Singularity Theorems

A critical appraisal

José M M Senovilla

UPV/EHU, Bilbao, Spain

Institute of Theoretical Physics, Warsaw, 11th December 2020

Outline

1. Introduction
2. The Penrose singularity theorem
   - Curvature condition
   - Incompleteness
   - Trapped spheres
3. Singularity theorems
   - Ideas behind the theorems
   - Contemporary singularity theorems
4. The devil is in the detail
   - Uncovering the mysteries of black holes
   - Curvature divergence?
   - Singularity-free spacetimes
   - Remarks on extensions
   - BKL or null singularity?
5. Conclusions and final comments

Introduction

Penrose's singularity theorem won the 2020 Nobel prize in Physics!

- It is a well deserved award. The theorem
  1. is sheer beauty
  2. contained novel ingredients along with fruitful ideas
  3. prompted multiple developments in theoretical relativity
  4. involved stunning physical consequences (incompleteness, failure of the theory?)
- In particular, the notion of trapped spheres is fundamental, a key idea in black hole physics, numerical relativity, mathematical relativity, cosmology, gravity analogs, etc.
- Its influence is endless, and its prolific range of applications keeps growing

Modern GR

- As I proposed some time ago (and has been echoed by the Nobel Committee) singularity theorems constitute the first post-Einsteinian content of relativity.
- What had happened before the theorem (deflection of light, universal expansion, Cauchy's problem, ADM formulation, Kerr's solution, Wheeler-de Witt equation, ...)
- as well as most of what came next (background radiation, gravitational lenses, radiation in binary systems, direct observation of waves, etc)
- All of this had been explicitly predicted, known, anticipated, or foreseen in one way or another by Einstein.
- In contrast, the singularity theorems and their consequences were (surely) not even suspected by the founder of GR.

In 1965 GR left adolescence behind, emancipated from its creator, and became a mature physical theory full of vitality and surprises.

Penrose's theorem (1965)

The Penrose singularity theorem

Theorem (Penrose singularity theorem)

If

- *the null convergence condition holds*
- *there is a non-compact Cauchy hypersurface* Σ
- *and a closed future-trapped surface,*

then there are future-incomplete null geodesics.

Two important novelties here:

1. Characterization of singularities by geodesic incompleteness
2. Concept of closed trapped surface

The convergence condition

Focusing of geodesics

- The Raychaudhuri equation

$$v^\nu \nabla_\nu (\nabla_\mu v^\mu) + \nabla_\mu v^\nu \nabla_\nu v^\mu - \nabla_\mu (v^\nu \nabla_\nu v^\mu) + R_{\rho\nu} v^\rho v^\nu = 0.$$

- If v^μ is tangent to a pencil of geodesics emanating either
  1. from a point
  2. orthogonal to a (hyper)surface

  then $v^\nu \nabla_\nu v^\mu = 0$ and $\nabla_\mu v_\nu = \nabla_{(\mu} v_{\nu)}$ and

$$v^\nu \nabla_\nu (\nabla_\mu v^\mu) = -\nabla_\mu v^\nu \nabla_\mu v^\nu - R_{\rho\nu} v^\rho v^\nu \leq 0.$$

- $\nabla_\mu v^\mu|_p < 0$ and $R_{\rho\nu} v^\rho v^\nu \geq 0 \quad \Longrightarrow \quad \nabla_\mu v^\mu \to -\infty$ in finite affine parameter. These are *caustics, or focal points*. *Geodesics stop maximizing the interval if they encounter a focal point.*

Convergence, or energy, condition?

- This is the Focusing effect. One needs the geometric condition

$$R_{\rho\nu}v^{\rho}v^{\nu} \geq 0$$

usually called the (null, timelike) convergence condition.

