Predicting the cosmological constant from the causal entropic principle

Raphael Bousso,[1] Roni Harnik,[2] Graham D. Kribs,[3] and Gilad Perez[4]

[1]*Center for Theoretical Physics, Department of Physics, University of California, Berkeley, California 94720-7300, USA*
and Lawrence Berkeley National Laboratory, Berkeley, California 94720-8162, USA
[2]*Stanford Linear Accelerator Center, Stanford University, Stanford, California 94309, USA*
and Physics Department, Stanford University, Stanford, California 94305, USA
[3]*Department of Physics and Institute of Theoretical Science, University of Oregon, Eugene, Oregon 97403, USA*
[4]*C. N. Yang Institute for Theoretical Physics, State University of New York, Stony Brook, New York 11794-3840, USA*

We compute the expected value of the cosmological constant in our universe from the causal entropic principle. Since observers must obey the laws of thermodynamics and causality, the principle asserts that physical parameters are most likely to be found in the range of values for which the total entropy production within a causally connected region is maximized. Despite the absence of more explicit anthropic criteria, the resulting probability distribution turns out to be in excellent agreement with observation. In particular, we find that dust heated by stars dominates the entropy production, demonstrating the remarkable power of this thermodynamic selection criterion. The alternative approach—weighting by the number of "observers per baryon"—is less well-defined, requires problematic assumptions about the nature of observers, and yet prefers values larger than present experimental bounds.

I. INTRODUCTION

The discovery that the universe is in a period of accelerated expansion [1,2], combined with an accurate accounting of the total, matter, and radiation components of the energy density [3], provide overwhelming evidence for dark energy. These measurements are completely consistent with the interpretation of dark energy as a nonzero cosmological constant Λ. This has undermined the hope that the energy of our vacuum is uniquely determined by fundamental theory. Instead, it lends credence to the hypothesis that the cosmological constant is an environmental variable, which takes on different values in widely separated regions of the universe.

The observed vacuum energy density[1]

$$\rho_\Lambda = \frac{\Lambda}{8\pi} = (1.25 \pm 0.25) \times 10^{-123}, \qquad (1.1)$$

is at least 55 orders of magnitude smaller than what would be expected from the standard model of particle physics (see, e.g., Ref. [5] for a recent review). The environmental approach does not assert that this tiny value is inevitable, or even typical among all possible values. Rather, it aims to show that it is not atypical among values measured by observers.

A number of conditions must be satisfied for the environmental approach to work. The first, obviously, is that the

fundamental theory must *admit* the observed value of the vacuum energy. This can happen without explicit tuning if the theory gives rise to an enormous number N of different vacua. Of course, typical values will be of order unity, but if they are randomly distributed, they can form a dense spectrum, or "discretuum," with average spacing of order $1/N$. If $N \gg 10^{123}$, then it is likely that the observed value is included in the spectrum. Thus, the approach really depends on whether fundamental theory (which, presumably, is more or less unique) admits a sufficiently dense discretuum.

The second condition is that the observed value must be dynamically *attainable*, starting from generic initial conditions. With $N \gg 10^{123}$ possibilities, there is no reason for the universe to start out in a vacuum like ours. The environmental approach therefore depends on a means to start from some generic initial value and later realize the observed value, either as a branch in the wave function or as a particular spacetime region embedded in a vast universe.

Finally, the environmental approach requires an explanation of why we happen to *observe* such an unusually small vacuum energy. Most values of ρ_Λ in the discretuum will be of order unity, and only a fraction (in the simplest case, a fraction of order 10^{-123}) will have a magnitude as small as the observed value. It is not enough to show that the small value given by Eq. (1.1) is possible; one must also show that it is *not unlikely* to be observed.

The first condition appears to be satisfied by string theory, which admits as many as 10^{500} long-lived metastable vacua [6–10] (see Ref. [11] for a discussion of earlier work). The second condition can then be met be-

---

[1]Unless indicated otherwise, all observed values in this paper are taken from Ref. [4]. Where no explicit units are given, we set $\hbar = G_N = c = k_B = 1$.

cause the vacua are metastable and can decay into one another.[2]

In this paper we address the third condition. We will use a novel approach, the causal entropic principle, to argue that the observed value of ρ_Λ is not unlikely. Our main result is shown in Fig. 1.

The causal entropic principle is based on two ideas: any act of observation increases the entropy, and spacetime regions that are causally inaccessible should be disregarded. It assumes that on average, the number of observations will be proportional to the amount of matter entropy produced in a causally connected region, ΔS. Vacua should be weighted by this factor to account for the rate at which they will be observed.

Crucially, the size of the causal diamond is inversely proportional to the vacuum energy, so smaller values of ρ_Λ allow for greater complexity. This compensates for the scarcity of vacua with small ρ_Λ. As a result, ρ_Λ prefers to take a value such that vacuum energy begins to dominate near the time of peak entropy production.

We will find that entropy production in the causal diamond is dominated by dust particles heated by stars. This is an important result in its own right: our weight ΔS is a simple physical quantity that turns out to be sensitive to the existence of galaxies, stars, and heavy elements.

We will show that the entropy production rate peaked approximately 2 to 3.5×10^9 yr after the big bang. It is this time scale, rather than the time of galaxy formation, which governs our prediction of the cosmological constant, and it prefers a range of ρ_Λ that is in very good agreement with observation.

The same result also explains the so-called coincidence problem or ''why now'' problem. According to the causal entropic principle, typical observers will exist when most of the entropy production in the causal diamond occurs. Our result ensures that this happens during the era when the matter and vacuum energy densities are comparable.

Outline.—Historically, discussions of the third condition—why do we observe an unusually small ρ_Λ—have focused on anthropic selection effects, which we discuss in Sec. II. Long before the discovery of the string landscape, it was noted that not all values of ρ_Λ are compatible with the existence of observers [14–17]. This culminated in Weinberg's successful prediction [18] that a small nonzero value of ρ_Λ would be observed, which we review in Sec. II A. Weinberg's assumption was relatively modest: Observers require galaxies. But astronomers have since
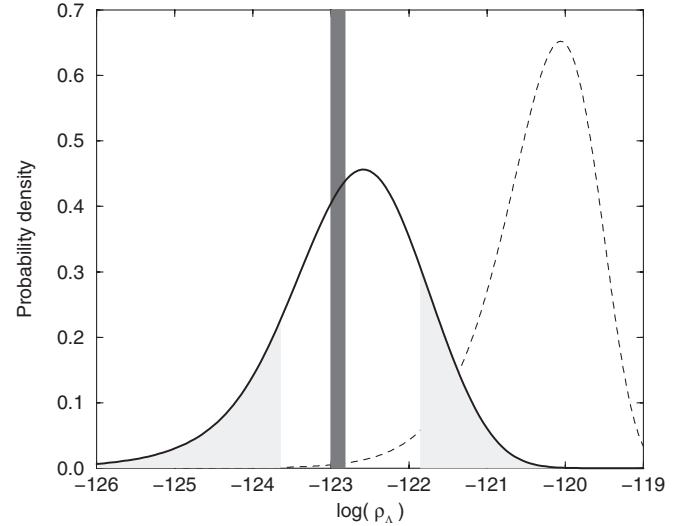

FIG. 1. The probability distribution of the vacuum energy measured by typical observers, computed from the causal entropic principle, is shown as a solid curve. The values consistent with present cosmological data, in the shaded vertical bar, are well inside the 1σ region (shown in white), and hence, not atypical. For comparison, the dashed line shows the distribution derived by estimating the number of observers per baryon. Unlike our curve, it assumes that galaxies are necessary for observers; yet, the observed value is very unlikely under this distribution. For more details about both curves, see Figs. 2 and 8.

discovered galaxies that would have formed even if ρ_Λ had been more than a thousand times larger than the observed value. This leaves a large discrepancy between theory and observation.

A possible resolution is to ask not only about the existence of observers, but to weight vacua by the *number* of observers they contain. This number is generically infinite or zero, so a regularization scheme is needed. A popular approach is to weight by the number of observers per baryon. In Sec. II B, we argue that this approach is both poorly motivated and, in a realistic landscape, poorly defined. Moreover, it does not resolve the conflict with observation. To mitigate the discrepancy, one is forced to posit increasingly specific conditions for life, such as the chemical elements required. Indeed, to do reasonably well, one must suppose that observers can only arise in galaxies as large as ours—a very strong assumption, for which there appears to be no evidence.

In Sec. II C, we motivate and discuss a different approach to this problem. The causal entropic principle weights each vacuum by the amount of entropy, ΔS, produced in a causally connected region [19]. This is the largest spacetime region that can be probed and across which matter can interact. Since observation requires free energy, it is natural to expect that the number of observers will scale, on average, with ΔS. In other words, we demand nothing more than that observers obey the laws of thermodynamics. This is far weaker even than Weinberg's crite-

---

[2]In general, long-lived metastability implies that all matter is diluted before the next decay occurs, so the mechanism depends on efficient reheating in the new region. This rules out models that reduce the cosmological constant gradually [12,13]. In the string landscape, the vacuum preceding ours was likely to have had an enormous cosmological constant. Its decay acted like a big bang for the observed universe and allowed for efficient reheating [6].

rion, and one might be concerned that it will not be sufficiently restrictive. Yet, in combination with the causal cutoff, this minimal requirement becomes a powerful predictive tool.

This is demonstrated in Sec. III, where we use the causal entropic principle to derive the probability distribution for ρ_Λ over universes otherwise identical to ours. We begin, in Sec. III A, by computing the geometry of the causal diamond for general ρ_Λ. We are mainly interested in its comoving volume as a function of time, $V_c(t)$. In Sec. III B, we consider important mechanisms by which entropy is produced within our causal diamond. We estimate contributions from stars, quasars, supernovae, and other processes. We find that the leading contribution to ΔS comes from infrared photons emitted by interstellar dust heated by stars.

In Sec. III C, we analyze this leading contribution in more detail. We compute $dS/(dV_c dt)$, the rate at which entropy is produced per unit comoving volume per unit time. This rate will depend on ρ_Λ because large ρ_Λ disrupts galaxy formation and thus star formation. Interestingly, however, it turns out that this dependence is not important for our final result. In Sec. III D, we integrate the rate found in Sec. III C over the causal diamond determined in Sec. III A. This yields the weight factor, $\Delta S(\rho_\Lambda)$. We display the resulting probability distribution for ρ_Λ, and we note that the observed vacuum energy lies in the most favored range.

In Sec. IV, we pinpoint the origins and discuss some implications of our main findings. We also identify important intermediate results.

