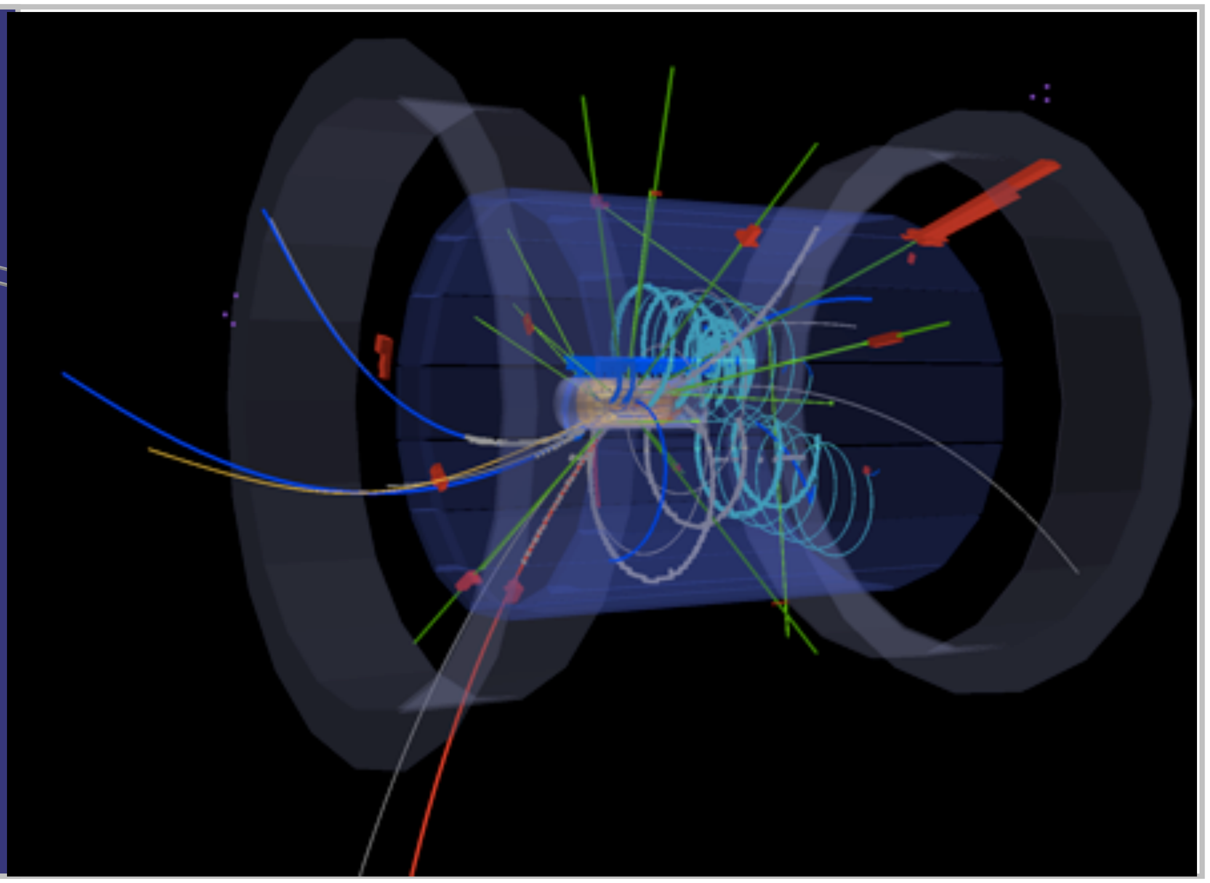
Belle II
Physics Analysis Model

Phillip Urquijo
Bonn University
CHEP, Amsterdam
October 2013



Belle II

universitätbonn

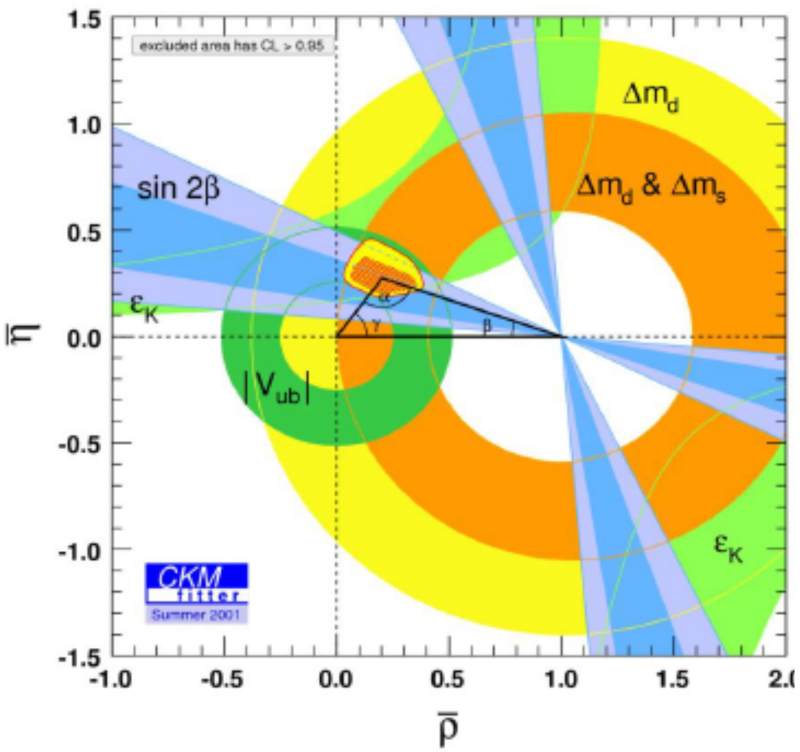
B-factories: A Success Story

- Measured **CKM** matrix **elements** and **angles** of the Unitarity Triangle (→2008 Nobel Prize)

  Many searches for **N**ew **P**henomena in flavour.

- **Possible due to unique capabilities of B factories clean events, detection of neutrals & neutrinos.**
- SM holds, precision limited → next generation expt.

EPS 2001



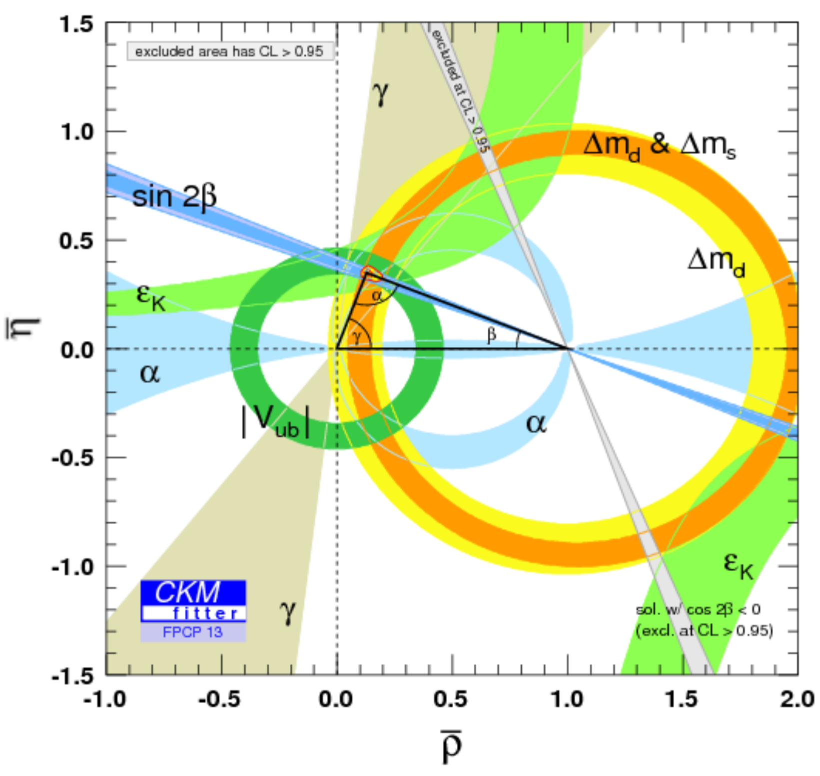
B-factories: A Success Story

- Measured **CKM** matrix **elements** and **angles** of the Unitarity Triangle (→2008 Nobel Prize)

  Many searches for **N**ew **P**henomena in flavour.

- **Possible due to unique capabilities of B factories clean events, detection of neutrals & neutrinos.**
- SM holds, precision limited → next generation expt.

FPCP2013

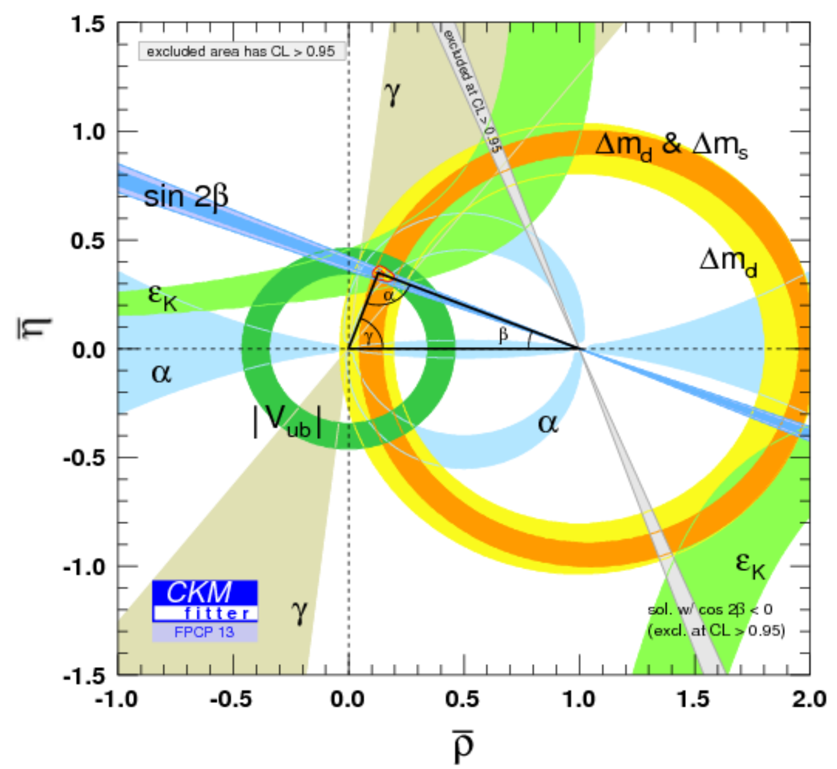


B-factories: A Success Story

- Measured **CKM** matrix **elements** and **angles** of the Unitarity Triangle (→2008 Nobel Prize)

  Many searches for **N**ew **P**henomena in flavour.

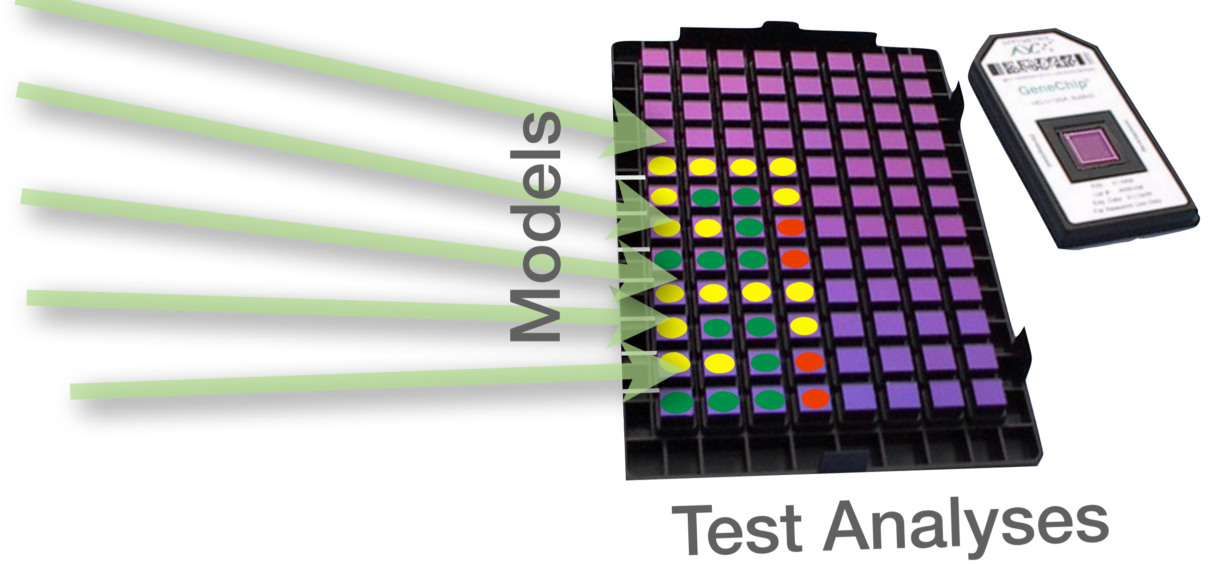
FPCP2013



- **Possible due to unique capabilities of B factories clean events, detection of neutrals & neutrinos.**
- SM holds, precision limited → next generation expt.

- Belle II's new regime of NP searches
  - E_{miss} :$B\to\tau/\mu\nu$, $B\to X\tau\nu$, $B\to h\nu\nu$
  - Radiative: $B\to X_{s,d}\gamma$, $B\to X_s ll$
  - CPV/Neutrals: $B\to K_s\,\pi^0\,\gamma$, $B_s\to\gamma\gamma$
  - Tau LFV
  - Quarkonium (minimal trigger impact)
  - Charm ...

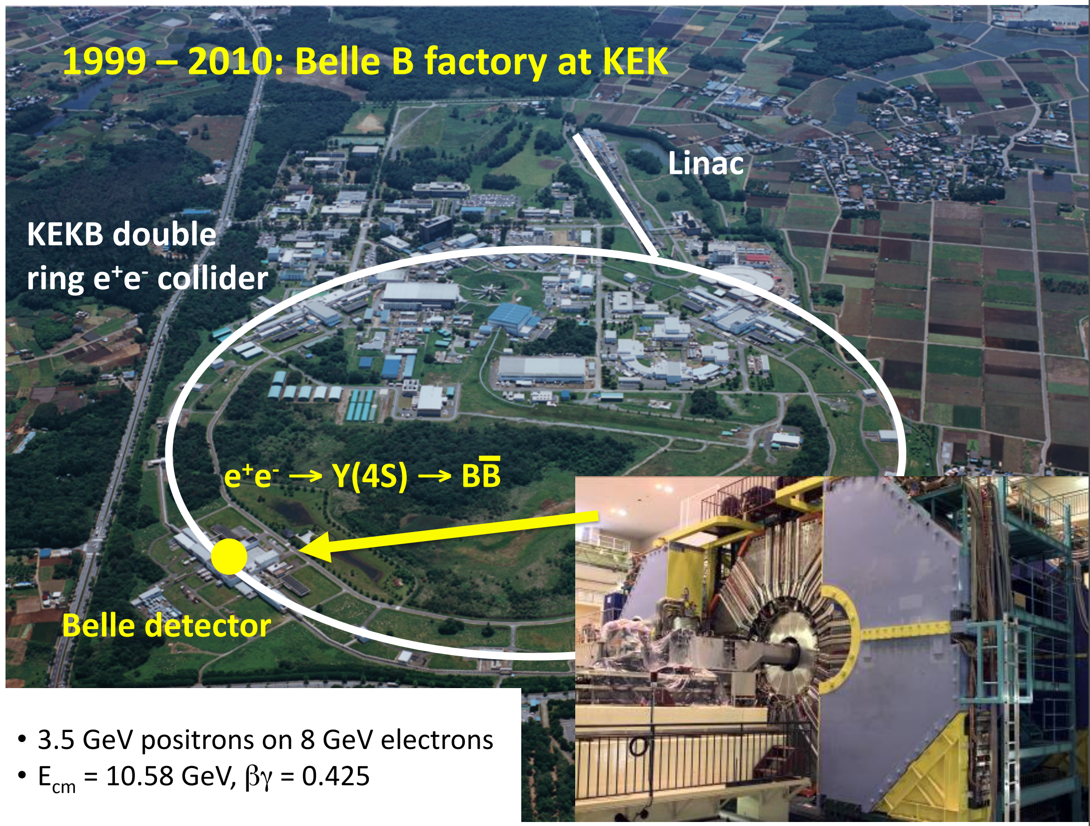
New Physics "DNA Chip"

