Cambridge University Press
978-1-107-03918-6 - Revealing the Heart of the Galaxy: The Milky Way and Its Black Hole
Robert H. Sanders
Excerpt
More information

1

Introduction: The Luminous Pathway

Nothing can be more spectacular than the nocturnal moonless sky, far from big city lights and haze. This is especially true in the summer, or so it seemed to me in the small Texas town of my childhood. Even in that soupy semitropical climate near the Gulf of Mexico there were nights when the sky was brilliant with stars of all brightness and colors. It appeared to me not as a flat canopy but as a void with depth – three-dimensional – the brighter objects so close that I could almost touch them and the fainter ones fading away to infinity. And through it all ran the luminous band of the Milky Way flowing north to south from Cassiopeia down to Sagittarius with a conspicuous bifurcation halfway between in Cygnus. What was it? What comprised this shining ribbon, present in the same form night after night in the summer sky? I think that it was the appearance of the Milky Way that stimulated my early obsession with astronomy, an obsession that I never outgrew.

Primitive peoples gazed upon this same celestial spectacle and, given their intimate proximity to nature and the absence of interfering human sources of light, were certainly more aware of its appearance and constancy than are we. Although we have no idea of the mystical or superstitious or anthropomorphic associations prehistoric humans assigned to this phenomenon, the mythology of ancient peoples does present several consistent images invoked to explain, or describe, the Milky Way. The first is that of spilt or wasted milk, but milk with a luminous or magical quality. A story out of Greek mythology is that Zeus, having fathered a son, Hercules, with the mortal Alemene wished him to be endowed with god-like qualities, and so, as his wife Hera was sleeping, placed the infant on her breast to suckle the milk of gods. Hera, apparently not an abnormally heavy sleeper, awoke and pushed the strange infant away, spraying the divine liquid across the heavens. For the Egyptians the Milky Way was also a pool of cow milk produced by a

Cambridge University Press
978-1-107-03918-6 - Revealing the Heart of the Galaxy: The Milky Way and Its Black Hole
Robert H. Sanders
Excerpt
More information

heavenly herd evidenced by the stars. And then there is the image of a stream or river. For Hindus it was the Ganges of heaven; for the Australian aboriginals it was the sky river with creature dwellings along its bank. The image I prefer is that of a pathway or heavenly street as described by Ovid in Metamorphosis:

> When the nighttime sky is clear, there can be seen
> a winding highway visible in heaven, named
> the Milky Way, distinguished for its whiteness.
> Gods take this path to the royal apartments
> of Jove the Thunderer; on either side
> are the palaces with folding doors flung wide,
> and filled with guests of their distinguished
> owners; plebeian gods reside in other sections,
> but here in this exclusive neighborhood,
> the most renowned of heaven's occupants
> have *their* own household deities enshrined;
> and if I were permitted to speak freely,
> I would not hesitate to call this enclave
> the Palatine of heaven's ruling class.
> (translation by Charles Martin)

So the Milky Way is the main street of an extremely upscale neighborhood – a silvery Mulholland Drive, meandering its way through a celestial Beverly Hills where the stars dwell.

When I was about nine or ten years old, I read that the appearance of the Milky Way was actually due to uncountable faint stars along the line of sight in the great disk of our Galaxy – stars too faint to be seen individually but, combined, created the appearance of a continuous band of light. But this knowledge remained rather theoretical until age twelve, when my father gave me a small refracting telescope (40-mm aperture). When I pointed the telescope at the Milky Way I actually saw these countless stars filling the field of view. I can only image how startled Galileo must have been when he turned his small self-made telescope to the Milky Way and became the first human to discern its true nature.

