Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **508,** 1189–1194 (2021)
Advance Access publication 2021 September 17

https://doi.org/10.1093/mnras/stab2650

The causal effect of environment on halo mass and concentration

Corentin Cadiou [1]★ Andrew Pontzen,[1] Hiranya V. Peiris[1,2] and Luisa Lucie-Smith [3]

[1]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[2]*The Oskar Klein Centre for Cosmoparticle Physics, Department of Physics, Stockholm University, AlbaNova, SE-106 91 Stockholm, Sweden*
[3]*Max Planck Institute for Astrophysics, Karl-Schwarzschild-Strasse 1, D-85748 Garching, Germany*

Accepted 2021 September 10. Received 2021 September 9; in original form 2021 July 7

ABSTRACT

Understanding the impact of environment on the formation and evolution of dark matter haloes and galaxies is a crucial open problem. Studying statistical correlations in large simulated populations sheds some light on these impacts, but the causal effect of an environment on individual objects is harder to pinpoint. Addressing this, we present a new method for resimulating a single dark matter halo in multiple large-scale environments. In the initial conditions, we 'splice' (i.e. insert) the Lagrangian region of a halo into different Gaussian random fields, while enforcing consistency with the statistical properties of Lambda cold dark matter. Applying this technique, we demonstrate that the mass of haloes is primarily determined by the density structure inside their Lagrangian patches, while the haloes' concentration is more strongly affected by environment. The splicing approach will also allow us to study, for example, the impact of the cosmic web on accretion processes and galaxy quenching.

Key words: methods: numerical – galaxies: formation – galaxies: haloes – dark matter.

1 INTRODUCTION

The growth of dark matter haloes and galaxies can be most accurately computed using numerical simulations. Understanding the physical origin of environmental quenching (e.g. Kauffmann et al. 2004; Peng et al. 2010), intrinsic alignments (e.g. Tempel, Stoica & Saar 2013; Chisari et al. 2015), or colour gradients in the cosmic web (Kraljic et al. 2018; Laigle et al. 2018) are some of the most fundamental open problems in galaxy formation. However, attaining a physical understanding of these effects of cosmological environment on individual galaxies is complicated by the wide variety of possible configurations that are generated by the Gaussian random initial conditions (ICs).

Currently, the main approach to disentangling the impact of environmental factors on galaxy formation is statistical in nature (Aubert, Pichon & Colombi 2004; Codis et al. 2012; Danovich et al. 2012; Kraljic et al. 2019; Martizzi et al. 2020). Analytical models can provide hypotheses for the causal relationships between ICs and final haloes (e.g. Press & Schechter 1974; Sheth, Mo & Tormen 2001; Hahn et al. 2009; Codis, Pichon & Pogosyan 2015; Musso et al. 2018), but it is difficult to test these hypotheses at the level of individual haloes (Borzyszkowski et al. 2017; Lucie-Smith, Peiris & Pontzen 2019).

In this work, we extend the 'genetic modification' (GM) technique (Roth, Pontzen & Peiris 2016), which is designed specifically to construct controlled experiments in cosmological galaxy and halo formation. Previously, GM has been used to control the mass, merger history (Pontzen et al. 2017; Rey & Pontzen 2018), and angular momentum (Cadiou, Pontzen & Peiris 2021) of individual objects. Our extension aims to manipulate instead the large-scale environment, while leaving the density structure of a target object's Lagrangian patch untouched.

We extend the code GENETIC (Stopyra et al. 2021) to embed the ICs that will eventually collapse into a halo into new environments. This can be seen as a 'gene-splicing' operation, combining two Gaussian random fields into a single realization. We apply this technique to investigate how the mass and concentration of haloes in dark matter simulations are affected by environment.

The paper is structured as follows: We first present qualitatively the gene-splicing method and the set of numerical simulations used throughout the paper in Section 2. We then present their analysis in Section 3. Finally, we summarize and discuss our findings in Section 4. A more detailed mathematical derivation of the gene-splicing method can be found in Section A.

2 METHODS

In this section, we first present the 'gene-splicing' technique; a more formal derivation can be found in Section A. We will then discuss how it has been applied to produce a suite of simulations for this first study.

The splicing operation is applied to the linear ICs, which we generate at $z = 100$. We start from two Gaussian random fields representing the overdensity of independent realizations, denoted a and b, and select an arbitrary region Γ. To obtain the results in this paper, we choose Γ to be the Lagrangian region of a $z = 0$ halo (i.e. the region that its constituent particles occupied at $z = 100$).

