Sequential Experimentation by Evolutionary Algorithms

Ofer M. Shir

Tel-Hai College and Migal Research Institute, Upper Galilee, Israel

Thomas Bäck

Natural Computing Group, Leiden University, The Netherlands



ermission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee vermined that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the all citation on the first page. Copyrights for third-party components of this work must be honored.

For all other uses, contact the owner/author(s).

ECCO '19 Companion, July 13–17, 2019, Prague, Czech Republic

D 2019 Copyright is held by the owner/author(s).

(CM ISBN 978-1-4503-6748-6/1907.

**SERVER OF THE OWNER OF THE OWNER OWN

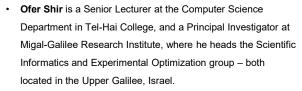




Agenda

- 1. What do we mean by "Sequential Experimentation"?
- 2. Examples of what has been done
- 3. Potential Application Areas
- 4. Reference: Statistical Design of Experiments
- 5. Case-Study: Quantum Control Experiments
- 6. Hot off the lab-bench: Protein Expression
- 7. Discussion: Conclusions and Open Questions

Instructors





Thomas Bäck is Professor of Computer Science at the Leiden Institute of Advanced Computer Science (LIACS), Leiden University, The Netherlands, where he is head of the Natural Computing group since 2002.



Contributors and former-instructors:

- · Joshua Knowles, University of Birmingham, UK.
- · Richard Allmendinger, University of Manchester, UK.

.

What do we mean by ...

SEQUENTIAL EXPERIMENTATION











"Typical" Characteristics

- Experiments are time-consuming
- Experiments are expensive
- Evaluations can also be subjective (human experts)
- Only few experiments are possible
- There are exceptions as well!

Quantum Control: Case-Study

- Evolution "in the loop"
- Thousands of experiments possible ("kHz regime")

Further Challenges

- Noise and uncertainty of measurements
- Multiple objectives
- Dynamically changing requirements of experimentalists / stakeholders!
- Dynamically changing (resource) constraints
- Cost choices during optimization
 - → Some experiments may cost more than others
- · Unusual constraints on population sizes, other hyperparameters

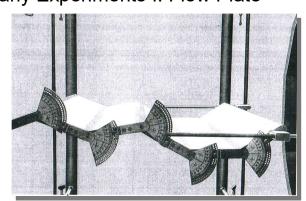


- Flow Plate
- Bended Pipe
- Nozzle
- Nutrient Solutions
- Coffee Formulations
 - **Quantum Control**
- Protein Expression



Early Experiments I: Flow Plate



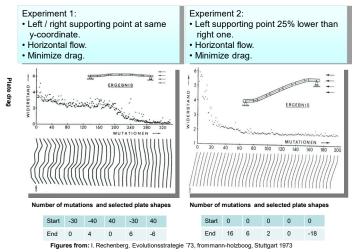


- · A plate with 5 controllable angle brackets
- Measurable air flow drag (by a pitot tube)

Figure from: I. Rechenberg, Evolutionsstrategie '73, frommann-holzboog, Stuttgart 1973

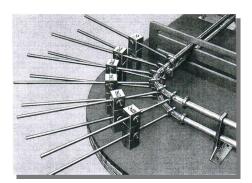
Early Experiments I: Flow Plate





Early Experiments II: Bended Pipe





- · A flexible pipe with 6 controllable bending devices
- · Minimize bend losses of liquid flow
- Measure drag by pitot tube

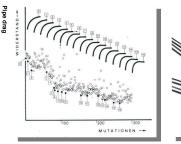
Figure from: I. Rechenberg, Evolutionsstrategie '73, frommann-holzboog, Stuttgart 1973

10

Early Experiments II: Bended Pipe



11





Number of mutations and selected pipe shapes

Initial (a) and optimized (b) pipe shape

- Bend loss of final form reduced by 10%
- · Including drag a total reduction of 2%

Figure from: I. Rechenberg, Evolutionsstrategie '73, frommann-holzboog, Stuttgart 1973

Early Experiment III: Nozzle



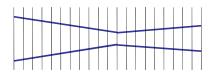
- What can be done if physics, (bio-) chemistry, ... of process unkown?
- No model or simulation program available!
- · Idea: Optimize with the real object
- "Hardware in the loop"
- Example: Supersonic nozzle, turbulent flow, physical model not available.







