Large-scale flow in a cubic Rayleigh-Bénard cell: Long-term turbulence statistics and Markovianity of macrostate transitions

Priyanka Maity[1*]
[1] *Institute of Thermodynamics and Fluid Mechanics, Technische Universität Ilmenau, Postfach 100565, D-98684 Ilmenau, Germany*

Péter Koltai[2†]
Department of Mathematics, Freie Universität Berlin, Arnimallee 6, D-14195 Berlin, Germany

Jörg Schumacher[1,3‡]
[1] *Institute of Thermodynamics and Fluid Mechanics, Technische Universität Ilmenau, Postfach 100565, D-98684 Ilmenau, Germany*
[3] *Tandon School of Engineering, New York University, New York, NY 11201, USA*

We investigate the large-scale circulation (LSC) in a turbulent Rayleigh-Bénard convection flow in a cubic closed convection cell by means of direct numerical simulations at a Rayleigh number $Ra = 10^6$. The numerical studies are conducted for a single flow trajectory up to 10^5 convective free-fall times to obtain a sufficient sampling of the four discrete LSC states and the two crossover configurations which are taken in between for short periods. It is found that the statistics and time history depends strongly on the Prandtl number Pr of the working fluid which takes values of 0.1, 0.7, and 10. It changes from very rapid switches for the lowest Prandtl number to the spontaneous lock in one of the four states for the whole period for the largest one. Alternatively, we run ensembles of up to 1800 short-term simulations to study the transition probabilities between the discrete LSC states. This second approach is also used to probe the Markov property of the dynamics. The ensemble analysis revealed that the sample size might still be too small to conclude firmly the Markovianity of the transition process from one LSC state to another even though it is indicated.

PACS numbers: 47.20.Bp, 47.27-i., 02.50.Ga

* priyanka.maity@icts.res.in
† peter.koltai@fu-berlin.de
‡ joerg.schumacher@tu-ilmenau.de

I. INTRODUCTION

Turbulent convection is a classic example of dynamical system driven out of equilibrium which is omnipresent in nature; astrophysical [1], atmospheric, and oceanic convection [2] are the primary contributing factors for observed phenomenona in planets and stars. Contrary to the notion of presence of multiple scales and vigorous mixing due to turbulence, natural turbulent convection frequently exhibits formation of large scale structures and patterns, such as in clouds [3] and supergranules in Sun [4, 5]. On a physically attainable scale, such natural turbulent convection can be investigated using the simplified Rayleigh–Bénard convective (RBC) system, where a layer of fluid is confined between two plates with a thermal gradient [6–8]. Thermal convection originates due to density differences in the fluid as a result of the constant thermal gradient across the fluid layer, represented by the dimensionless Rayleigh number, which is given by

$$Ra = \frac{\alpha g\, \delta T H^3}{\nu \kappa}\,, \tag{1}$$

where α represents the isobaric thermal expansion coefficient, g the acceleration due to gravity, $\delta T = T_{\mathrm{bottom}} - T_{\mathrm{top}}$ the temperature difference maintained along the fluid layer of thickness H, and ν and κ being the kinematic viscosity and thermal diffusivity of the fluid, respectively. Convection begins when the temperature difference between the plates exceeds the critical value, $\delta T > \delta T_c$, which corresponds to the critical Rayleigh number of $Ra_c = 1708$ when the impermeable plates satisfy the no-slip boundary condition for the velocity [6]. The onset is generally characterized by laminar structures. Subsequent increment of Ra leads to a transition from laminar to turbulent region, which is characterized by thinning of boundary layers and generation of spatially extended coherent structures which are now denoted as superstructures of convection [9].

Turbulent RBC in confined geometries, such as cubic or cylindrical cells, also leads to an alignment of the rising and falling thermal plumes to form a coherent structure known as the large scale circulation or the mean wind of thermal convection [10–15]. The alignment of the plumes into an LSC and the increasing confinement towards a space direction can significantly influence the heat transport in the system [16–18]. In such cases the aspect ratio, which denotes the ratio of the horizontal extension (diameter or side length) to the height H, is decreased. Daya and Ecke took a further perspective on this subject and asked how the specific geometry at same aspect ratio alters the heat transfer [19]. Cubic closed cells are very specific since the LSC cannot drift azimuthally, but appears in a finite number of discrete macroscopic flow states [20–24]. In detail, the LSC appears preferentially in the form of circulation rolls along the two diagonals. This results in four states if one takes two orientations along each diagonal into account. Fast switches between these states proceed via four transient LSC states that are aligned with the side faces of the cell. While the configurations parallel to the diagonals are considered as stable, longer living LSC states of the turbulent convection flow, the ones along the side faces of the closed cube are unstable, short-living convection states.

One such attempt has been demonstrated in Giannakis *et al* [23] by a data-driven analysis of the eigenfunctions of the linear Koopman operator that describes the unitary time evolution of observables of the dynamical system rather than the nonlinear evolution of the states [25, 26]. The leading eigenvectors could be assigned to the 4 stable LSC states, the subsequent eigenstates were related to secondary flow structures in form of corner vortices that drive the system from one LSC state to another. However, the simulation time was still too short to obtain a sufficient number of LSC switches with 10^4 free-fall times, the characteristic convective time unit of the flow. A detailed analysis of these re-orientations sets the motivation of the present work.

The question that consequently arises is whether the re-orientation process of the LSC is indeed random and all discrete LSC states appear with the same probability. A further point is how this dynamics depends on the dimensionless Prandtl number, which is given by

$$Pr = \frac{\nu}{\kappa}\,, \tag{2}$$

and quantifies the ratio of viscous to temperature diffusion in the fluid.

In this work, we want to study these points by three-dimensional direct numerical simulations in two different ways. First, we follow a single long-term trajectory through the phase space of the convection flow for 10^5 free-fall times and determine the transitions between the different LSC states. This analysis is conducted for three different flow cases at a Rayleigh number $Ra = 10^6$ and Prandtl numbers $Pr = 0.1, 0.7$ and 10. We demonstrate a strong dependence of the large-scale flow behavior on Pr at the fixed Ra. Secondly, in the case of $Pr = 0.7$ an alternative approach to the LSC dynamics is presented which we term an *ensemble simulation*. Therefore, we take a coarse grid long-term trajectory that advances through phase space and start short-term ensemble simulations runs at full resolution from different outputs along the coarse grid run. The assumption in this approach (which cannot be proven) is that the coarse-grid trajectory "shadows" the true system evolution sufficiently well [27]. In the present study, we can actually take fully

resolved trajectory from the first analysis part, in studies with higher Rayleigh numbers this would not be possible. We then test if a Markov State Model (MSM) is able to describe the transitions between different LSC states [28, 29]. A time-discrete MSM describes a hopping process where the probability to get into a future state depends on the current state only, see [30, 31] for theory, [32] for applications, and [33, 34] for software implementations. This implies that we have to determine the transition probabilities among the different LSC states which will be described more specifically further below. The study also requires to determine the correct sampling time along the individual short-term trajectories that form an ensemble. Application of MSMs to stochastic system such as molecular kinetics [30] and protein folding [29] are known to show promising results, and recently it has been applied to RBC experimental data as well [35].

The manuscript is organized as follows. In section 2, we discuss in brief the numerical simulation model. Section 3 analysis the large-scale flow along the individual long trajectory. In section 4, we explain the details of the MSM and probe the Markov state property for the present system. The last section contains a summary and a brief outlook.