The splicing operation finds a new field f that satisfies $f(x) = a(x)$ inside Γ, but which closely approximates $b(x)$ elsewhere in the simulation volume. It is not possible to set $f(x) = b(x)$ outside Γ because this would cause discontinuities on the boundary; such

★ E-mail: c.cadiou@ucl.ac.uk

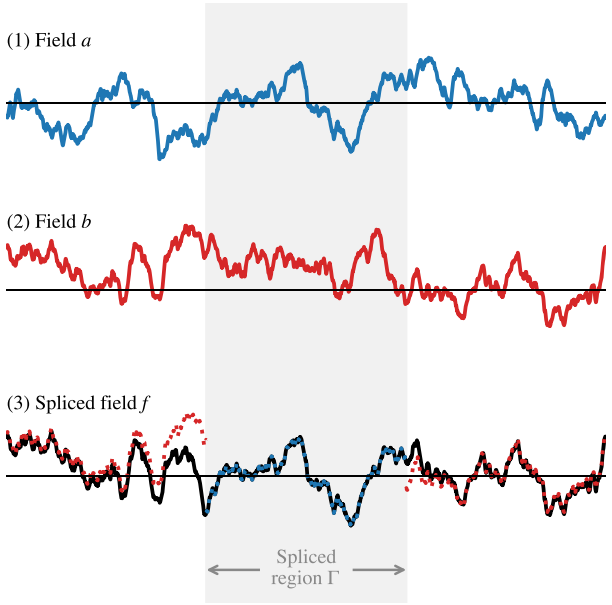

Figure 1. Illustration of the gene-splicing procedure applied to 1D ICs. We draw a field a (in blue) and another independent field b (in red). We obtain the new ICs (in black) by 'splicing' a given region of a into b. The spliced field has the value of a in the spliced region and rapidly converges to the value of b outside it, while remaining maximally consistent with the Gaussian random field statistics.

discontinuities are incompatible with the assumption of a Gaussian random field. Instead, we minimize the χ^2 of the field difference $f(x) - b(x)$. This approach has been motivated at length by Roth et al. (2016) and Rey & Pontzen (2018), and leads to fields that are maximally likely in the Gaussian random ensemble under the constraints. Given the spliced density, we then use the Zel'dovich approximation to generate a corresponding set of particles with new positions and velocities. These are used as ICs for a new N-body simulation.

The algorithm described above is equivalent to altering the field $b(x)$ with a list of modifications specifying the new value of $f(x)$ at every point x_i in Γ. However, applying the existing GM algorithm to this problem becomes quickly impractical as the number of points in Γ increases, requiring $\mathcal{O}\left(N^d \times N_{\rm pt}\right)$ memory, where N is the number of cells in each direction, d is the number of dimensions, and $N_{\rm pt}$ is the number of constrained points. To circumvent this problem, we instead solve the χ^2 difference minimization iteratively using a gradient descent method. We have implemented the method within the code GENETIC (Stopyra et al. 2021) in v1.3 (Pontzen et al. 2021). Further details can be found in Section A.

A one-dimensional (1D) example of a spliced Gaussian random field is illustrated in Fig. 1; the splicing region Γ is indicated by grey shading. The independent fields a and b are shown in the top two panels; the spliced field f is shown in the bottom panel (solid line) along with the relevant portions of the original fields for comparison (dotted lines). The spliced field f can be seen to obey our requirements: It traces a perfectly inside Γ; is continuous on the boundary of Γ; and closely approximates b at large distances from Γ. The rate at which f converges to b depends both on the correlation function (or equivalently the power spectrum) and on the difference between fields a and b around the splicing region boundary. In this test, the reduced χ^2 of realizations a, b, and f are 1.00, 1.03, and

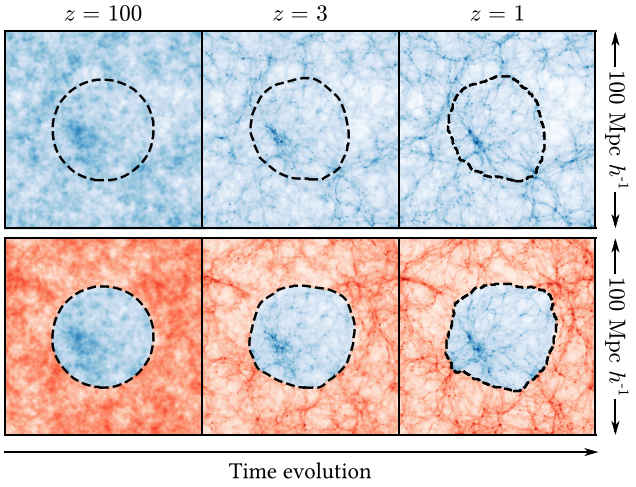

Figure 2. Slices of the dark matter density field evolved from an unmodified set of ICs (top row) and corresponding spliced ICs (bottom row). Regions evolving from the original ICs are coloured in blue; the new external region is coloured in red. The sphere (dashed lines) is tidally distorted over cosmic time, leading to differences between the two simulations in terms of the shape of the boundary. Structures in the spliced region (bottom row, in blue) can be mapped on to their counterparts in the original simulation (top row). Conversely, outside this region, the matter density fields in the two simulations bear no resemblance to each other.

