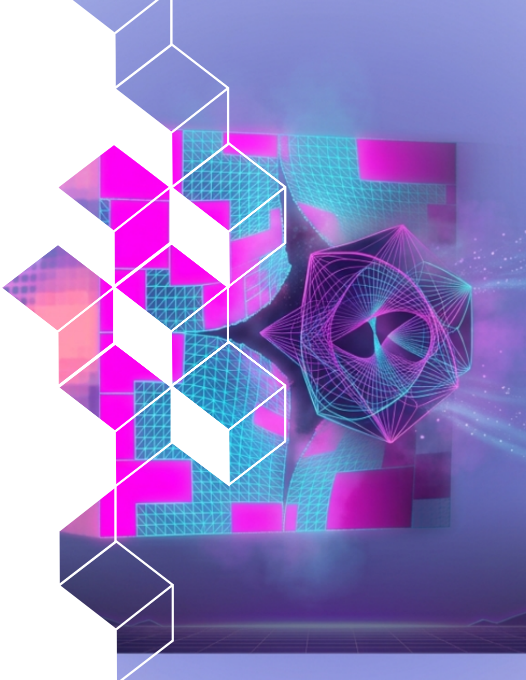
Data Field Theory

A Functional Renormalization Group Approach to Anomaly Detection via Dimensional Phase Transitions

Riccardo FINOTELLO

13th International Conference on the Exact Renormalization Group

4 September 2026



The People

CEA Paris-Saclay

Riccardo
FINOTELLO

Vincent
LAHOCHE

Parham
RADPAY

U. of Abomey-Calavi

Dine
OUSMANE
SAMARY

Cover image generated with Google's *Nano Banana 2* model.

Table of contents

1. **The Problem**
   The BBP Transition of Spiked Models
   Extensive-Rank Signals and the Extensive Spike Model
2. **The Field Theory**
   The Empirical Spectrum as an EFT
   The FRG: Wetterich, Litim, LPA
3. **Dimensional Phase Transition**
   The Canonical Dimensions
   Detecting the Instability
   The Onsager Benchmark
4. **Implications**
   A Realistic Data Example
   Open Questions and Outlook

1. The Problem

The Baik–Ben Arous–Péché Transition

Spiked Matrix Model

Let us consider a **spiked matrix model**:

$$X = \beta\, u v^\top + Z \in \mathbb{R}^{n \times p}, \qquad Z \sim \mathcal{N}(0, \mathbb{1}), \quad u \in \mathbb{R}^n, \quad v \in \mathbb{R}^p, \quad \text{s.t.} \quad \|u\| = \|v\| = 1,$$

with a **covariance matrix**:

$$C = n^{-1} X^\top X \in \mathbb{R}^{p \times p}.$$

The spiked matrix model is related to the spiked covariance model by $X = \Sigma^{1/2} Z$, where $\Sigma = \mathbb{1} + \beta\, u u^\top$ and $Z = \mathbb{1} + W / \sqrt{n}$, from which $C = n^{-1} X^\top X = n^{-1} Z^\top \Sigma Z = \mathbb{1} + \beta\, u v^\top + \frac{2}{\sqrt{n}} W + \mathcal{O}(\beta / \sqrt{n})$. Hence, $C - \mathbb{1} \simeq \beta\, u v^\top + \frac{2}{\sqrt{n}} W$, i.e. a spiked matrix model.

The Baik–Ben Arous–Péché Transition

Spiked Matrix Model

Let us consider a **spiked matrix model**:

$$X = \beta\, u v^\top + Z \in \mathbb{R}^{n \times p}, \qquad Z \sim \mathcal{N}(0, \mathbb{1}), \quad u \in \mathbb{R}^n, \quad v \in \mathbb{R}^p, \quad \text{s.t.} \quad \|u\| = \|v\| = 1,$$

with a **covariance matrix**:

$$C = n^{-1} X^\top X \in \mathbb{R}^{p \times p}.$$

`Data Scientist:` *Under what conditions can we recover the signal v from the noisy observation X (i.e., when is the signal detectable)?*

`Mathematician:` *Under what conditions does the largest eigenvalue of C separate from the bulk of the spectrum?*

The spiked matrix model is related to the spiked covariance model by $X = \Sigma^{1/2} Z$, where $\Sigma = \mathbb{1} + \beta\, u u^\top$ and $Z = \mathbb{1} + W/\sqrt{n}$, from which $C = n^{-1} X^\top X = n^{-1} Z^\top \Sigma Z = \mathbb{1} + \beta\, u v^\top + \frac{2}{\sqrt{n}} W + \mathcal{O}(\beta/\sqrt{n})$. Hence, $C - \mathbb{1} \simeq \beta\, u v^\top + \frac{2}{\sqrt{n}} W$, i.e. a spiked matrix model.

The Baik–Ben Arous–Péché Transition

Decomposing signal

Decompose into **bulk + spike**:

$$n\,C = \underbrace{\beta^2 v v^\top + \beta \left(v u^\top Z + Z^\top u v^\top \right)}_{\text{spike contribution}} + \overbrace{Z^\top Z}^{\text{bulk}},$$

The Baik–Ben Arous–Péché Transition

Decomposing signal

Decompose into **bulk + spike**:

$$n\,C = \underbrace{\beta^2 v v^\top + \beta\left(v u^\top Z + Z^\top u v^\top\right)}_{\text{spike contribution}} + \overbrace{Z^\top Z}^{\text{bulk}},$$

and consider **spectra** for $n,\ p \to \infty$ and $p/n \to \gamma \in (0, 1]$:

$$n^{-1}\left\{\#\lambda_i\right\}_{\text{bulk}} \xrightarrow{\text{a.s.}} \mu_{\text{MP}}(\lambda) = \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{2\pi\gamma\lambda}\Theta(\lambda - \lambda_-)\Theta(\lambda_+ - \lambda),$$

where

$$\lambda_\pm = (1 \pm \sqrt{\gamma})^2$$

are the edges of the **Marčenko–Pastur** bulk.

