A PYTHIA-Based Study of Isospin Symmetry Violation in Elementary Hadronic Collision Systems

David Qian

August 12, 2025

Abstract

Isospin symmetry holds to a good approximation in the strong interaction. In collisions with equal numbers of protons and neutrons, isospin symmetry dictates that the number of produced charged kaons should be equal to the number of neutral ones ($R_K = 1$). Recent experimental results from the NA61/SHINE experiment at CERN, however, recorded an unexpectedly large charged to neutral kaon yield ratio of about 1.233, a large violation of isospin symmetry. In this study, we establish a generator-level baseline in analogous elementary collision systems using the Pythia 8.3 event generator to simulate pp, np, and nn collisions over a range of center-of-mass energies between 50 and 5000 GeV, and analyze the resulting kaon yields. We observe a consistent excess of charged over neutral kaons in most configurations, with the largest deviations from expectation at lower energies and in proton-rich systems. These deviations are attributed to internal model mechanisms, including quark mass differences and the decay of the $\phi(1020)$ meson. These results establish a generator-level baseline for R_K in elementary collisions. These findings, when compared with experimental data, highlight the importance of further high-precision experimental and theoretical efforts to understand large isospin symmetry breaking.

Keywords: Isospin symmetry breaking, QCD, kaon production

1 Introduction

Isospin symmetry is an approximate symmetry of the strong interaction, based on the near-equal mass and QCD couplings of the up and down quarks. In hadron–hadron collisions, this symmetry implies that the yields of charged and neutral kaons should be equal under charge-symmetric conditions. Specifically, the charged-to-neutral kaon yield ratio, $R_K \equiv \frac{\langle K^+\rangle+\langle K^-\rangle}{\langle K^0\rangle+\langle \bar{K}^0\rangle}$, is expected to be close to unity in collisions involving equal numbers of protons and neutrons.

Recent measurements by the NA61/SHINE experiment at CERN reported $R_K \approx 1.23$ in Ar+Sc collisions at $\sqrt{s_{NN}} = 11.9$ GeV (Brylinski, Aduszkiewicz, Ali, & others (NA61/SHINE Collaboration), 2024), a value that significantly exceeds theoretical expectations and cannot be fully explained by established sources of isospin violation such as quark mass differences or electromagnetic corrections. These known effects, including the larger mass of the neutral kaons relative to charged kaons, electromagnetic coupling differences between u and d quarks, and $\phi(1020)$ meson decay branching ratios, are estimated to increase R_K only to ≈ 1.03 in symmetric systems (Brylinski et al., 2024; Bryliński & Collaboration, 2023).

The exact isospin-symmetric expectation $R_K = 1$ can be derived from the charge-symmetry operator $\hat{C} = e^{i\pi \hat{I}_y}$, which exchanges $u \leftrightarrow d$ quarks (Brylinski et al., 2024). In collisions where the total isospin projection I_z is zero, this symmetry enforces $\langle K^+\rangle = \langle K^0\rangle$ and $\langle K^-\rangle = \langle \bar{K}^0\rangle$, leading directly to $R_K = 1$. While pp and nn collisions are not strictly charge-symmetric, their kaon ratios are still expected to remain near unity, with small deviations arising from valence quark asymmetries.

Given the large discrepancy between experimental results and theoretical expectations, it is important to establish a generator-level baseline for R_K in simple systems where medium effects are absent. In this study, we use the PYTHIA 8.3 event generator to simulate pp, np, and nn collisions across $\sqrt{s} = 50$–5000 GeV, analyzing both inclusive production and kaons within a mid-rapidity kinematic window. By comparing the results with theoretical predictions and experimental measurements, we aim to determine whether the observed enhancement in R_K originates from genuine medium effects or from intrinsic features of the hadronization process.

