Reversible Jump MCMC for Pile-up Decomposition

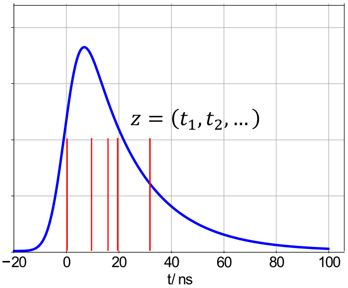
Yiyang Wu, Benda Xu

CIDeR-ML meeting

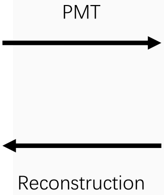
April 11, 2025

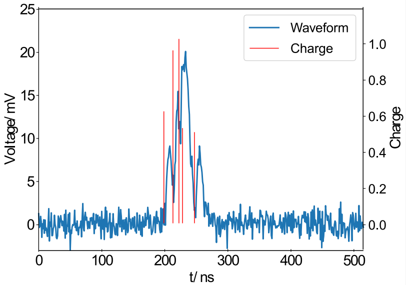
Motivation: pileup is common in observations

PE pulses pile up in waveform (1-D time pileup)



Sample from Poisson process



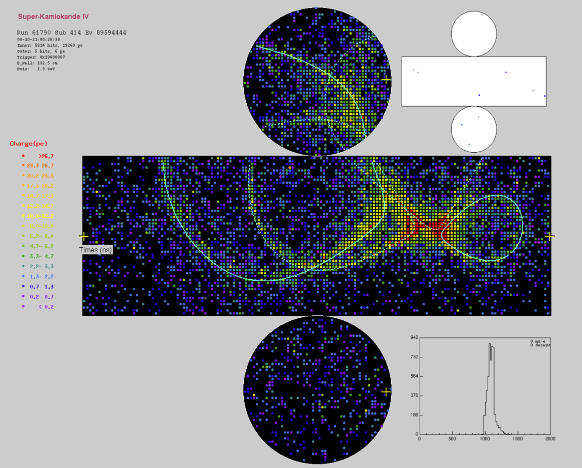


Convolution with single PE waveform

Number of PE (hit time t_i) is a **variable**, $\{N; t_1, \dots, t_N\} \in \mathbb{L}_N = \{N\} \otimes \mathbb{R}^N$

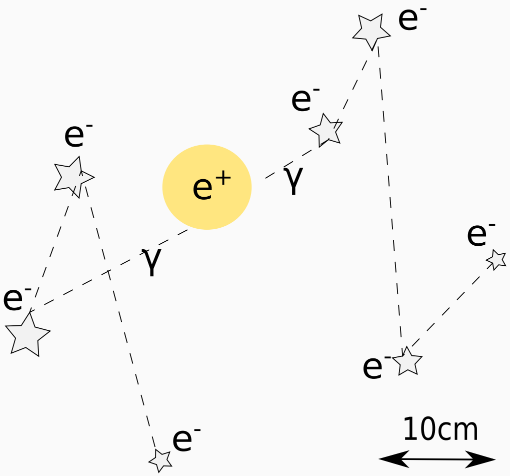
Concerning charge of each PE: $\{N; t_i, q_i\} \in \mathbb{L}_N = \{N\} \otimes \mathbb{R}^{2N}$

Multiple charged particles in a GeV neutrino event o proton decay



Number of rings is a **variable.** $\{N; t_i, \vec{r}_i, \vec{p}_i, \mathrm{PID}_i\} \in \mathbb{L}_N = \{N\} \otimes \mathbb{R}^{8N}$

MeV physics: β^-/$2\beta^-$/4β^-/γ/β^+ discrimination



- β^-: single point source. E.g. solar neutrino CC capture induced single β decay (^{130}Te + $\nu_e \longrightarrow e^-$ + ^{130}I*, ^{130}I $\longrightarrow \beta$ + $\bar{\nu}_e$ + ^{130}Xe); ^{8}B solar neutrino ES. Background of $0\nu\beta\beta$
- $2\beta^-$: $2\nu\beta\beta$, Majorana neutrino $\rightarrow$ $0\nu\beta\beta$
- $4\beta^-$: Dirac neutrino $\rightarrow 0\nu4\beta$
- γ: radioactive calibration source or background (external ^{40}K, ^{208}Tl; n capture γ); NCQE
- β^+: $\bar{\nu}_e$ IBD; β decay with γ

