

# A weakly modelled view of the joint compact-binary mass plane: population structure and spectral-siren cosmology

Viola De Renzis, Alessandro Agapito, Michele Mancarella



*Neuvième assemblée générale du GdR Ondes Gravitationnelles,  
Lyon, 06/10/2026*

# Cosmology with gravitational waves

Hubble law in the local universe

$$H_0 \stackrel{z \ll 1}{\sim} \frac{cz}{d_L}$$

Not measurable from GWs alone (degeneracy with mass)

Directly measured from GW signal

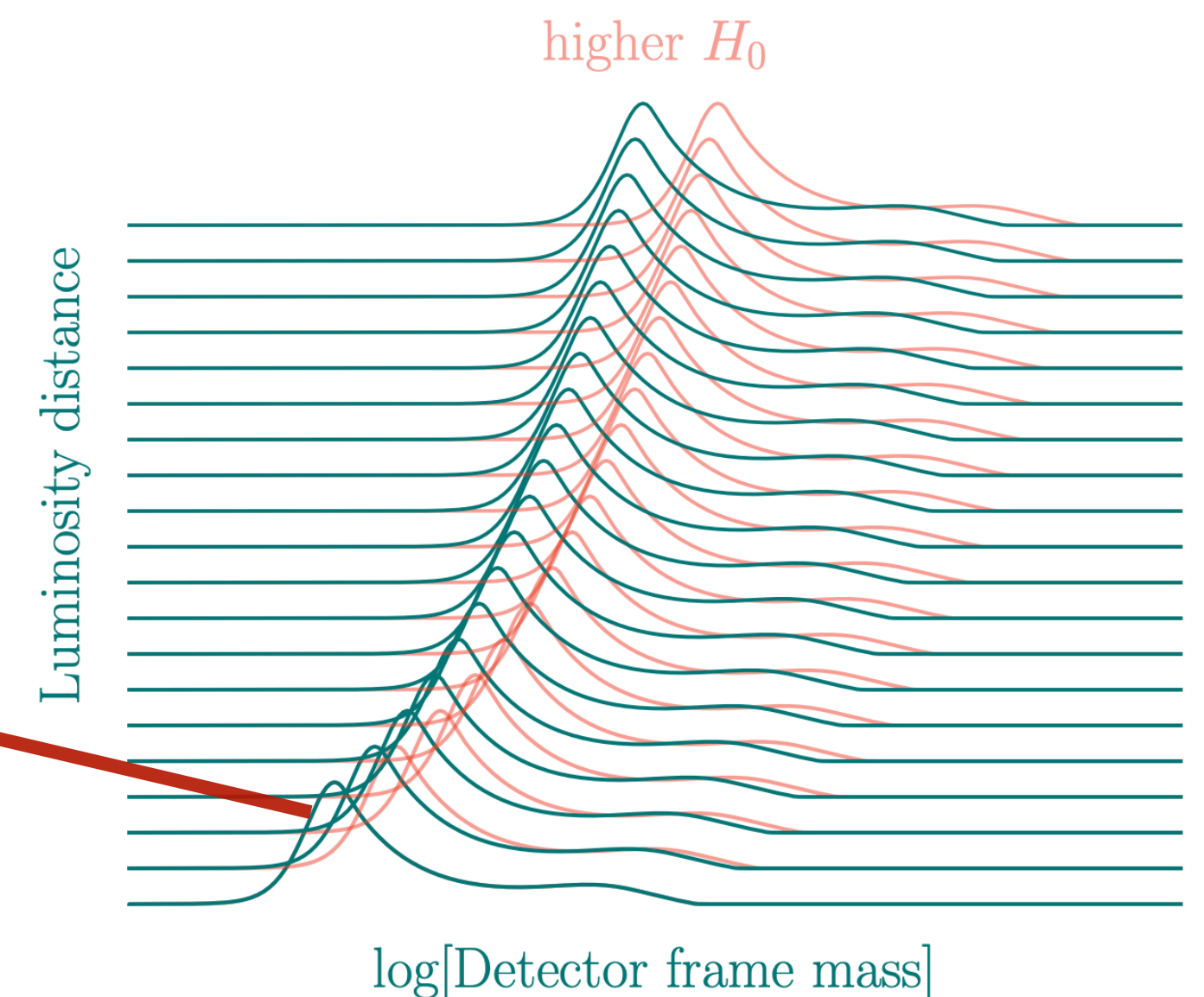
$$h(f) \propto \frac{\mathcal{M}_z^{5/3} f_{\text{obs}}^{2/3}}{d_L}$$

## Standard sirens with CBCs

We need **independent measurement** of redshift to do cosmology  
(bright, dark, spectral sirens.)

- ✱ **Spectral sirens:** Use features in the source-frame mass/redshift distribution to break the mass-redshift degeneracy

$$m_{\text{det}} = m_{\text{source}}(1 + z)$$



# Motivations of this work

## 1 An accurate mass modelling is essential for cosmology

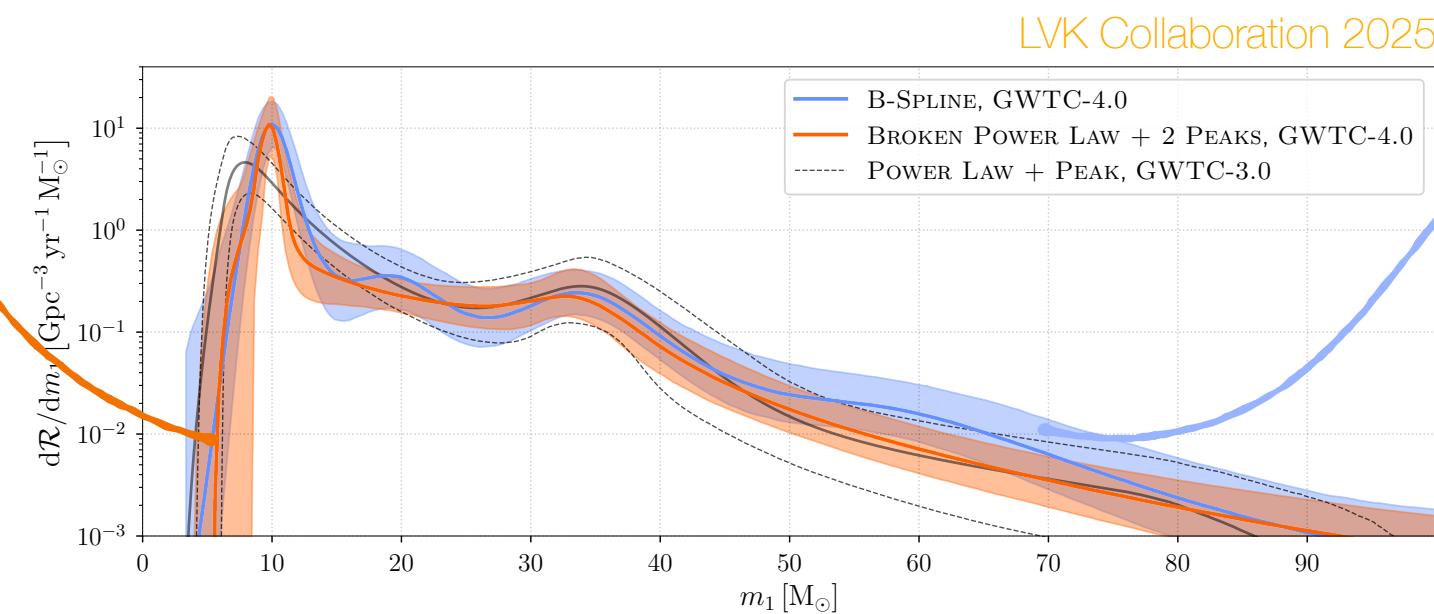
$p_{\text{pop}}(m_1, m_2 | \lambda_m)$   $\rightarrow$  Mismodelling the **source-frame** mass model propagates into the inference of  $z$  and  $H_0$

## 2 Population inference is moving towards more flexible models

### Strongly parametrised methods

Easy interpretability but low flexibility

$\lambda_m \rightarrow$  direct astrophysical interpretation  
(e.g. peaks, cutoffs, slopes)



### Flexible, data-driven methods

High flexibility but low interpretability

$\lambda_m \rightarrow$  lack a direct physical interpretation

Recent spectral-siren analyses have used 1D splines

Tagliazzucchi+2026, Gennari+2026,  
Bertheas+2026, J.Li+2026

Our goal: reconstruct the joint 2D mass distribution with a model that lets the data determine where local structure is needed

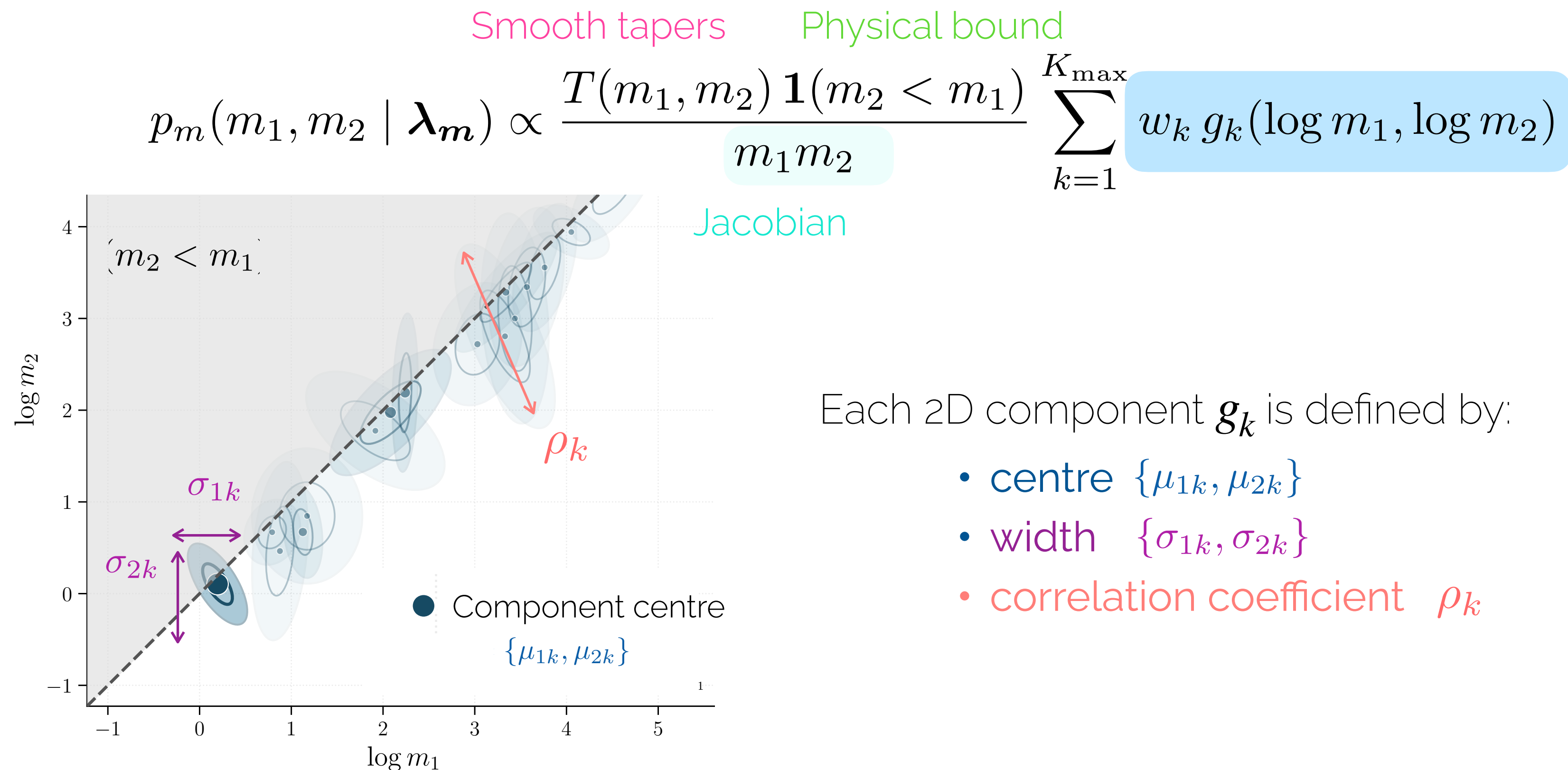
## 3 What is the definition of a feature in a weakly parametrised model?

