深層強化学習による噴流の混合制御

Mixing Control of Free Jet using Deep Reinforcement Learning

○ 辻本 公一，三重大学大学院, 三重県津市栗真町屋町 1577, E-mail：tuujimoto@mach.mie-u.ac.jp
田ノ上 飛翔，三重大学大学院, 三重県津市栗真町屋町 1577, E-mail：419d051@m.mie-u.ac.jp
Koichi Tsujimoto, Mie University, 1577 Kurimamachiya-cho, Tsu, Mie
Tsubsa Tanoue, Mie University, 1577 Kurimamachiya-cho, Tsu, Mie

In order to develop a new jet mixing procedure, we explore the possibility of DRL (deep reinforcement learning). First, we conduct some cases of open-loop control in which a main jet is manipulated by a pair of sub jet being actuated at the inlet of main jet, and examine the effect of actuating frequency on the mixing performance of main jet. Then, we select a DDPG (deep deterministic policy gradient) scheme among of the present DRL schemes, and apply it to the above-mentioned jet control problem. Compared the the results of DDPG with that of open-loop control, the DDPG scheme turns out the useful performance of jet mixing control, *i.e.*,the entrainment of fluid from surroundings is enhanced through the DRL while the main jet behaves like a flapping jet.

1. 緒言

工学機器における混合・拡散を促進するため, 噴流の制御手法の検討が行われている．これまでに受動的手法（非円形ノズル, リブ, タブ, 同軸噴流, 共鳴噴流など）[1] と能動的手法 (音響励起, ノズルの回転, 微小噴流の吹き付けなど）に関する多くの制御手法 [2]~[8] が提案されている．これらの制御目標の多くは混合を活発にするため，噴流を下流方向に著しく拡散させることを目指している．上記の能動制御の場合，制御方法として噴流における流れの不安定性を積極的に利用することに主眼が置かれてきた．軸対称噴流の場合, ノズル近傍領域の大規模構造を特徴付ける不安定なモードは，周方向に一様な軸対称モード (varicos) とヘリカル (helical) モードであり，この 2 つが支配的であることは線形安定解析の結果から明らかにされている．これらのモードを励起することにより噴流の拡散や混合を制御し [2][3]，さらにこれらの基本モードの組み合わせでより複雑な噴流を作ることができる．例えば 2 つのヘリカルモードの周期と振幅を同じにして励起すると フラッピング (flapping) モードが生じる. さらにこのモードに軸対称モードを追加し, 両モードの励起周波数を変えると分岐 (bifurcating) モードや開花 (blooming) モードなどの噴流制御の可能が実験で確かめられている [2]. これら能動制御は DNS(direct numericalsimulation) でも検討され [4]~[6], 強い噴流拡散を生じることが再現されている. 一方，不安定モードを励起するのではなく噴流を幅広く噴射する方法として，流体素子をノズル部に備え，ノズル出口で噴流を発振させる方法 [7] や，これらの発振噴流に機械的な旋回を加えた方法 [7] も提案されている．能動制御については，人為的にパラメータを設定せざるを得ないことからさらに機能向上するにはあらたな方策が必要である．例えば，随伴法による最適化手法によれば，様々な機能向上のパラメータを見出すことが可能であり，乱流制御において適用され，有効な知見が見出されている [10]．しかし，随伴方程式を解くことの処置に関するコーディングの手間，随伴方程式を解く際に必要な流れ場データの保存，計算時間の長大化，また，初期値依存性の問題もある．

最近，機械学習に注目が集まっている．関連する書籍 [11] も多数出版されるようになり，分野外の研究者でも比較的知識を吸収しやすい環境が整ってきている．その機械学習の分野のうちでも，正解のない，教師なしの問題に対応できる強化学習（RL: reinforcement learning）に対し，ニューラルネットワークを組み込み込んだ深層強化学習 (DRL : deep reinforcement learning) は従来の最適制御の抱える問題を緩和できる可能性がある．最近，その可能性を示唆する成果として，小泉ら [12] により円柱周りの揚力低減に DRL を適用，興味深い特性が明らかにされている．

本研究では，以上の背景を受け，噴流の混合性能を向上させるために深層強化学習を利用する．具体的には，2 次元噴流の噴流出口に，アクチュエータとして一対の副噴流を配置し，このアクチュエータを深層強化学習により，学習，同時に開ループ制御した結果と比較し，DRL の有効性について評価した結果を報告する．