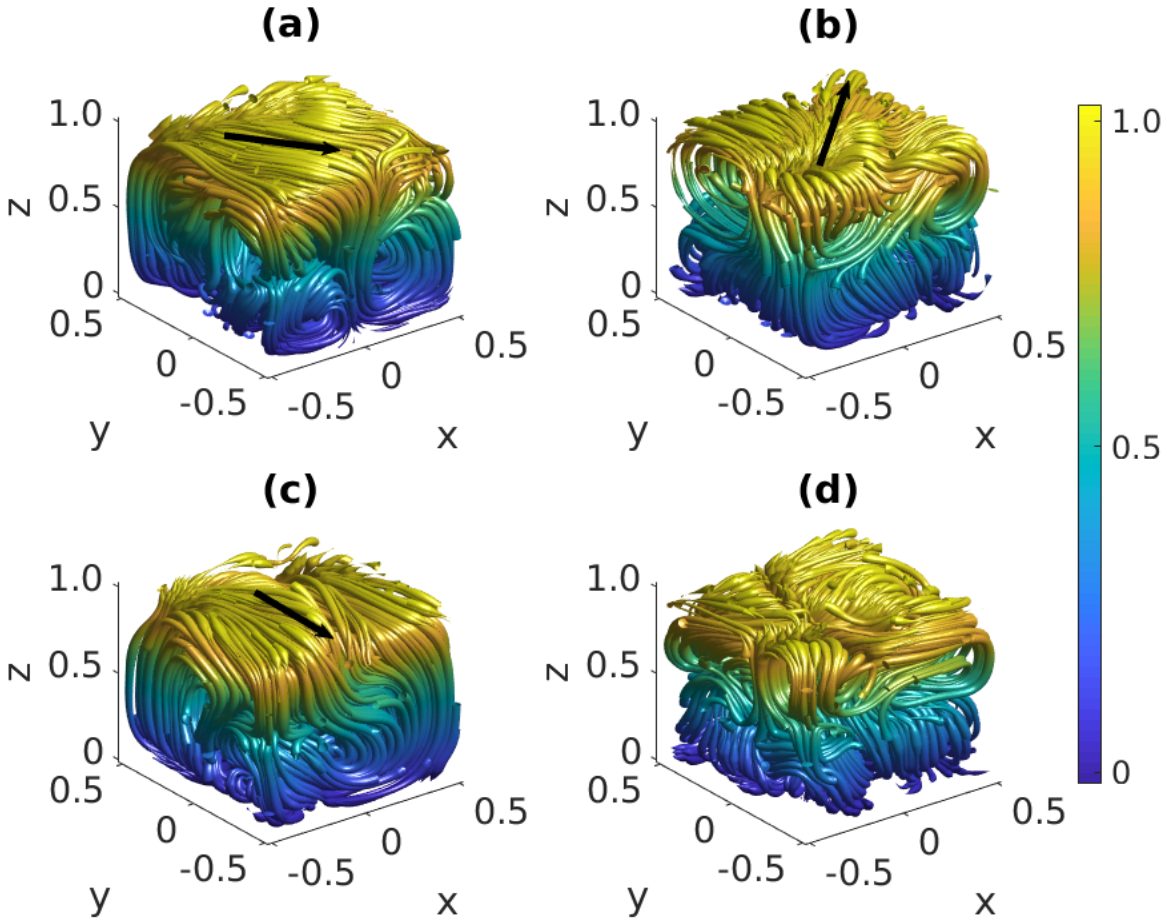

FIG. 1: Instantaneous velocity streamtubes representing different configurations of the large-scale circulation. (a) Stable large-scale circulation state S_{135} aligned along a diagonal. (b) Stable state S_{45} along the other diagonal. (c) Unstable large-scale circulation state $S_{n\pi/2}$ which is aligned along one pair of opposite side faces. (d) Decoherent or null state S_0 without any well-defined large scale circulation. All data are for a fluid with $Pr = 0.7$. The color coding is with respect to the z-plane as given in the legend.

Run	Ra	Pr	Nu	Re	$u_{\rm rms}$	$t_{\rm total}/t_f$
1	10^6	0.1	6.08	1550.5	0.49	10^5
2	10^6	0.7	6.73	437.5	0.37	10^5
3	10^6	10.0	6.92	26.3	0.08	5×10^4

TABLE I: Parameters of the simulations. These are the Rayleigh and Prandtl numbers, followed by the Nusselt number Nu, the Reynolds number Re, the root mean square velocity $u_{\rm rms}$, and the total integration time in units of the free-fall time t_f.

II. NUMERICAL MODEL

In the Rayleigh–Bénard model, the dimensionless equations of motion of an incompressible fluid undergoing thermal convection are given by

$$\partial_t \mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u} = -\nabla p + \sqrt{\frac{Pr}{Ra}}\nabla^2\mathbf{u} + T\hat{z}\,, \tag{3}$$

$$\partial_t T + (\mathbf{u}\cdot\nabla)\mathbf{T} = \frac{1}{\sqrt{RaPr}}\nabla^2 T\,. \tag{4}$$

$$\nabla\cdot\mathbf{u} = 0\,. \tag{5}$$

where $\mathbf{u}(x,y,z,t) \equiv (u_x(x,y,z,t), u_y(x,y,z,t), u_z(x,y,z,t))$ is the velocity vector field of the fluid, $T = T(x,y,z,t)$ denotes the scalar temperature field, and $p \equiv p(x,y,z,t)$ is the scalar pressure field. We employ the Boussinesq approximation, whereby the density of the fluid ρ_0 is assumed to be constant except in the buoyancy term that leads to the last term on the right hand side in (3). The equations are made dimensionless by scaling the length scales by the thickness of the fluid d, velocity by the free-fall velocity $U_f = \sqrt{\alpha g \delta T d}$, and time by the free-fall time scale $t_f = \sqrt{d/(\alpha g \delta T)}$. The Rayleigh number Ra and the Prandtl number Pr are two control parameters of the system. The aspect ratio of the cube is one. In table I, we summarize some important statistical quantities of the three simulation runs.

We performed direct numerical simulations (DNS) of the eqns. (3)–(5) for three different fluids corresponding to Prandtl numbers of $Pr = 0.1, 0.7$, and 10 and sustain a Rayleigh number to be $Ra = 10^6$ in each case. Consequently, the effective Reynolds number and the Kolmogorov scales change in each case, allowing us to study the effect of Reynolds number on re-orientations between the large scale circulations. The Reynolds number measures the turbulent momentum transfer in RBC and is given by

$$Re = \sqrt{\frac{Ra}{Pr}}\,u_{\rm rms} \quad \text{with} \quad u_{\rm rms} = \sqrt{\langle u_x^2 + u_y^2 + u_z^2\rangle_{V,t}}\,, \tag{6}$$

where $\langle\cdot\rangle_{V,t}$ denotes a combined volume and time average. The variation of Pr alters also the magnitude of the Nusselt number Nu, the dimensionless measure of the turbulent heat transfer which is given by

$$Nu = 1 + \sqrt{RaPr}\langle u_z T\rangle_{V,t}\,, \tag{7}$$

see again Table I.

