High1 Workshop on Particle, String and Cosmology
January 12~18, 2025, High1 Resort, Jeongseon, Korea

Machine learning studies for dark sector at future e+e- colliders

2025. 1. 16

Kihyeon Cho[1, 2]
[1]University of Science and Technology,
[2]Korea Institute of Science and Technology Information

Kihong Park, Kyungho Kim, Alexei Sytov and Kihyeon Cho, JKPS 84, 403-426 (2024).

Working with

- Kihong Park* (UST/KISTI)
- Kyungho Kim (KISTI)
- Alexei Sytov (INFN)

* Current address: IBS

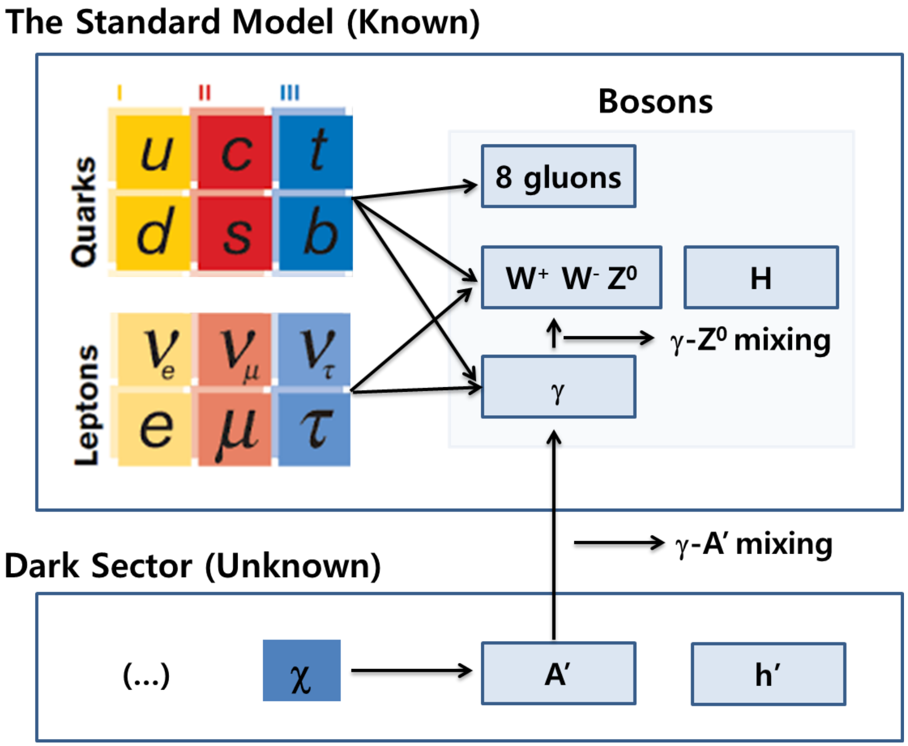
Contents

1. Theoretical motivation

▪ Dark photon

- A hypothetical particle **mediating the Standard Model (known) and dark sector (unknown)**
- Could decay into the **SM particles which are detectable**

<Dark sector scheme with the SM [1]>

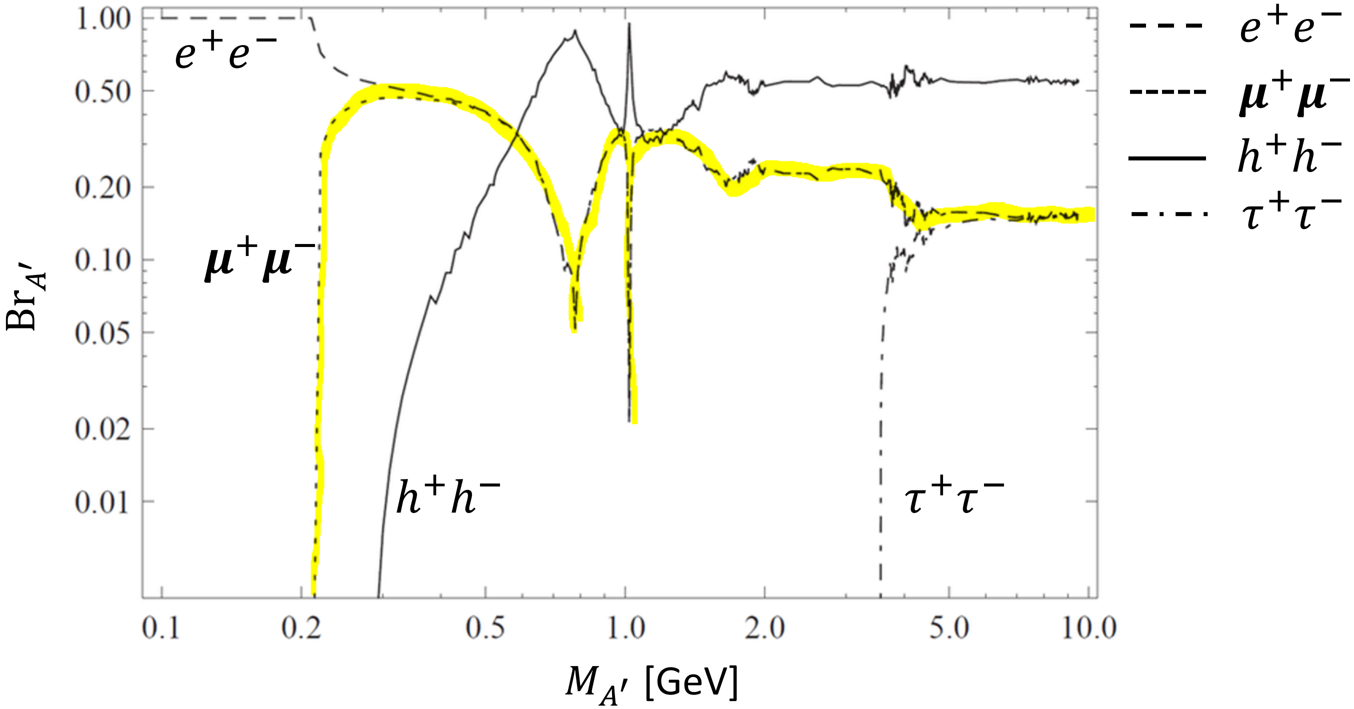


[1] Insung Yeo and Kihyeon Cho, J. Astron. Space Sci. 35, 67-74 (2018).

Dark photon decaying into the SM particles

- **Through kinetic mixing** between dark photon and photon

<Branching fraction of dark photon decaying into the SM particles [1]>



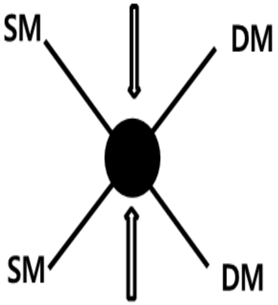
- Dark photon (A') decaying into a **muon pair ($\boldsymbol{\mu^+\mu^-}$)** [1, 2]

[1] Brian Shuve and Itay Yavin, Phys. Rev. D, 89, 113004 (2014).
[2] Insung Yeo and Kihyeon Cho, J. Astron. Space Sci. 35, 67-74 (2018).

Signal modes

❖ Simplified Model

<The simplified model between UV and EFT [1]>



Top-down approach

- Ultraviolet (UV) model: SUSY, ExtraDim, …
  $\{\mathcal{L}; m_{DM}, m_1, m_2, \ldots, g_1, g_2, \ldots\}$
- **Simplified model: SM+DM+Mediator particles**
  $\{\mathcal{L}; m_{DM}, m_{med}, g_{DM}, g_q\}$
- Effective field Theory (EFT) model: SM+DM particles
  $\{\mathcal{L}; m_{DM}, M_*\}$

Bottom-up approach

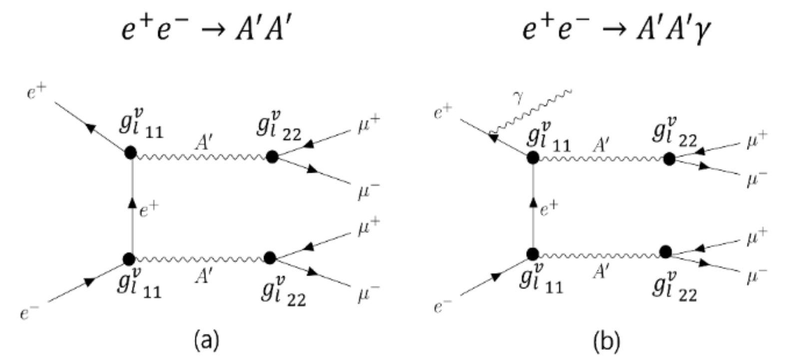
Kentarou Mawatari, KAIST-KAIX workshop (2019.7.15)

❖ Double dark photon modes

$e^+ e^- \to A'A'$ and $e^+ e^- \to A'A'\gamma$

$\mu^+\mu^-$ $\mu^+\mu^-$

$\mu^+\mu^-$ $\mu^+\mu^-$


$e^+e^- \to A'A'$ $e^+e^- \to A'A'\gamma$

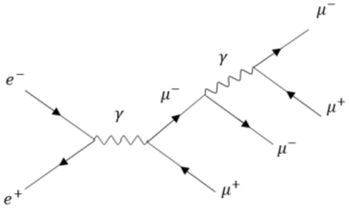
(a) (b)

K. Park and K. Cho, J. Astron. Space Sci. 38(1), 55-63, 38(1), 55-63 (2021).
K. Park, K. Kim and K. Cho, J. Astron. Space Sci. 39(1), 1-10 (2022).