Extensions.—In the interest of time and clarity, we have limited our task. We use the causal entropic principle solely to compute a probability distribution over positive values of the vacuum energy, holding all other physical parameters fixed. This is the case most frequently studied in the literature, making it straightforward to compare our result with those obtained from the traditional method of weighting by observers-per-baryon [20–23].

In other words, our distribution is conditioned on the assumptions that $\rho_\Lambda > 0$, and that all low-energy physics is the same as in our vacuum. We ask only about the probability distribution of ρ_Λ in this subspace of the landscape. This is a valid consistency check: Suppose that the observed value were disfavored on a subspace picked out by other observed parameters. Then it would only become less likely when the distribution is extended over the entire landscape, and so the model would conflict with observation.

Because negative values of ρ_Λ are tightly constrained by standard (though questionable) anthropic arguments, the main challenge for environmental approaches has been to suppress large positive values of ρ_Λ. For this purpose, it suffices to concentrate on the subset of vacua with $\rho_\Lambda > 0$. This simplifies our analysis, since negative ρ_Λ lead to a different class of metrics. (Interestingly, a preliminary analysis indicates that negative values will be somewhat favored by the causal entropic principle, though not by enough to render the observed value unlikely.)

It will be interesting to use the causal entropic principle to compute the probability distribution of other parameters, such as the amplitude of primordial density perturbations, $\delta\rho/\rho$, the spatial curvature, k, or the baryon-to-photon ratio, η. A crucial task will be to estimate the probability distribution over multiple parameters at once, since this is a much more stringent test for the environmental approach. For example, consider a distribution over two parameters, ρ_Λ and $\delta\rho/\rho$. When weighting by observers-per-baryon, the upper bound on ρ_Λ arises from the requirement that galaxies can form. Hence, it would seem highly favorable to increase $\delta\rho/\rho$, since ρ_Λ could then be increased by the third power of the same factor. This would render the observed values of both ρ_Λ and $\delta\rho/\rho$ exceedingly unlikely. As we will discuss in Sec. IV B, however, galaxy formation is not a significant constraint on ρ_Λ in our approach. Hence, we expect this problem to be virtually absent.

Given a multivariate distribution, one can ask about the probability distribution over one parameter (say, $\log\rho_\Lambda$) with other parameters integrated out. As more parameters are allowed to vary, the distribution for $\log\rho_\Lambda$ is likely to become broader after they are integrated out. Yet, the observed value must remain typical if the environmental approach is to succeed. The most radical choice is to study the distribution of $\log\rho_\Lambda$ after integrating out all other parameters characterizing the landscape. This would be tantamount to deriving the typical range of $\log\rho_\Lambda$ from fundamental theory alone, without reference to parameters specific to our own vacuum. This would have been impossible in conventional approaches. But as we discuss in Sec. IV A, ΔS may depend simply on ρ_Λ when averaged over the entire landscape. Hence, the causal entropic principle puts this task within our reach.

II. APPROACHES TO WEIGHTING VACUA

A. The Weinberg bound

Weinberg [18] estimated the range of ρ_Λ compatible with galaxy formation. No galaxies form in regions where ρ_Λ exceeds the matter density ρ_m at the time when the corresponding density perturbations become nonlinear (assuming otherwise identical physics). If we grant that galaxies are a prerequisite for the existence of observers, then these regions will not contain observers, and such values of ρ_Λ will not be observed. Combined with a similar argument[3] for negative values of ρ_Λ, Weinberg [18] found that

---

[3]With $\rho_\Lambda < 0$, the universe will recollapse after a time of order $|\rho_\Lambda|^{-1/2}$. If one assumes that most observers emerge only after several billion years, then an upper bound on $(-\rho_\Lambda)$ results by requiring that the universe should not recollapse too soon [16]. We will consider only positive values of Λ in this paper.

only values in the interval

$$-1 < \frac{\rho_\Lambda}{\rho_{\rm m}} < 550 \qquad (2.1)$$

will be observed.

The upper bound is larger than 1 because the matter density today, $\rho_{\rm m}$, has been diluted since galaxy formation, by the redshift factor $(1 + z_{\rm gal})^3$. Weinberg used $z_{\rm gal} \approx 4.5$, but in the meantime, dwarf galaxies have been discovered at redshifts as high as $z_{\rm gal} \approx 10$ [24], raising the upper bound:

$$-1 < \frac{\rho_\Lambda}{\rho_{\rm m}} < 5000. \qquad (2.2)$$

The observed ratio of vacuum to matter energy density is much smaller than the upper bound:

$$\frac{\rho_\Lambda}{\rho_{\rm m}} \approx 2.3. \qquad (2.3)$$

In other words, it would appear that the observed value of ρ_Λ is in fact quite unlikely, even allowing for the anthropic constraint, Eq. (2.2).

Let us rephrase Weinberg's argument in a more general language. The probability distribution for observed values of ρ_Λ can be written as

$$\frac{dP}{d\rho_\Lambda} \propto w(\rho_\Lambda) \frac{dp}{dN} \frac{dN}{d\rho_\Lambda}, \qquad (2.4)$$

where N is the number of vacua with vacuum energy smaller than ρ_Λ, and p is the total prior probability for these vacua. Since $\rho_\Lambda = 0$ is not a special point, vacua in the landscape are uniformly distributed when averaged over intervals of ρ_Λ of order 10^{-123} or smaller near $\rho_\Lambda = 0$:

$$\frac{dN}{d\rho_\Lambda} = {\rm const.} \qquad (2.5)$$

Before anthropic selection, it is reasonable to assume that all vacua are equally likely[4]:

$$\frac{dp}{dN} = {\rm const.} \qquad (2.6)$$

The anthropic condition of galaxy formation assigns a weight $w = 1$ to vacua in the range of Eq. (2.2), and $w = 0$ to all other vacua. Thus, all values of ρ_Λ in the interval of Eq. (2.2) are equally likely, $dP/d\rho_\Lambda = {\rm const.}$

Restricting to vacua with $\rho_\Lambda > 0$, it is instructive to consider the probability distribution as a function of $\log\rho_\Lambda$,

$$\frac{dP}{d\log\rho_\Lambda} = \rho_\Lambda \frac{dP}{d\rho_\Lambda} \propto \rho_\Lambda w(\rho_\Lambda). \qquad (2.7)$$

With the above, "binary" weight, this distribution will be a growing exponential of $\log\rho_\Lambda$, with a sharp cutoff at $\log\rho_\Lambda \approx -120$ from the upper bound in Eq. (2.2). The observed value, $\log\rho_\Lambda \approx -123$, is suppressed by about 3 orders of magnitude compared to values near the cutoff.

B. Weighting by observers per baryon

In order to reduce this discrepancy, one would need to go beyond the binary question of whether or not there are observers. Surely the *number* of observers will depend on the cosmological constant and will begin to decrease for values of ρ_Λ smaller than the upper bound in Eq. (2.2). If we weight vacua by this number, perhaps this will be more effective at suppressing large values of ρ_Λ than a simple binary filter.

Unfortunately, this strategy is not well-defined without a regularization scheme. The dynamics of eternal inflation results in a multiverse containing an infinite number of regions for every value of ρ_Λ. Each region is an open universe with infinite spatial volume at all times. (The hyperbolic spatial geometry reflects the symmetries preserved by a vacuum bubble formed in a first-order phase transition from a higher metastable vacuum.) If a vacuum admits any observers at all, their number will be infinite.

Various authors [20–22] have proposed to weight vacua by the number of observers per baryon, or per photon, or per unit matter mass. But none of these choices are particularly well motivated. If there are infinitely many baryons, why should it matter how efficiently they are converted to observers? Why is a vacuum less likely to be observed if a smaller fraction of its mass becomes observers, as long as there are infinitely many of them?

More importantly, these regularization methods are not universally defined. This makes them inapplicable in a rich landscape, where we will eventually be forced to consider vacua with very different low-energy physics. Two vacua may have different baryon-to-photon ratios, so that the above weighting methods are inequivalent; which should we choose? Indeed, it seems unlikely that a standard definition of "baryon" can be given that would be meaningful in all vacua.[5]

These difficulties arise because the regularization refers to a reference particle species such as "baryons." But it also refers to "observers," and this leads to additional problems. It seems virtually impossible to define what an observer is in vacua with different low-energy physics. Even in our own universe, it is unclear how to estimate the number of observers per baryon. One approximation is to assume that it will be proportional to the fraction of baryons that end up in galaxies. But this fraction depends strongly on the minimum mass of a galaxy capable of harboring observers, M_*, which is not known.

---

[4]Dynamical effects can modify this flat prior [25]. We shall assume that the resulting distribution remains effectively flat, at least for small $|\rho_\Lambda|$. Models violating this assumption are unlikely to solve the cosmological constant problem.

[5]Reference [26] proposes an interesting method for defining a unit comoving volume in different vacua, in the limit where bubbles preserve an exact $SO(3, 1)$ symmetry.

Figure 2 shows probability distributions for ρ_Λ, under various assumptions for M_*. Dwarf galaxies as small as $M_* \sim 10^7 M_\odot$ have been detected. With this choice, the observed value of Λ is nearly 3 orders of magnitude, or 3.5σ, below the median value.

There is some evidence that galaxies with mass below $10^9 M_\odot$ will not retain the heavy elements produced in supernova explosions [23]. Under the *additional assumption* that such elements are required for life, one may then set M_* to this larger value. But the resulting prediction remains unsatisfactory. The median exceeds the observed value by a factor of more than 2 orders of magnitude, or 2.9σ.

One can speculate that for some reason, life requires a galaxy as large as our own, or perhaps even a larger group [20,23] ($M_* = 10^{12} M_\odot$). Then the observed value is about 1.8σ, or a factor of 22, below the median of the predicted distribution. However, at present we can see no compelling arguments for this extreme choice. Thus, the weighting by observers-per-baryon leads to a dilemma: Either it requires overly specific and questionable assumptions, or else it faces a real conflict with the data.

