GENUS TOPOLOGY OF STRUCTURE IN THE SLOAN DIGITAL SKY SURVEY: MODEL TESTING

J. Richard Gott III,¹ D. Clay Hambrick,¹ Michael S. Vogely,² Juhan Kim,³ Changbom Park,³ Yun-Young Choi,³ Renyue Cen,³ Jeremiah P. Ostriker,¹ and Kentaro Nagamine³

Received 2006 October 25; accepted 2007 October 5

ABSTRACT

We measure the three-dimensional topology of large-scale structure in the Sloan Digital Sky Survey (SDSS). This allows the genus statistic to be measured with unprecedented statistical accuracy. The sample size is now sufficiently large to allow the topology to be an important tool for testing galaxy formation models. For comparison, we make mock SDSS samples using several state-of-the-art N-body simulations: the Millennium run of Springel et al. (10 billion particles), the Kim & Park CDM models (1.1 billion particles), and the Cen & Ostriker hydrodynamic code models (8.6 billion cell hydro mesh). Each of these simulations uses a different method for modeling galaxy formation. The SDSS data show a genus curve that is broadly characteristic of that produced by Gaussian random-phase initial conditions. Thus, the data strongly support the standard model of inflation where Gaussian random-phase initial conditions are produced by random quantum fluctuations in the early universe. But on top of this general shape there are measurable differences produced by nonlinear gravitational effects and biasing connected with galaxy formation. The N-body simulations have been tuned to reproduce the power spectrum and multiplicity function but not topology, so topology is an acid test for these models. The data show a “meatball” shift (only partly due to the Sloan Great Wall of galaxies) that differs at the 2.5 σ level from the results of the Millenium run and the Kim & Park dark halo models, even including the effects of cosmic variance.

Subject headings: cosmology: observations — large-scale structure of universe — methods: numerical

1. INTRODUCTION

The topology of large-scale structure in the universe is an important physical property of the matter density field that can be compared with the prediction of the simple inflationary models (Guth 1981; Linde 1983) where Gaussian random phase initial conditions are generated from quantum fluctuations in the early universe. The analytic tools used in the present work for quantitatively analyzing the topology of large-scale structure in three dimensions have been developed during the past 20 years (Gott et al. 1986, 1987, 1989; Hamilton et al. 1986; Vogely, et al. 1994; Park et al. 2005a, 2005b). The distribution of galaxies in space is smooth to allow the topology to be an important tool for testing galaxy formation models. For comparison, we make mock SDSS samples using several state-of-the-art N-body simulations: the Millennium run of Springel et al. (10 billion particles), the Kim & Park CDM models (1.1 billion particles), and the Cen & Ostriker hydrodynamic code models (8.6 billion cell hydro mesh). Each of these simulations uses a different method for modeling galaxy formation. The SDSS data show a genus curve that is broadly characteristic of that produced by Gaussian random-phase initial conditions. Thus, the data strongly support the standard model of inflation where Gaussian random-phase initial conditions are produced by random quantum fluctuations in the early universe. But on top of this general shape there are measurable differences produced by nonlinear gravitational effects and biasing connected with galaxy formation. The N-body simulations have been tuned to reproduce the power spectrum and multiplicity function but not topology, so topology is an acid test for these models. The data show a “meatball” shift (only partly due to the Sloan Great Wall of galaxies) that differs at the 2.5 σ level from the results of the Millenium run and the Kim & Park dark halo models, even including the effects of cosmic variance.

Studies of many observational samples have been conducted by our group and others, which have shown in every case a spongelike topology as expected from inflation. (By contrast, as shown by Gott et al. [1987], a Voronoi tessellation where galaxies are located on the honeycomb walls produces a median density contour, which shows isolated voids rather than a spongelike topology.) For notable examples, see Gott et al. (1986, 1989), Moore et al. (1992), Vogely et al. (1994), Canaveses et al. (1998), Hikage et al. (2002, 2003), and Park et al. (2005b). Vogeley et al. found that in 3D the CfA sample was spongelike at the median density contour. This is the sample from which the de Lapparent et al. 2D slice was drawn. Voids were found to be connected by low-density tunnels. In

¹ Princeton University Observatory, Princeton, NJ 08544.
² Department of Physics, Drexel University, Philadelphia, PA 19104.
³ Korea Institute for Advanced Study, CheongNyangNi, Seoul 130-722, Korea.
⁴ Department of Physics, University of Nevada, Las Vegas, 89154.
addition, in all cases the observed genus curve was reasonably well fit by the Gaussian random-phase theoretical curve in equation (4). Perhaps the most spectacular such agreement was seen in the Canavese et al. (1998) analysis of the 15,000 galaxy Point Source Catalogue (PSCz) redshift survey. This study showed quite a good fit (within the noise) to the random-phase curve at a variety of smoothing lengths. The IRAS galaxies in this sample are primarily low-mass spiral and irregular galaxies and so may suffer less biasing effects than galaxies from an optically selected sample. See Park et al. (2007) for a discussion of the strong dependence of morphological fraction on density in the SDSS.

Interestingly, in hindsight, one can note that a number of surveys investigating the “cell” structure model in fact gave topological descriptions that emphasized finding “strings” of galaxies connecting clusters (Einasto et al. 1984) and noting its “filamentary” nature. Of course, clusters connected by a network of “strings” of galaxies is in fact a sponglike topology. Numerical studies confirmed this. For example, Klypin & Shandarin (1983) used \(32^3 = 32,768\) particles in their cosmological simulation and found “chains” of galaxies. Even though they were working under the assumption that “pancakes” formed first followed by the establishment of a network, in their computer simulations they found that “the level surfaces constructed over a large space reveal that there exists a system of chains (or filaments) extending over the whole volume of the model...none of the discussed methods were successful in finding the pancakes.” These findings are consistent with sponglike topology: filaments connecting clusters of galaxies, and low-density tunnels connecting voids.

Although the present work restricts itself to genus topology, we should mention briefly that another method for probing topology and connectedness that is powerful and interesting in its own right is percolation theory (Shandarin 1982). If high-density regions are connected, then they will percolate early—that is, if one draws spheres around individual galaxies and links spheres that overlap, the spheres percolate or make chains covering the entire volume using sphere radii that are smaller than those that would be required to percolate if the distribution of galaxies were Poisson. One can measure the critical radii required for percolation and compare the data with simulations. Such studies showed conclusively that the high-density regions were connected (Shandarin 1983; Einasto et al. 1984; Zel’’ dovich et al. 1982; Yess & Shandarin 1996). However, as the authors note, percolation studies “do not distinguish between a network made of filaments (quasi-one-dimensional objects), or pancakes or sheets (quasi-two-dimensional objects)” (Yess & Shandarin 1996). Thus, they do not distinguish between a Swiss cheese topology and a sponglike topology, both of which have connected high-density regions and both of which percolate early. Indeed, Yess & Shandarin (1996) find in their simulations that at the median density contour the largest high-density region and the largest low-density region occupy almost all of the space so the structures “can be labeled a sponge topology, as predicted by Gott, Melott, & Dickinson 1986.”