The Baik–Ben Arous–Péché Transition

The BBP Transition Regime

The **spike contribution**

$$\beta^2 v v^\top + \beta \left(v u^\top Z + Z^\top u v^\top \right)$$

completely controls the **largest eigenvalue** λ_{signal} (i.e. the *signal* we want to detect):

$$\lambda_{\text{signal}} = (1 + \beta) \left(1 + \frac{\gamma}{\beta} \right) > \lambda_+ \quad \Longleftrightarrow \quad \beta_{\text{critical}} > \sqrt{\gamma}$$

Baik, Ben Arous, Péché (2005)

The Baik–Ben Arous–Péché Transition

The BBP Transition Regime

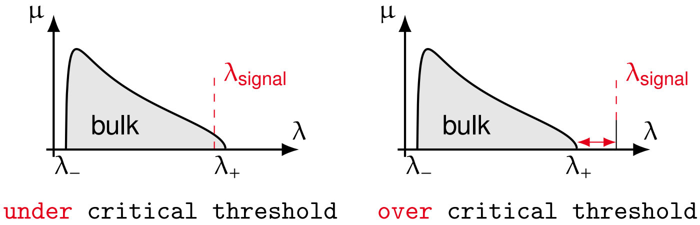
The **spike contribution**

$$\beta^2 v v^\top + \beta \left(v u^\top Z + Z^\top u v^\top \right)$$

completely controls the **largest eigenvalue** λ_{signal} (i.e. the *signal* we want to detect):

$$\lambda_{\text{signal}} = (1 + \beta) \left(1 + \frac{\gamma}{\beta} \right) > \lambda_+ \quad \Longleftrightarrow \quad \beta_{\text{critical}} > \sqrt{\gamma}$$

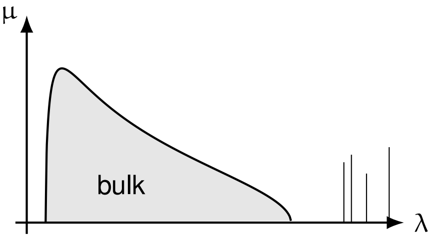
Two **regimes** appear:



Baik, Ben Arous, Péché (2005)

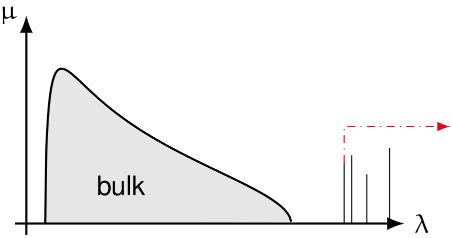
Failure of the BBP Transition
Extensive-rank Signals

However, let us consider:



Failure of the BBP Transition
Extensive-rank Signals

However, let us consider:

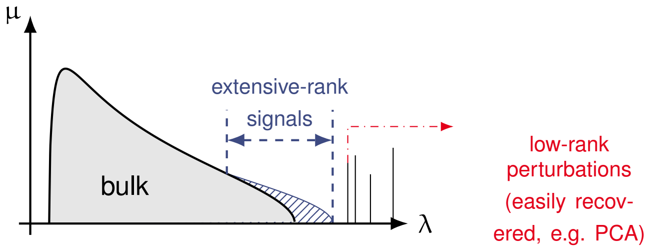


low-rank perturbations (easily recovered, e.g. PCA)

Failure of the BBP Transition

Extensive-rank Signals

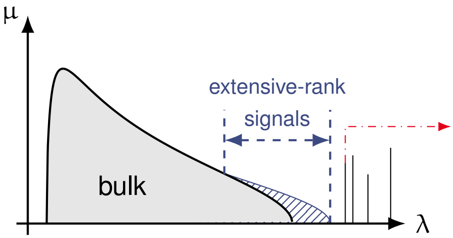
However, let us consider:



Failure of the BBP Transition

Extensive-rank Signals

However, let us consider:



low-rank perturbations (easily recovered, e.g. PCA)

For **extensive-rank signals**:

- We lose the ability to determine the largest eigenvalue of the bulk.
- We are no longer able to detect when the signal separates from bulk.
- BBP transition fails to give a separation threshold

Extensive-rank Signals
The Challenge

Let us consider an **extensive-rank** model:

$$X = \beta S + Z \in \mathbb{R}^{n \times p},$$

where

$$1 \ll \operatorname{rank} S < \min(n, p).$$

Landau, Mel, Ganguli (2023)

Extensive-rank Signals

The Challenge

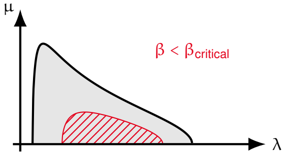
Let us consider an **extensive-rank** model:

$$X = \beta S + Z \in \mathbb{R}^{n \times p},$$

where

$$1 \ll \operatorname{rank} S < \min(n, p).$$

Intuitively:



Landau, Mel, Ganguli (2023)

Extensive-rank Signals
The Challenge

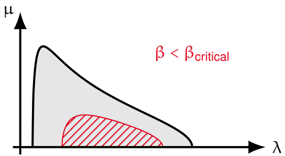
Let us consider an **extensive-rank** model:

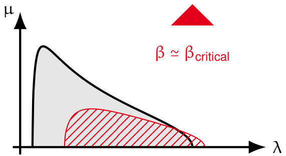
$$X = \beta S + Z \in \mathbb{R}^{n \times p},$$

where

$$1 \ll \operatorname{rank} S < \min(n, p).$$

Intuitively:





Landau, Mel, Ganguli (2023)

Extensive-rank Signals

The Challenge

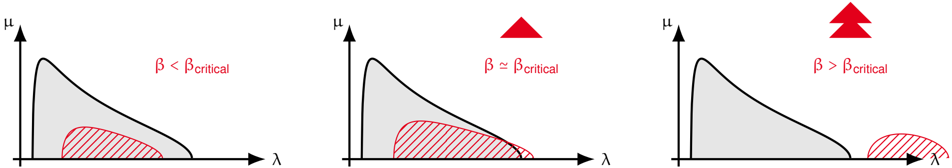
Let us consider an **extensive-rank** model:

$$X = \beta S + Z \in \mathbb{R}^{n \times p},$$

where

$$1 \ll \operatorname{rank} S < \min(n, p).$$

Intuitively:



Landau, Mel, Ganguli (2023)

Extensive-rank Signals
The Challenge

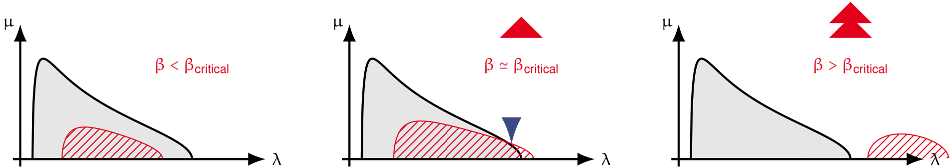
Let us consider an **extensive-rank** model:

$$X = \beta S + Z \in \mathbb{R}^{n \times p},$$

where

$$1 \ll \operatorname{rank} S < \min(n, p).$$

Intuitively:



How to detect the extensive-rank inflection/critical point?
How to follow the flow of the extensive-rank signal?

