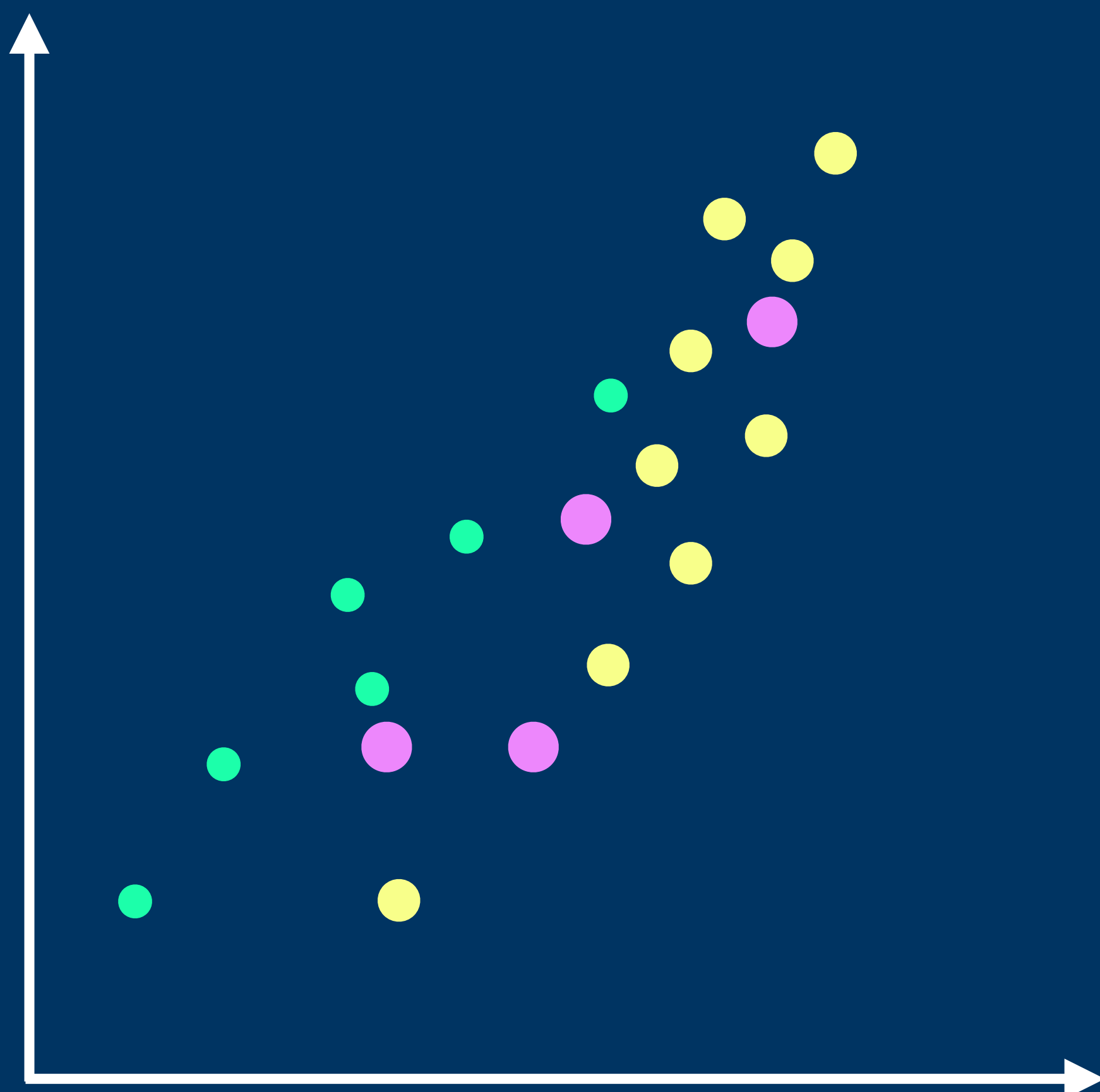
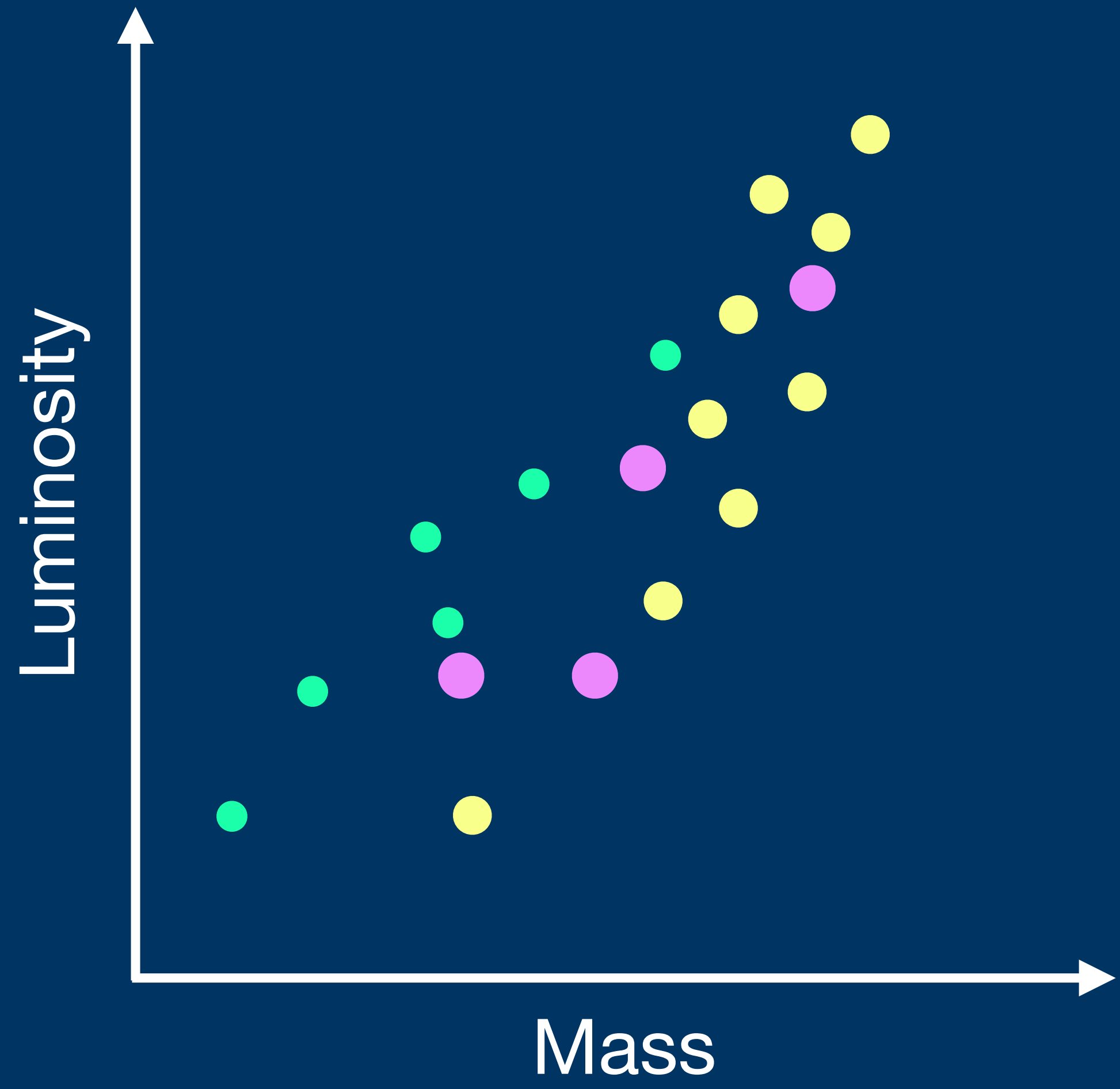


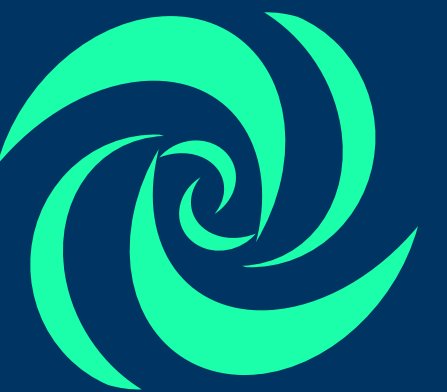
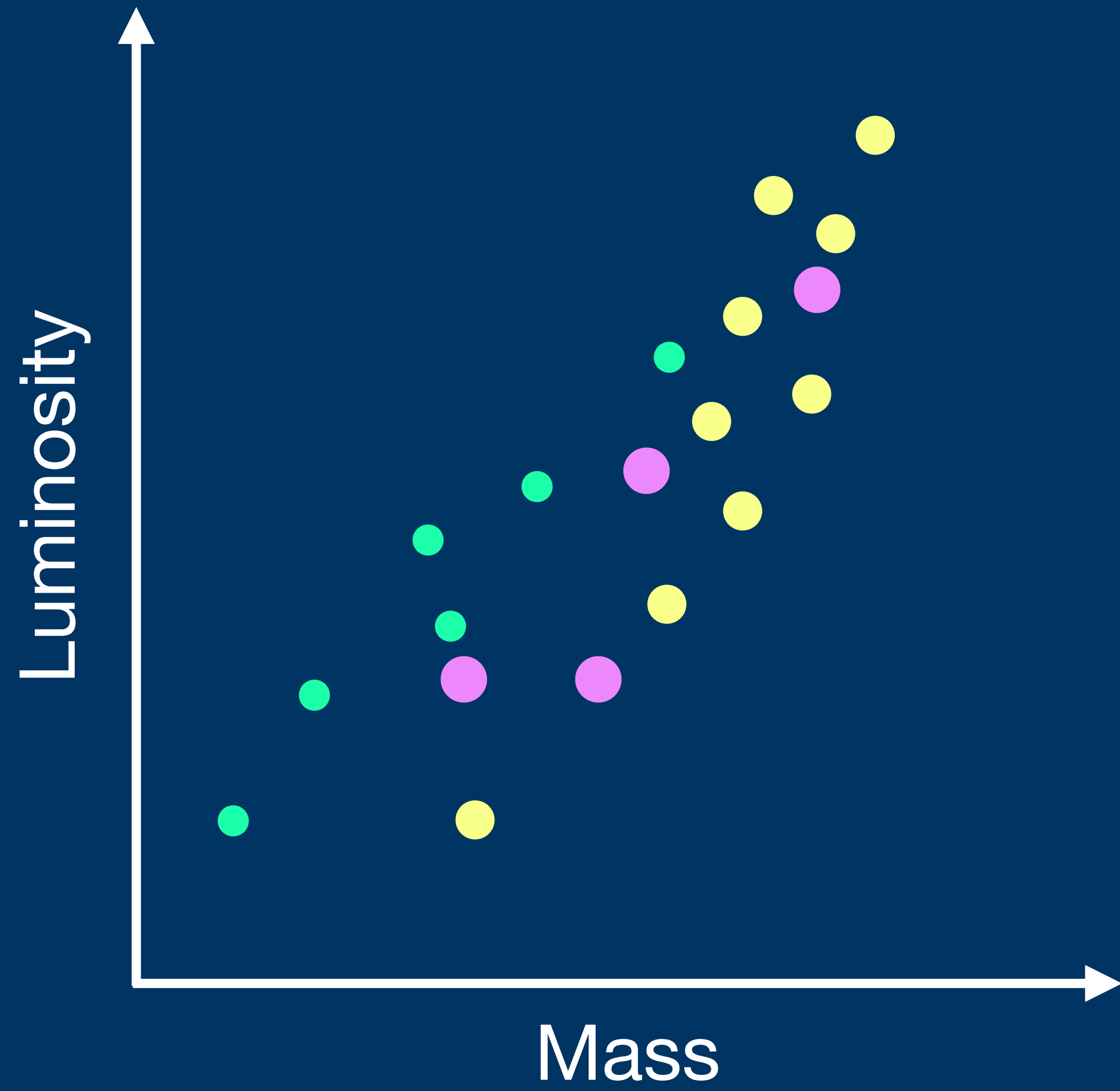
Hierarchical Bayesian Models: What are they?

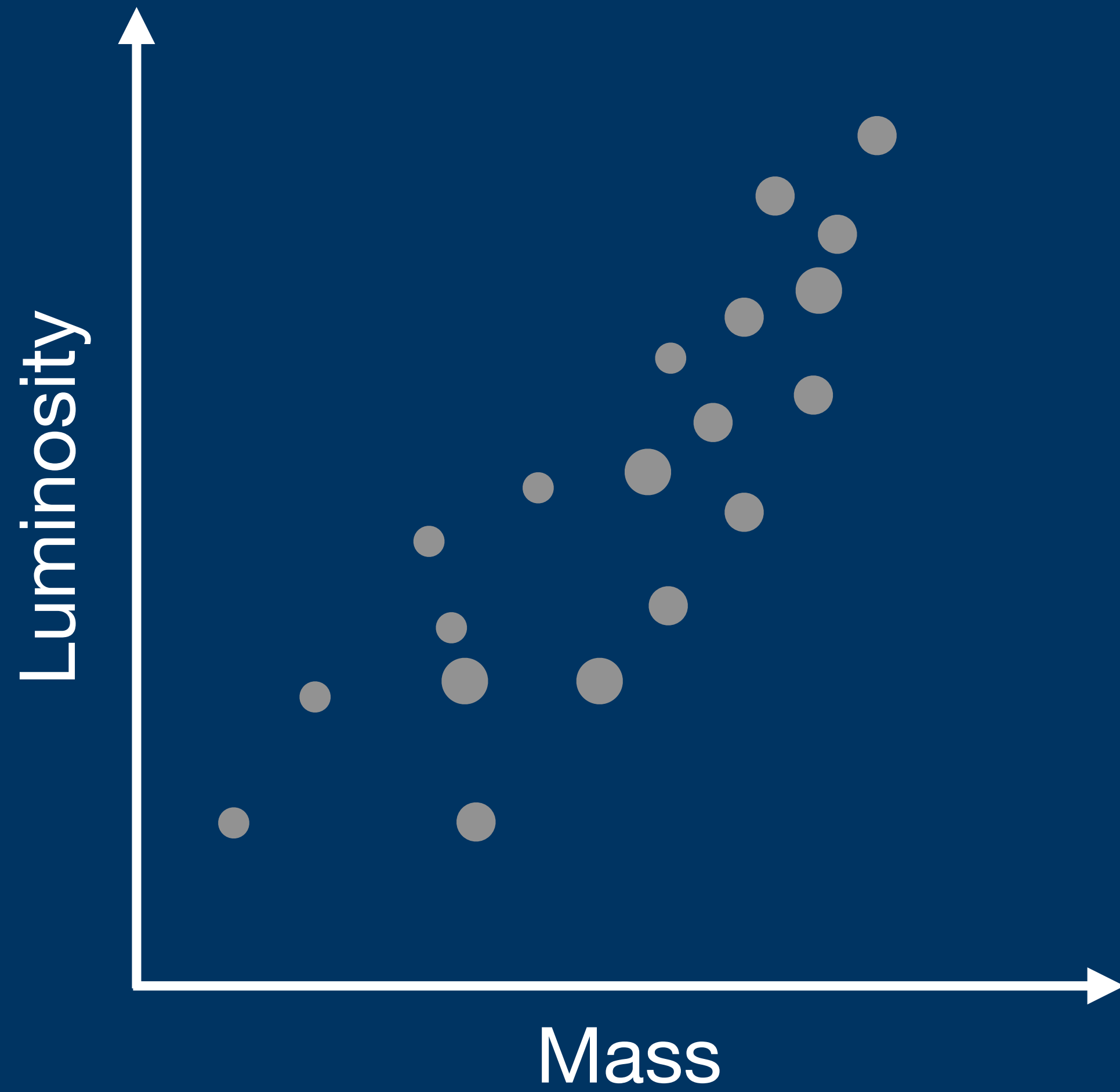
Examples from astrophysics

J. Michael Burgess
MPE

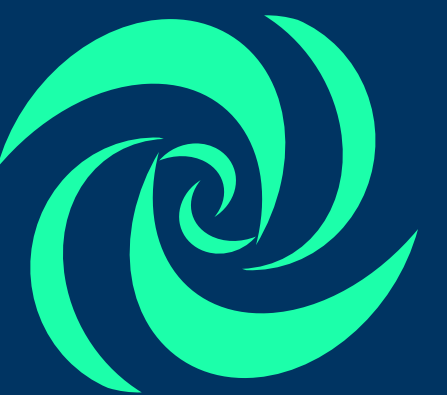




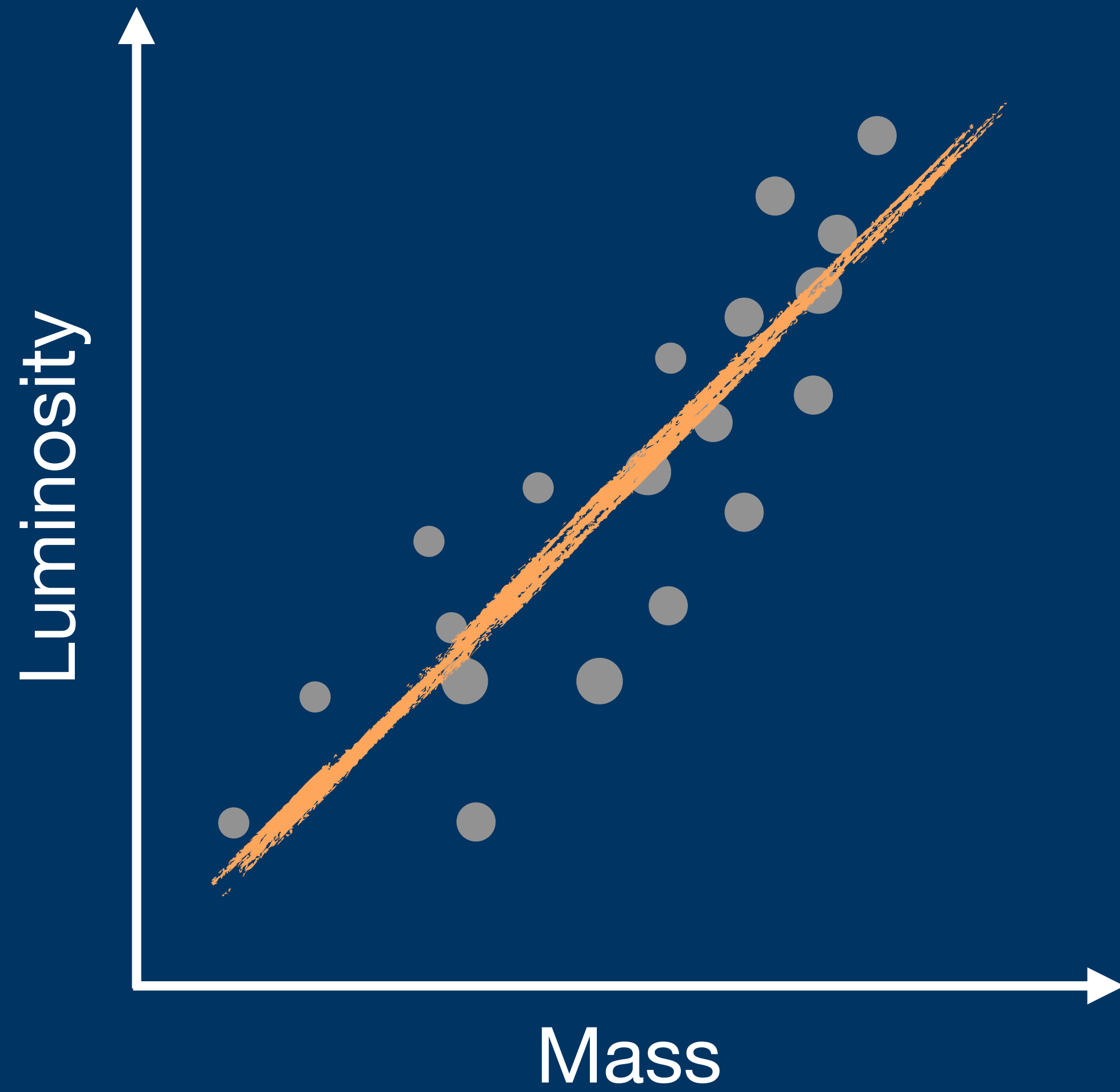




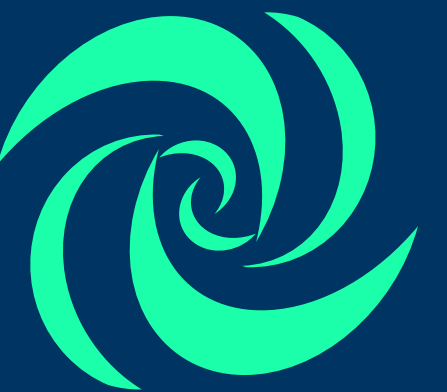
$$L = \alpha M + \beta$$



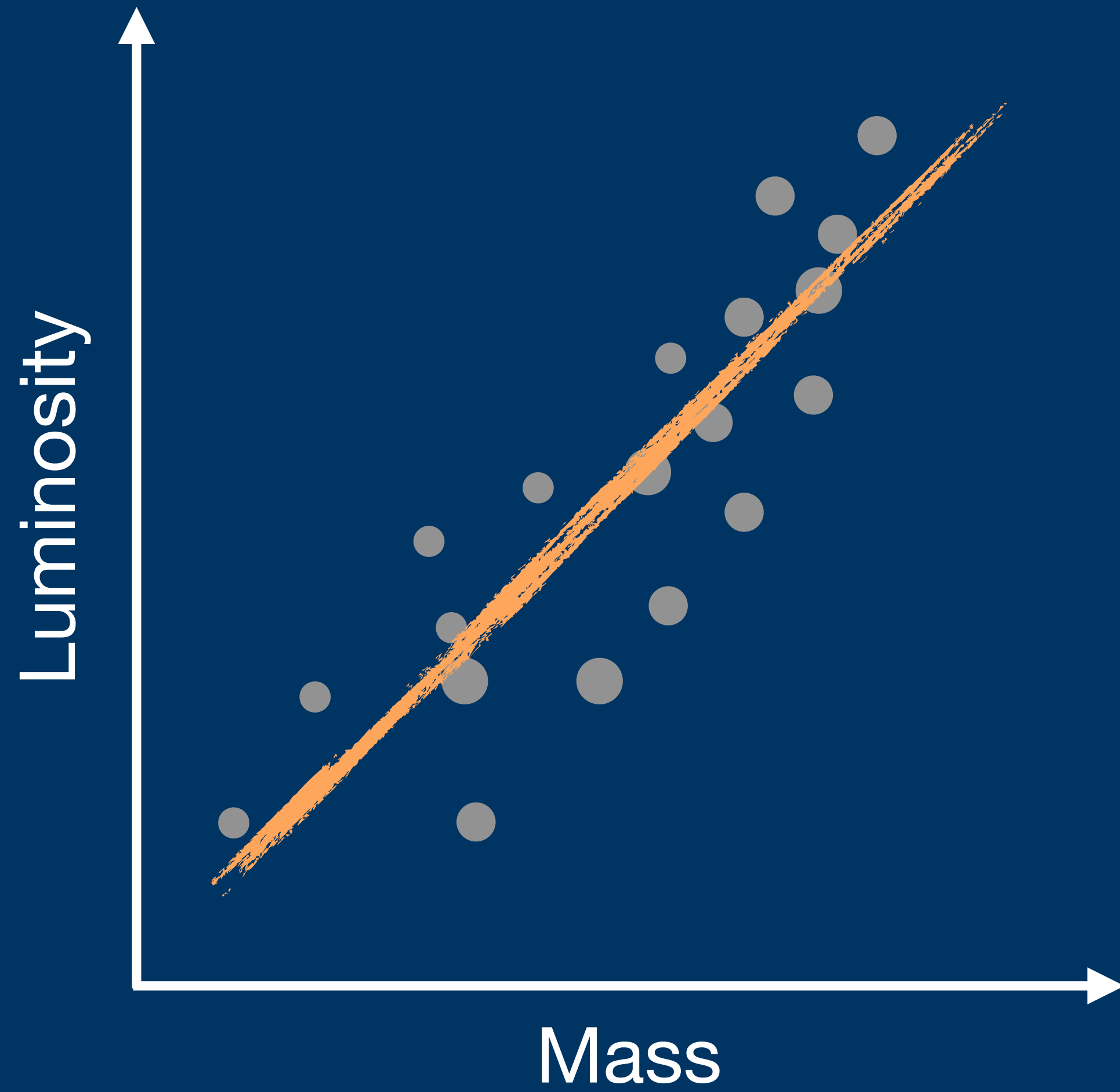
How do we get global information?



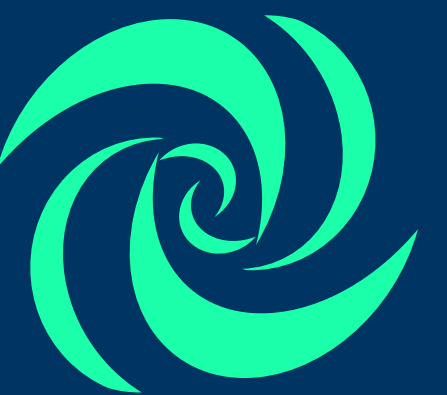
$$L = \alpha M + \beta$$



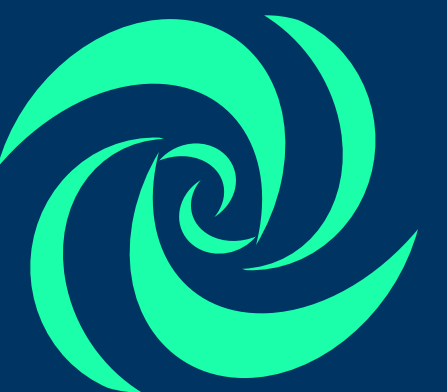
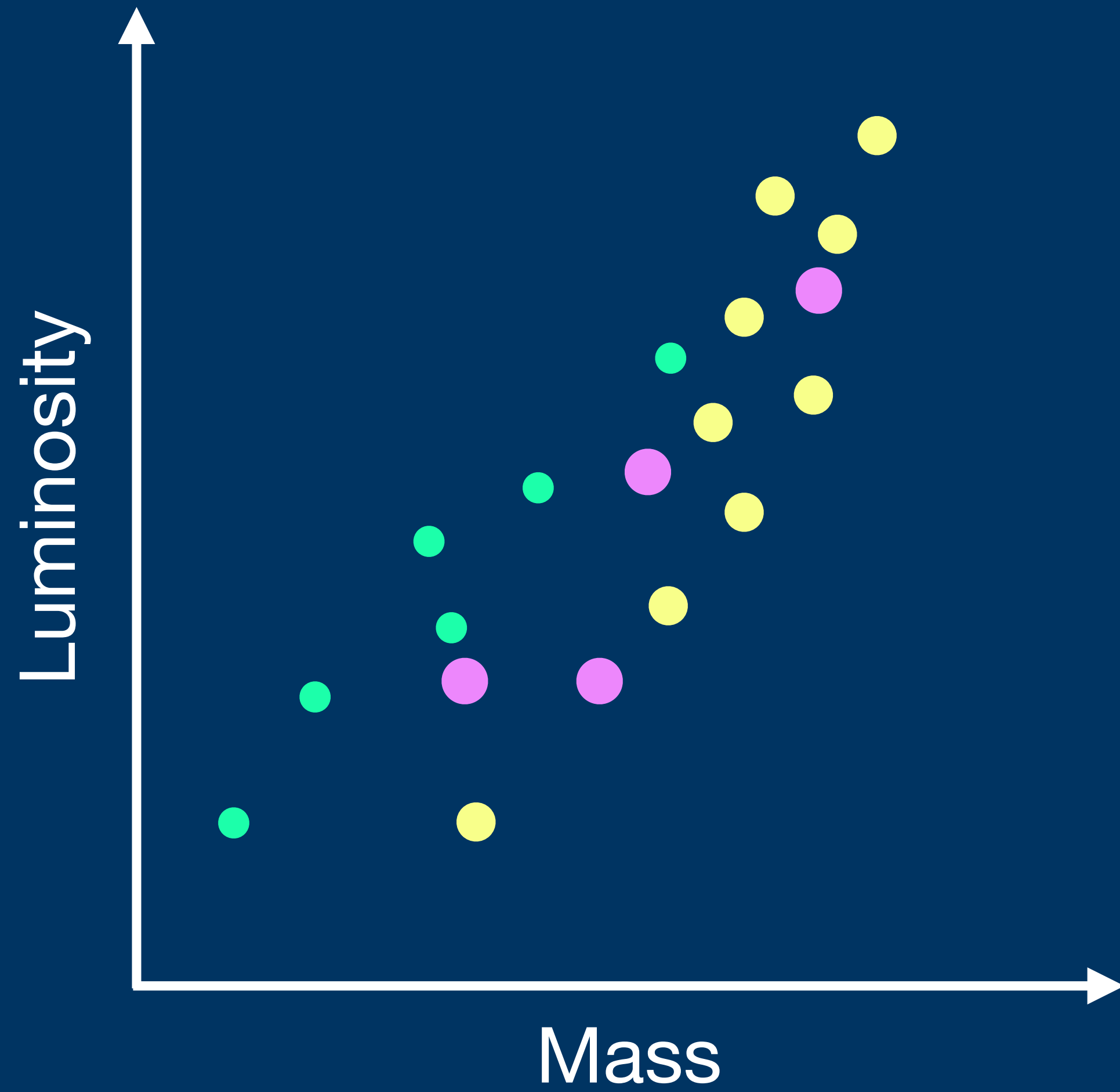
How do we get global information?



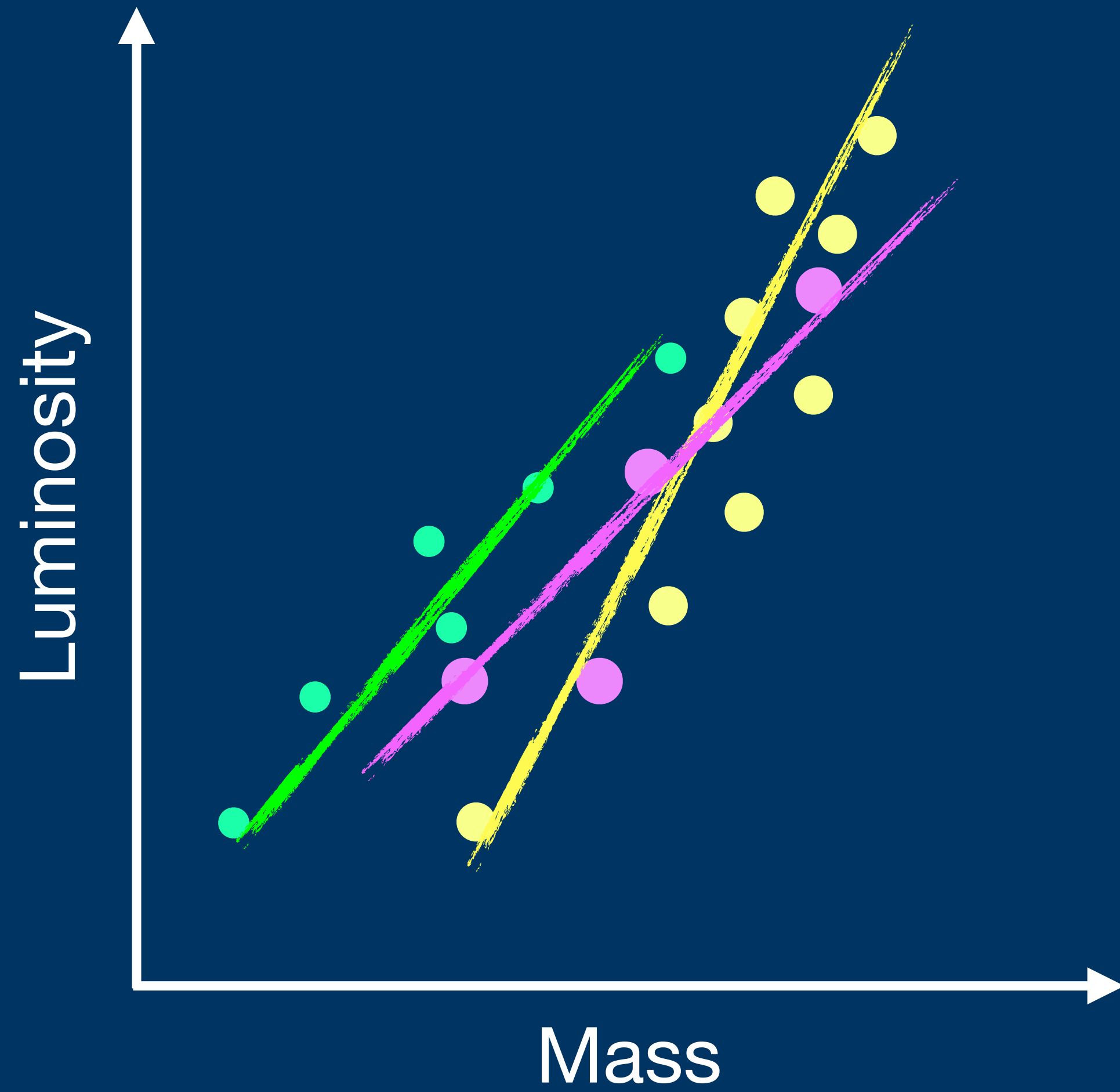
$$L = \alpha M + \beta$$



How do we get global information?



How do we incorporate all the information?



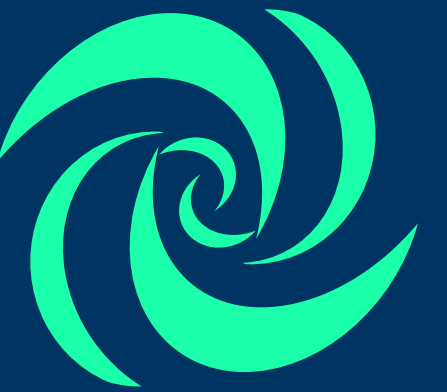
$$L = \alpha M + \beta$$



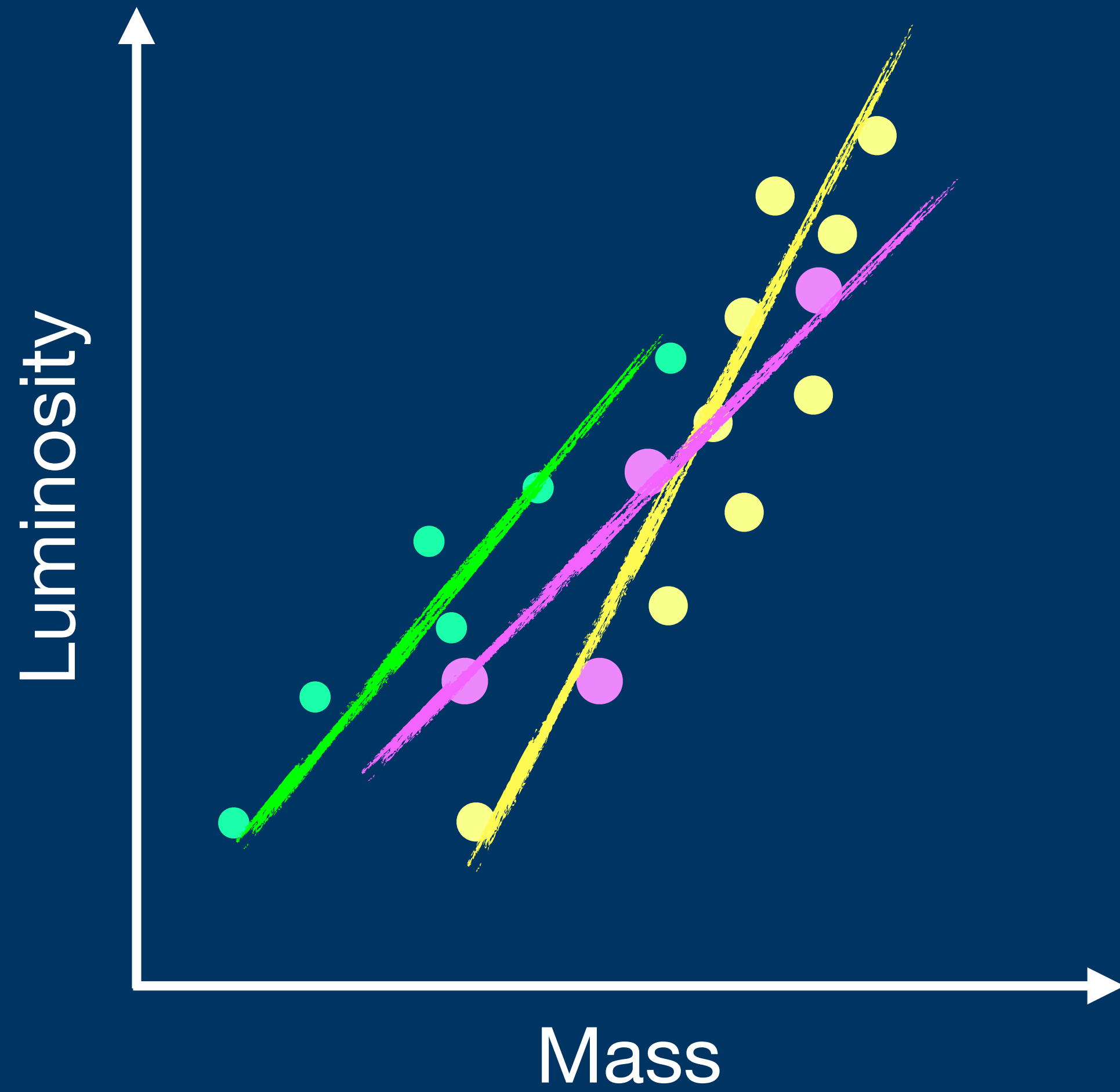
$$L = \alpha M + \beta$$



$$L = \alpha M + \beta$$



How do we incorporate all the information?