2. 計算方法

2.1 基礎式および離散化

支配方程式として，2 次元での非圧縮場を仮定した連続の式，運動方程式，エネルギー方程式を用いる．

$$\frac{\partial u_i}{\partial x_i} = 0 \tag{1}$$

$$\frac{\partial u_i}{\partial t} + h_i = -\frac{\partial p}{\partial x_i} + \frac{1}{\mathrm{Re}}\frac{\partial^2 u_i}{\partial x_j \partial x_j} \tag{2}$$

$$(h_i = \epsilon_{ijk}\omega_j u_k,\ \omega_j : \text{vorticity})$$

$$\frac{\partial T}{\partial t} + u_i\frac{\partial T}{\partial x_i} = \frac{1}{\mathrm{RePr}}\frac{\partial^2 T}{\partial x_i \partial x_i} \tag{3}$$

ここで，対流項は回転型とし，代表長さに噴流出口直径 D，代表速度に主流方向速度 V_0 を用いて無次元化した. 計算領域は長方形領域とし，主流方向を y，主流と直交する方向を x，z とする．空間の離散化は，主流方向に 6 次精度の Compact Scheme[9] を主流と直交する方向には sin，cos 級数展開を用いた．また，時間進行には 3 次精度の Adams-Bashforth 法を用いた．

2.2 計算条件

主流方向の流入速度分布 $V_m(r,t)$ を示す

$$V_m(r) = \frac{V_0}{2} - \frac{V_0}{2}\tanh\left[\frac{1}{4}\frac{R}{\theta_0}\left(\frac{r}{R} - \frac{R}{r}\right)\right] \tag{4}$$

ここで，V_0 は噴流中心速度，$R(=D/2)$ は噴出口半径，θ_0 は初期せん断層の運動量厚さ，r は各噴流中心軸からの半径方向距離である．更に，混合制御のために上述の噴流出口に 1 組の補助噴流を配置する．レイノルズ数は $\mathrm{Re}=V_0D/\nu = 100$，プラントル数は $\mathrm{Pr}=0.707$ とする．計算領域は主流と直交する方向は $H_x = 15D$，主流方向に $H_y = 20D$ とする．さまざまな状況下で，制御アルゴリズムの特性を調査するため，可能な限り少ない格子数での計算行うため，格子数は x, y 方向それぞれに 64×101 とした．

2.3 深層強化学習

強化学習では，環境，エージェント，行動，報酬という概念を組み合わせて学習が行われる．すなわち，エージェントがある環境下で，ポリシー（政策）に基づいて行動選択し，状態と報酬を得る．環境は，そのエージェントからの行動により状態を変化させ，それに応じてエージェントに報酬を与える．そして行動による報酬を通して報酬を最大化する政策を見つけ出す．強化学習では，行動，状態，報酬についてマルコフ過程を仮定する．したがって，ある状態での行動は次の状態へ状態遷移確率で遷移することを仮定する．以上についてより厳密で丁寧な説明は先に挙げた書籍 (11) や文献 (12) を参考いただきたい．これらの概念は分野外の研究者には理解しづらいが，例えば，今回の研究において対応させると，行動は瞬時ごとのアクチュエータである副噴流の噴出速度，方策はアクチュエータの駆動のさせ方，環境は，噴流の状況をモニタリングしている場所の速度となる．さらに本研究では，混合性能を高めることが目的であることから，報酬として主噴流によるエントレインメントを考えることができる．

深層強化学習のスキームは日進月歩で発展しているが，先ほどの文献から，ゲームなどとの学習とは違い、今回扱う研究対象が連続的な行動の取扱いをすることから，Deep Deterministic Policy Gradient (DDPG)(13) を採用する．

DDPG では Critic ネットワーク，Actor ネットワークの 2 つから成り，これらは相互に補完しながら学習をすすめる．ここでは，オリジナルの方法通りの DDPG のアルゴリズムを実装した．

DDPG(13) のアルゴリズム

(1) Critic ネットワーク $Q(s,a|\theta^Q)$ と Actor ネットワーク $\mu(s|\theta_\mu)$ のそれぞれのネットワークの重み θ_Q, θ_μ に乱数を与えて初期化する．