Need for a **reproducible** definition of population features

# A flexible model for the source-frame masses

$$p_{\text{pop}}(m_1, m_2, z \mid \boldsymbol{\lambda}) = p_m(m_1, m_2 \mid \boldsymbol{\lambda}_m) p_z(z \mid \boldsymbol{\lambda}_z, \boldsymbol{\lambda}_c)$$

**Mass model:** Weighted mixture of correlated 2D Gaussian components in the joint plane defined in the space  $\mathbf{x} = (\log m_1, \log m_2)$



## Redshift model

Madau-Dickinson merger rate evolution

$$p_z(z \mid \boldsymbol{\lambda}_z, \boldsymbol{\lambda}_c) \propto \psi(z \mid \boldsymbol{\lambda}_z) \frac{dV_c}{dz} \frac{1}{1+z}$$

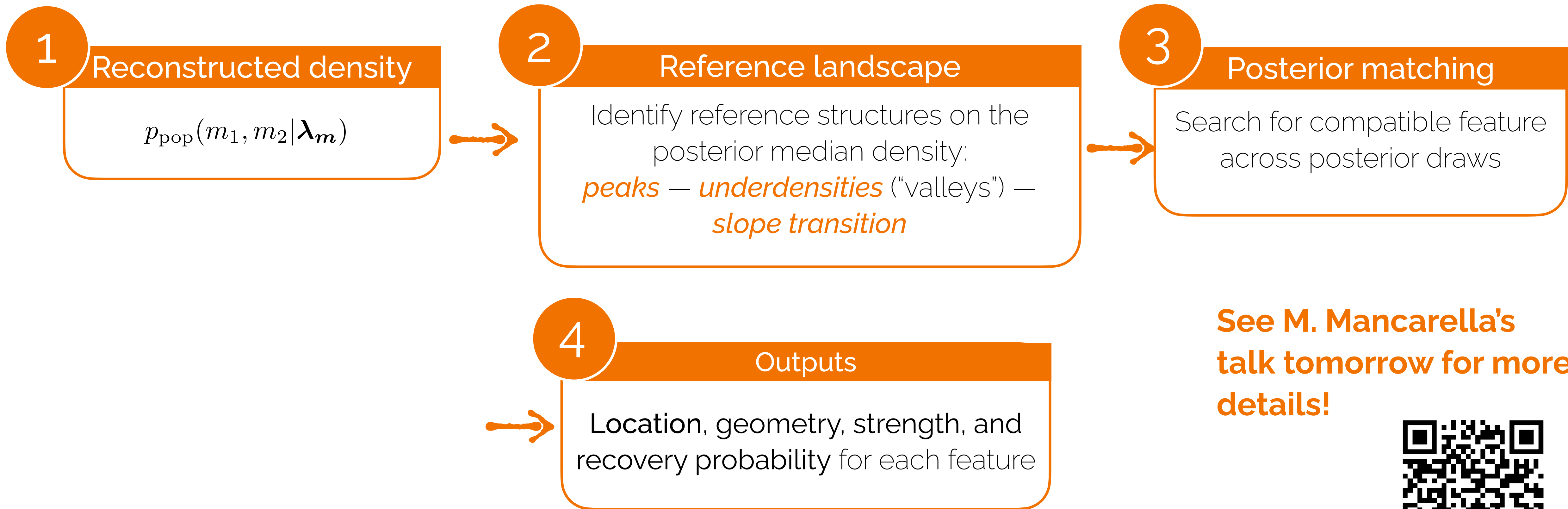
Cosmology is jointly inferred with the mass distribution

- No mass-redshift evolution
- Spin distribution follows the parameter estimation prior

**How to define a population feature in a weakly modelled reconstruction?**

# From density reconstruction to population features

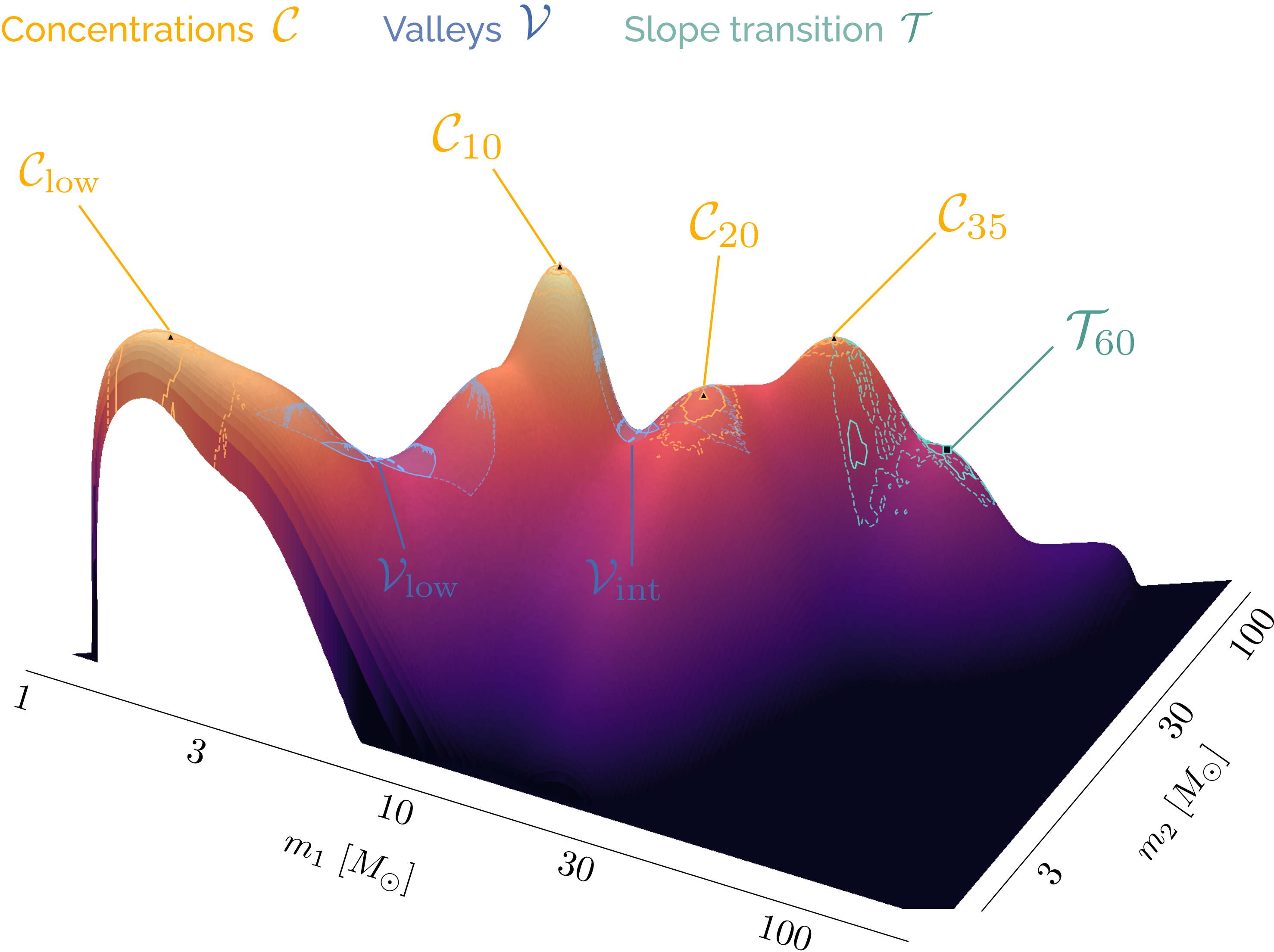
**posterior-landscape:** deterministic and reproducible identification of 2D population features.  
Independent 1D catalogues are constructed from the  $m_1$  and  $m_2$  marginals



See M. Mancarella's  
talk tomorrow for more  
details!