- In General Relativity, one can relate the Ricci tensor to the energy-momentum tensor T via Einstein's field equations

$$R_{\mu\nu} - \frac{1}{2}R\,g_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}$$

- Thereby, the convergence condition can be rewritten in terms of physical quantities. This is why $R_{\rho\nu}v^{\rho}v^{\nu} \geq 0$, when valid for all time-like v^{μ}, is called the strong energy condition.
- One should bear in mind, however, that this is a condition on the Ricci tensor (a geometrical object)

Incompleteness

A simple example: plane waves

- The line-element of an electromagnetic plane wave propagating in the z direction reads

$$ds^2 = -c^2dt^2 + dz^2 + dx^2 + dy^2 - \frac{4\pi G}{c^4}(x^2+y^2)E^2(ct-z)(cdt-dz)^2$$

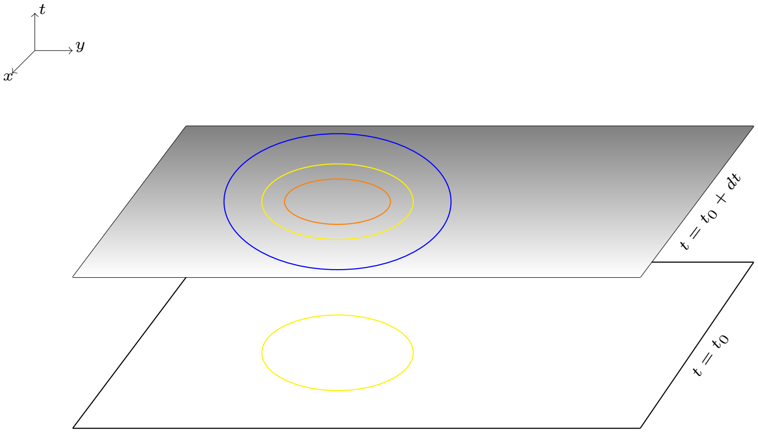
  where E is the electric (or magnetic) field amplitude.
- <u>All</u> curvature scalar invariants vanish.
- Yet, spacetime may be singular if the electric field misbehaves.
- Incompleteness of geodesics is a good way to signal the singularity.
- By the way, these spacetimes (also the geodesically complete ones) do not possess any Cauchy hypersurface (Penrose 1965).

(Now) standard definition of singularity

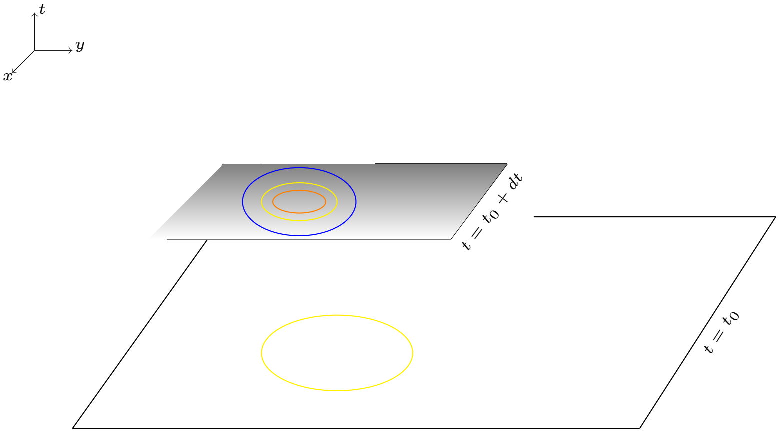
- Spacetime fails at singularities (ergo they are not in spacetime!)
- Penrose's idea was to use physical curves that do belong to the spacetime: note that curves are good pointers.
- Thus, if these curves cannot be continued they are pointing towards a problem: the singularity.
- The curves do not need to be geodesic, and as a matter of fact there are known examples (Geroch) of geodesically complete space-times with incomplete time-like curves of everywhere bounded acceleration.
- It must be remarked, however, that all singularity theorems prove merely the existence of geodesic incompleteness, which of course is a sufficient condition for incompleteness (existence of singularities)

Closed trapped surfaces

Untrapped surfaces, “Normal situation”



Possible trapping in contracting worlds



A little geometry

- Let ζ be any spacelike submanifold (of any co-dimension) in spacetime, and denote by A_ζ its "area, volume, ...".
- Choose an arbitrary vector field $\vec{\xi}$ and deform ζ along its flow. The initial variation of A_ζ due to this deformation is

$$\delta_\xi A_\zeta = \int_\zeta \left(\mathrm{div}\xi^T + H^\mu \xi_\mu\right)$$

