Cloud Scale, Machine Learning with FlexGP

Una-May O'Reilly
Evolutionary Design and Optimization Group
Computer Science and Artificial Intelligence Lab
MIT


HELP

imagination at work

We bring good things to life.

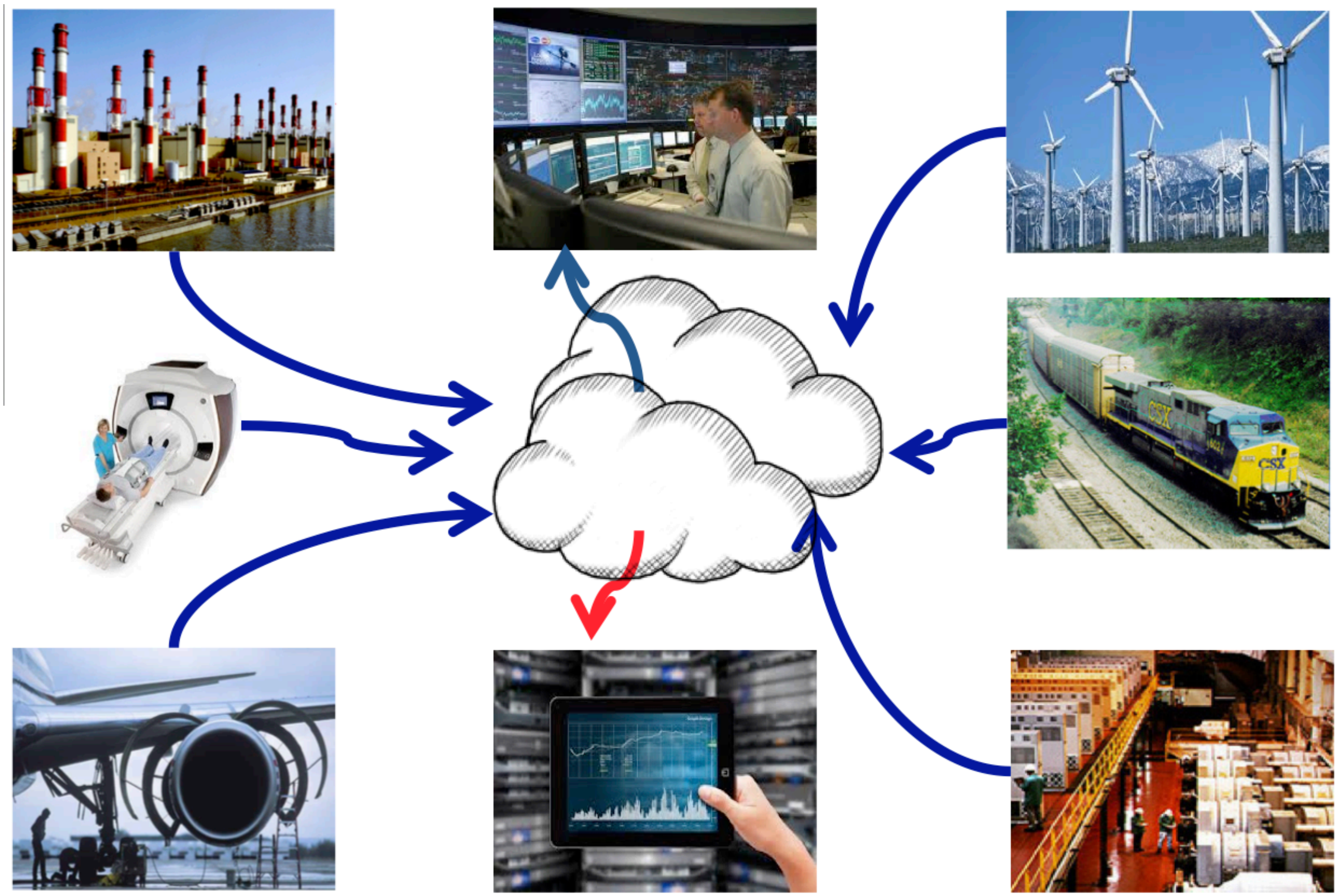
Lots of Data Everywhere

Engineering

EvoDesignOpt

CSAIL

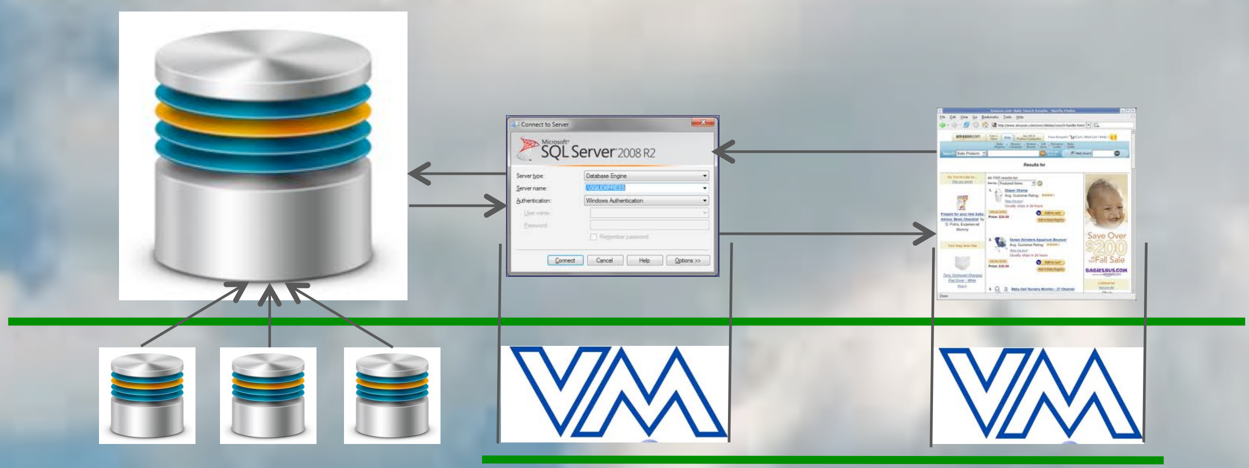
Lots of Data Everywhere



EvoDesignOpt

CSAIL

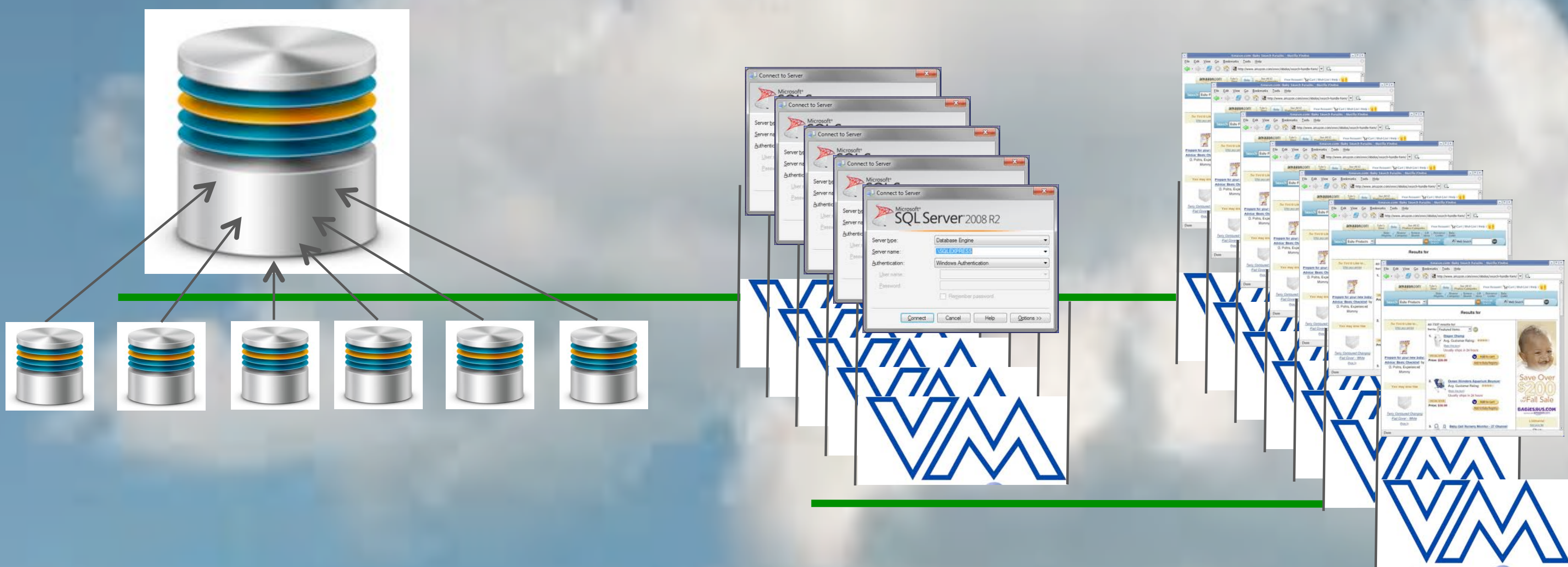
The Cloud's Role



EvoDesignOpt

CSAIL

The Cloud's Role



+Elasticity
+Infinite resources on demand
+Budget and time choice space

-robustness
-Time to scale up
-Need interim solutions
-algorithms need to exploit the positives

Agenda

- Strategies for cloud-scale machine learning with massive data
- FlexGP
  - Flexibly factored, flexibly scaled machine learning with Genetic Programming (GP)
  - Deeper Dives
    - Launch
    - Genetic programming learning engines for ML
- Beyond FlexGP

EvoDesignOpt

CSAIL

Strategies for Machine Learning

- **Scaled up, existing algorithms are not completely sufficient**

People who are really *serious about software* should make their own hardware. (Allan Kay)

The hardware is the cloud

ML algorithms should be designed with the assumption of infinite resources

Ensembles of Diverse Learners

Ensembles

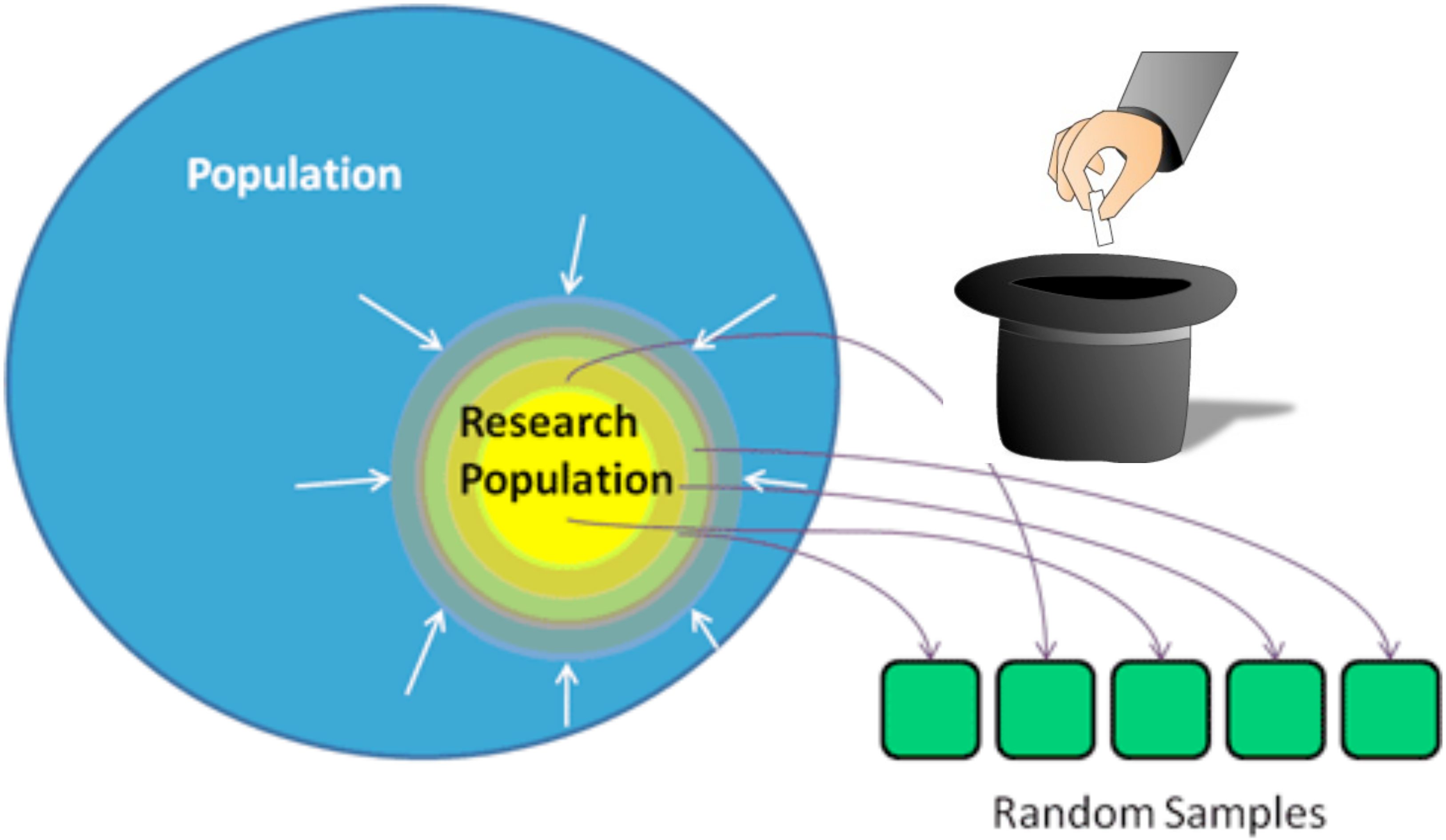
- factoring
  - » Heterogeneous learning engines
    - Training data D
    - Within Algorithm (PI)
      - Model structure
      - Objective
      - Indicators/Explanatory vars
    - Across algorithms
- filtering
  - » Diverse models or classifiers or clusters
- Fusion
  - » A robust result

