Dynamical Instability of Self-Gravitating Membranes

Huan Yang,†, Béatrice Bonga,‡ and Zhen Pan

1Perimeter Institute for Theoretical Physics, Ontario, N2L 2Y5, Canada
2University of Guelph, Guelph, Ontario N1G 2W1, Canada
3Institute for Mathematics, Astrophysics and Particle Physics, Radboud University, 6525 AJ Nijmegen, The Netherlands

We show that a generic relativistic membrane with in-plane pressure and surface density having the same sign is unstable with respect to a series of warping mode instabilities with high wave numbers. We also examine the criteria of instability for commonly studied exotic compact objects with membranes, such as gravastars, AdS bubbles and thin-shell wormholes. For example, a gravastar which satisfies the weak energy condition turns out to be dynamically unstable. A thin-layer black hole mimicker is stable only if it has positive pressure and negative surface density (such as a wormhole), or vice versa.

Introduction. The detection of binary black hole (BH) mergers with ground-based gravitational-wave (GW) detectors [1–4], images of the supermassive BHs M87 & Sgr A* with radio interferometry [5, 6], and the observation of S stars orbiting a small dark region in the galactic center [7], all point to the existence of BHs, of which a description was first obtained by Karl Schwarzschild more than one hundred years ago using General Relativity. The study of BHs is not limited to astrophysics and General Relativity, but also plays a role in other major areas of physics, such as quantum fields and strings, condensed matter physics and quantum information. Because of its unparalleled conceptual and observational importance, it is paramount to test the more refined features of BHs against its unparalleled conceptual and observational importance, it is paramount to test the more refined features of BHs against all viable alternatives allowed by the laws of nature. Any signal, e.g. ringdown quasinormal modes [8, 9], that favors a BH mimicker over BHs themselves would represent a fundamental breakthrough/revolution in physics. In the coming decades the third-generation ground-based GW detectors [10, 11], the space-borne GW detectors [12, 13] and the next-generation Event Horizon Telescope, will likely improve the precision of such tests by many orders.

Horizonless compact objects are important candidates for BH mimickers [14]. One class of them, such as boson stars, has smooth distributions of matter/fields that are convenient for stability analysis and numerical simulations. However, it appears difficult to construct stable configurations of these compact stars that approach the compactness of BHs. For example, a fluid star with causal equation of state can achieve maximum compactness at around $M/R \approx 0.355$ [15, 16], with $M$ its mass and $R$ its radius as measured from the surface area. The bound for boson stars is around 0.44 [17]. There are proposals for constructing compact stars with anisotropic stress [18–24] to increase the maximum compactness, but they often feature problems such as superluminal sound speed, violation of energy conditions and lack of stability analysis. A recent study showed that the bound can be improved to $\sim 0.376$ by including various prescriptions of elastic stress [25].

Another class of compact objects often include a (or multiple) membrane(s) that separates spacetime regions, such as gravastars [26] and thin-shell wormholes [27, 28]. These are interesting because this type of construction allows the transition to the exterior spacetime, which is the same as the BH spacetime, to be arbitrarily close to the horizon of a corresponding BH. Therefore, these models can have compactness arbitrarily close to that of BHs. In addition, membranes are often invoked if there is interesting physics happening near a certain surface, such as proposals considering hard structures near BH horizons motivated by firewalls or 2–2 holes [29, 30]. Moreover, compact objects with membranes are expected to have distinct strong-gravity dynamic behavior from more uniform compact objects. For this latter reason, two-dimensional domain walls have been extensively studied in cosmology.

Here, we present a perturbation study of self-gravitating membranes with nontrivial energy and stress. We find that if the signs of the in-plane pressure and surface density in the membrane are the same, there is a generic warping instability for modes with sufficiently high wave numbers. We apply these results to commonly studied compact objects, and find that a significant portion of the parameter space of gravastars — which are usually modeled by a de Sitter interior and Schwarzschild exterior with a spherical shell of matter at the boundary — and AdS bubbles (with anti-de Sitter interiors) are dynamically unstable. Static thin-shell wormholes always have positive pressure and negative surface density, so that they are free from these instabilities. Therefore, requiring membranes to have negative pressure (for positive surface density) and positive pressure (for negative density) becomes a powerful qualifier for the stability of compact objects.

Throughout this work, we adopt geometric units ($c = G = 1$).

Membrane instability. Let us consider a self-gravitating membrane with intra-surface pressure surrounded by vacuum. If the pressure is positive, any local vertical displacement results in an “anti-restoring” force that pushes the mass element away from equilibrium, see Fig.1. On the other hand, the gravitational attraction from surrounding mass elements tends to bring it back to equilibrium. We shall show that the anti-spring force always wins in the eikonal limit, leading to a series of instabilities with high wave number. To illustrate the basic picture, we present the analysis in the Newtonian regime first before proceeding to the relativistic case.