(2) 上記それぞれのターゲットネットワーク Q',μ' の重み $\theta_Q{}'$, θ'_μ を θ_Q, θ_μ とそれぞれ同じにして初期化する．

(3) エピソードを開始する．その際，初めに乱数を行動 a に与え，初期の状態 s_1 を与える．

(4) 以下の (a)∼(h) を繰り返し，時間ステップ t を 1 から T まで進める．

(a) 現在のポリシーに従い，ノイズを与えて，時刻 t での行動 a_t を選ぶ．
$a_t = \mu(s_t|\theta^\mu) + \mathcal{N}_t$

(b) a_t を実行し，報酬 r_t ならびに新しい状態 s_{t+1} を得る．

(c) バッファに状態の組 (s_t,a_t,r_t,s_{t+1}) 保存する．

(d) バッファからランダムに N 個の状態の組を抽出する．

(e) $y_i = r_i + \gamma Q'(s_{i+1},\mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$ を計算する．

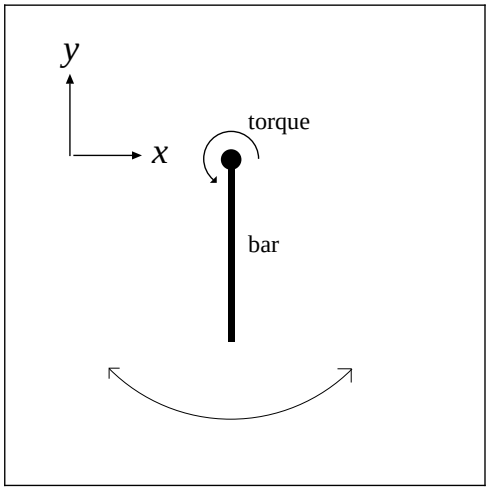
(f) 損失 L を最小化して Critic ネットワークを更新する．

$$L = \frac{1}{N}\sum_i (y_i - Q(s_i,a_i|\theta^Q))^2$$

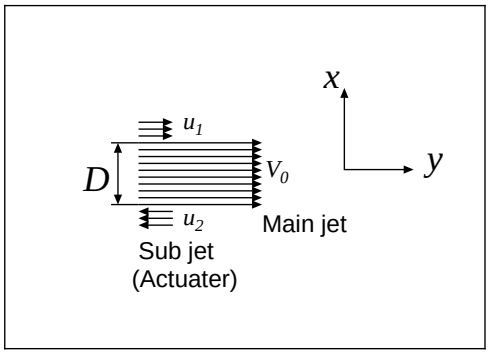
(g) サンプルされたポリシーの勾配を用いて Actor ネットワークを更新する．

$$\nabla_{\theta^\mu} J \approx \frac{1}{N}\sum_i \nabla_a Q(s_i,a|\theta^Q|_{s=s_{i,a=\mu(s_i)}} \nabla_{\theta^\mu}\mu(s|\theta^\mu)|_{s_i}$$

$$\theta^\mu_{t+1} = \theta^\mu_{t+1} + \nabla_{\theta^\mu} J$$



(a) rotating pendulum



(b) active controlled jet

Fig. 1: Schematics of the problems being solved by machine learning

(h) ターゲットネットワークを更新する．
$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$

(5) 以上について (3) に戻り，次のエピソードを実行する．

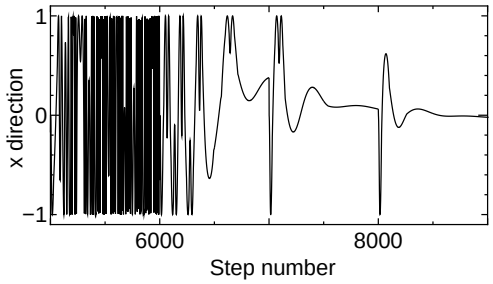
上記のアルゴリズムでは，明記されていないが，事前に設定しなければならないパラメータは多くある．実施にあたり，ニューラルネットワークの階層の数，ニューラルの数，活性化関数の選定，探索に必要なノイズの与え方（強度，種類，持続性），アルゴリズム中の学習率 γ，混和率 τ，報酬の内容など，学習が始まってしまうと計算機任せになるが，探索を飽和さないためにもこれらパラメータの選択を注意深く行う必要がある．