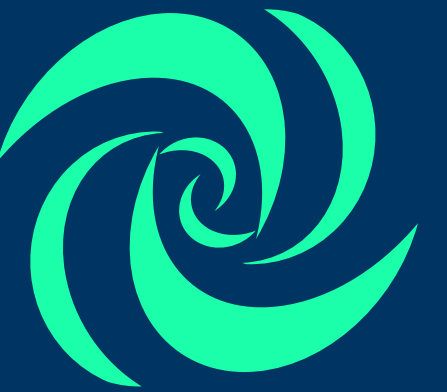
$$L = \alpha M + \beta$$



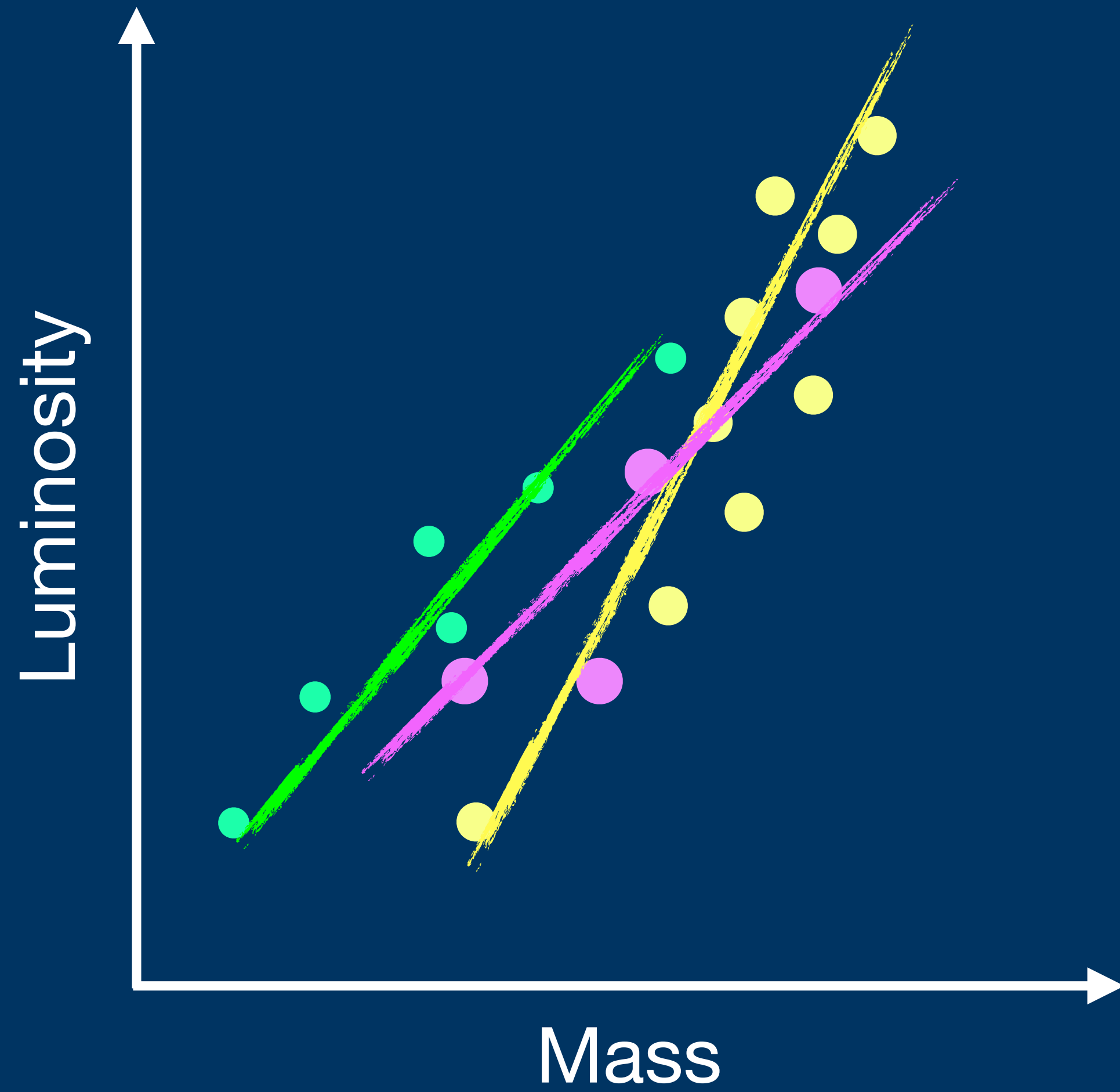
$$L = \alpha M + \beta$$



$$L = \alpha M + \beta$$



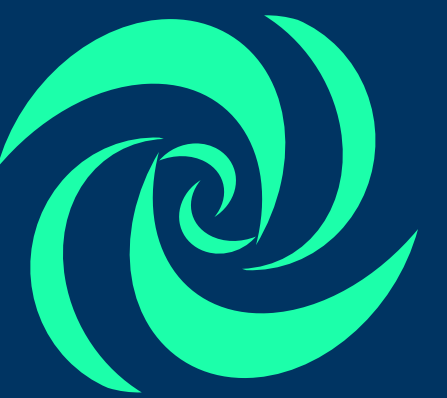
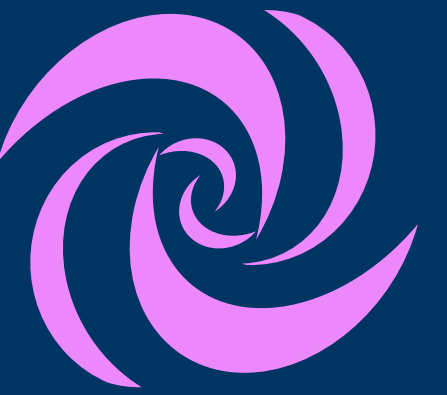
How do we incorporate all the information?



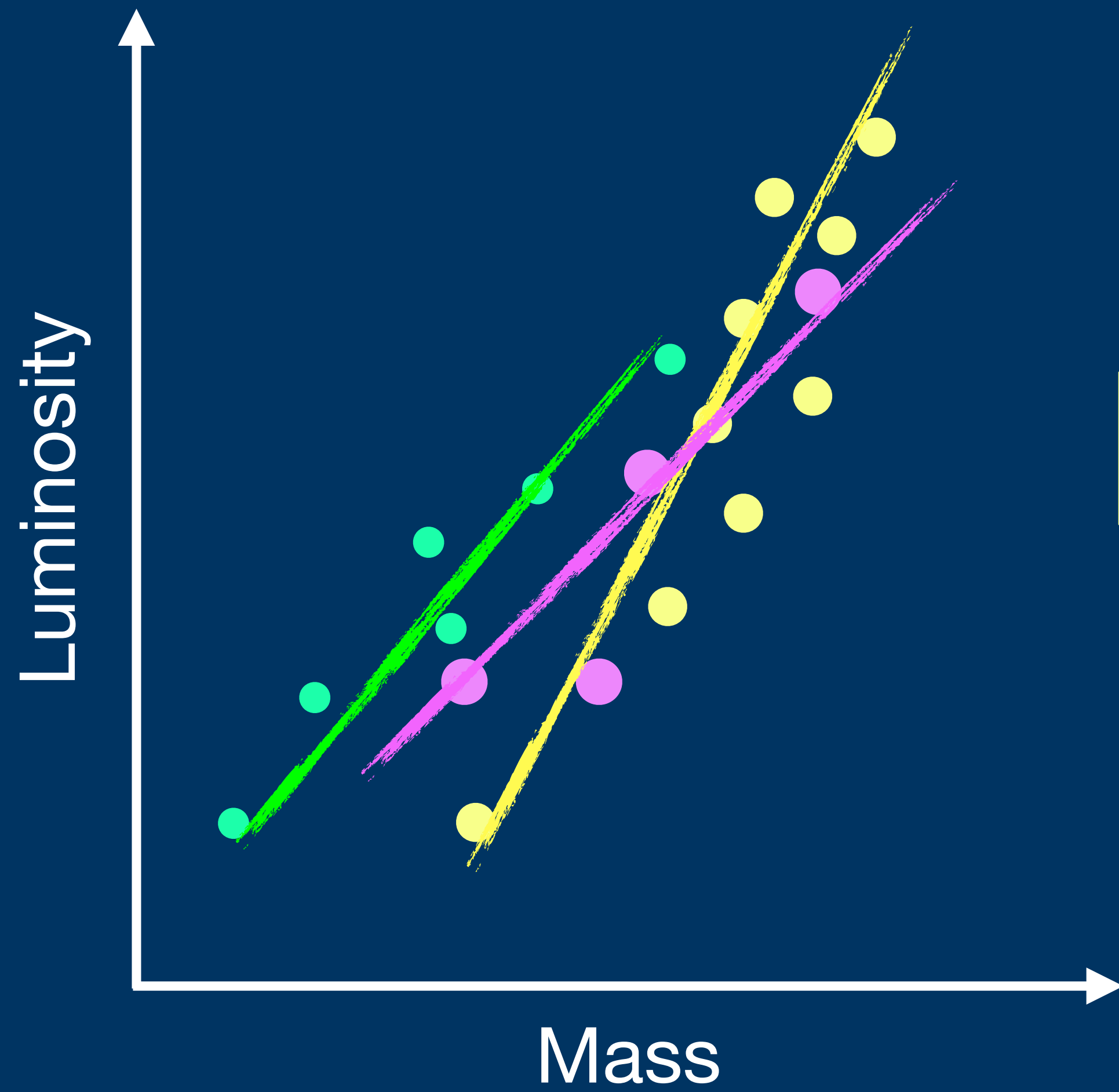
$$L = \alpha M + \beta$$

$$L = \alpha M + \beta$$

$$L = \alpha M + \beta$$



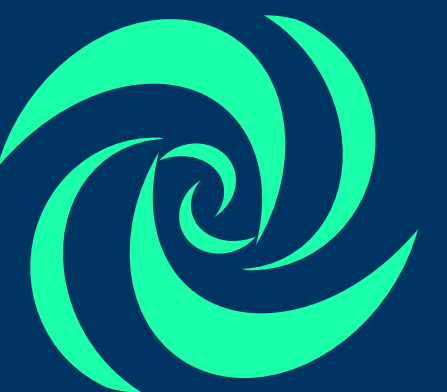
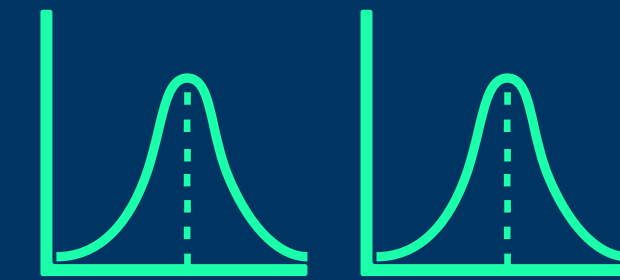
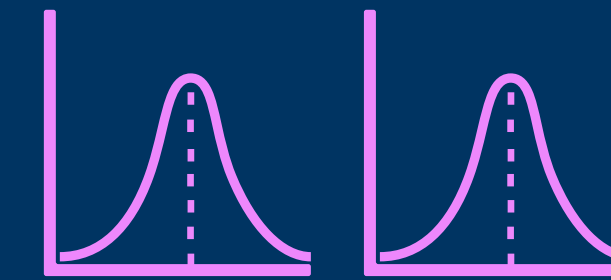
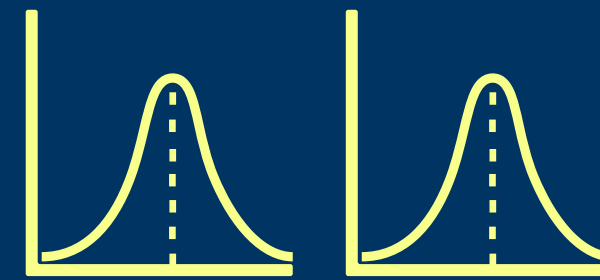
How do we incorporate all the information?



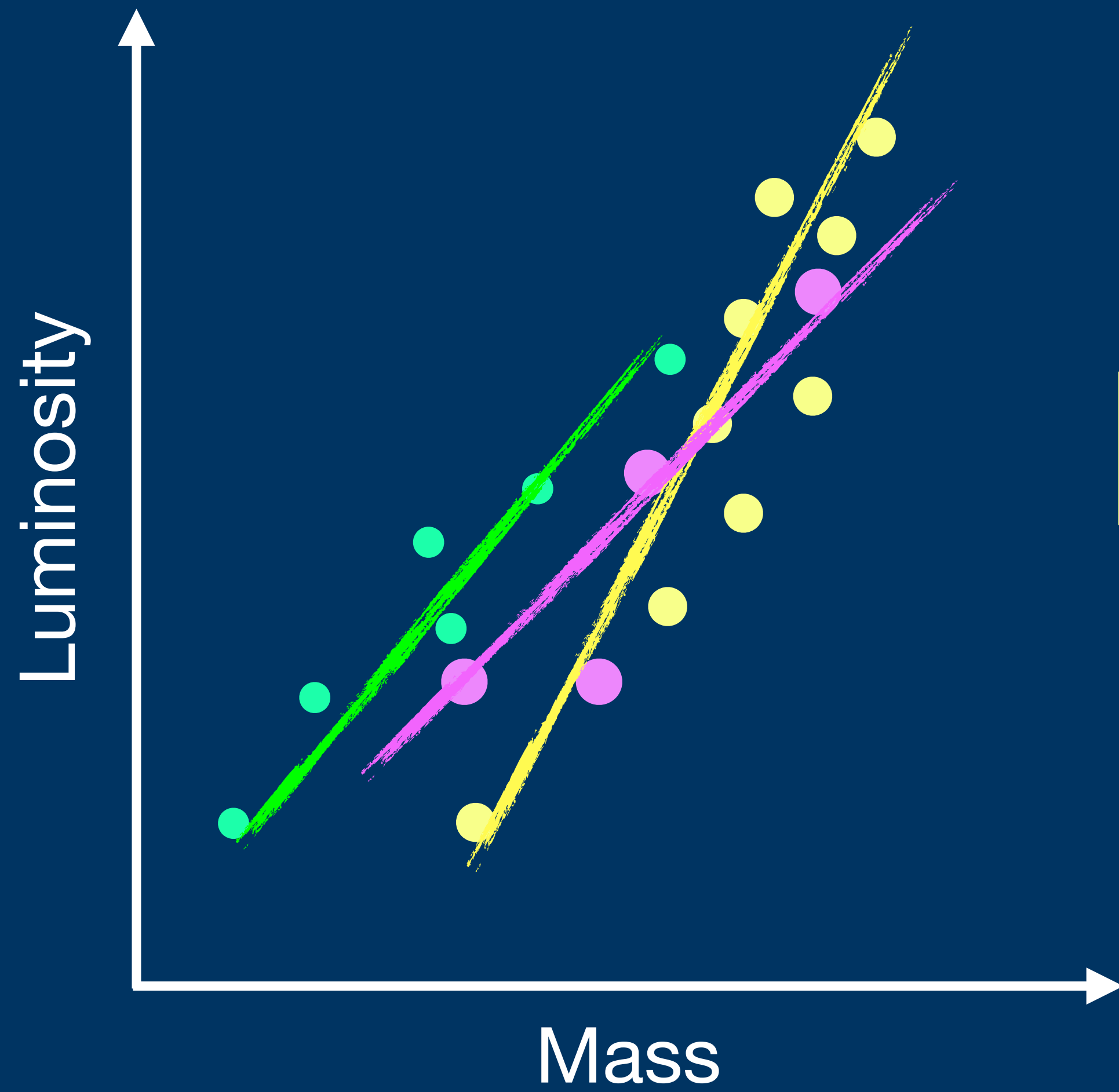
$$L = \alpha M + \beta$$

$$L = \alpha M + \beta$$

$$L = \alpha M + \beta$$



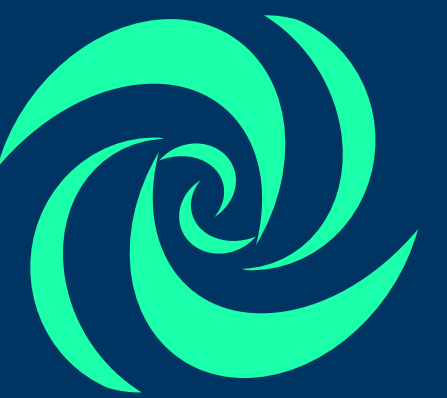
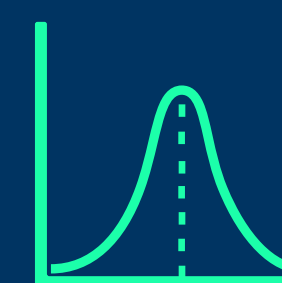
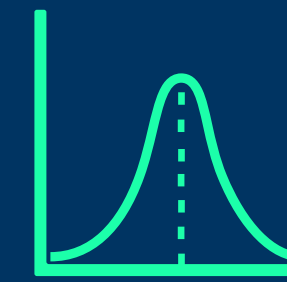
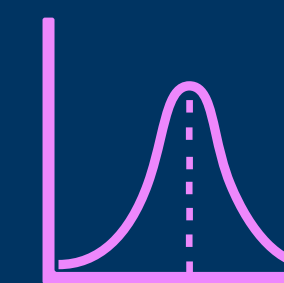
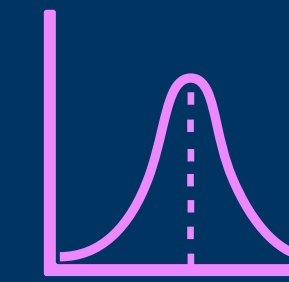
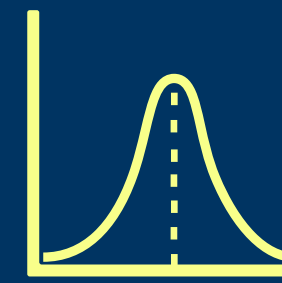
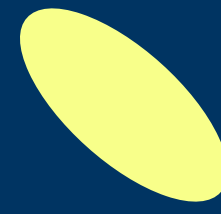
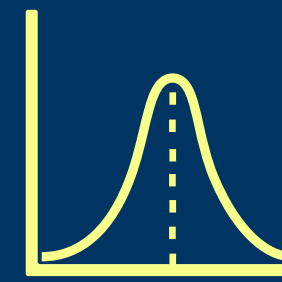
How do we incorporate all the information?



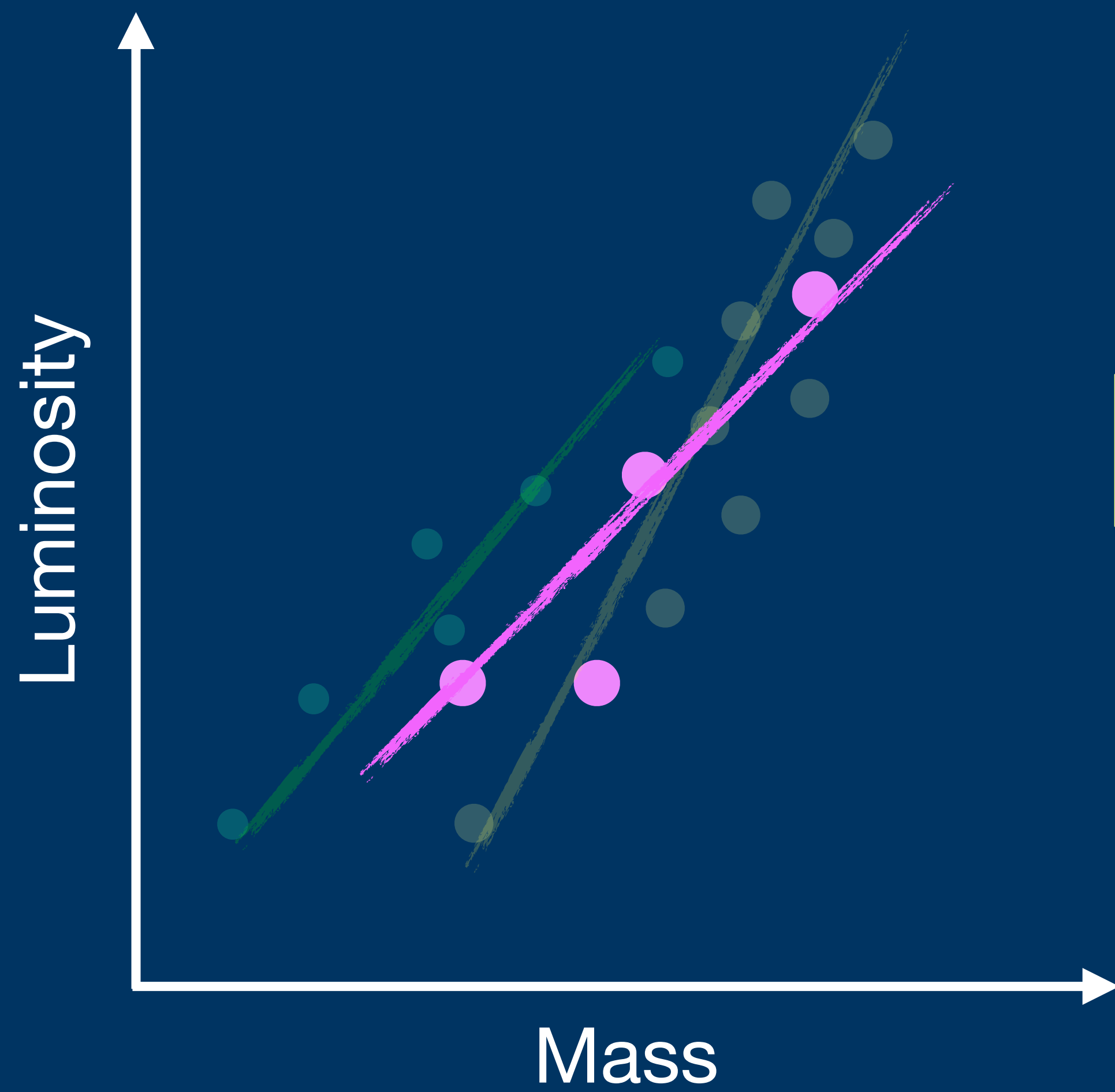
$$L = \alpha M + \beta$$

$$L = \alpha M + \beta$$

$$L = \alpha M + \beta$$



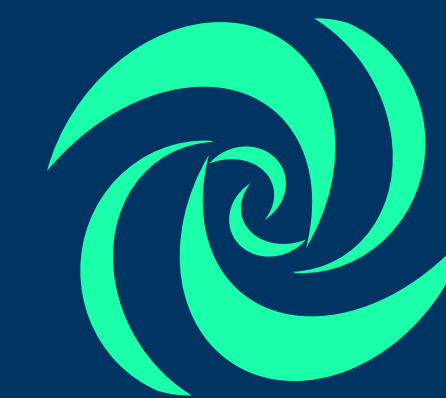
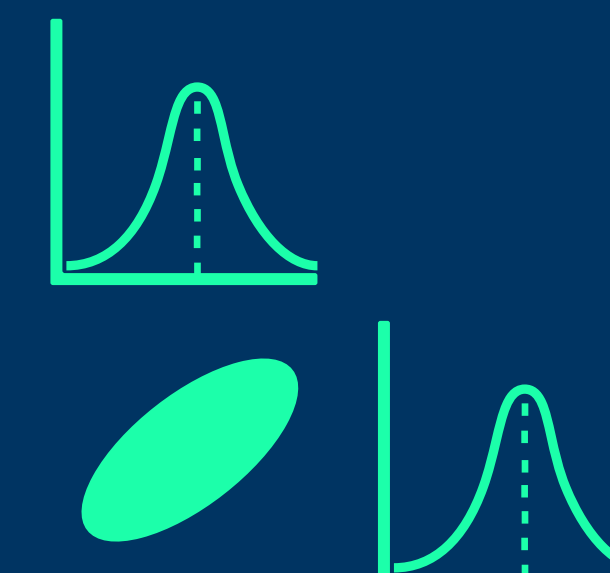
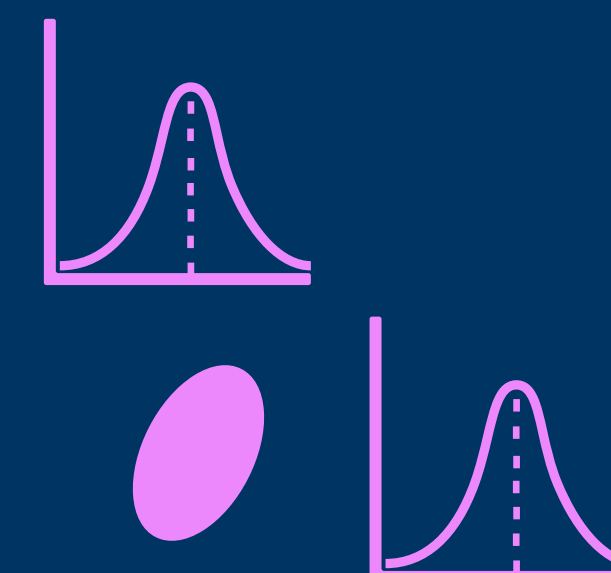
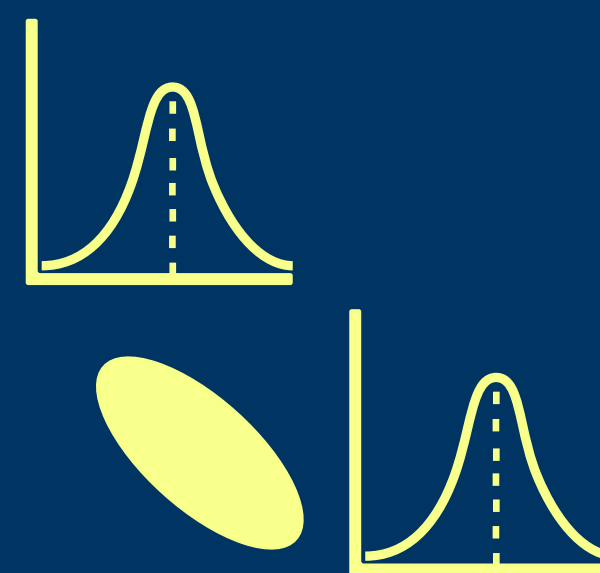
How do we incorporate all the information?



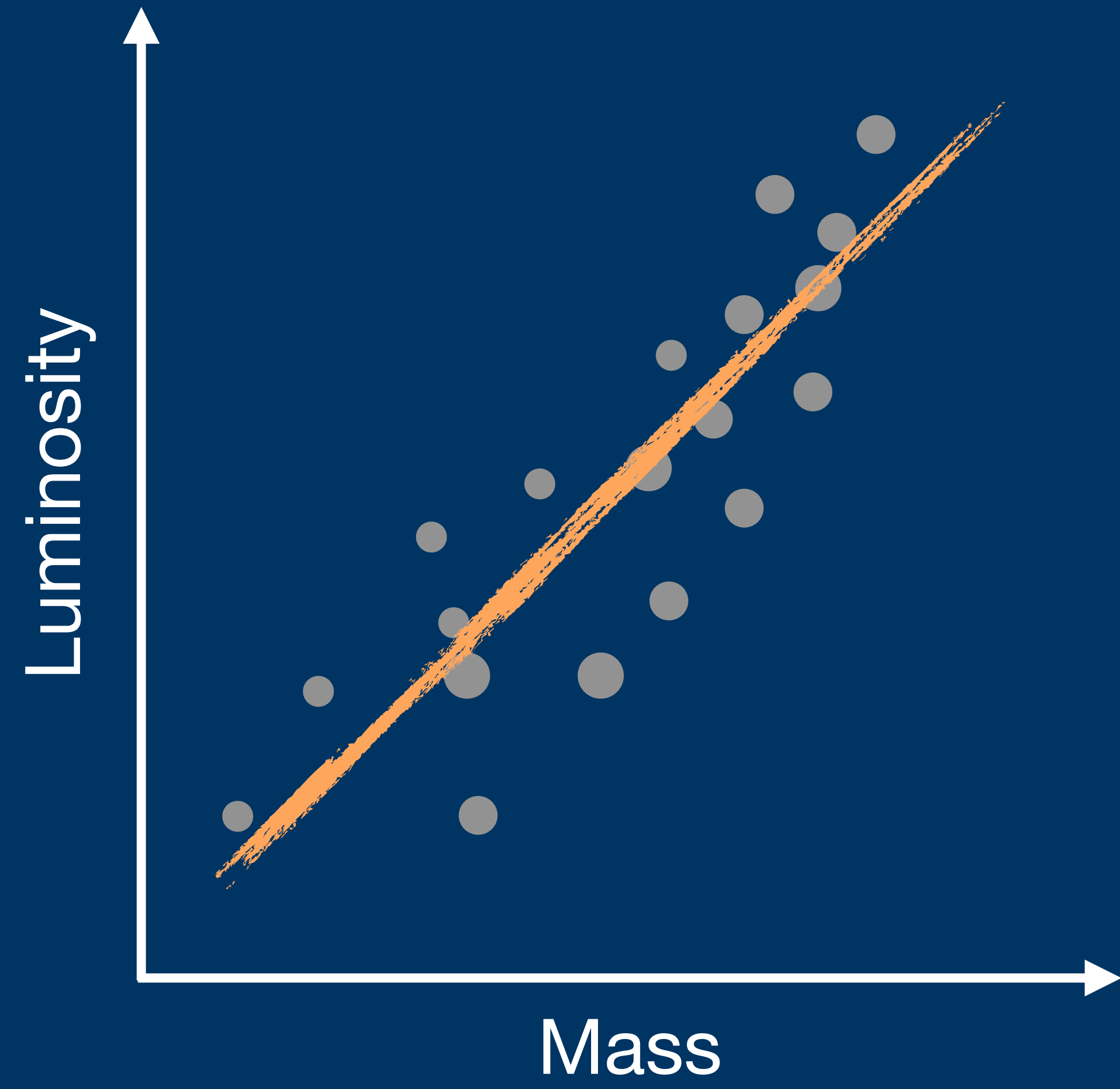
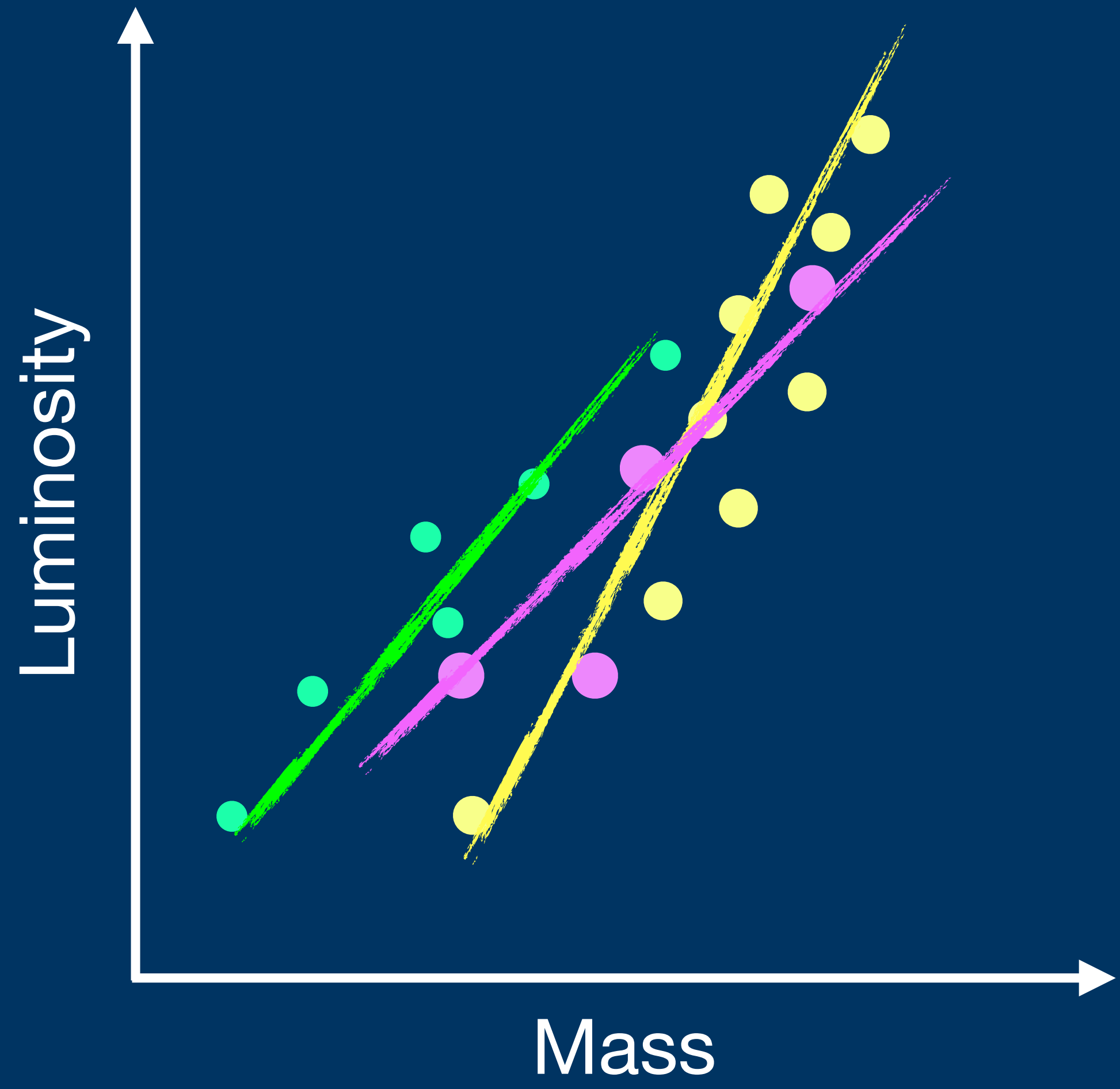
$$L = \alpha M + \beta$$