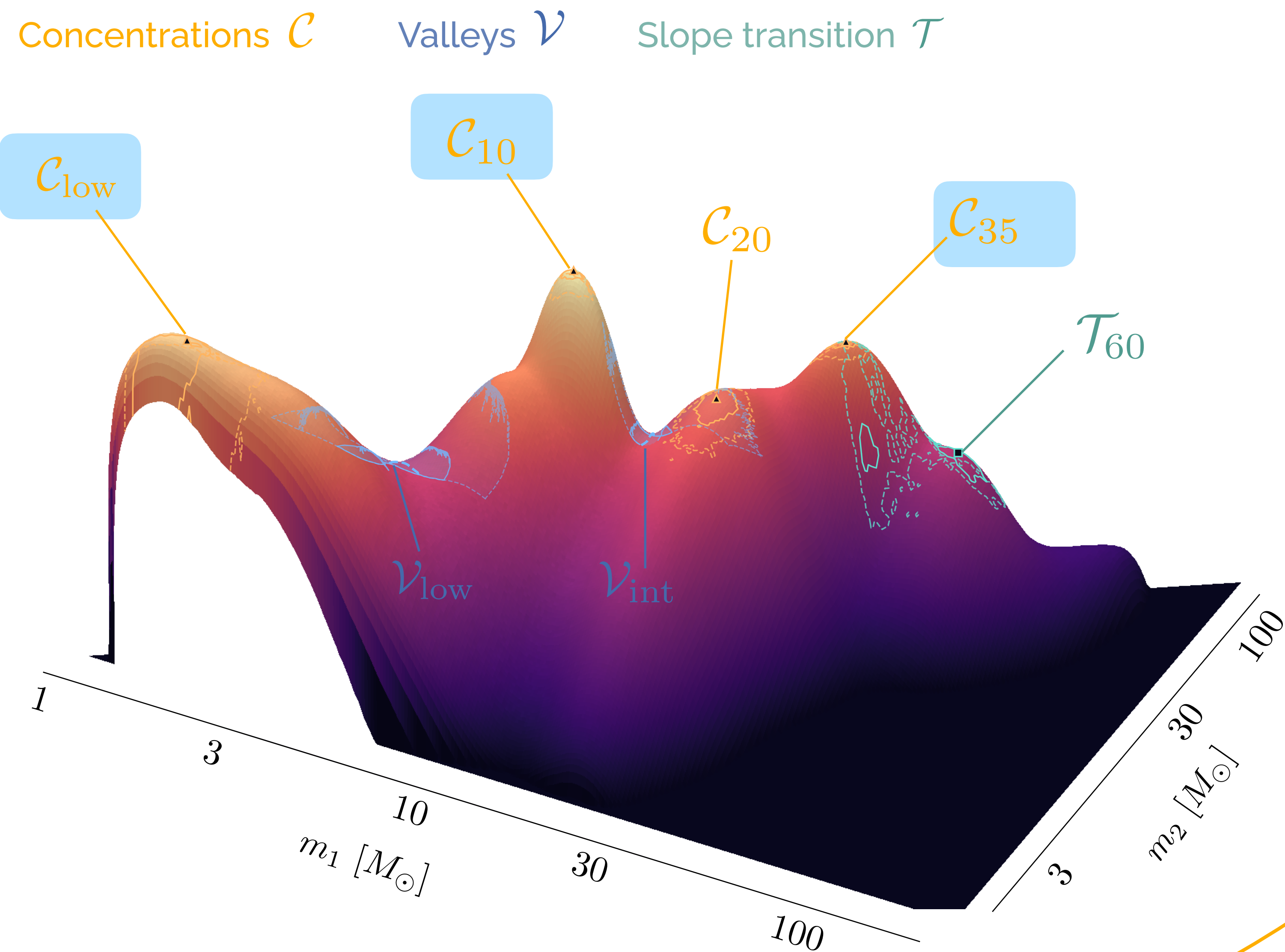


# Population features in the GWTC-5.0 mass distribution



| Feature                    | $P_{\text{rec}}$ | Strength                   |
|----------------------------|------------------|----------------------------|
| $\mathcal{C}_{\text{low}}$ | 99.1%            | $A_p \simeq 0.985$         |
| $\mathcal{V}_{\text{low}}$ | 52.2%            | $A_V \simeq 0.886$         |
| $\mathcal{C}_{10}$         | 99.7%            | $A_p \simeq 0.997$         |
| $\mathcal{V}_{\text{int}}$ | 73.5%            | $A_V \simeq 0.933$         |
| $\mathcal{C}_{20}$         | 95.4%            | $A_p \simeq 0.551$         |
| $\mathcal{C}_{35}$         | 99.9%            | $A_p \simeq 0.857$         |
| $\mathcal{T}_{60}$         | 36.7%            | $\Delta\alpha \simeq 4.92$ |

# Population features in the GWTC-5.0 mass distribution



| Feature                    | $P_{\text{rec}}$ | Strength                   |
|----------------------------|------------------|----------------------------|
| $\mathcal{C}_{\text{low}}$ | 99.1%            | $A_p \simeq 0.985$         |
| $\mathcal{V}_{\text{low}}$ | 52.2%            | $A_V \simeq 0.886$         |
| $\mathcal{C}_{10}$         | 99.7%            | $A_p \simeq 0.997$         |
| $\mathcal{V}_{\text{int}}$ | 73.5%            | $A_V \simeq 0.933$         |
| $\mathcal{C}_{20}$         | 95.4%            | $A_p \simeq 0.551$         |
| $\mathcal{C}_{35}$         | 99.9%            | $A_p \simeq 0.857$         |
| $\mathcal{T}_{60}$         | 36.7%            | $\Delta\alpha \simeq 4.92$ |

**\*\*Stable and prominent features**

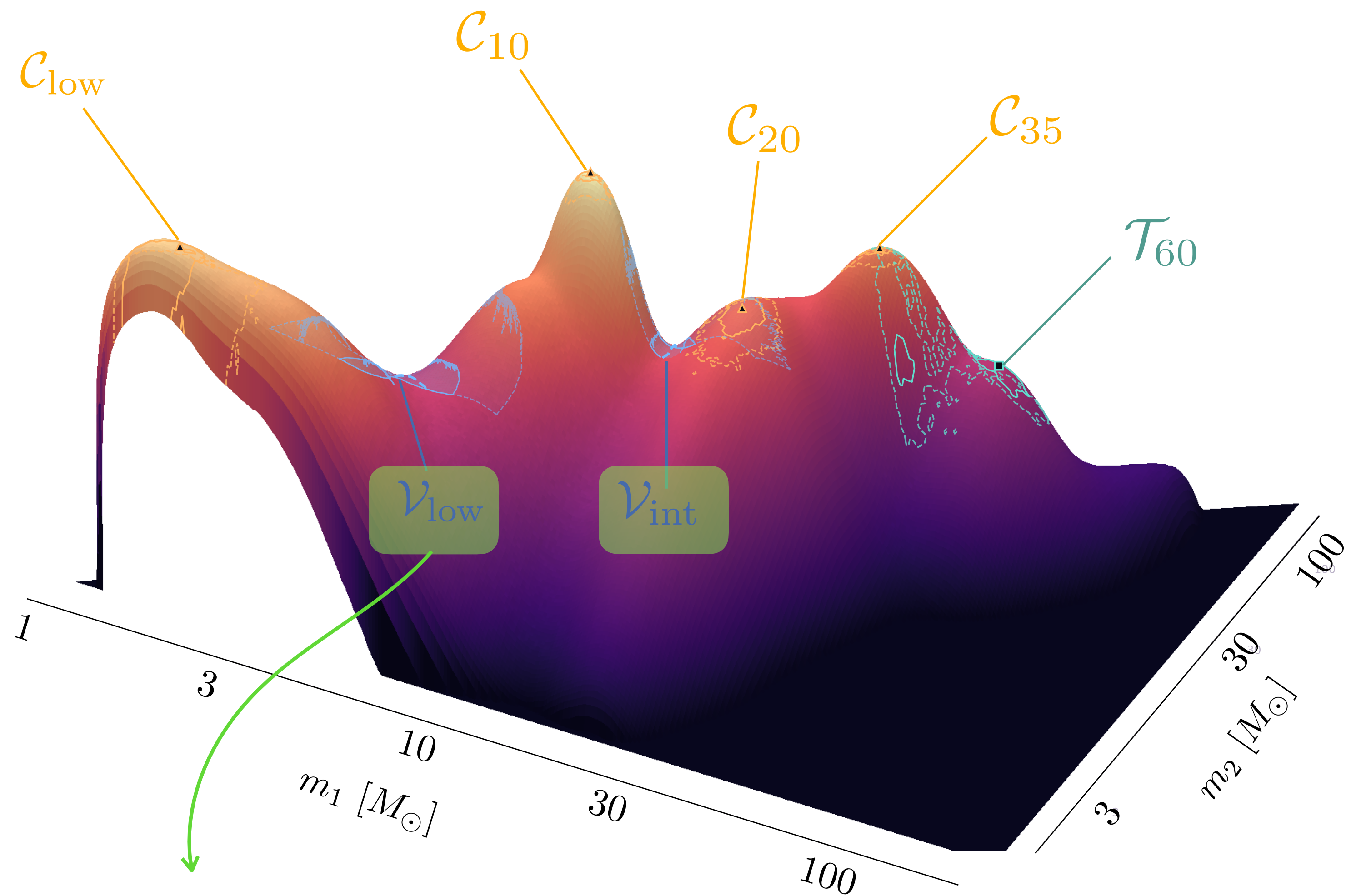
$m_1$  projection: peak (50.5%), slope change (49.5%)  
 $m_2$  projection: peak (99%)

# Population features in the GWTC-5.0 mass distribution

Concentrations  $\mathcal{C}$

Valleys  $\mathcal{V}$

Slope transition  $\mathcal{T}$



$m_1$  and  $m_2$  projections: stable dips ( $P_{\text{rec}} > 99\%$ )

| Feature                    | $P_{\text{rec}}$ | Strength                   |
|----------------------------|------------------|----------------------------|
| $\mathcal{C}_{\text{low}}$ | 99.1%            | $A_p \simeq 0.985$         |
| $\mathcal{V}_{\text{low}}$ | 52.2%            | $A_V \simeq 0.886$         |
| $\mathcal{C}_{10}$         | 99.7%            | $A_p \simeq 0.997$         |
| $\mathcal{V}_{\text{int}}$ | 73.5%            | $A_V \simeq 0.933$         |
| $\mathcal{C}_{20}$         | 95.4%            | $A_p \simeq 0.551$         |
| $\mathcal{C}_{35}$         | 99.9%            | $A_p \simeq 0.857$         |
| $\mathcal{T}_{60}$         | 36.7%            | $\Delta\alpha \simeq 4.92$ |

