Ex astris scientia: what can we learn about gravity from the stars?

Savvas Nesseris

IFT UAM-CSIC, Madrid, Spain

https://members.ift.uam-csic.es/savvas.nesseris/

https://github.com/snesseris

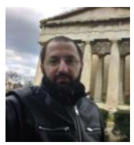
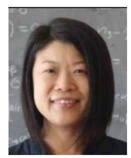
Outline

- Machine Learning (ML) and Artificial Intelligence (AI)
- Overview of the Hellings-Downs (HD) curve & its extensions
- Astrometry, GAIA & the stochastic GW background (SGWB)
- Two applications:
  - A comparison (Bayesian & ML) of extended HD models
  - Neural Network (NN) constraints on the SGWB
- Conclusions

Based on the following papers

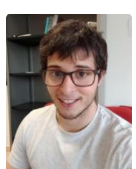
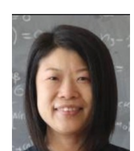
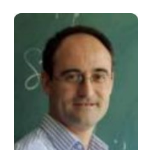
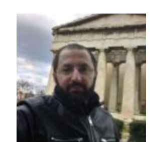
Comparative analysis of the NANOgrav Hellings-Downs as a window into new physics
arXiv: 2412.12975

Ruben + Savvas + Sachiko

Astrometric constraints on stochastic gravitational wave background with neural networks
arXiv: 2412.15879

Marienza + Gonzalo + Santiago + Sachiko + Juan + Savvas

CHAPTER 1
AI & MACHINE LEARNING

AI Nobel prize controversy!

KUNGL. VETENSKAPS-AKADEMIEN

THE ROYAL SWEDISH ACADEMY OF SCIENCES

8 October 2024

The Royal Swedish Academy of Sciences has decided to award the Nobel Prize in Physics 2024 to

John J. Hopfield
Princeton University, NJ, USA

Created the Hopfield network, a type of artificial network made of binary neurons that can be 'on' or 'off'. He extended his formalism to continuous activation functions.

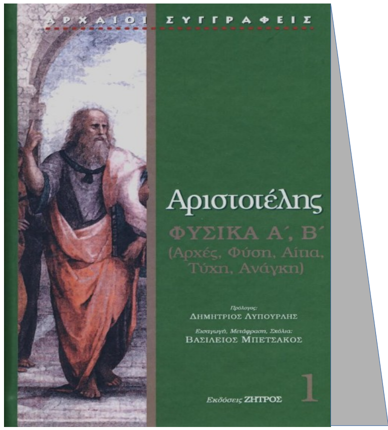
Geoffrey E. Hinton
University of Toronto, Canada

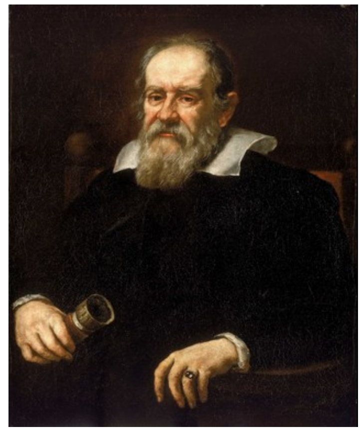
A Boltzmann machine, is a spin-glass model with an external field, that is a stochastic Ising model. It is also classified as a Markov random field and can learn to recognize characteristic elements in a given type of data.

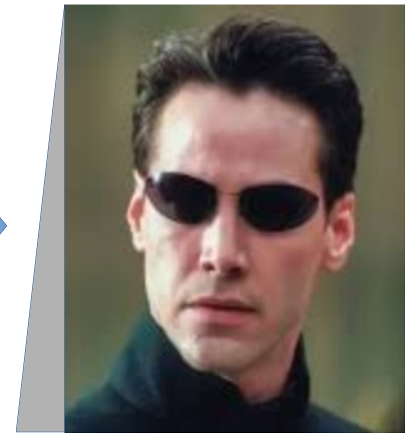
"for foundational discoveries and inventions that enable machine learning with artificial neural networks"

They trained artificial neural networks using physics

Physics has always been evolving…







Theory + Logic
~350 BC

Experiments
~1600 AD

Simulations+AI
~1950 AD

What is Machine Learning?

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia

Article Talk Read Edit View history Search Wikipedia

Machine learning

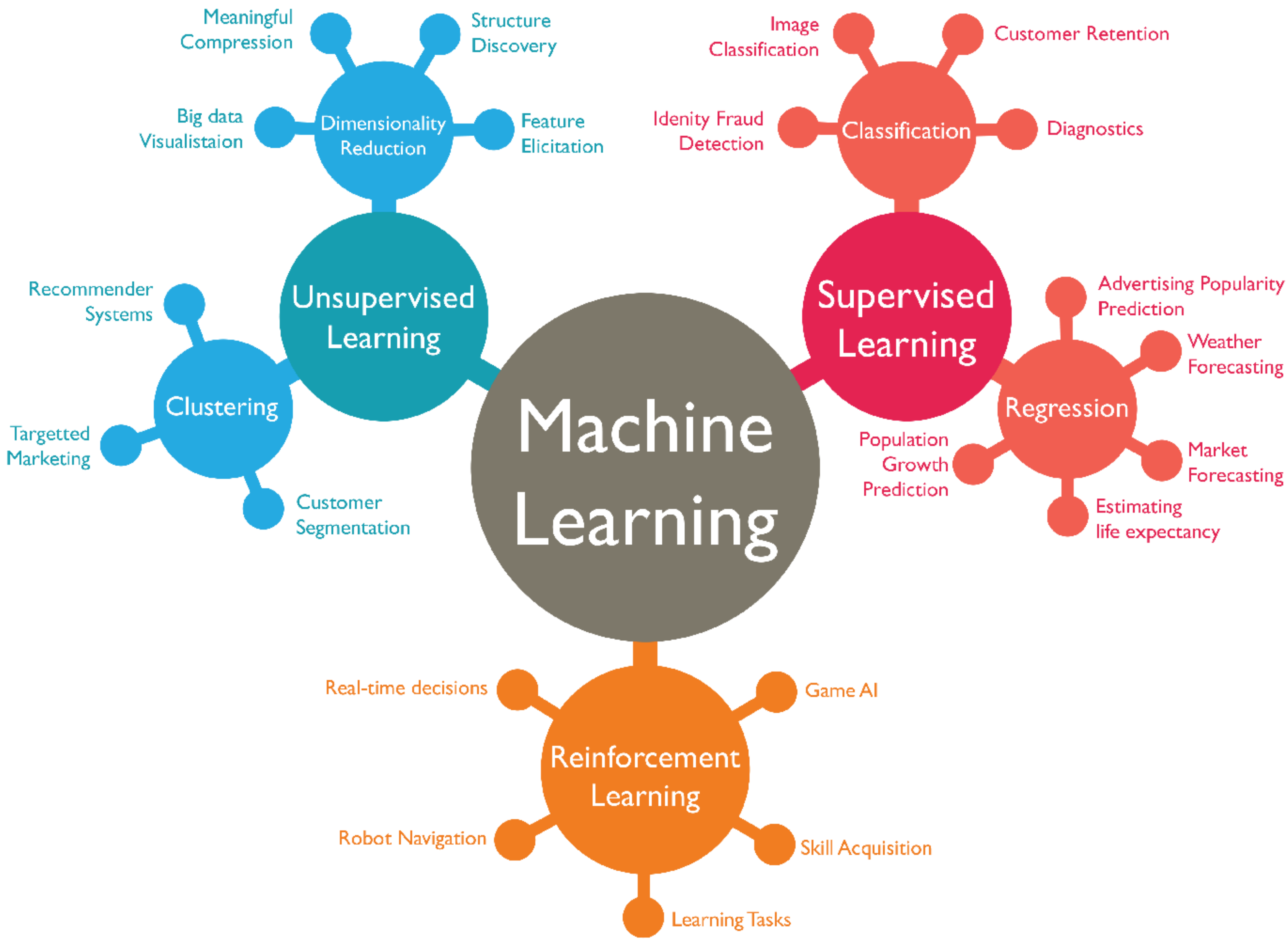
From Wikipedia, the free encyclopedia

For the journal, see Machine Learning (journal).
"Statistical learning" redirects here. For statistical learning in linguistics, see statistical learning in language acquisition.

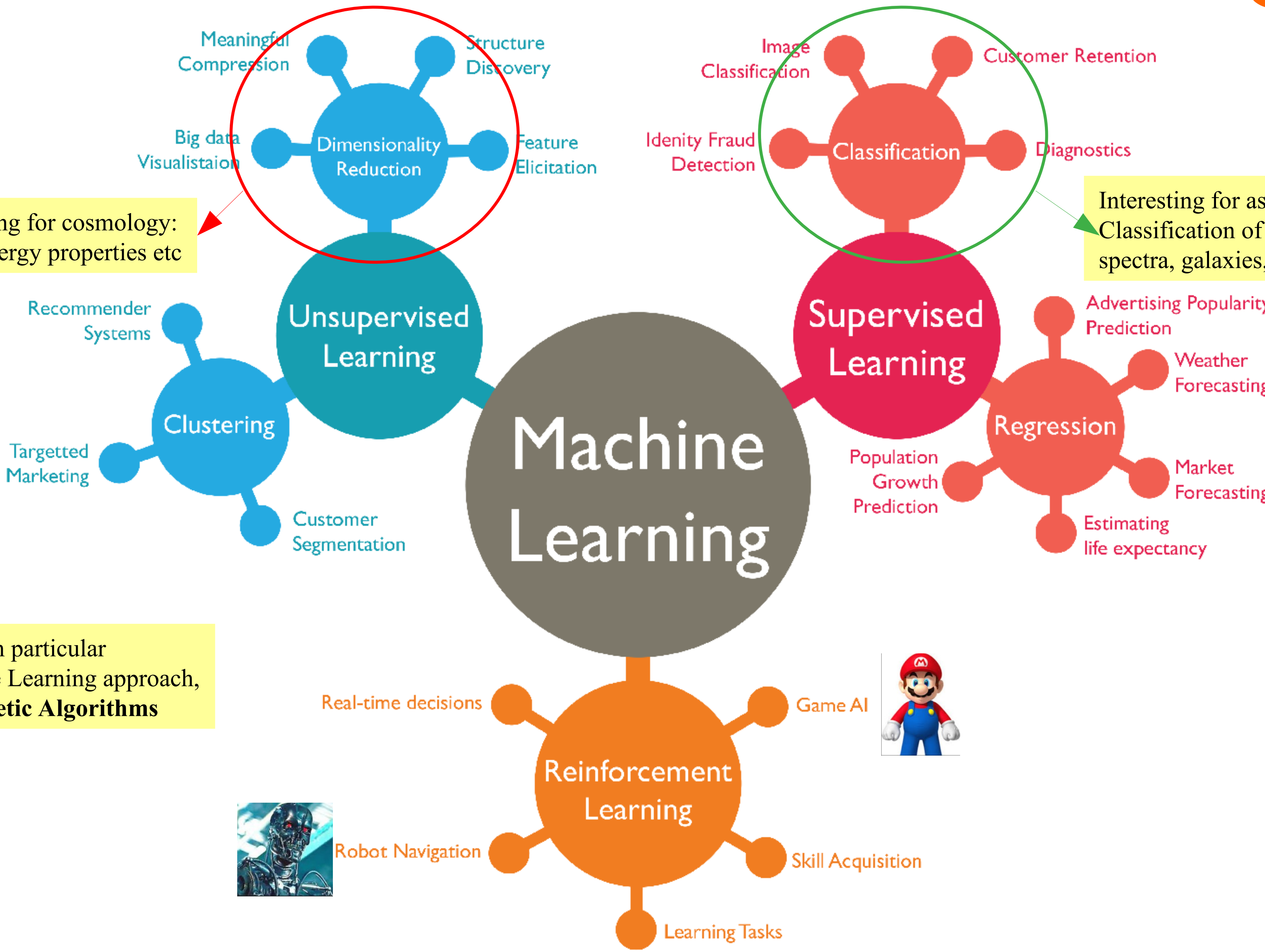
Machine learning (**ML**) is the study of computer algorithms that improve automatically through experience.[1] It is seen as a subset of artificial intelligence.

Part of a series on
Machine learning

What is Machine Learning?



What is Machine Learning?



Interesting for cosmology:
Dark Energy properties etc

Interesting for astronomy:
Classification of
spectra, galaxies, models etc

Focus on particular
Machine Learning approach,
the **Genetic Algorithms**

Neural Networks (NNs)

KUNGL.
VETENSKAPS-
AKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

8 October 2024

The Royal Swedish Academy of Sciences has decided to award the Nobel Prize in Physics 2024 to

John J. Hopfield
Princeton University, NJ, USA

Created the Hopfield network, a type of artificial network made of binary neurons that can be 'on' or 'off'. He extended his formalism to continuous activation functions.

Geoffrey E. Hinton
University of Toronto, Canada

A Boltzmann machine, is a spin-glass model with an external field, that is a stochastic Ising model. It is also classified as a Markov random field and can learn to recognize characteristic elements in a given type of data.

