THE EUROPEAN
PHYSICAL JOURNAL C

ToMCCA: a Toy Monte Carlo Coalescence Afterburner

Maximilian Mahlein[1,a] , **Chiara Pinto**[1,2], **Laura Fabbietti**[1]

[1] Physics Department, TUM School of Natural Sciences, Technical University of Munich, James-Franck-Straße 1, 85748 Garching b. München, Germany

[2] European Organisation for Nuclear Research CERN, Geneva, Switzerland

Abstract Antinuclei in our Galaxy may stem either from annihilation or decay of dark matter, or from collisions of cosmic rays with the interstellar medium, which constitute the background of indirect dark matter searches. Understanding the formation mechanism of (anti)nuclei is crucial for setting limits on their production in space. Coalescence models, which describe the formation of light nuclei from final-state interaction of nucleons, have been widely employed in high-energy collisions. In this work, we introduce ToMCCA (**To**y **M**onte **C**arlo **C**oalescence **A**fterburner), which allows for detailed studies of the nuclear formation processes without the overload of general-purpose event generators. ToMCCA contains parameterizations of the multiplicity dependence of the transverse momentum distributions of protons and of the baryon-emitting source size, extracted from ALICE measurements in pp collisions at $\sqrt{s} = 5 - 13$ TeV, as well as of the event multiplicity distributions, taken from the EPOS event generator. ToMCCA provides predictions of the deuteron transverse momentum distributions, with agreement of $\sim 5\%$ with the experimental data. The results of ToMCCA show that the coalescence mechanism in pp collisions depends only on the event multiplicity, not on the collision system or its energy. This allows the model to be utilized for predictions at lower center-of-mass collision energies, which are the most relevant for the production of antinuclei from processes related to dark matter. This model can also be extended to heavier nuclei as long as the target nucleus wavefunction and its Wigner function are known.

1 Introduction

The study and potential detection of antinuclei in cosmic rays has emerged as an exciting field of research, offering a unique opportunity to explore possible signals from physics beyond the Standard Model. Antinuclei, consisting of antiprotons and antineutrons, are exceedingly rare in cosmic rays relative to their matter counterparts, and their presence in our Galaxy could be due to annihilation or decay of dark matter (DM) candidates [1,2]. Indeed, the antinuclear flux expected from Standard Model processes due to collisions of high-energy cosmic rays (CRs) and the interstellar medium (ISM) is expected to be two to four orders of magnitude lower than the dark matter flux at low kinetic energies ($E_{\text{kin}} < 1$ GeV/A) [3,4]. Both CRs and the ISM are mostly made of protons ($\sim 90\%$) and ^{4}He nuclei ($\sim 9\%$), and only include a small percentage of heavier nuclei. Hence, the main channels contributing to the Standard Model background are collisions between protons and ^{4}He nuclei (p–p, p–^{4}He, ^{4}He–p, ^{4}He–^{4}He), with center-of-mass energies of $\mathcal{O}(10)$ GeV [5]. Understanding the formation mechanism of antinuclei is crucial for estimating the production cross-section from the Standard Model and dark matter contributions and for interpreting any future measurements of antinuclear flux from cosmic rays. Different experimental efforts are ongoing to measure the flux of antinuclei in cosmic rays, such as the Alpha Magnetic Spectrometer (AMS)-02 [6], General AntiParticle Spectrometer (GAPS) [7], and Balloon-borne Experiment with Superconducting Spectrometer (BESS)-Polar [8] experiments. However, antinuclei heavier than antiprotons have not yet been measured.

Coalescence models have been widely used to study the formation of light nuclei starting from the production of individual nucleons in high-energy hadronic collisions [9–15]. In these models, light nuclei are formed through the coalescence of nucleons that satisfy specific conditions, such as small relative momentum, spatial proximity, and spin–isospin alignment. In the simplest implementations of coalescence, only the momentum correlations are considered, and the bound states are formed if the difference in momentum among the

a e-mail: maximilian.horst@tum.de (corresponding author)

nucleons is below a given threshold, namely the coalescence momentum p_0. In the state-of-the-art implementations of the coalescence approach, the quantum-mechanical properties of the constituent baryons and those of the final bound states are taken into account, and the coalescence probability is calculated from the overlap between the wavefunctions of individual (point-like) baryons and the Wigner density of the final-state cluster [13,15–19]. Monte Carlo (MC) simulations have proved to be powerful tools for investigating nuclear dynamics and particle production in high-energy hadronic collisions, as they offer a stochastic treatment of the complex dynamics in high-energy particle collisions while maintaining the full information on the spacetime position and momenta of individual particles. By combining the principles of MC event generators with coalescence models, recent descriptions of the momentum distributions of nuclei have drastically improved [16,19].

In light of these advancements, we propose a new approach to MC-based coalescence, called ToMCCA (Toy Monte Carlo Coalescence Afterburner)[1] ToMCCA is a toy Monte Carlo that allows for a detailed study of the nuclear formation process without the overload of general-purpose event generators. In ToMCCA, the production of bound states is implemented through a coalescence afterburner based on the Wigner function formalism. The event generation is customizable, and the kinematics of the produced nucleons are tuned using parameterizations of (i) the multiplicity dependence of their transverse momentum (p_T) distributions, (ii) their angular momentum correlations, (iii) the baryon-emitting source size, and (iv) the multiplicity distributions. Using such parameterizations and computing the probability of coalescence employing the Wigner function of the nucleus, ToMCCA provides predictions for the transverse momentum distributions of deuterons, for arbitrary multiplicity intervals covered by pp collisions. The parameterizations included in ToMCCA, based on measurements carried out with ALICE, make use of results from proton–proton (pp) collisions at a center-of-mass energy of $\sqrt{s} = 5$ TeV [20] and of those selected with the high-multiplicity (HM) trigger in pp collisions at $\sqrt{s} = 13$ TeV [21]. The performance of ToMCCA is tested against the deuteron p_T distributions of pp collisions at energies not included in the parameterizations, namely $\sqrt{s} = 13$ TeV [22] (with a minimum-bias trigger) and $\sqrt{s} = 7$ TeV [23]. The ToMCCA model can be easily extended to heavier nuclei, as long as the nucleus wavefunction is known and its Wigner function is computed, and to lower center-of-mass collision energies, of the order of 10 GeV, which are the most relevant for the production of antinuclei from processes related to dark matter [5].

