Accelerating Evolutionary Algorithms to Solve High-dimensional Expensive Problems via Autoencoders

MengChu Zhou, Dist. Professor
Fellow of IEEE, IFAC, AAAS, CAA & NAI
New Jersey Institute of Technology
zhou@njit.edu

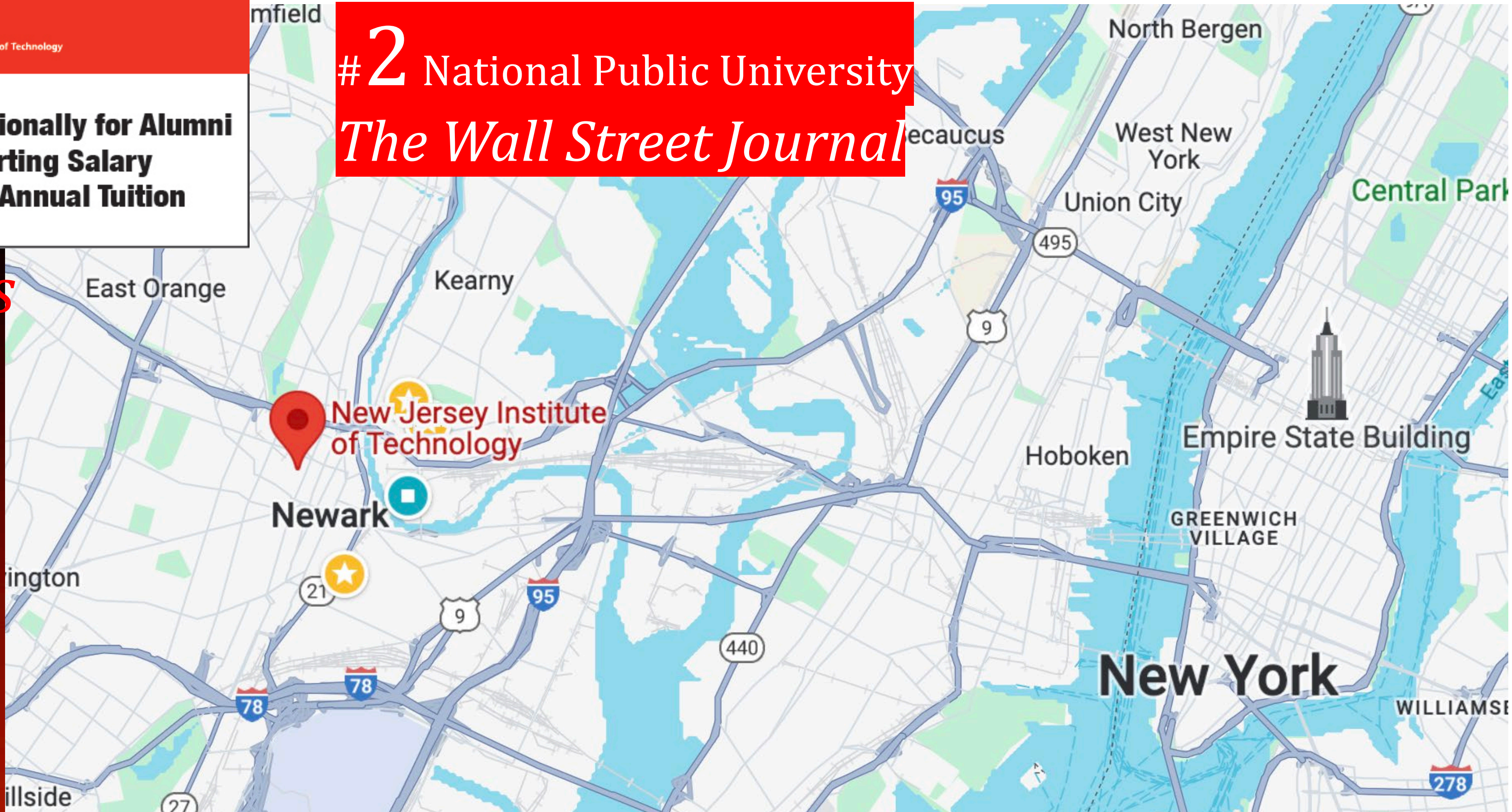
NJIT New Jersey Institute of Technology

 # 1 Nationally for Alumni Starting Salary vs. Annual Tuition

Forbes

2 National Public University

The Wall Street Journal



Top #71 in Best Universities for Electrical and Electronic Engineering Globally (#9 in the USA)

U.S. News & World Report

Proposals for Special Sessions
January 21, 2024

Workshop and Tutorial Proposals
April 8, 2024

Regular, Special Session, Industrial, BMI Workshop Papers
April 8, 2024

NJIT
New Jersey's Science & Technology University

Special Issue on Internet of Things, Internet of Behaviors, and Industry 5.0

Call for Papers

Taylor & Francis
Taylor & Francis Group

International Journal of Production Research

Internet of Things, Internet of Behaviors and Industry 5.0

https://www.callforpapers.co.uk/internet-of-behaviors

VIEW JOURNAL & SUBMIT

Guest Editors

- Mengchu Zhou, New Jersey Institute of Technology, USA. Email: zhou@njit.edu
- Yixiong Feng, Guizhou University, China
- Qinglin Zhao, Macau University of Science and Technology, China
- Maria Pia Fanti, Polytechnic of Bari, Italy

Important Dates

- Submission: June 1, 2024
- Revision: Aug 1, 2024
- Final: Oct. 1, 2024
- Publication: Jan 2025

Main Topics

- •Emerging methods, paradigms and architecture in IoT and IoB for Industry 5.0
- •IoT/IoB-enabled efficient human-robot interaction and collaboration
- •Intelligent perception methods, architectures and platforms in Industry 5.0
- •Big data analytics and machine learning techniques for intelligent manufacturing
- •Human factors and ergonomics for Industry 5.0
- •Technology and application of human digital twin modelling for Industry 5.0
- •Collaborative design and optimization in Industry 5
- •IoT/IoB-enabled flexible Job shop scheduling
- •Cloud, fog, and edge computing for Industry 5.0
- •Multi-objective optimization methods for Industry 5
- •Supply chain for Industry 5.0
- •IoT/IoB-driven collective decision-making and intelligent control
- •IoT/IoB-driven manufacturing, assembly, disassem and remanufacturing

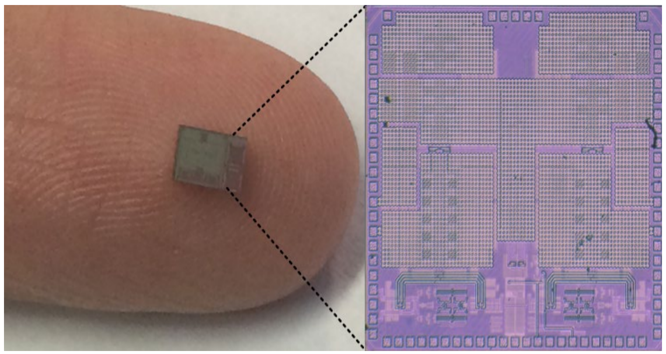
NJIT New Jersey's Science & Technology University

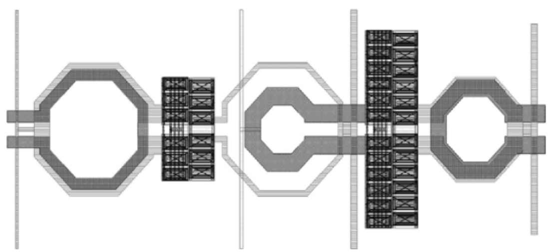
Layout

- Background
- Proposed Algorithms
- Experimental Results
- Future Research
- Conclusions

Background

Background



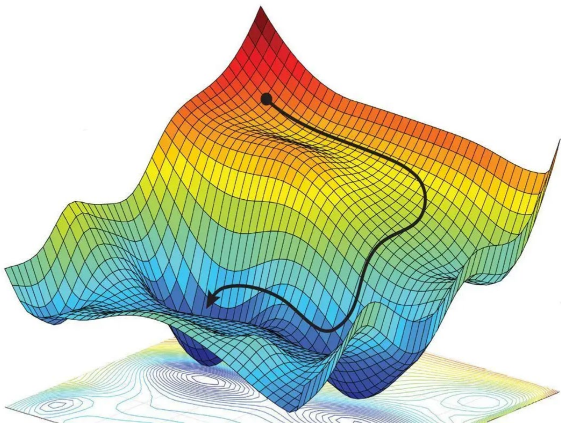


mm-wave integrated circuit design problem

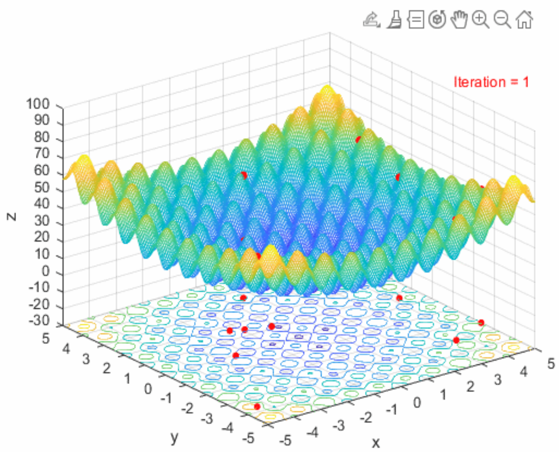
SPACEX related design/ optimization problem

Background

Mathematical Optimization V.S. Evolutionary Optimization



Mathematical optimization



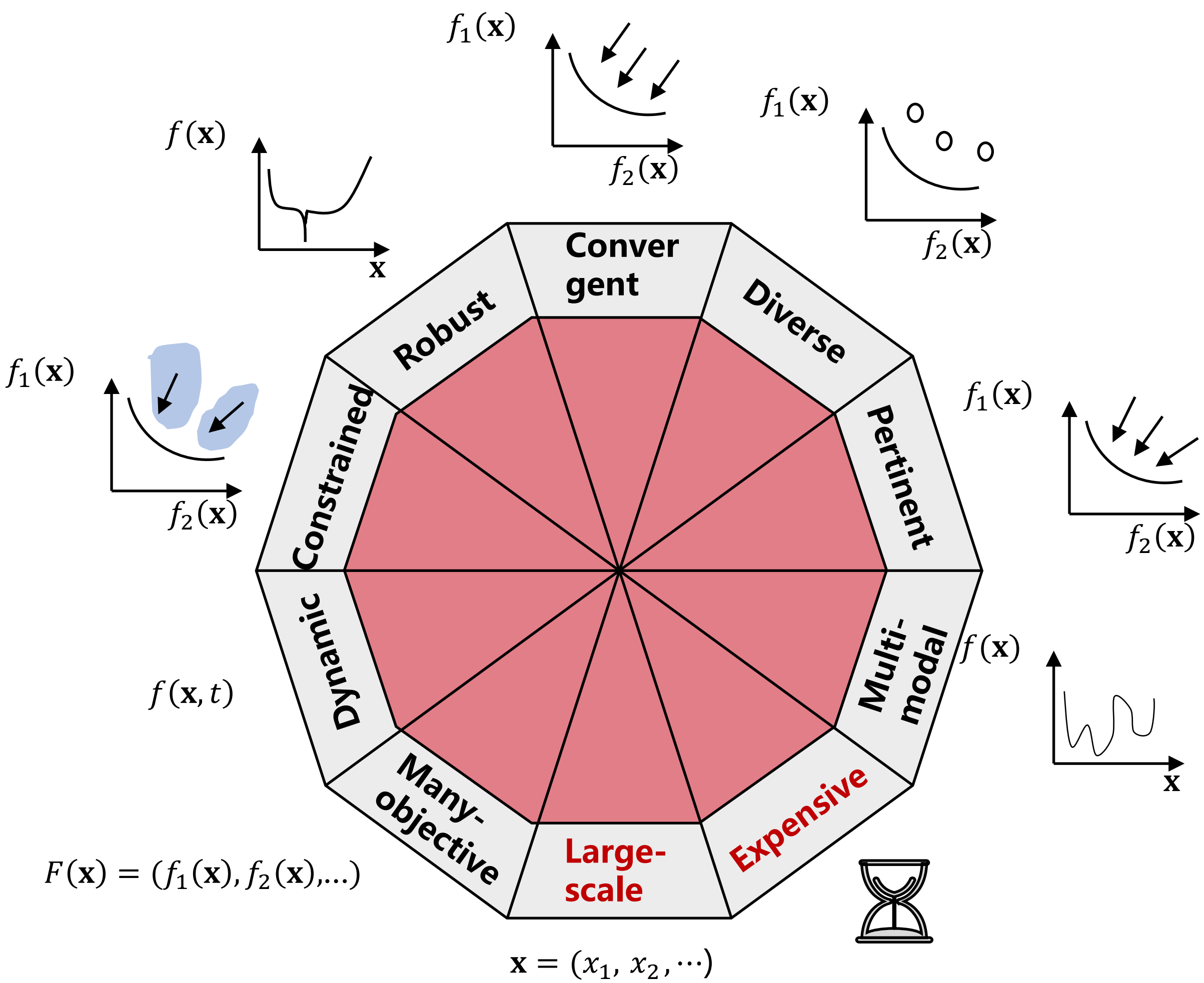
Evolutionary optimization

- Gradient-based method
- Fast convergence speed

- Problem-independent method
- GPU-based parallel computing

Background

Research Directions of Evolutionary Optimization



Background

Large-scale decision variables

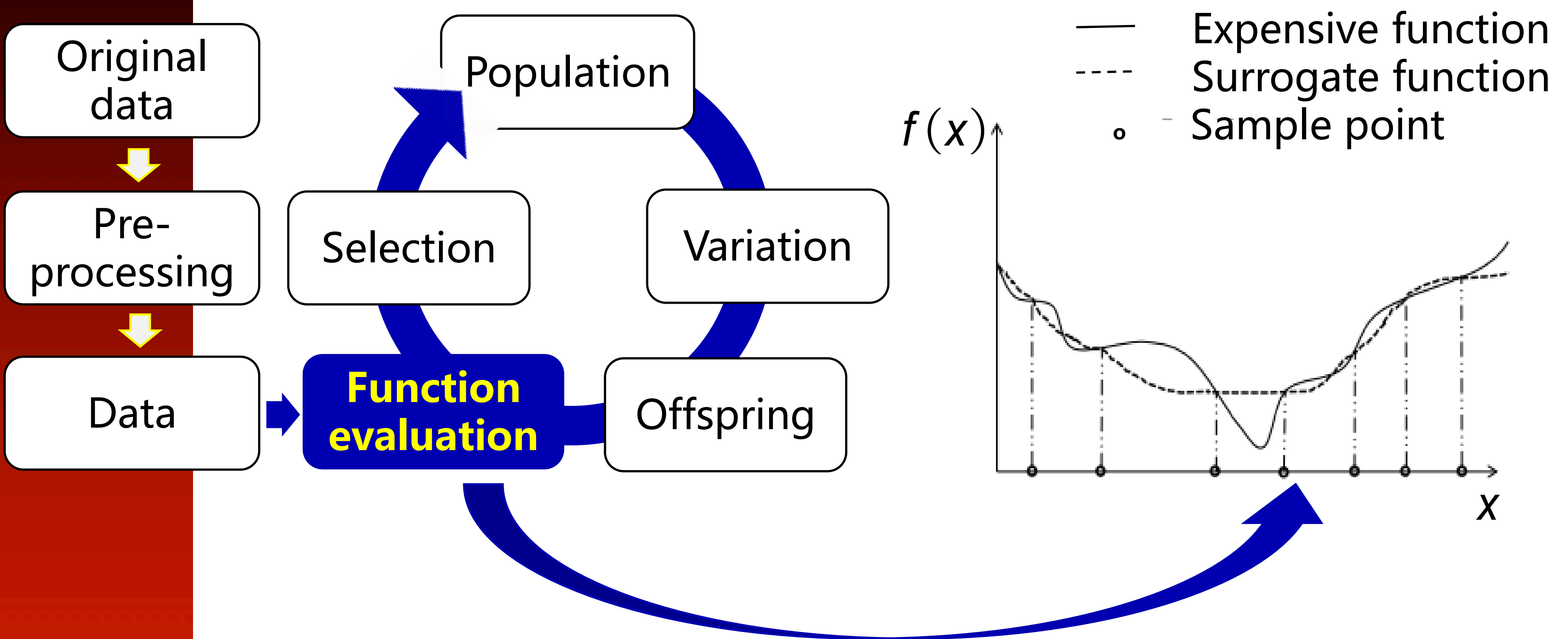
- Expanding search space
- Complex relations among variables
- Changing problem characteristics

