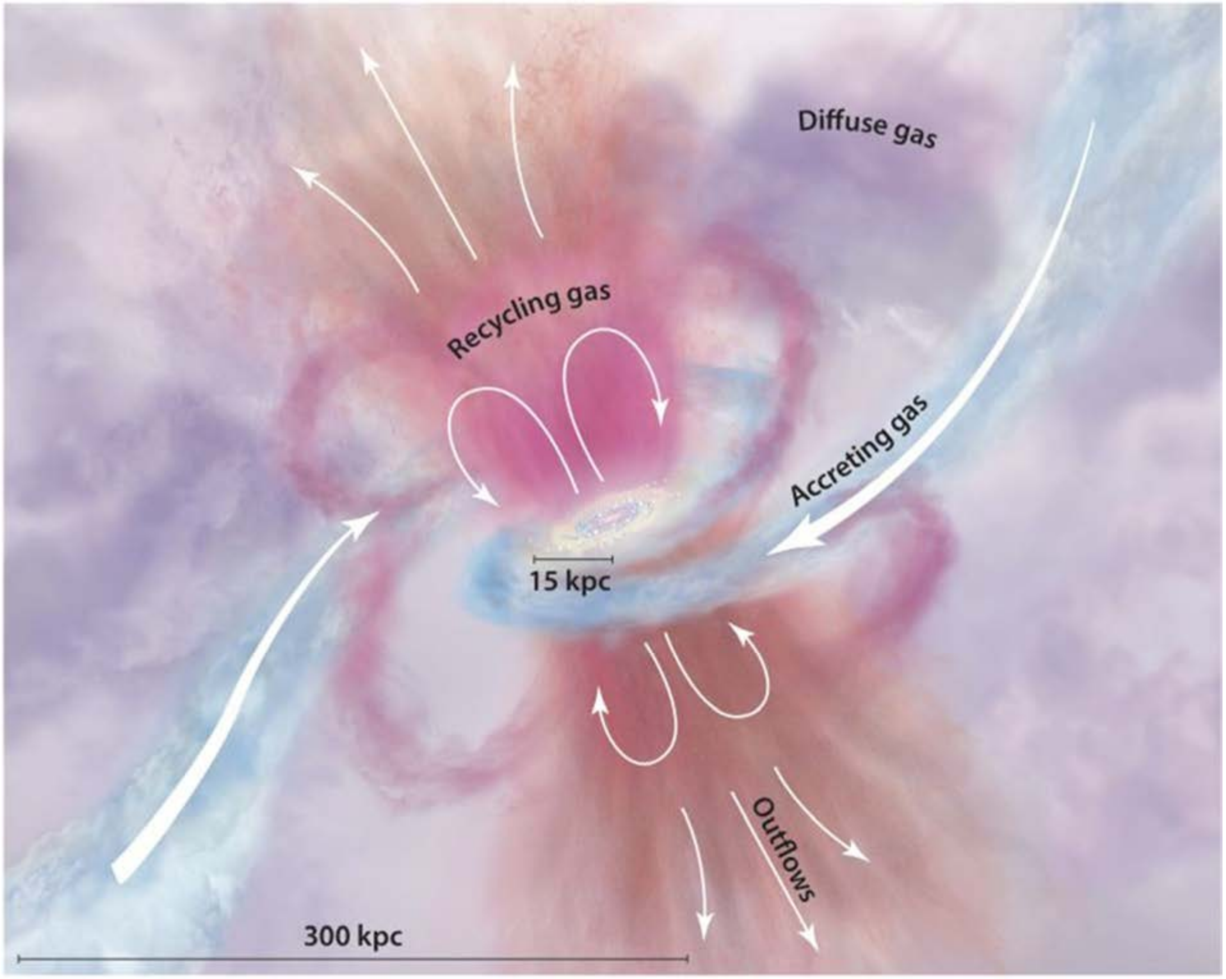

FIGURE 2.3.1 Illustration of the circumgalactic medium, where intense outflows from galaxies—driven by supernovae and black holes—interact with inflowing pristine gas from intergalactic space. SOURCE: Republished with permission of Annual Reviews, Inc., from Tumlinson, Peeples, and Werk, 2017, The circumgalactic medium, *Annual Review of Astronomy and Astrophysics* 55: 389-432; permission conveyed through Copyright Clearance Center, Inc.