3. 計算結果

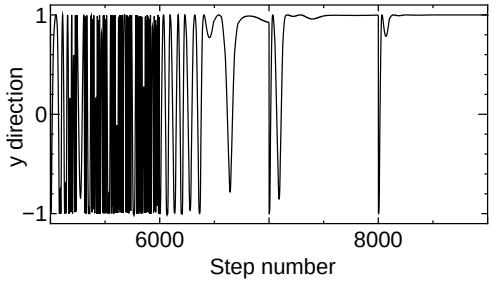
3.1 回転振り子による開発コードの検証

通常，深層強化学習を含む機械学習では，言語として Paython を用い，機械学習のソフトウエアとして Tensorflow などがたいてい利用される．しかしながら，学習中における環境として，流体解析コードが必要なことから，自前の解析コードとの親和性より，Fortran77 でのコード開発を行った．その妥当性の検証のため，ここでは，Fig. 1(a) に示すような回転軸周りに旋回する回転振り子について深層強化学習を行った．

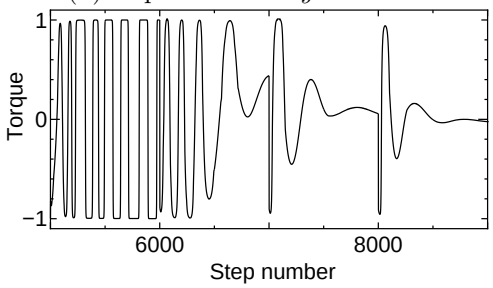
回転振り子では初期には棒は重力に従い，下を向いて静止している．その後，適当なトルクを与えて，棒を倒立させ，かつ静止させることが目標となる．深層学習では，恣意的に都合のよい条件は与えず，トルクの大きさ



(a) displacement of x direction



(b) displacement of y direction



(c) torque

Fig. 2: Motions of a rotating pendulum

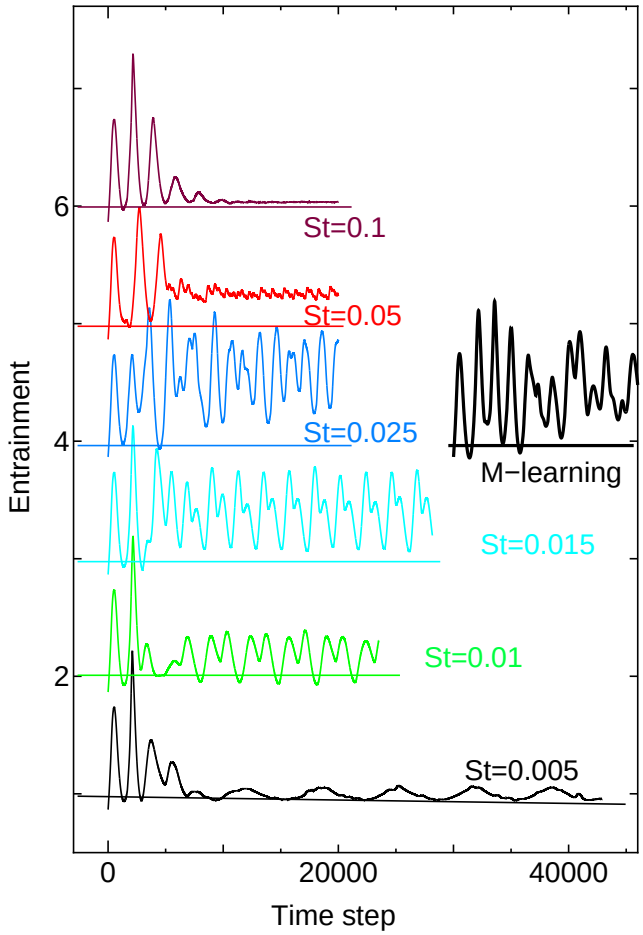


Fig. 3: Time evolution of entrainment at various condition. open-loop control (St = 0.005, 0,01, 0,.015, 0.025, 0.05, and 0.1); M-learning (DDGP) .

