Self-interactions and Spontaneous Black Hole Scalarization

Caio F. Macedo,†,∗ Jeremy Sakstein,‡,† Emanuele Berti,§,† Leonardo Gualtieri,¶,§ Hector O. Silva,¶,† and Thomas P. Sotiriou¶,†,‡∗∗

1Campus Salinópolis, Universidade Federal do Pará, Salinópolis, Pará, 68721-000, Brazil
2Center for Particle Cosmology, Department of Physics and Astronomy, University of Pennsylvania, 209 S. 33rd St., Philadelphia, PA 19104, USA
3Department of Physics and Astronomy, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA
4Dipartimento di Fisica “Sapienza” Università di Roma & Sezione INFN Roma1, Piazzale Aldo Moro 5, 00185, Roma, Italy
5ExtreMe Gravity Institute, Department of Physics, Montana State University, Bozeman, MT 59717 USA
6School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, UK
7School of Physics and Astronomy, University of Nottingham, University Park, Nottingham, NG7 2RD, UK

(Dated: March 19, 2019)

It has recently been shown that nontrivial couplings between a scalar and the Gauss-Bonnet invariant can give rise to black hole spontaneous scalarization. Theories that exhibit this phenomenon are among the leading candidates for testing gravity with upcoming black hole observations. All models considered so far have focused on specific forms for the coupling, neglecting scalar self-interactions. In this work, we take the first steps towards placing this phenomenon on a more robust theoretical footing by considering the leading-order scalar self-interactions as well as the scalar-Gauss-Bonnet coupling. Our approach is consistent with the principles of effective field theory and yields the simplest and most natural model. We find that a mass term for the scalar alters the threshold for the onset of scalarization, and we study the mass range over which scalarized black hole solutions exist. We also demonstrate that the quartic self-coupling is sufficient to produce scalarized solutions that are stable against radial perturbations, without the need to resort to higher-order terms in the Gauss-Bonnet coupling function. Our model therefore represents a canonical model that can be studied further, with the ultimate aim of developing falsifiable tests of black hole scalarization.

I. INTRODUCTION

The era of gravitational-wave observations has arrived. For the first time we can see the universe in gravitational waves as well as optically, and this new window affords us the opportunity to test gravity in extreme spacetimes for the first time. The LIGO/Virgo collaboration has already detected ten black hole (BH) mergers [1, 2] and one neutron star merger [3]. The latter has proved incredibly powerful for testing and constraining infrared modifications of gravity [4–10] (if the modifications are important for the late-time cosmology [11]), but ultraviolet (UV) modifications are more difficult to test. This is partly due to the numerical and theoretical challenges that arise when extending computations of merger events to theories beyond general relativity (GR), but also due to a theoretical roadblock: the no-hair theorems [12–14] (see e.g. [15–19] for reviews). These preclude the existence of nontrivial scalar hair (or scalar charges) for BHs, and so the dynamics of theories including new scalar degrees of freedom (i.e. scalar-tensor theories) is similar to GR. One possible way forward is to instead use neutron stars as probes of UV modifications of GR [15, 20–28]. These are far more complicated objects since the equation of state for nuclear matter is presently unknown, and, unlike BHs, neutron stars have higher-order multipole moments that give rise to strong tidal effects. On the observational side, LIGO/Virgo has observed more BH mergers than neutron star mergers [2], and this may well remain the case, even as more gravitational-wave detectors come online and the existing ones are upgraded to improved sensitivities.

A. Black hole spontaneous scalarization and effective field theory

The considerations above have motivated a theoretical effort to find UV-modifications of GR that can circumvent the no-hair theorems by violating some of their assumptions. Some of these theories exhibit spontaneous BH scalarization [29, 30], a phenomenon where both the GR BH solution and novel BH solutions with scalar hair can exist. The phenomenon has been predicted for static [29, 30] and, more recently, charged [31, 32] BHs. This allows for the possibility that, even if all LIGO/Virgo detections to date have been compatible with GR, future detections could be consistent with scalarized BH solutions.

The fundamental interaction responsible for scalarization is the coupling between a scalar field $\phi$ and the Gauss-Bonnet invariant $G = R^2 - 4R_{ab}R^{ab} + R_{abcd}R^{abcd}$, so that the action has the form

$$S = \frac{1}{2} \int d^4x \sqrt{-g} \left[ \frac{R}{8\pi G} - \frac{1}{2} \nabla_a \phi \nabla^a \phi + f(\phi)G \right],$$

where we are using units where $\hbar = c = 1$, so that the Planck mass $M_P = (8\pi G)^{-1/2}$. In subsequent sections we will switch to units where $8\pi G = c = 1$, which is more suited to (and more common in) the study of BH solutions. In Eq. (1) we have
chosen the same normalization and conventions as in Ref. [33] (modulo an overall sign in the definition of the Riemann tensor) and in Refs. [30, 34–36], while the scalar field $\phi^{\text{DY}}$ in Refs. [29, 37] has a different normalization: $\phi^{\text{DY}} = \phi / 2$. A canonically normalized scalar field $\phi^{\text{can}}$ is such that $\phi^{\text{can}} = \phi / \sqrt{2}$.

Reference [30] proved a no-hair theorem for scalar-Gauss-Bonnet theories under certain conditions. Scalarization may occur when these conditions are violated. The essential requirement is that the coupling function $f(\phi)$ has at least one stationary point at some $\phi = \hat{\phi}$ such that $f(\hat{\phi}) = 0$. GR BHs correspond to solutions with $\phi = \hat{\phi}$, but this configuration may be unstable for certain BH masses or model parameters. When this is the case, the field rolls away, and the BH acquires scalar hair. Apart from this requirement, there is no guiding principle for choosing $f(\phi)$. The patent choice $f(\phi) = \phi^2 / 2M^2$ [30] (where $M$ is a new mass scale) produces scalarized BHs that are unstable to radial perturbations [37]. This can be resolved by including higher-order terms, in particular $f(\phi) = \phi^2 / 2M^2 + c\phi^4 / M^4$ [35, 36], or by assuming a more complicated function: for example, exponential couplings $f(\phi) = \exp(\beta \phi^2 / 2M^2)$ with both positive and negative signs for $\beta$ have been considered in the literature [29, 38].

