

The second law of black hole mechanics in effective field theory

Harvey Reall

DAMTP, Cambridge University

Stefan Hollands, Aron Kovacs and HSR, 2205.15341

Iain Davies and HSR, 2212.09777, 2312.07659

Black hole mechanics

The laws of black hole mechanics hold for General Relativity coupled to matter obeying suitable energy conditions.

First law: $dM = TdS + \Omega_H dJ$ where $T = \kappa/(2\pi)$ and $S = A/(4G)$

Second law: S is non-decreasing function of time.

Wald (1993) showed that a first law can be proved in any theory of gravity arising from a diffeomorphism invariant Lagrangian. (Typically contains terms with more than two derivatives.)

This gives a definition of black hole entropy that applies to any *stationary* (time-independent) black hole: the Wald entropy.

Can the second law also be extended to this class of theories?

Does there exist a definition of S , depending only on the geometry of a horizon cross-section, that satisfies a second law, and reduces to the Wald entropy in equilibrium?

Second law beyond GR

Iyer and Wald (1994) proposed a definition of entropy of a dynamical black hole in any theory arising from a diffeomorphism invariant Lagrangian.

Example: Einstein-Gauss-Bonnet

$$16\pi G\mathcal{L} = -2\Lambda + R + kL_{GB} \quad L_{GB} \propto R_{abcd}R^{abcd} - 4R_{ab}R^{ab} + R^2$$

Iyer-Wald entropy of horizon cross-section C is $\frac{1}{4G} \int_C (1 + kR[C])$

Various nice properties (free of ambiguities, agrees with Wald entropy for a stationary black hole, satisfies first law) but: unclear whether it satisfies a second law, not invariant under field redefinitions

Jacobson, Kang and Myers (1995) proved a second law for $f(R)$ theories. Their entropy differs from the Iyer-Wald entropy.

Perturbative approach

Consider perturbations of a stationary black hole:

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \epsilon h_{\mu\nu}^{(1)} + \epsilon^2 h_{\mu\nu}^{(2)} + \dots$$

This could describe a black hole “relaxing to equilibrium”.

Wall (2015) presented an algorithm for defining an entropy that satisfies a second law to *linear order* in perturbations around a stationary black hole, for any diffeo-invariant theory of vacuum gravity.

Wall entropy

Wall's entropy is the Iyer-Wald entropy plus new "Wall terms".

The simplest Wall terms are of the (schematic) form $K\bar{K}$ where K_{AB} and \bar{K}_{AB} are matrices describing the expansion+shear of the outgoing and ingoing null geodesics orthogonal to a horizon cut.

To linear order, the second law says $\delta\dot{S} = 0$ (if $\delta\dot{S} > 0$ then multiply perturbation by -1 to reach $\delta\dot{S} < 0$)

To see an entropy *increase* we need to go beyond linear perturbations of a stationary black hole.

(Wall also considers coupling gravity to "matter" described by $T_{\mu\nu}$ satisfying null energy condition, this gives $\dot{S} \geq 0$.)

Our approach

The main idea: work within the framework of *effective field theory* (EFT). This provides the most convincing motivation for considering theories with higher derivatives.

Starting from a “UV” theory, “integrate out” massive degrees of freedom to obtain an EFT for the light degrees of freedom, valid at energy below a scale associated with UV physics.

An EFT Lagrangian is ordered by terms with increasing numbers of derivatives. If only light field is the metric then the Lagrangian is

$$16\pi G\mathcal{L} = -2\Lambda + R + a_1\ell^2 R^2 + a_2\ell^2 R_{ab}R^{ab} + a_3\ell^2 L_{GB} + \dots$$

Assume coefficients of k -derivative terms ($k > 2$) are proportional to ℓ^{k-2} where ℓ is the “UV length scale” e.g. string scale or Planck scale.

Einstein-Maxwell EFT: integrate out electron from QED in curved spacetime: ℓ is Compton wavelength of electron.