が指定する大きさを越えない条件のみを課す．また報酬として，トルクをできるだけ与えず，かつ上方に位置するように垂直方向位置ができるだけ大きくなるように報酬を設定する．多くの最適値の評価関数と同様，直接達成したい目標値と制約条件をペナルティとして混合する．Fig. 2 は深層学習した結果である．横軸はステップ数を示し，ここでは，一つのエピソードに 1000 ステップ割りてる．Fig. 2(a)(b) は，棒の先端位置の座標で，x 方向では $x = 0$ に y 方向には $y = 1$ となることが目標となる．図からもわかるように，6 エピソード目まではどのデータも極めて振動的な挙動を示している．これは Fig. 2(c) のトルクの挙動が原因で，トルクの on-off が連続的に繰り返され，off の状態では棒が空回りしていることに対応いている． 6 エピソード目の途中から学習の効果が突如表れ始める．7 エピソード以降では，各エピソードが開始されるとトルクが特徴的な滑らかな山と谷の分布をとり，速やかに目標状態が達成される．このような，機械学習特有の突如収束する特性や，深層学習の効果による短い学習での目標状況への到達の様子から，開発したコードが，深層強化学習の研究に対応できる性能を有していることが確かめられた．

3.2 噴流混合

噴流の混合性能を高めるために深層強化学習の可能性を探る．今回行う噴流の混合制御は文献 [14] を参考に，噴流噴出口の両側に副噴流を設置する (Fig. 1(b))．各副噴流は逆位相で駆動する．そのことにより適当なアクチュエータの条件により主噴流が上下方向にフラッピングを生じさせることで噴流混合を活性化する．そこで，混合性能として噴流自体の周囲流体を巻き込み量である下流の境界における主流方向速度の積分値を混合性の目安とした．

$$S = \int_0^{H_x} v|_{y=H_y} dx \tag{5}$$

3.2.1 開ループ制御 学習により得られた結果を評価するため，副噴流の噴出速度を正弦波上に入力する開ループ制御を行う．その際，速度振幅は一定（$u_0 = 0.2$) として周期的に変化させた．

$$u_1 = -u_2 = u_0 sin(\omega t) \tag{6}$$

Fig. 3 に開ループで制御された積分値 S の時間発展を示す．図中の記号 St は主噴流の流入速度，主噴流径で無次元化したストローハル数である．また図中の横線は積分量 S が 2 となる線を表しており，条件間での性能を比較する上での目安となる．噴流は落ち着いた状態から，突如，アクチュエータが作用するため，すべての場合にいて，アクチュエータ印加後，強いエントレインメントが生じている．また積分位置である下流断面では，渦が計算領域外に放出される際に主流とは逆向きの強い流れが生じることから計算領域での下流長さも短いことも影響し，振動的な結果を生んでいる．準定常的な挙動に落ち着

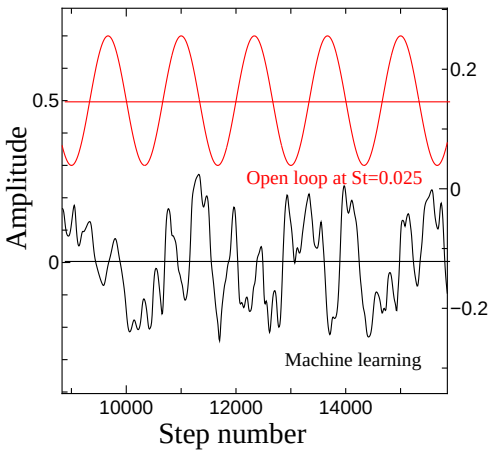


Fig. 4: Amplitude of actuator input

いた以降で積分量を比較すると，極めて低い St = 0,005 から St = 0.025 当たりに向けて積分量は増大し，それ以降，ここには図示していないが，より低い積分量となる．文献 [14] でも同様の開ループ制御を行い，その際，最適な St 数として St = 0.03 を検討した．その値は本研究で得られた結果ともほぼ一致している．また，噴流のプリモード St = 0.3∼0.5 と比べると非常に低い St 数であることに気づく．さらに噴流の混合が最も活性化している St = 0.025 では，他の条件のように準定常状態には移行していない．