These solutions are somewhat unsatisfactory from a theoretical perspective. Since we lack a UV-completion for these models, it would be more appropriate to construct the theory using the principles of effective field theory (EFT) [39–41]. From this perspective, relying on higher-order corrections to the coupling function in order to stabilize the BH solutions implies that higher-dimensional operators are competing with (supposedly leading) lower-dimensional operators. This suggests that operators that have been omitted can be just as important, and therefore these solutions are outside the range of validity of the EFT. Moreover, without any enhanced symmetry protecting the form of special choices of the coupling functions (and the action in general), it is likely that these theories are radiatively unstable. We note that there is currently no known enhanced symmetry of the exponential or quartic couplings, though this is by no means a proof that there cannot be one. Similarly, it is possible that such couplings arise as a truncation of a UV-complete theory and just appear fine-tuned from an IR perspective [42].

In the coming decade and beyond, LIGO/Virgo will be upgraded to higher sensitivities and additional detectors will come online. Hundreds or thousands of detections are anticipated, and it therefore behooves us to make theoretical predictions from robust models that are stable from a QFT point of view. The main purpose of this paper is to take a first step towards placing the phenomenon of BH spontaneous scalarization on a more robust theoretical foundation by constructing the theory using EFT principles.

When viewed as an EFT, spontaneous scalarization is a phenomenon occurring in theories where a $\mathbb{Z}_2$-symmetric scalar (i.e. the action is invariant under $\phi \to -\phi$) is coupled to a massless spin-2 particle. We should therefore build our action out of operators that are invariant under this symmetry. In particular, the leading-order (relevant and marginal) operators are not Gauss-Bonnet couplings, but rather a mass term and a quartic self-interaction. One should supplement these with irrelevant operators suppressed by some cut-off scale $\mathcal{M}$, which will include a quadratic scalar-Gauss-Bonnet coupling at lowest order. For this reason, we will mainly study the action

$$S = \frac{1}{2} \int d^4x \sqrt{-g} \left[ M_0^2 R - \frac{1}{2} \nabla_a \phi \nabla^a \phi - \frac{1}{2} \mu \phi^2 - \frac{1}{2} \phi^4 + \frac{\phi^2}{2M^2} G \right].$$

Later on, we will also include the quartic scalar-Gauss-Bonnet coupling in order to provide a different and well-studied stable model against which we can compare the effects of the self-interactions. This coupling is higher order and it will not give the leading order effect from an EFT perspective, but it is included for the purpose of comparison with previous studies [29, 30, 36, 37]. Boson star and BH solutions have recently been studied in a similar class of theories [43, 44]. Scalar-tensor theories with a self-interacting potential, but without a Gauss-Bonnet term, were considered in [45–47]. Of course, there are other operators that one could write down, such as a term $\propto \phi^6$ in the potential or scalar-curvature couplings such as $\phi^2 R$, but in this work we will restrict our focus to understanding how scalarization works when only the leading-order operators (including the leading-order scalar-Gauss-Bonnet coupling) are included, since this is the minimal input required to produce the phenomenon. We postpone the more arduous task of determining the unique set of dimension-six operators that contribute to this theory, and a full explorative study of the resultant parameter space, for future work.

**B. Executive summary**

In this article we study the existence, stability, and properties of scalarized BHs in the theory defined by the action (2), including the subtleties and conceptual issues that arise due to the inclusion of a mass for the scalar.

We find (as previously noted in [44]) that including a mass term for the scalar alters the threshold for the onset of scalarization. Most notably, we find that the quartic self-interaction is sufficient to stabilize some scalarized BHs, and higher-order scalar-Gauss-Bonnet couplings are not required. For this reason, and because the theory is a robust EFT, the action (2) represents the leading canonical model with which to study spontaneous BH scalarization.

The action in Eq. (2) uses units where $\hbar = c = 1$, which are useful for understanding the theory from an EFT perspective. For the purposes of calculating, it is more convenient to use geometrized units where $8\pi G = c = 1$. Furthermore, we will

\footnote{The leading-order scalar-graviton coupling is $\phi^2 R$, but this does not lead to BH scalarization, so we will not include it in this work. This operator does not contribute to the scalar’s equation of motion on a Ricci-flat GR solution, which means it cannot alter the threshold for the onset of BH scalarization. Note however that it can contribute to the effective mass on a scalarized BH or a neutron star background.}
rescale the field so that $\phi$ is dimensionless by defining (before the change of units) $\phi = M_{\text{Pl}} \varphi$. In the new units, the action (2) reads

$$S = \frac{1}{2} \int d^4x \sqrt{-g} \left[ R - \frac{1}{2} \nabla_a \varphi \nabla^a \varphi - V(\varphi) + f(\varphi) \xi \right],$$

(3)

where the potential is

$$V(\varphi) = \frac{1}{2} \mu^2 \varphi^2 + \frac{1}{2} \lambda \varphi^4,$$

(4)

the coupling function is

$$f(\varphi) = \frac{1}{8} \eta \varphi^2,$$

(5)

and $\eta$ has units of $[\text{Length}]^2$. Note that the parameters $\mu$ and $\lambda$ appearing in the potential (4) have units of $[\text{Length}]^{-1}$ and $[\text{Length}]^{-2}$, respectively. In order to compare our results with those of [36], in part of our analysis we will also include a quartic term in the coupling function, i.e. $f(\varphi) = (\eta \varphi^2 + \zeta \varphi^4)/8$, where $\zeta$ has units of $[\text{Length}]^2$ (see Appendix B). Finally, note that in this work we will only consider $\mu^2 > 0$. One could consider $\mu^2 < 0$, which would give a global minimum of the potential at some $\varphi \neq 0$. This sign choice would require the addition of a cosmological constant to cancel the net vacuum energy at the new minimum in order for the theory to admit asymptotically-flat spacetimes. Since the aim of this work is to discern the effects of the scalar self-interactions on the canonical model of spontaneous scalarization, we prefer not to include this more technical, and quantitatively different possibility.

