Strongly coupled quark-gluon plasma in heavy ion collisions

Edward Shuryak

Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York 11794-3800, USA

(published 19 July 2017)

A decade ago, a brief summary of the field of the relativistic heavy ion physics could be formulated as the discovery of strongly coupled quark-gluon plasma, sQGP for short, a near-perfect fluid with surprisingly large entropy-density-to-viscosity ratio. Since 2010, the LHC heavy ion program added excellent new data and discoveries. Significant theoretical efforts have been made to understand these phenomena. Now there is a need to consolidate what we have learned and formulate a list of issues to be studied next. Studies of angular correlations of two and more secondaries reveal higher harmonics of flow, identified as the sound waves induced by the initial state perturbations. As in cosmology, detailed measurements and calculations of these correlations helped to make our knowledge of the explosion much more quantitative. In particular, their damping had quantified the viscosity. Other kinetic coefficients—the heavy-quark diffusion constants and the jet quenching parameters—also show enhancements near the critical point $T \approx T_c$. Since densities of QGP quarks and gluons strongly decrease at this point, these facts indicate large role of nonperturbative mechanisms, e.g., scattering on monopoles. New studies of the pp and pA collisions at high multiplicities reveal collective explosions similar to those in heavy ion AA collisions. These "smallest drops of the sQGP" revived debates about the initial out-of-equilibrium stage of the collisions and mechanisms of subsequent equilibration.

DOI: 10.1103/RevModPhys.89.035001

CONTENTS

I. Introduction 2
A. Before RHIC 2
B. RHIC runs in 2000–2005 and the strongly coupled QGP paradigm 3
C. The first runs of the LHC 4
D. Why is sQGP so unusual? 4
II. The Main Issues in QCD and Heavy Ion Physics 9
III. Sounds on Top of the "Little Bang" 9
A. Comments on hydrodynamics 9
B. Hydrodynamical response to perturbations 10
C. Acoustic systematics: The viscous damping 11
D. Waves from a point perturbation and harmonic spectra 12
E. Detecting the interactions between harmonics 14
F. Event-by-event v_n fluctuations and correlations 15
G. The map of the sounds 15
H. Sounds in the loops 16
IV. The Preequilibrium State, Global Observables, and Fluctuations 17
A. Perturbative versus nonperturbative models 17
B. Centrality, $E_\perp$, and fluctuations 18
C. Anisotropy and the boundaries of hydrodynamics 19
V. The Smallest Drops of QGP 20
A. Collectivity in small systems 21
B. Pedagogical digression: Scale invariance of sQGP and small systems 21
C. Comparison of the peripheral AA, central pA, and high-multiplicity pp 22
D. The size and radial flow puzzle for central pA 23
E. Radial flow in high multiplicity pp 25
F. Can flows in small systems be "fake"? 26
G. Shape fluctuations: Central pA vs peripheral AA 26
VI. Equilibration in QCD-based Models 27
A. CGC and turbulent GLASMA 27
B. From GLASMA to hydro 28
C. The initial state and angular correlations 29
D. Multistring dynamics 31
VII. Holographic Equilibration 32
A. Near equilibrium 32
B. Out of equilibrium 1: The shocks 33
C. Out of equilibrium 2: The falling shell 33
D. Out of equilibrium 3: Anisotropic plasma 34
E. Out of equilibrium 4: Rapidity-independent collisions 34
VIII. Collisions in Holography 35
A. "Trapped surfaces" and the entropy production 35
B. From holographic to QCD strings 36
C. Holographic Pomeron 37
D. Collisions at ultrahigh energies 39
IX. Electromagnetic Probes 40
A. Brief summary 40
B. New sources of photons and dileptons: Multigluon or phonon + magnetic field 41
C. Dilepton polarization and the (early time) pressure anisotropy 41
X. Heavy Quarks and Quarkonia as sQGP Probes 42
A. Quarkonium suppression 42
B. Quarkonia and lattice correlation functions 43
C. Quarkonia and real-time QFT formalism 44
D. Observed charmonium composition and chemical equilibration 45
E. Are there some stationary quarkonia states in a strongly coupled medium? 45
XI. Jet Quenching 46
A. Is jet quenching dominated by the near-T_c matter? 46
B. "Fixed points" of the jet distributions 47

XII. Near the Phase Boundary: Fluctuations and Freezeouts 47
A. Chemical freezeouts 47
B. From chemical to kinetic freezeouts 48
C. The search for the critical point and the RHIC low energy scan 50
XIII. Summary and Discussion 51
A. Progress on the big questions 51
B. Sounds 51
C. The conflicting views of the initial state 51
D. The smallest drops of sQGP 52
E. Heavy quarks and quarkonia 52
F. Jets 52
1. Bjorken flow 56
2. Gubser flow 56
Acknowledgments 53
Appendix A: Heavy Ion Terminology 53
Appendix B: Relativistic Hydrodynamics 55
Appendix C: Introduction to Gauge-gravity Duality 58
References 59

I. INTRODUCTION

The history of the ultrarelativistic heavy ion collisions can be roughly divided into three periods as follows:

(i) *formulation of the scientific goals and forming the community*: from the 1950s to the year 2000, the start of operation of the Relativistic Heavy Ion collider (RHIC);

(ii) *crucial experimental and theoretical developments* during 2000–2004, reaching the goal of quark-gluon plasma (QGP) production and first estimates of its properties; and

(iii) the *last decade*, 2005–now.

This review attempts to provide theoretical and experimental summaries in the last decade. It is mostly addressed to people belonging to the heavy ion community. Some pedagogical material on heavy ion terminology, relativistic hydrodynamics, and holographic models is placed in the Appendix sections at the end of the paper.

Experimental results, from RHIC and now the Large Hadron Collider (LHC), continue to provide answers to old question and add new puzzles. While looking through the talks at the latest quark matter conferences, one observes a relative weakness of the theory: this review tries to elucidate current trends. It is not a comprehensive summary of the field. Several important directions are not covered because (in my opinion) they have not yet matured to a conclusive stage. One such direction is related to fluctuations in the gauge topology and chiral magnetic effect. Another is the RHIC beam energy scan, with an emphasis on critical fluctuations aimed at a possible location of the critical point. There are also other topics not covered simply because of space limitations and interests.

A. Before RHIC

Perhaps the first question one asks in the field is whether a tiny system created in high energy collisions, of nucleons or nuclei, can indeed be large enough to be treated as macroscopically large. An affirmative answer would mean that some new form of matter is produced, close to its equilibrium. A negative answer would imply that one deals with a multiparticle system far from equilibrium.

Already in the 1950s, when only the very first hints of multiparticle production reactions were detected in cosmic ray events, three famous physicists conjectured that the answer to the previous question would be positive. Fermi (1951) argued that a strongly interacting system of particles should rapidly equilibrate and predicted that multiplicity should grow with the center-of-mass (c.m.) energy $\sqrt{s}$ as $N \sim s^{1/4}$. Pomeranchuk (1951) replied immediately: if the interaction is strong enough for rapid equilibration, then the particles would be able to leave the system only at what we now call the freezeout conditions. Landau (1953) connected the initial stage of Fermi with the final stage of Pomeranchuck via relativistic hydrodynamics. He pointed out that in the expected quasiadiabatic expansion the entropy must be preserved, and thus Fermi's prediction for entropy and multiplicity generation should hold.

In a later review, Landau related these expectations of very strong coupling at high momenta scales, as expected from QED and other quantum field theories (QFTs) known at the time. This last argument was dramatically reversed two decades later, when asymptotic freedom of non-Abelian gauge theories and QCD was discovered by Gross, Politzer, and Wilczek in 1973. This discovery initiated studies of the high temperature T (or large chemical potential μ, or both) limit of hadronic matter using perturbative diagrams and their resummations. It was found that, unlike vacuum fluctuations of the gluon fields providing the asymptotic freedom or antiscreening, thermal fluctuations of both gluon and quark fields lead to screening in matter; thus it was called QGP. For early reviews on finite-T QCD, see Shuryak (1980) and Gross, Pisarski, and Yaffe (1981). Very hot QGP is now referred to as weakly coupled (wQGP). While the simplest from a theoretical perspective, it is not yet reached even at the highest temperatures at the LHC collider.

While during the 1970s and 1980s the field and its conferences were mostly run by theorists, appropriate experimental observables were high on the agenda. Collective hydrodynamical explosion in the transverse plane and penetrating probes (photons and dileptons) were emphasized (Shuryak, 1980). Important QGP signals suggested in the 1980s were the jet quenching (Bjorken, 1982) and subsequent melting of charmonium and bottonium states (Matsui and Satz, 1986).

Two "experimental homes" of the field are the European Center for Nuclear Research near Geneva, known by its French abbreviation CERN, and Brookhaven National Laboratory (BNL) on Long Island, New York. The first round of fixed target experiments started in both places soon led to the realization that the collision energy is insufficient and thus heavy ion colliders are needed. The decision of the U.S. nuclear physics community was to build RHIC, completed in 2000, initially with four detectors: the larger STAR and PHENIX and the smaller PHOBOS and BRAMHS.

Prior to the RHIC era, it was widely assumed that the wQGP regime extends down to the phase transition point. Numerical simulations of QCD-like theories on supercomputers, using space- (Euclidean) time discretization—"lattice

QCD" for short—had provided first-principle results on QCD thermodynamics.[1] A crossover transition at

$$T_c \approx 155 \text{ MeV} \tag{1}$$

separates the confining phase with the broken chiral symmetry from the QGP region at $T > T_c$. The same value is observed experimentally as "chemical freezeout" defining the composition of outgoing hadrons.

Most QCD theorists were skeptical, predicting that the RHIC program, aimed at the production of a new form of matter, would basically fail. Perturbative estimates in the framework of various parton cascades indicated an unrealistically long equilibration time and predicted "fireworks of minijets" rather than a collective explosion.

B. RHIC runs in 2000–2005 and the strongly coupled QGP paradigm

From the very first RHIC data, it became clear that what is seen is a rather spectacular collective explosion, with spectra and especially the elliptic flows found to be in excellent agreement with predictions of relativistic hydrodynamics. Especially successful were its versions supplemented by hadronic cascade at freezeout (Teaney, Lauret, and Shuryak, 2001a, 2001b; Hirano *et al.*, 2006). All relevant dependences (as a function of[2] transverse momentum $p_\perp$, centrality, particle mass, rapidity, and collision energy) were well reproduced.

It took a few years to document the extent of jet quenching, but by 2004 the case for a "quark-gluon plasma discovery" had been officially made, in a theory workshop at Brookhaven and then in summarizing "white papers" produced by all four RHIC collaborations. Those, especially by PHENIX (Adcox *et al.*, 2005) and STAR (Bellwied, 2005), are written as extensive reviews and both are recommended as a pedagogical introduction to the field.

Perturbative parton cascades and hydrodynamics made very different predictions for the elliptic flow parameter[3] v_2. Let us mention one of them: the dependence on the transverse momentum $v_2(p_\perp)$. The parton model describes colliding nucleons and nuclei in terms of "partons," quarks, and gluons with a certain fraction of the total momentum x and "resolution scale" Q. For large enough $Q \gg 1$ GeV parton-parton scattering locally knows only the direction of the impact parameter $\vec{b}$ toward the parton it scatters with, but not the overall impact parameter of two nuclei. So high $p_\perp$ partons are, to first approximation, distributed isotropically in azimuthal angle ϕ. Only the low $p_\perp$ partons, with wavelength comparable to the size of the overlapping regions of both nuclei $1/p_\perp \sim R$ may have a substantial anisotropy in ϕ. So, in the first approximation, such models have no reason for v_2 to increase with $p_\perp$. Furthermore, since next-order inelastic production of new partons dominates over their absorption, it was predicted that more partons appear in the longest direction of the overlap "almond" $\phi = \pm\pi/2$, so v_2 was predicted to be negative.

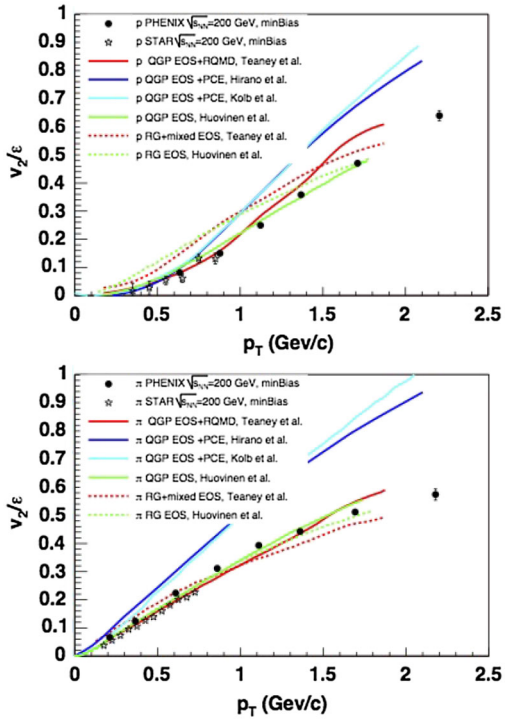

FIG. 1. Elliptic flow parameter v_2, the experimental data vs the hydrodynamical predictions for pions (upper plot) and protons (lower plot). From Adcox *et al.*, 2005.

Hydrodynamics, on the other hand, predicts an anisotropic explosion driven by the pressure gradient. The largest gradient is in the shorter direction of the almond $\phi = 0, \pi$, so v_2 was predicted to be positive. Furthermore, larger $p_\perp$ particles originate from the edge of the fireball moving toward the observer. The consequence of such "splash" geometry is that v_2 was predicted to increase linearly with $p_\perp$. The observed dependences of the elliptic flow parameter v_2 for pions and protons on the particle transverse momentum are shown in Fig. 1: so v_2 does grow linearly with p_t, reaching the large value ~0.2 as predicted by hydrodynamics.

Note that while the agreement is quite good through the whole kinematic region shown in this plot, theoretical curves stop at $p_\perp \approx 2$ GeV. Indeed, even the bravest theorists at the time had not dared to venture beyond it, since it was commonly expected that above such $p_\perp$ one would enter a jet-dominated regime. Yet a hydrodominated region was later found to extend to $p_\perp \sim 4$ GeV, and the transition to a powerlike jet regime in fact is observed only at $p_\perp > 10$ GeV. It turned out that the jet yield, and with it contributions to hard particle spectra, are strongly suppressed by jet quenching. This phenomenon is demonstrated by Fig. 2, also from a PHENIX white paper (Adcox *et al.*, 2005).

These observations and conclusions were of course scrutinized in the years to come, and the range and degree of the agreement with this picture in fact only increased. In particular, the ellipticity parameter v_2 has been measured using correlations of several (4, 6, etc.) particles, confirming the

[1]The thermodynamic quantities such as energy density and pressure are briefly introduced in Appendix A.

[2]For the interested readers, some explanation of what those variables are can be found in Appendix A.

[3]See Appendix A for the definition.

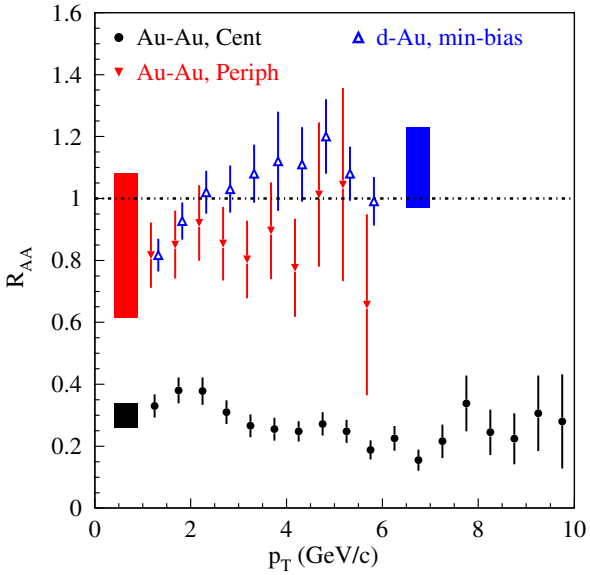

FIG. 2. .The ratio R_{AA} of the yield of high $p_\perp$ hadrons in d-Au and Au-Au collisions to the expected one according to parton model scaling. Strong deviations from one of this ration indicate the jet quenching phenomenon. From Adcox *et al.*, 2005.

conclusion that this is truly a collective phenomenon. Indeed thousands of secondaries do share the same anisotropic distribution in a given event.

These data, and their comparison with theoretical approaches, had created a new paradigm called "strongly coupled" quark-gluon Plasma (sQGP). In order to quantify it, one needs to introduce certain kinetic coefficients, parameters describing deviations from ideal hydrodynamics in an expansion in gradients. The first one, the shear viscosity entering the Navier-Stokes (NS) first-order term, is usually mentioned as a ratio to the entropy density, since both are $O(T^3)$ in scale-invariant QGP. Since the viscosity is inversely proportional to the scattering cross section times the density, we prefer to use the inverse ratio, entropy-density-to-viscosity. Its value turned out to be unexpectedly large

$$\frac{s}{\eta} \approx 5\text{–}10 \gg 1. \tag{2}$$

Thus the sQGP is among "the most perfect fluids" known.

The only "competitor" to sQGP, with comparably large entropy-to-viscosity ratio, is the so-called "unitary gas" of trapped fermionic atoms, with the scattering length $a \to \infty$. Both fluids demonstrated spectacular elliptic flows. Note that those two fluids are at the opposite end of the scale known to physics: sQGP has T measured in a fraction of GeV while T of the unitary gas is measured in nK, or 22 orders of magnitude smaller. Yet the scale hardly matters: in fact, both of them are nearly scale invariant by themselves. The unusual kinetic parameters of them are two major puzzles of modern many-body physics.

To understand why it was unexpected, it is convenient to compare it to various predictions. Kinetic theory (valid for weak coupling) interprets this dimensionless ratio as the product of the particle density, transport cross section, and inverse power of the mean velocity

$$\frac{s}{\eta} \sim \frac{n\sigma_{\text{transport}}}{T\bar{v}} \tag{3}$$

or simply as the ratio of interparticle separation to the mean free path. In a kinetic regime the latter is larger, so this ratio is expected to be small. More specifically, in weakly coupled plasma (Arnold, Moore, and Yaffe, 2003) it is

$$\frac{s}{\eta} = \frac{g^4 \log(2.42/g)}{5.12}. \tag{4}$$

Since it was derived assuming the "electric scale" is small compared to momenta $M_D \sim gT \ll T$, it should perhaps be used only[4] for $g < 1$. Equation (4), valid at high T, was not intended to be used at $T \sim T_c$, where the densities of quarks and gluons are suppressed by confinement, specifically by the vanishing Polyakov line $\langle P(T \to T_c)\rangle \to 0$. The contribution of the gluon-gluon scattering to the (inverse) ratio η/s including this effect is shown in the lower part of Fig. 3.

The nonperturbative models (see Fig. 3) predict otherwise: s/η has a peak at T_c. Qualitatively similar behavior of kinetic coefficients is known for other fluids near their phase transitions [see the examples in Csernai, Kapusta, and McLerran (2006)]. The peak in shear viscosity correlates with similar peaks claimed for other kinetic parameters: the heavy-quark diffusion constant discussed in Sec. X and the jet quenching parameter $\hat{q}$ discussed in Sec. XI; see Fig. 57.

C. The first runs of the LHC

The European nuclear and particle physicists decided to share the same LHC. One detector, ALICE, was built specially for heavy ion conditions, capable to work with high multiplicities reaching 10 000 or so in an event. Two other collaborations, CMS and ATLAS, although built mostly for high energy physics goals, both include subgroups focused on heavy ions as well.

The first runs of the LHC were expected with obvious interest. On the one hand, the energy of this collider is about a factor of 20 higher than that of RHIC, which leads to about twice larger multiplicity. Hydrodynamics predicted stronger flows: indeed, at the LHC relatively "stiff" QGP play a larger role than "soft" matter near and below the critical point T_c. On the other hand, entering a higher T domain means a higher momentum scale, reaching well over 1 GeV, where many colleagues expected the onset of asymptotic freedom and perturbative regime.

The first LHC run took place in 2010; radial and elliptic flows were observed to be even stronger than at RHIC, confirming the hydrodynamic predictions once again. More recently the energy of the LHC has been doubled, and collective effects such as elliptic flow get further enhanced as well.

D. Why is sQGP so unusual?

Attempts to answer this question led to significant progress in our understanding of the finite-T QCD. Its important part are large-scale lattice gauge simulations, which not only

[4]See Fig. 4 and the discussion of this point.

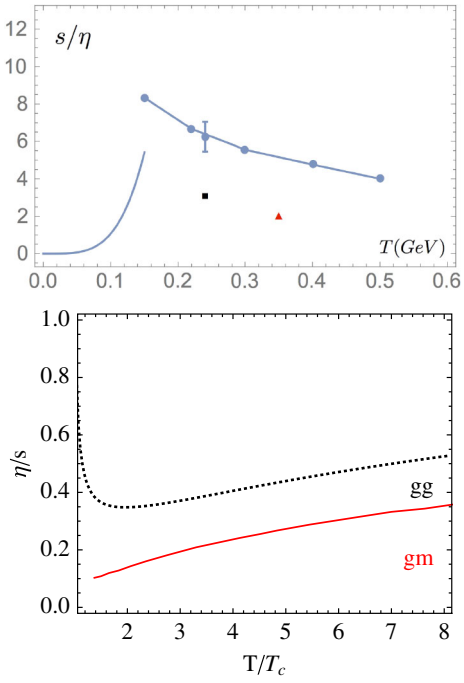

FIG. 3. Upper plot: The entropy-density-to-viscosity ratio s/η vs the temperature T (GeV). The upper range of the plot $s/\eta = 4\pi$ corresponds to the value in infinitely strongly coupled $\mathcal{N} = 4$ plasma (Policastro, Son, and Starinets, 2001). The curve without points on the left side corresponds to pion rescattering according to chiral perturbation theory (Prakash *et al.*, 1993). The triangle (red) corresponds to the molecular dynamics study of classical strongly coupled colored plasma (Gelman, Shuryak, and Zahed, 2006a), and the square (black) corresponds to a numerical evaluation (Nakamura and Sakai, 2005) on the lattice. The single point with error bar corresponds to the phenomenological value extracted from the data; see the text. The series of points connected by a line correspond to gluon-monopole scattering (Ratti and Shuryak, 2009). Lower plot: The inverse ratio η/s as a function of the temperature normalized to its critical value T/T_c. The solid line corresponds to gluon-monopole (gm) scattering (Ratti and Shuryak, 2009), the same as in the upper plot, and the dashed line shows the perturbative gluon-gluon (gg) scattering.

quantified the thermodynamical quantities of QGP and located confinement and chiral symmetry breaking phase transitions, but also elucidated physics of certain nonperturbative and topological effects. Presenting all of these results systematically goes well beyond the present paper, which is focused on heavy ion collisions. Instead we outline in this section the main ideas proposed on a nontechnical intuitive level.

At high T the asymptotic freedom limits the coupling at scale T, while plasma screening reduces the interaction below the momentum scale gT: so the hot QGP regime is amenable to a perturbative description. As T decreases, toward the end of the QGP phase at T_c, the effective coupling grows. Opinions differ on how one should describe matter in this domain. Different schools of thought can be classified as (i) perturbative, (ii) semiclassical, (iii) dual magnetic, and (iv) dual holographic ones.

(i) A "perturbative school" suggests that as one reduces temperature from its high values down to $T \approx T_c$ nothing crucial happens. The basic conclusions based on perturbative diagrams and their resummation hold. Quark and gluon quasiparticles are dominant, together with plasma-related phenomena such as plasmon and plasmino excitation modes. Perturbative formulas qualitatively hold even when the running coupling reaches values of $\alpha_s = g^2/4\pi \sim 0.5$, $g \approx 2.5$ or so. In particular, using this logic for viscosity calculated perturbatively (Arnold, Moore, and Yaffe, 2003) one gets the ratio mentioned ~2, only a factor of 2–5 lower than the data. Considering higher order processes one can recover this factor; see, e.g., the discussion in Xu and Greiner (2005).

The boundary of perturbative and nonperturbative domains is an important subject, to which, unfortunately, not enough attention has been devoted. The perturbative (pQCD) diagrams provide the answer in the form of a series in powers of $\alpha_s/\pi = g^2/4\pi^2$, so naively it can be used for say $\alpha_s < 1/2$. And yet, looking at the Particle Data Group plot of the running QCD coupling, one finds that all but one coupling determinations are at the scale $Q \approx 10$ GeV and above. The only exception, the τ lepton decays, defining $\alpha_s(Q = m_\tau)$ is based on a carefully selected combination of the vector and axial spectral densities, in which the nonperturbative effects are maximally suppressed. In general, studies of point-to-point correlation functions, from phenomenology or lattice, are the most direct source of information about the relative magnitude of perturbative and nonperturbative effects. The most relevant to the question of perturbative treatment of gluons is correlators of the type

$$\langle G^2_{\mu\nu}(x) G^2_{\mu\nu}(0)\rangle, \qquad \langle G_{\mu\nu}\tilde{G}_{\mu\nu}(x) G_{\alpha\beta}\tilde{G}_{\alpha\beta}(0)\rangle.$$

As originally observed by Novikov *et al.* (1981), their perturbative behavior sets at a much higher scale for $Q^2 > 10$ GeV^2. Lattice studies confirm that and in general show that the lowest glueball masses with the most quantum numbers are about 2–3 GeV, while the perturbative continuum in the spectral densities is setting in well above that.

The perturbation theory at finite T is different from its vacuum version: resummations of the new IR divergent series become mandatory, and the resulting series goes in powers of g rather than α_s/π. Where the perturbative series is convergent depends on higher order coefficients, which are rarely known. An exceptional case is the QGP free energy, to which much effort has been invested. The first perturbative corrections, calculated by myself in the 1970s, was an order of magnitude smaller than the zeroth order, so the beginning of a perturbative series looked promising. Many years of effort resulted in further terms [we use those from Kajantie *et al.* (2003)] plotted as a function of the coupling g in Fig. 4, as ratios to the first $O(g^2)$ perturbative effect. From this plot one can judge at which coupling these series are reliable. Unfortunately, we do not yet know the magnitude of perturbative corrections to kinetic coefficients.

(ii) What can be called the "semiclassical direction" focuses on evaluation of the path integral over the fields using generalization of the saddle point method. The extrema of its integrand are identified and their contributions evaluated. It is so far most developed in the quantum-mechanical models, for which two- and even three-loop corrections have been

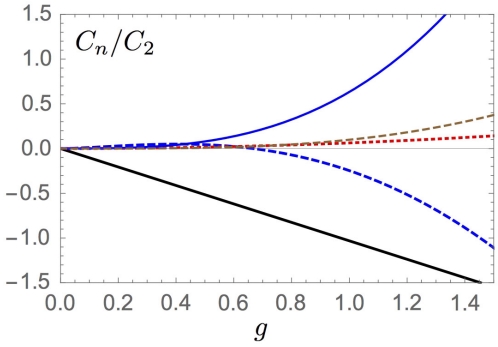

FIG. 4. Ratio of subsequent perturbative corrections to the first one $O(g^2)$: g^3, $g^4\log(g)$, g^4, g^5, and $g^6\log(g)$ are shown vs g by the thick black, thick dashed (blue), dotted (red), and thin blue and brown dashed curves, respectively.

calculated. In the case of gauge theories extrema are "instantons," a complementing perturbative series by terms of $\sim\exp(-\text{const}/g^2)$ times the so-called "instanton series" in g^2. This result in the so-called *trans-series*, which are not only more accurate than perturbative ones, but they are supposed to be free from ambiguities and unphysical imaginary parts which perturbative and instanton series separately have.

For the finite-temperature applications, plugging logarithmic running of the coupling into such exponential terms one finds some power dependences of the type

$$e^{-S}\sim\exp\left(-\frac{\text{const}}{g^2(T)}\right)\sim\left(\frac{\Lambda}{T}\right)^{\text{power}}. \tag{5}$$

As a result, these effects are not important at high T but explode, as inverse powers of T, near T_c.

In the 1980s–1990s it was shown how instanton-induced interactions between light quarks break the chiral symmetries, the $U_A(1)$ explicitly and $SU(N_f)$ spontaneously. The latter is understood via collectivization of fermionic zero modes; for a review, see Schafer and Shuryak (1998). Account for nonzero average Polyakov line or nonzero vacuum expectation value of the zeroth component of the gauge potential[5] $\langle A_0\rangle$ requires redefined solitons, in which this gauge field component does not vanish at large distances. Account for this changed instantons into a set of N_c instanton constituents, the so-called Lee–Li–Kraan–van Baal (LLKvB) instanton dyons, or instanton -monopoles (Kraan and van Baal, 1998; Lee and Lu, 1998). It was recently shown that those, if dense enough, can naturally generate both confinement and chiral symmetry breaking; see Liu, Shuryak, and Zahed (2015) and Larsen and Shuryak (2016), and for a recent review see Shuryak (2016). These works are, however, too recent to have impact on heavy ion physics, and we will not discuss it in this review.

(iii) A "dual magnetic" school consists of two distinct approaches. A "puristic" point of view assumes that at the momentum scale of interest the electric coupling is large $\alpha_s\gg 1$, and therefore there is no hope to progress with the usual "electric" formulation of the gauge theory, and therefore one should proceed with building its "magnetic" formulation, with weak "magnetic coupling" $\alpha_m=1/\alpha_s\ll 1$. Working examples of effective magnetic theory of such kind were demonstrated for supersymmetric theories (Seiberg and Witten, 1994). For applications of the dual magnetic model to QCD flux tubes see Baker, Ball, and Zachariasen (1997).

A more pragmatic point of view, known as the "magnetic scenario," starts with the acknowledgment that both electric and magnetic couplings are close to 1, $\alpha_m\sim\alpha_e\sim 1$. So, neither perturbative, semiclassical, nor dual formulation will work quantitatively. Effective masses, couplings, and other properties of all coexisting quasiparticles—quarks, gluons, and magnetic monopoles—can be deduced only phenomenologically from the analysis of lattice simulations. We will discuss this scenario next in this section.

(iv) Finally, very popular during the last decade were "holographic dualities," connecting strongly coupled gauge theories to a string theory in the curved space with extra dimensions. As shown by Maldacena (1999), in the limit of the large number of colors $N_c\to\infty$, it is a duality to a much simpler, and weakly coupled, theory, a modification of classical gravity. Such duality relates problems we want to study holographically to some problems in general relativity (GR). In particular, the thermally equilibrated QGP at strong coupling is related to certain black hole (BH) solutions in five dimensions, in which the plasma temperature is identified with the Hawking temperature, and the QGP entropy with the Bekenstein entropy. For interested readers who need terminological introduction see Appendix C. Section VII summarizes studies of the out-of-equilibrium settings, in which a bulk black hole is initially absent, but then is dynamically generated. Holographic models of the anti–de Sitter (AdS) and QCD types also lead to new views on the QCD strings, Reggeons, and Pomerons; see Sec. VIII.C.

Completing this round of comments, we now return to (iii), the approach focused on magnetically charged quasiparticles, and provide more details on its history, basic ideas, and results.

J. J. Thompson, the discoverer of the electron, noticed that something unusual already happened for static electric and magnetic charges existing together. While both the electric field $\vec{E}$ (pointing from the center of the electric charge e) and the magnetic field $\vec{B}$ (pointing from the center of the magnetic charge g) are static (time independent), the Pointing vector

$$S=[\vec{E}\times\vec{B}]$$

indicates that the electromagnetic field energy rotates. In fact, requiring the resulting angular momentum to be quantized, to an integer times $\hbar$, one get the Dirac condition (see later).

A. Poincaré went further, allowing one of the charges to move in the field of another. The Lorentz force

$$m\ddot{\vec{r}}=-eg\frac{[\dot{\vec{r}}\times\vec{r}]}{r^3} \tag{6}$$

is proportional to the product of two charges, electric e and magnetic g. The total angular momentum of the system includes the field contribution

[5]Note that in gauge theory at finite temperatures there is a preferable frame, in which matter is at rest. Therefore this expectation value does not contradict Lorentz invariance of the vacuum.

$$\vec{J} = m[\vec{r} \times \dot{\vec{r}}] + eg\frac{\vec{r}}{r}. \quad (7)$$

Its conservation leads to unusual consequences: unlike in the case of the usual potential forces, in this case the particle motion is not restricted to the scattering plane, normal to $\vec{J}$, but to a different 2D surface, called the Poincaré cone.

In the quantum-mechanical setting the problems involving a pair of electrically and magnetically charged particles provide further surprises. The angular momentum of the field must take values proportional to $\hbar$ with integer or semi-integer coefficient: this leads to the famous Dirac quantization condition (Dirac, 1931)

$$eg = \tfrac{1}{2}\hbar c n \quad (8)$$

(where we keep $\hbar$, unlike most other formulas) with an integer n on the right-hand side (rhs). Dirac himself derived it differently, arguing that the unavoidable singularities of the gauge potential of the form of the Dirac strings should be pure gauge artifacts and thus invisible. He emphatically noted that this relation was the first suggested reason for electric charge quantization.

Many outstanding theorists, Dirac and Tamm among them, wrote papers about a quantum-mechanical version of the problem of a monopole moving in a field of a charge, yet this problem was fully solved only decades later (Boulware *et al.*, 1976; Schwinger *et al.*, 1976). It is unfortunate that this problem is not, to our knowledge, part of any textbooks on quantum mechanics. The key element was substitution of the usual angular harmonics $Y_{lm}(\theta, \phi)$ by other functions, which for large l, m replicates the Poincaré cone rather than the scattering plane.

The resurfaced interest to monopoles in the 1970s was inspired by the discovery of the 't Hooft–Polyakov monopole solution ('t Hooft, 1974; Polyakov, 1974) for the Georgi-Glashow model, with an adjoint scalar field complementing the non-Abelian gauge field. In the $(2+1)$-dimensional theories such monopoles play the role of instantons: their long-range interaction was used by Polyakov (1977) to prove confinement, in a gauge theory in this dimension of space-time.

In the real world, with $3+1$ space-time dimensions, the monopoles are quasiparticles. A different confinement mechanism has been conjectured (Mandelstam, 1976; 't Hooft, 1978): monopoles may undergo Bose-Einstein condensation, provided their density is large enough and the temperature sufficiently low. These ideas, known as the "dual superconductor" model, were later strongly supported by lattice studies.

The monopole story continued at the level of QFTs, with another fascinating turn. Dirac considered the electric and magnetic charges e and g to be some fixed parameters: but in QFTs the charges run as a function of the momentum scale. So, to keep the Dirac condition, $e(Q)$ and $g(Q)$ must be running in the opposite directions, keeping their product fixed. In QCD-like theories the electric coupling is small in UV (large Q) but increase toward IR (small Q).

One example was provided by the $\mathcal{N}=2$ supersymmetric theory for which a partial solution was found by Seiberg and Witten (1994). In this theory, possessing adjoint scalar fields, the monopoles do exist as particles with well-defined masses. Furthermore, for certain special values of the vacuum expectation value (VEV) of the Higgs field, they do indeed become massless and weakly interacting, while the electric ones, gluons and gluinos, are very heavy and strongly interacting. The corresponding low energy magnetic theory is (supersymmetric) QED, and its beta function, as expected, has the opposite sign to that of the electric theory.

More examples are provided by the four-dimensional conformal theories, such as $\mathcal{N}=4$ super-Yang-Mills. Those theories are electric-magnetic self-dual. This means that monopoles, dressed by all fermions bound to them, form the same spin multiplet as the original fields of the "electric theory." Therefore, the beta function of this theory should be equal to itself with the minus sign. The only solution to that is that it must be identically zero, so the theory is conformal.

Completing this brief pedagogical update, let us return to Liao and Shuryak (2007), considering properties of a classical plasma, including both electrically and magnetically charged particles.[6] Let us proceed in steps of complexity of the problem, starting from three particles: a pair of $\pm q$ static electric charges, plus a monopole which can move in their "dipole field." Numerical integration of the equation of motion showed that the monopole's motion takes place on a curious surface, interpolating two Poincaré cones with ends at the two charges: so to say, two charges play ping-pong with a monopole, without even moving. Another way to explain it is by noting that an electric dipole is "dual" to a "magnetic bottle," with magnetic coils, invented to keep electrically charged particles inside.

The next example was a cell with eight alternating static positive and negative charges, modeling a grain of salt. A monopole, which is initially placed inside the cell, has formidable obstacles to get out of it: hundreds of scattering with the corner charges happen before it takes place. The Lorentz force acting on a magnetic charge forces it to rotate around the electric field. Closer to the charge the field grows and thus rotation radius decreases, and eventually two particles collide.

Finally, multiples (hundreds) of electric and magnetic particles were considered by Liao and Shuryak (2007), moving according to a classical equation of motions. It was found that their paths essentially replicate the previous example, with each particle being in a "cage," made by its dual neighbors. These findings provide some explanation of why electric-magnetic plasma has an unusually small mean free path and, as a result, an unusually perfect collective behavior.

At the quantum-mechanical level the many-body studies of such plasma are still to be done. Therefore one has to rely on kinetic theory and binary cross sections. Those for gluon-monopole scattering were calculated by Ratti and Shuryak (2009). It was found that gluon-monopole scattering dominates over the gluon-gluon one, as far as transport cross sections are concerned, and produces values of the viscosity

[6]We are not aware of other attempts to study such setting, although it is hardly possible that nobody thought of it.

quite comparable with that observed in sQGP experimentally, as shown in Fig. 3. What is also worth noting is that it does predict a maximum of this ratio at $T = T_c$, reflecting the behavior of the density of monopoles.

Returning to QCD-like theories which do not have powerful extended supersymmetries that would prevent any phase transitions and guarantee smooth transition from UV to IR, one finds a transition to confining and chirally broken phases. Those have certain quantum condensates which divert the renormalization group (RG) flow to a hadronic phase at $T < T_c$. Therefore the duality argument must hold at least in the plasma phase, at $T > T_c$. We can follow the duality argument and the Dirac condition only halfway, until $e^2/4\pi\hbar c \sim g^2/4\pi\hbar c \sim 1$. This is a plasma of coexisting electric quasiparticles and magnetic monopoles.

One can summarize the picture of the so-called magnetic scenario by a schematic plot shown in Fig. 5, from Liao and Shuryak (2007). At the top (the high T domain) and at the right (the high density domain) one finds weakly coupled or "electrically dominated" regimes (wQGP). On the contrary, near the origin of the plot, in vacuum, the electric fields are subdominant and confined into the flux tubes. The vacuum is filled by the magnetically charged condensate, known as a dual superconductor. The region in between (relevant for matter produced at RHIC and the LHC) is close to the "equilibrium line," marked by $e = g$ on the plot. In this region both electric and magnetic coupling are equal and thus $\alpha_{\text{electric}} = \alpha_{\text{magnetic}} = 1$: so neither the electric nor magnetic formulations of the theory are simple.

Do we have any evidence of a presence or importance for heavy ion physics of magnetic objects? Here are some arguments for that based on lattice studies and phenomenology, more or less in historical order.

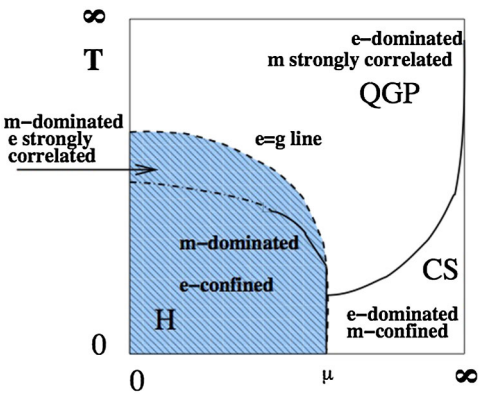

FIG. 5. A schematic phase diagram on a (compactified) plane of temperature and baryonic chemical potential $T - \mu$. The shaded (blue) region shows the magnetically dominated region $g < e$, which includes the deconfined hadronic phase as well as a small part of the QGP domain. The unshaded region includes the electrically dominated part of QGP and the color superconducting (CS) region, which has e-charged diquark condensates and is therefore "magnetically confined." The dashed line ($e = g$) is the line of electric-magnetic equilibrium. The solid lines indicate true phase transitions, while the dash-dotted line is a deconfinement crossover line. From Liao and Shuryak, 2007.

(i) In the RHIC-LHC region $T_c < T < 2T_c$ the VEV of the Polyakov line $\langle P \rangle$ is substantially different from 1. Hidaka and Pisarski (2008) argued that $\langle P \rangle$ must be incorporated into a density of thermal quarks and gluons and thus suppress their contributions. They called such matter "semi-QGP" emphasizing that only about one-half of the QGP degrees of freedom should actually contribute to thermodynamics at such T. And yet, the lattice data insist that the thermal energy density remains close to the T^4 trend nearly all the way to T_c.

(ii) The magnetic scenario (Liao and Shuryak, 2007) proposed to explain this puzzle by ascribing "another half" of such contributions to the magnetic monopoles, which are not subject to $\langle P \rangle$ suppression because they do not have the electric charge. A number of lattice studies found magnetic monopoles and showed that they behave as physical quasiparticles in the medium. Their motion definitely shows Bose-Einstein condensation at $T < T_c$ (D'Alessandro, D'Elia, and Shuryak, 2010). Their spatial correlation functions are plasmalike. Even more striking is the observation (Liao and Shuryak, 2008) revealing magnetic coupling which grows with T, being indeed an inverse of the asymptotic freedom curve.

The magnetic scenario also has difficulties. Unlike the instanton dyons, lattice monopoles so far defined are gauge dependent. The original t'Hooft–Polyakov solution requires an adjoint scalar field, absent in the QCD Lagrangian, but perhaps an effective scalar can be generated dynamically. In the Euclidean time finite-temperature setting this is not a problem, as A_0 naturally takes this role, but it cannot be used in real-time applications required for kinetic calculations.

(iii) Plasmas with electric and magnetic charges show unusual transport properties: The Lorenz force enhances the collision rate and reduces viscosity (Liao and Shuryak, 2007). Quantum gluon-monopole scattering leads to a large transport cross section (Ratti and Shuryak, 2009), providing small viscosity in the range close to that observed at the RHIC and the LHC.

(iv) The high density of (noncondensed) monopoles near T_c leads to compression of the electric flux tubes, perhaps explaining curious lattice observations of very high tension in the potential energy (not free energy) of the heavy-quark potentials near T_c (Liao and Shuryak, 2007); see Sec. X.

(v) Last but not least, the peaking density of monopoles near T_c seems to be directly relevant to jet quenching; see Sec. XI.

Completing this introduction to monopole applications, it is impossible not to mention the remaining unresolved issues. Theories with adjoint scalar fields, such as, e.g., the celebrated $\mathcal{N} = 2$ Seiberg-Witten theory, naturally have particlelike monopole solutions. However, in QCD-like theories without scalars the exact structure of the lattice monopole are not yet well understood. There are indications that most, if not all, of the monopole physics can be taken care of via the instanton dyons previously mentioned: in this case the role of the adjoint "Higgs" is played by the time component of the gauge potential A_4. The dyon solution is well defined and real in Euclidean time, but becomes imaginary in the Minkowski continuation: so it is not a "particle" in the ordinary sense.

