LLNL-POST-xxxxxx This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

A SPECTRAL APPROACH FOR THE DESIGN OF EXPERIMENTS IN HIGH DIMENSIONS

J.J. Thiagarajan (LLNL), B. Kailkhura (LLNL), P.-T. Bremer (LLNL)

Computational Sciences increasingly depend on exploring high-dimensional parameter spaces. In this project, we develop new sampling techniques that achieve provably better space-filling properties, thus improving subsequent analysis significantly. For the first time, we obtain theoretical bounds on achievable performance for a given sample budget in high dimensions. Empirical studies on surrogate modeling and hyper-parameter search for deep neural networks show that our method saves up to 50% on sample budgets, when compared to existing approaches.

Improved Sample Quality is Critical to the Success of Data Analysis Pipelines

Save millions of CPU hours while providing more reliable and more informative results – Directly impacts all mission-critical areas

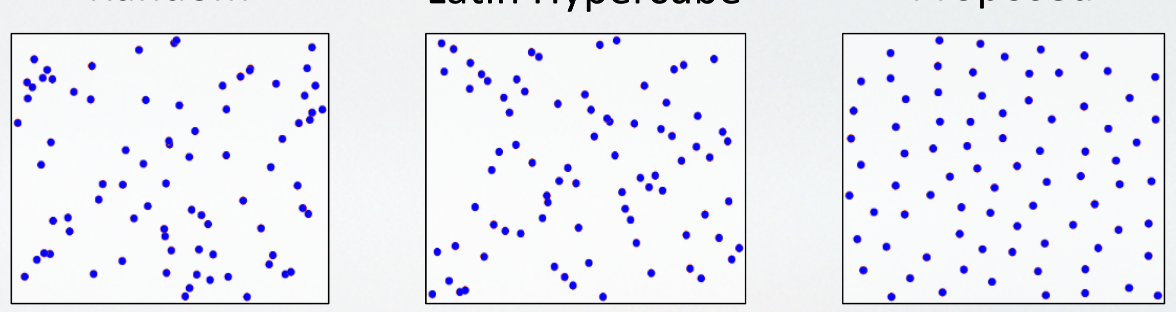
Our sampling methods are actively adopted in **several ongoing efforts** – Cognitive simulations LDRD, material informatics LDRD, performance optimization, collaborative autonomy, cancer modeling (pilot 2)

Provably effective for fundamental problems in data analysis that are applicable across the missions – Surrogate modeling, Monte-Carlo integration and Auto ML