With these conventions in place, we now summarize the main results of this work. First, we introduce the dimensionless mass and scalar charge of the solutions

$$\tilde{M} \equiv M/\eta^{1/2}; \quad \tilde{Q} \equiv Q/\eta^{1/2},$$

(6)

as well as the dimensionless coupling parameters

$$\tilde{\mu} \equiv \mu \eta^{1/2}; \quad \tilde{\lambda} \equiv \lambda \eta; \quad \tilde{\zeta} \equiv \zeta/\eta.$$

(7)

For the purpose of understanding the changes introduced by the scalar potential, it will be useful to introduce an effective potential, which is spacetime dependent:

$$V_{\text{eff}}(\varphi) = \frac{1}{2} \left( \mu^2 - \frac{\eta}{4} \xi \right) \varphi^2 + \frac{1}{2} \lambda \varphi^4,$$

(8)

The equation of motion for the scalar is $\Box \varphi = V_{\text{eff}}(\varphi)$. In particular, there is an effective mass for the scalar about the point $\varphi = 0$

$$m_{\text{eff}}^2 = \mu^2 - \frac{\eta}{4} \xi; \quad m_{\text{eff}} = m_{\text{eff}}/\eta^{1/2},$$

(9)

where we have defined a dimensionless effective mass for later convenience. Close to the BH, the contribution of the Gauss-Bonnet invariant dominates and the Schwarzschild solution is unstable due to a tachyonic instability (recall that $\xi \sim M^2/r^6$ for the Schwarzschild metric). This gives rise to spontaneous scalarization. Further away, the Gauss-Bonnet contribution is negligible, and the effective mass is positive.

A summary of our results is as follows:

- **Effects of the mass term (left panel of Fig. 1):** The main effect of the mass term is to alter the threshold for the onset of scalarization, as already noted in [44]. The dimensionless mass threshold $\tilde{M}$, below which scalarization is possible, is studied in Sec. III, and it is plotted as a function of $\tilde{\mu}$ in the left panel of Fig. 1 for solutions where the radial profile of the scalar field has no nodes (black, solid line) and one node (red, dashed line). Only the nodeless solutions are radially stable. Note that $\tilde{M}$ is a decreasing function of $\tilde{\mu}$. This can be qualitatively understood from the effective potential (8): $m_{\text{eff}}$ grows with $\tilde{\mu}$ (at fixed $\xi$), so that the tachyonic instability...
The existence of a global stable minimum should resolve the concerns raised in reference [48], where it was shown that quantum fluctuations could trigger the tachyonic instability during inflation. In our model, the field would begin, and remain, at the global minimum for the duration of inflation and play no role in its dynamics (the field’s mass would be much larger than the Hubble scale so that the field does not fluctuate). That being said, inflation occurs at energies far higher than the cut-off of the effective field theory for spontaneous scalarization (10^{-20} GeV for scalarized solar mass BHs), and it is not clear that the range of validity of any current model exhibiting scalarization can be extended to the early Universe.

**Effects of the quartic self-interaction (right panel of Fig. 1):** The quartic self-interaction stabilizes scalarized BH solutions with respect to radial perturbations. At fixed $\hat{\mu}$, all scalarized solutions with $\hat{\lambda} < \hat{\lambda}_{\text{crit}}$ are unstable, and we conjecture that gravitational collapse will generally lead to a Schwarzschild solution since these are always stable for $\hat{M} > \hat{M}_c$. This corresponds to the region on the right of the dotted vertical lines in Fig. 5 below. When $\hat{\lambda} > \hat{\lambda}_{\text{crit}}$, stable scalarized BH solutions are possible: these are the solid lines on the left of the dotted vertical lines in Fig. 5, and we conjecture that they are the end-state of gravitational collapse. The threshold value $\hat{\lambda}_{\text{crit}}(\hat{\mu})$ is shown in the right panel of Fig. 1. Qualitatively, this can be understood as follows: For scalarized solutions, the effective mass for the scalar [Eq. (9)] is tachyonic, at least in some region of spacetime, and therefore the scalar tends to grow from its scalarized value. Introducing a quartic term in the effective potential (8) bounds the effective potential from below, so that there is a stable minimum about which the effective mass (9) is positive and the solution is globally stable. This is also the reason why a quartic Gauss-Bonnet coupling can stabilize the scalarized solutions [35, 36], although in the Gauss-Bonnet case the coefficient of the $\varphi^4$ term is also spacetime-dependent.

The existence of a global stable minimum should resolve the concerns raised in reference [48], where it was shown that quantum fluctuations could trigger the tachyonic instability during inflation. In our model, the field would begin, and remain, at the global minimum for the duration of inflation and play no role in its dynamics (the field’s mass would be much larger than the Hubble scale so that the field does not fluctuate). That being said, inflation occurs at energies far higher than the cut-off of the effective field theory for spontaneous scalarization (10^{-20} GeV for scalarized solar mass BHs), and it is not clear that the range of validity of any current model exhibiting scalarization can be extended to the early Universe.