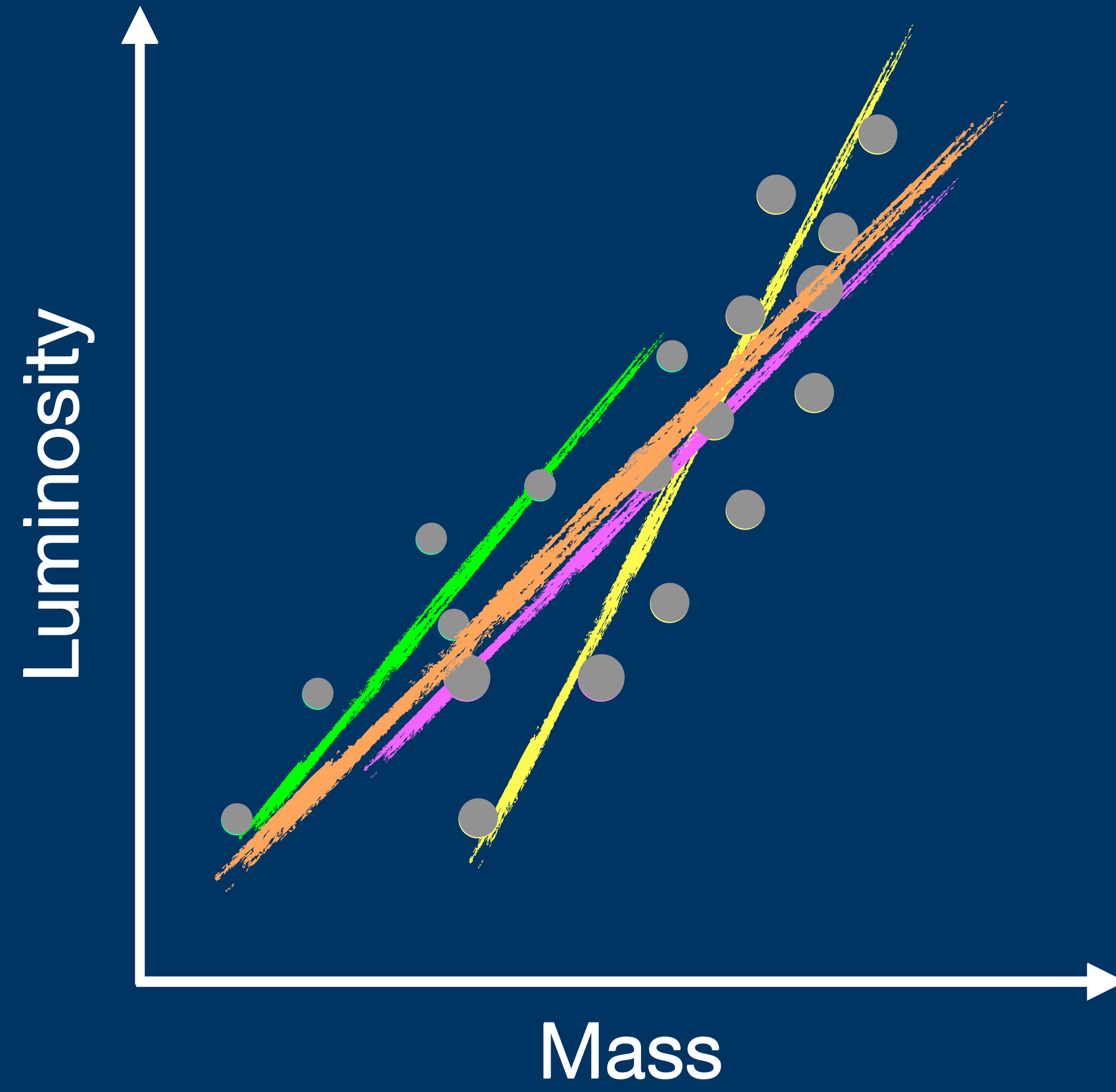
$$L = \alpha M + \beta$$

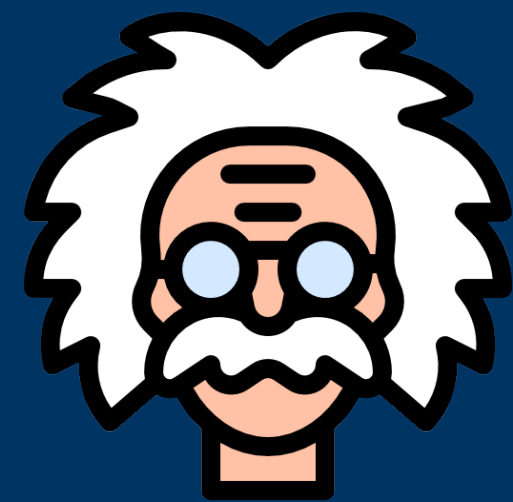
$$L = \alpha M + \beta$$

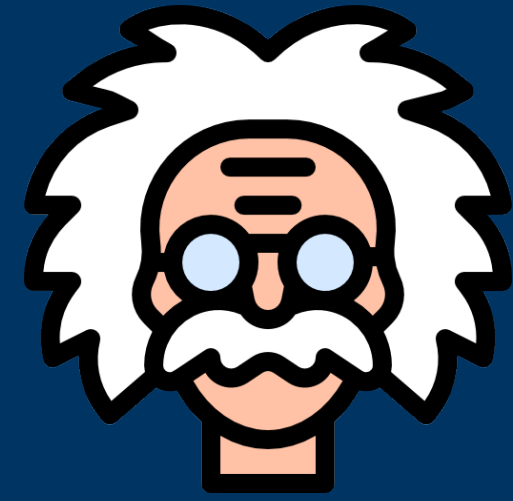
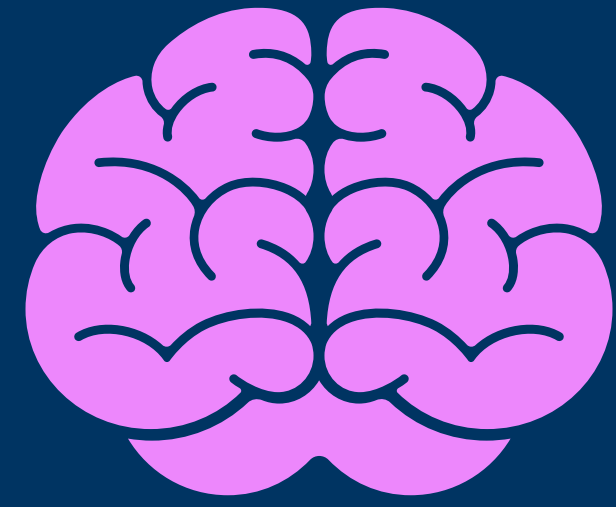


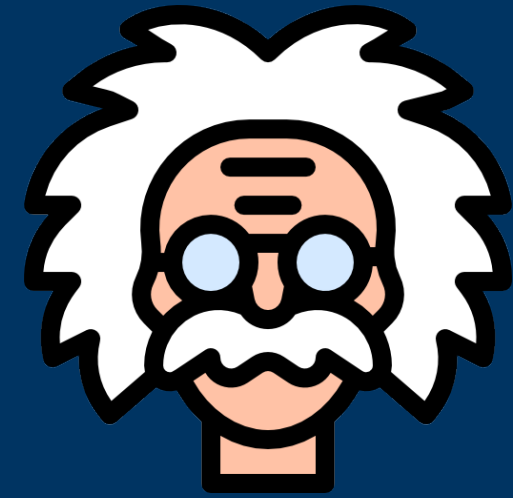
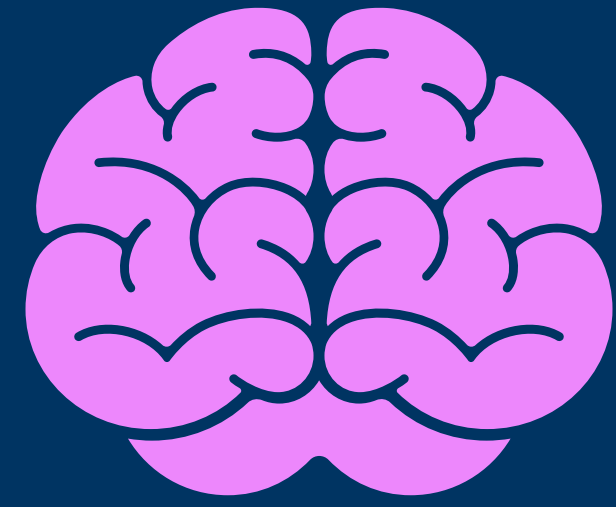
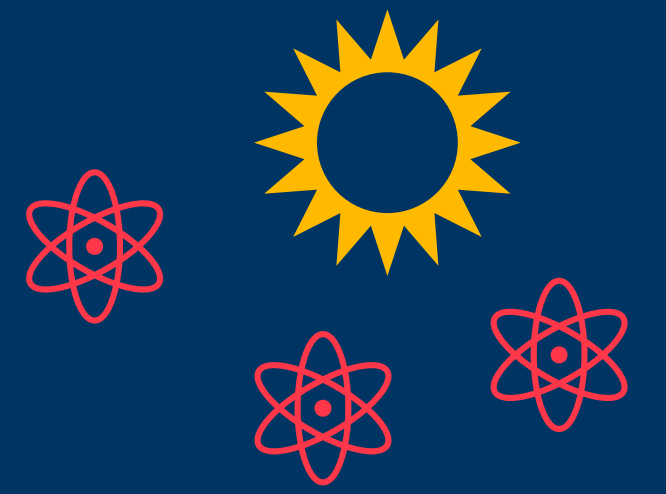
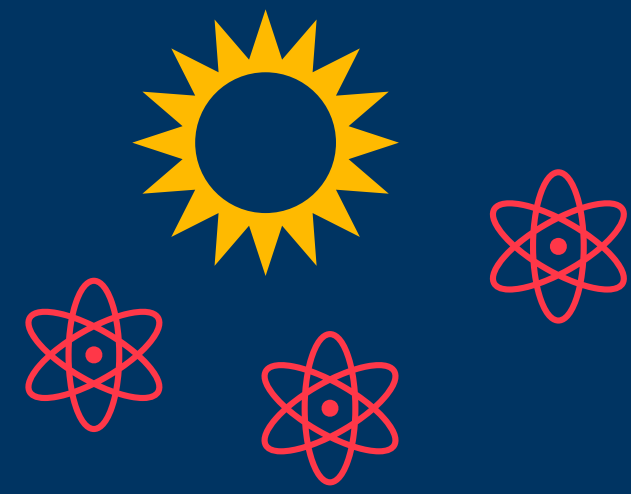
How do we incorporate all the information?

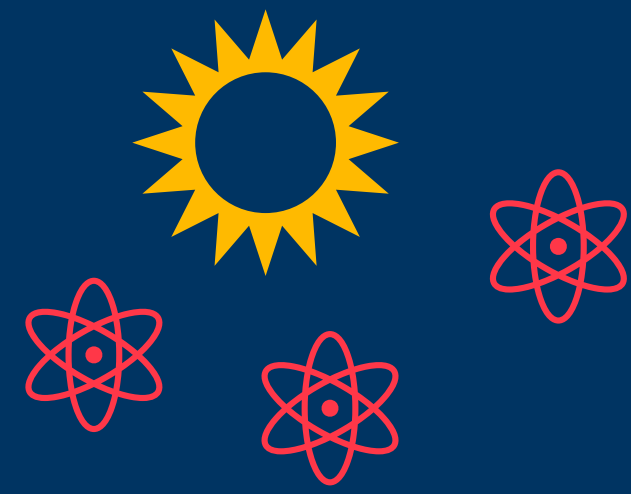






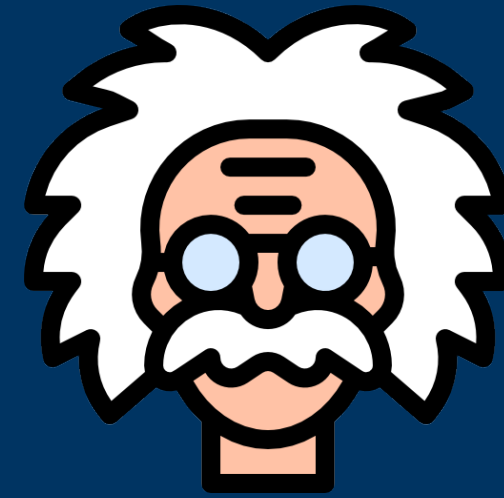
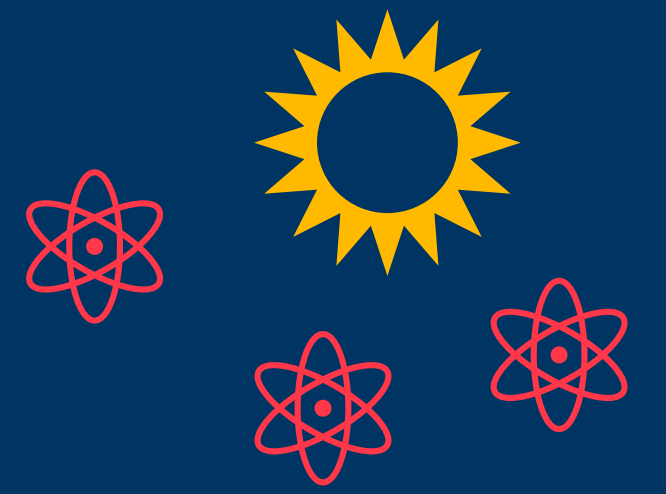
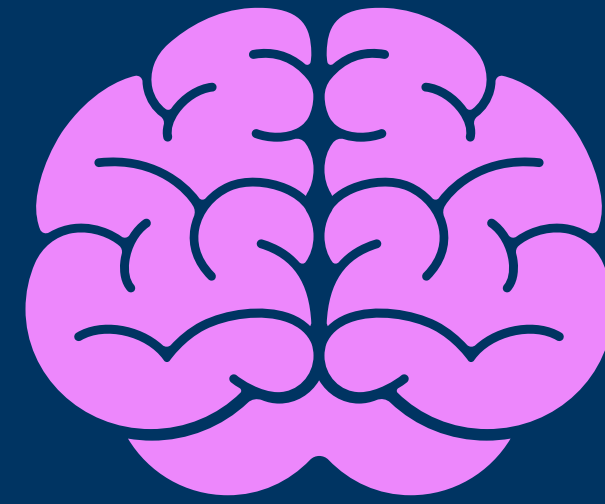


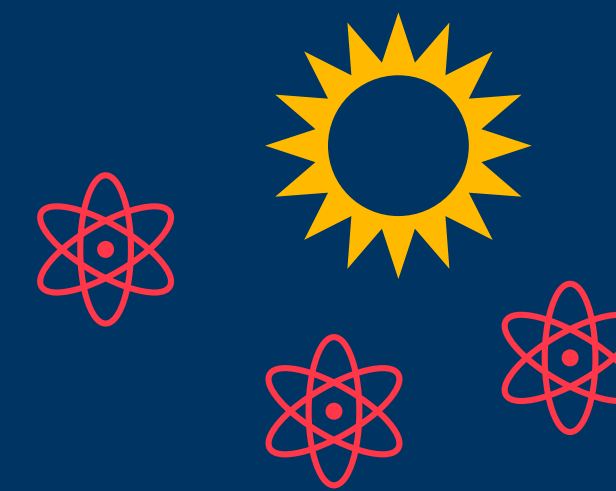
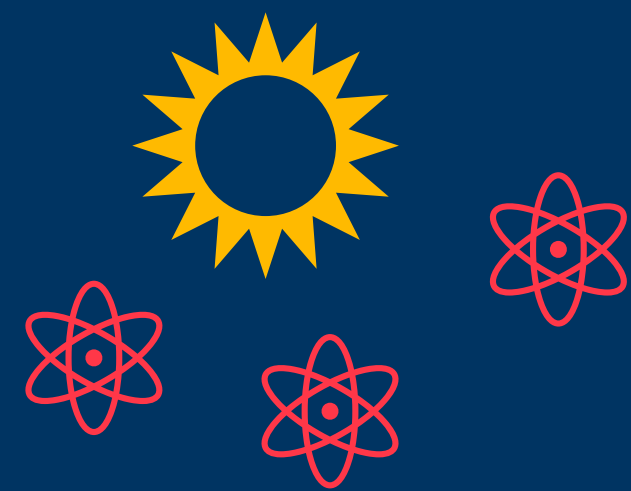


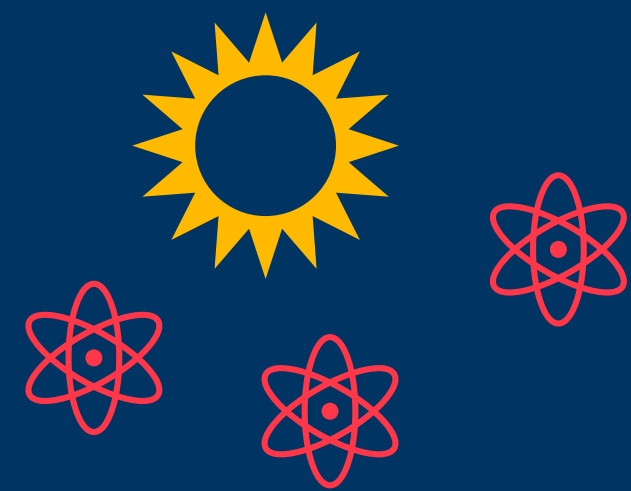
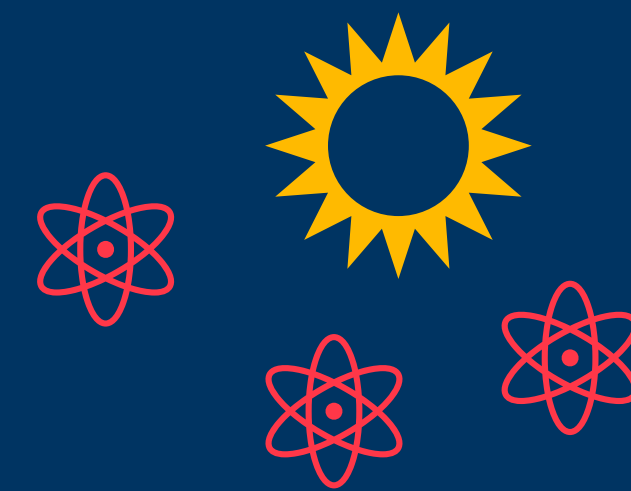
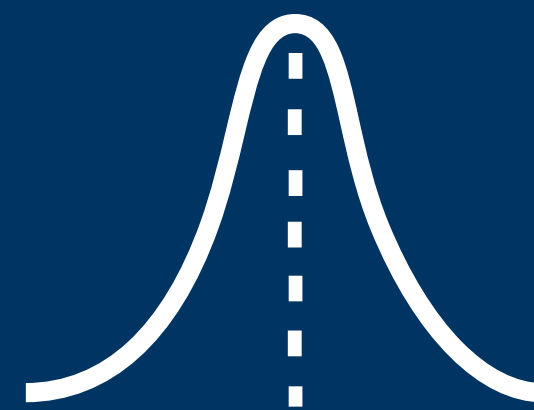
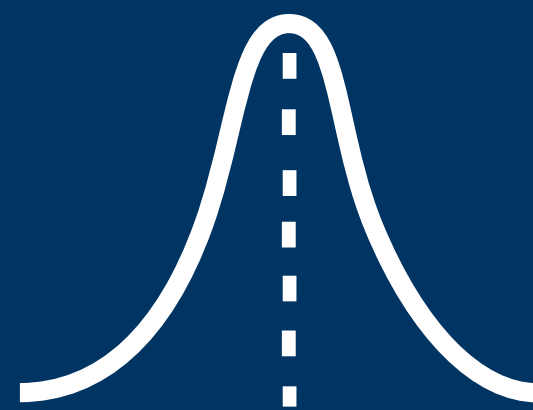


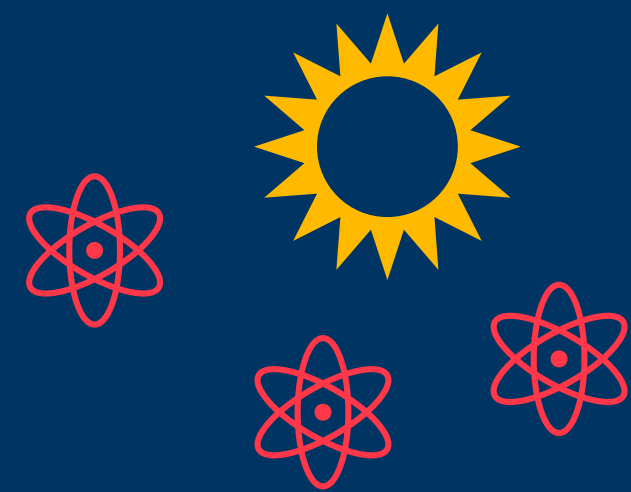
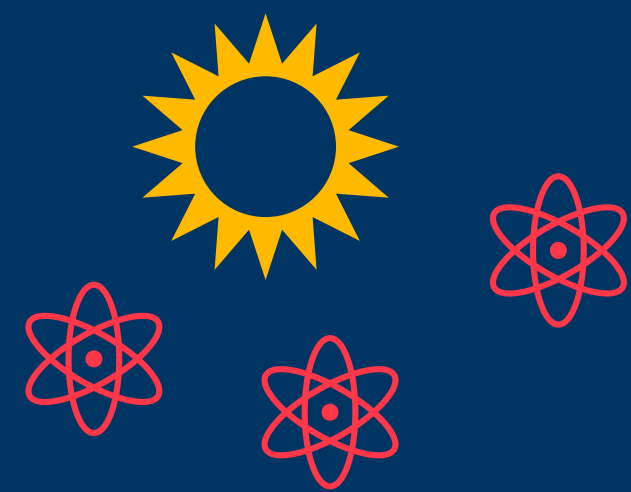
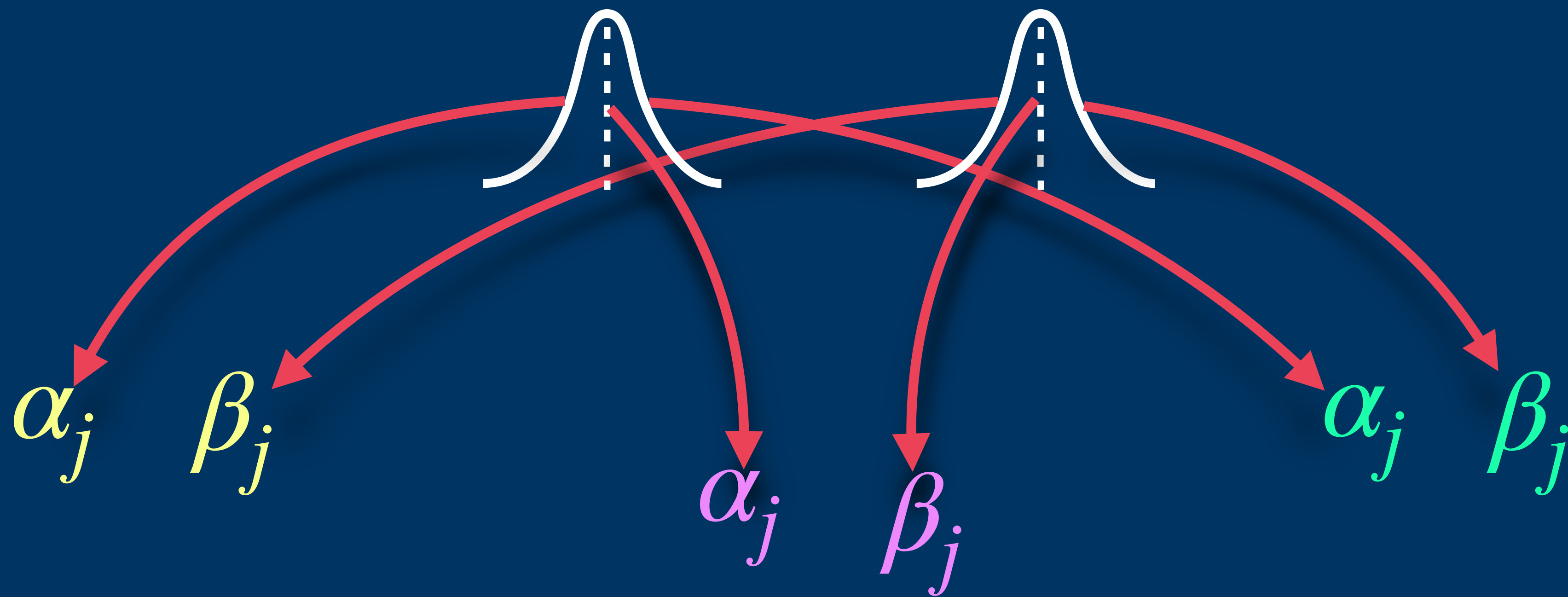
α

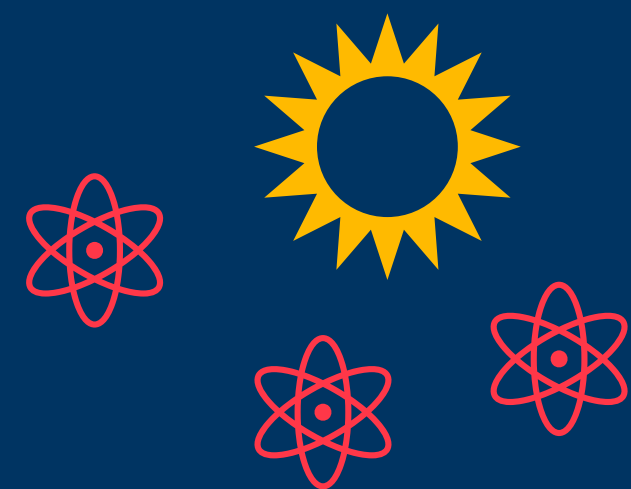
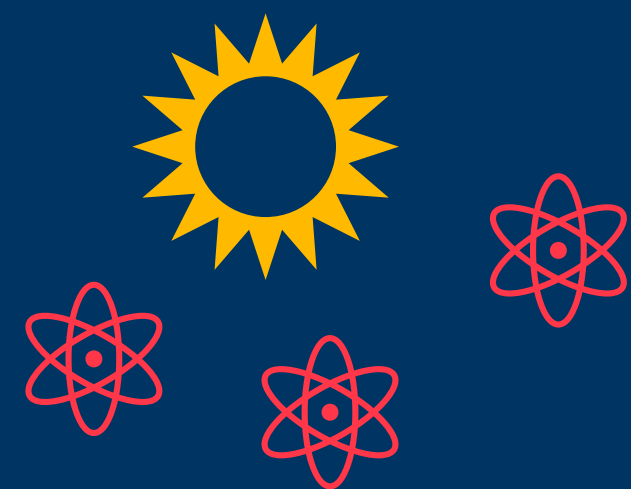
β





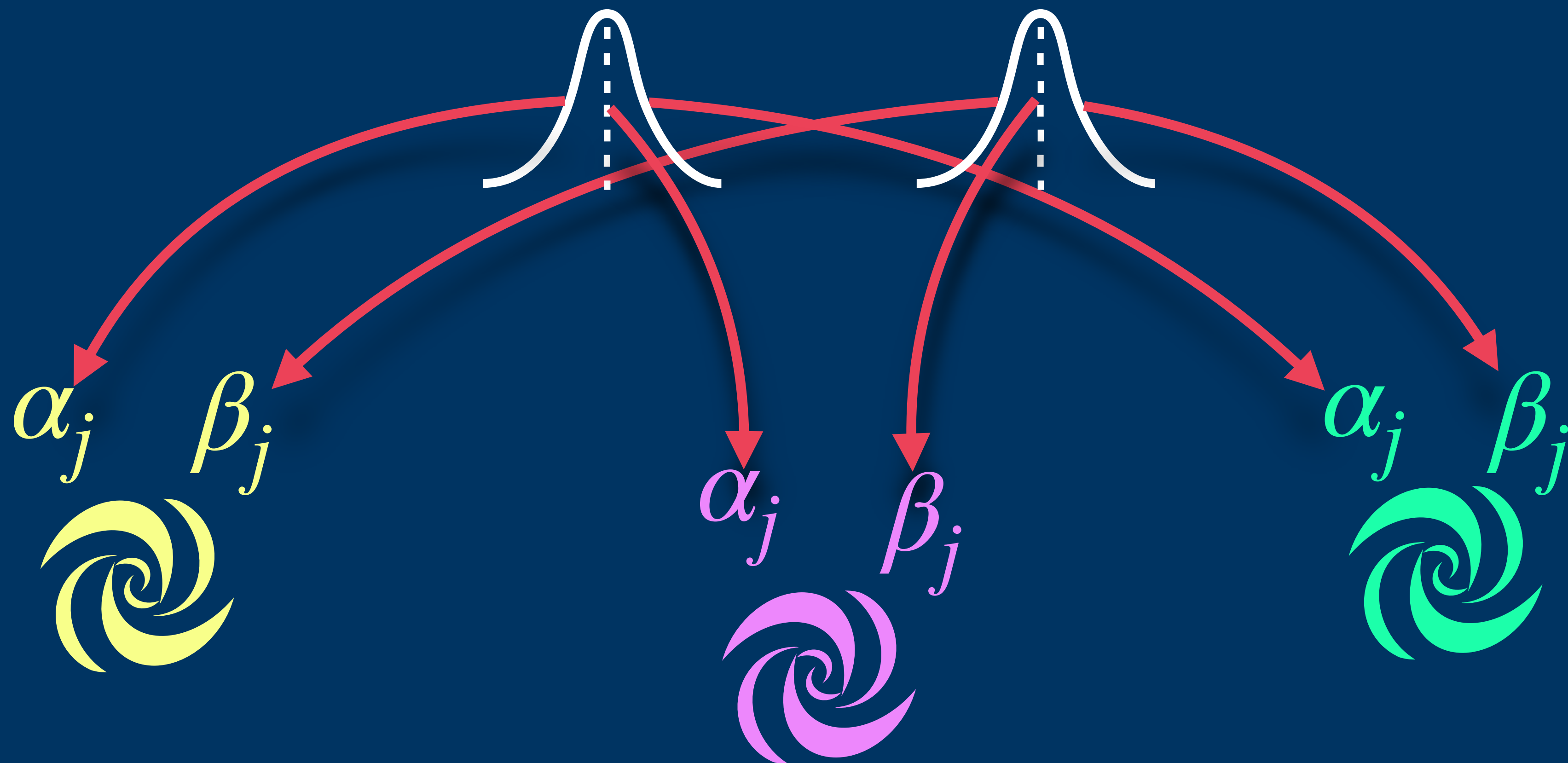
 μ_α σ_α μ_β σ_β α β 

 μ_α σ_α μ_β σ_β α β 



μ_α σ_α μ_β σ_β

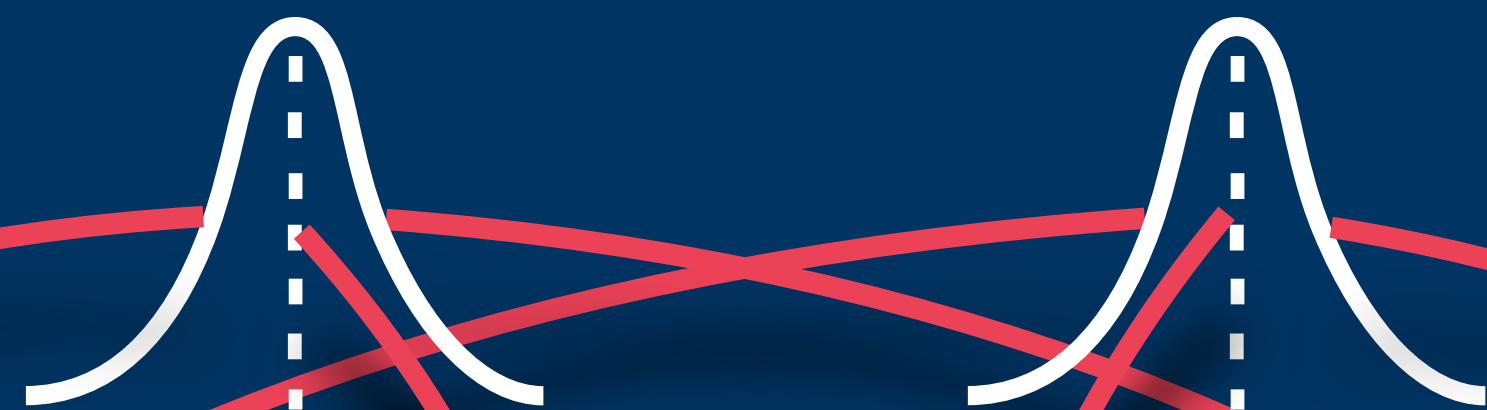
α β





α

β



α_j β_j



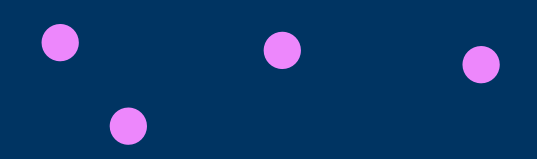
$$L_i = \alpha_j M_i + \beta_j$$



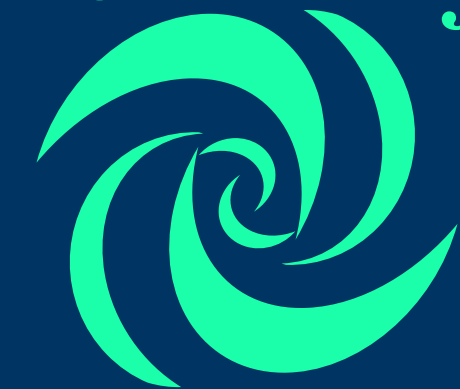
α_j β_j



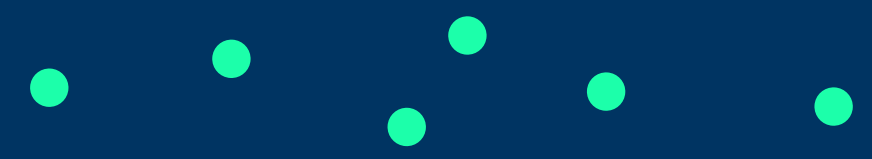
$$L_i = \alpha_j M_i + \beta_j$$

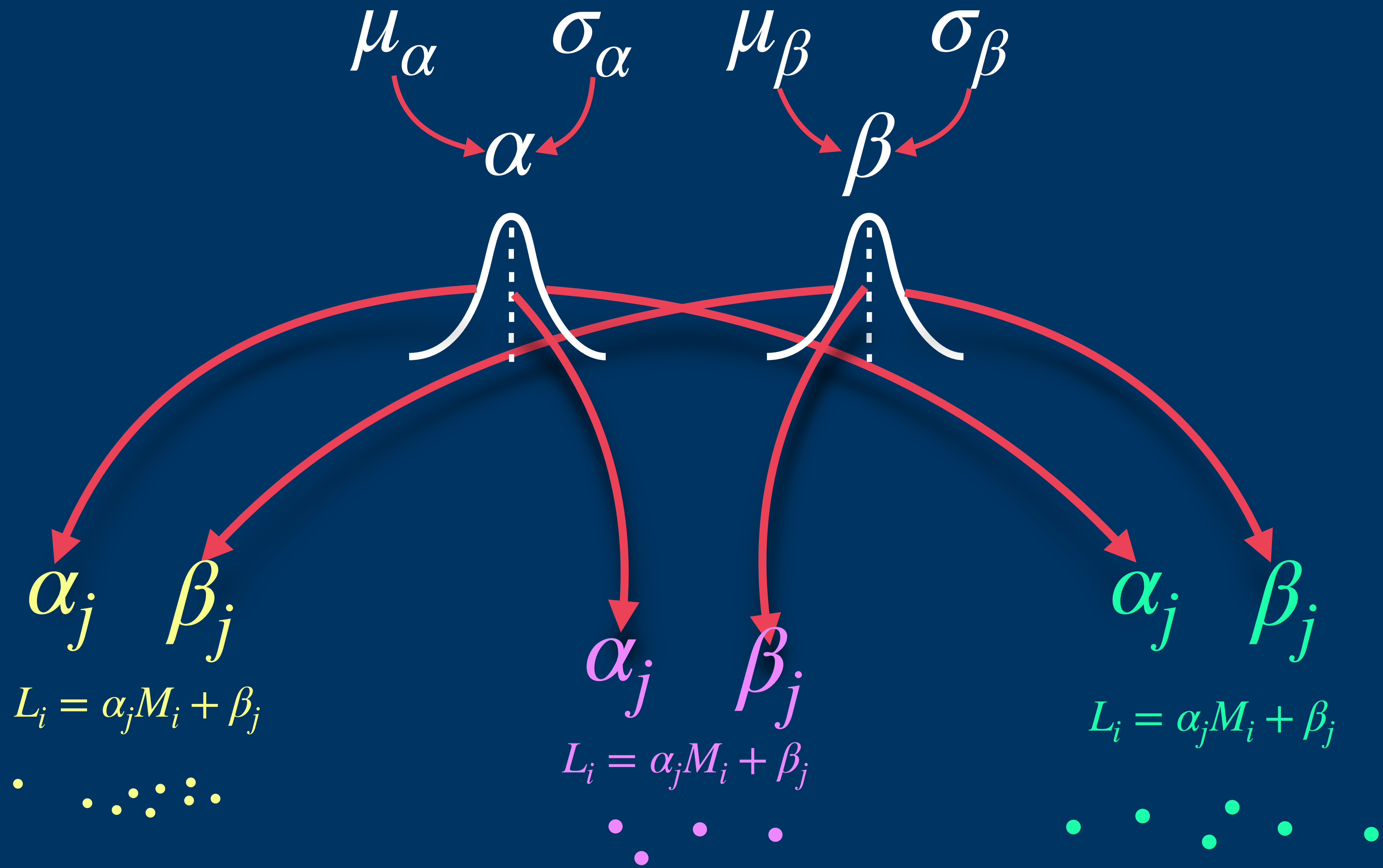


α_j β_j

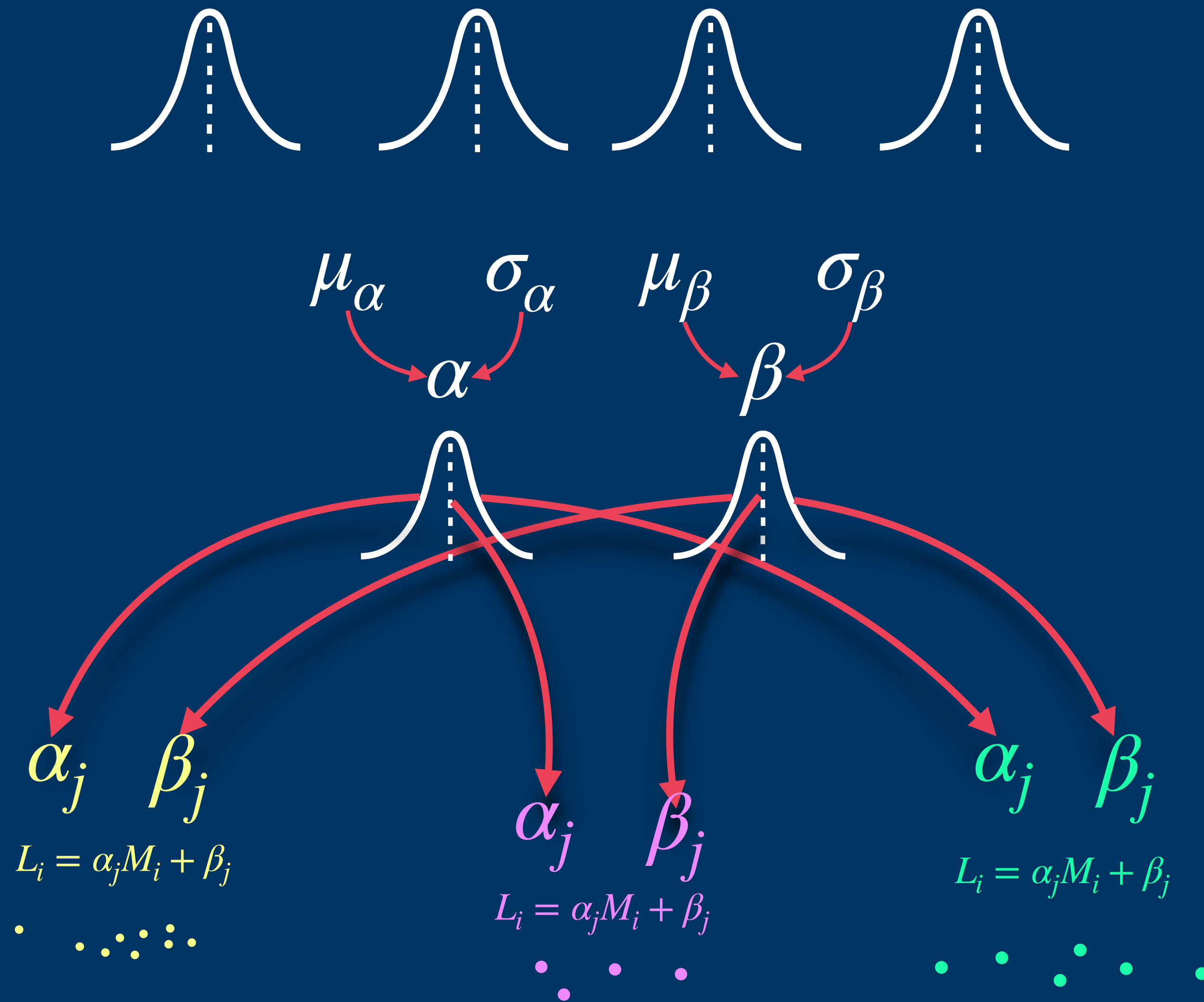


$$L_i = \alpha_j M_i + \beta_j$$





$$\mathcal{L}(\hat{L}_i, \hat{M}_i | L_i, M_i) = \pi(\hat{L}_i, \hat{M}_i | L_i, M_i)$$



