

A question-led introduction to cosmology

The universe and its history

Evidence, inflation and unanswered questions

David Elliman

Neuro-Symbolic

Modern physics through questions

Volume 4

Preface

Modern physics is fascinating, often surprising, and sometimes hidden behind unfamiliar mathematics. Popular accounts can awaken curiosity without showing enough of the reasoning to satisfy it. University textbooks supply that reasoning, but often assume mathematical fluency that a reader has not yet acquired, or has not used for some time.

At Neuro-Symbolic, we needed material to help able graduates understand the physics relevant to their work. These four books are our attempt to meet that need. They may also interest undergraduates, researchers entering a new field, and readers who want to understand more than a memorable analogy.

The books use questions and answers. We found this approach useful in workshops: a concrete puzzle gives people a reason to examine an idea and discuss their predictions. Each group of questions begins with the physical setting. Explanations use experiments, diagrams and simple arithmetic before introducing symbols. Optional mathematical sections connect that reasoning to the notation used in more advanced texts.

Our aim is to explain what the theories say, why physicists take them seriously, and where their explanations remain incomplete. Established results, assumptions of an example and unsettled proposals are identified separately. A useful calculation need not explain every ingredient that went into it.

We have taken care to check the physics and the examples, but that is no guarantee against error. Mistakes are our responsibility, and we welcome corrections from researchers and readers. The most useful feedback often begins with the exact sentence where an explanation stopped making sense.

Acknowledgement

OpenAI's Codex assisted with organisation, prose, diagrams and teaching checks. Responsibility for the content rests with the author.

How to use this volume

How could anyone know what happened before the first stars? Could the region we see really have grown from something smaller than a proton? Does faster-than-light expansion break relativity, and do other universes follow? These questions deserve explanations before equations.

We begin with records: spectra, nuclear abundances, background radiation and the distribution of matter. We then ask how far a cosmological model can take us backwards, and where an additional hypothesis enters. The hot Big Bang history, an inflationary mechanism and a multiverse proposal do not have the same evidential status.

Each part opens with physical context. Read it, then the short transition before each question. Make a prediction before continuing. Numerical examples illustrate a relationship; their invented values are not measurements of the early universe. Worked answers and a glossary are at the back.

The companion *Space, time and gravity* explains clocks, light cones and Einstein's equation. The quantum and particle volumes supply the background for amplitudes, fields, nuclei and matter.

There is no requirement to accept an entire cosmological story as one package. Ask what was observed, what was inferred within a model and what remains a proposal. Common-sense objections are useful when they help locate a missing assumption.

Contents

Part one: Reading the cosmic record	5
1 Why is the universe expanding?	5
2 What does the hot Big Bang actually explain?	7
Part two: What inflation would explain	10
3 Why propose inflation in the first place?	10
4 Could our universe once have been smaller than a proton?	12
5 Can cosmic expansion outrun light?	13
6 Did inflation create matter and energy from nothing?	15
7 What could convince us that inflation happened?	17
8 Does inflation mean there are other universes?	19
Part three: What the dark components mean	21
9 Why infer so much matter we cannot see?	21
10 Is dark energy the same thing as inflation?	23
11 Why is the gravitational vacuum contribution so small?	24
Part four: Why the history has a direction	26
12 Why was the early universe so orderly?	26
13 Why does time have a direction?	28

For checking: Answers to the Try questions	30
For returning to: Words and sources	32

Part one

Reading the cosmic record

Astronomy gives us records of an earlier universe. Light takes time to travel. Its spectrum describes how much light arrives at different wavelengths, just as a rainbow separates visible colours. Atoms emit or absorb characteristic wavelength patterns, called spectral lines. Comparing those patterns with laboratory measurements can reveal a shift towards longer wavelengths, called redshift. This is one part of the evidence for cosmic expansion.

Cosmologists combine those records with the cosmic microwave background, a faint glow of long-wavelength radiation arriving from across the sky. They also measure light-element abundances, meaning how much hydrogen, helium and other light elements there are, and the distribution of matter. Their large-scale description averages over stars and galaxies; it does not require every atom or ruler to expand.

We start with what this evidence supports before asking about an earlier inflationary phase. A reconstruction of a hot past and a theory of an absolute beginning are different achievements.

1 Why is the universe expanding?

Before reading on. If every large-scale separation increases by one per cent, do nearby and distant pairs gain the same extra distance?

Take two separations in an ideal homogeneous universe: 10 units and 20 units. A one per cent increase adds 0.1 unit to the first and 0.2 unit to the second. More distant pairs separate faster because the same fractional change acts on a larger distance.

Cosmology describes this behaviour with a scale factor. It tells us how distances between locations following the idealised cosmic flow change. Locations following that flow are called comoving: their map labels stay

fixed while their physical separation changes. Nearby galaxies can also have motions relative to the flow, so not every measured velocity follows the simple expansion rule. Homogeneity means statistical similarity from place to place on large scales; it does not deny galaxies, voids or local structure.

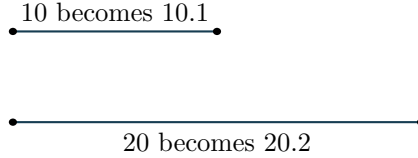


Figure 1: The same fractional growth increases a longer separation by a larger amount. The units and one-per-cent change are illustrative. This is a distance account, not an explosion from a marked centre.

