“Search for long-lived particles using displaced vertices and missing momentum in proton-proton collisions at $\sqrt{s}$ = 13 TeV”

Student Project for AEPSHEP 2024, Discussion Group E

Standard Model particles

up quark

charm quark

top quark

photon

down quark

strange quark

bottom quark

gluon

electron

muon

tau

weak bosons

electron neutrino

muon neutrino

tau neutrino

Higgs boson

Supersymmetric (SUSY) particles

sup squark

scharm squark

stop squark

photino

sdown squark

sstrange squark

sbottom squark

gluino

selectron

smuon

stau

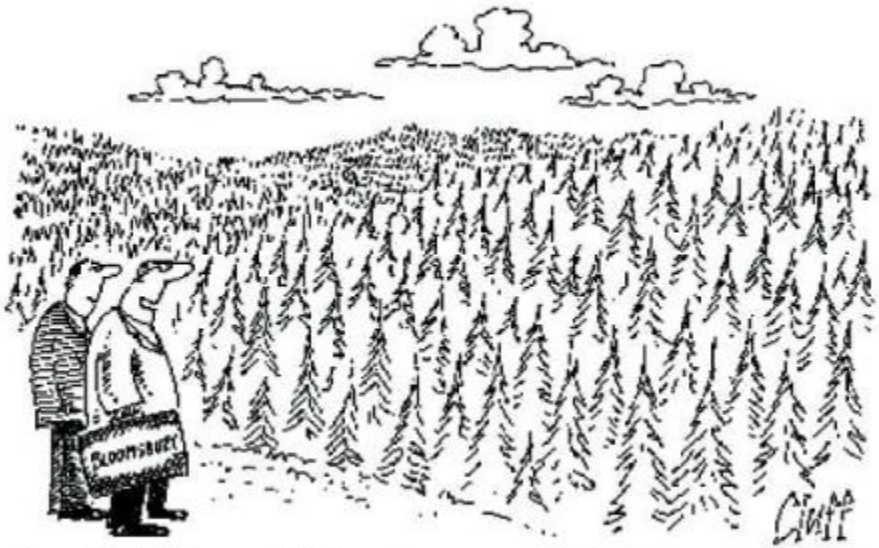
zino · wino

selectron sneutrino

smuon sneutrino

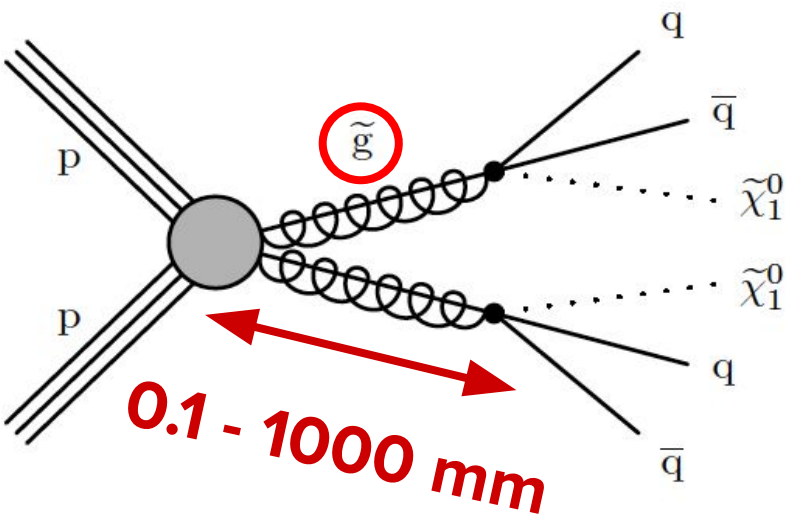
stau sneutrino

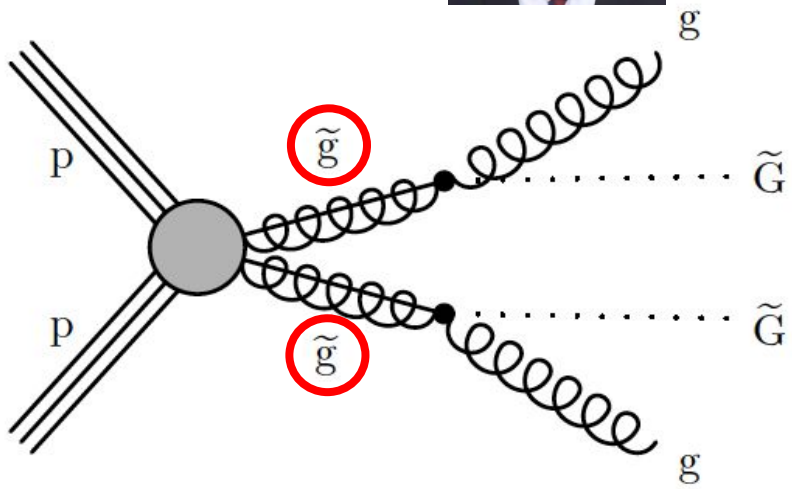
higgsino



"One day, all of these will be supersymmetric phenomenology papers."