Priority Science Area: Unveiling the Hidden Drivers of Galaxy Growth

By its very nature the *Cosmic Ecosystems* theme is very rich, exploring interconnected processes ranging over a billionfold range of linear scales, from individual supernovae to the virial radii of galaxies and beyond, Among this abundance of science objectives, the survey has chosen as its priority science area Unveiling the Drivers of Galaxy Growth. The pathway towards achieving this goal is bracketed by two sets of transformational observational opportunities: first the revolutionary measurements of the comic history of galaxy growth to come very soon from JWST, and culminating with the revolutionary capabilities for ground-based and space-based imaging and spectroscopy from the 24-40 m ELTs and a large IR/optical/UV space observatory, respectively.

An essential milestone on the pathway to revealing these drivers is a complete understanding of the formation and buildup of galaxies and their structures, stellar populations, metals, and central black holes over cosmic time. JWST is poised to revolutionize this subject, by imaging and identifying galaxies out to some of the first generations during the epoch of reionization, and by obtaining deep redshifted ultraviolet – visible spectra for thousands of a galaxies spanning the period from reionization to the present. When combined with observations from ALMA, NOEMA, and the JVLA it should be possible to trace the transformation of baryons from interstellar molecules to stars and metals within a self-consistent framework. This information will be complemented by wide-field galaxy surveys from the Vera Rubin Observatory, the SPHEREx Explorer mission, the ESA Euclid mission, and the Roman Space Telescope, which will provide large statistical inventories of galaxies and rich target lists for future massively multiplexed spectroscopic surveys on the ground. Observations of nearby galaxies and our galaxy with JWST and WFIRST will also provide new insights into the physical processes driving star formation and the feedback processes powering the ecosystem.

Beyond that, the key missing link in unveiling the physics driving galaxy growth is to measure the properties of the diffuse gas within, surrounding, and between galaxies. These are the sites where baryons accrete on to galaxies, star formation is triggered, central black holes accrete and grow, and the feedback processes regulating galaxy growth are manifested. The key observational probes of all of these processes are emission and absorption-line spectroscopy of the diffuse gas, which contain a wealth of diagnostic information on the physical conditions, compositions, and dynamics of the gas. Current telescopes, whether located in space (HST) or on the ground (6-10 m apertures) are only capable of probing the densest regions in emission and occasional single sightlines through rare galaxies which happen to be superimposed in front of a bright quasar. These sporadic observations have been sufficient to demonstrate the power of spectroscopy for probing the baryon cycle but far too little to build up a robust physical picture of the processes at work. A large-aperture UV/optical space telescope, however, also envisaged for addressing Pathways to Habitable Worlds, would transform this subject. The combination of 6 m-class aperture and a high-efficiency spectrograph with modern detectors would provide thousands of potential sightlines to nearby galaxies, enabling “tomographic” studies of their circumgalactic and interstellar media, as well as rich new observations of the intergalactic gas clouds along the same lines of sight. With a versatility comparable to that of HST but an aperture comparable to JWST such an observatory would carry out groundbreaking observations of galaxy growth and address a majority of the science questions identified in the survey overall.

Major inroads to addressing this problem can also be met by the next generation of ELTs. This would include similar absorption-line spectroscopy (but targeted at the redshifted spectra of more distant galaxies observed at earlier cosmic epochs), as well as studies of the evolution of active galactic nuclei and the growth of the corresponding central black holes over cosmic time. The giant-aperture telescope alone will be able to obtain spectra for the faintest galaxies imaged by JWST, including some of the first generations of objects formed during reionization. The same spectroscopic capabilities, when trained on ancient stars in the halos of the Milky Way and in its nearest companion galaxies, can provide a detailed record of the assembly and chemical enrichment of these galaxies in their formative stages. More generally, the vast majority of faint galaxies revealed by deep JWST images will be too faint for follow-up spectroscopy either with the largest current ground-based telescopes or with JWST itself; the ELTs will play major roles in carrying out deep spectroscopic observations of these targets, probably including some of the most important “Rosetta stones” of the first generations of galaxies, to confirm their redshifts and measure their physical properties. Imaging observations with the ELTs (with AO) and with a IR/optical/UV space telescope of the resolved stellar populations of nearby galaxies out to the Virgo cluster can extend these fossil records to the full demographic population of galaxies today.

As with our other two priority science areas, these large facilities, though transformational in their potential impact, cannot address these critical scientific questions by themselves. Over the coming decade, major contributions to these investigations will come from multi-wavelength observations with the JVLA, ALMA, VRO, Euclid, Roman, 3.5-10 m OIR telescopes, Chandra, and toward the end of this period by the ESA Athena mission. Athena’s emphasis on *The Hot and Energetic Universe* will provide especially unique and powerful constraints on the cosmic feedback cycle, especially from AGNs, and it underscores the critical importance of future X-ray and infrared missions if these scientific questions are to be fully addressed. Finally, perhaps as much as in any branch of astrophysics, progress in this field critically rests on a vibrant program of theoretical modelling and numerical simulations of galaxies, over the full dynamic range of scales over which the ecosystems operate.

