1

Hochschule für Technik
und Wirtschaft Berlin

University of Applied Sciences

Monte Carlo Pathfinding in Radio Astronomy

Hermann Heßling

GLOWSKA

3. 6. 2016

SKA Data Challenges

- Planning the German National SKA Data Centre
  - Cost of computing
    - Strong impact on construction costs
    - Determined mostly by processing power and growth of archive

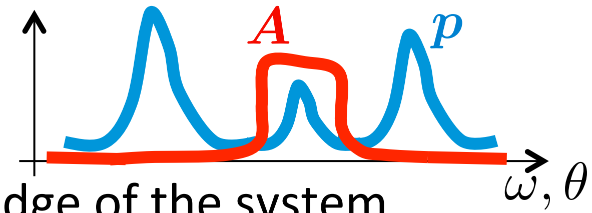
$$\tfrac{2}{3}\, C_{tot} \sim \alpha\, C_{CPU} + (1-\alpha)\, C_{arch}$$

  - Impact
    - Only a small fraction of all "raw data" taken by the antennas can be stored in long-term archives
- Reducing the data volume
  - Developing algorithms for weeding out data of interest in near-realtime
  - Each key science project (EoR, pulsar, …) demands its own solution
- Data analytics
  - Simply copying software from LOFAR not effective
  - SKA much more complex than LHC
    - Growth of archive / "raw data rate": **SKA $\gtrsim$ 100 LHC**

Monte Carlo Simulation

- Outcome of measurements is compared to expectation values of observables A

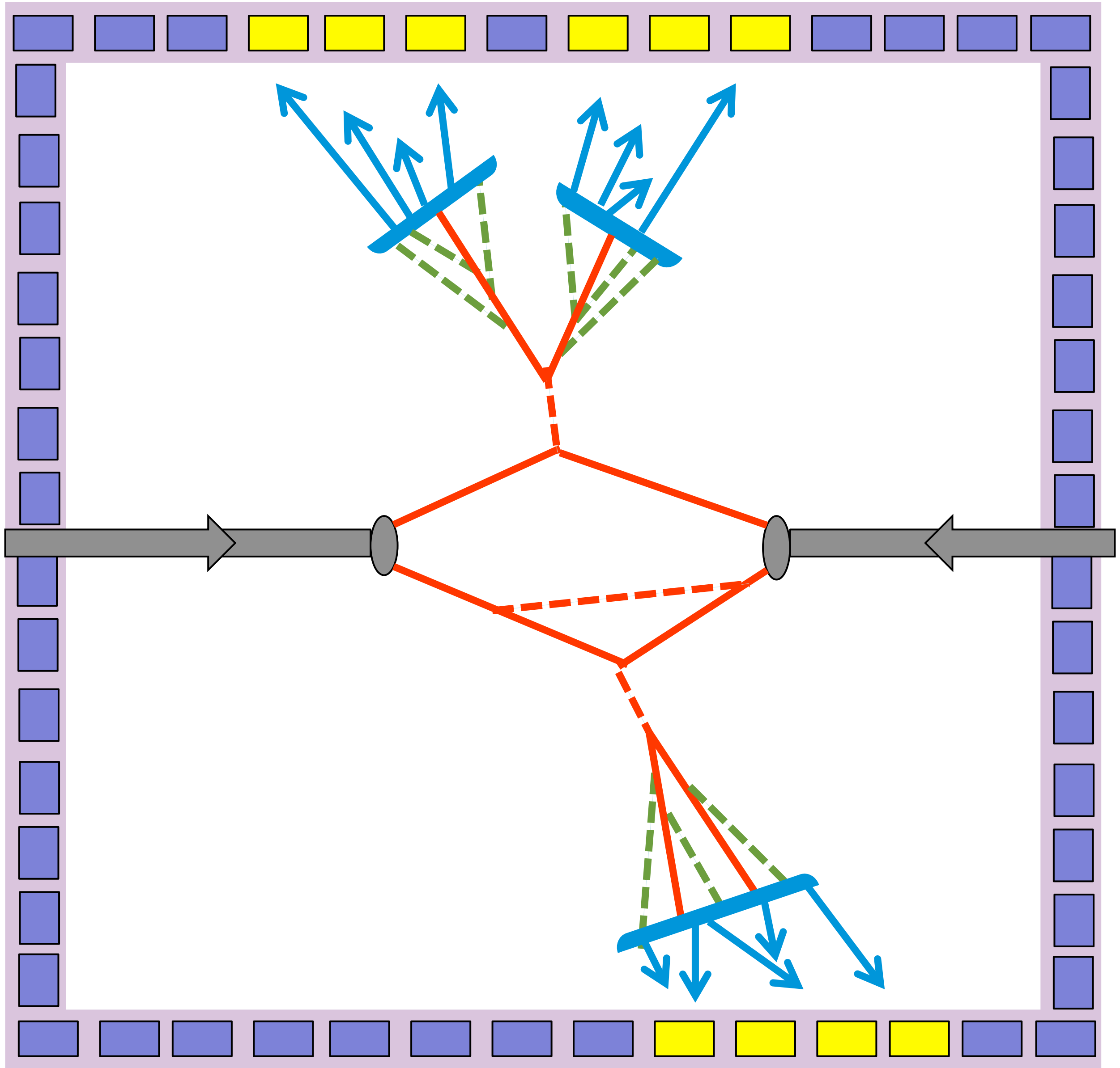
$$\langle A \rangle = \int A(\omega, \theta)\, p(\omega, \theta)\; \mathrm{d}\omega \mathrm{d}\theta$$



