# Interactions resolve state-dependence in a toy-model of AdS black holes

**Adam Bzowski,**[a,b] **Alessandra Gnecchi**[c] **and Thomas Hertog**[a]

[a] *Institute for Theoretical Physics, KU Leuven,*
 *Celestijnenlaan 200D, Leuven, Belgium*

[b] *Institut de Physique Théorique, CEA Saclay,*
 *Orme des Merisiers bâtiment 774, Gif-sur-Yvette, France*

[c] *Theoretical Physics Department, CERN,*
 *Geneva, Switzerland*

 *E-mail:* adam.bzowski@ipht.fr, alessandra.gnecchi@cern.ch,
 thomas.hertog@kuleuven.be

ABSTRACT: We show that the holographic description of a class of AdS black holes with scalar hair involves dual field theories with a double well effective potential. Black hole microstates have significant support around both vacua in the dual, which correspond to perturbative degrees of freedom on opposite sides of the horizon. A solvable toy-model version of this dual is given by a quantum mechanical particle in a double well potential. In this we show explicitly that the interactions replace the state-dependence that is needed to describe black hole microstates in a low energy effective model involving the tensor product of two decoupled harmonic oscillators. A naive number operator signals the presence of a firewall but a careful construction of perturbative states and operators extinguishes this.

# Contents

## 1 Introduction

Holography has provided a fruitful new perspective on the black hole information paradox first formulated by Hawking [1]. Specifically it has enabled an expression of the essence of the paradox in dual quantum field theoretic terms. This has further sharpened some of the underlying assumptions, and it has led to novel suggestions for its resolution. Key elements that have emerged in this recent discussion include the following;

- *Firewalls:* according to [2–5] a generic black hole microstate exhibits a firewall, i.e. that its horizon is not a smooth surface. In terms of field theory data the paradox is stated as an incompatibility of the following four conditions: $i$) unitarity, $ii$) validity

of the semiclassical approximation for an asymptotic observer, $iii$) existence of black hole microstates visible to an asymptotic observer as states with exponentially small energy differences, and $iv$) existence of a firewall-free number operator for an infalling observer.

- *State-dependence:* in contrast with this [6–9] proposed a dual description of the black hole interior by introducing state-dependent operators. The black hole horizon remains smooth, but such operators depend on the specific black hole microstate and hence go beyond the standard paradigm of quantum field theory. In this way one avoids many pitfalls presented by the firewall arguments such as existence of creation-annihilation operators satisfying the 'wrong sign' commutation relations.

- *Vacuum structure:* Hawking's original argument has been recast as a no-go theorem [10, 11] that states that quantum gravity effects cannot prevent information loss if they are confined to within a given scale and if the vacuum of the theory is assumed to be unique. This is a particularly sharp puzzle in the context of holography given the apparent uniqueness of the vacuum of dual CFTs. It can be seen as one motivation for recent investigations into a possible non-trivial vacuum structure as in [12].

The reformulation of the information paradox in terms of purely field theoretic data elucidates the nature of the underlying issues. Devoid of a geometrical interpretation one can ask whether there exists any tractable well-defined quantum theories modeling black holes and satisfying basic properties required by the aforementioned papers.

In this paper we put forward a new holographic toy-model for AdS black holes that incorporates in a toy-model fashion a specific proposal for the nature of non-perturbative quantum gravity corrections to black hole physics. The model consists of a quantum mechanical particle in a double well potential.

Our motivation to advance this as a toy-model for black holes in AdS is twofold. First, there is a class of single-sided black hole solutions in global AdS with scalar hair outside the horizon whose dual description involves a field theory with a double well effective potential [13–16]. We review these black holes, which are solutions in truncations of AdS supergravity with so-called designer gravity boundary conditions, in section 2. The potential barrier in their dual description separates the perturbative degrees of freedom on both sides of the horizon, but they are coupled through multi-trace interactions. We argue that black hole microstates are states with significant support around both perturbative vacua. The quantum mechanical model we put forward can be viewed as a toy-model for systems of this kind since it amounts to two harmonic oscillators coupled through a 'non-perturbative' interaction modeled as a double well potential. Outside the context of holography Giddings [17, 18] has studied how novel, non-local interactions (in the bulk) can resolve the information paradox.

Secondly, recent work in the context of two-sided black holes in AdS has advocated that not only entanglement but also interactions between the two boundary CFTs are needed to describe the bulk [19–22]. Some implications of adding a specific example of such interactions were explored in nearly $AdS_2$ in [22–27]. These studies yield a different

motivation for our toy-model in which the two perturbative vacua are thought of as being dual to the two asymptotic regions on both sides of the horizon. The potential barrier in the dual toy-model amounts to a proposal for a specific interaction between a single pair of left and right modes in the bulk, $|n_k\rangle_L$ and $|n_k\rangle_R$, with fixed frequency $\omega$ and fixed wave vector $k$ and related to boundary states by some form of the HKLL construction [28, 29]. For small values of $n_k$ the left and right modes are essentially separated by a potential barrier. By contrast, for sufficiently large occupation numbers left and right modes interact strongly and the semiclassical approximation breaks down. In the context of holography [30] argued that such non-perturbative effects can be sufficient to resolve the information paradox.

Motivated by these developments we consider a quantum mechanical particle with the following potential

$$V(x) = \frac{1}{32\lambda} - \frac{1}{4}\omega^2 x^2 + \frac{\lambda}{2}\omega^4 x^4 \tag{1.1}$$

governed by the standard Hamiltonian

$$H = \frac{1}{2}p^2 + V(x) \tag{1.2}$$

This system is characterized by a dimensionless coupling constant $\lambda$ and a frequency $\omega$ of the approximate harmonic oscillators around the semiclassical vacuum states $\varphi_0^L$ and $\varphi_0^R$ at $x_{R,L} = \pm 1/(2\omega\sqrt{\lambda})$. The vacua in our model correspond to the left and right, or interior and exterior, semiclassical vacua in the bulk. Excited states from the standpoint of observers in one of these vacua then naturally correspond to perturbative states $\varphi_n^L$ and $\varphi_n^R$. Black hole microstates finally are linear combinations of perturbative states in both vacua. We will argue that typical microstates correspond to states with significant support around both perturbative vacua. By contrast, bulk spacetimes without a black hole correspond to states with support around one of the vacua only.

To make contact with the usual perturbative expansion in semiclassical gravity we introduce a dimensionless parameter $N$ as

$$\lambda = \frac{1}{N^2}. \tag{1.3}$$

Hence perturbation theory in $\lambda$ models the usual large $N$ expansion. The height of the potential barrier equals $V_* = V(0) = 1/(32\lambda)$. In the limit $\lambda \to 0$ the barrier grows, and the two minima move apart. In the exact $\lambda = 0$ limit the excitations around both perturbative vacua decouple completely and the system reduces to two decoupled harmonic oscillators with frequency $\omega$. In the bulk, with designer gravity boundary conditions, this decoupling limit corresponds precisely to a limit in which the horizon of the hairy black holes becomes singular.

In this paper we carefully study how states and operators in the full interacting toy-model relate to quantities in a low energy effective theory involving the tensor product of two decoupled harmonic oscillators. At the effective theory level our model captures many of the usual paradoxes associated with the semiclassical approximation of black hole physics in a remarkably precise manner. A major advantage of our model is that it is solvable. This enables us to sharpen the limitations of the perturbative description of black

holes, to explore dynamical processes, and to understand in this concrete toy-model setup how non-perturbative interactions resolve the paradoxes. In particular we show explicitly that the interactions eliminate the state-dependence that is needed to describe black hole microstates in the effective low energy dual. We also find that a naive number operator signals the presence of a firewall, but that a careful construction of perturbative states and operators in the full model extinguishes this. Finally, when it comes to dynamical processes, we point out that tunneling near the potential maximum corresponds to Hawking radiation in the bulk, and that the scattering of classical waves nicely captures the behavior of shock waves in the bulk.

We conclude this introduction with an important caveat. Evidently our model is not suitable for the analysis of properties of black holes that depend on a collection of modes. This in particular encompasses all thermodynamical properties that rely on the existence of an ensemble of modes. In this context it would be necessary to consider more complicated models such as matrix or tensor models.

## 2 Dual description of hairy AdS black holes

In this section we review the dual description in terms of a field theory with a double well effective potential of a class of single-sided asymptotically $AdS_4$ static black hole solutions with scalar hair. The perturbative degrees of freedom on both sides of the horizon correspond in the dual description to excitations around two distinct perturbative vacua. However, multi-trace interactions in the dual imply a non-perturbative coupling between both sides. As such this setup motivates the quantum mechanical particle in a double well potential as a toy-model for black holes in AdS.

The black hole solutions we construct are variations of the solutions found in [13–16, 31–33] and recently in [34, 35]. Consider the low energy limit of M theory with $AdS_4 \times S^7$ boundary conditions. The massless sector of the compactification of $D = 11$ supergravity on $S^7$ is $\mathcal{N} = 8$ gauged supergravity in four dimensions. It is possible to consistently truncate this theory to include only gravity and a single scalar with action

$$S = \int d^4x \sqrt{-g} \left[ \frac{1}{2}R - \frac{1}{2}(\nabla\phi)^2 + 2 + \cosh(\sqrt{2}\phi) \right] \qquad (2.1)$$

where we have set $8\pi G = 1$ and chosen the gauge coupling so that the AdS radius is one. The potential has a maximum at $\phi = 0$ corresponding to an $AdS_4$ solution with unit radius. It is unbounded from below, but small fluctuations have $m^2 = -2$, which is above the Breitenlohner-Freedman bound $m_{BF}^2 = -9/4$ so with the usual boundary conditions $AdS_4$ is stable. Consider global coordinates in which the $AdS_4$ metric takes the form

$$ds_0^2 = \bar{g}_{\mu\nu}dx^\mu dx^\nu = -(1 + r^2)dt^2 + \frac{dr^2}{1 + r^2} + r^2 d\Omega_2 \qquad (2.2)$$

In all asymptotically AdS solutions, the scalar $\phi$ decays at large radius as

$$\phi(r) \sim \frac{\alpha}{r} + \frac{\beta}{r^2}, \qquad r \to \infty \qquad (2.3)$$

where $\alpha$ and $\beta$ are functions of $t$ and the angles. To have a well-defined theory one must specify boundary conditions at spacelike infinity. The standard choice of boundary condition corresponds to taking $\alpha = 0$. However one can consider more general 'designer gravity' boundary conditions [14] with $\alpha \neq 0$ that are specified by a functional relation $\beta(\alpha)$ in (2.3). The backreaction of the $\alpha$-branch of the scalar field and its self-interaction modify the asymptotic behavior of the gravitational fields. Writing the metric as $g_{\mu\nu} = \bar{g}_{\mu\nu} + h_{\mu\nu}$ the corresponding asymptotic behavior of the metric components is given by

$$h_{rr} = -\frac{(1 + \alpha^2/2)}{r^4} + \mathcal{O}(1/r^5), \quad h_{rm} = \mathcal{O}(1/r^2), \quad h_{mn} = \mathcal{O}(1/r) \tag{2.4}$$

Nevertheless, the Hamiltonian generators of the asymptotic symmetries remain well-defined and finite when $\alpha \neq 0$ [13, 36, 37]. They acquire however an explicit contribution from the scalar field. For instance, the conserved mass of spherical solutions is given by

$$M = \mathrm{Vol}(S^2)\left[M_0 + \alpha\beta + W\right] \tag{2.5}$$

where $M_0$ is the coefficient of the $1/r^5$ term in the asymptotic expansion of $g_{rr}$, and where we have defined the function

$$W(\alpha) = \int_0^\alpha \beta(\tilde{\alpha})d\tilde{\alpha}, \tag{2.6}$$

which defines the choice of boundary conditions.

Consider now a specific class of boundary conditions defined by the following relation

$$\beta_{bc}(\alpha) = -c_1\alpha^2 + c_2\alpha^3 \tag{2.7}$$

where $c_1$ and $c_2$ are constants. With these, the conserved mass (2.5) is given by

$$M = 4\pi \left(M_0 - \frac{4}{3}c_1\alpha^3 + \frac{1}{6}c_2\alpha^4\right) \tag{2.8}$$

Both the vacuum and the dynamical properties of the theory — as well as the possible black hole endstates of gravitational collapse — depend significantly on $W$ [14, 38]. In the context of the AdS/CFT correspondence, adopting designer gravity boundary conditions defined by a function $W \neq 0$ corresponds to adding a potential term $\int W(\mathcal{O})$ to the dual CFT action, where $\mathcal{O}$ is the field theory operator that is dual to the bulk scalar [39, 40]. This is generally a complicated multi-trace interaction. Certain deformations $W$, including those corresponding to boundary conditions of the form (2.7), give rise to field theories with additional, possibly metastable vacua.

The AdS/CFT correspondence relates the expectation values $\langle\mathcal{O}\rangle$ in different field theory vacua to the asymptotic scalar profile of regular static solitons in the bulk. The precise correspondence between solitons and field theory vacua is given by the following function [37],

$$\mathcal{V}(\alpha) = -\int_0^\alpha \beta_s(\tilde{\alpha})d\tilde{\alpha} + W(\alpha) \tag{2.9}$$

where $\beta_s(\alpha)$ is obtained from the asymptotic scalar profiles of spherical soliton solutions with different values $\phi(0)$ at the origin $r = 0$. This curve was first obtained in [14] for the theory (2.1) and is plotted in figure 1(a).
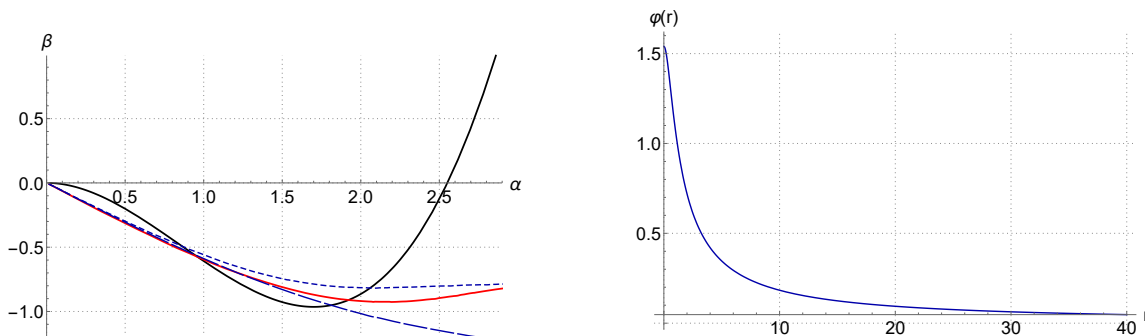
**Figure 1**. Left: the functions $\beta(\alpha)$ that characterize the set of solitons (red) in the theory, and hairy black holes with horizon size $R_e = .2$ (blue dotted) and $R_e = 1$ (blue dashed). The black curve shows the boundary condition function $\beta_{bc}(\alpha) = -\alpha^2 + 0.392\alpha^3$ which, together with the characteristics of the solitons, leads to the dual double well effective potential shown in figure 2(a). Right: the scalar radial profile of the soliton associated with the second intersection point of $\beta_{bc}(\alpha)$ with the soliton curve $\beta_s(\alpha)$.

Given a choice of boundary condition $\beta(\alpha)$, the allowed solitons are simply given by the points where the soliton curve intersects the boundary condition curve: $\beta_s(\alpha) = \beta(\alpha)$. For any $W$ the location of the extrema of $\mathcal{V}$ in (2.9) yield the vacuum expectation values $\langle \mathcal{O} \rangle = \alpha$, and the value of $\mathcal{V}$ at each extremum yields the energy of the corresponding soliton. Hence $\mathcal{V}(\alpha)$ can be interpreted as an effective potential for $\langle \mathcal{O} \rangle$. This led [14] to conjecture that (a) there is a lower bound on the gravitational energy in those designer gravity theories where $\mathcal{V}(\alpha)$ is bounded from below, and that (b) the solutions locally minimizing the energy are given by the spherically symmetric, static soliton configurations.

For the boundary conditions (2.7) the effective potential is generally of the form shown in figure 2, indicating the emergence of a second, metastable vacuum.[1] The AdS/CFT correspondence then suggests that the bulk theory (2.1) with such boundary conditions satisfies the Positive Mass Theorem, and that empty AdS remains the true ground state.[2] However the constants in (2.7) can be tuned so that the new vacuum has precisely the same energy as the AdS vacuum. For this choice of boundary conditions the effective potential in the dual takes the form of a double well potential with two vacua at equal energy, separated by a barrier.

One can also consider excitations around each of these perturbative vacua. A particular class of excitations corresponds to 'adding' a black hole at the centre of the soliton. When non-linear backreaction is included, these are spherical static black hole solutions with scalar hair. Black holes of this kind were found numerically in [13, 15] for boundary conditions similar to (2.7). Regularity of the event horizon $R_e$ implies the relation

$$\phi'(R_e) = \frac{R_e V_{,\phi_e}}{1 - R_e^2 V(\phi_e)} \tag{2.10}$$

---

[1]In the bulk this new vacuum corresponds to the second intersection point of $\beta_s(\alpha) = \beta(\alpha)$ in figure 1(a). The first intersection point corresponds to unstable solitons associated with the local maximum of $\mathcal{V}(\alpha)$.

[2]See [41] for a stability analysis of this theory (with more stringent conditions on $W$) using purely gravitational arguments.
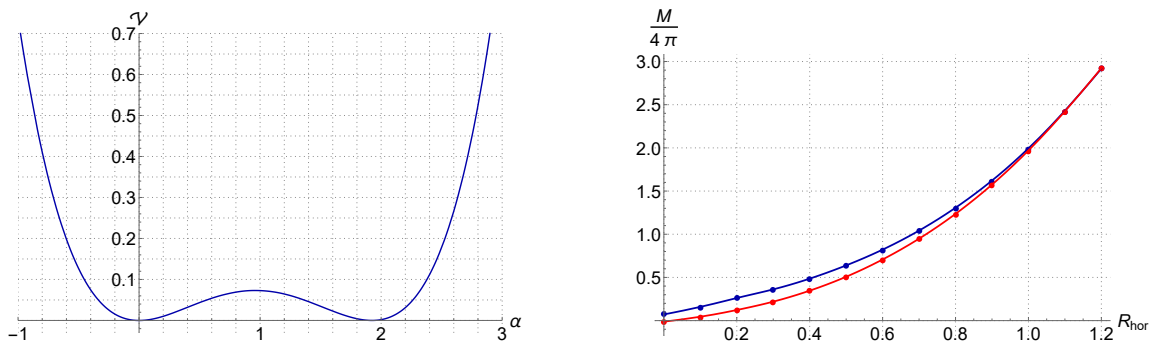
**Figure 2**. Left: the effective potential $\mathcal{V}(\alpha)$ for the vacuum expectation values $\langle \mathcal{O} \rangle$ in the dual field theory with deformation $W = -\frac{1}{3}\alpha^3 + 0.098\alpha^4$. The parameters are chosen in such a way that the potential exhibits two minima of equal depth. Right: the mass of the hairy black holes that obey these boundary conditions. The red line gives the masses of the second (perturbatively stable) branch of solutions, which are associated with the second intersection point of the curves $\beta_{R_e}(\alpha)$ with $\beta_{bc}(\alpha)$, and hence have more hair.

where $\phi_e \equiv \phi(R_e)$. The usual Schwarzschild-AdS black holes with $\phi = 0$ everywhere are still valid solutions of the theory with boundary conditions (2.7), since the curve $\beta(\alpha)$ intersects the origin. However in addition the theory admits black holes with scalar hair at and outside the horizon. The scalar asymptotically behaves again as (2.3), so we obtain a point in the $(\alpha, \beta)$ plane for each combination $(R_e, \phi_e)$. Repeating for all $\phi_e$ gives a curve $\beta_{R_e}(\alpha)$. In figure 1(a) we show this curve for hairy black holes of two different sizes. As one increases $R_e$, the curve decreases faster and reaches larger (negative) values of $\beta$. Given a choice of boundary conditions $\beta(\alpha)$, the allowed black hole solutions are given by the points where the black hole curves intersect the boundary condition curve: $\beta_{R_e}(\alpha) = \beta(\alpha)$. It follows immediately that for boundary conditions (2.7) there are two hairy black holes of a given horizon size provided $R_e$ is sufficiently small. Each branch of hairy black holes tends to one of the two spherical static solitons in the limit $R_e \to 0$. The mass (2.8) of both branches of black holes is shown in figure 2(b).