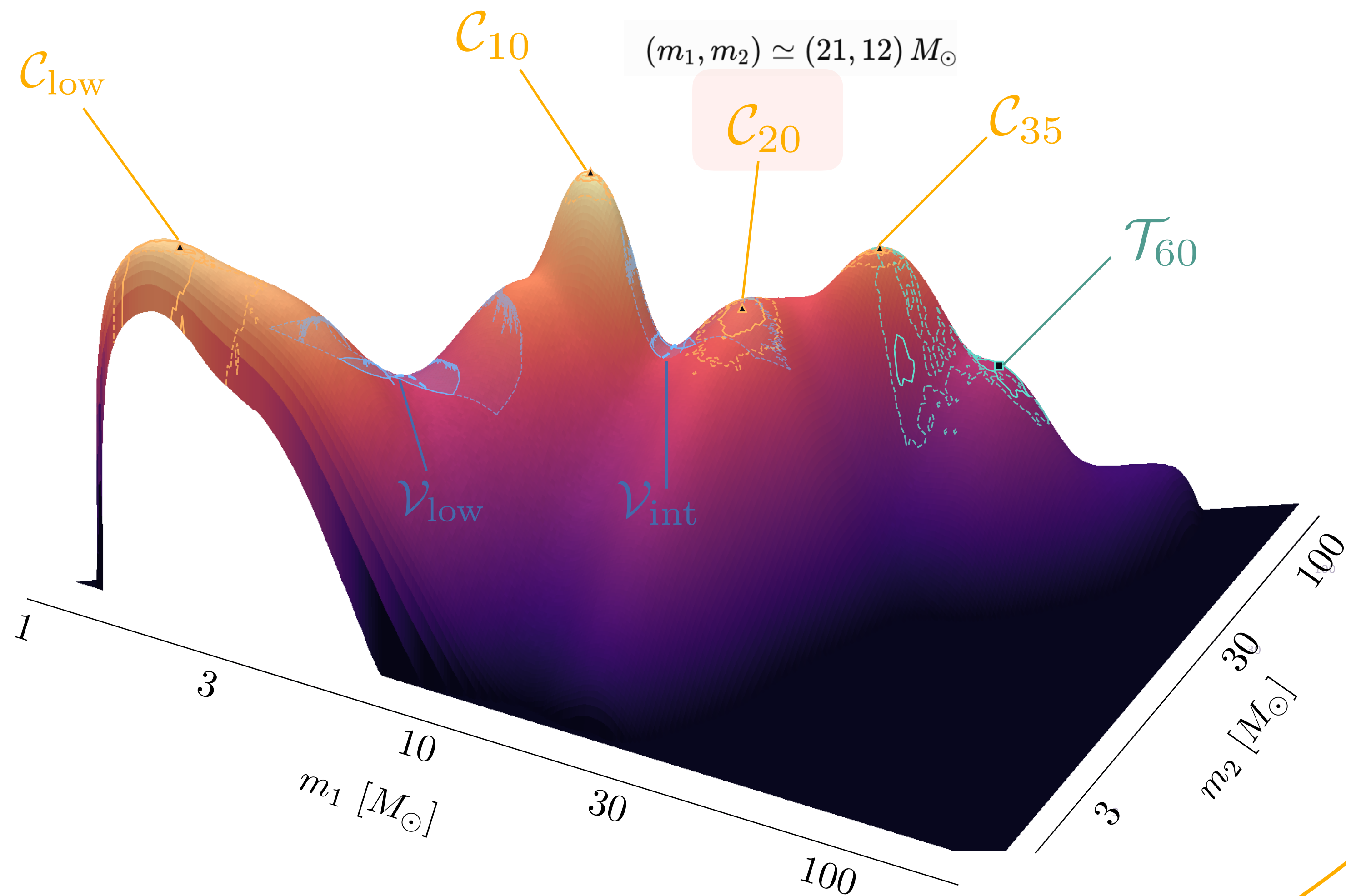
\*\*Less stable but deep

# Population features in the GWTC-5.0 mass distribution

Concentrations  $\mathcal{C}$

Valleys  $\mathcal{V}$

Slope transition  $\mathcal{T}$



$m_1$  projection: peak (89.9%), slope change (10.1%)

$m_2$  projection: slope change (86%), peak (14%)

| Feature                    | $P_{\text{rec}}$ | Strength                   |
|----------------------------|------------------|----------------------------|
| $\mathcal{C}_{\text{low}}$ | 99.1%            | $A_p \simeq 0.985$         |
| $\mathcal{V}_{\text{low}}$ | 52.2%            | $A_V \simeq 0.886$         |
| $\mathcal{C}_{10}$         | 99.7%            | $A_p \simeq 0.997$         |
| $\mathcal{V}_{\text{int}}$ | 73.5%            | $A_V \simeq 0.933$         |
| $\mathcal{C}_{20}$         | 95.4%            | $A_p \simeq 0.551$         |
| $\mathcal{C}_{35}$         | 99.9%            | $A_p \simeq 0.857$         |
| $\mathcal{T}_{60}$         | 36.7%            | $\Delta\alpha \simeq 4.92$ |

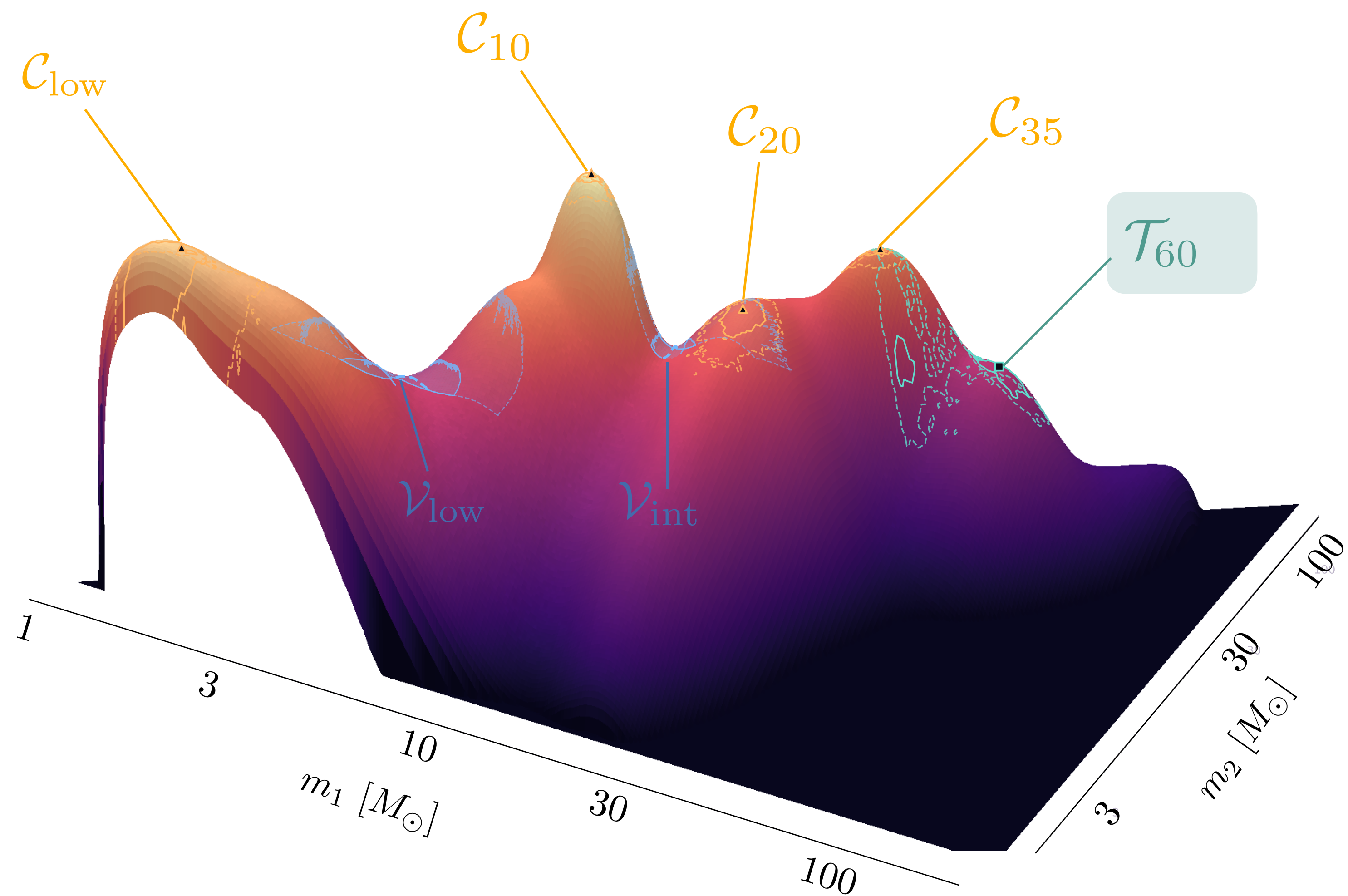
**\*\*Recurrent but weak feature**

# Population features in the GWTC-5.0 mass distribution

Concentrations  $\mathcal{C}$

Valleys  $\mathcal{V}$

Slope transition  $\mathcal{T}$



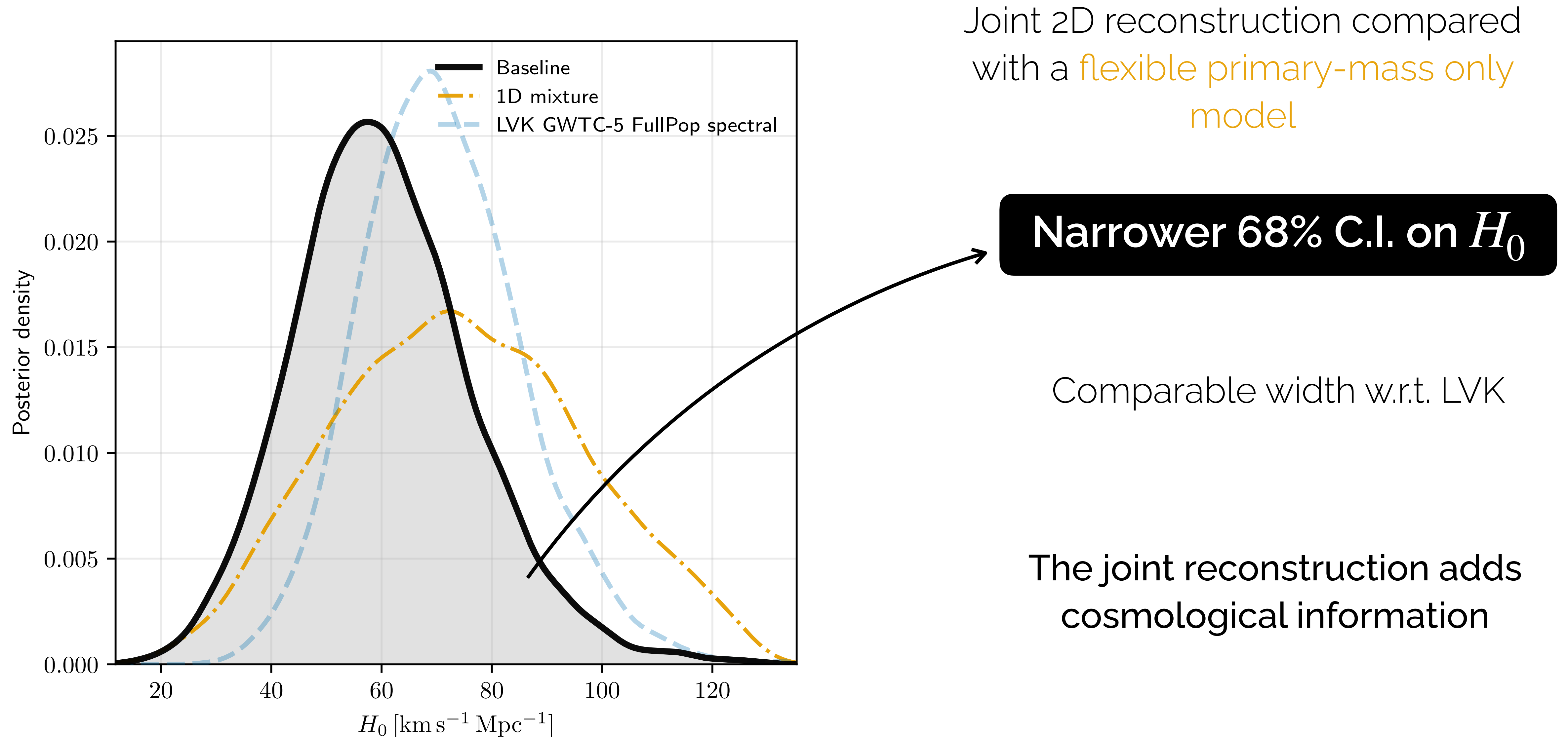
$m_1$  projection: slope change (86.4%)

