Unveiling the Darkest Secrets in the Universe: Notes on the 2020 Nobel Prize in Physics

Zhiyuan Li[1] & Feng Yuan[2,3]

[1]School of Astronomy and Space Science, Nanjing University, Nanjing 210023, China (lizy@nju.edu.cn)

[2]Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China (fyuan@shao.ac.cn)

[3]University of Chinese Academy of Sciences, Beijing 100049, China

ABSTRACT

The century-long story of black holes, started soon after the birth of Einstein's theory of gravity, is now widely spread. The first half of the story was mostly mathematical and abstract. It is into its second half that the story found its stage in the physical world. Central to the stage is a massive compact object residing in (and in fact, defining) the center of our Milky Way galaxy, considered by many as the most convincing case of an astrophysical black hole. In response to the recent excitement brought by the 2020 Nobel Prize in Physics, this article aims to highlight some of the milestones toward the establishment of black holes as a physical reality.

INTRODUCTION

On October 6, 2020, the anticipated announcement of the 2020 Nobel Prize in Physics was made by the Royal Swedish Academy of Sciences. This year's prize is divided, with one half going to Roger Penrose, for his discovery that *"black hole formation is a robust prediction of the general theory of relativity"*, and the other half is equally shared by Reinhard Genzel and Andrea Ghez, for their discovery of *"a supermassive compact object at the centre of our galaxy"*. In this article we aim to highlight some of the milestones toward these achievements and the establishment of black holes as a physical reality.

BLACK HOLES: A GENERIC PREDICTION BY GENERAL RELATIVITY (GR)

The concept of black hole in the pre-GR era

John Michell, a British astronomer, first proposed the concept of a black hole in the 1880s. He imagined a celestial body in the Universe, which would have the same density as the Sun but with a radius 500 times the solar radius. He then used Newtonian mechanics to calculate the escape velocity of this celestial body and found that it would be equal to the speed of light. This meant that even light would not escape from the surface of the object, which made Michell's concept the embryonic form of a black hole. Thirteen years later, the French naturalist Pierre-Simon Laplace independently put forward a similar idea in a paper, and even calculated the radius of a "black hole".

Black holes in general relativity: the ideal case of point-mass symmetry

These pioneering conjectures were based on Newtonian mechanics, but Newtonian mechanics are no longer applicable when gravity becomes extraordinarily strong. In November 1915, Einstein proposed his revolutionary theory of general relativity. Only two months later, in January 1916, the German astronomer Karl Schwarzschild obtained the exact solution of Einstein's field equation in the case of spherical symmetry [1]. The solution describes the geometric structure of space-time around a point mass. Schwarzschild found that space-time would be distorted around the point mass, such that a critical radius, commonly known as the Schwarzschild radius, $R_S = 2GM/c^2$ (here G is the gravitational constant, c the speed of light, and M the mass), exists to separate the space-time inside and outside it. No matter, including light, can escape from inside the Schwarzschild radius. This is the *black hole* in strict general relativity, and the sphere defined by the Schwarzschild radius is the surface of the black hole, also known as the *event horizon*.

Coincidentally, the radius of the black hole calculated by general relativity is exactly the same as that calculated by Newton's theory of gravity. The solution has the famous characteristic feature of a singularity inside the event horizon.

Schwarzschild's solution to Einstein's field equation is purely mathematical. In 1939, the American physicist J. Robert Oppenheimer and his student Hartland Snyder began to consider the physical meaning of the solution. They studied the collapse of a spherically symmetric gas cloud under its own gravity, and found that at the end of the collapse, the case of the Schwarzschild solution would take place, that is, the collapsing cloud would inevitably form a singularity wrapped by the event horizon [2]. This result had the important implication that stars in the Universe may eventually form black holes as described by the Schwarzschild solution. However, Oppenheimer's calculation strongly presupposed that the collapsing cloud would be strictly spherically symmetric, but this could not be the case in reality. At that time, many physicists believed that if this idealized assumption was abandoned, a singularity would not form at the end of the collapse. In fact, Einstein himself did not believe in black holes throughout his lifetime.

