Eur. Phys. J. C (2019) 79:883
https://doi.org/10.1140/epjc/s10052-019-7393-0

THE EUROPEAN PHYSICAL JOURNAL C

Regular Article - Theoretical Physics

An explanation for dark matter and dark energy consistent with the standard model of particle physics and General Relativity

Alexandre Deur[a]

University of Virginia, Charlottesville, VA 22904, USA

Received: 16 August 2018 / Accepted: 10 October 2019 / Published online: 29 October 2019

Abstract Analyses of internal galaxy and cluster dynamics typically employ Newton's law of gravity, which neglects the field self-interaction effects of General Relativity. This may be why dark matter seems necessary. The universe evolution, on the other hand, is treated with the full theory, General Relativity. However, the approximations of isotropy and homogeneity, normally used to derive and solve the universe evolution equations, effectively suppress General Relativity's field self-interaction effects and this may introduce the need for dark energy. Calculations have shown that field self-interaction increases the binding of matter inside massive systems, which may account for galaxy and cluster dynamics without invoking dark matter. In turn, energy conservation dictates that the increased binding must be balanced by an effectively decreased gravitational interaction outside the massive system. In this article, such suppression is estimated and its consequence for the Universe's evolution is discussed. Observations are reproduced without need for dark energy.

1 Introduction

For the last 20 years, observations have shown that the Universe's expansion is presently accelerating. The first solid indication came from measurements of the apparent magnitude of supernovae [1,2]. The leading explanations for the origin of the acceleration are either a non-zero cosmological constant Λ, or exotic fields [3]. This article investigates another possibility which does not require $\Lambda \neq 0$, exotic fields, or a modification of General Relativity (GR). This alternative is a direct consequence of a mechanism that can explain the missing mass problem in galaxies and galaxy clusters without requiring dark matter nor modifying gravity/dynamical laws [4,5]. The phenomenology stems from GR's field self-interaction, which causes GR's non-linear behavior.[1] The consequences of such field self-interaction are well-studied in Quantum Chromodynamics (QCD) which Lagrangian has a similar structure to that of GR.

GR's Lagrangian density is:

$$\mathcal{L}_{\rm GR} = \frac{1}{16\pi G}\sqrt{\det(g_{\mu\nu})}\, g_{\mu\nu}R^{\mu\nu}, \tag{1}$$

where G is the Newton constant, $g_{\mu\nu}$ the metric and $R_{\mu\nu}$ the Ricci tensor. The deviation of $g_{\mu\nu}$ from a constant reference metric $\eta_{\mu\nu}$ defines the gravity field, $\psi_{\mu\nu} = g_{\mu\nu} - \eta_{\mu\nu}$. Expanding in $\psi_{\mu\nu}$ and rescaling[2] the field as $\varphi_{\mu\nu} = \psi_{\mu\nu}/\sqrt{M}$ yields the field Lagrangian [6]:

$$\begin{aligned}\mathcal{L}_{\rm GR} = {} & [\partial\varphi\partial\varphi] + \sqrt{16\pi MG}\,[\varphi\partial\varphi\partial\varphi] \\ & + 16\pi MG\left[\varphi^2\partial\varphi\partial\varphi\right] + \cdots,\end{aligned} \tag{2}$$

where $\left[\varphi^n\partial\varphi\partial\varphi\right]$ denotes a sum over Lorentz-invariant terms of the form $\varphi^n\partial\varphi\partial\varphi$, and M is the system mass. The $n > 0$ terms cause field self-interaction, i.e. the non-linearities that distinguish GR from Newton's theory. This latter is given by the $n = 0$ term, $\mathcal{L}_{\rm Newton} = [\partial\varphi\partial\varphi]$.

QCD's field Lagrangian is:

$$\mathcal{L}_{QCD} = [\partial\phi\partial\phi] + \sqrt{\pi\alpha_s}\left[\phi^2\partial\phi\right] + \pi\alpha_s\left[\phi^4\right], \tag{3}$$

with ϕ^a_μ the gluonic field and α_s the QCD coupling. In the bracket terms, contractions of the color charge indices a are understood in addition to the sums over Lorentz-invariant

[a] e-mail: deurpam@jlab.org

[1] "Self-interaction" is used rather than the less specific "non-linear" denomination: non-linearities in GR or QCD arise from field self-interaction. In contrast, pure-field QED is a linear theory. Non-linearities appear in QED (e.g. photon-photon scattering) once matter is introduced. To distinguish between these two cases, "self-interaction" is used.

[2] The magnitude of the gravity field $\psi_{\mu\nu}$ being proportional to the quantity of matter, $\psi^2 \propto M$, the rescaled field $\sqrt{M}\varphi_{\mu\nu} = \psi_{\mu\nu}$ is the field originating from a unitary mass. This notation emphasizes that effectively, self-interaction terms couple as $\sqrt{GM}$. The rescaling does not affect the results since it amounts to rescaling $\mathcal{L}_{GR}$ by $1/M$ and for classical systems, $\mathcal{L}$ can be rescaled without physical effects.

terms. As in Eq. (2), field self-interaction arises from the terms beside $[\partial\phi\partial\phi]$. Those stem from the color charges carried by the gluonic field. Likewise, GR's self-interaction originates from its field's energy-momentum, the tensor-charge to which gravity couples.

In QCD, self-interaction effects are conspicuous because α_s is large, typically $\simeq 0.1$ at the transition between QCD's weak and strong regimes [7]. A crucial consequence is an increased binding of quarks, which leads to their confinement. In GR, self-interaction becomes important for $\sqrt{GM/L}$ large enough (L is the system characteristic scale), typically for $\sqrt{GM/L} \gtrsim 10^{-3}$ as discussed in Ref. [5] or exemplified by the Hulse-Taylor binary pulsar [8], the first system in which GR was experimentally tested in its strong regime, which has $\sqrt{GM/L} = 10^{-3}$. As in the case of QCD, self-interaction increases the binding compared to Newton's theory. Since the latter is used to treat the internal dynamics of galaxies or galaxy clusters, its neglect of self-interaction may contribute to – or even create – the missing mass problem [4,5,9]. In Ref. [4] a non-perturbative numerical calculation based on Eq. (2) is applied in the static limit to spiral galaxies and clusters. A non-perturbative formalism (lattice technique) – rather than a perturbative one such as the post-newtonian formalism – was chosen because in QCD, confinement is an entirely non-perturbative phenomenon, unexplainable within a perturbative approach. The results of Refs. [4,5] indicate that self-interaction increases sufficiently the gravitational binding of large massive systems such that no dark matter nor ad-hoc gravity/dynamical law modifications are needed to account for the galaxy missing mass problem. Self-interaction also explains galaxy cluster dynamics and the Bullet cluster observation [10]. Finally, the Tully–Fisher relation [11,12], an important observation difficult to explain in the dark matter context, was shown in Ref. [4] to be the GR analog to QCD's Regge trajectories [13]. Accounting for self-interaction automatically yields flat rotation curves for disk galaxies when those are modeled as homogeneous disks of baryonic matter with exponentially decreasing density profiles, which is a good approximation of the observations. In contrast, dark matter halo profiles must be specifically tuned for each galaxy to make its rotation curve flat.

