Search for squarks and gluinos in final states with jets and missing transverse momentum using 139 fb⁻¹ of \( \sqrt{s} = 13 \) TeV \( pp \) collision data with the ATLAS detector

09.April.2021
Reinterpretation: Auxiliary Material Presentation (RAMP) kickoff meeting
Kenta Uno
The University of Tokyo \( \rightarrow \) Niigata university
Introduction

Kenta Uno

• 2017 – 2020: Worked on SUSY Strong 0-lepton analysis
  • Main analyzer of BDT analysis
• 2020 – Present: Move to Belle II experiment as post-doc

I use figures in the following Refs.

• Ref1: SUSY Strong 0-lepton paper [ link ]
• Ref2: My Doctoral thesis [ link ]

→ If you have question, please contact me, uno@hep.sc.niigata-u.ac.jp
SUSY search at LHC-ATLAS

LHC : Proton-Proton collision at $\sqrt{s} = 13$ TeV

- Gluino pair production has large x-section (Strong interaction)
  - Golden channel in LHC SUSY search!

Gluino production at LHC

![Gluino production diagram]

![Cross section graph]

$pp, \sqrt{s} = 13$ TeV, NLO+NLL - NNLO$_{approx}$+NNLL

- $\tilde{g}\tilde{g}$
- $\tilde{g}\tilde{g}$ (higgsino)
- $\tilde{g}\tilde{g}$ (wino)
- $\tilde{g}\tilde{g}$
- $\tilde{g}\tilde{g}$ (wino)
- $\tilde{g}\tilde{g}$
- $\tilde{g}\tilde{g}$
- $\tilde{g}\tilde{g}$
- $\tilde{g}\tilde{g}$
- $\tilde{g}\tilde{g}$
- $\tilde{g}\tilde{g}$

2021/4/9
Gluino decay process

- Gluino direct decay ($\tilde{g} \rightarrow qq\tilde{\chi}_1^0$): **Simplest process**
- Gluino one-step decay ($\tilde{g} \rightarrow qqW\tilde{\chi}_1^0$):

**Event Topology**
- Multi-jet + Large missing transverse momentum (MET)
Analysis approach: BDT

Key variable in the previous analysis: $M_{\text{eff}}$

$$M_{\text{eff}} = \sum_i |p_T(i)| + E_T^{\text{miss}}$$

- However, it is not efficient as $\tilde{\chi}_1^0$ mass is large

Using correlation between variables is still efficient.

→ BDT analysis is useful for SUSY search
How to train for gluino search?

**Training method**

Training → BDT score → Apply

- Use gluino events and SM background events
- Add 10 – 12 input variables in BDT (e.g. $p_T$, $\eta$, $E_T^{\text{miss}}$, $M_{\text{eff}}$, …)

*We don’t know gluino/neutralino mass.*

- The kinematics depends on these masses.

→ We have to prepare BDT score per each gluino/neutralino mass.

- Need many BDT scores: Not possible.
Determine training samples
The jet activity depends on $\Delta M(\tilde{g}, \tilde{\chi}_1^0)$

- Similar $\Delta M$ ↔ Similar kinematics

→ Prepare 4 categories based on $\Delta M$.

By using category 1–4,
- All signal probing point can be covered.
Overtraining

- Samples are divided into 2 subsets
  - If use half sample 1 as training, use sample 2 as testing.
  - The training is also done to another one.
  - It can keep the statistics in the final fit.

Comparison b.t.w Train and Test

Comparison b.t.w BDT1 and BDT2

No effect on the bias due to overtraining
Signal region definition

Define SRs per each training category: 8 SRs

Gluino direct decay

Gluino onestep decay

Target ΔM

Target ΔM

A good separation power between sig/bkg
Background estimation

SUSY search: Use tail in the background distribution

- MC simulation should be corrected by using the data
- Use Control regions (CRs)

\[
N_{SR} = N_{SR}^{MC} \times \left[ \frac{N_{CR}^{data}}{N_{CR}^{MC}} \right],
\]

\(N_{SR}\): # predicted bkg events in SR
\(N_{SR}^{MC}\): # bkg events in SR (MC simulation)
\(N_{CR}^{MC}\): # bkg events in CR (MC simulation)
\(N_{CR}^{data}\): # bkg events in CR (data)