The hairy black holes reviewed here are solutions where a normal Schwarzschild-AdS black hole interior solution is smoothly glued at the horizon onto a scalar soliton solution outside, slightly modified by the non-linear backreaction of the black hole. Hence the usual AdS vacuum outside the black hole is essentially replaced by a solitonic vacuum.[3] In this way one separates to first approximation the excitations that make up the black hole interior from the degrees of freedom outside the horizon, but without introducing a second boundary. This separation is clearly manifest in the dual description of the black holes which involve a double well effective potential of the form shown in figure 2(a). In this description, the soliton corresponds to the ground state wave function around the new vacuum whereas the black hole degrees of freedom correspond to excitations around the original vacuum.

---

[3]The upper branch of more massive hairy black holes in figure 2(b) corresponds to black holes glued onto the unstable soliton associated with the maximum of $\mathcal{V}(\alpha)$. Those black holes are, like the soliton, unstable.

However, there is evidently also an important coupling between both vacua. On the field theory side the vacuum structure emerges from a complicated set of multi-trace interactions. In the bulk the coupling is at the semiclassical level encoded in the regularity condition at the horizon. Note furthermore that one can consider a one-parameter family of boundary conditions of the form (2.7) for which the second vacuum is always at zero energy, but is gradually taken further away. In the limit of large separation in which both vacua decouple, the hairy black holes become singular on the horizon.

These features of the dual description of hairy black holes form the basic motivation to put forward a quantum mechanical particle in a double well potential as an extremely simplified — but solvable — toy-model for (this class of) black holes in AdS. In the remainder of this paper we study this toy-model, and its connection to black hole physics.

## 3 Quantum mechanics in a double well and black hole microstates

### 3.1 Canonical quantization and Hilbert spaces

Consider the 1D quantum mechanical system of a particle in the double well potential[4] (1.1). One can carry out the procedure of canonical quantization either around $x_L$ or around $x_R$ by ignoring all interaction terms. This leads to two separate Fock spaces $\mathcal{F}_L$ and $\mathcal{F}_R$ which come equipped with two pairs of creation and annihilation operators $b_L, b_L^+$ and $b_R, b_R^+$. However, while formally independent, these two Fock spaces must be related since they arise from the same system.

We first discuss the perturbative structure around the minumum at $x_R$. When expanded around $x_R$, the Hamiltonian can be written as $H = H_R^{(0)} + H_R^{(1)}$, with $H_R^{(0)}$ the Hamiltonian of a harmonic oscillator and $H_R^{(1)}$ its (perturbative) correction,

$$H_R^{(0)} = \frac{1}{2}p^2 + \frac{1}{2}y_R^2, \tag{3.1}$$

$$H_R^{(1)} = \sqrt{\lambda}y_R^3 + \frac{\lambda}{2}y_R^4, \tag{3.2}$$

where $y_R = x - x_R$. Standard canonical quantization based on $H_R^{(0)}$ around $x_R$ yields a Fock space $\mathcal{F}_R$ with a set of basis states $|n\rangle_R$, together with a pair of creation-annihilation operators $b_R, b_R^+$ satisfying

$$b_R|0\rangle_R = 0, \qquad |n\rangle_R = \frac{1}{\sqrt{n!}}(b_R^+)^n|0\rangle_R . \tag{3.3}$$

A similar analysis is applicable to the left minimum at $x_L$[5] and gives another Fock space $\mathcal{F}_L$ spanned by the states $|n\rangle_L$ and with creation and annihilation operators $b_L, b_L^+$. One can

---

[4]For simplicity we set $\omega = 1$ from now on and restore $\omega$ only where it is illuminating.

[5]In particular, we can define analogous Hamiltonians $H_L^{(0)}$ and $H_L^{(1)}$ which differ from their $x_R$ counterparts but satisfy by definition $H_L^{(0)} + H_L^{(1)} = H_R^{(0)} + H_R^{(1)} = H$. It was argued in [42] that, when modelling an eternal black hole, $H_R$ itself should be regarded as the total Hamiltonian of the theory, instead of $H_R \pm H_L$. Our model realizes this intuition since there is only a single Hamiltonian $H$. This Hamiltonian can be split into its free and interacting part as in (3.1) and (3.2) to better describe an experience of the right and the left observer. But nevertheless $H_R = H_L = H$ is always the same operator.

associate to these Fock spaces two observers, left and right. The right observer perceives the state $|0\rangle_R$ as the natural semiclassical vacuum and the excited state $|n\rangle_R$ as an $n$-particle state. Similary, the left observer regards $|0\rangle_L$ as the semiclassical vacuum and $|n\rangle_L$ as an $n$-particle state.

The above canonical quantizations eliminate any relation between both Fock spaces. In particular expressions such as $[b_L, b_R]$ make no sense as the operators involved act on different Hilbert spaces. To relate $\mathcal{F}_L$ and $\mathcal{F}_R$ we have to embed these into the total Hilbert space $\mathcal{H} = L^2(\mathbb{R}; \mathbb{C})$ of complex-valued, square-integrable wave functions. Consider the set $\{\varphi_n\}_{n \in \mathbb{N}}$ of normalized eigenfunctions of the Hamiltonian of a harmonic oscillator,

$$\varphi_n(x) = \frac{1}{\pi^{1/4}\sqrt{2^n n!}} H_n(x) e^{-\frac{x^2}{2}}, \qquad x \in \mathbb{R}, \tag{3.4}$$

where $H_n$ denotes standard Hermite polynomials. We can define two morphisms $F_L$ and $F_R$ between the Fock spaces $\mathcal{F}_L$, $\mathcal{F}_R$ and $\mathcal{H}$ as

$$F_R: \qquad \mathcal{F}_R \ni |n_R\rangle \mapsto \varphi_n^R \in \mathcal{H}, \qquad \varphi_n^R(x) = \varphi_n(x - x_R), \tag{3.5}$$

$$F_L: \qquad \mathcal{F}_L \ni |n_L\rangle \mapsto \varphi_n^L \in \mathcal{H}, \qquad \varphi_n^L(x) = (\Theta \varphi_n^R)(x) = (-1)^n \varphi_n(x - x_L). \tag{3.6}$$

where the CPT operator $\Theta$ acts on elements $\psi \in \mathcal{H}$ as $(\Theta \psi)(x) = \psi^*(-x)$, and the asterisk denotes complex conjugation. We have introduced an additional factor $(-1)^n$ in the definition of $\varphi_n^L$ which allows us to relate left and right modes as CPT conjugates of each other. The maps $F_R$ and $F_L$ are obviously isomorphisms that can be thought of as two different choices of basis of $\mathcal{H}$ associated with harmonic oscillator eigenstates around either $x_L$ or $x_R$. Hence the total Hilbert space $\mathcal{H}$ is isomorphic to each Fock space $\mathcal{F}_L$ and $\mathcal{F}_R$ *separately*,

$$\mathcal{H} \cong \mathcal{F}_R \cong \mathcal{F}_L; \tag{3.7}$$

There is no tensor product. The interactions provide a non-trivial identification of the two Fock spaces within a single $\mathcal{H}$.

On the other hand, the Fock spaces $\mathcal{F}_R$ and $\mathcal{F}_L$ have distinct sets of creation and annihilation operators. This means that an expression such as $b_L|n\rangle_R$ makes *a priori* no sense, since $b_L$ acts on $\mathcal{F}_L$, whereas $|n\rangle_R$ is a state in $\mathcal{F}_R$. However the isomorphisms (3.5) and (3.6) can be used to define new annihilation operators $a_R, a_L$ constructed from $b_R, b_L$ that do have a well defined action in $\mathcal{H}$. In particular, defining the operators

$$a_R = F_R b_R F_R^{-1}, \qquad a_L = F_L b_L F_L^{-1}, \tag{3.8}$$

together with their conjugates $a_R^+$ and $a_L^+$, we get the following actions,

$$a_R \varphi_n^R = \sqrt{n} \varphi_{n-1}^R, \qquad\qquad a_R^+ \varphi_n^R = \sqrt{n+1} \varphi_{n+1}^R, \tag{3.9}$$

$$a_L \varphi_n^L = \sqrt{n} \varphi_{n-1}^L, \qquad\qquad a_L^+ \varphi_n^L = \sqrt{n+1} \varphi_{n+1}^L, \tag{3.10}$$

In particular, the action of $a_L$ is related to the action of $a_R$ by the parity operator,

$$a_L = \Theta a_R \Theta \tag{3.11}$$

which is the standard relation between left and right creation-annihilation operators featuring in black hole physics, e.g., [43]. Expressions such as $a_L \varphi_n^R$ are now meaningful because both pairs of operators $a_L, a_L^+$ and $a_R, a_R^+$ act on the same Hilbert space $\mathcal{H}$. Their relation to the fundamental field operator $x$ is simply that of a shifted harmonic oscillator,

$$y_R = x - x_R = \frac{1}{\sqrt{2}}(a_R + a_R^+), \tag{3.12}$$

$$y_L = \Theta y_R \Theta = x_L - x = \frac{1}{\sqrt{2}}(a_L + a_L^+). \tag{3.13}$$

Since the two oscillators are related by a displacement (up to a sign), it follows that the creation-annihilation operators are related as well,

$$a_L = -\frac{1}{\sqrt{2\lambda}}\mathbb{I} - a_R = -\frac{N}{\sqrt{2}}\mathbb{I} - a_R. \tag{3.14}$$

A striking feature of this expression is that it does not possess a finite decoupling limit $N \to \infty$ as an operator statement. Instead of approximating the free field creation-annihilation operators $b_L, b_L^+, b_R, b_R^+$, these operators diverge in the decoupling limit.[6]

As we can see, the $\lambda \to 0$ limit should be taken with care. The low occupancy modes $\varphi_n^R$ and $\varphi_m^L$, $n, m, \ll N$ nearly decouple for small $\lambda$ and in the limit $\lambda \to 0$, the two sets of modes decouple completely. We end up with two separate harmonic oscillators with the tensor product Hilbert space $\mathcal{H}_0 \cong \mathcal{F}_L \otimes \mathcal{F}_R$. On the other hand the interaction Hamiltonian $H_R^{(1)}$ vanishes for $\lambda = 0$. From the point of view of the vacuum at $x_R$, the second vacuum moves away and a single harmonic oscillator Hilbert space $\mathcal{F}_R$ remains. Hence from this viewpoint every state $\varphi_n^L$ disappears as $\lambda \to 0$. In the language of mathematical perturbation theory [44] this limit is singular.

This is also the case in the bulk for the black holes with scalar hair described in the Introduction and reviewed in section 2. The distance between the vacua in the dual is related to the boundary conditions in the bulk, which in turn determine the amount of scalar hair. Increasing the distance between both vacua, keeping the black hole mass constant, also increases the value of the scalar hair on the horizon. In the limit in which the vacua in the dual theory decouple, a curvature singularity at the horizon develops, effectively dividing the inside and outside regions in two separate spacetimes.

In black hole physics one usually assumes at the outset that the Hilbert space splits in a tensor product of Fock spaces associated with modes inside and outside the horizon. In particular this is a fundamental assumption behind much of the discussion of the information

---

[6]Another indication that the natural creation-annihilation operators $a_L, a_L^+, a_R, a_R^+$ are not to be identified with the perturbative creation-annihilation operators are their commutation relations. Using (3.14) we can find that

$$[a_L, a_R] = [a_L^+, a_R^+] = 0, \qquad [a_L, a_R^+] = [a_R, a_L^+] = -1 . \tag{3.15}$$

Hence $a_L$ and $a_R^+$ do not commute, in sharp contrast with the usual situation in black hole physics, where one expects the left and right creation-annihilation operators to commute as a consequence of the locality of semiclassical physics near the horizon.
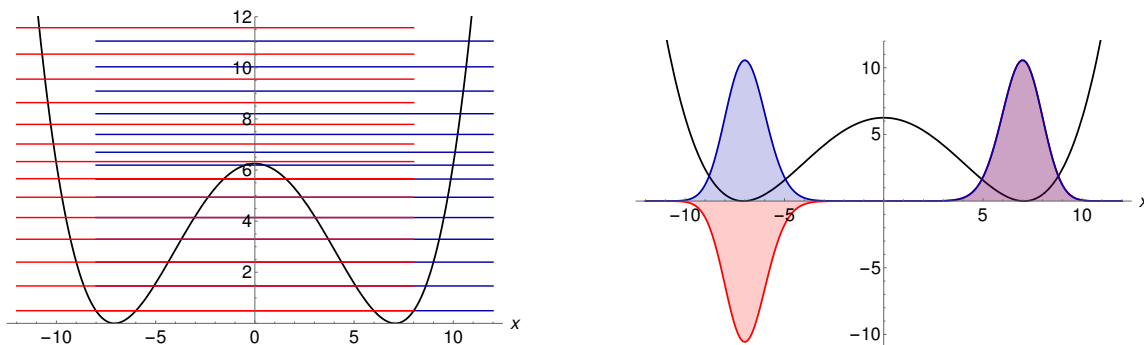
**Figure 3**. The structure of energy levels (left) and the profile of the two lowest energy eigenstates $\Psi_0^+$ and $\Psi_0^-$ (right). For energies lower than the top of the barrier $V_*$ odd eigenstates $\Psi_n^-$ (red) have slightly larger energy than even eigenstates $\Psi_n^+$ (blue).

paradox (see e.g., [2–5, 10, 11, 45–47]). Our toy-model shows that small, non-perturbative interactions can drastically change this structure.[7]

## 3.2 Black hole microstates

Another important characteristic of our toy-model is that it is solvable. The energy eigenstates and the corresponding energies can be computed numerically to arbitrary precision by various methods, most notably the standard Ritz-Rayleigh method.[8]

We will say that $f(\lambda)$ is *non-perturbatively small* if $f \sim 0$ as $\lambda \to 0^+$, where $\sim$ denotes the asymptotic expansion. Equivalently, $f(\lambda)$ is non-perturbatively small if $f(\lambda) = o(\lambda^n)$ for all $n \geq 0$ as $\lambda \to 0^+$. We will denote any non-perturbatively small terms by $o(\lambda^\infty)$. Finally we say that two quantities are *equal in perturbation theory* or *equal up to non-perturbative terms* if they have identical asymptotic expansions. These definitions are needed when considering, for example, the energy eigenstates of the full system, to which we now turn.

Since the double well potential is invariant under $x \to -x$ the operator $\Theta$ commutes with the Hamiltonian. Hence every energy eigenstate has definite parity, and the Hilbert space can be decomposed as $\mathcal{H} = \mathcal{H}^+ \oplus \mathcal{H}^-$, where $\Theta \mathcal{H}^+ = \mathcal{H}^+$ and $\Theta \mathcal{H}^- = -\mathcal{H}^-$. We denote energy eigenstates by

$$H\Psi_n^\pm = E_n^\pm \Psi_n^\pm , \qquad (3.16)$$

where $\Psi_n^\pm$ are even/odd eigenstates.[9] The corresponding energies satisfy $E_n^+ < E_n^-$ and their difference $\Delta E_n = E_n^- - E_n^+$ is exponentially small. In particular, the energy difference between the ground state and the first excited state is dominated by the 1-instanton

---

[7]Interactions in the form of shock waves between the left and right Hilbert spaces of two-sided eternal black holes in AdS were recently analyzed in [26, 27].

[8]One cannot rely on perturbation theory around a single minimum, since perturbative series diverge and lead to asymptotic expansions around $\lambda = 0$.

[9]We choose all eigenstates $\Psi_n^\pm$ to be real and normed to one. Overall signs are such that for $x \to \infty$, $\Psi_n^\pm \approx 2^{-1/2}\varphi_n^R$. The two lowest energy eigenstates $\Psi_0^\pm$ can be seen in figure 3.
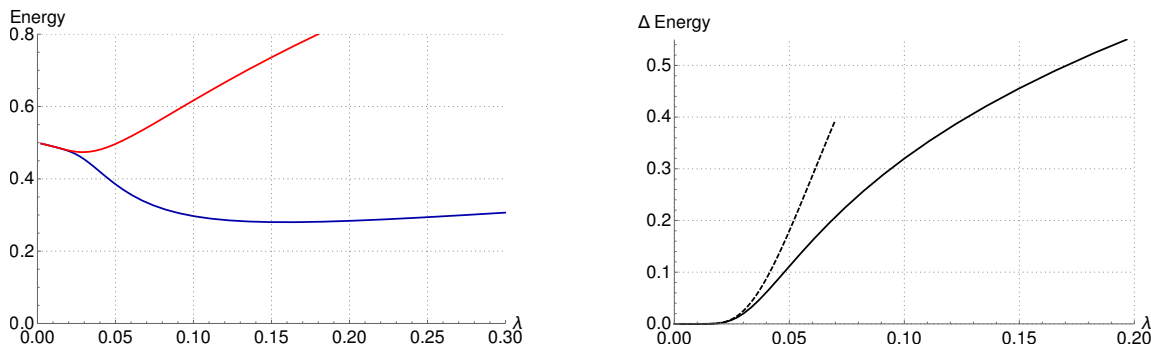
**Figure 4**. Left: energy of the vacuum (blue) and the first excited state (red) as a function of the coupling $\lambda$. Right: the difference between the two energies as a function of $\lambda$. The dashed line represents the leading 1-instanton approximation given by (3.17).

effect [48] and satisfies

$$E_0^- - E_0^+ = \frac{2}{\sqrt{\pi\lambda}} e^{-\frac{1}{6\lambda}} \left[1 + O(\lambda)\right]. \tag{3.17}$$

The vacuum energy $E_0^+$ is always smaller than $1/2$ — the energy of an unperturbed ground state of the harmonic oscillator.

