thanks for
the logo btw

Updates on Galapy

a fast API for modelling galaxy SEDs with Bayesian sampling

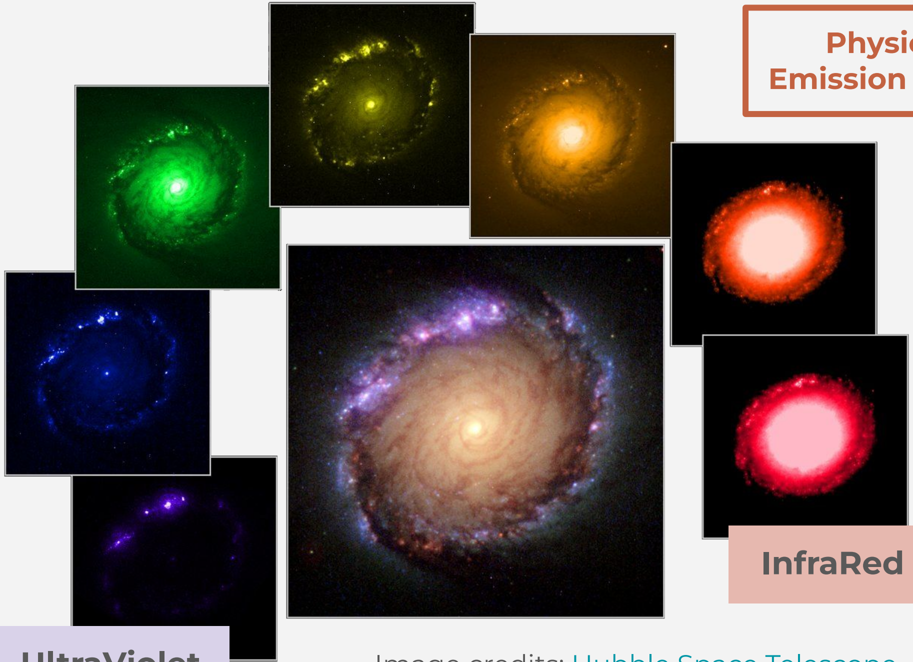
Tommaso Ronconi

@Spoke3-WP1&2 monthly meeting

tronconi@sissa.it

Extract (astro-)Physical Info from the SED of galaxies

At different **wavelengths**
➔ different **components** dominate

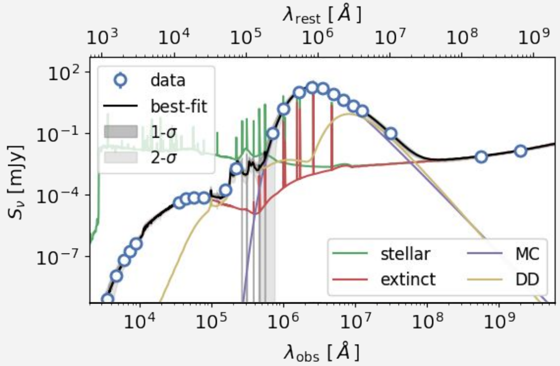


UltraViolet

InfraRed

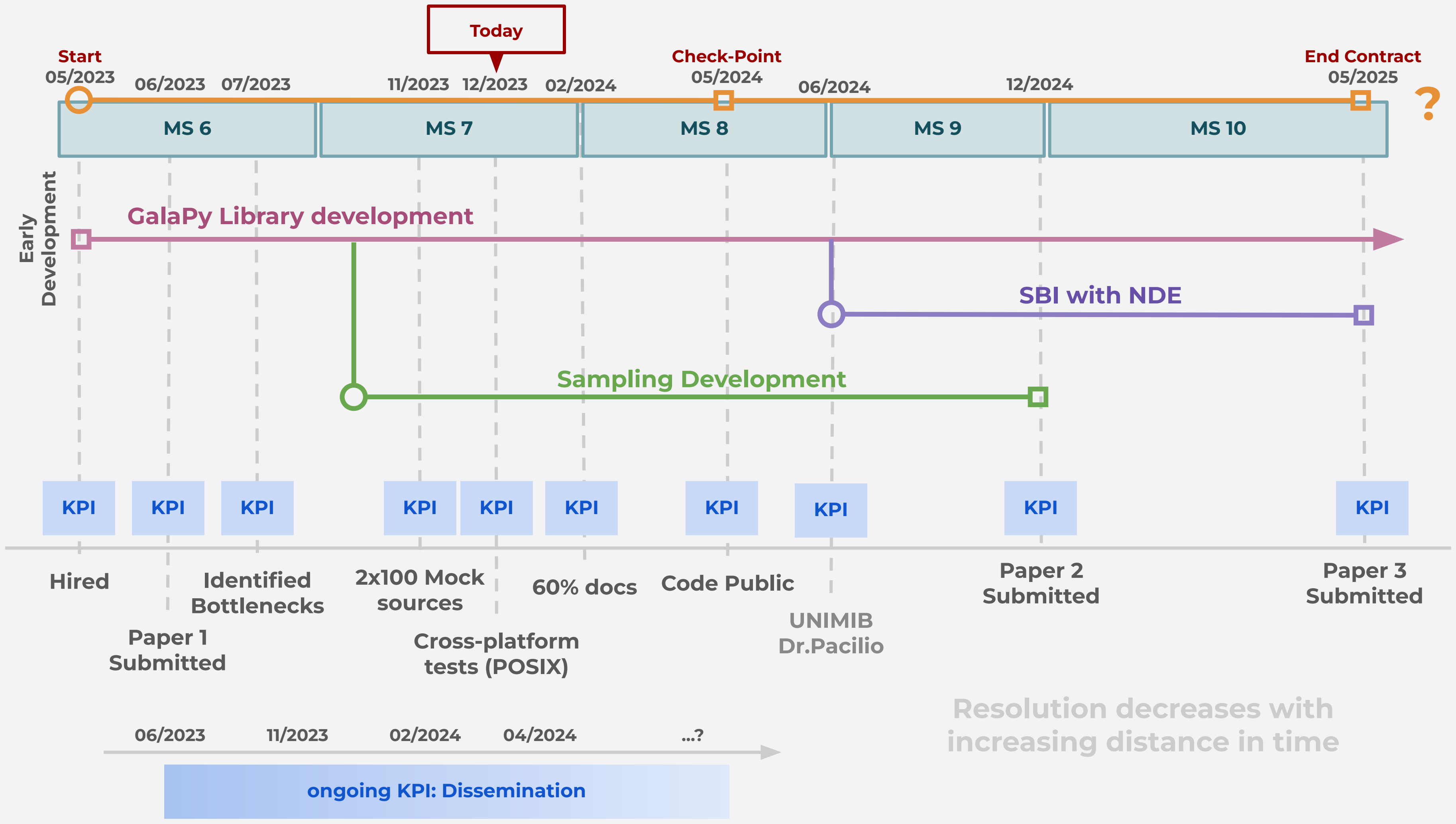
Image credits: Hubble Space Telescope

Physical Emission Model



Spectral Energy Distribution (SED)

➔ **1 zettabyte** of data per year!


Today
Start
05/2023
06/2023
07/2023
11/2023
12/2023
02/2024
Check-Point
05/2024
06/2024
12/2024
End Contract
05/2025
?
MS 6
MS 7
MS 8
MS 9
MS 10
Early Development
GalaPy Library development
SBI with NDE
Sampling Development
KPI
Hired
Paper 1 Submitted
Identified Bottlenecks
2x100 Mock sources
Cross-platform tests (POSIX)
60% docs
Code Public
UNIMIB Dr.Pacilio
Paper 2 Submitted
Paper 3 Submitted
06/2023
11/2023
02/2024
04/2024
...?
ongoing KPI: Dissemination
Resolution decreases with increasing distance in time

Key Performance Indicators

MS6 & MS7

GalaPy presentation paper in review

KPI

SUBMITTED!