Target: Long-lived Gluinos





quark and neutralino final state (split SUSY)

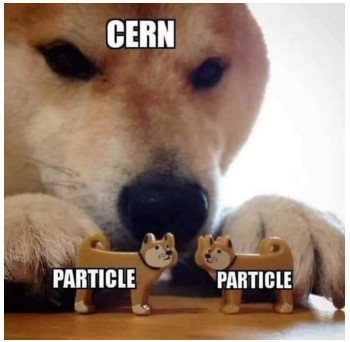
gluon and gravitino final state (GMSB SUSY)

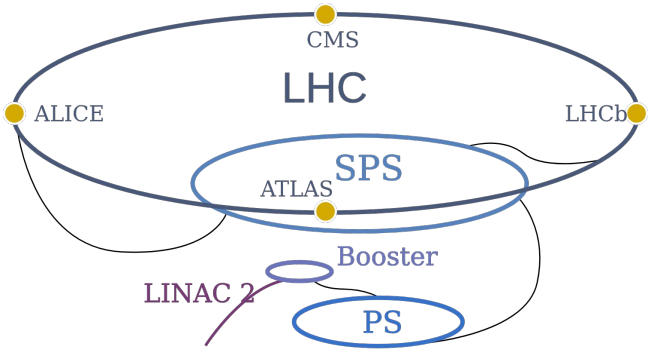
The Large Hadron Collider: LHC

The Large Hadron Collider (LHC) is the biggest particle accelerator located at CERN

- protons and heavy ions
- several acceleration stages
- centre of mass energy $\sqrt{s}$ = 13 TeV

Four interaction points: **CMS**, ATLAS, LHCb, ALICE.

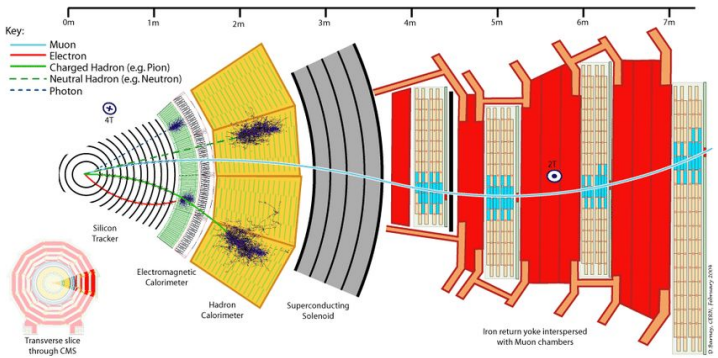


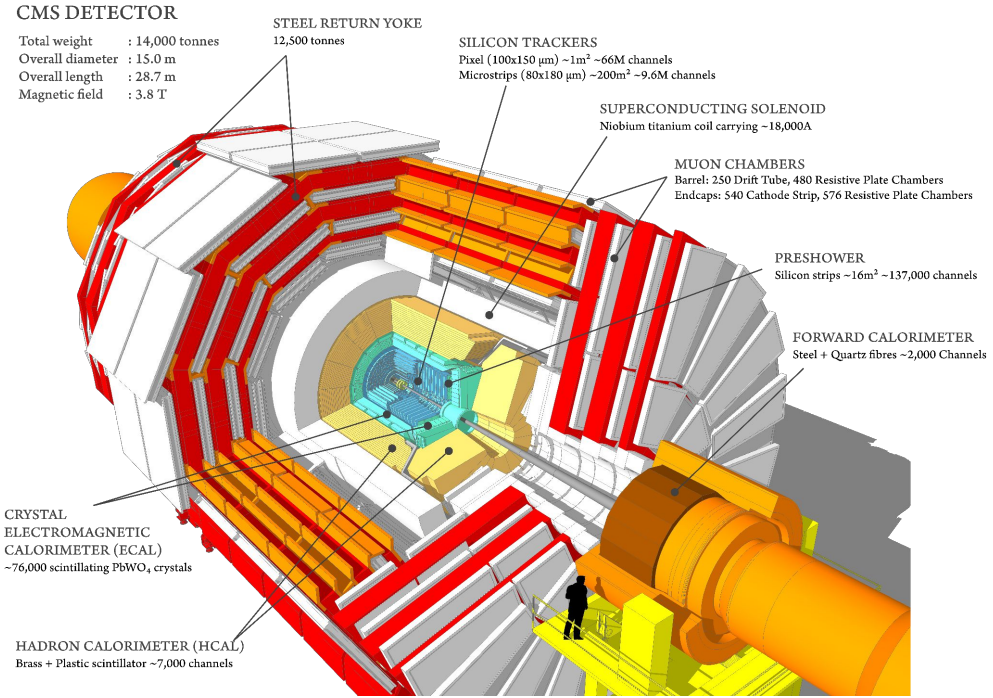


The Compact Muon Solenoid: CMS

The **Silicon Pixel** detector is used to track particles and to reconstruct the vertices.

The **ECAL** aims to reconstruct photons and electrons.





Event Selection & Background Sources

Event Selection

- events with: $p_T^{miss} > 200\,GeV/c$
- collected data in 2016-2018 (analyzed separately)

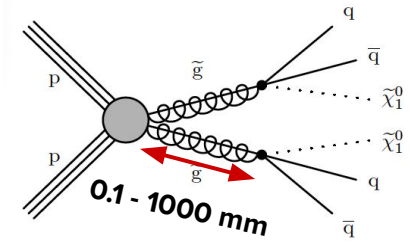
Background Processes

- Dominant
  - QCD
  - $t\bar{t}$
- Additional
  - W/Z boson
  - single top
  - diboson