Besides quark confinement, the other principal feature of QCD is a dearth of strong interaction outside of hadrons (i.e., quark bound states) because color confinement keeps the (colored) gluonic field in the hadron. As shown in numerical lattice calculations [14], the field lines – which for a free-field spread isotropically from the source to infinite distances – are for a self-interacting field rearranged in a finite volume roughly contained between the quarks: the collapsed field lines between two quarks form an approximately one dimensional "flux-tube" in which flux lines do not spread. Their density, i.e., the force acting between quarks, is hence constant with the quark separation r. While this confined field produces a binding energy stronger than in the free-field case, such field concentration inside the hadron causes a field depletion outside. This conforms to energy conservation: compared to the free-field case, the increased binding energy in the hadron is compensated by a near absence of potential energy outside the hadron since the field lines have been pulled-in due to self-interaction. Increases of binding energy have also been calculated, with the same numerical lattice technique, for gravity and massive structures [4,5]. This increased binding must, by energy conservation, weaken the action of gravity at larger scale. This can then be mistaken for a repulsion, i.e., dark energy. Specifically, the Friedmann equation for an isotropic and homogeneous universe is (for a matter-dominated flat Universe) $H^2 = 8\pi G\rho/3$, with H the Hubble parameter and ρ the density. As massive structures coalesce, gravity is effectively suppressed at scales larger that these structures. This weakening with time results in a larger than expected value of H at early times, as seen by the observations suggesting the existence of dark energy.

An important point for the present article is that the morphology of the massive structures in which gravity may be trapped determines how effective the trapping is: the less isotropic and homogeneous a system is, the larger the trapping is. For example, this implies a correlation between the missing mass of elliptical galaxies and their ellipticities. The correlation was predicted in [4] and subsequently verified in [9]. The role of the system spacial symmetry is also supported by the relation $J = \epsilon M^\gamma$ describing both the galactic Tully–Fisher observation and the hadronic Regge trajectories.[3] Here J is the angular momentum, M the system mass, and ϵ a constant depending on the type of galaxy or hadron family considered. Inside a less symmetric system, the force is more enhanced and γ is larger than that of a more symmetric system, as observed: Regge trajectories apply to hadrons (flux tubes of 1-dimension) and have $\gamma = 2$, while for the Tully–Fisher relation which applies to disk galaxies (2-dimensional systems), $\gamma = 1.26 \pm 0.07$.

The possible effects of the Universe's inhomogeneity have been discussed in the past to explain cosmological obser-

[3] The Tully–Fisher relation is usually expressed as $L \propto V^x$, with L the absolute luminosity of the galaxy (proportional to its visible mass M), V the rotation speed and $x = 3.9 \pm 0.2$. Since $MV \propto J$ with J the disk angular momentum, the Tully–Fisher relation can then be re-expressed as $J \propto M^{1.26\pm0.07}$, of the same form as Regge trajectories $J \propto M^2$. Regge trajectories stem from the increase of the quark binding energy necessary to compensate for the increased centrifugal force at higher angular momentum J. The potential determining the binding energy (essentially the mass for light hadrons) is proportional to r, which yields $J \propto M^2$. A similar picture holds for the Tully–Fisher relation in the self-interaction framework. The difference is the shape of the system in which the force is confined, *viz* the 2-dimensional galaxy disk rather than the 1-dimensional flux tube. In two dimensions, flux line density, i.e. force, falls as $1/r$, yielding in a $\ln(r)$ potential not as steep as the 1-dimensional potential proportional to r.

vations without requiring dark energy [15]. In particular, the possible importance of backreactions, of same origin as field self-interaction, has been pointed out [16]. The calculations carried out so far are typically perturbative. Thus they are blind to the non-perturbative phenomena that are critical in the analogous QCD phenomenology. Previous non-perturbative attempts have been inconclusive [17]. The present approach, while remaining within GR's description of the universe evolution, see Sect. 3, folds the effects of inhomogeneities into a generic function D_M that expresses the large distance consequences of the non-perturbative effects, and which functional form is modeled from general considerations, see Sect. 4. That this approach differs from others using backreactions or inhomogeneity is illustrated by the identification of an explicit mechanism (field trapping) that is not perturbative, and by the direct connection between dark energy and dark matter that this work exposes.

In summary, traditional analyses of internal galaxy or cluster dynamics employ Newton's gravity that neglects the self-interaction terms in Eq. (2), and this may explain the need for dark matter [4,5]. Traditional analyses of universe evolution do use GR, but under the approximations of isotropy and homogeneity, which suppress the effects of the self-interaction terms [4,9], and would by definition disregard any local phenomenon that could affect gravity's field, such as field trapping. The weakening of gravity at large distance due to these terms is thus neglected, which may be why dark energy seems necessary. This was conjectured in Ref. [4] and the present article investigates this possibility.

2 Field depletion outside massive structures

As just discussed, energy conservation implies that the increased binding energy in massive non-isotropic systems, e.g., galaxies or galaxy clusters, should decrease gravity's influence outside these systems.[4] The consequence on the Universe's dynamics can be folded in a depletion factor D. We will show in Sect. 3 that such factor naturally appears in the Universe's evolution equations once the approximations that the universe is isotropic and homogeneous are lifted. Before this, it is useful to first get an idea of its form and magnitude. Its more thorough determination is given in Sect. 4.

When self-interaction effects are small (low mass) or suppressed (symmetric system) $D \simeq 1$ and the traditional treatment of gravity applies. When gravity's field is trapped in a massive system, gravity is suppressed outside the system. This is accounted for by having $D < 1$, with $D = 0$ if the field is fully trapped. Since the spacial distributions of matter, radiation, and dark energy differ, separate D factors must be considered for these quantities, D_M, D_R and D_Λ, respectively. Electromagnetic radiation does not clump and couples weakly to gravity, so $D_R \simeq 1$. Presumably, $D_\Lambda = 1$ for the same reason but in any case, we assume $\Lambda = 0$ throughout this article. Since self-interaction effects disappear for homogeneous isotropic systems, $D_M \simeq 1$ for the early Universe. Then D_M decreases as structure formation renders the universe less homogeneous. Thus, D_M depends on time, i.e. on the redshift z, and this dependence is driven by large structure formation. In particular, significant field trapping occurs, i.e. the transition from $D_M(z) \simeq 1$ to $D_M(z) < 1$, when galaxies formed and became massive enough so that the $\sqrt{GM}$ coupling in Eq. (2) enables self-interaction. This happens in the range $2 \lesssim z \lesssim 10$ since typically, present ($z = 0$) galaxies have $\sqrt{GM/L} \approx 10^{-3}$ and a large structure mass increases as $(1+z)^{-1}$. (L grows slower so we ignore its z-dependence in the assessment.) $D_M(z)$ then continues to change as groups and clusters form. $D_M(z)$ may not always decrease with time even if the structures' masses increase, since trapping also depends on the homogeneity and symmetry of the structures. For instance, some galaxies had filament shapes for $z \gtrsim 2$, which favors field trapping and thus small D_M. They then grew to disks or ellipsoids, i.e., more symmetric morphologies for which field self-interaction effects tend to cancel out. Likewise, the elliptical/disk galaxy ratio is continuously increasing [18–20]. This implies that, all other things being equal (e.g., ignoring other rearrangements such as in clusters), D_M may rise at small z.

