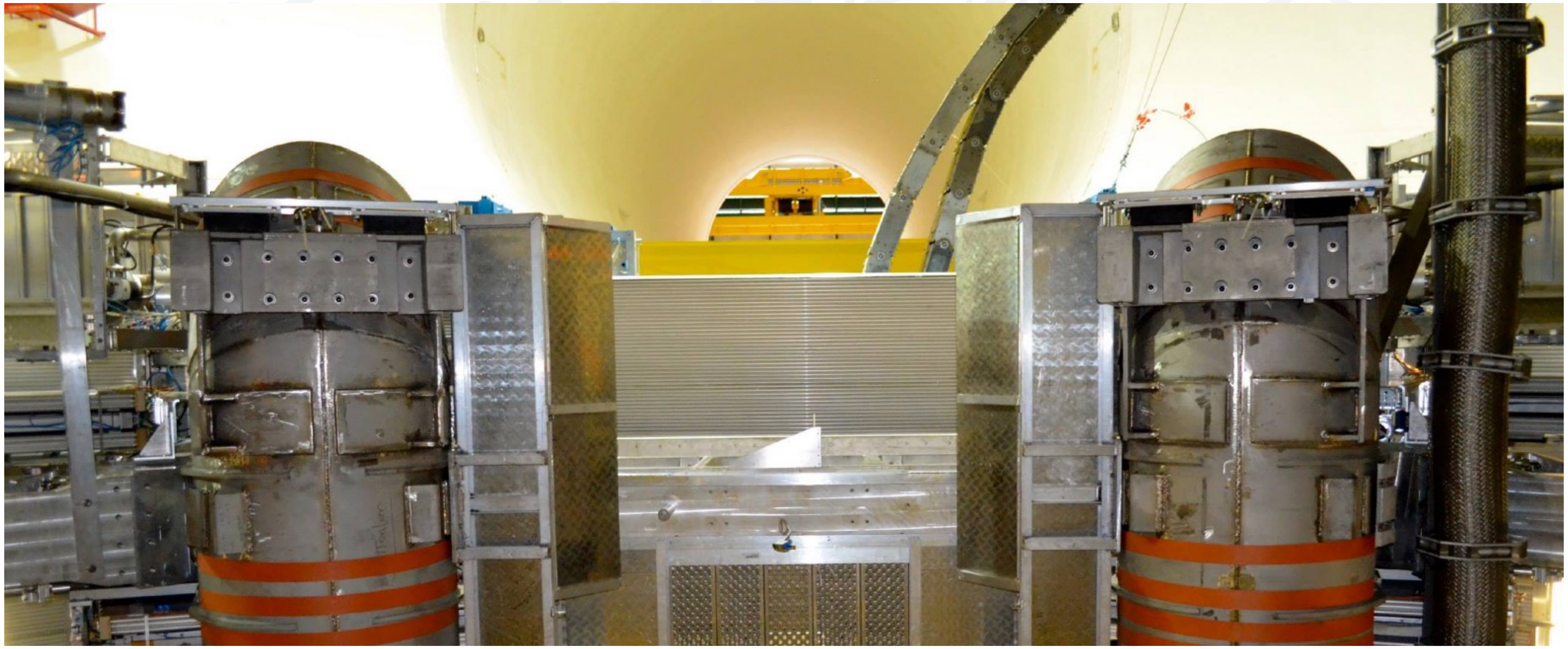
Efficient tracking in the search for rare processes

13th December 2018

Rebecca Carney



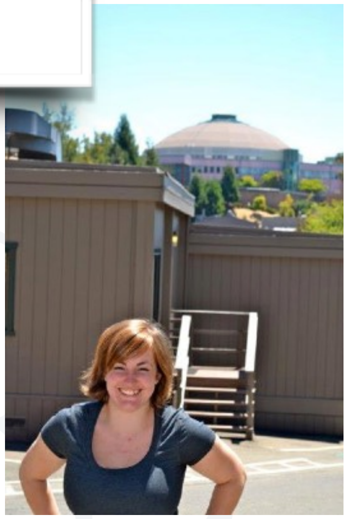
Berkeley Lab

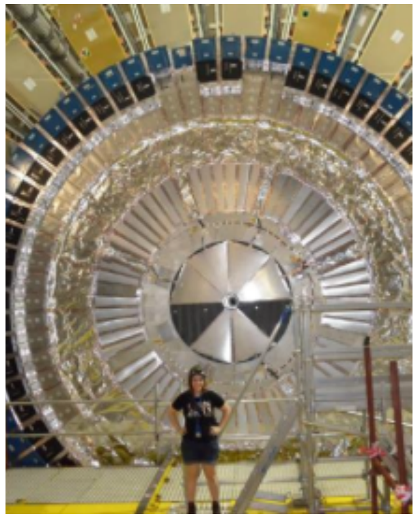
rcarney@lbl.gov

About my degree program

Student at Stockholm university

Graduate RA at LBNL





PhD in Instrumentation Physics

- Four year PhD programme with Instrumentation Physics dpt at SU.
- Student of SU but based at Berkeley Lab.
- Unique opportunity: work in a US National Lab and near industry, whilst on a European PhD.

Sam Silverstein

Sara Strandberg

Maurice Garcia-Sciveres

Paolo Calafiura

Laura Jeanty

Simone Pagan Griso

Lead advisor **Advisor**

Stockholm University

Advisor, technical supervisor

Technical supervisors

BERKELEY LAB

rcarney@lbl.gov

About my degree projects

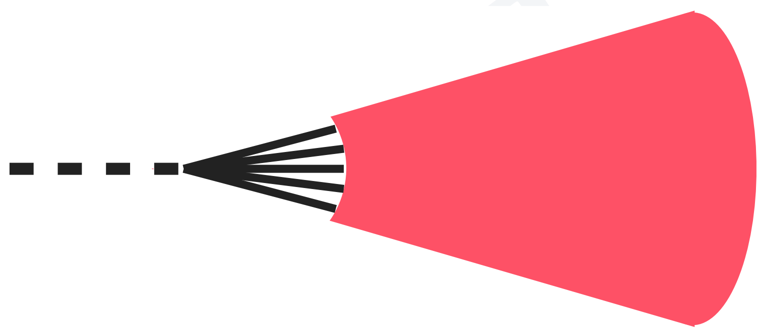
- First two years of degree concluded by a Swedish Licentiate thesis and defense with external committee and opponent.
- My program's focus is instrumentation; data analysis work is not a required part of the degree.
- However, I requested adding a data analysis component to my thesis.

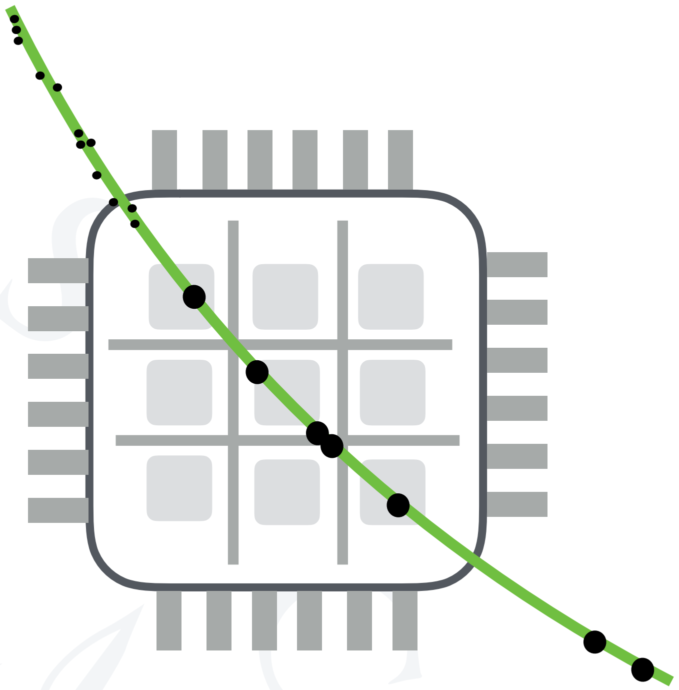
Instrumentation for silicon tracking at the HL-LHC

Licentiate thesis

Many topics in PhD, common theme of tracking

Outline



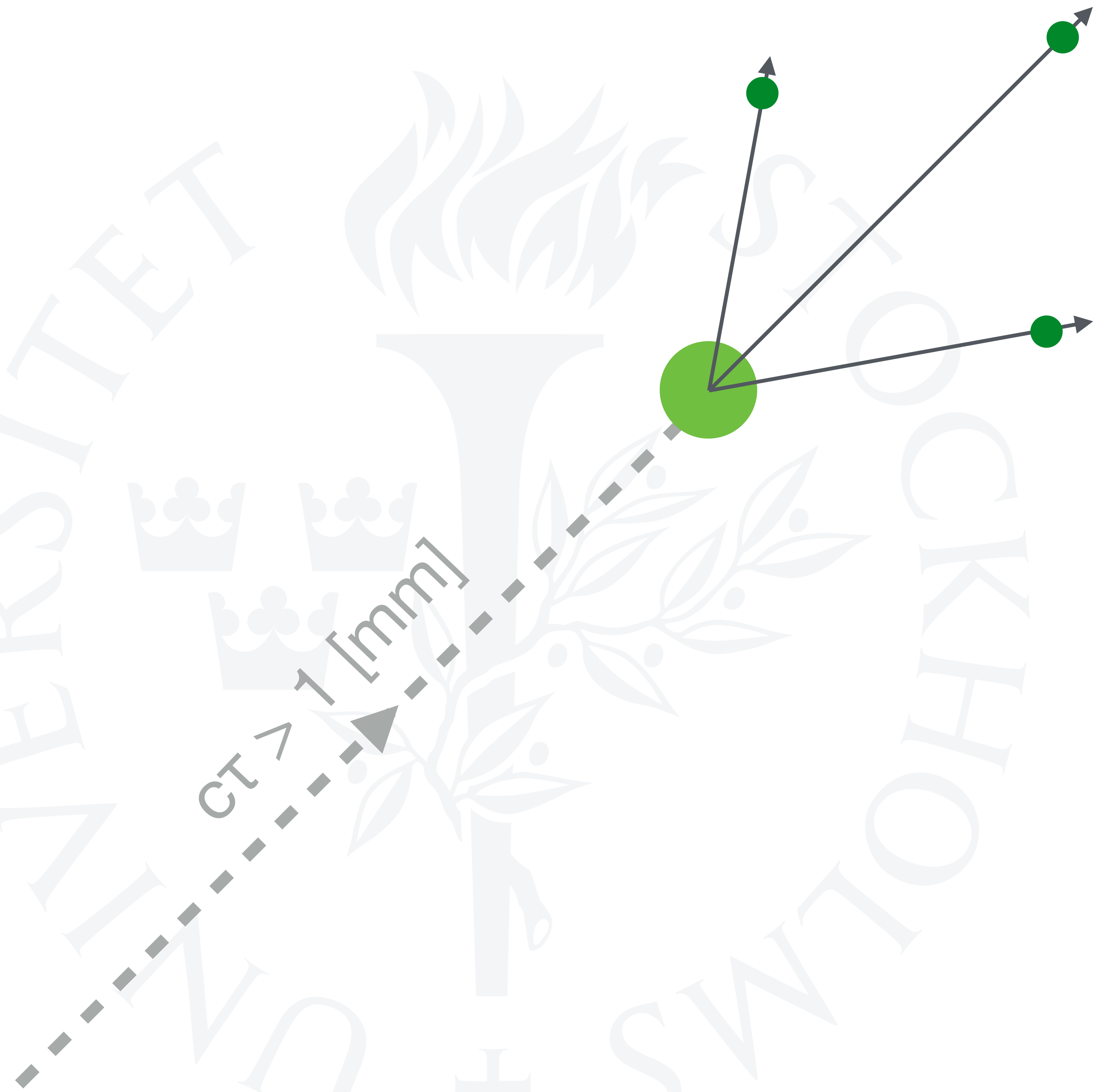


(i) Search for physics beyond the Standard Model with displaced vertices decaying in the Inner Detector to, and produced in association with, high-energy jets using 13 TeV ATLAS Run 2 data

(ii) Implementation of a discrete spiking Kalman filter on IBM's TrueNorth neuromorphic chip

rcarney@lbl.gov

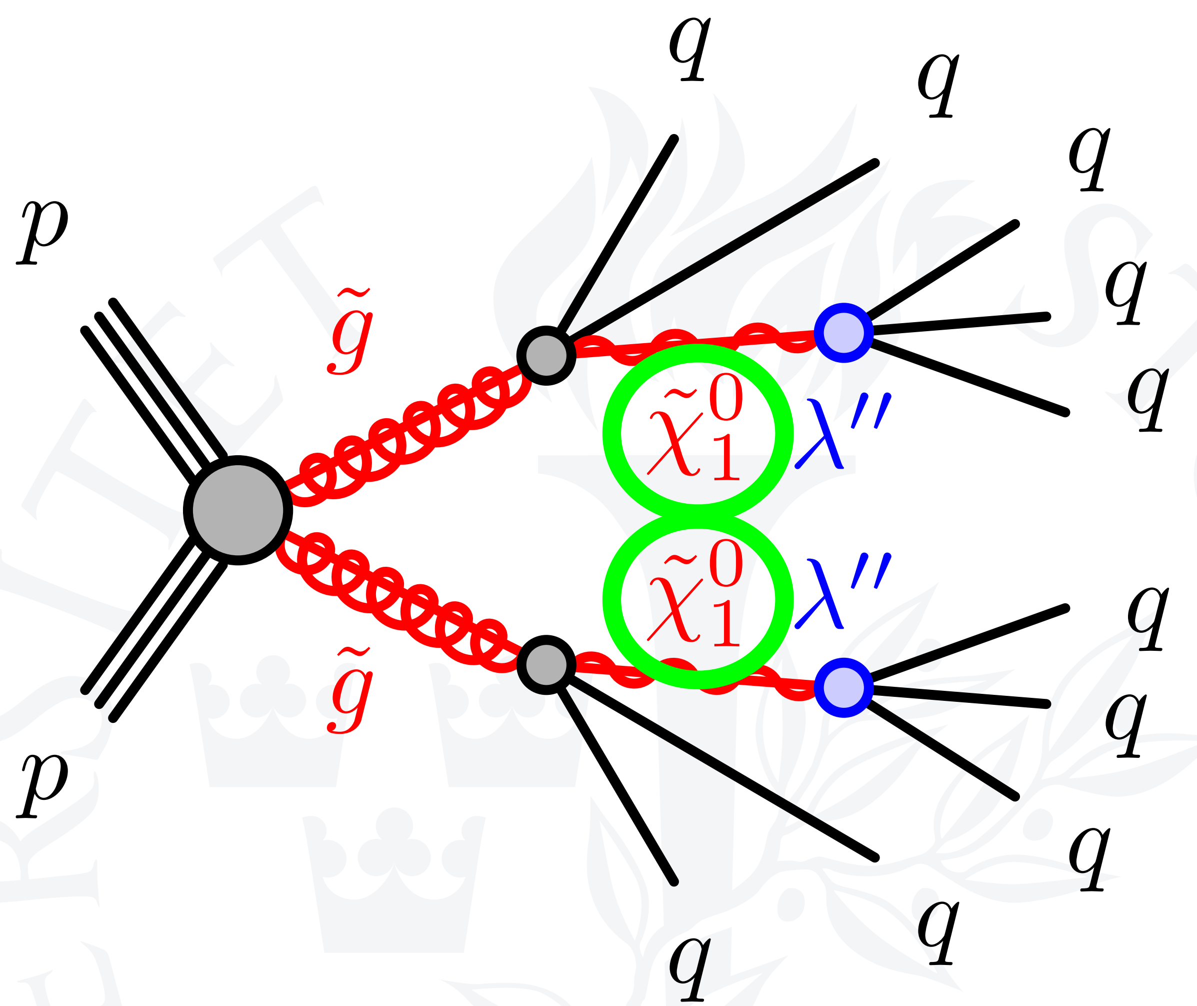
Long-lived particles



Searches for long-lived particles beyond the Standard Model

rcarney@lbl.gov

Long-lived particle mechanisms



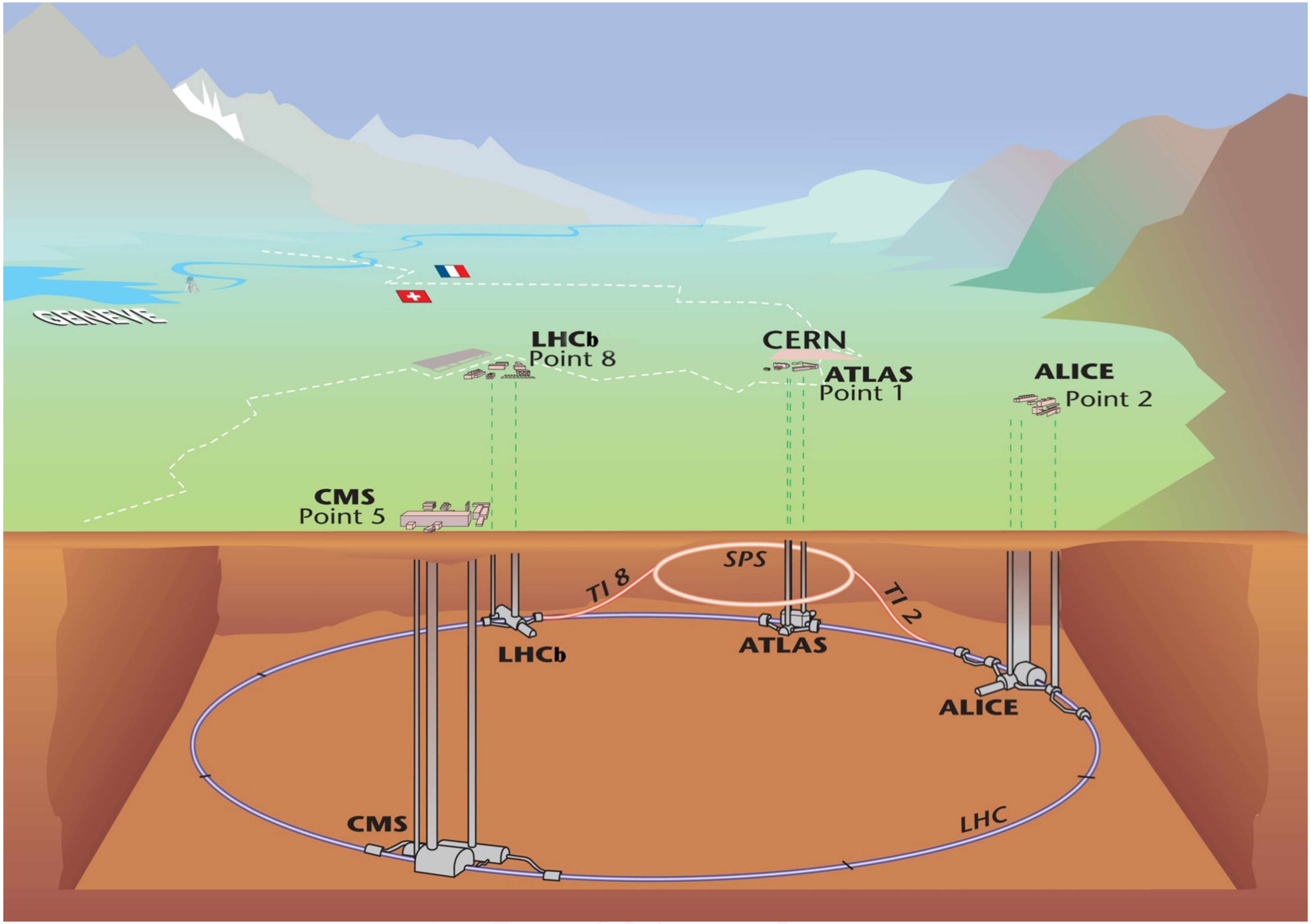
Long-lived particles circled

- Small coupling (e.g B-hadron decay)
- Nearly degenerate masses (e.g. free neutron decay)
- Heavy mediator (e.g. charged pion decay)

My analysis: targeting this signal model, long-lived neutralino decaying via RPV coupling to, and produced in association with, high-energy jets.

How do we search for this process?

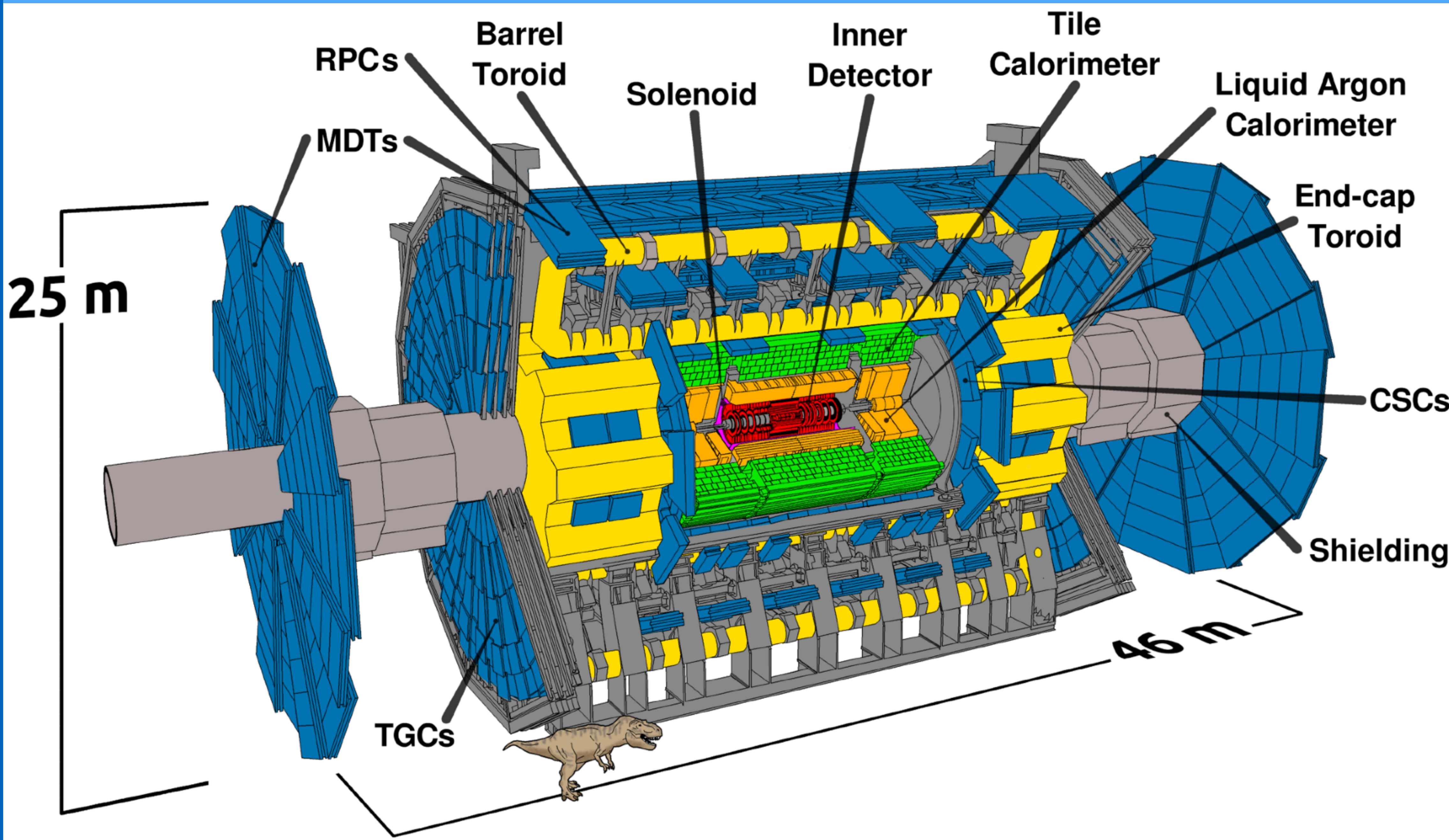
LHC @ CERN



https://cds.cern.ch/journal/CERNBulletin/2008/38/News%20Articles/1125888

The Large Hadron Collider accelerates protons which collider in detectors shown above with a centre-of-mass energy of 13 TeV.

The **ATLAS** detector

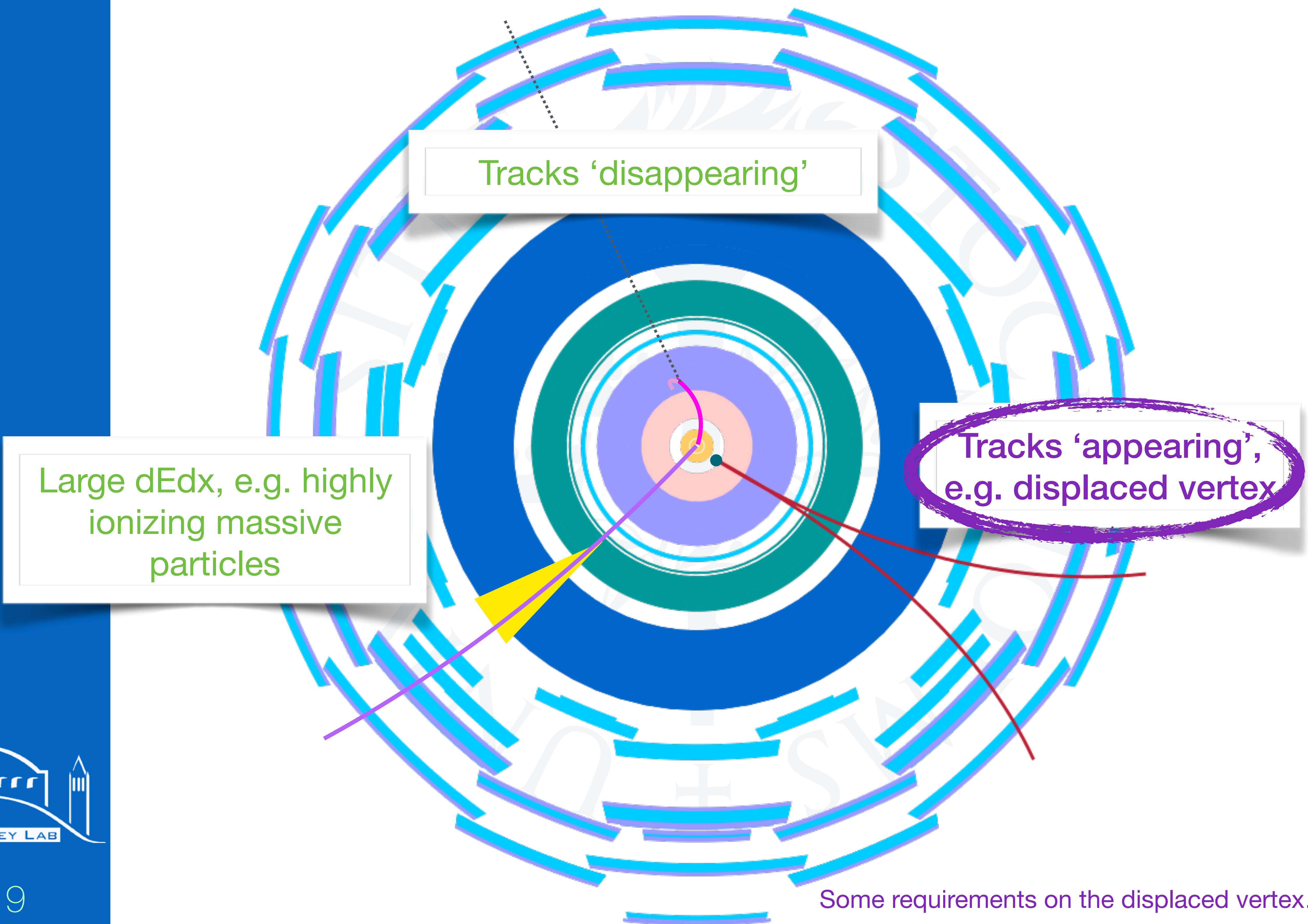


https://cds.cern.ch/record/1449722/files/CERN-THESIS-2012-053.pdf

Inner Detector	LAr Calorimeters	Muon detector
Solenoid	Tile Calorimeter	Toroid

What do long-lived particles look like in the detector?

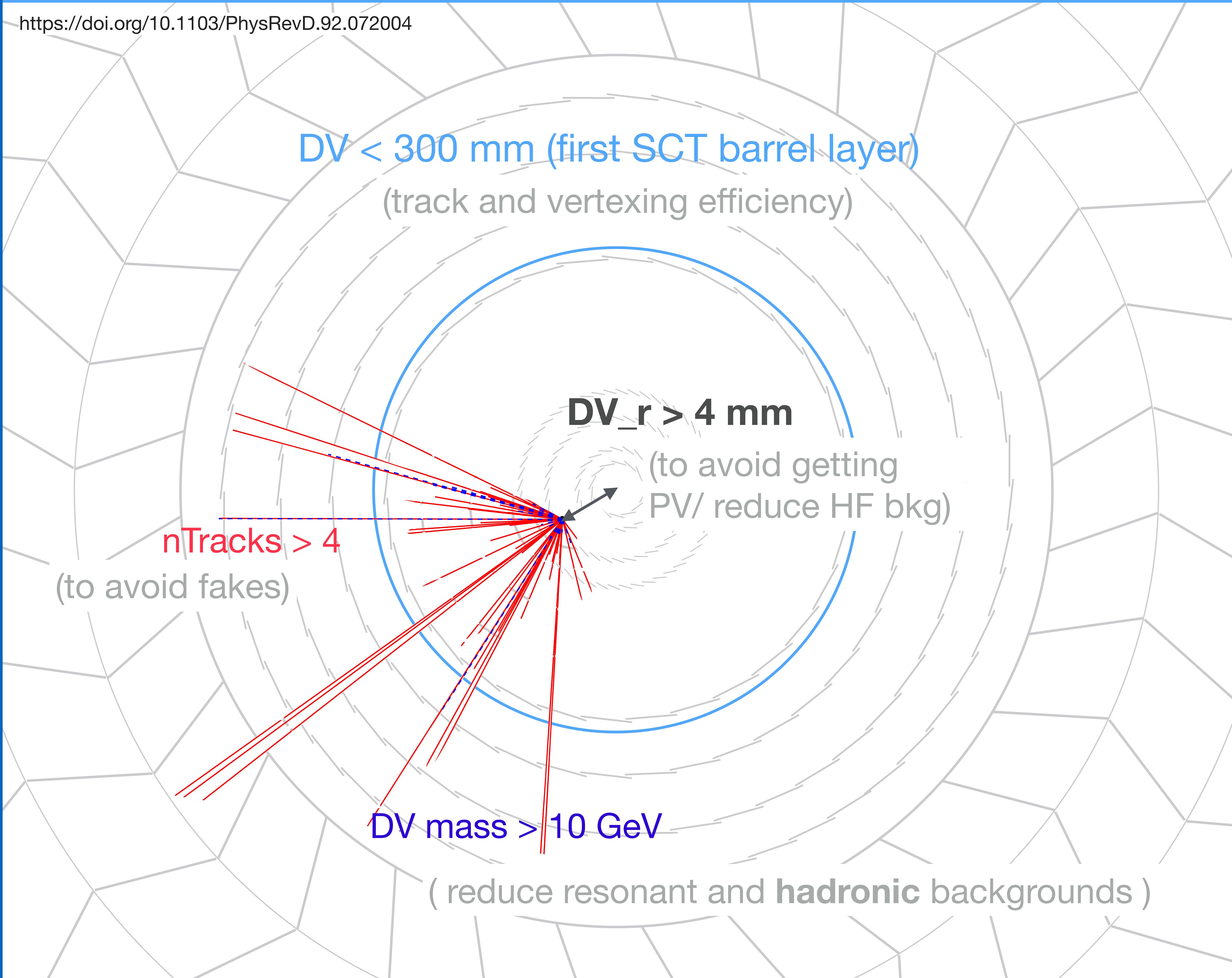
Long-lived particle signatures



Some requirements on the displaced vertex…

DV **Selections**

https://doi.org/10.1103/PhysRevD.92.072004



rcarney@lbl.gov

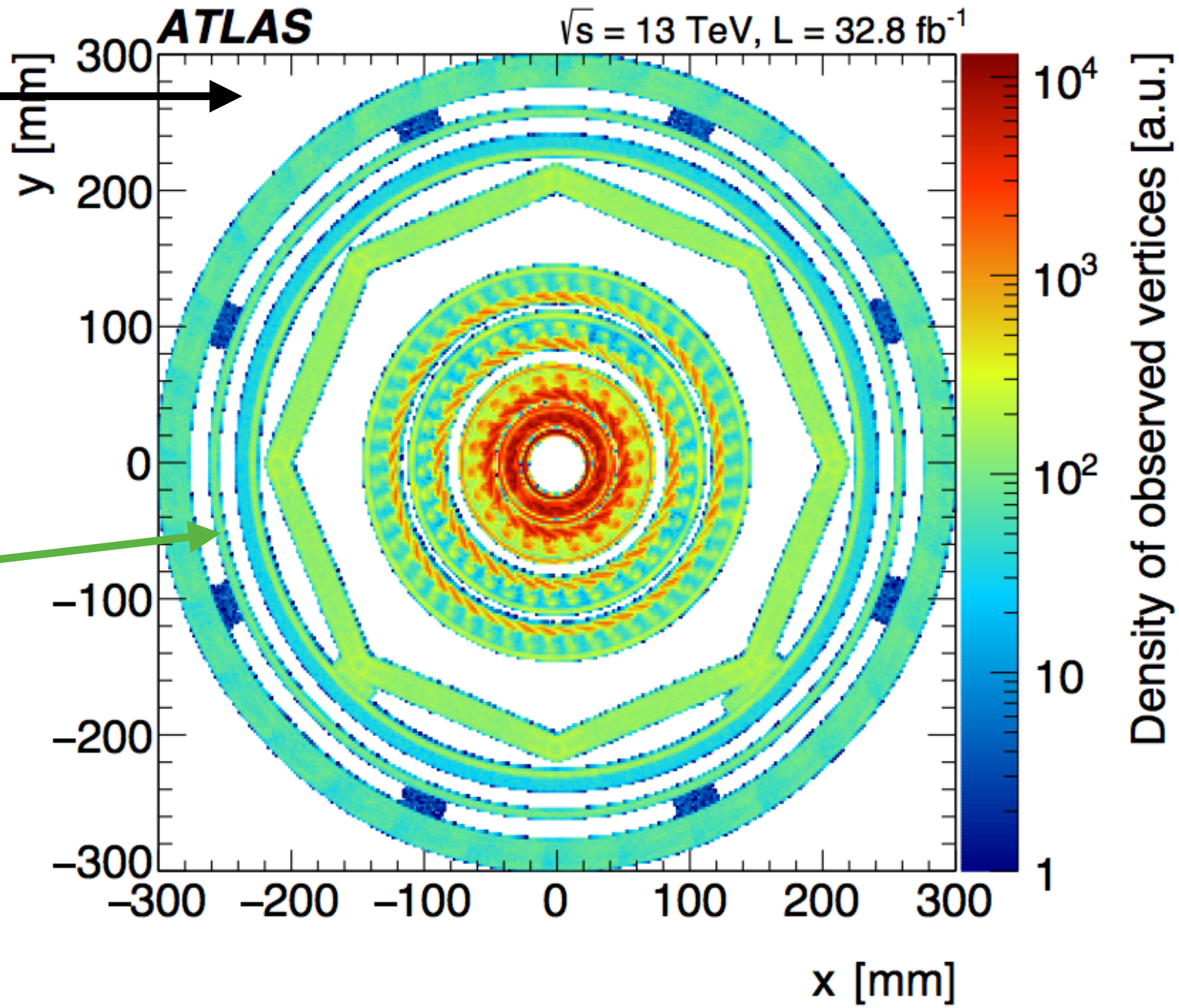
Data-driven material-veto map

Fiducial region

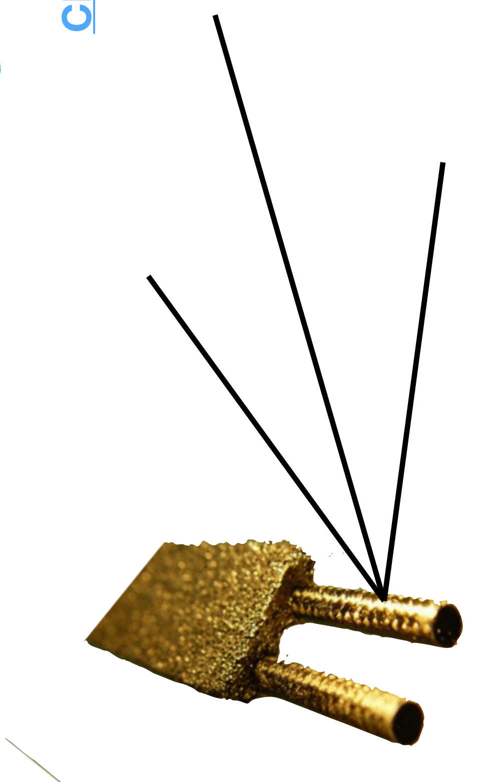
Vetoed region

Hadronic interactions with material is a significant background for DV's.

Veto map excludes around 42% of fiducial volume!