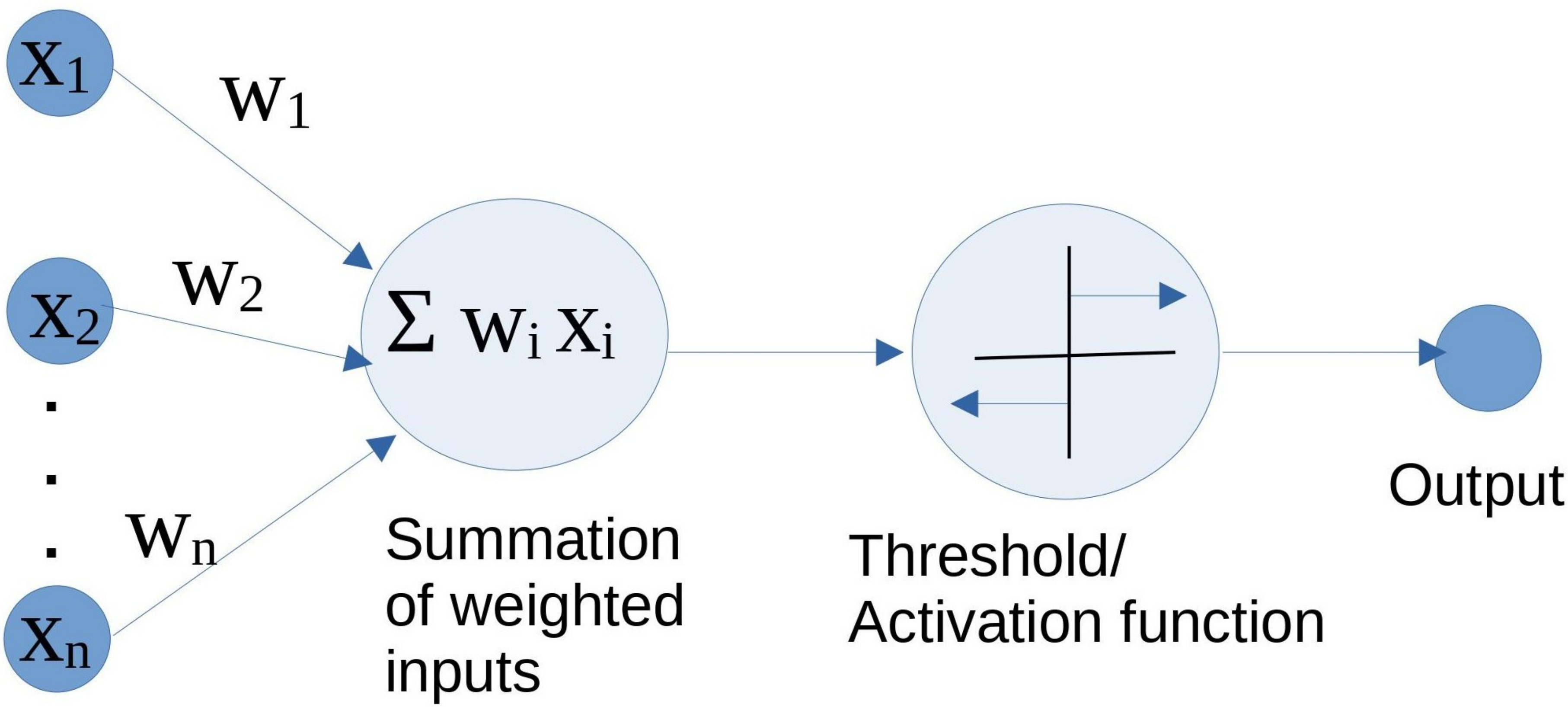
"for foundational discoveries and inventions that enable machine learning with artificial neural networks"

They trained artificial neural networks using physics

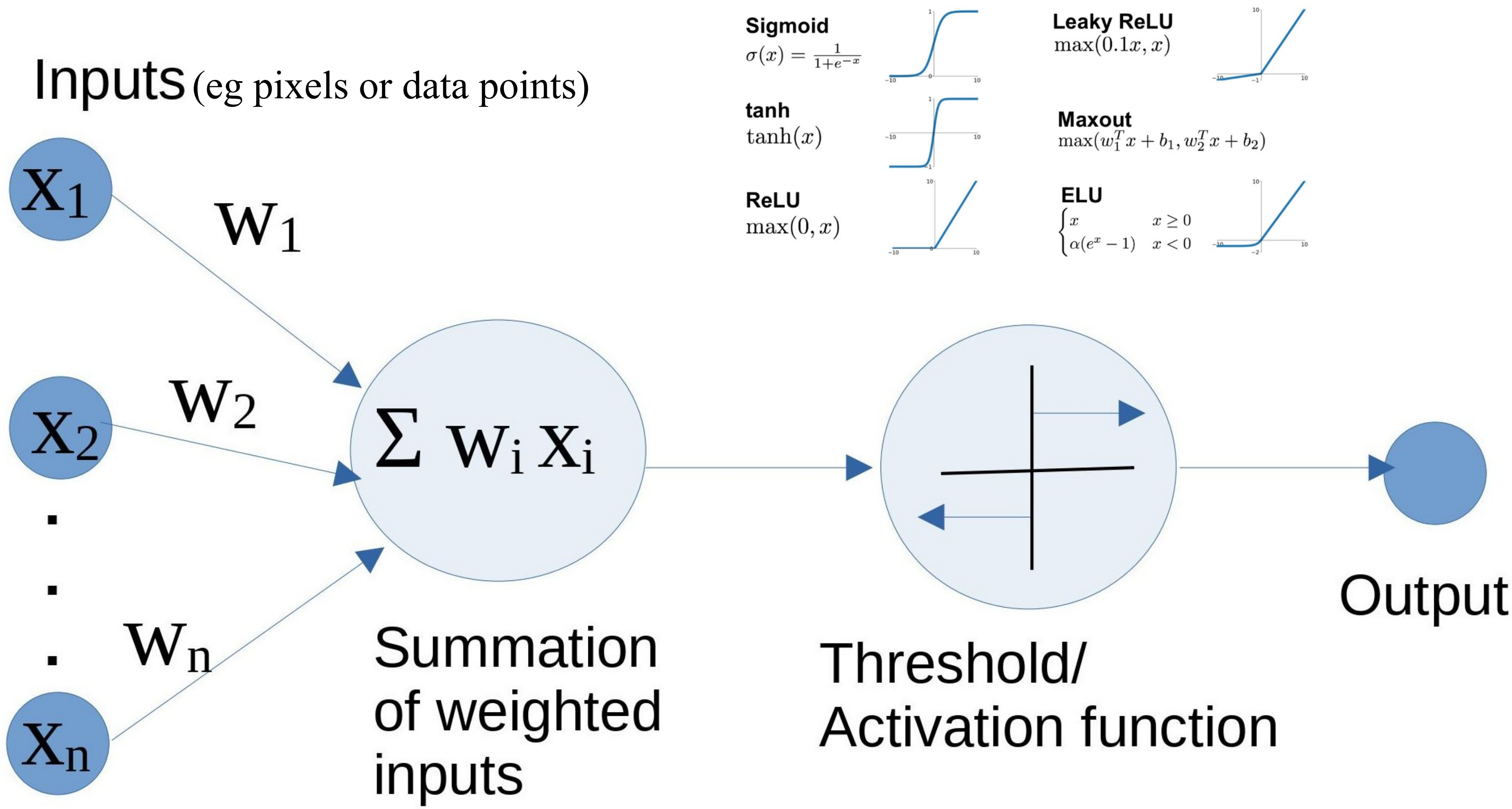
Neural Networks (NNs)

Carleo, Cirrac, et al. 1903.10563

Inputs (eg pixels or data points)



Neural Networks (NNs)