Consider a membrane placed in the $(x – y)$ plane, with surface density $\sigma$ and surface pressure $P$. The displacement field...
for the analysis presented in both the Newtonian and relativistic regime, we use a coordinate transformation to map the membrane to the equatorial plane of the new coordinates to facilitate the derivation of the equation of motion.

\[ \xi(x, y) \] can be decomposed as

\[ \xi = \xi_x \hat{e}_x + \xi_y \hat{e}_y + \xi_z \hat{e}_z. \]  

(1)

We use \( \delta \) to denote Eulerian perturbations and \( \Delta \) to denote Lagrangian perturbations. For example, the Eulerian density fluctuation is given by \( \delta \sigma = -\nabla \cdot (\sigma \xi) \), where \( \nabla \) here operates on the two horizontal directions, and the Lagrangian density perturbation is given by \( \Delta \sigma = \delta \sigma + \xi \cdot \nabla \sigma \). For the purpose of this analysis, we only need to consider the case with \( \xi_z \) nonzero, in which case the local area change of mass elements is second order in \( \xi \), i.e., \( \sigma = \Delta \sigma = 0 \) at linear order.

The equation of motion for three-dimensional fluid elements, in terms of Lagrangian variables, can be written as [31]

\[ \rho_0 \left( \frac{\partial^2 \xi}{\partial t^2} + \nabla \Delta U - (\nabla \cdot \xi) \nabla U_0 \right) = \nabla \cdot \mathbf{F}, \]  

(2)

where \( \rho_0 \) is the unperturbed mass density, \( U_0 \) is the unperturbed gravitational potential and \( \mathbf{F} \) is the stress tensor so that the right hand side represents the hydrodynamical force acting on the fluid element. In other words, the left hand side of the equation is the kinetic term and the right hand side of the equation represents the external force. Similarly, for a mass element on a two-dimensional disk, we can write down the equation of motion as

\[ \sigma_0 \left( \frac{\partial^2 \xi}{\partial t^2} + \nabla \Delta U - (\nabla \cdot \xi) \nabla U_0 \right) = F_{\text{disk-in}} + F_{\text{disk-out}}, \]  

(3)

where \( \sigma_0 \) is the unperturbed surface mass density and \( \Delta U \) is the Lagrangian potential perturbation. Since we only consider the vertical displacement, \( \xi \) is divergence-free \( \nabla \cdot \xi = 0 \), i.e., there are no density perturbations. In the equilibrium case, \( U_0 \) satisfies \( \nabla^2 U_0 = 0 \) except at the disk plane, where the vertical derivative is discontinuous:

\[ \frac{\partial U_0}{\partial z_+} - \frac{\partial U_0}{\partial z_-} = 4\pi \sigma_0. \]  

(4)

The right hand side of Eq. (42) can be obtained by integrating Eq. (2) across the membrane (for details, see the Supplementary Material ). It can also be derived directly from the membrane configuration — as we motivate here — since it physical represents external forces. The external force is given by two components of disk forces. The in-plane component is generated by the pressure variation and the tilt of the disk plane:

\[ F_{\text{disk-in}} = -\nabla \Delta P + (\nabla P \cdot \mathbf{n}) \hat{e}_z, \]  

(5)

where \( \mathbf{n} = \hat{e}_z - \partial_x \xi \hat{e}_x - \partial_y \xi \hat{e}_y = \hat{e}_z + \Delta \mathbf{n} \) is the normal vector to the disk. The pressure perturbation is related to the density perturbation through the disk equation of state: \( \Delta P / P = \Gamma_1 \Delta \sigma / \sigma [32] \), where \( \Gamma_1 \) depends on the equation of state and the nature of the perturbation (e.g., adiabatic or isothermal). Therefore, the Lagrangian pressure perturbation is zero for vanishing \( \Delta \sigma \). The off-plane disk force is due to the warping of the disk. If we imagine the local disk surface has a radius of curvature \( R \), then the magnitude of out-of-plane force is just \( 2P/R \). For general mean curvature \( \kappa \), we have

\[ F_{\text{disk-out}} = -P \kappa \mathbf{n}, \quad \text{with} \quad \kappa = \frac{\partial^2 \xi}{\partial x^2} + \frac{\partial^2 \xi}{\partial y^2}. \]  

(6)

In order to compute the potential perturbation \( \Delta U \), in particular, its value and derivatives on the disk plane, we make a coordinate transformation so that \( z' = z - \xi \), with \( x, y \) coordinates unchanged. The disk is mapped to the \( z' = 0 \)-plane in this new coordinate system, which is more convenient for solving the boundary value problem. Using

\[ \frac{\partial}{\partial x'} = \frac{\partial}{\partial x} + \partial_x \xi \frac{\partial}{\partial z}, \quad \frac{\partial}{\partial y'} = \frac{\partial}{\partial y} + \partial_y \xi \frac{\partial}{\partial z}, \quad \frac{\partial}{\partial z'} = \frac{\partial}{\partial z}, \]  

(7)

the original Laplace equation \( \nabla^2 U = 0 \) (for \( z \neq \xi_z \)) becomes

\[ \nabla'^2 U = 2 \partial_x \xi_z \frac{\partial^2 U}{\partial x^2} + 2 \partial_y \xi_z \frac{\partial^2 U}{\partial y^2} + \left( \partial_x^2 \xi_z + \partial_y^2 \xi_z \right) \frac{\partial U}{\partial z'}, \]  