Landau, Mel, Ganguli (2023)

2. The Field Theory

Empirical Spectrum and Maximum Entropy

The MP Universality Class

MP Universality Class

C belongs to the MP universality class if:

1. $\mu_{\text{data}} = n^{-1} \left\{ \# \lambda_i \right\}_{\text{bulk}} \xrightarrow{\text{a.s.}} \mu_{\text{MP}}(\lambda)$ (i.e. the empirical distribution must "look like" the MP distribution),
2. $\int \left(\mathrm{d}\mu_{\text{data}} - \mathrm{d}\mu_{\text{MP}} \right) \leq K \left\| p_{\text{data}} - p_{\text{MP}} \right\|$ (i.e. the two distributions must not be "too far off"),

where $K > 0$, and p is the set of parameters which define the spectrum.

Empirical Spectrum and Maximum Entropy

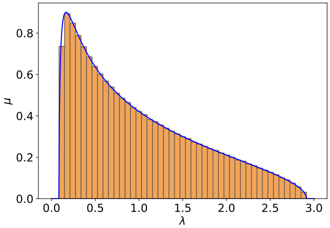
The MP Universality Class

MP Universality Class

C belongs to the MP universality class if:

1. $\mu_{\text{data}} = n^{-1} \{\#\lambda_i\}_{\text{bulk}} \xrightarrow{\text{a.s.}} \mu_{\text{MP}}(\lambda)$ (i.e. the empirical distribution must "look like" the MP distribution),
2. $\int (\mathrm{d}\mu_{\text{data}} - \mathrm{d}\mu_{\text{MP}}) \leq K \left\| p_{\text{data}} - p_{\text{MP}} \right\|$ (i.e. the two distributions must not be "too far off"),

where $K > 0$, and p is the set of parameters which define the spectrum.



Build field $\widetilde{\Phi} = \left(\tilde{\phi}_1 \ldots \tilde{\phi}_p\right)^\top$ s.t.
$G^{(2)} = \langle \widetilde{\Phi} \widetilde{\Phi}^\top \rangle = C$ and $C = U^\top \Lambda U \Rightarrow \widetilde{\Phi} = U^\top \Phi$:

$$P(\Phi) \overset{\text{max entropy}}{\propto} \exp\left(-\frac{1}{2} \Phi^\top \Lambda^{-1} \Phi\right)$$

Empirical Spectrum and Maximum Entropy

Gaussian Field Theory and MP Distribution

Using max entropy we have a minimally constrained distribution:

$$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p) \quad \text{s.t.} \quad \lambda_i = \frac{1}{k_i^2 + m^2}, \quad \text{for} \quad i = 1, \ldots, p,$$

where $\lambda_1 > \lambda_2 > \ldots > \lambda_p > 0$ and $m^2 = \lambda_1^{-1}$.

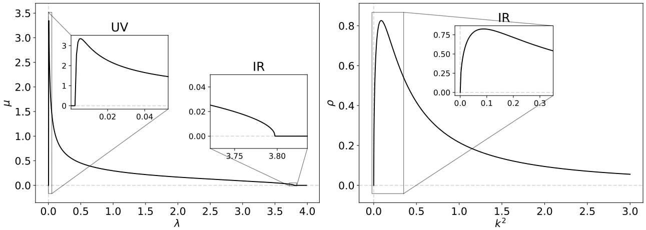
Empirical Spectrum and Maximum Entropy

Gaussian Field Theory and MP Distribution

Using max entropy we have a minimally constrained distribution:

$$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p) \quad \text{s.t.} \quad \lambda_i = \frac{1}{k_i^2 + m^2}, \quad \text{for} \quad i = 1, \ldots, p,$$

where $\lambda_1 > \lambda_2 > \ldots > \lambda_p > 0$ and $m^2 = \lambda_1^{-1}$.



The momentum distribution $\rho_{\mathrm{MP}}(k^2)$ defines:

UV: `small eigenvalues` (high energy/k^2)

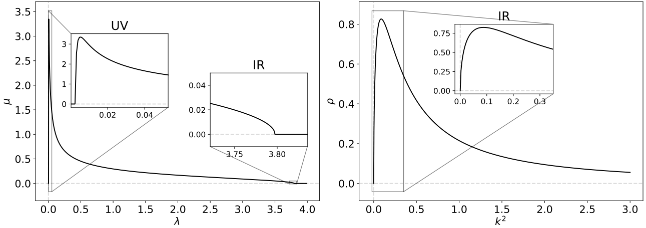
IR: `large eigenvalues` (low energy/k^2)

Empirical Spectrum and Maximum Entropy
Gaussian Field Theory and MP Distribution

Using max entropy we have a minimally constrained distribution:

$$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p) \quad \text{s.t.} \quad \lambda_i = \frac{1}{k_i^2 + m^2}, \quad \text{for} \quad i = 1, \ldots, p,$$

where $\lambda_1 > \lambda_2 > \ldots > \lambda_p > 0$ and $m^2 = \lambda_1^{-1}$.



The momentum distribution $\rho_{\mathrm{MP}}(k^2)$ defines:

UV: `small eigenvalues` (high energy/k^2)

IR: `large eigenvalues` (low energy/k^2)

Finotello, Lahoche, Radpay, Ousmane Samary (2026)

Small Perturbations to the Gaussian Model

Interacting Local Field Theory

The Gaussian model has a strong limitation:

$$\langle \phi_{i_1} \phi_{i_2} \dots \phi_{i_{2n}} \rangle = C_{i_1 i_2} C_{i_3 i_4} \dots C_{i_{2n-1} i_{2n}} + \text{permutations}, \quad \text{and} \quad \langle \phi_{i_1} \phi_{i_2} \dots \phi_{i_{2n+1}} \rangle = 0.$$