A two-dimensional variant of the genus statistic can also be applied to redshift slices, sky maps, and CMB maps (Coles 1988; Melott et al. 1989; Gott et al. 1990; Park et al. 1998). In this case, \(G(\nu) = \) number of hot (or high-density) spots minus number of cold (or low-density) spots, and the Gaussian random-phase hypothesis implies \(g(\nu) \propto \nu \exp(-\nu^2/2)\). All of these studies (redshift slices: Park et al. 1992; Colley 1997; Hoyle et al. 2002a, 2002b; sky maps: Gott et al. 1992; Park et al. 2001; CMB maps: Smoot et al. 1994; Colley et al. 1996; Kogut et al. 1996; Colley & Gott 2003; Park 2004; Spergel et al. 2006; Gott et al. 2007) indicate consistency with the Gaussian random-phase hypothesis. The CMB maps are a particularly powerful test of the Gaussian random-phase hypothesis because the fluctuations are still firmly in the linear regime. The dramatic agreement between the 2D CMB results and the Gaussian random-phase hypothesis strongly supports idea of the standard theory of inflation and that the initial conditions were truly Gaussian random-phase.

The three-dimensional topology data on galaxy clustering are particularly interesting because they allow us not only to confirm their general Gaussian random-phase nature (on large scales), but also to test for nonlinear processes and bias involved with galaxy formation. Matsubara (1994) has discussed how the genus curve may be altered by second-order nonlinear gravitational clustering effects, which can show up at small smoothing lengths. Vogele et al. (1994) explicitly measured the diminution of the genus amplitude caused by these effects. Park et al. (2005a) have studied other important alterations that can occur by nonlinear gravitational evolution, redshift space distortion, and biasing associated with galaxy formation.

In this paper we compute genus curves for volume-limited samples of the largest galaxy redshift survey to date, the Data Release 5 (DR5) of the SDSS, and for mock samples from state-of-the-art N-body simulations. Our goal is to examine whether the models for galaxy formation represented by these simulations are consistent with the observations. This is a potentially powerful test, because the input parameters of the flat \(\Lambda\)CDM model used in these simulations were determined by fitting to a host of other observations—CMB anisotropy data, large-scale power spectrum and correlation function of galaxies, SN Ia luminosity-distance data, cluster abundances and baryon fraction, etc.—but not topology. Thus, this comparison provides an independent test of the model for structure formation.

2. THE GENUS AND RELATED STATISTICS

The genus is a measure of the topology of the large-scale distribution of galaxies. We first smooth the point distribution of galaxy positions (we use only volume-limited samples in the analysis below) with a Gaussian smoothing ball of radius \(\lambda\):

\[
W(r) = \frac{1}{(2\pi)^{3/2}} e^{-r^2/2\lambda^2},
\]

where \(\lambda\) is chosen to be greater than or equal to the correlation length. In this paper we choose \(\lambda = 6 h^{-1} \text{Mpc}\), which is approximately equal to the galaxy correlation length. This smallest scale yields the highest resolution measure of the three-dimensional topology and the greatest statistical power because of the large number of resolution elements. This scale also gives the greatest amount of information about nonlinear gravitational effects and biasing involved in galaxy formation.

We establish density contour surfaces labeled by \(\nu\), where the volume fraction on the high-density side of the density contour surface is \(f\):

\[
f = \frac{1}{\sqrt{2\pi}} \int_{\nu}^{\infty} e^{-x^2/2} dx.
\]
The genus as a function of \( \nu \) is given by

\[
G(\nu) = \text{number of donut holes} - \text{number of isolated bounding surfaces} \tag{3}
\]

(Gott et al. 1986). Thus, an isolated, solid cluster has a genus of \(-1\) by this definition. (Notice that this definition differs by 1 from the usual definition of the genus, under which an isolated, solid object has genus 0. Notice also that the genus is a property of surfaces and not volumes: a spherical shell has genus 0, not \(-2\), by our definition.) We have shown that \( G(\nu) \) is also equal to minus the integral of the Gaussian curvature over the area of the contour surface divided by \( 4\pi \), which enables us to measure the genus with a computer program (CONTOUR 3D; see Gott et al. 1986, 1987). This comes from the Gauss-Bonnet theorem and the fact that \( G(\nu) = -1/2EC \), where EC is the Euler characteristic that is, for example, equal to 2 for a single solid object.

For a Gaussian random-phase density field, the genus per unit volume, \( g(\nu) \equiv G(\nu)/V \), is given by

\[
g(\nu) = A(1 - \nu^2)e^{-\nu^2/2}, \tag{4}
\]

where the amplitude \( A = (\langle k^2 \rangle / 3)^{1/2}/(2\pi^2) \) depends only on the average value of \( k^2 \) integrated over the smoothed power spectrum (Doroshkevich 1970; Adler 1981; rederived and applied to the \( \nu = 0 \) regime by Hamilton et al. 1986). Thus, the amplitude \( A \) [units: genus/(4\pi^2)] can tell us about the primordial power spectrum. For a Gaussian random field, the median density contour \( (\nu = 0, f = 50\% \text{ volume enclosed}) \) exhibits a spongelike topology (many holes and no isolated regions); the \( f = 7\% \text{ high-density contour (}\nu = 1.5\text{)} \) shows isolated clusters, while the \( f = 93\% \text{ density contour (}\nu = -1.5\text{)} \) is dominated by isolated voids. We call the curves \( G(\nu) \) and \( g(\nu) \) (which differ only by a constant factor for a given sample) the “genus curves.”

For the purpose of examining departures of the observed genus curve from the Gaussian random-phase prediction, we parameterize the genus curve by several derived quantities. First is the best-fit amplitude,

\[
A = \text{amplitude of the genus curve}, \tag{5}
\]

which we measure by a least-squares fit of the theoretical random-phase curve to the data, fitting only in the range \(-1 < \nu < 1\). For the random-phase case, this amplitude is proportional to \( (\langle k^2 \rangle / 3)^{1/2}/(2\pi^2) \) of the smoothed power spectrum and so gives information about the primordial power spectrum. For observations, this amplitude appears lower because of nonlinear clustering and biasing due to coalescence of structures (Park & Gott 1991b; Vogeley et al. 1994; Canavezes et al. 1998).