As these difficulties demonstrate, weighting by observers-per-baryon may not be the correct way to compute probabilities in the landscape. We will now argue for a
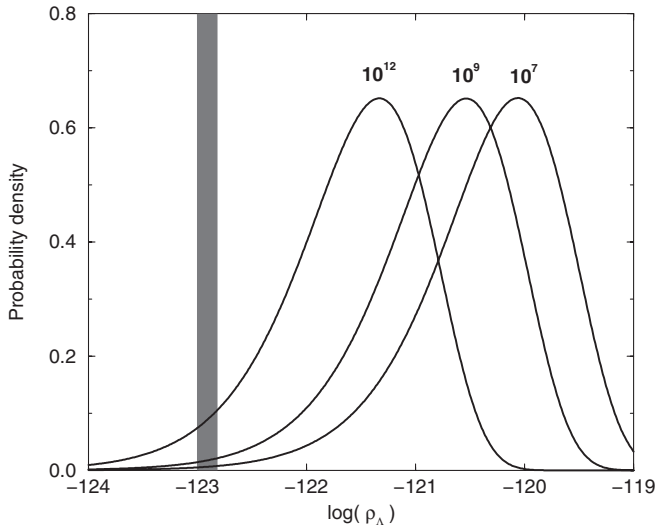

FIG. 2. Weighted by "observers per baryon," the probability distribution for ρ_Λ depends strongly on specific assumptions about conditions necessary for life. Three curves are shown, corresponding to different choices for the minimum required mass of a galaxy: $M_* = (10^7, 10^9, 10^{12})M_\odot$. In neither case is the observed value (vertical bar) in the preferred range. The choice $M_* = 10^7 M_\odot$ (also shown in Fig. 1) corresponds to the smallest observed galaxies. The choice $M_* = 10^{12} M_\odot$ minimizes the discrepancy with observation but amounts to assuming that only the largest galaxies can host observers. By contrast, the causal entropic principle does not assume that observers require structure formation, let alone galaxies of a certain mass; yet its prediction is in excellent agreement with the observed value (see Fig. 8).

different approach, which is always well-defined. It will allow us to assume nothing more about observers than that they respect the laws of thermodynamics.

C. Weighting by entropy production in the causal diamond

Causal entropic principle.—In this paper we will compute the probability distribution for ρ_Λ based on the causal entropic principle, which is defined by the following two conjectures [19]:

(1) The universe consists of one causally connected region, or "causal diamond." Larger regions cannot be probed and should not be considered part of the semiclassical geometry.

(2) The number of observations is proportional to ΔS, the total entropy production in the causal diamond.

Before motivating the two conjectures, let us first clarify the key terms—causal diamond and entropy—they refer to.

The causal diamond [27] is the largest region of space-time causally accessible to a single observer. It is defined by the intersection of the past light-cone of a late-time point on the worldline with a future light-cone of an early-time point, shown in Fig. 3. We choose this time to be the time of reheating, since no matter existed before then. (In vacua with no reheating, ΔS vanishes independently of the choice of the causal diamond. This case does not arise here, since we are holding all parameters other than ρ_Λ fixed.) Only after reheating can matter begin to interact and commence the formation of complex structures, at most at the speed of light.

In a vacuum with negative cosmological constant, the tip of the top cone would be on the big crunch. In any meta-
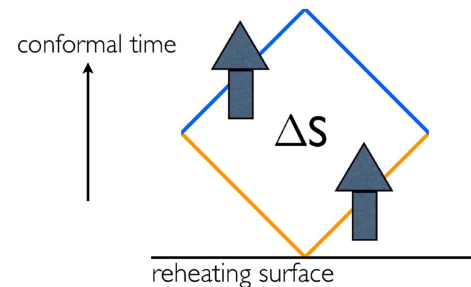

FIG. 3 (color online). A causal diamond is the largest space-time region over which matter can interact. It is delimited by a future light cone from a point on the reheating surface (orange/light), and by a past light cone from a late-time event (blue/dark); in the case of de Sitter vacua this is the cosmological horizon. A vacuum should be weighted by the number of observations made in this spacetime region. Since observation requires free energy, we expect that on average, this number will be proportional to the amount of entropy, ΔS, produced in the causal diamond. Entropy entering through the bottom cone (bottom arrow), such as the CMB, does not contribute to this entropy difference.

stable vacuum with positive cosmological constant, like ours, the top cone is given by the de Sitter event horizon.

One may question the universality of "reheating surface" or our use of an event horizon (a global concept) in a vacuum with finite lifetime, so let us give a more careful definition. A causal diamond can be triggered (that is, a bottom cone drawn) as soon as there is entropy in the matter sector. Reheating is a special case: during inflation, all entropy is in the gravitational sector (the growing horizon of the inflationary universe), but reheating generates matter entropy (mostly radiation, which we include in this class). After vacuum domination in a metastable de Sitter region, the diamond empties out at an exponential rate. If we enlarge the diamond by choosing a later point for the tip, and the additional spacetime volume contains no matter (which will be generic at late times), there is no point in going further. The total amount of matter enclosed by the top-cone will be the same as if the vacuum had been completely stable. Once the metastable vacuum decays, reheating may again occur, in which case a new diamond is triggered. This definition implies that if a decay or phase transition happens while there is still matter around, the two vacua should not be considered separate, and a single causal diamond will enclose both. Thus, we are fundamentally defining the range of a causal diamond in terms of the presence of excitations in the matter sector. This is well-defined in the entire regime of semiclassical gravity, independently of the details of the particle physics.

In a de Sitter vacuum, the cosmological horizon has entropy given by one quarter of its surface area. The relevant area is the cross section of the top cone of the causal diamond, which grows as matter crosses the horizon. This production of Bekenstein-Hawking entropy would contribute enormously to ΔS. However, it is difficult to see the relevance of horizon entropy to the existence of observers. For the same reason, we will ignore the Bekenstein-Hawking entropy produced in black hole formation. This is also natural since we have defined the causal diamond through this distinction. As we have emphasized above, horizons are a gravitational phenomenon and can always be distinguished from the matter sector in the semiclassical regime. Hence, this restriction does not affect the universality of our method.[6]

To summarize, we will consider exclusively the production of entropy in ordinary matter. We will weight a vacuum by the total entropy increase in the causal diamond:

$$w = \Delta S. \qquad (2.8)$$

Motivation.—From a pragmatic point of view, one can simply regard this proposal as an attractive alternative to weighting by observers-per-baryon. The causal diamond is defined independently of low-energy physics, and the entropy increase is a well-defined quantity that replaces more specific assumptions about observers. However, there are additional, more fundamental reasons to embrace the causal entropic principle. Let us discuss each of the two conjectures put forward at the beginning of this subsection.

The first conjecture is motivated by the study of black holes (see [19,27,28] for details). There is now considerable evidence that black hole formation and evaporation is a unitary process [29,30]. This means that there will be two copies of the initial state at the same instant of time, one inside the black hole, and the other in the Hawking radiation outside. This would appear to violate the linearity of quantum mechanics. However, no actual observer can verify the problem [31,32], since the two copies reside in causally disconnected regions of the spacetime. Hence, we can escape from the apparent paradox by abandoning the global viewpoint, and be content with merely predicting the observations of (any) one observer.

However, it would be unnatural for such a radical revision of our view of spacetime to be confined to the context of black holes. In many cosmological solutions, a single observer can access only a small portion of the global spacetime. Hence, it is equally important to restrict our attention to only one (cosmological) observer at a time, which is what we do in this paper. Descriptions of cosmology from the local viewpoint can be found, for example, in Refs. [27,33–36]. Its implications for eternal inflation were studied in Ref. [37].

The second conjecture replaces far more specific conditions assumed to be necessary for observers, such as the existence of galaxies, stable planetary orbits, suitable chemistry, etc., which were needed in the observers-per-baryon approach. The basic idea is that observers, whatever their form, have to obey the laws of thermodynamics. Observation requires free energy and is clearly incompatible with thermal equilibrium or an empty universe. The free energy, divided by the temperature at which it is radiated, can be regarded as a measure of the potential complexity arising in a spacetime region. This is equal to the number of quanta produced, or more fundamentally, the entropy increase ΔS.

It seems plausible that there might be an absolute minimum complexity necessary for observers, so that subcritical weights $\Delta S < \Delta S_{\text{crit}}$ should be replaced by zero. For example, it seems likely that vacua with ρ_Λ of order unity, which can contain only a few bits, have strictly zero probability of hosting observers (see also Ref. [38]). However, any such cutoff does not appear to play an important role for the questions studied here. We choose the weight factor to be simply ΔS, with no cutoff.

---

[6]A different question is whether the exclusion of black hole horizon entropy actually makes a qualitative difference to the results of this paper. It seems likely that it would not. According to the causal entropic principle, the preferred value of ρ_Λ is set by the time of maximum entropy production. The growth of supermassive black holes, and of their entropy, is largest during the era a few Gyr after the big bang and eventually slows down. Thus, it appears to set a similar time scale to the one we obtain from stellar entropy production.

To avoid confusion, we stress that our weighting has nothing to do with the Hartle-Hawking amplitude, $\exp(S_{dS})$, which describes the number of quantum states associated with a de Sitter horizon [39]. The number of observers, and of observations, is naturally proportional to an entropy difference, ΔS, not to an absolute entropy or the exponential of an entropy. (Despite our appropriation of the term "entropic principle," for which we apologize to the authors, there is also little relation to Ref. [40].)

We could have adopted only the first conjecture, and continued to estimate the number of observers by more explicit anthropic criteria. This would not have changed our final result significantly. But why make a strong assumption if a more conservative one suffices? Our results suggest that the poor predictions from weighting by observers-per-baryon stemmed not from a lack of specificity in characterizing observers, but from the regularization scheme ("per baryon"). The causal diamond cutoff solves this problem, and allows us to weaken anthropic assumptions to the level of an inevitable thermodynamic condition.

Moreover, ΔS is a well-defined weight in any vacuum, however different from ours. It will be a powerful tool in future work, when parameters other than ρ_Λ are allowed to vary simultaneously. We are thus motivated to use our second conjecture throughout. We will find that in our own universe, it captures conditions for observers remarkably well.

III. COMPUTING ρ_Λ FROM THE CAUSAL ENTROPIC PRINCIPLE

In this section, we will compute the probability distribution over ρ_Λ, holding all other physical parameters fixed. We do this in four steps. First, we will compute the geometry of the causal diamond as a function of ρ_Λ. Next, we will identify important effects that produce entropy within the causal diamond. Then we will determine the time-dependence of the entropy production rate per unit comoving volume, as a function of ρ_Λ. Finally, we will fold this together with the time-dependence of the comoving volume contained in the causal diamond to obtain the weight factor, $\Delta S(\rho_\Lambda)$.

A. Metric and causal diamond

Current data are consistent with a spatially flat universe. Hence, we will assume that since the time of reheating, the large scale structure of our universe is described by a spatially flat Friedman-Robertson-Walker (FRW) model:

$$ds^2 = -dt^2 + a(t)^2 d\mathbf{x}^2. \qquad (3.1)$$

(Actually, we are making the stronger assumption that the cosmological constant dominates before curvature does, for *all* values of ρ_Λ considered here. Thus, we are assuming that the universe is flatter than necessarily required by current constraints.)

