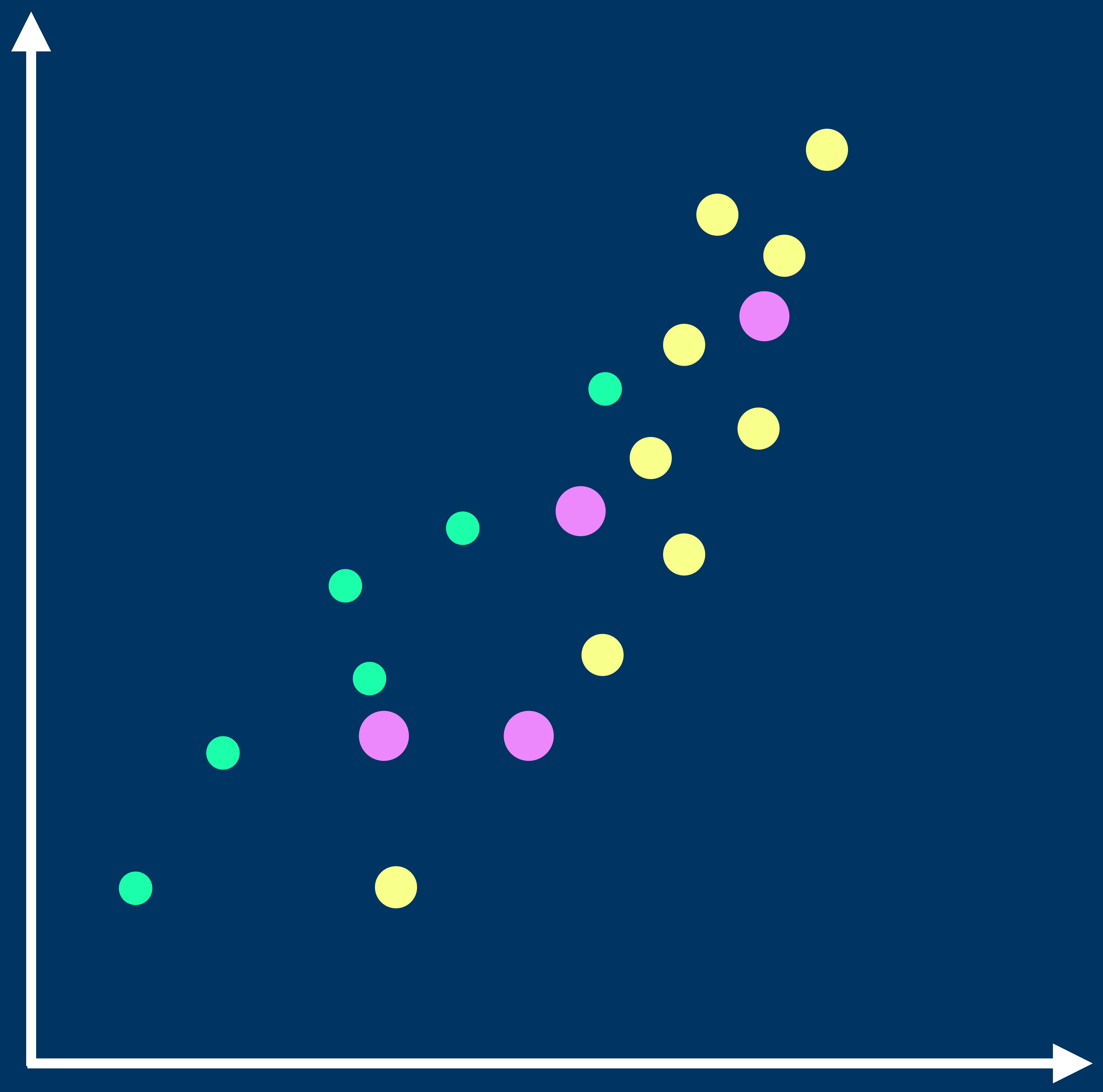
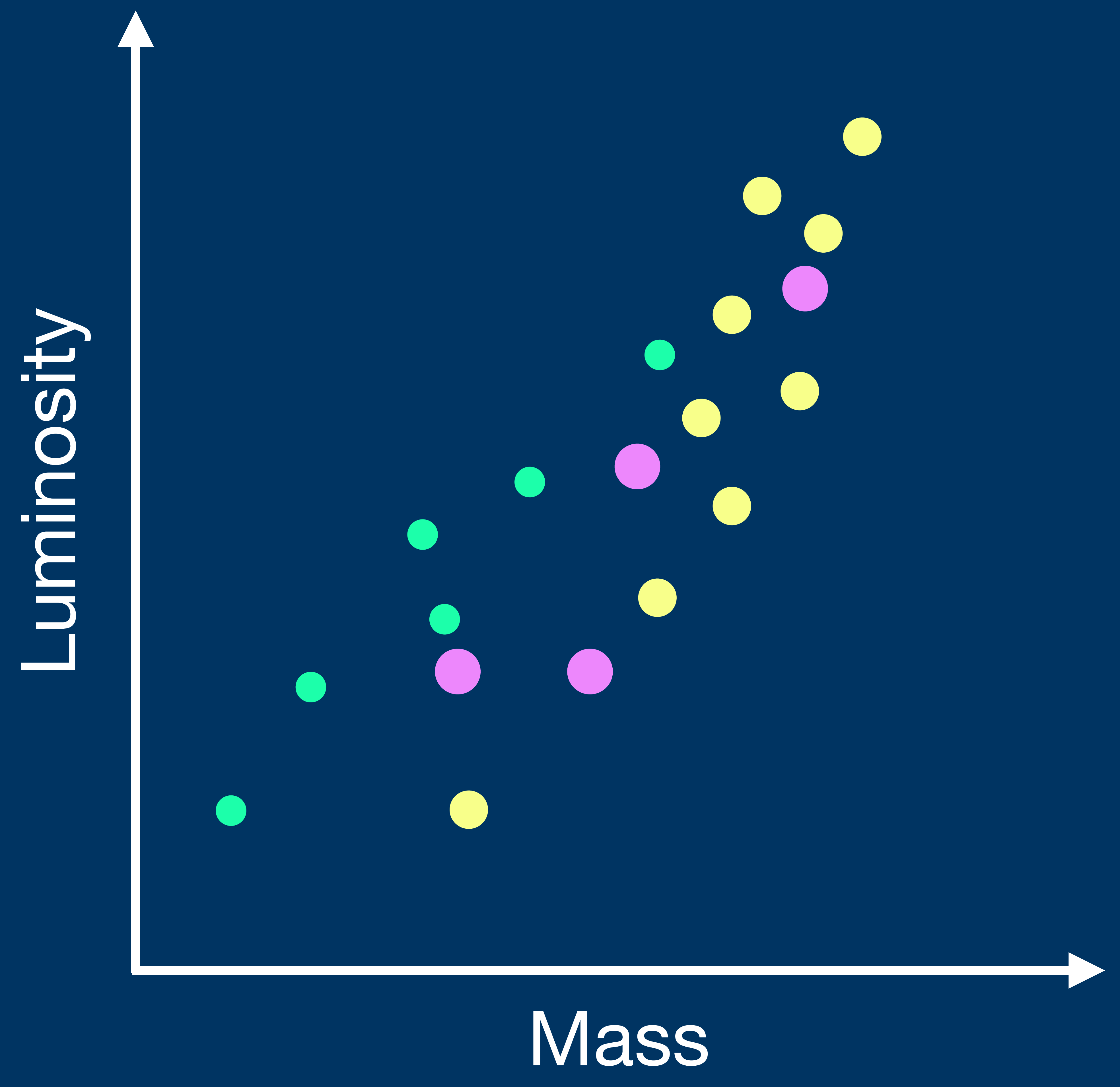
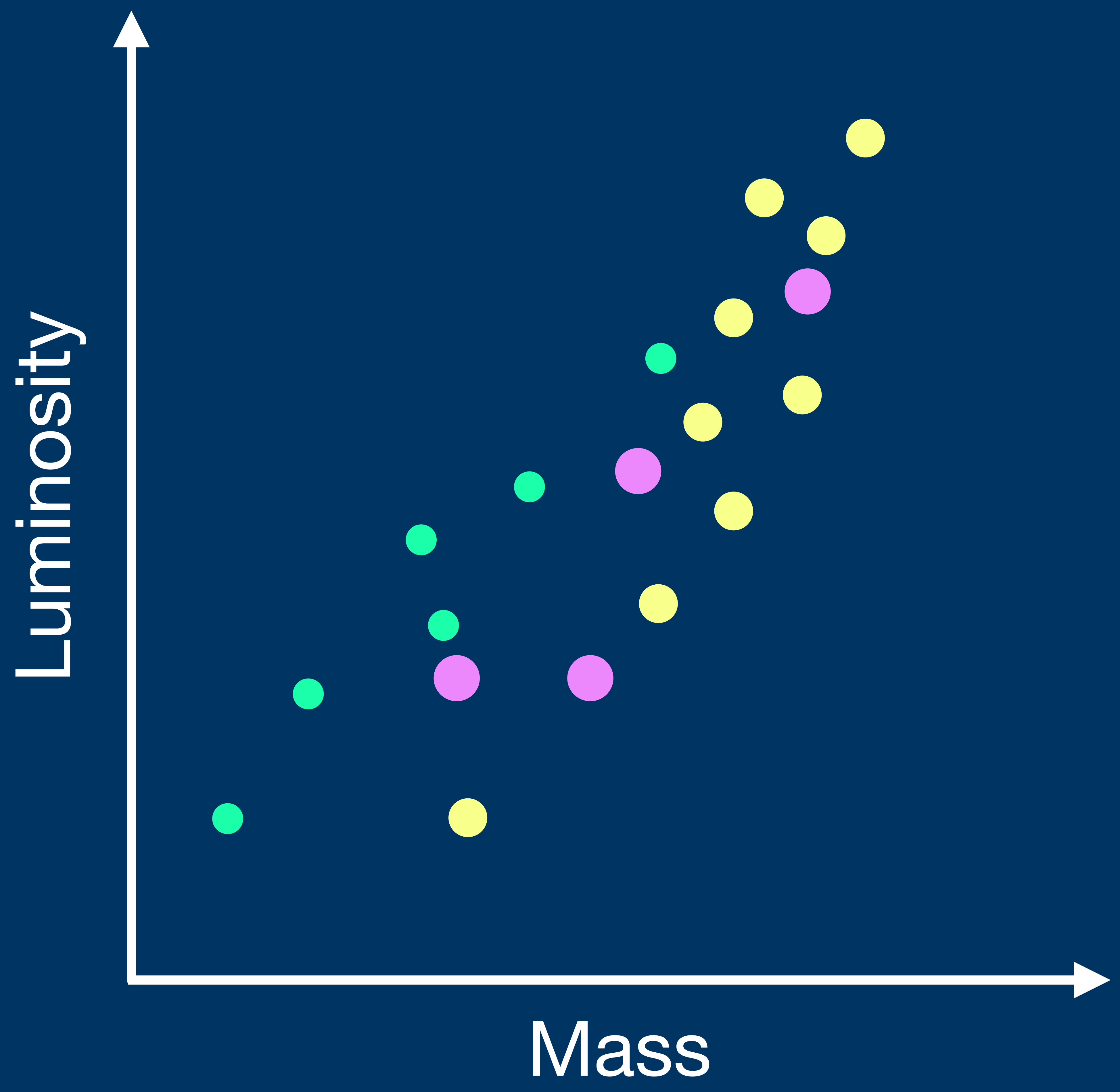
Hierarchical Bayesian Models: What are they?

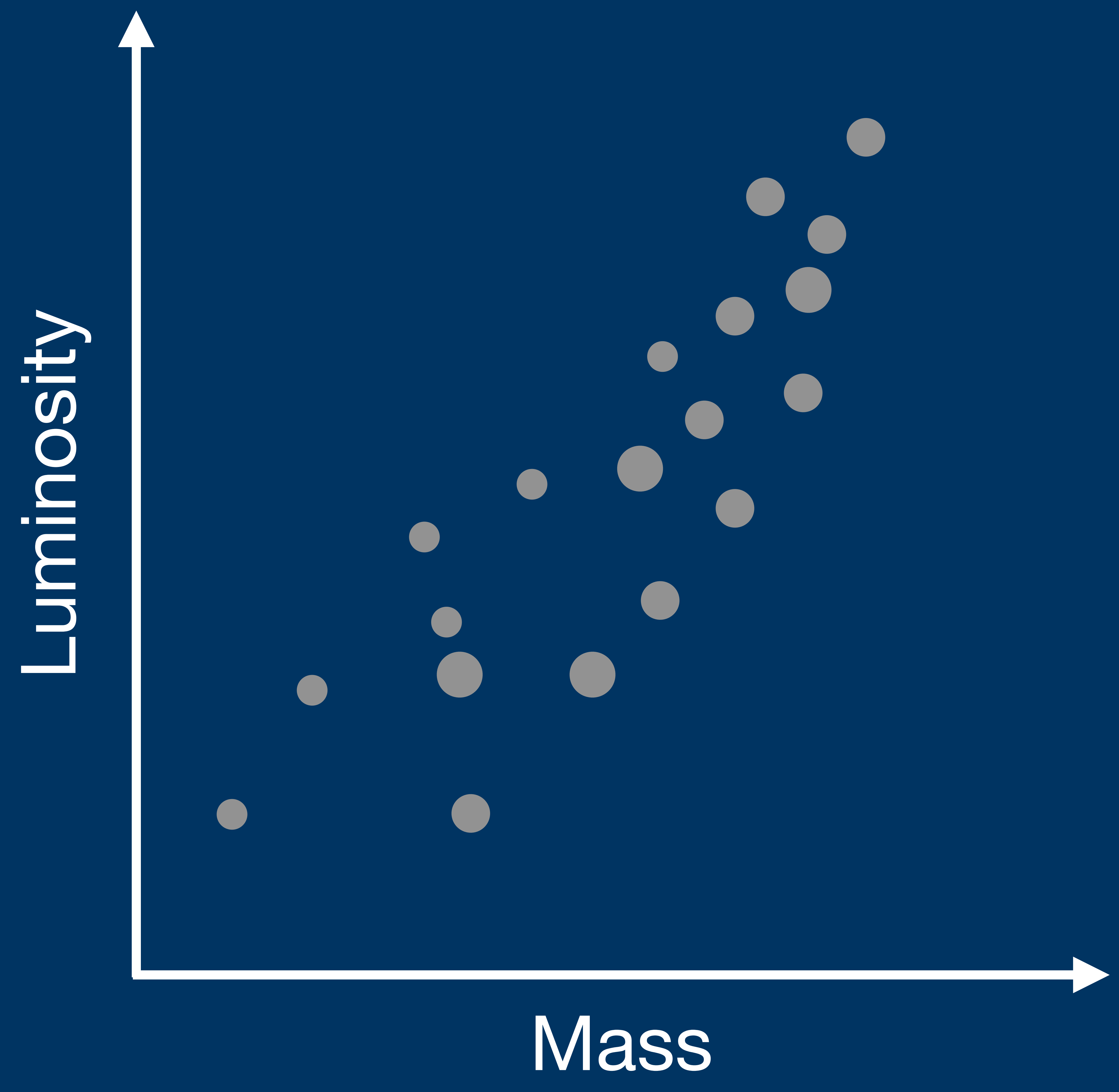
Examples from astrophysics

J. Michael Burgess
MPE



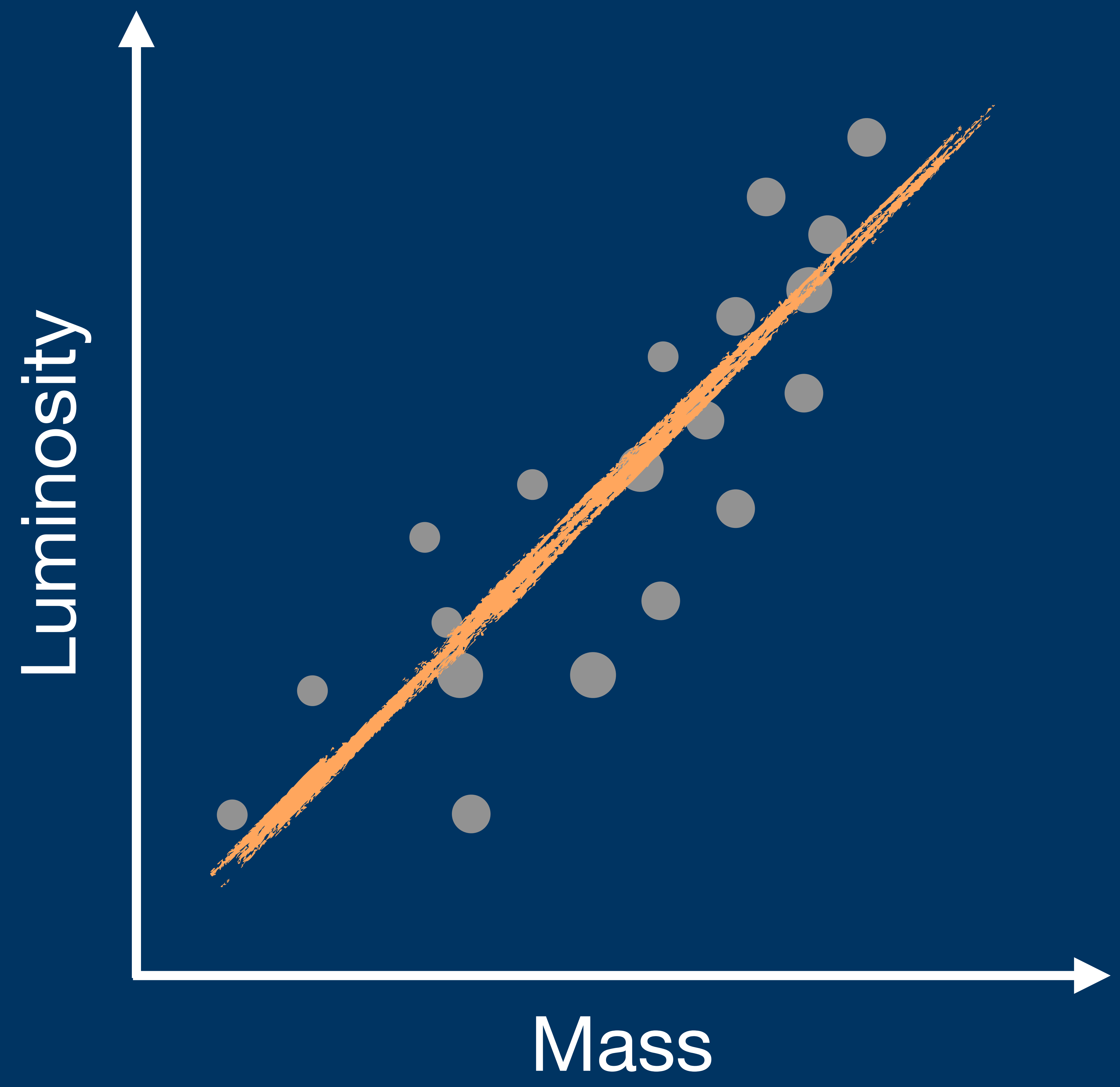

Luminosity
Mass


Luminosity
Mass



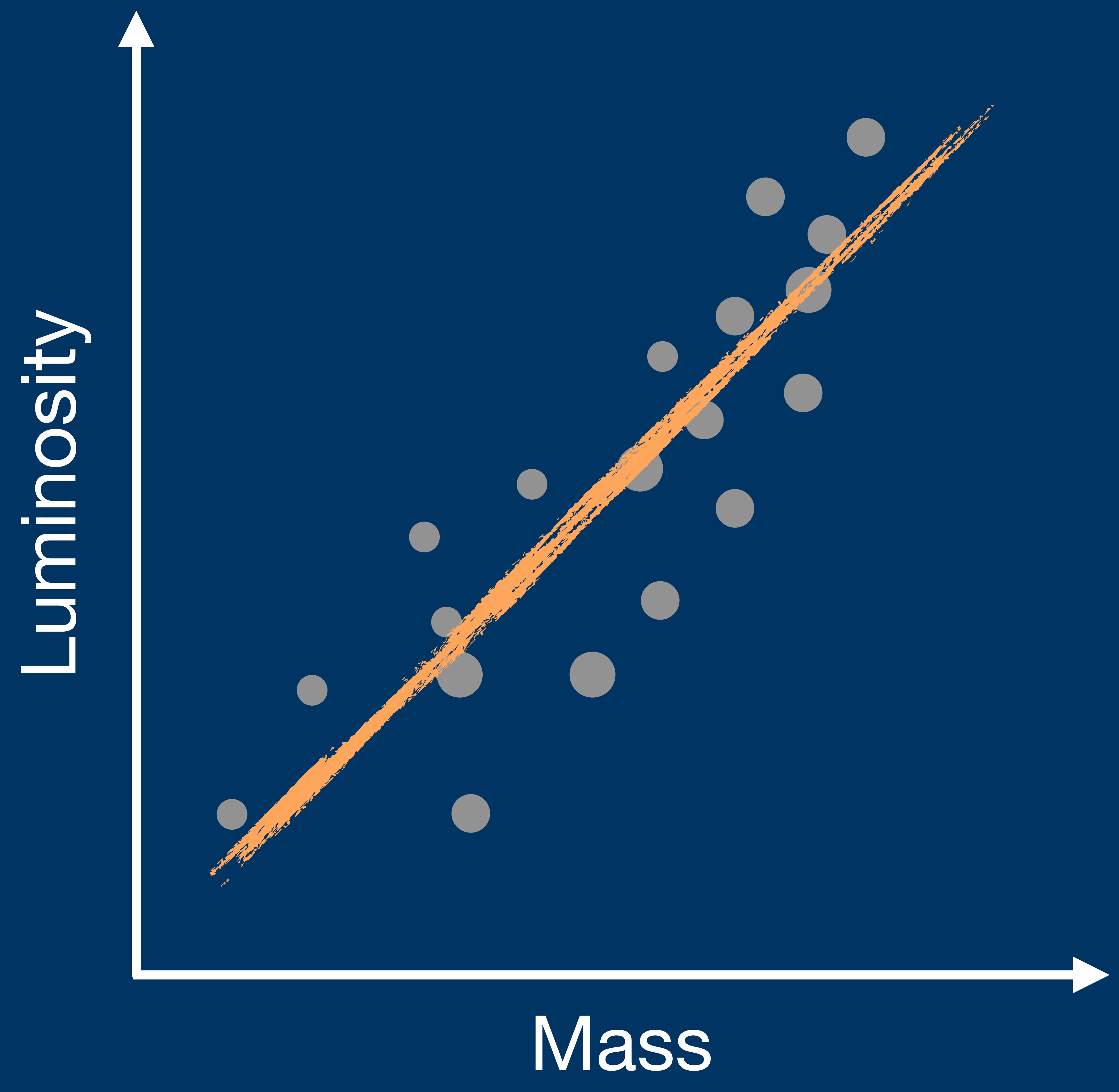
$$L = \alpha M + \beta$$

How do we get global information?

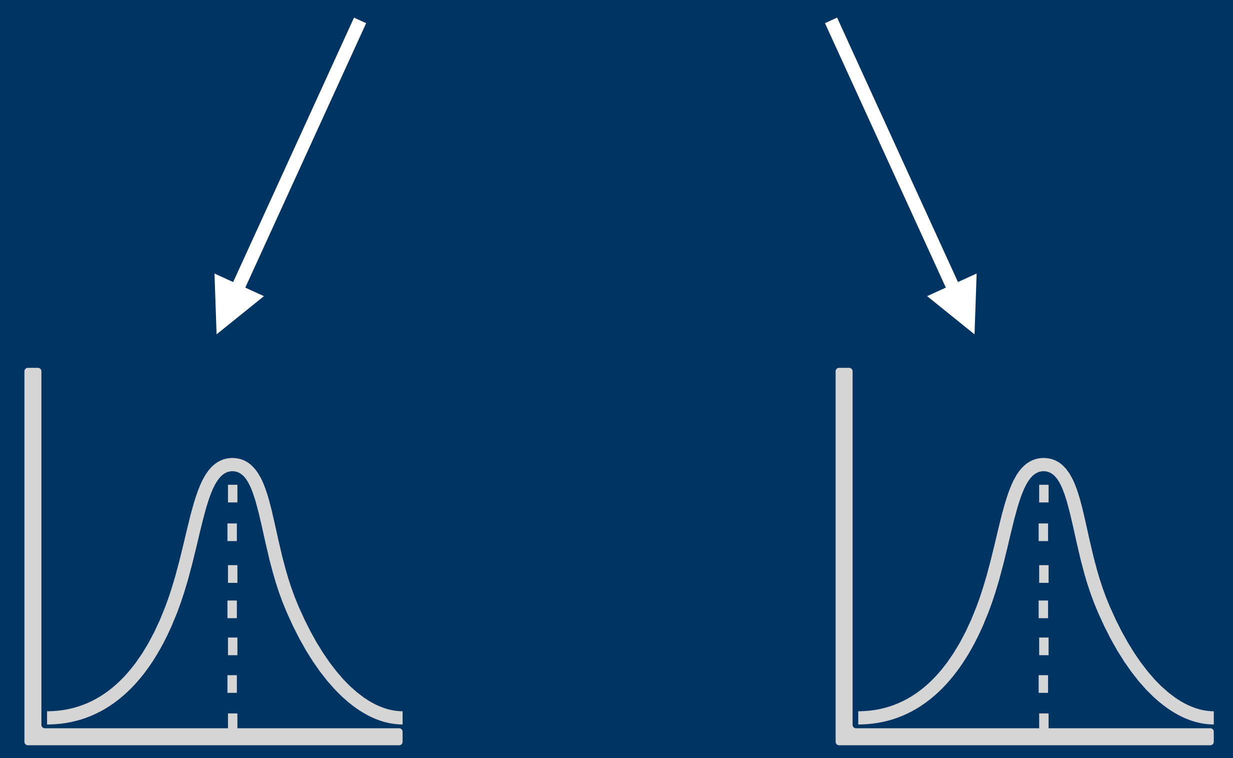


$$L = \alpha M + \beta$$

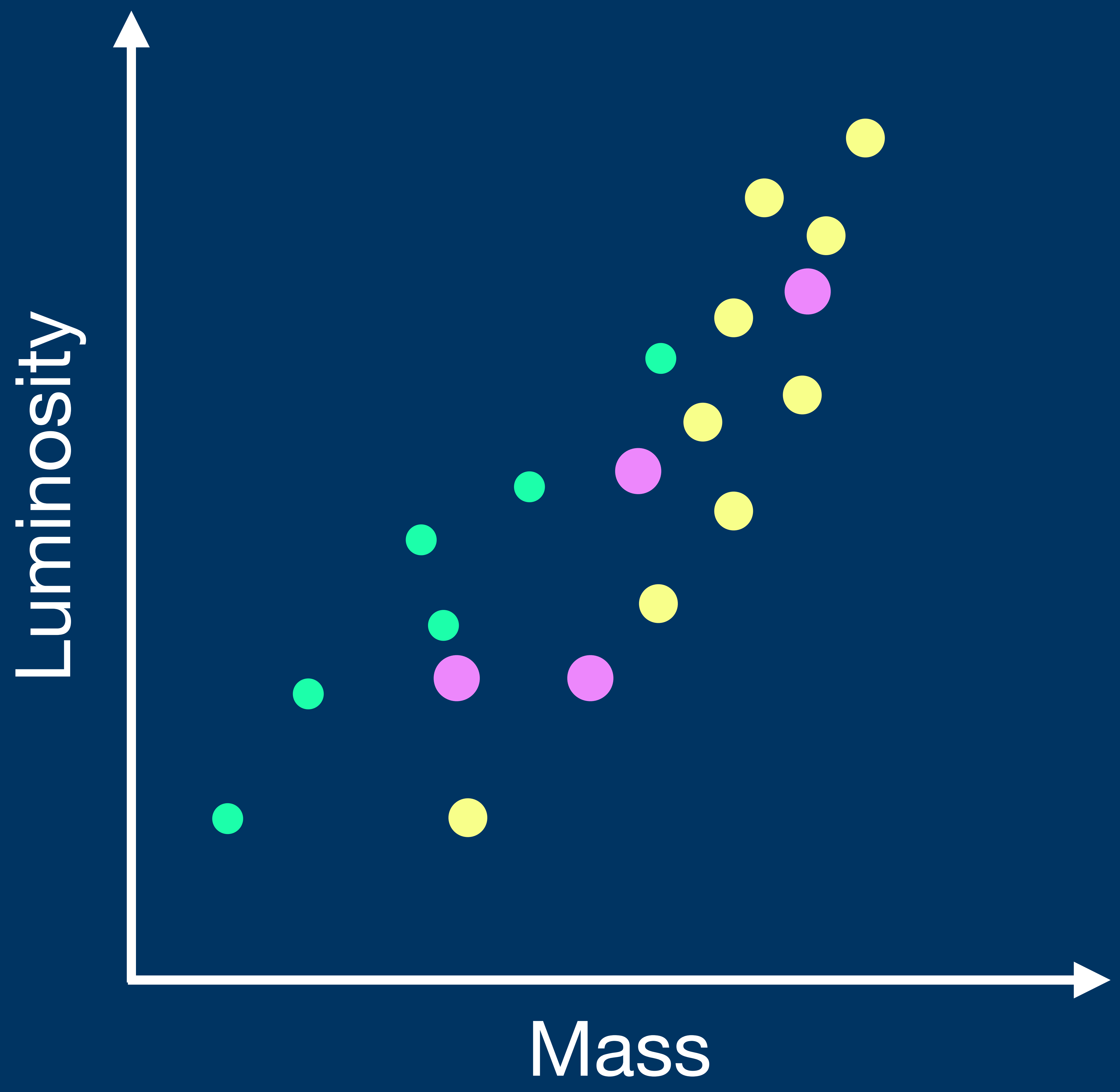
How do we get global information?



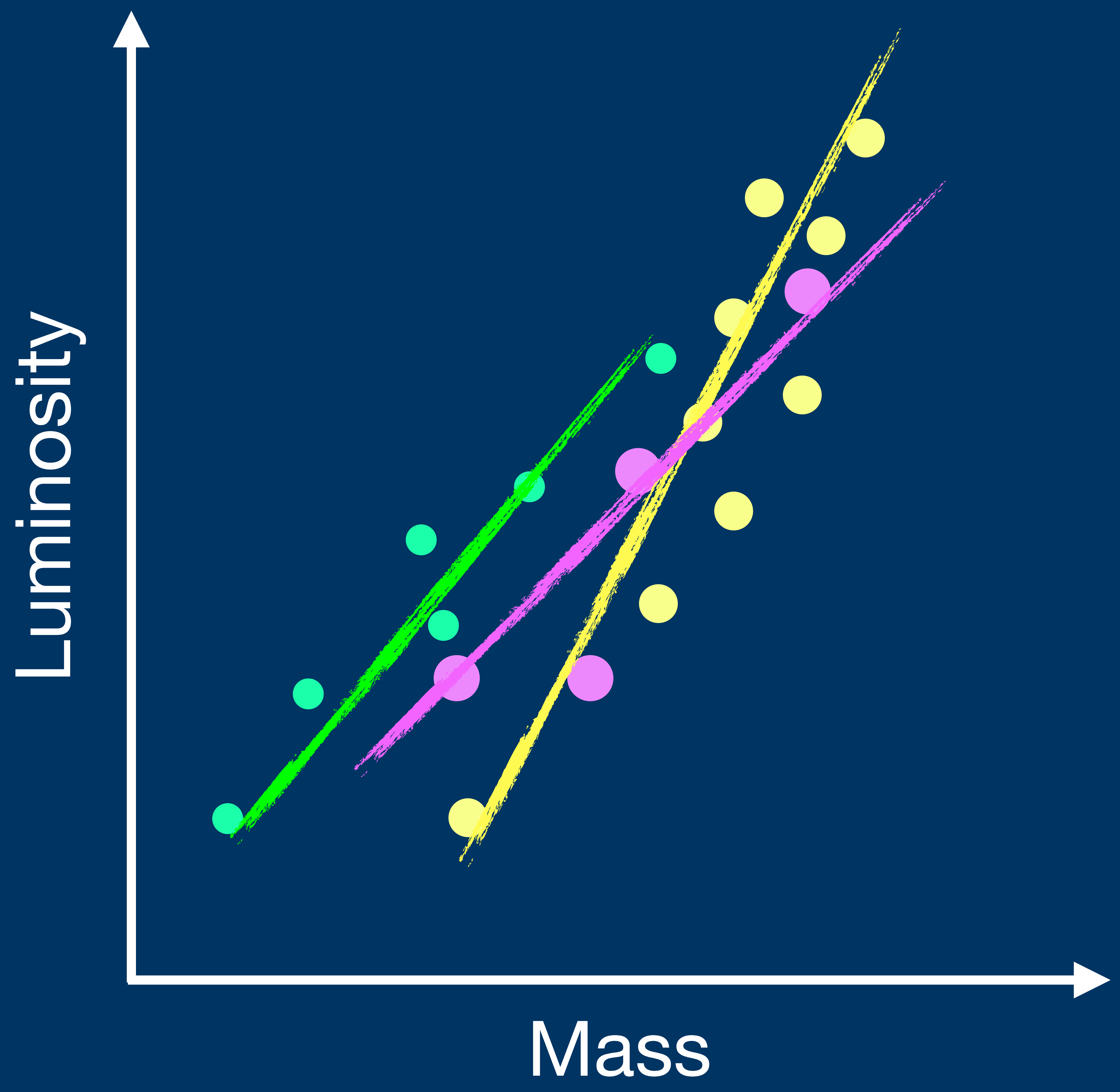
$$L = \alpha M + \beta$$



How do we get global information?



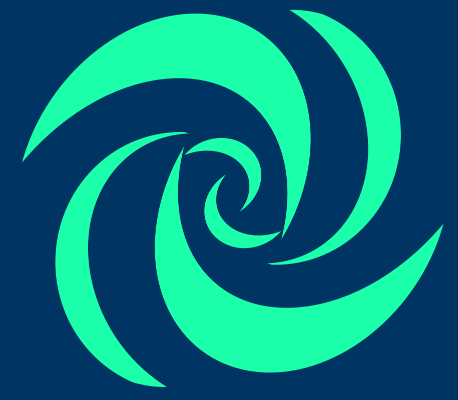
How do we incorporate all the information?