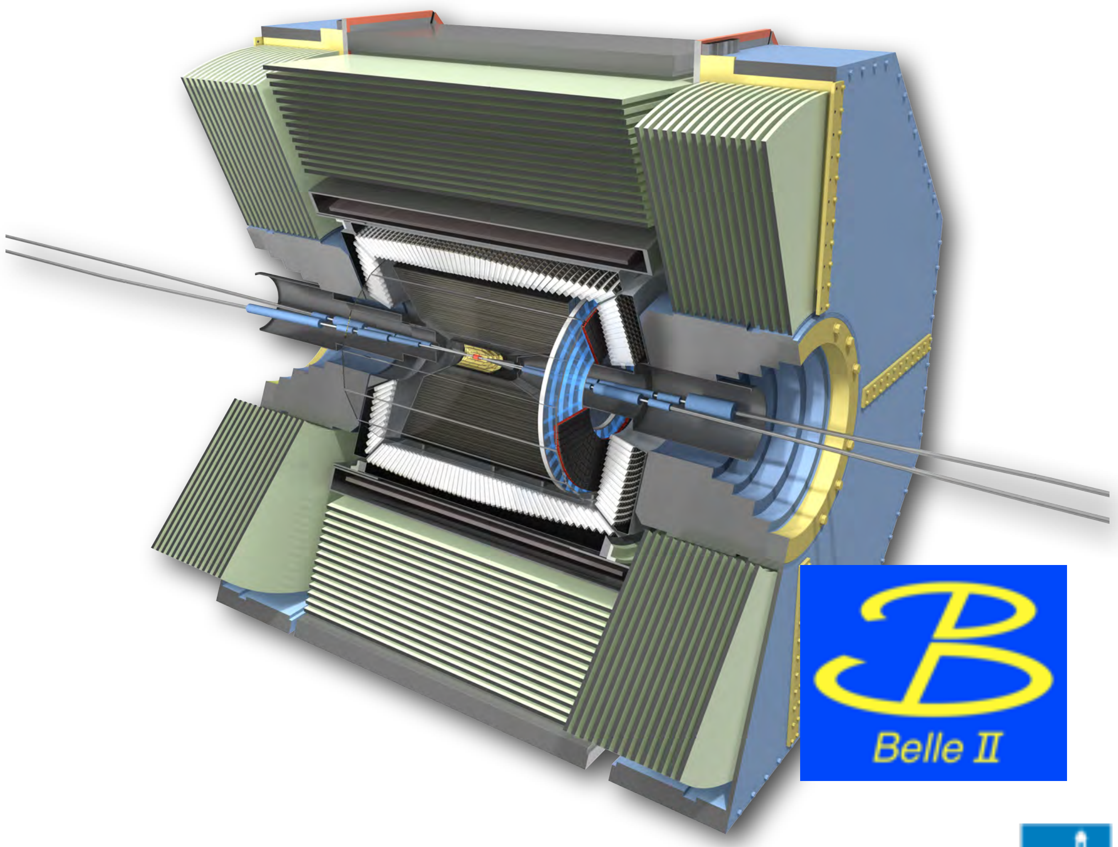


Belle→Belle II

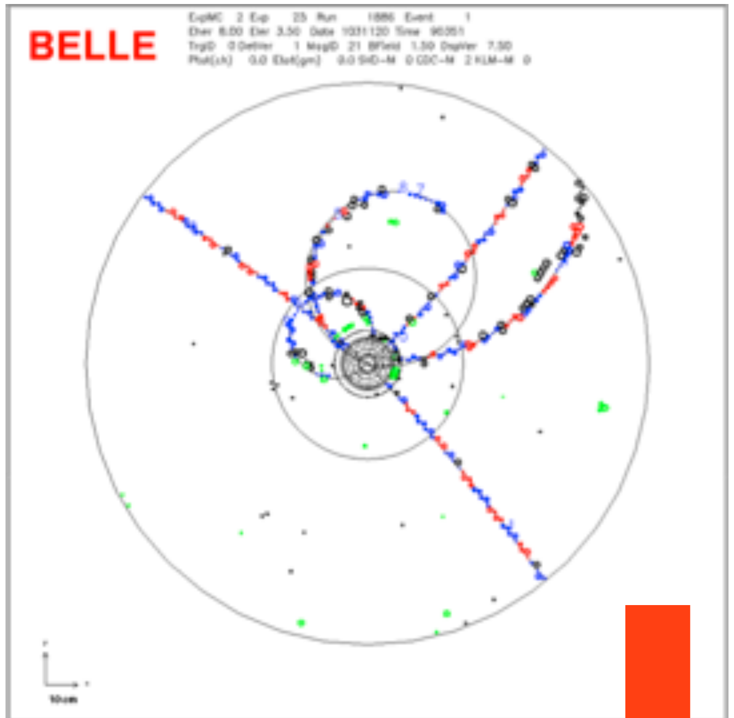
- **Belle** experiment @ KEKB e^+e^- collider
  1ab^{-1} data, mostly Υ(4S) (1999-2010)
  - B, D, τ ++.

- **Belle II** experiment **@ SuperKEKB (online in 2015)** [500+ collaborators, 96 institutes]
  - 40x collision rate: L=8 · **10^{35} cm^{-2}s^{-1}**
    50x data sample: 10 ab-1/year
  - New detector technology throughout.
    - Pixel detector alone: 2 GB/s!





Belle→Belle II Challenges



Higher background (×10-20)

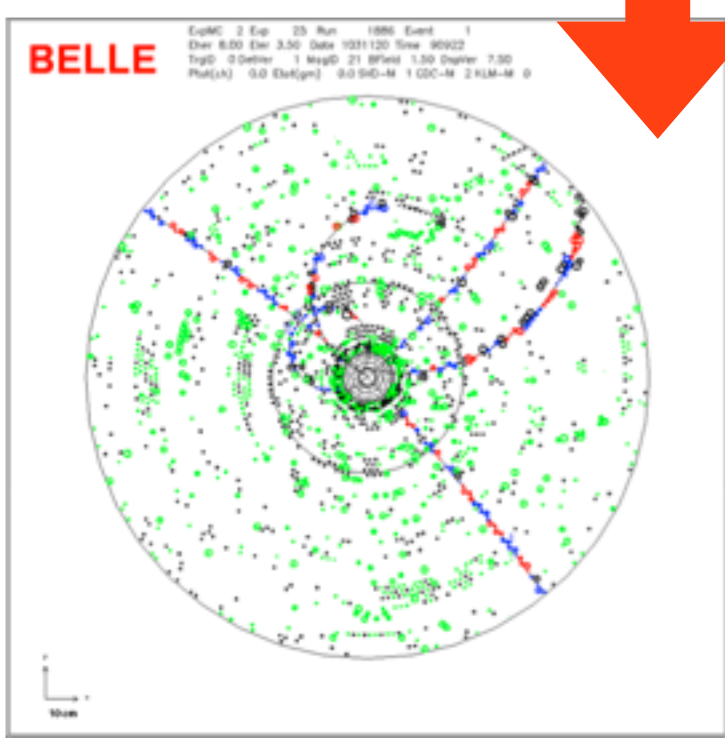
Occupancy, fake hits and pile-up noise

Higher event rate (×10)

Trigger, DAQ and computing

Special features

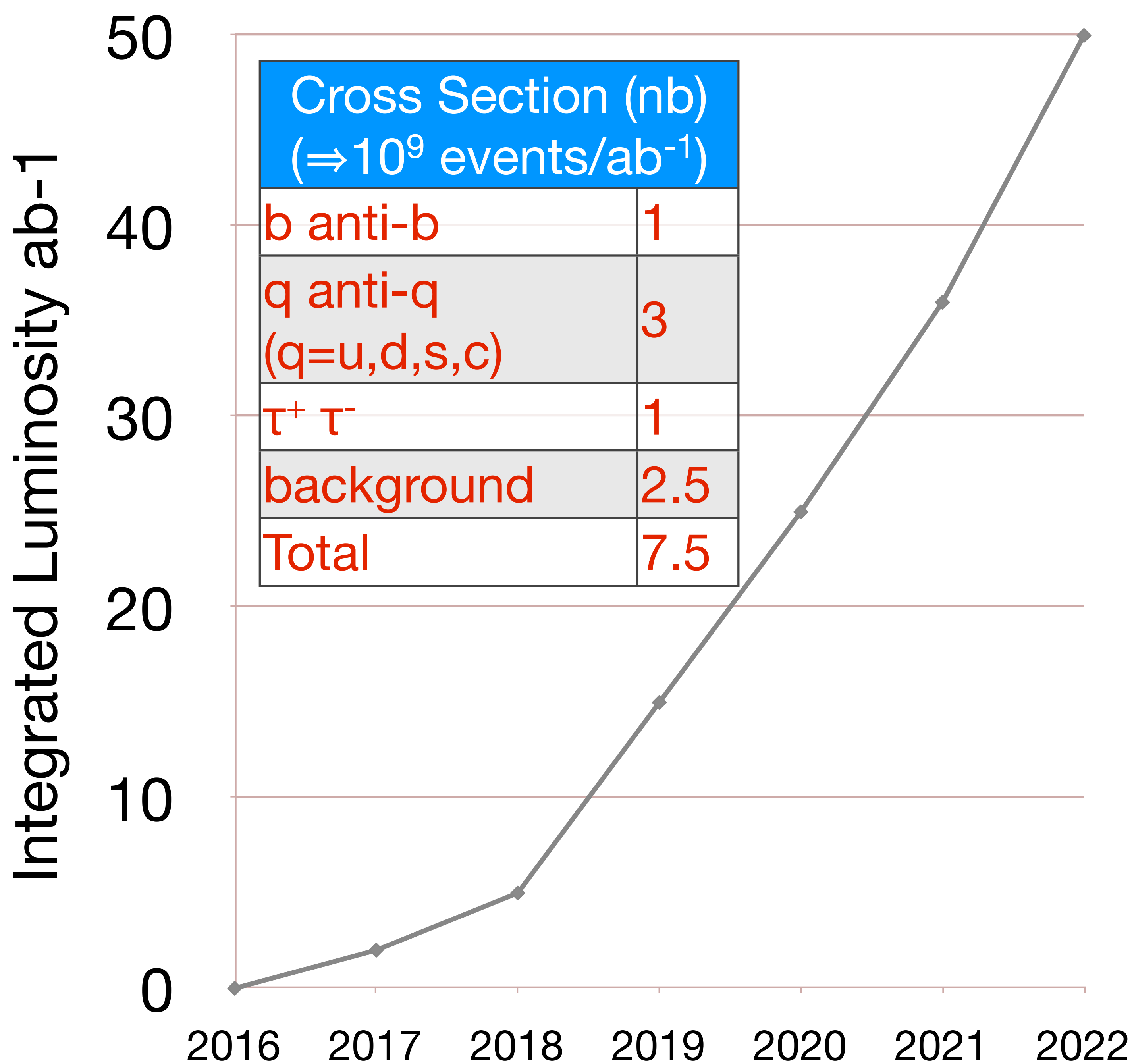
Better at low $\mathbf{p_T}$, & hermeticity



Rising to the challenge

- Pixel detector.
- Better particle ID
- Faster readout
- New computing system
- **Entirely new analysis model.**

Data Volume



Cross Section (nb) ($\Rightarrow 10^9$ events/ab^{-1})	
b anti-b	1
q anti-q (q=u,d,s,c)	3
$\tau^+ \tau^-$	1
background	2.5
Total	7.5

Belle II / 50 ab^{-1}		
Events	**3.8 x 10^{11}**	
Raw/DST Storage	Raw data (pB)	200
mDST storage	mDST data (pB)	8.7
	mDST MC (pB)	42
User/ Skims	**1. Group μDST** **2. User μDST** **3. nTuples**	20**

Belle II TDR, KEK Report 2010-1, arXiv:1011.0352

***Under development*

Belle II Software Requirements

- **Key Reconstruction algorithms:**
  - Low p_T tracking (~100 MeV) & μID
  - Precise(calibrated), low noise Calorimeter response
  - Precise vertexing, flavour tagging
  - Full reconstruction of B and D mesons
  - Beam background suppression (10-20x)

- **Changes to the physicist's interface:**
  Paradigm shift to centralised processing (faster/precise)
  - Grid based model
  - Persistent analysis object class "layer"
  - Centralised "plug-in" analysis tools (minimise user tools on Grid)
    - Greater precision & measurement repeatability
  - Centrally managed signal MC w/provenance.

Reconstruction Performance

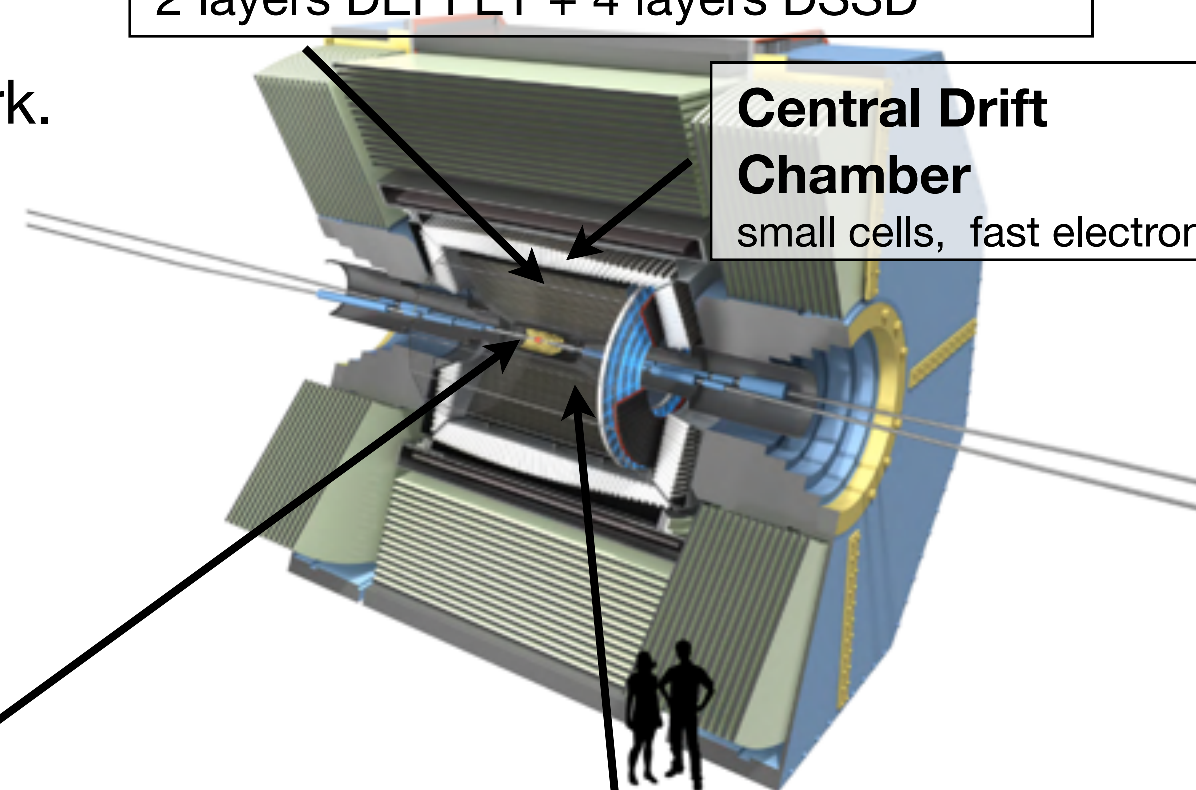
Further details in backup

- **Tracking (under development)**
  - **General: D**eterministic **A**nnealing **F**ilter,
  - **Si** only: Cellular automaton / Hopfield network.