$$\mathcal{L}(\hat{L}_i, \hat{M}_i | L_i, M_i) = \pi(\hat{L}_i, \hat{M}_i | L_i, M_i)$$

$$\pi(\mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta \mid \hat{L}_i, \hat{M}_i) \propto \int \pi(\hat{L}_i, \hat{M}_i \mid L_i, M_i; \alpha_j, \beta_j) \pi(\alpha_j \mid \mu_\alpha, \sigma_\alpha) \pi(\beta_j \mid \mu_\beta, \sigma_\beta) \pi(\mu_\alpha) \pi(\sigma_\alpha) \pi(\mu_\beta) \pi(\sigma_\beta)$$

posterior

$$\pi(\mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta \mid \hat{L}_i, \hat{M}_i) \propto \int \pi(\hat{L}_i, \hat{M}_i \mid L_i, M_i; \alpha_j, \beta_j) \pi(\alpha_j \mid \mu_\alpha, \sigma_\alpha) \pi(\beta_j \mid \mu_\beta, \sigma_\beta) \pi(\mu_\alpha) \pi(\sigma_\alpha) \pi(\mu_\beta) \pi(\sigma_\beta)$$

posterior likelihood

$$\pi(\mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta \mid \hat{L}_i, \hat{M}_i) \propto \int \underbrace{\pi(\hat{L}_i, \hat{M}_i \mid L_i, M_i; \alpha_j, \beta_j)}_{\text{likelihood}} \underbrace{\pi(\alpha_j \mid \mu_\alpha, \sigma_\alpha) \pi(\beta_j \mid \mu_\beta, \sigma_\beta)}_{\text{group level priors}} \pi(\mu_\alpha) \pi(\sigma_\alpha) \pi(\mu_\beta) \pi(\sigma_\beta)$$

posterior

$$\pi(\mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta \mid \hat{L}_i, \hat{M}_i) \propto \int \underbrace{\pi(\hat{L}_i, \hat{M}_i \mid L_i, M_i; \alpha_j, \beta_j)}_{\text{likelihood}} \underbrace{\pi(\alpha_j \mid \mu_\alpha, \sigma_\alpha) \pi(\beta_j \mid \mu_\beta, \sigma_\beta)}_{\text{group level priors}} \underbrace{\pi(\mu_\alpha) \pi(\sigma_\alpha) \pi(\mu_\beta) \pi(\sigma_\beta)}_{\text{hyper priors}}$$

posterior

$$\pi(\mu_\alpha, \sigma_\alpha, \mu_\beta, \sigma_\beta \mid \hat{L}_i, \hat{M}_i) \propto \int \underbrace{\pi(\hat{L}_i, \hat{M}_i \mid L_i, M_i; \alpha_j, \beta_j)}_{\text{likelihood}} \underbrace{\pi(\alpha_j \mid \mu_\alpha, \sigma_\alpha) \pi(\beta_j \mid \mu_\beta, \sigma_\beta)}_{\text{group level priors}} \underbrace{\pi(\mu_\alpha) \pi(\sigma_\alpha) \pi(\mu_\beta) \pi(\sigma_\beta)}_{\text{hyper priors}}$$

posterior



Stan



PYMC3

cosmology

$$E(z) = \sqrt{\Omega_m (1+z')^3 + \Omega_k (1+z')^2 + \Omega_\Lambda (1+z')^{3(1+w)}}$$

luminosity distance

$$d_L = (1+z) \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')}$$

distance modulus

$$\mu_C = 5 \log \left[\frac{d_L}{10 \text{pc}} \right]$$

μ_{obs}

z

cosmology

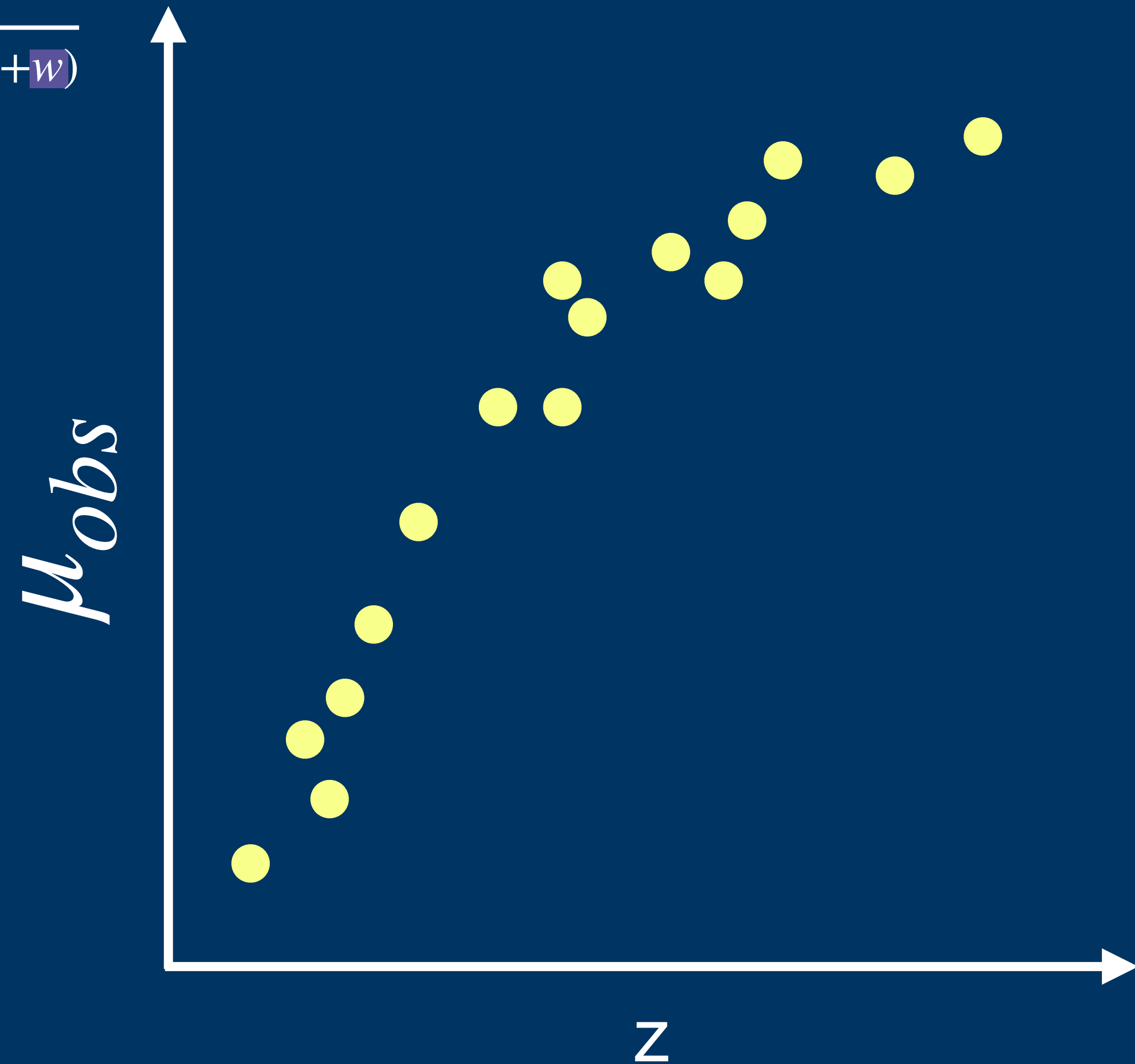
$$E(z) = \sqrt{\Omega_m (1+z')^3 + \Omega_k (1+z')^2 + \Omega_\Lambda (1+z')^{3(1+w)}}$$

luminosity distance

$$d_L = (1+z) \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')}$$

distance modulus

$$\mu_C = 5 \log \left[\frac{d_L}{10 \text{pc}} \right]$$



cosmology

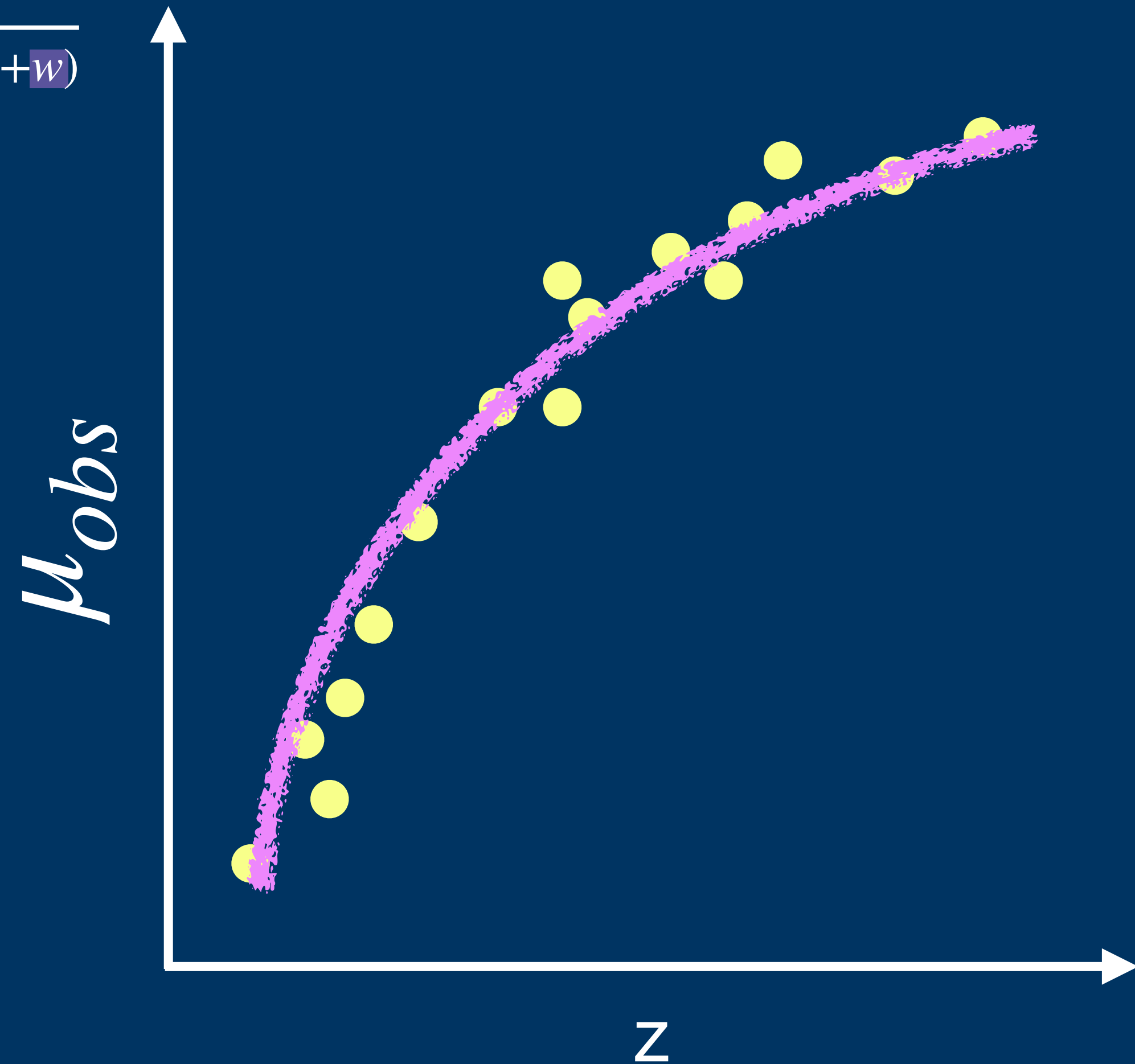
$$E(z) = \sqrt{\Omega_m (1+z)^3 + \Omega_k (1+z)^2 + \Omega_\Lambda (1+z)^{3(1+w)}}$$

luminosity distance

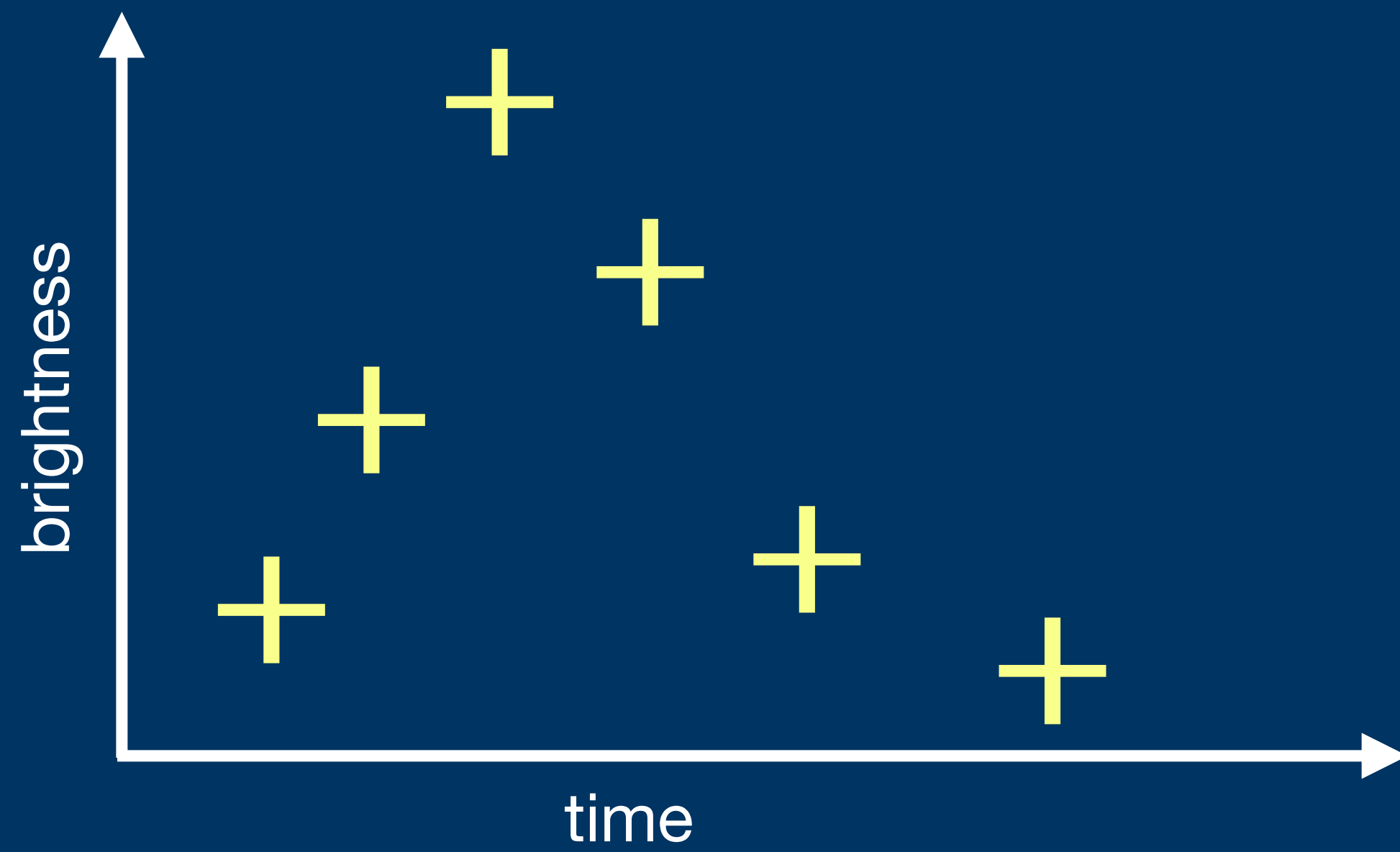
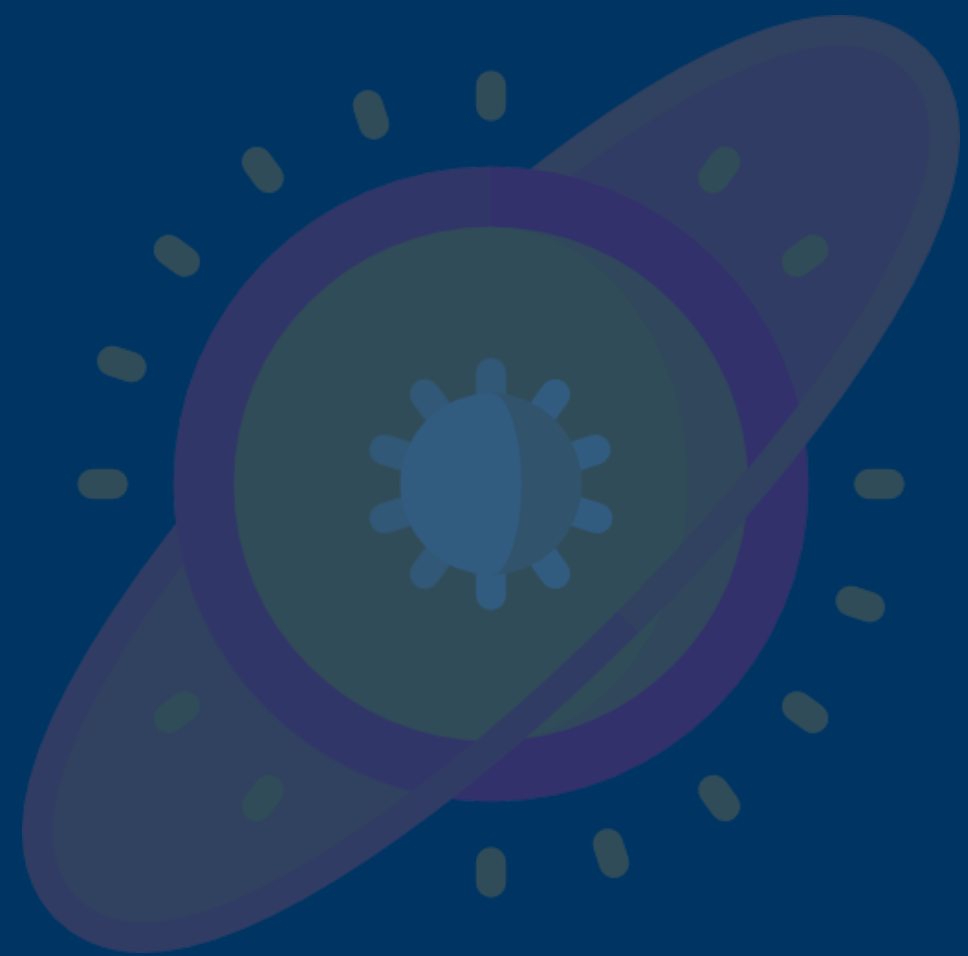
$$d_L = (1+z) \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')}$$

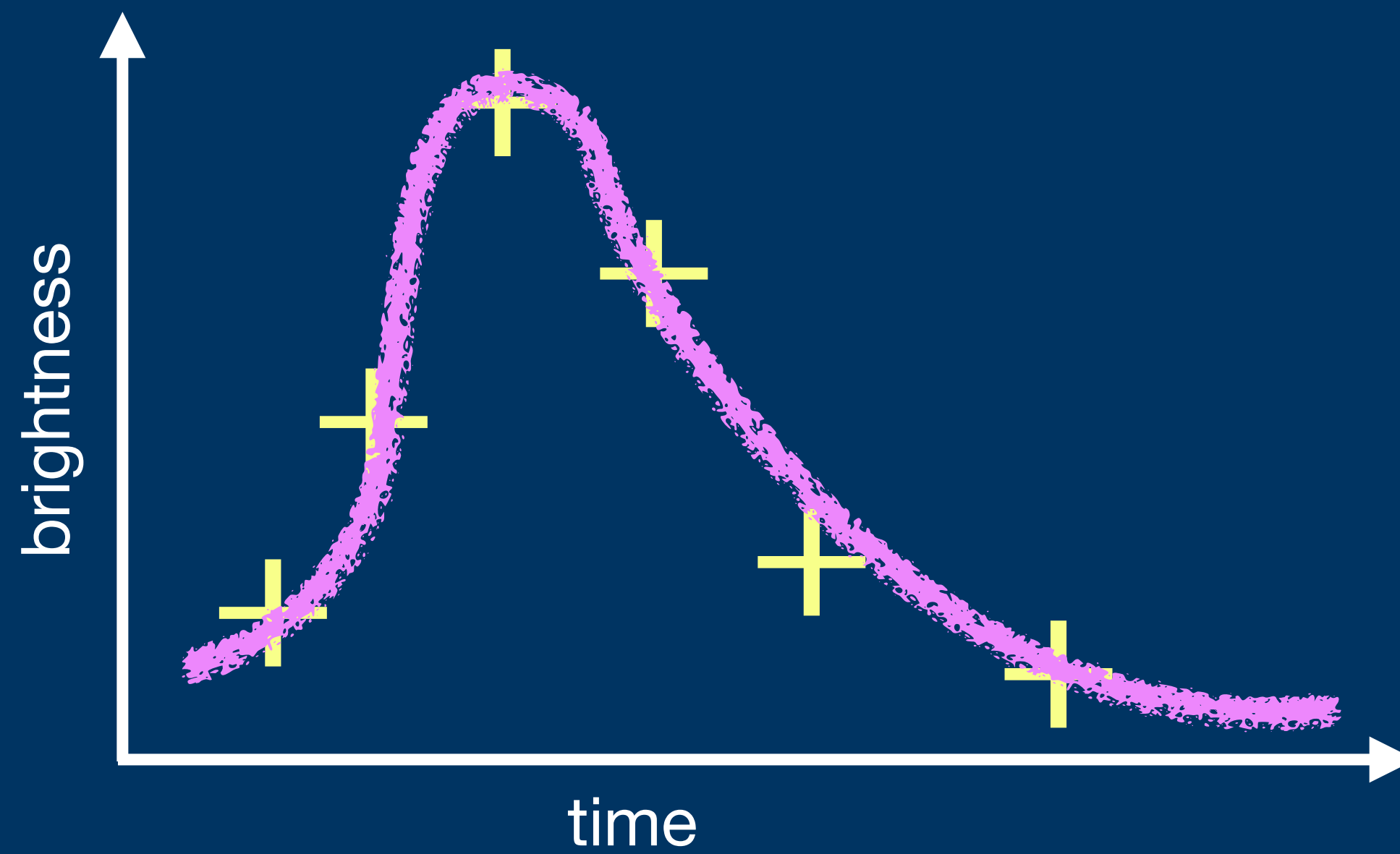
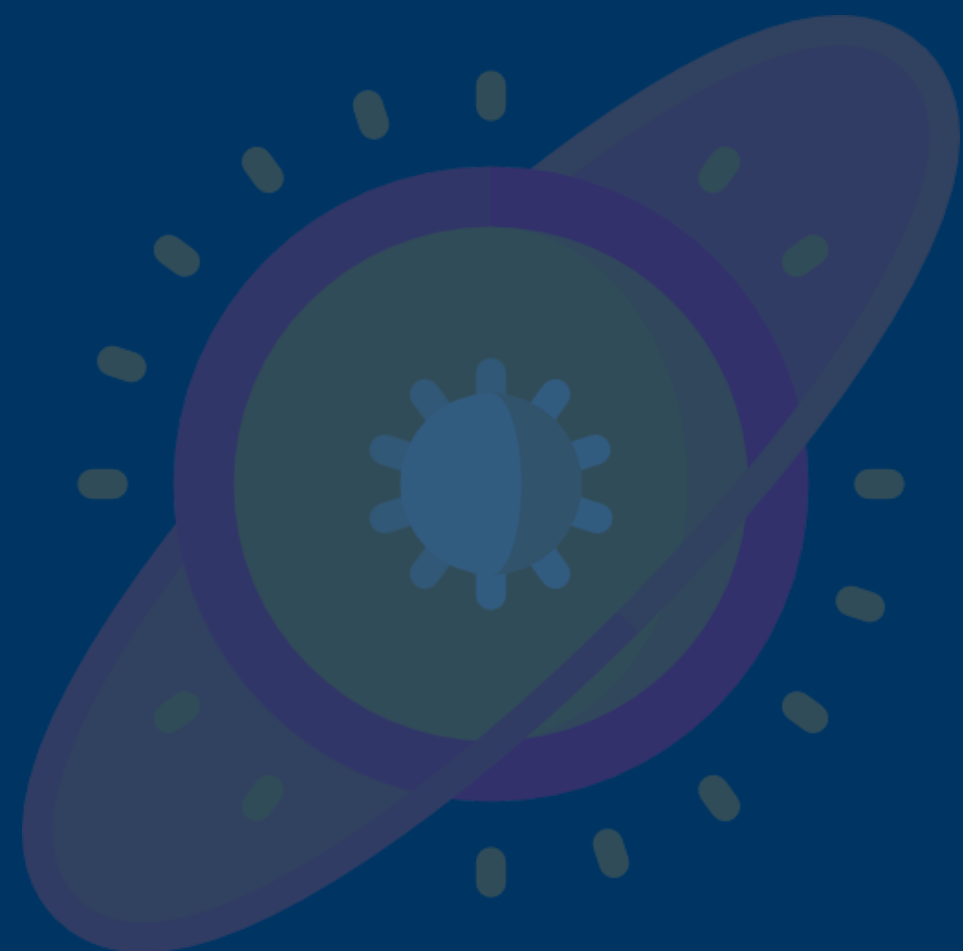
distance modulus

$$\mu_C = 5 \log \left[\frac{d_L}{10 \text{pc}} \right]$$





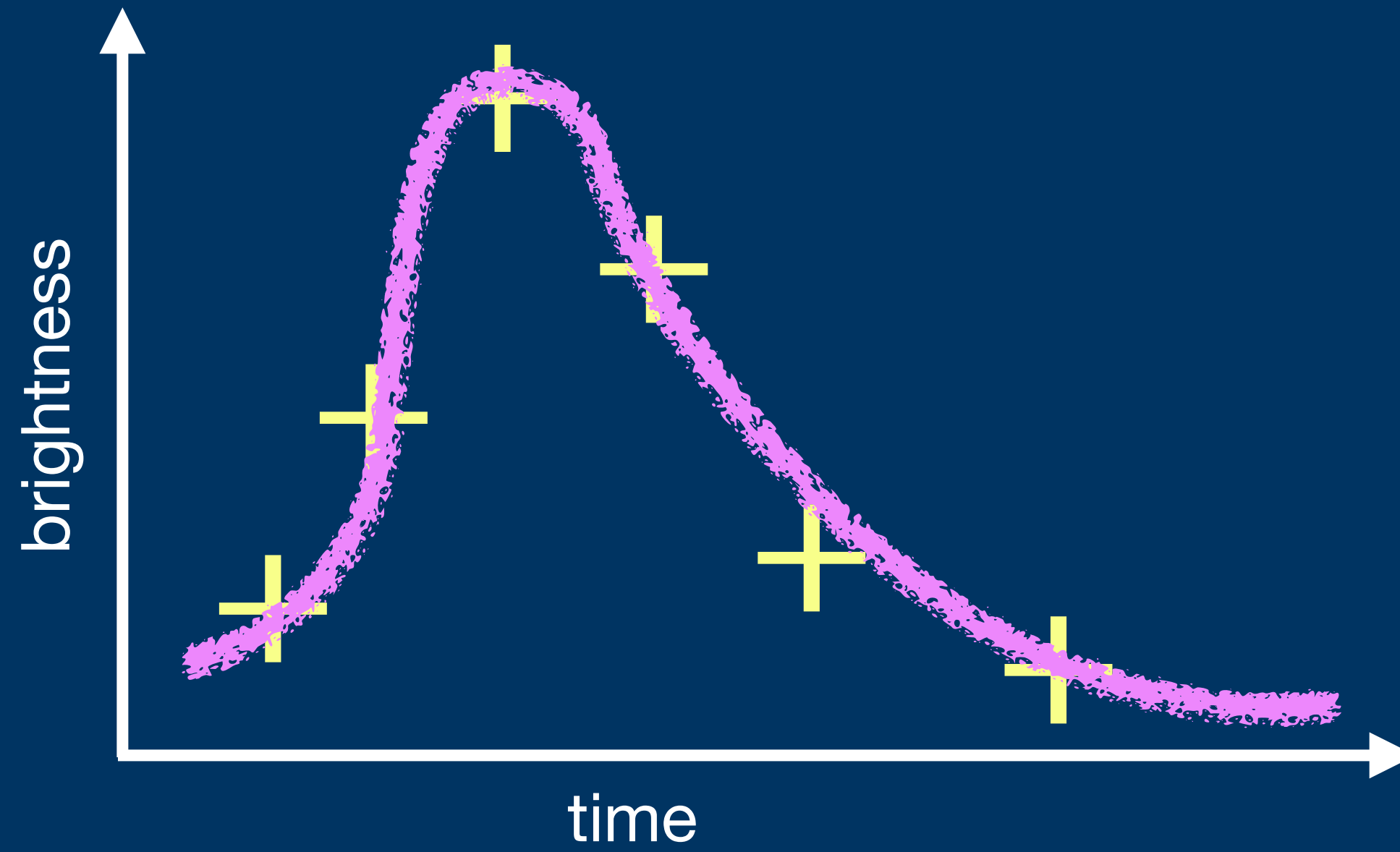
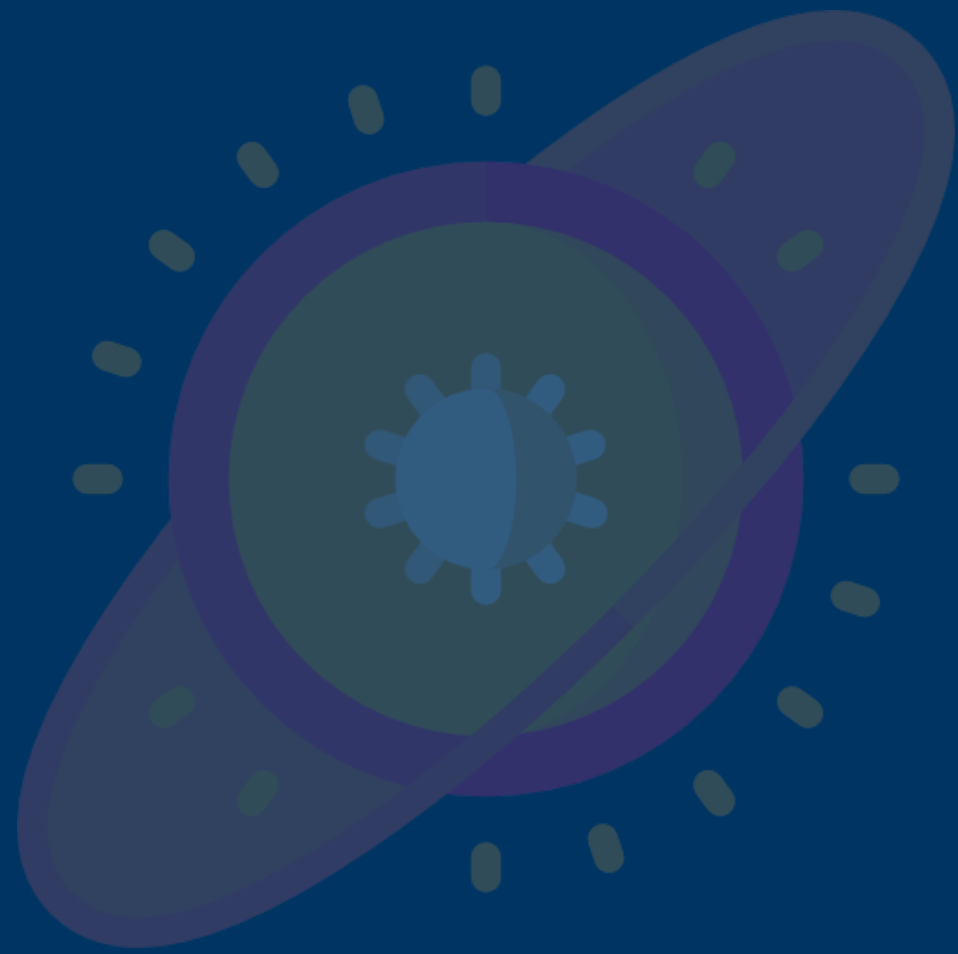