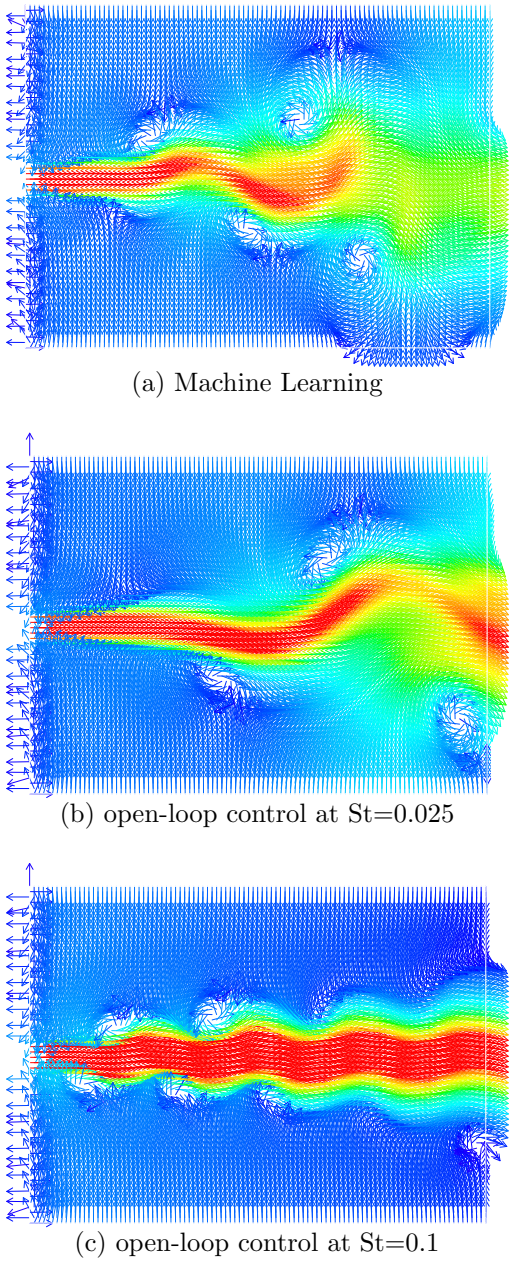
3.2.2　DDPG による制御　アクチュエータと混合が起こる場所とは離れており，したがって観測点での状態と行動には時間遅れが生じる．小泉ら [12] は時間遅れが生じる場合，時間遅れのあるマルコフ決定過程の取り扱いを指摘し，観測点での時間遅れした情報とさらに行動の時間遅れを状態量に組み込む方法を提案した．

本研究では，噴流の主流方向が明確であることから，時間遅れの情報の代わりに下流位置で複数の観測点を配置することで対応する．具体的には，噴流中心軸から両側の出口半径の距離となる位置についてそれぞれ，下流にむけて $2D$ ごとに計 20 点の観測点を配置した．学習による行動の切り替えは噴流計算の 1 ステップの 10 倍である無次元時間で $\Delta t = 0.3$ とし，1 エピソードに学習に 2 万ステップ，60 エピソード程度で学習を終了させた．アルゴリズムに記載されているパラメータについては，$\gamma = 0.9$, $\tau = 0.1$,　バッファから抽出する状態数を 32 個とした．ニューラルネットワークについては，Actor ネットワークの入力層は 20 ノード，出力は行動の 1 ノード，隠れ層は第 1 層が 120 ノードの全結合層，第 2 層が　240 ノードの全結合層である．活性化関数は Rectified Linear Unit (ReLU) を選んだ．Critic ネットワークの入力層は状態を 20 ノードと行動を 1 ノード，同様に 120 ノード，240 ノードの全結合層が隠れ層としてあり，それぞれ ReLU を活性化関数とした．報酬の与え方は小泉ら [12] を参考に，次式のように与える．すなわち，エントレインメントを強化する右辺第 1 項に加え，作用点の入力速度が出来るだけ小さいことが望ましいので，右辺第 2 項をペナルティ項として報酬に加えている．更に，作用点の速度変化が激しいことは物理的に望ましくないことからも右辺第 3 項を加える．

$$r_t = C_1 * (S/S_{max} - 1)) - C_2 .* {a_t}^2 - C_3 * (\Delta a_t)^2 \qquad (7)$$

ここで $C_1 = 1000, C_2 = 500, C_3 = 10, S_{max} = \pi/4$ とする．

Fig. 3 に開ループ制御の結果と DDPG の結果を比較する．DDPG は 63 エピソードの時の結果で，波形の違



(a) Machine Learning

(b) open-loop control at St=0.025

(c) open-loop control at St=0.1

Fig. 5: Instantaneous velocity vector plot at the end of each computation. open-loop control (St = 0.005, 0,01, 0,.015, 0.025, 0.05, and 0.1); M-learnin (DDGP) .