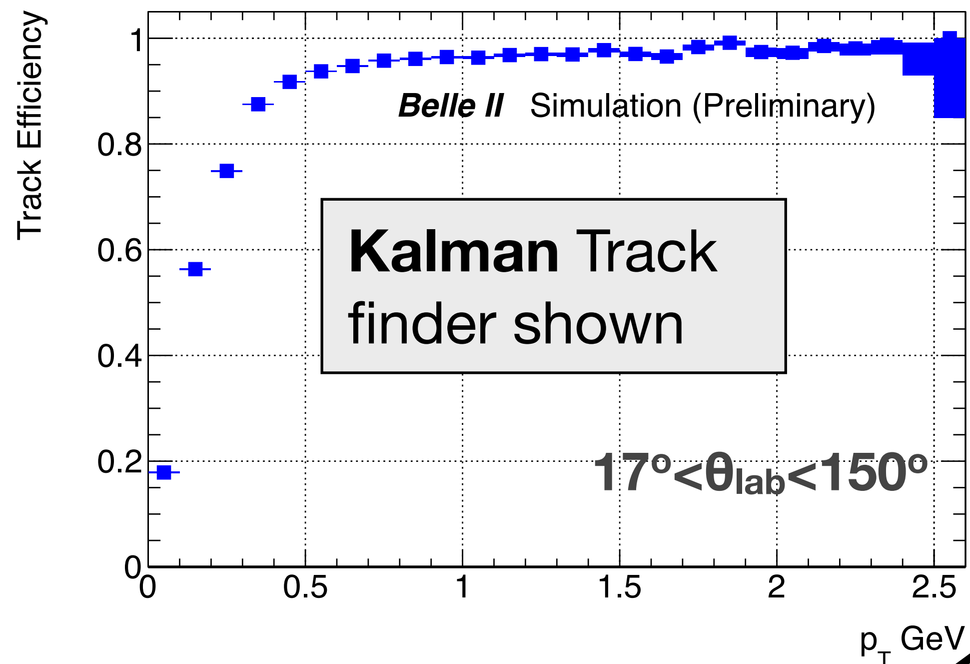
J Lettenbichler *et al* 2012 *J. Phys.: Conf. Ser.* 396 022030

Vertex Detector
2 layers DEPFET + 4 layers DSSD

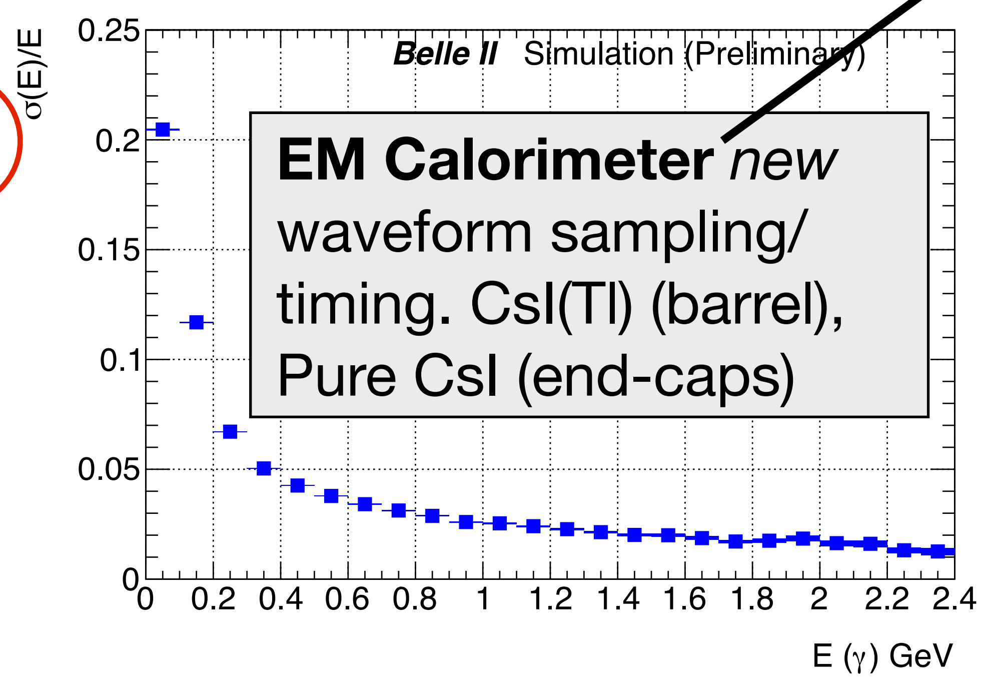
Central Drift Chamber
small cells, fast electronics



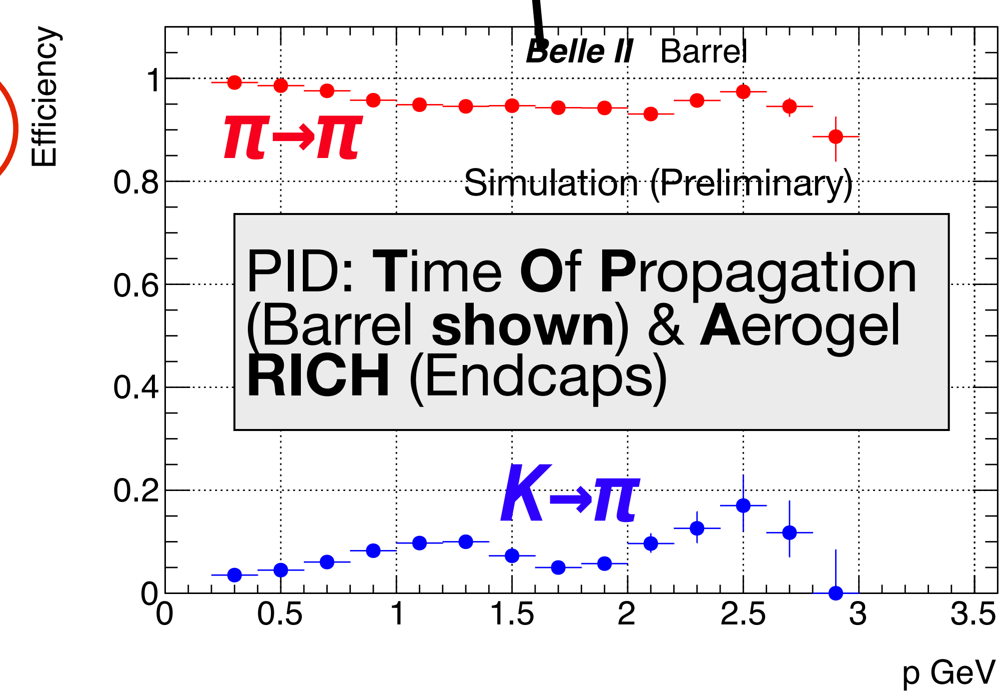
1



2

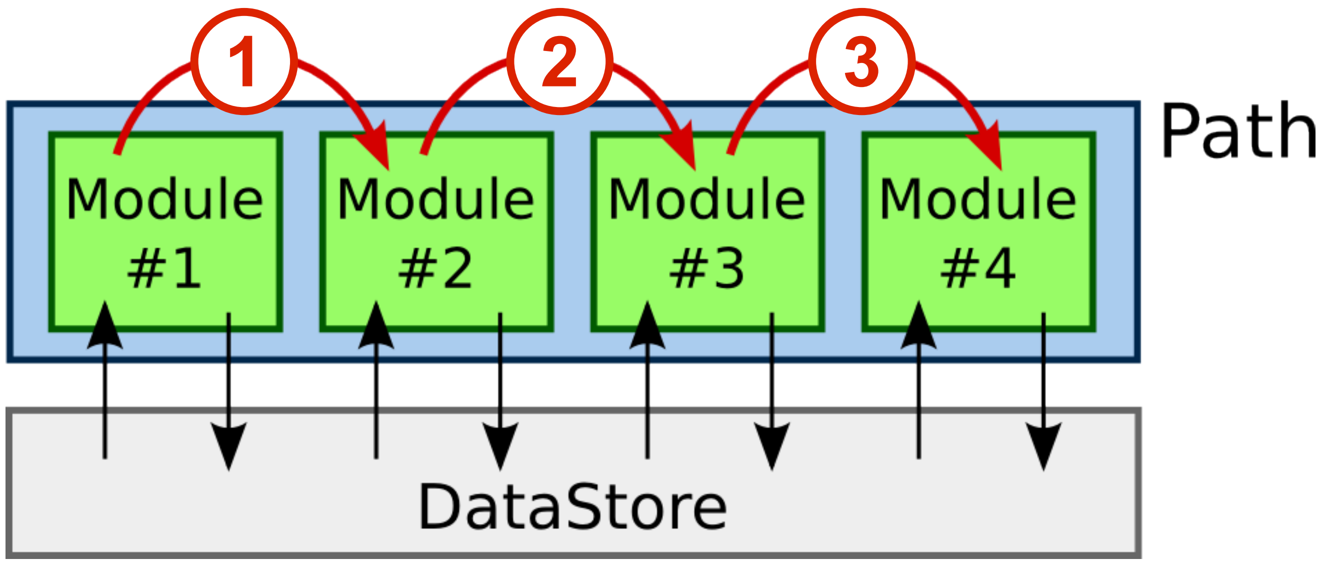


3



Belle II Analysis Software Framework (BASF2)

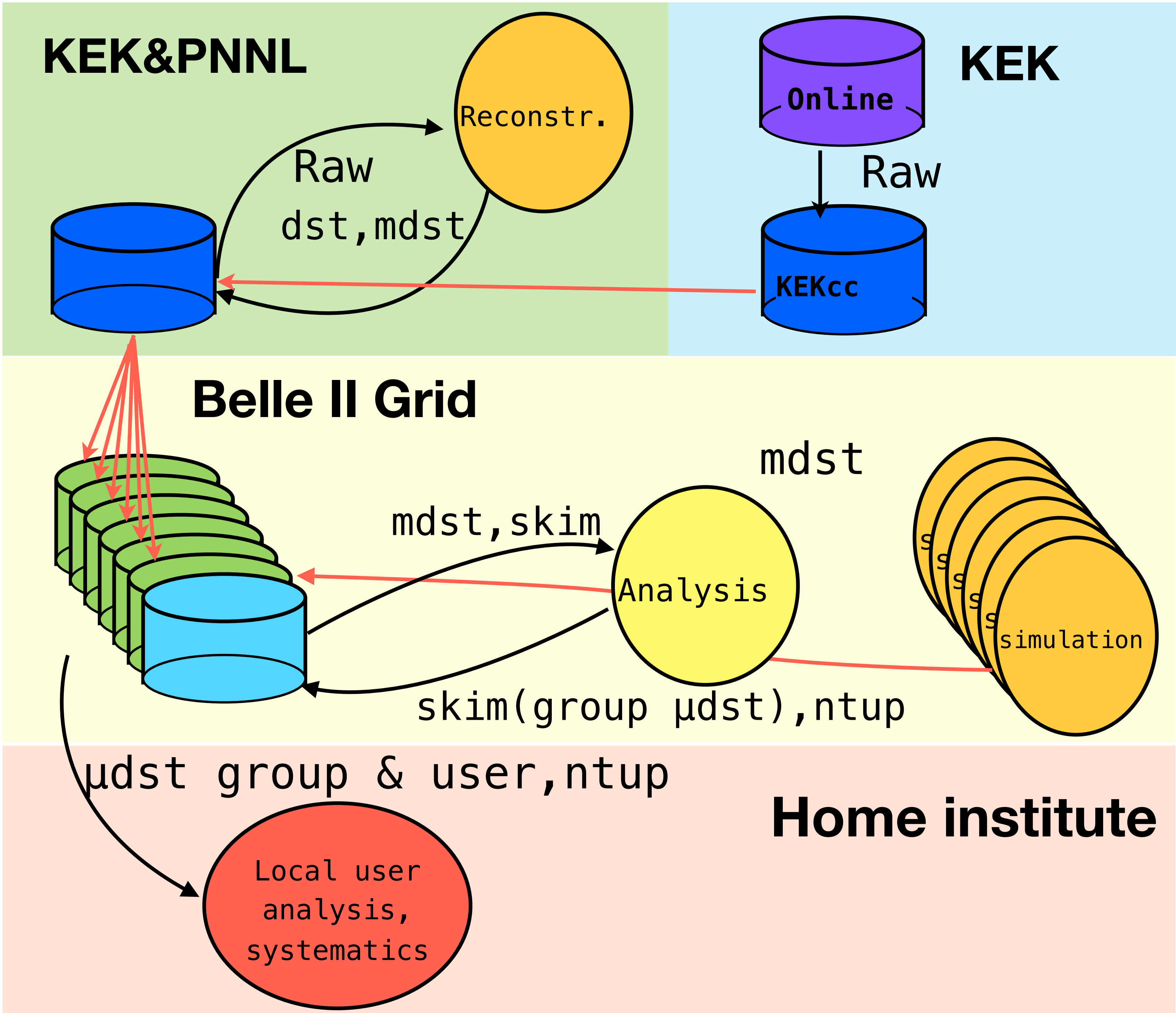
- All algorithms are modules (base class), minimal core overhead.
- `DataStore` manages all data loaded or created during event processing
- Can store any class that inherits from `TObject` and has a ROOT dictionary



- *Where possible keep external tools (Evtgen, GenFit, DAF) as externals.*
- Analyses use compiled C++ algorithms with Python settings
- Data can be grid provenance tracked.

Analysis Flow

Analysis flow (Physics Trigger → nTuple)



- *Very* many physics channels.
  - from low(rare) stats to high(precision) .
- **See || talk by T. Kuhr on Belle II Grid & index files**

Data model

Analysis Flow

Detector response: Hits, clusters, timing, trigger	Tracks, Photons, PID, Summaries	Particle candidates & Event topology
400 kB	*40 kB*	*few kB*

RAW/DST
- ARICH
- CDC
- ECL
- Tracks
- PXD
- SVD
- TOP
- Trigger

Pixel (PXD) data reduction applied.