Validity of EFT

Many solutions of higher derivative theories exhibit pathological behaviour. (cf radiation reaction problem)

Example: for $d = 4$ consider the EFT of vacuum gravity truncated to retain the leading corrections to GR:

$$16\pi G\mathcal{L} = R + a_1\ell^2 R^2 + a_2\ell^2 R_{ab}R^{ab}$$

Linearize around flat spacetime: in addition to the usual graviton, this theory describes heavy fields with $(\text{mass})^2 \propto \ell^2$. Hence there exist solutions that exhibit oscillations or exponential growth on the UV time scale ℓ . To describe phenomena with time variation on this scale requires a UV theory, i.e., these solutions *lie outside the regime of validity of the EFT*.

EFT eqs of motion admit solutions exhibiting length/time variation on the UV scale ℓ . These solutions are *unphysical* so there is no reason why they should satisfy a second law.

Validity of EFT criterion

Consider an EFT with UV length scale ℓ

We should consider only those black hole solutions lying within the regime of validity of EFT: Let L be “smallest length/time scale over which solution varies”. We require that $\ell/L \ll 1$.

More precisely: consider a 1-parameter family of solutions, with parameter L , for which there exists coordinates such that $\partial^k g_{\mu\nu} = O(L^{-k})$. Terms with more derivatives are less important

We do *not* assume that solution is an expansion in ℓ/L because this would be too restrictive e.g. it would exclude quasinormal modes $\propto e^{-i\omega t}$ with $\omega = L^{-1} + \ell^2 L^{-3} + \dots$

Second law in EFT

We never know the Lagrangian of an EFT exactly, but only the terms with up to N derivatives. The EFT equation of motion is

$$E_{\mu\nu} = \mathcal{O}(\ell^N)$$

where $E_{\mu\nu} = \Lambda g_{\mu\nu} + R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \dots$ includes the known terms with up to N derivatives. The RHS is really $\mathcal{O}(\ell^N/L^{N+2})$ but I will suppress L .

We should not aim to prove a second law that holds exactly, instead *the second law should hold only to the same order of accuracy as the theory itself*, i.e., modulo terms of order ℓ^N .

From $E_{\mu\nu}$ we want to construct an entropy S that satisfies $\dot{S} \geq -\mathcal{O}(\ell^N)$. By increasing N we increase the accuracy to which the EFT is known and the accuracy to which the second law is satisfied.

Our assumptions

1. The event horizon is *smooth*: expect this to be valid *after* gravitational collapse or black hole merger. We use Gaussian null coordinates (GNCs) adapted to the horizon.
2. The black hole settles down to equilibrium at late time. Quantities that vanish on a stationary black hole horizon decay at large affine time in our GNCs.
3. The validity of EFT assumption holds in our GNCs.

Our results

We show how to define an entropy S that satisfies a *non-perturbative* second law, i.e., $\dot{S} \geq -\mathcal{O}(\ell^N)$ holds for arbitrary (smooth) perturbations of a stationary black hole.

In equilibrium our entropy reduces to the Wald entropy. To linear order in perturbations around a stationary BH it agrees with Wall's entropy (hence satisfies first law).

Our entropy is constructed by adding extra terms to Wall's entropy. For $E_{\mu\nu}$ containing up to N derivatives, our entropy S will contain up to $N - 2$ derivatives. Our new terms involve at least 4 derivatives so they appear only for $N \geq 6$.

A theory containing only the leading $\mathcal{O}(\ell^2)$ higher derivative corrections to Einstein gravity has $N = 4$. For such a theory, our result implies that the Wall entropy satisfies a non-perturbative second law in the sense of EFT: $\dot{S} \geq -\mathcal{O}(\ell^4)$

Example: Einstein-Gauss-Bonnet theory

$$16\pi G\mathcal{L} = -2\Lambda + R + kl^2 L_{GB}$$

View as an EFT for which coefficients of all higher derivative terms are zero. Consider horizon cross-section C . Entropy density $s/4G$.