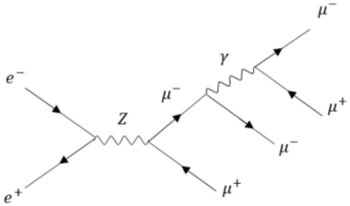
Background modes

$$e^+ e^- \rightarrow \mu^+\mu^-\mu^+\mu^-$$

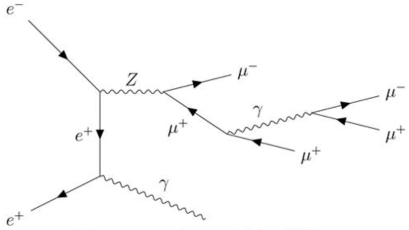
$$e^+ e^- \rightarrow \mu^+\mu^-\mu^+\mu^-\gamma$$



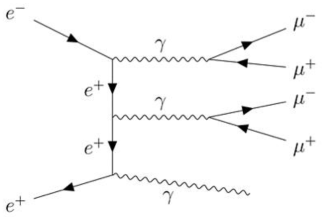
γ interaction



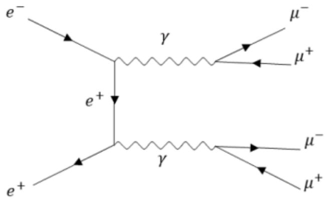
Z interaction



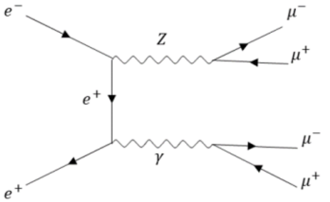
Z interaction with ISR γ



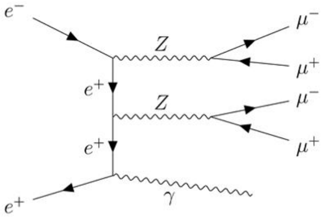
γ-γ interaction with ISR γ



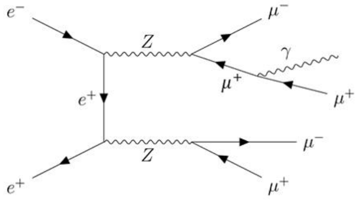
γ-γ interaction



Z-γ interaction



Z-Z interaction with ISR γ



Z-Z interaction

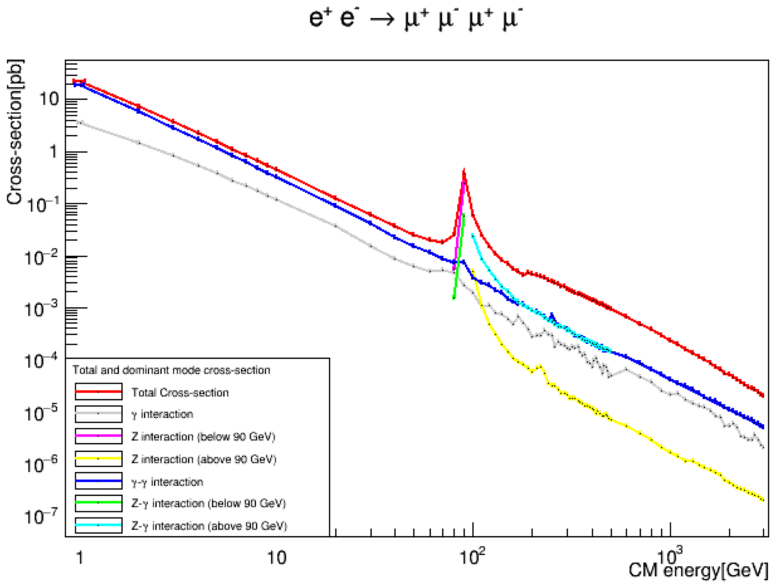
- Dominant the SM background ($\sigma_{SM} \gg \sigma_{DS}$)

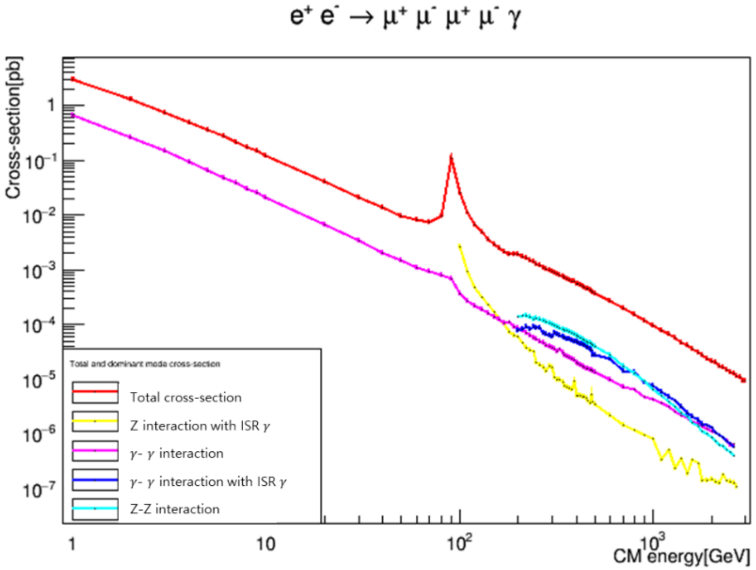
Kihong Park, Kyungho Kim and Kihyeon Cho, J. Astron. Space Sci. 39(1), 1-10 (2022).

- Background modes:

  $e^+e^- \rightarrow \mu^+\mu^-\mu^+\mu^-$ and $e^+e^- \rightarrow \mu^+\mu^-\mu^+\mu^-\gamma$

  - Cross-section depending on the center-of-mass energy ($\sqrt{s}$)





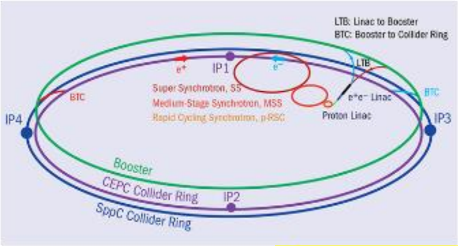
- Strategy

  - Complex process & sufficient event generation ⇒ **KISTI-5 supercomputer**
  - Should be removed as many as possible ⇒ **Machine learning**

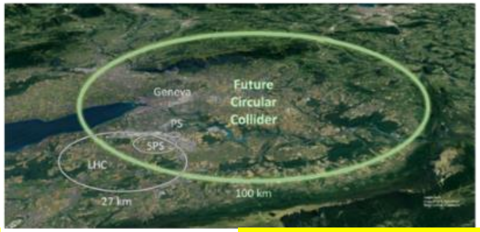
Kihong Park, Kyungho Kim and Kihyeon Cho, J. Astron. Space Sci. 39(1), 1-10 (2022).

▪ Future electron-positron collider experiments

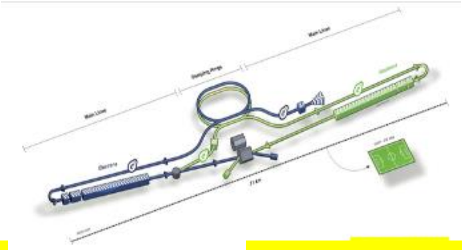
<table>
<tr><th>Spec.</th><th>CEPC/CEPC</th><th>FCC-ee/IDEA</th><th colspan="3">ILC/ILD</th></tr>
<tr><td rowspan="2">Accelerator/ Detector</td><td>Circular Electron-Positron Collider (CEPC)/CEPC</td><td>Future Circular Collider (FCC)-ee/ Innovative Detector for Electron-positron Accelerators (IDEA)</td><td colspan="3">International Linear Collider(ILC)/ International Linear Detector(ILD)</td></tr>
<tr><td>(2030~, IHEP/China)



Circular Electron-Positron Collider http://cepc.ihep.ac.cn/</td><td>(2038~, CERN/Switzerland)



Future Circular Collider https://home.cern/science/accelerators/future-circular-collider</td><td colspan="3">(2034~, Japan)