**Mass range for scalarization and maximum scalar charge (Fig. 2):** For any given choice of the theory parameters $(\hat{\mu}, \hat{\lambda})$, marginally stable scalarized BH solutions correspond to a minimum in the BH mass $\hat{M}$ and a maximum in the scalar charge $\hat{Q}$; cf. again Fig. 5 below. (This maximum charge refers to stable BHs; unstable BHs can have larger charges, but they are unphysical.) In the left panel of Fig. 2 we focus on nodeless solutions, and we plot: (i) the $\hat{\lambda}$-independent threshold mass $\hat{M} = \hat{M}_c(\hat{\mu})$ below which scalarization is possible (thick, gray line); (ii) the minimum dimensionless mass $\hat{M}_{\text{min}}(\hat{\mu})$, below which both Schwarzschild and scalarized BH solutions are unstable, for selected values of $\hat{\lambda}$. The mass range in which stable, scalarized BH solutions can exist becomes narrower as $\hat{\mu}$ increases.

Summarizing: for $\hat{M} > \hat{M}_c$, the Schwarzschild solution is stable, while scalarized BH solutions are unstable to radial perturbations; for $\hat{M}_\text{min} < \hat{M} < \hat{M}_c$ there is at least one stable $n = 0$ scalarized BH, while the Schwarzschild solution (and the $n > 0$ scalarized BH solutions) are unstable; finally, for $\hat{M} < \hat{M}_\text{min}$, all BH solutions are unstable. The existence of a minimum BH mass is a common feature in theories with scalar-Gauss-Bonnet coupling (see e.g. the cases of Einstein-dilaton Gauss-Bonnet gravity [33, 49] and of shift-symmetric Gauss-Bonnet gravity [50]), although in theories that do not exhibit scalarization (such as these) the minimum mass is due to the inability to satisfy a regularity condition at the horizon.

The right panel of Fig. 2 shows the maximum dimensionless scalar charge $\hat{Q}_{\text{max}}(\hat{\mu})$ for selected values of $\hat{\lambda}$. The most relevant feature here is that, for all values of $\hat{\lambda}$ that we investigated, $\hat{Q}_{\text{max}}(\hat{\mu})$ has a local maximum $\sim 0.15$: this near-universal maximum value of the scalar charge is of phenomenological interest, because the dipolar radiation in BH binaries (which is potentially measurable by gravitational-wave interferometers) is proportional to the difference between the BH charges [15, 51, 52].
C. Plan of the paper

The paper is organized as follows. In section II we present the equations of motion resulting from the action (3) and analyze their properties. In section III we investigate the effect of a nonzero scalar mass on the threshold for the onset of scalarization. We accomplish this by studying the limit in which the scalar is decoupled from the metric equations of motion, i.e., we consider the linearized field equations for a scalar field propagating on a Schwarzschild background. In section IV we move beyond this “decoupling limit” and solve the coupled metric-scalar equations numerically in order to confirm the results of our linear analysis. We also calculate the properties of the scalarized solutions, including their stability to radial perturbations. The study of radial perturbations in scalar-Gauss-Bonnet theories is by now standard (cf. [35–37]), so we do not rederive the formalism in this work. In section V we summarize our results and discuss possible directions for future work.

II. FIELD EQUATIONS AND SCALARIZED SOLUTIONS

The modified Einstein equations can be obtained by extremizing the action (3) with respect to the metric and the scalar field, with the result

\[ G_{ab} = T^\varphi_{ab} - \frac{1}{2} K_{ab}, \]  

(10)

\[ \Box \varphi = V_{\varphi} - f_{\varphi} \mathcal{G}, \]  

(11)

where

\[ T^\varphi_{ab} = \frac{1}{2} \delta_{\alpha \beta} \partial_a \partial_b \varphi - \frac{1}{2} g_{ab} \left[ \frac{1}{2}(\partial_c \varphi)^2 + V(\varphi) \right], \]  

(12)

\[ K_{ab} = 2 g_{(a} \epsilon_{b)\mu} \epsilon^{\nu \mu j k} \nabla_j f^{\nu} \epsilon_{\varphi \varphi}, \]  

(13)

where \( V(\varphi) \) is given in equation (4), \( f(\varphi) = (\eta \varphi^2 + \zeta \varphi^4)/8 \), and \( * R_{\mu \nu} = \epsilon^{a b c d} R_{\mu b c d} \).

As discussed in [29, 30], Schwarzschild solutions exist in scalar Gauss-Bonnet theories provided that there is some \( \bar{\varphi} \) such that \( f'(\bar{\varphi}) = 0 \). In our model, \( \bar{\varphi} = 0 \). Allowing for a nonzero value of the background scalar may have important phenomenological consequences for gravitational wave astronomy, as pointed out in the context of Einstein-Maxwell-dilaton theory [53], and we plan to revisit this assumption in future work.

We focus on static, spherically symmetric BHs. In this case the line element and the scalar field read

\[ ds^2 = -A(r) dr^2 + B(r)^{-1} dr^2 + r^2 d\Omega^2, \]  

(14)

\[ \varphi = \varphi_0(r), \]  

(15)

where \( d\Omega = d\theta^2 + \sin^2 \theta d\phi^2 \) is the line element on a 2-sphere.

The field equations can be obtained by substituting Eqs. (14) and (15) into Eqs. (10) and (11). We show the equations in Appendix A. We also make them available online through a Mathematica notebook [54].