II. THE MAIN ISSUES IN QCD AND HEAVY ION PHYSICS

Let us start with a few super questions and comments on them, which are common to the entire strong interaction physics, extending well beyond the boundaries of the heavy ion field.

(I) Can one locate the "soft-to-hard" boundary, in whatever observables under consideration, where the transition from weak- to strong-coupling regimes take place?

(II) Can one locate the "micro-to- macro" boundary, where some transition in the value of the mean-free-path happens, from large (ballistic) to small (hydrodynamic) regime? In particular, which experimental observables best display this transition.

(III) Can we experimentally locate the QCD critical point, by following higher order fluctuations or correlations?

Some brief comments on these points are as follows:

(Ia) We already mentioned that a large variety of correlation functions has been studied on the lattice and phenomenologically, locating transitions between pQCD and nonperturbative regimes in various channels. The closest to it in experiments is hard exclusive processes, e.g., the pion and nucleon form factors. To the highest $Q^2 \sim 4$ GeV measured so far, neither of them had reached quantitative agreement with the pQCD predictions. Because of its importance, the experimental studies should be continued to higher momenta, until such agreement is observed.

In the heavy ion field there exists a hotly debated "minijet" issue. While the identified jets have rather large momenta, $p_\perp > 20$ GeV or so, it is generally assumed that the parton description is valid down to much smaller momenta. How much smaller? Following the Dokshitser-Gribov-Lipatov-Altarelli-Parisi (DGLAP) evolution toward small Q^2 all the way to ~ 1 GeV^2 one eventually reaches a negative gluon density. This and other arguments tell us that at this scale, 1 GeV, pQCD cannot be used. At which scale $Q_{\rm min}$ one has to stop is defined by the "higher twist effects," not yet studied to the extent to provide a quantitative answer.

(Ib) The fundamental process is pp scattering. Its total cross section and elastic amplitude is described by Pomeron phenomenology. The elastic amplitude is a function of the momentum transfer $t = -q^2$, and its Bessel-Fourier transform is the profile function $F(b)$, depending on the impact parameter b. Small b is understood via a perturbative Balitsky-Fadin-Kuraev-Lipatov (BFKL) Pomeron, while large b is understood via some string-exchange models. In this case the experimental data actually do indicate a sharp transition between these regimes. Attempts to understand both regimes in a single AdS-QCD framework have been successful (Stoffers and Zahed, 2013). Furthermore, it has been suggested that the critical b is related to critical temperature T_c of the phase transition in the gauge theory; we discuss this in Sec. VIII.C.

(Ic) Proceeding from elastic to inelastic collisions, when should we describe the initial snapshots of hadrons and nuclei in terms of partons (quarks and gluons) or nonperturbative effective objects (monopoles, strings)? As we will discuss, these initial effective objects produce fluctuations, which, via long-lived hydrosound modes, are visible to the detectors. Therefore, their number becomes experimentally observable.

(IIa) Heavy ion (AA) collisions are now complemented by "small systems" pA and pp collisions. At high enough multiplicity they display collective phenomena: radial, elliptic, and even triangular flows. One wants to quantify the regime change, if possible, experimentally and theoretically. Unfortunately, so far no sharp transitions as a function of multiplicity are detected.

(IIb) Where exactly is the boundary between the microtheories and macrotheories? The textbook answer is that one can compare the micro or "mean-free-path" scale l to the size of the system

$$L \gg l \tag{9}$$

and if the l/L ratio is small one can use the macroscopic theories. Small phenomenological viscosity suggests that the mean free path in sQGP is a few times smaller than the interparticle distance. By observing the smallest exploding systems, one checks if this is indeed the case.

III. SOUNDS ON TOP OF THE "LITTLE BANG"

A. Comments on hydrodynamics

As emphasized in the Introduction, hydrodynamical explosion of QGP is well documented; see, e.g., Heinz and Snellings (2013). The interest has now shifted from a description of the bulk of the data to special cases, with an emphasis on the limits of the hydrodynamical description.

When one thinks of ideal hydrodynamics plus viscous corrections, it is sometimes stated that the latter, the viscosity times the velocity gradients, should be smaller than the main terms. In fact it is not so: it is the terms second (and higher order) in gradients that are neglected and thus assumed small. Navier-Stokes hydrodynamics can successfully describe anisotropic flows. Studies of the "anisotropic hydrodynamics" as well as an exact solution of the Boltzmann equation in Gubser setting are discussed in Sec. IV.C.

Another direction is the so-called higher order hydrodynamics, attempting resummation of certain gradient terms (Lublinsky and Shuryak, 2009); see Sec. VII.E.

All formulations of improved hydrodynamics are supposed to shift their initiation to a somewhat earlier time or promise to treat somewhat smaller systems. Yet, while the out-of-equilibrium initial stage gets reduced, of course it can never be eliminated. The distinction between the initial and the equilibrated stages is a matter of definition: but physical outputs, e.g., the total amount of entropy produced, should ideally be independent of that. Unfortunately, in practice we are still far from this ideal scenario: studies of entropy generation at an initial stage is still in its infancy.

A few other issues remains open, related to the boundary of hydrodynamical description.

One is the boundary at high p_t. The region in which hydrodynamical predictions describe the data goes roughly up to $p_\perp \sim 3$ GeV. (Note that it includes more than 99.9% of all secondaries.) Collective flows decrease above this momentum, and one needs to understand why. High $p_\perp$ particles come from an edge of the fireball, where the magnitude of the flow is maximal. Using the saddle point method for the Cooper-Frye integral (Blaizot and Ollitrault, 1990), one can

see that the region from which such particles come shrinks, as $p_\perp$ grows. We will return to this point in connection to high-multiplicity pp collisions and radii extracted by Hanbury-Brown–Twiss (HBT) interferometry, see Sec. V.E. Eventually this region shrinks to a single hydrodynamic cell of the size of a mean-free path, and hydrodynamics can no longer be applied.

The viscosity effects lead to anisotropy of particle distributions induced by flow gradients (Teaney, 2003): these effects should be enhanced at the fireball's edge. Gradients add an extra power of $p_\perp$, and the deviation in the flows from ideal hydrodynamics should be negative. This picture qualitatively agrees with observations, but quantitative theory in the window $4 < p_\perp < 10$ GeV is still missing. Above it one finds a completely different, jet-dominated, physics, which is again under theoretical control.

Another issue is the "end of hydrodynamics" at the end of the collisions. Because different secondaries have very different cross sections, it has been argued that they have sequential freezeouts depending on their value (Hung and Shuryak, 1998). Traditionally, after chemical freezeout the practical models switch from hydrodynamics to hadronic cascade, which implements it in detail. But recent data, especially from ALICE, put such an approach into question. The particle yields are described by chemical equilibrium so well, up to light nuclei, that little space is left to inelastic rescatterings.

The hydrodynamical paradigm states that out of all individual properties of the secondary hadrons, only one, their masses, is important, as one translates from the distribution over collective flow velocities to the observed momenta. All one needs to know is that an object of mass m in a flow with velocity v has the momentum mv, plus thermal motion which also depends only on m. Let us check it by a direct comparison of the spectra for a pair of hadrons with the same mass, with otherwise completely different quantum numbers and cross sections, e.g., p and ϕ. They can hardly be more different, a nonstrange baryon versus a $\bar{s}s$ meson, so any "afterburner" code shows their late-time dynamics to be different. And yet, as the data shown in Fig. 6 demonstrate, their spectra are practically identical, up to $p_\perp \sim 4$ GeV$/c$.

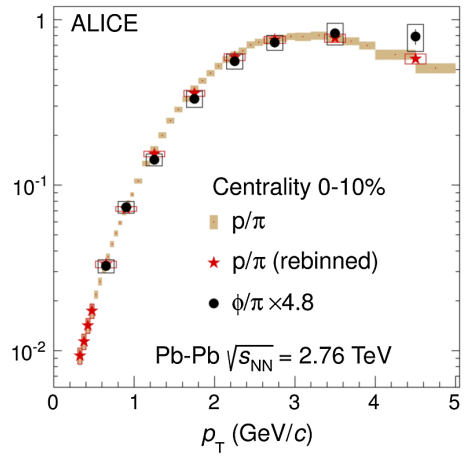

FIG. 6. Proton-to-pion and (rescaled) ϕ/π ratios, as a function of transverse momentum p_T.

B. Hydrodynamical response to perturbations

Now, assuming that the average pattern of the fireball explosion has been well established, we are going to add perturbations to it. The induced fluctuations and their correlations is thus the next topic of our discussion.

The first method is the so-called event-by-event hydrodynamics, solved for some ensemble of initial conditions. However, most of what was learned from these studies can also be understood from a simpler approach, in which one adds small and elementary perturbations on top of a smooth average fireball.[7] Using an analogy, instead of beating a drum forcefully with both hands and all fingers, one may touch it gently with a drumstick, at different locations, recording the spectra and intensities of the sounds produced. Eventually, summing up all the relevant modes of excitations, one gets a complete Green's function, from which a (linearized) description of any initial conditions follow.

In order to summarize what we learned from fluctuation and correlation studies, one needs to return to the data and to the results of multiple hydrodynamical calculations, separating their essence from unimportant complications. A simple pocket formula, revealing the systematics, will help us do so.

Before we get into the details, let us formulate the main points using the drum analogy. First, perturbations on top of the sQGP fireball basically are sounds, as those propagating on the drum. The main phenomenon is their viscous damping; the value of viscosity will be the main output. Unlike the drum, the fireball is not static but exploding: therefore an oscillating behavior is superimposed on the dynamical time dependence of the amplitudes. Different excitations are excited if the drum is struck at different places; similarly we find excitation of different modes depending on their origination point.

These calculations typically start from the angular deformations of the initial state. In Fig. 7(a) one finds the dependence of the mean harmonics (eccentricities)

$$\epsilon_n = \langle \cos(n\phi) \rangle, \tag{10}$$

where n is an integer and ϕ is the azimuthal angle. The angular brackets indicate an average over events, usually for a particular centrality bin (indicated in the upper left corner as a fraction of the total cross section, which scales as bdb). The bin 0%–0.2% is called the "ultracentral" one, $b \approx 0$, and 50%–60% are peripheral collisions. The first comment to this plot is that the $n = 2$ harmonics is special; it peaks for peripheral bins, due to collision geometry. However other harmonics, and in fact all of them for the central bins, are basically independent of n and centrality. What this tells us is that statistically independent "elementary perturbations" (or "bumps") have small angular size $\delta\phi \ll 2\pi$, so one sees here an angular Fourier transform of the delta function.

[7]By no means does the latter approach undermine good work in which the former approach does. Development of stable and causal second-order hydrodynamic codes and ensembles of initial conditions is a significant achievement. The averaging hydrodynamic results over thousands of configurations, with complicated shapes, is a lot of work, resulting in the v_n moments. Here one needs to focus on the essence of the issue, in the simplest setting possible.

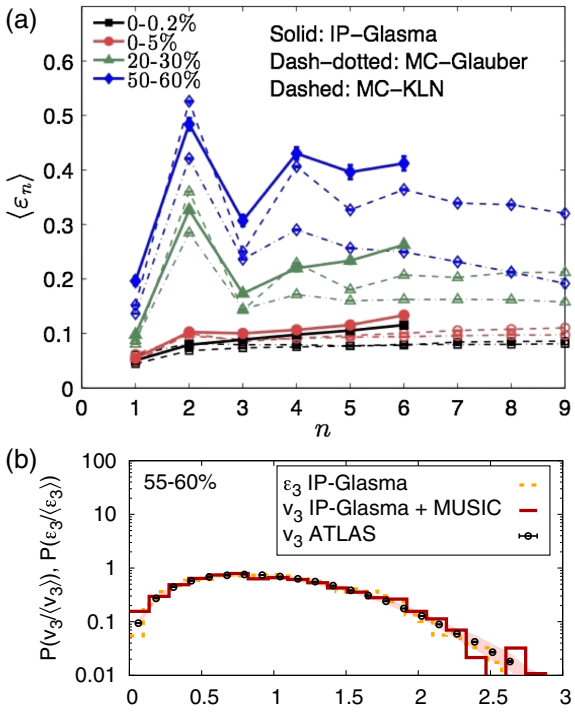

FIG. 7. (a) Average initial stage deformations $\langle\epsilon_n\rangle = \langle\cos(n\phi)\rangle$ for various centralities using the models indicated. From Heinz and Snellings, 2013. (b) Data points correspond to the event- by-event distribution of v_3 measured by the ATLAS Collaboration, compared to the distributions of initial eccentricities in the IP-GLASMA model and the distributions of v_3 from fluid dynamic evolution with IP-GLASMA initial conditions. From Schenke and Venugopalan, 2014.

The next observation is that the deformations are smaller for central collisions. This is also natural: central collisions produce larger fireballs, which have more particles and thus fluctuate less

$$\langle\epsilon_n\rangle \sim \frac{1}{\sqrt{N_{\rm cells}}}, \tag{11}$$

where $N_{\rm cells} = A_{\rm cell}/A_{\rm fireball}$ is the number of statistically independent cells.

Models of the initial state give not only the average deformation but also their distributions and correlations. Remarkably, the experimentally observed distributions over flows $P(v_n)$ directly reflect the distributions of the angular anisotropies $P(\epsilon_n)$ at the initial time; see, e.g., the ϵ_3 and v_3 distributions in Fig. 7(b). In other words, apparently no extra noise is generated during the hydrodynamic evolution, from the initial state ϵ_n to the final state v_n.

C. Acoustic systematics: The viscous damping

There is a qualitative difference between radial flow and higher angular harmonics. While the former monotonically grows with time, driven by the outward pressure gradient with a fixed sign, the latter are basically sounds, or density oscillations. Therefore the signal observed should, on general grounds, be the product of the two factors: (i) the amplitude reduction factor due to viscous damping and (ii) the phase factor containing the oscillation at the freezeout. (We will discuss the effects of the phase in the next section.)

Let us start with the "acoustic systematics" which includes only the viscous damping factor. Somewhat surprisingly, this simple expression describes both the data and the hydrodynamic calculations. More specifically, it reproduces the dependence on the viscosity value η, the size of the system R, and the harmonic number n.

The expression can be motivated as follows. We had already mentioned "naive" macroscales and microscales (9): now we define it a bit more accurately, by inserting the viscosity-to-entropy ratio $\eta/s = lT$ into it

$$\frac{l}{L} = \frac{\eta}{s}\frac{1}{LT}. \tag{12}$$

This "true micro-to-macroratio," corresponding to the mean-free path in kinetic theory, defines the minimal size of a hydrodynamic cell.

One effect of viscosity on sounds is the damping of their amplitudes. The acoustic damping formula (Staig and Shuryak, 2011a) is

$$\frac{v_n}{\epsilon_n} \sim \exp\left[-Cn^2\left(\frac{\eta}{s}\right)\left(\frac{1}{TR}\right)\right], \tag{13}$$

where C is some constant. The harmonic number n appears squared because the damping includes a square of the gradient, the momentum of the wave. It gives the following predictions: (i) the viscous damping is exponential in n^2, (ii) the exponent contains the product of two small factors η/s and $1/TR$, and (iii) the exponent contains $1/R$ which should be understood as the largest gradient in the system, often modeled[8] as $1/R = 1/R_x + 1/R_y$.

An extensive comparison of this expression with the AA data, from central to peripheral, was done by Lacey *et al.* (2013) from which we borrow Figs. 8 and 9. Figure. 8(a) shows the well-known centrality dependence of the elliptic and triangular flows. v_2 is small for central collisions due to the smallness of ϵ_2, and also small in the very peripheral bin because viscosity is large in small systems. Figure 8(b) shows the $\ln(v_n/\epsilon_n)$. As a function of the inverse system's size $1/R$, both elliptic and triangular flows show perfectly linear behavior. Further issues [the n^2 dependence as well as the linear dependences of the $\log(v_m/\epsilon_m)$ on the viscosity value] are also very well reproduced; see Fig. 9. Note that this expression works all the way to rather peripheral AA collisions with $R \sim 1$ fm and multiplicities comparable to those in the highest pA bins. It also seems to work to the largest n so far measured.

The acoustic damping provides correct systematics of the harmonic strength. This increases our confidence that, in spite of somewhat different geometry, the perturbations observed are actually just a form of sound waves.

[8]Remember that x and y axes are transverse to the beam, and x is along the impact parameter. Thus for peripheral collisions R_x is dominant in this combination.

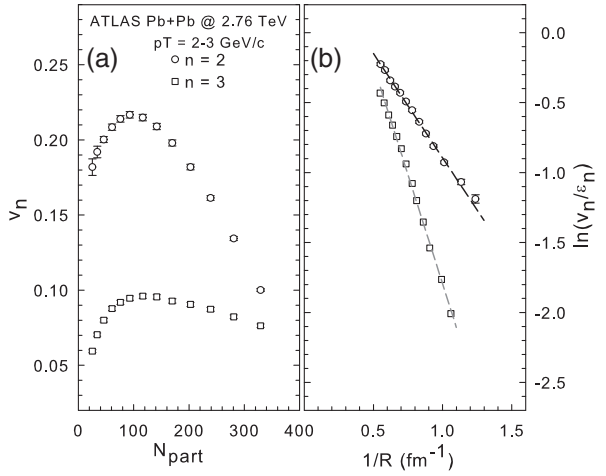

FIG. 8. (a) ATLAS data for v_2 and v_3 vs $N_{\rm part}$. (b) $\ln(v_n/\epsilon_n)$ vs $1/R$ for the same data. From Lacey *et al.*, 2013.

Since we are interested not only in large AA systems but also in new, pA and pp, much smaller fireballs, one may use the systematics to compare it with the new data. Using the acoustic damping formula, one can estimate how many flow harmonics can be observed in these cases. For central PbPb at LHC collisions with

$$\frac{1}{TR} = \mathcal{O}(1/10) \tag{14}$$

its product of η/s is $O(10^{-2})$. One can immediately see from this expression why harmonics up to $n = O(10)$ can be observed. Proceeding to smaller systems, by keeping a similar initial temperature $T_i \sim 400$ MeV $\sim 1/(0.5$ fm) but a smaller size R, results in a macro-to-microparameter that is no longer small, or $1/TR \sim 1$. Note that for a usual liquid or gas, with $\eta/s > 1$, there would not be any small parameter left, and one would have to conclude that hydrodynamics is inapplicable for such a small system. However, since the quark-gluon plasma is an exceptionally good liquid with a very small η/s,

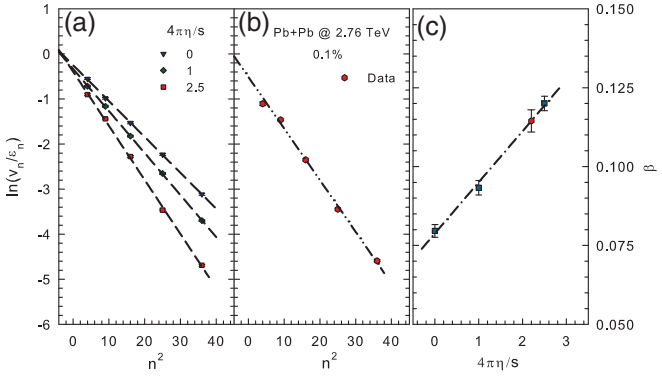

FIG. 9. (a) ATLAS data $\ln(v_n/\epsilon_n)$ vs n^2 from viscous hydrodynamical calculations for three values of specific shear viscosity as indicated. (b) $\ln(v_n/\epsilon_n)$ vs n^2 for Pb + Pb data. The $p_\perp$-integrated v_n results in (a) and (b) are from ATLAS 0.1% central Pb + Pb collisions at nucleon-nucleon $s_{NN} = 2.76$ TeV; the curves are linear fits. (c) Exponent vs viscosity-to-entropy ratio $4\pi/s$ for curves shown in (a) and (b). From Lacey *et al.*, 2013.

one can still observe flow harmonics up to $m = O(\sqrt{10}) \sim 3$. And indeed, v_2 and v_3 have been observed already in the first round of measurements (for later data, see Fig. 22).

D. Waves from a point perturbation and harmonic spectra

The event-by-event hydrodynamics appears to be a very complicated problem: events have multiple shapes, described by multidimensional probability distributions $P(\epsilon_2, \epsilon_3 \ldots)$, and different spatial shapes lead to complicated deformations of the secondary spectra. Yet the analysis shows that all those shapes are however mostly due to a statistical noise, and the problem is much simpler than it naively appears to be.

The point is that the individual rows of nucleons, located at different places in the transverse plane, by causality cannot possibly know about fluctuations of other rows at different locations; their fluctuations are statistically independent. Therefore it is sufficient to study one "elementary excitation" produced by a delta-function source in the transverse plane (in reality, of the size of a nucleon). In other words, one needs to find the Green's function of the linearized hydrodynamic equations.

A particular model of the initial state expressing locality and statistical independence of bumps has been formulated by Bhalerao and Ollitrault (2006): the correlator of fluctuations is given by the Poisson local expression

$$\langle \delta n(x) \delta n(y) \rangle = \bar{n}(x) \delta^2(x - y), \tag{15}$$

where $\bar{n}(x)$ is the average matter distribution. The immediate consequence of this model is that, for the central collisions (on which we focus now), ϵ_m are the same for all $m < m_{\rm max} = O(10)$ (until the bump size gets resolved).

In order to calculate perturbations at later times one needs to apply the Green's functions twice, describing perturbation propagating from the original source O to the observation points x and y as shown in Fig. 10(a). This was first done by Staig and Shuryak (2011b) analytically, for Gubser flow (see Appendix B for details). One finds that the main contributions come from two points in Fig. 10, where the "sound circle" intersects the fireball boundary. In single-body angular distributions those two points correspond to two excesses of particles at the corresponding two azimuthal directions. The angle between them at Fig. 10(a) is about $\Delta\phi \approx 120°$ or 2 rad. The azimuthal correlation function (Staig and Shuryak, 2011b) is shown in Fig. 11(a). One of its features is a peak at zero $\delta\phi = 0$: it is generated when both observed particles come from the same azimuthal enhancement. If two particles come from two different locations, there peaks displaced by $\Delta\phi = \pm 2$ rad. [As shown in Fig. 10, if one shifts the position of perturbation from (a) to (b), the peak angle $\Delta\phi$ changes toward its maximal value, πrad, or 1/2 of the circle.]

This calculation has been made and presented[9] before the experimental data were shown. The experimental correlation function from ATLAS, for the "supercentral bin" with the fraction of the total cross section 0%–1%, is shown in

[9]At the first day of the Quark Matter 2011 Conference at Annecy.

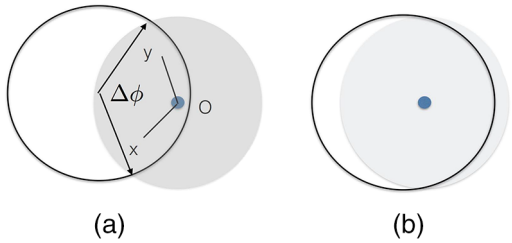

FIG. 10. The perturbation is shown by a small (blue) circle at point O: its time evolution to points x and y is described by the Green's function of linearized hydrodynamics shown by two lines. The perturbed region, shown by the gray circle, is inside the sound horizon. The sound wave effect is maximal at the intersection points of this area with the fireball boundary: $\Delta\phi$ angle is the value at which the peak in the two-body correlation function is to be found. Shifting the location of the perturbation, from (a) to (b), results in a rather small shift in $\Delta\phi$.

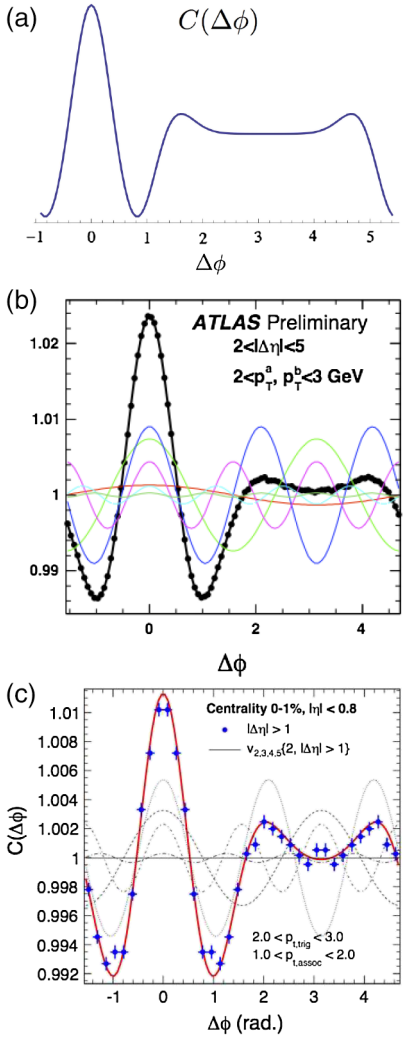

FIG. 11. (a) Calculated two-pion distribution as a function of the azimuthal angle difference $\Delta\phi$, for viscosity-to-entropy ratios $\eta/s = 0.134$. From Staig and Shuryak, 2011b. (b) From ATLAS, Jia, 2011. (c) From ALICE, Aamodt *et al.*, 2011b. All for ultracentral collisions.

Fig. 11(b). The shape of the correlator was predicted strikingly well.

While there is no need to use Fourier harmonics, we insist that the correlation function as a function of the relative angle may teach us more than harmonics, separately studied. Note that for the ultracentral collisions we now discuss the largest harmonics is v_3 [see Fig. 11(b)], not v_2. Since starting deformations ϵ_n are basically the same for all n, the difference must come from hydrodynamics, and it does. As previously explained using the notion of the sound horizon, the angular distance between enhancements is about 120°, not 180°.

A very similar phenomenon takes place for the big bang sound perturbations. All hydrodynamic harmonics get excited at the initial time $t = 0$ by the big bang, and all got frozen out at the same time as well. The acquired phases depend on the harmonic number n, because at larger n they oscillate more rapidly. The binary correlator is proportional to $\cos^2(\phi^n_{\text{freezout}})$ of these phases and harmonics with the optimal phases close to $\pi/2$ or $3\pi/2$ values show maxima, with minima in between. Planck Collaboration data on the power spectrum of the cosmic microwave big bang perturbations, shown in Fig. 12 as a function of the harmonic number, display a number of such maxima and minima.

The first calculation of the harmonic spectrum (Staig and Shuryak, 2011b) similarly showed such oscillations, with the first peak close to $n = 3$ and the minimum around $n = 7$; see Fig. 13(a). A subsequent study of analytic linearized perturbation on top of Gubser flow (Gorda and Romatschke, 2014) produced more information about the minimum; see Fig. 13(b). While the minima are clearly there for one pointlike source, its location depends on the radial location of the original source r_0. Gorda and Romatschke further studied the question and found that for $r_0 \sim 5-6$ fm the minimum shifts to $m = 10$. The questions whether such minima can survive realistic ensemble average, and perhaps ever be observed experimentally, remain open. To date, high harmonics $n > 6$ remains out of reach for statistical reasons.

Currently a number of groups developed sophisticated event-by-event hydrodynamic calculations and calculated the magnitude of harmonics: however they do so with

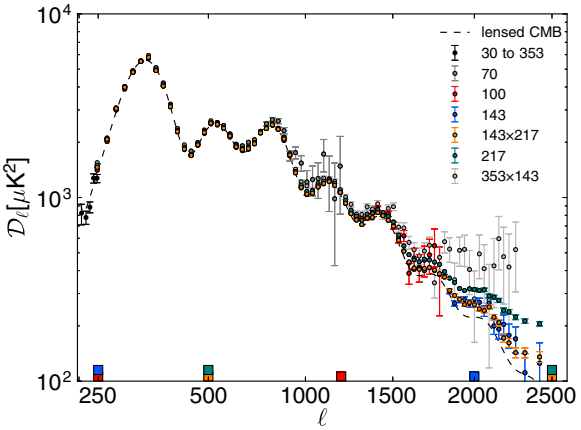

FIG. 12. A power spectrum of cosmic microwave background radiation measured by the Planck Collaboration. From Ade *et al.*, 2014.

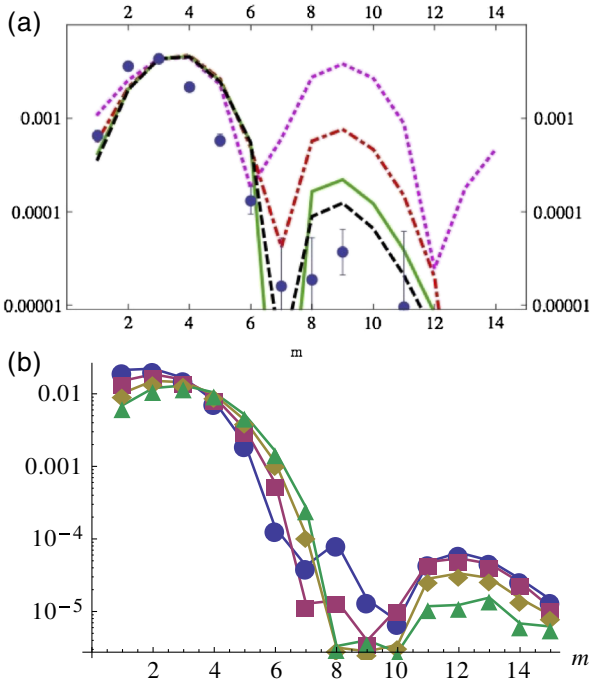

FIG. 13. (a) The lines are from analytic linearized hydrodynamic calculations of the correlation function harmonics v_m^2, based on a Green's function from a point source, for four values of viscosity $4\pi\eta/s = 0$, 1, 1.68, and 2 (top to bottom at the right). The closed circles are the ATLAS data for the ultracentral bin. From Staig and Shuryak, 2011b. (b) Calculation of harmonic flow spectra from analytic linearized hydrodynamics at $p_\perp = 1$ GeV demonstrates the dependence of the minimum on the location of the perturbation $r_0 = 7.5$, 8.0, 8.5, and 9 fm, by the circles (blue), squares (red), diamonds (brown), and triangles (green), respectively. From Gorda and Romatschke, 2014.

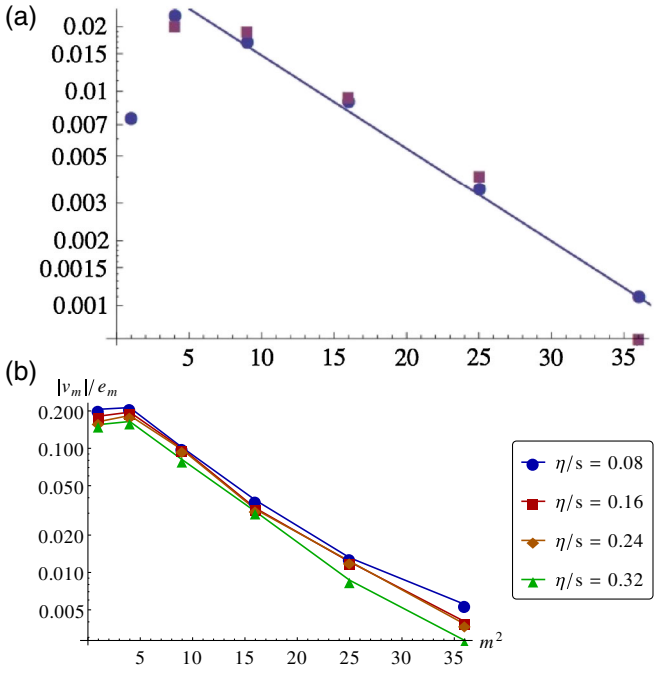

FIG. 14. (a) $v_n\{2\}$ vs n^2. The solid (blue) circles correspond to viscous event-by-event hydrodynamics, in the "IPGLASMA+Music" model, with viscosity value $\eta/s = 0.14$. The straight line, shown to guide the eye, demonstrates that acoustic systematics does in fact describe the results of this heavy calculation quite accurately. The CMS data for the 0%–1% centrality bin, shown by the squares (red), in fact display larger deviations, perhaps oscillatory ones. From Rose *et al.*, 2014. (b) The harmonics induced by a fixed perturbation at $r_0 = 4$ fm for variable viscosity. From Gorda and Romatschke, 2014.

$n = 2$–6 that is below the discussed minimum. It is interesting that their results are well described by $\exp(-n^2 \times \text{const})$ dependence, expected from the acoustic damping discussed previously; see, for example, Rose *et al.* (2014) and Fig. 14(a). Gorda and Romatschke also saw this dependence; see Fig. 14(b), but they do not agree that the coefficient is proportional to the viscosity.

Another question about these harmonic spectra is whether there is a maximum? This question reduces to the relation between amplitudes of the harmonics $m = 2$ and 3.[10] The experiments show that $v_3 > v_2$: see the ultracentral ATLAS data shown in Fig. 13(a), as well as the CMS central bin data in Fig. 13(b). (The latter include slightly larger impact parameters and thus feed more geometry-related contributions to v_2.) The same conclusion stems from both calculations just discussed (Staig and Shuryak, 2011b; Gorda and Romatschke, 2014) at large r_0. Surprisingly, all sophisticated event-by-event studies led to the opposite conclusion, namely, $v_3 < v_2$ for the ultracentral bin.

[10]The harmonics $m = 1$ is known to be especially small due to the vanishing dipole, and so it should be removed from consideration. We now discuss central collisions only, in which all effect comes from fluctuations, not the geometry.

The "flow harmonics," solutions of linearized equations on top of average smooth flow, should make a complete set of all possible perturbations. The functions of course depend not only on the azimuthal angle $\sim e^{im\phi}$, but on other coordinates r and η as well. For Gubser flow (see Appendix B) one can use the comoving coordinates ρ, θ, ϕ, η, derive linearized equations for perturbations, and separate the dependence on all four coordinates, with analytic expressions for all harmonics. The flow in the transverse plane r, ϕ is combined into standard angular harmonics $Y_l(\theta, \phi)$, combining the azimuthal angle ϕ and the radial coordinate r into θ. The waves in the rapidity direction η are simple plane waves.

Can one similarly define a complete set of independent harmonics for a generic non-Gubser setting? And, even more importantly, can those be observed? A step in this direction was made by Mazeliauskas and Teaney (2015) using "subheading harmonics" of flow, extracted from experimental data. Figure 15 indicates a difference between the leading and subleading triangular flows. Note that the latter gets a sign change along the radial direction, unlike the former one.

E. Detecting the interactions between harmonics

We already argued that the individual sources, or bumps as we call them, are uncorrelated, and so one should not pay attention to their interferences. The suggested picture is similar to a number of stones thrown into the pond: the

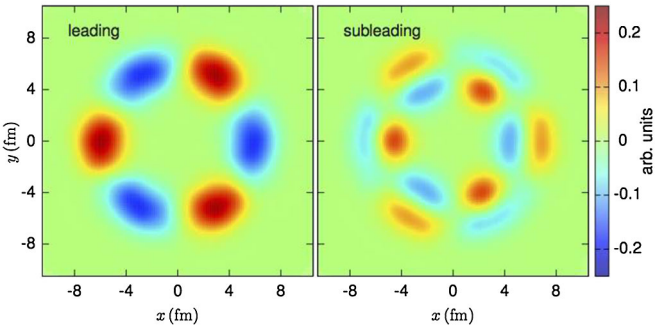

FIG. 15. The leading (left) and subleading (right) harmonics of triangular flow. From Mazeliauskas and Teaney, 2015.

produced expanding circles visibly interfere, but they do not really interact with each other since the amplitude is small.

Now we look at the details of this picture. First, how many such bumps are there? For central collisions the circumference of the fireball is $2\pi R_A \approx 40$ fm. The correlation length is perhaps the typical impact parameter in NN collisions, which at the LHC energies is $b \sim \sqrt{\sigma/\pi} \sim 1.6$ fm. As such there are not more than

$$N_{\text{sources}} \approx \frac{2\pi R_A}{b} \approx 25 \tag{16}$$

independent sources.

The absolute scale of correlator harmonics in experiments is of the order of a percent; see, e.g., Fig. 11(b). Of course, it comes from an incoherent sum over the number of sources, so one individual bump contributes only $O(10^{-3})$ into it. In the wave amplitude one needs to take the square root, since one deals with perturbation of the order of $1/30$.

With such small amplitudes one might conclude that the linear theory is completely sufficient, and all nonlinear effects can be ignored. It is basically correct, but not near the fireball edge. Small waves can produce large effects at the large p_t end. Indeed, at $p_t \sim 3$ GeV the elliptic flow gets large $v_n \sim 0.2$ which makes angular distribution 100% asymmetric. Similarly, the nonlinear interactions of flows at the edge, large p_t, are non-negligible.

For example, v_4 received a contribution proportional to ϵ_2^2, v_6 from ϵ_3^2, etc. Detailed studies of such effects can be found in Teaney and Yan (2012). Nonlinear effects include a particularly curious case: v_1 harmonics generated by nonlinear $\epsilon_2 \cdot \epsilon_3$ interaction

These nonlinear effects do not originate from the nonlinear terms in the hydrodynamic equation, but from an expansion of the Cooper-Frye exponent $\exp(p^\mu u_\mu/T)$, containing flow velocity, in powers of these perturbations. Obviously they become more important at high p_t. Furthermore, while the linear terms are also linear in p_t, the nonlinear effects mentioned are quadratic $\sim p_\perp^2$, etc. (Teaney and Yan, 2012).

F. Event-by-event v_n fluctuations and correlations

At the beginning of Sec. III we already emphasized that the main source of the v_n fluctuations is that of the original perturbations ϵ_n themselves; see, e.g., Fig. 7(b).

Now we return to the question: Why do the ratios v_n/ϵ_n, evaluated by hydro, have such a small spread? While the practitioners of the event-by-event hydrodynamics use a large variety of initial configurations, it turns out that just one parameter ϵ_n is sufficient to predict v_n. If there would be some spread in values, the distributions in v_n and ϵ_n would not match that well.

Even adopting the minimalistic model, that all perturbations come from incoherent pointlike sources, it is surprising that event-by-event fluctuations of strength and locations of the bumps do not create any additional spread. (Or, using the drum analogy, does one get the same sound when the drum is hit in a different radius?)

Trying to understand this dependence, let us return to Fig. 10. The source located at the fireball edge, Fig. 10(a), produces correlation at $\Delta\phi \approx 2$ rad. As emphasized previously, projected into harmonics, it will excite the $m = 3$ one, since 2 rad is about $1/3$ of 2π. As the source moves inward, Fig. 10(b), the overlap of the sound circle and the fireball edge moves to $t\,\Delta\phi \sim \pi$, and the leading excitation becomes elliptic $m = 2$. The calculations show that in the latter case the correlation gets much weaker. The observed shape of the correlator for ultracentral collisions does have a minimum at $\Delta\phi = \pi$. Finally, as the source moves farther toward the fireball center (not shown in Fig. 10), the correlation appears at all angles equally, and its contributions to $m \neq 0$ harmonics vanishes. In summary, the harmonics we see comes mostly from the sources located near the boundary of the fireball. The angular correlations they induce have one universal form.

The study of flow harmonics and their correlations is a rapidly expanding field. Correlations can be divided into those sensitive to relative phases of the harmonics, and those which are not. An example of the latter is

$$SC(m,n) = \langle v_m^2 v_n^2 \rangle - \langle v_m^2 \rangle \langle v_n^2 \rangle. \tag{17}$$

ALICE provided data for $SC(4,2)$ and $SC(3,2)$, observing that $SC(4,2) > 0$ but $SC(3,2) < 0$: both qualitatively reproduced by hydrodynamical models.

This development requires the initial state model which is good enough not only to predict the mean ϵ_n, but their fluctuations and even respective correlations. Suppose we do so using the Bhalerao-Ollitrault relation (15): the results depend on integrals such as $\int d^2 r r^P \bar{n}(r)$ with large powers $P = 6, 8$. Thus the correlations are again coming from the very surface of the fireball at the initial time. Their absolute magnitude suffers from significant uncertainties, but the angular shapes and normalized correlators of harmonics can still be under control.

G. The map of the sounds

The number of gauge field harmonics needed to describe the initial state is rather large, counted in hundreds, much exceeding the number of sound modes we detect at freezeout. In Fig. 16 we show a map of those, in terms of momentum (rather than angular momentum). The curved line, corresponding to the acoustic systematics discussed previously, shows their lifetime. This curve crosses the freezeout time: smaller k waves can be observed at freezeout. Larger k cannot: they are

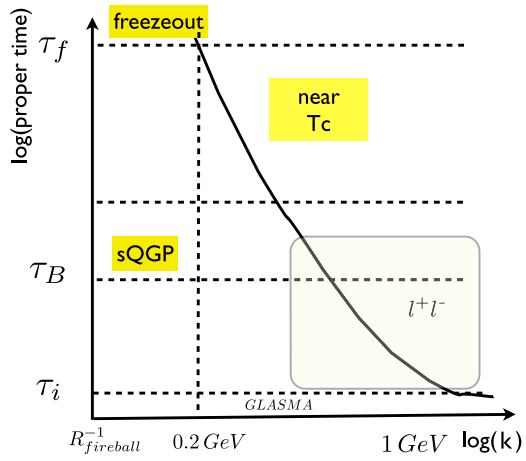

FIG. 16. The log-log plane proper time τ, sound momentum k. The solid curve indicates the amplitude damping by a factor of e: only small-k sounds thus survive until freezeout. The shadowed region on the right corresponds to that in which the magneto-sonoluminescence effect may produce extra dileptons.

weakened due to viscous damping. [A suggestion to detect those via the magnetosonoluminescence (MSL) process is discussed in Sec. IX.B.]

The fluctuation-dissipation theorem tells us that while the initial perturbations are damped, new ones should be produced instead. There should therefore be some noise, producing sounds continuously during the whole hydrodynamical evolution. Studies of the resulting "hydrodynamics with noise" have been demonstrated by Young *et al.* (2015). Unfortunately, they have not separated the "initial time sounds" from the "late-time ones." Perhaps that can be experimentally separated via studies of the azimuth + rapidity correlations, exploring the fact that the former lead to rapidity-independent modes, and the latter are rapidity localized.

Shuryak and Staig (2013b) suggested that late-time fluctuations may be generated by collapsing QGP clusters inside the hadronic phase. Those collapse events should happen because the QGP phase gets unstable in the bulk, once the temperature cools to $T < T_c$. This phenomenon is similar to the celebrated bubble collapse studied by Rayleigh.

Shocks and sounds are also expected to be created by jets depositing their energy into the ambient matter. How those are propagated via sounds was worked out by Shuryak and Staig (2013a).

H. Sounds in the loops

The hydrodynamical longitudinal pressure waves, the sounds, are the best quasiparticles we have. They are Goldstone modes, related to the spontaneous breaking of the translation invariance by matter, and thus their interaction follows a certain pattern familiar to pion physics. For large wavelengths they have a long lifetime, exceeding the freeze-out time. Therefore, in both the little and the big bangs, one can observe "frozen" traces of the initial perturbations, provided one looks at large enough wavelengths.