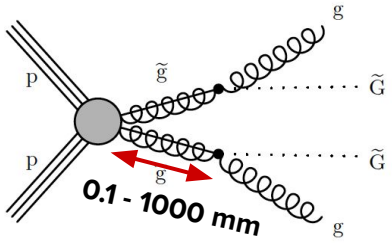
Signal Simulation

- proper decay lengths: 0.1-1000 mm

Split SUSY



GMSB SUSY



With Mass

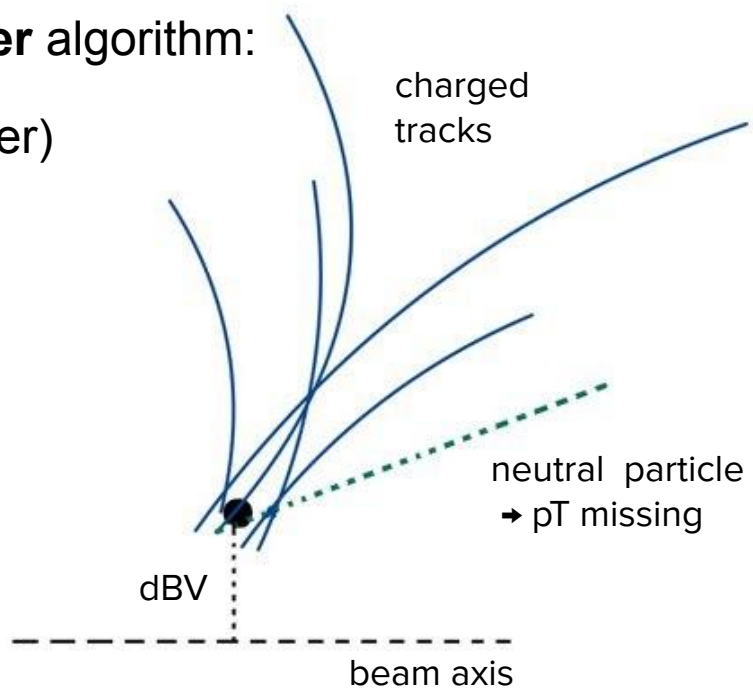
1200-2600 GeV

1800-2800 GeV

Vertex reconstruction

Secondary vertex reconstruction with **Kalman Filter** algorithm:

- Track quality selections (pT and n hits in tracker)
- Tracks from displaced vertices
  - $|d_{xy}/\sigma_{xy}| > 4$ → Remove prompt SM tracks.
- Vertex selection:
  - $\chi^2/\mathrm{dof} < 5$
  - $|d_{BV}| < 2\,\mathrm{cm}$ → decays within the beam pipe
  - $\sigma_{d_{BV}} < 25\,\mu\mathrm{m}$ → suppress contribution from boosted B and K decays
  - $n_{tracks} \geq 3$



What next?


PROMPT
AND OVERLAPPED
VERTICES
BOOSTED
B AND K
MESON VERTICES
RANDOM
COINCIDENCES
imgflip.com
memegenerator.net

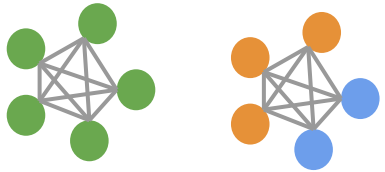
Machine Learning Strategy

Training Data

LLP (1) QCD (0)

Tracking information:
- Momentum
- Impact parameters (both transverse and longitudinal)

