Two-Sheeted Universe, Analyticity and the Arrow of Time

Latham Boyle\textsuperscript{1} and Neil Turok\textsuperscript{1,2}

\textsuperscript{1}Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada, N2L 2Y5
\textsuperscript{2}Higgs Centre for Theoretical Physics, University of Edinburgh, Edinburgh, Scotland, EH8 9YL

(Dated: August 2021)

Our universe seems to be radiation dominated at early times, and vacuum energy dominated at late times. When we consider the maximal analytic extension of this spacetime, its symmetries and complex analytic properties suggest a picture in which spacetime has two sheets, exchanged by an isometry which, in turn, picks a preferred (CPT-symmetric) vacuum state for quantum fields on the spacetime. Previously\textsuperscript{[1, 2]}, we showed how this line of thought provides new explanations for dark matter, the matter-antimatter asymmetry, the absence of primordial vector and tensor perturbations, and the phase of the primordial scalar perturbations; and additional testable predictions. In this paper, we develop this picture in several respects and, in particular, point out that it also provides a new explanation for why the thermodynamic arrow of time points away from the bang.

\textbf{Introduction.} Observations reveal that, seconds after the Big Bang, the universe was described by a highly ordered state: a spatially-flat radiation-dominated FRW metric (plus tiny gaussian, adiabatic, purely-growing-mode scalar perturbations described by a nearly-scale-invariant power spectrum)\textsuperscript{3}. What is this profound clue trying to tell us? The conventional interpretation is that the radiation dominated era was preceded by an earlier hypothetical epoch of accelerated expansion called inflation. However, for every property of the universe that inflation claims to explain or predict, there are numerous published inflationary models that claim to explain or predict the opposite. We worry that inflation may not be a “good explanation” in the sense articulated in Ch. 1 of\textsuperscript{4}, and hope to find an alternative explanation that is harder to vary.

In Refs.\textsuperscript{[1, 2]}, we began to explore a different approach. Instead of imagining that the simple radiation-dominated era we see was preceded by earlier unseen chaos that must be cleaned up by inflation, we tried to take the observed simplicity of the early universe at face value and suppose that it extends all the way back to the bang, to explore where that line of thought naturally leads. We found that the maximally-extended spacetime so obtained has a new reflection isometry around the bang, and that taking this symmetry seriously yields elegant new explanations and predictions for several observed properties of the early universe, for the dark matter, and for the neutrino sector.

In this paper, we go further: in addition to being radiation-dominated in the past, our universe will be vacuum energy dominated in the future. We start by studying the simplest cosmology of this type – a universe with just radiation and vacuum energy – paying special attention to the structure (symmetries and complex-analytical properties) of its maximal analytic extension, which we regard as key clues. We then add matter, requiring it to respect this structure. We point out that, in addition to the features and predictions pointed out in\textsuperscript{[1, 2]}, this picture provides other insights, including a new solution to an old problem: why the thermodynamic arrow of time points away from the bang\textsuperscript{[5, 6].}

\textbf{The background spacetime.} Consider the general FRW line element

\[ ds^2 = a^2(\eta)(-d\eta^2 + \gamma_{ij}dx^i dx^j) \]

where \( \gamma_{ij} \) is the metric on a maximally-symmetric space of curvature \( \kappa \). The Friedmann equation is

\[ \frac{\dot{a}^2}{a^4} = \frac{\rho}{3} - \frac{\kappa}{a^2} \]

where we have taken \( \hbar = c = 8\pi G = 1 \) and a dot denotes \( \dot{a} = da/d\eta \).

Since we would like to study the analytic behavior of the scale factor as \( a \to 0 \) (where radiation dominates) and as \( a \to \infty \) (where vacuum energy dominates), we start with the simplest model that has the correct asymptotic behavior at both extremes – i.e. a universe which only contains radiation and vacuum energy: \( \rho = r/a^4 + \lambda \), where \( r \) and \( \lambda \) are positive constants.