We quantify shifts and deviations of the genus curve from the shape of the random-phase curve using the following three variables. We measure vertical shifts of the central part of the genus curve with

\[
\Delta \nu = \frac{\int_{-1}^{1} g(\nu)\nu\,d\nu}{\int_{-1}^{1} g(\nu)\,d\nu}, \tag{6}
\]

where \( g(\nu) \) is the genus of the random-phase curve following the formula in equation (4), using the fitted amplitude \( A \) above. The theoretical curve (eq. [4]) has \( \Delta \nu = 0 \). A negative value of \( \Delta \nu \) is called a “meatball shift,” as it is caused by a greater prominence of isolated connected high-density structures that push the genus curve to the left. A positive value of \( \Delta \nu \) is called a “bubble shift,” as it can be caused by a greater prominence of isolated voids and might be produced by isolated explosions (Ostriker & Cowie 1981) as opposed to inflation. A slight, statistically significant meatball shift \( (\Delta \nu < 0) \) was observed first by Gott et al. (1989), who examined the CFA, Giovanelli & Haynes, and Tully data sets. In hindsight, one can see a slight meatball shift in the very first genus curve ever measured (Gott et al. 1987), and this meatball shift was also seen for brighter galaxies in an analysis of an earlier sample of the SDSS (Park et al. 2005b). This shift is presumably due to nonlinear galaxy clustering and bias associated with galaxy formation.

To quantify departures of the observed genus from the random-phase prediction in the region of the genus curve where isolated voids should dominate, we measure

\[
A_v = \frac{\int_{-1}^{1} g(\nu)\,d\nu}{\int_{-1}^{1} g(\nu)\,d\nu}, \tag{7}
\]

where \( g(\nu) \) is again the genus of the best-fit random-phase curve following the formula in equation (4) (see Park et al. 2005a for an explanation of the choice of range in \( \nu \)). As shown in Park et al. (2005a), a value of \( A_v < 1 \) can be the result of biasing in galaxy formation because voids are very empty and can coalesce into a few larger voids. \( A_v \) is sensitive to the number of isolated voids, but the density contour (at \( \nu = -1.7 \)), for example) is given by the volume fraction, so if \( A_v \) is less than 1, and by definition there is the same volume in the low-density regions being measured, there must therefore be fewer but larger voids. Nonlinear clustering alone at these scales predicts a value of \( A_v > 1 \) for the power spectrum of the CDM model we adopt (see Fig. 1 of Park et al. 2005a), so observing \( A_v < 1 \) may be an indication of bias in galaxy formation.

Similar to \( A_v \), we measure a quantity \( A_r \), that characterizes departure from random-phase behavior in the part of the genus curve expected to be sensitive to the number of isolated high-density regions (clusters),

\[
A_r = \frac{\int_{-1}^{1} g(\nu)\,d\nu}{\int_{-1}^{1} g(\nu)\,d\nu}. \tag{8}
\]

A value of \( A_r < 1 \) may occur because of nonlinear clustering, when clusters collide and merge. Also if there is a single large connected structure like the Sloan Great Wall, this can also lower the value of \( A_r \). Also, as Park et al. (2005a) have shown, the Matsubara (1994) formula for second-order gravitational nonlinear effects has the result that \( A_r + A_v = 2 \) at all scales, so if we observe both \( A_v \) and \( A_r \), it is less than 1 (as we find below to be the case), biased galaxy formation must be involved.

3. N-BODY SIMULATIONS OF LARGE-SCALE STRUCTURE

Before we confront results of current simulations of the flat \( \Lambda \)CDM model with the best observations of the topology of the galaxy distribution currently available, it is instructive to consider the remarkable success to date of large \( N \)-body simulations in modeling large-scale structure. It is encouraging that as the volume and resolution in \( N \)-body simulations have grown with time, the Matsubara (1994) formula for second-order gravitational nonlinear effects has the result that \( A_r + A_v = 2 \) at all scales, so if we observe both \( A_v \) and \( A_r \), it is less than 1 (as we find below to be the case), biased galaxy formation must be involved.

Peebles did the first large \( N \)-body simulation for cosmology using 1000 dark matter particles with \( \Omega_m = 1 \) and Poisson initial conditions. It showed clusters like the Coma cluster forming
from random fluctuations by gravitational instability and a reasonable covariance function. Aarseth et al. (1979) used 4000 particles with initial conditions that had more power on large scales than Poisson (index $n = -1$). The Aarseth et al. N-body simulations included $n = 0$ power spectra (Poisson initial conditions), and $n = -1$ power spectra generated by placing the galaxies initially randomly on randomly placed rods. That made the initial covariance function positive and falling off like $1/r^2$, which gave it an $n = -1$ power spectrum. The latter initial conditions were not Gaussian random-phase but did have more power on large scales than Poisson. These simulations were started cold and allowed to develop velocity perturbations (a technique that Gunn & Gott [1972] showed gives similar results to a more proper treatment that starts particles with small velocity perturbations based on the Zel’dovich approximation to simulate growing modes—later pioneered by Klypin & Shandarin [1983]). They found power-law covariance functions quite like those observed even for models with $\Omega_m < 1$ and $n = -1$ (Gott et al. 1979; Gott & Turner 1979) as originally proposed theoretically by Gott & Rees (1975). (Indeed, inflationary flat $\lambda$ models popular today have $\Omega_m < 1$ and more power on large scales than Poisson initial conditions, just as these early simulations suggested.) They also found voids as large as those observed. The largest voids had volumes such that at the mean density they would have contained as much mass as the Coma type clusters contained. This was reasonable from theoretical considerations of non-linear clustering, considering cluster (Gunn & Gott 1972) and void (Bertschinger 1985; Fillmore & Goldreich 1984) formation from small fluctuations via gravitational instability. In Gaussian random-phase initial conditions, isolated over-and-under-dense regions in the initial conditions should be equal in mass leading to equal mass great clusters and empty voids lacking the same amount of mass. They also found that such $\Omega_m < 1$ models with more power at large scales than Poisson produced better multiplicity functions than $\Omega_m = 1$ Poisson models (Gott & Turner 1977; Bhavsar et al. 1981).

The advent of inflation brought for the first time realistic theoretical power spectra to input into N-body models. For example, Klypin & Shandarin (1983) made an N-body simulation using $32^3 = 32,768$ particles. They found a network of filaments connecting clusters (a network of filaments connecting clusters gives a sponge-like topology as we have discussed above). Together, inflation and cold dark matter specified reasonable initial conditions. A suite of such simulations by Davis et al. (1985) provided an impressive match to many aspects of the observed large-scale structure.

Just as theory seemed to be converging on the now-disproven “standard CDM model” with $\Omega_m = 1$, the observations provided a shock. De Lapparent et al. (1986) found many voids 50 $h^{-1}$ Mpc across. This caused a number of people to abandon Gaussian random-phase initial conditions and gravitational instability, favoring explosions to produce the voids instead (Ostriker & Cowie 1981). Then Geller & Huchra (1989) discovered the CfA Great Wall of galaxies, which surprised everyone. To some it appeared to be a fatal blow for random-phase initial conditions, since one expected the covariance function to die at a scale of about 30 $h^{-1}$ Mpc and here was a structure that was 150 $h^{-1}$ Mpc long.