If we trust this theory up to N derivatives then:

$$N = 2: s = 1$$

satisfies 2nd law modulo $\mathcal{O}(\ell^2)$

$$N = 4: s = 1 + kl^2 R[C]$$

satisfies 2nd law modulo $\mathcal{O}(\ell^4)$ (Wall terms vanish for this theory)

$N = 6$: (d is spacetime dimension)

$$s = 1 + kl^2 R[C] + \frac{1}{2}k^2 \left[(6-d) KK\bar{K}\bar{K} - KK\bar{K}^{AB}\bar{K}_{AB} + 4KK^{AB}\bar{K}_A{}^C\bar{K}_{BC} + \right. \\ \left. (-14+2d) KK^{AB}\bar{K}\bar{K}_{AB} - 2K^{AB}K_A{}^C\bar{K}_B{}^E\bar{K}_{CE} - 2K^{AB}K^{CE}\bar{K}_{AC}\bar{K}_{BE} + \right. \\ \left. (6-d)K^{AB}K^{CE}\bar{K}_{AB}\bar{K}_{CE} + 4K^{AB}K_A{}^C\bar{K}\bar{K}_{BC} - K^{AB}K_{AB}\bar{K}\bar{K} \right]$$

satisfies 2nd law modulo $\mathcal{O}(\ell^6)$ (K_{AB}, \bar{K}_{AB} : describe expansion and shear of outgoing/ingoing null geodesics orthogonal to C , $K = K^A{}_A, \bar{K} = \bar{K}^A{}_A$)

Example: $d = 4$ Riemann cubed

For $d = 4$ vacuum gravity field redefinitions can be used to eliminate 4-derivative terms and simplify 6-derivative terms to:

$$16\pi G\mathcal{L} = -2\Lambda + R + \ell^4(k_1\mathcal{L}_{\text{even}} + k_2\mathcal{L}_{\text{odd}})$$

$$\mathcal{L}_{\text{even}} = R_{\mu\nu\kappa\lambda}R^{\kappa\lambda\chi\eta}R_{\chi\eta}{}^{\mu\nu}$$

$$\mathcal{L}_{\text{odd}} = R_{\mu\nu\kappa\lambda}R^{\kappa\lambda\chi\eta}R_{\chi\eta\rho\sigma}\epsilon^{\mu\nu\rho\sigma}$$

Here $N = 6$ and our entropy density is (schematically)

$$s = 1 + k_1\ell^4 (\text{Riemann}^2 + \text{Riemann}K\bar{K} + KK\bar{K}\bar{K}) + k_2\ell^4\epsilon_2(\text{similar})$$

Bekenstein-Hawking + (Iyer-Wald + Wall + Hollands-Kovacs-HSR)

Satisfies second law modulo $\mathcal{O}(\ell^6)$

Gauge invariance

Our definition, and also Wall's, is formulated using GNCs.

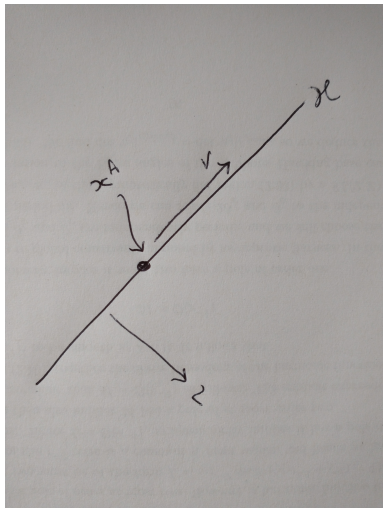
Does the definition depend on the choice of GNCs? If so then it is not fully geometrical.

We proved that Wall's definition does not depend on this choice.

Our definition does not depend on this choice for $N \leq 6$: for such theories we have a fully geometrical definition of black hole entropy that satisfies a non-perturbative second law in the sense of EFT.

For $N \geq 8$, our definition does depend on the choice of GNCs. Disappointing, but why would we ever need 8 derivatives?

Gaussian null coordinates



We assume: black hole spacetime with *smooth* horizon that “settles down to equilibrium” at late time.

Gaussian null coordinates (v, r, x^A) so horizon is at $r = 0$ and generators have constant x^A and affine parameter v .

$$ds^2 = -r^2 \alpha dv^2 + 2r \beta_A dv dx^A + 2dr dv + \mu_{AB} dx^A dx^B$$

Define *boost* as $v \rightarrow av$, $r \rightarrow r/a$ for constant a .

A quantity has *boost weight* b if it scales as a^b under a boost.

α, β_A and μ_{AB} have boost weight 0.