0.99, respectively (with 1499 degrees of freedom), indicating that f is a likely draw from the underlying distribution despite being constructed artificially.

Having shown how splicing works in a 1D example, we next illustrate in Fig. 2 the cosmological evolution of a three-dimensional (3D) spliced field. The top left panel shows our reference ICs at redshift $z = 100$; we use a 256^3 grid in a domain of size $100\,{\rm Mpc}\,h^{-1}$ for a mass resolution of $M_{\rm DM} = 7.7 \times 10^9\,{\rm M}_\odot$. The transfer function is computed using CAMB (Lewis, Challinor & Lasenby 2000) and cosmological parameters consistent with the values of Planck Collaboration VI (2020). The ICs are then evolved using RAMSES (Teyssier 2002), as illustrated in the top row. Gravity is solved using a particle-mesh approach on an adaptive mesh. We allow the mesh to be refined wherever it contains more than eight dark matter particles. The effective minimal force resolution reached by the simulation is 9 kpc physical.

Next, we select a region Γ in the ICs of the reference simulation. As an illustrative example, in Fig. 2 we splice a sphere of comoving radius $25\,{\rm Mpc}\,h^{-1}$. Finally, we draw an independent overdensity field, and splice the sphere into it to form the new ICs; the result is shown in the bottom left panel. The region that is identical to the original ICs is shown in blue, while the external region is shown in red. We evolve the new ICs using an identical simulation configuration to the original.

The time evolution of the sphere in the reference (top row) and spliced (bottom row) simulations can now be compared. We indicate the edge of the sphere (dashed black line), defined by the set of particles that it contains in the ICs as a function of time. The edge of the region is deformed by non-linear structure formation, becoming less spherical with time. This deformation depends on the long-range tidal effect of the region outside the sphere and so the shape of the patches increasingly differs between the two simulations.

The density field within the sphere is identical, by construction, in the two sets of ICs. The subsequent interior gravitational evolution is similar; however, it has small differences, due to the differing large-

scale gravitational forces. The impact of these changes on haloes is the focus of this paper. By contrast, far from the sphere, the ICs are unrelated between the two simulations, and structures in one simulation cannot be mapped to the other. In the case illustrated, a large cosmic void is present in the rightmost region of the unaltered simulation, while a massive filament forms in the spliced simulation.

In the remainder of the paper, we will study how the large-scale environment contributes to setting the mass and concentration of dark matter haloes, as an example of the gene-splicing technique's promise. For this purpose, we performed a reference simulation with identical cosmological and numerical parameters to the example described above, in a domain of size $50\,\mathrm{Mpc}\,h^{-1}$, for a mass resolution of $M_{\rm DM} = 9.7 \times 10^8\,\mathrm{M}_\odot$ and an effective minimal force resolution of 2 kpc physical. From this unmodified simulation, we selected six dark matter haloes with masses between[1] 10^{13} and $10^{14}\,\mathrm{M}_\odot$ at $z = 0$. We select all their member particles as computed by the halo finder – including those in any of their subhaloes – and trace these back to the ICs to obtain the Lagrangian patch. At this point, we have six patches that will eventually form a dark matter halo in the reference simulation. We separately spliced each of these 6 patches into 10 independent realizations of the box, for a total of 60 new ICs that were evolved to $z = 0$.

We extract halo catalogues using ADAPTAHOP (Aubert et al. 2004) and the parameters presented in Tweed et al. (2009) with the 'Most massive Substructure Method' and a minimum number of 200 particles per halo. We analyse the catalogues using TANGOS (Pontzen & Tremmel 2018), which we employ to extract the virial radius R_{200c}, virial mass M_{200c}, and concentration parameter c as we will describe below.

3 RESULTS

We now investigate the effect of environment on dark matter haloes' masses. Our set of 60 simulations corresponds to 10 environmental realizations around each of the 6 central haloes. For each of the six haloes, we compute the mean virial mass $\langle M_{200c}\rangle$ over the ten realizations. We then calculate, for each realization, the ratio of its mass to this mean as follows:

$$r = \frac{M_{200c}}{\langle M_{200c}\rangle}. \tag{1}$$

This yields 60 measurements of r, which are plotted as a histogram in Fig. 3; the masses are scattered by ±15 per cent around the halo's mean value.