Computationally expensive evaluations

- Heavy computing resources
- Intensive physical resources

Background

Surrogate-assisted Evolutionary Algorithms



- Model selection: problem independent
- Model construction: curse of dimensionality
- Model management: non-determinacy

Background

Surrogate-assisted Evolutionary Algorithms

Algorithm	Specific strategy	EA + Surrogate	Max_D
GPEME	DR + Local surrogate	DE + GP	50D
SA-COSO	Global + Local surrogate	PSO, SL-PSO + RBF, FES	200D
SHPSO	Multi-swarm	PSO + RBF	100D
ESAO	Global + Local surrogate	DE + RBF	200D
GSGA	2 Global + 1 Local surrogate	GA + GP, RBF	100D
SAMSO	Multi-swarm, dynamic swarm size adjustment	TLBO+SPSO, RBF + GP	200D

GPEME: Gaussian process surrogate model assisted evolutionary algorithm
SA-COSO: Surrogate-assisted cooperative swarm optimization algorithm
SHPSO: Surrogate-assisted cooperative swarm optimization algorithm
EASO: Evolutionary sampling assisted optimization
GSGA: Generalized surrogate-assisted genetic algorithm
SAMSO: Surrogate-assisted multiswarm optimization

Background

Disadvantages of Surrogate-assisted Optimization

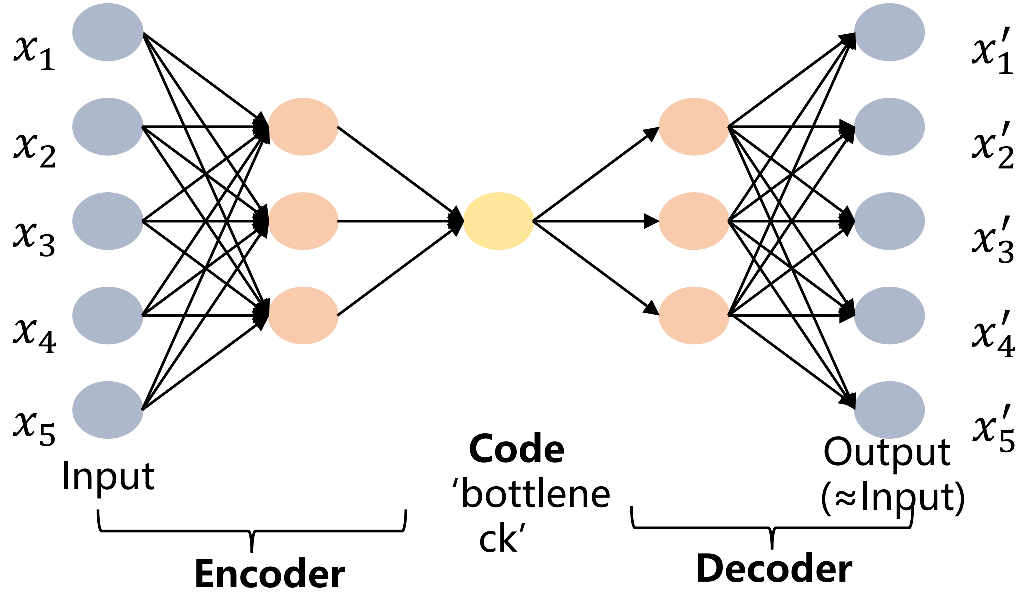
- Low surrogate model accuracy
- High training time complexity

Motivations

- Make full use of expensive fitness evaluation
- Generate offspring in low-dimensional space

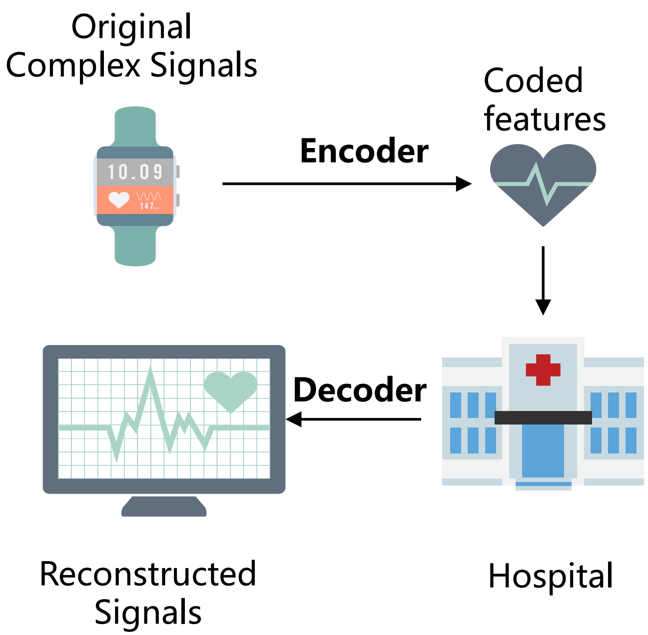
Autoencoder (AE)

- Unsupervised artificial neural networks
- Learn efficient representations of input data
- Reconstruct outputs (≈inputs) from that representations (a restorable process)



Autoencoders

➢ Typical Applications

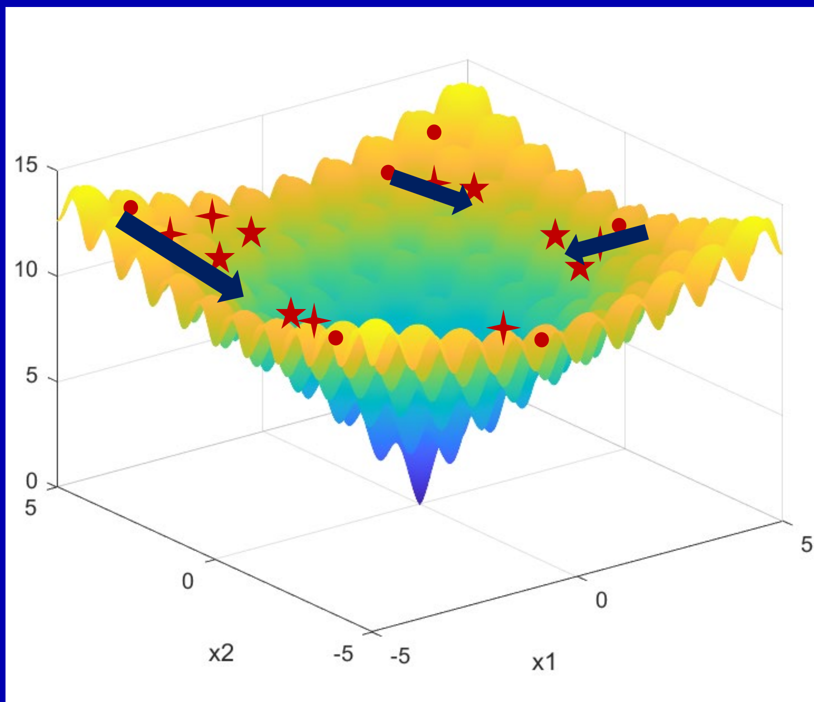


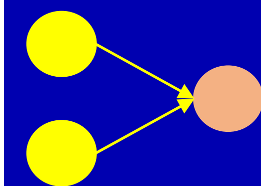
Electrocardiograms (ECG) compression

➢ Advantages

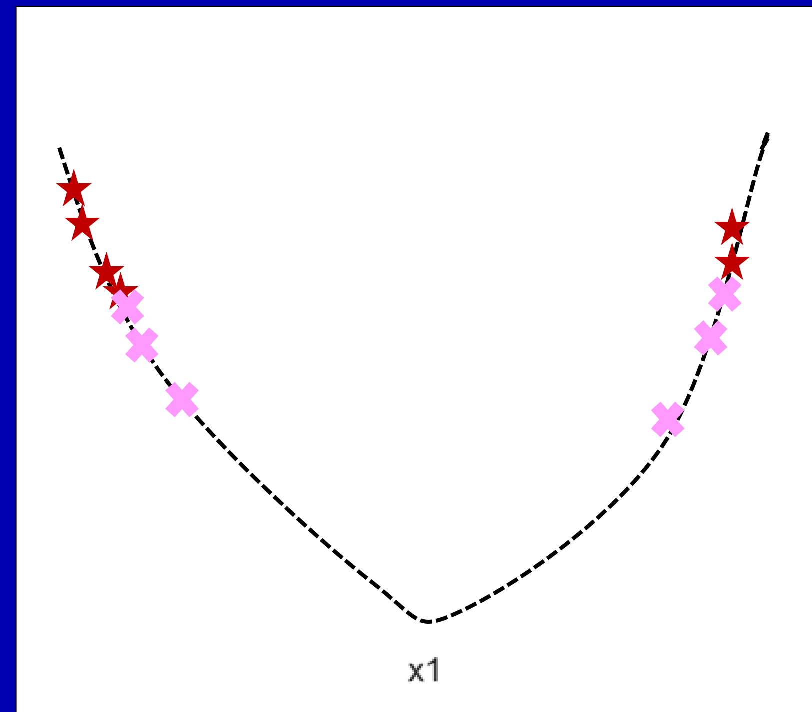
- Suitable to deal with large-scale data
- Restoration ability without extra effort