Small Perturbations to the Gaussian Model

Interacting Local Field Theory

The Gaussian model has a strong limitation:

$$\langle \phi_{i_1} \phi_{i_2} \dots \phi_{i_{2n}} \rangle = C_{i_1 i_2} C_{i_3 i_4} \dots C_{i_{2n-1} i_{2n}} + \text{permutations}, \quad \text{and} \quad \langle \phi_{i_1} \phi_{i_2} \dots \phi_{i_{2n+1}} \rangle = 0.$$

The natural extension becomes:

$$H[\varphi] = \sum\!\!\!\!\!\!\int_k \varphi(k)(k^2 + m^2)\varphi(-k) + U[\varphi],$$

where

$$U[\varphi] = \sum_{n=3}^{\infty} \frac{u_{2n}}{(2n)!\,p^{n-1}} \underbrace{\delta(k_1 + k_2 + \dots + k_{2n})}_{\texttt{momentum conservation}} \underbrace{\varphi(k_1)\varphi(k_2)\dots\varphi(k_{2n})}_{\texttt{local interactions}}.$$

The empirical spectrum is always imposed as constraint: $G(k_i^2) = \lambda_i$, for $i = 1, \dots, p$. Moreover, we have the continuous limit:

$$\sum\!\!\!\!\!\!\int_k \longrightarrow \int_0^k \mathrm{d}\ell\, \ell\, \rho_{\text{empirical}}(\ell^2).$$

Small Perturbations to the Gaussian Model

Self-energy and LPA

In the interacting theory:

$$G(k^2) = \frac{1}{k^2 + m^2} + \frac{1}{k^2 + m^2} \Sigma(k^2) \frac{1}{k^2 + m^2} + \ldots = \frac{1}{k^2 + m^2 - \Sigma(k^2)}$$

This ultimately leads to $\rho_{\text{empirical}} \neq \rho_{\text{MP}}$.

Small Perturbations to the Gaussian Model

Self-energy and LPA

In the interacting theory:

$$G(k^2) = \frac{1}{k^2 + m^2} + \frac{1}{k^2 + m^2}\Sigma(k^2)\frac{1}{k^2 + m^2} + \dots = \frac{1}{k^2 + m^2 - \Sigma(k^2)}$$

This ultimately leads to $\rho_{\text{empirical}} \neq \rho_{\text{MP}}$.

Using the Local Potential Approximation (LPA):

$$G(k^2) \simeq \frac{1}{Z\, k^2 + m^2_{\text{eff}}},$$

where:

$$Z = 1 - \Sigma'(0) \simeq 1, \qquad m^2_{\text{eff}} = m^2 - \Sigma(0).$$

This restores the Gaussian approximation $m^2_{\text{eff}} = \lambda^{-1}_{\text{max}}$.

Lahoche *et al.* (2020)

Small Perturbations to the Gaussian Model

Self-energy and LPA

In the interacting theory:

$$G(k^2) = \frac{1}{k^2 + m^2} + \frac{1}{k^2 + m^2}\Sigma(k^2)\frac{1}{k^2 + m^2} + \ldots = \frac{1}{k^2 + m^2 - \Sigma(k^2)}$$

This ultimately leads to $\rho_{\text{empirical}} \neq \rho_{\text{MP}}$.

Using the Local Potential Approximation (LPA):

$$G(k^2) \simeq \frac{1}{Z\, k^2 + m^2_{\text{eff}}},$$

where:

$$Z = 1 - \Sigma'(0) \simeq 1, \quad m^2_{\text{eff}} = m^2 - \Sigma(0).$$

This restores the Gaussian approximation $m^2_{\text{eff}} = \lambda^{-1}_{\text{max}}$.

Data Field Theory

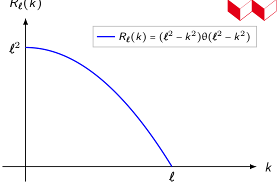
This **Data Field Theory** (DFT) is built as a field theory whose **propagator** reproduces the **empirical spectrum** of the **covariance matrix**.

Lahoche *et al.* (2020)

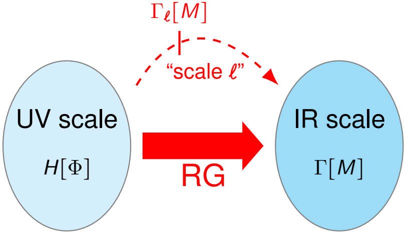
The FRG Machinery

The Average Effective Action

The DFT is “inherently IR” coming from observational data:



$$H[\varphi] \xrightarrow{\text{reg.}} H[\varphi] + \underbrace{\frac{1}{2} \sum\!\!\!\!\!\!\int_k \varphi(k)\, \overbrace{R_\ell(k)}^{\texttt{Litim}}\, \varphi(-k)}_{\Delta H_\ell[\varphi]},$$



then:

$$\Gamma_\ell[M] = \int_x J(x)M(x) - W[J] - \Delta H_\ell[\varphi],$$

where $W[J]$ is the connected generator.

Wetterich (1993) Morris (1993)

The FRG Machinery

The Wetterich-Morris Approach

This ultimately leads to the Wetterich-Morris equation:

$$\ell \frac{\partial \Gamma_\ell[M]}{\partial \ell} = \frac{\ell}{2} \sum\!\!\!\!\!\!\int_k \frac{\partial R_\ell(k)}{\partial \ell} \left(\frac{\delta^2 \Gamma_\ell[M]}{\delta M(k) \delta M(-k)} + R_\ell(k) \mathbb{1} \right)^{-1} (k, -k).$$

The FRG Machinery

The Wetterich-Morris Approach

This ultimately leads to the Wetterich-Morris equation:

$$\ell \frac{\partial \Gamma_\ell[M]}{\partial \ell} = \frac{\ell}{2} \sumint_k \frac{\partial R_\ell(k)}{\partial \ell} \left(\frac{\delta^2 \Gamma_\ell[M]}{\delta M(k) \delta M(-k)} + R_\ell(k) \mathbb{1} \right)^{-1} (k, -k).$$

which can be used in exact LPA:

$$\Gamma_\ell[M] \simeq \frac{1}{2} \sumint_k k^2 M(k) M(-k) + \mathcal{U}_\ell[M],$$

where

$$\mathcal{U}_\ell[M] = \frac{1}{2} \sumint_k u_2 M(k) M(-k) + \sum_{n=3}^{\infty} \frac{u_{2n}}{(2n)! p^{n-1}} \underbrace{\delta\left(k_1 + k_2 + \ldots + k_{2n}\right)}_{\texttt{momentum conservation}} \underbrace{M(k_1) M(k_2) \ldots M(k_{2n})}_{\texttt{local interactions}}.$$

The FRG Machinery
The Flow Equations

From the Wetterich-Morris equation we can derive the flow equations for couplings u_{2n} **as usual**. However,

$$u_2(\ell),\ u_4(\ell),\ u_6(\ell), \dots \xleftarrow{\text{contain}} L(k) = \int_0^k \mathrm{d}\ell\, \ell\, \rho(\ell^2)$$

scale dependence introduced!