Expansion/shear of horizon generators described by $b = +1$
quantity $K_{AB} \equiv (1/2) \partial_v \mu_{AB}$

Expansion/shear of ingoing null geodesics described by $b = -1$
quantity $\bar{K}_{AB} \equiv (1/2) \partial_r \mu_{AB}$

All positive b quantities vanish on event horizon of a stationary black hole (assuming zeroth law).

Why area law doesn't hold

Rate of increase of horizon area: $\dot{A} = \int d^{d-2}x \sqrt{\mu} K_A^A$

Raychaudhuri eq: $\partial_\nu K_A^A = -K_{AB} K^{AB} - R_{\nu\nu}$

Usual proof: assume null convergence condition $R_{\nu\nu} \geq 0$ so $\partial_\nu K_A^A \leq 0$. "Late time equilibrium" implies $K_A^A \rightarrow 0$ as $\nu \rightarrow \infty$. Hence $K_A^A \geq 0$ hence $\dot{A} \geq 0$.

With higher derivatives, $R_{\nu\nu} = \mathcal{O}(\ell^2)$ need not have a good sign.

When EFT is valid, higher derivative terms are small so usually expect $K_{AB} K^{AB}$ to dominate. However, in some situations, e.g. black hole settling down to equilibrium, K_{AB} will also be small. Higher derivative terms like $\ell^2 \bar{K}^{AB} \partial_\nu K_{AB}$ (linear in K_{AB}) might then dominate, giving possible area decrease.

Wall's approach fixes this problem for linear perturbations of a stationary black hole.

Entropy current

An *entropy current* is a vector field (s^ν, s^A) defined on the horizon. Standard GR has $(s^\nu, s^A) = (1, 0)$.

Entropy of horizon cross-section $C(\nu)$ is $S(\nu) = \frac{1}{4G} \int_{C(\nu)} \sqrt{\mu} s^\nu$.

Divergence of entropy current is

$$\nabla \cdot s \equiv \frac{1}{\sqrt{\mu}} \partial_\nu (\sqrt{\mu} s^\nu) + D_A s^A$$

(Standard GR: $\nabla \cdot s = K_A^A$.) Then

$$\dot{S} = \frac{1}{4G} \int_{C(\nu)} \partial_\nu (\sqrt{\mu} s^\nu) = \frac{1}{4G} \int_{C(\nu)} \sqrt{\mu} \nabla \cdot s$$

Copying usual GR strategy would aim to show $\partial_\nu \nabla \cdot s \leq 0$ then, assuming late time equilibrium, we have $\nabla \cdot s \rightarrow 0$ as $\nu \rightarrow \infty$ so $\nabla \cdot s \geq 0$ hence $\dot{S} \geq 0$. Unfortunately this doesn't work.

Wall's approach can be formulated in terms of entropy current (Bhattacharyya *et al* (2019)): if eq of motion is $E_{\mu\nu} = 0$ then

$$\partial_\nu \nabla \cdot s = -E_{\nu\nu} - F$$

holds off-shell where F is quadratic (or higher) order in positive b quantities (so vanishes to linear order in perturbation theory).

This is a “generalized Raychaudhuri” equation. 2-derivative part of F contains usual $K_{AB}K^{AB}$ term.

We showed that F can be rearranged “nicely”, at the expense of going on-shell and invoking EFT. (Hollands, Kovacs & HSR 22)

Assume EFT equations of motion known up to N derivatives

$$E_{\mu\nu} = \mathcal{O}(\ell^N)$$

Working order by order in derivatives, and using equations of motion we show that new terms can be added to s^ν to define an “improved” entropy density S^ν and entropy current (S^ν, S^A) satisfying on-shell equation

$$\partial_\nu \nabla \cdot S = -(K_{AB} + X_{AB})(K^{AB} + X^{AB}) - D_A Y^A + \mathcal{O}(\ell^N)$$

where X_{AB} and Y^A arise from the higher derivative terms. The new terms in S^ν are quadratic in positive boost weight quantities. Similarly for Y^A .

