Offline Organization: Simulations & Analysis

Michael A Kagan, Dylan J Temples

22 February 2023

MAGIS Science & Simulation Meeting

What is the plan for official MAGIS-100 publications?

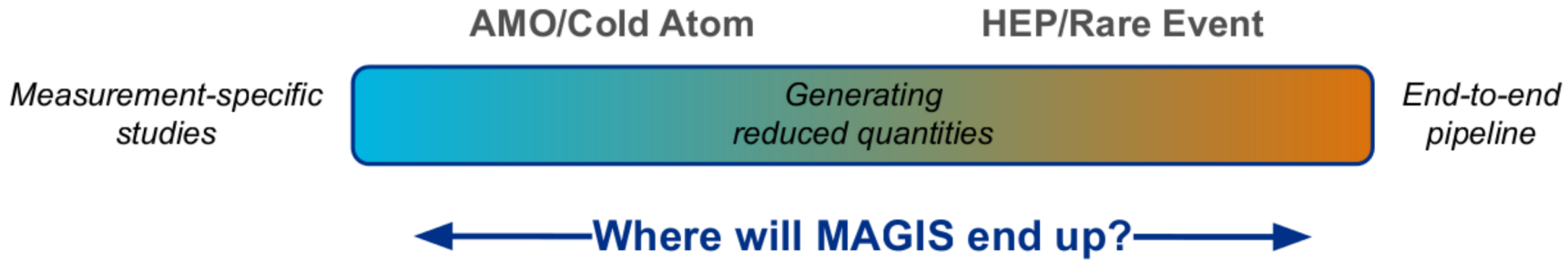
What will it take to convince ourselves?

What will it take to convince the AMO community?

What will it take to convince the DM or GW communities?

Simulation philosophy

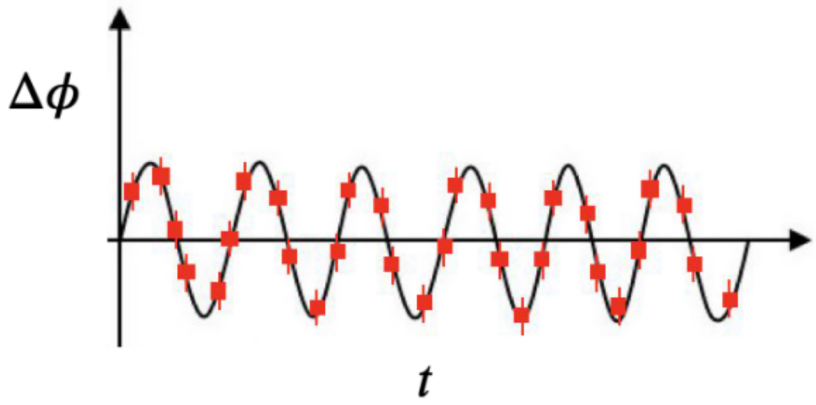
- Mechanistic model of the data generation process
- Can incorporate detailed experimental and theoretical effects
- Enables sampling plausible outcomes from likelihood
  - Generate a campaign's worth of data under a signal hypothesis (happy hunting!)
- Inference through comparing simulation and real data



Why bother?

Ultimately, our goal will be to:

- Determine if there is a signal in the data
- Extract physics parameters of that signal
- Or, constrain parameters in its absence



These are statistical inference questions

- For inference, we need a model to compare
- Comparison via test statistic, e.g. likelihood ratio

$$r(\theta) = \frac{p(\Delta\phi|\theta_1)}{p(\Delta\phi|\theta_0)}$$

  - θ_i: physics parameters
  - $p(x|\theta)$: likelihood of data

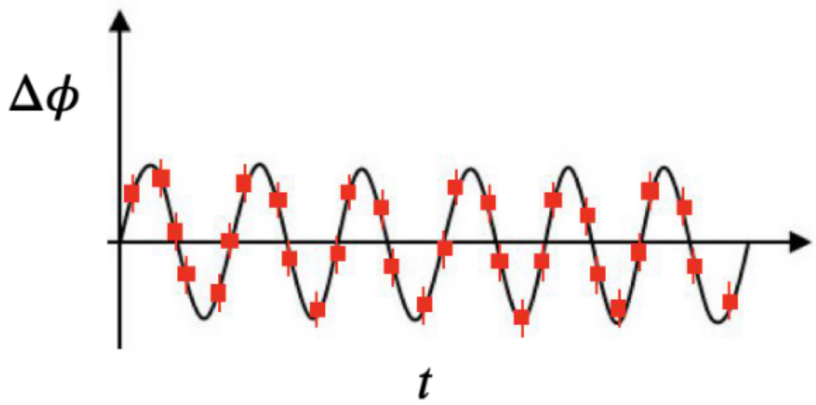
Community belief in a discovery claim in the event of signal (or a limit, in absence) requires robust statistical analysis

Why bother?