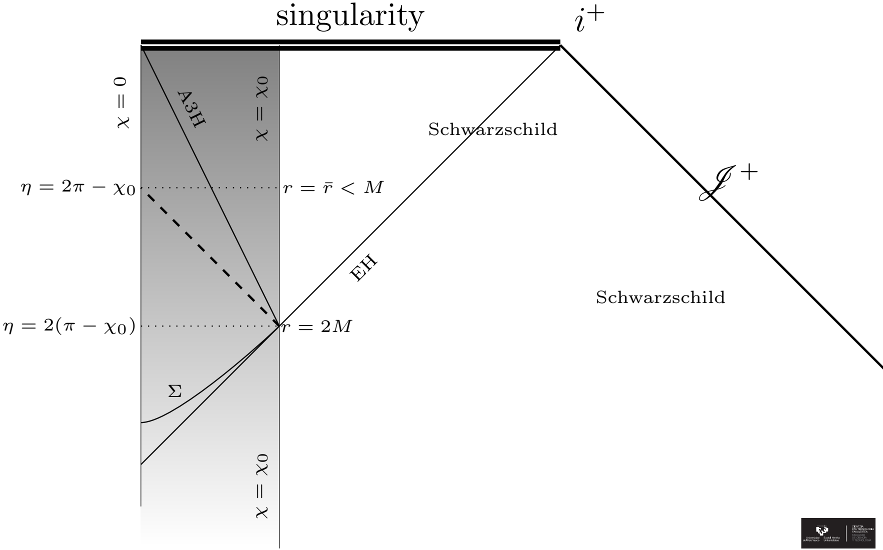
  where H^μ is the mean curvature vector of ζ, that is, the trace of its second fundamental form (or shape tensor). Notice that H^μ is orthogonal to ζ.
- If ζ is compact the first term disappears and one simple has

$$\delta_\xi A_\zeta = \int_\zeta H^\mu \xi_\mu$$

A little geometry

- This is a classical result in Riemannian geometry from where minimal submanifolds are characterized by $H^\mu = 0$. This is the only distinguished case for positive-definite metrics.
- However, in Lorentzian geometry if, say, H^μ is future timelike on ζ then **the variation of A_ζ along any future $\vec{\xi}$ is strictly negative!**
- These are precisely the trapped submanifolds: those having H^μ future timelike. (Idem past-trapped)
- If on the other hand H^μ is future lightlike then ζ is *marginally future trapped*.
- Observation: stationary spacetimes cannot have *compact* trapped submanifolds (Mars-Senovilla 2003)
- For each $\vec{n}$ normal to ζ, $\theta_n := H^\mu n_\mu$ is called expansion along $\vec{n}$. Hence, for a future-trapped ζ all possible future expansions θ_n are negative.

Recall the Oppenheimer-Snyder model



Trapped submanifolds are stable

- The notion of trapped submanifold is independent of coordinates, bases, existence of symmetries or preferred surfaces, etc.
- A decisive point is that its definition is given by *inequalities* and therefore trapped submanifolds are stable.
- The Einstein-Euler field equations describing a perfect fluid in GR are hyperbolic, and thus possess the property of continuous dependence of the solution on the initial conditions.
- The initial conditions of the Oppenheimer-Snyder collapse leads to a trapped sphere within a finite time, hence initial conditions which are sufficiently close to O-S will also lead to the formation of closed trapped surfaces within the same time interval, *regardless of symmetries*.

Classical singularity theorems

A Pattern Singularity Theorem

- Since 1965, many singularity theorems have been proven
- some of them are applicable to cosmological situations, some to star or galaxy collapse, and others to the collision of gravitational waves, to cite prominent situations
- All singularity theorems share a well-defined skeleton, the very same pattern. This is, succintly, as follows

Theorem (Pattern Singularity Theorem)

If a space-time of sufficient differentiability satisfies

1. *a condition on the curvature*
2. *a causality condition*
3. *and an appropriate initial and/or boundary condition*

then there are null or time-like inextensible incomplete geodesics.

The simplest singularity theorem

Theorem (Hawking)

If

1. *there is a Cauchy hypersurface Σ such that*
2. *the trace K of its second fundamental form satisfies $K \geq b > 0$*
3. *and the convergence condition $R_{\rho\nu}v^{\rho}v^{\nu} \geq 0$ holds along the timelike geodesic congruence v^{μ} orthogonal to Σ*

then all timelike geodesics are past incomplete.