Ensembles of Diverse Learners

EvoDesignOpt

CSAIL

Distributed sampling approaches



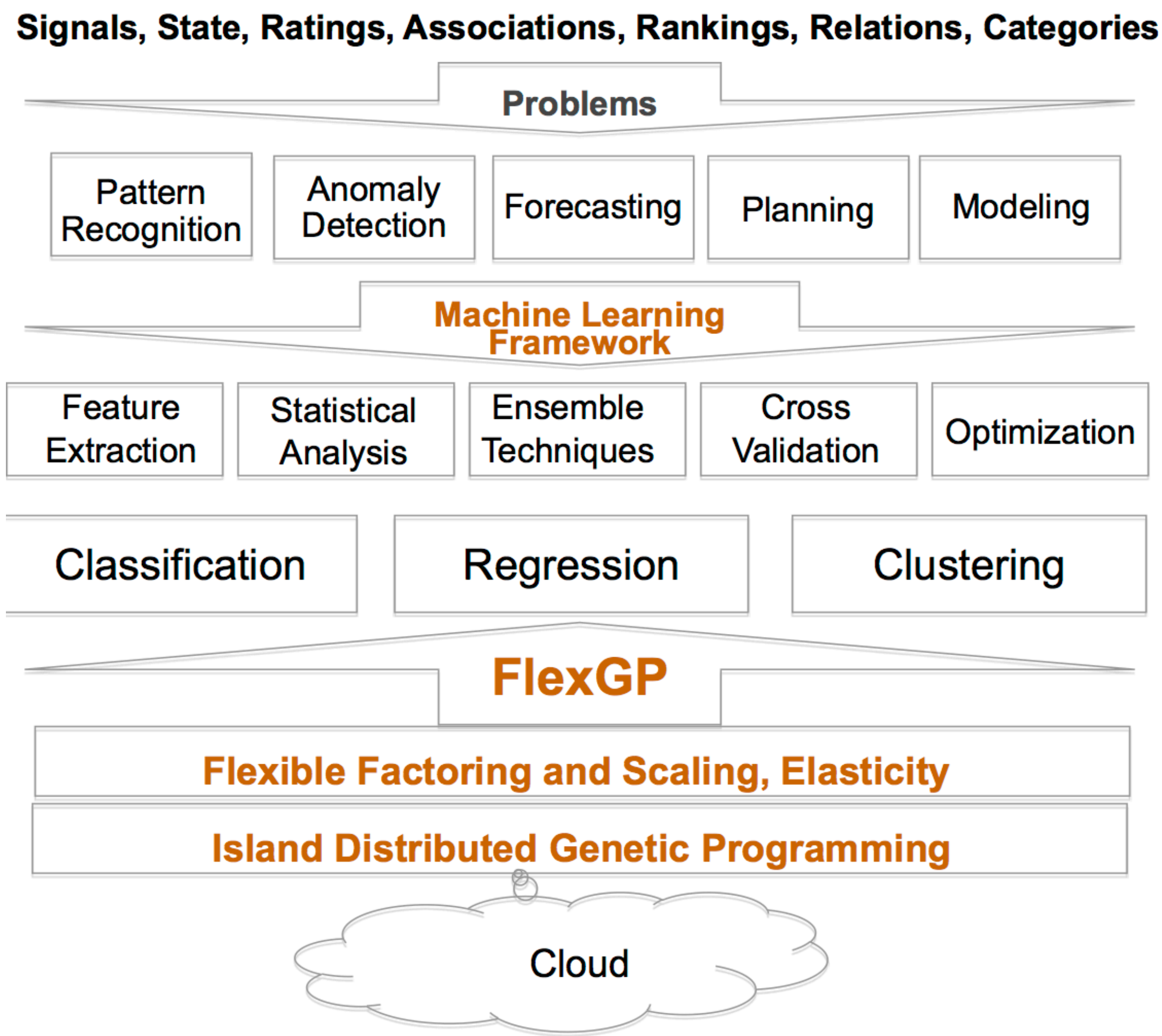
Agenda

- Strategies for cloud-scale machine learning
- FlexGP
  - Flexibly factored, flexibly scaled machine learning with Genetic Programming (GP)
  - Deeper Dives
    - Launch
    - Genetic programming learning engines for ML
- Beyond FlexGP

EvoDesignOpt

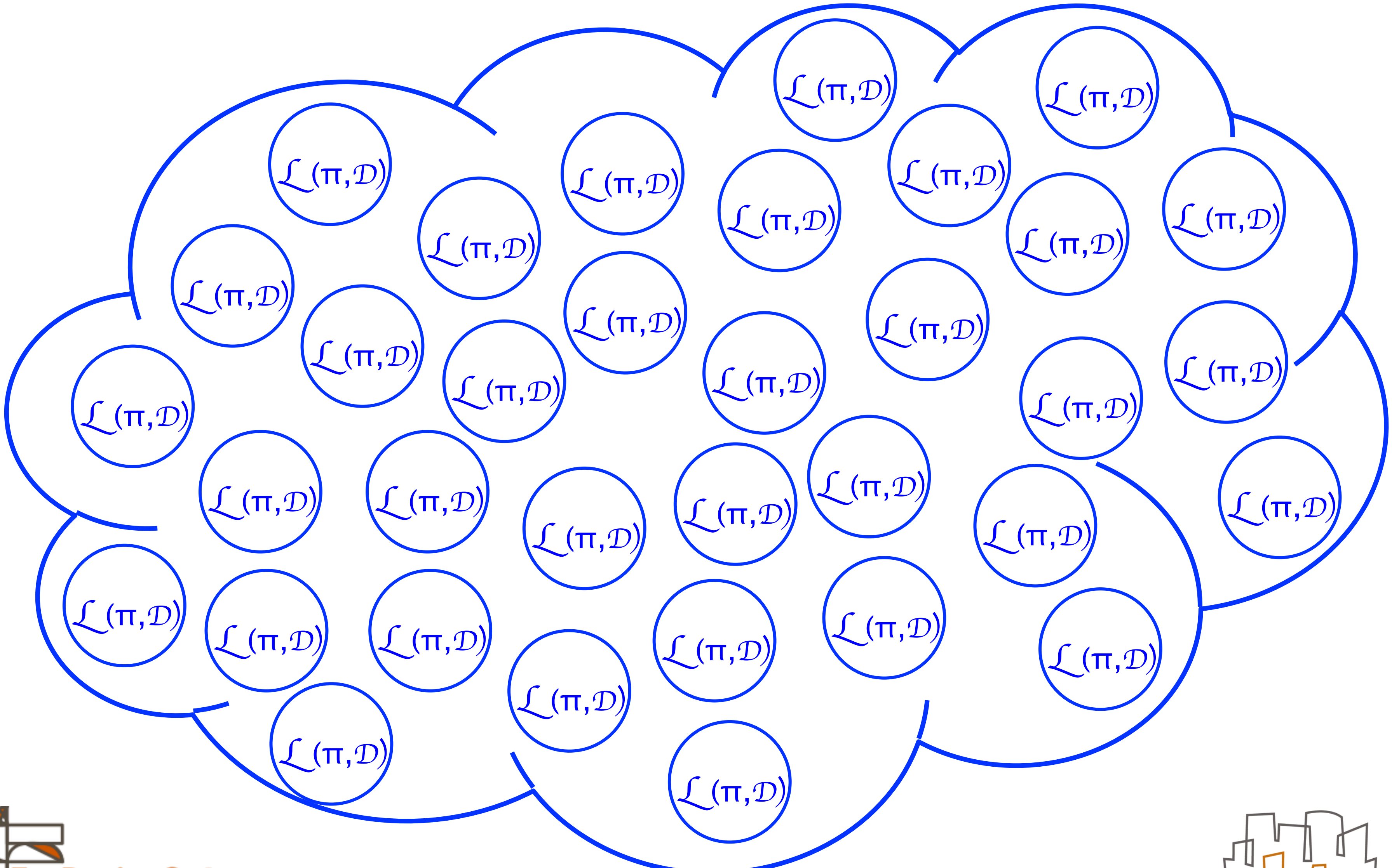
CSAIL

FlexGP

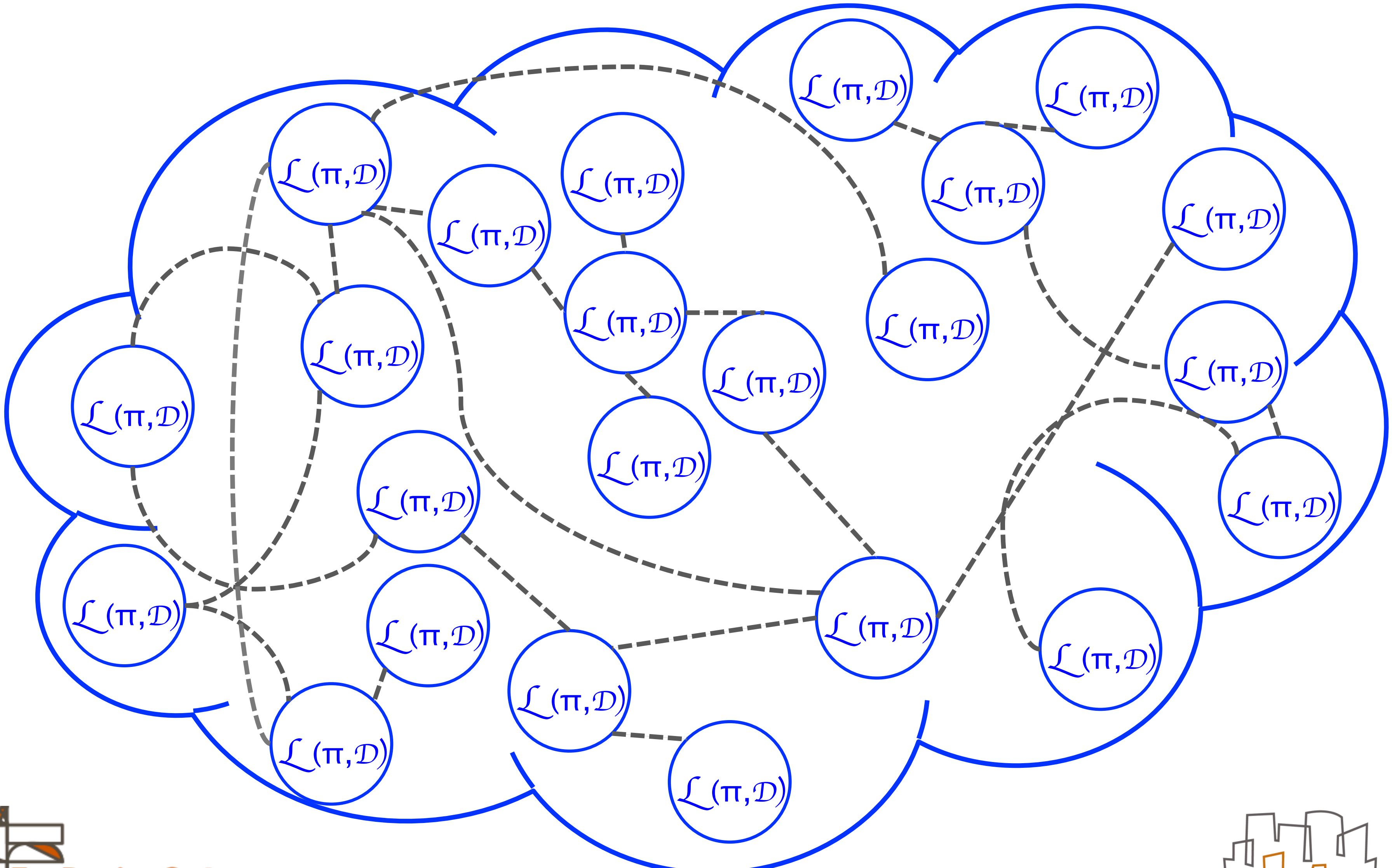


FlexGP Overview

Cloud with Learners



Cloud with Networked Learners



EvoDesignOpt

CSAIL

FlexGP Learning Engines

$$\mathcal{L}(\pi, \mathcal{D})$$

π_1 = { + - * / sin cos tan sqrt} Model operators

π_2 = L3 Objective function

π_3 = (x2 x3 x4) Explanatory vars

Data

N	T	t	PMZ for $w =$ 0.2	0.4	0.6	0.8	Ave. comp. time (sec.)	Ave. no. of Pareto optimal schedules
50	10	2	11.30	13.25	11.92	9.54	0.06	11
		4	10.30	12.16	10.58	9.26	0.06	12
	14	2	12.78	14.36	13.06	11.24	0.04	8
		4	11.44	13.32	12.50	10.10	0.05	9
	18	2	15.81	16.96	15.68	13.82	0.03	6
		4	14.88	15.61	15.14	12.26	0.03	7
100	10	2	10.46	11.48	10.35	9.22	0.08	14
		4	10.00	10.86	9.89	8.60	0.08	16
	14	2	10.18	11.75	10.54	9.30	0.06	12
		4	9.80	11.06	10.10	9.01	0.07	14
	18	2	11.66	13.59	12.44	10.30	0.05	9
		4	11.32	12.76	11.10	9.62	0.05	10
150	10	2	8.88	9.06	8.20	7.62	0.09	15
		4	8.22	8.60	7.96	6.99	0.10	17
	14	2	8.50	9.75	8.86	7.52	0.08	13
		4	7.88	9.03	8.38	7.10	0.09	15
	18	2	9.76	10.96	10.19	8.60	0.07	11
		4	9.85	10.20	9.64	7.82	0.08	13
200	10	2	6.96	8.19	7.10	5.66	0.13	20
		4	6.25	7.80	6.76	5.28	0.14	22
	14	2	7.12	8.62	7.28	6.32	0.12	18
		4	6.55	8.26	6.98	5.69	0.13	20
	18	2	8.19	9.49	8.63	7.08	0.10	17
		4	8.39	9.67	8.58	6.35	0.11	18

$$\frac{\cos(x_4)}{\tan(x_2)+x_2} + \mathrm{sqrt}(x_3)$$

Model or classifier

EvoDesignOpt

CSAIL

FlexGP Learning Engines

$\pi, \mathcal{D}$

π_1 = { + - * /}

π_2 = mean squared error (L2)

π_3 = (x1, x2, x3, x5)