Soon public on GitHub and PyPI

New modelling

- **Analytical solutions for Galaxy Evolution**:
  realistic modelling of dust absorption/emission
  Gas, Stars, Metals and Dust
  (Lapi et al., 2018; 2020, Pantoni et al., 2019)
- **2-components dust model**:
  as seen in GRASIL (Silva et al., 1998)
  but AGE DEPENDENT ENERGY BALANCE
  (no radiative transfer + physical temperature)

Performance

- **Hybrid C++/CPython** implementation (performance + user friendly)
- First release: **shared memory parallelization** from Python

Bayesian framework

First release:

- **Markov Chain Monte Carlo** with `emcee` Foreman-Mackey et al., 2013
- **Dynamic Nested Sampling** with `dynesty` Speagle, 2019

Validation on 2x100 mock sources

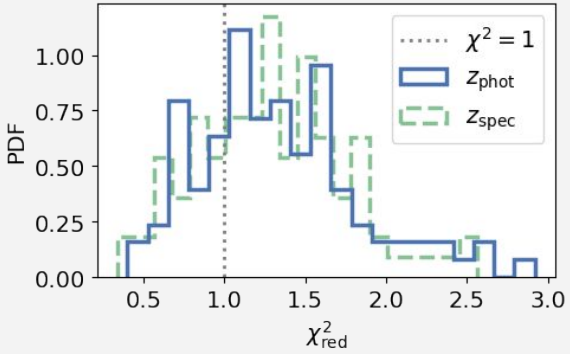
KPI

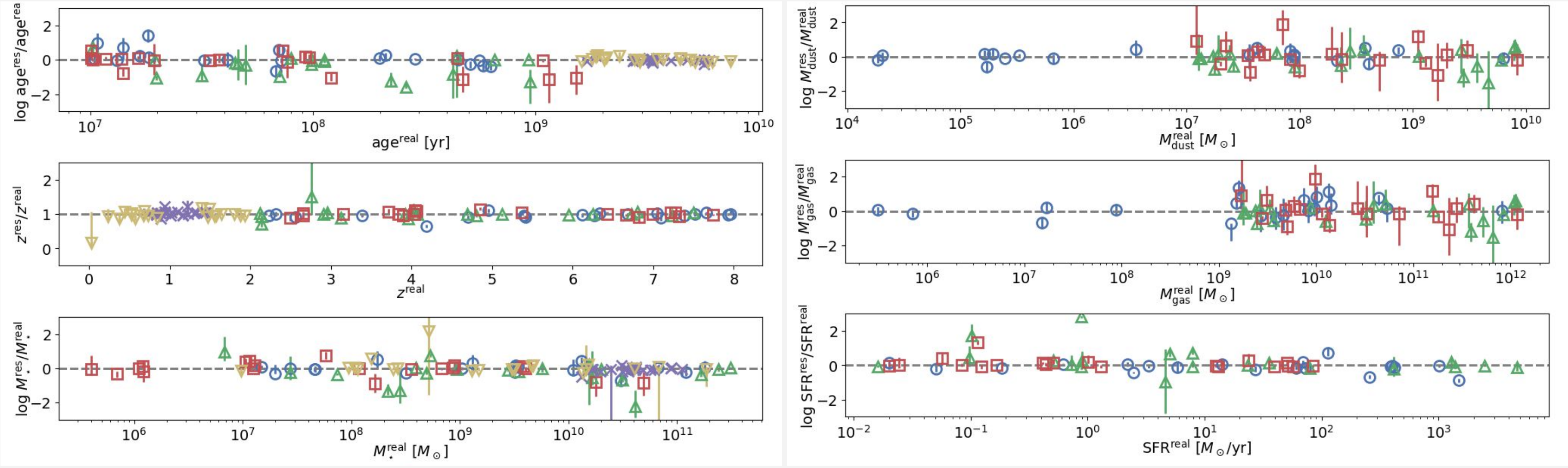
Different evolution phases

Different Models

	ACTIVE	PASSIVE
In-Situ		
Delayed-Exp.		
Constant		

- **+ Uninformative priors**
- **+ Different experiments configuration**





1st Bottleneck = CSP computation

KPI

SSP

- wavelength **Nλ >2000**
- age of SSP **Nτ >100**
- metallicity **NZ ~10**

>10^6 double precision ⇒ **~10 MB**

⊗

SFH

- at a given **AGE**
  - SFR(**τ<AGE**)
  - Z_{stars}(**τ<AGE**)
  - idx_Z (**τ<AGE**)