The classical Hawking-Penrose singularity theorem

The paradigmatic case was the celebrated Hawking-Penrose theorem (1970), which since then has been considered the singularity theorem *par excellence*.
Comparing it with the Pattern theorem, it reads:

Theorem (Hawking and Penrose)

If the convergence, causality and generic conditions hold and if there is one of the following:

- *a closed achronal set without edge,*
- *a closed trapped surface,*
- *a point with re-converging light cone*

then the space-time is causal geodesically incomplete.

The "causality" and "generic" conditions are standard.

Ideas behind the proofs

- The "curvature condition"
  1. usually referred to as "energy" and "generic" conditions, but as explained above this assumption is of a geometric nature
  2. it is absolutely indispensable: no singularity theorem can be proven without some sort of curvature condition
  3. it enforces the geodesic focusing via the Raychaudhuri equation.
- The causality condition
  1. basically, it ensures the existence of maximal geodesics between any two (causally related) events in appropriate domains of dependence
  2. these maximal geodesics cannot have focal points (caustics).

Ideas behind the proofs

- Recapitulating, one has
  - focusing of all causal geodesics —ergo the existence of caustics and focal points—
  - together with the existence of geodesics of maximal proper time —hence necessarily without focal points— joining **causally related events** of the space-time
- A contradiction starts to glimmer if all geodesics are complete.
- **However**, there is no such contradiction yet!
- this is because we have not enforced a finite upper bound for the proper time of selected families of time-like geodesics (and analogously for null geodesics).
- To get the contradiction with geodesic completeness one needs to add the initial/boundary condition, which happens to be absolutely essential in the theorems.
- Later, I will discuss several examples of physically reasonable singularity-free space-times for illustration.

(Almost) contemporary versions

- Recently the theorems have been proven with the minimum allowable differentiability (Kunzinger, Steinbauer, Vickers, Graf, et al)
- Cosmological cases where local black holes have formed (Vilenkin, Wall). They imply the existence of causally disconnected regions.
- Using mathematical results of volume comparison (Treude, Grant)
- Replacing trapped surfaces with marginally trapped, or *outer* trapped surfaces (Andersson et al, Eichmair et al)
- Suppressing some of the hypotheses to see what happens, and what is the "robust" part of the theorems (Costa e Silva, Flores, Galloway, Vega)
- Relating incomplete geodesics to causal edge points and curvature divergences (Ashley, Whale, Scott).

(Almost) contemporary versions 2

- In higher dimensions, where the topology can be richer (Galloway, Andersson, Cai, ...)
- Using trapped submanifolds of any dimension (Galloway, Senovilla)
- For warped-product spacetimes (Cipriani, Senovilla)
- Weakening the curvature conditions in order to incorporate inflation (Edge-Guth-Vilenkin)
- Ditto to support $\Lambda > 0$ (Galloway)
- Ditto to incorporate quantum corrections (Tipler, Roman, Fewster, Kouton, Galloway, Kuipers, Calmet)
- Taking into account other quantum effects (Ford, Bojowald, Wall)
- Applicable to black holes evaporating by quantum effects –Hawking radiation– (Minguzzi)
- Based on averages (Raychaudhuri, Senovilla)

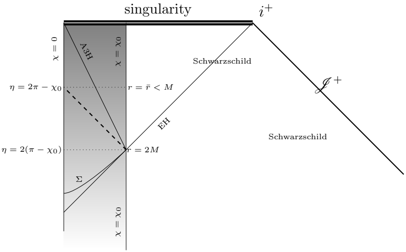
Critical appraisal

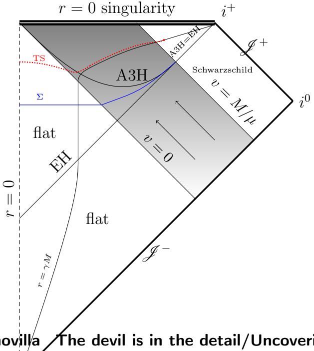
The devil is in the detail

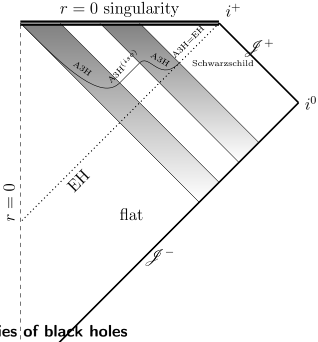
What Penrose's theorem says

- Sometimes the Penrose theorem is interpreted as definite proof that black holes form in gravitational collapse
- The actual fact is more subtle ...
- The assumption of the existence of a trapped sphere does not state anything on the formation of black holes
- rather uncovers (some of) what happens inside the BHs once they are formed!