$\mathcal{D}$

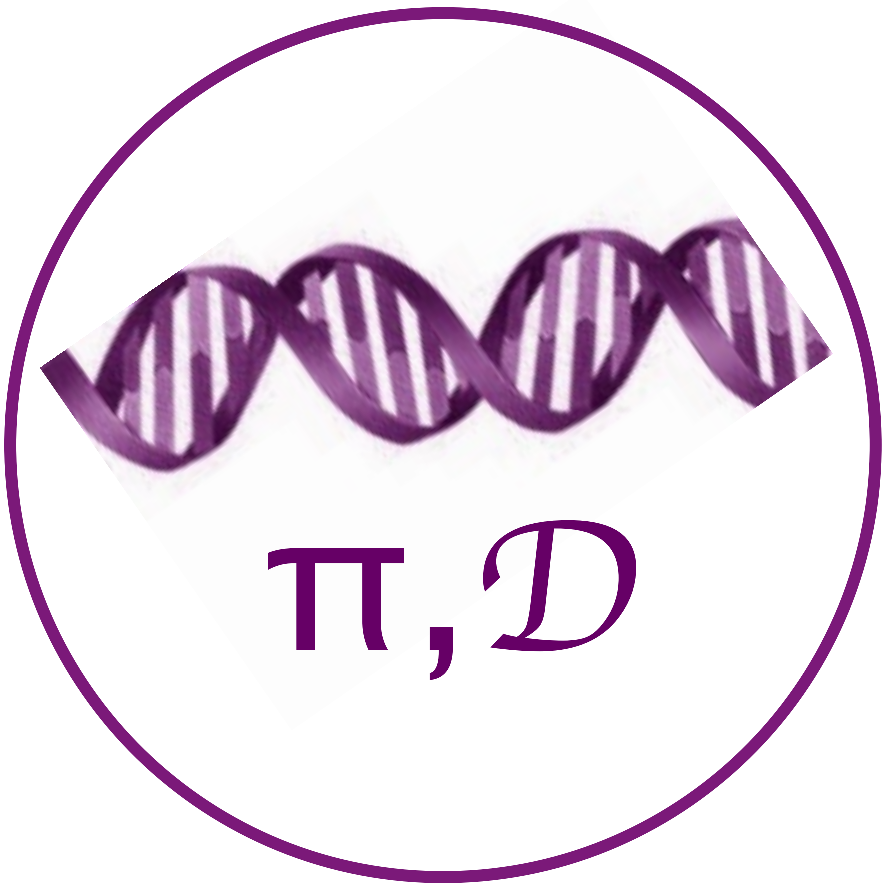
$3x_1 + x_2 x_5 - \frac{x_3^2}{\ldots}$

N	T	t	PMZ for $w =$ 0.2	0.4	0.6	0.8	Ave. comp. time (sec.)	Ave. no. of Pareto optimal schedules
50	10	2	11.30	13.25	11.92	9.54	0.06	11
		4	10.30	12.16	10.58	9.26	0.06	12
	14	2	12.78	14.36	13.06	11.24	0.04	8
		4	11.44	13.32	12.50	10.10	0.05	9
	18	2	15.81	16.96	15.68	13.82	0.03	6
		4	14.88	15.61	15.14	12.26	0.03	7
100	10	2	10.46	11.48	10.35	9.22	0.08	14
		4	10.00	10.86	9.89	8.60	0.08	16
	14	2	10.18	11.75	10.54	9.30	0.06	12
		4	9.80	11.06	10.10	9.01	0.07	14
	18	2	11.66	13.59	12.44	10.30	0.05	9
		4	11.32	12.76	11.10	9.62	0.05	10
150	10	2	8.88	9.06	8.20	7.62	0.09	15
		4	8.22	8.60	7.96	6.99	0.10	17
	14	2	8.50	9.75	8.86	7.52	0.08	13
		4	7.88	9.03	8.38	7.10	0.09	15
	18	2	9.76	10.96	10.19	8.60	0.07	11
		4	9.85	10.20	9.64	7.82	0.08	13
200	10	2	6.96	8.19	7.10	5.66	0.13	20
		4	6.25	7.80	6.76	5.28	0.14	22
	14	2	7.12	8.62	7.28	6.32	0.12	18
		4	6.55	8.26	6.98	5.69	0.13	20
	18	2	8.19	9.49	8.63	7.08	0.10	17
		4	8.39	9.67	8.58	6.35	0.11	18

EvoDesignOpt

CSAIL

FlexGP Learning Engines



π_1 = { + - * / sin cos tan sqrt}

π_2 = L3

π_3 = (x2 x3 x4)

$\mathcal{D}$

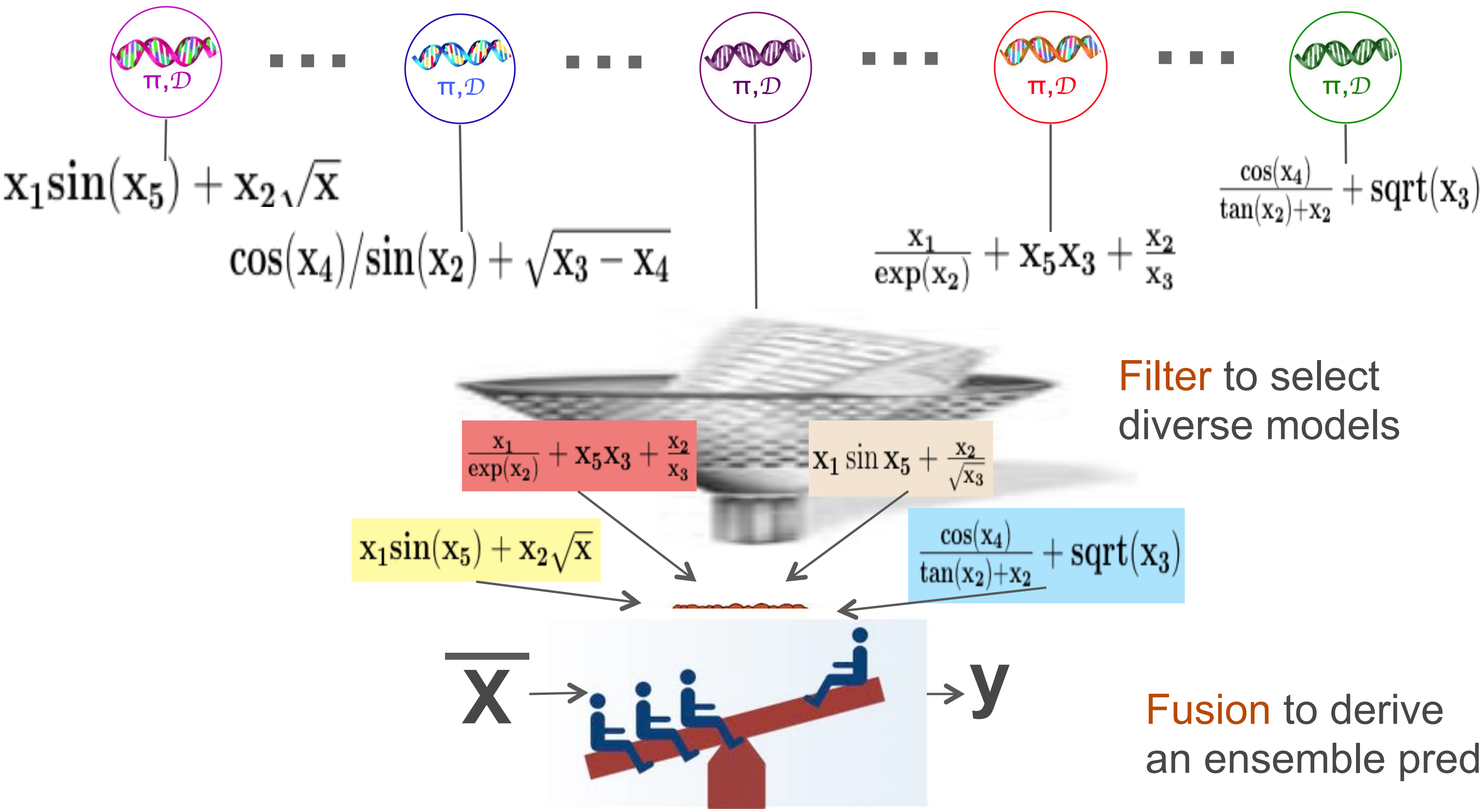
$$\frac{\cos(x_4)}{t} + \mathrm{sqrt}(x_2)$$

N	T	t	PMZ for $w =$ 0.2	0.4	0.6	0.8	Ave. comp. time (sec.)	Ave. no. of Pareto optimal schedules
50	10	2	11.30	13.25	11.92	9.54	0.06	11
		4	10.30	12.16	10.58	9.26	0.06	12
	14	2	12.78	14.36	13.06	11.24	0.04	8
		4	11.44	13.32	12.50	10.10	0.05	9
	18	2	15.81	16.96	15.68	13.82	0.03	6
		4	14.88	15.61	15.14	12.26	0.03	7
100	10	2	10.46	11.48	10.35	9.22	0.08	14
		4	10.00	10.86	9.89	8.60	0.08	16
	14	2	10.18	11.75	10.54	9.30	0.06	12
		4	9.80	11.06	10.10	9.01	0.07	14
	18	2	11.66	13.59	12.44	10.30	0.05	9
		4	11.32	12.76	11.10	9.62	0.05	10
150	10	2	8.88	9.06	8.20	7.62	0.09	15
		4	8.22	8.60	7.96	6.99	0.10	17
	14	2	8.50	9.75	8.86	7.52	0.08	13
		4	7.88	9.03	8.38	7.10	0.09	15
	18	2	9.76	10.96	10.19	8.60	0.07	11
		4	9.85	10.20	9.64	7.82	0.08	13
200	10	2	6.96	8.19	7.10	5.66	0.13	20
		4	6.25	7.80	6.76	5.28	0.14	22
	14	2	7.12	8.62	7.28	6.32	0.12	18
		4	6.55	8.26	6.98	5.69	0.13	20
	18	2	8.19	9.49	8.63	7.08	0.10	17
		4	8.39	9.67	8.58	6.35	0.11	18

EvoDesignOpt

CSAIL

FlexGP Ensemble Fusion

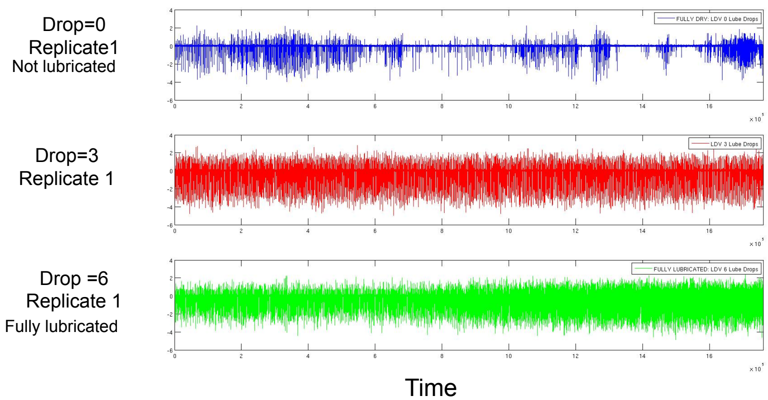


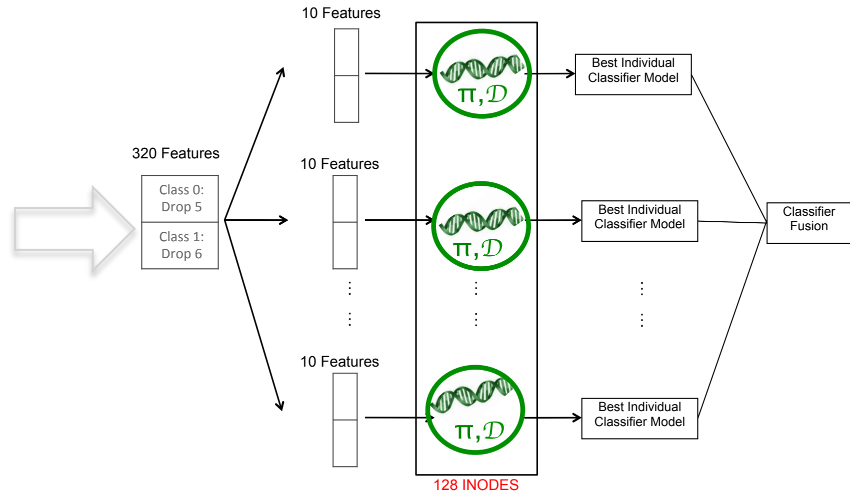
Filter to select diverse models

Fusion to derive an ensemble prediction

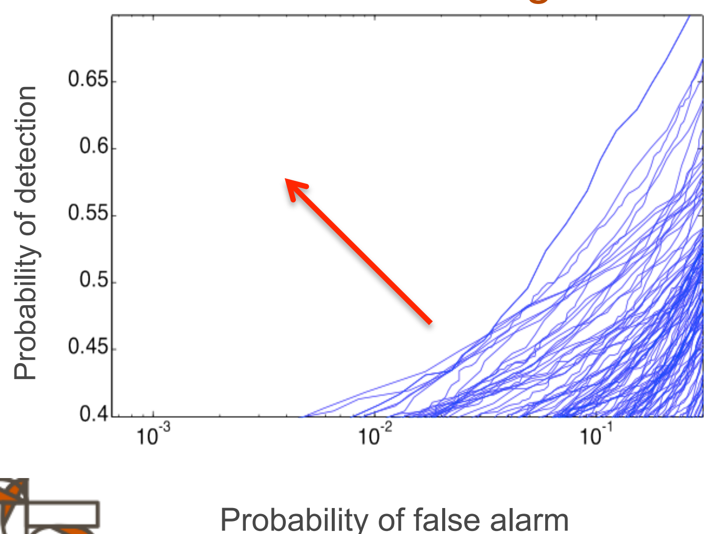
FlexGP Demonstrated

Build a classifier that can discriminate between these signals?

