

Black holes and causal nonlinear electrodynamics

Jorge G. Russo ^{a,b} and Paul K. Townsend ^c

^aInstitució Catalana de Recerca i Estudis Avançats (ICREA),
Pg. Lluis Companys, 23, 08010 Barcelona, Spain.

^bDepartament de Física Cuántica i Astrofísica and Institut de Ciències del Cosmos,
Universitat de Barcelona, Martí Franquès, 1, 08028 Barcelona, Spain.

^cDepartment of Applied Mathematics and Theoretical Physics,
Centre for Mathematical Sciences, University of Cambridge,
Wilberforce Road, Cambridge, CB3 0WA, U.K.

e-mail: jorge.russo@icrea.cat, pkt10@cam.ac.uk

ABSTRACT

For generic theories of nonlinear electrodynamics (NLED) we investigate the restrictions imposed by causality on spherically-symmetric charged black-hole solutions of the Einstein-NLED equations. For a large class of (acausal) Born-type NLED theories, we find that the Reissner–Nordström (RN) metric is an exact, but unstable, solution for some dyonic black holes. For all causal NLED we show that there are no regular black holes, and that the entropy of extremal black holes is less than the RN entropy for fixed charge. We also find the conditions for a parameter-space transition between RN-type and Schwarzschild-type global structure. For the transition from Schwarzschild-type to naked singularity, which occurs at finite mass, we show that the metric at the transition point is a Barriola-Vilenkin global monopole.

Contents

1	Introduction	1
2	NLED in general spacetimes	6
2.1	ADM coordinates and the Hamiltonian	6
2.2	Self-dual NLED	9
3	Einstein-NLED equations and black holes	10
3.1	Asymptotically-flat charged black holes	13
3.2	Near-singularity expansions	15
3.2.1	Born-Infeld	17
3.3	Electric field and energy	18
3.4	Black holes for self-dual NLED	20
3.4.1	A Lagrangian perspective	21
3.5	NLED theories with exact Reissner-Nordström black holes	22
3.5.1	Black hole instability for acausal NLED	24
4	Global structure for causal NLED	26
4.1	Causality implies singularity	27
4.2	Killing horizons	29
4.3	The vanishing Cauchy horizon	30
4.4	Extremal charged black holes	34
4.5	Deficit solid angle and the Born particle	36
5	Summary and Discussion	37

1 Introduction

The Einstein-Maxwell field equations have as solutions the three-parameter family of Reissner-Nordström (RN) charged black holes, with a spacetime metric that is asymptotically-flat and spherically-symmetric. The parameters are the black-hole mass M and its electric and magnetic charges (q_e, q_m), although the RN metric depends only on the two parameters M and

$$Q = \sqrt{q_e^2 + q_m^2}. \quad (1.1)$$

An important property of the Einstein-Maxwell field equations of relevance to the classical physics of black holes is that they are second-order partial differential equations, and this remains true if Maxwell electrodynamics is replaced by any theory of nonlinear electrodynamics (NLED) defined by a scalar Lagrangian function $L(F)$ of the Faraday field strength 2-form $F = dA$ (for 1-form gauge potential A) but not derivatives of F . It is therefore natural to consider black holes as solutions of the more general Einstein-NLED field equations.

Here we consider only NLED theories with a weak-field expansion and a conformal weak-field limit. Simple examples are Born’s original 1933 NLED theory [1] and the subsequent Born-Infeld modification of it [2]. In both cases the Lagrangian depends on a constant T with dimensions of energy density (this is the “Born tension”, which is the square of the “Born constant”). The $T \rightarrow \infty$ limit for fixed fields is a weak-field limit to Maxwell, and the weak-field expansion is a formal series expansion of $L(F)$ in powers of $1/T$. More generally we can expect the interactions in $L(F)$ to depend on a number of dimensionless parameters and a single parameter T which sets the scale for field-energy densities at which the interactions become important.

A special subclass of NLED theories is composed of those with a Hamiltonian invariant under a $U(1)$ electromagnetic duality group. These are “self-dual”; the free-field Maxwell theory is an example, as is the Born-Infeld theory. Our derivation of static charged black-hole solutions of the Einstein-NLED equations makes it manifest that for self-dual NLED the spacetime metric can depend on the charges (q_e, q_m) only through its dependence on Q , and more generally that duality symmetries of the Hamiltonian imply corresponding symmetries in the (q_e, q_m) parameter space of the spacetime metric. This result is perhaps not surprising but we know of no previous proof of it. Another example is the Z_2 -duality invariance of the original Born theory, which implies isometric electric and magnetic black hole spacetimes.

One of the expected effects of replacing the free-field Maxwell theory by some interacting NLED theory is that asymptotically-flat static black-hole solutions of the Einstein-NLED equations will be only asymptotic to the RN solutions of the Einstein-Maxwell equations. The mass M and charge Q of the RN spacetime determine two distinct length scales; these are GM and $\sqrt{G}Q$, where G is Newton’s constant. For any non-conformal NLED theory the Born tension T determines an additional “gravitational-Born” length scale:

$$\ell_{gB} \sim 1/\sqrt{GT}. \quad (1.2)$$

Notice that this is zero in the $T \rightarrow \infty$ limit, which is a weak-field limit. We can expect (and it is true) that the existence of an event horizon requires $GM \gtrsim \sqrt{G}Q$. This suggests that significant modifications to the RN metric occur when

$$\ell_{gB} \gtrsim \sqrt{G}Q, \quad (1.3)$$

which can viewed as a “small charge” condition.

Despite the expectation of deviations from the RN metric, a natural question is whether there are NLED theories for which the RN metric, for some choice of charges (q_e, q_m) , is an exact solution. The answer is no for causal theories but yes for some acausal theories; Born’s original (and acausal [3]) theory is an example since the RN solution is exact for $q_e = q_m$. In fact, we show that exact dyonic RN black-hole solutions exist for an entire class of Born-type theories, all acausal for sufficiently strong fields, and we use the example of the dyonic RN Born black hole to examine the effects of

strong-field acausality on charged black-holes. One might suppose that they are only significant behind the event horizon, but a simple calculation based on Minkowski spacetime results of [4], and valid for $\ell_{gB} \gg \sqrt{G} Q$, shows that an instability against perturbations that are not spherically-symmetric can set in outside the event horizon.

Generic NLED theories are acausal because they allow superluminal propagation of shock-wave discontinuities in smooth electromagnetic backgrounds [5, 6] or (equivalently) of small-amplitude perturbations of stationary homogeneous backgrounds, which can be viewed as anisotropic optical media [7]. The necessary and sufficient conditions on $L(F)$ for this not to be possible, i.e. for causality, were found in [3]. They comprise conventional convexity conditions which are necessary for weak-field causality but also an additional condition needed for strong-field causality, which we have rederived, discussed and applied in recent works [8–12]. Our principal aim here is to deduce implications of causality for spherically-symmetric charged black-hole solutions of the Einstein-NLED equations¹.

There are several issues of interest for which causality plays a role in restricting possibilities. One is whether the central spacetime singularity of a charged static black-hole solution of the Einstein-NLED equations can be absent for some choice of NLED theory [14–22]. A restriction on this choice is inherent in the Hawking-Penrose theorem to the effect that singularities are inevitable, once a trapped surface appears, whenever the matter stress-energy tensor satisfies the Strong Energy Condition (SEC) [23]. Thus, regular black-hole solutions require an SEC-violating NLED theory. However, one consequence of the strong-field causality condition mentioned above is that it implies the SEC [10], assuming non-positive vacuum energy. The possibility of asymptotically-flat (or adS) regular black holes is thus ruled out for causal NLED theories. This leaves open the possibility of regular charged black holes in an ambient de Sitter spacetime but it is not difficult to see that a non-zero cosmological constant has no effect on singularities. We also clarify and complete some more direct arguments of Bronnikov that there are no regular charged spherically-symmetric black holes for any causal NLED theory [15, 16].

Only radial electric/magnetic fields, depending only on a radial coordinate r , are compatible with spherical symmetry and time-independence, which means that the NLED Hamiltonian function H becomes a “charged black-hole” function $H(r)$. Many features of spherically-symmetric Einstein-NLED black holes depend crucially on the behaviour of this function near the central singularity: $r = 0$ in Schwarzschild coordinates. For all explicitly known causal NLED theories that we are aware of, the leading term as $r \rightarrow 0$ is a power of r ; i.e.

$$H(r) \sim r^{-4\nu} \quad (r \rightarrow 0) \tag{1.4}$$

for some constant ν . For Maxwell $\nu = 1$ because the electric/magnetic fields diverge as $1/r^2$ and H is a quadratic function of these fields.

¹Such solutions are necessarily static by Birkhoff’s theorem; see [13].

The function $H(r)$ determines two other functions that are important to properties of static black-hole spacetimes. One is a function $\mathcal{E}(r)$ (introduced as $U_{\text{self}}(r)$ in [20]) which gives the electromagnetic energy outside a sphere of radius r . The other is an “effective charge” function $Q_{\text{eff}}(r)$ that approaches the Maxwell charge Q as $r \rightarrow \infty$ (assuming that Maxwell is the weak-field limit). The importance of these functions is that they are restricted in a very simple way by causality. Specifically, $\mathcal{E}(r)$ must be convex [20], and $Q_{\text{eff}}^2(r)$ must be concave (as we show here). The combination of these two convexity/concavity constraints fixes the range of ν to be

$$\frac{1}{2} \leq \nu \leq 1. \quad (1.5)$$

The value of ν also determines (for any causal NLED) whether the electric field is finite at $r = 0$ (it is iff $\nu = \frac{1}{2}$) and whether the total electromagnetic energy is finite (it is iff $\frac{1}{2} \leq \nu < \frac{3}{4}$).

An assumption was made in [20] about the behaviour of $\mathcal{E}(r)$ as $r \rightarrow 0$. This assumption agrees with what we find from (1.4) for $\nu = \frac{1}{2}$, which includes Born-Infeld. However, it disagrees with what we find for $\frac{1}{2} < \nu < \frac{3}{4}$; i.e. the cases for which the electric field diverges as $r \rightarrow 0$ but the electromagnetic energy is finite). This disagreement is significant because it explains a discrepancy that we find with the results of [20] on the global structure of spherically-symmetric charged black holes for generic causal NLED theories. Our results are also based on an assumption: that $H(r)$ has a power-law behaviour as $r \rightarrow 0$, but this assumption is backed up by the existence of an explicit causal (and self-dual) NLED theory with a Hamiltonian function that yields a charged black-hole function $H(r)$ with the assumed power-law behaviour (irrespective of the electric/magnetic charge ratio because of self-duality). This demonstrates, at the very least, that the assumption of [20] is not always valid.

As observed above, the function $H(r)$ can be found on a case-by-case basis, given any Hamiltonian function H of a Lorentz invariant NLED theory, and a choice of the black-hole charges (q_e, q_m) . However, one can do better for self-dual theories because the restriction to static fields means that H is a function of a single scalar variable s (the free-field energy density), and $2s = Q^2/r^4$. In these cases the form of $H(s)$ as $s \rightarrow \infty$ directly determines the form of the function $H(r)$ as $r \rightarrow 0$. Moreover, causality conditions for self-dual NLED reduce to simple conditions on $H(s)$ [11]. As expected, they imply (1.5), and this simple derivation of (1.5) can also be applied to non-self-dual NLED if we restrict to purely electric black holes. This is because the restriction to a static electric field reduces the Hamiltonian function H of any NLED to the one-variable function $H(s)$ of some self-dual NLED. The electric black holes are therefore the same as that of the self-dual NLED, as are their properties except that strong-field causality (and hence stability) is no longer guaranteed.

A major aim of this paper is to explore the implications of causality on the global structure of spherically-symmetric and asymptotically-flat Einstein-NLED charged black-holes. One implication is that there cannot be more than two Killing horizons; this

was established in [20] under the assumption of finite electromagnetic energy but this assumption is not required. It should be noted that black-hole spacetimes with more than two Killing horizons are easily found if causality is not required [24].

Whenever there are two Killing horizons they can be identified as the event horizon and an (interior) Cauchy horizon. In the special case that these two Killing horizons coincide there is only one (degenerate) horizon for an “extremal” black hole, and the mass-to-charge ratio takes its lowest value compatible with the existence of an event horizon. This lowest mass-to-charge ratio depends on the choice of NLED theory. We show that the nonlinearities of all causal NLED theories necessarily reduce this ratio. Thus the extremal RN black hole has greater mass, for fixed charge, than the extremal black-hole of any interacting causal NLED theory. This leads to the conclusion that NLED interactions reduce the entropy of an extremal black-hole, at fixed charge.

For Maxwell, and the non-extremal RN metric, the geometry is such that a test-particle freely falling through the event horizon will then pass through the Cauchy horizon before reaching a timelike singularity at the origin ($r = 0$). This is true for any value of the charge Q but for generic NLED theories we should expect deviations for ‘small charge’ black holes: those for which (1.3) holds. In fact, as first shown long ago by Oliveira [14], the Cauchy horizon is absent for a family of NLED theories that includes Born-Infeld black holes, when the charge is sufficiently small. A recent observation of Hale et al. is that this phenomenon requires finite electromagnetic energy [20]. However, as we show, here, their analysis fails to include *all* causal NLED theories with finite electromagnetic energy; there is a much larger class of NLED theories for which this is true (those for which $\nu < \frac{3}{4}$). In the class of theories considered by Hale et al. ($\nu = \frac{1}{2}$) there is a critical charge below which no Cauchy horizon can exist irrespective of the value of the mass. In the most general class of causal NLED with finite electromagnetic energy, RN-type black holes exist for arbitrarily small charge, provided the mass is greater than the extremal mass and lower than the total electromagnetic energy.

For the special theories with $\nu = \frac{1}{2}$, below the critical charge the singularity is spacelike (naked) for a mass greater (lower) than the total electromagnetic energy. Our main result is that the spacetime geometry at the transition point is that of the Barriola-Vilenkin global monopole, which has a non-flat conical singularity. We also comment on some unusual global features of charged black holes with a Schwarzschild-type global structure, and we comment briefly on some surprising quantum effects due to Hawking radiation.

The organisation is as follows. We begin with a review of generic NLED theories in general spacetimes, including self-dual NLED as a special case. Restricting to the spherically-symmetric black-hole spacetimes of most interest here we then show how the Einstein-NLED field equations reduce to a single first-order ODE involving the Hamiltonian of the chosen NLED theory, and we discuss the implications of causality based mainly on convexity/concavity properties required by causality for two functions related to the electromagnetic energy and the “effective” charge. The NLED classifi-

cation involving the number ν explained above is used to illustrate results. We also consider some acausal theories, in particular Born's original theory, with sample calculations to show how acausality manifests itself as black hole instabilities. We then move on to aspects of the spacetime geometry, singularities, Killing horizons, the existence of a Cauchy horizons, the nature of the transition point where it ceases to exist, and the absence of charge in charged black holes without a Cauchy horizon.