2 Methodology

PYTHIA 8.3 is a Monte Carlo event generator that simulates high-energy particle collisions by stochastically modeling each stage of an event based on QCD probability distributions. It begins by sampling an initial hard scattering process (e.g., $q\bar{q} \to q\bar{q}$, $gg \to q\bar{q}$) according to leading-order matrix elements and parton distribution functions, which determine the likelihood of finding specific partons inside the colliding hadrons. Initial-and final-state parton showers are then simulated to model gluon radiation before and after the collision, generating cascades of softer emissions governed by DGLAP evolution. In hadron-hadron collisions, multiple parton interactions and beam remnants contribute to the event structure.

Hadronization is next, converting color-charged partons into observable hadrons using the Lund string fragmentation model. In this model, color-connected partons are treated as being linked by a "string" of energy, which breaks from quark-antiquark pair production as it stretches, thus forming mesons and baryons. The flavor composition and momentum distribution of the hadrons are determined from empirical fragmentation functions. Resonance decays are then simulated based on known branching ratios. For example, mesons like the $\phi(1020)$ meson decay preferentially into charged kaons. Because each event is generated using randomized sampling, PYTHIA's results reflect a variety of possible outcomes rather than determined predictions.

In this study, we employed the Pythia 8.3 event generator on the Monash tune to simulate $100,000$ minimum-bias inelastic events for each of the following nucleon–nucleon configurations: pp, np, and nn at four center-of-mass energies: $\sqrt{s} = 50,\ 200,\ 2000,\ 5000$ GeV (Skands, Carrazza, & Rojo, 2014; Sjöstrand, Ask, Christiansen, & et al., 2015). Events are generated with hadron-level output enabled and weak decays turned off for kaons, pions, and light neutral mesons to isolate primary hadron production and focus on hadronization and isospin effects without contamination from decay.

For each event, we count all final-state kaons before decay, distinguishing between charged (K^+, K^-) and neutral (K^0, $\bar{K}^0$) species. Two samples are generated: an inclusive sample where all produced kaons are counted without kinematic cuts and a mid-rapidity window where only kaons within $0.2 < p_T < 2$ GeV and $|\eta| < 0.5$ are counted.

The principal observable is the charged-to-neutral kaon yield ratio:

$$R_K = \frac{K^+ + K^-}{K^0 + \bar{K}^0} \tag{1}$$

We compute R_K by averaging over the $100,000$ events.

To assess the statistical significance of the observed charged-to-neutral kaon yield ratio, we estimate the standard deviation of R_K using standard error propagation techniques. The kaon ratio is defined as

$$R_K = \frac{N_{\text{ch}}}{N_{\text{neu}}} = \frac{K^+ + K^-}{K^0 + \bar{K}^0}, \tag{2}$$

where N_{ch} and N_{neu} are the total numbers of charged and neutral kaons, respectively, summed over $100,000$ PYTHIA-generated events. Since both N_{ch} and N_{neu} are independent count-based quantities, we assume they follow Poisson statistics, with uncertainties of $\sigma_{N_{\text{ch}}} = \sqrt{N_{\text{ch}}}$ and $\sigma_{N_{\text{neu}}} = \sqrt{N_{\text{neu}}}$.

The uncertainty in the ratio R_K is then given by the standard propagation of uncertainty for a ratio of independent variables:

$$\sigma_{R_K} = R_K \cdot \sqrt{\left(\frac{\sigma_{N_{\text{ch}}}}{N_{\text{ch}}}\right)^2 + \left(\frac{\sigma_{N_{\text{neu}}}}{N_{\text{neu}}}\right)^2} = R_K \cdot \sqrt{\frac{1}{N_{\text{ch}}} + \frac{1}{N_{\text{neu}}}}. \tag{3}$$

By evaluating $R_K - 3\sigma_{R_K}$ for each collision system and energy, we test whether the observed deviations from the isospin-symmetric expectation ($R_K = 1$) remain significant at the 99.7% confidence level. In many cases, these lower bounds remain greater than 1, reinforcing the robustness of the observed enhancement in charged over neutral kaon production.