Interaction Network



Graph Representation

GNN-based model

Output Score S_{ML}

Loss

Binary Cross Entropy

+

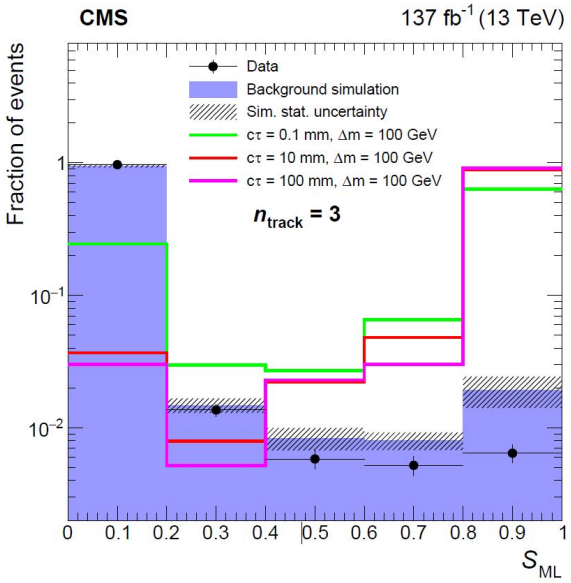
L2 Regularization

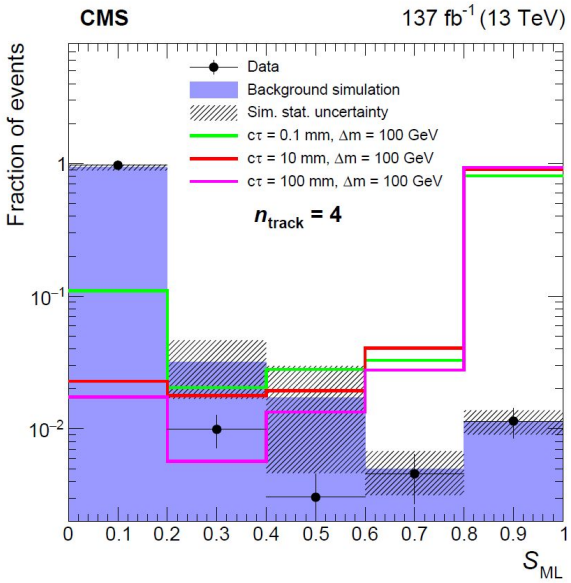
+

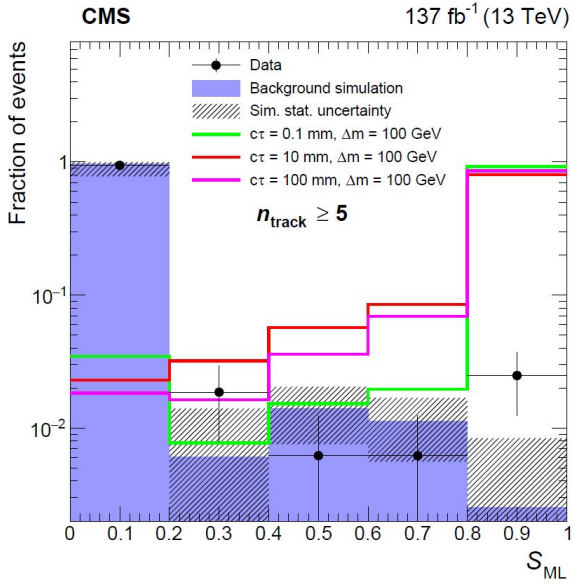
Distance Correlation

Decorrelation to number of tracks in displaced vertex

Model Performance







$S_{ML} > 0.2$ ⇒ Identify ~97% of Signals & reject ~97% of Backgrounds

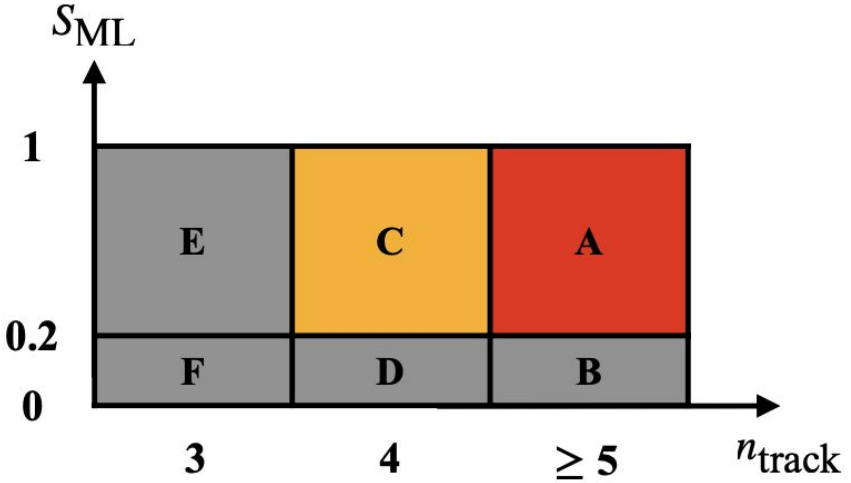
Background Estimation

- S_{ML} and n_{track} are stat. independent
- Number of background events:

$$N_{\mathrm{bkg}}^{\mathrm{A}} = \frac{N_{\mathrm{bkg}}^{\mathrm{B}} N_{\mathrm{bkg}}^{\mathrm{E}}}{N_{\mathrm{bkg}}^{\mathrm{F}}},$$

$$N_{\mathrm{bkg}}^{\mathrm{C}} = \frac{N_{\mathrm{bkg}}^{\mathrm{D}} N_{\mathrm{bkg}}^{\mathrm{E}}}{N_{\mathrm{bkg}}^{\mathrm{F}}}.$$

- C region validates the Method



Red - Signal region,
Yellow - Validation region,
Grey - Control region

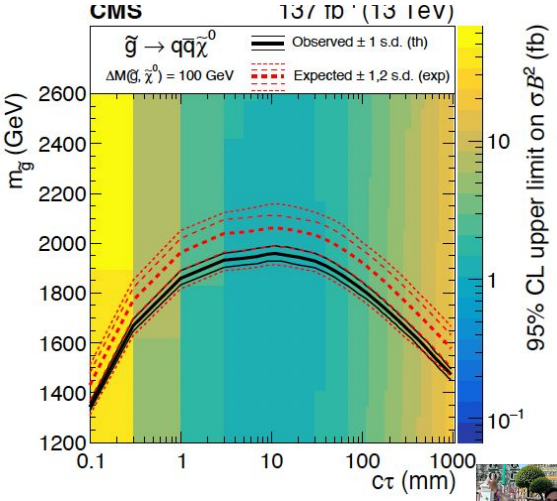
Systematic Uncertainties

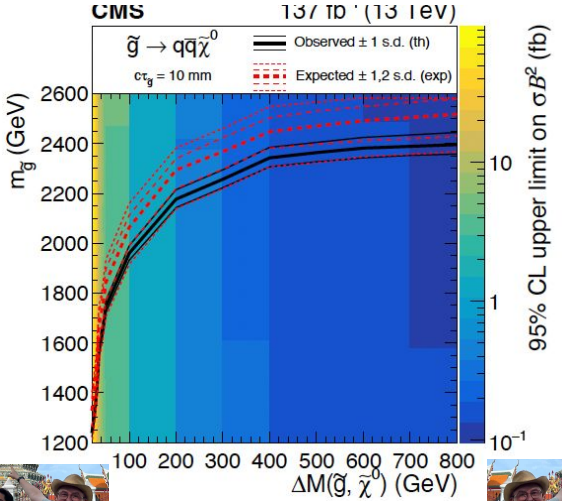
No systematic uncertainties from background estimation

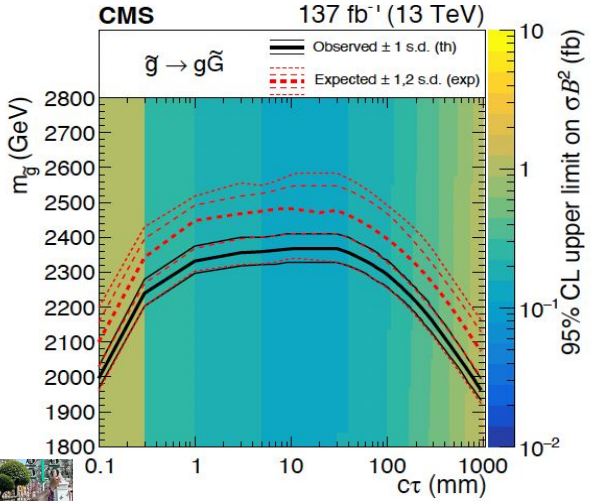
Systematic uncertainty	Magnitude (%)
Track reconstruction	6–21
Vertex reconstruction	3–20
ML tagging	≤24
$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ selection	≤8
PDF uncertainty	1–85
Trigger efficiency	1–6
Pileup	2–15
Integrated luminosity	1–3
L1 trigger inefficiency	≤1
Total	8–91