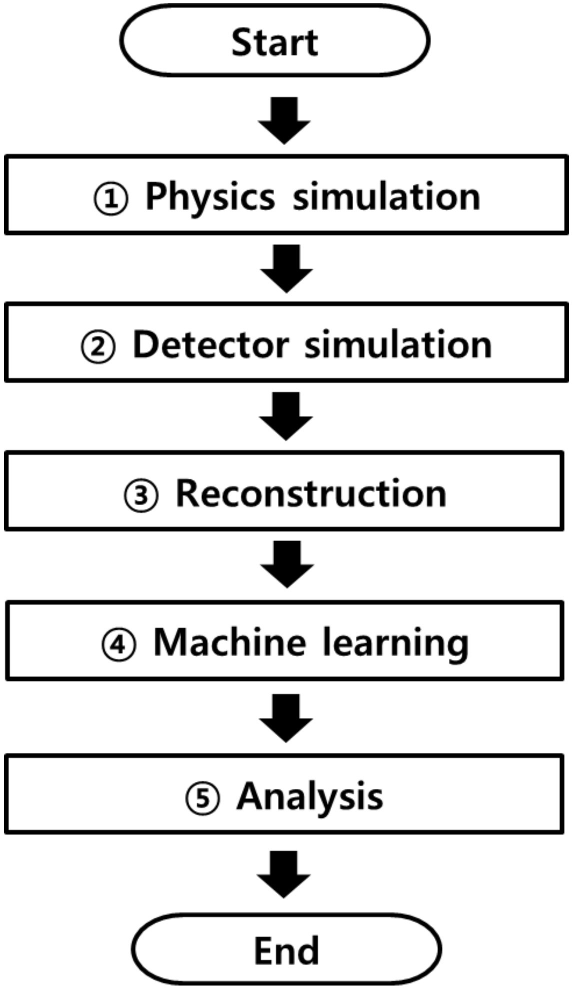


International Linear Collider https://linearcollider.org/</td></tr>
<tr><td>Circum. or length [km]</td><td>100</td><td>97.75</td><td>20.5</td><td>31</td><td>40</td></tr>
<tr><td>$\sqrt{s}$ [GeV]</td><td>91, 160, **240**</td><td>91, 160, 250, 350</td><td>250</td><td>500</td><td>1000</td></tr>
<tr><td>Place</td><td>IHEP/China</td><td>CERN/Switzerland</td><td colspan="3">Japan</td></tr>
<tr><td>Data taking</td><td>2030~</td><td>2038~</td><td colspan="3">2034~</td></tr>
</table>

Zyla P. A. et al., Review of Particle Physics, Prog. Theor. Exp. Phys. **2020(8)**, (2020).

2. Methodology

▪ Flowchart of the study



Start

① Physics simulation

② Detector simulation

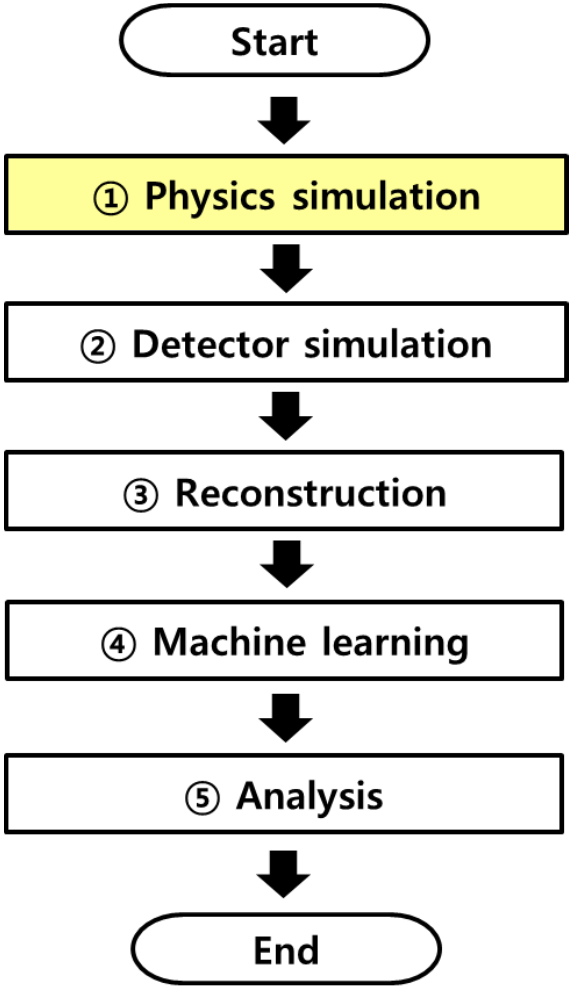
③ Reconstruction

④ Machine learning

⑤ Analysis

End

① Physics simulation



Start

① Physics simulation

② Detector simulation

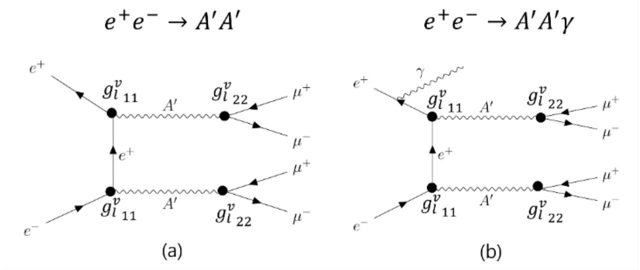
③ Reconstruction

④ Machine learning

⑤ Analysis

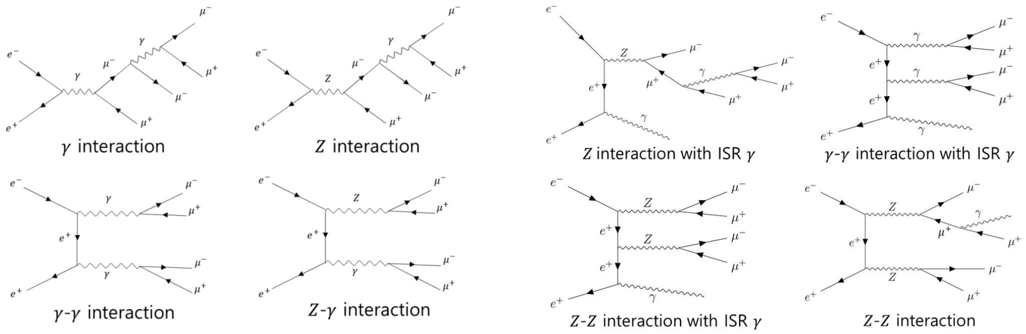
End

- Signal event generation using MadGraph5



(a) (b)