2 Working principle of ToMCCA

2.1 Coalescence model

The coalescence afterburner model built in ToMCCA is based on the Wigner function formulation of coalescence. For more details, see Refs. [16,19] and references therein. For convenience, we summarize here the basics of the coalescence model used in ToMCCA. The general idea is that the momentum distribution of deuterons in the final state (d^3N/dP^3) can be expressed by projecting the phase space of the unbound proton–neutron (p–n) system (W_{np}) onto the Wigner function of the deuteron $D(\vec{q}, \vec{r})$

$$\frac{d^3N}{dP^3} = S \int d^3q \int d^3r_d \int d^3r \, D(\vec{q}, \vec{r}) \\ \times W_{np}(\vec{P}_d/2 + \vec{q}, \vec{P}_d/2 - \vec{q}, \vec{r}_d, \vec{r}), \qquad (1)$$

where S is the spin–isospin statistics factor, which equals 3/8 for deuterons, $\vec{P}_d$ is the total momentum, $\vec{q}$ is the relative momentum, $\vec{r}_d$ is the center-of-mass position, and $\vec{r}$ the distance of the p–n pair. The deuteron Wigner function depends only on the wavefunction of the deuteron, for which there are multiple options, e.g., Gaussian, Hulthen [24], double Gaussian, or Argonne v_{18} [25]. In this work, we use the Argonne v_{18} wavefunction, which was shown to have the best agreement with experimental data in Ref. [19]. For the deuteron Wigner function, we use the parameterization reported in the appendix of Ref. [19] for the Argonne v_{18} wavefunction[2] We make the assumption that one can factorize the two-nucleon phase space into its space and momentum terms. The p–n phase space assumes the form

$$W_{np}(\vec{P}_d/2 + \vec{q}, \vec{P}_d/2 - \vec{q}, \vec{r}_d, \vec{r}) = H_{np}(\vec{r}, \vec{r}_d; r_0) \\ \times G_{np}(\vec{P}_d/2 + \vec{q}, \vec{P}_d/2 - \vec{q}). \qquad (2)$$

Folding the deuteron Wigner function with the nucleon phase-space distribution, it is possible to obtain the coalescence probability $\mathcal{P}(r_0, \vec{q})$ as a function of the source size r_0 and $\vec{q}$

$$\mathcal{P}(\vec{q}, r_0) = \int d^3r \int d^3r_d \, D(\vec{q}, \vec{r}) H_{np}(\vec{r}, \vec{r}_d; r_0). \qquad (3)$$

The source size is described in Sect. 3. This probability can then be applied on any p–n pair, on a pair-by-pair basis, to predict deuteron transverse momentum distributions.

2.2 The ToMCCA toy Monte Carlo

The use of a toy MC instead of traditional MC event generators, such as Pythia [26,27] or EPOS [28], allows for much

---

[1] ToMCCA is available at https://github.com/HorstMa/ToMCCA-Public.

[2] The deuteron Wigner function for the Argonne v_{18} wavefunction is available in the GitHub project at https://github.com/HorstMa/ToMCCA-Public.

faster computation, as only processes strictly necessary for coalescence are computed. Furthermore, such a lightweight generator can be easily modified and customized by the end-user, allowing for many more differential studies. In this work, we developed a toy Monte Carlo generator that reproduces the deuteron momentum distributions employing parameterizations of the following parameters:

– Charged-particle multiplicity of events
– Nucleon momentum distributions
– Angular correlations among nucleons
– Hadron-emitting source size
– Nuclear wavefunctions

The parameterizations included in ToMCCA are based on experimental measurements from the ALICE Collaboration, except for the parameterization of the charged-particle multiplicity distributions, which is based on EPOS 3, as described in detail in Sect. 2.2.4.

2.2.1 Event multiplicity

The starting point of ToMCCA is the determination of the charged-particle multiplicity of the event that is generated. This number is usually given by the number of charged particles per unit of pseudorapidity η, e.g., measured in a range of $|\eta| < 0.5$ $((\mathrm{d}N_{ch}/\mathrm{d}\eta)_{|\eta|<0.5})$. Particle spectra are often measured in a different kinematic range, such as $|y| < 0.5$, where y is the rapidity relative to the beam axis. To transform from $|\eta| < 0.5$ to $|y| < 0.5$, a scaling constant α has to be determined such that $\alpha \cdot (\mathrm{d}N_{ch}/\mathrm{d}\eta)_{|\eta|<0.5} = (\mathrm{d}N_{ch}/\mathrm{d}y)_{|y|<0.5}$ [29]. This can be computed using event generators. For the case of ToMCCA, a value of $\alpha = 1.192$ is determined using EPOS 3 [28].

There are multiple ways to reproduce each given $\mathrm{d}N_{ch}/\mathrm{d}\eta$ value. Two simple approaches are included in ToMCCA: (i) fixed multiplicity, i.e., generating exactly $\alpha \mathrm{d}N_{ch}/\mathrm{d}\eta$ particles per event, and (ii) Poissonian multiplicity, i.e., generating a Poissonian distribution around the chosen multiplicity value. However, since coalescence is a nonlinear process, i.e., it depends on the number of pairs rather than single particles, the overall shape of the distribution plays a significant role. Thus, a more realistic approach is needed. An improvement can be reached by using an event generator, such as EPOS 3. Using an unbiased triggering method, such as the one employed by the ALICE experiment [21], the chosen mean charged-particle multiplicities can be triggered outside of the nucleus measurement region ($-3.7 < \eta < -1.7$ and $2.8 < \eta < 5.1$). The shapes of the resulting midrapidity distributions can be well described using the Erlang distribution [30], given by

Table 1 Parameterization of the Erlang distribution using power law functions

Parameter	k	λ
A	0	0
B	0.77544	0.72653
C	0.74669	-0.23343

$$f(x; k, \lambda) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!}. \tag{4}$$

While technically the Erlang distribution is only defined for $k \in \mathbb{N}^+$, it can be extended to $k \in \mathbb{R}_{\geq 0}$ without introducing any poles, as long as $x > 0$. This further requires one to generalize the factorial $(k-1)!$ to the gamma function $\Gamma(k)$, valid for $k \in \mathbb{R}_{\geq 0}$. With these modifications, the multiplicity dependence of the parameters of the Erlang function k and λ can be modeled with a power law function (given by Eq. 13). The parameters of the power law function can be found in Table 1.

2.2.2 The ToMCCA event

A schematic view of the event loop in ToMCCA is shown in Fig. 1. First, the number of nucleons (protons and neutrons separately) in a given event is calculated starting from the particle multiplicity of the event (Sect. 2.2.3). In panel (1) of Fig. 1, an exemplary situation in which four protons (red) and four neutrons (blue) are produced is depicted. Then, the three-dimensional momentum of the first proton and all neutrons is determined (Sect. 2.2.4), as well as their relative distances (Sect. 3), as sketched in panel (2) of Fig. 1. Using these momenta (q) and distances (r_0), the coalescence probability $\mathcal{P}(q, r_0)$ is calculated for each pair using Eq. 3. A random number $r \in [0, 1]$ is drawn, and if $r < \mathcal{P}(q, r_0)$, the pair is accepted as a deuteron. Otherwise, the proton is checked against other neutrons. If multiple neutrons satisfy the condition of coalescence with the same proton, only the first pair forms the deuteron. However, this is a very rare occurrence comparable to the ^{3}He and ^{3}H production rates ($\sim 1/10^{-7}$ events [21]). After all p–n pairs have been combined, a new proton and a completely new set of neutrons are created, and the same procedure is repeated. If a deuteron is created in the first loop, the number of neutrons available for the following protons in this event is reduced by 1. In panel (3) of Fig. 1, a sketch of a deuteron being formed is shown. The subsequent event in Fig. 1 (4) shows the second proton with now only three, and completely new, neutrons.