The difference $\Delta E_n \sim 0$ in general is a non-perturbative effect and hence exponentially small for energies well below the maximum of the potential, $E_n^\pm < V_* = 1/(32\lambda)$. For energies larger than this, non-perturbative effects are numerically large and the difference $\Delta E_n \simeq 1/2$, as one can observe in figure 3. In this sense every pair of energy eigenstates $\Psi_n^\pm$ corresponds to two microstates with exponentially small energy splitting due to non-perturbative effects.

Motivated by the dual description of the black holes with scalar hair in section 2, where the (perturbative) degrees of freedom on both sides of the horizon correspond to excitations around two different perturbative vacua, we interpret states with significant support around both minima of the potential as the dual description in our toy-model of a *black hole microstate*. By contrast, semiclassical states centered around one of the two vacua only are interpreted as spacetimes without a black hole.[10] We consider microstates of any energy[11] $E \ll V_*$, but the lowest energy states are of a particular interest. Denote

$$\mathcal{M} = \{\alpha_+ \Psi_0^+ + \alpha_- \Psi_0^- \; : \; \alpha_\pm \in \mathbb{C}\}. \tag{3.18}$$

and consider any normalized state $\mu \in \mathcal{M}$. The energy of any such state is non-perturbatively close to the ground state energy, $\langle \mu | H | \mu \rangle = E_0^+ + o(\lambda^\infty)$. Hence, in perturbation theory the ground state is degenerate. For this reason we will refer to $\mathcal{M}$ as

---

[10]At first sight this description of microstates is at odds with the intuition that black holes should be high-energy states $E \sim N^2$, compared to the vacuum energy, as perceived by an asymptotic observer. However for an interacting system a number operator and the Hamiltonian are in general very different; whereas the Hamiltonian of the system is a unique operator once the time direction is chosen, a number operator is an inherently semiclassical object that depends on a choice of a semiclassical vacuum. We return to this point below.

[11]This restriction on the energies is needed for a perturbative description to be meaingful and resonates with the bulk where there are no hairy black holes above a certain mass (cf. figure 2(b)).

the subspace of *perturbative vacua* and each element $\mu \in \mathcal{M}$ will be called a *perturbative vacuum*. Roughly we have in mind a correspondence between the (degenerate) energy $E_n$ of the states and the mass of the black holes.

A relation between microstates and macrostates can be twofold. We first discuss this from the viewpoint of a single, say right, asymptotic observer with easy access to the right portion of the wave function only. This is the natural perspective if we consider our model to be a dual toy-model description of the single-sided hairy black holes discussed earlier. In this context the right portion of the wave function specifies a macrostate, and a set of microstates differing in the shape of the wave function around the left minimum can be considered.

A second characterisation of macrostates follows from considerations of a pair of observers in two distinct asymptotic regions as e.g., in the case of eternal black holes. From the point of view of two such perturbative observers a macrostate is given by two independent pieces of the wave function: the left portion, $\psi_L$, and the right portion, $\psi_R$. We can regard a macrostate as being represented by a tensor product $\psi_L \otimes \psi_R$, while a set of corresponding microstates is given by all states of the form $\alpha_L \psi_L + \alpha_R \psi_R$ for $\alpha_L, \alpha_R \in \mathbb{C}$. Each microstate is a specific continuation through the potential barrier that eludes both observers. We will discuss the relation between such macrostates and microstates in detail in section 4.

This interpretation also fits in the black hole paradigm of [49] and with a more general quantum perspective on black holes [10, 50, 51] according to which, from the point of view of a single asymptotic observer, the Hilbert space $\mathcal{H}$ factors as $\mathcal{H} \cong \mathcal{H}_{\text{coarse}} \otimes \mathcal{H}_{\text{fine}}$. The coarse degrees of freedom $\mathcal{H}_{\text{coarse}}$ are clearly distinguishable by the asymptotic observer within perturbation theory whereas $\mathcal{H}_{\text{fine}}$ contains fine degrees of freedom that require non-perturbative effects to identify. In our model $\mathcal{H}_{\text{fine}} \cong \mathbb{C}^2$ is a two-dimensional space, which can be identified with the space of perturbative vacua $\mathcal{M}$. The energy difference of any two microstates is then non-perturbatively small, and hence our model satisfies postulate 3 of [2, 52].

Finally the fact that in perturbation theory various microstates cannot be distinguished also leads to the Boltzmann entropy,

$$S_B = \log \dim \mathcal{H}_{\text{fine}} = \log 2. \tag{3.19}$$

This is the Bolzmann entropy associated with a single pair of harmonic oscillators.[12]

### 3.3 Firewalls?

We now explore further the implications of the above identification of black hole microstates in our toy-model. The model has two natural number operators that describe (perturbative) excitations from one or the other asymptotic viewpoint,

$$N_L = H_L^{(0)} = a_L^+ a_L, \qquad N_R = H_R^{(0)} = a_R^+ a_R \tag{3.20}$$

However, we are also interested in the description of observations from the viewpoint of an infalling observer with easy access to perturbative physics in both asymptotic regions,

---

[12]To get an area factor as in black holes, one should consider an ensemble of oscillators with different frequencies [43].
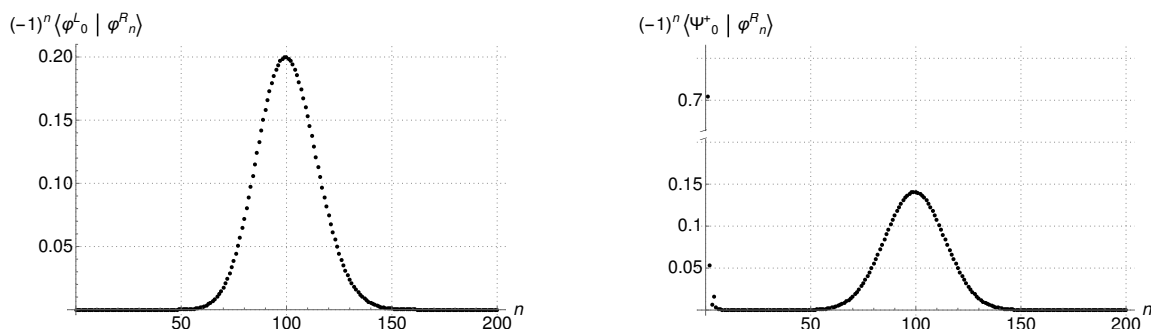
**Figure 5**. Left: matrix coefficients (3.29) between $\varphi_0^L$ and $\varphi_n^R$ as a function of $n$. The position of the maximum is always around $n \sim 1/(2\lambda)$. Right: matrix coefficients between the true ground state $\Psi_0^+$ and $\varphi_n^R$ as a function of $n$. Recall that in the leading order the ground state is given by (3.30). Hence the plot shows excitations of a few initial modes $\varphi_n^R$ for $n \sim 0$ as well as a wide peak around $n \sim 1/(2\lambda)$. In both figures $\lambda = 1/200$.

inside and outside the horizon of the black hole. A first guess for this is to consider the following number operator

$$N_A = N_L + N_R + O(\sqrt{\lambda}) = a_L^+ a_L + a_R^+ a_R + O(\sqrt{\lambda}), \tag{3.21}$$

possibly up to small corrections in $\lambda$.[13] Indeed, if $\lambda = 0$ this operator counts a sum of excitations of two decoupled harmonic oscillators. One would expect that with a small coupling $\lambda$, the sum $N_L + N_R$ recieves corrections of order $O(\sqrt{\lambda})$ in such a way that $N_A$ vanishes (or at least is small) in the new vacuum $\Psi_0^+$. However, to the contrary, the expectation value of $N_A$ in a generic state turns out to be very large,

$$\langle \psi | N_A | \psi \rangle \gtrsim \frac{1}{2} N^2 \left[ 1 + O\left(\frac{1}{N}\right) \right]. \tag{3.22}$$

Since this diverges as $N$ grows, in the language of [2] the microstate appears to exhibit a *firewall*. In particular even very low energy states including semiclassical vacua exhibit firewalls. Indeed, while

$$\langle \varphi_0^L | N_L | \varphi_0^L \rangle = \langle \varphi_0^R | N_R | \varphi_0^R \rangle = 0, \tag{3.23}$$

as expected, relation (3.14) leads to a large expectation value

$$\langle \varphi_0^L | N_R | \varphi_0^L \rangle = \langle \varphi_0^R | N_L | \varphi_0^R \rangle = \frac{1}{2\lambda} = \frac{1}{2} N^2. \tag{3.24}$$

By contrast, the energy of the state remains small,

$$\langle \varphi_0^L | H | \varphi_0^L \rangle = \langle \varphi_0^R | H | \varphi_0^R \rangle = \frac{1}{2} + \frac{3}{8} \lambda. \tag{3.25}$$

---

[13]Formally, the number operator for the infalling observer is of this form. In fact, for a fixed mode, the number operator for the infalling observer is non-perturbatively close to the number operator for the asymptotic observer, since $\langle N_A \rangle \sim e^{-8\pi\omega M}$ with $M$ of order $N$.

The large expectation values (3.24) are a manifestation of the fact that from the viewpoint of, say, the right minimum, the state $\varphi_0^L$ is a highly excited state. Indeed, since the full Hilbert space $\mathcal{H}$ is isomorphic to $\mathcal{F}_R$, both sets $\{\varphi_n^R\}_n$ and $\{\varphi_n^L\}_n$ span the entire Hilbert space $\mathcal{H}$ separately. We can decompose the left modes $\varphi_n^L$ in terms of right modes $\varphi_n^R$ as,

$$\varphi_m^L = \sum_{n=0}^{\infty} \langle \varphi_n^R | \varphi_m^L \rangle \varphi_n^R. \tag{3.26}$$

The value of the matrix element $\langle \varphi_n^R | \varphi_m^L \rangle$ can be calculated by noticing that the left and right modes are related by a displacement, up to a sign,

$$\varphi_n^L(x) = (-1)^n \varphi_n^R(x + a), \qquad a = 2x_R = \lambda^{-1/2}. \tag{3.27}$$

The matrix element then reads

$$\langle \varphi_m^L | \varphi_n^R \rangle = \langle \varphi_n^L | \varphi_m^R \rangle = \langle \varphi_n^R | \Theta | \varphi_m^R \rangle = \langle \varphi_m^L | \Theta | \varphi_n^L \rangle$$
$$= (-1)^n \sqrt{\frac{m!}{n!}} e^{-\frac{1}{4\lambda}} (2\lambda)^{\frac{1}{2}(m-n)} L_m^{(n-m)} \left( \frac{1}{2\lambda} \right), \tag{3.28}$$

where $L_m^{(\alpha)}$ denotes Laguerre polynomials. This expression simplifies for the semiclassical vacuum state $m = 0$, for which $L_0^{(n)} = 1$,

$$\langle \varphi_0^L | \varphi_n^R \rangle = \frac{(-1)^n e^{-\frac{1}{4\lambda}}}{\sqrt{2^n \lambda^n n!}}. \tag{3.29}$$

Figure 5 shows numerical values of these matrix elements as a function of $n$. For small $\lambda$ the distance between the minima is large and one can use Stirling's formula to find that (3.29) attains its maximum at $n = 1/(2\lambda)$. This shows that, in order to write a semiclassical vacuum state $\varphi_0^L$ as a superposition of the semiclassical states around the right minimum, $\varphi_n^R$, one needs to excite highly energetic states, namely those with $n$ of order $N^2$. Low energy states of the left asymptotic observer are detected as highly excited states by the right asymptotic observer and *vice versa*.

The above conclusions are directly applicable to the lowest energy eigenstates. It follows from perturbation theory that to leading order in the coupling $\lambda$ we have

$$\Psi_0^\pm = \frac{1}{\sqrt{2}} (\varphi_0^R \pm \varphi_0^L) + O(\sqrt{\lambda}). \tag{3.30}$$

Hence these states are all highly populated both with respect to $N_L$ and $N_R$. In particular equation (3.24) implies that

$$\langle \Psi_0^\pm | N_R | \Psi_0^\pm \rangle = \langle \Psi_0^\pm | N_L | \Psi_0^\pm \rangle = \frac{1}{4} N^2 \left[ 1 + O\left( \frac{1}{N} \right) \right]. \tag{3.31}$$

By considering a general microstate of the form $\mu \in \mathcal{M}$ the definition (3.21) leads to a firewall expressed by equation (3.22).

Thus we find that even the ground state $\Psi_0^+$ is a highly populated state of very small energy as measured by the total Hamiltonian (1.2). This seems paradoxical but is in fact
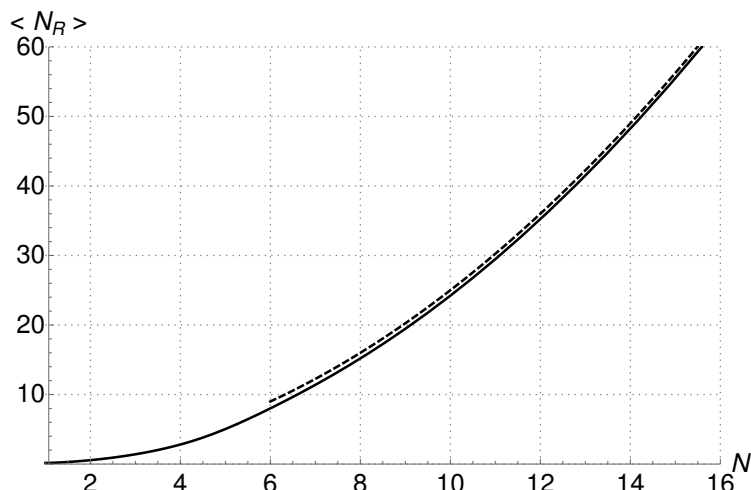
**Figure 6**. The expectation value of the naive number operator $N_R$ in the ground state $\Psi_0^+$ as function of $N = \lambda^{-1/2}$. The dashed line represents the leading term $N^2/4$ in equation (3.31).

just a consequence of the interacting nature of the system and related to non-perturbative effects. Actual numerical values of the matrix elements $\langle \Psi_0^+ | \varphi_n^R \rangle$ as a function of $n$ are shown in figure 5, while the expectation value of the right number operator $N_R$ in the ground state $\Psi_0^R$ is presented in figure 6.

To conclude, we have identified black hole microstates in our dual toy-model as relatively low energy states in the dual theory that are nevertheless heavily populated from the point of view of an asymptotic observer. The microstates are indistinguishable by an asymptotic observer with access to the perturbative physics only. A microstate structure emerges in our model as a consequence of non-perturbative level splitting (3.17) in the presence of interactions. This splitting can then be regarded as a source for the entropy (3.19).

At first sight the results of this section would seem to support the conclusions of [2, 3], i.e., the presence of firewalls. This, however, will turn out to be false. A caveat is that in the decoupling limit $\lambda \to 0$ the creation-annihilation operators $a_L, a_L^+, a_R, a_R^+$ do not approach the perturbative operators $b_L, b_L^+, b_R, b_R^+$ in any sense. In fact, as indicated by equation (3.14) such a limit is ill-defined. In section 4, starting from the interacting model, we carefully identify the perturbative degrees of freedom from the point of view of both observers, and we construct well-defined perturbative operators. In particular we will construct another set of creation-annihilation operators on $\mathcal{H}$, which will have a well-defined decoupling limit.

### 3.4 Limitations of perturbation theory

Before we proceed we pause to formulate precisely the limitations of the validity of perturbation theory around one of the minima in our model. This will be important in what follows.

a. *Perturbation theory breaks down when the overlap between the left and right semiclassical modes, $\langle \varphi_m^L | \varphi_n^R \rangle$ becomes significant when $n \sim m \sim N^2$.*

Notice that all matrix elements (3.28) are exponentially damped by a factor of

$e^{-1/(4\lambda)}$. In other words one could write $\langle \varphi_m^L | \varphi_n^R \rangle = O(e^{-1/(4\lambda)}) = o(\lambda^\infty)$. This is the correct behavior for an amplitude associated with the tunneling process, but it does not imply that the amplitude remains small for all states. Indeed, a degree of a Laguerre polynomial $L_n^{(\alpha)}(z)$ is equal to $n$, and the leading term is $(-1)^n z^n / n!$. Hence, for $n = m$ the matrix element becomes

$$\langle \varphi_n^L | \varphi_n^R \rangle = \frac{e^{-\frac{1}{4\lambda}}}{(2\lambda)^n n!} \left[ 1 + O(\lambda) \right]. \tag{3.32}$$

For $n \sim m \sim 1/(4\lambda)$ the Stirling's formula indicates that the denominator vanishes faster than the numerator. Therefore the non-perturbative terms become numerically large.

b. *Time-independent perturbation theory breaks down for states with occupancy numbers $m, n \sim N^2$.*
Since the potential (1.1) is quartic, $\langle \varphi_m^R | H | \varphi_n^R \rangle = 0$ if $|n - m| > 4$. Furthermore, $\langle \varphi_m^R | H | \varphi_n^R \rangle = O(\sqrt{\lambda})$ if $m \neq n$. Hence we can concentrate on the energy of the $n$-th perturbative state $\varphi_n^R$,

$$\langle \varphi_n^R | H | \varphi_n^R \rangle = \frac{1}{2} + n + \frac{3}{8}\lambda(2n^2 + 2n + 1). \tag{3.33}$$

Clearly, if $n$ is of order $1/\lambda$ the correction is of the same order than the unperturbed part. This is one of the many indications that the perturbative methods break down for states with occupancy numbers of order $1/\lambda$.

c. *The subleading terms in the commutation relation*

$$[H, a_R] = -a_R - \frac{3}{\sqrt{2}}\sqrt{\lambda} y_R^2 - \lambda\sqrt{2} y_R^3 \tag{3.34}$$

*become relevant whenever applied to states with occupancy numbers $n \sim N^2$.*
In the commutation relation (3.34) the correction is formally of order $\sqrt{\lambda}$. However, when applied to the state $\varphi_n^R$ with $n \sim 1/\lambda$ the unperturbed part is of the same order than the perturbation.

d. *Time-dependent perturbation theory breaks down for times $t \sim N$ for any state.*
Breakdown of perturbation theory can also be seen in time evolution. For example, time-dependent perturbation gives a matrix element

$$\langle \varphi_n^R | e^{-itH} | \varphi_m^R \rangle = \delta_{nm} - it \langle \varphi_n^R | H_R^{(1)} | \varphi_m^R \rangle + \ldots \tag{3.35}$$

When $|n - m| > 4$, the matrix element in the second term is identically zero. Hence such term becomes relevant in the expansion in $\lambda$, when either $n$, $m$ are of order $1/\sqrt{\lambda}$. Even for low occupancy states with $m, n$ of order 1 in $\lambda$, the perturbation theory breaks down after time $t \sim 1/\sqrt{\lambda} \sim N$. Notice that this is significantly shorter than the exponentially large tunneling time. In particular it agrees with the scrambling time of [23, 53] with entropy (3.19) and the mass $M \sim N$. We will discuss time-dependent processes in section 6 in more detail, where we will also recover the relation $M \sim N$ for our toy-model.