Signal region VS control region

Assume mismodeling in CRs is same as SRs
- CRs are designed to be kinematically close to corresponding SRs
Fit result

Perform a profile Log Likelihood Ratio (LLR) approach

- Normalization factors obtained from the fit
- Apply them to each MC sample in SR and Validation region (VR)
  - VR is to validate background estimation
  - It is kinematically close to SR, but orthogonal SR/CR

![Graph showing fit results and normalization factors](image)

Confirmed background estimation in BDT analysis is fine

2021/4/9
Result in SRs

<table>
<thead>
<tr>
<th>Signal Region</th>
<th>GGd1</th>
<th>GGd2</th>
<th>GGd3</th>
<th>GGd4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total bkg pre-fit</td>
<td>29</td>
<td>56</td>
<td>253</td>
<td>348</td>
</tr>
<tr>
<td>Fitted background events</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diboson</td>
<td>3.0 ± 0.9</td>
<td>4.9 ± 1.4</td>
<td>21 ± 5</td>
<td>26 ± 7</td>
</tr>
<tr>
<td>Z/γ*+jets</td>
<td>20 ± 4</td>
<td>33 ± 5</td>
<td>139 ± 14</td>
<td>180 ± 18</td>
</tr>
<tr>
<td>W+jets</td>
<td>7.1 ± 2.6</td>
<td>13 ± 4</td>
<td>48 ± 8</td>
<td>52 ± 9</td>
</tr>
<tr>
<td>tH(EW) + single top</td>
<td>0.1 ± 0.3</td>
<td>0.6 ± 0.8</td>
<td>16 ± 5</td>
<td>39 ± 11</td>
</tr>
<tr>
<td>Multi-jet</td>
<td>0.1 ± 0.1</td>
<td>–</td>
<td>0.1 ± 0.1</td>
<td>0.1 ± 0.1</td>
</tr>
<tr>
<td>Total bkg post-fit</td>
<td>30 ± 5</td>
<td>52 ± 6</td>
<td>223 ± 17</td>
<td>298 ± 23</td>
</tr>
<tr>
<td>Observed</td>
<td>34</td>
<td>68</td>
<td>227</td>
<td>291</td>
</tr>
</tbody>
</table>

| (εσr)_{95}^{obs} [fb] | 0.13 | 0.24 | 0.33 | 0.36 |
| S^{95}_{obs}           | 18   | 33   | 46   | 50   |
| S^{95}_{exp}           | 15^{+5}_{-4} | 20^{+5}_{-6} | 44^{+17}_{-12} | 54^{+15}_{-12} |
| p_{0} (Z) | 0.30 (0.51) | 0.05 (1.60) | 0.44 (0.15) | 0.50 (0.00) |

<table>
<thead>
<tr>
<th>Signal Region</th>
<th>GGo1</th>
<th>GGo2</th>
<th>GGo3</th>
<th>GGo4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total bkg pre-fit</td>
<td>7</td>
<td>25</td>
<td>111</td>
<td>177</td>
</tr>
<tr>
<td>Fitted background events</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diboson</td>
<td>0.6 ± 0.2</td>
<td>2.2 ± 0.6</td>
<td>6.6 ± 2.2</td>
<td>6.8 ± 2.1</td>
</tr>
<tr>
<td>Z/γ*+jets</td>
<td>3.8 ± 1.3</td>
<td>10.9 ± 1.9</td>
<td>35 ± 6</td>
<td>39 ± 7</td>
</tr>
<tr>
<td>W+jets</td>
<td>0.9 ± 0.5</td>
<td>3.9 ± 1.3</td>
<td>16 ± 4</td>
<td>27 ± 6</td>
</tr>
<tr>
<td>tH(EW) + single top</td>
<td>0.2 ± 0.2</td>
<td>1.3 ± 0.8</td>
<td>28 ± 6</td>
<td>85 ± 14</td>
</tr>
<tr>
<td>Multi-jet</td>
<td>–</td>
<td>–</td>
<td>0.1^{+0.1}_{-0.1}</td>
<td>0.7^{+0.7}_{-0.7}</td>
</tr>
<tr>
<td>Total bkg post-fit</td>
<td>5.5 ± 1.5</td>
<td>18 ± 2.4</td>
<td>85 ± 9</td>
<td>159 ± 16</td>
</tr>
<tr>
<td>Observed</td>
<td>6</td>
<td>25</td>
<td>80</td>
<td>135</td>
</tr>
</tbody>
</table>