2 NLED in general spacetimes

The general NLED with a weak-field limit has a Lagrangian density \mathcal{L} that is a function of two independent Lorentz scalars quadratic in the field-strength 2-form $F = dA$. If we allow for a generic spacetime with metric g , then

$$\mathcal{L} = \sqrt{|g|} L(S, P), \quad (2.1)$$

where $|g| = -\det g$, and L is a scalar function of the (pseudo)scalars

$$S = -\frac{1}{4}g^{\mu\rho}g^{\nu\sigma}F_{\mu\nu}F_{\rho\sigma}, \quad P = -\frac{1}{8\sqrt{|g|}}\varepsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}. \quad (2.2)$$

Maxwell electrodynamics corresponds to $L = S$ in units for which $c = 1$, where c is the *in vacuo* speed of light. The NLED stress-energy tensor can be found from the Hilbert formula

$$T_{\mu\nu} = -\frac{2}{\sqrt{|g|}}\frac{\partial \mathcal{L}}{\partial g^{\mu\nu}}. \quad (2.3)$$

This yields

$$T_{\mu\nu} = L_S T_{\mu\nu}^{\text{Max}} - (SL_S + PL_P - L)g_{\mu\nu}, \quad (2.4)$$

where the Maxwell stress-energy tensor is

$$T_{\mu\nu}^{\text{Max}} = g^{\rho\sigma}F_{\mu\rho}F_{\nu\sigma} + Sg_{\mu\nu}. \quad (2.5)$$

Since the Maxwell stress-energy tensor is traceless (with respect to the metric g) we have

$$\Theta := g^{\mu\nu}T_{\mu\nu} = -4(SL_S + PL_P - L) \leq 0, \quad (2.6)$$

where the inequality follows from the requirement of causality, and the assumption of zero vacuum energy (required for asymptotically-flat spacetimes). This is because $SL_S + PL_P \geq L$ for any convex function $L(S, P)$ such that $L(0, 0) = 0$, and convexity of $L(S, P)$ is a necessary condition for causality, as briefly explained below.

2.1 ADM coordinates and the Hamiltonian

To pass to the Hamiltonian formulation it is useful to choose coordinates $x^\mu = (t, x^i)$ for which the spacetime metric takes the following (ADM) form:

$$ds^2(g) = -\mathcal{N}^2 dt^2 + h_{ij} (dx^i + u^i dt) (dx^j + u^j dt) \quad (\Rightarrow |g| = \mathcal{N}^2 \det h). \quad (2.7)$$

In this case the (pseudo)scalars (S, P) are

$$S = \frac{1}{2} \left\{ \frac{1}{\mathcal{N}^2} |E - u \times B|^2 - \frac{1}{\det h} |B|^2 \right\}, \quad P = \frac{1}{\mathcal{N} \sqrt{\det h}} E_i B^i, \quad (2.8)$$

where

$$E_i := F_{i0} = \partial_t A_t - \partial_t A_i, \quad B^i := \frac{1}{2} \varepsilon^{ijk} F_{jk} \equiv \varepsilon^{ijk} \partial_j A_k, \quad (2.9)$$

and

$$(u \times B)_i := \varepsilon_{ijk} u^j B^k, \quad (\varepsilon^{ijk} \varepsilon_{lmn} = 6 \delta^i_{(l} \delta^j_m \delta^k_{n)}). \quad (2.10)$$

The norm $|..|$ is taken using the 3-metric h ; e.g.

$$|E|^2 = h^{ij} E_i E_j, \quad |B|^2 = h_{ij} B^i B^j. \quad (2.11)$$

For this choice of spacetime coordinates, the dielectric displacement field is

$$D^i := \frac{\partial \mathcal{L}}{\partial E_i} = L_S \frac{\sqrt{\det h}}{\mathcal{N}} h^{ij} (E - u \times B)_j + L_P B^i. \quad (2.12)$$

Solving this equation for E as a function of D allows us to find the Hamiltonian density $\mathcal{H}(D, B)$ by Legendre transform of $\mathcal{L}(E, B)$. The solution will be unique whenever $\mathcal{L}(E, B)$ is a strictly convex function of E . This is equivalent to strict convexity of the function $L(S, P)$ combined with the condition $L_S > 0$ [25], which are necessary conditions for causality of any NLED with a weak-field limit [3], and sufficient for weak-field causality [8].

As it is not always possible to find $\mathcal{H}(D, B)$ explicitly from $\mathcal{L}(E, B)$, a useful alternative starting point is the ‘phase-space’ Lagrangian density:

$$\mathcal{L}' = E_i D^i - \mathcal{H}(D, B), \quad (2.13)$$

where D is now an independent field and \mathcal{H} is the Hamiltonian density. The integral of \mathcal{H} over any volume V at fixed time t is the electromagnetic energy in V :

$$\mathcal{E}_{\text{em}}(V) = \int_V d^3x \mathcal{H}(D, B). \quad (2.14)$$

The field equations that follow from \mathcal{L}' imply that

$$\dot{\mathcal{H}} = -\partial_i (E \times H)^i, \quad H_i := \frac{\partial \mathcal{H}}{\partial B^i}. \quad (2.15)$$

For any NLED that is (locally) Lorentz invariant, we have

$$(E \times H)^i = g^{ij} p_j, \quad p_i := (D \times B)_i, \quad (2.16)$$

where p_i is the momentum density; this is zero for static electromagnetic fields, and therefore $\mathcal{E}_{\text{em}}(V)$ is time-independent for static fields. Notice that both D and B are divergence-free since $\partial_i B^i = 0$ is an identity and $\partial_i D^i = 0$ is the constraint imposed by

the Lagrange multiplier A_t (in the absence of electric charges). Variation of D in \mathcal{L}' yields

$$E_i = \frac{\partial \mathcal{H}}{\partial D^i}. \quad (2.17)$$

This equation will uniquely determine D as a function of E if \mathcal{H} is a strictly convex function of D , which is required for causality of any non-conformal NLED with a weak-field expansion². Thus, D is an auxiliary field in this context.

For the general spacetime metric of (2.7) we have

$$\mathcal{H}(D, B) = \mathcal{N}\sqrt{\det h} H(D, B) - u^i p_i, \quad (2.18)$$

where $H(D, B)$ may be expressed, assuming rotation invariance, as a function $H(x, y, z)$ of the following three 3-space rotation scalars:

$$x = \frac{1}{2\det h} h_{ij} D^i D^j, \quad y = \frac{1}{2\det h} h_{ij} B^i B^j, \quad z = \frac{1}{\det h} h_{ij} D^i B^j. \quad (2.19)$$

We may now rewrite (2.17) as

$$E_i = \left(\frac{\mathcal{N}}{\sqrt{\det h}} \right) h_{ij} (H_x D^j + H_z B^j). \quad (2.20)$$

Using this to eliminate D from \mathcal{L}' yields a Lagrangian density $\mathcal{L}(E, B)$ but this will be expressible in the ‘Plebanski’ form of (2.1) only if the (local) Lorentz invariance condition (2.16) is satisfied; for Hamiltonian functions $H(x, y, z)$ this condition is equivalent to [4]

$$H_x H_y - H_z^2 = 1. \quad (2.21)$$

If this condition is satisfied and if \mathcal{H} is a strictly convex function of D then $\mathcal{H}(D, B)$ is the Legendre dual of the Lagrangian function $\mathcal{L}(E, B)$ found by elimination of D from \mathcal{L}' . For example, the free-field Maxwell case is $H = x + y$, and elimination of D yields $\mathcal{L} = \sqrt{|g|} S$, with S given by (2.8).

We remark that the Hilbert formula of (2.3) for the stress-energy tensor can also be used for \mathcal{L}' , in which case it is equivalent to the relations

$$\begin{aligned} T_{tt} - 2u^i T_{ti} + u^i u^j T_{ij} &= -\frac{\mathcal{N}^2}{\sqrt{\det h}} \frac{\partial \mathcal{L}'}{\partial \mathcal{N}} = \mathcal{N}^2 H \\ T_{ti} - u^j T_{ij} &= -\frac{\mathcal{N}}{\sqrt{\det h}} \frac{\partial \mathcal{L}'}{\partial u^i} = -\frac{\mathcal{N}}{\sqrt{\det h}} (D \times B)_i \\ T_{ij} &= -\frac{2}{\mathcal{N} \sqrt{\det h}} \frac{\partial \mathcal{L}'}{\partial h^{ij}} = 2 \frac{\partial H}{\partial h^{ij}} - h_{ij} H. \end{aligned} \quad (2.22)$$

For static electromagnetic fields on a static spacetime we have $u = 0$ and $|D \times B| = 0$, and all non-zero fields and metric components are time-independent. In this case (of relevance here) we have

$$H = \mathcal{N}^{-2} T_{tt} = T^t_t. \quad (2.23)$$

²If the convexity is not strict then there will be Lagrangian constraints and no weak-field limit.

2.2 Self-dual NLED

For the subset of Lorentz-invariant NLED that are also $U(1)$ electromagnetic-duality invariant (i.e. self-dual) the Lagrangian function L must satisfy a particular nonlinear first-order differential equation [7]. In terms of new scalar variables (U, V) defined by [26, 27]

$$S = V - U, \quad P = \sqrt{4UV} \quad (V \geq U \geq 0), \quad (2.24)$$

the self-duality equation is $L_U L_V = -1$ and the general solution (given in [28]) is

$$L(U, V) = \ell(\tau) - \frac{2U}{\dot{\ell}(\tau)}, \quad \tau = V + \frac{U}{[\dot{\ell}(\tau)]^2}, \quad (2.25)$$

where $\ell(\tau)$ is a “CH-function” which has a Taylor expansion in powers of τ for self-dual NLED theories with a weak-field expansion:

$$\ell(\tau) = e^\gamma \tau + \mathcal{O}(\tau^2) \quad (\tau \geq 0). \quad (2.26)$$

The absence of a constant term implies zero vacuum energy, and causality of the weak-field limit. The linear term yields the conformal Maxwell/ModMax family with

$$\mathcal{L} = e^\gamma V - e^{-\gamma} U \equiv (\cosh \gamma) S + (\sinh \gamma) \sqrt{S^2 + P^2}, \quad (2.27)$$

and causality requires $\gamma \geq 0$ [29]. The higher-order powers of τ introduce the non-conformal interactions of the weak-field expansion. NLED theories defined by ℓ in this way are causal iff [9]

$$\dot{\ell} \geq 1 \quad \ddot{\ell} \geq 0, \quad (2.28)$$

where the overdot indicates a derivative with respect to the independent variable. The first inequality is an equality only for Maxwell. The second inequality is an equality only for the Maxwell/ModMax family, which means that ℓ is a strictly convex function for any non-conformal causal self-dual NLED.

Since $(U, V) = (0, S)$ when $B = 0$, which implies $P = 0$, we have

$$L(S, 0) = \ell(S) \quad (B = 0 : \Rightarrow S \geq 0). \quad (2.29)$$

This can be used as a definition of the function ℓ for generic NLED theories. In this general context, the conditions (2.28) become necessary conditions for causality, equivalent to convexity of L as a function of E when $B = 0$, but they only guarantee strong-field causality for self-dual NLED.

In the Hamiltonian formulation, self-duality can be made manifest by restricting H to be a function of the two duality invariant variables

$$s := x + y, \quad p := \sqrt{4xy - z^2} \equiv |D \times B|. \quad (2.30)$$

However, a function $H(s, p)$ will not define a Lorentz invariant NLED unless it satisfies a particular nonlinear first-order differential equation. In terms of new variables (u, v) defined by

$$s = v + u, \quad p = \sqrt{4uv} \quad (v \geq u \geq 0), \quad (2.31)$$

the general solution of this Lorentz-Invariance condition is

$$H(u, v) = \mathfrak{h}(\sigma) + \frac{2u}{\mathfrak{h}'(\sigma)}, \quad \sigma = v - \frac{u}{[\mathfrak{h}''(\sigma)]^2}, \quad (2.32)$$

where \mathfrak{h} is a “Hamiltonian CH-function”, and the prime indicates a derivative with respect to the independent variable. The conditions for causality may now be expressed as the following conditions on this function [11]:

$$0 < \mathfrak{h}'(\sigma) \leq 1, \quad \mathfrak{h}''(\sigma) \leq 0, \quad \mathfrak{h}'(\sigma) + 2\sigma\mathfrak{h}''(\sigma) > 0. \quad (2.33)$$

Notice that for $p = 0$ we have $(u, v) = (0, \sigma)$, and hence

$$H(s) = \mathfrak{h}(s) \quad (p = 0) \quad (2.34)$$

This may be taken as a definition of \mathfrak{h} for generic NLED, in which context the conditions (2.33) become necessary conditions for causality that are sufficient only for self-dual NLED.

For any non-conformal NLED theory, $\ell(\tau)$ and $\mathfrak{h}(\sigma)$ are related as follows [11, 12]

$$\ell(\tau) = 2\sigma\mathfrak{h}'(\sigma) - \mathfrak{h}(\sigma), \quad \tau = \sigma [\mathfrak{h}']^2, \quad (2.35)$$

or, equivalently,

$$\mathfrak{h}(\sigma) = 2\tau\dot{\ell}(\tau) - \ell(\tau), \quad \sigma = \tau [\dot{\ell}(\tau)]^2. \quad (2.36)$$

These relations imply that

$$\dot{\ell}(\tau)\mathfrak{h}'(\sigma) = 1. \quad (2.37)$$

For any Lorentz invariant self-dual NLED with Hamiltonian $H(s, p)$ its conformal weak-field limit is (if it exists) either Maxwell or ModMax [29]; the Maxwell/Modmax family has $\mathfrak{h}(\sigma) = e^{-\gamma}\sigma$ and

$$H_{MM}(s, p) = (\cosh \gamma)s - (\sinh \gamma)\sqrt{s^2 - p^2}, \quad (2.38)$$

where $\gamma \geq 0$ is a non-negative dimensionless coupling constant. Maxwell is the free-field $\gamma = 0$ case (the $\gamma < 0$ cases are acausal).

3 Einstein-NLED equations and black holes

The Einstein field equation for the spacetime metric g is

$$G_{\mu\nu} = (8\pi G) T_{\mu\nu}, \quad (3.1)$$

where $G_{\mu\nu}$ and $T_{\mu\nu}$ are, respectively, the Einstein tensor and the NLED stress-energy tensor, and G is Newton’s constant. Here we restrict to static asymptotically-flat spacetimes associated to charged (generically dyonic) black holes. In the context of the general spacetime metric of (2.7) this means that $u^i = 0$ and both \mathcal{N} and h_{ij} are

time-independent. If we further assume spherical symmetry, and choose Schwarzschild radial coordinates, we arrive at a family of metrics parameterised by functions $\mathcal{N}(r)$ and $h_{rr}(r)$:

$$ds^2 = -\mathcal{N}^2(r)dt^2 + h_{rr}(r)dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2). \quad (3.2)$$

The electromagnetic fields must also be static and compatible with spherical symmetry. This implies that the only non-zero components of (D, B) are the radial components (D^r, B^r) , which must be t -independent. They must also be r -independent by the divergence-free conditions, and scalar densities on the unit 2-sphere. Therefore

$$D^r = q_e \sin \theta, \quad B^r = q_m \sin \theta, \quad (3.3)$$

for constants (q_e, q_m) that can be identified as the electric and magnetic charges. The 3-space scalars (x, y, z) are now³

$$x = \frac{q_e^2}{2r^4}, \quad y = \frac{q_m^2}{2r^4}, \quad z = \frac{q_e q_m}{r^4}. \quad (3.4)$$

From (2.22) we see that the only non-zero components of the NLED stress-energy tensor for static field configurations are

$$T_{tt} = \mathcal{N}^2 H, \quad T_{ij} = 2 \frac{\partial H}{\partial h^{ij}} - h_{ij} H. \quad (3.5)$$

The further restriction to spherical symmetry implies that only the diagonal components of T_{ij} are non-zero, and because the expressions of (3.4) are independent of h_{rr} , we have

$$T_{rr} = -h_{rr} H. \quad (3.6)$$

The other non-zero components are

$$T_{\theta\theta} = (\sin \theta)^{-2} T_{\phi\phi} = r^2 [2(xH_x + yH_y + zH_z) - H]. \quad (3.7)$$

The right-hand side can be simplified by using the fact that

$$4(xH_x + yH_y + zH_z) = -rH'(r), \quad (3.8)$$

where, on the right-hand side, $H(r)$ is the function⁴ found from $H(x, y, z)$ via the r -dependence of the variables (x, y, z) as given in (3.4). We thus find that

$$T_{\theta\theta} = (\sin \theta)^{-2} T_{\phi\phi} = -\frac{1}{2} r(r^2 H)' . \quad (3.9)$$

Recalling the definition of Θ as the trace of the stress-energy tensor, it follows from the above results that

$$\Theta = -(rH' + 4H), \quad (3.10)$$

³Since $4xy - z^2 \equiv |D \times B|^2$, which must be zero for a static solution of the NLED field equations, we have $4xy = z^2$.

