Design Optimization of Tokamak Fusion Reactor through Deep Reinforcement Learning

Fusion Reactor design optimization with single-step reinforcement learning

¹⁾Jinsu Kim and ^{2)*}Jaemin Seo

¹⁾ Department of Mechanical and Aerospace Engineering, Princeton University (wlstn5376@gmail.com)

²⁾ Department of Physics, Chung-Ang University (jseo@cau.ac.kr)

2024.05.03

Contents

Design Optimization

- Objectives of design optimization
- How to approach
- Tools for design optimization
 - ✓ Brute force algorithm: Grid search
 - ✓ Generic algorithm: Stochastic global search optimization algorithm
 - ✓ Reinforcement learning: Neural combinatorial optimization through RL

Related works

- Airfoil Design Optimization
- Chip Design Optimization

Fusion Reactor Design for DEMO

- Aspects of nuclear physics and engineering for designing the Tokamak reactor: Blanket, Coil, Armour
- Aspects of plasma physics for designing the optimal state operation: T, P, n, Ip, fbs,q, nG, beta, Tau, Pw
- Verification of the design parameters computation code

Design Optimization



Design Optimization

Design variables

- Input parameter
- Must not depend on each other
- X = [x1,x2,x3,x4,...]
- The optimizer must be free to choose the elements of x independently.
- Each x should be bound for physical constraints

minimizef(x) $x_0 \rightarrow 0$ by varying $\underline{x}_i \leq x_i \leq \overline{x}_i$ $i = 1, \dots, n_x$ xsubject to $g_j(x) \leq 0$ $j = 1, \dots, n_g$ $h_l(x) = 0$ $l = 1, \dots, n_h$



Design Optimization - Objectives

Objectives (targets) of design optimization

- System performance
- Cost

 \checkmark

 \checkmark

- Uncertainty
- Physical consistency

What we need

- Objective function: A quantity that determines if one design is better or not
- Constraints: Restrict design variables to be set on a feasible region





Design Optimization - Issues

Multi-modality



Multidisciplinary Design Optimization

Design Optimization - How to approach



Design Optimization - How to approach

Single-step reinforcement learning for design optimization

- Surrogate models based on ML / Genetic algorithm: High computational cost
- Discrete design parameters vs Continuous design parameters: RL can handle both variables
- RL training process => directly finding the optimal solutions

ALGORITHM 1: Proximal policy optimization.

1 Initialize policy model parameters θ_0 , value model parameters ϕ_0 ; 2 For each episode, i = 1, ..., N do; 3 Based on current policy $\pi_{\theta,i}$ for T time steps, collecting trajectories $\{\tau\} = \{s_t, a_t, r_t\};$ 4 Estimate advantage values $\hat{A}_t = \sum_{t'>t} \gamma^{t'-t} r_{t'} - V_{\phi}(s_t);$ 5 $\sum_{t'>t} \gamma^{t'>t} r_{t'}$ is the rollout policy; 6 V_{ϕ} is given by the value network; 7 $\gamma \in (0, 1)$ is the discount factor, which represents the influence of future states on the current state; 8 Update θ by a gradient method (e.g., Adam) with respect to J_{θ} : $J_{\theta} = \sum_{t=1}^{T} \min \left[\frac{\pi_{\theta}(a_t/s_t)}{\pi_{old}(a_t/s_t)} \hat{A}_t, clip \left(\frac{\pi_{\theta}(a_t/s_t)}{\pi_{old}(a_t/s_t)}, 1 - \varepsilon, 1 + \varepsilon \right) \hat{A}_t \right],$ where ε is a hyperparameter, $\varepsilon = 0.2$;

9 Update ϕ by a gradient method (e.g., Adam) with respect to L_{ϕ} :

$$L_{\phi} = -\sum_{t=1}^{T} \left(\sum_{t'>t} \gamma^{t'-t} r_{t'} - V_{\phi}(s_t) \right)^2$$



Figure 4: **Degenerate reinforcement learning framework used in this paper**. One episode consists of a single control from the agent: the same initial observation is provided to the agent at the start, which in return provides an action to the environment. The environment returns a reward value to the agent, and the episode is terminated.