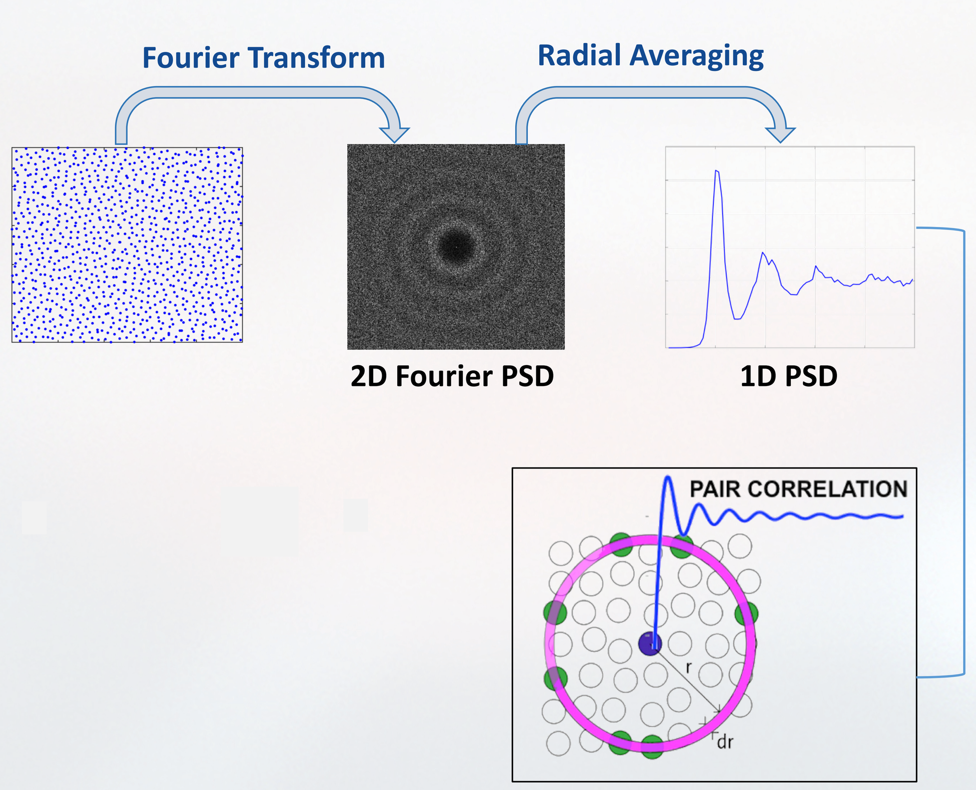
Characterizing Sample Quality

Existing metrics for evaluating space-filling designs are grossly insufficient.

Random | Latin Hypercube | Proposed



Methods that optimize only for the spatial properties result in poor coverage at low sampling rates and are plagued by variance in the solution.



Key Finding:
A direct connection exists between 1D *Power Spectral Density* and *Pair Correlation Function* in any dimension

Implication:
Sample designs that jointly optimize spatial and spectral properties

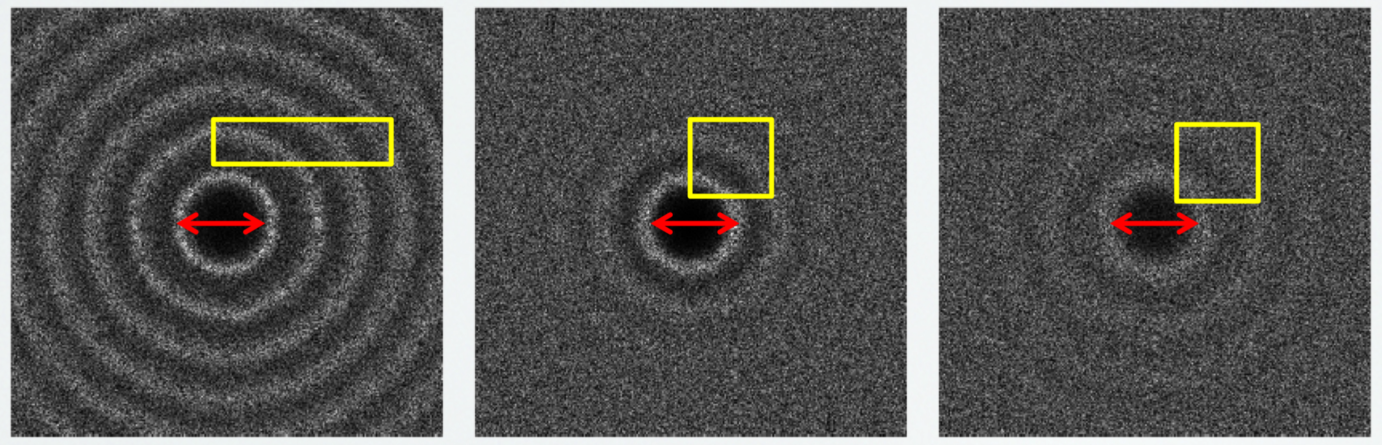
A new way to quantify the space-filling property, which enables us to systematically trade-off uniformity and randomness,

Going Beyond Spatial Heuristics

What makes a good spectral sample?