3 Accounting for field depletion in evolution of the universe

Once structures have enough mass so that GR's self-interaction cannot be neglected anymore, field trapping diminishes the effect of gravity at scales larger than the structures. One can thus presume that the effect can be embodied by a function $D(z)$ factoring the gravity magnitude G. We demonstrate it in this section by tracking the terms that disappear from the universe evolution equation under the hypotheses of homogeneity and isotropy, and by identifying these terms with the effect of field trapping.[5] First, we recall the traditional evolution equation obtained assuming homogeneity and isotropy.

[4] In GR, total energy is not necessarily conserved since its definition may excludes gravitational energy. However, here (and in the previous work of Ref. [5]) gravitational energy is included. In any cases, regardless of its definition, energy is conserved in static cases or for asymptotically flat space-time, which are the cases treated here and in [5].

[5] The same method is used in QCD: when symmetry assumptions are lifted, new structures terms appear, e.g., in cross-section expressions. These terms are then parameterized using measurements or calculated non-perturbatively e.g. with lattice methods.

3.1 Evolution equation for an homogeneous, isotropic universe

The universe evolution equation is derived using the Einstein field equation:

$$R_{\mu\nu} = -8\pi G S_{\mu\nu}, \tag{4}$$

with $S_{\mu\nu}$ the energy-momentum tensor. Assuming an homogeneous and isotropic universe reduces $R_{\mu\nu}$ and $S_{\mu\nu}$ to diagonal tensors. Eq. (4) then yields:

$$R_{00} = \frac{3\ddot{a}}{a}, \tag{5}$$

$$R_{ij} = -\left[2K + 2\dot{a}^2 + a\ddot{a}\right]g_{ij}, \tag{6}$$

$$R_{0i} = 0, \tag{7}$$

$$S_{ij} = \frac{1}{2}(\rho - p)a^2 g_{ij}, \tag{8}$$

where a is the Robertson–Walker scale factor, K the space curvature sign, ρ the density, p the pressure, and latin indices denote spacial components only. Combining Eqs. (5–8) yields the traditional Friedmann equation:

$$\dot{a}^2 + K = 8\pi G\rho a^2/3. \tag{9}$$

3.2 Evolution equation for an inhomogeneous, anisotropic universe

Structure formation causes spatial inhomogeneities and, once those are massive enough, field trapping is induced. Terms in $R_{\mu\nu}$ and $S_{\mu\nu}$ that vanish under the approximations of isotropy and homogeneity now appear with the formation of structures. We show here that within GR's formalism, these terms can be regrouped in an overall term $\mathbf{D}(z)$ factoring the right hand side of Eq. (9).

If the assumptions of isotropy and homogeneity are lifted, new terms, including off-diagonal ones, appear in $R_{\mu\nu}$ and S_{ij}. Eqs. (5–8) then change to:

$$R_{00}(1+\alpha) = \frac{3\ddot{a}}{a}, \tag{10}$$

$$R_{ik}(\delta^k_j + \beta^k_j) = -\left[2K + 2\dot{a}^2 + a\ddot{a}\right]g_{ij}, \tag{11}$$

$$R_{0i} = \gamma_i, \tag{12}$$

$$S_{ik} = \frac{1}{2}(\rho - p)a^2 g_{ij}(\delta^j_k + \theta^j_k), \tag{13}$$

where δ_{ij} is the Kronecker delta, and α, β_{ij}, γ_i and θ_{ij} are functions representing the anisotropic components of $R_{\mu\nu}$ and S_{ij}, i.e. the components vanishing when isotropy and the Robertson–Walker metric are assumed. Combining Eqs. (4) and (10) yields:

$$\frac{3\ddot{a}}{a} = -4\pi G(\rho + 3p)(1+\alpha). \tag{14}$$

Equations (4) and (11) together bring:

$$-8\pi G S_{ik}(\delta^k_j + \beta^k_j) = -\left[2K + 2\dot{a}^2 + a\ddot{a}\right]g_{ij}. \tag{15}$$

Combining Eqs. (13) and (15) gives:

$$4\pi G(\rho - p)(1+\omega) = \left[\frac{2K}{a^2} + \frac{2\dot{a}^2}{a^2} + \frac{\ddot{a}}{a}\right], \tag{16}$$

where $\omega \equiv g_{il}(\beta^l_j + \theta^l_j + \theta^l_k\beta^k_j)(g^{-1})^{ij}$, i.e., ω is an average of the anisotropy factors. Defining $D(z) \equiv \left[(1 + \frac{3\omega+\alpha}{4}) + \frac{3p}{4\rho}(\alpha - \omega)\right]$, Eqs. (14) and (16) yield:

$$\dot{a}^2 + K = 8\pi G D(z)\rho a^2/3. \tag{17}$$

We assumed here that the effect of field trapping are represented by $D(z)$ which we thus identify to the depletion function.[6] This fulfills the expectation that $D(z)$ factors G and that $D(z) \to 0$ for an homogeneous isotropic universe.

For simplicity we have not distinguished so far between non-relativistic matter and radiation/relativistic matter. The anisotropy factors for the latter, ω_R and α_R, are negligible. So, in the early Universe, when all content is relativistic, $(\alpha_R - \omega_R) \approx 0$. In the latter Universe, $\rho \gg p$ so $\frac{3p}{4\rho}(\alpha_M - \omega_M) \ll [1 + (3\omega_M + \alpha_M)/4]$. ($\omega_M$ and α_M are the anisotropy factors for non-relativistic matter.) Accounting for this simplifies the depletion function to:

$$D(z) \equiv \left[\left(1 + \frac{3\omega + \alpha}{4}\right) + \frac{3p}{4\rho}(\alpha - \omega)\right] \approx 1 + (3\omega + \alpha)/4. \tag{18}$$

That anisotropy factors differ for relativistic and non-relativistic contents and for Λ is formalized by transforming D and ρ into vectors in Eq. (17), with $\rho \to \boldsymbol{\rho} = (\rho_M, \rho_R, \rho_\Lambda)$ and $D \to \mathbf{D} = (D_M, D_R, D_\Lambda)$. After the matter-radiation equilibrium epoch, $z \ll z_{eq} \simeq 3400$, and assuming $\Lambda = 0$, one has $\mathbf{D} \simeq (D_M, 1, 1)$ and $\boldsymbol{\rho} \simeq (\rho_M, 0, 0)$.