Paper: Direct Shape Optimization through Deep Reinforcement Learning



(a) Best shape with 4 points, 1 free point (3 d.o.f.s)



(b) Best shape with 4 points, 3 free points (9 d.o.f.s)



(c) Best shape with 4 points, 4 free points (12 d.o.f.s)



(a) Instantaneous and moving-average reward history using 4 free points



(b) Moving-average reward history using 1, 3 and 4 free points



(d) Computed v_x velocity field at $Re \sim 600$ around shape 5c (the domain is cropped).



Paper: Direct Shape Optimization through Deep Reinforcement Learning



Interpolation for Airfoil Design Configuration

C Shape generation using Bézier curves

This section describes the process followed to generate shapes from a set of n_s points provided by the agent. Once the points are collected, an ascending trigonometric angle sort is performed (see figure 11a), and the angles between consecutive points are computed. An average angle is then computed around each point (see figure 11b) using:

$$\theta_i^* = \alpha \theta_{i-1,i} + (1-\alpha)\theta_{i,i+1}$$

with $\alpha \in [0, 1]$. The averaging parameter α allows to alter the sharpness of the curve locally, maximum smoothness being obtained for $\alpha = 0.5$. Then, each pair of points is joined using a cubic Bézier curve, defined by four points: the first and last points, p_i and p_{i+1} , are part of the curve, while the second and third ones, p_i^* and p_i^{**} , are control points that define the tangent of the curve at p_i and p_{i+1} . The tangents at p_i and p_{i+1} are respectively controlled by θ_i^* and θ_{i+1}^* (see figure 11c). A final sampling of the successive Bézier curves leads to a boundary description of the shape (figure 11d). Using this method, a wide variety of shapes can be attained.

Paper: Direct Shape Optimization through Deep Reinforcement Learning





Figure 4: **Degenerate reinforcement learning framework used in this paper**. One episode consists of a single control from the agent: the same initial observation is provided to the agent at the start, which in return provides an action to the environment. The environment returns a reward value to the agent, and the episode is terminated.

- Determine 4-points
- Interpolation for designing airfoil
- Compute the design performance (drag / lift coefficient)



$$f_d = \int \sigma \, n e_x dS \quad f_l = \int \sigma \, n e_y dS$$

• Evaluate the reward w.r.t design objectives

$$r_t = <\frac{C_l}{|C_d|} > - <\frac{C_l}{|C_d|} >_{cyl}$$

• Optimize the policy through Single-step PPO algorithm

Paper: DRL for engineering design through topology optimization of elementally discretized design domains

- Topology optimization through DRL for 2D material design
- Pixel -> discrete actions
- Input: 2D pixels => output: change of pixel flips





Fig. 6. Deep convolutional neural network architecture.

Fusion Power Plant (FPP) System



Designing a Nuclear Fusion Reactor



Designing a Nuclear Fusion Reactor

Objectives for designing a nuclear fusion reactor

Poloidal plane of a Tokamak system



Plasma operation condition

Reactor wall configuration

Basic concept of our research



Reward = R(cost-params) + R(beta) + R(safety-factor) + R(density) + R(TBR) + R(bootstrap) + R(Q-factor)

Reward contains design objectives (constraints from plasma physics and nuclear engineering) and performances (Q-factor + TBR)

How to apply

- Definition of the custom rewards that represent the objectives (Cost + TBR + Operation limits + Q-factor)
- Training the policy network to learn the optimal policy to determine the best control parameters

Step 1: Development of a Fusion Reactor Design Simulator



Example of the simulation result

Major radius R : 4.630 m Minor radius a : 1.543 m Armour : 0.100 m Blanket : 1.276 m Shield : 0.100 m TF coil : 0.522 m total thickness : 3.541 m Magnetic field : 16.000 T Elongation : 1.700 Aspect ratio : 3.000 Thermal efficiency : 0.400 Electric power : 1000.000 MW TBR : 1.362 beta : 6.034 tau : 1.159 s Ip : 15.345 MA q : 3.855 f bs : 0.534 Q-parallel : 513.03 MW-T/m T_avg : 20.00 keV n_avg : 1.09x10^20 #/m^3 p avg : 8.39 atm Greenwald density : 2.051, operation density : 1.092 | 0 g-kink : 2.000, operation g : 3.855 | 0 Troyon beta : 6.307, operation beta : 6.034 | 0 Neoclassical f bs : 0.651, operation f bs : 0.534 | 0 Lawson nTau : 1.812 , operation n*Tau: 1.266 | X Cost params : 0.757