Experimental Setup: Nozzle



- Production of differently formed conic nozzle parts (pierced plates).
- Form of nozzle part is value of decision variable.

choosing conic nozzle parts (by EA) clamping of conic nozzle parts (manually) steam under high pressure passed into nozzle degree of efficiency is measured!

"simulator replacement"

Nozzle Experiment (I)



device for clamping nozzle parts

Figures courtesy of Hans-Paul Schwefel

collection of conical nozzle parts



Tel-Hai College

Tel-Hai College

Nozzle Experiment (II)

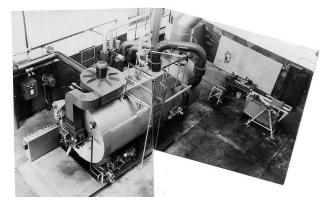


Hans-Paul Schwefel while changing nozzle parts

Figures courtesy of Hans-Paul Schwefel



Nozzle Experiment (III)



Figures courtesy of Hans-Paul Schwefel

steam plant / experimental setup



Tel-Hai

Nozzle Experiment (IV)





the nozzle in operation ...

... while measuring degree of efficiency Figures courtesy of Hans-Paul Schwefel

Nozzle Results (I)

• Illustrative Example: Optimize Efficiency

- Initial:



- Evolution:

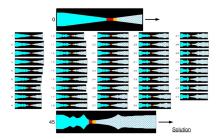


• 32% Improvement in Efficiency!



Tel-Hai

Nozzle Results (II)

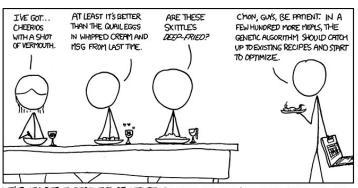


- · 250 experiments were made.
- · 45 improvements found.
- Discrete ring segments, variable-dimensional optimisation
- · Gene duplication and deletion as additional operators.

J. Klockgether and H.-P. Schwefel, "Two-phase nozzle and hollow core jet experiments," in Proceedings of the 11th Symposium on Engineering Aspects of Magneto-Hydrodynamics, Caltech, Pasadena, California, USA, 1970.

Experiment: Coffee Formulations



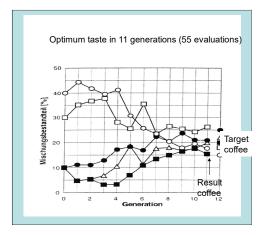


WE'VE DECIDED TO DROP THE CS DEPARTMENT FROM OUR WEEKLY DINNER PARTY HOSTING ROTATION.

M. Herdy: Beiträge zur Theorie und Anwendung der Evolutionsstrategie, PhD Thesis, Technical University of Berlin, Germany, 2000.

Coffee Formulations: Results





M. Herdy: Beiträge zur Theorie und Anwendung der Evolutionsstrategie, PhD Thesis, Technical University of Berlin, Germany, 2000.

EXPERIMENTAL

OPTIMIZATION:

FUNDAMENTALS

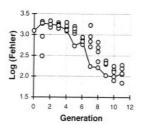
21

2

Coffee Formulations: Results



- · Coffee mixture differs a lot from target coffee!
- Taste is identical!
- · Multiple realizations, but cost optimal!
- Approximation of cubic polynomial: 35 evals.



M. Herdy: Beiträge zur Theorie und Anwendung der Evolutionsstrategie, PhD Thesis, Technical University of Berlin, Germany, 2000.

22

Experimental Requirements (for an Optimizer)



- 1. Speed: fast convergence is required
- 2. Reliability: reproducibility of results within a margin
 - Environmental parameters often hidden (temperature, pressure, ...)
- 3. Robustness: manufacturing feasibility
- Reference solution (recommended):
 pre-designed reference item, robust and stable, having a known objective function value

23



College

Convergence Speed

· Experiments are typically expensive:

Goal: Drive the system towards finding large improvements with as few experiments as possible.

- Practical solutions: "greedy" variants of evolutionary algorithms, e.g.,
 - Derandomized evolution strategies
 - ParEGO
 - Often "stochastic gradient search"
 - Need to support parallel execution!