However, the jump to abandon random-phase initial conditions ignored the fact that no N-body simulations had been done by that time that were large enough to properly model structures as large as the CfA Great Wall. When Park (1990) did such simulations using 4 million particles, simulating such a volume for the first time, the results showed that such great walls form routinely. In fact, a 20” thick slice survey through the simulations was a near-perfect visual match to the Geller & Huchra survey. These simulations used a standard peak biasing scheme and included both standard CDM and $\Omega_m = 0.4$ models. Narrow 6” thick slices showed prominent large voids like those in the De Lapparent et al. slice, and great walls appeared in 20” thick slices. This simulation showed weak narrow walls and filaments of galaxies inside the voids, as seen in the CfA data (Park et al. 1992).

In similar fashion, N-body dark matter simulations large enough to mimic the deep pencil beam surveys of Broadhurst et al. (1990) showed apparently regular spikes (walls) of galaxies just like those observed (Park & Gott 1991a). And large N-body simulations (Park 1991) showed great attractors just like those observed by Lynden-Bell et al. (1988). Great repulsors are not seen because such peaks in the gravitational potential occur in the middle of large voids where there are too few tracer galaxies. N-body simulations including hydrodynamics have been successful in modeling the Ly$\alpha$ forest (Cen et al. 1994; Hernquist et al. 1996).

Prior to this study we did an analysis of the topology of a large N-body computer simulation made to mimic the Sloan Digital Sky Survey (Colley et al. 2000). This 54 million particle simulation was observed from one location to simulate what will be seen by the Sloan Digital Sky Survey, to produce sky maps, slices, and 3D topology maps, to show the power of the survey. The sky map looked astonishingly like real sky maps made to similar depth, and the slice maps looked quite like similar survey maps made in the Las Campanas Survey (Kirshner et al. 1981) and now seen in the SDSS. The cosmological model for this simulation was the flat ΛCDM model, which remains in favor.

Even larger simulations are available today, and we are interested to see how they fare in their ability to model the topology of large-scale structure. The “Millennium Run” (MR), using over 10 billion (2160$^3$) dark matter particles (Springel et al. 2005) and surveying a cube of side length 500 $h^{-1}$ Mpc, has shown structures remarkably like the Great Wall found by Geller & Huchra (1989), and even wall complexes somewhat resembling the Sloan Great Wall that Gott et al. (2007) measured to span 1.37 billion light years (Springel et al. 2006). Indeed, Figure 1 in Springel et al. (2006) shows a remarkable visual agreement between what is seen in the MR and in slices of the CfA, the 2dF survey, and the SDSS. The most noticeable difference is that the Sloan Great Wall looks much more visually prominent and coherent than the longest chain of walls found in the MR. (In this simulation, the box size of 500 $h^{-1}$ Mpc cuts off the power spectrum at larger scales. If a simulation were to be made with a larger box size, it would have more power at these larger scales and therefore could more easily produce large coherent structures like the Sloan Great Wall.) The MR computes dark matter halo formation merger trees and uses a semianalytic model to simulate the galaxy formation process where star formation and feedback are modeled by simple analytic physical models. Croton et al. (2006) have produced mock galaxy samples of their cube that include galaxies brighter than the Magellanic clouds, including absolute magnitudes on the SDSS system, which allow us to make mock SDSS galaxy samples.

Park et al. (2005a) produced 8.6 billion particle (2048$^3$) simulations that cover volumes of (1024 $h^{-1}$ Mpc)$^3$ and (5632 $h^{-1}$ Mpc)$^3$. These simulations employ a halo occupation distribution method to place an appropriate number of galaxies in heavy halos identified by the PSB (Kim & Park 2006) and FoF techniques. These simulations were used to analyze the effects of galaxy formation and bias on topology by Park et al. (2005a). More recently, Kim & Park have produced 1.1 billion particle
simulations covering a volume of \((614 \, h^{-1} \, \text{Mpc})^3\). Here they use a new technique to identify physically bound dark matter halos (not tidally disrupted by larger structures) at the present epoch and identify these with galaxies (we call these the DH simulations, for dark [matter] halos). They too have produced magnitudes for these mock galaxies on the SDSS system by matching the halo mass function with the luminosity function of the SDSS galaxies.

R. Cen & K. Ostriker (2008, in preparation) have run hydrodynamic simulations covering a smaller cube of \((120 \, h^{-1} \, \text{Mpc})^3\) on a side using an 8.6 billion \((2048^3)\) cell hydro mesh, with \((1024^3)\) dark matter particles. Here the galaxy formation process is simulated with a hydrodynamic code that identifies collapsing regions, calculates star formation rates, and includes radiative cooling/heating, UV background radiation with local attenuation, and supernova feedback associated with star formation. Again, some assumptions about star formation are made, but this model has one of the most detailed and direct physical calculations of the galaxy formation process available for any simulation that spans a cosmologically interesting volume. For further details of the simulation, we refer the readers to Cen et al. (2005). Nagamine has produced the mock catalogs giving absolute magnitudes on the SDSS system from the R. Cen & K. Ostriker (2008, in preparation) hydro simulations.

The MR simulation, the Kim & Park DH simulation, and the Cen & Ostriker hydro simulation represent state-of-the-art simulations for different schemes to mimic galaxy formation. While the astronomical community seems to be converging on a standard model for cosmology—the flat \(\Lambda\)CDM model (see, e.g., Reiss et al. 1998; Perlmutter et al. 1999; de Bernadis et al. 2000; Spergel et al. 2006)—galaxy formation remains an unsolved problem. This means that since only one cosmological model need be simulated, larger N-body runs exploring different galaxy formation scenarios from different teams can be run. Since the parameters in the semianalytic models have been tuned to account for other features such as covariance function and multiplicity function, and topology was not considered, topology is a particularly stringent test. If the models produce the right topology automatically, it would constitute dramatic evidence that their galaxy formation scenarios were on the right track. In any case, a successful model must show the universe in all its features, including topology.

4. SLOAN DIGITAL SKY SURVEY DATA

The SDSS (York et al. 2000; Stoughton et al. 2002; Adelman-McCarthy et al. 2006) is a survey to explore the large-scale distribution of galaxies and quasars by using a dedicated 2.5 m telescope (Gunn et al. 2006) at Apache Point Observatory. The photometric survey has imaged roughly \(\pi\) steradians of the Northern Galactic Cap in five photometric bandpasses denoted by \(u, g, r, i,\) and \(z\) centered at 3551, 4686, 6165, 7481, and 8931 \(\AA\), respectively, by an imaging camera with 54 CCDs (Fukugita et al. 1996; Gunn et al. 1998). The limiting magnitudes of photometry at a signal-to-noise ratio of 5:1 are 22.0, 22.2, 22.2, respectively, by an imaging camera with 54 CCDs (Fukugita et al. 1996; Gunn et al. 1998). The limiting magnitudes of photometry at a signal-to-noise ratio of 5:1 are 22.0, 22.2, 22.2, 21.3, and 20.5 in the five bandpasses, respectively. The median width of the PSF is 1.4\(''\), and the photometric uncertainties are 2% rms (Abazajian et al. 2004). See Ivezić et al. (2004) for details of assessment of photometric quality and Tucker et al. (2006) for discussion of the monitor telescope pipeline employed for calibration.