Step 2: Implementation of a RL based Design Optimization Code



Decision making = Optimal design parameters estimation

Reactor structure and configuration

Default structure design and configuration – fixed during the optimization



• Main concepts for our fusion reactor

- Liquid type blanket: Coolant + Tritium breeding + relatively simple design
- Avoidance of the operation limits + High Q-factor + TBR > 1 + Low volume for reducing cost
- D-T fusion + ECRH & LHCD heating + H-mode plasma

Core Design

Plasma core profile: n, T, P

Profile effects on high Q subignited Tokamak Fusion Plasmas

- The Q-value clearly expresses its dependence on the radial profile of T,N,P
- Same configuration for radial profiles of T,N,P referred from Friedberg et al.
- Parabolic profiles with numerical coefficient

$$T = \bar{T}(1 + \nu_T)(1 - \rho^2)^{\nu_T} = 2\bar{T}(1 - \rho^2), \qquad \nu_T = 1$$

$$p = \bar{p}(1 + \nu_p)(1 - \rho^2)^{\nu_p} = 2.5\bar{p}(1 - \rho^2)^{3/2}, \qquad \nu_p = 3/2$$

$$n = \bar{n}(1 + \nu_n)(1 - \rho^2)^{\nu_n} = 1.5\bar{n}(1 - \rho^2)^{1/2}. \qquad \nu_n = 1/2$$

Effect of elongation

• Plasma current and elongation

Both of these advantages have been demonstrated experimentally up to elongations of k = 2.35: Set initial K = 1.7

Vertical elongation of the plasma cross-section brings the increased plasma current



 $\beta_m \leq \beta_n \frac{\eta_p}{q_n^R}$

- Due to the increase of plasma current, the energy confinement time can also increase
- Beta and elongation
 - Vertically elongated and D-shape cross-section allow much high beta values than circular ones
 - Troyon beta limit: High elongation plasma allow the beta value higher





Fig 1. From F.Hofmann, Physical Review Letters, Volume 81, 14, 1998

Blanket Design

Breeder design and material

Materials : liquid-type

- Lithium leads (Li17Pb83)
- Versatile: Coolant + Tritium breeding
- Lead : Neutron multiplier
- Corrosion: middle
- Chemical stability: less than other liquid materials
- Inboard shield thickness reduction / but the thickness of the blanket may increase
- $\circ~$ Structures for additional coolant system would not be necessary

Structures

- Structure type of TBM: CLAM designed by EU
- \circ Breeder Inlet / Outlet: 300 / 480
- Breeder Maximum temperature: 543 + 273 K -> High enough Pb-17Li Inhet pipe+manifold
- o Breeder coolant pressure drop : less than 0.8 MPa
- TBR : 0.43 -> using lead as a multiplier



TABLE I. Physical properties of eutectic LiPb compare							
with other possible breeding materials.							

	n ouner poss	lore creeding		
Liquid Breeder	Li	Li ₁₇ Pb ₈₃	Flibe	Li ₂₀ Pb ₈₀
Melting Point (°C)	180	235	459	320
Density (g/cm ³) 873K	0.48	8.98	2.0	6.0
Li Density (g/cm ³) 873K	0.48	0.061	0.28	0.09
Breeding property	Good	Fairly good	Neutron multipler required	Neutron multipler required
Chemical stability	Active	Middle	Almost stable	Almost stable
Corrosion	Severe	Middle	HF exist severe	?
Tritium release form	HT, T ₂	HT, T ₂	HT,T2 TF	HT, T ₂
Tritium solubility (atom fracPa ^{-0.5} T=873K)	Very high 7.49x10 ⁻³	Very low 1.93x10 ⁻⁸	Very low HT/T ₂ 1.77x10 ⁻¹¹ TF 1.77x10 ⁻¹¹	Middle 2x10 ⁻⁷ - 1x10 ⁻⁵
Tritium diffusivity order (m ² /s) (873K)	Relatively high 10 ^{.9}	Relatively high 10 ^{.9}	Relatively high 10 ⁻⁹	Relatively high 10 ^{.9}
Thermal conductivity]	L1>L1 ₂₀ Sn ₈₀ > L1	i ₁₇ Pb ₈₃ > Fli	be
Dynamic viscosity		Flibe>Li ₂₀ Sn ₈₀	~Li ₁₇ Pb ₈₃ >I	Li