The present-time critical density $\boldsymbol{\rho_{c0}}$ is defined by setting $K = 0$ in the vector version of Eq. (17):

$$\boldsymbol{\rho_{c0}}\mathbf{D}(0) \equiv \frac{3H_0^2}{8\pi G}, \tag{19}$$

with $H_0 \equiv \dot{a}_0/a_0$. The densities of matter, radiation and Λ evolve as usual:

$$\boldsymbol{\rho} = \left(\rho_{0M}\left(\frac{a_0}{a}\right)^3, \rho_{0R}\left(\frac{a_0}{a}\right)^4, \rho_{0\Lambda}\right). \tag{20}$$

[6] Most generally, the inhomogeneity/anisotropy term $D(z)$ may contain effects other than field-trapping (like e.g., a pure QCD calculation of a structure function would not include electromagnetic effects on hadron structure). We neglect this possibility.

Defining $\Omega^*_M(z) \equiv \frac{8\pi G D_M(z)}{3H_0^2}\rho_{0M}$, $\Omega^*_R \equiv \frac{8\pi G D_R}{3H_0^2}\rho_{0R}$ and $\Omega^*_\Lambda \equiv \frac{8\pi G D_\Lambda}{3H_0^2}\rho_{0\Lambda}$, Eqs. (17) and (20) yield:

$$\boldsymbol{\rho}\mathbf{D}(0) = \frac{3H_0^2}{8\pi G}\left[\Omega^*_M\left(\frac{a_0}{a}\right)^3 + \Omega^*_R\left(\frac{a_0}{a}\right)^4 + \Omega^*_\Lambda\right]. \tag{21}$$

A "screened" density fraction Ω^* has the form $\Omega^* = \Omega D$ where Ω corresponds to the traditional definition: $\Omega \equiv \frac{8\pi G}{3H_0^2}e_0$. The Ω^* are not directly comparable with the mass-energy census of the Universe. They are relevant to densities assessed from the universe dynamical evolution. With the definition of Ω^*, $\boldsymbol{\rho}$ is explicitly independent of $\mathbf{D}(z)$. Equation (17) yields for present time:

$$1 = [D_M(0)\Omega_M + D_R\Omega_R + D_\Lambda\Omega_\Lambda] - \frac{K}{a_0^2H_0^2}, \tag{22}$$

which leads to $\Omega_K \equiv -\frac{K}{a_0^2H_0^2}$, as usual. Due to the D_M term in Eq. (22), that $\Omega_M = 1$, $\Omega_R \simeq 0$ and $\Omega_\Lambda = 0$ does not imply $\Omega_K = 0$. This does not necessarily disagree with the WMAP result that $\Omega_K \approx 0$ [24] since it depends on the universe dynamical evolution, which is modeled differently in Eq. (22).

Equation (17) yields the usual expression for $\mathcal{D}_L$, the luminosity distance of a source with redshift z, except that z-dependent density fractions Ω^*_i now enters $\mathcal{D}_L$:

$$\mathcal{D}_L(z) = \frac{1+z}{H_0\sqrt{\Omega_K}}\sinh\left[\sqrt{\Omega_K}\int_{(1+z)^{-1}}^{1}\frac{dx}{x^2\sqrt{\Omega_K x^{-2} + \Omega^*_M(z)x^{-3} + \Omega^*_R(z)x^{-4} + \Omega^*_\Lambda(z)}}\right] \tag{23}$$

with $x \equiv 1/(1+z)$. Likewise, the universe age is given by:

$$t_0 = \frac{1}{H_0}\int_0^1 \frac{dx}{x\sqrt{\Omega_K x^{-2} + \Omega^*_M(z)x^{-3} + \Omega^*_R(z)x^{-4} + \Omega^*_\Lambda(z)}}. \tag{24}$$

The luminosity distance, Eq. (23), is the quantity used to interpret the large-z supernova data. We now need to model $D_M(z)$ which enters Eq. (23) *via* the screened density Ω^*_M. (We have $\Omega^*_R = \Omega_R$, and we assume a zero cosmological constant, $\Omega_\Lambda = 0$.)

4 Construction of $D_M(z)$

In this section, we apply our current knowledge of the evolution of large structures to quantitatively model the depletion function $D_M(z)$. It is driven by:

- the timeline of the mass growth of large structures;
- the masses involved in these structures relative to the total universe mass;
- the geometry (mass distribution) of these structures.

The timeline and its effect on $D_M(z)$ is as follow: During the matter-radiation epoch, for $z_{eq} \lesssim 3400$, the universe is nearly homogeneous and isotropic, $D_M(z_{eq}) \approx 1$. From $15 \gtrsim z \gtrsim 0$, galaxies form and evolve to their present morphologies. At $z \simeq 10$, about 10% of the baryonic matter has coalesced into highly asymmetric protogalaxies [21,22], with field trapping inside galaxies becoming important. At $z \simeq 2$, galaxies evolve to more symmetric shapes and the elliptical/disk galaxy ratio increases [18–20]. These two developments release some of the field trapped inside the galaxies [5,9]. During the evolution of galaxy morphologies, from $10 \gtrsim z \gtrsim 2$, larger structures coalesce: galaxies gather to form groups and protoclusters. At $z \simeq 6$, most of the baryonic matter is in these structures, with field trapped between pairs of galaxies. From $2 \gtrsim z \gtrsim 1.2$, protoclusters evolve to clusters. For $z \lesssim 1.2$, clusters start arranging themselves into more homogeneous superstructures: filaments and sheets. This releases some of the field trapped in between clusters.

The different families of structure evolve with different timelines, e.g., galaxies form before groups or clusters. Furthermore, these families have vastly different shapes and masses. Hence, we separate $D_M(z)$ into a galactic part and a group/cluster/supercluster part:[7]

$$D_M(z) = \xi\left[R_g D_g(z) + R_c\right]D_c(z), \tag{25}$$

where g stands for galaxy, and c for cluster, group or supercluster. The normalization factor ξ, close to 1, accounts for possible field trapping before galaxies start forming. R_i (with $i = g$ or c) is the fraction of the baryonic mass contained in the family i at $z = 0$. Such fractions should vary with z but this is factored in $D_i(z)$, the depletion function for the family i.