If we want to know what signals really look like in our experiment…

And have a principled way to do statistical inference…

We may need large scale simulation.



End-to-end simulation allows us to:

- Characterize and study our apparatus
- Vet our analysis tools in advance of real data, for quick “blessed” results
- Mitigate bias via data salting

What is needed?

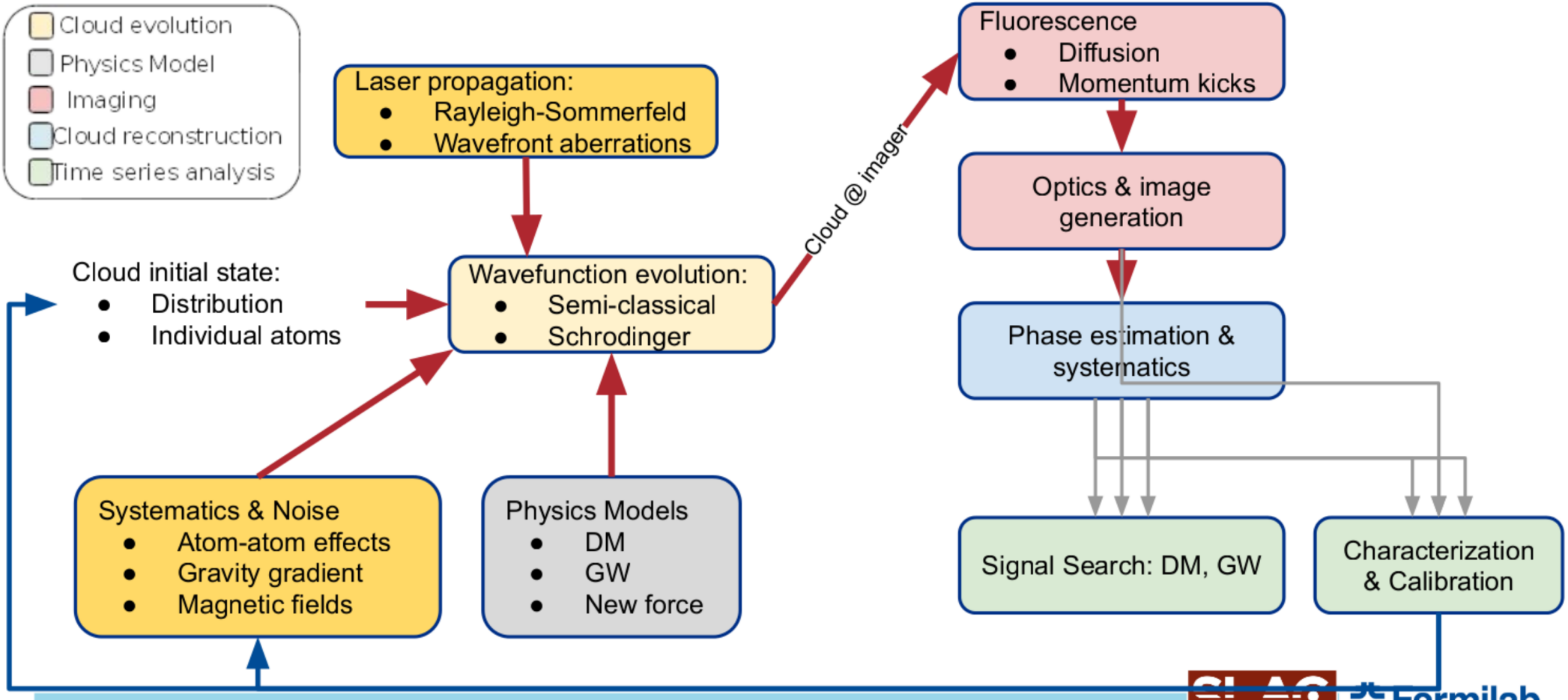
To identify the critical components of the theory / experiment that are needed to generate plausible data

A framework to guide the simulation, connect different simulation pieces, interface with databases, and potentially enable parallel processing

How plausible should the simulated data be?