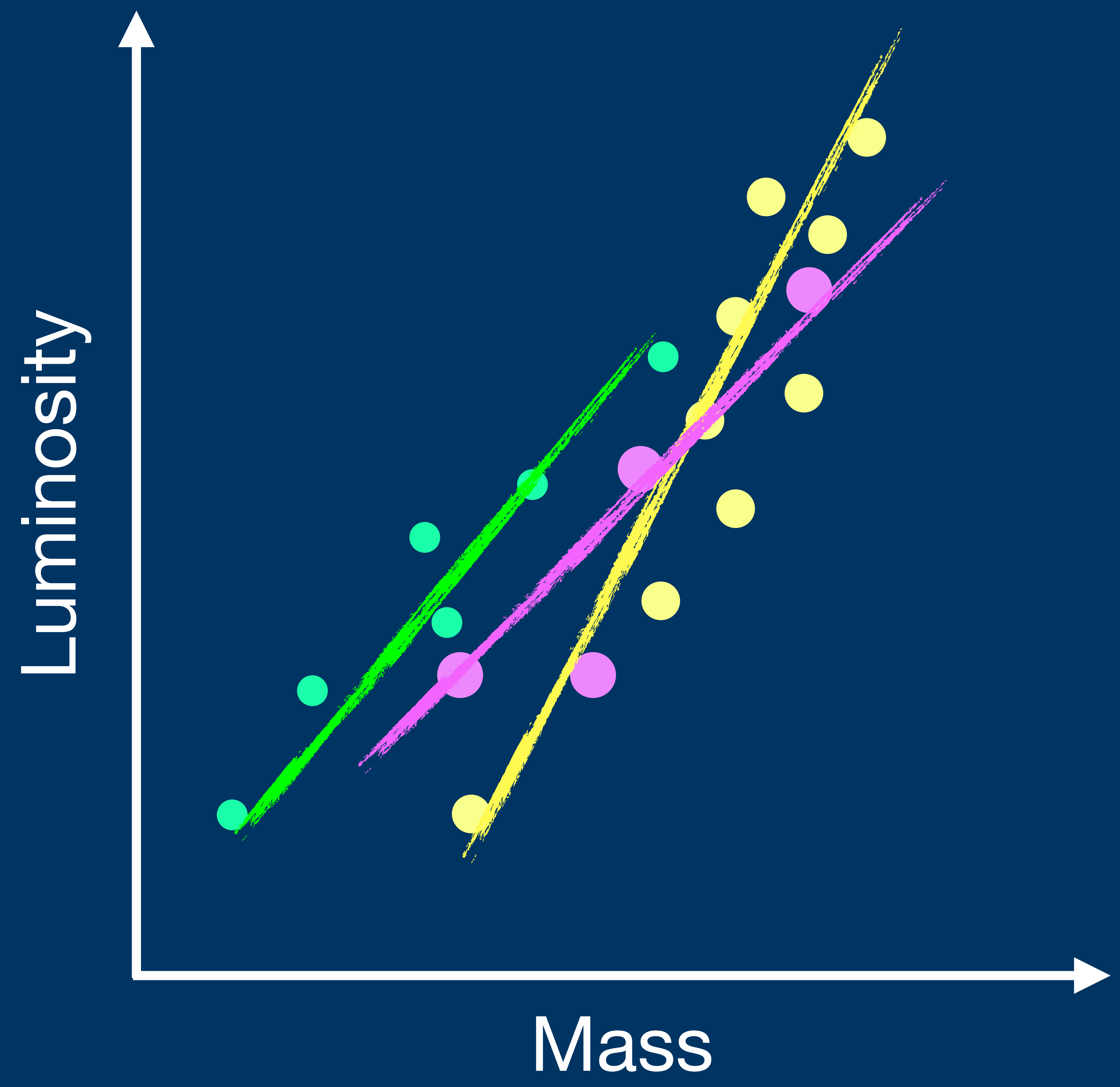
$$L = \alpha M + \beta$$

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

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



How do we incorporate all the information?

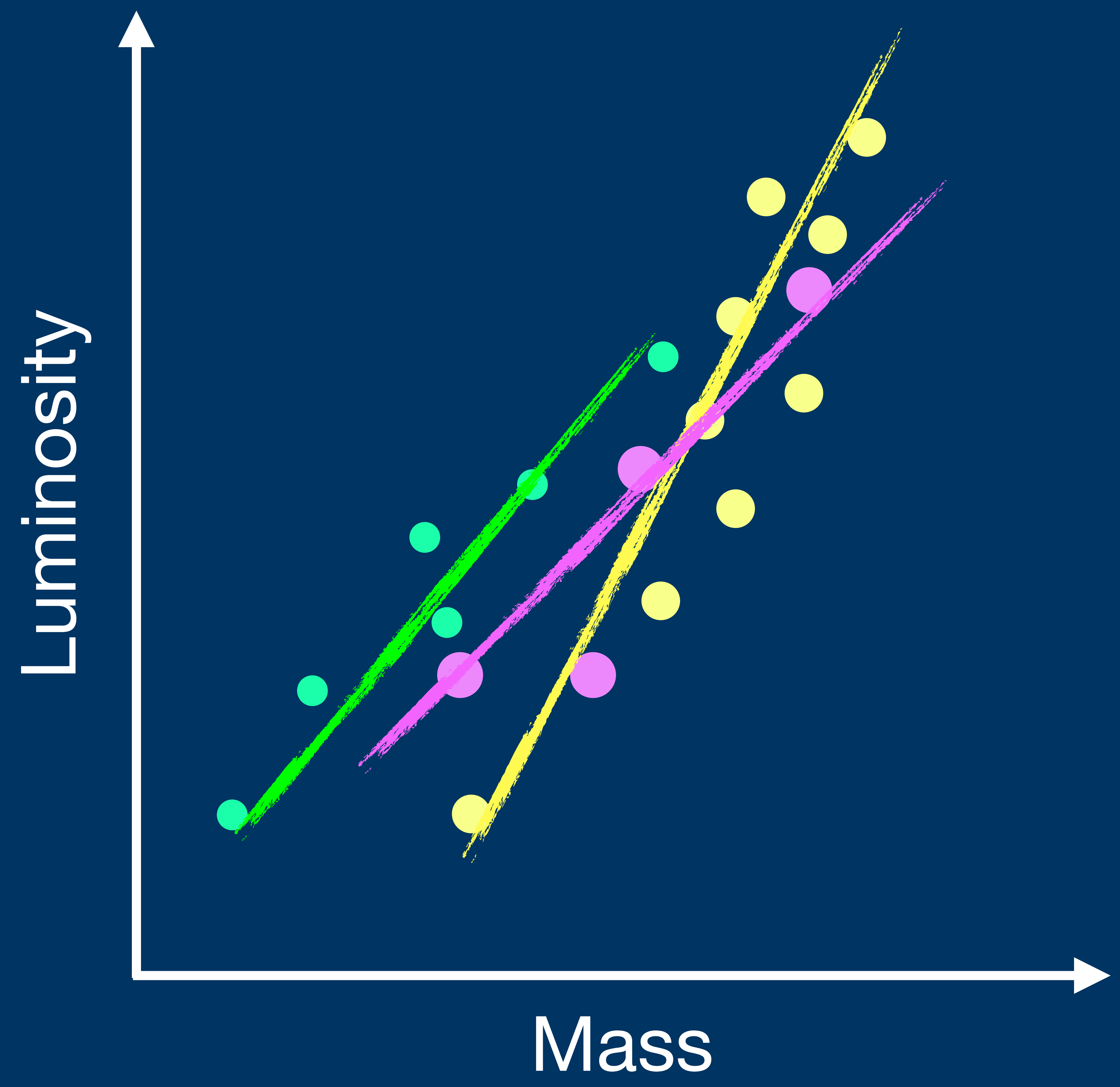


$$L = \alpha M + \beta$$

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

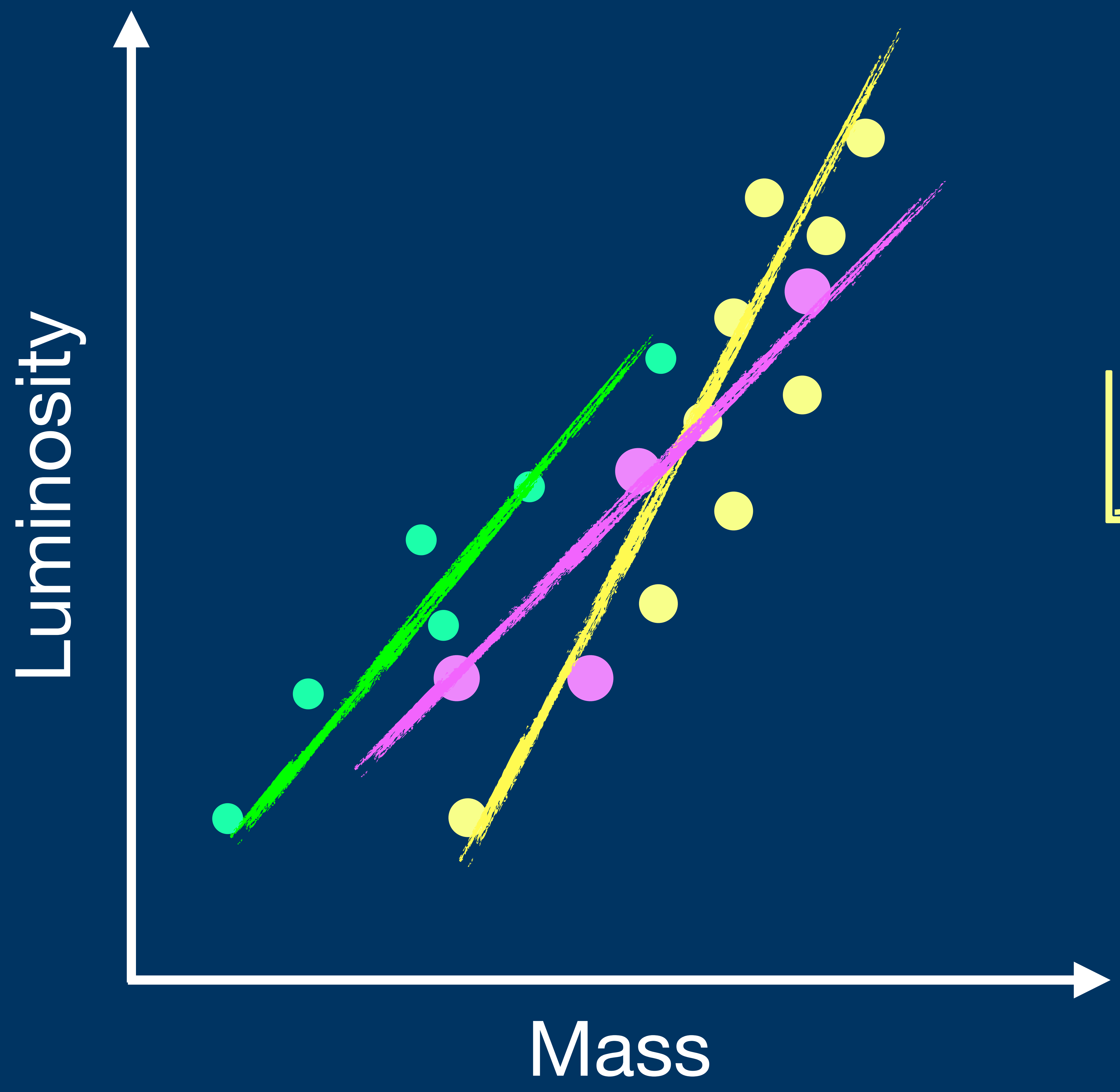
$$L = \alpha M + \beta$$

How do we incorporate all the information?

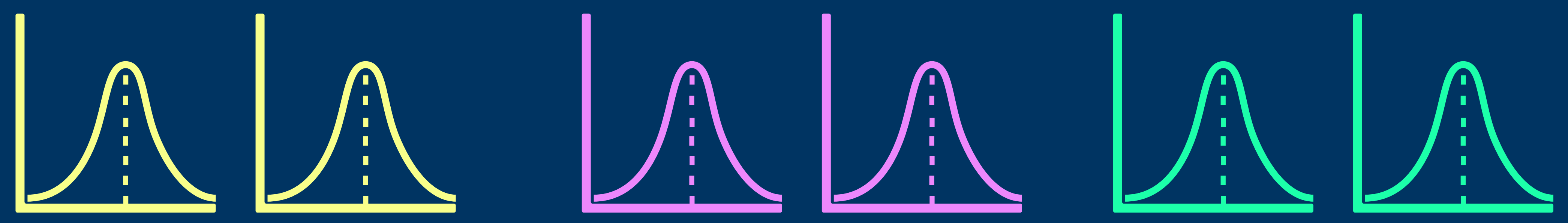


$L = \alpha M + \beta$ $L = \alpha M + \beta$ $L = \alpha M + \beta$

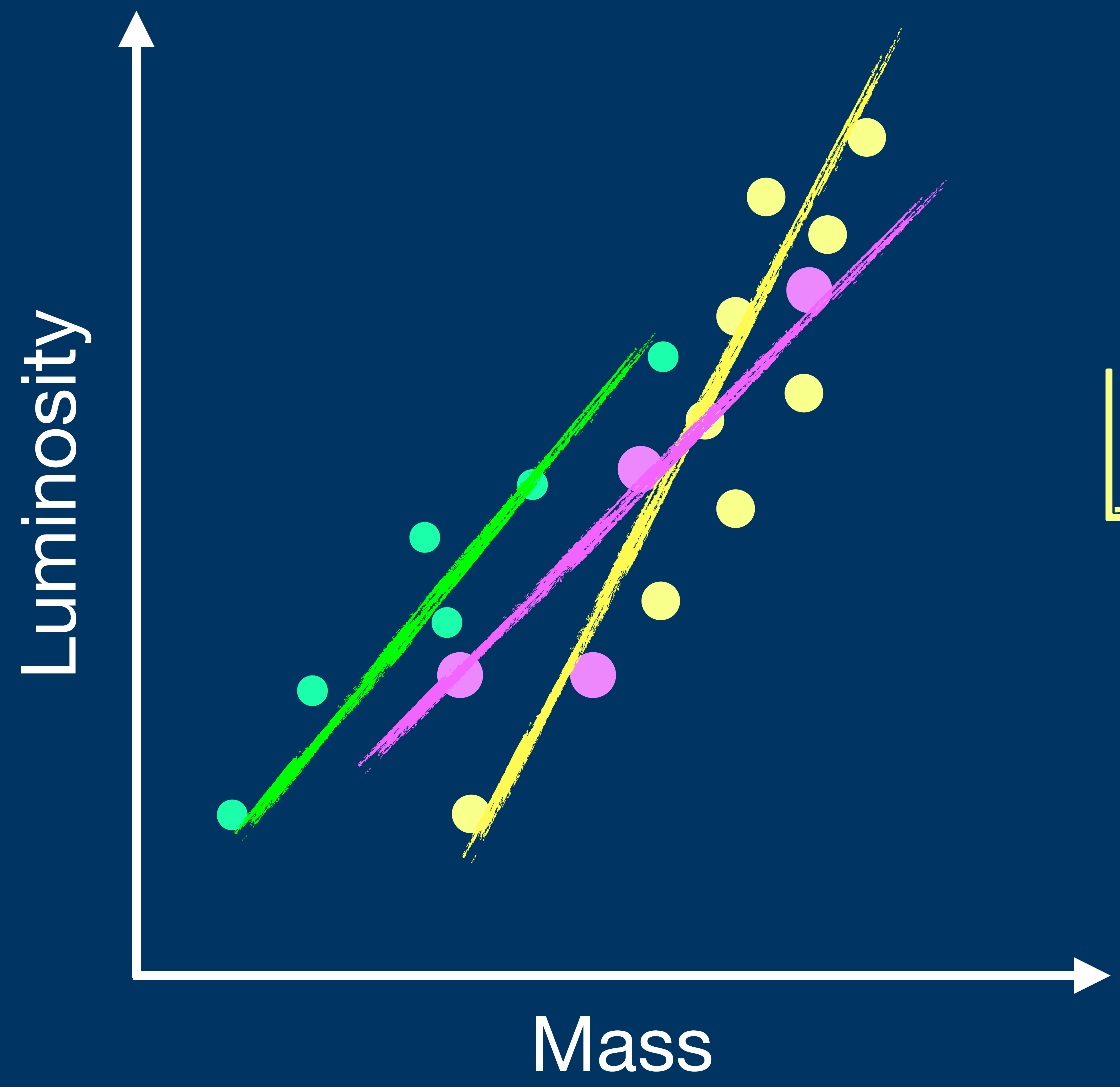
How do we incorporate all the information?



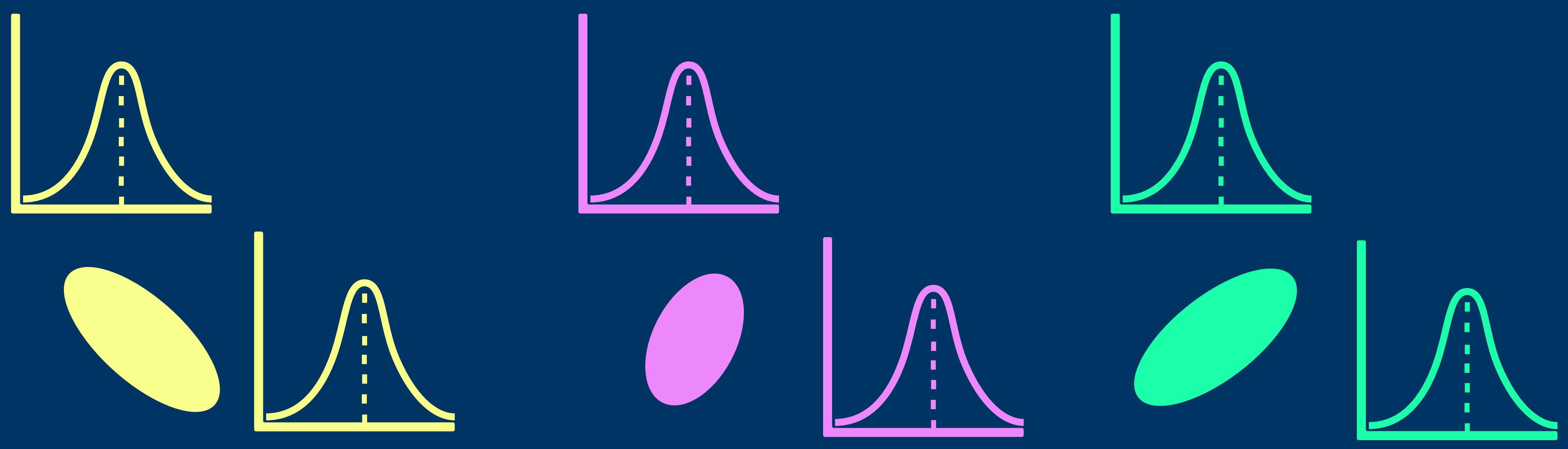
$L = \alpha M + \beta$ $L = \alpha M + \beta$ $L = \alpha M + \beta$



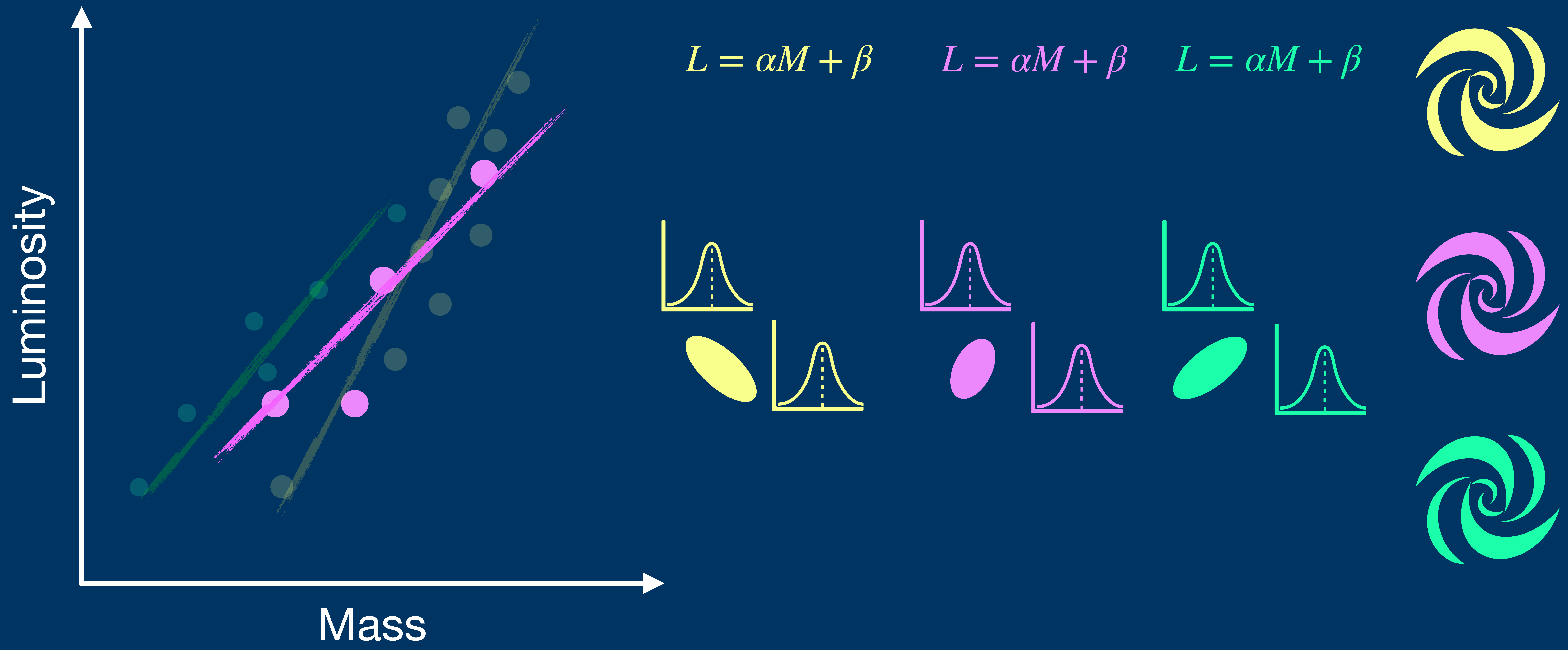
How do we incorporate all the information?



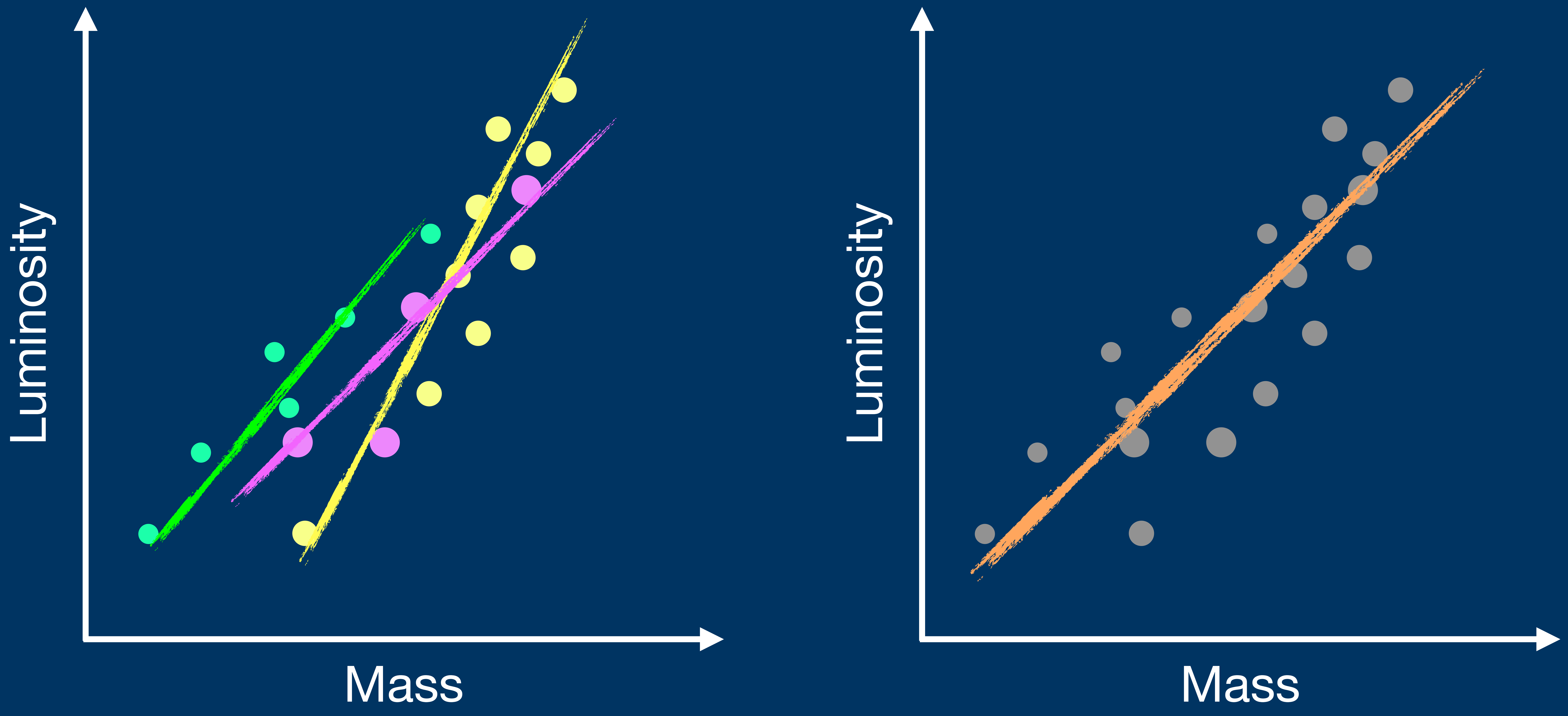
$L = \alpha M + \beta$ $L = \alpha M + \beta$ $L = \alpha M + \beta$

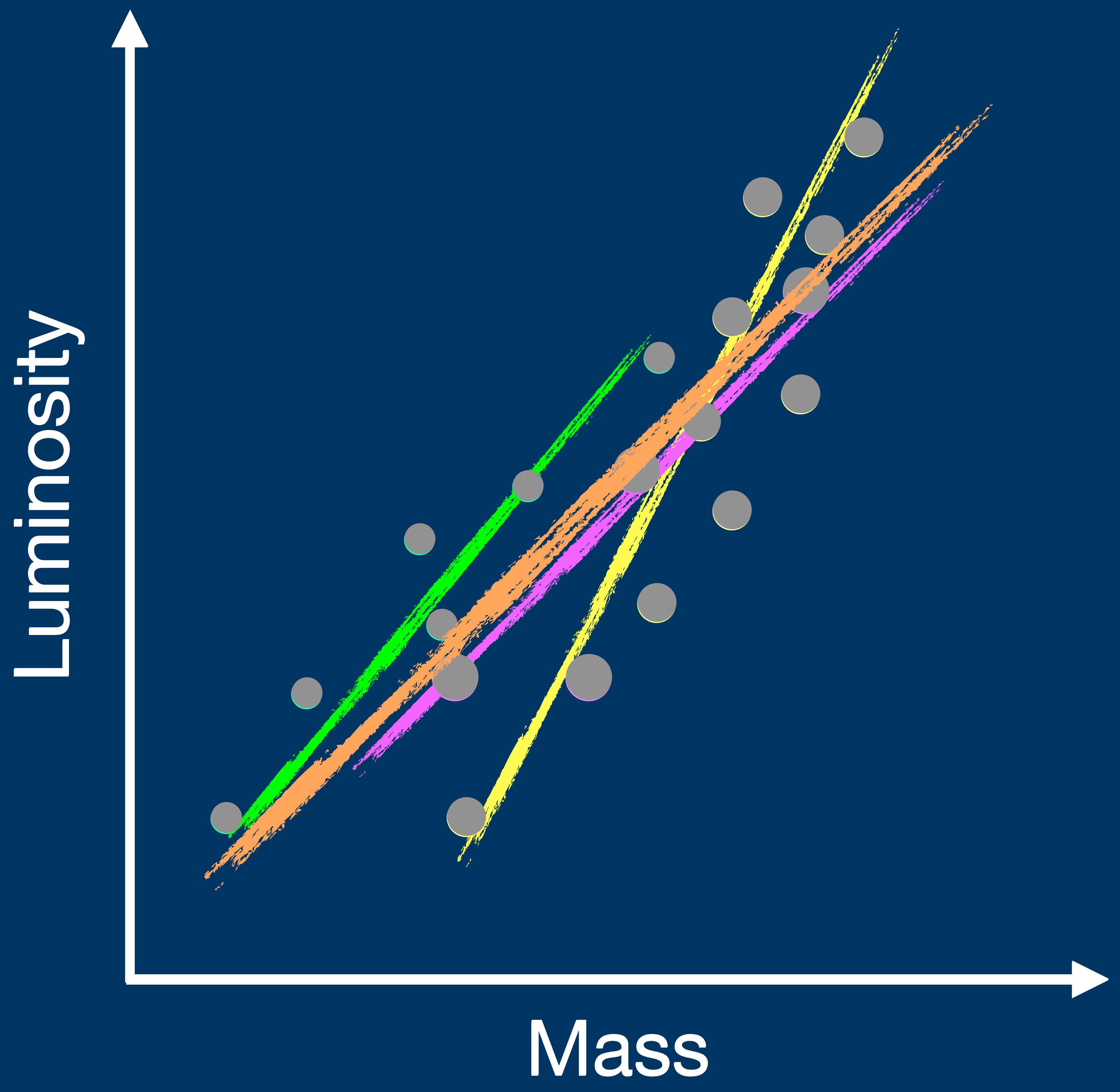


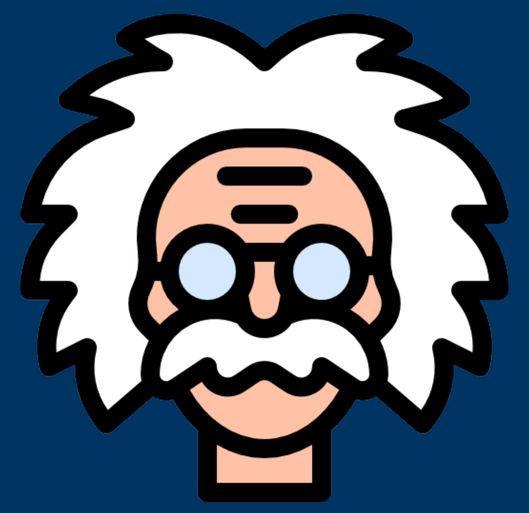
How do we incorporate all the information?

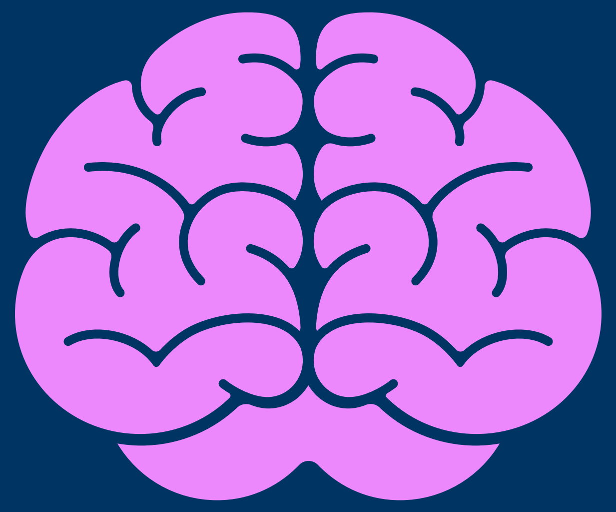
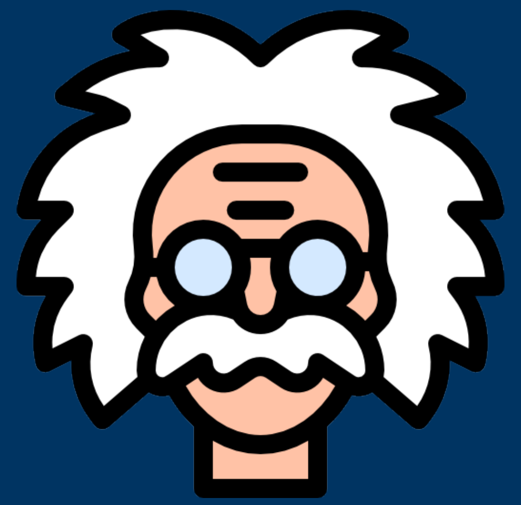


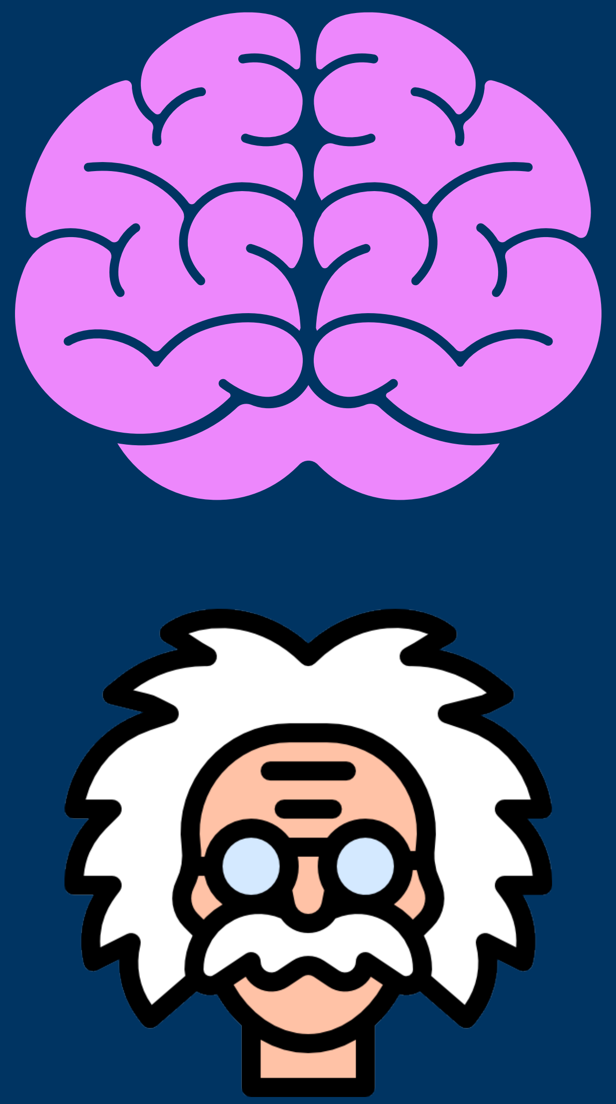
How do we incorporate all the information?