After image processing (Lupton et al. 2001; Stoughton et al. 2002; Pier et al. 2003) and calibration (Hogg et al. 2001; Smith et al. 2002), targets are selected for spectroscopic follow-up ob-

The spectroscopic survey is planned to continue through 2008 as the Legacy survey and yield about \(10^6\) galaxy spectra. The spectra are obtained by two dual fiber-fed CCD spectrographs. The spectral resolution is \(\lambda/\Delta\lambda \sim 1800\), and the rms uncertainty in redshift is \(\sim 30 \, \text{km} \, s^{-1}\). Because of the mechanical constraint of using fibers, no two fibers can be placed closer than 55' on the same tile. Mainly due to this fiber collision constraint, incompleteness of the spectroscopy survey reaches about 6% (Blanton et al. 2003a) in such a way that regions with high surface densities of galaxies become less prominent even after adaptive overlapping of multiple tiles. This angular variation of sampling density is accounted for in our analysis.

The SDSS spectroscopy yields three major samples: the main galaxy sample (Strauss et al. 2002), the luminous red galaxy sample (Eisenstein et al. 2001), and the quasar sample (Richards et al. 2002). The main galaxy sample is a magnitude-limited sample with apparent Petrosian \(r\)-magnitude cut of \(m_r, \text{lim} \approx 17.77\) which is the limiting magnitude for spectroscopy. It has a further cut in Petrosian half-light surface brightness \(m_{\text{hls}}\lim = 24.5 \, \text{mag} \, \text{arcsec}^{-2}\). More details about the survey can be found on the SDSS Web site.\(^5\)

In our study, we use a subsample of SDSS galaxies known as the New York University Value-Added Galaxy Catalog (NYU-VAGC; Blanton et al. 2005). This sample is a subset of the recent SDSS Data Release 5. One of the products of the NYU-VAGC used here is Large-Scale Structure sample DR4plus (LSS-DR4plus). We use galaxies within the boundaries shown in Figure 1 of Park et al. (2007), which improves the volume-to-surface area ratio of the survey (important when smoothing). There are also three stripes in the Southern Galactic Cap observed by SDSS. Density estimation is difficult within these narrow stripes, so we do not use them.

The remaining survey region covers 4471 \(\text{deg}^2\) (1.362 sr). The primary sample of galaxies used here is a subset of the LSS-DR4plus sample referred to as void0, which is further selected to have apparent magnitudes in the range \(14.5 < r < 17.6\) and redshifts in the range \(0.001 < z < 0.5\). These cuts yield a sample of 312,338 galaxies. The roughly 6% of targeted galaxies that do not have a measured redshift due to fiber collisions are assigned the redshift of their nearest neighbor.

Completeness of the SDSS is poor for bright galaxies with \(r < 14.5\) because of both the spectroscopic selection criteria (which exclude objects with large flux within the 3" fiber aperture; the cut at \(r = 14.5\) is an empirical approximation of the completeness limit caused by that cut) and the difficulty of obtaining correct photometry for objects with large angular size. For these reasons, analyses of SDSS galaxy samples have typically been limited to \(r > 14.5\); using the magnitude limits of the void0 sample, the range of absolute magnitude is only 3.1 at a given redshift.

The comoving distance and redshift limits of the volume-limited sample we analyze are determined from absolute magnitude limits obtained by using the formula

\[ m_{r, \text{lim}} - M_{r, \text{lim}} = 5 \log[(1 + z)r] + 25 + \bar{K}(z) + \bar{E}(z), \quad (9) \]

where \(\bar{K}(z)\) is the mean \(K\)-correction, \(\bar{E}(z)\) is the mean luminosity evolution correction, and \(r\) is the comoving distance corresponding to redshift \(z\). We adopt a flat \(\Lambda\)CDM cosmology with density parameters \(\Omega \Lambda = 0.73\) and \(\Omega_m = 0.27\) to convert redshift to

\(^5\) See http://www.sdss.org/dr5/.
comoving distance. To determine sample boundaries we use a polynomial fit to the mean $K$-correction,

$$\bar{K}(z) = 3.0084(z - 0.1)^2 + 1.0543(z - 0.1) - 2.5 \log (1 + 0.1).$$

We apply the mean luminosity evolution correction given by Tegmark et al. (2004), $E(z) = 1.6(z - 0.1)$. The rest-frame absolute magnitudes of individual galaxies are computed in fixed bandpasses, shifted to $z = 0.1$, using Galactic reddening corrections (Schlegel et al. 1998) and $K$-corrections as described by Blanton et al. (2003b). This means that a galaxy at $z = 0.1$ has a $K$-correction of $-2.5 \log (1 + 0.1)$, independent of its SED.

From this sample, we construct a volume-limited sample containing galaxies brighter than absolute magnitude $M_r = -20.2$ and fainter than $M_r = -21.7$, and spanning comoving distance from $171.3$ to $344.5 h^{-1}$ Mpc (corresponding to $z = 0.0578 - 0.1178$). This observational sample is similar to the BEST sample studied in Park et al. (2005b) but now larger in extent. This volume-limited sample contains 70,781 galaxies (before overlap correction) and has a mean galaxy separation of $6.097 h^{-1}$ Mpc, so we can safely apply a Gaussian smoothing length of $6 h^{-1}$ Mpc. Numerous numerical experiments have shown that if the smoothing length is smaller than $1/\sqrt{2} \approx 0.71$ times the mean intergalaxy separation there can be a “meatball shift” due to the algorithm picking out individual galaxies as isolated high-density regions (Gott et al. 1987, 1989). In this sample the smoothing length is approximately equal to 0.98 times the mean interparticle separation, so this shot noise effect should be small. In any case, this is not critical for our analysis because we compare the observations directly with mock galaxy catalogs from $N$-body simulations. These mock catalogs are constructed to cover exactly the same range in absolute magnitude as seen in the observations and so contain very nearly (within a few percent) the same total number of galaxies in the sample. Because the techniques being applied to the observations and the $N$-body simulation mock catalogs are identical, the results should be identical (within statistical variation) if the $N$-body simulations are correctly modeling the distribution of galaxies.

5. TOPOLOGY OF LARGE-SCALE STRUCTURE IN THE SDSS

In a previous paper (Park et al. 2005b) we analyzed the three-dimensional topology of large-scale structure in the SDSS at a range of smoothing lengths and compare this with theoretical expectations. In the present paper we focus on results with a smoothing length of $6 h^{-1}$ Mpc, which yields the most resolution elements and gives the most important information on galaxy formation. The sample of galaxies available has now grown significantly larger, and so we are now able to make direct comparison of this sample with large $N$-body simulations and their various methods of modeling of galaxy formation. As we see below, the observational sample is now large enough that the topology, as measured by the genus curve $g(v)$, is now a powerful tool for testing models of galaxy formation.