(8)

with the matching conditions that \( \partial_+ U|_{z'} = \partial_- U|_{z'} = 4\pi \sigma \) and \( U|_+ = U|_- \). Because \( \xi \) is an infinitesimal displacement, we can write \( U \) as \( U_0 + U_1 \), with \( U_0 \) satisfying \( \nabla^2 U_0 = 0 \) together with \( \partial_+ U_0|_{z'} = \partial_- U_0|_{z'} = 4\pi \sigma \) and \( U_0|_+ = U_0|_- \). The solution of \( U_0 \) is obviously known, and \( U_1 \) may be obtained by solving

\[ \nabla'^2 U_1 = 2 \partial_x \xi_z \frac{\partial^2 U_0}{\partial x^2} + 2 \partial_y \xi_z \frac{\partial^2 U_0}{\partial y^2} + \left( \partial_x^2 \xi_z + \partial_y^2 \xi_z \right) \frac{\partial U_0}{\partial z'}, \]  

(9)

with \( \partial_+ U_1|_{z'} - \partial_- U_1|_{z'} = 0 \) and \( U_1|_+ = U_1|_- \), so that \( U_1 \) is completely regular in the entire spacetime. In particular, \( U_1 \)
evaluated on the disk surface can be mapped back to $\Delta U$ with $\Delta U := U(\zeta') - U_0(\zeta) = U_0(\zeta') + U_1(\zeta') - U_0(\zeta)$, and $\nabla \Delta U$ is the gravitational backreaction described in Eq. (42).

At this point, we consider a planar mode with $\xi \propto e^{k \cdot \mathbf{x}}$ in the eikonal limit, that is, $|k| \gg 1$. The right hand side of Eq. (42) is dominated by $F_{\text{disk-out}}$, which is proportional to $k^2$. On the other hand, as $U_1$ is also proportional to $e^{k \cdot \mathbf{x}}$ and the source term for $U_1$ in Eq. (9) is dominated by the term proportional to $k^2$, we have $U_1 \propto k^2 \nabla U_1 \propto k$ and $\partial_\zeta U_1 \propto k$ ($\partial_\zeta \propto k$ as $1/k$ is the only length scale in the problem). So the gravitational restoring force is subordinate compared to the anti-restoring force by the warping disk. The dispersion relation is approximately (with $\partial_\zeta \rightarrow -i \omega$)

$$\omega^2 \approx -P/\sigma_0 k^2 $$

which leads to exponential mode growth if $P/\sigma_0 > 0$.

**Relativistic case.** In the relativistic setting, we consider a model problem for compact objects with a membrane: an infi-

te plane with an infinite mass density. We discuss all the steps of the derivation of the equation of motion of the

membrane perturbations. Detailed manipulations are relegated to the Supplementary Material.

If we consider the spacetime of a gravastar or a thin-shell wormhole, the metric can be expressed as

$$ds^2 = \delta_{\mu\nu} dx^\mu dx^\nu = -U(z) dt^2 + U_\zeta(\zeta) dx^2 + U_p(\zeta)(dx^2 + dy^2),$$

where $x = \theta \cos \phi$ and $y = \theta \sin \phi$. This local representation of the membrane metric is generic. The Israel boundary conditions on the membrane relate the extrinsic curvature $K_{ij}$ to

the surface-layer property by [33] ($d\tau = \sqrt{U} dt$)

$$K_{i\zeta}^- = K_{j\zeta}^- = -\frac{1}{\sqrt{U_p}} \left| \frac{\partial P}{\partial \zeta} \right| = -4\pi \sigma,$$

$$K_{i\zeta}^+ = \frac{1}{\sqrt{U_p}} \left( \frac{\partial P}{\partial \zeta} \right) = 8\pi (P + \sigma/2)$$

where $|^+$ indicates the difference between $0_+$ and $0_-$ of the membrane in the $z$-direction. Since we can always rescale $z$

in the vertical/radial direction, in the rest of the discussion we shall set $U_\zeta = 1$.

Let us now assume the membrane is perturbed with vertical displacement $\xi_{\zeta} = \xi(x, y, t)$. The membrane stress energy tensor is
given by

$$\tau^{\mu\nu} = \delta(\zeta - \xi) \left[ (\sigma + P) n^\mu n^\nu + P (g^{\mu\nu} - n^\mu n^\nu) \right],$$

where $n$ is the normal vector of the membrane. It is given by $e(z - h_{zz}/2) - \sum_{\alpha=x,y} \xi_{\alpha} e_{\alpha}/8\alpha(0)$ and $e_\zeta$ is given by $\frac{\partial}{\partial \zeta}$ (similarly for $e_x$ and $e_y$), where $h_{zz}$ is sourced by the membrane motion (compare with the right hand side of Eq. (8)).

In order to derive the equation of motion for $\xi_{\zeta}$, we transform to the coordinate system with $\zeta' = z - \xi$, $t' = t$, $x' = x$, $y' = y$, such that the membrane is mapped back to the "equatorial" plane in the new coordinates. The spacetime metric in the new coordinates can be written as $g_{\mu\nu} = g_{\mu\nu}^{(0)} + \xi_{\mu\nu} + \xi_{\mu\zeta} + h_{\mu\nu} = g_{\mu\nu}^{(0)} + h_{\mu\nu}$, where $\mu$ represents the covariant derivative with respect to $g_{\mu\nu}^{(0)}$.