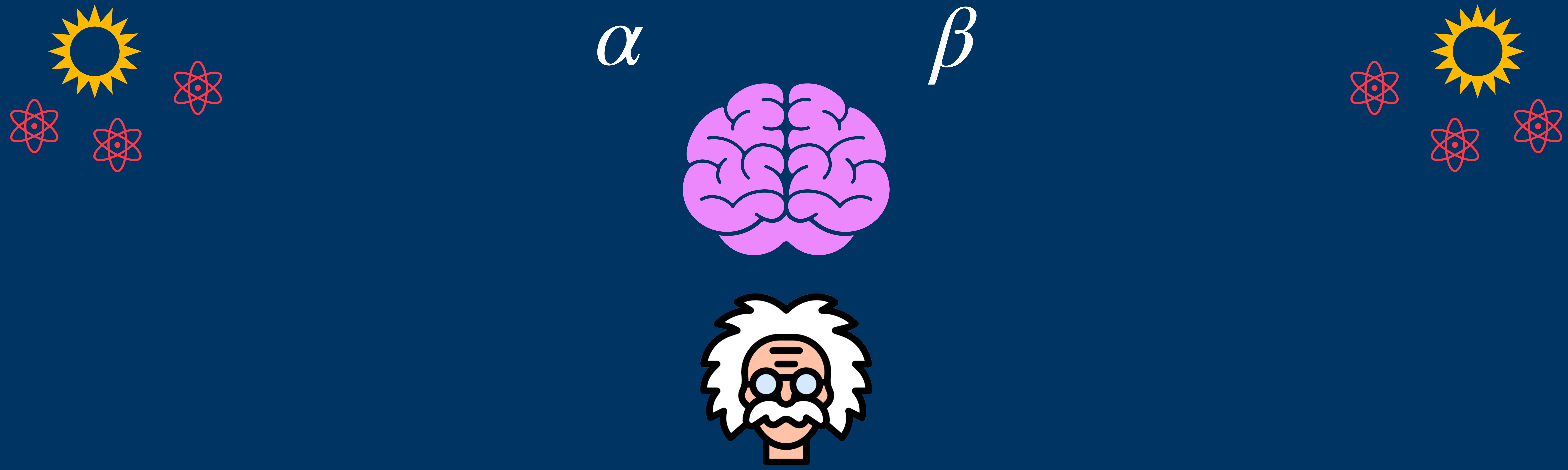

Luminosity
Mass
Luminosity
Mass

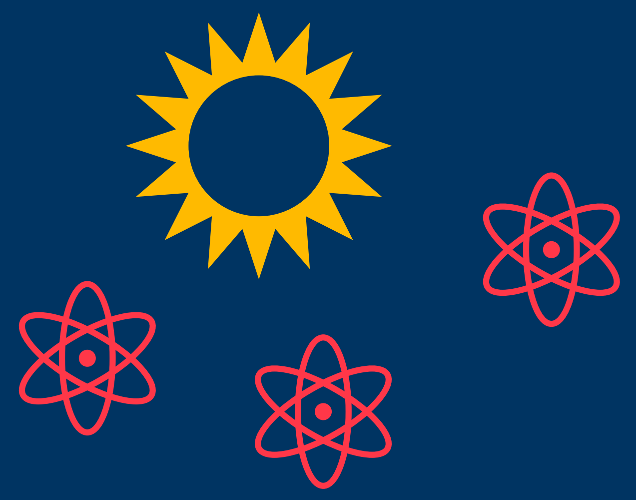

Luminosity
Mass



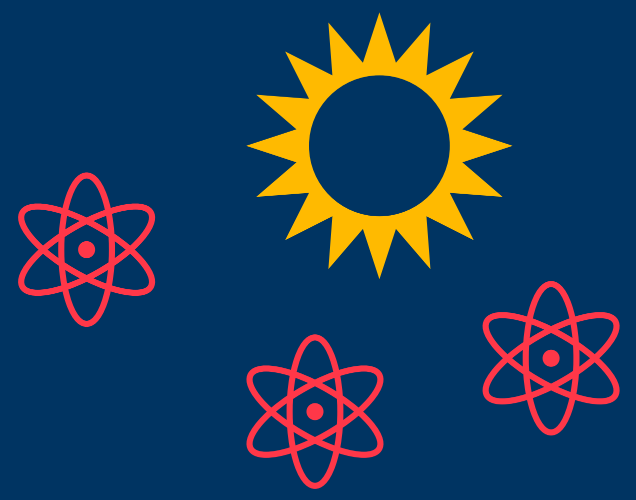






α
β

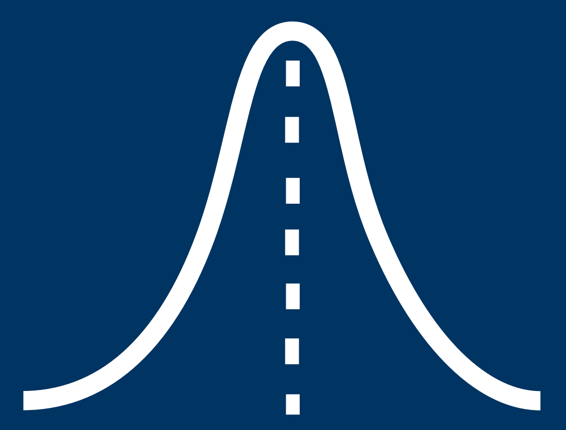


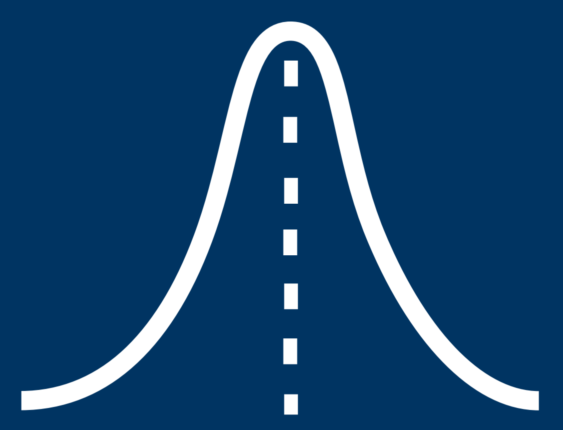

α
β

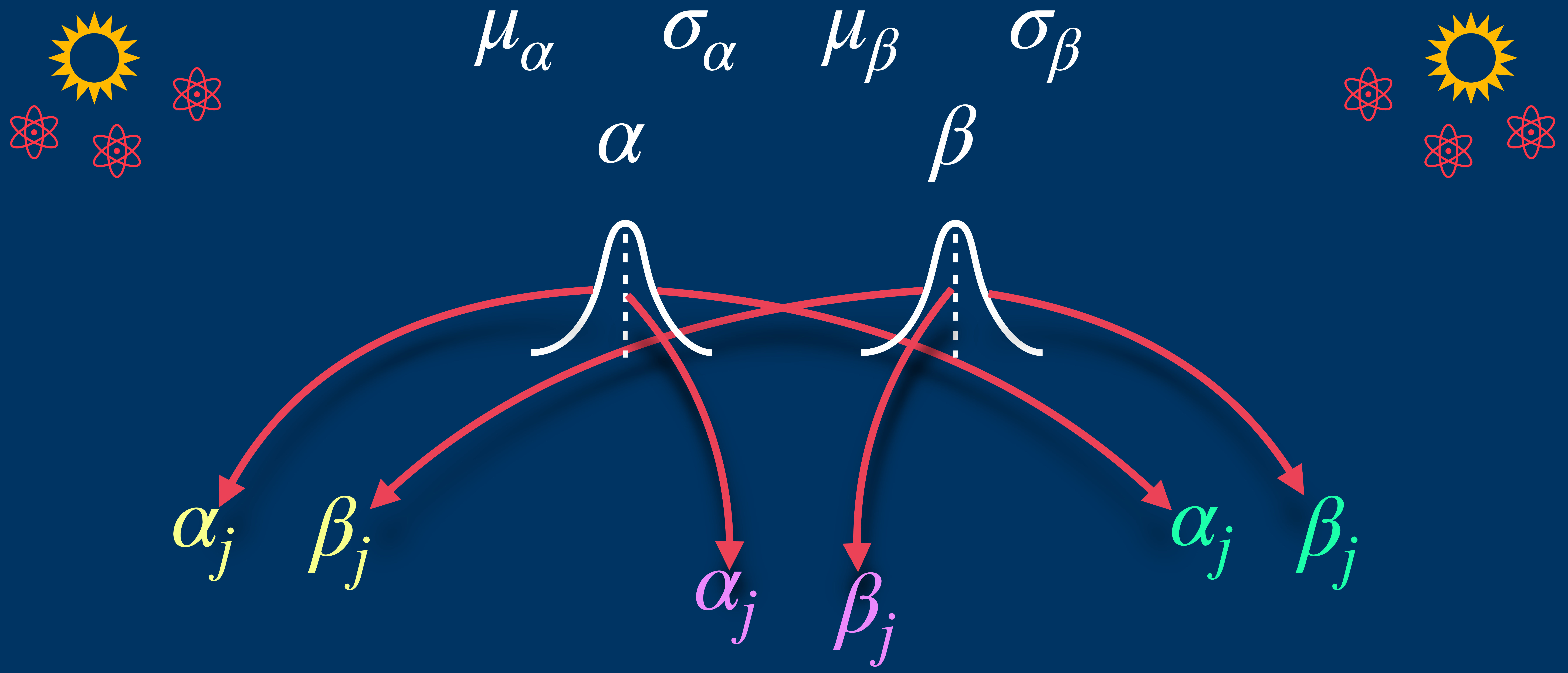


μ_α σ_α μ_β σ_β

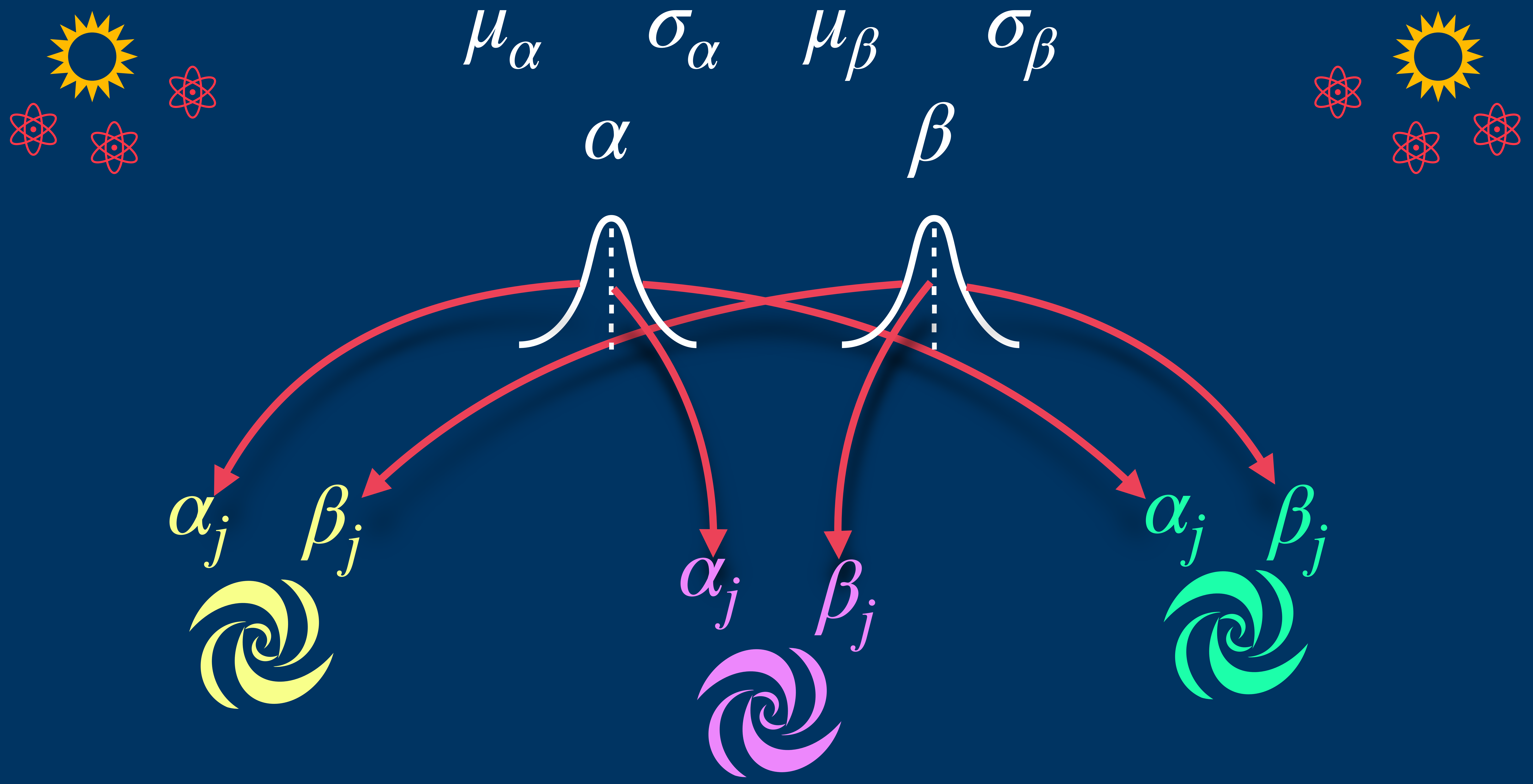
α β



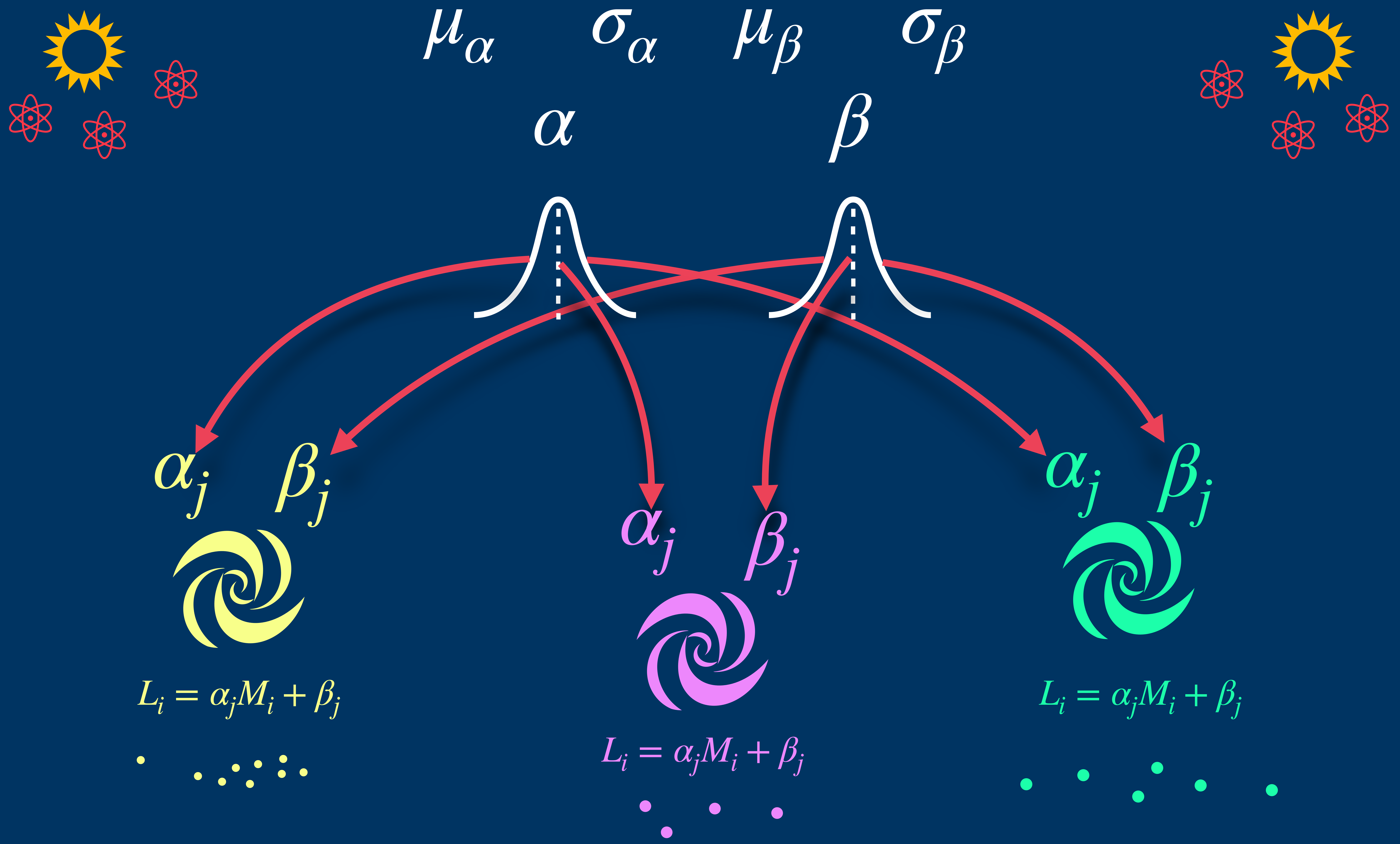




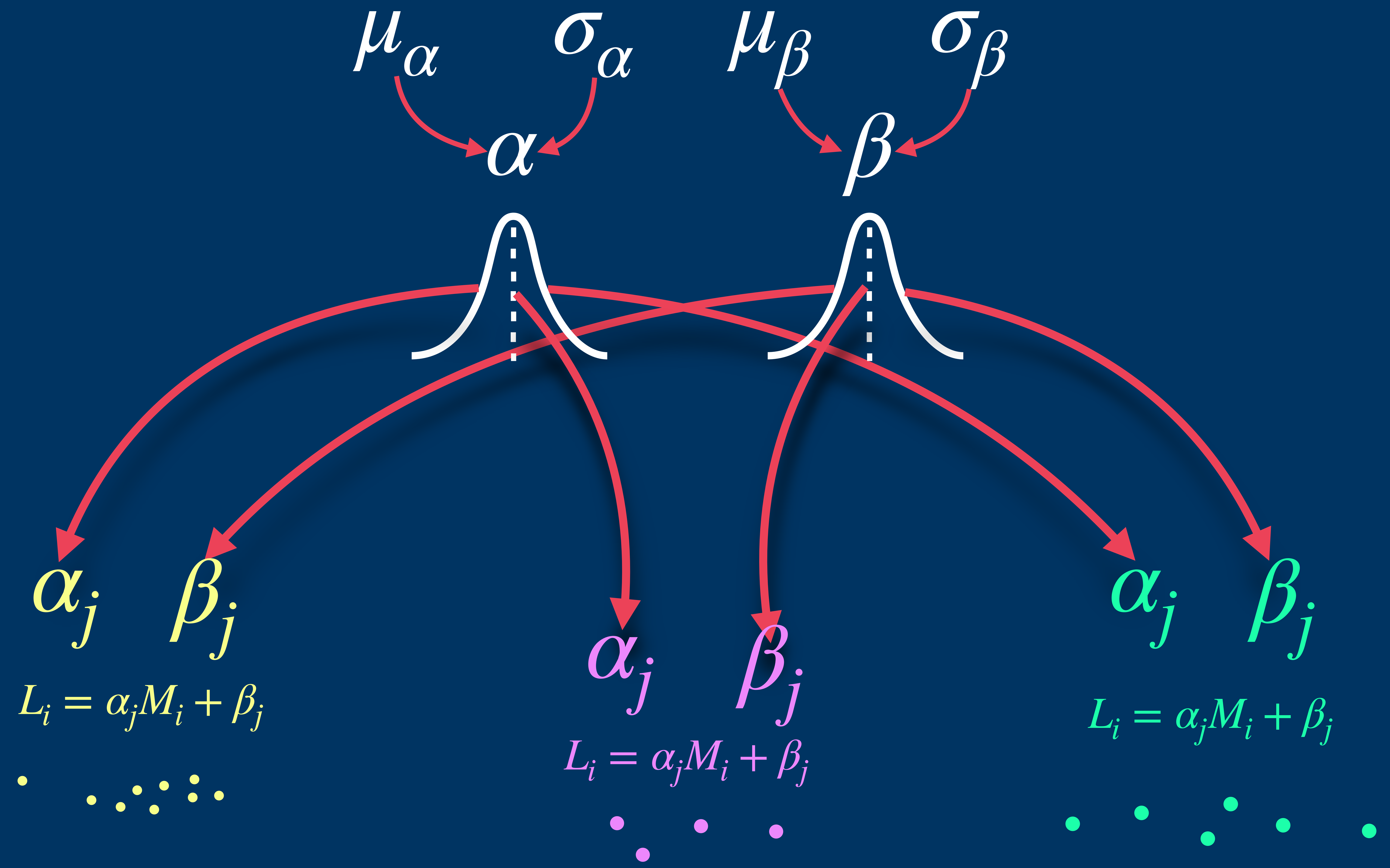
μ_α
σ_α
μ_β
σ_β
α
β
α_j
β_j
α_j
β_j
α_j
β_j



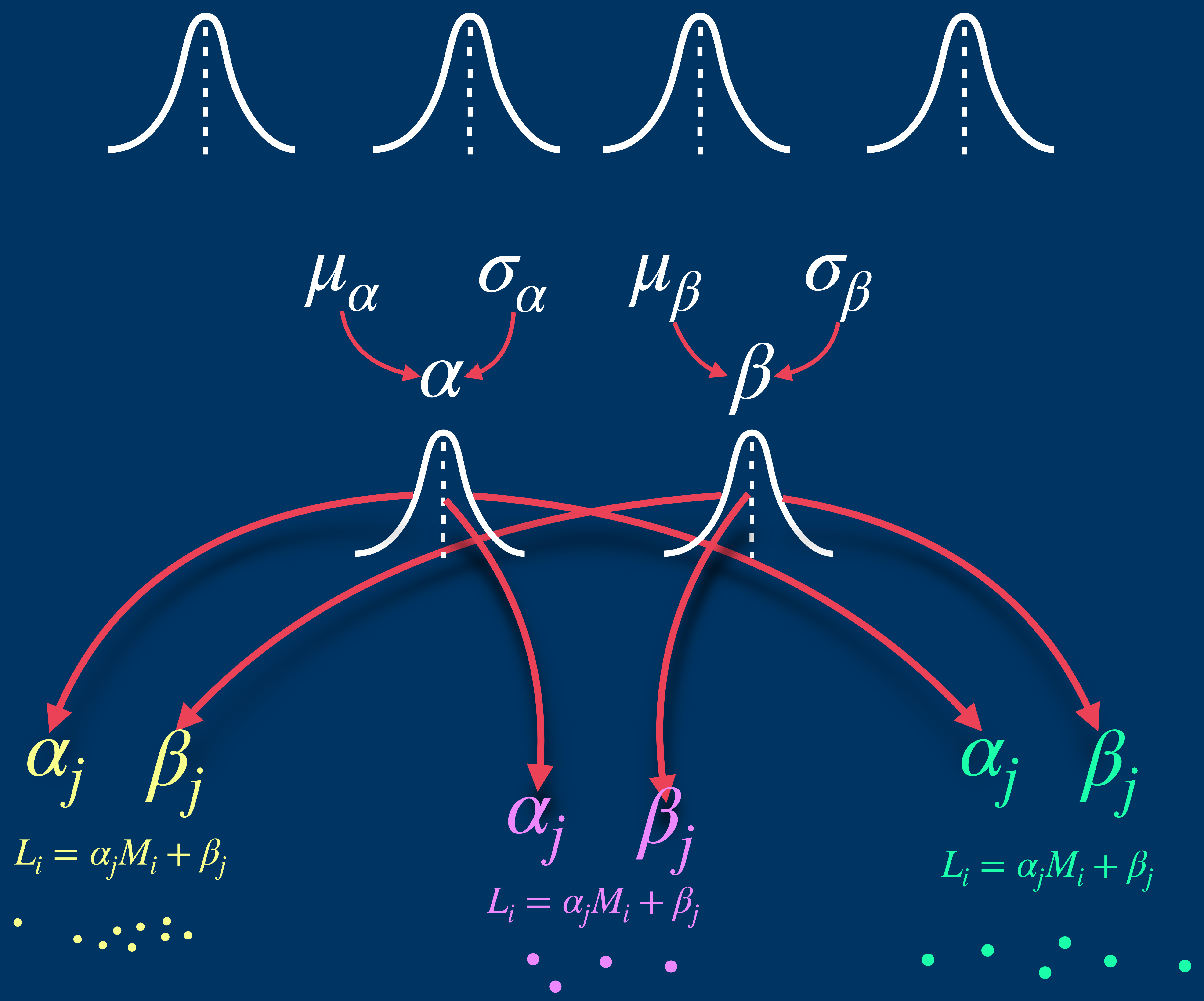
μ_α σ_α μ_β σ_β
α
β
α_j β_j
α_j β_j
α_j β_j



μ_α σ_α μ_β σ_β
α
β
α_j β_j
α_j β_j
α_j β_j
$L_i = \alpha_j M_i + \beta_j$
$L_i = \alpha_j M_i + \beta_j$
$L_i = \alpha_j M_i + \beta_j$



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



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

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

posterior

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

posterior

likelihood

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

posterior — likelihood — group level priors

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

posterior | likelihood | group level priors | hyper priors

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

posterior — likelihood — group level priors — hyper priors

Stan

PyMC3

cosmology

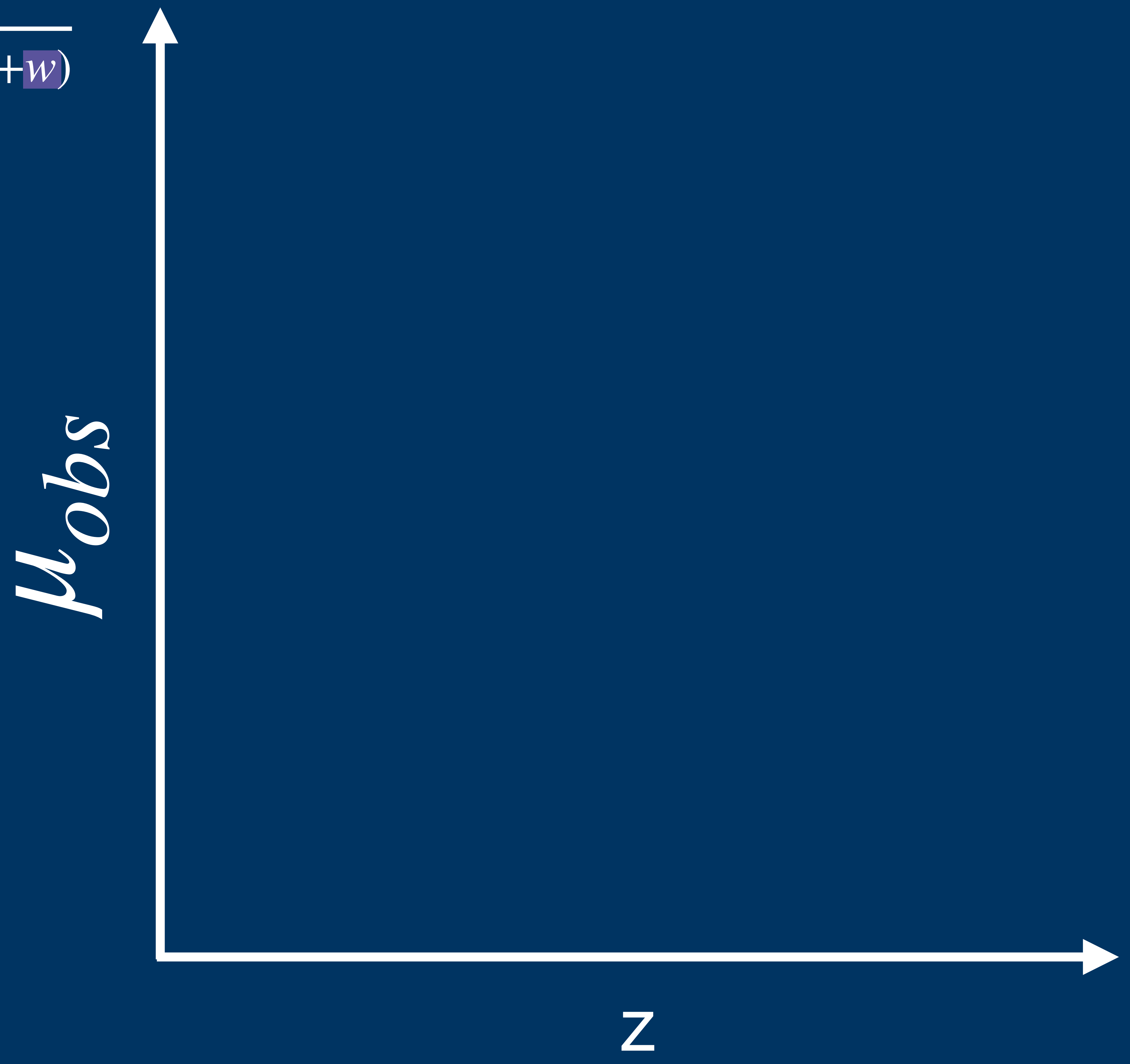
$$E(z) = \sqrt{\Omega_m (1+z')^3 + \Omega_k (1+z')^2 + \Omega_\Lambda (1+z')^{3(1+w)}}$$

luminosity distance

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

distance modulus

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



cosmology

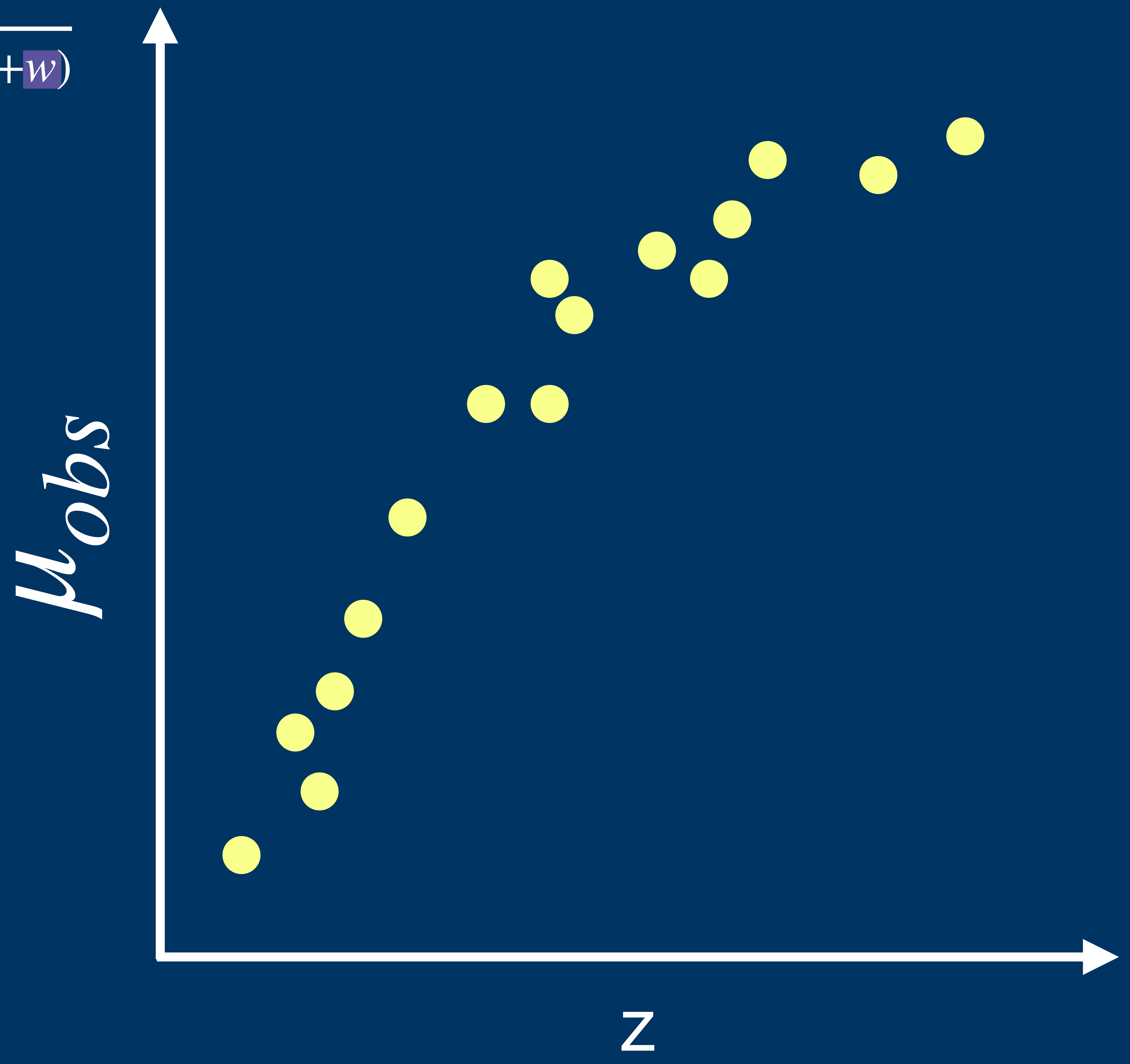
$$E(z) = \sqrt{\Omega_m (1+z')^3 + \Omega_k (1+z')^2 + \Omega_\Lambda (1+z')^{3(1+w)}}$$