Now the Friedmann equation reads:

\[ \dot{a}^2 = \frac{1}{3}(r - 3\kappa a^2 + \lambda a^4). \]
We choose the root \( m \) from the different ways to slice the same de Sitter ("dS") solution; representing three physically distinct solutions.

The solution of (3) is given by

\[
\eta = \frac{1}{3} \left( 3r^2 + 4\kappa b^2 + \lambda \right),
\]

(4)

i.e., (3), with \( a \to b \) and \( r \leftrightarrow \lambda \).

The choice of roots \( m \), \( \alpha \) and \( \beta \) is just a convention: for definiteness, when \( (3\kappa)^2 < 4\lambda \) (so \( m \) is complex) we choose the root \( m \) in the upper half plane; and when \( (3\kappa)^2 > 4\lambda \) (so \( m \) is real) we choose the root \( m < 1 \) (and imagine it to have an infinitesimal positive imaginary part). Then \( \alpha \) and \( \beta \) are fixed by taking \( \sqrt{2} \) to have its standard branch cut along the negative real axis.

Note that in the absence of radiation, the three options for \( \kappa \) (positive, negative, and zero) just represent three different ways to slice the same de Sitter ("dS") solution; but when we include radiation, these three options for \( \kappa \) represent three physically distinct solutions.

The limit \( (3\kappa)^2 / (3\lambda \eta) \to 0 \) yields the \( \kappa = 0 \) solution.

The limit \( (3\kappa)^2 / (3\lambda \eta) \to 0 \) yields the \( \kappa = 0 \) solution.

The nature of the solution \( a(\eta) \) depends, first, on the sign of the spatial curvature \( \kappa \), and second, on whether the universe is close to flat, \( (3\kappa)^2 < 4\lambda \), or far from flat, \( (3\kappa)^2 > 4\lambda \). In all four cases, as we move along the real \( \eta \) axis, the solution \( a(\eta) \) is also real, and the universe described by \( a > 0 \) starts at \( \eta = 0 \) and ends at \( \eta = \eta_\ast \), where \( \eta_\ast \) is given in the following table:
In the three cases satisfying $(3\kappa) < \sqrt{4\lambda r}$, $a(\eta)$ describes a universe that grows from a simple zero at $\eta = 0$ (the bang) to a simple pole at $\eta = \eta_*$ (dS asymptopia). In the fourth case $(3\kappa) > \sqrt{4\lambda r}$, corresponding to the lower right-hand corner of the table, $a(\tau)$ describes a universe that grows from a simple zero at $\eta = 0$ (the bang) to a finite maximum size, before recollapsing to another simple zero at $\eta = \eta_*$ (the crunch), while $a(\tau - iK'/\beta)$ describes a universe that shrinks from a simple pole at $\eta = 0$ (past dS asymptopia) to a minimum size, before re-expanding to another simple pole at $\eta = \eta_*$ (future dS asymptopia).

These various solutions are shown in Fig. 1.

In all cases, the universe described by $a(\eta) > 0$ extends from $\eta = 0$ to $\eta = \eta_*$, but the maximal analytic extension of $a(\eta)$ is periodic, with period $2\eta_*$, and describes a two-sheeted universe, with $a > 0$ on one sheet and $a < 0$ on the other, and with $a(\eta)$ odd under temporal reflection across either the bang at $\eta = (2n)\eta_*$ or the dS boundary at $\eta = (2n+1)\eta_*$ $(n \in \mathbb{Z})$. We can visualize this as saying that $\eta$ is curled into a circle, with $\eta$ and $\eta + 2\eta_*$ identified, so that the universe forms the two-sheeted cylinder shown in Fig. 2 with an isometry that swaps adjacent points on the inner and outer cylinder and sends $a \rightarrow -a$.