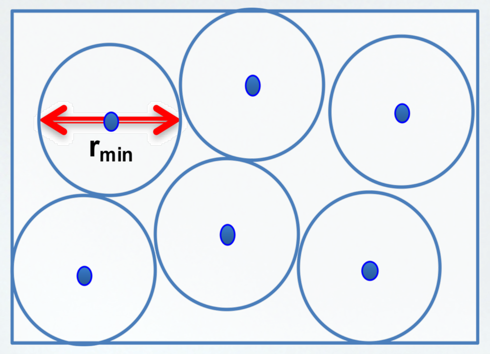
- Large *Zero Region* – Alias-free low-frequency region
- Lesser *Oscillations* – Reduced aliasing artifacts

Farthest Point | CCVT | Dart Throwing



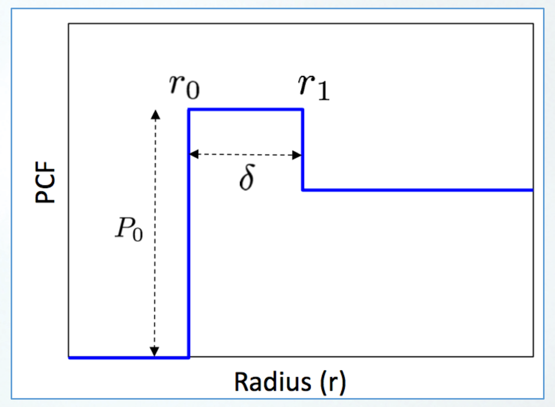
Spectral properties can be used to compare different distributions, though only their spatial characteristics are optimized by existing sampling techniques.

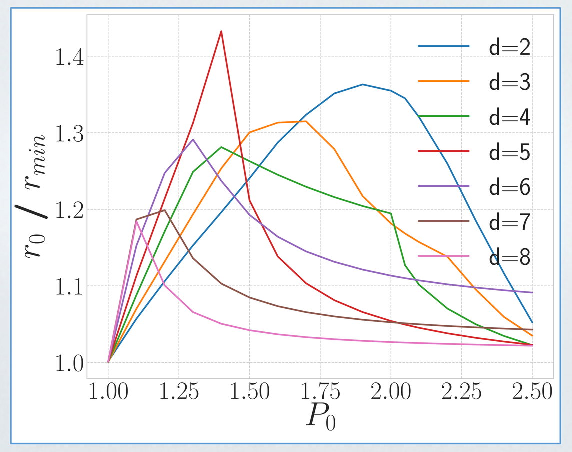
Space-Filling Spectral Designs



"Photo-receptors in the retina of the monkey eye are distributed according to a Poisson disk distribution"

PCF of the sample design is characterized by a stair – Except for a single peak, does not compromise uniformity

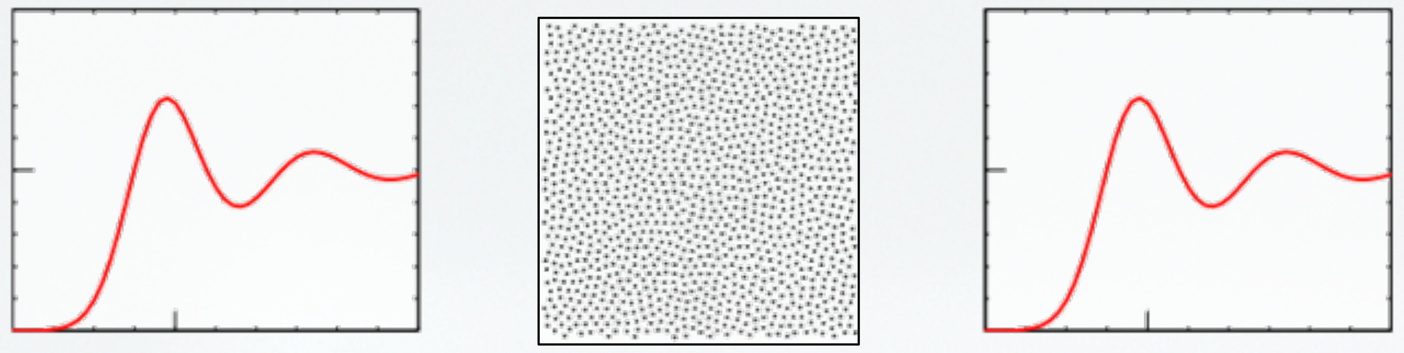




Surprisingly, samples produced by all existing approaches are much larger than the theoretically required sample size