2.2.3 Particle production

In ToMCCA, the exact composition of charged particles is not important. Only the final number of protons and neutrons that
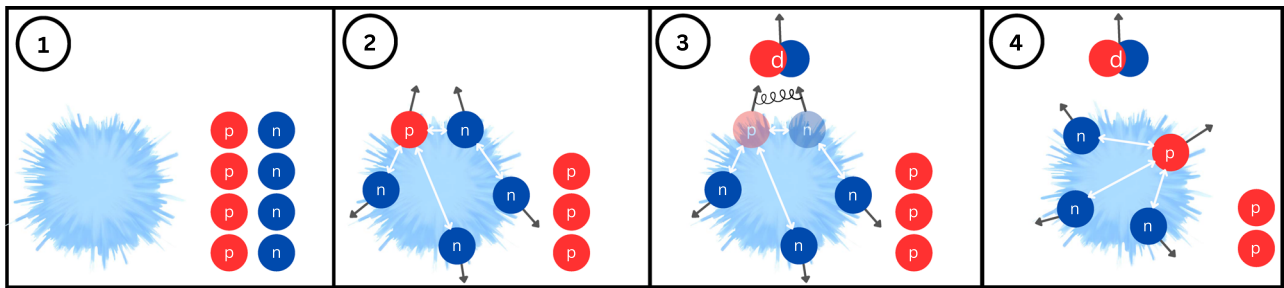
Fig. 1 (1) Starting situation after the number of protons (red) and neutrons (blue) have been determined. (2) For each proton, a new set of neutrons is generated. (3) One proton–neutron pair that satisfies the coalescence condition forms a deuteron. (4) Next proton of the event gets a new set of neutrons. The number of neutrons has been reduced by 1, since one neutron was bound to a proton to form the deuteron

are generated is of interest. There are four methods implemented in ToMCCA to determine the number of nucleons: uncorrelated emission, string fragmentation, quark recombination, and tuned emission. The methods used for particle production are described in detail below.

1. Uncorrelated emission

In this method, for every charged particle created, there is a probability for it to be a proton given by

$$p_{prot} = (\mathrm{d}N/\mathrm{d}y)/(\mathrm{d}N_{ch}/\mathrm{d}y), \tag{5}$$

where $\mathrm{d}N/\mathrm{d}y$ is the production *yield* of protons, anchored to the measured production spectra, as will be discussed in Sect. 2.2.4. A mirroring technique is used to determine the number of neutrons, where N_{ch} neutral particles are created and assigned to be neutrons in the same manner as protons. Notably, this method creates no correlation between the event-by-event number of protons and that of neutrons, and therefore between protons and deuterons. The impact of this could be quantified by studying the deuteron number fluctuations and correlations between nucleons and deuterons on an event-by-event basis, as has been done in Ref. [31] for Pb–Pb collisions.

2. String fragmentation

The second method implemented in ToMCCA is called string fragmentation. This method mimics much more complex hadronization methods employed by general-purpose event generators, such as Pythia or EPOS. Here, one large string is created, and either a u($\bar{\mathrm{u}}$) or a d($\bar{\mathrm{d}}$) quark is placed at one end, and its antimatter counterpart at the other end. This string is fragmented in a given number of locations, which equals the total number of particles created $\mathrm{d}N_{ch}/\mathrm{d}y$. On each fragmentation location, a quark–antiquark (q$\bar{\mathrm{q}}$) or diquark–antidiquark (dq$\overline{\mathrm{dq}}$) pair is created. The order of the pairs is determined so that a color-neutral particle is created at all times, i.e., diquarks are paired with quarks, and single quarks are paired with single antiquarks. To avoid tetraquark configurations, no diquark can be created directly after another diquark. After all the fragmentations, the triplets created from
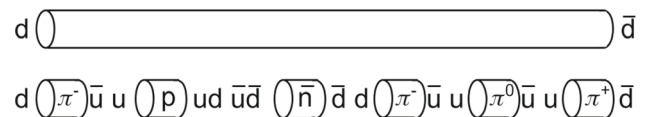
Fig. 2 An example of the fragmentation of a d$\bar{\mathrm{d}}$ string. Mesons are indicated for clarity but are ignored for coalescence in ToMCCA

diquarks are assigned to the corresponding particles depending on the quark content. The combinations ddd and uuu are assigned to neutrons and protons, respectively, since these are Δ^- and Δ^{++} resonances, which decay exclusively into the assigned nucleons. Mesons that would be created with this method are ignored for simplicity. An example of a string with an initial d$\bar{\mathrm{d}}$ configuration is shown in Fig. 2.

This method determines the total proton yield starting from the probability of creating a diquark. Due to edge effects in the fragmentation algorithm, baryon production at low multiplicities is suppressed. The diquark probability needs to be modified and parameterized as a function of multiplicity, in order to recover the production yield measured at low multiplicity. This probability is parameterized such that the proton yields as a function of multiplicity correspond to the one from the uncorrelated emission method through the equation

$$\begin{aligned} p_{\mathrm{diquark}} &= p_{\mathrm{prot,modified}}(\mathrm{d}N/\mathrm{d}y) \\ &= p_{\mathrm{prot}}(\mathrm{d}N/\mathrm{d}y) \cdot (A + B\,(\mathrm{d}N/\mathrm{d}\eta)^C), \end{aligned} \tag{6}$$

where p_{prot} is given by Eq. 5. The obtained parameters can be found in Table 2.

3. Quark recombination

The quark recombination method is based on the idea that a sea of u and d quarks is created and recombined randomly to create hadrons. The total number of each quark species is $N_d = N_u = \mathrm{d}N_{ch}/\mathrm{d}y$, where $\mathrm{d}N_{ch}/\mathrm{d}y$ is the number of charged particles as described above. First, a u and a d quark are combined randomly, and then, with probability $p_{\mathrm{prot,modified}}$ given by Eq. 6, a third u(d) quark is added to the ud pair, creating a proton (neutron). This is repeated until no ud pairs are left. The fit parameters of Eq. 6, for the

Table 2 Fit parameters of Eq. 6, for the parameterization of the diquark probability (string fragmentation method) and of the probability of adding a quark (quark recombination method)

Parameter	String Fragmentation	Quark recombination
A	2.13957998	0.87627309
B	-3.8557517	1.01566945
C	-2.0928103	-0.54537776

Table 3 The Pearson correlation coefficient between protons and neutrons for different hadronization methods

	Unc. Em.	String Fragm.	Quark Rec.	Tuned Em.
ρ_{pn}	0	-0.052	-0.058	-0.024

parameterization of the probability of adding one quark to the ud pair, are also reported in Table 2.