"reconstruction"

mDST
- Track
- PID (TOP, ARICH)
- ECL (eID, π0, Gamma, Shower)
- dE/dx
- KLM (KL, Mu)
- Vertex (Ks, Lambda)
- Trigger

"skim"

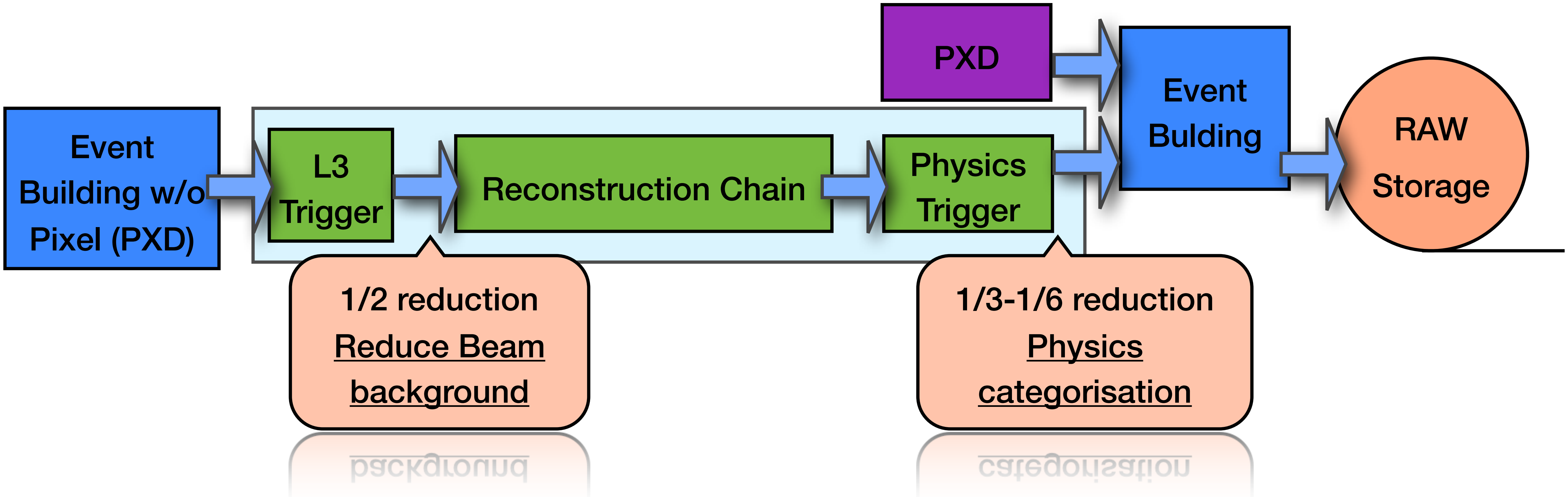
µDST (+ mDST)
- Particle(Lists)
- *PID likelihood*
- Vertex
- Topology (Continuum suppression)
- Tag (B,D,CP)
- RestOfEvent

µDST special options
- systematics mdst (trk/ECL)
- calibration dst (hits)

new to Belle II

Physics Trigger

- Applied at end of online reconstruction
  - Calibration constants not tuned, PXD not available.



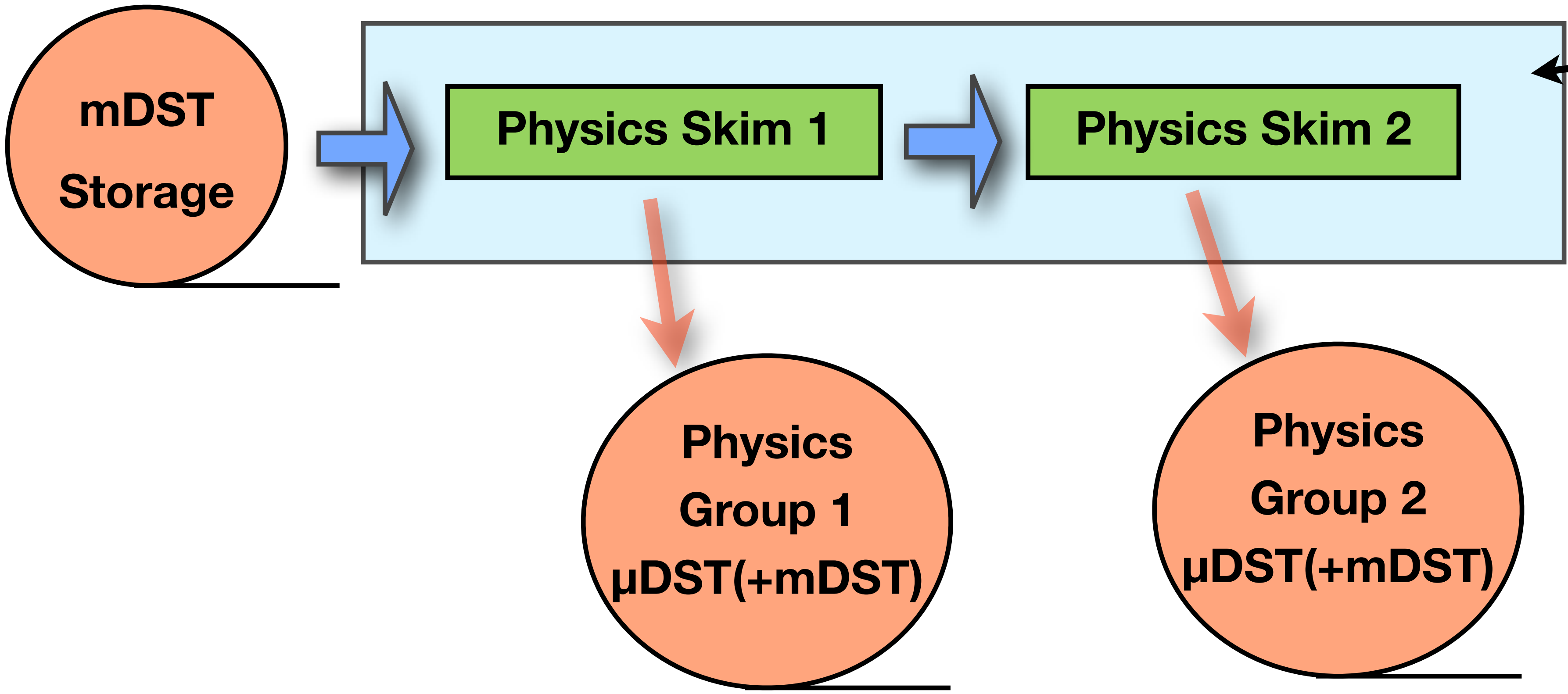
- ***Based on offline physics analysis model & steering mechanism.***

Trigger	Prescaled?
(rad.) Bhabha	Y
μ-pairs	Y
τ-pairs	N
2-photons	Y
Hadronic (B,D)	N
Salvage	Y

+ Finer Categorisation

Physics Skim

- Selection of interesting events, **after** `mDST` reconstruction.
  - 2 data formats: **`index`** (event pointers), and **derived data** `μDST(w/mDST)`.
  - Centralised reconstruction of **Particle** candidates



1. **Group μDST** production, with overlapping analyses,
2. **User μDST** Reduced from **group μDST** for local user analysis (stored locally).

- Python selection archived between metadata & BelleII repository!
- ***Content delivery: μDST subscription or distribution service is to be determined.***

Estimate of resources

Trigger Skims	pB/ab^{-1}	Skim Group (Grid)	pB/ab^{-1}	User Physics Skims (Local)
Hadron Trigger	1	B_{tag}	0.07	V_{ub}
				$B \to \tau/\mu/e\ \nu$
				$B \to D^{*}\tau\ \nu$
		Radiative	0.04	$B \to X_{s,d}\ \gamma$
				$B \to K^{*}/\rho\ \gamma$
		J/ψ	0.02	Charmonium
				B CPV
		Charm (D,D_s,Λ_c)	0.07	D Mixing
				D CPV
		>50× Exclusive	0.20	Many
Tau, Bhabha, 2 Photons	0.1	Various	0.05	Many
Systematics	0.1	Systematics	0.05	Many
		Calibration	0.05	
Total	**1.2**	**0.5 (0.2-0.4 PB per skim group)**		**0.2 pB (~3TB/analysis)**

*Exclusives skims will be further grouped
*Skim optimisation tools prepared

Estimates include MC (average of 2.5 streams)

Analysis Flow

Physics Analysis Framework

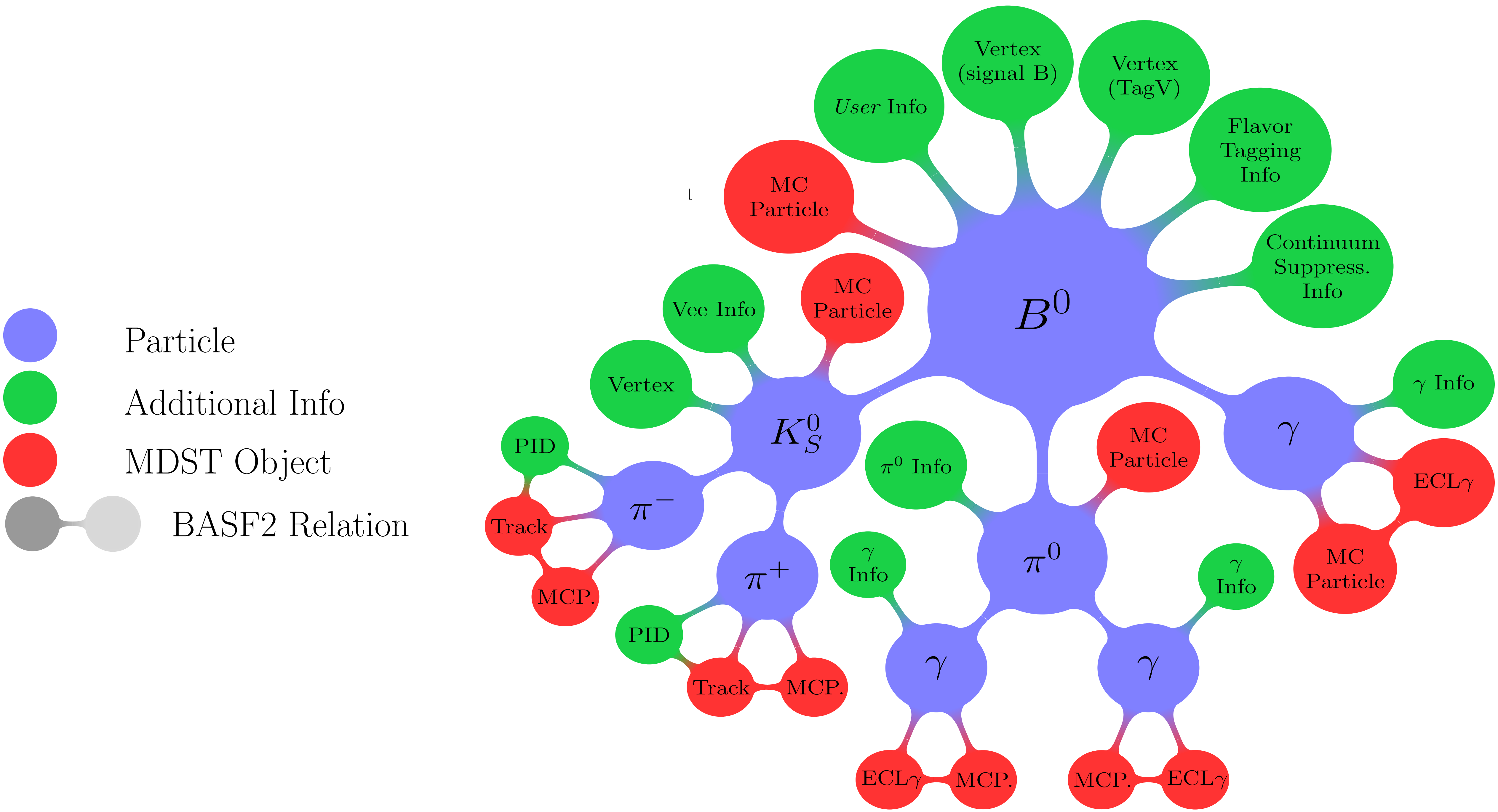
Concurrent processing of >100 analyses

Belle II

Physics Analysis Framework

3 main components

- ***Analysis* data types:** that extend *Reconstruction* data types.
- **Tools** for selection, combinations, tagging, vertexing etc.
- Preparation of **"derived" data formats**, and automatic **ntuples**.



Belle II Particles

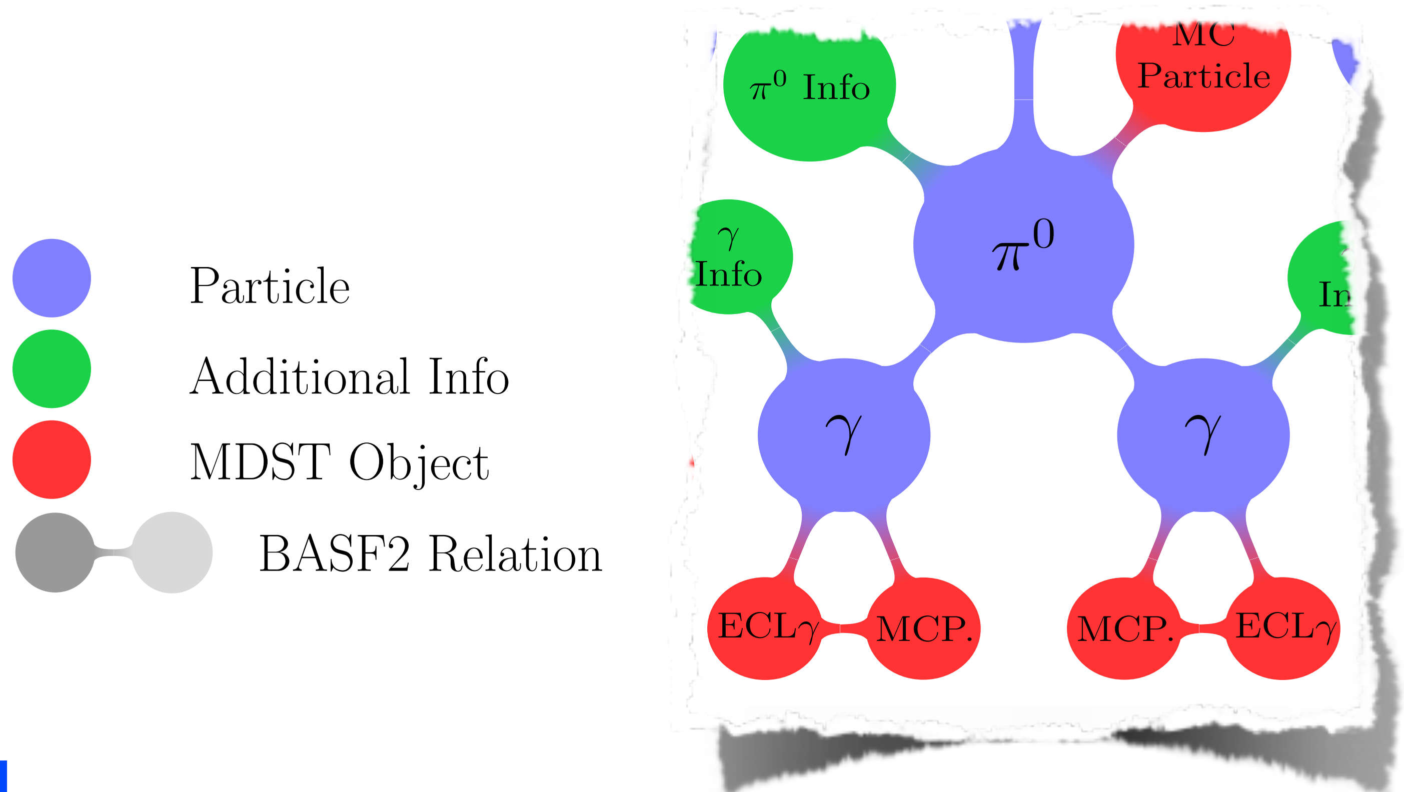
Analysts' "Interface".

- **Persistent.**
- Collected in **ParticleLists**,

  final state: K, π, e ...

  composite: D→Kπ ...