Black holes cannot be seen, nor felt (outside)







Closed trapped surfaces do not stick out

- There is a theorem (Claudel, Hawking) stating that, in asymptotically flat spacetimes, no closed trapped surface can be seen from $\mathscr{J}^+$
- In other words, they are enclosed beyond the event horizon EH.
- **Therefore, Penrose's theorem informs us of what happens *beyond the horizon*. It is a result about the interior of black holes.**
- It is telling us that, if BHs form, inside them there will probably be closed trapped surfaces and, therefore, incompleteness of the spacetime follows (classically and under certain conditions).
- From the viewpoint of black hole formation the question is quite another: to know if closed trapped surfaces form from innocuous initial data (Christodoulou 2009, Reiterer-Trubowitz, Klainerman-Rodnianski)

Incompleteness = curvature problems?

- What is the relation between geodesic incompleteness and curvature problems, if any?
- Surprisingly, limits on curvature growth can be placed on (maximal) geodesics (Tipler, Newman, Kánnár-Rácz, Szabados).
- If v^μ is the vector tangent to the geodesic, $R_{\alpha\beta\mu\nu}v^\beta v^\mu$ cannot grow more than $(\tau - \hat{\tau})^{-2}$ when approaching a singularity at $\tau = \hat{\tau}$, where τ is the affine parameter.
- There are some (few) partial results that point, in some cases, to the existence of curvature divergences (Clark).
- Recent results seem to indicate that divergence of the curvature scalars is compatible with the spacetime being geodesically complete (Olmo et al).

When spacetimes can be geodesically complete?

- Sometimes the idea is turned around and geodesic completeness is assumed to discern what kind of models support it.
- In the stationary case, for globally hyperbolic spacetimes, it can be shown that if the convergence condition holds then

$$R_{\mu\nu}\xi^{\mu}\xi^{\nu}/(\xi^{\mu}\xi_{\mu}) \sim k/\rho^2$$

  ($\vec{\xi}$=Killing, ρ= appropriate spatial distance between any two events). (Garfinkle-Harris)
- In the dynamical globally hyperbolic case, it can be proven that if the convergence condition holds and the expansion K of a Cauchy hypersurface Σ is positive then at least one of the following three quantities must be non-positive
  - Λ
  - the averaged energy density on Σ
  - minus the averaged scalar curvature of Σ

  (Senovilla)

Singularity-free spacetimes

Devil in the detail: Einstein static universe

The Einstein universe metric is $(\Lambda > 0)$

$$ds^2 = -c^2 dt^2 + \frac{1}{\Lambda}(d\chi^2 + \sin^2\chi d\Omega^2)$$

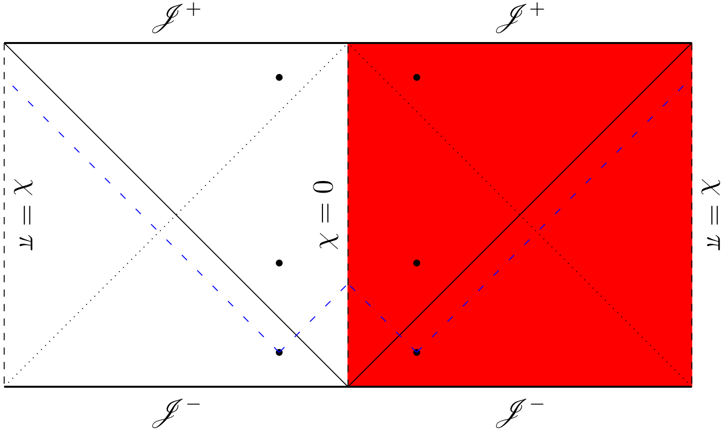
with $8\pi G\varrho = 2c^4\Lambda$. This spacetime

1. Satisfies the convergence condition (SEC)
2. is globally hyperbolic, any $t =$const. hypersurface is a compact Cauchy slice
3. has points with reconverging light cones
4. yet, it is geodesically complete.