The observations supporting an expanding hot-universe history include galaxy redshifts, the cosmic microwave background and light-element abundances. One distance–redshift plot alone would not establish the entire cosmological model. The evidence works together.

Energy density means energy per unit volume. General relativity relates the evolution of the scale factor to energy density, pressure, spatial curvature and initial conditions. It does not say that every possible universe must expand. There are expanding and contracting solutions, and different matter contents give different histories.

Expansion is not an explosion into a surrounding empty room. Within the homogeneous description there need be no preferred central galaxy from which everything was fired. Every location can describe sufficiently distant comoving locations as receding.

Bound systems do not simply grow with the cosmic scale factor. Atomic binding, planetary dynamics and galactic gravity matter locally. The cosmic description is a large-scale approximation, not a command that every ruler must stretch.

Light travelling between comoving emitter and observer is cosmologically redshifted. If the scale factor doubles during its journey, its wavelength doubles and its photon energy halves. For light in vacuum, photon energy is inversely proportional to wavelength. The energies here are those measured locally at emission and reception by observers following

the cosmic flow; they need not agree across an evolving universe. That is one way the expansion connects geometry with the cooling of radiation.

Try 1. A separation of 30 units increases by one per cent. What extra distance does it gain?

Following the expansion backwards suggests a hotter and denser past. Nuclear abundances and the microwave background test parts of that history. The phrase Big Bang can then mislead us into picturing a single explosion, so we need to separate the tested history from questions about its earliest origin.

2 What does the hot Big Bang actually explain?

Before reading on. Does evidence for an early hot, dense universe by itself reveal an absolute first instant or an explosion into pre-existing space?

The hot Big Bang model describes an early hot, dense phase and its subsequent evolution. As the universe expands, its radiation cools and its reactions change. The model is not, by itself, an answer to every question about an ultimate beginning.

A plasma is matter in which electrons and charged nuclei can move separately, rather than remaining bound into neutral atoms. In the early hot plasma, energetic collisions continually changed which particles and bound structures could survive. During the first minutes, nuclear reactions bound many of the available neutrons into helium and produced small amounts of other light nuclei. Most protons remained as hydrogen nuclei. Much later, about 380,000 years into the standard hot-universe history, electrons could remain bound to nuclei in neutral atoms. With far fewer free electrons to scatter from, photons travelled much farther between scatterings. The microwave background is the redshifted relic of that later release of radiation, not a photograph of the first instant. The long wavelengths we observe now explain its name: cosmic microwave background.[8]

The distinction between nuclei and atoms is important. Making a helium nucleus does not immediately attach electrons to it in a stable neutral atom. The plasma can remain ionised long after nuclear formation.

There is evidence from more than one epoch. Light-element abundances test nuclear processes in the early hot plasma. The microwave background tests the later release of radiation and the small departures from a perfectly uniform state, called perturbations, present at that time. Galaxy observations test still later evolution. A successful history must connect these records with the same physical model.

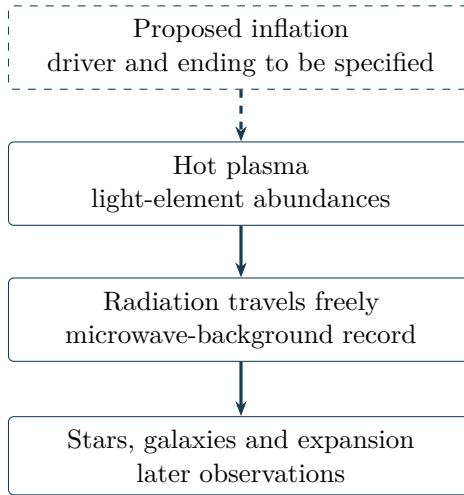


Figure 2: Different evidence probes different stages. The proposed inflationary stage precedes the tested hot history. The spacing of the boxes is not a scale of elapsed time, and the earliest box is not an observed first instant.

Big Bang can also refer to the singular limit of particular classical solutions extrapolated backwards. A singularity in that description signals a limit of the classical account; it is not an observation of creation from nothing.

Nothing in the hot-history evidence requires an explosion into pre-existing surrounding space. Expansion changes distances within the cosmological geometry. Nor does it identify the complete size of the universe from the region visible to us.

The question of spatial flatness and the origin of the primordial variations comes next. Inflation is one developed attempt to account for those conditions. Primordial means inherited from the early universe. The term labels the epoch, not an established origin mechanism. Inflation's evidence and assumptions must be assessed separately from the well-supported thermal history.

Try 2. Why is forming light nuclei different from forming neutral atoms? Does either event identify an absolute beginning of the universe?

Part two

What inflation would explain

The hot Big Bang account starts with a hot, expanding universe. It does not, by itself, explain why widely separated regions had very similar conditions or why the large-scale spatial curvature is so small. Inflation proposes an earlier period of accelerated expansion to address those puzzles and supply the initial variations from which structure could grow.

Several claims are often compressed into one story: a tiny beginning, expansion faster than light, creation of matter, and other universes. They are not interchangeable. We will examine the physical meaning and evidence for each before deciding what follows.

