DE-FG02-09ER25877

DOE Award Number DE-SC0002097

PRINCETON UNIVERSITY, Princeton, NJ

Prof. Ioannis G. Kevrekidis, PI (also Prof. C. W. Gear and Prof. A. Singer)

(Prof. R. R. Coifman, Yale University)

EQUATION-FREE AND VARIABLE FREE MODELING FOR COMPLEX/MULTISCALE SYSTEMS:

Coarse-grained computation in science and engineering using fine-grained models

February 2017

FINAL REPORT

for Dr. Sandy Landsberg

attn. Ms. Teresa Beachley

Office of Advanced Scientific Computing Research

Office of Science

US DOE

PROGRESS REPORT

The best way to start this report is to repeat the *PROJECT OBJECTIVE* in the original proposal:

We propose to integrate two recent developments in the data-driven modeling of complex/multiscale systems in a seamless computational framework. We will link the Kevrekidis group computational research on Equation-Free modeling of complex systems with the Coifman group theoretical and computational work on non-linear feature extraction from complex systems data through (diffusion) operators on graphs. We will validate the new algorithmic toolkit by extensively exploring its mathematical underpinnings, benefits and limitations.

This has been our goal and the backbone of our efforts for the entire period of the work. There have been several different directions, form algorithm development to the exploration of potential applications. At the end of the report I have added a list of publications (archival journal as well as a few refereed proceedings) from the work. I will start by discussing those –more or less completed- efforts, and I will proceed to discuss some items completed/appeared after the grant finishehd. As I present the work I will also mention the graduate students, postdoctoral fellows, and the US and international collaborators that have been involved in the various projects.

Algorithm Development

A paper most representative of our effort is [7] that appeared in the journal *Processes.* It contains our first effort to link manifold learning techniques (and, in particular, diffusion maps) with solving dissipative PDEs. It is representative both in terms of our ultimate goal and because of the authors involved. Prof. Gear is a world class numerical analyst and is also partially supported through this program; and finally, the grad student advisor is the PI of the grant. Instead of using traditional (e.g. Fourier or finite element) basis functions, or even semi-empirical ones based on principal component analysis of simulation data (PODs), we collect simulation data and then parametrize the low-dimensional, nonlinear manifold on which the long term dynamics of the PDE lie through Diffusion Maps. The paper then discusses how to formulate the low-dimensional evolution equations on this manifold.

Ben Sonday, a DOE-CSGF Fellow that obtained his PhD with research supported through this grant was the first author of another paper involving a collaboration with Prof. A. Singer: [8] that appeared in *CAMWA.* Here the issue is the simulation of systems with symmetry, and the important question is the reduction of the data by factoring out of these symmetries especially in the presence of noise. This is especially important in the context of multiscale simulations when the symmetry is only present at the macroscopic level (e.g. a particle density is translationally invariant); the fine-scale, particle based simulation is a natural source of noise in estimating particle density, and the tools developed by Prof. Singer (the `"eigenvector method" in the paper as well as Vector Diffusion Maps) provide a systematic way to perform this reduction. It is also worth mentioning an interesting little paper ([5]) about initializing equation-free computations, in which Ben Sonday again played an important role, and which involved an international collaboration with K.U.Leuven and a long-term collaborator from Russia (A.Makeev).

One of the first projects we pursued with Prof. Gear has appeared in the publication [2] – this short paper explores systems whose solutions do not lie on a smooth manifold, but are rather slowly evolving

distributions. In this paper we develop the idea of a *Virtual Slow Manifold* that parametrizes these distributions, and that can be usefully exploited for numerical computations such as coarse projective integration. A project currently being pursued with Prof. Gear (and postdoctoral fellow Arta Jamshidi, a mathematics PhD) involves "mesh free" equation-free computations based on the use of Radial Basis Functions (RBFs). When the macroscopic fields we evolve are observed in terms of a finite number of RBFs, the parameters of the RBFs (width, amplitude, center location) become the quantities to project forward in time in coarse projective integration. To avoid the ubiquitous problem of discontinuities in a snapshot-by-snapshot fits using RBFs, we propose and develop a class of Linearly Time-Varying RBFs tailor-made for (coarse) projective integration in such systems. This is a paper that has also appeared.

