Thermodynamic solution of the homogeneity, isotropy and flatness puzzles
(and a clue to the cosmological constant)

Latham Boyle\(^1\) and Neil Turok\(^{1,2}\)

\(^1\)Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada, N2L 2Y5
\(^2\)Higgs Centre for Theoretical Physics, University of Edinburgh, Edinburgh, Scotland, EH8 9YL

(Dated: October 2022)

We obtain the analytic solution of the Friedmann equation for fully realistic cosmologies including radiation, non-relativistic matter, a cosmological constant \(\lambda\) and arbitrary spatial curvature \(\kappa\).

The general solution for the scale factor \(a(\tau)\), with \(\tau\) the conformal time, is an elliptic function, meromorphic and doubly periodic in the complex \(\tau\)-plane, with one period along the real \(\tau\)-axis, and the other along the imaginary \(\tau\)-axis. The periodicity in imaginary time allows us to compute the thermodynamic temperature and entropy of such spacetimes, just as Gibbons and Hawking did for black holes and the de Sitter universe. The gravitational entropy favors universes like our own which are spatially flat, homogeneous, and isotropic, with a small positive cosmological constant.

INTRODUCTION

Soon after the discovery of black hole thermodynamics\([1-3]\), Gibbons and Hawking\([5]\) showed that one could elegantly compute the temperature and entropy of a general (charged, spinning) black hole, and of de Sitter (dS) spacetime, by noticing that these spacetimes are periodic in the imaginary time direction (see also the preceding work of Gibbons and Perry\([6,7]\)).

In this paper, we find the exact solution of the Friedmann equation for a general FRW universe including arbitrary amounts of radiation, non-relativistic matter (including baryons and dark matter), a cosmological constant and spatial curvature\([8,9]\). Noticing that the scale factor \(a(\tau)\) is periodic in the imaginary \(\tau\) direction, we are able to compute the corresponding temperature and gravitational entropy.

Remarkably, the entropy obtained in this way favors universes like our own: homogeneous, isotropic and spatially flat, with small positive cosmological constant.

In earlier papers\([10,11]\), we obtained similar results in the context of a simplified cosmology with radiation, a cosmological constant \(\lambda\), and spatial curvature \(\kappa\) but without non-relativistic matter. Our new result strengthens the case that we have been developing\([10-15]\) for a simpler theory of the universe, not requiring inflation.

GENERAL SOLUTION FOR \(a(\tau)\)

The action for Einstein gravity coupled to matter is

\[
S = \int d^4x \sqrt{-g} \left( \frac{R}{2} - \lambda + L_{\text{matter}} \right)
\]

where \(\lambda\) is the dark energy (or cosmological constant). We shall use Planck units with \(\hbar = c = k_B = 8\pi G_N = 1\) throughout. Now take the FRW line element

\[
ds^2 = a^2(\tau)(-n^2d\tau^2 + \gamma_{ij}dx^idx^j)
\]

where \(n\) is the lapse and \(\gamma_{ij}\) is the metric on a maximally-symmetric 3-space of constant curvature \(\kappa\); and take the matter to consist of radiation with energy density \(r/a^4\) and non-relativistic matter (baryons and dark matter) with energy \(\mu/a^2\), where \(r\) and \(\mu\) are positive constants. Then the action becomes

\[
S = V_3 \int d\tau \left[ -3\frac{\dot{a}^2}{n} + nV(a) \right]
\]

where \(V_3\) is the comoving spatial volume\([28]\), \(\dot{a} = da/d\tau\), and

\[
V(a) = -\lambda a^4 + 3\kappa a^2 - \mu a - r
\]

Varying with respect to \(a\) (and then choosing the gauge \(n = 1\)) yields the Friedmann equation

\[
3\ddot{a} + V'(a) = 0
\]

while varying with respect to \(a\) (and again taking \(n = 1\)) yields the acceleration equation

\[
6\ddot{a} + V''(a) = 0.
\]

First consider the critical “Einstein static universe” (ESU) solutions, with constant scale factor \(a_{\text{esu}}\): these require \(\kappa > 0\) and \(\lambda > 0\), with the parameters related by

\[
\frac{2\kappa^3}{\mu^2} = \frac{(1+8x)(1+\frac{8}{3}x)\sqrt{1+\frac{32}{3}x}}{1+3\sqrt{1+\frac{32}{3}x}} (x \equiv \frac{\kappa}{\mu}).
\]