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

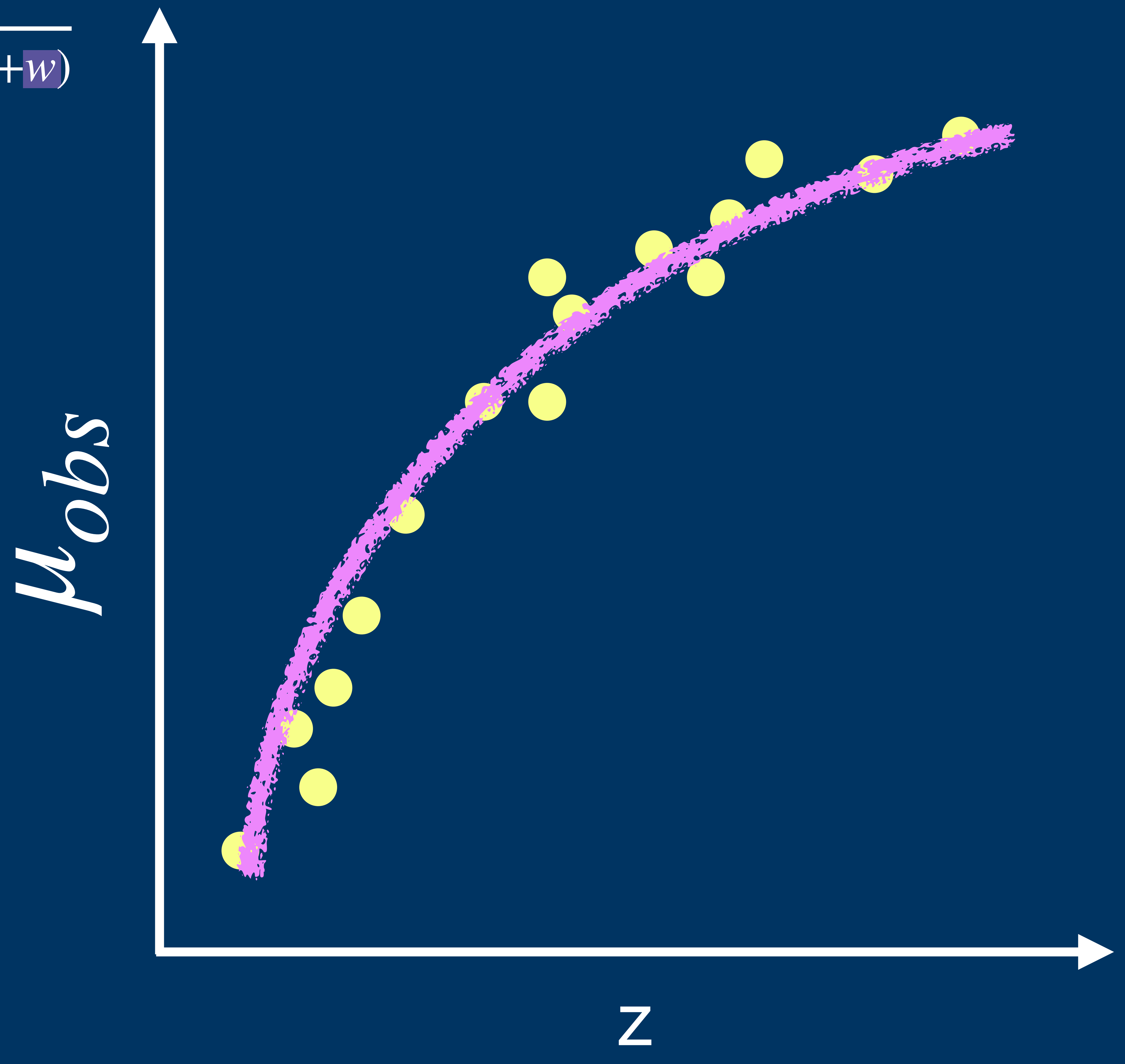
$$E(z) = \sqrt{\Omega_m (1+z')^3 + \Omega_k (1+z')^2 + \Omega_\Lambda (1+z')^{3(1+w)}}$$

luminosity distance

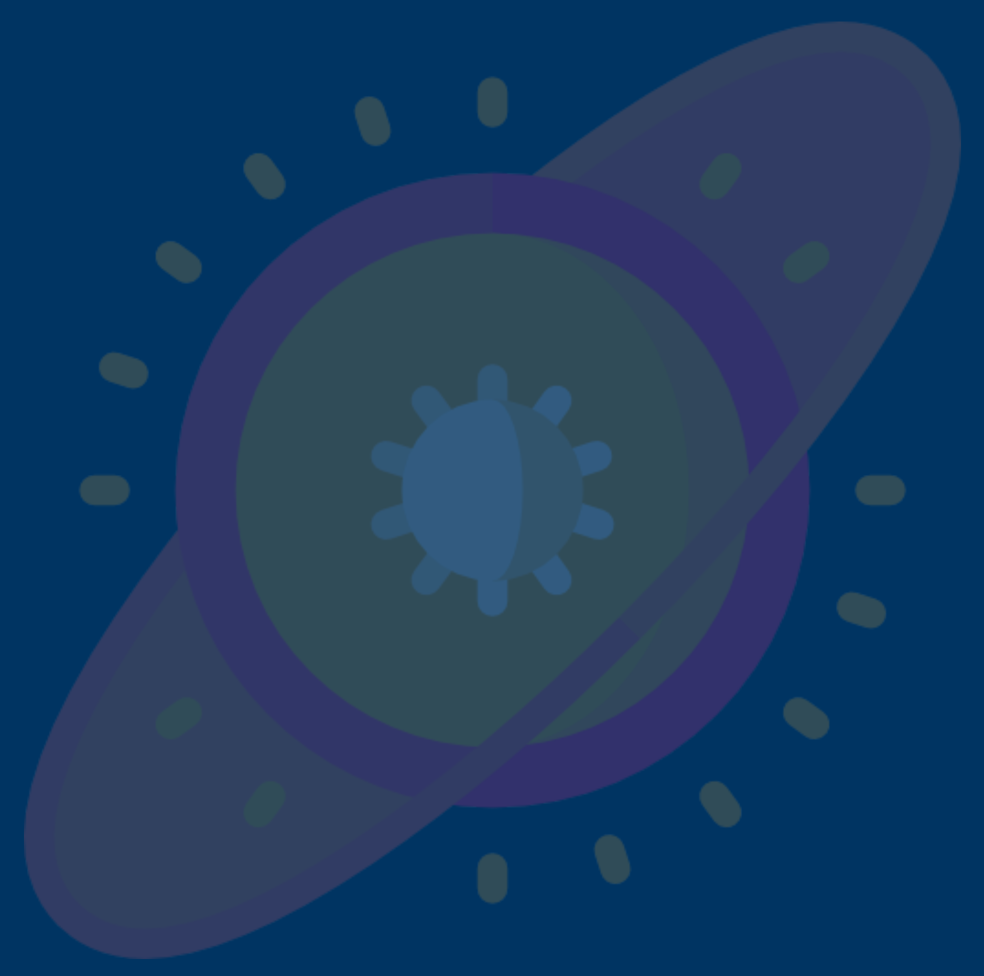
$$d_L = (1+z)\frac{c}{H_0}\int_0^z \frac{dz'}{E(z')}$$

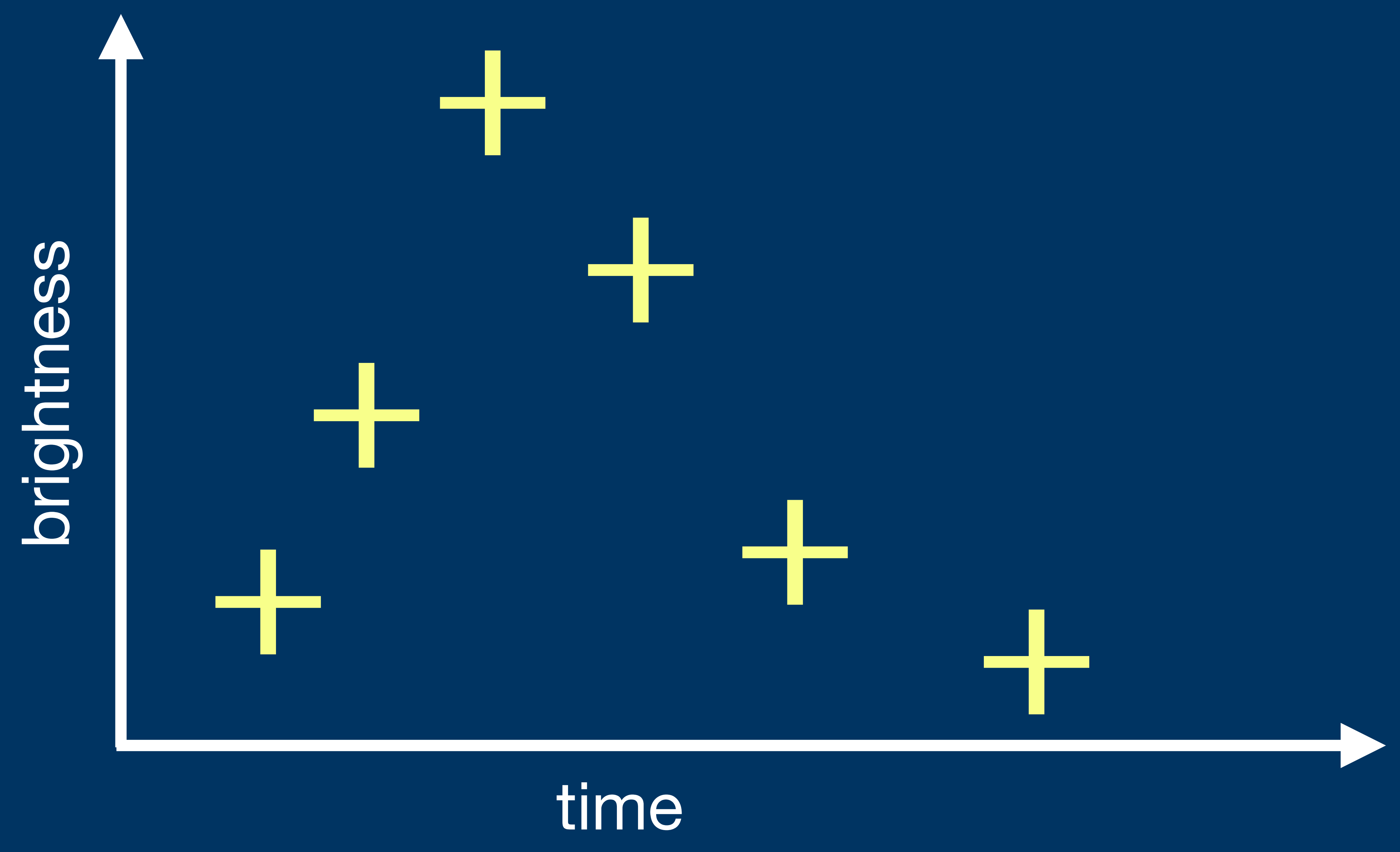
distance modulus

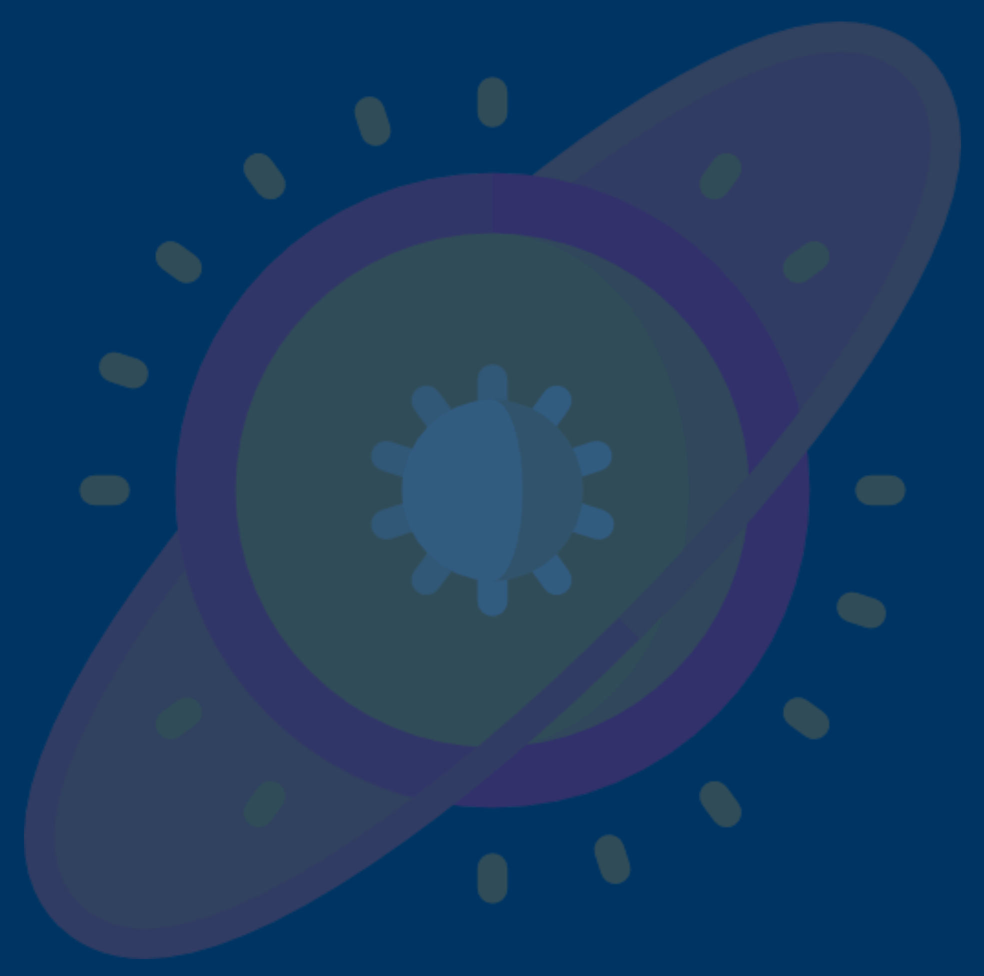
$$\mu_C = 5\log\left[\frac{d_L}{10\text{pc}}\right]$$