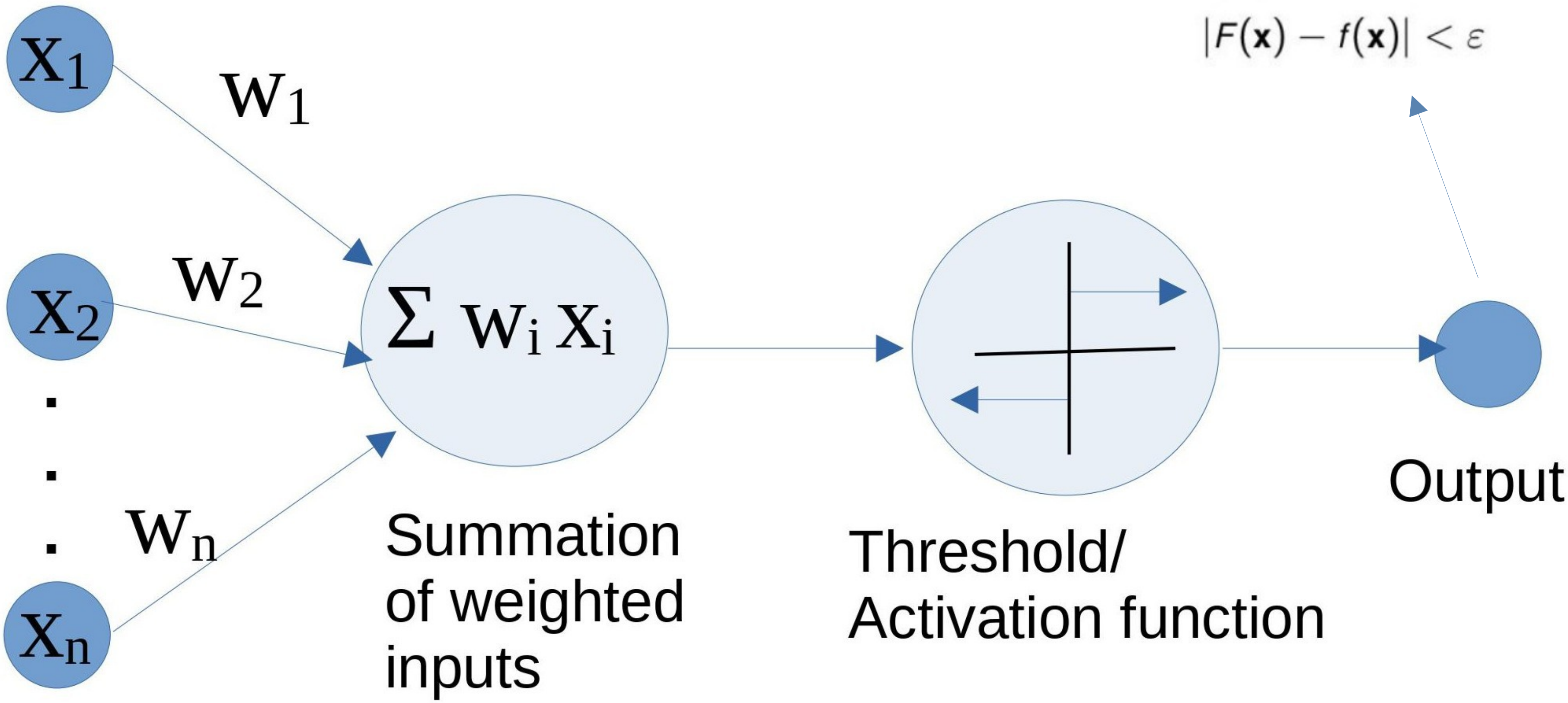


NNs as universal function approximators

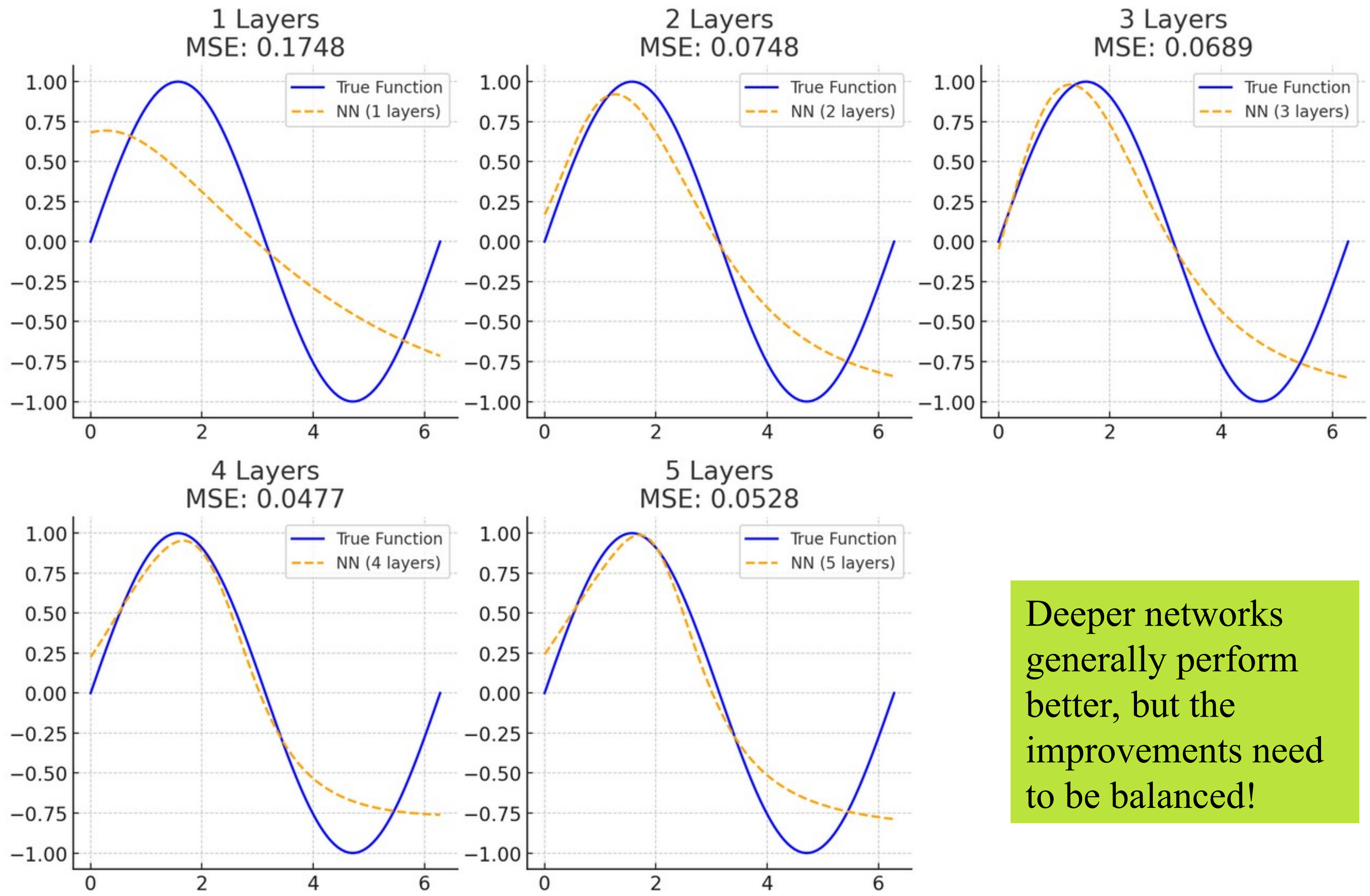
$$F(\mathbf{x}) = \sum_{i=1}^{N} v_i \varphi(\mathbf{w}_i^T \mathbf{x} + b_i) \quad \text{with}$$

$$|F(\mathbf{x}) - f(\mathbf{x})| < \varepsilon$$

Inputs (eg pixels or data points)



NNs as universal function approximators

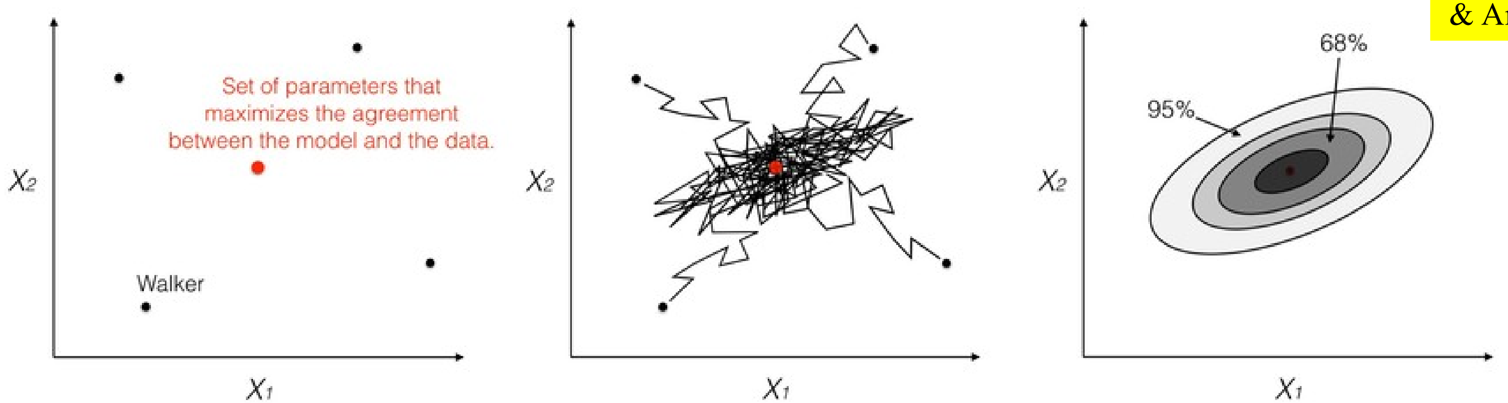


Deeper networks generally perform better, but the improvements need to be balanced!

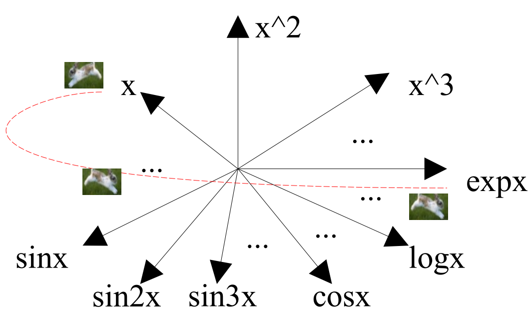
The Genetic Algorithms (GA)

The GA is a stochastic (i.e. random) optimization and symbolic regression method, not very different from an MCMC:

John Holland, (1975)
Adaptation in Natural
& Artificial Systems.



MCMCs do a random walk in parameter space (X_1, X_2 …), while the GA does a random walk in an abstract functional space!

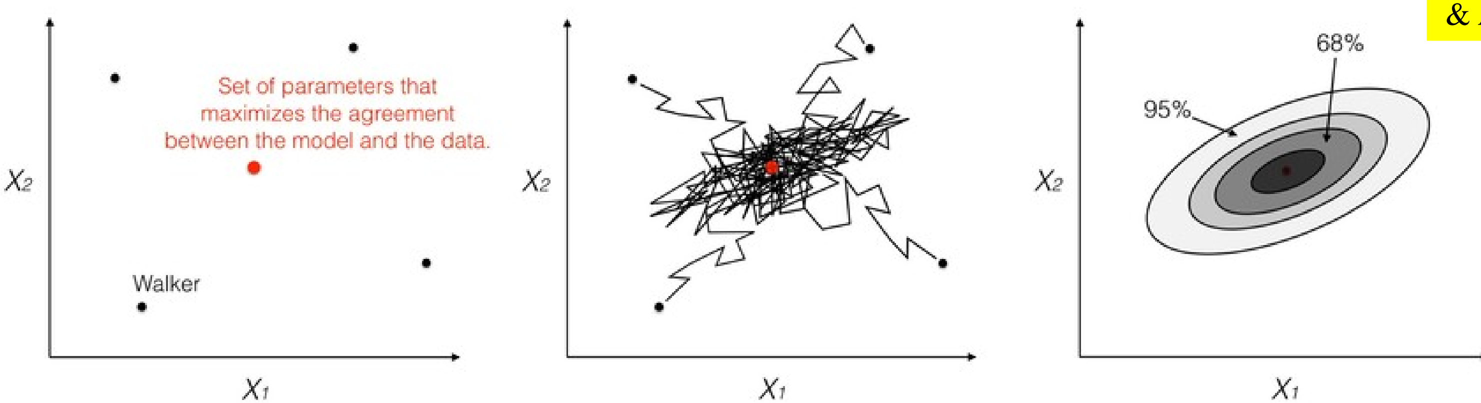


f(x) = x + 2*x^3 * (1+e^x) etc

The Genetic Algorithms (GA)

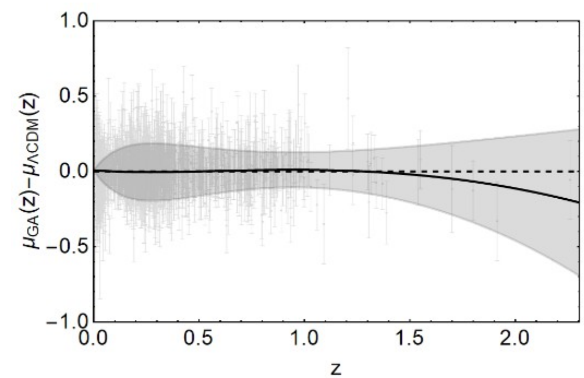
The GA is a stochastic (i.e. random) optimization and symbolic regression method, not very different from an MCMC:

John Holland, (1975) Adaptation in Natural & Artificial Systems.

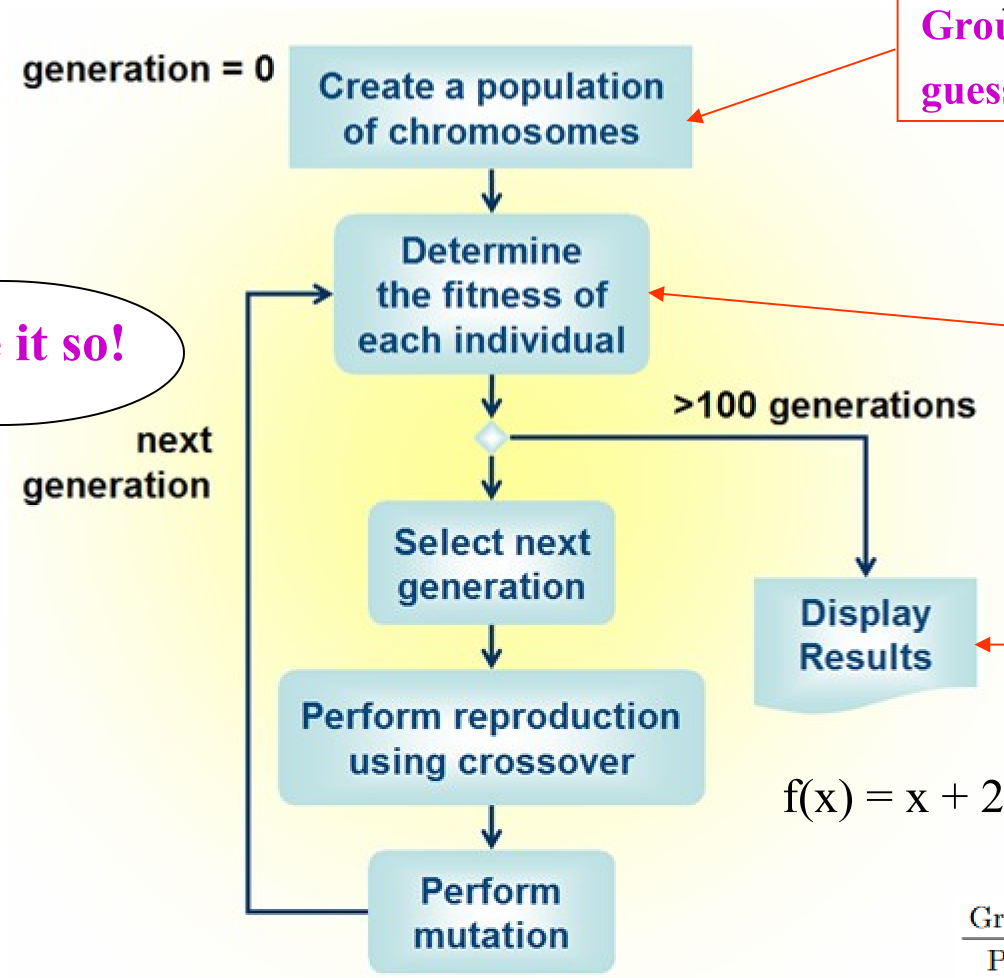


MCMCs do a random walk in parameter space (X_1, X_2 …), while the GA does a random walk in an abstract functional space! In a nutshell:

<u>MCMC</u>		<u>GA</u>
parameters	→	analytic functions
best fit (point)	→	best fit (function)
confidence contours	→	error regions



Genetic Algorithms: Outline



Group of initial random guesses, (aka the "grammar").