Because the sounds have long lifetimes and travel far, one may ask how an ensemble of sounds would behave, given such long times. In other fields of physics a theory developed for this questions is called the "acoustic turbulence." Furthermore, one may add to the hydrodynamical equations a Langevin-type noise term, with some Gaussian distribution, and formulate the resulting theory as a path integral, in the QFT-like form. Progress in this direction was summarized in a review by Kovtun (2012). Discussion of formal issues cannot be made in this review, however, and thus I illustrate the physics involved by one example only, also from Kovtun, Moore, and Romatschke (2011).

Recall that matter viscosity can be defined via a certain limit of the stress tensor correlator, known as the Kubo formula. Kovtun *et al.* calculated "loop corrections" to this correlator induced by the equilibrium sounds. Technically the calculation is done as follows: in the $\langle T^{\mu\nu}T^{\mu'\nu'}\rangle$ correlator one substitutes a hydrodynamical expression for the stress tensor containing sound perturbation velocities and makes it into a loop diagram with the "sound propagators"

$$\Delta^{mn} = \int d^4x e^{-ip_\alpha x^\alpha}\langle u^m(x)u^n(0)\rangle \tag{18}$$

for two pairs of the velocities. (We use Latin indices indicating that they are only spacelike here. For shear viscosity those used are $m = x$, $n = y$.) Skipping the derivation to the answer obtained from this calculation, which can be put into the form of a loop correction to the viscosity

$$\delta\eta_{\rm loop} = \frac{17}{120\pi^2}\frac{p_{\rm max}T(\epsilon+p)}{\eta_0}, \tag{19}$$

which is UV divergent and thus includes $p_{\rm max}$, the largest momentum which still makes sense for sound. What is important here is that the zeroth-order viscosity enters into the denominator. This should not be surprising: a good liquid with small η_0 supports very long-lived sounds, which can transfer momentum at relatively large distances, which means they produce a large contribution[11] to the effective viscosity.

The same correlator of stress tensors, in the "sound approximation" similar to that previously used, has been used (Kalaydzhyan and Shuryak, 2015) to calculate the on-shell rate of sound + sound → gravity waves, from sounds generated by QCD and/or electroweak cosmological phase transition.

Completing this section we remind the interested reader about the existence of other hydrodynamical modes, the rotational[12] ones. Rotational modes on top of the Bjorken flow were discussed by Floerchinger and Wiedemann (2011): under certain conditions one hydrodynamic mode does become unstable. Other unstable hydrodynamical modes appear for noncentral collisions with rotation (Csernai, Becattini, and Wang, 2014). More studies of these instabilities

[11]Other examples of the most penetrating modes dominating transport are ballistically moving phonons in liquid helium or neutrinos in a supernova.

[12]It is well known that those are central, e.g., for the atmospheric turbulence.

are however necessary to see whether such instabilities can indeed be observed.

IV. THE PREEQUILIBRIUM STATE, GLOBAL OBSERVABLES, AND FLUCTUATIONS

A. Perturbative versus nonperturbative models

A theory of the early stage should be able to (i) specify certain wave functions of the colliding particle, in a wide rapidity range; (ii) explain what happens during the collision time; and (iii) explain how the produced state evolves into the final observed hadronic state.

It is perhaps fair to say that approaches based on the weak coupling (pQCD) have been able to explain (i) and (ii), but not (iii). Strong-coupling approaches, AdS/CFT (conformal field theory) especially, explain (iii) but not the first two.

More specifically, the pQCD regime is natural for hard processes, for which the QCD running coupling is weak. Already in the 1970s pQCD developed a factorization framework, which divided production amplitudes into "before," "during," and "after" parts. The before and after parts are treated empirically, by structure (or distribution) and fragmentation functions. The during, near-instantaneous, part is described by the explicit partonic processes under consideration.

The strength of this approach is based on the separation of hard and soft scales, by some normalization scale μ, on which the final answer should not depend. The dependence of the parton distribution functions (PDFs) and fragmentation functions on μ is described by the renormalization group tool, the DGLAP evolution. Using it, one can tune the hard scale Q to pertinent kinematics. It works well for truly hard processes such as jet production at $Q > 10$ GeV. In the minijets domain, at Q given by a few GeV, higher twist $1/Q^n$ corrections are large and not yet under theoretical control.

The described pQCD framework has also serious restrictions as well. The factorized PDFs by definition describe the average nucleon (or nucleus). As soon as a particle is touched, e.g., the impact parameter (multiplicity bin) is selected, factorization theorems are no longer applicable. The absence of good practical models describing partonic states with fluctuations remains a problem: e.g., for understanding pp collisions with multiplicity several times the average. As we discuss in detail, pQCD can hardly be used for assessing the transverse plane distributions or correlations of partons.

For "baseline" soft processes, minimally biased pp, pA collisions with low multiplicity, the phenomenological models describing QCD string production and fragmentation are rather successful. The Lund model has branched into various "event generators," such as PYTHIA, popular among experimentalists. Their key feature is independent string fragmentation. However, new experiments focused on high-multiplicity events found correlation phenomena clearly going beyond the reach of these event generators.

In the case of very high multiplicity, e.g., central pA, AA, the initial conditions for hydrodynamics are smooth and given by the nuclear shape. The main parameter one needs to know about the preequilibrium stage is the total amount of the entropy generated. So far this is treated with some empirical coefficients (entropy per parton) not yet derived.

In the next approximation one accounts for quantum fluctuations, in the positions of the nucleons as well as in the cross sections, via versions of the Glauber eikonal models. They provide well-defined and reasonable predictions for initial state perturbations ϵ_n, generating flow harmonics via hydrodynamics. The GLASMA-based models include more fluctuations, resolving partonic substructure of the nucleons: its relevance for the results is at the moment unclear.

Partonic description of the initial state of the collision at asymptotically high parton density evolved into the so-called color-glass-condensate (CGC)-GLASMA paradigm (McLerran and Venugopalan, 1994). Since at any transverse location the number of colored objects involved can be considered large, the color charge fluctuation should also become large, producing strong gauge fields, the CGC. When the gluonic fields become so strong that the occupation numbers reach $O(1/\alpha_s)$, the nonlinear commutator term in non-Abelian gauge fields is as large as linear ones, and so one should use classical nonlinear Yang-Mills equations. GLASMA is a state made of such random classical fields, starting from CGC at the collision time and then evolving as the system expands, until the occupation numbers reduce to $O(1)$. The model remains valid provided the scale of the (two-dimensional) parton density $n \sim Q_s^2$, known as a saturation momentum, remains large compared to the nonperturbative QCD scale. At early time the charges in each "GLASMA cell" separate longitudinally, producing longitudinal electric and magnetic fields. Cells of area $\sim 1/Q_s^2$ are statistically independent and fluctuate with their own Poisson-like distributions. The explicit modeling of the resulting field, from cells in the transverse plane, is now known as an impact-parameter (IP) GLASMA model.

High-multiplicity initial states then evolve into sQGP, which undergo hydrodynamical expansion. Reducing the multiplicity, one expects to find a regime in which the system is too small to have the hydrodynamical stage. How would one see that?

Let us illustrate it by one particular observable, the elliptic flow v_2. Suppose there is no sQGP stage: partons, gluons and quarks, simply become minijets after the collision, more or less like the Weizsacker-Williams photons do in QED. Correlations in the collision, related to the impact parameter plane, will produce certain v_n moments. Let us discuss their $p_\perp$ dependence.

Hard partons at large momentum scale $Q > Q_s$, exceeding the saturation scale of GLASMA, cannot possibly know about other cells and geometry: those would be produced isotropically in the transverse plane and do not contribute to v_n. If they reinteract later, the resulting showers contribute a negative correction to v_2. Softer partons, with momenta $Q \sim 1/R \ll Q_s$, will know about the "overlap almond" shape of the initial state: their distribution will be anisotropic, perhaps even with v_2 of the order of several percent as observed. Thus the prediction would be basically flat $v_2(p_\perp)$ below Q_s and decreasing as $p_\perp > Q_s$, perhaps to negative values. Unfortunately, in practice implementation of these ideas is difficult. The Q_s value for small systems in question is in the range of 2.5–3 GeV, close to the maxima produced be the flow. Experiments show v_n extending to higher $p_\perp \sim 5$ GeV, confusing the situation.

Strong-coupling models of the initial stage and equilibrated matter fall into two categories. One is classical strongly coupled plasmas (Gelman, Shuryak, and Zahed, 2006a). Its main feature is that the so- called plasma parameter is large

$$\Gamma = \frac{V_{\text{interaction}}}{T} > 1 \tag{20}$$

and therefore the potential energy of a particle exceeds its kinetic energy. Simulations and experiments with QED strongly coupled plasmas show that for relevant $\Gamma = 1$–10 one deals with strongly correlated liquids. Screening in this regime was studied by Gelman, Shuryak, and Zahed (2006b) and viscosity and diffusion constants by Liao and Shuryak (2007). The second, much wider known, strong-coupling framework is based on holography and AdS/CFT correspondence.

B. Centrality, $E_\perp$, and fluctuations

Let us briefly remind one of some basic facts about the global observables and their fluctuations. One of the first practical questions for AA collisions is the determination of centrality classes, related to observables such as the number of participant nucleons N_p, correlated to total multiplicity N or transverse energy $E_\perp$. The N_p is defined via measurements of near-forward moving neutrons, by forward-backward calorimeters, complementing observables determined from the central midrapidity detectors. Correlation plots between all of them and precise cuts defining the centrality classes are defined by each collaboration in their technical reports.

Historically, the ratio of the $E_\perp$ rapidity distributions for AA and pp collisions were fitted by a parametrization

$$\frac{dE_\perp^{AA}}{d\eta}\Big/\frac{dE_\perp^{pp}}{d\eta} = (1-x)\frac{N_p}{2} + xN_{\text{coll}} \tag{21}$$

with a parameter x interpreted as an admixture of the "binary collisions" N_{coll} to the main soft term, proportional to the number of participants. However, we now know that a hard interpretation of this last term is questionable: the hard powerlike component of the particle spectra is actually orders of magnitude smaller than would be needed for it.

One possibility can be that such interpretation is correct at an early time, yet with subsequent equilibration jets are quenched and disappear from spectra; however, an extra entropy generated by them may still survive and contribute to the total multiplicity.

Another interesting interpretation for the multiplicity and $E_\perp$ distributions was provided by Tannenbaum (2014). The notion of a participant nucleon is substituted by the "participant quark" N_{qp}. The model, an incarnation of the additive quark model of the 1960s, views a nucleon as a set of three constituent quarks, which interact separately. Defining the number of "quark participants" N_{qp} he showed that, within a 1% accuracy, it is proportional to the rhs of Eq. (21). In particular, the $E_\perp$ is perfectly linear in N_{qp}; see Fig. 17(b). If so, each participant quark is connected by the QCD string to the other one, and those strings are the "clusters" or ancestors for the observed secondaries. We return to the "wounded quarks" concept in the discussion of the Pomeron in Sec. VIII.C.

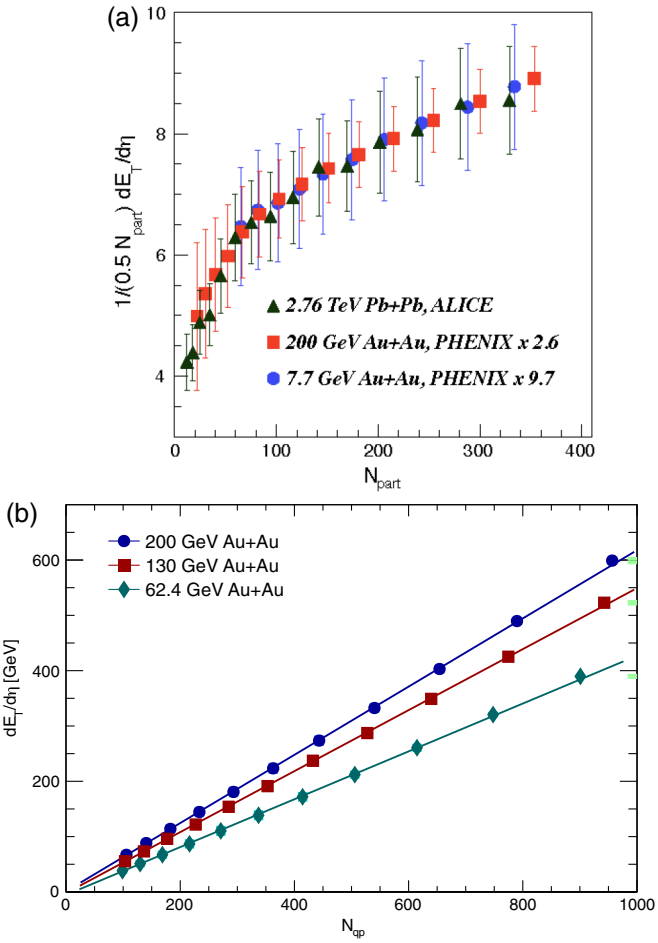

FIG. 17. Distributions over participant (a) nucleons and participant (b) quarks. From Adler *et al.*, 2014 and Tannenbaum, 2014.

On the other hand, the additive quark model does not agree with the CGC-GLASMA picture, in which cells or flux tubes[13] of the size $1/Q_s \times 1/Q_s$ are the statistically independent sources producing the secondaries. Presumably the CGC-GLASMA picture should be valid at a high density regime, while simpler Lund-type models with QCD strings (and, perhaps, constituent quarks at their ends) are true in the low-density regime. The problem is in the data we do not experimentally detect any sharp transition between the two regimes.

Let us seek further guidance from phenomenology. Note that if the number of independently decaying clusters is N, the width of the observed distributions should scale as $O(N^{-1/2})$. The multiplicity distributions have long tails toward large values, which are usually fitted by the negative binomial or similar distributions with two parameters, or some similar convolutions of two random processes with different parameters. Its second moment should tell us how many

[13]These McLerran-Venugopalan "flux tubes" should not be confused with the QCD strings: the former exist in a dense deconfined phase, are classical and not quantized, have an arbitrary fluctuating field strength, and thus do not have a universal tension.

"progenitors" (clusters, clans, ancestors, cells) the system goes through.

Let us discuss three models: (i) the usual Glauber in which N is the number of participant nucleons N_p; (ii) the Tannenbaum modification, based on the number of participant constituent quarks N_{pq}; and (iii) the CGC-GLASMA, and calculate the fluctuations.

In the last case the number of cells is

$$\mathrm{N_{GLASMA}} \sim (\pi R^2) Q_s^2. \tag{22}$$

For central AA the area is geometrical area $= 100$ fm^2 and

$$N_p \sim 400, \qquad N_{pq} \sim 1000, \qquad N_{\mathrm{GLASMA}} \sim 10^4. \tag{23}$$

For central pA the area is given by the NN cross section $\sigma \sim 100$ mb $= 10$ fm^2. As such one gets a very different number of clusters

$$N_p \sim 16, \qquad N_{pq} \sim 40, \qquad N_{\mathrm{GLASMA}} \sim 10^3. \tag{24}$$

Therefore these models predict vastly different fluctuations.

A brief summary of such a comparison with data is as follows. The participant quark model describes AuAu and dAu data extremely well, while for pp it underpredicts the tail of the distribution. Even six participant quarks, the maximal of the model, is not enough; there seems to be more clusters than that. The lesson is perhaps that the highest multiplicity pp is indeed the first case when soft models become insufficient. The models which have pQCD gluons in the wave function and hard scatterings are doing better on the "tails." In particular, PYTHIA (pQCD + strings) describes the high-multiplicity tail of pp reasonably well.

An alternative approach to initial state fluctuations is provided by the angular deformations ϵ_n. We argued that those are created by a number of statistically independent small-size sources (or bumps). For simplicity, as before, let us focus on central collisions only, where ϵ_2 is fluctuation induced and of the same magnitude as all other harmonics. We return to these in Sec. VI.C.

C. Anisotropy and the boundaries of hydrodynamics

The partonic initial state leads to the initial out-of-equilibrium stage of the collision which is highly anisotropic in momentum distribution. However, during the collisions, partons are naturally separated in time according to different rapidities and create "floating matter," in which a spread of longitudinal momenta is smaller than the transverse one. At later hydrodynamical stage, the viscosity effects reduce such anisotropy. Knowledge of the viscosity value allows one to calculate the anisotropy at this stage, provided its initial value is known.

What happens in between is still a matter of debate. Weak-coupling approaches, partonic cascades, predict anisotropy to be rising to large values, while the strong-coupling (holographic) approaches lead to rapid convergence to small values, consistent with hydrodynamics. For a discussion, see, e.g., Martinez and Strickland (2010) and subsequent papers.

The issue of anisotropy has two practical aspects. The experimental one, to which we return in Sec. IX.C, is a question of how one can experimentally monitor the anisotropy of matter at various stages of the evolution. The theoretical question is whether one can extend the hydrodynamical description for strongly anisotropic matter. Recently there was significant development along the lines of the so-called anisotropic hydrodynamics (aHydro). The idea (Martinez and Strickland, 2010; Florkowski and Ryblewski, 2011) is to introduce the asymmetry parameter into the particle distribution, and then, from the Boltzmann equation, derive an equation of motion for it, complementing the equations of the hydrodynamics. Solutions of various versions of hydrodynamics were compared to the exact solution of the Boltzmann equation itself, derived for the Gubser geometrical setting by Denicol *et al.* (2014). This paper contains many instructive plots, from the normalized temperature shown in Fig. 18 and the sheer stress Π^{ξ}_{ξ} shown in Fig. 19. In both cases the pairs of points correspond to small and very high viscosity values, separated by 2 orders of magnitude and roughly representatives of the strongly and weakly coupled regimes.

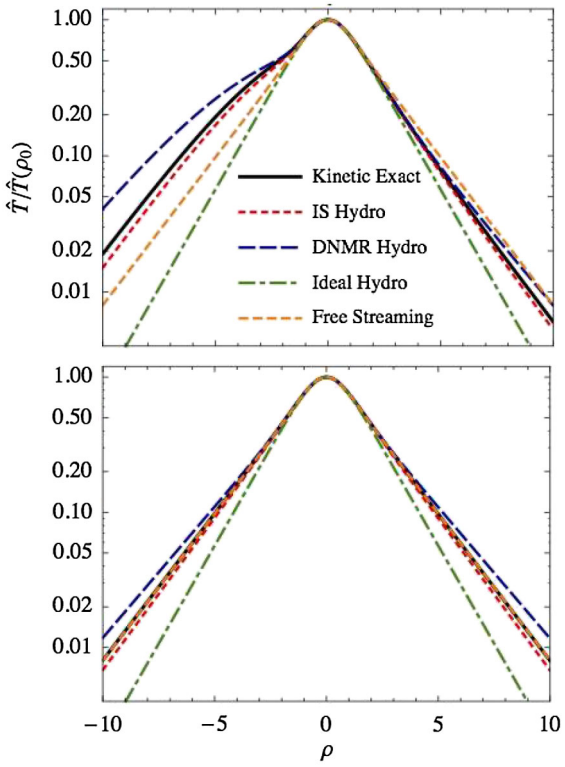

FIG. 18. The normalized temperature for $4\pi\eta/s = 1$, 100, upper, and lower plots, respectively. The meaning of the different curves is explained in the upper plot. From Denicol *et al.*, 2014.

Gubser's variable ρ is the "time" coordinate. In all four plots one can see that all curves coincide in the interval $-2 < \rho < 2$, but deviate from each other both at large negative values, corresponding to the very early stages, and at large positive ones, corresponding to very late times. In fact, all practical applications of hydrodynamics were indeed made inside this interval of ρ, with other regions being "before formation" and "after freezeout."

Solutions for two, very different, viscosities show similar trends. Israel-Stuart hydrodynamics seems to follow the

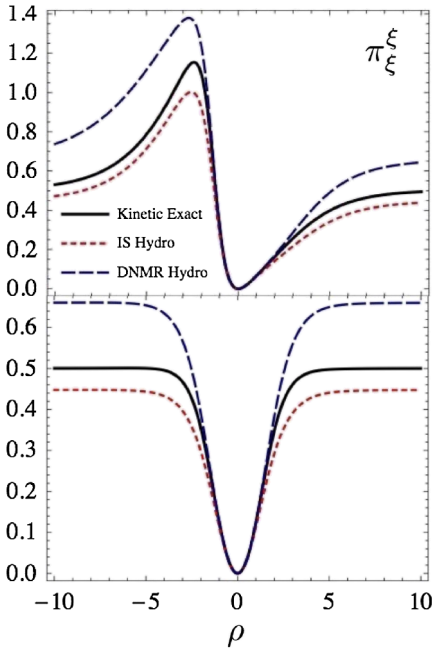

FIG. 19. The shear stress π^{ξ}_{ξ} vs coordinate ρ, for viscosity $4\pi\eta/s = 1$, 100, the upper, and the lower plots, respectively. From Denicol *et al.*, 2014.

solution of the Boltzmann equation quite well. Even the free streaming regime is not very far from all hydros and exact Boltzmann equations: this would be surprising if we did not already know that the radial flow, unlike higher harmonics, can indeed be "faked." If they would calculate the elliptic and higher flows, the results would be quite different. One would expect that for $4\pi\eta/s = 100$ all flow harmonics would be completely obliterated. The plots for the shear stress show different behaviors for small and large viscosities, but again all curves coincide inside the "hydro window" of $-2 < \rho < 2$. Even going well outside that domain, we never see discrepancies between them by more than 20%.

The overall conclusion one can draw is quite simple: all versions of hydrodynamics used in practice are accurate enough for realistic viscosities $4\pi\eta/s \approx 2$ and the times the hydrodynamics is actually applied.

V. THE SMALLEST DROPS OF QGP

We emphasized a certain gap that still exists between weak and strong coupling estimates of the equilibration time and viscosity. Such issues should play an enhanced role in experiments with systems smaller than *AA* collisions and should clarify the limits of hydrodynamics.

Let us start this discussion with another look at the spatial scale corresponding to the shortest sound wavelength, for the highest n of the v_n observed. Azimuthal harmonics are waves propagating along the fireball surface. Therefore, a successful description of the nth harmonics implies that hydrodynamics is still applicable at a wavelength scale $2\pi R/n$. Taking the nuclear radius as $R \sim 6$ fm and the largest harmonics so far observed $n = 6$, one finds that the scale under consideration is larger than $O(1\text{ fm})$. However, it is still not small enough to resolve the nucleon substructure. That is why there is little difference between the initial states of the Glauber model (described in terms of nucleons) from those generated by GLASMA models (operating on a parton level).

Why do we not see harmonics with larger $n > 6$? Higher harmonics suffer stronger viscous damping. The limitation is due to current statistical limitations of the data sample and is thus unrelated to the limits of hydrodynamics.

In principle, one could have studied *AA* collisions with smaller and smaller A, decreasing the system size smoothly, while preserving the overall geometry. However, historically the development was not that smooth: an unexpected discovery of the so-called "ridge" at the LHC in very small systems, pp with a high-multiplicity trigger, provided a look at the opposite extreme case.

Before we go into detail, let us see how small these systems really are. At freezeout the size can be directly measured, using the so-called *femtoscopy* method (see Appendix A for some explanations). The corresponding data are shown in Fig. 20, which combines the traditional two-pion with more novel three-pion correlation functions of identical pions.

An overall growth of the freezeout size with multiplicity, roughly as $\langle N_{\rm ch}\rangle^{1/3}$, is expected if the freezeout density is a universal constant. While for *AA* collisions this simple idea indeed works, the pp, pA data apparently form a different line, with significantly smaller radii. Apparently those two systems get frozen at higher density, compared to *AA*, but why? To understand that recall the following freezeout

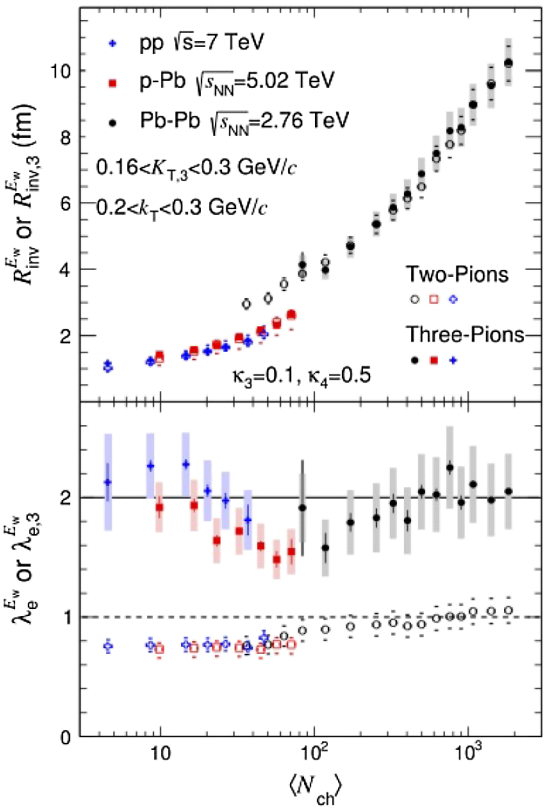

FIG. 20. (Upper) Alice data on the femtoscopy radii and (lower) "coherence parameter" as a function of multiplicity, for pp, pPb, PbPb collisions. From Grosse-Oetringhaus, 2014.

condition: "the collision rate becomes comparable to the expansion rate"

$$\langle n\sigma v\rangle = \tau_{\rm coll}^{-1}(n) \sim \tau_{\rm expansion}^{-1} = \frac{dn(\tau)/d\tau}{n(\tau)}. \qquad (25)$$

Higher density means a larger left-hand side (lhs), and thus a larger rhs. So, pp, pA high-multiplicity systems are more "explosive," with a larger expansion rate. We indeed argue that this conclusion is also confirmed by radial flow measured by particle spectra, as well as the HBT radii.

But why are those systems more explosive? Where is the room for that, given that even the measured final size of these objects is smaller than in peripheral AA (which shows only a rather modest radial flow). The only space left is at the beginning: those systems must obviously start accelerating earlier, from an even smaller size than seen by femtoscopy, to produce a strong collective flow at freezeout.

Another puzzle is why central pA (a collision of a proton with about 16 others) appears to be so similar in size and collective flows to (same multiplicity) pp, a collision of only two protons. We turn to its discussion in Sec. V.C.

A. Collectivity in small systems

The first discovery, in the very first LHC run, was due to the CMS Collaboration (Khachatryan *et al.*, 2010) which found a ridge correlation in high-multiplicity pp events. A special trigger was required because unfortunately the effect was first seen only in events with a probability $P \sim 10^{-6}$.[14] Switching to central pA collisions, the CMS observed a similar ridge there, now with much higher probability, a few percent instead of $P \sim 10^{-6}$ (Chatrchyan *et al.*, 2013). By subtracting central minus peripheral correlations, ATLAS CMS and ALICE groups soon all found that the observed ridge is accompanied by the "antiridge" in the other hemisphere, concluding the phenomenon is a familiar elliptic flow $v_2\cos(2\phi)$.

The PHENIX Collaboration at RHIC also found a ridgelike correction in central dAu collisions, with the v_2 value about twice larger than in pPb at the LHC. This difference was soon attributed to different initial conditions, for d and p beams, since the former produces a "double explosion" by its two nucleons. Quantitative hydrodynamical studies, such as Bozek (2012), confirmed this simple idea. It later was additionally confirmed by collisions of He^3Au as well.

The set of data, which established collectivity of the flow in pA beyond a reasonable doubt, came from CMS (de Cassagnac *et al.*, 2014). Their v_2 measurements from 2, 4, 6, and 8 particle correlators are shown in Fig. 21.

Taking collectivity for granted, one can further ask if the v_2 observed is caused by the precollision correlations or by the after-collision collective flows. A nice control experiment testing this is provided by dA and ^{3}HeA collisions. Two

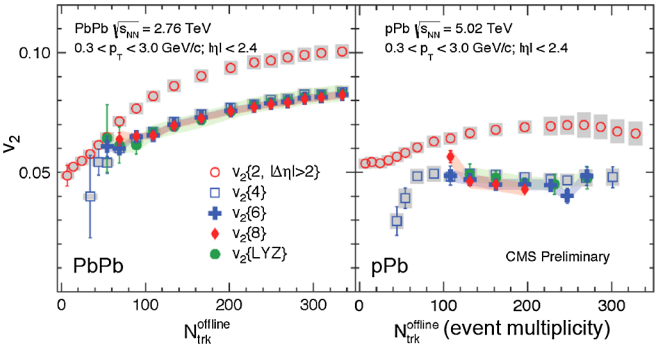

FIG. 21. CMS data for v_2 calculated using 2, 4, 6, and 8 particle correlations, as well as Lee-Yang zeros (basically all particles). Good agreement between those manifest collectivity of the phenomenon. From de Cassagnac *et al.*, 2014.

nucleons in d are on average far from each other and 2 MeV binding is so small that one surely can ignore their initial state correlations. As a result, whatever the "initial shape" effect in pp, in dA it should be reduced by $1/\sqrt{2}$ because two shapes cannot be correlated. It should be reduced further in ^{3}HeA by $1/\sqrt{3}$, if the same logic holds.

Hydrodynamical predictions are opposite: double (or triple) initial explosions still lead to one common fireball, with the initial anisotropies larger than in pA. Data from RHIC by PHENIX and STAR on dAu, ^{3}HeAu do indeed show such an increase of the v_2, v_3, relative to pAu, again in quantitative agreement with hydrodynamics (Bozek and Broniowski, 2014; Nagle *et al.*, 2014).

ATLAS was able to perform the first measurements of higher harmonics $v_n, n = 2$–5 in central pPb; see Fig. 22. Except at very high $p_\perp$, those two harmonics seem to be comparable in magnitude: it is the first contradiction to "viscous damping" systematics.

B. Pedagogical digression: Scale invariance of sQGP and small systems

The acceptance of the hydrodynamical treatment of "small system explosions" is psychologically hard for many. One asks how is it possible to treat a fireball, of size less than 1 fm, as a macroscopic one.[15]

So let us take a step back from the data and consider the issue of scales. If one takes smaller and smaller cells of ordinary fluid, such as water or air, eventually one reaches the atomic scale, beyond which water or air as such do not exist: at some scale one resolves the individual molecules. But QGP is not like that: it is made of essentially massless quarks and gluons which have no scale of their own. The relevant scale is given by only one parameter T; thus QGP is approximately scale invariant. (The second scale $\Lambda_{\rm QCD}$ enters only via a logarithmic running of the coupling, which is relatively slow and can in some approximation be ignored.)

As lattice numerical calculations show (see, e.g., Fig. 64), at $T > 200$ MeV the QGP thermodynamics is approximately

[14]Dividing the cost of the LHC, $\sim 10^{10}$ by the number of recorded pp events $\sim 10^{10}$ one finds that an event costs about a dollar each. Therefore, high-multiplicity pp events under consideration cost about a million dollars each, and one needs thousands of them to construct a correlation function.

[15]Note that just 15 years ago the same question was asked about systems of 6 fm size.

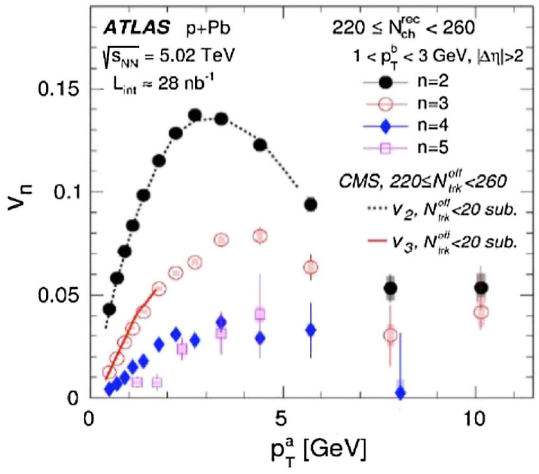

FIG. 22. v_n for $n = 2, 3, 4, 5$ vs $p_\perp$ in GeV, for the high-multiplicity bin indicated in the figure. The points are from ATLAS, the lines from CMS (presentation at QM2015).

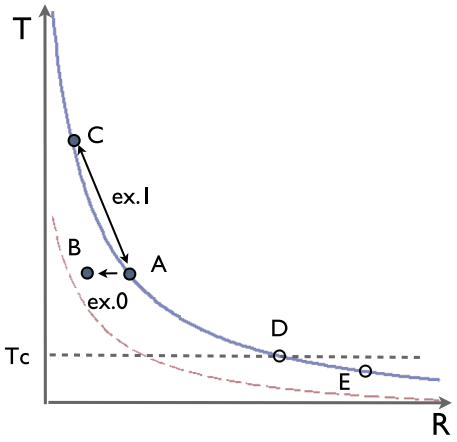

FIG. 23. Temperature T vs the fireball size R plane. The solid blue line is the adiabatic $S = \text{const}$, approximately $TR = \text{const}$ for sQGP.

scale invariant[16]: ϵ/T^4, p/T^4 are approximately T independent. The comparison of the LHC to RHIC data further suggests that similar scaling $\eta/T^3 \sim \text{const}$ holds for viscosity as well (although with less accuracy so far). Thus, as a first approximation one may assume that QGP does not have a scale of its own. This means that it would show exactly the same behavior if conditions related by the scale transformation

$$R_A/R_C = \xi, \qquad T_A/T_C = \xi^{-1} \tag{26}$$

are compared.

Consider a thought experiment 1, in which we compare two systems on the same adiabatic A and C; see Fig. 23. For scale-invariant sQGP the points A and C are related by this scale transformation mentioned previously and have the same entropy.[17] Assuming the scale transformation is an approximate symmetry, one expects the same dynamical evolution. A smaller-but-hotter plasma ball C will explode in exactly the same way as its larger-but-cooler version A.

Let us now proceed to thought experiment 2, which is the same as the previous but in QCD, with a running coupling. In the sQGP regime it leads to (a very small, as lattice tells us) running of s/T^3, some (unknown) running of η/T^3, etc. The most dramatic effect is however not the running coupling *per se*, but the lack of supersymmetry, which allows for the chiral or deconfinement phase transition out of the sQGP phase at $T > T_c$ to the hadronic phase at $T < T_c$. The end of the sQGP explosion D thus has an absolute scale, not subject to a scale transformation.

[16]At $T < 200$ MeV there is no scale symmetry: there is significant change instead. It is of course taken into account in hydrodynamics and will be discussed later in the section on HBT radii.

[17]Note that the holographic models interpret the "RG scale" as an extra fifth coordinate. The evolution in scale is thus depicted as "gravitational falling" of particles, strings, fireballs, etc., along this coordinate. In this language, our two systems fall similarly in the same gravity, but since the smaller system starts "higher," it gets larger velocity at the same "ground level," given by a fixed scale T_c.

Let us consider two systems A and C of the same total entropy and multiplicity, both initiated in the sQGP phase, with initial conditions related by the scale transformation. Their evolution would be the same, until the larger or cooler one reaches $T \approx T_c$, where scale invariance of their evolution ends. The final results of the explosions are not the same. In fact, the smaller and hotter system has an advantage over the larger and cooler one, since the larger ratio of the initial and final scales T_i/T_f allows it to be accelerated more.

The hydrodynamic expansion does not need to stop at the phase boundary D. In fact, large systems as obtained in central AA collisions are known to freezeout at $T_f < T_c$, down to the 100 MeV range (and indicated in the sketch by the point E). However small systems obtained in peripheral AA or central pA seem to freezeout at D.

In summary, we expect hydrodynamics to work as well in smaller systems, due to approximate scale invariance of sQGP. Including deviations from scale invariance, one finds that small systems should explode even more violently, compared to larger or cooler systems, since a larger fraction of time is spent in the sQGP phase.

Another meaning of the term "small systems" is applied when not the actual size but entropy or multiplicity is reduced. Deviations from hydrodynamics in such cases are seen as higher viscous corrections. For a recent discussion of those issues see Romatschke (2016) and Spalinski (2016).

C. Comparison of the peripheral *AA*, central *pA*, and high-multiplicity *pp*

From thought experiments with some ideal systems, let us return to reality. We do it in two steps, starting in this section with naive estimates for the three cases at hand, based on standard assumptions about the collision dynamics, and then returning to a more model-dependent discussion later.

We want to evaluate the initial transverse radii and parton densities, not that of the fireball at freezeout, after a hydrodynamic expansion. The multiplicity is however the final one, but due to (approximate) entropy conservation

during hydrodynamics we think of it as a proxy for the entropy at all times. Entropy generated by viscosity during expansion is relatively small and can be corrected for, if needed.

(i) Our most studied case, the central AuAu or PbPb, is the obvious benchmark. With the total multiplicity about $N_{AA} \approx 10^4$ and transverse area of nuclei $\pi R_A^2 \approx 100$ fm^2 one gets the density per area

$$n_{AA} = \frac{N}{\pi R_A^2} \sim 100 \text{ fm}^{-2}. \tag{27}$$

This can be transformed into entropy if needed, in a standard way.

(ii) Central pA (up to a few percent of the total cross section) has a CMS track multiplicity of about 100. Accounting for the unobserved range of $p_\perp, y$ and neutrals increases it by about a factor of 3, so $N_{pA}^{\text{central}} \sim 300$. The area now corresponds to the typical impact parameter b in pp collisions, or $\pi\langle b^2\rangle = \sigma_{pp} \approx 10$ fm^2. The density per area is then

$$n_{pA}^{\text{central}} = \frac{N_{pA}^{\text{central}}}{\sigma_{pp}} \sim 30 \text{ fm}^2 \tag{28}$$

or $1/3$ of that in central AA. Using the power of the LHC luminosity CMS can reach as a fluctuation with the probability 10^{-6} another increase of the multiplicity, by a factor of 2.5 or so, up to the density $N_{pA}^{\max}/\sigma_{pp}$ in AA. Another approach used is a comparison of central pA with peripheral AA of the same multiplicity, or more or less the same number of participants, or similar matter density.

(iii) Now we move to the last (and most controversial) case, of the high-multiplicity pp collisions. (Needless to say the density is very low for minimum bias events.) "High multiplicity" at which CMS famously discovered the ridge starts from about $N_{pp}^{\max} > 100 \times 3$ (again, 100 is the number of CMS recorded tracks and 3 is extrapolation outside the detector covered).

The big question now is what is the area? Unlike in the case of central pA, we do not utilize standard Glauber and full cross section (maximal impact parameters). We instead address a fluctuation which has small probability. In fact, nobody knows the answer to this question. Based on the profile of pp elastic scattering (to be discussed in Sec. VIII.C), it should correspond to the impact parameter b in the black disk regime. If so $\pi b_{b.d.}^2 \sim 1/2$ fm^2, which leads to density per area

$$n_{pp}^{\max} \approx \frac{N_{pp}^{\max}}{\pi b_{b.d.}^2} \sim 600 \text{ fm}^2. \tag{29}$$

Other evidences about the glue distribution in a proton comes from diffractive electroproduction, especially of $\gamma \to J/\psi$. They also suggest a rms radius of only 0.3 fm, less than one-half of the electromagnetic radius.

In summary, in terms of the initial entropy density one expects the following order of the densities per area involved:

$$\frac{dN_{\text{maximal}}^{pA}}{dA_\perp} \sim \frac{dN_{\text{peripheral}}^{AA}}{dA_\perp} \ll \frac{dN_{\text{central}}^{AA}}{dA_\perp} \ll \frac{dN_{\text{maximal}}^{pp}}{dA_\perp}. \tag{30}$$

One may expect that the radial flow follows the same pattern: yet the data show it is not the case.

D. The size and radial flow puzzle for central pA

The simplest consequence of the radial flow is an increase in mean transverse momentum. CMS data as a function of multiplicity are shown in Fig. 24(a). While pp and pA data are shown by points, the AA ones (from ALICE) are shown by shaded areas. (The most central ones correspond to the upper edge of this shaded region.)

The next experimental signature of the radial flow is that the blueshifts it induces modify spectra of secondaries of different mass differently. While light pions retain their exponential shape of the p_t spectra, only with a blueshifted slope

$$T' = T_f e^\kappa, \tag{31}$$

where κ is the transverse rapidity of the flow. Note that the spectra of massive particles change their shape. Eventually, for very heavy particles (e.g., d or other nuclei), their thermal motion is negligible and their spectra depend completely on collective velocities. Their distribution has a characteristic peak at the fireball's edge, with $v = v_{\max}$, and thus the $p_\perp$ spectra develop a peak at $p_\perp = m v_{\max}$.

More specifically, flow creates a violation of the $m_\perp$ scaling (Shuryak and Zhirov, 1980). The $m_\perp$ slopes T', defined by the exponential form (above certain p_t)

$$\frac{dN}{dy dp_\perp^2} = \frac{dN}{dy dm_\perp^2} \sim \exp\left(-\frac{m_\perp}{T'}\right), \tag{32}$$

are sensitive indicators of the radial flow. A sample of such slopes for pA collisions, from CMS, is shown in Fig. 24. Note that for low-multiplicity bins (marked by 8 and 32 at the bottom right) one sees the same T' for all secondaries: this is the $m_\perp$ scaling: the flow is absent. This behavior is natural for independent string fragmentation, rescattering, or GLASMA models.

Flow manifests itself at higher multiplicity bins, in which the slopes T' are mass dependent. As seen from Fig. 24(b), they are growing approximately linearly with the particle mass. The effect gets more pronounced with multiplicity: this is the sign of stronger collective flow. Note that for the most central pA bin this slope exceeds[18] those in central PbPb collisions at the LHC, the previous record-violent explosion.

This gives rise to what we call the *radial flow puzzle*. Indeed, naive estimates of densities in the previous section may suggest that the explosion in the highest multiplicity pA case should still be weaker than in AA. Indeed, the system is smaller and the initial entropy density seems to be smaller as

[18]Predicted to happen before the data: check version v1 of Shuryak and Zahed (2013).

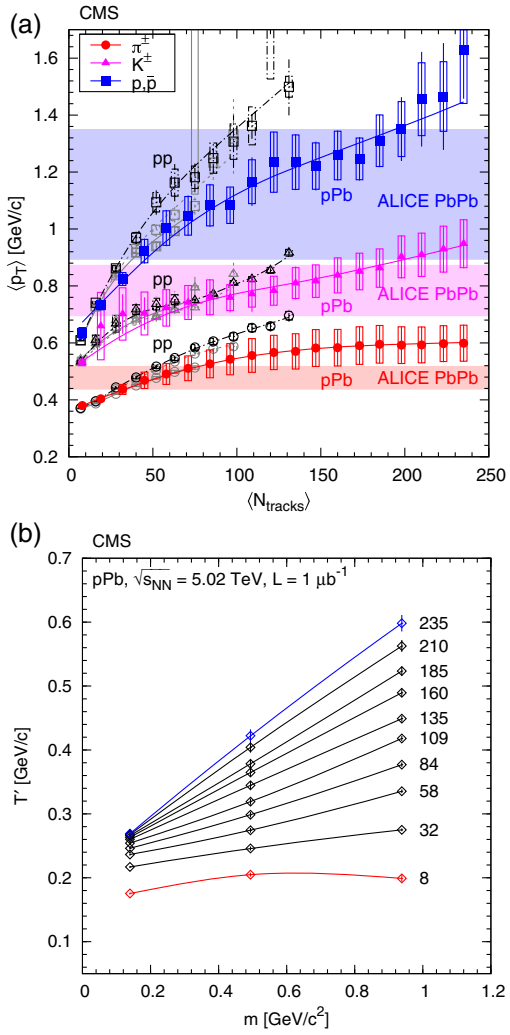

FIG. 24. (a) Average transverse momentum of identified charged hadrons (left panel: pions, kaons, and protons) and ratios of particle yields (right panel) in the range $|y| < 1$ as a function of the corrected track multiplicity for $|\eta| < 2.4$, for pp collisions (open symbols) at several energies, and for pPb collisions (filled symbols) at $\sqrt{s_{NN}} = 5.02$ TeV. (b) The slopes of the $m_\perp$ distribution T' (in GeV) as a function of the particle mass. The numbers on the right of the lines give the track multiplicity. From Chatrchyan *et al.*, 2014.