- Background event generation using MadGraph5



γ interaction

Z interaction

Z interaction with ISR γ

γ-γ interaction with ISR γ

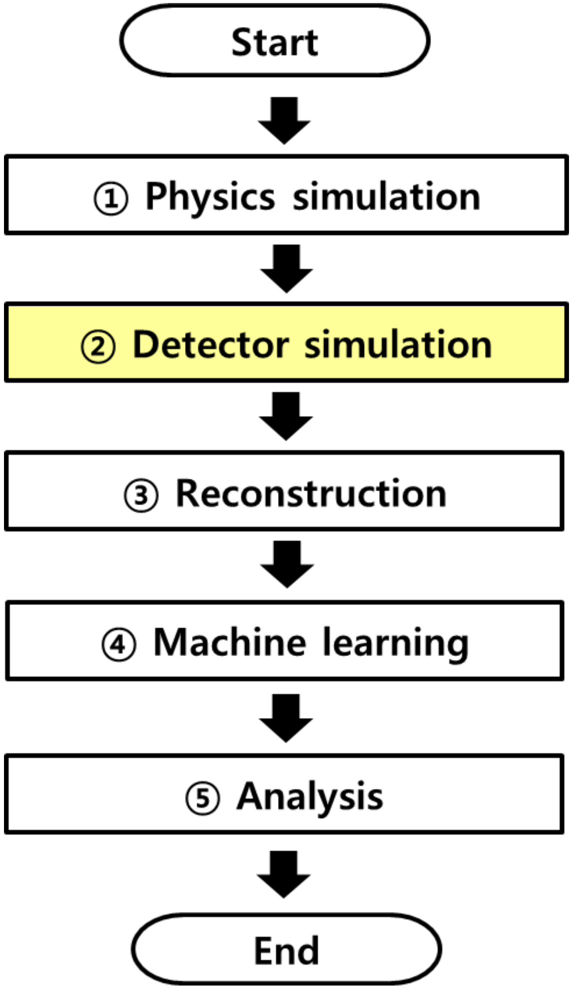
γ-γ interaction

Z-γ interaction

Z-Z interaction with ISR γ

Z-Z interaction

② Detector simulation



Start

① Physics simulation

② Detector simulation

③ Reconstruction

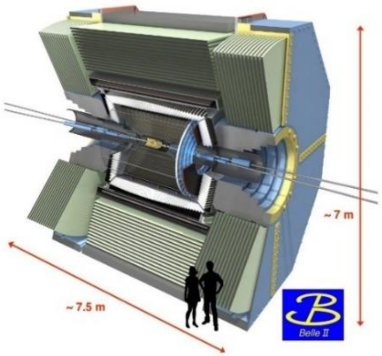
④ Machine learning

⑤ Analysis

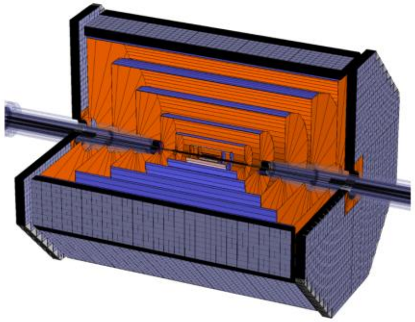
End

- Using **Delphes** for Belle II, CEPC, FCC-ee and ILC

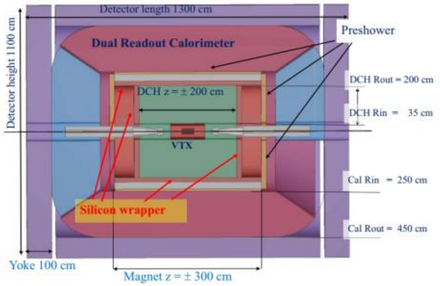
SuperKEKB/**Belle II**



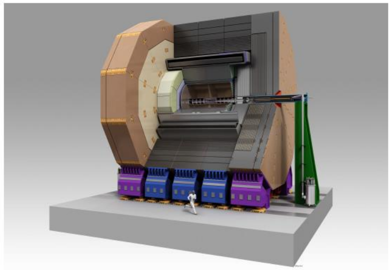
CEPC/**CEPC**



FCC-ee/**IDEA**



ILC/**ILD**



③ Reconstruction

Start

① Physics simulation

② Detector simulation

③ Reconstruction

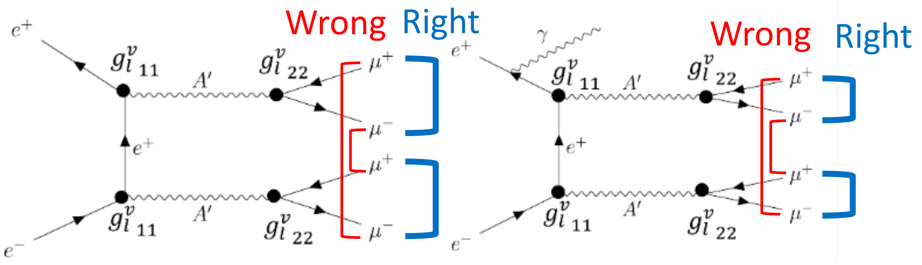
④ Machine learning

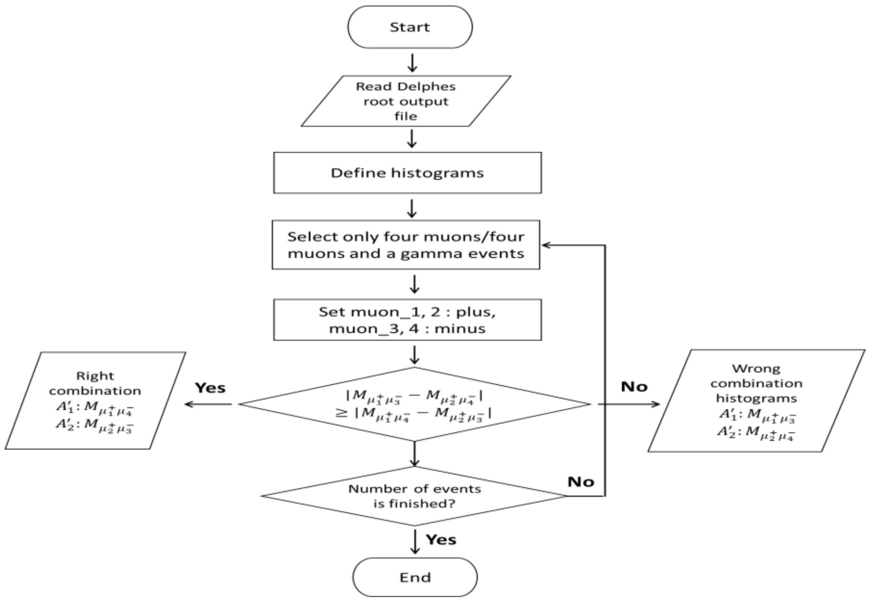
⑤ Analysis

End

- Dark photon selection
  - by choosing Right combination of muon pair which its invariant mass difference is smaller than the other.
  - If $\left|M_{\mu_1^+\mu_3^-} - M_{\mu_2^+\mu_4^-}\right| \geq \left|M_{\mu_1^+\mu_4^-} - M_{\mu_2^+\mu_3^-}\right|$ in an event, then $\mu_1^+\mu_4^-$ and $\mu_2^+\mu_3^-$ are right combination.
  - Meanwhile, $\mu_1^+\mu_3^-$ and $\mu_2^+\mu_4^-$ are wrong combination.

$e^+e^- \to A'A'$ $\qquad$ $e^+e^- \to A'A'\gamma$





④ Machine learning

Start

① Physics simulation

② Detector simulation

③ Reconstruction

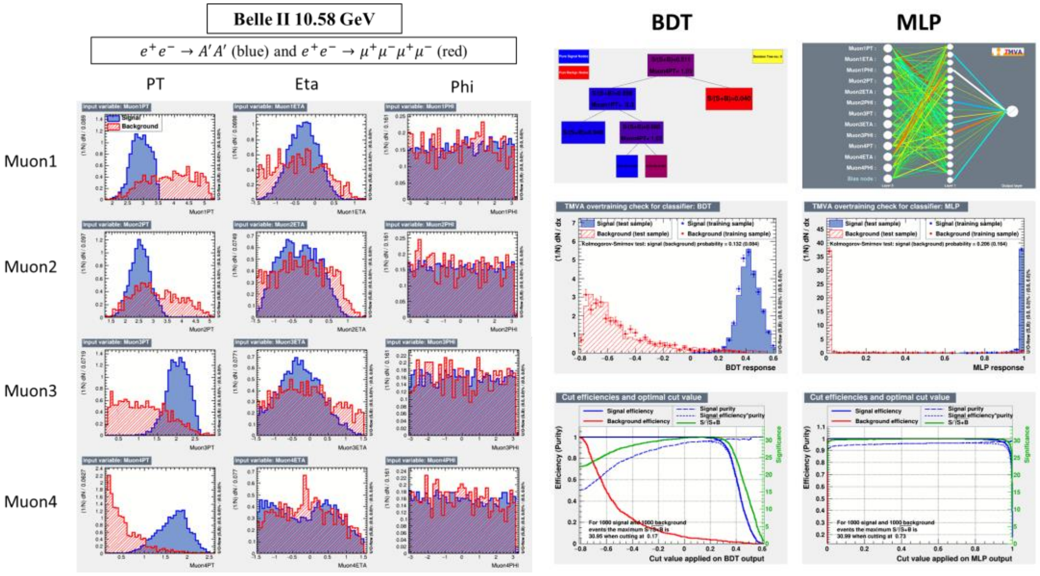
④ Machine learning

⑤ Analysis

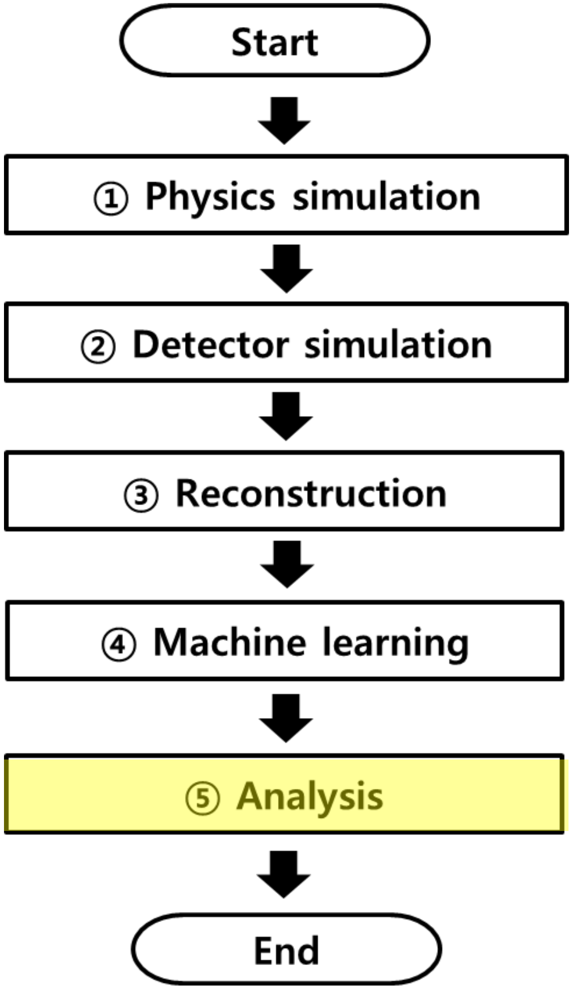
End

- Examined Kaggle and TMVA method
- Compared classifiers performance
- Choose the **BDT classifier** based on **TMVA method**