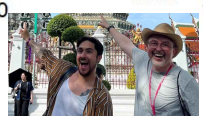
Results & Statistical Interpretation

	$n_{\text{track}} = 3$	$n_{\text{track}} = 4$	$n_{\text{track}} \geq 5$
Predicted $S_{\text{ML}} > 0.2$	— (E)	38.0 ± 6.0 (C)	5.2 ± 0.5 (A)
Observed $S_{\text{ML}} > 0.2$	203 (E)	38 (C)	9 (A)
Observed $S_{\text{ML}} < 0.2$	6327 (F)	1276 (D)	152 (B)



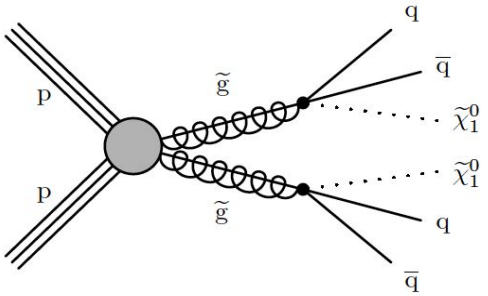


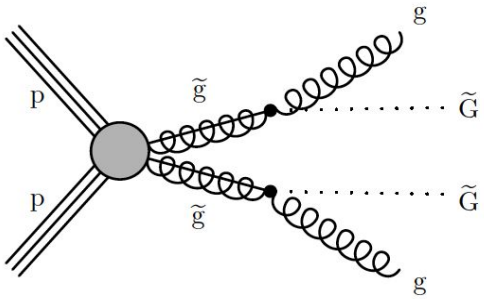




Summary

- The CMS studied LLPs.
- **Most stringent limits for split and GMSB model to date!**
- No SUSY …yet ?
- Still hope …?