The gravitational perturbation is more conveniently computed in the original ($t, z, x, y$) coordinate system:

$$\tilde{h}_{\mu\nu} + 2R_{\mu\rho\sigma\zeta} \tilde{h}^{\rho\sigma} = 0,$$

with the trace-reversed $\tilde{h}_{\mu\nu} := h_{\mu\nu} - \frac{1}{2} h g^{(0)}_{\mu\nu}$ and assuming the Lorenz gauge condition $\tilde{h}_{\mu\zeta}|^{\omega} = 0$, which is preserved along the evolution driven by the wave equations if it is initially satisfied. The waves should be outgoing at infinity and the matching condition at the membrane leads to

$$\int_{-\infty}^{+\infty} dz \tilde{h}_{\mu\nu} \xi_{\zeta} = 8\pi \delta \tau_{\mu\nu}$$

or, more explicitly,

$$\frac{\sigma + P}{2U} (2\tilde{h}_{\zeta\zeta} - \tilde{h}_{tt}) = -\frac{P}{2U} \tilde{h}_{tt} + \frac{P}{2U_p} \tilde{h}_{tx} + \tilde{h}_{yz}$$

$$+ \frac{P \tilde{h}_{\zeta\zeta}}{U} - \frac{P \tilde{h}_{\zeta\zeta}}{U_p}.$$
for the gravitational backreaction. The relevant terms in the eikonal limit are described by the terms on the right-hand side, which all scale as $k$ according to the discussion under Eq. (15). Therefore similar to the Newtonian case, the relativistic antispring force effect scales as $k^2$ and gravitational backreaction scales as $k$. In the eikonal limit, we therefore find

$$\omega^2 \approx -\frac{U(0)P}{U_p(0)} k^2,$$  

(19)

which signals an instability if $P/\sigma > 0$. This result can be straightforwardly extended to cases for which the upper and lower spacetime have different cosmological constants.

**Gravastars.** A gravastar can be modeled by a spherical membrane separating an inner de Sitter spacetime and an outer Schwarzschild spacetime. If the inner region is an anti-de Sitter (AdS) spacetime, it is usually called an AdS Bubble [35, 36]. Defining $\rho$ as the “energy density” or cosmological constant in the inner space, $a$ as the radius of the membrane, $\sigma$ as the membrane surface energy density and $P$ as its pressure, the total mass $M$ of the spacetime is (following the notation in [37])

$$M = M_\rho + M_\sigma \sqrt{1 - \frac{2M_\rho}{a} + \frac{M_\sigma^2}{2a}},$$  

(20)

where $M_\rho = 4\pi a^2 \rho$ is the thin-shell mass and $M_\sigma = 4\pi a^3 / 3$ is the volume energy within the shell. The pressure within the shell is related to these masses through

$$P = \frac{1}{8\pi a} \left[ -\frac{1 - 4M_\rho/a}{\sqrt{1 - 2M_\rho/a}} + \frac{1 - M/\rho}{\sqrt{1 - 2M/\rho}} \right]$$  

and

$$P = \frac{1}{8\pi a} \left[ \frac{3M_\rho/a}{\sqrt{1 - 2M_\rho/a}} - \frac{1 - M_\rho/a}{\sqrt{1 - 2M_\rho/a}} + \frac{1 - M/\rho}{\sqrt{1 - 2M/\rho}} \right].$$  

(21)

To ensure meaningful values for $P$, we require that $M/\rho \leq 1/2$ and $M_\rho/\rho \leq 1/2$. If the gravastar satisfies the weak energy condition, the surface density $\sigma$ and $M_\sigma$ are both positive. We notice that $M \geq M_\rho > 0$ according to Eq. (20) (note $(1 - x)/\sqrt{1 - 2x}$ is a monotonically increasing for $0 \leq x \leq 1/2$). From the second line of Eq. (21), it is straightforward to see that the pressure is always positive. Intuitively it can be viewed as a consequence of the outer spacetime squeezing the inner spacetime, as the outer spacetime has larger effective pressure than the inner spacetime (also with the self-gravitation of the membrane). Although the analysis in the previous section was with topology $\mathbb{R}^2$ while the membrane of gravastars has topology $S^2$, this distinction is irrelevant as we consider local perturbations in the eikonal limit. This simple observation, together with the analysis of the warping mode instabilities, immediately suggests that *gravastars satisfying the weak energy condition are unstable*. The instability timescale is determined by Eq. (19) and depends on the prescription for $P$ and $\sigma$.

In the more general setting, as we consider both de Sitter and AdS interiors and surface density with arbitrary sign, the warping instability applies part of the parameter space of gravastars and AdS Bubbles, as shown in Fig. 2.

The modal instability of gravastars was initially studied in [37], which explicitly computed the quasinormal mode frequency for $\ell = 2$ axial and polar perturbations. However, the analysis in [37] treats the membrane as the provider of the matching condition between the inner and outer spacetime, in the same spirit as Eq. (4), but did not incorporate the membrane oscillations into the coupled mode equations. An explicit discussion of the gravastar mode analysis is included in the Supplementary material. It is indeed the membrane modes that destabilize the whole system in the eikonal limit.

**Thin-shell wormholes.** There are other horizonless compact objects generally considered in the literature as BH mimickers, or as candidates sourcing gravitational wave echoes. For example, thin-shell wormholes are commonly studied objects with compactness arbitrarily close to a BH. Consider two Schwarzschild solutions of the same mass $M$ attached at radius $r_0$ [33, 38], the corresponding thin-shell pressure and density at the wormhole throat are

$$P = \frac{1}{4\pi r_0} \frac{1 - M/r_0}{\sqrt{1 - 2M/r_0}}, \quad \sigma = \frac{1}{2\pi r_0} \frac{1}{\sqrt{1 - 2M/r_0}}$$  

(22)

so that the pressure is positive and the density is negative, which implies that static thin-shell wormholes are free from the warping instability. On the other hand, the empty shell models (which have positive surface energy density) as considered in [39, 40] naturally require positive in-shell pressure to support against gravity, and are all unstable to warping perturbations in the eikonal limit [41].