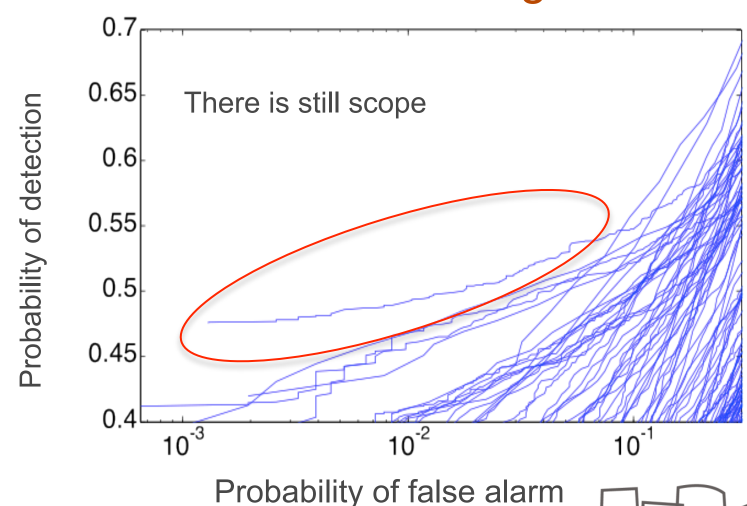




Before Learning



After Learning



EvoDesignOpt

CSAIL

Agenda

- Strategies for cloud-scale machine learning
- FlexGP
  - Flexibly factored, flexibly scaled machine learning with Genetic Programming (GP)
  - Deeper Dives
    - Launch
    - Genetic programming learning engines for ML
- Beyond FlexGP

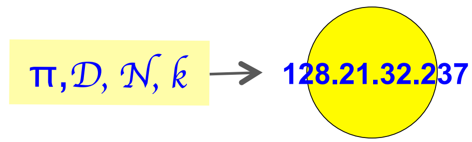
Cascading, Asynchronous Launch

"Start" node initiates recursive local launches

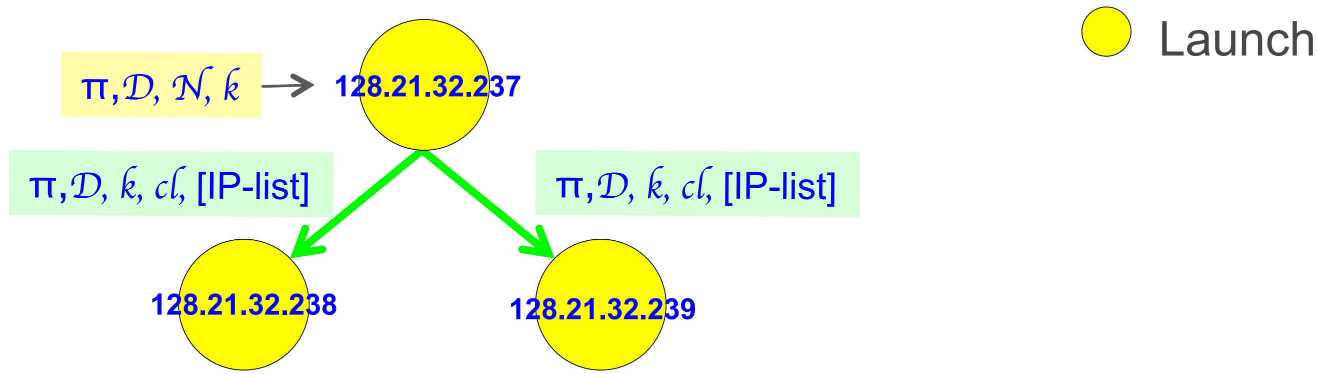
  - Inputs are distributions of $\pi,\mathcal{D}$ and cascading values: $\mathcal{N},k$ -> cl

Each node

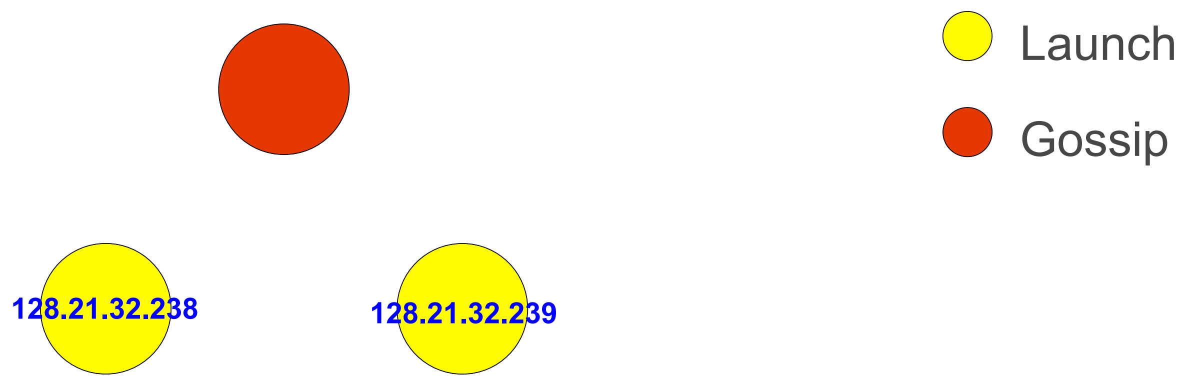
- **Phase 1: launch k other nodes if cl >0**
  - Each child is sent distributions $\pi,\mathcal{D}$ and k, $cl=cl-1$
  - Each child is sent ancestors' IPs: IP-list
- **Phase 2:**
  - Thread 1: global IP discovery through gossip
    - » Select an IP, dispatch IP-list
    - » Return IP-list to any sender
  - Thread 2: $\mathcal{L}(\pi,\mathcal{D})$ after sampling from distributions



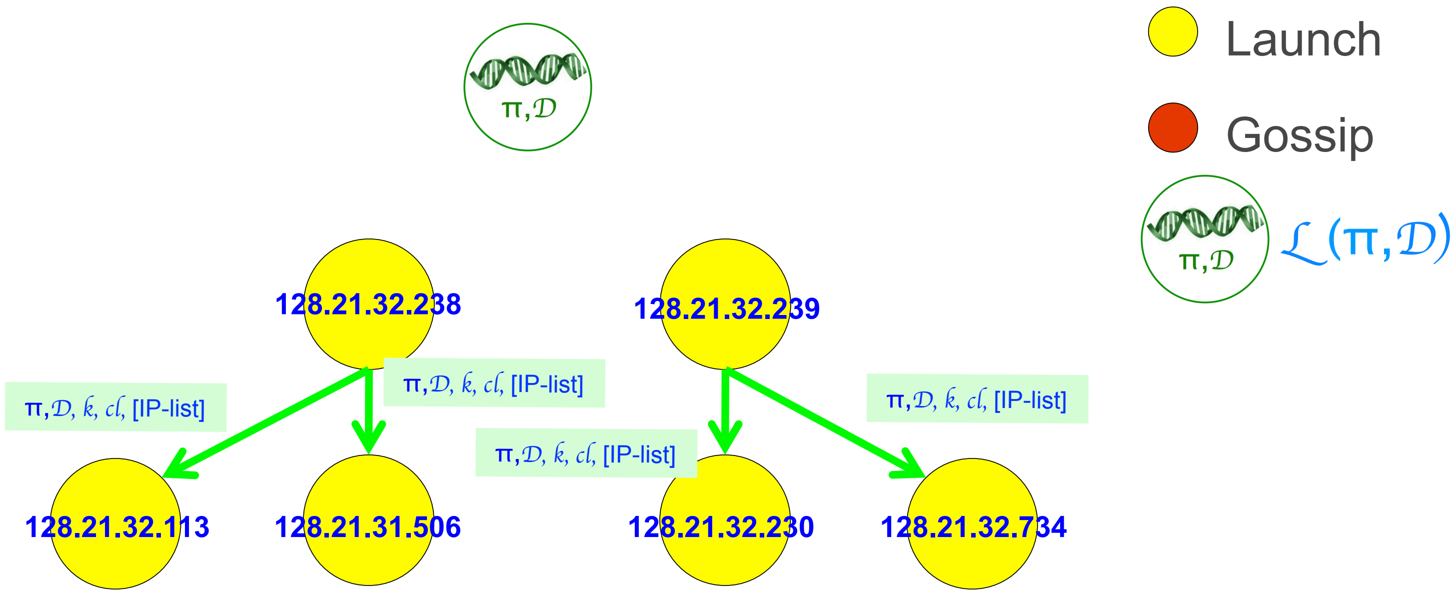
Launch



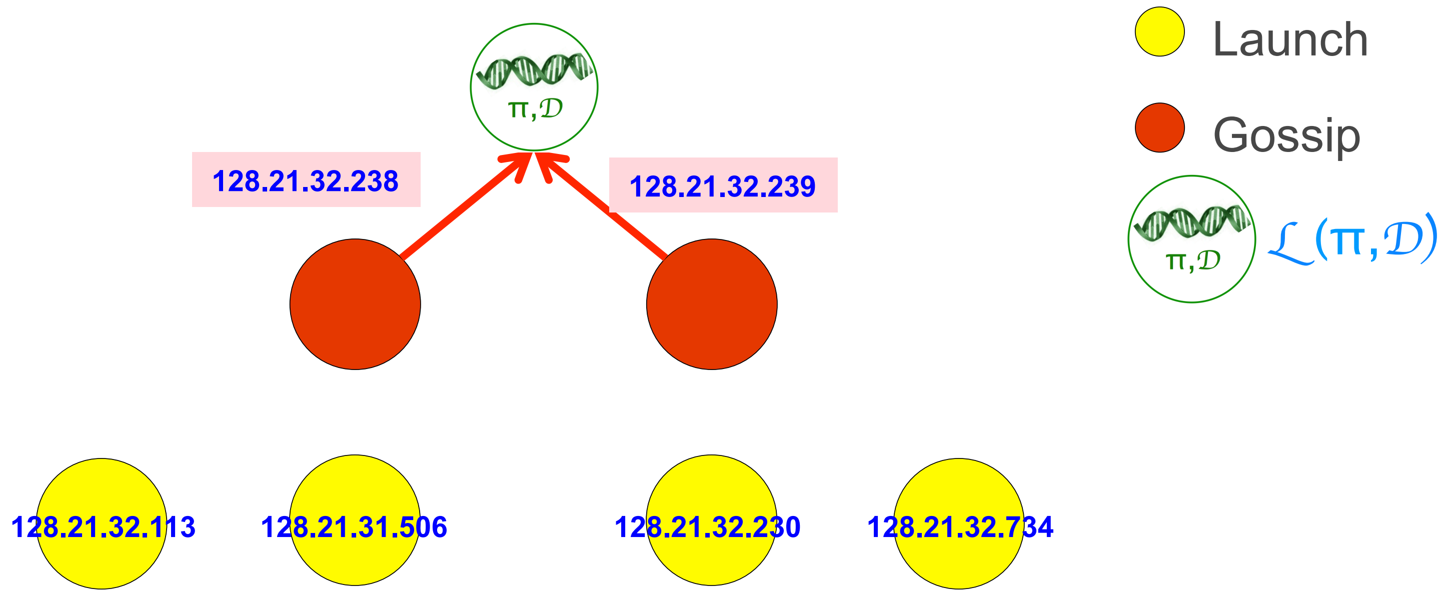
FlexGP Launch



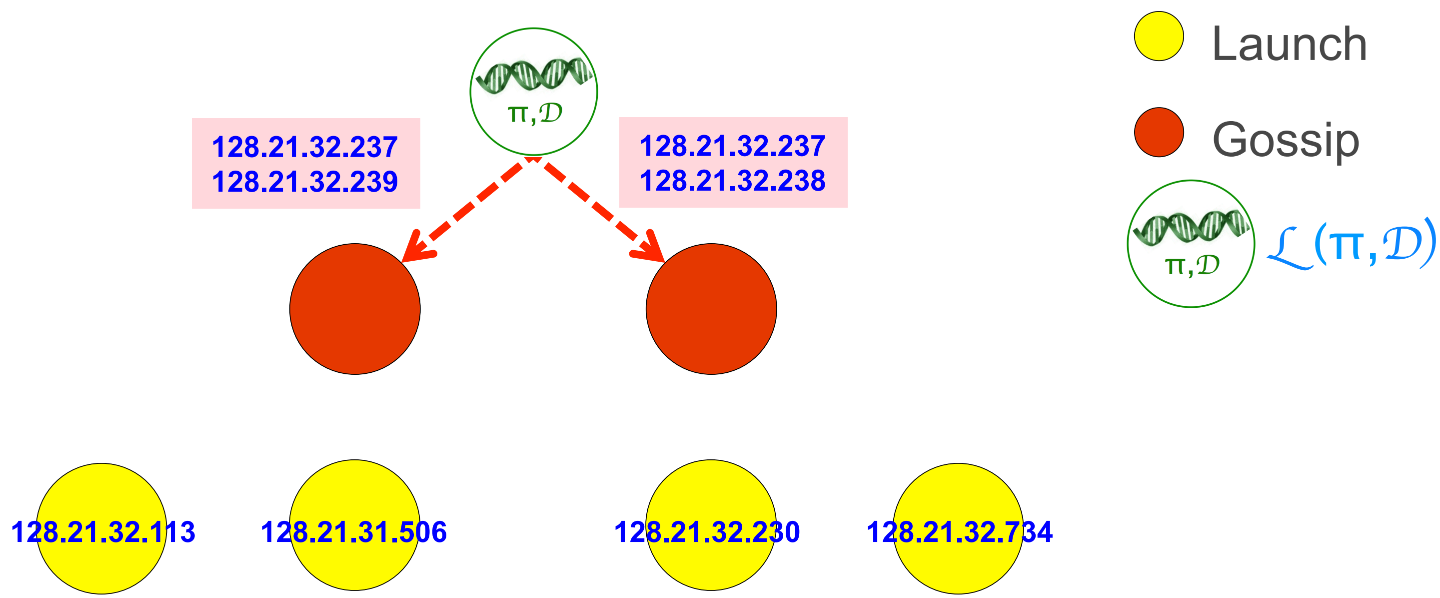
128.21.32.238
128.21.32.239
Launch
Gossip



