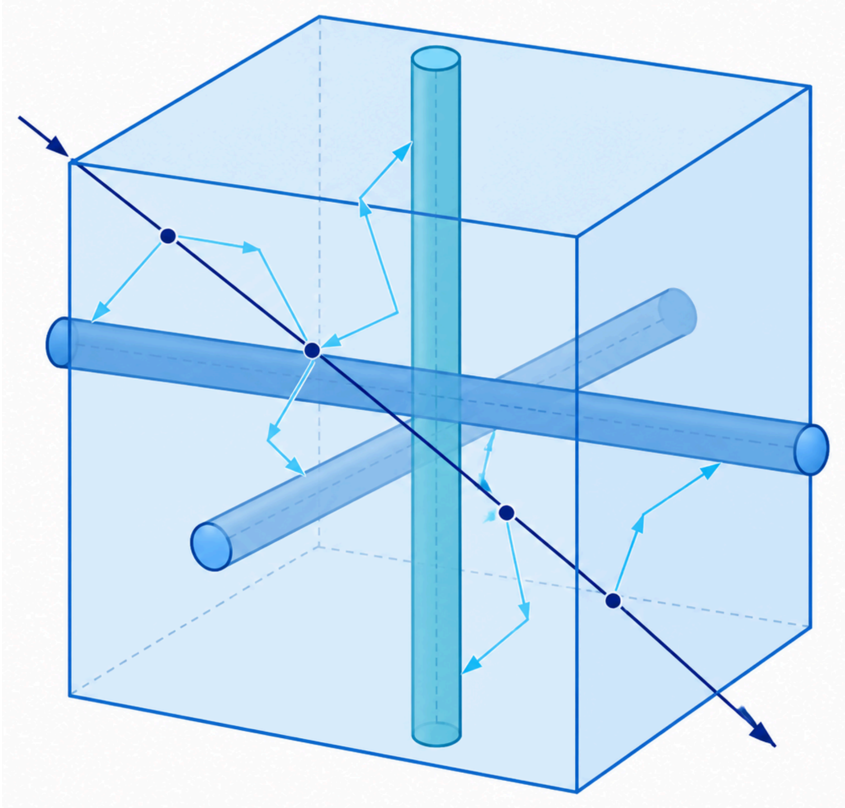
FastSimu: Constraint-Preserving Generative Fast Simulation of Optical-Photon Response at Wavelength-Shifting-Fiber Entry

From detector reconstruction to a modular stochastic simulation framework

Jiacheng Wu
Shanghai Jiao Tong University · Undergraduate researcher
ML4Jets 2026

> Today: how a detector-reconstruction problem became a generative-simulation problem.



↓

STOCHASTIC BOUNDARY RESPONSE

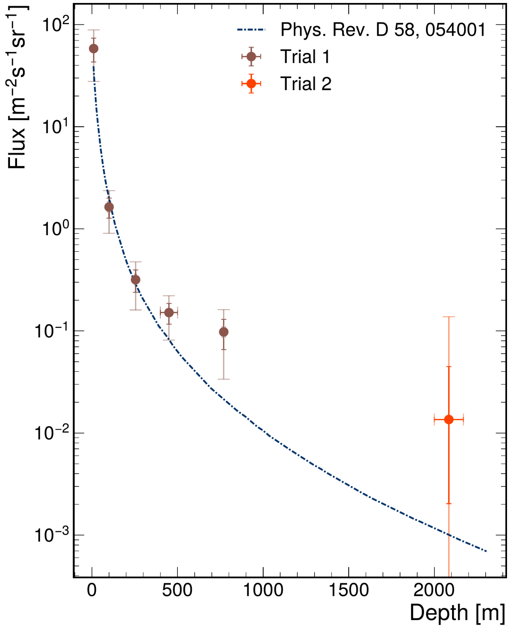
COUNT	TIMING	AXIAL
N	$\Gamma(\alpha, \beta)$	$p(z)$

MuonSLab measured the flux; MuonCube made track reconstruction the problem

The detector question changed from “how many muons?” to “what trajectory produced these hits?”