(An example) Belle II with A’A’ mode based on TMVA



⑤ Analysis



Start

① Physics simulation

② Detector simulation

③ Reconstruction

④ Machine learning

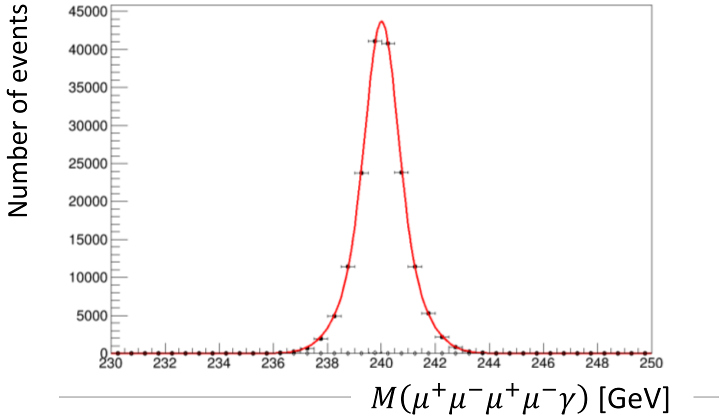
⑤ Analysis

End

- Detector resolution ($\boldsymbol{\sigma}$)
- Estimate number of signal events ($\boldsymbol{N_{exp}}$)
- Detector efficiency ($\boldsymbol{\varepsilon_{det}}$)

(An example) CEPC 240 GeV $A'A'\gamma$

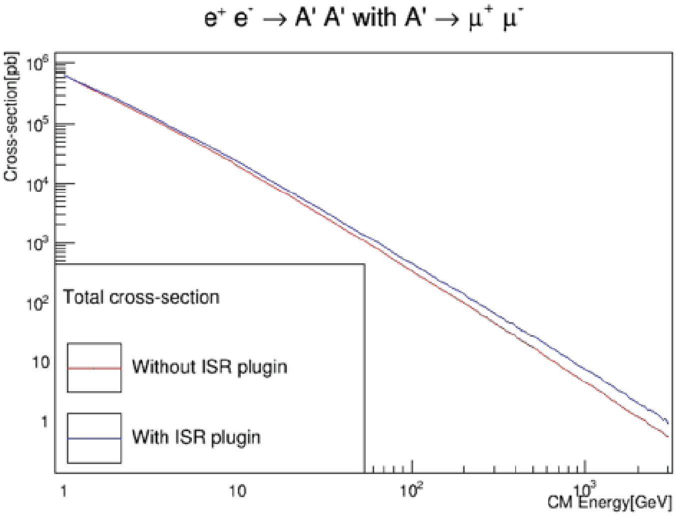
Fit to $M(\mu^+\mu^-\mu^+\mu^-\gamma)$ signal events after machine learning

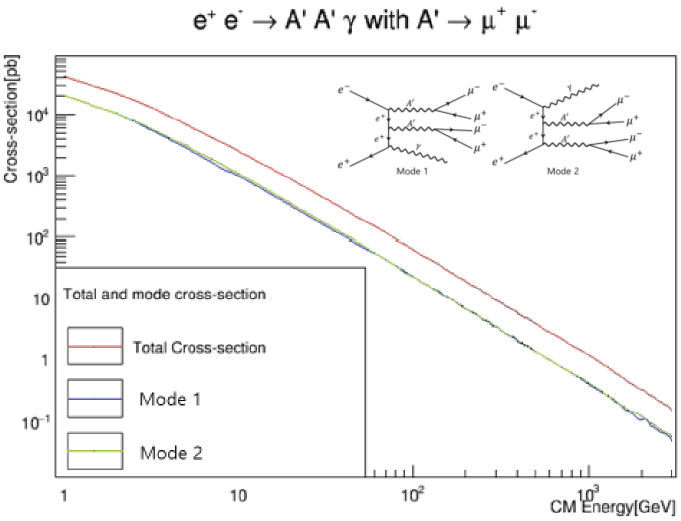


- Fitting: Double Gaussian function
- $m = 240.0 \pm 0.0$
- $\sigma_1 = 0.5614 \pm 0.0066$
- $\boldsymbol{\sigma_2 = 1.1232 \pm 0.0062}$
- $\boldsymbol{N_{exp} = 168879 \pm 2503}$
- $\boldsymbol{\varepsilon_{det}}$ = **33.8±0.5 %**

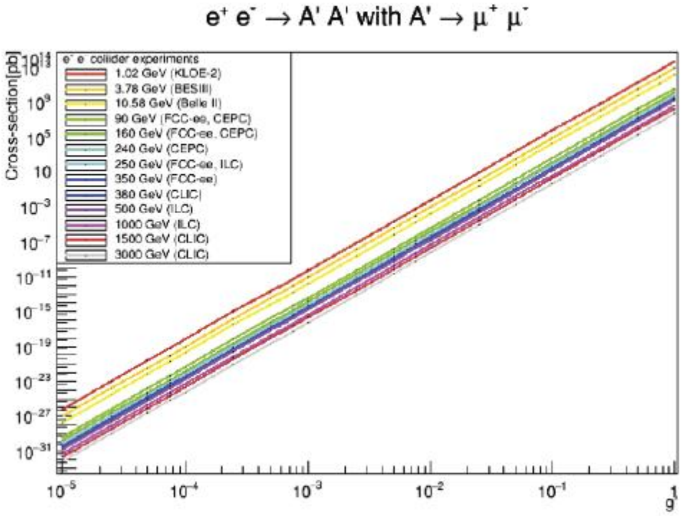
3. Cross-section dependence

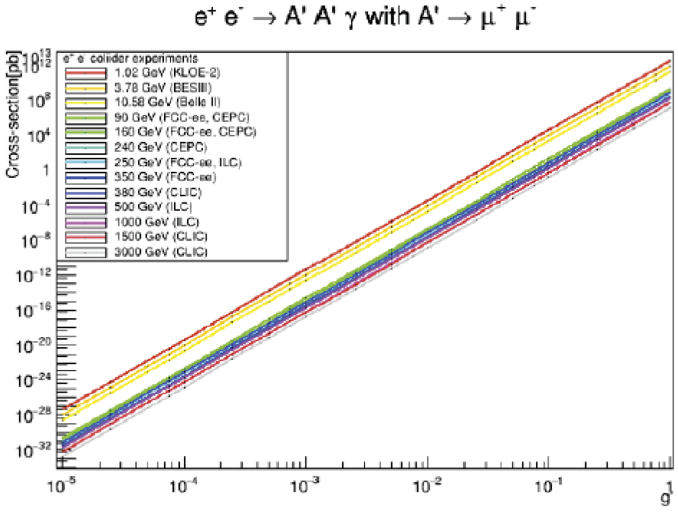
- Cross-section depending on the **Center-of-mass energy ($\sqrt{s}$)**


$e^+ e^- \to A' A'$ with $A' \to \mu^+ \mu^-$


$e^+ e^- \to A' A' \gamma$ with $A' \to \mu^+ \mu^-$

- Cross-section depending on the **coupling constant (g')**

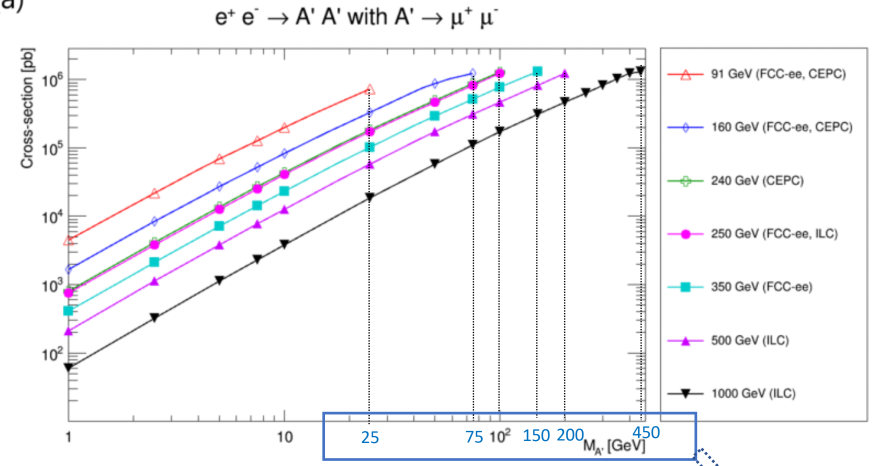

$e^+ e^- \to A' A'$ with $A' \to \mu^+ \mu^-$


$e^+ e^- \to A' A' \gamma$ with $A' \to \mu^+ \mu^-$

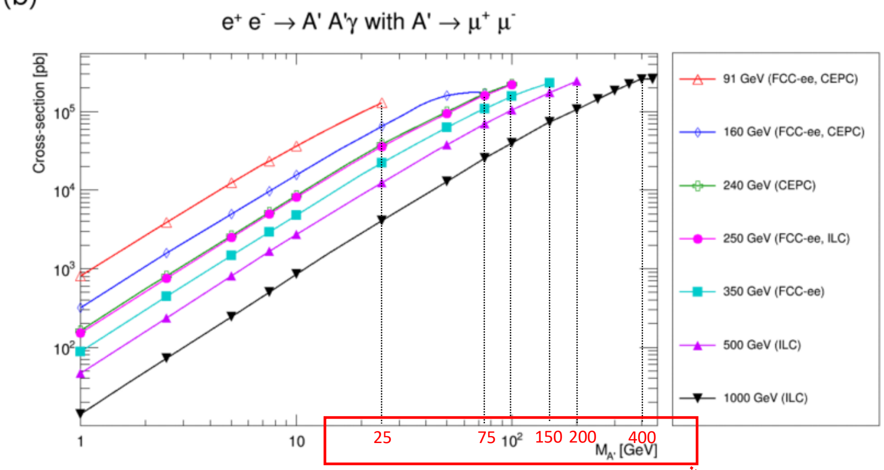
Kihong Park, Kyungho Kim and Kihyeon Cho, J. Astron. Space Sci. 39(1), 1-10 (2022).