Motivation: pileup is common in observations

- Dark noise PEs mixing with event PEs (sub MeV in LS; MeV in water)
- Cascade event: $\pi/\mu \to e$, ^{212}Bi-^{212}Po, ^{85}Kr, Kaon ($p \to \nu + K$, $K \to \mu/\pi \to e$) in a DAQ window (1-D time pileup)
  $\{x, y, z; N; t_i, E_i\} \in \mathbb{R}^3 \otimes \mathbb{L}_N = \mathbb{R}^3 \otimes \{N\} \otimes \mathbb{R}^{2N}$
- Accidental ^{14}C pileup with other event in organic LS
- Multiple segments in a track
- Multiple tracks
- Bundle muons vs single muon

Mixture model and RJMCMC

Mixture model, and the frequentist way

Master likelihood $\mathcal{L}(\theta; x) = p(x|\theta),\ \theta \in \{N\} \otimes \mathbb{R}^{mN}$, x is observed data (e.g. waveform ADC; PMT charge)

Mixture model, and the frequentist way

Master likelihood $\mathcal{L}(\theta; x) = p(x|\theta),\ \theta \in \{N\} \otimes \mathbb{R}^{mN}$, x is observed data (e.g. waveform ADC; PMT charge)

When N is fixed, problem simplifies to a mixture model:

Mixture model, and the frequentist way

Master likelihood $\mathcal{L}(\theta; x) = p(x|\theta),\ \theta \in \{N\} \otimes \mathbb{R}^{mN}$, x is observed data (e.g. waveform ADC; PMT charge)

When N is fixed, problem simplifies to a mixture model:

$\{N\} \otimes \mathbb{R}^{mN} \Rightarrow \mathbb{R}^{mN}$

MLE is performed in $\mathbb{R}^{mN}$ subspace. Techniques like Expectation-Maximization can be useful to further simplify the MLE process.

Mixture model, and the frequentist way

Master likelihood $\mathcal{L}(\theta; x) = p(x|\theta),\ \theta \in \{N\} \otimes \mathbb{R}^{mN}$, x is observed data (e.g. waveform ADC; PMT charge)

When N is fixed, problem simplifies to a mixture model:

$\{N\} \otimes \mathbb{R}^{mN} \Rightarrow \mathbb{R}^{mN}$

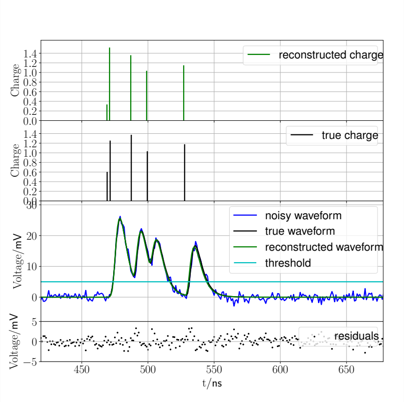
MLE is performed in $\mathbb{R}^{mN}$ subspace. Techniques like Expectation-Maximization can be useful to further simplify the MLE process.

Then we compare mixture models with different N, e.g. likelihood ratio with penalties to constraint number of free parameters.

E.g. SK `fiTQun` adds a penalty term to $\log \mathcal{L} -= a_0 + a_1 \cdot E_{\mathrm{new}},\ a_0 > 0,\ a_1 \geqslant 0$ to discourage adding new ring

Can we use statistical model to replace penalty term?

$$p(w_j|\{t_i, q_i\}, N) = \prod_{j=1}^{W} \frac{1}{\sqrt{2\pi}\sigma_w} \exp\left\{-\frac{[w_j - \sum_{i=1}^{N} q_i V_{\mathrm{PE}}(t_j - t_i)]^2}{2\sigma_w^2}\right\}$$

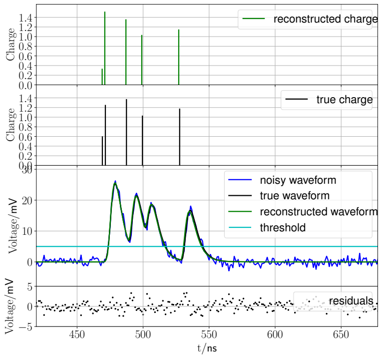


$j \in 1, \cdots, W$ time grids, w_j waveform, σ_w Gaussian white noise

$V_{\mathrm{PE}} \otimes \{q_i@t_i\} = \sum_{i=1}^{N} q_i V_{\mathrm{PE}}(t_j - t_i)$ expected wave

Can we use statistical model to replace penalty term?