25



Environmental Parameters

- As many as possible physical conditions should be recorded during the experiment
- Ideally, sensitivity of the system to the environment should be assessed
- Basic starting points: recording Signal/Noise, extracting power spectrum of the noise, etc.

Reliability of Results

- Mostly algorithm-dependent
- · Attained results must be reproducible
- Scenarios of recording *experimental outliers* must be avoided (elitism is tricky...)
- Perceived result versus a posteriori result
- · Possible solutions:
 - Employing comma (non-elitist) strategies
 - In ES, the recombination operator assists in treating noise (The Genetic Repair (GR) Hypothesis, Beyer)
 - Increasing sampling rate of measurements ("signal averaging")

26



Manufacturing Feasibility

- Mostly system-dependent
- Realization of the prescribed decision parameters of the experiment to equivalent systems, e.g., in a manufacturing stage
- To this end, sensitivity of the system must be assessed (electronics, for instance)
- Upon obtaining reproducible results, they should be verified on equivalent systems



Noise "Colors"

Autocorrelation of the noise spectrum indicates the "memory property" of the disturbance -

• White Noise: $^1\!\!/_{f^0} \to \delta(t)$ (no correlation)

• Pink (Flicker) Noise: $\frac{1}{f^1}$ \rightarrow unknown

• Red (Brownian) Noise: $1/f^2 \rightarrow e^{-\lambda t}$ (exp. distribution)

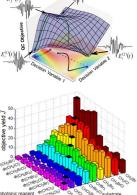
Tip: Assess the stability of your system by extracting the Power Spectral Density of its signal-free state.

M. Roth, J. Roslund, and H. Rabitz, "Assessing and managing laser system stability for quantum control experiments", Rev. Sci. Instrum. 77, 083107 (2006)

Tel-Hai

Basic Science: Discoveries as **Combinatorial Optimization Problems**

- A problem shared by scientists is to achieve optimal behavior of their systems and arrive at new discoveries while searching over an array of parameters
- It is commonly visualized in terms of a 'landscape': a candidate solution is mapped onto a 'position', its quality onto an 'altitude'
- The task becomes to efficiently navigating within this search-space, which scales exponentially with the number of variables



Kell, D.B., Scientific discovery as a combinatorial optimisation problem: How best to navigate the landscape of possible experiments? BioEssays, 2012. 34(3): 236-244.

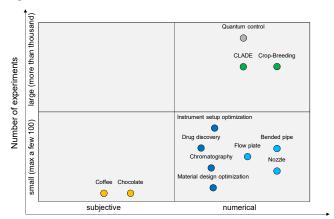
APPLICATION AREAS







A Classification



Selection/evaluation



Potential Application Areas

- · Cosmetics / Detergent Formulation Optimization
- Catalyst Formulation Optimization (Cost, Effectiveness, ...)
- · Subjective Evaluation Applications based on Human Taste or other Senses
- · Engineering Applications Requiring Real-World **Experiments for Measurement**
- · Concrete Formulation Optimization
- · Glue Formulation Optimization
- · Plant Startup Process
- Chemical Compound Synthesis Processes (e.g., Drugs)
- Instrument Setup Optimization

Reference/State-of-the-Art:

STATISTICAL DESIGN OF **EXPERIMENTS**







Introductory charts courtesy of Joshua Knowles.



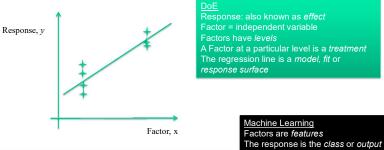


DoE: The Industry's Golden Standard



Experimentation terminology





actors are *decision variables* onse is *objective value, cost, benefit, utility* or *fitnes*

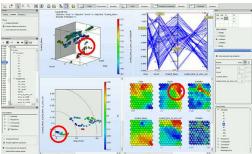


College Tel-Hai

Modern experiment 1

- N factors, N>>2, e.g. genes
- M effects, M>1, e.g. disease, + other effects
- P>1 nuisance factors, ages, gender, etc
- Possible Research Questions: which genes are most responsible for the disease, which groups of genes work together, and are other effects involved in explaining the disease?

