A machine learning approach for mass composition analysis with TALE-SD data

Ryuhei Arimura[1,*] on behalf of the Telescope Array Collaboration
(a complete list of authors can be found at the end of the proceedings)

[1]*Graduate School of Science, Osaka City University, Osaka, Japan*

Abstract. The TALE experiment is a Telescope Array Low-energy Extension constructed to observe cosmic rays with energies down to $10^{16.5}$ to clarify the origin of the second knee and the energy of the galactic to extragalactic CRs transition. TALE consists of 10 high-elevation fluorescence detectors and 80 scintillation counters in an area of 21 km^2. The key of data interpretation is the mass composition of cosmic rays, and we report on a machine learning approach of mass composition analysis that utilizes waveform data of TALE scintillation counters.

1 Introduction

1.1 Telescope Array experiment

The Telescope Array (TA) experiment is the largest cosmic ray observatory in the northern hemisphere designed to detect ultra high energy cosmic rays is deployed in Millard County, Utah, USA. It mainly observes ultra high energy (above $\sim 10^{18}$ eV) cosmic rays, using both Fluorescence Detectors (FD) and scintillator Surface Detectors (SD). There are 507 SDs with about 700 km^2 effective detection area and the 3 FD stations, which are called Black Rock Mesa (BRM), Long Ridge (LR) and Middle Drum (MD). The FD telescopes cover the sky above the SD array from 3° to 31° in elevation. The general map of the TA detectors is shown in Fig. 1a.

1.2 Telescope Array Low-energy Extension (TALE)

The Telescope Array Low-energy Extension (TALE) detectors are designed for the energy threshold of the experiment to be well below $10^{16.5}$ eV. We call the bend in the energy spectrum around 10^{17} eV as the "second knee" and consider that the feature suggests a galatic-to-extragalactic transition of cosmic ray origin. The motivation of the TALE experiment is to clarify the origin of the feature measuring the energy spectrum and mass composition of cosmic rays around 10^{17} eV. To observe lower energy showers, TALE utilizes 10 FDs with elevation angles higher than that of TA-FD, from 31° up to 59°. In addition 80 SDs with denser grids are also installed near the MD site. The full details of the detectors may be found in [1]. For this analysis, we report on a machine learning approach of mass composition analysis that utilizes waveform data of TALE-SD array. The layout of TALE detectors is shown in Fig. 1b.

*e-mail: m21sa003@st.osaka-cu.ac.jp

2 TALE-SD mass composition analysis

2.1 Monte Carlo (MC) simulation and Event selection criteria

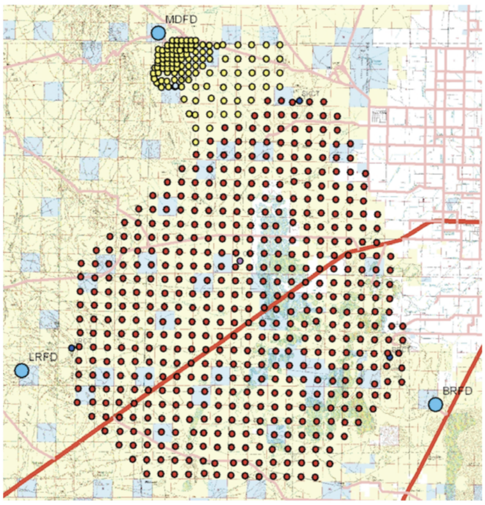
We generate cosmic ray air showers using CORSIKA-based MC simulation code developed for TA [2] in this analysis. QGSJETII-04 [3] and Geant 4 are used for air shower simulation and detector response simulation, respectively. The MC input parameters are given in Table 1. The simulated air shower is reconstructed using TALE-SD software [4]. Only events that pass the event selection criteria shown in Table 2a below are used in the analysis. The number of Monte-Carlo events after the selection are 17,121 events (proton) and 17,262 (iron), respectively.