Next, as an example of a more detailed structural property of haloes, we measure the concentration parameter using the approach presented by Klypin et al. (2016); see their equations (18)–(20). The NFW concentration parameter, c, is estimated using the implicit solution to

$$\frac{V^2_{\rm circ,max}}{V^2_{200c}} = \frac{c}{x_{\rm max}}\frac{f(x_{\rm max})}{f(c)}, \tag{2}$$

$$f(x) \equiv \ln(1+x) - \tfrac{1}{1+x}, \tag{3}$$

$$x_{\rm max} = 2.163. \tag{4}$$

Here, $V^2_{\rm circ}(r) = GM(<r)/r$ is the circular velocity, $V_{\rm circ,max}$ is its maximum value for $0 \le r \le R_{200c}$, and $V_{200c} = V_{\rm circ}(R_{200c})$. We measure the circular velocities in 100 logarithmically spaced radial bins between $R_{200c}/100$ and R_{200c}. We use this procedure because it is much more stable than fitting the NFW profile directly through χ^2 optimization, which suffers from significant degeneracies. We verified the numerical stability of the Klypin et al. (2016) estimator by calculating the change in c for all our haloes between two adjacent time-steps, finding that its r.m.s. variation is only ±10 per cent. This is negligible compared to the population scatter that we will discuss below.

[1]The individual masses are $\{3.2, 3.3, 5.3, 5.9, 7.2, 8.6\} \times 10^{13}\,\mathrm{M}_\odot$.

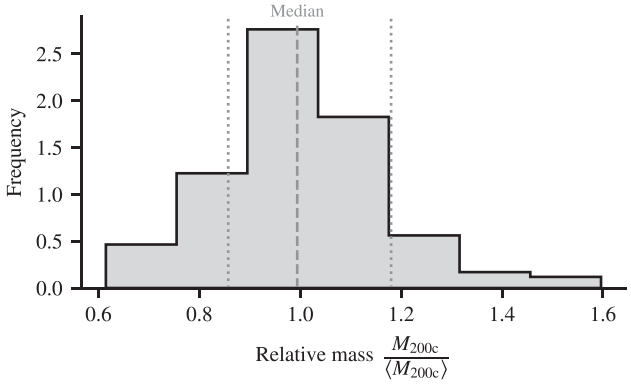

Figure 3. When splicing haloes into a new realization, their mass changes due to environmental effects. For our six haloes, each simulated in ten different environments, we find that the change in mass is modest. The histogram shows the new mass divided by the mean over the 10 realizations. Vertical lines indicate the median (dashed) and 68 per cent credible interval (dotted), showing that the mass typically scatters only by ±15 per cent.

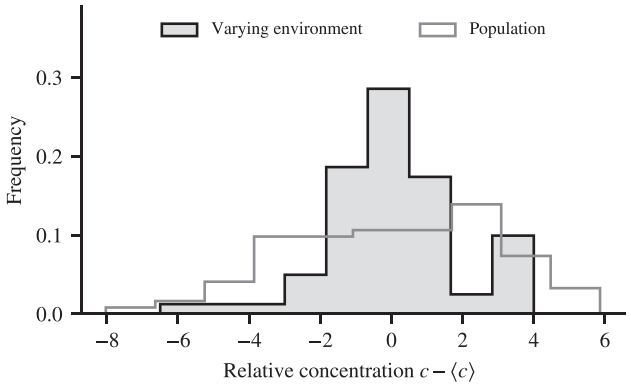

Figure 4. The scatter in the concentration induced by placing haloes in a new environment is highly significant. For each halo, we calculate the scatter around its mean concentration in the 10 environments. The shaded histogram shows the resulting distribution for all six haloes, which can be compared to the scatter in concentration within the population (light histogram). At least half the scatter of the concentration can be attributed to the effect of environment.

For each of the six halo families, we compute $\langle c\rangle$, where the average is taken over the ten environments. We then calculate a distribution of $c - \langle c\rangle$ over all 60 simulations. To contextualize this distribution, we create a second ensemble, consisting of all 88 haloes in the original reference run in the same mass window as the 6 reference haloes, $10^{13} < M_{200c}/\mathrm{M}_\odot < 10^{14}$. We then calculate $c - \langle c\rangle$ over this entire second population. The difference in the statistics of these two ensembles captures the effect of the environment.