MuonSLab underwater muon-flux measurement

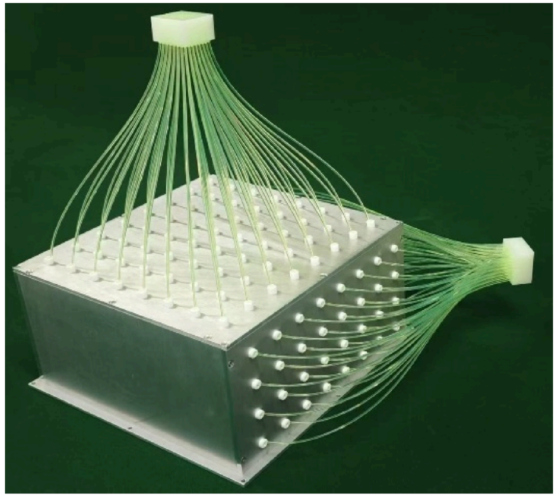
deployment at sea



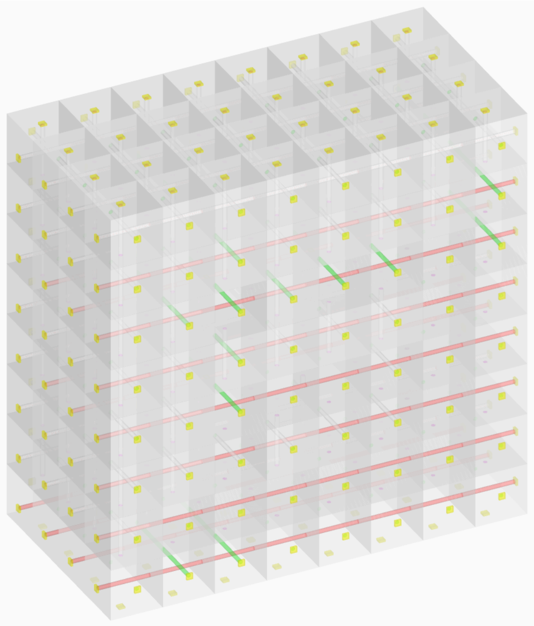
flux vs depth

J. Wu et al., JINST 20 (2025) P07035 · arXiv:2501.17639

MuonCube pixelated scintillator tracker



hardware



simulation model

MuonCube Technical Status Report · Jan. 2026 · project hardware + Geant4 geometry

Track reconstruction pushed the question inside one scintillator voxel.

Two questions pointed me to the same microscopic layer

Array-level reconstruction pushed me below the voxel scale, while a controlled Geant4 diagnostic showed that optical transport accounts for about 93% of event wall time.

Reconstruction
What limits the track reconstruction?

✔	**Timing**	direction of travel
—	**Angular precision**	not sharpened

→ **look inside the voxel**

MuonCube reconstruction study · 2026

Simulation
What limits the simulation speed?

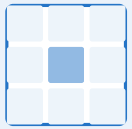
~93% of event wall time

optical transport

→ **replace microscopic tracking**

controlled Geant4 diagnostic

FastSimu R2.1 controlled diagnostic



single-voxel optical response

Geant4: S. Agostinelli et al., NIM A 506 (2003) 250; J. Allison et al., NIM A 835 (2016) 186.

Direct counters changed the intuition: the gain comes from replacing trajectories, not discarding photons

About two thirds of generated optical photons reach the FastSimu boundary; the acceleration opportunity is to generate that boundary ensemble directly instead of replaying every microscopic path.

Generated optical photons	→	FASTSIMU 1.0 BOUNDARY First cladding entry	→	WLS absorbed	→	WLS re-emitted
56.9k		37.8k		24.4k		19.5k
				64.6% of boundary		80.1% of absorbed

66.4% reach this boundary

first cladding entry ≠ WLS absorption ≠ WLS re-emission

median/event · N = 980

THE CORRECTION

NOT

throwing most photons away

INSTEAD

replace tens of thousands of microscopic trajectories with one stochastic boundary response

The first modeling choice was what information should survive

Before choosing a network, I filtered per-photon truth by downstream relevance, symmetry / redundancy, and ensemble compressibility.

Microscopic photon record

- fiber ID
- wavelength λ
- entry time t
- incoming direction Ω
- axial position z
- full path history

too detailed to make every variable an ML target

What should survive?

- **Downstream relevance**
- **Symmetry / redundancy**
- **Ensemble compressibility**

BOUNDARY INTERFACE

FastSimu 1.0

boundary response

per fiber · × 3 orthogonal fibers

COUNT	N	boundary photon count
TIMING	$\Gamma(\alpha, \beta)$	shape + rate
AXIAL SPACE	$p_1 \cdots p_{25}$	25-bin probability profile

This is the released FastSimu 1.0 boundary-response representation — not a universal downstream restart state.

