

# Introduction to Cosmology and Olbers' Paradox

## Introduction to Cosmology

*Cosmology* is our attempt to understand the overall structure, history, and future of the entire Universe

Some of the “big questions” of cosmology include the following:

- How old is the Universe?
- What is the size and shape of the Universe?
- Is the Universe changing with time?
- What will be the fate of the Universe?
- Where did it all come from?

We don't have fully satisfying answers to all these questions, but we have made a lot of progress.

Our general approach here will be to use observations and physical laws to address these questions scientifically.

Our best scientific theory of cosmology is the *hot Big Bang model*.

The basic idea is that the Universe started expanding from a hot, dense state in an event called the Big Bang, and it has been expanding ever since.

The Big Bang event created all matter, energy, and space itself.

We will see that a variety of seemingly disparate observational phenomena can be explained with a common foundation in the hot Big Bang model.

Notably, some of the questions of cosmology seem to border on those of religion and philosophy.

This has been the situation throughout history, as shown in many historical cultural works.

Indeed, we will see that some conclusions from the *science* of cosmology seem, by their nature, to remove them from the realm of science. They become essentially untestable, at least at present. So we will be at the very borders of human understanding.

## **Darkness at Night: Olbers' Paradox**

One of the most basic and important observations of cosmology is that the sky is dark at night.

This is true when we observe the sky ourselves on a dark night, and it remains true when one considers the most sensitive optical images of the Universe ever made by the Hubble Space Telescope. There is "black space" between the galaxies.

Scientists such as Johannes Kepler (in about 1610) and Edmund Halley (in the 1700's) realized the importance of this observation quite early.

The German astronomer Heinrich Olbers popularized this issue around 1826, and it has come to be known as Olbers' paradox (even though Olbers did not make major progress in resolving the paradox).

Why is the sky dark at night? To understand why this is a deeply profound question, let's consider the case of a universe that is infinitely large, infinitely old, and uniformly filled with stars.

In such a universe, one's line of sight in any direction should always intersect a star's surface eventually. So we expect that the entire night sky should be as bright as the surface as a typical star. The night sky would be blazingly bright!

One might be tempted to say that distant stars will be fainter, and that this will resolve the paradox. This is not the case:

- Distant stars will indeed have lower fluxes, but their projected surface areas decrease at the same rate with

distance (as distance squared). So their *surface brightness*, which is flux per unit area, remains the same and is independent of distance. This is why distant stars still appear as bright points of light in the night sky.

- Furthermore, the volume of space sampled becomes larger as the considered distance becomes greater. Even if an individual star may be invisible to the eye due to its great distance, its light would combine with the light of an infinite number of other individually imperceptible stars to light the night sky.

An analogous situation is looking around when you are far into a large forest. No matter where you look, your line of sight will always intersect a tree.

Possible obscuring dust in the Universe does not help the paradoxical situation. The intense starlight would heat the dust until it glowed equally brightly (or evaporated away).

So the night sky should be blazingly bright, provided our Universe is infinitely large, infinitely old, and uniformly filled with stars.

Since this is clearly not the case, one of the assumptions must fail. The possibilities are that

- Our Universe has a finite size.
- Our Universe has a finite age. The light from the most distant stars has not yet reached us.
- Our Universe is infinite but not uniformly filled with stars. For example, there might be no stars at great distances.

Any one of these possibilities would have profound and far-reaching implications!

Olbers' paradox hung over cosmology until well into the 20<sup>th</sup> century.

We now know that stars are not uniformly distributed but are clumped into galaxies, but this is not the solution to the paradox. You can replace the word "star" with the word "galaxy" in the arguments above, and the paradox still holds. We do not find any "special"

clustering of stars or galaxies that can resolve the paradox, as will be demonstrated in detail later in this course.

So we are left with a Universe that is either finite in size or finite in age.

We now know that the Universe has a finite age of about 13.7 billion years, and this is the main solution to Olbers' paradox.

Because of the finite age, there are a finite number of stars in the observable Universe. We are able to look back to a time before stars had formed.

Edgar Allen Poe put this basic solution forward in 1848 in an essay named *Eureka*, which he considered to be his career masterpiece. Part of this essay is the following:

“Were the succession of stars endless, then the background of the sky would present us an uniform luminosity, like that displayed by the Galaxy - since there could be absolutely no point, in all that background, at which would not exist a star. The only mode, therefore, in which, under such a state of affairs, we could comprehend the voids which our telescopes find in innumerable directions, would be by supposing the distance of the invisible background so immense that no ray from it has yet been able to reach us at all.”

# The Size and Age of the Universe

## Introduction and the Parallax Method

Being able to measure cosmic distances reliably is central to the topic of cosmology. Without this, we cannot set the scale of the Universe in which we live.

An enormous amount of effort in the astronomical community is spent trying to get the distances to cosmic objects.

Robin Ciardullo, a professor in the Department, has spent much of his astronomical career working on this.

Let's first talk about how we have determined the size of our Galaxy.

For example, how have we determined that the distance to the Galactic Center is 8.5 kpc?

The most basic method to measure the distances of nearby objects is *parallax*.

Demo with finger, closing and opening each eye

- Finger at ~ 4 inches from nose
- Finger at arm's length

The basic rule is that far away objects move less.

Astronomers have become very skilled at using this exact method to get distances. Have calibrated it, and in fact this is the origin of the parsec (pc) unit:

$$(\text{Distance to an object in pc}) = \frac{1}{(\text{Parallax shift angle in arcsec})}$$

Can use this method with two telescopes located far apart on Earth.

Or, even better, can use the motion of the Earth around the Sun to get a long baseline of 2 AU = 299.2 million km.

The science of measuring celestial positions to high precision is known as astrometry.

The Hipparcos (High-Precision Parallax Collecting Satellite) mission is the best astrometric mission we've had so far. It operated from 1989-1993.

- Named after Hipparchus of Nicea (today's Turkey), who made the first astrometric catalog of about 1000 stars. Hipparchus lived from about 190-125 BC.
- The satellite got above the atmosphere, where the stars no longer "twinkle." This allowed very precise positional measurements.
- Could measure positions with incredible accuracy
  - ≈ 1/1000 arcsec
  - ≈ 0.0000003°
  - ≈ The angular size of an astronaut on the Moon
- Measured parallax distances for ≈ 120,000 stars out to 500 light years. This is still only 0.15 kpc, about 1/50 of the distance to the Galactic center. So other methods are needed to measure larger distances.
- The derived distances are typically reliable to within 5-30%.

There are plans for even better astrometric satellites in the future, such as one called GAIA. This would measure distances to about one billion stars throughout much of our Galaxy. GAIA would aim for an astrometric accuracy of a few tens of microarcseconds.

## Measuring Cosmological Distances

Most other methods involve a common principle. I'll call it the "2 of 3 principle."

If you know 2 of the 3 following things, you can get the other:

1. Luminosity
2. Apparent brightness (flux)

### 3. Distance

For example, imagine driving a car at night. There is a red light ahead. You know (1) luminosity and can observe (2) apparent brightness, so you can estimate the distance to the light.

The mathematical relation is just the inverse-square law for radiation:

$$\text{Flux} = \frac{\text{Luminosity}}{4\pi (\text{Distance})^2}$$

All other distance methods involve using some trick to get

1. Luminosity *and can then measure*
2. Flux *to get*
3. Distance

One of the main methods involves *Cepheid variable stars*.

Some stars, as they evolve, develop instabilities in their envelopes that cause them to pulsate in luminosity. Basically, the star's size and temperature change back-and-forth in a regular way, and this causes the luminosity to pulsate in a periodic manner. The size changes by 10-20%.

The most relevant variable stars for us are the Cepheid variables, named after their prototype Delta Cephei (the fourth brightest star in the constellation of Cepheus). John Goodricke noted Delta Cephei to be variable in 1784.

Cepheid variable stars have

- 2-60 day periods
- Characteristic “shark fin” shaped light curves
- Large luminosities of  $\approx 100\text{-}20,000 L_{\odot}$
- Masses of  $\approx 5\text{-}10 M_{\odot}$
- Radii of  $\approx 100\text{-}1000 R_{\odot}$
- Yellow colors, corresponding to surface temperatures around 6000 K

Why do Cepheid variables pulsate? A simple, stripped-down explanation is the following:

- When a Cepheid is compressed, it becomes opaque.
- Photons are trapped inside, heating the gas and increasing its pressure.
- The high-pressure gas expands, becoming transparent.
- Photons escape, the gas cools, the pressure drops.
- As the pressure drops, the Cepheid is compressed by gravity. Thus, the cycle repeats.

The reason that Cepheid variables pulsate, while other stars do not, deals with the detailed nature of the opacity in them.

Henrietta Leavitt of Harvard made a key discovery about Cepheid variables around 1908-1912. She systematically studied Cepheids in the Magellanic Clouds.

She found the *Period-Luminosity Relation*:

(Average luminosity of Cepheid) proportional to (Pulsation period)

High-luminosity Cepheids pulsate slowly

Low-luminosity Cepheids pulsate quickly

A Cepheid with a period of 3 days is about 800 times more luminous than the Sun.

A Cepheid with a period of 30 days is about 10,000 times more luminous than the Sun.

So, you can watch a Cepheid to determine its pulsation period and average flux. Then you can calculate the luminosity with the Period-Luminosity relation. Then you have luminosity and flux, so you can get the distance using the “2 of 3 principle.”

People use exactly this method to get distances throughout our Galaxy and to other nearby galaxies.

It sounds a little crazy to use some pulsating star as a “distance machine”!

But it actually works well, especially since Cepheids are so luminous that they can be seen far away.

Astronomers have now been able to calibrate the Period-Luminosity relation for Cepheid variables directly using Hipparcos parallax distances. There are about 20 reliable distances for Cepheid variables from Hipparcos (more than 200 Cepheids were measured, but the majority of the measurements lacked sufficient precision for reliable distance work).

Cepheids have long been used for our Galaxy, the Magellanic Clouds, the Andromeda Galaxy (M31), and other nearby galaxies.

With the Hubble Space Telescope, it is possible to study Cepheid variables out to distances of  $\approx 20$  Mpc. This allows one to reach the Virgo Cluster.

We can also use other types of stars where we can estimate the luminosity:

- RR Lyrae stars
- Planetary nebulae
- Novae
- Main-sequence stars

Now, Cepheid variables can get one out to  $\approx 20$  Mpc, but this still is a small distance compared to size of Universe. Beyond this even Cepheids are too faint to see.

We see distant galaxies and quasars  $\sim 30,000$  Mpc or more away, and we certainly cannot use Cepheids for these.

We can try to get the luminosities of other objects, more luminous than Cepheids, in galaxies.

For example, the *brightest globular cluster* in a typical spiral typically has  $\approx 10^6 L_{\odot}$ , about 100 times the luminosity of a Cepheid. So assume all brightest globular clusters have similar luminosities.

Also can use *Type 1a Supernovae*

- Explosions of white dwarf stars in binary systems
- Push the white dwarf over the Chandrasekhar mass via accretion, and it will explosively carbon burn
- People have discovered that the light-curve rate of decay correlates with the peak luminosity
- So can use the “2 of 3 principle”

Ultimately, these methods and others allow one to calibrate the use of *redshift* as a distance indicator. I will say much more about the redshift later.

## The Age of the Universe

In addition to the scale of our Universe in size, it is equally important to understand its scale in time.

I'll tell you the answer – we think the Universe is about 13.7 billion years old.

Good to start with the Earth and other objects in our Solar System, since we can study them in detail. They set a lower bound to the age of the Universe.

We can measure ages of Earth rocks, Moon rocks, and meteorites. We use a technique called isochron dating, which uses radioactive elements such as Rb, K, Sm, Lu, Re, Th, U. These have half-lives of 1.25-48.8 billion years.

Isochron dating is a robust and well-established technique. Cross-checking for consistency can be done by different labs around the world.