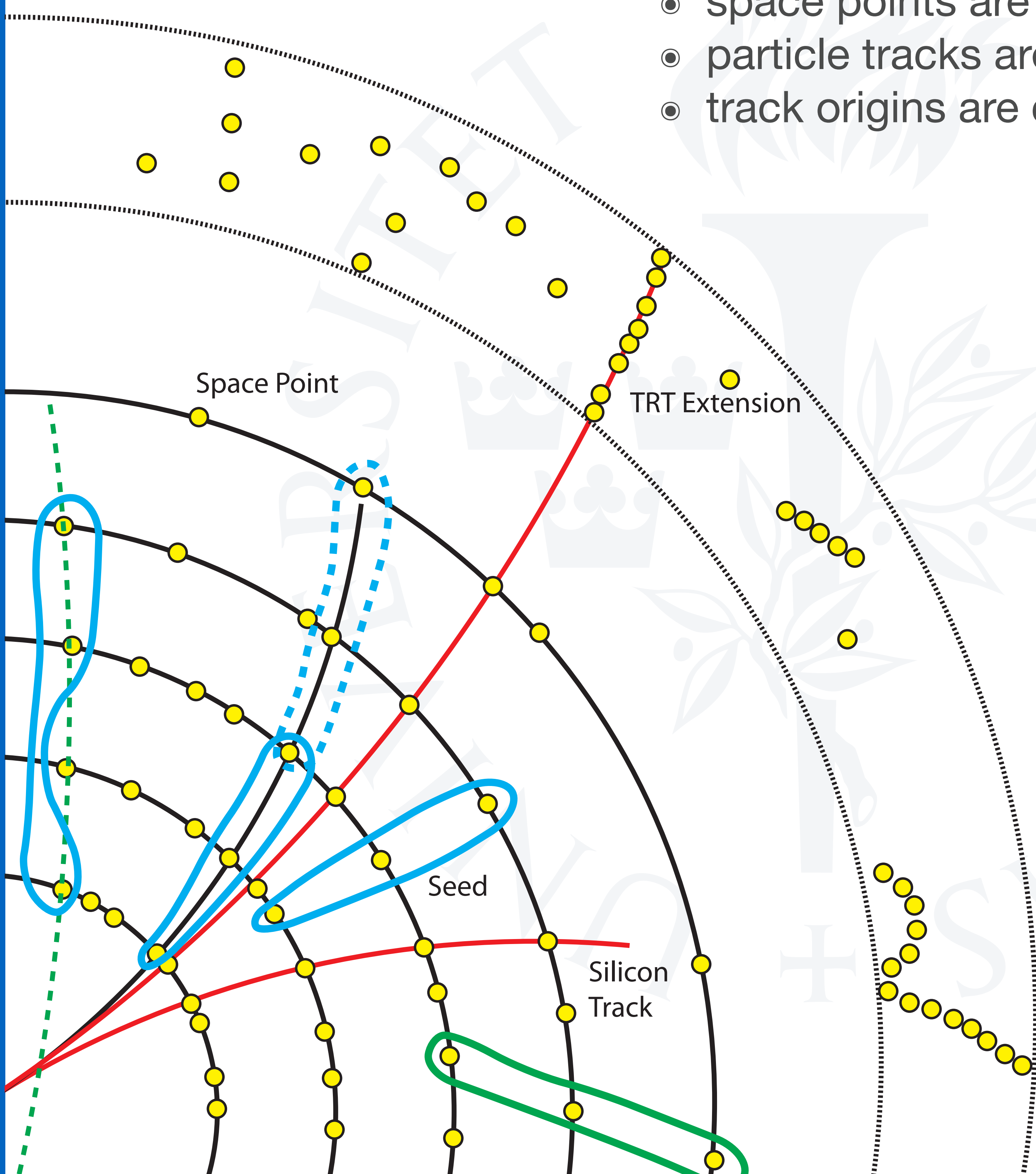
CERN-EP-2017-202



Let's go back to how we construct vertices in the first place….

Tracking and reconstruction

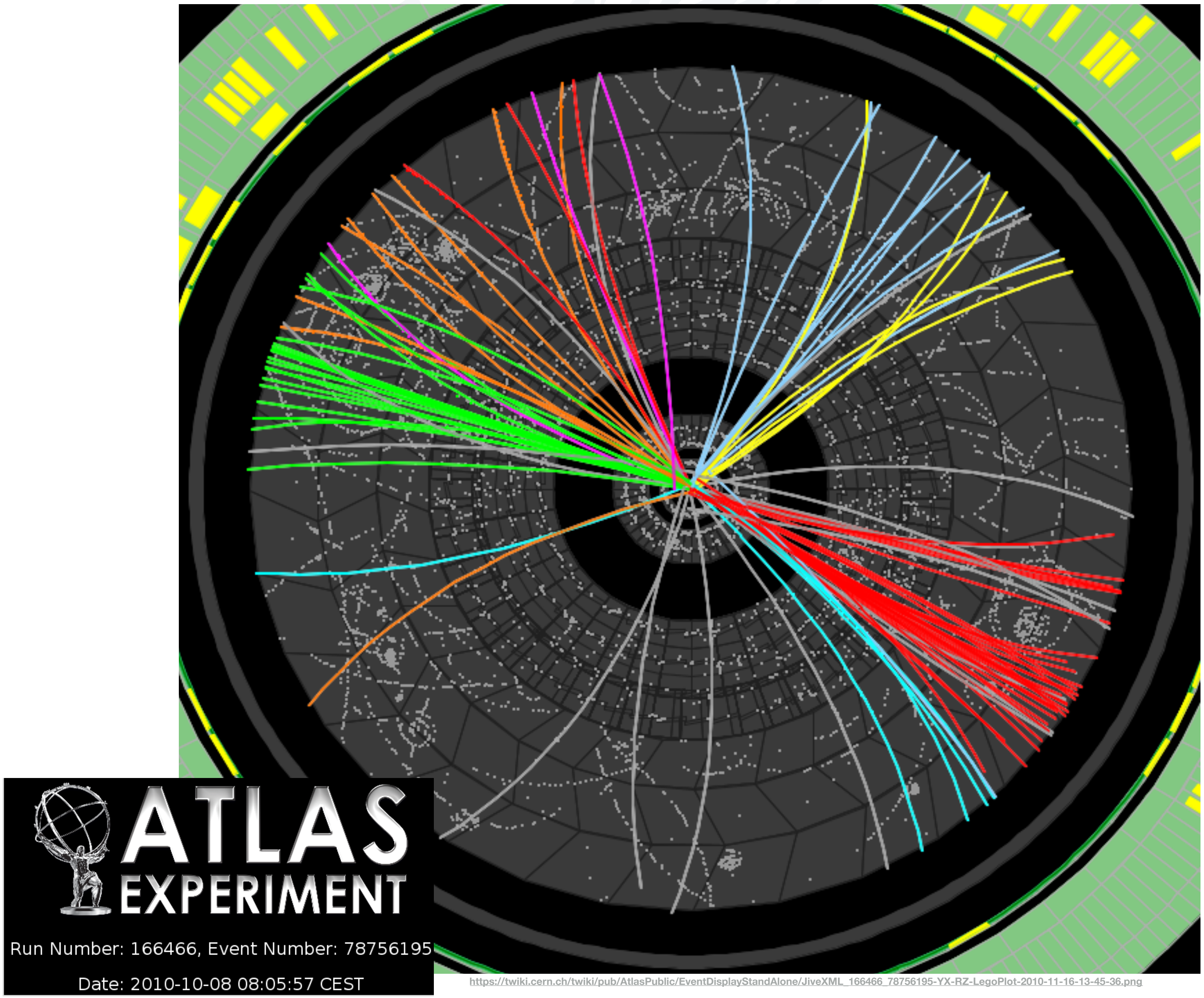
- space points are created from charge clusters
- particle tracks are created from space-points
- track origins are clustered into a 'vertex'



cartoon courtesy of Heather Gray

rcarney@lbl.gov

With so many tracks, must reduce combinatorics by imposing restrictions



https://twiki.cern.ch/twiki/pub/AtlasPublic/EventDisplayStandAlone/JiveXML_166466_78756195-YX-RZ-LegoPlot-2010-11-16-13-45-36.png

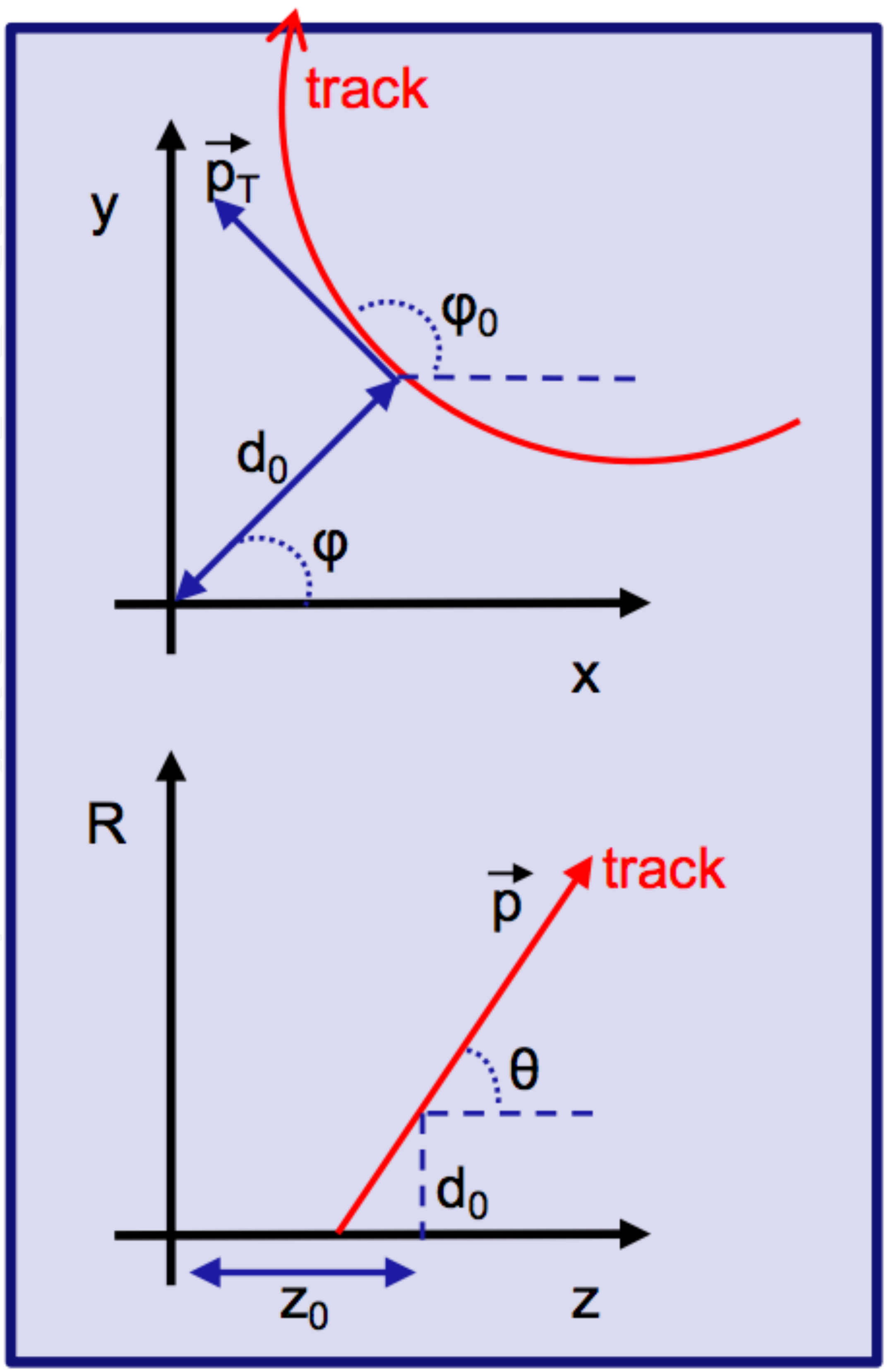
BERKELEY LAB

rcarney@lbl.gov

Restrict combinatorics

A track can be characterized by a set of parameters…

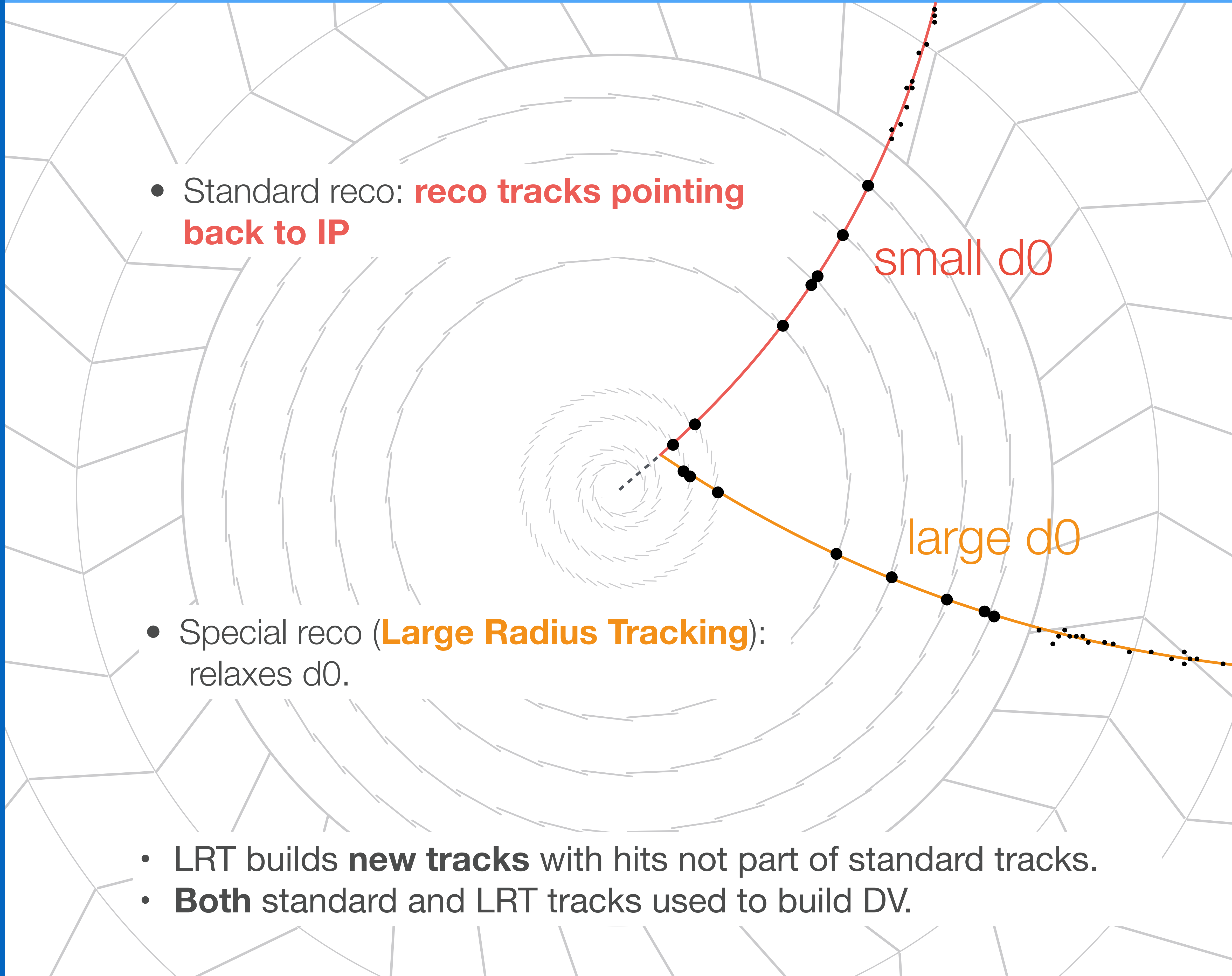
- We expect most particles to originate from centre of detector
- small distance of closest approach (d0, z0)
- restricting this value helps reduce combinatorics



Perigee parameters

rcarney@lbl.gov

DV reconstruction

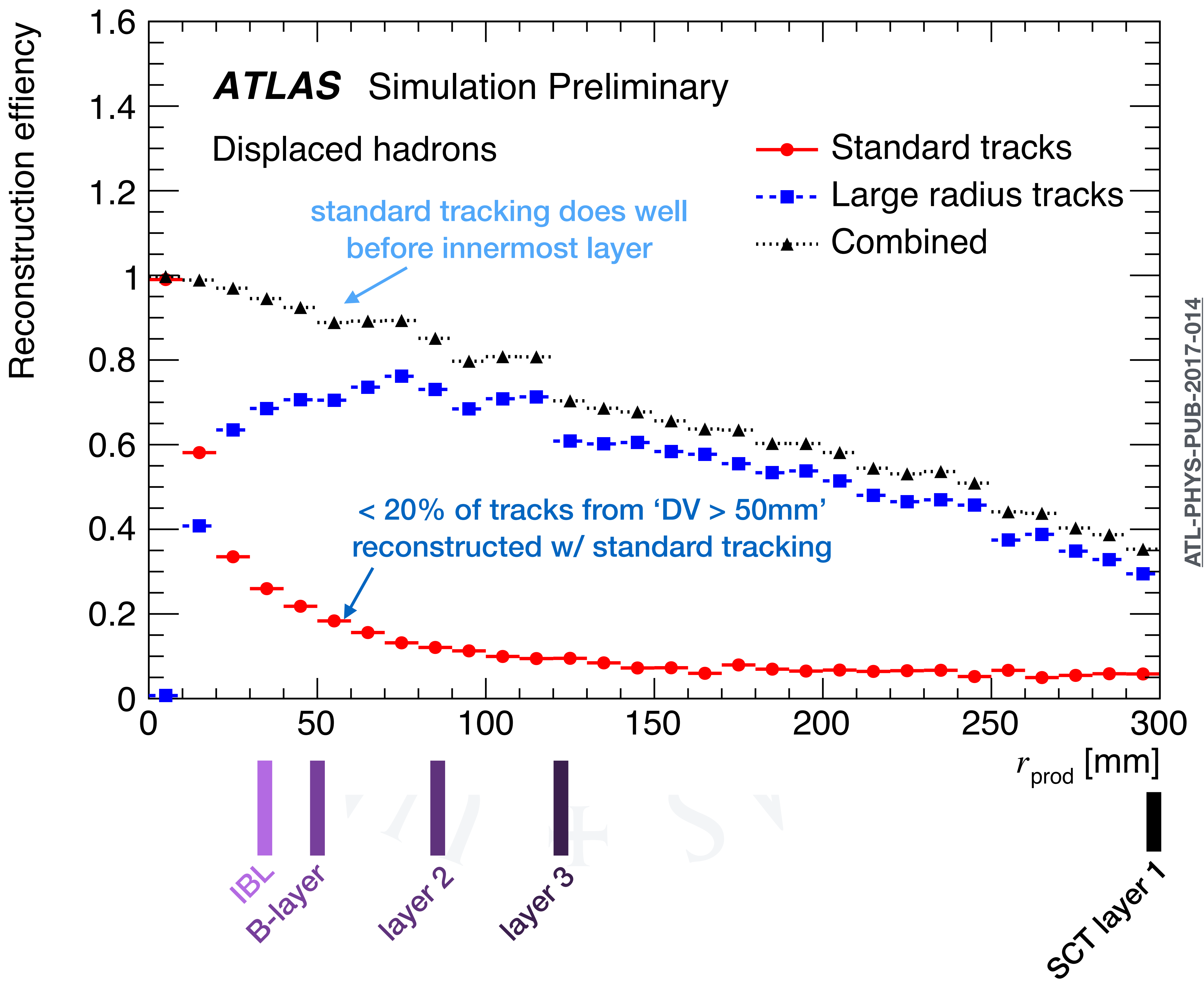


- Standard reco: **reco tracks pointing back to IP**
- Special reco (**Large Radius Tracking**): relaxes d0.

- LRT builds **new tracks** with hits not part of standard tracks.
- **Both** standard and LRT tracks used to build DV.

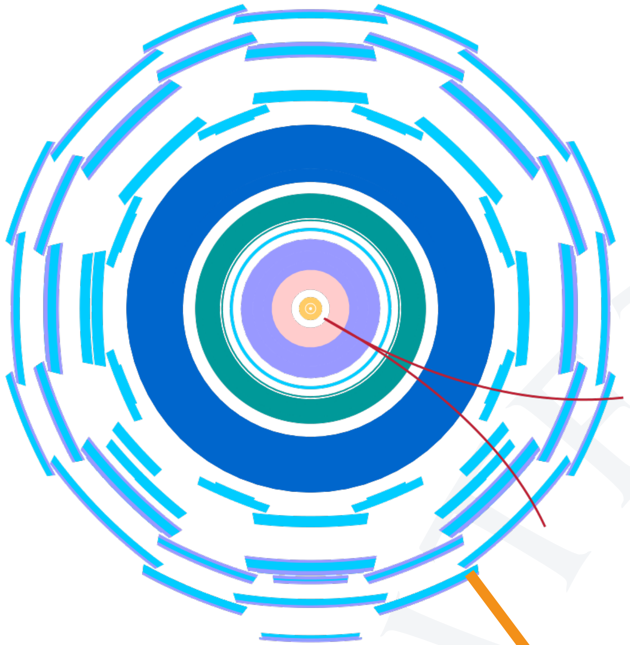
BERKELEY LAB

Large Radius tracking

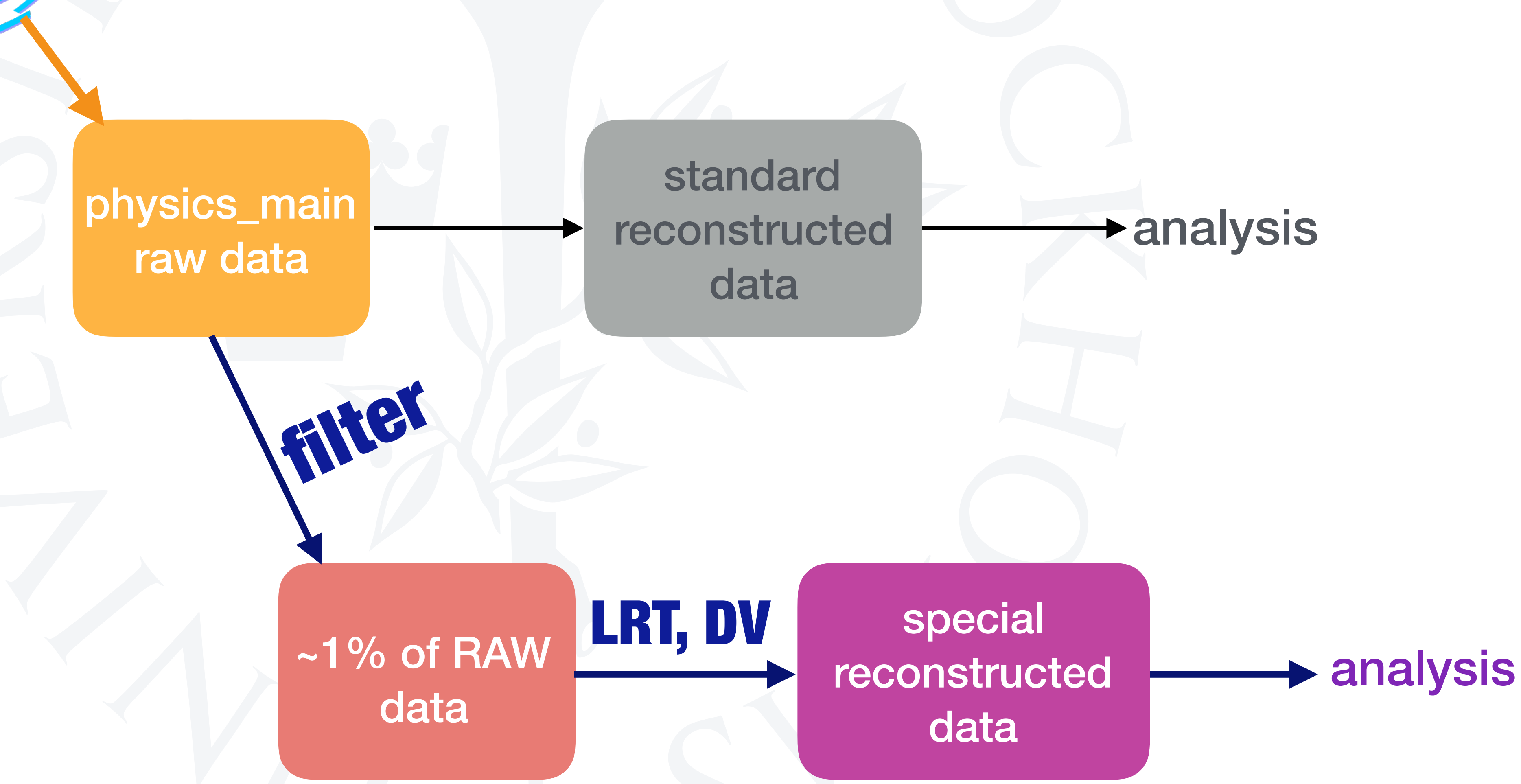


rcarney@lbl.gov

Data flow



Must access RAW data to run Large Radius Tracking (LRT) and Displaced Vertex (DV) reconstruction algorithms.
Expensive & time consuming…



… so filter out interesting events before reconstruction.

rcarney@lbl.gov

Filter example DV Multijet

DV_MultiJetFinalTracklessFilter =

next slide

{ tracklessJets &

[(2j & 3j) ^ 4j ^ 5j ^ 6j ^ 7j] &

DVMultiJetTriggerFlags }

# jets	Lowest unprescaled triggers in Run2 [GeV]
3	175
4	85
5	55
6	45
7	45

# jets	Jet pT cut [GeV]
2	220
3	120
4	100
5	75
6	50
7	45

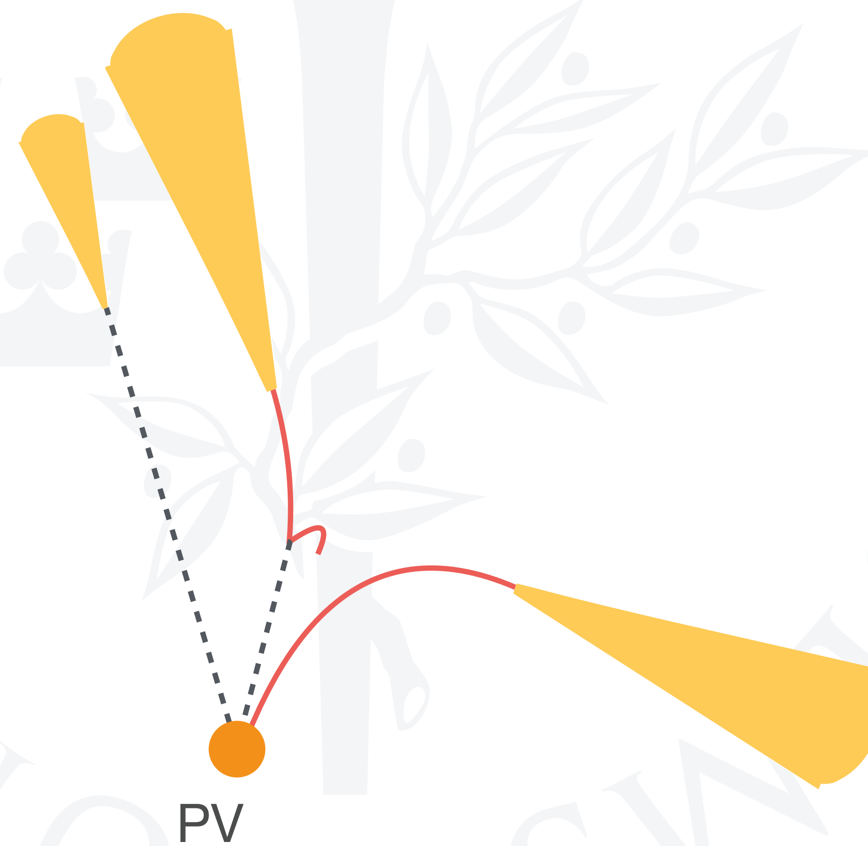
Filter example DV Multijet

DV_MultiJetFinalTracklessFilter =

{ tracklessJets &

[(2j & 3j) ^ 4j ^ 5j ^ 6j ^ 7j] &

DVMultiJetTriggerFlags }

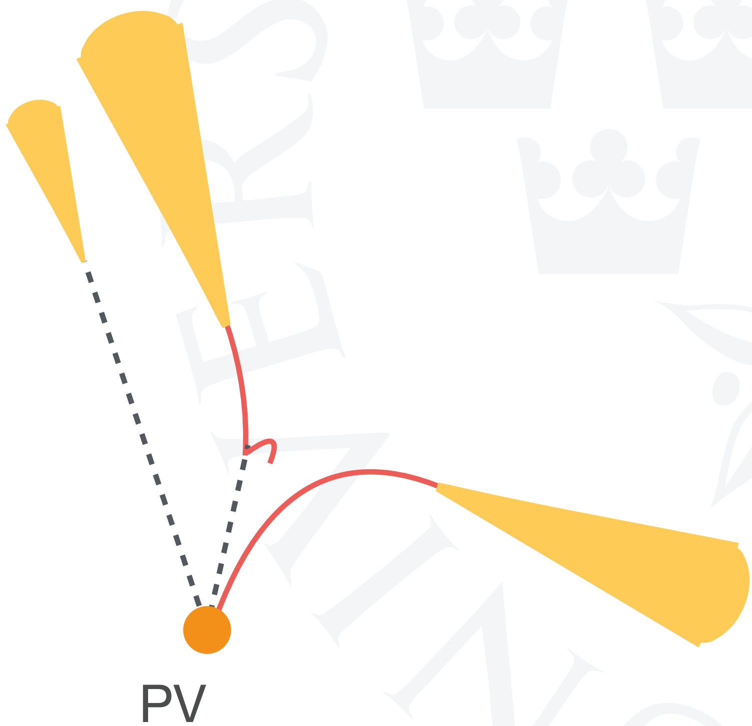


jet is 'trackless' if Σp_T^{track} < 5 GeV

tracks matched to PV & jet

Filter example DV Multijet

DV_MultiJetFinalTracklessFilter =
{ tracklessJets &
[(2j & 3j) ^ 4j ^ 5j ^ 6j ^ 7j] &
DVMultiJetTriggerFlags }

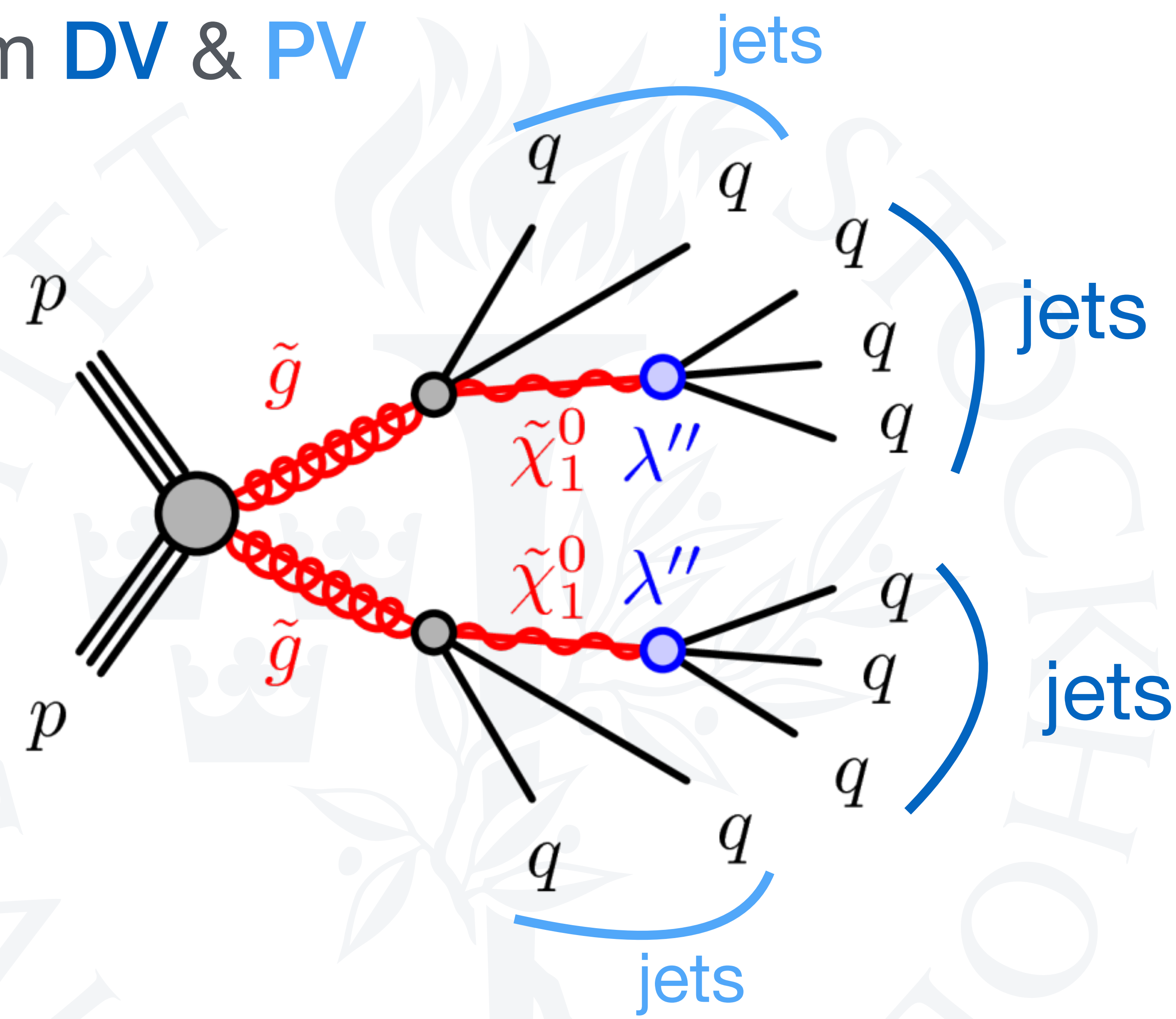


# jets	Jet pT cut [GeV]	Eta cut
single	70	2.5
double	25	2.5

tracklessJets = single **OR** double

To pass the filter, a triggered events has to have at least:
1 or 2 trackless jets (single or double)

Jets from DV & PV

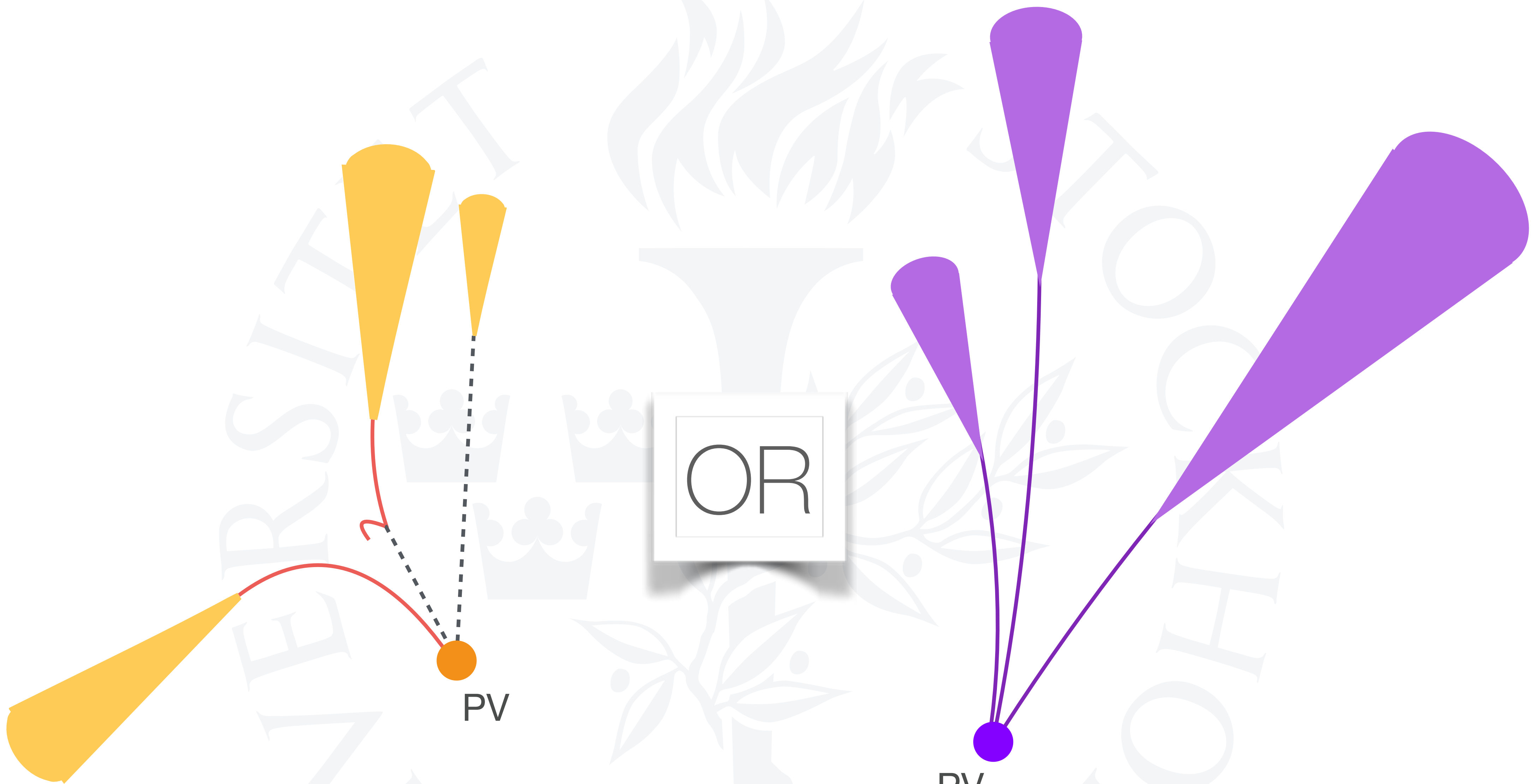


- During the analysis, the filters were studies for inefficiencies.
- If neutralino lifetime is *short* (<100 ps), jets from neutralino decay are largely inside first pixel layer, so jets are very likely to have tracks associated with them: trackless jet **requirement** is inefficient.

Also inefficient for LL R-hadrons

rcarney@lbl.gov

New DV multi-jet filter



Trackless jets

Single or double

High pT jets

No trackless jet requirement

Pre-existing filter had this logic:

# jets	Jet pT cut [GeV]
2	220
3	120
4	100
5	75
6	50
7	45

DV_MultiJetFinalTracklessFilter =

{ tracklessJets &
[(2j & 3j) ^ 4j ^ 5j ^ 6j ^ 7j] &
DVMultiJetTriggerFlags }

2015 + 2016 data

Updated filter for DV+Jets search:

# jets	Jet *High* pT cut [GeV]
2	500
3	180
4	220
5	170
6	100
7	75

{ DV_MultiJetFinalTracklessFilter ^
[(2j_h & 3j_h) ^ 4j_h ^ 5j_h ^ 6j_h^ 7j_h] }

2017 + 2018 data

these cuts increase the rate by no more than 1 Hz

So how much of our signal can we recover with this addition?