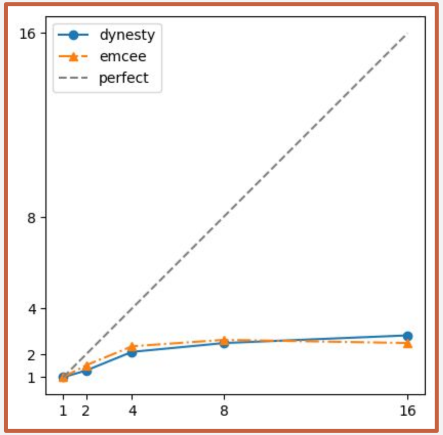
$$L_{\rm CSP}(\lambda_i, \tau_{\rm GXY}) = \sum_{\forall\, j>0\,|\,\tau_j \lesssim \tau_{\rm GXY}} \frac{\tau_j - \tau_{j-1}}{2} \times$$

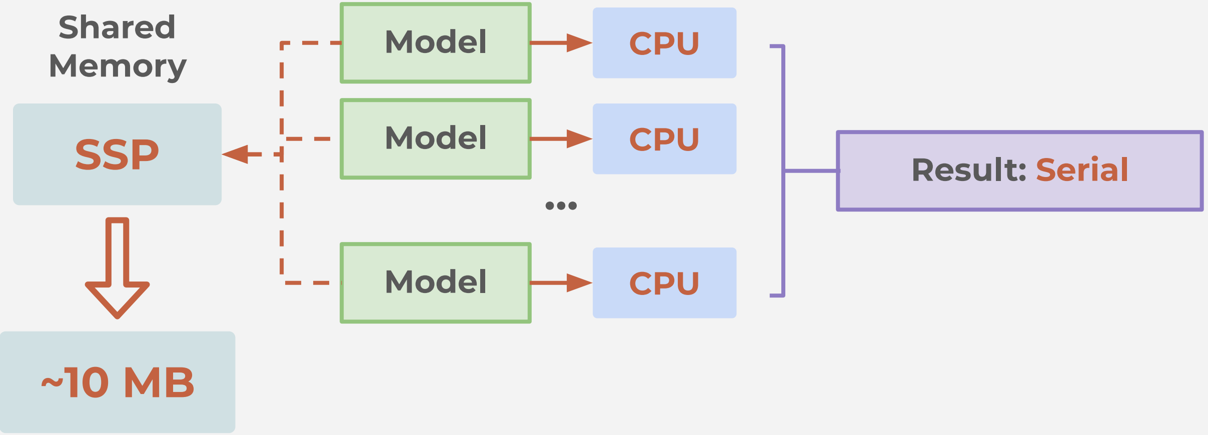
$$\times \left\{ \psi(\tau_j)\; P^{(1)}_{L_{\rm SSP}}\left[\lambda_i, \tau_j, Z_\star(\tau_{\rm GXY} - \tau_j)\right] + \right.$$

$$\left. + \psi(\tau_{j-1})\; P^{(1)}_{L_{\rm SSP}}\left[\lambda_i, \tau_{j-1}, Z_\star(\tau_{\rm GXY} - \tau_{j-1})\right] \right\}$$

2nd “Bottleneck” = scaling, but why?

KPI






```
$ lscpu
[...]
Caches (sum of all):
  L1d:     256 KiB (8 instances) ->  32 KiB
  L1i:     256 KiB (8 instances) ->  32 KiB
  L2:      2 MiB (8 instances)   -> 256 KiB
  L3:      16 MiB (1 instance)
[...]
```

- It is still bugging me the way memory is dealt by the processes in particular, **access to the SSP matrix**
- my two cents: since the matrix cannot be loaded all in L1/L2 private caches, other processes **trigger cache misses**
- is it therefore a problem of **cache invalidation?**