NO NEW PHYSICS OBSERVED

→ LOOK FURTHER/DEEPER

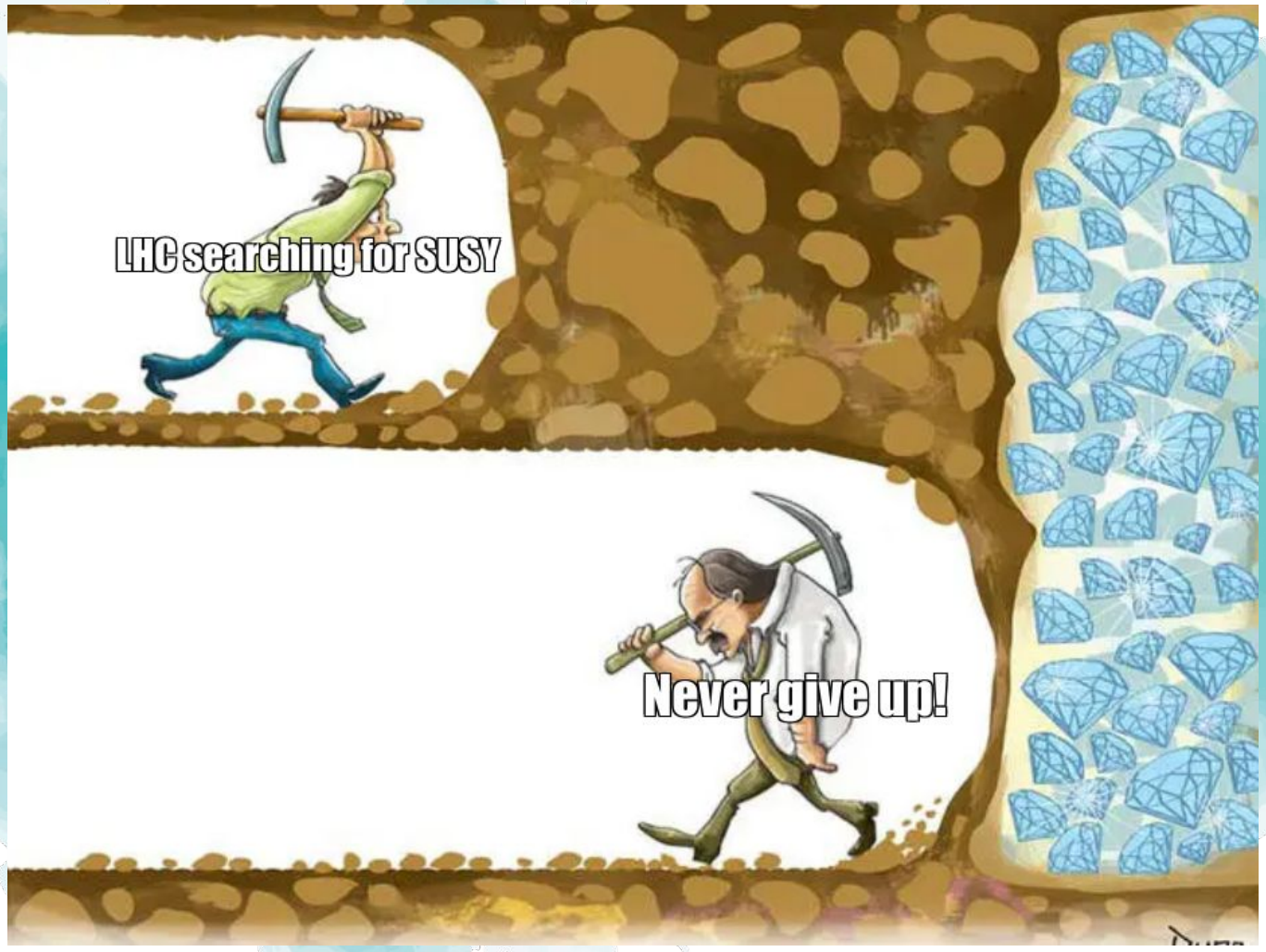
Explore new tools and techniques

The highest energy possible

The highest luminosity possible

As low backgrounds as possible

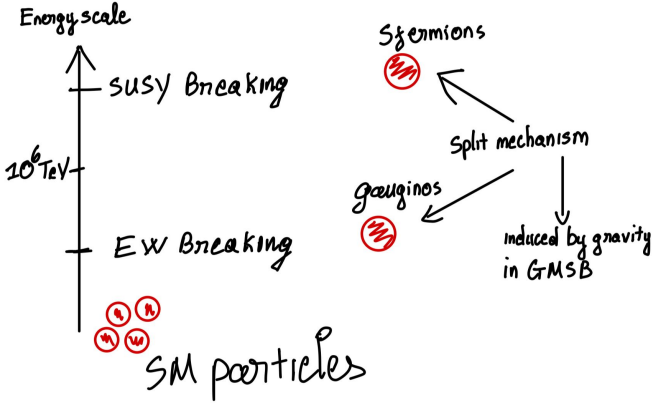
Thank You

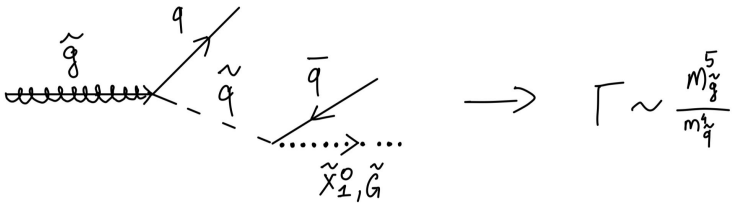

LHC searching for SUSY
Never give up!

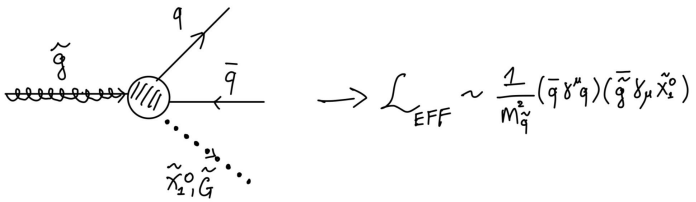
Backup slides

Long-lived Gluino in Split-SUSY and GMSB

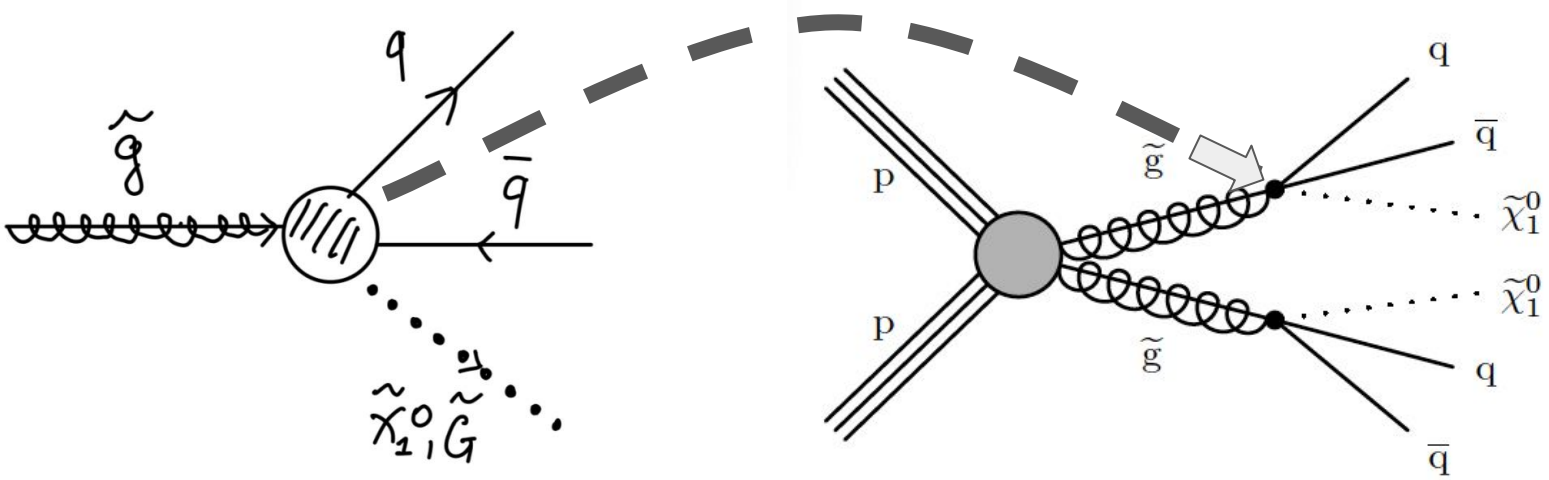
- A mechanism generates the SUSY breaking term
- Sfermions in very high energy scale, gauginos in weak-scale
- Gluino decay into LSP and pair quarks mediated by extremely heavy squark ($m_{\tilde{q}} \gg 10^6\ \mathrm{TeV}$), make it a long-lived particle able to traced at current detectors.