After reheating, the universe will be dominated first by radiation, then by matter, and finally by vacuum energy. The reheating temperature will not be important; in fact, we will neglect the radiation era altogether. Instead, we extrapolate the matter-dominated era all the way back to the big bang ($t = 0$), where we will place the bottom tip of the causal diamond. This approximation is justified because the radiation era contributes only a small fraction of conformal time in the range of values of ρ_Λ that have any significant probability. This is shown in detail in the appendix.

Thus, we treat the universe as containing only pressureless matter and a cosmological constant. At early times (matter domination), the scale factor will be proportional to $t^{2/3}$, independently of ρ_Λ. At late times (vacuum domination) it will grow like $\exp(t/t_\Lambda)$, where

$$t_\Lambda = \sqrt{\frac{3}{\Lambda}} = \sqrt{\frac{3}{8\pi\rho_\Lambda}}. \qquad (3.2)$$

(In our universe, with ρ_Λ given by Eq. (1.1), we have $t_\Lambda \approx 0.98 \times 10^{61} = 16.7$ Gyr.)

An exact solution (aside from the neglected radiation era) that includes both regimes is

$$a(t) = \left[t_\Lambda \sinh\left(\frac{3}{2}\frac{t}{t_\Lambda}\right) \right]^{2/3}, \qquad (3.3)$$

$$\rho(t) = \rho_\Lambda + \rho_m = \rho_\Lambda \left[1 + \frac{1}{\sinh^2(\frac{3}{2}\frac{t}{t_\Lambda})} \right]. \qquad (3.4)$$

The prefactor, $t_\Lambda^{2/3}$, is arbitrary and can be changed by rescaling the spatial coordinates. Our choice ensures that for solutions with different values of Λ, the scale factors will agree at early times not only by diffeomorphism, but explicitly. This is convenient because Λ is dynamically irrelevant at early times.

Vacuum energy begins to dominate the density when $\rho_\Lambda = \rho_m$, at $t = 0.59t_\Lambda$ (in our universe, 9.8 Gyr after the big bang). Acceleration begins earlier, when $\rho + 3p = \rho_m - 2\rho_\Lambda = 0$, at $0.44t_\Lambda$ (in our universe, at 7.3 Gyr).

In order to compute the boundaries of the causal diamond, it is convenient to work in conformal time, $\tau = \int dt/a(t)$. The metric becomes

$$ds^2 = a(\tau)^2[-d\tau^2 + d\mathbf{x}^2]. \qquad (3.5)$$

Light-rays obey $ds = 0$, and hence for radial light-rays, $d\tau = \pm dr$, where $r = |\mathbf{x}|$.

Using our solution, Eq. (3.3), conformal time will be given by

$$\tau(t) = -t_\Lambda^{1/3} \frac{1}{\cosh^{2/3}(\frac{3t}{2t_\Lambda})} F\left(\frac{5}{6}, \frac{1}{3}, \frac{4}{3}, \frac{1}{\cosh^2(\frac{3t}{2t_\Lambda})}\right) \qquad (3.6)$$

$$= -\frac{t_\Lambda}{a(t)} F\left(\frac{1}{3}, \frac{1}{2}, \frac{4}{3}, \frac{-t_\Lambda^2}{a(t)^3}\right). \quad (3.7)$$

It has finite range, running from

$$\tau(0) = -\frac{\Gamma(\frac{4}{3})\Gamma(\frac{1}{6})}{\Gamma(\frac{1}{2})} t_\Lambda^{1/3} \approx -2.804 t_\Lambda^{1/3} \quad (3.8)$$

at the big bang, to $\tau(\infty) = 0$ at asymptotically late times. The total conformal lifetime of the universe is $\Delta\tau = \tau(\infty) - \tau(0) = 2.804 t_\Lambda^{1/3}$.

The causal diamond is given by the region delimited by

$$r = \frac{\Delta\tau}{2} - \left|\frac{\Delta\tau}{2} + \tau\right|. \quad (3.9)$$

The comoving volume inside the causal diamond at conformal time τ is simply

$$V_c = \frac{4\pi}{3} r^3(\tau). \quad (3.10)$$

This is shown in Fig. 4, as a function of proper time, for several values of ρ_Λ. The maximum comoving volume occurs at the edge of the causal diamond at conformal time $\tau(0)/2$, or equivalently,

$$t_{\text{edge}} \approx 0.23 t_\Lambda \quad (3.11)$$

(in our universe, $t_{\text{edge}} \approx 3.9$ Gyr). The maximum comoving volume itself is $V_c(\tau(0)/2) \approx 11.6 t_\Lambda$.

At late times, the comoving volume goes to zero. This reflects the exponential dilution of all matter, which is expelled from the diamond by the accelerated expansion. The physical radius approaches a constant, t_Λ, the horizon radius of the asymptotic de Sitter space. Note that the "comoving four-volume,"

$$V_4 \equiv \int_0^\infty V_c(t) dt, \quad (3.12)$$

is finite. It is proportional to t_Λ^2, and hence, inversely proportional to the cosmological constant. This will be important, since it means that smaller values of ρ_Λ are rewarded by a larger causal diamond, and thus, potentially greater complexity. This can compensate for their rarity.

B. Major sources of entropy production

To calculate $dP/d(\log\rho_\Lambda)$, we need to calculate the total entropy production in the causal diamond as a function of the cosmological constant, $\Delta S(\rho_\Lambda)$. We have determined above how the causal diamond depends on ρ_Λ, but we must also understand how the rate of entropy production depends on ρ_Λ. We begin by identifying the major sources contributing to ΔS in the causal diamond in our own vacuum.

First, let us discuss sources which can be neglected. Because the causal diamond is small at early times ($\frac{\tau}{\tau_\infty} \ll 1$), and empty at late times ($1 - \frac{\tau}{\tau_\infty} \ll 1$), the most impor-

tant contributions will be produced in the era between 0.1 Gyr to 100 Gyr, when the comoving volume shown in Fig. 4 is large. Hence, we can disregard the entropy produced at reheating, at phase transitions, or by any other processes in the early universe. Virtually all of this entropy (in particular, the cosmic microwave background) entered the causal diamond through the bottom cone and does not contribute to ΔS.

For the same reason, we neglect the entropy in Hawking radiation produced by supermassive black holes. (One might contemplate the possibility of a "planet" orbiting such an object and exploiting its very low temperature radiation as a source of free energy.) However, the time scale for the evaporation of a black hole is enormous (M^3). By the time this entropy would be produced, a typical causal diamond, on which the measure for prior probabilities is based in the local approach [19], will be completely empty.

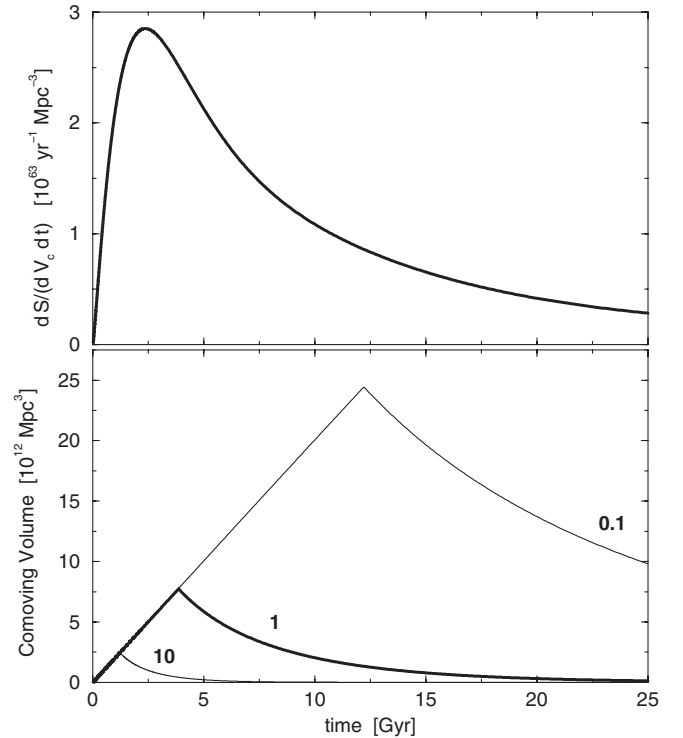Having dismissed effects at small comoving volume, we turn to processes which operate from 0.1 to 100 Gyr, when

FIG. 4. The lower plot shows comoving volume in the causal diamond, V_c, as a function of time for $\rho_\Lambda = (0.1, 1, 10)$ times the observed value given in Eq. (1.1). The kink in the comoving volume corresponds to the "edge" of the causal diamond, where the top and bottom cone meet (see Fig. 3). The upper plot shows the rate of entropy production computed in Sec. III C, which peaks around 2 to 3.5 Gyr. As explained in Sec. IV B, the causal entropic principle prefers values of ρ_Λ such that the (ρ_Λ-dependent) peak of the comoving volume coincides with the (nearly ρ_Λ-independent) peak of the entropy production rate (see Fig. 9).

the comoving volume is large. In this era, entropy is produced by baryonic systems, and we can gauge the importance of a given process by its total entropy production per baryon, or equivalently, per unit mass, or unit comoving volume,[7] dS/dV_c. This can be estimated as the ratio between the amount of energy released per baryon, and the temperature at which this energy is dominantly radiated.

Stars.—Ten percent of baryons end up in galaxies, and thus in stars. Approximately 10% of those baryons actually burn in the course of the lifetime of the star. The energy released is about 7 MeV for each baryon in the reaction $4p \rightarrow {}^4He$, or about 0.7×10^{-2} of the rest mass of the proton. Stars radiate at varying temperatures corresponding largely to optical wavelengths (about 0.2 to 3 eV). However, more than half of this radiation is reprocessed by dust [41]. It is reemitted in the infrared, at approximately 60 μm, or 20 meV. Hence, there will be more than 100 infrared photons per optical photon [42], and the infrared emission will dominate the entropy production. In summary, stellar burning produces an entropy of order 10^6 per baryon, after reprocessing by dust.

Active galactic nuclei.—Active galactic nuclei (AGNs) appear to be the main competitor to stars in terms of luminosity and entropy production. Approximately 10^{-3} of the total stellar mass (i.e., 10^{-4} of baryonic mass) ends up in supermassive black holes, and perhaps 10% of this energy is released during the violent accretion process. This suggests that AGNs release at most about one sixth of the energy as compared with stellar burning. The effective temperature will again be 20 meV, since a large fraction of the radiation is reprocessed by dust. Hence, AGNs appear to produce somewhat less total entropy per baryon than stars, and we will neglect their contribution in the present paper.