⁴We use italic font for the Hamiltonian function and roman font for the associated function of r .

which is a result that we shall use later.

For convenience, we now set

$$\mathcal{N}(r) = e^{\alpha(r)}, \quad h_{rr}(r) = e^{2\beta(r)}, \quad (3.11)$$

which gives us the following expression for the general spherically-symmetric spacetime metric:

$$ds^2 = -e^{2\alpha(r)}dt^2 + e^{2\beta(r)}dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2). \quad (3.12)$$

A computation of the Einstein tensor for this metric shows that the non-zero components match with the non-zero components of the NLED stress-energy tensor, with the result that the Einstein field equations are equivalent to the following component equations:

$$\begin{aligned} 0 &= G_{tt} - (8\pi G)T_{tt} = e^{2\alpha} \left\{ r^{-2} \left[1 - (re^{-2\beta})' \right] - (8\pi G)H \right\} \\ 0 &= G_{rr} - (8\pi G)T_{rr} = -e^{2\beta} \left\{ r^{-2} \left[1 - e^{-2(\alpha+\beta)}(re^{2\alpha})' \right] - (8\pi G)H \right\} \\ 0 &= G_{\theta\theta} - (8\pi G)T_{\theta\theta} = e^{-2\beta} \left[e^{-2\alpha} (r^2 e^{2\alpha} \alpha')' - (r + r^2 \alpha')(\alpha + \beta)' \right] \\ &\quad + (4\pi G)r (r^2 H)' . \end{aligned} \quad (3.13)$$

The $\phi\phi$ -component yields nothing new since

$$G_{\phi\phi} - (8\pi G)T_{\phi\phi} = \sin^2\theta [G_{\theta\theta} - (8\pi G)T_{\theta\theta}]. \quad (3.14)$$

The tt and rr components are compatible iff

$$\alpha + \beta = 0, \quad (3.15)$$

and then both equations reduce to the one equation

$$1 - (re^{2\alpha})' = (8\pi G)r^2 H. \quad (3.16)$$

The remaining independent ($\theta\theta$) equation is now

$$(r^2 e^{2\alpha} \alpha')' = -(4\pi G)r (r^2 H)', \quad (3.17)$$

but this is implied by (3.16), which is therefore the only equation that we need solve; we may rewrite it as

$$(re^{2\alpha})' = 1 - (8\pi G)r^2 H. \quad (3.18)$$

Recall that $H(r)$ is the function found from $H(x, y, z)$ by using (3.4), but this is unchanged by any duality symmetry of $H(D, B)$, which acts linearly on (q_e, q_m) . It follows that any *electromagnetic-duality invariance of the NLED Hamiltonian implies a corresponding symmetry in the (q_e, q_m) parameter subspace of spherically-symmetric black hole spacetime metrics*. In particular, for self-dual theories the metric only depends on (q_e, q_m) through the $U(1)$ -duality invariant charge

$$Q = \sqrt{q_e^2 + q_m^2}. \quad (3.19)$$

For NLED theories with a $D \leftrightarrow B$ discrete duality symmetry the spacetime metrics for the purely electric and purely magnetic black holes will be the same.

3.1 Asymptotically-flat charged black holes

For what follows it is convenient to define a new function $\mathcal{E}(r)$ by setting

$$re^{2\alpha} = r - 2G[M - \mathcal{E}(r)] , \quad (3.20)$$

where the constant M is a free parameter. The key equation of (3.18) now becomes the following simple equation ODE $\mathcal{E}(r)$:

$$\mathcal{E}'(r) = -(4\pi)r^2H(r) . \quad (3.21)$$

We may omit the integration constant when integrating this equation because the freedom it represents is precisely the freedom to choose the constant M in (3.20). We thus arrive at the relation

$$\mathcal{E}(r) = 4\pi \int_r^\infty dr [r^2H(r)] . \quad (3.22)$$

We are assuming here that there is no constant term in $H(r)$ arising from a constant term Λ in the Hamiltonian function H because the integral would not then be defined. This amounts to the assumption that there is no cosmological constant, as required for the existence of asymptotically flat black-hole spacetimes, in which case the constant M is the ADM mass. We are also anticipating that $\mathcal{E}(r) \rightarrow 0$ as $r \rightarrow \infty$; i.e. that the asymptotic spacetime is not only flat but also empty, which implies that the $r \rightarrow \infty$ limit is also a weak-field limit. This limit is either Maxwell or ModMax for a self-dual NLED. For NLED theories that are not self-dual there may be other conformal weak-field limits but we shall ignore this possibility here.

The function $\mathcal{E}(r)$ has a simple physical interpretation. To see this we observe that (3.15) implies

$$\mathcal{N}\sqrt{\det h} = r^2 \sin \theta , \quad (3.23)$$

which is the standard volume density for a 2-sphere of radius r in Euclidean 3-space. Using this in the formula of (2.14) for electromagnetic energy, now specialised to the radial fields of a spherically-symmetric black hole, we find that the total electromagnetic energy *outside* the sphere of radius r is precisely $\mathcal{E}(r)$. The total electromagnetic energy is therefore

$$\mathcal{E}_{\text{em}} = \lim_{r \rightarrow 0} \mathcal{E}(r) , \quad (3.24)$$

but this is finite only if $\mathcal{E}(r)$ is finite at $r = 0$.

It was shown in [20] that the Einstein field equations imply that $\mathcal{E}(r)$ is a convex function of r if the NLED stress-energy tensor satisfies the SEC. This can be seen as follows: the SEC is a condition on the stress-energy tensor that is equivalent, given the Einstein field equations, to $R_{tt} \geq 0$, but the Einstein field equations of (3.13) imply that

$$R_{tt} = \frac{\mathcal{N}^2 G}{r} \mathcal{E}''(r) , \quad (3.25)$$

and hence that the SEC implies, via the Einstein field equations, that

$$[\mathcal{E}(r)]'' \geq 0. \quad (3.26)$$

The significance of this result is that the SEC is implied by the NLED strong-field causality condition [10]. Convexity of $\mathcal{E}(r)$ is therefore a property of all causal NLED theories; this fact was used in much of the analysis of [20]. It should be appreciated, however, that the SEC does not guarantee strong-field causality; the original Born theory is an example of a NLED theory for which the SEC is satisfied but the strong-field causality condition is violated.

Another important function for our purposes is the “effective charge” function $Q_{\text{eff}}^2(r)$ defined by

$$2\pi Q_{\text{eff}}^2(r) \equiv r \mathcal{E}(r). \quad (3.27)$$

This function is more directly related to the deformation of the RN metric due to NLED interactions, since

$$\mathcal{N}^2 \equiv e^{2\alpha} = 1 - \frac{2GM}{r} + \frac{(4\pi G)Q_{\text{eff}}^2(r)}{r^2}, \quad (3.28)$$

with $Q_{\text{eff}} = Q$ for Maxwell. It is an important function, for our purposes, because causality imposes a very simple condition on it, which can be deduced as follows.

The first-order equation of (3.21) implies the following second-order equation

$$[Q_{\text{eff}}^2(r)]'' = -2r^2[rH' + 4H], \quad (3.29)$$

which can be rewritten, using (3.10), in the form

$$[Q_{\text{eff}}^2]'' = 2r^2\Theta, \quad (3.30)$$

where (we recall) Θ is the trace of the stress-energy tensor; in fact, this equation is the trace of the Einstein field equation of (3.1). Recalling additionally that causality requires $\Theta \leq 0$ (with equality only for conformal theories) we deduce that

$$[Q_{\text{eff}}^2(r)]'' \leq 0, \quad (3.31)$$

where equality holds only for conformal NLED theories. This is equivalent to the statement that the function $Q_{\text{eff}}^2(r)$ is concave, and strictly concave for any non-conformal NLED theory.

The function $Q_{\text{eff}}^2(r)$ must also approach a constant as $r \rightarrow \infty$, which is Q^2 if the weak-field limit is Maxwell. If the weak-field limit is ModMax then the constant is less than Q^2 . To see this we observe that the Hamiltonian function H_{MM} of (2.38) reduces for $p = 0$ (i.e. static fields) to

$$H_{MM} = e^{-\gamma}s \quad (p = 0), \quad (3.32)$$

which yields (for spherically-symmetric black holes)

$$r^2 H_{MM}(r) = e^{-\gamma} \frac{Q^2}{2r^2}, \quad (3.33)$$

and hence

$$Q_{\text{eff}}^2(r) = e^{-\gamma} Q^2 \leq Q^2, \quad (3.34)$$

where the inequality follows from the $\gamma \geq 0$ requirement for causality [29]. The only effect of the conformally invariant interactions of ModMax is to reduce the value of the Maxwell charge.

Moving on to non-conformal NLED, we observe that since the function $Q_{\text{eff}}^2(r)$ is strictly concave and approaches a constant as $r \rightarrow \infty$, this constant must be its maximum value, which it must approach monotonically. This can be seen more directly by integration of (3.30) subject to the boundary condition that the left-hand side approaches zero as $r \rightarrow \infty$:

$$[Q_{\text{eff}}^2(r)]' = -8\pi G \int_r^\infty dr r^2 \Theta > 0. \quad (3.35)$$

We shall use this monotonicity property later to obtain several significant restrictions imposed by causality on black hole solutions of generic NLED theories⁵.

To summarise: we have introduced two functions in this subsection. A strictly convex function $\mathcal{E}(r)$ which equals the electromagnetic energy outside a sphere of radius r in a spherically-symmetric black-hole spacetime, and a strictly concave function $Q_{\text{eff}}^2(r)$, which is asymptotic to Q^2 (for a Maxwell weak-field limit) as $r \rightarrow \infty$. Both functions are useful in an analysis of the global structure of static black-hole solutions, and much can be learned about the implications of causality from their behaviour near $r = 0$. This is the topic of the following subsection.

3.2 Near-singularity expansions

As we have seen, the Hamiltonian function H of any particular NLED yields, in the context of spherically-symmetric black holes, a function $H(r)$ that appears on the right-hand side of the Einstein field equations. In a variety of simple examples the leading term of this function as $r \rightarrow 0$ is a power of r :

$$H(r) \sim r^{-4\nu}, \quad (3.36)$$

for some number ν , with $\nu = 1$ for Maxwell. While most of our results follow directly from causality, it is nevertheless useful to understand how specific properties of static charged black holes depend on the value of ν in those cases for which (3.36) is valid.

⁵A monotonicity property for the mass-to-charge ratio of *extremal* black holes was derived from causality in [18] but it is unclear to us how, or whether, it is related to the more general monotonicity property established here for $Q_{\text{eff}}^2(r)$.

For example, using (3.36) in (3.21) we see that

$$[\mathcal{E}(r)]' \sim -r^{2(1-2\nu)} \Rightarrow [\mathcal{E}(r)]'' \sim -(1-2\nu)r^{1-4\nu}, \quad (3.37)$$

and hence that convexity of $\mathcal{E}(r)$ requires

$$\nu \geq \frac{1}{2}, \quad (3.38)$$

which is therefore necessary for causality.

Integrating the expression of (3.37) for $\mathcal{E}'(r)$ we get

$$r \rightarrow 0 : \quad \mathcal{E}(r) \sim \begin{cases} c_1 - r^{3-4\nu} & \nu < \frac{3}{4} \\ -\ln r & \nu = \frac{3}{4} \\ r^{-(4\nu-3)} & \nu > \frac{3}{4} \end{cases} \quad (3.39)$$

Multiplicative constants have been ignored here, except for their signs, and the integration constant c_1 is included only when it is the leading term if non-zero. Notice that $\mathcal{E}(r)$ is singular at $r = 0$ unless $\nu < \frac{3}{4}$. This implies that the total electromagnetic energy $\mathcal{E}_{\text{em}} := \mathcal{E}(0)$ is finite iff

$$\nu < \frac{3}{4}. \quad (3.40)$$

This includes $\nu = \frac{1}{2}$ and hence Born-Infeld.

Next, we use (3.39) to get the following expressions for $Q_{\text{eff}}^2(r)$:

$$r \rightarrow 0 : \quad Q_{\text{eff}}^2(r) \sim \begin{cases} c_1 r - r^{4(1-\nu)} & \nu < \frac{3}{4} \\ -r \ln r & \nu = \frac{3}{4} \\ r^{4(1-\nu)} & \nu > \frac{3}{4} \end{cases} \quad (3.41)$$

Taking two derivatives yields

$$r \rightarrow 0 : \quad [Q_{\text{eff}}^2(r)]'' \sim \begin{cases} -r^{-2(2\nu-1)} & \nu < \frac{3}{4} \\ -r^{-1} & \nu = \frac{3}{4} \\ -r^{-2(2\nu-1)} & 1 > \nu > \frac{3}{4} \\ 0 & \nu = 1 \\ r^{-2(2\nu-1)} & \nu > 1 \end{cases} \quad (3.42)$$

From this we see that concavity of $Q_{\text{eff}}^2(r)$, and hence causality, requires

$$\nu \leq 1, \quad (3.43)$$

with equality only for conformal theories. The case $\nu = 1$ includes Maxwell theory but also special interacting theories with a similar behavior near $r = 0$. In this case inspection of (3.41) shows that Q_{eff}^2 is generically a non-negative constant Q_0^2 at $\nu = 1$, and (3.35) implies $Q_0^2 \leq Q^2$, with equality only in the Maxwell case.

An obvious implication of (3.27) is

$$\mathcal{E}_{\text{em}} < \infty \Rightarrow Q_{\text{eff}}^2(0) = 0. \quad (3.44)$$

The above expressions for these functions near $r = 0$ are of course consistent with this implication but they also show that $Q_{\text{eff}}^2(0) = 0$ when $\frac{3}{4} \leq \nu < 1$, for which \mathcal{E}_{em} is not finite. Thus, $Q_{\text{eff}}^2(0) = 0$ does not imply finite \mathcal{E}_{em} .

To summarise: a necessary condition for causality is (1.5), i.e.

$$\frac{1}{2} \leq \nu \leq 1, \quad (3.45)$$

and $\nu < \frac{3}{4}$ is required for finite \mathcal{E}_{em} . We show in subsection 3.4 that there exists a causal self-dual NLED theory for every value of ν in the above range.