Measure of how well it fits the data

Best-fit function that describes the data!

f(x) = x + 2*x^3 * (1+e^x) etc

Grammar type	Functions
Polynomials	$c,\ x,\ 1+x$
Fractions	$\frac{x}{1+x}$
Trigonometric	$\sin(x),\ \cos(x),\ \tan(x)$
Exponentials	$e^x,\ x^x,\ (1+x)^{1+x}$
Logarithms	$\ln(x),\ \ln(1+x)$

S. Nesseris et al, arXiv: 0903.2805, 1205.0364, 1910.01529, 2001.11420, 2106.00428 etc.

The Genetic Algorithms (GA)

Ideal for emulators: eg emulate the sound horizon at the drag redshift

Aizpuru, Arjona, Nesseris: arXiv: 2106.00428

$$r_{\rm s}\left(z_{\rm d}\right)=\frac{1}{H_0}\int_{z_{\rm d}}^{\infty}\frac{c_{\rm s}(z)}{H(z)/H_0}\,{\rm d}z,$$

EH ~2%

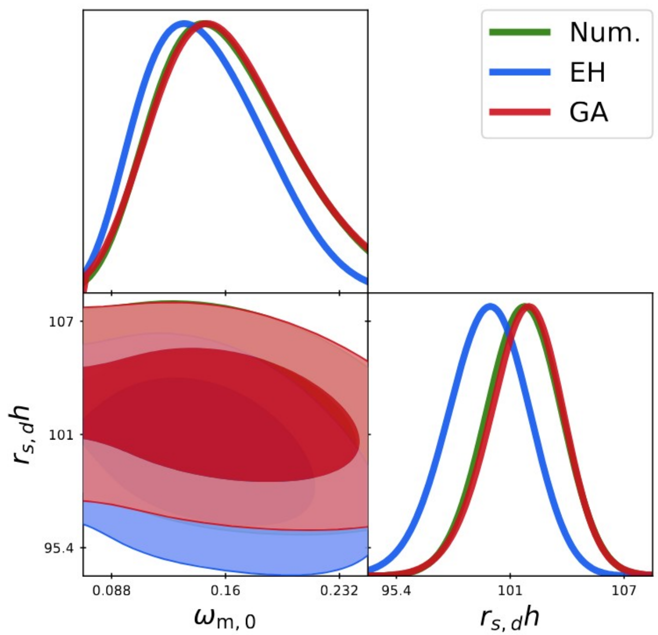
$$r_{\rm s}\left(z_{\rm d}\right)\simeq\frac{44.5\ln\left(\frac{9.83}{\omega_m}\right)}{\sqrt{1+10\,\omega_b^{3/4}}}{\rm Mpc},$$

GA ~0.003%

$$r_{\rm s}(z_{\rm d})=\frac{1}{a_1\omega_b^{a_2}+a_3\omega_m^{a_4}+a_5\omega_b^{a_6}\omega_m^{a_7}}{\rm Mpc}$$

$a_1=0.00785436, a_2=0.177084, a_3=0.00912388,$
$a_4=0.618711, a_5=11.9611, a_6=2.81343,$
$a_7=0.784719.$

EH expression biases BAO analyses by ~0.5σ!



The Genetic Algorithms (GA)

Ideal for null tests: eg test the duality/Etherington relation

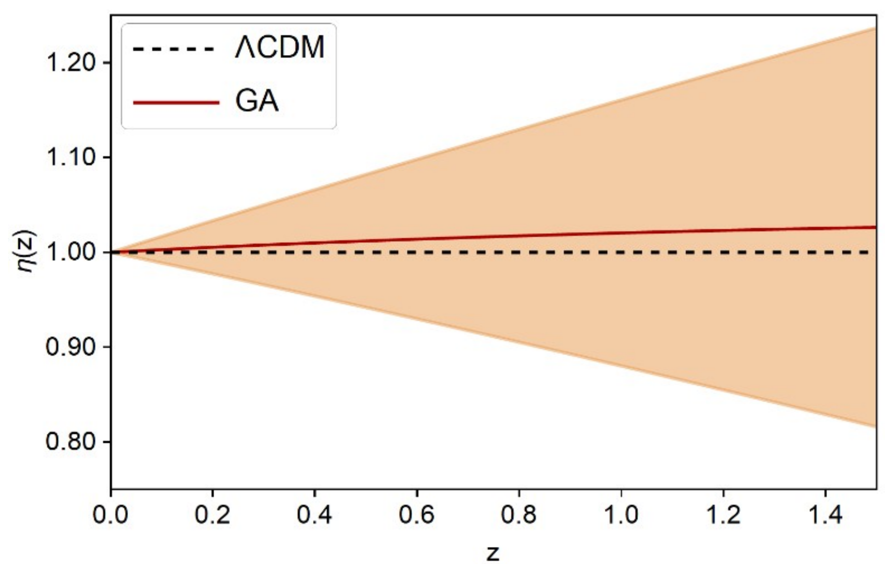
Nesseris et al, arXiv: 2007.16153

$$\eta(z) = \frac{d_{\mathrm{L}}(z)}{d_{\mathrm{A}}(z)(1+z)^2}$$

$= 1$ in GR

$\neq 1$ if number of photons not conserved or not a metric theory of gravity.

Use SnIa to get $d_L(z)$ and BAO to get $d_A(z)$



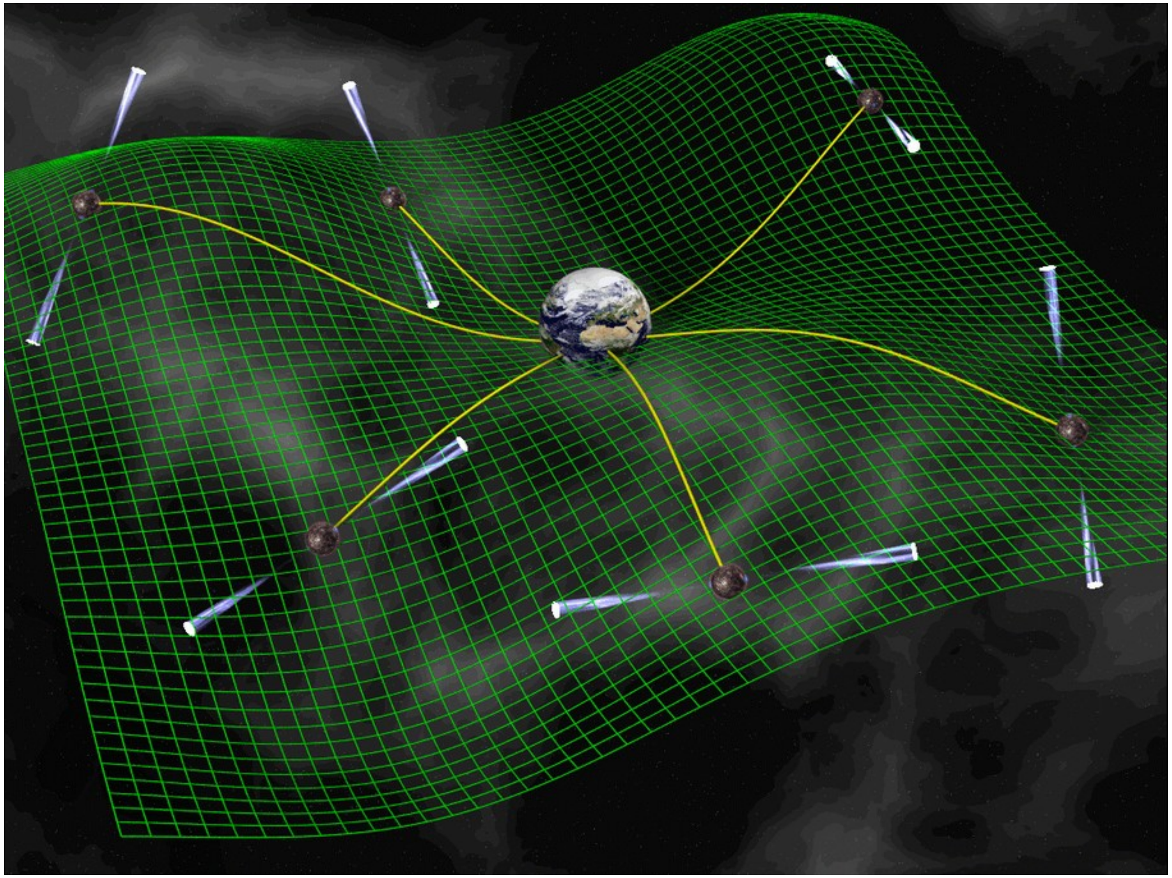
Pantheon SnIa
SDSS etc BAO

No/minimal assumptions on curvature, DM, DE!

CHAPTER 2
THE HELLINGS-DOWNS CURVE

The Hellings-Downs (HD) curve

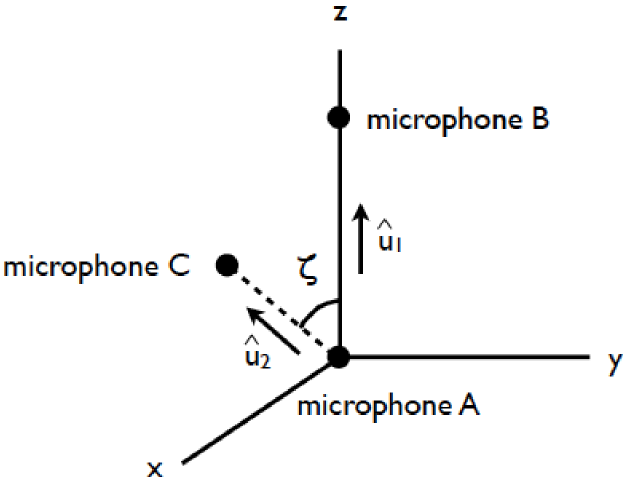
Searching for correlated signatures in the pulsar arrival times on Earth, we get hints for the stochastic gravitational wave background (SGWB):



The Hellings-Downs (HD) curve

The SGWB is similar to the acoustic noise in a bar:
There are several point sources and the correlations can be measured!)

Jenet & Romano, arXiv:1412.1142

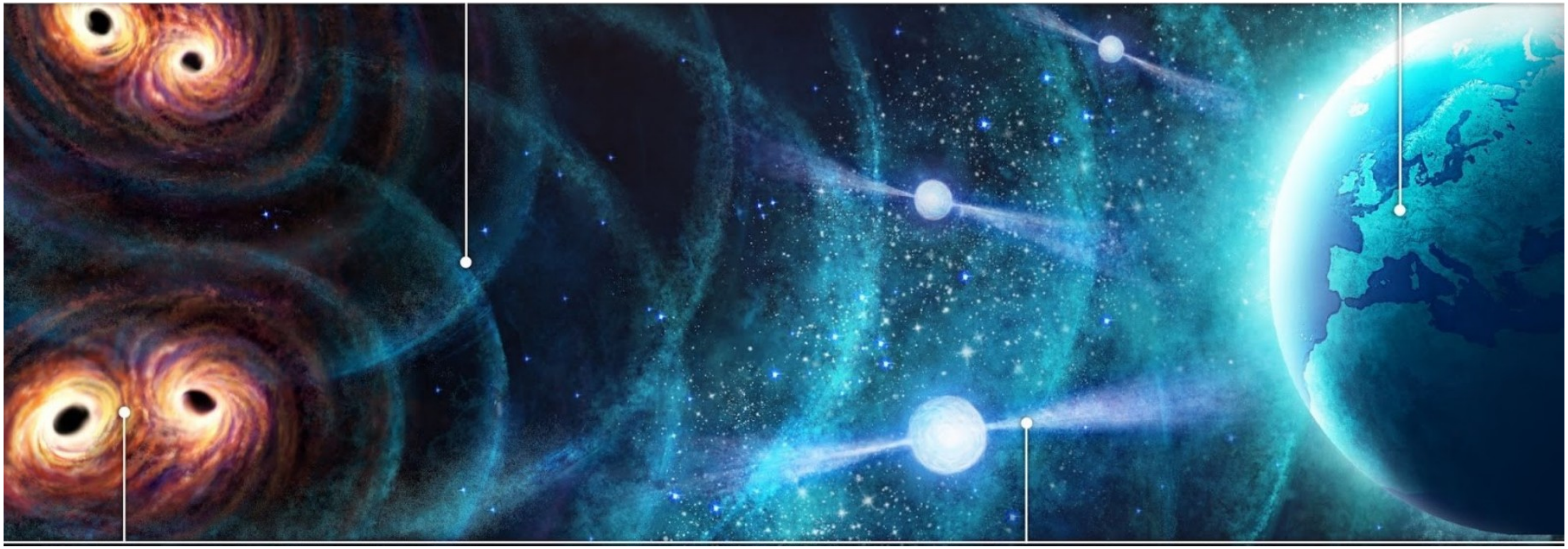


$$\chi(\zeta) \equiv \Gamma_{12}(\omega)$$
$$= \mathrm{Re}\left\{G^2 \frac{1}{4\pi}\int_{S^2} \mathrm{d}^2\Omega_{\hat{\mathbf{k}}} \left[1 - e^{i\mathbf{k}\cdot\mathbf{x}_B} - e^{-i\mathbf{k}\cdot\mathbf{x}_C} + e^{i\mathbf{k}\cdot(\mathbf{x}_B - \mathbf{x}_C)}\right]\right\}$$
$$\simeq G^2$$

The Hellings-Downs (HD) curve

The SGWB produces a distinctive GW signature:

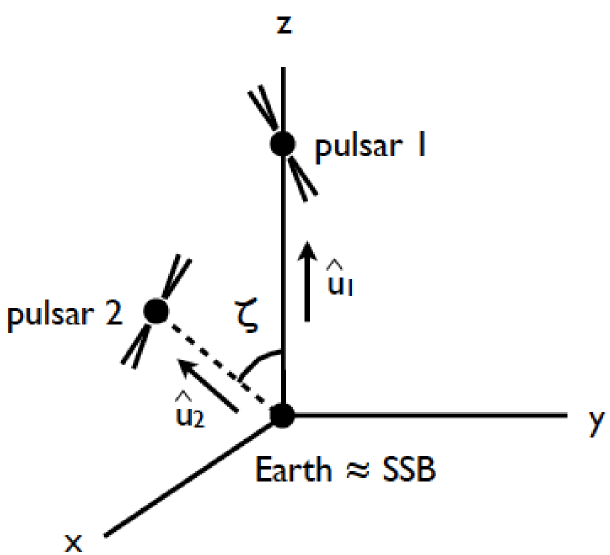
A quadrupolar & higher multipolar spatial correlation between arrival times of pulses, that depends only on the angular separation in the sky.