The FRG Machinery

The Flow Equations

From the Wetterich-Morris equation we can derive the flow equations for couplings u_{2n} **as usual**. However,

$$u_2(\ell),\ u_4(\ell),\ u_6(\ell),\ldots \xleftarrow{\text{contain}} L(k) = \int_0^k \mathrm{d}\ell\, \ell\, \rho(\ell^2) \quad \text{scale dependence introduced!}$$

The Scale-dependent Canonical Dimensions

We need to go from the usual “time” $t = \ln k$ to

$$\tau = \ln L(k) \Longrightarrow \frac{\mathrm{d}}{\mathrm{d}\tau} = \frac{\mathrm{d}t}{\mathrm{d}\tau}\frac{\mathrm{d}}{\mathrm{d}t} = \underbrace{\frac{L(k)}{k\rho(k^2)}\frac{\mathrm{d}}{\mathrm{d}k}}_{\text{correction}} \Longrightarrow \frac{\mathrm{d}\overline{u}_{2n}}{\mathrm{d}\tau} = -\underbrace{\mathrm{dim}_\tau(u_{2n})}_{\text{scale-dependent}}\ \overline{u}_{2n} + \beta_{2n}^{(\text{quant})}(\overline{u}_4, \overline{u}_6, \ldots)$$

The FRG Machinery

The Flow Equations

From the Wetterich-Morris equation we can derive the flow equations for couplings u_{2n} **as usual**. However,

$$u_2(\ell),\ u_4(\ell),\ u_6(\ell), \ldots \xleftarrow{\text{contain}} L(k) = \int_0^k \mathrm{d}\ell\, \ell\, \rho(\ell^2) \quad \text{scale dependence introduced!}$$

The Scale-dependent Canonical Dimensions

We need to go from the usual "time" $t = \ln k$ to

$$\tau = \ln L(k) \Longrightarrow \frac{\mathrm{d}}{\mathrm{d}\tau} = \frac{\mathrm{d}t}{\mathrm{d}\tau}\frac{\mathrm{d}}{\mathrm{d}t} = \underbrace{\frac{L(k)}{k\rho(k^2)}\frac{\mathrm{d}}{\mathrm{d}k}}_{\texttt{correction}} \Longrightarrow \frac{\mathrm{d}\overline{u}_{2n}}{\mathrm{d}\tau} = -\underbrace{\dim_\tau(u_{2n})}_{\texttt{scale-dependent}}\ \overline{u}_{2n} + \beta_{2n}^{\text{(quant)}}(\overline{u}_4, \overline{u}_6, \ldots)$$

For a field theory in dimension d, we have:

$$\rho(k^2) \sim (k^2)^{\frac{d}{2}-1} \Longrightarrow L(k) \sim (k^2)^{\frac{d}{2}}.$$

This can ultimately be used to define the dimensionless couplings by simple rescaling. For instance, we have $\overline{u}_4 = u_4 (k^2)^{\frac{D-4}{2}} \Rightarrow \dim(u_4) = 4 - D$.

3. Dimensional Phase Transition

The Scale-Dependent Canonical Dimension

The Canonical Dimensions in Practice

We finally need to work with the following scale-dependent canonical dimension:

$$
\begin{aligned}
\dim_\tau(u_2) &= 2\frac{\mathrm{d}t}{\mathrm{d}\tau}, \\
\dim_\tau(u_4) &= 2\left(\frac{\mathrm{d}^2 t}{\mathrm{d}\tau^2}\left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^{-1} + \left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)\left(\frac{1}{2}\frac{\mathrm{d}\ln\rho}{\mathrm{d}t} - 1\right)\right), \\
\dim_\tau(u_{2m}) &= (2-m)\dim_\tau(u_2) + (m-1)\dim_\tau(u_4).
\end{aligned}
$$

The Scale-Dependent Canonical Dimension

The Canonical Dimensions in Practice

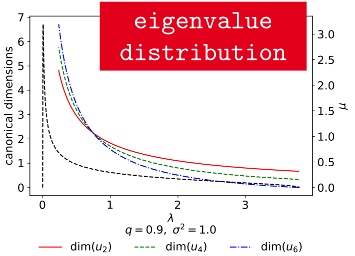
We finally need to work with the following scale-dependent canonical dimension:

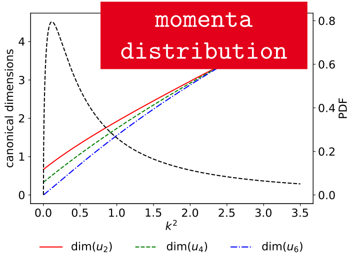
$$\dim_\tau(u_2) = 2\frac{\mathrm{d}t}{\mathrm{d}\tau},$$
$$\dim_\tau(u_4) = 2\left(\frac{\mathrm{d}^2 t}{\mathrm{d}\tau^2}\left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^{-1} + \left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)\left(\frac{1}{2}\frac{\mathrm{d}\ln\rho}{\mathrm{d}t} - 1\right)\right),$$
$$\dim_\tau(u_{2m}) = (2-m)\dim_\tau(u_2) + (m-1)\dim_\tau(u_4).$$

Consider

$$Z \in \mathbb{R}^{n\times p} \sim \mathcal{N}(0, \mathbb{1})$$

i.e. the MP universality class





The Scale-Dependent Canonical Dimension

The Canonical Dimensions in Practice

We finally need to work with the following scale-dependent canonical dimension:

$$\dim_\tau(u_2) = 2\frac{\mathrm{d}t}{\mathrm{d}\tau},$$

$$\dim_\tau(u_4) = 2\left(\frac{\mathrm{d}}{\mathrm{d}}\right.$$

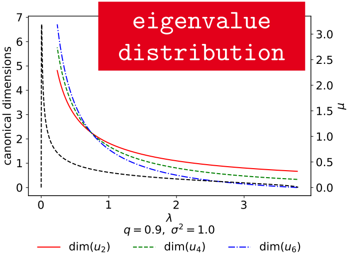
$$\dim_\tau(u_{2m}) = (2-m)\dim_\tau(u_2) + (m-1)\dim_\tau(u_4).$$

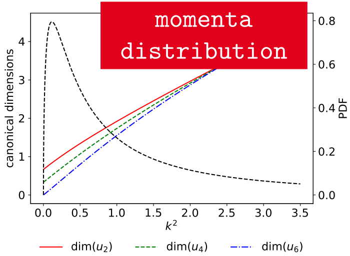
u_2 and u_4 are the only **relevant** couplings in the deep **IR**. u_6 is asymptotically **marginal**.