Role of Autoencoders



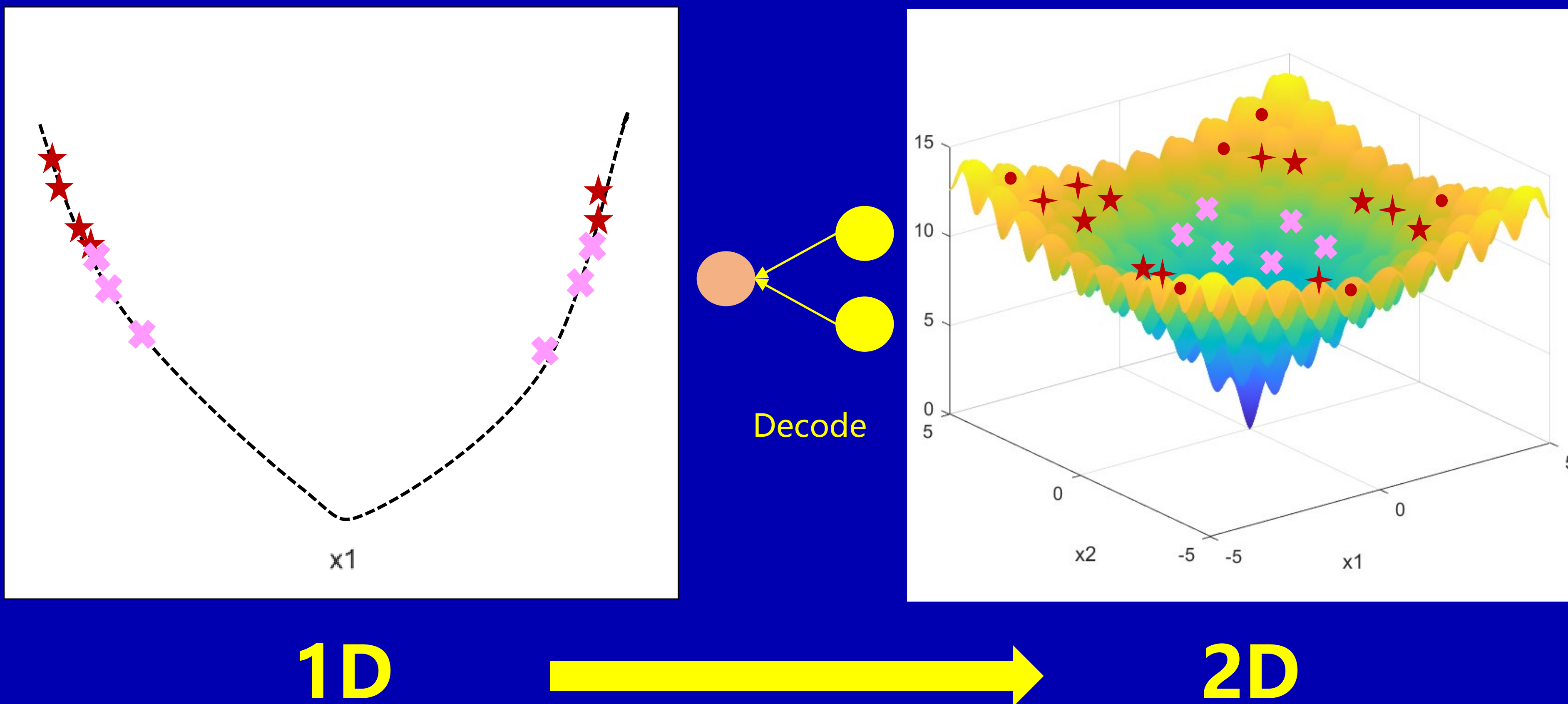


Encode



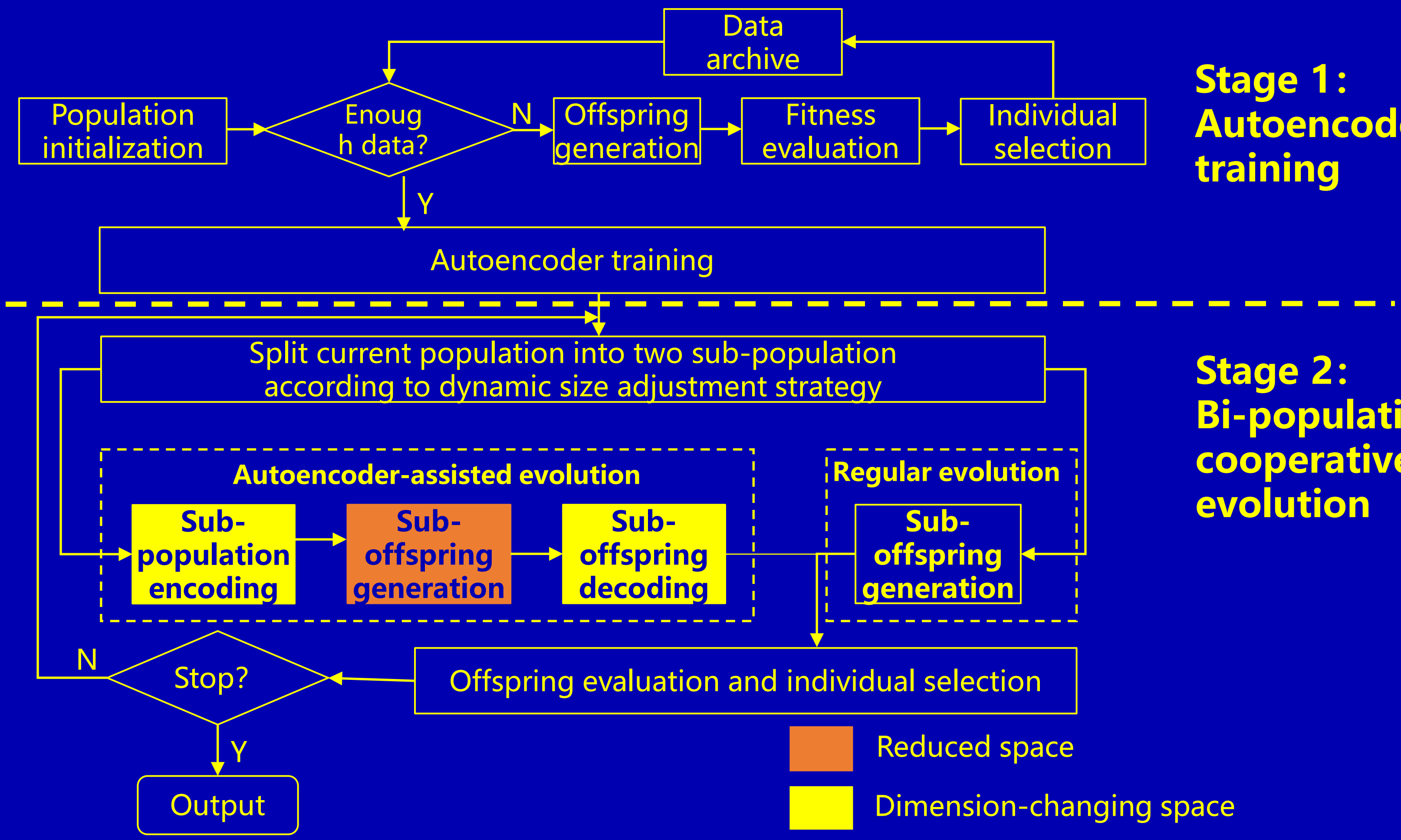
2D → **1D**

Role of Autoencoders



- Learning promising evolution directions
- Compressing search space

Autoencoder-embedded Optimization(AEO)



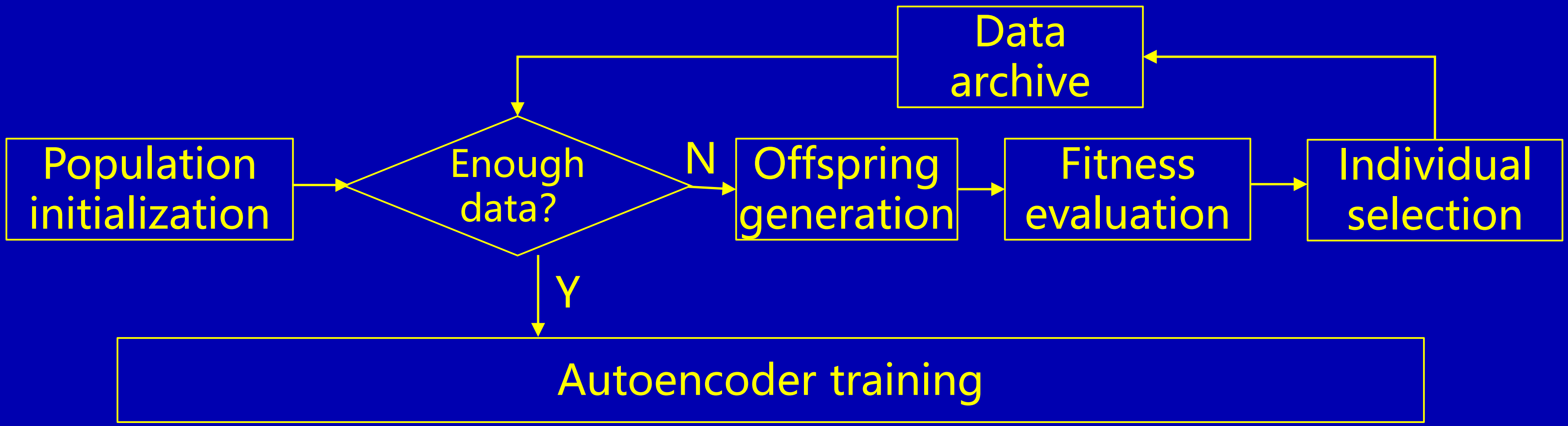
Stage 1: Autoencoder training

Stage 2: Bi-population cooperative evolution

AEO Framework

Autoencoder-embedded Optimization

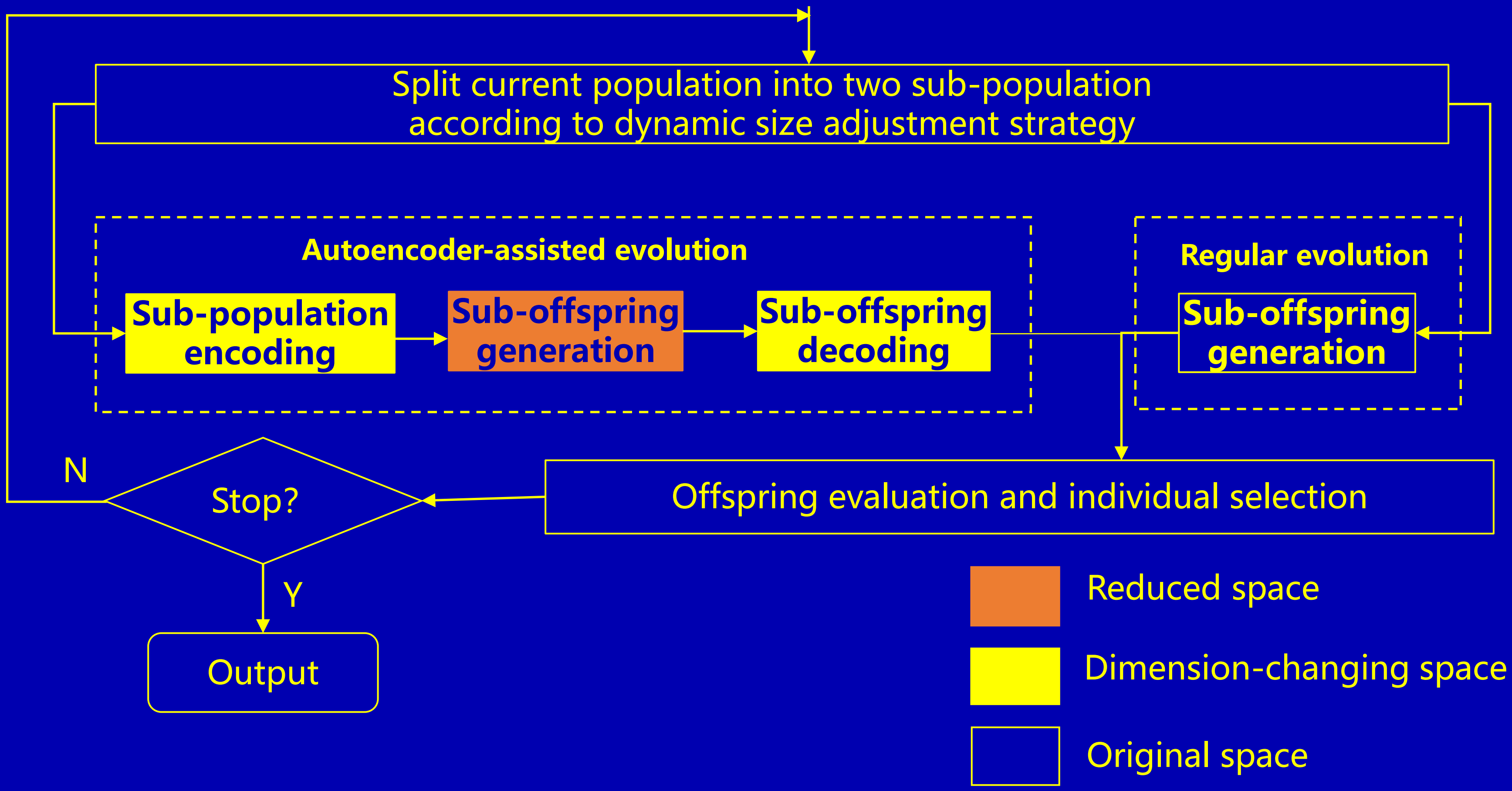
Stage 1: Autoencoder training



- Autoencoder training
  - Compressing search space
  - Learning promising evolution directions
  - Possessing restoration ability

Autoencoder-embedded Optimization

Stage 2: Bi-population co-operation evolution



- Bi-population cooperative evolution
  - Offspring generated in different dimensional spaces
- Dynamic population information exchange

Autoencoder-embedded Evolutionary Optimization

- ✓Finding high-quality solutions in a short time
- ✓Autoencoder for dimension reduction
- ✓Swarm algorithm as a baseline optimizer, e.g., Teaching-learning-based optimizer (TLBO) and Gray Wolf Optimizer (GWO).
- ✓Bi-population coevolution

Two-phase Teaching-learning-based Optimizer (TTLBO)

- ✓ The process of teaching and learning
- ✓ Less parameters and fast convergence

Teaching Phase (TLBO-T): Global exploration

$$\mathbf{x}$$

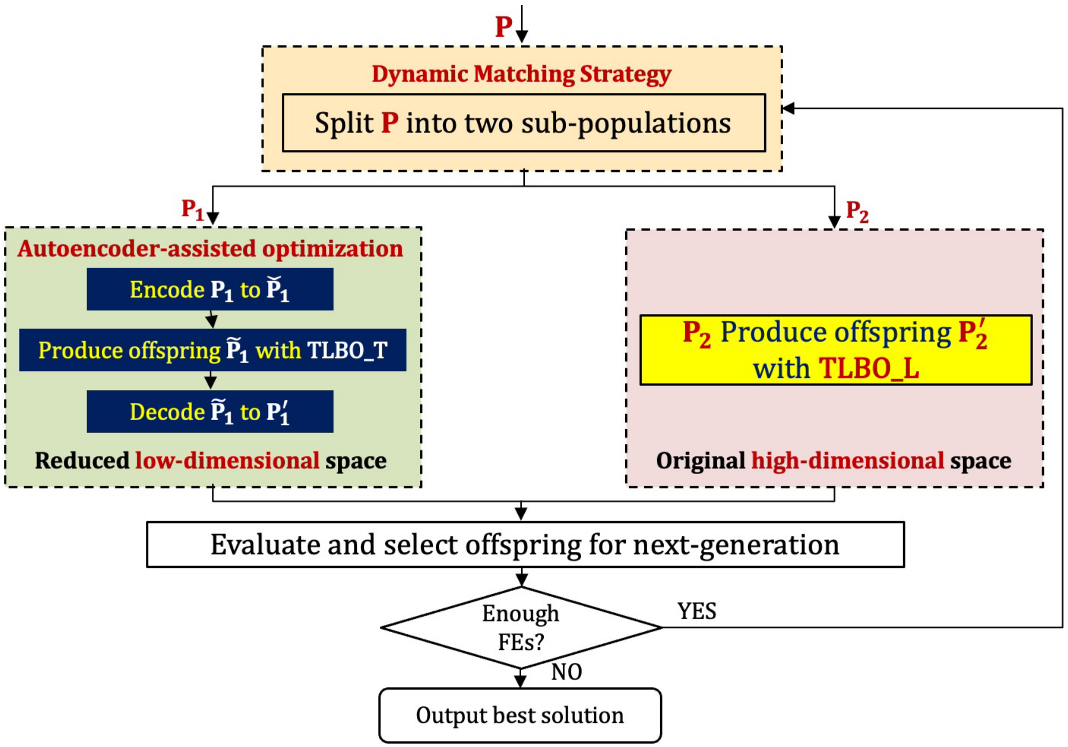
Learning Phase (TLBO-L): Local exploitation

$$\mathbf{x}$$

$$\mathbf{x}$$

$\mathbf{x}^*$ − teacher with the best fitness value
$\bar{\mathbf{x}}$ − mean of fitness values
r_i − random number in range $[0, 1]$
T_F − teaching factor 1 or 2

Autoencoder-embedded TTLBO (ATLBO)



- ✓ **Exploration and exploitation balance**
  - TLBO_T in low-dimensional space
  - TLBO_L in high-dimensional space
- ✓ **Dynamic matching bi-population**

Multi-swarm Gray-wolf-optimizer based on Genetic Learning (MGG)

Classic GWO: the predatory behavior simulation of gray wolf packs

- ✓ **Social hierarchy:** $\alpha, \beta, \delta, \omega$ from the fittest to worst
- ✓ **Hunting:** $\vec{\boldsymbol{X}}' = \dfrac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3}$

$$\vec{D}_\alpha = |\vec{C}_1 \cdot \vec{X}_\alpha - \vec{X}|, \quad \vec{D}_\beta = |\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}|, \quad \vec{D}_\delta = |\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}|,$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot \vec{D}_\alpha, \quad \vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot \vec{D}_\beta, \quad \vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot \vec{D}_\delta$$