2.4 SYNTHESIS AND CONNECTIONS

The science themes envisioned for the next decade and beyond reflect an interconnected cosmos. The hierarchies of structure inherent in the universe are dependent on interactions across this wide range of scales. Planet formation and evolution is influenced by the local stellar environment; star formation is affected by galactic environments and strongly influences them through supernovae and other feedback processes; and galaxy formation and evolution depends on the circumgalactic and intergalactic environment and interactions therein as well as on the energy output produced by massive black hole growth in galactic nuclei. Measurements of the positions and properties of individual stars in nearby galaxies enables an extraction of their evolution and life history from the larger forces at play in the galaxy's structure, formation, and evolution. Disentangling this web requires understanding the complex local interdependencies. The interconnectedness even extends to the power of applying multiple wavelength observations, multiple messengers, and novel observational techniques being applied to vastly different science areas.

The need for observations across the electromagnetic spectrum is common for all of these science themes, as the complexity of these questions requires a multiplexed effort. The interconnected nature of cosmic ecosystems necessitates an observational and theoretical program in which traditional boundaries between disciplines (e.g., radio, X-ray, star formation theory, galaxy formation theory) give way to a more comprehensive approach. Improved multi-wavelength observational capabilities are particularly important for probing the full range of physical conditions, from cold, dusty gas and synchrotron emitting cosmic-rays observed in the radio and infrared, to optical and UV observations of stellar and black hole radiation, to the hot interstellar and circumgalactic medium observed in the UV and X-ray and by the Sunyaev-Zeldovich effect on the CMB. The study of radio jets in galaxies informs the mechanical and radiative inputs to galaxy structure and details the large scale consequence of supermassive black holes at a galaxy's center, itself amenable to study at high energies, ultraviolet, optical and infrared wavelengths. Research into aspects of star formation requires a panchromatic approach: deeply embedded sources only appear in the infrared, while magnetically active young stars emit copious high energy radiation, and accretion processes appear at far infrared, optical, ultraviolet, and even high energy wavelengths. Compact objects can emit at the longest and shortest wavelengths of the electromagnetic spectrum, as well as potentially be a source of other particles like neutrinos or cosmic rays.

The universe is not static, and the time domain is important to many aspects of astrophysics. Constraints on the local value of the expansion of the universe rely on the identification and use of variable stars as one avenue for measurements, while a different route uses time delays in variability from multiple gravitationally lensed images of quasars to provide the constraint. Most of the detection methods for exoplanets use some means of detecting changes in stellar properties over time to infer the presence of a planet. Many objects in the universe change their intensity with time, in a way that illuminates the astrophysics of the object itself or makes it amenable to study some other phenomenon: witness the monitoring of quasar spectra over time to illuminate the innermost regions near the center of a galaxy around the central supermassive black hole, as well as studies of the lifetime and distribution of starspots through long-term precision light curves. The time domain also encompasses eruptions and explosions, from nova outbursts of episodic accretion to the merging of two neutron stars, which produce both electromagnetic and gravitational wave signals during the final coalescence, to the high energy flares that may prevent the development of life on otherwise Earth-like planets orbiting magnetically active stars.

In the same way that answering the science questions of the next decade requires a multi-wavelength and multi-messenger approach, these objectives also require the synergy of space, ground, and even underground facilities. In the gravitational wave arena, space-based detection of inspiralling intermediate mass black hole mergers can signal the need for ground-based gravitational wave observation of the final merger, with searches for possible electromagnetic counterparts. Transients in electromagnetic emission will come by the millions per night from the Rubin Observatory once it is operational, with follow-up in other ground- and space-based telescopes needed for further characterization of the most interesting transient events. Neutrino observatories are opening an entirely new window for probing the most energetic processes in the universe. Exoplanet atmosphere characterization needs the combination of high precision spectro-photometric measurements obtainable from space with the high spectral resolving power available from large aperture ground facilities, to probe planets in the habitable zone around low-mass stars.