Accelerated expansion means that the scale factor grows by increasing amounts in successive equal intervals of cosmic time. Cosmic time is measured by ideal clocks following the average cosmic flow, with their readings referred to a common cosmological history. It is not a claim that relativity has restored one absolute time for every moving observer. A microscopic model must specify what drives that change, how it ends, and how the subsequent hot phase arises. The word inflation names a family of proposals, not a directly observed film of the earliest universe.

3 Why propose inflation in the first place?

Before reading on. If two distant regions had no time to exchange signals in a proposed history, would that history explain why they reached the same temperature by ordinary mixing?

No. A successful explanation through mixing requires time for interactions to establish it. The microwave background is extremely uniform, with small, structured variations. In a simple hot-universe history continued backwards without an earlier accelerated phase, sufficiently

distant parts of the observed background had no shared causal past of the needed kind. This is the horizon problem.

Inflation changes that history. Under suitable initial conditions, the region that became our observable patch can descend from a much smaller region with causal contact. Accelerated expansion stretches it, followed by a hot phase in which the expansion evolves differently. This can explain why regions now far apart share similar conditions without requiring superluminal messages.

The same expansion can reduce the importance of spatial curvature over the region we observe. A spatial slice is a map of locations at one chosen cosmic time. Spatial flatness means that distances on this three-dimensional map obey Euclidean geometry: for example, a sufficiently large triangle has angles adding to 180 degrees, within observational uncertainty. It does not mean the universe is a flat two-dimensional sheet or that spacetime has no curvature.

Inflationary models also give an account of the small primordial variations. Quantum fluctuations are the nonzero spread and correlations of field measurements even in a specified quantum state. In inflationary models, they can produce small variations in the later matter and geometry from which gravitational clustering grows. Calculating their statistical pattern connects a proposed early dynamics with measured temperature and polarisation patterns. Polarisation describes the orientation of the light's electric-field oscillations; a sky map can record that orientation as well as brightness.

These are substantial reasons to investigate inflation. They are not proof that a unique inflationary model has been identified. The model still needs suitable initial conditions, a physically specified field or other driver, and a successful exit. It does not automatically explain why any suitable initial state existed.

The observational test concerns the combination of predictions. A story that merely makes space very large is insufficient. It must also produce the right kind of small variations and a plausible subsequent thermal history. Alternative early histories face that same requirement.

Try 3. What is missing from an explanation that says two regions equalised their temperatures but gives them no time to exchange signals?

Inflation can connect the region we see to a much smaller earlier region. The next question asks exactly what region is meant when an account compares the young universe with the size of a proton.

4 Could our universe once have been smaller than a proton?

Before reading on. If the part of a landscape visible from one window was once very small in a model, would that establish the past size of the entire landscape?

First distinguish the observable universe from the whole universe. Our observable region is limited by which signals could have reached us by now in the specified cosmic history. That is different from the question of which future signals might eventually arrive. It is not known to be the complete universe. The whole spatial universe could be much larger, and some models allow it to be infinite throughout the expansion.

An assertion about a proton-sized beginning commonly refers to the earlier physical size of the region that grew into our observable patch. It need not mean that all of existence was a little ball surrounded by an empty room. In an infinite homogeneous model, every specified finite patch can shrink when followed backwards while the whole spatial universe stays infinite.

The calculation uses an expansion history. Follow the same idealised comoving region backwards, dividing its present linear size by the intervening growth factors. For an illustration, ten doublings multiply a length by 1024. Twenty multiply it by 1,048,576. Enormous changes can result from repeated fractional growth; the arithmetic itself is not impossible.

But the starting size obtained in this way depends on the proposed duration and energy scale of inflation, its ending and the subsequent thermal history. There is no unique observed sub-proton diameter or universal initial time supplied by the phrase inflation happened.

Nor is the proton a little container known to exist at that proposed early epoch. Its present size is being used as a familiar length comparison: the proton's extent is of order one millionth of a billionth of a metre. Different measures of that extent need not give exactly the same number. The matter state in a sufficiently energetic early universe would not consist of today's intact protons.

Why take such a remote claim seriously at all? Physics often reconstructs earlier conditions from surviving records. Here those records include nuclear abundances, background radiation and cosmic structure. Confidence weakens as the extrapolation enters less tested physics. A responsible account identifies that boundary instead of presenting the entire backward extrapolation as a measurement.

You need not accept a vivid size slogan on authority. Ask which region, which model and which observations constrain the inference. Those questions turn an astonishing claim into something that can be assessed.

Try 4. How much does a length grow after ten doublings? Does that arithmetic establish the initial size of the whole universe?

Large growth factors explain how an initially small patch could become large. They raise a separate concern about relativity: whether distant regions can separate faster than a light signal travels locally.

5 Can cosmic expansion outrun light?

Before reading on. Must the rate of change of a distance between very distant comoving locations obey the same bound as a nearby object's speed measured in one laboratory?

They are different measurements. The relativistic speed limit concerns local motion and signalling: an object or message cannot overtake a neighbouring vacuum light signal in its local inertial frame. A distant separation requires a convention for comparing locations across an extended spacetime.

In the usual cosmological description, we compare distances on slices of common cosmic time. The proper distance is the distance assigned by that same-time spatial geometry, as if local ruler lengths along the route

were added together. It is not the distance obtained by multiplying a photon's travel time by its speed. Between comoving locations it changes as the scale factor changes. Its rate can exceed the invariant light speed for sufficiently distant locations, without either location passing a local light ray.