It is important to note that we compute R_K using the total kaon yields across all events rather than averaging the per-event ratios. This avoids statistical bias that can arise due to fluctuations in small denominators. If one were to instead calculate $R_K = \langle K_{\text{ch}}/K_{\text{neu}} \rangle$ on an event-by-event basis, the result would be higher than the true ratio, especially when the mean yields are low. For example, for two Gaussian-distributed variables X and Y with the same mean μ and standard deviation $\sigma \ll \mu$, the expectation value of the ratio is approximately $\langle X/Y \rangle \approx 1+\sigma^2/\mu^2$, leading to an artificial increase of about 4% for $\sigma = 0.2\mu$. By summing over large event samples, we avoid this bias in our calculation of R_K.

3 Results

Inclusive production: Across the measured energies, the simulation yields the following ranges of kaon ratios: $1.04 < R_K^{pp} < 1.11$, $1.04 < R_K^{np} < 1.05$, and $0.97 < R_K^{nn} < 1.05$. The largest ratio appears at the lowest energy ($\sqrt{s} = 50$ GeV) in the pp collision system, where $R_K^{pp} \approx 1.10$. The ratio remains greater than 1, except for a slight dip to $R_K^{nn} \approx 0.97$ at low energy.

Mid-rapidity production: Applying the cut reduces multiplicities and yields similar or reduced values of the kaon ratios, as expected due to resonance decay into kaons which is often distributed across a wider rapidity range. Still, values of R_K persist across all energies with the following ranges of kaon ratios from the data: $1.03 < R_K^{pp} < 1.08$, $1.04 < R_K^{np} < 1.06$, and $1.04 < R_K^{nn} < 1.06$. The largest ratio again appears at the lowest energy ($\sqrt{s} = 50$ GeV) in the pp collision system, where $R_K^{pp} \approx 1.07$. The ratio remains greater than 1 throughout all systems.

To test whether the observed deviations from the isospin-symmetric expectation ($R_K = 1$) are statistically robust, we calculate the significance level of the deviation of R_K from unity ($\frac{R_K-1}{\sigma_{R_K}}$) as listed in the tables below, with a larger value signifying a more significant deviation. In nearly all configurations, the highest-energy cases have the most significant deviations. For example, at $\sqrt{s} = 50$ GeV in the pp collision, $R_K = 1.1043$ with $\sigma_{R_K} = 0.0046$, giving $\frac{R_K-1}{\sigma_{R_K}} = 22.2$ while at $\sqrt{s} = 5000$ GeV, $R_K = 1.0486$ with $\frac{R_K-1}{\sigma_{R_K}} = 38.26$.

Using $\frac{(R_K-1)}{\sigma_{R_K}}$ makes it clear that the excess of charged over neutral kaons is statistically significant in nearly all configurations, with the exception of the low-energy nn case, which shows a small but significant deficit. This indicates that the excess of charged over neutral kaons is not the result of random fluctuations, but rather a statistically significant effect emerging from the underlying hadronization dynamics modeled by PYTHIA. The systematic ordering $R_K^{pp} > R_K^{np} > R_K^{nn}$ and the decreasing magnitude of $(R_K - 1)/\sigma_{R_K}$ with $\sqrt{s}$are consistent with expectations from valence-quark content and its diminishing role at higher energies.

Table 1: Kaon yields and ratios from PYTHIA pp collisions (100k events, full phase space)

$\sqrt{s}$ (GeV)	$\langle K^+ \rangle$	$\langle K^- \rangle$	$\langle K_S \rangle$	$\langle K_L \rangle$	$\langle K^0 \rangle$	$\langle \bar{K}^0 \rangle$	$\frac{K^+ + K^-}{K^0 + \bar{K}^0}$	σ_{R_K}	$\frac{R_K-1}{\sigma_{R_K}}$
5000	7.052	6.905	6.859	6.866	6.706	6.604	1.0486	0.0012	38.2629
2000	4.986	4.832	4.824	4.817	4.715	4.635	1.0501	0.0015	32.9876
200	2.089	1.949	1.963	1.963	1.955	1.860	1.0585	0.0023	24.4597
50	0.649	0.519	0.543	0.541	0.563	0.494	1.1043	0.0046	22.2361

Table 2: Kaon yields and ratios from PYTHIA nn collisions (100k events, full phase space)