The field equations must be supplemented by boundary conditions. Spherically symmetric BHs have an event horizon \( r_h \) where the functions \( A \) and \( B \) vanish and the scalar field tends to a constant:

\[ A(r \approx r_h) \approx a_1 (r - r_h) + O[(r - r_h)^2], \]  

(16)

\[ B(r \approx r_h) \approx b_1 (r - r_h) + O[(r - r_h)^2], \]  

(17)

\[ \varphi_0(r \approx r_h) \approx \varphi_{0h} + O[(r - r_h)]. \]  

(18)

These conditions impose a restriction on the derivative of the scalar field at the horizon:

\[ \frac{d\varphi_0}{dr}\big|_{r=r_h} = a^{-1} \left( b + c \sqrt{\Lambda} \right), \]  

(19)

where \( a, b, \) and \( c \) are functions of \( r_h, \varphi_{0h}, \) and of the parameters of the theory. The explicit expression of these functions is given in Appendix A. The important quantity is \( \Lambda \), which is given by

\[ \Lambda = 1 - 6 \frac{\varphi_{0h}}{r_h^2} \left( \eta + \zeta \varphi_{0h}^2 \right)^2 \left\{ 1 - \frac{1}{2} \frac{\varphi_{0h}^2}{r_h^2} \left( \eta + \zeta \varphi_{0h}^2 \right) \left( \mu^2 + 2 \lambda \varphi_{0h}^3 \right) \right. \]

\[ - \left. \frac{r_h^2}{6} \frac{\varphi_{0h}^2}{r_h^2} \left( \mu^2 + \lambda \varphi_{0h}^2 \right) \left( 1 + \frac{1}{16r_h^2} \eta \varphi_{0h}^3 + \zeta \varphi_{0h}^4 \right)^2 \right\}. \]  

(20)

When \( \Lambda < 0 \) it is not possible to enforce regularity at the horizon. Ref. [50] studied this regularity condition for shift-symmetric scalar-Gauss-Bonnet gravity, showing that there is a naked singularity when the condition is violated. Thus, \( \Lambda > 0 \) is a necessary condition for the existence of BH solutions.

By expanding the field equations for large \( r \) we obtain

\[ A(r \gg r_h) \approx 1 - 2M/r, \]  

(21)

\[ B(r \gg r_h) \approx 1 - 2M/r, \]  

(22)

\[ \varphi_0(r \gg r_h) = Q e^{-\mu r}/r, \]  

(23)

where \( M \) is the ADM mass, \( Q \) is an integration constant, and we have set the cosmological value of the scalar field to zero. In the \( \mu \to 0 \) limit, the scalar field decays like \( \varphi \sim 1/r \), and the constant in front of \( 1/r \) is typically referred to as the “scalar charge.” Strictly speaking, \( Q \) is not a conserved charge (even when \( \mu = 0 \)), but we will follow conventions and refer to it as such from here on.

Typically, in scalar-tensor theories one must set the scalar field’s mass such that the force range is sub-micron (for \( O(1) \) couplings), or else the theory will fail laboratory and solar system tests of GR [55–63]. Therefore one would expect the spacetime outside the BH to rapidly approach the Schwarzschild metric, thereby suppressing any deviations from GR. This logic follows from scalar-gravity couplings of the form \( \varphi R \), which, in the absence of any screening mechanisms, give rise to Yukawa forces. The coupling considered in our model \( \eta \varphi^2 \mathcal{G} \)
is expected to appear at high post-Newtonian order in the weak-field limit (provided $\eta/M^2_\odot$ is not too large), and therefore the theory is compatible with Solar System tests of GR [64, 65]. Furthermore, since it is unlikely that weakly gravitating objects like the Sun and the Earth are scalarized, gravity in the Solar System should behave identically to GR. For these reasons, we will not place any restrictions on the mass of the field in this work. One could imagine completing the EFT by adding a term proportional to $\varphi^2 R$ into the action, which is not forbidden by the symmetries, and which we have ignored in this work for the sake of simplicity. Such couplings could give rise to Yukawa-like forces, but (again) only if the Sun or the Earth is scalarized, which is unlikely to be the case, with the exception of extreme couplings [66, 67]. The situation would be different if the asymptotic field value were different from zero.

III. SCHWARZSCHILD RADIAL STABILITY AND THE SCALARIZATION THRESHOLD

In Sec. IV we will explore the BH solutions of the theory. Before doing so, we first wish to understand whether such solutions can exist as a result of instabilities of the ordinary Schwarzschild solution to linear perturbations.

The Schwarzschild metric with a vanishing scalar field is a solution of Eqs. (10) and (11). We can study the radial stability of the Schwarzschild spacetime by considering perturbations of the field equations of the form $\varphi = \epsilon \varphi_1(r)e^{-i\omega t}/r$, where $\epsilon$ is a small bookkeeping parameter. From the scalar field equation (11) we find

$$\frac{d^2 \varphi_1}{dr^2} + (\omega^2 - V_{\text{eff}}) \varphi_1 = 0,$$

where

$$V_{\text{eff}} = \left(1 - \frac{r_h}{r}\right) \left(\frac{r_h}{r^3} + \mu^2 - \frac{3 \eta G}{r^6}\right).$$

This equation involves only the field’s mass $\mu$ and the strength of the Gauss-Bonnet coupling $\eta$. Therefore, higher-order terms in the scalar potential and in the coupling function do not have any influence on the stability of the Schwarzschild spacetime. In particular, the threshold for scalarization is independent of $\lambda$.

To investigate the radial stability of the Schwarzschild spacetime, we solve Eq. (24) by requiring that the field vanishes at the BH horizon and at infinity [36]. Since the equation is real, the eigenvalue $\omega^2$ is also real, and $\omega^2 < 0$ corresponds to unstable modes [36, 37, 68]. The critical threshold value $\tilde{M} = \tilde{M}_t$ for which scalarization can occur corresponds to solutions of Eq. (24) with eigenvalue $\omega = 0$, indicating a transition between stable and unstable states. The condition $\omega^2 = 0$ is satisfied by different values of $\tilde{M}$, corresponding to scalarized solutions with $n = 0, 1, \ldots$ nodes in the scalar field profile. We denote the threshold value for the $n = 0$ solution by $\tilde{M}_t$.