- Cross-section depending on the **dark photon mass ($m_{A'}$)**

(a)



(b)



- Dark photon masses with the highest cross-section

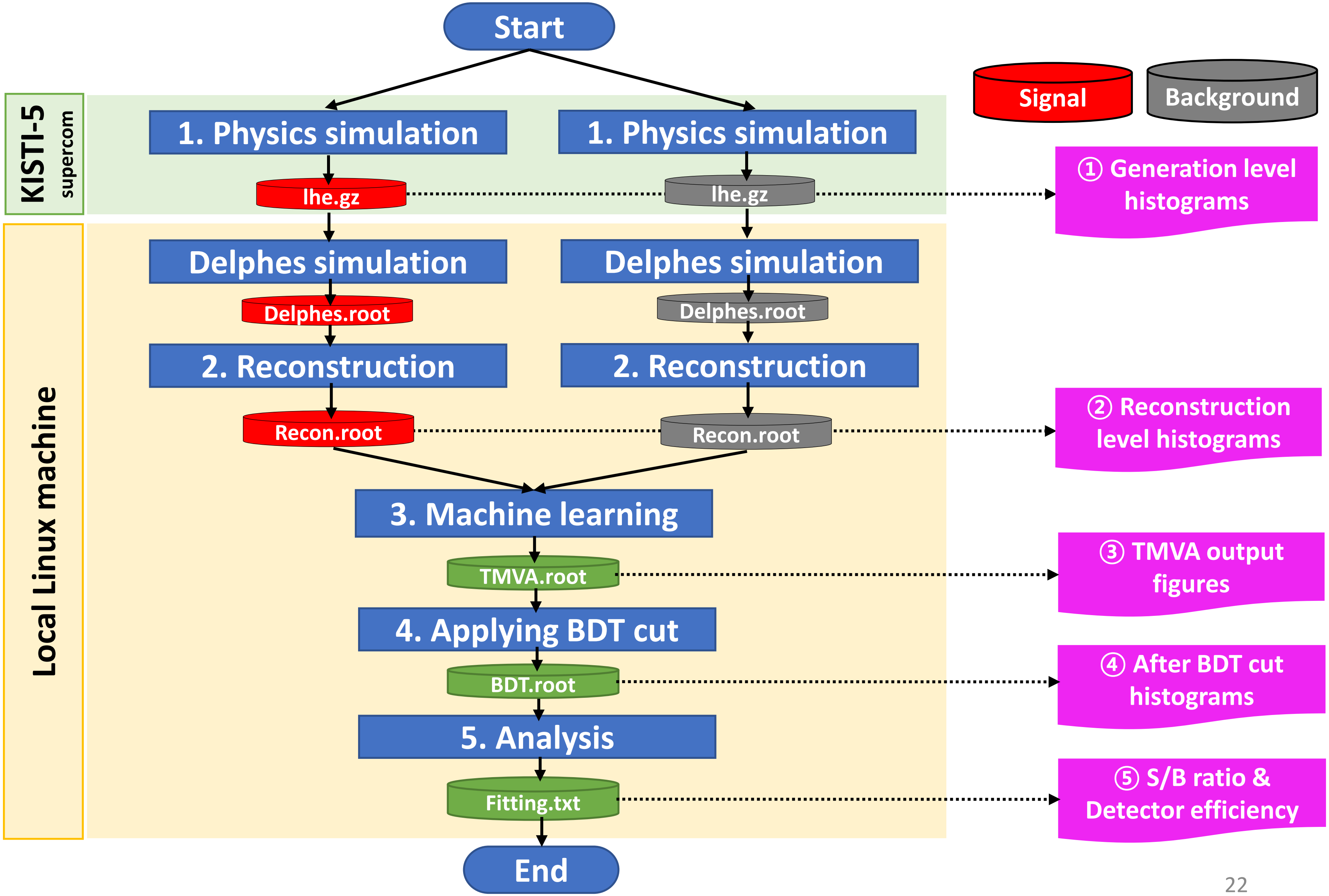
Accelerator/Detector	**$\sqrt{s}$ [GeV]**	**$m_{A'}$ [GeV]**	
		$e^+e^- \to A'A'$	$e^+e^- \to A'A'\gamma$
CEPC/CEPC	91	25	25
	160	75	75
	240	100	100
FCC-ee/IDEA	91	25	25
	160	75	75
	250	100	100
	350	150	150
ILC/ILD	250	100	100
	500	200	200
	1000	450	400

Kihong Park, Kyungho Kim, Alexei Sytov and Kihyeon Cho, JKPS 84, 403-426 (2024).

4. Study of $e^+e^- \to A'A'$ and $A'A'\gamma$ using machine learning

Flowchart of the study

- TMVA: Toolkit for Multivariate Data Analysis
- BDT: Boosted Decision Trees



Start

Signal

Background

KISTI-5 supercom

1. Physics simulation

1. Physics simulation

lhe.gz

lhe.gz

① Generation level histograms

Local Linux machine

Delphes simulation

Delphes simulation

Delphes.root

Delphes.root

2. Reconstruction

2. Reconstruction

Recon.root

Recon.root

② Reconstruction level histograms

3. Machine learning

TMVA.root

③ TMVA output figures

4. Applying BDT cut

BDT.root

④ After BDT cut histograms

5. Analysis

Fitting.txt

⑤ S/B ratio & Detector efficiency

End

Configuration of the study

- Configuration of generation and machine learning

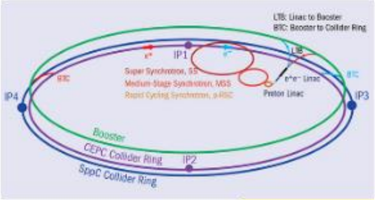
Process	Parameters	Signal modes $e^+e^- \to A'A'(\gamma)$	Background modes $e^+e^- \to \mu^+\mu^-\mu^+\mu^-(\gamma)$
Generation	Event generator	MadGraph5 ver. 2.6.6	
	Model	The simplified model	The Standard Model
	No. of events	**1,000,000**	~**1,000,000**
	$\sqrt{s}$ [GeV]	91, 160, 240, 250, 350, 500, 1000	91, 160, 240, 250, 350, 500, 1000
	Decay width [GeV]	6.7×10^{-6}	-
	Coupling constants $(g^{\nu}_{l}{}_{11}, g^{\nu}_{l}{}_{22})$	**0.1**	-
Machine learning (TMVA)	Model	Boosted Decision Trees	
	Input variables	$p_{T_{\mu_1}}, p_{T_{\mu_2}}, p_{T_{\mu_3}}, p_{T_{\mu_4}}, \eta_{\mu_1}, \eta_{\mu_2}, \eta_{\mu_3}, \eta_{\mu_4},$ $\phi_{\mu_1}, \phi_{\mu_2}, \phi_{\mu_3}, \phi_{\mu_4}, m_{A'_1}, m_{A'_2}, p_{T_{A'_1}}$ and $p_{T_{A'_2}}$ **(16 variables)** **for both $e^+e^- \to A'A'$ and $e^+e^- \to A'A'\gamma$**	
		$p_{T_{\gamma}}, \eta_{\gamma}, \phi_{\gamma}$ and E_{γ} (**4 more variables** for γ-included decay modes) **for $e^+e^- \to A'A'\gamma$**	