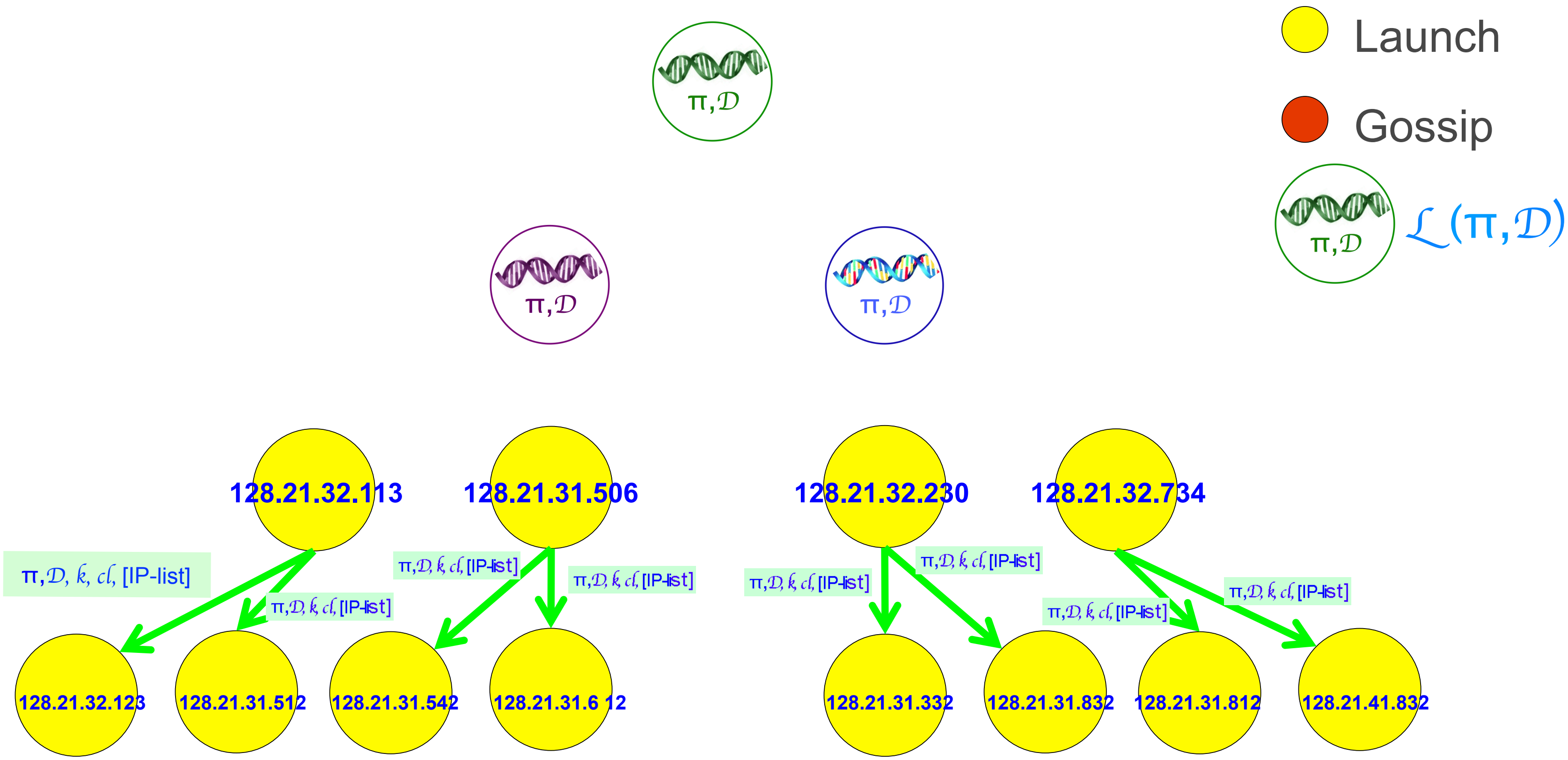
FlexGP Launch



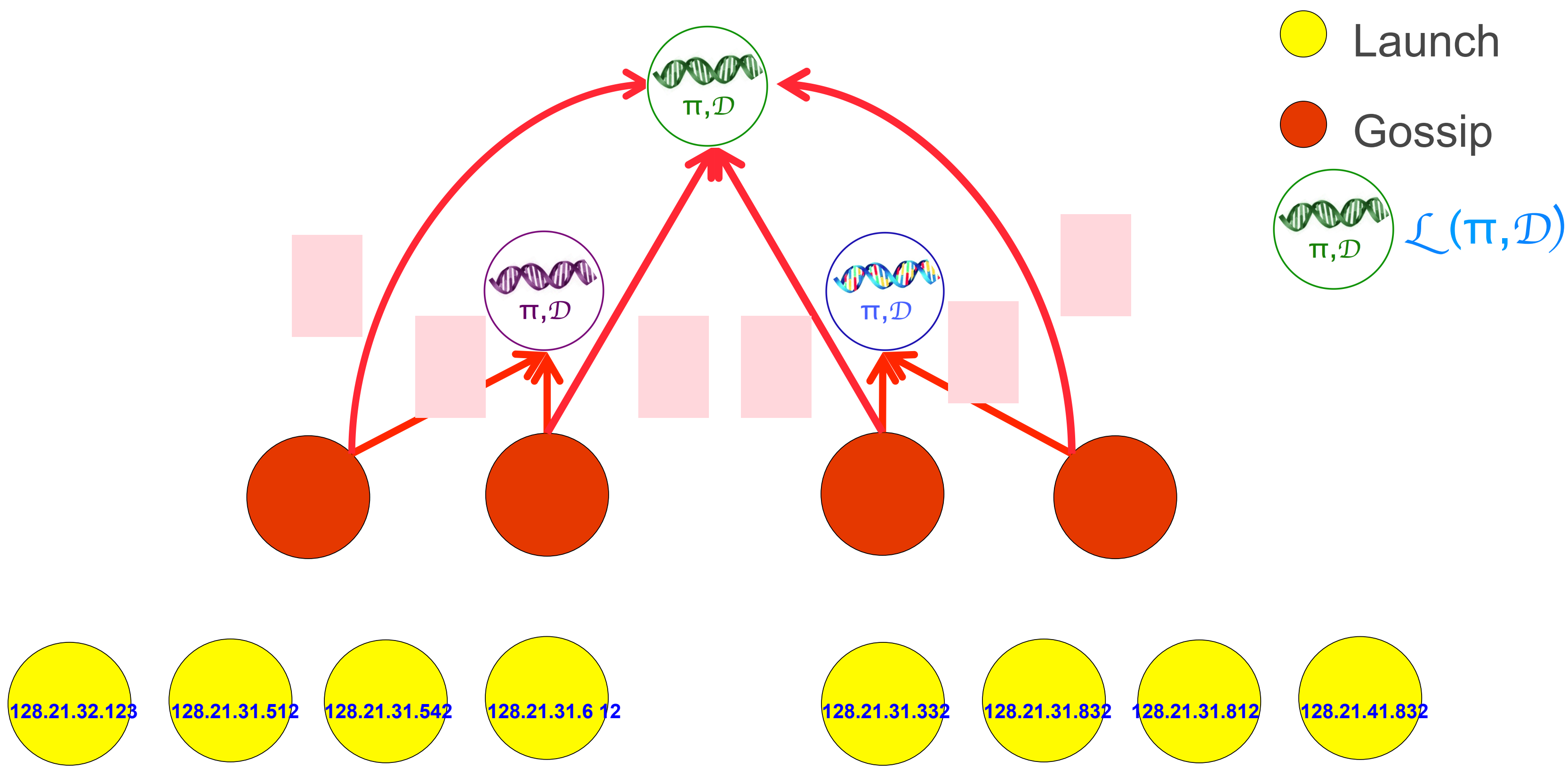
128.21.32.238
128.21.32.239
π,𝒟
128.21.32.113
128.21.31.506
128.21.32.230
128.21.32.734
Launch
Gossip
π,𝒟
ℒ(π,𝒟)