As we discussed the microstates $\Psi_0^\pm$ are highly excited. Hence, according to point $a$ above, perturbation theory is expected to break down whenever 'black hole microstate' effects are probed from the point of view of one of the semiclassical vacua. A very similar conclusion was reached in [54, 55] on the basis of gravitational (bulk) arguments. However this does not invalidate perturbation theory in general, which remains valid for excitations close to the semiclassical vacuum. In this sense postulate 2 of [2, 52] holds in our model.

## 4 Low energy excitations

In the previous section we identified several aspects of the holographic dictionary that relate our quantum mechanical toy-model to black hole physics in a dual bulk spacetime. In particular we established a notion of asymptotic observers, perturbative vacua, semiclassical states and their Fock spaces as well as dual black hole microstates. We have shown how non-perturbative effects enter in the picture leading to the breakdown of perturbation theory when it comes to the fine-grained features of microstates. We have also shown that the natural (naive) creation and annihilation operators (3.8) do not possess a well-defined decoupling limit $\lambda \to 0$, and therefore cannot represent creation-annihilation operators associated with the asymptotic regions. As a consequence a firewall (3.22) emerged.

In this section we correctly identify perturbative degrees of freedom as perceived by the asymptotic observers. We are able to distinguish perturbative and non-perturbative physics and to define suitable creation and annihilation operators with a well-defined decoupling limit.

### 4.1 Extinguishing firewalls

In section 3.3 we have shown that any microstate exhibits a firewall as measured by the naive number operator (3.21). We have established that the source of the firewall is the fact that the creation and annihilation operators $a_L, a_L^+, a_R, a_R^+$ act on both left and right perturbative states $\varphi_n^L$ and $\varphi_n^R$. A natural resolution would seem to be to define a different set of creation-annihilation operators $\hat{a}_L, \hat{a}_L^+$ and $\hat{a}_R, \hat{a}_R^+$ such that $\hat{a}_R$ and $\hat{a}_R^+$ act only on $\varphi_n^R$, while $\hat{a}_L$ and $\hat{a}_L^+$ act only on $\varphi_n^L$,

$$\hat{a}_R \varphi_n^R \overset{?}{=} \sqrt{n}\varphi_{n-1}^R, \qquad\qquad \hat{a}_R \varphi_n^L \overset{?}{=} 0, \qquad (4.1)$$

$$\hat{a}_R^+ \varphi_n^R \overset{?}{=} \sqrt{n+1}\varphi_{n+1}^R, \qquad\qquad \hat{a}_R^+ \varphi_n^L \overset{?}{=} 0 \qquad (4.2)$$

A new number operator $\hat{N}_A$ defined by means of the hatted operators

$$\hat{N}_A = \hat{a}_L^+ \hat{a}_L + \hat{a}_R^+ \hat{a}_R \qquad (4.3)$$

would then act on the energy eigenstates $\Psi_0^\pm$ according to (3.30) as

$$\hat{N}_A |\Psi_0^\pm\rangle = \frac{1}{\sqrt{2}}(a_R^+ a_R \varphi_0^R \pm a_L^+ a_L \varphi_0^L) + O(\sqrt{\lambda})$$
$$= 0 + O(\sqrt{\lambda}). \qquad (4.4)$$

Thus, no firewall!

The only problem with this reasoning is that the operators satisfying (4.1) or (4.2) cannot exist. Since the full Hilbert space $\mathcal{H}$ is isomorphic to any single Fock space associated to a minimum, the set $\{\varphi_n^L, \varphi_n^R\}_n$ constitutes an overcomplete basis. Given an action of $a_R$ on all right modes $\varphi_n^R$, its action on left modes $\varphi_n^L$ is fixed by means of (3.26). One could however hope to achieve relations (4.1) and (4.2) approximately for low energy modes $\varphi_n^L$ and $\varphi_n^R$ with $n \ll N$. In fact, according to (3.28), the overlap between $\varphi_m^L$ and $\varphi_n^R$ for $m, n \ll N$ is exponentially small, and the subset of modes $\{\varphi_m^L, \varphi_n^R\}_{m,n \ll N}$ constitutes an 'almost' orthonormal basis. Hence, we expect that the low energy physics should be well-approximated by the tensor product $\mathcal{F}_L \otimes \mathcal{F}_R$, where the excitations on the left and the right become independent.

We now give a specific proposal for 'orthogonalizing' the overcomplete basis $\{\varphi_n^L, \varphi_n^R\}_n$ in such a way that the hatted operators (4.1) or (4.2) can be successfully defined. To be more precise, we will split the total Hilbert space $\mathcal{H}$ into two orthogonal components, $\mathcal{H} = \mathcal{H}_L \oplus \mathcal{H}_R$. The left and right hatted annihilation operators $\hat{a}_L, \hat{a}_R$ can then be defined as projections of the unhatted operators $a_L, a_R$ onto the appropriate subspaces. The 'orthogonalization' is highly non-unique, but all ambiguities are non-perturbative and hence inaccessible in perturbation theory around any minimum. In the context of black hole physics the problem of overcompleteness of the basis has been pointed out in [56].

To resolve the overcompleteness of the set $\{\varphi_n^L, \varphi_n^R\}_n$ consider symmetric and antisymmetric combinations of all energy eigenstates,

$$\Psi_n^L = \frac{1}{\sqrt{2}}(\Psi_n^+ - \Psi_n^-), \qquad \Psi_n^R = \frac{1}{\sqrt{2}}(\Psi_n^+ + \Psi_n^-) \tag{4.5}$$

and consider two Hilbert subspaces of $\mathcal{H}$, spanned by $\Psi_n^L$ and $\Psi_n^R$ respectively,

$$\mathcal{H}_L = \mathrm{span}\{\Psi_n^L\}_n, \qquad \mathcal{H}_R = \mathrm{span}\{\Psi_n^R\}_n. \tag{4.6}$$

From (4.5) we have that $\langle \Psi_m^L | \Psi_n^R \rangle = 0$ for all $n, m$ and hence $\mathcal{H}_L$ and $\mathcal{H}_R$ are orthogonal to each other. The full Hilbert space splits into a *direct sum*,

$$\mathcal{H} = \mathcal{H}_L \oplus \mathcal{H}_R, \qquad \mathcal{H}_L \perp \mathcal{H}_R, \qquad \Theta\mathcal{H}_L = \mathcal{H}_R, \qquad \Theta\mathcal{H}_R = \mathcal{H}_L. \tag{4.7}$$

In other words $\Theta$ is a polarization of $\mathcal{H}$. We refer to $\mathcal{H}_L$ and $\mathcal{H}_R$ as *left* and *right perturbative Hilbert spaces* respectively.[14] Furthermore by $P_L$ and $P_R$ we denote canonical orthogonal projections of $\mathcal{H}$ onto $\mathcal{H}_L$ and $\mathcal{H}_R$ respectively.

Define $\hat{a}_R$ as $a_R$ restricted to $\mathcal{H}_R$ and similarly define $\hat{a}_L$ as $a_L$ restricted to $\mathcal{H}_L$,

$$\hat{a}_L = P_L a_L P_L, \qquad \hat{a}_R = P_R a_R P_R. \tag{4.8}$$

Define $\hat{a}_L^+$ and $\hat{a}_R^+$ as their Hermitian conjugates. We repeat this prescription to define a number operator $\hat{N}_A$, taking into account the fact that it should count excitations both in $\mathcal{H}_L$ and $\mathcal{H}_R$,

$$\hat{N}_A = P_L N_L P_L \oplus P_R N_R P_R = P_L a_L^+ a_L P_L + P_R a_R^+ a_R P_R. \tag{4.9}$$

---

[14]We will discuss the precise meaning of the word *perturbative* in the next section. For now notice that all low occupancy states $\Psi_n^R$ for $n \ll N$ are localized around the right minimum only, whereas $\Psi_n^L$ are localized around the left vacuum.
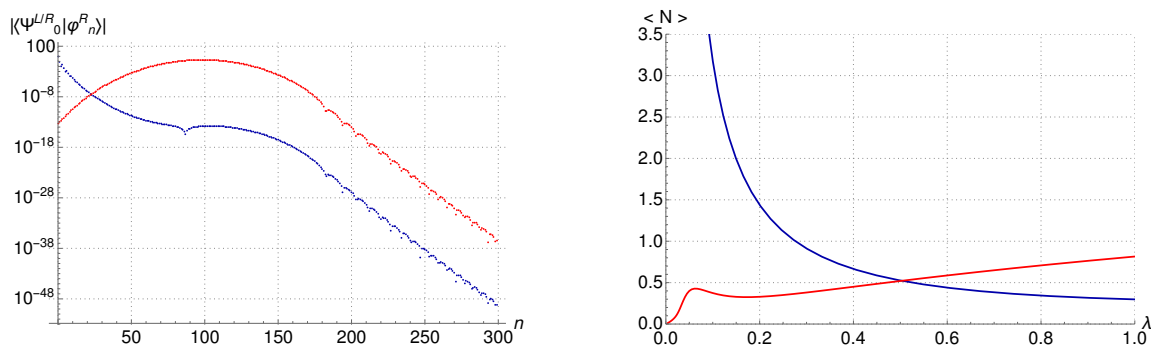
**Figure 7**. On the left: a comparison between matrix elements $\langle \Psi_0^R | \varphi_n^R \rangle$ (blue) and $\langle \Psi_0^L | \varphi_n^R \rangle$ (red). On the right: the expectation value $\langle \Psi_0^+ | N_A | \Psi_0^+ \rangle$ for number operators in the true vacuum state. The blue line shows the expectation value for the operator $N_A = a_L^+ a_L + a_R^+ a_R$. This operator leads to a firewall at $\lambda \to 0$. The red line shows the expectation value of the operator $\hat{N}_A = P_L a_L^+ a_L P_L + P_R a_R^+ a_R P_R$. The result approaches zero at $\lambda \to 0$ in a characteristic non-perturbative fashion and exhibits a small non-vanishing value for $\lambda > 0$ due to the interactions.

This operator is now defined globally on the entire $\mathcal{H}$. It counts excitations on top of microstates from the point of view of both perturbative vacua. In particular, we argue that its expectation value in a state $\psi = \alpha_L \varphi_m^L + \alpha_R \varphi_n^R$ representing approximately $m$ particles on the left and $n$ particles on the right, is given by

$$\langle \psi | \hat{N}_A | \psi \rangle = |\alpha_L|^2 m + |\alpha_R|^2 n + O(\sqrt{\lambda}). \tag{4.10}$$

Hence $\hat{N}_A$ is a natural candidate for a global, firewall-free number operator.[15] Specifically in any perturbative vacuum $\mu \in \mathcal{M}$, $\langle \mu | \hat{N}_A | \mu \rangle = O(\sqrt{\lambda})$.[16] The numerical plot of the expectation value of the number operator in the ground state $\Psi_0^+$ as a function of $\lambda$ is given in figure 7.

While mathematically we defined hatted operators in (4.8) using projectors, it may be more physically accurate *not* to specify the action of $\hat{a}_R$ on $\mathcal{H}_L$ nor $\hat{a}_L$ on $\mathcal{H}_R$. Similarly, we could define left and right number operators $N_L$ and $N_R$ on $\mathcal{H}_L$ and $\mathcal{H}_R$ only, as their

---

[15]There are two slightly different choices here. One can define $\hat{a}_L$ and $\hat{a}_R$ with their images unrestricted or restricted to the corresponding subspaces $\mathcal{H}_L$ and $\mathcal{H}_R$. The latter definition is $\hat{a}_R = P_R a_R P_R$ as we have defined, the former means that $\hat{a}_R = a_R P_R$, and similarly for $\hat{a}_L$. The difference, $P_L a_R P_R$, is however non-perturbatively small, as we will argue in section 4.3, and hence invisible in perturbation theory. For the same reason one can consider another number operator built up with hatted creation-annihilation operators

$$\hat{N}_A' = \hat{a}_L^+ \hat{a}_L \oplus \hat{a}_R^+ \hat{a}_R = P_L a_L^+ P_L a_L P_L + P_R a_R^+ P_R a_R P_R. \tag{4.11}$$

While formally different, the fact that $P_L a_R P_R = o(\lambda^\infty)$ as well as $P_R a_R P_L = o(\lambda^\infty)$ implies that the two number operators can differ by non-perturbative terms only, $\hat{N}_A = \hat{N}_A' + o(\lambda^\infty)$. Hence, in perturbation theory the two operators are indistinguishable. We will stick to the definition (4.9), which is slightly more convenient for numerical calculations.

[16]In [6–9] the Authors insist on a number operator $\hat{N}_A$ which satisfies $\hat{N}_A | \Psi_0^\pm \rangle = 0$ exactly. From the point of view of the QFT this seems unnecessarily strong, since interactions do create particles. Nevertheless, one can define the appropriate number operator $\hat{N}_A = \hat{A}_L^+ \hat{A}_L \oplus \hat{A}_R^+ \hat{A}_R$, where $\hat{A}_L$ is defined on $\mathcal{H}_L$ as $\hat{A}_L \Psi_n^L = \sqrt{n} \Psi_{n-1}^L$, and $\hat{A}_R$ is defined on $\mathcal{H}_R$ as $\hat{A}_R \Psi_n^R = \sqrt{n} \Psi_{n-1}^R$.
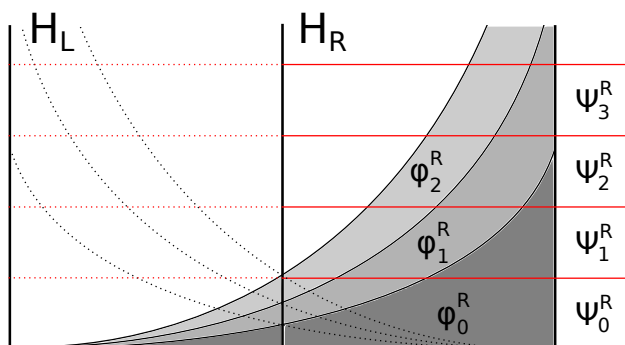
**Figure 8**. A (very) schematic structure of the Hilbert space. The full Hilbert space $\mathcal{H}$ is split into a simple sum $\mathcal{H} \cong \mathcal{H}_L \oplus \mathcal{H}_R$. 1-dimensional subspaces spanned by $\Psi_n^R$ do not align exactly with semiclassical states $\varphi_n^R$. One the other hand, the states $\varphi_n^R$ are not entirely contained in $\mathcal{H}_R$, as they have a small non-perturbative overlap with states in $\mathcal{H}_L$. The structure of $\mathcal{H}_L$ and coresponding states $\Psi_n^L$ and $\varphi_n^L$ is symmetric and denoted by dotted lines.

physical meaning is associated with perturbative physics percieved by the corresponding observers. We can either refuse to act with perturbative operators on non-perturbative states or accept the fact that natural perturbative observables from the point of view of a given observer become non-perturbative from the point of view of the another observer. The total number operator (3.21), however, remains globally defined.

In order to conclude the proof of (4.10) we have to study a relation between states $\varphi_n^R$ and $\Psi_n^R$, or equivalently between $\mathcal{H}_R$ and $\mathcal{F}_R$. The perturbation theory [48] implies that $\Psi_n^R = \varphi_n^R + O(\sqrt{\lambda})$, and hence

$$P_R \varphi_n^R = P_R \left[ \Psi_n^R + O(\sqrt{\lambda}) \right] = \Psi_n^R + O(\sqrt{\lambda}) = \varphi_n^R + O(\sqrt{\lambda}). \tag{4.12}$$

This implies that

$$\hat{a}_R^+ \hat{a}_R \varphi_n^R = P_R a_R^+ a_R P_R \varphi_n^R = n \varphi_n^R + O(\sqrt{\lambda}) \tag{4.13}$$

and (4.10) follows.

Equation (4.12) suggests that the familiar semiclassical state $\varphi_n^R$ is not an element of the right perturbative space $\mathcal{H}_R$. Indeed, if some $\varphi_n^R$ was an element of $\mathcal{H}_R$, then $\Theta \varphi_n^R = \varphi_n^L$ would belong to $\mathcal{H}_L$. But the two states $\varphi_n^R$ and $\varphi_n^L$ are not orthogonal as their scalar product (3.28) is non-vanishing. Hence $\varphi_n^R \notin \mathcal{H}_R$.

While no $\varphi_n^R$ belongs to $\mathcal{H}_R$, for each $n$ a difference between the state $\varphi_n^R$ and its projection $P_R \varphi_n^R$ on $\mathcal{H}_R$ is non-perturbatively small. Equivalently, by following Example 6 of section XII.3 of [44] one can argue that $\|P_L \varphi_n^R\| = o(\lambda^\infty)$ and $\|P_R \varphi_n^L\| = o(\lambda^\infty)$ for any $n$. Hence, while $\varphi_n^R$ is not an element of the perturbative Hilbert space $\mathcal{H}_R$, its projection $P_R \varphi_n^R \in \mathcal{H}_R$ is non-perturbatively close to $\varphi_n^R$. In perturbation theory, one cannot distinguish the two states. A schematic relation between various states is presented in figure 8.

By defining left and right perturbative spaces $\mathcal{H}_L$ and $\mathcal{H}_R$ we effectively resolved the overcompleteness of the set $\{\varphi_n^L, \varphi_n^R\}_n$. While stricktly speaking no semiclassical state $\varphi_n^R$ belongs to $\mathcal{H}_R$, there exist states $P_R \varphi_n^R$ non-perturbatively close to $\varphi_n^R$ lying in $\mathcal{H}_R$. We

have found 'approximate isomorphisms'

$$\mathcal{F}_L \cong \mathcal{H}_L + o(\lambda^\infty), \qquad \mathcal{F}_R \cong \mathcal{H}_R + o(\lambda^\infty) \qquad (4.14)$$

up to non-perturbative terms.

## 4.2 Perturbative states

A notion of a perturbative state is crucial for the discussion of the information paradox. An intuitive idea is that its support is concentrated around a single minimum. As an example consider basis states $\varphi_n^R$. Are all these states perturbative with respect to the right minimum or only those with $n \ll N^2$, so that their support is concentrated around the minimum? In the language of [6–9] one would call a state $\varphi_n^R$ perturbative only if $n \ll N^2$ is small. In this paper, however, we will introduce a weaker definition that allows for a wider range of perturbative states.