Eq.\([7]\) defines an important boundary between different dynamical phases. If the lhs of Eq.\([7]\) is greater than the rhs (which can only happen when \(\kappa\) and \(\lambda\) are both positive), we call it a “turnaround” universe: its curvature \(\kappa\) is sufficiently positive to cause a non-singular reversal from expansion to re-contraction, or contraction to re-expansion. Otherwise (i.e., if \(\lambda > 0\) and \(\kappa < \kappa_{\text{esu}}\), the positive, critical value set by Eq.\([7]\)), the universe expands monotonically from the bang \((a = 0)\) to the dS-like boundary \((a \to \infty)\), or the reverse.
To find the general solutions for \(a(\tau)\), first re-arrange the Friedmann equation \([5]\) as

\[
da/\sqrt{-V(a)} = d\tau/\sqrt{3}.
\]

(8)

Then, express \(V(a)\) in terms of its four roots as

\[
V(a) = -\lambda \prod_{i=1}^{4} (a - a_i),
\]

(9)

where \(\{a_1, a_2, a_3, a_4\} \equiv \{a_+, a_+, a_-, a_-\}\) and

\[
a_{\pm} = \frac{1}{2} \left[ (e_{\pm})^{\pm} - \sqrt{-(e_{\pm})^{2} - 2\mu/(e_{\pm}) + 6\kappa} \right],
\]

(10)

where we define \(e_{\pm} \equiv \pm(2\kappa + z - P/z)^{1/2}\), with \(z = (-Q + \sqrt{Q^2 + P^2})^{1/3}\), \(P = -\frac{4}{3}\kappa^2 - \kappa^2\), and \(Q = -\frac{2}{3}\kappa^2 + \frac{2}{3}\kappa^2\); and on \(a_{\pm}\), the first \(\pm\) subscript refers to \(e_{\pm}\) while the second \(\pm\) subscript is the sign in front of the square root in \((10)\). Now we define \(a_{ij} \equiv a_i - a_j\),

\[
\sin \varphi \equiv \sqrt{(a_{13}(a_2 - a))/(a_{23}(a_1 - a))},
\]

\[
m = \frac{a_{23}a_{14}}{a_{13}a_{24}},
\]

(11a)

\[
\zeta = \frac{1}{2} \left( \frac{\lambda a_{13}a_{24}}{2} \right)^{1/2}.
\]

(11b)

Then we integrate \([5]\), (e.g., using Eq. 5 in Sec. 3.147 of Ref. \([14]\)) to obtain

\[
F(\varphi, m) = \zeta \tau + C,
\]

where \(F(\varphi, m)\) is the elliptic integral of the first kind \([17]\), and \(C\) is a constant. We invert this to obtain the Jacobi amplitude \(\varphi = \text{am}(\zeta \tau + C, m)\). Finally, taking sin of both sides and using \(\sin(\text{am}(u, m)) = \text{sn}(u, m)\), where \(\text{sn}\) is a Jacobi elliptic function \([17]\), we obtain the general solution

\[
a(\tau) = \frac{a_{23}a_{14} - a_{13}a_{24} \text{sn}^2(\zeta \tau + C, m)}{a_{31} - a_{32} \text{sn}^2(\zeta \tau + C, m)}.
\]

(12)

This solution is valid in general, even when the roots \(a_i\) are complex. Assuming \(\lambda > 0, C = 0\) for a “turnaround” universe and \(C = \frac{1}{2} K(1 - m)\) otherwise, with \(K(m)\) the complete elliptic integral of the first kind \([17]\).

The general solution for \(a(\tau)\) is thus an elliptic function – meromorphic and doubly periodic in the complex \(\tau\)-plane, with one period oriented along the real \(\tau\)-axis and the other oriented along the imaginary \(\tau\)-axis. As shown in Fig. \([1]\) these periods form the two sides of a rectangle, containing two poles of opposite residue; the solution \(a(\tau)\) may be regarded as the rectangular tiling of the complex \(\tau\)-plane by copies of this rectangle. Alternatively, we can think of \(a(\tau)\) as a meromorphic function on the torus formed by identifying the opposite edges of the rectangle (with the open circles indicating the two poles). The imaginary \(\tau\) period is

\[
\Delta \tau = \pm 2i K(1 - m)/\zeta.
\]