- May want a high-fidelity simulation with all the bells and whistles… even if slow
- Fast simulation… what we use most of the time, but lower fidelity

Pipeline from 10,000 feet: Simulation & Analysis Chain



What is needed on Day 1 of running?

Ability to simulate sources of background and noise (e.g. magnetic effects) and quantify performance of the apparatus.

- Integrate real measurements back (magnetic field & laser models) into sim
- Fast simulation chain with a working steering framework

Access to databases.

An informed approach to data taking.

What else?

Where are we now?

Met with Jason and Tim to understand simulation from precision AMO perspective

Compiling a list of groups/individuals working on simulation tools/components

- We've asked PIs to encourage students & post-docs to reach out to us
- Following up with developers soon with technical questions
- Thank you to those that have responded already

Looking for involvement in defining the pipeline (beyond the individual components)

We've reached out to other experiments to learn about their sim/ana frameworks

- From roughly MAGIS scale up to few hundred people
- Various fields: particle DM, telescopes, spectral searches
- Tentative discussion at Mar 8 Science & Simulation meeting

Where do we hope to be soon?

- Current tools and developers identified and interfaces/assumptions understood
- Sketch out components for analytic (fast) simulation approach vs full numeric (atom-by-atom?) approach
- Collaboration consensus on simulation philosophy and path forward
- Work with offline infrastructure coordinators to understand computing resources at Fermilab

Starting small: integrate a couple core components

- Balance computational efficiency vs implementation complexity
- Distribution-level, semi-classical tools

Open Questions to the Collaboration

What is the plan for official MAGIS-100 publications?

Will we have official tools reviewed by the collaboration to “bless” results?

> **Michael & Dylan: collaboration supported and developed simulation tools and framework will strengthen our ability to do analysis and publish results**

Do we need to coalesce into a single language for simulation?

How do we want to approach collaborative code development?

- Should we define standards for code maintainability and comp. efficiency?

What does the online data model look like?

- Want to emulate structure in the simulation event data model

SLAC Stanford University U.S. DEPARTMENT OF ENERGY

Fermilab U.S. DEPARTMENT OF ENERGY Office of Science

Thank You!

MAGIS-100

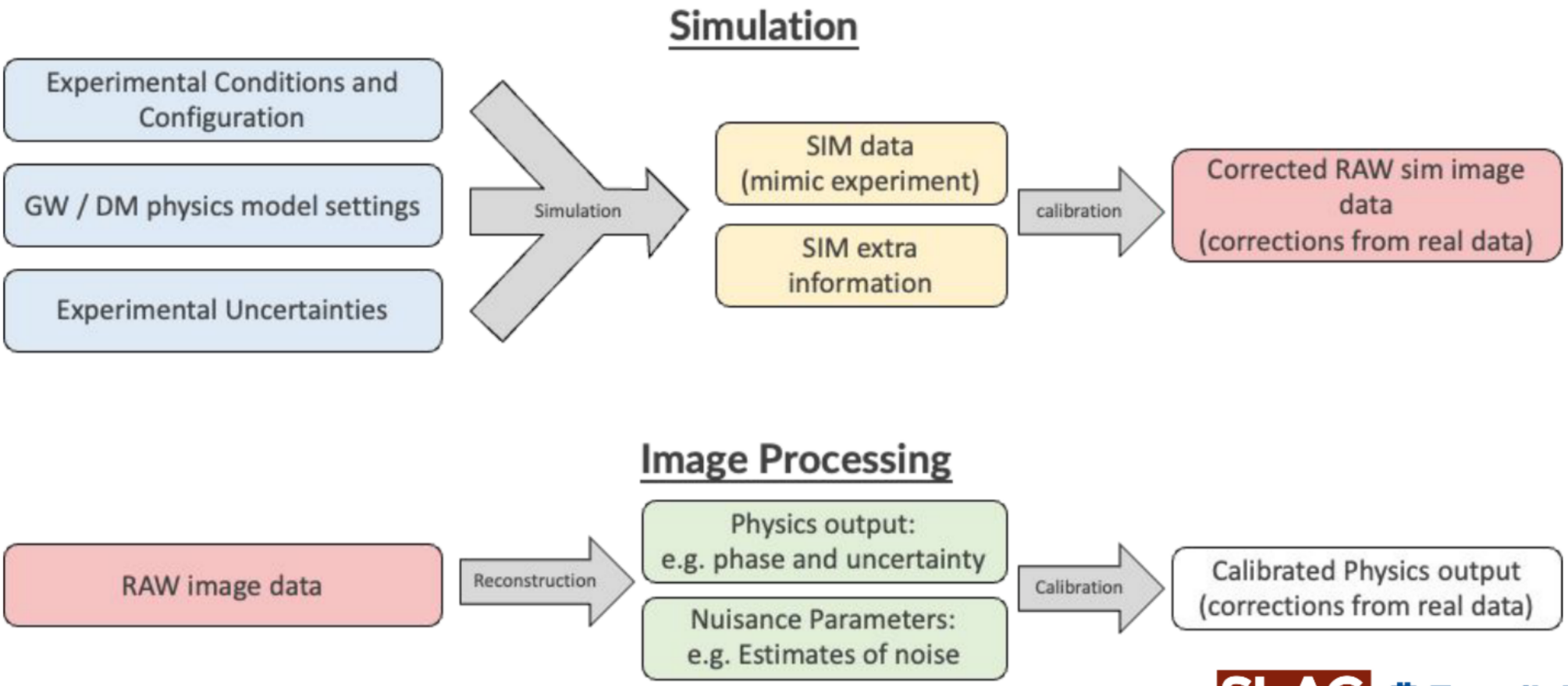
U.S. DEPARTMENT OF ENERGY Office of Science