Our definition of a perturbative state deals with the behavior of the state as the coupling $\lambda$ approaches zero. Therefore, instead of a single state $\psi$, we consider a family $\{\psi_\lambda\}_{\lambda>0}$ labeled by the coupling. Essentially all states we consider depend on $\lambda$ in some implicit way. For example the perturbative states $\varphi_n^R$ and $\varphi_n^L$ are defined by (3.5)–(3.6) and hence they implicitly depend on $\lambda$. For that reason we will refer to the elements of the family $\{\psi_\lambda\}_{\lambda>0}$ as a state $\psi_\lambda \in \mathcal{H}$.

We say that the state $\psi_\lambda$ is *perturbative with respect to the right minimum* if $F_R^{-1}\psi_\lambda$ converges in norm in $\mathcal{F}_R$ when $\lambda \to 0^+$. Here $F_R^{-1} : \mathcal{H} \xrightarrow{\cong} \mathcal{F}_R$ is the inverse of the isomorphism (3.5) between $\mathcal{H}$ and $\mathcal{F}_R$. Analogously we define states perturbative with respect to the left minimum.

First notice that all states $\varphi_n^R$ are mapped to $|n\rangle_R \in \mathcal{F}_R$, which are $\lambda$-independent in $\mathcal{F}_R$. Hence all $\varphi_n^R$ are trivially perturbative with respect to the right minimum. Consider now a state such as a ground state $\Psi_0^+$, which possesses two bumps around both left and right minimum. By going to $\mathcal{F}_R$ we may simply position ourselves at $x = x_R$ and send $\lambda$ to zero. The right portion of the wave function then concentrates around the right minimum and approaches $\varphi_0^R$. As the left minimum moves away to $-\infty$, the left portion of the wave function is lost in the decoupling limit $\lambda = 0$. Indeed, neither $\Psi_0^+$ nor $\Psi_0^-$ are perturbtive with respect to any minimum.

On the other hand, all states $\Psi_n^R$ and $\Psi_n^L$ as defined in (4.5) are perturbative with respect to right and left minima respectively and we have, [44], $\lim_{\lambda\to0^+} F_R^{-1}\Psi_n^R = |n\rangle_R$ and $\lim_{\lambda\to0^+} F_L^{-1}\Psi_n^L = |n\rangle_L$ with the convergence in norm.[17] Hence every element of $\mathcal{H}_L$ is perturbative with respect to the left minimum and every element of $\mathcal{H}_R$ is perturbative

---

[17]Even if a state $\psi_\lambda$ is non-perturbative with respect to, say, right minimum, one can still define its decoupling limit. We will say that a state $|\psi_0\rangle \in \mathcal{F}_R$ is a *decoupling limit with respect to the right minimum* of $\psi_\lambda \in \mathcal{H}$ if $|\psi_0\rangle = \text{wlim}_{\lambda\to0^+} F_R^{-1}\psi_\lambda$, where wlim denotes the weak limit. In this sense we have the following decoupling limits,

$$\text{wlim}_{\lambda\to0^+} F_R^{-1}\Psi_n^R = |n\rangle_R, \qquad \text{wlim}_{\lambda\to0^+} F_R^{-1}\Psi_n^L = 0, \qquad \text{wlim}_{\lambda\to0^+} F_R^{-1}\Psi_n^\pm = \frac{1}{\sqrt{2}}|n\rangle_R. \qquad (4.15)$$

Numerical values of the matrix elements $\langle\Psi_0^L|\varphi_n^R\rangle$ and $\langle\Psi_0^R|\varphi_n^R\rangle$ as functions of $n$ can be seen in figure 7.

with respect to the right minimum. This justifies their names as left and right *perturbative* Hilbert spaces $\mathcal{H}_L$ and $\mathcal{H}_R$. We can also sharpen our definition of a *right (left) observer* by declaring $\mathcal{H}_R$ ($\mathcal{H}_L$) as the Hilbert space available to the observer.

Notice that our definition of a perturbative state depends only on what happens with the state when $\lambda$ approaches to zero. For example, for a fixed value of $\lambda > 0$ the support of $\varphi_n^R$ is concentrated around the right minimum only for $n \ll \lambda^{-1} = N^2$. Nevertheless, according to our definition, all $\varphi_n^R$ are perturbative. As $\lambda$ approaches zero, each $\varphi_n^R$ concentrates around the right minimum, since for each $n$ there exists $\lambda$ so small that $n \ll \lambda^{-1}$.

In the language of [6–9] the space of perturbative states was finitely dimensional, as the condition $n \ll N^2$ on the occupancy numbers was imposed. In particular such a space was not generated by a genuine algebra acting on a cyclic vector: the issue that led the Authors of [7] to use a concept of 'algebras with a cut-off'. In our approach such issues are completely avoided. The 'small algebra' $\mathcal{A}_R$ associated with the right minimum is generated by $\hat{a}_R$ and $\hat{a}_R^+$ and the perturbative Hilbert space is then $\mathcal{H}_R = \mathcal{A}_R |\Psi_0^R\rangle$. No cut-offs of any sort are required and all states in $\mathcal{H}_R$ are perturbative with respect to the right minimum.

### 4.3 Perturbative operators

Having defined perturbative states, one can also define perturbative operators. These should be operators which: (i) preserve the decoupling of the potential wells up to non-perturbative effects; and (ii) have a well-defined decoupling limit. If we write an operator $A$ in the matrix form

$$A = \begin{pmatrix} A_{LL} & A_{LR} \\ A_{RL} & A_{RR} \end{pmatrix} \; : \; \mathcal{H}_L \oplus \mathcal{H}_R \to \mathcal{H}_L \oplus \mathcal{H}_R \tag{4.16}$$

with $A_{IJ}$ mapping $\mathcal{H}_I$ into $\mathcal{H}_J$, $I, J \in \{L, R\}$, then: (i) $A_{LR}$ and $A_{RL}$ must be non-perturbatively small, i.e.,

$$A_{LR} = o(\lambda^\infty) \text{ and } A_{RL} = o(\lambda^\infty); \tag{4.17}$$

and (ii) the decoupling limits $\lambda \to 0$ of $A_{LL}$ in $\mathcal{H}_L$ and $A_{RR}$ in $\mathcal{H}_R$ must exist. We will call such an operator $A$ as *perturbative*.

All hatted operators such as creation-annihilation operators $\hat{a}_L, \hat{a}_L^+, \hat{a}_R, \hat{a}_R^+$ or the number operator $\hat{N}_A$ (4.9) are by construction perturbative with vanishing off-diagonal terms. Their unhatted versions usually fail to satisfy condition (ii), as indicated by (3.14) or by the firewall in (3.22). On the other hand, one expects that condition (i) holds, since $P_R a_R P_R \varphi_n^R = \sqrt{n}\varphi_{n-1}^R + o(\lambda^\infty)$ or equivalently $P_L a_R P_R = o(\lambda^\infty)$. Figure 9 shows the norm of the state $P_L a_R \Psi_0^R$ with a characteristic exponential fall-off around $\lambda = 0$. In this sense hatted and unhatted creation-annihilation operators agree on their corresponding perturbative Hilbert spaces. Schematically, $a_R = \hat{a}_R + o(\lambda^\infty)$ on $\mathcal{H}_R$ and $a_L = \hat{a}_L + o(\lambda^\infty)$ on $\mathcal{H}_L$ together with their conjugates. Only on $\mathcal{H}_L$, the complement of $\mathcal{H}_R$, the operators $a_R$ and $\hat{a}_R$ differ significantly.

The unhatted creation-annihilation operators satisfy commutation relation (3.15). However, in black hole physics, locality at the level of the effective bulk theory requires
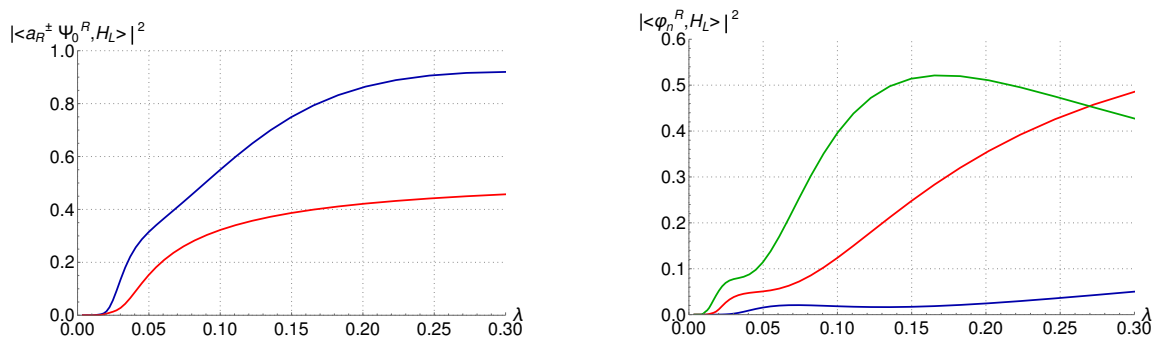
**Figure 9**. Left: a measure of an overlap of the states $a_R \Psi_0^R$ (blue) and $a_R^+ \Psi_0^R$ (red) with the left Hilbert space $\mathcal{H}_L$ as function of the coupling $\lambda$. Right: a measure of an overlap of the perturbative states $\varphi_n^R$ with $\mathcal{H}_L$ for $n = 0$ (blue), $n = 1$ (red) and $n = 2$ (green). The overlap of a state $\psi \in \mathcal{H}$ with $\mathcal{H}_L$ is defined as $|\langle \psi, \mathcal{H}_L \rangle|^2 = \|P_L \psi\|^2$. For $\lambda \to 0$ the overlaps approach zero in a characteristic non-perturbative fashion.

that left and right operators commute. This is indeed the case for the hatted operators, where we find

$$[\hat{a}_L, \hat{a}_L^+] = [\hat{a}_R, \hat{a}_R^+] = 1 + o(\lambda^\infty), \qquad [\hat{a}_L, \hat{a}_R] = [\hat{a}_L, \hat{a}_R^+] = 0. \qquad (4.18)$$

The canonical commutation relations on $\mathcal{H}_L$ and $\mathcal{H}_R$ are altered by a non-perturbative factor, while left and right operators commute.[18] Hence locality is maintained up to non-perturbative effects as predicted by a number of papers, e.g., [6, 7, 49, 57, 58]. This, however, comes at the cost of the action of, say, right perturbative operators on the left perturbative states to be either unrelated to the action of unhatted creation-annihilation operators (extremely 'non-local' as in [29, 59]), or simply ill-defined [60].

In this way we have resolved the problem of 'fitting' the left creation-annihilation operators into the full Hilbert space. By identifying the right perturbative Hilbert space as $\mathcal{H}_R$ and the right perturbative operators as $\hat{a}_R$ and $\hat{a}_R^+$ we found the proper set of degrees of freedom and operators associated with a perturbative (asymptotic) observer in the right vacuum. In particular every state in $\mathcal{H}_R$ is perturbative with respect to the right minimum, but non-perturbative from the point of view of the left observer.

In the context of black holes a number of papers [61–64] have suggested to remove 'half' of the states by considering an antipodal identification of the spacetime. The total Hilbert space then consists of only parity even or odd wave functions. Such an approach would remove a degeneracy within perturbation theory leading to each black hole having a single microstate. In our model we do not find a support for such an identification, but we also do not find any obstacles in its implementation. From the point of view of a single observer, the two situations are indistinguishable as long as perturbative processes of small energies are considered. Only at higher energies one would be able to notice 'missing' states in the total Hilbert space.

---

[18]Slightly different definitions of creation-annihilation operators as indicated in section 4.1 may result in non-perturbative corrections to the locality condition $[\hat{a}_L, \hat{a}_R] = [\hat{a}_L, \hat{a}_R^+] = o(\lambda^\infty)$. In perturbation theory operators are required to be defined uniquely only up to non-perturbative terms.

Finally let us point out that the original Hamiltonian $H$ is a perturbative operator. The off-diagonal elements $P_R H P_L = o(\lambda^\infty)$ and $P_L H P_R = o(\lambda^\infty)$ are related to the tunneling rate and can be calculated by standard methods within the WKB approximation. With our definition of non-perturbative effects this statement remains true for all energies, even if actual matrix elements become numerically large. We return to time-dependent processes in section 6.

## 5 Low energy effective theory: state-dependence

In the previous section we have constructed states and operators in the full theory that are the natural perturbative objects from the standpoint of semiclassical (or asymptotic) observers. In this section we relate these operators (observables) in the full theory to operators (observables) in the effective low energy theory. In doing so we will find that the resulting operators in the effective theory are state-dependent.

### 5.1 Perturbative observables

In semiclassical black hole physics one usually takes the Hilbert space to be a tensor product $\mathcal{F}_L \otimes \mathcal{F}_R$ of the degrees of left/right freedom on both sides of the horizon. The creation-annihilation operators $b_L, b_L^+$ in $\mathcal{F}_L$ and $b_R, b_R^+$ on $\mathcal{H}_R$ give rise to creation-annihilation operators in the tensor product: $\mathbb{I} \otimes b_R, \mathbb{I} \otimes b_R^+$, *etc.* The bulk of the Hermitian operators are of the form $A_{LL} \otimes \mathbb{I} + \mathbb{I} \otimes A_{RR}$. The black hole microstate is not a vacuum state, but rather an excited state.

Due to the large potential barrier, low occupancy states in our toy-model exhibit an approximate tensor product structure $\mathcal{H}_L \otimes \mathcal{H}_R$. Indeed, states $\Psi_{mn} = \Psi_m^L + \Psi_n^R$ for $m, n \ll N$ resemble elements of the tensor product, as they are combinations of two, nearly decoupled states centered around, respectively, the left and right perturbative vacuum, with a number of excitations given by $\langle \Psi_{mn} | \hat{N}_A | \Psi_{mn} \rangle = m + n + O(\sqrt{\lambda})$ as expected.

Beyond this, however, the two systems differ. A linear structure of the direct sum $\mathcal{H}_L \oplus \mathcal{H}_R$ and the tensor product $\mathcal{H}_L \otimes \mathcal{H}_R$ is vastly different and we cannot hope to reproduce all perturbative operators on the tensor product in the full theory. In particular the action of linear operators in the tensor product is different and there are significantly 'more' linear operators on the tensor product.

Consider now an operator $B = A_{LL} \otimes \mathbb{I} + \mathbb{I} \otimes A_{RR}$ on $\mathcal{H}_L \otimes \mathcal{H}_R$ and assume $\chi_m^L$ and $\chi_n^R$ form a set of normalized eigenfunctions of $A_{LL}$ and $A_{RR}$ respectively, with eigenvalues $\lambda_m^L$ and $\lambda_n^R$. This means that

$$B(\chi_m^L \otimes \chi_n^R) = (\lambda_m^L + \lambda_n^R)(\chi_m^L \otimes \chi_n^R). \tag{5.1}$$

We can use the identification of the direct sum with the direct product $\mathcal{H}_L \oplus \mathcal{H}_R \cong \mathcal{H}_L \times \mathcal{H}_R$ to define an operator $A$ on $\mathcal{H}_L \oplus \mathcal{H}_R$ as

$$A(\chi_m^L, \chi_n^R) = (\lambda_m^L + \lambda_n^R)(\chi_m^L, \chi_n^R). \tag{5.2}$$

The operator $A$ acts on a state $\chi_m^L + \chi_n^R$ in the same way as $B$ acts on $\chi_m^L \otimes \chi_n^R$. Furthermore the expectation values match,

$$\langle \chi_m^L \otimes \chi_n^R | B | \chi_m^L \otimes \chi_n^R \rangle = \lambda_m^L + \lambda_n^R = \langle (\chi_m^L, \chi_n^R) | A | (\chi_m^L, \chi_n^R) \rangle. \tag{5.3}$$

In this sense $A$ realizes the same relations on $\mathcal{H}_L \oplus \mathcal{H}_R$ as $B$ on $\mathcal{H}_L \otimes \mathcal{H}_R$.

But the operator $A$ is not linear. It is not even very clear how to extend it to the entire Hilbert space $\mathcal{H}_L \oplus \mathcal{H}_R$. For instance, we may consider projections $A_L$ and $A_R$ of $A$ on $\mathcal{H}_L$ and $\mathcal{H}_R$, so that $A = A_L \oplus A_R$. Operators $A_L$ and $A_R$ can be extended separately to bilinear operators by linearity in each argument. For example, we find that for any $\psi_R \in \mathcal{H}_R$

$$A_R(\chi_m^L, \psi_R) = P_R A(\chi_m^L, \psi_R) = (\lambda_m^L \mathbb{I} + A_{RR}) \psi_R. \tag{5.4}$$

This is clearly a linear operator with respect to the right portion of the state, $\psi_R \in \mathcal{H}_R$. The appearance of $\lambda_m^L$, however, amounts to a form of 'state-dependence': the value of the right operator $A_R$ depends on the 'hidden' left portion of the state. In this state-dependent sense we can think about $A_R$ as an operator on $\mathcal{H}_R$ only, and we can write $A_R^{\psi_L}(\psi_R) = A_R(\psi_L, \psi_R)$. That is, we can regard $A_R^{\psi_L}$ as an observable of the right observer. Its value on $\psi_R$, however, depends on the 'black hole microstate' $\psi_L$, i.e., on the shape of the wave function on the other side of the potential barrier.

In section 5.3 we relax our assumption that the action of the operator $A$ takes the form (5.2). Starting from the weaker requirement that expectation values of operators in the full theory agree with those of the corresponding operators in the effective theory, up to non-perturbative terms, we will show there that the corresponding global operators again cannot be defined in a 'state-independent' manner. However, the 'state-dependence' then no longer pertains to the entire portion of the 'hidden' wave function $\psi_L$, but instead is restricted to a choice of perturbative vacuum state $\mu \in \mathcal{M}$. More generally, regardless of the details, we find one cannot realize perturbative observables in the full theory as linear operators. Mathematically this is a consequence of the incompatibility of the linear structure of the direct sum and that of the tensor product. Physically, this means that perturbative interpretations of various operators depend on microstates of the system, as argued in [6, 7].

## 5.2 The tensor product: loss of information

In the previous section we have argued that it is impossible to represent all perturbative observables on the tensor product $\mathcal{H}_L \otimes \mathcal{H}_R$ by linear operators in the full theory $\mathcal{H}_L \oplus \mathcal{H}_R$. On the other hand, the low energy physics in our model does effectively take place on the tensor product. One therefore expects that an effective theory based on the tensor product should capture the low energy physics, up to non-perturbative effects.

