Why Are Siegel Modular Forms Relevant for Black Hole Physics?

Ashoke Sen

International Center for Theoretical Sciences, Bengaluru, India

Bengaluru, July 2025

Translation: Physics $\to$ Mathematics

Theory $\to$ A set of differential equations

Black hole $\to$ A solution to these equations

Mass M, charges Q = $(Q_1, Q_2, \cdots)$ $\to$ integration constants

Entropy f(M,Q) $\to$ A number computed from the solution

Temperature $\to$ Another number computed from the solution

Supersymmetric theory $\to$ A special set of differential equations

Extremal black hole $\to$ special choice of integration constants

– temperature = 0

– M is fixed in terms of Q, Entropy = F(Q)

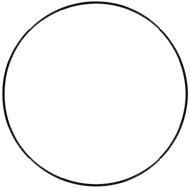
In a theory of gravity, possibly coupled to other fields, e.g. electromagnetic fields, black holes are classical solutions to the equations of motion:

$$R_{\mu\nu} - \frac{1}{2} R\, g_{\mu\nu} = \cdots, \qquad \cdots$$

Simplest black holes are spherically symmetric

– surrounded by a spherical surface that acts as a one way membrane – from outside to the inside

– known as the horizon



In quantum theory the black hole behaves as a thermal object with temperature, entropy etc.

The ‘entropy’ S is given by a simple formula Bekenstein; Hawking

$$S = \frac{A\,c^3}{4\,G\,\hbar}$$

A: area of the horizon

c: speed of light

G: Newton’s gravitational constant

$\hbar$: Planck’s constant

Other than A, all other quantities in this formula are universal constants of nature

This formula is correct when the black hole is very big, but when this is not so, there are corrections involving log(A), 1/A etc

A in turn is determined from the mass and charges carried by the black hole.

This gives a relation between entropy and mass and charges

$$S = f(M, Q)$$

Q should be regarded as a collection of numbers $(Q_1, \cdots .Q_n)$ labelling different types of charges carried by the black hole

In statistical physics, the entropy of a system has an interpretation as the result of a counting problem

e^S: Dimension of the Hilbert space of the system.

Question: Does $e^{f(M,Q)}$ have a similar interpretation?

In order to make the question precise we need to refine it a bit.

1. A generic black hole has temperature and as a result radiates away its energy

– a time dependent system

Hilbert space of such a system will also evolve with time due to interaction with the environment

In order to avoid this we consider extremal black holes in a supersymmetric theory

– zero temperature black hole

For this the mass is fixed in terms of the charges $\Rightarrow$ M=g(Q)

$$S = F(Q), \qquad F(Q) := f(M = g(Q), Q)$$

2. Supersymmetric theories of the type we consider have moduli spaces

– a set of parameters labelling the theory

In two regions of the moduli space, the weak coupling region and the black hole region, the total Hilbert space of the theory factorizes

$$H_{total} = H_{BH} \otimes H_{environment}$$

In between the two regions the factorization fails.

Weak coupling		Black hole
$H_{total} = H'_{BH} \otimes H'_{environment}$	$H_{total} \neq H_{BH} \otimes H_{environment}$	$H_{total} = H_{BH} \otimes H_{environment}$

As a result, $\dim(H'_{BH})$ can differ from $\dim(H_{BH})$

Weak coupling **Black hole**

$H_{total} = H'_{BH} \otimes H'_{environment}$ $H_{total} \neq H_{BH} \otimes H_{environment}$ $H_{total} = H_{BH} \otimes H_{environment}$

$\dim(H'_{BH})$ can be computed explicitly but not $\dim(H_{BH})$

$e^{F(Q)}$ should be compared to $\dim(H_{BH})$

$\Rightarrow$ we cannot directly compare $e^{F(Q)}$ with the result of counting states.

Remedy: Study the index

Witten; Bachas, Kiritsis; Gregori, Kiritsis, Kounnas, Obers, Petropoulos, Pioline

$$I(Q) = Tr_{H_{total},Q}(\sigma)$$

– trace over states in H_{total} carrying fixed charges Q

σ: takes value 1 for bosonic states and -1 for fermionic states

1. I(Q) is independent of the moduli

2. I(Q) can be computed at weak coupling

3. In the black hole regime, up to computable corrections, I(Q) gets contribution only from the states in H_{BH}

Each state in H_{BH} can be shown to have $\sigma = 1$ A.S.