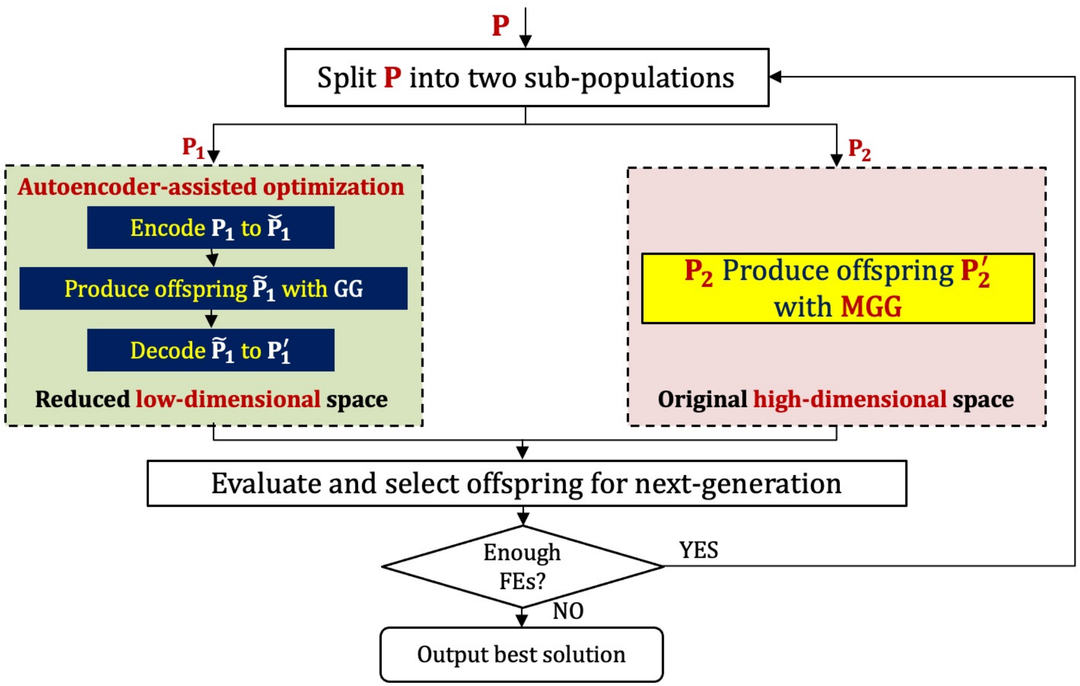
$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a}, \quad \vec{C} = 2\vec{r}_2$$

$r_1, r_2 -$ random number in range $[0, 1]$

MGG: enhance local search ability and guarantee population diversity

- ✓ **Divide population into many equal sub-populations**
- ✓ **Genetic operators**

Autoencoder-embedded MGG (AMGG)



- ✓ **Dynamic-subpopulation Number**

  Beginning: decreasing subpopulation count

  End: subpopulation count reduced to one

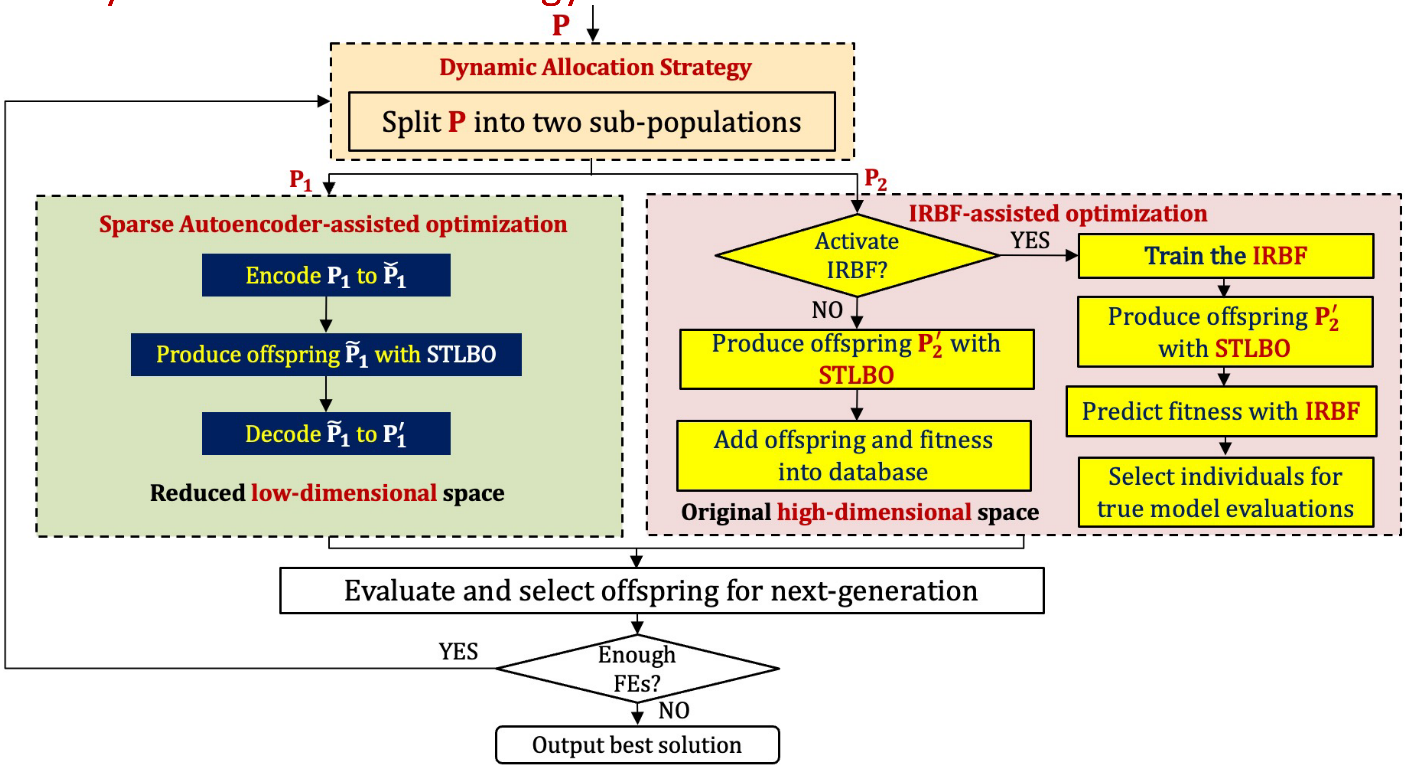
- ✓ **Subpopulation Stochastic Recombination**

- ✓ **GG:** Genetic-learning GWO

  MGG: Multi-swarm GG

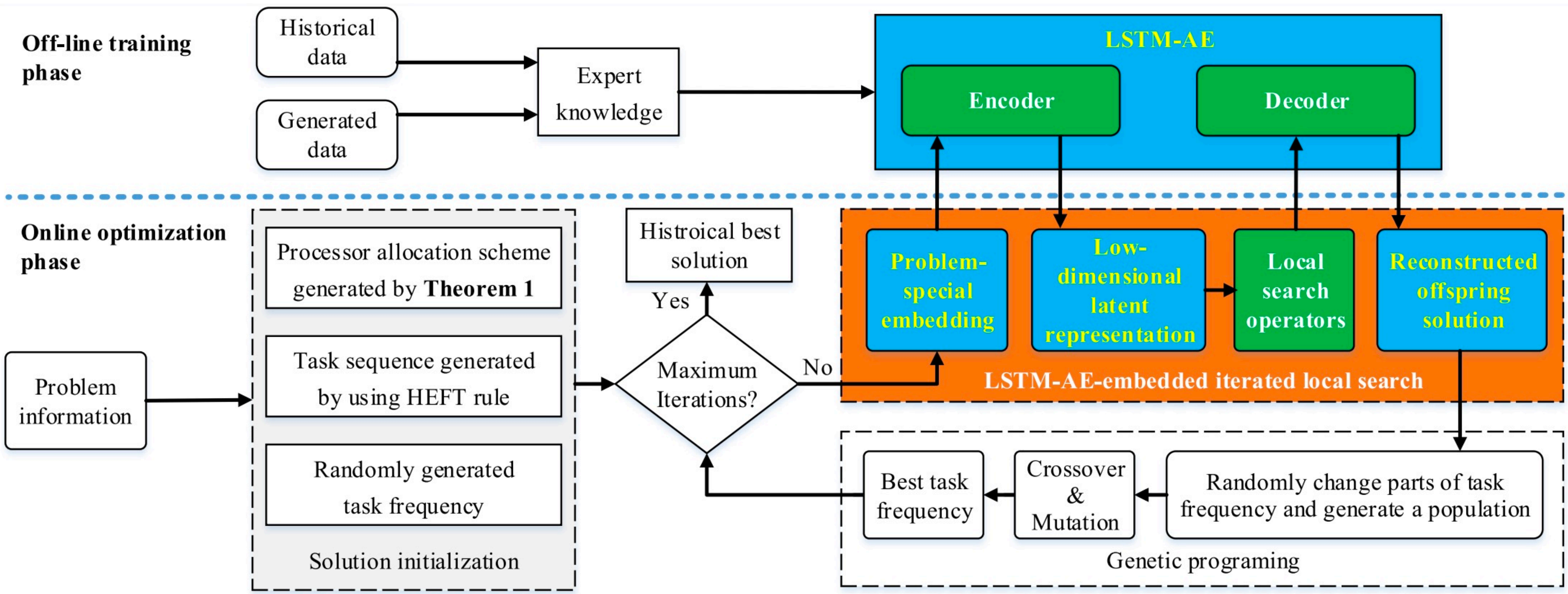
Self-adaptive TLBO with IRBF and a sparse Autoencoder (STORA)

- ✓ Self-adaptive TLBO **(STLBO)** as a baseline optimizer
- ✓ Improved Radial basis function **(IRBF)** as surrogates to predict fitness values
- ✓ Bi-population dynamic allocation strategy



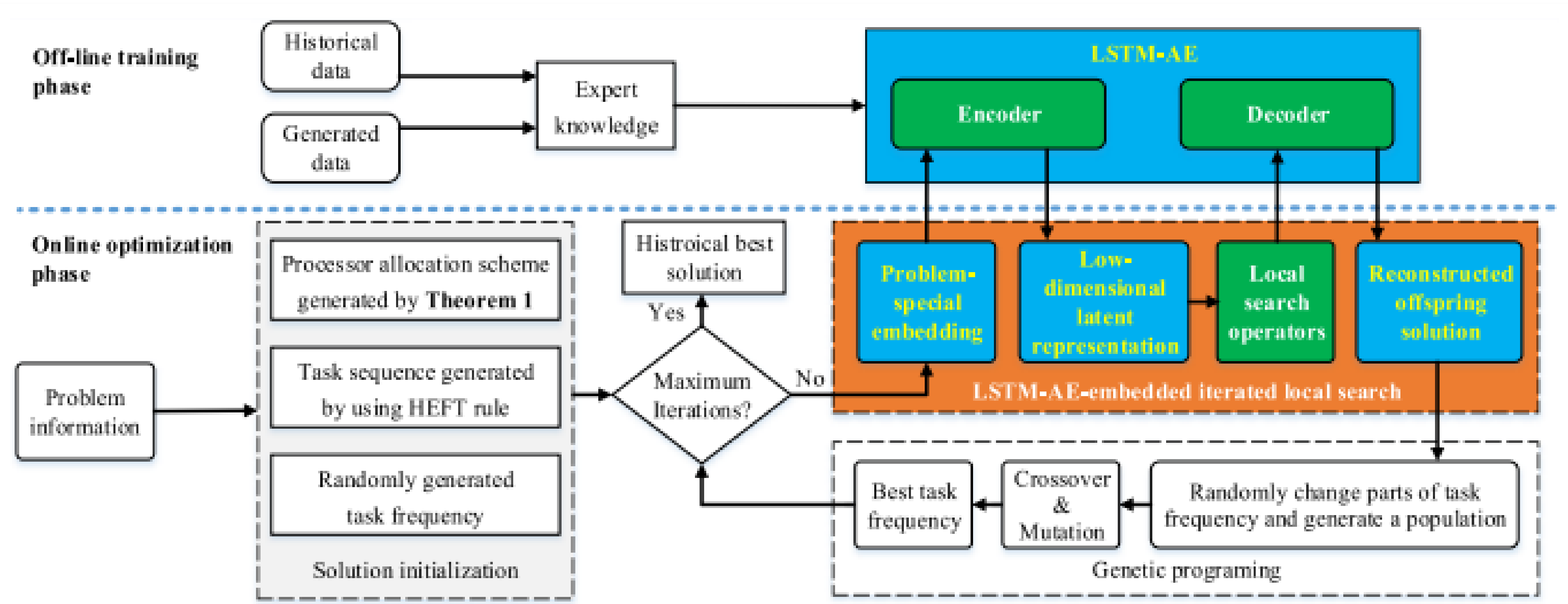
Autoencoder-embedded Iterated Local Search (AILS)

- ✓ Long Short-Term Memory-based Autoencoder (LSTM-AE)



AEO-based Method for Combinatorial Optimization

LSTM-AE-embedded Evolutionary Algorithm



LSTM= Long Short-Term Memory network

Applications to Function Optimization

Experimental Design

- Functions to be optimized

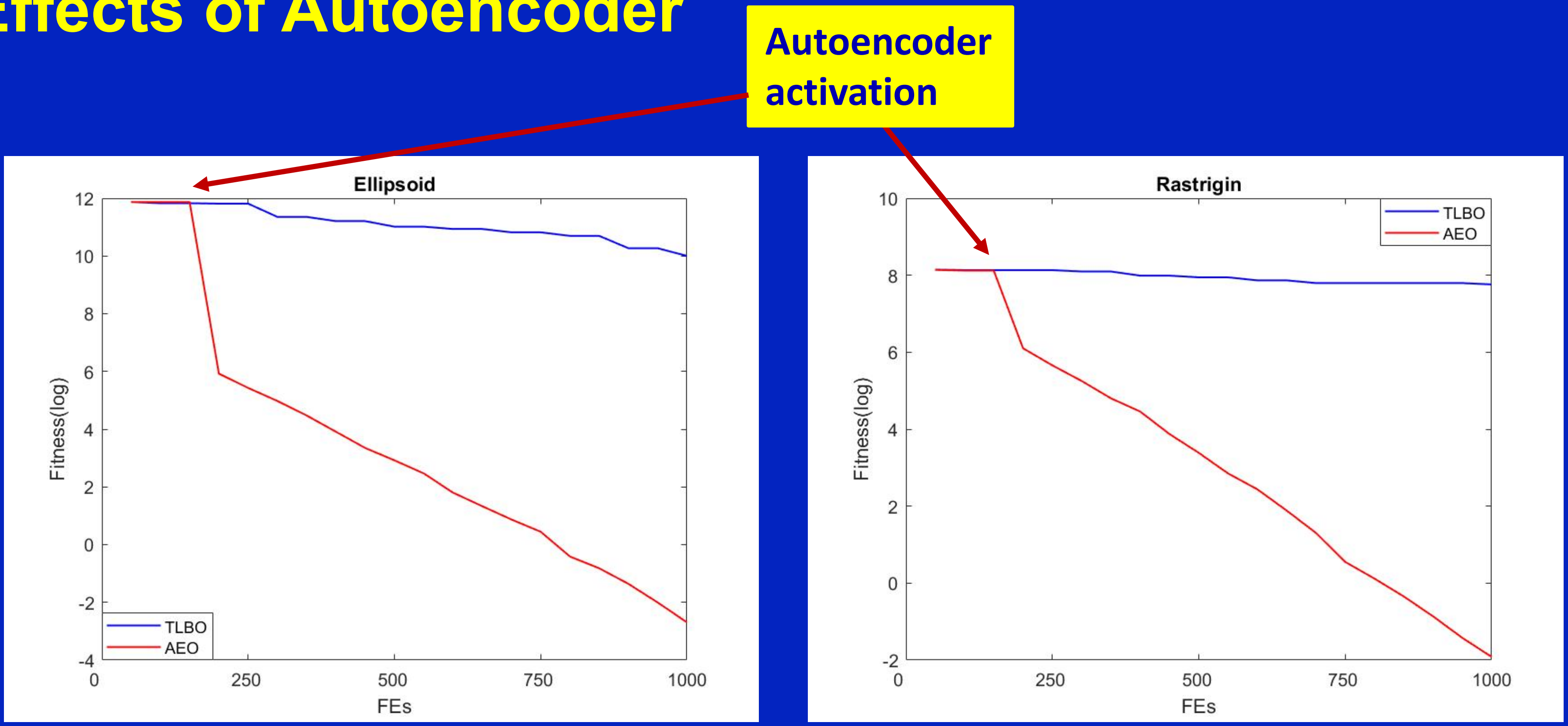
Fun	Name	Design space	$f^{*\dagger}$	Property
F1	Ellipsoid	$[-5, 5]^D$	0	Unimodal
F2	Rosenbrock	$[-2, 2]^D$	0	Multimodal with narrow valley
F3	Ackley	$[-32, 32]^D$	0	Multimodal
F4	Griewank	$[-600, 600]^D$	0	Multimodal
F5	Rastrigin	$[-5, 5]^D$	0	Multimodal
F6	Shifted rotated F5	$[-5, 5]^D$	-330	Multimodal & Complex
F7	Hybrid function‡	$[-5, 5]^D$	10	Multimodal & Complex

† f^* means global optimum.
‡ Rotated hybrid composition function with a narrow basin for the global optimum.

