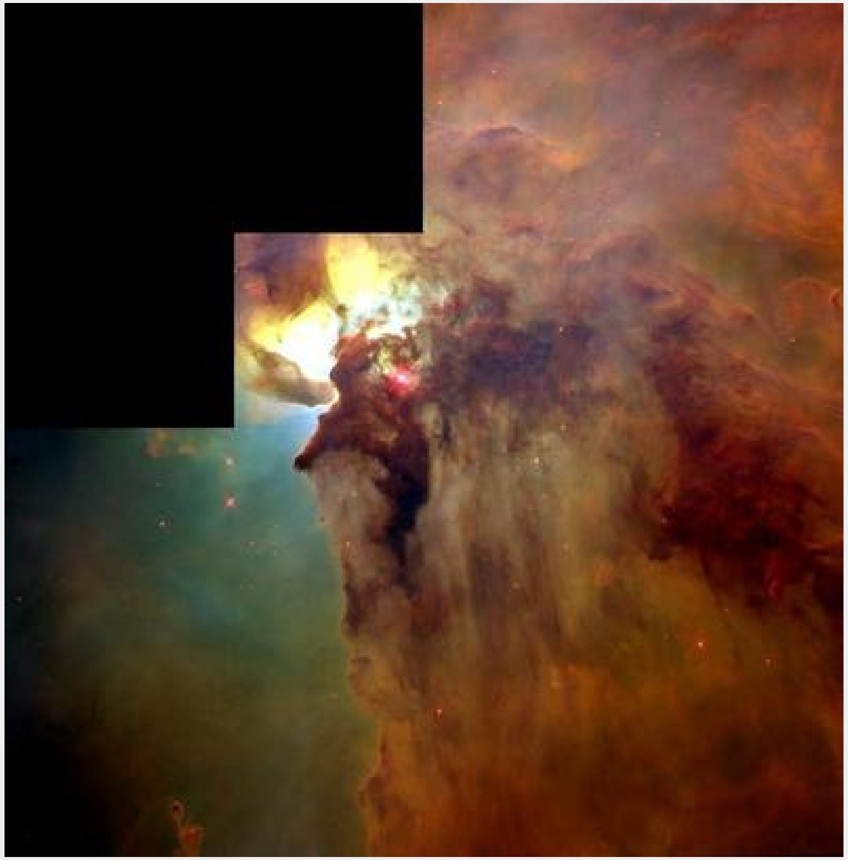
Theories of Galaxy Formation: How the Universe Got its Lumps



Credit: A. Caulet (ST-ECF, ESA) and NASA

Summary

In the final weeks of this Unit we will focus on the formation and evolution of galaxies and the place galaxies hold in our understanding in the overall structure and evolution of the Universe.

In this Activity, we will see how these ideas relate to the **formation of structure in the Universe**.

In particular, we will look at:

- The theory of **gravitational instability** to explain the Universe's structure;
- outstanding problems that motivate continuing research into theories about the formation of the Universe's structure.

Introduction

Once upon a time, when the Universe was young (300,000 years after the Big Bang), space was filled by a uniform hot plasma. All parts of the Universe looked almost identical.

Nowadays, fourteen billion or so years later, the Universe is extremely different, characterised by of all sorts of lumps, such as the Earth, the Milky Way and galaxy clusters.

In this Activity, we discuss the gravitational instability theory and it's explanation of how these lumps came to exist.

In the process, we will revisit some of the ideas we discussed in the context of modelling the formation of the Milky Way, in the Activity *How Old is the Milky Way?*

How did the Universe become so lumpy?

The increase in lumpiness is truly astounding. The densest places in the early Universe were less than 0.01% denser than the least dense places. Compare that to the situation today: the air in your room is $\sim 10^{20}$ times denser than the air 100km above your head!

On larger scales, the density of matter in the middle of a galaxy is thousands of times greater than in deepest intergalactic space. Somehow, most of the matter in the Universe has been swept up into lumps during the last ten billion years or so.

The galaxy distribution: cosmic foam

Here is an animation of large scale galaxy distribution. Note that the distribution looks “foamy”: galaxies lie in sheets separated by bubble-like voids.

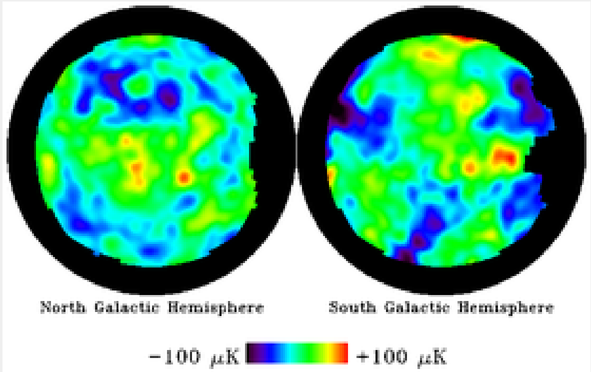
Each dot in this animated ‘fly through’ of the data shows the position of one galaxy in the 2 Degree Field Galaxy Redshift Survey of 250,000 galaxies. The researchers measured the position and velocity of every galaxy down to a certain brightness limit in two strips across the southern sky. This allowed them to construct a 3D map of galaxies within these strips (which look like two wedges in the animation).