Table 1: MC simulation dataset

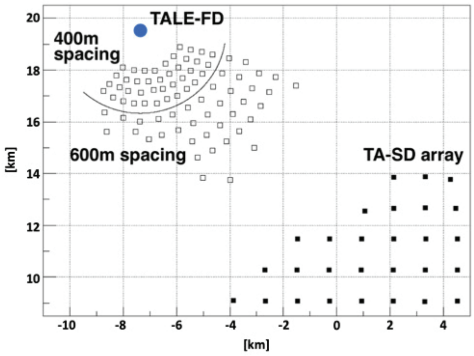
primary	proton	iron
MC simulation	CORSIKA	
interaction model	QGSJETII-04	
primary energy	$E = 10^{18.0}$ eV (fixed)	
zenith angle	$\theta = 30.0°$ (fixed)	
azimuthal angle	$0° \leqq \varphi \leqq 360°$	
TALE-SDs	78SDs, 2019/10/01-2020/01/29	
core position	refer to Fig. 2	
Number of generated event	100,000	100,000
Number of selected event	17,121	17,262

2.2 Composition-sensitive parameters

The advantages of the TALE-SD array are its high statistics and high uniform sensitivity compared to FD. If we analyze mass composition analysis using TALE-SD data, we can apply the result to cosmic ray energy spectrum and



(a)



(b)

Figure 1: (a)General map of the TA experiment site. The locations of the TA-SDs are shown as red circles and the locations of the three TA-FD stations are indicated by light blue circles. Yellow circles show the location of the TALE-SDs. The TALE-FD station is located at the MD site. (b)The layout of the TALE detectors. Open square boxes (□) represent the locations of the TALE-SDs and blue circle (•) correspond to the MD / TALE-FD station. Black boxes (■) repesent the locations of the TA-SDs.

anisotropy analysis.
A schematic view of an air shower arriving at the surface is shown in Fig. 3. As shown in this figure, air showers arrive at the surface with 3 characteristics: lateral distribution, curvature, and thickness. Using TALE-SD software and simulated proton and iron air showers, we search and extract 22 parameters [5, 6] that exhibit 3 characteristics. The parameters' histograms are shown in Fig. 4.

Table 2: (a)Event selection criteria (b)$N_{\text{thickness}}$ selection (SD selection)

(a)

$N_{\text{SD}} \geqq 5,\ N_{\text{thickness}} \geqq 1$
$\chi^2_{\text{geometry}}/\text{d.o.f.} \leqq 4,\ \chi^2_{\text{ldf}}/\text{d.o.f.} \leqq 2$
$\left(\sigma_\theta^2 + \sin^2\theta\,\sigma_\phi^2\right)^{0.5} \leqq 2.5$ deg.
$\sigma_{S_{600}}/S_{600} \leqq 0.25$
$0^\circ \leqq \theta_{\text{rec.}} \leqq 45^\circ$

(b)

$t_{\text{recorded waveform}} \leqq 2.56\,\mu$s (128 bin)
$N_{\text{bin}}(\geqq 15$ FADC count) is more than 2.
$N_{\text{bin}}(\geqq 45$ FADC count) is more than 1.
No saturation in recorded signal (upper/lower)
within the range of 400 m $\leqq r \leqq$ 700 m.

https://doi.org/10.1051/epjconf/202328302011

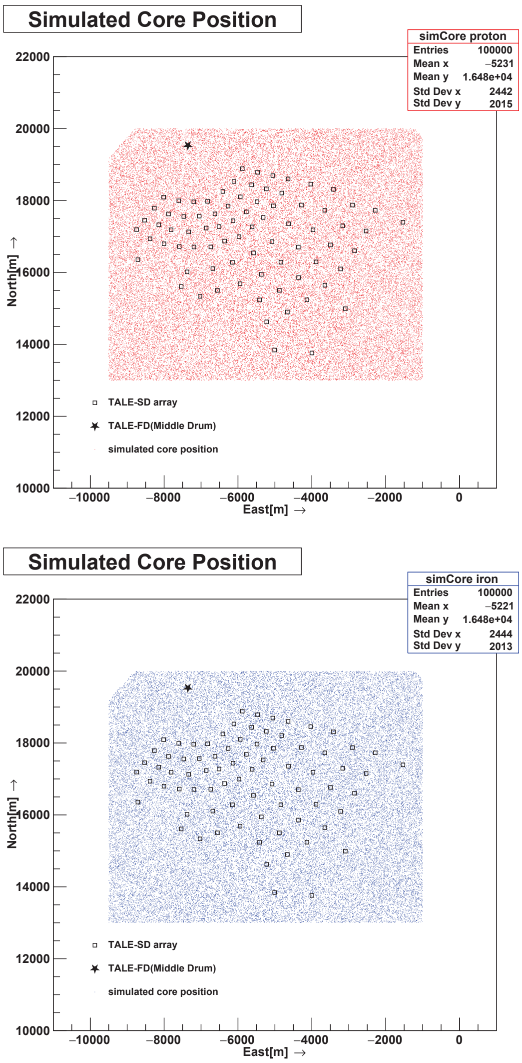


Figure 2: Simulated core position distribution for TALE-SD array. □, ⋆ and •/• indicates TALE-SD array, TALE-FD and simulated core positions, respectively. red one is (a)proton and blue one is (b)iron. "Entries" indicates the number of generated events.