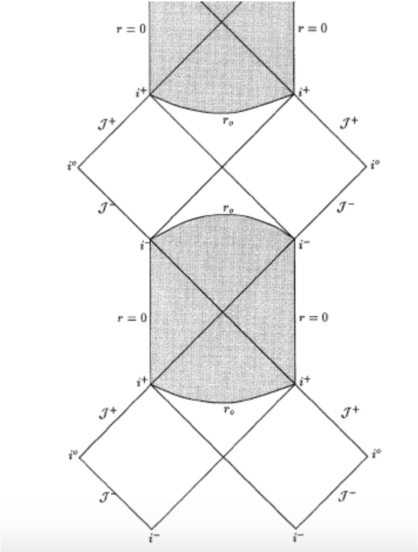
The only assumption in the Hawking-Penrose theorem that is not met is the *generic condition*, which can certainly be seen to fail for some specific timelike geodesics.

Devil in the detail: de Sitter

$$ds^2 = -c^2dt^2 + \lambda^2 \cosh^2(ct/\lambda)(d\chi^2 + \sin^2\chi d\Omega^2), \qquad \lambda^2 = 3/\Lambda$$



Devil in the detail: regular black holes



Topology change.

Incompleteness at the Cauchy horizon. Extensions!

A singularity-free perfect-fluid with $p = \varrho/3$

Take $(\mathbb{R}^4, g)$ in cylindrical coordinates $\{t, \rho, \varphi, z\}$ with

$$ds^2 = \cosh^4(at)\cosh^2(3a\rho)(-dt^2 + d\rho^2) + \\ + \frac{1}{9a^2}\cosh^4(at)\cosh^{-2/3}(3a\rho)\sinh^2(3a\rho)d\varphi^2 \\ + \cosh^{-2}(at)\cosh^{-2/3}(3a\rho)dz^2,$$

- Solution for a perfect fluid ($\Lambda = 0$) with:

$$\frac{8\pi G}{c^4}\varrho = 15a^2\cosh^{-4}(at)\cosh^{-4}(3a\rho) \qquad \boxed{p = \frac{1}{3}\varrho.}$$

- This solution is geodesically complete and **globally hyperbolic**. Each $t =$const. hypersurface is a non-compact global Cauchy hypersurface.
- The trace of their 2nd fundamental forms (expansion) is

$$K = \nabla_\mu u^\mu = 3a\frac{\sinh(at)}{\cosh^3(at)\cosh(3a\rho)} \quad (> 0 \text{ for } t > 0).$$

Devil in the detail...

- This space-time satisfies the stricter convergence and curvature conditions (dominant, strict strong) hence the focussing effect takes place fully,
- At the same time there are maximal timelike geodesics, **without focal points**, between any two (causally related) points of the manifold
- This may happen because the expansion is not bounded from below by a positive constant: $\lim_{\rho\to\infty} K = 0$.
- This subtle difference allows for the model to be geodesically complete avoiding Hawking's theorem
- The Raychudhuri-Komar theorem does not apply because there is acceleration, that is, pressure gradients.
- In summary, none of the possibilities for the initial/boundary condition holds in this model.

The problem of extensions

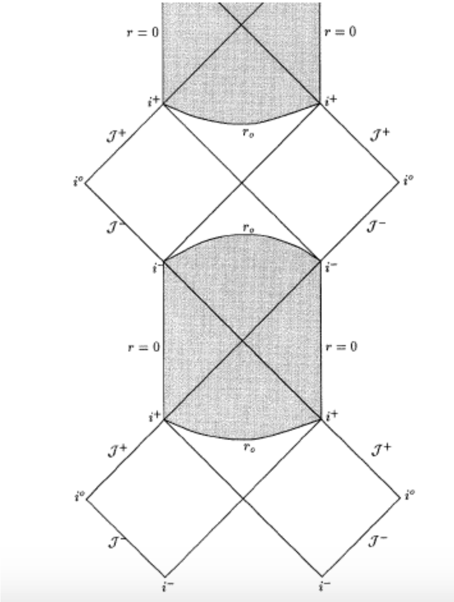
- Singularities may indicate *not* a problem with the incomplete curve when approaching the edge, but rather incompleteness of the manifold itself (e.g. excising regular points).
- This is why usually the study of singularities is restricted to inextensible manifolds.
- The physical problem, however, is hidden under the carpet with this “solution” because: what are we supposed to do with given extensible space-times?
- The answer may seem simple and easy: just extend them until you cannot extend it anymore...
- However, this is not so simple ...