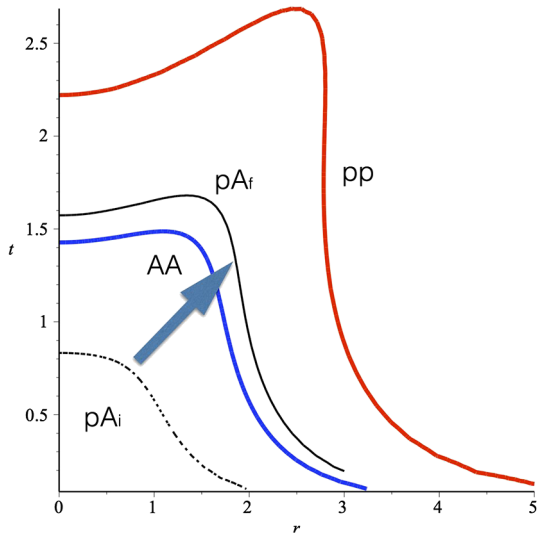

FIG. 25. The freezeout surface in universal dimensionless time t and radial distance r coordinates. The thick solid (blue) line in the middle corresponds to central AA (PbPb) collisions, and the thick solid (red) line on the top to the highest multiplicity pp. The two (black) thin lines correspond to the central pPb case, before and after collapse compression, marked pA_i and pA_f, respectively. The arrow connecting them indicates the effect of multistring collapse.

well. Yet the data show the opposite: the observed radial flow strength (expressed via the magnitude of the transverse rapidity) follows a different pattern

$$y_\perp^{AA,\text{central}} < y_\perp^{pA,\text{central}} < y_\perp^{pp,\text{highest}}. \tag{33}$$

Hydrodynamics is basically a bridge between the initial and the final properties of the system. For the radial flow dependence on the size of the system it is convenient to follow Shuryak and Zahed (2013) based on Gubser's flow; see Appendix B.2. One single analytic solution describes all cases considered: we proceed from the dimensional variables $\bar{\tau}, \bar{r}$ with the bar to the dimensionless variables

$$t = q\bar{\tau}, \qquad r = q\bar{r} \tag{34}$$

rescaled by a factor q, with the dimension of an inverse length. In such variables there is a single Gubser solution of ideal relativistic hydrodynamics, for the transverse velocity Eq. (B29) and the energy density Eq. (B30).

Recall our thought experiment 1 of Sec. V.B: two collisions which are conformal copies of each other merge into a single one in these dimensionless variables. In fact, the line (blue)[19] marked AA in Fig. 25 corresponds not only to central PbPb collisions but actually to any other AA collisions at the LHC. The two black lines are for the pPb case; they both have $T_f = 170$ MeV and the same multiplicity but different scale parameters: $q = 1/1.6$ fm for the lower dotted line but twice smaller initial size $q = 1/0.8$ fm for the upper (thin solid black) line. As the arrow indicates, in order to explain the observed higher explosion velocity one has to move the curve above the AA (blue) line. This implies that hydrodynamics must be initiated from a smaller "compressed" size, according to the "spaghetti collapse" scenario discussed in Sec. V.B. If this is done, the freezeout surface "jumps over" our AA benchmark blue line, and its radial flow gets stronger. The maximal transverse velocities on these curves (located near the turn of the freezeout surface downward) are

$$v_\perp^{p\text{Au}} = 0.56 < v_\perp^{AA} = 0.81 < v_\perp^{p\text{Au},f} = 0.84. \tag{35}$$

The upper red line is our guess for the maximal multiplicity pp collisions, assuming its $q = 1/0.5$ fm: it has even stronger radial flow, with maximal $v_\perp^{pp} \approx 0.93$. Paradoxically, small systems are in fact larger than AA in the appropriate

[19]For the record, its parameters are $q = 1/4.3$ fm, $\hat{e}_0 = 2531$, and $T_f = 120$ MeV.

dimensionless variables, and that is why their radial flow is better developed.

In summary, the observed pattern of radial flow magnitude can be explained if the initial size of the pA system is significantly reduced compared to the naive estimates of the preceding section.

E. Radial flow in high multiplicity pp

According to our estimates of the densities per area, it is much higher for high multiplicity pp than for AA collisions. The initial state must be in a GLASMA state, if there is one. Unfortunately we have little theoretical guidance about the size. After all, in this section we discuss fluctuations with a probability $\sim 10^{-6}$. Lacking good theory guidance, one may invert our logic and proceed phenomenologically as follows:

(i) The first phenomenological input, the mean $p_\perp$ and spectra of the identified particles, is shown in Fig. 24. The absolute magnitude of the flows at freezeout (radial, v_2, v_3) can thus be evaluated from the data.

(ii) Then one can solve hydrodynamics backward and determine which initial conditions are required to generate it.

(iii) For consistency, one can calculate the absolute values of the radii provided by the femtoscopy. The observed radii show a decrease with the increase of the (total) transverse momentum $\vec{k}_1 + \vec{k}_2 = \vec{k}_\perp$ of the pair (Makhlin and Sinyukov, 1988). Modification of the argument is explained in Fig. 26. At small $k_\perp$ the detector sees hadrons emitted from the whole fireball, but the larger $k_\perp$ is, the brighter its small (shaded) part becomes in which the radial flow is (a) maximal and (b) has the same direction as $\vec{k}_t$. This follows from maximization of the Doppler-shifted thermal spectrum $\sim \exp\left(p^\mu u_\mu / T_{\text{freezeout}}\right)$.

This effect was calculated (Hirono and Shuryak, 2015) and compared with the ALICE HBT data (Aamodt *et al.*, 2011a) shown in Fig. 27. Strong flow in high-multiplicity pp collisions is directly visible in the data. The effect is best seen in the "out"-directed radius R_{out} (the top plot). While low-multiplicity data (connected by the dashed blue lines) are basically independent of the pair momentum, at high multiplicity (stars and dashed red lines) they decrease by a rather large factor. Another consequence of the flow is anisotropy of radii. In the bottom plot the ratio of two radii is shown. At small multiplicity it is always 1 (that is, the source is isotropic)

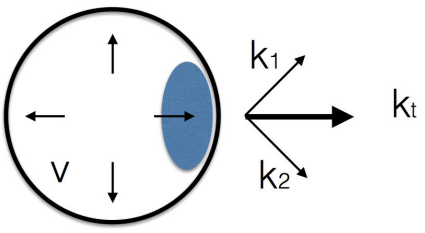

FIG. 26. Sketch of the radial flow (arrows directed radially from the fireball center) explaining how it influences the HBT radii. At small k_t the whole fireball (the large circle) contributes, but at larger k_t one sees only the part of the fireball which is comoving in the same direction as the observed pair. This region, shown by the shaded ellipse, has a smaller radii and anisotropic shape, even for central collisions.

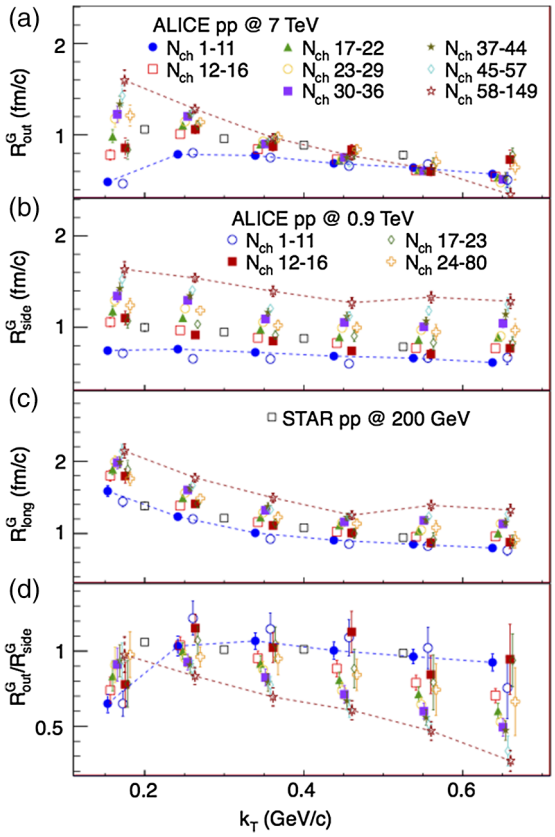

FIG. 27. HBT radii vs the pair transverse momentum k_T, for various multiplicities of the pp collisions, from ALICE. From Aamodt *et al.*, 2011a.

but at high multiplicity the source becomes anisotropic, the radii in two directions are quite different, with their ratio dropping to about $1/3$ at the largest k_t. Thus, a direct consequence of the flow is that only $1/3$ of the fireball emits pairs of such momenta.

In Fig. 28 we show a series of calculations in which the initial QGP stage of the collision is modeled by a numerical hydrodynamics solution close to the Gubser analytic solution

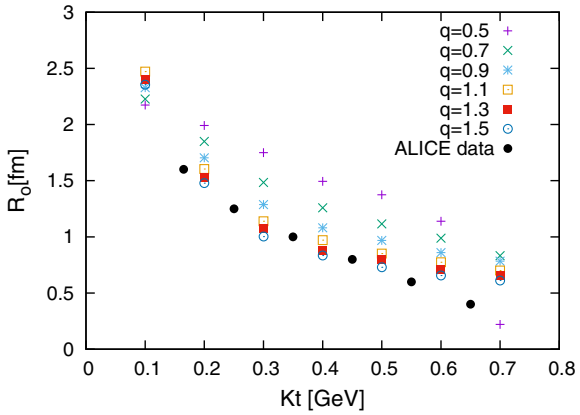

FIG. 28. HBT radii compared to ALICE data (solid circles), for solutions starting with different initial size of the fireball, indicated by the Gubser scale parameter q (which is inversely proportional to the size). From Hirono and Shuryak, 2015.

with variable parameter q. [The late stages need to deviate from the Gubser solution since near T_c the equation of state (EOS) is very different from conformal $\epsilon = 3p$ assumed in Gubser's derivation.]

In summary, unlike central pA the highest multiplicity pp events are significantly denser and hotter than central AA. A strong radial flow, seen in the spectra of identified particles and HBT radii, requires very small, subfemtometer, initial size of the system. In spite of the high cost associated with those events, their studies are justified because here we produce the most extreme state of matter ever created in the laboratory.

F. Can flows in small systems be "fake"?

The question of what we call the "fake flow" and subsequent development is due to Romatschke (2015) who considered consequences of a scenario in which quarks and gluons at the QGP phase have no interactions, they "free stream" from the point of the initial scattering to the hadronization surface. At this surface the system switches to the hadronic phase, treated by a standard hadronic cascade code.

In Fig. 29 one sees a comparison of the radial flow profiles of the two cases, with and without interaction at the QGP phase. One can see that the profiles are in fact very similar, becoming linear Hubble-like as time goes on. In fact free streaming generates even a bit stronger flow, because free streaming uncouples from the longitudinal direction, and an equilibrated medium does not. Comparing particle spectra and HBT radii Romatschke showed that this fake radial flow is indeed indistinguishable from the hydrodynamics.

What about flow harmonics? The results for v_2 in PbPb collisions are shown in Fig. 30. As one can see, without hydrodynamics they nearly disappear. This is not surprising since if there is no hydrodynamics then there are no sound waves, and initial bumps are simply dissolved without a trace.

In fact it is an interesting question how any v_n can be generated in the free streaming. The initial momentum distribution of partons is isotropic, and so it must be related to the interaction after hadronization. Romatschke found that indeed before hadronization they are absent. However, two components of $T_{\mu\nu}$, the flow $\sim u_\mu u_\nu$ part and the dissipative $\Pi_{\mu\nu}$ part, still have nonzero values which cancel each other in sum. After hadronization the hadronic interaction kills the second component $\Pi_{\mu\nu} \to 0$ and reveals the effect of the first one.

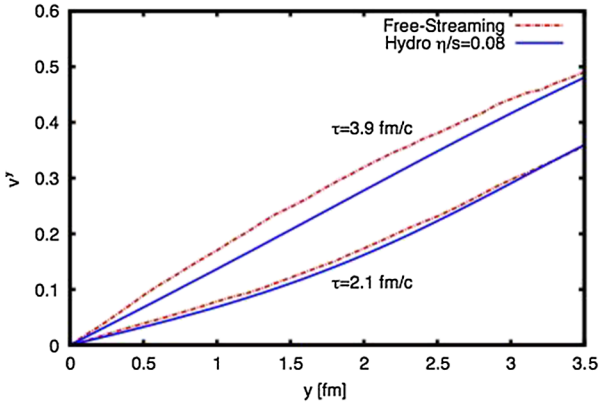

FIG. 29. Comparison of the flow profile, for hydrodynamics and free streaming. From Romatschke, 2015.

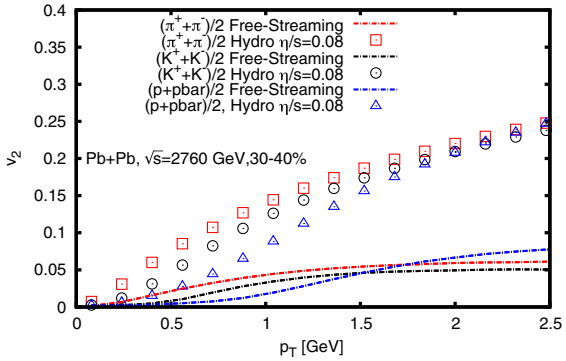

FIG. 30. Comparison of the flow harmonic for hydrodynamics and free streaming for PbPb collisions. From Romatschke, 2015.

Not only are the fake flow harmonics small, they do not show two important features of the "true" hydrodynamical ones: (i) they do not show a strong increase with $p_\perp$, and (ii) they do not show a strong decrease with the number $\sim \exp(-n^2)$ induced by the viscosity during the time before hadronization.

Unlike the radial flow, higher harmonics in large (PbPb) systems cannot be faked. What about smaller systems? Romatschke gives the results for pPb at the LHC and dAu and ^{3}HeAu for RHIC energies. We show the first case in Fig. 31. Again the free streaming model seems to be failing for v_2, but is somewhat marginally surviving for v_3.

In summary, flow harmonics are not faked. Yet, for small systems, taking into account remaining uncertainties of the initial stage models and thus ϵ_n values, this conclusion is not as robust as for the AA. Perhaps some scenarios, intermediate between equilibrated hydrodynamics and free streaming, may still fit these data.

G. Shape fluctuations: Central pA vs peripheral AA

Scaling relations between central pA and peripheral AA were suggested by Basar and Teaney (2014). Step one of their paper was prompted by the fact (noticed in the CMS paper

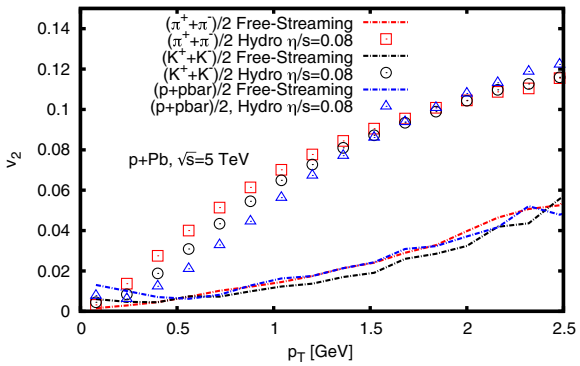

FIG. 31. Comparison of the flow harmonic for hydrodynamics and free streaming in the pPb central bin. From Romatschke, 2015.

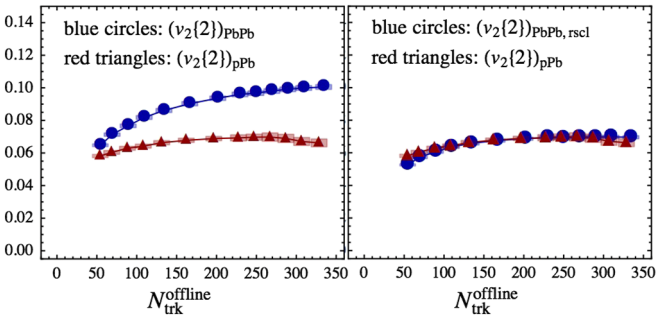

FIG. 32. The integrated $v_2\{2\}$ for PbPb and pPb vs multiplicity. Left: Original values. Right: The fluctuation dependent elliptic flow, with the geometrical part subtracted. This geometrical part was calculated using the Phobos Glauber model and is not a fit. From Basar and Teany, 2014.

already) that at the same multiplicity v_3 in central pA and peripheral AA are basically the same. Some called for new paradigms based on this fact, but in fact it is hardly surprising: equal multiplicity means an equal number of participant nucleons, and thus equal fluctuations of the shape. After the geometrical contribution to v_2 in peripheral AA is removed, the remaining, fluctuation-driven, part of the elliptic flow is exactly the same in both cases; see Fig. 32.

Their second proposal is that the $p_\perp$ dependence of (the fluctuating part) of the v_n has a universal shape, and AA and pA data are different only by a scale of mean $p_\perp$:

$$v_n^{pA}(p_t) = v_n^{pA}\left(\frac{p_\perp}{\kappa}\right), \tag{36}$$

where the scaling factor is defined as

$$\kappa = \frac{\langle p_\perp \rangle_{p\mathrm{Pb}}}{\langle p_\perp \rangle_{\mathrm{PbPb}}} \approx 1.25 \tag{37}$$

and is due to the difference in the radial flow.

VI. EQUILIBRATION IN QCD-BASED MODELS

A. CGC and turbulent GLASMA

The idea of continuity, from a state before collision to early time after it, is most directly realized in the so-called CGC-GLASMA approach. Technically it is based on the argument (McLerran and Venugopalan, 1994) that high density of partons leads to large color charge fluctuations, which should create strong color fields. If fields are strong enough, then classical Yang-Mills (YM) equations are sufficient, and those can be solved numerically. It is important that at this stage the fields get strong, the occupancy of gluons $n_g \sim 1/\alpha_s \gg 1$, and, by rescaling them, one can get the coupling out of the equations. This means that GLASMA is nonperturbative, in spite of weak coupling. It remains so until gluon occupation numbers drop to their thermal magnitude $n_g \sim 1$.

When the density of gluons gets large enough and nonlinear effects become important, the GLASMA is in its "dense regime." Its boundary, shown in Fig. 33, is defined by the saturation momentum $Q_s(x)$, separating it from the dilute

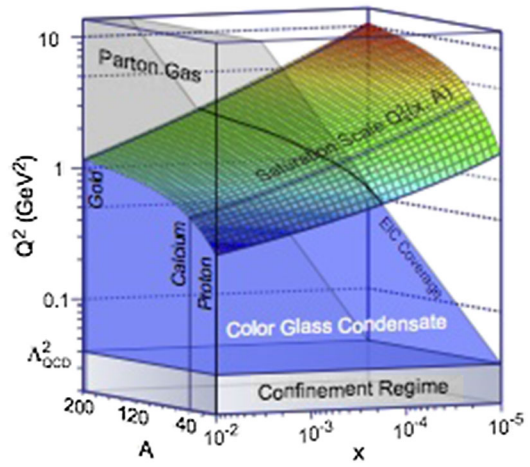

FIG. 33. The CGC phase diagram: the saturation momentum Q_s as a function of the fraction of momentum x and the atomic number A. From Lappi, Dumitru, and Nara, 2014.

partonic phase. $Q_s(x)$ is expected to grow with collision energy (smaller x) and higher atomic number A. At the highest LHC energies and atomic numbers its value is as large as $Q_s^2 \sim 10$ GeV^2, believed to be in the perturbative domain. However, another boundary, of the "confining regime" at the bottom of the figure, is indicated by extremely small $Q^2 < 0.03$ GeV^2. This is unacceptable: the boundary of pQCD is in fact at least a factor of 30 or more higher. There are no gluons with virtuality below 1 GeV.[20] Modern lattice simulations show that the gluon effective mass in QGP is of the order of 1 GeV at $T = T_c$ and grows further at $T > T_c$.

The theoretical study of parton equilibration in a weak-coupling domain has a long history. The "bottom-up" approach (Baier *et al.*, 2001) was based on soft gluons radiated by scattered hard partons. The name reflects the fact that thermal occupation starts from the IR end. (Note that it is opposite to the "top-down" equilibration in holographic models we discuss next.) The main predictions of that model were the equilibration time and the initial temperature scaling with the coupling

$$\tau_{\mathrm{eq}} \sim 1/\alpha_s^{13/5} Q_s, \qquad T_i \sim \alpha_s^{2/5} Q_s. \tag{38}$$

Some details were changed later. Weibel, Nielsen-Olesen, and other instabilities that occur in the model were incorporated. Its validity domain is restricted by its core assumption of small-angle scattering of the gluons, justified by large impact parameters of the order of the inverse (perturbative) Debye mass. Perturbative means $M_D \approx gT \ll T$ or small $g \ll 1$. For its edge values $g = 1$, $\alpha_s = 1/4\pi$ the equilibration time is predicted to be very long $\tau_{\mathrm{eq}} Q_s \sim 700$, exceeding the duration of the collisions.

In the last few years several groups performed numerical studies of parton equilibration using both the Boltzmann and

[20]Recall that already the Nambu–Jona-Lasinio model of 1961 had strong nonperturbative forces at $Q < 1$ GeV creating chiral symmetry breaking.

the YM field equations. Typically, in such studies the coupling constant is taken to be extremely small. In fact, so small that one can treat not only powers of α_s but even its log as a large parameter $\log(1/\alpha_s) \gg 1$, allowing the total GLASMA evolution scale

$$\tau_{\mathrm{GLASMA}} \sim \frac{\log^2(1/\alpha_s)}{Q_s} \tag{39}$$

to be considered large.

Significant progress in this direction was induced by the incorporation of certain ideas from the general theory of turbulent cascades. Not going into its long history, we just mention the Kholmogorov-Zakharov stationary powerlike solutions for Boltzmann equations for a number of systems with various waves. Another general advance is the existence of the time-dependent self-similar solutions. J. Berges and collaborators developed it for scalar and gauge fields, pointing out different regimes for UV- and IR-directed cascades and identifying such regimes in impressive numerical simulations.

This body of work resulted in the following new scenario: the preequilibrated stage is dominated by a nontrivial turbulent attractor, a certain self-similar power solution, in which it spent some time before progressing toward the thermal equilibrium. An important signature of that is large momentum anisotropy, measured by the ratio $p_l/p_\perp$. One group (Epelbaum, 2014) performed a next-order GLASMA simulation, with $g = 0.5$, in which $p_l/p_\perp$ kept approximately a constant value during the entire time of the simulation $\tau Q_s = 10$–40. Another group (Berges *et al.*, 2015) found that, at $g = 0.3$, the longitudinal pressure p_l/ϵ remains close to zero at similar times. Both results are shown in Fig. 34.

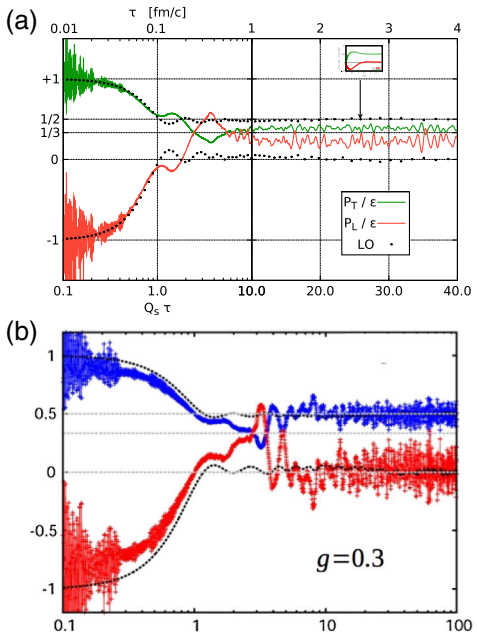

FIG. 34. (a) The upper green curve is $p_\perp/\epsilon$, and the lower red is p_l/ϵ, as a function of time in units of saturation scale τQ_s at $g = 0.5$. From Epelbaum, 2014. (b) The upper blue curve is $p_\perp/\epsilon$, and the lower red is p_l/ϵ at $g = 0.3$. From Berges *et al.*, 2015.

Cascade simulations at larger couplings (Kurkela and Zhu, 2015) produce the results shown in Fig. 35. The two sets of trajectories, shown by solid and dashed lines, start from two different initial distributions. At zero coupling (upper left curve) the longitudinal momenta of particles gets very small compared to the transverse, and anisotropy steadily increases. This is a scalinglike classical regime with a nontrivial fixed point. However, all other paths stay more or less at the same initial anisotropy and then rapidly turn downward to locally isotropic distributions (marked by crosses at the bottom). Unfortunately no simulations were done with λ between 0 and 0.5: perhaps at some critical coupling a bifurcation of the trajectories happens, separating those who proceed toward the new and the equilibrium fixed points. Yet the issue is rather academic, since the realistic relevant coupling value $\alpha_s = g^2/4\pi \approx 0.3$ corresponds to the 't Hooft coupling constant $\lambda = \alpha_s N_c 4\pi \approx 10$, which is the largest value shown in this figure (bottom right). The corresponding curve rapidly approaches the equilibrium point, with coordinates (1,1) in this figure.

B. From GLASMA to hydro

The weak-coupling cascades discussed previously predict highly anisotropic pressure $p_l \ll p_\perp$. The question is about the time during which this feature persists, before the viscous hydrodynamics becomes valid. Recent research focused on a "hydrodynamization," a convergence of the stress tensor calculated using GLASMA or parton cascade simulations to a form appropriate for hydrodynamics. In other words, the issue is the relaxation mechanism or time of the nonhydrodynamical modes; see, e.g., Keegan *et al.* (2016). Perturbative parton cascades propose time of several fm/c, while strong-coupling approaches such as AdS and QCD suggest an order of magnitude shorter time, around 0.5 fm/c.

To know which value is the case one needs to do more calculations using both approaches, especially those that can be directly compared to the data. Surprisingly, there is little discussion of how to measure the anisotropy time in

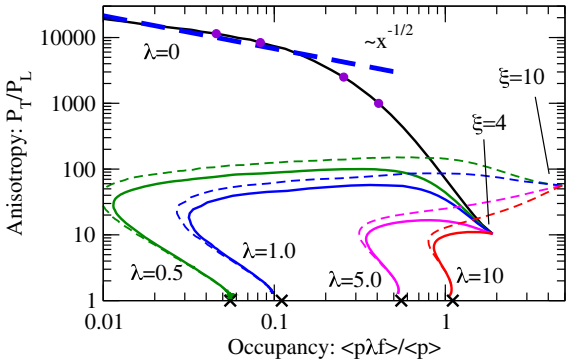

FIG. 35. Trajectories of the systems on the occupancy-anisotropy plane for various settings. The parameters λ near the curves show the corresponding 't Hooft coupling constant $\lambda = g^2 N_c$. All solid lines originate from one initial distribution characterized by the anisotropy parameter $\xi = 4$, and the dashed lines originate from a different point with $\xi = 10$. From Kurkela and Zhu, 2015.

experiment. One particular suggestion, via dilepton polarization, is discussed in Sec. IX.C.

Since the longitudinal pressure p_l changes most, one can perhaps study the effect of the longitudinal pressure on the rapidity distribution. Historically, the original papers of Landau focused on the longitudinal expansion. But Landau's initial condition, the instantaneous stopping, is rather unrealistic in QCD. To quantify a realistic initial rapidity distribution, from PDFs or GLASMA theory, is still to be done.

Since the beginning of the RHIC era, the time of hydrodynamization has been empirically derived from a comparison to data on the radial and especially the elliptic flows. It was shown (Molnar and Huovinen, 2005) that the parton cascades tend to effectively dissipate anisotropies needed for the elliptic flows. This is especially true for high $p_\perp \sim 3$ GeV and peripheral collisions. Even for rather extreme values of the parton cross section, a cascade and a hydrodynamical evolution diverge at rather early time (Molnar and Huovinen, 2005). The current round of studies based on parton cascades and kinetic equations needs to directly address these issues.

Figure 36 illustrates what is done in the IP GLASMA approach. An important feature of GLASMA is independent fluctuations of color in different cells, which seeds the harmonic flows. At certain proper time, 0.2 fm/c in this example, GLASMA evolution is stopped and the stress tensor is matched to that of an ideal fluid. For technical reasons the value of the viscous tensor is put to zero. How important is the selected time 0.2 fm/c? Note that the second picture is hardly different from the first, except the overall scale of the energy density is reduced. Indeed, 0.2 fm is a small distance relative to the nuclear size, and all one finds at this time is dilution due to longitudinal stretching. By starting hydrodynamics right from the second picture, Schenke and collaborators implicitly assumed that hydrodynamics cells can indeed be as small as 0.2 fm, and that their code can cope with large gradients between the cells. (Typically hydrodynamics starts at a few times later time, 0.6 fm/c or more.)

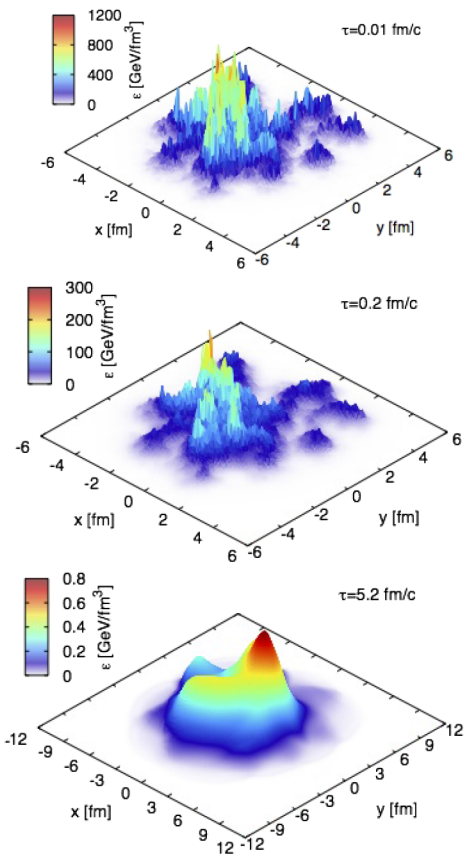

FIG. 36. Transverse energy profile from the IP-GLASMA model for a semiperipheral ($b=8$ fm) Au+Au collision at $s = 200A$ GeV, at times $\tau = 0.01$, 0.2, and 5.2 fm/c. From $\tau =$ 0.01 to 0.2 fm/c the fireball evolves out of equilibrium according to the GLASMA model. From Schenke and Venugopalan, 2014.

Following the evolution to the bottom figure of Fig. 36, at time 5.2 fm/c, one finds it to be very different. The original bumps have disappeared and instead a new one at another location is formed. Indeed, sound perturbations cannot stand still and must move with the speed of sound. At intersections of "sound circles" from the primary bumps random enhancements of the density are observed. Yet since the bumps are statistically uncorrelated, those should get averaged out, at least in two-particle correlations, and only correlations from the same circle will stay.

How many harmonics are needed to describe pictures such as those shown in Fig. 36? Taking 0.2 fm as a resolution and 4 fm as the fireball size, one finds that the upper picture requires about $20 \times 20 = 400$ pixels to be represented by certain stress tensor components. At the freezeout there are only several angular harmonics observed, so 99% of the information shown in those pictures does not survive until the freezeout. In the hydrodynamic simulation just described those disappear predictably, via viscous damping. It is possible that these systematics will fail at shorter wavelengths: so it is worth trying to measure higher harmonics. Other ways to observe density waves can perhaps be invented: one of such is the potential observation of those in the dileptons to be discussed next.

C. The initial state and angular correlations

The role of the initial state is greater for "small systems." When a nucleon is going along the diameter of large-A nucleus the mean number of participant nucleons is

$$\langle N_p \rangle = n_0 \sigma_{NN} 2R_A, \tag{40}$$

so for pPb at the LHC one gets $\langle N_p \rangle \approx 16$.

The question however is, where exactly in the transverse plane is the deposited energy located?

In Fig. 37 we sketched two opposite models of the initial state. Figure 37(a) shows each of the N_p participants represented by N_g gluons (ignoring sea quarks and antiquarks) from their PDFs each, so the total number of partons is $N_p N_g$. We assume that these gluons are randomly distributed in a spot of the size of the pp cross section. Figure 37(b) shows an alternative picture, the stringy Pomerons, in which there are no gluons but $2N_p$ QCD strings instead. Since those are cold (unexcited), they are shown by straight lines.

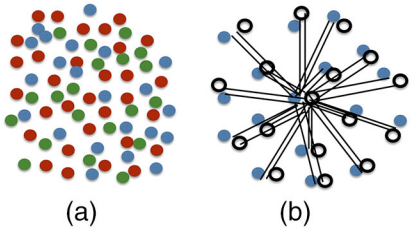

FIG. 37. Sketch of the initial state in central pA collisions. (a) The IP-GLASMA model, with colored circles representing multiple gluons. (b) $N_p = 16$ Pomerons, each represented by a pair of cold strings. The open circles are quarks and filled blue circles are diquarks.

Let us estimate the deformation of the initial state in central collisions, for which there are no geometrical effects and all deformations come from the fluctuations. As discussed, for all n one expects the same magnitude

$$\epsilon_n \sim \frac{1}{\sqrt{N}}, \tag{41}$$

where $N = N_p N_g$, the number of participant nucleons times the number of gluons per participant, for Fig. 37(a) and $N = N_p$ for Fig. 37(b). Evaluating N_g from PDFs at LHC energy includes integration from $x_{\min} \sim 10^{-3}$ to 1: one roughly gets the ratio

$$\frac{\epsilon_n^{(b)}}{\epsilon_n^{(a)}} \sim \frac{1}{\sqrt{N_g}} \sim 4. \tag{42}$$

The elliptic and triangular flows in very peripheral AA and central pA studied by Schenke and Venugopalan (2014) demonstrated that the IP-GLASMA model does a good job for the former case, but strongly underpredicts fluctuations in the latter case; see Fig. 38.

As already discussed, in the peripheral AA ϵ_2 is large, $O(1)$, in any model, and in order to get the right v_2 one has to have the correct viscosity—which apparently Schenke and Venugopalan have. The central pA is indeed the test case: we argued that the density is not yet large enough to apply the IP-GLASMA model, while the stringy Pomeron model should be applicable instead. If so, using Eq. (42) we should increase the v_2 by a factor of 4, which brings it to an agreement with the CMS measurements. We thus conclude that the stringy model of Fig. 37(b) is preferable over Fig. 37(a), the uncorrelated gluons.

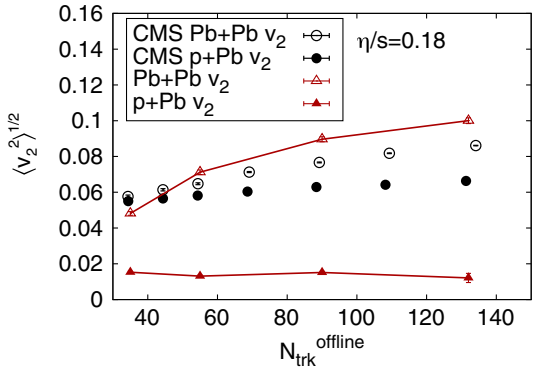

FIG. 38. Multiplicity dependence of the root-mean-square elliptic flow coefficient v_2 in Pb + Pb (open symbols) and p + Pb collisions (solid symbols) from the IP-GLASMA + music model (connected triangles) compared to experimental data by the CMS Collaboration. From Schenke and Venugopalan, 2014.

Previously we simplistically assumed a Glauber picture in which each wounded nucleon (or a participant) interacts with the projectile proton by a single Pomeron. Note that one gluon exchange generates (at least) two strings. If strings are simply stretched longitudinally, until they fragment independently, the rapidity distributions would be flat (rapidity independent) for all centrality classes. This is not the case, as seen in the ATLAS data shown in Fig. 39. As one can see, the peripheral bins have flat rapidity distribution: few strings are produced, and those are extended through all rapidities. Yet central bins for pPb have rather asymmetric distributions, with larger multiplicity at the Pb side.

In the Pomeron language, this is explained by the "fan diagrams," in which one Pomeron can split into two. The "triple Pomeron vertex" is, however, small, preventing development of extensive "Pomeron cascades." The multiplicity difference between the rhs and the lhs of the plot is not too dramatic; it certainly is not proportional to N_p scaling. For example, for the most commonly used centrality bin 1%–5% the rapidity density $dn_{\rm ch}/d\eta$ changes from about 35 to 55, across the rapidity interval shown in this figure. If on the Pb end there are say $N_s > 2N_p \approx 40$ strings, then on the p end there are not one or a few, but approximately 20 strings. Since the area on the lhs is reduced by an order of magnitude or so, and the number only by a factor of 2, it is by far a more dense system than the rhs. One may further ask if flows, and the development of collectivity, depend on rapidity. So far we do not see any evidence of that. For example, the famous v_2 ridge is rapidity independent.

Finally, we briefly mention the case of high-multiplicity pp collisions. We do not yet know ϵ_n in this case. Theoretical predictions for pp cover the whole range: from elongated transverse strings (Bjorken, Brodsky, and Scharff Goldhaber, 2013) predicting large $\epsilon_2 \sim 1$ to a IP-GLASMA or "stringball" picture (Kalaydzhyan and Shuryak, 2014c) which predicts

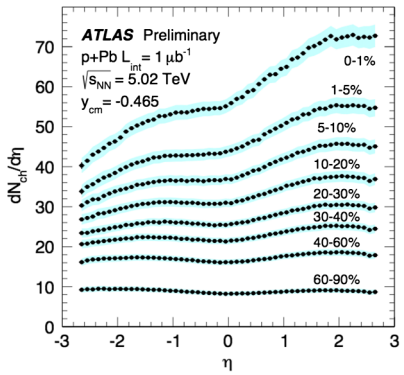

FIG. 39. Rapidity distribution in pPb collisions for different centrality classes, from "Measurement of the centrality dependence of the charged particle pseudorapidity distribution in proton-lead collisions," ATLAS-CONF-2013-096, https://cds.cern.ch/record/1599773.

very small ϵ_2 instead. Experiment gives v_2 and v_3 for 2 particle correlations, but 4, 6 particle correlators are still beyond the reach for statistical reasons.

D. Multistring dynamics

A version of the initial state theory, an alternative to the GLASMA picture at high density, is the old Lund model, represented by event generators such as PYTHIA. It is supposed to be applicable for lower matter density, remaining in the confined phase. Multiple color charges, moving relativistically from each other after collisions, are connected by multiple QCD strings. As they are rapidly stretched longitudinally, the strings become nearly parallel to each other.

Note that in both GLASMA and string pictures the color fields have a similar longitudinal structure: one difference though is that GLASMA also has longitudinal magnetic fields. GLASMA state dynamics leads to interesting oscillations shown in Fig. 40: we will return to its analog in the string model using holography later.

Transition between the two pictures (GLASMA and a multistring state called a "spaghetti") is expected when the string diluteness parameter becomes of the order of 1, so they can no longer be separated. This is expected to happen at

$$\frac{N_{\rm string}}{\rm area} \sim \frac{1}{\pi r_{\rm string}^2} \sim 10\ {\rm fm}^2, \tag{43}$$

where in the numerical value we use the field radius in the string $r_s \approx 0.17$ fm ~ 1 GeV^{-1} from lattice measurements.

Collective interaction between the QCD strings in a spaghetti state was studied by Kalaydzhyan and Shuryak (2014c). An analysis of the lattice data made there confirmed that the string interaction at large distances is mediated by the lightest scalar σ, similar to long-distance forces between nucleons. Specifically, the shape of the quark condensate around the string is well described by

$$\frac{\langle \bar{q}q(r_\perp) W\rangle}{\langle W\rangle \langle \bar{q}q\rangle} = 1 - CK_0(m_\sigma \tilde{r}_\perp) \tag{44}$$

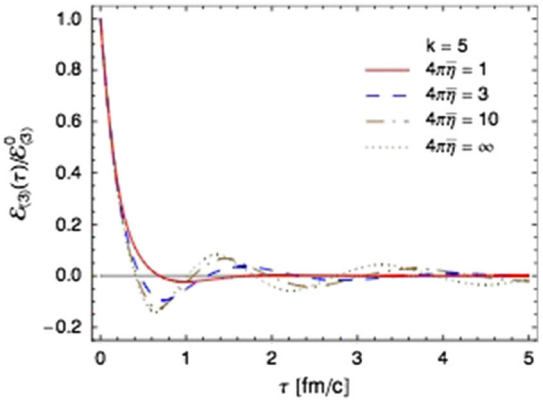

FIG. 40. Oscillation of the energy density in simulations starting from "GLASMA"-like initial conditions. $k = 5$ is the number of fluxes through the flux tubes. From Florkowski and Ryblewski, 2014.

where K_0 is the modified Bessel function and the "regularized" transverse distance $\tilde{r}_\perp$ is

$$\tilde{r}_\perp = \sqrt{r_\perp^2 + s_{\rm string}^2}, \tag{45}$$

which regulates the Coulomb singularity $\sim \ln(r_\perp)$ at small $r_\perp$; see Fig. 41. The σ mass value used is $m_\sigma = 600$ MeV.

Since the strings are almost parallel to each other, the problem is reduced to the set of point particles in a 2D plane with the 2D Yukawa interaction. From the fit (44) one can see (Kalaydzhyan and Shuryak, 2014c) that the main parameter of the string-string interaction (in string tension units) is numerically small,

$$g_N \sigma_T = \frac{\langle\sigma\rangle^2 C^2}{4\sigma_T} \ll 1, \tag{46}$$

typically in the range 10^{-1}–10^{-2}. Therefore it was correctly neglected in the situations for which the Lund model was originally invented—when only $\mathcal{O}(1)$ strings are created.

The collective interaction plays a role when this smallness can be compensated by a large number of strings. As seen from Fig. 41, a magnitude of the quark condensate $\sigma = |\langle \bar{q}q\rangle|$ at the string position is suppressed by about 20% of its vacuum value. In a spaghetti state one should think of the quark condensate suppression of about 0.2 times the diluteness, which is still less than 1.