Figure 1 shows the progression by date of survey of the 3D topology of selected galaxy redshift surveys. All have similar smoothing lengths of $5-6 h^{-1}$ Mpc and show the median density contour surface with the high-density regions shown as filled areas and the low-density regions as open. According to standard inflationary theory, this median density contour should be spongelike. The small cube on the left shows the 3D region studied by Gott et al. (1986). The Earth is at the lower front right corner of the cube. The topology is spongelike, and the Virgo cluster is included in the high-density region. The larger region of isodensity contours in the center of this figure is from the CfA redshift survey (Vogeley et al. 1994). The Earth is at the center, and the upper fan-shaped region is in the North Galactic Hemisphere while the lower fan-shaped region lies within the South Galactic Hemisphere. The Great Wall noted by Geller & Huchra (1989) can be seen connecting high-density regions across the top fan-shaped region. Again, the topology is spongelike, with the high-density regions all connected together, and the low-density regions also connected in an interlocking pattern.

Finally, the portion of the SDSS data now available (in 2006, a full 20 years after the first figure) is shown on the right of Figure 1. This is the largest region yet studied for topology and contains nearly 400,000 galaxies in total. The location of the Earth is at the back. The horizontal slice extending out toward
us is the northern equatorial slice of the SDSS and includes the Sloan Great Wall (Gott et al. 2005). The upper slice is a second contiguous thick region in the northern hemisphere of the SDSS. (When SDSS-II is complete in 2008, the gap between these two slices will be filled in.) It is easy to see that the topology of this median density contour is spongelike. The high-density regions (taking up half the volume) form one multiply connected region (filled area) and the low-density regions (taking up the other half of the volume) also form one multiply connected region that is interlocking with the high-density region.

Figure 2 shows the same regions of the SDSS, but with iso-density contours in different colors for different volume fractions. The 7% high-density regions—containing the highest density 7% of the volume—are red. The end of the Sloan Great Wall can be seen as the red structure snaking from the left to the right in the Equatorial slice. This red contour also shows isolated high-density regions (clusters) as expected from the random-phase genus curve. The 50% high-density contour is shown in transparent green—this contour is a multiply connected spongelike surface that divides the high-density half of the sample from the low-density half. The 7% low-density regions are shown as solid blue and show isolated voids. The red and blue regions lie on opposite sides of the transparent green spongelike surface.

To facilitate viewing the three-dimensional nature of the density contours, Figure 3 shows a stereo pair of the same SDSS contours. This offers our best picture yet of the 3D topology of large-scale structure in the universe.

Figure 4 shows the genus curve of this volume-limited SDSS sample smoothed at \( \lambda = 6 \, h^{-1} \) Mpc. In this figure we compare the observed genus curve with results for mock surveys produced from the N-body simulations, which we describe in Section 6 below. Also shown in Figure 4 is the random-phase genus curve that best fits the SDSS genus curve. The data approximately follow this random-phase curve, as expected from inflation. However, there are measurable departures that are likely to have been caused by nonlinear effects and galaxy formation. We characterize these differences from the random-phase curve using the measures \( \Delta \nu, A, \Delta f, A_c \), and \( A_e \) described above and plot these values in Figures 5 and 6.

Figure 2.—The 7% low (blue), 50% (green), and 7% high (red) volume contours in our SDSS sample. The Sloan Great Wall is visible as the long red structure in the lower slice.
Great Wall, has by itself a value of $\Delta \nu = -0.08$. Thus, the meatball shift in the data is a general phenomenon and is not due solely to the Sloan Great Wall.

Such a meatball shift has been seen in observational samples before. It was first noticed and commented on by Gott et al. (1989), who examined the CfA, Giovanelli & Haynes, and Tully samples. Gott et al. (1996) found that hydrodynamic simulations predicted a meatball shift for (early-type) elliptical galaxies relative to (late-type) spiral galaxies (elliptical galaxies tend to congregate more in isolated rich clusters), an effect later observed in the 2D topology analysis of the SDSS by Hoyle et al. (2002b), where a meatball shift in red (early-type) galaxies was seen relative to blue (late-type) galaxies. Thus, it is clear that galaxy formation processes can produce meatball shifts relative to the random-phase curve, as we observe with high statistical significance. The question is whether our galaxy formation models accurately reproduce this effect. Below we discuss whether they successfully model this shift in the genus curve.

Figure 6 shows the $(A_v, A_e)$ plane with the SDSS data again shown as a filled blue square, with the random-phase prediction shown as a black circle. Again, crosses indicate measurements for the two regions of the SDSS considered separately. The data have values of $(A_v, A_e)$ that depart from the random-phase values of $(1, 1)$ due to biased galaxy formation and nonlinear effects.

6. TOPOLOGY OF SDSS DATA VERSUS SIMULATIONS

The genus curve for the SDSS sample shows a clear shift toward a “meatball” topology and behavior in the void and cluster-dominated tails that indicate fewer isolated voids and clusters than expected from either a Gaussian random-phase distribution or from perturbation theory (see Matsubara 1994; Park et al. 2005b). In this section we examine whether current simulations...
of large-scale structure reproduce these features and use the simulations to assess the statistical significance of these departures from random-phase behavior.

Our approach is to construct mock SDSS redshift samples from each of the simulations that mimic the observational selection effects caused by the survey geometry, sampling density of structure, and redshift-space distortions. In the DH and MR cases, we construct many such mock surveys, smooth and compute the genus curves, and compute the variables \( A, A_v, A_c \) for each. Then we compute the mean and standard deviations of those statistics, as plotted in Figures 5 and 6.

In considering the predictions of the various \( N \)-body simulations it is important to estimate the cosmic variance one may encounter. We start with the DH simulation. This simulation has a volume of \((614 \ h^{-1} \text{ Mpc})^3\), which is a bit over 16 times the volume of the SDSS sample, allowing roughly that many independent mock surveys. We create 100 such surveys of the SDSS in order to fully sample the structure in the cube with the irregular shape of the SDSS; they are clearly not independent, but the mean and distribution function of the genus statistics are not affected by this. We show the mean and standard deviations of the genus quantities for the Park et al. simulations in Figures 5 and 6 with a green hexagon and associated uncertainty limits. It is clear that the SDSS is more than \( 3 \sigma \) away from the mean value of \( A \). In fact, none of the 100 mock surveys has a value of \( A \) as negative as the observations. It could be argued that the SDSS contains the Sloan Great Wall, which is so unusual that this region should be excluded from the analysis. However, this is exactly the purpose of the 100 mock surveys: to examine the range of values we expect in SDSS-sized surveys. So the fact that the SDSS is outside the \( 2 \sigma \) error bars from the cosmic variance expected shows that the DH simulations are not successful in predicting the observed meatball shift. Also, recall that the values of \( \Delta \nu \) for both SDSS subsamples (one not containing the Sloan Great Wall) are outside these limits.