Devil in the detail: problem of extensions

This is not so simple for several reasons:

1. physically meaningful extensions are far from obvious
2. extensions are not unique; usually there are infinite inequivalent choices.
3. analytical ones are not possible, nor advisable, many times
4. for a given extensible space-time, there are usually inequivalent extensions leading to
   - i new extensible space-times,
   - ii singular (incomplete) space-times,
   - iii to singularity-free (complete) and inextensible spacetimes
5. It might seem obvious that one should choose iii, but this is not the case! —if the singularity-free extension violates a physical condition, such as causality or energy positivity, then other extensions will be preferred.

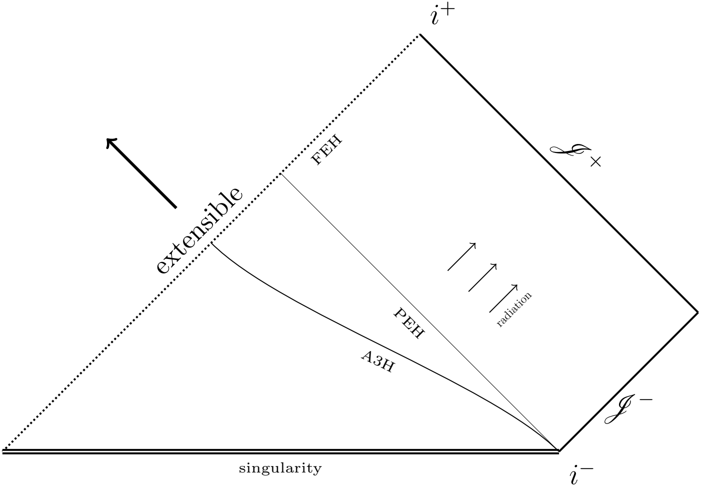
Elementary example: regular BHs



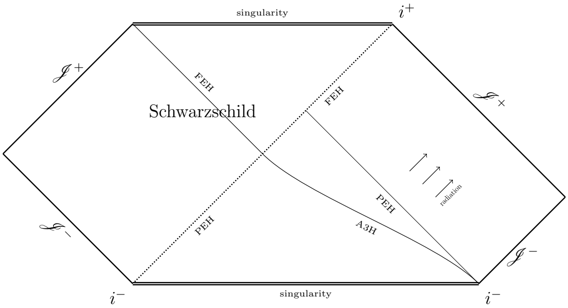
Inequivalent extension of Schwarzschild.
Notice what Penrose's theorem says

Vaidya's radiating metric

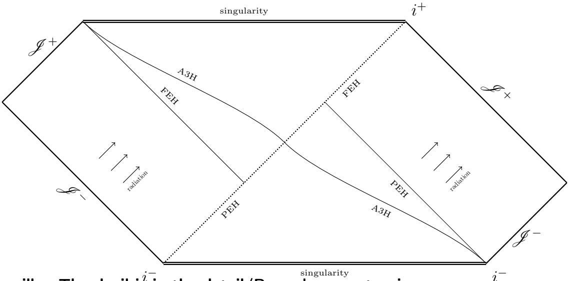
$$ds^2 = -\left(1 - \frac{2m(u)}{r}\right) du^2 - 2du dr + r^2 d\Omega^2, \quad \dot{m} \leq 0$$