The results are shown in Fig. 4. The two distributions are non-Gaussian; in order to compare them quantitatively, we compute the 68 and 90 per cent credible intervals. The 68 per cent interval

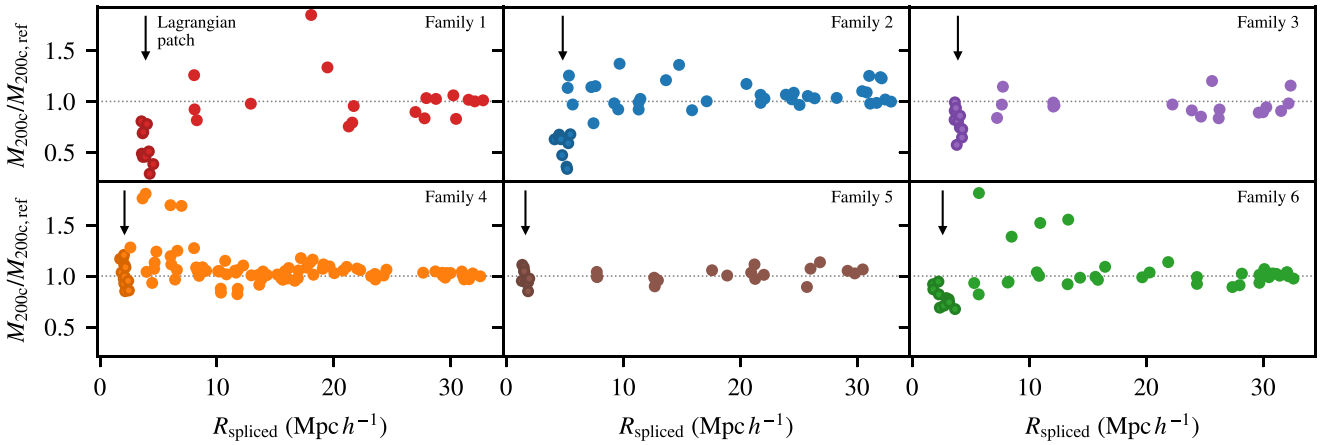

Figure 5. The ratio of the virial mass M_{200c} of the spliced haloes to the reference halo, for the six reference haloes. We highlight the simulations where the spliced region includes only the Lagrangian patch (darker symbols) and their mean R_{spliced} (black arrow); all other simulations use a splicing that has been expanded. The mass converges to the reference mass with increasing size of the spliced region at $z = 0$, R_{spliced}.

for the spliced distribution (shaded histogram), characterizing the impact of varying environment alone, is $[-1.0, 1.8]$. By contrast, the corresponding credible interval of the concentration of the entire population is $[-3.2, 2.7]$. When using 90 per cent credible intervals, the ranges expand to $[-2.1, 3.9]$ (spliced population) and $[-4.5, 4.0]$ (entire population). Therefore, between half and 70 per cent of the scatter in the concentration at fixed mass can be attributed to the effect of environment.

Having looked at the effect of splicing the Lagrangian patch of haloes into new environments, we now consider splicing larger regions. As discussed in Section 2, the size and shape of the spliced region can be chosen arbitrarily. Physically, one would expect that as the size of the spliced region expands, the influence of the external environment must become negligible because of the finite correlation length in Lambda cold dark matter (ΛCDM). Accordingly, we expect the variation between environments of any measured halo property to become small.

We performed an additional set of 211 simulations, using 3 outer realizations around the same 6 inner families, but expanding the spliced region to progressively include all matter within some distance from the Lagrangian patch. As the region is expanded, it becomes progressively more spherical. We quantify the size of the resulting spliced regions at $z = 0$ by an effective radius R_{spliced}, where

$$R^3_{\mathrm{spliced}} = \frac{3}{4\pi}\frac{M_{\mathrm{region}}}{\langle\rho\rangle_{\mathrm{v}}}, \tag{5}$$

where M_{region} is the total mass in the spliced region and $\langle\rho\rangle_{\mathrm{v}}$ is the volume-weighted mean density in the region at $z = 0$.

The results are shown in Fig. 5. Each panel uses simulations from one of our six families, showing how the final halo mass divided by the reference (unspliced) halo mass changes as the patch is expanded. Qualitatively, the halo mass converges towards the reference value as the splice radius becomes larger, as expected. This agrees with the work of Lucie-Smith et al. (2019), who found that the information relevant to determining the mass of haloes is localized within scales that are somewhat larger than their Lagrangian patches. However, we caution that a quantitative measure of the convergence radius using our method would require a considerably larger box size than the 50 Mpc h^{-1} used in this study.

4 DISCUSSION AND CONCLUSIONS

We have presented 'gene splicing', a method for resimulating a chosen halo within a variety of environments, while respecting the statistical properties of ΛCDM ICs. This is an extension of the 'GM' approach (Roth et al. 2016; Rey & Pontzen 2018), in which controlled experiments are carried out on a target halo, while the environment is minimally changed.