3.2.1 Born-Infeld

Since Born-Infeld is an important special case we give below some details of the functions $H(r)$, and $\mathcal{E}(r)$ and $Q_{\text{eff}}^2(r)$ for this case, including both the large r and small r expansions. We start from the Born-Infeld Hamiltonian function:

$$H_{\text{BI}} = \sqrt{(T+2x)(T+2y)-z^2} - T \equiv \sqrt{T^2 + 2Ts + p^2} - T. \quad (3.46)$$

Setting $p = 0$ leads to

$$H_{\text{BI}}(r) = T \left\{ \sqrt{1 + \frac{Q^2}{Tr^4}} - 1 \right\}, \quad (3.47)$$

and expanding for large r we find that

$$r^2 H_{\text{BI}}(r) = \frac{Q^2}{2r^2} \left\{ 1 - \frac{Q^2}{4Tr^4} + \mathcal{O} \left[\left(\frac{Q^2}{Tr^4} \right)^2 \right] \right\}. \quad (3.48)$$

This yields the following results:

$$\mathcal{E}(r) = \frac{2\pi Q^2}{r} \left\{ 1 - \frac{Q^2}{20Tr^4} + \mathcal{O} \left[\left(\frac{Q^2}{Tr^4} \right)^2 \right] \right\}, \quad (3.49)$$

and

$$(4\pi G)Q_{\text{eff}}^2(r) = (4\pi G)Q^2 \left\{ 1 - \frac{Q^2}{20Tr^4} + \mathcal{O} \left[\left(\frac{Q^2}{Tr^4} \right)^2 \right] \right\}. \quad (3.50)$$

As expected we have a deformation of the RN metric with $Q_{\text{eff}}^2(r) < Q^2$.

Expanding (3.47) for small r we find that

$$r^2 H_{\text{BI}}(r) = -Tr^2 + \sqrt{T}Q \left\{ 1 + \mathcal{O} \left(\frac{Tr^4}{Q^2} \right) \right\}, \quad (3.51)$$

which yields

$$\mathcal{E}(r) = \mathcal{E}_{\text{em}} + \frac{4\pi}{3}Tr^3 - 4\pi\sqrt{T}Qr \left\{ 1 + \mathcal{O} \left(\frac{Tr^4}{Q^2} \right) \right\}. \quad (3.52)$$

The second term represents minus the total energy inside a sphere of radius r , contributed by the constant $-T$ term in the Hamiltonian (the role of the second term is to subtract the extra energy in \mathcal{E}_{em} to give rise to $\mathcal{E}(r)$, which contains the electromagnetic energy from r to infinity). The higher corrections come from inverse powers of s in the expansion of the Hamiltonian at large s . The effective charge then has the behavior

$$(4\pi G)Q_{\text{eff}}^2 = (2G\mathcal{E}_{\text{em}})r + \frac{8\pi}{3}GTr^4 - (8\pi G)\sqrt{T}Qr^2 \left\{ 1 + \mathcal{O}\left(\frac{Tr^4}{Q^2}\right) \right\}. \quad (3.53)$$

The precise value of \mathcal{E}_{em} is obtained by integration from 0 to infinity, but it is finite because the integrand of the integral defining it in (3.24) is sufficiently well-behaved as $r \rightarrow 0$ (see (4.15) below).

3.3 Electric field and energy

Born's original NLED theory has the feature that it puts an upper bound on the strength $|E|$ of the electric field; specifically $|E|^2 \leq T$. This is also true of the Born-Infeld theory, although the maximum value is larger for non-zero magnetic field. This property is potentially of importance to charged black holes since the electric field strength blows up at the central singularity of the RN black-hole solution to the Einstein-Maxwell equations, whereas this does not occur if Maxwell electrodynamics is replaced by Born-Infeld. We now aim to investigate the behaviour of the electric field near the central singularity of an electrically charged black hole for a generic NLED theory.

For zero magnetic field, the expression of (2.12) for D as the partial derivative of $\mathcal{L}(S, P)$ with respect to the electric field E reduces to

$$D^i = \frac{\sqrt{\det h}}{\mathcal{N}} h^{ij} L_S E_j, \quad (3.54)$$

where now $S = \frac{1}{2}\mathcal{N}^{-2}|E|^2$ and $P = 0$. For a static and spherically-symmetric electric field configurations in the general black hole spacetime metric of (3.12), we find that

$$\frac{D^r}{r^2 \sin \theta} = L_S E_r, \quad S = \frac{1}{2}E_r^2. \quad (3.55)$$

Since spherical symmetry requires $D^r = q_e \sin \theta$, we have

$$E_r = \frac{q_e}{r^2 L_S}, \quad (3.56)$$

and taking a derivative with respect to r on both sides one finds that

$$[L_S + 2SL_{SS}] \frac{dE_r}{dr} = -\frac{2q_e}{r^3}. \quad (3.57)$$

We may assume here, without loss of generality, that $q_e > 0$, in which case $E_r \geq 0$.

Two necessary conditions for causality of any NLED theory are [3]

$$L_S > 0, \quad L_{SS} \geq 0. \quad (3.58)$$

However, since $L_S \geq 1$ in the conformal weak-field limit (with equality when this is Maxwell), L_S must satisfy the slightly stronger causality conditions⁶

$$L_S \geq 1, \quad L_{SS} \geq 0, \quad (3.59)$$

with equality (for a Maxwell weak-field limit) as $S \rightarrow 0$ (i.e. $r \rightarrow 0$ in the current black-hole context). Both pairs of conditions imply that L_S is a positive non-decreasing function of S , but its minimal value at $S = 0$ is more strongly constrained by (3.59).

The implications for the electric field are illustrated in Fig. 1. For example, E_r is a monotonically decreasing function of r in any causal NLED; this follows immediately from (3.57) since causality requires $(L_S + 2SL_{SS}) > 0$ when $B = 0$. In fact, $(L_S + 2SL_{SS}) > 1$ for positive S , and this has the further implication that the graph of the function $E_r(r)$ for any non-conformal causal NLED lies below this graph for Maxwell (or ModMax if that is the weak-field limit) except asymptotically as $r \rightarrow \infty$ (where they meet). This implies that the rate of increase of E_r as $r \rightarrow 0$ is less than it is for Maxwell; in particular, it can happen that E_r remains finite as $r \rightarrow 0$.

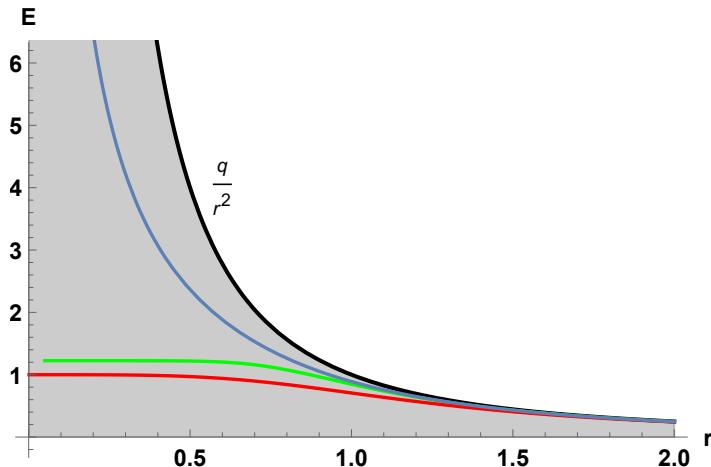


Figure 1: The electric field $E_r(r)$ for Maxwell (black), Born-Infeld (red) and two other examples of causal theories with $E_r \sim 1/r$ (blue) and $E_r \rightarrow E_0$ (green). For any causal NLED, $E_r(r)$ is a monotonically decreasing function of r that is less than Q/r^2 for all finite r , and hence lies entirely in the grey-shaded area.

A Hamiltonian version of the formula (3.56) for the electric field can be found by taking (2.17), with $B = 0$, as the starting point. Instead of (3.54) we now have

$$E_i = \left(\frac{\mathcal{N}}{\sqrt{\det h}} \right) h_{ij} H_x D^j, \quad (3.60)$$

⁶Recall that these have been derived for $B = 0$. In fact, they are the causality conditions of (2.28) for self-dual NLED generalized (as necessary causality conditions) to generic NLED.

which leads (since $H_x = H_s$ for $y = z = 0$) to

$$E_r = H_s \frac{q_e}{r^2}, \quad (3.61)$$

where now $s = \frac{1}{2}(D^r)^2$. A comparison of this result with (3.56) shows that (cf. (2.37))

$$L_S H_s = 1, \quad (3.62)$$

which is essentially the relation of (2.37) (expressed there in terms of derivatives of the CH functions of self-dual NLED theories).

As for self-dual NLED in the previous subsection, if $H(s) \sim s^\nu$ then $H_s \sim s^{\nu-1}$ and we can convert this into a function of r using (3.65). This leads to the following behaviour of E_r near the central singularity:

$$E_r(r) \sim \frac{1}{r^{2(2\nu-1)}}. \quad (3.63)$$

Given the $\nu \geq \frac{1}{2}$ restriction required by causality, the electric field will remain finite as $r \rightarrow 0$ iff $\nu = \frac{1}{2}$, even though (as we saw above) the total electromagnetic energy remains finite provided $\nu < \frac{3}{4}$. The Born-Infeld theory illustrates this since, in this special case,

$$E_r = \frac{q_e \sqrt{T}}{\sqrt{Tr^4 + q_e^2}} \rightarrow \sqrt{T}. \quad (3.64)$$

In contrast, for $\frac{1}{2} < \nu < \frac{3}{4}$, i.e. the remainder of the range of the exponent ν for which the electromagnetic energy is finite, $E_r(r)$ blows up as $r \rightarrow 0$. This is a reminder that finite energy does *not* require a finite electric field at $r = 0$. Finally, for $\frac{3}{4} \leq \nu < 1$ both $E_r(r)$ and $\mathcal{E}(r)$ blow up as $r \rightarrow 0$ (as also occurs in the $\nu = 1$ limit, which includes the Maxwell case).

3.4 Black holes for self-dual NLED

For self-dual NLED it is simple to find the static black-hole function $H(r)$ from the Hamiltonian function H because the latter must be a function of (s, p) only and $p = 0$ for static field configurations. Thus, H can be viewed as a function of s only, which we call $H(s)$, and then $H(r)$ is found by the substitution

$$s = \frac{Q^2}{2r^4}. \quad (3.65)$$

We see from this that the $s \rightarrow \infty$ limit of $H(s)$ corresponds to the $r \rightarrow 0$ limit of $H(r)$. If we now suppose that

$$H(s) \sim s^\nu \quad (s \rightarrow \infty), \quad (3.66)$$

then we find the power-law behaviour of $H(r)$ postulated in (3.36).

We summarised in subsection 2.2 the construction of both the Lagrangian and Hamiltonian functions of Lorentz-invariant self-dual NLED theories in terms of one-variable CH-functions, and we explained how the causality constraints on generic NLED

theories then reduce to much simpler constraints on these functions. In particular, the Hamiltonian function $H(s, p)$ is expressed in terms of the Hamiltonian CH-function \mathfrak{h} such that

$$\mathfrak{h}(s) = H(s, 0) \equiv H(s). \quad (3.67)$$

This allows us to rewrite the causality constraints of (2.33) as

$$0 < H_s \leq 1, \quad H_{ss} \leq 0, \quad H_s + 2sH_{ss} > 0. \quad (3.68)$$

Assuming the power-law behaviour of (3.66), the first two inequalities are satisfied if $0 < \nu \leq 1$, and the third one then requires $\nu \geq \frac{1}{2}$, so we again recover the causality restriction of (3.45) from the large s behaviour of H .

Now consider the following choice of Hamiltonian CH-function [11]:

$$\mathfrak{h}(\sigma) = T \left\{ \left(1 + \frac{\sigma}{\nu T} \right)^\nu - 1 \right\} = \sigma + \mathcal{O}(\sigma^2). \quad (3.69)$$

Notice that the $\nu = 1$ case is Maxwell and the $\nu = \frac{1}{2}$ case is Born-Infeld. This CH-function yields, via (2.32), the Hamiltonian function $H(s, p)$ of a Lorentz-invariant self-dual NLED theory, with

$$H(s) \equiv \mathfrak{h}(s) \sim s^\nu \quad (s \rightarrow \infty). \quad (3.70)$$

A necessary condition for causality of the NLED defined by $H(s, p)$ is therefore the constraint (3.45) on the range of ν . A very special property of the particular self-dual NLED defied by the Hamiltonian CH-function (3.69) is that the full causality conditions of (3.68) are *equivalent* to the restriction on the range of ν .

In this sense, (3.69) defines a ‘minimal’ causal self-dual NLED theory. There are no further causality conditions to satisfy, so we have *for each allowed value of ν an explicit causal NLED theory for which $H(s) \sim s^\nu$ as $s \rightarrow \infty$* . This is important because it shows that the power-law behaviour that we assumed for the “charged black-hole” function $H(r)$ as $r \rightarrow 0$ is actually realised in a simple model for all values of ν allowed by causality.

3.4.1 A Lagrangian perspective

We have now seen how the behaviour of $H(r)$ as $r \rightarrow 0$ is related to the behaviour of $H(s)$ as $s \rightarrow \infty$ or, equivalently, to the behaviour of the Hamiltonian CH-function $\mathfrak{h}(\sigma)$ as $\sigma \rightarrow \infty$, but how is it related to the Lagrangian CH-function $\ell(\tau)$? We summarised in subsection 2.2 the relation between the Hamiltonian and Lagrangian CH-functions. Using the relations (2.35) and/or (2.36), we find that, as $\sigma \rightarrow \infty$,

$$\ell(\tau) \sim (2\nu - 1)\sigma^\nu, \quad \tau \sim \sigma^{2\nu-1}, \quad (3.71)$$

We see again that the $\nu = \frac{1}{2}$ case is special, so we postpone discussion of it.

For $\nu > \frac{1}{2}$ we have

$$\sigma \rightarrow \infty \Rightarrow \tau \rightarrow \infty \Rightarrow \ell(\tau) \sim \tau^\beta, \quad (3.72)$$

where ν and β related by $2\beta\nu = \beta + \nu$. For example

$$\beta = \nu/(2\nu - 1) = \begin{cases} \beta = 1 & \nu = 1 \\ \beta = \frac{3}{2} & \nu = \frac{3}{4} \\ \beta \rightarrow \infty & \nu \rightarrow \frac{1}{2} \end{cases} \quad (3.73)$$

From this we see that the electric field diverges as $r \rightarrow 0$ for all values of β , but the electromagnetic energy is finite for $\beta > \frac{3}{2}$. Examples of Lagrangian CH-functions defining causal self-dual NLED theories are

$$\ell(\tau) = T \left\{ \left[1 + \frac{\tau}{\alpha T} \right]^\beta - 1 \right\}, \quad \beta \geq 1. \quad (3.74)$$

The $\nu = \frac{1}{2}$ case (which includes all theories with finite electric field at $r = 0$) is very different. There is some maximum value of τ ($\tau_{\max} > 0$) such that

$$\sigma \rightarrow \infty \Rightarrow \tau \rightarrow \tau_{\max} : \ell(\tau) \rightarrow \ell(\tau_{\max}). \quad (3.75)$$

Examples of Lagrangian CH-functions of this type defining causal self-dual NLED theories are [9]

$$\ell(\tau) = T \left\{ 1 - \left[1 - \frac{\tau}{\alpha T} \right]^\alpha \right\}, \quad 0 < \alpha \leq 1. \quad (3.76)$$

For these examples $\tau_{\max} = \alpha T$, and $\ell(\tau_{\max}) = T$. The Born-Infeld theory is the $\alpha = \frac{1}{2}$ case:

$$\ell_{BI}(\tau) = T \left\{ 1 - \sqrt{1 - \frac{2\tau}{T}} \right\}. \quad (3.77)$$

Black hole solutions for self-dual NLED theories with CH-functions of the form (3.74) and (3.76) have been recently found in [21], where their thermodynamic properties have also been discussed.