We live in a spiral galaxy, a great disk of stars turning ponderously about its center supported against its own gravity by centrifugal force. All the stars that we can see at night belong to this vast stellar system; the stars we can individually discern are close to the Sun and do not appear to be confined to the disk. My early perception of depth in the sky was somewhat of an illusion, because the brighter stars are not generally closer; they are more luminous. The disk contains not only stars but also a very tenuous gas out of which new stars are continuously being formed. Within this gas there are small solid particles, dust that obscures the light of stars behind. It is this interstellar dust in the plane of the Milky Way that creates the appearance of the "great rift" in Cygnus – the apparent splitting of the Milky Way. This dust obscures all visible light beyond a few hundred light years.

Up to the early twentieth century, before astronomers knew about the dust and appreciated its importance in dimming the star light, they perceived that we were near the center of this disk, a somewhat privileged position. Distributed more uniformly around the sky, not lying in the plane of the disk, were various other fuzzy objects such as globular star clusters and "spiral nebulae." But the Milky Way disk with us at the center was perceived to be the principal constituent of the Universe in this rather non-Copernican world-view.

Now that astronomical distance can be determined with precision we understand that the Milky Way is only one of these "spiral nebulae," or rather *galaxies* – actually great star systems themselves lying at previously unimaginable distances from our Galaxy. Moreover, now that the effect of the interstellar dust is understood and quantified, we realize that we are not at the center but at the outer edge of the disk; the center lies 26 000 light years away from us in the constellation of Sagittarius. We can lift this veil; we can peer through the dust by looking at electromagnetic radiation of longer wavelength than visible light – radio waves and infrared radiation – and directly observe the large-scale structure of the Milky Way. As in other distant spiral galaxies we see that the disk of the Milky Way is only one component of the Galaxy. Surrounding the center is a spheroidal system of stars, the "bulge," which is not rotating – at least not so rapidly as the disk; the stars of the bulge are moving with high random velocity.

Proceeding to the center of the Galaxy, the density of stars in the disk and bulge increases. Locally, there is about one star per 30 cubic light years; if this stellar material were smoothed out through space it would amount to about one atom per cubic centimeter – an incredible vacuum. But in the inner parsec of the galaxy, the density of stars has increased to about 300 000 per cubic light year, all moving with random speeds of about 100 km/s. The density of stars is so high here that actual collisions of stars must occur, although not as frequently as we might expect. In the very central region there is one such collision about once per 10 000 years, and the probability of any given star undergoing a collision is small. Even though a collision might be a tremendously violent event, the average rate of producing luminous energy from stellar collision is rather modest: less than 1000 times the luminosity of the Sun or equivalent to a star with ten times the mass of the Sun.

But other even more bizarre phenomena are present in the inner region of the galaxy. There are several thousand massive bright young stars in this region, which implies intense star formation in the past few million years; this is very peculiar given the relatively low density of gas in the inner parsec. And at the very center there is a point-like source of radio emission – emission over a range of microwave frequencies, so-called continuum emission. This source has been designated “Sagittarius A*” or Sgr A* (commonly pronounced “Sag A star”). Although the total power of this radio source is not large, it is very compact – essentially smaller than 100 astronomical units (100 times the distance from Earth to the Sun). Moreover, after correcting for the orbital motion of the Sun about the center of the Galaxy, its position does not change; it remains rock solid at the center. Several of the young bright stars are clearly in elliptical orbits about this radio source just as the planets are in orbit about the Sun. By tracing the orbits over time we can estimate the mass of this object, and it turns out to be more than four million times the mass of the Sun – four million times the mass of the Sun in a region smaller than 100 Astronomical Units (AU) Sgr A* is without doubt the most peculiar object in the Milky Way Galaxy and, indeed, is associated with one of the weirdest constructs conceived by theoretical physics – a massive black hole.