In the left panel of Fig. 1 we show $\tilde{M}_t$ as a function of the mass $\tilde{\mu}$ of the field (we also show the threshold values for

![FIG. 3. Frequency of the unstable modes of Schwarzschild BHs in our theory.](image)

the $n = 1$ solution, which, as discussed in the introduction, is always smaller than the threshold mass $\tilde{M}_t$ for $n = 0$). One can see that the threshold for scalarization $\tilde{M}_t$ decreases with increasing $\tilde{\mu}$. This can be understood by considering the effective mass for the scalar given in Eq. (9): larger values of $\mu$ require the product $\eta G$ to be larger in order to induce the tachyonic instability. The instability is therefore harder to realize for larger scalar masses.

By solving Eq. (24) we can also investigate the instability time scale as a function of $\tilde{M}$. In Fig. 3 we show the normalized frequency for unstable modes, $2M\omega_j$, as a function of the parameter $\tilde{M}$. The Schwarzschild solution is stable ($\omega_j < 0$) in the region $\tilde{M} > \tilde{M}_h$, where $\tilde{M}_h$ is the value corresponding to the intersection of $2M\omega_j$ with the x-axis of this plot. The three cases studied here correspond to the blue dots in the left panel of Fig. 1. It is therefore plausible that hairy solutions should exist in the region $\tilde{M} < \tilde{M}_h$, where the Schwarzschild BH is unstable. This expectation will be confirmed in Sec. IV below.
solutions with marginal stability, which correspond to the minimum mass, maximally charged scalarized BH for the given solutions: the upper branch (dashed lines) is unstable to radial perturbations, whereas the lower branch (solid lines) is stable. Blue dots mark solutions with marginal stability, which correspond to the minimum mass, maximally charged scalarized BH for the given ($\hat{\mu}$, $\lambda$).

IV. SCALARIZED BLACK HOLE SOLUTIONS AND RADIAL STABILITY

In this section we solve the fully nonlinear equations to construct scalarized solutions, and check their stability under linear radial scalar and tensor perturbations. This is accomplished as follows. First, we integrate the field equations onwards starting from the horizon, where we impose the conditions (16)–(19). By matching the numerical solutions with Eqs. (21)–(23) in the far region ($r \gg r_h$), we can extract the BH mass $\hat{M}$ and the scalar charge $\hat{Q}$. This procedure gives us the unperturbed solution. Next we check stability. The linearized field equations for radial perturbations follow from the unperturbed solution. Next we check stability. The linearized field equations for radial perturbations follow from the ansatz

$$\varphi = \varphi_0 + \epsilon \varphi_1,$$

$$ds^2 = [A + \epsilon E_r(t, r)]dr^2 + [B^{-1} + \epsilon F_r(t, r)] + r^2d\Omega^2,$$

where ($A$, $B$, $\varphi_0$) are functions of $r$ which satisfy the zeroth-order (background) field equations. By inserting Eqs. (26) and (27) into the field equations (10) and (11) and expanding up to first order, one can show that the equations for the perturbation functions reduce to a single second-order equation of the form

$$h(r) \frac{\partial^2 \varphi_1}{\partial t^2} - \frac{\partial^2 \varphi_1}{\partial r^2} + k(r) \frac{\partial \varphi_1}{\partial r} + p(r) \varphi_1 = 0,$$

(see Appendix A and the supplemental Mathematica notebook [54]) where the coefficients ($h$, $k$, $p$) depend only on the background quantities and on $r$ (cf. [33, 36, 37]). Eq. (28) can be further manipulated to reduce it to a Schrödinger-like form, but since this step is not necessary to analyze the stability of the system, and generates more complicated coefficients, we prefer not to display it here (see [37] for details). A mode analysis can be performed by looking for solutions of the form

$$\varphi_1(t, r) = \varphi_1(r)e^{-\lambda t},$$

and by imposing the requirement that $\varphi_1(\lambda)$ vanishes at the horizon and at infinity when searching for unstable modes. These requirements (as in Sec. III) result in an eigenvalue problem for $\omega^2 < 0$.

Before applying this process in general, it is instructive to perform a preliminary comparative study in order to discern how self-interactions affect the stability of scalarized solutions. In Fig. 4 we fix $\hat{\mu} = 0.05$ and we compare the normalized imaginary mode for the scalarized solutions with the corresponding calculation for the Schwarzschild case, as presented in Fig. 3. When $\hat{\lambda} \leq 0.2$, both the modes of the scalarized solutions (dashed red) and the Schwarzschild modes (solid gray) converge to zero when $\hat{M} = \hat{M}_c$. However, for $\hat{\lambda} > 0.2$ the modes tend to zero when $\hat{M} = \hat{M}_{\text{min}}$ and $\hat{Q} = \hat{Q}_{\text{max}}$, and we found no unstable modes for BHs with $\hat{M} > \hat{M}_{\text{min}}$ and $\hat{Q} < \hat{Q}_{\text{max}}$. We note also that the unstable mode frequencies typically decrease as $\hat{\lambda}$ increases, implying stability on longer time-scales. Qualitatively similar conclusions apply to other values of $\hat{\mu}$.

The main results of our integrations are presented in Fig. 5, where we show scalarized solutions in the $(\hat{M}, \hat{Q})$ plane for representative values of $\hat{\mu}$ and $\hat{\lambda}$. The dotted vertical line represents the threshold for the stability of the Schwarzschild solution, $\hat{M} = \hat{M}_c$. Solid lines correspond to radially stable solutions, while dashed lines correspond to radially unstable solutions. Note that we use different conventions for radial stability with respect to Refs. [36, 37], where solid and dashed lines have the opposite meaning.