A numerical illustration makes the distinction visible. Suppose each of ten consecutive intervals gains one unit of length per unit of cosmic time. The full separation gains ten units per unit time. This adds changes across an extended region; it does not describe one body travelling through each interval at that total rate.

Inflation is not defined simply by one recession speed exceeding light speed. Even a decelerating expansion can have sufficiently distant locations receding faster than light in this sense. Inflation is an accelerated phase: the scale factor grows by increasing amounts in equal intervals of cosmic time. These are different tests of the expansion history.

Whether a particular light signal can eventually reach us depends on the whole intervening expansion history. A galaxy's superluminal recession rate at one time is not, by itself, a complete visibility test. Light always follows the appropriate local causal paths.

The phrase space expands faster than light is therefore a shorthand that can cause trouble. An expansion rate as a fractional change has units of inverse time, not speed. To get a recession speed, one also specifies a separation. No claim that information outruns the local light cone is needed.

Try 5. If ten intervals each lengthen by one unit per unit time, how fast does their total length grow? Why is that not the measured speed of one local object?

Separating recession from local speed removes one apparent conflict. It does not explain where the hot matter came from. For that we need the proposed energy-bearing field and its interaction with other fields. A field assigns physical variables throughout space. Its excitations can carry energy and, in suitable circumstances, appear as particles.

6 Did inflation create matter and energy from nothing?

Before reading on. If one field transfers energy into excitations of other fields, is that the same claim as creating a universe from an absence of fields, laws and state?

No. In a familiar class of inflationary models, a field called the inflaton supplies the energy density that drives the accelerated expansion. The model already assumes that field, its dynamics, gravity and an appropriate state. This is considerable physical structure.

In the simplest examples the inflaton is a scalar field: its local value has no spatial direction attached, unlike an electric-field arrow. Its energy includes a potential contribution determined by that value, as well as contributions from changes in space and time. If the field is sufficiently smooth and changes slowly, potential energy can dominate.

That state has negative pressure. Positive pressure pushes outwards on a surface; negative pressure has the sign of a tension. This is a useful sign comparison, not a proposal that space is made of stretched rubber. In general relativity pressure as well as energy density contributes to the expansion dynamics. A sufficiently negative pressure can make the expansion accelerate. For a potential-dominated scalar field, pressure is approximately minus its energy density in consistent units; the gravitational equations give the accelerated behaviour. It does not follow from treating the field as an ordinary cold or empty gas.

In many models, inflation ends as the field evolves out of that regime. Through its couplings it can transfer energy into other fields. Their excitations interact and can form a hot plasma. Repeated interactions distribute energy among the excitations; approaching a thermal distribution in this way is called thermalisation. The broad transfer process is

called reheating, although its detailed mechanism depends on the model. The name does not by itself establish a hot phase before inflation.

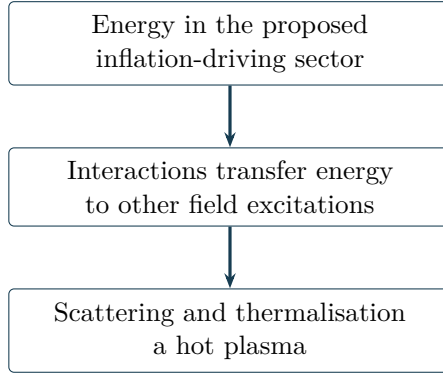


Figure 3: A proposed sequence, not a measurement of the inflaton. The hot phase has observational tests; the earlier driver and transfer mechanism need their own physical specification.

As an accounting illustration, transferring six of ten energy units from one component leaves four there and puts six elsewhere. Real reheating also includes expansion and the evolving fields. The point is the need for a transfer law, not those invented numbers.

There is a further subtlety. In an expanding universe, a region's matter energy does not generally obey the same global conservation rule as a sealed box in a time-independent laboratory. General relativity retains local stress-energy conservation, but a general cosmological spacetime need not have a unique conserved total energy of the familiar kind.

For nearly constant inflationary energy density, the energy assigned to a growing comoving volume can increase. The cosmological conservation relation includes pressure work, the term accounting for how a component's energy changes as its volume changes. For nearly constant energy density, twice the comoving volume contains twice the assigned field energy. This is allowed by the local conservation equation with negative pressure; there need not be a separate global energy account with the laboratory-box form. It is misleading either to call this ordinary energy borrowing or to declare that negative gravitational energy has universally cancelled everything in a uniquely defined total.

Reheating addresses how a specified early state could produce the later hot contents. It does not answer why the field, laws or initial state existed. A quantum vacuum is a state of a theory, not philosophical nothing. Those origin questions should remain visible rather than disappear inside that one word.

Try 6. In the ten-unit transfer example, what remains after six units move to the other component? Which additional ingredient is needed to predict an actual reheating process?

We have separated the proposed driver, expansion and transfer into matter. An explanation earns support by predicting surviving records. We can now ask which measurements discriminate between possible early histories.

7 What could convince us that inflation happened?

Before reading on. If several models produce a nearly uniform universe, would uniformity alone identify which one describes its early history?

One shared feature cannot do that job. The stronger test is a pattern of predictions that must fit together without independently adjusting every observable. Cosmologists compare the distribution of primordial variations across sizes, their correlations and their effect on later structure.