**Discussion.** We have discovered a generic instability for thin-layer structures in General Relativity, if $P/\sigma > 0$, with applications highlighted in exotic compact objects with membranes. One may imagine various ways to “cure” these systems so that they are free from warping instabilities. One possible way is to add additional rigidity against warping for the
membrane, e.g., a new term in the action with

$$S = \alpha \int d^3\xi \ K^{ij} K_{ij}$$  \hspace{1cm} (23)$$

where $\xi$ is the parameterization for the “world tube” of the membrane, $K^{ij}$ is the extrinsic curvature and $\alpha$ is a positive constant characterizing the rigidity. A possible caveat is that such an additional term in the action may lead to higher-order derivative terms in the equation of motion, which may raise concerns regarding well-posedness of the problem. Moreover, adding dissipation to the system does not cure the instability. This is because the anti-spring force causes run-away behavior of the displacement instead of oscillations. If the displacement were to saturate at some value, the dissipation becomes zero as there is zero velocity, but the anti-spring force continues to drive the displacement to larger values, i.e., there is no saturation point. On the other hand, if we replace the membrane with a shell of matter of thickness $d$, this can remove the instability. The thickness essentially adds a spatial frequency cutoff $k \sim 1/d$ in the above analysis. The caveat is that $d$ has to be sufficiently large so that the anti-spring in Eq. (10) becomes sub-dominant. Note that in this case the description for the dynamic behavior of matter with anisotropic stress is highly nontrivial and currently unknown.

A membrane with density and pressure having the same sign generically prefers configurations with higher surface curvature as they are associated with a lower energy state, if gravitational backreaction is neglected. For example, a membrane with an ellipsoidal shape has lower potential energy than that with a spherical shape. Mathematically the potential energy is $\propto Q_z^2 / (2 \lambda)$ where $Q_z$ is the mass quadrupole moment and $\lambda$ the tidal Love number. Negative potential energy means that $\lambda$ is negative. Even with gravitational backreaction included, if it is weaker than the anti-spring force such that the potential energy is still negative, the Love number $\lambda$ will also be negative [42]. Therefore the warping instability is connected to the negativity of tidal Love numbers, which applies to generic deformations with any $\ell \geq 2$ [43]. In the eikonal limit, the tidal Love number $\lambda_\ell$ has to be negative if $P / \sqrt{\sigma} > 0$.

Acknowledgements—We thank Luis Lehner, Eric Poisson and Yanbei Chen for helpful discussions. HY and ZP are supported by the Natural Sciences and Engineering Research Council of Canada and in part by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Colleges and Universities.

FIG. 3. This figure illustrates the physical meaning of $F_{\text{disk-in}}$ (top) and $F_{\text{disk-out}}$ (bottom).

SELF-GRAVITATING MEMBRANE IN THE NEWTONIAN SETTING

The treatment of membrane motion in the Newtonian limit offers an intuitive example for various driving terms in the membrane’s equation of motion, especially the dominant disk force and the gravitational back-reaction. By integrating the fluid variables from 0 to the $z_0$ side of the membrane, it is straightforward to obtain the left hand side of Eq. (3) in the main text. Here, we will present a derivation for the right hand side of Eq. (3), followed by explanations for the physical meaning of the different parts.

The anisotropic stress of the membrane can be written as

$$t_{ij} = -P(\delta_{ij} - n_i n_j) \delta[z - \xi(t)]$$  \hspace{1cm} (24)$$

where $n$ is the normal vector of the membrane. After performing the average from the bottom to the top side of the membrane, the $\delta(z - \xi)$ function in $\Delta t$ is removed and $\nabla \cdot \Delta t$ becomes $\nabla\| \cdot \langle \Delta t \rangle^\|$, where $\nabla\| = \nabla - n(\n \cdot \nabla)$ is the gradient within the surface of the membrane. With nonzero $\xi$, the normal vector becomes

$$n \rightarrow \hat{e}_z + \Delta n = \hat{e}_z - \partial_z \xi \hat{e}_x - \partial_z \xi \hat{e}_y$$  \hspace{1cm} (25)$$

so that

$$\nabla\| = \partial_z \hat{e}_z + \partial_z \hat{e}_z + (\partial_z \xi \hat{e}_x + \partial_z \xi \hat{e}_y) \hat{e}_c + (\partial_z \xi \hat{e}_x + \partial_z \xi \hat{e}_y) \partial_z .$$  \hspace{1cm} (26)$$