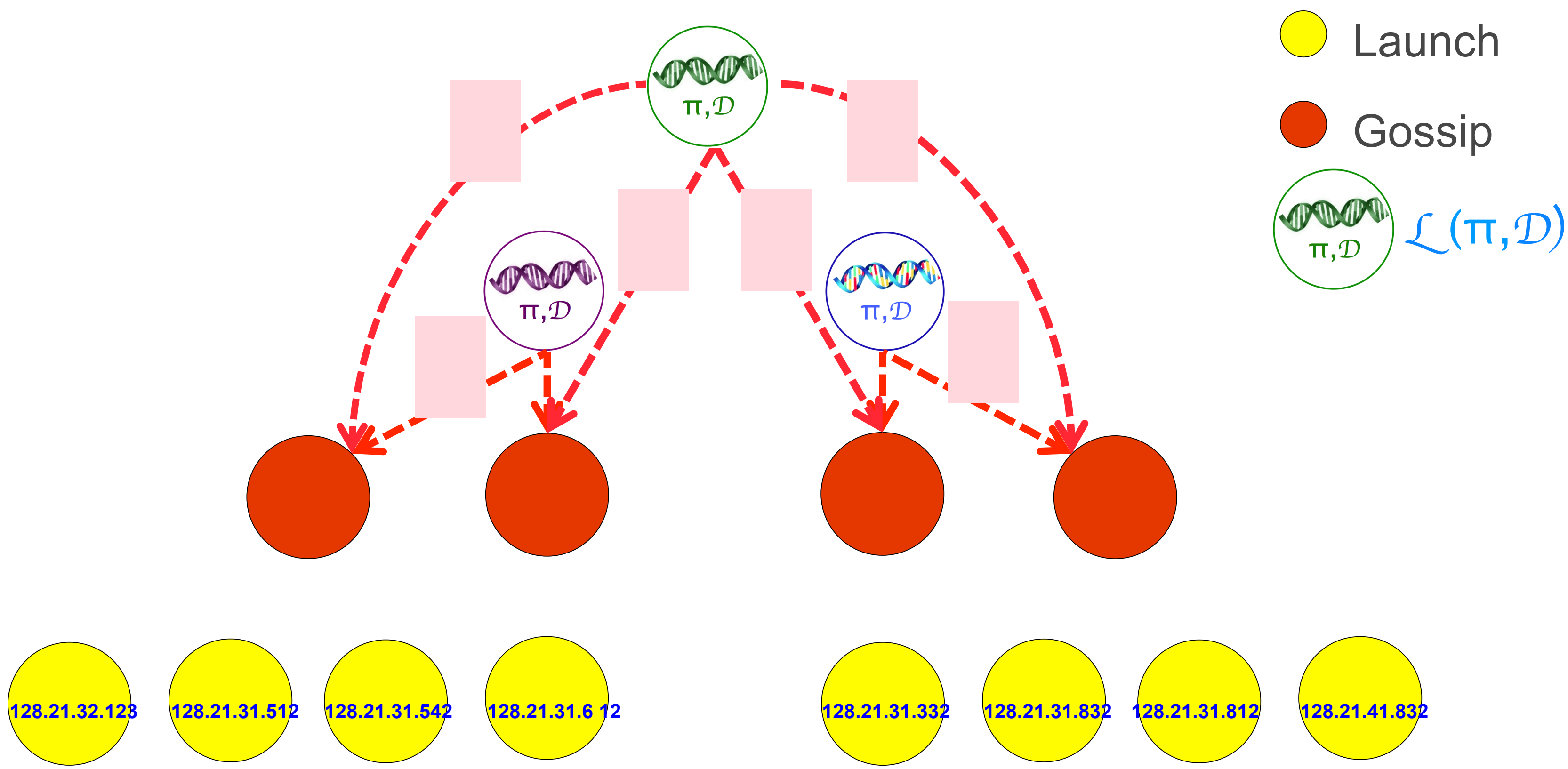
FlexGP Launch



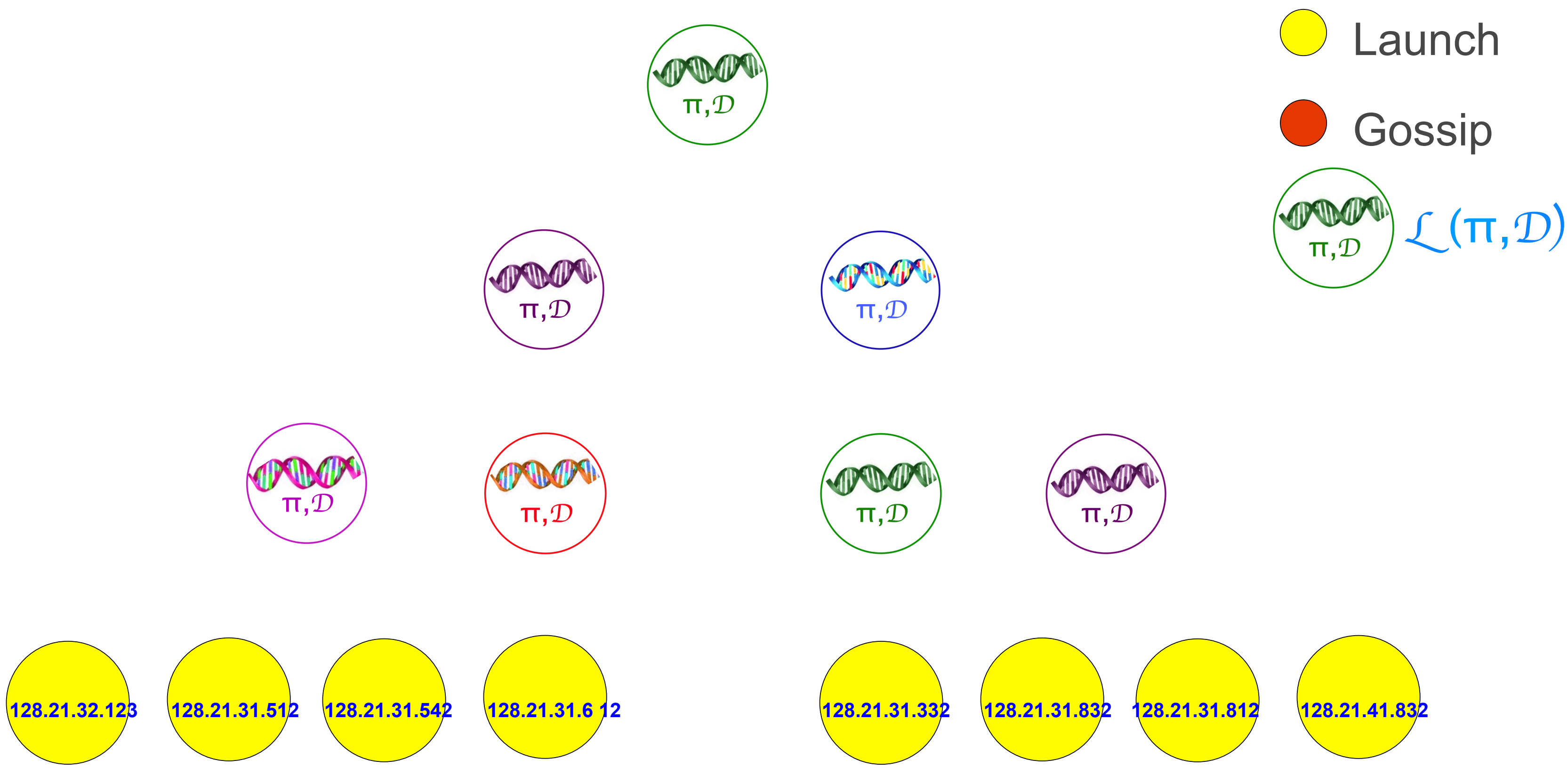
FlexGP Launch



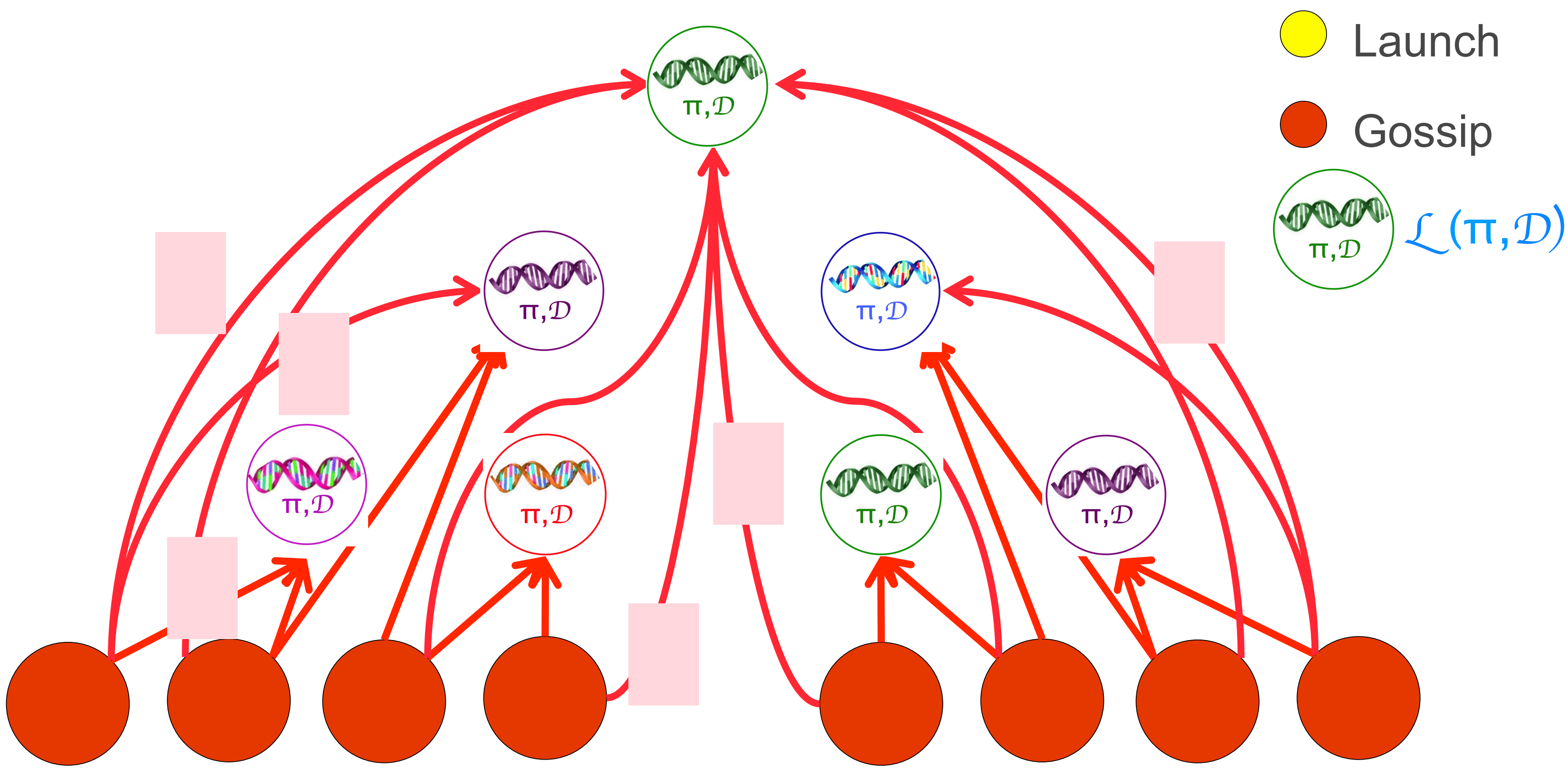
FlexGP Launch



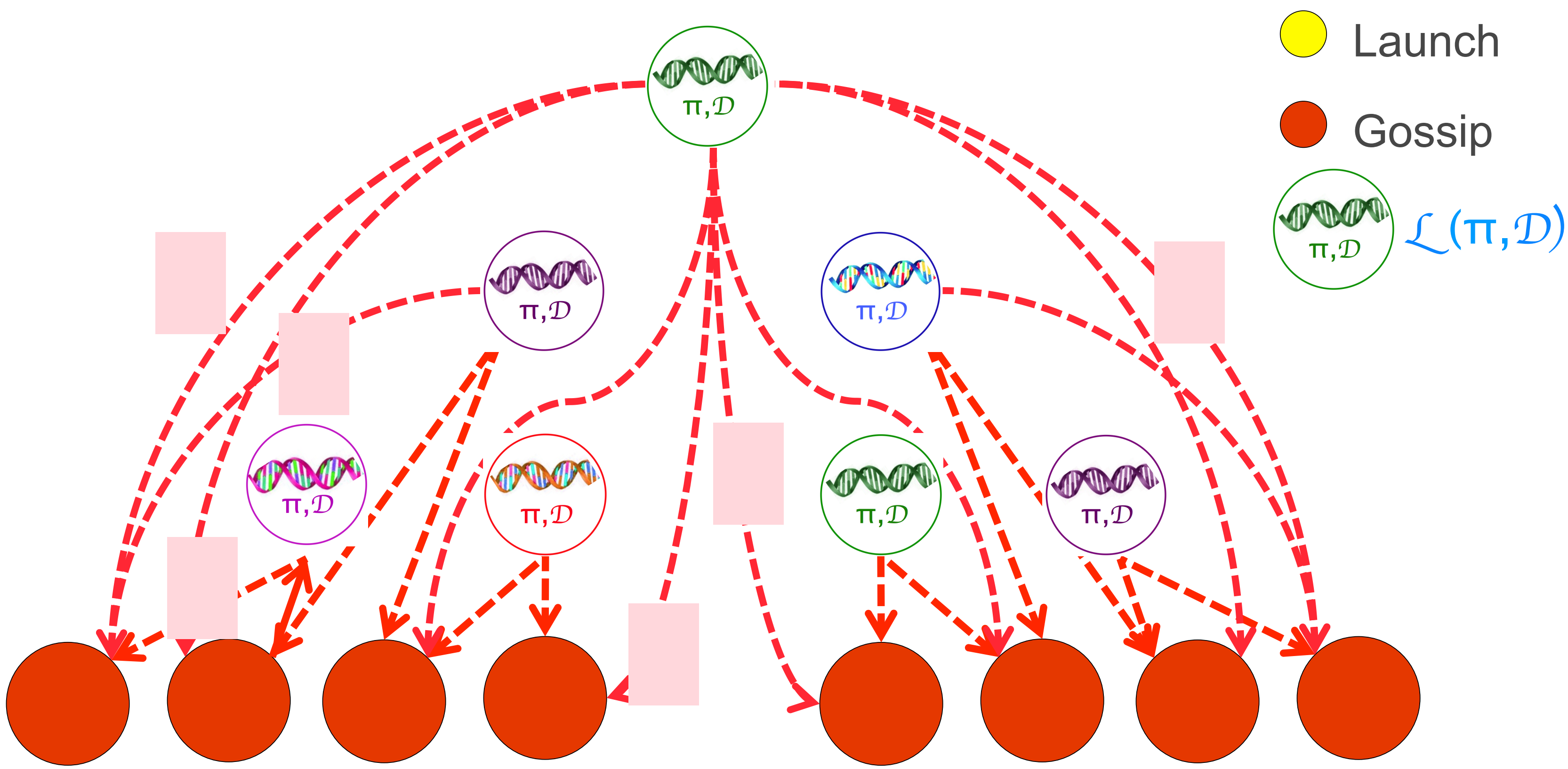
FlexGP Launch



FlexGP Launch



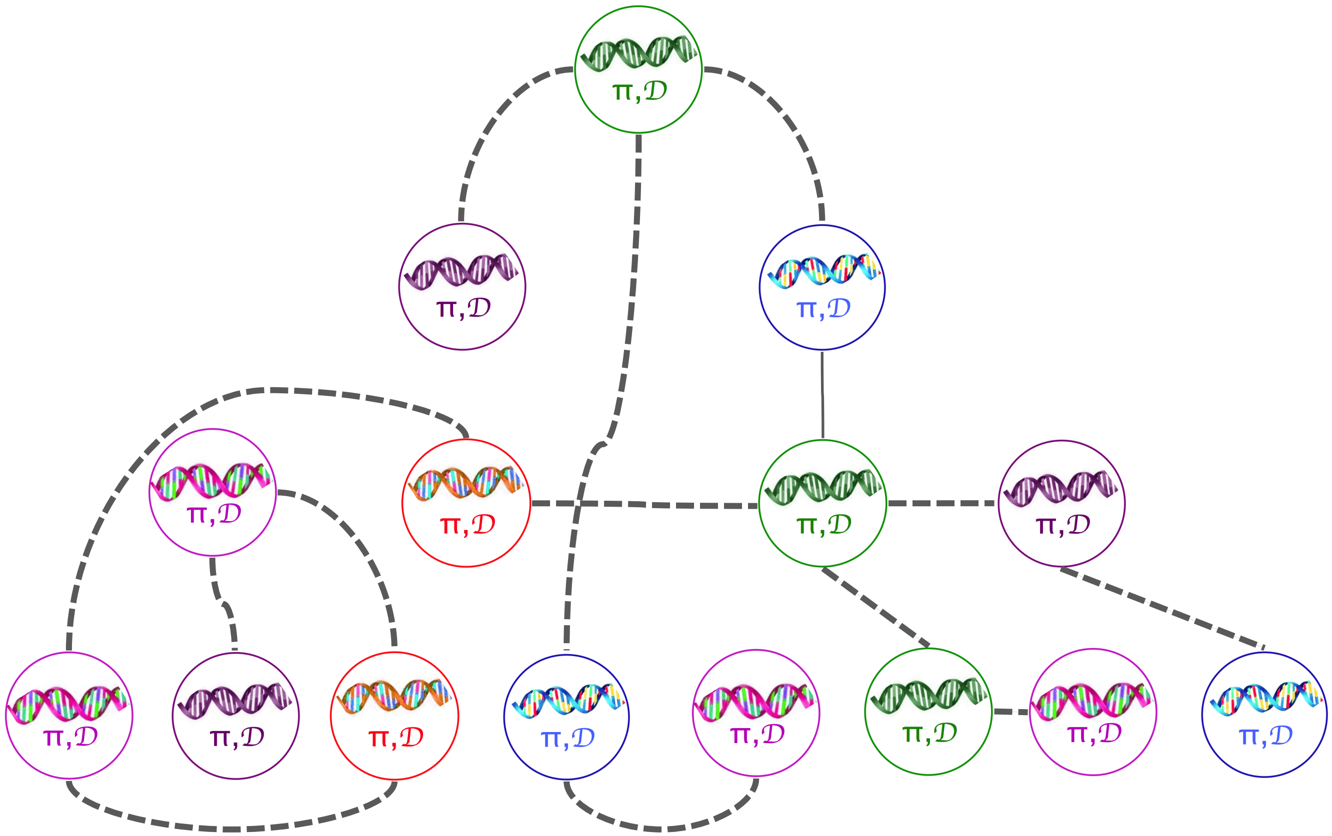
FlexGP Launch



FlexGP Launch

Launch complete!