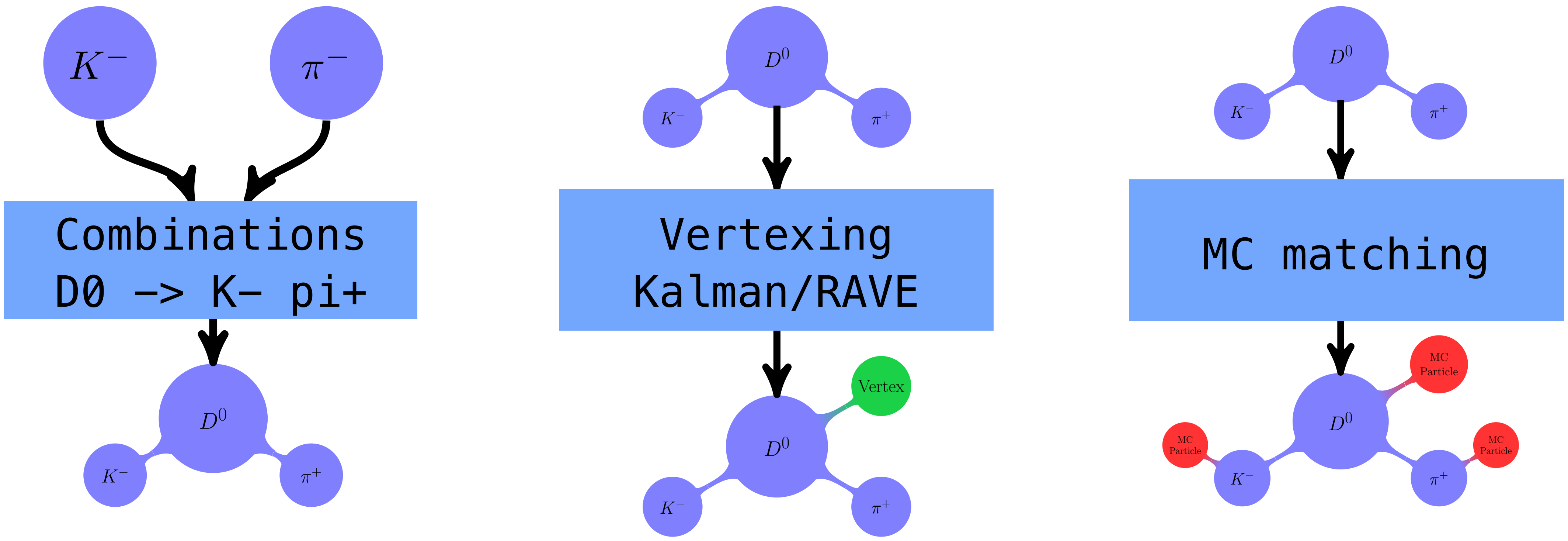
  Predefined lists used for skimming & concurrent analyses.



Particle Object

- PDG ID & flavour content
- Mass & p vectors
- Position & error
- Particle type (Track, Gamma, pi0, composite)
- mDST index (to Track/ Cluster)
- PID info "relation"
- Daughter "relation"

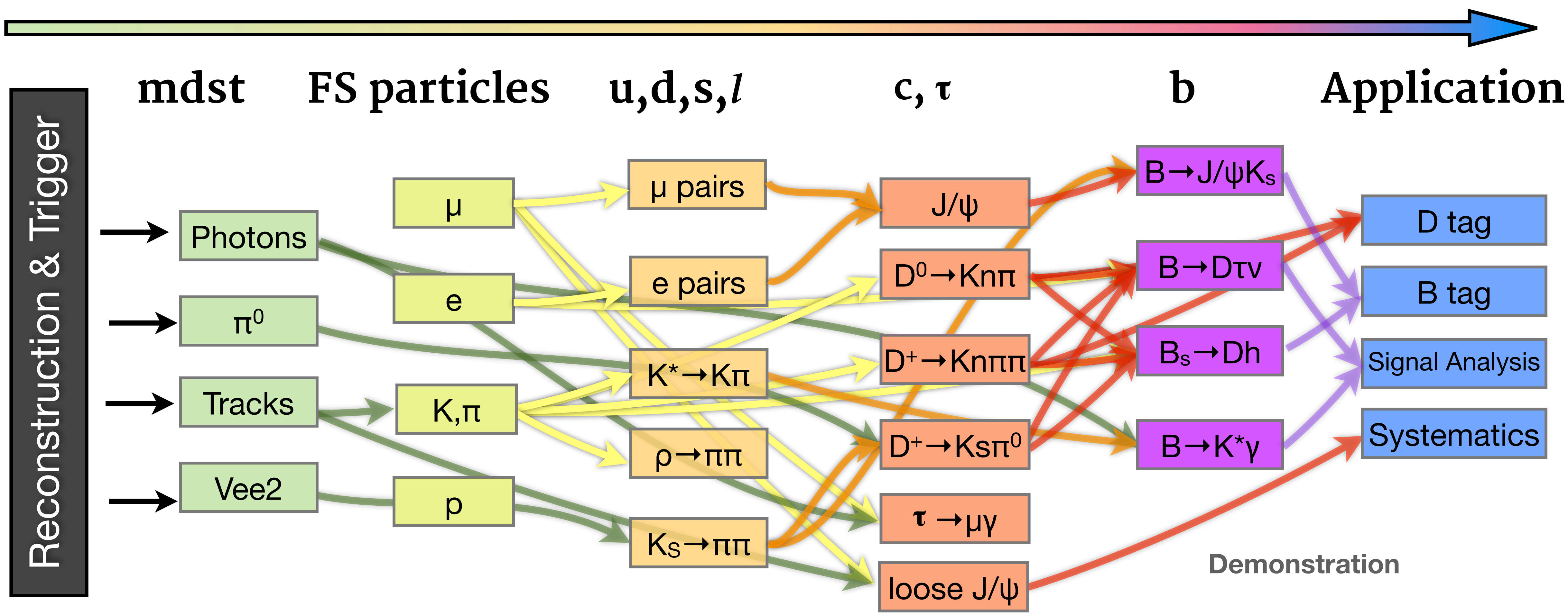
Algorithms & Objects



‣ `Combiners`	Particle combination.
‣ `MC-Matching`	Non-trivial MC matching.
‣ `Vertex`	Suite of fitting methods.
‣ `Topology`	Continuum (*q anti-q*) suppression
‣ `Tag`	B, D tag candidates with MVA classifiers
‣ `Rest of Event`	Extra tracks, looping tracks, photons

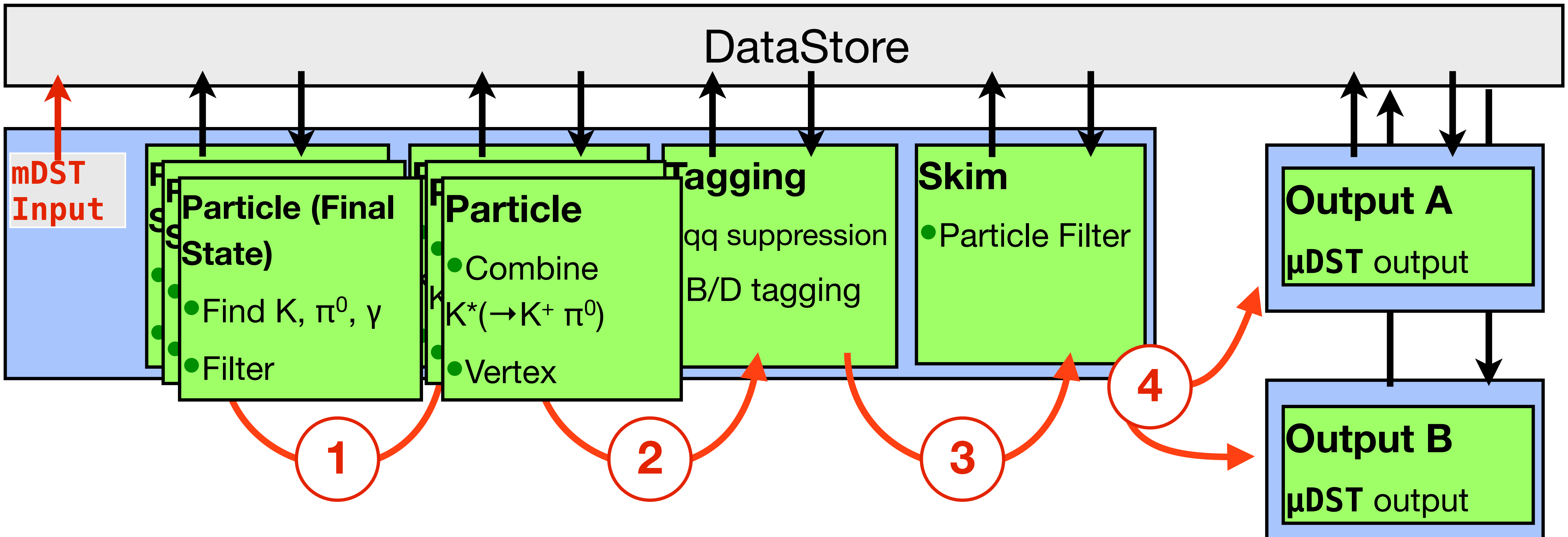
(Persistent) Particle Layer Approach

- Efficient *concurrent processing of* many analyses.
  - “**Particle Zoo Map**” done in Grid “production” mode
  - Reconstruct layers: same intermediate candidates in many modes.

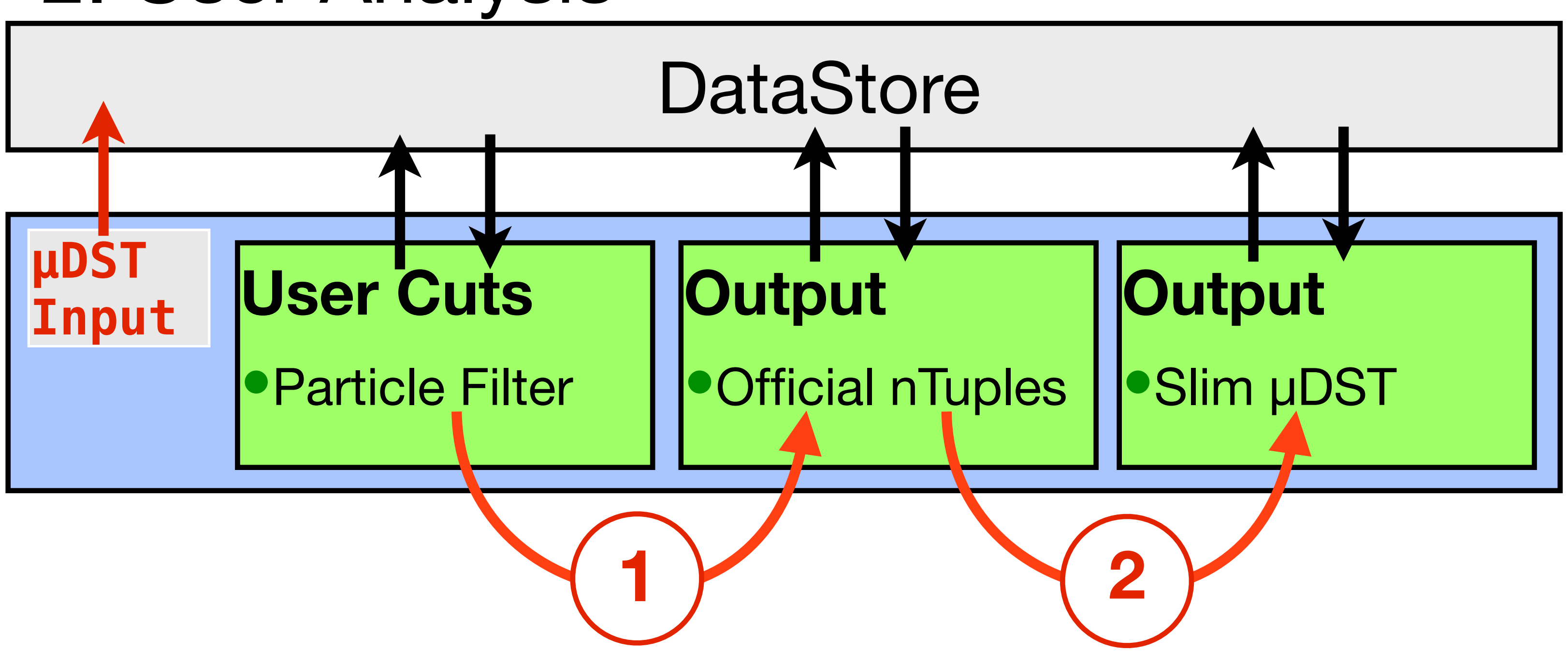


Particle Analysis Sequence

1. Grid based "production"



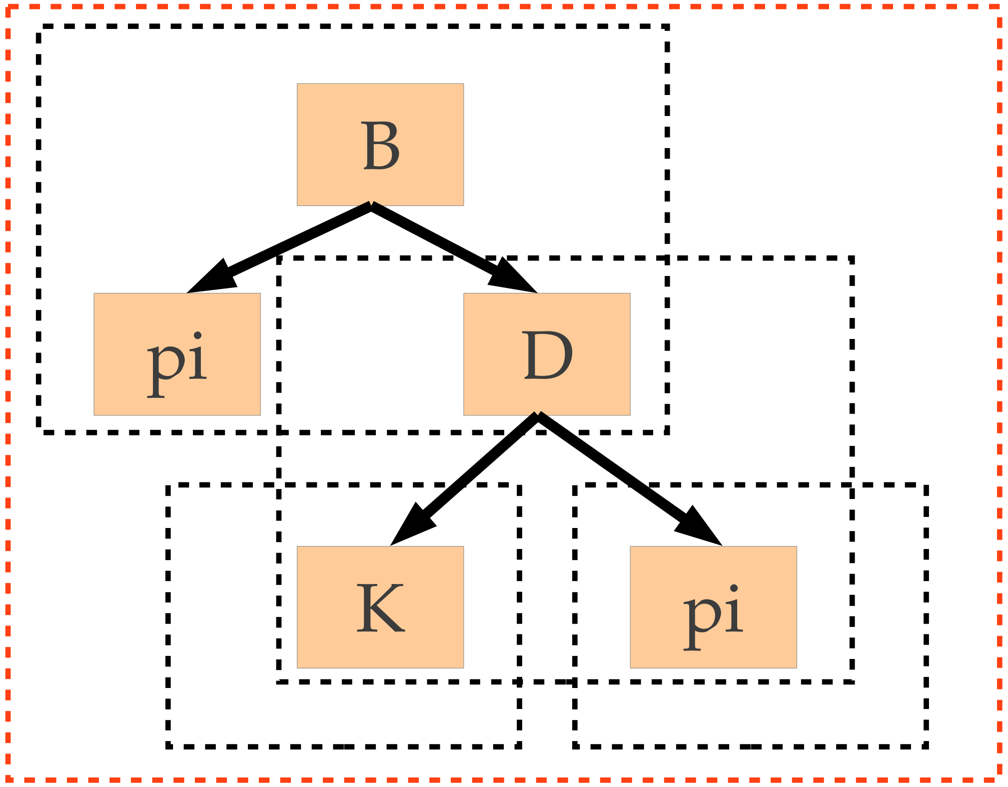
2. User Analysis



- **Persistent Particle approach reduces post-skim analysis time *e.g.* $D^0 \to K\pi$, $K\pi\pi\pi$ & $D^\pm \to K\pi\pi$ >6x faster!**

Steering: Analysis Parser

- Use common **selection, combination**, **nTuple** functions.
- Based on *decay description*
  - `B+ -> D0 (-> K+ pi- ) pi-`
- "Natural" grammar (*Boost* library)
  - `Arrows, PDG names, Math symbols, analysis functions, mapping.`
  - Checks against unphysical analyses.
- *Official Skims will use this approach: easy debugging, transparent.*



Functions@steering level (e.g.)

- **Particles, Vertices**
  `P,PT,M,Q,MASS,…`
- **Particle identification**
  `eID,μID,KID,…`
- **Topological**
  `R2, Thrust …`

NTuple tools

- Book with standard tool for common parameters.
  - **"Plug-ins"** for Events, Particles, MC-matching, Vertices, Continuum suppression, Tagging, Vetoes. (minimises user bugs)

Module Parameters

'B- -> D0 pi- (^D0 -> ^K+ pi-)'
Selected ParticleList
Plug-in List

Plug-in Tools

- Topology & continuum variables
- Momentum etc.
- Tracking χ2, extra info, etc.
- PID information, e, μ, K, π, p, etc.
- ...