The DH simulations are perfect in amplitude relative to the data. Importantly, the DH simulations (as well as the Millenium Run and the Cen and Ostriker simulations) include the effects of baryon oscillations, which give the correct initial power spectrum. A similar DH simulation with 8.6 billion particles but without baryon oscillations had an amplitude of \( 209 \pm 11 \), which is \( 1.6 \sigma \) high relative to the data (191). This shows that the topology can detect the presence of the baryon oscillations by measuring the slope of the power spectrum; simulations without them produce

![Fig. 4](image1.png)

**Fig. 4.**—Genus curves for the SDSS sample, hydro sample, and random samples drawn from the 100 DH and 50 MR studies. The Gaussian random curve is shown for comparison. Notice that we plot \( G(\nu) \) (i.e., do not divide by the sample volume) in order to show the genus of the entire sample at each \( \nu \).

![Fig. 5](image2.png)

**Fig. 5.**—Plot of the genus shift parameter \( \Delta \nu \) vs. the genus-curve amplitude \( A \) for our samples. The black circle corresponds to Gaussian random phase. The blue square is the SDSS sample, with the two blue crosses representing each of the two SDSS subregions (normalized in amplitude to the volume of the whole sample). The green hexagon is the mean of the 100 DH mock Sloans, with error bars representing the standard deviation of that sample. The red pentagon is the mean of the 50 MR mock Sloans, with error bars showing that standard deviation. The pink cross denotes the hydrodynamic simulation of Cen & Ostriker; the error bars correspond to the cosmic variance of \((120 \ h^{-1} \text{ Mpc})^3\) subregions within a \((500 \ h^{-1} \text{ Mpc})^3\) box.

![Fig. 6](image3.png)

**Fig. 6.**—Same as Fig. 5, but plotting \( A_v \) and \( A_c \). The upper left blue cross corresponds to the SDSS region without the Sloan Great Wall.
an amplitude of the genus curve that is too high. The values of $\Delta \nu = +0.020 \pm 0.027$ for those simulations without the baryon oscillations were not appreciably different from the ones that include the baryon oscillations.

For the MR simulation, we have a volume of $(500 \, h^{-1} \, \text{Mpc})^3$, which enables us to make roughly 10 independent mock surveys of the SDSS. Again, to fully sample the cube, we make 50 mock surveys at random position and orientation. The mean and standard deviation of these mock surveys are shown in Figures 5 and 6 as a red pentagon with associated error bars. This indicates the cosmic variance seen from SDSS-sized mock surveys drawn from the $(500 \, h^{-1} \, \text{Mpc})^3$ survey region.

However, we must also consider the effect of cosmic variance on the scale of the simulation region: error bars on simulations with smaller box sizes will be somewhat underestimated with respect to larger (e.g., the MR with respect to the DH), and with respect to the real universe (which of course has infinite box size). To estimate this added variance, we take eight subcubes of volume $(512 \, h^{-1} \, \text{Mpc})^3$ out of the larger DH simulation of J. Kim & C. Park (2008, in preparation) mentioned above, of volume $(1024 \, h^{-1} \, \text{Mpc})^3$, and make redshift maps of them; i.e., we give the galaxies their correct $x$ and $y$ coordinates but put them at a $z$ coordinate equal to $z + v_z H_0^{-1}$ as if we were viewing a redshift space map of the survey region from a great distance. (Note that even with the different power spectrum, the variance of the genus statistics should still be correct.) We then compute the 3D topology for each of the eight subcubes and measure the standard deviation of the 64 values of the parameters ($\Delta \nu$, $A$, and $A_c$). This cosmic variance from the 64 subcubes produces the error bars surrounding the one value from the Cen et al. simulation (shown as a magenta cross with error bars). We can then add the extra cosmic variance as per the MR data point to produce the standard deviations corresponding to a volume of $(1024 \, h^{-1} \, \text{Mpc})^3$ (again given in Table 1—the extra s. d. is $\leq 1\%$ for all statistics). These error bars are very large; this one simulation is not very constraining of the topological properties. Roughly speaking, the volume of the Cen & Ostriker simulation is an eighth that of the SDSS, so we expect error bars that are roughly $\sqrt{8} \approx 2.8$ times as large as for the cosmic variance seen in the SDSS. This ratio is approximately correct as Figures 5 and 6 show.

In Figure 4 we show the genus curve for the Cen & Ostriker simulation. It is not a good fit to the SDSS observations. The top of the genus curve is chopped off, and it has a meatball shift in the void region that is not seen in the observations. In the central region $-1 < \nu < 1$, where $\Delta \nu$ is measured, the curve is too fat and also has a small negative value of $\Delta \nu = -0.01$. This is more negative than either the MR or the DH simulations but is still not very close to the observed value of $\Delta \nu = -0.08$. The amplitude of the Cen & Ostriker simulation is much lower than the other three data sets. Overall, the Cen et al. simulation does not give a good fit to the data; however, the volume of the simulation is small and the large error bars show that the observations are within $2\sigma$ in both $\Delta \nu$ and $A$. (To be fair, other hydro simulations by Cen and his collaborators—e.g., Gott et al. [1996]—have produced good-looking genus curves, so this one may be suffering from a bit of bad luck.)

So, currently, the Cen & Ostriker hydro simulations are too small. They need to be increased in volume by about a factor of 10 to be fairly tested against the SDSS. In 1975 the largest $N$-body cosmological simulation had 4000 particles (Aarseth et al. 1979). In 1990 the largest $N$-body cosmological simulation had 4 million particles (Park 1990). This led Gott to predict in 1990 that by 2005 the record would be 4 billion particles.

---

**Table 1**

| Name          | Amplitude | $\Delta \nu$ | $A_0$  | $A_c$  |
|---------------|-----------|--------------|--------|--------|
| SDSS          | 190.48    | -0.080       | 0.747  | 0.804  |
| DH            | 190.79 ± 11.65 | 0.022 ± 0.028 | 0.806 ± 0.052 | 0.811 ± 0.067 |
| MR            | 175.42 ± 9.69(± 9.86) | 0.010 ± 0.023(± 0.036) | 0.845 ± 0.057(± 0.081) | 0.862 ± 0.063(± 0.097) |
| Hydro         | 146.23(±31.9)(±31.9) | -0.008[±0.106](±0.107) | 0.783[±0.328](±0.328) | 1.016[±0.218](±0.219) |

**Notes.**—Errors not in parentheses are the standard deviations of the population in question (not given for the real SDSS or the Hydro simulation, where there is only one sample). Errors in parentheses also include the effect of cosmic variation within a $1024 \times 1024$ box size within a $500 \times 500$ box. We would have liked to make a table for both sets of values). We would have liked to make a table for both sets of values). We would have liked to make a table for both sets of values). We would have liked to make a table for both sets of values). We would have liked to make a table for both sets of values).
Springel et al. (2005) did even better than this with over 10 billion particles. An increase of a factor of 10 every 5 years is just what would be predicted by Moore’s law (a doubling every 18 months). So given Moore’s law we should expect that Cen and his colleagues will have a hydro code simulation with volume 10 times larger by 2011. Then we can see if it outperforms the MR or the DH simulations. Just one additional hydro run with the same parameters would also be interesting—would it be closer to the DH simulations. Of the 100 DH mock surveys, the one closest to fitting the observations because the band of MR simulations is farther away from either the MR or DH simulations. Of the 100 DH mock surveys, the one closest to the observations in terms of the four variables and their error regions for the (a) 100 DH and (b) 50 MR samples, compared with SDSS and Gaussian random phase.