| (εσr)_{95}^{obs} [fb] | 0.05 | 0.12 | 0.16 | 0.18 |
| S^{95}_{obs}           | 7.1  | 17   | 22   | 25   |
| S^{95}_{exp}           | 6.9^{+2.3}_{-1.6} | 11^{+5}_{-2} | 25^{+10}_{-7} | 37^{+14}_{-10} |
| p_{0} (Z) | 0.49 (0.01) | 0.10 (1.28) | 0.50 (0.00) | 0.50 (0.00) |

No excess over the background events
Interpretation

Obtained significant improvement

By combining other analysis method, the excluded limits are

- Gluino direct decay: $m_{\tilde{g}} < 2.3$ TeV, $m_{\tilde{\chi}_1^0} < 1.2$ TeV at 95% CL.
- Gluino one-step decay: $m_{\tilde{g}} < 2.2$ TeV, $m_{\tilde{\chi}_1^0} < 1.2$ TeV at 95% CL.

Most stringent gluino limit
Discussion point 1

- The auxiliary material provided for the BDT reinterpretation
  - We published HepData materials [Link]
    - Provided xml files of BDT classifiers
      → ZeroLepton2018-SRBBDT-weight.tar.gz
    - Also provided analysis snipped code
      → ZeroLeptonBDT2018.cxx
      → By using this code, you can use same BDT classifiers.
    - Full likelihood is provided in the hepdata material

※ Signal events: fast sim (reco)
※ Background events: full reco

We provide data, SM total and signal yields in each BDT score bin and the acceptance and efficiencies of BDT SRs for simplified GG models.

→ I think it might help everyone.
Discussion point 2

- Discussing for example truth vs reco weights and BDT weight limits of applicability for reinterpretations.
  - Since we use the tail in distribution, “reco” samples are useful.
  - Produced many “reco” signal events for training: ~50k
  - We didn’t study “truth” samples : (if they are fine or not)
- BDT analysis is useful for the signals with similar kinematics
  - Applicable to evaluate other signal events

If jet multiplicity is same, the reinterpretation is fine in my analysis

Eg. Squark onestep: similar gluino direct

2021/4/9

ATLAS
\(\sqrt{s}=13\text{ TeV}, 139\text{ fb}^{-1}\)
0-leptons, 2-6 jets
All limits at 95% CL
Summary

SUSY is one of the promising theory
• New particles are predicted in this theory.

Gluino is one of the primary targets.
• Their pair production has large x-section. (strong interaction)

ML technique (BDT) is key to improve sensitivities.
• Developed training focused on $\Delta M$ (SUSY specific feature)
  • It can lead to improve all phase spaces.

No significant excess over the background estimation
• Gluino mass ($< 2.3$ TeV) is excluded at 95% CL.
Introduction
Discovered Higgs particle at LHC-ATLAS/CMS experiments

• However, there are still some remaining problems
  • Eg. Hierarchy problem, dark matter ... 

Symmetry between boson and fermion \(\rightarrow\) Super Symmetry

• Possible to solve dark matter, for instance.
LHC ATLAS 36.1 fb\(^{-1}\): Cut and count analysis

- Gluino direct decay \((\tilde{g} \rightarrow qq\tilde{\chi}_1^0)\): \(m_{\tilde{g}} < 2.0\) TeV, \(m_{\tilde{\chi}_1^0} < 1.0\) TeV
- Gluino one-step decay \((\tilde{g} \rightarrow qqW\tilde{\chi}_1^0)\): \(m_{\tilde{g}} < 1.9\) TeV, \(m_{\tilde{\chi}_1^0} < 0.90\) TeV
Training method

- Samples are divided into 2 categories: $iHalf=0$ or 1
  - If the events with $iHalf=0$ are used as training, the other events are used as testing.
- The training is also done to another one.
  - It can keep the statistics in the final fit.

- We don’t use training sample for the analysis itself.
  - Almost no effect on the bias due to overtraining.