Click to Download.
Media: 2dfQTmov.mov
Size: 3.590 Mb
Credit: Paul Bourke, © Swinburne University of Technology

Evidence for gravitational instabilities

- The COBE satellite discovered that there are indeed small lumps in the microwave background: just big enough to have grown by gravitational instability into the galaxy clusters we see today.

Credit: NASA/GSFC/COBE Team



Credit: GSFC, NASA

Problems with the theory

The Gravitational instability explanation is somewhat inadequate, given its rather arbitrary starting conditions. Why should the Universe have some bits denser than others?

Problems and puzzles

Questions that the gravitational instability model needs to answer are:

- How can gravitational instability produce quasars and other such extreme objects early enough?
- Why does gravitational instability produce the bubble-like distribution of galaxies we observe today?
- Why does it produce elliptical and spiral shaped lumps?
- Why doesn't it just produce gas clouds? Or black holes? Or planet sized rocks? Why is there a universal preference for stars gathered into galaxies?
- Where did the very slightly denser regions that started the whole business come from?

Gravitational collapse theory

Theoretical astronomers have been trying to solve these problems for decades. Some of the greatest minds of the 20th century have worked on it.

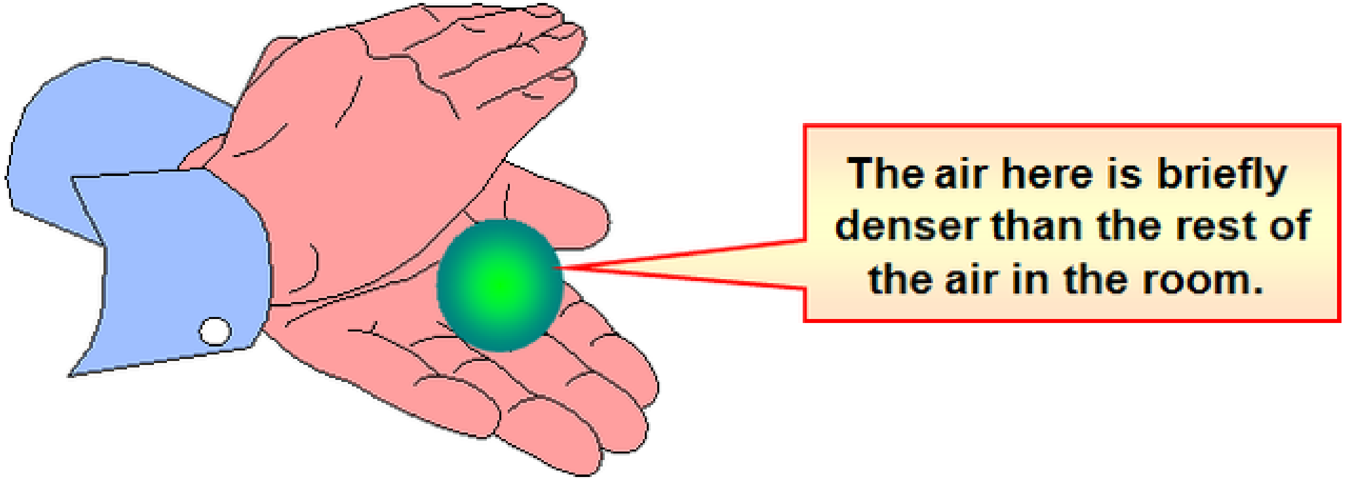
The premise of the problem sounds simple: blobs of gas collapsing under their own gravity. Unfortunately, modelling a solution turns out to be really quite complex and difficult (which is probably just as well, since you couldn't get the wonderfully complex Universe we live in out of a simple process).

For the rest of this Activity, we'll look at some of the achievements and problems of this theory. In the next Activity, we'll get to the incredible recent progress that has been made possible by supercomputers.

Why do overdense regions collapse?

The gravitational instability theory says that regions that are slightly denser than average attract matter towards them (due to their gravity), and hence become denser.

But why doesn't this happen to the air in your room? Clap your hands: this briefly compresses the air between them. For an instant, the air between your hands is denser than the rest of the air in the room... so why doesn't its gravity slowly suck in the rest of the air in the room?