Others on MS7

KPI

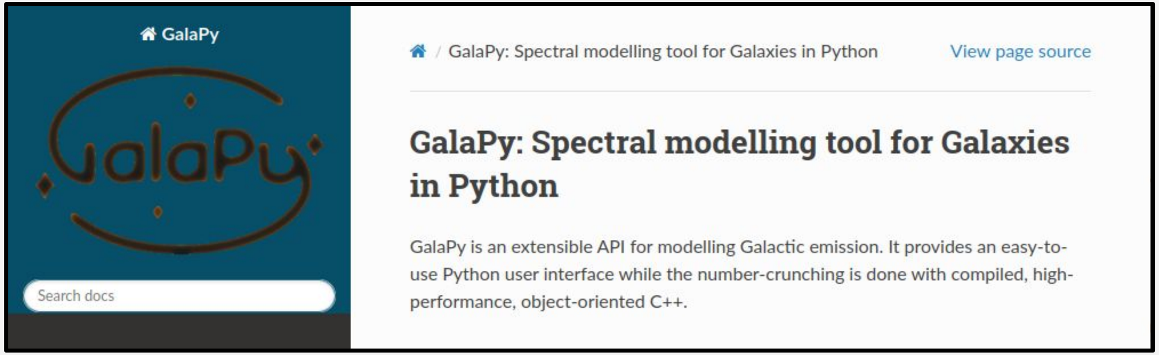
Cross-Platform testing

Currently:
- Linux
- MacOS

Missing:
- Windows (..sigh)