Consider

$$Z \in \mathbb{R}^{n\times p} \sim \mathcal{N}(0, \mathbb{1})$$

i.e. the MP universality class





The Scale-Dependent Canonical Dimension

Extensive-rank Signals

What happens when

$$X = \beta S + Z, \qquad \beta > 0,$$

with S an extensive-rank signal?

The Scale-Dependent Canonical Dimension

Extensive-rank Signals

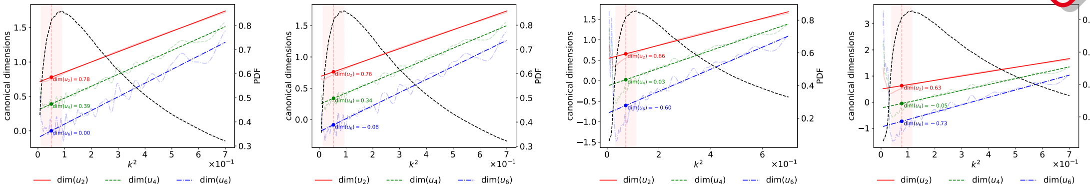
SPOILER ALERT!

What happens when

$$X = \beta\, S + Z, \qquad \beta > 0,$$

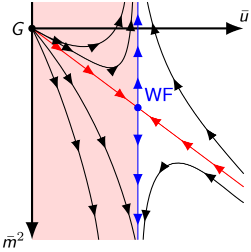
with S an extensive-rank signal?

increasing β

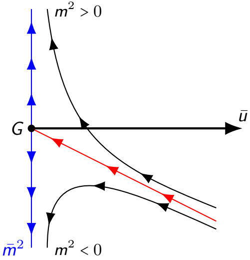


decreasing $\dim_\tau$

Let us consider the canonical $D = 4 - \varepsilon$ expansion:



$\varepsilon > 0$

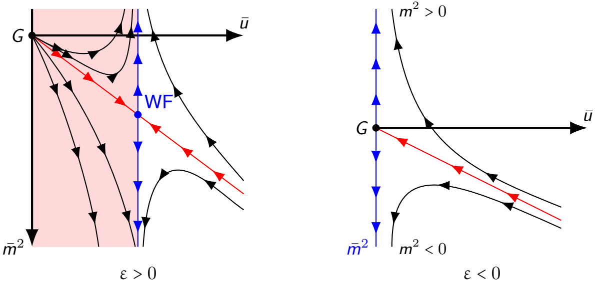


$\varepsilon < 0$

Detecting the Transition

The Fixed Point and Asymptotic Canonical Dimensions

Let us consider the canonical $D = 4 - \varepsilon$ expansion:



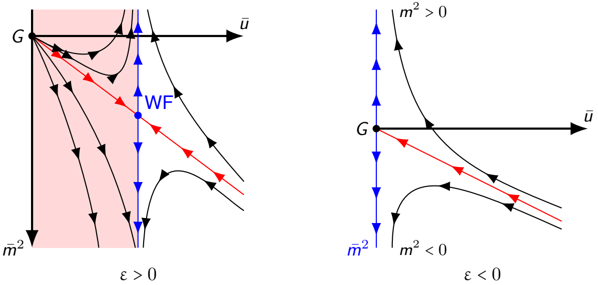
The “pure noise” (MP) regime is described by $D = 3$ (presence of a WF fixed point):

Increasing β effectively increases the spacetime dimension D of the EFT, approaching the Gaussian fixed point.

Detecting the Transition

The Fixed Point and Asymptotic Canonical Dimensions

Let us consider the canonical $D = 4 - \varepsilon$ expansion:



The “pure noise” (MP) regime is described by $D = 3$ (presence of a WF fixed point):

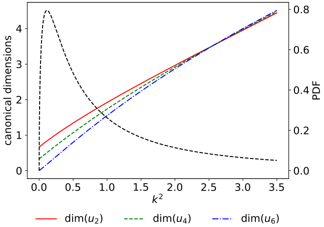
Increasing β effectively increases the spacetime dimension D of the EFT, approaching the Gaussian fixed point.

The higher the signal, the more irrelevant $u_{n>2}$ become! We can detect the instability threshold by looking at the canonical dimensions of the couplings, then!

Signal Thresholds

A Model-agnostic Definition of Limit of Detection

Using the **character** of u_4 and u_6, we can define a model-agnostic notion of **signal thresholds**:

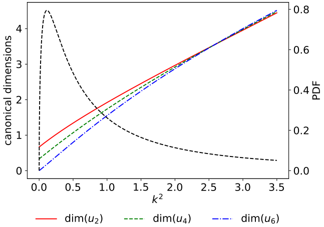


Symbol	Definition
β_t	**Limit of detection.** $\beta > 0 \mid \dim_\tau(u_6) \neq 0$.
β_c	**Critical threshold.** $\beta > 0 \mid \dim_\tau(u_4) = 0$.
β_O	**Optimal threshold.** $\beta > 0 \mid \dim_\tau(u_4)$ first minimum

Signal Thresholds

A Model-agnostic Definition of Limit of Detection

Using the **character** of u_4 and u_6, we can define a model-agnostic notion of **signal thresholds**:



Symbol	Definition
β_t	**Limit of detection.** $\beta > 0 \mid \dim_\tau(u_6) \neq 0$.
β_c	**Critical threshold.** $\beta > 0 \mid \dim_\tau(u_4) = 0$.
β_O	**Optimal threshold.** $\beta > 0 \mid \dim_\tau(u_4)$ first minimum

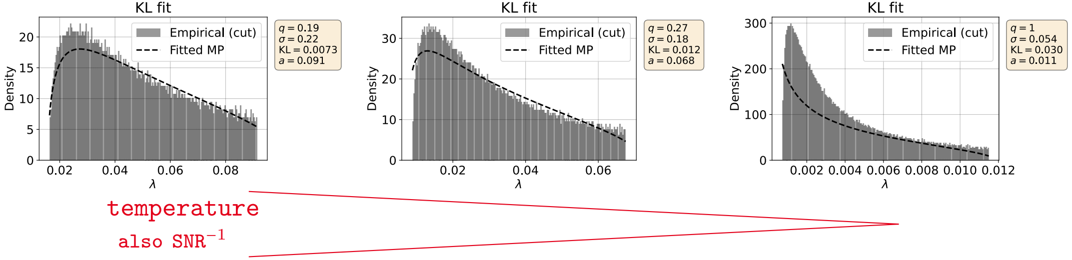
In chemometrics terms, we can roughly identify β_t with the limit of detection and β_c with the limit of quantitation.