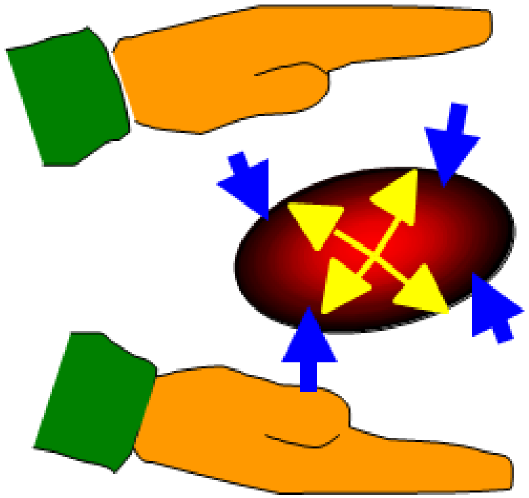
It all depends on gas pressure

Gas pressure depends on the density of the gas, and on its temperature: a hot dense gas has a higher pressure than a cold, thin one.

As you bring your hands together, the air between them is compressed.

The air therefore gets hotter and denser, so its pressure rises (blue arrows).

Gravity tries to suck the surrounding air in. But the increased gas pressure pushes the gas outwards harder (yellow arrows), resisting gravity.



What about space?

This is why the air in your room doesn't collapse into a lump every time you clap your hands: as the gas is compressed, its density and temperature rise, which in turn makes the pressure rise so that the air being sucked in by gravity is pushed rapidly out by the pressure.

Why doesn't this same process occur in space? Let's assume that there were regions of the early Universe with slightly more than the average density of matter. Sure enough, their gravity will suck matter in. But won't their pressure rise to push the gas straight back out again?

The Jeans Mass

If you look at the equations controlling how strongly gravity attracts matter into dense regions, and the equations controlling how strongly their pressure pushes back, you can show that for small objects (like the gas between your hands) pressure wins. For really large objects, however, gravity wins: it overcomes pressure and can make things collapse.

This is because gravity is a long range force and pressure is only a short range force: as you make something bigger, its gravity goes up more than its pressure.

The critical mass above which things collapse is called the Jeans Mass after the British mathematical physicist James Jeans and its value depends on the type of gas and the temperature.

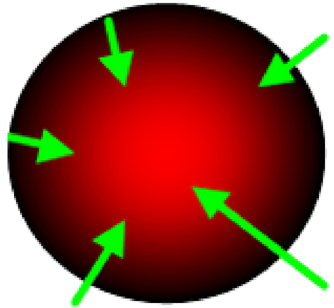
What Collapses, and When?

When you apply Jeans's equations to the gas left over from the Big Bang, you find that the earliest it can collapse is about 300,000 years after the Big Bang (redshift z 1400). At this point, only lumps with masses of more than one million times the mass of the Sun can collapse: smaller lumps are held up by their gas pressure.

By this calculation, most of the matter left over from the Big Bang should have started to collapse into objects of mass 10^6 times the mass of the Sun. Objects with just this mass are common in the local Universe: globular clusters!

What happens next?

Our theory tells us that lots of ‘globular-cluster-mass’ bits of the early Universe start to collapse under the influence of their own gravity. What happens next? Do they keep shrinking until they form black holes? Do they turn into enormous stars? Globular clusters, perhaps? And where do galaxies fit into all this?



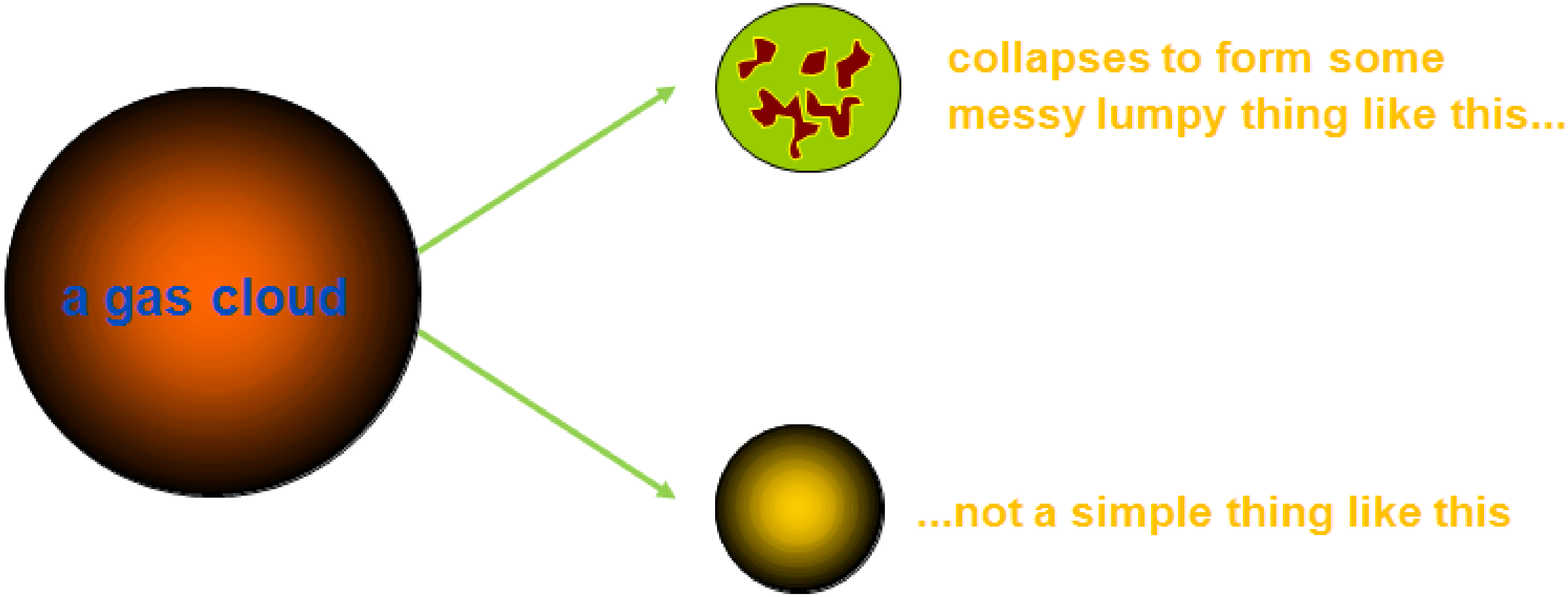
If the primordial lumps were perfectly uniform spheres, then this would be an easy question to answer: the lumps would keep on shrinking until they had produced massive black holes.

