

Early dynamical world models: A historical review

Helge Kragh

Department for Studies of Science and Science Education,
University of Århus,
Bygning 1110, C.F. Møllers Allé 8, 8000 Århus, Denmark
email: helge.kragh@ivs.au.dk

Abstract. Models of the universe, in the sense of solutions to the cosmological field equations, took their start in 1917 with Einstein's closed universe. During the next two decades they were developed to comprise evolving models, some of them cyclic and some of them with a definite age. The history of this development, as it occurred up to the mid 1930s, is reviewed. It is argued that in 1930-31, cosmology experienced a kind of paradigm shift.

Keywords. Expanding universe, cyclic models, history of cosmology, Friedmann, Einstein

1. Introduction

Modern cosmology has its roots in the first decades of the 20th century, when a combination of observational and theoretical advances led to a picture of the universe entirely different from the one that astronomers were used to entertain. At about 1910, many astronomers were convinced that the material or “*sidereal*” universe, principally made up of stars, was restricted to the Milky Way and its most immediate surroundings. Outside this system, there was nothing except empty space and perhaps endless ether. Models of the Milky Way varied, but experts such as Jacobus Kapteyn, Karl Schwarzschild and Hugo von Seeliger largely agreed that the Milky Way was a huge disk which in the galactic plane spanned over a distance of about 40,000 light years, with the Sun placed in the neighbourhood of the centre. However, there was no consensus about either the structure or size of the stellar universe, nor could the rival *island universe* theory be ruled out (Berendzen, Hart & Seeley 1984).

As far as physical theory was concerned, Newton's celebrated law of gravitation was taken to be the obvious and solid basis not only of astronomy, but also of cosmology. Apparent anomalies turned up in both areas, but without seriously questioning the authority of the law. One problem, going back to 1859, was the anomalous motion of Mercury's perihelion; another one was Seeliger's demonstration in 1895 of the so-called gravitational paradox, that an infinite (and Euclidean) universe uniformly populated with stars could not be brought into agreement with Newton's law. A few astronomers, and particularly Schwarzschild in 1900, contemplated that physical space might be curved in accordance with the theories of Riemann and other mathematicians, but such ideas remained speculative and peripheral until General Relativity changed the game of cosmological thinking.

Newtonian gravitation was not the only fundamental law of physics that was applied to the universe at large in pre-Einsteinian cosmology. During the last third of the 19th century this was also the case with the laws of thermodynamics. Several scientists and philosophers argued from the second law, not only that the universe would end in a *heat death* of maximum entropy, but also that it must have had a beginning in time (Kragh 2008). Because, if the world had existed in an eternity, wouldn't the entropy already have

reached its maximum? Although the controversial question of cosmic thermodynamics received little attention among physicists, and even less among astronomers, it is an episode worth recalling. It was the first attempt to argue from a fundamental law of physics that the entire universe is in a state of irreversible change.

This essay deals with selected areas of what in retrospect can be recognized as early mainstream theoretical cosmology, namely relativistic models in the period from about 1920 to the mid 1930s. It was during this period that the framework of cosmological models, as we know them today, was founded; and it was during the same period that revolutionary concepts such as the expanding universe and the big bang universe were introduced. Understandably, this phase in the history of cosmology has been described in detail by historians of science (e.g., North 1965; Kerzberg 1989). We may tend to think that theoretical cosmology in the period was identical to relativistic cosmology, but this would be a mistake that reads modern knowledge into the past. Although relativistic models obtained a kind of mainstream status, they had to compete with models based on classical physics, some of them qualitative and others quantitative. The most important and sophisticated of these non-relativistic models was undoubtedly the one that E. Arthur Milne started developing in 1932 and which for a decade or so attracted much interest. However, for the sake of focus and brevity I shall restrict my essay to models based on Einstein's field equations.