In the same algorithm-development direction there are efforts that involve several international collaborators: with Prof. Tony Roberts from Adelaide we have been working towards equation-free computation in stochastic problems (the Proceedings [15]) and in a joint collaboration with Prof. Jinqiao Duan of IIT Chicago for the computation of stochastic approximate inertial manifolds. The latter paper has also now appeared. With Dr. Giovanni Samaey of KU Leuven in Belgium, Prof. Gear and graduate student Ping Liu we have been working towards (a) a patch dynamics scheme for coarse-grained computations with an agent-based model and (b) for the same agent-based model, in a particular parameter regime, the use of diffusion maps for the coarse computational study of rare events. Another effort with graduate student K. Rajendran has as its goal the coarse-graining of problems where the evolving objects are graphs – and the dynamics are either dynamics *of* graphs or dynamics of quantities (e.g. real numbers representing opinion propagation} *on* a network. A first manuscript from this effort has appeared [9]; it does use diffusion map related techniques to devise appropriate coarse-grained variables that can accurately describe evolution of functions on networks.

There were three important directions ongoing in the work of Mr. (now Dr.) Rajendran. One is in collaboration with Prof. Singer, and has to do with the use of graph kernels to obtain in a systematic way good observables for coarse-graining evolving networks. The second introduces a (probably new) connection between equation-free computation and tools for uncertainty quantification (like polynomial chaos expansions). The main idea is that *heterogeneity* in networks can be treated as one or more uncertain parameters, and polynomial chaos representations may provide good coarse observables for coarse-graining the dynamics of networks of heterogeneous nodes. The third is a new direction for us that arose after the grant started and involves a collaboration with Prof. C. Floudas in my Department (the paper [19] has already electronically appeared from this work). One of the most important steps in equation-free computation is *lifting*: how to construct fine scale realizations consistent with given macroscopic observables. This work studied algorithms (based on mathematical programming, and, in particular, MILP) in order to do this when the system whose dynamics we are coarse-graining is a network. The work along these lines with Dr. Balazs Rath has also led to a publication.

Our Yale collaborators have also been working on algorithm development. There has been an ongoing effort with Dr. Haddad and Prof. Coifman which included Mr. Sonday and will most probably continue with Mr. Rajendran and Prof. Gear, about "out-of-sample" extension using diffusion maps. It turns out that this step is particularly important in equation-free computation, where new "interesting" points for data collection are suggested by our algorithms in diffusion map space, yet the simulations have to be initialized in physical space (the "lifting" problem again). During this period of our work Prof. Coifman and his group have addressed two fundamental challenges in our proposal, the first of which relates to the

extraction of efficient intrinsic parameters describing complex systems. This effort had as goal to extend the methodologies of prior work on a nonlinear version of principal component analysis, and inverse parameter extraction [17] to enable an automatic insertion of new data into the corresponding geometry. In this paper by Coifman and Haddad a general nonlinear projector is introduced, which automatically relates any data to existing observed or simulated structures. This result was an essential task that needed to be resolved in order to enable the goal of equation-free simulations. We have been finally able, in paper [7] to accomplish this important eextsion task in our work.

The second challenge that the Yale part of the team faced involved the development of both theory and algorithms for adaptive coarse graining through analytic adaptive sampling grids [18]. Since typically data generated in simulations is high dimensional and has unknown or unmodeled internal structure, it is necessary to enable a flexible subsampling strategy that permits synthesis of band limited, or other classes of functions on the data to be extrapolated with controlled accuracy. In particular we may require that for each band the grid be efficient or optimal for that band. We provide an analytical multiscale coarse graining which is dimensionally independent. Our construction provides the sampling point of view of diffusion geometry embeddings and parametric intrinsic coordinates described above. The points that we identify in the original data are directly related to points in the nonlinear diffusion map coordinate systems which are well sampled there. We are now comparing the merits of both strategies. It should be mentioned in passing, that the multiscale analytic grid strategy described in the relevant paper by Averbuch, Bermanis and Coifman [18] is valuable even in the setting of some boundary value problems in low dimensions as a way of discretizing the boundary without ever needing to find a parameterization for it. In the images below we sketch the multiscale sampling strategy where each grid is designed to extrapolate with Gaussians at the prescribed scale (or with band limited functions with the corresponding band). We also note that the whole construction is at optimal efficiency, as the SVD performed to compute the sparse grids are local in nature and are computed using a number of random codes needed locally in the data.