Reality, however, is much messier...

The Astrophysicists' Curse!

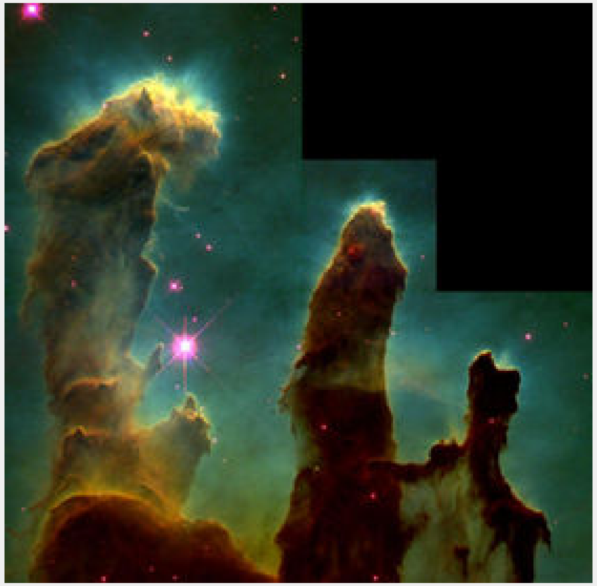
We run into an infamous problem, one that crops up in many branches of astronomy: thermal instability.

As any gas cloud shrinks, it tends to fragment. Instead of forming a uniform smaller dense cloud, it almost invariably turns into a small, but complex and messy mixture of cold dense clouds and hot diffuse gas.



Example: Nebulae

You only have to look at nebulae in our own galaxy to see that they are incredibly complex, lumpy mixtures of high and low density gas, hot and cold gas, ionised and neutral gas. These regions are so complex that the details of even the nearest and best studied nebulae defy detailed analysis.

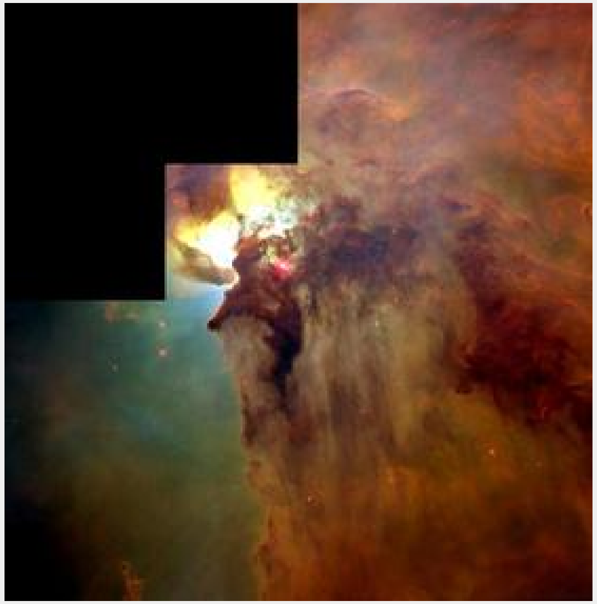


Credit: NASA, ESA, STScI, J. Hester and P. Scowen (Arizona State University)

If even small, nearby galactic nebulae are this complex and confusing, what hope do we have of understanding the much larger and stranger gas clouds that formed during the collapse of primordial lumps?

NGC 6188 nebula

Credit: M. Bessell, MSSSO



The Lagoon nebula: not a simple structure!