$\sqrt{s}$ (GeV)	$\langle K^+\rangle$	$\langle K^-\rangle$	$\langle K_S\rangle$	$\langle K_L\rangle$	$\langle K^0\rangle$	$\langle \bar{K}^0\rangle$	$\frac{K^+ + K^-}{K^0+\bar{K}^0}$	σ_{R_K}	$\frac{R_K-1}{\sigma_{R_K}}$
5000	6.978	6.895	6.880	6.864	6.741	6.587	1.0409	0.0012	32.3892
2000	4.934	4.847	4.850	4.849	4.780	4.632	1.0392	0.0015	26.1278
200	2.034	1.945	1.986	2.003	2.010	1.868	1.0260	0.0023	11.2488
50	0.584	0.514	0.580	0.576	0.631	0.498	0.9724	0.0041	-6.7043

Table 3: Kaon yields and ratios from PYTHIA np collisions (100k events, full phase space)

$\sqrt{s}$ (GeV)	$\langle K^+\rangle$	$\langle K^-\rangle$	$\langle K_S\rangle$	$\langle K_L\rangle$	$\langle K^0\rangle$	$\langle \bar{K}^0\rangle$	$\frac{K^+ + K^-}{K^0+\bar{K}^0}$	σ_{R_K}	$\frac{R_K-1}{\sigma_{R_K}}$
5000	7.009	6.890	6.881	6.863	6.730	6.605	1.0423	0.0012	33.4755
2000	4.954	4.845	4.841	4.835	4.754	4.632	1.0440	0.0015	29.1822
200	2.056	1.946	1.980	1.973	1.984	1.858	1.0416	0.0023	17.7007
50	0.616	0.514	0.557	0.555	0.592	0.492	1.0415	0.0044	9.3724

Table 4: Kaon yields and ratios from PYTHIA pp collisions (100k events, $0.2 < p_T < 2.0$, $|\eta| < 0.5$)

$\sqrt{s}$ (GeV)	$\langle K^+\rangle$	$\langle K^-\rangle$	$\langle K_S\rangle$	$\langle K_L\rangle$	$\langle K^0\rangle$	$\langle \bar{K}^0\rangle$	$\frac{K^+ + K^-}{K^0+\bar{K}^0}$	σ_{R_K}	$\frac{R_K-1}{\sigma_{R_K}}$
5000	0.5165	0.5185	0.5116	0.5132	0.4969	0.4975	1.0408	0.0046	8.8339
2000	0.4273	0.4265	0.4243	0.4230	0.4130	0.4105	1.0368	0.0050	7.2659
200	0.3387	0.3386	0.3339	0.3346	0.3271	0.3219	1.0436	0.0057	7.6067
50	0.1823	0.1772	0.1735	0.1743	0.1660	0.1697	1.0709	0.0081	8.7226

Table 5: Kaon yields and ratios from PYTHIA nn collisions (100k events, $0.2 < p_T < 2.0$, $|\eta| < 0.5$)

$\sqrt{s}$ (GeV)	$\langle K^+\rangle$	$\langle K^-\rangle$	$\langle K_S\rangle$	$\langle K_L\rangle$	$\langle K^0\rangle$	$\langle \bar{K}^0\rangle$	$\frac{K^+ + K^-}{K^0+\bar{K}^0}$	σ_{R_K}	$\frac{R_K-1}{\sigma_{R_K}}$
5000	0.5203	0.5198	0.5136	0.5137	0.4975	0.4984	1.0444	0.0046	9.5852
2000	0.4263	0.4282	0.4199	0.4181	0.4060	0.4073	1.0507	0.0051	9.8422
200	0.3412	0.3357	0.3323	0.3358	0.3259	0.3218	1.0451	0.0057	7.8481
50	0.1765	0.1770	0.1758	0.1746	0.1695	0.1691	1.0440	0.0079	5.5430

Table 6: Kaon yields and ratios from PYTHIA np collisions (100k events, $0.2 < p_T < 2.0$, $|\eta| < 0.5$)