Penrose's theorem of singularity and the generic prediction for black holes

So, does the singularity exist in reality? In 1965, Roger Penrose, a British physicist, was attracted by this question [3]. His motivation came in part from the discovery of quasars (short for *quasi-stellar objects*) in 1963, which was considered one of the four breaking discoveries in astronomy of the 1960s (the other three include cosmic microwave background radiation, pulsars, and interstellar organic molecules). Some quasars were first detected in radio surveys of the sky performed in the 1950s, which manifested themselves as a bright compact radio source. However, the then large positional uncertainties in these radio sources hindered the identification of their optical counterparts and the determination of their distances. In 1960, the American astronomer Allan Sandage used the 5-meter Hale telescope at the Palomar Observatory to identify the optical counterpart of one of these compact radio sources, known as 3C48. The spectrum of this source exhibited several broad emission lines but at unfamiliar wavelengths, which led Sandage to suggest it as some unusual type of stellar object. The breakthrough came in February 1963, when the Dutch-born American astronomer Martin Schmidt took the spectrum of another radio source, 3C273. Also finding broad emission lines at unusual wavelengths (but different from those in the spectrum of 3C48), Schmidt realized that they could be understood as the Balmer series of hydrogen shifted by a considerable distance in the direction of red light; that is, 3C273 has a large *redshift* [4]. Following the case of 3C273, redshifted emission lines were soon identified in many more quasars, including those without significant radio emissions.

Initially, three explanations were proposed for the large redshifts of quasars: (i) fast-moving objects in or near our Galaxy, (ii) gravitational redshifts, and (iii) cosmological redshifts due to the expansion of the Universe. The debate was soon settled by a consensus with the third explanation, and most astronomers agreed that quasars lie at cosmological distances. Strikingly, this means that the typical luminosity of quasars is more than 100 times the luminosity of our Galaxy! A prominent question immediately followed: where does the enormous energy of quasars come from? Among the early answers to this question, the American astrophysicists Edward Salpeter and the Russian physicists Yakov Zel'dovich independently provided an insightful solution: quasars are powered by accretion of surrounding gas onto "massive collapsed objects" [5, 6] (it is noteworthy that at this time the nomenclature of "black hole" was not yet invented). In this scenario, gas falling toward the collapsed object releases its gravitational energy, which is then converted into internal and kinetic energy, raising the gas temperature to millions of degrees to even tens of billions of degrees. As a result, these high-temperature plasmas emit strong multi-wavelength electromagnetic radiation. Because of the deep gravitational potential well, the efficiency of this energy conversion is high enough to explain the observed high luminosity of quasars.

It was the problem of quasar energetics that prompted Penrose to reconsider the reality of singularity in the framework of general relativity, that is, the real existence of black holes. In order to study the reality of singularity, the hypothesis of strictly spherical symmetry needs to be abandoned from the very beginning. Penrose, with his exceptional talent in mathematics, successfully invented a new mathematical method making full use of the knowledge of topology [3]. One of the key concepts he proposed is the so-called trapped surface, a two-dimensional closed surface. All the light perpendicular to this surface is converging along the direction toward the future. This is contrary to a spherical surface in flat space, where outward-directed light-rays diverge. Penrose proved that no matter what kind of disturbance a star encounters during the collapse and whether the star is spherically symmetric, the trapped surface always exists. Moreover, in the framework of general relativity, once this surface is formed, the star will collapse irresistibly to the center to form a singularity, as illustrated in Figure 1. Of course, physical quantities are divergent at the singularity, which is unphysical. This is now understood as general relativity still belonging to the category of classical physics and only applicable to macro-scales. At micro-scales related to a singularity, a combined theory of general relativity and quantum mechanics is required. A successful theory of this kind has not yet been achieved. Nevertheless, this would be only a local correction of the singularity and does not affect the overall correctness of the Penrose theorem.

The Penrose theorem convincingly proves, for the first time, that the existence of black holes is a robust prediction of general relativity. Hence it is recognized as the most important progress in the field of general relativity since Einstein.