The microwave background is particularly useful. Its small temperature variations and polarisation contain information about the earlier plasma and its initial perturbations. In a specified model, the statistical character of the initial pattern can be calculated from the inflationary dynamics.

Planck's published inflation analysis finds support for several features predicted by simple models, including an almost scale-invariant pattern with a small measured departure from exact scale invariance. Here the initial curvature fluctuations contribute roughly comparable power in wavelength ranges spanning the same ratio, such as 10 to 100 and 100

to 1000 in chosen units. Power measures their mean squared size. This does not mean that the later galaxy pattern or every feature of the microwave sky looks identical at every scale. The analysis constrains particular model families; it does not identify a unique inflaton or directly photograph inflation.[3]

A primordial gravitational-wave contribution would be another important clue. It can produce a characteristic part of the background polarisation pattern, called B modes. Polarisation describes the orientation structure of the radiation, and the B-mode classification describes a pattern across the sky, not a new kind of photon.

Galactic dust and gravitational lensing also produce relevant patterns. They must be separated from a claimed primordial signal. For a concrete published example, the BICEP/Keck 2018 data analysis reported an upper limit rather than a detection of primordial gravitational waves. The technical name tensor signal refers to that gravitational-wave contribution, not to a separate type of observed matter. That dated result is a constraint, not a forecast of what later experiments must find.[4]

Failure to detect a large signal can exclude models that require one while leaving models predicting a smaller signal viable. That is not permission to alter any model without cost: its assumptions and predicted pattern must be specified before comparison.

The appropriate judgement is graded. The hot expanding history is strongly supported. Inflation offers a developed and testable account of important earlier conditions. Its detailed mechanism and some more remote consequences remain unsettled. Confidence in one layer should not be silently transferred to all the others.

Try 7. Why does an upper limit on primordial gravitational waves test some inflationary models without rejecting every possible inflationary history?

Tests concern the history and records accessible to us. Some extensions of inflation describe regions beyond those records. Before calling them other universes, we need to see which extra assumptions lead there.

8 Does inflation mean there are other universes?

Before reading on. If a theory allows an eternally inflating regime for some choices of dynamics and state, does every successful inflationary model have to realise that regime?

No. Inflation and eternal inflation are not the same claim. Inflation refers to an accelerated phase. Eternal-inflation scenarios add a mechanism by which inflation ends in some regions while continuing to generate an increasing inflating volume elsewhere.

There are different proposed mechanisms. In one class, a metastable field state is long-lived but can change to a lower-energy state. Regions making that change form expanding bubbles while other regions continue inflating. The bubbles are regions with a different field state, not soap films in surrounding empty space. In another, quantum fluctuations compete with the field's evolution so that some regions remain in an inflating regime. Their realisation depends on the dynamics and the range where the model is trusted.

Regions where inflation ends can have separate hot histories. They are sometimes called pocket universes. That phrase need not mean completely unrelated mathematical realities. Within the proposal they belong to a specified larger spacetime construction, although communication between them may be impossible.

The possibility is a legitimate consequence to investigate in a model. It is not an observation of other universes. Nor does explaining our microwave background automatically establish the behaviour of a field in remote regimes that those data do not probe.

Probabilities become especially troublesome when a model produces infinitely many regions. Saying an outcome occurs somewhere does not predict how often an observer should expect it. A rule for comparing

populations, usually called a measure, is needed. Different rules can give different answers. This is a substantive limitation, not merely an awkward choice of words.[5]

Have cosmologists lost touch with common sense? Everyday intuition is useful for detecting an unexplained step, but it is not a reliable law of physics outside familiar conditions. The fair questions are whether the mathematics is consistent, whether the assumptions are justified, and what could distinguish the proposal from its alternatives.

It is reasonable to find this part of the subject speculative. One can take the tested hot-universe history seriously, investigate inflation, and remain unconvinced by a particular multiverse claim. Those are compatible judgements, not an obligation to accept or reject cosmology as a package.

Try 8. Name two additional problems that must be addressed before a possible eternally inflating model becomes a prediction about other universes or observers.

Part three

What the dark components mean

Stars reveal only part of the cosmic energy account. Motions, gravitational lensing and the expansion history provide information that a photograph of luminous matter alone cannot supply.

Dark matter and dark energy name different roles in a cosmological model. The first helps account for clustering and gravitational structures. The second describes a component associated with the late accelerated expansion. Neither name identifies a microscopic constituent.

The next questions separate these inferences from ordinary unlit matter, the proposed inflaton and formal vacuum-energy sums. Agreement with a measured abundance is a constraint on a theory, not an explanation of why that abundance occurs.

9 Why infer so much matter we cannot see?

Before reading on. Does dark matter mean all ordinary matter that happens to emit little visible light?

Ordinary matter is often called baryonic because most of its mass is in protons and neutrons, which are baryons. Electrons are included in this ordinary-matter account but are not themselves baryons. Some ordinary matter is dim: gas, faint objects and other material can be hard to see. The cosmological dark-matter problem is broader. Within the standard cosmological model, the combined evidence requires much more matter than the baryonic inventory can supply.

The evidence includes galactic and cluster dynamics, gravitational lensing, the microwave background and the growth of structure. These probes do not all use the same physical assumptions in the same way. A successful alternative must address their combination, not only fit one rotation curve.