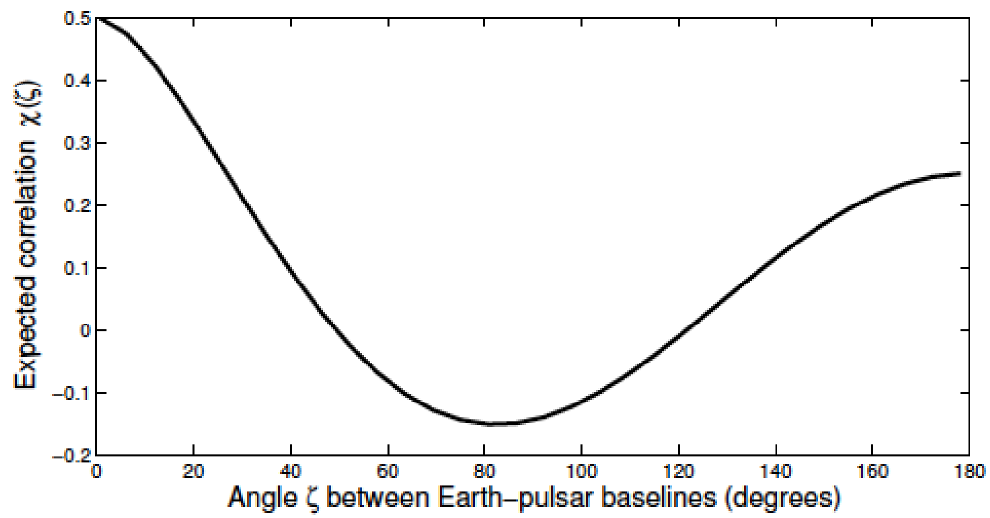
The Hellings-Downs (HD) curve

The signal from the SGWB will be correlated across the sightlines of pulsar pairs, while that from the other noise processes will not.

In GR with standard matter content (baryons + CDM) → the HD curve



Jenet & Romano, arXiv:1412.1142

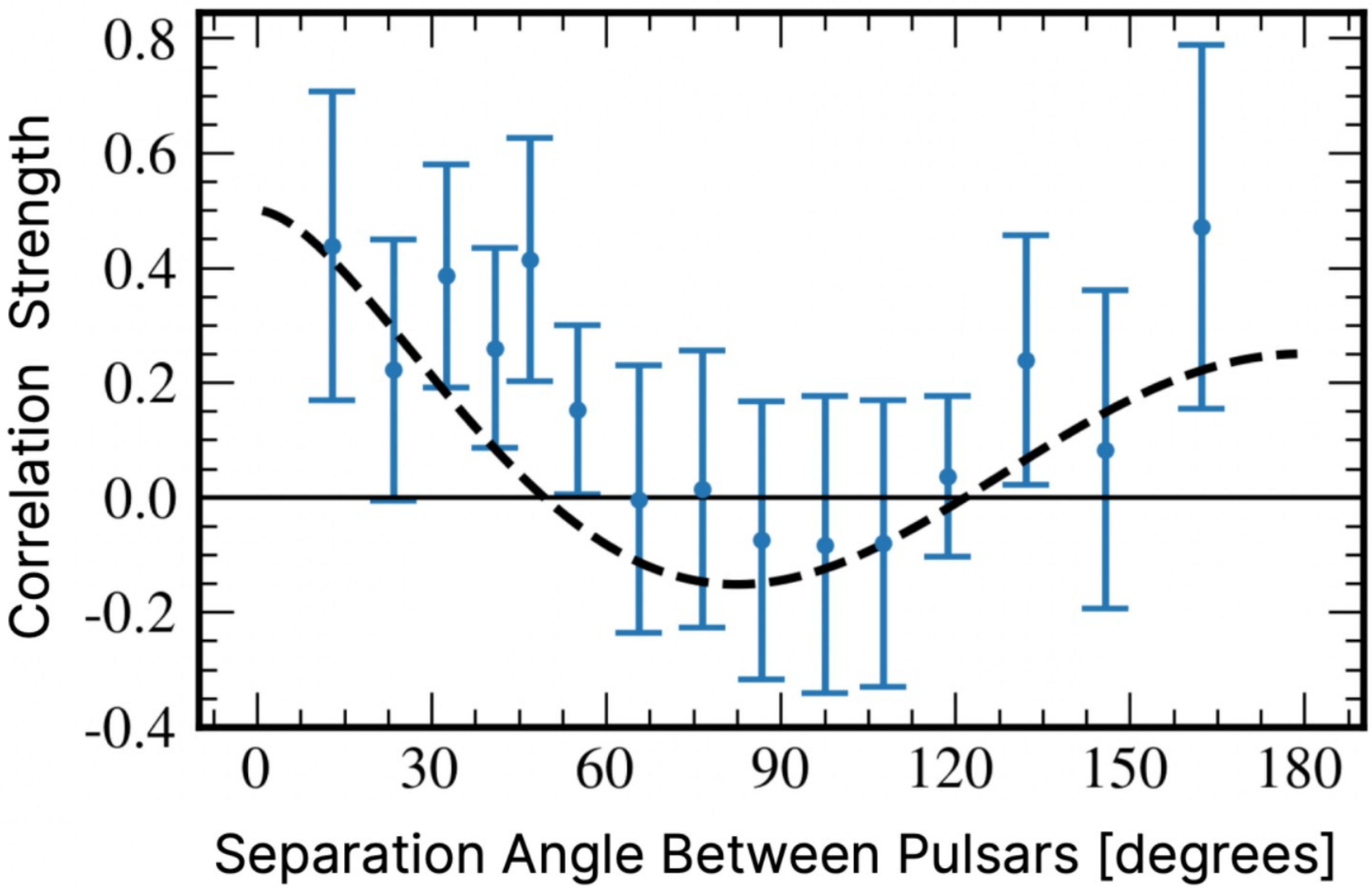


$$\chi(\zeta) = \frac{1}{2} - \frac{1}{4}\left(\frac{1-\cos\zeta}{2}\right) + \frac{3}{2}\left(\frac{1-\cos\zeta}{2}\right)\ln\left(\frac{1-\cos\zeta}{2}\right)$$

Hellings & Downs, Astrophys. J. , vol. 265, pp. L39-L42, 1983

The Hellings-Downs (HD) curve

NANOGrav 15-year data set results (arXiv:2306.16213)



The Hellings-Downs (HD) curve

Some "popular" extensions:

A. Ultralight vector dark matter (Omiya, Nomura & Soda, PRD108, 104006, 2023):
Coherent oscillation of an ultralight bosonic field → fluctuations in the gravitational potential → affect the timing residuals of pulsars.

$$\Gamma_{\text{eff}}(\xi,\mu) = \frac{\Phi_{\text{GW}}(\mu/\pi)}{\Phi_{\text{GW}}(\mu/\pi)+\Phi_{\text{DM}}}\left[\Gamma_{\text{HD}}(\xi)+\frac{\Phi_{\text{DM}}}{\Phi_{\text{GW}}(\mu/\pi)}\Gamma_{\text{DM}}(\xi)\right] \qquad 10^{-24}\text{eV} \lesssim \mu \lesssim 10^{-23}\text{eV}$$

B. Spin-2 ultralight dark matter (Armaleo, Nacir, Urban, arXiv: 2005.03731):
Based on bimetric theory with a massive spin-2 field that is free of the Boulware-Deser ghost

$$\Gamma_{\text{eff}}(\xi,m,\alpha) = \frac{\Phi_{\text{GW}}(m/2\pi)}{\Phi_{\text{GW}}(m/2\pi)+\Phi_{\text{DM}}(\alpha)}\Gamma_{\text{HD}}(\xi) + \frac{\Phi_{\text{DM}}(\alpha)}{\Phi_{\text{GW}}(m/2\pi)+\Phi_{\text{DM}}(\alpha)}\Gamma_{\text{DM}}(\xi)$$

$$\Phi_{\text{GW}}(m/2\pi) \sim 1\times 10^{-32}\text{yr}^2\left(\frac{m}{10^{-22}\text{eV}}\right)^{-\frac{13}{3}}\left(\frac{15\text{yr}}{T_{\text{obs}}}\right)$$

The Hellings-Downs (HD) curve

C. Massive Gravity (Liang & Trodden, PRD 104, 084052, 2021):
Additional polarizations (tensor, vector, and scalar) must be considered

$$\Gamma_{\rm eff}(\xi, A, \Omega) = \Gamma_T(\xi, A) + \Omega \cdot \Gamma_V(\xi, A) + \Omega \cdot \Gamma_S(\xi, A),$$

$$A = \frac{|\boldsymbol{k}|}{k_0} \quad k_0^2 = |\boldsymbol{k}|^2 + m^2$$

D. Non-Gaussian component to SGWB (higher-order correlations between pulsars), see Jiang & Piao, arXiv:2401.16950

$$\Gamma\left(\zeta_{ab}\right) \propto$$

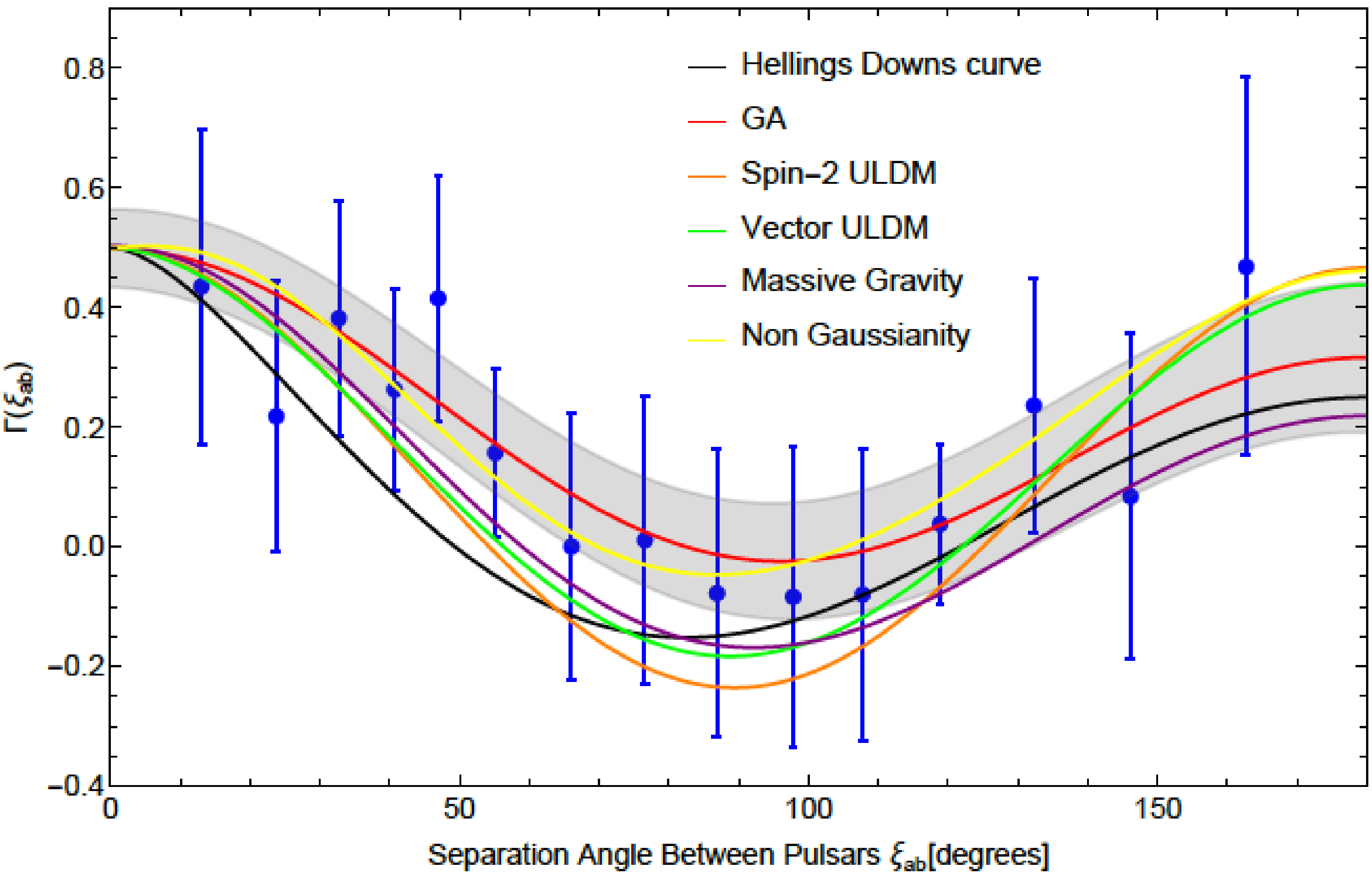
$$\sum_{A=+,\times} \int_{S^2} \mathrm{d}\hat{\mathrm{n}} \left\{ \frac{\bar{E}_a^A \bar{E}_b^A}{\left(1+\hat{\mathrm{n}}_a \cdot \hat{\mathrm{n}}\right)\left(1+\hat{\mathrm{n}}_b \cdot \hat{\mathrm{n}}\right)} \right.$$

$$\left. + \frac{4\alpha\left[\frac{9}{16} \bar{E}_a^A \bar{E}_a^A \bar{E}_b^A \bar{E}_b^A + \frac{5}{8}\left(\bar{E}_a^A \bar{E}_a^A \bar{E}_a^A \bar{E}_b^A + \bar{E}_a^A \bar{E}_b^A \bar{E}_b^A \bar{E}_b^A\right)\right]}{\left(1+\hat{\mathrm{n}}_a \cdot \hat{\mathrm{n}}\right)\left(1+\hat{\mathrm{n}}_b \cdot \hat{\mathrm{n}}\right)} \right\}$$

$$\bar{E}_a^A = \mathrm{e}_{ij}^A \, \hat{\mathrm{n}}_a^i \, \hat{\mathrm{n}}_a^j$$

Comparison of HD extensions

A Bayesian analysis of the models given the NANOGrav data (just playing with the theory!)