The term $(\partial_z \xi \partial_z + \partial_z \xi) \hat{e}_z \cdot t$ in $\nabla\| \cdot \Delta t$ is $O(\xi^3)$ and therefore does not contribute at linear order. The $(\partial_x \xi \hat{e}_x + \partial_y \xi \hat{e}_y) \partial_z \cdot \Delta t$ term also does not contribute because after performing the average across the vertical direction of the membrane $\langle \Delta t \rangle^\|$ has no explicit dependence on $z$, i.e., $\nabla\| \cdot \Delta t = (\partial_x \hat{e}_x + \partial_y \hat{e}_y) \cdot \Delta t$. Therefore, we find that
where $\nabla_i$ denotes the gradient along the membrane. This is equal to

$$\nabla_i \cdot (\kappa \Delta t^\gamma) \rightarrow -\nabla_i \left\{ (P_0 + \Delta P) \left[ \delta_{ij} - (\tilde{e}_{zj} + \Delta n_j) \tilde{e}_{zi} \right] \right\} - P_0 \left[ \delta_{ij} - \tilde{e}_{zi} \tilde{e}_{zi} \right], \quad (27)$$

Let us now proceed to discuss the physical meaning of different terms. For simplicity, we can neglect the $y$-dependence and visualize the two torus in Fig. 3. $F_{\text{disk-in}}$ is shown in the top panel, with its first term corresponding to the additional force generated by the pressure perturbation and the second term is produced by the tilt of the orbital plane, so that the pressure force $\nabla P$ points to a slightly different direction. $F_{\text{disk-out}}$ is shown in the bottom panel and is generated by the warping of the membrane. For example, in this effective one-dimensional scenario for Fig. 3, $\partial^2 \xi = -\xi/R$, with $R$ being its local curvature radius. It is straightforward to see that the net force is along the $z$-direction, with force density $P/R$ as consistent with the expression for $F_{\text{disk-out}}$.

**SELF-GRAVITATING MEMBRANE IN THE RELATIVISTIC SETTING**

In this section, we will show how the equation of motion for $\xi$ in Eq. (18) in the main text derives from Eq. (16), which we repeat here for convenience:

$$(\sigma + P)(u^i u^i)_{,\sigma} = P(n^i n^i)_{,\sigma} . \quad (29)$$

Let us start by expanding the left hand side, but first note that in $(t, z', x, y)$ coordinates $g_{\mu\nu} = g_{\mu\nu}^{(0)} + \xi_{\mu\nu} + \xi_{\mu \gamma} + \xi_{\nu \gamma}$, where $g_{\mu\nu}^{(0)} = \text{diag}\{- U(z), 1, U_p(z), U_p(z)\}$. $\xi_{\mu\nu} = (0, \xi_{t, y}(t, y), 0)$ and $u^\mu = (u^t, u^x, u^y, u^x) = (u^t, 0, 0, 0)$, where $u^i$ is constrained by its normalization. In particular, $u^i$ accordingly to the linear order is

$$u^i = \frac{1}{\sqrt{U(z) - \tilde{h}_{ti}}} = \frac{1}{U^{1/2}} + \frac{1}{2} \tilde{h}_{ti} . \quad (30)$$

Substituting these expressions, we obtain

$$u^i_{,\sigma} u^i_{,\sigma} = (u^i_{,\sigma} + u^i \Gamma^i_{,\gamma} u^\gamma)_{,\sigma} = u^i_{,\sigma} \Gamma^i_{,\sigma} = \frac{1}{U} \left( \tilde{h}_{i,\sigma} - \frac{1}{2} \tilde{h}_{n,\sigma} \right) . \quad (31)$$

and $u^i u^i_{,\sigma} = 0$, where we have used the reflection condition $U_{+,} + U_{-,} = 0$. Therefore, the left hand side of Eq. (29), up to linear order in the perturbations, is given by

$$(\sigma + P)(u^i u^i)_{,\sigma} = \frac{\sigma + P}{U} \left( \tilde{h}_{i,\sigma} - \frac{1}{2} \tilde{h}_{n,\sigma} \right) . \quad (32)$$

For expanding the right hand side, we start with the normal vector $n$ that is perpendicular to the surface $z' = 0$ so that $n_{\mu} \propto (0, 1, 0, 0)$ and

$$n^\mu = (n^t, n^r, n^\rho, n^\nu) = \left( \tilde{h}_{\mu}, \frac{1}{U} - \frac{1}{2} \tilde{h}_{\nu}, -\tilde{h}_{n,\mu}, U_p - \frac{\tilde{h}_{n,\mu}}{U_p} \right) , \quad (33)$$

accurate to the linear order, where we have normalized the vector $g_{\mu\nu} n^\nu = 1$. Then it is straightforward to show that

$$n^t n^t_{,\sigma} = n^r n^r_{,\sigma} + n^\rho n^\rho_{,\sigma} + n^\nu n^\nu_{,\sigma} = \left( \tilde{h}_{n,\sigma}, \frac{1}{U} \tilde{h}_{n,\sigma} + \frac{1}{U} \tilde{h}_{n,\sigma} , \frac{1}{U} \tilde{h}_{n,\sigma} + \tilde{h}_{n,\gamma} \right) . \quad (34)$$

and $n^r n^r_{,\sigma} = 0$. Consequently, the right hand side of Eq. (29) is

$$P(n^r n^r)_{,\sigma} = -\frac{P}{2U} \tilde{h}_{n,\sigma} + \frac{P}{2U_p} \left( \tilde{h}_{n,\sigma} + \tilde{h}_{n,\sigma} \right) + \frac{P}{U} \tilde{h}_{n,\sigma} - \frac{P}{U_p} \left( \tilde{h}_{n,\sigma} + \tilde{h}_{n,\sigma} \right) . \quad (35)$$