The Ising Phase Transition

Grounding Signal Thresholds in Physics

Consider a Monte Carlo simulation of a 2D $N \times N$ Ising model: (start random, evolve via Metropolis-Hastings)

$$M = \sum_{i=1}^{N} S_i \quad \longrightarrow \quad \underbrace{K = 1 - \frac{\langle M^4 \rangle}{3 \langle M^2 \rangle^2}}_{\text{Binder cumulant}} .$$



Finotello, Lahoche, Radpay, Ousmane Samary (2026)

The Ising Phase Transition

Detecting Onsager's Critical Temperature via Anomaly Detection

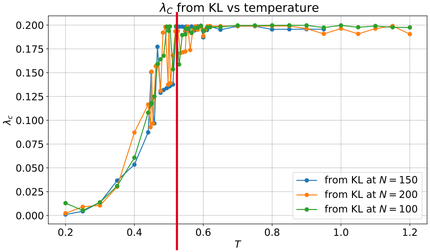
Two methods to detect T_c (departure from Gaussianity – Onsager solution):

The Ising Phase Transition

Detecting Onsager's Critical Temperature via Anomaly Detection

Two methods to detect T_c (departure from Gaussianity – Onsager solution):

1. **Information Theoretic (`baseline`):** use the $\lambda_{\max}$ of the KL divergence fit with a MP distribution



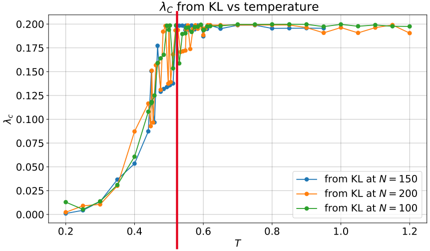
$T_c = 0.53 \pm 0.07$

The Ising Phase Transition

Detecting Onsager's Critical Temperature via Anomaly Detection

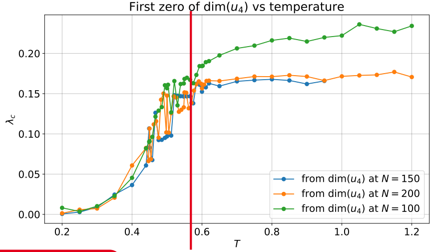
Two methods to detect T_c (departure from Gaussianity – Onsager solution):

1. **Information Theoretic (`baseline`):** use the $\lambda_{\max}$ of the KL divergence fit with a MP distribution
2. **RG-based (`proposed`):** use the critical threshold β_c of the corresponding EFT (i.e. u_4 becomes marginal)



$T_c = 0.53 \pm 0.07$

$T_c^{(\text{Onsager})} = 0.57$



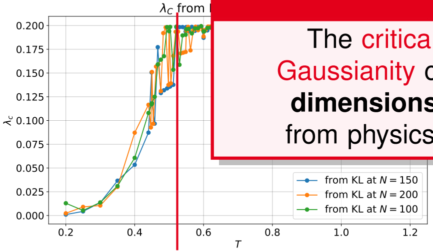
$T_c = 0.58 \pm 0.01$

The Ising Phase Transition

Detecting Onsager's Critical Temperature via Anomaly Detection

Two methods to detect T_c (departure from Gaussianity – Onsager solution):

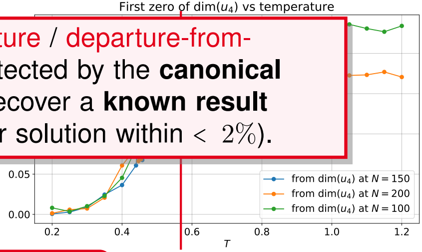
1. **Information Theoretic (`baseline`):** use the $\lambda_{\max}$ of the KL divergence fit with a MP distribution
2. **RG-based (`proposed`):** use the critical threshold β_c of the corresponding EFT (i.e. u_4 becomes marginal)



$T_c = 0.53 \pm 0.07$

Dimensional Criterion

The critical temperature / departure-from-Gaussianity can be detected by the **canonical dimensions**, which recover a **known result** from physics (Onsager solution within $< 2\%$).



$T_c^{(\text{Onsager})} = 0.57$

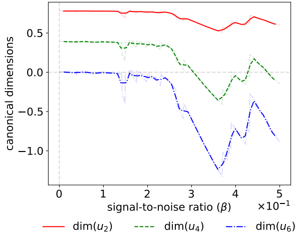
$T_c = 0.58 \pm 0.01$

4. Implications

Realistic Data

Detecting Signal and Anomaly Sources

Consider a realistic image:



$$X \sim \beta \times \text{ image} + \mathcal{N}(0, \mathbb{1})$$

leads to:

- $\beta_t \approx 0.15$ (limit of detection)
- $\beta_c \approx 0.32$ (critical threshold)
- $\beta_O \approx 0.37$ (optimal threshold)

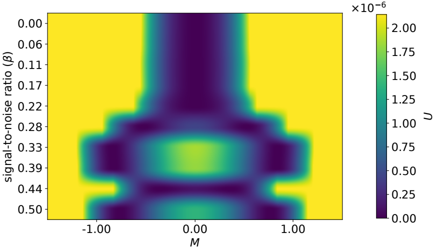
against the **BBP transition** $\beta_{\text{BBP}} \sim 0.97$ ($q = 0.9$):

*detection is **still possible** even below the BBP transition!*

Realistic Data

Detecting Signal and Anomaly Sources

Consider a realistic image:



$$X \sim \beta \times \text{ image} + \mathcal{N}(0, \mathbb{1})$$

leads to:

- $\beta_t \approx 0.15$ (limit of detection)
- $\beta_c \approx 0.32$ (critical threshold)
- $\beta_O \approx 0.37$ (optimal threshold)

against the **BBP transition** $\beta_{\text{BBP}} \sim 0.97$ ($q = 0.9$):

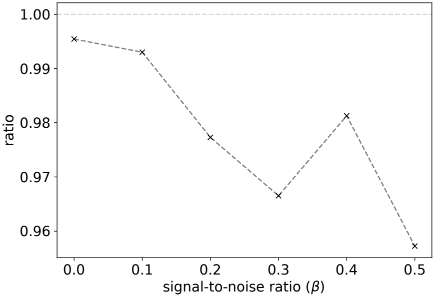
*detection is **still possible** even below the BBP transition!*

$$U \simeq \frac{1}{2} u_2^{(\text{IR})} M^2 + \frac{1}{4!} u_4^{(\text{IR})} M^4$$

Realistic Data

Detecting Signal and Anomaly Sources

Consider a realistic image:



Porter-Thomas delocalisation of the eigenvectors (UV / IR ratio)!

Finotello, Lahoche, Ousmane Samary (2025)

$$X \sim \beta \times \text{ image} + \mathcal{N}(0, \mathbb{1})$$

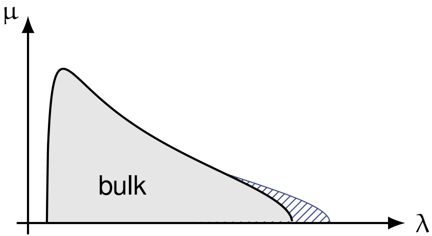
leads to:

- $\beta_t \approx 0.15$ (limit of detection)
- $\beta_c \approx 0.32$ (critical threshold)
- $\beta_O \approx 0.37$ (optimal threshold)

against the **BBP transition** $\beta_{\text{BBP}} \sim 0.97$ ($q = 0.9$):

*detection is **still possible** even below the BBP transition!*

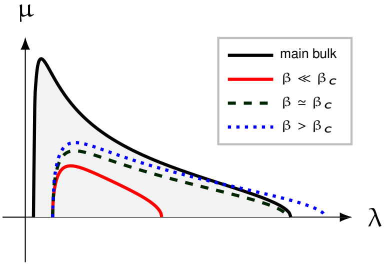
Back to Random Matrix Theory
Multiple Signal/Anomaly Sources



The original idea was to detect **when and if** a continuous **extensive rank distribution** starts to **emerge from the bulk**.

Back to Random Matrix Theory

Multiple Signal/Anomaly Sources

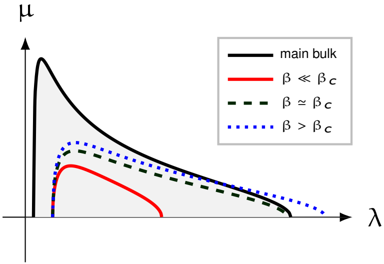


The original idea was to detect **when and if** a continuous **extensive rank distribution** starts to **emerge from the bulk.**

We now have a relation between **signal-to-noise ratio** and this phenomenon, and a way to relate this to a **dimensional phase transition** of a corresponding EFT.

Back to Random Matrix Theory

Multiple Signal/Anomaly Sources



The original idea was to detect **when and if** a continuous **extensive rank distribution** starts to **emerge from the bulk.**

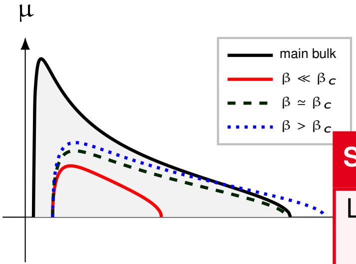
We now have a relation between **signal-to-noise ratio** and this phenomenon, and a way to relate this to a **dimensional phase transition** of a corresponding EFT.

This works for **all** extensive-rank signals/anomalies

$$S = S_0 + \underbrace{\sum_{i=1}^{L} \tilde{S}_i(\omega_i)}_{\text{sensor response, other noise sources, etc.}} \quad \Rightarrow \quad Y = \beta S_0 + \left(Z + \beta \sum_{i=1}^{L} \tilde{S}_i(\omega_i) \right) = \beta S_0 + \tilde{Z}_L(\beta)$$

Back to Random Matrix Theory

Multiple Signal/Anomaly Sources



The original idea was to detect **when and if** a continuous **extensive rank distribution** starts to **emerge from the bulk.**

y have a relation between **signal-to-noise ratio** s phenomenon, and a way to relate this to a corresponding

Signal Sources

Looking at the progression of the **canonical dimensions**, we can estimate L (the number of **extensive rank sources**)!

This works for **all** extensive-

$$S = S_0 + \underbrace{\sum_{i=1}^{L} \tilde{S}_i(\omega_i)}_{\text{sensor response, other noise sources, etc.}} \quad \Rightarrow \quad Y = \beta S_0 + \left(Z + \beta \sum_{i=1}^{L} \tilde{S}_i(\omega_i) \right) = \beta S_0 + \tilde{Z}_L(\beta)$$

Some Conclusions and Future Directions

From Here to Real-World Applications

The **FRG** enables (with a strong grounding in physics):

Anomaly Detection: via a dimensional phase transition of the corresponding EFT, we are able to detect a perturbation of the MP universality class

Signal Characterisation: following the evolution of the SNR, we can characterise the presence of different signal/anomaly sources, and their relative magnitude.

Some Conclusions and Future Directions

From Here to Real-World Applications

The **FRG** enables (with a strong grounding in physics):

Anomaly Detection:	via a dimensional phase transition of the corresponding EFT, we are able to detect a perturbation of the MP universality class
Signal Characterisation:	following the evolution of the SNR, we can characterise the presence of different signal/anomaly sources, and their relative magnitude.

We need to further explore its potential applications. For instance:

- The FRG defines a notion of **distance** to detect the anomaly: how can this be used for **uncertainty quantification**?
- **GenAI** models are trained on high-dimensional data to retrieve/reconstruct a signal: can we use the FRG to **quantify the quality of the reconstruction**?

Thank you!

CEA SACLAY
91 191 Gif-sur-Yvette Cedex
France
riccardo.finotello@cea.fr