Theory, simulations, and laboratory measurements are just as crucial as new observations in making headway on these important lines of inquiry. A core challenge is understanding how to model systems in which processes on a wide range of time and length-scales interact to produce the universe that we observe. This includes both the need to model how the smallest objects (stars, supernovae, and black hole accretion disks) interact with their environment to understand the universe on large scales (e.g., galaxies and the circumgalactic medium), as well as the need to include small-scale physical processes (e.g., dust-gas instabilities in protostellar disks) to understand astronomical systems (planet formation, in this example). Theoretical mechanisms can explain the variety of transients observed and anticipated in the coming decade. General relativistic simulations of black hole accretion advance state-of-the-art predictions of the behavior of jets and winds from these compact objects, which can be compared with observations. From theory and simulations, a more complete knowledge of stars, as far as their rotation, binarity, and the impact of magnetic fields, improves the ability to model and interpret the expected ionizing output from massive stars. These parameters are also critical for understanding the evolution of massive stars and their feedback effects on galactic ecosystems. Likewise, simulations of convection, rotation, and magnetic field generation illuminate the dynamic nature of stars. Computational models and laboratory experiments, along with observations of gas and dust in planet-forming disks, are needed to understand planet formation in a more holistic manner. Knowledge of atomic and molecular properties gleaned from the laboratory can be essential for fully understanding the microscopic processes that have macroscopic consequences for stars, galaxies, and the cosmos.

TABLE 2.1 Science Panel Questions

Question	Theme(s)
Panel on Compact Objects and Energetic Phenomena	
What are the mass and spin distributions of neutron stars and stellar mass black holes?	New Messengers and New Physics
What powers the diversity of explosive phenomena across the electromagnetic spectrum?	New Messengers and New Physics
What do some compact objects eject material at nearly-light-speed jets, and what is that material made of?	New Messengers and New Physics
What seeds supermassive black holes and how to they grow?	New Messengers and New Physics, Cosmic Ecosystem
Panel on Cosmology	
What set the hot Big Bang in motion?	New Messengers and New Physics
What are the properties of dark matter and the dark sector?	New Messengers and New Physics
What physics drives the cosmic expansion and the large-scale evolution of the universe?	New Messengers and New Physics
How will measurements of gravitational waves reshape our cosmological view?	New Messengers and New Physics
Panel on Galaxies	
How did the intergalactic medium and the first sources of radiation evolve from cosmic dawn through the epoch of reionization?	Cosmic Ecosystem
How do gas, metals, and dust flow into, through, and out of galaxies?	Cosmic Ecosystem
How do supermassive black holes form and how is their growth coupled to the evolution of their host galaxies?	Cosmic Ecosystem, New Messengers and New Physics
How do the histories of galaxies and their dark matter halos shape their observable properties?	Cosmic Ecosystem
Panel on Exoplanets, Astrobiology, and the Solar System	
What is the range of planetary system architectures, and is the configuration of the solar system common?	Worlds and Suns in Context
What are the properties of individual planets, and which processes lead to planetary diversity?	Worlds and Suns in Context
How do habitable environments arise and evolve within the context of their planetary systems?	Worlds and Suns in Context
How can signs of habitable life be identified and interpreted in the context of their planetary environments?	Worlds and Suns in Context
Panel on the Interstellar Medium and Star and Planet Formation	
How to star-forming structures arise from, and interact with, the diffuse ISM?	Cosmic Ecosystem
What regulates the structures and motions within molecular clouds?	Cosmic Ecosystem
How does gas flow from parsec scales down to protostars and disks?	Cosmic Ecosystem
Is planet formation fast or slow?	Worlds and Suns in Context, Cosmic Ecosystem
Panel on Stars, the Sun, and Stellar Populations	
What are the most extreme stars and stellar populations?	Worlds and Suns in Context, Cosmic Ecosystem
How does multiplicity affect the way a star lives and dies?	Worlds and Suns in Context
What would stars look like if we view them like we do the Sun?	Worlds and Suns in Context
How do the Sun and other stars create space weather?	Worlds and Suns in Context

TABLE 2.2 Science Panel Discovery Areas

Discovery Area	Theme(s)
Panel on Compact Objects and Energetic Phenomena:	New Messengers and New Physics
Transforming our View of the Universe by Combining New Information from Light, Particles, and Gravitational Waves	
Panel on Cosmology:	New Messengers and New Physics
The Dark Ages as a Cosmological Probe	
Panel on Galaxies:	Cosmic Ecosystem
Mapping the Circumgalactic Medium and Intergalactic Medium in Emission	
Panel on Exoplanets, Astrobiology, and the Solar System:	Worlds and Suns in Context
The Search for Life on Exoplanets	
Panel on the Interstellar Medium and Star and Planet Formation:	Worlds and Suns in Context
Detecting and Characterizing Forming Planets	
Panel on Stars, the Sun, and Stellar Populations:	Worlds and Suns in Context, Cosmic Ecosystem
“Industrial Scale” Spectroscopy	