| Feature                    | $P_{\text{rec}}$ | Strength                   |
|----------------------------|------------------|----------------------------|
| $\mathcal{C}_{\text{low}}$ | 99.1%            | $A_p \simeq 0.985$         |
| $\mathcal{V}_{\text{low}}$ | 52.2%            | $A_V \simeq 0.886$         |
| $\mathcal{C}_{10}$         | 99.7%            | $A_p \simeq 0.997$         |
| $\mathcal{V}_{\text{int}}$ | 73.5%            | $A_V \simeq 0.933$         |
| $\mathcal{C}_{20}$         | 95.4%            | $A_p \simeq 0.551$         |
| $\mathcal{C}_{35}$         | 99.9%            | $A_p \simeq 0.857$         |
| $\mathcal{T}_{60}$         | 36.7%            | $\Delta\alpha \simeq 4.92$ |

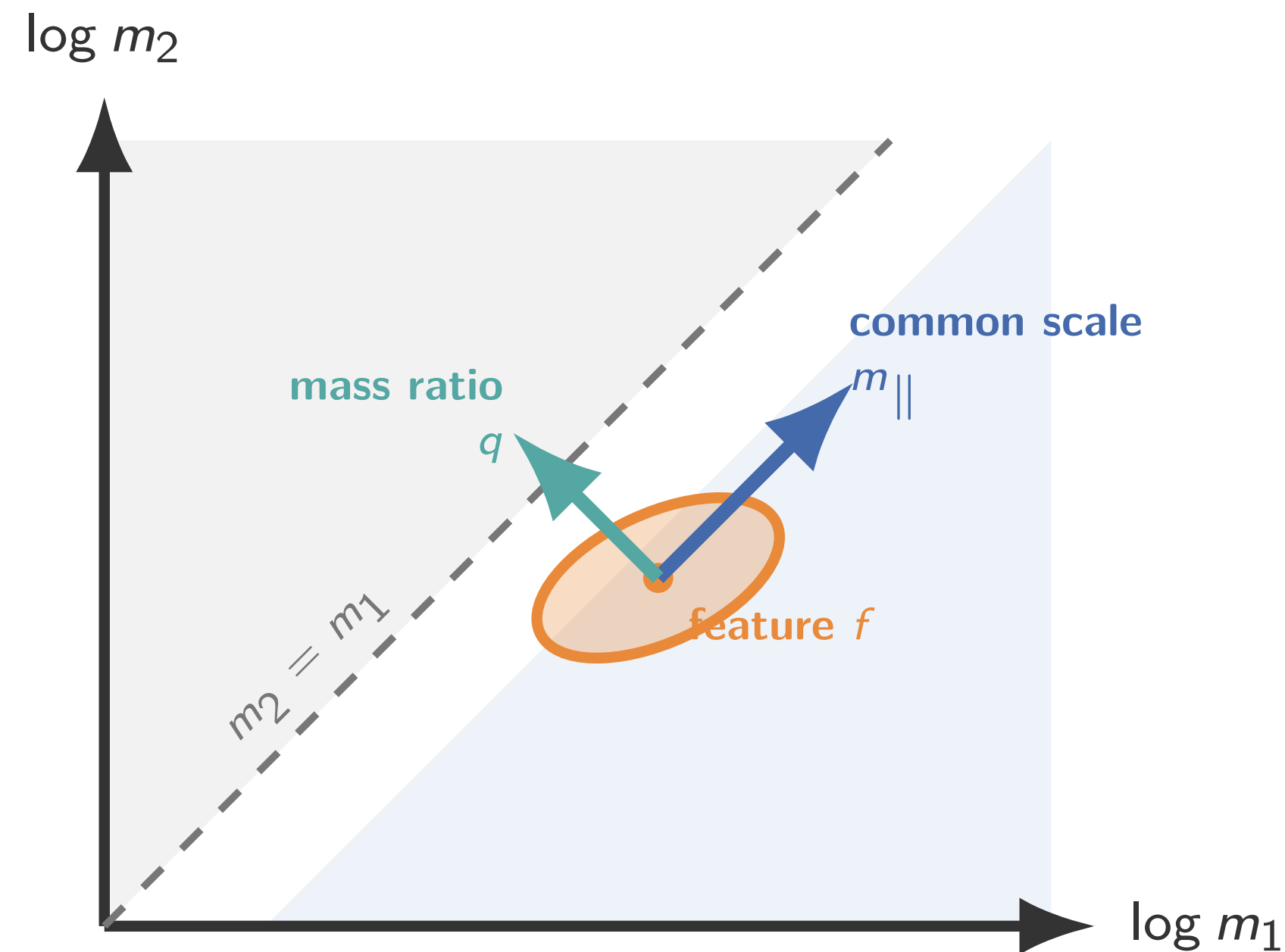
Tentative high mass slope change

**The joint mass plane is needed for the correct interpretation of 1D features**

# Joint structure improves spectral-siren precision



# The joint mass plane can carry information on $H_0$ beyond a coherent rescaling of the two masses



$$m_{\parallel,f} = \sqrt{x_{1,f} x_{2,f}}, \quad q_f = \frac{x_{2,f}}{x_{1,f}}.$$

**Common scale  $m_{\parallel}$ :** coherent shift of both component masses.

**Mass ratio  $q$ :** position along the orthogonal direction.

## Dominant common-scale anchors

$$\mathcal{C}_{10} > \mathcal{C}_{35}$$

Their dependence on  $H_0$  is carried mainly by a coherent displacement of both component masses.

$$\mathcal{T}_{60}$$

is an additional common-scale anchor when its joint morphology is recovered.

## Residual mass ratio information

$$\mathcal{C}_{20}, \quad \mathcal{V}_{\text{int}}$$

$\mathcal{C}_{20}$  and  $\mathcal{V}_{\text{int}}$  are weaker overall, but contain the clearest genuinely two-dimensional contribution.

# Take home messages

## I. **A flexible joint reconstruction**

A correlated Gaussian mixture reconstructs the joint source-frame distribution, inferred simultaneously with cosmology

## II. **A reproducible interpretation of population structure**

**Posterior-landscape:** a new framework to identify population structures directly from the reconstructed density

## III. **The joint mass plane contains cosmological information**

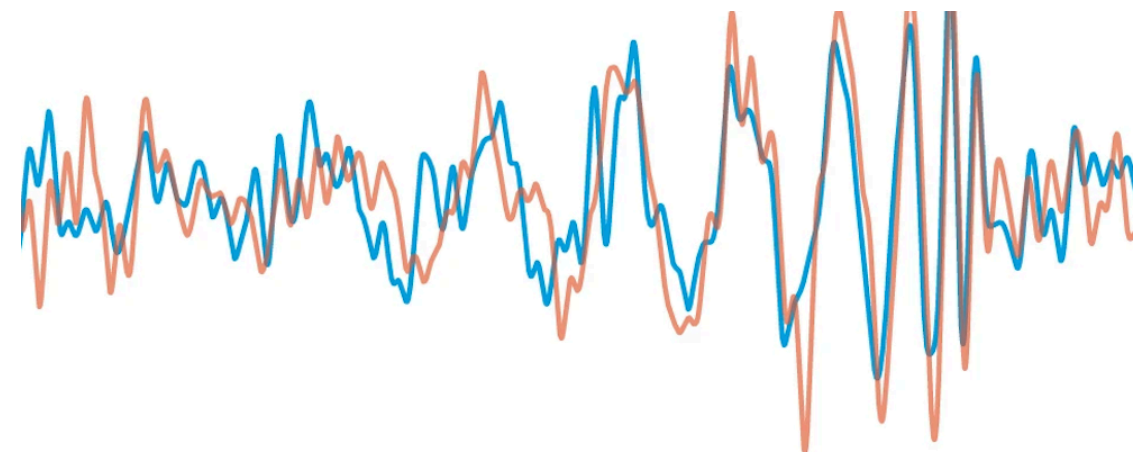
A 2D flexible reconstruction retains much of the spectral-siren precision of strongly parameterized models

The 2D geometry can contain information that is not captured by a coherent rescaling of both masses.

**Back up slides**

# Cosmology with gravitational waves

Thanks to GR, **GWs** provides direct measurement of  $d_L$



$$h(f) \propto \frac{\mathcal{M}_z^{5/3} f_{\text{obs}}^{2/3}}{d_L}$$

GW signal

$$\mathcal{M}_z = \mathcal{M}_c(1 + z)$$

Redshifted  
chirp mass

Assuming a flat LCDM model:

$$d_L = \frac{(1 + z)c}{H_0} \int_0^z \frac{dz'}{\sqrt{\Omega_m(1 + z')^3 + \Omega_\Lambda}} \quad \begin{matrix} z \ll 1 \\ \approx \end{matrix} \boxed{\frac{cz}{H_0}}$$

Hubble law in  
the local  
universe

From GWs alone, we can not measure  $z$   
(degeneracy with mass)



**Need independent measurement of redshift  
to do cosmology**

# Standard sirens with CBCs

$$H_0 \sim \frac{cz}{d_L}$$

need **complementary methods** to measure  $z$ !