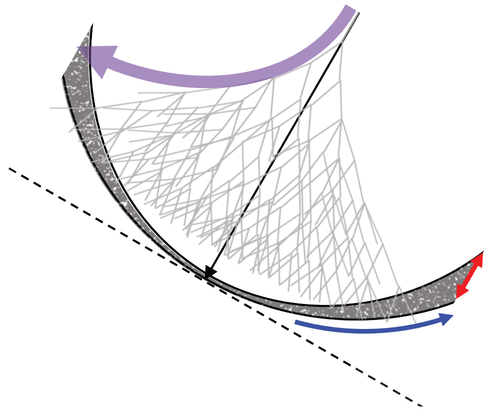


Figure 3: A schematic view of the development of an air shower around the surface. Dashed line and black arrow are shower plane and shower axis, respectively. Purple, red and blue arrow indicates "lateral distribution", "thickness" and "curvature" of air showers, respectively.

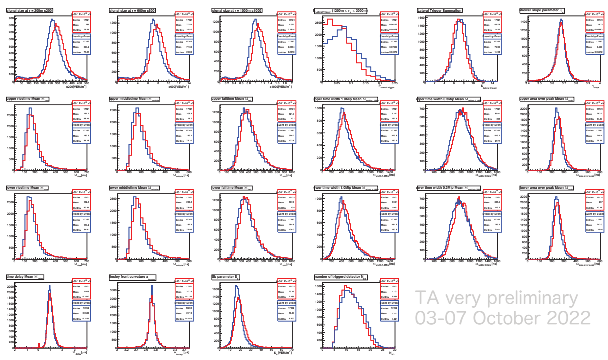


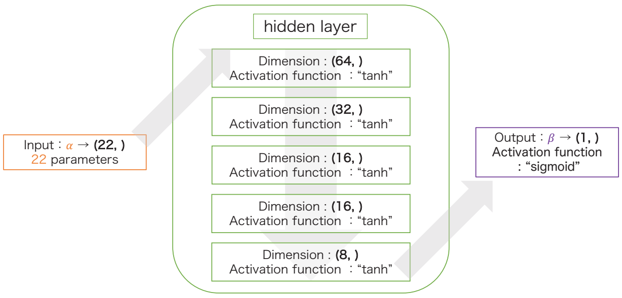
Figure 4: 22 parameters by the TALE-SD array histograms. Proton MC is shown with red lines and iron MC is shown with blue them. "Entries" indicates the number of events.

3 Machine learning approach

3.1 Neural network details

We have developed a machine learning model to take advantage of the high-statistics TALE-SD data for mass composition analysis. The machine learning model that we use is a binary classification to discriminate between proton and iron cosmic ray events by MC simulation as shown the Fig. 5.
The input vector α consists of 22 normalized parameters. Output value β is a numerical values from 0 to 1, with 0 being labeled as proton and 1 as iron. The output value β is determined to be proton if it is less than 0.5, and iron if it is greater than 0.5.
The machine learning model details are shown in Table 6.

Figure 5: Machine learning model outline



We use two activation functions (a hyperbolic tangent and a sigmoid). To eliminate learning bias, the same number of events are prepared for proton and iron air showers. All data are split into training, validation, and test data at a ratio of 8:1:1.

Figure 6: Machine learning model details

primary	proton	iron
All data	17120	17120
Training data	13696	13696
Validation data	1712	1712
Test data	1712	1712
Optimizer	Adam	
Loss function	Cross-Entropy Error (CEE)	

3.2 The prediction of the machine learning model

The performance of the trained machine learning model was evaluated using all test data. The result is shown in Fig. 7. The red line represents the model's prediction on test data where the answer is proton ($\beta = 0$), and the blue line represents the model's prediction on test data where the answer is iron ($\beta = 1$). Classification matrix for all test data are shown in Table 3. The percentage of those that predicted that the answer is proton is 68.9%, and the percentage of those that predicted that the answer is iron is 65.4%. The overall accuracy is determined by the average of these values, which is 67.1%.