Another small but interesting effect has to do with halo or galaxy identification: nearby halos (or galaxies in the hydro simulation) may get merged together and only counted as one point. To estimate this effect, we calculated the genus of SDSS after merging together all galaxies closer than 100 kpc (an overestimate of the actual scale of the problem). This does seem to move SDSS slightly closer to the simulations: the amplitude decreased by ~0.5%, \( \Delta \nu \) increased from \(-0.080\) to \(-0.078\), and \( A_c \) increased by ~1%, while \( A_t \) decreased by about the same amount.

Figure 4 shows the genus curve of the observations versus one of the 50 MR mock surveys picked at random, one of the 100 DH mock surveys picked at random, and the hydro mock survey. The simulation looks the best. The top of the genus curve near \( \nu = 0 \) is not cut off, and it looks most like the observations. Both the MR mock and the hydro mock run are cut off at the top in the same way, the hydro more so. To get a better picture of the cosmic variance, Figure 7 shows the observations compared to hatched bands showing the 1σ variation in the mock runs from the (a) DH and (b) MR simulations. It is clear from this that the one random MR mock survey shown in Figure 4 was worse than average at fitting the top of the curve, but it is also clear that the DH simulation is still better than the MR at fitting the observations because the band of MR simulations are lower in this region. The place where the MR mocks and the DH mocks fail the worst is in the region \( 0.4 < \nu < 1.2 \), where the observations are consistently shifted to the left with respect to the simulations.

We conclude that the simulations do an adequate job of representing the topology in all the variables except \( \Delta \nu \), which for the SDSS data lies more than 2.5σ away from either the MR or DH simulations. Of the 100 DH mock surveys, the one closest to the observations in terms of the four variables and their error bars is mock catalog 94, which has \( \Delta \nu = -0.02 \), \( A = 191 \), \( A_c = 0.80 \), and \( A_t = 0.77 \), which are OK in all except \( \Delta \nu \), where it is still far off the observational value. In fact, of the 100 mock DH simulations the two most negative in \( \Delta \nu \) are between \(-0.05\) and...
−0.04, while the observations show −0.08. The four most negative of the 50 mock MR simulations are between −0.03 and −0.02. Interestingly, the higher spatial resolution and semi-analytical modeling of the MR does not yield a better fit to the observed topology of large-scale structure than the dark matter, halo-finding DH simulations.

7. CONCLUSIONS

The SDSS data set has now become large enough that the topology of large-scale structure can be used for more detailed model testing. We find that the SDSS observations have a sponge-like median contour and follow fairly closely the genus curve expected from Gaussian random-phase initial conditions predicted by inflation. We quantify departures from this theoretical curve that provide key tests of models for galaxy formation, as represented by the several simulations that we examine.

The amplitude of the genus curve is in agreement with that predicted from the standard ΛCDM model (with the WMAP parameters) with baryon oscillations included. If baryon oscillations were not included, the fit to the amplitude would be significantly worse (1.6 σ). The observed values of $A_1$ and $A_2$ are predicted well by the MR and DH simulations. Both show the effects of nonlinear gravitational evolution and biased galaxy formation. The Cen & Ostriker hydro simulations are consistent with the data, but their small volume gives them large error bars, and they are currently not giving values closer to the observations than the MR or DH simulations.

The most notable feature of the observations is a meatball shift $\Delta \nu = −0.08$ showing a slight prominence of isolated high-density regions over isolated voids. The SDSS Great Wall is one large, connected, isolated high-density region and contributes to this effect, but the effect also shows up in the northern part of the SDSS, which does not contain the Sloan Great Wall. If the Sloan Great Wall were entirely responsible for this result, one might argue that it was produced by rare objects whose frequency of occurrence was determined by the power in the initial conditions at very large scales and that even larger simulations $> (1024 \, h^{-1} \, Mpc)^3$ would be needed to properly test for this effect. But this is not the case. Negative values of $\Delta \nu = −0.08$ show up even in the part of the survey that does not include the SDSS Great Wall. Also, slice surveys of the MR simulation show great walls that look quite impressive—if not quite as dramatic as the Sloan Great Wall. The observed $\Delta \nu$ values are more than 2.5 $\sigma$ away from the values found in the MR and the dark matter halo simulations. This is a severe test for large N-body simulations and their heuristic galaxy formation scenarios because these were not tuned to account for topology.

The slight meatball shift seen in the observations has been noticed in previous observational samples with similar smoothing lengths ($6 \, h^{-1} \, Mpc$), being mentioned first by Gott et al. (1989). The large survey by Canaveses et al. (1998) of the IRAS galaxies looks Gaussian random-phase on all larger smoothing lengths, but does have a slight meatball shift with a smoothing length of $5 \, h^{-1} \, Mpc$. Hydrodynamic simulations suggest that early-type galaxies should show more of a meatball shift than late type galaxies (Gott et al. 1996), and this effect has already been observed in the equatorial slice of the SDSS in a 2D topology survey by comparing the relative meatball shift between red and blue galaxies (Hoyle et al. 2002b).

As the SDSS is completed, the gap between the northern and equatorial slices will be closed giving us one large continuous volume, where the fraction of the sample one throws away because of closeness to the edge will be diminished. This will approximately double the effective volume of the sample and give us a still better test. Also, studies with smoothing lengths of 10 and $20 \, h^{-1} \, Mpc$ will be possible with high precision, allowing more direct tests of the Gaussian random-phase hypothesis on scales where the galaxy formation effects are less important.

It would be interesting to see $N$-body simulations covering larger volumes that would have more power at large scales (because they would not be artificially cut off at the box size). This would make for more accurate modeling of the structure and frequency of occurrence of structures like the Sloan Great Wall (Gott et al. 2005).

The results here suggest that in order to account for the observed topology some changes in galaxy formation scenarios are called for. We look forward to improvements in the $N$-body simulations. Of particular interest is how well larger hydrodynamic simulations will perform when compared with larger samples, and whether there will be a convergence of predictions as both hydrodynamic and merger tree, and dynamical halo occupation methods are improved. Galaxy formation is not yet a solved problem in cosmology and the 3D topology offers a strong test of models that is independent of other measures.