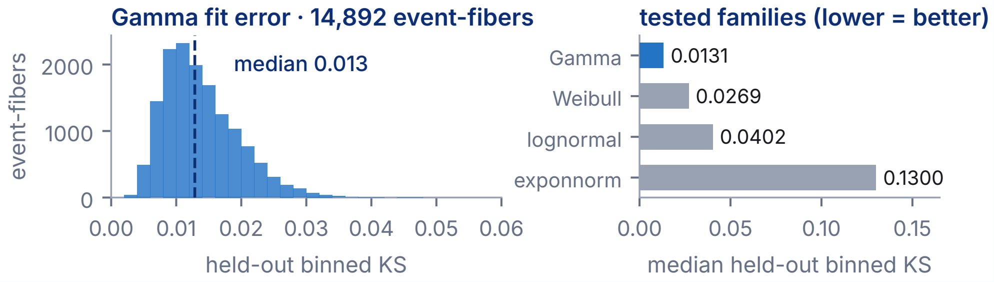
Later restart audit also requires wavelength and incoming direction.

The photon ensemble was more compressible than I expected

Timing compresses cleanly to Gamma parameters; spatial structure was conservatively retained in 25 bins, with later evidence suggesting further compression may be possible.

Timing compressed to a compact Gamma family

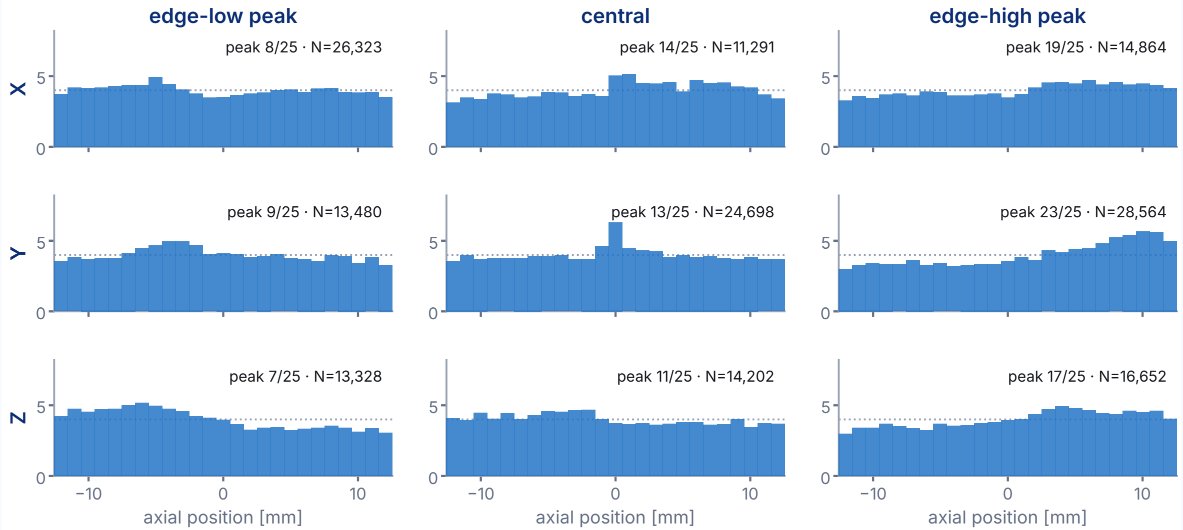
per fiber · held-out photon protocol



$\Gamma(\alpha, \beta)$ median held-out binned KS $\approx$ **0.013**

Two parameters capture the dominant arrival-time shape.

Spatial: preserve the structure first



Event-specific peak positions and local structure motivated conservative 25-bin retention.
Later presentation diagnostics suggest it may compress further.

Release choice: $\Gamma(\alpha, \beta)$ for timing · $p_1 ... p_{25}$ for spatial

FastSimu 1.0 maps a DepositState to 28 values per fiber

The released D→B module represents each fiber with one boundary-photon count, two Gamma timing parameters, and a 25-bin axial probability profile.

DepositStateV1

8 numbers, after energy deposition

position	**x, y, z**
direction	**dx, dy, dz**
track length	L
deposited energy	E_{dep}

→

D → B RELEASED

BoundaryResponseV1 · three orthogonal fibers

- **X** — N — $\Gamma(\alpha, \beta)$ — $p_1 \dots p_{25}$
- **Y** — N — $\Gamma(\alpha, \beta)$ — $p_1 \dots p_{25}$
- **Z** — N — $\Gamma(\alpha, \beta)$ — $p_1 \dots p_{25}$

Optional observation

$H \sim \mathrm{Multinomial}(N, p)$

$\sum H = N$ **exactly**

Optional sampled integer histogram — observation semantics, not part of the required representation.