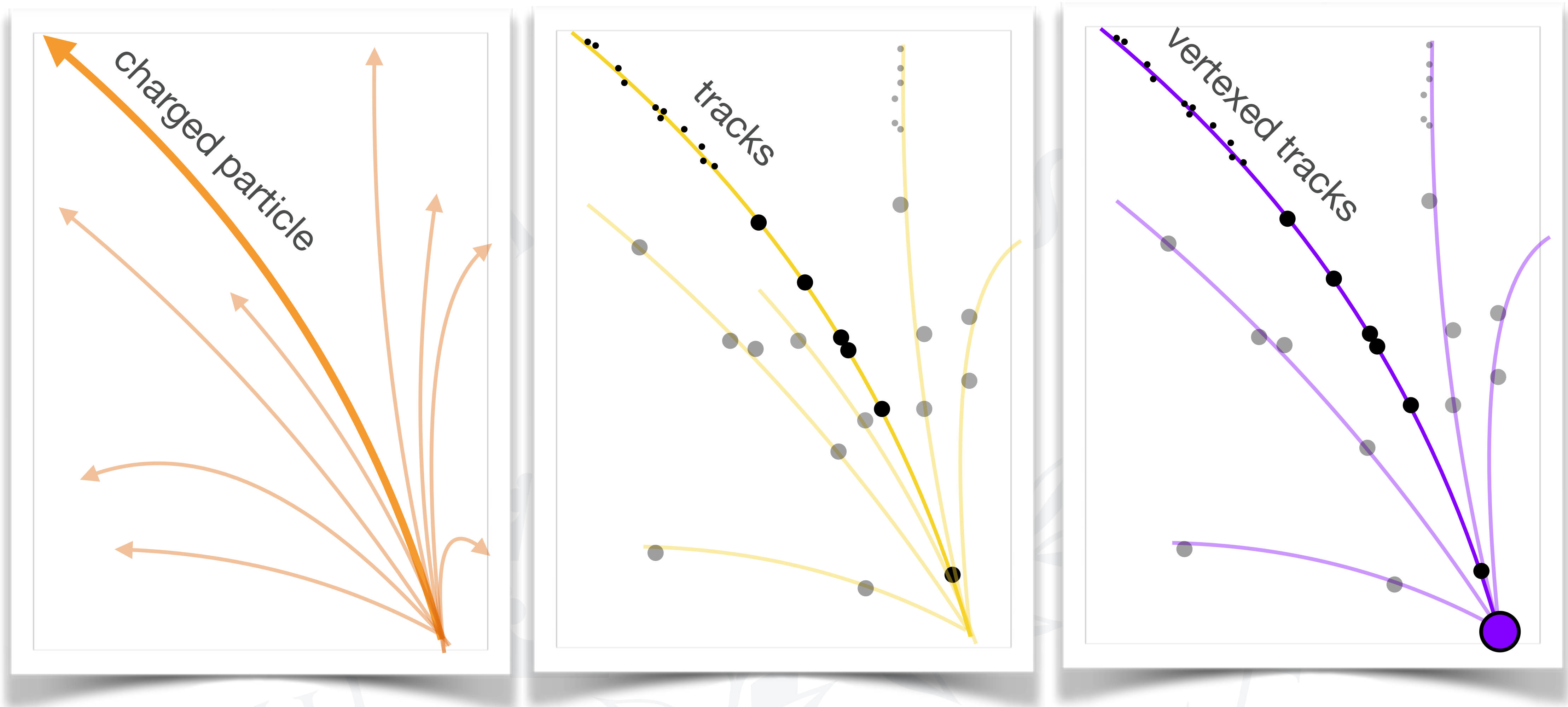
BERKELEY LAB

Fundamentally, these filters seek to minimize unnecessary computations

How has adding the high pT filters changed the event acceptance rate?

rcarney@lbl.gov

DV **summary**



- Aside from studying the efficiency of the DRAW filter, part of my work on this analysis has included studying the efficiency of the DV reconstruction, analysis software framework & sim studies.

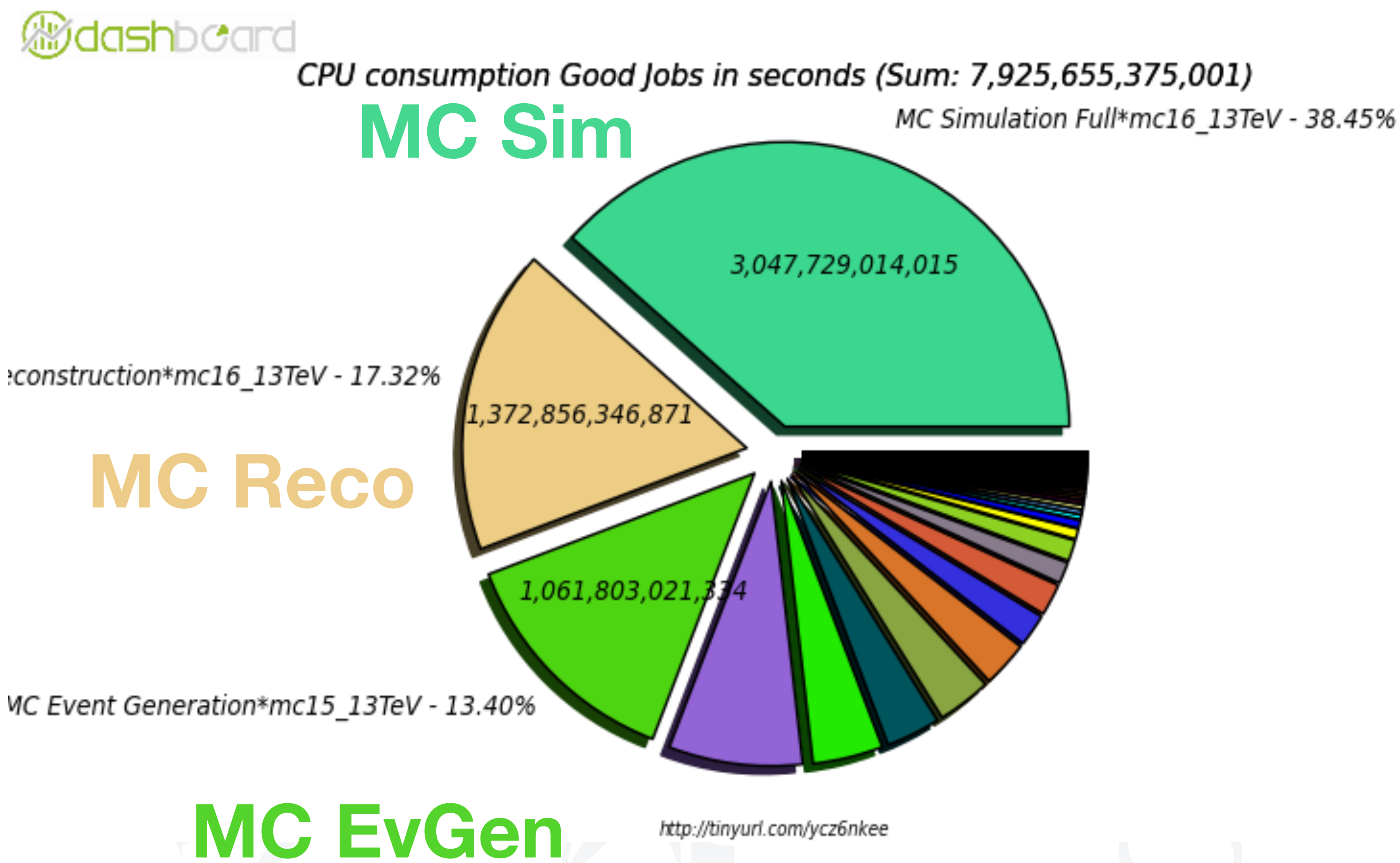
- Searching for unique signatures from rare processes requires special reconstruction: additional processing.

Does it really matter?

Fundamentally, these filters seek to minimize unnecessary computations

As we prepare for HL-LHC, measures like these will become paramount

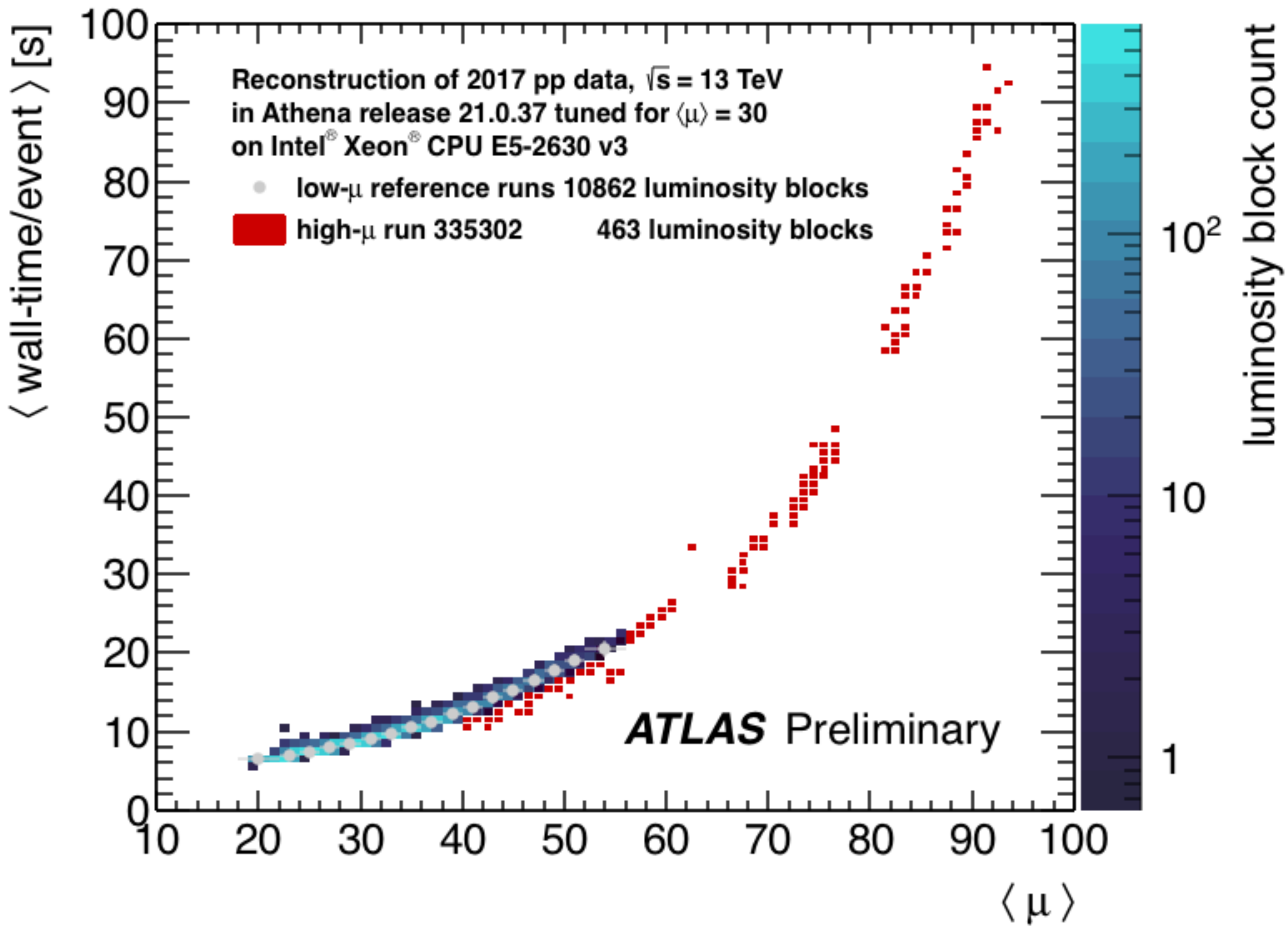
Current **CPU consumption**



Currently, largest CPU consumption from Sim: G4 calorimetry - massive progress in last couple of years to reduce this! However, in HL-LHC a different process will take over…

Future **CPU consumption**

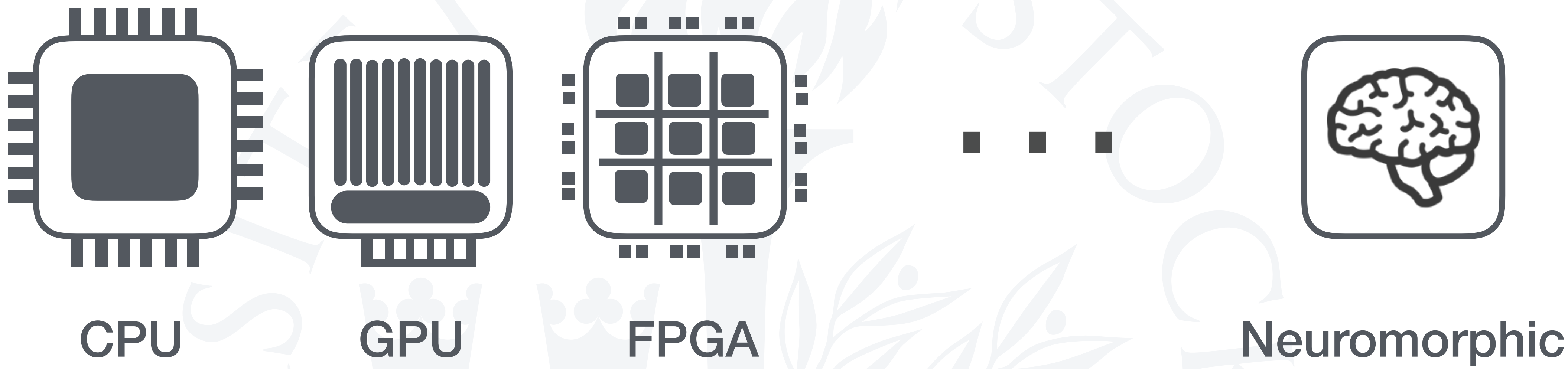
Reconstruction takes over



- During HL-LHC: **dominated by reconstruction**,
- Our computing requirements will outstrip our resources by a factor of 3
- A different approach is needed: supplement the grid with HPC and parallelize.

BERKELEY LAB

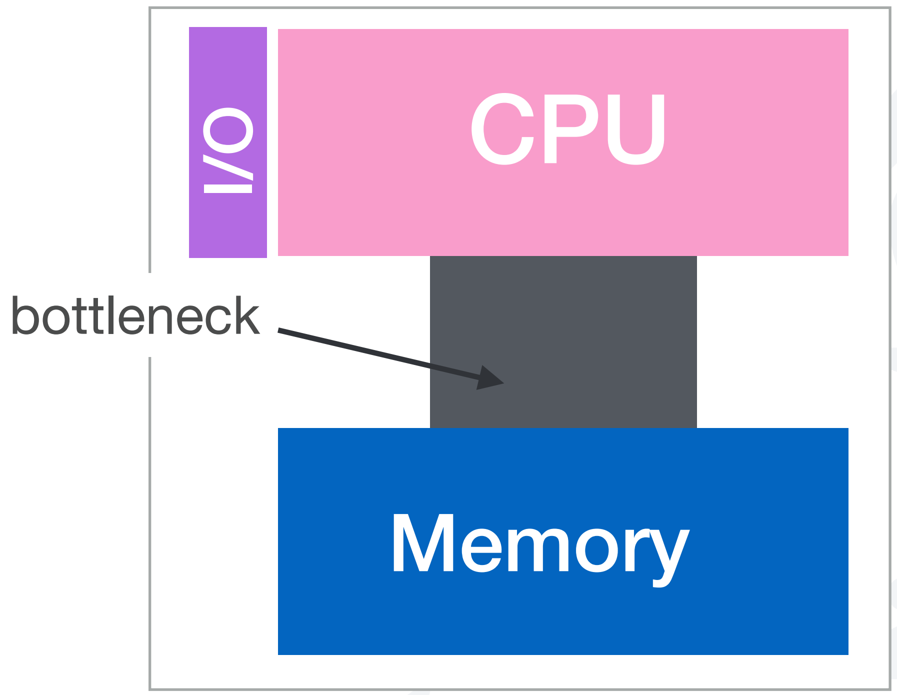
Future **commodity computing**



- HPC is trending towards parallelism.
- Hardware efforts include exploring FPGA’s, GPU’s, multi-core processors, and cutting edge architecture

What about neuromorphic computing?

Neuromorphic computing

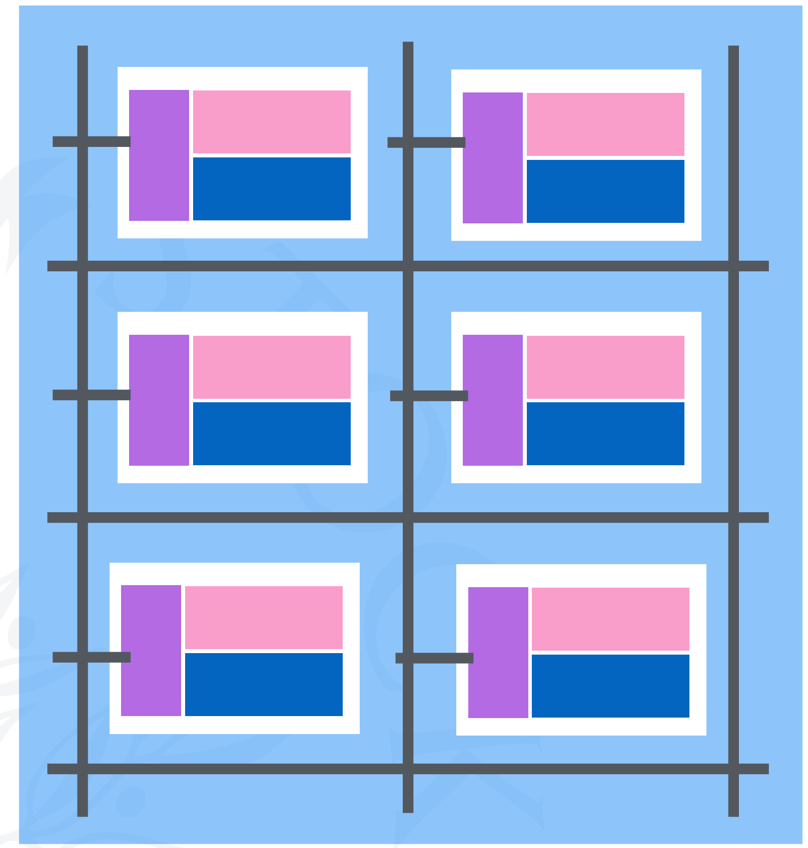


Traditional von Neumann architecture

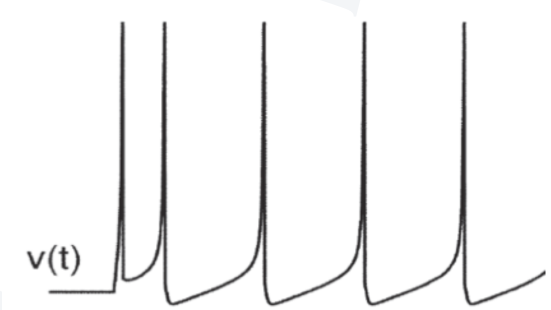
01000011

Binary n-bit words

Instruction set



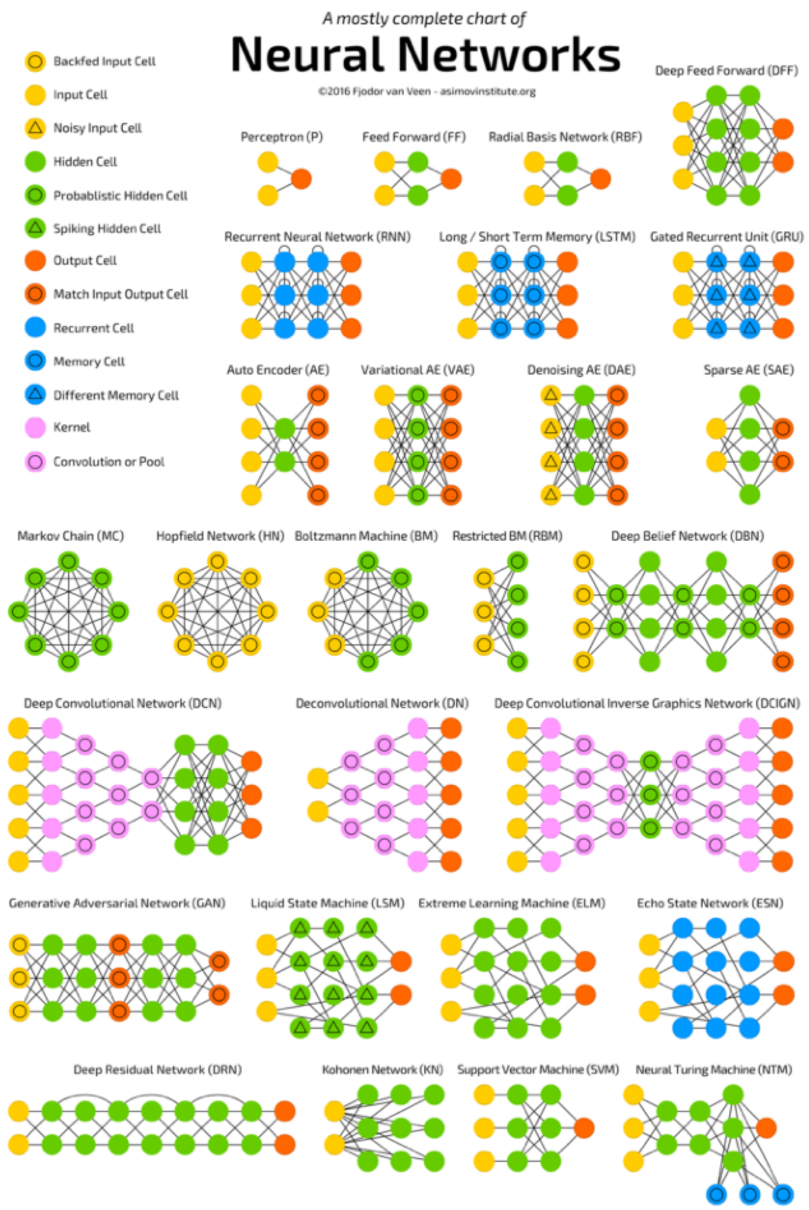
Distributed, parallel architecture

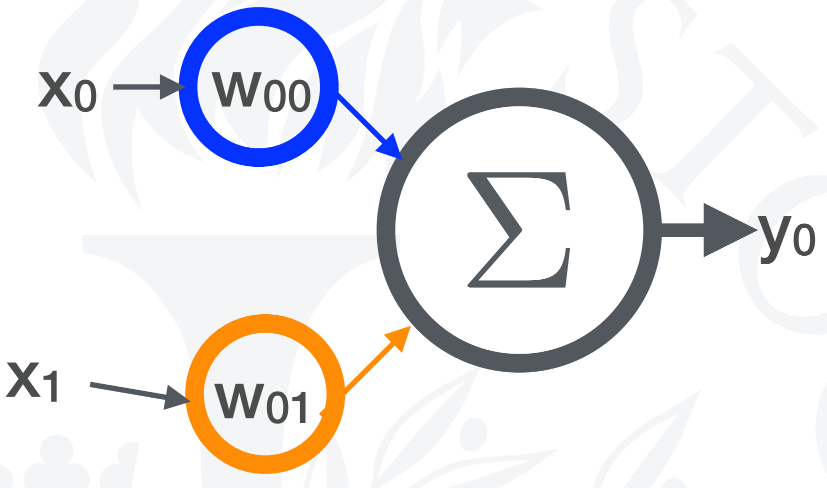


Temporally separated spikes

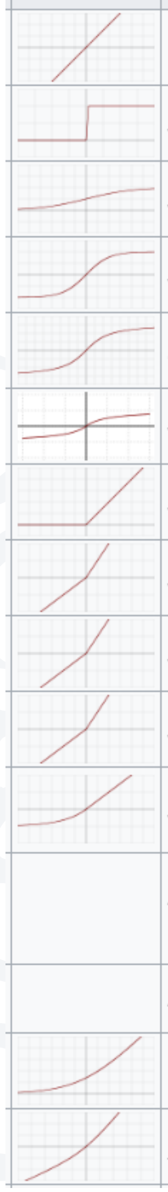
Connectivity and weight model file

Neuromorphic computing





$$y_j = \sum_i w_{ij} x_i$$



Plot	Equation
	$f(x) = x$
	$f(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 & \text{for} \quad x \geq 0 \end{cases}$
	$f(x) = \frac{1}{1+e^{-x}}$
	$f(x) = \tanh(x) = \frac{2}{1+e^{-2x}} - 1$
	$f(x) = \tan^{-1}(x)$
	$f(x) = \frac{x}{1+\lvert x \rvert}$
	$f(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ x & \text{for} \quad x \geq 0 \end{cases}$
	$f(x) = \begin{cases} 0.01x & \text{for} \quad x < 0 \\ x & \text{for} \quad x \geq 0 \end{cases}$
	$f(\alpha, x) = \begin{cases} \alpha x & \text{for} \quad x < 0 \\ x & \text{for} \quad x \geq 0 \end{cases}$
	$f(\alpha, x) = \begin{cases} \alpha x & \text{for} \quad x < 0 \\ x & \text{for} \quad x \geq 0 \end{cases}$ [1]
	$f(\alpha, x) = \begin{cases} \alpha(e^x - 1) & \text{for} \quad x < 0 \\ x & \text{for} \quad x \geq 0 \end{cases}$
	$f_{t_l,a_l,t_r,a_r}(x) = \begin{cases} t_l + a_l(x - t_l) & \text{for} \quad x \leq t_l \\ x & \text{for} \quad t_l < x < t_r \\ t_r + a_r(x - t_r) & \text{for} \quad x \geq t_r \end{cases}$ t_l, a_l, t_r, a_r are parameters.
	$f(x) = \max(0, x) + \sum_{s=1}^{S} a_i^s \max(0, -x + b_i^s)$
	$f(x) = \ln(1 + e^x)$
	$f(x) = \frac{\sqrt{x^2+1}-1}{2} + x$

Topology

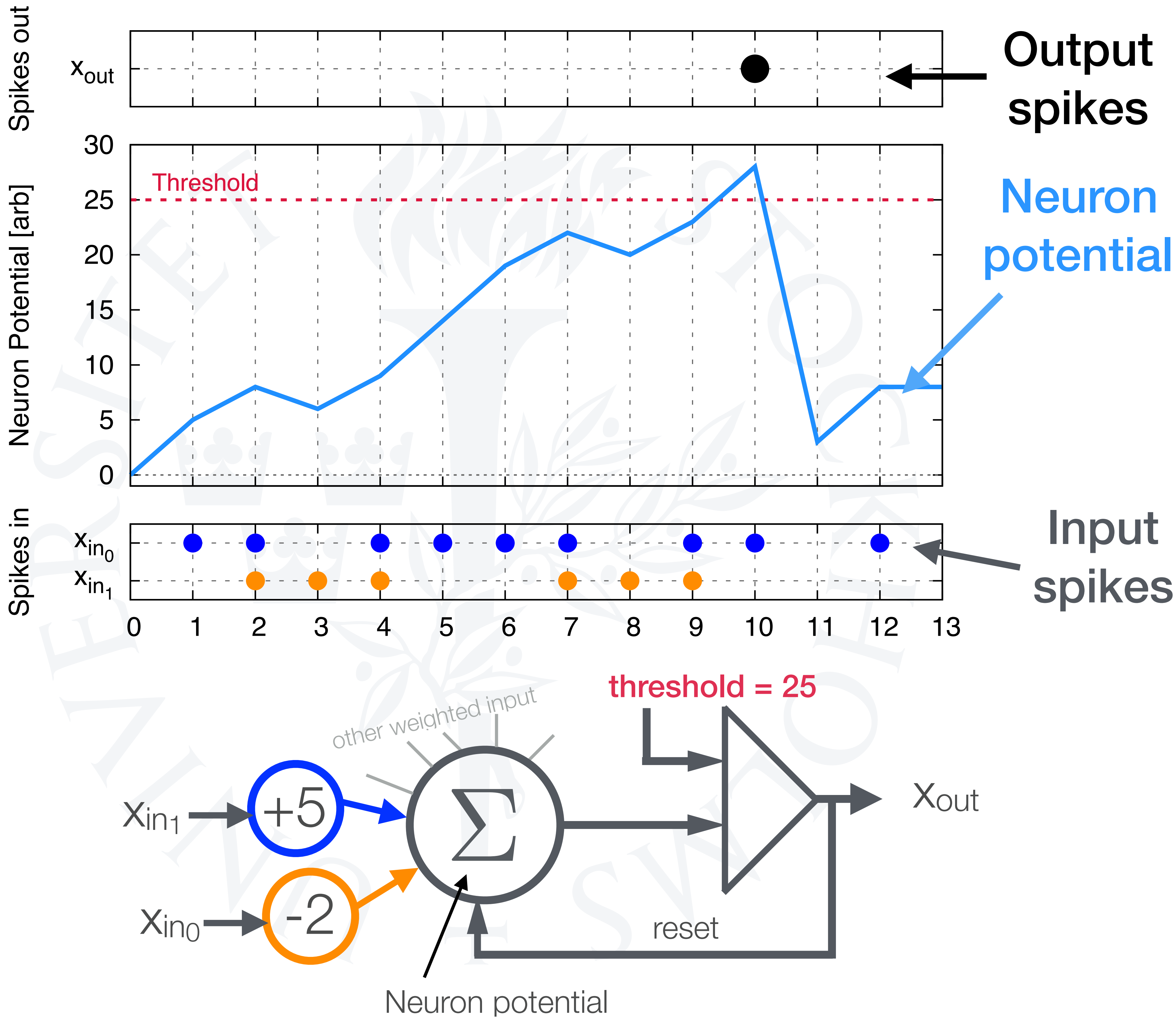
Sum of weighted inputs

Activation function

ANN's and neuromorphic networks have much in common.

BERKELEY LAB

Neuron **computation**

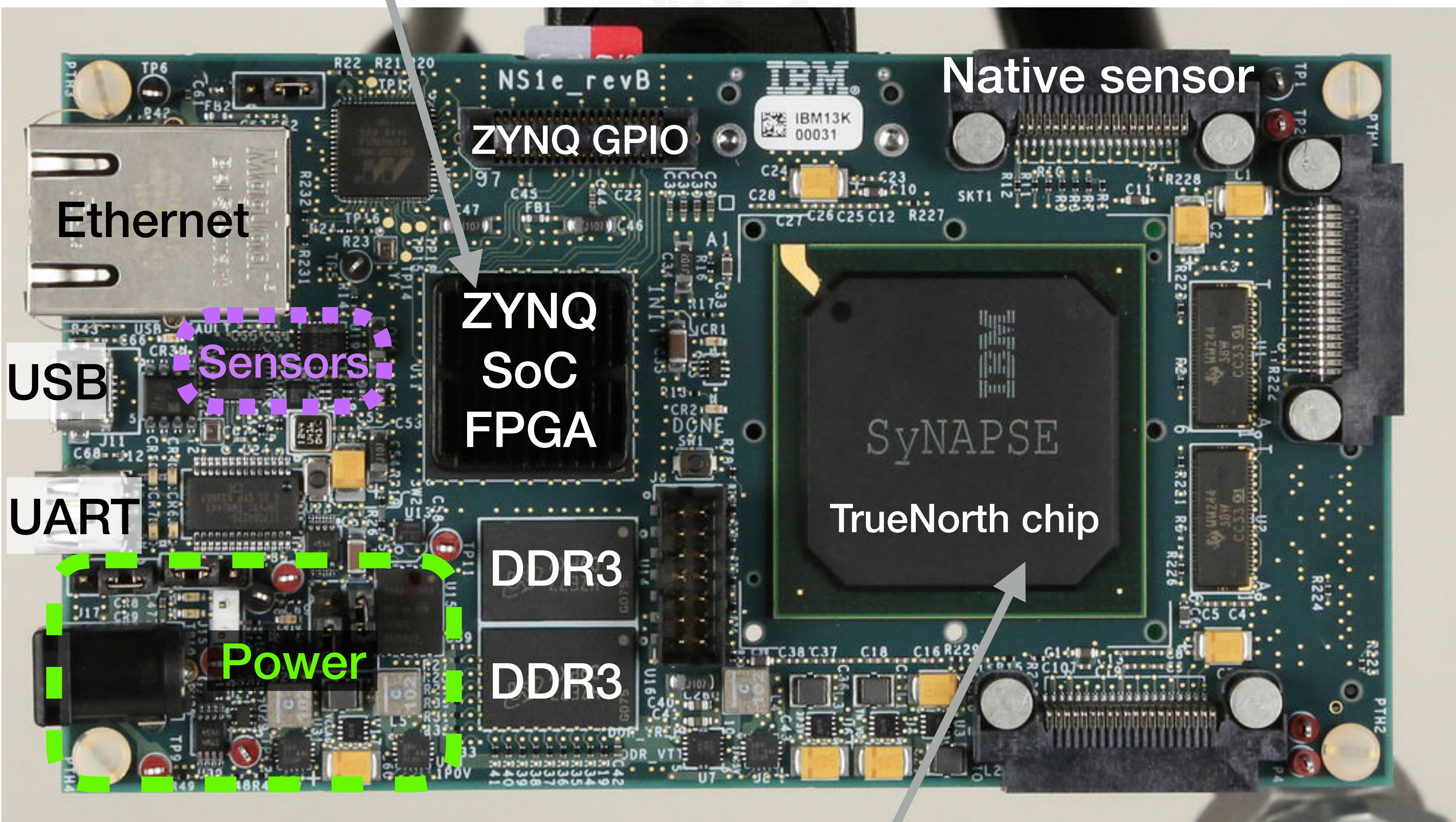


Time & encoding are fundamental concepts in neuromorphic computing

rcarney@lbl.gov

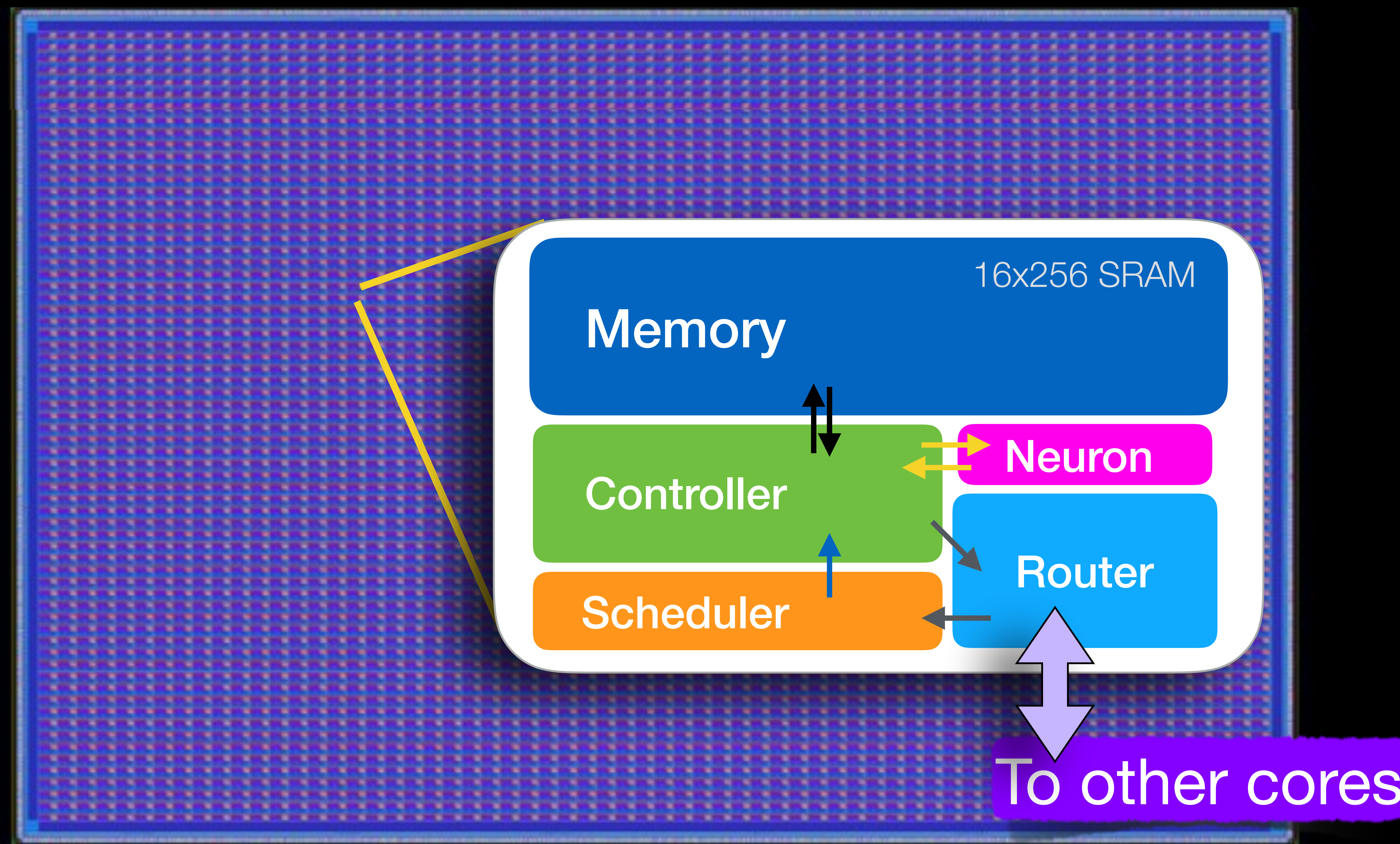
TrueNorth testboard

spike transduction



2015 IBM released TrueNorth test boards to US national labs

TrueNorth architecture



64x64 cores

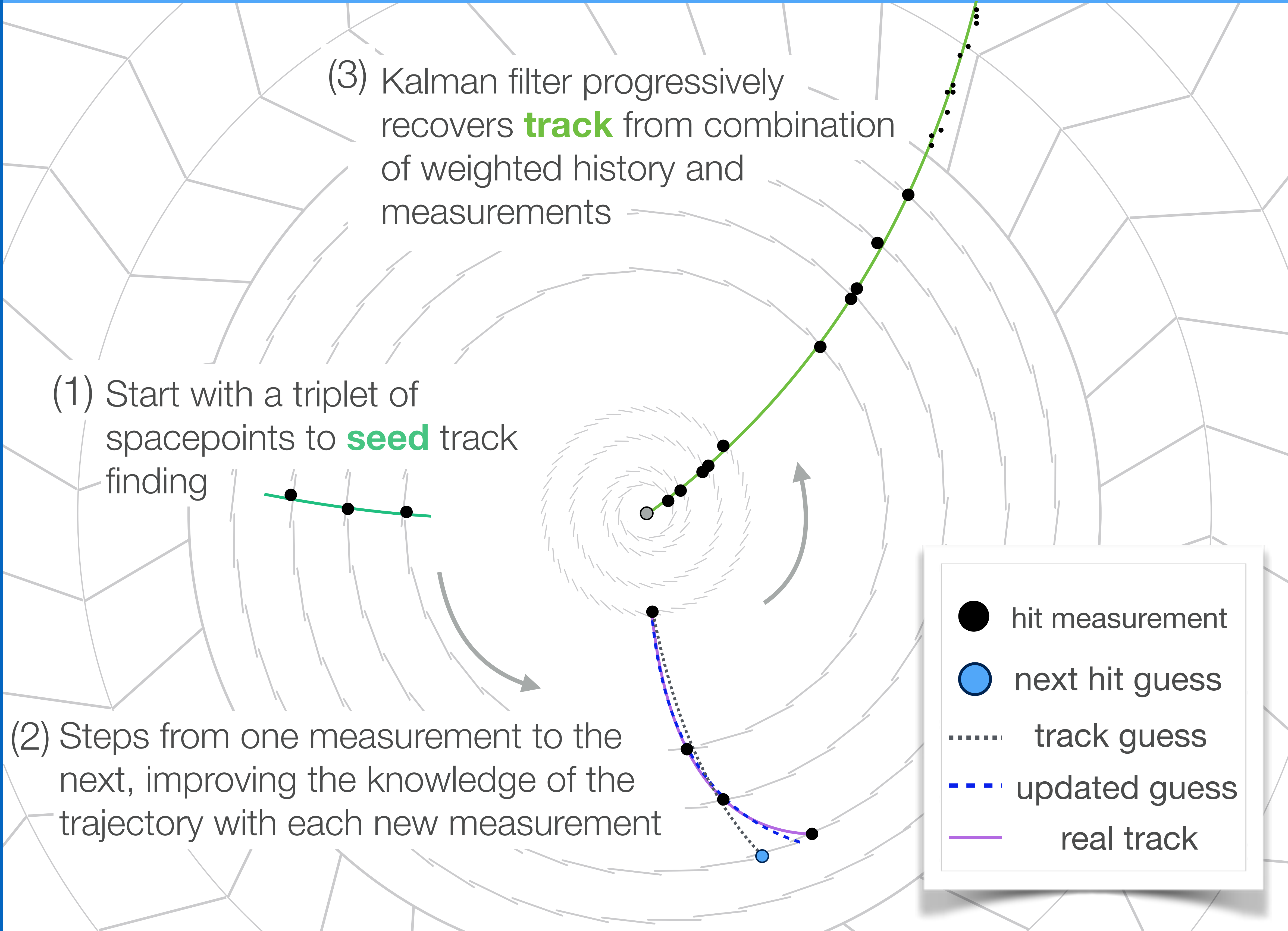
Mixed synchronous-asynchronous logic, global 'tick' is synchronization point (1 kHz), 10's-100's of mW