- nTuples provenance tracked (with Belle-II metadata)

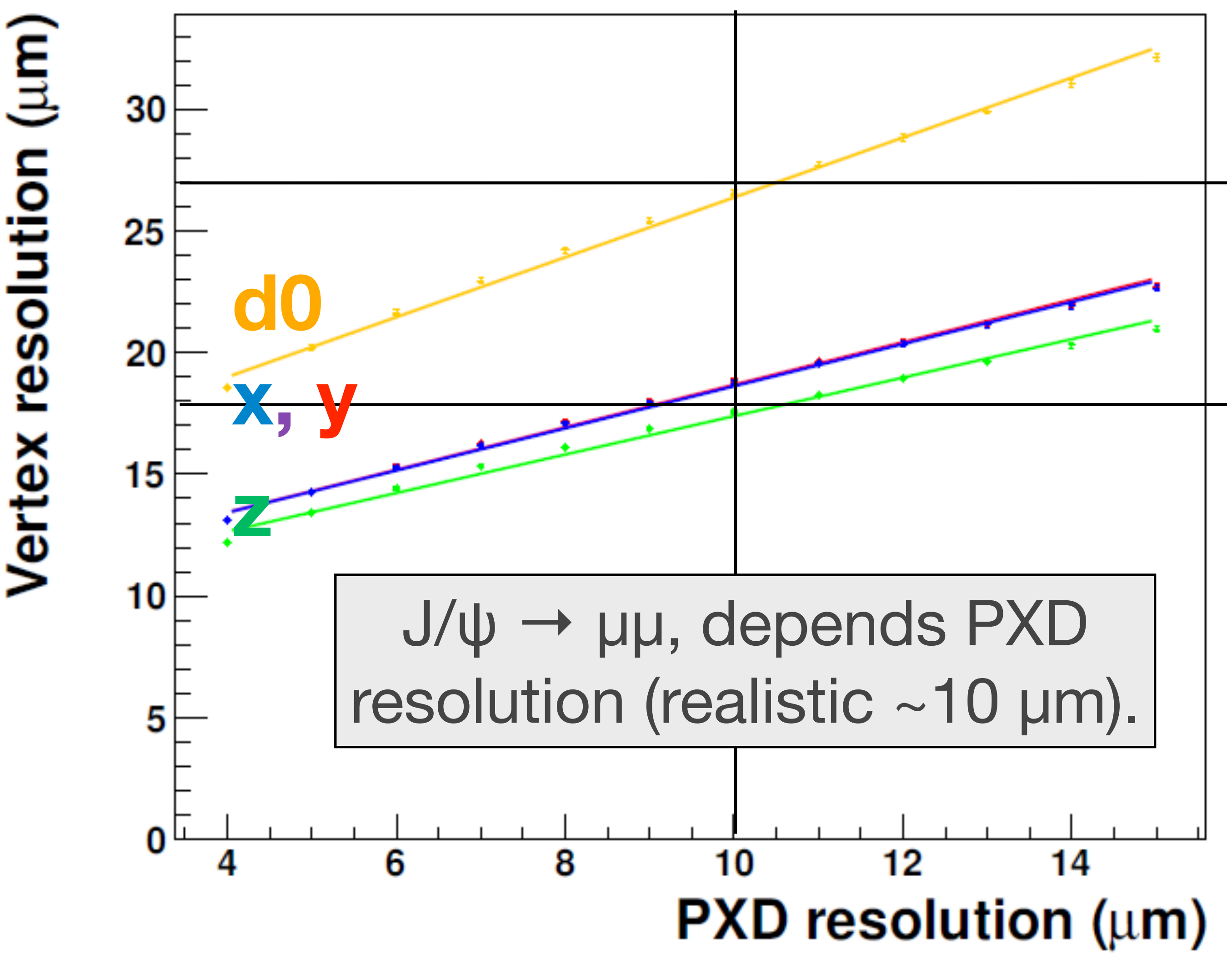
Advanced physics toolkit

Novel Algorithms Under Development

Tagging & Vertexing

New Interface for various fitting techniques (e.g.):

- **RAVE vertexing**: active vertex refitter (based on annealing used in tracking) & full magnetic field map.
- Full decay chain refitting & constraints for neutrals.



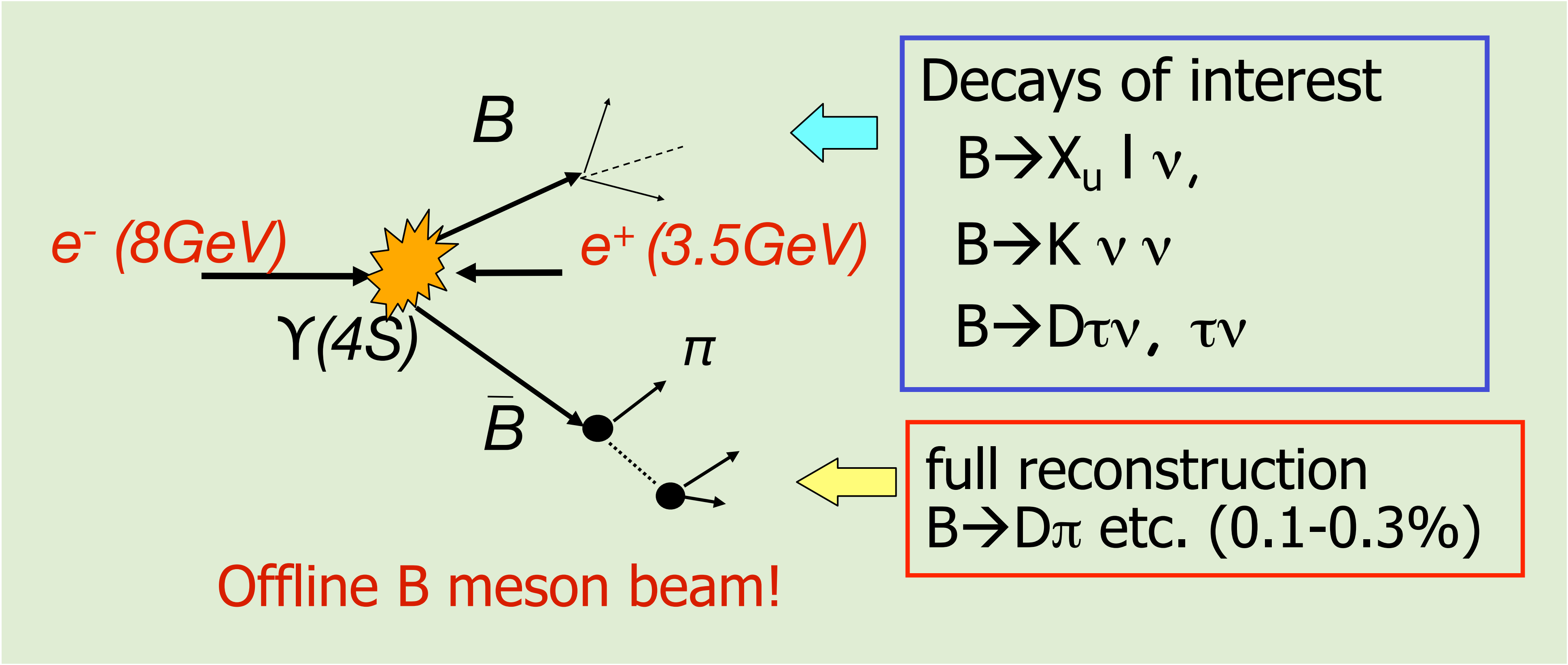
Simple case (No IP constraint)

σ(z) on Vertex
Belle II~18μm *c.f.* Belle~61μm

Full Reconstruction (B & D)

Most computationally intensive task in analysis computing

Fully reconstruct one B/D to tag flavour, charge, momentum

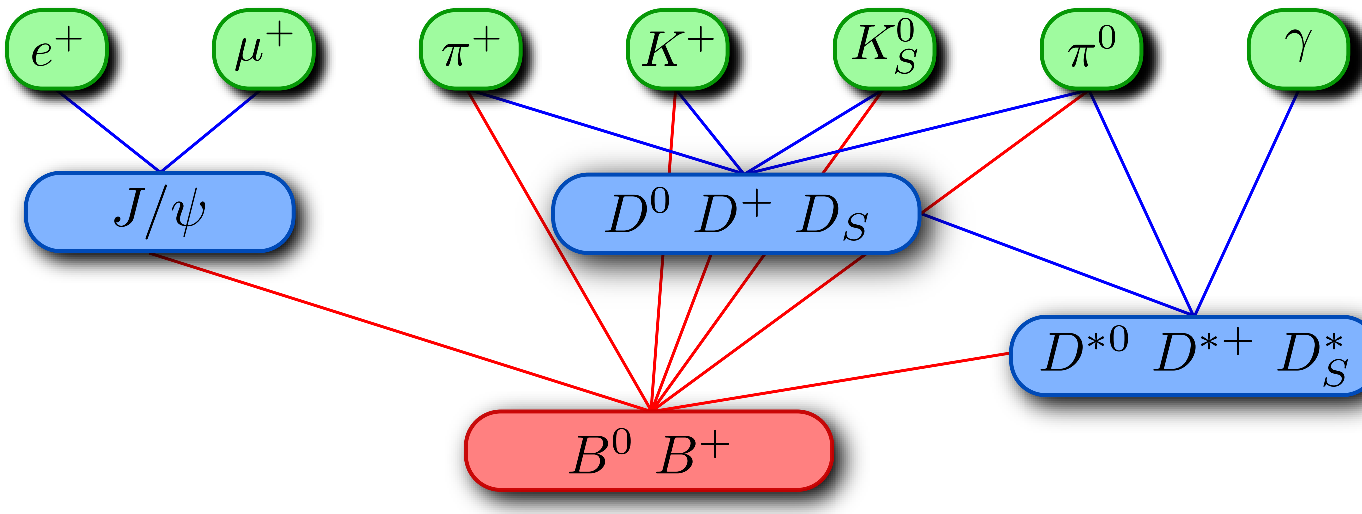


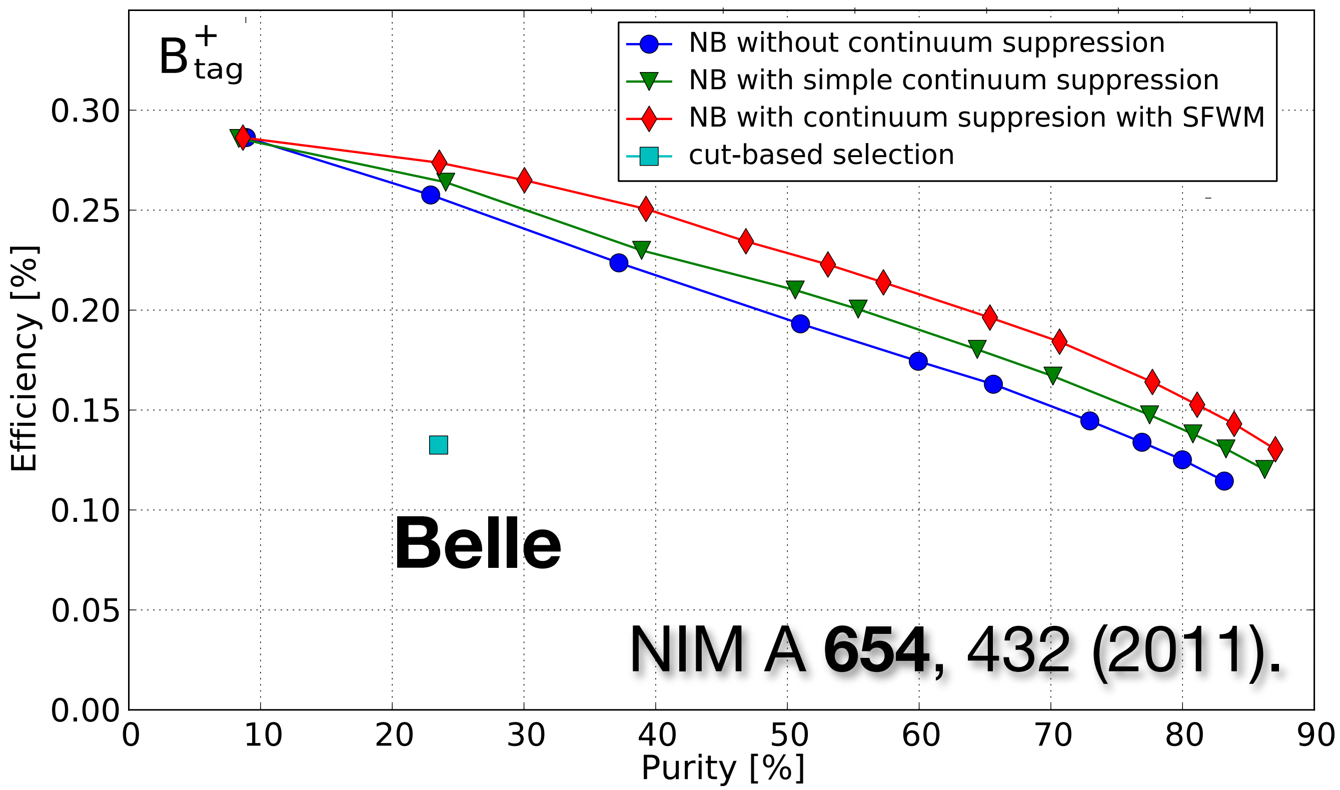
Unique tool for Belle II, for decays with neutrinos

Low efficiency, requiring MVAs

Tagging Algorithms

	e.g. Belle
Method	**NeuroBayes** optimised, 4 Particle selection **layers** with individual classifiers
Modes	1000
Eff%	0.2-0.3
	High purity, classifier output

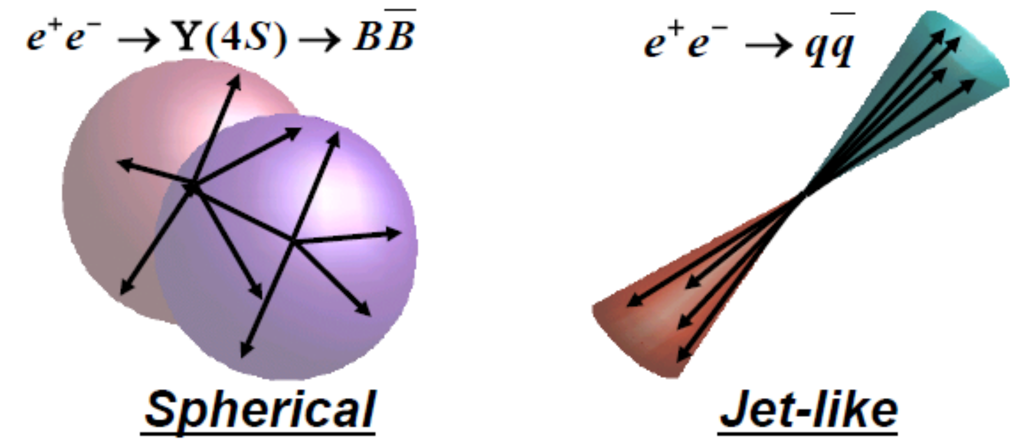




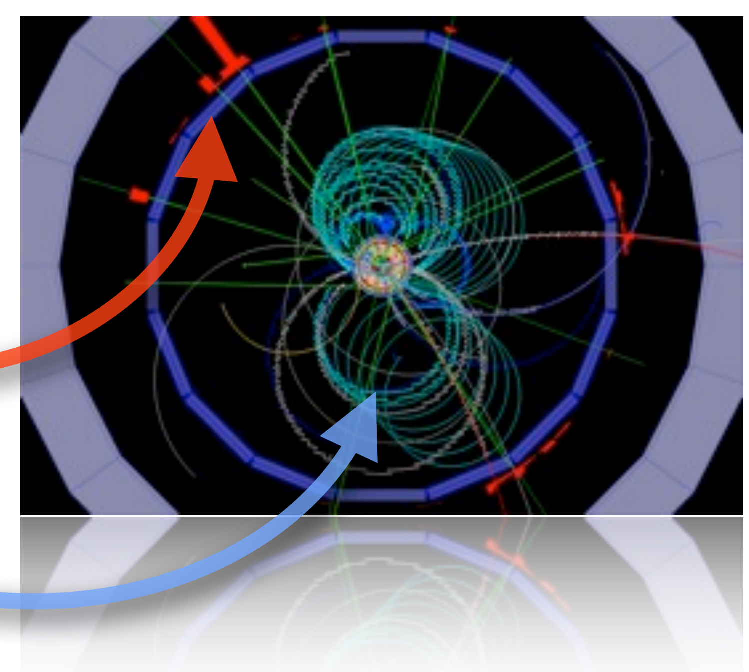
- *Belle II: Integrate with new Particle layer.*