In Ref. [43], the intrinsic luminosities of all AGNs above observational limits were compiled to create a quasar luminosity function applicable back to about 1 Gyr after the big bang. This luminosity function suggests that the entropy production rate from quasars is more narrowly peaked than that estimated for stars in Fig. 6. But even at the peak, near a redshift of $z = 2$, the quasar luminosity is a factor of 3 lower than the *present* stellar luminosity [44] (which, in turn, is about 1 order of magnitude lower than the peak stellar luminosity). This seems to rule out the possibility that AGNs ever dominated the entropy production rate.

Other potentially important contributions come from galaxy cooling and from supernovae. Even assuming reprocessing by dust, neither of these phenomena can compete with stellar burning, because they run on less energy but not at lower temperature.

Supernovae.—We focus on core collapse supernovae since they are more abundant than type Ia supernovae (by a factor of 4–5) while producing comparable luminosity per event. They occur in all stars with more than eight solar masses, which constitute 1% of the total stellar mass [see Eq. (3.17)]. The collapse of an iron core into a neutron star releases gravitational binding energy not much smaller than the core mass (1.4 solar masses). Thus roughly 10% of the total mass of the progenitor is released. Most of this energy is carried away by high energy neutrinos, producing little entropy. Only 1% produces optical photons, and is reprocessed by dust into infrared photons. Altogether, supernovae thus convert a fraction of 10^{-5} of stellar baryonic mass into soft photons.

Further quantitative analysis confirms the above estimate. We find that entropy production from supernovae is more than an order of magnitude below the contribution from dust heated by stars.

Galaxy cooling.—A typical galaxy like ours has a mass-to-radius fraction of approximately 10^{-6}. This fraction of the galaxy mass is converted into kinetic energy at virialization. This energy, about 1 keV per stellar baryon, is converted into radiation as the galaxy cools. The virial temperature is about 10^5 K, or 10 to 100 times greater than the temperature of a star. Even assuming reprocessing by dust, galaxy cooling will produce less than 10^4 photons per baryon. This is more than 2 orders of magnitude below the entropy production from dust heated by stars.

C. Entropy production rate

We have argued that the dominant source of entropy production in our universe is dust heated by starlight. In this subsection we will consider the rate at which this entropy is produced. We will ask how it depends on time and on ρ_Λ.

Time dependence.—Deriving $dS/dV_c dt$ from first principles would require a detailed description of all of the dynamics that led up to stellar burning, such as nonlinear evolution, gas cooling, and disk fragmentation. Instead, we will take advantage of observations that quantify the time-dependence of the star formation rate. This will allow us to obtain the entropy production from dust heated by stars. We will then estimate how this rate changes in universes with different cosmological constants. Surprisingly, this latter dependence will not be important for our final result.

In recent years, observations of the extragalactic background radiation in a large range of wavelengths have improved our understanding of the galaxy luminosity function. This has allowed astronomers to infer the star formation rate (SFR) in galaxies, and its evolution in time [45,46] (see [47] for a review). The SFR is defined as the rate of stellar mass production per comoving volume

$$\dot{\rho}_\star(t) \equiv \frac{\partial^2 M_\star}{\partial V_c \partial t}. \tag{3.13}$$

---

[7]The choice of reference unit does not affect our weighting, because it drops out after integrating over the causal diamond.

This function is constrained by observation through a variety of techniques. For example, uv emission from star-forming galaxies is dominated by massive stars that are short-lived. Because of the short lifetime, luminosities in these wavelengths track the birth rate of stars [48]. In addition, detailed surveys of the local universe constrain the SFR at low redshift [49,50]. Bounds on the rate of type II supernovae from Super-Kamiokande and KamLAND also place an indirect bound on stellar birth [51].

The combination of these measurements constrain the SFR back to redshifts of about $z \sim 6$, when the universe was 1 Gyr old. Since then, the SFR may be grossly described as a smooth function that peaks at around 2.5 Gyr and subsequently decreases exponentially with time. The SFR today is roughly 1 order of magnitude less than its peak value. As we shall see, the era of peak stellar formation, which we will call $t_\star$, will play an essential role in determining the cosmological constant using the causal entropic principle.

Several authors have postulated models or functional forms for the SFR, fitting model parameters to the data, for example [52–55]. Variations between the fits lead to a range for $t_\star \sim 1.5$–3 Gyr.

We will illustrate our computation using two different SFRs, in order to illustrate the systematic dependence of our calculation on the time dependence of star formation. The first SFR is from Ref. [54] (labeled "N" in plots) and has $t_\star \sim 1.7$ Gyr. The second SFR, from Ref. [55] (labeled "H"), peaks at $t_\star \sim 2.8$ Gyr. We will find in both cases
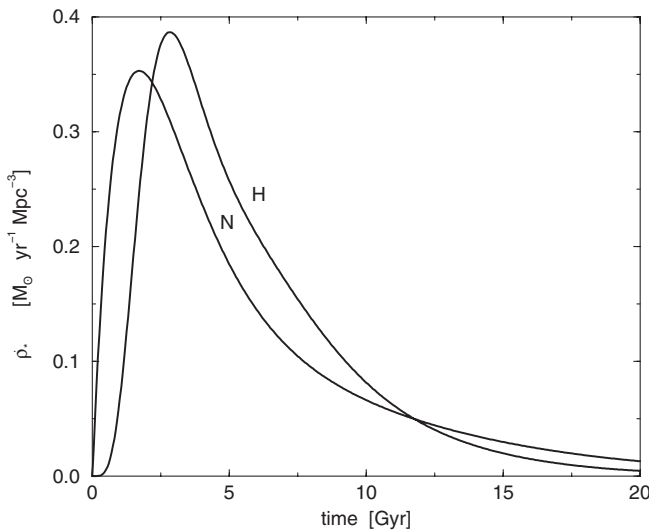

FIG. 5.   The star formation rate as a function of time from Nagamine *et al.* [54], labeled N, and from Hopkins and Beacom [55], labeled H. They are peaked at $t_\star \sim 1.7$ and 2.8 Gyr, respectively. We have normalized both SFRs such that the stellar luminosity density calculated below agrees with the bolometric luminosity observed today. It should be noted, however, that our result is not sensitive to this renormalization.

that the observed value of ρ_Λ lies well inside the 1σ region of the resulting probability distribution.

Both SFRs are shown in Fig. 5 renormalized. One of the biggest uncertainties regarding the determination of the star formation rate is the overall normalization of the curve. Fortunately our result is *completely* insensitive to this overall normalization; it depends only on the shape of the SFR.

For concreteness, we have normalized the SFR such that the implied stellar luminosity today matches the observed bolometric luminosity from low-redshift stars, $10^{8.6} L_\odot / \mathrm{Mpc}^3$. Our normalizations differs from those in the original works since we have used rather rudimentary formulas to compute the present luminosity from the SFR. However, the shape of the total entropy production rate derived with these simple formulas agrees well with that in Ref. [54] when this SFR is used.

With the rate of star formation in hand, one can estimate the rate of entropy production by stars. Let us first consider a single star of mass M born at a time t'. Its entropy production rate is the stellar luminosity divided by the temperature at which photons are radiated,

$$\frac{d^2 s}{dN_\star dt}(M) = \frac{L_\star}{T_{\mathrm{eff}}} \sim \frac{1}{2}\left(\frac{M}{M_\odot}\right)^{3.5} \frac{L_\odot}{20 \text{ meV}}$$
$$= \frac{1}{2}\left(\frac{M}{M_\odot}\right)^{3.5} 3.7 \times 10^{54} \text{ yr}^{-1}. \quad (3.14)$$

Here we have assumed a mass-luminosity relation

$$L_\star \propto M^{3.5}. \quad (3.15)$$

(We use $M_\star$ to refer to total stellar mass, and M to refer to the mass of a specific star. $N_\star$ denotes star number.) The effective temperature of 20 meV is that of the dust which reprocesses about one-half (hence the prefactor) of the starlight and dominates photon number.

The star will only produce entropy over a finite time $t_\star$ that also depends on M:

$$t_\star(M) \sim \left(\frac{M_\odot}{M}\right)^{2.5} 10^{10} \text{ yr}. \quad (3.16)$$

Now let us consider a whole population of stars that are formed at a time t'. At birth, stellar masses are observed to be distributed according to a universal, time-independent function known as the initial mass function (IMF). The Salpeter distribution [56], which goes as $M^{-2.35}$, agrees reasonably with observation, but has since been refined by many authors. In the present calculation we will use a modified Salpeter IMF of the form [55]

$$\xi_{\mathrm{IMF}}(M) \equiv \frac{dN_\star}{dM} = \begin{cases} C_1 M^{-2.35} & \text{for } M \geq 0.5 M_\odot \\ C_2 M^{-1.5} & \text{for } M < 0.5 M_\odot, \end{cases}$$
$$(3.17)$$

where the constants $C_{1,2}$ are set by requiring that the IMF is

continuous and that it integrate to one over the range $0.08 M_\odot < M < 100 M_\odot$.

We now have all the ingredients in place to calculate $d^2 S/(dM_\star dt)$, the contribution of a stellar population that is born at time t' to the entropy production rate at some later time $t > t'$, per unit initial stellar mass. This rate is a function only of the time difference $t - t'$

$$\frac{d^2 S}{dM_\star dt}(t - t') = \frac{1}{\langle M \rangle} \int_{M_{\min}}^{M_{\max}(t-t')} dM \xi_{\mathrm{IMF}} \frac{d^2 s}{dN_\star dt}(M),$$
(3.18)

with the average initial mass (defined at t'), $\langle M \rangle$, defined as

$$\langle M \rangle = \int_{0.08 M_\odot}^{100 M_\odot} dM \xi_{\mathrm{IMF}} M \approx 0.48 M_\odot.$$
(3.19)

The time dependence enters Eq. (3.18) through the death of stars of various masses at different times and thus appears in the upper limit of the upper integral. It is derived by inverting Eq. (3.16):

$$M_{\max}(t - t') = \begin{cases} 100 M_\odot & \text{for } t - t' < 10^5 \text{ yr} \\ \left(\frac{10^{10} \text{ yr}}{t-t'}\right)^{2/5} M_\odot & \text{for } t - t' > 10^5 \text{ yr.} \end{cases}$$
(3.20)

The lower limit of the integral is set by the minimal mass of stars that can sustain nuclear burning,

$$M_{\min} = 0.08 M_\odot.$$
(3.21)

These stars burn for ~5000 Gyr, living well into vacuum domination in our universe.