(13)

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig1.png}
\caption{Top: The real and imaginary parts of \(a(\tau)\), with (black) rectangular fundamental domain. (For this example, \(r = \frac{1}{3}, \mu = \frac{1}{2}, \kappa = 1, \lambda = 1\): a “turnaround” cosmology.) The thick horizontal blue (resp. green) line shows the Lorentzian solution along which the universe expands from a bang and re-collapses to a crunch, (resp. collapses from past dS infinity and re-expands to future dS infinity). The vertical red line is the integration contour used to compute the gravitational entropy \(S_g\) and temperature \(T_g\). Bottom: We can regard \(a(\tau)\) as a meromorphic function with two poles, on the torus obtained by identifying opposite edges of the fundamental domain. The red contour may be deformed to any topologically equivalent contour.}
\end{figure}

\section*{Explicit Formulae for Gravitational Entropy \(S_g\) and Temperature \(T_g\)}

We can express the quantum amplitude to go from initial state \(|i\rangle\) to final state \(|f\rangle\) in time \(\Delta t\) in two ways

\[
\langle f | e^{-i H \Delta t} | i \rangle = \int \mathcal{D}[\varphi] \exp(i S[\varphi]),
\]

(14)

where the right-hand side is the usual path integral over all interpolating configurations \(\varphi(x)\). Now, by a standard argument \([5]\), if we identify the initial and final states \(|f\rangle = |i\rangle\), sum over all states \(|i\rangle\), and Wick rotate to Euclidean time \(\Delta t = -i \Delta t_E = -i/T\), \((14)\) becomes

\[
\sum_i \langle i | e^{-H/T} | i \rangle = \int \mathcal{D}[\varphi] \exp(-S_E[\varphi]),
\]

(15)

with \(-S_E[\varphi] \equiv i S[\varphi]\), so the lhs – the partition function at temperature \(T\) – may be expressed as the the amplitude to propagate along the Euclidean time direction \(t_E\) by an amount \(\Delta t_E = 1/T\) and return to the initial state, while the rhs is the path integral over configurations periodic in Euclidean time, with period \(\Delta t_E\).

For a cosmological spacetime, the full (gravity plus matter) Hamiltonian \(\hat{H}\) vanishes due to time reparameterization invariance, so the lhs of \((15)\) just evaluates to
the total number of states or, in other words, exp(Sg), where Sg is the gravitational entropy. On the other hand, just as in the black hole and de Sitter spacetimes \[5\], the rhs may be evaluated in the semiclassical approximation, yielding \(\exp(-S_E) = \exp(iS)\), where S is the action for the classical spacetime, evaluated over a full period in imaginary time. We conclude that

\[ S_g = iV_3 \int_0^{\Delta \tau} d\tau \left[-3a^2 + V(a)\right], \tag{16} \]

where the sign of \(\Delta \tau\) in \[13\] must be chosen so that \(S_g > 0\), since \(\exp(S_g)\) – the number of microstates corresponding to a given macroscopic spacetime – is \(\geq 1\).

The integration contour is depicted by the red curve in Fig.\[1\] as shown there, it is a non-contractible loop winding once around the torus. Cauchy’s theorem guarantees that the integral \(S_g\) is invariant under contour deformations which avoid the two poles. Of all the topologically-distinct equivalence classes of contours we can choose, the physically correct contour is the one that sticks to the region where the scale factor \(a\) (or more precisely its real part) remains large and positive (since the formula for the energy density of non-relativistic matter, \(\mu/a^3\), may be only trusted as long as \(a\) is neither too close to the bang nor negative). When \(\lambda > 0\), there is always precisely one such equivalence class.

To evaluate \[16\], first imagine we are in the “turnaround” case, so that \(\kappa > 0\), \(\lambda > 0\) and the roots \(\{a_1, a_2, a_3, a_4\}\) are real, with \(a_1 > a_2 > a_3 > a_4\). Using Eq. \[5\], we can rewrite \[16\] as

\[ S_g = 4\sqrt{3}V_3 \int_{a_2}^{a_1} da \sqrt{V(a)}. \tag{17} \]