$\Rightarrow I(Q) = Tr_{H_{BH},Q}(1) = \dim(H_{BH,Q}) = e^{F(Q)}$

We compare ln I(Q) with F(Q)

$$\mathbf{F(Q)=\ln I(Q)?}$$

F(Q) is calculated from geometry of black hole space-time

$$\mathbf{F(Q)} = \frac{\mathbf{A\,c^3}}{\mathbf{4\,G}\,\hbar} + \textbf{corrections}$$

I(Q) is computed from counting of states

Siegel modular forms enter in the computation of I(Q) for a class of string theories known as N=4 supersymmetric theories

Dijkgraaf, Verlinde, Verlinde; Jatkar, A.S.; David, Jatkar; A.S.

We shall first explain the results and then outline the derivation

In the N=4 supersymmetric string theories we have a large group of duality symmetries that allow us to label a state using three charges (Q_1, Q_2, Q_3)

The index $I(Q_1, Q_2, Q_3)$ is a function of these charges

We define the generating function of the index:

$$f(\rho, \sigma, \nu) = \sum_{Q_1, Q_2, Q_3} I(Q_1, Q_2, Q_3)\, e^{2\pi i(\rho Q_1 + \sigma Q_2 + \nu Q_3)}$$

Result: $f(\rho, \sigma, \nu)$ is given by the inverse of a meromorphic Siegel modular form

Definition of $\mathrm{Sp}(2,\mathbb{Z})$ group:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \mathrm{Sp}(2,\mathbb{Z})$$

if A, B, C, D are 2×2 matrices with integer entries, satisfying

$$AD^T - BC^T = I, \qquad AB^T = BA^T, \qquad CD^T = DC^T$$

For N=1,2,3,5,7, we define $G_N \subset \mathrm{Sp}(2,\mathbb{Z})$ as the group generated by $g_0 g_1 g_0^{-1}$, $g_0 g_2 g_0^{-1}$ and $g_0 g_3 g_0^{-1}$ with

$$g_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 \end{pmatrix}, \quad g_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \quad g_3(\lambda,\mu) = \begin{pmatrix} 1 & 0 & 0 & \mu \\ \lambda & 1 & \mu & 0 \\ 0 & 0 & 1 & -\lambda \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \lambda,\mu \in \mathbf{Z}$$

$$g_2(a,b,c,d) = \begin{pmatrix} a & 0 & b & 0 \\ 0 & 1 & 0 & 0 \\ c & 0 & d & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad ad - bc = 1, \quad c=0 \bmod N, \quad a,d=1 \bmod N$$

$$G_1 = \mathrm{Sp}(2,\mathbb{Z})$$

$\Phi_k(\rho, \sigma, \nu)$ is a meromorphic Siegel modular form of weight k of G_N if

1. $\Phi_k(\rho, \sigma, \nu)$ is a meromorphic function in the Siegel upper half plane

$$\rho_2 > 0, \quad \sigma_2 > 0, \quad \rho_2\sigma_2 - \nu_2^2 > 0$$

$$(\rho_2, \sigma_2, \nu_2) := \mathrm{Im}(\rho, \sigma, \nu), \qquad (\rho_1, \sigma_1, \nu_1) := \mathrm{Re}(\rho, \sigma, \nu)$$

2. If we define

$$\Omega = \begin{pmatrix} \rho & \nu \\ \nu & \sigma \end{pmatrix}$$

then

$$\Phi_k((A\Omega + B)(C\Omega + D)^{-1}) = \det(C\Omega + D)^k \Phi_k(\Omega) \quad \text{for} \quad \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in G_N$$

There is a class of N=4 supersymmetric string theories known as CHL models, labelled by a prime number N=1,2,3,5,7

Chaudhuri, Hockney, Lykken

For this theories,

$$f(\rho, \sigma, \nu) = \Phi_k\left(\rho, \sigma, \nu + \frac{1}{2}\right)^{-1}, \qquad k = \frac{24}{N+1} - 2$$