Oldest rocks found on Earth, from western Greenland, have ages of 3.7-3.8 billion years. Very old rocks also found in southern Africa,

western Australia, and the Great Lakes region of North America. So the Earth must be at least this old, and likely older since it takes time for rocks to solidify from the originally molten Earth.

Oldest Moon rocks are from the lunar highlands. They were returned by the Apollo missions. They have ages of 4.4-4.5 billion years.

The majority of about 70 well-dated meteorites have ages of 4.4-4.6 billion years. These are fragments of asteroids and represent some of the most primitive material in the Solar System.

So, we think the Solar System is about 4.6 billion years old, and this is a lower bound to the age of the Universe.

We can also study the ages of the oldest star clusters, exploiting our understanding of how long stars live. By studying the globular clusters in our Galaxy, people have deduced that they are at least 10-12 billion years old.

We can also estimate the ages of the oldest white dwarfs. White dwarfs glow from the residual heat from when they were the cores of stars like our Sun. The oldest white dwarfs will thus be the coldest and faintest ones. We can calculate how long it takes for white dwarfs to cool, and we deduce that the coldest white dwarfs (located in globular clusters) are at least 12 billion years old.

So there is a pleasing coincidence of ages here, at least considering the orders of magnitude involved (thousands vs. millions vs. billions vs. trillions). We are getting ages of billions of years consistently. If our methods lacked validity, we would not expect this coincidence of ages.

The size of the Universe also sets constraints upon its age. For example, the Andromeda Galaxy is about 2.5 million light years away. So we see it as it was 2.5 million years ago. So the Universe must be at least 2.5 million years old. Similarly, we see the most distant quasars and galaxies (at  $z \sim 6.4$ ) as they were about 12.8 billion years ago. The details of this argument rest upon a deeper understanding of cosmology, and we will return to this later in the course.



# The Hubble Law and the Expansion of the Universe

## Basic Concepts and the Discovery of the Hubble Law

When we observe galaxies using a telescope equipped with a spectrograph, we can see characteristic emission and absorption lines in their spectra.

These spectral lines are made by atomic transitions going on in the stars and gas of the galaxy. Each element has a characteristic line pattern, which can be reproduced in labs on Earth.

A remarkable finding, first made in  $\approx 1912$  by the American astronomer Vesto Slipher, is that almost all galaxies (aside from a few nearby ones) have “redshifted” spectra. Slipher was working at the Lowell Observatory in Flagstaff, Arizona.

Their spectral lines are shifted toward the red part of the spectrum.

This effect can initially be interpreted as being due to the Doppler effect, although this interpretation is not exactly correct as we’ll see later.

For sound:

Train approaching — higher frequency

Train receding — lower frequency

Same effect applies for light:

Object approaching — higher frequency — bluer

Object receding — lower frequency — redder

So it appears that galaxies in the Universe are undergoing *universal recession*.

At the time when Slipher discovered this, people weren’t entirely sure what galaxies were. Some still thought they might be forming planetary systems, for example.

However, the velocities Slipher found for some of his receding galaxies were astonishingly large — hundreds of km / s.

These were the largest velocities measured in astronomy at that time. Such velocities alone convinced many astronomers that the nebulae could not be in our Galaxy.

Slipher won the gold medals of the Royal Astronomical Society and the Astronomical Society of the Pacific for his discovery.

Now, the history that followed Slipher's discovery is quite complex with many claims, counterclaims, false starts, scientific duels, etc.

I'll just focus on the main ideas, and you can read the fascinating history in the book:

*Man Discovers the Galaxies*  
R. Berendzen, R. Hart, D. Seeley  
Columbia University Press (1984)

The next major advance occurred around 1929, after people had shown that the galaxies were outside our Galaxy and had measured the distances to ~ 20 galaxies.

Edwin Hubble, working with Milton Humason, noted in this year that the recessional velocities of galaxies depended on the distances to them.

Now, I need to define a few mathematical terms used by astronomers.

First redshift =  $z$

Defined to be

$$z = \frac{\text{Shift in } \lambda \text{ of light}}{\text{The } \lambda \text{ of what the light should be}}$$

For example, hydrogen makes a line at 650 nm.

We observe this line at longer  $\lambda$  (redder  $\lambda$ ) due to redshift, say at 680 nm.

Then

$$z = \frac{680-650}{650} = \frac{30}{650} = 0.0462$$

Now, the redshift can also be easily transformed into velocity, provided the velocity is small compared to speed of light.

In particular, for  $v \ll c$

$$\begin{aligned} \text{Velocity} &= (\text{Speed of light}) z \\ &= (300,000 \text{ km / s}) z \\ &= (300,000 \text{ km / s}) (0.0462) \\ &= 13,800 \text{ km / s} \end{aligned}$$

Second, Hubble made a law relating distance and velocity in his plot. This law was a linear relation.

In particular, today we write

$$(\text{Recession velocity in km/s}) = H_0 (\text{Distance in Mpc})$$

$$v = H_0 D \quad H_0 \text{ is called the "Hubble constant" or "Hubble parameter"}$$

Or we can write

$$H_0 = v / D = cz / D$$

$$\text{So } H_0 \text{ has units } \frac{\text{km/s}}{\text{Mpc}}$$

We've been able to test this relation quite well today — can see the "Hubble flow."

Note the relation is not perfect, there is some "scatter" around the line of exact proportion.

This is because galaxies, aside from their motion via the Hubble Law, have their own small velocities too.

These small velocities explain the negative, approaching velocities for a few nearby galaxies. We'll talk more about this later.

However, overall the law works quite well.

Empirically, this means that we can get the distance to a galaxy just by measuring its recessional velocity.

$$D = H_0 / v$$

Provided we can determine  $H_0$  accurately.

There have been great efforts to measure  $H_0$  well. Over the years, these have led to major revisions of  $H_0$  downward as some important systematic errors have been uncovered. Hubble's original values for  $H_0$  were too large by a factor of  $\sim 10$ .

Our best measurement for  $H_0$  today is  $H_0 \approx 73 \frac{\text{km/s}}{\text{Mpc}}$

The uncertainty on this value is only about 4% due to improved observations in recent years.

This technique for getting distances is much easier than studying Cepheid variables, supernovae, etc. Only one spectrum is needed.

## **The Expansion of the Universe**

Aside from just its technical utility, however, Hubble's Law has a deep and fundamental meaning.

It implies that the Universe is not steady and unchanging on the largest scales. Rather its contents are in motion, an organized motion in fact.

Now, why should all the galaxies be receding from us? Do we occupy a special place in the Universe, at the center of some grand expansion?

This seems unlikely given the Copernican principle!

The best interpretation we have for the Hubble Law is that space itself is expanding, carrying the galaxies with it.

This can be understood via some analogies, although all these analogies have limitations and should not be taken too seriously.

One good analogy is a loaf of raisin bread expanding uniformly.

Expansion of the bread means that the raisins are moving away from one another — like the galaxies do.

Also, raisins further apart move apart faster since there is more expanding dough between them. Like the Hubble Law.

Also, the expansion is not centered on any one raisin — no special privileged location of the expansion.

Another analogy is to consider an ant on the surface of a balloon.

This basic picture can also explain the observed redshifts of galaxies.

As light travels through expanding space from a distant object, its wavelength is expanded. This means redder light.

This is the correct interpretation of the cosmological redshift. It is not really a Doppler shift, because the galaxies are not moving relative to the Universe.

## **General Relativity and the Expansion of the Universe**

The best theory we have to describe space, time, and gravity is General Relativity, which was developed by Einstein from 1907-1915.

Almost immediately after the publication of Einstein's General Relativity, people started trying to derive its implications for the Universe as a whole.

This was done first by Einstein in 1917, and later by scientists including Willem de Sitter, Alexander Friedmann, and Georges Lemaitre.

These efforts all made an important simplifying assumption known as the *cosmological principle*. This is the assumption that, on large cosmic scales, the Universe is homogeneous.

Einstein found that the most straightforward version of his equations seemed to predict a dynamic Universe, one that was in a state of either collapse or expansion.

Note that during Einstein's early work, people still did not understand the relation of other galaxies to our own. Edwin Hubble did not show the existence of other galaxies until about 1924. Many thought that our Galaxy was the entire Universe, and it was known that our Galaxy was not collapsing or expanding. This was of concern to Einstein.

Hence, Einstein introduced the *cosmological constant*. This constant could be added into General Relativity in a mathematically legitimate way. It described a repulsive "force" that counteracted the tendency of gravity to cause the Universe to collapse. The effects of the cosmological constant were negligibly small on the scale of the Solar System and only became significant over cosmic distances.

With the cosmological constant, Einstein was able to construct a "static" Universe that seemed to be in agreement with the (confusing and misleading) observations of 1917. However, this static Universe was not very satisfying, since it was unstable to collapse. If one arranged the cosmological constant and gravity to counterbalance each other at a given time, they would not remain counterbalanced at later times.

Also in 1917, the Dutch astronomer Willem de Sitter derived alternative models for the Universe using General Relativity. He

considered nearly matter-free Universes that expanded exponentially due to the cosmological constant.

Somewhat later (1922-1927), Alexander Friedmann and Georges Lemaitre developed their own models of dynamic universes.

Eventually, in about 1929, it became clear through the work of Hubble that the Universe is indeed dynamic and in a state of expansion. Einstein then tried to “retract” the cosmological constant in 1931, thinking of it as a major blunder.

Had Einstein stood by the most straightforward version of his equations, he could have predicted the dynamic nature of our Universe before Hubble observed this.

Notably, the cosmological constant is still with us today, so it may not have been such a blunder. This will be discussed in detail later in the context of dark energy.

## The Age of the Universe

The observed expansion implies the Universe was smaller in the past. At some point in the past, it must have been in a very small state. So the Universe cannot be infinitely old. How old is the Universe?

We think the Universe started from a very compact state in an event called the “Big Bang.”

Let’s first of all look at the units of the Hubble constant.

$$H_0 = 73 \frac{\text{km/s}}{\text{Mpc}}$$

(=D / v)

$$H_0 = 73 \frac{\text{km}}{\text{s Mpc}}$$

(1 Mpc =  $3.08 \times 10^{19}$  km)

$$H_0 = 2.37 \times 10^{-18} \text{ 1 / s}$$

So the Hubble constant has dimensions of 1/time — remember this!

Now let's consider two galaxies in the Universe today. Take them to be separated by a distance  $D$ . In the past, these galaxies were closer to each other.

At one time they were touching, and this is the time of the Big Bang.

So following Hubble law galaxies are separating with velocity  $v = H_0 D$

Now, the time until touching =  $D/v = D/H_0 D = 1/H_0$

So the quantity “ $1/H_0$ ” is roughly the age of the Universe.

Note the calculation above is true for any two arbitrary galaxies. It doesn't depend upon the chosen value of  $D$ .

$$\begin{aligned} \text{OK, so } 1/H_0 &= 1 / 2.37 \times 10^{-18} \text{ 1/sec} = 4.21 \times 10^{17} \text{ sec} \\ &= 1.34 \times 10^{10} \text{ years} \\ &= 13.4 \text{ billion years} \end{aligned}$$

The best professional estimate for the age of the Universe is about 13.7 billion years. So our simple estimate is not too bad!

Furthermore, this estimate agrees fairly well with the ages of the oldest known objects in the Universe (the oldest stars in our Galaxy, for example), even though these ages were determined via completely independent methods. If the Big Bang model were wrong, there would be no reason for such good agreement here. So there is a pleasing sense of overall consistency.

## **Some Common Questions and Concerns**

*Is the expansion of the Universe making me expand?*

No. Internal electromagnetic forces that easily overcome the expansion hold you together.

Similarly, the gravitational forces holding the Solar System together easily overcome the expansion, so the Solar System is not expanding.

*Where was the Big Bang, and is there an edge to the Universe?*

The Big Bang was not some “explosion” in an otherwise empty Universe.

Rather, the Big Bang involved the entire Universe, not just the matter and radiation within it. Space itself was created at the time of the Big Bang.