When I finished high school, the only aspects of my life that I was certain of were: (1) I wanted to get away from high school and (2) unlike many of my teenage cohorts, I was not ready to settle down to serious life (job, wife, babies, and so forth). So I looked around and decided to attend the only reasonable and inexpensive university I could find in my neighborhood; that was Rice University in Houston. I remained obsessed with astronomy, but to my father, a solid down-to-earth type, that seemed incredibly fanciful and impractical (little did he realize the long-term pernicious effects of that seemingly innocent gift of a refracting telescope). So I had to hide, or rather camouflage, this interest. I majored in physics, which at least appeared in some respects to be applicable to the real world of jobs and money. At Rice, after flirting with chemistry, mathematics, and philosophy (there was very little else for a heterosexual male to flirt with at Rice in those days) I stayed with physics and, in fact, became quite fascinated with the more theoretical and mathematical aspects (thanks to several motivating teachers such as Stephen Baker). But I remained a closet astronomer. And when it came time for graduate school, I “came out” and made the jump to astronomy at Princeton.

The Princeton University Observatory was a different sociological world than I had known before. It was my introduction to a scientific community that I have previously compared to an extended family. In those days, the senior and dominant figures at the observatory were two remarkable characters: Lyman Spitzer and Martin Schwarzschild. They were both extremely eminent scientists with different but complementary talents. To me, arriving from Texas to the big time world of Ivy League science, Spitzer and Schwarzschild were imposing and, initially at least, a bit scary because to a large extent they had defined modern post–World War II astrophysics. But, at the same time (I quickly discovered), they were very personable and absolutely without arrogance. At Princeton in those days there was much the spirit that we were all marching together sweeping forward the boundaries of knowledge (and graduate students, although foot soldiers, were certainly part of that march).

Astrophysics is, in a real sense, applied physics: the concepts of modern physics are applied to astronomical problems. Moreover, explanations of astronomical phenomena are sought in terms of *known* physics. Occasionally, very rarely, new physics may be discovered, or at least hinted at, by means of astronomical observations. But the first impulse is to seek an explanation in the context of textbook physics. This is probably a good thing because otherwise quite fanciful and unnecessary excursions arise while constructing "theories" underlying observations having a completely conventional explanation. The point is that there is a built-in conservatism to astrophysics; new physics, when it comes, most often comes from the physics community.

So it was at Princeton. The staff at the observatory set very high standards for astrophysics but were inherently conservative with respect to new physical constructs. Across the street, however, at the Palmer Physics Lab, it was a different story. There was a group of wild insurgents led by the chief radical, John Wheeler. Now, John Wheeler seemed nothing at all like a extremist; he was soft-spoken, extremely polite, and, in class, treated all students with great respect, even when they didn't deserve it. But in the realm of ideas he was a true revolutionary. He worked on and coined the expression "black hole."

At that time – 1966, 1967 – the idea of black holes seemed an extremely bizarre concept. Only several groups in the world were thinking about such hypothetical objects: at Caltech, Cambridge and Oxford, Moscow, and Princeton. Now, in the modern world, everyone who is reasonably sentient, has heard the term "black hole" and has a rough idea of what it is: a mass that is so concentrated that nothing, not even light, can escape from it. But actually these objects have even more peculiar properties. In this period of the late 1960s several significant theorems on the nature of black holes were being developed, the most important of which was the singularity theorem of Roger Penrose at Oxford. He proved that, in the context of General Relativity, mathematical singularities may exist in space–time and, in a sense, are inevitable. To physicists this seemed a particularly strange concept. A singularity is a well understood concept where, at a point in space, a function, a mathematical object, "blows up" (approaches infinity) or becomes undefined. That's fine in mathematics, where the function is a pure abstract entity, but that singularities should actually exist in nature, where the function is a real physical quantity such as force or density, is very odd indeed. Penrose had proven that if General Relativity is the correct theory of gravity, and there were and are no experimental contradictions to this theory, then such bizarre points must exist (of course, General Relativity may in some sense be "correct," but it is certainly an incomplete reflection of deeper theory).