Let $s : \mathcal{H}_L \times \mathcal{H}_R \to \mathcal{H}_L \otimes \mathcal{H}_R$ denote the canonical bilinear map $s(\chi_L, \chi_R) = \chi_L \otimes \chi_R$. Since $\mathcal{H}_L \oplus \mathcal{H}_R \cong \mathcal{H}_L \times \mathcal{H}_R$ we can assign to any state $\chi = \chi_L + \chi_R$, with $\chi_L \in \mathcal{H}_L$ and $\chi_R \in \mathcal{H}_R$ an effective state $\chi^{\text{eff}}$ as

$$\chi = \chi_L + \chi_R \ \longmapsto \ \chi^{\text{eff}} = \mathcal{N} s(\chi) = \mathcal{N} \, \chi_L \otimes \chi_R, \tag{5.5}$$

where $\mathcal{N}$ is a normalization. If $s(\chi) \neq 0$, we choose

$$\mathcal{N}^2 = \frac{\|\chi_L + \chi_R\|^2}{\|\chi_L \otimes \chi_R\|^2} = \frac{\|\chi_L\|^2 + \|\chi_R\|^2}{\|\chi_L\|^2 \|\chi_R\|^2}, \tag{5.6}$$

so that the norms of $\chi$ and $\chi^{\mathrm{eff}}$ are equal.

First notice that if either $\chi_L = 0$ or $\chi_R = 0$, then $\chi^{\mathrm{eff}} = 0$. In other words if a state $\chi \in \mathcal{H}$ is perturbative with respect to any minimum, then $\chi^{\mathrm{eff}} = 0$. We will say that a state is *typical* if it is represented by a non-vanishing effective state in the effective theory. Clearly, every typical state $\chi = \chi_L + \chi_R$ is generic in the sense that upon 'random choice' of $\chi_L$ and $\chi_R$ it is unlikely to end up with an atypical state. Only typical states can be represented in the effective theory.

The image of the map (5.5) consists of a set $\mathcal{H}_{\mathrm{pert}}$ of all simple tensors $\psi_L \otimes \psi_R \in \mathcal{H}_L \otimes \mathcal{H}_R$. A pre-image of a given simple tensor $\psi_L \otimes \psi_R$, however, is not unique. Every normed state $\psi^{\mathrm{eff}} \in \mathcal{H}_{\mathrm{pert}}$ can be written as $\psi^{\mathrm{eff}} = e^{i\theta} \psi_L \otimes \psi_R$ with $\|\psi_L\| = \|\psi_R\| = 1$ and an irrelevant phase $\theta$. The most general form of $\psi \in \mathcal{H}$ mapped onto $\psi^{\mathrm{eff}}$ is

$$\psi = \alpha_L \psi_L + \alpha_R \psi_R, \text{ with } |\alpha_L|^2 + |\alpha_R|^2 = 1, \tag{5.7}$$

where $\arg \alpha_L + \arg \alpha_R = \theta + 2\pi n$. We will refer to any effective state $\psi^{\mathrm{eff}}$ as a *macrostate* whereas all corresponding $\psi$ are *microstates*. Many different states $\psi$ are mapped onto the same macrostate in the effective theory.

In our toy-model, all possible microstates corresponding to a given macrostate $\psi^{\mathrm{eff}}$ are parametrized by a unit vector $(\alpha_L, \alpha_R) \in \mathbb{C}^2$. Equivalently, this ambiguity amounts to a choice of a normed perturbative vacuum $\mu \in \mathcal{M}$. Physically this means that knowledge of the left and right portions $\psi_L$ and $\psi_R$ of the wave function is not sufficient to reconstruct the full wave function $\alpha_L \psi_L + \alpha_R \psi_R$. Instead a prescription for the continuation of the wave functions through the potential barrier is needed. This is clearly a non-perturbative effect, and hence invisible in the effective theory.

Given a simple tensor $\psi_L \otimes \psi_R \in \mathcal{H}_{\mathrm{pert}}$, one can define its projection on, say, $\mathcal{H}_R$ as $\psi_L \otimes \psi_R \mapsto \psi_R$. From the point of view of the right vacuum these are pure states since, when traced over $\mathcal{H}_L$, they lead to the pure state density matrix $|\psi_R\rangle\langle\psi_R|$. Every other state $\psi^{\mathrm{eff}} \in \mathcal{H}_L \otimes \mathcal{H}_R$ that does not belong to $\mathcal{H}_{\mathrm{pert}}$ can be regarded as a mixed state with the density matrix $\rho_R = \mathrm{Tr}_L |\psi^{\mathrm{eff}}\rangle\langle\psi^{\mathrm{eff}}|$.

Note also that from the point of view of the right asymptotic observer, only states that lie in $\mathcal{H}_R \subseteq \mathcal{F}_R$ are perturbative. States beyond $\mathcal{H}_R$ lack any perturbative interpretation. Hence states in $\mathcal{F}_L \otimes \mathcal{F}_R$ that do not belong to $\mathcal{H}_L \otimes \mathcal{H}_R$ lack a perturbative description from the standpoint of both asymptotic regions, even as mixed states.

## 5.3 Operators in the effective theory

Given an operator $A$ in the full theory, we would also like to construct an operator $A^{\mathrm{eff}}$ in the effective theory such that

$$\langle\psi|A|\psi\rangle = \frac{1}{Z}\langle\psi^{\mathrm{eff}}|A^{\mathrm{eff}}|\psi^{\mathrm{eff}}\rangle + o(\lambda^\infty), \tag{5.8}$$

assuming $\psi$ and $\psi^{\text{eff}}$ are normalized to one. The proportionality factor $Z$ should correspond to the number of microstates $\psi$ represented by an identical macrostate $\psi^{\text{eff}}$. The relation (5.8) is known as the equilibrium condition [42]. It is in this sense that the full theory is realized by the effective theory up to non-perturbative terms.

Note that in order for (5.8) to hold, one must have $\psi^{\text{eff}} \neq 0$ if $\psi$ is non-zero, i.e., the state $\psi$ must be typical. Secondly, off-diagonal elements of $A$ must be non-perturbatively small, i.e., the operator $A$ must be perturbative. Given a perturbative operator $A$ a natural guess for its effective counterpart $A^{\text{eff}}$ would be

$$B = A_{LL} \otimes \mathbb{I} + \mathbb{I} \otimes A_{RR}. \tag{5.9}$$

Many operators considered in the effective theory are expected to be of this form. Unfortunately, $B$ does not satisfy relation (5.8). Indeed, consider a normed state $\psi = \alpha_L \psi_L + \alpha_R \psi_R$ with $\|\psi_L\| = \|\psi_R\| = \|\psi\| = 1$. The effective state is $\psi^{\text{eff}} = e^{i\theta} \psi_L \otimes \psi_R$ with an irrelevant overall phase $\theta$, which drops out from the expectation value. The left hand side of (5.8) then reads

$$\langle \psi | A | \psi \rangle = |\alpha_L|^2 \langle \psi_L | A_{LL} | \psi_L \rangle + |\alpha_R|^2 \langle \psi_R | A_{RR} | \psi_R \rangle + o(\lambda^{\infty}) \tag{5.10}$$

while the right hand side is

$$\langle \psi^{\text{eff}} | B | \psi^{\text{eff}} \rangle = \langle \psi_L | A_{LL} | \psi_L \rangle + \langle \psi_R | A_{RR} | \psi_R \rangle. \tag{5.11}$$

The mismatch is not surprising: the correlator $\langle \psi | A | \psi \rangle$ clearly distinguishes specific microstates of the system, whereas $\langle \psi^{\text{eff}} | B | \psi^{\text{eff}} \rangle$ depends on the overall macrostate only.

In order for the condition (5.8) to hold one possibility is for the system to be in a special 'equilibrium' state, and with an operator $A$ that does not distinguish between microstates. Mathematically, if $\Theta \psi = \pm \psi$ and $A_{LL} = \Theta A_{RR} \Theta$, then indeed $A^{\text{eff}} = B$ and we find

$$\langle \psi | A | \psi \rangle = \frac{1}{2} \langle \psi^{\text{eff}} | A^{\text{eff}} | \psi^{\text{eff}} \rangle + o(\lambda^{\infty}). \tag{5.12}$$

The $Z$ factor accounts for the degeneracy of the macrostate, $Z = e^{S_B} = 2$, where $S_B$ is the Boltzmann entropy (3.19).

What are equilibrium states in our model? Notice that the condition $\Theta \psi = \pm \psi$ means $\alpha_R = \pm \alpha_L = 2^{-1/2}$. Hence every energy eigenstate $\Psi_n^{\pm}$ is an equilibrium state. Furthermore every state of fixed parity, even or odd, is also an equilibrium state. Operators satisfying $A_{LL} = \Theta A_{RR} \Theta$ are those that act on both sides of the potential well 'in the same way' regardless of a specific microstate. This is closely related to the definition of operators satisfying the Eigenstate Thermalization Hypothesis, e.g., [65–67]. In particular the total number operator and the Hamiltonian are of this form.

By contrast, if the microstate $\psi$ is not an equilibrium state, then (5.10) depends on the specific values of $\alpha_L$ and $\alpha_R$. Given a global operator $A$, we can construct a class of effective operators $A_\mu^{\text{eff}}$, which depend on these parameters. Indeed, by comparing (5.10) with (5.11) we see that we need

$$A_\mu^{\text{eff}} = A_{LL} \otimes |\alpha_R|^2 \mathbb{I} + |\alpha_L|^2 \mathbb{I} \otimes A_{RR}. \tag{5.13}$$

In this case $Z = 1$, as the right hand side of (5.8) produces the expectation value of $A$ within a single, given microstate $\psi$. The family of the operators $A_\mu^{\text{eff}}$ is parametrized by a unit vector $(\alpha_L, \alpha_R) \in \mathbb{C}^2 \cong \mathcal{M}$, or equivalently, by a perturbative vacuum state $\mu \in \mathcal{M}$. In particular every state of the form

$$\sum_{n=0}^{\infty} \left( \gamma_n^L \frac{(\hat{a}_L^+)^n}{\sqrt{n!}} + \gamma_n^R \frac{(\hat{a}_R^+)^n}{\sqrt{n!}} \right) \mu, \qquad \sum_{n=0}^{\infty} |\gamma_n^L|^2 = \sum_{n=0}^{\infty} |\gamma_n^R|^2 = 1 \qquad (5.14)$$

is characterized by the same vector $(\alpha_L, \alpha_R)$ with $\mu = \alpha_L \Psi_0^L + \alpha_R \Psi_0^R$.

It is a defining property of an effective theory that it should give an approximate description of the full theory within its region of validity. In the context of quantum theory this means that the correlation functions calculated within the effective theory should approximate those in the full theory. Given an operator $A$ in the full theory, the operators $A_\mu^{\text{eff}}$ defined in (5.13) satisfy this condition. This is a family of operators together with a 'fake' type of state-dependence as described in [42]. The parameter $\mu$ can be thought of as parametrizing degenerated perturbative vacua. Indeed, each $A_\mu^{\text{eff}}$ is a perfectly well-defined linear operator on $\mathcal{H}_L \otimes \mathcal{H}_R$, since the numbers $\alpha_L$ and $\alpha_R$ are fixed parameters from the point of view of the effective theory.

One can however revert this last construction. Starting from an operator $B$ as given in (5.9), one can try to construct a global operator $A$ such that (5.8) is satisfied. This is a weaker condition than what we have considered in section 5.1, since here we only demand the agreement between the expectation values of the operators (up to non-perturbative effects). It is obvious that a definition of $A$ must depend on the specific microstate of the full system. We can construct a family of operators $A_\mu$ in the full theory such that their action on a microstate $\psi = \alpha_L \psi_L + \alpha_R \psi_R$ with $\psi_L \in \mathcal{H}_L$, $\psi_R \in \mathcal{H}_R$ and $\|\psi\| = \|\psi_L\| = \|\psi_R\| = 1$ reads

$$A_\mu \psi = \left( \frac{A_{LL}}{|\alpha_L|^2} + \frac{A_{RR}}{|\alpha_R|^2} \right) \psi. \qquad (5.15)$$

With this definition (5.8) holds with $Z = 1$. The functions $A_\mu$ are now non-linear, i.e., 'state-dependent', as they implicitly depend on the parameters $\alpha_L$ and $\alpha_R$ of the state $\psi = \alpha_L \psi_L + \alpha_R \psi_R$ they act on. In other words they depend on the choice of the microstate within a given macrostate $\psi^{\text{eff}} = e^{i\theta} \psi_L \otimes \psi_R$.

This is an explicit construction of state-dependent operators in the theory in the spirit of [6–9]. Operators of the form (5.9) corresponding to naive perturbative observables become non-linear functions or, equivalently, state-dependent operators. In the context of [68] their matrix elements represent certain conditional probabilities. Furthermore, since the time evolution generically mixes various microstates, comparison of their matrix elements at different times seems problematic.

## 5.4 Global time evolution

It is common in the context of quantum field theory to calculate time-dependent field operator correlation functions such as $\langle 0|\phi(t_1)\phi(t_2)\ldots\phi(t_n)|0\rangle$. To consider such correlators in our toy-model, we must decide what are the corresponding state $|0\rangle$, the field operator $\phi$, and the Hamiltonian driving the time evolution.

From the point of view of the full theory, $|0\rangle$ is the vacuum state $\Psi_0^+$, $\phi = x$ is the field operator, and time evolution is governed by the full Hamiltonian $H$. In the context of black hole physics, however, one typically considers correlation functions in the perturbation theory corresponding to the viewpoint of a single, say right, asymptotic observer. In perturbation theory $|0\rangle$ corresponds to the right perturbative vacuum $\Psi_0^R$, and $\phi$ corresponds to the right perturbative field operator $\hat{y}_R = P_R y_R P_R = (\hat{a}_R + \hat{a}_R^+)/\sqrt{2}$.

What remains to be analyzed is the time evolution governed by the unitary operator $U(t) = e^{itH}$. Its effective version $U^{\text{eff}}(t)$ should be such that the evolution of the states and the operators in the full theory matches with that in the effective theory. Assume that $U(t)$ satisfies the condition (4.17), i.e., $U_{LR}(t) = o(\lambda^\infty)$ and $U_{RL}(t) = o(\lambda^\infty)$, and that the diagonal elements $U_{LL}(t)$ and $U_{RR}(t)$ remain unitary, at least up to non-perturbative effects. Since the Hamiltonian $H$ satisfies (4.17), this is the case for times $t$ sufficiently small for the tunneling effects to be insignificant. Now assign

$$U(t) = U_{LL}(t) + U_{RR}(t) \;\longmapsto\; U^{\text{eff}}(t) = U_{LL}(t) \otimes U_{RR}(t). \tag{5.16}$$

Since $U_{LL}$ and $U_{RR}$ are unitary, they preserve the norm of the states and hence for any $\psi \in \mathcal{H}$ we have

$$[U(t)\psi]^{\text{eff}} = U^{\text{eff}}(t)\psi^{\text{eff}}, \qquad\qquad (U^+)^{\text{eff}}(t) = (U^{\text{eff}})^+(t). \tag{5.17}$$

This means that the states evolve in the same way both in the full and in the effective theory. In particular

$$A_\mu^{\text{eff}}(t) = \left[U^+(t) A_\mu U(t)\right]^{\text{eff}} = (U^{\text{eff}})^+(t) A_\mu^{\text{eff}} U^{\text{eff}}(t). \tag{5.18}$$

with $A_\mu^{\text{eff}}$ defined as in the previous section. Therefore the condition (5.8) holds for time-dependent operators as well,

$$\langle\psi|A(t)|\psi\rangle = \frac{1}{Z}\langle\psi^{\text{eff}}|A_\mu^{\text{eff}}(t)|\psi^{\text{eff}}\rangle + o(\lambda^\infty) \tag{5.19}$$

With (5.16) the effective theory correctly describes the expectation values of time-dependent operators as long as the perturbation theory remains valid.

## 5.5 Time reversal and the 'wrong sign' commutation relations

In the previous section we have defined a notion of time evolution in the effective theory that is specified by the full theory, based in particular on the global time $t$ inherited from the full Hamiltonian $H$ in the Hilbert space $\mathcal{H}$. From the perspective of a single observer (say the right one), however, the common practice is to consider the doubled Hilbert space built out of the right Hilbert space $\mathcal{H}_R$. In this case, the effective theory as constructed by the right observer lives on $\mathcal{H}_R \otimes \mathcal{H}_R = (\Theta\mathcal{H}_L) \otimes \mathcal{H}_R$, with effective states defined as

$$\chi = \chi_L + \chi_R \;\longmapsto\; \tilde{\chi}^{\text{eff}} = \mathcal{N}\,(\Theta\chi_L) \otimes \chi_R \in \mathcal{H}_R \otimes \mathcal{H}_R \tag{5.20}$$

where $\mathcal{N}$ is the same normalization as in (5.5). This does not change the analysis of previous sections in any significant way. Simply, every operator acting on $\mathcal{H}_L$ must now be accompanied by a conjugation by $\Theta$. For example, equation (5.13) would read

$$\tilde{A}_\mu^{\text{eff}} = \Theta A_{LL}\Theta \otimes |\alpha_R|^2\mathbb{I} + |\alpha_L|^2\mathbb{I} \otimes A_{RR} \tag{5.21}$$

and so on. For the Hamiltonian $H$ of the full theory we have $H_{LL} = \Theta H_{RR} \Theta$ and hence we find its effective version

$$\tilde{H}_\mu^{\text{eff}} = H_{RR} \otimes |\alpha_R|^2 \mathbb{I} + |\alpha_L|^2 \mathbb{I} \otimes H_{RR}. \tag{5.22}$$

Consider now the time evolution operator $U(t) = e^{itH}$. Due to the additional factor of $i$ we have $\Theta U_{LL}(t) \Theta = U_{RR}(-t)$. Hence, in order to maintain (5.17), we have to include an additional conjugation by $\Theta$ in (5.16), i.e., to define

$$\tilde{U}^{\text{eff}}(t) = \Theta U_{LL}(t) \Theta \otimes U_{RR}(t) = U_{RR}(-t) \otimes U_{RR}(t). \tag{5.23}$$

From the point of view of the right observer, in the effective theory on $\mathcal{H}_R \otimes \mathcal{H}_R$ time directions are opposite in both component spaces $\mathcal{H}_R$. In particular the Hermitian generator $\tilde{h}$ of $\tilde{U}^{\text{eff}}(t)$ satisfying $\tilde{U}^{\text{eff}}(t) = e^{it\tilde{h}}$ is

$$\tilde{h} = -H_{RR} \otimes \mathbb{I} + \mathbb{I} \otimes H_{RR}. \tag{5.24}$$

This is a Hermitian generator of time translations in the effective theory on $\mathcal{H}_R \otimes \mathcal{H}_R$, but it is not bounded from below. Furthermore, it is not proportional to any $\tilde{H}_\mu^{\text{eff}}$, the family of effective operators corresponding to the full Hamiltonian. Finally, $\tilde{h}$ exhibits the famous 'wrong' commutation relations [2, 5–7, 49] with left creation-annihilation operators satisfying,

$$[\tilde{h}, \hat{a}_L \otimes \mathbb{I}] = +\hat{a}_L \otimes \mathbb{I} + O(\sqrt{\lambda}), \qquad [\tilde{h}, \hat{a}_L^+ \otimes \mathbb{I}] = -\hat{a}_L^+ \otimes \mathbb{I} + O(\sqrt{\lambda}). \tag{5.25}$$

This is in no contradiction with unitarity or any other property of a well-defined quantum theory. The effective theory is merely designed to mimic the full theory within the regime of its validity. It recognizes the full Hamiltonian as $\tilde{H}_\mu^{\text{eff}}$ defined in (5.22). The time evolution of the effective theory is however driven by $\tilde{h}$ in such a way that (5.19) holds. This is different from the effective theory on $\mathcal{H}_L \otimes \mathcal{H}_R$ (or $\mathcal{H}_L \otimes \mathcal{H}_L$ as perceived by the left observer). Different effective viewpoints realize observables in different ways, despite describing the same theory. However expectation values of corresponding operators in corresponding states are equal in all effective theories we have constructed here.