Table 2. From A.Fraile, Molecular dynamics simulations of lead and lithium in liquid phase

Armor and Shielding

Armor design and material

- Material: Tungsten
 - o Thickness : 10 cm
 - o Good resistance for physical, chemical sputtering
 - o Tungsten exhibit pronounced surface morphology changes under He plasma exposure
 - High radiational loss can be occurred due to high atomic number and induce instabilities
 - Tungsten surface energy limit : 0.5MJ/m^2

Shielding design and material

- Material: Graphite
 - Thickness : 10cm
 - \circ Neutron reflector and moderator for shielding the materials (TF coil)
 - $\circ~$ Located between the blanket and TF coils
 - $\circ~$ Good neutron-moderating properties and availability in large quantities
 - $\circ~$ Neutron reflection outside the blanket : enhancing TBR

Table 2 Typical densities of carbon and gray	phite products.	Table 3 Vaporization data for carbon [20].		
Material	Bulk density, g/cm ³	Temperature, K	Pressure, atm	
Pyrolytic carbon	1.2-2.2	2000	1.31 × 10 ⁻¹⁰	
Single crystals (theoretical density)	2.26	2500	1.05×10^{-6}	
Nuclear graphite-moderator and reflector	1.5-1.7	3000	5.38 × 10 ⁻⁴	
Porous electrocarbons and graphites	0.6-1.3	3500	5.30×10^{-2}	
Carbon felt	0.08-0.17	3800	1.0	









Fig 2. From D.E.Baker, Graphite as a neutron moderator and reflector materials

TF coils and PF coils





Common steel Ref. c = 0.97 [m] Maximum Tensile Strength : 600 [MPa] Maximum current density : 20 [MA/m²]

2800 Maraging Steel Ref. c = 0.66 [m] Maximum Tensile Strength : ~ 2700 [MPa]

HTS coil Ref. c = 0.44 [m] Maximum current density : ~ 200 [MA/m²]

TF, PF coils CAN BE MADE THINNER !

Heating and Current Drive Sources

Heating and Current Drive – ECRH and ICRH

ECRH

\circ Functions

- o Plasma startup
- \circ $\,$ Heating for access to H-mode $\,$
- \circ q profile control
- \circ $\,$ CD for steady-state $\,$
- \circ Requirement
 - Frequency : ~200GHz
 - \circ $\$ Power : determined by tokamak design
 - o Gamma : 0.15

LHCD

- \circ Functions
 - \circ $\;$ Current is driven by lower hybrid waves in our cases
 - \circ High efficiency
- \circ Requirement
 - \circ $\,$ Considered coupling of the edge plasma and generator power : 0.4 $\,$
 - $\circ~$ P > 4MW -> no experiment has ever coupled ICRF power into an H-mode
 - $\circ \quad \text{Efficiency of LH klystrons}:~~50\%$
 - \circ $\$ Power : determined by tokamak design
 - o Gamma : 0.3 0.4

$$\begin{split} \eta_{CD} &= \frac{R_0 \bar{n} I_{CD}}{P_{CD}} \approx \frac{1.2}{n_{\parallel}^2} \ ,\\ n_{\parallel} &\approx \frac{\omega_{pe}}{\Omega_e} + \left(1 + \frac{\omega_{pe}^2}{\Omega_e^2}\right)^{1/2} \left(1 - \frac{\omega_{LH}^2}{\omega^2}\right)^{1/2} \end{split}$$