Credit: A. Caulet (ST-ECF, ESA) and NASA

Conservation Laws

When in doubt, use conservation laws! This might well be the motto of theoretical astrophysics: whenever things get so messy and complex that detailed calculations are impossible, resort to using the great conservation laws of physics:

- Conservation of Mass
- Conservation of Energy
- Conservation of Momentum

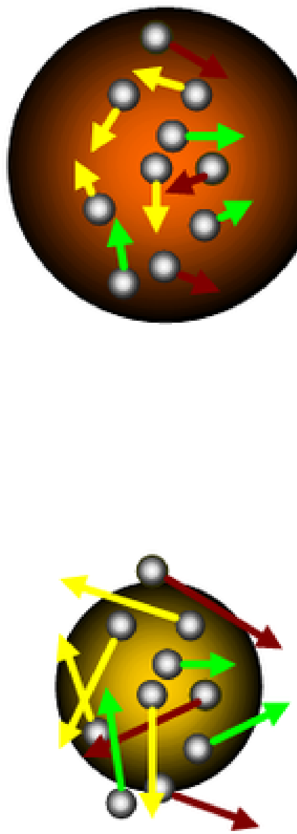
No matter how messy, lumpy and chaotic the gas gets, these laws must still be obeyed. The final result of the collapse must have the same mass, energy and momentum as the initial cloud, regardless of what happened during the collapse.

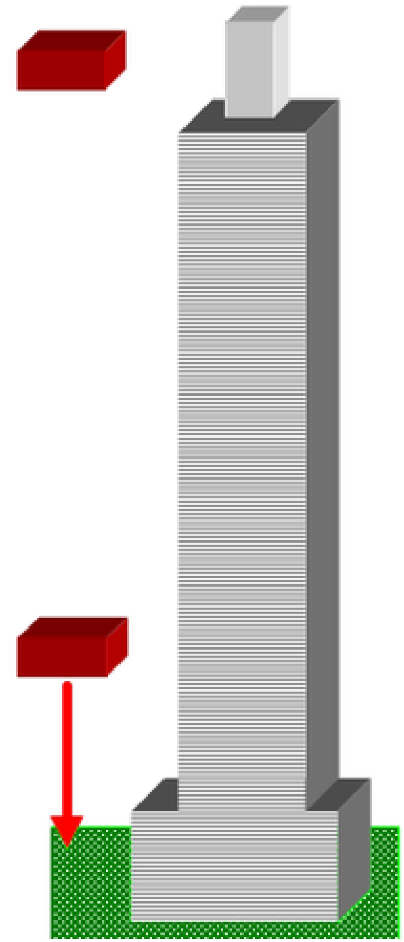
Conservation of Energy

The huge, diffuse initial dense region has a lot of gravitational potential energy (just as a brick on top of a skyscraper has lots of gravitational potential energy).

As it collapses, the potential energy drops. As energy is conserved, however, it must turn into kinetic energy: the individual particles (gas atoms, stars or particles of dark matter) must start moving rapidly.

Similarly, the brick, by the time it reaches the bottom of the skyscraper, will have lost a lot of potential energy, and hence must be moving fast (ouch!)



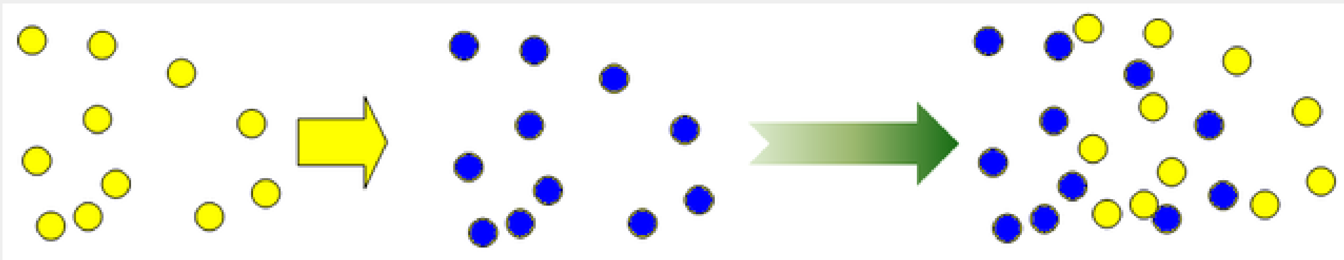


Collisional and Collisionless Materials

If something starts shrinking under the influence of its own gravity, the particles within it must start moving faster. How does this affect the collapsing object?