Two additional algorithm-development efforts should be mentioned: one that involved a constrained approach to multiscale Gillespie-type simulations [10], in collaboration with Dr. Erban in Oxford and two of his postdocs; and one that involves the linking of diffusion maps with established statistical mechanical algorithms (and, in particular, umbrella sampling and histogram reweighting) [6]. The latter was the result of serendipity: an excellent graduate student at Princeton working with P. Debenedetti and A. Panagiotopoulos, on whose Thesis committee I served, decided to follow up some suggestions I had on his work – he took a course by A. Singer, brought me and his advisors together, in effect changed the direction of his Thesis work and now as a postdoc at MIT and future faculty member at Illinois, Dr. A. Ferguson lists nonlinear data reduction as his main research field. In addition to the paper [6] there was a PNAS from this collaboration [4] and an invited *Chem. Phys. Lett.* Frontiers article [16].

The most important challenge for algorithm development in the last (third) year is the clear understanding of the issues and the algorithms (and the error control) associated with *extension* and *lifting* from diffusion map space to physical space. This is a vital challenge, since there is no direct formula for this process, as opposed to, say, Principal Component Analysis, where the lifting is straightforward. If we are successful in this, there is a host of applications (with those in statistical mechanics and exploration of effective free energy surfaces / rare events being the most obvious ones) where our methodologies have the potential for breakthroughs. Tailoring both our equation-free approach and data-mining approaches to the

evolution of (and on) networks is a new and truly interesting direction and we were finally able to accomplish this in the *Processes* paper.

Applications

In addition to algorithm development we have also been pursuing a number of new applications in which we plan to apply the combination of equation free/ variable free numerics. These include Monte Carlo self-assembly simulations [11,12,14], neuronal network complex dynamics [13], molecular dynamics simulations for penta-alanine (postdoctoral fellow M. Amat and graduate student L. Nedialkova), as well as Monte Carlo dynamics of driven interfaces (work completed by B. Sonday, [1]), Monte Carlo simulations of particle coagulation and sintering [3], molecular dynamics simulations of cluster formation (graduate student and now Dr. Carmeline Dsilva, who also is a DOE CSGF Fellow!). We also pursued applications where the fine scale simulator is an agent-based model (graduate student Ping Liu), evolving network with heterogeneous nodes (graduate student, now Dr. K. Rajendran, collaborator Prof. C. Siettos) and a mesoscopic Lattice-Boltzmann simulator of multiphase flow (postdoctoral fellow C. Colosqui, collaborator Prof. A. Papathanasiou).

Before concluding with the publications acknowledging the grant I would like to add that there have been many contributed, invited and plenary presentations from the work these three years –including a plenary at 2011 SIAM Control Conference in July in Baltmore. The PI was honored through (a) selection as a SIAM Fellow in 2010; (b) the R. Wilhelm award of the American Institute of Chemical Engineers in 2010 and (c) the selection as Gutzwiller Fellow at the Max Planck Institute for Complex Systems in Dresden, Germany for 2010-2011. I am also happy to report that one of our international collaborators (Prof. C. Laing from New Zealand) came on sabbatical in Princeton and that another collaborator (Prof. C. Siettos from NTUA in Athens) came for an extended visit, supported in part through a Fulbright Fellowship.

I also add three representative figures, two from Princeton and one from Yale, from our work in the next page.

It is important to note several papers that appeared after the end of the grant, but which embodied many of the main items we were pursuing, and are indicative of the success of the effort. The first had to do with the ability to do gauge invariant data mining – that is find ways to merge different observations of the same system, and thus create nonlinear observers of hidden states. This is something that we will keep actively pursuing in our research in the future. The second item is the ability to reduce complex multiscale stochastic problems by discovering what the true slow variables are – and how to parsimoniously parametrize the resulting effective models. These papers appeared after the end of the grant, but the seeds for them were laid during the grant period.

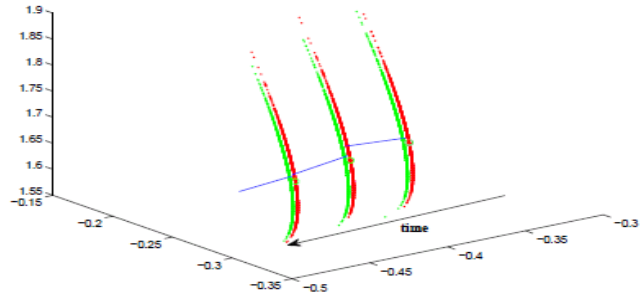
Overall, we have been, I feel, quite successful in achieving the originally stated goal – to link equation free computations with variable free data mining – in effect, find a new and important link between machine learning and scientific computation / modeling.