$1 + 2 + 25 = 28$ **/ fiber** $\rightarrow$ 3 **fibers** $\rightarrow$ **84 required distributional values / voxel**

N is the boundary-photon count at the FastSimu boundary — not all scintillation photons generated in the plastic.

Boundary-response state — not a full downstream restart state.

Different output geometries need different generative heads

FastSimu 1.0 uses separate count, timing, and spatial heads; a separate observation layer gives exact-count integer histograms.

DepositStateV1

8 numbers

same DepositState conditioning for all heads

Count head

Negative Binomial
$D \to (\mu_f, r_f) \to N_f$
discrete · nonnegative

Timing head

$\Gamma(\alpha_f, \beta_f)$
correlated $\log \alpha$, $\log \beta$ · full 6×6 correlation
positive parameters

Spatial head

$\pi_f \in \Delta^{24}$
probability simplex
ILR (24D) → diffusion model → inverse ILR → deterministic $\pi_{f(D)}$

→ **BoundaryResponseV1**

D → B RELEASED

OPTIONAL OBSERVATION

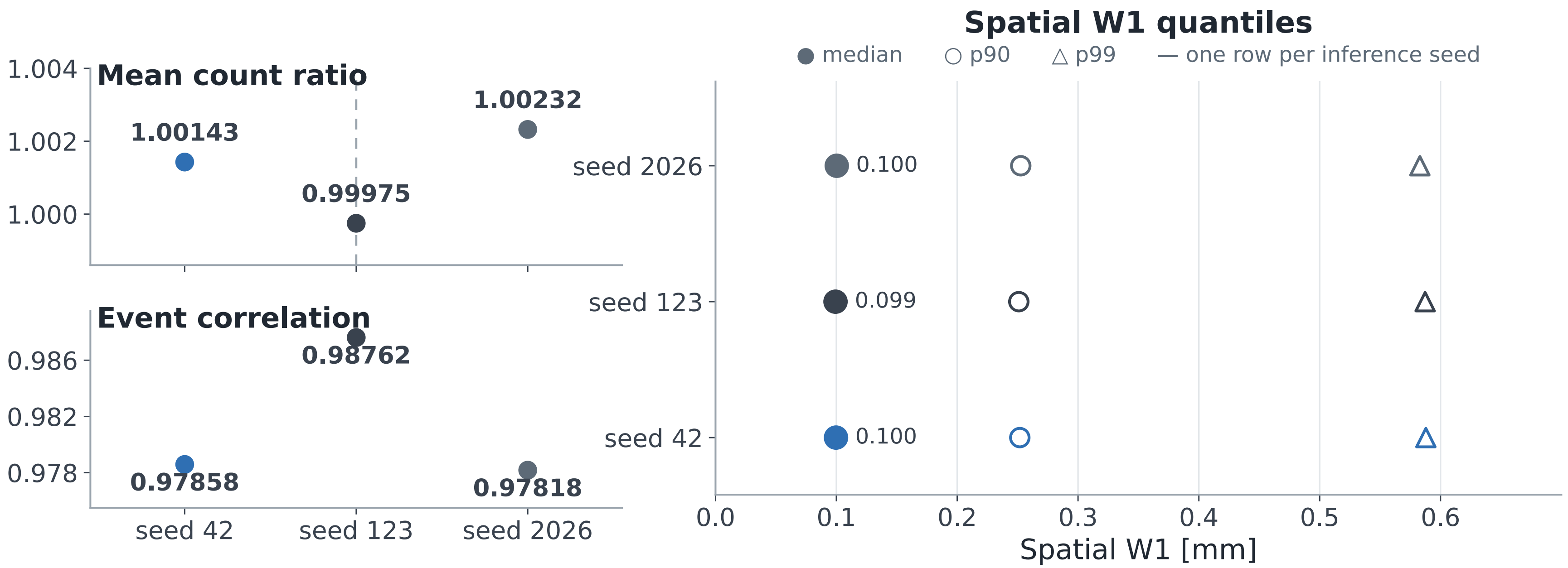
$N_f + \pi_{f(D)} \to \text{Multinomial} \to H_f \cdot \sum H_f = N_f$ **exactly**

observation layer · not a head

ILR: J. J. Egozcue et al., Mathematical Geology 35 (2003) 279–300 · Diffusion: J. Ho et al., NeurIPS 2020; J. Song et al., ICLR 2021

FastSimu 1.0 passed a one-shot final blind audit

On 19,869 valid events from 20 reserved files, the frozen D→B surrogate passed all 12 / 12 predefined audit conditions while preserving exact output constraints.