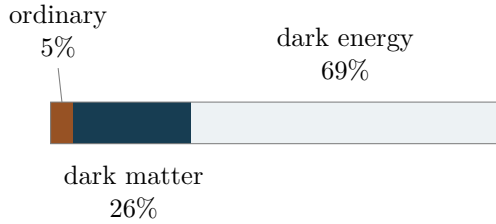


Figure 4: A rounded present-day benchmark within the standard cosmological model: 5 per cent ordinary matter, 26 per cent dark matter and 69 per cent dark energy. These are model-dependent energy-density fractions, not timeless proportions or a count of particles.

Using those rounded numbers, the dark-to-ordinary matter ratio is $26/5 = 5.2$. That is the origin of the rough statement about five times as much. The precise fitted fractions depend on the model and datasets. They should not be promoted into exact integers.

Cold dark matter behaves, on the relevant large scales, approximately as matter with little pressure. Cold here means that its particles move slowly compared with light during the relevant era; it is not a claim about a temperature one could feel. It clusters gravitationally. Dark energy has a very different role in the expansion. Combining both under the word dark does not make them the same physical substance.

A candidate dark particle needs more than invisibility. Its production history, abundance, stability and clustering must fit observations. If it moves too rapidly at the relevant epoch, it can erase structure that cold matter would help build.

The inferred gravitational component does not by itself identify its microscopic constituents. Searches for particle candidates and tests of gravitational alternatives address different parts of the question. The evidence for extra gravitational structure must be kept distinct from evidence for one particular particle.

Try 9. Using the rounded benchmark, what is the dark-to-ordinary matter ratio? Does that arithmetic identify a dark-matter particle?

Dark matter helps form gravitational structures. The expansion history points to another component with a different role. Its name resembles neither a particle identification nor an explanation of the early inflationary phase.

10 Is dark energy the same thing as inflation?

Before reading on. If two epochs both involve accelerated expansion, must the same microscopic field be responsible for both?

No. Similar large-scale behaviour does not uniquely identify the underlying cause. Inflation describes a proposed early accelerated phase. Dark energy names the component used to describe the much later acceleration within the usual cosmological model.

The evidence for late acceleration combines measurements of distances and expansion history, including supernovae, the microwave background and large-scale structure. Inferring a component involves the gravitational model and the interpretation of those observations. Naming it dark energy does not reveal its microscopic identity. Type Ia supernovae, a particular class of stellar explosion, are particularly useful: their light curves can be calibrated to estimate intrinsic luminosity, then compared with observed brightness to infer distance. That calibration and other astrophysical effects must be checked.

A cosmological constant is the simplest common description. Its energy density remains constant as the cosmic scale factor changes, and its negative pressure enters the gravitational equations. Other models allow an evolving component. Comparing their predictions requires data and assumptions, rather than the word dark.

Cold dark matter behaves differently. For a fixed noninteracting population, doubling every linear separation increases the volume eightfold and reduces its density by a factor of eight. A cosmological constant keeps the same density. The two cannot be interchanged merely because neither shines like an ordinary star.

The early and late accelerated phases would also have very different energy scales and histories. The early phase must end in a way compatible

with the hot universe. The late component must fit the measured recent history. A model connecting them needs to demonstrate both tasks.

There are such proposals, but the similarity of the expansion behaviour is not their proof. We should not treat the inflaton, dark energy and the quantum vacuum as three established names for one discovered substance.

This distinction prepares the vacuum-energy question. If quantum fields contribute to the gravitational source, why is the observed net effect so small? Writing cosmological constant into the model can fit a value without explaining it.

Try 10. If a noninteracting cold-matter population occupies eight times its former volume, what happens to its density? Would a cosmological constant follow the same rule?

A quantum vacuum is a field's reference state, not a statement that there are no fields or correlations. When gravity is included, a vacuum contribution can affect cosmic expansion. The problem is to account for its small net gravitational effect, not just to add a list of formal oscillator energies.

11 Why is the gravitational vacuum contribution so small?

Before reading on. If adding the calculated energies of many field modes gives an enormous total, does that raw sum automatically equal the observable cosmological constant?

Quantum fields have a vacuum state with structure and correlations. A mode is one component oscillation used to describe the field. Even a quantum oscillator in its lowest-energy state need not have zero energy; this is its zero-point contribution. Adding such contributions up to an imposed highest frequency, called a cutoff, can produce a very large energy density. The famous enormous discrepancy depends on which cutoff and comparison are used.

In a gravitational theory, the physical issue is the small net vacuum contribution to the stress-energy and cosmological expansion. The calculation combines parameters supplied in the model with quantum contributions. Renormalisation is the prescription for relating that combination to finite physical quantities. The separate bare parameters and corrections depend on the chosen prescription; the observable result must be consistently defined. A raw divergent sum is not by itself a directly observed energy density.

This qualification does not make the problem disappear. We still need to understand why the net gravitational effect is small and stable under the relevant physical contributions. Choosing a subtraction to fit one measured value is not an explanation of that stability.

The cosmic dilution rules help distinguish possible components. Double the scale factor. A fixed population of cold particles is spread through eight times the volume, so its energy density falls by eight. Freely travelling radiation also loses half the energy per photon through redshift, so its density falls by sixteen. A cosmological constant retains constant density.