## 6 Dynamics

With all elements of our toy-model in place we now turn to a number of dynamical processes that are both tractable and have a clear dual interpretation in terms of black hole physics. Before we proceed, let us point out the obvious: our toy-model with Hamiltonian (1.2) is unitary. Since it also has a unique vacuum state, $\Psi_0^+$, it would seem that our model violates Hawking's theorem, as stated in [11]. However, the vacuum state in our effective theory is doubly degenerate, $\Psi_0^\pm$, and, since the theorem describes the semiclassical situation, its assumptions are not satisfied.

First we investigate tunneling through the potential barrier. Depending on the energy range, this can be viewed either as the decay of one of the perturbative vacua, the Hawking radiation process, or scattering of waves off of the black hole. Then we consider the evolution of a classical particle. We identify signatures of the chaotic behavior and scrambling.
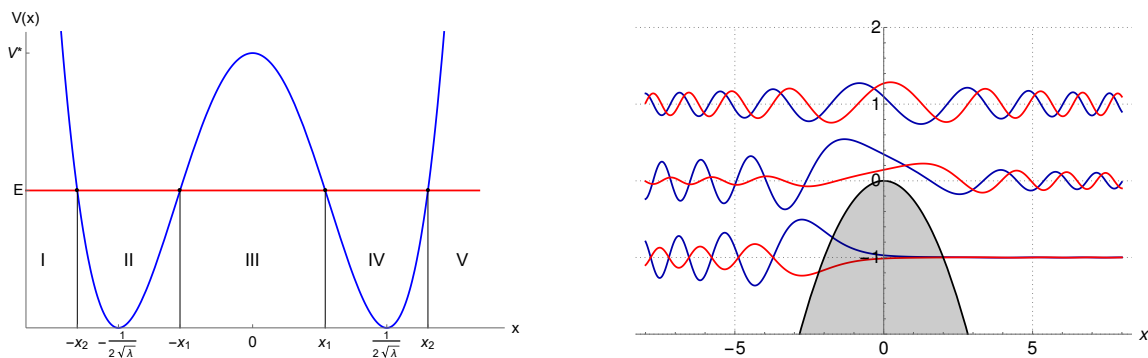
**Figure 10**. Left: the tunneling probability is computed in region III in the WKB approximation. Right: the shape of the real part (blue) and imaginaty part (red) of the (generalized) energy eigenfunctions representing waves incident from the left and scattered by the potential of the inverted harmonic oscillator. The functions correspond to $\nu^2 = 1/2$ and energies $\Delta = -1, 0, 1$.

## 6.1 Tunneling and Hawking radiation

The evolution operator $e^{itH}$ restricted to the space of perturbative vacua $\mathcal{M}$ slowly mixes the left and right states $\Psi_0^L$ and $\Psi_0^R$, as

$$e^{itH} \begin{pmatrix} \Psi_0^L \\ \Psi_0^R \end{pmatrix} = e^{itE_0} \begin{pmatrix} \cos\left(\frac{1}{2}t\Delta E_0\right) & -i\sin\left(\frac{1}{2}t\Delta E_0\right) \\ -i\sin\left(\frac{1}{2}t\Delta E_0\right) & \cos\left(\frac{1}{2}t\Delta E_0\right) \end{pmatrix} \begin{pmatrix} \Psi_0^L \\ \Psi_0^R \end{pmatrix} \tag{6.1}$$

where

$$E_0 = \frac{1}{2}(E_0^+ + E_0^-), \qquad \Delta E_0 = E_0^+ - E_0^-. \tag{6.2}$$

Assume that at $t = 0$ the system is in the state $\Psi_0^R$. As the system evolves, the wave function slowly leaks into the left minimum, evolving into 'more typical' black hole microstates. This is reminiscent of known instabilities of perturbative vacua [69, 70] evolving into black holes and envisioned e.g., in the formation of fuzzballs [51, 71].

On the other hand, as the wave functions builds up on the left, the right observer perceives the inflow of highly excited particles. Indeed, according to the results of section 3.3, when decomposed in terms of semiclassical modes $\varphi_n^R$, $\Psi_0^L$ is a highly-excited state. We can interpret these particles as the Hawking radiation. The same remains true for any low energy black hole state $\psi$. We can identify some features of the Hawking radiation in our model by tracing back these high occupancy modes through the potential barrier.

In the WKB approximation the tunneling probability $\Lambda$ for a particle of energy $E$ in a double well potential $V$ is given by

$$\Lambda = \int_{-x_1}^{x_1} \sqrt{2(V(x) - E)} dx \,, \tag{6.3}$$

where $\pm x_1$ are the turning points as illustrated in figure 10, and $E < V_* = V(0) = 1/(32\lambda)$. The tunneling time $\tau$ and the tunneling rate $\Gamma$ are then $\tau \sim \Gamma^{-1/2} \sim e^\Lambda$. In our toy-model the tunneling probability (6.3) within the WKB approximation can be expressed in terms

of complete elliptic integrals,

$$\Lambda = \frac{1}{6\lambda}\sqrt{1+\sqrt{\epsilon}}\left[\mathrm{E}\left(\sqrt{\frac{1-\sqrt{\epsilon}}{1+\sqrt{\epsilon}}}\right) - \sqrt{\epsilon}\,\mathrm{K}\left(\sqrt{\frac{1-\sqrt{\epsilon}}{1+\sqrt{\epsilon}}}\right)\right], \tag{6.4}$$

where $\epsilon = E/V_* = 32\lambda E$ with $0 < \epsilon < 1$, and K denotes the complete elliptic integral of the first kind.

For energies $E$ close to $V_*$ we have $\epsilon \sim 1$ and then (6.4) reduces to

$$\Lambda(\lambda,\delta) = \sqrt{2}\pi\delta + 3\sqrt{2}\pi\lambda\delta^2 + O(\delta^3), \quad \delta = V_* - E. \tag{6.5}$$

The first term in (6.5) corresponds to tunneling in an inverted harmonic oscillator with the potential $V_{\mathrm{iho}} = -\omega^2\frac{x^2}{4}$, for a particle of energy $E = \omega\delta$. The tunneling rate becomes

$$\Gamma_{\mathrm{iho}}(E = \omega\delta) = \exp\left(-2\sqrt{2}\pi\delta - 6\sqrt{2}\pi\lambda\delta^2 + O(\delta^3)\right). \tag{6.6}$$

This resonates with [72] where it was shown that Hawking radiation can be viewed as a tunneling effect between regions near both sides of the horizon. Furthermore, specific deviations from an exact thermal spectrum were found in such a process, in a range of different black hole backgrounds. A comparison with [72] shows that the Hawking radiation in our toy-model corresponds roughly to a single pair of modes of a field of frequency $\omega \sim \sqrt{\lambda}\delta$ in a black hole state of mass $M \sim 1/\sqrt{\lambda} = N$.

Building on the analysis of [72] it has been shown in [73] (see also [74]) that the scattering matrix of an inverted harmonic oscillator also appears in the investigation of scattering of shock waves from in the black hole background. Indeed, this is precisely the conclusion from our toy-model. Around $x = 0$ energy eigenfunctions are approximated by wave functions of the inverted harmonic oscillator with energy $E$. The approximate Hamiltonian reads

$$H_{\mathrm{iho}} = \frac{p^2}{2} - \frac{1}{2}\nu^2 x^2 \tag{6.7}$$

and in our specific toy-model $\nu^2 = \omega^2/2 = 1/2$. To analyze quantum scattering, we consider oscillating wave functions $\psi$ satisfying the Schrödinger equation with the Hamiltonian (6.7),

$$\frac{\partial^2}{\partial\xi^2}\psi(\xi) + \left(\frac{1}{4}\xi^2 + \frac{\Delta}{\nu}\right)\psi(\xi) = 0 \tag{6.8}$$

where $\Delta = E - V_*$ and $\xi = \sqrt{2\nu}x$. A pair of solutions exists for each energy value $\Delta$, which can be expressed in terms of hypergeometric functions, see [75]. With an appropriate normalization the two solutions represent waves incident from the left and from the right, as shown in figure 10. By expanding the wave functions in the asymptotic regions transmission and reflection coefficients can be found,

$$T = \frac{e^{\pi\Delta/2\nu}}{\sqrt{2\pi}}\Gamma\left(\frac{1}{2} - \frac{i\Delta}{\nu}\right), \qquad R = -i\frac{e^{-\pi\Delta/2\nu}}{\sqrt{2\pi}}\Gamma\left(\frac{1}{2} - \frac{i\Delta}{\nu}\right), \tag{6.9}$$

which satisfy $|T|^2 + |R|^2 = 1$. Loosely speaking these coefficients describe the behavior in our toy-model of excitations near a black hole horizon. It is remarkable that the same form

of the transmission and reflection coefficients (6.9) was recovered there from the analysis of the dynamics of shock waves traveling in the black hole background in [73]. To leading order the result follows simply from the repulsive nature of the inverted harmonic oscillator and depends on a single parameter $\nu$ in (6.7).

## 6.2 Classical evolution and chaos

Consider now a classical particle of energy $E \geq 0$. Such a particle will stay on a closed orbit. If $E < V_* = V(0)$, then the orbit remains 'outside the black hole', i.e., it does not cross the maximum of the potential at $x = 0$. The period of the classical motion is then

$$T_{\text{trapped}} = \frac{8}{\sqrt{1 - \sqrt{\epsilon}} + \sqrt{1 + \sqrt{\epsilon}}} \, \text{K}\left(\sqrt{\frac{1 - \sqrt{1 - \epsilon}}{1 + \sqrt{1 - \epsilon}}}\right), \qquad (6.10)$$

When expanded around $\epsilon = 0$ one finds as expected, $T_{\text{trapped}} = 2\pi + O(\epsilon)$. On the other hand the period diverges logarithmically when $E \to V_*$,

$$T_{\text{trapped}} = \sqrt{2}\log\left(\frac{2}{\lambda\delta}\right) + O(\epsilon), \qquad \delta = V_* - E. \qquad (6.11)$$

This is a sign of a critical behavior. When analyzed from the point of view of the time-dependent position $x(t)$ of the particle, it is a manifestation of chaos. To see it explicitly, assume that $\delta$ is small enough for the particle to be located close to the tip of the potential for a long time. Hence we can neglect interaction terms and consider the potential of the inverted harmonic oscillator only. The solution to the equations of motion is then simply $x(t) = x_0 \cosh(\nu t) + v_0/\nu \sinh(\nu t)$, where $\nu$ is the 'frequency' of the inverted harmonic oscillator as defined in (6.7). Parameters $x_0$ and $v_0$ are initial position and velocity at $t = 0$ and under their variation $\delta x(t) \sim e^{\nu t}(\delta x_0 + \delta v_0/\nu)$ for large times. This is by definition chaotic behavior with the Lapunov exponent $\nu = \omega/\sqrt{2} = 1/\sqrt{2}$.

As in the previous section the chaotic behavior is driven by the inverted harmonic oscillator and parametrized by a single parameter $\nu$. The same conclusion arises from the analysis of shock waves in a black hole background in [23, 24, 26, 27]. Perhaps unsurprisingly, our analysis shows a connection between the classical chaotic behavior and quantum Hawking radiation from the point of view of the tunneling process as analyzed in the previous section. Both processes are described by the same parameter $\nu$ characterizing the 'frequency' of the inverted harmonic oscillator (6.7).

One can also analyze the quantum evolution of the system numerically and show how the classical evolution becomes inaccurate when $E$ approaches $V_*$. To do so, we consider the evolution of suitable initially coherent states. The resulting evolution is illustrated in figure 11. As we can see, at low energies the original coherent state (blue curve on the left) bounces back and retains its Gaussian shape after a full period (green). A small change in the shape is caused by interactions of order $\sqrt{\lambda}$. The wave function leaks slightly into the left minimum, creating barely visible ripples in the density $|\psi|^2$. On the other hand a coherent state of energy only slightly lower than $V_*$ rapidly turns into a highly quantum oscillating wave. The original classical particle (blue curve on the right) starts
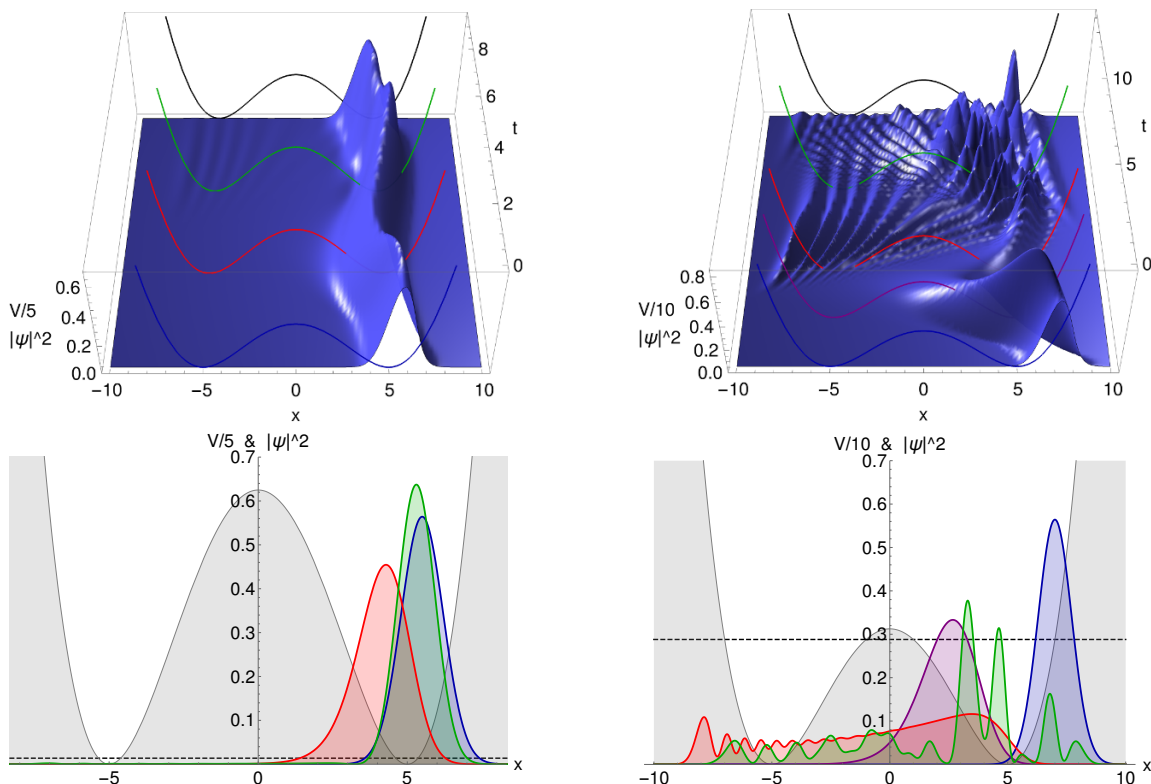
**Figure 11**. Evolution of the density $|\psi|^2$ of an initially classical system represented as a coherent state. Plots on the left show the evolution of a low energy wave packet on a classical path starting at $x = 5.5$ (classical energy $E = 0.14$). Plots on the right present the evolution of a wave packet of energy just below the energy $V_*$ of the tip of the potential ($E = 2.88$, $V_* = 3.125$), with the initial position at $x = 7$. In both cases $\lambda = 1/100$. The outlines of the potential in the 3D plots are placed at times: $t = 0$ (blue), quarter of the classical period (purple, only on the right), half of the period (red), and the full classical period (green). The 2D plots show the shapes of the wavefunctions at the corresponding times. The dashed lines indicate the classical energy of the wave packet.

moving towards the maximum of the potential. At the quarter of the period the peak widens (purple curve). By the time it bounces back at half the period it already breaks into a highly oscillating quantum wave (red curve). After the full classical period passes, it becomes completely scrambled and spread out over the entire domain (green curve). It cannot be viewed as a classical, localized state any more.

# 7 Summary and conclusions

Motivated by the holographic description of certain classes of black holes in AdS we have put forward a new and solvable dual toy-model of black holes in terms of a quantum mechanical particle in a double well potential. The effective low energy description involves the tensor product of two decoupled harmonic oscillators representing the degrees of freedom on both sides of the horizon, or in both asymptotic regions. At this level our model captures many

of the usual paradoxes of semiclassical black hole physics expressed here in quantum field theoretical language without explicit reference to geometry.

The effective low energy description of the system as a pair of decoupled oscillators is altered drastically by non-perturbative interactions. We have carefully explored how states and operators in the effective theory emerge from and relate to corresponding quantities in the full model. This elucidates how holographic black hole models involving non-perturbative interactions between two decoupled low energy theories resolve some of the paradoxes of semiclassical black hole physics. Our key findings are the following:

- *Firewalls:* black hole states in our model are represented by wave functions with significant support in both minima.[19] At first sight the low energy theory predicts a firewall. This is because the expectation value in black hole states of the naive number operator (3.21) is large. This number operator has the same form in the decoupling limit and hence it is usually assumed that it should represent the number operator for a small coupling $\lambda$ as well. However, by carefully identifying and disentangling the perturbative degrees of freedom we have shown that this is incorrect. We have constructed a different number operator (4.9) that is well-defined in the full model and perturbative in a precise sense. In the decoupling limit this operator correctly reproduces (3.21). We found this does not predict a firewall. Hence our model satisfies all four postulates of [2].

- *State-dependence:* our toy-model describes both sides of the horizon in terms of a single well-defined, local, unitary quantum theory. Hence it does not require any 'state-dependent' operators to describe behind the horizon physics. However we have shown that a clear notion of 'state-dependence' emerges when one relates perturbative operators in the full theory to observables in the effective low energy model on the tensor product. Mathematically this is because the linear structure of the direct sum is very different from that of the tensor product. Physically, the state-dependence accounts for the dependence of perturbative operators in the full theory on the black hole microstate, i.e., on the shape of the portion of the wave function behind the barrier that is inaccessible to a given asymptotic observer.