● **Final blind audit PASS** 19,869 events | integer/nonnegative/conservation = 100% | 0.935 ms/event @ batch 1024

three frozen inference seeds: 42 · 123 · 2026

FINAL BLIND AUDIT PASS

12 / 12

frozen audit conditions

20 reserved files · 19,869 valid events

RELEASED

D → B

EXACT INTEGRITY

integer · nonnegative
conservation = **100%**

The surrogate is fast — with the speed claim kept scoped

In the matched same-boundary benchmark, FastSimu gives 4,397× amortized throughput acceleration; the released module runs separately at 0.9347 ms/event at batch 1024.

Matched same-boundary benchmark

Geant4
optical boundary
5.829
± 0.103 s / event

→

FastSimu matched pipeline
batch 100
1.3257
ms / event

→

4,397×
amortized throughput

same worker · same boundary · batch 100 · amortized throughput

SEPARATE ABSOLUTE RUNTIME

0.9347

ms / event

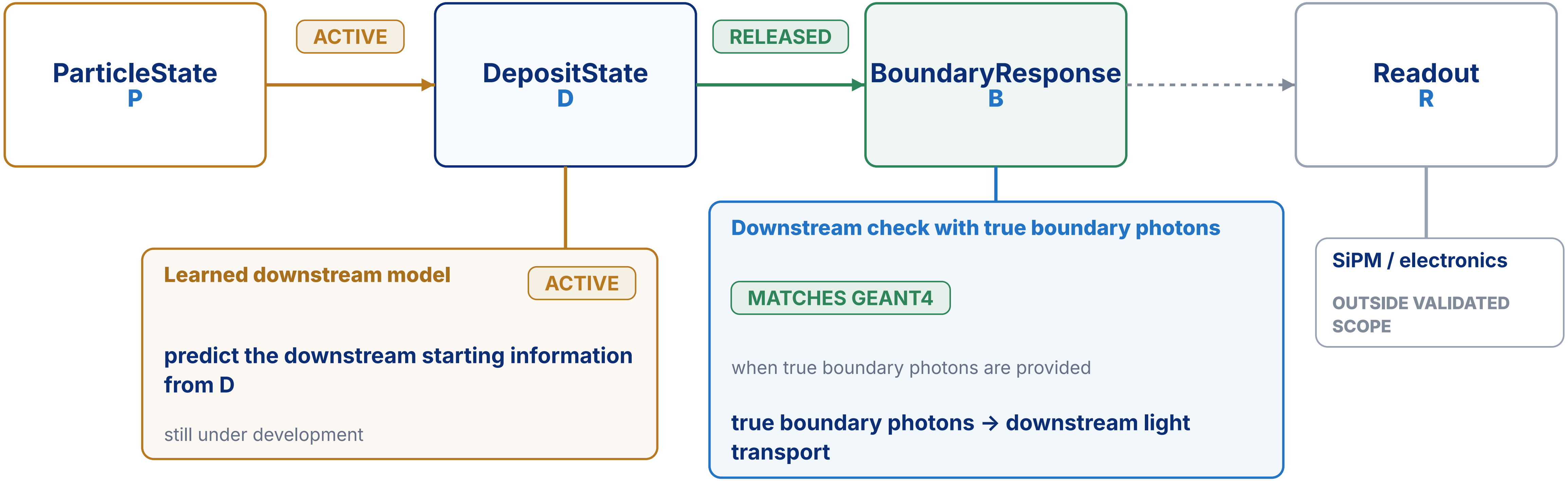
batch 1024

different benchmark protocol
not the speedup denominator

Scope: D→B boundary surrogate; not end-to-end P→R.

FastSimu 1.0 is useful, but it is not particle-to-readout simulation

The released D→B module is useful on its own; upstream deposition and downstream continuation remain active research, while full electronics readout stays outside the current scope.