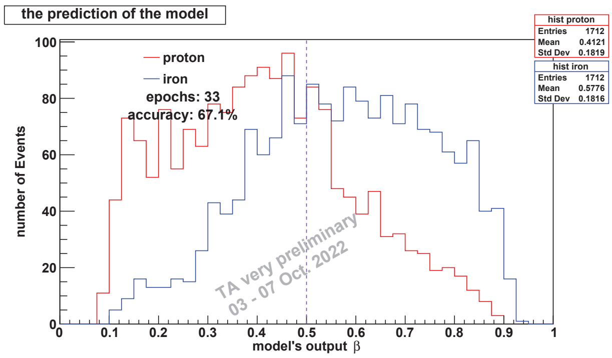


Figure 7: The response distribution of the test data to the machine learning model. The purple dashed line is drawn at $\beta = 0.5$.

4 Conclusion

We search and extract 22 parameters that depend on primary cosmic ray for cosmic ray mass composition analysis using TALE-SD array. Using these parameters, we have developed a binary classification machine learning model to discriminate proton and iron cosmic rays by MC simulation. The classification accuracy of the machine learning model is currently 67.1%. To improve the accuracy, it is planned to add information for each detector, search for new parameters, and use a Graph Neural Network.

Acknowledgements

The Telescope Array experiment is supported by the Japan Society for the Promotion of Science(JSPS) through Grants-in-Aid for Priority Area 431, for Specially Promoted Research JP21000002, for Scientific Research (S) JP19104006, for Specially Promoted Research JP15H05693, for Scientific Research (S) JP19H05607, for Scientific Research (S) JP15H05741, for Science Research (A) JP18H03705, for Young Scientists (A) JPH26707011, and for Fostering Joint International Research (B) JP19KK0074, by the joint research program of the Institute for Cosmic Ray Research (ICRR), The University of Tokyo; by the Pioneering Program of RIKEN for the Evolution of Matter in the Universe (r-EMU); by the U.S. National Science Foundation awards PHY-1607727, PHY-1712517, PHY-1806797, PHY-2012934, and PHY-2112904; by the National Research Foundation of Korea (2017K1A4A3015188, 2020R1A2C1008230, & 2020R1A2C2102800) ; by the Ministry of Science and Higher Education of the Russian Federation under the contract 075-15-2020-778, IISN project No. 4.4501.18, and Belgian Science Policy under IUAP VII/37 (ULB). This work was partially supported by the grants ofThe joint research program of the Institute for Space-Earth Environmental Research, Nagoya University and Inter-University Research Program of the Institute for Cosmic Ray Research of University of Tokyo. The foundations of Dr. Ezekiel R. and Edna Wattis Dumke, Willard L. Eccles, and George S. and Dolores Doré Eccles all helped with generous donations. The State of Utah supported the project

Table 3: Classification matrix for all test data in Fig. 7. The percentage of cosmic-ray events considered to be of proton origin if the output value β is less than 0.5, and of iron origin if the output value is greater than 0.5, is shown in brackets.

answer \ prediction	proton($0 \leqq \beta < 0.5$)	iron($0.5 \leqq \beta \leqq 1$)	total
proton($0 \leqq \beta < 0.5$)	1180 (68.9%)	532 (31.1%)	1712
iron ($0.5 \leqq \beta \leqq 1$)	593 (34.6%)	1119 (65.4%)	1712
total	1773	1651	3424

through its Economic Development Board, and the University of Utah through the Office of the Vice President for Research. The experimental site became available through the cooperation of the Utah School and Institutional Trust Lands Administration (SITLA), U.S. Bureau of Land Management (BLM), and the U.S. Air Force. We appreciate the assistance of the State of Utah and Fillmore offices of the BLM in crafting the Plan of Development for the site. Patrick A. Shea assisted the collaboration with valuable advice and supported the collaborations efforts. The people and the officials of Millard County, Utah have been a source of steadfast and warm support for our work which we greatly appreciate. We are indebted to the Millard County Road Department for their efforts to maintain and clear the roads which get us to our sites. We gratefully acknowledge the contribution from the technical staffs of our home institutions. An allocation of computer time from the Center for High Performance Computing at the University of Utah is gratefully acknowledged.

References

[1] S. Ogio, for the Telescope Array collaboration, "Telescope Array Low-energy Extension (TALE) hybrid", *PoS* **ICRC2019** (2019) 375.