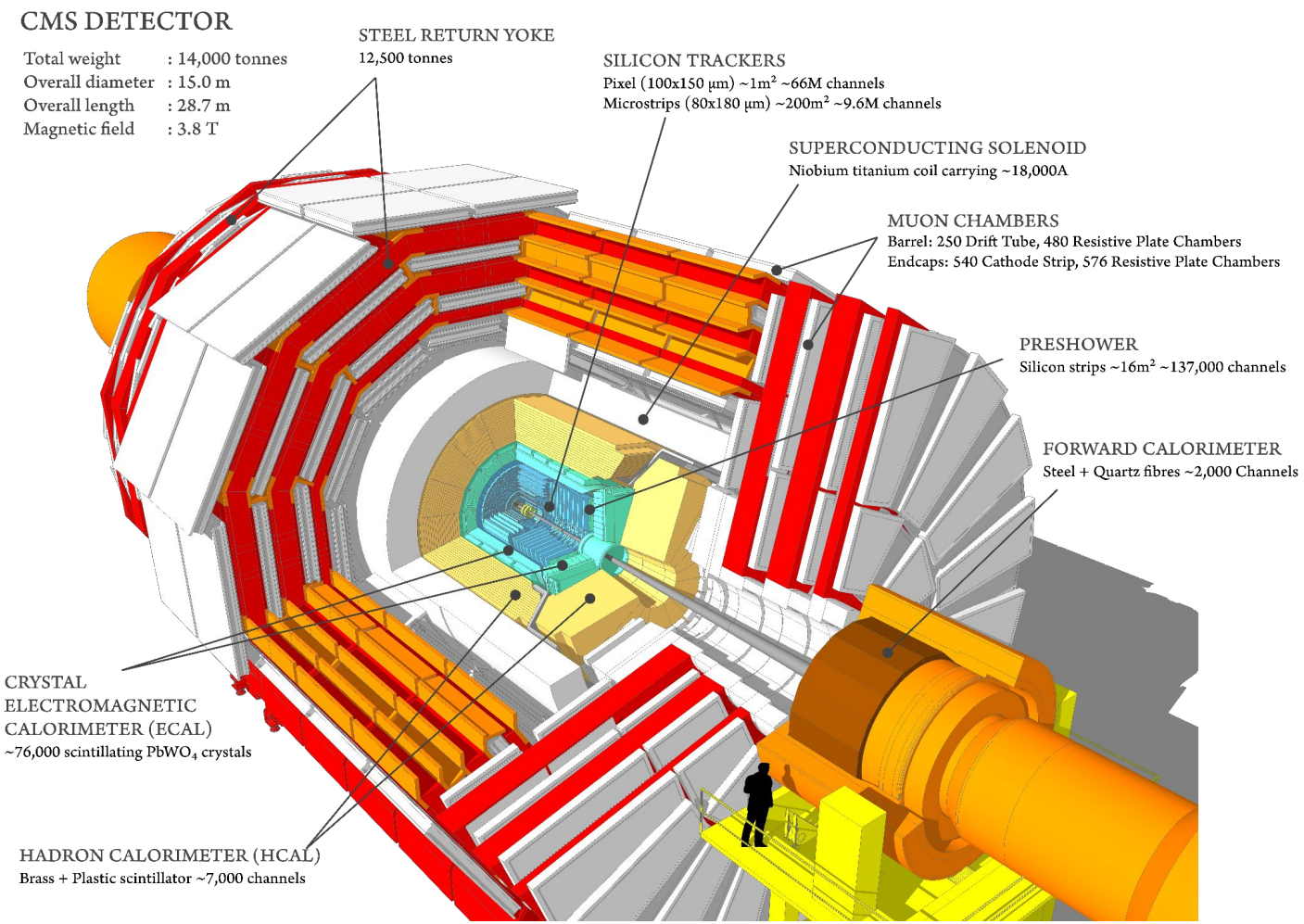




Long-lived Gluino in Split-SUSY and GMSB



$$\mathcal{L}_{EFF} \sim \frac{1}{M_{\tilde{q}}^2}(\bar{q}\gamma^\mu q)(\bar{\tilde{g}}\gamma_\mu \tilde{\chi}_1^0)$$


CMS DETECTOR
Total weight : 14,000 tonnes
Overall diameter : 15.0 m
Overall length : 28.7 m
Magnetic field : 3.8 T
STEEL RETURN YOKE
12,500 tonnes
SILICON TRACKERS
Pixel (100x150 μm) ~1m² ~66M channels
Microstrips (80x180 μm) ~200m² ~9.6M channels
SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying ~18,000A
MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 540 Cathode Strip, 576 Resistive Plate Chambers
PRESHOWER
Silicon strips ~16m² ~137,000 channels
FORWARD CALORIMETER
Steel + Quartz fibres ~2,000 Channels
CRYSTAL
ELECTROMAGNETIC
CALORIMETER (ECAL)
~76,000 scintillating $PbWO_4$ crystals
HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator ~7,000 channels