Simplest example: Igusa cusp form, N=1, k=10

$$\mathbf{F}(\mathbf{z},\tau) := 8\left[\left(\frac{\vartheta_2(\mathbf{z}|\tau)^2}{\vartheta_2(0|\tau)^2}\right) + \left(\frac{\vartheta_3(\mathbf{z}|\tau)^2}{\vartheta_3(0|\tau)^2}\right) + \left(\frac{\vartheta_4(\mathbf{z}|\tau)^2}{\vartheta_4(0|\tau)^2}\right)\right]$$

ϑ_i: Jacobi theta functions

Define $c(s), s \in \mathbb{Z}$ via

$$F(z,\tau) = \sum_{n,b\in\mathbb{Z}} c(4n-b^2)\, e^{2\pi i n\tau + 2\pi i b z}$$

Then

Borcherds; Gritsenko, Nikulin

$$\Phi_{10}(\rho,\sigma,\nu) = e^{2\pi i(\rho+\sigma+\nu)} \prod_{\substack{k,\ell,j\in\mathbb{Z}\\ k,\ell\geq 0,\ j<0 \text{ if } k=\ell=0}} \left(1 - e^{2\pi i(k\rho+\ell\sigma+j\nu)}\right)^{c(4\ell k-j^2)}$$

Similar explicit expressions exist for the other Φ_k's

David, Jatkar, A.S.

An outline of the derivation

Shih, Strominger, Yin; David, A.S.

In the weak coupling regime H'_{BH} becomes the product of many bosonic and fermionic harmonic oscillators

Bosonic harmonic oscillator: Hilbert space states are labelled by a non-negative integer n

The n-th state carries charge $(Q_1, Q_2, Q_3) = (nq_1, nq_2, nq_3)$ for some q_1, q_2, q_3

All states have $\sigma = 1$

Generating function:

$$\sum_{n=0}^{\infty} e^{2\pi i(nq_1\rho + nq_2\sigma + nq_3\nu)} = \left(1 - e^{2\pi i(q_1\rho + q_2\sigma + q_3\nu)}\right)^{-1}$$

Fermionic harmonic oscillators: Hilbert space states are labelled by an integer n taking values 0,1

The n-th state carries charge $(Q_1, Q_2, Q_3) = (nq_1, nq_2, nq_3)$ for some q_1, q_2, q_3

The states have $\sigma = (-1)^n$

Generating function:

$$\sum_{n=0}^{1} e^{2\pi i(nq_1\rho + nq_2\sigma + nq_3\nu)}(-1)^n = \left(1 - e^{2\pi i(q_1\rho + q_2\sigma + q_3\nu)}\right)$$

We label the oscillators by the index α with the charges $(\mathbf{q}_1^\alpha, \mathbf{q}_2^\alpha, \mathbf{q}_3^\alpha)$ and define

$$\sigma_\alpha = \begin{cases} 1 \text{ for bosonic oscillator} \\ -1 \text{ for fermionic oscillator} \end{cases}$$

Then the total partition function is

$$\sigma_0\, e^{2\pi i(Q_1^{(0)}\rho + Q_2^{(0)}\sigma + Q_3^{(0)}\nu)} \prod_\alpha \left(1 - e^{2\pi i(q_1^\alpha \rho + q_2^\alpha \sigma + q_3^\alpha \nu)}\right)^{-\sigma_\alpha}$$

$(Q_1^{(0)}, Q_2^{(0)}, Q_3^{(0)})$: Charges carried by the state even when all oscillators have $n_\alpha = 0$

$\sigma_0 = \pm 1$ depending on whether the state with all $n_\alpha = 0$ is a boson or a fermion.

After inclusion of the contribution from all the oscillators, this product magically becomes the inverse of a Siegel modular form

How do we compare F(Q) and ln I(Q)?

I(Q) is known exactly but F(Q) is known only approximately as an expansion in 1/Q

Leading area term for large Q_1, Q_2

$$F(Q) = \pi\sqrt{4\,Q_1\,Q_2 - Q_3^2}$$

Next correction scales as Q^0

etc.