3.5 NLED theories with exact Reissner-Nordström black holes

For NLED theories that are not self-dual, the black hole metric will depend separately on the charges q_e and q_m . For the original Born theory, for example, we have

$$H_{\text{Born}} = \sqrt{(T+2x)(T+2y)} - T = \sqrt{\left(T + \frac{q_e^2}{r^4} \right) \left(T + \frac{q_m^2}{r^4} \right)} - T. \quad (3.78)$$

This is *not* $U(1)$ -duality invariant but there is a Z_2 duality invariance group that exchanges x with y , and hence q_e with q_m . In this case, the purely electric ($q_m = 0$) and purely magnetic ($q_e = 0$) black hole spacetimes will be identical, with

$$H_{\text{Born}}(r) = T \left\{ \sqrt{1 + \frac{Q^2}{Tr^4}} - 1 \right\} \quad (q_e q_m = 0), \quad (3.79)$$

where either $Q = q_e$ or $Q = q_m$. This is also identical to (3.47) but only for the special cases of purely electric or purely magnetic charge, as expected because the Born and BI Hamiltonian functions coincide for purely electric or purely magnetic fields.

The coincidence of the BI and Born black hole solutions for $q_e q_m = 0$ does not apply to dyonic ($q_e q_m \neq 0$) black holes. Consider, for example, the following special case:

$$q_e = q_m = Q/\sqrt{2} \quad \rightarrow \quad H_{\text{Born}}(r) = \frac{Q^2}{2r^4}. \quad (3.80)$$

This yields the RN metric! Thus, the RN dyonic black hole with $q_e = q_m$ is an exact solution of the Einstein-Born field equations, but *not* of the Einstein-BI field equations. An important point to appreciate here is that the stability properties of any specific black-hole solution will generally depend on the particular field equations that it solves; we shall argue in the following subsection that the $q_e = q_m$ dyonic black-hole solution of BI is stable but becomes unstable as a solution of the Born theory.

The Born theory is not the only NLED theory for which the Einstein-NLED field equations have exact dyonic RN black-hole. Remarkably, there is an entire class of (Lorentz invariant) NLED theories for which this is true, as we now show. Consider the class of NLED theories defined by $H(x, y)$. Since $H_z = 0$ the Lorentz-invariance condition (2.21) reduces to

$$H_x H_y = 1. \quad (3.81)$$

The solution of this equation can be expressed in terms of the function

$$\hat{H}(x, y, \varphi) = e^{-\varphi} x + e^{\varphi} y + V(\varphi), \quad (3.82)$$

where φ is an auxiliary field and V a (potentially arbitrary) function of it. A function $H(x, y)$ solving (3.81) is now found by eliminating φ by means of the equation $\hat{H}_\varphi = 0$, which is

$$V'(\varphi) = e^{-\varphi} x - e^{\varphi} y. \quad (3.83)$$

Of course, the solution of this equation must be unique, and it will be if $V(\varphi)$ is a strictly monotonically-increasing function for $\varphi \geq \varphi_0$, where φ_0 is the value of φ in the vacuum; i.e. the solution of $V'(\varphi) = 0$. To see that the function $H(x, y)$ constructed in this way solves (3.81) we observe that

$$d\hat{H} = e^{-\varphi} dx + e^{\varphi} dy + \hat{H}_\varphi d\varphi, \quad (3.84)$$

but since φ is found as a function of (x, y) by solving $\hat{H}_\varphi = 0$, we deduce that

$$H_x = e^{-\varphi} \quad H_y = e^{\varphi} \quad \Rightarrow \quad H_x H_y = 1. \quad (3.85)$$

When $V(\varphi)$ is an even function of φ , i.e. $V(-\varphi) = V(\varphi)$, the function \hat{H} is invariant under $x \leftrightarrow y$ if we simultaneously take $\varphi \rightarrow -\varphi$. This implies that the Hamiltonian function H resulting from the elimination of φ will be invariant under $x \leftrightarrow y$. In other words, an even function $V(\varphi)$ yields a Hamiltonian function $H(x, y)$ with a Z_2

electromagnetic duality symmetry $D \leftrightarrow B$; Born's theory is a simple example since it corresponds to $V(\varphi) = T(\cosh \varphi - 1)$. For such NLED theories,

$$\hat{H}(x, x) = 2(\cosh \varphi)x + V(\varphi), \quad V'(\varphi) = -2(\sinh \varphi)x, \quad (3.86)$$

and $\varphi = 0$ is obviously a solution of the auxiliary field equation because $V'(\varphi)$ is odd when $V(\varphi)$ is even. If this is the unique solution (which it will be for V with the above mentioned monotonicity property) then

$$H(x, x) = 2x \quad \rightarrow \quad H(r) = \frac{q^2}{r^4} \quad (q = q_e = q_m). \quad (3.87)$$

As for the special case of Born's theory, this leads to the RN metric.

When we recall that Born's theory is acausal for sufficiently strong fields [3, 8], an obvious question is whether any of these 'Born-type' generalisations of it could be causal. The answer is no: they are all similarly acausal. To see why, let us return to the phase-space Lagrangian density of (2.13); for the Hamiltonian function $\hat{H}(x, y; \varphi)$ we have

$$\mathcal{L}' = D^i(E - u \times B)_i - \mathcal{N}\sqrt{\det h} [e^{-\varphi}x + e^{\varphi}y + V(\varphi)]. \quad (3.88)$$

Eliminating φ and then D yields the Lagrangian density $\mathcal{L} = \sqrt{|g|}L$ corresponding to the Hamiltonian function $H(x, y)$, but we may first eliminate D ; its field equation is equivalent to

$$D^i = e^{\varphi} \frac{\sqrt{\det h}}{\mathcal{N}} h^{ij} (E - u \times B)_j, \quad (3.89)$$

and back-substitution yields

$$L = e^{\varphi}S - V(\varphi), \quad (3.90)$$

where S is given by (2.8). We must now solve, for φ as a function of S , the auxiliary field equation

$$e^{-\varphi}V'(\varphi) = S. \quad (3.91)$$

The solution is implicit, generically, but it yields φ as some function of S only, which implies that L is also a function only of S . This confirms Lorentz invariance, but all NLED theories of this type ($L_P \equiv 0$) are known to be acausal because they fail to satisfy the strong-field causality condition found in [3] and applied in various contexts in [9], [10], [8].

A more general construction considers any theory described by the Hamiltonian function of (3.82), which yields the general Lorentz invariant NLED with $H_z \equiv 0$. Now the dyonic RN black hole with $q_m = e^{-2\varphi_0}q_e$ will be an exact solution provided $V'(\varphi) = 0$ has a unique solution at some φ_0 where $V(\varphi_0) = 0$.

3.5.1 Black hole instability for acausal NLED

Both the Born and Born-Infeld theories satisfy the conditions that guarantee causality for *weak* fields; i.e. for fields with energy densities much less than T . However, the

Born theory in Minkowski spacetime allows acausal propagation of perturbations to stationary homogeneous electromagnetic backgrounds that are sufficiently strong [3,8]; specifically acausal propagation is possible when $|D||B| > T$ [8]. In contrast, the propagation of such perturbations for the BI theory are always causal [], and this remains true for any self-dual NLED that is causal for weak fields [9].

These results cannot be applied directly to NLED theories in the spacetime background of a black hole. However, let us consider the wave-propagation of some perturbation of the radial electric and magnetic fields of a spherically-symmetric charged black hole within a spherical shell bounded by the spheres $r = r_0 \pm \epsilon$ for $r_0 \gg r_H$ and $\epsilon \ll r_0 - r_H$. A small-amplitude but high frequency wave-packet centered at $r = r_0$ and propagating in a direction tangent to the radial unit vector will remain within the shell for long enough for it to be approximated as a small-amplitude plane wave moving in the constant static electromagnetic background defined by the fields at $r = r_0$. In this approximation the spacetime metric within some small volume of the shell centered around a choice of radial vector of length r_0 is

$$ds^2 \approx -\mathcal{N}_0^2 dt^2 + \mathcal{N}_0^{-2} dr^2 + r_0^2 d\ell^2(\mathbb{E}^2) \quad [\mathcal{N}_0 = \mathcal{N}(r_0)]. \quad (3.92)$$

This is a Minkowski metric but with rescalings of the coordinates.

Let us consider the RN solution of the Born theory for $q_e = q_m = q$. A small volume of the shell at radius $r = r_0$ in a fixed direction with unit vector \mathbf{n} will have approximately constant and uniform vector densities (D, B) in the direction of \mathbf{n} with scalar magnitudes q/r_0^2 . We may now use the Minkowski spacetime results of [] to find the dispersion relations of small amplitude plane waves in this background for wave vectors \mathbf{k} orthogonal to \mathbf{n} . Taking into account the rescaling of the Minkowski coordinates one finds the following two dispersion relations for the two polarisations (\pm) of waves with angular frequencies ω_{\pm} and wave-vector magnitude k :

$$\omega_{+}^2 = \mathcal{N}_0^2 k^2, \quad \omega_{-}^2 = \mathcal{N}_0^2 \left[\frac{Tr_0^4 - \mathcal{N}_0^{-2} q^2}{Tr_0^4 + \mathcal{N}_0^{-2} q^2} \right] k^2. \quad (3.93)$$

These expressions may be compared with the Minkowski spacetime result obtained in [8]. For Born's theory, a wave propagating in a uniform, static electromagnetic background, has the dispersion relation:

$$\omega_{+}^2 \Big|_{\text{Mink}} = k^2, \quad \omega_{-}^2 \Big|_{\text{Mink}} = \left[\frac{T^2 - B^2 D^2}{(T + B^2)(T + D^2)} \right] k^2. \quad (3.94)$$

where $|B|$ and $|D|$ are here the Minkowski spacetime scalar magnitudes of the constant uniform electromagnetic background. We see that these formulas agree with (3.93) in the limit $\mathcal{N}_0^2 \rightarrow 1$, upon setting $B^2 = D^2 = q^2/r_0^4$.

Returning to (3.93), we see that $\omega_{-}^2 \geq 0$ requires

$$r_0^2 \mathcal{N}_0 \geq \frac{q}{\sqrt{T}}. \quad (3.95)$$

For simplicity, let us focus on the extreme RN metric, for which $\mathcal{N} = 1 - GM/r$. Since $q = Q/\sqrt{2}$ in the current context, and $Q = \sqrt{\frac{G}{4\pi}}M$ for the extremal RN black hole, the above inequality simplifies to

$$r_0^2 - GM r_0 - \frac{GM}{\sqrt{(8\pi G)T}} \geq 0, \quad (3.96)$$

with equality for the critical value r_c of r_0 at which $\omega_- = 0$. This critical value is

$$r_c = \frac{1}{2}GM \left[1 + \sqrt{1 + \frac{4}{GM\sqrt{(8\pi G)T}}} \right]. \quad (3.97)$$

For $r_0 < r_c$ we have $\omega_-^2 < 0$, which indicates an instability of the static electric/magnetic fields against non-spherical perturbations.

As $r_c \rightarrow GM$ in the $T \rightarrow \infty$ limit there is no instability outside the horizon, at $r = GM$, as expected because the $T \rightarrow \infty$ limit of Born is Maxwell. However, for finite T there is an instability with $r_c > GM$, i.e. not “hidden” behind the horizon as one might have thought possible. It is an instability against certain perturbations of the radial fields away from spherical symmetry. This conclusion assumes the validity of the approximations made in arriving at the formula (3.97). We expect these approximations to be valid when $r_c \gg GM$, which occurs when the “gravitational Born length” $1/\sqrt{GT}$ is much greater than the horizon radius; equivalently

$$\ell_{gB} \gg \sqrt{GQ}. \quad (3.98)$$

Comparison with (1.3) shows that the RN black hole is one with very strong fields, but we think it likely that this restriction is simply due to our approximations and that there will always be an instability outside the event horizon for an acausal NLED.

4 Global structure for causal NLED

We now turn to an analysis of how the global structure of static Einstein-NLED black-hole spacetime depends on the detailed properties of the NLED theories, and the causality constraints on these properties. One important issue is whether there is necessarily a spacetime singularity at $r = 0$.

Another important issue is the number of Killing horizons of the timelike vector field, ∂_t . These are hypersurfaces of constant r , with the constants given by the positive zeros of the function $g_{tt}(r)$. There are examples of NLED theories for which there are three, or more, Killing horizons [24] but we show here (generalising a result of [22] that was restricted to theories with finite electromagnetic energy) that at most two Killing horizons are possible for any causal NLED.

In the case of two horizons, they can be identified as the event horizon and an interior Cauchy horizon, exactly as for the generic RN black hole. As parameters are varied the

Cauchy horizon may coincide with the event horizon, leading to an extremal charged black hole analogous to the extremal RN black hole. However, it is also possible for the Cauchy horizon to move to $r = 0$; this is actually a parameter-space transition point between a RN type global structure to a Schwarzschild-type global structure, a possibility first noticed by Oliveira [14]. Here we focus on three issues. One is the circumstances for which a transition point exists, another is the geometry at the transition point, and third is an observation on Schwarzschild-type charged black holes.

4.1 Causality implies singularity

For NLED theories defined by a Lagrangian function $L(S)$, it was shown by Bronnikov many years ago that any regular (singularity-free) spherically-symmetric charged black-hole solution of the Einstein-NLED equations can have only magnetic charge [15]. For any spherically-symmetric black-hole solution of the Einstein-NLED equations the scalar S becomes a function of r only, and $L(S)$ therefore becomes a function of r , which we write as $L(r)$. Bronnikov also showed in [15] that a necessary condition for the existence of a regular magnetically charged black-hole solution is that $L(r)$ is finite as $r \rightarrow 0$.

As we now know, all Born-type NLED theories, i.e. those defined by a Lagrangian function $L(S)$ (excepting Maxwell) are acausal [3]. The possibility of regular charged black holes within specific examples in the larger class of NLED theories defined by a Lagrangian function $L(S, P)$ with a weak-field limit was investigated relatively recently by Bokulic et al. [17] but only additional constraints on L were found. Around the same time, Bronnikov extended his earlier results to the same larger class of NLED theories [16]. It remains true that any *regular* spherically-symmetric charged black hole must have magnetic charge only (which immediately excludes the possibility of regular black-hole solutions for any self-dual NLED theory).