m_B

x_1

c



$$\mu_{obs} = m_B - M_B$$

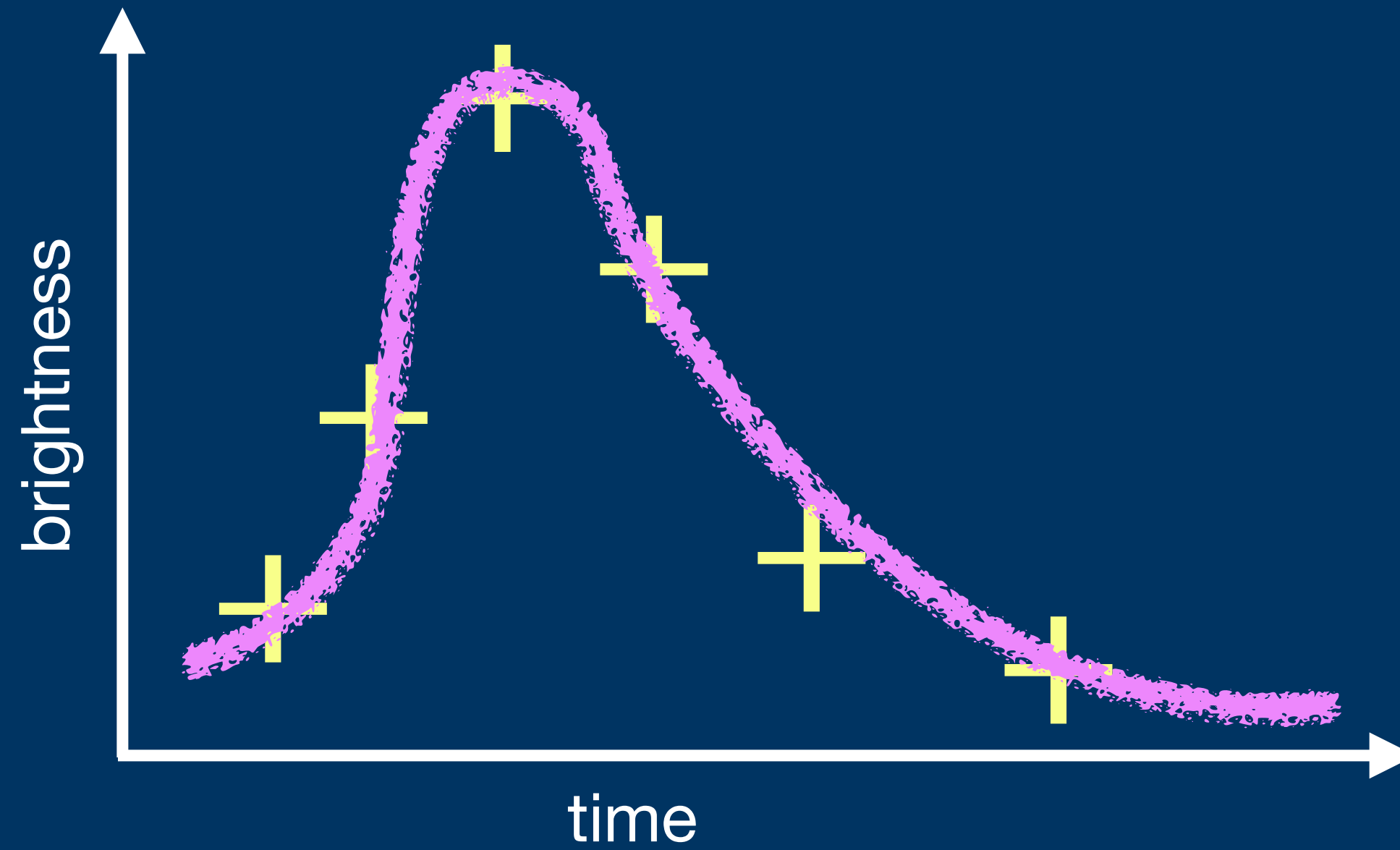
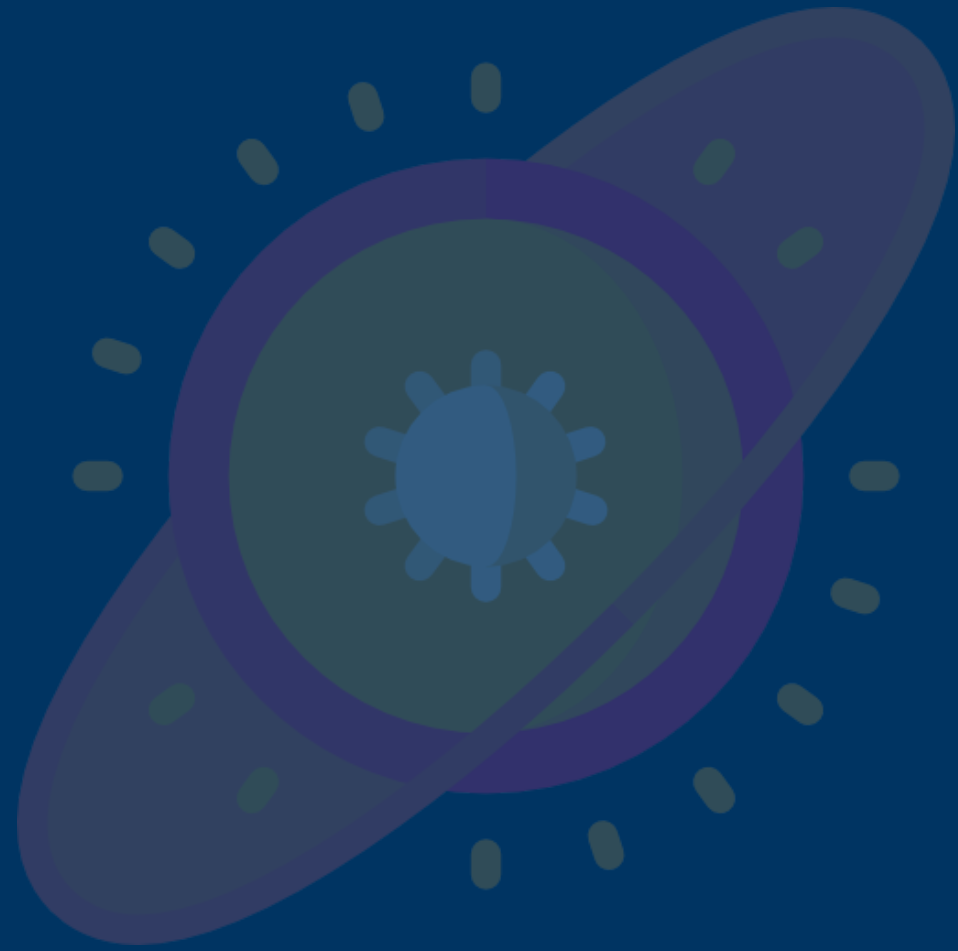
average brightness

apparent brightness

m_B

x_1

c



$$\mu_{obs} = m_B - M_B$$

average brightness

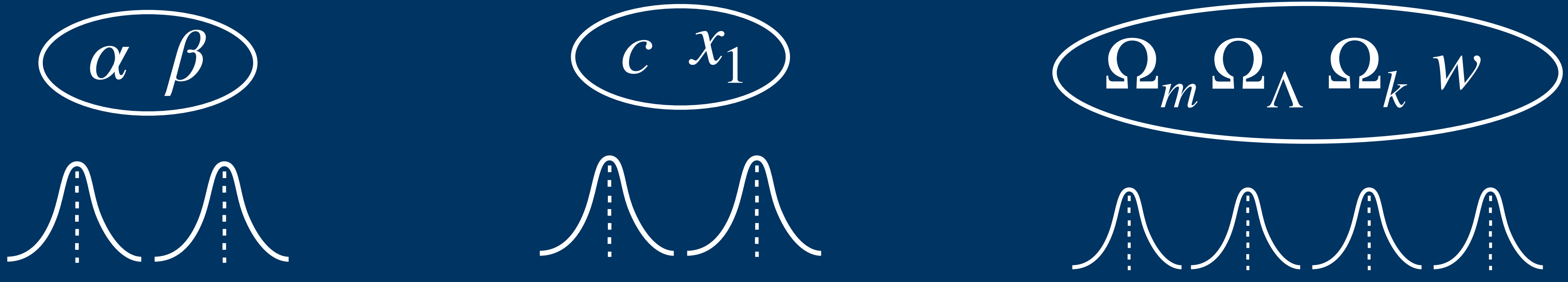
apparent brightness

m_B

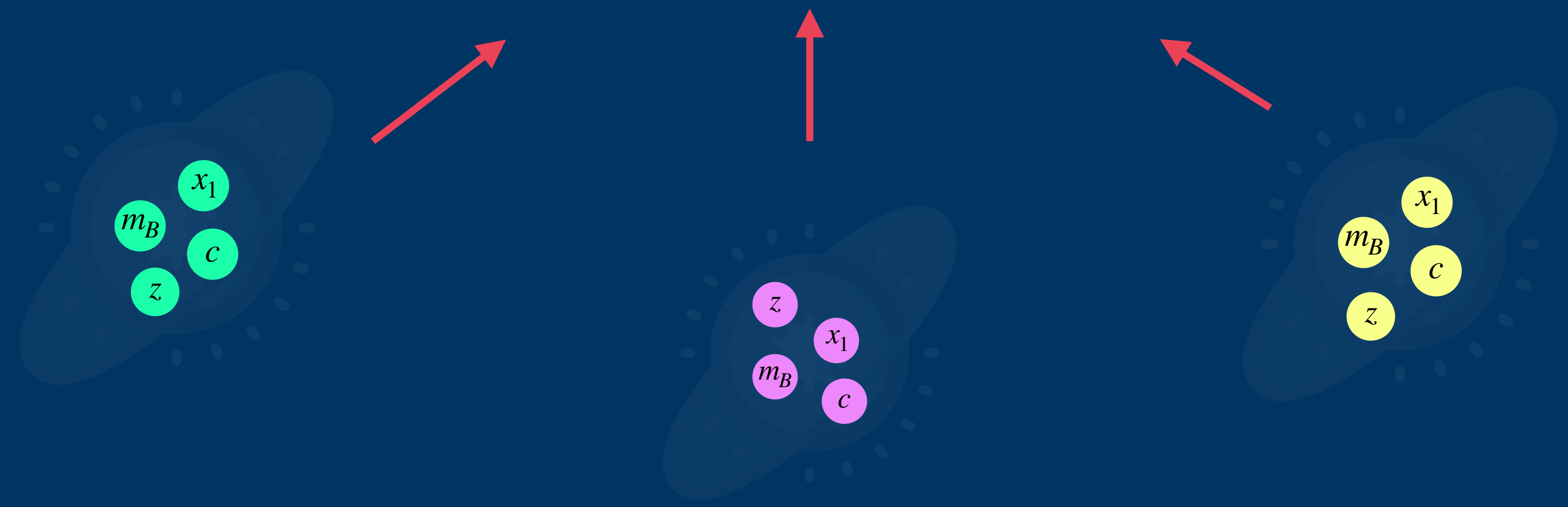
x_1

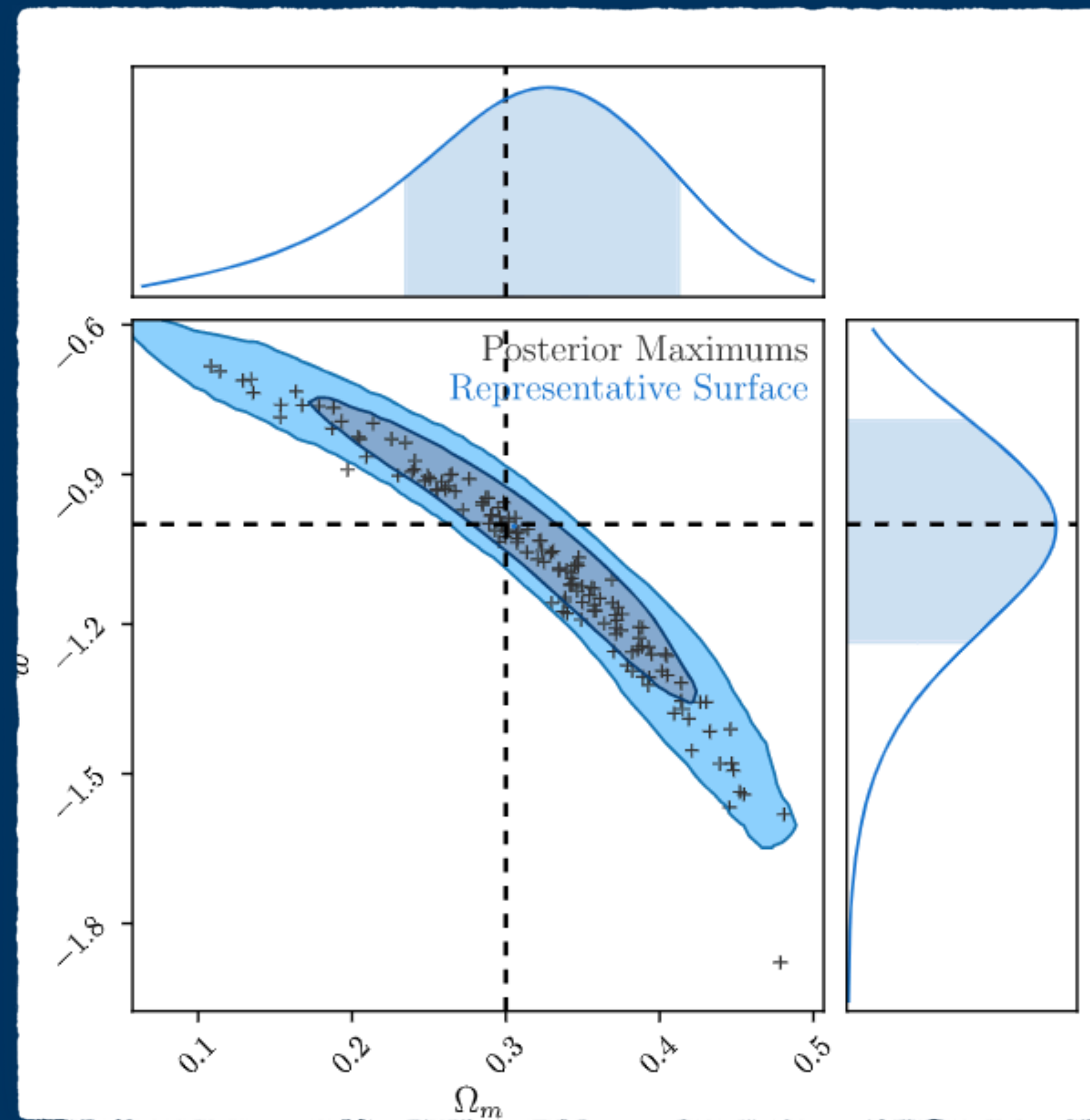
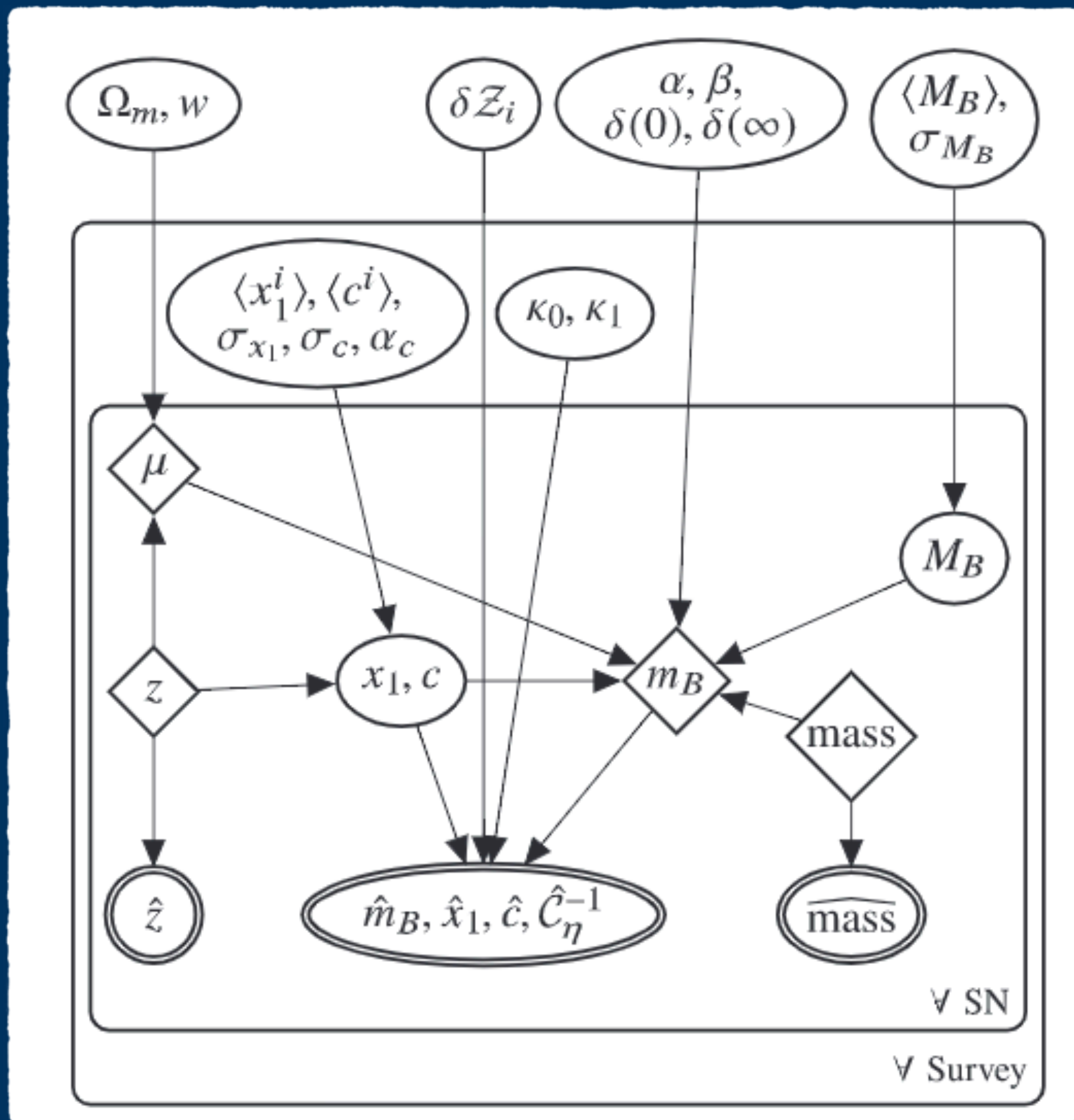
c

$$\mu_{obs} = m_B + \alpha x_1 - \beta c - M_B$$



$$m_B^i = + \alpha x_1^i - \beta c^i - M_B - \mu^i$$



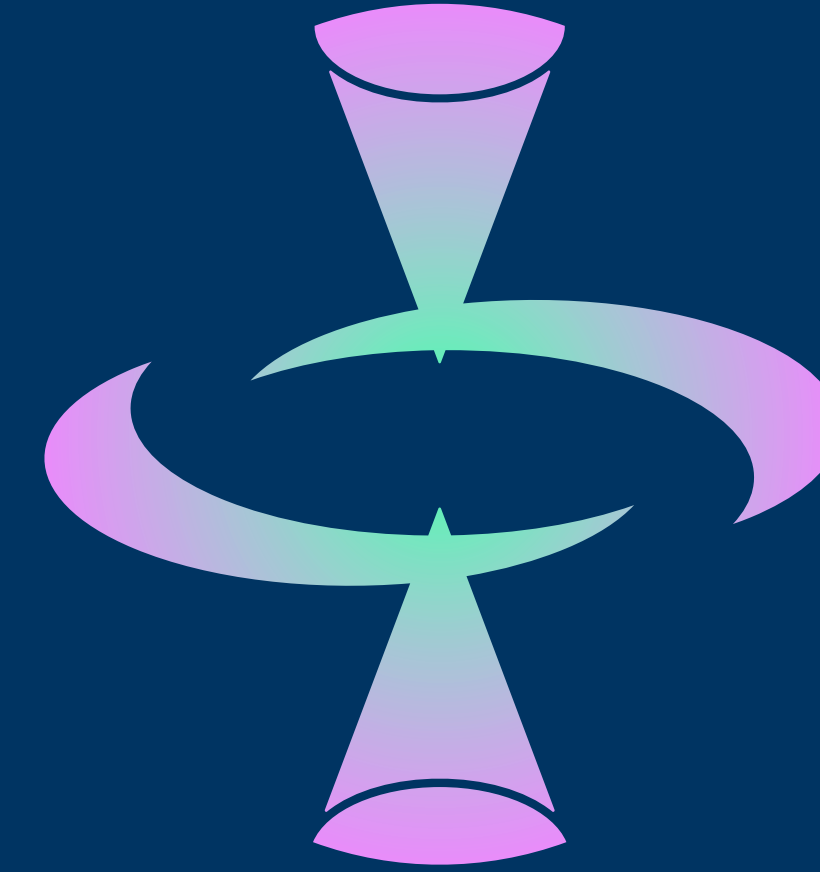
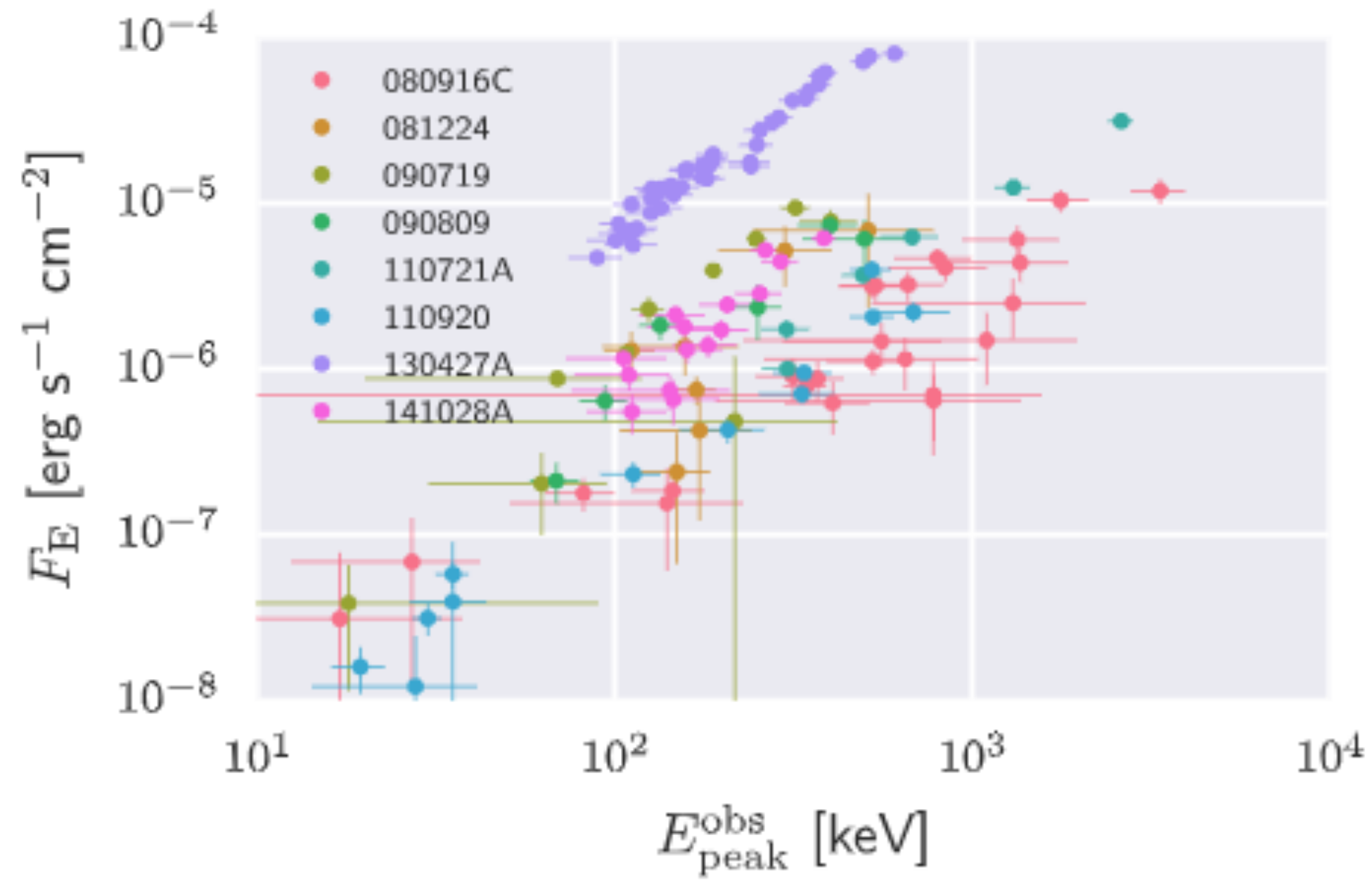


The rest-frame Golenetskii correlation via a hierarchical Bayesian analysis

J. Michael Burgess^{1,2★}

¹The Oskar Klein Centre for Cosmoparticle Physics, SE-106 91 Stockholm, Sweden

²Department of Physics, KTH Royal Institute of Technology, AlbaNova, SE-106 91 Stockholm, Sweden



$$L = N_{\text{rest}} \left(\frac{E_{\text{peak}}^{\text{rest}}}{100 \text{keV}} \right)^{\gamma} \text{ erg s}^{-1}$$

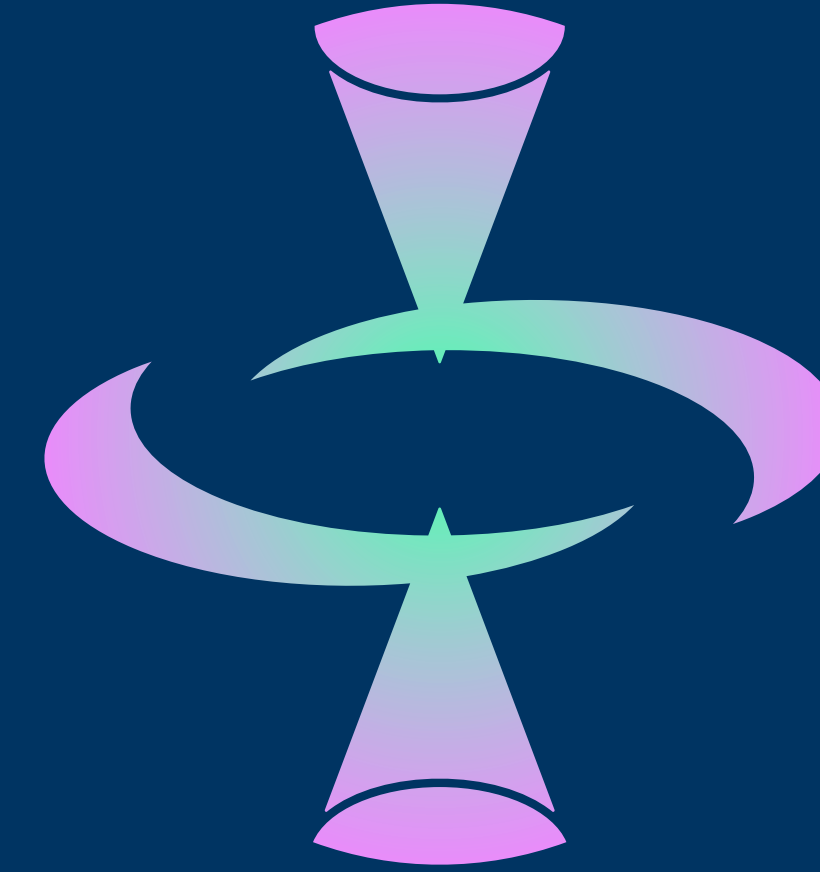
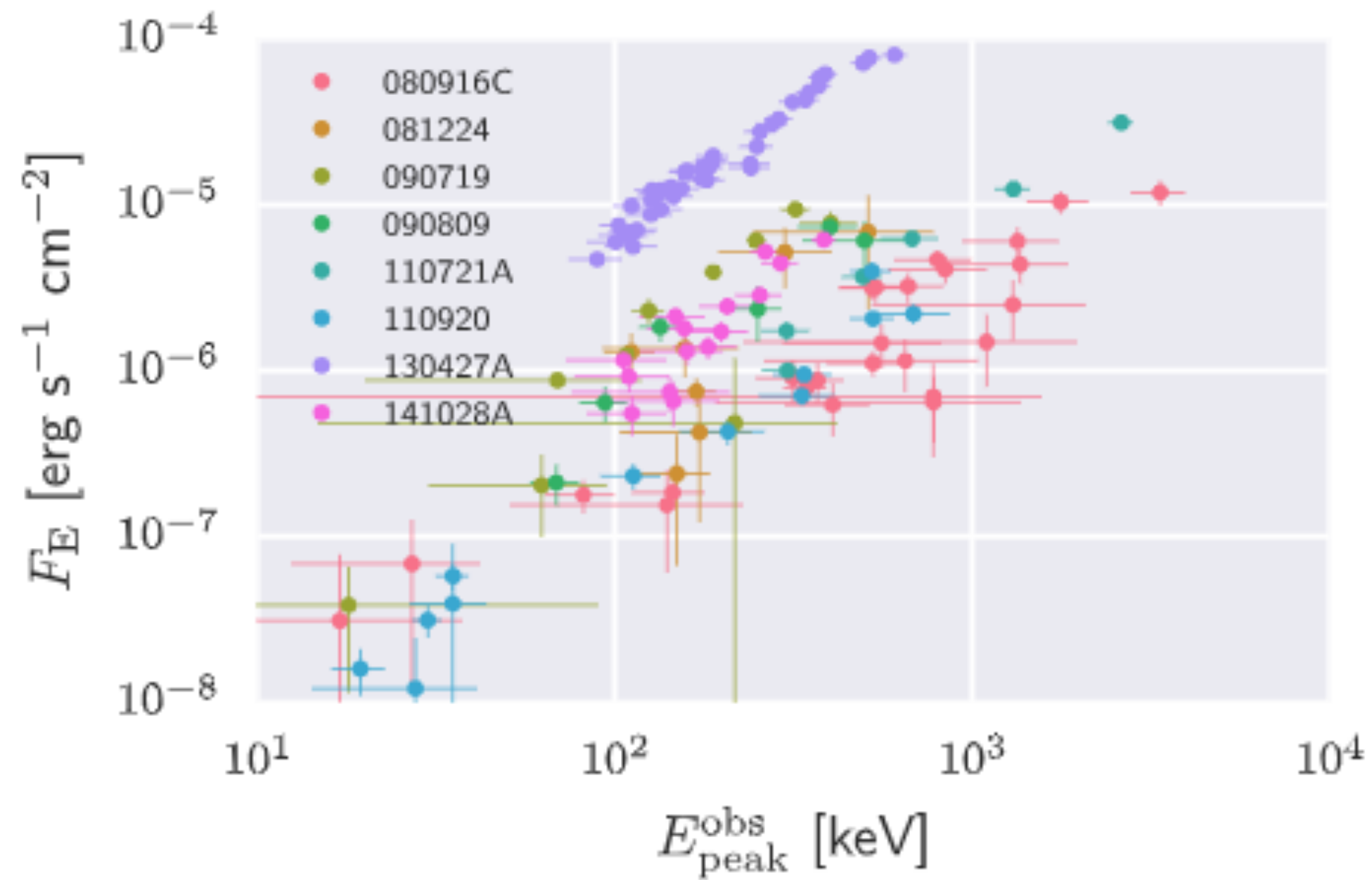
$$F_E = \frac{N_{\text{rest}}}{4\pi d_L^2(z)} \left(\frac{E_{\text{peak}}^{\text{obs}} (1+z)}{100 \text{keV}} \right)^{\gamma} \text{ erg s}^{-1} \text{ cm}^{-2}$$

The rest-frame Golenetskii correlation via a hierarchical Bayesian analysis

J. Michael Burgess^{1,2★}

¹The Oskar Klein Centre for Cosmoparticle Physics, SE-106 91 Stockholm, Sweden

²Department of Physics, KTH Royal Institute of Technology, AlbaNova, SE-106 91 Stockholm, Sweden



$$L = N_{\text{rest}} \left(\frac{E_{\text{peak}}^{\text{rest}}}{100 \text{keV}} \right)^{\gamma} \text{ erg s}^{-1}$$

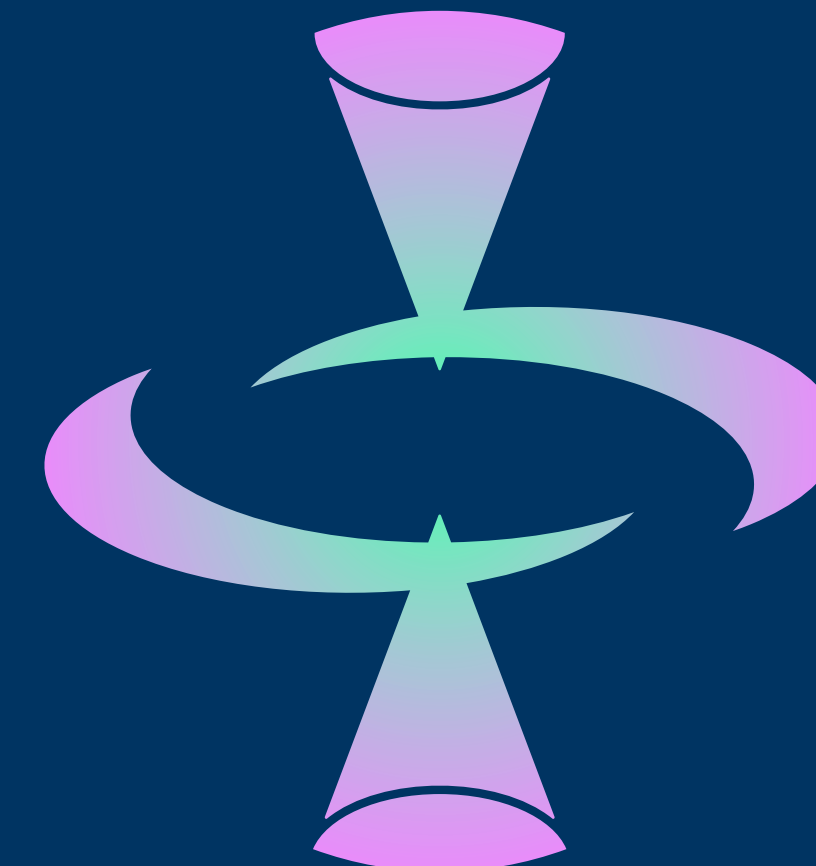
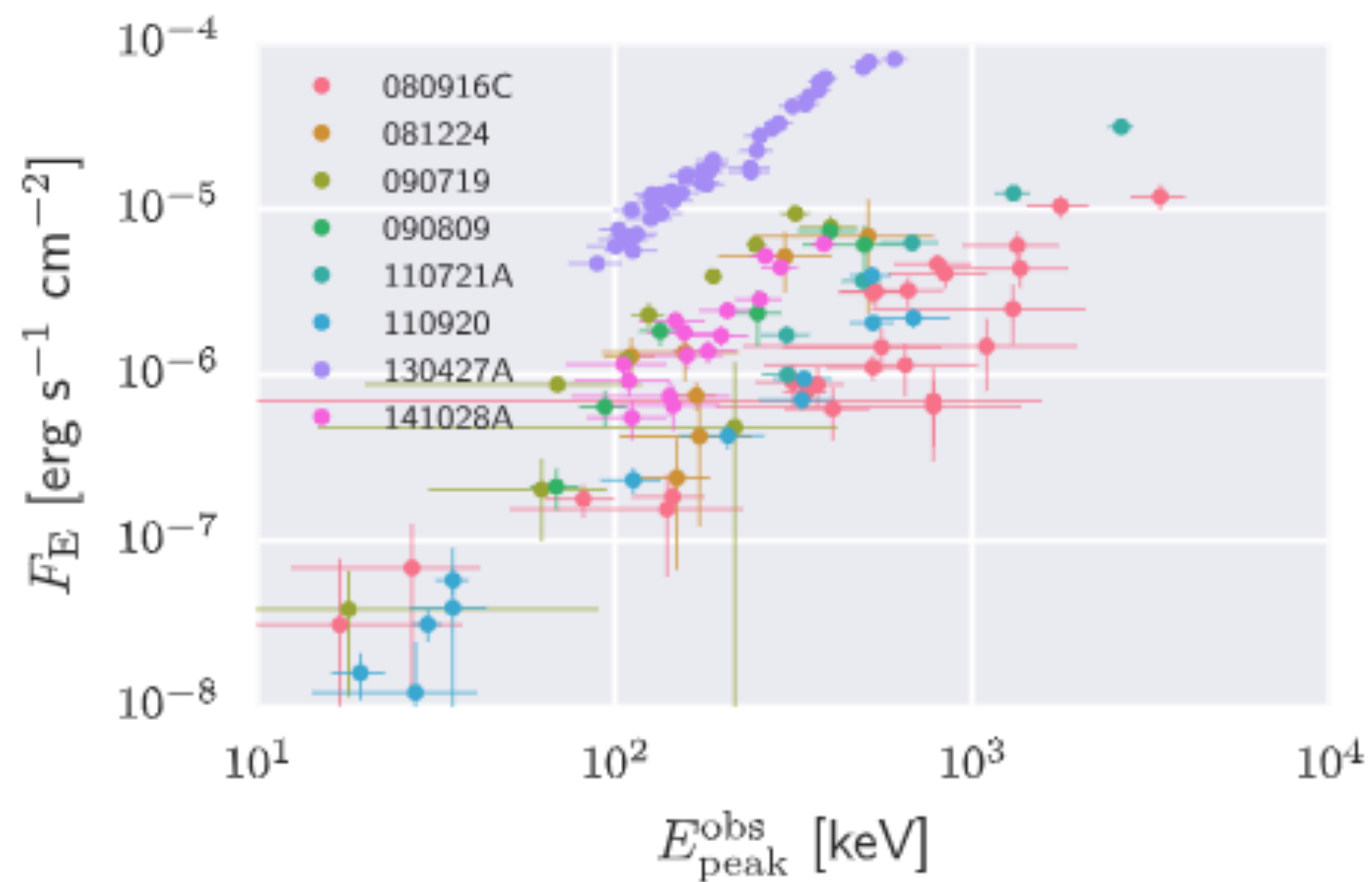
$$F_E = \frac{N_{\text{rest}}}{4\pi d_L^2(z)} \left(\frac{E_{\text{peak}}^{\text{obs}} (1+z)}{100 \text{keV}} \right)^{\gamma} \text{ erg s}^{-1} \text{ cm}^{-2}$$