- Problem dimensions: 50-200
- Optimization algorithm: teaching-learning based optimization
- Stopping criterion: 1000 fitness evaluations

Experimental results

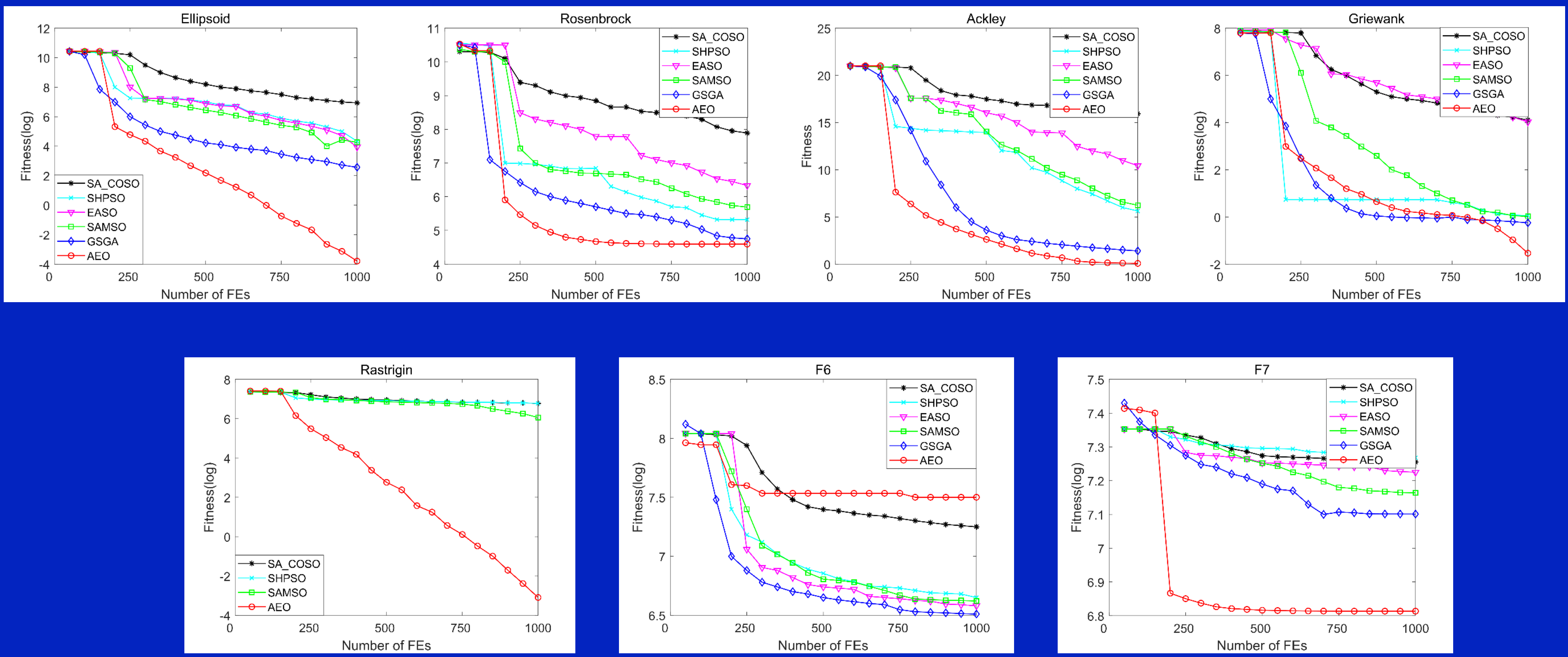
Effects of Autoencoder



Convergence curves of 200D Ellipsoid and 200D Rastrigin

Experimental results

Experimental results on benchmark functions



Convergence curves of AEO and other algorithms to deal with 100D problems

Experimental results

Experimental results on benchmark functions

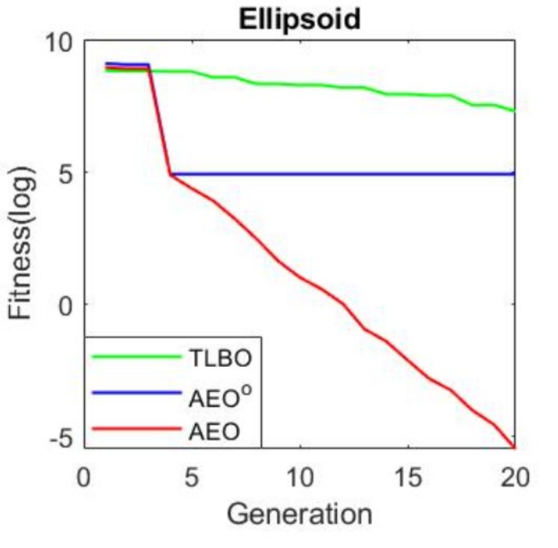
COMPARATIVE RESULTS WITH COMPARED ALGORITHMS ON 200D FUNCTIONS

Fun	SA_COSO	EASO	SAMSO	AEO
F1	1.63e+04(2.98e+03) +	1.76e+04(1.17e+03) +	1.52e+03(2.12e+02) +	**7.01e-02(6.46e-02)**
F2	1.64e+04(4.09e+03) +	4.31e+03(2.84e+02) +	1.15e+03(1.16e+02) +	**1.98e+02(7.84e-02)**
F3	1.78e+01(2.23e-02) +	1.46e+01(2.19e-01) +	1.20e+01(4.00e-01) +	**1.42e-01(4.28e-02)**
F4	5.77e+02(1.01e+02) +	5.72e+02(3.60e+01) +	9.03e+00(1.33e+00) +	**1.79e-01(2.04e-01)**
F6	**3.92e+03(2.72e+02)** −	5.38e+03(1.56e+02) +	4.96e+03(1.38e+02) +	4.80e+03(2.19e+02)
F7	1.34e+03(2.46e+01) +	1.45e+03(2.04e+01) +	1.34e+03(2.43e+01) +	**9.10e+02(1.00e-02)**
$+/\approx/-$	5/0/1	6/0/0	6/0/0	**N/A**

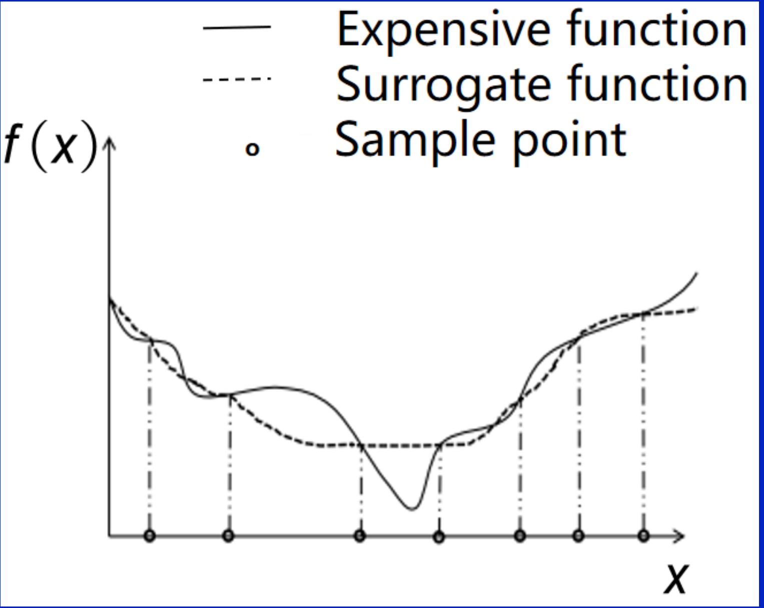
Meiji Cui, Li Li*, Mengchu Zhou*, Jiankai Li, Abdullah Abusorrah. A Bi-population Cooperative Optimization Algorithm Assisted by an Autoencoder for Medium-scale Expensive Problems. ***IEEE/CAA Journal of Automatica Sinica***, 2022. DOI: 10.1109/JAS.2022.105425

Surrogate-assisted AEO (SAEO)

➤ Will AEO ***further converge*** given more FEs?



(a) D=50



Example of surrogate model

- Incorporate surrogate models into AEO

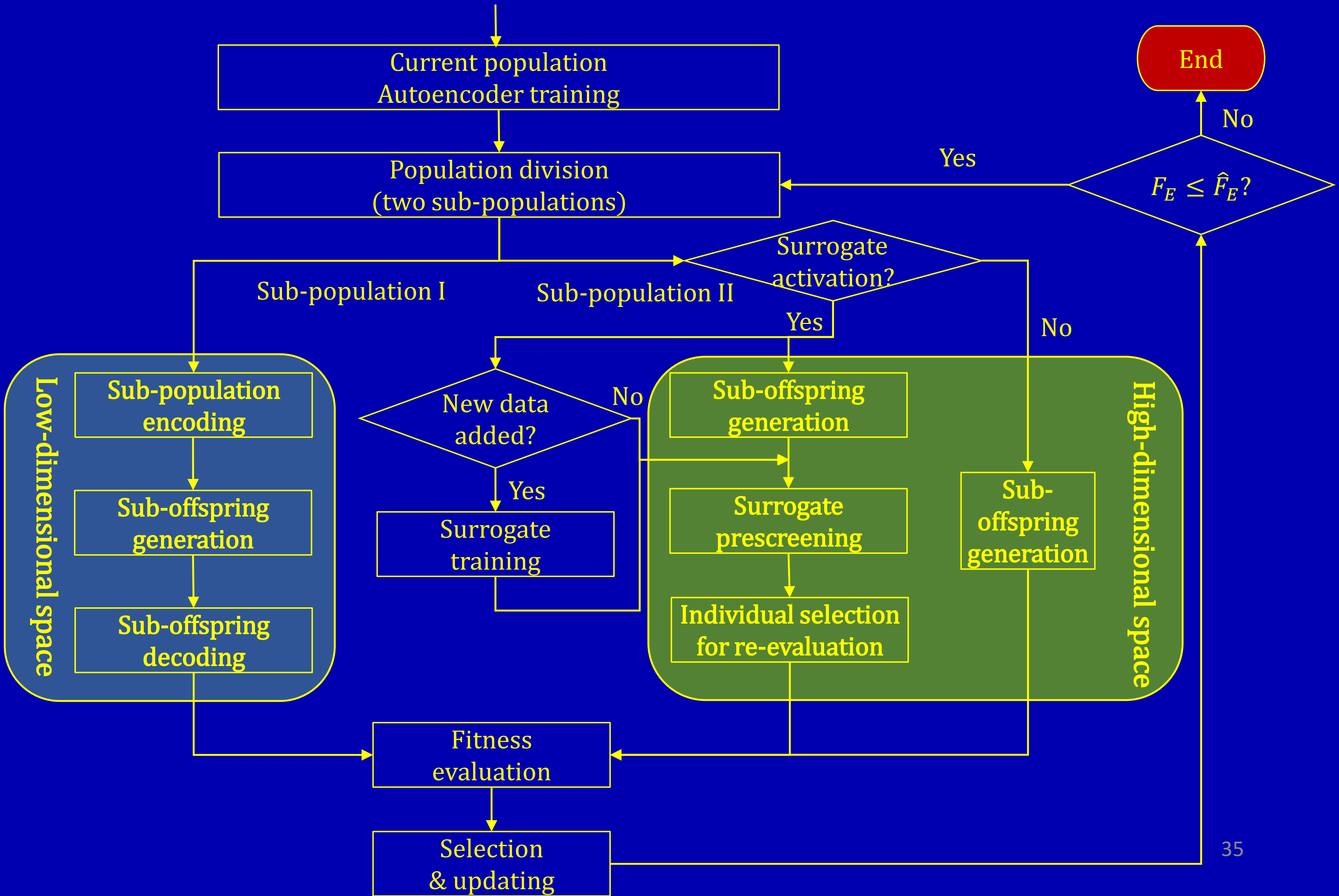
Surrogate-assisted AEO (SAEO)

- **Surrogate selection**
  - Gaussian Processes (GPs)
  - Random forest (RF)
  - Radial basis function (RBF)

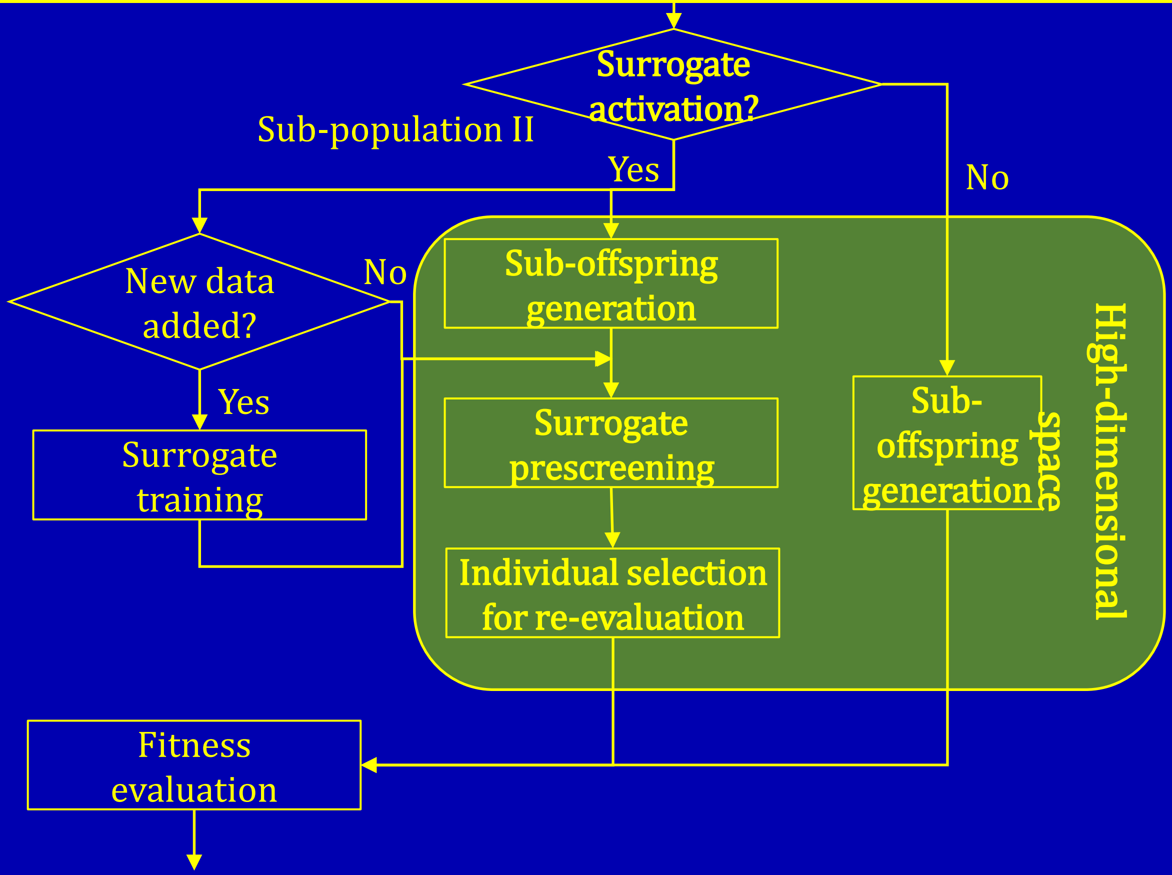
- **Surrogate construction**
  - Dimension reduction technique
  - Surrogate ensembles