1. $e^{+}\, e^{-} \to A'A'$

$\quad\quad\quad\quad\quad\quad\quad\quad \llcorner\!\rightarrow \mu^{+}\mu^{-}$

$\quad\quad\quad\quad\quad\quad \llcorner\!\rightarrow \mu^{+}\mu^{-}$

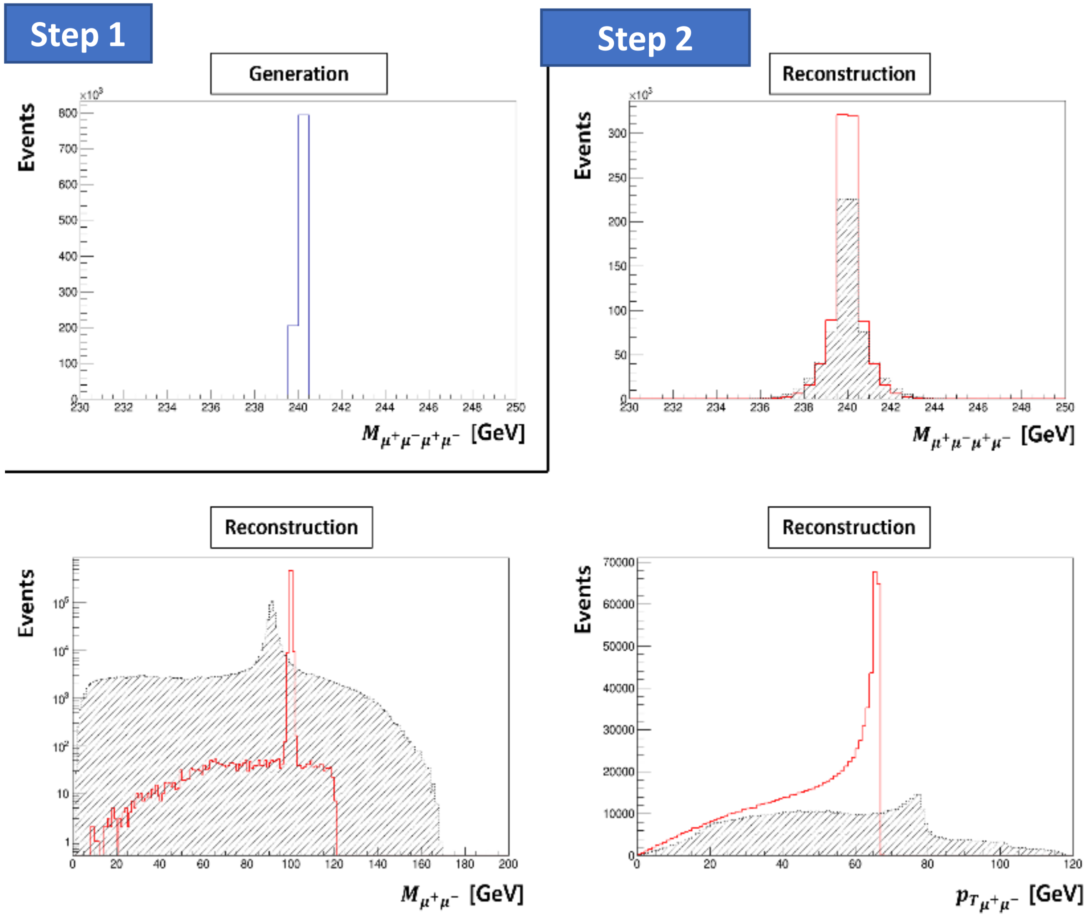
(2030~, IHEP/China)



Circular Electron-Positron Collider

Image copyright IHEP

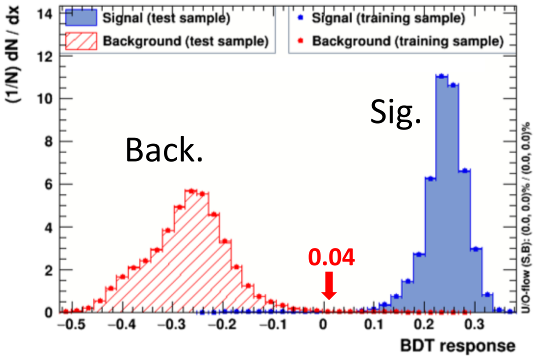
- **(An example) CEPC at $\sqrt{s}$ = 240 GeV, $m_{A'}$ = 100 GeV**



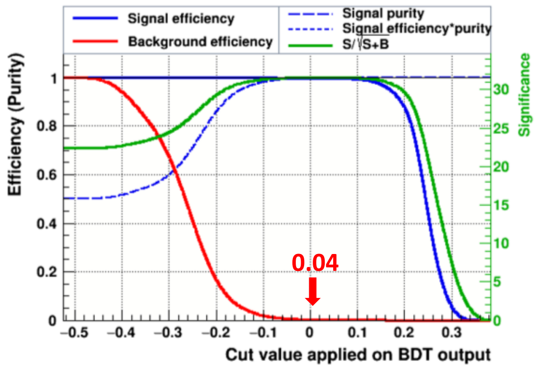
K.Park, K, Kim, A. Sytov and K. Cho, *JKPS* 84 (2024) 403-425.

- CEPC at $\sqrt{s}$ = 240 GeV, $m_{A'}$ = 100 GeV

Step 3 BDT response

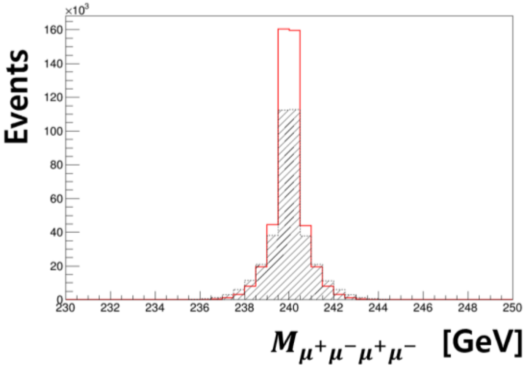


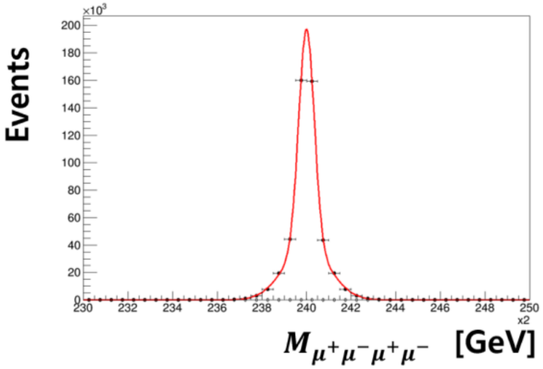
Cut efficiency



- Figure of merit: S/$\sqrt{S+B}$
  - S: Number of signal
  - B: Number of background
- Optimal BDT cut: 0.04
  - BDT where a figure of merit is maximized

Step 4 Before BDT cut → After BDT cut





Step 5

- Fitting: Double Gaussian function
- $m = 240.0 \pm 0.0$
- $\sigma_1 = 0.3466 \pm 0.0012$
- $\boldsymbol{\sigma_2 = 1.0183 \pm 0.0025}$
- $\boldsymbol{N_{exp} = 470580 \pm 1598}$
- ε_{det} = 94.1±0.3 %

	Before BDT cut	After BDT cut	S/B ↑
Signal	472,155	470,580	-
Background	388,956	**596**	-
Sig. / Back.	1.2	789.6	**658.0**

K.Park, K, Kim, A. Sytov and K. Cho, *JKPS* 84 (2024) 403-425.

▪ The summary of $e^+e^- \to A'A'$ mode

- Performed same procedure to CEPC(91, 160, 240 GeV), FCC-ee(91, 160, 250, 350 GeV) and ILC(250, 500, 1000 GeV)
- Signal($\mu^+\mu^-\mu^+\mu^-$)/Background($\mu^+\mu^-\mu^+\mu^-$) improved **a factor of O(100~10000).**

Shown as an example

Accelrator/ Detector	$\sqrt{s}$ [GeV]	$m_{A'}$ [GeV]	S/B improvement	Signal $M(\mu^+\mu^-\mu^-\mu^-)$		
				N_{exp}	Width [GeV]	Detector efficiency [%]
CEPC/CEPC	91	25	847.3	**461860 $\pm$ 1024**	0.4524 $\pm$ 0.0007	92.4 $\pm$ 0.2
	160	75	680.6	**471670 $\pm$ 1069**	0.5599 $\pm$ 0.0004	94.3 $\pm$ 0.2
	240	100	658.0	**470580 $\pm$ 1598**	1.0183 $\pm$ 0.0025	94.1 $\pm$ 0.3
FCC-ee/IDEA	91	25	11781.2	**438260 $\pm$ 662**	0.0887 $\pm$ 0.0001	87.7 $\pm$ 0.1
	160	75	876.3	**395660 $\pm$ 629**	0.2196 $\pm$ 0.0002	79.1 $\pm$ 0.1
	250	100	1455.8	**417540 $\pm$ 646**	0.5892 $\pm$ 0.0006	83.5 $\pm$ 0.1
	350	150	486.0	**408960 $\pm$ 639**	1.0838 $\pm$ 0.0012	81.8 $\pm$ 0.1
ILC/ILD	250	100	713.6	**355080 $\pm$ 1287**	1.0901 $\pm$ 0.0030	71.0 $\pm$ 0.3
	500	200	709.6	**366552 $\pm$ 1776**	3.0506 $\pm$ 0.0133	73.3 $\pm$ 0.4
	1000	450	288.1	**369880 $\pm$ 1113**	6.4568 $\pm$ 0.0162	74.0 $\pm$ 0.2

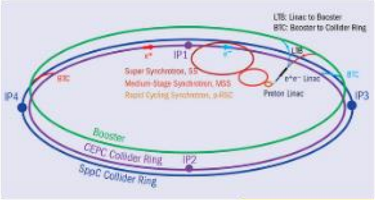
K.Park, K, Kim, A. Sytov and K. Cho, *JKPS* 84 (2024) 403-425.