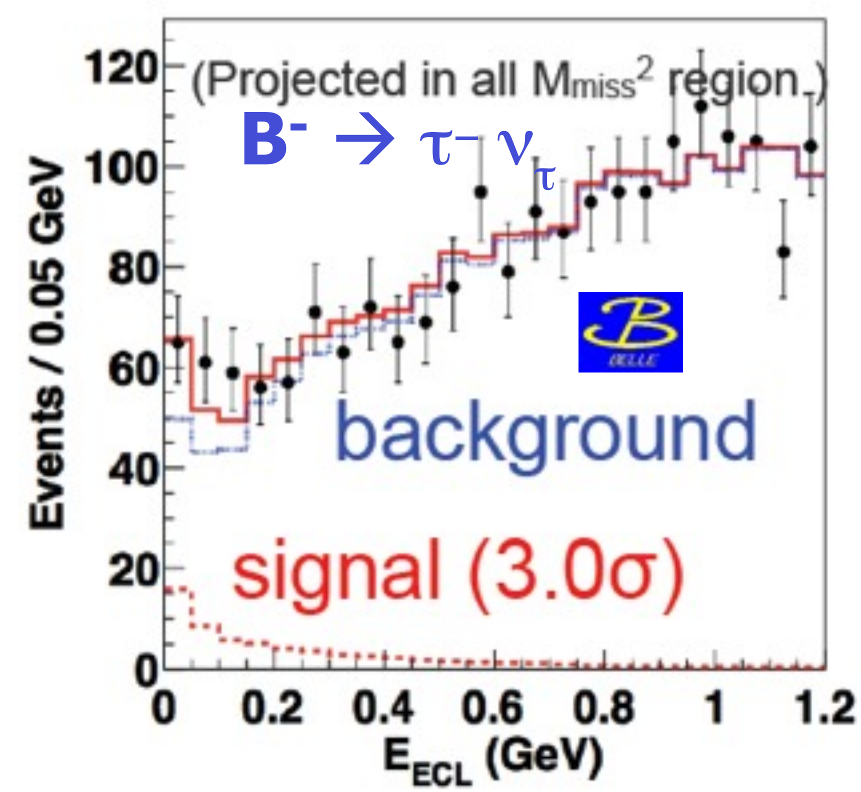
Background suppression

- **Continuum Suppression**
  - (inspired by B-tagging) Novel general purpose **Particle selection MVA class** for layers
  - **Layers:** Hadron, Lepton, Photon, *Event Shapes*



- Residual **energy & tracks**
  - Highly tuned account of low E clusters, K_L, (curling) tracks.
  - May require MVA *classifier* for event **cleaning**.





Other

- **Systematics**
  e.g. ($J/\psi \rightarrow (\mu\mu, ee)$, $D^* \rightarrow D(K\pi)\pi$) for PID, Tracking, A_{CP}
  - Tuned sample formats. Metadata allow grid lookup of parent DST files.
  - Results stored in software tags/releases or DB.
  - Application through plug-in correction tools at *x*DST “object” level.

- **Validation**
  - We hold regular tutorials & nightly physics validation of the framework.

Conclusion

- Belle II will collect 10 ab^{-1}/year, requiring a well designed analysis model.

- New layer of analysis tools & objects designed to improve precision, repeatability, reduce computation time for all analyses.
  - Centralised particle selection.
  - Analysts should see >1 order of magnitude speed improvement.

- To fully exploit Belle II data, complex tagging tools are key: framework is being built for them.

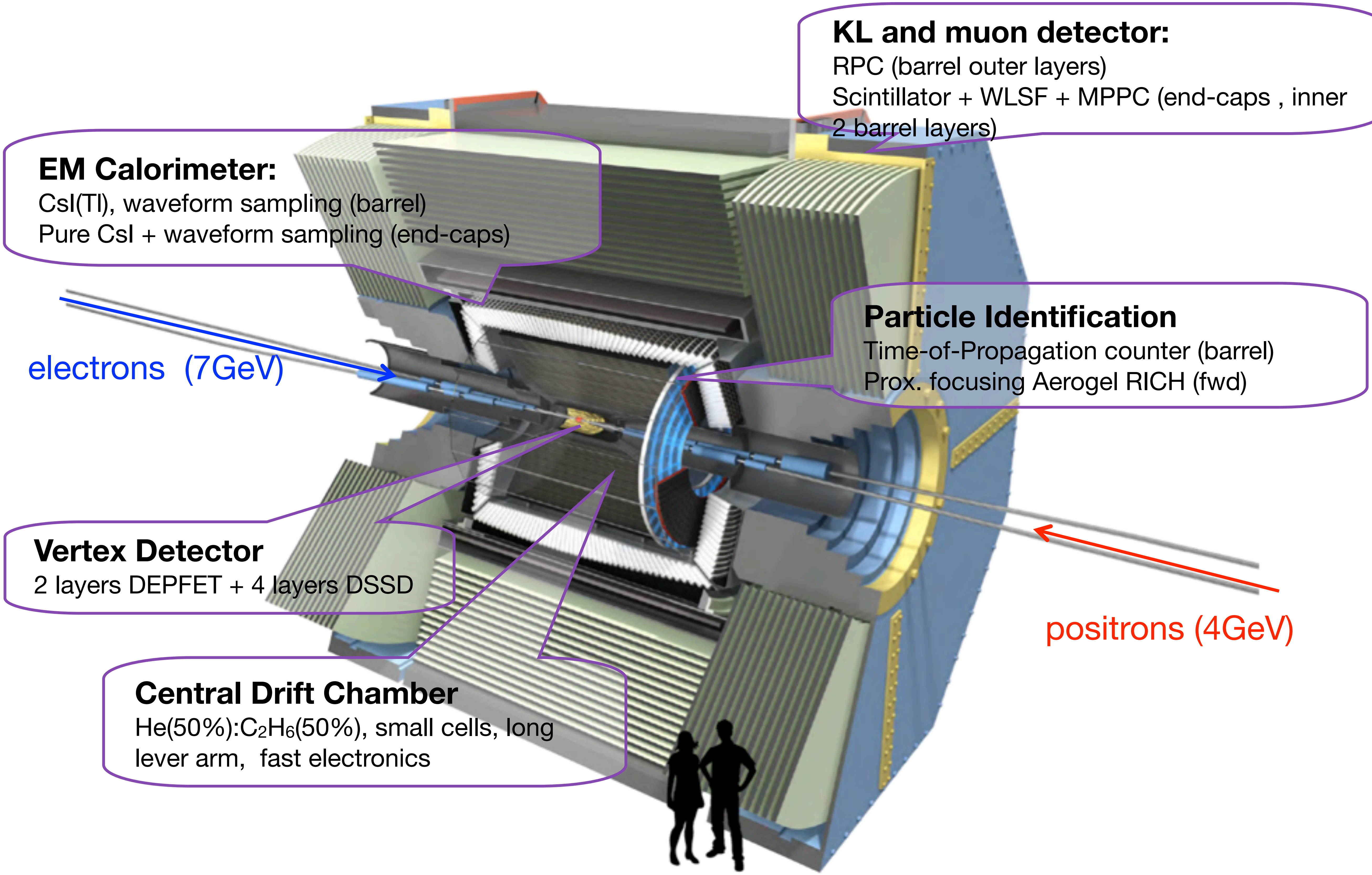
- 2 years of R&D with benchmarking until first collision data.

Backup

Problem-solution checklist

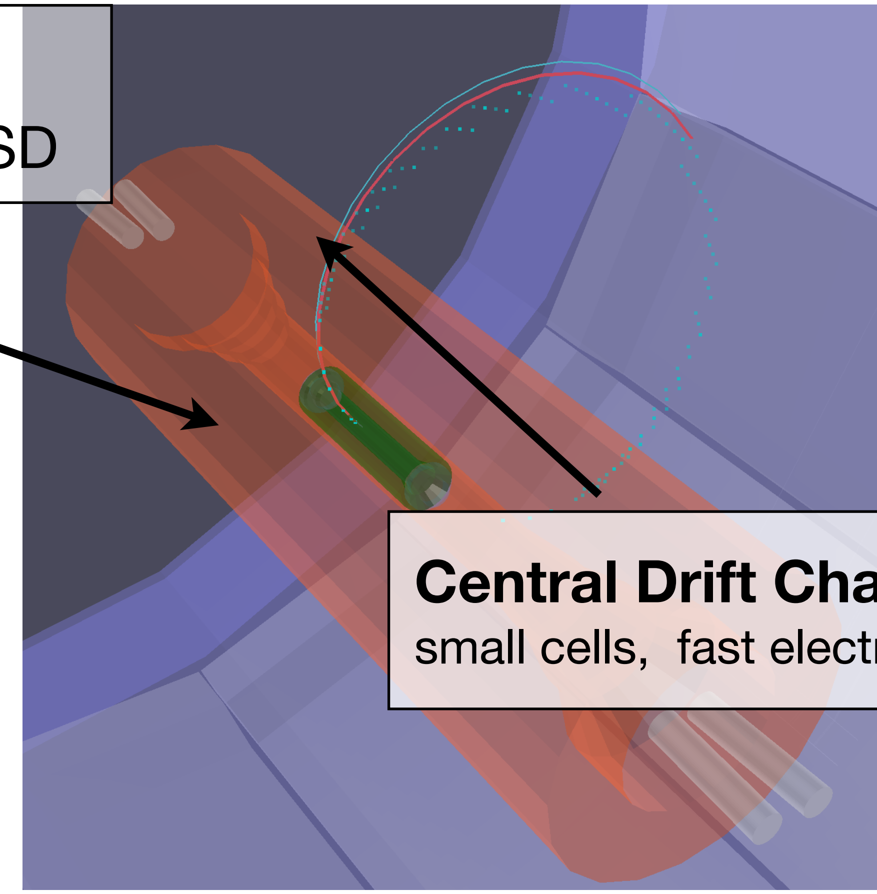
Belle	Belle II
Skimming/filtering often done repeatedly per analysis, per person on the full data set (and often not done carefully enough). Time consuming step.	Centralised skimming framework.
Many analyses share common reconstruction steps, e.g. D, D*, τ modes for B reconstruction. But privately shared reconstruction code and ran separate jobs.	Persistent Particles: Standardised vertexing, Rapid-accurate systematics, Common Particles.
MC produced privately was very slowly, and with more bugs with respect to coordinated production.	Centralised, book keeping
Minimal plug in systematics tools.	Dedicated systematics skims with provenance. Plug-in to BASF2 root DST format.

Belle II Detector



Tracking

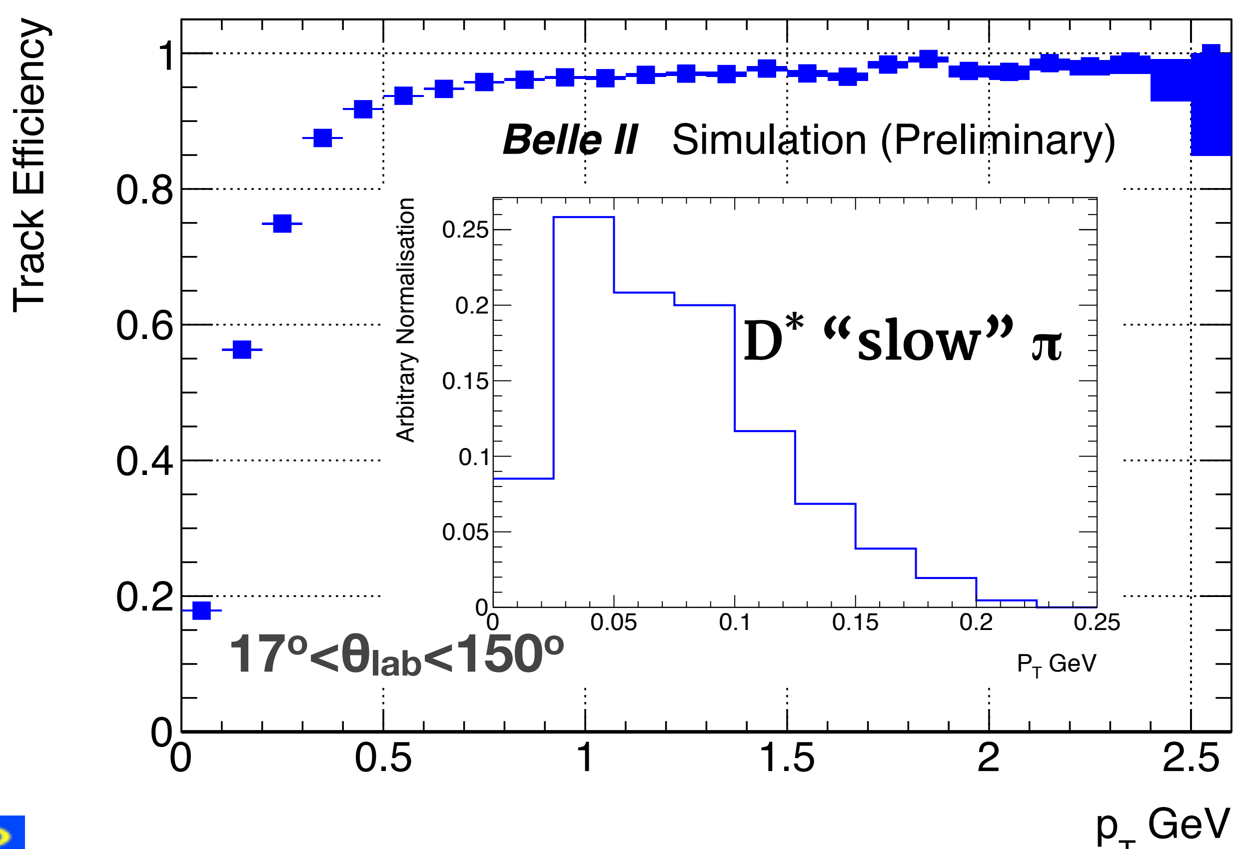
Vertex Detector
2 layers DEPFET + 4 layers DSSD

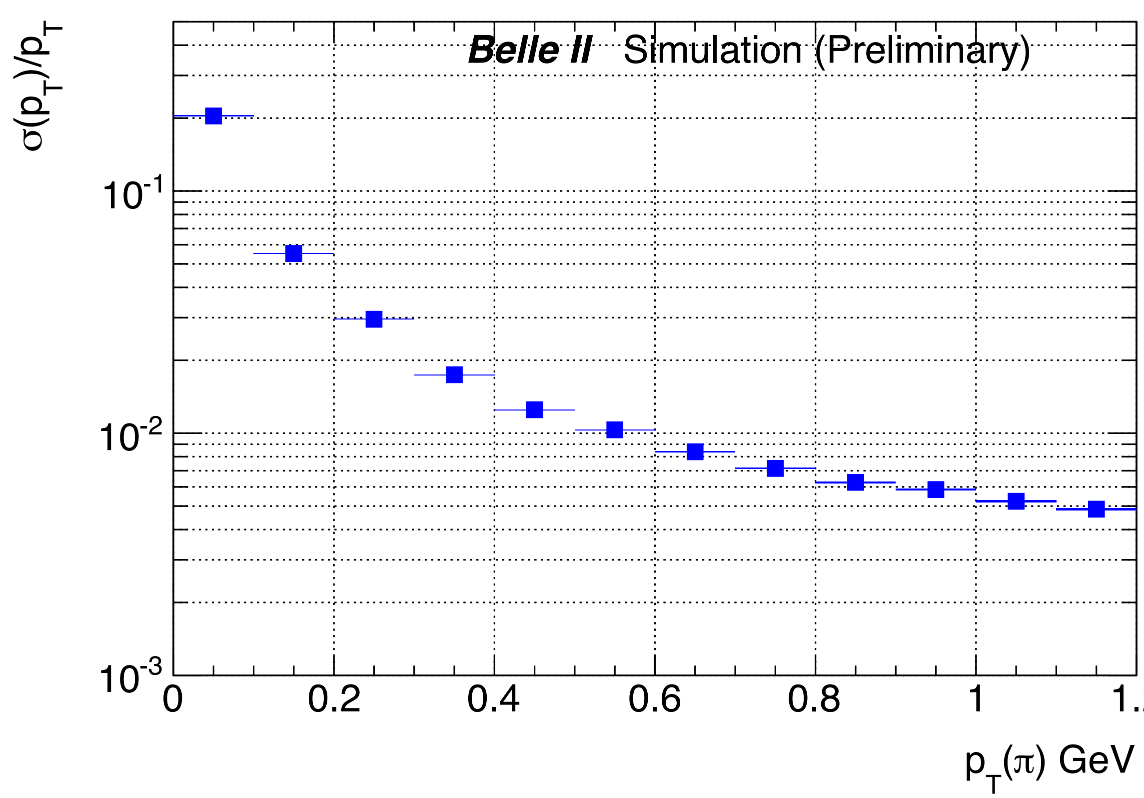


Central Drift Chamber
small cells, fast electronics

- Algorithms under development
  - **General: D**eterministic **A**nnealing **F**ilter,
  - **Si** only (slow **π**): cellular automaton with Hopfield network **NN**.