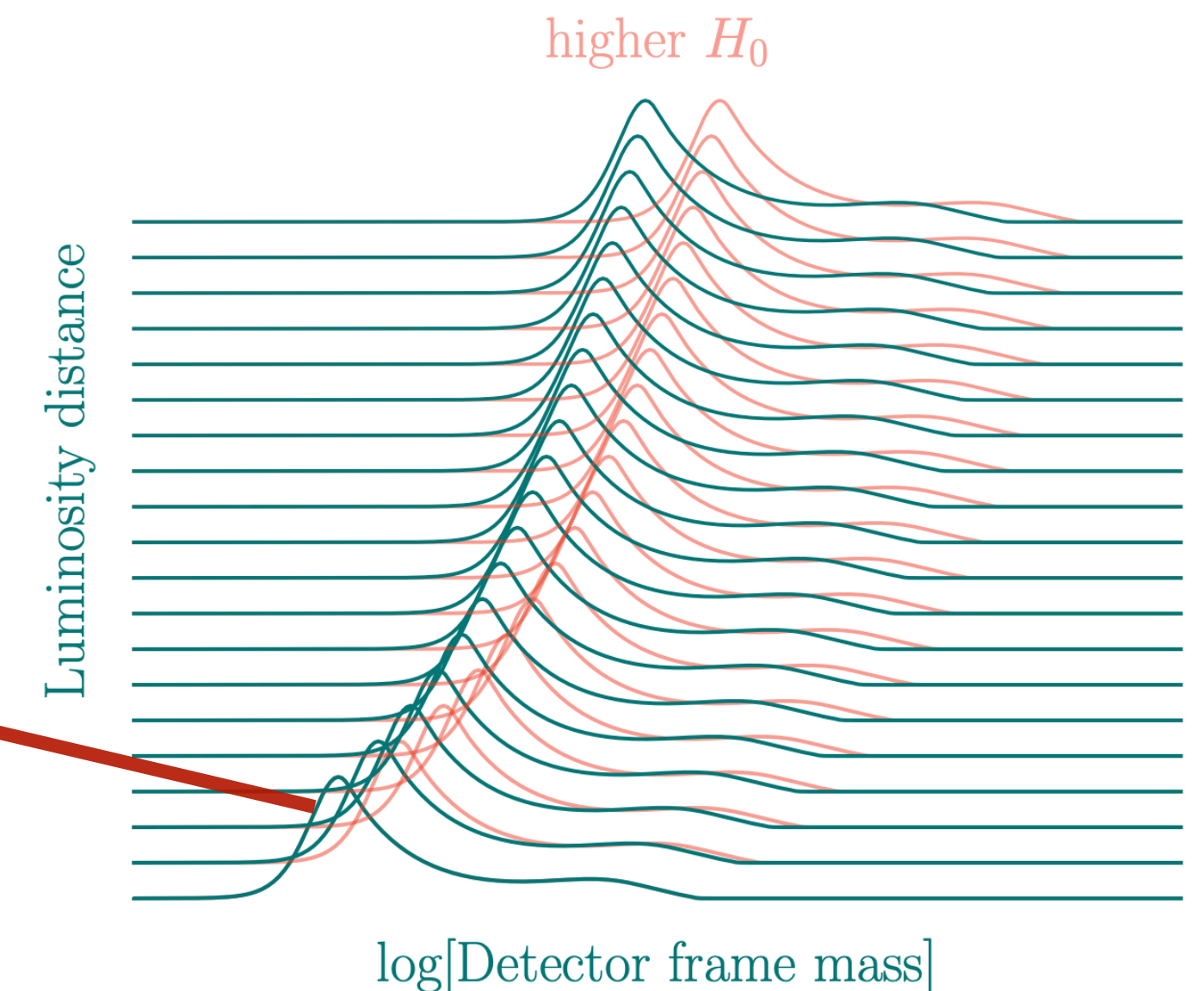
from GW signal

- ✱ **Bright sirens:** identification of the host galaxy through an associated EM counterpart
- ✱ **Dark sirens:** Galaxy surveys can be used to identify possible hosts in the GW localization volume

- ✱ **Spectral sirens:** Use features in the source-frame mass/redshift distribution to break the mass-redshift degeneracy

$$m_{\text{det}} = m_{\text{source}}(1 + z)$$

GW-only



# Hierarchical bayesian inference

$\theta$  **Single-event parameters:** masses, spins, distance

$\lambda$  **Population parameters:** spectral index of mass distribution, cutoffs,  $H_0$ ,  $\Omega_m$

$$p(\lambda \mid \{\mathcal{D}\}) \propto \frac{\pi(\lambda)}{\xi(\lambda)^{N_{\text{obs}}}} \prod_{i=1}^{N_{\text{obs}}} \int d\theta_i \frac{p(\theta_i \mid \mathcal{D}_i)}{\pi_{\text{PE}}(\theta_i)} p_{\text{pop}}(\theta_i \mid \lambda)$$

*Population prior*      *Single-event posterior*  
*Single-event prior*

Mandel+ 2019, Vitale+ 2020

*Selection effects*

$$\xi(\lambda) = \int p_{\text{pop}}(\theta \mid \lambda) p_{\text{det}}(\theta) d\theta$$

*Population model*

$$p_{\text{pop}}(\theta \mid \lambda) = p(m_1, m_2 \mid \lambda_m) p(z \mid \lambda_z, \lambda_c) \pi_{\text{PE}}(\Omega)$$

*Masses*

*Redshift*

$$\lambda_c = (H_0, \Omega_m)$$

They both  
require MC  
integration

# What if you don't marginalize over $\theta$ ?

$\theta$  **Single-event parameters:** masses, spins, distance

$\lambda$  **Population parameters:** spectral index of mass distribution, cutoffs,  $H_0$ ,  $\Omega_m$

$$p(\lambda \mid \{\mathcal{D}\}) \propto \frac{\pi(\lambda)}{\xi(\lambda)^{N_{\text{obs}}}} \prod_{i=1}^{N_{\text{obs}}} \int d\theta_i \frac{p(\theta_i \mid \mathcal{D}_i)}{\pi_{\text{PE}}(\theta_i)} p_{\text{pop}}(\theta_i \mid \lambda)$$

Sampling the full dimensional posterior using Hamiltonian Monte Carlo:

$$p(\lambda, \theta \mid \{\mathcal{D}\}) \propto \frac{\pi(\lambda)}{\xi(\lambda)^{N_{\text{obs}}}} \prod_{i=1}^{N_{\text{obs}}} p(\theta_i \mid \mathcal{D}_i) \frac{p_{\text{pop}}(\theta_i \mid \lambda)}{\pi_{\text{PE}}(\theta_i)}$$

Mancarella+2025

Still MC integrals

Continuous GMM

**Dimensionality:**  $N_\theta \times N_{\text{obs}} + N_\lambda$

# Prior construction for the mass mixture

## 1. Mixture complexity

We use a finite stick-breaking construction with

$$K_{\max} = 15.$$

The component weights are generated as

$$v_k \sim \text{Beta}(1, \alpha_{\text{DP}}), \quad w_k \propto v_k \prod_{\ell < k} (1 - v_\ell),$$

with concentration hyperprior

$$\log(\alpha_{\text{DP}}^{-1}) \sim \mathcal{N}(-1.5, 0.5^2).$$

Although  $K_{\max} = 15$  components are available, the stick-breaking prior allows unnecessary components to receive negligible weight. The effective complexity is therefore inferred from the data.

## 2. Component locations

The Gaussian centres are defined in the  $(\log m_1, \log m_2)$  plane:

$$\mu_{2k} = \mu_{1k} + \log q_{\mu,k}, \quad 0 < q_{\mu,k} < 1.$$

For  $\theta \in \{\mu_1, q_\mu\}$ , bounded priors are implemented through

$$\theta_k = \theta_{\text{low}} + (\theta_{\text{up}} - \theta_{\text{low}}) \text{sigmoid}(u_{\theta,k}),$$

$$u_{\theta,k} \sim \mathcal{N}(0, 1).$$

The broad data-adaptive bounds define the accessible mass region without imposing where population features should appear.

## 3. Component widths

For  $j \in \{1, 2\}$ , the component widths are

$$\sigma_{jk} = L_j + \tau_j e^{\epsilon_{jk}}, \quad \epsilon_{jk} \sim \mathcal{N}(0, 0.5^2).$$

The minimum width is tied to the typical single-event uncertainty:

$$L_j = 0.25 \, Q_{25\%} \left( \left\{ \text{Std}[\log m_j]_i \right\} \right).$$

This resolution floor prevents the reconstruction from fitting structures much narrower than the mass resolution supported by the observations.

## 4. Correlation and shape

The correlation coefficients are generated through

$$r_k \sim \text{Beta}(3, 3),$$

and mapped to

$$\rho_k = \sqrt{1 - 0.25^2} (2r_k - 1).$$

This allows each component to describe structures that are not aligned with either mass axis.

A soft aspect-ratio regularization,

$$a_{\max} = 1, \quad s_a = 0.5, \quad \lambda_a = 0.5,$$

discourages extremely elongated or needle-like components without forcing the Gaussians to be circular.

# A flexible model for the source-frame masses

Mass model:

$$p_m(m_1, m_2 \mid \boldsymbol{\lambda}_m) \propto \frac{T(m_1, m_2) \mathbf{1}(m_2 < m_1)}{m_1 m_2} \sum_{k=1}^{K_{\max}} w_k g_k(\log m_1, \log m_2)$$