The simulations were performed using an open source code nek5000 based on a spectral element method [36, 37]. Our system consists of a cubic container of side unity in each direction. The system is uniformly heated from below at $z = 0$ and cooled from at $z = 1$, i.e., Dirichlet conditions for the temperature field apply. We assume no-slip boundary conditions, which translates to $u_x = u_y = u_z = 0$ at all boundaries. We also assumed insulating boundaries for temperature field at the four side faces, i.e., $\mathbf{n}\cdot\nabla T = 0$. We took 16 elements along each direction for simulations and the order of the Lagrangian interpolation polynomials along each space direction and on each spectral element is 5. The vertical profiles of the mean kinetic energy dissipation rate were analysed to verify that this spectral resolution is sufficient. For the LSC flow analysis, we interpolate the vector and scalar fields spectrally onto a uniform mesh.

In agreement with previous studies [22, 23], we identify four stable LSC states along the diagonals. There are denoted as S_{45}, S_{135}, S_{225}, and S_{315}. Once the flow gets into one of these stable states, it stays here for a considerable amount of time before re-orienting into another stable configuration. The re-orientations between the stable large scale circulations transition via the unstable large scale structures aligned along the edges of the cube. These four individual states are summarized to a fifth LSC configuration which is denoted as $S_{n\pi/2}$ with $n = 0, 1, 2, 3$. In addition to the stable and unstable LSC states, we identified a sixth state which does not belong to any of the previous states.

LSC state	Orientation angle θ	Stability
S_{45}	45	stable
S_{135}	135	stable
S_{225}	225	stable
S_{315}	315	stable
$S_{n\pi/2}$	0, 90, 180, 270	unstable
S_0	–	unstable

TABLE II: Six large-scale flow states in the cubical convection cell. The orientation angle θ is given in figure 2. Stability is meant rather in terms of a mean living time of the state than a linear stability property of the flow configuration.

In this state, termed as the *decoherent state* or *Null state* S_0, the turbulent system do not have any distinct large scale circulation. Figure 1 provides a visualization of typical stable and unstable large scale circulations for a Prandtl number of 0.7. Table II also summarizes the 6 LSC states. The identification of the specific states is described in the subsequent section.

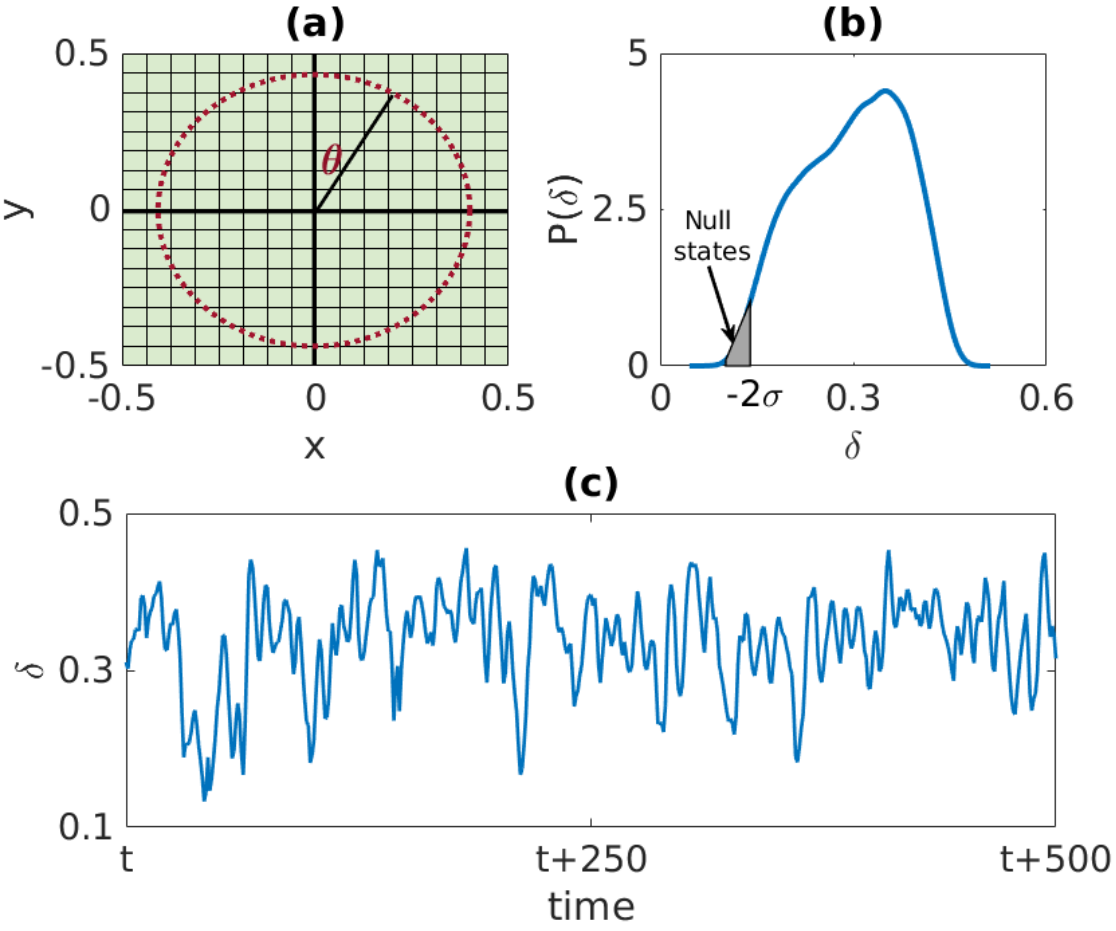

FIG. 2: (a) Schematic representation of the interpolated grid on the circle from data of the uniform rectangular grid at the mid plane $z = 0.5$. The rectangular and circular grids are superimposed on each other for clarity in comprehension. The angle of the LSC (θ) is measured with respect to the origin in clockwise direction. (b) Probability distribution function (PDF) of the fraction of energy δ (see eq. (8)) carried by the largest Fourier mode. The shaded portion represents the region in which the events were considered to be the Null or decoherent states. (c) Shows the temporal variation of δ over a short window.

III. LARGE-SCALE CIRCULATION STATES ALONG SINGLE LONG-TERM TRAJECTORY

We began our simulations from a random initial condition and waited for the flow to become fully turbulent. Once the system reached a steady state, we performed the simulations for another 10^5 free fall times and output the data at each free fall time (in cases $Pr = 0.1, 0.7$). For the purpose of measuring the orientation angles of the LSC, we isolated the mid-plane $z = 0.5$ and computed the vertical velocities at all grid points in this plane, as shown in figure 2(a). Thereafter, we interpolated the vertical velocities to a circular grid with fixed radius of $r = 0.45$ and an angle θ (measured with respect to the y-axis in a clockwise manner) varying by 5° for each subsequent grid point.