To understand how these design features affect usability of neuromorphic chip

Implement a tracking algorithm in TrueNorth

rcarney@lbl.gov

Kalman filter tracking



(1) Start with a triplet of spacepoints to **seed** track finding

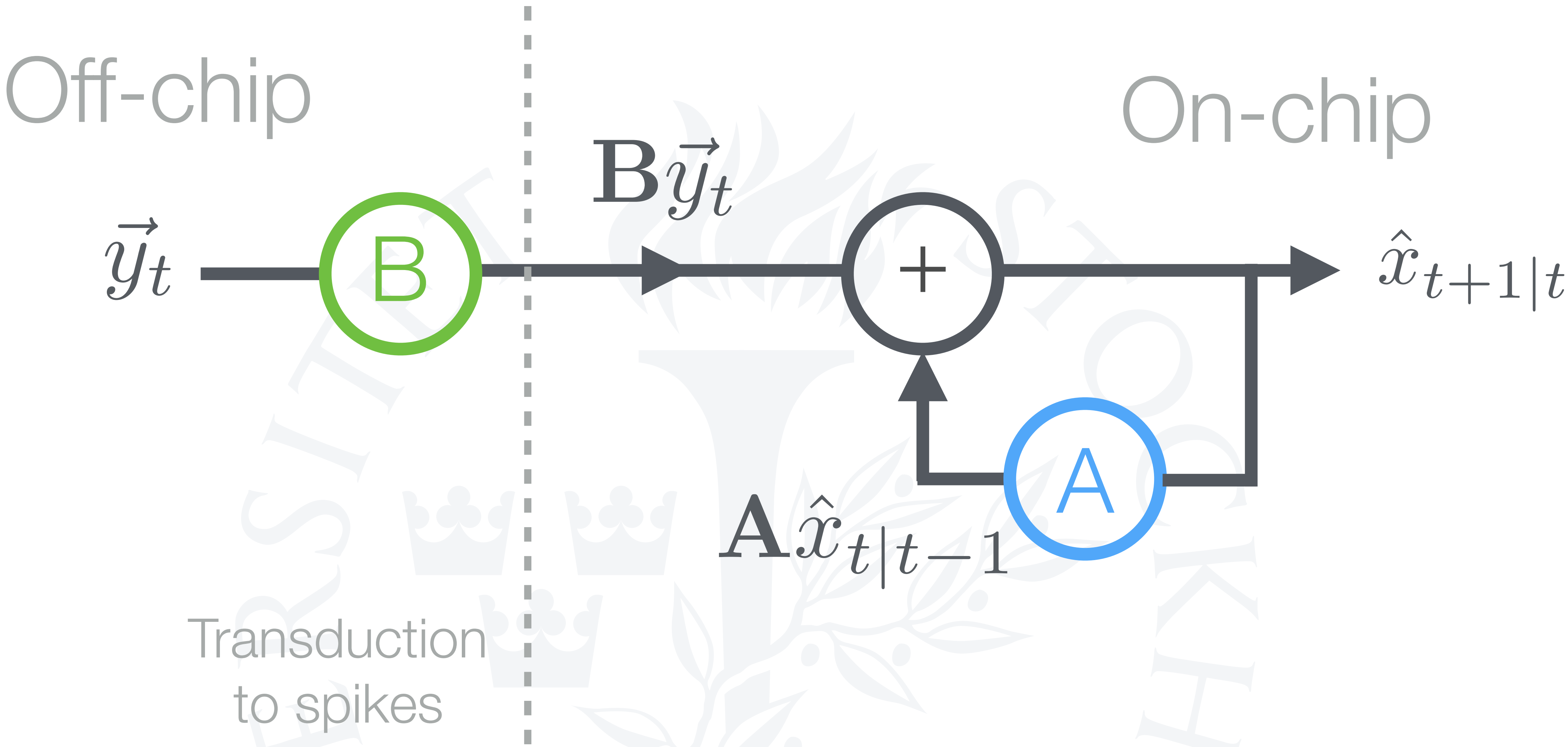
(2) Steps from one measurement to the next, improving the knowledge of the trajectory with each new measurement

(3) Kalman filter progressively recovers **track** from combination of weighted history and measurements

Kalman filter recovers underlying state from noisy measurements

Berkeley Lab

Steady-state Kalman filter



Discrete, steady state Kalman filter update

$$\hat{x}_{t+1|t} = (\mathbf{F} - \mathbf{L})\hat{x}_{t|t-1} + \mathbf{L}\vec{y}_t$$
$$= \mathbf{A}\hat{x}_{t|t-1} + \mathbf{B}\vec{y}_t$$

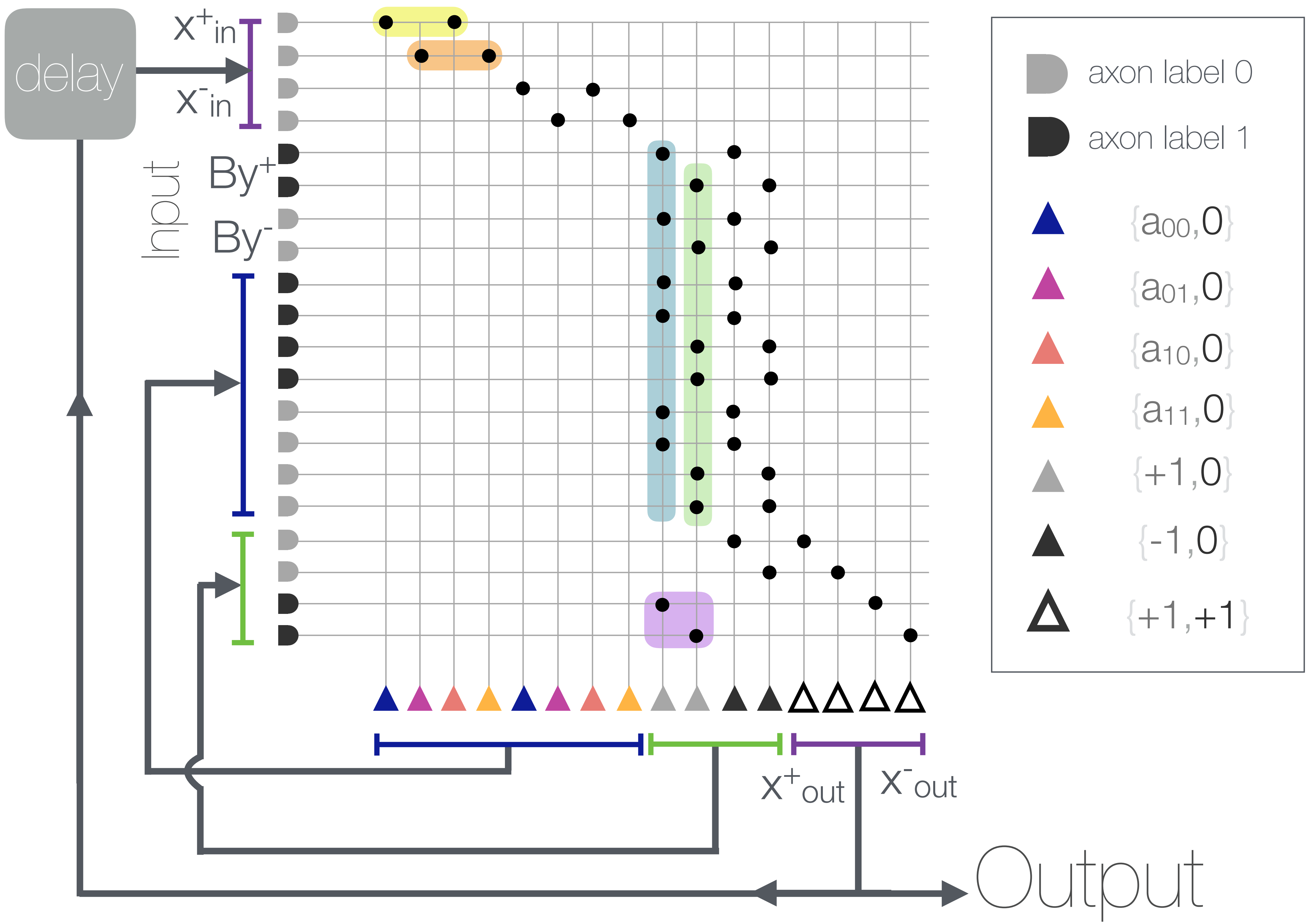
y = state measurement
x = state prediction
F = transition matrix
L = Kalman gain
w = process noise
v = measurement noise

A state estimate and a noisy measurement

$$\vec{x}_{t+1} = \mathbf{F}\vec{x}_t + \vec{w}_t \quad , \quad \vec{y}_t = \vec{x}_t + \vec{v}_t$$

Kalman filter **in TrueNorth**

Example of implementation

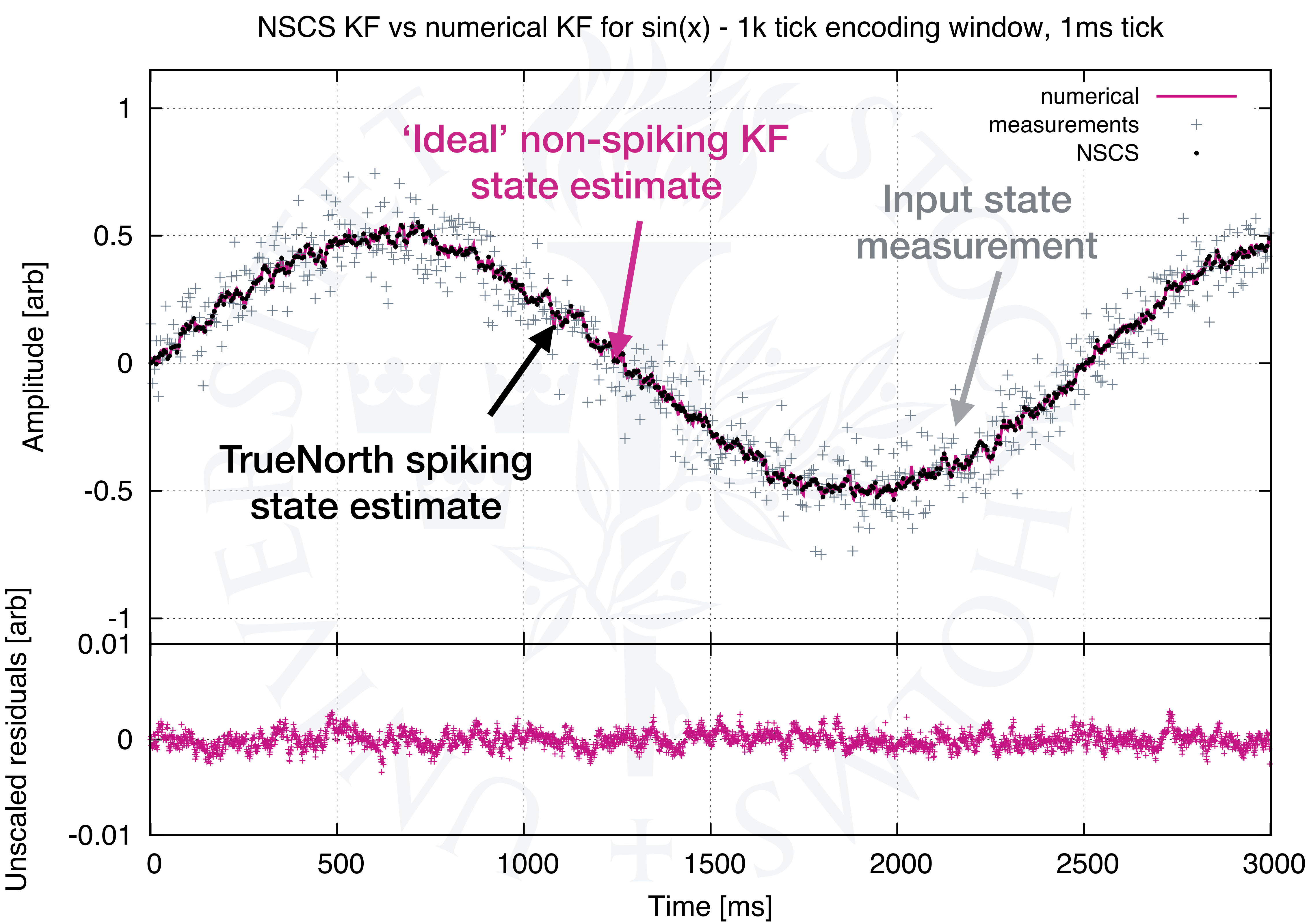


Does it work?

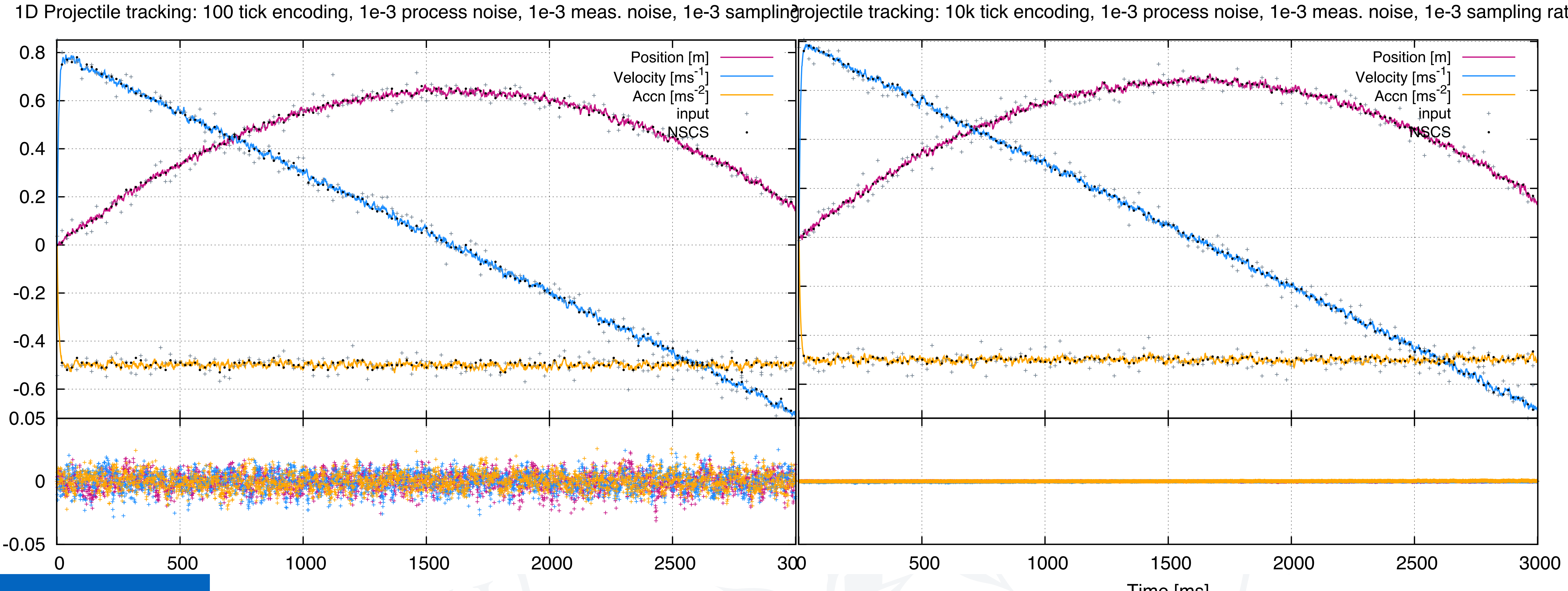
(rate encoded discrete steady-state, 2-dim KF)

Kalman filter in TrueNorth: **results**

Tracking a noisy sine wave



TrueNorth challenges: encoding precision



100 tick encoding window

10,000 tick encoding window

Encoding (i.e. input precision) massively affects performance!

Spiking data **encoding**

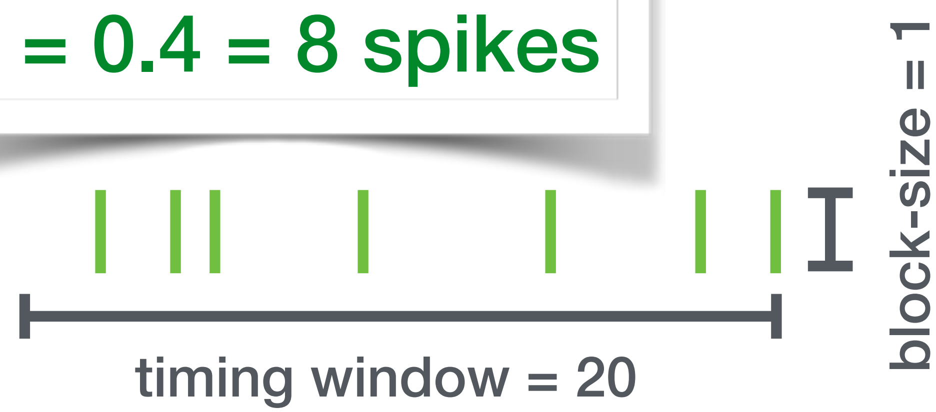
In binary: 100 != 010 != 001

In unary: 100 == 010 == 001

Rate encoding

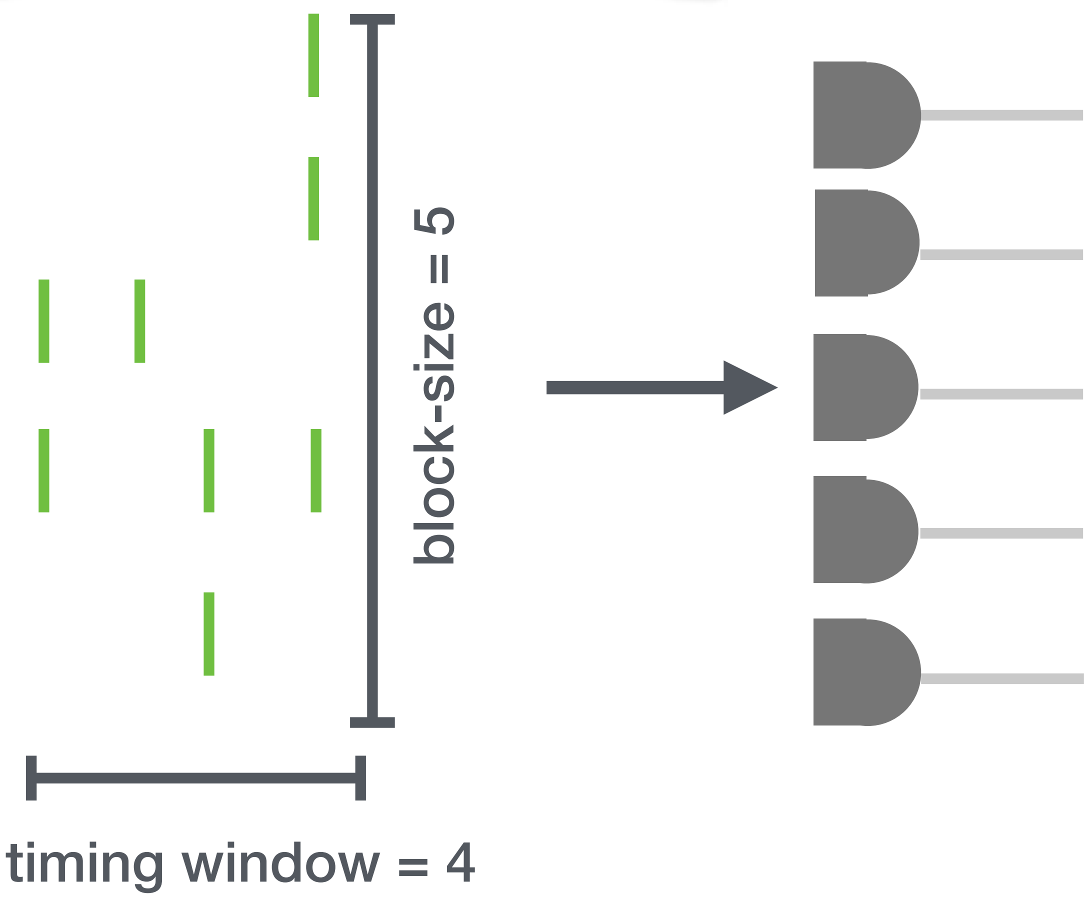
x = 1.0 = 20 spikes

x = 0.4 = 8 spikes



Spatial & rate encoding

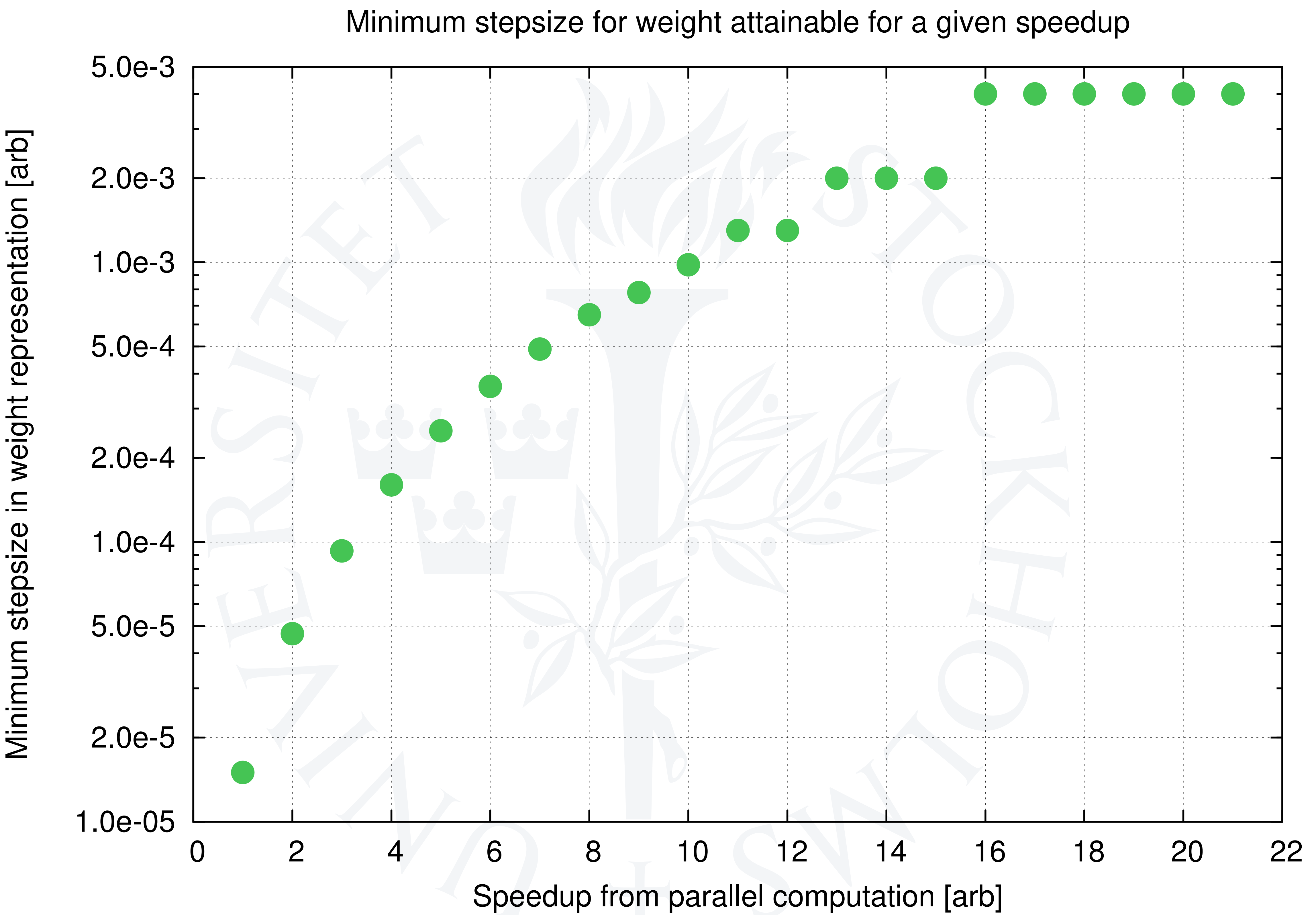
x = 0.4 = 8 spikes



But if axons used for parallel inputs, can't be used to compensate for other things…

TrueNorth challenges

Limits on **core size**

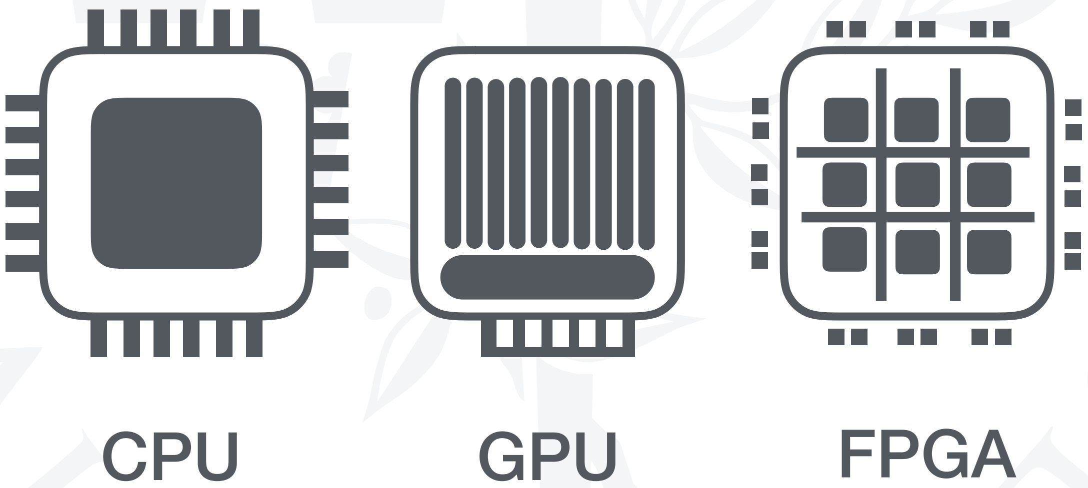


A compromise must be struck between accuracy and speedup: necessary because of the chips design and the way information is encoded

TrueNorth Summary

Key points

- TrueNorth compromises flexibility and neuron complexity for low power.
- **Slow clock**: 2 kHz max clock: cannot be sped up
- **Small cores**: Parallelism of multiplications per core limited to 21x.
- **Complicated encoding**: spike encoding/decoding + tick must be done off-chip..
- **Fixed functionality at runtime**: cannot implement either real-time weight updates or back-prop.



Could neuromorphic hardware feature in our toolkit in 10 years time? **It seems more likely we would use a hardware ANN in a GPU or FPGA.**

Conclusions

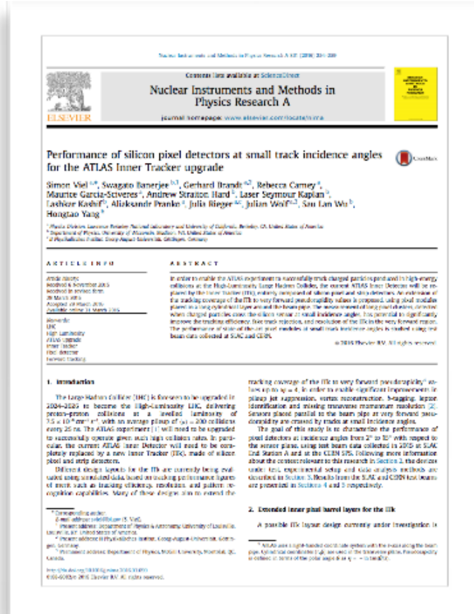
- Expanding our tracking capabilities open us up to interesting BSM physics signatures such as those seen with long-lived particles.
- High multiplicity events lead to non-trivial event reconstruction, particularly track reconstruction.
- Understanding how to minimize cpu usage with both algorithms and hardware is necessary to compensate for our burgeoning computing requirements.

Beyond PhD…

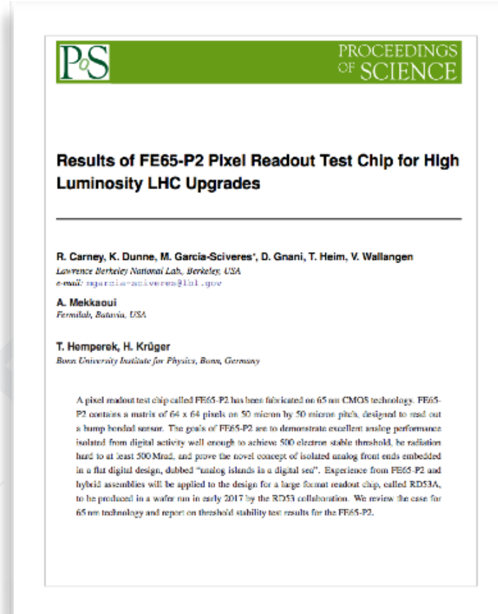
- During my postdoc I'd like to continue to tackle interesting problems that require a strong understanding of the detector.
- I want to do this by taking a leading role in ITk Pixel production, and by applying the skills and insight from DV+Jets to new challenges in data analysis!

Papers published during PhD

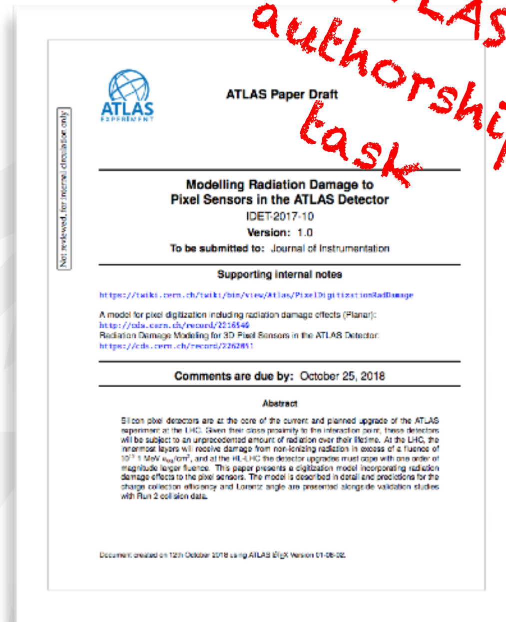
Papers and **contributions**


Nuclear Instruments and Methods in Physics Research A

Performance of silicon pixel detectors at small track incidence angles for the ATLAS Inner Tracker upgrade

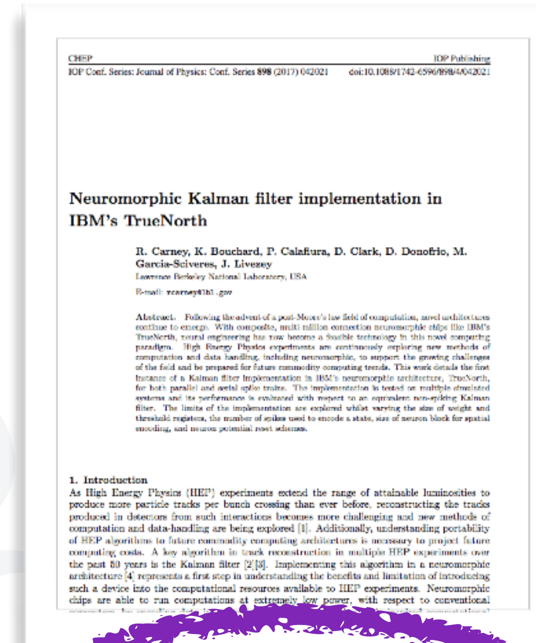

PROCEEDINGS OF SCIENCE

Results of FE65-P2 Pixel Readout Test Chip for High Luminosity LHC Upgrades


ATLAS Paper Draft

Modelling Radiation Damage to Pixel Sensors in the ATLAS Detector

ATLAS authorship task


Neuromorphic Kalman filter implementation in IBM's TrueNorth

Performance of Silicon Pixel Detectors at Small Track Incidence Angles for ATLAS ITk	Results of FE65-P2 Pixel Readout Test Chip for High Luminosity LHC Upgrades	Modelling Radiation Damage to Pixel Sensors in the ATLAS Detector	Neuromorphic Kalman filter implementation in IBM's TrueNorth
Published NIM A 11/15	Published PoS ICHEP 11/16	ATLAS Phase 2 approval	Published J. Phys Conf. 09/17
https://doi.org/10.1016/j.nima.2016.03.099	https://doi.org/10.22323/1.282.0272	https://glance.cern.ch/atlas/analysis/papers/details.php?id=10403	https://doi.org/10.1088/1742-6596/898/4/042021
Prepared modules, ran testbeam, wrote analysis framework	**Digital verification of FE, prepared & ran chips during testbeam and irradiation**	**Wrote new digitizer for rel. 21 and rel. 22, restructured PixDig pkg for rel. 22**	**Designed and implemented KF in TN, ran studies, wrote paper**

And search for new physics using Run 2 data currently in the works…

rcarney@lbl.gov

Backup

DV+JETS

Long-lived particle mechanisms

More detail: **Standard Model**

small coupling

$$\Gamma \sim y^2 \left(\frac{m}{M}\right)^n m$$

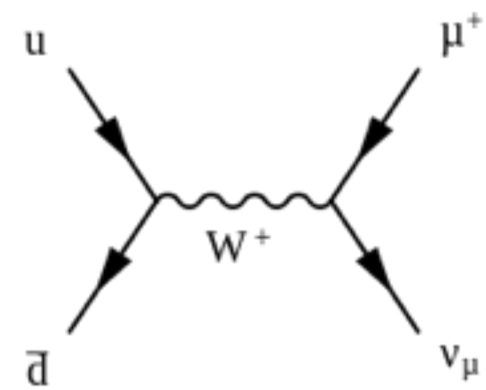
Set by symmetry structure, typically n ≥ 4

$m \ll M$

hierarchy of scales

- Heavy mediator
  (e.g. charged pion decay)

$$\pi^+ \to \mu^+ \nu_\mu$$



$$\Gamma \sim g_2^2 \left(\frac{m}{m_W}\right)^4 m$$

$$\Gamma \sim g_2^2 \left(\frac{m_n - m_p}{m_W}\right)^4 (m_n - m_p)$$

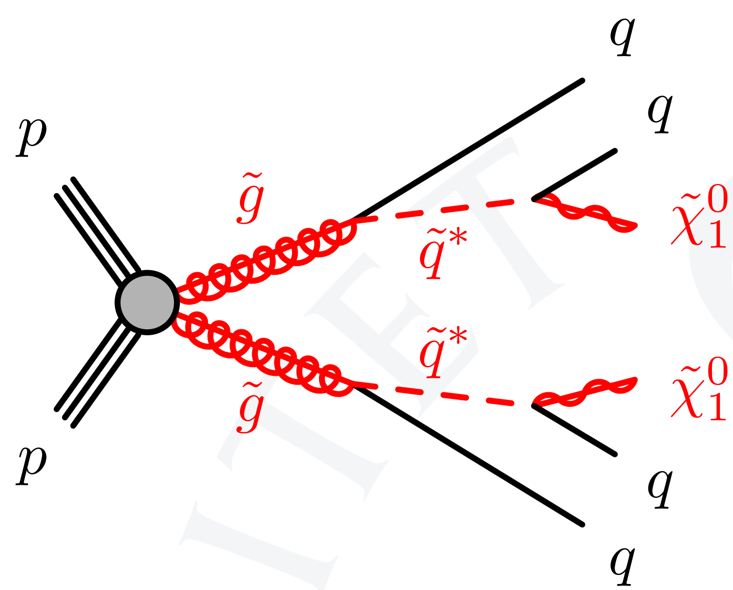
$$n^0 \to p^+ e^- \nu_e$$

- Nearly degenerate masses (e.g. free neutron decay)