For purely magnetic black holes we have

$$S = -\frac{q_m^2}{2r^4}, \quad P = 0, \tag{4.1}$$

and $L(S) := L(S, 0)$ reduces to a function $L(r)$ which is the same as it would be for the simpler NLED theory defined by $L(S)$. In [16] Bronnikov also extends to generic NLED theories⁷ his earlier conclusion that regular magnetic black holes are possible only for NLED theories for which $L(S)$ is finite at $r = 0$. Notice that this condition is equivalent to

$$\lim_{S \rightarrow \infty} L(S) < \infty. \tag{4.2}$$

Here we complete the proof that there are no regular spherically-symmetric black-hole solutions of the Einstein-NLED equations for any causal NLED with a weak-field

⁷Actually, he considers Lagrangian functions of the form $L(S, J)$ where J is a scalar quartic in F , but since $J = 4(2S^2 + P^2)$ the ‘new’ class of NLED theories is the subset of the ‘old’ class defined by $L(S, P)$ that preserve parity, which is a restriction that has no relevance here.

limit by showing that (4.2) cannot be satisfied by any causal NLED theory. The argument is similar to the one presented in [16] for Lagrangian functions $L = L(S)$, but now valid for arbitrary NLED theories defined by any function $L(S, P)$ with a weak-field expansion. For simplicity, we assume in what follows that Maxwell is the weak-field limit. Of course, the full causality conditions on $L(S, P)$ must be satisfied by any causal NLED but it suffices for our purposes to show that any one of these causality conditions is violated if (4.2) is true. We focus on the condition

$$L_{SS} \geq 0, \quad (4.3)$$

where equality holds only for conformal NLED theories. We additionally use the condition $L_S > 0$ but if we assume a Maxwell weak-field limit then $L_S(0) = 1$ (in our conventions, which involve a unit speed of light) and hence

$$L_S \geq 1, \quad L_{SS} > 0, \quad (4.4)$$

for all non-conformal NLED theories, where $L_S = 1$ only in the $S \rightarrow 0$ limit. These are the same causality conditions that we derived earlier for zero magnetic field⁸ but now we see that they also apply for zero electric field. The main difference is that for $E = 0$ we have $S \leq 0$.

To summarise: the function $L(S)$ appearing in (4.2) is finite at $S = 0$ and increases as S increases. For $S < 0$ this implies that $L(S)$ decreases as $|S|$ increases. Moreover, the rate of decrease is positive, which implies that $L(S) \rightarrow -\infty$ as $S \rightarrow -\infty$. It follows that the condition (4.2) cannot be satisfied, and therefore *there are no purely magnetic spherically symmetric charged black hole solutions of the Einstein-NLED equations for any causal NLED with a weak-field expansion*.

Notice that we have not assumed asymptotic flatness here, which would be equivalent to the assumption of zero cosmological constant. The irrelevance of the cosmological constant in this context is explained briefly in [16]. From our perspective, the reason is that the addition of a cosmological constant Λ is equivalent to the addition of Λ to the Hamiltonian function appearing in the key Einstein-NLED equation of (3.18). The effect of this is to add an r^2 term to the metric coefficient g_{tt} (as we shall see in subsection 4.3). This has no effect on the behaviour of the metric as $r \rightarrow 0$.

The above result, taken together with the no-go result of Bronnikov for regular black holes with non-zero electric charge, settles the issue of the existence of regular charged black holes for any causal NLED theory with a weak-field expansion: they do not exist.

⁸For $B = 0$ the inequalities (4.4) are equivalent to the causality conditions of (2.28) since, as explained in the Introduction, the “CH-function” ℓ arising in the construction of Lagrangian functions for self-dual NLED has an interpretation for any NLED in terms of L at zero magnetic field.

4.2 Killing horizons

Horizons are determined by the equation $g_{tt}(r) = 0$, i.e.

$$0 = 1 - \frac{2G[M - \mathcal{E}(r)]}{r} = 1 - \frac{2MG}{r} + \frac{4\pi G Q_{\text{eff}}^2(r)}{r^2}. \quad (4.5)$$

For $r \neq 0$ we may rewrite this as

$$(r - MG)^2 = (MG)^2 - 4\pi G Q_{\text{eff}}^2(r). \quad (4.6)$$

The solutions of this equation are the values of r at which the graphs of the following two functions intersect:

$$f_L(r) = (r - MG)^2, \quad f_R(r) = (MG)^2 - 4\pi G Q_{\text{eff}}^2(r). \quad (4.7)$$

The graph of f_L is a parabola with a minimum at $r = MG$. The function f_R is convex because Q_{eff}^2 is concave (for a causal NLED theory). Because it also approaches a constant as $r \rightarrow \infty$, it must decrease monotonically as r increases, starting from $(GM)^2$ for $\nu \neq 1$ (in which case $f_L = f_R$ at $r = 0$) and a smaller value for $\nu = 1$. Therefore, there are at most two $r \neq 0$ intersection points of the graphs of f_L and f_R , and hence *at most two Killing horizons*. The same result was found in [20] but by a different argument that required a restriction to cases with finite energy density. We see now that the result is a completely general consequence of causality.

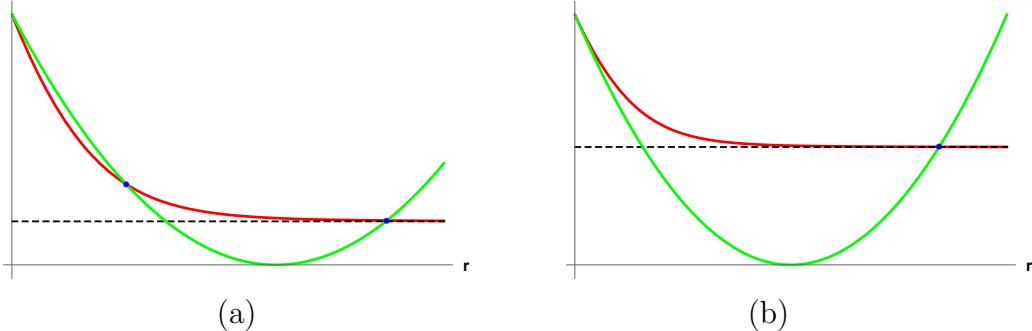


Figure 2: Location of horizons for Born-Infeld at the intersection of $f_L \equiv (r - GM)^2$ (green) and $f_R \equiv (MG)^2 - 4\pi G Q_{\text{eff}}^2(r)$ (red), which is asymptotic to the horizontal dashed line: $f_R(\infty) = (MG)^2 - 4\pi G Q^2$. a) $Q = 1$, $\mu = 1.1$ (here $G = 1$, $T = 1/100$). b) $Q = 0.6$, $\mu = 1.1$. In this case there is no Cauchy horizon and the global structure of the black hole is Schwarzschild type with a space-like singularity.

In the case of two Killing horizons at $r = r_{\pm}$ (with $r_+ \geq r_-$), the horizon at $r = r_+$ is the event horizon and r_- is the (interior) Cauchy horizon. The global structure is similar to that of the RN spacetime. This possibility is illustrated in fig. 2a for Born-Infeld. However, in this Born-Infeld case the Cauchy horizon is absent when the mass M exceeds a critical value, as illustrated in fig. 2b. This phenomenon, which occurs when $r_- \rightarrow 0$, was discussed in [14] and [20].

The Cauchy horizon is also absent in the “extremal” limit $r_+ - r_- \rightarrow 0$. This is a merger of the Cauchy horizon with the event horizon to form a “degenerate” event horizon with zero surface gravity (zero Hawking temperature in the quantum theory). Consider fig. 2a, and how it could change as the mass M is decreased to the point at which the horizontal dashed line (the asymptote for $f_R(r)$ as $r \rightarrow \infty$) has descended to coincide with the r -axis, which is the horizontal tangent to the graph of $f_L(r)$, as illustrated in fig. 3a). As M is further decreased the asymptote to $f_R(r)$ will fall below the r axis and the intersection points will continue to approach until $r_- = r_+$, as illustrated in fig. 3b). At this merger point we have $(GM)^2 < 4\pi GQ^2$. This is a necessary condition for the existence of a merger point because $r_+ - GM$ and $r_- - GM$ have opposite signs when $(GM)^2 > 4\pi GQ^2$.

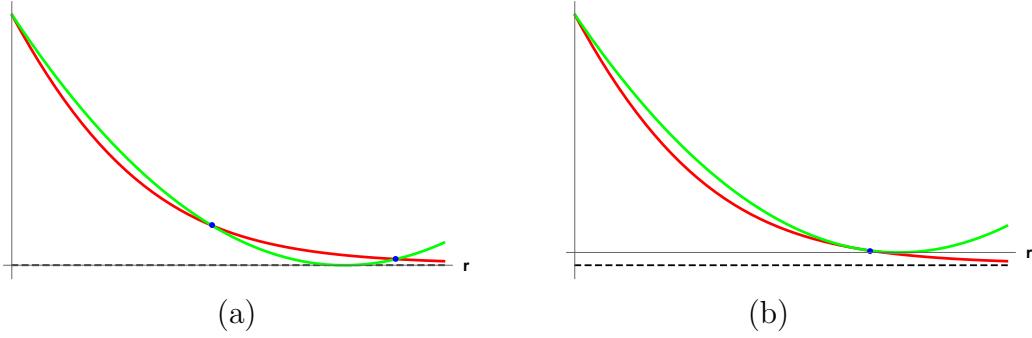


Figure 3: The figures show that $\mu < 1$ for any NLED different from Maxwell theory. a) For $\mu = 1$, the horizons have not yet merged (here $G = 1$, $T = 1/100$, $Q = 0.8$). Extremality requires a lower value of μ . b) Extremal solution, occurring at $\mu = 0.974$ (same conventions as fig. 2).

To summarise: the condition for there to exist an extremal charged black hole is⁹

$$M^2 \leq 4\pi GQ^2, \quad (4.8)$$

and for this extremal solution $M = M_{\text{ext}}$, the mass of the extremal black hole. This puts an upper bound on M_{ext} for fixed charge Q , which we may write as the inequality

$$\mu \leq 1, \quad \mu := \frac{GM_{\text{ext}}}{\sqrt{4\pi G} Q}. \quad (4.9)$$

where equality holds for Maxwell; see fig.3. Notice that μ is a dimensionless parameter that is proportional to the mass-to-charge ratio at extremality.

In the following two subsections we discuss in more detail these two ways in which the Cauchy horizon can disappear.

4.3 The vanishing Cauchy horizon

As the ratio $\sqrt{G} Q / \ell_{gB} = GQ\sqrt{T}$ is decreased (for example, by decreasing the charge), the NLED interactions become relevant near the Cauchy horizon. At this point, the

⁹We allow equality here because this is the Maxwell case.

existence of a Cauchy horizon depends on the theory and on the value of parameters. As parameters vary, it can happen that $r_- \rightarrow 0$. In fact, there is critical hypersurface in parameter space separating two distinct global structures for the black-hole interior. On one side the global structure is similar to that of the RN spacetime and $r_- \rightarrow 0$ as the critical hypersurface is approached. On the other side the global structure is similar to the Schwarzschild spacetime.

To understand when and how this transition occurs, we begin by recalling that a charged black hole with infinite electromagnetic energy \mathcal{E}_{em} always has a Cauchy horizon. This means that we must focus on charged black holes for which \mathcal{E}_{em} is finite. We have already seen that the electromagnetic energy is finite when $H \sim r^{-4\nu}$ as $r \rightarrow 0$ for $\frac{1}{2} \leq \nu < \frac{3}{4}$. In particular, this is the case when the electric field E_r is bounded ($\nu = \frac{1}{2}$) but it is also the case in theories where the electric field diverges as¹⁰ $E_r \sim r^{-k}$ with $0 < k < 1$.

We consider first the $\nu = \frac{1}{2}$ case. In the BI case we have

$$r \rightarrow 0 : \quad H(r) \sim \frac{\sqrt{T}Q}{r^2} - T, \quad (4.10)$$

where the constant $-T$ here arises from the $-T$ term in $H(s)$ that ensures zero vacuum energy as $r \rightarrow \infty$; it leads to the r^3 term in the expansion of $\mathcal{E}(r)$ of (3.52), and this is the leading non-linear term in this expansion. If we allow for a cosmological constant then its effect on the small- r expansion of $\mathcal{E}(r)$ is to change the coefficient of the r^3 term.

On dimensional grounds, we expect (4.10) to apply to any causal NLED theory that has $\nu = \frac{1}{2}$ charged black holes, irrespective of the charges (q_e, q_m) except that $Q(q_e, q_m)$ is generically some homogeneous function of degree-1. Generically, we should also expect factors of positive dimensionless functions of any additional dimensionless parameters multiplying each term in the expansion. Using (4.10) in (3.21) we now find the following small- r expansion of $\mathcal{E}(r)$:

$$\mathcal{E}(r) = \mathcal{E}_{\text{em}} - b_1 \sqrt{T} Q r + b_2 r^k + \dots \quad (4.11)$$

where $b_1 > 0$ and $k \geq 3$. In general we expect $k = 3$ and $b_2 > 0$ for the reasons given above, but b_2 will be negative for a sufficiently large negative vacuum energy, and $k > 3$ if the constant term in the small- r expansion of $H(r)$ is absent. This yields the following expansion small- r expansion of $-g_{tt}$:

$$1 - \frac{2G[M - \mathcal{E}(r)]}{r} = -\frac{2G(M - \mathcal{E}_{\text{em}})}{r} + \left(1 - \frac{Q}{Q_{\text{cr}}}\right) + 2Gb_2r^{k-1} + \dots \quad (4.12)$$

where

$$Q_{\text{cr}} = \frac{1}{2b_1 G \sqrt{T}}. \quad (4.13)$$

¹⁰This is (3.63) with $k = 2(2\nu - 1)$.

For $M \approx \mathcal{E}_{\text{em}}$, there is a horizon at

$$r_h \approx \frac{(M - \mathcal{E}_{\text{em}})}{b_1 \sqrt{T} (Q_{\text{cr}} - Q)}, \quad (4.14)$$

provided $r_h > 0$. This can be an event horizon or a Cauchy horizon depending on the region in parameter space. The typical phase diagram is shown in fig. 4 for the case of Born-Infeld black holes. In this case we have [14]

$$\mathcal{E}_{\text{em}} = \frac{2}{3} \Gamma\left(\frac{1}{4}\right)^2 \sqrt{\pi} T^{\frac{1}{4}} Q^{\frac{3}{2}}, \quad Q_{\text{cr}} = \frac{1}{8\pi G \sqrt{T}}. \quad (4.15)$$

At the tricritical point, $r_h = 0$ and

$$\mu = \mu_{\text{tcr}} \equiv \frac{\Gamma\left(\frac{1}{4}\right)^2}{6\sqrt{2\pi}} \approx 0.874. \quad (4.16)$$

Thus, when $\nu = \frac{1}{2}$, extremal black holes only exists with μ in the interval $\mu_{\text{tcr}} \leq \mu < 1$, as shown in fig. 4.