Thereafter, we obtained the discrete Fourier transform of the vertical velocities interpolated on the circular grid in the mid-plane. In presence of a distinct LSC structure in the cubic cell, the largest Fourier mode will possess considerable amount of energy and the corresponding phase will give a measure of the angle of orientation (θ) of the LSC. Figure 2(c) shows the ratio δ of the energy carried by the largest Fourier mode to that of the total energy of the system as a function of time which is given by

$$\delta(t) = \frac{\max_{k_\theta} |\hat{u}_z(k_\theta, t)|^2}{\sum_{k_\theta} |\hat{u}_z(k_\theta, t)|^2} \,. \tag{8}$$

The majority of the time, the system will be either in one of the four stable states or in the unstable LSC states $S_{n\pi/2}$. When there is no distinct LSC (Null state), the ratio δ will be minuscule. This low value of δ in the Null state S_0 is obviously expected as in absence of any LSC structure, the total energy is almost equally distributed among all the Fourier modes. These Null states are extremely rare and hence to identify them, we first calculate the probability distribution function (PDF) of the ratio δ which is shown in figure 2(b). We then calculated the standard deviation σ of the PDF of δ. All the events which fall below -2σ and thus have extremely low values of δ are identified as the Null states. It is noted that all events larger than $+2\sigma$ still fall into the category of distinct LSC structures. A visualization of the flow structure in a typical Null state is shown in figure 1(d).

As mentioned in the previous sections, we performed the single trajectory analysis in case of fluids with $Pr = 0.1, 0.7$ for 10^5 free fall time units. We assigned an angle of orientation (θ) of the LSC structure with each time frame obtained at every free fall time. The classification is conditioned to both, the values of θ and δ. Figures 3(a) and (b) show the temporal evolution of the LSC orientation angle (θ) for the cases of $Pr = 0.1$ and $Pr = 0.7$, respectively. In both cases, we classified the flow trajectory into 6 distinct "macrostates" or basic states. The plots underline that the runtime is long enough to capture all states equally. The first four macrostates correspond to stable LSC orientations parallel to the diagonals, with an angle of orientation of $\theta = 315^\circ, 225^\circ, 135^\circ$, and 45° and value of $\delta > 2\sigma$. All the LSC orientations which have θ values $\pm 20^\circ$ to the diagonals are also clubbed in the same macrostate. Therefore, a particular macrostate can have many distinct microstates in the trajectory. The fifth macrostate is the unstable LSC state parallel to the edges having $\theta = n \times 90^\circ \pm 20^\circ$ for $n = 0, 1, 2, 3$ and $\delta > 2\sigma$. We do not distinguish between different n and summarize all LSC states that are aligned with the side faces into $S_{n\pi/2}$. The sixth and the final macrostate S_0 is indicated by the black dots in both panels. Its occurrence is rare.

For both Prandtl numbers in figure 3, the stable LSC states appear to occur approximately equal number of times. The overall pattern differs slightly. The case at $Pr = 0.7$ can be characterized by less frequent switches between the stable states. For the lowest Prandtl number in our series at $Pr = 0.1$, the large-scale flow indicates a stronger decoherence for the present Rayleigh number. One possible reason could be the difference of the mean viscous and thermal boundary layer thicknesses which drive the thermal plume formation and thus LSC jointly. The derivation of clear trends in the Prandtl number dependence of the LSC in this specific geometry would require a generation of additional long-term trajectories at further higher Rayleigh numbers.

The unstable LSC states parallel to the edges, $S_{n\pi/2}$ are seen to fill up the trajectory in both cases as the transition to any stable state occurs via these unstable states. It is also be noted that data in figure 3 are averaged over 20 time frames to obtain a lucid representation. The Null state appears when there is a transition from one state to another and the flow re-orients itself. In the case of $Pr = 10$ once the flow reaches a stable state, it remains locked there until and unless subjected to strong perturbation. We followed this state for 5×10^4 free-fall times without a switch. Therefore, this trajectory at the highest of the three Prandtl numbers in our DNS series is not shown here.

In the following, we will intend to focus on an alternative analysis approach to the large-scale flow, namely by the application of Markov State Models. In such an approach, the main objective is to calculate the macrostate transition probability matrix. Using this information, we can predict how the large-scale statistics of the system evolves in the future, *without* requiring to observe a trajectory that reaches statistical equilibrium.

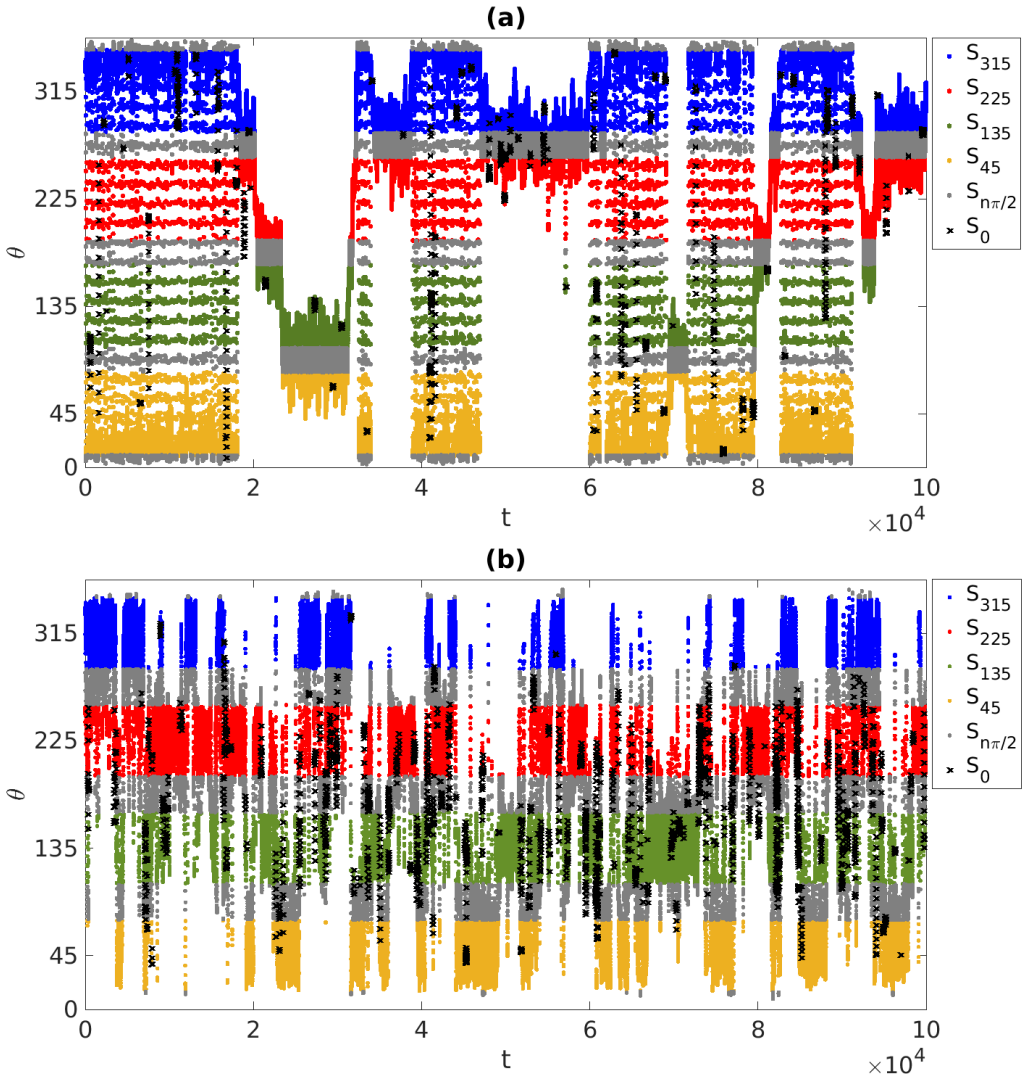

FIG. 3: Trajectories of the orientation angle θ of the large-scale circulation for $Pr = 0.1$ (top panel) and $Pr = 0.7$ (bottom panel). Time t is measured in units of the free-fall time t_f. States $S_{n\pi/2}$ are given in gray, Null states S_0 in black. In both cases, an sliding time-windowed average with a width of 20 t_f is applied.