- The **probability distribution** p characterizes the knowledge of the system
  - Thermal equilibrium: $p \sim e^{-H/kT}$
  - Fitting models to data: $p(\omega, \theta) = p(\omega|\theta)\, p(\theta)$
    - $p(\omega|\theta)$ is the likelihood function of some model distribution
    - The prior $p(\theta)$ describes the initial knowledge of the parameter set θ

- Importance of Monte Carlo methods for evaluating expectation values
  - Scaling of error $\sim 1/\sqrt{N}$, N= number of **random points**: $(\omega_1, \theta_1), \ldots, (\omega_N, \theta_N)$
    - Independent of dimension of integral
  - Parallel computing
    - Many Monte Carlo algorithms are parallelizable
  - Finding “rare events” in large search spaces
    - Importance sampling

Monte Carlo Event Generators at LHC



"MCTruth" data

- Hard interaction
- Parton shower
- Hadronization

⇒ Probabilistic simu-lation of the final state

Simulated data

- Detector simulation

⇒ Comparison with experimental data

Monte Carlo Event Generators at LHC

- Theory
  - Predicting final state
  - Simulating background / noise

- Experiment
  - Optimizing trigger algorithms

    ⇒ reducing data volume
  - Determining influence of detector (e.g. dead material)

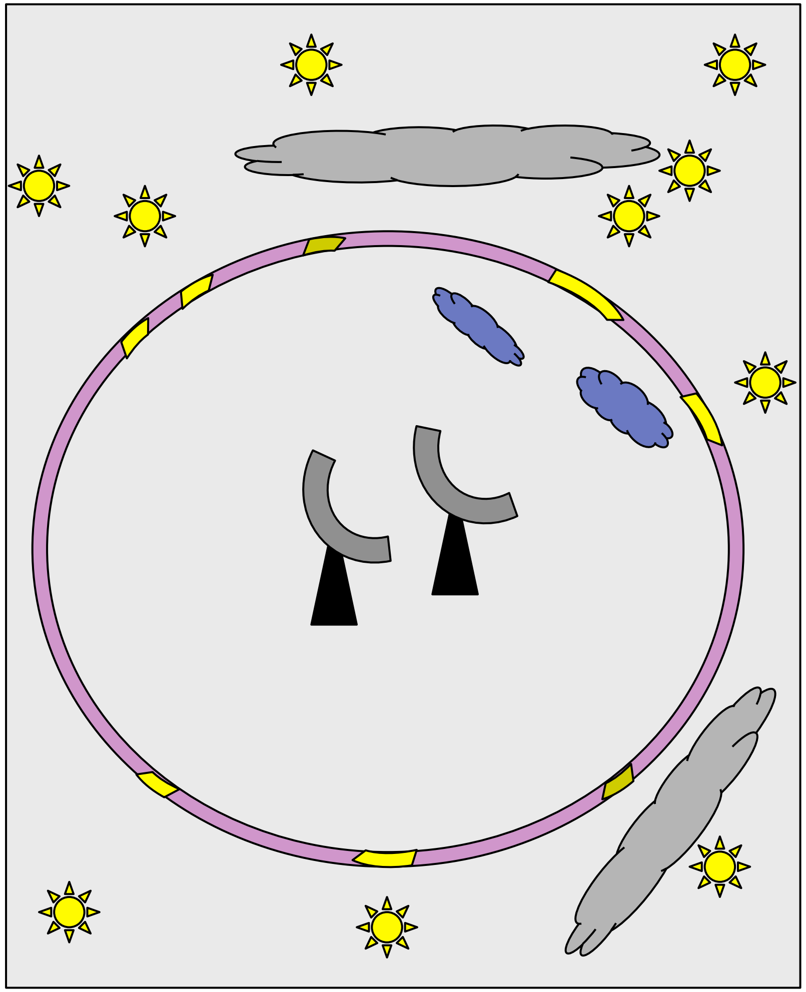
    ⇒ estimating acceptance corrections
  - Unfolding measured data to “hadron level”