Since structures grow linearly with $(1+z)^{-1}$ and since D_i varies inversely to the structure mass: the larger the mass, the smaller D_i, a simple choice for D_i is to take $D_i \propto z$ during the growth process and D_i constant otherwise. A better functional form is chosen in view of the following considerations: once a structure reaches a mass $m(z)$ larger than a critical mass m_{crit} (that depends on the structure geometry) at z_{crit},

[7] The separation is approximative since the baryonic contents of galaxies and intracluster medium interact. Also, finer separations could be considered, such as distinguishing between group, cluster and supercluster, or between disk, elliptical and irregular galaxies. We assume here that separating $D_M(z)$ into two components is enough.

field lines collapse, presumably quickly,[8] and the field gets trapped inside the structure. Hence, for a single growing system, $D_i(z)$ should essentially be a Heaviside step-function, $D_i(z) = \begin{cases} \varepsilon \text{ for } z \geq z_{crit} \\ 0 \text{ for } z < z_{crit} \end{cases}$, where ε is the ratio of the system baryonic mass to the universe baryonic mass. Since elements of a family, e.g. galaxies, reach m_{crit} at different z_{crit}, and since structures grow as $(1+z)^{-1}$, the overall $D_i(z)$ is the convolution of m_{crit}-weighted step-functions with the m_{crit} probability distribution. This one is a Gaussian of width τ and centered at average $\langle z_{crit} \rangle$ because of the initial ($z \gg z_{crit}$) normal distribution of mass inhomogeneities. If the m_{crit} and z_{crit} distributions are not strongly correlated, the resulting convolution is similar to a Fermi–Dirac (FD) function (see Fig. 2), which we will use to conveniently model and study $D_i(z)$. In addition to the FD function that encompasses the process of trapping fields into the systems, an exponential term is added to account for a possible release of the trapped field, e.g. as galaxies evolve to more symmetric morphologies or as superclusters form. An exponential form is chosen because the release process is the reverse of the trapping process modeled by the FD function, and $\mathrm{FD}(x) \to e^x$ for large x. In all, the form for $D_i(z)$ that we will use is:

$$D_i(z) \simeq \left[1 - \left(1 + e^{(z-z_{i0})/\tau_i}\right)^{-1}\right] + \left[A_i e^{-B_i z}\right], \tag{26}$$

where the parameters z_{i0}, τ_i, A_i and B_i are determined from the timeline of the formation of the structures of family i. Their interpretation is as follow: $z_{i0} \equiv \langle z_{crit} \rangle$ is the average z at which the set of structures i is forming; τ_i is the average duration (in z's scale) that such formation takes; A_i quantifies the relative amount of structures i that evolved their shapes into more isotropic ones; and B_i quantifies how fast this process is.

First we determine these parameters for $D_g(z)$. Approximating that galaxies grow most of their mass between $z_{g,b} = 15$ and $z_{g,e} = 3$ (b stands for "begin" and e for "end"), and evolve to their more symmetric shapes for $z < z_{g,e}$, the FD function is centered at $z_{g0} = (z_{g,b} + z_{g,e})/2 = 9 \pm 1$. The ± 1 comes from assuming 10% uncertainties on $z_{g,b}$ and $z_{g,e}$. The parameter τ_g characterizes the transition width, with $\mathrm{FD}(z_0 - \tau) \approx \mathrm{FD}(z_0)/2$. Setting $2\tau_g = z_{g0} - z_{g,e}$, i.e. with $\mathrm{FD}(z_{g,e}) \approx 0.1$ so that there, the trapping has essentially ended, yields $\tau_g = 3 \pm 0.5$. At $z < z_{g,e}$, galaxies become more symmetric. For example, at $z \simeq 3$, the ratio of elliptical to disk galaxies is negligible, growing to about 50% at $z = 0$ [21,22]. This releases some of the field trapped in the galaxies which results in a restrengthening of gravity. Considering that a disk galaxy traps most of its field, while it is mostly released in elliptical ones, and that the ratio of elliptical to disk galaxies is about 0.5, one has $D_g(z=0) \simeq 0.5$, which corresponds to $A_g \simeq 0.4$. However, elliptical galaxies usually belong to clusters and the released intragalactic field may be re-trapped between galaxy pairs. Choosing $A_g = 0.1 \pm 0.1$ accounts for this. It yields $D_g(z=0) \simeq 0.2 \pm 0.1$. Choosing $B_g = (4 \pm 1) z_{ge}$ makes restrengthening significant only for $z \lesssim 0.1$. Finally, the last galactic parameter in Eq. (25) is R_g, the present baryonic mass fraction contained in galaxies. One has $R_g = 0.15 \pm 0.10$.

We now turn to $D_c(z)$: if groups and clusters contained perfectly homogeneous and isotropically distributed gas, extragalactic field would not be trapped inside groups or clusters. Then, $D_c(z)$ would represent the field trapped between groups or clusters rather than inside these structures, and one would have $R_c = 1 - R_g$. To this relation, we add a term β to account for gas anisotropy, $R_c = 1 - R_g - \beta$. We assume $\beta = R_g$, i.e. that the effect concerns a mass similar to that which has already coalesced in galaxies. This is a small correction since most field trapping occurs between the groups and clusters. Lastly, one needs to consider that groups and clusters are now arranging themselves in superstructures more homogeneous than their uniformly scattered initial distribution. This releases some of the field trapped between groups and clusters. Thus in all, $D_c(z)$ has the same form as $D_g(z)$, given by Eq. (26). As for the galaxy case, the parameters for $D_c(z)$ are $z_{c0} \simeq (z_{c,b} + z_{c,e})/2$ and $\tau_c \simeq (z_{c0} - z_{c,e})/2$. Setting $z_{c,b} = 10 \pm 1$ (when groups/clusters start to coalesce) and $z_{c,e} = 1.2$ (when superclusters start forming) yields $z_{c0} = 5.6 \pm 1$ and $\tau_c = 2.2 \pm 0.5$. As for galaxies, $B_c = 4 z_{g,e} = 4.8 \pm 1.6$. A_c is difficult to assess. $A_c = 0.3 \pm 0.15$ is tentatively chosen.

Finally, ξ in Eq. (25) accounts for possible field trapping before galaxies started forming, i.e., for $z \gg 15$, e.g., field trapping in between the homogeneities that would latter trigger the growth of large structures, or that which lead to the first (non-galactic) supermassive stars at $z \approx 15$. Since ξ represents a small effect, if any, we assume $\xi = 0.9 \pm 0.1$.