Gordon and Betty MOORE FOUNDATION

UKRI Science and Technology Facilities Council

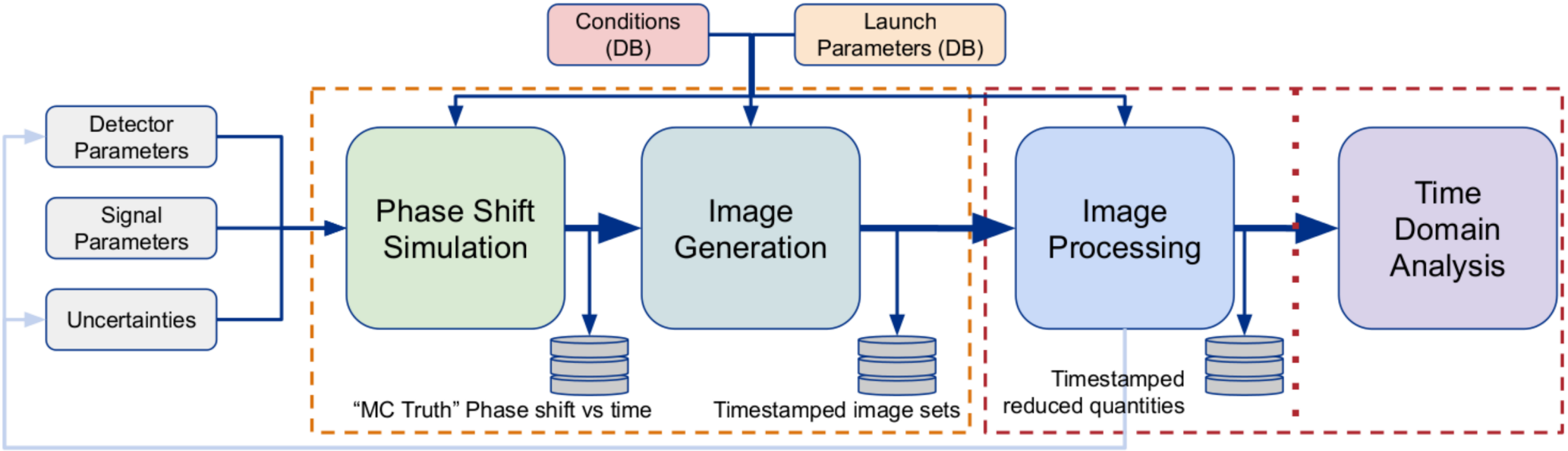
THE KAVLI FOUNDATION

Broad Look


Simulation
Experimental Conditions and Configuration
GW / DM physics model settings
Experimental Uncertainties
Simulation
SIM data (mimic experiment)
SIM extra information
calibration
Corrected RAW sim image data (corrections from real data)
Image Processing
RAW image data
Reconstruction
Physics output: e.g. phase and uncertainty
Nuisance Parameters: e.g. Estimates of noise
Calibration
Calibrated Physics output (corrections from real data)