- Mass distribution modelled as a weighted mixture of finite components defined in  $\boldsymbol{x} = (\log m_1, \log m_2)$
- $g_k$ : each component  $k$  is a **correlated 2D Gaussian** in the joint plane

$$g_k(x_1, x_2) = \mathcal{N}(x_1 \mid \mu_{1k}, \sigma_{1k}^2) \mathcal{N}(x_2 \mid \mu_{2|1,k}(x_1), \sigma_{2|1,k}^2),$$

where

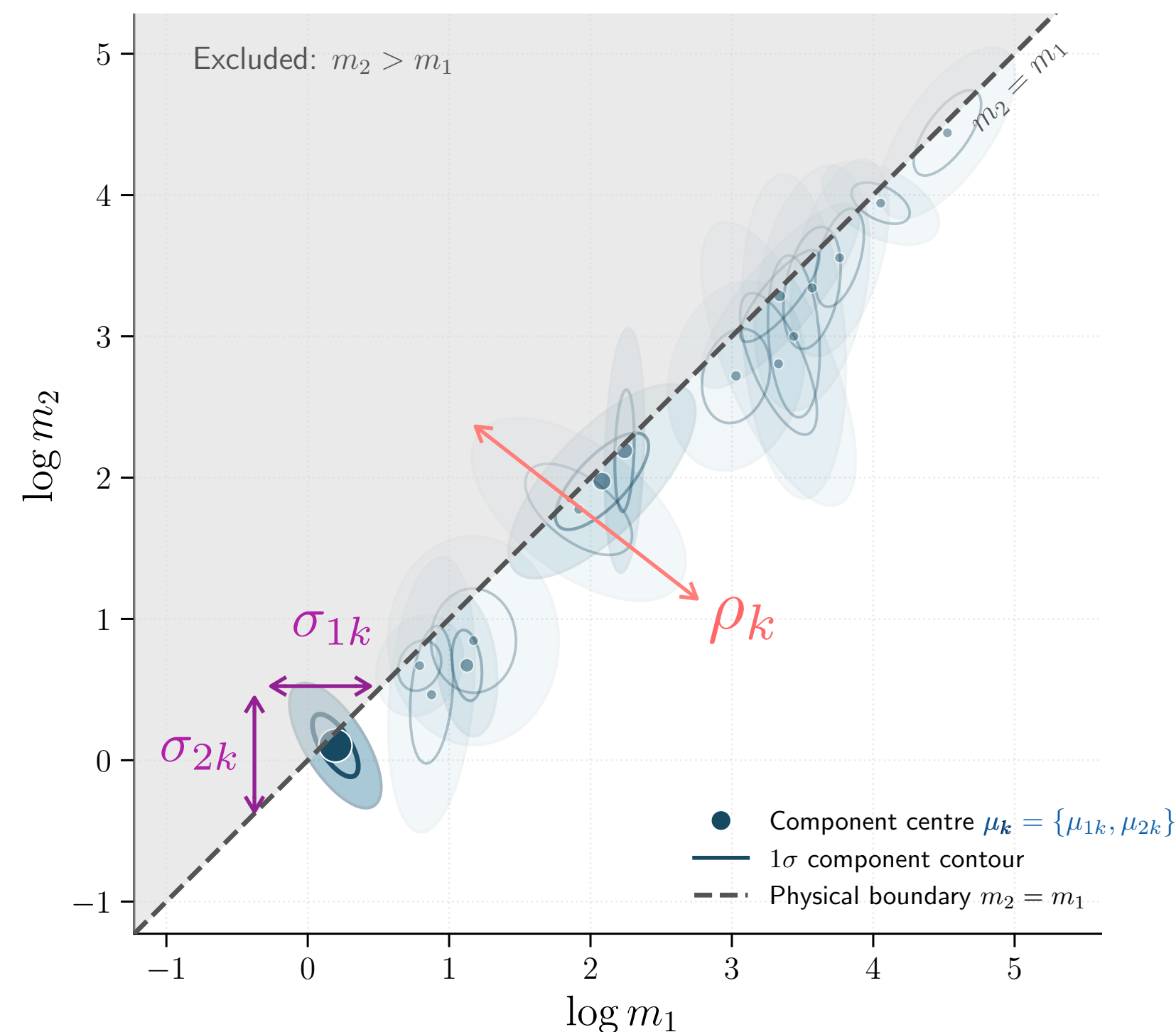
$$\mu_{2|1,k}(x_1) = \mu_{2k} + \rho_k \frac{\sigma_{2k}}{\sigma_{1k}} (x_1 - \mu_{1k})$$

$$\sigma_{2|1,k} = \sigma_{2k} \sqrt{1 - \rho_k^2}$$

Centre of the 2D gaussian

Width of the 2D gaussian

Correlation coefficient

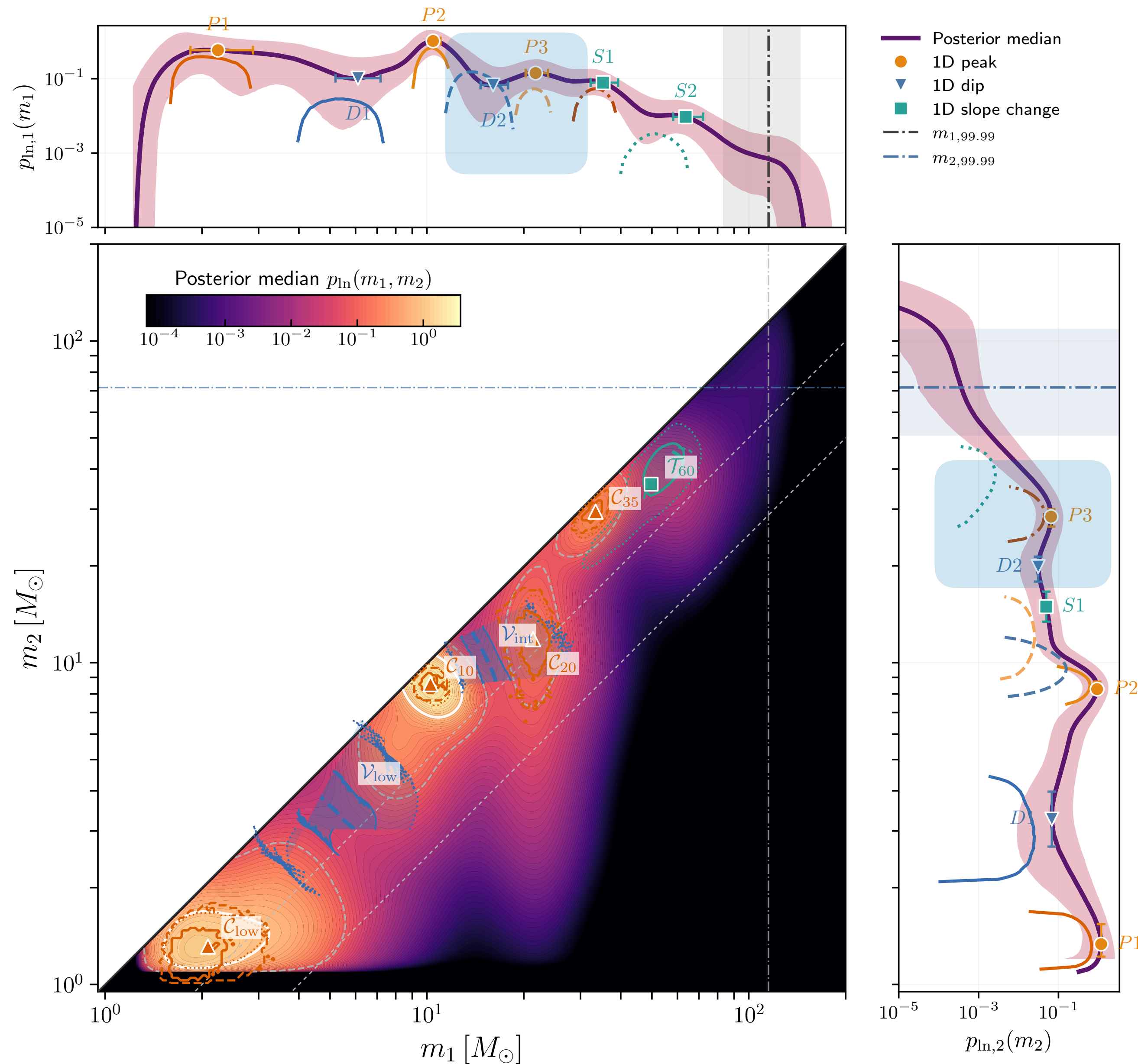


The secondary coordinate of the 2D Gaussian is parametrised w.r.t. the primary one as:

$$\mu_{2k} = \mu_{1k} + \log q_{\mu,k}, \quad 0 < q_{\mu,k} < 1$$

- The mixture weights do not correspond directly to physical population fractions since they are defined before applying mass tapers and the condition  $m_2 < m_1$

# Projections on 1D marginals



Distinct features in 2D that are merged in 1D

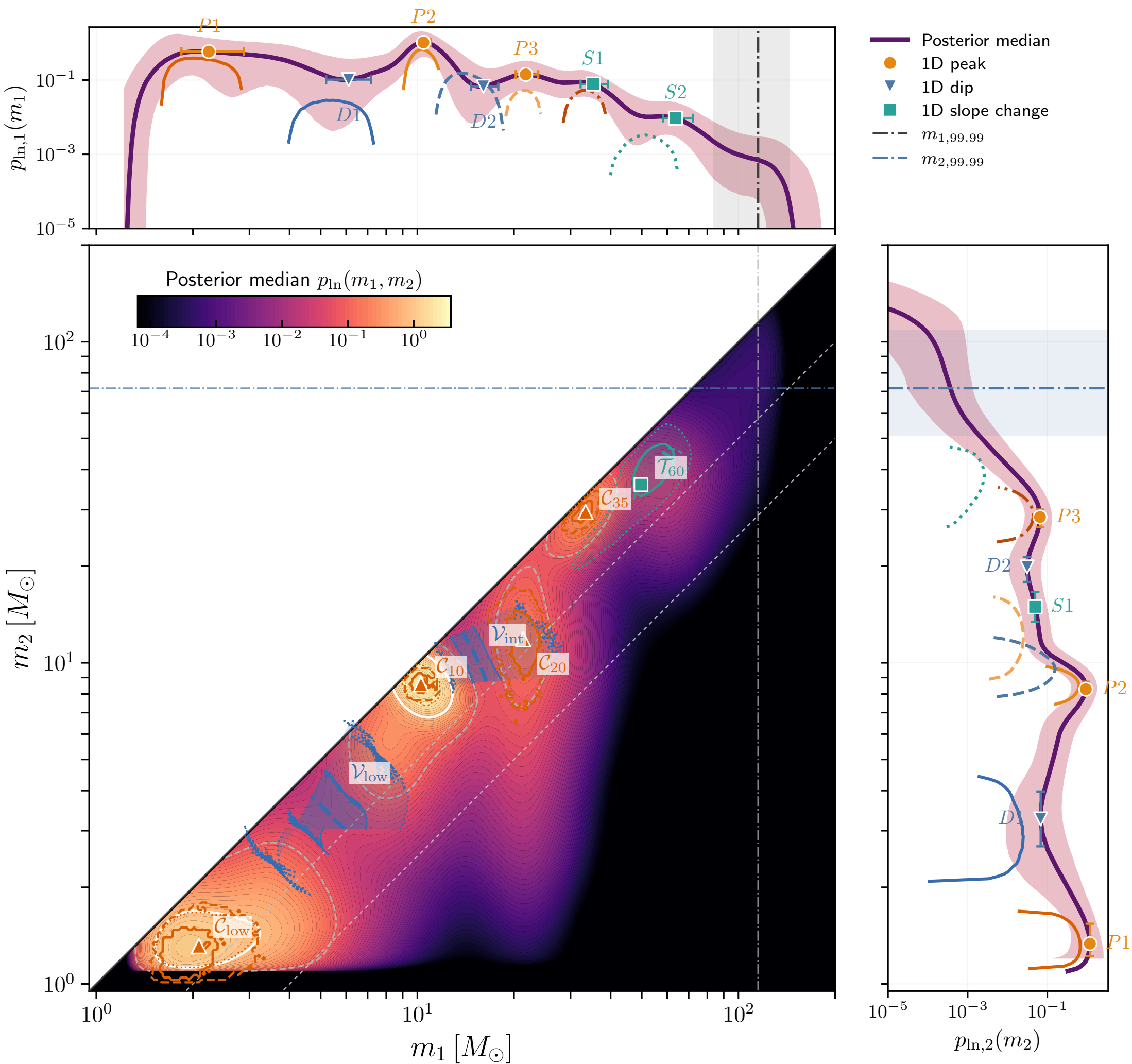
$\mathcal{V}_{\text{int}}$

Extended oblique deficit  
(clearly distinct from  $\mathcal{C}_{20}$  in 2D)

$m_1$  recovery: dip  $D_2$  which forms a dip-excess pattern with the higher-mass feature  $P_3$

Marginalization can merge distinct 2D structures or emphasize features associated mainly with one component mass.