The galaxies are not flying apart into the rest of the Universe. Rather, space itself is expanding, carrying the galaxies along for the ride.

So there is no single location where the Big Bang happened. It involved the whole Universe, and it happened everywhere at once.

# Large-Scale Structures in the Universe and Homogeneity on the Largest Scales

## Groups and Clusters

I'd now like to talk about how galaxies are arranged throughout the Universe on large scales.

### *Groups*

I've told you that we live in a group of galaxies, the Local Group. There are many other groups too, of course.

Groups typically have 3-10 large galaxies in a region  $\sim 1$  Mpc in size.

One might wonder—what is the nearest group to ours?

The answer is probably the IC 342/Maffei group. It is named after the three largest galaxies in the group: IC 342, Maffei 1, and Maffei 2.

There are lots of other known groups too, such as the Hickson Compact Groups.

The galaxies in a group are gravitationally bound to each other, but they do move around and can collide.

An important characteristic timescale is the length of time it takes for a galaxy to cross a group. Galaxies are typically moving within groups at a few hundred km/s.

Size  $\approx 1$  Mpc  $\approx 3 \cdot 10^{24}$  cm  
Velocity  $\approx 300$  km/s  $\approx 3 \cdot 10^7$  cm/s

So crossing time =  $1 \cdot 10^{17}$  sec  
 $\approx 3$  billion years

Now, the Universe is only about 13.7 billion years old. So galaxies have only had time for a few “orbits” in the group.

This means that groups are “dynamically young” and are only now becoming fully “mixed.” There have not been hundreds of orbits and collisions, but only a few.

It is known that many galaxy groups contain a lot of matter in a fundamentally different form—hot X-ray emitting gas.

- Gas has a temperature of a few million degrees.
- Typical density of this gas very low—only 1 particle in the size of a coffee cup.
- Glows in X-rays because it is so hot.
- Can study the chemical composition of this gas. We find that it has a significant amount of metals in it (for example, iron). Not as high a fraction as in our Sun, rather about 1/3 as much.

This is important since it implies that much of the gas in the group is not primordial (in which case it would be made of just hydrogen and helium). It came out of the galaxies, perhaps via supernova explosions that drove large-scale galactic winds.

- In many groups the hot gas has as much mass as in all the galaxies put together.
- There is also “dark matter” in groups, as we’ll discuss later.
- It looks likely that groups contain most of the visible mass in the Universe. They are the “typical” place in the Universe.

### *Clusters*

Now galaxies often come in much larger concentrations too, known as clusters.

For example, the Virgo Cluster doesn’t hold just ~ 10-30 galaxies but more like ~ 2500.

Typical size is about 3 Mpc, and again gravity holds it together.

The Virgo Cluster is at a distance of about 20 Mpc.

If we lived near the center of the Virgo Cluster, there would be  $\approx 100$  large galaxies within 1 Mpc rather than only M31.

The galaxy in the center of the Virgo Cluster is the giant elliptical M87, which is about 400 kpc in size. M87 is  $\approx 20$  times as big and  $\approx 300$  times as massive as the Milky Way Galaxy

About half of clusters have giant ellipticals in the middle, and some of these are significantly larger than M87.

Galaxies are moving in the Virgo Cluster, at  $\approx 1500$  km/s, and again we can calculate a characteristic timescale of crossing.

Size  $\approx 3$  Mpc  $\approx 9 \cdot 10^{24}$  cm  
Velocity  $\approx 1500$  km/s  $\approx 1.5 \cdot 10^8$  cm/s

So crossing time =  $6 \cdot 10^{16}$  sec  
 $\approx 2$  billion years

So again a typical galaxy in a cluster has only made a few “orbits” and clusters are “dynamically young.”

Again, like groups, clusters have hot X-ray emitting gas in them.

- The gas temperature is  $\sim 20$ -100 million kelvin.
- Metal enriched
- Almost always more mass in the X-ray gas than in all the galaxies put together.

With all those galaxies and gas, there is a huge amount of mass. Let's estimate this roughly...

We know the stellar mass of the Milky Way Galaxy  
 $\approx 10^{11}$  stars  $\cdot 1 M_{\odot}$   
 $\approx 10^{11} M_{\odot}$

Now we have  $\sim 2500$  galaxies in Virgo cluster  
So  $2500 \cdot M_{\text{Milky Way}} \approx 2.5 \cdot 10^{14} M_{\odot}$

Also the hot X-ray gives  $\approx 5 \cdot 10^{14} M_{\odot}$

So the visible mass of the Virgo Cluster is

$\sim 7.5 \times 10^{14} M_{\odot}$

$\sim 750$  trillion  $M_{\odot}$

$\sim 1.5 \times 10^{45}$  kg

There is also even more mass in dark matter, as we'll discuss later. This increases the mass by a factor of several more.

With all that mass, the Virgo Cluster gravitationally pulls on galaxies near it. In fact, even our own Local Group of galaxies is being pulled toward the Virgo cluster. Our best estimates are that, in the distant future, our Local Group will be pulled into the Virgo cluster.

In fact, the Virgo Cluster also pulls on other groups near it, and for this reason people often speak of the "Virgo Supercluster" or "Local Supercluster." This supercluster is  $\sim 20$ - $30$  Mpc in size, and we're at the periphery.

There are many other clusters we know about too—thousands in fact.

A few additional examples...

- Coma Cluster at  $\approx 140$  Mpc
- Hercules Cluster at  $\approx 160$  Mpc
- Abell 3528 – a cluster merger

Clusters, with their great masses, can also act as gravitational lenses. Light rays passing near a cluster are deflected via gravity. This process magnifies, brightens, and distorts images of objects that lie beyond the cluster. By modeling the galaxy "arcs" one can determine the amount of mass present in the cluster, including the dark matter.

## Large-Scale Structures

Aside from groups and clusters, are there even larger configurations of matter in the Universe?

Yes. These have been found by making “cartographic” maps that show where galaxies are located in 3D space. The study of this large-scale structure is an extensive field of research in astronomy.

There are other superclusters, like the Virgo Supercluster, that have been found. Some of the relatively nearby ones, within about 200 Mpc, are the Perseus Supercluster, the Centaurus Supercluster, and the Shapley Supercluster.

Astronomers have now mapped the spatial distributions of galaxies, groups, and clusters out to  $\approx 1000$  Mpc  $\approx 1$  Gpc.

To do this well, one has to collect redshifts for 100,000-1,000,000 carefully selected galaxies, and then apply the Hubble Law to get galaxy distances. This is an enormous effort, but modern technology has made it feasible with dedicated surveys. For example, the two largest galaxy surveys to date are the 2dF Survey and the Sloan Digital Sky Survey (SDSS). The SDSS will obtain spectroscopic redshifts for about 1,000,000 galaxies.

On such large scales, astronomers find galaxies to be arranged in a network of “filaments” surrounding large, relatively empty regions of space called “voids.” The largest voids measure  $\approx 100$  Mpc across.

Even surveys like 2dF and the SDSS only reach redshifts of  $z \sim 0.2$ . To probe structures at higher redshifts, one does “pencil beam” redshift surveys in well-studied fields such as the Hubble Deep Field-North. Such surveys require use of some of the largest telescopes on Earth, since the distant galaxies under study are optically faint.

In addition to measuring the properties of these large-scale structures carefully, astronomers also run enormous computer calculations trying to calculate this structure from physical principles. These calculations must include many effects including

- The law of gravity
- The physics of gas motions
- The amount of dark matter present
- The expansion rate of the Universe

- “Feedback” from galaxies that have formed

The largest calculation to date includes about 10 billion particles. It has been run on a 512-processor supercomputer and used about 350,000 processor hours of time.

There is now good quantitative success in explaining the large-scale structure with computer calculations. Such success has only really been achieved since about 1995-2000.

The comparison between observations and computer calculations is made using the “power spectrum.” This quantifies the amount of structure (that is, “lumpiness”) on a given spatial scale.

Imagine placing a series of spheres of a given radius (say, 20 Mpc) at random in the Universe and counting the number of galaxies in each one. Because galaxies are clustered, that number will vary from one sphere to another. The variation in the number of galaxies is a measure of the lumpiness of the galaxy distribution on a scale, in this case, of 20 Mpc. Astronomers repeat this exercise with spheres of various radii to measure the lumpiness at different scales, and the final result is the power spectrum.

The agreement between the observations and computer calculations shows that galaxies and large-scale structures almost certainly formed via the continuous action of gravity over the history of the Universe. Nothing exotic is needed, just gravity.

There is also apparently some significant “feedback” from galaxies that have formed, via supernova-driven winds and perhaps also active-galaxy winds.

## **Homogeneity and Isotropy on the Largest Scales**

When we study the Universe on the largest possible scales (about 5 Gpc and larger), we find good evidence for homogeneity. There do not appear to be even larger scale structures.

This is a good thing, since homogeneity is a key assumption built into models of the Universe that use Einstein’s General Relativity.

For example, we can use radio-emitting galaxies and quasars to trace the largest scales since they can be detected and studied to great distances. The distribution of these on the largest scales is highly isotropic.

We can also study the isotropy of the Cosmic Microwave Background. We find that it is isotropic to better than 1 part in 10,000, as will be explained in more detail later.

# The Birth and Evolution of Galaxies

## Basic Concepts

How do we view the birth of galaxies? How do we watch them evolve?

The age of the Universe is known to be  $13.7 \pm 0.2$  billion years.

Looking out in distance lets us look back in time.

- We see Sun as it was  $\sim 8$  min ago.
- We see nearest stars as they were a few years ago.
- We see Andromeda as it was  $\sim 2.5$  million years ago.
- We see distant galaxies as they were billions of years ago.

So, by looking at very distant objects, we can see back to a time when the Universe was much younger.

One of the most sensitive optical images ever taken is the Hubble Deep Field-North (HDF-N). People have measured redshifts for many of the galaxies in the HDF-N. This has taken a large investment of time on the world's largest telescopes, such as the Keck Telescope.

Using the Hubble Law, you can get the "lookback time" for each distant galaxy in the HDF-N. There is a direct, unambiguous relation between redshift and lookback time.

Newspapers often call this lookback time the distance. But the concept of distance becomes somewhat tricky when considering such vast distances, since the Universe has been expanding significantly since the light was emitted.

- Do we mean the distance when the light was emitted?
- Do we mean the distance now?
- Do we mean some intermediate distance?

It is not as ambiguous, however, to talk about how long it took for light to make its trip.

Since we can see galaxies when the Universe was younger, we should be able to see them being born and evolving with time.

We should be able to see what analogs of our Galaxy and other local galaxies looked like long ago, assuming those distant galaxies ultimately evolved into the galaxies around us in the local universe.

## **Observations of Distant Galaxies**

Now, this is an observationally challenging endeavor. Since more distant galaxies are fainter and subtend smaller angles. It is hard to make out details of their shapes and structures.

This is why people want large telescopes above Earth's atmosphere, so they can make out details of distant galaxies to the greatest extent possible.

The Hubble Space Telescope has been good for this, but in the future we want even more powerful telescopes such as the planned James Webb Space Telescope.

There are other "tricks" that people have also exploited to study distant galaxies. For example, they utilize gravitational lensing by clusters of galaxies.

A cluster can "bend" light from a distant galaxy in the background. This has the effects of distorting, magnifying, and brightening the image of the background galaxy.

It allows us to investigate galaxies that would otherwise be too faint to study effectively. While the shapes of the galaxies are distorted by lensing, their spectra are not affected. So we can use their spectra to measure the chemical elements present, the rate of star formation, the dynamics of galactic gas, and other key physical quantities.

Via intensive work, we have presently managed to find galaxies and quasars out to a redshift of  $z \sim 6.5$ . The corresponding lookback time is 12.8 billion years, so the Universe was only 0.9 billion years old

when the light from these objects was emitted. So we are seeing them as they were when the Universe was only  $\sim 7\%$  of its present age.