Manipulating Gaussian random fields in order to obtain insight into structure formation is an increasingly important tool (Aragon-Calvo 2016; Pontzen et al. 2016; Sawala et al. 2021). Because structure formation is localized, it is often desirable to make modifications in real space. This, however, requires a careful treatment to maintain consistency with ΛCDM correlation structure. Our approach to doing so follows in a long tradition of solving linear constrained systems in cosmological contexts (Bertschinger 1987; Hoffman & Ribak 1991; Bond, Kofman & Pogosyan 1996; van de Weygaert & Bertschinger 1996; Romano-Diaz et al. 2006; Elsner & Wandelt 2013).

As a first demonstration of the gene-splicing method, we showed that at least half the scatter in the mass–concentration relation can be attributed to the effect of the large-scale environment. This complements the results of Roth et al. (2016), where it was shown that the time of collapse (encapsulated by the local density field) is not able on its own to account for the scatter in this relation. We also showed that as the size of the spliced patch increases, the variation in mass decays towards zero in accordance with physical expectations. However, due to running a large number of simulations (274) we used a relatively small box of 50 Mpc h^{-1}. With the splicing approach, larger boxes would be needed to robustly measure the size of the region that contains information about halo collapse.

While the focus of this paper was on mass and concentration of dark matter haloes, many properties of haloes and galaxies are affected by their environment, and in future work, we will explore the underlying causal connections. For example, there is an observed correlation between galaxy quenched fraction and closeness to the nearest cosmological filament (Laigle et al. 2018; Kraljic et al. 2019), whose causal origin is as yet unclear (Romano-Díaz et al. 2017; Musso et al. 2018; Song et al. 2021). In future work, we intend to test these models by splicing a galaxy at different distances from a cosmic filament.

The gene-splicing method may also prove useful in the study of the secondary bias problem (Gao & White 2007; Dalal et al. 2008; Hahn et al. 2009). In particular, it enables direct tests of how the anisotropy in the environment affects the relationship between bias and concentration (Paranjape, Hahn & Sheth 2018). In this paper, we have applied the splicing operation to exactly fix the density field in a *finite* region of space. This, however, does not mean that other fields, such as the velocity field or the tidal shear, are fixed as well in the region. In order to fix the initial shear, we could apply the gene-splicing operation to the potential rather than the density field. This would still fix the density field in the spliced region but also the tidal shear, since both can be computed from local derivatives of the potential. However, since the potential is a smoother field (i.e. has longer range correlations) than the density, the gene-splicing operation will be less localized in this alternative formulation. Consequently, while the large-scale structure surrounding the spliced region is minimally affected when we splice overdensity (as in this work), it may be more strongly changed when splicing potential. This extension to splicing potential rather than density will be explored in future work.

ACKNOWLEDGEMENTS

CC thanks S. Codis and M. Musso for stimulating discussions. LLS thanks E. Komatsu for useful comments on the manuscript. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 818085 GMGalaxies. HVP's work was partially supported by the research project grant 'Understanding the Dynamic Universe', funded by the Knut and Alice Wallenberg Foundation under Dnr KAW 2018.0067. AP was supported by the Royal Society. This work used computing equipment funded by the Research Capital Investment Fund (RCIF) provided by UKRI, and partially funded by the UCL Cosmoparticle Initiative. The analysis was carried out using COLOSSUS (Diemer 2018), JUPYTER notebooks (Kluyver et al. 2016), MATPLOTLIB (Hunter 2007), NUMPY (Harris et al. 2020), PYNBODY (Pontzen et al. 2013), PYTHON, TANGOS (Pontzen & Tremmel 2018), and YT (Turk et al. 2011).

The main roles of the authors were, using the CRediT (Contribution Roles Taxonomy) system (https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/credit.html), as follows:

CC: conceptualization; methodology; validation; investigation; data curation; formal analysis; writing – original draft; visualization.

AP: conceptualization; methodology; software; validation and interpretation; writing – review and editing; funding acquisition.

HVP: conceptualization; validation and interpretation; writing – review and editing.