- **Surrogate management**
  - Performance-based indicator
  - Uncertainty-based indicator
  - Both of them

Surrogate-assisted AEO (SAEO)



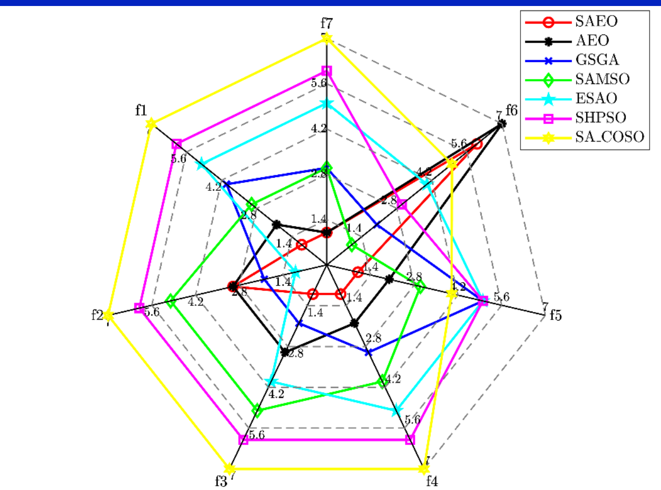
Surrogate-assisted AEO (SAEO)



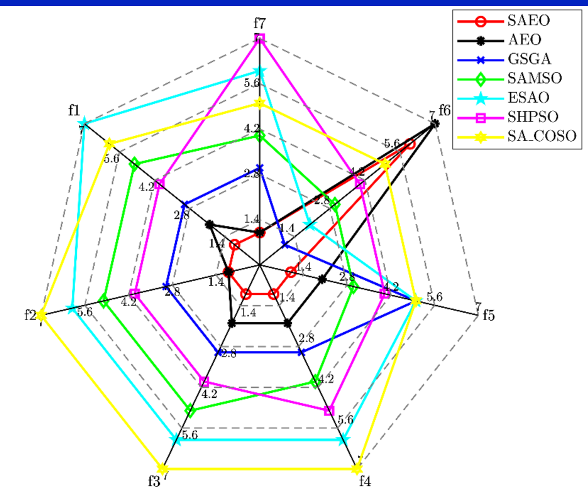
- Model construction: surrogate activation condition (balance model accuracy and construction time)
- Model management: re-evaluate individuals whose predicted values better than historical ones (guarantee convergence speed)

Experimental results

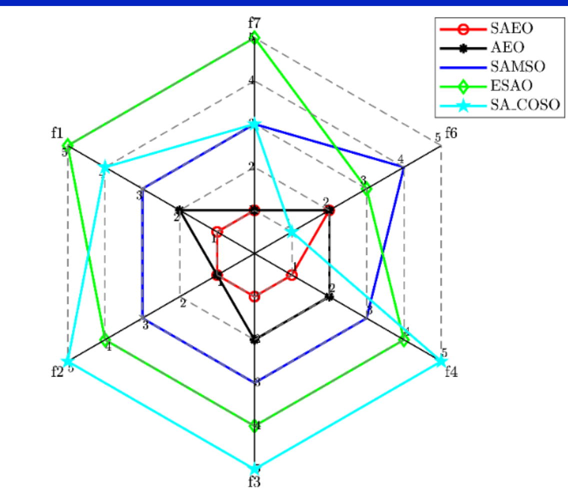
Experimental results on benchmark functions



50D



100D

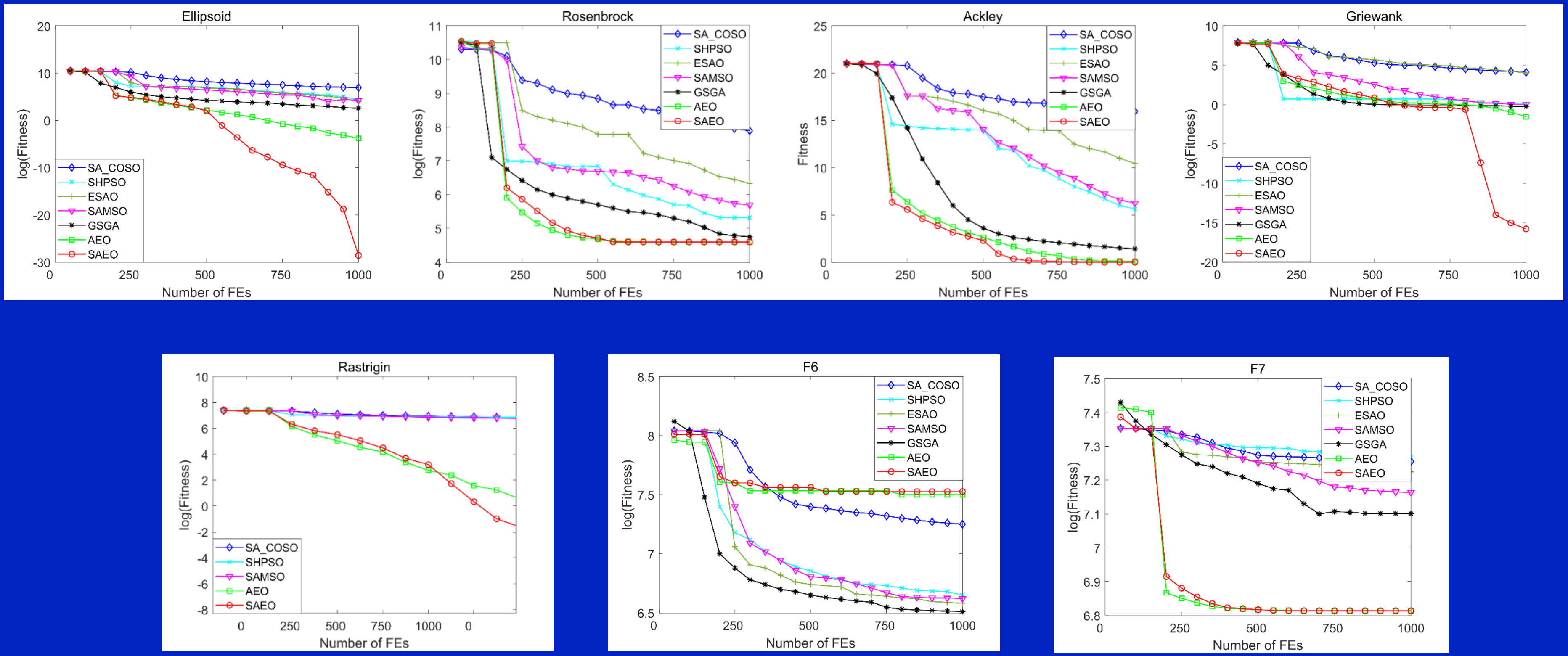


200D

Radar figures of different algorithms

Experimental results

Experimental results on benchmark functions



Convergence curves of SAEO and other algorithms to deal with 100D problems

Experimental results

Experimental results on benchmark functions

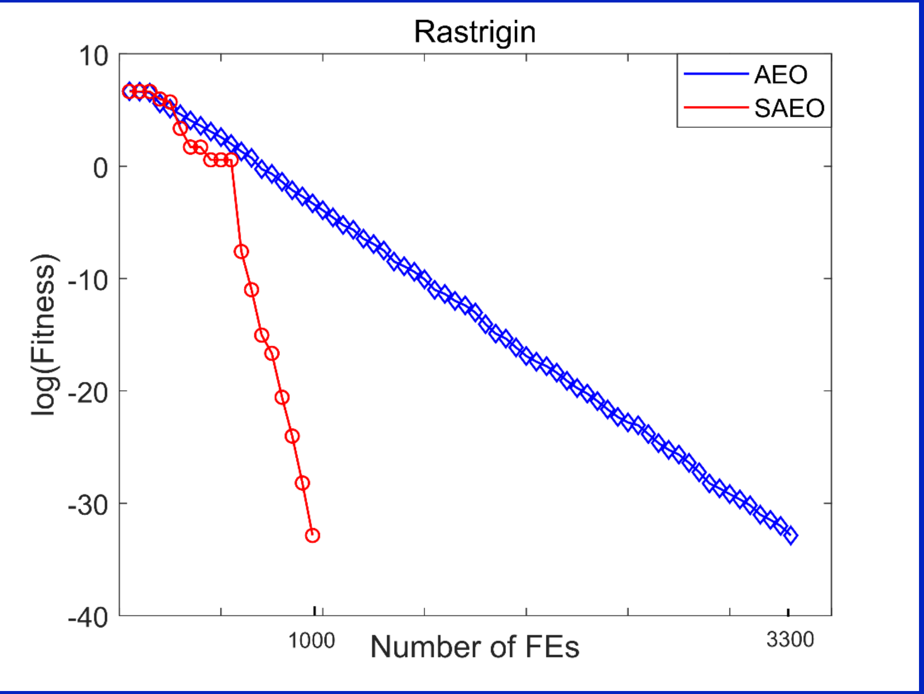
Comparative results of different algorithms on 500D problems

Fun	SAMSO	SAEO
F1	3.28e+04(3.66e+03)	**6.60e-09(3.70e-09)**
F2	2.72e+03(3.39e+02)	**4.98e+02(1.90e-02)**
F3	1.30e+01(5.71e+00)	**3.08e-08(1.02e-08)**
F4	4.43e+02(8.40e+01)	**1.98e-09(1.21e-09)**
F5	4.55e+03(1.13e+02)	**1.69e-07(2.95e-07)**
F6	1.35e+04(1.11e+02)	**1.28e+04(1.29e+04)**
F7	1.17e+03(2.17e+01)	**9.10e+02(1.01e+01)**

Meiji Cui, Li Li*, Mengchu Zhou*, Abdullah Abusorrah. Surrogate-assisted Autoencoder-embedded Evolutionary Optimization Algorithm to Solve High-dimensional Expensive Problems. ***IEEE Transactions on Evolutionary Computation***, 2021. DOI: 10.1109/TEVC.2021.3113923

Experimental results

AEO V.S. SAEO



Comparative convergence curves of AEO and SAEO

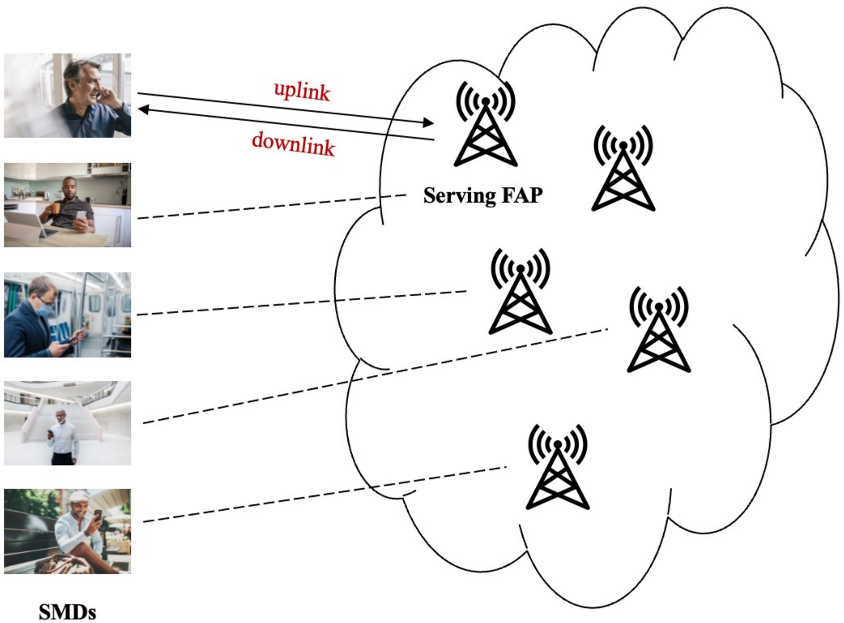
- General framework
- Suitable to deal with large-scale expensive problems

Applications

- Scheduling Mobile Edge Computing Systems
- Task scheduling for Human-cyber-physical Systems
- Job-shop and flow-shop scheduling in discrete manufacturing

Mobile Edge Computing Systems (MEC)

Energy Consumption Minimization (ECM)



SMDs: *smart mobile devices*
FAPs: *femto access points*

- ✓ The limited energy, computing, and storage resources of smart mobile devices (SMDs).
- ✓ Providing more scalable performance, reducing network load and hastening data transmission.
- ✓ High-dimensional and resource-intensive computing data
- ✓ How to effectively reduce energy consumption.