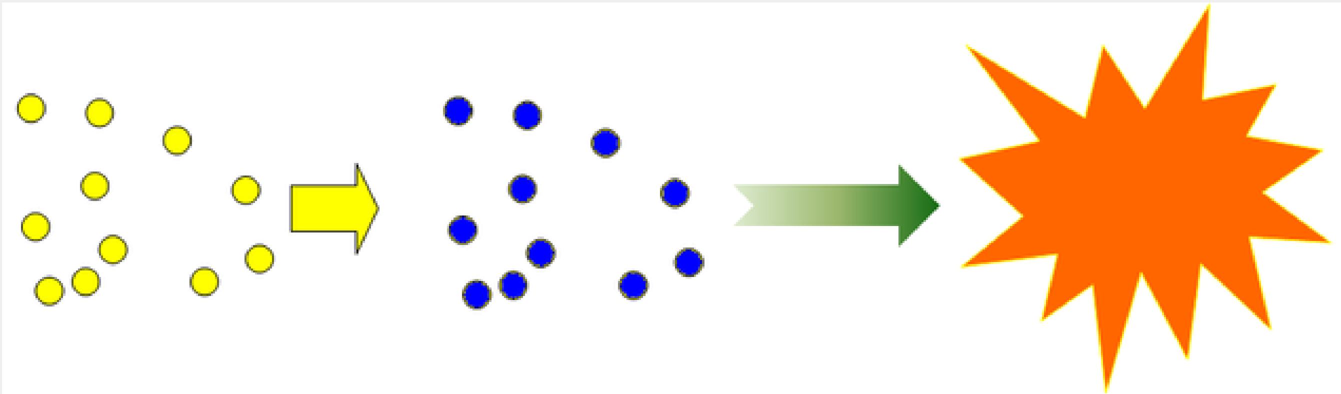
It depends on whether the object is made of collisional or collisionless material.

What does this mean? If you get two clouds of collisionless material and you fling them at each other, they won't collide; they will just pass straight through each other.



Collisionless

On the other hand, if two clouds of collisional material are flung towards each other, they will collide and something violent and messy will happen:



Collisional

Most things on Earth are collisional (car insurance would be much cheaper if they were collisionless). In space, however, things are much more evenly split between the two categories.

Types of collisions in space

Collisional materials

Gas clouds: there is no way a gas cloud can pass through another unscathed.

Collisionless materials

Dark matter: most forms of dark matter that have been hypothesized are collisionless. For example, two clouds of WIMPS, neutrinos or MACHOs could pass easily through each other.

Stars: This may seem strange, but stars are so far apart that even if you collided two dense star clusters together, it is unlikely that any of the stars would ever touch.

Were the primordial lumps collisionless?

Galaxies today are mostly made of dark matter. Therefore, the primordial lumps which collapsed to form present day galaxies were most likely dominated by dark matter too, and dark matter is collisionless.

The rest of the matter starts off as gas, which is collisional. If it stays as gas, it remains collisional. If and when this gas turns into stars, it becomes collisionless.

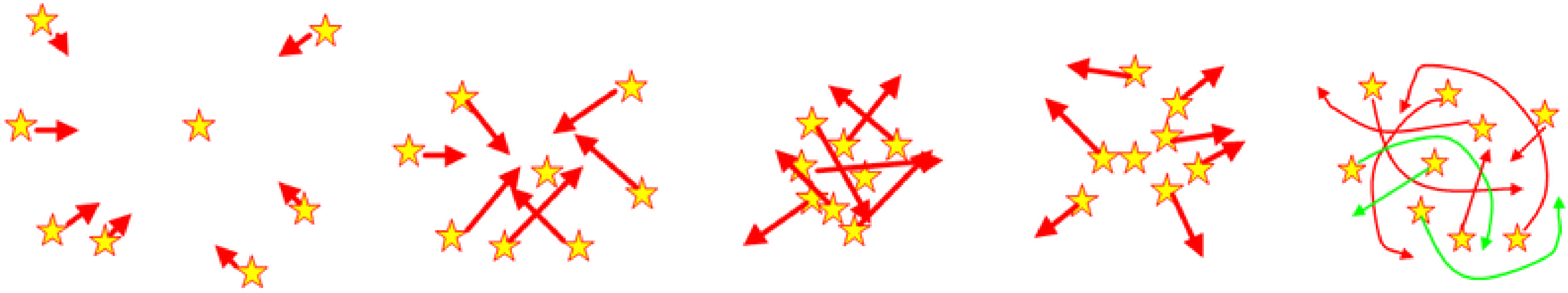
Gas (collisional) until it turns into stars (collisionless)

A collisionless collapse

As energy is conserved, when a cloud of collisionless material collapses, the individual particles start moving faster and faster.

To begin with this motion is inward, but then they fly through the centre and out.

They fall in again, fly through and out, over and over again, until eventually they settle down as a fuzzy ball of particles speeding in and out, round and round, never hitting each other.



A real example!

Elliptical galaxies and globular clusters are fuzzy balls, and the stars in them are flying around, in and out, with exactly these motions!

Could this be how elliptical galaxies and globular clusters formed? Some primordial lump forms stars (thus becoming collisionless) and then collapses to form a diffuse fuzzy blob of rapidly moving stars?