IV. ENSEMBLE ANALYSIS OF TRANSITION PROBABILITIES

We proceed with the aim of calculating the probability of transition from one LSC state to another. This analysis is conducted for the case of $Pr = 0.7$ only. Needless to say that we shifted from the single trajectory approach to a more comprehensive statistical ensemble approach. The ensemble averaging approach is more prudent than the single trajectory approach, as it cancels out any inaccuracy that might arise due to observation of rare events for a single trajectory run. For direct numerical simulations at higher Rayleigh numbers than used here, it might be the only way to sample the phase space sufficiently well, namely to advance with a coarse grid run for a long time interval and to start short-term highly-resolved simulations from the coarse grid outputs. In the present case, the long-term single trajectory run was already sufficiently well resolved such that an upscaling is not necessary.

As mentioned in preceding section, we classified the system in six basic states – four stable LSC configurations (S_{45}, S_{135}, S_{225}, and S_{315}), the unstable LSC configuration ($S_{n\pi/2}$), and the "Null" or undefined configuration sans any LSC (S_0). Consequently, the transition probability matrix will be a matrix of dimension 6×6, with the matrix elements A_{ij} representing the probability of transition from state i to state j. The accuracy of the computed matrix depends upon two pivotal parameters: (i) The number of short term simulations for ensemble averaging, (ii) the duration of a single ensemble run for determining the transition probability. Also for this method, we have to stick to a finite and attainable amount of runs for ensemble averaging.

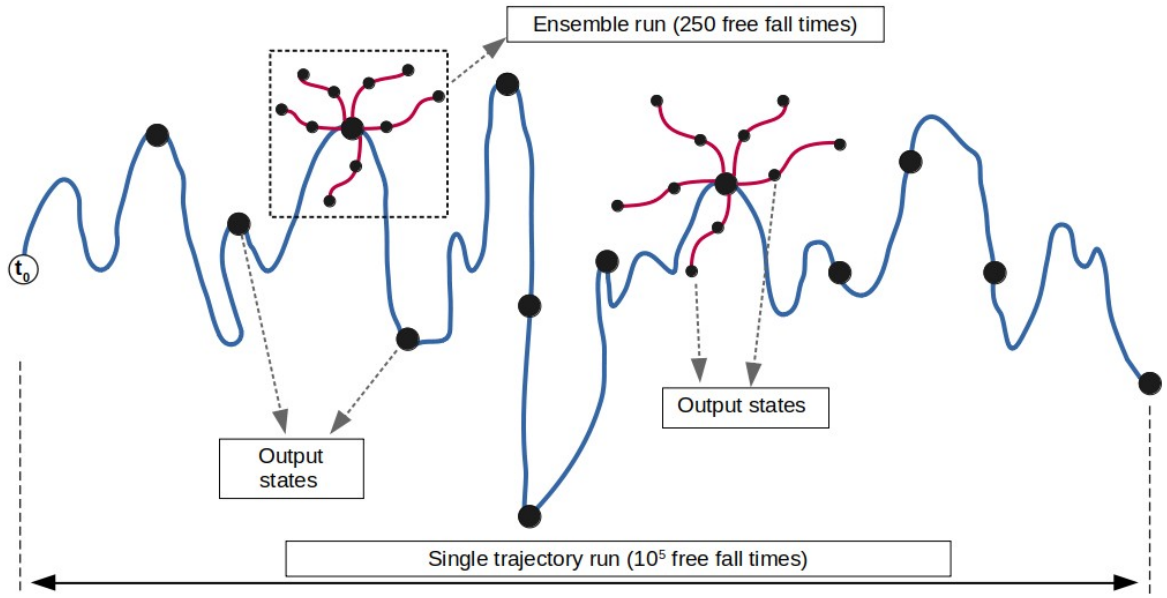

FIG. 4: Schematic representation of the ensemble averaging process to obtain the transition probability matrix. The blue line denotes high resolution long single trajectory run (such as the ones shown in fig. 3 (a) and (b)), where the output states (black '•') are printed at every free fall time. The curves in the dashed box represents a typical shorter ensemble run, where starting from a randomly chosen initial condition, the system is perturbed and allowed to evolve in time.

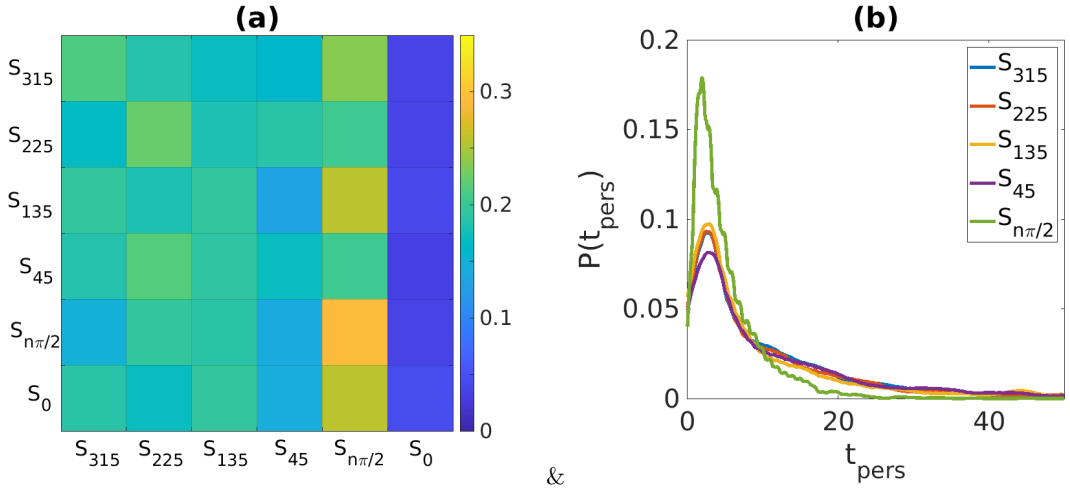

FIG. 5: Ensemble run analysis to test the Markov state model. (a) Color coded transition probability matrix A_{ij} for a long runtime of 2500 free-fall times. (b) Probability distribution function (PDF) of time t_{pers} for which a system persists in a specific large-scale flow configuration (macrostate). 1800 individual trajectories were used to determine the statistics in panel (a).