LLS: conceptualization; validation and interpretation; writing – review and editing.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Aragon-Calvo M. A., 2016, MNRAS, 455, 438
Aubert D., Pichon C., Colombi S., 2004, MNRAS, 352, 376
Bertschinger E., 1987, ApJ, 323, L103
Bond J. R., Kofman L., Pogosyan D., 1996, Nature, 380, 603
Borzyszkowski M., Porciani C., Romano-Díaz E., Garaldi E., 2017, MNRAS, 469, 594
Cadiou C., Pontzen A., Peiris H. V., 2021, MNRAS, 502, 5480
Chisari N. et al., 2015, MNRAS, 454, 2736
Codis S., Pichon C., Devriendt J., Slyz A., Pogosyan D., Dubois Y., Sousbie T., 2012, MNRAS, 427, 3320
Codis S., Pichon C., Pogosyan D., 2015, MNRAS, 452, 3369
Dalal N., White M., Bond J. R., Shirokov A., 2008, ApJ, 687, 12
Danovich M., Dekel A., Hahn O., Teyssier R., 2012, MNRAS, 422, 1732
Diemer B., 2018, ApJS, 239, 35
Elsner F., Wandelt B. D., 2013, A&A, 549, A111
Gao L., White S. D. M., 2007, MNRAS, 377, L5
Hahn O., Porciani C., Dekel A., Carollo C. M., 2009, MNRAS, 398, 1742
Harris C. R. et al., 2020, Nature, 585, 357
Hoffman Y., Ribak E., 1991, ApJ, 380, L5
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Kauffmann G., White S. D. M., Heckman T. M., Ménard B., Brinchmann J., Charlot S., Tremonti C., Brinkmann J., 2004, MNRAS, 353, 713
Kluyver T. et al., 2016, in Loizides F., Scmidt B., eds, Positioning and Power in Academic Publishing: Players, Agents and Agendas. IOS Press, Amsterdam, p. 87, available at https://eprints.soton.ac.uk/403913/
Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340
Kraljic K. et al., 2018, MNRAS, 474, 547
Kraljic K. et al., 2019, MNRAS, 483, 3227
Laigle C. et al., 2018, MNRAS, 474, 5437
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Lucie-Smith L., Peiris H. V., Pontzen A., 2019, MNRAS, 490, 331
Martizzi D., Vogelsberger M., Torrey P., Pillepich A., Hansen S. H., Marinacci F., Hernquist L., 2020, MNRAS, 491, 5747
Musso M., Cadiou C., Pichon C., Codis S., Kraljic K., Dubois Y., 2018, MNRAS, 476, 4877
Paranjape A., Hahn O., Sheth R. K., 2018, MNRAS, 476, 3631
Peng Y.-j. et al., 2010, ApJ, 721, 193
Planck Collaboration VI, 2020, A&A, 641, A6
Pontzen A., Tremmel M., 2018, ApJS, 237, 23
Pontzen A., Roškar R., Stinson G., Woods R., 2013, Astrophysics Source Code Library, record ascl:1305.002
Pontzen A., Slosar A., Roth N., Peiris H. V., 2016, Phys. Rev. D, 93, 103519
Pontzen A., Tremmel M., Roth N., Peiris H. V., Saintonge A., Volonteri M., Quinn T., Governato F., 2017, MNRAS, 465, 547
Pontzen A., Stopyra S., Cadiou C., Roth N., Rey M. P., 2021, Pynbody/genetIC: Version 1.3, available at https://doi.org/10.5281/zenodo.5079937
Press W. H., Schechter P., 1974, ApJ, 187, 425
Rey M. P., Pontzen A., 2018, MNRAS, 474, 45
Romano-Diaz E., Faltenbacher A., Jones D., Heller C., Hoffman Y., Shlosman I., 2006, ApJ, 637, L93
Romano-Díaz E., Garaldi E., Borzyszkowski M., Porciani C., 2017, MNRAS, 469, 1809
Roth N., Pontzen A., Peiris H. V., 2016, MNRAS, 455, 974
Sawala T., Jenkins A., McAlpine S., Jasche J., Lavaux G., Johansson P. H., Frenk C. S., 2021, MNRAS, 501, 4759
Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1
Song H. et al., 2021, MNRAS, 501, 4635
Stopyra S., Pontzen A., Peiris H., Roth N., Rey M. P., 2021, ApJS, 252, 28
Tempel E., Stoica R. S., Saar E., 2013, MNRAS, 428, 1827
Teyssier R., 2002, A&A, 385, 337
Turk M. J., Smith B. D., Oishi J. S., Skory S., Skillman S. W., Abel T., Norman M. L., 2011, ApJS, 192, 9
Tweed D., Devriendt J., Blaizot J., Colombi S., Slyz A., 2009, A&A, 506, 647
van de Weygaert R., Bertschinger E., 1996, MNRAS, 281, 84

APPENDIX A: DERIVATION OF THE SPLICING TECHNIQUE

In this appendix, we present an outline of the technique by which a given initial density field can be 'spliced' into another while remaining consistent with a ΛCDM power spectrum.

We assume the ICs to be in the linear regime so that we can write any density field as

$$\rho(\boldsymbol{x}) = \rho_0(1 + \delta(\boldsymbol{x})), \tag{A1}$$

where $\delta(\boldsymbol{x})$ is a Gaussian random field. The problem of generating ICs in the case of numerical simulations involves a list of N discretized cells located at position $\boldsymbol{x}_i$. In order to draw ICs, we need to draw N values from a multivariate normal distribution with N dimensions with mean value $\boldsymbol{\mu}$ and covariance C. For ΛCDM ICs, $\boldsymbol{\mu} = 0$, while the effect of C on any vector can be computed by performing a discrete Fourier transform, multiplying by the power spectrum, and transforming back to real space. The effect of C^{-1} is obtained through a similar sequence, dividing by the power spectrum instead of multiplying. Note that the splicing method is not limited to the density field and can be applied to any Gaussian random field, such as the gravitational potential field.