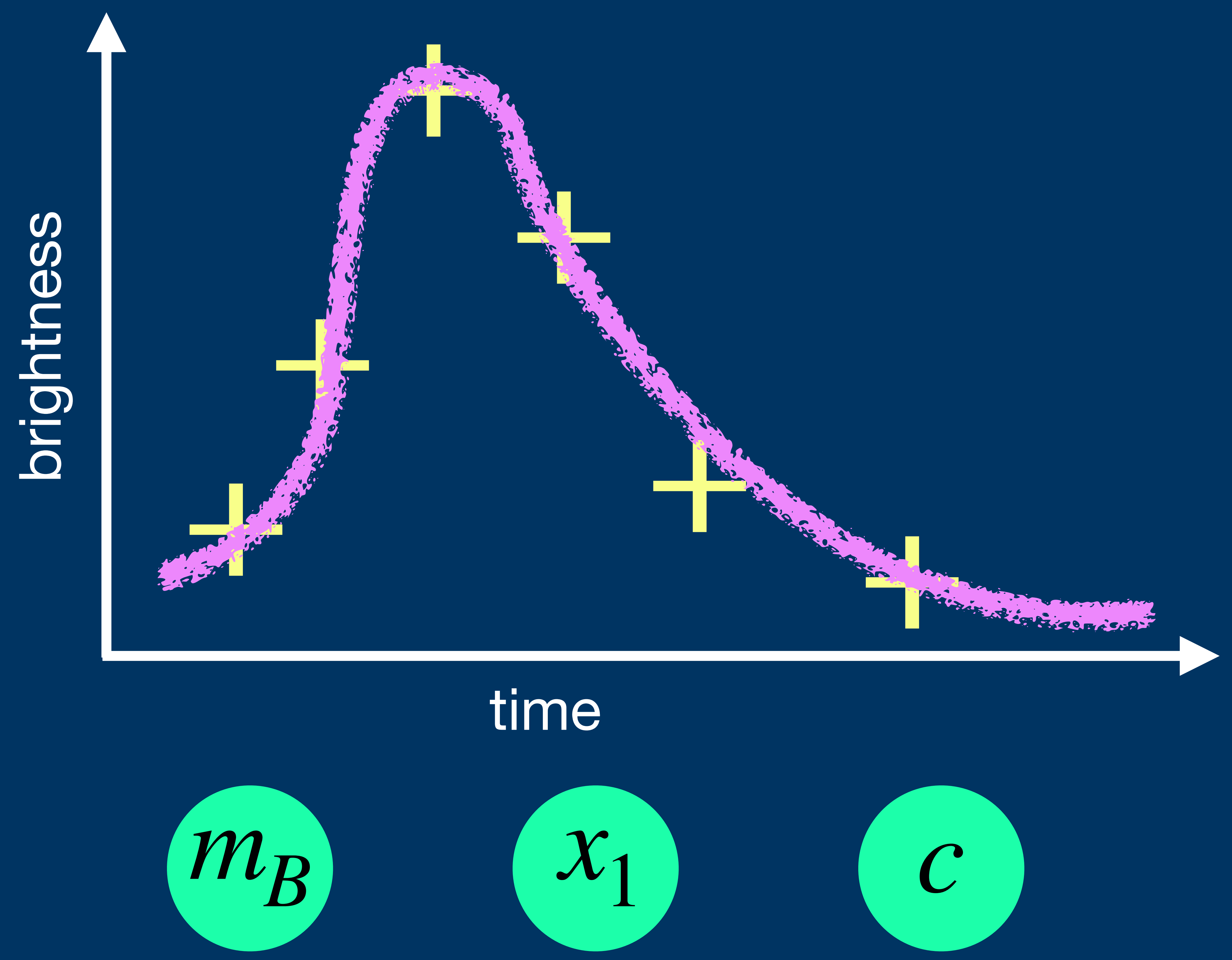


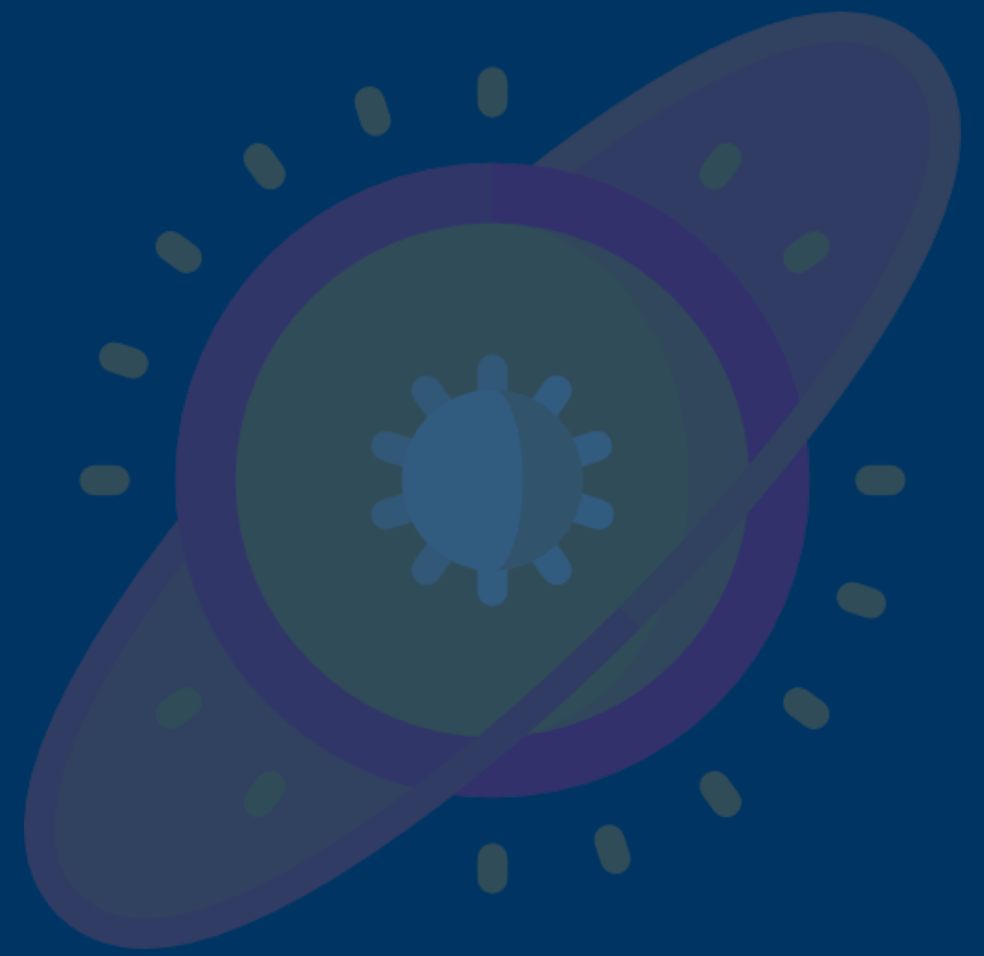

brightness
time

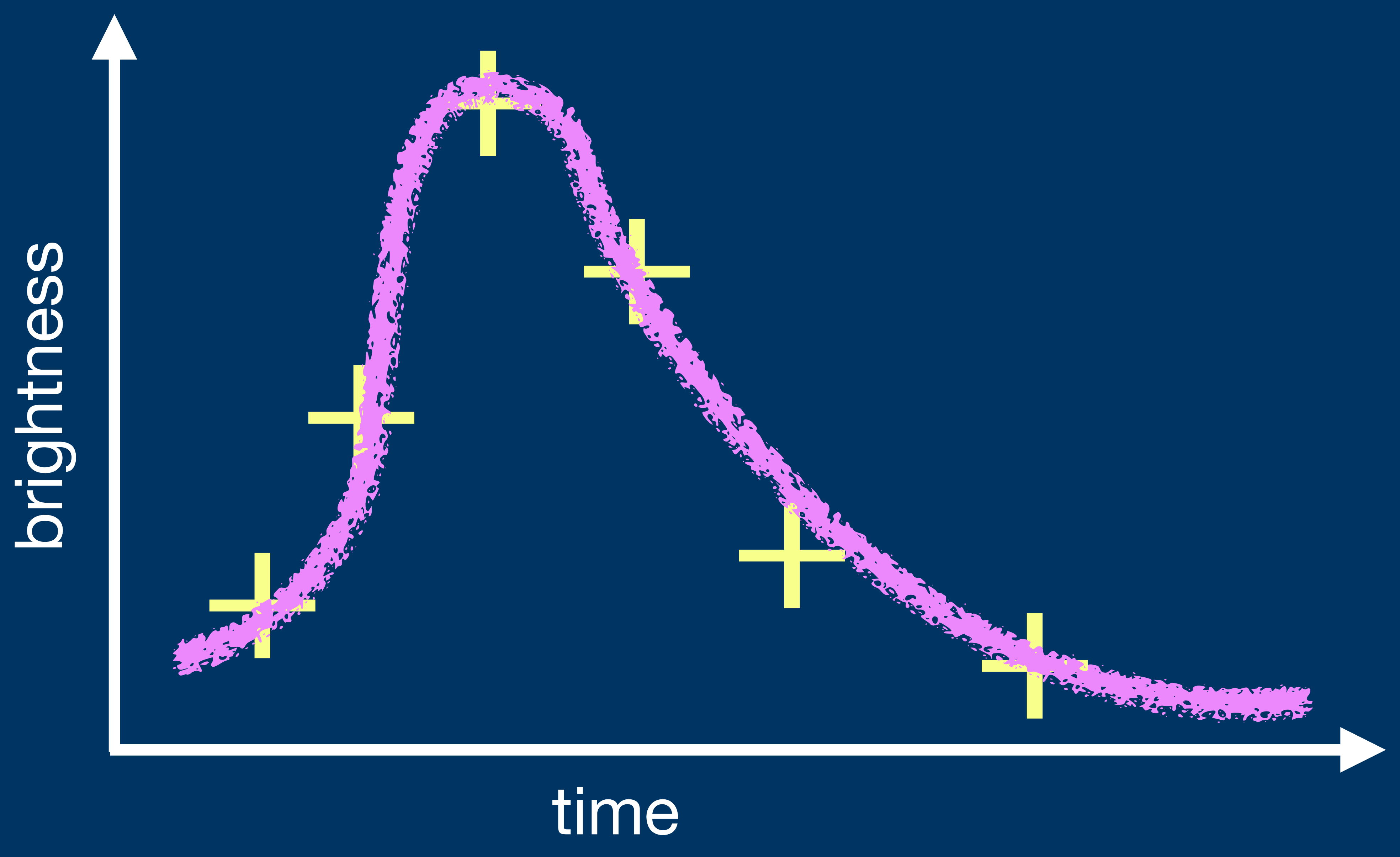



brightness
time




brightness
time
m_B
x_1
c



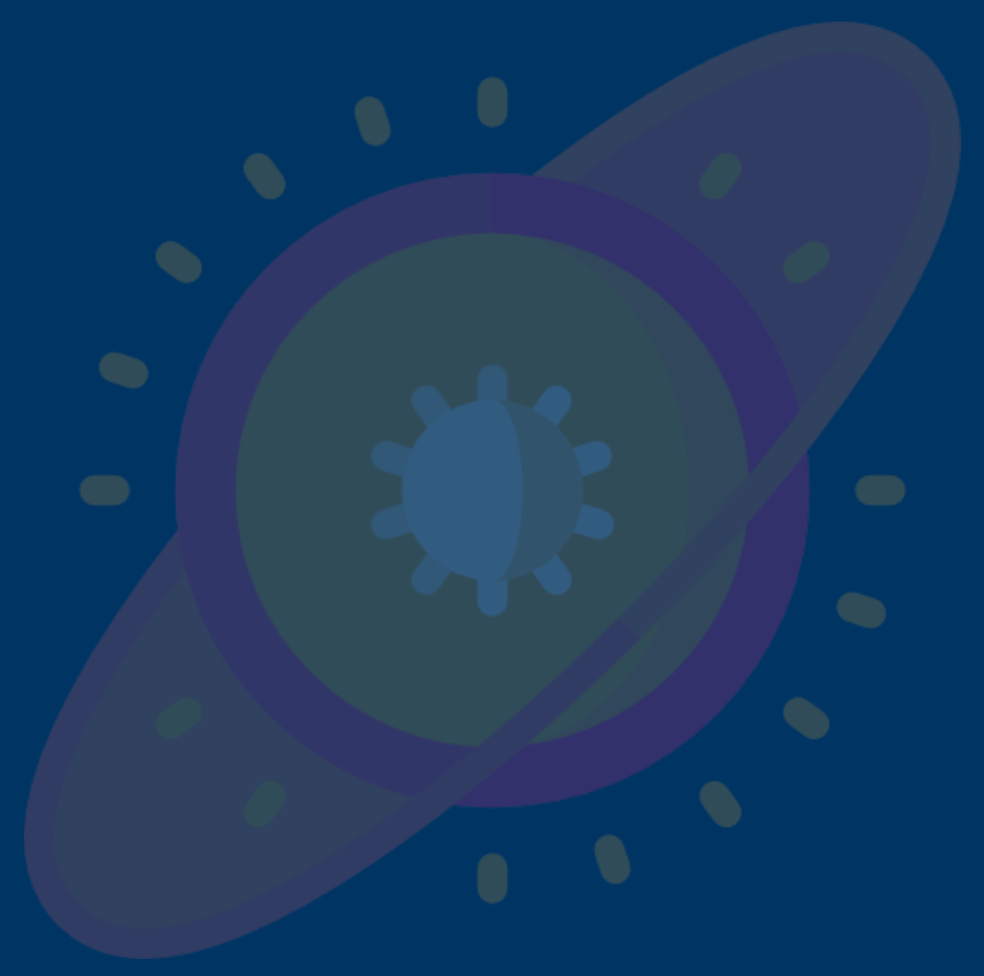


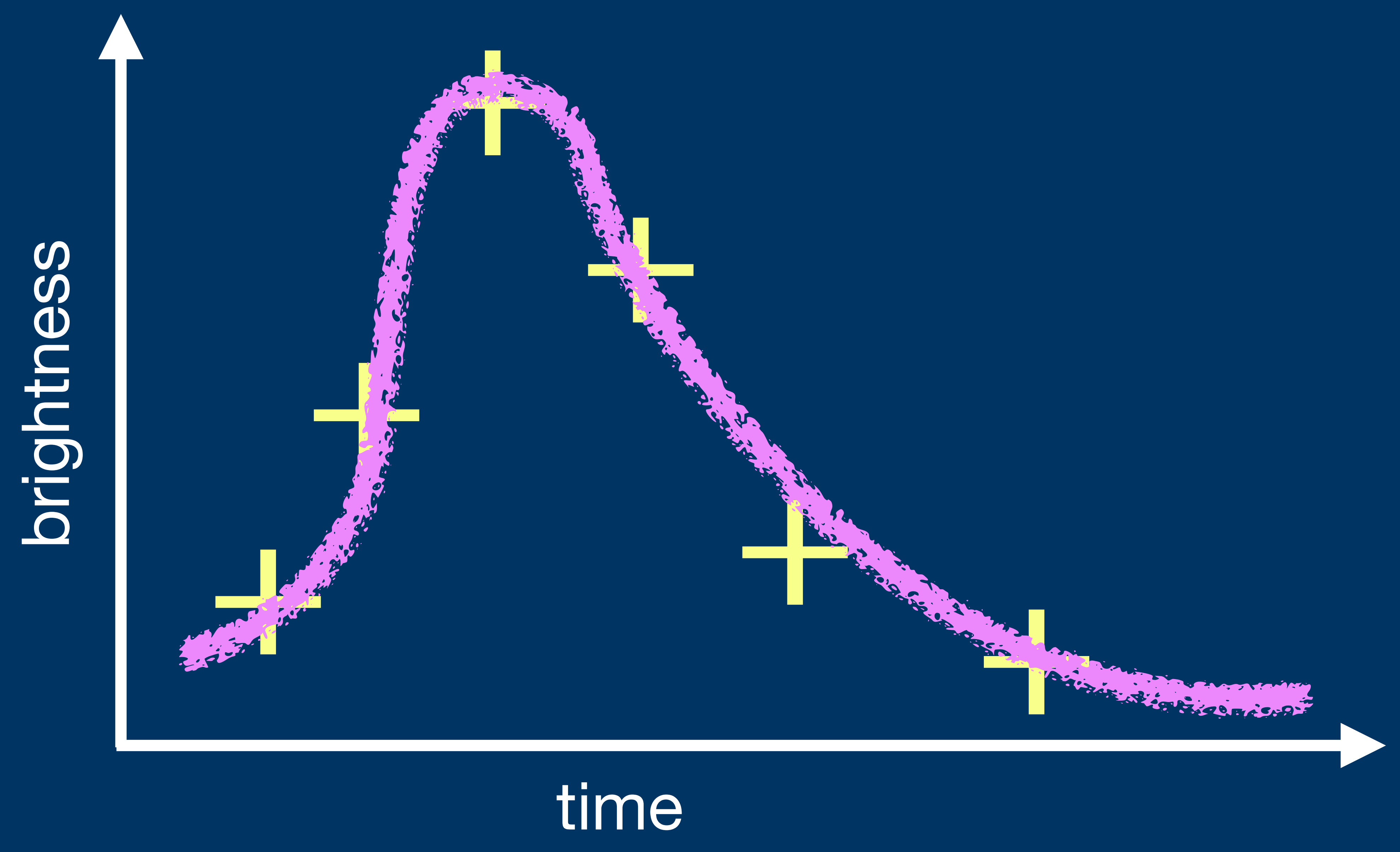
m_B x_1 c

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

average brightness

apparent brightness





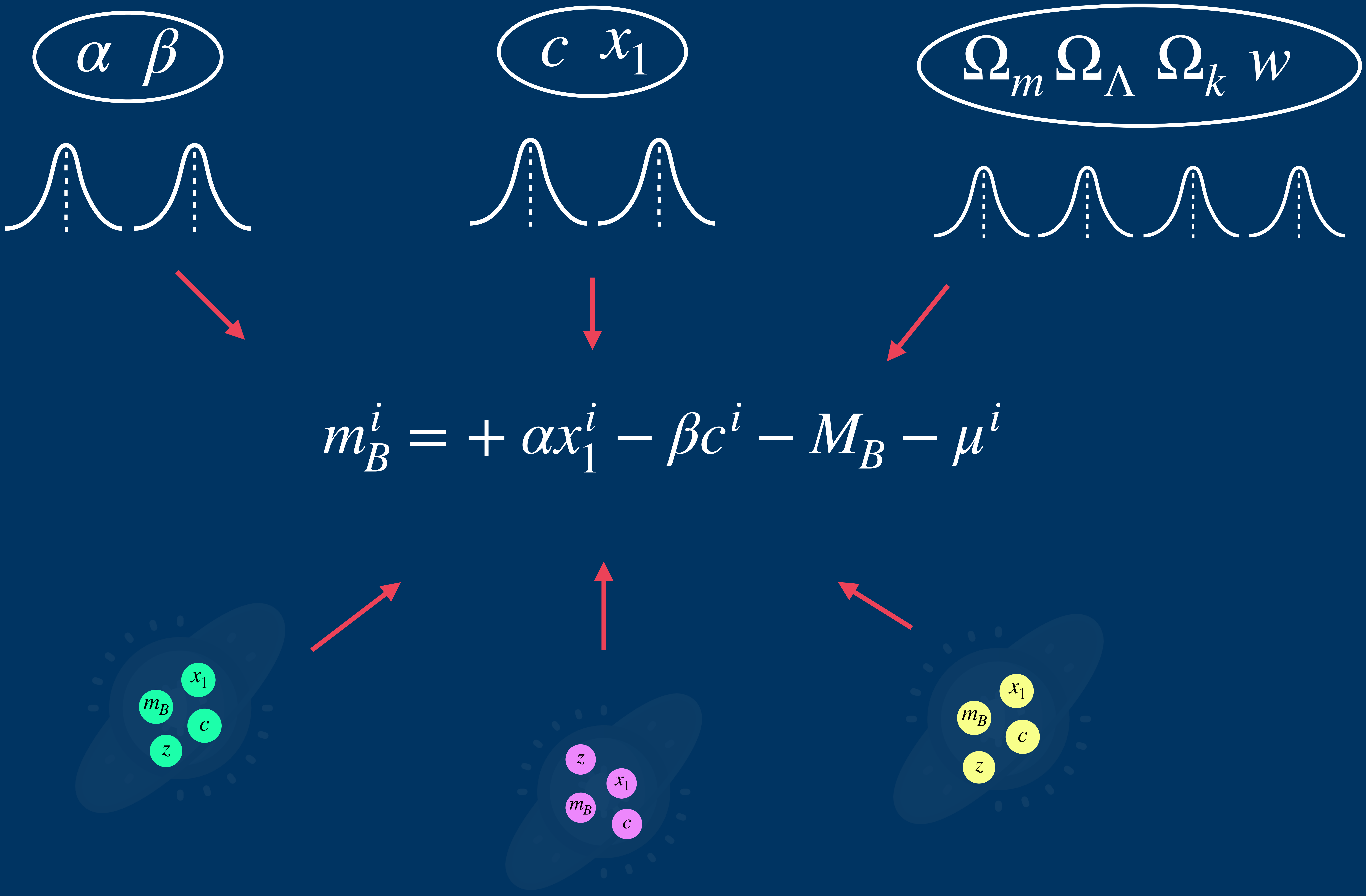
m_B x_1 c

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

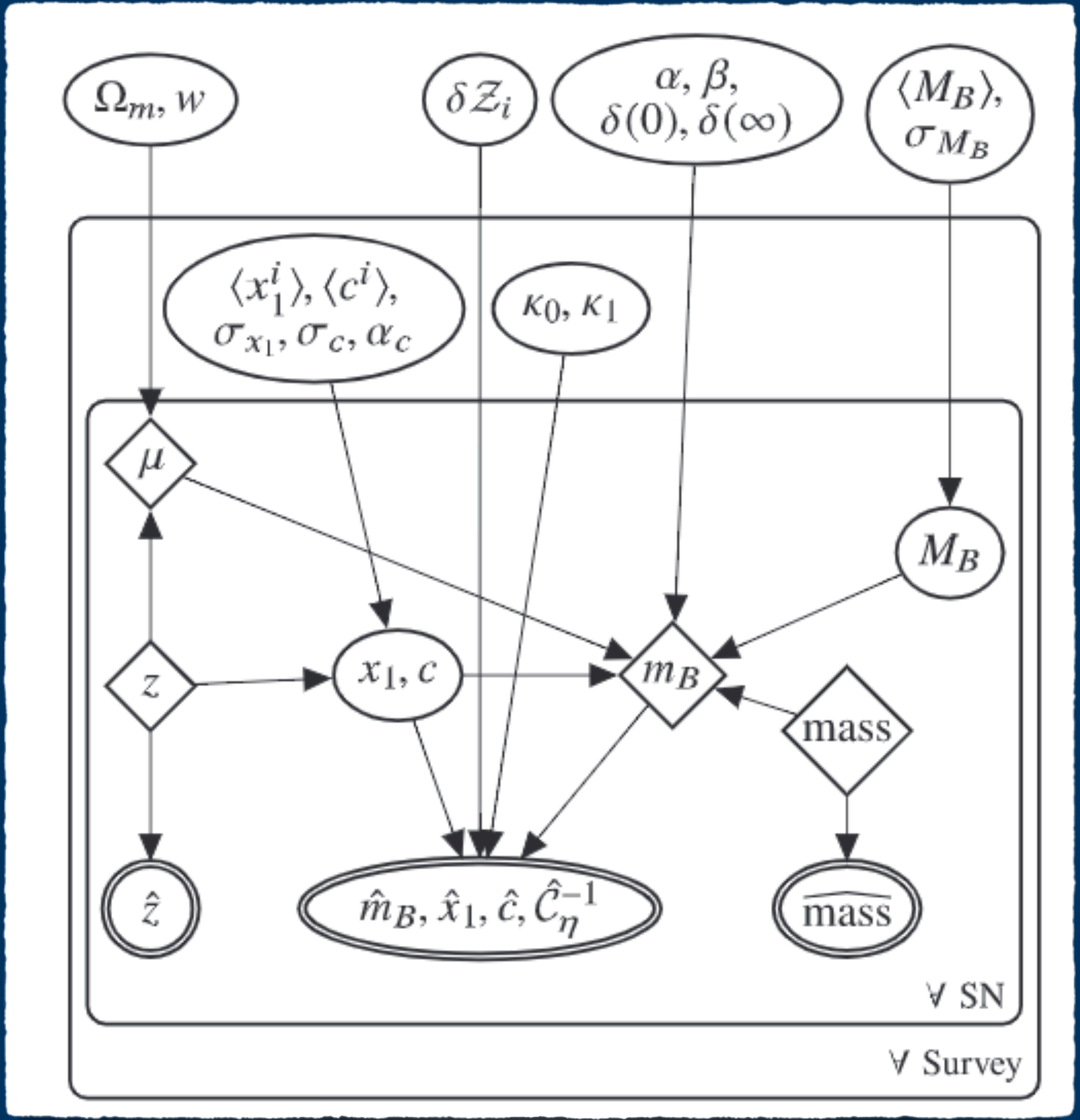
average brightness

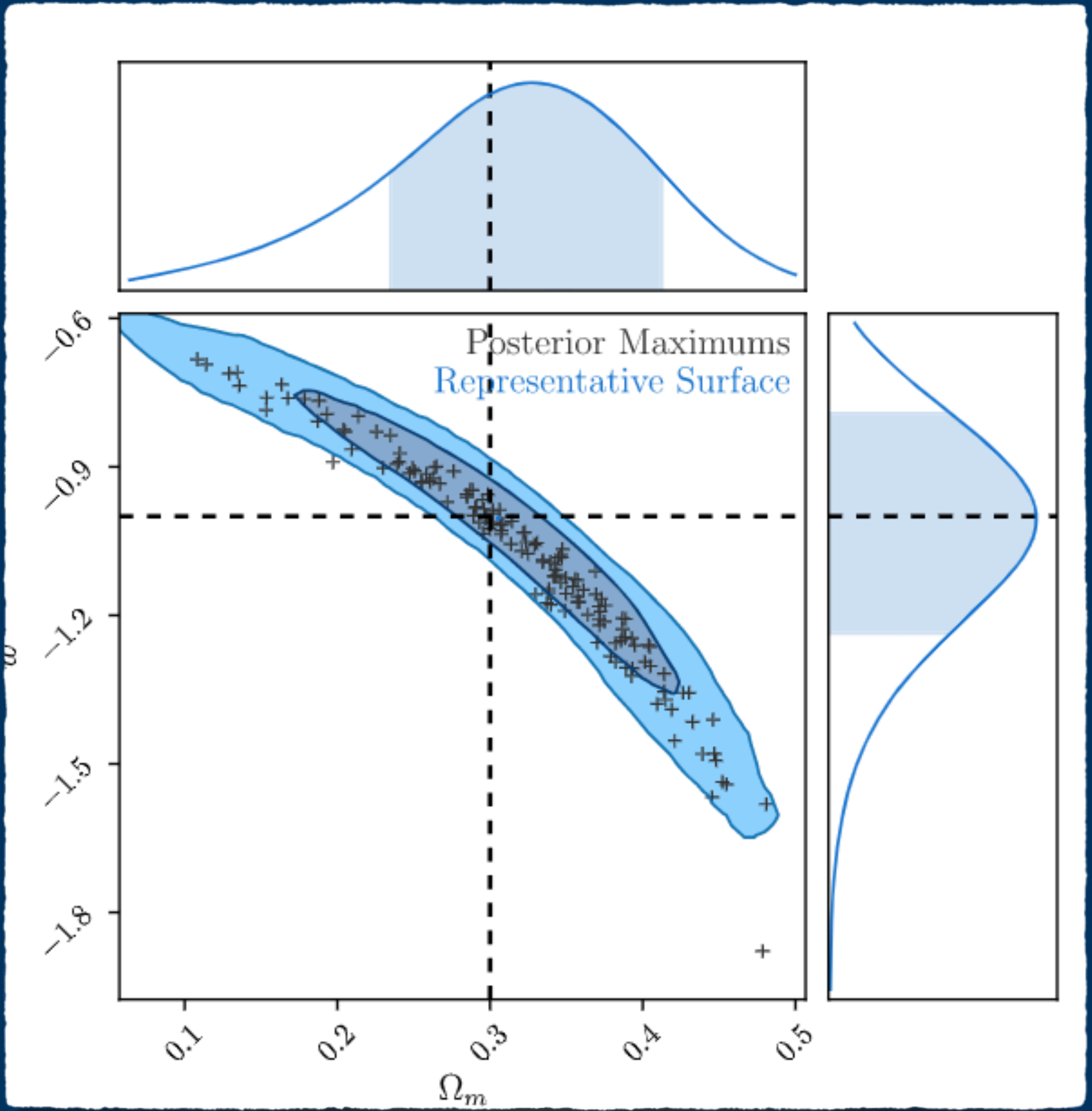
apparent brightness

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



$\alpha \quad \beta$
$c \quad x_1$
$\Omega_m \, \Omega_\Lambda \, \Omega_k \, w$
$$m_B^i = +\alpha x_1^i - \beta c^i - M_B - \mu^i$$
x_1 m_B c z
z x_1 m_B c
x_1 m_B c z



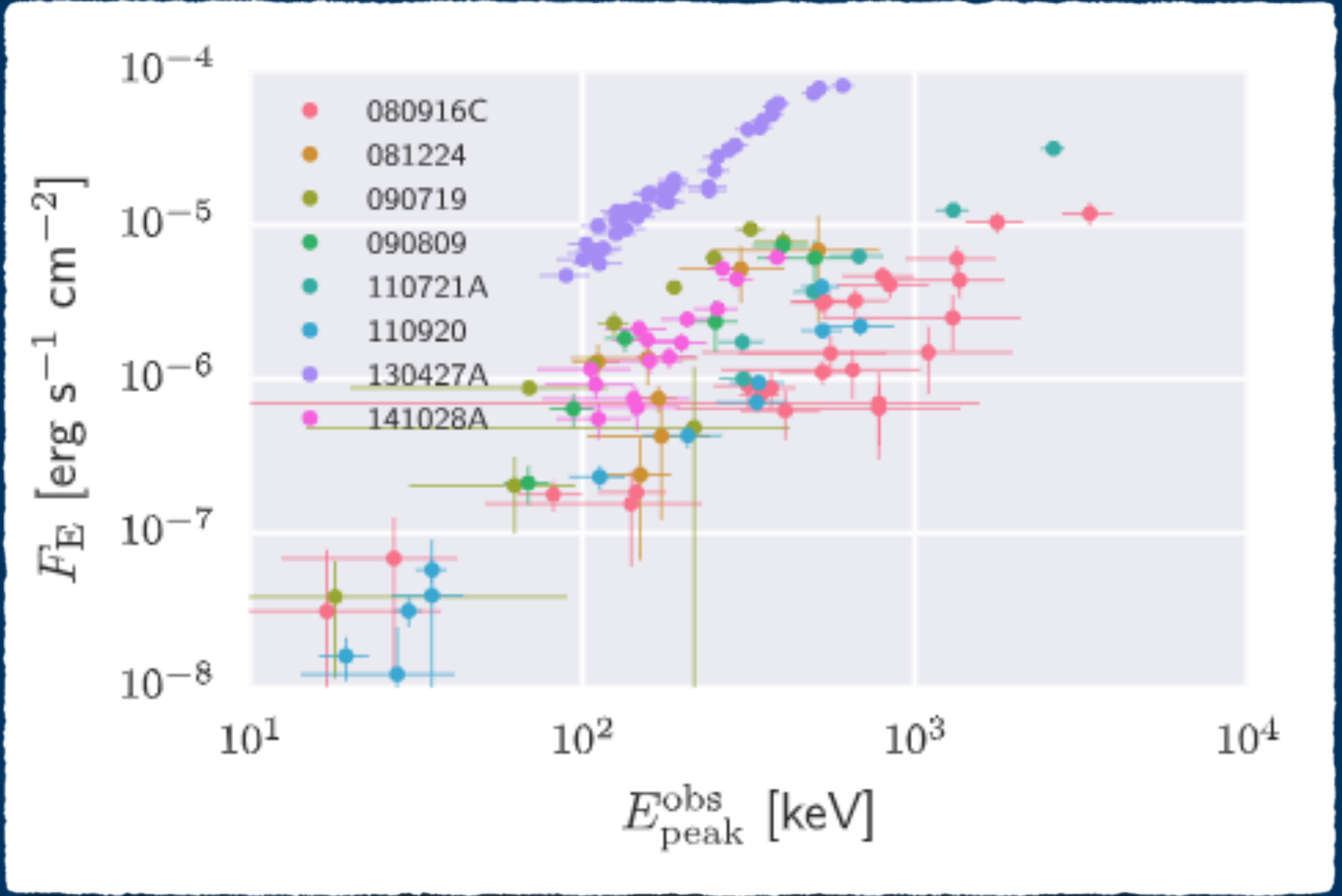


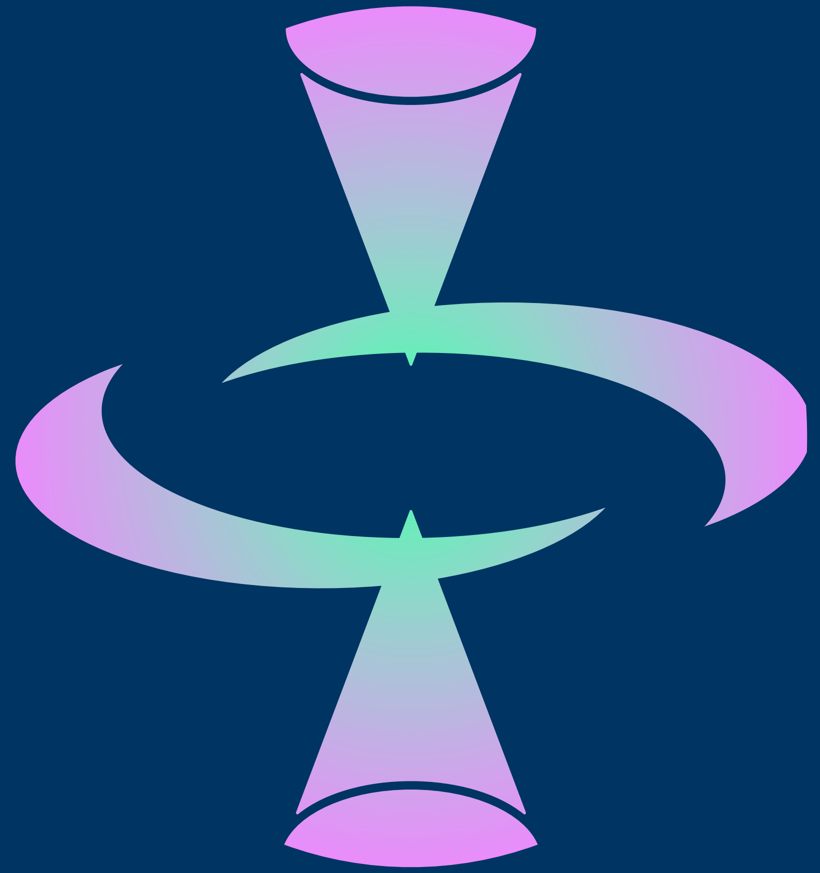
Hinton et al. 2019


Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY
MNRAS **485**, 1262–1277 (2019)
Advance Access publication 2017 May 15
doi:10.1093/mnras/stx1159

The rest-frame Golenetskii correlation via a hierarchical Bayesian analysis

J. Michael Burgess[1,2★]
[1]*The Oskar Klein Centre for Cosmoparticle Physics, SE-106 91 Stockholm, Sweden*
[2]*Department of Physics, KTH Royal Institute of Technology, AlbaNova, SE-106 91 Stockholm, Sweden*





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

$$F_{\mathrm{E}} = \frac{N_{\mathrm{rest}}}{4\pi d_{\mathrm{L}}^{2}(z)} \left(\frac{E_{\mathrm{peak}}^{\mathrm{obs}}\,(1+z)}{100\mathrm{keV}} \right)^{\gamma} \mathrm{erg\ s^{-1}\ cm^{-2}}$$

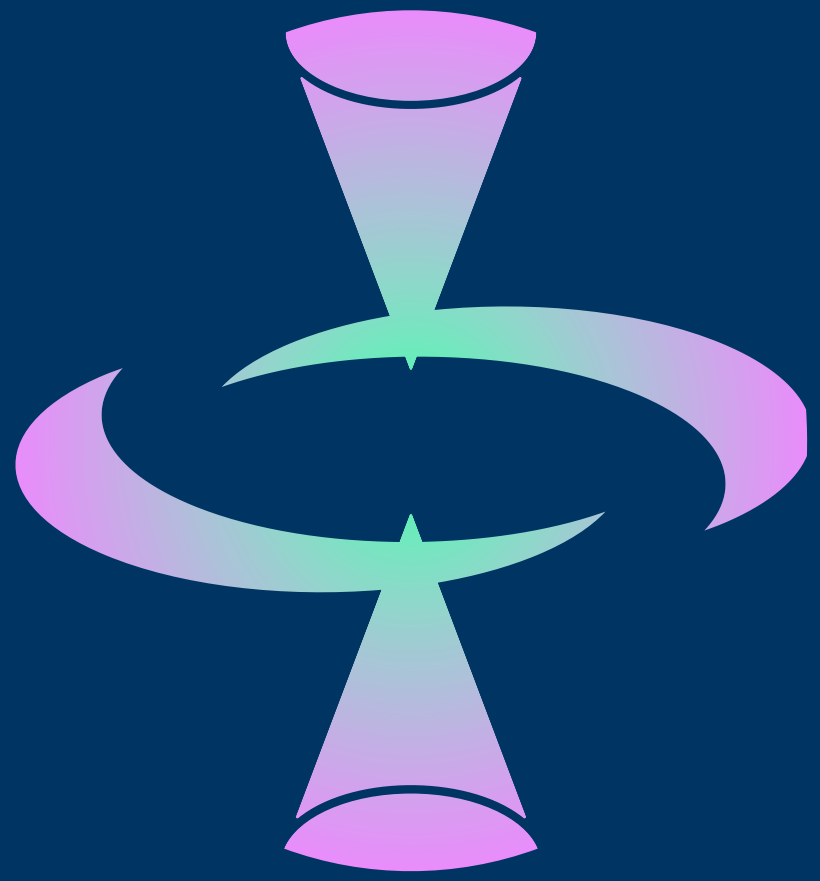
Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

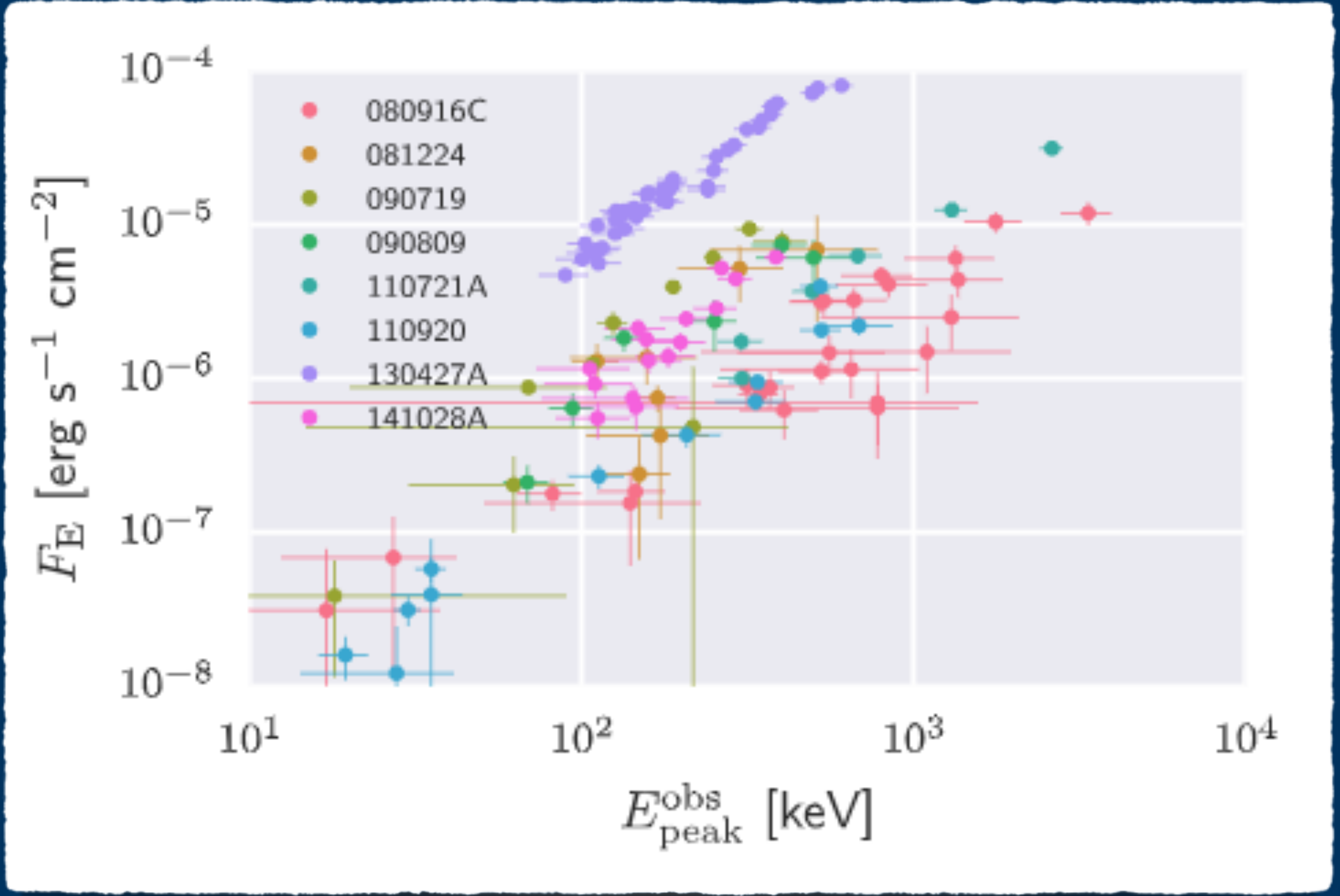
MNRAS **485**, 1262–1277 (2019) doi:10.1093/mnras/stx1159
Advance Access publication 2017 May 15

The rest-frame Golenetskii correlation via a hierarchical Bayesian analysis

J. Michael Burgess[1,2]★

[1]*The Oskar Klein Centre for Cosmoparticle Physics, SE-106 91 Stockholm, Sweden*
[2]*Department of Physics, KTH Royal Institute of Technology, AlbaNova, SE-106 91 Stockholm, Sweden*





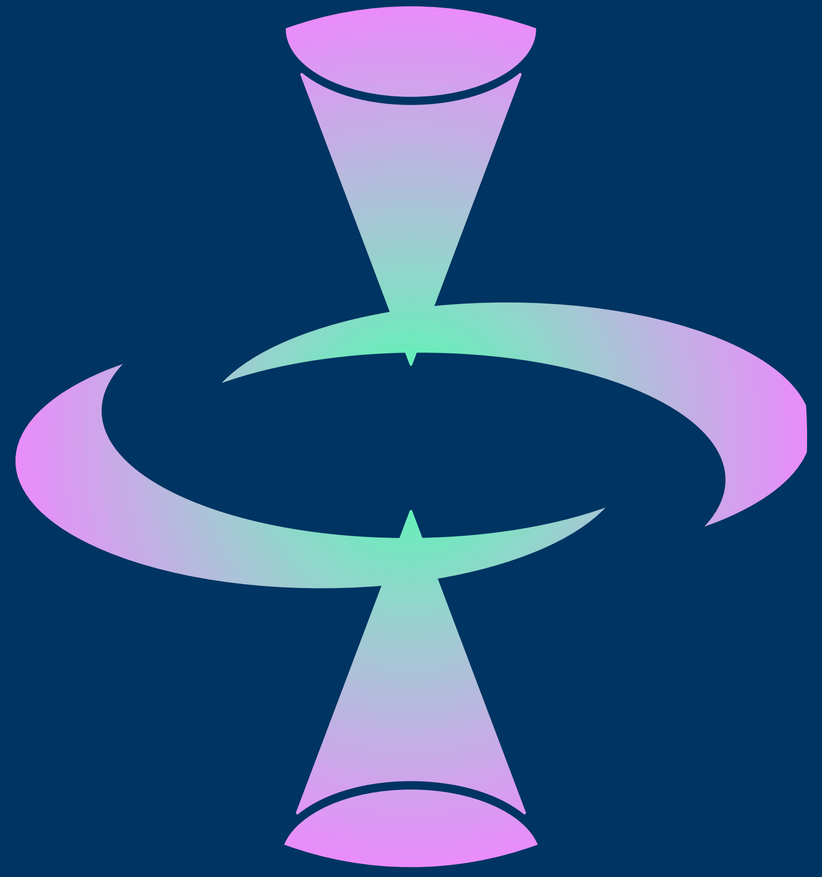
$$L = N_{\mathrm{rest}} \left(\frac{E_{\mathrm{peak}}^{\mathrm{rest}}}{100\mathrm{keV}} \right)^{\gamma} \ \mathrm{erg\ s^{-1}}$$

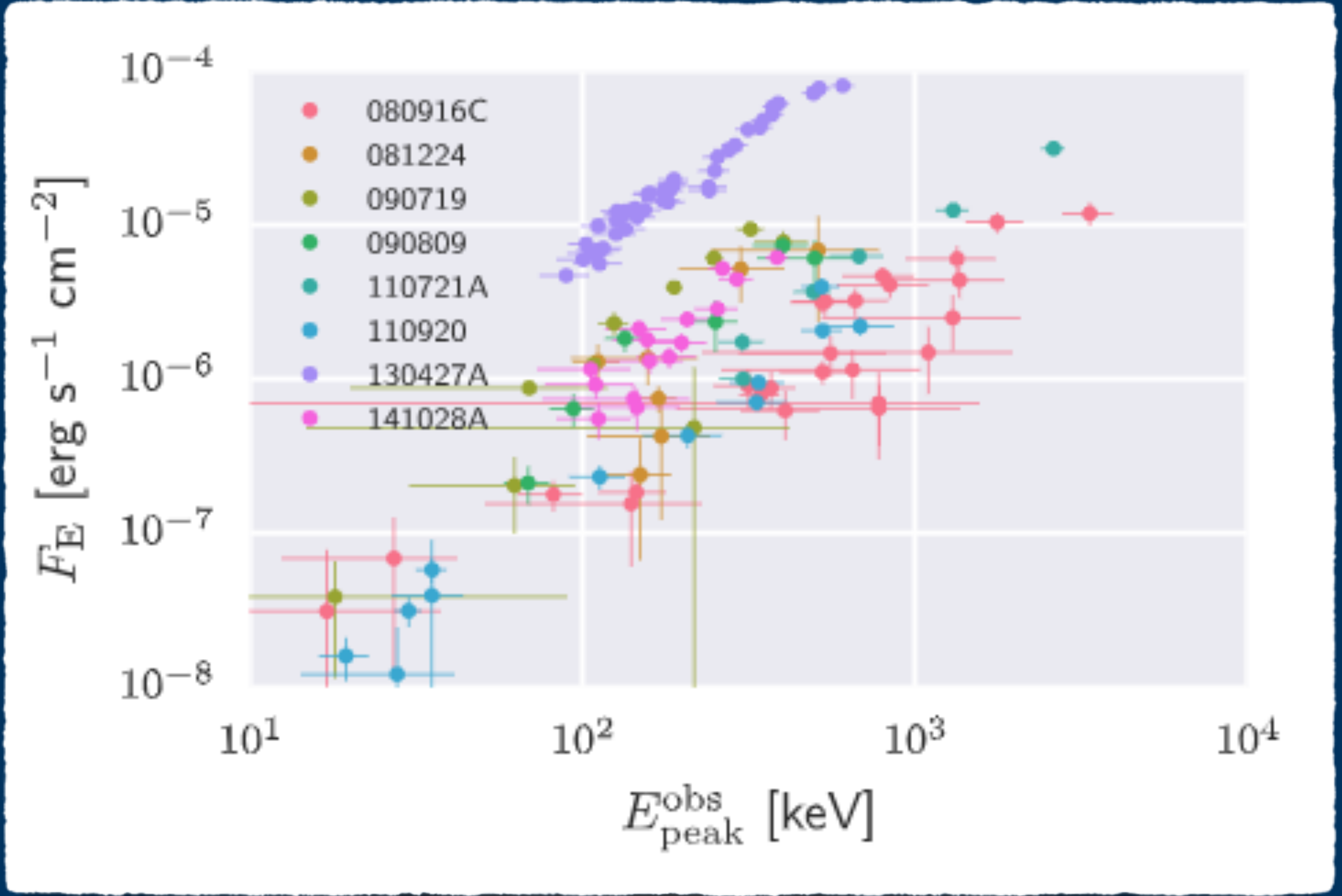
$$F_{\mathrm{E}} = \frac{N_{\mathrm{rest}}}{4\pi d_{\mathrm{L}}^{2}(z)} \left(\frac{E_{\mathrm{peak}}^{\mathrm{obs}}\,(1+z)}{100\mathrm{keV}} \right)^{\gamma} \ \mathrm{erg\ s^{-1}\ cm^{-2}}$$

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY
MNRAS **485**, 1262–1277 (2019)
Advance Access publication 2017 May 15
doi:10.1093/mnras/stx1159

The rest-frame Golenetskii correlation via a hierarchical Bayesian analysis

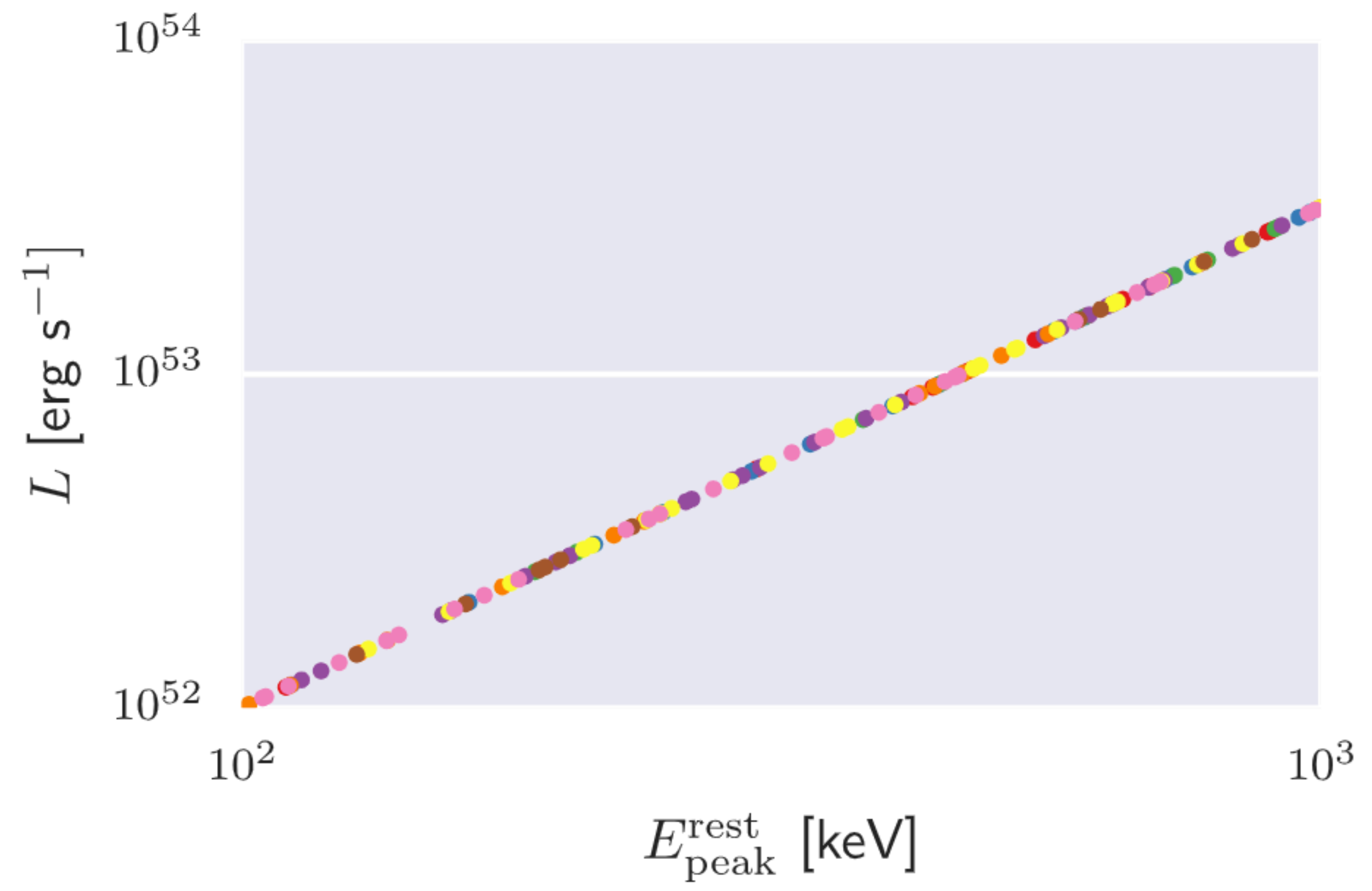
J. Michael Burgess[1,2]★
[1]The Oskar Klein Centre for Cosmoparticle Physics, SE-106 91 Stockholm, Sweden
[2]Department of Physics, KTH Royal Institute of Technology, AlbaNova, SE-106 91 Stockholm, Sweden