On the other hand, about five overlapping strings are enough to eliminate the condensate and restore the chiral symmetry. If $N_s > 30$ strings implode into an area several times smaller than $\sigma_{\rm in}$, then the chiral condensate will be

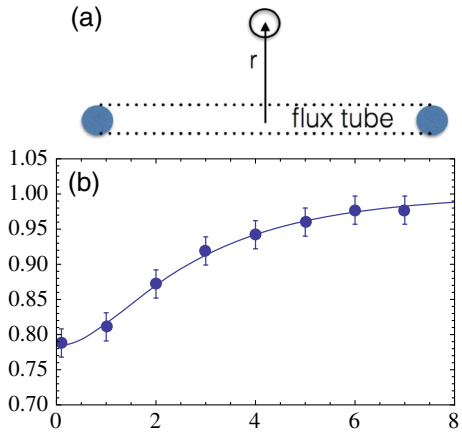

FIG. 41. (a) A static quark-antiquark pair is indicated by shaded circles: those are connected by the flux tube (QCD string). At distance r from the tube the local value of the quark condensate $\bar{q}q(r)$ is measured. (b) Normalized chiral condensate as a function of the coordinate r transverse to the QCD string (in lattice units). Points are from the lattice calculation (Iritani, Cossu, and Hashimoto, 2014). The curve is Eq. (44) with $C = 0.26$, $s_{\rm string} = 0.176$ fm.

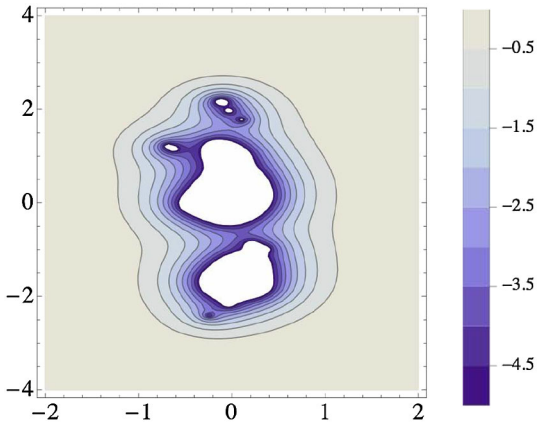

FIG. 42. Instantaneous collective potential in units $2g_N\sigma_T$ for an AA configuration with $b = 11$ fm, $g_N\sigma_T = 0.2$, and $N_s = 50$ at the time $\tau = 1$ fm/c. White regions correspond to the chirally restored phase.

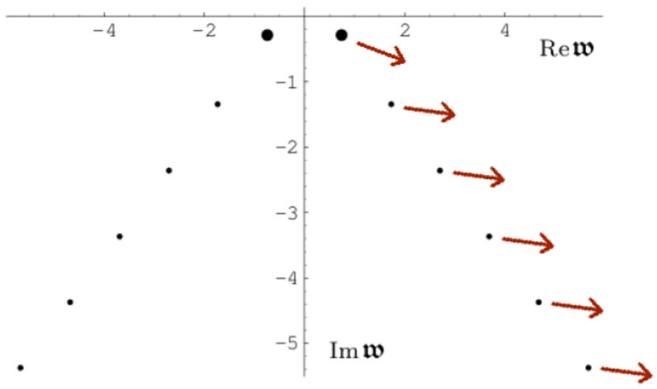

FIG. 43. A set of frequency modes, on the ω complex plane(. The dots are for a particular wave vector k, and the arrows indicate the direction of motion as k increases. From Kovtun and Starinets, 2005.

eliminated inside a larger region of 1 fm in radius, or about 3 fm^2 in area. This is nothing but a hot QGP fireball.

As discussed, the strings can be viewed as a 2D gas of particles (in transverse plane) with unit masses at positions $\vec{r}_i$. The forces between them are given by the derivative of the energy (44), and

$$\ddot{\vec{r}}_i = \vec{f}_{ij} = \frac{\vec{r}_{ij}}{\tilde{r}_{ij}}(g_N\sigma_T)m_\sigma 2K_1(m_\sigma\tilde{r}_{ij}) \tag{47}$$

with $\vec{r}_{ij} = \vec{r}_j - \vec{r}_i$ and regularized $\tilde{r}$, Eq. (45). In the simulations a classical molecular dynamics code was used.

The evolution consists of two qualitatively distinct parts: (i) early implosion, which converts potential energy into kinetic and has its peak when a fraction of the particles "gravitationally collapses" into a tight cluster; and (ii) the subsequent approach to a "minigalaxy" in virtual quasiequilibrium. Only the first one is physical, as the imploded spaghetti has density sufficient for the production of a QGP fireball, and after that explodes hydrodynamically. The entire scenario thus resembles the supernovae: an implosion, leading to a more violent explosion later.

Figure 42 shows an example of the instantaneous collective potential produced by the strings in the transverse plane. The white regions correspond to the values of the potential smaller than $-5 \times 2g_N\sigma_T(\text{fm}^{-1}) \approx -400$ MeV, i.e., the chiral symmetry can be completely restored in those regions. A sufficiently strong gradient of this potential can cause quark pair production, similar to the Schwinger process in an electric field. One particle may fall into the well and another may fly away, a phenomenon analogous to Hawking radiation near the black hole.

VII. HOLOGRAPHIC EQUILIBRATION

A. Near equilibrium

The holographic equilibrium setting includes the so-called "AdS-black hole" metric, with its horizon located at the fifth coordinate $z_h = 1/\pi T$, so the gauge theory located at the $z = 0$ boundary feels the Hawking radiation temperature T.[21]

Gravity waves propagating in the AdS background metric have certain dispersion relations $\omega(\vec{k})$ with calculable real and imaginary parts; an example is shown in Fig. 43. Such quasinormal[22] modes are known for various examples of BHs for a long time; these particular ones were calculated by Kovtun and Starinets (2005). In this channel, the lowest eigenvalue, shown by larger dots, is close to the origin and describes the sound mode. For reference we mention several known terms at small k (Lublinsky and Shuryak, 2009):

$$\frac{\omega}{2\pi T} = \pm\frac{\tilde{k}}{\sqrt{3}}\left[1 + \left(\frac{1}{2} - \frac{\ln 2}{3}\right)\tilde{k}^2 - 0.088\tilde{k}^4\right] - \frac{i\tilde{k}^2}{3}\left[1 - \frac{4 - 8\ln 2 + \ln^2 2}{12}\tilde{k}^2 - 0.15\tilde{k}^4\right], \tag{48}$$

where $\tilde{k} \equiv k/2\pi T$. First, note that at small k the imaginary viscous term is very small[23] $\text{Im}\omega \sim k^2$. Second, the dispersive correction to the speed of sound (the k^2 term in the first square bracket) has positive coefficient. Thus AdS/CFT predicts that one sound wave can decay into two. Third, note that higher order corrections to viscosity are both negative. This is in contrast to some popular second-order *ad hoc* schemes such as the Israel and Stuart.

All other "nonhydrodynamical" modes have large $\text{Im}(\omega)/(2\pi T) = O(1)$. During the time of the order of

[21]Some may be confused by the known fact that Hawking radiation leads to the evaporation of black holes. Indeed, one BH, placed in an asymptotically flat background, cannot heat up an infinite universe, and it does evaporate. But the AdS metric is basically a finite box, and in this case the BH can be in a static equilibrium state with the "heated" universe.

[22]As with wave functions of the α-decaying nuclei, when energy is complex the wave function grows in space and is not normalizable, thus the name. In nuclear physics they are called quasistationary states.

[23]We used that fact in the section on the "acoustic damping" phenomenology.

$z_h \sim 1/(2\pi T)$ they all disappear since they "fall into the black hole."[24] The essence of the AdS explanation for the rapid equilibration is thus simple: any objects (nonhydro modes) become invisible as they are absorbed by the black hole. The only[25] remaining memory is their total mass, which BH transforms into the appropriate amount of Bekenstein entropy.

B. Out of equilibrium 1: The shocks

Shocks are classic examples of out-of-equilibrium phenomena. They traditionally are divided into two categories: weak and strong. In the former case the difference between matter before and after the shock is small. Weak shock can be treated hydrodynamically, e.g., using the NS approximation. Strong shocks have finite jumps in matter properties. Their profiles have large gradients: so one needs some more powerful means to solve the problem, not relying on hydrodynamics, which is just an expansion in gradients.

The reason we put this example as number one is because it is the only one which can be considered in a stationary approximation. Indeed, in the frame which moves with the velocity of the shock, its profile is time independent.

Strong shocks in the AdS/CFT setting were discussed by Shuryak (2012b). A hydrodynamical example with the Navier-Stokes profile is shown in Fig. 44. Fluxes of the total energy and momentum are tuned to be the same, from the left to the right side of the picture. One may think of it as a low-density QGP entering on the right with higher rapidity, which gets suddenly excited into a higher density QGP, floating out more slowly. It resembles a picture seen from a cockpit of a supersonic jet.

In the AdS/CFT setting, one can solve the problem from the first principles, by solving the Einstein equations. Since the setting has an extra holographic dimension z, even the static solution depends not on one but on two variables: the longitudinal coordinates x and z. Not going into detail, the surprising conclusion was that all corrections of the Navier-Stokes profile of the shock happen to be small, at a scale of a few percent, even without any apparent small parameter in the problem.

Another tool used to correct the NS solution was the so-called "resummed hydrodynamics" (Lublinsky and Shuryak, 2009): it also lead to corrections at the percent level. Unfortunately, the accuracy we had on the AdS/CFT solution was insufficient to tell whether both agree or not.

The lesson is that all higher order gradient corrections to the NS solution have a strong tendency to cancel each other.

C. Out of equilibrium 2: The falling shell

This setting of a falling matter shell was proposed by Lin and Shuryak (2008b). It is in a way complementary to the previous one: there is dependence on time t but no dependence on space x, because the shell's motion occurs along the holographic fifth direction z.

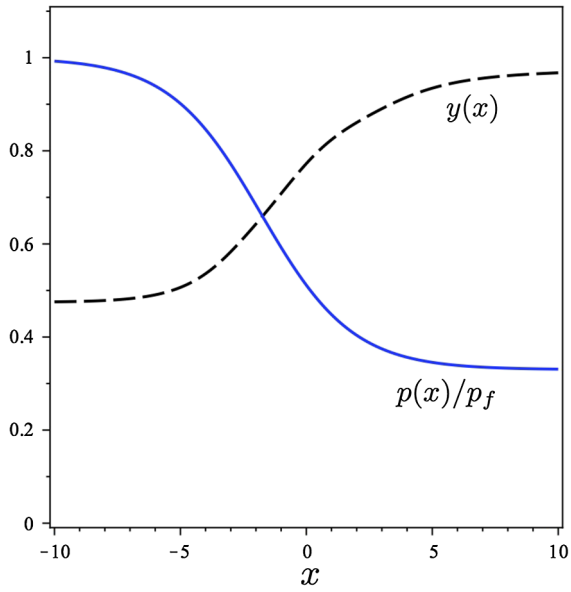

FIG. 44. Profile of a strong shock in QGP in its rest frame, according to the Navier-Stokes hydrodynamics, as a function of the coordinate normal to the shock front. The time goes right to left, so the left-hand side shows the final values of the observables, while the right-hand side shows the initial ones. The pressure, shown by the solid (blue) line, is in units of its asymptotic value: thus the curve jumps to 1 on the left side. The flow rapidity, shown by the dashed (black) line, is reduced. The process is thus a rapid formation of hotter denser QGP by an influx of a cooler and more dilute one. From Shuryak, 2012b.

The physical meaning of this motion is as follows. First, recall that the fifth coordinate $z = 1/r$ corresponds to a momentum "scale." Small values near the boundary (large r) correspond to the UV end of the scales, while large z, small r correspond to the IR or small momenta. Since everything happens much quicker in the UV as compared to the IR, the equilibration process naturally proceeds from UV to IR, also known as "top-down" equilibration. The gravity force in the AdS is directed accordingly.

One can imagine that this process can in some sense be reduced to a thin "equilibration shock wave," propagating in the z direction. The key idea (Lin and Shuryak, 2008b) was that this shock can be thought of as certain external objects, a shell or an elastic membrane, falling under its own weight [see Fig. 45(a)]. If this is the case, the total energy of the membrane is conserved (potential energy goes into kinetic). The consequences of this are very important: while the shell is falling toward the AdS center, the metric, both above and below the membrane, is actually time independent, as it depends only on its total mass. As such there is no need to solve the Einstein equations.[26] In the case of an extreme black hole at the AdS center [the dot (blue) in Fig. 45(a)] the solution consists of (i) a thermal Schwarzschild-AdS metric above the shell and (ii) "empty vacuum" or the AdS_5 solution below it.

[24]It is amusing to note that a puzzling process of QGP equilibration is, in the AdS/CFT setting, analogous to the first problem in physics, the Galilean stones falling in gravity field.

[25]Plus all conserved charges, if they are there.

[26]It is instructive to recall Newton's proof that a massive sphere has the outside field the same as a point mass, and that there is no gravity inside the sphere. This is also true if the sphere is falling. It also remains true in general relativity.

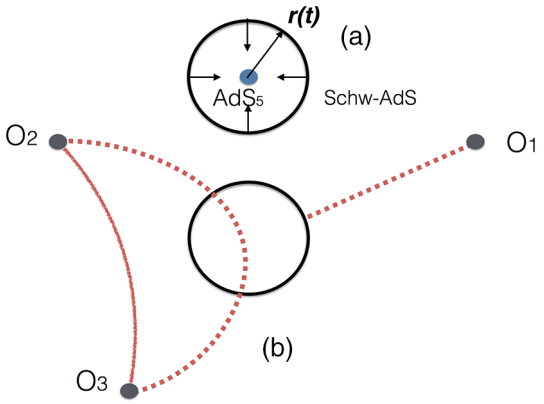

FIG. 45. (a) A sketch of a falling shell geometry. Its radius is $r(t) = 1/z(t)$ used in the text. (b) Single-point observer O_1 and the two-point observers O_2 and O_3 near a collapsing shell shown by a circle.

The only equations that need to be solved are those describing motion of the shell itself, $r(t) = 1/z(t)$. It is not so trivial to derive, since the coordinates used below and above the shell are discontinuous. Fortunately, a thin shell collapse was already solved in the general relativity: in the literature it is the so-called "Israel junction" condition. The shell equation of motion in time of the distant observer t is given by

$$\frac{dz}{dt} = \frac{\dot{z}}{\dot{t}} = \frac{f\sqrt{(\kappa_5^2 p/6)^2 + (3/2\kappa_5^2 p)^2(1-f)^2 - (1+f)/2}}{(\kappa_5^2 p/6) + (3/2\kappa_5^2 p)(1-f)}, \tag{49}$$

where $\kappa_5^2 p$ is the product of the five-dimensional gravity constant and the shell elastic constant, and $f = 1 - z^4/z_h^4$ is the standard BH function of the thermal-AdS background.

The shell starts falling with zero velocity from a certain height and then gets accelerated to nearly the speed of light. Finally, near the horizon position $z \to z_h$ there appears a "braking phenomenon": the shell slows down to velocity zero. This braking is a standard feature stemming from the use of the distant observer time, familiar from the Schwarzschild metric.

After a solution is found, one can calculate what different observers at the boundary, that is, in the gauge theory, will see. In particular, one may ask if or how such an observer can tell a static black hole (the thermal state with stationary horizon) from that with a falling shell?

A "one-point observer" O_1 [Fig. 45(b)] would simply see a stress tensor perturbation induced a gravitational propagator indicated by the dashed red line. Since the metric above the shell is thermal AdS, such an observer will see the time-independent temperature, pressure, and energy density, corresponding to the static final equilibrium. Yet more sophisticated "two-point observers" O_2 and O_3 can measure certain correlation functions of the stress tensors. They will see contributions both from gravity waves propagating along the line shown by the solid line above the shell, that is in the thermal metrics, and from waves propagating along the path shown by the dashed line which penetrate below the shell: those would notice deviations from equilibrium. Solving for various two-point functions in the background with a falling shell or membrane we found these deviations. They happen to be oscillating as a function of the wave frequency. This observation, first puzzling, is explained (Lin and Shuryak, 2008b) by finite "echo" times due to a signal reflected from the shell. Thus, one can experimentally observe an echo, coming from the fifth (nonexisting) dimension.

For further discussion of the scenarios of top-down equilibration, with infalling scalar fields, etc.; see, e.g., Balasubramanian *et al.* (2011).

D. Out of equilibrium 3: Anisotropic plasma

Our next example, due to Chesler and Yaffe (2014), is a setting in which one starts with some anisotropic but homogeneous metric and follows its relaxation to equilibrium. The metric is diagonal, with time-dependent but space-independent components, and the resulting Einstein equations are solved numerically.

Rapid relaxation to an equilibrium thermal-AdS solution is observed. A number of initial states can be compared, selected with the same equilibrium energy density (or horizon, or T) at late time. While at early time the momentum asymmetry can be very large (say, an order of magnitude) it becomes exponentially small with time. Any deviations from equilibrium are strongly redshifted as they approach the horizon.

There are no hydrodynamical modes since the setting is homogeneous. The lowest mode has frequency $\omega = (2.74 + i3.12)\pi T$. Thus, the strongly coupled QGP has the "isotropization time" as short as

$$\tau_{\text{isotropization}} \sim \frac{0.1}{T}. \tag{50}$$

E. Out of equilibrium 4: Rapidity-independent collisions

The picture of "debris" created in the bulk after a high energy collision, forming a small black hole falling toward the AdS center (Shuryak, Sin, and Zahed, 2007), related the exploding and cooling fireball in the real world to a hologram of the black hole horizon moving away from the boundary. The specific solution discussed in that paper was spherically symmetric and thus more appropriate for cosmology than for heavy ion applications.

A more appropriate setting with "falling" horizons, corresponding to the Bjorken hydro solution (see Appendix B), was developed by Janik and Peschanski (2006). In this case, the horizon is rapidity independent and has a time-dependent location $z_h(\tau)$. At late time the solution of the Einstein equation reproduces a Bjorken hydro, because at $\tau \to \infty$

$$z_h(\tau) = 1/\pi T(\tau) \sim \tau^{1/3}. \tag{51}$$

The variable $w = T\tau$ has the meaning of the macroscale-to-microscale ratio. At late time $\tau \to \infty$ it grows, indicating that the system becomes more macroscopic and hydrodynamics becomes more accurate. The question is when exactly does the hydrodynamical description become valid and with what accuracy.

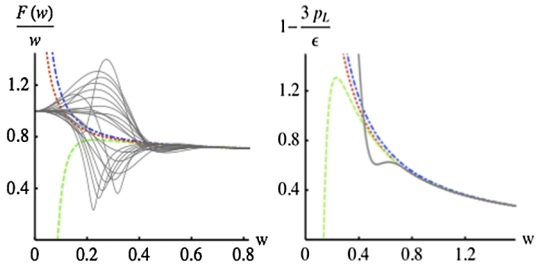

FIG. 46. (Left) The temperature evolution combination $d\log(w)/d\log\tau$ for different initial conditions (thin black curves) converging into a universal function of $w=T\tau$, compared to the hydrodynamical prediction. (Right) The pressure anisotropy for one of the evolutions compared to first- (NS), second-, and third-order hydrodynamics. From Heller, Janik, and Witaszczyk, 2012.

Figure 46 shows the time evolutions of many initial states, all approaching the same hydrodynamical solution. Figure 46 (left) shows that this happens via convergence to certain universal functions of the variable $w=\tau T$ defined by

$$\frac{dw}{d\ln\tau}=F(w). \tag{52}$$

Existence of such universality is the essence of the "resummed hydro" (Lublinsky and Shuryak, 2009). Depending on the required degree of accuracy, one may assign a specific initial value of w at which the "hydrodynamics starts," in the range $w_i=0.4$–0.6. The plot on the right demonstrates that at such time the anisotropy is still large and viscosity is important.

The lesson from this work can be better explained by comparing its result to naive expectations that hydrodynamics starts when macrotimes and microtimes are the same $w=\tau T>1$, and that the accuracy of hydrodynamics should be bad $O(100\%)$. Calculations show instead that at a twice smaller time $w\sim 1/2$ the accuracy of (the lowest-order Navier-Stokes) hydro suddenly becomes quite good, to a few percent.

Why is it so? While gradients are not yet small at that time, the combined effect of all of them is. The Lublinsky-Shuryak resummation provides an answer: the higher-gradient series has alternating signs, it can be Padé resummed *à la* a geometric series to a decreasing function.

The issue has its practical aspect, related to one of the first observations made at the first LHC PbPb run. It was found that the (charged hadron) multiplicity in PbPb collisions grows with energy a bit more rapidly than in pp:

$$\frac{dN^{\rm PbPb}}{dy}(y=0,s)\sim s^{0.15},\qquad \frac{dN^{pp}}{dy}(y=0,s)\sim s^{0.11}. \tag{53}$$

From the RHIC energy ($E=0.2$ TeV) to the LHC, the double ratio

$$\frac{\frac{dN}{d\eta}|_{\rm PbPb,LHC}/\frac{dN}{d\eta}|_{pp,\rm LHC}}{\frac{dN}{d\eta}|_{\rm AuAu,RHIC}/\frac{dN}{d\eta}|_{pp,\rm RHIC}}=1.23 \tag{54}$$

shows a noticeable change with the energy, which calls for an explanation.

A simple form for the function $F(w)$ was proposed by Lublinsky and Shuryak (2011). If known, one can calculate the entropy produced from the time w_i on: it turns out to be about 30%. Furthermore, we get the following expression for the contribution to this double ratio:

$$\approx 1+\frac{3[\bar\eta({\rm LHC})-\bar\eta({\rm RHIC})]}{2w_i+3\bar\eta({\rm RHIC})}$$

and show that the observed growth can be naturally explained by the viscosity entropy production, from RHIC to the LHC, predicted by a number of phenomenological models.

VIII. COLLISIONS IN HOLOGRAPHY

A. "Trapped surfaces" and the entropy production

The simplest geometry to consider is the wall-on-wall collisions, in which there is no dependence on the two transverse coordinates, and only the remaining three (time, longitudinal (rapidity), and the holographic direction) remain in play. Needless to say, it is a formidable problem, solved by Chesler and Yaffe (2014) via "nesting" of Einstein equations.

Collisions of finite size objects are even more difficult to solve, but those historically bring a discussion of the important issues of trapped surface formation and entropy production. It was pioneered by Gubser, Pufu, and Yarom (2008) who considered head-on (zero impact parameter) collisions of point black holes. The setting is shown in Fig. 47(a).

"Trapped surface" is a technical substitute for the horizon and its appearance in the collision basically means that there exists a black hole inside it. Classically, all information trapped inside it cannot be observed from outside, and lost information is entropy. For known static black hole solutions its area does give the black hole Bekenstein entropy.

Locating this surface allows one to limit the produced entropy from below, by simply calculating its area. The reason why this entropy estimate is from below is because the trapped surface area is calculated at the early time $t=0$ of the collision, not at its end. No particle can get out from a trapped surface, but some can get into it during the system's evolution, increasing the black hole mass and thus its entropy.

Gubser *et al.* denoted the distance separating a colliding black hole from the boundary by L; we will discuss its physical meaning later. Naively, central collisions have only axial O(2) symmetry in the transverse plane $x_\perp$, but using global AdS coordinates they found a higher O(3) symmetry of the problem, which becomes manifest if a coordinate

$$q=\frac{\vec{x}_\perp^2+(z-L)^2}{4zL} \tag{55}$$

is used. It was shown that the 3D trapped surface C at the collision moment is in this coordinate a 3-sphere, with some radius q_c. If q_c is determined, the relation between the CM collision energy and Bekenstein entropy can be calculated.

For large q_c these are given by

$$E\approx\frac{4L^2q_c^3}{G_5},\qquad S\approx\frac{4\pi L^3q_c^2}{G_5}, \tag{56}$$

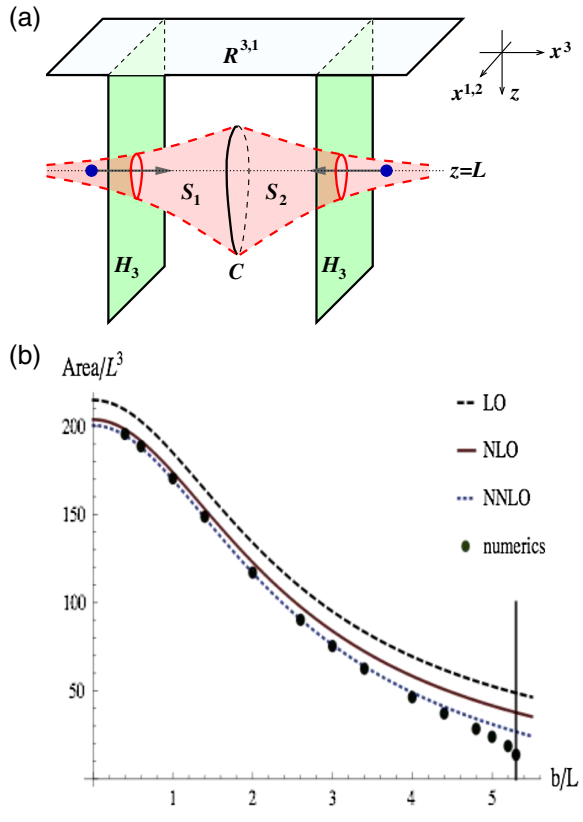

FIG. 47. (a) A projection of the marginally trapped surface onto a fixed time slice of the AdS geometry. From Gubser, Pufu, and Yarom, 2008. (b) The area of the trapped surface vs the impact parameter, with the comparison of the numerical studies (Lin and Shuryak, 2009) shown by points and analytic curves. The vertical line shows the location of the critical impact parameter b_c beyond which there is no trapped surface. From Gubser, Pufu, and Yarom, 2009.

from which by eliminating q_c one finds the main conclusion, the entropy grows with the collision energy as

$$S \sim E^{2/3} L^{5/3}. \tag{57}$$

Note that this power is in general $(d-3)/(d-2)$, so it is directly related to the $d=5$, the dimension of AdS space. Note also that it is different from the 1950s prediction of Fermi and Landau who predicted $S \sim E^{1/2}$ as well as from the data, which according to Eq. (53) indicates the power of about 0.30.

Let us now return to the meaning of the parameter L. Gubser *et al.* related the "depth" of the colliding objects with the nuclear size $L \sim 1/R_A$ which cannot depend on the energy. An alternative idea suggested by Lin and Shuryak (2009) ascribed L to the (inverse) "saturation scale," the typical parton's momenta in the wave function of the colliding objects. In this case it is related to the collision energy by

$$L \approx \frac{1}{Q_s(E)} \sim E^{-\alpha}, \tag{58}$$

where $\alpha \approx 1/4$ is an empirical index of the PDFs. It is especially clear if one considers wall-on-wall collisions, in which the nuclear size R_A goes to infinity, while Q_s characterize the material the wall is made of, and remains fixed. L is not of the $O(10\text{ fm})$ scale, but rather 2 orders of magnitude smaller $O(0.1\text{ fm})$. Furthermore, it is expected to decrease with the energy $L \sim 1/Q_s(E)$. Including this modifies Eq. (57) to

$$S \sim E^{(2/3)-(5/3)\alpha} \sim E^{0.25}, \tag{59}$$

which is in reasonable agreement with the observed multiplicity.

The generalization of this theory to noncentral collisions (Lin and Shuryak, 2009) leads to the results shown in Fig. 47(b). The figure shows the dependence of the trapped surface area on the impact parameter. Specifically, numerical results from Lin and Shuryak (2009) (points) are compared with the analytic series of curves (Gubser, Pufu, and Yarom, 2009), which are in excellent agreement.

From the gravity point of view the qualitative trend shown is clear: two colliding objects may merge into a common black hole provided only that the impact parameter is less than some critical value $b_c(E)$, depending on the collision energy. Indeed, with b rising, the trapped energy decreases while the total angular momentum increases, so at some point the Kerr parameter exceeds 1 and thus no black hole can be formed. Interestingly, the calculation shows that it happens with a finite jump—a first-order transition[27] as a function of the impact parameter.

Just a bit below the critical value of the impact parameter, the trapped surface and black hole exist, and nothing indicates that at larger b none is formed. At $b < b_c$ the creation of QGP fireball happens, while for peripheral $b > b_c$ collisions the system remains in the hadronic phase. A jump in entropy as a function of the impact parameter is rather surprising, but in fact the experimental multiplicity-per-participant plots do indeed show rapid change between non-QGP small systems and QGP-based not-too-peripheral AA collisions. It would be interesting to compare it with all available information on small systems undergoing transition to an explosive regime.

Lin and Shuryak (2009) also pointed out that the simplest geometry of the trapped surface would be that for a wall-wall collision, in which there is no dependence on transverse coordinates x^2 and x^3. Thus a sphere becomes just two points in z, above and below the colliding bulk objects. We elaborated on this, considering the collision of two infinite walls made of a material with different "saturation scales" (Lin and Shuryak, 2011), and studied conditions for trapped surface formation.

B. From holographic to QCD strings

AdS/CFT is a duality with a string theory, so fundamental strings are naturally present in the bulk. Already the first calculation made in Maldacena's original paper, the "modified Coulomb law," was based on the evaluation of the shape and

[27]The first-order transition stems from the large N_c approximation: this conclusion may perhaps be modified at finite N_c. Furthermore, the problem of a trapped surface in quantum gravity is much too complicated, and not studied so far.

total energy of a "pending string," sourced by "quarks" on the boundary.

Extension to nonstatic strings was done in Lin and Shuryak (2008a, 2008c). First the shape of a falling string with ends moving away from each other with velocities $\pm v$ was derived, and second its hologram (stress tensor distribution) at the boundary was calculated. This study can be thought of as a strongly coupled version of Ampere's law, with two currents rather than charges, or as a strongly coupled version of e^+e^- annihilation into two quarks. The hologram showed a near-spherical explosion, historically an early indication that there are no jets at strong coupling.

These works used the setting associated with conformal gauge theory: in AdS_5 string falling continues forever. This is of course not what we observe in the real world, in which there is confinement and there are jets. Modern strong-coupling models moved into what is collectively known as AdS/QCD; for a review, see, e.g., Gursoy and Kiritsis (2008) and Gursoy, Kiritsis, and Nitti (2008). In contrast with the original AdS/CFT, the background metric is not conformally invariant and incorporates both confinement in the IR and the asymptotic freedom in the UV. These models use an additional scalar ("dilaton") field, which is also given a phenomenological potential depending on the fifth coordinate.

In such settings, the bulk strings can "levitate" at some position z_*, at which the downward gravity force is compensated by the uplifting dilaton gradient. The hologram of such a levitating string at the boundary is the QCD string. Its tension, width, and stress tensor distribution are all calculable. The potential between point charges is still given by a pending string. In the AdS/QCD background its energy changes from the Coulombic potential at small r to the linear potential at large r, showing confinement. Furthermore, allowing fundamental fermions in the bulk, via certain brane constructions, and including their backreaction in a consistent manner, one can get the so-called Veneziano limit of QCD (N_c, $N_f \to \infty$, $x = N_f/N_c = \text{fixed}$) (Arean *et al.*, 2013).

Since in the UV these models possess a weak coupling regime, one can also model perturbative GLASMA, by putting a certain density of color sources on the two planes, departing from each other. In such a setting there is a smooth transition between two alternatives descriptions of the initial state we discussed previously—from the perturbative GLASMA to a spaghetti made of the QCD strings. When time τ is small [Fig. 48(a)], strings are in the UV domain (at small distance $z \sim 1/Q_s$ from the boundary) and their hologram is Coulombic or GLASMA-like. When strings fall farther and reach the "levitation point" z^* they start oscillating around it (Iatrakis, Ramamurti, and Shuryak, 2015), Fig. 48(b). This is similar to oscillations discussed by Florkowski (see Sec. VI.D) without AdS/QCD.

AdS/QCD predictions for string-string interactions were derived by Iatrakis, Ramamurti, and Shuryak (2015). We already discussed this issue in the QCD context, concluding that its long-range attraction is dominated by the σ meson exchanges (just like between nucleons, in nuclear forces). The question is whether it is also the case in the AdS/QCD. AdS/QCD has very few fields in the bulk—gravitons, dilatons, and (quark-related) "tachyons." Their quantization along the fifth coordinate generate towers of four-dimensional hadronic states. Hadronic masses are just quantized fifth momentum. From this approach one can calculate not only the masses but also the wave functions in a scale space, as well as mixing between the fields. A specific issue studied by Iatrakis, Ramamurti, and Shuryak (2015) is the mechanism of hadronic flavorless scalars, which includes the σ and others. Without any changes in the setting of AdS/QCD we found a good description of mixing patterns of the scalars, which puzzled spectroscopists for decades. Since strings are gluonic objects and σ interacts strongly with quarks, the understanding of such mixing is crucial for obtaining realistic string-string forces.

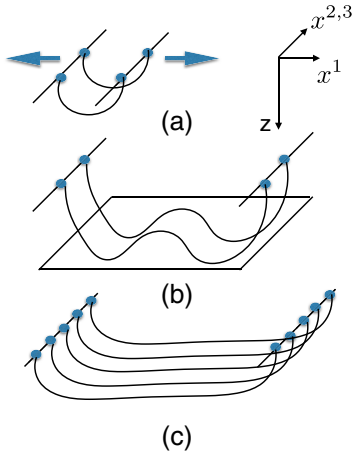

FIG. 48. (a) An early time snapshot of a pair of strings created after one color exchange. The coordinates are explained on the right: the colliding objects move with a speed of light away from each other and strings are stretched. (b) Later time snapshot, in AdS/QCD background. After strings reach the "levitation surface, " shown by a rectangular shape, they start to oscillate around it. (c) In the case of high density of many strings they can be approximated by a continuous membrane.

Note that so far there is no temperature or entropy in the problem: the dynamics is given by classical mechanics of strings moving in certain backgrounds. If the number (or density) of strings becomes high enough [Fig. 48(c)], one should include the backreaction of their gravity and dilaton field, or even include mutual attraction of strings. Such an AdS/QCD version of multistring dynamics (Iatrakis, Ramamurti, and Shuryak, 2015) is the holographic version of spaghetti collapse, discussed in Sec. VI.D following Kalaydzhyan and Shuryak (2014a). The bulk strings, if too many, produce a gravitational collapse.

C. Holographic Pomeron

The description of hadronic cross sections and elastic amplitudes using Reggeons and Pomerons originates from the phenomenology as well as from the Veneziano amplitudes. While originally derived following the resonance duality in different channels, a nearly forgotten pre-QCD ideology of the 1960s, these expressions were historically important, as they gave us the first hints of the existence of QCD strings.

The Pomeron is an effective object corresponding to the "leading" Regge trajectory $\alpha(t)$ which dominates the high

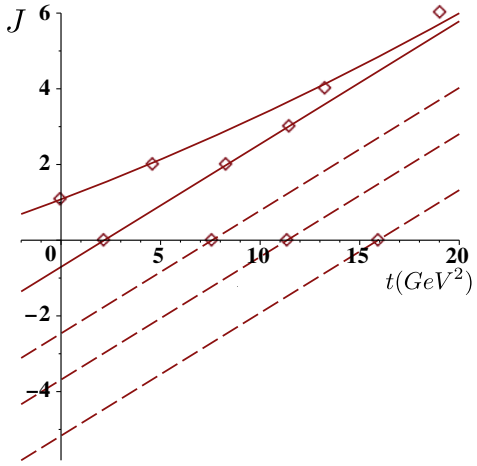

FIG. 49. Glueball masses calculated on the lattice (diamonds) organized in Regge trajectories (lines). From Shuryak and Zahed, 2014.

energy asymptotics of the hadron-hadron cross sections. Figure 49 is a version of the Regge plot (angular momentum J versus the mass squared $m^2 = t$) for glueballs. The Pomeron corresponds to scattering and thus has small nonphysical mass $t < 0$ and a noninteger J slightly above 1: the trajectory $\alpha(t)$ has of course physical states as well. It enters the elastic cross section in a form

$$\frac{d\sigma}{dt} \approx \left(\frac{s}{s_0}\right)^{\alpha(t)-1} \approx e^{\ln(s)\{[\alpha(0)-1]+\alpha'(0)t\}}. \tag{60}$$

The two main parameters of the Pomeron have a very different origin. The intercept $\alpha(0)$ is a dimensionless index, describing the power with which the total cross section rises. A perturbative description of the Pomeron by the Balitsky-Fadin-Kuraev-Lipatov (BFKL) model (Kuraev, Lipatov, and Fadin, 1977) provides a perturbative $O(\alpha_s)$ value for it. The slope[28] $\alpha'(t = 0)$ has dimension $[\text{mass}^{-2}]$ and is nonperturbative. Its value is related to the string tension $\alpha' \approx 1/(2\pi\sigma)$, roughly twice that of the slope observed in meson and baryon Reggeons, related to open strings.

The Pomeron phenomenology includes elements of perturbative and nonperturbative physics. To make it less confusing, let us consider scattering as a function of the impact parameter b. For small color dipoles and small b the amplitude should be perturbative, due to gluon exchanges. At large $b \sim 1$ fm it should be nonperturbative: the model description of it is given in terms of a (double) string exchange.

The holographic models collectively known as AdS/QCD combine a strongly coupled regime in the IR (large values of the fifth coordinate z) with a weakly coupled regime in the UV (Gursoy and Kiritsis, 2008). The topic of this section, the "holographic Pomeron," includes in fact two different approaches to the Pomeron, which we discuss subsequently.

[28]The "string scale" in the fundamental string theory is traditionally called α' still, as a historic remnant of QCD phenomenology left in it.

The first idea is to approximate the Pomeron properties via analytic continuations of the bulk gravitons. It may look exotic, but note that Regge trajectories are natural consequence of AdS/QCD models, and that the closest state to the Pomeron along the trajectory is a $J^{PC} = 2^{++}$ glueball, rather well described by these models. Recent application of the Hilbert-Einstein action to the Pomeron-Pomeron 2^{++} glueball triple vertex successfully described the double diffractive production data which were puzzling for a long time (Iatrakis, Ramamurti, and Shuryak, 2016). Even more recent are successful applications of the tensor Pomeron for the description of the RHIC polarized pp scattering; see Ewerz *et al.* (2016) and subsequent works.

We will however follow in more detail the second idea by Basar *et al.* (2012) and Stoffers and Zahed (2013), providing derivation of the Pomeron-induced scattering amplitude in a AdS/QCD, including both soft and hard regimes, and thus providing interpolation between the "stringy" and the BFKL limits of the Pomeron. The semiclassical derivation of the Pomeron amplitude is given in terms of closed string production, similar to Schwinger pair production in an electric field. The string world volume has the shape depicted in Fig. 50(a): it is a "tube" connecting two flat strips, the world volume of propagating color dipoles. In Fig. 50(b) we sketch a Pomeron in a collision of two nucleons, consisting of three quarks and three strings, joined into a string junction. In this case a Pomeron tube "punctures" only one of the three surfaces. This produces one "wounded quark," as we

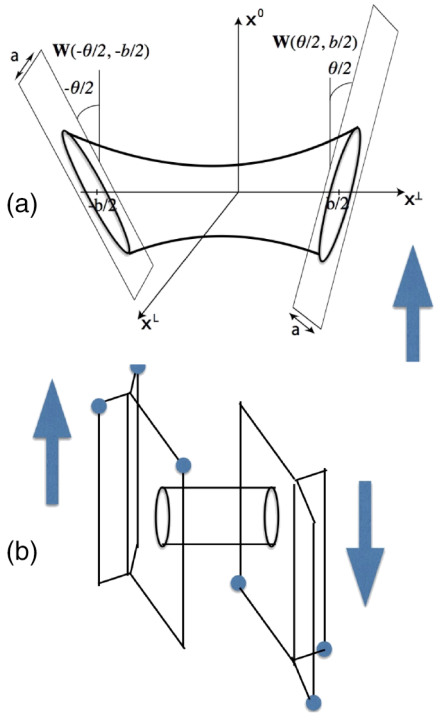

FIG. 50. (a) Dipole-dipole scattering configuration in Euclidean space. The dipoles have size a and are b apart. The dipoles are tilted by $\pm\theta/2$ (Euclidean rapidity) in the longitudinal x_0-x_L plane. (b) A sketch illustrating Pomeron exchange for baryon-baryon scattering: only one pair of quarks become wounded quarks.

discussed in connection to Tannenbaum's description of fluctuations.

Direct semiclassical derivation of the scattering amplitude (Basar *et al.*, 2012) is based on the Nambu-Goto action (the tube's area). Figure 50(a) indicates a Euclidean setting in which a difference in rapidity is represented by twisted angle θ between the propagation vectors of the two dipoles. Figure 50(b) illustrates a baryon-baryon scattering, in which the Pomeron tube can be connected to any of the available dipoles, explaining the concept of the wounded quarks mentioned in Sec. IV.B. More than one Pomeron means more tubes, perhaps connecting other quarks.

The classical action of this configuration provides the α' term, while the intercept $\alpha(0)-1$ is in this approach generated by the next-order (one-loop) corrections due to string vibrations. The elastic amplitude squared can be cut in half by the unitarity cut. The corresponding tube configurations, when cut longitudinally, provide two strings of certain length and shape. Those are two physical strings that jump from under the barrier and appear in the Minkowskian world. They should be used as the initial conditions for Minkowskian real-time evolution.

The same expression for the scattering amplitude has an alternative derivation from the string diffusion equation, in the 5D bulk. The fifth coordinate in it has a meaning of the dipole size parameter, so motion in this coordinate is dual to the DGLAP evolution. Scattering pp data as well as deep-inelastic ep (DESY data) are well reproduced by this model (Basar *et al.*, 2012).

It was further argued by Shuryak and Zahed (2014) that because the tube has a periodic variable resembling the Matsubara time, its fluctuations take the thermal form. The appearance of an effective temperature and entropy was a new aspect to the Pomeron problem: but once recognized the analogy to thermal strings can be exploited. It was argued that two known regimes of the Pomeron we already mentioned (long string exchange at large impact parameter b and perturbative gluon exchange at small b) should be joined by a third distinct regime, in which strings are highly excited due to the Hagedorn phenomenon. This is what happens near critical temperature in gluodynamics. The third regime is known as the mixed phase, between the confined and the deconfined phases.

Figure 51(a) shows the elastic scattering profile, defined as the Fourier-Bessel transform of the imaginary part of the elastic amplitude. Shown by the dashed line is the prediction of the model with the basic first-order transition occurring at fixed impact parameter, and a fixed size of the colliding dipoles. The circles show a more realistic model, including averaging over a certain distribution over the dipole sizes. The solid line is the empirical fit to the elastic scattering amplitude measured at the LHC. The difference between this empirical curve and the model (circles) is relatively small in Fig. 51(a), although it becomes more noticeable in Fig. 51(b) showing its Fourier-Bessel transform as a function of the momentum transfer Q. As one can see, beyond the minimum good agreement is lost. Apparently the "transition edge" in b predicted by the first-order transition model is sharper than the corresponding change in the experimental data.