2. From the static to the expanding universe

When Einstein wrote his pioneering paper on *Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie* in early 1917, he wanted to formulate a model of the world that was not only consistent with his new covariant theory of gravitation, but also agreed qualitatively with contemporary astronomical knowledge. This knowledge indicated that the proper motions of the stars were small and that the universe as a whole was therefore static. The result of Einstein's heroic efforts was an unbounded, but spatially closed and static universe filled uniformly with a finite amount of highly diluted pressureless matter. To obtain this result he had to introduce the famous cosmological term as a *generalizing modification of the gravitational field equations*. The corresponding constant played two crucial roles in Einstein's world model: First, it provided a hypothetical cosmic repulsion that kept the universe from collapsing as a result of the gravitational attraction; second, it expressed the mean density of the matter in the closed universe, namely by the simple relation $2\Lambda = \kappa\rho$, where κ is Einstein's constant of gravitation. Because Λ was also a measure of the curvature, it served as the link that connected the density of matter in the universe with the constant radius of curvature.

It is sometimes stated that Einstein missed the expansion of the universe because of his introduction of the cosmological constant as a means to keep the universe static. If he had only had the courage to believe in his original field equations, he might have predicted the expansion of the universe more than a decade before it was found observationally. However, this is plainly a misunderstanding, as the cosmological term in no way prevents space to expand. After all, the first expanding models, as constructed by Friedmann, Lemaître and Eddington, made use of the constant.

Einstein was convinced that his solution to the cosmological problem was unique in the sense that it was the only solution to the field equations that made physical sense. When he was confronted with Willem de Sitter's alternative solution later in 1917, he therefore received it with distrust. At that time Einstein was still under the spell of Mach's principle, which in Einstein's view precluded such a monstrous thing as a universe without matter. He continued to think of de Sitter's solution as artificial and unphysical until the

early 1930s, when the expansion of space changed the whole framework of theoretical cosmology.

During the period 1917–1930, the scene of relativistic cosmology was dominated by the two static models of Einstein and de Sitter, at the time generally known as the *A model* and the *B model*, respectively. Einstein's closed *A model* was relatively easy to visualise, for which reason –and also because of Einstein's unparalleled fame– it became well known in the public sphere. As far as philosophers and other lay persons were concerned, the relativistic universe was the same as Einstein's universe. Whereas the more abstract and mathematically sophisticated *B model* of de Sitter was absent from the public discussions, it attracted much interest among mathematicians and mathematically-inclined physicists and astronomers such as F. Klein, H. Weyl, C. Lanczos, H.P. Robertson and A.S. Eddington.

One reason why experts in cosmology found the empty de Sitter solution to be of great interest was that it offered a much needed connection to observational astronomy. After V. Melvin Slipher in 1912 had found that the spectral lines of the Andromeda nebula were shifted towards the blue, his further work showed that the light from the spiral nebulae was redshifted rather than blueshifted, which he interpreted as the result of a Doppler effect. This interpretation was generally accepted, and by the early 1920s astronomers agreed that most spiral nebulae exhibited large radial velocities, some of them surprisingly large. Although static, de Sitter's theory (but not Einstein's) included a cosmological redshift that might possibly explain Slipher's data, and it also suggested that there was a definite relationship between redshifts and distances. However, it was difficult to justify such a redshift-distance relation observationally, in particular because the distances to the nebulae were only known with great uncertainty or not known at all. This unsatisfactory situation began to change after Edwin Hubble in early 1925 announced his discovery of cepheid variables in the Andromeda and thereby opened up for a method to determine the distances to other spiral nebulae as well.

This work in observational astronomy culminated in the early spring of 1929, when Hubble's famous paper on “*A Relation Between Distance and Radial Velocity Among Extra-Galactic Nebulae*” appeared in the *Proceedings of the US National Academy of Sciences*. The essence of Hubble's paper was data that correlated the redshifts or *apparent velocities* of 46 spiral nebulae with their distances, thereby providing empirical evidence for the relationship $v = Hr$, where H came to be known as the Hubble constant or parameter†. The value of the constant was found to be approximately $500 \text{ km s}^{-1} \text{ Mpc}^{-1}$. From a historical point of view it is important to realize that Hubble did not conclude that the universe is expanding or that the observed redshifts were solely caused by the recession of the nebulae. A somewhat timid empiricist, he wanted to keep to his data and leave their interpretation to others. He did not oppose the theory of the expanding universe, but throughout his life he felt that its empirical basis was not quite sufficient. Strangely from a modern point of view, Hubble's work could even be used as an argument against the expansion of the universe. Commenting on Hubble's monograph, *The Realm of the Nebulae*, his colleague Nicholas Mayall wrote him in 1937 how pleased some astronomers were *that your interpretation of the nature of the red-shifts casts doubt upon the validity of theories of the expanding universe* (Kragh & Smith 2003).