In order to calculate the entropy production rate in the universe at time t we must convolve $d^2 S/(dM_\star dt)$ with the SFR. That is, we integrate over the birth times t' of all populations of stars born before the time t:

$$\frac{d^2 S}{dV_c dt}(t) = \int_0^t dt' \frac{d^2 S}{dM_\star dt}(t - t') \frac{d^2 M_\star}{dV_c dt'}(t')$$
$$= \int_0^t dt' \frac{d^2 S}{dM_\star dt}(t - t') \dot{\rho}_\star(t').$$
(3.22)

This function is plotted in Fig. 6 for the two forms of the SFR. It is a smooth function that peaks when the universe is about 2.3 Gyr old and 3.3 Gyr old for the two curves plotted. This time scale is set by the peak of the star formation rate and the mean lifetime of a star in our universe. The entropy production rate decreases as stars born during the peak of the SFR begin to die. But due to the high abundance of long-lived low mass stars, $\partial_{V_c} \dot{S}$ maintains a finite value long after star formation has ceased.

In our approximation, the entropy production rate is half the luminosity of stars at the time t, divided by the effective temperature (dust at 20 meV). Modulo this rescaling, Fig. 6 thus also shows our estimate for the luminosity, which was used for normalization as described above.
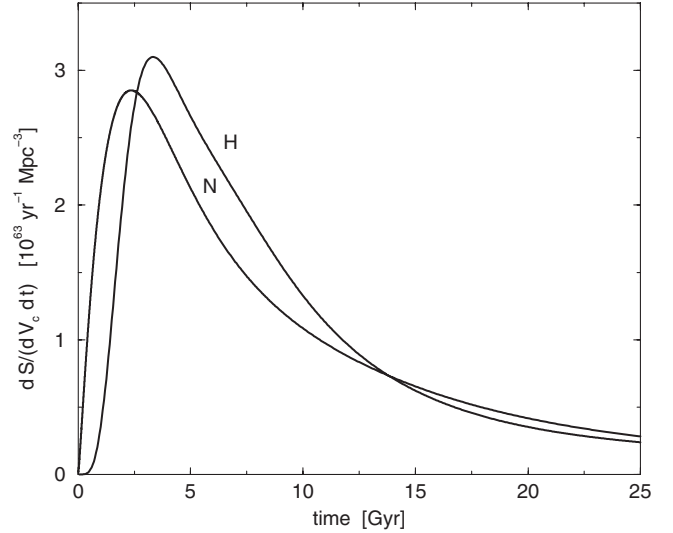

FIG. 6. The entropy production rate in our universe as a function of time, from dust heated by starlight. The two curves shown correspond to different models of the star formation history of the universe [54,55]; see Fig. 5.

We caution that we assumed overly simplistic formulas for the luminosity and the lifetime of a star. For example, we ignored the dependence on metallicity, and a dependence of the exponent in Eq. (3.15) on the mass. This will likely affect the shape of the entropy production rate somewhat, as will corrections from other sources of entropy. However, our prediction for ρ_Λ depends only on the roughest features: the width and position of the peak of the entropy production rate. Hence, it is unlikely that our result would be qualitatively affected by our simplifications.

Dependence on ρ_Λ.—The entropy production rate calculated above is that in our universe. In order to determine $\Delta S(\rho_\Lambda)$, we will need to calculate this function for universes with different values of the cosmological constant. Interestingly, this dependence will be of little importance for our final result.

Stars have decoupled from the global expansion of the universe, so their internal dynamics is unaffected by variations of ρ_Λ. However, the value of ρ_Λ affects the fraction of baryons that are incorporated in halos large enough to form stars at any given time, thus affecting the star formation rate, and ultimately the entropy production. For example, if ρ_Λ is large enough to violate Weinberg's anthropic bound, no baryons will be in star-forming halos and ΔS will be very suppressed.

In a universe with a vacuum energy density ρ_Λ, the fraction of matter that is incorporated in halos of a mass M_G by time t or above, $F(\rho_\Lambda, M_G, t)$, is easily calculated in the Press-Schechter (PS) formalism [57]. The formulas for the PS fraction are summarized in [4]. (This fraction is also a function of the amplitude of density perturbations and the temperature at matter-radiation equality, but since these are held fixed in this calculation we will suppress them.)

For the purpose of our calculation, we will consider a galaxy to be star-producing if the mass of its host halo is $10^7 M_\odot$ or above. Note that this choice involves no speculation about what observers need. The causal entropic principle requires us to compute the entropy production rate as a function of ρ_Λ. This rate depends on whether or not stars, the dominant contributors of entropy, actually form. For $M_G > 10^7 M_\odot$, the virial temperature of the halo is above 10^3 K, enough to support rapid line cooling and efficient stellar production. (The first generation of stars—Population III—were formed in galaxies with even lower masses, but since these stars have been neither observed nor accounted for in the observation-based SFRs, we will not consider them further.) In any case, as we shall see, our final result will be quite insensitive to this choice.

Based on the SFR $\dot{\rho}_\star(t)$ in our universe, we will now estimate the SFR in a universe with a different vacuum energy, $\dot{\rho}_\star(\rho_\Lambda, t)$. The SFR in our universe is peaked at about $t_\star \sim 1.5$–3 Gyr, which is still in the matter-dominated era. The cosmological constant played no dynamical role and cannot have anything to do with this peak. Rather, this time scale is set by astrophysical dynamics, such as galaxy formation, cooling, and feedback.

Therefore, the star formation rate depends mainly on the mass fraction in star-forming galaxies at the critical time $t_\star$. Strictly speaking, the SFR will be sensitive to the PS fraction at times before $t_\star$ because of the cooling period that is required between the time a baryonic structure goes nonlinear and the time it collapses into stars. Baryons that burn during the peak of the SFR actually fell into nonlinear halos a cooling time earlier. We leave a more careful analysis of these and other effects to further investigation.

Let us define $F_\star(\rho_\Lambda)$ as the PS fraction that is most relevant for star formation in a universe with cosmological constant ρ_Λ, namely, that evaluated at $t_\star$, with a minimum mass of $10^7 M_\odot$. In order to capture the mild sensitivity of the SFR to changes of the cosmological constant, we rescale the SFR by the appropriate $F_\star$:

$$\dot{\rho}_\star(\rho_\Lambda, t) = \dot{\rho}_\star(t) \times \frac{F_\star(\rho_\Lambda)}{F_\star(1.25 \times 10^{-123})}. \qquad (3.23)$$

Because the observed value of ρ_Λ is far from disturbing the formation of $10^7 M_\odot$ galaxies, small variations of ρ_Λ will barely affect the mass fraction $F_\star$; see Fig. 7. But they do affect the geometry of the causal diamond. This is the reason why the latter effect will be important for our final result, while Eq. (3.23) gives only a tiny correction.

D. Total entropy production in the causal diamond

From the above results, we can compute the total entropy production

$$\Delta S(\rho_\Lambda) = \int_0^\infty dt\, V_c(\rho_\Lambda, t)\partial_{V_c}\dot{S}(\rho_\Lambda, t). \qquad (3.24)$$

Here, V_c is the comoving volume in the causal diamond at the time t, given in Eq. (3.10). $\partial_{V_c}\dot{S}$ is the rate of entropy
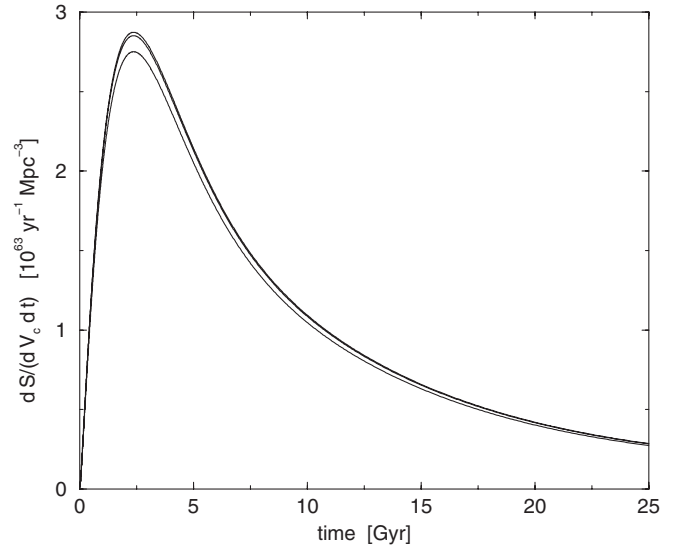

FIG. 7. The entropy production rate of Fig. 6 depends only mildly on the vacuum energy. Hence, dependence on the vacuum energy enters our calculation mainly through the size of the causal diamond (see Figs. 4 and 9). The rate of Ref. [54] is shown here for $\rho_\Lambda = (0.1, 1, 10)$ times the observed value (from top to bottom) using the approximation of Eq. (3.23).

production per unit comoving volume, given in Eq. (3.22). The dependence on ρ_Λ enters mainly through V_c. It is straightforward to perform the integrals numerically.

By Eqs. (2.7) and (2.8), the probability distribution $dP/d(\log\rho_\Lambda)$ is proportional to $\rho_\Lambda \Delta S(\rho_\Lambda)$. We show this distribution in Fig. 8 for both the Nagamine *et al.* [54] as
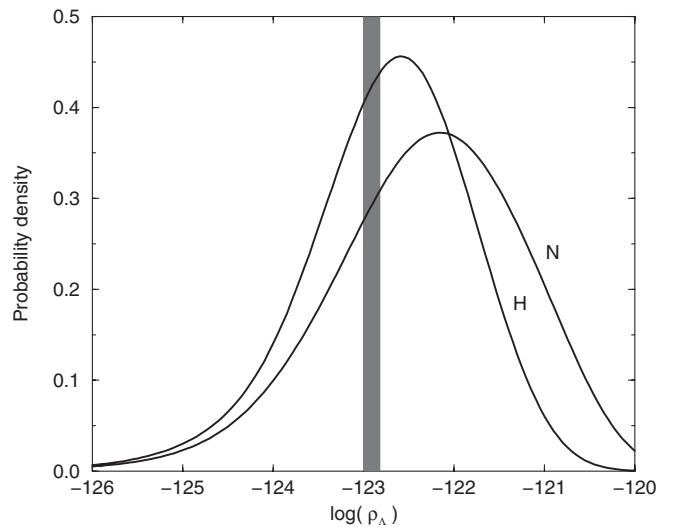

FIG. 8. The probability distribution over $\log(\rho_\Lambda)$ computed from the causal entropic principle. The two curves shown arise from two different models of the star formation rate (see Figs. 5 and 6). Their differences hint at the systematic uncertainties in our calculation that arise since the history of entropy production is not known to arbitrary precision. These uncertainties are apparently irrelevant to our main conclusion: either way, the observed value of ρ_Λ (vertical bar) is not unlikely.

well as the Hopkins *et al.* [55] SFRs. For the Nagamine *et al.* SFR, the median value is $\rho_\Lambda = 5.6 \times 10^{-123}$ with the 1σ error band between 4.2×10^{-124} to 5.8×10^{-122}. For the Hopkins *et al.* SFR, we find the median value $\rho_\Lambda = 2.1 \times 10^{-123}$ with the 1σ error band between 2.4×10^{-124} to 1.4×10^{-122}.