The initial conditions for the short term simulations were chosen randomly from the output of the single long-term trajectory (as depicted in fig. 4). We chose sixty turbulent flow configurations each belonging to a specific LSC configuration (macrostate). This results in 360 initial configurations that were picked along the single long-term trajectory in total. Each of the 360 initial conditions is then perturbed five times distinctively. Therefore, in effect, we have 300 of such short-term simulations, starting from each LSC configuration or macrostate. The ensemble analysis is consequently based on a total of 1800 short-term trajectories. Unlike the case of single trajectory approach, the

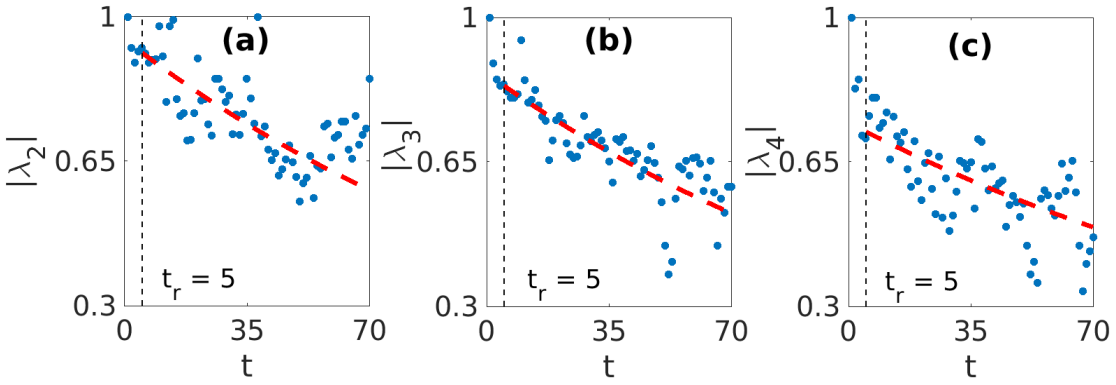

FIG. 6: Plot of absolute values of second, third, and fourth-largest eigenvalues of the transition probability matrix as a function of time (blue dots). The eigenvalues are computed from the matrices $\mathbf{A}_t$ for each time frame and the red dashed curve represents the fitted exponential curve. The data for the fit are the eigenvalues for $t > t_r = 5$.

output is now coarser with respect to time. The turbulence fields are written out after a finite lag time and not at every free-fall time in order to keep the amount of data limited.

The appropriate choice of the output time along the trajectories has to be determined first. If the lagtime is too small then one would not observe transitions to another LSC state. For long lagtimes, one would be missing information about intermediate transitions. Also, the longer the lagtime, the larger the ensemble needs to be to resolve the transition statistics, as these spread over more macrostates. Figure 5(a) represents the color-coded transition probability matrix for a lagtime along the short-term trajectories of 2500 free-fall times t_f to give a first example. Each box in the figure signifies the value of an element A_{ij} of the matrix. The average is taken over all 1800 trajectories. As it can be inferred from the figure, all states except the Null state S_0 get populated more or less equally, irrespective of the initial condition. The S_0 state, obviously, has the lowest probability of occurrence. This, however, might not portray the correct picture because in the course of its dynamical evolution the system might have undergone transitions and the ensemble trajectories might have diverged strongly in phase space of each other. We conclude that this chosen lagtime is by far too long.

To evaluate the lagtime, we evaluate the time for which the system remains in a state before transitioning to a different state and label this time as the persistence time $t_{\rm pers}$. The persistence times $t_{\rm pers}$ for all possible stable and unstable states ($S_{45}, S_{135}, S_{225}, S_{315}$, and $S_{n\pi/2}$) were computed separately, using the data from the original single trajectory run, except the rarely occurring S_0 state. We, then, plotted the probability distribution functions (PDFs) of these persistence times as shown in figure 5(b). The PDFs for all macrostates peak around 5 t_f, indicating that for most initial conditions a first transition can occur at a lagtime as short as 5 free fall times t_f. These PDFs are obtained using the raw data without any averaging and hence this short persistence time is connected to the turbulent fluctuations of the convection flow. Note also that all PDFs have fat tails up to 100 free fall times. The latter time would correspond to the persistence time of a macrostate, if we average out small scale fluctuations, as observed from figure 3. We proceed by taking a lagtime of 10 t_f which is twice the persistence time $t_{\rm pers}$. Thus more than 50% of the cases would be transition to the subsequent macrostate. Thereafter, we used again the method described in section III to identify the state of the system at the corresponding time and to determine A_{ij} (which are always averaged over the ensemble of 1800 trajectories).

V. PROBING THE MARKOV STATE PROPERTY

Based on the ensemble run data, we now proceed to test the Markovianity of the estimated model. This is done by means of the eigenvalue spectrum of the transition probability matrices A_{ij} at different output times which are multiples of 10 t_f as discussed in the last section. To begin with, we calculated the ensemble averaged transition probability matrix $\mathbf{A}_t$ at different lagtimes $t \geq t_r$ where $t_r = 5t_f$ is a reference time for Markov analysis. The eigenvalues $\lambda_i(t)$, $i = 1, \ldots, 6$ are always ordered according to descending magnitude. The largest eigenvalue λ_1 will remain unity at all times, since the matrices are all stochastic. The subsequent eigenvalues may remain unity for times $t < t_r$ until they start to leave the initial macrostates and decrease thereafter.

If $\mathbf{A}_t$ defines the transition matrix up to lagtime t then for a truly Markovian process the so-called *Chapman–*

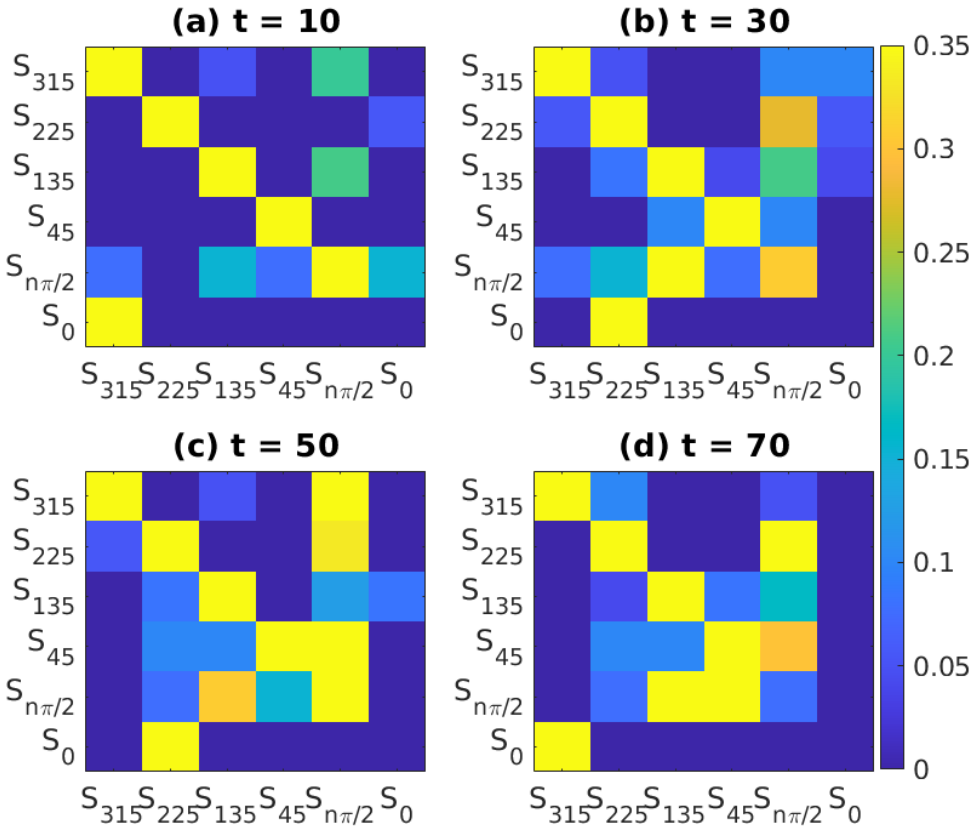

FIG. 7: Time evolution of the transition probability matrix. The color coded transition probability matrix are plotted for four different values of time instants (a) t = 10, (b) t = 30, (c) t = 50, and (d) t = 70 free fall time units.