When $\hat{\lambda} = 0$, all scalarized solutions are in the region $\hat{M} > \hat{M}_c$, where the Schwarzschild solution is stable. These scalarized solutions are radially unstable, and it is plausible that Schwarzschild BHs will be the end-state of gravitational collapse. As $\hat{\lambda}$ increases, the solutions move into the region where $\hat{M}_{\text{min}} < \hat{M} < \hat{M}_c$; the minimum mass $\hat{M}_{\text{min}}$ corresponds to the blue dots in Fig. 5. Schwarzschild BHs are unstable in this region, so the BH can support a nontrivial scalar provided the scalarized solutions are stable. For $\hat{M} < \hat{M}_{\text{min}}$, both Schwarzschild and scalarized BHs are unstable.

Our analysis reveals that the quartic self-interaction can stabilize scalarized solutions with a quadratic scalar-Gauss-
Bonnet coupling up to some maximum scalar charge \( \hat{Q} \), beyond which the solutions are unstable. Interestingly, it is possible to have two scalarized solutions (in addition to the unstable Schwarzschild solution) at fixed \( \hat{M} \), provided that \( \hat{A} \) is large enough. In such cases, the solution with larger \( \hat{Q} \) is unstable, and is expected to decay to the solution with smaller \( \hat{Q} \), which is stable.

The main result of this section is that we do not need more exotic scalar-Gauss-Bonnet couplings to stabilize the scalarized solutions: leading-order scalar self-interactions are sufficient. From an EFT perspective, these models are better-motivated.

V. DISCUSSION AND CONCLUSIONS

Black hole spontaneous scalarization is so far the only known mechanism that allows BHs to possess scalar hair only if their mass is below a certain threshold. Theories that allow for this phenomenon are prime candidates for modelling deviations from GR that have so far avoided detection but can be tested using current and future gravitational wave observations. It therefore behooves the theoretical community to devise robust, stable theories that exhibit BH scalarization. To date, all studies in the literature are not consistent effective field theories since they ignored leading-order terms that are compatible with the underlying symmetries of the theory. The aim of the present work is to take the first steps towards realizing the phenomenon within robust and well-motivated theories.

In this paper, we have presented the simplest model that exhibits spontaneous scalarization by viewing the theory as one of a \( \mathbb{Z}_2 \)-symmetric scalar and writing down all of the leading-order (relevant and marginal) operators, as well as the leading-order coupling of the scalar to the Gauss-Bonnet invariant required to produce scalarized BHs. In practice, this is tantamount to including a mass and quartic self-interaction for the scalar, so that the theory includes a massive scalar with a \( \phi^4 \)-potential and a quadratic coupling of the scalar to the Gauss-Bonnet invariant.

Our analysis has revealed that spontaneous scalarization persists in this bottom-up construction. We have demonstrated that static scalarized solutions exist and, furthermore, that they are stable to radial perturbations. This model (possibly augmented by a \( \phi^2 R \) coupling, and other dimension-six operators) therefore represents the leading candidate model with which to explore spontaneous scalarization. In future studies, we intend to take this program forward by studying rotating BHs and neutron stars, and by understanding the stability and dynamics of these compact objects in full generality. The ultimate aim of this program is to predict theoretically sound observational signatures that can be used to test GR in the strong-field regime with upcoming gravitational wave observations.

ACKNOWLEDGMENTS

This work was supported by the H2020-MSCA-RISE-2015 Grant No. StronGrHEP-690904 and by the COST action CA16104 “GWverse”. H.O.S was supported by NASA grants NNX16AB98G and 80NSSC17M0041. J.S. was supported by funds provided to the Center for Particle Cosmology by the University of Pennsylvania. E.B. is supported by NSF Grants No. PHY-1841464, AST-1841358, PHY-090003, and NASA ATP Grant No. 17-ATP17-0225. C.F.B.M. would like to thank Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for partial financial support, and the Johns Hopkins University for kind hospitality during the preparation of this work and the American Physical Society which funded the visit through the International Research Travel Award Program.

Appendix A: Spherical black holes in scalar Gauss-Bonnet gravity

From the nontrivial components of the zeroth-order Einstein equations, we obtain

\[
(t, t) : \quad B \left\{ (\phi_0')^2 \left[ 16(B - 1)f_{\phi_{0}00} - r^2 \right] + 16(B - 1)\phi_{00}'' f_{\phi_{0}0} - 4 \right\} - 4B' \left[ 2(1 - 3B)\phi_{0}0 f_{\phi_{0}} + r \right] - r^2 V(\phi_{0}) = 0, \tag{A1}
\]

\[
(r, r) : \quad \frac{1}{4} A \left[ B \left[ 4 - r^2 (\phi_0')^2 \right] + r^2 V(\phi_{0}) - 4 \right] + BA' \left[ 2(1 - 3B)\phi_{0}0 f_{\phi_{0}} + r \right] = 0, \tag{A2}
\]

\[
(\theta, \theta) : \quad - rBA'' (r - 4B\phi_{0}0 f_{\phi_{0}}) + 4rB^2 A' \phi_{00}'' f_{\phi_{0}} - \frac{1}{2} r^2 A \left[ B\phi_{0}^2 + V(\phi_{0}) \right]
+ \frac{1}{2A} \left[ rBA^2 (r - 4B\phi_{0}0 f_{\phi_{0}}) \right] + A' \left[ rB \left[ 4B\phi_{0}0 f_{\phi_{0}} + (B - 1)BA'' f_{\phi_{0}} \right] - \frac{1}{2} rB' \left( r - 12B\phi_{0}0 f_{\phi_{0}} \right) \right] - rAB' = 0, \tag{A3}
\]

where a prime indicates differentiation with respect to \( r \). The equation for the background scalar field is

\[
AB\phi_{0}'' + \frac{1}{2} A' \left[ BA' + A \left( \frac{4B}{r} + B' \right) \right] + \frac{4}{r^2} \left( (B - 1)BA'' f_{\phi_{0}} \right)
- \frac{1}{2r^2 A} \left\{ 4A' \left[ (B - 1)BA' + A(1 - 3B)B' \right] f_{\phi_{0}} + r^2 A^2 V(\phi_{0}) \right\} = 0. \tag{A4}
\]
The equations above can be recast as a system of three differential equations: two first-order equations for $A$ and $B$, and one second-order equation for $\varphi_0$. These are integrated as explained in Sec. II, in units where $r_h = 1$, changing the parameters $(\eta, \zeta, \lambda, \mu)$ in each integration. As noted in Ref. [30], only some values of the parameters allow for scalarized solutions. With these we construct the background solutions shown in Fig. 5.