There are "systematic uncertainties" in our calculation of the probability distribution, which come from our lack of knowledge of the precise history of entropy production in our universe. The comparison between the two models for the SFR gives a good sense of their size. It shows that these uncertainties do not affect our conclusion: with either choice, the observed value is well within the 1σ region, and hence, not unlikely.

IV. DISCUSSION

A. ΔS captures complexity

Our main quantitative result is the probability distribution for ρ_Λ. However, we have also discovered an important intermediate result: in our own causal diamond, the dominant contribution to entropy production comes from the infrared glow of dust heated by starlight.[8]

This is remarkable. It shows that a seemingly primitive quantity, ΔS, captures many of the conditions that are usually demanded explicitly by anthropic arguments. ΔS would drop sharply if galaxies, stars, or heavy elements were absent. According to the causal entropic principle, the weight carried by such a vacuum would be suppressed.

For example, consider a universe like ours, except without heavy elements. (This could be arranged by adjustments in the standard model.) Galaxies would still form, and stars would burn, but there would be no dust available to convert optical photons into a much larger number of infrared photons [41]. The causal entropic principle assigns a weight 100 times larger to our vacuum than to this one— simply based on the entropy production, without knowing anything about the potential advantages of heavy elements often claimed in anthropic arguments.

This demonstrates that ΔS can be an effective and very simple substitute for a number of dubious anthropic conditions. More importantly, our result lends credibility to ΔS as a weighting factor for vacua with very different low-energy physics. Estimating the number of observers in such vacua, even averaged over a large parameter space, appears wholly intractable, but estimating ΔS may not be.

Thus, the causal entropic principle may allow us, for the first time, to predict probability distributions over the entire landscape, rather than just over subspaces constrained to coincide with much of our low-energy physics. As discussed in Sec. I, this could lead to a breakthrough on

---

[8]It is amusing to note that this class of entropy producers includes the authors and the reader, in the sense that the Earth, like dust, is made of heavy elements and also absorbs starlight and reemits in the infrared.

extracting predictions directly from the underlying theory (the string landscape), without conditioning on parameters specific to our own vacuum.

B. Understanding our distribution

Now, let us turn to our main result. In our approach, the most likely range of $\log\rho_\Lambda$ is set not by the time of galaxy formation, but by the time at which the *rate* of entropy production, per unit time and unit comoving volume, is largest. This can be understood as follows.

Consider, for the sake of argument, an entropy production rate that is independent of time and of ρ_Λ. Then the total entropy ΔS produced within the causal diamond is proportional to $\int V_c(t)dt$, where V_c is the comoving volume (or equivalently, the mass) present inside the causal diamond at the time t. This integral is the area under the curves shown in the lower panel of Fig. 4.

At small times (near the bottom tip) the causal diamond is small and V_c is negligible. After vacuum domination, at a time of order $t_\Lambda = (3/8\pi\rho_\Lambda)^{1/2}$, the top cone, which contains one de Sitter horizon region, quickly empties out and $M(t)$ vanishes exponentially. Thus, only the era around the time of matter/vacuum equality contributes significantly to the integral.

Up to a ρ_Λ-independent factor of order unity, the above integral is therefore the product of $t_\Lambda M(t_\Lambda)$. But $M(t_\Lambda)$ is just the total mass inside the horizon around the time when Λ begins to dominate. This is again of order t_Λ and thus proportional to $\rho_\Lambda^{-1/2}$. Hence the total entropy produced, ΔS, scales like $1/\rho_\Lambda$ in our hypothetical case of a constant entropy production rate. This is also clear by inspecting the area under the different curves in Fig. 4.

Assuming a flat prior ($dp/d\rho_\Lambda = $ const, or equivalently, $dp/d\log\rho_\Lambda \propto \rho_\Lambda$), the observer-weighted probability distribution is

$$\frac{dP}{d\log\rho_\Lambda} \propto w(\rho_\Lambda)\rho_\Lambda, \qquad (4.1)$$

and the causal entropic principle states that the weight is

$$w = \Delta S. \qquad (4.2)$$

For the hypothetical, constant entropy production rate, we have $w \propto \rho_\Lambda^{-1}$, and hence

$$\frac{dP}{d\log\rho_\Lambda} = \text{const.} \qquad (4.3)$$

The weight ΔS in this case takes a prior distribution that was flat in ρ_Λ into an observer-weighted distribution that is flat in $\log\rho_\Lambda$, showing no preference between, say, $\rho_\Lambda = 10^{-121}$ and $\rho_\Lambda = 10^{-123}$.

In the prior distribution, there are more vacua at large ρ_Λ, so exponentially small values of ρ_Λ are very unlikely. The above example shows that the causal entropic principle captures an important compensating factor: vacua with

smaller ρ_Λ give rise to a larger causal diamond, i.e., to a bigger de Sitter horizon and a longer time until vacuum domination. This allows for greater complexity and compensates for the rarity of such vacua.

Next, let us consider the time-dependent entropy production rate we found in Sec. III C. We found that the entropy production due to stars has a fairly broad peak around $t_{\rm peak} \sim 2$ to 3.5 Gyr after the big bang. At earlier times, it is lower because fewer stars have formed; at late times, it is lower because few new stars form while older ones have burned out.

Because of the time dependence, $\rho_\Lambda \Delta S$ will no longer be constant. For large values of ρ_Λ, the causal diamond is small, and it will contain only a small comoving volume by the time the entropy production peaks [$t_{\rm peak} \gg t_{\rm edge} = 0.23 t_\Lambda$; see Eq. (3.11)]. In this regime, $dP/d\log\rho_\Lambda$ will decrease with ρ_Λ. For small ρ_Λ, the causal diamond is very large, but the entropy production rate peaks early, when the comoving volume is still relatively small ($t_{\rm peak} \ll t_{\rm edge}$). In this regime, $dP/d\log\rho_\Lambda$ will increase with ρ_Λ. This is illustrated in Fig. 9.

Therefore, $dP/d\log\rho_\Lambda$ will be maximal for values of ρ_Λ such that

$$t_{\rm edge}(\rho_\Lambda) = t_{\rm peak}. \qquad (4.4)$$

By Eqs. (3.2) and (3.11), the observed value of $\log\rho_\Lambda$ should be near

$$\log\rho_{\Lambda,{\rm peak}} \approx \log(0.006/t_{\rm peak}^2) \approx -123. \qquad (4.5)$$

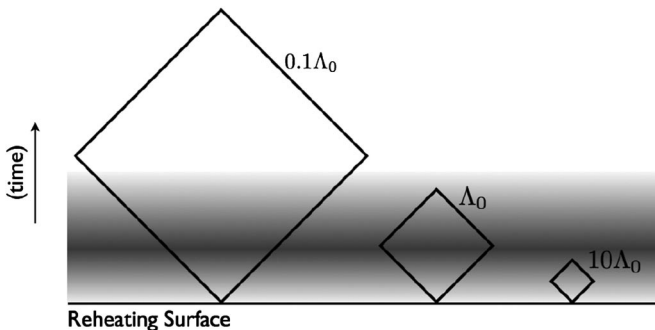This rough estimate is borne out by our more careful

FIG. 9. This cartoon demonstrates how the causal entropic principle leads to a preferred value of the cosmological constant. The horizontal band represents the rate of entropy production; darker areas correspond to a higher rate. Vacua are weighted by ΔS, the total amount of "darkness" inside a causal diamond. Vacua with large cosmological constant are plentiful in the landscape, but they lead to small causal diamonds, which capture virtually no entropy production (right). For some smaller value, the diamond will be just large enough to capture the bulk of the entropy production (center). This is the preferred cosmological constant. Larger diamonds may capture slightly more ΔS (left), but not in proportion to their size. They correspond to vacua with very small cosmological constant which are much rarer in the landscape. Therefore they will be suppressed.

calculation in Sec. III. The excellent agreement of the observed $\log\rho_\Lambda$, Eq. (1.1), with this prediction is reflected in Fig. 4, where it can be seen that the edge time and the peak time really coincide for our universe.

The width of our distribution can also be understood in this manner. Let $t_{\rm on}$ and $t_{\rm off}$ be the times at which the entropy production rate is at half of its peak rate. Using those values in Eq. (4.4) gives roughly the 1σ boundaries we found for our distribution in Sec. III D. To summarize, the peak and the width of the probability distribution for ρ_Λ are related to the peak and width of the entropy production rate by Eq. (4.4).

Our distribution has a greater width than the distribution obtained from the number of observers-per-baryon; this can be seen clearly in Fig. 1. This is also not hard to understand. In the traditional approach, nothing compensates for the exponential growth of $dP/d\log\rho_\Lambda$ with $\log\rho_\Lambda$, until a fairly sharp cutoff occurs when $\log\rho_\Lambda$ becomes large enough to disrupt galaxy formation. Hence, the preferred values of $\log\rho_\Lambda$ are squeezed into a narrow interval, and the observed value is strongly excluded. In our approach, the spacetime volume of the causal diamond depends inversely on ρ_Λ, canceling the pressure towards large values of ρ_Λ. The width of the probability curve is set only by the shape of the peak of the entropy production rate (Fig. 6), which is fairly wide.

In this discussion we have pretended that ρ_Λ does not affect the entropy production rate. In fact, this is an excellent approximation. In the vicinity of the observed value of ρ_Λ, the total entropy production depends on ρ_Λ mainly through the geometry of the causal diamond. The probability density decreases away from this maximum. As a result, values of ρ_Λ large enough to disrupt galaxy formation are highly suppressed even before we take into account the suppression of the entropy production rate resulting from this disruption.

This points at another crucial difference between weighting by entropy production in the causal diamond, and weighting by observers-per-baryon: the preferred ρ_Λ is set by completely different physical processes, and hence, by essentially unrelated time scales. In the latter approach, one assumes that observers require galaxies. Then the disruption of galaxy formation cuts off the exponential growth of $dP/d\log\rho_\Lambda$. As a result, the preferred $\log\rho_\Lambda$ is set by the time when galaxies first form, and this gives a value that is too large compared to Eq. (1.1).