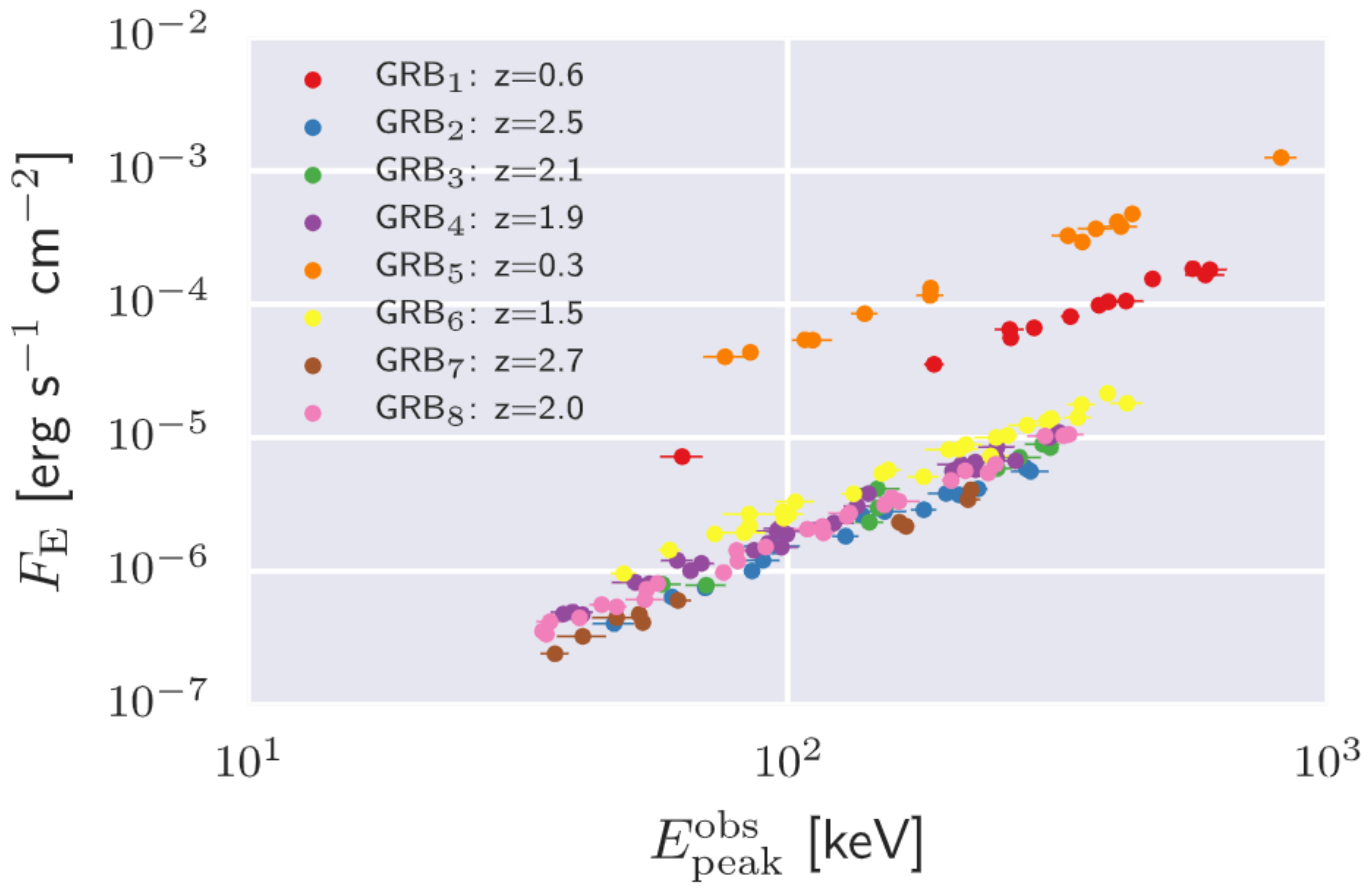




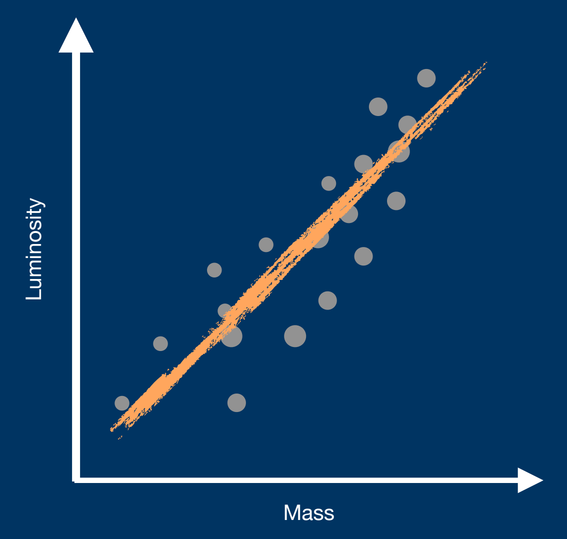
$$L = N_{\mathrm{rest}} \left(\frac{E_{\mathrm{peak}}^{\mathrm{rest}}}{100\mathrm{keV}} \right)^{\gamma} \ \mathrm{erg\ s^{-1}}$$

$$F_{\mathrm{E}} = \frac{N_{\mathrm{rest}}}{4\pi d_{\mathrm{L}}^{2}(z)} \left(\frac{E_{\mathrm{peak}}^{\mathrm{obs}}(1+z)}{100\mathrm{keV}} \right)^{\gamma} \ \mathrm{erg\ s^{-1}\ cm^{-2}}$$

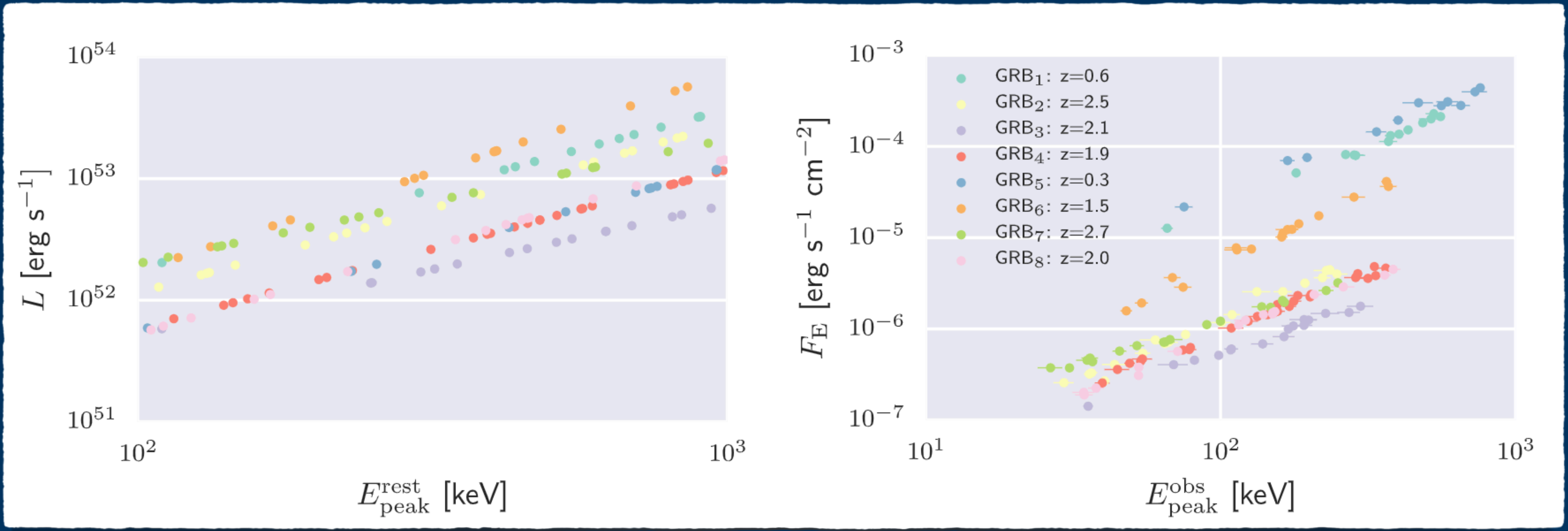




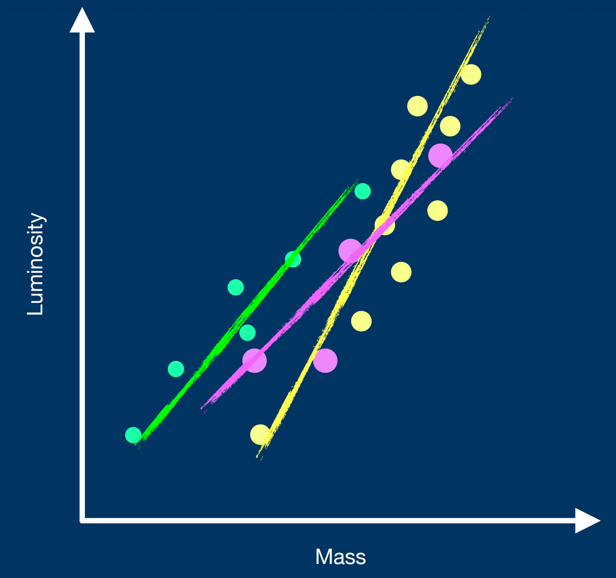
Rest Frame



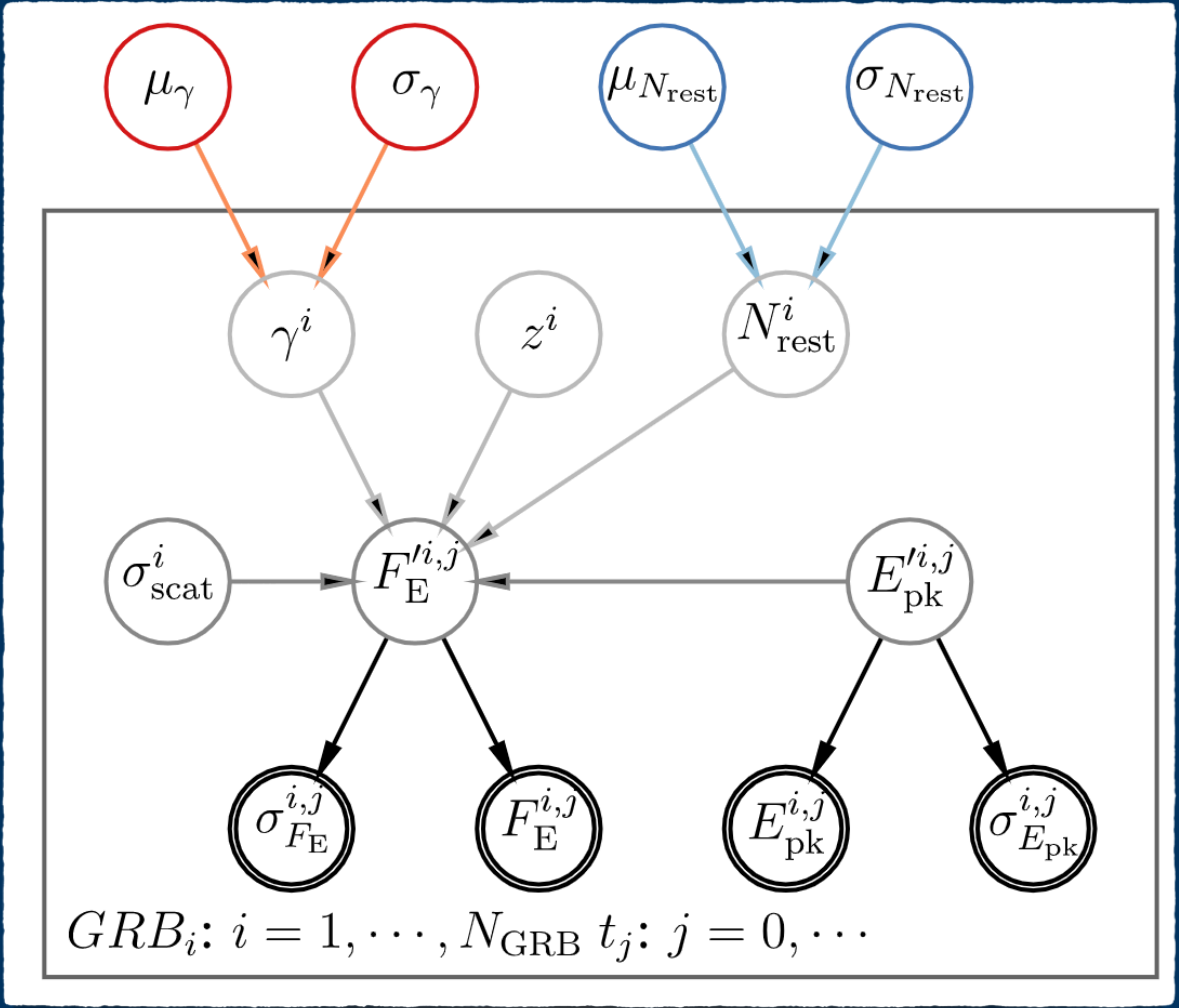
Observer Frame

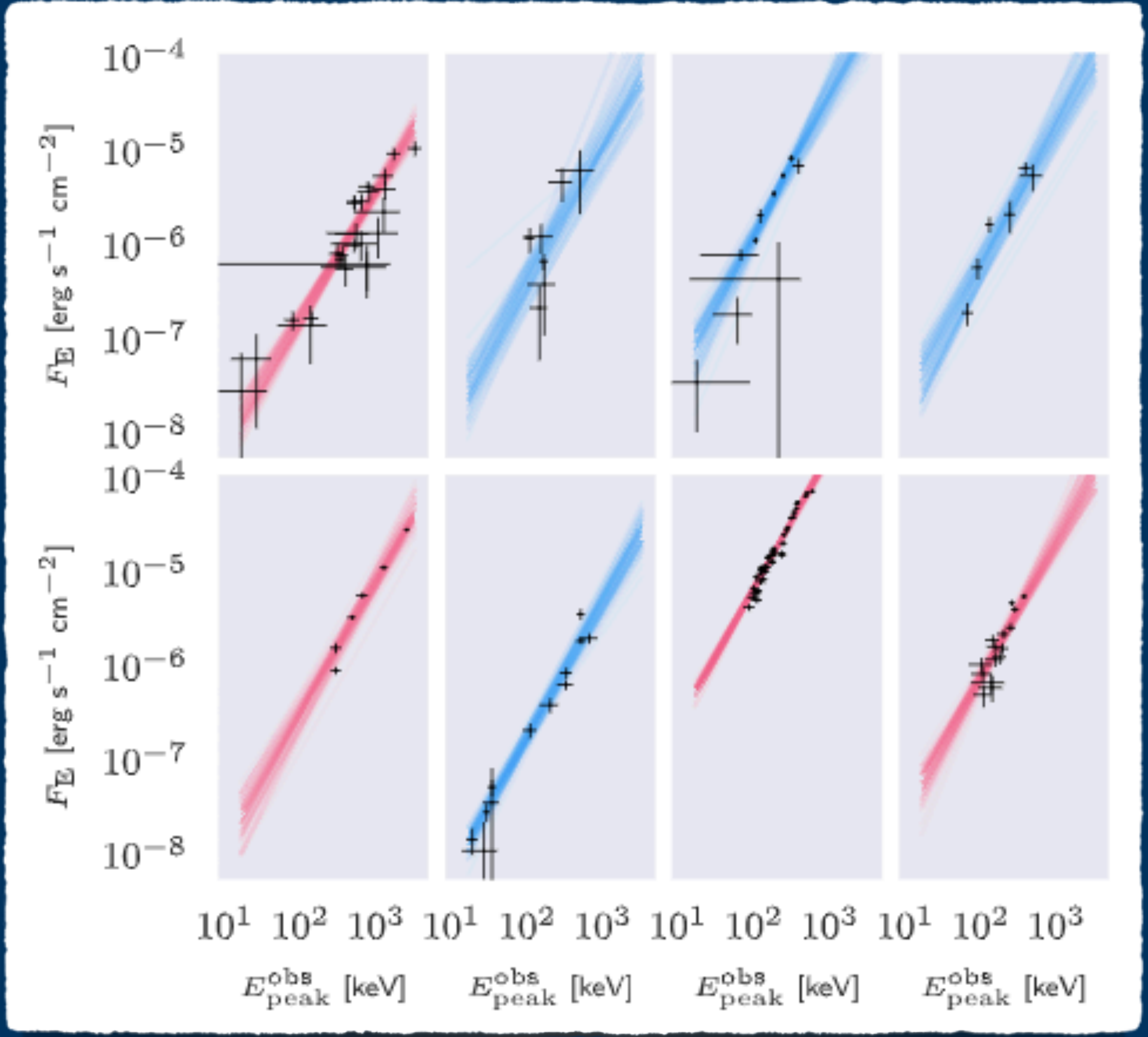


Rest Frame

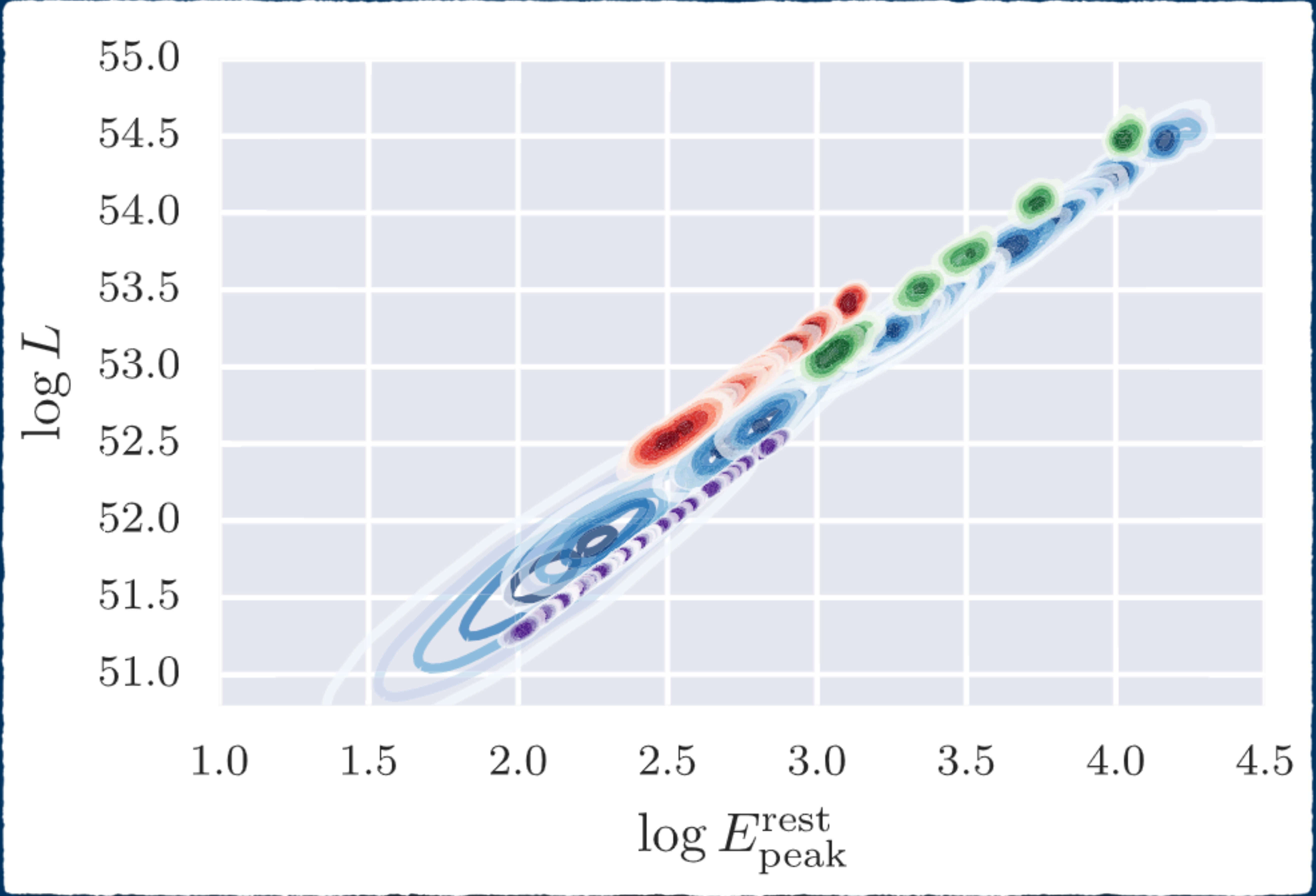


Observer Frame

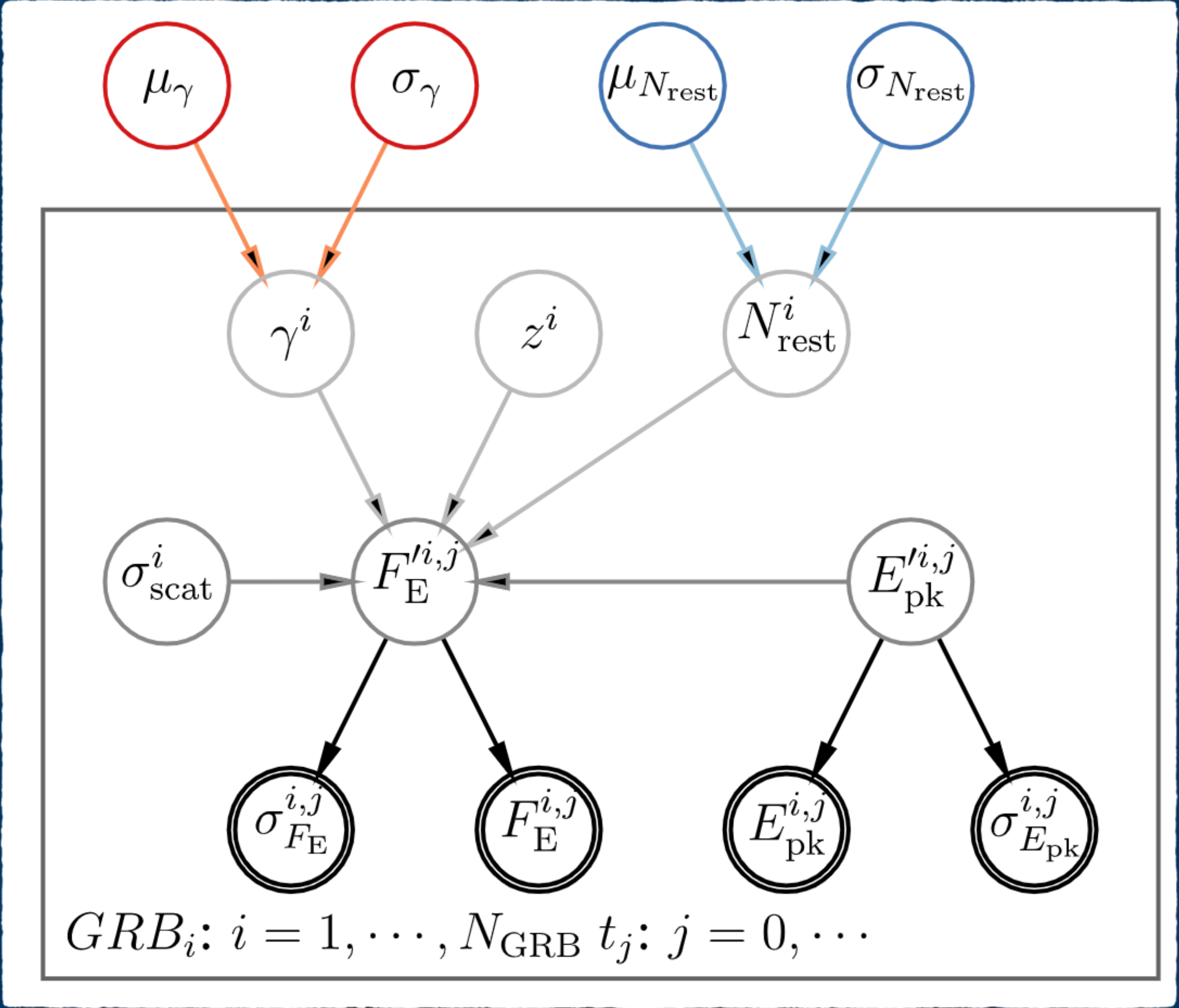

μ_γ
σ_γ
$\mu_{N_{\rm rest}}$
$\sigma_{N_{\rm rest}}$
γ^i
z^i
$N^i_{\rm rest}$
$\sigma^i_{\rm scat}$
$F'^{i,j}_{\rm E}$
$E'^{i,j}_{\rm pk}$
$\sigma^{i,j}_{F_{\rm E}}$
$F^{i,j}_{\rm E}$
$E^{i,j}_{\rm pk}$
$\sigma^{i,j}_{E_{\rm pk}}$
GRB_i: $i = 1, \cdots, N_{\rm GRB}$ t_j: $j = 0, \cdots$



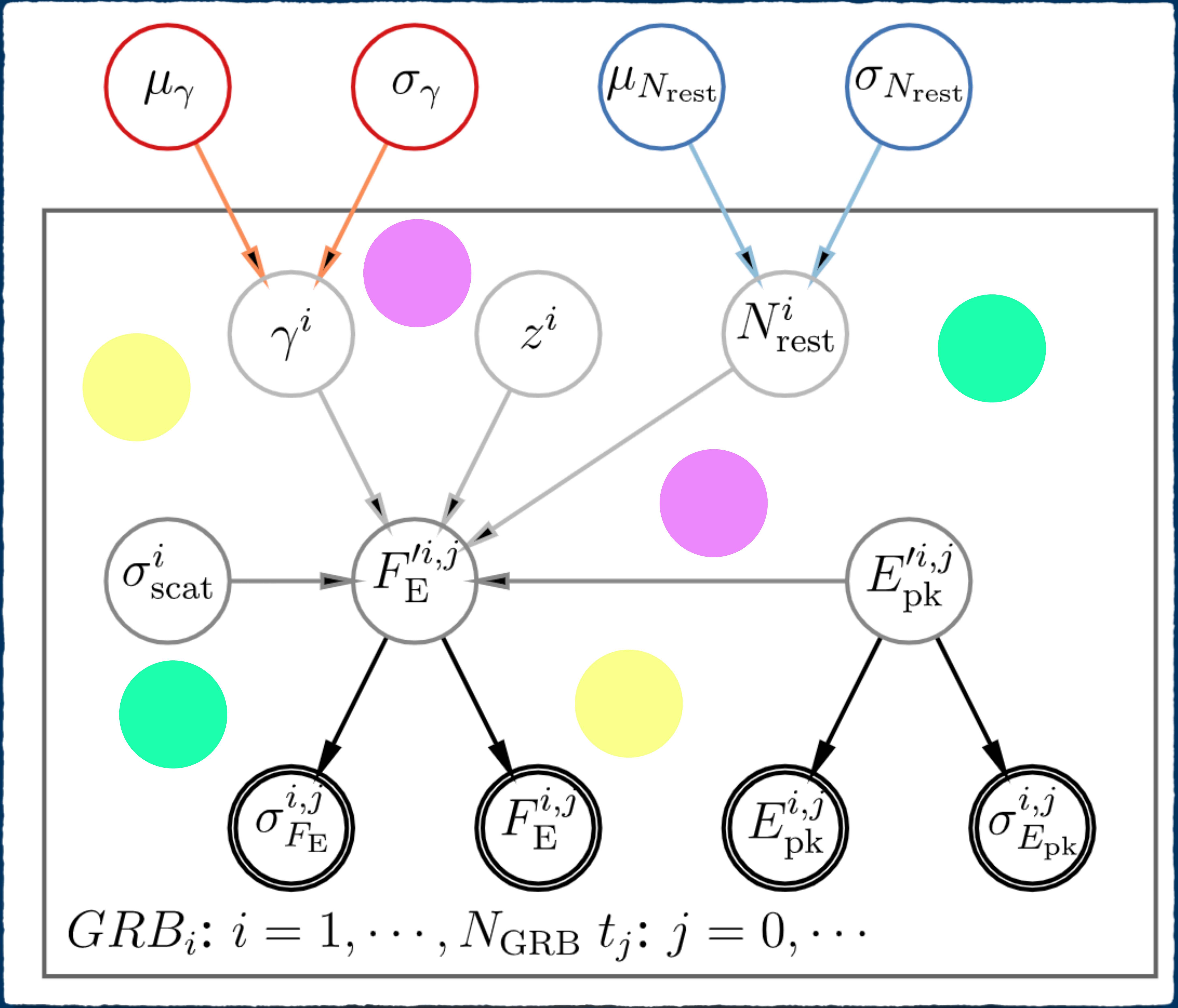
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
F_E [erg s^{-1} cm^{-2}]
10^1 10^2 10^3
$E_{\rm peak}^{\rm obs}$ [keV]



55.0
54.5
54.0
53.5
53.0
52.5
52.0
51.5
51.0
$\log L$
1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5
$\log E_{\rm peak}^{\rm rest}$

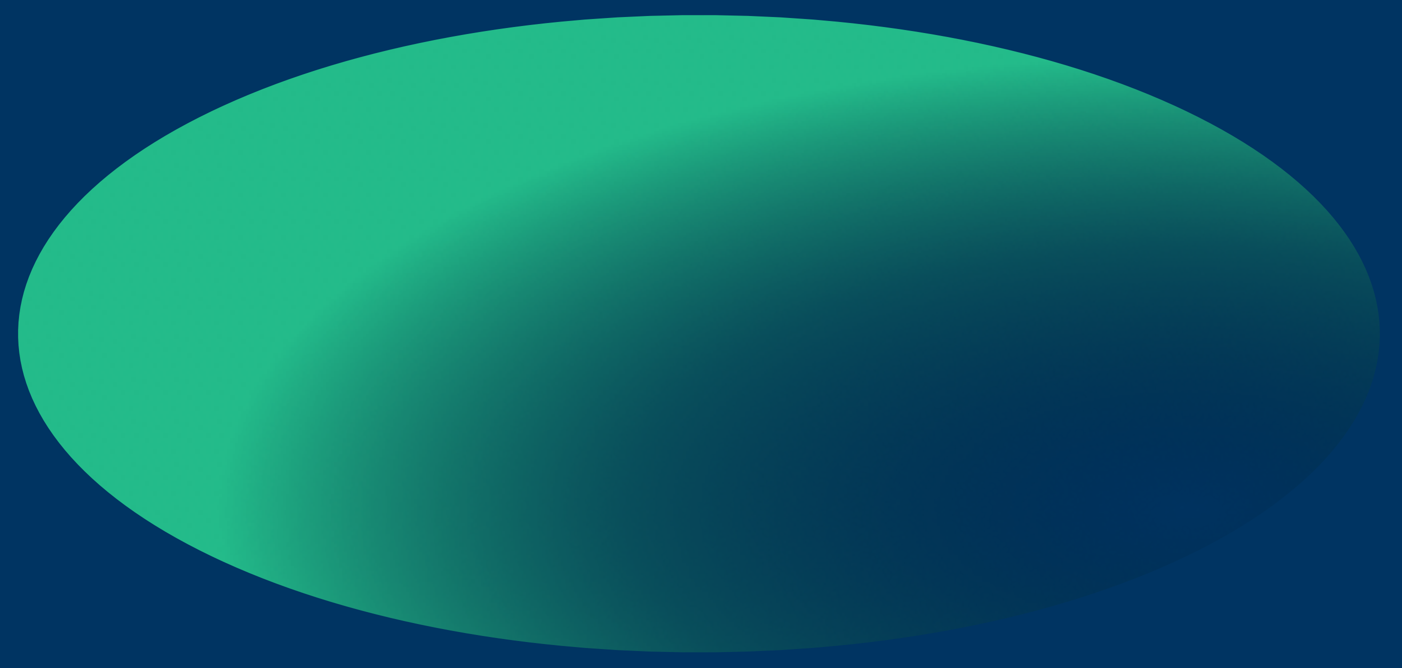


μ_γ
σ_γ
$\mu_{N_{\rm rest}}$
$\sigma_{N_{\rm rest}}$
γ^i
z^i
$N^i_{\rm rest}$
$\sigma^i_{\rm scat}$
$F'^{i,j}_{\rm E}$
$E'^{i,j}_{\rm pk}$
$\sigma^{i,j}_{F_{\rm E}}$
$F^{i,j}_{\rm E}$
$E^{i,j}_{\rm pk}$
$\sigma^{i,j}_{E_{\rm pk}}$
GRB_i: $i = 1, \cdots, N_{\rm GRB}$ t_j: $j = 0, \cdots$



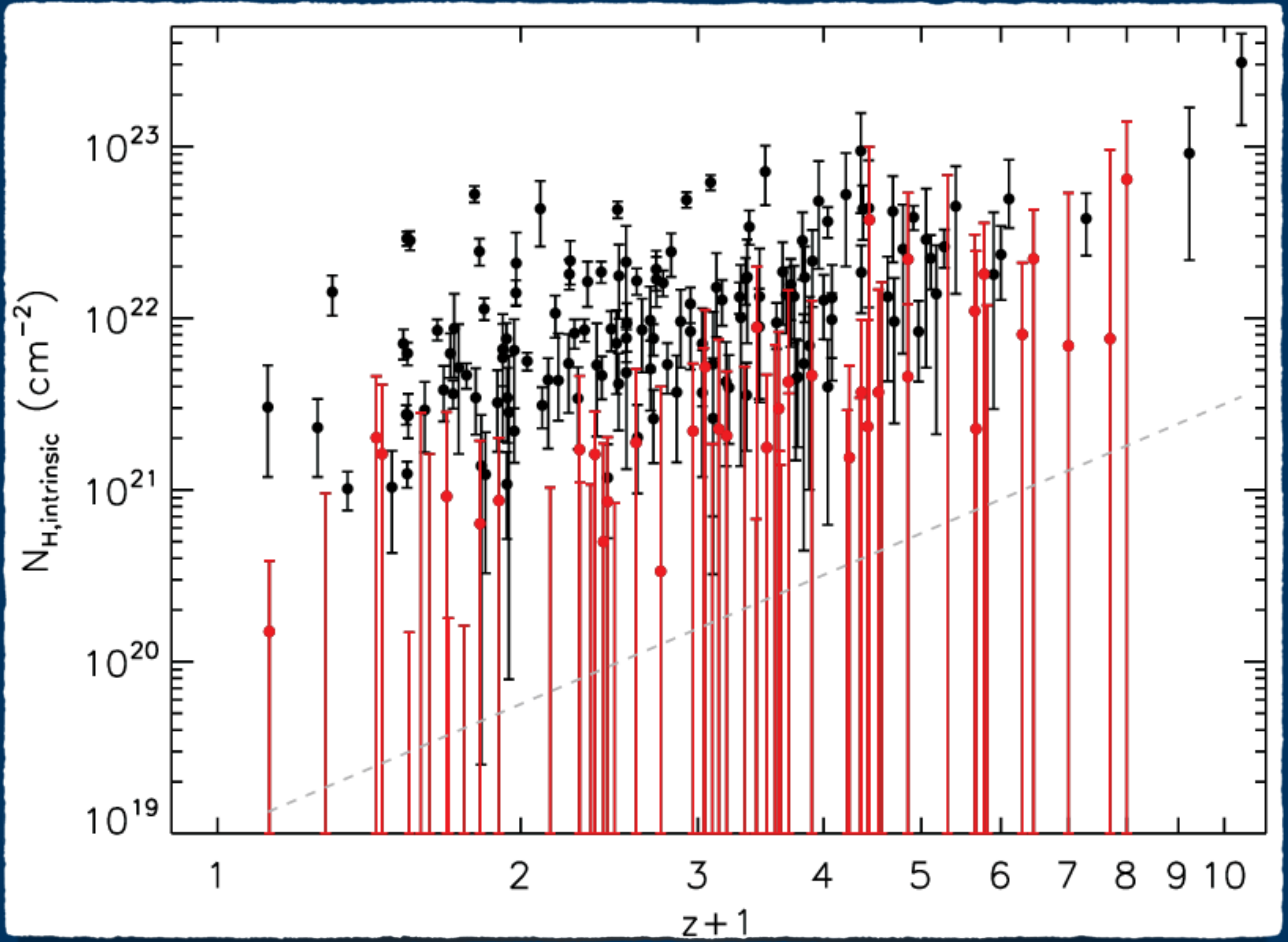
μ_γ
σ_γ
$\mu_{N_{\rm rest}}$
$\sigma_{N_{\rm rest}}$
γ^i
z^i
$N^i_{\rm rest}$
$\sigma^i_{\rm scat}$
$F'^{i,j}_{\rm E}$
$E'^{i,j}_{\rm pk}$
$\sigma^{i,j}_{F_{\rm E}}$
$F^{i,j}_{\rm E}$
$E^{i,j}_{\rm pk}$
$\sigma^{i,j}_{E_{\rm pk}}$
GRB_i: $i = 1, \cdots, N_{\rm GRB}$ t_j: $j = 0, \cdots$