$$p(w_j|\{t_i, q_i\}, N) = \prod_{j=1}^{W} \frac{1}{\sqrt{2\pi}\sigma_w} \exp\left\{-\frac{[w_j - \sum_{i=1}^{N} q_i V_{\mathrm{PE}}(t_j - t_i)]^2}{2\sigma_w^2}\right\}$$



$j \in 1, \cdots, W$ time grids, w_j waveform, σ_w Gaussian white noise

$V_{\mathrm{PE}} \otimes \{q_i@t_i\} = \sum_{i=1}^{N} q_i V_{\mathrm{PE}}(t_j - t_i)$ expected wave

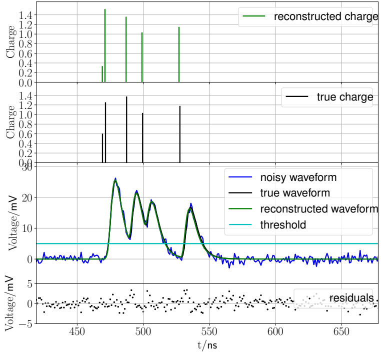
Prevent overfit w with many small q_i? Constraint q_i with charge model!

$p(\{q_i\}, w_j|\{t_i\}, N) \;=\; p(w_j|\{t_i, q_i\}, N) \prod_{i=1}^{N} p(q_i)$,

$\{q_i\}$ becomes part of model (frequentist view) or a prior (Bayesian view)

Can we use statistical model to replace penalty term?

$$p(w_j|\{t_i, q_i\}, N) = \prod_{j=1}^{W} \frac{1}{\sqrt{2\pi}\sigma_w} \exp\left\{-\frac{[w_j - \sum_{i=1}^{N} q_i V_\mathrm{PE}(t_j - t_i)]^2}{2\sigma_w^2}\right\}$$



$j \in 1, \cdots, W$ time grids, w_j waveform, σ_w Gaussian white noise

$V_\mathrm{PE} \otimes \{q_i@t_i\} = \sum_{i=1}^{N} q_i V_\mathrm{PE}(t_j - t_i)$ expected wave

Prevent overfit w with many small q_i? Constraint q_i with charge model!

$p(\{q_i\}, w_j|\{t_i\}, N) = p(w_j|\{t_i, q_i\}, N) \prod_{i=1}^{N} p(q_i)$, $\{q_i\}$ becomes part of model (frequentist view) or a prior (Bayesian view)

Modeling N is also useful: $p(N, \{q_i\}, w_j|\{t_i\}) = p(\{q_i\}, w_j|\{t_i\}, N)p(N)$, $N \sim \lambda(\mu)$

The dimension curse for likelihood ratio

Cannot direct compare probabilities across dimensions! $>$, $<$ and ratio is not well-defined when comparing likelihood in $\{N\} \otimes \mathbb{R}^{mN}$ across N!

$\prod_{i=1}^{N} p(q_i)\mathrm{d}q^N$ and $\prod_{i=1}^{N+1} p(q_i)\mathrm{d}q^{N+1}$ lives in different dimensions.

The dimension curse for likelihood ratio

Cannot direct compare probabilities across dimensions! $>$, $<$ and ratio is not well-defined when comparing likelihood in $\{N\} \otimes \mathbb{R}^{mN}$ across N!

$\prod_{i=1}^{N} p(q_i)\mathrm{d}q^N$ and $\prod_{i=1}^{N+1} p(q_i)\mathrm{d}q^{N+1}$ lives in different dimensions.

One last struggle: analytical integration

$p(w_j|\{t_i\}) = \int_{\mathbb{R}^N} \mathrm{d}q^N p(w_j|\{t_i, q_i\}, N) \prod_{i=1}^{N} p(q_i)$ but it only has a closed form if $p(q)$ is Gaussian. Let alone $p(w_j|N) = \int_{\mathbb{R}^N} \mathrm{d}q^N \int_{\mathbb{R}^N} \mathrm{d}t^N p(w_j|\{t_i, q_i\}, N) \prod_{i=1}^{N} p(q_i)p(t_i)$, but it only has a closed form if $p(q)$ is Gaussian.

The dimension curse for likelihood ratio

Cannot direct compare probabilities across dimensions! $>$, $<$ and ratio is not well-defined when comparing likelihood in $\{N\} \otimes \mathbb{R}^{mN}$ across N!