# Projections on 1D marginals



| Feature                    | $P_{\text{morph}}$ | Strength                              | $m_1$ manifestation                                                                                                                                                 | $m_2$ manifestation                                 |
|----------------------------|--------------------|---------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------|
| $\mathcal{C}_{\text{low}}$ | 99.1%              | $A_P = 0.985^{+0.014}_{-0.100}$       | peak at $2.24^{+1.14}_{-0.53}$ (99.3%)                                                                                                                              | peak at $1.33^{+0.38}_{-0.18}$ (> 99.9%)            |
| $\mathcal{V}_{\text{low}}$ | 52.2%              | $A_V = 0.886^{+0.078}_{-0.182}$       | dip at $6.11^{+1.74}_{-1.44}$ (99.2%)                                                                                                                               | dip at $3.26^{+1.32}_{-0.90}$ (> 99.9%)             |
| $\mathcal{C}_{10}$         | 99.7%              | $A_P = 0.977^{+0.021}_{-0.059}$       | peak at $10.45^{+0.98}_{-0.72}$ (100%)                                                                                                                              | peak at $8.28^{+0.61}_{-0.57}$ (100%)               |
| $\mathcal{V}_{\text{int}}$ | 73.5%              | $A_V = 0.933^{+0.036}_{-0.094}$       | dip at $16.05^{+5.32}_{-2.14}$ (92.2%); no unique marginal counterpart with $P_3$ quoted better part low                                                            |                                                     |
| $\mathcal{C}_{20}$         | 95.4%              | $A_P = 0.551^{+0.239}_{-0.275}$       | $P_3$ at $21.75^{+3.35}_{-2.56}$ (93.6%); feature at $14.94^{+2.93}_{-1.99}$ 89.9% peak, 10.1% slope change; $A_P^{1D} = 0.649^{+0.248}_{-0.418}$ when a valid peak | 86.0% slope change, 14.0% peak                      |
| $\mathcal{C}_{35}$         | 99.9%              | $A_P = 0.857^{+0.092}_{-0.134}$       | feature at $35.26^{+6.91}_{-4.70}$ (90.2%); 50.5% peak, 49.5% slope change                                                                                          | peak at $28.45^{+2.66}_{-2.90}$ (99.9%); 99.1% peak |
| $\mathcal{T}_{60}$         | 36.7%              | $\Delta\alpha = 4.92^{+3.13}_{-2.12}$ | high-mass transition at $63.64^{+15.24}_{-9.47}$ (86.4%)                                                                                                            | no secure direct counterpart                        |

# Recovery probability and strength

Recovery probability:  $P_{\text{morph},f} = \frac{1}{N_s} \sum_{s=1}^{N_s} Z_f^{(s)}$

Strength:

- *Concentrations:*  $A_P = 1 - \frac{p_{\text{ln,saddle}}}{p_{\text{ln,max}}}$
- *Valleys:*  $A_V = 1 - \frac{p_V}{\sqrt{p_L p_R}}$
- *Slope transitions:*  $\Delta\alpha = \alpha_{\text{pre}} - \alpha_{\text{post}}$

## Coordinates of a population feature

For each feature  $f$ , let

$$(x_{1,f}, x_{2,f})$$

denote its centroid position, or the representative front point for the high-mass transition.

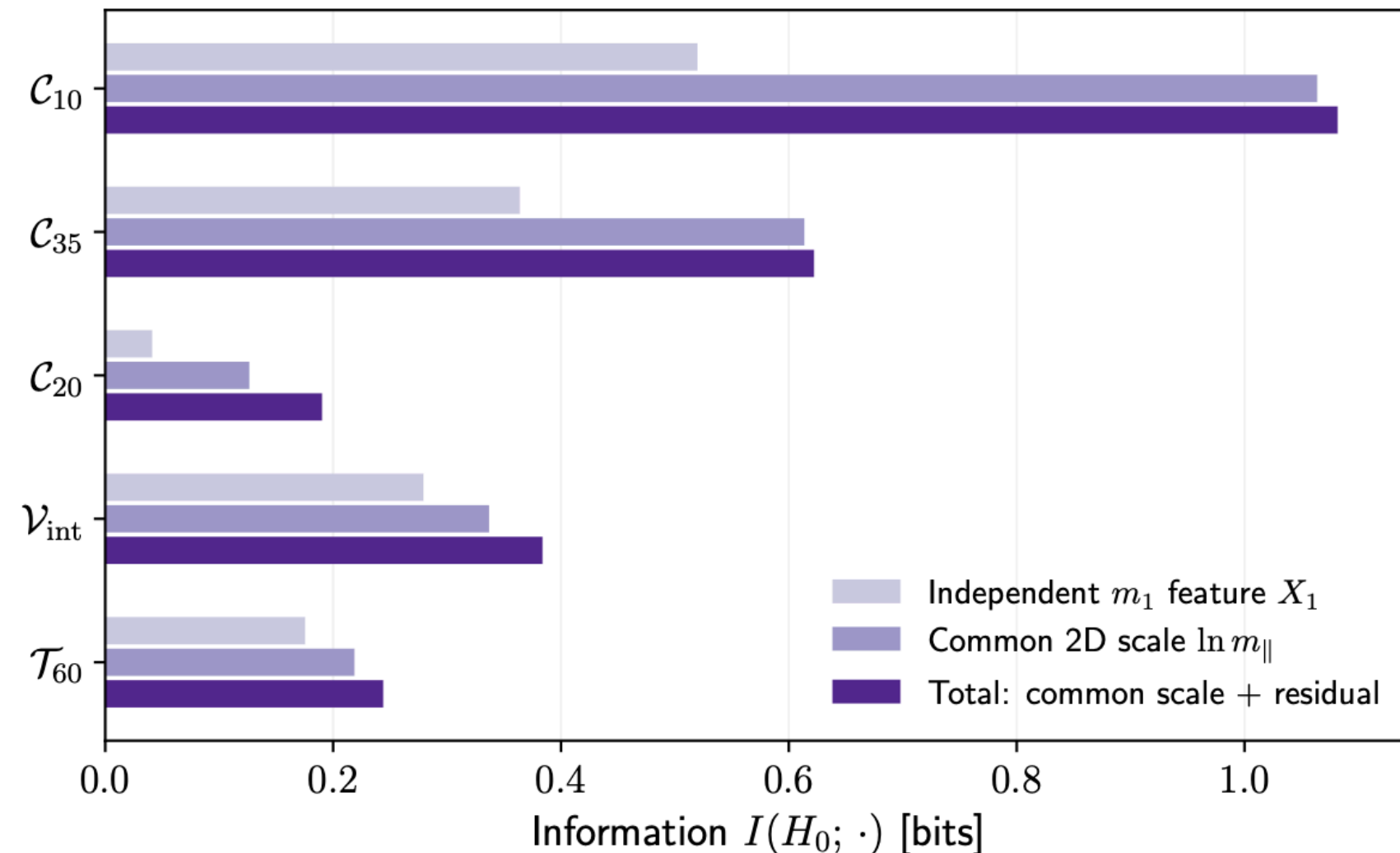
We decompose this position into

$$m_{\parallel,f} = \sqrt{x_{1,f} x_{2,f}} \quad q_f = \frac{x_{2,f}}{x_{1,f}}.$$

Equivalently,

$$\ln m_{\parallel,f} = \frac{\ln x_{1,f} + \ln x_{2,f}}{2},$$

$$\ln q_f = \ln x_{2,f} - \ln x_{1,f}.$$



## Three feature representations

For each feature, we compare:

$$X_{1,f}^{(1D)} = \ln m_{1,f}^{(1D)}, \quad \text{independent } m_1 \text{ feature,}$$

$$X_{\parallel,f} = \ln m_{\parallel,f}, \quad \text{joint common scale,}$$

$$\mathbf{x}_f^{(2D)} = \left( \ln m_{\parallel,f}, \ln q_f \right), \quad \text{full joint position.}$$

All comparisons use the same posterior realizations for which all three measurements are available.

## Mutual information with $H_0$

For a feature representation  $X$ ,

$$I(H_0; X) = \int p(H_0, X) \log_2 \left[ \frac{p(H_0, X)}{p(H_0)p(X)} \right] dH_0 dX.$$

It measures the reduction in uncertainty about  $H_0$  obtained by knowing  $X$ , including nonlinear dependence.

The information specific to the mass-ratio direction is

$$I\left(H_0; \ln q_f \mid \ln m_{\parallel,f}\right).$$

This asks whether  $q_f$  retains information about  $H_0$  after accounting for the common mass scale.

# Inference: full-dimensional hierarchical sampling



Full inference:

$$p(\boldsymbol{\lambda}, \boldsymbol{\theta} \mid \{\mathcal{D}\}) \propto \frac{\pi(\boldsymbol{\lambda})}{\xi(\boldsymbol{\lambda})^{N_{\text{obs}}}} \prod_{i=1}^{N_{\text{obs}}} p(\theta_i \mid \mathcal{D}_i) \frac{p_{\text{pop}}(\theta_i \mid \boldsymbol{\lambda})}{\pi_{\text{PE}}(\theta_i)}$$

- Single-event likelihoods are represented by continuous Gaussian-mixture surrogates.
- Sampling is performed with GPU-accelerated NUTS in NumPyro: four chains with  $\hat{r} \sim 1$

$$\sigma_{\ln \mathcal{L}}^2 < 1$$

Selection effects:

$$\xi(\lambda) = \int p_{\text{pop}}(\theta|\lambda) p_{\text{det}}(\theta) d\theta$$

- Selection effects are evaluated by reweighting detected MC injections
- We require the Monte Carlo variance of the estimated log likelihood to satisfy  $\sigma_{\ln \mathcal{L}}^2 < 1$