The W. T. and Idalia Reid Prize in Mathematics that was awarded to the PI Prof. Kevrekidis, in the annual SIAM meeting in 2016 is precisely (as the award citation mentions) for our equation-free work that has been supported under this grant – and for this support we are grateful.

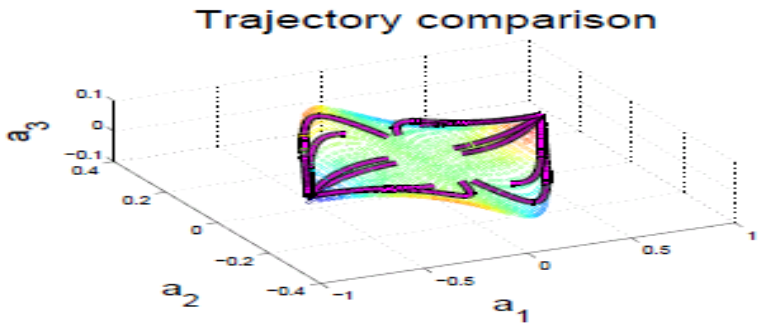
UNEXPENDED FUNDS There were be no unexpended funds (~ $2,000) by the end of the grant.

Three Representative Figures from the Research.

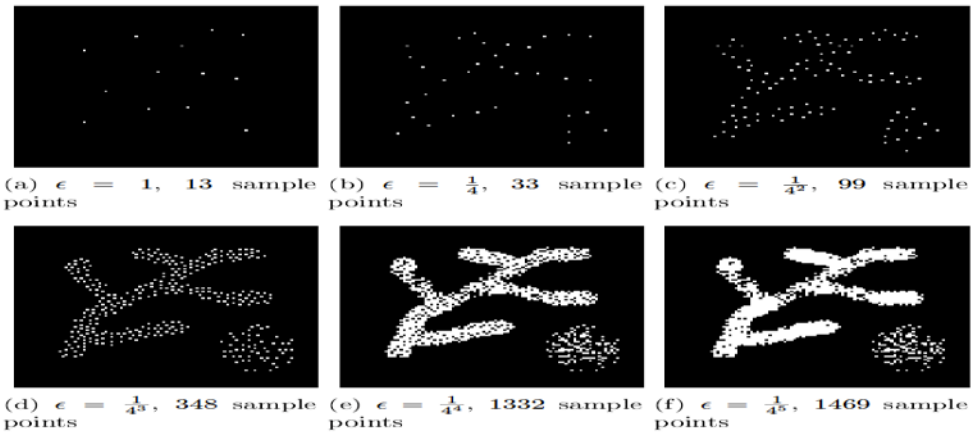
TOP: (from [2]) Coarse Projective Integration for systems where the slow evolving quantities are measures (clouds of points in the figure). The red and the green denote the "initial" and the "final" cloud used to construct each projective step (marked by the light blue linear segments).



MIDDLE: (from [7]) Visualization of trajectories on an Approximate Inertial Manifold (for the Chafee-Infante RD equation) in the space of the first three Fourier Components of the solution. The boldly colored trajectories on the manifold have been obtained from dynamics formulated in diffusion map coordinates parametrizing the manifold.



BOTTOM: (from [18]) An illustration of adaptive subsampling of data – a sequence of subsamples (referred to as adaptive grids) as well as a sequence of approximations to a given empirical function on the data are generated (along with its extension on newly arrived data points).