The rest-frame Golenetskii correlation via a hierarchical Bayesian analysis

J. Michael Burgess^{1,2★}

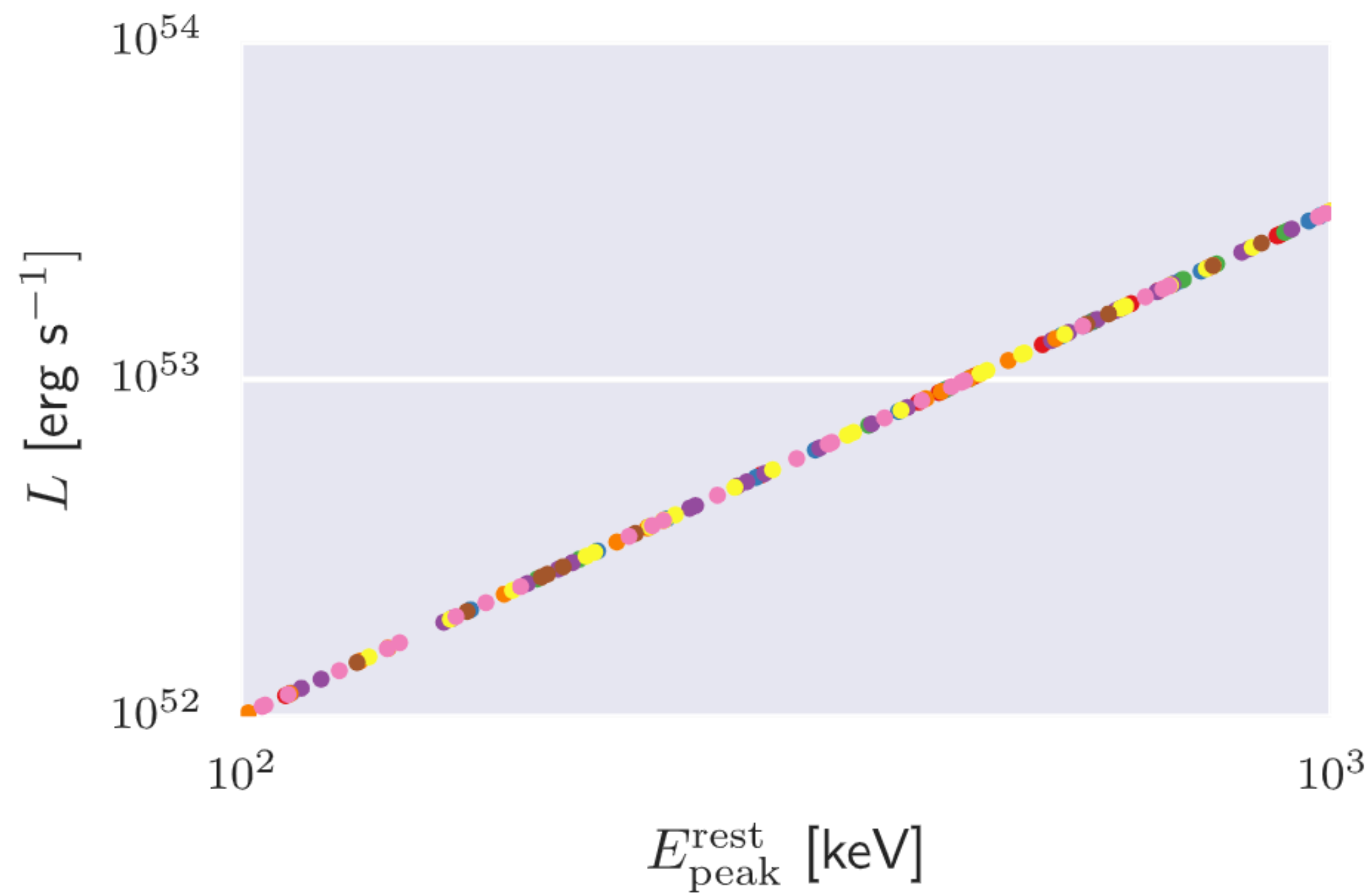
¹The Oskar Klein Centre for Cosmoparticle Physics, SE-106 91 Stockholm, Sweden

²Department of Physics, KTH Royal Institute of Technology, AlbaNova, SE-106 91 Stockholm, Sweden

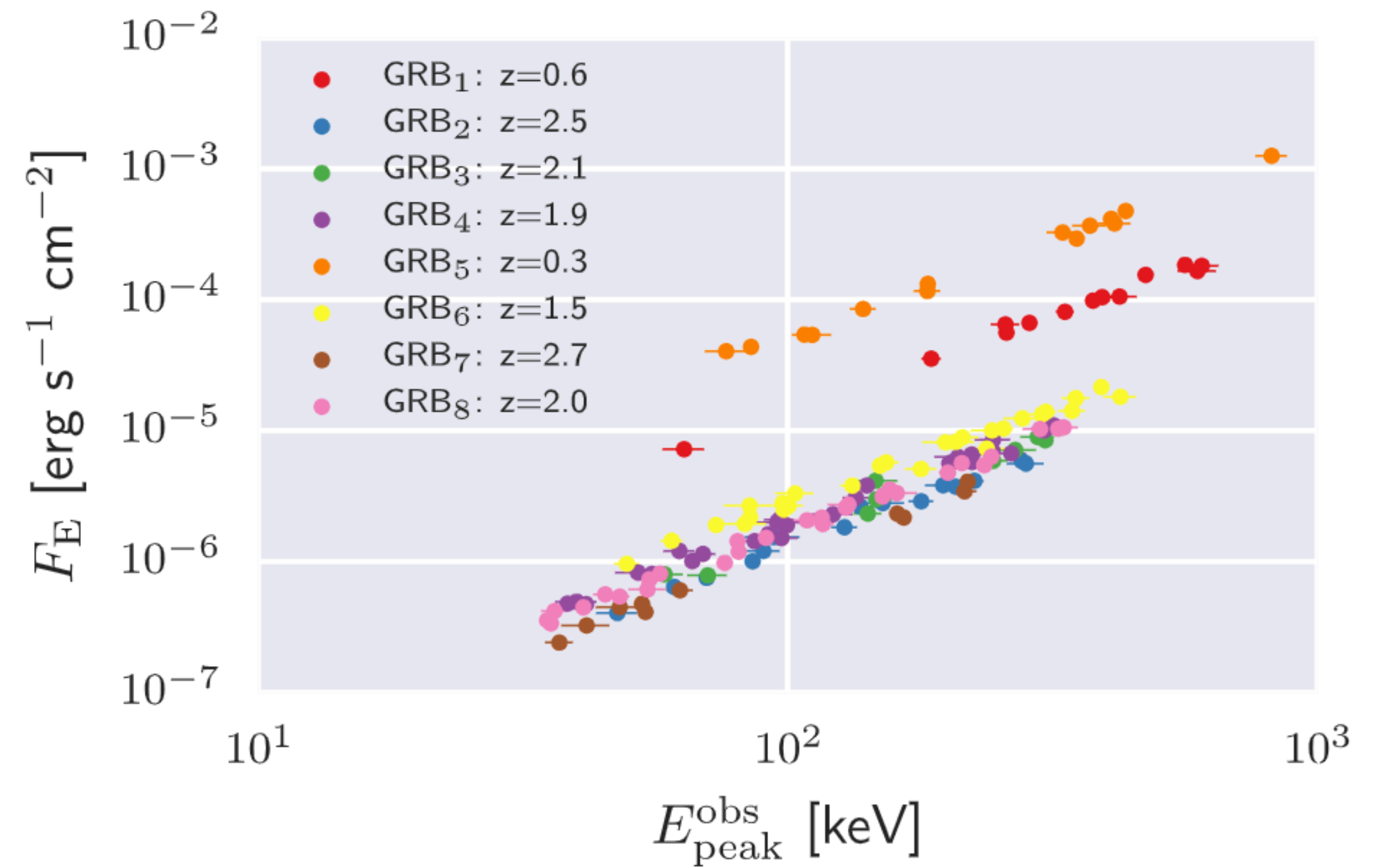


$$L = N_{\text{rest}} \left(\frac{E_{\text{peak}}^{\text{rest}}}{100 \text{keV}} \right)^{\gamma} \text{ erg s}^{-1}$$

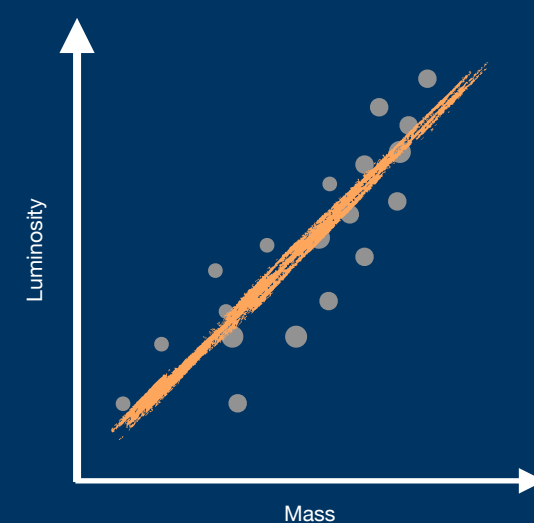
$$F_E = \frac{N_{\text{rest}}}{4\pi d_L^2(z)} \left(\frac{E_{\text{peak}}^{\text{obs}} (1+z)}{100 \text{keV}} \right)^{\gamma} \text{ erg s}^{-1} \text{ cm}^{-2}$$

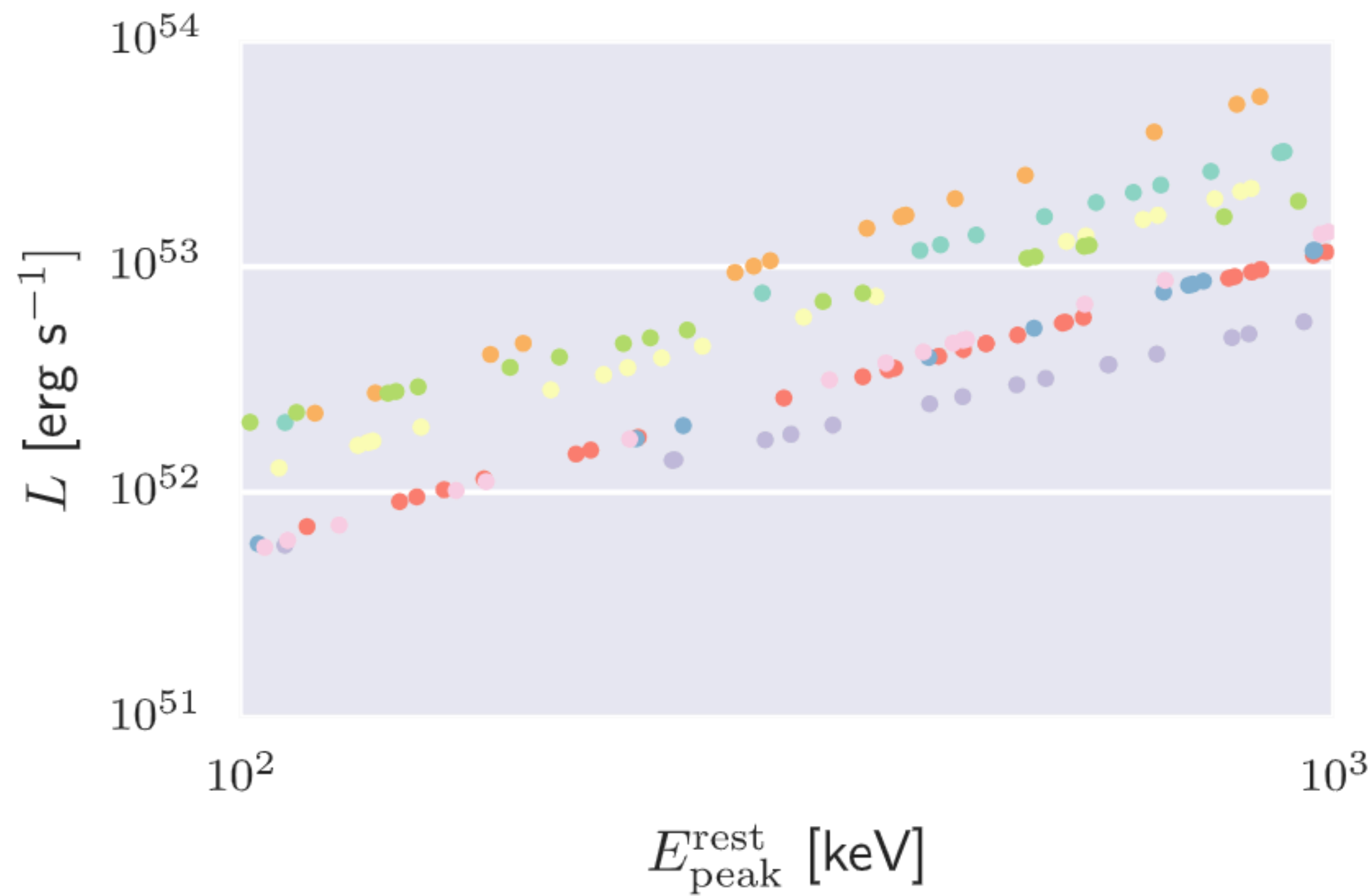


Rest Frame

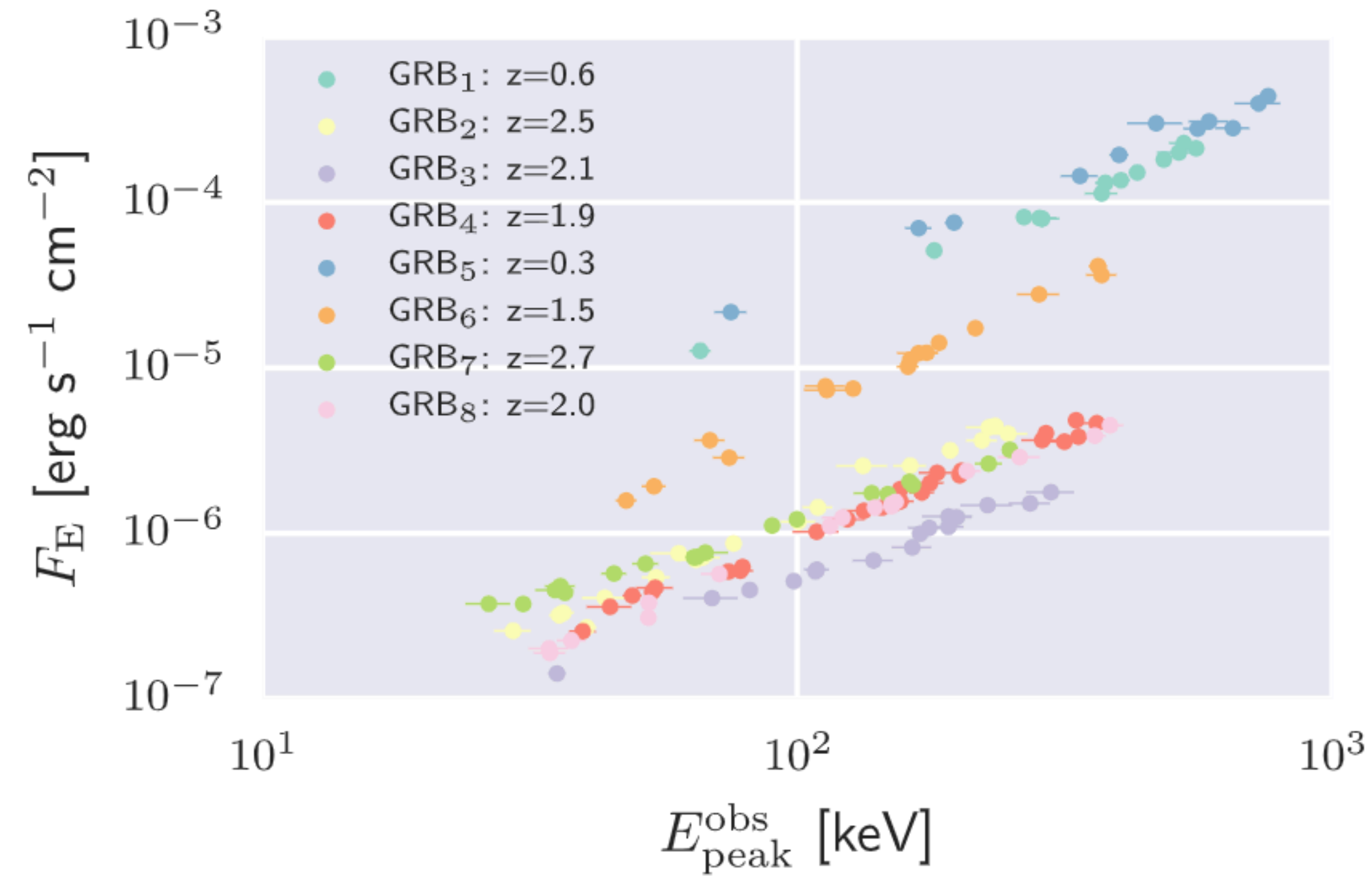


Observer Frame

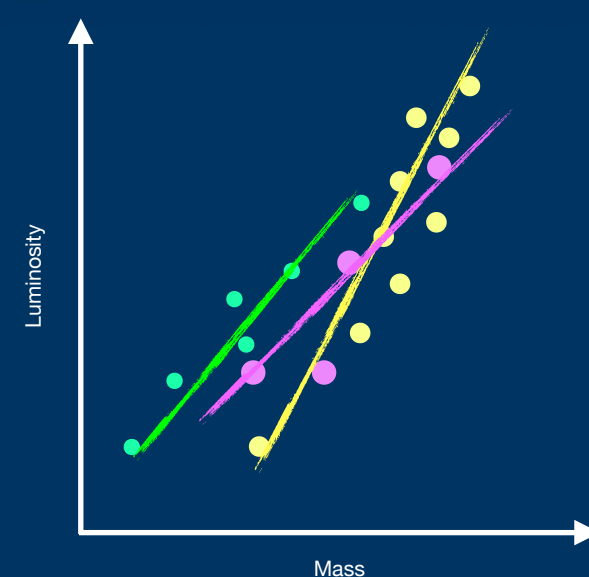


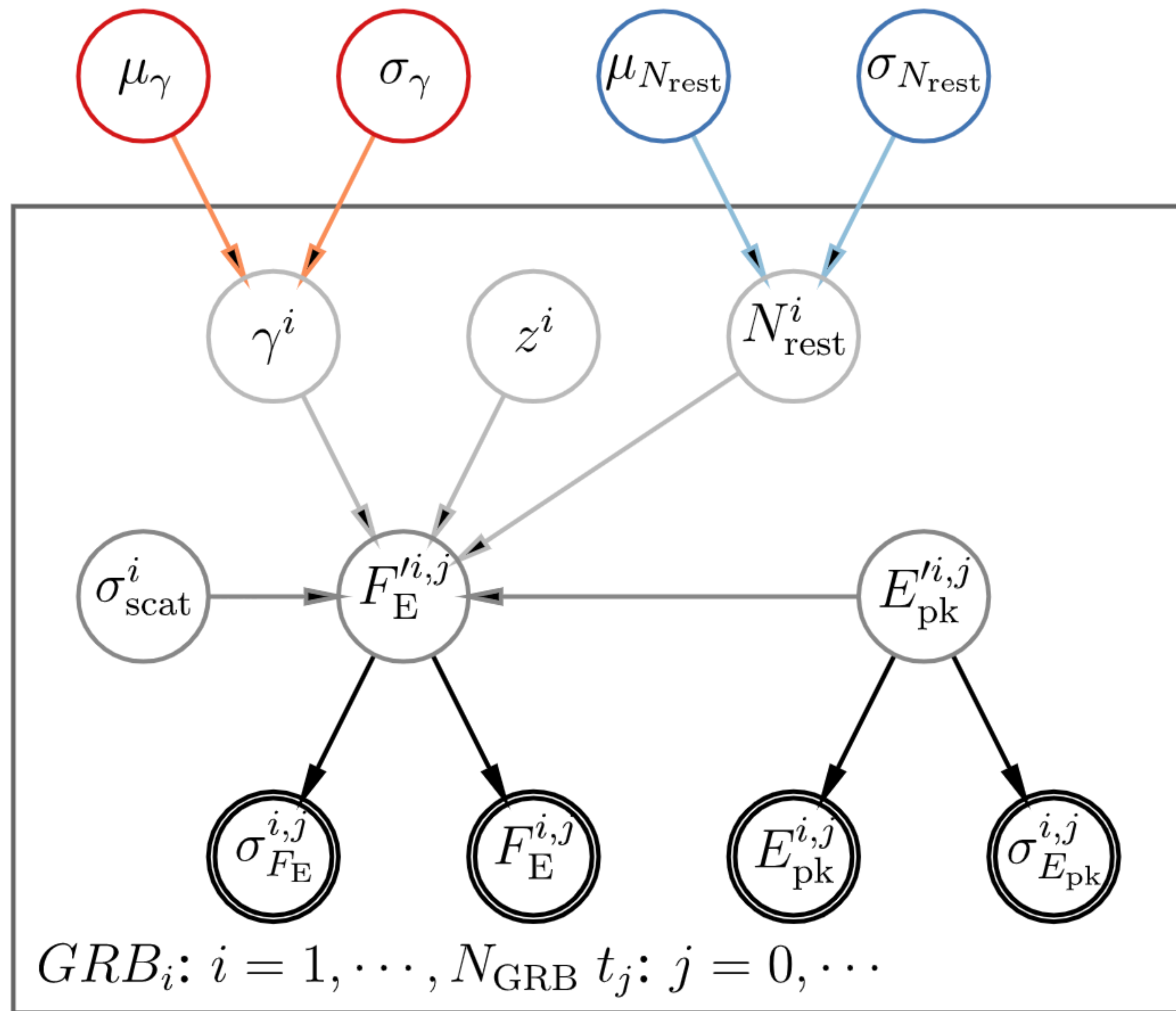


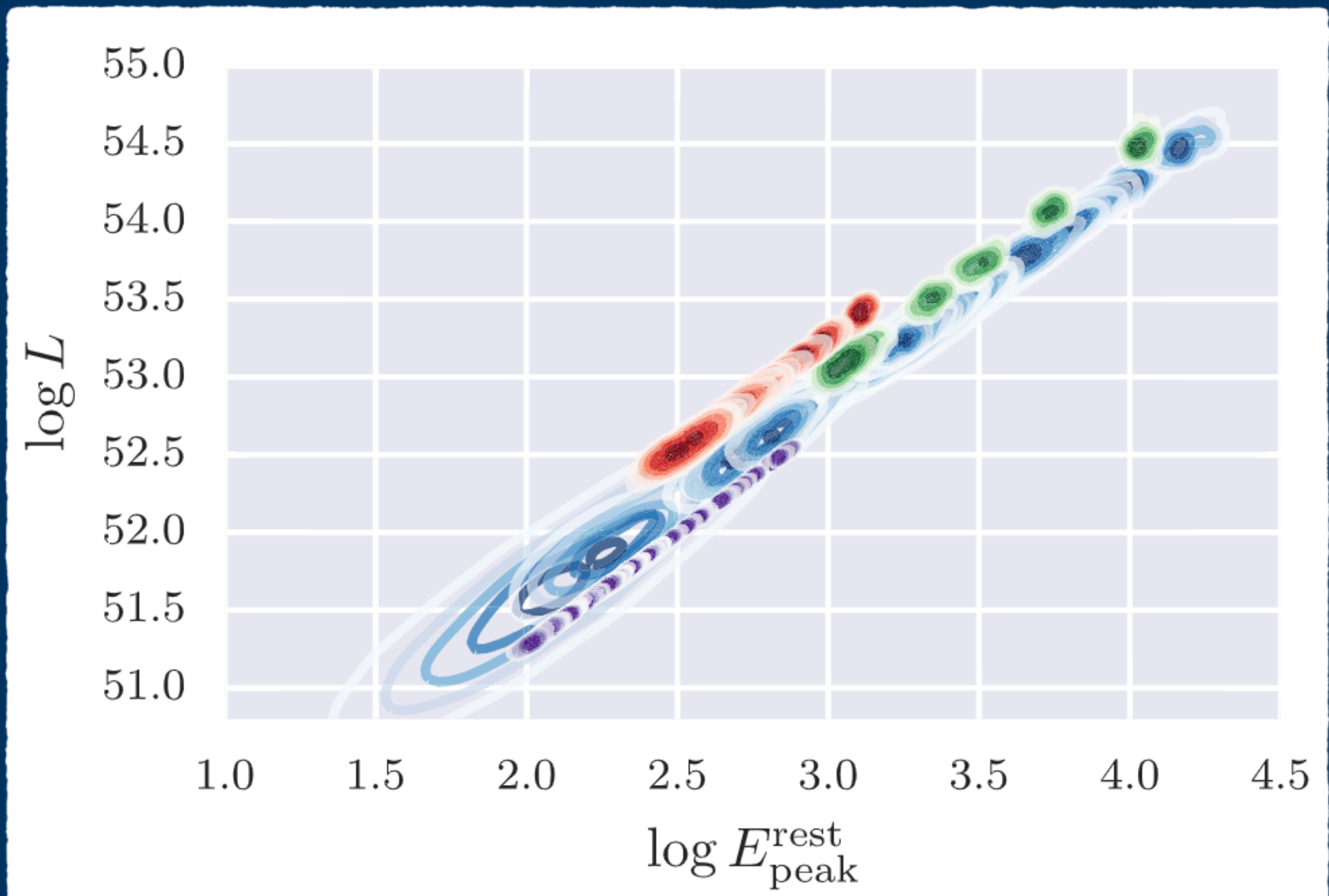
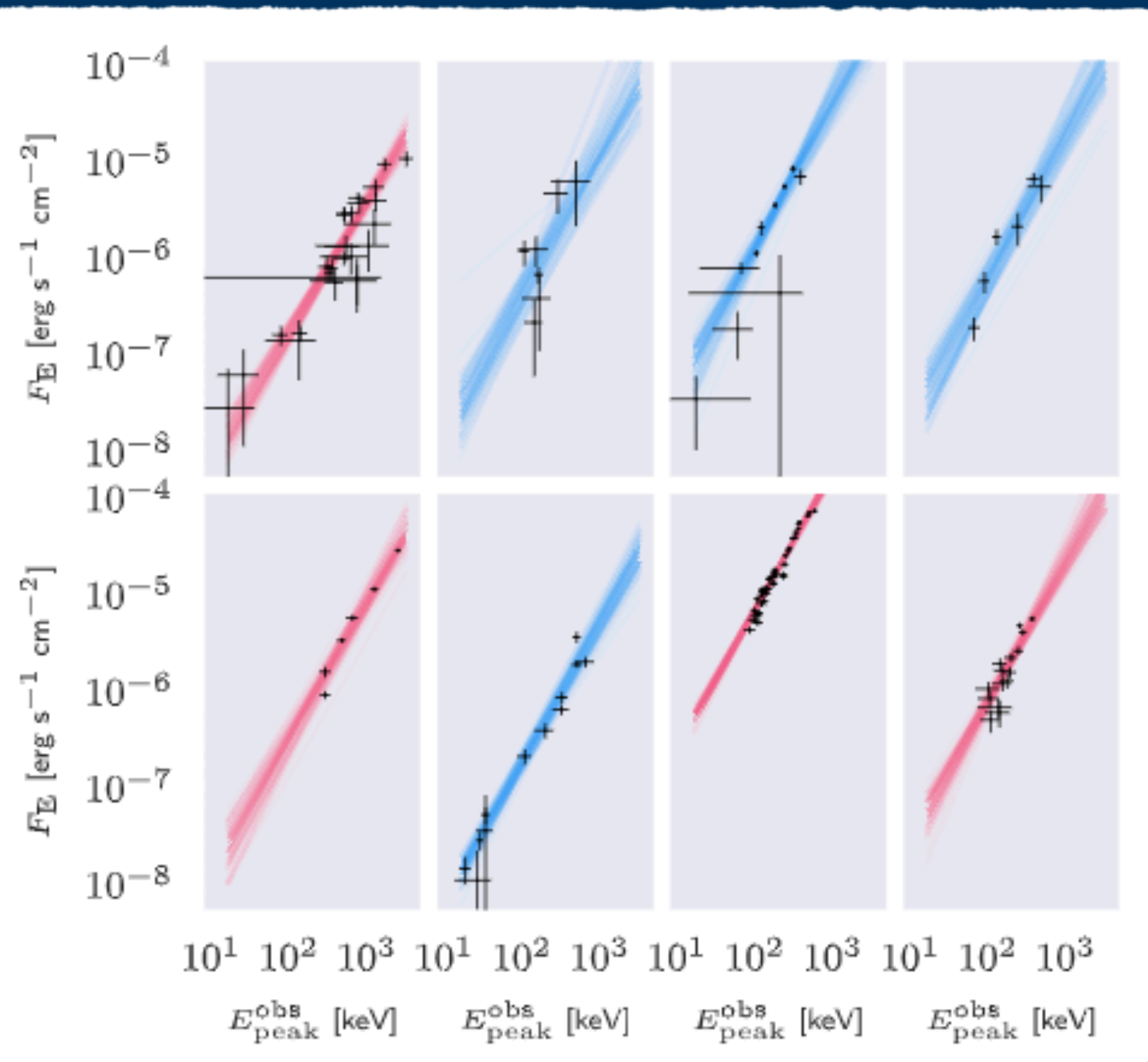
Rest Frame

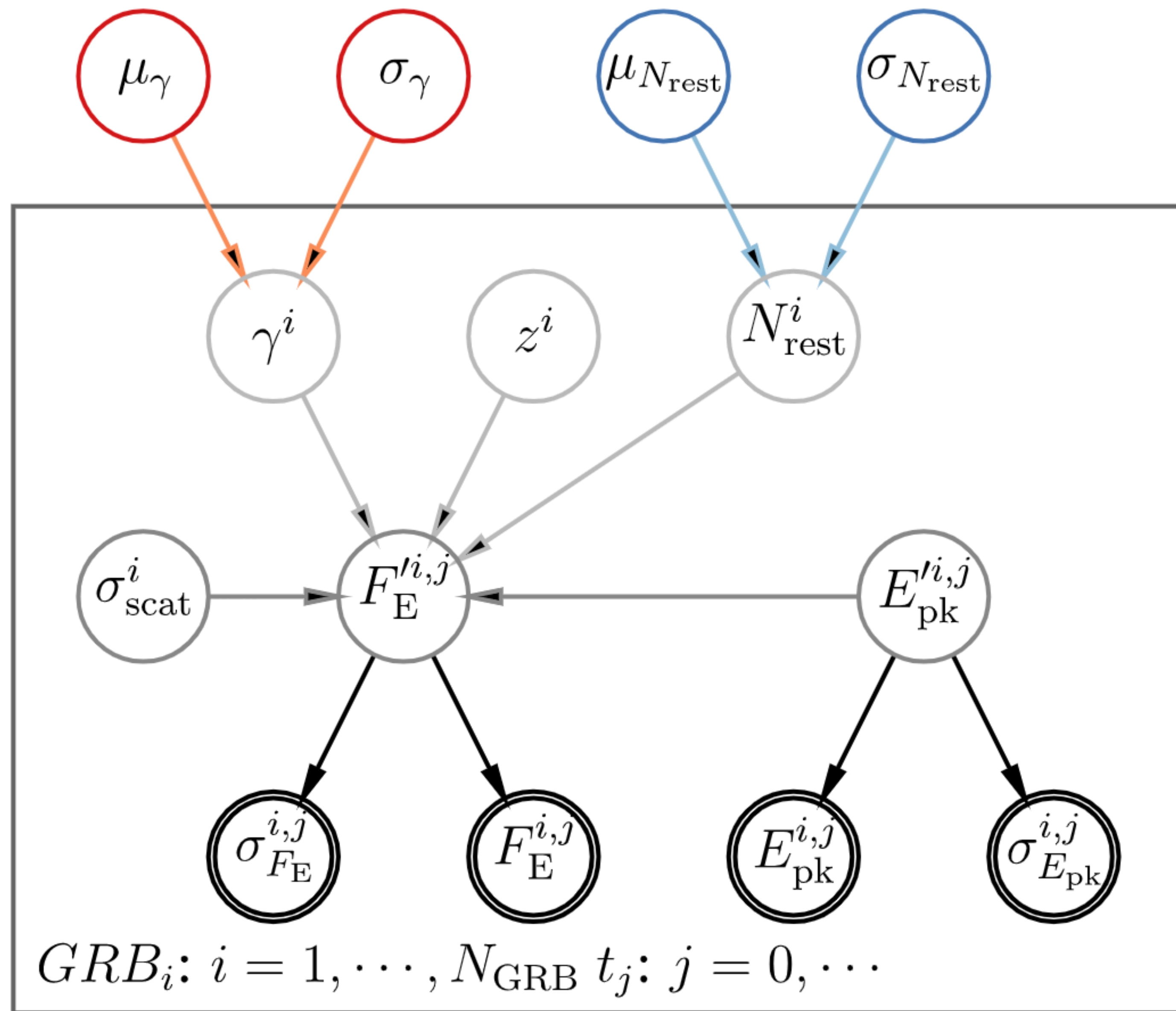


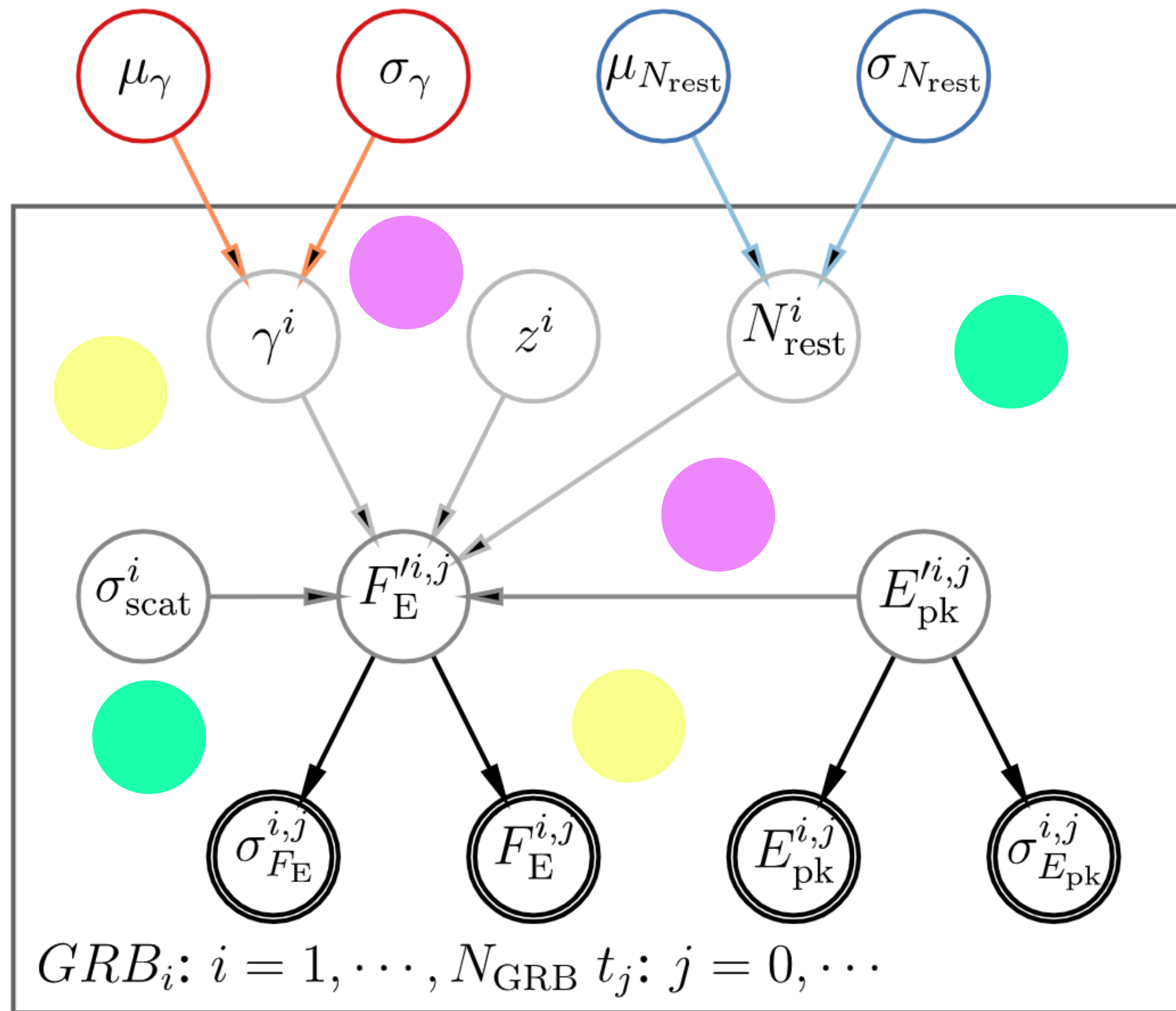
Observer Frame



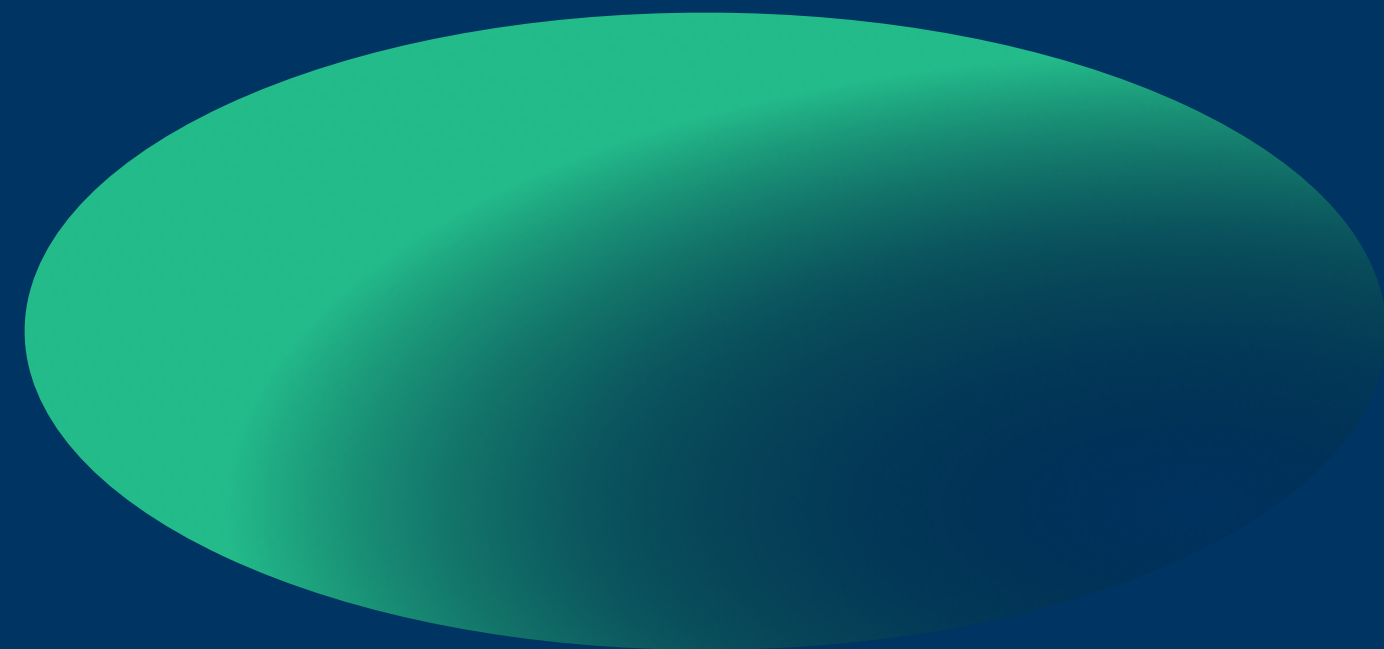




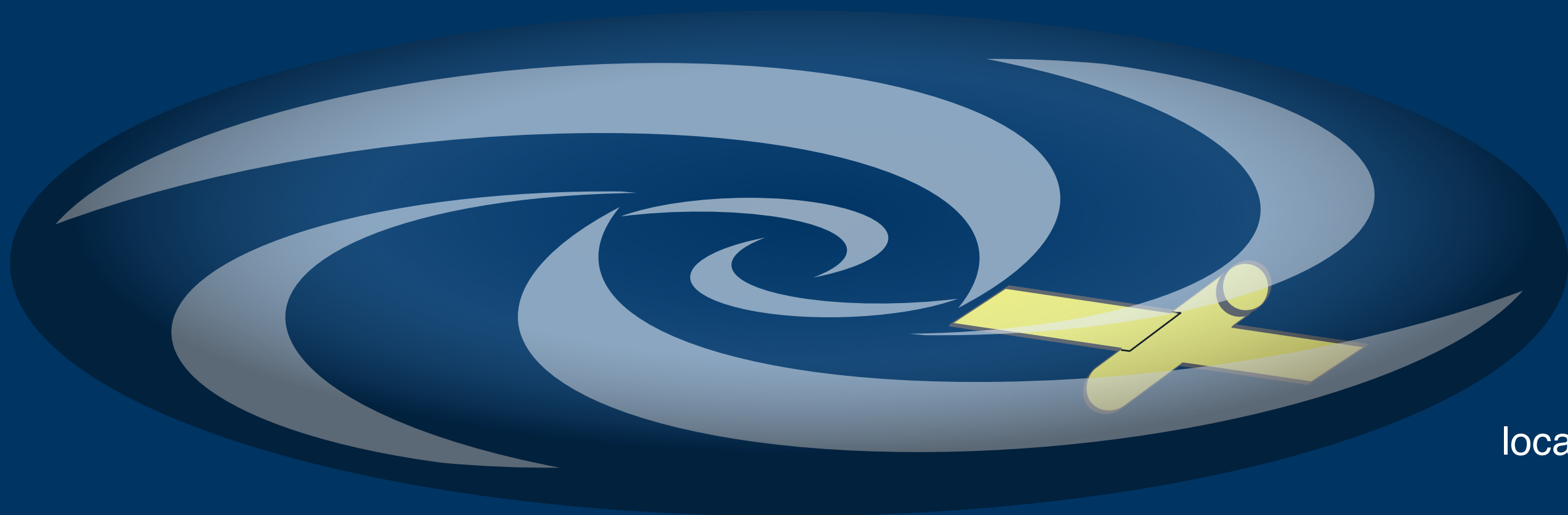




$$f_i(\varepsilon; \gamma, N_{\text{host},i}, N_{\text{mw},i}) = K_i \left(\frac{\varepsilon}{\varepsilon_{\text{piv}}} \right)^{\gamma_i} \eta(N_{\text{host},i}, \varepsilon(1+z_i)) \eta(N_{\text{mw},i}, \varepsilon)$$



host galaxy gas



local mw gas

$$f_i(\varepsilon; \gamma, N_{\text{host},i}, N_{\text{mw},i}) = K_i \left(\frac{\varepsilon}{\varepsilon_{\text{piv}}} \right)^{\gamma_i} \eta(N_{\text{host},i}, \varepsilon(1+z_i)) \eta(N_{\text{mw},i}, \varepsilon)$$