Putting together the elements of Eq. (25) produces the result shown in Fig. 1. The $D_M(z)$ obtained for the nominal values of parameters in Eqs. (25) and (26) is shown by the line. The width of the band comes from the uncertainties on these parameters, *viz* it reflects the current state of our knowledge of the evolution of large structures, and of the relative amount of matter associated with each structure type. We can see the individual effects of the parameters uncertainties in Fig. 2: it displays $D_M(z)$ for the nominal values of the parameters (central line), and for the upper and lower values of one of the parameters while the others are kept nominal (two other lines). The total width of the band is obtained

[8] This expectation is based the prompt transition seen in QCD between the string-like regime (collapsed field-lines) and the nearly free-field
Footnote 8 continued
regime (field lines spreading isotropically from a charge) once the field coupling has reached a critical value [7]. In fact, the field line collapse transition in QCD has been parameterized with a FD function, see Ref. [23].

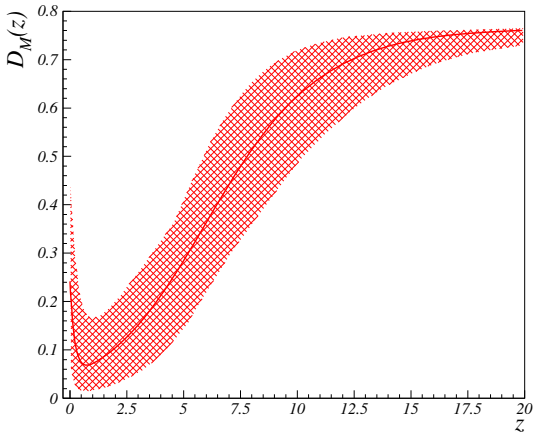

Fig. 1 Depletion factor $D_M(z)$, as constrained by the timeline of large structure formation and by the relative amount of baryonic matter pertaining to each type of structure. The width of the band represents the uncertainties on the value of the parameters of $D_M(z)$. The central line is obtained for the nominal values of the parameters

by adding the effects of the uncertainties.[9] We see that the uncertainties from τ_c, z_{g0}, R_g and ξ dominate.

In the bottom right panel of Fig. 2 we show the result of using the convolution of a Heaviside step-function with a Gaussian, instead of using a FD function. The same nominal values of the parameters are used (z_{i0} and τ_i now being, respectively, center and the full width of the gaussian for family i).

Also shown is the simplest choice for $D_i(z)$: to use, instead of a FD function, a function linear between $z_{i,b}$ and $z_{e,b}$ and constant otherwise.

The results using these different functional forms are close.

In the next section, the factor $D(z)$ just obtained is used in the luminosity distance, Eq. (23), to interpret the supernova observations without requiring $\Lambda \neq 0$ nor modifying laws of gravity or dynamics.

5 Comparison with observations

The compelling observations suggestive of dark energy are: (1) luminosity distance measurements with supernovae; (2) the age of the Universe; (3) large structure formation; (4) the cosmic microwave background (CMB); (5) baryon acoustic oscillations (BAO). We are concerned here with the foremost evidence, (1), and only sketch how observations (2)–(5) may also be explained. Addressing them in details is beyond the scope of a single article.

[9] Since the correlations between the observations used to determined the values and uncertainties of the parameters are unclear, we conservatively added linearly the uncertainties rather than quadratically.

5.1 Supernova observations

Explaining supernova observations with GR's self-interaction is the focus of this article. These observations are that the large-z ($0.1 \lesssim z \lesssim 1.5$) supernova apparent luminosities are dimmer, *viz* their apparent magnitudes are larger, than expected from a homogeneous and isotropic decelerating Universe. This is interpreted as evidence for an accelerating universe, i.e., for $\Lambda > 0$. However, we show in this section that lifting the approximations of homogeneity and isotropy can also explain the observations, while keeping $\Lambda = 0$.

From the luminosity distance $\mathcal{D}_L(z)$, Eq. (23), the apparent magnitude $\mathcal{D}_L(z)H_0$ of events can be calculated. Assuming $\Lambda = 0$, taking $H_0 = 68 \pm 1$ km/s/Mpc [3], and using the depletion factor $D_m(z)$ modeled in Sect. 4 (see Fig. 1), we compute the band shown in Fig. 3. Its width stems from propagating that of $D_M(z)$. Our calculation agrees well with the γ-ray bursts [25] and supernovae [26–29] data. There is no adjustment to these data, all the parameters in D_M being constrained by observations of large structure evolution. Also shown in the figure are the calculations for the cases of a homogeneous and isotropic universe with only matter (dotted line), that for an empty universe (continuous line) and the ΛCDM (dark energy, cold dark matter) model (dashed line).

The difference between the observations and the expectation from a homogeneous and isotropic universe with $\Lambda = 0$ is clearer by forming a residual apparent magnitude:

$$r(z) = 5\log\left(H_0\frac{\mathcal{D}_L(z)}{1+z}\right) - 5\log\left(H_0\frac{z+z^2/2}{1+z}\right), \qquad (27)$$

with the last term corresponding to the empty universe case. Positive values of $r(z)$ indicate fainter apparent luminosities than expected in the case of an empty Universe. They constitute the best evidence for $\Lambda > 0$. Our calculation of $r(z)$ with $\Lambda = 0$ agrees well with the observations, see Fig. 4.

We now outline how GR's self-interaction may also explain the observations providing less direct evidence for $\Lambda > 0$.

5.2 Age of the universe

Without $\Lambda > 0$, the calculated age of the universe would be 11.7 ± 0.2 Gyr for the standard $\Omega_M = 0.32$ value and for $H_0 = 68 \pm 1$ km/s/Mpc [3]. This conflicts with the measured age of the oldest stars, up to $\sim$ 13.5 Gyr. The ΛCDM model, with $\Omega_\Lambda = 0.68$ and the same H_0 and Ω_M values, yields 13.6 ± 0.2 Gyr. GR's self-interaction also solves this problem, while keeping $\Lambda = 0$: Eq. (24) yields a compatible universe age of 13.2 ± 1.7 Gyr.