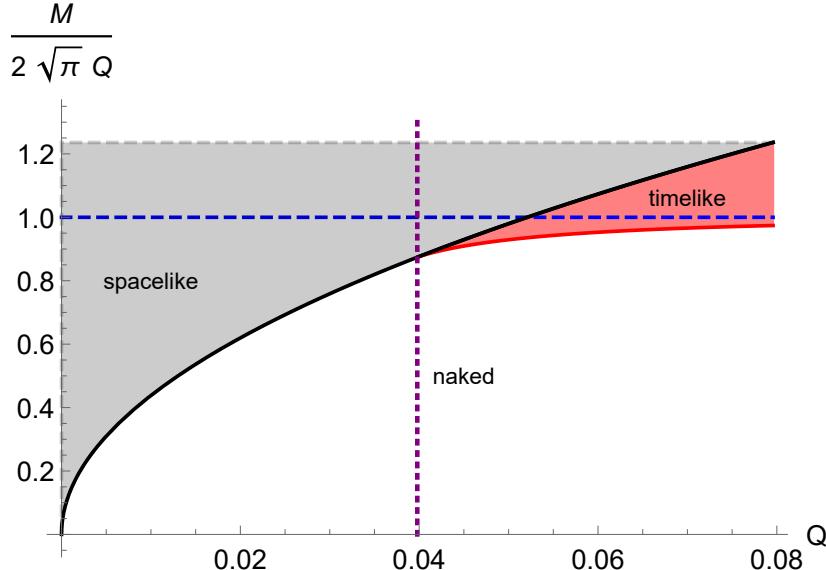


Figure 4: Phase diagram for Born-Infeld black hole, illustrating a $\nu = \frac{1}{2}$ case ($G = 1$).

In general, when $\nu = \frac{1}{2}$, we have the following geometries, according to the parameter regimes:

- | | |
|--|---|
| $Q > Q_{\text{cr}}$, $\mathcal{E}_{\text{em}} > M > M_{\text{ext}}$: | Timelike singularity, horizons at r_+ , r_- . |
| $Q > 0$, $M > \mathcal{E}_{\text{em}}$: | Spacelike singularity, single horizon at r_h . |
| $Q > Q_{\text{cr}}$, $M < M_{\text{ext}}$: | Naked singularity . |
| $Q < Q_{\text{cr}}$, $M < \mathcal{E}_{\text{em}}$: | Naked singularity . |
- (4.17)

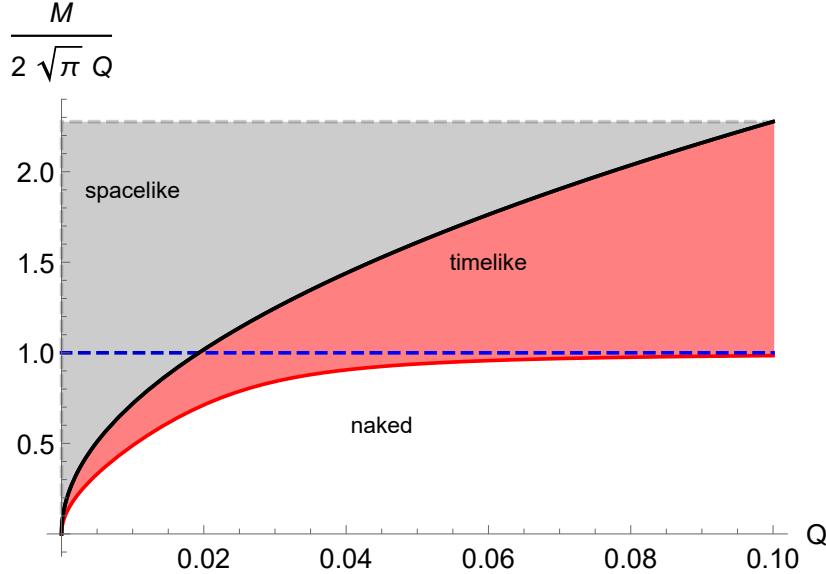


Figure 5: Phase diagram for a charged black hole in a theory with $\nu = \frac{5}{8}$.

We now turn to consider the remaining cases of NLED theories with finite electromagnetic energy, for which $\frac{1}{2} < \nu < \frac{3}{4}$. Now we have

$$\mathcal{E}'(r) \sim -\frac{Q^{2\nu} T^{1-\nu}}{r^{2(2\nu-1)}} \quad \Rightarrow \quad \mathcal{E}(r) - \mathcal{E}_{\text{em}} \sim -c_0 Q^{2\nu} T^{1-\nu} r^{3-4\nu} + \dots, \quad (4.18)$$

where c_0 is a numerical coefficient and the dependence on Q and T is uniquely determined by dimensional analysis. Therefore¹¹

$$1 - \frac{2G[M - \mathcal{E}(r)]}{r} = -\frac{2G(M - \mathcal{E}_{\text{em}})}{r} - \frac{2c_0 G T^{1-\nu} Q^{2\nu}}{r^{4\nu-2}} + 1 + \dots \quad (4.19)$$

Note that the term with coefficient c_0 is of order $1/r^k$, with $0 < k < 1$. The Cauchy horizon disappears for $M > \mathcal{E}_{\text{em}}$, irrespective of the value of the charge, but the geometry has two horizons for $M_{\text{cr}} > M > \mathcal{E}_{\text{em}}$. There is no analog of the critical charge appearing in the $\nu = \frac{1}{2}$ case. The phase diagram is shown in fig. 5 (which illustrates the case $\nu = \frac{5}{8}$). We see that there are RN-type black holes for all values of the charge, and all positive masses starting from $M = 0$, provided $M_{\text{ext}} < M < M_{\text{cr}}$.

As an explicit example, consider the self-dual theory defined by the Hamiltonian CH-function (3.69). In this case, $H(r) = \mathfrak{h}(s(r))$, with $s = Q/2r^4$. The integral (3.22) can be computed explicitly, giving a hypergeometric function. Expanding at small r , we find

$$\mathcal{E}(r) = \mathcal{E}_{\text{em}} - c_\nu \frac{Q^{2\nu} T^{1-\nu}}{r^{4\nu-3}} + \frac{4\pi}{3} T r^3 + \dots \quad (4.20)$$

¹¹It should be noted that the actual form (4.19) of the metric is different from that assumed in [20]. Given the form of the metric (4.19), for $\frac{1}{2} < \nu < \frac{3}{4}$, a Cauchy horizon can exist even for arbitrarily small black hole charges, as illustrated in Fig. 5.

where

$$\mathcal{E}_{\text{em}} = \frac{2^{\frac{5}{4}} \Gamma(\frac{1}{4}) \nu! \sin(\pi\nu) \Gamma(\frac{3}{4} - \nu)}{3 \nu^{\frac{3}{4}}} Q^{\frac{3}{2}} T^{\frac{1}{4}}, \quad c_\nu \equiv \frac{2^{2-\nu} \pi \nu^{-\nu}}{3 - 4\nu}, \quad (4.21)$$

and the term proportional to r^3 has the same origin as the similar term in (3.52). We see that the electromagnetic energy is indeed finite for $\nu < \frac{3}{4}$. For $\nu = \frac{1}{2}$, this expansion gives rise to the terms given in (4.15). For any ν in the interval $\frac{1}{2} < \nu < \frac{3}{4}$, it agrees with the generic expansion (4.18), but now the coefficient c_0 is given explicitly, $c_0 = c_\nu$.

This completes our survey of charged black holes with all possible electric and magnetic charges in all possible causal NLED theories.

4.4 Extremal charged black holes

The extremality condition ($r_\pm = r_h$) implies that the expression of (4.6) has a double zero at $r = r_h$, so its derivative with respect to r must also be zero at $r = r_h$. This gives us the following two equations:

$$r_h = 2G[M_{\text{ext}} - \mathcal{E}(r_h)], \quad 1 = -2G\mathcal{E}'(r_h). \quad (4.22)$$

The first of these equations determines M_{ext} in terms of its horizon radius r_h :

$$M_{\text{ext}} = \frac{r_h}{2G} + \mathcal{E}(r_h). \quad (4.23)$$

Let us rewrite the second equation as

$$F(r_H) = 1, \quad F(r) := -2G\mathcal{E}'(r). \quad (4.24)$$

Because the positive function $\mathcal{E}(r)$ is strictly convex (for causal NLED theories) the negative function $-2G\mathcal{E}(r)$ is strictly concave, and it approaches zero as $r \rightarrow \infty$. Its derivative $F(r)$ is therefore a positive monotonically decreasing function that also approaches zero as $r \rightarrow \infty$. Equivalently, $F(r)$ is a positive function that increases monotonically as r decreases. Therefore, as r is decreased from $r = \infty$ (where $F = 0$) the value of F increases monotonically as $r \rightarrow 0$ and we must distinguish between the following possibilities:

- $F(r) \rightarrow \infty$ as $r \rightarrow 0$. In these cases the equation $F(r) = 1$ will always have a (unique) solution with $r = r_H > 0$; i.e. at the (degenerate) event horizon of an extremal charged black hole.
- $F(0)$ is finite. There are now two subcases which depend on parameter ranges: either $F(0) > 1$ and the equation $F(r) = 1$ has a unique solution $r = r_H$, which is the event horizon of an extremal black hole, or $F(0) \leq 1$ and the equation $F(r) = 1$ has no solution with $r > 0$ and hence there is no extremal black hole.

Recalling the definition of \mathcal{E}' in (3.21), we have

$$F(r) = 8\pi Gr^2 H(r), \quad (4.25)$$

and, assuming the power-law behaviour for $H(r)$ in (3.36), we may use (3.37) to conclude that

$$F(r) \sim r^{2(1-2\nu)} \quad (r \rightarrow 0). \quad (4.26)$$

This shows that $F(r)$ diverges as $r \rightarrow 0$ for all $\nu > \frac{1}{2}$, and more rapidly for larger ν . For electric black holes these are the cases for which the electric field diverges as $r \rightarrow 0$, and we now see that they are also the cases for which there exists an extremal charged black hole for any value of the charge.

In contrast, $F(0)$ is finite for $\nu = \frac{1}{2}$, e.g. Born-Infeld¹², and the existence of an extremal black hole depends on which of the two subcases applies. Let us first consider this issue for Born-Infeld. The function F is

$$F(r) = -8\pi Gr^2 T \left(1 - \sqrt{1 + \frac{Q^2}{Tr^4}} \right) \Rightarrow F(0) = 8\pi G\sqrt{T}Q. \quad (4.27)$$

Therefore, for Born-Infeld there is an extremal black hole iff $8\pi G\sqrt{T}Q > 1$.

If we take T to be fixed (i.e. part of the definition of the ‘theory’) and Q to be variable (since it depends on the choice of black-hole solution) then the distinction between $F(0) \leq 1$ and $F(0) > 1$ becomes a distinction between “small-charge” and “large-charge”, with small-charge being needed to find geometry that differs qualitatively from that of the RN black hole. For Born-Infeld the small-charge to large-charge transition occurs when $8\pi G\sqrt{T}Q = 1$, which is a precise version of the approximate relation $\ell_{gB} \sim \sqrt{G}Q$, which we argued in the Introduction to be the condition for NLED interactions to have a significant effect on the spacetime geometry. For this reason, we expect a small-charge *vs* large-charge distinction to be relevant for generic causal NLED theories, but the precise values of parameters at which abrupt qualitative changes in the geometry occur will depend on the details of the NLED theory.

As noted above, the convexity of $\mathcal{E}(r)$ implies that $F(r)$ is a monotonic decreasing function. In addition, the concavity of $Q_{\text{eff}}^2(r)$ implies that

$$[r^2 F(r)]' \geq 0. \quad (4.28)$$

with equality only for conformal NLED theories; e.g. $r^2 F_{\text{Max}}(r) = 4\pi GQ^2$ for Maxwell. For any non-conformal NLED with Maxwell as the weak-field limit, the inequality of (4.28) is a strict one, except in the $r \rightarrow \infty$ limit. Thus, the graph of $F(r)$ for any non-conformal NLED lies below the analogous graph for Maxwell, for all finite r . This implies that

$$r_h \leq r_h^{\text{Max}}, \quad (4.29)$$

¹²Recall that the parameter ν applies to those black holes for which $H(r)$ has the power-law behaviour of (3.36) as $r \rightarrow 0$, and not to specific NLED theories except in the case of self-dual theories because $H(r)$ cannot then depend on the electric/magnetic charge ratio.

where r_h^{Max} is the horizon radius for an extreme RN black hole, and r_h is the horizon radius for any causal NLED; equality then holds *only* for Maxwell. Since the horizon area is proportional to r_h^2 , a corollary of the inequality (4.29) is that causal NLED interactions reduce (at fixed charge) the horizon area and hence (in the quantum theory) the entropy of an extremal black hole. Thus, *if we compare extremal Einstein-NLED charged black-holes for causal NLED theories at fixed charge, the one with the largest entropy is the extreme RN black hole.*

As an example, let us consider Born-Infeld. In this case $H(r)$ takes the form (3.47), with $Q = \sqrt{q_e^2 + q_m^2}$. Using this in the ‘extremality’ equation $F(r_H) = 1$, we find that

$$r_h^2 = 4\pi G (Q^2 - Q_{\text{cr}}^2) , \quad \sqrt{G}Q_{\text{cr}} := \frac{1}{8\pi\sqrt{GT}} \sim \ell_{gB} . \quad (4.30)$$

We see that a BI extreme black hole exists only if $Q > Q_{\text{cr}}$, and (since $\ell_{gB} = 0$ for Maxwell) that its entropy is less than the extreme RN black hole for fixed charge Q because $Q_{\text{cr}} = 0$ for Maxwell. As we showed earlier, using only the concavity property of Q_{eff}^2 for causal NLED theories, this entropy reduction is a general feature of NLED interactions.

4.5 Deficit solid angle and the Born particle

Here we analyse the spacetime geometry when $M = \mathcal{E}_{\text{em}}$ for $\nu = \frac{1}{2}$ black holes (e.g. those of Born-Infeld). Using (4.12) and setting $M = \mathcal{E}_{\text{em}}$, we find that the spacetime metric takes the following form near $r = 0$:

$$\begin{aligned} ds^2 \approx & - \left(1 - \frac{Q}{Q_{\text{cr}}} + 2Gb_2r^{k-1} + \dots \right) dt^2 \\ & + \left(1 - \frac{Q}{Q_{\text{cr}}} + 2Gb_2r^{k-1} + \dots \right)^{-1} dr^2 + r^2 d\Omega^2 . \end{aligned} \quad (4.31)$$

This metric exhibits a deficit solid angle, because $Q \neq 0$. The point $r = 0$ is the singularity at the tip of a “non-flat cone” because the Ricci curvature scalar is non-zero for $r > 0$:

$$R \approx \frac{2Q}{Q_{\text{cr}} r^2} . \quad (4.32)$$

Depending on the value of Q , the hypersurface $r = \epsilon$ can be spacelike, timelike or null. A vector field normal to this hypersurface is $n = g^{rr}\partial_r$ and its norm squared near (or at) $Q = Q_{\text{cr}}$ is¹³

$$g_{\mu\nu}n^\mu n^\nu = g^{rr} = \begin{cases} (1 - \frac{Q}{Q_{\text{cr}}})^{-1} & \text{if } Q \neq Q_{\text{cr}} \\ 2Gb_2\epsilon^{k-1} \rightarrow 0 & \text{if } Q = Q_{\text{cr}} \end{cases} \quad (4.33)$$

¹³For $Q \neq Q_{\text{cr}}$, the $O(r^{k-1})$ term can be neglected.