![Diagram of training and applying BDT scores](image)
BDT score distribution

\[ m(\tilde{g}) = 2.2 \text{ TeV}, \ m(\tilde{\chi}_1^0) = 400 \text{ GeV} \]
\[ \Delta M = 1.8 \text{ TeV} \]

\[ m(\tilde{g}) = 1.8 \text{ TeV}, \ m(\tilde{\chi}_1^0) = 1000 \text{ GeV} \]
\[ \Delta M = 800 \text{ GeV} \]

Better separation (~30 %)

Better separation (~30 %)

Category 1 BDT score distribution
( High gluino, Low LSP mass target)

Category 3 BDT score distribution
( middle gluino, high LSP mass target)

- Good separation power between signal and background.
- A good agreement of data/MC in BDT score distribution.
Input variables

Candidate: Jet $p_T$, $\eta$, $m_{\text{eff}}$, $E_T^{\text{miss}}$, Aplanarity.

- Check the sensitivity by adding the variables
- Try to reduce the number of input variables

By adding Jet $P_T$, the sensitivity can be improved: $\sim 50\%$ $(m_\tilde{g}, m_\tilde{\chi}_1^0) = (2000, 800)$ GeV
- Correlation among jet $p_T$ is effective

Use 10 - 12 input variables: jet $p_T, \eta, E_T^{\text{miss}}, m_{\text{eff}},$ Aplanarity
Z+Jet estimation

Z+Jet events have high MET

→ Use $\gamma$+Jet events as CR

Obtain similar kinematics!

Compute normalization factor for Z+Jet events: 1.05

$\sqrt{s}=13$ TeV

**Figure B.1**: BDT score distributions in CRY for all training categories. The hatched red error bands indicate the MC statistical uncertainties. The bottom panel is the ratio of the observed data and the MC simulation.

Same BDT cut as SR

Jet $\gamma$

$P_T>150$ GeV

Jet $Z$

$E_T^{miss}$

$\nu$

$\nu$
Validation of correlation

Use MC as training → Need to describe correlations by the data

For example, $M_{\text{eff}}$ and Leading jet

- Distribution with right tail
  → Has positive correlation
- Agreement with data

\[
\text{corr}(x, y)_i = \frac{(x_i - \langle x \rangle) \cdot (y_i - \langle y \rangle)}{\sigma_x \sigma_y}
\]

\[
\langle x \rangle = \frac{\sum_i^N x_i}{N}, \quad \langle y \rangle = \frac{\sum_i^N y_i}{N}
\]

\[
\sigma_x = \sqrt{\frac{1}{N} \sum_i^N x_i^2}, \quad \sigma_y = \sqrt{\frac{1}{N} \sum_i^N y_i^2}
\]

Checked all correlations → Data/MC is fine
Systematic uncertainty
A summary of systematic uncertainty about event yields

\( \sqrt{s} = 13 \text{TeV}, 139 \text{ fb}^{-1} \)

Major uncertainty: MC stat, CR stat, Theo and Exp uncertainty
- Low CR stat due to tight BDT score cut → dominant
Simplified Model

We would like to say “Exclude gluino ~XX GeV ”
• Limit MSSM parameter space ? (108 dimension)
  • It is impossible…

Almost parameters does not depend on sensitivity in exp. side
• Possible to do the reduction

• Look at the space which only determines the acceptance of experiment
  • For example, SUSY mass * branching ratio

Evaluate excl. limit for all decay chains
~ Limit MSSM 108 dimension

→ Simplified Model approach
BDT configuration

Weight in training

• Each signal is weighted with each cross section.
  → It can avoid the bias of different phase spaces.

• The specific weights are not added in backgrounds
  → All background components are equally treated

Hyper parameter

• Use TMVA
• nTree = 800
• MaxDepth = 3
• nCuts = 20
• MinNodeSize = 2.5
• Shrinkage = 0.20
• Gradient Boost
5.5 Boosted Decision Tree (BDT)

BDT is one of the most famous multivariate analysis (MVA) methods in particle physics\textsuperscript{[112, 113]}. This is a binary tree structured classifier and one of ensemble learnings. The ensemble learning is composed of many “weak” learners and constructs one learner, which is called “strong learner”. Outputs of the strong learner are obtained from majority vote of weak learners considering weights. In the BDT, a decision tree as shown in Figure 5.11 is used as the weak learners and the weights are updated by “Boosting” algorithm. A classification criteria of the decision trees is determined by residual sum of squares, entropy information or Gini coefficient: this analysis uses entropy information to determine a classification criteria of the decision trees. By using many decision trees, the phase space can be split into several regions and each region is classified as signal or background. This analysis uses 800 decision trees in order to make strong learner. The considering weights are determined by “Gradient Boosting” algorithm.