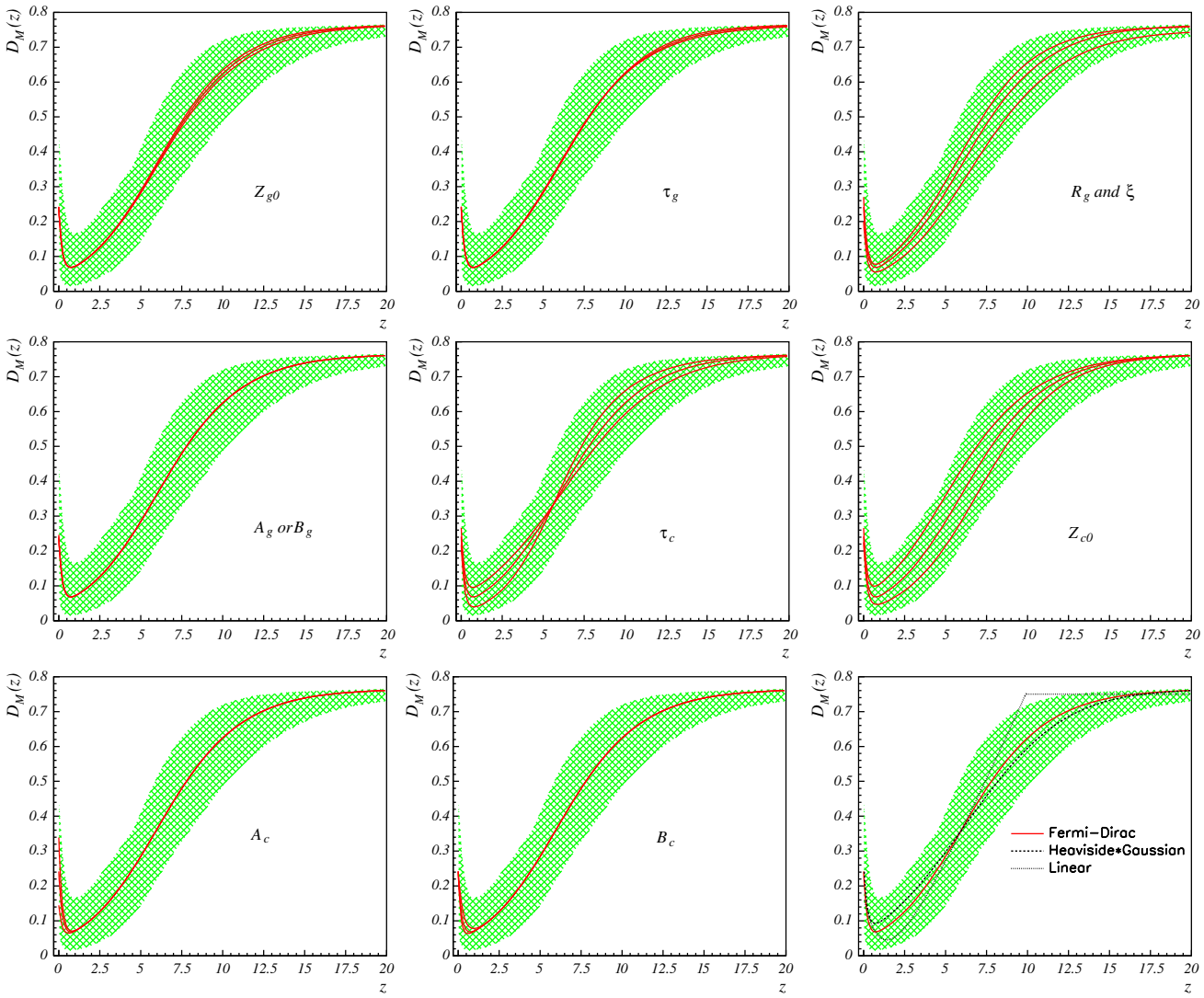

Fig. 2 Individual and total contributions of the parameter uncertainties to the depletion factor $D_M(z)$. The central line shows $D_M(z)$ for the nominal values of the parameters. The two other lines (sometimes masked by the central line) correspond to upper and lower values of the parameter labelled in the panel, the other parameters being kept at their nominal values. The band, the same as in Fig. 1, results from linearly propagating the effects of all the uncertainties. The bottom right panel shows the results of using the convolution of a Heaviside function with a Gaussian function (dashed line) or a simple linear function (dotted line) instead of a Fermi–Dirac function

5.3 Large structure formation

In a universe without gravitational self-interaction or dark matter, large structures do not have time to coalesce. What happens in the self-interaction framework can be sketched as follow: As $D_M(z)$ departs from 1, *viz* as gravity weakens globally, energy conservation demands that the global weakening is balanced locally by an increase of gravity within the structures themselves, thus speeding up their formation compared to a universe without self-interaction.

Since $D_M(z)$ evolves following the formation of large structures, gravity strengthens locally with the same timeline. Because strengthening reproduces the dynamics of galaxies and clusters [4], the local effect of self-interaction is equivalent to the effect of dark matter. Furthermore, the position of the peak of the matter power spectrum is now given by $k_{eq} = H_0\sqrt{2\Omega_M^*(0)/a_{eq}}$, with a_{eq} the scale parameter at z_{eq}. Assuming $\Omega_{Baryon} = \Omega_M$ (no dark matter) and using $\Omega_M^* = \Omega_M D_M$ yield $\Omega_M^*(0) \simeq 0.3$, i.e. $k_{eq} \simeq 0.014$, in agreement with observations [30]. This suggests that the present approach is compatible with large structure formation.

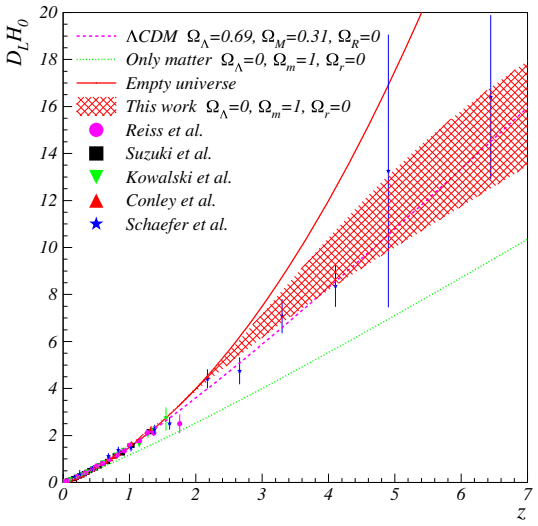

Fig. 3 Apparent magnitudes of γ-ray bursts (star symbol) and supernovae (other symbols). Larger $\mathcal{D}_L H_0$ values correspond to fainter observed events. The dashed line is the expectation from the ΛCDM model. The dotted line is the case of a universe with only matter and with the traditional approximation of homogeneity and isotropy. The continuous line shows the case of an empty Universe. The band is the present work (universe containing only baryonic matter, with gravity field partially trapped in massive systems due to field self-interaction). It has no free parameters adjusted to the γ-ray or supernova data

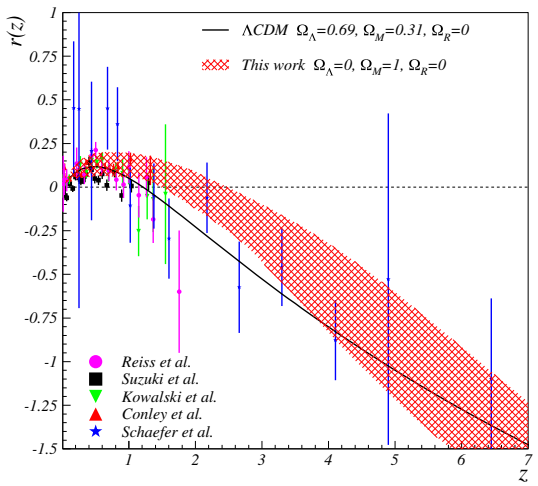

Fig. 4 Residual between observed apparent magnitudes (γ-ray bursts: star symbol. Supernovae: other symbols) and their expectation from an empty universe. The continuous line is for the ΛCDM model. The band is the present work, without any free parameters adjusted to the γ-ray or supernova data

5.4 CMB and BAO

The CMB main acoustic peak position depending on the Universe dynamical evolution, its calculation in the present framework involves Ω_M^* rather than Ω_M. Thus we have now $\theta \simeq \sqrt{\Omega_M^*/z_{rec}}$ (with $z_{rec} \simeq 1100$ at the recombination time), resulting in $\theta \simeq 0.8°$, which agrees with observations [24]. Predicting the smaller features of the CMB and the BAO is complex and, like for large structure formation, beyond the scope of this first article.