Comparison of HD extensions

Bayesian model comparison (via the Savage-Dickey formula)

$$B \equiv \frac{E_{M'}}{E_M}$$

The simpler model (HD)

A bigger (nested) model

$$\chi(\zeta) = \frac{1}{2} - \frac{1}{4}\left(\frac{1-\cos\zeta}{2}\right) + \frac{3}{2}\left(\frac{1-\cos\zeta}{2}\right)\ln\left(\frac{1-\cos\zeta}{2}\right)$$

B > 1 (logB > 0) HD wins
B < 1 (logB < 0) HD loses

$$\ln B \simeq -\frac{1}{2}\left(\chi'^2_{\min} - \chi^2_{\min}\right) + \ln|F_{pp}|^{1/2} + \ln\left[\frac{\Pi^p_{i=n'}\Delta\theta_i}{(2\pi)^{p/2}}\right]$$

B_{ij}	$\ln B_{ij}$	Evidence
< 2.5	< 1.1	Comparable
< 20	< 3	Weak
< 150	< 5	Moderate
> 150	> 5	Strong

Models	B	$\ln(B)$	χ^2
Hellings-Downs	-	-	10.25
Genetic Algorithms	-	-	2.83
Ultralight vector DM	0.29	-1.24	6.39
Spin-2 ultralight DM	0.80	-0.22	8.43
Massive gravity	1.51	0.41	6.91
Non-Gaussianity of SGWB	0.06	-2.85	3.17

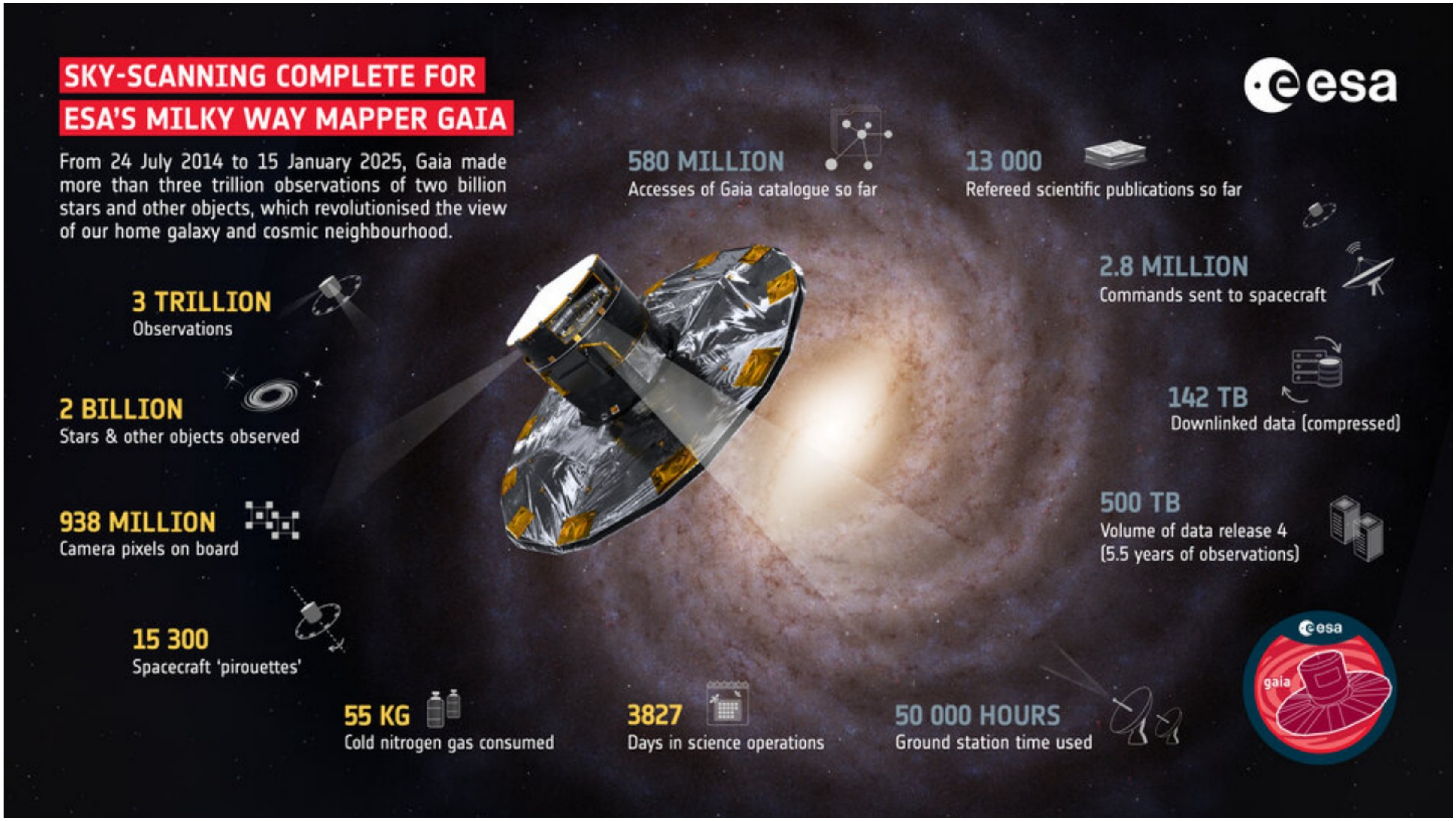
GA beats all!
But no Bayes...

Barely weak support!

CHAPTER 3
ASTROMETRY & GWs

Astrometry & GAIA

GAIA: 2 billion of the brightest stars, each star gets measured on average 75 times every five years plus proper motions.



Astrometry & GAIA

GWs in the vicinity of Earth induce correlated distortions in the apparent positions and proper motions of distant sources.

Astrometry & GAIA

The detection (or not) of a coherent behavior in astrometric data enables the measurement of the SGWB in $10^{-16} < f/Hz < 10^{-9}$

Astrometry & GAIA

Assuming an isotropic, unpolarised, stationary SGWB:

$$\Omega_{\rm GW}(f) = \frac{1}{\rho_c}\frac{d\rho_{\rm GW}}{d(\ln f)}$$

and this can be related to the stars' proper motion $\mu^2 = \mu_\delta^2 + \mu_\alpha^2 \cos\delta^2$

$$\langle\mu^2\rangle \approx \frac{H_0^2}{8\pi^2}\Omega_{\rm GW}\int_{f_{\rm min}}^{f_{\rm max}}\frac{df}{f^3}$$

$$\Omega_{\rm GW} \simeq \frac{6}{5}\frac{1}{4\pi}\frac{P_2}{H_0^2}$$

$$\Omega_{\rm GW} \lesssim \frac{\Delta\mu^2}{NH_0^2}$$

$$\Omega_{\rm GW}(f \sim 10^{-8}\ {\rm Hz}) \lesssim 0.4$$

where T is the total observing time and Δt the cadence (typical time between successive position measurements).

$$f_{\rm min} \sim 1/T$$

$$f_{\rm max} \sim 1/\Delta t$$

Darling et al., Astrophysical Journal, 861, 113

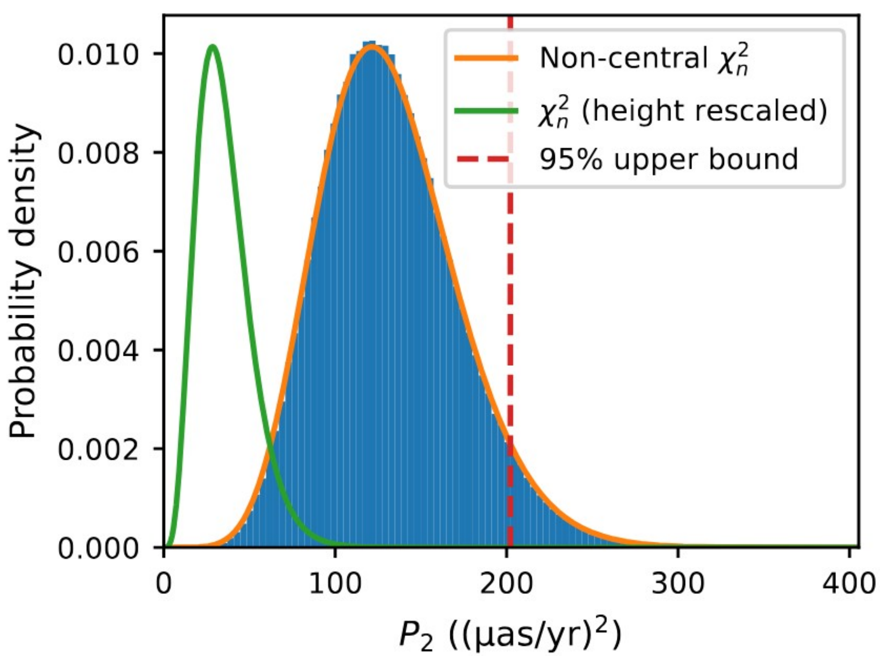
Astrometry & GAIA

Standard analysis constraints from GAIA DR3

S. Jaraba, J. Garcia-Bellido et al., arXiv: 2304.06350

$$h_{70}^2 \Omega_{\rm GW} \lesssim 0.087 \qquad 4.2 \times 10^{-18}\ \text{Hz} \lesssim f \lesssim 1.1 \times 10^{-8}\ \text{Hz}$$

the quadrupole power P2



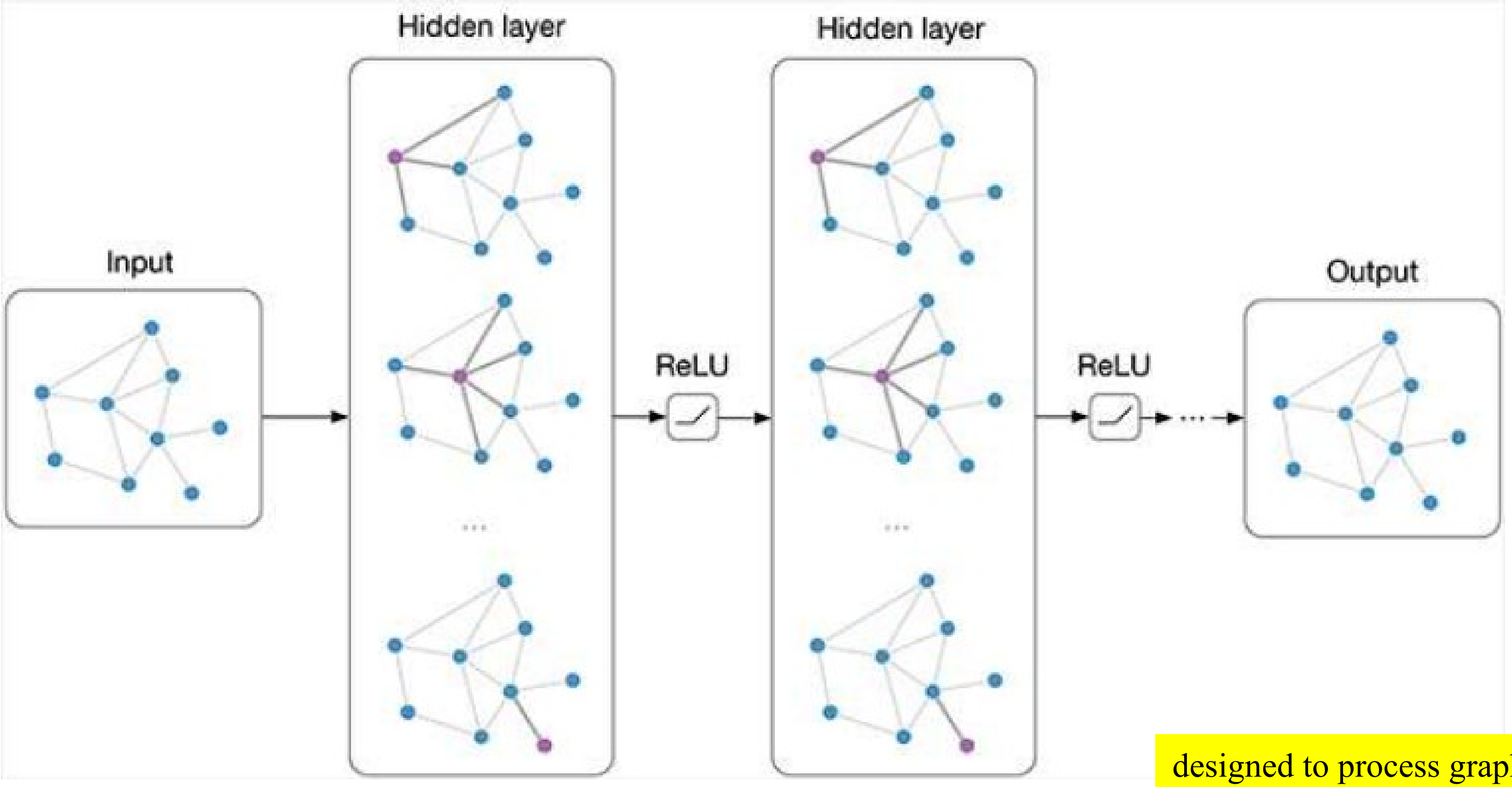
GAIA may reach proper motion of 200 μas/yr and $N = 5 \times 10$^5 thus → $\Omega_{\rm GW}$<10^-3