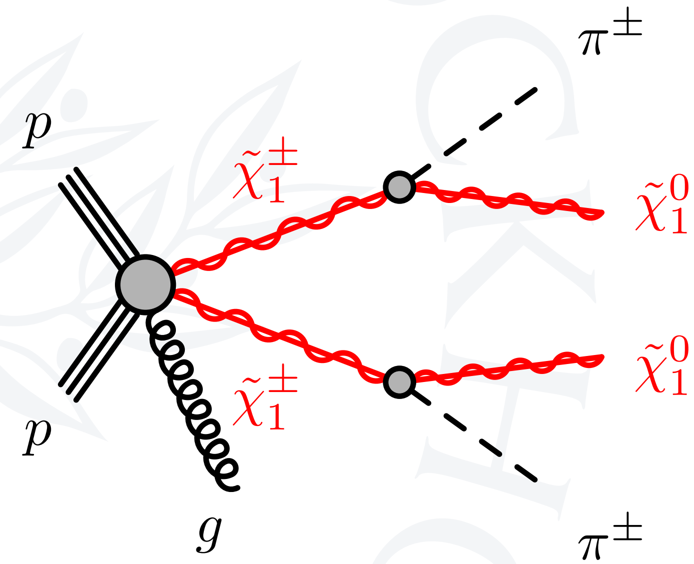
- Small coupling
  (e.g B-hadron decay)

$$\bar{b} \to cW^- \to cl\bar{l}$$

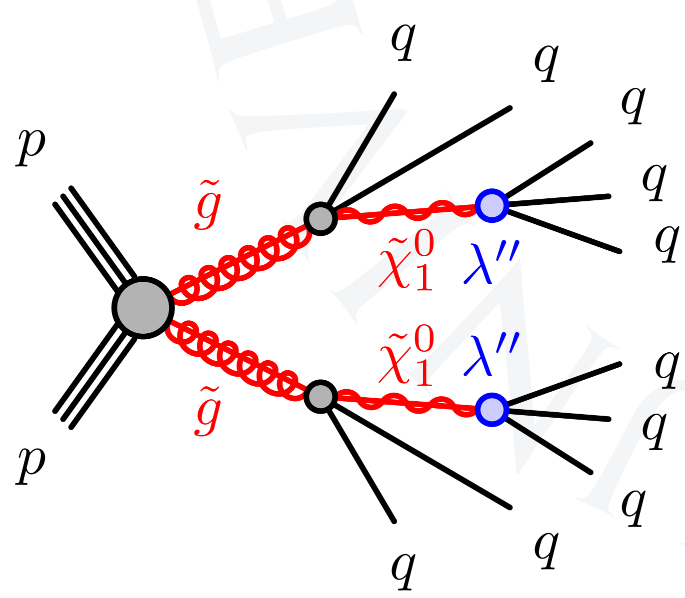
Long-lived particle mechanisms



Large mass difference



Degenerate mass



Small coupling

- R-parity is a discrete symmetry, defined as:

$$R_p = (-1)^{3B+L+2S}$$

- Standard Model particles have Rp = 1, and SUSY particles Rp = -1
- R-parity violating Lagrangian can be simplified as:

$$W_{\not{R}_p} = \underbrace{\lambda_{ijk} L_i L_j \bar{E}_k + \lambda'_{ijk} L_i Q_j \bar{D}_k}_{\text{lepton violating terms}} + \underbrace{\lambda''_{ijk} \bar{U}_i \bar{D}_j \bar{D}_k}_{\text{baryon violating terms}}.$$

rcarney@lbl.gov

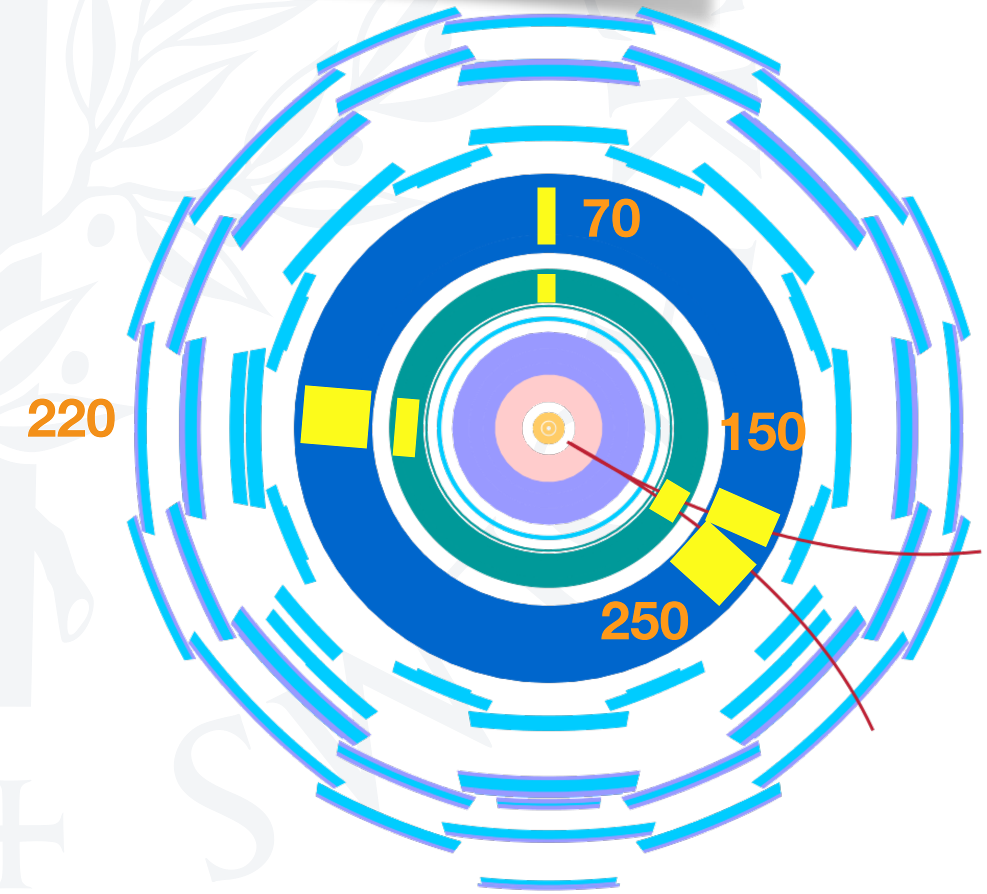
DRAW filter DV Multijet

Second part is jet multiplicity flags:

DV_MultiJetFinalTracklessFilter =
{ tracklessJets &
[(2j & 3j) ^ 4j ^ 5j ^ 6j ^ 7j] &
DVMultiJetTriggerFlags }

2015-2016 DV Multijet DRAW filter

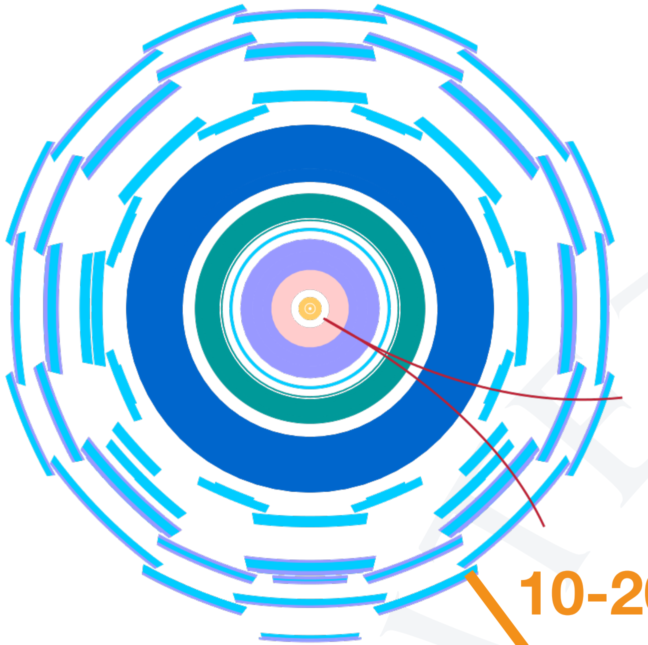
# jets	Jet pT cut [GeV]	Low. unpre. trig. [GeV]
2	220	-
3	120	175
4	100	85
5	75	55
6	50	45
7	45	45



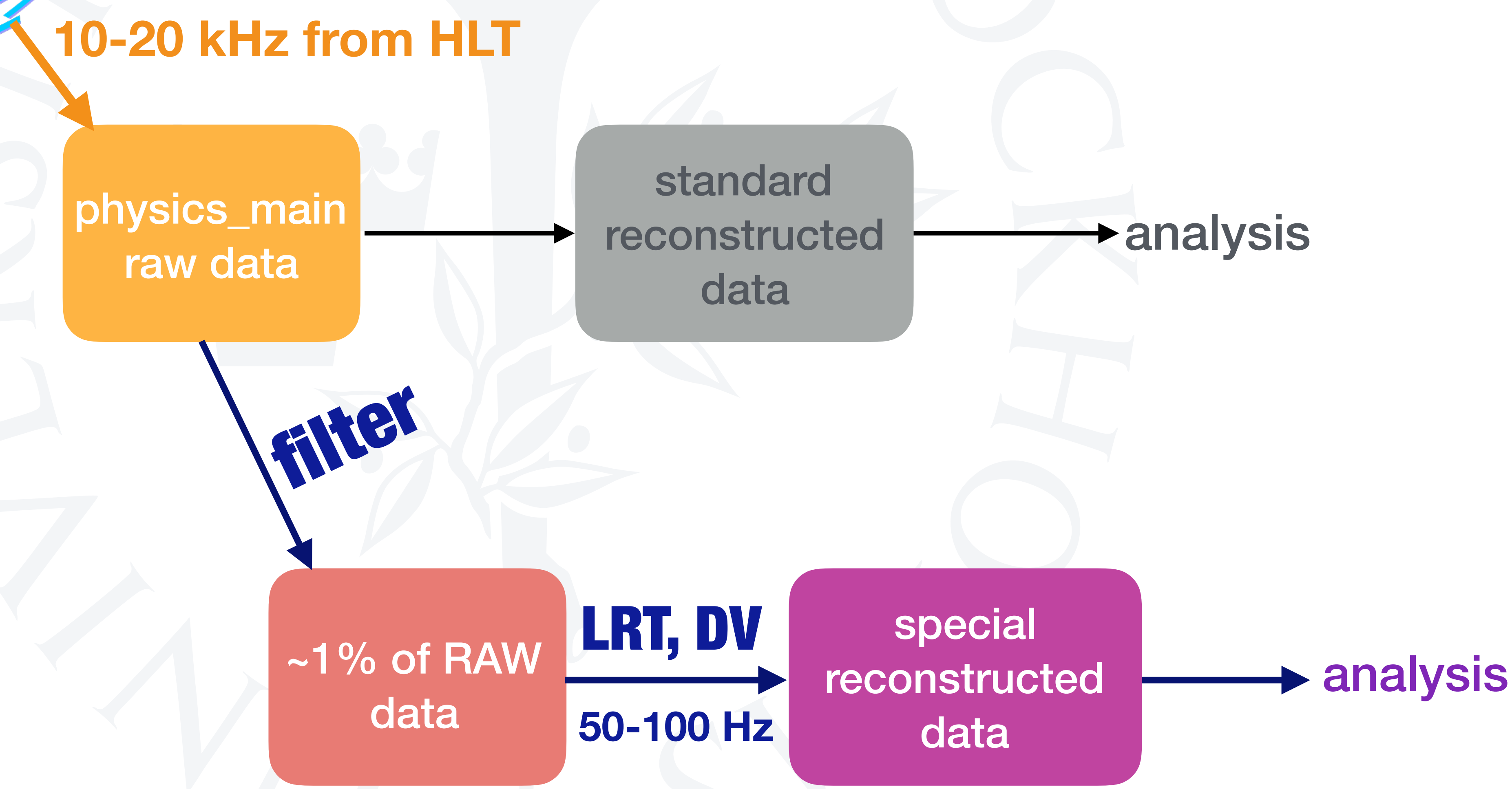
e.g. this scenario would pass the (2j & 3j) flag but not the 4j flag

Data flow

DRAW filter before special reco



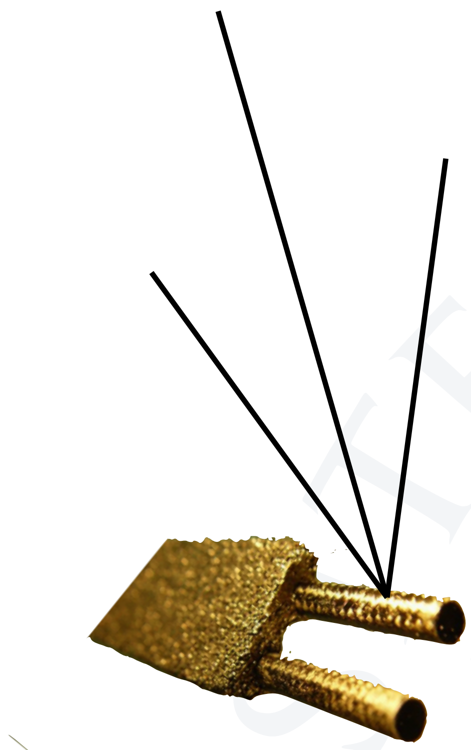
Must access RAW data to run Large Radius Tracking (LRT) and Displaced Vertex (DV) reconstruction algorithms.
Expensive & time consuming…



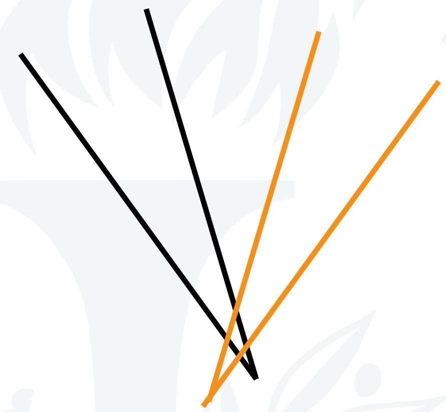
RPVLL_filter ~ 60% of bandwidth of similar RAW filters

rcarney@lbl.gov

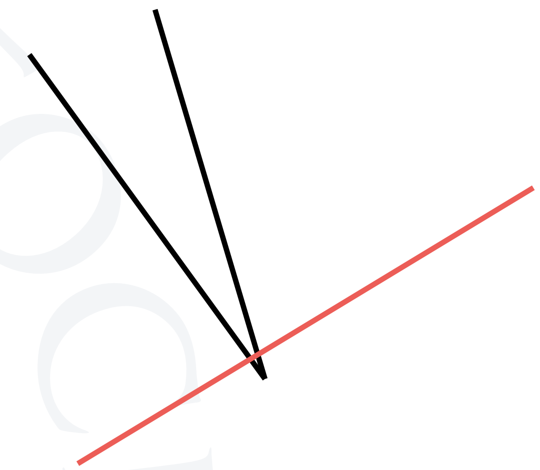
Backgrounds



Hadronic interactions



Merged low-mass vertices



Accidental crossings

Understanding backgrounds involves understanding how the vertices themselves are formed…

DV+Jets my contribution

Other DV analyses from Run 1 paper either published or pushing ahead:

DV + jets	**DV + MET**	**DV + muon**	**DV + dilepton**
Run 2 effort in progress…	Run 2 preprint	Run 2: EB formed	Run 2: glance phase 1 (SUSY-2017-04)

DV + Jets analysis:

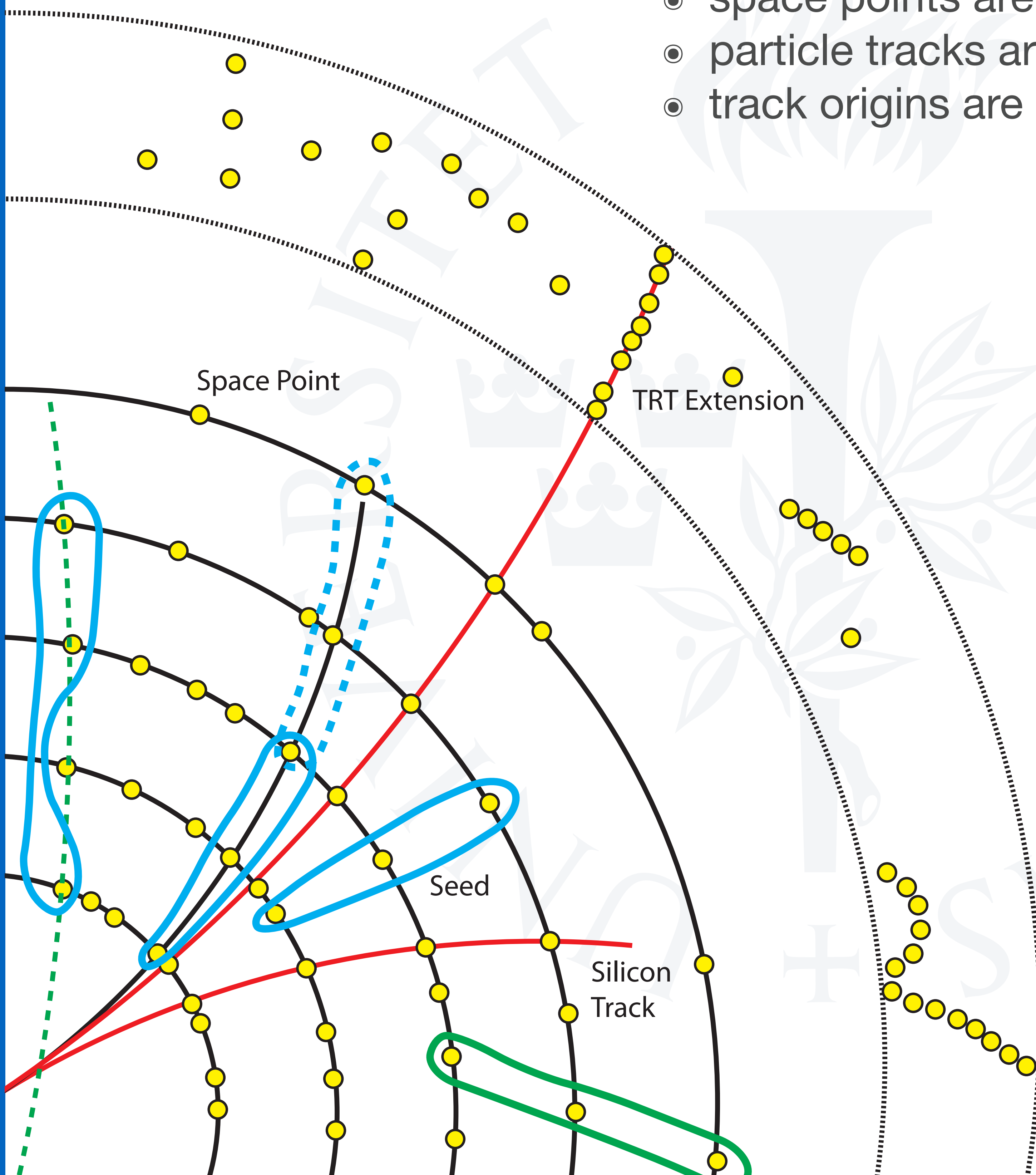
- I am one of 3 students on this analysis
- Developed analysis framework, signal selections
- Will now work on 1 of the 2 major backgrounds: accidental crossings
- Intend to write significant portion of note/paper
- Changes to DRAW, some of which discussed here

Backup: **Merged** vertices

- In ATLAS reconstruction software, two vertices are merged if their significance is smaller than a given *displacement significance.*
- Displacement significance is effectively the length of the displacement between the two vertices, scaled by projecting the uncertainties in the basis vectors (x,y,z) onto the distance vector between the 2 vertices.
- How often does it occur that two vertices are merged by default and promoted to a higher mass?

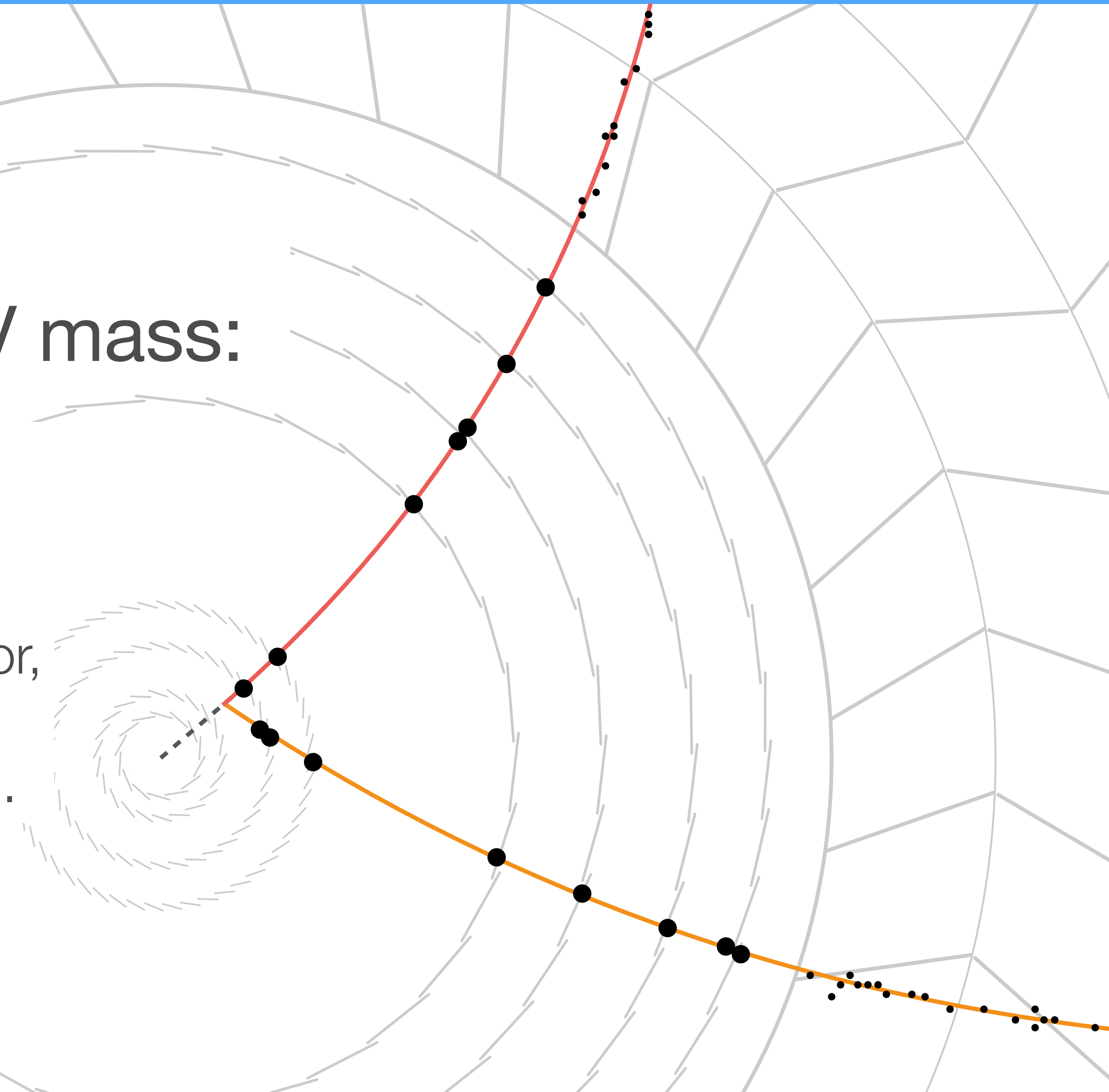
Tracking and reconstruction

- space points are created from charge clusters
- particle tracks are created from space-points
- track origins are clustered into a 'vertex'



For successful reconstruction of DV mass:

- Must reconstruct tracks
- Must correctly reconstruct track parameters with fit
- Must use tracks to form DV or, if missed in forming DV, re-attach them after it is formed.



Recover the tracks

Fit the tracks

(Re-)attach tracks to reco DV

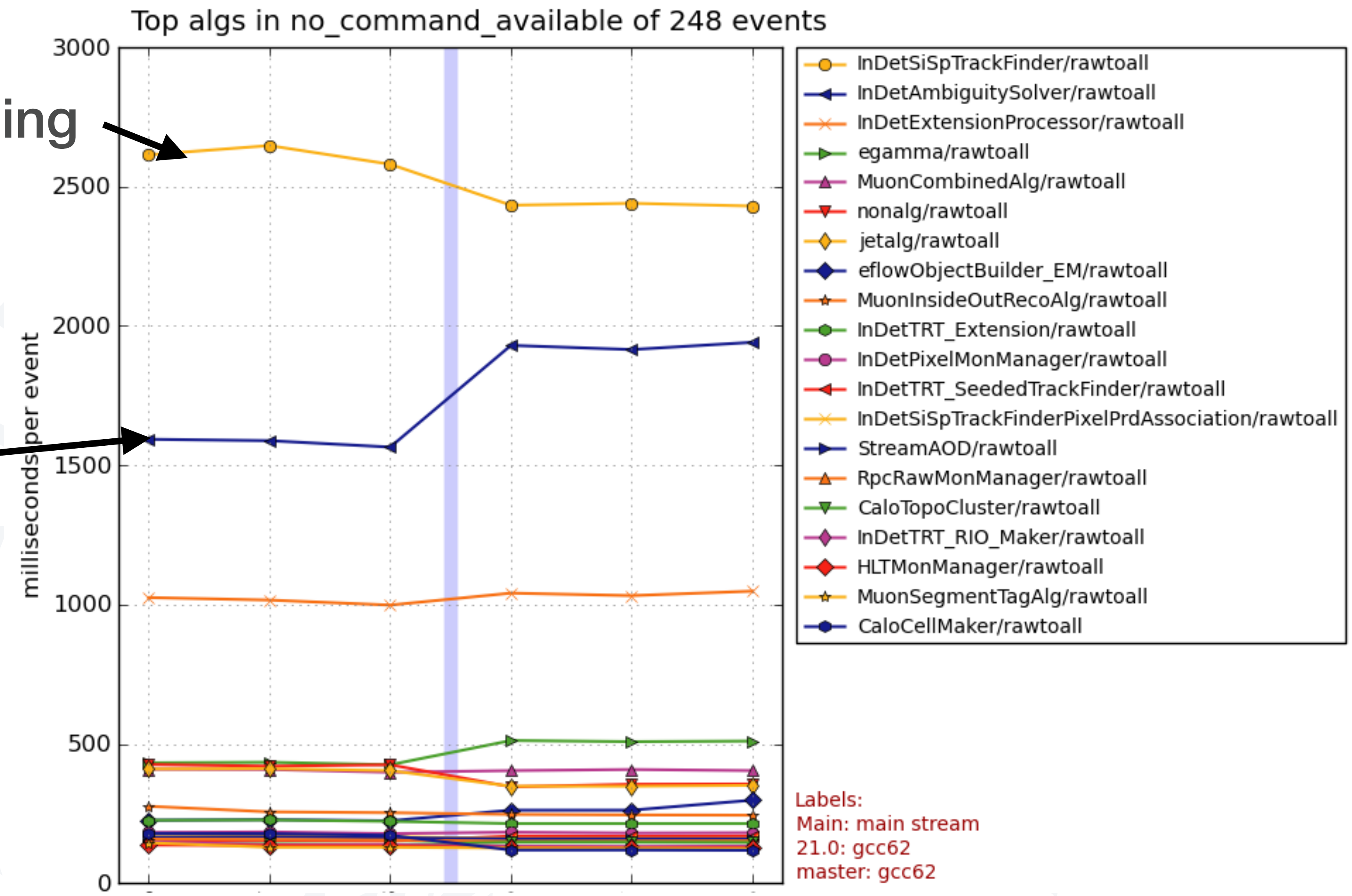
TrueNorth & HPC

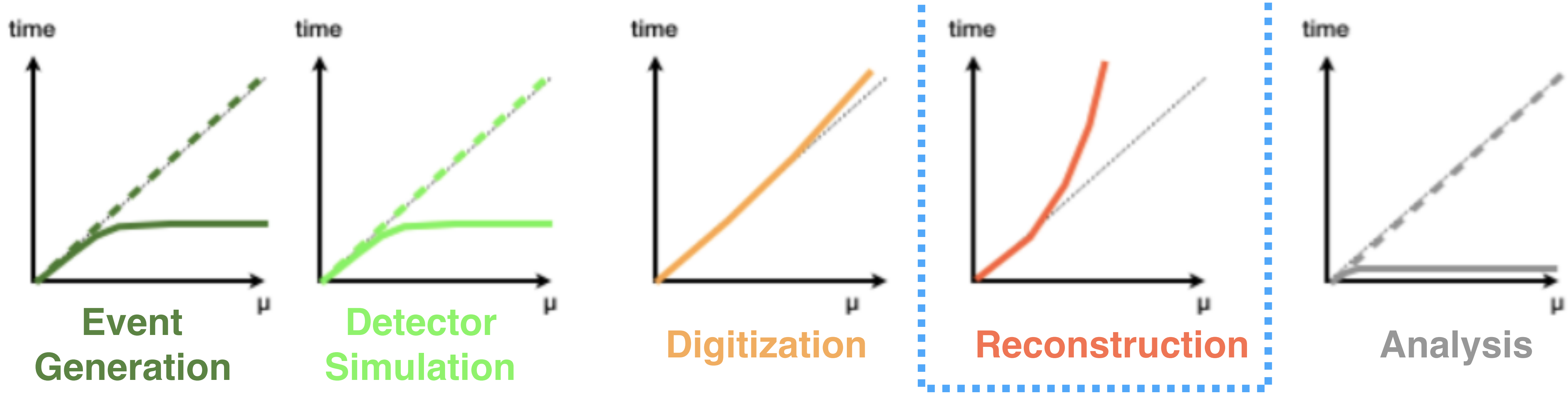
Future **CPU consumption**

How various processes scale

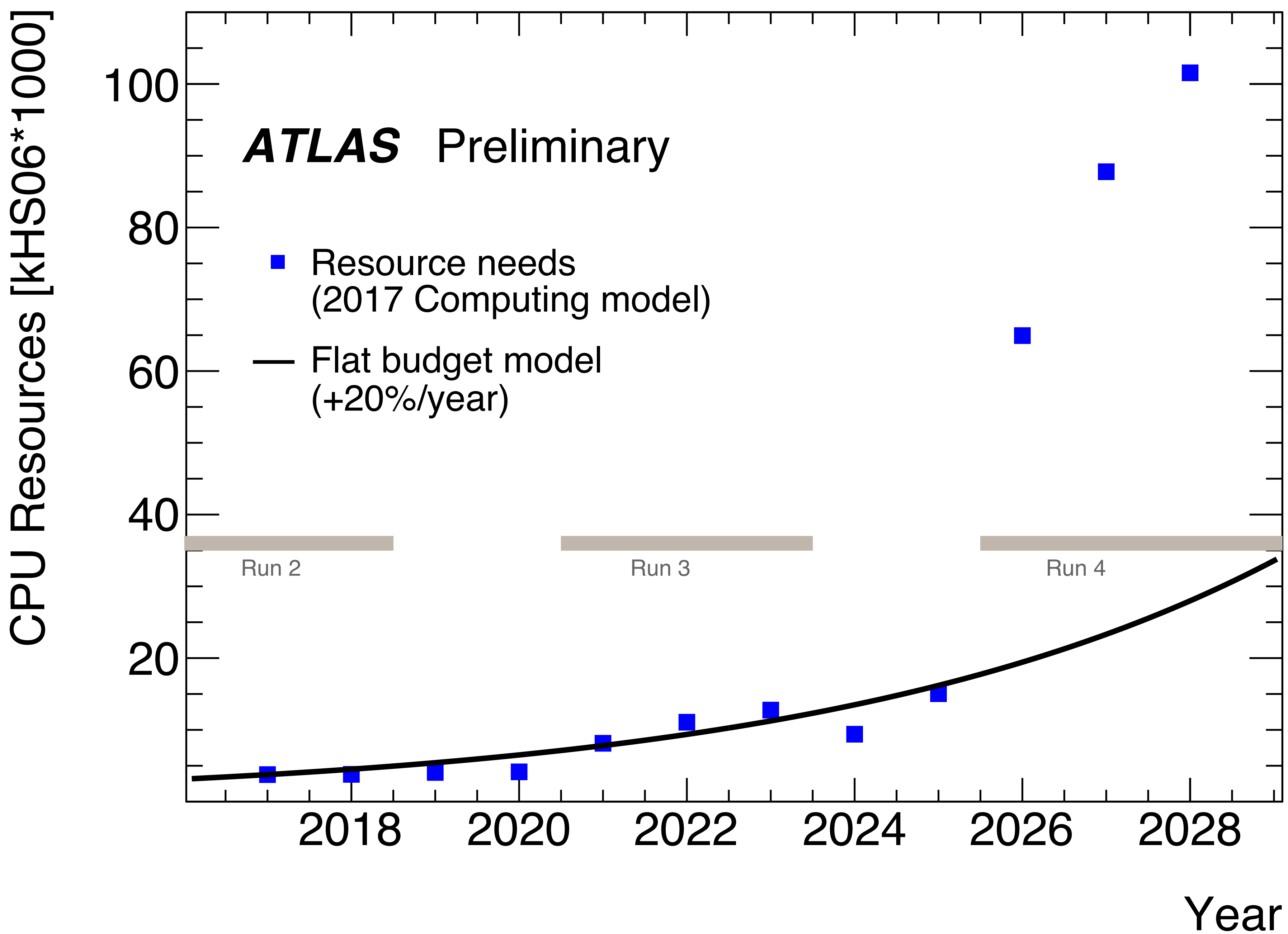
Silicon track finding

Ambiguity resolution

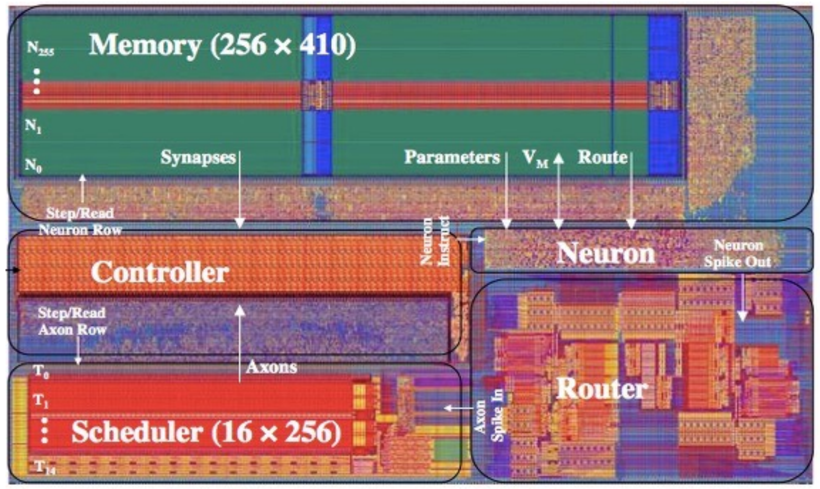


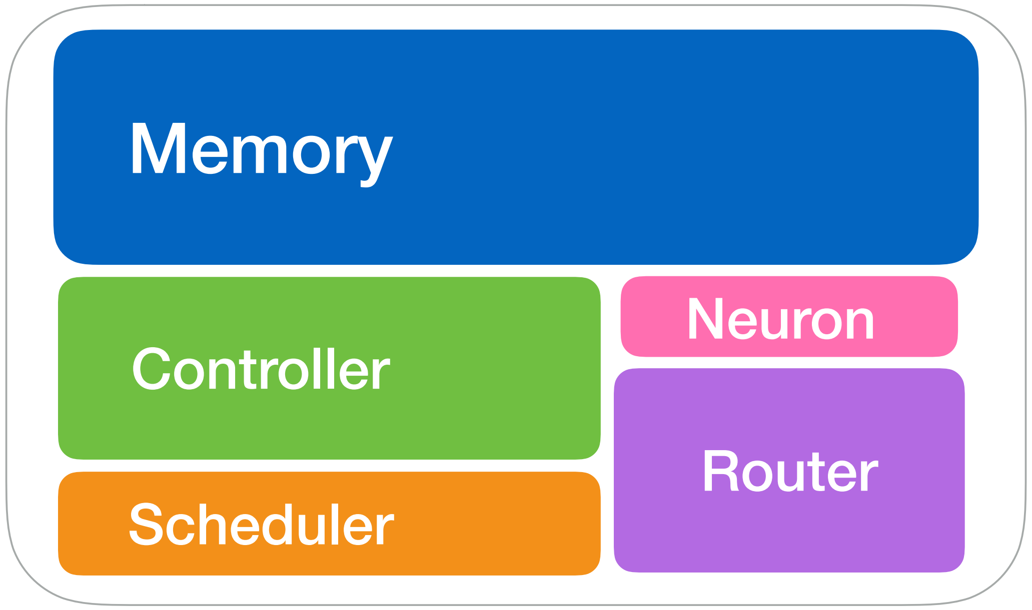


Future **CPU consumption**



- Our requirements will outstrip our resources by a factor of 3
- A different approach is needed: supplement the grid and parallelize.





Remainders in the neuron potential

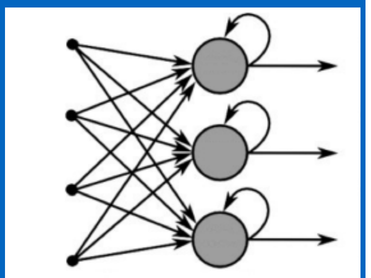
- **Scheduler error:** spike delivery is ensured to be longer than max latency between source and destination. So if chip has a large model implemented and tick is sped up - spike generator may refuse to create file.
- **Token ctrl error:** If the spikes can't be processed in time, because of the volume of spikes in a single core.
- A TrueNorth chip can only take so many spikes per port, and then there is a firmware and software pipeline that must match this **throughput.** We see on the order of **15 million spikes** per second total throughput, which is **480 Mbps** data rate between FPGA and TN.

We never managed to cause either of the first 2 failure modes with our simple serial kaman filter. However we did cause the last error to be thrown when we tried to increase the tick rate from 1kHz to 10kHz. A 2kHz tick was possible.
Throttling arose when using a serial spike train and < 0.1% of the chip area. Unlikely more complicated models could be sped up on this version of TrueNorth.

DOI: 10.1109/TCAD.2015.2474396

rcarney@lbl.gov

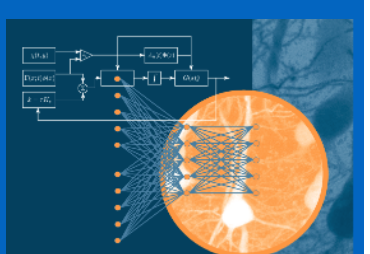
Backup: **Kalman filters** implementations



Things we couldn’t do

- Kalman filter implemented in ANN easily! Can perform covariance and Kalman gain matrix updates using recurrence &/or back-prop. depending on which KF. This is also true in FPGA and GPU implementations.
- **Cannot do this on TrueNorth because:** neuron parameters and connectivity cannot be changed at runtime and training is non-trivial (activation function is not differentiable). So would have to continuously stop and start chip to update model file.

Other implementations we couldn’t copy

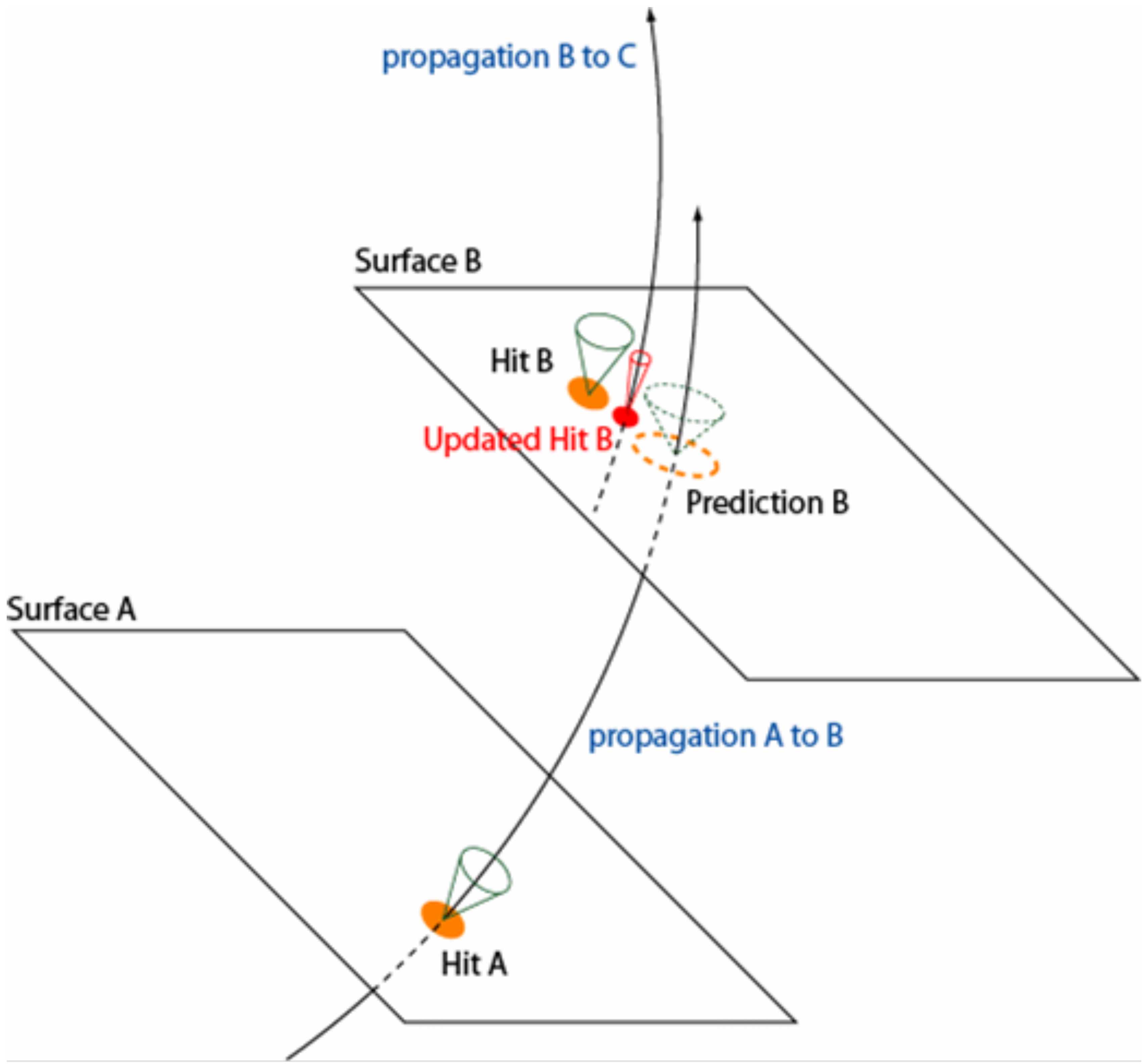


- Neural Engineering Framework implements linear control systems in *populations* of neurons. Perfect for a KF!
- **Cannot do this on TrueNorth because:** requires a more complex neuron model that includes non-linear leakage current & analogue spiking to shape post-synaptic current. Will not work for simple TrueNorth neuron model nor with digital spikes.