The last behaviour corresponds to negative pressure in the cosmological conservation equation. It does not mean that ordinary heat in a larger box automatically has the required equation of state, meaning the relation between its pressure, energy density and other thermodynamic variables. A proposed microscopic component must derive its pressure and energy density.

A finite cutoff could remove one formal infinity while leaving the physical smallness problem untouched. The net contribution still needs a consistent gravitational interpretation and an explanation of why the different terms produce such a small observable result.

Try 11. When the scale factor doubles, by what factors do cold-matter and radiation energy densities fall in the stated noninteracting models?

Part four

Why the history has a direction

The expansion history describes what changed. The thermodynamic arrow of time asks why familiar changes occur so much more readily in one direction than the reverse. Hot material can be nearly uniform and still occupy a special gravitational state.

Entropy is a measure tied to microscopic possibilities under stated macroscopic constraints. Macroscopic means the large-scale description, such as a gas's volume and total energy; microscopic means the detailed particle or quantum-state arrangements compatible with it. Entropy is not simply a measure of visible mess. We will use small counting examples to explain that distinction, without pretending they calculate the entropy of the universe.

The questions return to the initial-condition problem. Explaining how an unusual early state evolves is not the same as explaining why that state was realised. Inflation and ordinary thermodynamics must not be credited with that further result without an argument.

12 Why was the early universe so orderly?

Before reading on. Can a nearly uniform hot gas be close to thermal equilibrium while its gravitational configuration is very special?

Yes. Ordered and disordered are not enough unless we say which degrees of freedom and constraints are included. A uniform gas in a box with negligible gravity can be near a high-entropy state. Gravity changes the possibilities.

In a gravitating system, matter can clump, release energy and eventually form compact objects and black holes. The early universe was remarkably smooth on large scales, with small variations that later grew. That smoothness is associated with a low gravitational-entropy

condition compared with the much larger range of later gravitational states.

There is no single simple local entropy density for an arbitrary gravitational field that settles all such comparisons. Black-hole entropy provides a particularly sharp case; the general cosmological question remains more subtle. The qualitative contrast should not be disguised as a completed microscopic count.

Ordinary counting already illustrates why the specification matters. Four independent fair bits have sixteen possible joint values. Four bits constrained to all agree have only two: all zero or all one. Either single bit looks random, yet the joint possibilities differ sharply.

That example is not an entropy calculation for the early universe. It shows why a local appearance of disorder need not reveal the full organisation of a joint state. In cosmology the relevant gravitational constraints and physical histories must also be included.

An account of the early state must specify both its thermal and gravitational properties. Saying that the plasma was hot and well mixed addresses one part of the account. It leaves the unusually smooth geometry and the allowed gravitational states to be explained.

The question is therefore a real boundary-condition problem. Explaining why entropy typically increases from a special state is not yet explaining why the special state occurred.

Try 12. How many joint values are available to four independent fair bits, and how many if they must all agree?

A low-entropy preparation gives a direction to familiar processes: heat spreads and broken cups do not spontaneously rebuild. Yet many microscopic equations allow a reversed evolution. The issue is how the prepared state and the number of compatible microscopic arrangements influence what normally happens.

13 Why does time have a direction?

Before reading on. If a microscopic evolution can be reversed in principle, should a broken cup routinely reassemble itself?

Reversible equations do not imply that all macroscopic histories are equally likely under a given preparation. A broken cup has scattered fragments, vibrations, air motion and heat correlated in complicated ways. Reassembling it by reversal would require the appropriate coordinated microscopic state, not merely a wish that time run backwards. Reversible here means that the complete dynamics can be inverted in principle. The usual reverse-the-motion picture also assumes time-reversal symmetry of the interactions being used; these are related but distinct ideas.

Thermodynamics groups many microscopic states into the same macroscopic description. A small set of organised preparations can evolve towards macrostates compatible with many more microscopic arrangements. This gives statistical content to entropy increase, with the low-entropy preparation doing essential work.

Four labelled objects provide a small counting illustration. If each can occupy the left or right half of a box, there are sixteen assignments. Only one puts all four on the specified left side. Six put two on each side. The balanced macro-description corresponds to more assignments. Real thermodynamic systems have enormously more degrees of freedom.

Records fit into that physical history. A memory becomes correlated with an event, and its stable operation uses an environment and resources. Reversing a display alone does not reverse the surrounding correlations. This is the same distinction encountered with the route marker in the first volume.

Landauer's principle relates logically irreversible erasure to thermodynamic cost under specified conditions. In the standard ideal model, an initially unbiased bit occupies two equal-energy memory states, has no usable correlations with another record, and is reset to a specified value while coupled to a thermal reservoir. A thermal reservoir is a large environment whose temperature barely changes during the operation. Returning the rest of the device to its initial condition has a minimum average heat cost to that reservoir, approached in a reversible limit. The qualifications matter; this is not a fixed heat pulse required in every individual trial. The physical memory, temperature and cycle must be identified. Counting every event as an erasure does not establish that cost.

Microscopic time-reversal violation in some particle processes is another topic. It does not replace the low-entropy boundary-condition problem of ordinary thermodynamics. A complete account must distinguish the symmetry of the equations from the statistical history of the prepared universe.

Try 13. In the four-object example, why is the two-on-each-side macrostate more numerous than all-on-the-left?