There are also a few claims for objects with redshifts of  $z \sim 7$  to 10, but these have not been confirmed with reliable spectroscopic redshifts.

When we study the most distant galaxies, we don't find many large galaxies. Instead, we find many smaller objects that seem to be "pieces" of galaxies. We think these pieces ultimately coalesced to make larger galaxies like our Galaxy.

By measuring the emission-line properties of these "pieces", we infer that these most-distant galaxies are rapidly forming stars.

At intermediate distances, we see what appear to be intermediate objects between the early "pieces" and the fully formed galaxies around us today. We think we are seeing the "pieces" merging together, so it seems clear that this merging process is quite fundamental to the formation of galaxies.

Alongside the observers, theorists are trying to simulate galaxy formation from first physical principles using supercomputer calculations. These simulations are progressing rapidly, and there are now attempts to model galaxy evolution all the way from primeval "pieces" to present-day galaxies.

Our understanding of the cosmic formation history of galaxies and stars has changed dramatically over the past 5-10 years, as shown in plots of the cosmic star-formation rate in galaxies versus time.

We used to think that the cosmic star-formation rate peaked when the Universe was about 3-5 billion years old.

Recent work argues, however, that this rate keeps rising at early times. It seems to peak just about 1 billion years after the Big Bang. Work is ongoing to refine this picture.

One question we have not yet addressed is the origin of the first “pieces” mentioned above. This is mainly because we do not yet have good observational constraints on this matter. However, we do have good theoretical ideas about this, and these ideas will be discussed in detail later.

Finally, the clear detection of evolution of the galaxy population broadly supports the Big-Bang model. The Universe is clearly not in a “steady state” but is changing over time, as expected in the Big-Bang model.

## **Some Insights from Partially Analogous Local Galaxies**

Observations showing the importance of mergers in galaxy formation have given impetus to studies of nearby galaxy mergers. Since these likely hold clues into the details of how distant galaxies formed their stars.

In the nearby “Antennae” system, for example, we see how merging has led to rampant star formation. The merging causes clouds of gas to be shoved together, and these form stars. The stars themselves do not collide, since there is so much space between them.

Other notable examples of merging systems are the “Mice” and the “Tadpole.”

These merging systems also have relevance to our Galaxy. The Milky Way Galaxy and the Andromeda Galaxy will merge in ~ 6-7 billion years. When this event occurs, it will likely make tidal tails and other distortions. Ultimately, the whole Local Group will merge to form a single galaxy that is very large, probably a giant elliptical galaxy.

In some nearby galaxies, such as the famous galaxy M82, we see star formation that has been triggered by a close interaction, even though there was not a merger. We know that M82 interacted with M81 in the past, and that this triggered strong star formation. The star-formation rate of M82 is about 10 solar masses per year, compared to about 1 solar mass per year for our Galaxy. We believe

that similar interactions among high-redshift galaxies also led to lots of star formation.

Furthermore, we observe many “super star clusters” in M82 and other nearby interacting and merging galaxies. This seems to be an important mode of star formation, in addition to the more highly distributed star formation.

We observe a large-scale galactic wind flowing out of M82. This wind is mostly made of hot gas, as revealed in X-ray images of M82. The many supernova explosions occurring within the disk of this galaxy drive this wind. We think that such winds are a key way that the intracluster and intragroup medium is formed in clusters and groups of galaxies. They can largely deplete the interstellar medium of a galaxy, serving to halt its star formation.

Another important lesson is learned by considering the nearby merging system Arp 220. A huge amount of star formation is going on in this galaxy, but the amount is not easily apparent from just optical and near-infrared images of this galaxy. This is because there is a lot of dust in this galaxy (especially in its molecular clouds where stars form), and the dust absorbs most of the starlight. This light is ultimately re-radiated in the far-infrared, making Arp 220 an example of an ultraluminous infrared galaxy. Its far-infrared luminosity is about 100 times its optical luminosity.

Other famous examples of ultraluminous infrared galaxies include Markarian 231 and NGC 6240.

We believe many of the actively star forming galaxies at high redshift are similar to Arp 220, Markarian 231, and NGC 6240; they are highly dust obscured. It has recently become possible to study these high-redshift galaxies using submillimeter radiation. At high redshift, the rest-frame far-infrared emission they make gets redshifted into the submillimeter bandpass. We think the galaxies currently detected as submillimeter sources are the progenitors of massive elliptical galaxies in the local universe.

## The Evolution of Active Galaxies and the Role of Active Galaxies in Galaxy Evolution

Most of the above has focused on relatively normal galaxies, but it is also of interest to ask what we know about the evolution of active galaxies. These are galaxies where a nuclear supermassive black hole is actively accreting material, thereby producing lots of radiation.

The most luminous active galaxies are known as quasars, and quasars are so luminous that they can be studied well over essentially the entire history of the Universe.

We find clear evidence for evolution of the quasar population, at least in terms of their number density. There was a clear “epoch of the quasars” about 2-3 billion years after the Big Bang, when quasars were most numerous. They were much less numerous before and after this time.

While their number density seems to change, the properties of the individual quasar “unit” seem to be largely independent of cosmic time.

The quasars we see represent the growth of supermassive black holes via accretion. These supermassive black holes will not go away even after the quasar dies. So, we should be able to find the relic supermassive black holes today. Just like we can find the bones of the dinosaurs that lived long ago.

Indeed, when we study the nuclei of nearby galaxies, we almost always find supermassive black holes. Furthermore, there is good quantitative agreement between (1) the expected mass density of the black holes from the quasars and (2) the mass density observed in local galaxies.

Remarkably, strong relations have been found between the mass of the supermassive black hole and the properties (mainly mass, luminosity, and velocity dispersion) of the galactic bulge in which it lives.

This is amazing because the bulge is ~ 1000 times more massive and ~ 10 billion times larger than the supermassive black hole. So these two objects are related despite their enormously different scales.

This suggests that the black hole and bulge must have exerted “feedback” upon each other, to maintain the observed relations.

The nature of this feedback is poorly understood at present, but it was probably quite important in galaxy evolution. It is possible that radiation-driven winds from the vicinity of the supermassive black hole were a primary agent of feedback.

# The Cosmic Microwave Background

## The Three Main “Pillars” of the Hot Big Bang Model

The Hubble Law, with its implication of an expanding Universe, is the beginning of observational cosmology. It is often called one of the three main “pillars” of the hot Big Bang model.

Now we will start talking about the other two main “pillars.”

The other two “pillars” are the following:

1. The existence and properties of the cosmic microwave background.
2. Primordial nucleosynthesis and the observed light element abundances.

These two additional “pillars” tell us about conditions early in the history of the Universe, reaching as far back as minutes after the Big Bang occurred.

This lecture will focus on the cosmic microwave background.

## Recombination: The Plasma-to-Gas Transition of our Universe

We have learned that, by looking out into space, we can look back in time.

We can see distant galaxies as they were billions of years ago, and we can see galaxies evolving over the history of the Universe.

But, is there some limit to our ability to look back in time?

Yes, eventually we will start to hit an observational “wall.” This occurs when we look back to times before the galaxies had even formed. At such early times, the Universe was uniformly filled with hot plasma

that would ultimately form the galaxies (via cooling and then gravitational collapse).

When our line-of-sight strikes this hot plasma, we cannot see beyond it. It scatters radiation that tries to pass through it.

The situation is somewhat analogous to looking up in the sky on a cloudy day. One cannot see beyond the clouds, since they scatter the light from beyond them.

The Universe was like this until about a few hundred thousand years after the Big Bang. This corresponds to only about  $1 / 40,000$  of its present age.

To understand this in more detail, let's consider the nature and evolution of this hot plasma further. Since the plasma is filled with free charged particles, electrons and nuclei, it effectively scatters light (free electrons effectively scatter electromagnetic radiation). It is thus opaque.

As the Universe expands, this hot plasma cools. Similarly, expanding plasma or gas on Earth is one way of cooling it. Eventually the plasma cools to where the electrons and nuclei can combine together to form atoms (of hydrogen and helium); this corresponds to the plasma becoming a gas. At this point, there are no longer free charged particles filling the Universe, so light is able to stream freely throughout the Universe without being scattered. The Universe transitions from being opaque to being transparent.

This plasma-to-gas transition of the Universe is called "recombination." The "re" in "recombination" is somewhat inappropriate since this is the first time that the electrons and nuclei had ever combined together, but the name comes from atomic physics studies on Earth.

Recombination occurs when the Universe is about 380,000 years old. At this time, the plasma has a temperature of about 3000 K (somewhat cooler than the surface of the Sun).

The corresponding redshift is about  $z \sim 1100$ . Compare this to the most distant cosmic objects currently known, at  $z \sim 6.5$ .

## **Some Basic Predicted Properties of the Cosmic Microwave Background Radiation**

Hot, opaque objects radiate with a characteristic spectral shape known as a blackbody.

This applies for many familiar objects, such as the hot embers of a campfire, the Earth's surface, and the Sun's surface.

This spectral shape has a sharp drop at short wavelengths, and a more gradual drop at long wavelengths.

Note that the peak intensity of the blackbody shifts to longer wavelengths for cooler objects.

Since the plasma filling the early universe was hot and opaque, we expect the radiation within it to have a blackbody spectral shape as well.

When this radiation was released at recombination, it should have maintained its blackbody spectral shape.

Since the plasma temperature at recombination was about 3000 K, as previously noted, the light should have peaked in the red optical part of the spectrum.

However, as the photons of the cosmic microwave background fly to us from  $z \sim 1100$ , they are greatly redshifted by the expansion of the Universe (in the same way that light from distant galaxies is redshifted).

They get redshifted by a factor of  $\sim 1100$ .

Today the observed temperature of the cosmic microwave background is expected to be  $\sim 3000 / (1+z) \sim 2.7$  K.

This temperature of 2.7 K corresponds to radiation peaking the microwave part of the spectrum.

The microwave band lies between the infrared and radio bands. Microwave ovens emit radiation in this band.

So, if one could detect this cosmic microwave background and show that its spectral form matched the blackbody expectation, this would be a discovery of enormous importance. It would be direct evidence that the Universe had once been in a hot, dense state as predicted by the hot Big Bang model.

We would be “seeing” the Universe when it was only  $1 / 40,000$  of its current age.

## **The Early History of the Cosmic Microwave Background**

The history of the discovery of the cosmic microwave background is amusing and shows the role of serendipity in science.

People in Princeton, New Jersey worked out the basic ideas about the cosmic microwave background in the early 1960's; the people included Robert Dicke and Jim Peebles.

In fact, they were repeating and extending earlier work going back to the 1940's by George Gamow, Ralph Alpher, and Robert Hermann (the Princeton people did not know about this earlier work at the time). Gamow, Alpher, and Hermann predicted the cosmic microwave background as part of their efforts to understand nucleosynthesis in the early universe, as will be discussed in a later lecture.

The Princeton group was just starting a search for this radiation when they heard about some peculiar results from two scientists, Arno Penzias and Robert Wilson, working at Bell Labs in New Jersey.

These two scientists were carefully calibrating a radio telescope to make precise astronomical observations. They were working at a wavelength of 7.35 cm.

They found a “bothersome hiss” in their radio telescope that they could not make go away. It was detected no matter where they pointed their telescope, and it was detected at all times of day and night.

They tried to explain the “bothersome hiss” as being due to

- Atmospheric storms
- Ground interference
- Short circuits in the telescope electronics
- Pigeon droppings in the antenna

None of these explanations worked.

After discussions with theorists at Bell Labs and at Princeton, they realized they were seeing radiation left over from the fiery creation of the Universe itself. They published their results in 1965.

Penzias and Wilson won the 1978 Nobel Prize for their serendipitous discovery.

The fact that the temperature of the cosmic microwave background basically matched expectations from theory was a major point in favor of the hot Big Bang model.

## **Some Notable Asides about the Cosmic Microwave Background**

You can “see” the cosmic microwave background yourself. About 1% of TV static is due to it.