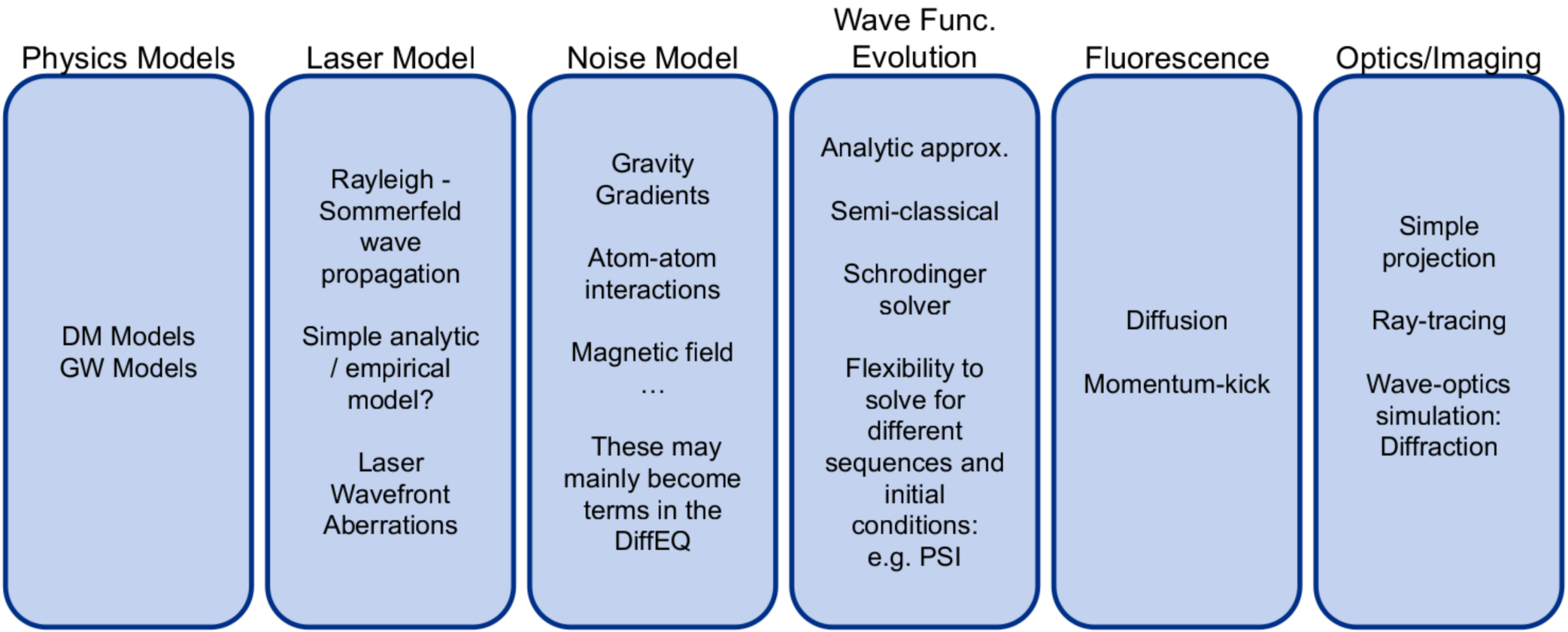
Production Simulations, Image Processing, and Analysis

Production simulation: pick some set of signals with chosen parameters, simulate the phase shift, detector effects, up to creating images (in same format as expected for the experiment)



Production analysis: starting from the images, fit the patterns and extract a phase shift ("image processing"). Should be able to run on both simulation and real data with minor user input. These algorithms should be used in both Online and Offline processing

Pipeline from 10,000 feet: Simulation Components



Assumptions

- Existing tools are flexible enough to be merged into a pipeline
- Plan to use Fermilab computing resources to produce simulated data campaigns
- Simulation data format will match data coming out of instruments
- Simulation tools will interface (read & write) with DBs in much the same way as online systems