Specializing to the quartic coupling and the quartic potential, the coefficients appearing in the condition on the scalar field derivative at the horizon of Eq. (19) are given by

$$a = \frac{\varphi_0}{r_h} \left( \eta + \zeta \varphi_0^2 \right) \left\{ -4 + 2 \varphi_0 \left( \zeta \varphi_0^2 + \eta \right) \left( 2 \lambda \varphi_0^2 + \mu^2 \right) + \frac{\varphi_0^2}{r_h^2} \left( \lambda \varphi_0^2 + \mu^2 \right) \right\}, \quad (A5)$$

$$b = \frac{\varphi_0^2}{r_h^2} \left( \lambda \varphi_0^2 + \mu^2 \right) - 4, \quad (A6)$$

$$c = 4 - \varphi_0^2 r_h^2 \left( \lambda \varphi_0^2 + \mu^2 \right) + \frac{3 \varphi_0^4}{r_h} \left( \zeta \varphi_0^2 + \eta \right) \left( 2 \lambda \varphi_0^2 + \mu^2 \right) - \frac{1}{4} \varphi_0 \left( \zeta \varphi_0^2 + \eta \right) \left( 2 \lambda \varphi_0^2 + \mu^2 \right)^2 - 4 \varphi_0 \left( \zeta \varphi_0^2 + \eta \right) \left( 2 \lambda \varphi_0^2 + \mu^2 \right)^2. \quad (A7)$$

The equations describing the perturbations can be obtained by expanding the Einstein-scalar system up to first order. The nonzero components of the perturbed Einstein equations are $(t, t), (t, r), (r, r),$ and $(\theta, \theta)$. Additionally, we have one more equation from the first-order expansion of the scalar field equation. These five equations can be manipulated to obtain (28) with a procedure similar to the one presented in [37, 68]. Instead of showing the explicit form of the differential equations, which are rather lengthy, we provide a companion Mathematica notebook which shows the nontrivial components of the first-order Einstein equations and the procedure to obtain Eq. (28) from these equations [54].

**Appendix B: Self-interactions within quartic Gauss-Bonnet coupling**

Ref. [36] showed that scalar-Gauss-Bonnet theories with $V = 0$ and a quartic coupling term

$$f(\varphi) = \frac{1}{8} \left( \eta \varphi^2 + \zeta \varphi^4 \right)$$

with $\zeta / \eta < 0$ can also generate stable BH solutions. A natural question is whether the combined effects—the quartic potential and the quartic coupling—can work together to stabilize BHs. While this is out of the scope of the EFT picture, with the quartic coupling being a sub-leading operator compared with the quartic self-interaction, here we investigate this issue as a complement to our main results.

In Fig. 6 we show the scalarized BH solutions considering $(\hat{\mu}, \hat{\lambda}) = (0.05, 0.4)$ for different values of $\hat{\zeta}$. As expected, the quartic term in the coupling still helps to generate stable BH solutions, even when the self-interaction potential is present. We note that this case also exhibits a minimum mass $M_{\text{min}}$ and a maximum charge $Q_{\text{max}}$. Unlike theories with $V = 0$: cf. Fig. 2 of [36].

[1] B. P. Abbott et al. (LIGO Scientific, Virgo), Phys. Rev. Lett. 116, 061102 (2016), arXiv:1602.03837 [gr-qc].
[2] B. P. Abbott et al. (LIGO Scientific, Virgo), (2018), arXiv:1811.12907 [astro-ph.HE].
[3] B. P. Abbott et al. (LIGO Scientific, Virgo), Phys. Rev. Lett. 119, 141101 (2017), arXiv:1709.09660 [gr-qc].
[4] J. Sakstein and B. Jain, Phys. Rev. Lett. 119, 251303 (2017), arXiv:1710.05893 [astro-ph.CO].
[5] J. M. Ezquiaga and M. Zumalacárregui, Phys. Rev. Lett. 119, 251304 (2017), arXiv:1710.05901 [astro-ph.CO].
[6] P. Creminelli and F. Vernizzi, Phys. Rev. Lett. 119, 251302 (2017), arXiv:1710.05877 [astro-ph.CO].
[7] T. Baker, E. Bellini, P. G. Ferreira, M. Lagos, J. Noller, and I. Sawicki, Phys. Rev. Lett. 119, 251301 (2017), arXiv:1710.06394 [astro-ph.CO].
[8] A. Dima and F. Vernizzi, Phys. Rev. D97, 101302 (2018), arXiv:1712.04731 [gr-qc].
[9] M. Crisostomi and K. Koyama, Phys. Rev. D97, 021301 (2018), arXiv:1711.06661 [astro-ph.CO].
[10] D. Langlois, R. Saito, D. Yamauchi, and K. Noui, Phys. Rev. D97, 061501 (2018), arXiv:1711.07403 [gr-qc].
[11] N. Franchini and T. P. Sotiriou, (2019), arXiv:1903.05427 [gr-qc].
[12] S. W. Hawking, Commun. Math. Phys. 25, 167 (1972).
[13] T. P. Sotiriou and V. Faraoni, Phys. Rev. Lett. 108, 081103 (2012), arXiv:1109.6324 [gr-qc].
[14] L. Hui and A. Nicolis, Phys. Rev. Lett. 110, 241104 (2013), arXiv:1202.1296 [hep-th].