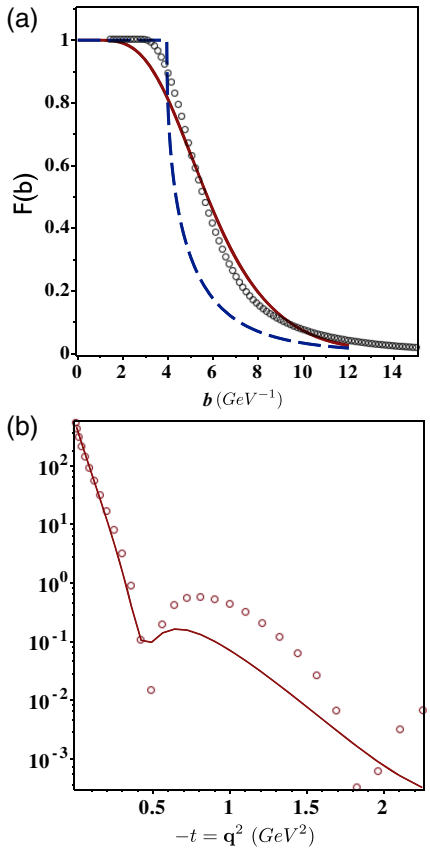

FIG. 51. (a) The solid line is the empirical LHC data parametrization. The dashed line is the shape corresponding to the "excited string" approximation for fixed sizes of the dipoles, while the circles correspond to the profile averaged over the fluctuating dipoles. (b) The corresponding elastic amplitude (the absolute value squared of the profile Bessel transform) as a function of the momentum transfer. Model prediction agrees with parametrization well at small t, up to the dip.

In summary, a decades-old Pomeron amplitude was finally derived in a stringy AdS/QCD setting. Surprisingly, the Pomeron is related to thermally excited strings, with an effective temperature proportional to the inverse impact parameter. Rapid "phase transition" in scattering amplitude at some impact parameter, from a nearly black disk to light gray, corresponds to the deconfinement transition. An intermediate regime corresponds to highly excited strings, known as the "stringball" regime. So far only the elastic amplitude has been calculated, due to the tunneling (Euclidean) stage of the evolution. To work out subsequent evolution of the system, extending this theory to a description of inelastic collisions remains to be done.

D. Collisions at ultrahigh energies

Discussions about Pomeron regimes ultimately drive us to the old question: what happens at the ultrahigh energies, well above those of the LHC? The highest observed energies, by Pierre Auger Observatory and similar cosmic ray detectors, go until

$$E_{\rm lab} \lesssim E_{\rm max} \sim 10^{20}\ {\rm eV}, \tag{61}$$

where they are limited by the so-called Greisen-Zatsepin-Kuzmin (GZK) bound, due to their inelastic interaction with the cosmic microwave background (CMB).

For comparison with the LHC, let us convert the laboratory energy into the center of mass frame and use a standard Mandelstam invariant. Assuming it is the pp collision, one finds

$$\sqrt{s_{\max}} = (2E_{\max}m_p)^{1/2} \approx 450 \text{ TeV}. \tag{62}$$

While significantly higher than the current LHC pp energy $\sqrt{s_{\rm LHC}} = 8$ TeV, the jump to it from the LHC is comparable to that from the Tevatron $\sqrt{s} = 1$ TeV or RHIC $\sqrt{s_{\rm RHIC}} = 0.5$ TeV. In view of smooth small-power s dependence of many observables, the extrapolation to the LHC worked relatively well, and one might think that any further extrapolations may work as well. However, smooth extrapolations using standard event generators do not reproduce experimental data from the Pierre Auger Collaboration.

Models aimed to resolved the contradiction were proposed. For example, Farrar and Allen (2013) suggested an exotic freezeout without chiral symmetry breaking, without multipion production. According to simulations presented in this work, if mostly nucleons are produced, the Pierre Auger data are explained.

A more modest (but still significant) change between the LHC and the so-called ultrahigh energies was proposed by Kalaydzhyan and Shuryak (2014b): at such energies $\sqrt{s_{\max}}$ even minimally biased pp collisions should be in the "explosive regime" as discussed previously for central pA or rare ($P \sim 10^{-6}$) high-multiplicity pp collisions at the LHC. This is simply caused by an increase in particle density dN/dy, by about a factor of 3. Another generic reason for this regime change is that both primary collisions and subsequent cascades in the Earth's atmosphere have as targets light N, O nuclei. Furthermore, the projectiles themselves are also most likely some mixture of nuclei, perhaps up to Fe.

Taking into account a large pp cross section at ultrahigh energies, ~150 mb, one finds that its typical impact parameters $b \approx 2$ fm. Thus the range of the interaction in the transverse plane is comparable to the radius of the light nuclei (oxygen $R_{\rm O} \approx 3$ fm) and therefore even in the pO collisions most of its 16 nucleons would become collision "participants." For light-light AA collisions such as OO the number of participants changes from 32 (central) to zero. Accidentally, the average number of participants is comparable to the average number of participant nucleons $\langle N_p \rangle \approx 16$ in central pPb collisions at the LHC.

IX. ELECTROMAGNETIC PROBES

A. Brief summary

The sources of the dileptons are split into the following categories:

(i) Instantaneous $\bar{q}q$ annihilation, known as the Drell-Yan partonic process.

(ii) $\bar{q}q$ annihilation at the preequilibrium stage, after the nuclei pass each other.

(iii) $\bar{q}q$ annihilation in the equilibrated sQGP.

(iv) Meson-meson annihilation at the (kinetically but not chemically) equilibrated hadronic stage.

(v) The so-called "cocktail" contribution, consisting of leptonic decays of unstable secondaries. Electromagnetic and electroweak decays all occur long after freezeout, so this component can be calculated from spectra and statistically subtracted. We do not discuss this component any longer.

The corresponding windows in the dilepton mass are, respectively, $M > 4$ GeV for (i), $1 < M < 3$ GeV for (iii), and $M < 1$ GeV for (iv). So the early stage dileptons mostly fall into the 3–4 GeV window.[29]

Among the CERN SPS dilepton experiments the most successful was NA60, which quantified with the largest statistics the following phenomena:

(a) Large enhancement over the cocktail at dilepton masses $M < m_\rho$. The resulting spectral density of the electromagnetic current in QGP was quantitatively measured.

(b) By plotting the $p_\perp$ slope as a function of the dilepton mass M it was demonstrated that light ($M < 1$ GeV) dileptons are produced when flow is fully developed, or near freezeout.

(c) The intermediate mass dileptons (IMD) in the mass range $1 < M < 3$ GeV are entirely different: they produced early, when the flow is still absent.

(d) The IMD are mostly "prompt," coming from QGP thermal radiation and not from the charm decays (as was suggested originally).

At RHIC:

(a) Low mass dileptons are also enhanced over the cocktail. The exact magnitude of it is still somewhat disputed between STAR and PHENIX, for the most central bin.

(b) IMD are well measured but the contribution from charm and bottom decays remains unknown. STAR expects to use $e-\mu$ correlations based on a new muon subsystem now in place.

(c) Direct photons have spectra consistent with standard rates and hydrodynamics in shape, but not in absolute magnitude.

(d) Unexpectedly large elliptic flow v_2 of direct photons persists.

At the LHC the dilepton measurements are not yet as developed as hadronic ones. Photons measurements by ALICE include confirmation of a relatively large (and puzzling) value of the direct photon ellipticity, consistent with that originally observed by PHENIX. Figure 52 shows a comparison to theoretical models (curves): poor agreement is known as the "direct photon puzzle."

Theory of the electromagnetic observables follows tradition set up long ago (Shuryak, 1980), mostly using the production rates based on binary collisions of partons or hadrons. Now we know that a similar approach fails to give viscosity, a heavy-quark diffusion coefficient, or a jet quenching parameter $\hat{q}$. Perhaps photon and dilepton production rates are larger numerically and of more complicated nonperturbative origin. If large v_2 of the photons (measured so far with large error

[29]This window of masses also contains prominent peaks due to $J/\psi, \psi', \psi''$ decays. Those, according to the definition used, belong to the cocktail that should also be subtracted.

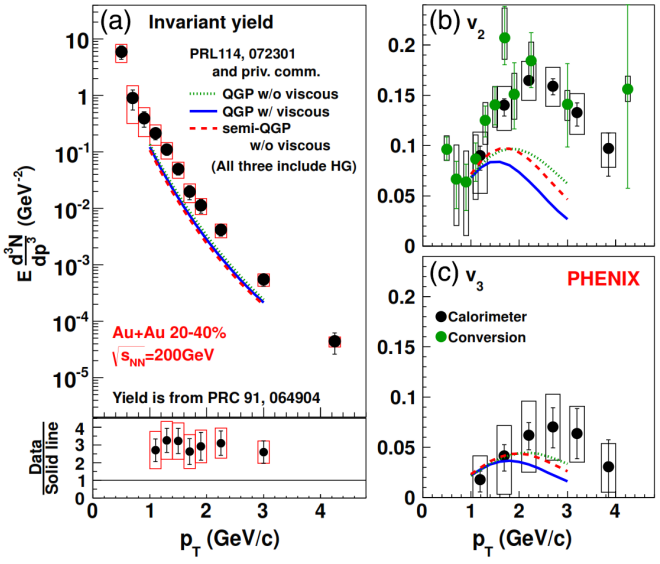

FIG. 52. Illustrations of the "direct photon puzzle." (a) The yield, (b) v_2, and (c) v_3 of the direct photons. Points are data and curves are from theory based on the hydrodynamical model, the reference indicated on the figure. From A. Drees, PHENIX presentation at QM2015.

bars) is confirmed, photons must be produced at the larger rate at the late stage of the collisions.

B. New sources of photons and dileptons: Multigluon or phonon + magnetic field

The production of photons and dileptons is tied to the presence of quarks, since gluons (and monopoles abundant near T_c) have no QED electric charge. The initial stages of the high energy collisions are believed to be dominated by gluons. Old perturbative arguments (Shuryak and Xiong, 1993) suggested that chemical equilibration via quark-antiquark pair production is relatively slow and should take place much later than thermal equilibration of the glue. This idea led to a "hot glue" scenario in which the quark-antiquark density at early stages is suppressed by powers of quark fugacity $\xi_q < 1$. The basic process of the dilepton production

$$q + \bar{q} \to \gamma^* \to l_+ + l_- \tag{63}$$

is expected to be suppressed quadratically, $\sim\xi_q^2$.

This scenario was recently challenged: higher order processes with virtual quark loops can produce electromagnetic effects as well, even without on-shell quarks. First, contrary to general expectations, the quark loop effect in GLASMA was suggested to be significant (Chiu *et al.*, 2013), enhanced due to multigluon, to virtual quark loop, to dilepton processes such as, e.g.,

$$ggg \to (\text{quark loop}) \to \gamma^* \to l_+ + l_-. \tag{64}$$

The magnitude of the correction to production rates due to these processes still needs to be quantified.

An explicit calculation of the rate of two gluon to two photon transition $gg \to \gamma\gamma$ was done by Basar, Kharzeev, and Shuryak (2014). One of the photons is assumed to be the ambient magnetic field at the time of collision, while both gluons are combined into a colorless stress tensor

$$T^\mu_\mu + \vec{B} \to (\text{quark loop}) \to \gamma^*(=\text{dileptons}). \tag{65}$$

The terminology introduced in this paper is as follows. The process in which glue appears as an average matter stress tensor $\langle T_{\mu\nu}\rangle$, producing photons (real or virtual) due to a time-dependent magnetic field, is called magneto-thermo-luminescence (MTL). The average stress tensor is nearly constant over the whole fireball, and therefore its Fourier decomposition has very small momenta $p \sim 1/R$.

Individual events, however, also possess fluctuations of the matter stress tensor $\delta T_{\mu\nu}$, with chaotic spatial distribution, and thus with the Fourier transforms with non-negligible momenta. Since those are referred to as "sounds," their interaction with the ambient magnetic field is called magneto-sono-luminescence (MSL). If observed, the MSL process tests both the amplitudes of the short-wavelength sounds and also the magnitude of the magnetic field. We already discussed many uses of sounds: we here comment only on the magnetic field. It is easy to evaluate its early values, resulting from Maxwell equations and the currents due to the spectators in peripheral collisions. Since sQGP is believed to be a good conductor, these fields are expected to create currents capturing a fraction of the field inside the plasma (Tuchin, 2013), perhaps lasting for many fm/c. Magnetic field evolution is important to know for other applications as well, e.g., chiral magnetic effects and the like.

So far luminosity and acceptance limitations have led experiments to focus on most luminous central collisions. In those, however, the ambient magnetic field is absent. Now it is perhaps time to look at dileptons in semiperipheral collisions as well. To tell the effect of the ambient magnetic field from others, RHIC considers runs of isotopes with similar A but different Z, N values.

C. Dilepton polarization and the (early time) pressure anisotropy

It is well known that when spin-1/2 particles (such as quarks) annihilate and produce lepton pairs, the cross section is not isotropic but has the following form:

$$\frac{d\sigma}{d\Omega_k} \sim (1 + a\cos^2\theta_k), \tag{66}$$

where the subscript corresponds to a momentum k of, say, the positively charged lepton and θ_k is its direction angle relative to the beam. The anisotropy parameter a in the Drell-Yan region [stage (i) in the terminology introduced at the beginning of this section] is produced by annihilation of the quark and antiquark partons, collinear to the beam. In this case, $a = 1$.

It was suggested (Shuryak, 2012a) that the parameter a can be used to control anisotropy of the early stages of the collision. In particular, if it is anisotropic so that longitudinal pressure is small relative to transverse $p_l < p_t$, the annihilating quarks should mostly move transversely to the beam, which leads to negative $a < 0$. Such a regime is expected due

to a "self-sorting" process, in which partons with different rapidities automatically become spatially separated after the collision.

For illustration one can use a simple one-parameter angular distribution of quarks over their momenta p in a form

$$W \sim \exp[-\alpha \cos^2\theta_p] \tag{67}$$

and calculate $a(\alpha)$ resulting from it. This shows that a may reach negative values as low as -0.2 at stage (ii), before it vanishes, when equilibration is over, at stages (ii)–(iv).

X. HEAVY QUARKS AND QUARKONIA AS sQGP PROBES

Physics of heavy-flavor quark or hadron production is a rather large area, to which we cannot give full justice here; for a review, see Prino and Rapp (2016). Heavy quarks provide interesting probes for matter properties, for a number of reasons. First, they are not produced by it, but by hard processes at the initial collision, which are under good theoretical control. Heavy quarks at the end are combined with light ones into heavy-flavor mesons and baryons, some identified by their decays. Leptons coming from c and b quark decays are also separately identified and studied. Thus, we know both the initial spectra, at the time of production, and the final ones.

Heavy-quark motion inside matter is described by a number of tools, such as Langevin or Fokker-Planck (FP) equations, or other kinetic approaches. There are two kinetic coefficients derived from these studies: (i) for small momenta, the diffusion coefficients $D_f(T)$, $f = c$, b and (ii) for large momenta, the (flavor dependent) quenching parameter $\hat{q}_f$. Not going into specifics of the fits of heavy-quark spectra, let us proceed directly to summary of the diffusion constant of the charm quark shown in Fig. 53. As with shear viscosity, the diffusion constant is inversely proportional to the scattering cross section. And, as with shear viscosity, the diffusion constant is expected to have a minimum at $T = T_c$.

There is a well-known but persistent puzzle associated with the diffusion constants which is worth mentioning. Perturbatively, the gluon quenching should be stronger than that of the quarks, by a factor of $9/4$, due to color Casimir operators. Contrary to this simple prediction, the data indicate about the same magnitude of the suppression, for gluonic, light quark, and even heavy quark-induced jets. The mechanism which would explain this puzzle is not yet identified.

A. Quarkonium suppression

Quarkonia, bound states of $\bar{c}c$ or $\bar{b}b$, are among the most discussed probes of matter properties. The initially discussed issue is the relative role of the real and imaginary parts of the potentials and binding them. Already the first paper on QGP signals (Shuryak, 1978) discussed an excitation process $J/\psi + g \to \bar{c}c$ in QGP, an analog of the photoeffect in atomic physics, with the conclusion that initially produced J/ψ would be partially destroyed by it. Matsui and Satz (1986) pointed out that, due to QGP screening, the real part of the potential is T dependent, and therefore quarkonia binding

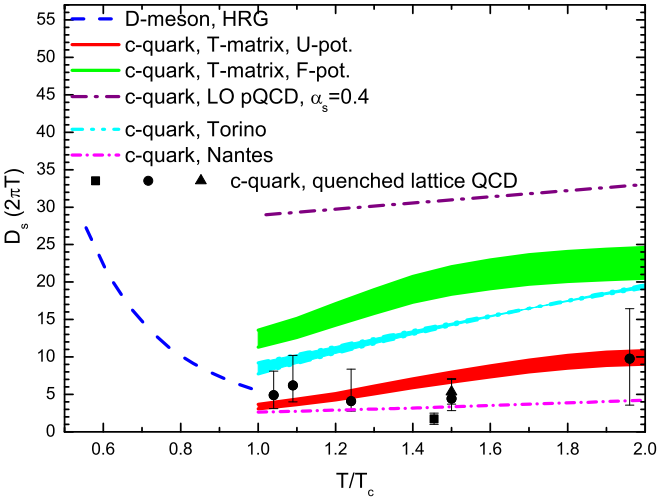

FIG. 53. The charm quark diffusion coefficients from quenched lattice QCD (circles, square, and triangle) compared to model calculations based on different elastic interactions in the QGP corresponding to the $A(p = 0)$ limit: T-matrix calculations with either free (green band) or internal energy (red band) as potential, pQCD Born calculations from HTL and pQCD matching using a reduced Debye mass and running coupling (Nantes, pink dash-dotted line) or with perturbative Debye mass and fixed coupling (Torino, cyan band), as well as schematic leading order (LO) pQCD with fixed coupling and Debye mass $mD = gT$ (purple dash-dotted line). The blue dashed line below T_c is a calculation of D-meson diffusion in hadronic matter from elastic scattering off various mesons, antibaryons, and baryons. From Prino and Rapp, 2016.

diminishes with increasing temperature. They predicted sequential melting of the charmonia states, from highest to the ground state. A quantitative study of these potentials in weakly coupled QGP was performed by Laine *et al.* (2007); for a review, see Brambilla *et al.* (2013). According to these works for $T > 300$ MeV the imaginary part of the potential exceeds the real part $\mathrm{Im}V > \mathrm{Re}V$ and becomes dominant.

However, ImV describes excitation to all other states combined, and its knowledge is not sufficient if one needs to follow the system more closely. Its usage assumes that excited states are gone forever, completely ignoring transitions back to the ground state. But, without the balance between direct and inverse reactions, one cannot formulate the concept of thermal equilibrium, which is crucial for the understanding of chemical freezeout.

Views on how QGP affects the quarkonia yield had changed in a complicated and confusing historical path. Instead of following it, we proceed from simpler to more complicated settings, namely, go thorough the following:

(0) static heavy-quark potentials,

(i) time-independent equilibrium state of charmonia,

(ii) equilibration processes and rates, and

(iii) heavy ion collisions.

(0) Static heavy-quark potentials have been extensively studied on the lattice, at variable temperatures. A sample of two-flavor QCD results is shown in Fig. 54. The vacuum potential at $T = 0$ is indicated for comparison. The free

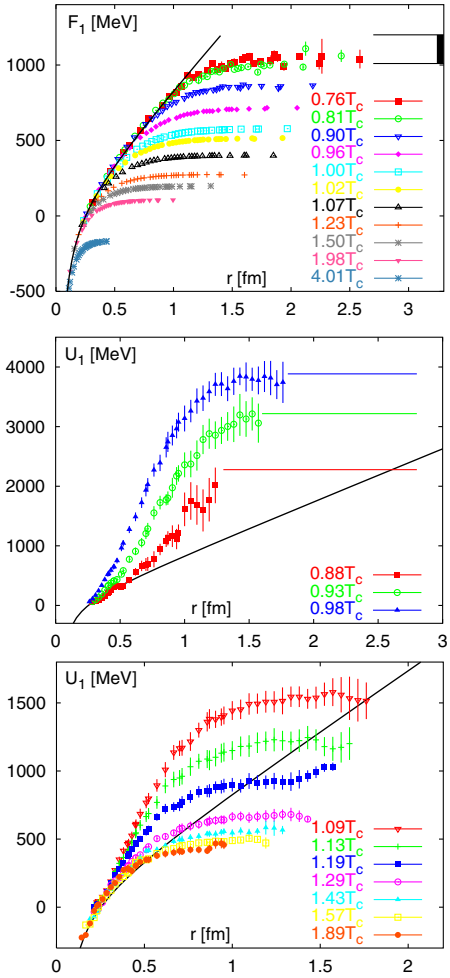

FIG. 54. Free energy singlet potentials $F_1(T,r)$ (top plot) and the potential energy $U_1(T,r)$ below and above T_c. Note that the vertical scale is different. From Kaczmarek and Zantow, 2006.

energy measured can be related to entropy and thus the internal energy $V(T,r)$, by standard thermodynamic relations

$$V(T,r) = F(T,r) - T\frac{\partial F}{\partial T} = F(T,r) + TS(T,r). \quad (68)$$

One can see from Fig. 54 that F and V are rather different. The force, also known as the string tension, has a maximum for U, reaching about 4 GeV in magnitude. What are their physical meanings? Which one should be used as the potential in the quarkonium problem?

One explanation of these lattice findings proposed by Liao and Shuryak (2010) is based on the observation that near T_c matter contains both "a dual superconductor" (Bose-condensed monopoles) and a "normal" Bose gas of monopoles. Both components create currents around the electric flux tubes, terminating the field outside them. The size and tension of the flux tube thus depend on the densities of both components, which in turn are strongly T dependent.

Its detailed discussion is out of context here, and we mention only a clarification of the entropy associated with the potential. Its generation is related with the level crossing phenomena, occurring while the separation between charges is changed. Suppose a pair of static charges are slowly moving apart in thermal medium at certain speed $v = \dot{L}$. For each fixed L, there are multiple configurations of the medium populated thermally. When L is changed, the energies of these configurations cross each other, and at each level crossing there is a certain probability to change or not to change the population of the states. These probabilities strongly depend on the speed $v = dL/dt$. For adiabatically slow motion all the level crossing processes happen with the probability of 1. The adiabatic limit is identified with the free energy $F(T,L)$.

If however the motion is very fast, transitions between crossing levels are suppressed. The internal energy $V(T,L)$, on the other hand, is different from $F(T,L)$ by subtracting the entropy term and thus it is probed in the extremely fast motion regime.

Such phenomena have multiple analogs in many other contexts in physics. The oldest example is the so-called Landau-Zener theory of electron terms in the vibrational motion of the nuclei in a diatomic molecule. Specific electron quantum states are defined at fixed L (the separation between the two nuclei) with energies $E_n(L)$, and certain levels cross each other as the value of L changes. Consider two levels with their energies given by $E_1(L) = \sigma_1 L + C_1$ and $E_2(L) = \sigma_2 L + C_2$ near the crossing point. When the two nuclei approach the crossing point adiabatically slowly $v = dL/dt \to 0$, the electrons always change from the lowest state to the lowest other state. But if dL/dt is finite, then the transition to both levels at the crossing point may happen, with certain probability. This is how initially a pure state becomes a mixture and entropy is produced. Landau and Zener gave the probability to remain in the original state at small velocity v in the following form:

$$P = \exp\left[-\frac{2\pi H_{12}}{v|\sigma_1 - \sigma_2|}\right], \quad (69)$$

where H_{12} is the off-diagonal matrix element of a two-level model Hamiltonian describing the transition between the two levels, and $\sigma_i = dE_i/dL$ are slopes of the crossing levels. In the opposite limit of rapid crossing, the system remains in the original state, and no entropy is produced again. Thus, there should be a maximum of entropy production at some speed. For a discussion of the "entropic forces" in a heavy-quark motion, see, e.g., Kharzeev (2014).

B. Quarkonia and lattice correlation functions

Suppose one puts one J/ψ in matter at some temperature T. Transitions from J/ψ to its excited states will happen first, eventually going into $\bar{D}D$, with the charm quark separated. Since $\bar{D}D$ can eventually occupy an infinite volume, the separated states will dominate over the bound states. Thus, given enough time any initial J/ψ will dissolve at any T. On the other hand, thermal transitions may also proceed in the opposite direction as well. Starting from a certain density of separated charm quarks, charmonia and their ground states J/ψ are constantly regenerated. Given sufficient time, an equilibrium between $\bar{D}D$ and J/ψ will be reached. LHC data suggest that this regime is in fact reached for charm by the chemical freezeout (see later).

Heavy quarks c and b are produced in hard processes, not in thermal reactions. Since the heavy ion collision time is small compared to weak decays, c and b quarks are conserved, and thus c and b chemical potentials can be introduced. (It is not the one coupled to the charm quantum number, but to the c and $\bar{c}$ numbers separately.) This μ is such that charm fugacity is large, defining the equilibrium production of charmonium states.

Can the density of certain quarkonia states be defined at $T > T_c$? We know for a fact that this is not so: the evidence come from lattice static potentials described previously. Already at $T \approx T_c$ there is a relatively large entropy associated with a quark $S = (U - F)(r \to \infty)/2T$, corresponding to a large number of states $N \sim e^S$.

A well-defined field theory object is the correlation function of local gauge invariant operators

$$K(x, y) = \langle \bar{c}\Gamma c(x)\bar{c}\Gamma c(y)\rangle, \tag{70}$$

where the average is over the heat bath, and x or time $x^0 = t$ can be Minkowskian or Euclidean. In both cases the correlation function is related to the same spectral density $\tilde{K}(\omega, k)$, characterizing the amplitude of the excitation of states with a given energy or momentum. At low T, one can find the individual states as certain peaks at the lines $\omega = \omega_i(k)$ in the spectral density, but with increasing T they all merge into a smooth continuum.

Such Euclidean correlation functions have been numerically calculated on the lattice; for a review, see Mocsy, Petreczky, and Strickland (2013). Unfortunately the problem of spectral density reconstruction from these is, in practice, very difficult. Highly accurate and expensive Euclidean correlation functions are converted into relatively poorly defined spectral densities. Even when the individual states are seen, as some peaks in the spectral density, their widths are hardly quantifiable. Above a certain T all peaks corresponding to charmonium states merge into one "near-threshold bump," the imprint of the Sommerfeld-Gamow enhancement due to an attractive potential $\sim e^{-V/T} > 1$.

C. Quarkonia and real-time QFT formalism

A more detailed, time-dependent, set of questions can be asked about transition rates in equilibrium matter. These have been addressed at least at three levels: (a) real-time QFT, (b) quantum mechanical, and (c) classical diffusion.

Real-time QFT, also known as the Schwinger-Keldysh formalism, can follow a system from some initial to some final state using the full Hamiltonian

$$\langle i|P\exp\left(-\int_i^f dtH\right)|f\rangle \tag{71}$$

which is viewed as a sum of the subsystem in question H_0 and matter perturbation V. Diagrammatic expansions, including two-time contours as well as a Matsubara portion of an Euclidean time for thermal media, are widely used in condensed matter problems, but they are not much used so far in the problem we discuss.

If H_0 corresponds to a nonrelativistic quantum-mechanical description of quarkonium, we call it the quantum-mechanical approach. One can evaluate matrix elements of V over various quarkonia states. The first considered reactions were J/ψ excitation into unbound states of $\bar{c}c$ due to photoeffectlike reactions of one gluon absorption $J/\psi + g \to \bar{c}c$. For heavy quarkonia the diagonal part of the real and imaginary parts of the perturbation V can be considered as a modified potential; for a review, see Brambilla *et al.* (2013). More generally, one can define transition rates between states and write a rate equation. The fundamental question here is of course whether the "matter perturbation" V is small or not. (We argued that at very low and very high T a perturbative approach may work, but at least for charmonium in the near-T_c matter the answer to this question is negative.)

Suppose the perturbation V is not small compared with the interparticle interaction: then quantum quarkonium states are no longer special and one can as well use for H_0 just free particles. Using mass as a large parameter, one can argue (Svetitsky, 1988; Moore and Teaney, 2005) that even in a strong-coupling setting the heavy-quark motion should be described by classical stochastic equations, the Langevin or Fokker-Plank type. Let us mention only two crucial consequences of the argument. First, motion is diffusive, with $x \sim \sqrt{t}$ as it happens in random uncorrelated directions. Second, each step in space is very small. Suppose a perturbation delivers a kick of the order T to a heavy quark of mass $M \gg T$. Its velocity is changed little, by $\Delta v \sim T/M$ and by the time the next kick comes $\Delta t \sim 1/T$ the shift in coordinate is small $\Delta x \sim 1/M$.

Suppose a quark needs to diffuse a distance large enough so that the gradient of the potential no longer pulls it toward the antiquark. From the energy potentials V one can see that the distance it needs to go is about 1.5 fm, or $\sim 10\Delta x$ jumps it can make. However, since it is moving diffusively, to get that far the quark would need $\sim 10^2$ jumps, which can be larger than the time available. A quantitative study of classical diffusion in a charmonium (Young and Shuryak, 2009) confirmed that to climb out of the attractive potential in multiple small steps is hard. Contrary to common prejudice, using the realistic charm diffusion constant we found that the survival probability of J/ψ is not small but is of the order of 50% or so.

Finally, in order to model the fate of heavy quarks and quarkonia in heavy ion collisions, one needs to follow them all the way from initial hard collisions to the freezeout. In the classical diffusion approach one starts with a pair distribution in the phase space, as defined by the parton model, and at the end project the resulting distribution to the charmonia states using their Wigner functions.

We finish this section by presenting two opposite answers to the question: What is the effect of the QGP production on the charmonium survival? In 1986 it was argued by Matsui and Satz (1986) that QGP, via the Debye screening of the color potential, destroys charmonium states sequentially, excited states first, and then eventually the ground states. In 2008 the opposite was argued by Young and Shuryak (2009): strongly coupled QGP helps to preserve charmonia. A small mean free path and a specific bottleneck in the Fokker-Planck solution prevents $Q\bar{Q}$ from moving far from each other. At the end of

the process, $\bar{c}c$ pairs are projected back to the bound states, and the ground states of charmonia dominate if the distance between them is small.

D. Observed charmonium composition and chemical equilibration

The simplest, and also most radical, model of charmonium composition is a picture of its statistical hadronization; see Andronic *et al.* (2007) and references therein. It assumes that, like all light hadrons, charmonium states are produced in a thermal equilibrium state at chemical freeze out. If so, issues discussed in the previous section are completely irrelevant: whatever charmonium history during the intermediate stages may be, it is simply forgotten in equilibrium.

The data at RHIC and the LHC show that this is only partially true, and in fact there are two components of the charmonium population: the "survived" one and the products of the "recombination." Observations of the latter component are among the most important results from heavy ion experiments.

Let us proceed to the next-order questions, related, in particular, to the relative population of charmonium states. If all of them come from the statistical hadronization at chemical freezeout, the consequence will be

$$\frac{N_{\psi'}}{N_{J/\psi}} = \exp\left(-\frac{M_{\psi'} - M_{J/\psi}}{T_{\rm ch}}\right) \tag{72}$$

with similar expressions for other states. However, if we have two out-of-equilibrium components, with different history, the answer should be different. The "survival" component, with its flow of probability from small to large r, should be richer in lower states. The recombination component flows the opposite way, and it should have more higher states instead. In general, two components have different centrality and p_t dependences, and in principle can be separated.

E. Are there some stationary quarkonia states in a strongly coupled medium?

As in all other parts of heavy ion theory, there exist two points of view, a weakly coupled and a strongly coupled one on the quarkonia dynamics. Which one is more adequate depends on the value of the quark mass. All approaches assume that it is very large compared to temperature $M \gg T$, so quarkonia are not thermally produced and are nonrelativistic, with $v/c \ll 1$.

If frequencies of the internal motion $\omega \sim v^2 M$ are small compared to those in ambient matter $\sim T$, one should be able to use static potentials from the lattice and calculate wave functions from the corresponding Schrödinger equation. Unfortunately, for both charmonia and bottomonia these frequencies are a few hundreds of MeV, comparable to energies of medium constituents, so this condition cannot really be fulfilled. As argued, this implies that the appropriate potential to be used should be somewhat in between the free energy $F(T, r)$ and the potential energy $V(T, r)$, corresponding to slow and fast limits.

The weakly coupled point of view is valid at parametrically large M. In this case one may argue that quarkonia interact weakly with the matter. Excitations are rare and by integrating an imaginary part of the potential ImV [summarized by Brambilla *et al.* (2013)] over the collision time one obtains quarkonia suppression.

The opposite picture is that very strong coupling assumes that transitions between quarkonium states are numerous. Large ImV exceeding the frequencies (distances between levels) makes the initial vacuum states meaningless. The spectral density of the correlator is smooth, without any peaks. Indeed, there are no two-particle bound states in a dense plasma, just certain spatial correlations between the charges. In practice, classical approaches such as Langevin or FP equations are used. If one starts with a close $\bar{c}c$ pair at $t = 0$, the solution to Langevin or FP equations describes positive flow toward the large relative distance $r \to \infty$. The opposite setting, describing quarkonium recombination, starts with originally unrelated $\bar{c}c$ at large r and calculates the diffusion current directed toward small r. Both are followed for some time, the sQGP era, and at freezeout the distributions obtained (in the phase space p, r) are projected to the vacuum quarkonia states using their Wigner functions (Young and Shuryak, 2009).

Both the inward and outward diffusions turned out to be rather slow. The reason for that is quite interesting. The spatial distribution rapidly reaches a certain shape which persists with only slow growth of its tail. The example is shown in Fig. 55. (Note that in this case the attractive Coulomb-like potential has been complemented by a repulsive quantum effective potential $\sim\hbar^2/mr^2$ which generates the hole in the distribution at small r and prevents classical falling of the charge on the center.)

We called solutions with a nearly permanent shape and small flux "quasiequilibrium" solutions. This concept has not yet been noticed in this particular field[30] but it deserves to be. Let us show how quasiequilibrium solutions appear in the quark diffusion problem, using the FP equation

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial \vec{r}} D\left(\frac{\partial P}{\partial \vec{r}} + \beta P \frac{\partial V}{\partial \vec{r}}\right), \tag{73}$$

where $P(t, \vec{r})$ is the distribution over the $\bar{c}c$ separation $\vec{r}$ at time t, D is the diffusion constant, $\beta = 1/T$, and $V(r)$ is the interquark potential. When $P \sim \exp(-\beta V)$ the two terms on the rhs bracket cancel, so the thermal equilibrium is time independent. Note further that if the bracket on the rhs is nonzero but constant, its divergence is still zero, which makes the lhs zero.

As a result one has a family of stationary solutions, characterized by constant fluxes. The direction depends on the sign of the constant; it can be from small to large r as in the

[30] An example in a different but similar context is the so-called globular clusters of stars inside galaxies. It is well known that they are not in thermal equilibrium, but in a certain dynamical quasiequilibrium, as the data show similar phase space distributions. Depending on the environment, some are slowly evaporating, and some are growing instead, but with the net flux small.

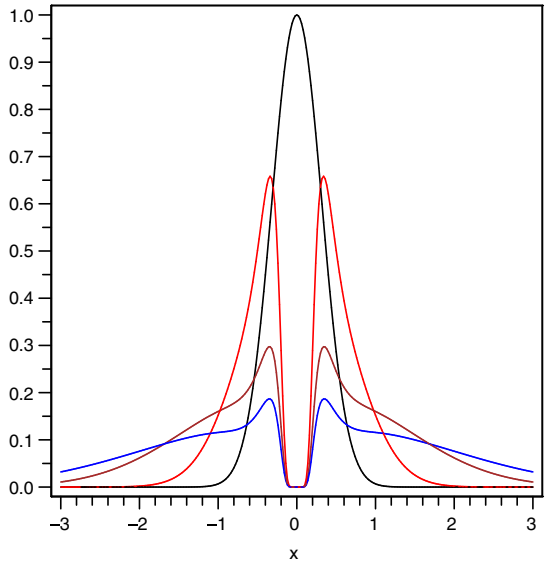

FIG. 55. One-dimensional Fokker-Planck equation for an interacting $\bar{c}c$ pair. The relaxation of the initial narrow Gaussian distribution is shown by curves (black, red, brown, green, blue, or top to bottom at $r = 0$) corresponding to times $t = 0$, 1, 5, and 10 fm, respectively. From Young and Shuryak, 2009.

charmonium suppression problem, or from large to small r for recombination. It is these stationary states that can be called "quarkonium states in matter." (Note, however, that constant flux needs to be supported at the ends: something should produce quarkonia at one end and destroy them on the other.)

XI. JET QUENCHING

Jets are produced by hard collisions of partons, with momentum transfer $Q \gg 1$ GeV, and constitute perturbative cascade from this scale down to on-shell hadrons. Hard collisions in QCD are described perturbatively, while the structure and fragmentation functions are nonperturbative objects treated so far phenomenologically.

By "jet quenching" one means modification in jet yields and shape, due to interaction of the leading parton and subleading cascade with ambient matter. The cascade modification has specific features, due to the so-called Landau-Pomeranchuck-Migdal (LPM) effect generalized from QED to QCD. We do not discus gluon radiation, in vacuum and/or in matter, except for Sec. XI.B on recent progress describing jets as turbulent flow fixed points.

What will be the central focus here is the possible usage of the jet quenching phenomenon as a matter probe. The jet quenching parameter $\hat{q}$ is defined as the mean square of transverse momentum to a jet given to it due to scattering per unit of length. It is a kinetic quantity proportional to the jet scattering cross section, in many respects similar to the inverse viscosity-to-entropy density discussed previously.

A. Is jet quenching dominated by the near-T_c matter?

Let us first consider a more general proposition: If the scattering of QGP quasiparticles and the scattering of jets (high $p_\perp$ partons) on the matter is similar, one may expect various related kinetic coefficients to have similar temperature dependence. In particular, one may expect that

$$\hat{q}(T) \sim \frac{s(T)}{\eta(T)}. \tag{74}$$

From hydrodynamical studies, the rhs seems to have a peak at $T = T_c$. The main message here (Liao and Shuryak, 2009; Xu, Liao, and Gyulassy, 2015) is that the lhs seems to have a peak at $T = T_c$ as well.

There were two experimental hints which eventually led us to this conclusion. The first was the angular dependence of the jet quenching. At the very beginning of the RHIC era it was noticed (Shuryak, 2002) that the simplest model, in which $\hat{q}$ was assumed to be a universal constant, cannot reproduce a direction-dependent (the ratio of in-reaction-plane to out-of-reaction-plane) data, or v_2(large $p_\perp$). The relation between them is simply

$$R_{AA}^{\text{in/out}} = R_{AA}(1 \pm 2v_2). \tag{75}$$

The experimental value of v_2 was, from the very first PHENIX and STAR measurements, about two times larger than all simple quenching models predict. It took years of slicing the matter distributions and trying various T-dependent $\hat{q}$ to find at least one possible solution to this puzzle (Liao and Shuryak, 2009): one can get close to the observed v_2 values if the jet quenching is strongly peaked near T_c. The reason being is that the angular asymmetry of the corresponding shell of matter is sufficiently large.

The second hint was provided by the LHC jet data, showing that quenching at the LHC is rather similar to that at RHIC, in spite of the fact that the multiplicity (and thus matter density) is two times larger there. Now we hope we understand that a shorter time spent near T_c at the LHC, as compared to RHIC, compensates for two times more scatterers.

Now let us have a look at the angular distribution of jet quenching (Fig. 56), showing v_2(large $p_\perp$) measurements at the LHC. There are three upper (red) theory curves, and three

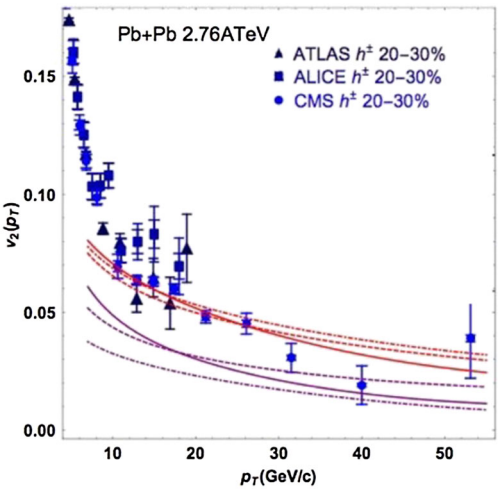

FIG. 56. Jet suppression and elliptic parameter v_2, data vs models. From Xu, Liao, and Gyulassy, 2015.

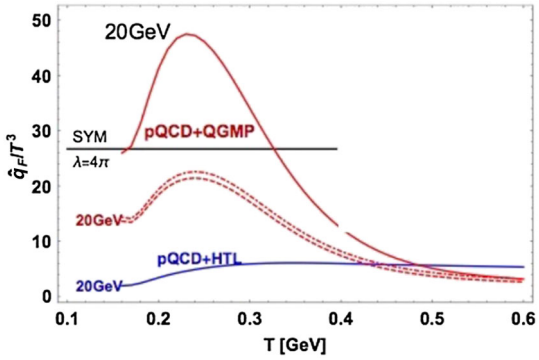

FIG. 57. The normalized dimensionless jet quenching parameter $\hat{q}/T^3$ vs the temperature T (GeV), for a 20 GeV jet. The lower curve marked pQCD + HTL is a standard perturbative contribution of quarks and gluons. Other curves with a maximum include contributions of scattering on monopoles. The horizontal line crudely corresponds to the viscosity/entropy ratio in the AdS/CFT approach.

lower (violet) ones. Only the former describe the data: those include the jet-monopole scattering. The latter, with $f_m = 0$ in the caption, ignore it and thus fail to describe the data. In Fig. 57 we show the $\hat{q}(T)$ of the model for a jet with 20 GeV. The red curve marked pQCD + QGMP has a peak reaching $\hat{q}/T^3 \sim 50$ in the left upper corner and includes the scattering on monopoles (this is what M in the matter description is). Note that the blue curve at the bottom marked pQCD + HTL, which includes quarks and gluons only in the hard thermal loop (HTL) approximation, has $\hat{q}/T^3 \sim 5$, an order of magnitude lower.

This is a concrete manifestation of what was discussed in the Introduction, where we compared perturbative and nonperturbative effects. The perturbative amplitude has the gluon charge and the coupling $\alpha_s \sim g^2$, one g from a jet and one from the "scatterer." The nonperturbative amplitude still has a factor g from a jet, times[31] $1/g$ from the field of the monopole: thus there is no coupling.

B. "Fixed points" of the jet distributions

There have been important developments relating jet in matter with a general turbulence theory (Blaizot, Mehtar-Tani, and Torres, 2015). For a large enough medium, successive gluon emissions can be considered as independent: multiple emissions can be treated as probabilistic branching processes, with the elementary branching rate. The inclusive gluon distribution function

$$\frac{dN}{d\log(x)d^2\vec{k}} = \frac{D(x,\vec{k},t)}{(2\pi)^2} \tag{76}$$

satisfies certain diffusion-branching equations. Integrating over transverse momentum one gets the zeroth moment $D(x,t) = \int_k D(x,k,t)$ on which we focus here, for simplicity. This moment satisfies

$$t_* \frac{\partial D(x,t)}{\partial t} = \int dz K(z) \left[\sqrt{\frac{z}{x}} D\left(\frac{x}{z}, t\right) - \frac{z}{\sqrt{x}} D(x,t) \right] \tag{77}$$

with the gain and loss terms on the rhs. Details such as the shape of the kernel K and time parameter t_* can be found in Blaizot, Mehtar-Tani, and Torres (2015). The central point is the analytic solution

$$D(x,t) = \frac{(t/t_*)}{\sqrt{x}(1-x)^{3/2}} \exp\left(-\frac{\pi t^2}{t_*^2(1-x)}\right) \tag{78}$$

which balances the gain and the loss.