$\sqrt{s}$ (GeV)	$\langle K^+\rangle$	$\langle K^-\rangle$	$\langle K_S\rangle$	$\langle K_L\rangle$	$\langle K^0\rangle$	$\langle \bar{K}^0\rangle$	$\frac{K^+ + K^-}{K^0+\bar{K}^0}$	σ_{R_K}	$\frac{R_K-1}{\sigma_{R_K}}$
5000	0.5162	0.5196	0.5112	0.5116	0.4965	0.4964	1.0432	0.0046	9.3253
2000	0.4287	0.4282	0.4222	0.4245	0.4118	0.4092	1.0437	0.0050	8.5786
200	0.3411	0.3388	0.3341	0.3303	0.3235	0.3206	1.0556	0.0058	9.5761
50	0.1760	0.1765	0.1733	0.1732	0.1673	0.1670	1.0544	0.0080	6.7630

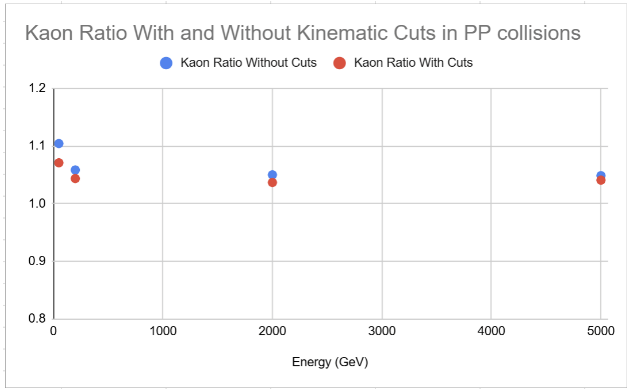

Figure 1: R_K with and without Kinematic Cuts in pp collisions

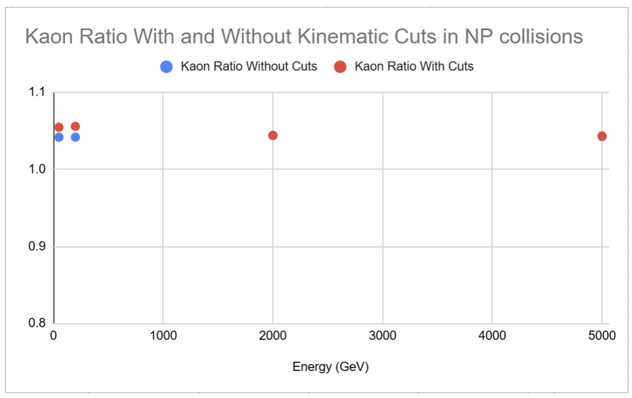

Figure 2: R_K with and without Kinematic Cuts in np collisions

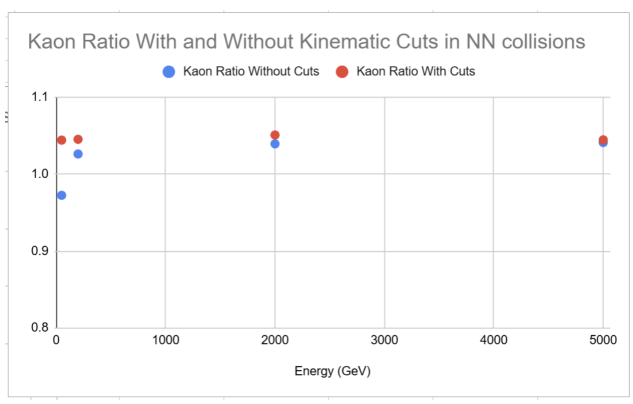

Figure 3: R_K with and without Kinematic Cuts in nn collisions

4 Conclusion

We predicted that in isospin symmetric systems, the charged to neutral kaon yield ratio R_K would equal or slightly exceed 1, up to around $R_K \approx 1.03$, accounting for isospin breaking effects. Yet the generated PYTHIA data exceeds this prediction, especially at lower energies, pointing to more pronounced hadronization and resonance effects. Since these values are collected from a Monte Carlo generator rather than an experiment, the observed $R_K > 1$ cannot be attributed to statistical noise or detector effects. Instead, several model-internal mechanisms could be responsible:

Quark masses: Since $m_d > m_u$, the production of lighter charged kaons is biased over heavier neutral kaons near the energy threshold.