To do this we show that equations of motion can be used to write everything in terms of “allowed quantities” of the form $\partial_\nu^p D^k K$, $\partial_r^p D^k \bar{K}$ or $D^k \beta$ e.g. $\partial_\nu \beta_A$ can be eliminated using $R_{\nu A} = \partial_\nu \beta_A + \dots$

$$\partial_\nu \nabla \cdot S = -(K_{AB} + X_{AB})(K^{AB} + X^{AB}) - D_A Y^A + \mathcal{O}(\ell^N)$$

Assuming late time equilibrium, $\nabla \cdot S \rightarrow 0$ as $\nu \rightarrow \infty$:

$$(\nabla \cdot S)(\nu_0) = \int_{\nu_0}^{\infty} d\nu' \left[(K_{AB} + X_{AB})(K^{AB} + X^{AB}) + D_A Y^A - \mathcal{O}(\ell^N) \right]$$

Integrate over horizon cross-section (Hollands, Kovacs & HSR 22)

$$\dot{S}(\nu_0) = \frac{1}{4G} \int d^{d-2}x \sqrt{\mu(\nu_0)} \int_{\nu_0}^{\infty} d\nu' \left[(K_{AB} + X_{AB})(K^{AB} + X^{AB}) + D_A Y^A - \mathcal{O}(\ell^N) \right]$$

Davies & HSR 23: Integrate final terms by parts: generates *bilocal* terms, depending on both ν_0 and ν' . But these can be rearranged into a nice form, generating further “improvement” terms in S :

$$\dot{S}_{\text{new}}(\nu_0) = \frac{1}{4G} \int d^{d-2}x \sqrt{\mu(\nu_0)} \int_{\nu_0}^{\infty} d\nu' \left[(K_{AB} + Z_{AB})(K^{AB} + Z^{AB}) - \mathcal{O}(\ell^N) \right]$$

where Z_{AB} is bilocal. Hence $\dot{S}_{\text{new}} \geq -\mathcal{O}(\ell^N)$: *non-perturbative* second law, in sense of EFT.

Matter fields

With matter fields, assume 2-derivative terms satisfy NEC then try to use them to help control higher derivative terms via completing the square as above.

This has been worked out in detail for the EFT of gravity coupled to a Maxwell field and a scalar field (Davies 2024).

Field redefinitions and uniqueness

Our definition of dynamical entropy appears not to be invariant under EFT field redefinitions.

For example take vacuum gravity $R_{\mu\nu} = 0$ and perform field redefinition $g_{\mu\nu} \rightarrow g_{\mu\nu} + c_1 \ell^2 R_{\mu\nu} + c_2 \ell^2 R g_{\mu\nu}$ (trivial on-shell!): generates R^2 and $R_{\mu\nu} R^{\mu\nu}$ terms in action and new terms in entropy that don't vanish on-shell.

Is this a problem? Thermodynamic entropy is not unique away from equilibrium. Example of fluid dynamics: for relativistic viscous fluid, there are multi-parameter families of entropy currents satisfying second law. Bhattacharyya *et al* 2008,2013, Romatschke 2009

Maybe field redefinitions just map between different possible definitions of the entropy, all of which satisfy a second law.

Related recent work

Hollands, Wald and Zhang 24: new definition of black hole entropy for diffeomorphism invariant theories, including GR.

Second law holds to linear order in perturbations around a stationary black hole (second order for GR). Simple relation to Wall entropy. In GR, entropy coincides with area of *apparent* horizon.

“Physical process” version of first law: increase in entropy between two horizon cuts determined by energy and angular momentum of matter crossing horizon between these cuts.

“Background structure” of stationary black hole is essential so a non-perturbative generalization seems unlikely.

Summary

We have introduced a procedure for defining black hole entropy in EFT such that a second law is satisfied non-perturbatively, in the sense of EFT, i.e., it holds for solutions that remain within the regime of validity of EFT, and it holds modulo terms of same size as the unknown higher order terms in the EFT equation of motion.

Open questions:

- ▶ Different types of matter fields
- ▶ Field redefinitions, non-uniqueness
- ▶ Event horizon as a characteristic surface in Einstein-Gauss-Bonnet (HSR 20)
- ▶ Defining entropy for non-smooth horizons (Gadioux & HSR 23)