KPI

~60% Documentation completed



Currently ~30%

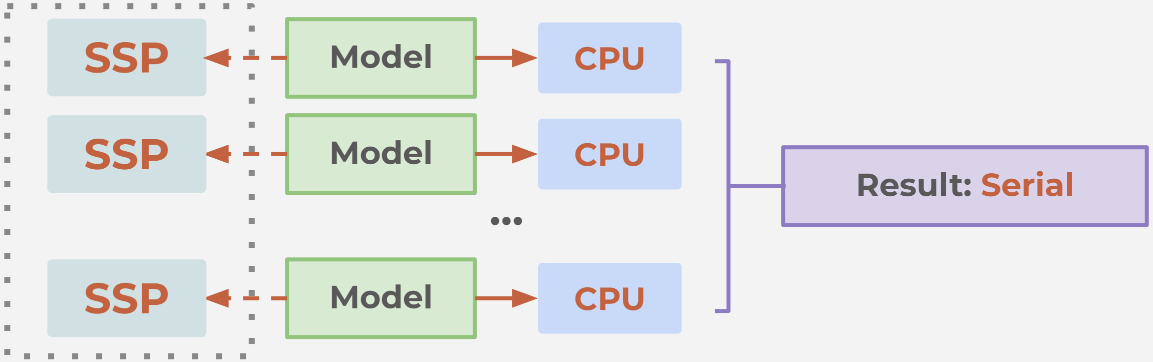
(foreseen) KPIs for MS8

- Paper 1 + Code public
- Parallelisation development

Foreseen KPIs for MS8

KPI

from 1 shared object to **MULTIPLE SHARED OBJECTS** **(1 per CPU)**

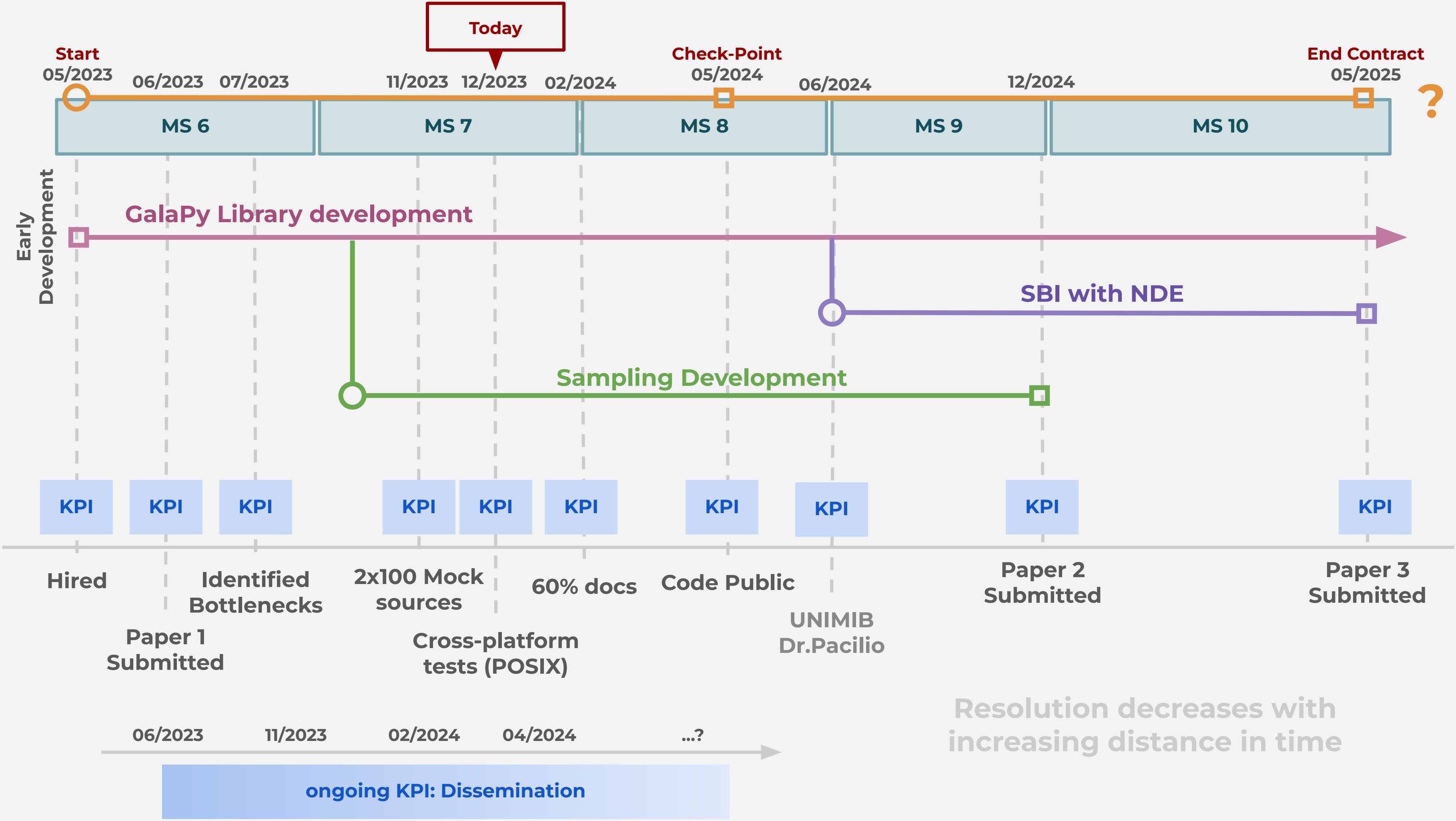


Managed by a Pool of processes

- **1 source: multiple positions** in parameter space
- **Multiple sources: multiple models** sampled serially

→ **Sampling devs** **KPI**

(extends up to end of MS9)


Today
Start
05/2023
06/2023
07/2023
11/2023
12/2023
02/2024
Check-Point
05/2024
06/2024
12/2024
End Contract
05/2025
?
MS 6
MS 7
MS 8
MS 9
MS 10
Early Development
GalaPy Library development
SBI with NDE
Sampling Development
KPI
KPI
KPI
KPI
KPI
KPI
KPI
KPI
KPI
KPI
Hired
Paper 1 Submitted
Identified Bottlenecks
2x100 Mock sources
Cross-platform tests (POSIX)
60% docs
Code Public
UNIMIB Dr.Pacilio
Paper 2 Submitted
Paper 3 Submitted
06/2023
11/2023
02/2024
04/2024
...?
ongoing KPI: Dissemination
Resolution decreases with increasing distance in time

ValaPy

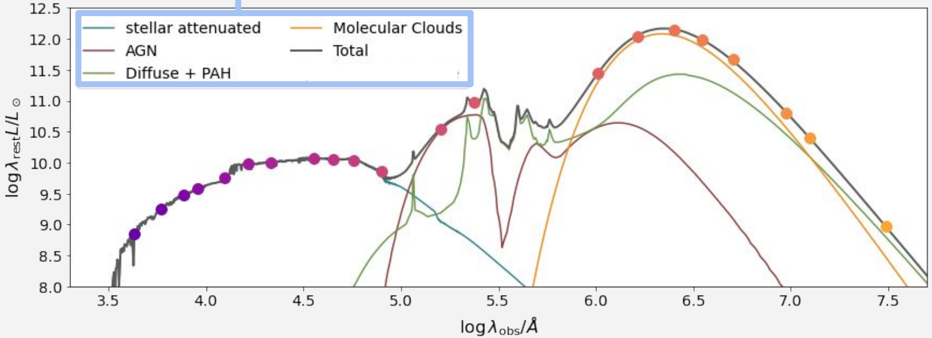
… and this is my last slide, thanx!