For checking

Answers to the Try questions

Try the questions before reading these answers. Compare the physical assumptions and the first step of the reasoning, not just the final number. A matching number can hide a different interpretation.

1. 0.3 unit. The fractional increase is the same for each separation, not the absolute increase.
2. A nucleus contains protons and neutrons; a neutral atom also needs bound electrons. The early plasma can contain nuclei while remaining ionised. Neither event establishes an absolute beginning.
3. A causal mechanism: interactions need time to carry energy or information between the relevant regions.
4. The length is multiplied by 1024. That illustrative arithmetic neither determines a particular early size nor identifies our observable patch with the whole universe.
5. Ten units per unit time. This is the sum of changes of distant separations on the chosen slices, not one locally measured object's velocity.
6. Four units remain. An actual model needs a specified field interaction, its dynamics and the evolving expansion history.
7. A limit rules out parameter choices requiring a larger signal. Models predicting a smaller signal may remain compatible with the same limit.
8. Examples include establishing the relevant field dynamics and initial regime, finding observable discriminants, and supplying a justified probability measure. Producing possible regions in a model does not observe them.
9. 5.2. The ratio compares fitted energy-density fractions; it does not identify a microscopic candidate.

- 10.** Its density falls by eight for the fixed cold-particle population. A cosmological constant retains its density.
- 11.** Cold-matter density falls by eight. Radiation density falls by sixteen because the volume grows eightfold and each photon energy halves.
- 12.** Sixteen for independent bits; two if all must agree. The joint constraint changes the counting.
- 13.** There are six choices of which two labelled objects go left, but only one assignment puts all four on the specified left.

For returning to

Words and sources

Words to return to

Observable universe

The region from which signals could have reached us by now in the specified cosmic history; not necessarily the whole universe.

Plasma

Matter with freely moving charged constituents, such as electrons and nuclei not bound into neutral atoms.

Comoving

Following the idealised cosmic flow. Comoving labels remain fixed while physical separations can change.

Scale factor

The function relating those labels to physical separations on cosmic-time slices.

Cosmic time

Time measured along the idealised average cosmic flow, using the shared cosmological history to relate clock readings.

Horizon

A limit on causal access defined by the spacetime history. Different horizons answer different questions.

Inflation

A proposed early accelerated expansion, requiring a driver, an ending and a subsequent hot history.

Inflaton

A field introduced to drive inflation in a specified class of models.

Reheating

Transfer from the inflation-driving sector into other excitations and a hot state; the mechanism is model dependent.

Perturbation

A small departure from a reference state, such as a density variation on an otherwise smooth background.

Dark matter

An inferred gravitating component that clusters, distinct from ordinary unlit baryonic matter.

Dark energy

A component used to account for late acceleration in a cosmological model; the name does not specify its microscopic origin.

Entropy

A physical measure related to the microscopic states or uncertainty compatible with a specified description.

Measure problem

The problem of assigning justified relative probabilities in a construction with indefinitely many regions or observers.

Sources and their scope

Tong's notes develop the expanding and hot universe, inflation and structure formation with more mathematics than is used here.[1] The book distinguishes the observational history from the proposed early dynamics even where a source uses more assertive introductory language.

The Planck parameter analysis supports the rounded composition benchmark used in the dark-matter question. It is a specified model and dataset, not an assertion that every later fit must agree exactly.[2] Planck's inflation analysis and the BICEP/Keck paper provide dated examples of how primordial predictions are constrained.[3, 4] Neither is cited as a direct observation of an inflaton.

Guth's paper explains mechanisms for eternal inflation and difficulties with probabilities. It is a theoretical account, not an observation of other universes.[5] Landauer and Bennett concern the physical cost of information processing under stated conditions; the simple bit examples are not calculations of cosmological entropy.[6, 7]

References

- [1] D. Tong, *Cosmology*, lecture notes, especially The Expanding Universe, The Hot Universe and Structure Formation. [Author's notes](#).
- [2] Planck Collaboration, N. Aghanim et al., Planck 2018 results. VI. Cosmological parameters, *Astronomy & Astrophysics* **641**, A6 (2020). doi:10.1051/0004-6361/201833910.
- [3] Planck Collaboration, Y. Akrami et al., Planck 2018 results. X. Constraints on inflation, *Astronomy & Astrophysics* **641**, A10 (2020). doi:10.1051/0004-6361/201833887. [Open manuscript](#).
- [4] BICEP/Keck Collaboration, P. A. R. Ade et al., BICEP/Keck XIII: Improved Constraints on Primordial Gravitational Waves using Planck, WMAP, and BICEP/Keck Observations through the 2018 Observing Season, *Physical Review Letters* **127**, 151301 (2021). doi:10.1103/PhysRevLett.127.151301. [Open manuscript](#).
- [5] A. H. Guth, Inflation and Eternal Inflation (2000). [Author's manuscript](#).
- [6] R. Landauer, Irreversibility and Heat Generation in the Computing Process, *IBM Journal of Research and Development* **5**, 183–191 (1961). doi:10.1147/rd.53.0183.
- [7] C. H. Bennett, The thermodynamics of computation—a review, *International Journal of Theoretical Physics* **21**, 905–940 (1982). doi:10.1007/BF02084158.
- [8] European Space Agency, The cosmic microwave background and inflation. [Planck science explanation](#).