What we today justifiably regard as a revolutionary contribution to cosmology, was at the time seen in a different light. In the year of 1929, Hubble's work was considered to be important, but not as observational proof that the universe is in a state of expansion. It took more than observations to effect a change in the world view from a static to an

† In fact, no such formula appeared in Hubble's paper.

expanding universe. Einstein was no exception. In 1929 he wrote an article on *Space-Time* for the *Encyclopedia Britannica* in which there was no trace at all of the possibility that the universe might be expanding.

3. Monotonously expanding world models

The birth of the expanding universe went through four phases. First, in 1922 Alexander Friedmann in the Soviet Union predicted the possibility, corresponding to a particular class of solutions to the cosmological field equations. Second, in 1927 Georges Lemaître in Belgium argued explicitly in favour of the expanding universe by combining relativity theory with astronomical data. Third, in 1929 Hubble in the United States of America provided the first evidence that the galaxies are receding from the Earth in agreement with $v = Hr$. All this was not enough to make the expansion of the universe a reality accepted or even considered by the astronomers. Then, fourth and finally, during the spring of 1930 the earlier works of Friedmann and Lemaître were reconsidered – *rediscovered* may be a better word – in connection with Hubble’s measurements, with the result that leading astronomers and physicists changed from the static to the expanding paradigm.

In his paper of 1922, Friedmann found a class of closed *monotonic worlds of the first class* characterised by a scale factor $R(t)$ that continued to increase from an original state formally corresponding to $R = 0$. Not only did he realize the possibility of an expanding universe, he also introduced for the first time in a scientific context the notion of the beginning, and hence the age, of the world. He even stated that this age was of the order of *tens of billions of ordinary years*, although it is unclear how he arrived at the result. Important as Friedmann’s analysis was, it was more a mathematical analysis of the solutions to Einstein’s field equations than an attempt to formulate a cosmological model of the universe. He seems to have been uninterested in the physical and astronomical consequences of his work, which he chose to neglect. There is no reason to assume that he thought that the one and only real universe was in a state of expansion.

Unaware of Friedmann’s paper, in 1927 Lemaître obtained many of the same results, including the differential equations for $R(t)$ often known as the Friedmann equations. However, in striking contrast to his unknown Russian colleague, Lemaître adopted a much more physical-astronomical approach, focusing on the expanding solution that appeared to agree best with the redshift data. This solution, he argued, was a closed universe that started expanding from an Einstein state of radius of curvature about 270 Mpc and which eventually would approach the empty de Sitter space. He emphasized that the redshifts were caused by the expansion of space, and demonstrated that this cosmological interpretation led to an approximately linear velocity-distance relation. For the constant of proportionality – what later would be known as the Hubble constant – he estimated a value of about $600 \text{ km s}^{-1} \text{ Mpc}^{-1}$. Remarkably, this was two years before Hubble found his somewhat smaller value. Lemaître’s theory, published in French in a rather obscure Belgian journal, attracted absolutely no attention. It was even more thoroughly neglected than Friedmann’s work.

Mainly due to the efforts of Eddington and de Sitter, the fate of the expanding universe changed drastically through the year of 1930. Eddington adopted the model of Lemaître (who was his former research student) and developed it into the slightly revised version known as the Lemaître-Eddington model. De Sitter was no less enthusiastic. In a letter of April 1930, he told Harlow Shapley that he considered Lemaître’s model to be “*the true solution, or at least a possible solution, which must be somewhere near the truth*” (Smith 1982). It took only a few years until the theory of the expanding universe was accepted by a majority of astronomers and also disseminated to a broader public by

means of newspaper articles and popular books. The shift was not complete, though, for throughout the 1930s several astronomers suggested alternative interpretations of the galactic redshifts.