An effect from the cosmic microwave background was observed as early as 1940-1941, although astronomers did not know how to interpret it.

There are gas clouds in our Galaxy that contain many molecules, including the cyanogen molecule. When a star lies behind one of these gas clouds, it is possible to study the molecules present using absorption-line spectroscopy.

For example, Andrew McKellar studied a spectrum of the star Zeta Ophiuchus and noted the R(1) line from rotationally excited cyanogen located in a gas cloud between the star and us.

This line should not have been present, unless there was a radiation field causing the rotational excitation of the cyanogen molecules. Using physics, one can derive that the required temperature of the radiation field is about 2.7 K (McKellar deduced 2.3 K). If this observation had been interpreted properly, the cosmic microwave background could have been discovered much earlier.

Astronomers have now extended these ideas to constrain the temperature of the cosmic microwave background using extragalactic gas clouds (where a distant quasar, rather than a star, provides the background source of illumination). If one can reliably study a gas cloud at high redshift, one should deduce a higher temperature for the cosmic microwave background (since, at higher redshift, the cosmic microwave background will not have “cooled” as much due to the expansion of the Universe).

This provides another fundamental way to test the hot Big Bang model. If the cosmic microwave background were not observed to be hotter at high redshift, this would seriously challenge our interpretation of this radiation.

Getting reliable measurements of high-redshift gas clouds is notoriously difficult. However, astronomers have now found one cloud, at  $z = 2.34$ , that allows the cosmic microwave background temperature to be constrained to be between 6.0 and 14 K. According to the Big Bang model, the cosmic microwave background at  $z = 2.34$  is expected to be 9.1 K, so the observations and theory are in agreement.

The astronomers used measurements of atomic fine structure in neutral and once-ionized carbon atoms to do this.

While this is not a high-precision constraint, it is of great importance since it convincingly shows that the Universe was hotter at earlier times.

## Observations of the Cosmic Microwave Background with the Cosmic Background Explorer (COBE) Satellite

The cosmic microwave background is so important that several satellites have now largely been dedicated to studying it.

The first USA satellite mission was the Cosmic Background Explorer (COBE), which was launched in 1989.

COBE carried instruments to measure the spectrum of the cosmic microwave background precisely as well as to map it spatially.

COBE found the spectrum to be an almost perfect blackbody with a temperature of 2.726 kelvin. This precise blackbody form was a clear success for the hot Big Bang model.

COBE also observed the cosmic microwave background to be highly isotropic, to within about 1 part in 800. This isotropy was also expected according to the hot Big Bang model and its “inflationary” extension (to be discussed in a later lecture).

The isotropy of the cosmic microwave background serves to vindicate the *cosmological principle* introduced by Einstein. It shows that, on the largest scales, the Universe is isotropic.

At more subtle levels, deviations from isotropy (“anisotropies”) were observed due to (1) the Earth’s motion through the Universe and (2) foreground emission from our Galaxy.

After subtracting off these “uninteresting” sources of anisotropy, the COBE team was able to detect *intrinsic* anisotropies in their maps of the cosmic microwave background. These were seen at levels of about 1 part in 25,000. The temperature of the cosmic microwave background was slightly higher or lower in some directions than in others.

This finding caused great excitement and was reported widely in the media, including on the front page of the *New York Times*. The

excitement arose because these small variations in temperature are associated with small variations in the density of the plasma that filled the early universe.

Over time, the small overdensities in the plasma grew under the action of gravity to form large-scale structures, clusters of galaxies, and galaxies.

So COBE was seeing the “seeds” of galaxies and other structures in the Universe, imprinted on the cosmic microwave background.

Where did these “seeds” originally come from? This is still a question at the forefront of modern cosmology, and we are not fully sure of the answer. Our best “educated guess” at present is that these “seeds” are the result of microscopic quantum-mechanical fluctuations that grew to macroscopic scales during an “inflationary” epoch the Universe experienced during its first small fraction of a second. This will be discussed in more detail in a later lecture.

COBE did have some significant limitations in its ability to map the cosmic microwave background. Its angular resolution was poor, about 7 degrees, and the signal-to-noise ratio of its maps was a limiting factor in detailed analyses.

But, since COBE was so successful, people wanted to build more satellites to study the cosmic microwave background.

## **Current Observations of the Cosmic Microwave Background**

Studies of the cosmic microwave background have become a major part of modern cosmology.

Not only do they allow one to see the “seeds” of galaxies and other structures in the Universe, but also they now provide precision constraints on cosmological parameters such as the Hubble constant and the density of the Universe.

Observations are being carried out from the ground, from long-duration balloons, and from space.

The current space mission is the Wilkinson Microwave Anisotropy Probe (WMAP), launched in 2001. It is located out past the Moon at the L2 Lagrangian point. As its name suggests, its primary purpose is to make high-quality maps of the anisotropy of the cosmic microwave background. It is still collecting data today.

Its angular resolution is about 0.3 degrees, much better than that of COBE, so it can see finer details of the anisotropy. It also produces maps with higher levels of signal-to-noise.

Similar to what is done in studies of large-scale structure, people work to measure the amount of variation of the cosmic microwave background on different angular scales.

Imagine placing a series of circles of a given angular radius (say, 1 degree) at random on a map of the cosmic microwave background and measuring the amount of flux in each one. Because the cosmic microwave background has anisotropies, the flux will vary from one circle to another. The variation in the amount of flux is a measure of the lumpiness of the cosmic microwave background on an angular scale, in this case, of 1 degree. Astronomers repeat this exercise with circles of various angular radii to measure the lumpiness at different scales, and the final result is the power spectrum.

They then compare their observations of the power spectrum with theoretical models of it to deduce information about the Universe.

The details of how this observation-theory comparison works are complicated, and it would take 1-2 lectures to explain them. So here I will only give the most basic ideas. To learn more, see “The Cosmic Symphony” in the February 2004 issue of *Scientific American*.

The “wiggles” in the power spectrum are known as “acoustic peaks.” These were created by sound waves propagating in the plasma that filled the early universe. The acoustic peaks show

that the slight hot and cold spots imprinted on the cosmic microwave background have characteristic sizes.

Scientists have gotten quite skilled at calculating the details of how sound waves and plasmas work. The physics of these are actually much simpler than the physics of galaxies, stars, and planets.

The basic idea behind the sound waves is that you have a “battle” between gravity and radiation pressure in the plasma.

If a region of plasma is cool and slightly overdense, gravity tends to pull it together. As the region contracts, the plasma heats. The additional pressure produced by the heating pushes outward, halting the contraction and reversing it. The resulting expansion cools the gas, reducing the pressure and allowing the cycle to repeat. This is basically a sound wave, albeit in an unfamiliar form.

The centroids, strengths, and widths of these peaks are set by the properties of the plasma through which the sound waves propagate. And the properties of this plasma depend upon cosmological parameters such as the overall matter density and the rate of cosmic expansion (i.e., the Hubble constant). So one can deduce cosmological parameters by the properties of these peaks.

The whole process is somewhat analogous to determining the construction of a musical instrument by carefully listening to its notes.

They can deduce a tremendous amount of precise cosmological information including the Hubble constant, the matter density, the baryonic matter density, the age of the Universe, the redshift of recombination, etc.

Largely on account of precise measurements of the power spectrum of the cosmic microwave background radiation, astronomers often say that we are now in an age of “precision cosmology.” Cosmology

has gone from being a data-starved enterprise to one where large amounts of precision data are available.

One exciting finding is that WMAP supports the need for “dark energy” in Universe. In the present day, more than 70% of the mass-energy of the Universe is in the form of a dark energy that has yet to be observed directly. This dark energy is currently causing the Hubble expansion to accelerate, as will be discussed further in a later lecture.

Another exciting finding is that WMAP supports some of the key predictions of the “inflationary” extension of the hot Big Bang model (to be discussed in a later lecture).

WMAP, and other observations of the cosmic microwave background, have also shown that it is slightly polarized. The detection of polarization has been very challenging, since the polarization signal is ~ 100 times weaker than the (already weak) anisotropy signal.

The polarization is expected as a secondary consequence of the anisotropy. Anisotropic photon bombardment sets the plasma’s electrons in motion, thus imposing a small net linear polarization on the photons they scatter.

The polarization is important since it provides further consistency checks on our overall interpretation of the cosmic microwave background. Its detection also allows for even more precise measurements of some cosmological parameters.

## **Future Observations of the Cosmic Microwave Background**

There is so much potential in space-based studies of the cosmic microwave background that the European Space Agency soon (hopefully in 2007) plans to launch another microwave background mission called Planck. This should allow even better studies of the anisotropies and polarization of the cosmic microwave background.

Another mission after Planck, often called CMBPOL, is also actively being discussed. It would significantly advance measurements of the polarization of the cosmic microwave background.

# The First 20 Minutes of Our Universe

## Introduction

The cosmic microwave background allows us to study the state of our Universe about 380,000 years after the Big Bang. As we have discussed, this represents the limit of our ability to “see” back in time using photons.

However, we can reliably probe conditions in the early universe back to about 30 seconds after the Big Bang. We will now discuss how this is done.

We will also discuss some less-certain ideas about our Universe at even earlier times, going back to about  $10^{-30}$  seconds after the Big Bang.

Throughout this lecture, we will be considering the Universe well before the formation of galaxies. The Universe at these times was filled with a “primordial soup” of elementary particles (a plasma).

## Primordial Nucleosynthesis

As has been mentioned before, primordial nucleosynthesis is one of the three main “pillars” of the hot Big Bang model.

This section could also be titled “cooking the light elements in the early universe.”

Most of the “heavy” elements in the Universe (such as C, O, Ne, Mg, Si, S, Fe) were made in the cores of stars via fusion burning.

However, the “light” elements (such as helium, deuterium, and lithium) could not be explained this way.

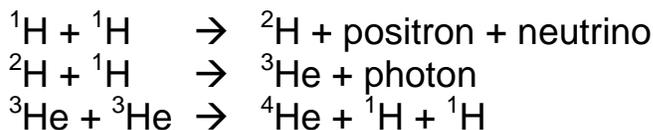
The problem is most striking for helium. There is simply too much helium in the Universe to be explained by nuclear fusion in stars. We always find  $> 25\%$  helium by mass in stars.

Our Sun creates  $5.8 \times 10^8$  tons of helium each second, which may sound like a lot. However, the Sun currently contains  $5 \times 10^{26}$  tons of helium. Given the rate of solar helium production, it would have taken about 29 billion years to make the amount of helium present. But this span of time is longer than both the age of the Universe (13.7 billion years) and the age of the Sun (4.6 billion years).

In the late 1940's, George Gamow and Ralph Alpher started to think about this problem in detail. They realized that large amounts of helium, and other light elements, could have been made by fusion reactions during the first ~ 20 minutes after the Big Bang (mainly during the first ~ 3 minutes). This key idea is known as primordial nucleosynthesis.

During the minutes after the Big Bang, the conditions in the plasma were hotter than in the core of a typical star. The plasma temperature was 900-300 million kelvin. However, like in the core of a star, nuclear fusion reactions could take place.

In the core of our Sun, hydrogen is converted into helium via the proton-proton chain:



The reactions that took place in the early universe were somewhat different, because the temperature and density conditions in the plasma were different.

The reactions were, in fact, quite complex. But, on account of intensive efforts to build fusion bombs and so on, physicists are very skilled at calculating nuclear reactions.

The relevant fusion reactions started about 30 seconds after the Big Bang. At earlier times the reactions could not occur, since deuterium kept being broken apart by the high-energy gamma rays throughout the hot, young Universe. This is known as the "deuterium bottleneck."

After ~ 20 minutes (and largely after ~ 3 minutes), the temperature and density of the plasma had dropped to levels where fusion could not proceed any further (on account of the expansion of the Universe). As a result of the short time available for fusion reactions (30 seconds to 20 minutes), only “light” elements were created.

Computer calculations of the relevant fusion physics predict that about 25% of the Universe (by mass) should have turned to He. About 75% of Universe (by mass) should have remained as H.