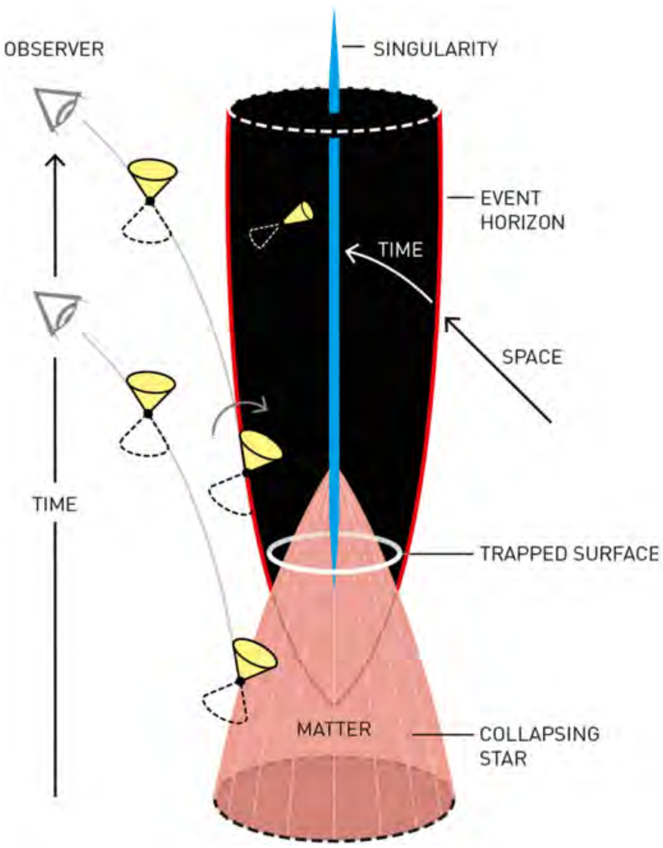

Fig. 1: An illustration of black hole formation from a collapsing star. Adapted from the Nobel Committee for Physics, 2020.

QUEST FOR A SUPER-MASSIVE BLACK HOLE IN THE GALACTIC CENTER

The existence of a super-massive black hole in the Galactic center (hereafter referred to as GCBH) is now beyond reasonable doubt. But it took several decades for the theoretical framework and convincing observational evidence to be established. It all started in the 1960s, with the discovery of quasars [4]; the recognition that accretion onto "massive collapsed objects" is the most likely energy source of the enormous radiation of quasars [5,6]; and the subsequent prediction that "dead quasars" exist in the nuclei of present-day galaxies, and in particular, in the nucleus of our own Galaxy [7, 8]. Observational searches for the predicted GCBH had mainly followed two lines: through the detection of electromagnetic signals coincident with the nucleus and through the kinematics of tracer particles (gas and/or stars) in response to the strong gravity.

The discovery of Sgr A*

By the 1960s, astronomers were certain about the direction of the Galactic center on the sky. This, however, had not been straightforward, since our view toward the inner Galaxy is necessarily interrupted by a large column density of cold interstellar medium. As a result, optical and ultraviolet emissions from the Galactic center are totally obscured. Therefore, astronomers used radio observations to first pinpoint the Galaxy's center with the bright radio source Sagittarius A [9, 10], which later decomposed into Sgr A East, a probable supernova remnant, and Sgr A West, an HII region also known as the "mini-spiral".

Beginning in the 1960s, the popularity of radio interferometry soon equipped astronomers with unprecedented angular resolutions down to the sub-arcsecond level. In the race for searching for the predicted signals from the putative GCBH, Balick & Brown [11] succeeded in making the first unambiguous detection of a compact (<~ 0.1") radio source coincident with the Galactic nucleus, based on a two-site interferometer of the National Radio Astronomical Observatory operating at 2695 MHz (11 cm) and 8085 MHz (3.7 cm) with a baseline of 35 km (the angular resolutions achieved at the two frequencies were 0.7" and 0.2", respectively). Balick & Brown [11] described their discovery:

"The unusual nature of the sub-arcsecond structure and its positional coincidence with the inner 1-pc core of the galactic nucleus

strongly suggests that this structure is physically associated with the galactic center (in fact, defines the galactic center)."