4. Tuned emission
Both the string fragmentation and quark recombination methods by construction have a fixed Pearson correlation coefficient ρ_{pn} (Eq. 7), which is a measure of the correlation between the number of protons and neutrons in a given event. A negative coefficient means that the presence of a proton suppresses the creation of new neutrons. To circumvent this feature, a fourth method is developed, namely the tuned emission, which contains a free parameter a to adjust the level of nucleon suppression. It is based on the uncorrelated emission, but the probability for a neutron p_{neut} is modified depending on the number of protons in the event via $p_{neut,tuned} = (a(N_p - \mathrm{d}N/\mathrm{d}y) + 1) p_{neut}$. Since the string fragmentation and quark recombination models show an unreasonably strong suppression of nucleon pairs at low multiplicities, the value a was set in such a way that its Pearson coefficient lies roughly in the middle between string fragmentation and uncorrelated emission. The exact parameterization is $a = 1.508 \, (\mathrm{d}N_{ch}/\mathrm{d}y)^{-1}$. The difference in the final deuteron yields between all these methods (1 to 4) ranges from $< 10\%$ for $\mathrm{d}N_{ch}/\mathrm{d}\eta|_{|\eta|<0.5} > 30$–$300\%$ for $\mathrm{d}N_{ch}/\mathrm{d}\eta|_{|\eta|<0.5} \sim 3$. This shows that the exact degree of suppression is a valuable input to this model. The parameter a can be tuned to a measurement of the Pearson coefficient of proton and deuteron ρ_{pd} as the correlation between proton and neutron is transferred into the correlation between proton and deuteron. Such measurements have been performed in the past on Pb–Pb data [32], but not yet in pp collisions. The Pearson correlation coefficient can be calculated as

$$\rho_{\mathrm{pn}} = \langle (n_p - \langle n_p \rangle)(n_n - \langle n_n \rangle) \rangle / \sqrt{\kappa_{2p} \kappa_{2n}}, \qquad (7)$$

where κ_{2a} is the second-order cumulant of the particle multiplicity distribution $\kappa_{2a} = \langle (n - \langle n \rangle)^2 \rangle$. The calculated proton–neutron Pearson coefficients ρ_{pn} for each method can be found in Table 3.

In the following, all results are obtained using the tuned emission method.

2.2.4 Kinematics

Once the number of protons and neutrons is calculated, their kinematics have to be determined. In the following, the procedure used to assign the three-dimensional momentum to nucleons and the angular correlations to the nucleon pairs is described.

1. Nucleon momentum distributions
The p_{T} distributions of nucleons are described using a Levy–Tsallis function (Eq. 16). However, the shape of the p_{T} distribution evolves with the multiplicity of the event [33]. Thus, ToMCCA requires a parameterization of the shape of the transverse momentum distributions of nucleons as a function of multiplicity. For this, the proton spectra measured by ALICE in pp collisions at a center-of-mass energy of $\sqrt{s} = 5$ TeV [20] and those selected with the high-multiplicity (HM) trigger in pp collisions at $\sqrt{s} = 13$ TeV [21] have been fitted using Eq. 16. The evolution of the Levy–Tsallis parameters ($\mathrm{d}N/\mathrm{d}y$, n, T) has been modeled as a function of $\mathrm{d}N/\mathrm{d}\eta$ according to a power law function (Eq. 13). The values of the parameters can be found in Table 4. Due to a high correlation between the parameters T and n, the latter was constrained to Eq. 13, only letting the parameter A vary slightly (7.3 ± 0.1) and fixing $C=-1$. This is to ensure that the T parameter evolves smoothly as a function of multiplicity. Furthermore, depending on the hadronization model used for the particle production (Sect. 2.2.3), the parameter T needs to be modified, since the models can show edge effects that alter the yield, especially at low multiplicities. For this, T is modified as

$$T_{\mathrm{mod}}(\mathrm{d}N/\mathrm{d}\eta) = T(\mathrm{d}N/\mathrm{d}\eta) \cdot (A + B(\mathrm{d}N/\mathrm{d}\eta)^C). \qquad (8)$$

The resulting parameterization can be found in Table 4. Since in the modification term $A \approx 1$ and $C < 0$, this correction only modifies low-multiplicity events, while the modification term tends to 1 for high-multiplicity events.

In addition to the transverse momentum, the rapidity y is drawn from a flat distribution $y \in [-0.5, 0.5]$. This is well motivated for measurements at midrapidity [34].

2. Angular correlations
To eventually fix the full three-dimensional momentum information, the relative angle between the proton and the neutron needs to be determined. Experimentally, this is accessed using the angular correlation function [35]

$$C(\Delta\varphi) = \frac{S(\Delta\varphi)}{B(\Delta\varphi)}, \qquad (9)$$

Table 4 Values of the parameters for the yield parameterization. The parameters $T_{\mathrm{mod}}^{\mathrm{QuarkReco}}$ and $T_{\mathrm{mod}}^{\mathrm{StringFrag}}$ refer to the modifications of the temperature parameter T of Eq. 8, for the quark recombination and string fragmentation hadronization methods, respectively

	dN/dy	n	T	$T_{\mathrm{mod}}^{\mathrm{QuarkReco}}$	$T_{\mathrm{mod}}^{\mathrm{StringFrag}}$
A	–	7.3230690	–	0.9953433	0.9947258
B	0.06152	2.0855418	0.10805	− 0.7380489	− 4.341114
C	0.95621	− 1	0.36889	− 2.0069334	− 4.555843

Table 5 Values for the parameterization of the angular correlation

Parameter	$B(\Delta\varphi)$
A	0.85747601
B	5.7197689
N_1	− 0.89304215
N_2	0.26041062
N_0	0.17701025

where $S(\Delta\varphi)$ is the same event $\Delta\varphi$ distribution, i.e., the angular distribution of the azimuthal angle between two particles in the same event. Conversely, $B(\Delta\varphi)$ is the $\Delta\varphi$ distribution of pairs from mixed events, i.e., relative angles between particles from different events. The latter removes trivial correlations from the same event distribution and reduces them to the genuine correlation. However, for a model like ToMCCA, the actual angle between particles in the same event is needed. Hence, only the same event angular distribution $S(\Delta\varphi)$ is required. In order to isolate this distribution, the measured $C(\Delta\varphi)$ distribution is multiplied by a $B(\Delta\varphi)$ obtained from event generators. This method is applicable since $B(\Delta\varphi)$ contains only trivial correlations due to collision geometry and energy conservation that are well reproduced in event generators. For ToMCCA, $B(\Delta\varphi)$ is obtained again from EPOS 3 and assumed to be independent of the event multiplicity. The $C(\Delta\varphi)$ for $pp + \overline{pp}$ pairs is measured by the ALICE Collaboration [35], as a function of multiplicity. However, since measurements in all multiplicity classes are compatible with each other within their uncertainties, the data measured in the 20–40% multiplicity class is used for all multiplicities. The final same event distribution $S(\Delta\varphi) = C(\Delta\varphi) \times B(\Delta\varphi)$ is expressed as

$$S(\Delta\varphi) = \underbrace{(N_0 \sin(\Delta\varphi - \pi/2) + 1)}_{C(\Delta\varphi)}$$
$$\times \underbrace{(N_1|\Delta\varphi| + B)(N_2 e^{-\Delta\varphi^2/A^2})}_{B(\Delta\varphi)}. \qquad (10)$$

The parameters can be found in Table 5.