Background

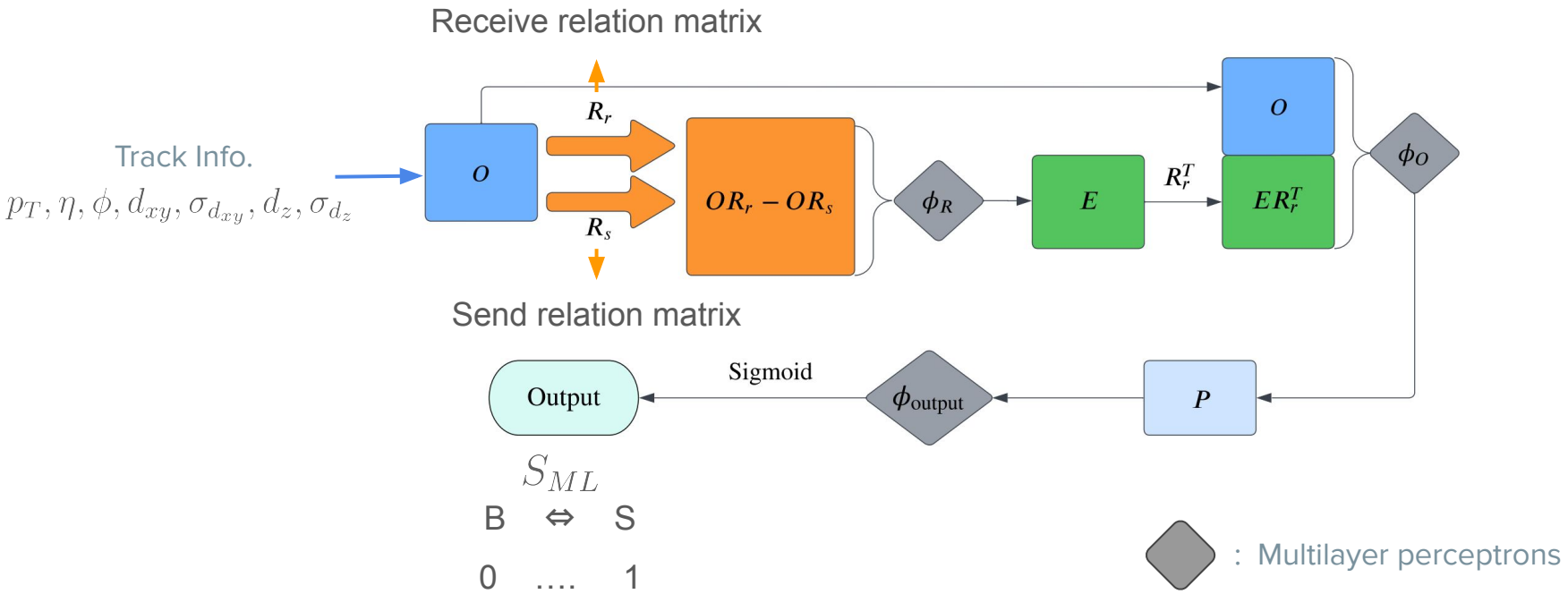
- Background Processes
  - dominated and additional background
  - generated by various simulation tools and parameters
    - LO (Leading Order)
    - NLO (next-to-LO)

	dominant		additional		
background	QCD	$t\bar{t}$	W/Z boson	single tops	diboson
generator	MADGRAPH5 aMC@NLO 2.6.5	MADGRAPH5 aMC@NLO 2.6.1	MADGRAPH5 aMC@NLO 2.6.5	POWHEG	PYTHIA 8.240
at	LO	NLO	LO	NLO	

Reconstruction

- Reconstruction
  - same algorithms as used for data
    - (using the CMS detector and GEANT4)
  - ensure that simulation accurately
  - reflects the characteristics of real data

Interaction Network



Distance Correlation (2001.05310)

Related to the dimension of X and Y

$$\mathrm{dCov}^2(X,Y) = \int d^p s d^q t \left| f_{X,Y}(s,t) - f_X(s) f_Y(t) \right|^2 w(s,t)$$

Weight function: uniquely determined by requiring invariance under shift, orthogonal and scale transformation.

$$\mathrm{dCorr}^2(X,Y) = \frac{\mathrm{dCov}^2(X,Y)}{\mathrm{dCov}(X,X)\mathrm{dCov}(Y,Y)}$$

Distance correlation: bounded in 0 and 1

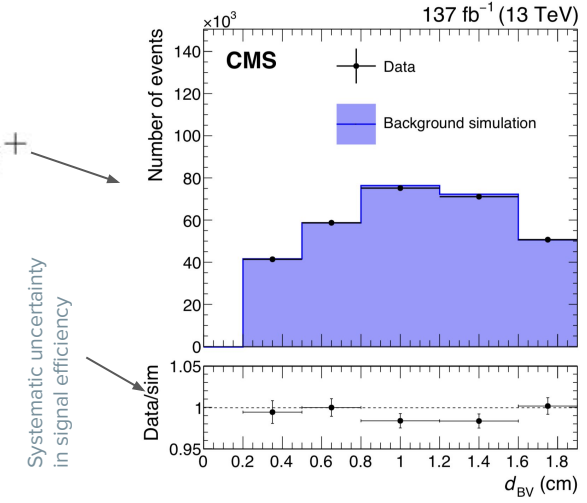
Train and test samples

The events used for training, validation, and testing are required to have at least one reconstructed vertex with $n_{\text{track}} \geq 3$ and pass all event preselection requirements except the offline $p_{\text{T}}^{\text{miss}}$ selection. Training and validation events are required to have $80 < p_{\text{T}}^{\text{miss}} < 200\,\text{GeV}$ to ensure orthogonality with testing events, which have $p_{\text{T}}^{\text{miss}} > 200\,\text{GeV}$. Within the training and validation events, 85% of them are used for training, and 15% are used for validation. To avoid any bias introduced by potential mismodeling of simulated events, 17 067 data events and 31 165 simulated background events are mixed together and labeled as background events during the training of the IN. The simulated background events are drawn from different simulated SM processes in proportion to their cross section. To avoid potential bias, data events that include vertices with $n_{\text{track}} \geq 5$ are excluded from the training and validation samples. Split-SUSY signal samples with different lifetimes and masses are combined and labeled as signal events in the training, and events from different data-taking periods are combined during training. Specifically, 91 013 events are used for training and 17 065 events are used for validation.

Event selection & Signal efficiency

- Signal region contains events
  - Passes preselection
  - At least one displaced vertex with $n_{track} \geq 5$, and have $S_{ML} > 0$
- Signal efficiency
  - Fraction of generated signal events pass signal region

$$K_S^0 \rightarrow \pi^- \pi^+$$



Vertex reconstruction & ML tagging efficiency

- Artificially displacing track in data
- Simulated background mimics displaced vertex as LLP

- Fraction of artificially displaced vertex reconstructed with in 200μm
- Fraction of reconstructed with S_{ML}>0.2