[2] B.T. Stokes, R. Cady, D. Ivanov, J.N. Matthews and G.B. Thomson, "Dethinning extensive air shower simulations" *Astroparticle Physics* **35** (2012) 759-766.

[3] S. Ostapchenko, "Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: QGSJET-II model", *Phys. Rev. D* **83** (2011) 014018.

[4] K. Sato, for the Telescope Array collaboration, "Cosmic ray energy spectrum in the 2nd knee region measured by the TALE-SD array", *PoS* **ICRC2021** (2021) 395.

[5] P. Abreu et al., "A Search for Photons with Energies Above 2×10^{17} eV Using Hybrid Data from the Low-Energy Extensions of the Pierre Auger Observatory", *Astrophys. J.* **424** (1994) 491.

[6] R. U. Abbasi et al., "Mass composition of ultrahigh-energy cosmic rays with the Telescope Array Surface Detector data", *Phys. Rev. D* **99** (2019) 022002.

Full Authors List: Telescope Array Collaboration

R.U. Abbasi[1], T. Abu-Zayyad[1,2], M. Allen[2], Y. Arai[3], R. Arimura[3], E. Barcikowski[2], J.W. Belz[2], D.R. Bergman[2], S.A. Blake[2], I. Buckland[2], B.G. Cheon[4], M. Chikawa[5], T. Fujii[6], K. Fujisue[5], K. Fujita[3], R. Fujiwara[3], M. Fukushima[5], G. Furlich[2], N. Globus[7*], R. Gonzalez[2], W. Hanlon[2], N. Hayashida[8], H. He[7], K. Hibino[8], R. Higuchi[5], K. Honda[9], D. Ikeda[8], N. Inoue[10], T. Ishii[9], H. Ito[7], D. Ivanov[2], H. Iwakura[11], A. Iwasaki[3], H.M. Jeong[12], S. Jeong[12], C.C.H. Jui[2], K. Kadota[13], F. Kakimoto[8], O. Kalashev[14], K. Kasahara[15], S. Kasami[16], S. Kawakami[3], K. Kawata[5], I. Kharuk[14], E. Kido[7], H.B. Kim[4], J.H. Kim[2**], J.H. Kim[2], S.W. Kim[12], Y. Kimura[3], I. Komae[3], Y. Kubota[11], V. Kuzmin[14***], M. Kuznetsov[14,17], Y.J. Kwon[18], K.H. Lee[12], B. Lubsandorzhiev[14], J.P. Lundquist[2,19], H. Matsumiya[3], T. Matsuyama[3], J.N. Matthews[2], R. Mayta[3], I. Myers[2], S. Nagataki[7], K. Nakai[3], R. Nakamura[11], T. Nakamura[20], T. Nakamura[11], Y. Nakamura[11], A. Nakazawa[11], E. Nishio[16], T. Nonaka[5], H. Oda[3], S. Ogio[5], M. Ohnishi[5], H. Ohoka[5], Y. Oku[16], T. Okuda[21], Y. Omura[3], M. Ono[7], A. Oshima[22], S. Ozawa[23], I.H. Park[12], M. Potts[2****], M.S. Pshirkov[14,24], J. Remington[2], D.C. Rodriguez[2], C. Rott[2,12], G.I. Rubtsov[14], D. Ryu[25], H. Sagawa[5], N. Sakaki[5], T. Sako[5], N. Sakurai[3], K. Sato[3], T. Seki[11], K. Sekino[5], P.D. Shah[2], Y. Shibasaki[11], N. Shibata[16], T. Shibata[5], J. Shikita[3], H. Shimodaira[5], B.K. Shin[25], H.S. Shin[5], D. Shinto[16], J.D. Smith[2], P. Sokolsky[2], B.T. Stokes[2], T.A. Stroman[2], K. Takahashi[5], M. Takamura[26], M. Takeda[5], R. Takeishi[5], A. Taketa[27], M. Takita[5], Y. Tameda[16], K. Tanaka[28], M. Tanaka[29], Y. Tanoue[3], S.B. Thomas[2], G.B. Thomson[2], P. Tinyakov[14,17], I. Tkachev[14], H. Tokuno[30], T. Tomida[11], S. Troitsky[14], R. Tsuda[3], Y. Tsunesada[3,31], S. Udo[8], T. Uehama[11], F. Urban[32], D. Warren[7], T. Wong[2], M. Yamamoto[11], K. Yamazaki[22], K. Yashiro[26], F. Yoshida[16], Y. Zhezher[5,14], and Z. Zundel[2]