Astrometric constraints with NNs

Two architectures to measure Ω_{GW}:

A graph NN (GNN)

GNNs are used in protein folding, social networks etc

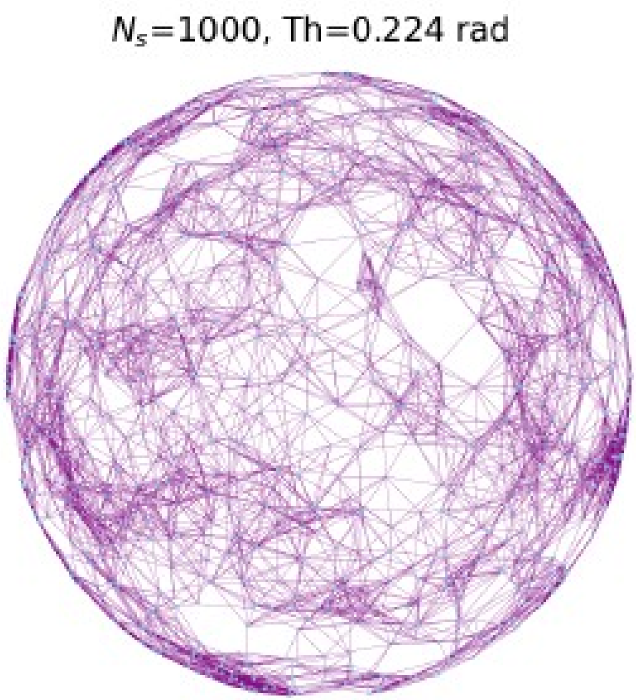


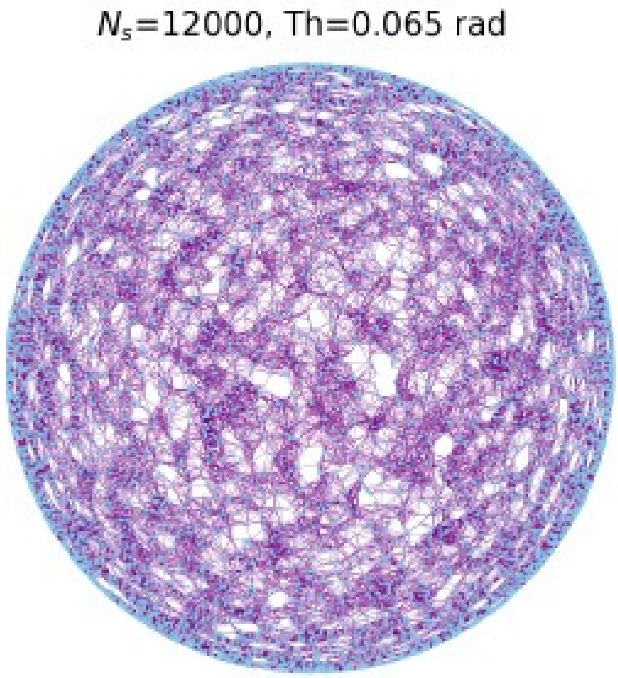
designed to process graph-structured data and capture relationships between elements

Astrometric constraints with NNs

Two architectures to measure Ω_{GW}:

A graph NN (GNN) with mock GAIA data for various Ω_{GW} values



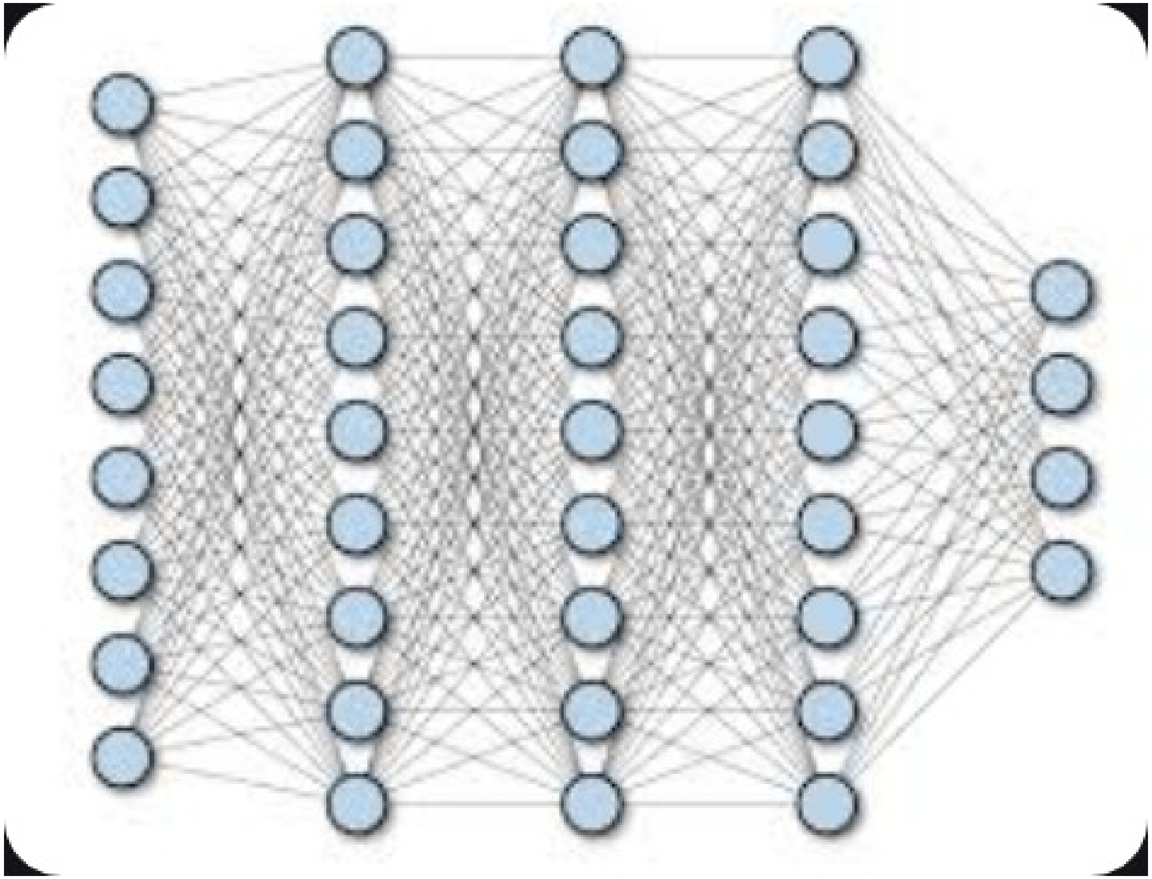


Different radial thresholds (correlation lengths) and number of stars.

Astrometric constraints with NNs

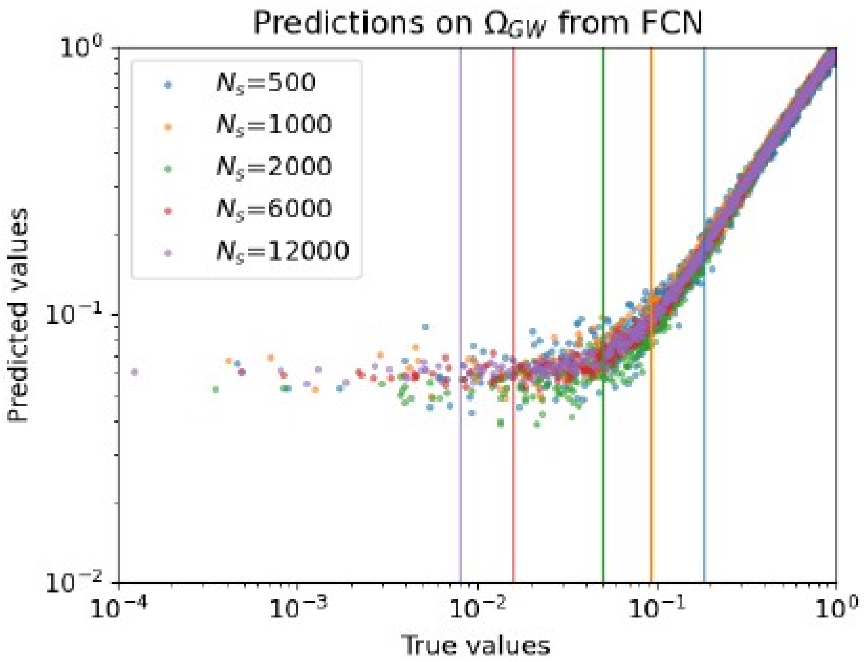
Two architectures to measure Ω_{GW}:

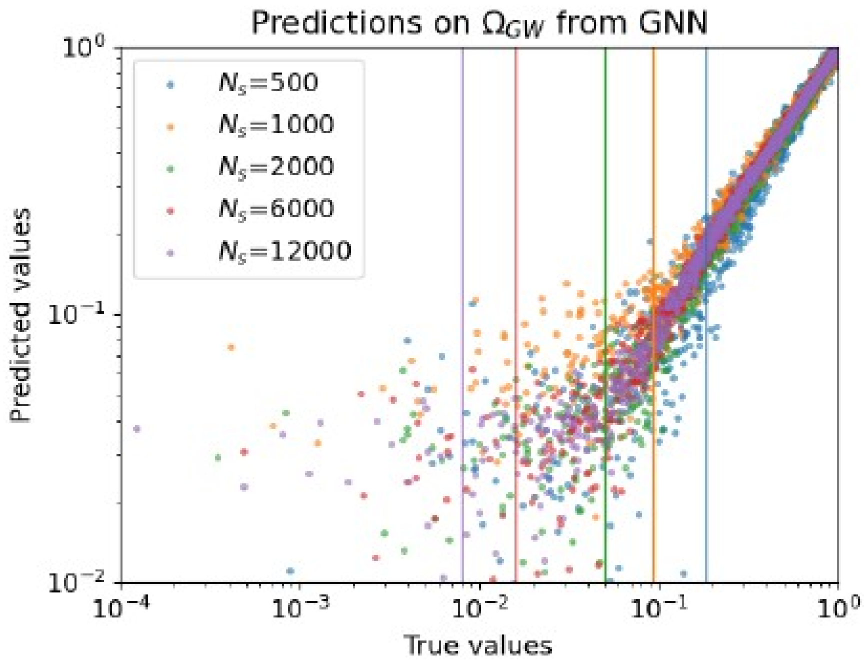
A fully connected network (FCN)