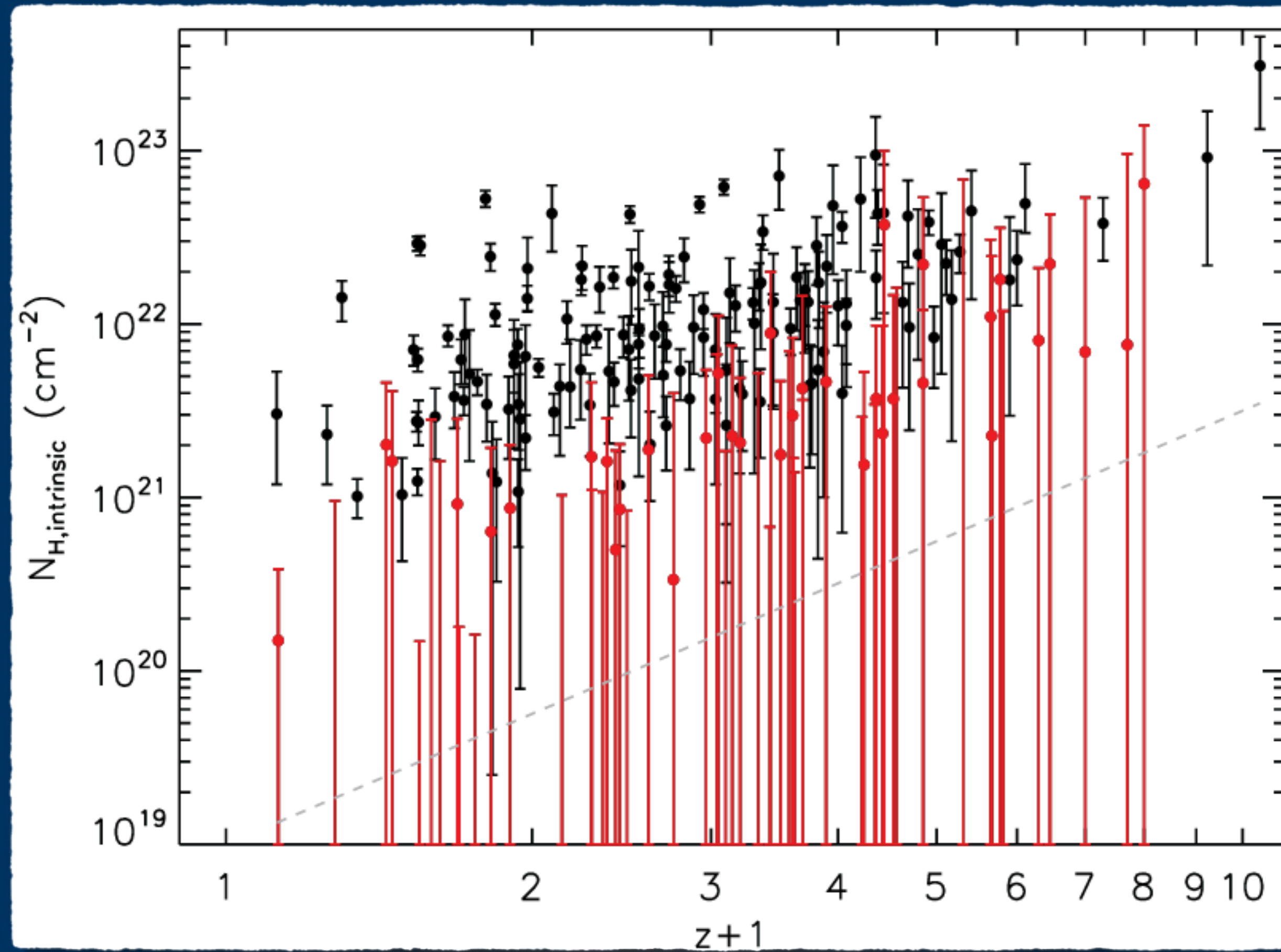


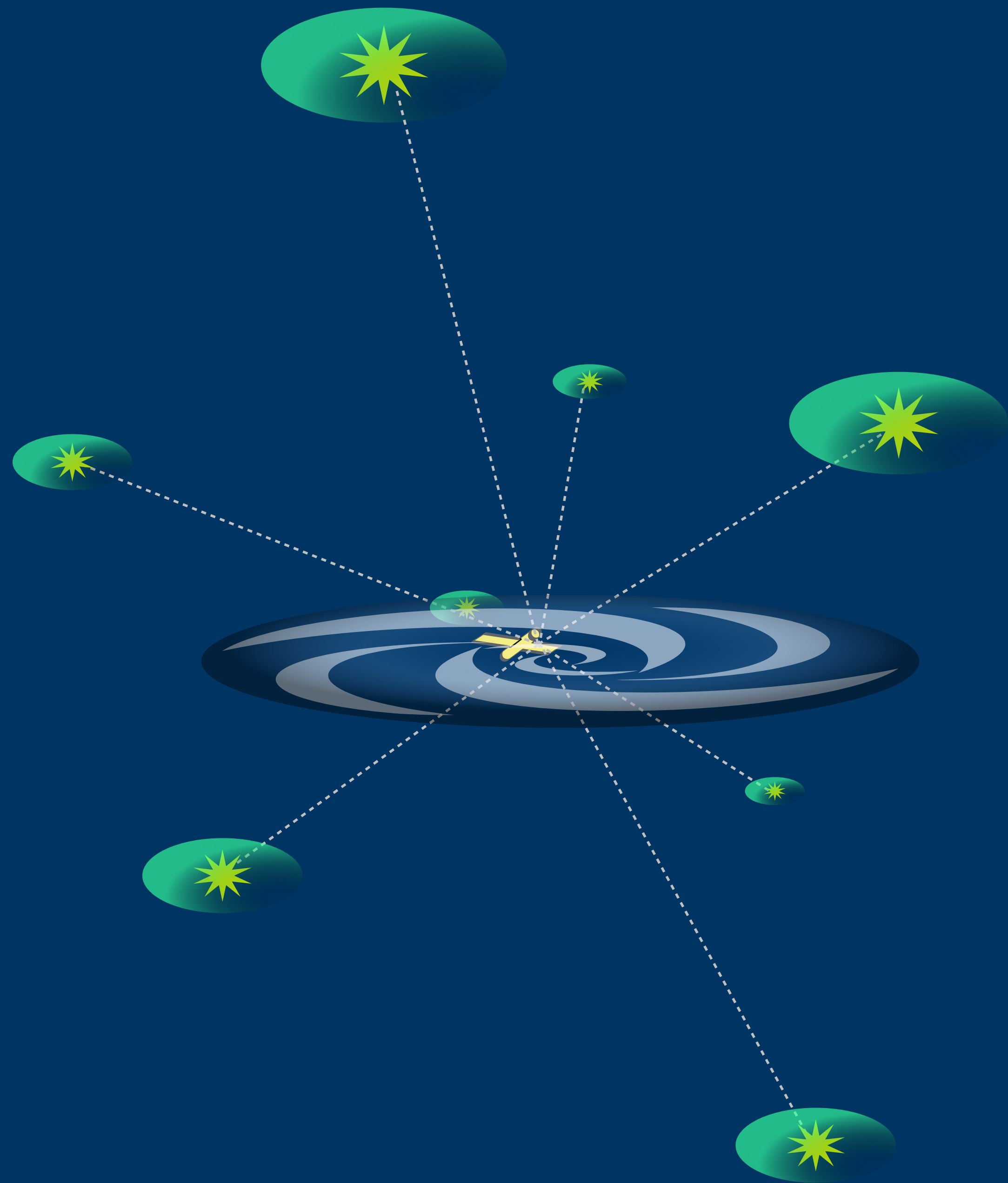
host galaxy gas



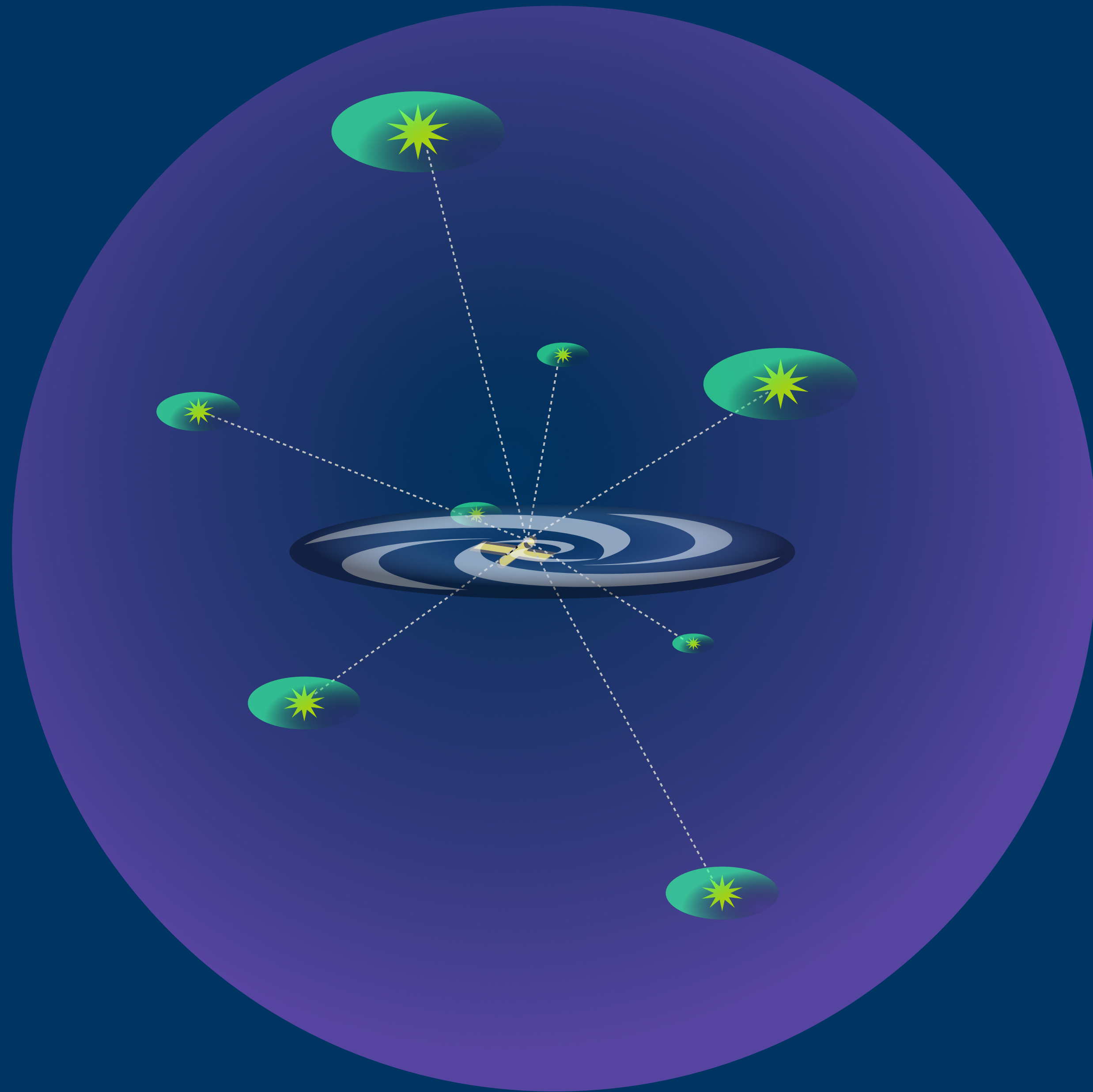
local mw gas

$$f_i(\varepsilon; \gamma, N_{\text{host},i}, N_{\text{mw},i}) = K_i \left(\frac{\varepsilon}{\varepsilon_{\text{piv}}} \right)^{\gamma_i} \eta(N_{\text{host},i}, \varepsilon(1+z_i)) \eta(N_{\text{mw},i}, \varepsilon)$$

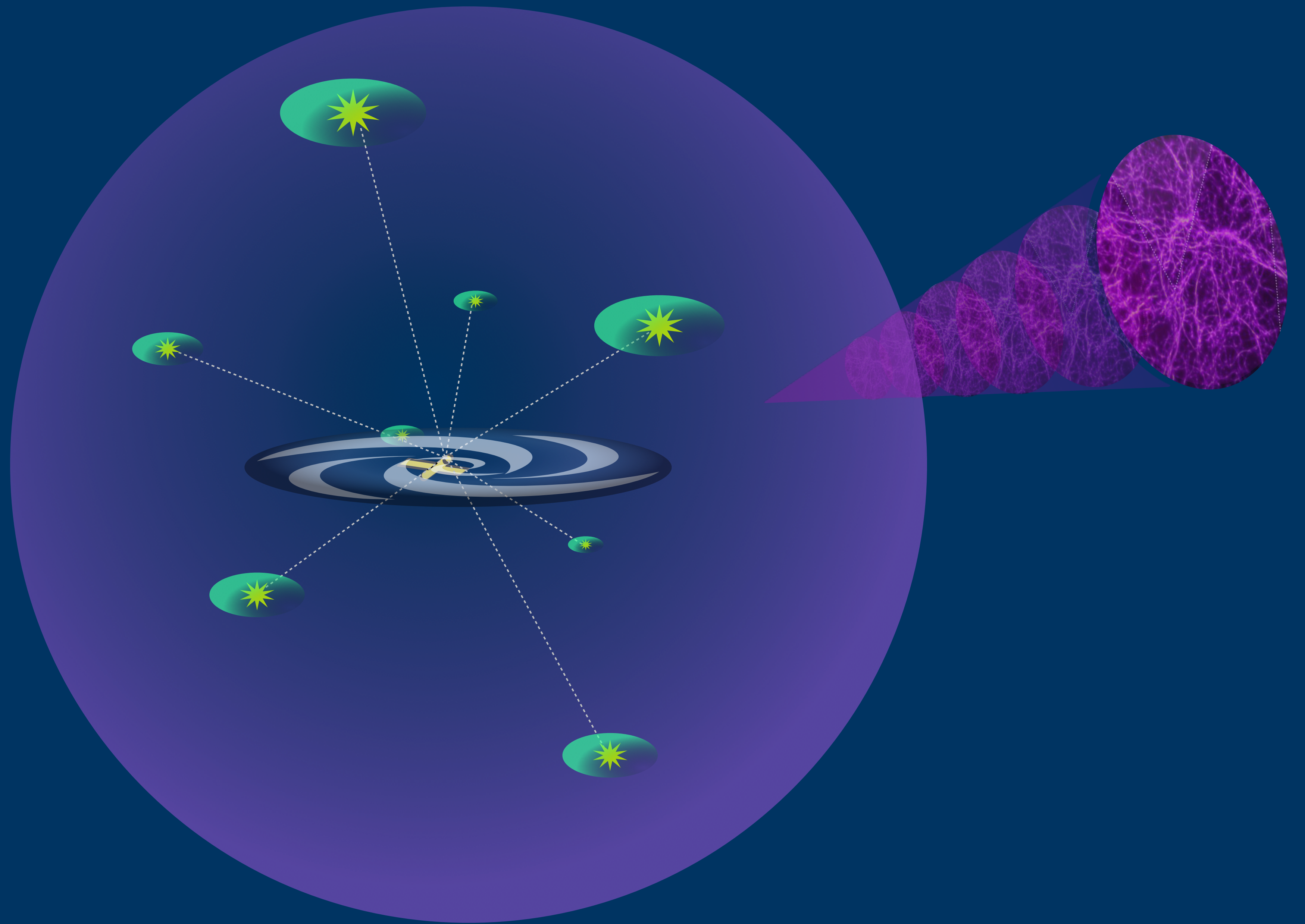




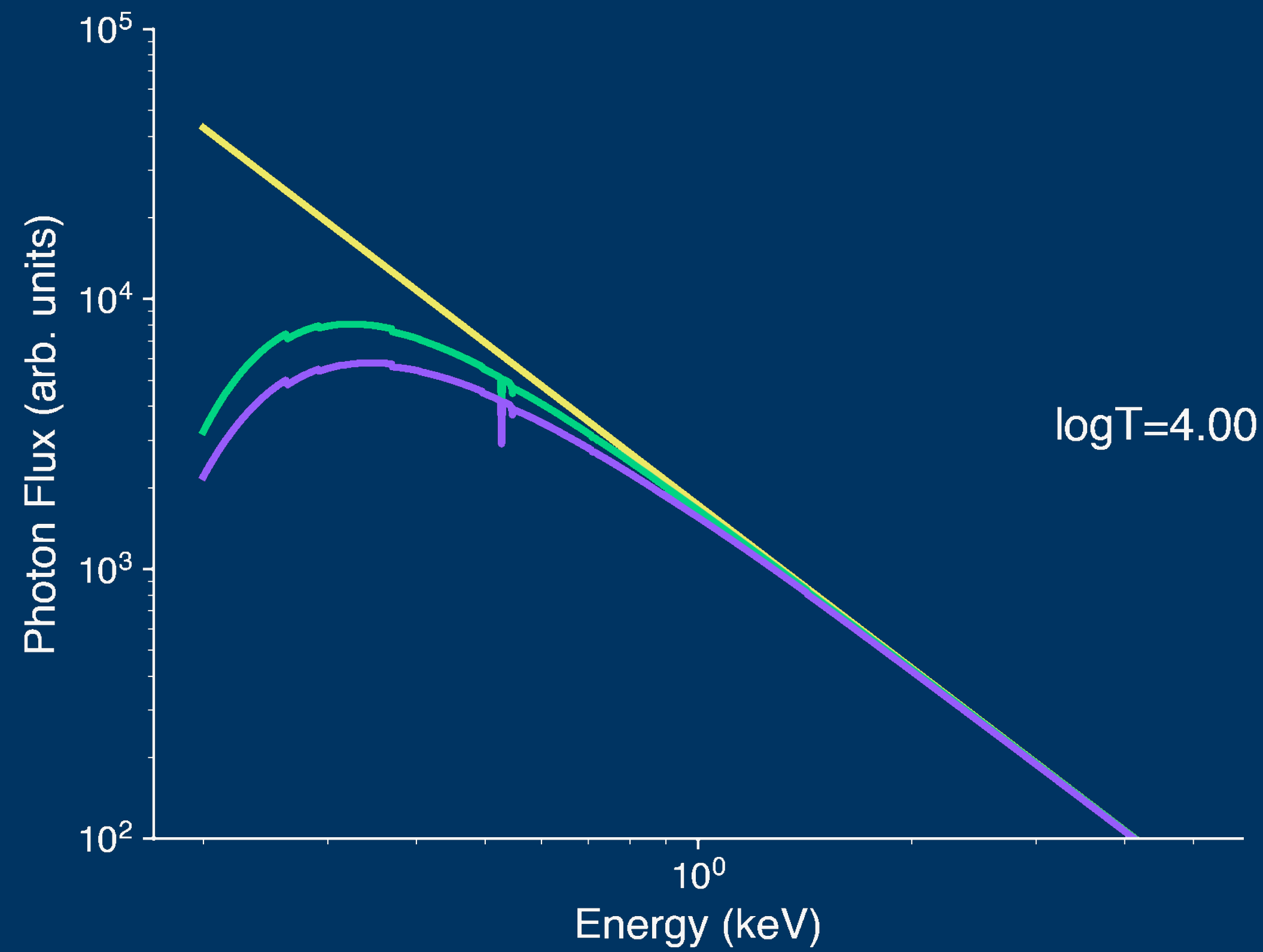
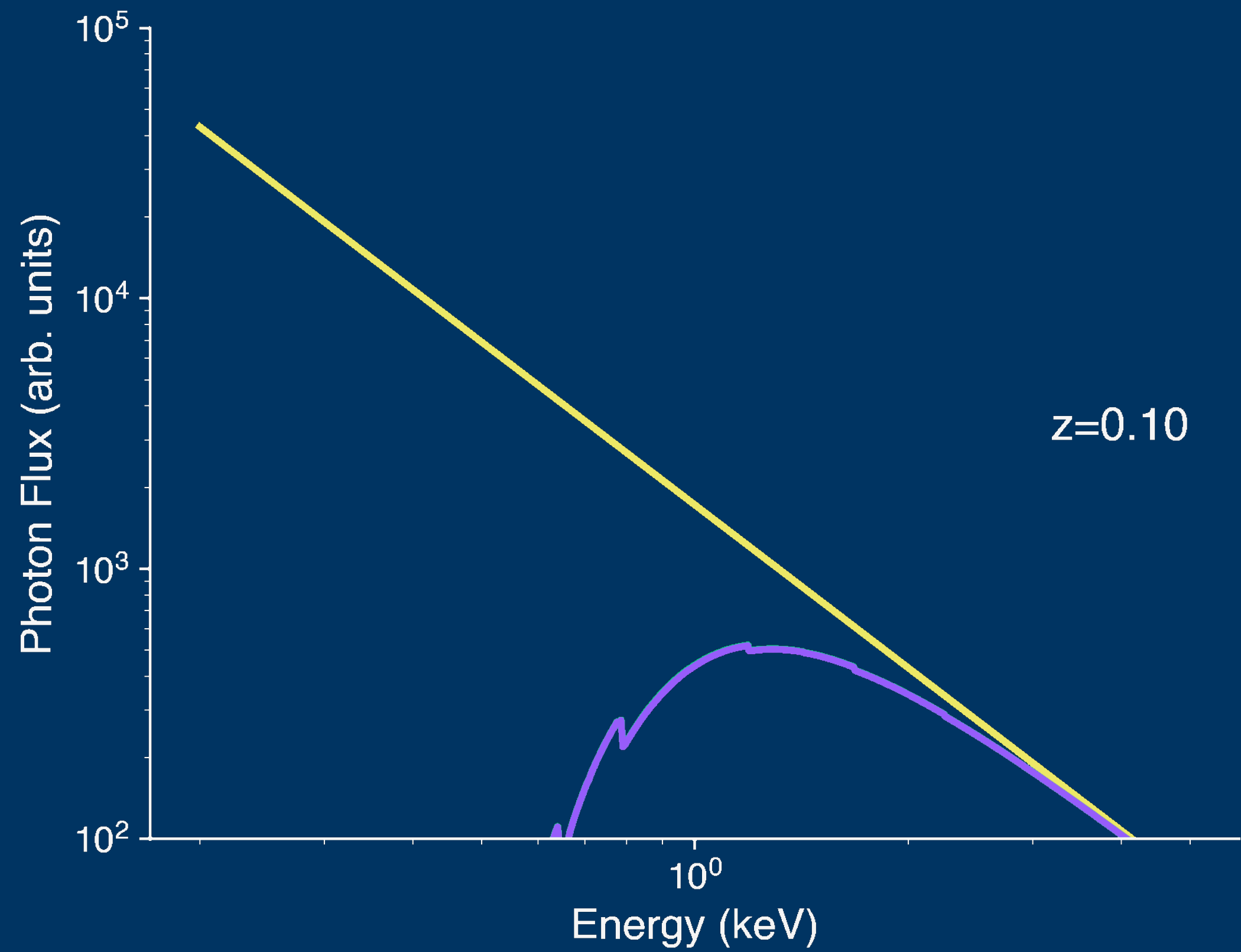
WHIM

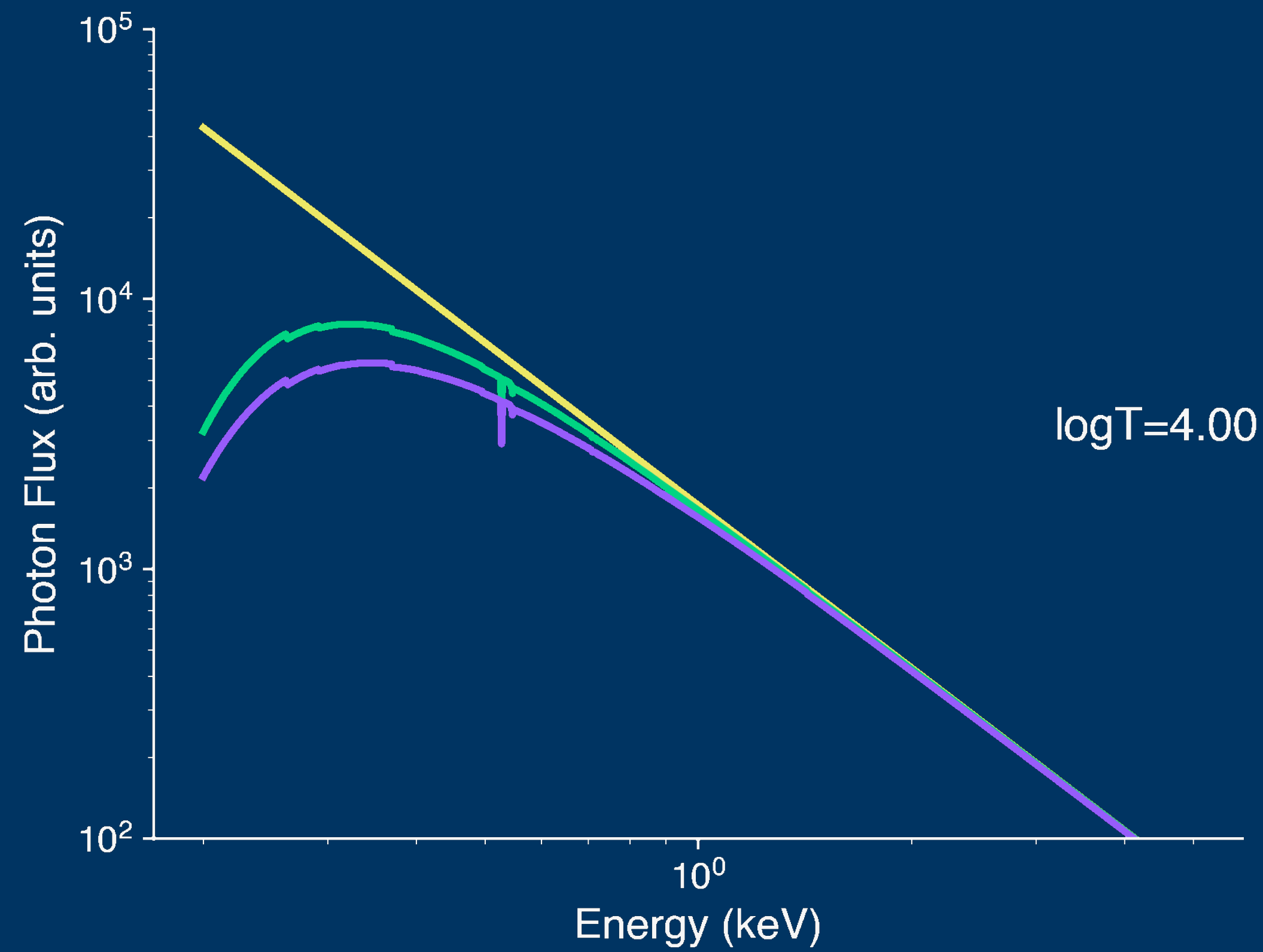
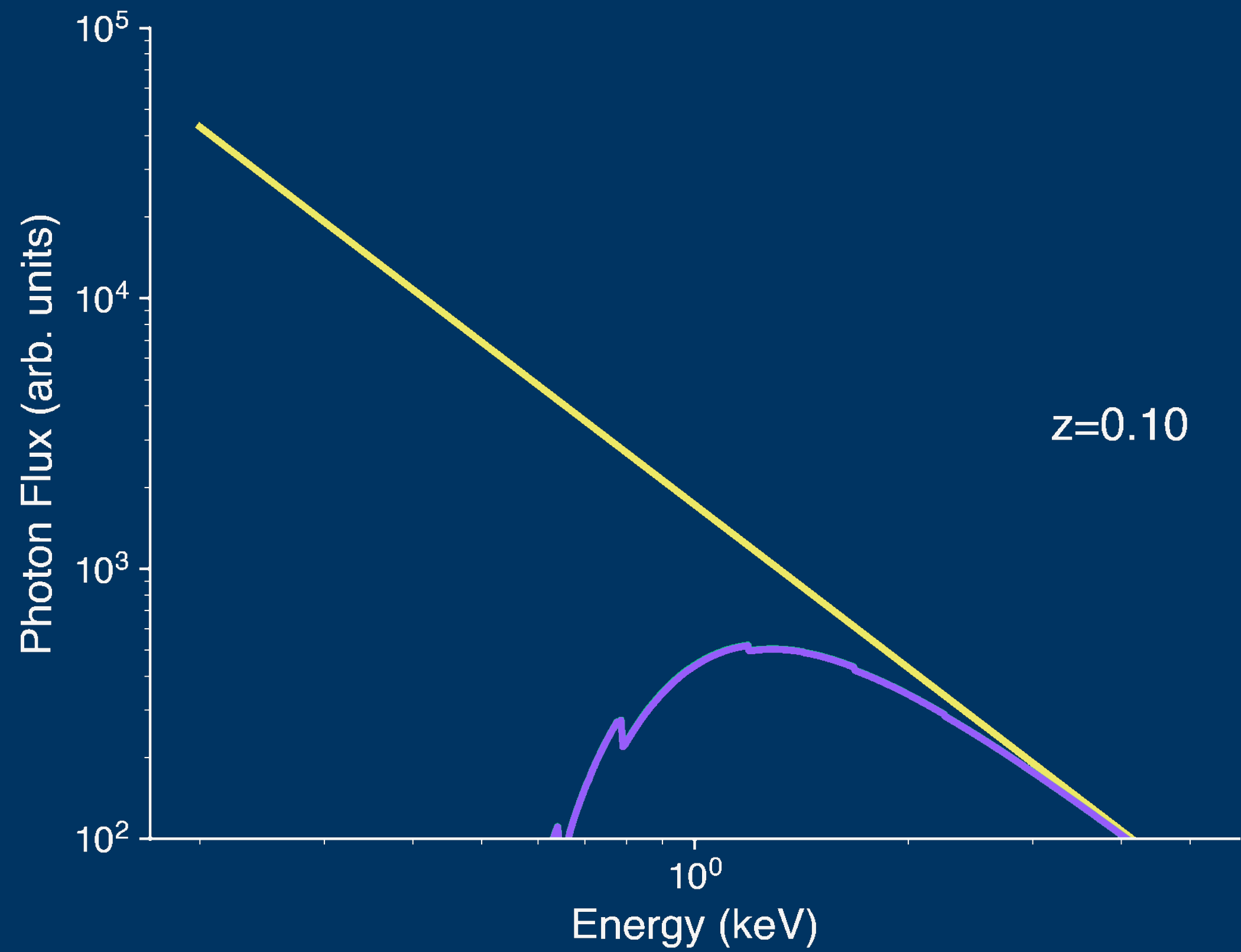