BERKELEY LAB

rcarney@lbl.gov

Kalman propagation

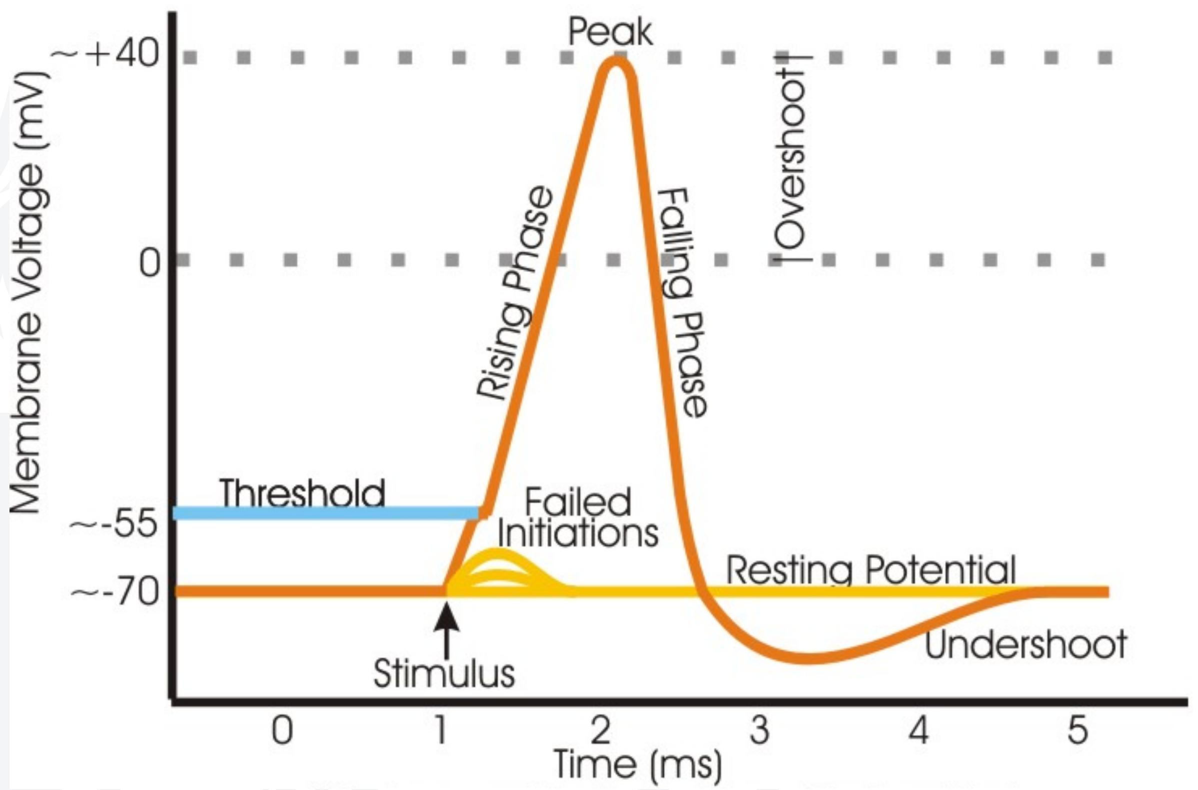


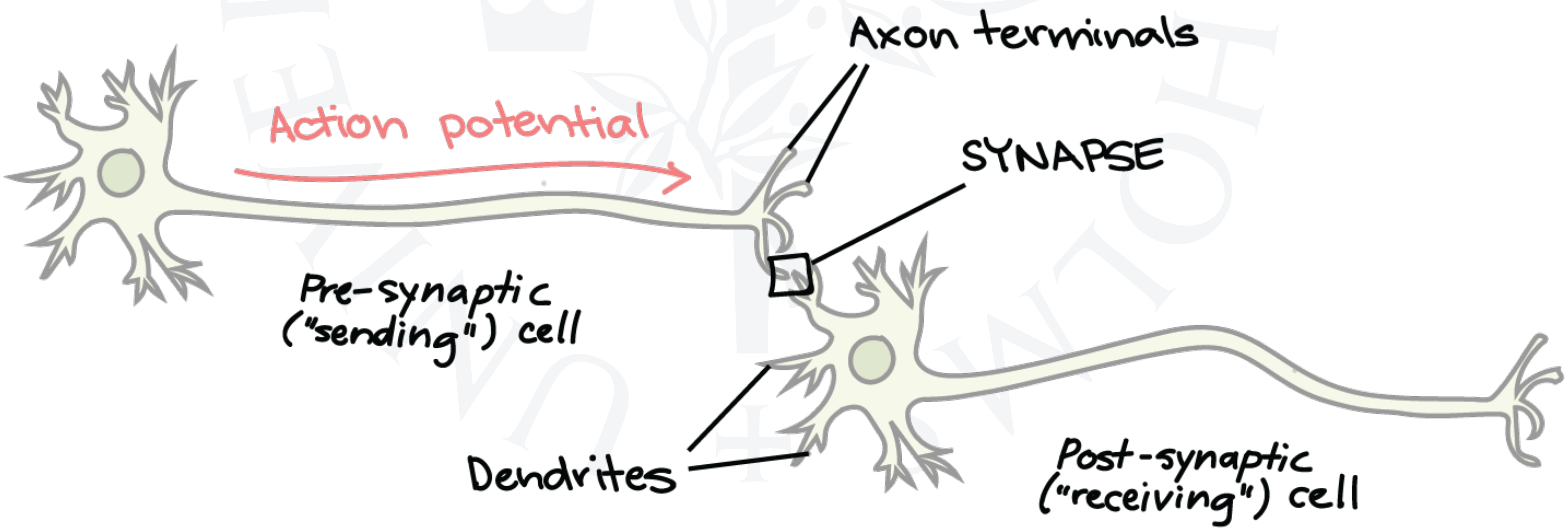
Taken from TrkKalmanFitter package documentation, author unknown. Internal link: https://gitlab.cern.ch/atlas/athena/blob/5aa7198ccd5f421e64aecb308ed93889c54f8391/Tracking/TrkFitter/TrkKalmanFitter/doc/images/KalmanPicture.png

Backup: **Neuromorphic computing**

Biological origins

- Neurons transmit information.
- Potential difference across inside & outside of cell.
- Incoming spikes change potential of membrane.
- Past a certain threshold spike is emitted.





McCulloch-Pitts neuron

The McCulloch-Pitts neuron is a simplified version of the perceptron, without a bias and with no Hebbian learning rules. It has a binary output based on a thresholded step function.

$$y = step\left(\sum_{i=1}^{N} n_i s_i\right)$$

Hodgkins-Huxley neuron

Hodgkins-Huxley model is a biologically plausible neuron model that accounts for leakage across the membrane from different ions and different rates of ions. The I_i term is a parameterized model of the conductance across the membrane, with at least 20 parameters.

$$C_m \frac{dV(t)}{dt} = -\sum_i I_i(t, V)$$

TrueNorth neuron

II. NEURON SPECIFICATION

Our neuron model is based on the leaky integrate-and-fire neural model with a constant leak, which we augmented in several ways. We begin by briefly reviewing the classic leaky integrate-and-fire neural model, followed by an in-depth description of our neuron model.

A. The Leaky Integrate and Fire (LIF) Neuron

The operation of the leaky integrate-and-fire (LIF) neuron model with a constant leak is described by five basic opera-

$$y = step\left(V_j(t-1) + \sum_{i=1}^{N=256} n_i(t) s_j^{G_i} w_{i,j} + \lambda_j\right)$$

Leaky Integrate-and-fire neuron

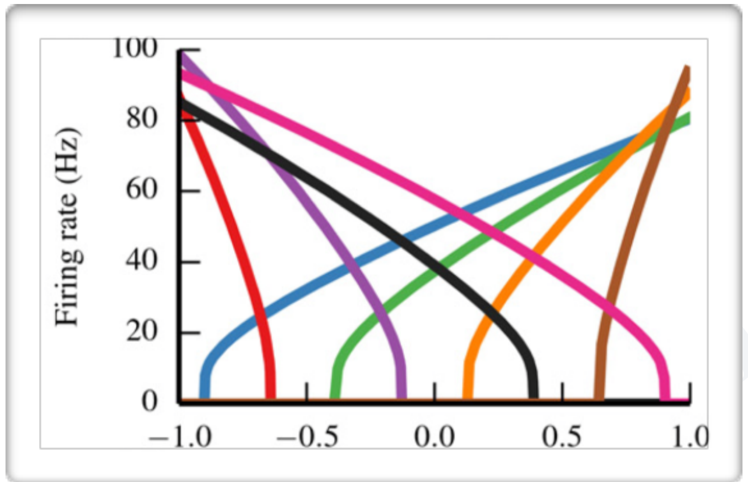
One of the simplifications to this model to make it tractable is the LIF neuron. Leaky integrate-and-fire is a reduced complexity subset of HH. Grouping any current independent of input as a constant leak.

$$C_m \frac{dV_m(t)}{dt} = I(t) - \frac{V_m(t)}{R_m}$$

Calling TN neurons LIF seems to be a misnomer. They are far closer McCulloch-Pitts neurons and have none of the time-dependent capabilities of LIF neurons (aside from the persistent neuron potential).

DOI: 10.1109/IJCNN.2013.6707077

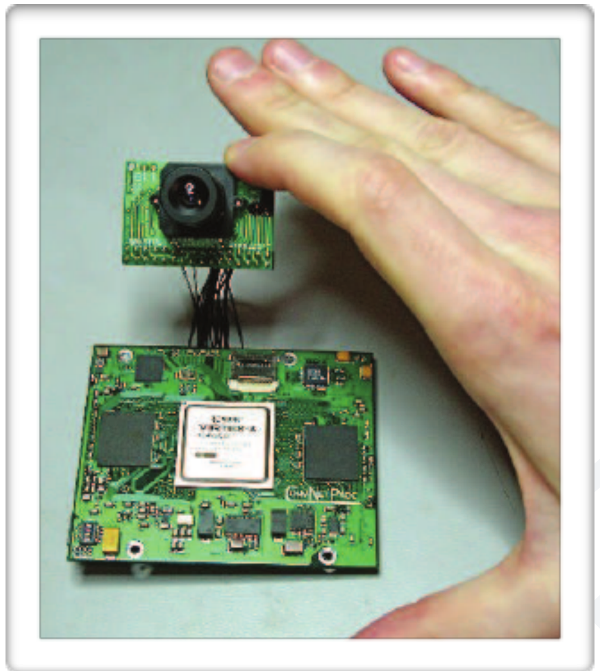
Neuromorphic hardware



2013-18: Neuromorph and Brainstorm @ Stanford

2016: Spinnaker 0.5M core machine @ HBP/Manchester

2011: Spikey @ Heidelberg



2008: NeuFlow @ NYU (J. LeCunn)

1. **Spinnaker**: scalable, low-power units
2. **Brainstorm**: NEF and populations of neurons
3. **Spikey**: analogue spikes, neuroplasticity, 10k speedup over real-time

.

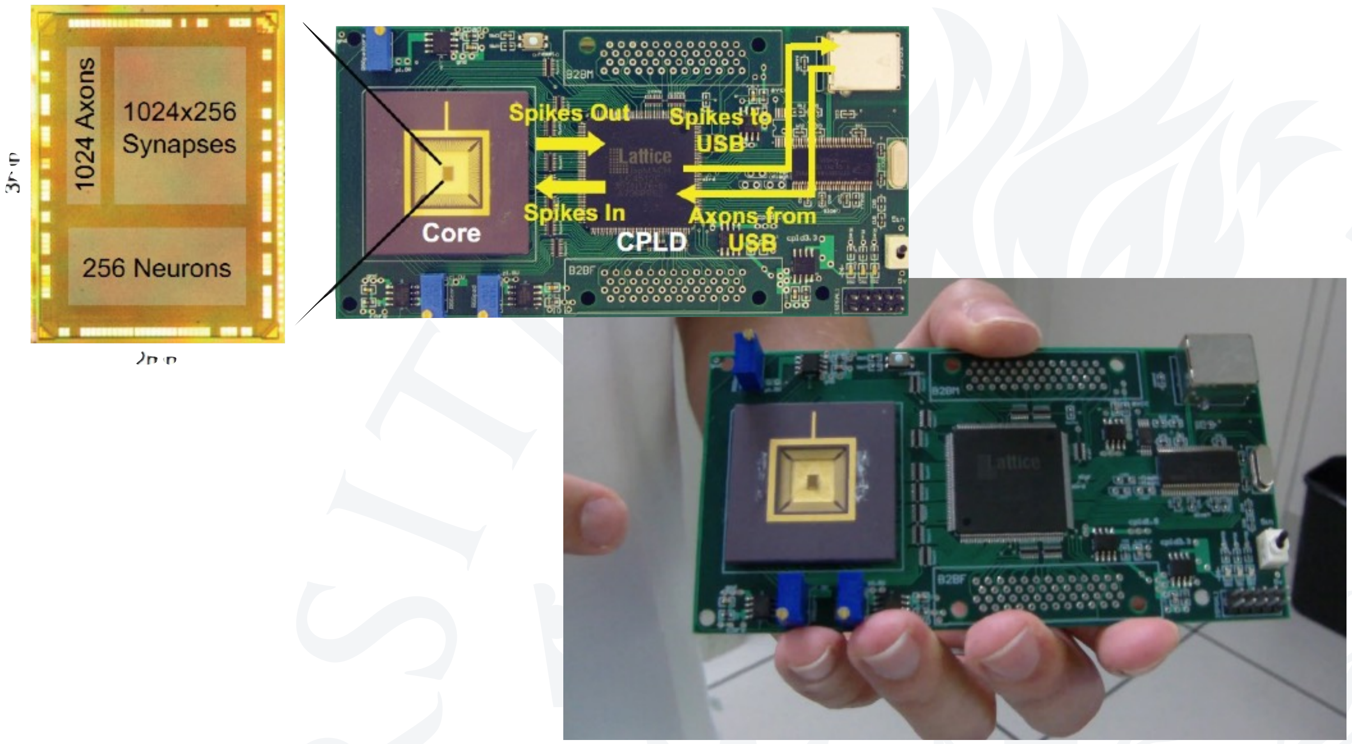
.

NeuFlow: Conv. neural network on-chip. Not really neuromorphic.. but an example of NN hardware!

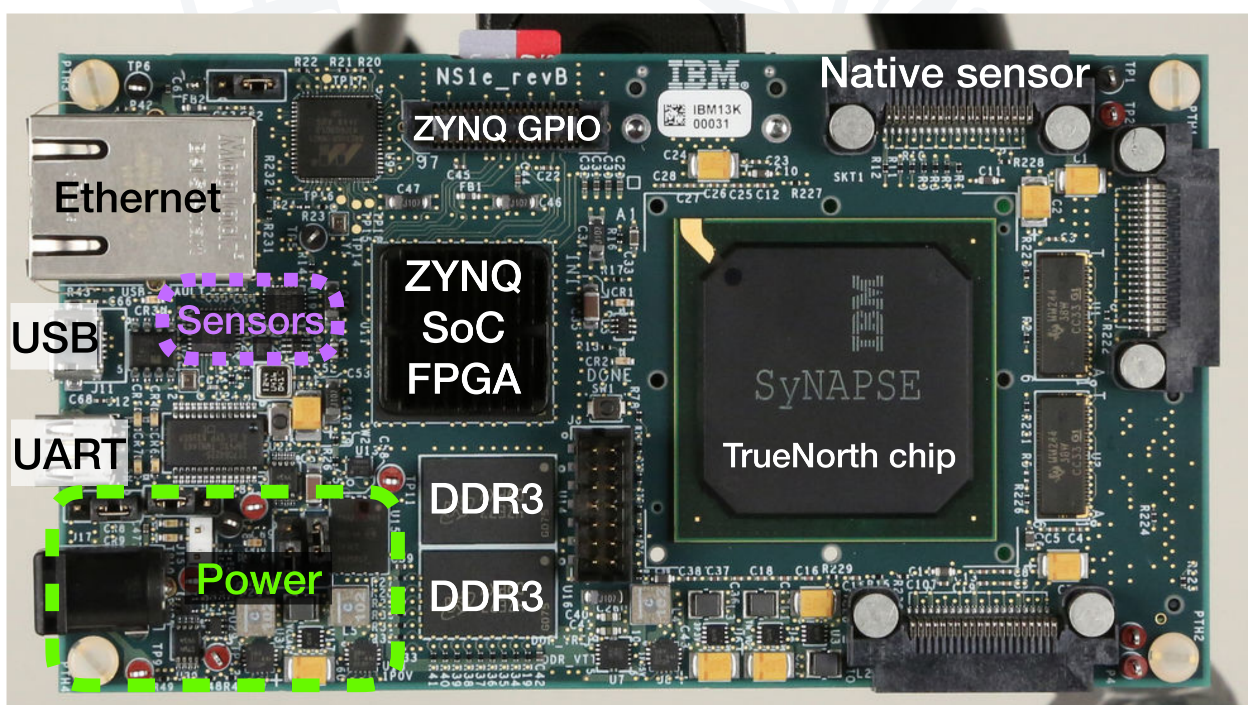
More neuron-inspired computing at NICE 2016 [link]

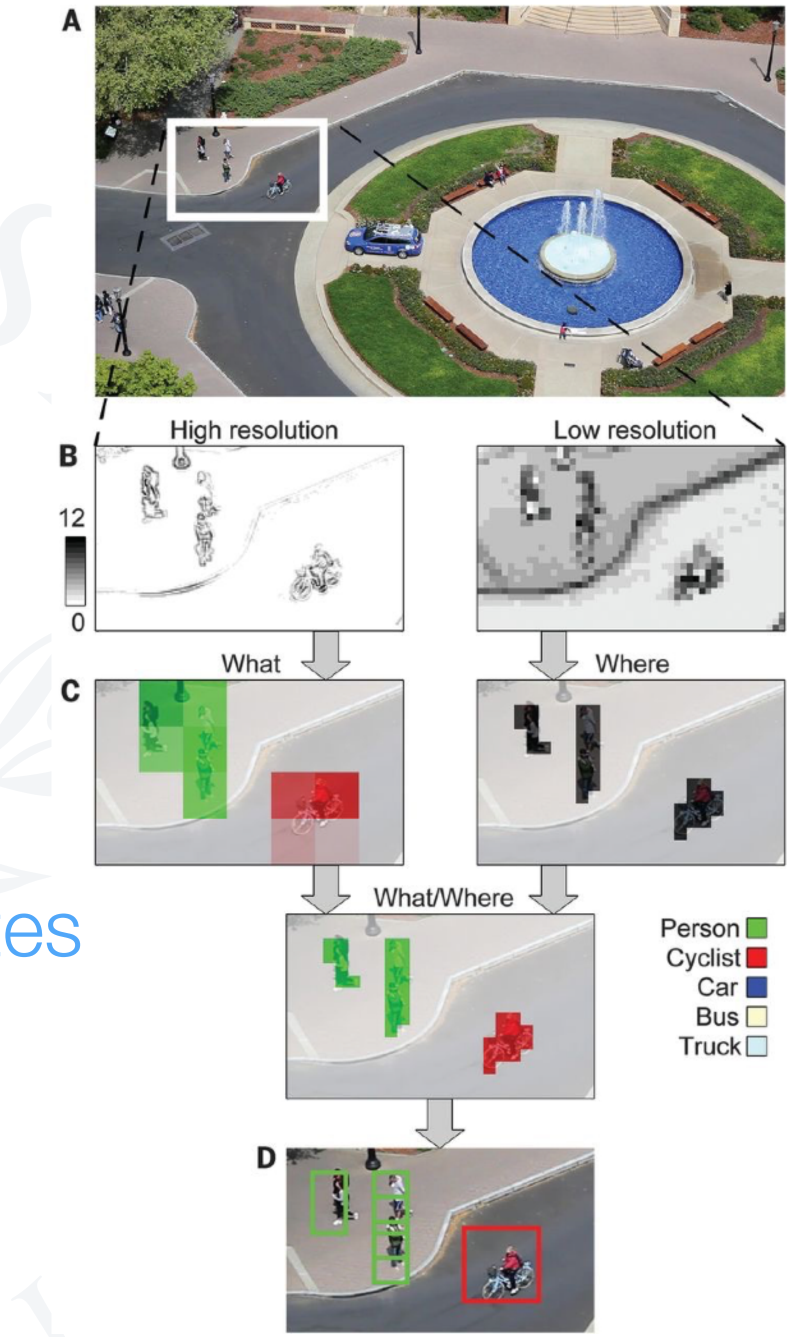
Brief history

2011 - a single core demonstrator



2014/15 - TrueNorth chip to select institutes





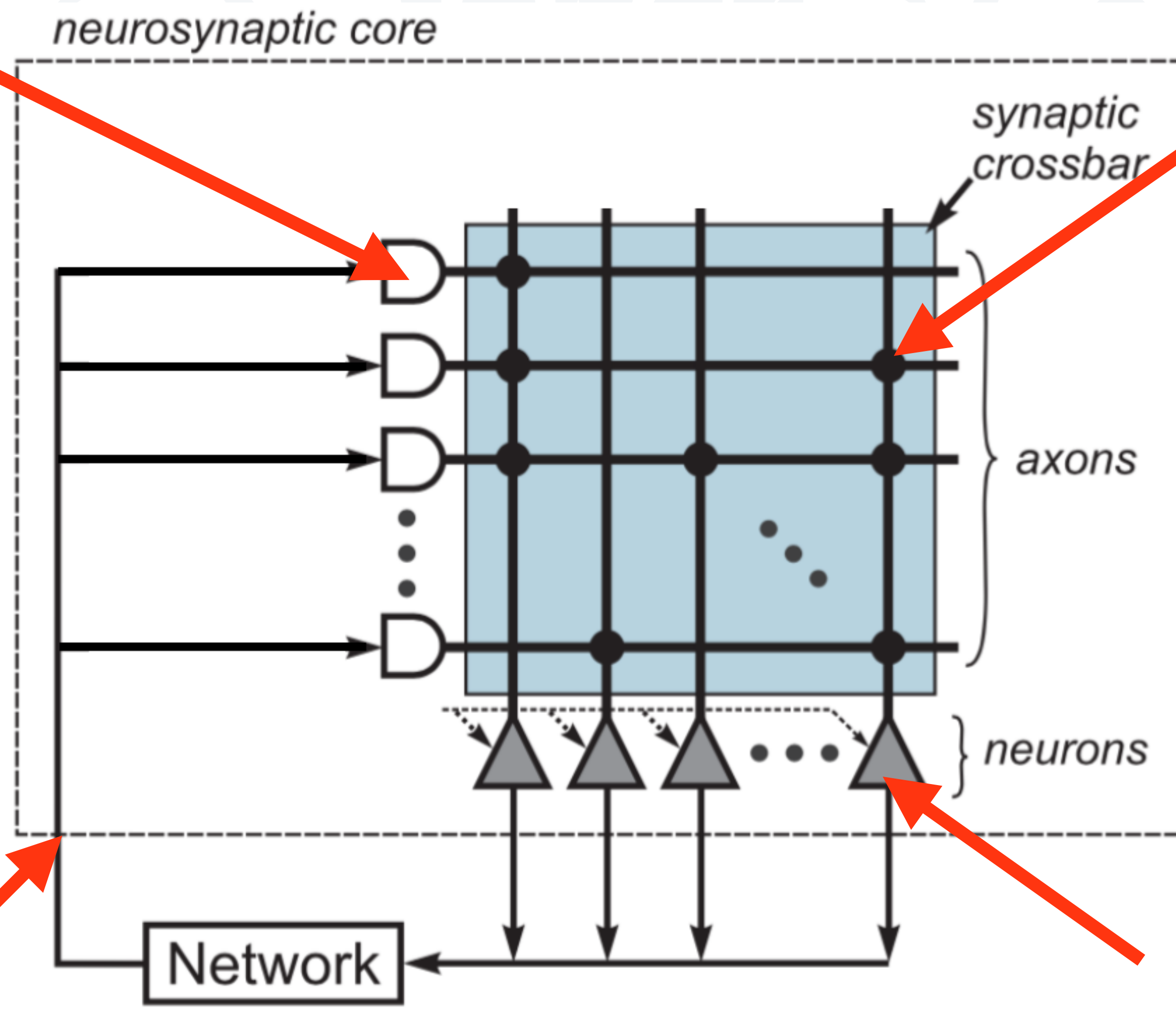
DARPA SyNAPSE: metric of one million neuron brain-inspired processor.

IBM

Core and crossbar

Axons are the xbar input. Also provide a label ∈ {1,2,3,4}

Synapses ∈ {0,1}



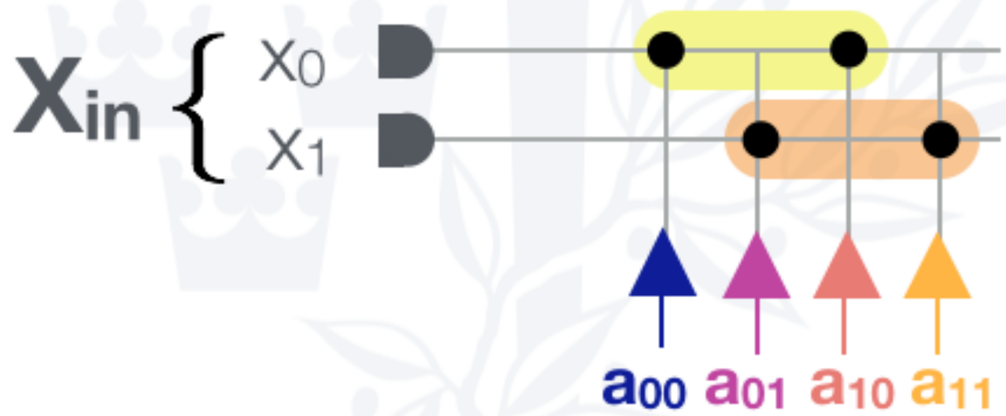
Connections are between 1 neuron and 1 axon. Fan out/in requires multiple neurons/axons.

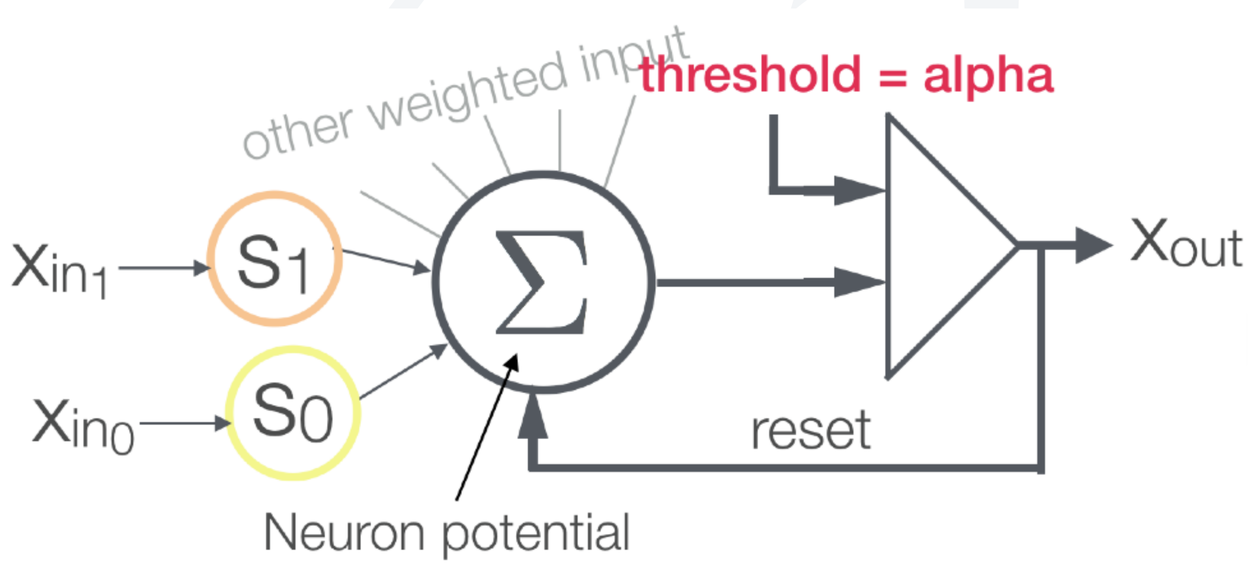
Weights and threshold applied in neuron. A signed 8-bit weight can be saved per axon *label.*

Berkeley Lab

Extrapolation of state vector

$$Ax = \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \end{pmatrix} = \begin{pmatrix} a_{00}x_0 + a_{01}x_1 \\ a_{10}x_0 + a_{11}x_1 \end{pmatrix}$$

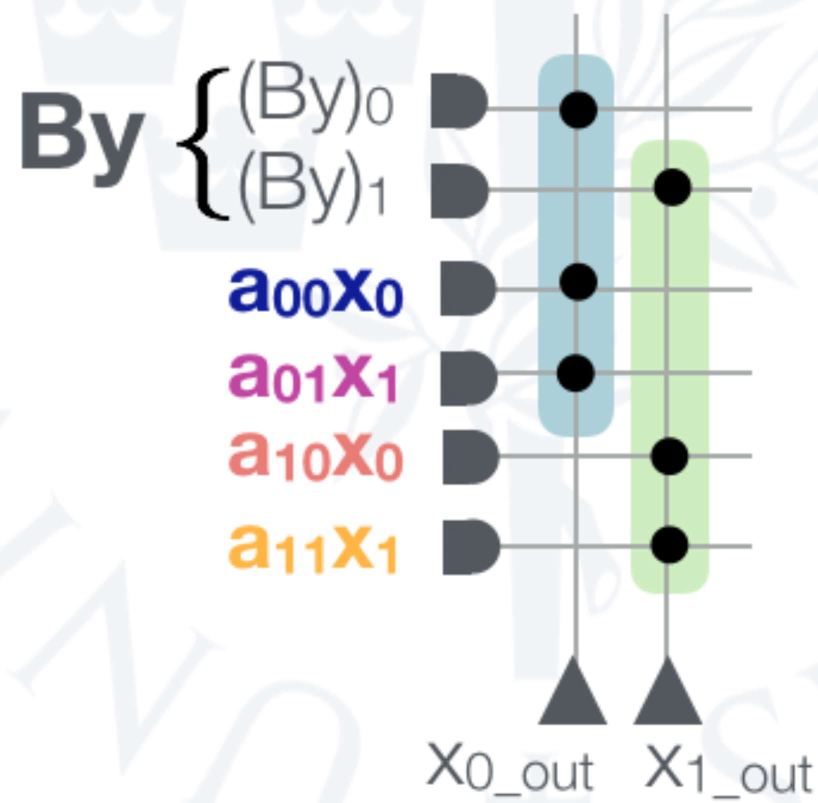


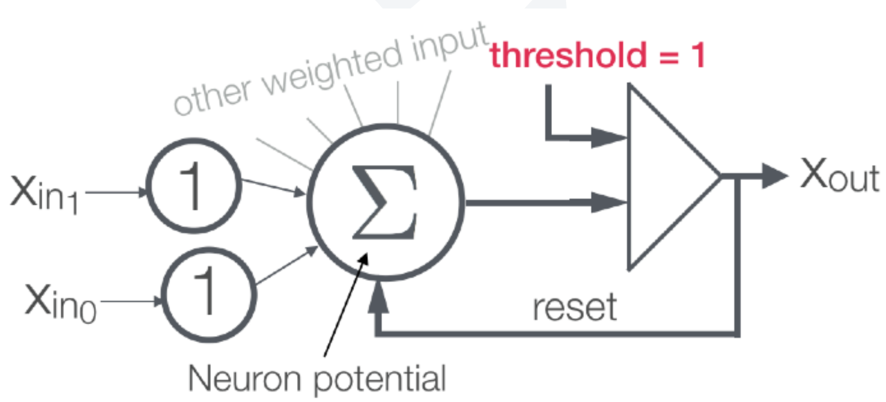


Kalman filter in TrueNorth

State vector update

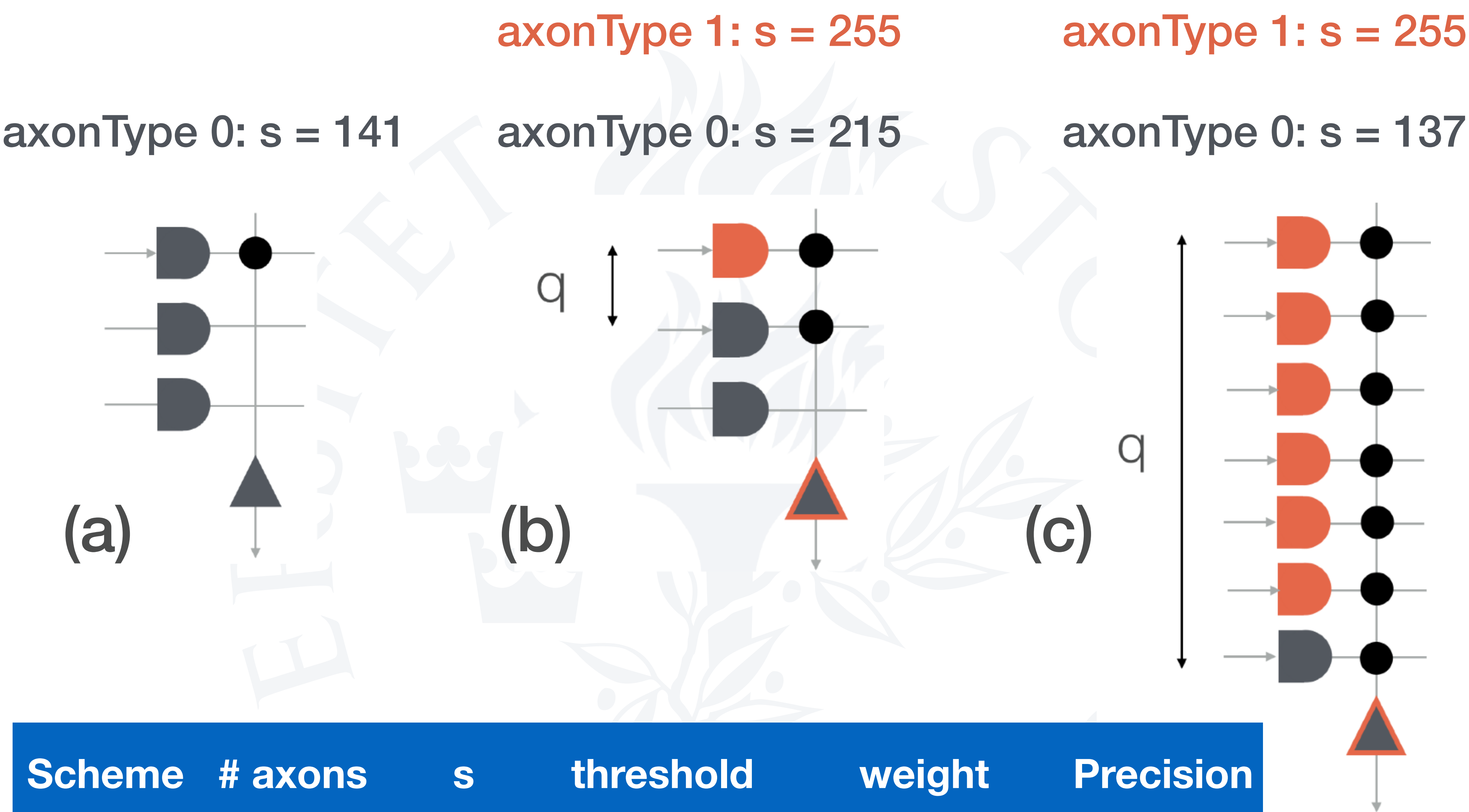
$$Ax + By = \begin{pmatrix} a_{00}x_0 + a_{01}x_1 \\ a_{10}x_0 + a_{11}x_1 \end{pmatrix} + \begin{pmatrix} (By)_0 \\ (By)_1 \end{pmatrix}$$