Efficient Algorithm for Sample Synthesis

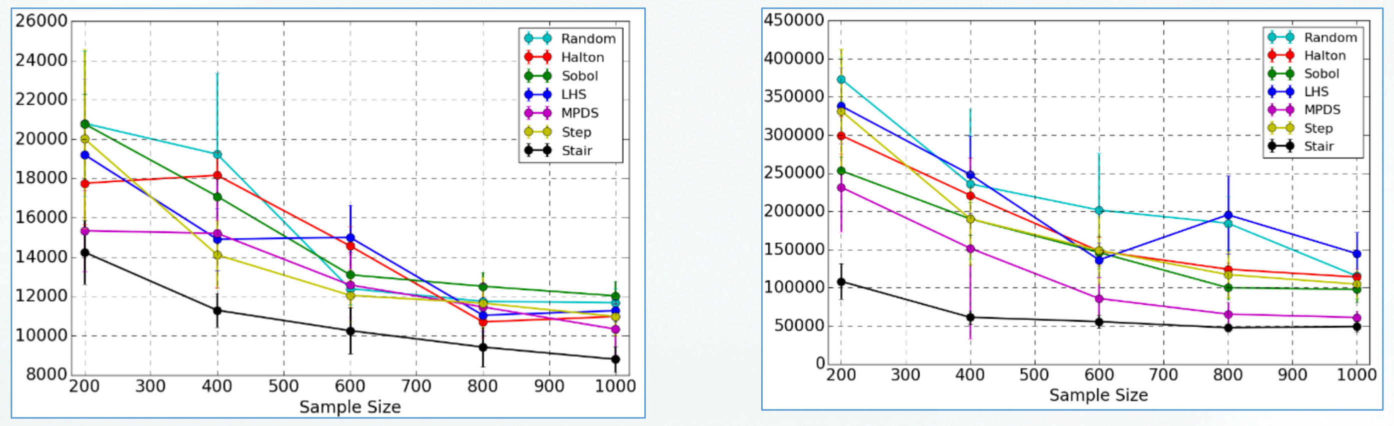
Desired PCF | Synthesis | Synthesized PCF



Our algorithm uses gradient descent optimization to match the Pair Correlation functions of the synthesized and desired distributions.

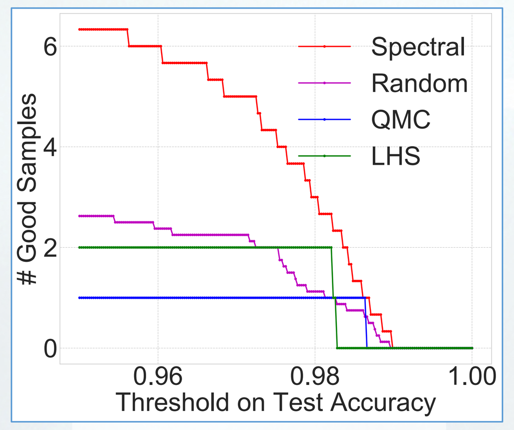
Performance Evaluation

Case Study: Surrogate modeling in moderately high dimensions (< 10D)

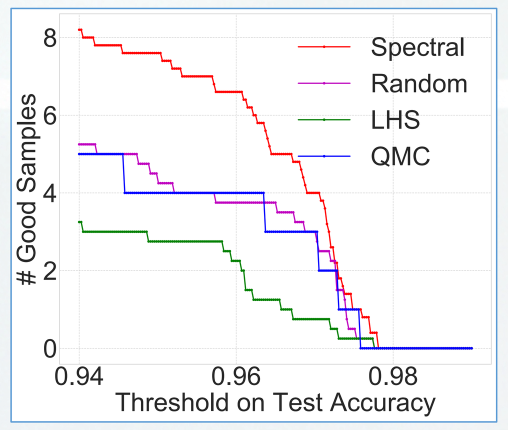


Significantly **lower MSE** compared to existing approaches

Case Study: Hyper-parameter search for deep neural networks



Can effectively recover optimal settings for training neural networks



Going Forward...

- Extensions to non-linear, **embedded manifolds** in high dimensions
- New strategies to balance **exploration** and **exploitation**
- Applications in **small-data deep learning** and **Auto ML**