$\prod_{i=1}^{N} p(q_i)\mathrm{d}q^N$ and $\prod_{i=1}^{N+1} p(q_i)\mathrm{d}q^{N+1}$ lives in different dimensions.

One last struggle: analytical integration

$p(w_j|\{t_i\}) = \int_{\mathbb{R}^N} \mathrm{d}q^N p(w_j|\{t_i, q_i\}, N) \prod_{i=1}^{N} p(q_i)$ but it only has a closed form if $p(q)$ is Gaussian. Let alone $p(w_j|N) = \int_{\mathbb{R}^N} \mathrm{d}q^N \int_{\mathbb{R}^N} \mathrm{d}t^N p(w_j|\{t_i, q_i\}, N) \prod_{i=1}^{N} p(q_i)p(t_i)$, but it only has a closed form if $p(q)$ is Gaussian.

However, MC sampling $\{N\} \otimes \mathbb{R}^N$ is possible: naively sample N and then t_i.

Sample posterior with RJMCMC

Posterior $p(N, \{t_i\}|w_j) \propto p(w_j|N, \{t_i\})\pi(N)\pi(t_i)^N$, easy to calculate but difficult to direct sample N and then t_i, can't even calculate $\int p(N|w) = \int_{\mathbb{R}^{\mathbb{N}}} \mathrm{d}t^N p(N, t|w)$

RJMCMC is used to sample in $\{N\} \otimes \mathbb{R}^{mN}$ without knowing marginal distribution $p(N)$ in advance.

$$\min\left\{1, \frac{L(\theta')\pi(\theta')}{L(\theta)\pi(\theta)}\frac{p(\theta' \to \theta)}{p(\theta \to \theta')}\right\} \xrightarrow[\text{RJMCMC}]{\text{From MH to}} \frac{L(\theta')\pi(\theta')}{L(\theta)\pi(\theta)}\frac{g(v)p(\text{choose } \theta' \to \theta)}{g(u)p(\text{choose } \theta \to \theta')}\left|\det\frac{\partial(\theta', v)}{\partial(\theta, u)}\right|$$

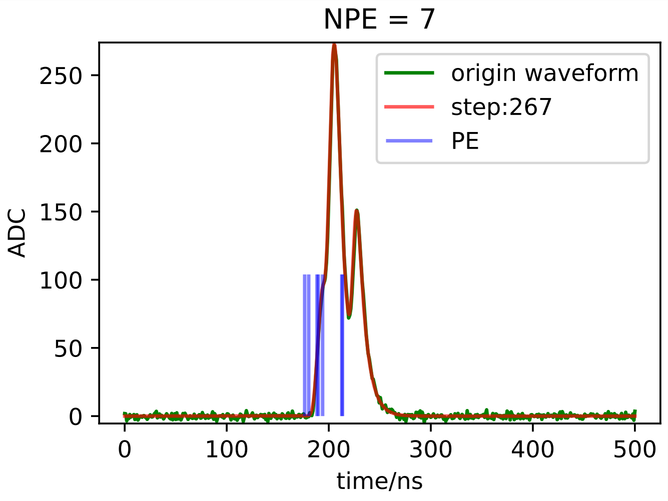
- θ: parameter to be fit; current step θ, next step θ'
- $p(a \to b)$: proposal probability to jump from parameter a to parameter b
- u, v: proposed auxiliary parameters satisfying $\dim u + \dim\theta = \dim v + \dim\theta'$
- $g(u)$ proposal probability to raise u (continuous part)
- $p(\text{choose } \theta' \to \theta)$: discrete part
- implemented jump $f : (\theta, u) \to (\theta', v)$, Jacobian term $|\det(\nabla f)|$

FSMP: RJMCMC on PE reconstruction from waveform

First glimpse on Fast Stochastic Matching Pursuit (FSMP)