5.5 Other consequence

Field trapping naturally explains the cosmic coincidence, i.e. that in the ΛCDM model, dark energy's repulsion currently nearly compensates matter's attraction, while repulsion was negligible in the past and attraction is expected be negligible in the future [6]. No natural explanation exists within ΛCDM for this apparently fortuitous coincidence. In the present approach, structure formation depletes attraction and thus, compensating it with a repulsion, *viz* dark energy, is unnecessary. Thus, there is no coincidence and hence no need for explanation.

The QCD analogy to the cosmic coincidence is that instead of accounting for the color field confinement in hadrons, one would introduce an exotic repulsive force to nearly counteract the strong force as it supposedly propagates outside hadrons.

6 Summary

The Lagrangian of General Relativity contains field self-interaction terms that become important for very massive systems. Their effects are unaccounted for in the studies of galaxies and galaxy clusters since the dynamical studies of these systems rely on Newton's law of gravity. Accounting for field self-interaction locally strengthens gravity's binding, thereby making dark matter superfluous. In turn, the stronger binding in the system must be balanced by a weakening of gravity outside the system, as demanded by energy conservation. This weakening is neglected in studies of the universe evolution because its equation is derived using assumptions –homogeneity and isotropy– that suppress the effects of self-interaction, and furthermore disregard by definition local phenomena that could affect gravity's field, such as field trapping.

In this article, a modified Friedmann equation effectively accounting for self-interaction is derived from Einstein's field equation. Then, the luminosity distance formula necessary to interpret the supernova data at large redshift z is derived. Gravity's weakening is folded into a global factor $D_M(z)$ that is modeled using a physically motivated function: it is constrained by the limit conditions $D_M(z \gg 1) \approx 1$ because of the homogeneity and isotropy of the early Universe, and $D_M(z \approx 0) \ll 1$ because the growth of large structures has enabled the effects of field self-interaction. The transition between the two limits is determined by considering that structures grow linearly with $(1+z)^{-1}$. The characteristic z when the transition occurs, and the length of the period over which it occurs, are constrained by the timeline of structure

formations. We used different functions for $D_M(z)$ that conform to the above constraints, and obtained similar results.

Using the luminosity distance accounting for gravity's weakening and the modeled $D_M(z)$, the large-z supernova and γ-ray burst data are explained without requiring dark energy. No free parameters are adjusted to these data: the effect of gravity's weakening is determined by our knowledge of large structure formation.

This approach thus explains the main observation suggestive of dark energy without requirements beyond the standard forces and laws of physics. The basic mechanism used here is in fact well-studied: a similar increase of force at short range, and its consequent suppression at long range, occurs for the strong nuclear interaction, another self-interacting force whose Lagrangian is similar to that of General Relativity. Other direct consequences of this approach are an explanation for the missing mass in galaxies and galaxy clusters without requiring dark matter, flat rotation curves for disk galaxies and the Tully–Fisher relation. The direct connection between galactic missing mass and the suppression of the Universe's deceleration eliminates the cosmic coincidence problem.

Acknowledgements The author thanks S. J. Brodsky, F. X. Girod-Gard, C. Munoz-Camacho, A. Sandorfi, S. Širca, E. Smith, B. Terzić and X. Zheng for useful discussions.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: This is a theoretical article, without data to be deposited.]

Funded by SCOAP[3].

References

1. A.G. Riess et al., AJ **116**, 1009 (1998)
2. S. Perlmutter et al., ApJ **517**, 565 (1999)
3. C. Patrignani, *et al.*, Chin. Phys. C, 40, (2016) (10000 and 2017 update)
4. A. Deur, PLB **676**, 21 (2009)
5. A. Deur, EPJC **77**(6), 412 (2017)
6. A. Zee, *Quantum Field Theory in a Nutshell* (Princeton University Press, Princeton, 2003)
7. A. Deur, S.J. Brodsky, G.F. de Teramond, PPNP 90, 1; 2016. PLB **757**, 275 (2016)
8. R.A. Hulse, J.H. Taylor, Astrophys. J. **195**, L51 (1975)
9. A. Deur, MNRAS **438**, 1535 (2014)
10. D. Clowe et al., ApJL **648**, 109 (2006)
11. R.B. Tully, J.R. Fisher, A&A **54**, 661 (1977)
12. S.S. McGaugh, J.M. Schombert, G.D. Bothun, W.J.G. de Blok, Astrophys. J. **533**, L99 (2000)
13. T. Regge, Nuovo Cim. **14**, 951 (1959)
14. G.S. Bali, K. Schilling, C. Schlichter, Phys. Rev. D **51**, 5165 (1995)
15. P.D. Mannheim, Prog. Part. Nucl. Phys. **56**, 340 (2006)
16. T. Buchert, Gen. Relat. Grav. **40**, 467 (2008)
17. T. Buchert, C. Nayet, A. Wiegand, PRD **87**(12), 123503 (2013)
18. A. Dressler et al., ApJ **490**, 577 (1997)
19. M. Postman et al., ApJ **623**, 721 (2005)
20. O.H. Parry, V.R. Eke, C.S. Frenk, MNRAS **396**, 1972 (2009)
21. S. van den Bergh, PASP **114**, 797 (2002)
22. T. Kuutma, A. Tamm, E. Tempel, A&A **600**, L6 (2017)
23. S.J. Brodsky, G.F. de Teramond, A. Deur, Phys. Rev. D **81**, 096010 (2010)
24. G. Hinshaw et al., ApJS **208**, 19 (2013)
25. B.E. Schaefer, ApJ **660**, 16 (2007)
26. G. Miknaitis et al., ApJ **666**, 674 (2007)
27. M. Kowalski et al., ApJ **686**, 749 (2008)
28. A. Conley et al., ApJS **192**, 1 (2011)
29. N. Suzuki et al., ApJ **746**, 85 (2012)
30. M. Tegmark et al., PRD **74**, 123507 (2006)