Globular cluster

Credit: Doug Williams, REU Program/NOAO/AURA/NSF

Elliptical galaxy

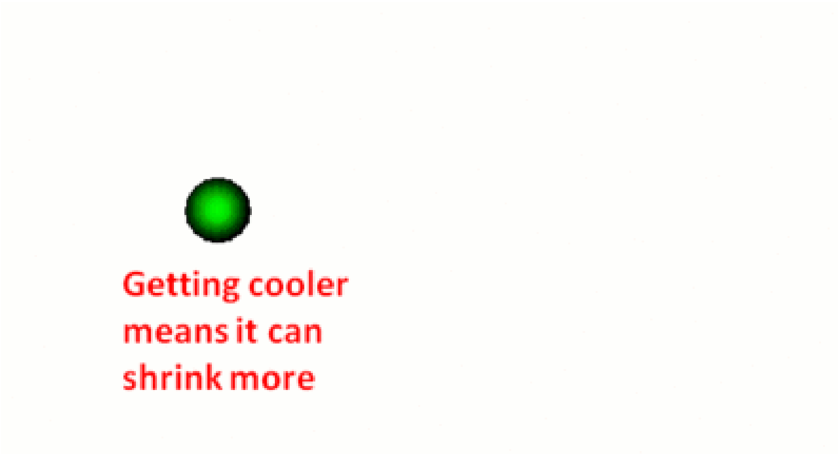
Credit: AAO

A collisional collapse

As energy is conserved, when a cloud of collisional material collapses, the individual particles start moving faster and faster (just as with collisionless matter). The faster they move, the hotter the cloud becomes.

Now comes the crucial difference: as the collisional material shrinks and heats up, the particles smash into each other, which makes them emit radiation: infrared or optical light.

This cloud radiates energy, so the particles move more slowly, allowing it to shrink much smaller.



Summary: two collapse scenarios

Collisional
As the cloud shrinks, the particles bang together and radiate energy. This slows them down, so the cloud can collapse further still. In principle, a collisional cloud can radiate all its initial potential energy and shrink all the way down to a dot.

Collisionless
As the cloud shrinks the particles move faster and faster, until their motions balance gravity, and you get a fuzzy ball of fast moving particles.

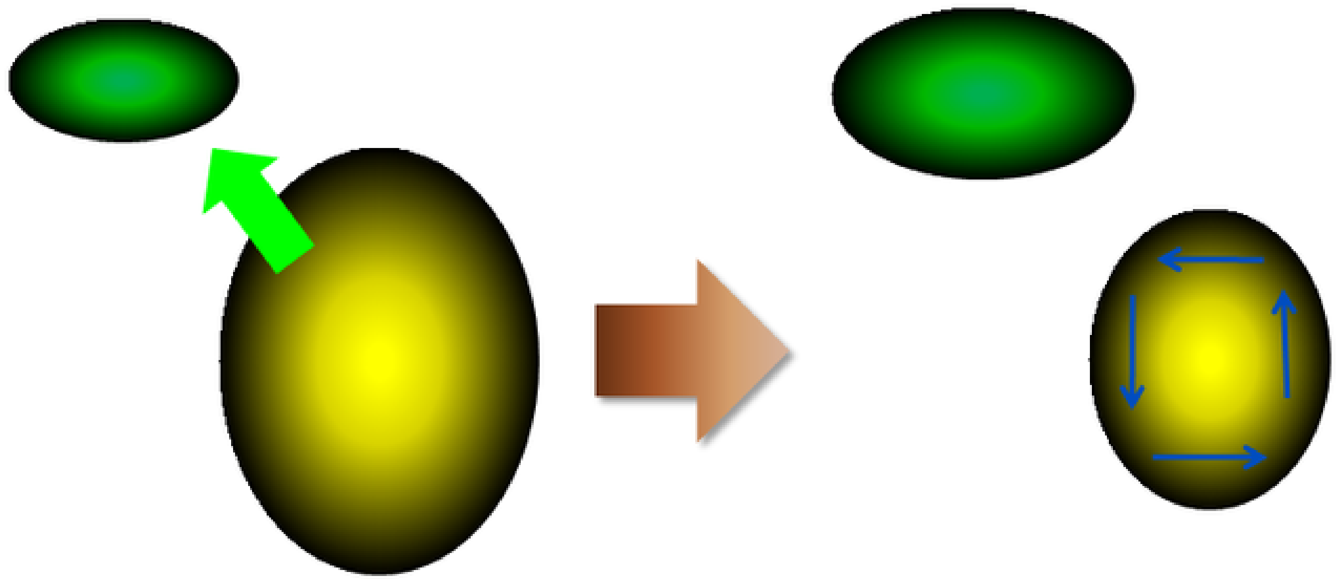
There is no way in which the particles can lose energy, so the cloud can never get smaller that this fuzzy ball.

Do collisional objects become black holes?

Collisional things (like gas clouds) can radiate energy, which means they can shrink much smaller than collisionless objects. What is to stop them shrinking all the way down to black holes?