Mobile Edge Computing Systems (MEC)

Energy Consumption Minimization (ECM)

- **High-dimensional Problem:** ***300***-Dimensionality in our work
- **Single-objective:** Minimize total energy consumed by all SMDs and edge servers while guaranteeing constrains for prolong battery life

$$\min_{f_l, P_t, \lambda} E(f_l, P_t, \lambda)$$

$$\begin{aligned} s.t.\ \ \mathbf{C1}&: L(f_l, P_t, \lambda) \le L_{\max} \\ \mathbf{C2}&: 0 \le \lambda \le 1 \\ \mathbf{C3}&: 0 \le P_t \le P_{t_{\max}} \\ \mathbf{C4}&: 0 \le f_l \le f_{l_{\max}} \end{aligned}$$

f_l — computational speed of SMD

P_t — transmit power of SMD

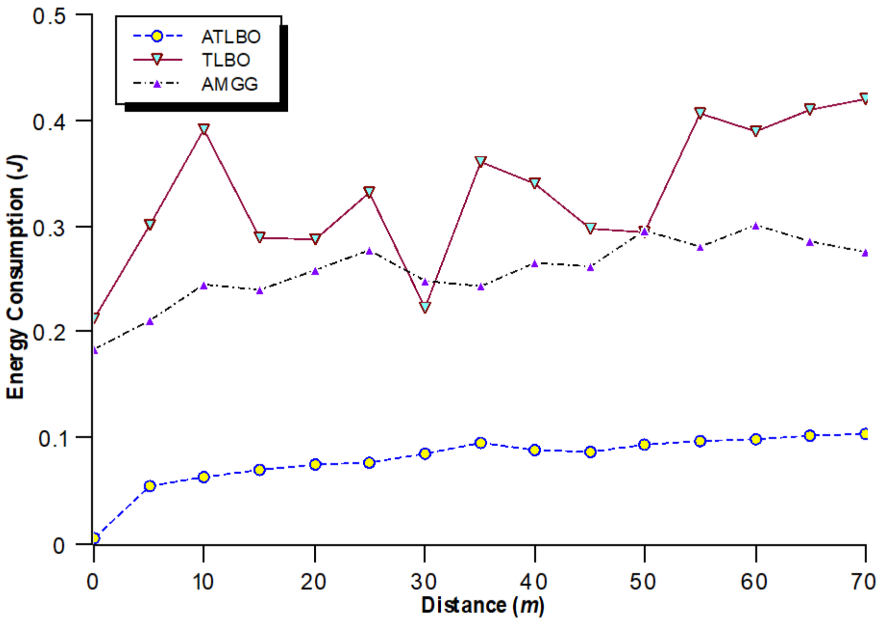
λ — ratio of locally executed amount of bits to the total input data bits

L — the latency to execute an application

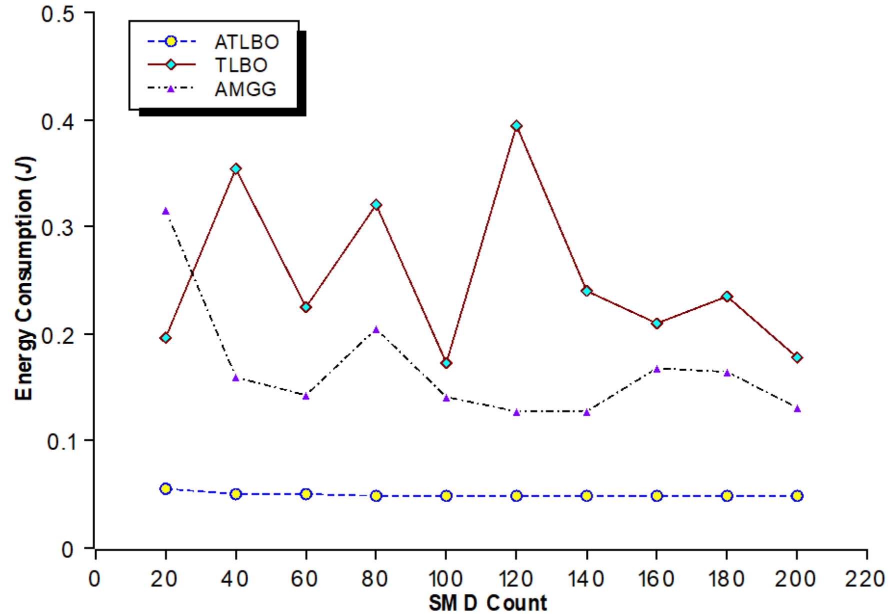
ECM Problem in Mobile Edge Computing Systems

ATLBO Performance

- ✓ ATLBO performs **three to six times better** than TLBO and AMGG.
- ✓ Main criterion: execution time



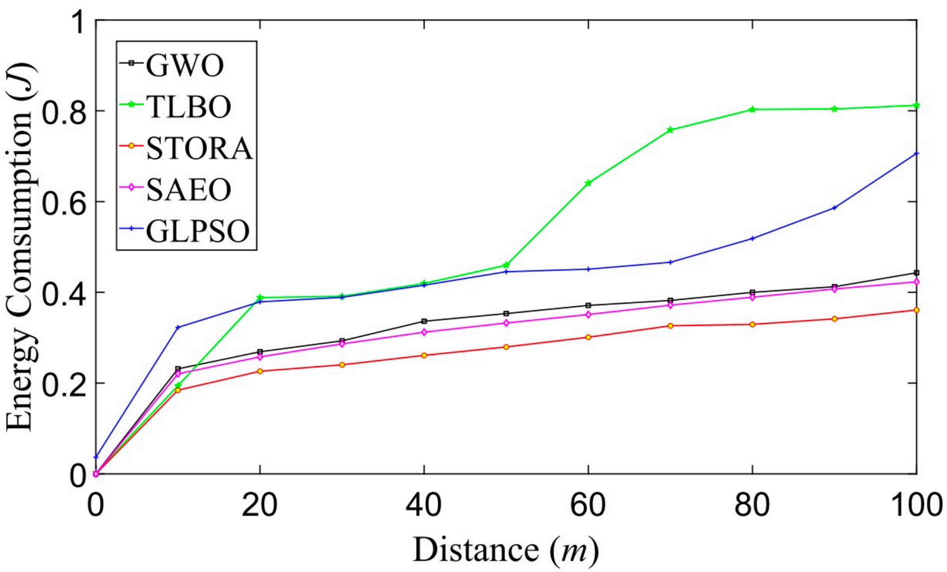
Average energy consumption v.s. distance



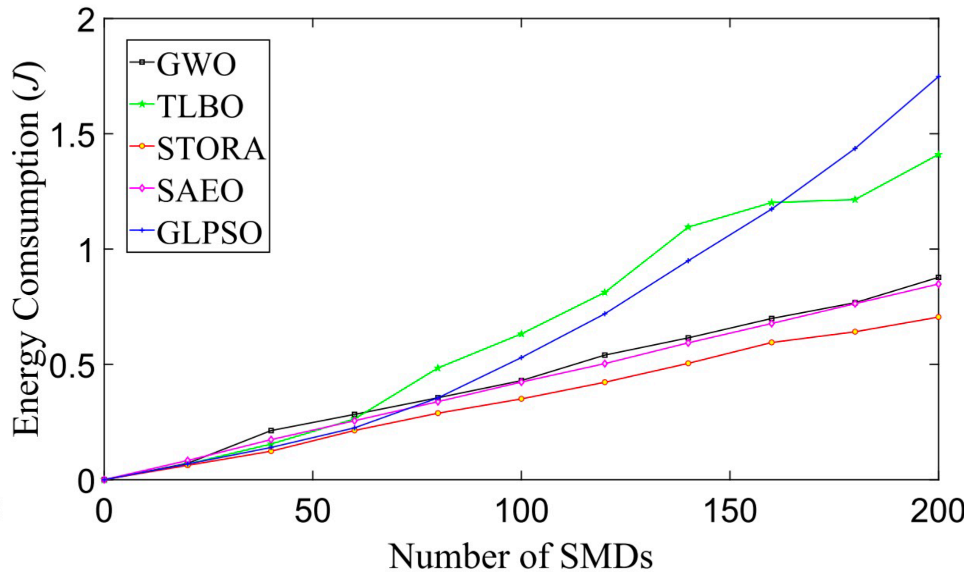
Average energy consumption v.s. SMD count

SMD: *smart mobile device*
Distance: *distance from SMD to its serving femto access points (FAPs)*

STORA Performance



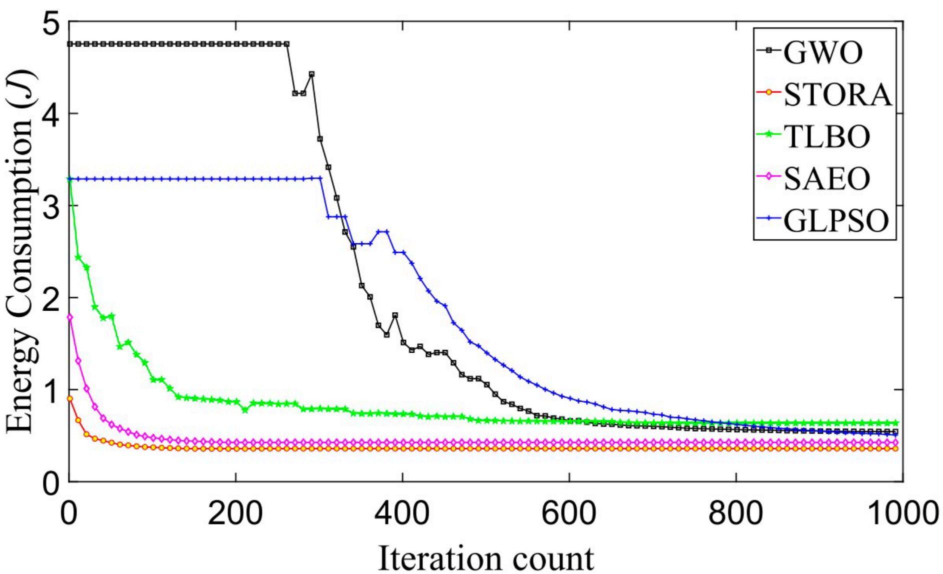
Average energy consumption v.s. distance



Average energy consumption v.s. SMD count

SMD: *smart mobile device*
Distance: *distance from SMD to its serving femto access points (FAPs)*

STORA Performance



Energy consumption in each iteration

Task scheduling in human-cyber-physical systems

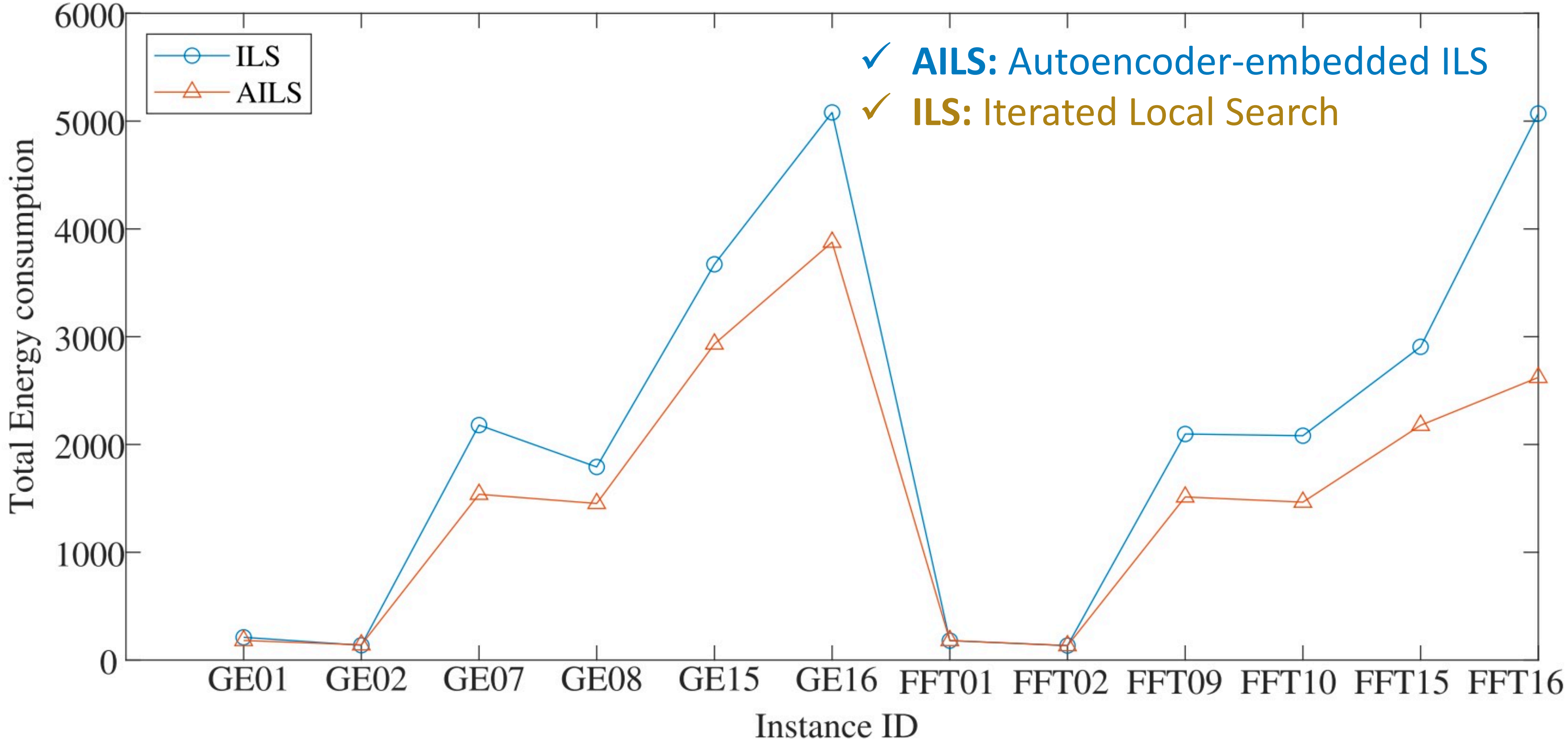
Execute *n* deadline-constrained tasks on *m* heterogeneous processors in, e.g., semiconductor manufacturing and smart logistics.

Tasks are represented by a group of directed acyclic graphs.

- **High-dimensional Problem**
- **Single-objective:** Minimize energy consumption during task scheduling
- **Limited computational resources**

ETSD Problem in Human-cyber-physical Systems

Average energy consumption of ILS and AILS



LSTM-AE-embedded Evolutionary Algorithm for Scheduling Hewlett-Packard's Post-printing Process

- 91 jobs and 10 machines
- The scheduling problem has about 2.376×10^{42} feasible solutions.