This compact radio source was later given the name of Sagittarius A* (Sgr A*), with the belief that it was the exciting source of the ionized gas in Sgr A West, and as an analogy to the nomenclature of atoms in an excited state [12].

That Sgr A* *defines the Galactic center* was more of a belief than a proof at the time of its discovery. But strong supporting evidence for a physical association of Sgr A* with the dynamical center of the Galaxy was soon gathered by the monitoring of its (two-dimensional) proper motion [13], which showed that the apparent motion of Sgr A* in the sky is consistent with the Sun's orbiting motion about the Galactic center. The most recent measurement of the proper motion of Sgr A* yields -0.58 ± 2.23 km/s in the direction of Galactic rotation and -0.85 ± 0.75 km/s toward the North Galactic Pole, after removing the Sun's orbital motion and for an assumed distance of 8.15 kpc [14].

It was also soon realized that Sgr A* is a variable source on timescales from days to years [15]. Sgr A* has been detected at progressively shorter wavelengths, through the centimeter, millimeter, infrared and X-ray bands. Substantial variability is found at all wavelengths, in particular, the X-ray (2-10 keV) flux of Sgr A* can frequently rise by a factor of 10-100 and decay back to its quiescent level within a timescale of minutes, a phenomenon dubbed "flares" [16]. This implies that the size of the X-ray-emitting region is as small as about 1 Astronomical Unit (1 AU$\approx1.5\times10^{13}$ cm). It is now generally thought that this flaring emission is due to synchrotron and/or synchrotron self-Compton radiation from high-energy electrons near the black hole event horizon, but the physical origin of these high-energy electrons remains to be understood.

An independent and more direct constraint on the size of Sgr A* comes from interferometric observations at millimeter wavelengths. While the apparent size of Sgr A* at centimeter wavelengths is dominated by interstellar electron scattering, this effect is wavelength-dependent (inversely proportional to the square of wavelength) and becomes sub-dominant at millimeter wavelengths. The intrinsic size of Sgr *A was thus determined to be 37 micro-arcsec (corresponding to about 0.3 AU at a distance of 8 kpc) at 1.7 mm [17].

Gas and stars orbiting a dark mass

The lack of motion of Sgr A* not only favors its coincidence with the dynamical center of the Galaxy, but also provides an interesting lower limit of its mass, 4×10^5 M$_\odot$, against Brownian perturbation by the surrounding stars [18].

Evidence for a concentration of stars in the innermost parsecs of the Galaxy, collectively called the nuclear star cluster, was first found by Becklin & Neugebauer [19], who pioneered near-infrared (NIR) observations toward this region. At wavelengths near 2 micron, some 10% of the starlight from the Galactic center can still peer through the interstellar medium as well as the Earth's atmosphere. The brightest stars in the nuclear star cluster were readily resolved [20], which primarily consist of late-type giants, supergiants and asymptotic giant branch stars. Interestingly, some of these stars have a compact radio counterpart, which makes them ideal benchmarks for aligning the NIR and radio image frames. This ultimately allows for the accurate positioning of Sgr A* with respect to the surrounding stars.

Early IR observations of an emission line from a forbidden transition of singly ionized neon at the rest-frame wavelength of 12.8 micron [21], also revealed the presence of streams of ionized gas in the innermost parsec, which is spatially coincident with Sgr A West. The kinematics of the ionized gas, traced by the velocity dispersion of the neon line, was used to derive the first estimate of dynamical mass enclosed within the central parsec. The resultant mass of a few times 10^6 M$_\odot$ was remarkable; however, it includes the cumulative mass of stars in the same region and thus should be considered an upper limit to the mass of the putative GCBH. An additional concern lies in that non-gravitational forces may play a non-negligible role in the gas kinematics, potentially biasing the estimate of the dynamical mass, although we now know that Keplerian motion is indeed a good approximation for the streams of ionized gas in Sgr A West [22].