Note that an essential singularity at $x = 1$ is expected: it is known as the Sudakov suppression factor. The remarkable news is that apart of the exponent, the shape of the x dependence remains the same at all times. Only the normalization changes. Let us see how it works in the most important small $x \ll 1$ region, where the leading $x^{-1/2}$ dependence is such that the gain and loss terms cancel. This is the quasiequilibrium solution for the jets. As a result, jets in matter are expected to approach some universal shape, not determined by the particular initial conditions, but by the quasiequilibrium solution to which it gets attracted as the process proceeds.

It is essentially the same phenomenon as seen in Fig. 55 for diffusing charmonium: the shape itself is dictated by the balance of the gain and loss. Both are quasiequilibrium attractor solutions: their main feature is constant flux of a certain quantity, from one end of the spectrum to the other. (The flux in decaying charmonium comes from small to large distance between quarks, in the jet case it comes from large to small x.) Once again, the constancy of the flux in such solutions is the key idea going back to Kolmogorov's theory of hydrodynamical turbulence. As Einstein once observed: "… the number of good ideas in physics is so small, that they keep being repeated again and again in various contexts."

XII. NEAR THE PHASE BOUNDARY: FLUCTUATIONS AND FREEZEOUTS

A. Chemical freezeouts

The concept of two separate freezeouts is based on the separation of elastic and inelastic rates at low T, in magnitude. Statistical equilibration is such a success story that we just show current ALICE data, with the corresponding thermal fits, in Fig. 58. For a simple statistical model with two parameters, the quality of the data description is extraordinary.

Note that even light nuclei d, t, He^3 and their antiparticles are also included. One may wonder how d, with its $B = 2$ MeV binding energy, can be found in an environment with ambient temperature $T \sim 160$ MeV $\gg B$. Should not d be instantly destroyed in it? The answer to this and similar questions is well known. Thermodynamics does not depend on the d lifetime. As one deuteron is destroyed, perhaps with some large rate, in equilibrium another one must be recreated, by the inverse process with the same rate. The average

[31]The term $1/g$ can also be called the magnetic coupling constant, related to the electric one by the Dirac condition.

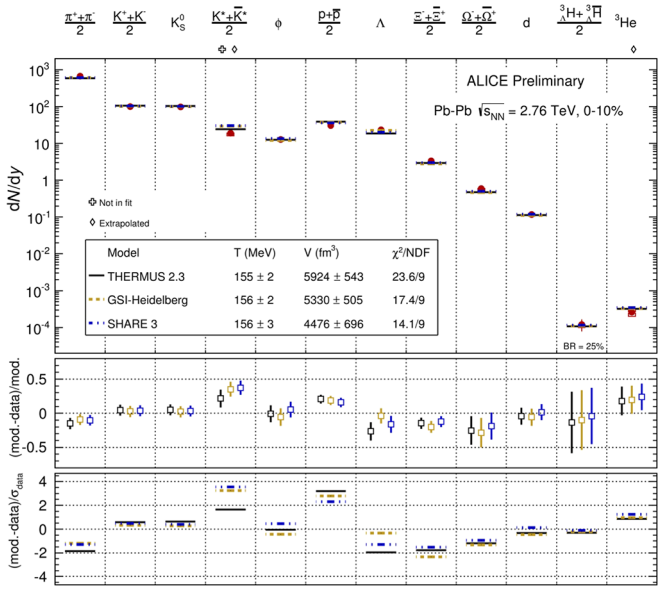

FIG. 58. ALICE data on particle yields compared to the thermal model. The main fit parameters are indicated in the upper plot. From Andronic *et al.*, 2007.

population is conserved, and it is what the statistical mechanics gives us.

Note that one deviation is K^*: the model predicts more than observed. This is expected: it is a short lived resonance which decays when the density is still non-negligible, the products can be rescattered, and their invariant mass moved out of the peak. Corrections to that can be made using any cascade codes.

Another deviation is for $p+\bar{p}$. Some argued that one should take into account possible annihilation on the way out. As available practical tools cascade codes were used (RQMD, URQMD, or similar) and got large effects. This puzzle was even much stronger at lower collision energies at which the density of baryons is large and antibaryons were predicted to get nearly extinct, contrary to observations.

The puzzle was resolved by Rapp and Shuryak (2001): all cascade codes included the annihilation $p+\bar{p}\to n\pi, n\sim 5$ but not their inverse reactions. Contrary to popular belief, the inverse reactions are not suppressed. In fact at equilibrium their rates are exactly the same as that of the direct one.

I will show the freezeout points on the phase diagram, which has been done many times. Let us just recall that these seem to be remarkably close to the phase boundary, defined on the lattice. Why should this be the case? An answer suggested by Braun-Munzinger, Stachel, and Wetterich (2004) is also related to the multiparticle reaction rates. Since these depend on a high temperature power, they all should decouple very close to the critical line

$$\frac{|T_{ch}-T_c|}{T_c}\ll 1. \tag{79}$$

The new trend in freezeout physics is the focus of susceptibilities. As pointed out by Shuryak (1998) before the RHIC era, information about them can be provided by measured event-by-event fluctuations. Since the susceptibilities are higher order derivatives of the free energy, over T or various chemical potentials, they are more sensitive to singularities. This idea was applied further toward the location of the critical point in Stephanov, Rajagopal, and Shuryak (1999), suggesting the low energy beam-scan program at RHIC. The results of the actual beam scan are, however, still not finalized enough to be reviewed here.

Both the measurements of event-by-event fluctuations in experiments and the calculation of susceptibilities on the lattice reached significant maturity. In fact comparison between them now allows the $T-\mu_B$ freezeout curve to be reconstructed, even without using any particle ratios. An example by Alba *et al.* (2014) is shown in Fig. 59. Figure 59(a) compares a new set of freezeout parameters (points) to earlier ones from the particle ratios. Figure 59(b) shows that as μ grows the consistency between different ratios becomes worse. It is generally believed that the hadron resonance gas model describes the QCD thermodynamics at chemical freezeout quite well.

B. From chemical to kinetic freezeouts

Separation in magnitude of the elastic and inelastic (low energy) hadronic reactions is the basis of the "two freezeouts" paradigm, with separate chemical $T_{\rm ch}$ and kinetic $T_{\rm kin}$ temperatures. Its effectiveness became even more clear at the LHC, illustrated in Fig. 60 containing the "blast wave" fitted

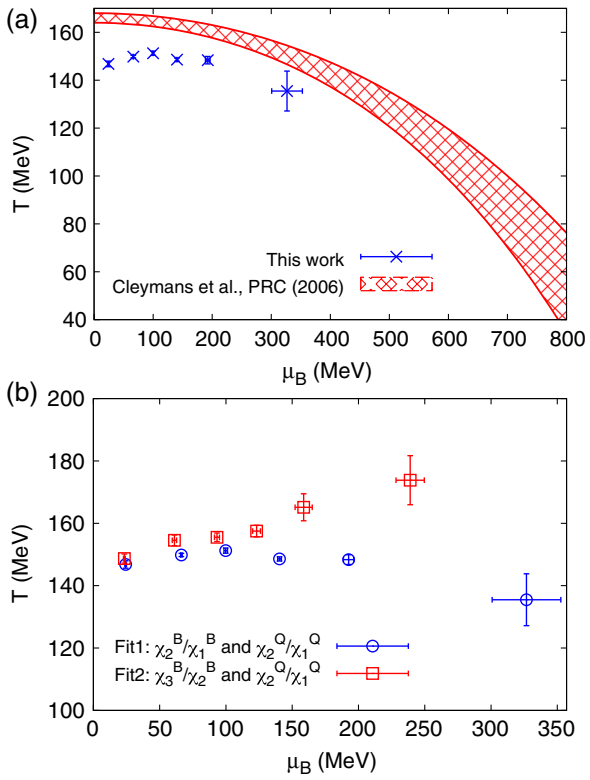

FIG. 59. (a) Freezeout parameters in the $T-\mu_B$ plane: comparison between the curve obtained in Cleymans *et al.* (red band) and the values obtained in the present analysis from a combined fit for net-electric charge and net protons (blue symbols). (b) The freezeout parameters fit from two different set of particle ratios.

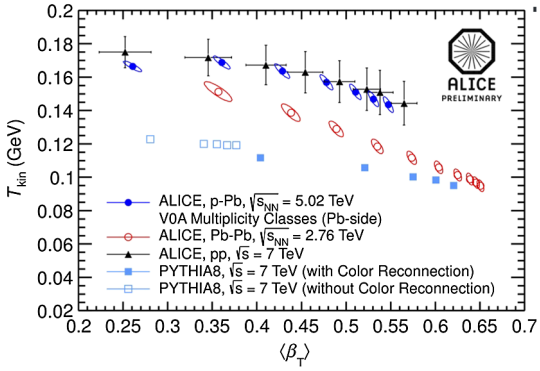

FIG. 60. The temperature of the kinetic freezeout vs mean velocity of the radial flow, fitted to ALICE spectra of the identified secondaries ($\pi, K, p, \Lambda, \Xi, \Omega$).

parameters to the ALICE spectra of identified secondaries. Unlike $T_{\rm ch}$, the kinetic $T_{\rm kin}$ strongly depends on the centrality of PbPb collisions, decreasing to values below 0.1 GeV for the most central bins. Cooling from 0.16 to 0.1 GeV may not look so dramatic but one should remember that pressure and energy density in this region are proportional to the high power of T, and so we talk about a change by a factor 20 or so.

Let us discuss the dependences displayed in Fig. 60 in more detail. The most central AA collisions produce the largest systems, which also have the highest T at the early stages. The fit displayed also shows that most central collisions produce the lowest T at the kinetic freezeout. Indeed, the larger the system, the smaller is its expansion rate, and thus its freezeout at a smaller collision rate or smaller density.

Consider now the pp, pA data of Fig. 60: for them both temperatures $T_{\rm ch}$ and $T_{\rm kin}$ are much closer. An explanation to that, discussed previously, is based on strong radial flow related to higher expansion rates. These collisions definitely cannot support the hadronic phase.

In summary, there is evidence that in central AA collisions matter cools deep into the hadronic phase, retaining kinetic, but not chemical, equilibrium. This raises some interesting questions related to the hadronic phase, which can now be addressed experimentally. Mentioning the relevant numbers: Cold fireballs created in central PbPb have several thousands of particles and their kinetic freezeout time reaches 15 fm. The highest multiplicities in pp, pA still correspond to fireballs with an order of magnitude less particles, and a freezeout time and size of 3 fm or so.

Since between $T_{\rm ch}$ and $T_{\rm kin}$ the particle numbers are conserved, one should introduce new nonzero chemical potentials, not associated with conserved quantum numbers such as charge and baryon number. In particular, there should be nonzero chemical potentials for pions. Whether there are nontrivial fugacity factors at the kinetic freezeout can be directly observed in the pion spectra, because in this case Bose enhancement can be measured. This idea is at least 20 years old (Bebie *et al.*, 1992): its experimental manifestation was also discussed by Hung and Shuryak (1998); see Fig. 61(a). (The reference SS collisions is a much smaller system than PbPb, and thus its chemical and kinetic freezeouts should be close.) Figure 61(b) shows that the same effect shows up, now at the LHC. The fit without chemical equilibrium, with

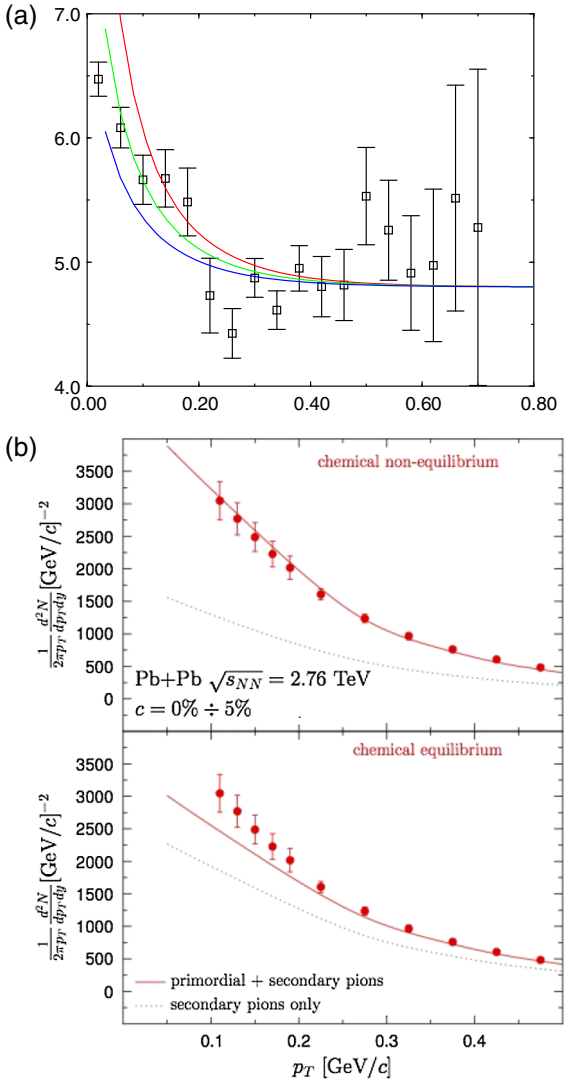

FIG. 61. (a) Points show the ratio of PbPb to SS spectra, from NA44. The three curves are for pion chemical potential $\mu_\pi = 60$, 80, and 100 MeV. From Hung and Shuryak, 1998. (b) Dots are ALICE transverse momentum spectra of pions in the low-p_t region, compared to the model with (upper plot) and without (lower plot) pion chemical potential. From Begun, Florkowski, and Rybczynski, 2014.

nonzero pion μ on top, provides a better description to the spectra at small $p_t < 200$ MeV.

It is interesting that the parameter of the fit in this last work gives $\mu_\pi \approx m_\pi$, so they speculated if the conditions for pion Bose-Einstein condensation (BEC) were actually reached. If this indeed becomes true, it has been many times suggested previously that the femtoscopy parameter λ should show it as it is sensitive to the "degree of coherence" of the pion source. Femtoscopy data on two and three identical pions from ALICE can be indeed fitted with a coherent source, with a fraction as large as 20%.

Do we actually witness the BEC formation at the LHC? In order to answer this question it is useful to recall the BEC discovery in experiments with ultracold atoms a decade ago. As the atomic system undergoes evaporating cooling and its temperature decreases, the measurements of the momentum

distribution (by switching off the trap) revealed the appearance of a new and much more narrow component. Unlike the usual thermal component, its width was independent of T and related to the inverse spatial size of the BEC cloud.

This indeed sounds like what is observed in heavy ion collisions: as one goes to the most central collisions and the kinetic freezeout $T_{\rm kin}$ gets below 100 MeV, the p_t spectra do become enhanced at small momenta. The difference, however, is in the shape: we do not see a new Gaussian, as in the atomic experiments, but only some deformation of the spectrum, by $\mu \neq 0$.

If the condensate is produced, it should be a separate component, with μ being exactly m_π, independent of T. If the BEC cloud contains about $1/4$ of all pions, its diameter should be large, at least of the order of 2–3 fm. The corresponding width of momentum distribution, from the uncertainty relation, should be as small as say $\langle p_t \rangle < 0.1$ GeV. Looking back to Fig. 61(b) one, however, finds that such soft secondaries seem to be outside of the acceptance. Thus, even if the BEC component is there, we so far cannot see it, neither with ALICE nor with any other LHC detectors. How then can we get their influence in the femtoscopy?

This issue can perhaps be clarified by a short dedicated run, in which the ALICE detector switches to a smaller (say $1/2$ of the current value) magnetic field, to improve the low p_t acceptance. Recalculating all the efficiencies is a lot of extra work, but perhaps it is worth clarifying this interesting issue.

C. The search for the critical point and the RHIC low energy scan

The main idea of a scan aimed at the QCD critical point (Stephanov, Rajagopal, and Shuryak, 1999) is well known. The critical point, if it exists, should enhance the event-by-event fluctuations, similar to critical opalescence, known in many cases. Technically, various effects given by diagrams can be classified according to a number of propagators of the critical modes, with each enhancing the effect due to large correlation lengths. It was quantified by Stephanov (2009). The n-particle correlators may contain up to n such propagators, three-particle correlators are $\sim\xi^6$, four-particle ones $\sim\xi^8$, etc.; see Fig. 62. The wavy line at zero 4-momentum is $\sim 1/m_\sigma^2 \sim \xi^2$: but the prediction is not just the power of the propagators because the coupling of critical modes by itself vanishes as certain power of ξ given by the critical indices. The quartic one in the diagram considered is $\sim 1/\xi$ so the total power is 7, not 8.

It is possible to tell the same story in a somewhat simpler nontechnical language. The critical field we call here σ should be viewed as some stochastic or fluctuating background field coupled to fluctuations in a particle number (circles with crosses in Fig. 62). One can view it being proportional to some stochastic potential $\Delta V(x)$ which enters the probability in the usual way $P \sim \exp[-\Delta V(x)/T]$, so that in its minima the probability is larger and more particles; e.g., four protons mentioned previously all gather there. The critical point is special in that the scale of the correlation length ξ of this potential increases, and thus more particles have a chance to get inside the correlation length.

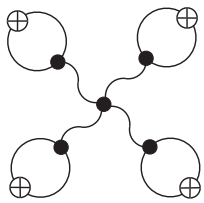

FIG. 62. The enhanced contribution to a four-particle correlator. From Stephanov, 2009.

Which particles should one observe? In principle σ is scalar isoscalar, coupled to any hadron. Stephanov, Rajagopal, and Shuryak (1999) and Stephanov (2009) considered the simplest coupling as an example, the $\sigma\pi\pi$ one, and so the particles were pions. Yet one can argue that the nucleons should work better. First, the powers of the baryon density $n_B = N_N - N_{\bar N}$ correlated together are the susceptibilities calculated on the lattice as derivatives over μ_B. Second, we know from the nucleon forces, e.g., the simplest version of the Walecka model. that σ is the main component of the attractive nuclear potential which binds the nuclei

$$\Delta V = \frac{g_{\sigma NN}^2}{4\pi r}\exp(-m_\sigma r). \tag{80}$$

In vacuum the typical mass $m_\sigma \sim 600$ MeV and the inter-nucleon distance $r \sim 1.5$ fm are combined into a small suppression factor $\sim \exp(-5) \ll 1$ explaining why the nuclear potential scale $\Delta V \sim -50$ MeV is much smaller than the nucleon mass, in spite of strong coupling. (At smaller distance r the repulsive ω contribution dominates the attractive σ one.)

Can it be so that at the QCD critical point $m_\sigma \to 0$, this small exponent disappears? If so, one should expect much deeper ΔV, perhaps even larger than the freezeout temperature T. Furthermore, if say $\xi = 2$ fm, the volume $4\pi\xi^3/3 \sim 40$ fm^3 is large enough to collect many nucleons, not just three or four, as Stephanov suggested. Thus, such clustering of the nucleons should produce large nuclear fragments, a new signal of the critical point.

This argument, unfortunately, is still a bit naive. The critical mode which gets long range is not just the σ field, but, because we are at nonzero density, a certain combination of scalar σ with vector ω fields. Therefore the repulsive forces between nucleons may become long -range as well. To discover what happens we need a reliable theory or some dedicated experiments. Fortunately, we can do it in the coming low energy scan. The isoscalar σ interacts with scalar, net baryon, density $n_s = N_N + N_{\bar N}$, while ω interacts with $n_B = N_N - N_{\bar N}$. The powers of these differ by the nondiagonal terms such as nucleon-antinucleon correlators $C^{m,n} = \langle N_N^m N_{\bar N}^n \rangle$ which can and should be measured. Perhaps restricting kinematics of all particles involved, rapidity and momenta differences, would further enhance the signal.

Let us now jump to the STAR data shown in Fig. 63. The proton four-point correlator has an interesting structure: a minimum at $\sqrt{s} = 20$–30 GeV and perhaps a maximum at low energy. Antiprotons have a similar shape but with much smaller amplitude. The structure qualitatively agrees with

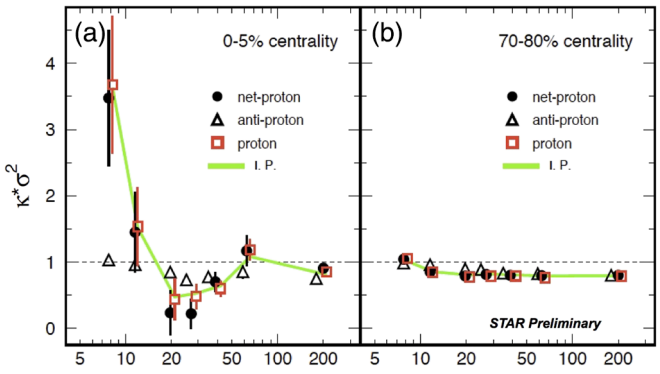

FIG. 63. The kurtosis, a four-particle correlator, in units of the width, as a function of the collision energy $\sqrt{s}$, GeV.

theoretical predictions of an oscillatory behavior of the kurtosis near the critical point.

However, before getting excited by the new large signal, with large error bar, let us recall that it appeared as a result of particle ID improvement, from $0.4 < p_\perp < 0.8$ GeV/c to now reaching $p_t = 2$ GeV/c. The newly open kinematic window should be sensitive to hydrodynamic flow and potentially to its fluctuations.

Also near the critical point one expects significant modifications of attractive (σ-related) nuclear forces: can those affect the four-proton (antiprotons) correlations in question?

Clearly more data at the low energies are needed to understand whether one has indeed located the QCD critical point or not.

XIII. SUMMARY AND DISCUSSION

A. Progress on the big questions

Before we summarize conclusions on the particular subjects, let us remind the reader "the big questions" mentioned in the Introduction:

(I) Can one locate the soft-to-hard boundary, where the transition from the strong to the weak coupling takes place?

(II) Can one locate the micro-to-macro boundary, where the transition from the ballistic to collective regime takes place?

(III) Can we experimentally identify signals of the QCD phase transition, in particular, locate the QCD critical point?

Somewhat surprisingly, the sharpest observed transition discussed is in the profile of the pp elastic amplitude shown in Fig. 51(a). Although indirectly, this sharp transition from nearly black to light gray profile is claimed to be related to the phase transition from the deconfined (gluonic) to confined (stringy) regimes of the Pomeron. On the one hand, the sharpness is surprising because it is associated with a quite small system—the Pomeron or a pair of strings, rather than a macroscopic system. On the other hand, the analogy originates from the phase transition in gluodynamics (strings at early stage are considered excitable but not breakable, so no quarks), which is the first-order transition.

The location of the micro-to-macro transition in pA and pp collisions, as a function of multiplicity, is debatable. Data on the mean $p_\perp$ and slopes shown in Fig. 24 indicate smooth growth of the radial flow: but we now know that the radial flow can be faked. At the same time, the $v_2\{2\}$ as a function of multiplicity are rather flat. Its version $v_2\{n\}, n > 2$, from multiparticle correlations, show changes, but their understanding is still missing. Calculations of v_n in dynamical models of the Pomeron are in progress: perhaps they will explain later the low-multiplicity side.

The theoretical justification of a successful hydrodynamical description for small systems is getting under control. A number of examples show rather effective cancellations of all higher-gradient corrections. The Navier-Stokes approximation seems to be rather accurate, even in situations in which one hardly expects it to work.

The low energy scan at RHIC shows clear experimental evidences for "the softest point." Attempts to locate the effects of the QCD critical point are intriguing but not yet conclusive.

B. Sounds

The first triumph of hydrodynamics, at the onset of the RHIC program, was the description of the "little bang." The magnitudes of the radial and elliptic flows were measured and calculated, as a function of $p_\perp$, centrality, rapidity, particle type, and collision energy. A successful description of higher azimuthal harmonics of the flow, with $m = 3$–6, had followed. As repeatedly emphasized, these are sounds, propagating on top of the exploding fireball. The damping of these modes agrees with acoustic-inspired formulas.

Another phenomenon, well known for the big bang perturbations, is due to a presence of the "phase factor." The common freezeout times for all harmonics imply m-dependent phases. As m grows, the phases rotate, so one should see maxima or minima in the power spectrum. These are not observed: the only experimental indication for that is the triangular flow $m = 3$ stronger than the elliptic one $m = 2$ for the ultracentral bin.

We emphasized that we have observed harmonics only with $m < 7$ because the higher ones are damped too much by the freezeout time. Yet at the initial time the Glauber model produced equally well harmonics up to $m = 20$ or so, and GLASMA-based models predict harmonics to the hundreds. Most of them do not survive until freezeout: so, are there any observable manifestations of their existence? The magneto-sono-luminescence process, converting them into electromagnetic signals, is an example of that.

The damping of harmonics with $m > 6$ is also an opportunity to observe the sources of sounds other than the initial state, in particular, from inhomogeneities at the phase transition.

Finally, even in equilibrium there must be fluctuations emitting sounds. These generate nontrivial "loop corrections" to hydrodynamics. Observation of "sound background" in hadronic matter is another challenge of the field.

C. The conflicting views of the initial state

Perhaps the most important conceptual controversy in the field remains the conflicting conclusions coming from weakly and strongly coupled scenarios of the initial state and equilibration.

Significant progress in the theory of a weakly coupled initial state is in the concept of turbulent cascades, with

stationary and time-dependent self-similar solutions. Both classical glue simulations and parton cascades came up with out-of-equilibrium attractors possessing power spectra with certain indices, which are qualitatively different from equilibrium. From a practical perspective, these studies suggest that the stress tensor remains anisotropic for a long time. However, more recent works indicate that the nontrivial attractor is only approached if the coupling is unrealistically small.

Strongly coupled approaches, especially based on AdS/CFT and related models, view equilibration as a process dual to the gravitational collapse resulting in black hole production in the bulk of AdS_5. As soon as some trapped surface (a black hole) is present, the equilibration is very rapid: any kind of debris simply falls into it. Mathematically, the nonhydro modes have imaginary parts comparable to the real ones, which numerically are quite large (50). In this scenario, there are no cascades or even quasiparticles, the only propagating modes are sounds (dual to gravitons).

Whether the stress tensor remains anisotropic beyond the short initial period or not is still an open question. Theoretical efforts to combine hydrodynamics with out-of-equilibrium parametrization of the stress tensor were discussed: the situation at any realistic anisotropy is thus under theoretical control.

In order to decide on the equilibration time and anisotropy, one needs to develop experimental observables sensitive to the early stage. Our specific proposal, the dilepton polarization, was discussed in Sec. IX.C.

D. The smallest drops of sQGP

The major experimental discovery from the first years of LHC operation was the observation of collective anisotropies in "small systems": central pA and high-multiplicity pp collisions.

One point of view, previously advocated, is that these are exploding fireballs. While smaller than those produced in AA collisions, they are still "macroscopically large" and can be described hydrodynamically. The hydrodynamical description of strong radial and elliptic flows in these systems is very successful.

The opposite point of view is that, from the smallest to the highest multiplicity bins, the pA and pp collisions produce microscopic systems which can be discussed dynamically. The models are the same as used for minimally biased pp, and the issue is sometimes known as "the shape of Pomeron." Recall that the Pomeron is based on the pQCD ladder diagrams in weak coupling, or strings in confining models. Both need to be developed much further to predict v_n correlations. While experiments at RHIC with d and ^{3}He beams disfavor such scenarios for large multiplicities, at lower ones they should be applicable.

A positive development is that groups working on all scenarios try now to figure out the limits of their approaches. Inside hydrodynamics, for example, one study is of higher gradients and their effect. In the string-based picture a discussion appears of a string-string interaction, ignored for a long time by event generators.

Meanwhile, phenomenologists describe the data. Hydrodynamical treatment of high-multiplicity pA, pp events seems to be rather successful. They require very small initial sizes and rather high temperatures, while we do not really understand how such systems can be produced. In particular, the case of central pA collisions is contested between the IP-GLASMA model and a string-based initial state picture. So far one has very little theoretical control over the initial state of the high multiplicity pp: if it exists anywhere, GLASMA should be there. It is difficult to study this system, for statistical reasons, but since this is the highest density system we have, it should be pursued.

E. Heavy quarks and quarkonia

The LHC data confirmed what has already been hinted at by the RHIC data: a significant fraction of the observed charmonia comes from recombination of charm quarks at the chemical freezeout. The "surviving charmonia" fraction continues to be reduced. Such a major change in charm quark behavior from "heavy like" to "light quark like" is clearly an important discovery.

It remains true that c and b quarks are produced differently from the light ones, namely, in the initial partonic processes. Yet their interaction with the ambient matter is strong. At large p_t we observe quenching $R_{AA}^{c,b}$ comparable to that of gluons and light quarks. At small p_t we observe an elliptic flow of open charm and changes in spectra.

Langevin and Fokker-Planck studies, however, suggest that c quarks are not moving with the flow. At early time c and b quarks are produced with large $p_\perp$ and start decelerating, due to drag, while the matter is slowly accelerating due to pressure gradients: their velocities move toward each other, yet they do not match even by the end. As a result, charm radial and elliptic flows are not given by the Cooper-Frye expression. The recombinant charmonia may perhaps be an exception. Whether they actually co-move with the flow still needs to be established.

On the theory front, Langevin and Fokker-Planck studies induced new conceptual developments. In particular, we discussed a new set of solutions of those for charmonia, the quasiequilibrium attractors with constant particle flux. These states, not the original bound states such as $J/\psi, \psi'$, etc., provide a convenient basis for evaluation of the speed of relaxation and out-of-equilibrium corrections to current charm hadronization models.

F. Jets

The theories of hard processes (jets, charm, and bottom production) were based on factorization theorems and a concept of structure functions. It is a solid foundation, but a very restrictive one. When one asks questions about jets in high-multiplicity bins of pp collisions, one soon realizes the corresponding structure functions do not exist: that concept has been defined only for the minimally biased (untouched) proton in a strictly inclusive setting. Universal structure functions, measured rather than calculated, have served us since the 1970s, but now they cannot be used anymore. If a

certain fluctuation of a nucleon is selected, new models and much more measurements are needed.

Unfortunately, there are severe practical limits: hard processes reduce the probability by several orders of magnitude, on top of 5% trigger for central pA. Yet high LHC luminosity plus specialized triggers should be enough to get to some of those issues in the near future. Current jet quenching data even for minimum bias pA remain to be understood. Scaling arguments, such as the ones we used for hydro in smaller-but-hotter systems, can and should be developed and compared with the data.

We argued that in AA collisions the jet quenching parameter $\hat{q}$ seems to be strongly enhanced at the near-critical $T \approx T_c$ region. Small systems are more explosive and pass near-freezeout stage rapidly: this should play a significant role in jet quenching.

ACKNOWLEDGMENTS

This paper is a summary from multiple conversations with colleagues, at seminars, workshops, and conferences. They are too many to attempt to name them here (see the long list of references), but still I need to thank them for patiently teaching me about this or that idea or experimental findings. This work was supported in part by the U.S. Department of Energy, Office of Science, under Contract No. DE-FG-88ER40388.

APPENDIX A: HEAVY ION TERMINOLOGY

"Ion" in physics refers to atoms with some of its electrons missing. While at various stages of the acceleration process the degree of ionization varies, all of it is unimportant for the collisions, which always are done with nuclei fully stripped.

By "heavy ions" we mean gold ^{197}Au (the only stable isotope in natural gold, and a favorite of the BNL) or lead ^{208}Pb (the double magic nucleus used at CERN). Some experiments with uranium U have also been done, but not because of its size but rather due to its strong deformation.

Collision centrality in physics is defined usually via an *impact parameter* b, the minimal distance between the centers of two objects. It is a classical concept, and in quantum mechanics channels with *integer* angular momentum $l = L/\hbar$ (in units of Plank constant) are used. However, collisions at very high energy have high angular momentum and uncertainty in b is small. The standard way of thinking about centrality is to divide any observed distribution, e.g., over the multiplicity P_n, into the so-called *centrality classes*, histogram bins with a fixed fraction of events rather than width. For example, many plots in this review say something like "centrality 20%–30%": This means that the total sum $\sum_n P_n$ is taken to be 100%, the events are split into say ten bins, numerated 0%–10%, 10%–20%, 20%–30%, etc., and only events from a particular one are used on the plot under consideration. The most central bins have the largest multiplicity and are always recorded, the more peripheral ones (say 80%–100%) often are not used or even recorded. While the observables, such as mean multiplicity, decrease with centrality b monotonically, it is not true for individual events. Multiple possible definitions of the centrality classes may sound complicated, but are not, and simple models such as Glauber nucleon scattering give quite good description of all these distributions, so in practice any centrality measure can safely be used.

The number of participant nucleons N_p plus the number of "spectators" is the total number of nucleons $2A$. The number of spectators (usually only the neutrons) propagating along the beam direction are typically recorded by special small-angle calorimeters in both directions. Two-dimensional distributions over signals of both such calorimeters are cut into bins of special shapes, also in a way that each bin keeps a fixed percentage of the total. Small corrections for nucleons suffering only small angle elastic and diffractive scatterings—not counted as "participants" are also made.

An overlap region is a region in the transverse space in which the participant nucleons are located at the moment of the collision. Note that due to relativistic contraction, high energy nuclei can be viewed as purely a 2D object, with longitudinal size reduced practically to zero: therefore the collision moment is well defined and is the same for all nucleons.

Flow harmonics are Fourier coefficients of the expansion in azimuthal angle ϕ:

$$\frac{dN}{dydp_\perp^2 d\phi} = \frac{dN}{dydp_\perp^2}\left[1 + 2\sum_m v_m(p_\perp)\cos(m\phi)\right]. \quad \text{(A1)}$$

Its measurements require knowing the direction of the impact parameter vector $\vec{b}$, from which the azimuthal angle ϕ is counted. The direction of $\vec{b}$ and the beam define the so-called *collision plane*. The direction of $\vec{b}$ in the transverse plane is traditionally denoted by x, the orthogonal direction by y, and the beam direction by z.

In practice this comes either from separate "near beam" calorimeters, recording "spectator" nucleons, or from correlations with other particles. The flow harmonics are often introduced as a response of the system to the asymmetry parameters ϵ_m describing Fourier components of matter distribution in ϕ. Note that v_m relates to momentum distribution and ϵ_m to that in space: connection between the two is nontrivial.

Collectivity of flow: Flow harmonics were originally derived from two-particle correlations in relative angle, to which they enter as a mean square

$$v_n^2\{2\} = \langle e^{in(\phi_1-\phi_2)}\rangle = \langle |v_n|^2\rangle. \quad \text{(A2)}$$

Alternatively, it can be derived from multihadron correlation functions: e.g., those for four and six particles mostly used are

$$v_n^4\{4\} = 2\langle |v_n|^2\rangle^2 - \langle |v_n|^4\rangle, \quad \text{(A3)}$$

$$v_n^6\{6\} = \tfrac{1}{4}(\langle |v_n|^6\rangle - 9\langle |v_n|^2\rangle\langle |v_n|^4\rangle + 12\langle |v_n|^2\rangle^3). \quad \text{(A4)}$$

By "collectivity" one means the fact that all such measurements produce nearly the same values of the harmonic v_n. In contrast to that, the "nonflow" effects, e.g., production of resonances, basically only affect the binary correlator $v_n\{2\}$ but not the others.

Soft and hard secondaries mentioned in the main text indicate their dynamical origin. "Soft" come from thermal heat bath, modified by collective flows, while the "hard" ones are from partonic reactions and jet decay. The boundary is not well established and depends on the reaction: soft are with $p_\perp < 4$ GeV while hard is perhaps with $p_\perp > 10$ GeV.

Rapidity y is defined mostly for longitudinal motion, via the longitudinal velocity being $v_z = \tanh(y)$. There is also the so-called space-time rapidity $\eta = (1/2)\log[(t+z)/(t-z)]$ (which should not be confused with viscosity, also designated by η) used in hydrodynamics. Both transform additively under the longitudinal Lorentz boost.

Sometimes one also uses transverse rapidity $y_\perp$ related to the velocity by $v_\perp = \tanh(y_\perp)$. A pseudorapidity variable is an approximate substitute for rapidity y, used when particle identification is not available.

Chemical and kinetic freezeouts refer to stages of the explosion at which the rates of the *inelastic* and *elastic* collisions become smaller than the rate of expansion. The chemical freezeout is called so because at this stage particle composition, somewhat resembling a chemical composition of matter, is finalized. The kinetic or final freezeout is where the last rescattering happens: it is similar to the photosphere of the Sun or to CMB photon freezeout in cosmology. The timelike surfaces of the chemical and kinetic freezeouts are usually approximated by isotherms with certain temperatures. The final particle spectrum is usually defined as the so-called Cooper-Frye integral of thermal distribution over the kinetic freezeout surface.

Femtoscopy or the HBT interferometry method came from radio astronomy: HBT is the abbreviation for Hanbury-Brown and Twiss who developed it there. The influence of Bose symmetrization of the observed meson wave function in particle physics was first emphasized by Goldhaber *et al.* (1960) and applied to proton-antiproton annihilation. Its use for the determination of the size and duration of the particle production processes was proposed back in the 1970s (Shuryak, 1973; Kopylov and Podgoretsky, 1974). With the advent of heavy ion collisions this "femtoscopy" technique grew into a large industry. Early applications for RHIC heavy ion collisions were in certain tension with the hydrodynamical models, although this issue was later resolved (Pratt, 2009).

QCD thermodynamics on the lattice is the calculation of the thermodynamical observables from the first principles, the QCD Lagrangian, using numerical simulations of the gauge and quark fields discretized on a four-dimensional lattice in Euclidean time. For a recent review, see, e.g., Ding, Karsch, and Mukherjee (2015), from which we took Fig. 64. The first thing to note is that quantities plotted are all normalized to corresponding powers of the temperature given by its dimension: scale-invariant matter corresponds to T-independent constants at this plot. And indeed, the curves seem to approach constant at its right side (high T). The second thing to note is that these constants seem to be lower than the dashed line at high temperatures, corresponding to a noninteracting quark-gluon gas. It is interesting that the value for infinitely strongly interacting supersymmetric plasma is predicted to be $3/4$ of this noninteracting value.

Confinement and deconfinement: The term is an abbreviated version of "color-electric confinement phenomenon,"

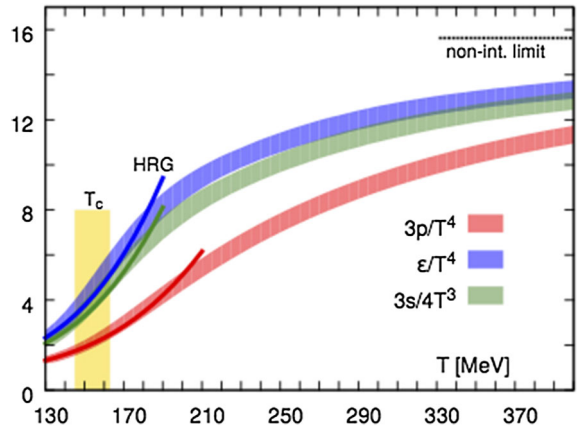

FIG. 64. Continuum extrapolated results for pressure, energy density, and entropy density. Solid lines on the low temperature side correspond to results obtained from hadron resonance gas (HRG) model calculations. The (yellow) band marked T_c indicates the phase transition region for deconfinement and chiral symmetry restoration.

a condition that any object with (electric) color charge cannot be in the spectrum of states in the QCD vacuum, since it must be produced accompanied by a flux tube which carries the electric flux to another, oppositely color charged, object. These flux tubes are dual to those observed in superconductors: they carry electric flux, not a magnetic one. The stability of flux tubes in superconductors against transverse expansion is produced by a "coil" with *supercurrent*, made by (electrically charged) Cooper pairs running around the tube. The QCD flux tubes also have a coil with supercurrent around them, also dual, made by magnetically charged objects. Detailed studies of these tubes have been made by lattice numerical simulations. Also similar to superconductivity, confinement goes away at sufficiently large temperatures $T > T_c$: this phenomenon is called "deconfinement."

Chiral symmetry breaking and restoration: The quark mass term is the only one in the QCD Lagrangian connecting right- and left-handed polarizations of the quark fields. For light u, d, s quarks one can as an approximation neglect the masses: in this case $U(3)$ flavor rotation symmetry is doubled to $U(3)_{\rm left} \times U(3)_{\rm right}$ larger symmetry. Extra new symmetry created by relative counterrotations of the left and right parts is known as the *chiral symmetries*, divided into an overall phase $U(1)_{\rm axial}$ and rotations $SU(3)_{\rm axial}$. The former one is not actually a symmetric since it is violated by axial anomaly (instantons). The second one is broken *spontaneously* in the QCD vacuum by quark-antiquark pairing. At sufficiently high temperatures $T > T_\chi$ the condensate created by pairing disappears, and this phenomenon is called "chiral symmetry restoration." In real-world QCD with quarks it turns out that T_c of deconfinement and T_χ are too close to tell the difference, so both are mentioned in the text simply as T_c. However, for QCD-like theories with a different number or color charge of the quarks lattice studies found that these two phase transitions can be separate, sometimes by a large factor.

Temperature range scanned in heavy ion experiments: The matter produced at RHIC and the LHC has the initial temperature $T \approx 2T_c$, and the final one, at the kinetic

freezeouts of the largest systems, is low as $T \approx 0.5T_c$. Thus the near-T_c phenomena play a significant role.

APPENDIX B: RELATIVISTIC HYDRODYNAMICS

Relativistic hydrodynamics is used to describe collective effects which are absent in elementary processes, such as, say e^+e^- annihilation into hadrons or (minimum bias) pp collisions. The explosions seen in heavy ion collisions are often called "the little bang," in analogy with the "big bang" in cosmology, with which it shares many concepts. One of them is the idea of "smooth average behavior" on top of which there are "perturbations," different on an event-by-event basis in the little bang and on specific locations in the big bang. The former is described by "full hydrodynamics" and the latter sometimes by a linearized version on top of the smooth solution.

The conceptual basis of the hydrodynamics is simple: it is just a set of local conservation laws for the stress tensor ($T^{\mu\nu}$) and for the conserved currents (J_i^μ),

$$\partial_\mu T^{\mu\nu} = 0, \qquad \partial_\mu J_i^\mu = 0. \tag{B1}$$

In equilibrium, $T^{\mu\nu}$ and J_i^μ are related to the bulk properties of the fluid by

$$\begin{aligned} T^{\mu\nu} &= (\epsilon + p)u^\mu u^\nu - pg^{\mu\nu}, \\ J_i^\mu &= n_i u^\mu. \end{aligned} \tag{B2}$$

Here ϵ is the energy density, p is the pressure, n_i is the number density of the corresponding current, and $u^\mu = \gamma(1, \mathbf{v})$ is the proper velocity of the fluid. In strong interactions, the conserved currents are isospin (J_I^μ), strangeness (J_S^μ), and baryon number (J_B^μ). For the hydrodynamic evolution, isospin symmetry is assumed and the net strangeness is set to zero; therefore only the baryon current J_B is considered.