- *Vacuum structure:* our model circumvents Mathur's no-go theorem [10], based on Hawking's original calculation, that states that the information paradox cannot be resolved by exponentially small corrections to correlation functions. This is because this theorem assumes there is a unique vacuum. In our toy-model the perturbative vacuum is degenerate, and the degeneracy is only broken by non-perturbative effects leading to a unique ground state $\Psi_0^+$ in the full theory. In our model the vacuum of an infalling observer is not represented by a semiclassical vacuum, but rather by a superposition of two semiclassical vacua in line with e.g., [76, 77]. Phrased differently, one could say our model takes seriously the doubled copy of the system and in fact realizes an ensemble of states in the full interacting theory that are to some extent similar to the thermofield double state.

---

[19]This resonates with the model in [76] which was also argued to remove the firewall.

- *Time evolution:* 'Wrong sign' commutation relations naturally emerge in the effective description of the system from the standpoint of an observer in one of the asymptotic regions. This is not in contradiction with unitarity of the full model since the Hamiltonian of the full theory is represented by a different operator (5.22) than the generator (5.24) of time translations in the effective theory around one of the perturbative vacua.

A major advantage of our toy-model is that it is solvable. This has enabled us to analyze the role of non-perturbative interactions in a number of interesting dynamical processes. The appearance of an inverted harmonic oscillator potential separating both wells as a toy-model for non-perturbative interactions leads to features, such as chaotic behavior, which have a natural analog or 'dual interpretation' in black hole backgrounds where similar behavior was obtained e.g., in the analysis of shock waves. Moreover, the breakdown of classical evolution of initially coherent states in our model shows that the evolution of an infalling object becomes highly quantum from the viewpoint of an external observer, in line with the principle of black hole complementarity [52]. It would be interesting to investigate the role of non-perturbative effects and the emergence of the effective theory in more complex models. These could include matrix and tensor models [78–83], where many features discussed in this paper emerge. Another direction includes the CFT analysis in the context of holography [30, 84, 85], where non-perturbative effects also become essential in the understanding of unitarity and locality.

## Acknowledgments

## References

[1] S.W. Hawking, *Breakdown of predictability in gravitational collapse*, *Phys. Rev.* **D 14** (1976) 2460 [INSPIRE].

[2] A. Almheiri, D. Marolf, J. Polchinski and J. Sully, *Black holes: complementarity or firewalls?*, *JHEP* **02** (2013) 062 [arXiv:1207.3123] [INSPIRE].

[3] A. Almheiri, D. Marolf, J. Polchinski, D. Stanford and J. Sully, *An apologia for firewalls*, *JHEP* **09** (2013) 018 [arXiv:1304.6483] [INSPIRE].

[4] R. Bousso, *Complementarity is not enough*, *Phys. Rev.* **D 87** (2013) 124023 [arXiv:1207.5192] [INSPIRE].

[5] D. Marolf and J. Polchinski, *Gauge/gravity duality and the black hole interior*, *Phys. Rev. Lett.* **111** (2013) 171301 [`arXiv:1307.4706`] [INSPIRE].

[6] K. Papadodimas and S. Raju, *State-dependent bulk-boundary maps and black hole complementarity*, *Phys. Rev.* **D 89** (2014) 086010 [`arXiv:1310.6335`] [INSPIRE].

[7] K. Papadodimas and S. Raju, *Black hole interior in the holographic correspondence and the information paradox*, *Phys. Rev. Lett.* **112** (2014) 051301 [`arXiv:1310.6334`] [INSPIRE].

[8] K. Papadodimas and S. Raju, *Remarks on the necessity and implications of state-dependence in the black hole interior*, *Phys. Rev.* **D 93** (2016) 084049 [`arXiv:1503.08825`] [INSPIRE].

[9] K. Papadodimas and S. Raju, *Local operators in the eternal black hole*, *Phys. Rev. Lett.* **115** (2015) 211601 [`arXiv:1502.06692`] [INSPIRE].

[10] S.D. Mathur, *What exactly is the information paradox?*, *Lect. Notes Phys.* **769** (2009) 3 [`arXiv:0803.2030`] [INSPIRE].

[11] S.D. Mathur, *The information paradox: a pedagogical introduction*, *Class. Quant. Grav.* **26** (2009) 224001 [`arXiv:0909.1038`] [INSPIRE].

[12] S.W. Hawking, M.J. Perry and A. Strominger, *Soft hair on black holes*, *Phys. Rev. Lett.* **116** (2016) 231301 [`arXiv:1601.00921`] [INSPIRE].

[13] T. Hertog and K. Maeda, *Black holes with scalar hair and asymptotics in $N = 8$ supergravity*, *JHEP* **07** (2004) 051 [`hep-th/0404261`] [INSPIRE].

[14] T. Hertog and G.T. Horowitz, *Designer gravity and field theory effective potentials*, *Phys. Rev. Lett.* **94** (2005) 221301 [`hep-th/0412169`] [INSPIRE].

[15] T. Hertog and G.T. Horowitz, *Holographic description of AdS cosmologies*, *JHEP* **04** (2005) 005 [`hep-th/0503071`] [INSPIRE].

[16] T. Hertog, *Towards a novel no-hair theorem for black holes*, *Phys. Rev.* **D 74** (2006) 084008 [`gr-qc/0608075`] [INSPIRE].

[17] S.B. Giddings, *Nonviolent nonlocality*, *Phys. Rev.* **D 88** (2013) 064023 [`arXiv:1211.7070`] [INSPIRE].

[18] S.B. Giddings, *Nonviolent unitarization: basic postulates to soft quantum structure of black holes*, *JHEP* **12** (2017) 047 [`arXiv:1701.08765`] [INSPIRE].

[19] B.D. Chowdhury, *Black holes versus firewalls and thermo-field dynamics*, *Int. J. Mod. Phys.* **D 22** (2013) 1342011 [`arXiv:1305.6343`] [INSPIRE].

[20] J. Maldacena and L. Susskind, *Cool horizons for entangled black holes*, *Fortsch. Phys.* **61** (2013) 781 [`arXiv:1306.0533`] [INSPIRE].

[21] L. Susskind, *Entanglement is not enough*, *Fortsch. Phys.* **64** (2016) 49 [`arXiv:1411.0690`] [INSPIRE].

[22] V. Balasubramanian, M. Berkooz, S.F. Ross and J. Simon, *Black holes, entanglement and random matrices*, *Class. Quant. Grav.* **31** (2014) 185009 [`arXiv:1404.6198`] [INSPIRE].

[23] S.H. Shenker and D. Stanford, *Black holes and the butterfly effect*, *JHEP* **03** (2014) 067 [`arXiv:1306.0622`] [INSPIRE].

[24] S.H. Shenker and D. Stanford, *Multiple shocks*, *JHEP* **12** (2014) 046 [`arXiv:1312.3296`] [INSPIRE].

[25] P. Gao, D.L. Jafferis and A. Wall, *Traversable wormholes via a double trace deformation*, *JHEP* **12** (2017) 151 [`arXiv:1608.05687`] [INSPIRE].

[26] J. Maldacena, D. Stanford and Z. Yang, *Diving into traversable wormholes*, *Fortsch. Phys.* **65** (2017) 1700034 [`arXiv:1704.05333`] [ɪɴSPIRE].

[27] R. van Breukelen and K. Papadodimas, *Quantum teleportation through time-shifted AdS wormholes*, `arXiv:1708.09370` [ɪɴSPIRE].

[28] A. Hamilton, D.N. Kabat, G. Lifschytz and D.A. Lowe, *Local bulk operators in AdS/CFT: a boundary view of horizons and locality*, *Phys. Rev.* **D 73** (2006) 086003 [`hep-th/0506118`] [ɪɴSPIRE].

[29] A. Hamilton, D.N. Kabat, G. Lifschytz and D.A. Lowe, *Holographic representation of local bulk operators*, *Phys. Rev.* **D 74** (2006) 066009 [`hep-th/0606141`] [ɪɴSPIRE].

[30] A.L. Fitzpatrick, J. Kaplan, D. Li and J. Wang, *On information loss in AdS$_3$/CFT$_2$*, *JHEP* **05** (2016) 109 [`arXiv:1603.08925`] [ɪɴSPIRE].

[31] C. Martinez, R. Troncoso and J. Zanelli, *Exact black hole solution with a minimally coupled scalar field*, *Phys. Rev.* **D 70** (2004) 084035 [`hep-th/0406111`] [ɪɴSPIRE].

[32] A. Aceña, A. Anabalón, D. Astefanesei and R. Mann, *Hairy planar black holes in higher dimensions*, *JHEP* **01** (2014) 153 [`arXiv:1311.6065`] [ɪɴSPIRE].

[33] A. Anabalon, *Exact black holes and universality in the backreaction of non-linear σ-models with a potential in (A)dS$_4$*, *JHEP* **06** (2012) 127 [`arXiv:1204.2720`] [ɪɴSPIRE].

[34] F. Faedo, D. Klemm and M. Nozawa, *Hairy black holes in N = 2 gauged supergravity*, in *Proceedings,* 14$^{th}$ *Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Astrophysics, and Relativistic Field Theories (MG14)*, Rome, Italy, 12–18 July 2015, vol. 4, World Scientific, Singapore, (2017), pg. 4204.

[35] A. Anabalón, D. Astefanesei, A. Gallerati and M. Trigiante, *Hairy black holes and duality in an extended supergravity model*, *JHEP* **04** (2018) 058 [`arXiv:1712.06971`] [ɪɴSPIRE].

[36] M. Henneaux, C. Martinez, R. Troncoso and J. Zanelli, *Asymptotically anti-de Sitter spacetimes and scalar fields with a logarithmic branch*, *Phys. Rev.* **D 70** (2004) 044034 [`hep-th/0404236`] [ɪɴSPIRE].

[37] T. Hertog and K. Maeda, *Stability and thermodynamics of AdS black holes with scalar hair*, *Phys. Rev.* **D 71** (2005) 024001 [`hep-th/0409314`] [ɪɴSPIRE].

[38] I. Papadimitriou, *Multi-trace deformations in AdS/CFT: exploring the vacuum structure of the deformed CFT*, *JHEP* **05** (2007) 075 [`hep-th/0703152`] [ɪɴSPIRE].

[39] E. Witten, *Multitrace operators, boundary conditions and AdS/CFT correspondence*, `hep-th/0112258` [ɪɴSPIRE].

[40] M. Berkooz, A. Sever and A. Shomer, *'Double trace' deformations, boundary conditions and space-time singularities*, *JHEP* **05** (2002) 034 [`hep-th/0112264`] [ɪɴSPIRE].

[41] T. Hertog and S. Hollands, *Stability in designer gravity*, *Class. Quant. Grav.* **22** (2005) 5323 [`hep-th/0508181`] [ɪɴSPIRE].

[42] D. Harlow, *Aspects of the Papadodimas-Raju proposal for the black hole interior*, *JHEP* **11** (2014) 055 [`arXiv:1405.1995`] [ɪɴSPIRE].

[43] D. Harlow, *Jerusalem lectures on black holes and quantum information*, *Rev. Mod. Phys.* **88** (2016) 015002 [`arXiv:1409.1231`] [ɪɴSPIRE].

[44] M. Reed and B. Simon, *Analysis of operators*, *Methods of modern mathematical physics* **4**, Academic Press, U.S.A., (1978).

[45] J.M. Maldacena, *Eternal black holes in anti-de Sitter*, *JHEP* **04** (2003) 021 [hep-th/0106112] [INSPIRE].

[46] B. Czech, J.L. Karczmarek, F. Nogueira and M. Van Raamsdonk, *Rindler quantum gravity*, *Class. Quant. Grav.* **29** (2012) 235025 [arXiv:1206.1323] [INSPIRE].

[47] M. Van Raamsdonk, *Evaporating firewalls*, *JHEP* **11** (2014) 038 [arXiv:1307.1796] [INSPIRE].

[48] J. Zinn-Justin and U.D. Jentschura, *Multi-instantons and exact results I: conjectures, WKB expansions and instanton interactions*, *Annals Phys.* **313** (2004) 197 [quant-ph/0501136] [INSPIRE].

[49] K. Papadodimas and S. Raju, *An infalling observer in AdS/CFT*, *JHEP* **10** (2013) 212 [arXiv:1211.6767] [INSPIRE].

[50] J. Hartle and T. Hertog, *Quantum transitions between classical histories*, *Phys. Rev.* **D 92** (2015) 063509 [arXiv:1502.06770] [INSPIRE].

[51] T. Hertog and J. Hartle, *Observational implications of fuzzball formation*, arXiv:1704.02123 [INSPIRE].

[52] L. Susskind, L. Thorlacius and J. Uglum, *The stretched horizon and black hole complementarity*, *Phys. Rev.* **D 48** (1993) 3743 [hep-th/9306069] [INSPIRE].

[53] Y. Sekino and L. Susskind, *Fast scramblers*, *JHEP* **10** (2008) 065 [arXiv:0808.2096] [INSPIRE].

[54] S. Ghosh and S. Raju, *Breakdown of string perturbation theory for many external particles*, *Phys. Rev. Lett.* **118** (2017) 131602 [arXiv:1611.08003] [INSPIRE].

[55] S. Ghosh and S. Raju, *Loss of locality in gravitational correlators with a large number of insertions*, *Phys. Rev.* **D 96** (2017) 066033 [arXiv:1706.07424] [INSPIRE].

[56] D.L. Jafferis, *Bulk reconstruction and the Hartle-Hawking wavefunction*, arXiv:1703.01519 [INSPIRE].

[57] D. Kabat and G. Lifschytz, *Finite N and the failure of bulk locality: black holes in AdS/CFT*, *JHEP* **09** (2014) 077 [arXiv:1405.6394] [INSPIRE].

[58] S. Raju, *Smooth causal patches for AdS black holes*, *Phys. Rev.* **D 95** (2017) 126002 [arXiv:1604.03095] [INSPIRE].

[59] A. Hamilton, D.N. Kabat, G. Lifschytz and D.A. Lowe, *Local bulk operators in AdS/CFT: a holographic description of the black hole interior*, *Phys. Rev.* **D 75** (2007) 106001 [*Erratum ibid.* **D 75** (2007) 129902] [hep-th/0612053] [INSPIRE].

[60] D. Kabat and G. Lifschytz, *Does boundary quantum mechanics imply quantum mechanics in the bulk?*, *JHEP* **03** (2018) 151 [arXiv:1801.08101] [INSPIRE].

[61] N.G. Sanchez and B.F. Whiting, *Quantum field theory and the antipodal identification of black holes*, *Nucl. Phys.* **B 283** (1987) 605 [INSPIRE].

[62] A. Chamblin and J. Michelson, *Alpha-vacua, black holes and AdS/CFT*, *Class. Quant. Grav.* **24** (2007) 1569 [hep-th/0610133] [INSPIRE].

[63] G. 't Hooft, *Black hole unitarity and antipodal entanglement*, *Found. Phys.* **46** (2016) 1185 [arXiv:1601.03447] [INSPIRE].

[64] G. 't Hooft, *The firewall transformation for black holes and some of its implications*, *Found. Phys.* **47** (2017) 1503 [arXiv:1612.08640] [INSPIRE].

[65] M. Srednicki, *Thermal fluctuations in quantized chaotic systems*, *J. Phys.* **A 29** (1996) L75 [chao-dyn/9511001] [INSPIRE].

[66] N. Lashkari, A. Dymarsky and H. Liu, *Eigenstate thermalization hypothesis in conformal field theory*, *J. Stat. Mech.* **1803** (2018) 033101 [arXiv:1610.00302] [INSPIRE].

[67] N. Lashkari, A. Dymarsky and H. Liu, *Universality of quantum information in chaotic CFTs*, *JHEP* **03** (2018) 070 [arXiv:1710.10458] [INSPIRE].

[68] D. Marolf and J. Polchinski, *Violations of the Born rule in cool state-dependent horizons*, *JHEP* **01** (2016) 008 [arXiv:1506.01337] [INSPIRE].

[69] H. Ooguri and C. Vafa, *Non-supersymmetric AdS and the swampland*, *Adv. Theor. Math. Phys.* **21** (2017) 1787 [arXiv:1610.01533] [INSPIRE].

[70] B. Freivogel and M. Kleban, *Vacua morghulis*, arXiv:1610.04564 [INSPIRE].

[71] S.D. Mathur, *Tunneling into fuzzball states*, *Gen. Rel. Grav.* **42** (2010) 113 [arXiv:0805.3716] [INSPIRE].

[72] M.K. Parikh and F. Wilczek, *Hawking radiation as tunneling*, *Phys. Rev. Lett.* **85** (2000) 5042 [hep-th/9907001] [INSPIRE].

[73] P. Betzios, N. Gaddam and O. Papadoulaki, *The black hole S-matrix from quantum mechanics*, *JHEP* **11** (2016) 131 [arXiv:1607.07885] [INSPIRE].

[74] G. Horowitz, A. Lawrence and E. Silverstein, *Insightful D-branes*, *JHEP* **07** (2009) 057 [arXiv:0904.3922] [INSPIRE].

[75] D. Bermudez and D.J.F. C, *Factorization method and new potentials from the inverted oscillator*, *Annals Phys.* **333** (2013) 290 [arXiv:1206.4519] [INSPIRE].

[76] T.J. Hollowood, *Schrödinger's cat and the firewall*, *Int. J. Mod. Phys.* **D 23** (2014) 1441004 [arXiv:1403.5947] [INSPIRE].

[77] N. Bao, S.M. Carroll and A. Singh, *The Hilbert space of quantum gravity is locally finite-dimensional*, *Int. J. Mod. Phys.* **D 26** (2017) 1743013 [arXiv:1704.00066] [INSPIRE].

[78] T. Banks, W. Fischler, S.H. Shenker and L. Susskind, *M theory as a matrix model: a conjecture*, *Phys. Rev.* **D 55** (1997) 5112 [hep-th/9610043] [INSPIRE].

[79] N. Iizuka, T. Okuda and J. Polchinski, *Matrix models for the black hole information paradox*, *JHEP* **02** (2010) 073 [arXiv:0808.0530] [INSPIRE].

[80] N. Iizuka and J. Polchinski, *A matrix model for black hole thermalization*, *JHEP* **10** (2008) 028 [arXiv:0801.3657] [INSPIRE].

[81] F. Ferrari, *Black hole horizons and Bose-Einstein condensation*, arXiv:1601.08120 [INSPIRE].

[82] J. Maldacena and D. Stanford, *Remarks on the Sachdev-Ye-Kitaev model*, *Phys. Rev.* **D 94** (2016) 106002 [arXiv:1604.07818] [INSPIRE].

[83] J. Polchinski and V. Rosenhaus, *The spectrum in the Sachdev-Ye-Kitaev model*, *JHEP* **04** (2016) 001 [arXiv:1601.06768] [INSPIRE].

[84] A.L. Fitzpatrick and J. Kaplan, *Conformal blocks beyond the semi-classical limit*, *JHEP* **05** (2016) 075 [arXiv:1512.03052] [INSPIRE].

[85] H. Chen, C. Hussong, J. Kaplan and D. Li, *A numerical approach to Virasoro blocks and the information paradox*, *JHEP* **09** (2017) 102 [arXiv:1703.09727] [INSPIRE].