In our approach, we do not assume that observers require galaxies. The size of the causal diamond depends inversely on ρ_Λ, allowing the preferred range of values for $\log\rho_\Lambda$ to be set by the time-dependence of the entropy production rate. The time of peak entropy production by dust heated by starlight picks out the value $\log\rho_\Lambda \approx -123$. The time scale when galaxies form does not enter directly. In our universe, the difference between the two time scales amounts to "only" 3 orders of magnitude in the preferred

ρ_Λ, but it is easy to imagine other vacua in the landscape where the era of peak entropy production is parametrically separated from the era when galaxy halos become non-linear (for example, by a large galaxy cooling time).

C. Statistical interpretation

It is worth emphasizing that it is entirely irrelevant whether the observed ρ_Λ is, say, 0.5σ above or 0.6σ below the median of our distribution. We get to make only one measurement. There is no reason to expect this one data point to be on the median (or on the peak) of the probability distribution. But we can expect that it will not be a very unlikely value. Any value in the 1σ region certainly qualifies as not unlikely. The success of the causal entropic principle, its formal advantages aside, is not that it predicts the precise value of ρ_Λ, but that our distribution shows that the observed value was *not unlikely* to have been observed.

Physicists have a great degree of confidence in certain theories that make only statistical predictions, even though we are unable to make more than a finite number of measurements, let alone test all the consequences of a theory. In this spirit, our result improves our confidence in the causal entropic principle and the underlying landscape. To improve our confidence further, we cannot repeat the measurement of the cosmological constant, but we can extract other predictions or postdictions and compare those to observation.

ACKNOWLEDGMENTS

APPENDIX: THE RADIATION ERA

In this appendix, we justify our neglect of the radiation era. The metric, conformal time, and density during this era are

$$a_{\mathrm{rad}}(t) = c(t - t_0)^{1/2}, \tag{A1}$$

$$\tau_{\mathrm{rad}}(t) = 2c^{-1}(t - t_0)^{1/2}, \tag{A2}$$

$$\rho_{\mathrm{rad}}(t) = \frac{3}{32\pi(t - t_0)^2}. \tag{A3}$$

The constants c and t_0 are determined by matching the Hubble constant and the scale factor to the metric Eq. (3.3), which becomes

$$a(t) = \left(\frac{3t}{2}\right)^{2/3} \tag{A4}$$

for $t \ll t_\Lambda$. They must agree at the time t_{eq}, when the matter and radiation densities are equal, i.e., when [4]

$$\rho_{\mathrm{rad}} = \rho_{\mathrm{eq}} \equiv 0.0026\xi^4 = 3.1 \times 10^{-113}, \tag{A5}$$

where

$$\xi \approx 3.3 \times 10^{-28} \tag{A6}$$

is the observed mass of pressureless matter per photon. This yields

$$t_0 = \frac{t_{\mathrm{eq}}}{4} = \frac{1}{6}\left(\frac{3}{8\pi\rho_{\mathrm{eq}}}\right)^{1/2}, \tag{A7}$$

$$c = \left(\frac{24}{\pi\rho_{\mathrm{eq}}}\right)^{1/12}. \tag{A8}$$

By Eq. (3.9), the size of the causal diamond is set by the total conformal time duration of the universe since reheating, which is finite. In Sec. III A, we neglected the radiation era and extended the matter/vacuum solution all the way back to the big bang ($t = 0$). This yielded a total conformal time

$$\Delta\tau = 2.804\left(\frac{3}{8\pi\rho_\Lambda}\right)^{1/6}, \tag{A9}$$

from Eqs. (3.2) and (3.8).

In order to correct for the presence of the radiation era, we should subtract the conformal time interval $\Delta\tau'$ corresponding to the era $0 < t < t_{\mathrm{eq}}$ that should be excised from the matter/vacuum solution. It should be replaced by the conformal time interval $\Delta\tau''$ corresponding to the radiation dominated era [the portion of the metric (A1) between reheating and matter domination].

Using the above results, however, it is easy to show that

$$2\Delta\tau'' < \Delta\tau' = \left(\frac{\rho_\Lambda}{\rho_{\mathrm{eq}}}\right)^{1/6} \frac{\Delta\tau}{2.804}. \tag{A10}$$

Thus, the corrections to the conformal time, and thus to the size of the causal diamond, are negligible for $\rho_\Lambda < \rho_{\mathrm{eq}}$. For example, with the observed value of ρ_Λ, the correction is less than 1%. The probability of values of $\rho_\Lambda > 10^{-120}$ almost vanishes according to our calculation; yet this is still 7 orders of magnitude below ρ_{eq}. Hence, our approximation is good in the entire range of ρ_Λ in which our probability distribution has support.

[1] S. Perlmutter *et al.* (Supernova Cosmology Project Collaboration), Astrophys. J. **517**, 565 (1999).

[2] A. G. Riess *et al.* (Supernova Search Team Collaboration), Astron. J. **116**, 1009 (1998).

[3] D. N. Spergel *et al.* (WMAP Collaboration), arXiv:astro-ph/0603449.

[4] M. Tegmark, A. Aguirre, M. Rees, and F. Wilczek, Phys. Rev. D **73**, 023505 (2006).

[5] J. Polchinski, arXiv:hep-th/0603249.

[6] R. Bousso and J. Polchinski, J. High Energy Phys. 06 (2000) 006.

[7] S. Kachru, R. Kallosh, A. Linde, and S. P. Trivedi, Phys. Rev. D **68**, 046005 (2003).

[8] F. Denef and M. R. Douglas, J. High Energy Phys. 05 (2004) 072.

[9] E. Silverstein, arXiv:hep-th/0106209.

[10] M. R. Douglas and S. Kachru, Rev. Mod. Phys. **79**, 733 (2007).

[11] A. N. Schellekens, arXiv:physics/0604134.

[12] J. D. Brown and C. Teitelboim, Phys. Lett. B **195**, 177 (1987).

[13] J. D. Brown and C. Teitelboim, Nucl. Phys. **B297**, 787 (1988).

[14] A. D. Sakharov, Zh. Eksp. Teor. Fiz. **87**, 375 (1984) [Sov. Phys. JETP **60**, 214 (1984); Sov. Phys. Usp. **34**, 409 (1991)].

[15] T. Banks, Nucl. Phys. **B249**, 332 (1985).

[16] J. D. Barrow and F. J. Tipler, *The Anthropic Cosmological Principle* (Clarendon Press, Oxford, 1986).

[17] A. D. Linde, Contemp. Concepts Phys. **5**, 1 (2005).

[18] S. Weinberg, Phys. Rev. Lett. **59**, 2607 (1987).

[19] R. Bousso, Phys. Rev. Lett. **97**, 191302 (2006).

[20] H. Martel, P. R. Shapiro, and S. Weinberg, Astrophys. J. **492**, 29 (1998).

[21] J. Garriga, M. Livio, and A. Vilenkin, Phys. Rev. D **61**, 023503 (1999).

[22] G. P. Efstathiou, Mon. Not. R. Astron. Soc. **274**, L73 (1995).

[23] A. Vilenkin, arXiv:astro-ph/0407586.

[24] A. Loeb, J. Cosmol. Astropart. Phys. 05 (2006) 009.

[25] D. Schwartz-Perlov and A. Vilenkin, J. Cosmol. Astropart. Phys. 06 (2006) 010.

[26] J. Garriga, D. Schwartz-Perlov, A. Vilenkin, and S. Winitzki, J. Cosmol. Astropart. Phys. 01 (2006) 017.

[27] R. Bousso, J. High Energy Phys. 11 (2000) 038.

[28] R. Bousso, arXiv:hep-th/0610211.

[29] A. Strominger and C. Vafa, Phys. Lett. B **379**, 99 (1996).

[30] J. M. Maldacena, Adv. Theor. Math. Phys. **2**, 231 (1998); Int. J. Theor. Phys. **38**, 1113 (1999).

[31] L. Susskind, L. Thorlacius, and J. Uglum, Phys. Rev. D **48**, 3743 (1993).

[32] J. Preskill, arXiv:hep-th/9209058.

[33] T. Banks, arXiv:hep-th/0007146.

[34] R. Bousso, J. High Energy Phys. 04 (2001) 035.

[35] L. Dyson, M. Kleban, and L. Susskind, J. High Energy Phys. 10 (2002) 011.

[36] R. Bousso, Phys. Rev. D **71**, 064024 (2005).

[37] R. Bousso, B. Freivogel, and I. S. Yang, Phys. Rev. D **74**, 103516 (2006).

[38] R. Harnik, G. D. Kribs, and G. Perez, Phys. Rev. D **74**, 035006 (2006).

[39] J. B. Hartle and S. W. Hawking, Phys. Rev. D **28**, 2960 (1983).

[40] S. Gukov, K. Saraikin, and C. Vafa, Phys. Rev. D **73**, 066010 (2006).

[41] Y. C. Pei, S. M. Fall, and M. G. Hauser, Astrophys. J. **522**, 604 (1999).

[42] H. Dole *et al.*, arXiv:astro-ph/0603208.

[43] P. F. Hopkins, G. T. Richards, and L. Hernquist, Astrophys. J. **654**, 731 (2007).

[44] A. Kashlinsky, Phys. Rep. **409**, 361 (2005).

[45] P. Madau, H. C. Ferguson, M. E. Dickinson, M. Giavalisco, C. C. Steidel, and A. Fruchter, Mon. Not. R. Astron. Soc. **283**, 1388 (1996).

[46] S. J. Lilly, O. Le Fevre, F. Hammer, and D. Crampton, Astrophys. J. **460**, L1 (1996).

[47] R. C. Kennicutt, Annu. Rev. Astron. Astrophys. **36**, 189 (1998).

[48] P. Madau, arXiv:astro-ph/9612157.

[49] S. Cole *et al.* (2dFGRS Collaboration), Mon. Not. R. Astron. Soc. **326**, 255 (2001).

[50] A. M. Hopkins *et al.* (SDSS Collaboration), Astrophys. J. **599**, 971 (2003).

[51] M. Fukugita and M. Kawasaki, Mon. Not. R. Astron. Soc. **340**, L7 (2003); K. Eguchi *et al.* (KamLAND Collaboration), Phys. Rev. Lett. **92**, 071301 (2004).

[52] S. Cole, C. G. Lacey, C. M. Baugh, and C. S. Frenk, Mon. Not. R. Astron. Soc. **319**, 168 (2000).

[53] A. Gal-Yam and D. Maoz, Mon. Not. R. Astron. Soc. **347**, 942 (2004).

[54] K. Nagamine, J. P. Ostriker, M. Fukugita, and R. Cen, Astrophys. J. **653**, 881 (2006).

[55] A. M. Hopkins and J. F. Beacom, Astrophys. J. **651**, 142 (2006).

[56] E. E. Salpeter, Astrophys. J. **121**, 161 (1955).

[57] W. H. Press and P. Schechter, Astrophys. J. **187**, 425 (1974).