(a) $\epsilon = 1$, 13 sample points (b) $\epsilon = \frac{1}{4}$, 33 sample points (c) $\epsilon = \frac{1}{4^2}$, 99 sample points

(d) $\epsilon = \frac{1}{4^3}$, 348 sample points (e) $\epsilon = \frac{1}{4^4}$, 1332 sample points (f) $\epsilon = \frac{1}{4^5}$, 1469 sample points

PUBLICATIONS acknowledging DOE support from this grant

1. Coarse-graining the dynamics of a driven interface in the presence of mobile impurities: effective description via diffusion maps, B. E.Sonday, M. Haataja and I. G. Kevrekidis, *Phys. Rev. E* **80** 031102 (2009)
2. Computing on virtual slow manifolds of fast stochastic systems, C. W. Gear, D. Givon and I. G. Kevrekidis, *JNAIAM* **5**(1-2) pp.61-72 (2010) pp.13597-13602 (2010)
3. Coarse-grained computation for particle sintering and coagulation processes by linking Quadrature Method of Moments with Monte Carlo, Y.Zou, M. E. Kavousanakis, I. G. Kevrekidis and R. O. Fox, *J. Comp. Phys.* **229** pp.5299-5314 (2010)
4. Systematic determination of order parameters for chain dynamics using diffusion maps, A. L. Ferguson, A. Z. Panagiotopoulos, P. G. Debenedetti and I. G. Kevrekidis, *Proc. Natl. Acad. Sci. USA* **107**(31) pp. 13597-13602 (2010)
5. A common approach to the computation of coarse-scale steady states and to consistent initialization on a slow manifold, C. Vandekerckhove, B. Sonday, A. Makeev, D. Roose and I. G. Kevrekidis, *Comp. Chem. Eng. **35**(10) pp.1949-1958 (2011)*
6. Integrating diffusion maps with umbrella sampling: application to alanine dipeptide. A. L. Ferguson, A. Z. Panagiotopoulos, P. G. Debenedetti and I. G. Kevrekidis, *J. Chem. Phys.* **134** 135103 (2011)
7. E. Chiavazzo, C. W. Gear, C. J. Dsilva, N. Rabin and I. G. Kevrekidis, Reduced Models in Chemical Kinetics via nonlinear data-mining, *Processes* **2**(1) pp.112-140 (2014)
8. Noisy dynamic simulations in the presence of symmetry: data alignment and model reduction, B. Sonday, A. Singer and I. G. Kevrekidis, submitted to *Physica D*, April 2010
9. Coarse graining the dynamics of heterogeneous oscillators in networks with spectral gaps, K. Rajendran and I. G. Kevrekidis, submitted to *Phys. Rev. E.,* April 2010
10. A constrained approach to multiscale stochastic simulation of chemically reacting systems, S. L. Cotter, K. C. Zygalakis, I. G. Kevrekidis and R. Erban, to *J. Chem. Phys.* April 2011
11. Lattice Monte-Carlo simulation of semiconductor nanocrystal synthesis in microemulsion droplets, S. R.Kuriyedtath, B. Kostova, I. G. Kevrekidis and T. J.Mountziaris, *Langmuir* **26**(13) pp.11355-11362 (2010).
12. Lattice Monte-Carlo simulation of cluster coalescence kinetics with application to template-assised synthesis of quantum dots, R. Kuriyedtath, B. Kostova, I. G. Kevrekidis and T. J.Mountziaris, *Ind. Eng. Chem. Res.* **49** pp.10442-10449 (2010)
13. Bifurcations of lurching waves in a thalamic neuronal network, T. M. Wasylenko, J. E. Cisternas, C. Laing and I. G. Kevrekidis, *Biol. Cybern.* **103** pp.447-462 (2010)
14. A comparison of implicit- and explicit-solven simulations of self-assembly in block-copolymer and solute systems, J. R. Spaeth, I. G. Kevrekidis, and A. Z. Panagiotopoulos, *J. Chem. Phys.* **134** 164902 (2011)
15. X. Kan, J. Duan, I. G. Kevrekidis and A. J. Roberts, "Simulating Stochastic Inertial Manifolds by a Backward-Forward Approach", *SIADS* **12**(1) pp.487-514 (2013).