The Lemaître-Eddington universe was continually expanding, but it included no singular state and could not be assigned a definite age. Thus, it was not a Big Bang model of the type that Friedmann had discussed as a mathematical possibility in his paper of 1922. From a more physical and realistic point of view the Big Bang, or something like it, was introduced by Lemaître in 1931, in the form of what he referred to as the *primeval atom* or *fireworks* hypothesis. He first suggested the idea, quite briefly and qualitatively, in a letter to *Nature* and then, in greater details, in a paper in the *Revue des Questions Scientifiques* (Lambert 2000).

Lemaître imagined that all matter in the universe had once – at the beginning of time – been concentrated in a kind of huge atomic nucleus, highly radioactive and with a density of about $10^{15} \text{ g cm}^{-3}$. This original and primitive state of matter was completely uniform and therefore devoid of physical properties. It was, in a sense, more metaphysical than physical. Following the initial explosion, the universe expanded rapidly until it reached a *stagnation phase*, after which it would enter a phase of accelerated expansion. This picture depended crucially on the assumption of a positive cosmological constant, a quantity Lemaître was a great advocate of. From the perspective of the 1930s the theory might look like nothing but a fairy tale, but Lemaître argued that it had testable consequences. The charged particles of the cosmic rays, he suggested, were fossils of the original explosion.

During the 1930s, the Lemaître-Eddington model was probably the most popular of the cosmological models based on the theory of General Relativity, whereas the explosive theory of Lemaître was either rejected or ignored. Another model, which came to play an important rôle in the later development, was constructed jointly by Einstein and de Sitter in 1932. Investigating a simple model with zero pressure, zero cosmological constant, and zero curvature, the two pioneering cosmologists showed that the matter density had to be *critical*, given by $\rho = 3H^2/8\pi G$. For this value of the density, gravitation precisely balances the cosmic expansion, meaning that the universe will continue to expand for ever (with R varying as $t^{2/3}$), but at a continually slower rate.

Although the Einstein-de Sitter model clearly is a Big Bang model with a definite age given by $2/3$ times the Hubble time, these features were not mentioned at all in the brief paper written by Einstein and de Sitter. Obviously, these were characteristics of the model that neither of the authors were comfortable with and which they preferred to ignore. Another feature that they refrained from commenting was the flat space of the model and its consequence, that the universe was spatially and materially infinite.

The continually expanding models mentioned so far were not the only ones that appeared in the decade. Milne proposed in 1932 his own version of the expanding universe, but this model was not founded on General Relativity. Yet another unorthodox model of the expanding universe was suggested by Paul Dirac in 1938, based on the assumption of a gravitational constant decreasing in time. Although both models belonged to the Big Bang class, none of them corresponded to solutions of Einstein's cosmological equations. In the case of Milne, the age of the universe was equal to the Hubble time, whereas Dirac's model predicted an age of only $1/3$ of this quantity.

4. Cyclic models from Friedmann to Tolman

The general idea of a cyclic or oscillating universe goes far back and time, and it played an important rôle in the controversy over the heat death in the late 19th century.

However, it was only in the 1920s, with the advent of relativistic cosmology, that the idea could be discussed from a strictly scientific point of view. In his paper of 1922, Friedmann introduced the notion of a closed *periodic world* with a period approximately given by $T = \kappa M/6\pi c$. The radius of curvature of the cyclic universe evolved from a singular state $R(0) = 0$ over $R = R_{max}$ to a final singularity $R(T/2) = 0$. Although Friedmann was aware that an endless series of cycles could not be justified by Einstein's field equations, he found such a model to be philosophically fascinating. In a semi-popular book of 1923 he discussed the idea, which *brings to mind what mythology has to say about cycles of existence* (Friedmann 2006).