Modern experiment 2



- Many factors
- Several effects
- Several nuisance variables
- Limited number of samples
- Noise (variance)
- Purpose: Optimize the effect

3





39

Classical DoE: topics

- OFAT
- · Full factorial/Fractional factorial
- LHS
- · Other designs

Handling multiple factors

- It is typical that we have *N*>1 factors to control
- The high-school solution to this is called

OFAT

(or one-factor-at-a-time)

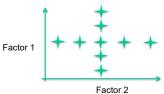
• You hold all but one factor constant and vary that. Then you go onto the second factor ... and so on





OFAT

An OFAT design in two variables



Weaknesses of OFAT

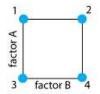
- 1. OFAT requires more* runs for the same precision in effect estimation
- 2. OFAT cannot estimate interactions between factors
- 3. OFAT can miss optimal settings of factors

42



44

Full Factorial design

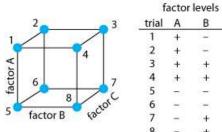


factor levels		
trial	Α	В
1	+	85
2	+	+
3	-	-
4	-	+



+ : alternative value

(-, ..., -) is the existing design

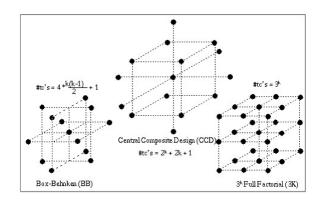


- ANOVA and linear regression
- All interactions can also be identified
- → linear model

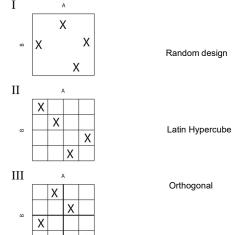
43

College

Other designs



Latin hypercube screening design

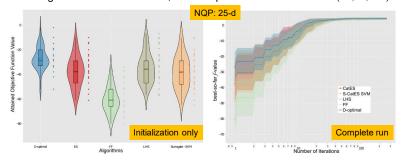


^{*}compared with experimental designs like Plackett-Burman



DoE vs. EAs on Combinatorial Optimization

- Comparing a Categorical ES, with/without surrogates, to modern DoEs
- Budget of ~2000 evaluations; discrete problems at dimensions {25,64,100}



Horesh, N., Bäck, T., Shir O.M.: Predict or Screen Your Expensive Assay? DoE vs. Surrogates in Experimental Combinatorial Optimization. In: Proceedings of the Genetic and Evolutionary Computation Conference, GECCO-2019, New York, NY, USA, ACM Press (2019) To Appear

Case-Study:

QUANTUM CONTROL EXPERIMENTS





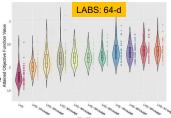


X Tel-Hai College

DoE vs. EAs on Combinatorial Optimization

Conclusions per the reported observations:

- Using surrogate-aided iterative search was observed to perform best on such setups with a small budget.
- DoE-initializations alone are inferior with respect to initializations that are followed by ES iterative search.
- There is no evident gain in granting more than 30% of the budget on DoEinitializations.
- D-Optimal was the most successful DoE methods on the low-dimensional setup, yet experiencing problem-dependent behavior.



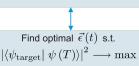
Horesh, N., Bäck, T., Shir O.M.: Predict or Screen Your Expensive Assay? DoE vs. Surrogates in Experimental Combinatorial Optimization. In: Proceedings of the Genetic and Evolutionary Computation Conference, GECCO-2019, New York, NY, USA, ACM Press (2019) To Appear

Tel-Hai (Land College (Land Co

Altering the Course of Quantum Phenomena

"Electric Field Design" Quantum Control Theory $i\hbar\frac{\partial}{\partial t}\left|\psi\left(t\right)\right\rangle = \mathcal{H}(t)\left|\psi\left(t\right)\right\rangle$ $\mathcal{H}\left(t\right) = \mathcal{H}_{0} - \vec{\mu}\cdot\vec{\epsilon}\left(t\right)$

Rabitz et al.