いがあるが定量的には，最適な開ループ制御の結果と遜色のない結果が得られている．この時のアクチュエータの入力波形を Fig. 4 に示す．また，改めて強調すべきではないかもしれないが，学習により開ループの単調な入力と比べて複雑な波形を生じる．ここには図示しないが，調査の過程で，アクチュエータの出力を抑えるペナルティ項の重みを高めると，波形は比較的単調な波形が得られた．一方，エントレインメントが抑えられ気味となる．この点からも，報酬は解に強い影響を持ち，学習したからと言って，最適な解を決定してくれるわけではないことに改めて気づく．最後に，流れ場の様子について確認する．Fig. 5 に開ループと機械学習した場合の結果を比較する．エントレインメントが活発になったパラメータでは大規模なフラッピングが生じ，周囲流体を巻き込むような渦構造が複数存在していることを確認でき

る．一方，不活性な場合，噴流は脈動するように下流に向けて変動している．機械学習により明らかに流れ構造が変調しており，生じている構造は，open-loop 制御と類似の構造をとる．

4. 結言

深層強化学習により噴流の混合制御を行った．学習により効果的な混合制御が行わることを明らかにしたが，これらの結果を得るまでには，複数のパレメータの影響を探りながら調整するなど，一定の成果が出るまで相当手間取った．振り子のような目標が明確なものについては，結果の把握は容易であるが，せん断流の制御では，どの時点で結果で納得するのか，さらに中高レイノルズ Re 数への適用への可能性など検討すべき課題は山積している．

参考文献

(1) Shakouchi, T., “Jet Flow Engineering - Fundamentals and Application-,” Morikita Pub. Co., Ltd.(2004)(in Japanese).

(2) Reynolds, W. C., Parekh, D. E., Juvet, P. J. D. and Lee, M. J. D., “Bifurcating and Blooming Jets, ”*Annu. Rev. Fluid Mech.* (2003), pp.295-315.

(3) Longmire, E.K. and Duong, L. H., “Bifurcating jets generated with stepped and sawtooth nozzles,” *Phys. of Fluids* **8** (1996), pp.978-992.

(4) Danaila, I. and Boersma, B. J., “Direct numerical simulation of bifurcating jets, ”*Phys. of Fluids* **12** (2000), pp.1255-1257.

(5) Hilgers, A. and Boersma, B. J., “Optimization of turbulent jet mixing, ”*Fluid Dyn. Res.* **29** (2001), pp.345-368.

(6) Silva, C. B. and Metais, O., “Vortex control of bifurcating jets: A numerical study,” *Phys. of Fluids* **14** (2002), pp.3798-3819.

(7) Nathan,G.J., Mi, J., Alwahabi, Z.T., Newbold,G.J.R. and Nobes, D.S., Impacts of a jet’s exit flow pattern on mixing and combustion performance, ”*Prog. Eng., Comb.,Sci.*, **32** (2006), pp.496-538.

(8) Yeh, Y.L., Hsu, C.C., Chiang, C.H. and Hsiao, F.B., “Vortical structures evolutions and spreading characteristics of a plane jet flow under antisymmetric long -wave excitation, ”*Exp. Therm. Fluid Sci.*, **33** (2009) pp.630-641.

(9) Lele, S. K., “Compact finite difference schemes with spectral-like resolution,” *J. Comp. Phys.* **103** (1992), pp.16-42.

(10) Yamamoto, A., Hasegawa, Y. and Kasagi, N., “Optimal control of dissimilar heat and momentum transfer in a fully developed turbulent channel flow,” J. Fluid Mech., 733 (2013) , pp. 189-220.

(11) 曽我部,“強化学習アルゴリズム入門 「平均」からはじめる基礎と応用,” オーム社 (2019).

(12) 小泉, 堤, 嶋,“円柱カルマン渦列の制御における深層強化習の試行,” ながれ, 37 （2018）, pp. 161-170.

(13) Lillicrap, T.P., Hunt, J.J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., Wierstra, D., “Continuous control with deep reinforcement learning, ” Int. Conf. Learn. Represent. (2016) arXiv:1509.02971v5.

(14) Yuan, C.C.L., Krstic, M. and Bewley, T.R., “Active control of jet mixing,” IEE Proceedings - Control Theory and Applications, 151-6 (2004), pp. 763 - 772.