SED fitting 101

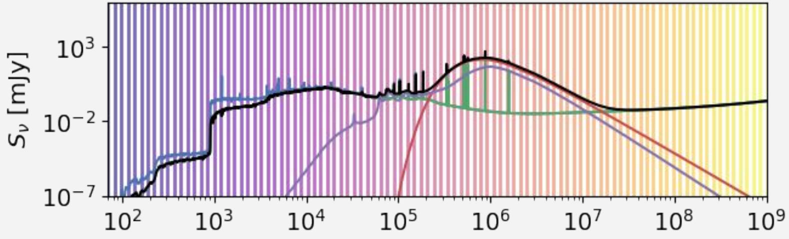
Spectrum Observation

Physical Emission Model

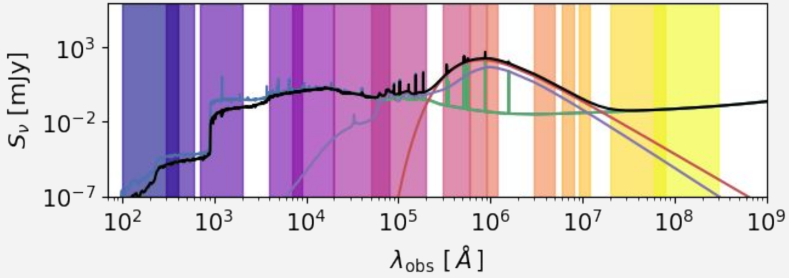
← **parameterized**



→ **Ideally:**



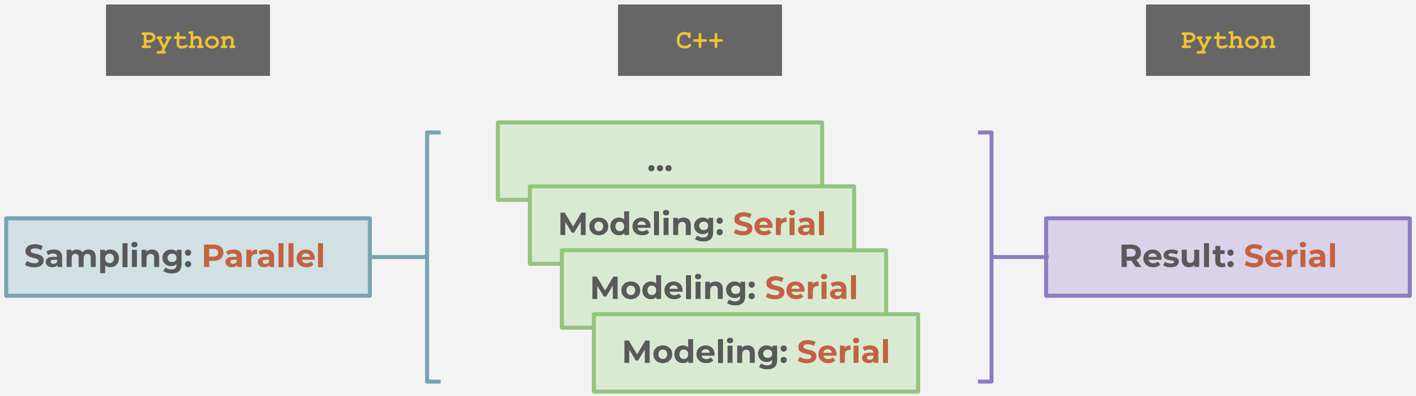
→ **Realistically:**



+ observation
+ parameterized model
+ sampling/fitting strategy

→ **optimal model**
representing your data
(hopefully)

Parallelisation scheme and why



Shared Memory parallel block

- **User-oriented:** I do want users to be able to tailor parallelization on their system easily, so this has to happen on the higher level interface
- **Future devs in the sampling strategy:** Especially considering a future development with **hierarchical bayesian sampling** I will need to parallelize sampling on multiple sources