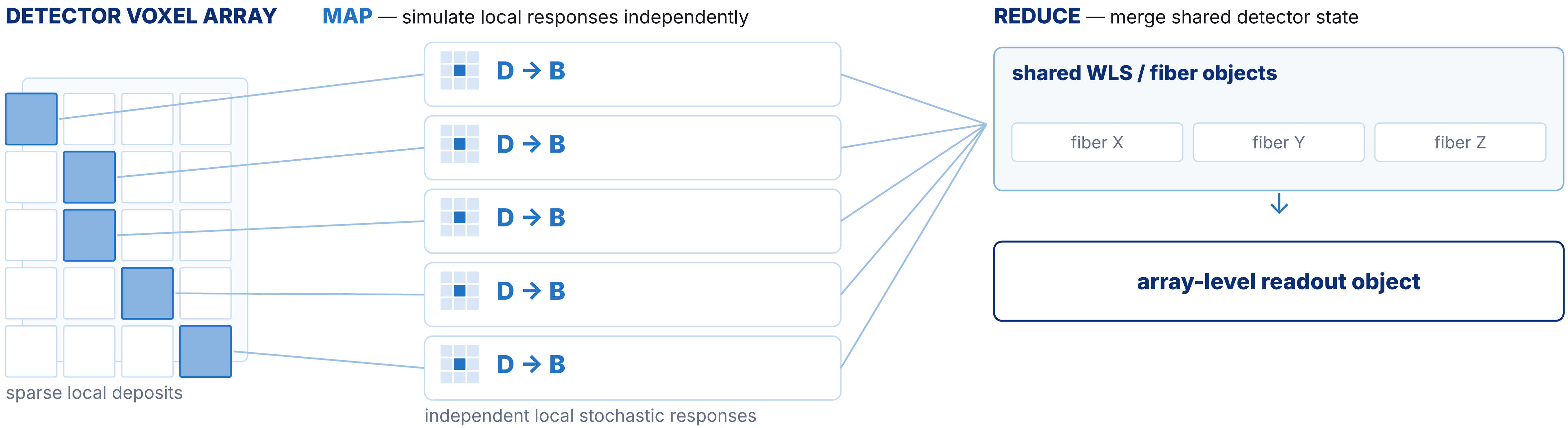
End-to-end P → R: NOT CLAIMED

Map–Reduce suggests how local voxel surrogates could scale to an array

A possible array architecture is to simulate sparse local voxel responses independently, then merge their contributions into shared fiber and readout state.

Same modular boundary → array-scale systems view

SYSTEMS VISION — not yet an array-level performance claim



REUSE local D → B abstraction **OPEN** inter-voxel transport · shared-fiber coupling · array-level timing · resource scaling

MapReduce: J. Dean & S. Ghemawat, OSDI 2004, 137–150.

Define the stochastic object before choosing the network

The transferable lesson is not one architecture: define the physical interface, measure irreducible stochasticity, respect the output geometry, and freeze validation before looking at the answer.

1

Choose the physical interface first.

Define what enters, what leaves, and what downstream needs.

- FastSimu example: first fiber-cladding entry boundary

2

Measure irreducible stochasticity.

Repeated conditions separate physical spread from model error.

- FastSimu example: same-particle-state Geant4 replicas

3

Match the model to the output geometry.

Counts, positive parameters, and profiles obey different constraints.

- FastSimu example: separate heads for different output spaces

4

Freeze validation before looking at the answer.

Lock splits, metrics, and pass/fail gates before evaluation.

- FastSimu example: one-shot frozen blind audit

FastSimu 1.0 is released; moving the boundary is the next research frontier

FastSimu 1.0 shows that a physically defined stochastic boundary can become a released fast-simulation module; extending that idea upstream, downstream, and eventually across an array is the next step.

Where FastSimu stands

RELEASED — **D → B · FastSimu 1.0**
frozen boundary-response surrogate

CHECKED WITH TRUE INPUT — **downstream WLS/fiber transport**
matches Geant4 when true boundary photons are supplied

ACTIVE — **particle → deposit modeling**
the same incoming particle can produce different deposits

ACTIVE — **learned downstream continuation**
still under development; not part of the current release

NOT PART OF THE CURRENT RELEASE — **SiPM / electronics**
full readout remains outside the validated scope

SUMMARY

01 BOUNDARY — **Replace microscopic trajectories with a stochastic boundary response.**

A physical interface turns expensive optical transport into a well-defined generative problem.

02 RELEASE — **FastSimu 1.0 turns that idea into a released D → B module.**

Frozen blind validation passed; the large speed gain is explicitly scoped to the same boundary.

03 FRONTIER — **The next challenge is moving the learned boundary.**

Upstream deposits, learned downstream transport, and array composition remain separate research problems.

Thank you — happy to take questions.