3 Source size

The last key ingredient for coalescence is the source size r_0, which is a measure of the distance between the particles of the pair. In ToMCCA, the source is assumed to be distributed according to a Gaussian function, and thus the distances r between particles follow the distribution [36]

$$S(r; r_0) = \frac{4\pi r^2}{(4\pi r_0^2)^{3/2}} e^{-\frac{r^2}{4r_0^2}}, \qquad (11)$$

from which distances are drawn at random. The source size in small systems has been measured as a function of the average transverse mass[3] in pp collisions at $\sqrt{s} = 13$ TeV with a high-multiplicity trigger [36]. These collision events have a mean charged-particle multiplicity of about $dN_{ch}/d\eta \sim 30$. On the other hand, minimum-bias pp collisions have a mean multiplicity of $dN_{ch}/d\eta \sim 7$, and it is expected that the source size depends on multiplicity [37]. Thus, a description of the source size for arbitrary multiplicity is needed to accurately describe the deuteron production. For this purpose, the coalescence model is reversed, leaving the source size as a free parameter and including the deuteron momentum distributions in the fitting procedure of ToMCCA. The deuteron spectra are taken from the ALICE measurements in pp collisions at $\sqrt{s} = 5$ TeV [20] and from the HM sample of pp collisions at $\sqrt{s} = 13$ TeV [21], spanning a range of $dN_{ch}/d\eta$ from 2.4 to 35.8 [21,22].

The fitting procedure consists of several steps, as described in the following. First, for a given multiplicity, the deuteron p_T distributions are generated for source sizes from $r_0 = 0$ to 1.8 fm. Then, for each value of p_T, the $\chi^2 = \frac{Y_{\mathrm{ToMCCA}} - Y_{\mathrm{ALICE}}}{\Delta Y_{\mathrm{ALICE}}}$ (with Y being the yield for a given p_T, either in ToMCCA or in data, and ΔY the uncertainty of the data points) is calculated for all tested source sizes. The resulting distribution of χ^2 as a function of the source size, $\chi^2(r_0)$, can be fitted with a polynomial function of second order, $y = a(x - c)^2$. The
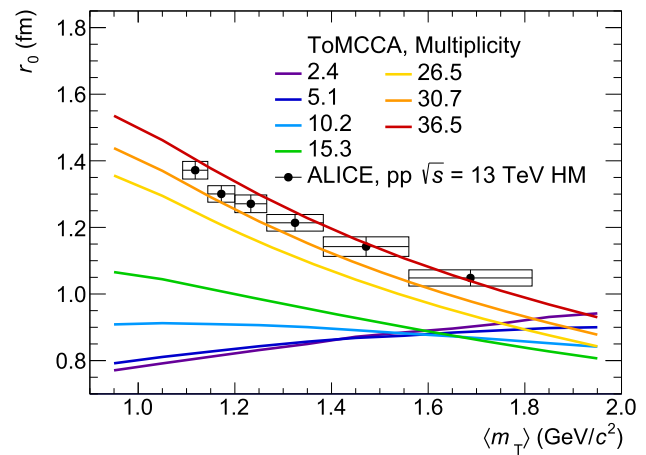
---

[3] The average transverse mass is defined as $\langle m_T \rangle = \sqrt{\left(\frac{p_{T,1} + p_{T,2}}{2}\right)^2 + \left(\frac{m_1 + m_2}{2}\right)^2}$ of the pair [36]. In the case of deuteron coalescence, the masses of the constituent nucleons are almost equal, $m_1 = m_2 = m_N \approx 0.938$ GeV/c^2, and the transverse momentum of the deuteron is given by the sum of those of the constituents, $p_{T,d} = p_{T,1} + p_{T,2}$.

Table 6 Parameterization as a function of multiplicity of the power law parameters describing the source size r_0

Parameter	$B(\mathrm{d}N/\mathrm{d}\eta)$	$C(\mathrm{d}N/\mathrm{d}\eta)$
A	0.59781	-0.85246
B	0.58983	0.83824
C	1.26481	4.25301
D	-0.48262	0.182242
N	–	0.99022
M	10.8889	–

Fig. 3 Scaling of the baryon-emitting source size as a function of $\langle m_T \rangle$, for arbitrary multiplicity values. The predictions of ToMCCA are compared with the measured source size corresponding to an event multiplicity of about 30 [36]

parameter c gives the best fitting source size, and $1/\sqrt{a}$ corresponds to the 1σ deviation. This procedure is repeated for every value of p_T covered by the measurement. The resulting source size scaling as a function of $\langle m_T \rangle$ is fitted using a power law (Eq. 13, with parameter $A = 0$). Finally, the procedure is repeated for all multiplicities in the datasets of pp collisions at $\sqrt{s} = 5$ TeV and HM at $\sqrt{s} = 13$ TeV. The resulting power law parameters B and C are fitted as a function of $\mathrm{d}N/\mathrm{d}\eta$ with Eqs. 14 and 15, respectively. The functional form of Eq. 14 has been chosen to reproduce the observation of $r_0 \propto N_{ch}^{1/3}$ at large multiplicities [38] and its saturation to a minimal source size $r_0 \sim R_p = 0.87$ fm [39] similar to the proton radius (R_p). The parameters are reported in Table 6. With this parameterization, it is possible to obtain the $\langle m_T \rangle$ scaling of the source size for arbitrary multiplicity values, shown as an example in Fig. 3. The predictions of ToMCCA are compared with the source size measured by ALICE [36] in the HM data sample of pp collisions, corresponding to an average multiplicity of about 30 charged particles, showing very good agreement with the results of the model. To date, there are no experimental measurements of the $\langle m_T \rangle$-scaling of the source radii as a function of the average charged-particle multiplicity in minimum-bias pp collisions at the Large Hadron Collider (LHC). Future measurements of this quantity with the LHC Run 3 data will be useful for a quantitative comparison with the predictions of ToMCCA.

The uncertainties associated with these predictions stem from the uncertainty of the measurement of the proton and deuteron production yields. The model uncertainties are estimated by repeating the fitting procedure using a parameterization of the lower and upper bounds of the $\mathrm{d}N/\mathrm{d}y$ parameter in Eq. 16 and taking the uncertainty from the χ^2 minimization described above. The obtained global uncertainties are 3.5% from the protons and 4.4% from the deuterons, which, added in quadrature, give $\delta B(\mathrm{d}N/\mathrm{d}\eta) = 5.7\%$. The C parameter is unchanged within these variations. Propagating this uncertainty to the predictions of the deuteron spectra, by varying $B(\mathrm{d}N/\mathrm{d}\eta)$ by 5.7%, changes the deuteron yield by $\pm 4.6\%$.
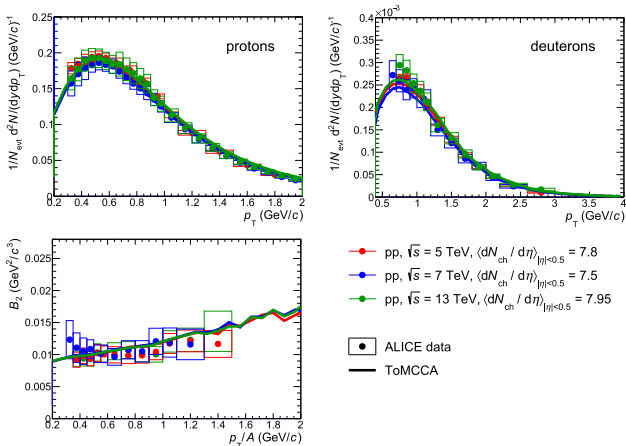
4 Results