**Imaginary time and thermality.** The analytic structure of $a(\eta)$ is shown in Fig. 3 for the physically relevant cases where $(3\kappa)^2 < 4\lambda r$ (“close-to-flat”). As evidence we should take this analytic structure seriously, note the physical time $t(\eta) = \int_0^\eta a(\eta')d\eta'$ is given by

$$t = \frac{\alpha}{\beta} \eta^{-1/2}\ln \left( \frac{\sinh(\beta\eta)}{\sinh(\beta)} - m^{1/2} \chi \frac{\cosh(\beta\eta)}{\cosh(\beta)} \right),$$

where $dn(z, m)$ and $cn(z, m)$ are also Jacobi elliptic functions. The integral $\int_0^{\eta_*} a(\eta')d\eta'$ depends on how the contour from $\eta' = 0$ to $\eta' = \eta$ winds around the poles of $a(\eta')$. This in turn corresponds to the fact that the answer (8) depends on $\ln(z)$, which is only defined modulo $2\pi i$, so that $t$ itself is only well defined modulo $\Delta t = 2\pi i (\alpha/\beta) m^{-1/2} = 2\pi i \sqrt{3/\lambda}$. But a system whose time coordinate is only defined modulo an imaginary shift $t \rightarrow t + \Delta t$ (or, in other words, whose $t$ is curled into a circle of circumference $\Delta t$ in the imaginary time direction) corresponds to a system at finite temperature $T = i/\Delta t$. In this way, we precisely recover the well-known expression for the de Sitter temperature

$$T_{dS} = H/2\pi$$

where $H = \sqrt{\lambda/3}$ is the asymptotic dS Hubble constant (the reciprocal of the asymptotic dS radius).

Note the generality of this result: $T_{dS}$ is independent of $\eta, \kappa, \omega$, or the radiation density, and determined by the residue of $a(\eta)$ at its pole, and hence by $\lambda$.

**Matter and the arrow of time.** Now let us add matter. For simplicity, we study a massive scalar field $\varphi$ and leave the details of higher-spin fields to future work.

The equation of motion for $\varphi$ is

$$\Box \varphi + \xi R \varphi + \mu^2 \varphi = 0$$

where the mass $\mu$ is a constant. For concreteness, we consider the conformal coupling $\xi = 1/6$, which is most natural \cite{8, 9}, but our conclusions apply for any value of $\xi$, including minimal coupling $\xi = 0$. If we define the $\eta$-independent conformal metric $\tilde{g}_{\mu\nu} = a^{-2} g_{\mu\nu}$ and corresponding field $\tilde{\varphi} = a \varphi$, Eq. (10) becomes

$$\Box \tilde{\varphi} + \xi \tilde{R} \tilde{\varphi} + \tilde{\mu}^2 \tilde{\varphi} = 0$$

where $\Box$ and $\tilde{R}$ are the wave operator and Ricci scalar of $\tilde{g}_{\mu\nu}$, and $\tilde{\mu} = a \mu$ is time-dependent: $\tilde{\mu} = \tilde{\mu}(\eta)$.

We have seen that the analytically-extended background metric may be interpreted as a two-sheeted cylinder, with a reflection symmetry (through the bang or through the dS boundary) which swaps adjacent points on the two sheets (see Fig. 2). We regard this as a clue to be taken seriously. We thus require that the matter fields living on this background share its structure. In particular, $\tilde{\varphi}$ should satisfy

$$\tilde{\varphi}(\eta, x) = \tilde{\varphi}(\eta + 2\eta_*, x)$$

(i.e. it lives on the cylinder shown in Fig. 2) and

$$\tilde{\varphi}(\eta, x) = \pm \tilde{\varphi}(-\eta, x)$$

$$\tilde{\varphi}(\eta, x) = \pm \tilde{\varphi}(2\eta_* - \eta, x),$$

(i.e. it is even or odd under reflection through either the bang or dS boundary: we explain how to resolve this sign ambiguity below). Note that these conditions are not independent: Eqs. (12) and (13a) imply Eq. (13b).