We need to study the asymptotic expansion of ln I(Q) for large charges

Cardoso, de Wit, Kappeli, Mohaupt; Jatkar, A.S.; David, Jatkar, A.S.

$$\Phi_k\left(\rho,\sigma,\nu+\frac{1}{2}\right)^{-1} = \sum_{Q_1,Q_3\in\mathbb{Z},Q_2\in\mathbb{Z}/N} I(Q_1,Q_2,Q_3)\, e^{2\pi i(\rho Q_1+\sigma Q_2+\nu Q_3)}$$

$$I(Q_1,Q_2,Q_3) = \frac{(-1)^{Q_3}}{N}\int_0^1 d\rho_1 \int_0^N d\sigma_1 \int_0^1 d\nu_1\, e^{-2\pi i(\rho Q_1+\sigma Q_2+\nu Q_3)}\Phi_k(\rho,\sigma,\nu)^{-1}$$

Since $\Phi_k(\rho,\sigma,\nu)^{-1}$ has poles it is necessary to specify the integration contour.

Go back to the generating function

$$\sum_{n=0}^{\infty} e^{2\pi i(nq_1\rho+nq_2\sigma+nq_3\nu)} = \left(1-e^{2\pi i(q_1\rho+q_2\sigma+q_3\nu)}\right)^{-1}$$

One finds that for most oscillators the convergence of lhs is ensured by taking

$$\rho_2,\sigma_2>0, \qquad \rho_2\sigma_2>\nu_2^2$$

$\Rightarrow$ we take the contour at large ρ_2,σ_2 with $\rho_2\sigma_2-\nu_2^2>0$.

$$I(Q_1, Q_2, Q_3) = \frac{(-1)^{Q_3}}{N} \int_0^1 d\rho_1 \int_0^N d\sigma_1 \int_0^1 d\nu_1 \, e^{-2\pi i(\rho Q_1 + \sigma Q_2 + \nu Q_3)} \Phi_k(\rho, \sigma, \nu)^{-1}$$

Contour at

$$\text{large } \rho_2, \sigma_2, \qquad \rho_2\sigma_2 - \nu_2^2 > 0$$

This does not remove all the ambiguities, but the remaining ambiguities do not affect large Q expansion

– to be discussed later

To study the large Q expansion, we deform the contour by lowering $\rho_2, \sigma_2, |\nu_2|$ keeping fixed their ratios

First pole is encountered at

$$\rho\sigma - \nu^2 + \nu = 0$$

We deform the contour through this, picking up residue at the pole.

The contribution from the deformed contour can be shown to be exponentially small compared to the residue at the pole.

Focus on the residue contribution

$\Rightarrow$ a two dimensional integral over ρ_1, σ_1

We can evaluate this integral using steepest descent method / Lefschetz thimble

Result:

$$I(Q) = \text{Exp}[\pi\sqrt{4Q_1Q_2 - Q_3^2} + \cdots]$$

The leading term in ln I(Q) agrees with the area contribution to the entropy F(Q).

$$I(Q) = \text{Exp}[\pi\sqrt{4Q_1Q_2 - Q_3^2} + \cdots]$$

The subleading terms in I(Q), denoted by $\cdots$, are relatively straightforward to compute using steepest descent method

But the subleading corrections to the black hole entropy F(Q) are more difficult to compute.

So far the comparison has been done up to the first subleading term.

An example: The subleading term in ln I(Q) for N=1 case

$$-24\ln|\eta(\tau)|^2 - 12\ln(2\tau_2), \qquad \tau = \tau_1 + i\tau_2 = \frac{Q_3}{Q_2} + i\frac{\sqrt{Q_1Q_2 - Q_3^2}}{Q_2}$$

agrees with F(Q) computed from black hole geometry.

For $N > 1$ we get more complicated functions

How to choose the initial contour?

A.S.; Dabholkar, Gaiotto, Nampuri; Cheng, Verlinde

$$\rho_2, \sigma_2 \ \text{large}, \qquad \rho_2\sigma_2 - \nu_2^2 > 0$$

But even in this region there are poles at

$$j\,\nu + n_1\,\sigma - m_1\,\rho + m_2 = 0$$

$$m_1 \in N\mathbb{Z}, \quad m_2, n_1 \in \mathbb{Z}, \quad j \in 2\mathbb{Z}+1, \quad m_1 n_1 + \frac{j^2}{4} = \frac{1}{4}$$

$$\mathbf{j}\,\nu + \mathbf{n_1}\,\sigma - \mathbf{m_1}\,\rho + \mathbf{m_2} = \mathbf{0}$$