    J Lettenbichler *et al* 2012 *J. Phys.: Conf. Ser.* 396 022030
  - **Shown: Kalman**



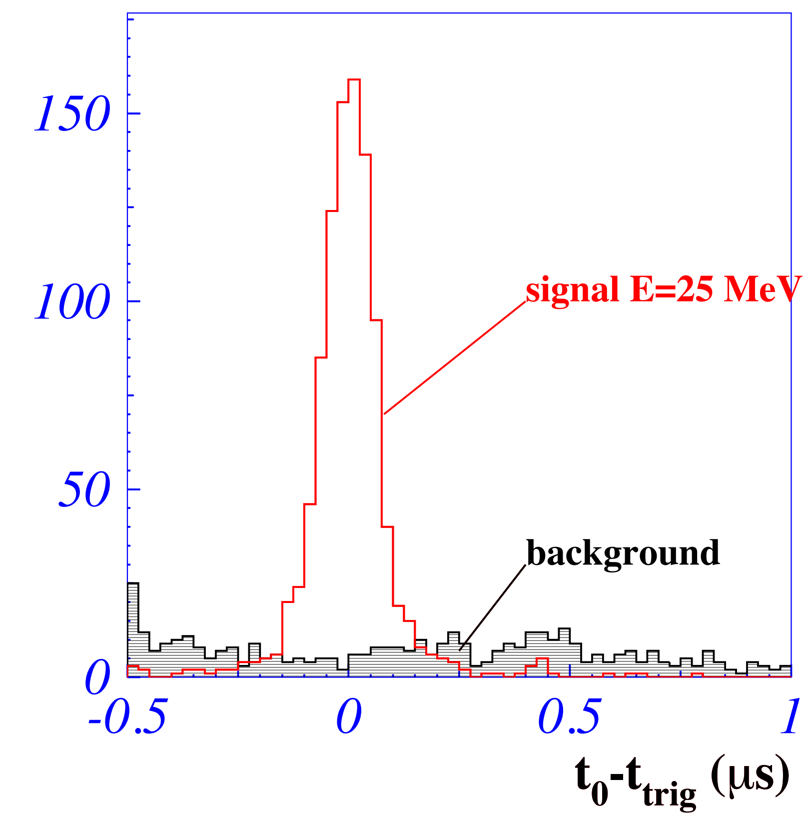


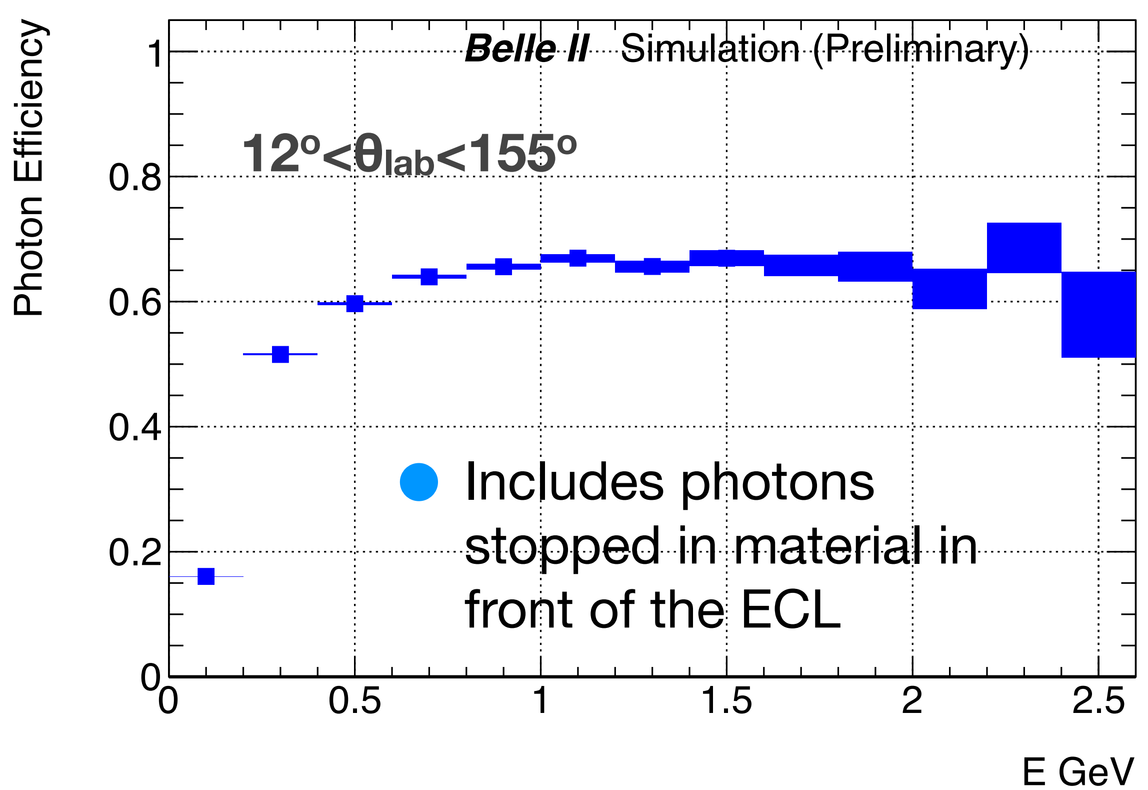
Calorimetry

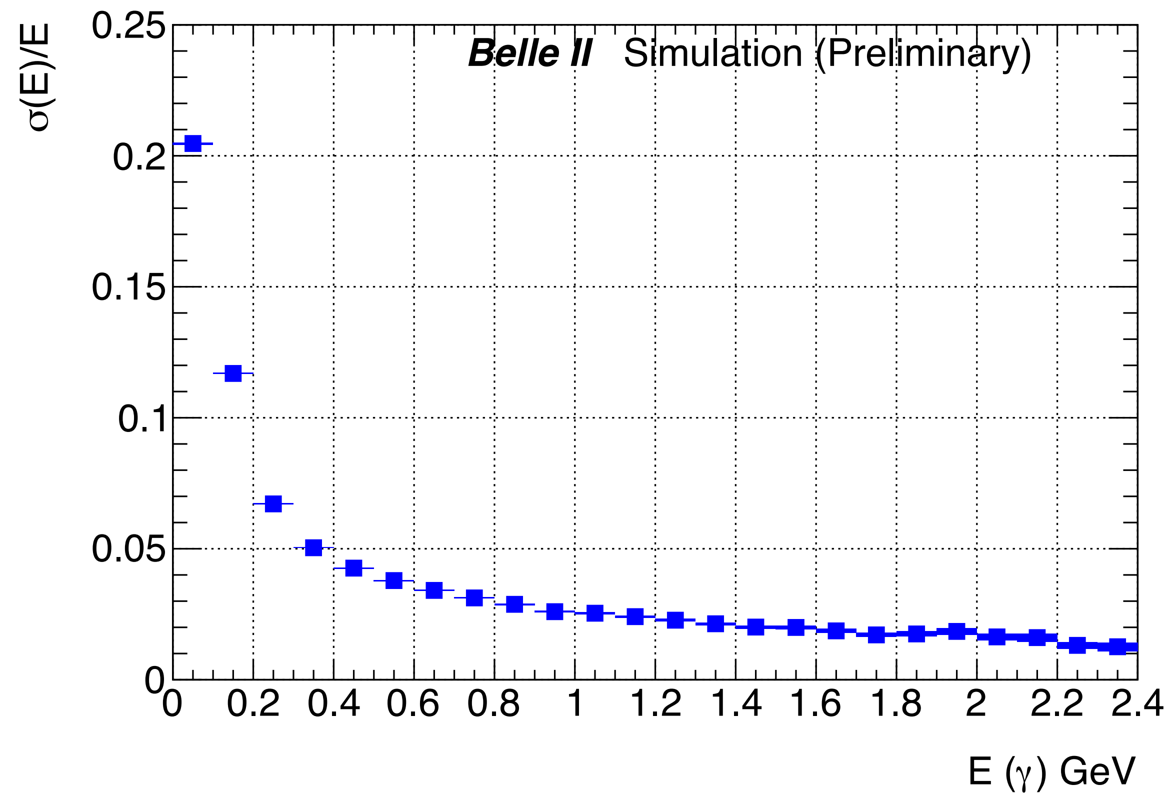
EM Calorimeter new waveform sampling:
CsI(Tl) (barrel), Pure CsI (end-caps)

I. Nakamura *et al* 2009 *J. Phys.: Conf. Ser.* 160 012003

- *Timing information → off-time rejection + calibration schema.*



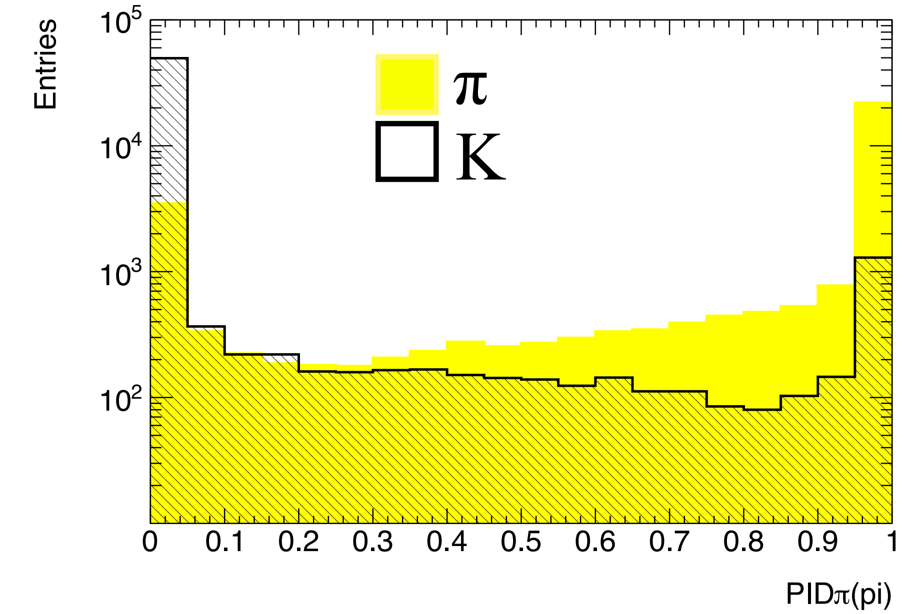




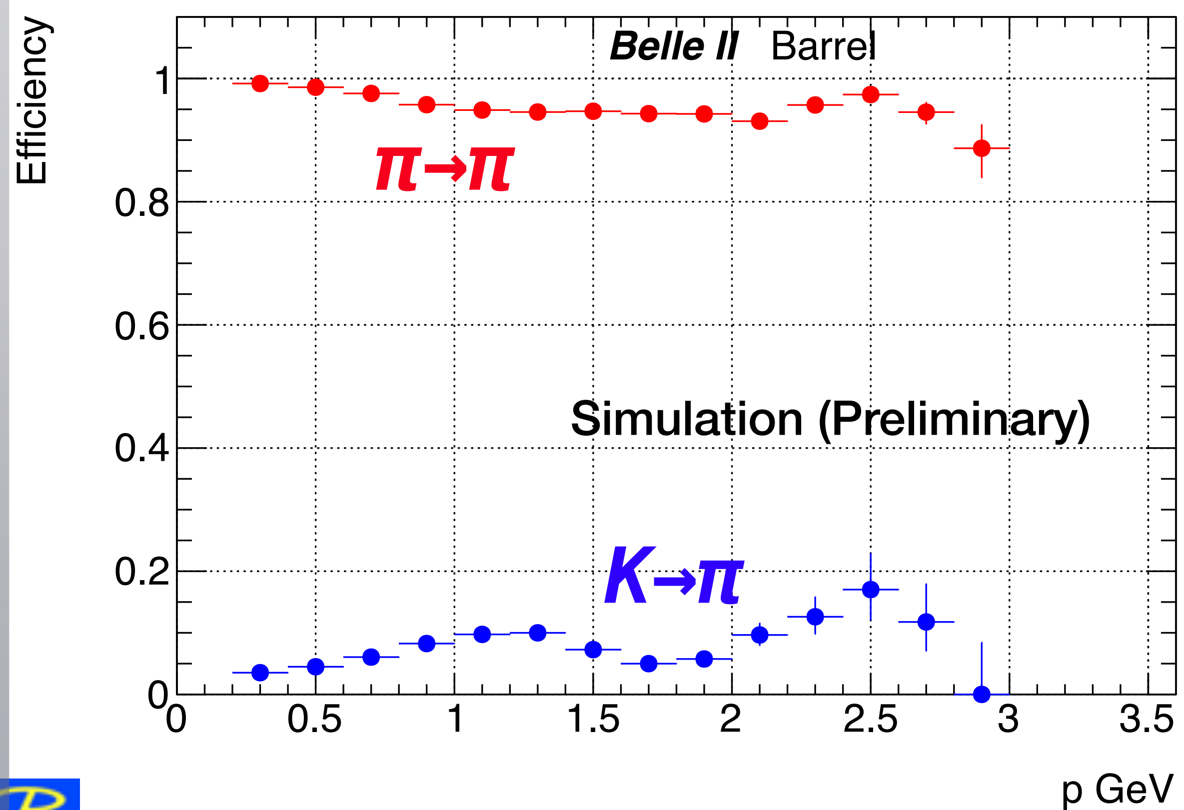
Particle Identification

Particle Identification
Time-**O**f-**P**ropagation counter (barrel)
Prox. focusing **A**erogel **RICH** (fwd)

- Likelihood of TOP or **A**RICH + dE/dX (CDC).
- *c.f.* CPV in $B\to\rho^0(\pi\pi)\gamma$, Background $B\to K^{*0}(K\pi)\gamma$



Barrel (TOP)



Endcap (ARICH)