    ⇒ publishing “true” results

Monte Carlo Simulation at SKA

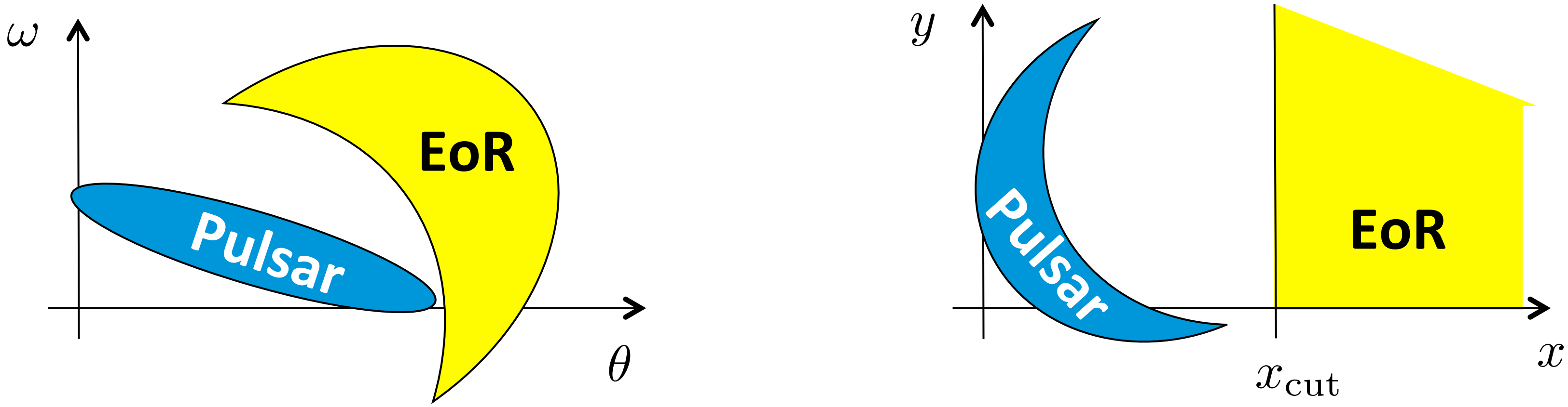
- LHC
  - Single source
    - Inside the detector
  - Theory: QCD + QED + EW
    - Initial state known
  - MC simulation of collision events
    - Several MC event generators

- SKA
  - Huge number of sources
    - Outside the celestial sphere (detectors)
  - Theory: Maxwell equation with sources
    - + GRT
    - Boundary value problem
  - Goal: MC simulation of the cosmos



Monte Carlo Simulation at SKA

- Probability distribution $p(\omega, \theta)$ on the phase space
  - Specify "subregions of interest" for every key science project



  - Determine deformation map $(\theta, \omega) \to (x, y)$ for applying "trigger cuts" on data
    - LHC: e.g. cut on transverse energy of calorimeter cells, $\Sigma_i \; E_i \sin \theta_i \geq E_T$
    - SKA: cuts on functions of visibility data, e.g. $f_{\text{EoR}}(u, v, w, V(u, v, w)) \geq x_{\text{cut}}$
- Monte Carlo Simulation
  - New physics by simulating the background in dedicated regions of phase space
  - Reducing data volume in near-realtime based on simulation of "trigger cuts"
    - Key science projects obtain a fixed share of long-term data archive