2. $e^+ e^- \to A'A'\gamma$

$\quad\quad\quad\quad\quad\quad\quad\quad \hookrightarrow \mu^+\mu^-$

$\quad\quad\quad\quad\quad\quad \hookrightarrow \mu^+\mu^-$

(2030~, IHEP/China)



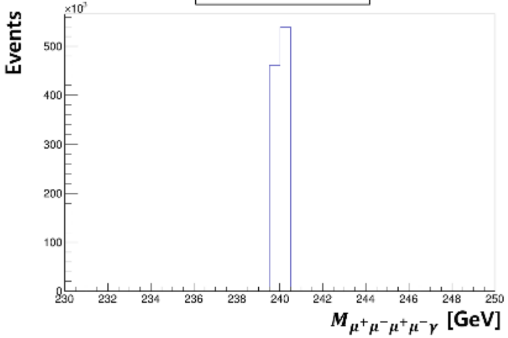
Circular Electron-Positron Collider

Image copyright IHEP

- **(An example) CEPC at $\sqrt{s}$ = 240 GeV, $m_{A'}$ = 100 GeV**

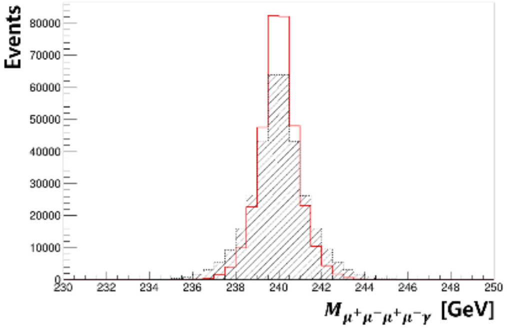
Step 1

Generation

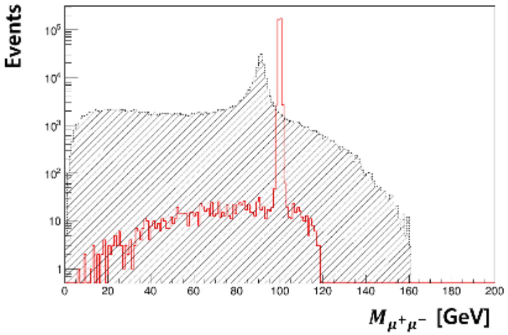


Step 2

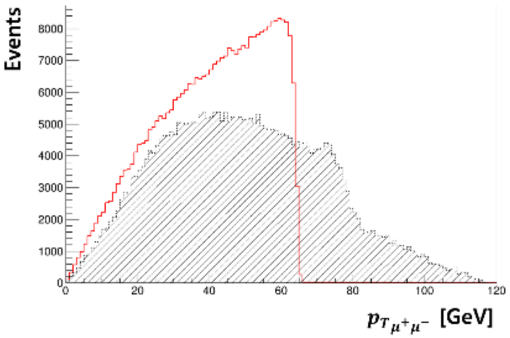
Reconstruction



Reconstruction



Reconstruction

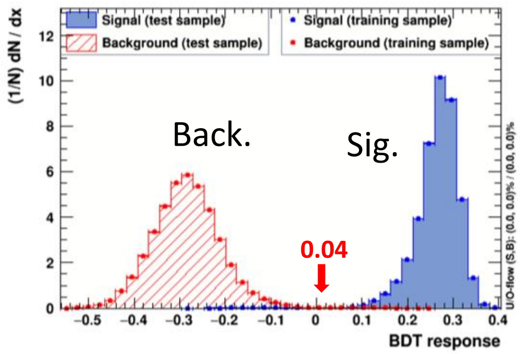


K.Park, K, Kim, A. Sytov and K. Cho, *JKPS* 84 (2024) 403-425.

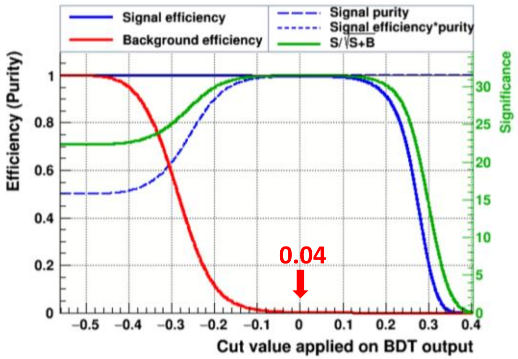
▪ CEPC at $\sqrt{s}$ = 240 GeV, $m_{A'}$ = 100 GeV

Step 3

BDT response



Cut efficiency

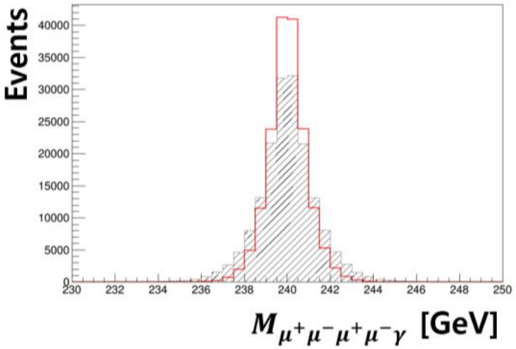


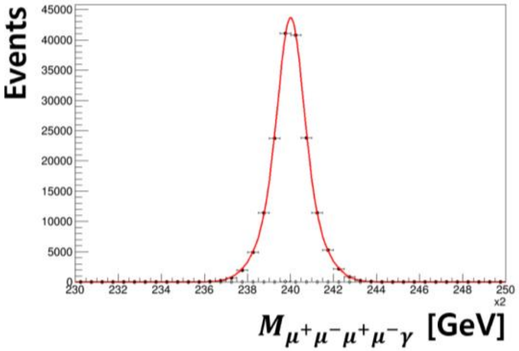
- Figure of merit: S/$\sqrt{S+B}$
  - S: Number of signal
  - B: Number of background
- Optimal BDT cut: 0.04
  - BDT where a figure of merit is maximized

Step 4

Before BDT cut

After BDT cut





Step 5

- Fitting: Double Gaussian function
- $m = 240.0 \pm 0.0$
- $\sigma_1 = 0.5614 \pm 0.0066$
- $\boldsymbol{\sigma_2 = 1.1232 \pm 0.0062}$
- $\boldsymbol{N_{exp} = 168879 \pm 2503}$
- ε_{det} = 33.8±0.5 %

	Before →	After	S/B ↑
Signal	169,580	168,879	-
Background	170,997	**238**	-
Sig. / Back.	1.0	709.6	**709.6**

K.Park, K, Kim, A. Sytov and K. Cho, *JKPS* 84 (2024) 403-425.

The summary of $e^+e^- \rightarrow A'A'\gamma$ mode

- Performed same procedure to CEPC(91, 160, 240 GeV), FCC-ee(91, 160, 250, 350 GeV) and ILC(250, 500, 1000 GeV)
- Signal($\mu^+\mu^-\mu^+\mu^-\gamma$)/Background($\mu^+\mu^-\mu^+\mu^-\gamma$) improved **a factor of O(100~1000)**

Shown as an example

Accelrator/ Detector	$\sqrt{s}$ [GeV]	$m_{A'}$ [GeV]	S/B improvement	Signal $M(\mu^+\mu^-\mu^-\mu^-\gamma)$		
				N_{exp}	Width [GeV]	Detector efficiency [%]
CEPC/CEPC	91	25	731.0	**161812 ± 4688**	0.8255 ± 0.0092	32.4 ± 0.9
	160	75	717.1	**97386 ± 2744**	0.7125 ± 0.0077	19.5 ± 0.5
	240	100	709.6	**168879 ± 2503**	1.1232 ± 0.0062	33.8 ± 0.5
FCC-ee/IDEA	91	25	409.2	**151360 ± 389**	0.3825 ± 0.0007	30.3 ± 0.1
	160	75	1328.9	**80929 ± 284**	0.3090 ± 0.0008	16.2 ± 0.1
	250	100	3694.7	**155180 ± 394**	0.6688 ± 0.0012	31.0 ± 0.1
	350	150	404.6	**152560 ± 391**	1.0709 ± 0.0019	30.5 ± 0.1
ILC/ILD	250	100	727.9	**27223 ± 3210**	1.1605 ± 0.0320	5.4 ± 0.6
	500	200	532.0	**52136 ± 1246**	2.7050 ± 0.0420	10.4 ± 0.2
	1000	450	350.9	**77053 ± 662**	6.9436 ± 0.0473	15.4 ± 0.1

K.Park, K, Kim, A. Sytov and K. Cho, *JKPS* 84 (2024) 403-425.

5. Conclusion

▪ Conclusion

- We studied dark photons using the simplified model at future e+e- colliders: CEPC, FCC-ee and ILC.
- Using BDT based on TMVA, we reduced the SM background significantly for the double dark photon modes:

  $e^+e^- \to A'A'$ and $e^+e^- \to A'A'\gamma$

- Estimated expected of signal events and detector efficiencies.

⇒ **For dark photon searches, this study provides:**

✓**A reference at future electron-positron experiments**

✓**A methodology at current experiments (ex. Belle II)**

ACKNOWLEDGMENTS

- The major institutional R&D program, KISTI (No. K25L2M2C3-01)
- This study was also supported by National Supercomputing Center with supercomputing resources including technical support (TS-2024-IN-0017).

Thank you for your attention.

cho@kisti.re.kr