[1] *Department of Physics, Loyola University Chicago, Chicago, Illinois 60660, USA*
[2] *High Energy Astrophysics Institute and Department of Physics and Astronomy, University of Utah, Salt Lake City, Utah 84112-0830, USA*
[3] *Graduate School of Science, Osaka Metropolitan University, Sugimoto, Sumiyoshi, Osaka 558-8585, Japan*
[4] *Department of Physics and The Research Institute of Natural Science, Hanyang University, Seongdong-gu, Seoul 426-791, Korea*
[5] *Institute for Cosmic Ray Research, University of Tokyo, Kashiwa, Chiba 277-8582, Japan*
[6] *The Hakubi Center for Advanced Research and Graduate School of Science, Kyoto University, Kitashirakawa-Oiwakecho, Sakyo-ku, Kyoto 606-8501, Japan*
[7] *Astrophysical Big Bang Laboratory, RIKEN, Wako, Saitama 351-0198, Japan*
[8] *Faculty of Engineering, Kanagawa University, Yokohama, Kanagawa 221-8686, Japan*
[9] *Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, Kofu, Yamanashi 400-8511, Japan*
[10] *The Graduate School of Science and Engineering, Saitama University, Saitama, Saitama 338-8570, Japan*
[11] *Academic Assembly School of Science and Technology Institute of Engineering, Shinshu University, Nagano, Nagano 380-8554, Japan*
[12] *Department of Physics, SungKyunKwan University, Jang-an-gu, Suwon 16419, Korea*
[13] *Department of Physics, Tokyo City University, Setagaya-ku, Tokyo 158-8557, Japan*
[14] *Institute for Nuclear Research of the Russian Academy of Sciences, Moscow 117312, Russia*
[15] *Faculty of Systems Engineering and Science, Shibaura Institute of Technology, Minato-ku, Tokyo 337-8570, Japan*
[16] *Department of Engineering Science, Faculty of Engineering, Osaka Electro-Communication University, Neyagawa-shi, Osaka 572-8530, Japan*
[17] *Service de Physique Thorique, University Libre de Bruxelles, Brussels 1050, Belgium*
[18] *Department of Physics, Yonsei University, Seodaemun-gu, Seoul 120-749, Korea*
[19] *Center for Astrophysics and Cosmology, University of Nova Gorica, Nova Gorica 5297, Slovenia*
[20] *Faculty of Science, Kochi University, Kochi, Kochi 780-8520, Japan*
[21] *Department of Physical Sciences, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan*
[22] *College of Engineering, Chubu University, Kasugai, Aichi 487-8501, Japan*
[23] *Quantum ICT Advanced Development Center, National Institute for Information and Communications Technology, Koganei, Tokyo 184-8795, Japan*
[24] *Sternberg Astronomical Institute, Moscow M.V. Lomonosov State University,*

Moscow 119991, Russia
[25] *Department of Physics, School of Natural Sciences, Ulsan National Institute of Science and Technology, UNIST-gil, Ulsan 689-798, Korea*
[26] *Department of Physics, Tokyo University of Science, Noda, Chiba 162-8601, Japan*
[27] *Earthquake Research Institute, University of Tokyo, Bunkyo-ku, Tokyo 277-8582, Japan*
[28] *Graduate School of Information Sciences, Hiroshima City University, Hiroshima, Hiroshima 731-3194, Japan*
[29] *Institute of Particle and Nuclear Studies, KEK, Tsukuba, Ibaraki 305-0801, Japan*
[30] *Graduate School of Science and Engineering, Tokyo Institute of Technology, Meguro, Tokyo 152-8550, Japan*
[31] *Nambu Yoichiro Institute of Theoretical and Experimental Physics, Osaka Metropolitan University, Sugimoto, Sumiyoshi, Osaka 558-8585, Japan*
[32] *CEICO, Institute of Physics, Czech Academy of Sciences, Prague 182 21, Czech Republic*

[*] Presently at: University of Californa - Santa Cruz and Flatiron Institute, Simons Foundation

[**] Presently at: Argonne National Laboratory, Physics Division, Lemont, Illinois 60439,USA

[***] Deceased

[****] Presently at: Georgia Institute of Technology, Physics Department, Atlanta, Geogia 30332,USA