Equating both sides, we obtain

$$\sigma \frac{U}{U} (h_{n,\sigma} + \xi_{\gamma}) + \sigma \frac{U}{U} h_{n,\sigma} = \frac{P}{2U} \left( h_{n,\sigma} + \tilde{h}_{n,\sigma} \right) + \frac{P}{U} \tilde{h}_{n,\sigma} - \frac{P}{U_p} \left( \tilde{h}_{n,\sigma} + \tilde{h}_{n,\sigma} \right) , \quad (36)$$

where we have used $\tilde{h}_{n,\sigma} = h_{n,\sigma} + \xi_{\mu\nu} + \xi_{\gamma\delta}$. Finally, using the reflection condition $h_{n,\sigma}|_+ + h_{n,\sigma}|_- = 0$, the above equation reduces to Eq. (18) in the main text:

$$\sigma \frac{U}{U} (\xi_{\gamma}, + \xi_{x, x} + \xi_{y, y}) = \frac{\sigma}{U} h_{n,\sigma} - \frac{P}{U} (h_{n,\sigma} + \tilde{h}_{n,\sigma}) . \quad (37)$$

**EXPLICIT APPLICATION TO GRAVASTAR PERTURBATIONS**

The derivation for the warping instability of thin-layers applies for general settings. To emphasize and illustrate its application for black hole mimickers, we discuss non-radial oscillations of gravastars, as initially presented in Ref. [37] of the main text, but now with the membrane motion properly taken into account. This example should provide intuition for the warping instability.

Let us consider a gravastar with the thin-shell located at $r = a$ with surface density $\sigma$ and surface pressure $P$. After performing the spherical decomposition, we focus on even
parity/polar metric perturbations in the interior and exterior region of the gravastar for which the non-zero metric perturbations are

\[ h_{tt} = f(r)H(t,r)Y_{t\ell m}, h_{rr} = H_1(t,r)Y_{t\ell m}, h_{\ell r} = H_2(t,r)/h(r)Y_{t\ell m}, \]

\[ h_{\ell \ell} = r^2K(t,r)Y_{t\ell m}, h_{\phi \phi} = r^2\sin^2\theta K(t,r)Y_{t\ell m}. \] (38)

The functions \( f(r), h(r) \) are \( f(r) = ah(r) = \alpha(1 - 2M/r^2/a^2) \) in the interior and \( f(r) = h(r) = 1 - 2M/r \) in the exterior, with \( \alpha = (1 - 2M/a^2)/(1 - 2M/r). \) At any time slice, the even-parity displacement of the membrane for a given spherical harmonic index \((\ell, m)\) can be written as

\[ \xi_{\ell m}^\alpha = [0, h(r)\zeta(r)Y_{t\ell m}, 0, 0], \]

\[ \xi_{\ell m}^{(2)} = [0, 0, \frac{w(t)}{r^2}Y_{t\ell m, \theta}, \frac{w(t)}{r^2\sin^2\theta}Y_{t\ell m, \phi}], \] (39)

for two independent vector spherical harmonic basis. To derive the matching conditions and the equations of motion of the membrane, it is convenient to perform a coordinate transformation to map each mass element to its original coordinate value, as implemented in Ref. [37]. So in the new coordinate we have \( \tilde{h}_{\mu \nu} = h_{\mu \nu} + h^{\ell}_{\mu \nu}, \tilde{g}_{\mu \nu} = g_{\mu \nu} + \tilde{h}_{\mu \nu}, \) where \( h^{\ell}_{\mu \nu} \) is

\[ h^{\ell}_{\mu \nu} = \begin{bmatrix} -f'h^z & 0 & 0 & 0 \\ h'z/h & z\partial_\theta & z\partial_\phi & 0 \\ 0 & 2rhz & 0 & 0 \\ 0 & 0 & 2rh^z\sin^2\theta & 0 \end{bmatrix} Y_{t\ell m} \]

\[ + \begin{bmatrix} 0 & 0 & \psi Y_{t\ell m, \theta} & \psi Y_{t\ell m, \phi} \\ 0 & \psi Y_{t\ell m, \theta} & -2\psi Y_{t\ell m, \phi} & 0 \\ 0 & 0 & 2\psi Y_{t\ell m, \phi} & 2\psi Y_{t\ell m, \phi} \\ 0 & 0 & 2\psi Y_{t\ell m, \phi} & 2\psi Y_{t\ell m, \phi} \end{bmatrix}, \] (40)

with * being the nonzero symmetry elements and | the covariant derivative on the 2-sphere.

In the new coordinates, as discussed in Ref. [37], the Israel junction conditions imply that

\[ \left[ \sqrt{h}\zeta \right]^+ = 0, \quad K^+ = 8\pi\sigma_0 \sqrt{h}\zeta, \quad H^+ = 8\pi(\sigma_0 + 2P_0) \sqrt{h}\zeta, \]

\[ \left[ \sqrt{h} \left( \frac{H}{a} - K \right) \right]^+ = \frac{2h'}{a^2} + \frac{h'}{a}, \sqrt{h}\zeta = 8\pi\sigma_0, \]

\[ \left[ \sqrt{h} \left( K' - H' + \frac{2H}{f} \right) \right]^+ = \left[ \sqrt{h} \left( 1 + \frac{af''}{f} \right) \frac{H}{a} \right]^+ + \frac{2h'}{a^2} + \frac{f'h'}{f} - \frac{f''h'}{f^2}, \sqrt{h}\zeta = 16\pi\delta P. \] (41)

At this point, it was assumed in Ref. [37] that both the interior and exterior metric perturbation satisfy the Regge-Wheeler gauge condition, so that the metric quantities are all functions of the master variable \( \Psi, \Psi'. \) As a result, one only needs two independent matching conditions, which can be obtained by solving the wave equation with the Israel matching conditions. The resulting metric can then be plugged back into Eq. (41) to obtain the values of \( z, \delta \sigma, \delta P, \) which are consequently no longer independent quantities.