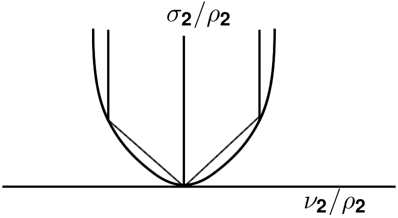
$$\mathbf{m_1} \in \mathbf{N}\mathbb{Z}, \quad \mathbf{m_2}, \mathbf{n_1} \in \mathbb{Z}, \quad \mathbf{j} \in \mathbf{2}\mathbb{Z} + \mathbf{1}, \quad \mathbf{m_1 n_1} + \frac{\mathbf{j^2}}{\mathbf{4}} = \frac{\mathbf{1}}{\mathbf{4}}$$

Imaginary part

$$\mathbf{j}\,\nu_2 + \mathbf{n_1}\,\sigma_2 - \mathbf{m_1}\,\rho_2 = \mathbf{0}$$

– divides up the ρ_2, σ_2 large, $\rho_2\sigma_2 - \nu_2^2 > 0$ region into many chambers separated by codimension one walls

We get different I(Q) in different chambers



Difference in I(Q) between neighboring chambers is determined by the residue at the pole at

$$\mathbf{j}\,\nu + \mathbf{n_1}\,\sigma - \mathbf{m_1}\,\rho + \mathbf{m_2} = \mathbf{0}$$

Residues of $\Phi_{\mathbf{k}}(\rho,\sigma,\nu)^{-1}$ at these poles give exponentially subleading contribution relative to the contribution from the pole at

$$\rho\sigma - \nu^{\mathbf{2}} + \nu = \mathbf{0}$$

$\Rightarrow$ this ambiguity does not affect the asymptotic entropy and comparison with black hole entropy

Nevertheless we need to understand the physical origin of this ambiguity

It turns out that the moduli space of N=4 supersymmetric string theories can also be divided up into many chambers, with different I(Q) in different chambers

There is one to one map between the chambers in the moduli space and the chambers in the $(\rho_2, \sigma_2, \nu_2)$ space.

The jump in I(Q) between two neighboring chambers in the moduli space can be computed from the residue of Φ_k^{-1} at the corresponding pole in the (ρ, σ, ν) space.

Why is there a jump in I(Q) between the neighboring chambers in the moduli space?

Explanation in the weak coupling side: The nature of some oscillators changes across the wall

Example:

Case 1. The n-th state of a bosonic oscillator carries $(Q_1, Q_2, Q_3) = \left(0, 0, n + \frac{1}{2}\right), \qquad n = 0, 1, 2, \cdots$

Generating function:

$$\sum_{n=0}^{\infty} e^{2\pi i\nu\left(n+\frac{1}{2}\right)} = e^{\pi i\nu}(1 - e^{2\pi i\nu})^{-1} = (e^{-\pi i\nu/2} - e^{\pi i\nu/2})^{-1}$$

Case 2. The n-th state of the bosonic oscillator carries $(Q_1, Q_2, Q_3) = \left(0, 0, -n - \frac{1}{2}\right), \qquad n = 0, 1, 2, \cdots$

Generating function:

$$\sum_{n=0}^{\infty} e^{-2\pi i\nu\left(n+\frac{1}{2}\right)} = e^{-\pi i\nu}(1 - e^{-2\pi i\nu})^{-1} = (e^{\pi i\nu/2} - e^{-\pi i\nu/2})^{-1}$$

Two such oscillators will have the same generating function but different I(Q)

$$\text{Case 1}: \qquad \left(\sum_{n=0}^{\infty} e^{2\pi i\nu\left(n+\frac{1}{2}\right)}\right)^2 = \left(e^{-\pi i\nu/2} - e^{\pi i\nu/2}\right)^{-2}$$

$$\text{Case 2}: \qquad \left(\sum_{n=0}^{\infty} e^{-2\pi i\nu\left(n+\frac{1}{2}\right)}\right)^2 = \left(e^{\pi i\nu/2} - e^{-\pi i\nu/2}\right)^{-2}$$

Even though the generating functions are the same, their contribution to I(Q) are different

First case:

$$\int_0^1 d\nu_1\, e^{-2\pi i Q_3 \nu_1} \left(e^{-\pi i\nu/2} - e^{\pi i\nu/2}\right)^{-2}\Big|_{\nu_2>0}$$

Second case

$$\int_0^1 d\nu_1\, e^{-2\pi i Q_3 \nu_1} \left(e^{\pi i\nu/2} - e^{-\pi i\nu/2}\right)^{-2}\Big|_{\nu_2<0}$$

Difference is given by the residue at the pole at $\nu = 0$.