This integral may be evaluated by following the algorithm explained in Sections 13.1-13.8 of Ref. \[18\]. The result is

\[ S_g = \frac{\sqrt{3\lambda}V_3a_{23}}{2\sqrt{a_{13}a_{24}}} \left[ C_K K(\bar{m}) + C_E E(\bar{m}) + C_{\Pi}\Pi\left(\frac{a_{12}}{a_{13}}, \bar{m}\right) \right] \] \tag{18} \]

where \(K(\bar{m})\), \(E(\bar{m})\) and \(\Pi(n, \bar{m})\) are the complete elliptic integrals of the 1st, 2nd and 3rd kinds, respectively, and

\[ C_K = a_{13}(a_{4}^{2} - a_{12}^{2} - \frac{4}{3}a_{14}a_{24}), \tag{19a} \]
\[ C_E = 8\kappa\frac{a_{13}a_{24}}{a_{23}}, \tag{19b} \]
\[ C_{\Pi} = (a_{14}^{2} - a_{24}^{2})(a_{14} - a_{23}), \tag{19c} \]
\[ \bar{m} = \frac{a_{12}a_{34}}{a_{13}a_{24}}. \tag{19d} \]

Although we derived this result for the “turnaround” case where the roots \(a_1 > a_2 > a_3 > a_4\) are all real, the resulting formula \[18\] is valid in general, even when some or all of the roots are complex. Our result for \(S_g\) is plotted in Fig.\[2\] (In this paper, the horizontal axes are always labeled by \(r, \mu \downarrow 0\) with \(\lambda\) and \(\kappa\) positive, we recover the standard de Sitter entropy.)

As two checks on our formula, in the limit \(r, \mu \downarrow 0\) with \(\lambda\) and \(\kappa\) positive, we recover the standard de Sitter entropy \(S_\lambda = 24\pi^2/\lambda\) is the standard de Sitter entropy.)

The physical time \(t = \int_0^\tau a(\tau)d\tau\) has three periods (one real and two imaginary) since, from Fig.\[1\], it depends on how many times the integration contour wraps around, or through, the hole in the torus, and also how it wraps around the poles in \(a(\tau)\). The first imaginary period, \(\Delta t = \int_0^\tau a(\tau)d\tau\), evaluated along the same contour used to compute \(S_g\), yields a global gravitational
“temperature,” illustrated in Fig. 3

\[ T_g \equiv i/\Delta t. \]  \hspace{1cm} (20)

Since the Euclidean geometry is not invariant under imaginary time translations, quantum field correlators defined in the Euclidean region will not be time translation invariant when they are continued to real time. As emphasized in Ref. [11], we are describing an out of equilibrium ensemble. Hence, \( T_g \) is a global quantity which is not locally measurable. The second imaginary period, \( \Delta \lambda = 2\pi i R \), where \( R = \pm \sqrt{3/\lambda} \) is the residue at the pole of \( a(\tau) \), yields the standard de Sitter temperature \( T_\lambda = i/\Delta \lambda = (2\pi)^{-1} \sqrt{\lambda/3} \), which can be interpreted as the temperature of quantum fields as we approach the dS asymptopia (the pole of \( a \)) [29].

Interestingly, the contour orientation needed to ensure \( S_g \geq 0 \) implies that \( T_g \) is positive for a “turnaround” universe, and negative otherwise. Negative temperatures can occur in systems with a finite number of accessible states [19, 20], as exemplified by the set of universes we study with finite \( S_g \). The negativity of \( T_g \) for universes like our own may be an important clue about how our macroscopic universe was born, as well as the microscopic ensemble that describes it.

For \( \lambda < 0 \), \( T_\lambda = (2\pi)^{-1} \sqrt{-3/\lambda} \) is imaginary, while \( T_g \) diverges in the \( \mu \to 0 \) limit and seems ill-defined for \( \mu \neq 0 \) (since one cannot find a suitable contour where \( \text{Re}(a) \) remains positive), casting doubt on any thermodynamic interpretation of FRW cosmology with negative \( \lambda \).