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

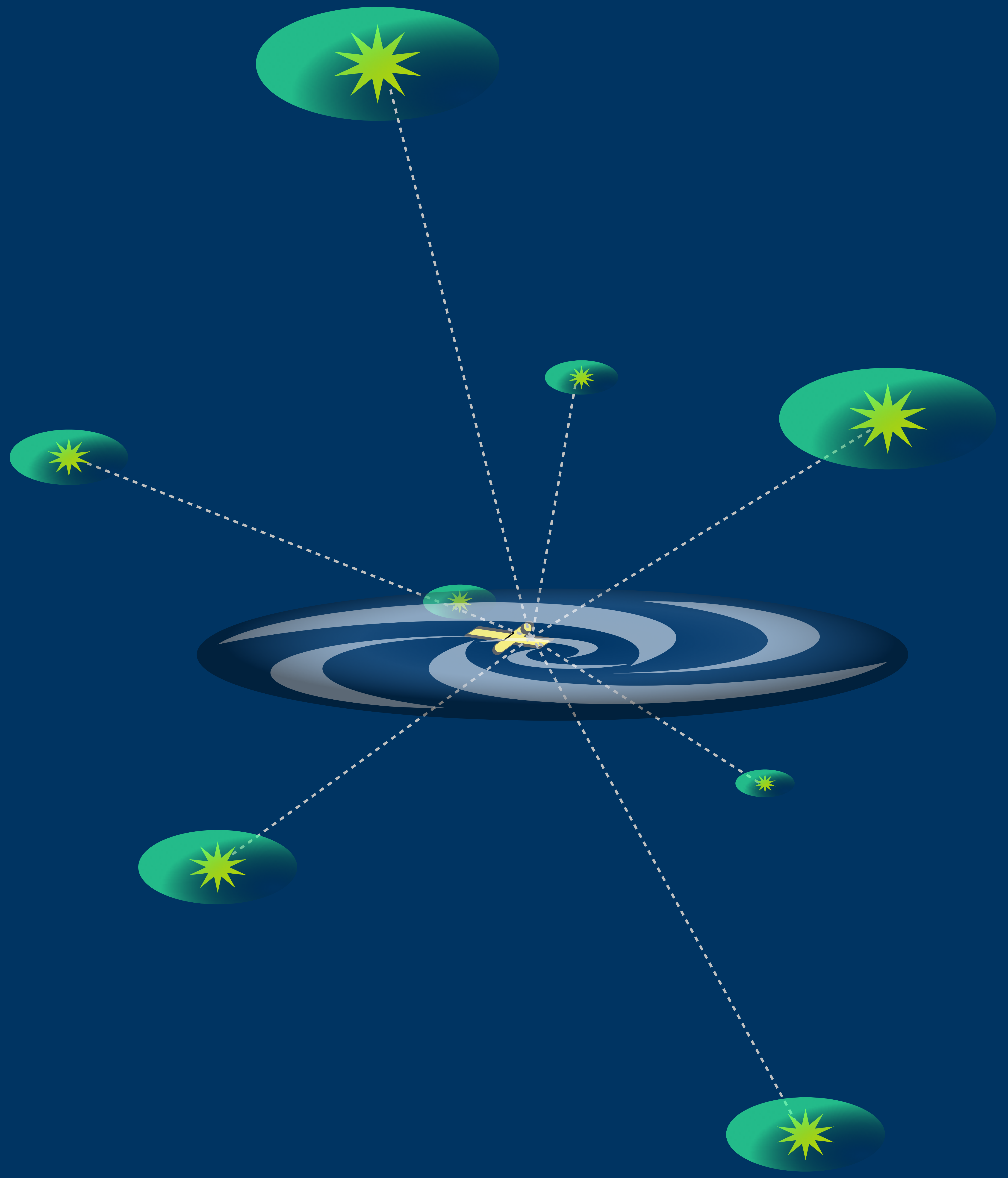


$$f_i\left(\varepsilon;\gamma,N_{\mathrm{host},i},N_{\mathrm{mw},i}\right)=K_i\left(\frac{\varepsilon}{\varepsilon_{\mathrm{piv}}}\right)^{\gamma_i}\eta\left(N_{\mathrm{host},i},\varepsilon\left(1+z_i\right)\right)\eta\left(N_{\mathrm{mw},i},\varepsilon\right)$$

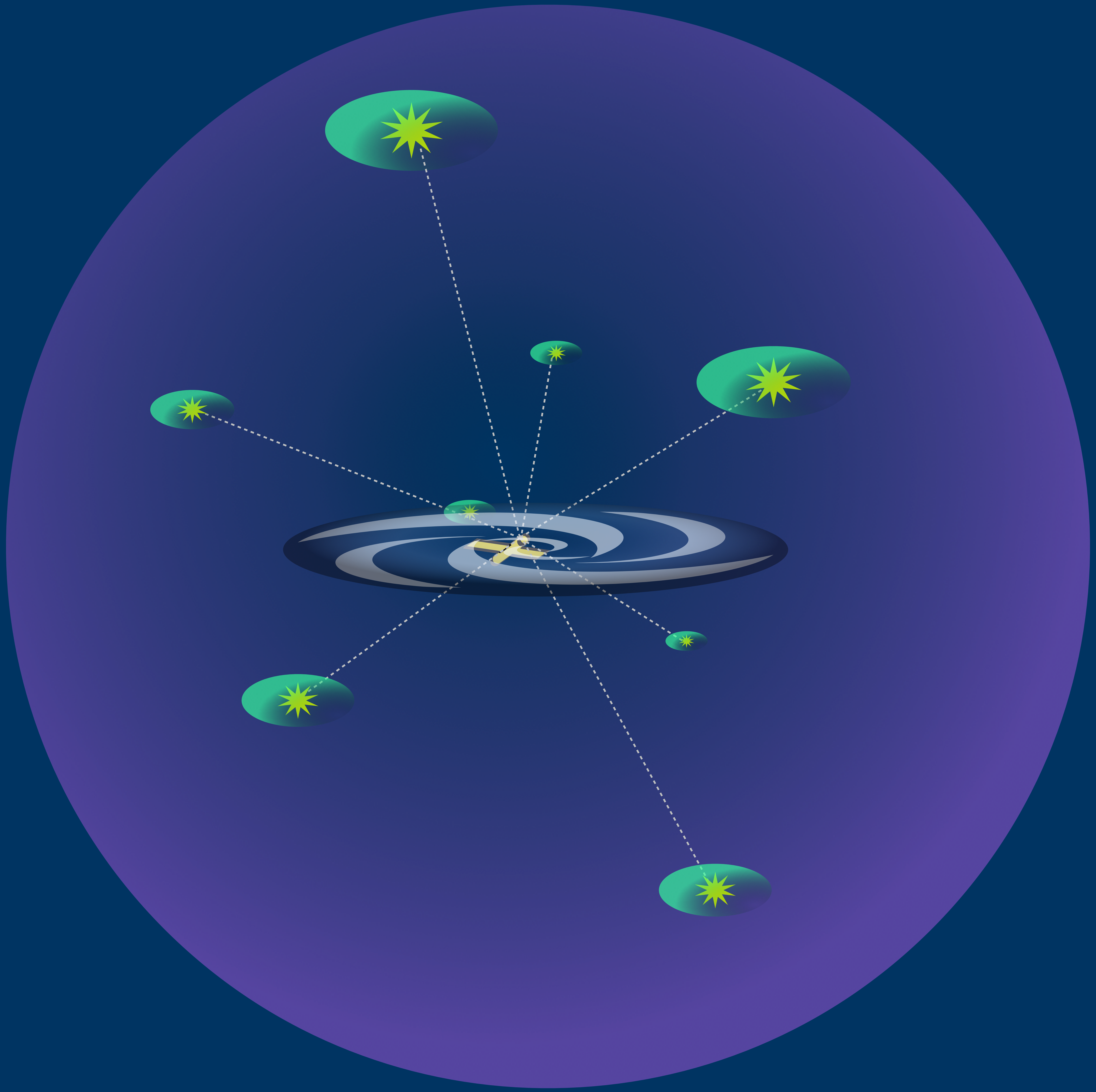
$$f_i\left(\varepsilon;\gamma,N_{\mathrm{host},i},N_{\mathrm{mw},i}\right)=K_i\left(\frac{\varepsilon}{\varepsilon_{\mathrm{piv}}}\right)^{\gamma_i}\eta\left(N_{\mathrm{host},i},\varepsilon\left(1+z_i\right)\right)\eta\left(N_{\mathrm{mw},i},\varepsilon\right)$$



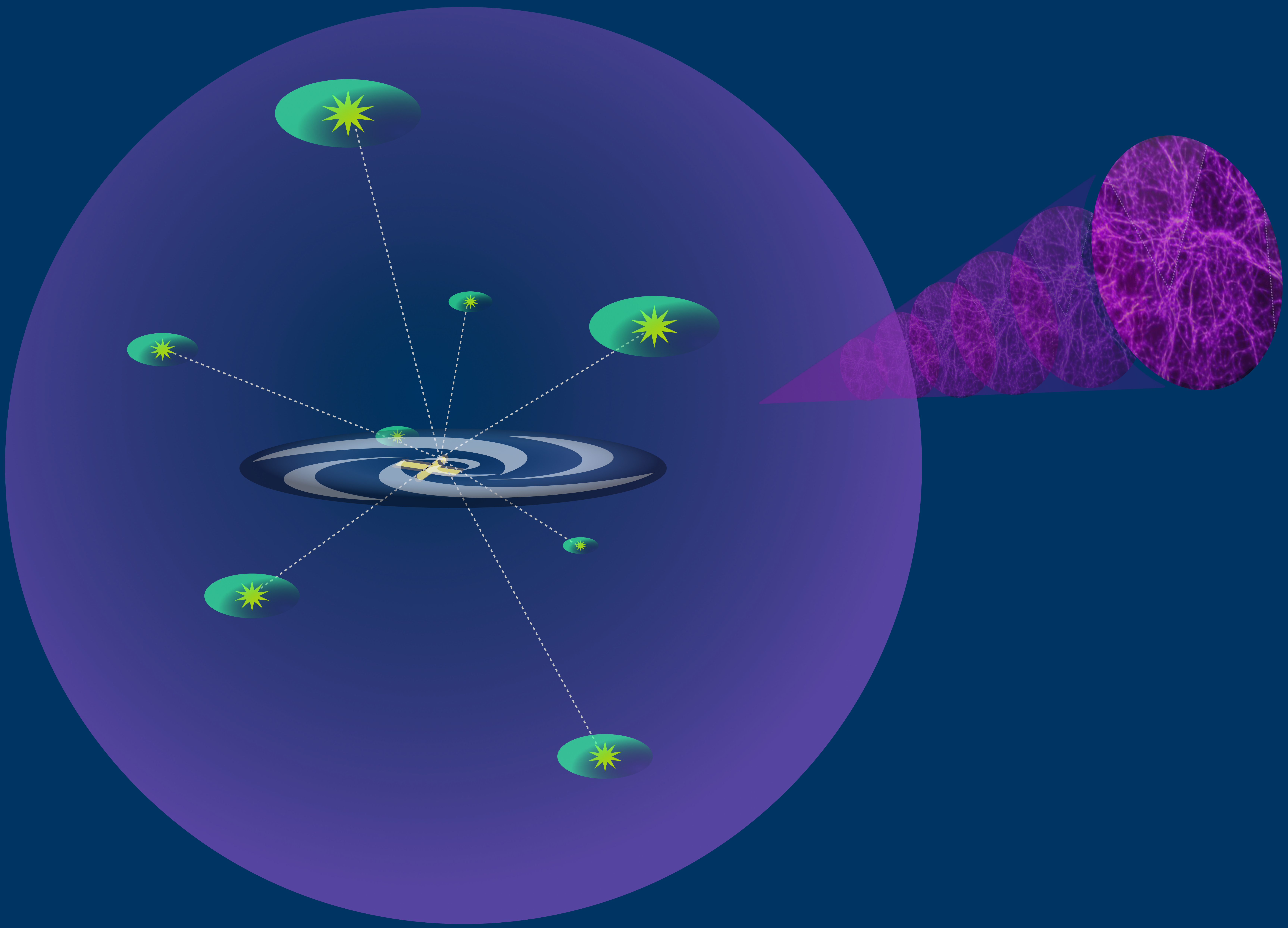
Starling+ 2013



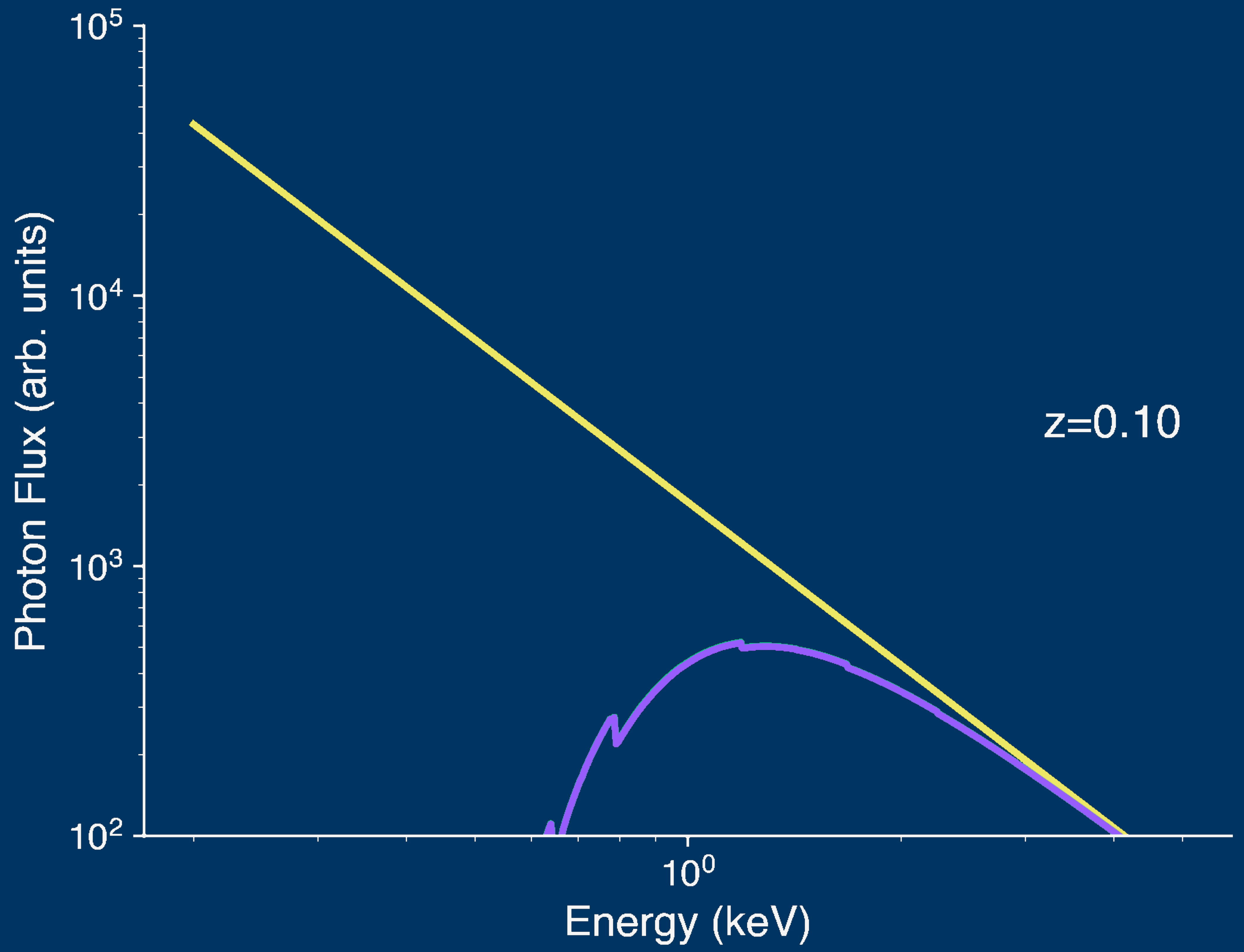
WHIM



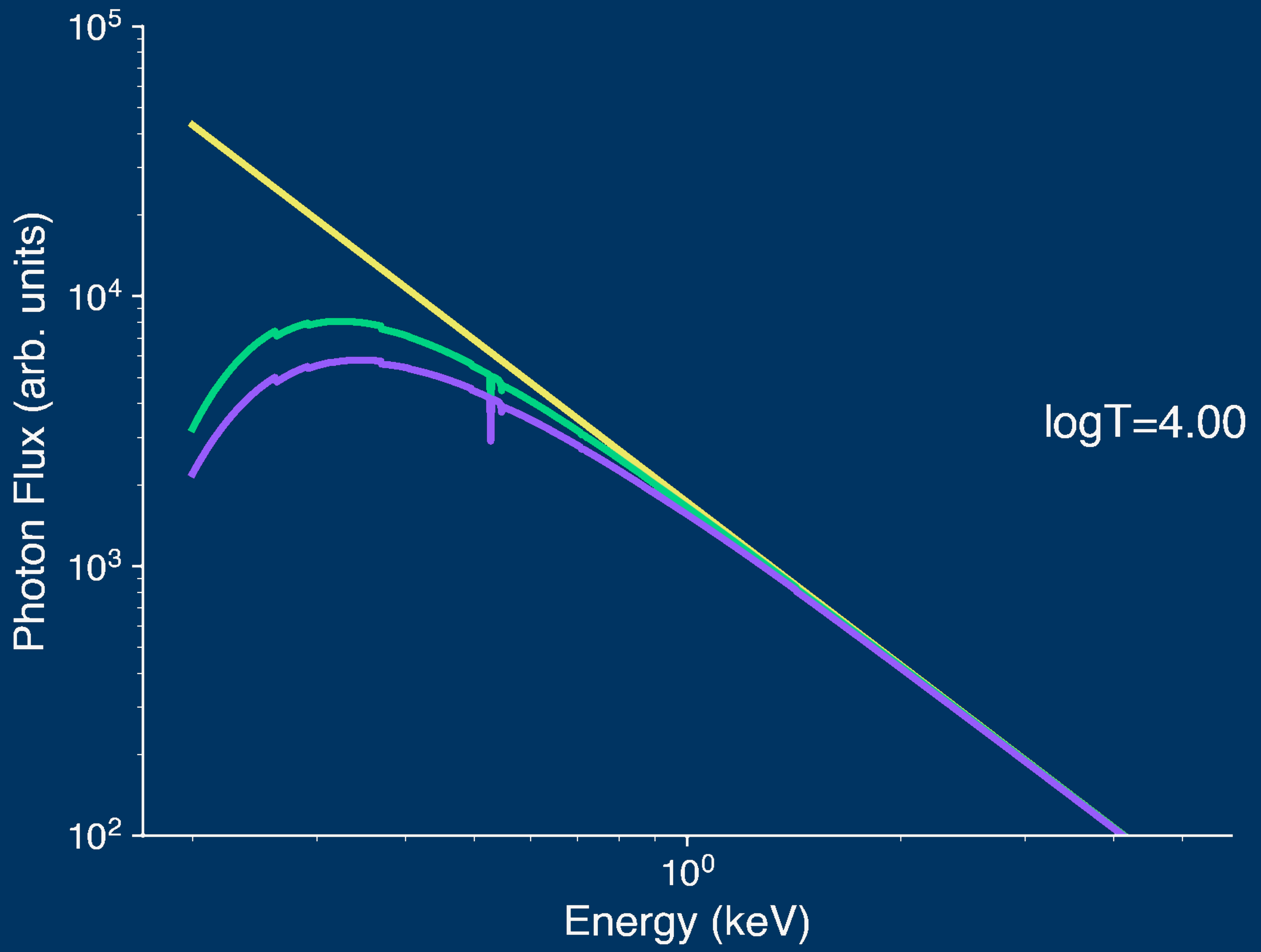
WHIM



WHIM

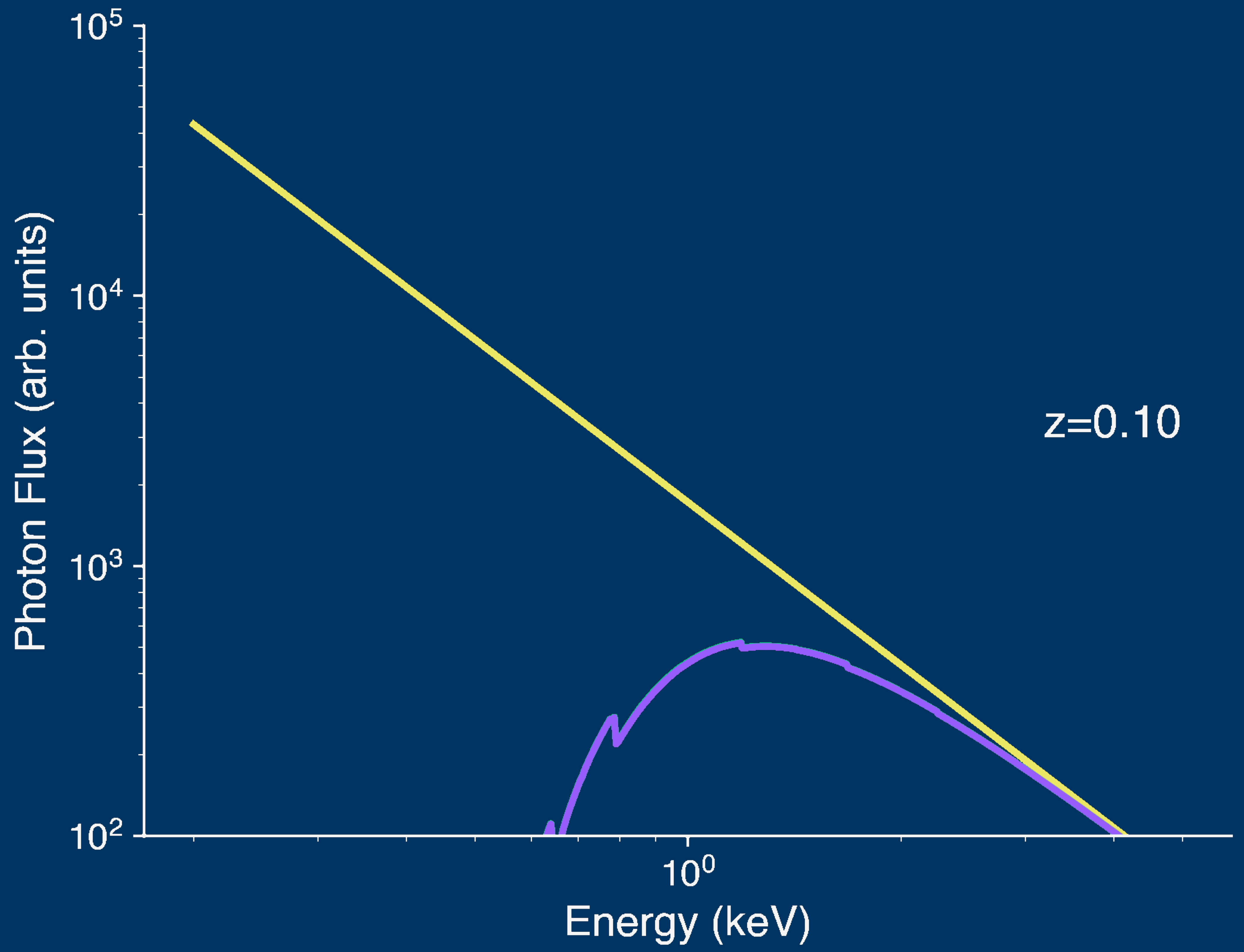


10^5
10^4
10^3
10^2
10^0
Photon Flux (arb. units)
Energy (keV)
z=0.10

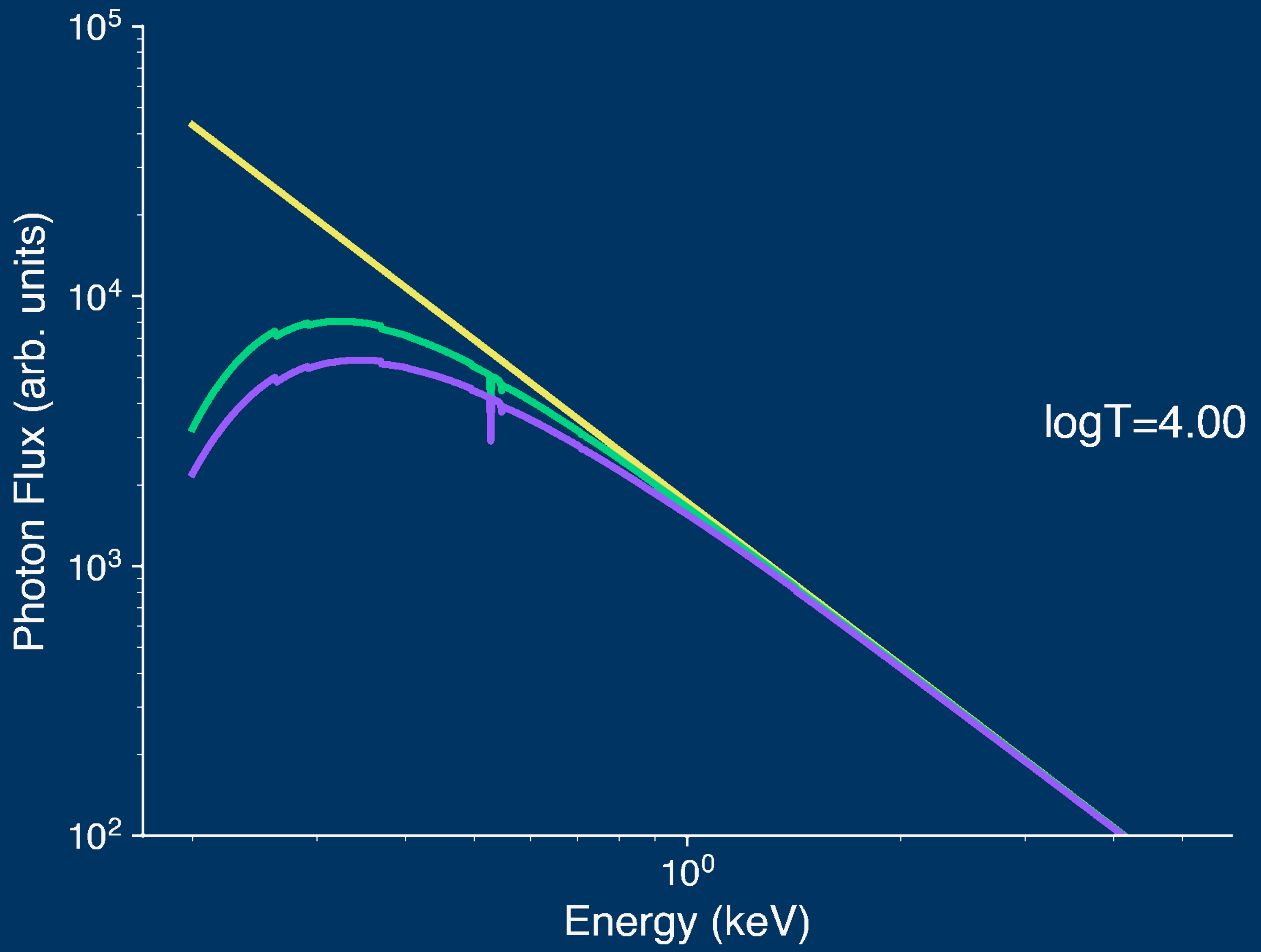


10^5
10^4
10^3
10^2
10^0
Photon Flux (arb. units)
Energy (keV)
logT=4.00

WHIM



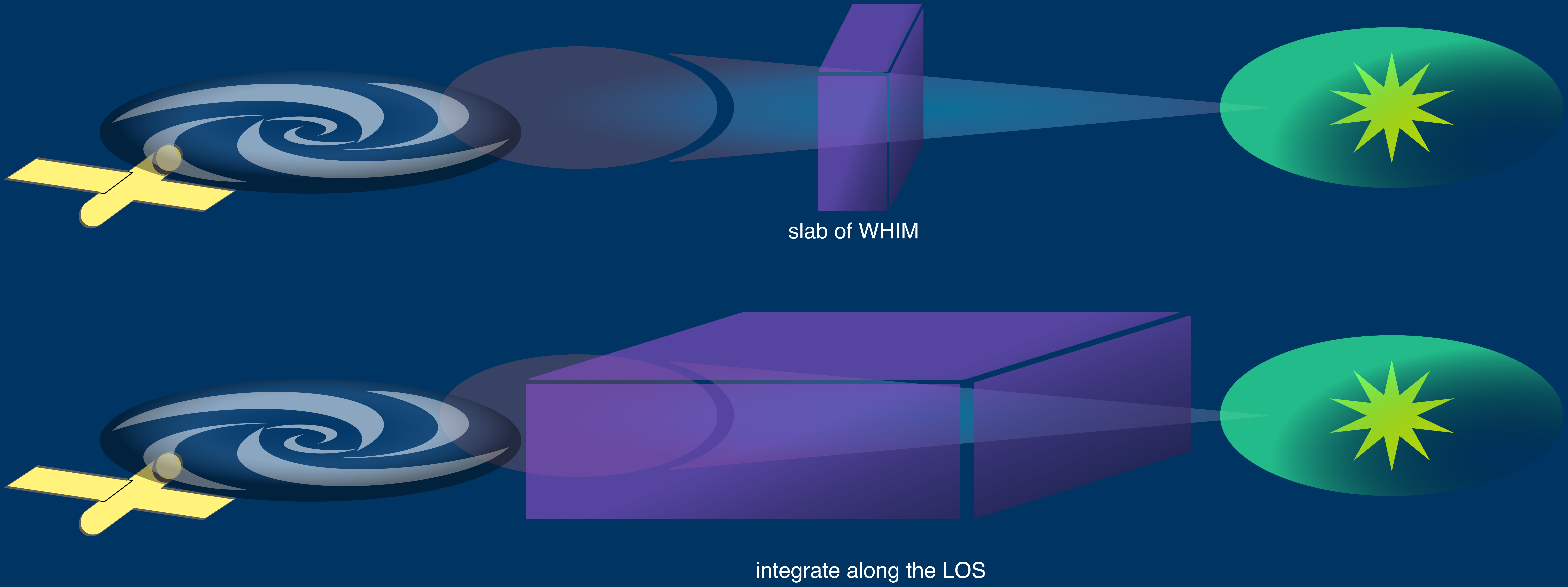
10^5
10^4
10^3
10^2
10^0
Photon Flux (arb. units)
Energy (keV)
z=0.10



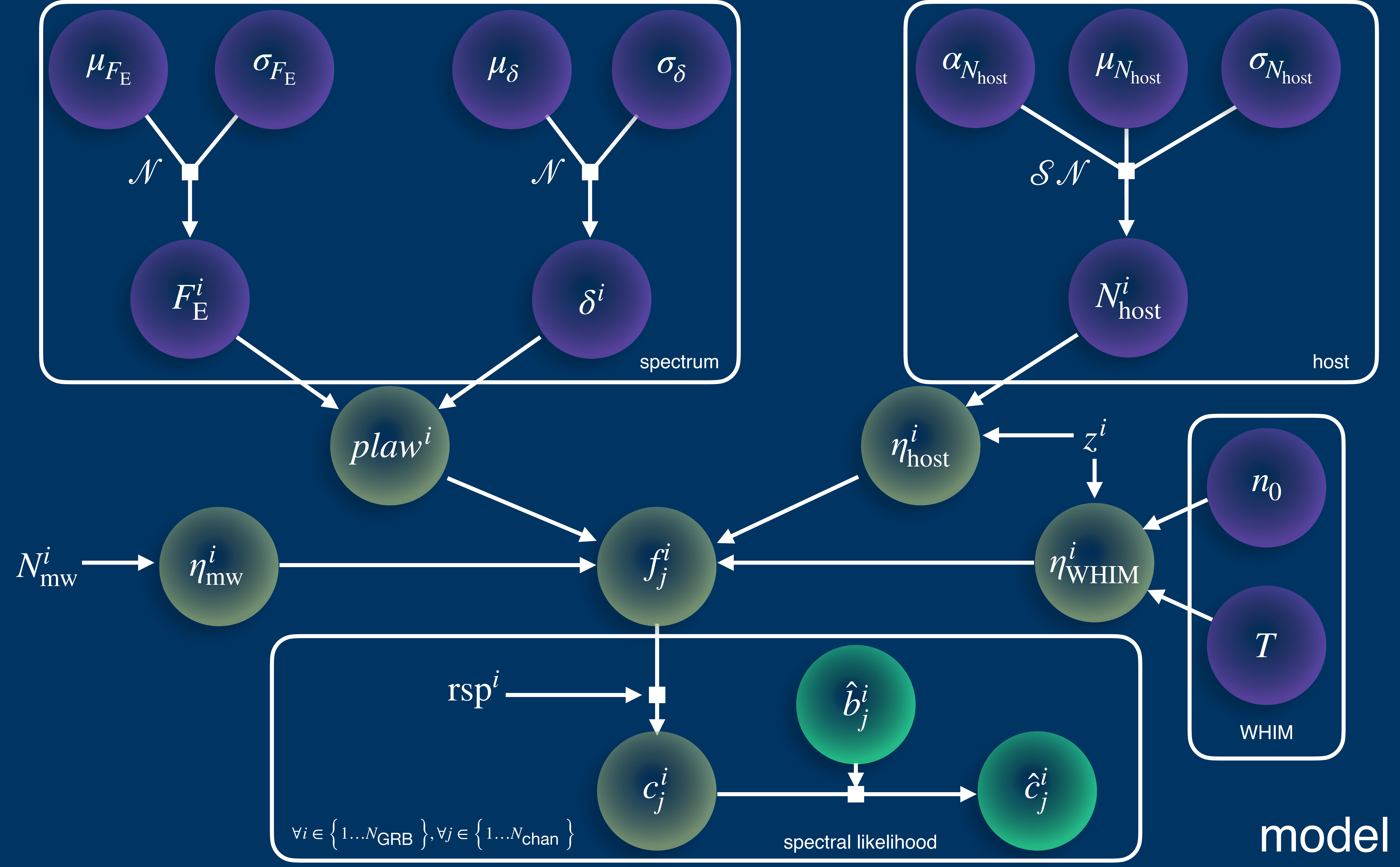
10^5
10^4
10^3
10^2
10^0
Photon Flux (arb. units)
Energy (keV)
logT=4.00