To see what the symmetries (12, 13) imply, let us examine the solutions of (11). For our purposes, it is sufficient to solve (11) near the bang and the dS boundary, where the gravitational backreaction of $\varphi$ can be neglected. [20]

Near the bang, where radiation dominates, we can neglect the vacuum energy and spatial curvature of the metric, so $a(\eta) = (r/3)^{1/2} \eta$, and $\tilde{\varphi}$ is a linear combination of terms of the form $\tilde{\varphi}_k(\eta) \exp(ik \cdot x)$, where

$$\tilde{\varphi}_k + (k^2 + \frac{r}{3} \eta^2 \mu^2) \tilde{\varphi}_k = 0,$$

with $k^2 = k \cdot k$, and the general solution is

$$\tilde{\varphi}_k = c_+ D_p(\gamma \eta) + c_- D_p(-\gamma \eta),$$

where $D_p(z)$ is the parabolic cylinder function \cite{13}, $c_\pm$ are constants, $p \equiv -\frac{1}{2} + i \sqrt{3} r k^2 / 2$, and $\gamma \equiv e^{-\pi i / (\sqrt{3} r^2 k^2)} / 4$.

Near the dS boundary, where vacuum energy dominates, we can neglect the radiation and spatial curvature of the metric, so $a(\eta) = 1/H(\eta_* - \eta)$, and $\tilde{\varphi}$ is a linear combination of terms $\tilde{\varphi}_k(\eta) \exp(ikx)$, where

$$\tilde{\varphi}_k + (k^2 + \frac{\mu^2}{H^2(\eta_* - \eta)^2}) \tilde{\varphi}_k = 0,$$
with the general solution
\[ \tilde{\varphi}_k = (k(a_1 - \eta))^{1/2}[c_1 J_k(k(a_1 - \eta)) + c_2 Y_k(k(a_1 - \eta))] \quad (17) \]

where \( J_k(z) \) and \( Y_k(z) \) are Bessel functions \[ 13 \], \( c_1 \) and \( c_2 \) are constants, and \( \nu \equiv \frac{1}{2}(1 - 4m^2/H^2)^{1/2} \).

From these solutions, we see a crucial difference between the behavior at the bang and at the dS boundary. At the bang, \( \tilde{\varphi}_k \) remains analytic (\( \varphi_k = a^{-1} \tilde{\varphi}_k \) remains meromorphic), and may be analytically continued across it. For this analytic continuation to be compatible with the reflection symmetry \[ 13 \], \( \tilde{\varphi} \) must satisfy the following boundary condition ("b.c.") at the bang:

\[ \tilde{\varphi}(0, x) = \pm \tilde{\varphi}(0, x), \quad (18a) \]

\[ \tilde{\varphi}(0, x) = \mp \tilde{\varphi}(0, x). \quad (18b) \]

By contrast, \( \tilde{\varphi} \) is not analytic at the dS boundary, and cannot be analytically extended across it, so it does not satisfy a corresponding b.c. there.

The fact that basic considerations of symmetry and analyticity force all fields to satisfy a restrictive boundary condition at one end of spacetime (the bang) but not at the other end (the dS boundary) is striking. It seems to suggest a fundamental new explanation for why the thermodynamic arrow of time (i.e. the direction in which entropy increases) points away from the bang.

The upper (lower) sign in \( 18 \) corresponds to a Neumann (Dirichlet) b.c. for \( \tilde{\varphi} \) at the bang. This is closely related to the b.c. at the boundary of anti de Sitter (AdS) spacetime \[ 14 \ 15 \]: there too, if one conformally maps the spacetime to a portion of the infinite static (Einstein) cylinder, one finds a simple reflecting (Neumann or Dirichlet) b.c. for the field \( \tilde{\varphi} \). In fact, the AdS b.c.s were originally derived \[ 14 \] by mapping AdS to a portion of the static Einstein cylinder, and then noting that it could be analytically extended to form a two-sheeted spacetime, with a natural \( Z_2 \) symmetry relating the sheets. Of course, this closely parallels our derivation, but with some notable differences: in the AdS case, one analytically continues in the radial direction, and the resulting b.c.s make the AdS boundary an ordinary mirror (a timelike surface which reflects radially); while in our case, one analytically continues in the time direction, and the resulting b.c.s make the bang an exotic mirror (a spacelike surface which reflects in the time direction).