This observation already hints at an inconsistency in the above analysis, as the membrane can also be assumed to have certain prescribed motion (e.g., with prescribed amplitude and frequency), and the metric perturbations in the spacetime should respond to the prescribed membrane motion by solving the wave equation with the Israel matching conditions. In other words, there are infinite ways to construct \( z, \delta P, \delta \sigma \) and find the corresponding metric quantities, as we have not specified the membrane equation of motion in the above analysis. Indeed this inconsistency comes from the Regge-Wheeler gauge condition. It is allowed to impose this gauge condition in both the interior and exterior space, but then the metric perturbations will not necessarily satisfy the continuity conditions as required in Eq. (A14)-(A20) in Ref. [37]. Put differently, if we were to impose the Regge-Wheeler gauge in the interior and use the matching conditions to compute the value and the derivatives of the metric quantities in the exterior side of the membrane, the resulting exterior perturbations generally do not satisfy the Regge-Wheeler gauge condition.

An appropriate procedure to compute the quasinormal mode can be that imposing the Regge-Wheeler gauge in the interior space, writing down the matching conditions and the membrane equations of motions (which give the prescription for \( z, \delta \sigma, \delta P, \delta \sigma = 0 \)) and after that the value and the derivatives of the master variable \( \Psi, \) which is gauge invariant. Finally by solving the wave equation of \( \Psi \) and imposing the outgoing condition at infinity, the quasinormal mode frequency can be determined.

For our discussion here, as we want to reveal the existence of the warp instability in the eikonal limit \( \ell \gg 1, \) a few observations can be made. First of all, as we focus on the radial oscillation mode, for reasons that will become clear later, the amplitude of \( w \) is \( 1/\ell \) times smaller than the amplitude of \( z, \) i.e., \( \mid w \mid \sim |z|/\ell. \) As the perturbation of the surface density is expected to be \( \delta \sigma \propto \|\Phi\|, \xi \propto \ell w, \) and \( \delta P \propto \ell^2 \delta \sigma, \) we find the following scaling \( \delta \sigma, \delta P, \sim O(\ell)^2 z. \) The matching condition in Eq. (41) implies that \( H, K, H_1 \sim O(\ell)^2 z \) as well. Second, we
can write down the membrane equation of motion
\[
\left[(\sigma_0 + \delta \sigma + P_0 + \delta P) u' w' \right]_\nu = \left[(P_0 + \delta P)(n^\mu n^\nu - g^\mu\nu)\right]_\nu
\]
(42)
where \(n_\nu \propto (0,1,0,0)\) and \(u' = (1/\sqrt{-g_\mu\nu}0,0,0,0)\). The equation of motion for \(z := \sqrt{r}z\) (which is continuous across the membrane) is obtained from the \(\mu = r\) component:
\[
\begin{align*}
\frac{4h}{r} h_{rr} + \frac{2c_0}{r^2} h_{r\theta} + \frac{2}{r^2} \left( h_{r\phi} + \frac{h_{\phi\phi}}{\sin^2 \theta} \right) - \frac{\sigma_0 h f'}{P_0 f} h_{rr} + \frac{\sigma_0 h f'}{P_0 f} \tilde{h}_{rr} + \frac{\sigma_0 h f'}{P_0 f} h_{\theta \theta} + \frac{\sigma_0 h f'}{P_0 f} \tilde{h}_{\theta \theta} \\
+ \frac{2}{r^2 \sin^2 \theta} \tilde{h}_{r \phi, \phi} - \frac{1}{r^2} \left( \tilde{h}_{\phi \phi, r} + \frac{1}{\sin^2 \theta} \tilde{h}_{\phi \phi, \theta} \right) + \frac{\sigma_0}{P_0 f} h_{r r, \theta} - \frac{4}{r P_0 f} \delta P + \frac{f'}{P_0 f} \delta \sigma = 0
\end{align*}
\]
(43)
where an average over plus and minus side of the membrane is implicitly performed. By assigning the derivatives an order \(O(\ell)\), we can read off the \(O(\ell^2)\) terms in the above equation
\[
-\frac{\sigma_0}{f} \frac{\ell^2}{a^2} P_0 z = O(\ell),
\]
so that \(\omega^2 \sim -\ell^2/P_0/\sigma_0\) which is consistent with the scaling we obtained in the main text. The equation of motion for \(w\) can be obtained by setting \(\mu = \theta\) in Eq. (42), where the relevant terms have order \(\omega^2 \tilde{w}, \ell^2 \tilde{w}\) and \(\partial h_{\mu \nu} \sim O(\ell)\). As a result, we have \(w \sim zO(\ell^{-1})\).