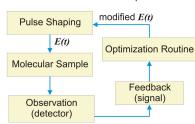


Hamiltonian required

PRA 37, 4950 (1988)

Judson and Rabitz

"Teaching Lasers to Control Molecules" Quantum Control Experiments



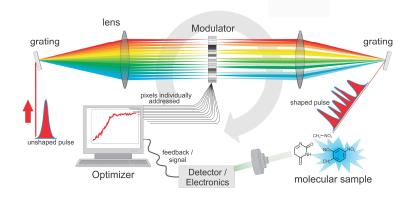
Hamiltonian not required

PRL 68, 1500 (1992)



Tel-Hai College

Quantum Control Experiments



The QCE Arena: The Optical Table

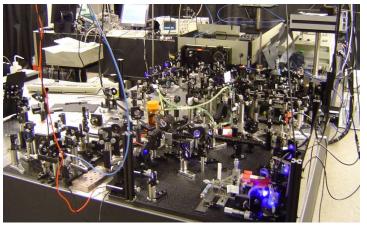


Figure courtesy of Jonathan Roslund

51

Tel-Hai College



52

The Optical Table: Shaping the Pulse

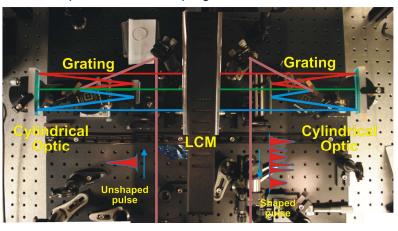
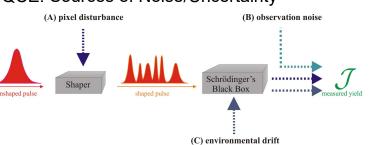


Figure courtesy of Jonathan Roslund

QCE: Sources of Noise/Uncertainty



- (A) $\tilde{\phi}(\omega) = (\phi(\omega_1) + \mathcal{N}_1(0, \epsilon_S^2), \dots, \phi(\omega_n) + \mathcal{N}_n(0, \epsilon_S^2))$
- **(B)** $\tilde{\mathcal{J}} = \mathcal{J} + \mathcal{N}\left(0, \epsilon_{\mathcal{J}}^{2}\right)$ Signal Averaging: $\left\langle \tilde{\mathcal{I}} \right\rangle = \mathcal{I}$, $\operatorname{VAR}\left[\tilde{\mathcal{I}}\right] = \frac{\epsilon_{\mathcal{J}}^{2}}{k}$
- (C) $\hat{\mathcal{J}}(t) = \tilde{\mathcal{J}} + \xi(t)$



Single-Objective QCE

- CMA-ES was observed to perform extremely well with small population sizes
- · Recombination is indeed necessary (GR, Beyer)
- · Robust, reproducible, reliable solutions

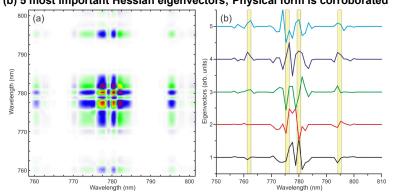


Roslund, J., Shir, O.M., Bäck, T., Rabitz, H.: Accelerated Optimization and Automated Discovery with Covariance Matrix Adaptation for Experimental Quantum Control. Physical Review A (Atomic, Molecular, and Optical Physics) **80**(4) (2009) 043415

54

Extended Features: Statistical Learning (FOCAL)

(a) Retrieving the Hessian by FOCAL for rank-deficient atomic Rubidium (b) 5 most important Hessian eigenvectors; Physical form is corroborated

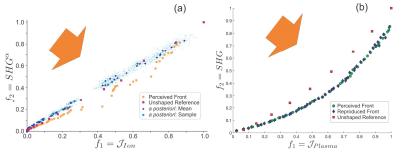


Shir, O.M., Roslund, J., Whitley, D., Rabitz, H.: Efficient retrieval of landscape Hessian: Forced optimal covariance adaptive learning. Physical Review E 89(6) (2014) 063306

Multi-Objective QCE



- (a) Experimental Pareto frontier for the Total Ionization problem approximated by MO-CMA-ES, displaying the perceived frontier of a single experiment, the **reference frontier** of the intensity based non-shaped pulse, as well as a sampling of the Pareto optimal set.
- (b) Experimental Pareto frontier for the Molecular Plasma Generation problem approximated by MO-CMA-ES remedied with occasional re-evaluation, displaying the perceived frontier, the reference frontier, and the reproduction of the Pareto optimal set.