4.1 Transverse momentum distributions and coalescence parameter

Using all the parameterizations as a function of multiplicity described in the previous sections and the coalescence afterburner based on the Wigner function, it is possible to obtain the proton and deuteron momentum distributions for any value of multiplicity. In Fig. 4, the resulting proton and deuteron momentum distributions obtained with ToMCCA are shown in the top panels and compared with the p_T distributions measured by ALICE [20,22,23,33,40]. The different colors refer to different multiplicity classes, as reported in the legend, taken from different pp collision energies. All results show excellent agreement with the data points, being able to reproduce data within 1σ for all the multiplicity classes covered by pp collisions (up to ~ 36). These results demonstrate that the coalescence model implemented in ToMCCA works for any collision energy and depends only on the chosen multiplicity class. This can be clearly seen in Fig. 5, where ToMCCA predictions are compared with results from three different energies of pp collisions, corresponding to similar multiplicity values ($\mathrm{d}N_{ch}/\mathrm{d}\eta \sim 8$). Consequently, the model presented in this article can be used for any collision energy of pp collisions, finding applicability for the astrophysics searches of indirect dark matter. Indeed, to correctly interpret future measurements of the flux of antinuclei from cosmic rays, modeling the flux of antinuclei produced in collisions between the ISM and the primary CRs is crucial, as they constitute the background for the DM searches. It was shown in [5] that antideuterons in cosmic rays most likely originate from collisions of center-of-mass energy of $\sqrt{s} = 20$–30 GeV. The multiplicities cor-
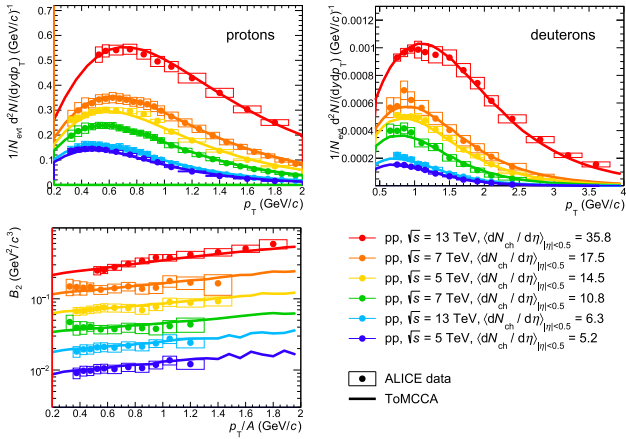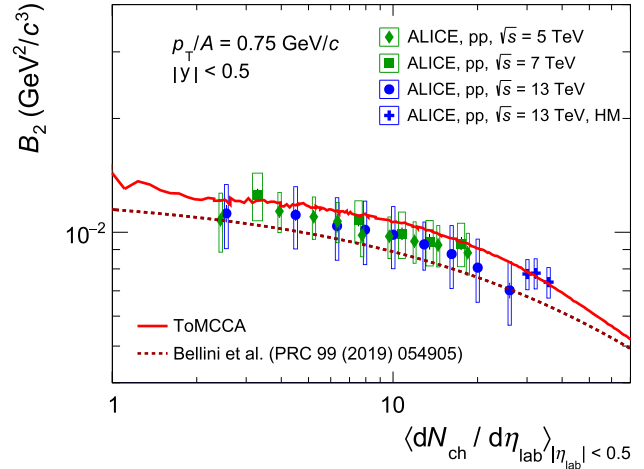
Fig. 4 Results of ToMCCA in different multiplicity classes, for the transverse momentum spectra of protons (top left) and deuterons (top right), and the coalescence parameter B_2 (bottom), compared with data measured by ALICE. The values of the average charged-particle multiplicity corresponding to each multiplicity class in the three different collision energies are indicated in the legend. The values of B_2 are scaled for better visibility, by the following factors, going from the top to the bottom: 42, 16, 8, 4, 2, 1

Fig. 6 Coalescence parameter B_2 as a function of multiplicity predicted by ToMCCA (red line), compared with the B_2 measurements from ALICE. The blue markers show the data included in the parametrizations, while the green markers show the ones not included. The dotted line is the model prediction of Ref. [37], where the coalescence parameter for deuterons with a size of 3.2 fm is computed

Fig. 5 Results of ToMCCA in the multiplicity class VI, VII, and VI+VII, respectively, for pp $\sqrt{s} = 5$, 13, and 7 TeV, for the transverse momentum spectra of protons (top left) and deuterons (top right), and the coalescence parameter B_2 (bottom), compared with data measured by ALICE. The values of the average charged-particle multiplicity corresponding to each multiplicity class (VI, VII, and VI+VII) in the three different collision energies are indicated in the legend

responding to these collision energies were measured by the Intersecting Storage Rings ISR) experiment [41] and correspond to values between $dN/d\eta = (1.50 \pm 0.01)$ and $dN/d\eta = (1.66 \pm 0.02)$, for energies between ~ 23.6 GeV and ~ 30.8 GeV. Hence, ToMCCA can be used to model the production of antideuterons at such multiplicities and to predict the expected flux of antinuclei due to the background.

Using the resulting proton and deuteron momentum distributions, the coalescence parameter B_A can be computed. It is obtained from the ratio of the invariant yield of the nucleus

with mass number A and that of the protons squared, assuming protons and neutrons have the same momentum distributions as they are isospin partners, and $p_T^p = p_T^A/A$

$$B_A = \left(\frac{1}{2\pi p_T^A} \left(\frac{d^2 N}{dy dp_T} \right)_A \right) \bigg/ \left(\frac{1}{2\pi p_T^p} \left(\frac{d^2 N}{dy dp_T} \right)_p \right)^A.$$

(12)

The resulting coalescence parameters are shown in the bottom panels of Figs. 4 and 5, in comparison with ALICE measurements. Also in this case, the predictions of ToMCCA are in excellent agreement with the experimental results ($< 1\sigma$).