In $\Phi_{\mathbf{k}}(\rho,\sigma,\nu)^{-1}$ the $(\mathbf{e}^{-\pi \mathbf{i}\nu/2}-\mathbf{e}^{\pi \mathbf{i}\nu/2})^{-2}$ is multiplied by the contribution from the other oscillators

But the reason for the jump is the same.

The jump across the wall

$$\mathbf{j}\,\nu_2+\mathbf{n}_1\,\sigma_2-\mathbf{m}_1\,\rho_2=0$$

has similar explanation with ν replaced by $\mathbf{j}\nu+\mathbf{n}_1\sigma-\mathbf{m}_1\rho$

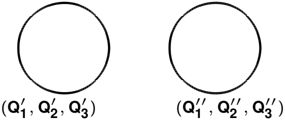
Explanation on the black hole side:

The index gets contribution not just from single black holes but also a system of two black holes with charges (Q'_1, Q'_2, Q'_3) and (Q''_1, Q''_2, Q''_3) with

$$(Q'_1, Q'_2, Q'_3) + (Q''_1, Q''_2, Q''_3) = (Q_1, Q_2, Q_3)$$

These 'two centered configurations' exist in some region of the moduli space but not everywhere

The index changes when we cross a wall across which some two centered configuration disappears.



Question: Can we write down the generating function for the index of single centered black holes only?

– need to subtract the contribution of two centered black holes

Mock Jacobi forms do the job to some extent but there are limitations

Dabholkar, Murthy, Zagier; Bhand, A.S., Singh; Benerjee, Bhutra, Singh

1. Need to use different generating function for different sets of charges

2. Fails to work for some charges even within this set

Goal: Write down a single generating function that captures the index of single centered black holes for all charges

Bhand, A.S., Singh, work in progress

– Mock Siegel modular forms?

Sign of the index

$$\mathbf{I(Q)} = \mathbf{Tr}_{\mathbf{H_{BH},Q}}(\sigma) = \mathbf{Tr}_{\mathbf{H_{BH},Q}}(\mathbf{1}) = \mathbf{dim}(\mathbf{H_{BH,Q}})$$

since for single centered black holes $\sigma = 1$

Thus after extracting the single centered black hole index from Φ_k we have positivity property of the index

– has been tested using the Fourier coefficients of $\Phi_k(\rho, \sigma, \nu)^{-1}$

A.S.; Bringmann, Murthy; Chattopadhyaya, David

Twisted index

There are subspaces of the moduli spaces of N=4 supersymmetric theories where the theory has extra $\mathbb{Z}_p$ symmetry generated by g for p=2,3,4,$\cdots$

– can be used to define twisted index

$$I_g(Q) = Tr_{H_{total},Q}(\sigma\, g) = Tr_{H_{BH},Q}(\sigma\, g)$$

– given by Fourier coefficients of a different set of Siegel modular forms

Black hole physics tells us that for large charges $I_g(Q)$ should grow as $\exp[S/p] = \exp[\pi\sqrt{4Q_1Q_2 - Q_3^2}/p]$

– has been verified by studying asymptotic growths of the twisted index computed from Siegel modular forms.

One also has positivity properties of the twisted index

e.g. for p=2, we have

$$\frac{1}{2}(I(Q) \pm I_g(Q)) = Tr_{H_{BH},Q}\left(\sigma\,\frac{1 \pm g}{2}\right) \simeq Tr_{H_{BH},Q}\left(\frac{1 \pm g}{2}\right) > 0$$

$\simeq$: after removing contribution from two centered black holes

– has also been tested in many examples.

Govindarajan, Samanta, Shanmugapriya, Virmani