… and ready to expand or contract
(gossiping intermittently)

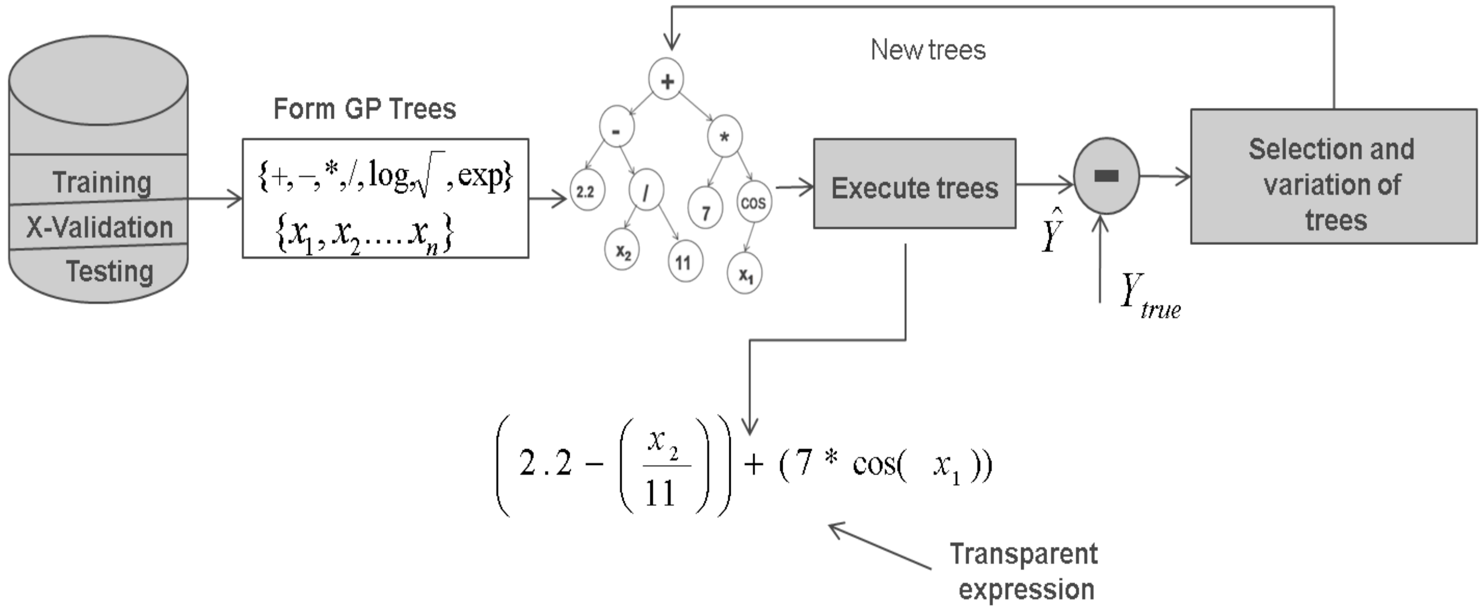


EvoDesignOpt

CSAIL

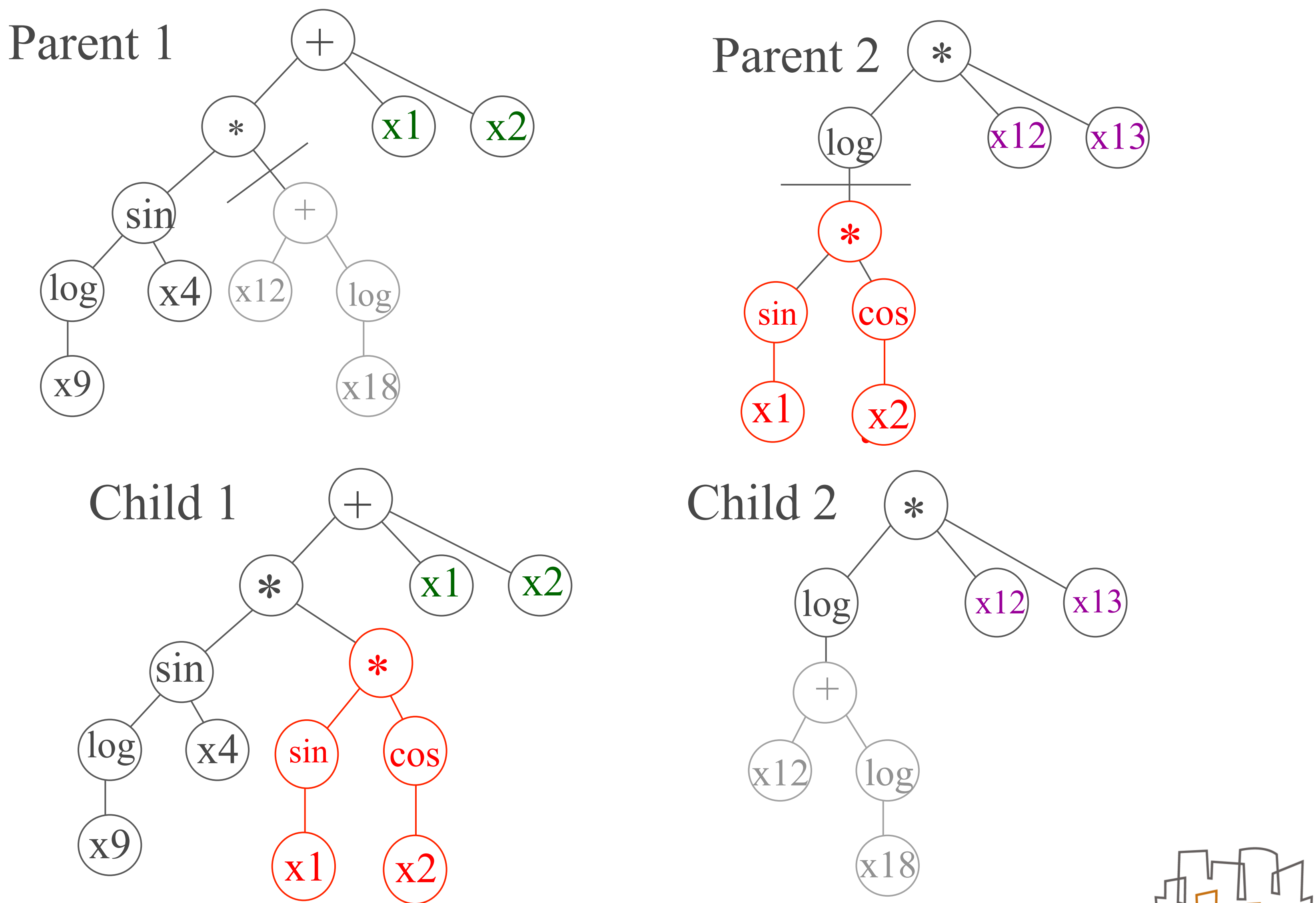
Genetic Programming

Goal: Model $y = f(x_1, x_2, \ldots x_n)$

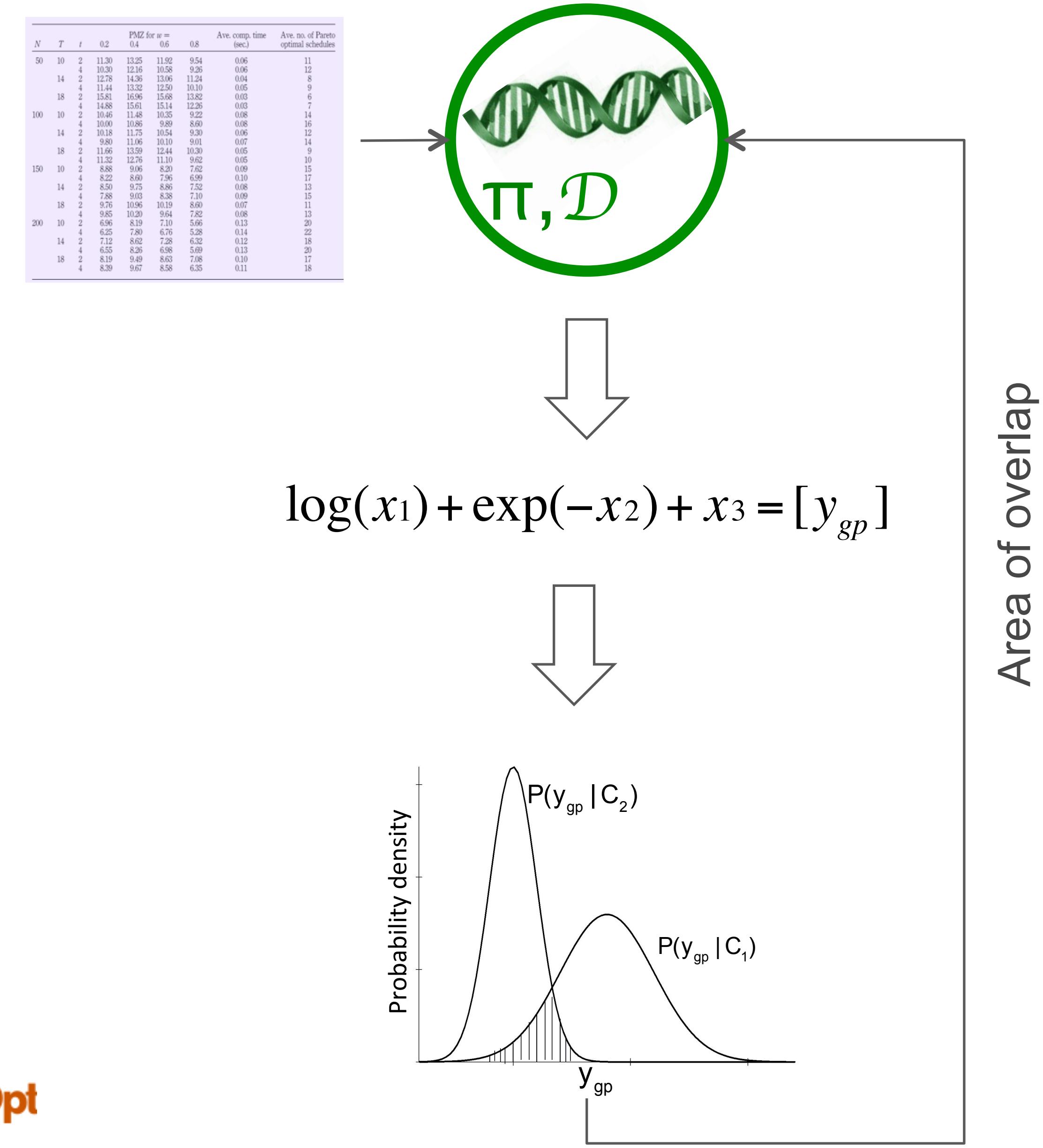


GP Learning Engine

GP Tree Crossover



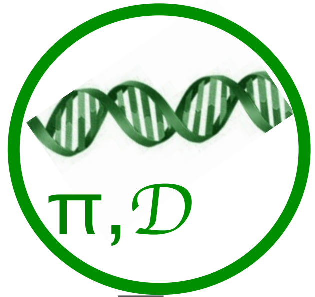
Learning a classifier



EvoDesignOpt

CSAIL

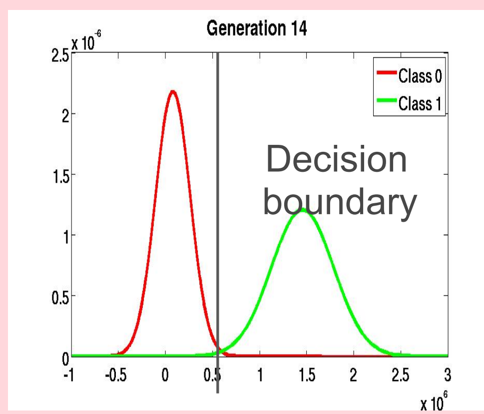
Learning a classifier



Final best model

$$x_1^2 + \exp(-x_2) + \log(x_3) = [y_{gp}]$$

Best classifier



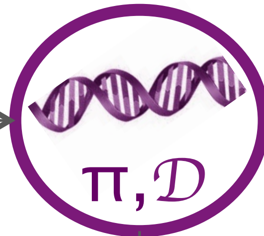
y_{gp}

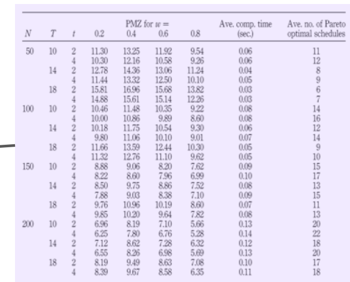
Likelihood ratio test

EvoDesignOpt

CSAIL

Learning a regression model





N	T	t	PMZ for w = 0.2	0.4	0.6	0.8	Ave. comp. time (sec.)	Ave. no. of Pareto optimal schedules
50	10	2	11.30	13.25	11.92	9.54	0.06	11
		4	10.30	12.16	10.58	9.26	0.06	12
	14	2	12.78	14.36	13.06	11.24	0.04	8
		4	11.44	13.32	12.50	10.10	0.05	9
	18	2	15.81	16.96	15.68	13.82	0.03	6
		4	14.88	15.61	15.14	12.26	0.03	7
100	10	2	10.46	11.48	10.35	9.22	0.08	14
		4	10.00	10.86	9.89	8.60	0.08	16
	14	2	10.18	11.75	10.54	9.30	0.06	12
		4	9.80	11.06	10.10	9.01	0.07	14
	18	2	11.66	13.59	12.44	10.30	0.05	9
		4	11.32	12.76	11.10	9.62	0.05	10
150	10	2	8.88	9.06	8.20	7.62	0.09	15
		4	8.22	8.60	7.96	6.99	0.10	17
	14	2	8.50	9.75	8.86	7.52	0.08	13
		4	7.88	9.03	8.38	7.10	0.09	15
	18	2	9.76	10.96	10.19	8.60	0.07	11
		4	9.85	10.20	9.64	7.82	0.08	13
200	10	2	6.96	8.19	7.10	5.66	0.13	20
		4	6.25	7.80	6.76	5.28	0.14	22
	14	2	7.12	8.62	7.28	6.32	0.12	18
		4	6.55	8.26	6.98	5.69	0.13	20
	18	2	8.19	9.49	8.63	7.08	0.10	17
		4	8.39	9.67	8.58	6.35	0.11	18