**HOMOGENEITY, ISOTROPY, FLATNESS, \( \lambda \)**

The ratio of curvature density to critical density is

\[ \Omega_\kappa(a) = -\frac{3k}{a^4} \left[ \frac{\hat{r}}{a^4} + \frac{\mu}{a^2} - \frac{3k}{a^2} + 1 \right]^{-1}. \]  \hspace{1cm} (21)

Setting the time derivative of this expression to zero, we see that \( \Omega_\kappa \) reaches its maximal value when the scale factor satisfies \( a = a_\star \), where \( a_\star \) is the positive real solution of \( a^4 + \frac{1}{2} \hat{r} a - \hat{r} = 0 \). Thus, \( \Omega_{\kappa_\text{max}} = \Omega_\kappa(a_\star) \) where \( a_\star = \frac{1}{2}(c_\star + \sqrt{c_\star^2 + \hat{r}/c_\star}) \), with \( c_\star = (z_\star - P_\kappa/\kappa_\star)^{1/2} \), \( z_\star = (-Q_\kappa + \sqrt{Q_\kappa^2 + P_\kappa^2})^{1/3} \), \( P_\kappa = (4/3)\hat{r} \), and \( Q_\kappa = -(1/8)\hat{r}^2 \). We plot \( \Omega_{\kappa_\text{max}} \) in Fig. 4.

Comparing Fig. 2 and Fig. 4 note that the contours have the same shape, and as \( S_g \) increases \( |\Omega_{\kappa_\text{max}}| \) decreases. Universes with entropy above some threshold \( S_g \) are universes with maximum curvature below some threshold \( |\Omega_{\kappa_\text{max}}| \); and increasing the threshold \( S_g \) decreases the threshold \( |\Omega_{\kappa_\text{max}}| \). In other words, the most entropically likely universes are those (like our own) in which the curvature never becomes significant throughout cosmic history. This solves the flatness problem.

To go beyond this “zeroth order” result, next add small inhomogeneities (i.e., tensor and scalar perturbations, \( h \) and \( \zeta \)) and anisotropies (i.e., tensor perturbations with wavelengths longer than the Hubble radius) to these preferred nearly-flat backgrounds. To quadratic order, these are described by the actions \( S_t \sim \int d^4x \frac{3}{2} (c_\star^2 - c_\star^2(\nabla \zeta)^2) \) for scalar perturbations and \( S_h \sim \int d^4x a^3 (h^2 - (\nabla h)^2) \) for tensor perturbations. As seen above, to ensure the leading order answer \( S_g \) is positive, the integration contour in this (and any non-turnaround) case must run up the imaginary \( \tau \) axis; and, as explained in Refs. [10, 12], the perturbations \( h \) and \( \zeta \) satisfy reflecting boundary conditions at the bang that ensure they are even functions of \( \tau \), and hence are real along the imaginary \( \tau \) axis. Together these facts imply \( S_{(2)}^h = 0 \) and \( S_{(2)}^\zeta \) both contribute negatively to \( S_g \) [11]. So for fixed values of the conserved quantities characterizing the cosmology (e.g., \( r \) and \( \mu \) or, more physically, the total entropy in radiation \( S_r \sim r^3/4V_3 \) and the total mass in non-relativistic matter \( M \sim \mu V_3 \)), inhomogeneities and anisotropies decrease \( S_g \), so are entropically disfavored.

So far, we have treated \( \lambda \) as a fixed constant, but now consider it as another parameter in the ensemble [23]. For \( \lambda < 0 \), the thermodynamic interpretation is suspect, and there is no reason to expect a large gravitational entropy. However, for \( \lambda > 0 \), the entropy \( S_g \) is huge, and increases as we decrease \( \lambda \). For example, for reasonably flat universes \( |\Omega_{\kappa_\text{max}}| \ll 1 \) we have \( S_g \sim S_r \lambda^{-1/4} \) when \( (\delta \hat{r}/9)^{1/2} \gg (\hat{r}/2)^{3/2} \) and \( S_g \sim M \lambda^{-1/2} \) when \( (4\delta \hat{r}/9)^{1/2} \lesssim (\hat{r}/2)^{3/2} \). From this we see that, for fixed \( S_r \) and \( M \), \( S_g \) increases as \( \lambda \) approaches zero from above, and the highest entropy \( S_g \) is achieved in the limit \( \lambda \to 0^+ \). In other words, the entropy \( S_g \) also seems to favor universes (like our own) with a tiny positive \( \lambda \). Our result echoes and extends to a realistic universe those of [25–27]. We will discuss the associated statistical ensemble in a forthcoming paper.
Acknowledgements: NT is supported by the STFC Consolidated Grant ‘Particle Physics at the Higgs Centre’ and the Higgs Chair of Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada, through Innovation, Science and Economic Development, Canada and the Province of Ontario through the Ministry of Research, Innovation and Science.