For a canonical scalar field, it seems most natural to choose the lower sign in \( 18 \) at the bang, so that \( \varphi = a^{-1} \tilde{\varphi} \) satisfies Dirichlet and Neumann boundary conditions, respectively, since in this case \( \varphi \) and \( \tilde{\varphi} \) both remain finite and analytic at the bang, consistent with the assumption of small gravitational back reaction.

The \( \pm \) option in \( 18 \) also appears in the b.c.s for higher spin fields (as in the AdS case \[ 15 \ 16 \]). Since \( a(\eta) \) is odd under \( \eta \rightarrow -\eta \), we can infer right choice for the tetrad: the conformal tetrad \( \tilde{e}_\mu^a \) (with \( \eta_{ab} \tilde{e}_\mu^a \tilde{e}_\nu^b = \tilde{g}_{\mu\nu} \)) satisfies a Neumann b.c. (or, equivalently, the symmetric condition \( \tilde{e}^a_\mu(\eta, x) = \tilde{e}^a_\mu(-\eta, x) \)) while the physical tetrad \( e^a_\mu(a^{-1} \tilde{e}^a_\mu \eta_{rat} e^r_b e^t_k = g_{\mu\nu} \) satisfies a Dirichlet b.c. (or, equivalently, the anti-symmetric condition \( e^a_\mu(\eta, x) = -e^a_\mu(-\eta, x) \)). As shown in \[ 11 \], this condition successfully explains: (i) the absence of primordial vector perturbations; (ii) the absence of primordial “decaying-mode” scalar and tensor perturbations (which synchronizes the scalar perturbations, explaining the observed oscillations in the CMB temperature spectrum and their phases); and (iii) the fact that bang seems to be a particularly mild singularity (a “Weyl singularity” – a non-singular conformal metric times an overall conformal factor \( a(\eta) \) which momentarily passes through zero).

**Remarks.** (i) In a forthcoming paper, we present a mechanism for generating the observed scale-invariant scalar perturbations in this framework \[ 17 \].

(ii) In our picture, instead of thinking of the universe evolving through the bang, we can think of the bang as a spacelike mirror – a surface of reflection symmetry. As with an electromagnetic mirror, we can either think of the bang as boundary of spacetime, satisfying special boundary conditions, or else (by the method of images), we can think of a mirror universe on the “other side” of the bang.

(iii) One might worry that, in deriving \( 5 \), we dropped the only term in the Friedmann equation (the matter density \( a^{-3} \)) that is odd under \( a \rightarrow -a \), while all other terms are even. Would including this term spoil the symmetry? No: if the matter fields satisfy the condition \[ 18 \] then, far from the bang, the matter density is actually \( a^{-3} \) (which is even under \( a \rightarrow -a \)); while near the bang it remains perfectly analytic.

(iv) We saw that the solution \( 17 \) generally has an essential singularity near the dS boundary. The exception is when the field is massless (\( \mu = 0 \)), in which case \( 17 \) remains analytic at the dS boundary \[ 21 \]. This leaves open the possibility that massless fields might obey a boundary condition there too, which would quantize the allowed comoving wavenumbers for massless fields \[ 17 \].

(v) It is natural to speculate that our classical symmetry conditions \( e^a_\mu(-\eta, x) = -e^a_\mu(\eta, x) \) and \( \varphi(-\eta, x) = \varphi(\eta, x) \) (along with analogous conditions for the various other fields) follow from a path integral formulation in which only paths satisfying these conditions are allowed \[ 22 \]. This suggests a natural wavefunction of the universe: namely, that the amplitude for a given spacelike 3-configuration of the geometry and matter fields is the path integral over all appropriately symmetric 4-configurations ending on the given 3-configuration.

**Acknowledgements.** Research at Perimeter Institute is supported by the Government of Canada, through Innovation, Science and Economic Development, Canada and by the Province of Ontario through the Ministry of Research, Innovation and Science. The work of NT is supported by the STFC Consolidated Grant “Particle Physics at the Higgs Centre.”
[1] L. Boyle, K. Finn and N. Turok, “CPT-Symmetric Universe,” Phys. Rev. Lett. 121, no.25, 251301 (2018) doi:10.1103/PhysRevLett.121.251301 [arXiv:1803.08928 [hep-ph]].