Minimize p-norm

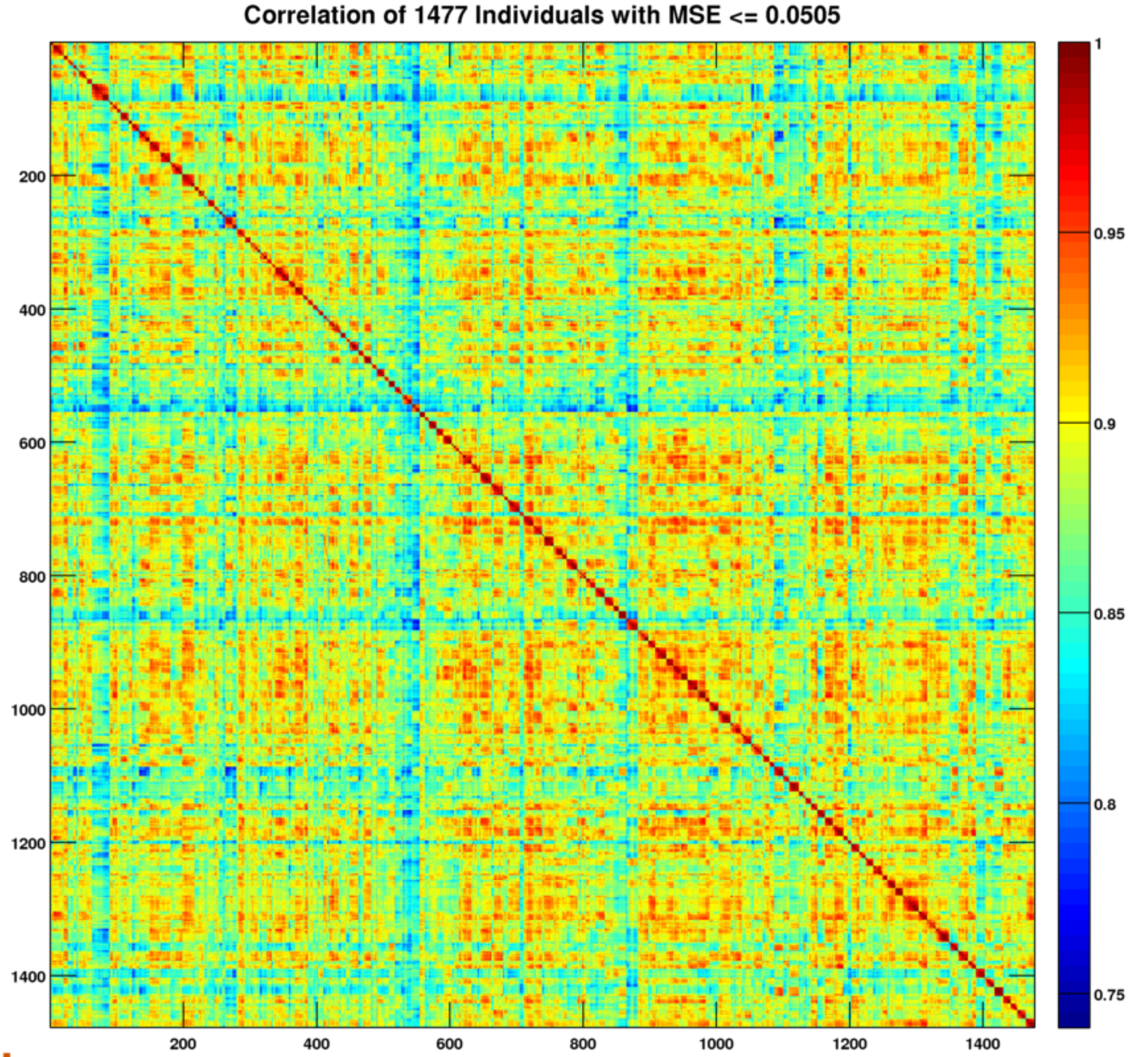
$$\log(x_1) + \exp(-x_2) + x_3 = [y_{gp}]$$

$$\hat{y}_s = \frac{y_{gp} - y_{gp\,min}}{y_{gp\,max} - y_{gp\,min}}$$

Scaling model
via linear regression

$$\left\| \hat{y}_s - y_s \right\|_p$$

FlexGP Regression Model Diversity



EvoDesignOpt

CSAIL

FlexGP…

Is:

Flexibly factored, aggregating ML system

- **Cascading launch**
- **Distributed scalable network protocol**
- **Cloud scale ensemble learning method**

Delivers:

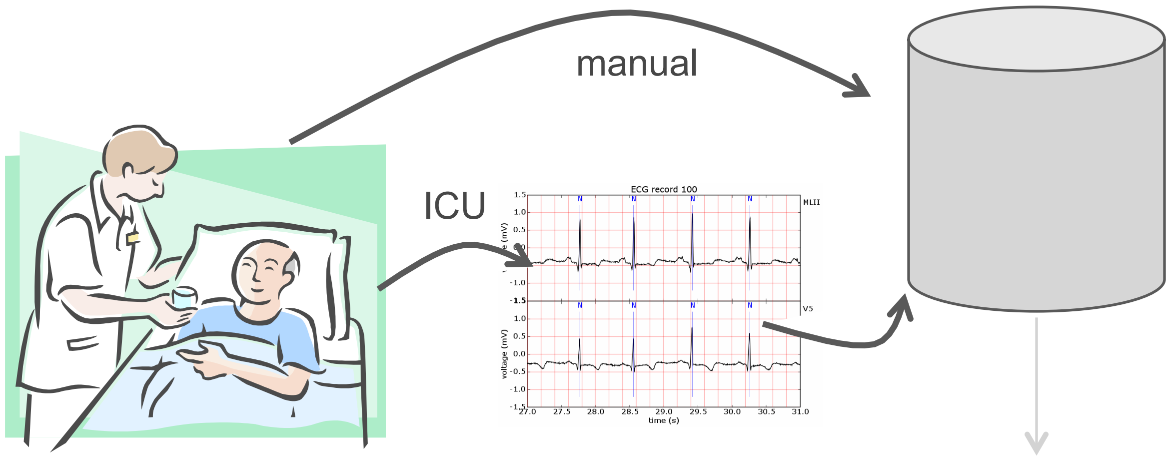
- **Elasticity**
- **Scalability in computation size**
- **Large data strategy**
- **Innovation in machine learning with evolutionary computation**

Automation

- **"In the end, the biggest bottleneck is not data or CPU cycles, but human cycles."**

Mass Customized Query Serving

Waveform database

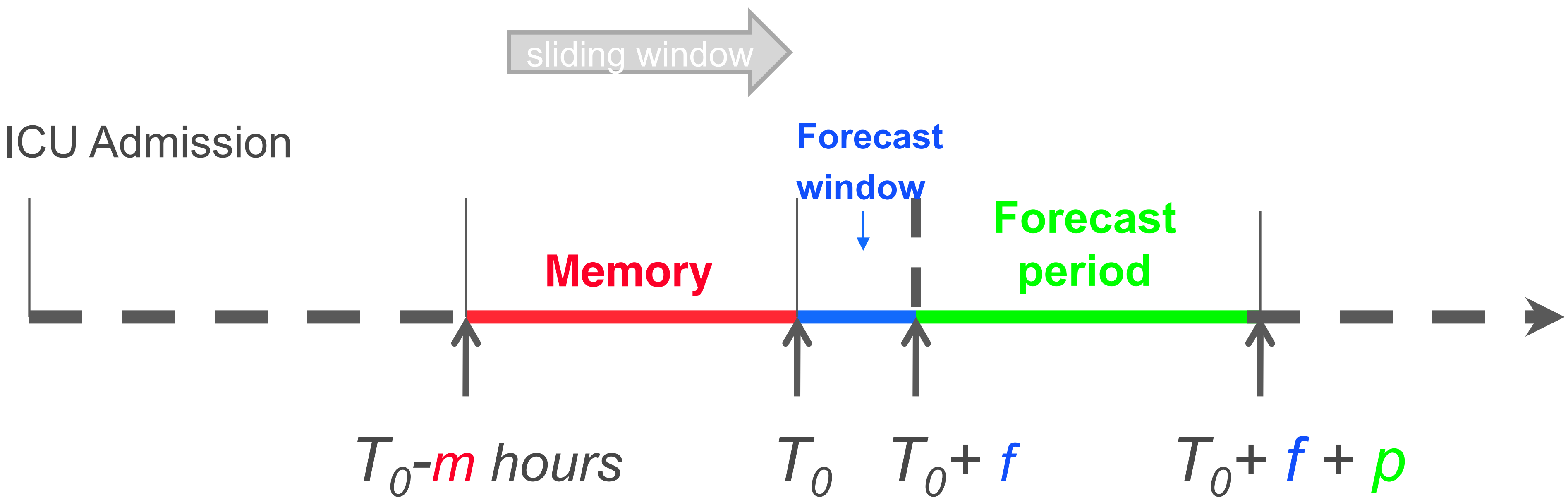


Feature	Value
Total size	4 Terra Bytes
Waveform types	22
Signal sampling frequency	125 samples/sec
Number of samples	500m

Personalized Query Serving

Parameterizations :

***m*- hours of past data used to forecast**
***f* – forecast window, lag**
***p*- period of forecast**



Fundamental Learning

- **When the data overwhelms us…**
  - **We bundle it up**
    - » **nb, this is not sampling!**
  - **We assume linearity and Gaussian distributions**

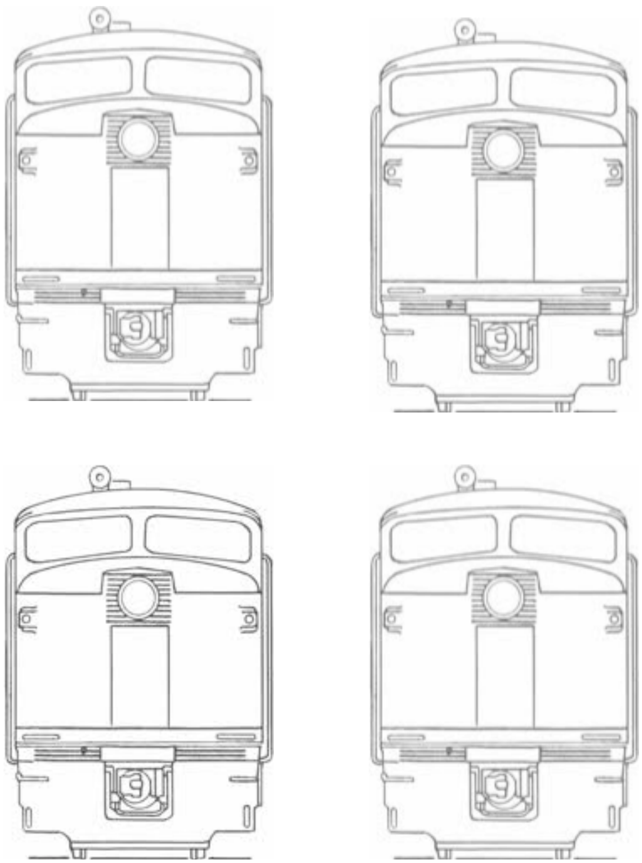
- **What are the intrinsic aggregations?**
- **What are the non-linearities and true distributions?**

- **Fundamental learning starts from the bottom up**
  - **Use unsupervised learning to propose features**
  - **Use features in a task**
  - **Pass performance feedback to feature learning**

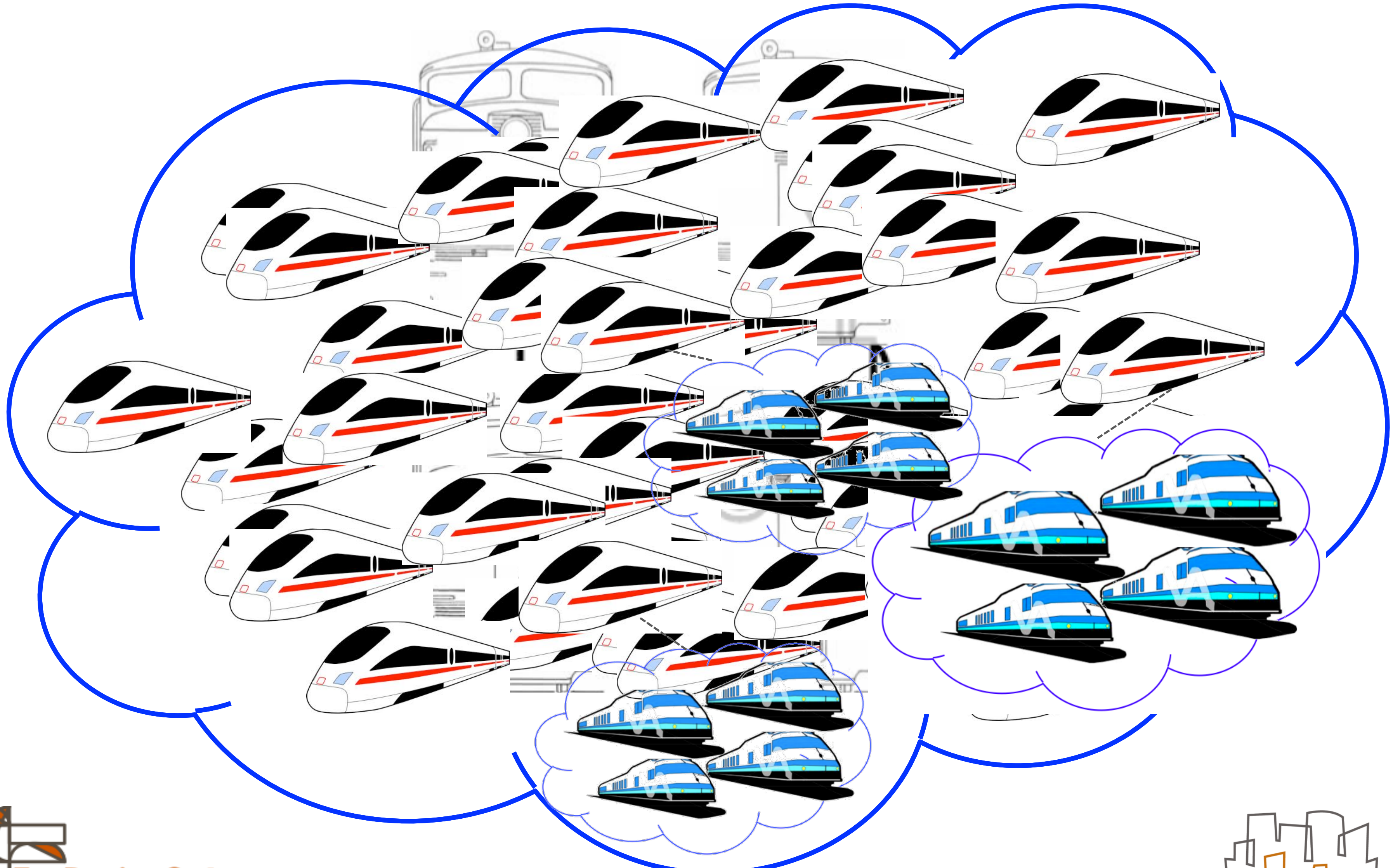
Requires BigData and Cloud-Scale ML

A time trajectory of GP-based machine learning

A time trajectory of GP-based machine learning



A time trajectory of GP-based machine learning



EvoDesignOpt

CSAIL

Acknowledgements

- **Members of the Evolutionary Design and Optimization Group**
  - **Past and present**
  - **Dr. Kalyan Veeramachaneni: Research Scientist**

GE Global Research

Industrial Machine Learning Lab, GEGR, Niskayuna

EvoDesignOpt

CSAIL