Stellar kinematics, which is much more immune to non-gravitational effects, soon superseded gas kinematics as a robust probe of the enclosed dynamical mass in the central parsec, thanks to the advent of high-sensitivity, large-format IR detectors and spectrometers in the 1980-90s. Early measurements of stellar radial velocities and velocity dispersions already provided a tight constraint on a presumed point mass at the center of the nuclear star cluster [23], but the possibility that the inferred mass,

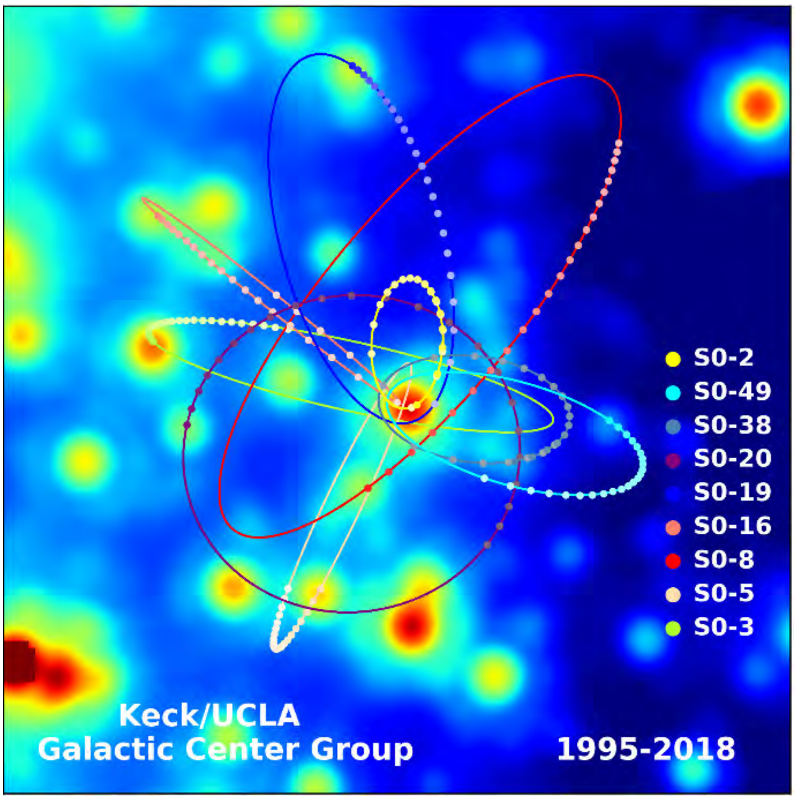

Fig. 2: More than two decades of monitoring of stars orbiting the GCBH. Adapted from the Keck/UCLA Galactic center group.

2.5×10^6 M⊙, is distributed across the central 0.1 parsec (≈3×10^4 AU), could not be ruled out due to an insufficient angular resolution. Diffraction-limited NIR images obtained with the 3.5-m New Technology Telescope improved the case further, deriving a dynamical mass of 2.5×10^6 M⊙ within 0.015 parsec from the position of Sgr A*, and a lower limit of 6.5×10^9 M⊙ pc^{-3} on the mass density [24]. The latter value is significantly higher than the density of normal star clusters, but still leaves some room for a compact cluster of stellar-mass black holes.

The diffraction-limited images were obtained against turbulence in the Earth's atmosphere, which tends to greatly reduce the image quality and resolving power (in such a case the image would be "seeing"-limited). To solve this problem, a technique called speckle imaging was first introduced. Since the typical timescale of turbulent motion is one second, this technique utilizes a sensitive detector to take short exposures on the order of one tenth of a second, so as to avoid the influence of turbulence and obtain high-quality images of stars. However, the short exposures mean that this method has low efficiency and is restricted to very bright stars. These limitations were finally overcome by a revolutionary technology called adaptive optics. The technology was invented as early as 1953 and successfully applied to astronomical observations in the late 1980s [25]. The working of adaptive optics requires the presence of a bright "natural" star near the target, or the use of laser beams to create an "artificial star" in the upper atmosphere. When the light of the reference star (natural or artificial) passes through the atmosphere, it carries the information of atmospheric turbulence. This information is captured by a deformable mirror and a wavefront sensor on the telescope and fed to a high-performance computer to make real-time correction of the star image, enabling a resolution at the diffraction limit.