4.2 Comparison with data as a function of multiplicity

The results shown in Sect. 4.1 can be obtained for arbitrary multiplicity. Hence, it is possible to scan the full range of average charged-particle multiplicity per event covered by pp collisions from ~ 1 up to ~ 70. Using the deuteron and proton p_T distributions obtained with ToMCCA, the coalescence parameter B_2 is computed, as well as the deuteron-to-proton ratio of the integrated yields (d/p). Selecting one value of p_T/A, the predictions for the B_2 of ToMCCA are shown as a function of multiplicity in Fig. 6, and compared with the data measured by ALICE. Similarly, the d/p integrated yield ratios are shown as a function of multiplicity in Fig. 7. The predictions from ToMCCA show excellent agreement with the data, within 1σ, for both the B_2 and the d/p ratio.
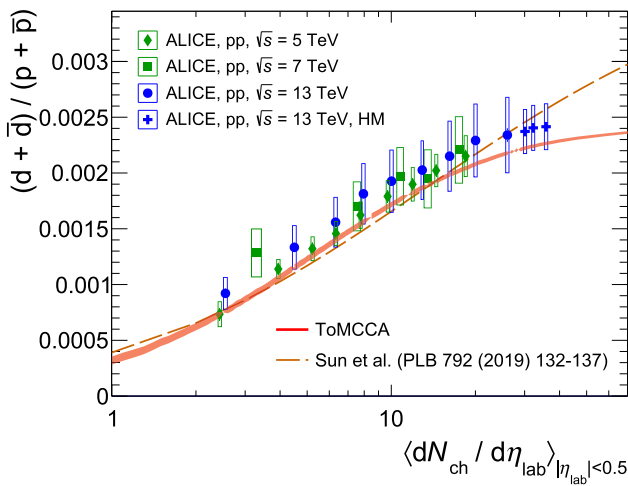
Fig. 7 Deuteron-to-proton ratio of integrated yields as a function of multiplicity obtained with ToMCCA (red solid line), compared with measurements of the ALICE Collaboration. The width of the model predictions reflects the statistical uncertainties. The dotted line is the model prediction of Ref. [14]
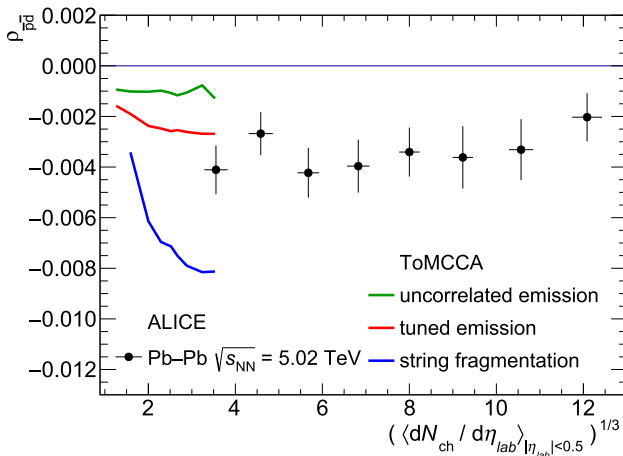
Fig. 8 Correlation coefficient for antiprotons and antideuterons ($\rho_{\bar{p}\bar{d}}$) as a function of the cubic root of the charged-particle multiplicity, predicted by ToMCCA, with the three different hadronization models, compared with the ALICE measurements in Pb–Pb collisions [42]

4.3 Prediction of the Pearson correlation coefficient

In Fig. 8, the predicted Pearson correlation coefficient for antiprotons and antideuterons ($\rho_{\bar{p}\bar{d}}$), as defined in Eq. 7, is shown. The predictions of ToMCCA are shown for the three hadronization models described in Sect. 2.2.3, i.e., tuned emission, string fragmentation, and uncorrelated emission, within a multiplicity range of up to $N_{ch} = 44$. The source size, determined by the fit from the tuned emission model and used consistently throughout this paper, is used to calculate the $\rho_{\bar{p}\bar{d}}$ regardless of the hadronization model, in order to isolate the effect of baryon number suppression on the $\rho_{\bar{p}\bar{d}}$. These predictions are compared with measurements by the

ALICE Collaboration in Pb–Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV [42]. The data points indicate that a nonzero $\rho_{\bar{p}\bar{d}}$ is necessary to match the observed suppression. However, it remains unclear whether the transition from pp to Pb–Pb collisions is smooth. Future measurements of $\rho_{\bar{p}\bar{d}}$ in pp collisions could provide constraints on $\rho_{\bar{p}\bar{d}}$.

5 Conclusions

In this paper, we have introduced ToMCCA (toy Monte Carlo coalescence afterburner), a toy Monte Carlo model for coalescence based on the Wigner function formalism. Our results show that ToMCCA is able to predict deuteron yields with precision better than 5% if all the required parameters, namely the charged-particle multiplicity, nucleon momentum distributions, angular correlations, and the two-nucleon source size, are provided from parameterizations of experimental measurements. The results of this model demonstrate that ToMCCA can be used for arbitrary center-of-mass energy of pp collisions, as the coalescence predictions depend only on the average multiplicity of the event rather than on the collision energy. This would be useful for astrophysical studies, enabling the prediction of antinuclear fluxes in cosmic rays at the multiplicities corresponding to the energies at which antinuclei are produced in the collisions of ISM and CRs. This is key for disentangling the contribution of dark matter annihilations from the Standard Model background, for correctly interpreting future measurements of antinuclear fluxes in cosmic rays.

Due to its lightweight nature and minimal complexity, ToMCCA can also be easily extended to study more complex phenomena related to light nuclei formation, either towards other collision systems, lower collision energies or lower multiplicities or towards heavier nuclei, as long as their wavefunction is known.

Data Availability Statement This manuscript has no associated data. [Author's' comment: Data sharing not applicable to this article as no datasets were generated or analysed during the current study.]

Code Availability Statement This manuscript has associated code/ software in a data repository. [Author's comment: ToMCCA is available at https://github.com/HorstMa/ToMCCAPublic.]

A Mathematical functions

Power law:

$$F(x) = A + B(x)^C. \tag{13}$$

Root-3-Sigmoid function, representing a smooth transition from cube-root behavior to zero-degree polynomial:

$$F(x) = \frac{A}{1 + e^{-B(x-M)}} x^{1/3}$$
$$+ (1 - \frac{1}{1 + e^{-B(x-M)}})C + D. \tag{14}$$

Sigmoid function:

$$F(x) = \frac{A}{(B + e^{D \cdot x - C})} + N. \tag{15}$$

Levy–Tsallis, describing the transverse momentum distributions of particles with mass m, described by a temperature parameter T, a shape parameter n, and with total yield dN/dy:

$$\frac{d^2 N}{dp_T dy} = \frac{dN}{dy} \frac{p_T}{nT} \frac{(n-1)(n-2)}{(nT + m(n-2))} \left(1 + \frac{m_T - m}{nT}\right)^{-n}. \tag{16}$$