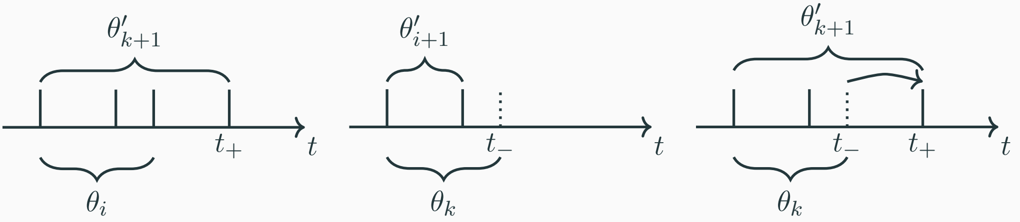
arXiv:2112.06913

arXiv:2403.03156



FSMP: a RJMCMC jumping N and t_i to match waveform

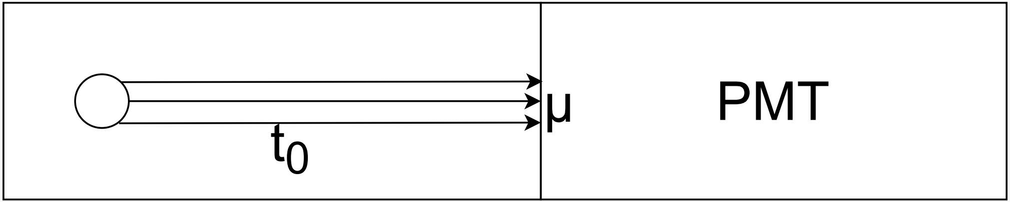
RJMCMC master formula



- $g(u)$: u can be drawn from smeared deconvolution result $h(t)$
- +1 PE $p(\text{choose } +1) = p_+ = 1/4$
- -1 PE at **the** t_-: $p(\text{choose } -t_-) = p_- \frac{1}{N} = 1/(4N)$
- +1 PE $\theta_{i+1} = (\theta_k, t_+) = f(\theta_k, u) = (\theta_k, u)$, f is i.d. transform so $|\det \nabla f| = 1$
- Move t_i from t_- to t_+: $p_{\text{move}} p(\text{choose } t_i) = 1/2 \times 1/N$
- $\alpha_+ = \frac{p(N_k, \{t_i\}_{k+1}|w_j)\pi(N_{k+1})\pi(\{t_i\}_{k+1})}{p(N_k, \{t_i\}_k|w_j)\pi(N_k)\pi(\{t_i\}_k)} \frac{1 \cdot p(\text{choose } -t_-)}{g(u)p(\text{choose } +1)} |\det \nabla f| = \cdots \times \frac{1}{(N+1)h(t_+)}$
- $\alpha_- = \cdots \times Nh(t_-)$

Toward particle reconstruction

A toy detector with one PMT: λ is energy, fly time t_0 is vertex.



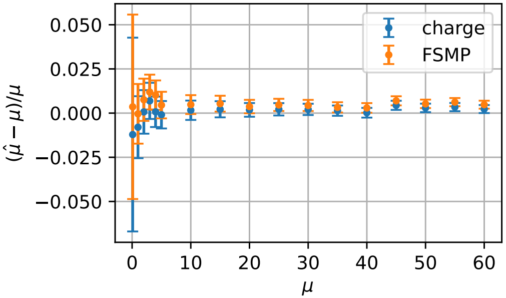
Reconstruct t_0 upon FSMP output vs first hit method

Reconstruct λ upon FSMP output vs $\frac{\text{charge}}{\text{gain}}$

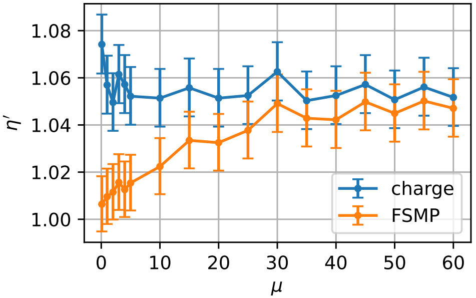
FSMP performance on energy

$$\eta_{\text{theory}} = \frac{1}{\sqrt{\mu}}$$ the theoretical limit of resolution

$$\eta = \frac{\sqrt{\mathrm{Var}[\hat{\mu}]}}{\mathrm{E}[\hat{\mu}]}$$, relative resolution $\eta' = \frac{\eta}{\eta_{\text{theory}}}$,



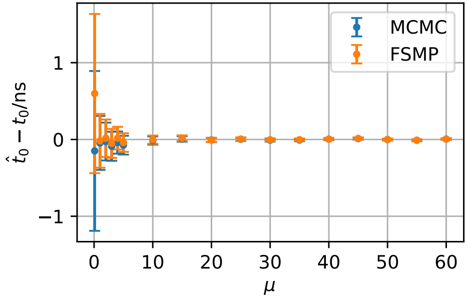
μ relative bias



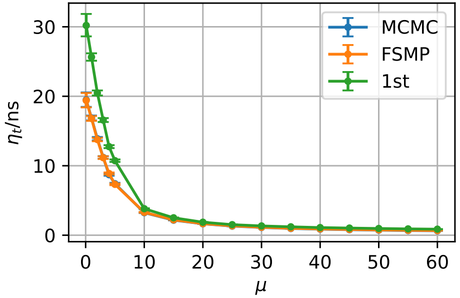
μ relative resolution

FSMP performance on vertex

MCMC: reconstruct t_0 with known PE truth, gives theoretical limit of t_0 resolution.