WHIM

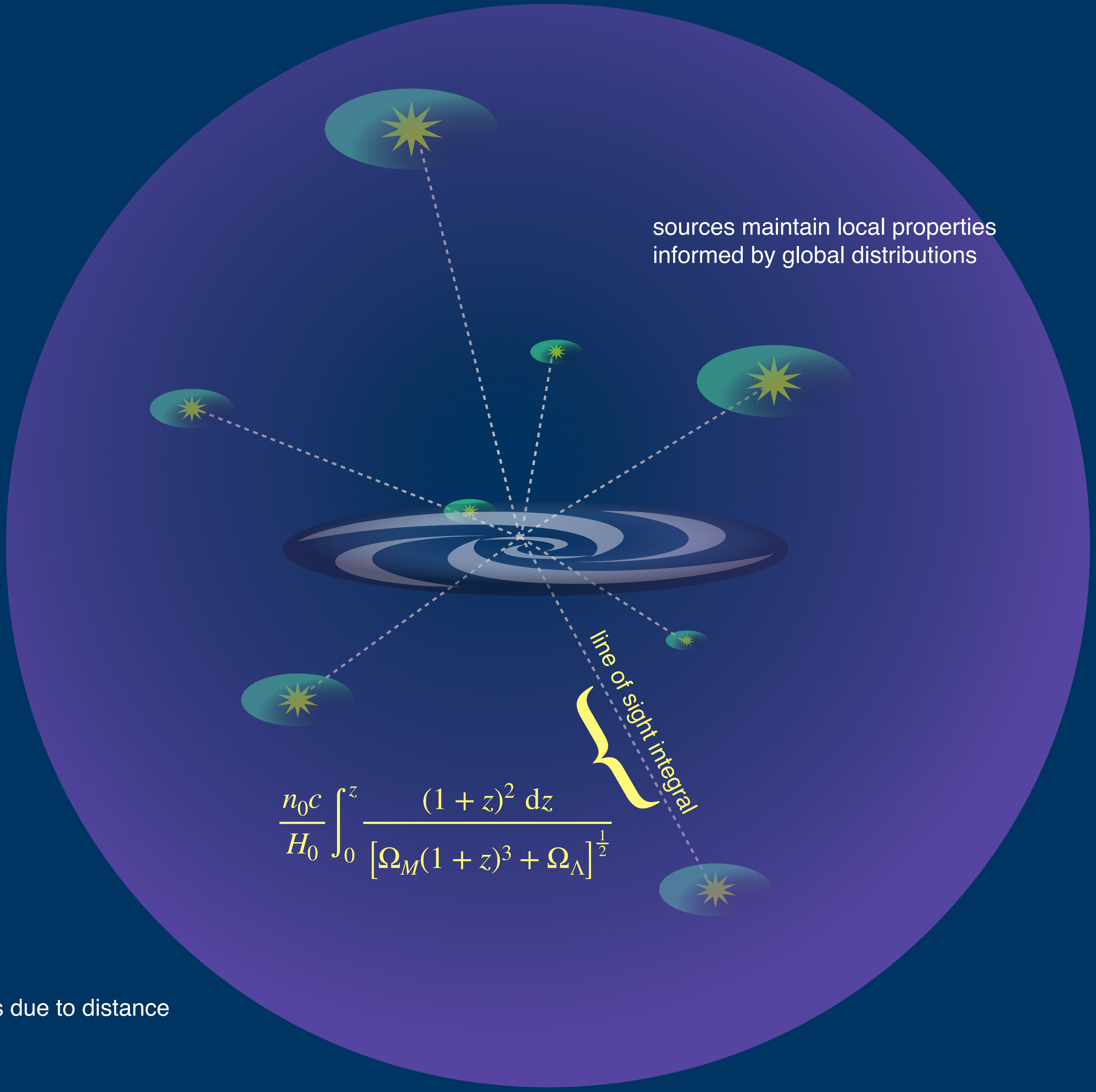
$$f_i\left(\varepsilon;\gamma,N_{\mathrm{host},i},N_{\mathrm{mw},i}\right)=K_i\left(\frac{\varepsilon}{\varepsilon_{\mathrm{piv}}}\right)^{\gamma_i}\eta\left(N_{\mathrm{host},i},\varepsilon\left(1+z_i\right)\right)\eta\left(N_{\mathrm{mw},i},\varepsilon\right)\eta\left(N_{0,i},z,\varepsilon\right)$$



WHIM

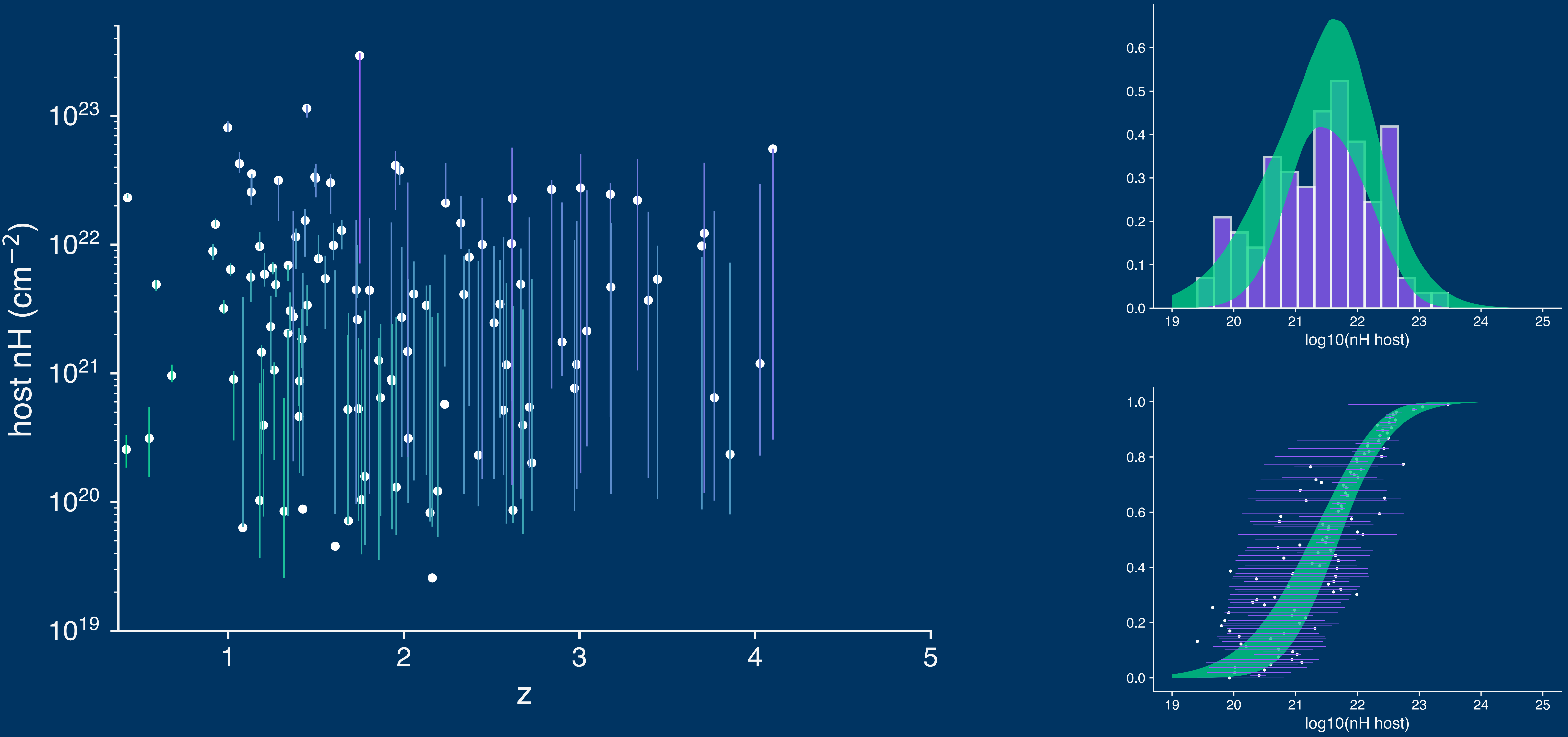


μ_{F_E}
σ_{F_E}
μ_δ
σ_δ
$\alpha_{N_{host}}$
$\mu_{N_{host}}$
$\sigma_{N_{host}}$
$\mathcal{N}$
$\mathcal{N}$
$\mathcal{SN}$
F_E^i
δ^i
N_{host}^i
spectrum
host
$plaw^i$
η_{host}^i
z^i
n_0
N_{mw}^i
η_{mw}^i
f_j^i
η_{WHIM}^i
T
WHIM
rsp^i
$\hat{b}_j^i$
c_j^i
$\hat{c}_j^i$
$\forall i \in \{1 \ldots N_{GRB}\}, \forall j \in \{1 \ldots N_{chan}\}$
spectral likelihood
model

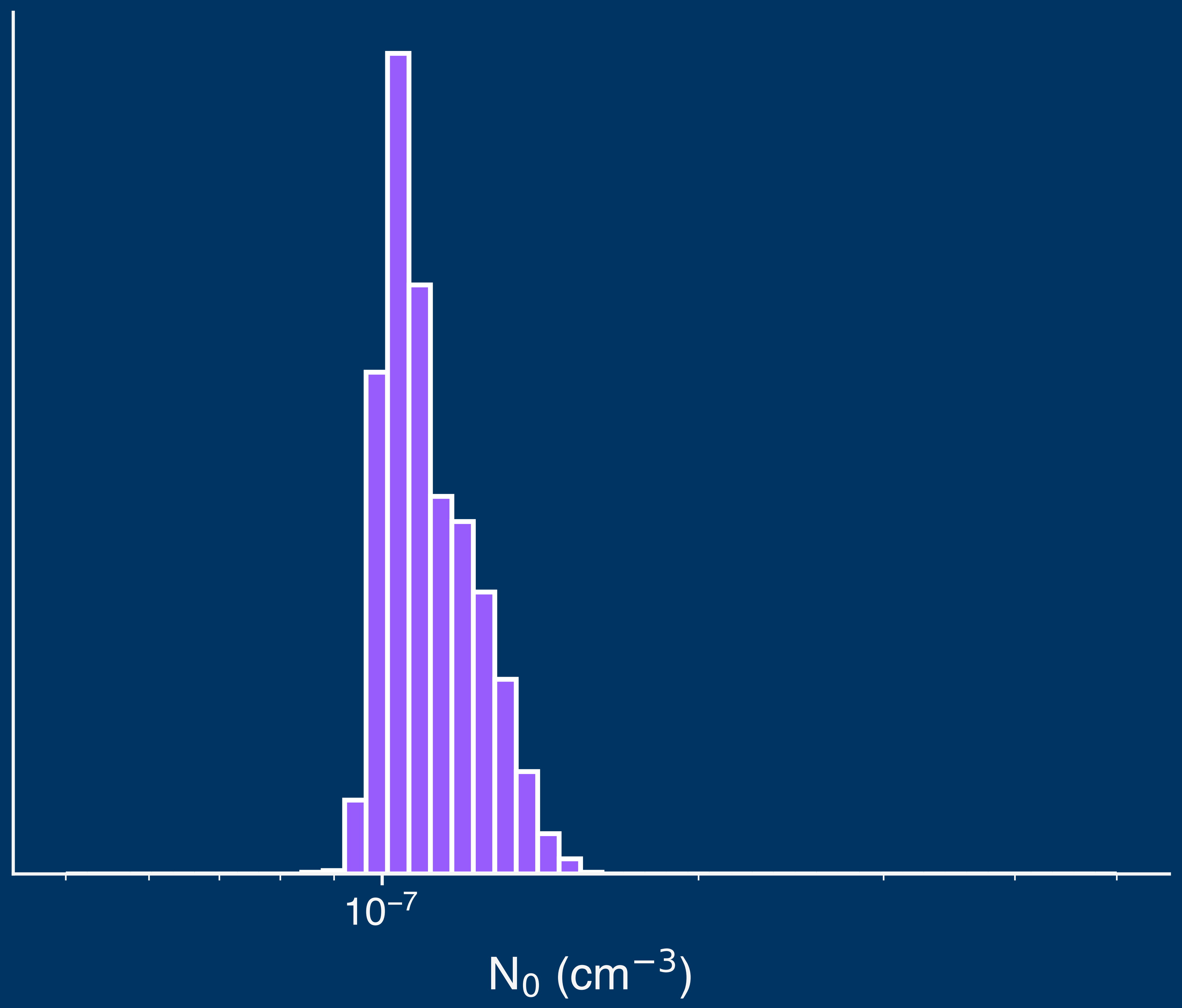


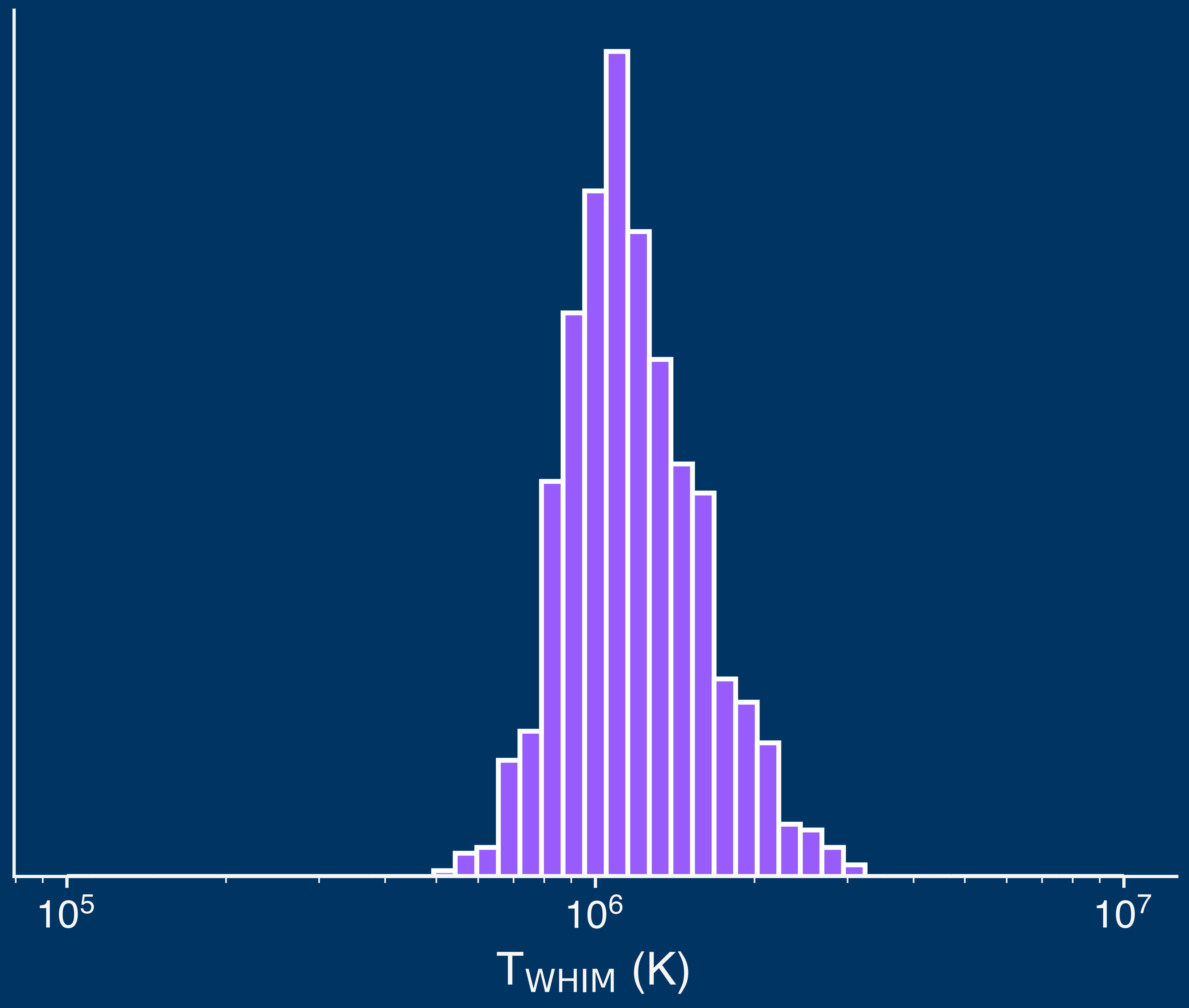
absorption varies due to distance
from observer

model



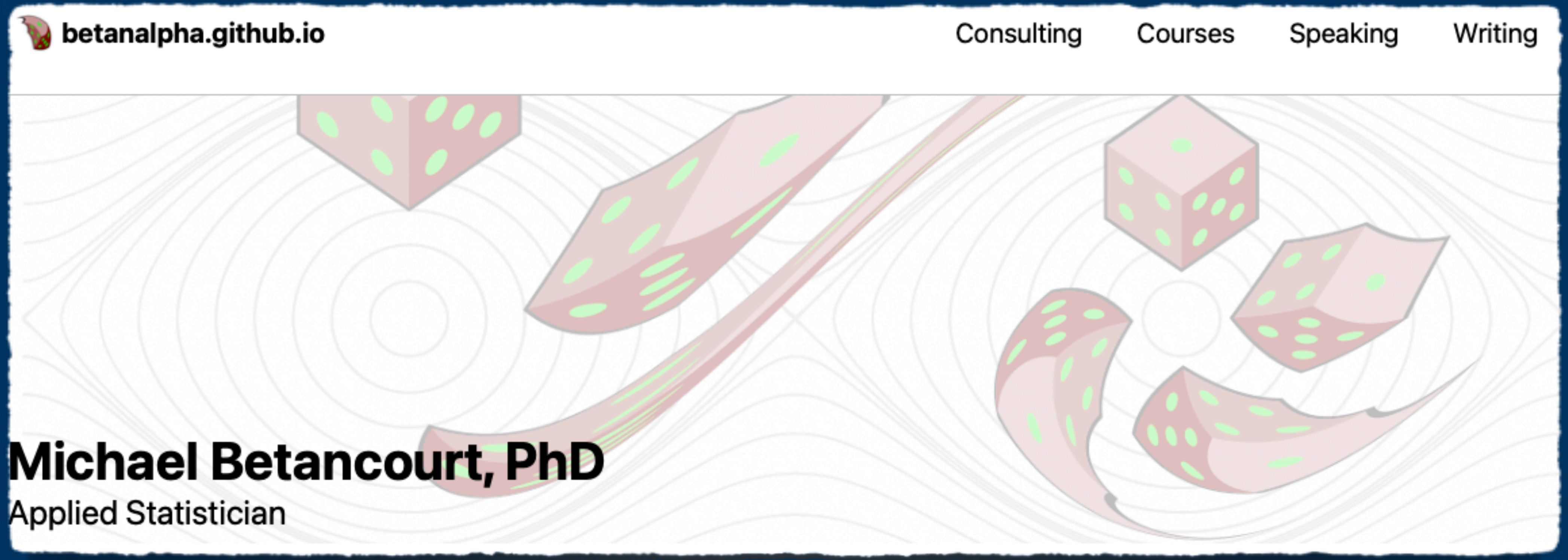
fitting the model with WHIM

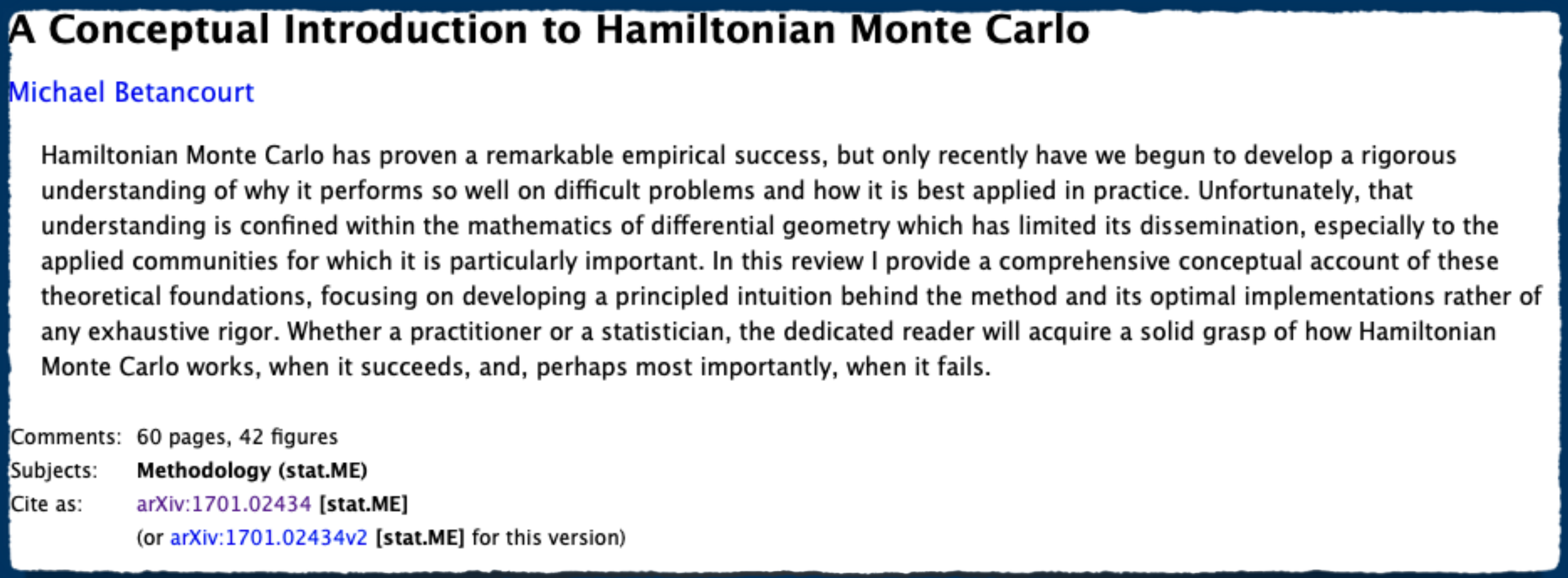




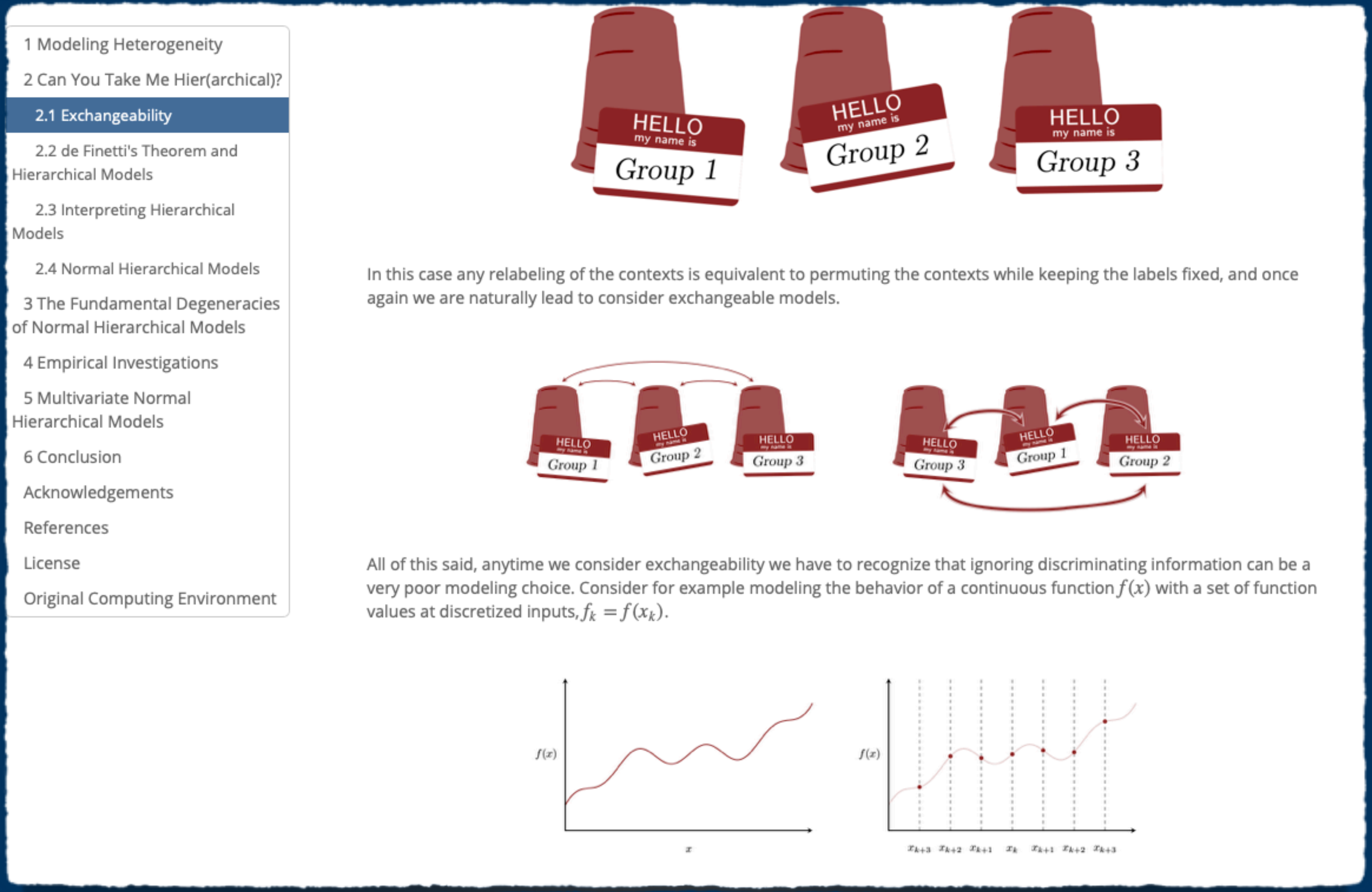
fitting the model with WHIM

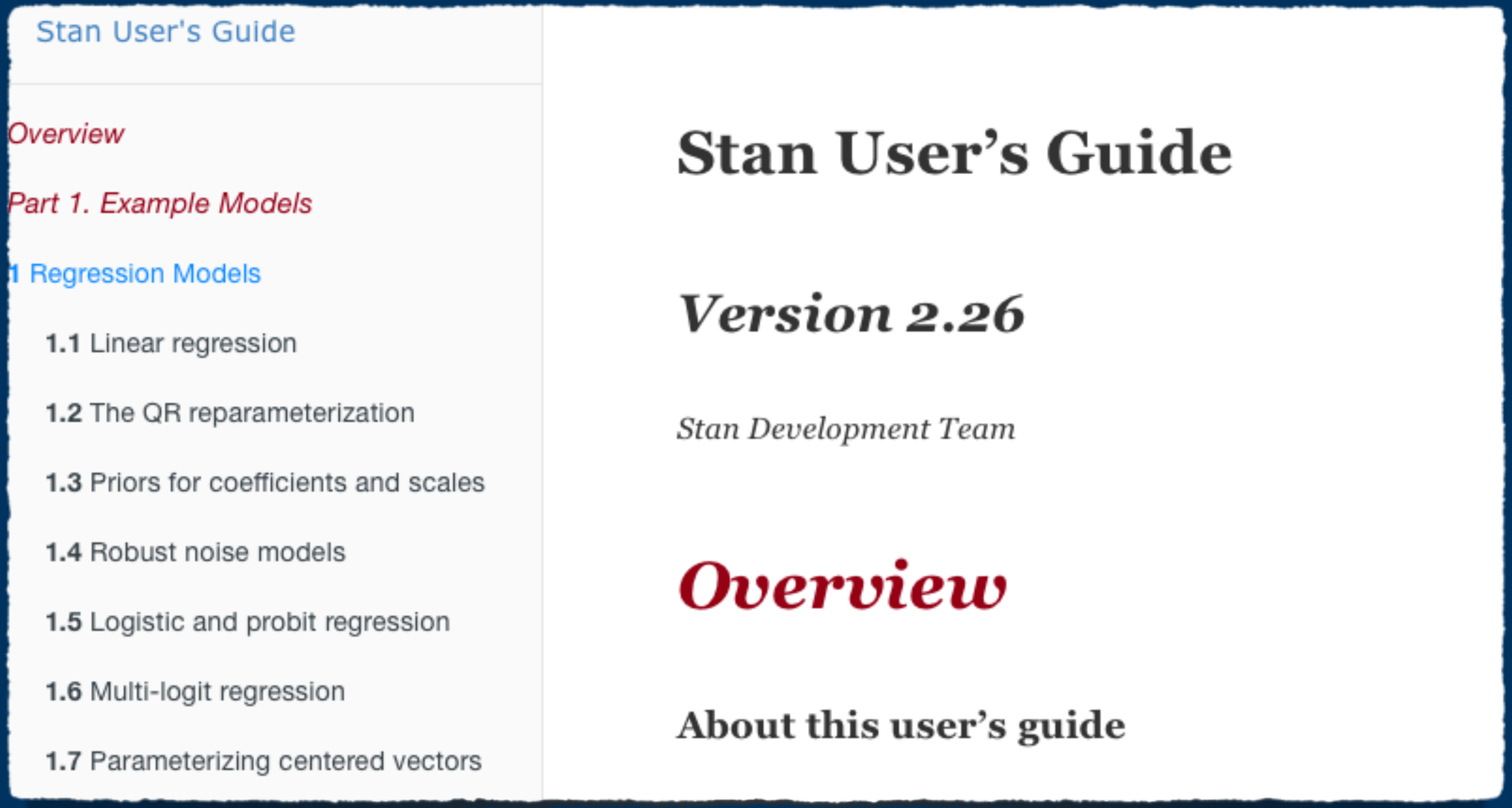
Resources



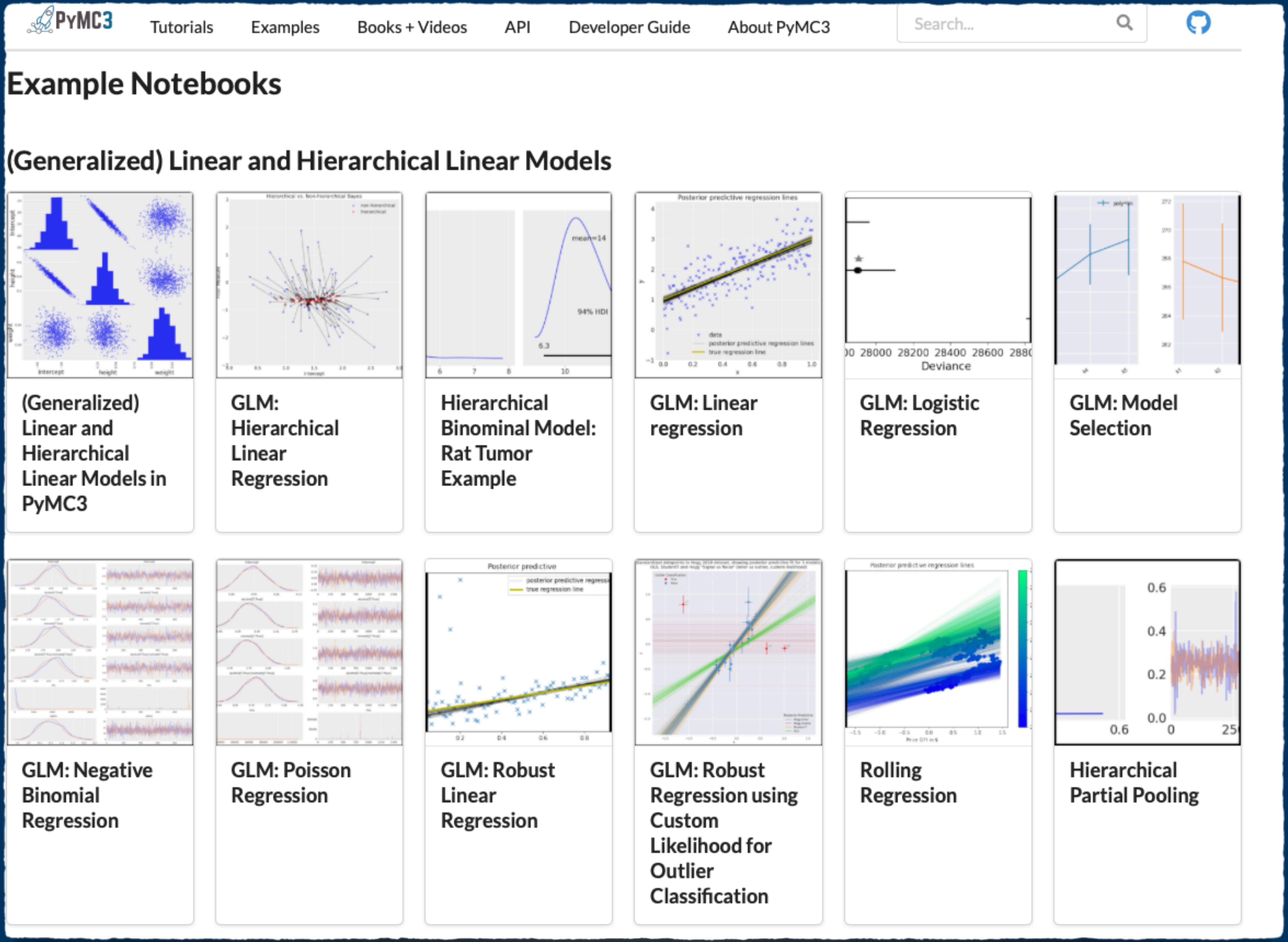


https://betanalpha.github.io/assets/case_studies/hierarchical_modeling.html





Resources


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**