Apart from these theoretical developments there had been for three decades the distinct possibility that these objects might actually exist in an astrophysical context. In 1930 Subramanyan Chandrasekhar, a young post-doctoral fellow at Cambridge, wrote his famous paper on the upper limit to the mass of white dwarf stars. White dwarfs are very faint small dense stars of about one solar mass (but only 1/100 of the solar radius) and are certainly the end stage of stellar evolution (the example of one teaspoon of white dwarf matter weighing one ton is almost a truism). Stars are stable objects in which two forces are in balance: the gravity force that pulls all matter of the star toward its center and a pressure force that pushes the other way and keeps the star from collapsing. In a normal star such as the Sun this outward force is provided by the usual thermal pressure of hot gas. But a star that has totally exhausted its nuclear fuel and has no way producing heat energy in its interior may also be supported entirely by the degenerate pressure of electrons, a pressure that exists because the electrons are packed so densely within the star. Chandrasekhar had discovered that it was not possible for this packing pressure to support a star if its mass were greater than about 1.4 times the mass of the Sun; there is no force available to support a more massive star against the force of gravity. A massive star at the end of its life that has exhausted its internal fuel source should therefore collapse to the point where light cannot escape; it should become a black hole.

Most stars in the Galaxy are less massive than about two solar masses, but there are a number of quite massive stars – going up to five, ten, or even 100 solar masses. Some of these massive stars certainly conclude their stellar existence in a spectacular supernova explosion. Supernovae are dramatic events in which a star suddenly brightens by many orders of magnitude (it may even outshine an entire galaxy) and then fades from view over an interval of a month or so. The entire star may be disrupted by this event so no remnant is left. But supernovae do not occur frequently enough; in the Galaxy there are perhaps one or two per century, insufficient to take care of all stars above the Chandrasekhar limit. What happens to all of these these massive stars when the nuclear fuel is gone? To avoid the black hole trap they must either lose mass or somehow get around the Chandrasekhar limit.

When I appeared at the Princeton Observatory in 1966, both of these possibilities were being considered. By looking at ultraviolet spectra of hot young stars, Don Morton had discovered that these stars were actively losing mass; perhaps they could lose enough mass during the course of their active lifetimes to push them below the Chandrasekhar limit before the fuel was exhausted. Jerry Ostriker, who had been a student of Chandrasekhar, was considering the effect of rotation on the mass limit; all stars rotate to some degree and perhaps this rotation could provide some additional support against gravity. This was, in general, the approach of astrophysicists: rather than immediately consider bizarre consequences of gravitational collapse, look for a conventional way out. There was the feeling that Nature, somehow, would avoid the bizarre "unphysical" black hole trap.

As it turned out, these escape exits were blind alleys; early mass loss is insufficient to significantly reduce the mass of young stars and rotation can only marginally extend the Chandrasekhar limit. Moreover, the subsequent discovery of neutron stars – compact objects supported by the degenerate pressure of neutrons rather than electrons – does nothing to alleviate the problem of gravitational collapse; neutron stars also have a mass limit that is comparable to that of white dwarfs. Finally, the development and application of X-ray astronomy in the 1970s produced convincing evidence that stellar mass black holes actually do exist; the evidence was in the form of X-ray emission from massive, otherwise dark objects – more massive than the white dwarf or neutron star limits – accreting gas from a nearby binary companion.

But here I am not going to bother with these puny stellar mass black holes, interesting though they are in establishing the existence of this class of objects. I am going to discuss massive, or supermassive, black holes that almost certainly exist in the dense nuclei of galaxies such as the Milky Way – objects of more than one million solar masses.

Back in 1963 when I was an undergraduate at Rice, Maarten Schmidt, a young Dutch astronomer working at Caltech, succeeded in identifying the spectral lines in a puzzling class of objects that had been discovered several years before. These objects were star-like in their appearance but conspicuous sources of radio emission, subsequently designated "quasi-stellar radio sources" or "quasars" (later it was found that most such objects have low or undetectable radio emission; they are radio quiet). Quasars are not preferentially in the plane of the Milky Way, but are distributed more or less uniformly across the sky, like the distant galaxies (an early hint that they are also extragalactic). The spectra had been measured: there were broad, conspicuous emission lines imposed on a bright continuum, but these lines had not been identified with any known substance. Schmidt's breakthrough was his realization that these lines are actually the common lines arising from ionized hydrogen (and nitrogen) but very highly redshifted – implying a recession velocity of 20% to 40% that of the speed of light. If these objects follow the usual Hubble law (distance proportional to redshift), this means that they are incredibly distant – hundreds of millions of light years – at the time, the most distant objects ever detected. Then, because we can see them at all, it also means that they are incredibly luminous – far more luminous than any source ever discovered, 100 or