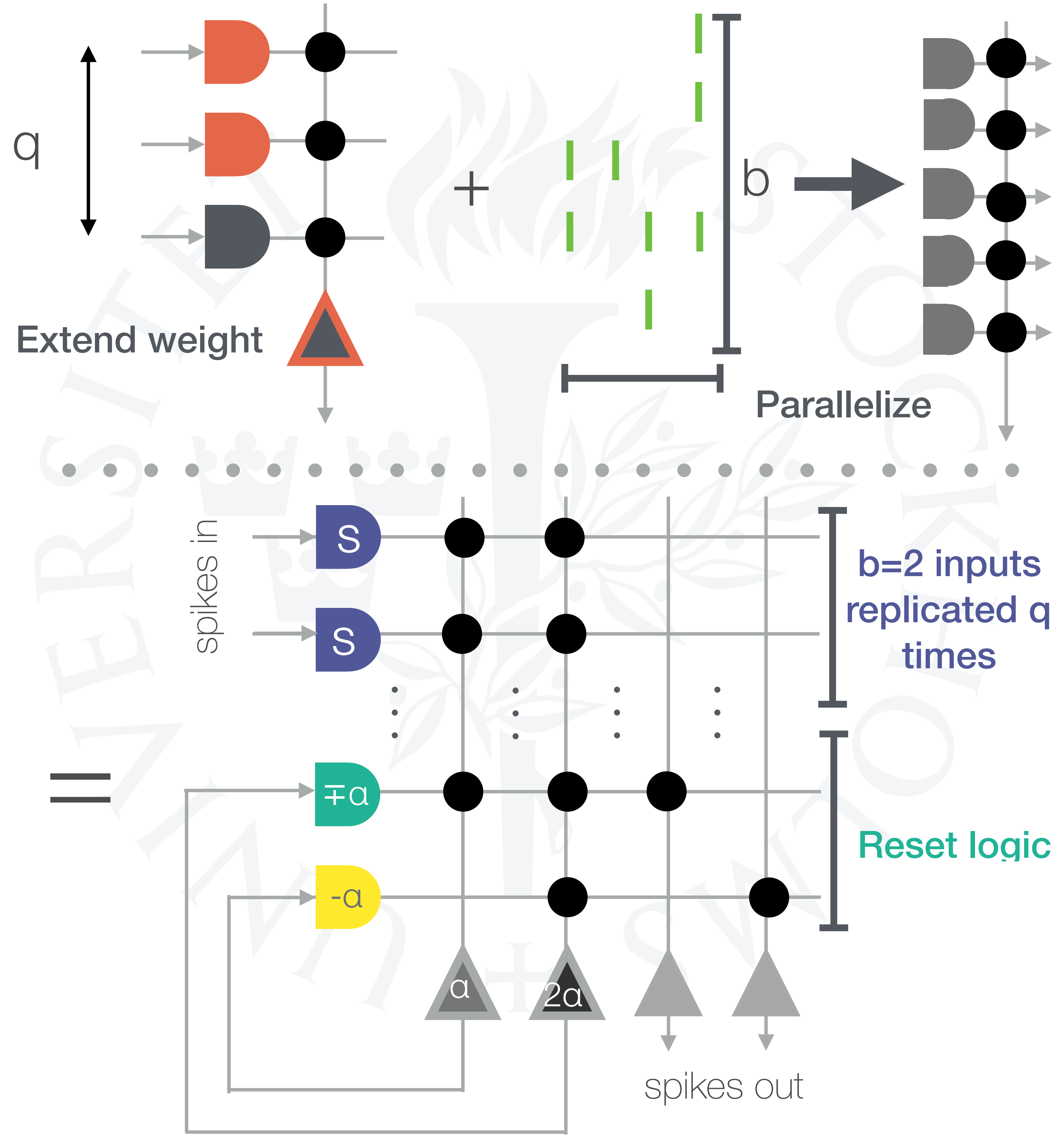
TrueNorth challenges

Weight representation

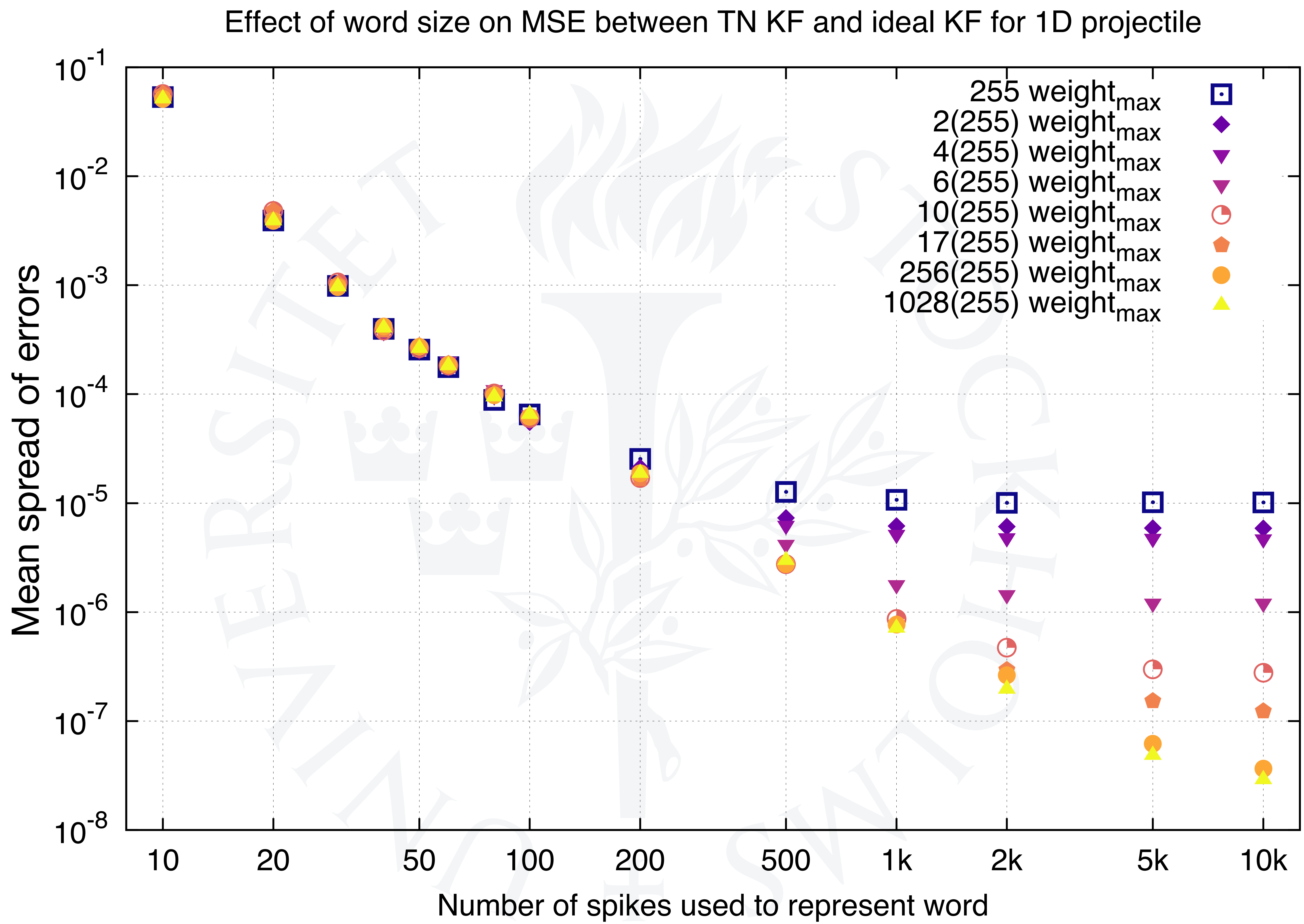


Scheme	# axons	s	threshold	weight	Precision
a	1	45	49	0.91837	3 s.f.
b	2	251	306	0.91830	4.s.f.
c	7	1922	2093	0.91829909	7.s.f.

Non-linear dependence

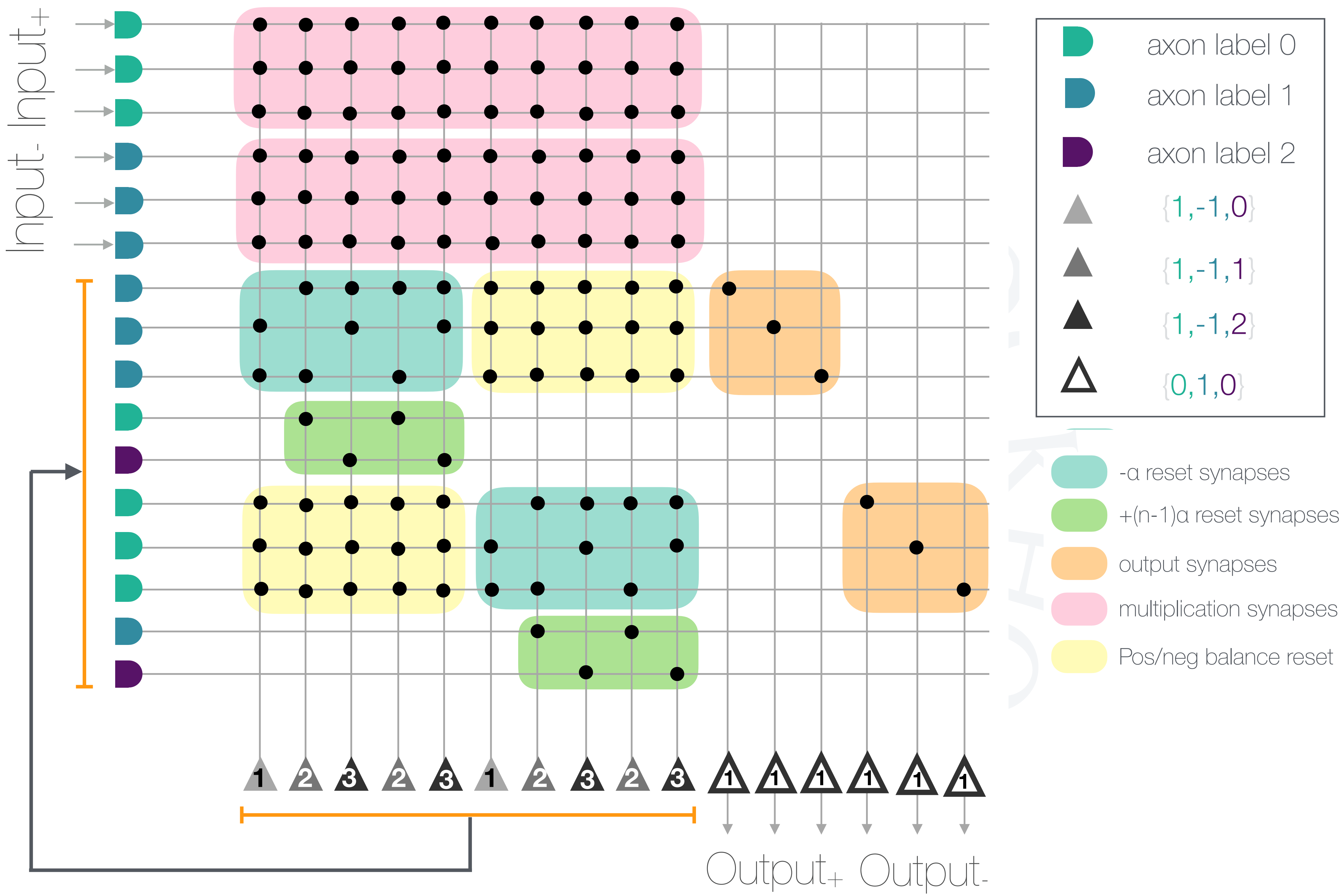


Effects of weight representation



A compromise must be struck between accuracy and speedup: necessary because of the chips design and the way information is encoded

Subtitle



(population encoded discrete steady-state, 3-dim KF)

rcarney@lbl.gov

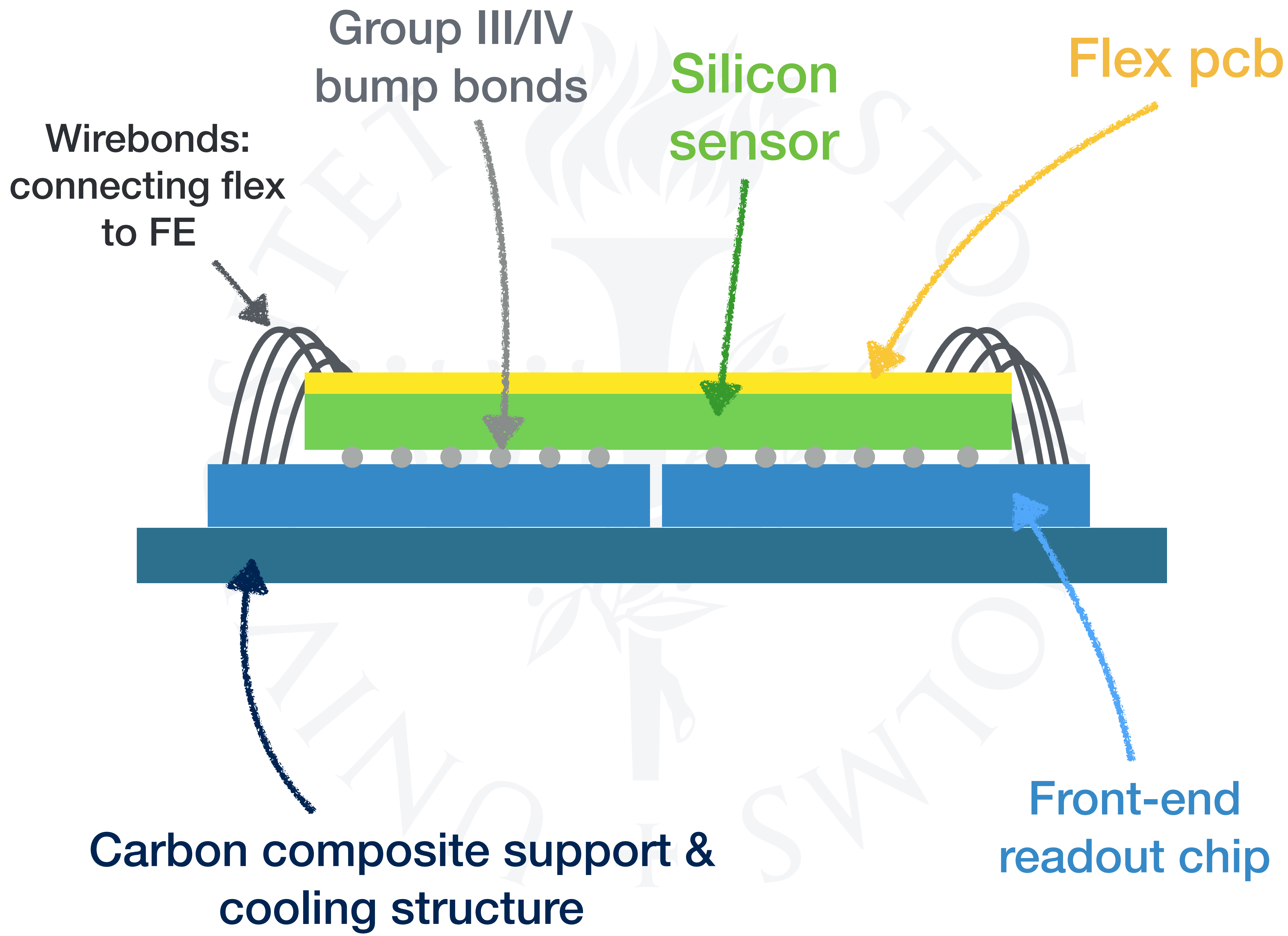
ITK &
FE65-P2

Conditions for ITk will be challenging…

	ITk Pixel: innermost layer (flat modules)	Run 2: IBL	Difference
Lifetime target integrated lumi.	2000 [fb^{-1}]*	300 [fb^{-1}]	x7
Lifetime fluence	1.3e16 [neq/cm^2]*	5e15 [neq/cm^2]	x2.6
TID	7 [MGy]*	2.5 [MGy]	x2.8
<Hits/chip/bc at target PU>	223 @ PU 200	27 @ PU 75	x8
<Density of hits>	0.6 [mm^{-2}]	0.08 [mm^{-2}]	x8

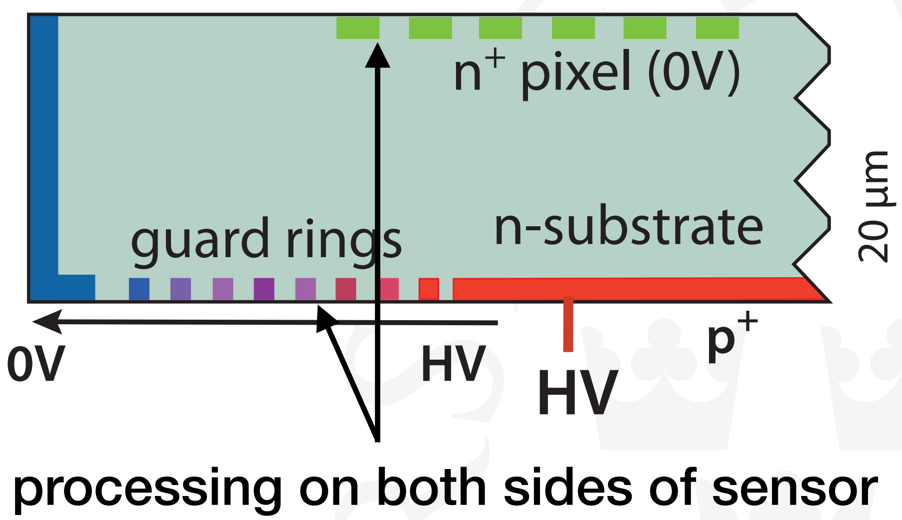
*assuming replacement after 2000 fb^{-1}

Module reminder

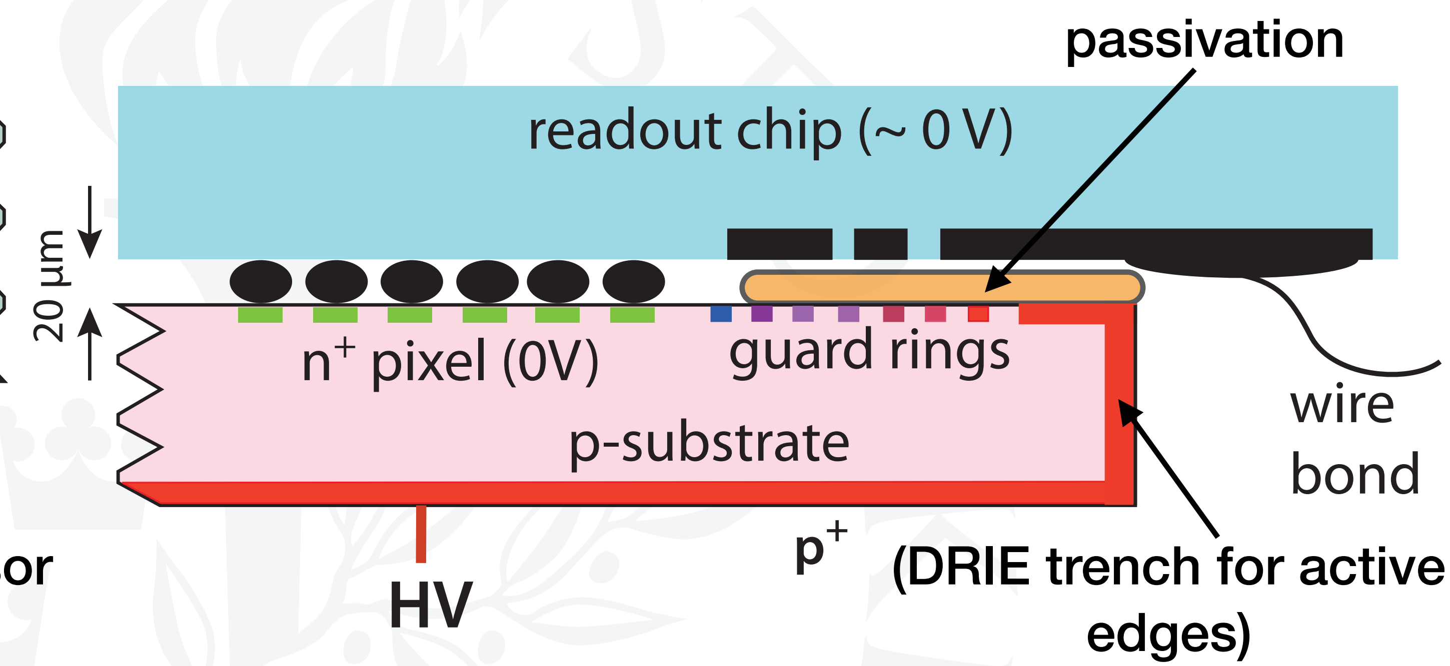


Planar sensors

- **n-in-p**:
  - ◉ single-sided processing: easier/cheaper to produce
  - ◉ needs passivation between FE chip and sensor



Current ATLAS Pixel sensors



ITk Pixel Sensors

- **Reduced thickness**:
  - ◉ better rad-hardness,
  - ◉ difficult to handle, > cost
  - ◉ current Pixel sensors: 200-250 um

ITk Layer	Sensor thickness [um]
0	3D sensor
1	100
2	100-150
3	150
4	150-300
EC	200

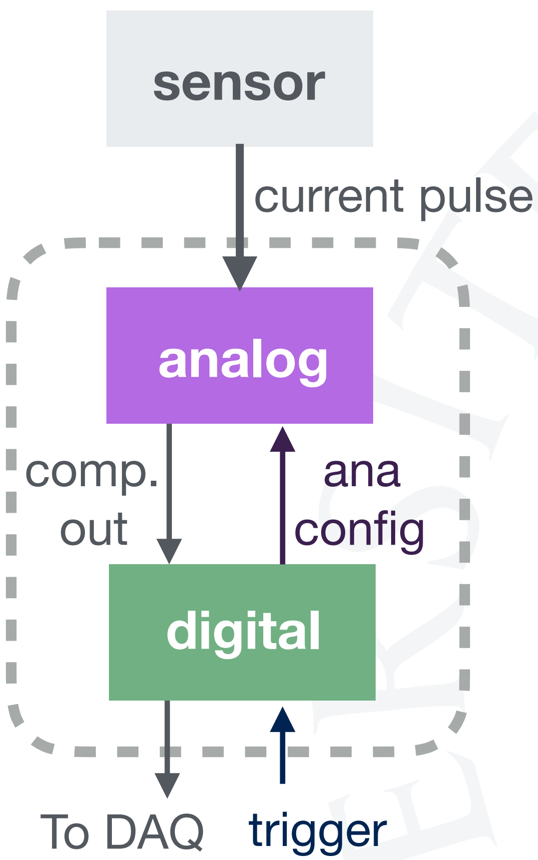
Lifetime fluence	1.2e16 [neq/cm^2]*	2.5 x Run 2

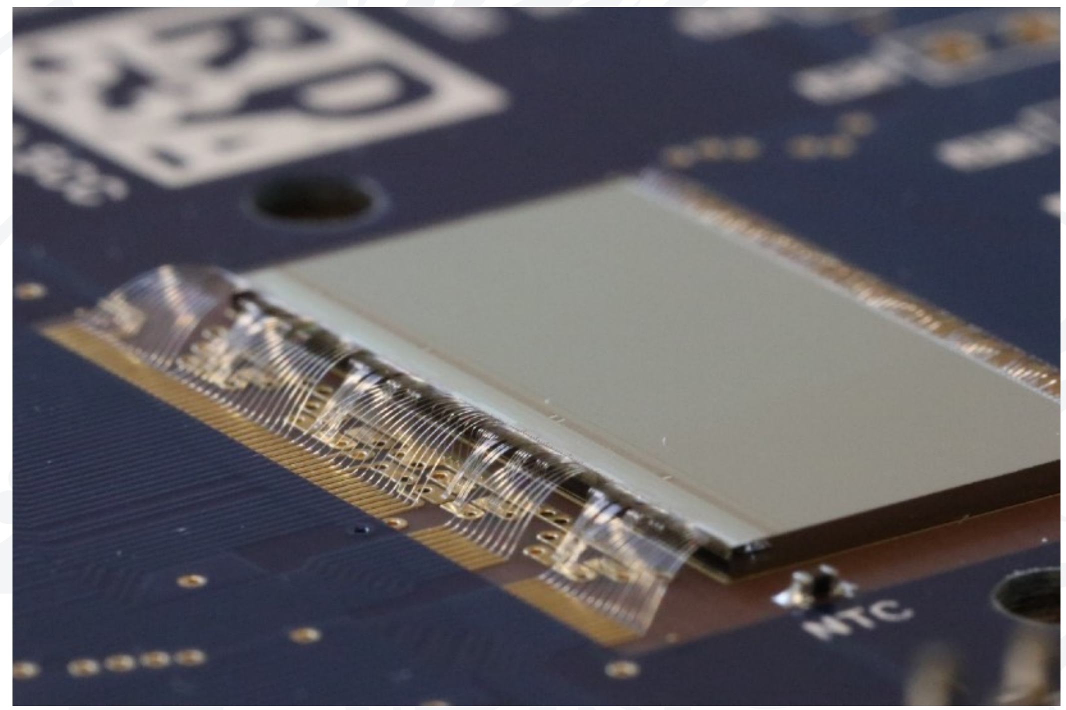
BERKELEY LAB

FE RD53A and beyond

https://twiki.cern.ch/twiki/bin/view/RD53/RD53APublicPlots

- analogue portion: integrates collected charge, amplifies
- digital portion: digitizes signal, stores till trigger, sends hit info to DAQ





RD53A

ITk FE (RD53B)

Coming soon

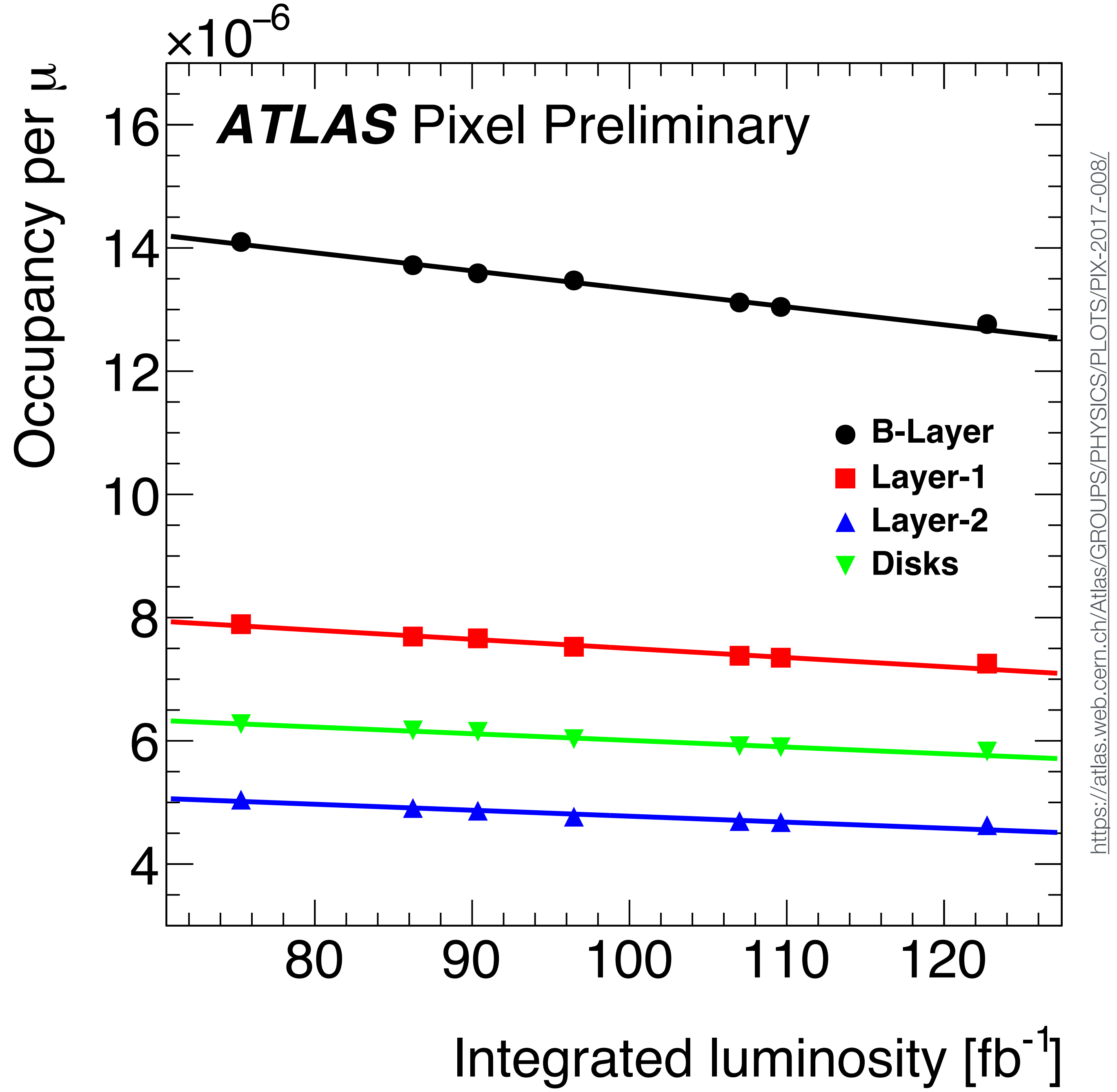
- Complex requirements & high hit rate = smaller CMOS.
- Equivalent rad-hardness, logic density @ 65nm is 3x of 130nm.
- Small test-chips to understand 65nm produced in 2015-2016.
- RD53A produced at end of 2017. Tests finalizing ana-FE and digital core used in final ITk chip are ongoing. **Current under review.**

<Hits/chip/bc>	223 @ PU 200	8 x Run 2
TID	7 [MGy]	2.8 x Run 2
# pixels/chip	153,600	5.7 x Run 2

Radiation Damage Simulation

Detector response changes with time

Less hits per pixel module as detector ages



Pixel sensor occupancy decreasing with integrated lumi.

Detector performance changes with time (fixed threshold)

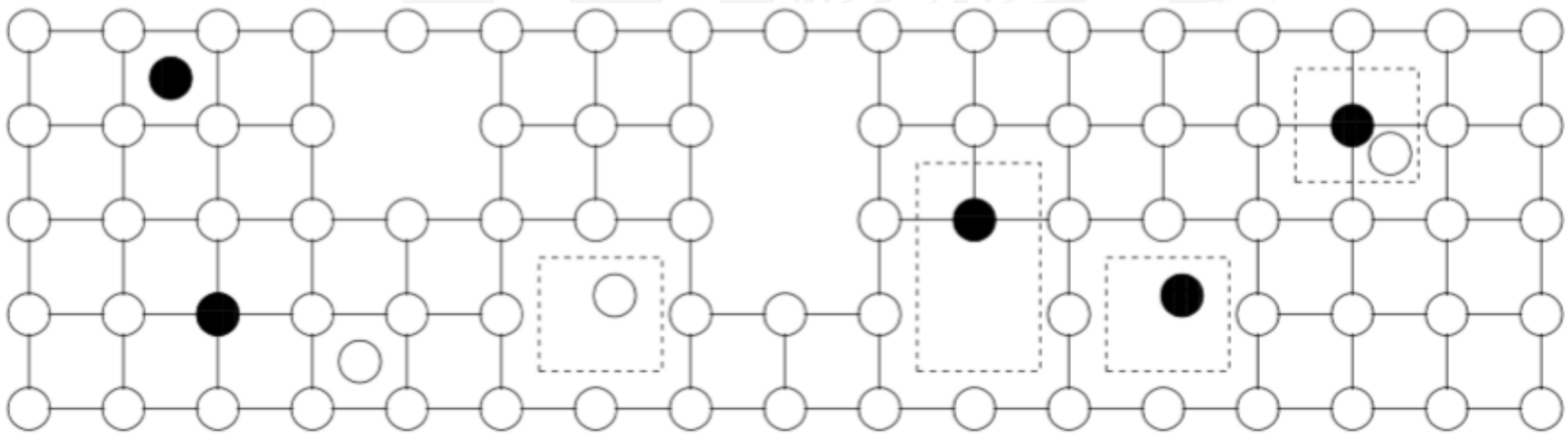
Need to understand these effects and how they change our reconstruction efficiency

Berkeley Lab

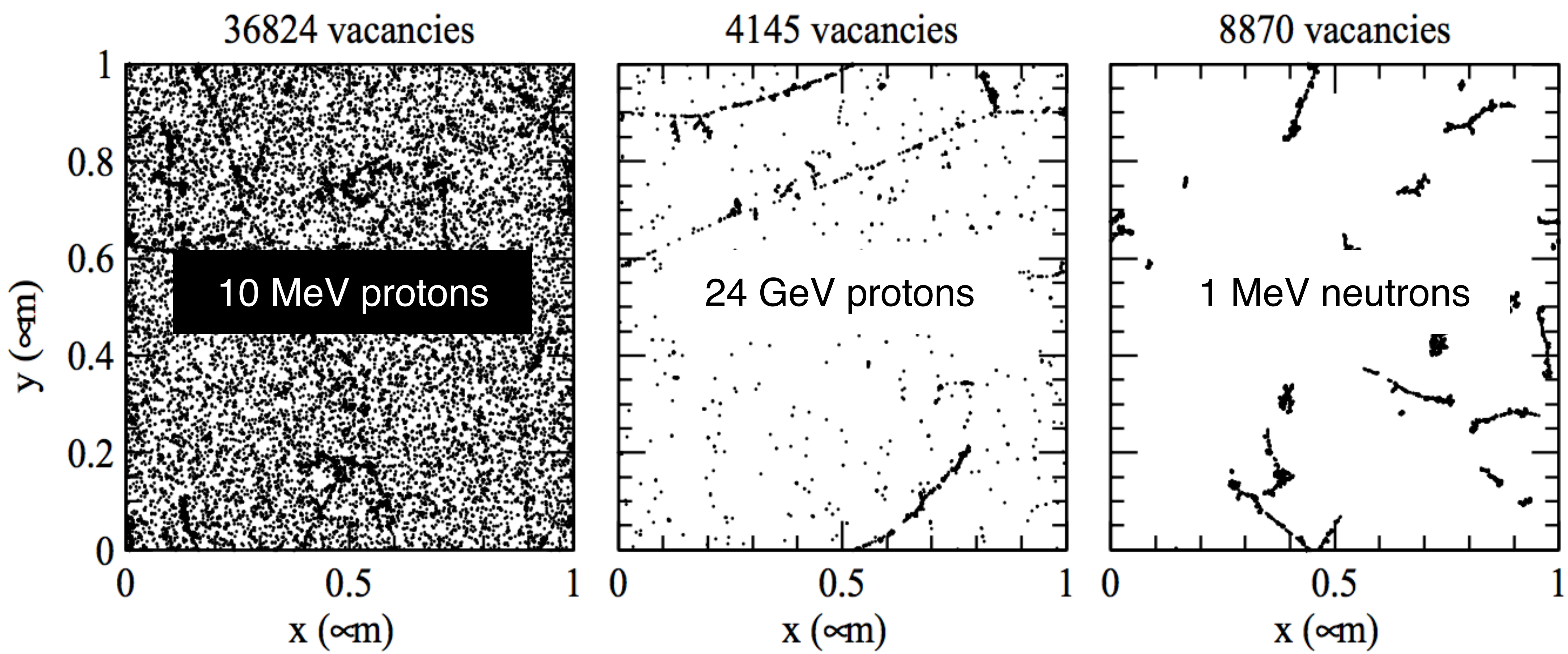
rcarney@lbl.gov

NIEL lattice deformation

- As particles pass through the detector material they can interact with it.
- Non-Ionizing Energy Loss (NIEL) damages sensor bulk
- Deformation of lattice introduces energy levels to silicon band-gap.
- Some of these levels can act as traps



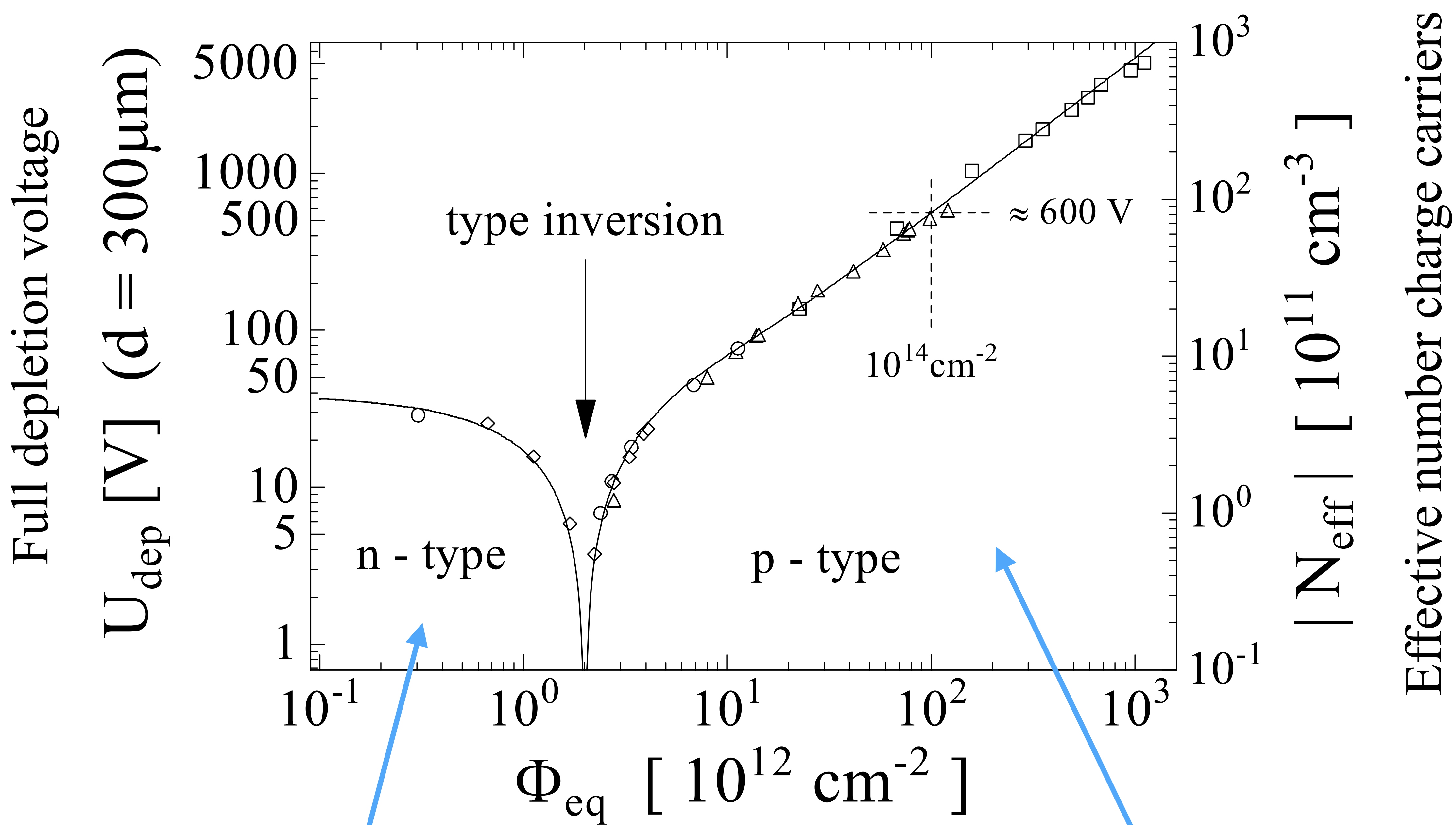
Backup: 1 MeV Neutron equivalent fluence



- The cluster deformations introduced by 24 GeV protons are very different to those introduced by 10 MeV protons.
- 24 GeV protons make comparable damage to 1 MeV neutrons.
- As such they have a hardness factor of 24.