Kolmogorov identity should be satisfied. It is given by

$$\mathbf{A}_{t+s} = \mathbf{A}_t \mathbf{A}_s \quad \text{for} \quad s, t > 0\,. \tag{9}$$

This means for the eigenvalues, that they must satisfy $\lambda_i(t) = \exp(-\omega_i t)$ for some *rate* $\omega_i \in \mathbb{C}$. Clearly, due to the coarse graining, as we discussed above, we can expect the Chapman–Kolmogorov identity to hold best for lagtimes $s, t \geq t_r$. Less dominant, i.e., faster decaying eigenvalues correspond to faster dynamical processes. Hence they are harder to estimate by the sampling procedure. We thus discard the two smallest eigenvalues λ_5 and λ_6 and test $|\lambda_i(t)| \approx \exp(-\alpha_i t)$ for $t > t_r$ for $i = 2, 3, 4$, and appropriate $\alpha_i > 0$.

The ensemble analysis is now conducted for the 1800 trajectories, each advanced in time for 250 free-fall times. For $t > t_r = 5$ we output each 10 free-fall times. These data are used to test Markovianity. The trend of variation of eigenvalues $\lambda_i(t)$ of $\mathbf{A}_t$ as a function of time yielded strong fluctuations, most likely due to the insufficient sample size. Hence, to obtain superior statistical averaging, we further divided these short term simulations into three different sections of a temporal length of 70 free-fall time each, separated by a window of 10 free-fall times. The window of 10 free fall time steps is chosen between the sections to decorrelate the time intervals. The three chunks of data are treated as independent trajectories, and the whole process of computing the matrices $\mathbf{A}_t$ and their eigenvalues $\lambda_i(t)$ was repeated. Figure 6 shows the magnitude of the second, third, and fourth-largest eigenvalues of $\mathbf{A}_t$ as a function of time. The figures indicate a better representation of an exponential decay and inferring to the fact that improving the statistics further with lead to an even better fit. Note that for the fits, we used the eigenvalues computed at every free-fall time.

We fitted an exponential function to the absolute values of the eigenvalues for $t > t_r$ and obtained decay coefficients ω_i from the fits. The reciprocal of ω_i provides an estimate for a characteristic time. The fits indicate that such a decay time is of the order of 150 free-fall time units. Computing the time for which $\lambda_2(t) < 0.01$ gives an idea of the time of convergence to within 1% of statistical equilibrium. Figure 7 shows the color-coded transition probability matrices $\mathbf{S}_t$ at different multiples of the lagtimes $t = 10t_f$. For the diagonal flow configurations, self-transitions appear

to be the most likely transitions. The transition probability is lowest to the S_0 state. As can be seen, the data is still accompanied by considerable noise, not allowing to come to any solid conclusion about the Markovianity of these transition matrices.

VI. SUMMARY AND CONCLUSION

The present numerical study of turbulent Rayleigh–Bénard convection was focused on the large-scale flow behavior in a closed cubic cell, a possible configuration of laboratory measurements. Contrary to the more frequently applied cylinder configuration the present geometry lacks an azimuthal symmetry, which generally causes a slow drift of the large-scale circulation configuration as a whole, see e.g. ref. [38]. The constrained flow in the cube causes 6 *discrete* macrostates of the convection flow and generates an approximate hopping dynamics between these different states in the high-dimensional phase space the Prandtl number is sufficiently low at a given Rayleigh number (which was here $Ra = 10^6$). In this respect, the RBC flow behaves similarly to molecular dynamics systems, such as proteins, that undergo rapidly conformational changes. This motivated the application of a Markov state model to the RBC system in this work.

First, we studied the dynamics along a single long-term trajectory that extends over 10^5 free-fall time units t_f. Along this trajectory each of the macrostates appeared sufficiently often which allowed us to determine their characteristic lifetime. It is found that this time depends strongly on the Prandtl number. For $Pr \lesssim 1$ our system switches rapidly between different macrostates. For $Pr \gg 1$, the LSC remains locked in one particular configuration for $5 \times 10^4 t_f$. This behavior will depend on the Rayleigh number, which was fixed to one value only here. As can be seen in Table I, the resulting momentum transfer at $Ra = 10^6$ and $Pr = 10$, which is quantified by the Reynolds number Re remains too small. The flow thus lacks sufficient fluctuations of the velocity field that are able to transfer the flow into a new configuration. For the case of $Pr = 0.7$, we analysed the transitions between the different macrostates in the form of an ensemble analysis and determined the corresponding transition matrix. It was shown that the elements of this matrix depend strongly on the lagtime that is used to probe a transition from a macrostate i to a macrostate j.

A subsequent analysis of the Markov State Model revealed that the ensemble size used to assemble the model might be still too small, and statistical errors overshadow dynamical properties. Further sampling beyond the one along the 1800 trajectories might be necessary to assess how a faithful Markovian model could be constructed. Another important factor that governs the success of a MSM is the Markovianity assumption, i.e., the independence of the distribution of the macrostate at time $t + t_r$ on the macrostate at time $t - t_r$, if the macrostate at time t is given ("memoryless property"). If this condition is not met, either the minimal lagtime t_r needs to be raised or the explicit incorporation of memory into the model is required. Alternatively, one could question the choice of macrostates, and methods could be considered that estimate surrogate macrostate variables directly from data. These are either tuned to give optimal representation of the state-space geometry [23, 35, 39], or to reproduce optimally certain statistical properties of the coarse-grained dynamics [40, 41].

The present work is thought as a first proof of concept. Several routes are possible to extend our present research. It would be desirable to increase the Rayleigh number to values of 10^7 or 10^8 which would be line with a modified hopping dynamics. We expect less coherent large-scale flows due to the enhanced level of velocity fluctuations and thus shorter mean lifetimes for the macrostates and for all Prandtl numbers. A second route would follow an analysis in larger-aspect-ratio configurations, i.e., for convection in flat domains. The switching between different large-scale patterns that might contribute differently to the heat transfer would be an interesting task for future work.

AUTHORS' CONTRIBUTIONS

P.K. and J.S. designed the research. P.M. conducted the simulations and analyzed the data. All three authors discussed the results and drafted the manuscript. All three authors gave final approval for publication.

COMPETING INTERESTS

We declare we have no competing interests.

FUNDING

Priyanka Maity is supported by the grants SCHU 1410/29-1 and SCHU 1410/30-1 of the Deutsche Forschungsgemeinschaft (DFG). Péter Koltai has been partially supported by Deutsche Forschungsgemeinschaft (DFG) through grant CRC 1114 "Scaling Cascades in Complex Systems", Project Number 235221301, Project A01 "Coupling a multiscale stochastic precipitation model to large scale atmospheric flow dynamics".

ACKNOWLEDGEMENTS

The comprehensive ensemble simulations were conducted at the MaPaCC4 compute cluster of the University Computing Centre at Technische Universität Ilmenau (Germany).

DISCLAIMER

Not applicable.