Now, of course, subsequent fusion in stars has made more He, but none should have been destroyed.

So we can test the Big Bang model by looking at He abundances in stars, gas clouds, and galaxies. One should always find > 25% by mass, or the Big Bang is wrong.

The Big Bang has survived this test beautifully. Our Sun, for example, is ~ 28% He by mass.

Amazingly, the overall composition of the Sun and the stars tells us about conditions just a few minutes after the Big Bang!

Primordial nucleosynthesis also predicts the amounts of deuterium and lithium. There is excellent quantitative agreement between the predictions and observations.

## **A “Soup” of Quarks and Gluons in the Early Universe**

We will now turn our attention to even earlier times in the history of the Universe. We will consider what likely happened during the first small fraction of a second after the Big Bang.

The conditions during this time were exceptional. Here we will be considering temperatures up to about  $10^{28}$  kelvin, and at earlier times the temperatures were even hotter.

Our ability to probe such times observationally is quite limited at present. As a result, our conclusions about this epoch are much less certain.

Furthermore, we are not confident of all the physics processes operating during this epoch. So we are now approaching the edge of current human knowledge, where it is possible to ask many questions that do not have clear answers at present.

At very early times, we think our Universe was filled with a “soup” of quarks, antiquarks, gluons, electrons, positrons, photons, etc.

Gluons are neutral, massless particles that carry the strong nuclear force, keeping quarks bound within protons and neutrons. There are eight of them.

Before about 10 microseconds had elapsed (corresponding to temperatures of about  $10^{12}$  kelvin or higher), the quarks had not yet combined to create protons and neutrons.

This state of matter is known as a “quark-gluon plasma” or “quark-gluon liquid.”

We are now, for the first time, starting to make this state of matter on Earth. This is mainly being done at Brookhaven National Lab located on Long Island in New York.

There, at the Relativistic Heavy Ion Collider (RHIC), physicists crash together two beams of heavy nuclei (often gold), each moving at more than 99.99% the speed of light.

The collisions cause the nuclei and their constituent protons and neutrons to melt, and a spray of quarks, antiquarks, and gluons are created.

The quarks, antiquarks, and gluons quickly recombine into protons, neutrons, and other hadrons. By studying the reaction products, physicists try to deduce properties of the soup of quarks, antiquarks, and gluons.

Some surprises have been found over the past few years. For example, the created soup is showing some liquid-like properties, rather than behaving as a perfect gas (as previously expected).

For details on this, see the article “The First Few Microseconds” in the May 2006 issue of *Scientific American*.

Real progress in our understanding of the early universe is coming from particle accelerators on Earth.

## **Baryogenesis: The Dominance of Matter over Antimatter**

From a human perspective, perhaps the most important thing that happened in the very early universe is baryogenesis.

The key idea here is that, during the first  $\sim 10^{-30}$  seconds of our Universe, matter somehow “won” over antimatter.

Antimatter is well established both theoretically and observationally. The positron was first predicted by Paul Dirac in 1928 and was discovered by Carl Anderson in 1932.

Antiquarks are the antimatter versions of quarks. They make up antiprotons, for example.

Baryons are particles consisting of three quarks, such as the proton and the neutron.

In particle reactions today in the Universe, the number of baryons is conserved (where antiparticles of baryons count negatively). For example, protons and antiprotons are always created in pairs, to conserve the number of baryons.

Since baryons are made of quarks, this also implies conservation of the number of quarks in a particle reaction (where antiquarks count negatively).

If this law has always held throughout all of cosmic history, then there should be equal amounts of matter and antimatter in the Universe today. In this case, the Universe would be a dangerous place due to matter-antimatter annihilations on large scales.

We do not observe large amounts of antimatter in the Universe today, so we believe that the law of baryon conservation must have slightly failed sometime in the early Universe.

This failing would have been very slight indeed, by only about 1 part in 10 billion.

So, for each excess quark of matter, about 10 billion other quarks would have annihilated with 10 billion antiquarks. All the quark-antiquark annihilations that occurred made the photons that we now see as the cosmic microwave background radiation.

The failing could have happened at the very instant of the Big Bang, but many scientists hope it occurred somewhat later, at about  $10^{-30}$  seconds after the Big Bang.

Our best theories of particle physics predict slight violations of the law of baryon conservation at sufficiently high particle energies. However, these theories are still uncertain and predict some phenomena that have not yet been observed (such as the long-timescale decay of the proton).

To date, we have not observed any particle interactions that violate the law of baryon conservation. However, our particle accelerators on Earth may simply not be probing sufficient energies.

Observations have revealed slight failings of the symmetries of time reversal, charge conjugation, and parity.

Andrei Sakharov, an eminent Soviet nuclear physicist, put forward three key conditions that must be satisfied for baryogenesis. Slightly restated, these are the following:

- Particle interactions must violate baryon number.
- Particle interactions must violate matter-antimatter symmetry.

- These interactions must be occurring out of thermal equilibrium.

Again, it is clear that better observations at particle accelerators on Earth will be important for understanding the early universe.

# Dark Matter, Dark Energy, and the Future of Our Universe

## Introduction

A startling and humbling fact is that astronomers do not know what composes more than 90% of the Universe.

We know that there is a lot of mass out there that we cannot see at *any* wavelength. This is called “dark matter.”

We deduce the need for dark matter using a method pervasive throughout astronomy. We “detect” the presence of something we cannot see based upon its gravitational influence upon things we can see.

This basic method is used to study

- Planets around Sun-like stars
- Black holes in binary star systems
- The supermassive black hole in the Galactic center
- Supermassive black holes in nearby galaxies

People, at least a few of them, started to become aware of dark matter as early as the 1930's.

It became a topic of intensive study in the 1960's and 1970's.

By ~ 1975 most astronomers took the existence of dark matter very seriously.

We also know that the expansion of the Universe seems to be accelerating. This is attributed to a “dark energy” throughout the Universe.

Dark energy has a much shorter history. It was only convincingly discovered in about 1998.

Today the study of dark matter and dark energy is one of the forefront topics in astronomy.

I will not follow a strictly historical approach today. Rather, I will follow a scheme broadly based upon distance from us:

- Dark matter in our Galaxy
- Dark matter in other galaxies
- Dark matter in clusters of galaxies
- Dark energy
- Dark matter, dark energy, and the future of our Universe

## **Evidence for Dark Matter in Our Galaxy**

Astronomers have intensively studied the structure and dynamics of our Galaxy.

In particular, I want to review some of the things astronomers have learned about the way our Galaxy rotates since this is a key piece of evidence for dark matter.

Our Galaxy is rotating about its center.

For example, the Sun orbits around the center of the Galaxy with a velocity of about 220 km / s.

We are about 8.5 kpc = 8500 pc =  $2.6 \times 10^{17}$  km from the center.

So, to make one orbit takes

$$P = 2\pi r / v = 2\pi (2.6 \times 10^{17}) / 220 = 7.4 \times 10^{15} \text{ sec} \\ = 250 \text{ million yrs}$$

This is a long time, so clearly we cannot watch the rotation occurring on human timescales.

The main way we quantify the rotation of our Galaxy is with a “rotation curve.” This shows the rotational speed of our Galaxy as a function of distance from its center.

We can plot rotation curves for many different objects, including merry-go-rounds, our Solar System, and our Galaxy.

Now, gravity is the key force that ultimately dictates how galaxies rotate, and the amount of gravity depends upon the amount of mass present. So, by studying the details of how galaxies rotate, we can deduce the amount of mass present in them.

For our Solar System, the rotation curve drops off quickly with distance from the Sun. This is because most of the mass of our Solar System is concentrated in the Sun. So, as you move away from the Sun, the strength of the gravitational field drops off quickly.

This basic idea is true more generally. For any cosmic object, once you get outside the radius where most of the mass lies, you expect the rotation curve to drop off quickly with distance.

Remarkably, however, this is not observed for our Galaxy if one believes the only mass present is that we can see.

Most of the visible mass of our Galaxy, including stars and gas clouds, lies within  $\sim 15$  kpc of its center.

However, when we study the stars in our Galaxy, we observe a “flat rotation curve” (at about  $220 \text{ km / s}$ ) going all the way out to  $40$  kpc or more. If anything, the rotation curve rises slightly from  $15$ - $40$  kpc.

This is taken as evidence for a large amount of matter that we cannot see directly. This matter is providing the additional gravity needed to keep the rotational velocities high at large radii.

We can measure velocities out to  $\sim 100$  kpc or more if we also use radio measurements of atomic hydrogen clouds. These emit a spectral line at  $21 \text{ cm}$  (in the radio band) and are found very far out from the center of our Galaxy.

At distances of  $\sim 100$  kpc we still measure velocities of  $\sim 200 \text{ km / s}$ . Using Newton’s law of gravity, we can deduce that the mass within  $100$  kpc is  $\sim 900$  billion solar masses.

Now, the mass interior to the Sun's orbit is only ~ 100 billion solar masses. So there is much more mass outside the Sun's orbit as within it.

Yet, there is relatively little light out at these large distances, hence the name "dark matter."

This dramatically changes our view of what our Galaxy is like. We only can see the "tip of the iceberg." A massive "halo" of dark matter dominates the mass of our Galaxy.

The Large and Small Magellanic Clouds, at distances of about 50 kpc, are moving inside our dark halo!

The dark-matter halo probably extends to even larger radii, and the best estimates today give a mass for our Galaxy of  $\sim 3 \times 10^{12}$  solar masses ~ 3000 billion solar masses.

## **What is the Dark Matter in Our Galaxy?**

Since 90% or more of our Galaxy is dark matter, it is critical to determine the nature of this dark matter.

One can think of lots of possibilities:

- Faint, red dwarf stars (fusion burning)
- Brown dwarfs with  $< 0.09$  solar masses (not fusion burning)
- Dead stars (white dwarfs, neutron stars, or black holes)
- Gas clouds, either cold or ionized

These things above are "normal" matter made out of atoms with protons and neutrons – called *baryonic*.

Baryons are particles consisting of three quarks, such as the proton and the neutron.

There might also be *nonbaryonic* dark matter made out of "stuff" other than "normal" matter:

- Massive neutrinos
- Weakly Interacting Massive Particles (WIMPS), such as the axion
- Cosmic strings

Many of these nonbaryonic dark matter candidates have never even been detected on Earth or in the cosmos. They are predicted by some theories of particle physics.

People have made substantial progress over the past ~ 10 years constraining these possibilities.

For example, very deep images made with the Hubble Space Telescope essentially rule out faint, red-dwarf stars and show that they make < 1% of the Galactic halo.

Brown dwarfs and dead stars are harder to constrain, since they need not emit much light at all. But their presence can be constrained with *gravitational microlensing*.

If there are billions of brown dwarfs or dead stars out in the halo, occasionally one will drift along our line of sight to a background star.

When this happens, the object's gravity will focus the background star's light almost directly at Earth. As a result, the star will appear to brighten for a period of days to weeks. Even a planet-mass object can act as an effective gravitational lens.

Microlensing events are expected to be rare. Only expect a few per million stars per year.

However, with dedicated telescopes and powerful computers, people can monitor millions of stars in the Large and Small Magellanic Clouds (and elsewhere).

One can discriminate between microlensing events and other types of stellar variability, since gravitational microlensing should cause no color changes during the event.

Many dozens of convincing microlensing events have now been discovered; the first was found in 1992.

Binary microlensing events have even been discovered.

The number of events per year tells one about the number of brown dwarfs or dead stars, and the event durations loosely constrain lens masses.

The lensing objects appear to have a range of masses, although the masses are usually not well constrained. Some have been argued to be black holes, while others have been argued to be brown dwarfs or free-floating planets.

While there are some massive, compact objects in the halo, they can only explain up to about 20% of the halo mass.

The other ~ 80% of the halo must be something else. There is increasing evidence that most of the dark matter in the Galactic halo is nonbaryonic, since we cannot find sensible baryonic candidates.