16. Y. Xue, P. J. Ludovice, M. A. Grover, L. V. Nedialkova, C. J. Dsilva and I. G. Kevrekidis, State reduction in molecular simulations, *Comp. Chem. Eng.* **51** pp.102-110 (2013)
17. Anisotropic diffusion on sub-manifolds with application to earth structure classification , D. Kushnir, A. Haddad and R.R. Coifman, submitted to *Appl. Comp. Harm. Anal.* 2011
18. Multiscale data sampling and function extension , A. Bermanis, A. Averbuch and R.R. Coifman , submitted to the Proceedins of SampTA (Sampling Theory and Applications 2011)
19. Generation of networks with prescribed degree-dependent clustering, C. Gounaris, K. Rajendran, I. G. Kevrekidis and C. A. Floudas, *Optim. Lett.* **5**(3) SI pp.435-451 (2011)
20. Efficient Coarse Simulation of a Growing Tumor at its Avascular Stage, M. E. Kavousanakis, P. Liu, A. G. Boudouvis J. Lowengrub and I. G. Kevrekidis, *Phys. Rev. E* **85** 031912 (2012)
21. C. Marschler, J. Starke, P. Liu and I. G. Kevrekidis, Coarse-graining a particle model for pedestrian flow using diffusion maps, *Phys Rev E* **89** 013304 (2014)
22. P. Liu, H. R. Safford, I. D. Couzin and I. G. Kevrekidis, Caorse-grained variables for particle-based models: diffusion maps and animal swarming simulations, J. Comp. Part. Mech. 1:425-440 (2014)
23. C. J. Dsilva, R. Talmon, N. Rabin, R. R. Coifman and I. G. Kevrekidis, Nonlinear Intrinsic Variables and State Reconstruction in Multiscale Simulations, J. Chem. Phys. 139 184109 (2013)
24. C. Dsilva, R. Talmon, R. R. Coifman and I. G. Kevrekidis, Parsimonious Representation of Nonlinear Dynamical Systems through Manifold Learning: a Chemotaxis Case Study, Appl. Comp. Harm. Anal. 13 July 2015
25. Dsilva, Carmeline J.; Talmon, Ronen; Gear, C. William; et al. Data-Driven Reduction for a Class of Multiscale Fast-Slow Stochastic Dynamical Systems, SIAM JOURNAL ON APPLIED DYNAMICAL SYSTEMS 15(3) pp.1327-1351 (2016)
26. L. V. Nedialkova, M. Amat, I. G. Kevrekidis and G. Hummer, Diffusion Maps, Clustering and Fuzzy Markov Modeling in Peptide Folding Transitions, J. Chem. Phys. 141 114102 (2014)
27. An equation-free approach to coarse-graining the dynamics of networks, K. A. Bold, K. Rajendran, B. Rath and I. G. Kevrekidis, J. Comp. Dyn 1(1) (2014)
28. Droplet and slug formation in polymer electrolyte membrane fuel cell flow channels" the role of interfacial forces. C. E. Colosqui, M. J. Cheah, I. G. Kevrekidis and J. B. Benziger, J. Power Sources 196(23) pp.10057-10068 (2011)
29. Mechanisms of wetting transitions in patterned surfaces: continuum and mesoscopic analysis, M. E. Kavousanakis, C. Colosqui, I. G. Kevrekidis and A G. Papathanasiou Soft Matter, 8 3928-3936 (2012)
30. C. E. Colosqui, M. E. Kavousanakis, A. G. Papathanasiou and I. G. Kevrekidis, "Mesoscopic models for microscale hydrodynamics and interfacial phenomena: slip, films and contact angle hysteresis", Phys. Rev. E 87 013302 (2013)

31. A. A. Jamshidi, C. W. Gear and I. G. Kevrekidis, Time varying radial basis functions, J. Comp. App. Math. 266 pp.61-72 (2104)
32. Stability and stabilization of the constrained runs schemes for equation-free projection to a slow manifold, A. Zagaris, C. Vandekerckhove, C. W. Gear, T. J. Kaper and I. G. Kevrekidis, Disc. Cont. Dyn. Syst. 32(8) pp.2759-2803 (2012)

It is gratifying to write that, in part as a result of this work, the PI has garnered the following distinctions over the last few years:

- Einstein Foundation and Zuse Institute, Distinguished Visiting Fellowship (2016-2018)
- Isaac Newton Institute, Cambridge, Rothschild Visiting Distinguished Fellowship (2016)
- Institute for Advanced Study, T. U. Muenchen and European Union – Hans Fischer Senior Fellow, 2015-2017
- American Institute of Chemical Engineers, Fellow, Elected 2016
- Academy of Athens, Corresponding Member, Elected 2015
- Institute for Advanced Study, Technical University of Munich, and EEC: Hans Fischer Senior Fellowship (2015-2018)
- Isaac Newton Institute, Cambridge, Microsoft Visitorship (2013)
- Max Planck Society, MPIPKS, Dresden, Martin Gutzwiller Fellowship (2010-2011)