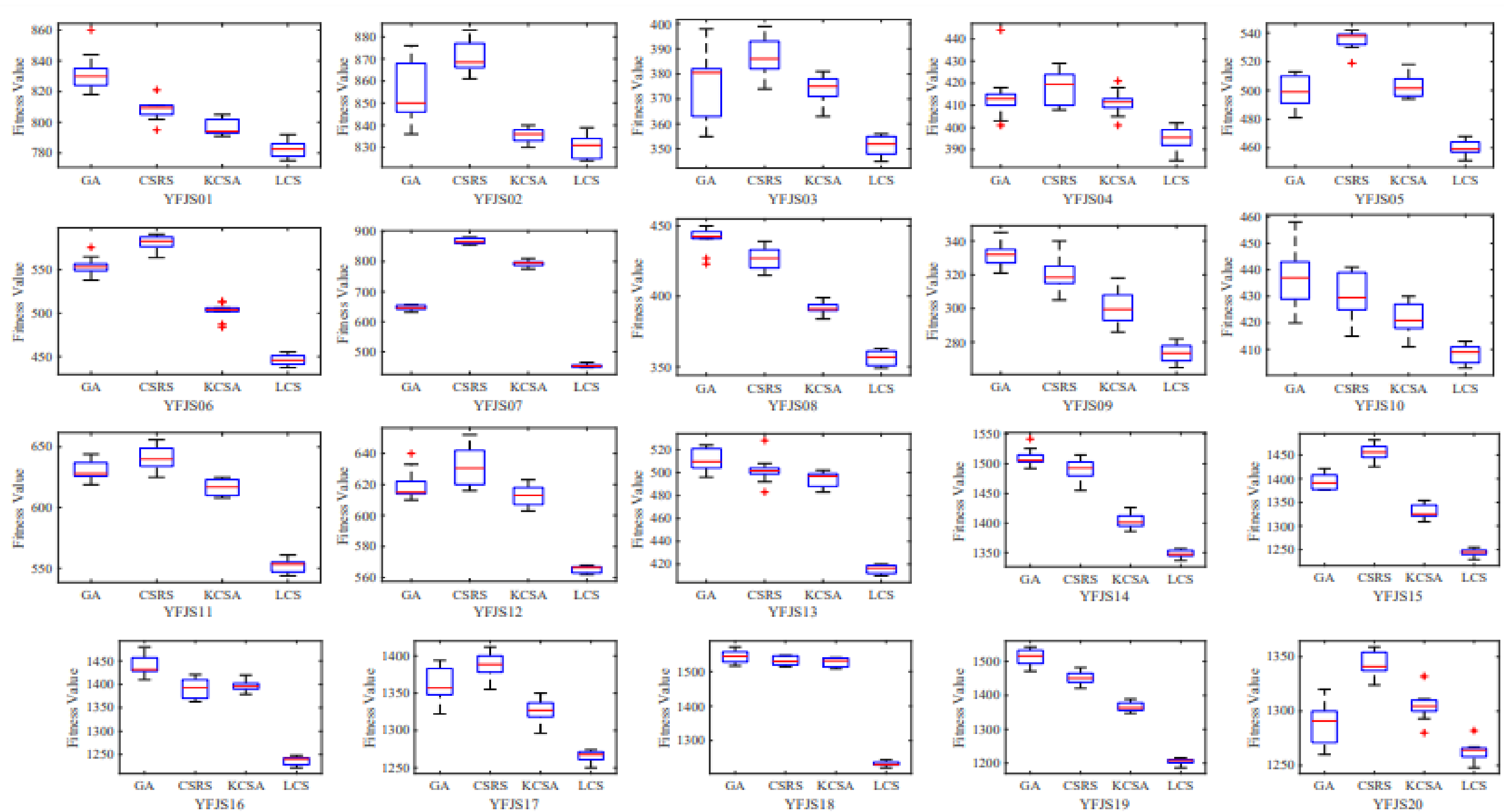
- Tianhe-2 can perform 5.49×10^{16} times fitness evolution per second. It will take 1.373×10^{16} years to do so.
- Note that lifespan of our universe = 1.5×10^{10} years.

Numerical simulation results

	Scenario (n, m)	IBM ILOG CPLEX 12.1 (1 hour)	LGWO without autoencoder	CPU time (second)	LSTM-AE-embedded EA	CPU time (second)
Small-scale Problem	DAFJS01 (26,5)	**257**	264	80	261	65
	DAFJS02 (25,5)	**289**	295	81	292	65
	DAFJS05 (39,5)	576	**401**	112	405	98
Medium-scale Problem	DAFJS07 (85,10)	565	524	275	**505**	231
	DAFJS11 (113,10)	708	697	272	**680**	221
	DAFJS12 (117,10)	720	730	312	**706**	251
Large-scale Problem	YFJS17 (289,26)	1622	2120	360	**1290**	360
	YFJS18 (289,26)	2082	2341	360	**1499**	360
	YFJS19 (289,26)	1525	2231	360	**1333**	360
	YFJS20 (289,26)	2020	3082	360	**1279**	360

Numerical simulation results

Autoencoder-embedded EA improves average fitness values by 10.8-16.1% over Genetic Algorithm, Cuckoo Search with reinforcement learning and surrogate modeling, and Knowledge-based Cuckoo Search.

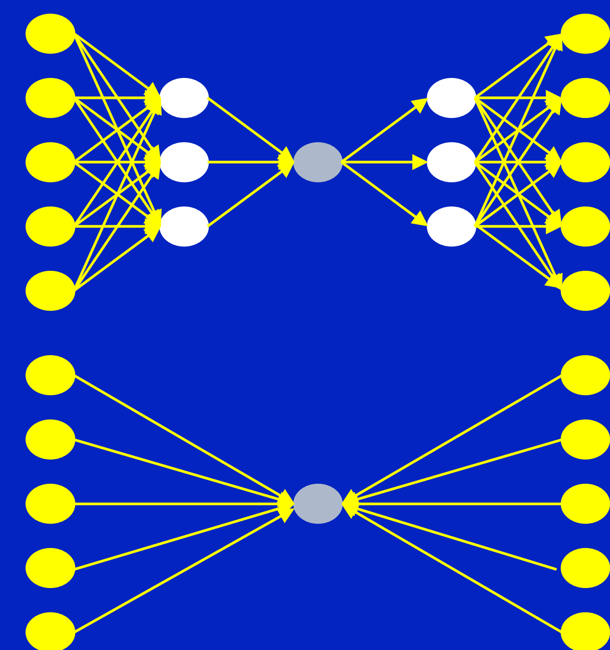


Future Research

- Will ***autoencoder architecture*** affect?
  - activation function
  - hidden layer

Relu	$f(x) = \max(0, x)$
Softmax	$p(y\|x) = \frac{\exp(f_y)}{\sum_{c=1}^{C} \exp(f_c)} = softmax(f)_y$
Elu	$f(x) = \begin{cases} -x & x > 0 \\ \alpha(\exp(x) - 1, & x \leq 0 \end{cases}$
Selu	$f(x) = \lambda \begin{cases} x & x > 0 \\ , \alpha(\exp(x) - 1x \leq softplus0 \end{cases}$
Tanh	$f(x) = \tanh\left(\frac{x}{2}\right)$
sigmoid	$f(x) = \frac{1}{1 + \exp(-\alpha x)}$

Activation functions



Hidden layers

Conclusions

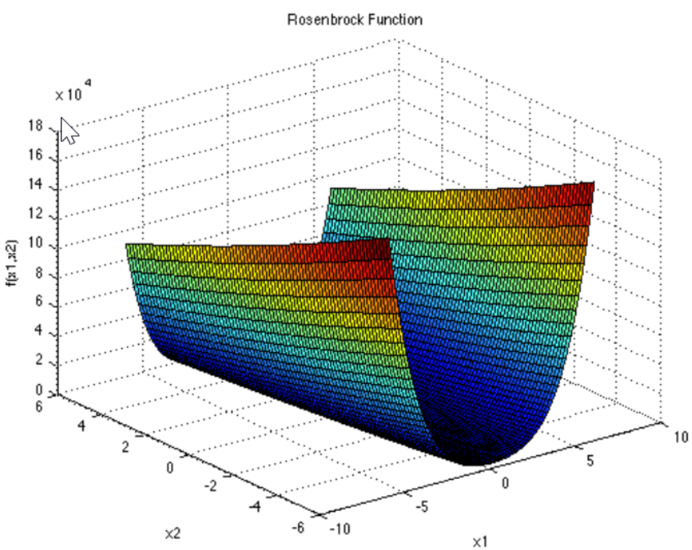
- AEO: suitable for large-scale expensive problems
- SAEO: enhance the performance of AEO given limited computational resources
- AEO and SAEO: general framework

Welcome your questions!

zhou@njit.edu

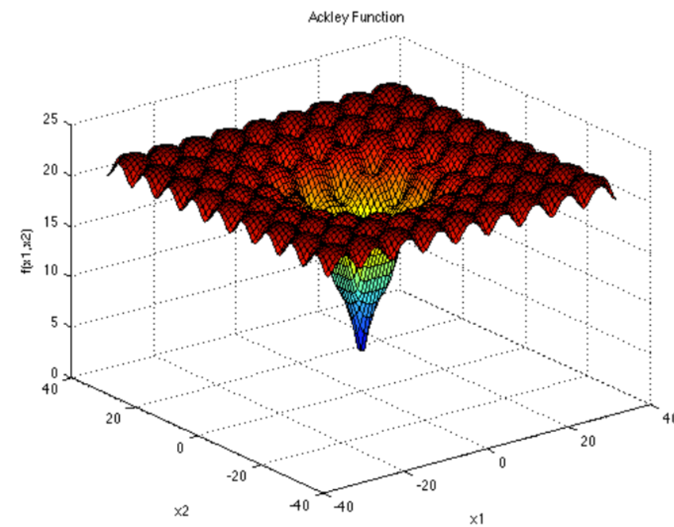
Functions

ROSENBROCK FUNCTION



$$f(\mathbf{x}) = \sum_{i=1}^{d-1} \left[100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2\right]$$

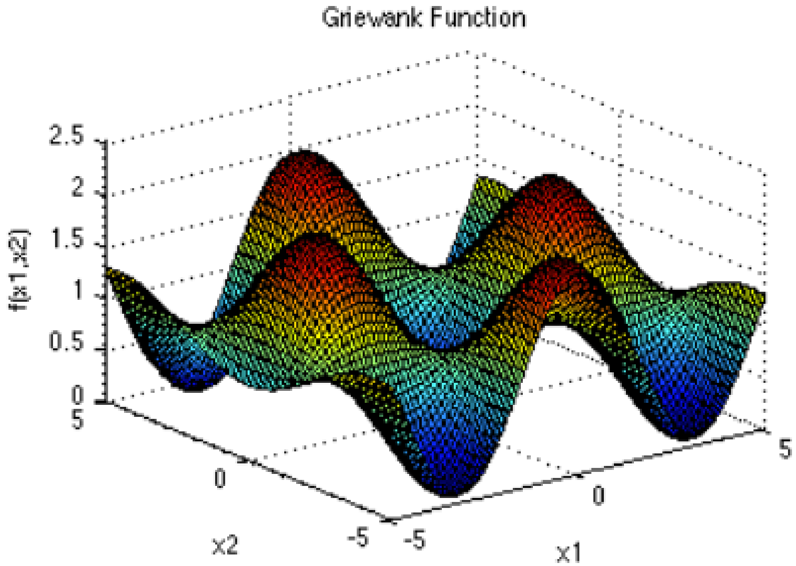
ACKLEY FUNCTION



$$f(\mathbf{x}) = -a \exp\left(-b\sqrt{\frac{1}{d}\sum_{i=1}^{d} x_i^2}\right) - \exp\left(\frac{1}{d}\sum_{i=1}^{d}\cos(cx_i)\right) + a + \exp(1)$$

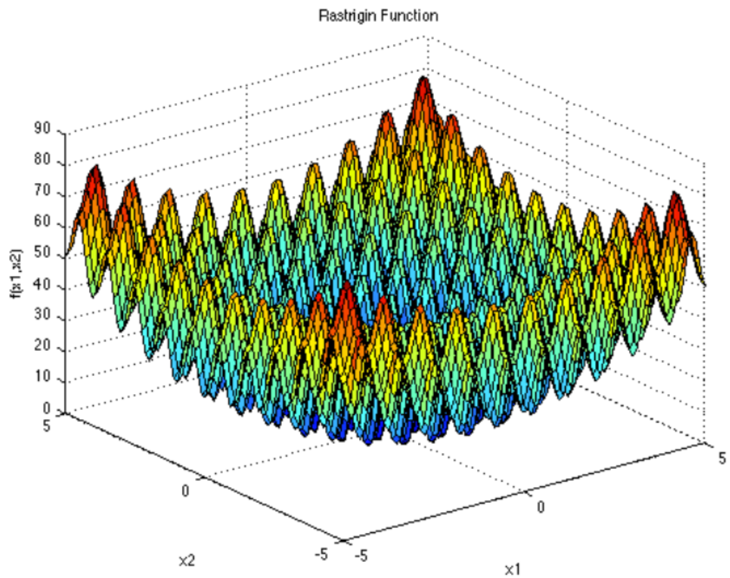
Functions

GRIEWANK FUNCTION



$$f(\mathbf{x}) = \sum_{i=1}^{d} \frac{x_i^2}{4000} - \prod_{i=1}^{d} \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$$

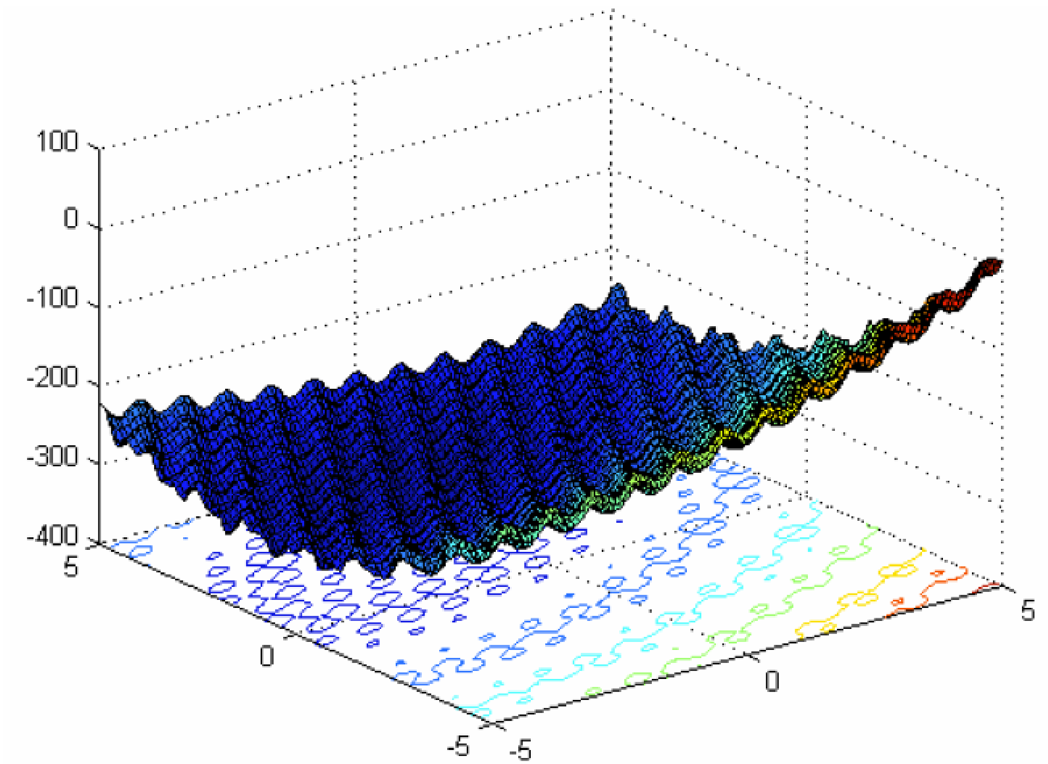
RASTRIGIN FUNCTION



$$f(\mathbf{x}) = 10d + \sum_{i=1}^{d} \left[x_i^2 - 10\cos(2\pi x_i)\right]$$

Functions

F6 Shifted Rotated Rastrigin



$$F_{10}(\mathbf{x}) = \sum_{i=1}^{D} (z_i^{\,2} - 10\cos(2\pi z_i) + 10) + f\_bias_{10}\,,\ \mathbf{z} = (\mathbf{x}-\mathbf{o})*\mathbf{M}\,, \mathbf{x} = [x_1, x_2, ..., x_D]$$

Functions

F7 Rotated Hybrid Composition Function with narrow basin global optimum