[2] L. Boyle, K. Finn and N. Turok, “The Big Bang, CPT, and neutrino dark matter,” [arXiv:1803.08930 [hep-ph]].

[3] P. A. R. Ade et al. [Planck Collaboration], “Planck 2015 results. XIII. Cosmological parameters,” Astron. Astrophys. 594, A13 (2016) doi:10.1051/0004-6361/201525830 [arXiv:1502.01589 [astro-ph.CO]].

[4] D. Deutsch, The Beginning of Infinity: Explanations that Transform the World, Penguin Books (2011).

[5] R. Penrose, “Singularities and time-asymmetry,” in General Relativity: An Einstein Centenary Survey, edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, 1979).

[6] R. Penrose, The road to reality: A complete guide to the laws of the universe, Random house (2005).

[7] M. Abramowitz and I. A. Stegun, Handbook of Mathematical Functions (1964).

[8] R. Penrose, “Conformal Treatment of Infinity,” in Relativity, Groups and Topology, Eds. C. DeWitt and B.S. DeWitt, Gordon and Breach, London (1964); republished as Gen. Rel. and Grav. 43, 901 (2011) doi:10.1007/s10714-011-1110-5.

[9] C. G. Callan, Jr., S. R. Coleman and R. Jackiw, “A New improved energy - momentum tensor,” Annals Phys. 59, 42-73 (1970) doi:10.1016/0003-4916(70)90394-5

[10] N. D. Birrell and P. C. W. Davies, Quantum fields in curved space, Cambridge University Press (1984).

[11] V. Mukhanov and S. Winitzki, Introduction to quantum effects in gravity, Cambridge University Press (2007).

[12] E.T. Whittaker, G.N. Watson, A course of modern analysis, Cambridge University Press (1952), Ch. 6

[13] M. Abramowitz and I. A. Stegun, Table of Integrals, Series and Products, 7th Edition.

[14] S. J. Avis, C. J. Isham and D. Storey, “Quantum Field Theory in anti-De Sitter Space-Time,” Phys. Rev. D 18, 3565 (1978) doi:10.1103/PhysRevD.18.3565

[15] P. Breitenlohner and D. Z. Freedman, “Stability in Gauged Extended Supergravity,” Annals Phys. 144, 249 (1982) doi:10.1016/0003-4916(82)90116-6 1497 citations counted in INSPIRE as of 27 Jun 2021

[16] S. W. Hawking, “The Boundary Conditions for Gauged Supergravity,” Phys. Lett. B 126, 175-177 (1983) doi:10.1016/0370-2693(83)90585-3

[17] L. Boyle and N. Turok, in preparation (2021).

[18] P. Horava and E. Witten, Nucl. Phys. B 460, 506-524 (1996) doi:10.1016/0550-3213(95)00621-4 [arXiv:hep-th/9510209 [hep-th]].

[19] P. Horava and E. Witten, Nucl. Phys. B 475, 94-114 (1996) doi:10.1016/0550-3213(96)00308-2 [arXiv:hep-th/9603142 [hep-th]].

[20] More generally, the solution $\tilde{\phi}$ is an arbitrary linear combination of separable solutions $\tilde{\phi}_k(\eta)\psi_k(x)$, where $\psi_k(x)$ is an eigenfunction of $\gamma^{ij}\tilde{\nabla}_i\tilde{\nabla}_j$; and, if we neglect the gravitational backreaction of $\phi$, so that $a(\eta)$ is given by (5), then (11) becomes the Lamé equation for $\tilde{\phi}_k(\eta)$, which can be solved in terms of Lamé functions [12].

[21] The fact that both massless and massive solutions are analytic at the bang, while only massless solutions are analytic at the dS boundary is due, physically, to the fact that both massless and massive particles reach the bang in finite proper time, while only massless particles reach the dS boundary in finite proper time.

[22] In this picture, the bang may alternatively be thought of as a kind of spacelike orbifold boundary, with some similarity to the timelike orbifold boundaries which appear in constructions like those in [13] [19].