References

1. F. Donato, N. Fornengo, P. Salati, Anti-deuterons as a signature of supersymmetric dark matter. Phys. Rev. D **62**, 043003 (2000)
2. T. Aramaki et al., Review of the theoretical and experimental status of dark matter identification with cosmic-ray antideuterons. Phys. Rep. **618**, 1–37 (2016)
3. M. Korsmeier, F. Donato, N. Fornengo, Prospects to verify a possible dark matter hint in cosmic antiprotons with antideuterons and antihelium. Phys. Rev. D **97**, 103011 (2018)
4. P. von Doetinchem et al., Cosmic-ray antinuclei as messengers of new physics: status and outlook for the new decade. J. Cosmol. Astropart. Phys. **2020**, 035 (2020)
5. L. Šerkšnytė et al., Reevaluation of the cosmic antideuteron flux from cosmic-ray interactions and from exotic sources. Phys. Rev. D **105**(8), 083021 (2022)
6. A. Kounine, The Alpha Magnetic Spectrometer on the International Space Station. Int. J. Mod. Phys. E **21**(08), 1230005 (2012)
7. C.J. Hailey, An indirect search for dark matter using antideuterons: the GAPS experiment. New J. Phys. **11**(10), 105022 (2009)
8. K. Abe et al., Measurement of the cosmic-ray antiproton spectrum at solar minimum with a long-duration balloon flight over Antarctica. Phys. Rev. Lett. **108**, 051102 (2012)
9. S.T. Butler, C.A. Pearson, Deuterons from high-energy proton bombardment of matter. Phys. Rev. **129**, 836–842 (1963)
10. J.I. Kapusta, Mechanisms for deuteron production in relativistic nuclear collisions. Phys. Rev. C **21**, 1301–1310 (1980)
11. R. Scheibl, U.W. Heinz, Coalescence and flow in ultrarelativistic heavy ion collisions. Phys. Rev. C **59**, 1585–1602 (1999)
12. W. Zhao, L. Zhu, H. Zheng, C.M. Ko, H. Song, Spectra and flow of light nuclei in relativistic heavy ion collisions at energies available at the BNL Relativistic Heavy Ion Collider and at the CERN Large Hadron Collider. Phys. Rev. C **98**, 054905 (2018)
13. K. Blum, M. Takimoto, Nuclear coalescence from correlation functions. Phys. Rev. C **99**(4), 044913 (2019)
14. K.-J. Sun, C.M. Ko, B. Dönigus, Suppression of light nuclei production in collisions of small systems at the Large Hadron Collider. Phys. Lett. B **792**, 132–137 (2019)
15. M. Kachelrieß, S. Ostapchenko, J. Tjemsland, Alternative coalescence model for deuteron, tritium, helium-3 and their antinuclei. Eur. Phys. J. A **56**(1), 4 (2020)
16. M. Kachelriess, S. Ostapchenko, J. Tjemsland, On nuclear coalescence in small interacting systems. Eur. Phys. J. A **57**(5), 167 (2021)
17. S. Mrówczyński, P. Słoń, Hadron–deuteron correlations and production of light nuclei in relativistic heavy-ion collisions. Acta Phys. Polon. B **51**(8), 1739–1755 (2020)
18. F. Bellini, K. Blum, A.P. Kalweit, M. Puccio, Examination of coalescence as the origin of nuclei in hadronic collisions. Phys. Rev. C **103**(1), 014907 (2021)
19. M. Mahlein, L. Barioglio, F. Bellini, L. Fabbietti, C. Pinto, B. Singh, S. Tripathy, A realistic coalescence model for deuteron production. Eur. Phys. J. C **83**(9), 804 (2023)
20. S. Acharya et al., Production of light (anti)nuclei in pp collisions at $\sqrt{s}$ = 5.02 TeV. Eur. Phys. J. C **82**(4), 289 (2022)
21. S. Acharya et al., Production of light (anti)nuclei in pp collisions at $\sqrt{s}$ = 13 TeV. JHEP **01**, 106 (2022)
22. S. Acharya et al., (Anti-)deuteron production in pp collisions at $\sqrt{s}$ = 13TeV. Eur. Phys. J. C **80**(9), 889 (2020)
23. S. Acharya et al., Multiplicity dependence of (anti-)deuteron production in pp collisions at $\sqrt{s}$ = 7 TeV. Phys. Lett. B **794**, 50–63 (2019)
24. U.W. Heinz, B.V. Jacak, Two particle correlations in relativistic heavy ion collisions. Annu. Rev. Nucl. Part. Sci. **49**, 529–579 (1999)
25. R.B. Wiringa, V.G.J. Stoks, R. Schiavilla, Accurate nucleon–nucleon potential with charge-independence breaking. Phys. Rev. C **51**, 38–51 (1995)
26. P. Skands, S. Carrazza, J. Rojo, Tuning PYTHIA 8.1: the Monash 2013 Tune. Eur. Phys. J. C **74**(8), 3024 (2014)
27. T. Sjöstrand, S. Ask, J.R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C.O. Rasmussen, P.Z. Skands, An introduction to PYTHIA 8.2. Comput. Phys. Commun. **191**, 159–177 (2015)
28. K. Werner, The hadronic interaction model EPOS. Nucl. Phys. B Proc. Suppl. **175–176**, 81–87 (2008)
29. S. Acharya et al., Multiplicity dependence of (multi-)strange hadron production in proton–proton collisions at $\sqrt{s}$ = 13 TeV. Eur. Phys. J. C **80**(2), 167 (2020)
30. F.-H. Liu, Unified description of multiplicity distributions of final-state particles produced in collisions at high energies. Nucl. Phys. A **810**, 159–172 (2008)
31. S. Acharya et al., First measurement of antideuteron number fluctuations at energies available at the Large Hadron Collider. Phys. Rev. Lett. **131**(4), 041901 (2023)

32. ALICE Collaboration, First measurement of antideuteron number fluctuations at energies available at the Large Hadron Collider. (2022)

33. S. Acharya et al., Multiplicity dependence of light-flavor hadron production in pp collisions at $\sqrt{s}$ = 7 TeV. Phys. Rev. C **99**(2), 024906 (2019)

34. I. Arsene et al., Rapidity dependence of deuteron production in Au + Au collisions at $\sqrt{s_{NN}}$ = 200 GeV. Phys. Rev. C **83**, 044906 (2011)

35. D. Ruggiano, Two-particle angular correlations of identified particles in pp collisions at $\sqrt{s}$ = 13 TeV with ALICE. 11 (2023)

36. S. Acharya et al., Search for a common baryon source in high-multiplicity pp collisions at the LHC. Phys. Lett. B **811**, 135849 (2020)

37. F. Bellini, A.P. Kalweit, Testing production scenarios for (anti-)(hyper-)nuclei and exotica at energies available at the CERN Large Hadron Collider. Phys. Rev. C **99**(5), 054905 (2019)

38. A. Kisiel, M. Gałażyn, P. Bożek, Pion, kaon, and proton femtoscopy in Pb–Pb collisions at $\sqrt{s_{NN}}$=2.76 TeV modeled in (3+1)D hydrodynamics. Phys. Rev. C **90**(6), 064914 (2014)

39. I. Sick, D. Trautmann, Proton root-mean-square radii and electron scattering. Phys. Rev. C **89**(1), 012201 (2014)

40. S. Acharya et al., Production of light-flavor hadrons in pp collisions at $\sqrt{s}$ = 7 and $\sqrt{s}$ = 13 TeV. Eur. Phys. J. C **81**(3), 256 (2021)

41. W. Thome et al., Charged particle multiplicity distributions in p p collisions at ISR energies. Nucl. Phys. B **129**, 365 (1977)

42. ALICE Collaboration et al., First measurement of antideuteron number fluctuations at energies available at the Large Hadron Collider. Phys. Rev. Lett. **131**, 041901 (2023)