https://mmoll.web.cern.ch/mmoll/thesis/

Backup: Space charge sign inversion



Net space-charge positive

Net space-charge negative

Radiation damage **simulation**

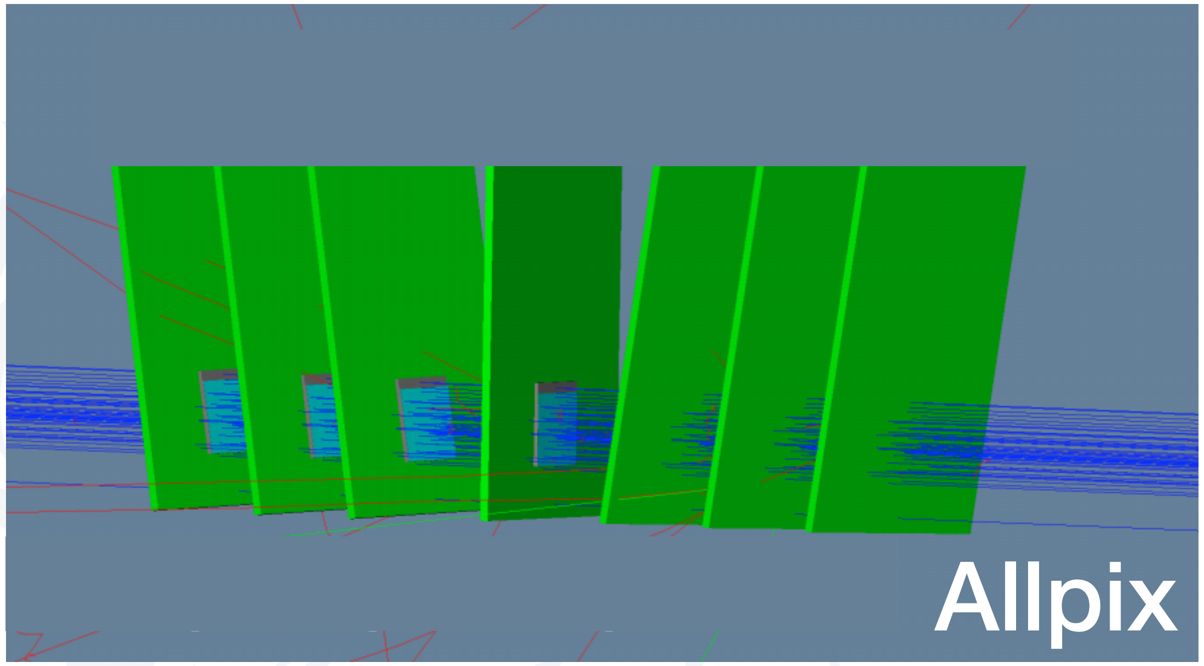
Allpix → Athena

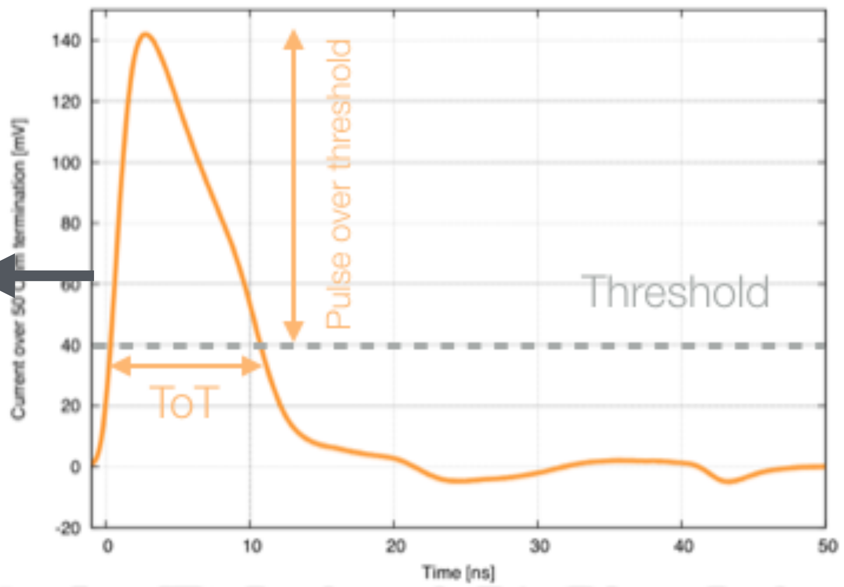
Generate pp interaction

Monte Carlo Detector

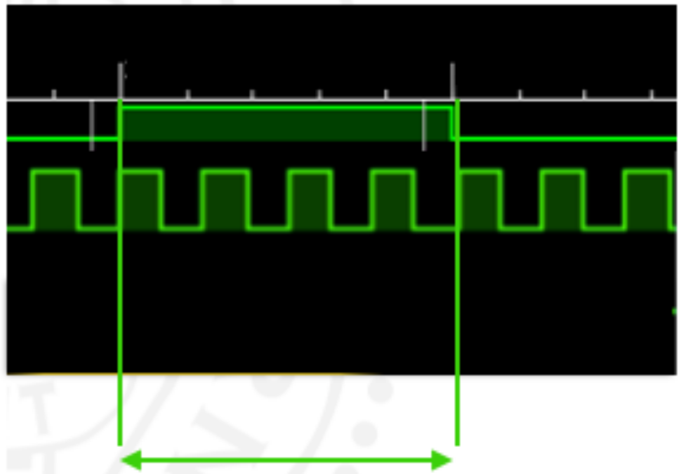
Detailed detector simulation

Event reconstruction





Analogue



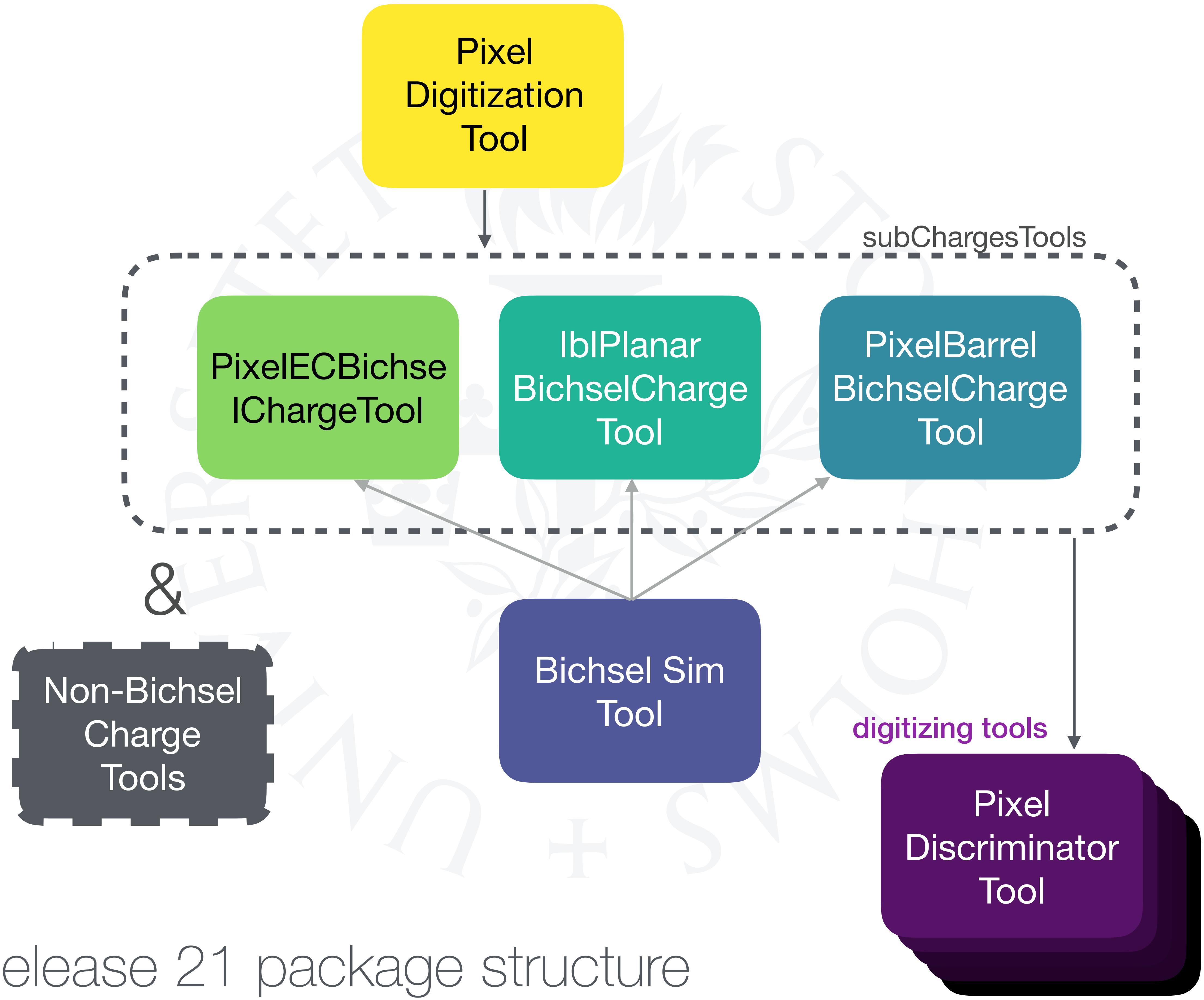
Duration of pulse = ToT

Digital

My task: implementing simulation in ATLAS framework

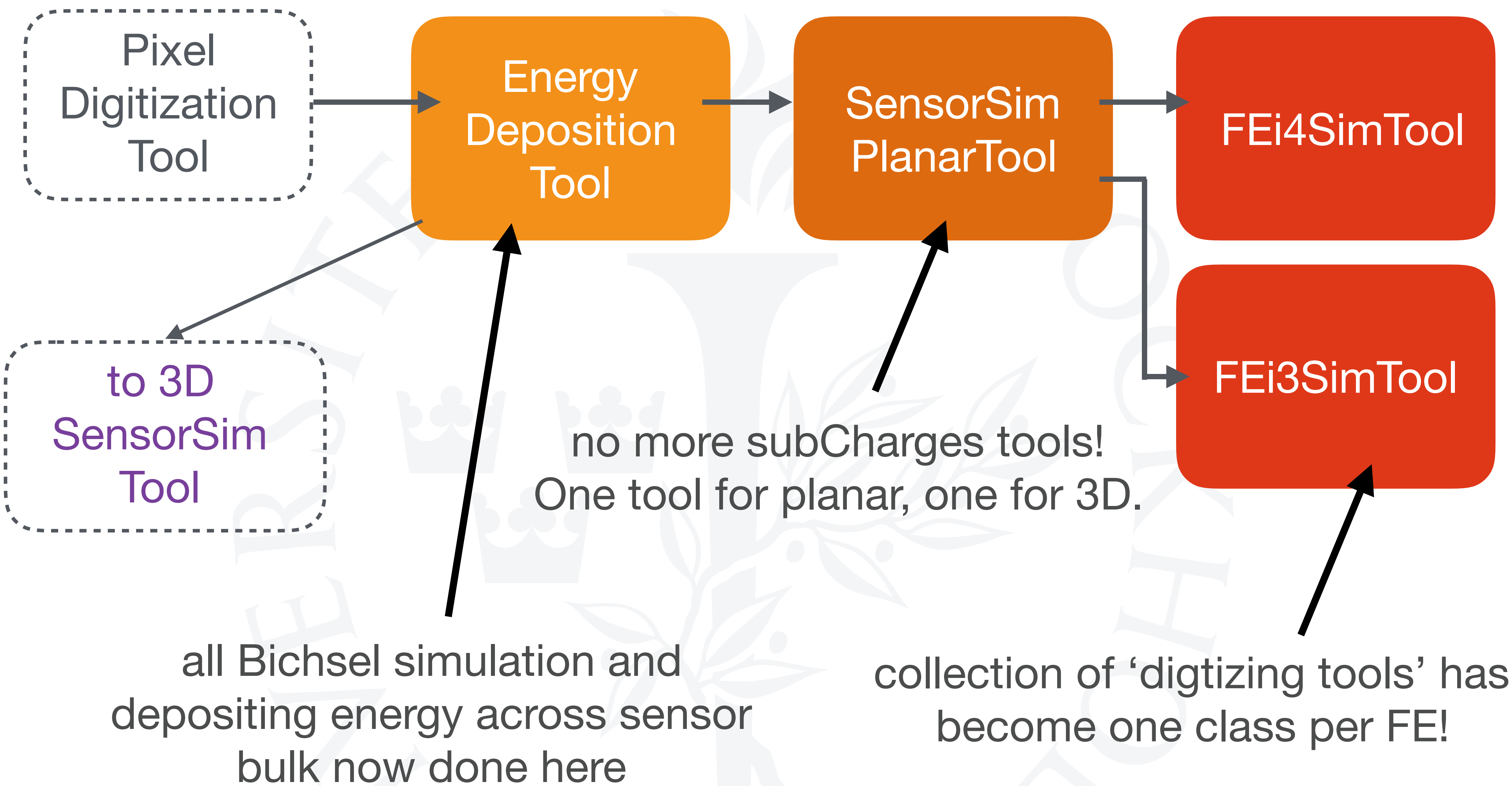
rcarney@lbl.gov

Pixel Digitization in rel. 21



Release 21 package structure

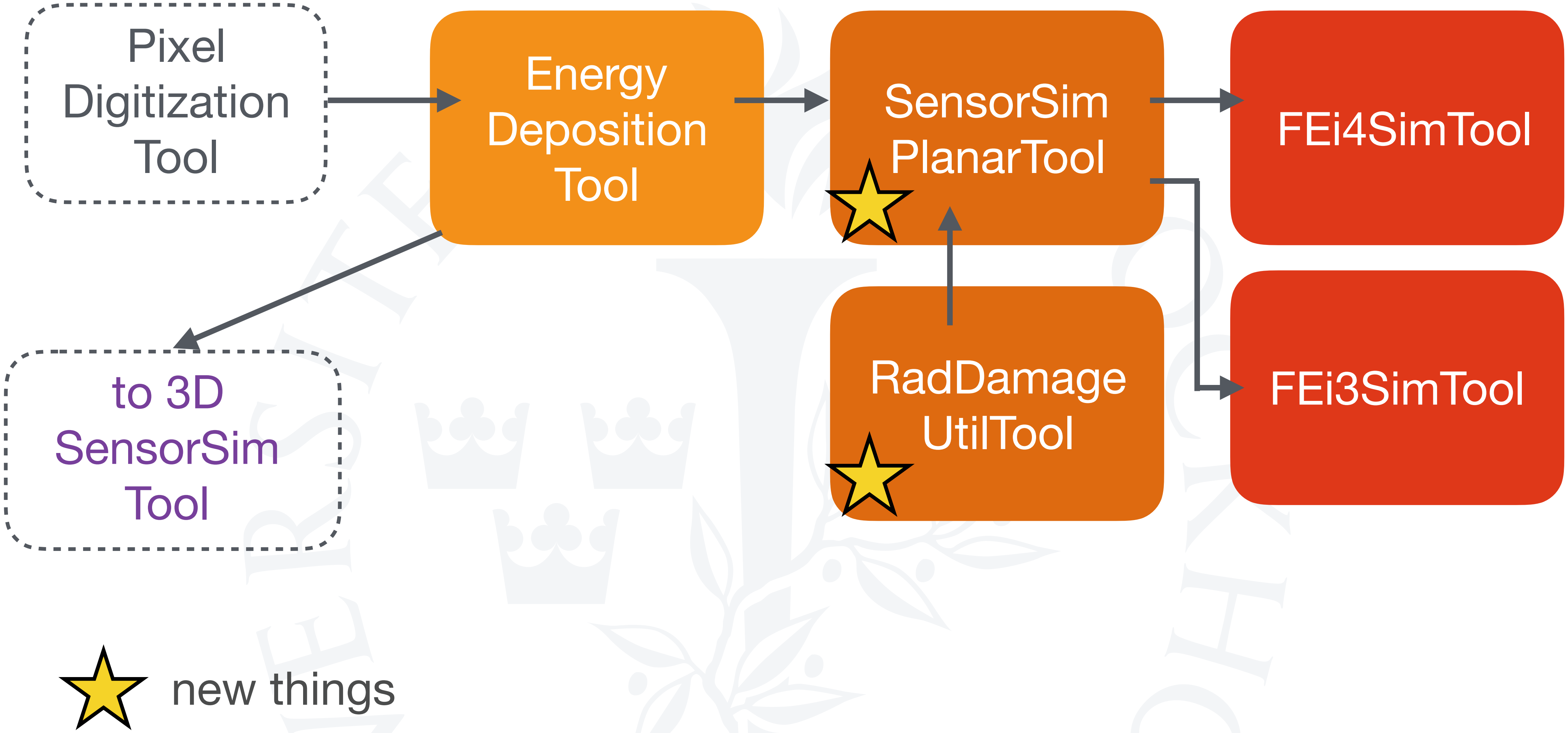
Rel. 21 and rel. 22 package differences



Release 22 package structure

ATLAS digitization package changes

Changes to incorporate radDamage

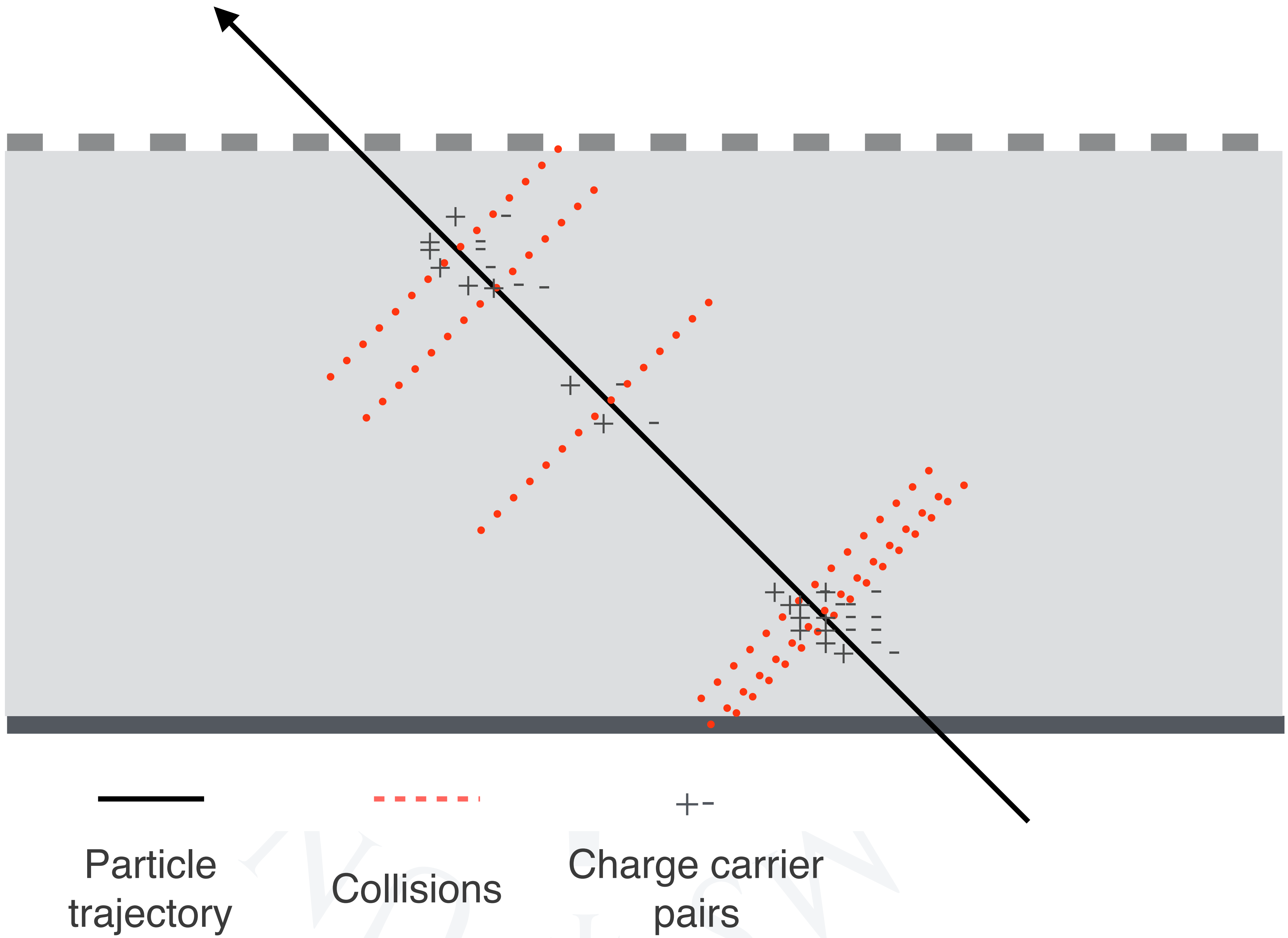


- One tool added (for neatness), sensorSim loop modified for trapping simulation
- Details of simulation strategy in paper: https://cds.cern.ch/record/2255825

Athena digi energy deposition in silicon

Energy deposition

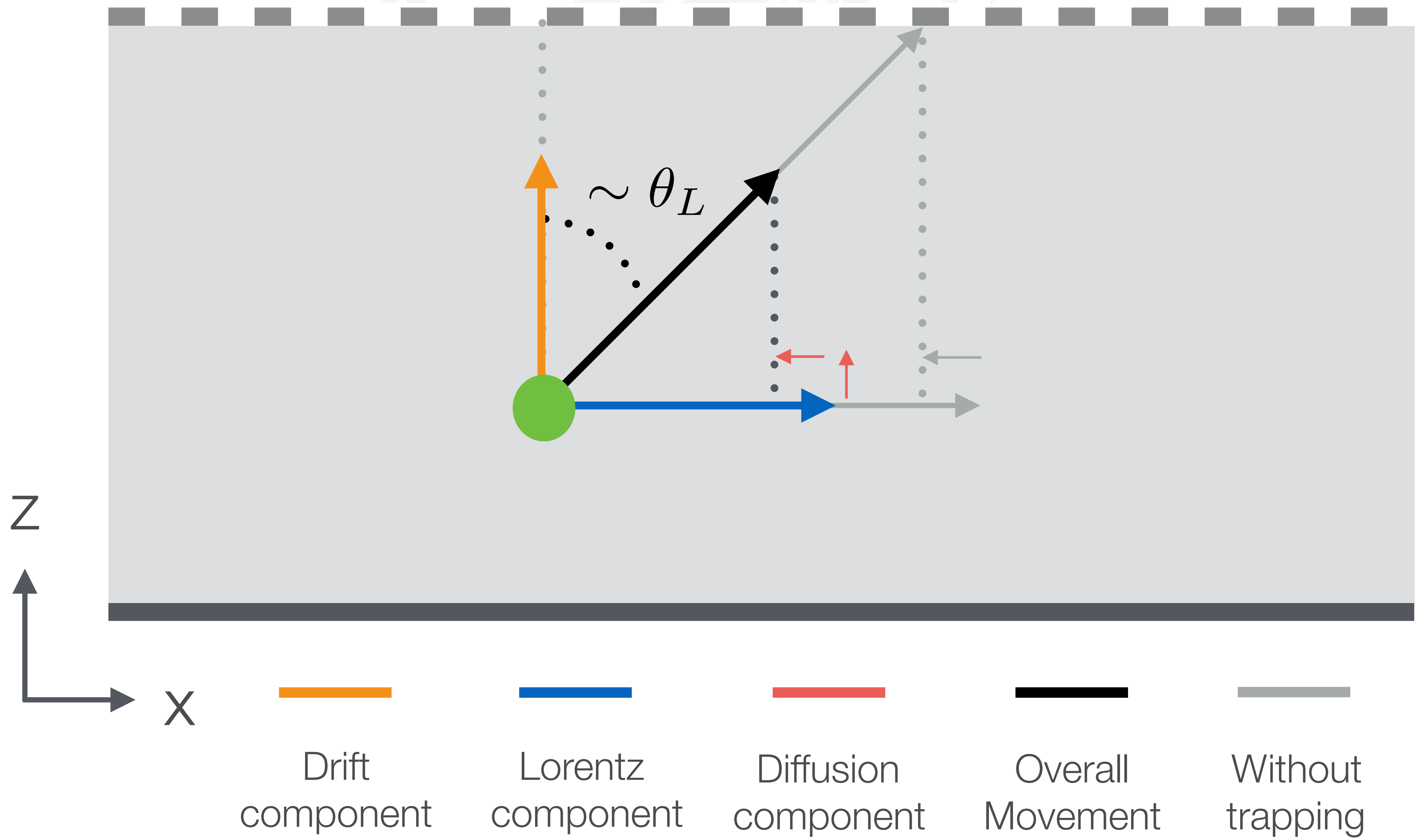
charged particle (betaGamma from G4)



Pre-existing: energy deposition, random collisions

Athena digi signal formation in pixels

(1) Is the charge trapped? (2) Where is trap? (3) What charge is induced?



New functionality: charge trapping, electric field shape

Radiation damage **simulation**

Is it trapped?

Fraction of charge carriers NOT trapped after time 't'

$$\frac{N(t)}{N(0)} = e^{-\frac{t}{\tau}} = e^{-\beta \Phi t}$$

Characteristic lifetime

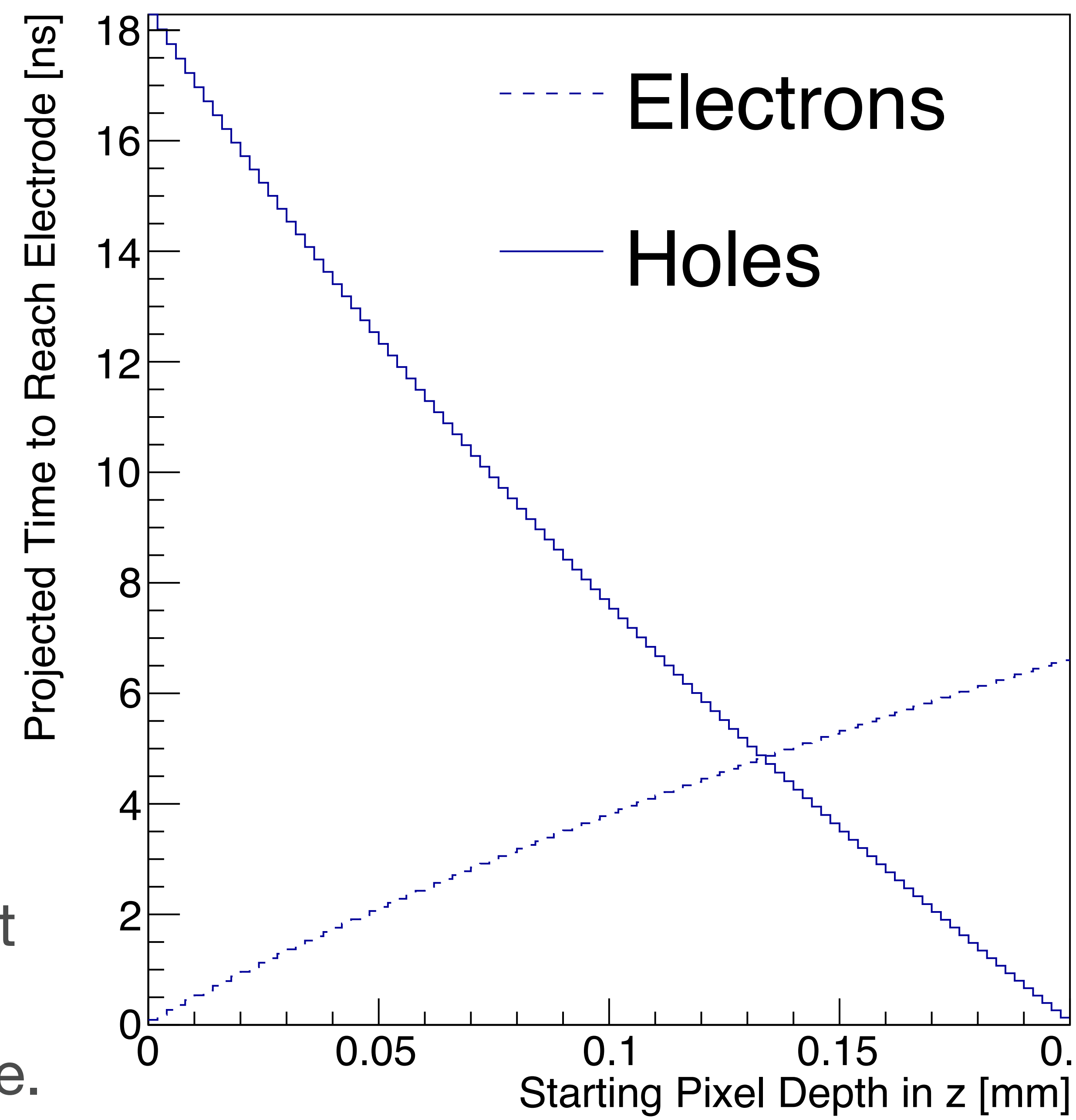
Beta is the 'damage parameter', Phi is the neq fluence

Random number thrown from a uniform distr.

$$t = -\frac{1}{\beta\Phi}\ln(u)$$

Trapping time

- t is dep. on fluence
- Throw rand. number to get trapping time.
- Compare to collection time.



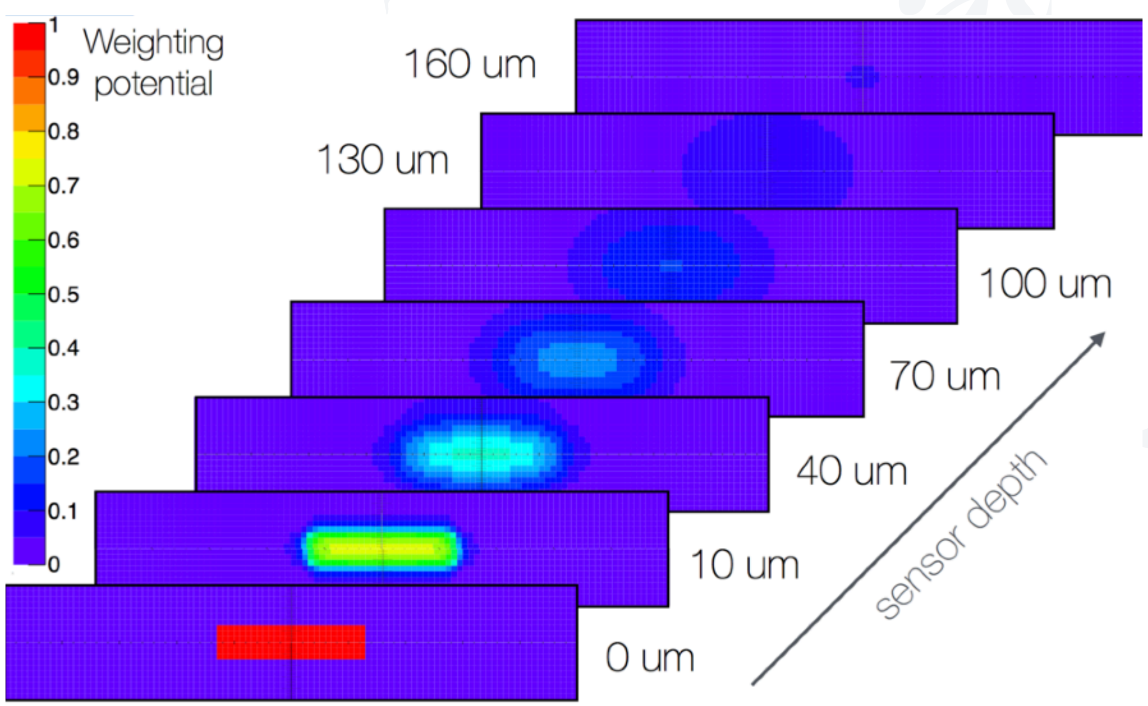
https://www.sciencedirect.com/science/article/pii/0920563293900528 (eq. 5 and 43)

Induced charge: lots of LUT

Find trap depth in bulk

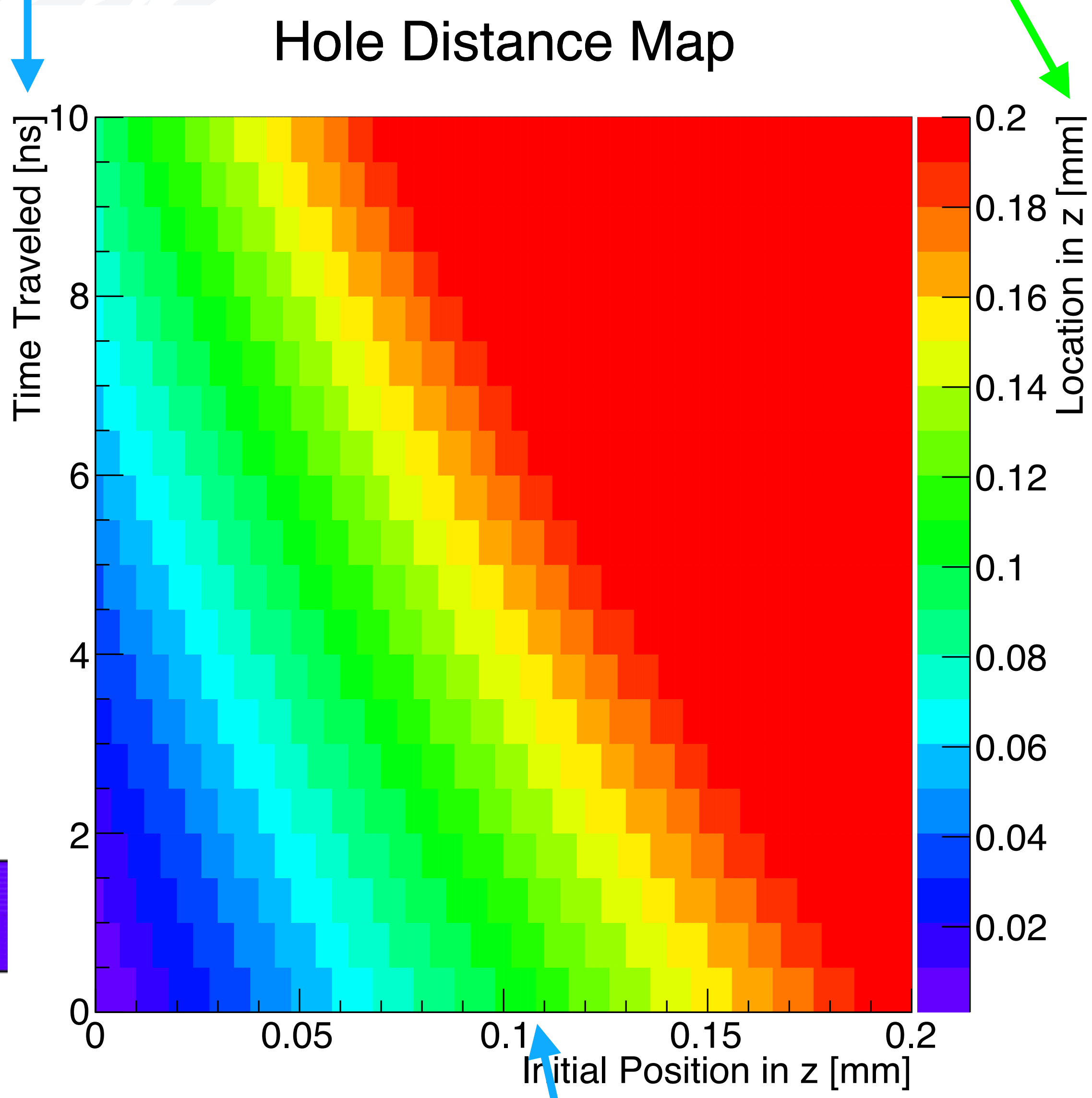
1) Look up trap position
2) Look up weight. pot. at trap position & starting position
3) Use difference to find charge induced on collecting electrode!

$$Q = q \cdot \Delta\phi(z)$$



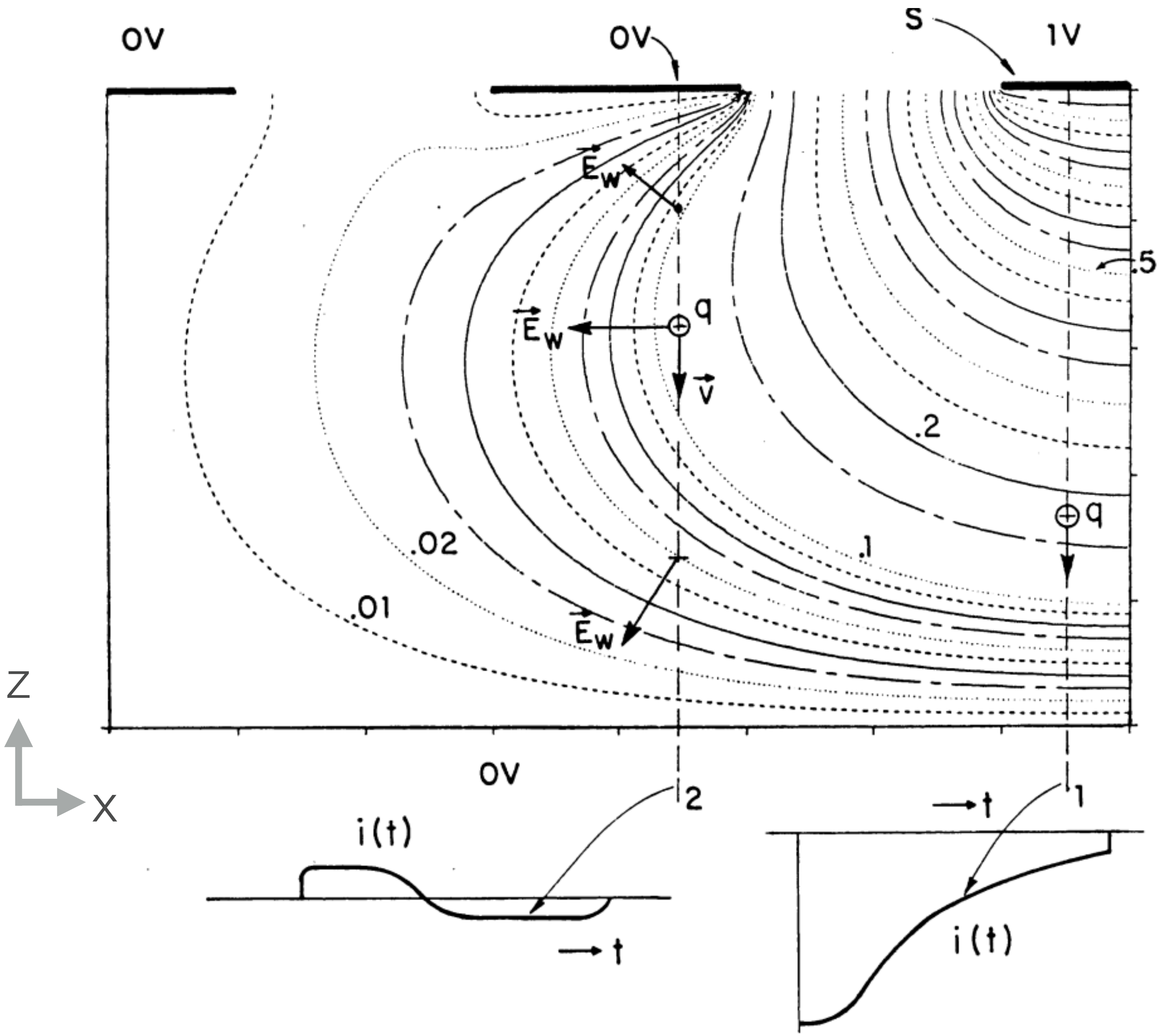
Input: time traveled till trapped

Output: trap position



Input: creation of charge carrier pair by ionization

Backup Athena sim: Neighboring pixel



Adjacent bipolar crosstalk signal

Signal at collection electrode