If they were all alone in the Universe, they probably would. Luckily for us, these lumps have neighbours.

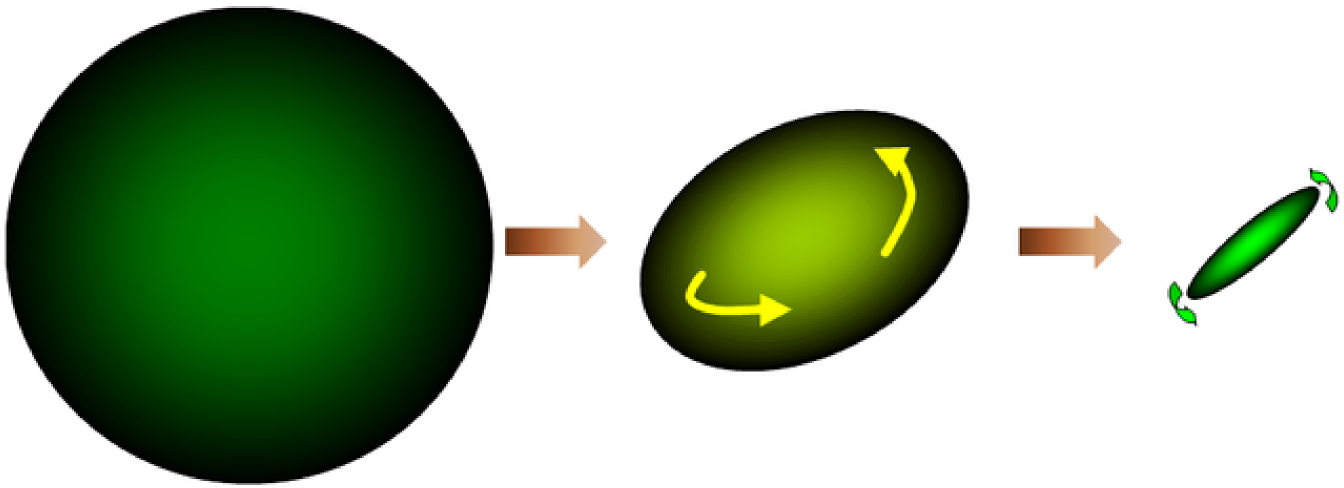
In the diagram below: as the yellow cloud shrinks, the gravity of the neighbouring cloud pulls at the top of the shrinking cloud (green arrow). This makes it rotate very slowly (blue arrows).



Rotation stops the collapse!

The induced rotation is very slow: you would think it was insignificant. But now, the law of conservation of angular momentum comes in. If a spinning object shrinks, it has to spin faster. Think of an ice-skater drawing his or her arms inwards while spinning - the speed of their spin increases dramatically.

As our collisional lump shrinks, its rotation must increase: the tinier it gets, the faster it spins, until the rotation balances gravity and our lump ends up as a spinning disk.



Making spiral galaxies?

Most of the stars in spiral galaxies lie in a very small flattened, rapidly spinning disk, just like this!

Could this be how spiral galaxies formed? A primordial gas cloud collapses before it has turned into stars (i.e. while it is still a collisional material). As it shrinks it cools, allowing it to shrink further, until it ends up as spinning disk of gas, which only then turns into stars.

Credit: The Hubble Heritage Team (AURA/STScI/NASA)

Two ways to collapse...

Have our two collapse models described the origin of spiral and elliptical galaxies?

The dark matter will always collapse collisionlessly, forming a fuzzy (and invisible) blob of dark matter.

The gas can do one of two things. If it turns into stars early in the collapse, they will shrink down collisionlessly and form a fuzzy blob of stars: an elliptical galaxy or globular cluster.

If, however, it stays in the form of gas, it can shrink much further because it can radiate energy. Indeed, the gas can shrink so far that its tiny initial rotation comes to dominate the gas, transforming it into a tiny, spinning disk. If star formation only occurs at this late stage, we have a spiral galaxy!

A tentative theory of galaxy types

We have just presented one of the popular rival theories that purports to explain why galaxies come in two major types: spirals and ellipticals. To recap,

- Elliptical galaxies: the gas turns into stars right at the beginning of the collapse of the primordial dense region.
- Spiral galaxies : gas only turns into stars right at the end of the collapse.

Why is there a difference? Perhaps elliptical galaxies originate in marginally denser regions of the early Universe, and this slightly higher density triggers star formation earlier?

Summary

In this Activity, we have discussed the currently accepted theory for how galaxies form. Regions of the early Universe are slightly denser than others (the reasons for this are poorly understood), these collapse under their own gravity to form galaxies.

The details of this model are still being investigated. If the collapsing regions form stars early, they might become elliptical galaxies, while late star forming regions might end up as spirals. Why they should behave like this is unknown, however.

In the next Activity, we will continue to investigate the theory of how the Universe got its lumps, including the dramatic new breakthroughs made possible by the use of supercomputers.