In order to close up this set of equations, one also needs the equation of state (EoS) $p(\epsilon)$. One should also be aware of two thermodynamical differentials

$$d\epsilon = Tds, \qquad dp = sdT, \tag{B3}$$

and the definition of the sound velocity

$$c_s^2 = \frac{\partial p}{\partial \epsilon} = \frac{s}{T}\frac{\partial T}{\partial s}, \tag{B4}$$

and that $\epsilon + p = Ts$. Using these equations and the thermodynamical relations in the form

$$\frac{\partial_\mu \epsilon}{\epsilon + p} = \frac{\partial_\mu s}{s}, \tag{B5}$$

one can show that these equations imply another nontrivial conservation law, namely, the conservation of the entropy

$$\partial_\mu (s u_\mu) = 0. \tag{B6}$$

Therefore in the idealized adiabatic flow all the entropy is produced only in the discontinuities such as shock waves.

In an arbitrary coordinate system the equations of motion can be written as

$$T^{mn}{}_{;m} = 0, \qquad j^m{}_{;m} = 0, \tag{B7}$$

where the semicolon indicates a covariant derivative. For tensors of rank 1 and 2 it reads explicitly

$$j^i_{;p} = j^i_{,p} + \Gamma^i_{pk} j^k, \tag{B8}$$

$$T^{ik}{}_{;p} = T^{ik}{}_{,p} + \Gamma^i_{pm} T^{mk} + \Gamma^k_{pm} T^{im}, \tag{B9}$$

where the comma denotes a simple partial derivative and the Christoffel symbols Γ^s_{ij} are given by derivatives of the metric tensor $g^{ab}(x)$:

$$\Gamma^s_{ij} = (1/2) g^{ks}(g_{ik,j} + g_{jk,i} - g_{ij,k}). \tag{B10}$$

As an example, let us do the following transformation from Cartesian to light cone coordinates:

$$\begin{aligned} x^\mu = (t, x, y, z) \to \bar{x}^m = (\tau, x, y, \eta), \\ t = \tau \cosh\eta, \qquad \tau = \sqrt{t^2 - z^2}, \end{aligned} \tag{B11}$$

$$z = \tau \sinh\eta, \qquad \eta = (1/2) \ln \frac{t+z}{t-z}. \tag{B12}$$

In the new coordinate system the velocity field (after inserting $v_z = z/t$) is given by

$$\bar{u}^m = \bar{\gamma}(1, \bar{v}_x, \bar{v}_y, 0) \tag{B13}$$

with $\bar{v}_i \equiv v_i \cosh\eta$, $i = x$, y, and $\bar{\gamma} \equiv 1/\sqrt{1 - \bar{v}_x^2 - \bar{v}_y^2}$.

Now we turn to the metric of the new system. We have

$$\begin{aligned} ds^2 = g_{\mu\nu} dx^\mu dx^\nu &= dt^2 - dx^2 - dy^2 - dz^2 \\ &= d\tau^2 - dx^2 - dy^2 - \tau^2 d\eta^2, \end{aligned} \tag{B14}$$

and therefore

$$g_{mn} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -\tau^2 \end{pmatrix}. \tag{B15}$$

The only nonvanishing Christoffel symbols are

$$\Gamma^\eta_{\eta\tau} = \Gamma^\eta_{\tau\eta} = \frac{1}{\tau}, \qquad \Gamma^\tau_{\eta\eta} = \tau. \tag{B16}$$

Dissipative corrections to the stress tensor and the current can be written as follows:

$$\delta T_{\mu\nu} = \eta(\nabla_\mu u_\nu + \nabla_\nu u_\mu - \tfrac{2}{3}\Delta_{\mu\nu}\nabla_\rho u_\rho) + \xi(\Delta_{\mu\nu}\nabla_\rho u_\rho), \tag{B17}$$

$$\delta J_\mu = k\left(\frac{\eta T}{\epsilon + p}\right)^2 \nabla_\mu(\mu_B/T), \tag{B18}$$

where the three coefficients η, ξ, and k are called the shear and the bulk viscosities and the heat conductivity, respectively. In these definitions the following projection operator onto the matter rest frame was used:

$$\nabla_\mu \equiv \Delta_{\mu\nu}\partial_\nu, \qquad \Delta_{\mu\nu} \equiv g_{\mu\nu} - u_\mu u_\nu. \tag{B19}$$

It is further useful to relate the magnitude of the viscosity coefficient η to a more physical observable. As such one can use the sound attenuation length. If a sound wave has frequency ω and wave vector $\vec{q}$, its dispersion law (the pole position) is

$$\omega = c_s q - \frac{i}{2}\vec{q}^2\Gamma_s, \qquad \Gamma_s \equiv \frac{4}{3}\frac{\eta}{\epsilon + p}. \tag{B20}$$

The Navier-Stokes term is the first-order expansion in gradients: it has some issues with causality and in practice some second-order hydrodynamic equations are used; for more information, see, e.g., Romatschke (2010). Attempts to do resummation of all gradient terms (Lublinsky and Shuryak, 2009) are discussed in the section on holography.

The original Landau paper focused on "longitudinal" flow and what we now call rapidity distribution. In the heavy ion domain we focus mostly on a "splash" in the transverse plane: a collective transverse velocity of up to $0.8c$ is observed, and thus it also requires relativistic hydrodynamics.

1. Bjorken flow

The idea of a rapidity-independent "scaling" distribution of secondaries originates from Feynman's early discussion of the parton model, around 1970. The existence of a rapidity-independent hydrodynamic solution was perhaps first noticed by Landau, who used a rapidity variable in his classic paper as a somewhat trivial case. The space-time picture connected with such scaling regime was discussed by Chiu, Sudarshan, and Wang (1975) and Gorenshtein, Zhdanov, and Sinjukov (1978) before Bjorken's famous paper (Bjorken, 1983) in which the solution was explicitly spelled out.

It is instructive first to describe it in the original Cartesian coordinates. There is no dependence on transverse coordinates x and y, only on time t and longitudinal coordinate z. The $1+1$D equations $\partial_\mu T^{\mu\nu} = 0$ can be rewritten in the following way:

$$\frac{\partial}{\partial t}(s\cosh y) + \frac{\partial}{\partial z}(s\sinh y) = 0, \tag{B21}$$

$$\frac{\partial}{\partial t}(T\sinh y) + \frac{\partial}{\partial z}(T\cosh y) = 0, \tag{B22}$$

where $u_\mu = (\cos h(y), \sin h(y))$, and T and s are the temperature and the energy density. The first equation manifests the entropy conservation.

The central point is the 1D Hubble ansatz for the 4-velocity

$$u_\mu = (t, 0, 0, z)/\tau, \tag{B23}$$

where $\tau^2 = t^2 - z^2$ is the proper time. Note that all volume elements are expanded linearly with time and move along straight lines from the collision point. The spatial $\eta = \tanh^{-1}(z/t)$ and the momentum rapidities $y = \tanh^{-1} v$ are just equal to each other. Exactly as in the big bang, for each "observer" (the volume element) the picture is just the same, with the pressure from the left compensated by that from the right. The history is also the same for all volume elements, if it is expressed in its own proper time τ. Thus one has $s(\tau)$, $T(\tau)$. Using this ansatz, the entropy conservation becomes an ordinary differential equation in proper time τ:

$$\frac{ds(\tau)}{d\tau} + \frac{s}{\tau} = 0 \tag{B24}$$

with a solution

$$s = \frac{\text{const}}{\tau}. \tag{B25}$$

So far all dissipative phenomena were ignored. Including the first dissipative terms into our equations one finds the following source for the entropy current:

$$\frac{1}{\epsilon + p}\frac{d\epsilon}{d\tau} = \frac{1}{s}\frac{ds}{d\tau} = -\frac{1}{\tau}\left(1 - \frac{(4/3)\eta + \xi}{(\epsilon + p)\tau}\right) \tag{B26}$$

with shear and bulk viscosities η and ξ, which tells us that one has to abandon ideal hydrodynamics at sufficiently early time.

Alternatively, one can start with curved coordinates τ and η from the beginning, and look for an η-independent solution. Those are comoving coordinates, in those $u_\mu = (1, 0, 0, 0)$ but the equations obtain an extra term from Christoffel symbols.

2. Gubser flow

The Gubser flow (Gubser, 2010; Gubser and Yarom, 2011) is a solution which keeps the boost invariance and the axial symmetry in the transverse plane of the Bjorken flow, but replaces the translational invariance in the transverse plane by symmetry under special conformal transformation. Therefore, one restriction is that the matter is required to be conformal, with the EOS $\epsilon = 3p$. Another is that the colliding systems has to be of a particular shape, corresponding to a conformal map of the sphere onto the transverse plane.

The solution of ideal hydrodynamics has three parameters: One is dimensional q, it defines the size of the system (and is roughly corresponding to the radii of the colliding nuclei). The other two are dimensionless, f^* characterizes the number of degrees of freedom in the matter, and $\hat{e}_0$ the amount of entropy in the system.

The original setting uses the coordinates the proper time, spatial rapidity, transverse radius, and azimuthal angle $(\bar{\tau}, \eta, \bar{r}, \phi)$ with the metric

$$ds^2 = -d\bar{\tau}^2 + \bar{\tau}^2 d\eta^2 + d\bar{r}^2 + \bar{r}^2 d\phi^2. \tag{B27}$$

The dimensionless coordinates $\bar{\tau} = q\tau$, $\bar{r} = qr$ are rescaled versions of the actual coordinates.

Looking for solutions independent of both "angles" η, ϕ and using transverse rapidity

$$u_\mu = (-\cosh\kappa(\tau,r), 0, \sinh\kappa(\tau,r), 0), \tag{B28}$$

Gubser obtained the following solution:

$$v_\perp = \tanh\kappa(\tau,r) = \left(\frac{2q^2\tau r}{1+q^2\tau^2+q^2r^2}\right), \tag{B29}$$

$$\epsilon = \frac{\hat{\epsilon}_0 (2q)^{8/3}}{\tau^{4/3}[1+2q^2(\tau^2+r^2)+q^4(\tau^2-r^2)^2]^{4/3}}, \tag{B30}$$

where $\hat{\epsilon}_0$ is the second parameter. Gubser and Yarom (2011) rederived the same solution by going into the comoving frame. In order to do so they rescaled the metric

$$ds^2 = \tau^2 d\hat{s}^2 \tag{B31}$$

and performed a coordinate transformation from the τ, r to a new set ρ, θ given by

$$\sinh\rho = -\frac{1-q^2\tau^2+q^2r^2}{2q\tau}, \tag{B32}$$

$$\tan\theta = \frac{2qr}{1+q^2\tau^2-q^2r^2}. \tag{B33}$$

In the new coordinates the rescaled metric reads

$$d\hat{s}^2 = -d\rho^2 + \cosh^2\rho(d\theta^2+\sin^2\theta d\phi^2) + d\eta^2 \tag{B34}$$

and we use ρ as the "new time" coordinate and θ as a new "space" coordinate. In the new coordinates the fluid is at rest.

The relation between the velocity in Minkowski space in the (τ, r, ϕ, η) coordinates and the one in the rescaled metric in $(\rho, \theta, \phi, \eta)$ coordinates corresponds to

$$u_\mu = \tau \frac{\partial \hat{x}^\nu}{\partial \hat{x}^\mu}\hat{u}_\nu, \tag{B35}$$

while the energy density transforms as $\epsilon = \tau^{-4}\hat{\epsilon}$.

The temperature [in the rescaled frame, $\hat{T} = \tau f_*^{1/4} T$, with $f_* = \epsilon/T^4 = 11$ as in Gubser (2010)] is now dependent only on the new time ρ, in the case with nonzero viscosity the solution is

$$\hat{T} = \frac{\hat{T}_0}{(\cosh\rho)^{2/3}} + \frac{H_0 \sinh^3\rho}{9(\cosh\rho)^{2/3}} {}_2F_1\left(\frac{3}{2},\frac{7}{6};\frac{5}{2};-\sinh^2\rho\right), \tag{B36}$$

where H_0 is a dimensionless constant made out of the shear viscosity and the temperature, $\eta = H_0 T^3$, and ${}_2F_1$ is the hypergeometric function. In the inviscid case the solution is just the first term of Eq. (B36), and it also conserves entropy in this case. The picture of the explosion is obtained by transformation from this expression back to τ and r coordinates.

Small perturbations to the Gubser flow obey linearized equations which have also been derived by Gubser and Yarom (2011). We start with the zero viscosity case, so that the background temperature (now called T_0) will be given by just the first term in Eq. (B36). The perturbations over the previous solution are defined by

$$\hat{T} = \hat{T}_0(1+\delta), \tag{B37}$$

$$u_\mu = u_{0\mu} + u_{1\mu}, \tag{B38}$$

with

$$\hat{u}_{0\mu} = (-1,0,0,0), \tag{B39}$$

$$\hat{u}_{1\mu} = (0, u_\theta(\rho,\theta,\phi), u_\phi(\rho,\theta,\phi), 0), \tag{B40}$$

$$\delta = \delta(\rho,\theta,\phi). \tag{B41}$$

Plugging Eqs. (B37) and (B38) into the hydrodynamic equations and only keeping linear terms in the perturbation, one can get a system of coupled first-order differential equations. Furthermore, if one ignores the viscosity terms, one may exclude velocity and get the following (second-order) closed equation for the temperature perturbation:

$$\frac{\partial^2\delta}{\partial\rho^2} - \frac{1}{3\cosh^2\rho}\left(\frac{\partial^2\delta}{\partial\theta^2} + \frac{1}{\tan\theta}\frac{\partial\delta}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2\delta}{\partial\phi^2}\right) + \frac{4}{3}\tanh\rho\frac{\partial\delta}{\partial\rho} = 0. \tag{B42}$$

Since the initial perturbations are assumed to be rapidity independent, we also ignored this coordinate here.

It has a number of remarkable properties: all three coordinates can be separated $\delta(\rho,\theta,\phi) = R(\rho)\Theta(\theta)\Phi(\phi)$ and a general solution is given by

$$\begin{aligned} R(\rho) &= \frac{C_1}{(\cosh\rho)^{2/3}} P^{2/3}_{-1/2+1/6\sqrt{12\lambda+1}}(\tanh\rho) \\ &\quad + \frac{C_2}{(\cosh\rho)^{2/3}} Q^{2/3}_{-1/2+1/6\sqrt{12\lambda+1}}(\tanh\rho), \\ \Theta(\theta) &= C_3 P_l^m(\cos\theta) + C_4 Q_l^m(\cos\theta), \\ \Phi(\phi) &= C_5 e^{im\phi} + C_6 e^{-im\phi}, \end{aligned} \tag{B43}$$

where $\lambda = l(l+1)$ and P and Q are associated Legendre polynomials. The parts of the solution depending on θ and ϕ can be combined in order to form spherical harmonics $Y_{lm}(\theta,\phi)$, such that $\delta(\rho,\theta,\phi) \propto R_l(\rho)Y_{lm}(\theta,\phi)$. This property should have been anticipated, as one of the main ideas of Gubser has been to introduce a coordinate which together with ϕ make a map on a 2D sphere.

Gubser flow was used as a theoretical laboratory ever since. A complete Green's function was constructed (Staig and Shuryak, 2011b), leading to pictures of sound circles as discussed at the beginning of this review. Generalization to perturbations by the quenching jets, with the sounds propagating in the rapidity direction, was done by Shuryak and Staig (2013a). For the second order (the Israel-Stuart version) of the hydrodynamics it was done by Marrochio *et al.* (2015) and Pang *et al.* (2015). A Boltzmann equation (in τ approximation) was also solved in such a setting; see Denicol *et al.* (2014) and the discussion in Sec. IV.C.

There are also a number of phenomenological applications. Without going into these, we just comment that we are limited by the fact that at large r the power tail of the solution is

completely inadequate for heavy ion collisions. As such, the Gubser solution is like an explosion in the atmosphere, while the real ones are in vacuum. As a result, in applications one basically needs to amputate the unphysical regions by hand.

APPENDIX C: INTRODUCTION TO GAUGE-GRAVITY DUALITY

The starting point of this development was (i) the discovery of AdS/CFT correspondence (Maldacena, 1999), which a decade later became an important tool for nuclear theorists, as well as a prime example of the "applications" in the string theory community. Here we briefly outline for nonspecialists the logics of several important developments, which is needed to understand the main text section devoted to gauge-gravity calculations. Those include (ii) the thermodynamics of strongly coupled $N=4$ plasma (Gubser, Klebanov, and Tseytlin, 1998) and (iii) the transport properties derived from the linearized hydrodynamics (Policastro, Son, and Starinets, 2002). Another significant achievement (iv) was the general derivation of the full nonlinear hydrodynamics from the gradient expansion of the Einstein equations (Natsuume and Okamura, 2008; Bhattacharyya, Hubeny, Loganayagam *et al.*, 2008; Bhattacharyya, Hubeny, Minwalla, and Rangamani, 2008).

(i) The AdS/CFT correspondence has been discovered in studies of certain string theory construction. While strings have one coordinate along them, and thus their world volume has codimension 2, usually parametrized by "internal coordinates" τ and σ, their dynamics is for certain consistency reasons studied in space-times with a much higher number of "external dimensions" D. String theory actually admits solitons called "branes" with a certain intermediate number of dimensions, e.g., D_3 branes with $3+1$ internal dimensions. The original construction contained N_c such D_3 branes stacked together at the same location in $D=10$ space-time.

Closed string massless excitations are known to include states with spin up to 2, described by certain supergravity. At large N_c the original stack generates a strong gravity described by classical GR. A particular solution of Einstein equation called $\mathrm{AdS}_5\times S^5$, where AdS_5 is anti–de Sitter five-dimensional space and S^5 is a five-dimensional sphere. The metric of AdS_5 does not depend on four coordinates of the brane space-time, but only on the fifth coordinate called z, and at $z=0$ it has a four-dimensional boundary. Maldacena (1999) had conjectured that since symmetries of conformal $\mathcal{N}=4$ supersymmetric Yang-Mills (SYM) theory on this 4D boundary and of the AdS_5 solution match uniquely, there must be certain "holographic" correspondence between them. In particular, gauge invariant colorless operators in the boundary theory (e.g., the stress tensor $T_{\mu\nu}$) should be related to fields in the five-dimensional "bulk" (such as the metric $g_{\mu\nu}$). Testing the conjecture was a popular occupation in the late 1990s: correlators in the $\mathcal{N}=4$ SYM theory were calculated and compared with those gravity calculations of certain propagators and multipoint Green's functions; all tests were positive and the conjecture was considered true. The important observation is that in certain limits the gauge theory has infinite 't Hooft coupling $\lambda=g^2N_c$ while the bulk theory is weakly coupled, since bulk fields are uncolored.

This discovery created a large industry which is divided into two directions. The "top-down" one looks for exact correspondence involving theories other than the $\mathcal{N}=4$ SYM theory: yet it seems to be impossible to find sufficiently convenient correspondence for nonsupersymmetric QCD-like theories. The "down-up" approach, also known as AdS/QCD, builds holographic models for any theories, without a string theory solution in the background. For a review of this approach, see, e.g., Gursoy and Kiritsis (2008). One bulk field in 5D, dilatons or gravity, generates many "radial excitations" of hadrons and their Regge trajectories. Such models include confinement and chiral symmetry breaking, and to certain accuracy they do reproduce the spectroscopy of mesons and glueballs.

(ii) The AdS_5 GR solution resembles the Schwarzschild one for a black hole, with a change of 4D Coulomb factors to 6D ones: these six dimensions are what is left from ten dimensions minus four on which no dependence is present. But the main difference is that string theory branes happen to have charges which make them "charged black holes," with certain vector fields added. Moreover, they are "extreme" black holes, with maximal allowed charge, as a result of which their horizon area (and thus entropy) vanishes. This is how it must be, since we intend to describe a vacuum state of the gauge theory on the boundary, and it should not be degenerate and have any entropy.

It was further found that excited states of the brane construction are described by nonextreme black holes with a nonzero horizon. Since the metric still depends on the fifth coordinate only, the horizon sits at some value z_h, and the physical region ranges from $z=0$ to z_h. Schwarzschild black holes emit Hawking radiation, and in a 4D flat universe they therefore must eventually emit all their energy and disappear. The "black brane" AdS_5 GR solution has a different fate: Hawking radiation heats up the Universe, including to its boundary, to certain equilibrium static case.

Projecting this solution to gauge theory one finds thermodynamics of strongly coupled $N=4$ plasma at nonzero temperature T. Unlike QCD-like theories (which develops a scale Λ_{QCD} via running coupling), the $N=4$ SYM has zero beta function and thus has no scale of its own. So its properties must obey trivial scaling given by the dimension, e.g., the energy density can only be $\epsilon=C(g)T^4$ with some dimensionless coefficient depending on the coupling. At zero coupling one has noninteracting gas or Stephan-Botzmann generalization of thermal radiation famously explained by Planck. It was found that (Gubser, Klebanov, and Tseytlin, 1998)

$$C(g^2N_c\to\infty)=\tfrac{3}{4}C(0). \tag{C1}$$

The coefficient $3/4$ is in fact in better agreement with the lattice equation of state, than 1 of ideal QGP.

(iii) The transport properties are derived in a number of ways. Small perturbations around the thermal AdS solution are elementary excitations of static plasma, such as sounds or transverse dissipative modes. Linearized perturbations of Einstein equations in the bulk correspond to solutions of the

linearized hydrodynamics on the boundary theory. Therefore one can find (Policastro, Son, and Starinets, 2001) the dissipation rate of these modes and thus the *shear viscosity*. (Bulk viscosity is zero because of scale invariance.) Since its dimension T^3 is the same as entropy density, most give the value of their ratio

$$\left.\frac{\eta}{s}\right|_{g^2N_c\to\infty} = \frac{1}{4\pi}. \tag{C2}$$

In a gravity setting it is clear what happens: the excitations are gravity waves which are simply falling into the black hole. On the gauge theory side we have no intuition of how this small number appears, but it is not far from experiment.

(iv) In the 19th century the hydrodynamics was an advanced theoretical field, teaching one how to work with partial differential equations, potentials, and rotational flows. Stokes was one of Maxwell's teachers, and electrodynamics clearly has benefited from hydrodynamical methods. Landau introduced relativistic hydrodynamics into the field of high energy collisions. Yet from the 1970s to about 2000 hydrodynamics was ridiculed by high energy theorists, as a simplistic approach incompatible with QCD and QFT's in general.

Apparently, this is no longer so, and one is allowed to mention it now. In particular, relativistic hydrodynamics of strongly coupled $N = 4$ plasma was derived by Natsuume and Okamura (2008) and Bhattacharyya, Hubeny, Loganayagam *et al.* (2008)) as a solution to the Einstein equation with the gradient expansion method. If the scale of inhomogeneity R is large compared to the horizon location z_h, one can think of smoothly varying horizon $z_h(x^\mu)$, $\mu = 0, 1, 2, 3$. This variation may or may not be small: in the latter case all nonlinear terms of hydrodynamics naturally appear as they should. And, last but not least, unlike phenomenological hydrodynamics (say, of water), all the kinetic coefficients, order by order, got their definite values. About a dozen of them have been calculated so far.

REFERENCES

Aamodt, K., *et al.* (ALICE Collaboration), 2011a, Phys. Rev. D **84**, 112004.

Aamodt, K., *et al.* (ALICE Collaboration), 2011b, Phys. Rev. Lett. **107**, 032301.

Adcox, K., *et al.* (PHENIX Collaboration), 2005, Nucl. Phys. A **757**, 184.

Ade, P. A. R., *et al.* (Planck Collaboration), 2014, Astron. Astrophys. **571**, A15.

Adler, S. S., *et al.* (PHENIX Collaboration), 2014, Phys. Rev. C **89**, 044905.

Alba, P., W. Alberico, R. Bellwied, M. Bluhm, V. Mantovani Sarti, M. Nahrgang, and C. Ratti, 2014, Phys. Lett. B **738**, 305.

Andronic, A., P. Braun-Munzinger, K. Redlich, and J. Stachel, 2007, "Proceedings, 4th International Workshop on Critical point and onset of deconfinement (CPOD07)," *Proc. Sci.* CPOD07, 044.

Arean, D., I. Iatrakis, M. Jarvinen, and E. Kiritsis, **2013**, J. High Energy Phys. 11, 068.

Arnold, P. B., G. D. Moore, and L. G. Yaffe, 2003, J. High Energy Phys. 05, 051.

Baier, R., A. H. Mueller, D. Schiff, and D. T. Son, 2001, Phys. Lett. B **502**, 51.

Baker, M., J. S. Ball, and F. Zachariasen, 1997, Phys. Rev. D **56**, 4400.

Balasubramanian, V., A. Bernamonti, J. de Boer, N. Copland, B. Craps, E. Keski-Vakkuri, B. Muller, A. Schafer, M. Shigemori, and W. Staessens, 2011, Phys. Rev. D **84**, 026010.

Basar, G., D. E. Kharzeev, and E. V. Shuryak, 2014, Phys. Rev. C **90**, 014905.

Basar, G., D. E. Kharzeev, H.-U. Yee, and I. Zahed, 2012, Phys. Rev. D **85**, 105005.

Basar, G., and D. Teaney, 2014, Phys. Rev. C **90**, 054903.

Bebie, H., P. Gerber, J. L. Goity, and H. Leutwyler, 1992, Nucl. Phys. B **378**, 95.

Begun, V., W. Florkowski, and M. Rybczynski, 2014, Phys. Rev. C **90**, 014906.

Bellwied, R. (STAR Collaboration), 2005, Nucl. Phys. A **752**, 398.

Berges, J., K. Boguslavski, S. Schlichting, and R. Venugopalan, 2015, arXiv:1508.03073.

Bhalerao, R. S., and J.-Y. Ollitrault, 2006, Phys. Lett. B **641**, 260.

Bhattacharyya, S., V. E. Hubeny, R. Loganayagam, G. Mandal, S. Minwalla, T. Morita, M. Rangamani, and H. S. Reall, **2008**, J. High Energy Phys. 06, 055.

Bhattacharyya, S., V. E. Hubeny, S. Minwalla, and M. Rangamani, 2008, J. High Energy Phys. 02, 045.

Bjorken, J. D., 1982, Report No. FERMILAB-PUB-82-059-THY, and Report No. FERMILAB-PUB-82-059-T.

Bjorken, J. D., 1983, Phys. Rev. D **27**, 140.

Bjorken, J. D., S. J. Brodsky, and A. Scharff Goldhaber, 2013, Phys. Lett. B **726**, 344.

Blaizot, J.-P., Y. Mehtar-Tani, and M. A. C. Torres, 2015, Phys. Rev. Lett. **114**, 222002.

Blaizot, J.-P., and J.-Y. Ollitrault, 1990, Adv. Ser. Dir. High Energy Phys. **6**, 393.

Boulware, D. G., L. S. Brown, R. N. Cahn, S. D. Ellis, and C.-k. Lee, 1976, Phys. Rev. D **14**, 2708.

Bozek, P., 2012, Phys. Rev. C **85**, 014911.

Bozek, P., and W. Broniowski, 2014, Phys. Lett. B **739**, 308.

Brambilla, N., M. A. Escobedo, J. Ghiglieri, and A. Vairo, **2013**, J. High Energy Phys. 05, 130.

Braun-Munzinger, P., J. Stachel, and C. Wetterich, 2004, Phys. Lett. B **596**, 61.

Chatrchyan, S., *et al.* (CMS Collaboration), 2013, Phys. Lett. B **718**, 795.

Chatrchyan, S., *et al.* (CMS Collaboration), 2014, Eur. Phys. J. C **74**, 2847.

Chesler, P. M., and L. G. Yaffe, **2014**, J. High Energy Phys. 07, 086.

Chiu, C. B., E. C. G. Sudarshan, and K.-H. Wang, 1975, Phys. Rev. D **12**, 902.

Chiu, M., T. K. Hemmick, V. Khachatryan, A. Leonidov, J. Liao, and L. McLerran, 2013, Nucl. Phys. A **900**, 16.

Csernai, L. P., F. Becattini, and D. J. Wang, 2014, J. Phys. Conf. Ser. **509**, 012054.

Csernai, L. P., J. Kapusta, and L. D. McLerran, 2006, Phys. Rev. Lett. **97**, 152303.

D'Alessandro, A., M. D'Elia, and E. V. Shuryak, 2010, Phys. Rev. D **81**, 094501.

de Cassagnac, R., *et al.* (CMS Collaboration), 2014, Nucl. Phys. A **931**, 22.

Denicol, G. S., U. W. Heinz, M. Martinez, J. Noronha, and M. Strickland, 2014, Phys. Rev. D **90**, 125026.

Ding, H.-T., F. Karsch, and S. Mukherjee, 2015, Int. J. Mod. Phys. E **24**, 1530007.

Dirac, P. A. M., 1931, Proc. R. Soc. A **133**, 60.
Epelbaum, T., 2014, Nucl. Phys. A **931**, 337.
Ewerz, C., P. Lebiedowicz, O. Nachtmann, and A. Szczurek, 2016, Phys. Lett. B **763**, 382.
Farrar, G. R., and J. D. Allen, 2013, EPJ Web Conf. **53**, 07007.
Fermi, E., 1951, Phys. Rev. **81**, 683.
Floerchinger, S., and U. A. Wiedemann, **2011**, J. High Energy Phys. 11, 100.
Florkowski, W., and R. Ryblewski, 2011, Phys. Rev. C **83**, 034907.
Florkowski, W., and R. Ryblewski, 2014, Nucl. Phys. A **931**, 343.
Gelman, B. A., E. V. Shuryak, and I. Zahed, 2006a, Phys. Rev. C **74**, 044908.
Gelman, B. A., E. V. Shuryak, and I. Zahed, 2006b, Phys. Rev. C **74**, 044909.
Goldhaber, G., S. Goldhaber, W.-Y. Lee, and A. Pais, 1960, Phys. Rev. **120**, 300.
Gorda, T., and P. Romatschke, 2014, Phys. Rev. C **90**, 054908.
Gorenshtein, M., V. A. Zhdanov, and Y. M. Sinjukov, 1978, Sov. Phys. JETP **74**, 833.
Gross, D. J., R. D. Pisarski, and L. G. Yaffe, 1981, Rev. Mod. Phys. **53**, 43.
Grosse-Oetringhaus, J. F. (ALICE Collaboration), 2014, Nucl. Phys. A **931**, 22.
Gubser, S. S., 2010, Phys. Rev. D **82**, 085027.
Gubser, S. S., I. R. Klebanov, and A. A. Tseytlin, 1998, Nucl. Phys. B **534**, 202.
Gubser, S. S., S. S. Pufu, and A. Yarom, 2008, Phys. Rev. D **78**, 066014.
Gubser, S. S., S. S. Pufu, and A. Yarom, **2009**, J. High Energy Phys. 11, 050.
Gubser, S. S., and A. Yarom, 2011, Nucl. Phys. B **846**, 469.
Gursoy, U., and E. Kiritsis, **2008**, J. High Energy Phys. 02, 032.
Gursoy, U., E. Kiritsis, and F. Nitti, 2008, J. High Energy Phys. 02, 019.
Heinz, U., and R. Snellings, 2013, Annu. Rev. Nucl. Part. Sci. **63**, 123.
Heller, M. P., R. A. Janik, and P. Witaszczyk, 2012, Phys. Rev. D **85**, 126002.
Hidaka, Y., and R. D. Pisarski, 2008, Phys. Rev. D **78**, 071501.
Hirano, T., U. W. Heinz, D. Kharzeev, R. Lacey, and Y. Nara, 2006, Phys. Lett. B **636**, 299.
Hirono, Y., and E. Shuryak, 2015, Phys. Rev. C **91**, 054915.
Hung, C. M., and E. V. Shuryak, 1998, Phys. Rev. C **57**, 1891.
Iatrakis, I., A. Ramamurti, and E. Shuryak, 2015, Phys. Rev. D **92**, 014011.
Iatrakis, I., A. Ramamurti, and E. Shuryak, 2016, Phys. Rev. D **94**, 045005.
Iritani, T., G. Cossu, and S. Hashimoto, 2014, *Proc. Sci.* LATTICE2013, 376.
Janik, R. A., and R. B. Peschanski, 2006, Phys. Rev. D **73**, 045013.
Jia, J. (ATLAS Collaboration), 2011, J. Phys. G **38**, 124012.
Kaczmarek, O., and F. Zantow, 2006, *Proc. Sci.* LAT2005, 192.
Kajantie, K., M. Laine, K. Rummukainen, and Y. Schroder, 2003, Phys. Rev. D **67**, 105008.
Kalaydzhyan, T., and E. Shuryak, 2014a, Phys. Rev. C **90**, 014901.
Kalaydzhyan, T., and E. Shuryak, 2014b, arXiv:1407.3270.
Kalaydzhyan, T., and E. Shuryak, 2014c, Phys. Rev. D **90**, 025031.
Kalaydzhyan, T., and E. Shuryak, 2015, Phys. Rev. D **91**, 083502.
Keegan, L., A. Kurkela, A. Mazeliauskas, and D. Teaney, **2016**, J. High Energy Phys. 08, 171.
Khachatryan, V., *et al.* (CMS Collaboration), **2010**, J. High Energy Phys. 09, 091.
Kharzeev, D. E., 2014, Phys. Rev. D **90**, 074007.
Kopylov, G. I., and M. I. Podgoretsky, 1974, Sov. J. Nucl. Phys. **18**, 336.
Kovtun, P., 2012, J. Phys. A **45**, 473001.
Kovtun, P., G. D. Moore, and P. Romatschke, 2011, Phys. Rev. D **84**, 025006.
Kovtun, P. K., and A. O. Starinets, 2005, Phys. Rev. D **72**, 086009.
Kraan, T. C., and P. van Baal, 1998, Phys. Lett. B **435**, 389.
Kuraev, E. A., L. N. Lipatov, and V. S. Fadin, 1977, Sov. Phys. JETP **45**, 199.
Kurkela, A., and Y. Zhu, 2015, Phys. Rev. Lett. **115**, 182301.
Lacey, R. A., Y. Gu, X. Gong, D. Reynolds, N. N. Ajitanand, J. M. Alexander, A. Mwai, and A. Taranenko, 2013, arXiv:1301.0165.
Laine, M., O. Philipsen, P. Romatschke, and M. Tassler, **2007**, J. High Energy Phys. 03, 054.
Landau, L. D., 1953, Izv. Akad. Nauk Ser. Fiz. **17**, 51.
Lappi, T., A. Dumitru, and Y. Nara, 2014, Nucl. Phys. A **931**, 354.
Larsen, R., and E. Shuryak, 2016, Phys. Rev. D **93**, 054029.
Lee, K.-M., and C.-h. Lu, 1998, Phys. Rev. D **58**, 025011.
Liao, J., and E. Shuryak, 2007, Phys. Rev. C **75**, 054907.
Liao, J., and E. Shuryak, 2008, Phys. Rev. Lett. **101**, 162302.
Liao, J., and E. Shuryak, 2009, Phys. Rev. Lett. **102**, 202302.
Liao, J., and E. Shuryak, 2010, Phys. Rev. D **82**, 094007.
Lin, S., and E. Shuryak, 2008a, Phys. Rev. D **77**, 085014.
Lin, S., and E. Shuryak, 2008b, Phys. Rev. D **78**, 125018.
Lin, S., and E. Shuryak, 2008c, Phys. Rev. D **77**, 085013.
Lin, S., and E. Shuryak, 2009, Phys. Rev. D **79**, 124015.
Lin, S., and E. Shuryak, 2011, Phys. Rev. D **83**, 045025.
Liu, Y., E. Shuryak, and I. Zahed, 2015, Phys. Rev. D **92**, 085007.
Lublinsky, M., and E. Shuryak, 2009, Phys. Rev. D **80**, 065026.
Lublinsky, M., and E. Shuryak, 2011, Phys. Rev. C **84**, 061901.
Makhlin, A. N., and Yu. M. Sinyukov, 1988, Z. Phys. C **39**, 69.
Maldacena, J. M., 1999, Int. J. Theor. Phys. **38**, 1113 [Adv. Theor. Math. Phys. **2**, 231 (1998)].
Mandelstam, S., 1976, Phys. Rep. **23**, 245.
Marrochio, H., J. Noronha, G. S. Denicol, M. Luzum, S. Jeon, and C. Gale, 2015, Phys. Rev. C **91**, 014903.
Martinez, M., and M. Strickland, 2010, Phys. Rev. C **81**, 024906.
Matsui, T., and H. Satz, 1986, Phys. Lett. B **178**, 416.
Mazeliauskas, A., and D. Teaney, 2015, Phys. Rev. C **91**, 044902.
McLerran, L. D., and R. Venugopalan, 1994, Phys. Rev. D **49**, 3352.
Mocsy, A., P. Petreczky, and M. Strickland, 2013, Int. J. Mod. Phys. A **28**, 1340012.
Molnar, D., and P. Huovinen, 2005, Phys. Rev. Lett. **94**, 012302.
Moore, G. D., and D. Teaney, 2005, Phys. Rev. C **71**, 064904.
Nagle, J., A. Adare, S. Beckman, T. Koblesky, J. O. Koop, D. McGlinchey, P. Romatschke, J. Carlson, J. Lynn, and M. McCumber, 2014, Phys. Rev. Lett. **113**, 112301.
Nakamura, A., and S. Sakai, 2005, Phys. Rev. Lett. **94**, 072305.
Natsuume, M., and T. Okamura, 2008, Phys. Rev. D **77**, 066014.
Novikov, V. A., M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, 1981, Nucl. Phys. B **191**, 301.
Pang, L.-G., Y. Hatta, X.-N. Wang, and B.-W. Xiao, 2015, Phys. Rev. D **91**, 074027.
Policastro, G., D. T. Son, and A. O. Starinets, 2001, Phys. Rev. Lett. **87**, 081601.
Policastro, G., D. T. Son, and A. O. Starinets, **2002**, J. High Energy Phys. 09, 043.
Polyakov, A. M., 1974, JETP Lett. **20**, 194.
Polyakov, A. M., 1977, Nucl. Phys. B **120**, 429.
Pomeranchuk, I. Ya., 1951, Dokl. Akad. Nauk Ser. Fiz. **78**, 889.
Prakash, M., M. Prakash, R. Venugopalan, and G. Welke, 1993, Phys. Rep. **227**, 321.

Pratt, S., 2009, Phys. Rev. Lett. **102**, 232301.
Prino, F., and R. Rapp, 2016, arXiv:1603.00529.
Rapp, R., and E. V. Shuryak, 2001, Phys. Rev. Lett. **86**, 2980.
Ratti, C., and E. Shuryak, 2009, Phys. Rev. D **80**, 034004.
Romatschke, P., 2010, Int. J. Mod. Phys. E **19**, 1.
Romatschke, P., 2015, Eur. Phys. J. C **75**, 429.
Romatschke, P., 2016, arXiv:1609.02820.
Rose, J.-B., J.-F. Paquet, G. S. Denicol, M. Luzum, B. Schenke, S. Jeon, and C. Gale, 2014, Nucl. Phys. A **931**, 926.
Schafer, T., and E. V. Shuryak, 1998, Rev. Mod. Phys. **70**, 323.
Schenke, B., and R. Venugopalan, 2014, Phys. Rev. Lett. **113**, 102301.
Schwinger, J. S., K. A. Milton, W.-y. Tsai, L. L. DeRaad, Jr., and D. C. Clark, 1976, Ann. Phys. (N.Y.) **101**, 451.
Seiberg, N., and E. Witten, 1994, Nucl. Phys. B **426**, 19.
Shuryak, E., 2012a, arXiv:1203.1012.
Shuryak, E., 2012b, Phys. Rev. C **86**, 024907.
Shuryak, E., 2016, arXiv:1610.08789.
Shuryak, E., S.-J. Sin, and I. Zahed, 2007, J. Korean Phys. Soc. **50**, 384.
Shuryak, E., and P. Staig, 2013a, Phys. Rev. C **88**, 054903.
Shuryak, E., and P. Staig, 2013b, Phys. Rev. C **88**, 064905.
Shuryak, E., and I. Zahed, 2013, Phys. Rev. C **88**, 044915.
Shuryak, E., and I. Zahed, 2014, Phys. Rev. D **89**, 094001.
Shuryak, E. V., 1973, Yad. Fiz. **18**, 1302.
Shuryak, E. V., 1978, Yad. Fiz. **28**, 796 [Phys. Lett. B **78**, 150 (1978)].
Shuryak, E. V., 1980, Phys. Rep. **61**, 71.
Shuryak, E. V., 1998, Phys. Lett. B **423**, 9.
Shuryak, E. V., 2002, Phys. Rev. C **66**, 027902.
Shuryak, E. V., and L. Xiong, 1993, Phys. Rev. Lett. **70**, 2241.
Shuryak, E. V., and O. V. Zhirov, 1980, Phys. Lett. B **89**, 253.
Spalinski, M., 2016, Phys. Rev. D **94**, 085002.
Staig, P., and E. Shuryak, 2011a, Phys. Rev. C **84**, 034908.
Staig, P., and E. Shuryak, 2011b, Phys. Rev. C **84**, 044912.
Stephanov, M. A., 2009, Phys. Rev. Lett. **102**, 032301.
Stephanov, M. A., K. Rajagopal, and E. V. Shuryak, 1999, Phys. Rev. D **60**, 114028.
Stoffers, A., and I. Zahed, 2013, Phys. Rev. D **87**, 075023.
Svetitsky, B., 1988, Phys. Rev. D **37**, 2484.
Tannenbaum, M., 2014, Nucl. Phys. A **931**, 877.
Teaney, D., 2003, Phys. Rev. C **68**, 034913.
Teaney, D., J. Lauret, and E. V. Shuryak, 2001a, arXiv:nucl-th/0110037.
Teaney, D., J. Lauret, and E. V. Shuryak, 2001b, Phys. Rev. Lett. **86**, 4783.
Teaney, D., and L. Yan, 2012, Phys. Rev. C **86**, 044908.
't Hooft, G., 1974, Nucl. Phys. B **79**, 276.
't Hooft, G., 1978, Nucl. Phys. B **138**, 1.
Tuchin, K., 2013, Phys. Rev. C **88**, 024911.
Xu, J., J. Liao, and M. Gyulassy, 2015, arXiv:1508.00552.
Xu, Z., and C. Greiner, 2005, Phys. Rev. C **71**, 064901.
Young, C., J. I. Kapusta, C. Gale, S. Jeon, and B. Schenke, 2015, Phys. Rev. C **91**, 044901.
Young, C., and E. Shuryak, 2009, Phys. Rev. C **79**, 034907.