[1] Spiegel EA. 1971 Convection in stars i. Basic Boussinesq convection. *Annu. Rev. Astron. Astrophys.* **9**, 323–352.
[2] Marshall J, Schott F. 1999 Open-ocean convection: Observations, theory, and models. *Rev. Geophys.* **37**, 1–64.
[3] Markson R. 1975 Atmospheric electrical detection of organized convection. *Science* **188**, 1171–1177.
[4] Hathaway DH. 2012 Supergranules as probes of solar convection zone dynamics. *Astrophys. J. Lett.* **749** L13.
[5] Schumacher J, Sreenivasan KR. 2020 Colloquium: Unusual dynamics of convection in the Sun. *Rev. Mod. Phys.* **92** 041001.
[6] Chandrashekhar S. 1961 *Hydrodynamic and Hydromagnetic instability.* Dover Publication, UK.
[7] Ahlers G, Grossmann S, Lohse D. 2009 Heat transfer and large scale dynamics in turbulent Rayleigh-Bénard convection. *Rev. Mod. Phys.* **81**, 503–537.
[8] Chillà F, Schumacher J. 2012 New perspectives in turbulent Rayleigh-Bénard convection. Eur. Phys. J. E **35**, 58.
[9] Pandey A, Scheel JD, Schumacher J. 2018 Turbulent superstructures in Rayleigh-Bénard convection. Nat. Commun. **9**, 2118.
[10] Niemela JJ, Skrbek L, Sreenivasan KR, Donelly RJ. 2001 The wind in confined thermal convection. *J. Fluid Mech.* **449**, 169–178.
[11] Sreenivasan KR, Bershadskii A, Niemela JJ. 2002 Mean wind and its reversal in thermal convection. *Phys. Rev. E* **65**, 056306.
[12] Parodi A, von Hardenberg J, Passoni G, Provenzale A, Spiegel EA. 2004 Clustering of plumes in turbulent convection. *Phys. Rev. Lett.* **92**, 194503.
[13] Puthenveettil BA, Arakeri JH. 2005 Plume structure in high-Rayleigh-number convection. *J. Fluid Mech.* **542** 217–249.
[14] Brown E, Ahlers G. 2006 Rotations and cessations of the large-scale circulations in turbulent Rayleigh-Bénard convection. *J. Fluid Mech.* **568**, 351–386.
[15] Xi H-D, Xia K-Q. 2007 Cessations and reversals of the large-scale circulations in turbulent thermal convection. *Phys. Rev. E* **75** 066307.
[16] Bao Y, Chen J, Liu B-F, She Z-S, Zhang J, Zhou Q 2015 Enhanced heat transport in partitioned thermal convection. *J. Fluid Mech. 784* R5.
[17] Wagner S, Shishkina O. 2013 Aspect-ratio dependency of Rayleigh-Bénard convection in a box shaped containers. *Phys. Fluids* **25** 085110.
[18] Chong KL, Huang S-D, Kaczorowski M, Ni R, Xia K-Q. 2015 Condensation of coherent structures in turbulent flows. *Phys. Rev. Lett.* **115** 264503.
[19] Daya ZA, Ecke RE. 2001 Does turbulent convection feel the shape of the container? *Phys. Rev. Lett.* **87**, 184501.
[20] Foroozani N, Niemela JJ, Armenio V, Sreenivasan KR. 2014 Influence of container shape on scaling of turbulent fluctuations in convection. *Phys. Rev. E* **90**, 063003.
[21] Bai K, Ji D, Brown E. 2016 Ability of a low-dimensional model to predict geometry-dependent dynamics of large-scale coherent structures in turbulence. *Phys. Rev. E* **93**, 023117.
[22] Foroozani N, Niemela JJ, Armenio V, Sreenivasan KR. 2017 Reorientations of the large-scale flow in turbulent convection in a cube. *Phys. Rev. E* **95**, 033107.
[23] Giannakis D, Kolchinskaya A, Krasnov D, Schumacher J. 2018 Koopman analysis of the long-term evolution in a turbulent convection cell. *J. Fluid Mech.* **847**, 735–767.
[24] Vasilev A, Frick P, Kumar A, Stepanov R, Sukhanovskii, Verma MK. 2019 Transient flows and reorientations of large-scale convection in a cubic cell. *Int. Commun. Heat Mass Transfer* **108**, 104319.
[25] Mezic I. 2013 Analysis of fluid flows via spectral properties of the Koopman operator. *Annu. Rev. Fluid Mech.* **45**, 357–378.

[26] Williams MO, Kevrekidis IG, Rowley CW. 2015 A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition. *J. Nonlinear Sci.* **25**, 1307–1346.

[27] Sauer T, Yorke JA. 1991 Shadowing trajectories of dynamical systems, In: Meyer K.R., Schmidt D.S. (eds) *Computer Aided Proofs in Analysis*. The IMA Volumes in Mathematics and Its Applications, vol 28. Springer, New York, NY, 229-234.

[28] Pande VS, Beauchamp K, Bowman GR. 2010 Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **52**, 99–105.

[29] Husic BE, Pande VS. 2018 Markov State Models: From an art to a science. *J. Am. Chem. Soc.* **140**, 2386–2396.

[30] Prinz JH, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera JD, Schütte C, Noé F. 2011 Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **134**, 174105.

[31] Schütte C, Sarich M. 2013 *Metastability and Markov State Models in Molecular Dynamics.* Courant Lecture Notes in Mathematics.

[32] Bowman GR, Pande VS, Noé F. 2014 *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation* Vol. 797, *Advances in Experimental Medicine and Biology*, Springer, Berlin.

[33] Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M, Plattner N, Wehmeyer C, Prinz JH, Noé F. 2015 PyEMMA 2: A software package for estimation, validation, and analysis of Markov models. *J. Chem. Theory Comput.* **11**, 5525–5542.

[34] Harrigan MP, Sultan MM, Hernández CX, Husic BE, Eastman P, Schwantes CR, Beauchamp KA, McGibbon RT, Pande VS. 2017 MSMBuilder: Statistical Models for Biomolecular Dynamics. *Biophys. J.* **112**, 10–15.

[35] Koltai P, Weiss S. 2020 Diffusion maps embedding and transition matrix analysis of the large-scale flow structure in turbulent Rayleigh–Bénard convection. *Nonlinearity* **33** 1723–1756.

[36] Fischer PF. 1997 An overlapping Schwarz method for spectral element solution of the incompressible Navier–Stokes equations. *J. Comp. Phys.* **133**, 84–101.

[37] Scheel JD, Emran MS, Schumacher J. 2013 Resolving the fine-scale structure in turbulent Rayleigh-Bénard convection. *New J. Phys.* **15**, 113063.

[38] Zürner T, Schindler F, Vogt T, Eckert S, Schumacher J. 2019 Combined measurement of velocity and temperature in liquid metal convection. *J. Fluid Mech.* **876**, 1108–1128.

[39] Berry, T, Giannakis, D, Harlim, J. 2015 Nonparametric forecasting of low-dimensional dynamical systems. *Physical Review E.* **91(3)**, 032915.

[40] Bittracher, A, Koltai, P, Klus, S, Banisch, R, Dellnitz, M, Schütte, C. 2018 Transition manifolds of complex metastable systems: Theory and data-driven computation of effective dynamics. *Journal of Nonlinear Science*, **28(2)**, 471–512.

[41] Pillaud-Vivien, L, Bach, F, Lelièvre, T, Rudi, A, Stoltz, G. 2020 Statistical estimation of the Poincaré constant and application to sampling multimodal distributions. In *International Conference on Artificial Intelligence and Statistics*, PMLR, 2753-2763.