The next one to discuss a cyclic universe within the framework of relativistic cosmology was a Japanese physicist, Tokio Takeuchi, who in September 1930 read a paper on the subject to the *Physico-Mathematical Society of Japan*. Apparently he was at the time unaware of Friedmann's earlier work. Takeuchi's paper, published in a not widely known Japanese journal, included the novelty of an endless series of cycles. By means of a complicated cyclic line element he succeeded to construct an eternally oscillating world model without singularities. Philosophically attractive as the model might be considered, it was clearly ad hoc and attracted very little attention.

Whereas Takeuchi's work on the cyclic universe largely went unnoticed, this was not the case with Einstein's slightly later contribution, dating from about the same time as Lemaître suggested his idea of an exploding universe. Einstein's paper of 1931 on the cyclic universe did not go much beyond Friedmann's earlier treatment, but it is worth noticing for at least three reasons. First, it marked Einstein's public abandonment of the cosmological constant, a quantity he had long been dissatisfied with. Second, he used the opportunity to formulate in explicit terms a version of what soon would be known as the cosmological principle. Third, Einstein's model belonged to the Big Bang (as well as Big Crunch) class, although he did not mention this feature specifically. He did, however, draw attention to the time-scale difficulty, the much discussed problem that the age of the model universe seemed to be much smaller than the age of the stars.

If cyclic models held no special significance for Einstein, they were taken much more seriously by Richard Tolman at Caltech. Starting in 1931, over the next couple of years he investigated this kind of cosmological model in great detail, introducing for the first time considerations of relativistic thermodynamics. He collected much of this work in his classical textbook of 1934, *Relativity, Thermodynamics and Cosmology*. Among the important conclusions of Tolman's investigation was that the heat death could not be ascribed any validity: within the framework of relativistic thermodynamics, the entropy of the universe would not increase irreversibly from a minimum to a maximum value. Tolman further proved that it was possible for a closed universe with $\Lambda \leq 0$ to undergo a continual series of expansions and contractions, only would the cycles not be identical. Although *strictly periodic solutions* could not be justified mathematically, he argued that such models of an ever-oscillating universe were allowed physically. Indeed, in his book of 1934 he boldly suggested that they were *physically inevitably necessary* (Tolman 1987).

Of course, much more could be said about the development of this class of cosmological models, which has continued to attract interest up to the present. For a detailed examination of the history of cyclic world models I refer to a forthcoming essay (Kragh 2011). Many of the early works on cosmological models, including most of those mentioned here, have been reprinted and translated into English (Lang & Gingerich 1979; Bernstein & Feinberg 1986). Unfortunately this is not yet the case with Einstein's 1931 paper on the cyclic world, written in German.

References

- Berendzen, R., Hart, R., & Seeley, D. 1984, *Man discovers the galaxies* (New York: Science History Publications)
- Bernstein, J. & Feinberg, G. 1986, *Cosmological constants* (New York: Columbia University Press)
- Friedmann, A. 2006, *Die Welt als Raum und Zeit* (Frankfurt am Main: Harri Deutsch)
- Kerzberg, P. 1989, *The invented universe* (Oxford: Clarendon Press)
- Kragh, H. 2008, *Entropic creation: Religious Contexts of Thermodynamics and Cosmology* (Aldershot: Ashgate)
- Kragh, H. 2011, in *Beyond Einstein*, Einstein Studies 12, D. Rowe (ed) (Basel: Birkhäuser), in press
- Kragh, H. & Smith, R. 2003, *Hist. Sci.*, 41, 141
- Lambert, D. 2000, *Un atome d'univers: La vie et l'œuvre de George Lemaître* (Bruxelles: Éditions Lessius)
- Lang, K.R. & Gingerich, O. 1979, *Source book in astronomy and astrophysics 1900-1975* (Cambridge: Harvard University Press)
- North, J. 1965, *The measure of the universe* (Oxford: Oxford University Press)
- Smith, R. 1982, *The expanding universe* (Cambridge: Cambridge University Press)
- Tolman, R.C. 1987, *Relativity, thermodynamics and cosmology* (New York: Dover)