Combined sim+sigproc?

Brett Viren

August 17, 2021

Topic

We may use the linearity of Fourier transforms to produce a **combined simulation and signal processing** algorithm which requires the same computation as simulation alone.

1. Explain the technique.
2. Give some reasons why we might want to do this.
3. List caveats and potential problems.

Simulation - drift + projection stages

Drift:

- ionization electrons at a point for each tracking step transported to near the wire planes
- electron absorption, diffusion, total charge fluctuation
- results in a truncated 2D Gaussian distribution for each.

Project:

- Gaussian distributions binned to a 2D **wire pitch** *vs* **drift step** coordinate system associated with each wire plane.
- pitch bins are $\frac{1}{10}^{th}$ pitch wide

Produces S^i, the 2D ionization charge distribution for each relative impact bin i for a given wire/strip plane.

Simulation - convolution stage

The detector response is applied to produce a "pre-measure" as an "interlaced-convolution" plus noise:

$$M_v = \sum_i R^i * S^i + N$$

- $S^i \sim (n_{wire}, n_{step})$ array giving counts of ionization electrons in each *impact bin i* counted relative to the nearest wire.
- $R^i \sim (n_{wire'}, n_{step'})$ array giving induced *response* from one electron falling on impact bin i relative to given wire and over one drift step.
- $N \sim (n_{wire}, n_{step})$ an array of noise[1]
- $M_v \sim (n_{wire}, n_{tick})$ represents a contribution to voltage from given wire during a given sampling period (tick).
- $\sum_i$ is over typically $n = 10$ impact bins per wire.

[1] We do not write it explicitly but N is subject to electronics response which is also forms part of R^i.

Simulation - digitization stage

And the final measure:

$$M = \mathcal{D}(CM_v)$$

- $C \sim (n_{chan}, n_{wire})$ channel map
- $M \sim (n_{chan}, n_{tick})$ represents ADC count on given channel at given sampling period (tick).[2]
- $\mathcal{D}()$ a linear scaling and integer-truncation function

For $C = \mathbb{I}$ (no wrapped wires) we approximate $M_v \to M \approx M_v$ up to the overall scaling.

[2] For convenience we choose *tick* and *step* to represent identical durations.

Marginalize over impact bins

Define an *average response*

$$R \equiv \frac{1}{n} \sum_i R^i$$

and an *average signal* $S \sim (n_{wire}, n_{step})$ such that

$$R * S \equiv \sum_i R^i * S^i$$

to allow rewriting

$$M = R * S + N$$

Estimate S

We may deconvolve M to produce an estimate of S using usual Fourier transform trick[3]:

$$\tilde{S} \approx \mathcal{F}^{-1}\left(\frac{\mathcal{F}(M)}{\mathcal{F}(R)}\right) \equiv \mathcal{F}^{-1}\left(\frac{\mathcal{M}}{\mathcal{R}}\right)$$

$$\tilde{S} \approx \sum_i \mathcal{F}^{-1}\left(\frac{\mathcal{R}^i}{\mathcal{R}}\mathcal{S}^i\right) \equiv \sum_i \mathcal{F}^{-1}\left(\tilde{\mathcal{R}}^i\mathcal{S}^i\right) = \sum_i \tilde{R}^i * S^i$$

Where $\tilde{\mathcal{R}}^i$ is the Fourier dual to a *residual response* $\tilde{R}^i$ for the i^{th} impact position.

- For Wire-Cell, $i \in \{1, ..., 10\}$.
- If there is no impact variation, $i \in \{1\}$, then $\tilde{S} \equiv S$!

[3] Noise is dropped here. We will come back to it.

What is the residual response $\tilde{R}^i$?

It represents the **relative deviation** of an impact response from the local wire-average response.

$$\mathcal{F}(\tilde{R}^i) \equiv \tilde{\mathcal{R}}^i = \frac{\mathcal{R}^i}{\sum_i \mathcal{R}^i}$$

- In Fourier interval space, $\tilde{R}^i$ will be mostly unity with small wiggles at the steps very near to the wires.
- Effectively, it produces a smearing of S^i (in addition to diffusion).
  - Potentially we may be able to replace $\sum_i \tilde{R}^i * S^i \rightarrow G * S$, where G is Gaussian smear.
  - This would give a further $n = 10\times$ speed up.
- Nature of the smearing will itself vary between detectors
  - Larger smearing expected using strips+holes than wires.

Noise

We dropped noise for simplicity. Adding it back now:

$$\tilde{S} \approx \sum_i \mathcal{F}^{-1}\left(\frac{\mathcal{R}^i\mathcal{S}^i + \mathcal{N}}{\mathcal{R}}\right) \equiv \sum_i \mathcal{F}^{-1}\left(\tilde{\mathcal{R}}^i\mathcal{S}^i + \tilde{\mathcal{N}}\right)$$

$$\mathcal{N} \equiv \mathcal{F}(N);\ \tilde{\mathcal{N}} \equiv \frac{\mathcal{N}}{\mathcal{R}}$$

Thus the noise[4] in SP is seen as being "filtered" by the inverse of the average response.

- Can we generate $\mathcal{N}$ directly in Fourier space to save an FFT?

[4] Expanding on a footnote above, $\mathcal{R}^i = \mathcal{R}^i_{FR}\mathcal{R}_{ER}$ and really $\mathcal{N} = \mathcal{R}_{ER}\mathcal{N}'$ where $\mathcal{N}'$ is some "unshaped" noise from wire capacitance and resistance before the shaper circuit. $\mathcal{R}_{ER}$ thus cancels from the noise term but this seems not particularly exploitable.

So what?

$\tilde{S}$ is equivalent to the output of the combination of the nominal simulation plus signal processing yet requires an amount of computation equivalent to simulation alone.

- Will speed up preparation of sim+sigproc data sets.
  - $time(SP) \approx time(sim) \Rightarrow 2\times$ speed up.
- May reap additional $10\times$ speedup of core part if we can replace the n–way interlaced fine-scale convolutions with a single coarse-scale convolution.
- May be a gateway to potentially novel techniques:
  - LS4GAN mapping $\tilde{S} \leftrightarrow S_{SP}^{det}$, $\tilde{S}$ is simpler than S_{SP}^{det} so easier to analyze/evaluate.
    - ★ Backwards LS4GAN map would give sort of 2D `GaussHit` reco.
    - ★ Push more and have LS4GAN map real data to 3D depos?
  - May construct an ensemble of $\tilde{S}$ each from an individual depo and use as a basis for "fitting" S_{SP}^{det}. Sort of a 2D `GaussHit`.

Caveats

- Does not apply to signal processing of real detector data.
- The effect of digitization is ignored.
  - Scaling is a trivial normalization but I do not know how to incorporate the integer truncation of the digitizer.
  - Approximate by adding a new term to $\mathcal{N}$?
- Filters are applied in signal processing and ignored here.
  - Directly include in a modified $\tilde{\mathcal{R}}^i$ definition.
- Advanced heuristic and AI/ML techniques are used in the real signal processing and can not (?) be accomodated here.
  - But, they tend to bring S_{SP}^{det} closer to $\tilde{S}$.
  - LS4GAN may still learn to bridge their residual differences
- For biploar induction $\mathcal{R}^i$ vanishes at zero frequency.
  - Forming denominator of $\tilde{\mathcal{R}}^i$ may need some "trick" or approximation at low frequency.