1000 times more luminous than all of the stars in a major galaxy like the Milky Way. Yet, they did not appear to be galaxies; they seemed to be star-like. So all of this electromagnetic energy arises in an extremely small region by astronomical standards, perhaps smaller than 1 light year. When it was discovered that the total flux of radiation from quasars varied on timescales of weeks or months, then, because they could not be larger than the distance traveled by light on this timescale, the size was even more drastically constrained to less than a light year.

At the time this appeared to stretch the known laws of physics. How could so much energy be emitted from such a compact region? What was the source of all of this electromagnetic radiation ranging, as we now know, from radio emission to gamma rays? Most astrophysicists at that time did what astrophysicists do: they tried to find an explanation in terms of conventional physics (not all did; some proposed that the phenomenon was due to totally new physics of unknown basis – matter recently created and shot out of galaxies, matter with an intrinsic redshift beyond that due to the Hubble expansion).

Lyman Spitzer had developed a “conventional” model. It was generally known that as an isolated stellar system (a star cluster or a galactic nucleus) evolves, it loses stars. Because the ejected stars carry away energy, the remaining system contracts and becomes denser; in time its density approaches infinity. Spitzer realized that before this happened, the stars in the system would actually begin to collide and this could possibly liberate a large amount of energy.

I was the third Spitzer student who worked on this problem (following Bill Saslaw and Mike Stone). Spitzer’s idea was that the collisions would, by and large, disrupt stars; the gas so liberated would cool by normal thermal radiation and collapse to the center of the system. Because of residual angular momentum, the gas would settle into a compact disk, new stars would form and diffuse out of the disk, and the entire system would contract further, accelerating the whole process (the end result of this process was not discussed). The hope was that quasar luminosities could be produced by such a mechanism. But a problem with Spitzer’s model was that the radiation produced was thermal emission from hot gas, whereas the continuum emission observed from quasars was nonthermal (due to relativistic electrons spiraling in a magnetic field). The radiation predicted did not match the radiation observed, and so the model required further “processing” of the radiation outside the dense stellar system.

Then, in 1967, Stirling Colgate, a nuclear physicist, pointed out that before stellar collisions became frequent and disruptive, they would become frequent and soft; these soft collisions would lead to coalescence, not disruption, of the colliding stars and more massive stars would be built up. Colgate estimated that this coalescence process would cease when the stars became more massive than about fifty solar masses (this would happen because the stars are bloated by the extra collision energy), and these fifty solar mass stars would then explode as supernovae on short order. The greatly enhanced supernovae rate would produce the high luminosity and all of the nonthermal emission observed from quasars. Colgate's model appeared to be quite plausible because it addressed the actual observations of quasars.

For my Ph.D. dissertation work, Spitzer asked me to look into this question: does a system of colliding stars lead generally to disruption of the stars and liberation of gas, or to coalescence with increasing average stellar mass? I found that the answer depends somewhat on the properties of the system; for a certain reasonable range of properties coalescence would indeed dominate, but I concluded, contrary to Colgate, that the growth would not stop at 50 solar masses; it would accelerate and not be terminated by usual supernovae explosions – a sort of runaway growth of stellar mass leading finally to one or several very massive stars. I did not speculate what would happen to these massive stars, but they are known to be highly unstable to gravitational collapse.

As later summarized by Martin Rees (Cambridge), this is only one of several possible paths leading to the formation of a massive black hole in a galactic nucleus. It would seem that the formation of black holes, and very big ones, is a natural development in dense galactic nuclei.