NNBI

ICRH

LHCD

ECRH Low-Medium (20-40%)

$$I_{dep}$$

$$I_{de$$

high

major radius R

deposition power of the beam P_{\perp}

low-medium

 $\eta_{CD} = \frac{I_{NBCD} n_e R}{P_e}$

Finding the optimal engineering parameters

Design optimization with single-step reinforcement learning

Single-step reinforcement learning and design optimization

- Policy gradient method to find out the best design choice of the tokamak
- Control parameters: H, armor thickness, T, betan, elongation, aspect ratio, B-field, RF recirculated efficiency, electric power
- Conventional optimization method: hard to find out the optimal configuration (Multi-objective + High dimensional space)



						1 episo
Initial state	Observation	Agent	Actions	Environment	Reward	Agent

Figure 4: Degenerate reinforcement learning framework used in this paper. One episode consists of a single control from the agent: the same initial observation is provided to the agent at the start, which in return provides an action to the environment. The environment returns a reward value to the agent, and the episode is terminated.

By customizing the reward function, we can optimize our tokamak design with respect to any design strategy!

Reward = R(cost-params) + R(beta) + R(safety-factor) + R(density) + R(TBR) + R(bootstrap) + R(Q-factor)

How to apply

- Definition of the custom rewards that represent the objectives (Cost + TBR + Operation limits + Q-factor)
- Training the policy network to learn the optimal policy to determine the best control parameters

Finding the optimal engineering parameters

Reward engineering for satisfying stabilities + maximum performance

□ <u>Stability</u>

- Kink instability
- Troyon beta limit
- Greenwald density limit

Design Performance

- High beta
- Low cost (proportional to volume)
- Bootstrap current ratio

$$R_{t} = a \times sigmoid\left(\frac{x}{x_{limit}} - 1\right) + R_{fail} \times \theta(1 - \frac{x}{x_{limit}})$$

$$R_t = a \times sigmoid\left(\frac{x}{x_{scale}} - 1\right)$$

Design Optimization Results

Comparison between the reference (Friedberg) and the optimized reactor

Major radius R : 5.346 m Minor radius a : 1.337 m Armour : 0.000 m Blanket : 0.899 m Shield : 0.100 m TF coil : 0.954 m total thickness : 3.289 m | Magnetic field : 13.000 T Elongation : 1.700 | Aspect ratio : 4.000 Thermal efficiency : 0.400 Electric power : 1000.000 MW TBR : 1.153 beta : 3.595 tau : 0.917 s Ip : 14.005 MA q : 3.398 | f bs : 0.612 Q-parallel : 550.25 MW-T/m T avg : 14.00 keV | n avg : 1.43x10^20 #/m^3 p avg : 7.67 atm Greenwald density : 2.495, operation density : 1.426 | 0 q-kink : 2.000, operation q : 3.398 | 0 Troyon beta : 4.007, operation beta : 3.595 | 0 | Neoclassical f bs : 0.464, operation f bs : 0.612 | X Lawson nTau : 2.507 , operation n*Tau: 1.308 | X Cost params : 0.756

Major radius R : 4.630 m Minor radius a : 1.543 m Armour : 0.100 m Blanket : 1.276 m Shield : 0.100 m | TF coil : 0.522 m total thickness : 3.541 m | Magnetic field : 16.000 T Elongation : 1.700 Aspect ratio : 3.000 Thermal efficiency : 0.400 Electric power : 1000.000 MW TBR : 1.362 beta : 6.034 | tau : 1.159 s Ip : 15.345 MA | q : 3.855 | f bs : 0.534 | Q-parallel : 513.03 MW-T/m | T avg : 20.00 keV n avg : 1.09x10^20 #/m^3 | p avg : 8.39 atm | Greenwald density : 2.051, operation density : 1.092 | C | q-kink : 2.000, operation q : 3.855 | 0 Troyon beta : 6.307, operation beta : 6.034 | 0 Neoclassical f_bs : 0.651, operation f_bs : 0.534 | 0 Lawson nTau : 1.812 , operation n*Tau: 1.266 | X Cost params : 0.757 _____



Design Optimization Results

Comparison between the reference (Friedberg) and the optimized reactor