Extending the description of Section 2, let us write the discretized independent random fields as $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively. We also introduce a mask matrix M that zeros pixels lying outside the selected region Γ, while leaving those inside the region untouched. We will also use the shorthand $\overline{\mathsf{M}} \equiv \mathbb{I} - \mathsf{M}$; functionally, $\overline{\mathsf{M}}$ zeros all pixels inside the mask, while retaining the value of those outside.

A spliced field $\boldsymbol{f}$ satisfies the defining relation:

$$\begin{aligned} &\text{minimize } Q = (\boldsymbol{b} - \boldsymbol{f})^\dagger \mathsf{C}^{-1}(\boldsymbol{b} - \boldsymbol{f}), \\ &\text{subject to } \mathsf{M}\boldsymbol{f} = \mathsf{M}\boldsymbol{a}. \end{aligned} \tag{A2}$$

To solve this constrained quadratic minimization, we split the problem into two systems – one of which is fully constrained, while the other is entirely unconstrained. Specifically, we write

$$\boldsymbol{\Delta} = \boldsymbol{f} - \boldsymbol{b}. \tag{A3}$$

With this definition, the constraint becomes $\mathsf{M}\boldsymbol{\Delta} = \mathsf{M}(\boldsymbol{a} - \boldsymbol{b})$. The solution for $\boldsymbol{\Delta}$ in the system (A2) then takes the form

$$\boldsymbol{\Delta} = \mathsf{M}(\boldsymbol{a} - \boldsymbol{b}) + \overline{\mathsf{M}}\boldsymbol{\alpha}. \tag{A4}$$

The constraint is satisfied by construction, while the vector $\boldsymbol{\alpha}$ is defined implicitly by the minimization of Q. Back substituting our definitions into Q, we have

$$Q = \boldsymbol{\alpha}^\dagger \overline{\mathsf{M}}\mathsf{C}^{-1}\overline{\mathsf{M}}\boldsymbol{\alpha} + \boldsymbol{\alpha}^\dagger \overline{\mathsf{M}}\mathsf{C}^{-1}\mathsf{M}(\boldsymbol{a} - \boldsymbol{b}) + \text{c.c.} + \text{const.} \tag{A5}$$

Here c.c. indicates the complex conjugate of the preceding term, while the constant term does not depend on the quantity we are now optimizing, $\boldsymbol{\alpha}$, and therefore does not need explicit calculation. Here, we have used the fact that all the matrices C, M, and $\overline{\mathsf{M}}$ are Hermitian.

Minimizing Q (now without any constraints) requires

$$\overline{\mathsf{M}}\mathsf{C}^{-1}\overline{\mathsf{M}}\boldsymbol{\alpha} = \overline{\mathsf{M}}\mathsf{C}^{-1}\mathsf{M}(\boldsymbol{b} - \boldsymbol{a}). \tag{A6}$$

This is an equation of the form $\mathsf{A}\boldsymbol{\alpha} = \boldsymbol{z}$ with $\mathsf{A} = \overline{\mathsf{M}}\mathsf{C}^{-1}\overline{\mathsf{M}}$ and $\boldsymbol{z} = \overline{\mathsf{M}}\mathsf{C}^{-1}\mathsf{M}(\boldsymbol{b} - \boldsymbol{a})$, which can be solved by standard conjugate gradient optimization methods. At each step of the optimization, a discrete Fourier transform and its inverse will be computed in order to multiply by C^{-1}.

Other than its null space, A has a similar spectrum to C^{-1}, and therefore we improve the convergence rate of the conjugate gradient method by pre-conditioning the problem with the matrix C. Once $\boldsymbol{\alpha}$ is obtained, we can back substitute it into equations (A3) and (A4) to find $\boldsymbol{f}$. The ICs are generated by finally computing particle displacements from the density field.

In our implementation, we stop the conjugate gradient iterations once the Euclidian norm of the residuals becomes smaller than $\|\boldsymbol{z}\|/10^6$. The convergence is typically achieved in a few steps in 1D, and a few hundred steps in 3D depending on the size and shape of the spliced region and the size of the grid it is spliced into. We find that the time to solution scales roughly as N^4, where N is the number of cells in each direction. For the simulations in this work, we obtained solutions to the splicing minimization within 4 min on 32 cores; given the scaling is barely worse than for a single FFT that scales as $N^3 \ln N$, the method can be applied to much larger simulations.

This paper has been typeset from a TEX/LATEX file prepared by the author.