t_0 absolute bias

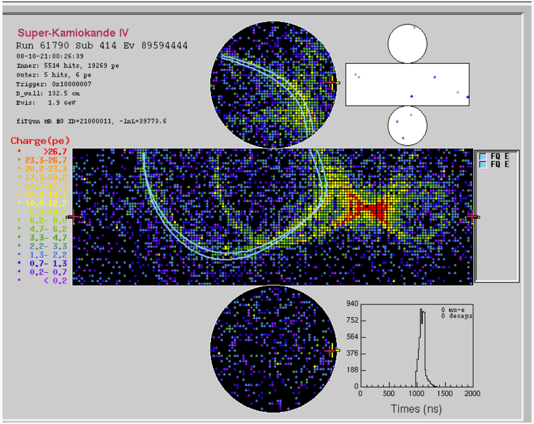


t_0 absolute resolution

RJMCMC-fiTQun

Development history and status until last month

Wiki page: `https://www-sk1.icrr.u-tokyo.ac.jp/sk-local/wiki/index.php/RJMCMC-fiTQun`

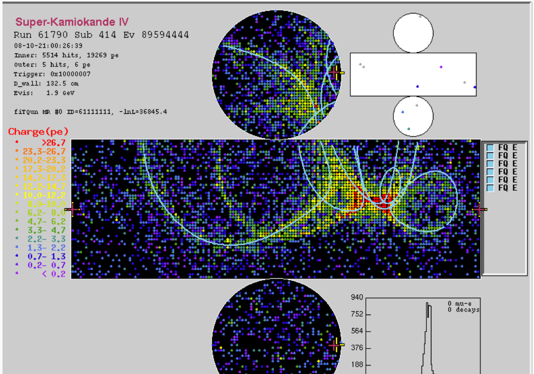


The RJMCMC is very hard to propose new rings with only split-merge steps

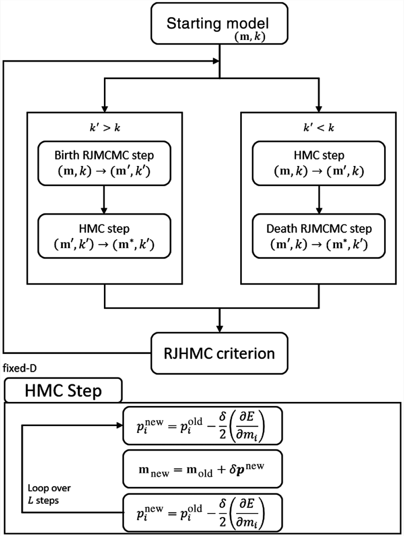
Progress in the last week

I improved the proposal algorithm $h(\vec{r})$ and correctly implemented the birth-death jump

Now RJMCMC can easily jump to multiple rings. But saving more than 6 rings to zbs somehow broke output format and event display fails, so only 6 rings at max.



A dream: RJHMC



Gradient information can be used to guide the MCMC chain: Hamiltonian MC[a].

Extend to cross-dimensional: RJHMC[b]

[a] Duane et al., “Hybrid Monte Carlo”.

[b] Sen and Biswas, “Transdimensional seismic inversion using the reversible jump Hamiltonian Monte Carlo algorithm”.

Detailed plans for workshop

Plan A: traditional RJMCMC written in C++

1. 4.11 Run `skdetsim`, shooting N 0.1-1 GeV electron/muon; check chain fidelity
2. 4.12 Calculate confusion matrix of `RJMCMC-fiTQun` to see the ring counting capabilities
3. 4.13 `fiTQun`1R result as initial state $\rightarrow$ MR result as initial state
4. 4.14-15 Tune prior and estimate chain convergence using Gelman-Rubin method

Plan B: (RJ)HMC written in python

1. 4.14 Prepare for the math formulas of RJHMC
2. 4.15-16 Implement a differentiable `fiTQun`likelihood with opticSIREN
3. 4.16-18 Implement the 1 ring HMC calling differentiable likelihood
4. 4.17-18 Run 1 ring HMC on forward simulation output
5. Future: multi-ring superposition forward simulation, RJHMC