The need for nonbaryonic dark matter is also indicated by (1) the power spectrum of the cosmic microwave background as measured by WMAP, and (2) the details of primordial nucleosynthesis. These indicate there is about five times as much nonbaryonic dark matter as “normal” matter (both dark and visible).

Unfortunately, however, we have poor constraints on the nature of the nonbaryonic dark matter at present.

Neutrinos are the one type of nonbaryonic dark matter that we have actually detected.

We have learned recently that neutrinos have small, but nonzero, masses. While each neutrino only has about one ten millionth the mass of an electron, they are very common. The total cosmic mass in neutrinos is comparable to that in all of the stars.

While the neutrino results are interesting, neutrinos appear to make up only a small fraction of the dark matter. Furthermore, neutrinos do

not have the right properties to help cosmic structures form gravitationally in the manner observed. The majority of the nonbaryonic dark matter has yet to be found.

There may well not be a single solution to the dark-matter problem. That is, there may be many types of nonbaryonic dark-matter particles awaiting discovery.

As a result, intensive efforts are now being made to search for nonbaryonic dark-matter particles. Such particles should be all around us, even in this room.

Scientists are building dark-matter detectors that aim to detect effects when a dark-matter particle interacts with normal matter. Such interactions may lead to small, but detectable, flashes of light or rises in temperature.

Since dark-matter particles interact only weakly with normal matter, interactions between dark-matter particles and normal matter are expected to be rare. It is thus critical to screen out unrelated background events.

For this reason, many dark-matter searches are located far beneath the Earth to provide shielding against cosmic rays. Cosmic rays are a major source of background for such searches.

Scientists also hope that increasingly powerful particle accelerators on Earth will be able to create dark-matter particles for study.

The next large particle accelerator activated will be the Large Hadron Collider at CERN in Europe. There is no guarantee that it will detect any new particles, but scientists' hopes are high.

## **Dark Matter in Other Galaxies**

As for our Galaxy, astronomers have measured rotation curves for other spiral galaxies and find them to be flat. They measure stellar motions (with optical spectra) or gas-cloud motions (using the 21 cm spectral line).

Dark matter is inferred to be an ubiquitous phenomenon in spiral galaxies. As for our Galaxy, scientists generally infer that  $> 90\%$  of the mass is in a dark form.

In elliptical galaxies it is somewhat harder to study dark matter. They contain little cold atomic gas so cannot use the 21 cm line. They also do not show organized rotation as for spirals.

However, dark matter is believed to be ubiquitous in elliptical galaxies as well. For example, many elliptical galaxies contain hot X-ray emitting gas. To keep this gas from “boiling” off into intergalactic space, there needs to be much more mass than is visible (to provide sufficient gravity to keep the gas bound).

## **Dark Matter in Clusters of Galaxies**

We will now start to consider dark matter on larger scales. Clusters of galaxies, for example, allow us to probe dark matter on scales of a few Mpc.

Clusters of galaxies have large amounts of visible mass in the constituent galaxies and also in the hot X-ray emitting gas located between the galaxies. Typically the visible mass is dominated by the hot X-ray emitting gas.

We deduce the need for large amounts of dark matter in galaxy clusters, and clusters are especially good environments for dark-matter studies.

In fact, the existence of dark matter was first inferred in the 1930's via studies of galaxy clusters. This was pioneering work done by the brilliant but unorthodox astronomer Fritz Zwicky.

Dark matter in galaxy clusters can be studied with three independent methods.

The first method involves the motions of the galaxies within the cluster; these can be measured using the Doppler shift. The cluster must have sufficient mass to keep the galaxies within it “bound” on

their orbits. If there were not sufficient mass, the cluster would quickly fly apart.

Using this argument combined with Newton's law of gravity, we deduce that the mass of a cluster is 70-90% due to its dark matter. This was first noted by Zwicky in the 1930's.

The second method involves the hot X-ray gas within the cluster. The cluster must have sufficient mass to keep this gas from "boiling" away (just like for elliptical galaxies).

The third method involves gravitational lensing of background galaxies by the cluster. The gravity of a cluster can bend the light from galaxies behind it to form multiple images. The amount of light bending depends upon the mass of the object doing the bending, and this allows one to deduce the mass of the cluster.

The masses deduced from these three different approaches are in respectable agreement, and this cross-checking gives us confidence that a large amount of dark matter is indeed present in galaxy clusters.

The typical mass composition of a galaxy cluster is 70-90% dark matter, 10-25% hot X-ray gas, and < 10% galaxies.

As is the case for individual galaxies, astronomers think that most of the dark matter in galaxy clusters is nonbaryonic.

## **Dark Energy**

Type Ia supernovae are thought to occur when a white dwarf star in a close binary system accretes so much matter from its companion that it collapses. This collapse triggers explosive nuclear fusion reactions throughout the white dwarf, causing it to explode.

In the 1990's, technology allowed a great increase in the number of supernovae with well-observed light curves. For Type Ia supernovae, it was possible to find and calibrate a relation between the peak luminosity of the supernova and the time it takes for the supernova to fade.

This allowed astronomers to use Type Ia supernovae to deduce cosmic distances (they had the peak luminosity and the peak flux, so they could solve for the distance). Since supernovae are so luminous, they can be seen to great distances (currently out to about  $z \sim 1.5$ ).

High-redshift Type Ia supernovae were found to be systematically fainter than expected, based on extrapolations from low-redshift supernovae and Hubble's Law. The implication was that these supernovae are further away than expected, if the Universe has been expanding at a constant rate. That is, the expansion of the Universe has been accelerating.

To explain this acceleration, some effect must be overcoming the tendency of gravity to slow down the expansion. This effect is associated with "dark energy" filling the Universe. The dark energy provides a negative pressure that drives the cosmic acceleration.

If the existence of dark energy depended only upon the Type Ia supernova results, then its existence would not be entirely secure. For example, it is possible that Type Ia supernovae seen billions of years ago could just be intrinsically less powerful than those today in the local universe. This would be physically surprising given our understanding of them, but perhaps it is not impossible.

However, the existence of dark energy is also now supported by precise measurements of the cosmic microwave background, clusters of galaxies, and large-scale structure.

When astronomers calculate the amount of mass-energy in the Universe associated with dark energy, they find it to be very large. About 70% of the mass-energy of the Universe is associated with it. So dark energy dominates the mass-energy budget of the Universe!

We understand the nature of dark energy even less than the nature of dark matter.

One possibility is that the dark energy is just Einstein's cosmological constant from 1917. This has the correct basic properties to cause the acceleration, although our current measurements are not refined

enough for a precision test. If the dark energy is the cosmological constant, then it would be constant throughout space and over time. Of course, such an explanation would not be entirely satisfying, since one hopes that some physical “agent” can be discovered locally and associated with the dark energy.

Other possibilities are that the dark energy is due to (1) quantum-mechanical vacuum energy associated with virtual particles, or (2) an exotic form of matter filling the Universe called “quintessence.” In the second case, the dark energy could vary over cosmic time.

Astronomers are now trying to make precise measurements of the acceleration history of the Universe over cosmic time. By doing this, they can discriminate between different models for the dark energy.

## **Dark Matter, Dark Energy, and the Future of Our Universe**

The Universe is currently expanding, with the galaxies receding from each other. Will this expansion continue forever?

The relevant factors to consider are the following:

- The amount of normal and dark matter. This provides gravity that tries to slow and reverse the expansion.
- The dark energy that is currently driving cosmic acceleration.

If the accelerated expansion continues indefinitely (as will be the case if dark energy is the cosmological constant), then the Universe will become increasingly diluted and cold in the distant future. All the stars will eventually burn out, and the Universe will “die by ice.” This is probably the preferred scenario right now, but we are not certain it is correct.

Since we do not understand the dark energy reliably, its nature could change in the future (as could happen if dark energy is quintessence). It could, in principle, change to drive a cosmic compression instead of a cosmic acceleration. Then the Universe would “die by fire” in a Big Crunch.

We cannot predict the fate of the Universe reliably until we understand the nature of the dark energy.

# The First Stars, Galaxies, and Quasars in Our Universe

## Introduction and the Cosmic Dark Ages

In the past lectures, we have covered most of cosmic history. We have seen that the very early Universe was filled with a highly uniform plasma, while the Universe today is clumpy and filled with galaxies.

In this lecture, we will discuss how the Universe went from being (1) uniform plasma to (2) clumpy with galaxies. Much of the action here happened during the first billion years of the Universe's history.

After the release of the cosmic background radiation, when the Universe went from being a plasma to being a gas, the Universe became a fairly dark place. The cosmic background radiation photons were still there, but they were shifted into the infrared due to the expansion of the Universe.

There were not yet any sources of visible light such as stars, since they had not yet formed from the gas.

The gas was almost entirely just hydrogen and helium that was made via primordial nucleosynthesis. Since stars had not yet formed, there were no gravitationally bound fusion reactors to make heavier elements.

This epoch without visible light lasted a few hundred million years and is often called the "cosmic dark ages".

## The Formation of the First Stars

Eventually, under the continual action of gravity, the dark matter and gas started to clump together to make protogalaxies.

The dark matter and gas clumped most in the locations where there had been slight overdensities in the plasma that filled the very early Universe. Since we can see the effects of these slight overdensities

imprinted on the cosmic microwave background today, we are indeed able to see the seeds of galaxies and other cosmic structures.

Since the mass of the dark matter substantially exceeds that of the normal matter, the dark matter dominated the gravitational effects. The dark matter clumped, and the gas got dragged along as a secondary effect.

For the gas to condense, it needed to cool. The gas was able to cool by forming hydrogen molecules; these can emit energy as they rotate and vibrate.

When the gas had condensed to a sufficient level, star formation was able to proceed in the densest regions formed. We now think this happened at redshifts of about 15. The protogalaxies formed at this time were only about 1 millionth the mass of our Galaxy.

Star formation is a very complex process. Even in the local universe, such as in the Orion Nebula, it is challenging to understand in detail. There are hundreds of astronomers working on this problem today. We still do not understand fully, for example, what controls the rate of star formation or how stars get the range of masses they have.

We cannot directly observe the first stars, at least at present. A single star at such great distances is too faint to be detectable with present observational facilities.

Thus, understanding the formation of the first stars is presently a theoretical endeavor. Theorists run enormous computer codes on powerful supercomputers to calculate how the formation proceeded. These codes include the relevant physical processes such as gravity, gas cooling, gas pressure, gas chemistry, transport of radiation, and the expansion of the Universe.

While such calculations might seem precarious without observations to guide them, there are a few helpful simplifications relevant to the formation of the first stars. Heavy elements did not exist, for example, and this makes the atomic-physics calculations much simpler. Furthermore, magnetic fields were probably not important, in contrast to the situation for local star formation.

The current calculations indicate the first stars would have been massive, about 100-1000 times the mass of our Sun. For comparison, the most massive stars we know about in the local universe are about 100 solar masses.

The high masses of the first stars are largely due to the different elements from which they form. With only hydrogen and helium present, there are no heavy metals that are more opaque to radiation. Thus, gas infall onto the first stars is not stopped as effectively when they start to produce radiation, and they can obtain higher masses.

The first stars would have had radii of 4-14 solar radii and would have emitted 1-30 million solar luminosities, mainly in the ultraviolet (due to their high temperatures).

Such massive and luminous stars have large effects on their environments. They ionize the gas around them and also “push” on this gas via radiation pressure. The presence of a single massive star in a small protogalaxy probably prevents any other stars from forming in that protogalaxy.

## **The Deaths of the First Stars**

In the local universe, we know that massive stars live very short lives before they explode as supernovae, and this is expected for the first stars as well. They probably only live for about 2 million years; this is  $1 / 5000$  the lifetime of the Sun.

When one of the first stars explodes as a supernova, it releases enough energy to expel all the gas from the protogalaxy in which it resides.