At about the turn of the millennium arrived the 8-10 m class telescopes, in particular the European Southern Observatory's Very Large Telescope (VLT) located on the mountain Cerro Paranal in Chile, and the W. M. Keck Telescope on Mauna Kea, Hawaii, both equipped with an adaptive optics system. The VLT and Keck telescopes greatly boosted studies of the Galactic center. Two teams, one using the VLT and led by Reinhard Genzel of the

Max Planck Institute for Extraterrestrial Physics (MPE) of Germany, and the other using Keck and led by Andrea Ghez of the University of California, Los Angeles (UCLA), of the United States, have intensely conducted dedicated observations of the central parsec for more than two decades.

The VLT and Keck observations have led to an ever-growing number of individually detected stars belonging to the nuclear star cluster. A large fraction of the newly discovered stars turned out to be early-type massive stars. Among them, a remarkable concentration of B-type stars, later referred to as the "S-stars" by the MPE group, or "S0-stars" by the UCLA group, was found within 1 arcsec (corresponding to a linear scale of only 0.04 parsec) from the position of Sgr A*.

The S-stars play a key role in making the GCBH a convincing case. Long-term monitoring of the trajectories of these stars show that they follow elliptical orbits on the sky, with a common focus nearly exactly coincident with the position of Sgr A* [26, 27], as illustrated in Figure 2. Among these stars, S2 (S0-2 in the convention of the UCLA group) has the shortest orbital period (16 years) and has recently completed its second pericentric passage since the launch of the MPE/UCLA monitoring programs [28, 29], where it is only 120 AU (or about 17 light-hours) away from Sgr A*. The latest analysis of the kinematic data of S2 (including two-dimensional trajectory and radial velocity) shows that a dark mass of 4.152 $(\pm 0.014)\times 10^6$ M⊙ is required to keep it on the elliptical orbit [30]. The uncertainty in the mass is now dominated by the correlated uncertainty in our knowledge about the Sun's distance from the Galactic center, which is involved in the conversion of the angular size of the stellar orbit into a physical size.

Taking into account the fact the size of Sgr A* is less than one AU, the density of the dark mass is constrained to be no less than 4×10^{23} M⊙ pc^{-3} [14]. This enormous mass density is many orders of magnitude higher than can be reached by any normal matter, and in fact, is not far from the mean density of a black hole of four million solar masses inside its Schwarzschild radius (1.7×10^{25} M⊙ pc^{-3}). Thus, the only reasonable explanation for this case is the existence of a massive black hole at the Galactic center.

SUMMARY

The research of black holes has taken a few key steps. It all started with Einstein's general theory of relativity published in 1915. In the next year, Schwarzschild obtained the black hole solution of Einstein's field equation under the assumption of spherical symmetry, in which a singularity is enclosed by an event horizon. In 1965, motivated by the problem of quasar energetics, Penrose proved that the existence of singularity and the formation of a black hole are a robust prediction of general relativity. With a decade-long effort into the 21st century, Genzel and Ghez independently made accurate measurements of stellar orbits and finally obtained overwhelmingly strong evidence for the existence of a massive black hole at the center of our Galaxy. Event Horizon Telescope (EHT) observations are now seeking to obtain a direct imaging of this black hole [31].

In parallel, the study of black holes has become one of the most active fields in modern astrophysics. We now know that black holes are ubiquitous in the Universe. For instance, it is thought that hundreds of millions of stars in a galaxy similar to the Milky Way will ultimately evolve into small black holes, and almost every such galaxy hosts a massive black hole at its center. Besides quasars, a wide array of astronomical phenomena and astrophysical processes are driven by black holes [32].