The two orthogonal samples are needed in the BDT: training and testing samples. The training sample is provided to calculate the entropy information and determine the classification criteria of decision trees. The strong learner produced by the training samples is called “BDT classifier” and considers the non-linear correlation as shown in Figure 5.12. The output of the BDT classifier is called “BDT score”. This BDT score is applied to the testing sample and used in the analysis.

If the effective mass is large for the signals, the signal events also have jets from initial state radiations and are close to planar event shape. Hense, the broad aplanarity distribution is obtained for large missing transverse momentum.
Table 5.5: Input variables for the gluino direct decay. Circle (⊙) indicates the variable is used in input variables. The hyphen (-) indicates the variable is not used in input variables.

<table>
<thead>
<tr>
<th>Input variables</th>
<th>GGo1</th>
<th>GGo2</th>
<th>GGo3</th>
<th>GGo4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_T^{\text{miss}}$ [GeV]</td>
<td>⊙</td>
<td>-</td>
<td>⊙</td>
<td>⊙</td>
</tr>
<tr>
<td>$p_T(j)$ [GeV]</td>
<td>1st-4th</td>
<td>1st-4th</td>
<td>1st-4th</td>
<td>1st-4th</td>
</tr>
<tr>
<td>$\eta(j)$</td>
<td>1st-4th</td>
<td>1st-4th</td>
<td>1st-4th</td>
<td>1st-4th</td>
</tr>
<tr>
<td>Aplanarity</td>
<td>⊙</td>
<td>⊙</td>
<td>⊙</td>
<td>-</td>
</tr>
<tr>
<td>$m_{\text{eff}}$ [GeV]</td>
<td>⊙</td>
<td>⊙</td>
<td>⊙</td>
<td>⊙</td>
</tr>
<tr>
<td>Total number of input variables</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 5.6: Input variables for the gluino one-step decay. Circle (⊙) indicates the variable is used in input variables. The hyphen (-) indicates the variable is not used in input variables.

<table>
<thead>
<tr>
<th>Input variables</th>
<th>GGo1</th>
<th>GGo2</th>
<th>GGo3</th>
<th>GGo4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_T^{\text{miss}}$ [GeV]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>⊙</td>
</tr>
<tr>
<td>$p_T(j)$ [GeV]</td>
<td>1st-5th</td>
<td>1st-4th</td>
<td>1st-3rd, 5th - 6th</td>
<td>1st-4th</td>
</tr>
<tr>
<td>$\eta(j)$</td>
<td>1st-5th</td>
<td>1st-4th</td>
<td>1st-3rd, 5th - 6th</td>
<td>1st-4th</td>
</tr>
<tr>
<td>Aplanarity</td>
<td>⊙</td>
<td>⊙</td>
<td>⊙</td>
<td>-</td>
</tr>
<tr>
<td>$m_{\text{eff}}$ [GeV]</td>
<td>⊙</td>
<td>⊙</td>
<td>⊙</td>
<td>-</td>
</tr>
<tr>
<td>Total number of input variables</td>
<td>12</td>
<td>10</td>
<td>12</td>
<td>10</td>
</tr>
</tbody>
</table>

Figure 5.15: Illustration of training categories for gluino direct decay (a) and gluino one-step decay (b). For gluino direct decay (gluino one-step decay), the red point corresponds to GGo1 (GGo4), the magenta point corresponds to GGo2 (GGo2), the gold point corresponds to GGo3 (GGo3) and the light blue point corresponds to GGo4 (GGo4) training category. The blue dotted line shows the expected exclusion limit on the previous ATLAS result with the data of 36.1 fb$^{-1}$ [34].
How to use the BDT classifier

I think TMVA user-guide can help it

• How to implement xml file into your C++ code
• If needed, I will show my code to github ..