The expelled gas is, however, now enriched in heavy elements including carbon, oxygen, neon, magnesium, silicon, sulfur, and iron. This is important, since when this gas re-condenses, the second generation of stars formed will be more similar to those in our local universe.

The deaths of the first stars probably also left stellar-mass black holes behind.

The fact that the first stars probably explode as supernovae may make them detectable in the future, since supernovae are very luminous. There are hopes that the James Webb Space Telescope or future large, ground-based telescopes may be able to detect these supernova events.

In addition, we now know that some fraction of supernova events have associated gamma-ray bursts. These are probably associated with jets formed by a nascent black hole. Gamma-ray bursts are also very luminous, and perhaps missions such as the Swift Gamma-Ray Burst Explorer will be able to detect gamma-ray bursts from the deaths of the first stars.

Gamma-ray bursts have already been found up to redshifts of 6.29, which is quite impressive. The most distant galaxies and quasars currently known only have redshifts slightly higher than this.

## **The First Supermassive Black Holes**

The metal-enriched gas created by the first stars eventually re-condensed to form galaxies with less-massive stars, and mergers of these galaxies were presumably common in the dense early universe.

In the densest parts of the early universe, the stellar-mass black holes left behind by the first stars were able to feed and grow efficiently. Given what we know about black-hole accretion, they probably were able to double their mass about every 50 million years.

Thus, via a series of about 25 mass doublings over about a billion years, a 100 solar-mass black hole was able to grow into a billion solar-mass black hole. This is the type of mass needed to power the most-distant quasars via accretion.

The basic numbers seem to fit together consistently, in that black holes of sufficient mass were able to grow rapidly enough to match what we observe in the distant universe.

Astronomers are now testing these ideas by measuring the properties of the most-distant quasars across the electromagnetic spectrum. The goal is to determine if these quasars are feeding and growing in the same way as nearer quasars.

The “bottom line” emerging from this work is that the observed properties of the first quasars are remarkably consistent with those in the local universe. The quasar core is insensitive to the fact that the Universe was a radically different place at  $z \sim 6$  compared to today.

# Inflation and the Possibility of a Multiverse

## Introduction

In this lecture, we will be discussing the earliest events after the Big Bang. Specifically, we will consider times about  $10^{-36}$  seconds after the Big Bang.

The times we will be considering are before baryogenesis (at perhaps  $10^{-30}$  seconds) and primordial nucleosynthesis.

The Universe at this time was extremely dense and hot. The temperature was about  $10^{28}$  kelvin.

We have never even approached such extreme conditions in laboratory experiments on Earth.

Given the extreme nature of the Universe at these early times, we do not understand what happened well. Thus, the contents of this lecture can be thought of more as informed speculation rather than hard fact.

We will be discussing the likely possibility that the Universe underwent a period of rapid expansion at very early times. This period of rapid expansion is known as “inflation.”

Inflation is an extension to the hot Big Bang model. It is not an integral part of the hot Big Bang model, however. If inflation were ultimately falsified or superseded, this would not significantly harm the hot Big Bang model itself.

Inflation is also not yet a complete idea. As you will see, there are key aspects of inflation that we simply do not understand well. However, the idea of inflation has some seductively powerful aspects. These lead many to think that inflation is a step in the right direction toward a deeper understanding of the Universe.

## Unresolved Issues in the Hot Big Bang Model

Inflation is considered important because it offers the prospect of explaining some unresolved issues in the hot Big Bang model and in particle physics. Here we will briefly discuss these issues.

### *The Horizon Problem*

First of all, inflation offers the prospect of explaining the amazing uniformity of the cosmic microwave background. The uniformity here seems too good to be by chance. In the laws of physics, there is no obvious compulsion for the Universe to be so uniform. This overall issue is known as the “horizon problem.”

When we look at two widely separated patches of sky, they are separated by billions of light years. Yet somehow they have synchronized the temperature of the cosmic microwave background to be very similar.

One might initially guess that, since the Universe was small at early times, the two patches could have synchronized at early times. However, in the standard hot Big Bang model, this is not the case. Calculations using General Relativity show that the ability to synchronize actually gets worse, not better, at earlier times. This is essentially because the rate at which the Universe expands outstrips the rate at which particles can synchronize (which is ultimately limited by the speed of light).

In the standard hot Big Bang model, two regions more than about 1.4 degrees apart on the sky could not have been synchronized at the time of recombination, when the cosmic microwave background became freed. Yet we observe quite precise uniformity even for patches of sky separated by 180 degrees.

It is as if an orchestra, composed of people who had never met or practiced together, spontaneously began to play almost perfect music!

### *The Structure Problem*

Another tightly related issue is the origin of the small fluctuations that served as the seeds of galaxies. We know that the cosmic microwave background is highly uniform, but it is not perfectly uniform.

What is the origin of these small fluctuations? The standard hot Big Bang model does not give a clear explanation of their origins.

This overall issue is known as the “structure problem.”

### *The Magnetic-Monopole Problem*

Magnets on Earth are always found to have two poles, usually called North and South. Thus, they are always “dipoles.”

Scientists have long searched for magnetic “monopoles” without success. These would be objects with just a magnetic North pole or just a magnetic South pole. If such monopoles exist at all, they are extremely rare.

Advanced theories of particle physics strongly suggest that magnetic monopoles should exist as common fundamental particles, just as electric dipoles do (such as the electron or proton). Thus, it is mysterious that such objects are observationally very rare, if they even exist at all.

This overall issue is known as the “magnetic-monopole problem.”

More generally, advanced theories of particle physics predict a variety of relic particles that have never been observed. What has happened to all these relic particles?

### *The Flatness Problem*

There is also another fundamental problem, known as the “flatness problem.”

Explaining this problem effectively is beyond the scope of this course, so we shall not discuss it here. It is covered in many books on cosmology, so please refer to these if interested.

## The Inflationary Extension to the Hot Big Bang Model

First I will describe the basic nature of inflation and how it offers a possible solution to the problems described above. Then I will address why inflation may have occurred.

Alan Guth originally developed inflation in 1979-1981, and many researchers have subsequently studied it. There are literally hundreds of research papers on inflationary cosmology, and further research is ongoing.

Inflation postulates that the Universe underwent a brief period of rapid exponential expansion very early in its history.

Inflation would have started about  $10^{-36}$  seconds after the Big Bang, and it would have lasted about  $10^{-32}$  seconds.

During this time, the Universe would have expanded by an amazing factor of  $\sim 10^{40}$  to  $10^{100}$  or more. Subatomic scales would have been inflated to be of cosmic size.

The expansion of space during inflation would have been faster than the speed of light, and this might appear problematic according to relativity theory. However, strictly speaking, relativity only says that one cannot propagate an information-carrying signal faster than light. The expansion of space during inflation did not carry any such signals, so there is not a conflict with relativity.

If such a cosmic inflation occurred, it offers a possible solution to the horizon problem. What is now the observable Universe began from a subatomic-scale region that was able to synchronize prior to the beginning of inflation. That is, our observable Universe came from a region much smaller than one would expect by simply extrapolating the standard hot Big Bang model back in time.

There would be many other such subatomic-scale regions that have also similarly inflated. So, when we observe out to our horizon, we

are only seeing part of what is a much larger Universe. Beyond our observable Universe is lots of unobservable Universe.

The unobservable Universe is expected to be much vaster than the observable one, making the total Universe gargantuan. The total Universe could be  $\sim 10^{100}$  times or more larger than the observable one.

There could be other intelligent creatures elsewhere observing out to a different horizon. Statistically speaking, their observable Universe would be similar in its overall properties to ours, since it arose from the same basic statistical processes.

A helpful analogy is two widely separated people looking out into a fog.

The structure problem can also be solved by inflation. The small fluctuations in the temperature of the cosmic microwave background (that are the seeds of galaxies and other structures in the Universe) are due to ultimately microscopic quantum-mechanical fluctuations. The largest structures in the Universe are nothing but quantum-mechanical fluctuations writ large!

Finally, inflation also offers a possible solution to the magnetic-monopole problem. If magnetic monopoles formed prior to inflation, then inflation would have diluted their density greatly, so that they are very rare in the Universe today. (Since baryogenesis occurs after inflation, baryons would not have been similarly diluted).

## **Why Did Inflation Occur?**

So inflation seems like a promising idea, but why might it have occurred? We are not entirely sure.

However, we can take a hint from the dark energy currently dominating the mass-energy of our Universe. Dark energy is currently driving an accelerated expansion of our Universe.

Perhaps there was a more extreme version of dark energy at work in the instants after the Big Bang, and this drove inflation. This “early dark energy” would have had to “decay” away so that the Universe did not continue inflating indefinitely.

The driver of inflation is commonly thought to be an elementary particle yet to be discovered; this particle is generically called the “inflaton.”

Clearly this is one area where the inflation idea is currently incomplete at a fundamental level. We do not have a respectable idea of which particle made it happen!

This is yet another reason why studying dark energy is of such great importance. Its nature may give us a clue about what drove inflation.

## **The Possibility of a Multiverse**

This section is highly speculative and almost philosophical. It is at the edge of current understanding.

The ideas behind inflation, when explored in depth, suggest that there might well be other Universes.

We do not presently have limits on the extent of the region of spacetime that inflated to become our Universe (both its observable and unobservable parts). However, it seems entirely plausible that there could have been many other regions that inflated to create other Universes that are disconnected from ours. The collection of these Universes is called the Multiverse.

Note this is an entirely different “ballgame” than just the observable and unobservable parts of our Universe. These other Universes are more separated from us than just the parts of our Universe that we cannot see at present.

It is also possible that new Universes are continually being created by inflationary events triggered by stellar collapses in our Universe (and others). When a sufficiently massive star collapses to form a black

hole, the resulting singularity is similar in nature to the hot, dense phase at the start of our Universe. Perhaps our Universe inflated from such an event in another Universe, and perhaps new Universes are continually being “spawned” from ours.

This is perhaps the ultimate extension of the Copernican Principle. Not even our Universe is special or unique, but it is just one of many.

What might these other Universes be like? One possibility is that there are deep laws of physics, yet undiscovered, that force these other Universes to be similar to ours in a statistical sense. The same laws of physics apply in these other Universes, they have the same number of dimensions, and the fundamental constants of nature have the same numerical values.

Alternatively, these other Universes could have different laws of physics, different numbers of dimensions, or different numerical values of the fundamental constants of nature.

We hope further work in fundamental physics, such as that on superstring theory, may provide insight here that is currently lacking. Presently this whole enterprise is highly speculative, although it raises many provocative and fundamental questions.

If there are many Universes, this could help to explain the surprisingly “biophilic” nature of our Universe that allows for the evolution of complex, sentient life.

There are many ways our Universe could have been “biohazardous” without complex, sentient life.

- It could have collapsed after just 1, 10, 100, 1000, etc. years, so that life had no time to get started.
- It could have expanded too fast, so that planets, stars, and galaxies never had a chance to form.
- The strong force could have been weaker than it is, so that no atoms other than hydrogen could have formed.

- The strong force could be stronger than it is, so that the diproton is stable. In this case, diprotons would have formed abundantly in the early universe, and there would be no hydrogen left. Thus there would be no fuel for ordinary stars, water would not exist, and complex chemistry seems unlikely.

In fact, the number of ways our Universe could have been biohazardous greatly exceeds the number of ways it could have been biophilic. So our Universe seems to be “finely tuned” to be biophilic.

If there are many Universes, each with different properties, this could help to explain the biophilic nature of our Universe without requiring fine-tuning. In the cosmic dart game, there were many chances to hit the “bull’s eye” of a biophilic Universe, and our Universe was the one that actually did hit it.

An analogous situation is that of an early philosopher pondering why the Earth is so biophilic in terms of its size and its distance from the Sun. There are many more ways a planet could be biohazardous than biophilic. We now know that there are many planets out there with many different properties, and Earth is just the planet that had the correct size and distance to be biophilic.

In the above scenario, the numerical values of the fundamental constants of nature would not have any fundamental meaning, other than that they allow for a biophilic Universe.