Shir, O.M., Roslund, J., Leghtas, Z., Rabitz, H.: Quantum Control Experiments as a Testbed for Evolutionary Multi-Objective Algorithms. Genetic Programming and Evolvable Machines 13(4) (2012) 445—491

Hot-off-the-lab-bench

PROTEIN EXPRESSION



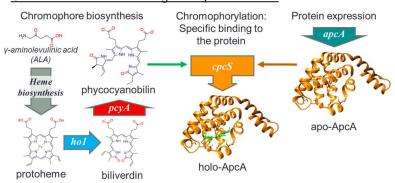






Heterologous Protein Expression

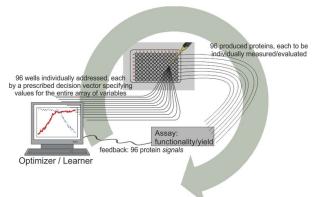
Four genes are required for ApcA heterologous expression in *E. coli*. Goal: maximize the heterologous expression level



Erlich, Ch.: Experimental combinatorial optimization of phycobiliproteins' expression in *E.coli*. Thesis Tel-Hai College (2019).



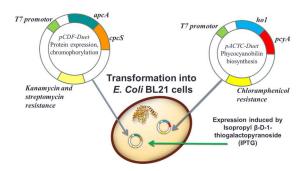
Closed Feedback Loop



A scheme of a typical single iteration (step) in a proposed optimization run of a production system. The input variables for each well are prescribed by the algorithm, whereas the output (feedback) of each produced molecule is provided by the assay - altogether closing a feedback loop.



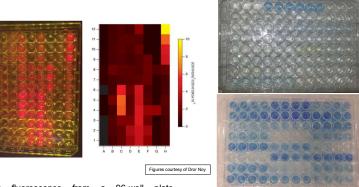
ApcA Expression System: Four Genes in Two Plasmids



- Controls (10 categorical decision variables):
 - 4 growth temp., 5 expression temp., 3 growth volumes, 6 IPTG concent., 5 O.D. values, 4 induction durations, 7 gamma-ALA concent., 2 gamma-ALA timings, 11 FeCl 3 concent., 3 Medium types
- Search-space cardinality: ~3×10⁶ possible combinations



ApcA Expression in E. coli: Assay



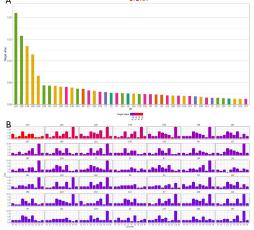
Left: fluorescence from a 96-well plate containing lysates of *E. coli* cells expressing ApcA. Right: evaluation of expression quality based on the ratio of absorption at 620nm vs. 280nm.

Images of the 1st and the 6th generation's collection plate, after cleaning and separating the proteins. The blue pigment is indicative of the expression strength.



Tel-Hai 🖟

ApcA Expression in *E. coli*: 6 generations



Top 25% combinations in terms of target values. A. Bar graph of the top 25% target values ordered from high to values, colored by generation. B. A gallery of the combinations of the top 25% objective function values throughout the experimental campaign, depicting explicit decision variables' values in bar-plots (the X-axis describes the decision variables and Y-axis shows their values). The color-map reflects the objective function

Erlich, Ch.: Experimental combinatorial optimization of phycobiliproteins' expression in *E.coli*. Thesis Tel-Hai College (2019).

DISCUSSION

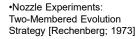




63

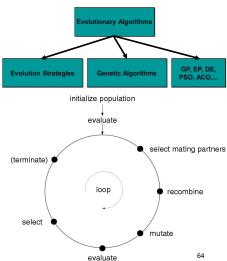
Tel-Hai

Evolutionary Algorithms Used

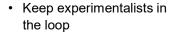


•Quantum Control Experiments: Derandomized Evolution Strategies [Hansen et al.; 1994-2008]

•Protein Expression Experiments: Categorical ES [unpublished]



Some Practical Principles for Closed-Loop Optimization



- Understand the experimental platform
- Simulate the platform, and compare algorithms
- Do it for real and get feedback

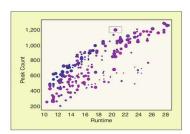




Tel-Hai College

Keep experimentalists in the loop

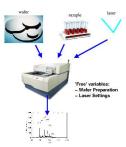
- Explain EAs, manage expectations of outcomes.
- Understand the variables and objectives. <u>Confirm 3</u> times at least.
- Still be prepared to change objectives half-way through!
- Enable them to use familiar software for viewing results.