Resonance decays: The $\phi(1020)$ meson decays twice as often into the charged kaons, increasing the kaon ratio.

String fragmentation biases: PYTHIA's Lund-string model includes flavor and charge dependent parameters that could introduce asymmetries in hadron production (Andersson, Gustafson, Ingelman, & Sjostrand, 1983).

***pp* and *nn* collisions:** A system is isospin symmetric when it has an equal number of up and down quarks. Since protons are made up of *uud* and neutrons are made up of *udd*, *pp* collisions begin with an excess of up quarks, biasing the formation of charged kaons, and *nn* collisions begin with an excess of down quarks, biasing the production of neutral kaons. Thus, the lack of isospin symmetry in the initial state could cause deviation from the expected kaon ratio.

Recently, the NA61/SHINE experiment observed $R_K \approx 1.23$ in Ar+Sc collisions (Bryliński & Collaboration, 2023). Combined with other available data, there was a significant deviation between theoretical models and experimental results in R_K of 5.5σ. With this significant deviation in mind, this study focused on examining R_K in elementary systems, *pp*, *np*, and *nn*.

Through this study, we have established a generator-level baseline in elementary nucleon–nucleon systems. Simulations show a persistent $R_K > 1$ in nearly all configurations, with deviations that exceed theoretical predictions from isospin-breaking sources (which suggest $R_K \approx 1.02$–1.03) and remain statistically significant well beyond the 3σ level in most cases. The enhancement is most prominent at low energies and in proton-rich systems such as *pp* at $\sqrt{s} = 50$ GeV.

This strongly suggests that the observed kaon asymmetries are not a result of statistical noise, but rather model-internal effects, for example particularly mass differences, resonance decay asymmetries, and flavor-dependent string fragmentation in PYTHIA. $R_K - 3\sigma_{R_K} > 1$ across multiple systems reinforces the idea that isospin symmetry breaking is a persistent feature of generator-level physics.

With some experimental data already showing the violations of isospin symmetry in kaon production, further comparative studies with generators and analytic modeling is warranted. Experimental confirmation through well-understood isospin-symmetric systems, such as $d+d$ collisions, would also help distinguish medium-induced effects from generator-specific biases.

Ultimately, the observed enhancement of R_K in both simulations and experiments suggests that isospin symmetry may be more fragile in hadronization than previously known. If future experimental results continue to show charged to neutral kaon ratios exceeding both theoretical and generator baselines, it may point to the emergence of a novel regime of isospin-violating QCD matter, with implications for our understanding of hadron formation in high-energy nuclear collisions. Thus, new high-precision measurements and theoretical efforts are warranted to further investigate large isospin-breaking effects.

References

Andersson, B., Gustafson, G., Ingelman, G., & Sjostrand, T. (1983). Parton fragmentation and string dynamics. *Phys. Rep.*, *97*, 31–145. doi: 10.1016/0370-1573(83)90080-7

Brylinski, K., Aduszkiewicz, A., Ali, Y., & others (NA61/SHINE Collaboration). (2024). Strong isospin breaking in kaon production from nucleus-nucleus collisions at sps energies. *arXiv preprint arXiv:2403.07920*. Retrieved from `https://arxiv.org/abs/2403.07920` (Submitted to Phys. Rev. Lett.)

Bryliński, K., & Collaboration, N. (2023). Large isospin symmetry breaking in kaon production at high energies. *arXiv preprint*. Retrieved from `https://arxiv.org/abs/2312.07176`

Sjöstrand, T., Ask, S., Christiansen, J. R., & et al. (2015). An introduction to pythia 8.2. *Comput. Phys. Commun.*, *191*, 159–177. doi: 10.1016/j.cpc.2015.01.024

Skands, P. Z., Carrazza, S., & Rojo, J. (2014). Tuning pythia 8.1: the monash 2013 tune. *Eur. Phys. J. C*, *74*, 3024. doi: 10.1140/epjc/s10052-014-3024-y