Astrometric constraints with NNs

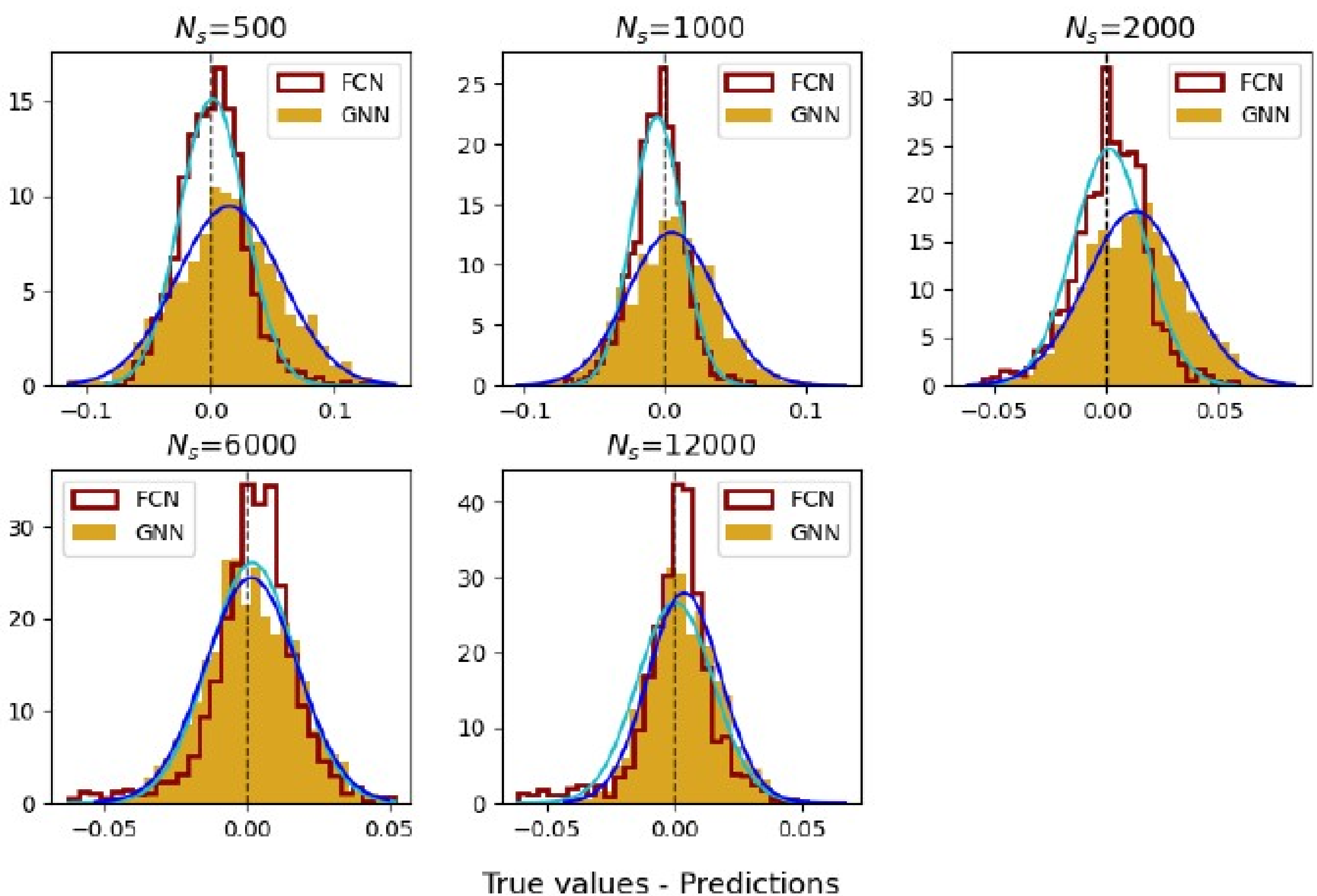
Try to predict GW density param Ω_{GW}





Astrometric constraints with NNs

Distributions of scatter



Summary

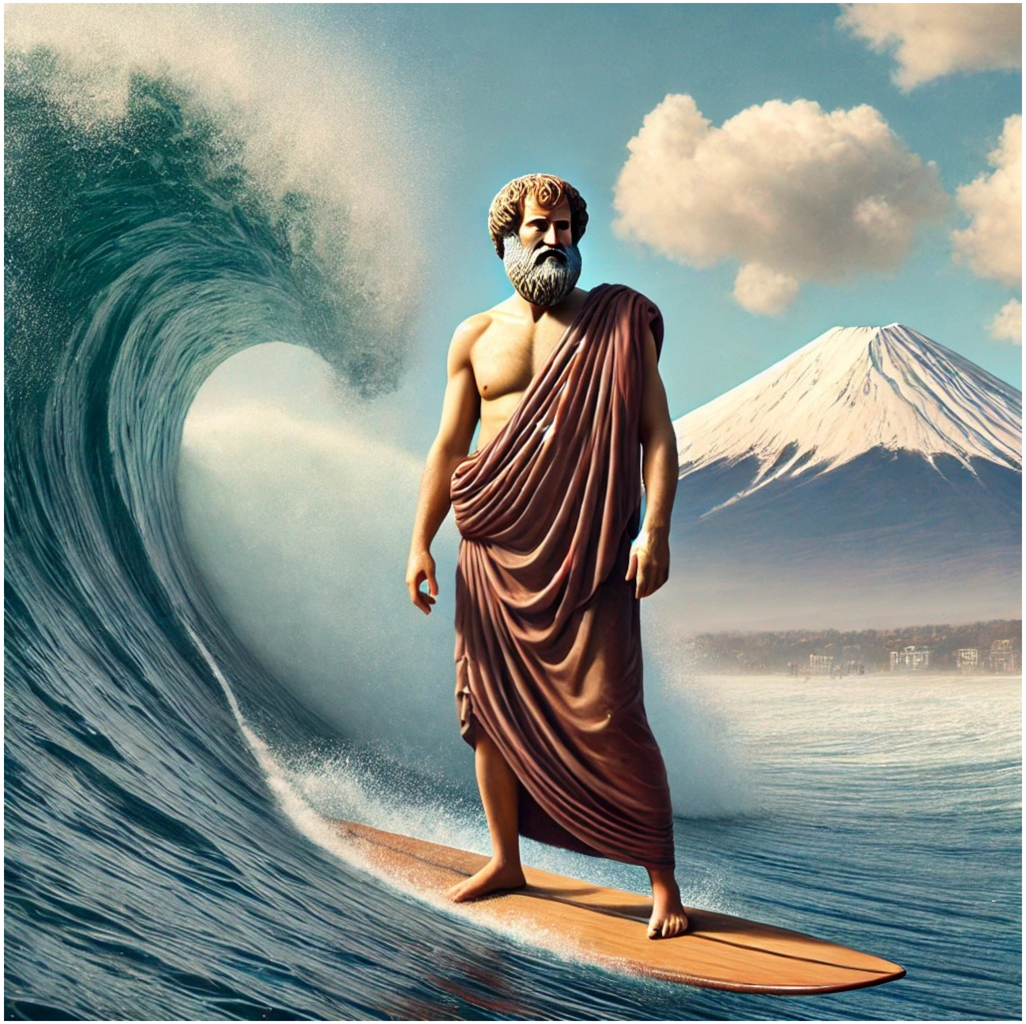
Physics nowadays has three pillars:
theory, experiment & simulations/AI.

New fascinating ML-AI tools that can help with complex (e.g. avoid specific modeling etc).

Constraints on extended HD models (early data, so nothing concrete and no claims of discovery).

NNs can help predict Ω_{GW} from GAIA observations (tests with small samples, they work well!).

Lots of exciting work to do in the near future!



Thanks!

Backup slides

Genetic Algorithms: Selection

There are two possible "Selection" methods:

1) **Roulette wheel selection** **(the selection probability is proportional to the fitness)**
   - widely used method, but…
   - a suboptimal solution may dominate the mating pool
   - premature convergence…

2) **Tournament selection** **(find best individual)**
   - easy to implement
   - in line with natural selection
   - premature convergence avoided

Model comparison and the data...

Main problems:

1) We can only test a limited number theories (e.g. Horndeski, f(R), extra dims, etc) as there are practically infinite number of possibilities. It's impossible to test everything!

2) Model bias: interpretation of the results depends on the chosen models+assumptions, e.g. using ΛCDM to find Ω_m=0.315 is a model-dependent statement!

Is there any (good) solution?

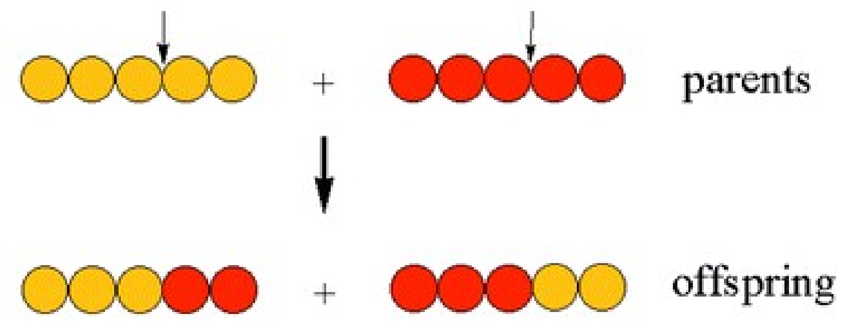
i) Use an EFT or effective fluid approach… → (top-down)

ii) Use machine learning methods (e.g. the GA) → (bottom-up)

Genetic Algorithms: Reproduction

Reproduction can be done in two ways:

1) Crossover:



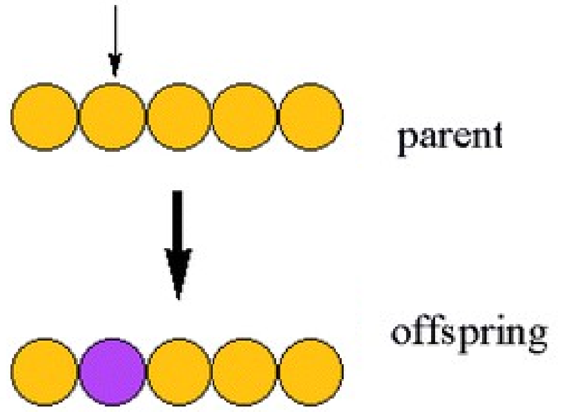
$$\mu_{GA,1}(z) = ln(z) \qquad \mu_{GA,2}(z) = -1 + z + z^2$$

$$\downarrow$$

$$\begin{aligned}\mu_{GA,1}(z) \oplus \mu_{GA,2}(z) &\to (\bar{\mu}_{GA,1}(z), \bar{\mu}_{GA,2}(z), \bar{\mu}_{GA,3}(z)) \\ &= (ln(z^2), -1 + ln(z^2), -1 + ln(z))\end{aligned}$$

2) Mutation:

$$\bar{\mu}_{GA,3}(z) = -1 + ln(z) \to -1 + ln(z^3)$$



Genetic Algorithms: Error estimation

S. Nesseris et al, arXiv: 1205.0364,1210.7652

1) Define a likelihood similarly to parametric approach :

$$\mathcal{L} = \mathcal{N} \exp\left(-\chi^2(\vec{a})/2\right)$$

$$\chi^2(\vec{a}) \equiv \sum_{i=1}^{N} \left(\frac{y_i - y(x_i; a_j)}{\sigma_i} \right)^2$$

$$\chi^2(\vec{a}) = \chi^2_{\text{min}} + (a - a_{\text{min}})_i \; F_{ij} \; (a - a_{\text{min}})_j$$

GA

$$\mathcal{L} = \mathcal{N} \exp\left(-\chi^2(f)/2\right)$$

$$\chi^2(f) \equiv \sum_{i=1}^{N} \left(\frac{y_i - f(x_i)}{\sigma_i} \right)^2$$

2) Normalize the likelihood by integrating over all functions, i.e. do a "path integral" **(in a parametric approach one integrates over all parameters)**

$$\int \mathcal{L} d\vec{a} = \int_{-\infty}^{\infty} \mathcal{N} \exp\left(-\chi^2(\vec{a})/2\right) d\vec{a} \; =$$

$$\mathcal{N} e^{-\chi^2_{\text{min}}/2} \int_{-\infty}^{\infty} e^{-1/2 (a-a_{\text{min}})_i \; F_{ij} \; (a-a_{\text{min}})_j} da_0 da_1 ... da_{M-1} \; = \; 1$$

$$\mathcal{N} = e^{\chi^2_{\text{min}}/2} \, |F|^{1/2} \, (2\pi)^{-M/2}$$

GA

$$\int \mathfrak{D}f \; \mathcal{L} = \int \mathfrak{D}f \; \mathcal{N} \exp\left(-\chi^2(f)/2\right) = 1$$

$$\mathfrak{D}f = \prod_{i=1}^{N} df_i$$

$$\mathcal{N} = \left((2\pi)^{N/2} \, |C|^{1/2} \right)^{-1}$$

Genetic Algorithms: Error estimation

3) Find the confidence interval by integrating around the best-fit ±1σ

$$
\begin{aligned}
CI(x_i,\delta f_i) &= \int_{f_{bf}(x_i)-\delta f_i}^{f_{bf}(x_i)+\delta f_i} df_i \frac{1}{(2\pi)^{1/2}\sigma_i} \exp\left(-\frac{1}{2}\left(\frac{y_i-f_i}{\sigma_i}\right)^2\right) \\
&= \frac{1}{2}\left(\mathrm{erf}\left(\frac{\delta f_i + f_{bf}(x_i) - y_i}{\sqrt{2}\sigma_i}\right) + \mathrm{erf}\left(\frac{\delta f_i - f_{bf}(x_i) + y_i}{\sqrt{2}\sigma_i}\right)\right) = \mathrm{erf}\left(1/\sqrt{2}\right)
\end{aligned}
$$

4) 1σ error region for $f_{bf}(x)$: $[f_{bf}(x)-\delta f(x), f_{bf}(x)+\delta f(x)]$

5) Generalize for correlated data:

$$\chi^2 = \sum^{N} (y_i - f(x_i))\, C_{ij}^{-1}\, (y_j - f(x_j))$$

$$\mathcal{L} = \frac{1}{(2\pi)^{N/2}\,|C|^{1/2}} \exp\left(-\chi^2(f)/2\right)$$

S. Nesseris et al, arXiv: 1205.0364