Objectives shown above were changed during optimization

Understand the experimental platform

- Variables, constraints, measurements, noise
- · Financial costs, time lags
- Resource constraints
- Batch size of platform dictates/constrains population size of EA



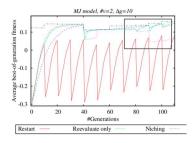
6

Tel-Hai

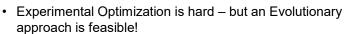


Simulations prior to the real thing

- Really helpful to manage expectations of stakeholders
- Tune your algorithms for weird and wonderful population sizes, constraints, budget limitations of real experimental platform
- If possible, use domain experts to design test problems that are similar to the real problem



Goals and Open Questions



- Fundamental research in EAs is much needed:
 - Given a budget of **k** experiments what strategy should be taken?
- NFL holds more than ever there will be no winner algorithm handling all experimental scenarios!
- How do statistical approaches perform in comparison?
 - Especially DoE
- The comparison presented earlier is a fine starting point

68



Al Prospects?



Broader Picture: Al

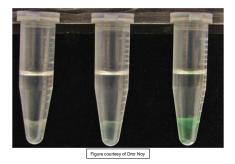
- The Artificial Intelligence (AI) and Machine Learning (ML) revolution already takes place.
- No doubt that ML may boost scientific research by applying pattern recognition. But is that it?
- Some universities target this direction in education already at the BSc/MSc levels (<u>CMU</u>), stating that "AI will drive more decisions in bio-experiments in the future"

70



Acknowledgments

- Joshua Knowles
- · Hans-Paul Schwefel
- Dror Nov
- Chen Erlich
- Naama Horesh
- · Richard Allmendinger
- Herschel Rabitz
- Jonathan Roslund



- The human/psychological factor among the experimentalists already plays a dominant role: shift the scientist/engineer's aim into explaining nature of solutions (mechanism!), rather than finding them
- But, existing hypotheses are already well-documented, plus there are established knowledge representation frameworks
- Next step?
 Al-based algorithm to formulate a scientific hypothesis and design experimentation

71



Bibliography

- Allmendinger, R. and Knowles, J. (2013). On handling ephemeral resource constraints in evolutionary search. Evolutionary Computation, 21(3):497–531.
- Bäck, T., Foussette, C., and Krause, P. (2013). Contemporary Evolution Strategies. Natural Computing Series. Springer-Verlag Berlin Heidelberg.
- Box, G. E., Hunter, J. S., and Hunter, W. G. (2005). Statistics for Experimenters: Design, Innovation and Discovery. John Wiley and Sons, Hoboken, NJ, USA, second edition.
- Eiben, A. E. and Smith, J. (2015). From evolutionary computation to the evolution of things. Nature, 521:476-482.
- Kell, D. B. (2012). Scientific discovery as a combinatorial optimisation problem: How best to navigate the landscape of possible experiments? BioEssays, 34(3):236–244.
- Knowles, J. (2006). ParEGO: A Hybrid Algorithm with On-Line Landscape Approximation for Expensive Multiobjective Optimization Problems. IEEE Transactions on Evolutionary Computation, 10(1):50–66.
- Knowles, J. (2009). Closed-loop evolutionary multiobjective optimization. IEEE Computational Intelligence Magazine, 4(3):77–91.
- Rechenberg, I. (2000). Case studies in evolutionary experimentation and computation. Computer Methods in Applied Mechanics and Engineering, 186(24):125–140.
- Roslund, J., Shir, O. M., Bäck, T., and Rabitz, H. (2009). Accelerated Optimization and Automated Discovery with Covariance Matrix Adaptation for Experimental Quantum Control. *Physical Review A*, 80(4):043415.
- Shir, O. M., Roslund, J., Leghtas, Z., and Rabitz, H. (2012). Quantum control experiments as a testbed for evolutionary multi-objective algorithms. Genetic Programming and Evolvable Machines, 13(4):445–491.
- Shir, O. M., Roslund, J., Whitley, D., and Rabitz, H. (2014). Efficient retrieval of landscape hessian: Forced optimal covariance adaptive learning. Physical Review E. 89:063306.

72