Acknowledgements: This work is supported in part by the National Key Research and Development Program of China (Grant Nos. 2016YFA0400704 and 2017YFA0402703), the Natural Science Foundation of China (Grants Nos. 11633006 and 11873028), and the Key Research Program of the Frontier Sciences of CAS (No. QYZDJSSW-SYS008).

References

[1] K. Schwarzschild, 1916, "Über das Gravitationsfeld enies Massenpunktes nach der Einstein'schen Theorie", Sitzungsberichte der Königlich-Preussischen Akademi der Wissenschaften, 1916.
[2] J. R. Oppenheimer & H. Snyder, Physical Review, 56, 455 (1939).
[3] R. Penrose, Physical Review Letters, 14, 57 (1965).
[4] M. Schmidt, Nature, 197, 1040 (1963).
[5] E. E. Salpeter, ApJ, 140, 796 (1964).
[6] Ya. B. Zeldovich, Dokl. Akad. Nauk SSSR, 155, 67 (1964).
[7] D. Lynden-Bell, Nature, 223, 690 (1969).

[8] D. Lynden-Bell & M. J. Rees, MNRAS, 152, 461 (1971).
[9] J. H. Piddington & H. C. Minnett, Aust. J. Sci. Res., A4 495 (1951).
[10] R. X. McGee & J. G. Bolton, Nature, 173, 985 (1954).
[11] B. Balick & R. L. Brown, ApJ, 194, 265 (1974).
[12] R. L. Brown, ApJ, 262, 110 (1982).
[13] D. C. Backer & R. A. Sramek, ApJ, 260, 512 (1982).
[14] M. J. Reid & A. Brunthaler, ApJ, 892, 39 (2020).
[15] R. L. Brown & K. Y. Lo, ApJ, 253, 108 (1982).
[16] F. K. Baganoff et al., Nature, 413, 45 (2001).
[17] S. S. Doeleman et al., Nature, 455, 78 (2008).
[18] M. J. Reid & A. Brunthaler, ApJ, 616, 872 (2004).
[19] E. E. Becklin & G. Neugebauer, ApJ, 151, 145 (1968).
[20] E. E. Becklin & G. Neugebauer, ApJ, 200, L71 (1975).
[21] E. R. Wollman et al., ApJ, 205, L5 (1976).
[22] J.-H. Zhao et al., ApJ, 699, 186 (2009).
[23] M. T. McGinn et al., ApJ, 338, 824 (1989).
[24] A. Eckart & R. Genzel, Nature, 383, 415 (1996).
[25] J. M. Beckers, ARA&A, 31, 13 (1993).
[26] A. M. Ghez et al., ApJ, 689, 1044 (2008).
[27] S. Gillessen et al., ApJ, 692, 1075 (2009).
[28] Gravity Collaboration, A&A, 615, L15 (2018a).
[29] T. Do et al., Science, 365, 664 (2019).
[30] Gravity Collaboration, A&A, 625, L10 (2018b).
[31] K. Asada & M. Nakamura, AAPPS Bulletin, 30(3): 6-11 (2020).
[32] G. Kang, AAPPS Bulletin, 28(1): 25-29 (2020).

Zhiyuan Li is a Professor of Astronomy at Nanjing University. He obtained his Ph.D. degree at the University of Massachusetts in 2009. He was then a postdoctoral fellow at the Smithsonian Astrophysical Observatory and an Assistant Researcher at the Physics and Astronomy Department of UCLA, before joining the faculty of Nanjing University in 2013. His research focuses on high-energy astrophysical processes on both stellar and galactic scales, using a multi-wavelength approach. Recently he's also spent time in high-energy instrumentation.

Feng Yuan is a professor of astrophysics, Shanghai Astronomical Observatory, Chinese Academy of Sciences. After receiving his PHD in University of Science and Technology of China in 1997, he worked in Nanjing Normal University, Nanjing University, Max-Planck Institute for Radio Astronomy, Harvard University, and Purdue University before joining Shanghai Astronomical Observatory in 2005. His research field is black hole astrophysics.