We thus have

$$\begin{aligned} Q > Q_{\text{cr}} : \quad & r = \epsilon \text{ is spacelike} \\ Q < Q_{\text{cr}} : \quad & r = \epsilon \text{ is timelike} \\ Q = Q_{\text{cr}} : \quad & r = \epsilon \text{ is null} \end{aligned} \quad (4.34)$$

In terms of appropriately rescaled time and radial coordinates, the $Q \neq Q_{\text{cr}}$ metrics are

$$ds^2 \approx -dt^2 + d\rho^2 + \left(1 - \frac{Q}{Q_{\text{cr}}}\right)\rho^2 d\Omega^2 , \quad \rho \approx 0, \quad Q < Q_{\text{cr}} , \quad (4.35)$$

$$ds^2 \approx dt^2 - d\rho^2 + \left(\frac{Q}{Q_{\text{cr}}} - 1\right)\rho^2 d\Omega^2 , \quad \rho \approx 0, \quad Q > Q_{\text{cr}} . \quad (4.36)$$

The metric (4.35) is known in the literature as the Barriola-Vilenkin (BV) geometry [30], describing a global monopole (see also [31]). The BV metric is

$$ds_{\text{BV}}^2 = -dt^2 + d\rho^2 + (1 - 8\pi G\zeta^2)\rho^2 d\Omega^2 , \quad (4.37)$$

where ζ^2 represents the topological charge of the monopole [31]. This may be compared with the black hole geometry in BI theory near $r = 0$, where Q_{cr} is given in (4.15):

- $M = \mathcal{E}_{\text{em}}$ and $Q < Q_{\text{cr}}$, the metric is (4.37) with the identification $\zeta^2 = \sqrt{T}Q$ and it has a time-like conical singularity (representing a naked singularity of the full metric).
- $M = \mathcal{E}_{\text{em}}$ and $Q > Q_{\text{cr}}$. The metric is given by (4.36), with a spacelike singularity, and it describes the near singularity region of a black hole because the full metric has an event horizon.

Thus, for $Q < Q_{\text{cr}}$ we have a geometry which near $r = 0$ is equivalent to the BV global monopole geometry, with a timelike conical defect. Remarkably, the conical singularity occurs precisely when the ADM mass is given by the total electromagnetic energy. This may be viewed as an implementation of Born's original idea that the mass of a charged point particle should be identified with the energy in its electric field.

Finally, at the tri-critical point where $Q = Q_{\text{cr}}$ (in addition to $M = \mathcal{E}_{\text{em}}$) the $O(r^{k-1})$ term in $\mathcal{E}(r)$ must be included; recall that generically we expect $k = 3$, as this term arises from a residual vacuum energy in the small- r region. The metric has a singularity at $r = 0$:

$$ds^2 \approx -(2Gb_2r^{k-1})dt^2 + (2b_2r^{k-1})^{-1}dr^2 + r^2 d\Omega^2 , \quad r \approx 0 . \quad (4.38)$$

5 Summary and Discussion

In the context of General Relativity (GR), nonlinear theories of electrodynamics allow a natural generalisation of the Einstein-Maxwell equations, and hence of the Reissner-Nordström (RN) charged black-hole solutions. Although this generalisation introduces

a new classical length scale via the “gravitation-Born length” $\ell_{gB} \sim 1/\sqrt{GT}$ associated to the Born tension T , it does so without changing the basic character of the equations (second-order PDE) that one expects of a semi-classical approximation to a quantum theory in which quantum effects become important at the Planck length. However, the propagation of superluminal signals in some backgrounds is still possible, so causality conditions on the NLED Lagrangian or Hamiltonian are necessary. This fact has been appreciated since the late 1960s, and but the necessary and sufficient conditions, which include a novel strong-field causality condition, were found relatively recently [3], and many implications of them are much more recent. Our aim here has been to uncover the implications for charged static black-hole solutions of the Einstein-NLED equations.

A feature of the RN spacetime is a central timelike singularity, hidden behind a Cauchy horizon, which in turn is hidden behind an event horizon. It is generally supposed that this singularity is resolved by quantum effects at the Planck scale ℓ_{Pl} (or possibly string-theory effects at a larger string scale) but there is in principle a possibility that it could be resolved by the nonlinear electromagnetic interactions. An indirect argument to exclude this possibility is that a regular black hole would violate one of the Hawking-Penrose singularity theorems [23] unless the NLED stress-energy tensor violates the SEC, but this is possible only for acausal NLED theories [10].

More direct methods led previously to partial results restricting the type of NLED theory and/or black hole; Bronnikov’s no-go theorem for black holes with electric charge and his condition for the existence of a regular magnetic black hole are notable examples [15, 16]. Using these results we have completed a direct proof that there is no *regular* charged spherically-symmetric black hole solution of the Einstein-NLED equations for any causal NLED theory with a Lagrangian function that has a weak-field expansion.

Although NLED interactions cannot resolve charged black-hole singularities, it has long been known that they can change their character by changing the global structure of the black-hole interior; specifically, for sufficiently large mass-to-charge ratio the interior Cauchy horizon can be absent [14]. The global structure is then similar to that of the Schwarzschild black-hole spacetime. The initial discovery by Oliveira of this possibility was in the context of a one-parameter family of theories generalising BI. The issue was recently revisited by Hale et al. in the context of generic NLED theory [20]. Using the fact that the Strong Energy Condition (SEC) is required for causality [10], and an assumption about the behaviour of the metric near the central singularity, they show that the Cauchy horizon can disappear for sufficiently weak charge, at fixed mass, in a class of theories where the total electromagnetic energy is finite. However, the assumption made is correct only for a special subclass of causal NLED theories. In fact, the electromagnetic energy is finite for a much larger class of theories, which we make explicit, and we give explicit theories in this larger class that are not considered in [20]. This has the important consequence that there exist static charged black holes where the global structure is similar to the Reissner-Nordström black hole for arbitrarily small charge: i.e. there is an interior Cauchy horizon.

Many of the properties of black-hole solutions of the Einstein-NLED equations

arise from special properties of certain functions of the radial coordinate r . One such function is the Hamiltonian expressed as a function of r using the r -dependence of the electric and magnetic fields in the context of a particular black hole solution. This “charged-black-hole” function $H(r)$ determines two other functions that are simply constrained by causality. One is a function $\mathcal{E}(r)$, which is the electromagnetic energy outside a sphere of radius r ; it was introduced (using different notation) in [20] where it was shown to be a strictly convex function as a consequence of the SEC (which is required for causality, as mentioned above).

The other function is an effective charge function $Q_{\text{eff}}^2(r)$ which is asymptotic to Q as $r \rightarrow \infty$ and is required by causality to be concave. This concavity property has been used here to prove that spherically symmetric and asymptotically-flat charged black holes have at most two Killing horizons. This result was proved in [20] under an assumption of finite electromagnetic energy, which is not needed in our proof. We have also used concavity of Q_{eff}^2 to show that causal NLED interactions lead to a reduction in the event-horizon radius of extremal black holes of fixed charge. In the quantum theory this result implies that the entropy of a zero-temperature charged black hole is maximised, at fixed charge, by the Reissner-Nordström black hole.

We have mentioned above that some NLED theories lead to charged black-hole spacetimes without a Cauchy horizon for sufficiently small charge. As shown in [14, 22], as the mass M of an RN-type charged black hole is decreased the radial distance r_- of the inner Cauchy horizon also decreases, and $r_- = 0$ when $M = \mathcal{E}_{\text{em}}$, the total electromagnetic energy. As the mass M is further decreased we have a Schwarzschild-type small-charge black hole. An interesting aspect of the Schwarzschild-type charged black holes concerns the flux lines of the divergence-free vector density fields (D, B) on a constant-t spacelike hypersurface. Recalling that the complete Schwarzschild spacetime has four regions, and that regions I and IV are connected by an Einstein-Rosen bridge across a minimum two-sphere at the horizon radius, we see that the flux lines must pass smoothly through region I into region IV, where they presumably expand out to another spatial infinity. If this is the case then the sign of the charge in region IV is opposite to its sign in region I. This appears to be example of “charge without charge”.

Here we have studied in some detail the nature of the spacetime geometry at the transition point between RN-type and Schwarzschild-type geometry, when $M = \mathcal{E}_{\text{em}}$, and we find that the geometry near $r = 0$ develops a conical singularity. This provides a curious gravitational realization of Born’s idea that a charged particle’s mass may arise entirely from its electromagnetic field [1].

In the quantum theory we expect any non-extremal charged black hole to evaporate by Hawking radiation, implying a gradual reduction of its mass. Thus, we might expect a $Q < Q_{\text{cr}}$ black hole with $M > \mathcal{E}_{\text{em}}$ to evaporate until until $r_h \rightarrow 0$, at which point it becomes a zero-entropy “Born particle” with $M = \mathcal{E}_{\text{em}}$. In reality we should expect quantum gravity effects when r_h reaches the Planck length.

Consider now a $Q > Q_{\text{cr}}$ black hole with $M > \mathcal{E}_{\text{em}}$. Initially, the singularity is spacelike. As the black hole evaporates, it loses mass until M decreases to \mathcal{E}_{em} . At

this critical point, a drastic change in topology occurs: the spacelike singularity transitions into a timelike singularity, purely as a consequence of evaporation. The global structure of the black hole then resembles that of the standard Reissner–Nordström solution, and evaporation continues until the extremal limit is reached, $M = M_{\text{ext}}$. It is striking that a smooth process well-described within the semiclassical approximation, such as evaporation far from extremality and occurring outside the event horizon, can nevertheless precipitate dramatic transformations within the black hole interior!

Acknowledgements

JGR acknowledges financial support from the Spanish MCIN/AEI/10.13039/501100011033 grant PID2022-126224NB-C21.

References

- [1] M. Born, “Modified field equations with a finite radius of the electron,” *Nature* **132** (1933) no.3329, 282.1
- [2] M. Born and L. Infeld, “Foundations of the new field theory,” *Nature* **132** (1933) no.3348, 1004.1
- [3] G. O. Schellstede, V. Perlick and C. Lämmerzahl, “On causality in nonlinear vacuum electrodynamics of the Plebański class,” *Annalen Phys.* **528** (2016) no.9-10, 738-749 [arXiv:1604.02545 [gr-qc]].
- [4] L. Mezincescu, J. G. Russo and P. K. Townsend, “Hamiltonian birefringence and Born-Infeld limits,” *JHEP* **02**, 186 (2024) [arXiv:2311.04278 [hep-th]].
- [5] G. Boillat, “Nonlinear electrodynamics - Lagrangians and equations of motion,” *J. Math. Phys.* **11** (1970) no.3, 941-951
- [6] J. Plebański, “Lectures on non-linear electrodynamics”, (The Niels Bohr Institute and NORDITA, Copenhagen, 1970).
- [7] I. Bialynicki-Birula, “Nonlinear Electrodynamics: Variations on a theme by Born and Infeld”, in *Quantum Theory of Particles and Fields*, eds. B. Jancewicz and J. Lukierski, (World Scientific, 1983) pp. 31-48.
- [8] J. G. Russo and P. K. Townsend, “Born again,” *SciPost Phys.* **16**, no.5, 124 (2024) [arXiv:2401.04167 [hep-th]].
- [9] J. G. Russo and P. K. Townsend, “Causal self-dual electrodynamics,” *Phys. Rev. D* **109**, no.10, 105023 (2024) [arXiv:2401.06707 [hep-th]].

- [10] J. G. Russo and P. K. Townsend, “Causality and energy conditions in nonlinear electrodynamics,” *JHEP* **06**, 191 (2024) [arXiv:2404.09994 [hep-th]].
- [11] J. G. Russo and P. K. Townsend, “Dualities of self-dual nonlinear electrodynamics,” *JHEP* **09**, 107 (2024) [arXiv:2407.02577 [hep-th]].
- [12] J. G. Russo and P. K. Townsend, “Simplified self-dual electrodynamics,” *JHEP* **10** (2025), 120 [arXiv:2505.08869 [hep-th]].
- [13] K. A. Bronnikov, “Comment on ‘Linear superposition of regular black hole solutions of Einstein nonlinear electrodynamics’,” *Phys. Rev. D* **101** (2020) no.12, 128501 [arXiv:1912.03149 [gr-qc]].
- [14] H. P. de Oliveira, “Nonlinear charged black holes,” *Class. Quant. Grav.* **11** (1994), 1469-1482.
- [15] K. A. Bronnikov, “Regular magnetic black holes and monopoles from nonlinear electrodynamics,” *Phys. Rev. D* **63** (2001), 044005 [arXiv:gr-qc/0006014 [gr-qc]].
- [16] K. A. Bronnikov, “Regular black holes sourced by nonlinear electrodynamics,” [arXiv:2211.00743 [gr-qc]].
- [17] A. Bokulić, I. Smolić and T. Jurić, “Constraints on singularity resolution by nonlinear electrodynamics,” *Phys. Rev. D* **106** (2022) no.6, 064020 [arXiv:2206.07064 [gr-qc]].
- [18] Y. Abe, M. Médevielle, T. Noumi and K. Yoshimura, “Causality Constraints on Black Hole Thermodynamics in Nonlinear Electrodynamics,” [arXiv:2505.23483 [hep-th]].
- [19] L. Croney, R. Gregory and C. J. Ramírez-Valdez, “Thermodynamics of dyonic black holes in non-linear electrodynamics,” *JHEP* **10** (2025), 013 [arXiv:2506.06437 [hep-th]].
- [20] T. Hale, R. A. Hennigar and D. Kubizňák, “Excising Cauchy Horizons with Non-linear Electrodynamics,” [arXiv:2506.20802 [gr-qc]].
- [21] H. Babaei-Aghbolagh, K. Babaei Velni, S. He and F. Isapour, “A Unified Causal Framework for Nonlinear Electrodynamics Black Hole from Courant-Hilbert Approach: Thermodynamics and Singularity,” [arXiv:2511.17407 [hep-th]].
- [22] T. Hale, R. A. Hennigar and D. Kubizňák, “Rotating Extremal Black Holes in Einstein-Born-Infeld Theory,” [arXiv:2509.13099 [gr-qc]].
- [23] S. W. Hawking and R. Penrose, “The Singularities of gravitational collapse and cosmology,” *Proc. Roy. Soc. Lond. A* **314** (1970), 529-548.

- [24] C. Gao, “Black holes with many horizons in the theories of nonlinear electrodynamics,” Phys. Rev. D **104** (2021) no.6, 064038 [arXiv:2106.13486 [gr-qc]].
- [25] I. Bandos, K. Lechner, D. Sorokin and P. K. Townsend, “ModMax meets Susy,” JHEP **10** (2021), 031 [arXiv:2106.07547 [hep-th]].
- [26] G. W. Gibbons and D. A. Rasheed, “Electric - magnetic duality rotations in non-linear electrodynamics,” Nucl. Phys. B **454**, 185-206 (1995) [arXiv:hep-th/9506035 [hep-th]].
- [27] M. K. Gaillard and B. Zumino, “Nonlinear electromagnetic selfduality and Legendre transformations,” [arXiv:hep-th/9712103 [hep-th]].
- [28] R. Courant and D. Hilbert, “Methods of Mathematical Physics”, Vol.II (Wiley Interscience, 1962) pp.91-94.
- [29] I. Bandos, K. Lechner, D. Sorokin and P. K. Townsend, “A non-linear duality-invariant conformal extension of Maxwell’s equations,” Phys. Rev. D **102** (2020), 121703 [arXiv:2007.09092 [hep-th]].
- [30] M. Barriola and A. Vilenkin, “Gravitational Field of a Global Monopole,” Phys. Rev. Lett. **63** (1989), 341.
- [31] H. Tan, J. Yang, J. Zhang and T. He, “The global monopole spacetime and its topological charge,” Chin. Phys. B **27** (2018) no.3, 030401 [arXiv:1705.00817 [gr-qc]].