But black holes are, after all, black. They are not supposed to shine. How can they be the source of all of this electromagnetic radiation from quasars? The answer is that we do not, of course, see the black hole itself; we see matter being accreted onto a black hole. Before the accreted mass disappears from view, it can radiate; in fact, it may emit up to 10% or 20% of its rest mass as electromagnetic radiation. But luminosities of 100 times that of an entire galaxy would still require something like the mass of ten suns disappearing down the hole every year. So not only is the black hole necessary for the quasar phenomenon, but also a copious supply of "fuel" must be available in the near environs of the hole. Moreover, if quasars live for ten million years (a reasonable supposition given the size of the extended radio structure surrounding quasars) then the black hole must have a mass of one hundred million solar masses.

In 1969, Donald Lynden-Bell (Cambridge), in a seminal paper published in the journal *Nature*, stressed the observational fact that the co-moving number density of quasars (i.e., after taking into account the expansion of the Universe) seemed to be higher in the past than in the present (more and more quasars are seen at large redshift or far in the past). If the quasar phenomenon is due to accretion onto massive black holes in galactic nuclei, then where are all of the dark black holes now? Because massive black holes do not disappear or lose mass, the nuclei of normal galaxies should also contain black holes. Because they are not shining, they must be underfed; these objects are suffering from a fuel crisis.

This, at the time, was quite a radical suggestion. It was known that 2% or 3% of nearby galaxy nuclei are indeed brilliant and are likely to host massive black holes. In 1943 the American astronomer Carl Seyfert had discovered and catalogued several spiral galaxies with bright star-like nuclei and broad emission lines coming from a compact central region. When quasars were later discovered, these Seyfert galaxies seemed to belong to this same class but of lower power. But Lynden-Bell was proposing that most galaxies have gone through Seyfert or quasar episodes. Perhaps black holes are a *normal* component of galactic nuclei and emerge naturally during the course of galaxy formation or evolution. Perhaps the Milky Way contains one of the bizarre objects at its center.

After Princeton, I went on to my first actual job as a post-doctoral fellow at Columbia University in New York City. The staff at Columbia was an eclectic collection of very bright but somewhat cynical characters who created quite a different atmosphere than at Princeton (at the time, I had tired of the generally wholesome boy scout attitude at Princeton and was quite ready for this dose of healthy cynicism). The caustic tone of Columbia was probably related to the surrounding atmosphere of New York, where one continually faced potential assaults on body and soul – but it was exciting. My immediate boss was Kevin Prendergast, a native New Yorker and a charter member of the cynics club, who also happened to be a brilliant astrophysicist and applied mathematician. He had developed a very clever technique for simulating gas dynamics on a computer, the "beam scheme," and he asked me to adapt and apply this method to a specific problem: the focusing of explosions in thin gaseous disks. At the time, explosive phenomena in galaxies – apparently evidenced by so-called "radio galaxies" (in which the optical galaxy was found between two large lobes of radio emission), the more nearby example M82 (supposedly an "exploding galaxy"), and large noncircular gas motions observed in the Milky Way in the direction of the Galactic Center – were quite in fashion (it was an explosive time politically as well). The immediate question was: Could a violent explosion in the center of a thin disk be narrowly focused in two opposite directions to explain the morphology of radio sources? Or could such an event excite noncircular gas motions several thousand light years from the Galactic Center? As it turned out, these speculations were something of a detour in understanding the dynamics of both normal and active galaxies, but this was not apparent at the time. The main road led directly to black holes and their off-and-on fueling.

If Lynden-Bell's suggestion that normal galaxies contain massive black holes were true then even our Galaxy, the Milky Way, will contain a massive underfed black hole and efforts should be made to find it. But if it doesn't shine, how can we find the black hole in the center of the Milky Way? Lynden-Bell suggested that we should look for dynamical evidence for a mass concentration at the Galactic