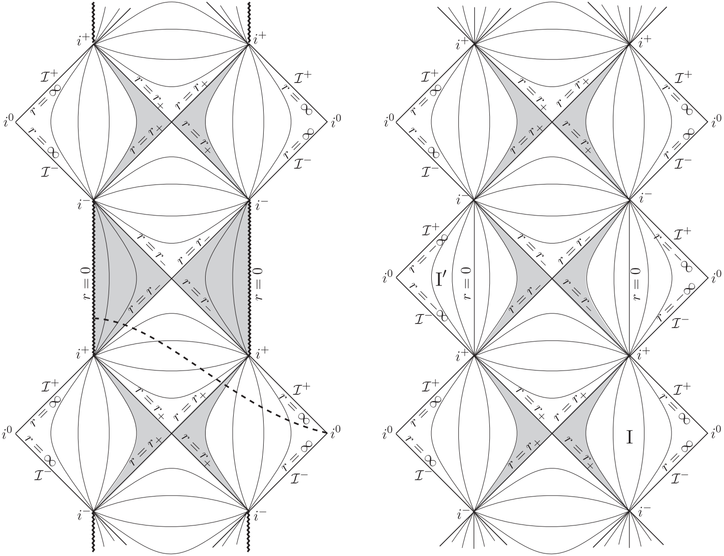
Two inequivalent possibilities



(Fayos-M-Prats-Senovilla)



The (maximally extended) Kerr BH



From Griffiths-Podolsky, Exact Spacetimes in Einstein's General Relativity (C.U.P)

BKL picture

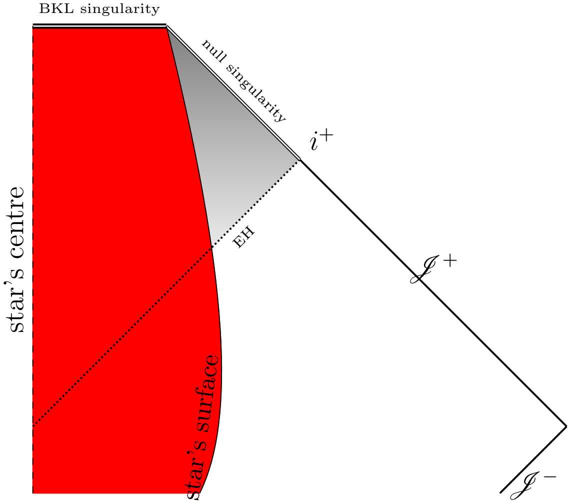
- The singularity theorems give very little information on the nature of singularities
- They merely state that for some unspecified reason some particle or light ray ends
- However, Belinskii, Khalatnikov and Lifschitz (BKL) argued, based on the field equations, that one can concentrate on dominant terms (time derivatives) to get a faithful idea of the character of the singularity
- The BKL conclusion can be summarized as *singularities are spacelike, local, oscillatory (Mixmaster), and 'matter does not matter'*.
- Even though there remain some technical problems, there exist theoretical (Ringström) as well as numerical (Berger, Moncrief, Garfinkle) results supporting the BKL picture.

BKL or Null singularity?

- Though the BKL picture is consistent with the field equations there are other competing possibilities on the market
- The main competitor is the idea of *Null singularities* (Poisson-Israel), based on the instability of Cauchy horizons
- The inner (Cauchy) horizon of a Reissner-Nordström, or a Kerr, or a regular black hole is an unstable null hypersurface in the sense that small perturbations blow up there.
- These perturbations should turn the horizon into a singularity, but this would still *retain its null character*.
- Invoking the BH uniqueness theorems, this picture should also be a good description of the singularity inside a black hole.
- The null singularity picture is supported by theoretical arguments (Ori-Flanagan), numerical simulations (Brady-Smith) and mathematical results (Dafermos).

Is this the right picture?

Thus, there is some tension between BKL and Null pictures! Can the singularity depend on the observer?



Summary, with some final remarks

- All gravitational systems are regular and accurately described with GR and its (post-Newtonian) limits, with the possible exception of the Universe and black holes.
- The singularities (incompleteness) in these two exceptions are a distinctive feature of GR
- The singularity theorems provide supporting evidence for the need of (quantum?) corrections to GR well inside black holes
- The problems of extensions and of the type of singularity (naked, censored..., if censored, BKL or Null) remain
- The most powerful theorems (Hawking-Penrose) have little application
- There is no theorem that can "predict" the singularity in the (maximally extended) Kerr black hole. And probably there will never be!

Thanks

dziękuję !

Thank you for your attention!