WHIM

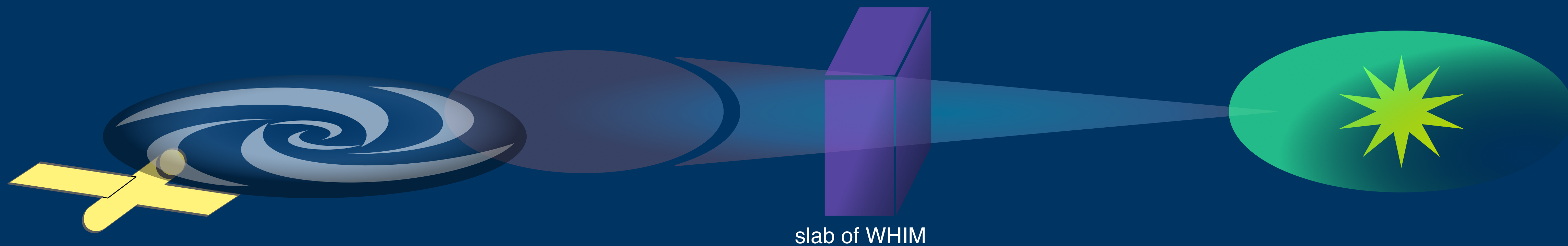


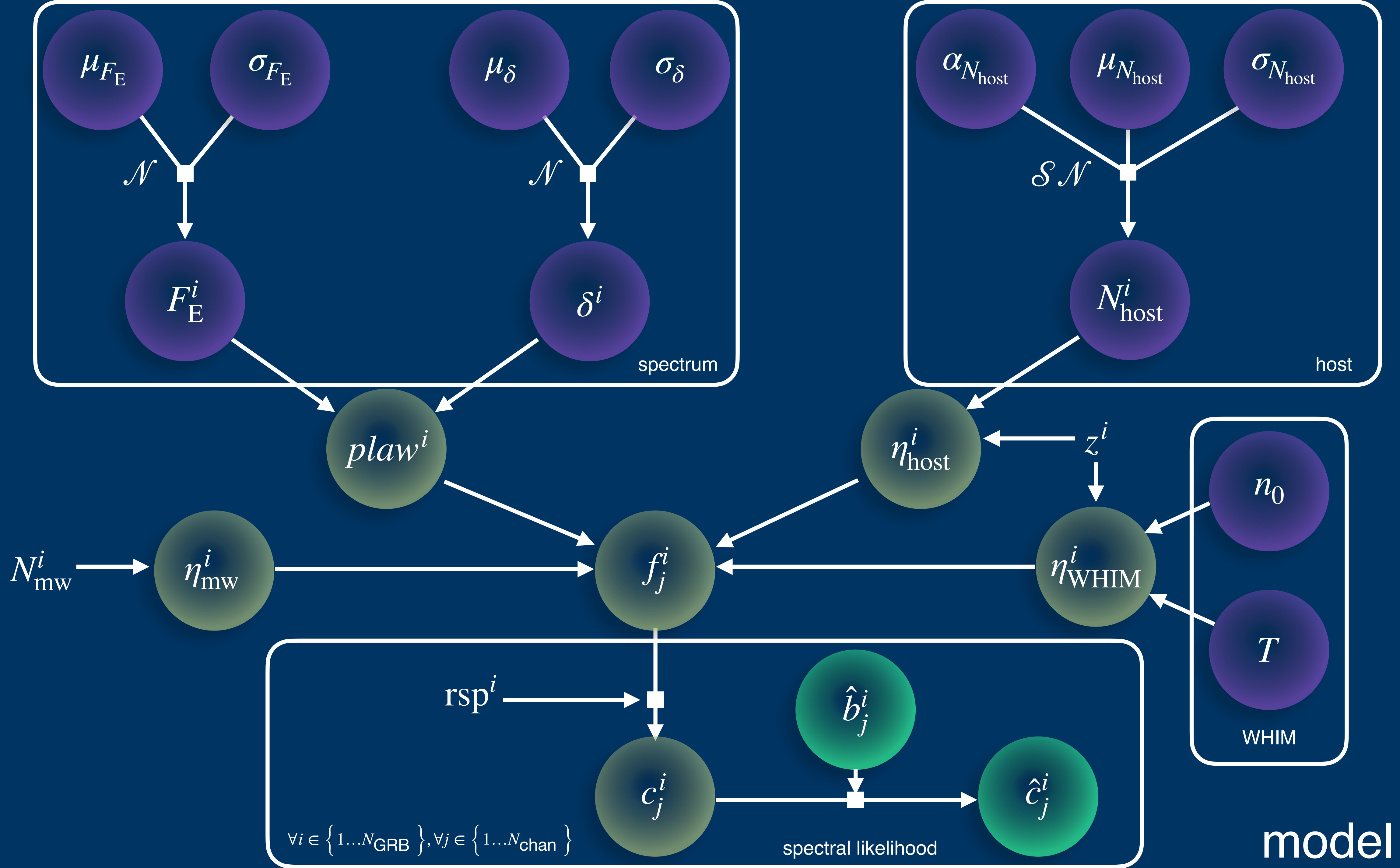
WHIM

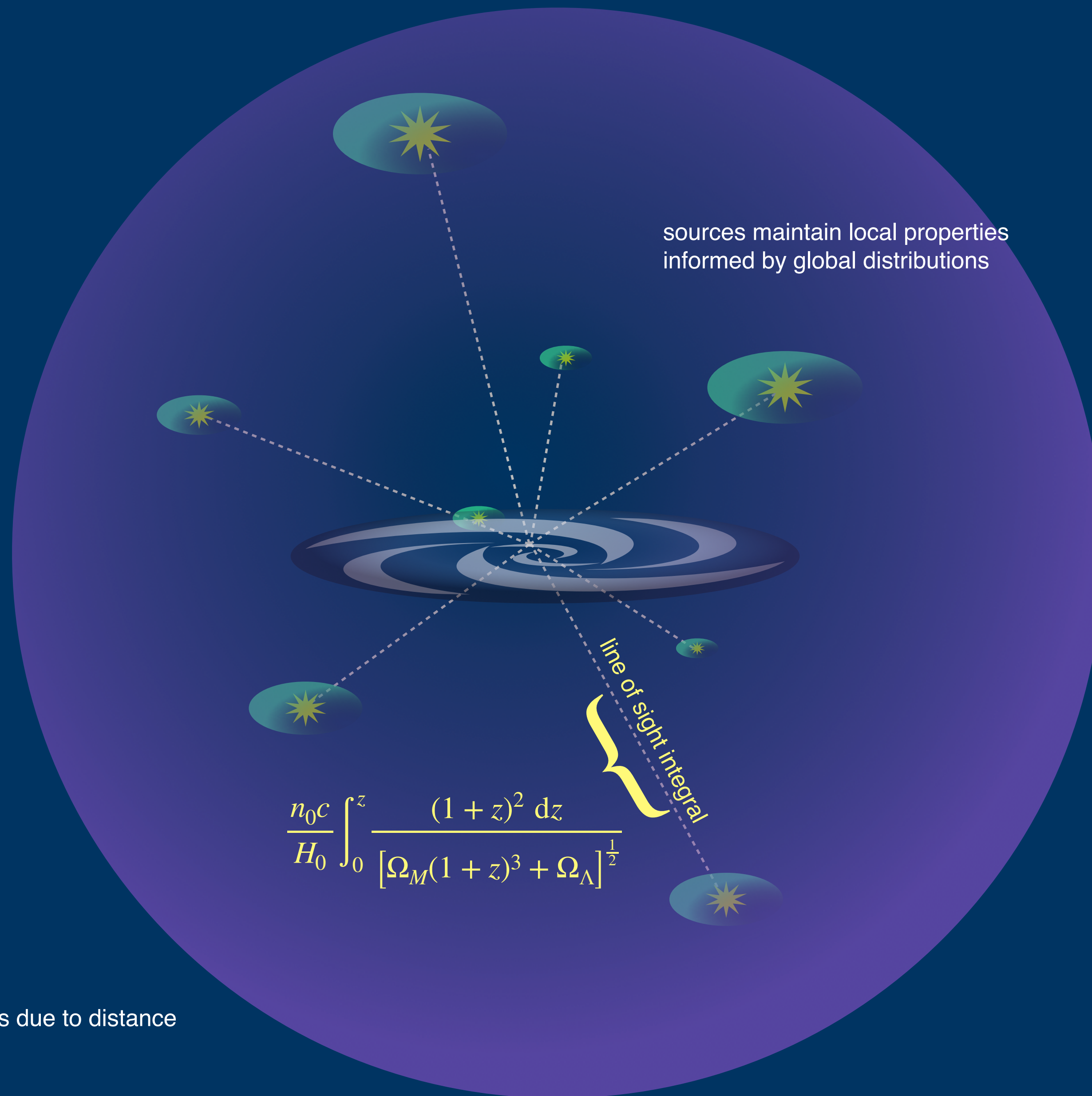




$$f_i(\varepsilon; \gamma, N_{\text{host},i}, N_{\text{mw},i}) = K_i \left(\frac{\varepsilon}{\varepsilon_{\text{piv}}} \right)^{\gamma_i} \eta(N_{\text{host},i}, \varepsilon(1+z_i)) \eta(N_{\text{mw},i}, \varepsilon) \eta(N_{0,i}, z, \varepsilon)$$





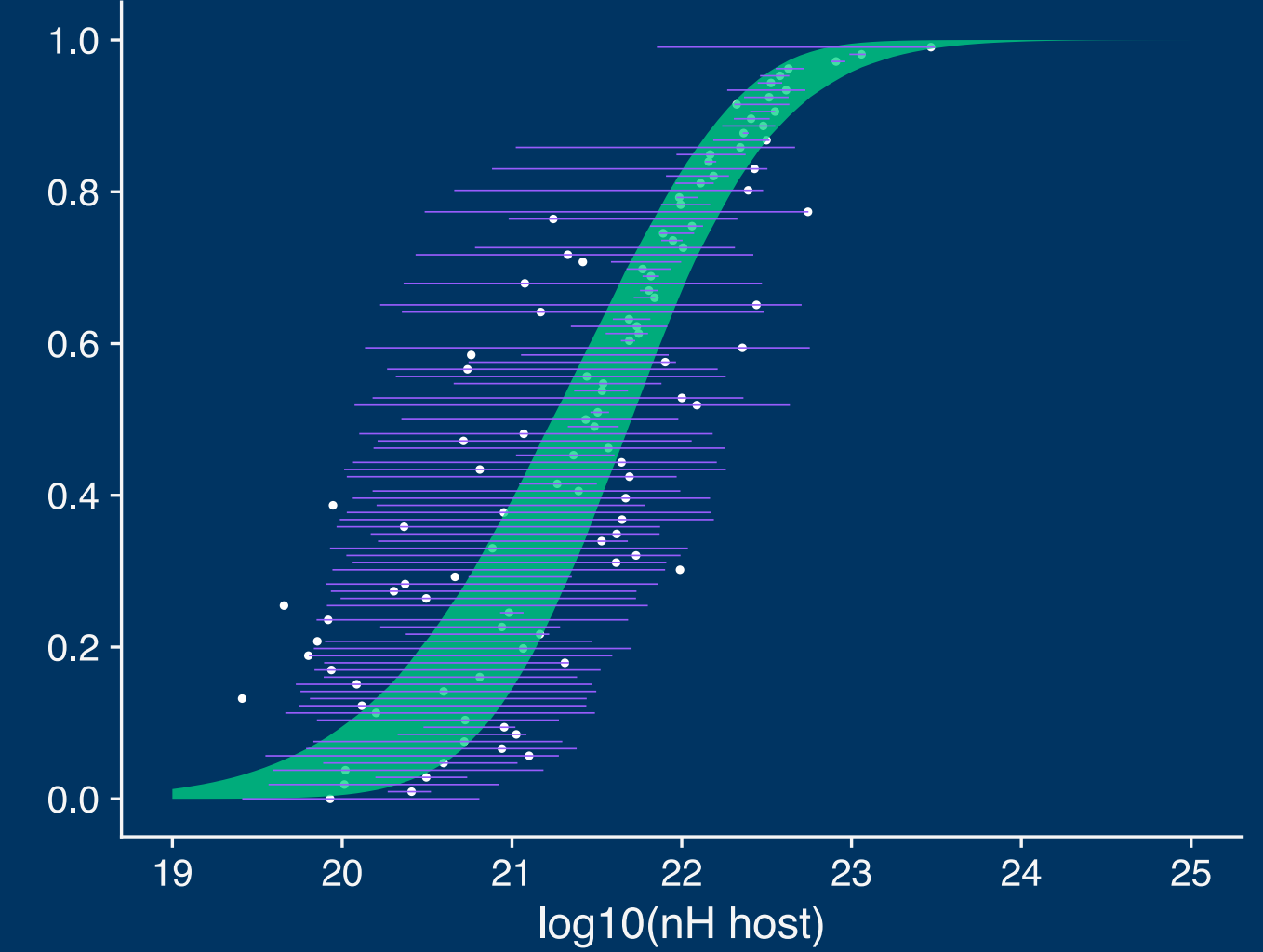
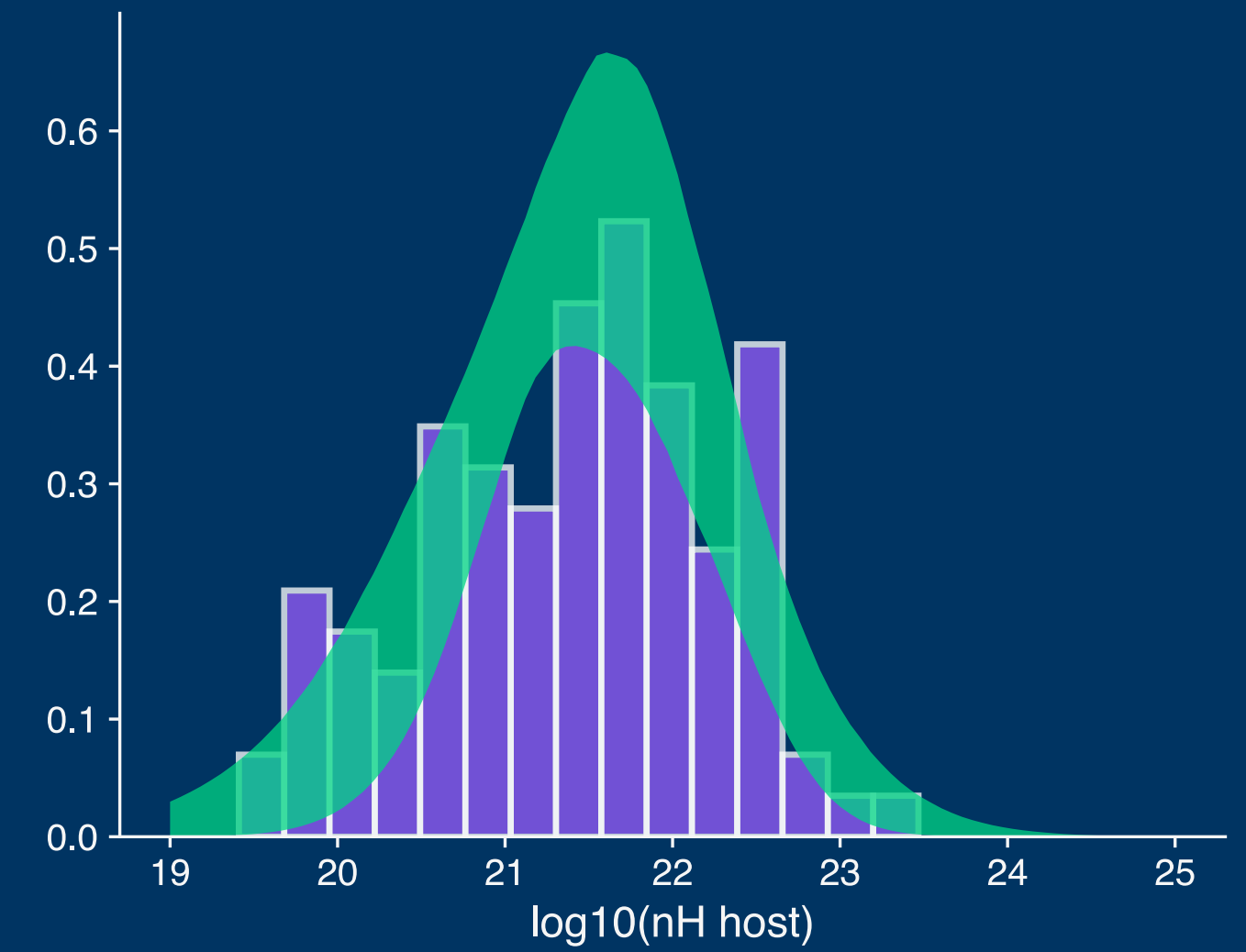
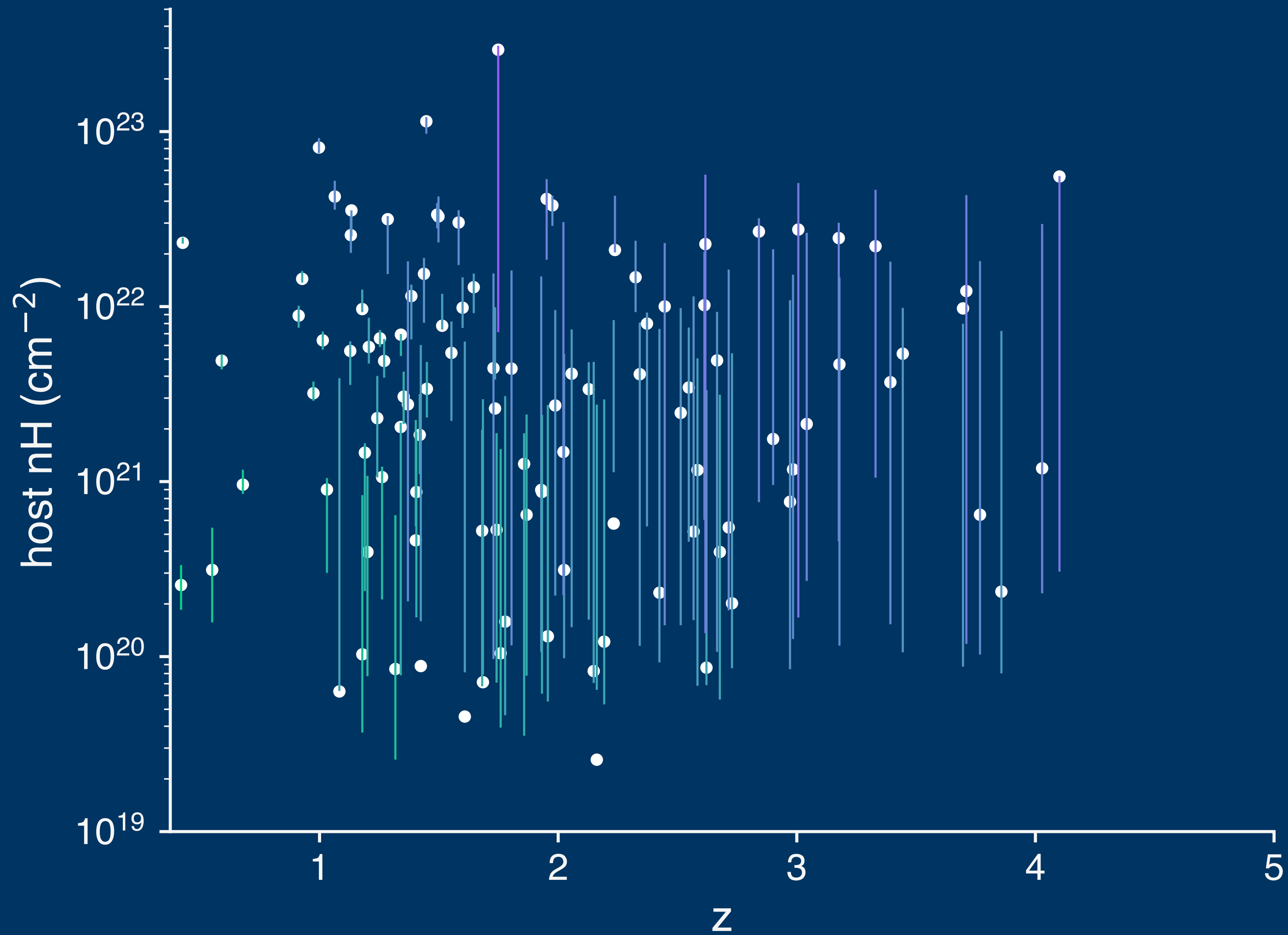


sources maintain local properties
informed by global distributions

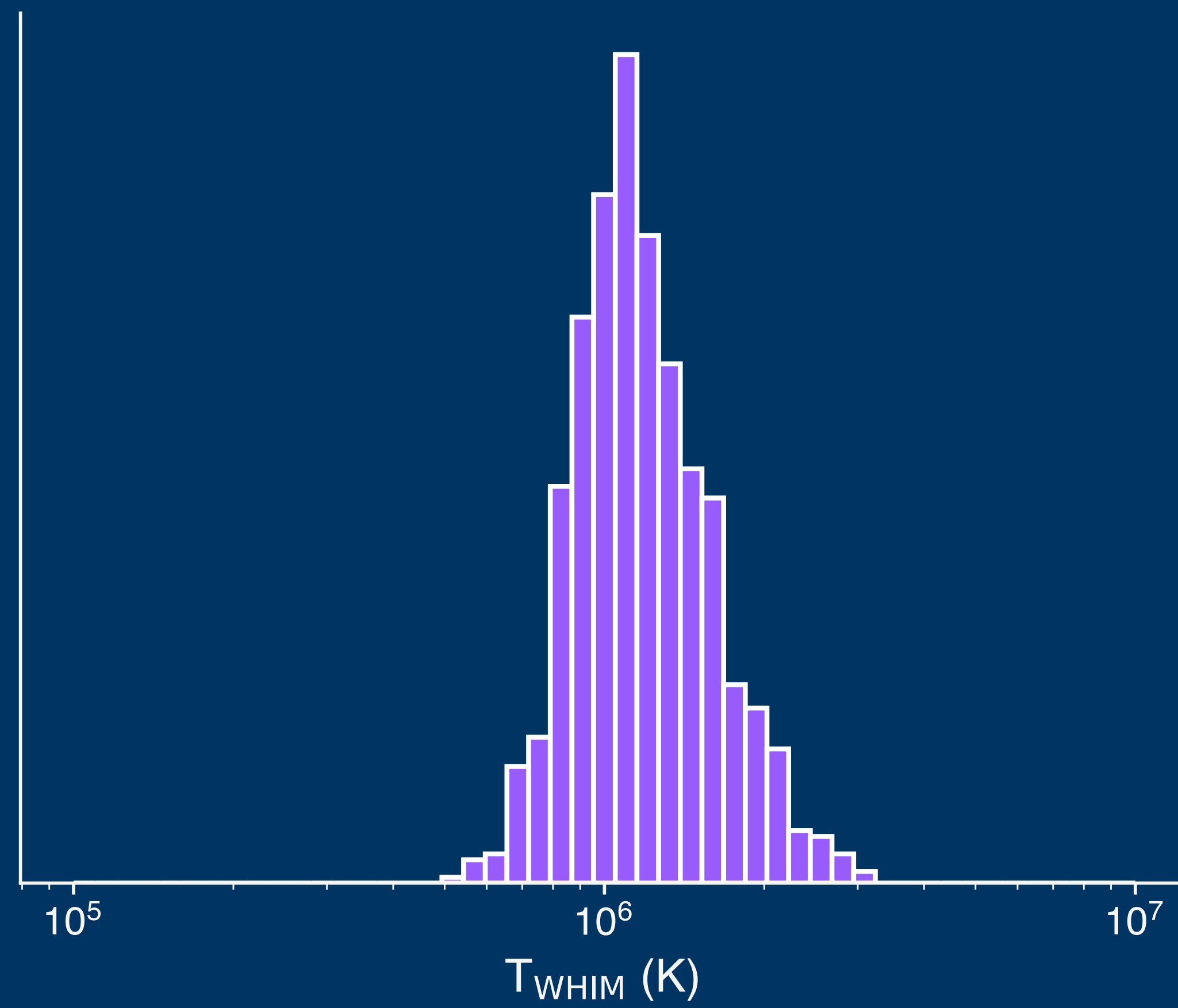
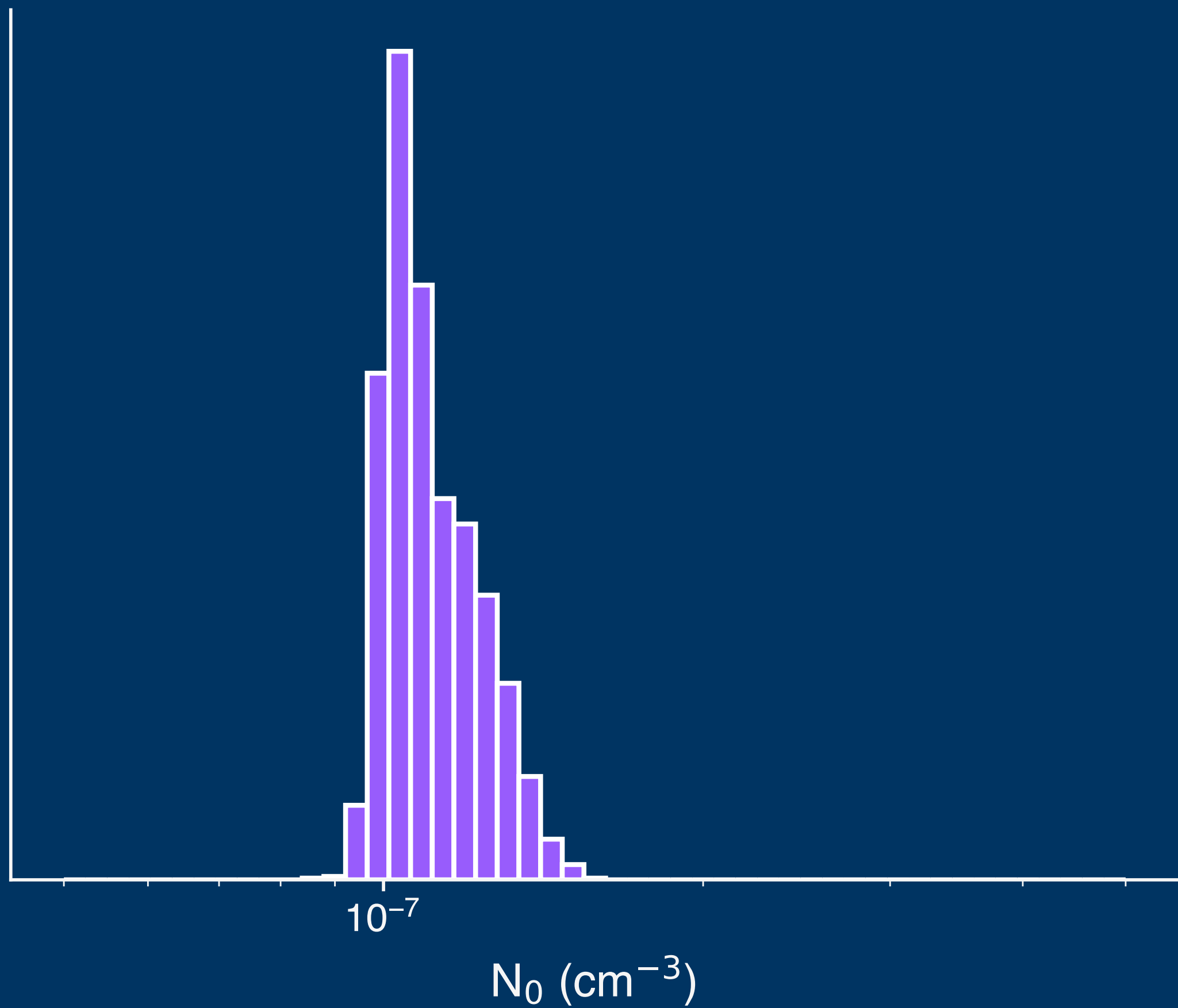
$$\frac{n_0 c}{H_0} \int_0^z \frac{(1+z)^2 dz}{[\Omega_M(1+z)^3 + \Omega_\Lambda]^{\frac{1}{2}}}$$

line of sight integral

absorption varies due to distance
from observer



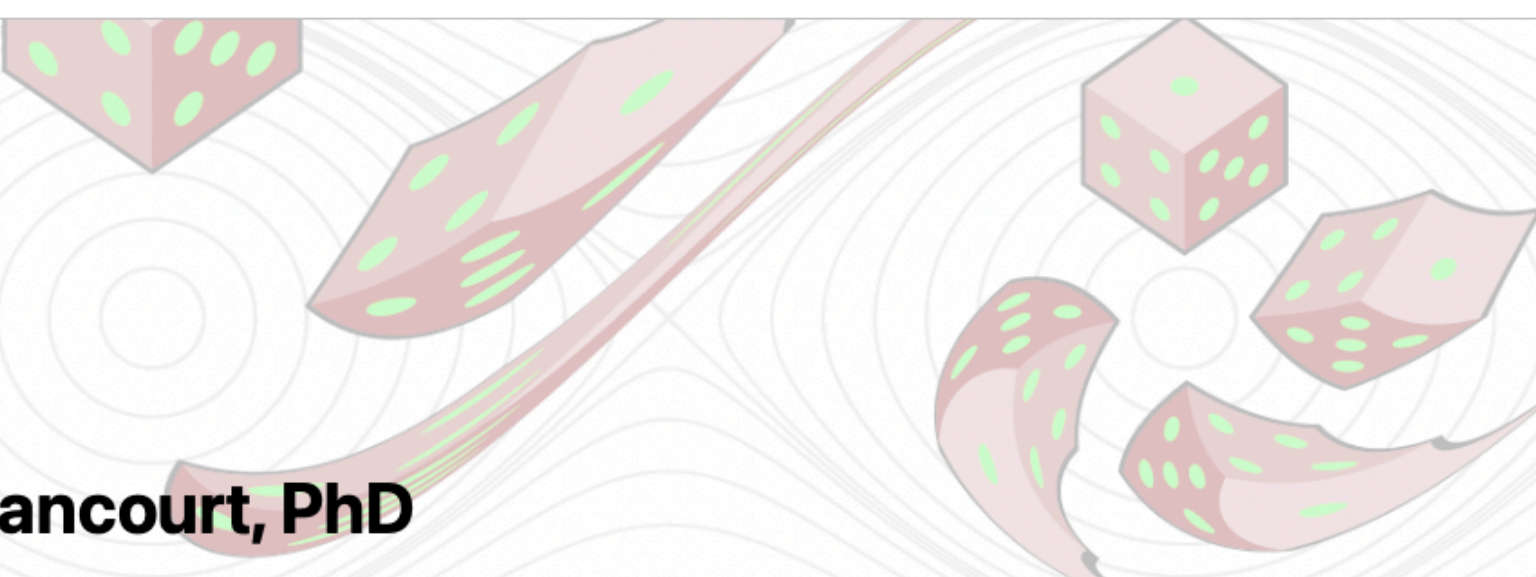
fitting the model with WHIM



fitting the model with WHIM

Resources

betanalpha.github.io Consulting Courses Speaking Writing



Michael Betancourt, PhD
Applied Statistician

https://betanalpha.github.io/assets/case_studies/hierarchical_modeling.html

A Conceptual Introduction to Hamiltonian Monte Carlo

Michael Betancourt

Hamiltonian Monte Carlo has proven a remarkable empirical success, but only recently have we begun to develop a rigorous understanding of why it performs so well on difficult problems and how it is best applied in practice. Unfortunately, that understanding is confined within the mathematics of differential geometry which has limited its dissemination, especially to the applied communities for which it is particularly important. In this review I provide a comprehensive conceptual account of these theoretical foundations, focusing on developing a principled intuition behind the method and its optimal implementations rather than any exhaustive rigor. Whether a practitioner or a statistician, the dedicated reader will acquire a solid grasp of how Hamiltonian Monte Carlo works, when it succeeds, and, perhaps most importantly, when it fails.

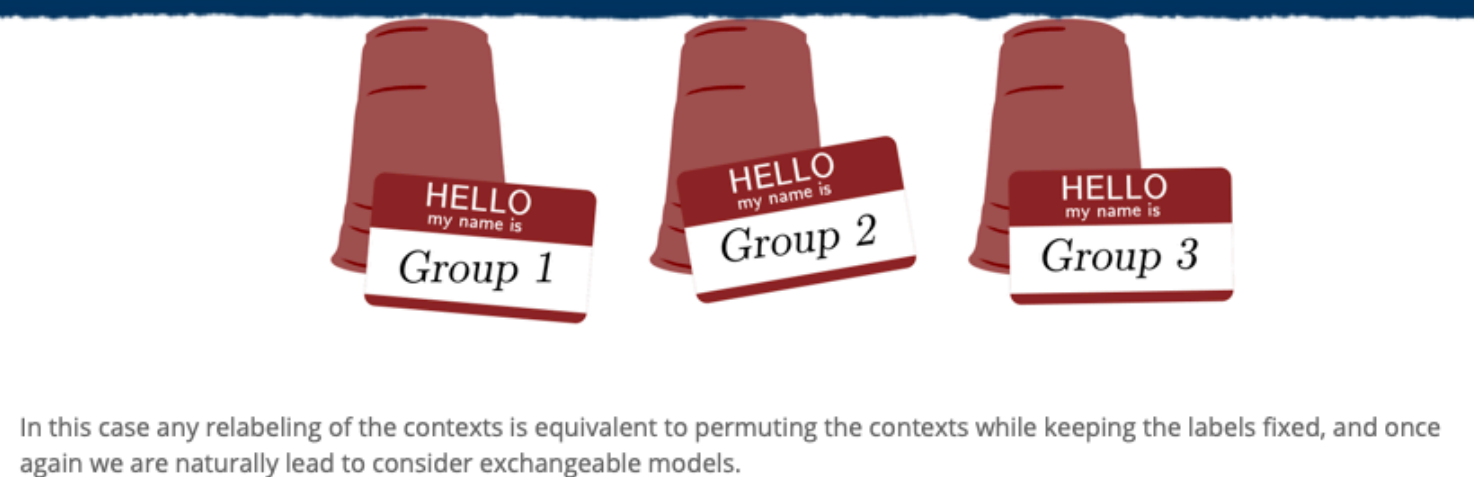
Comments: 60 pages, 42 figures

Subjects: **Methodology (stat.ME)**

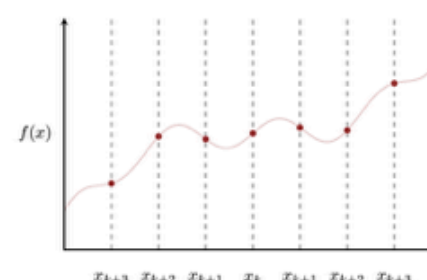
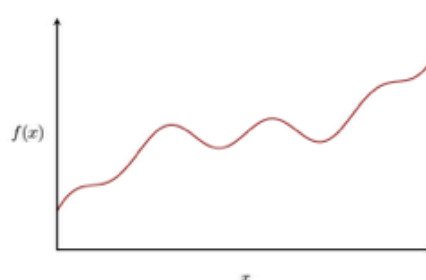
Cite as: [arXiv:1701.02434](https://arxiv.org/abs/1701.02434) [stat.ME]

(or [arXiv:1701.02434v2](https://arxiv.org/abs/1701.02434v2) [stat.ME] for this version)

- 1 Modeling Heterogeneity
- 2 Can You Take Me Hier(archical)?
 - 2.1 Exchangeability
 - 2.2 de Finetti's Theorem and Hierarchical Models
 - 2.3 Interpreting Hierarchical Models
 - 2.4 Normal Hierarchical Models
- 3 The Fundamental Degeneracies of Normal Hierarchical Models
- 4 Empirical Investigations
- 5 Multivariate Normal Hierarchical Models
- 6 Conclusion
- Acknowledgements
- References
- License
- Original Computing Environment



All of this said, anytime we consider exchangeability we have to recognize that ignoring discriminating information can be a very poor modeling choice. Consider for example modeling the behavior of a continuous function $f(x)$ with a set of function values at discretized inputs, $f_k = f(x_k)$.



Stan User's Guide

Overview

Part 1. Example Models

1 Regression Models

- 1.1 Linear regression
- 1.2 The QR reparameterization
- 1.3 Priors for coefficients and scales
- 1.4 Robust noise models
- 1.5 Logistic and probit regression
- 1.6 Multi-logit regression
- 1.7 Parameterizing centered vectors

Stan User's Guide

Version 2.26

Stan Development Team

Overview

About this user's guide

Resources

The image shows a screenshot of the PyMC3 website's 'Example Notebooks' section. The website header includes the PyMC3 logo, navigation links for 'Tutorials', 'Examples', 'Books + Videos', 'API', 'Developer Guide', and 'About PyMC3', a search bar, and a GitHub icon. The main heading is 'Example Notebooks', followed by a sub-heading '(Generalized) Linear and Hierarchical Linear Models'. Below this, there is a 2x6 grid of notebook thumbnails. Each thumbnail contains a small plot and a title. The titles are: (Generalized) Linear and Hierarchical Linear Models in PyMC3, GLM: Hierarchical Linear Regression, Hierarchical Binominal Model: Rat Tumor Example, GLM: Linear regression, GLM: Logistic Regression, GLM: Model Selection, GLM: Negative Binomial Regression, GLM: Poisson Regression, GLM: Robust Linear Regression, GLM: Robust Regression using Custom Likelihood for Outlier Classification, Rolling Regression, and Hierarchical Partial Pooling.

PyMC3 Tutorials Examples Books + Videos API Developer Guide About PyMC3 Search...

Example Notebooks

(Generalized) Linear and Hierarchical Linear Models

(Generalized) Linear and Hierarchical Linear Models in PyMC3	GLM: Hierarchical Linear Regression	Hierarchical Binominal Model: Rat Tumor Example	GLM: Linear regression	GLM: Logistic Regression	GLM: Model Selection
GLM: Negative Binomial Regression	GLM: Poisson Regression	GLM: Robust Linear Regression	GLM: Robust Regression using Custom Likelihood for Outlier Classification	Rolling Regression	Hierarchical Partial Pooling

Summary

- Hierarchical modeling is more than statistics, its a way to think about problems
- More parameters != more flexibility
- It is easy to fit 1000s of parameters!
- There is no universal hierarchical model
- Hierarchical modeling is **everywhere**

Summary

- Hierarchical modeling is more than statistics, its a way to think about problems
- More parameters != more flexibility
- It is easy to fit 1000s of parameters!
- There is no universal hierarchical model
- Hierarchical modeling is **everywhere**

