Custom Design and Control of Extreme Beams with Deep Reinforcement Learning

Auralee Edelen, Claudio Emma, Brendan O'Shea, Adi Hanuka, Spencer Gessner, Alex Scheinker, Glen White

FACET-II PAC Meeting
26-29 October *2020*

U.S. DEPARTMENT OF ENERGY
Office of Science

SLAC NATIONAL ACCELERATOR LABORATORY

Relationship to Other ML Application Areas

Note: all are collaborative, complementary efforts!

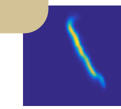
Prediction

surrogate models

virtual diagnostics

adaptively-tuned physics models

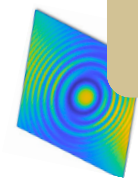
readings that are always available → ML model or tuned physics model → beam prediction



C. Emma, A. Edelen, A. Hanuka, B. O'Shea, A. Scheinker, S. Gessner, et al. E-327

Diagnostic analysis

e.g. edge radiation

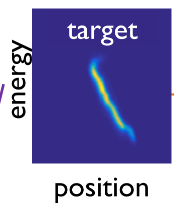


complicated beam diagnostic → ML-based analysis → value of interest (e.g. emittance)

B. O'Shea, C. Emma, A. Edelen, A. Scheinker, S. Gessner, et al. E-326

Control

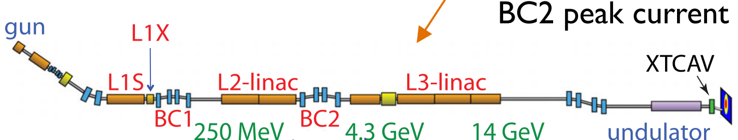
inverse model, then local feedback



ML inverse model → Suggested initial settings

L1S phase
BC2 peak current

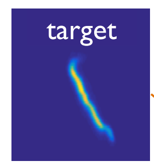
C. Emma, A. Edelen, B. O'Shea, G. White, A. Scheinker, et al. E-327

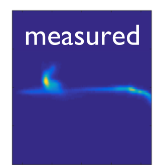


Local optimizer

A. Scheinker, S. Gessner, C. Emma, A. Edelen, B. O'shea, et al., E-325

reinforcement learning controller

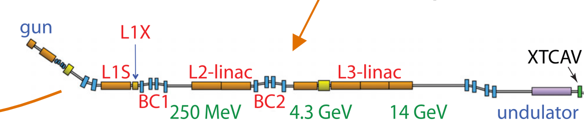




ML system model

Present Settings → ML control policy → Change in settings

Δ L1S phase
Δ BC2 peak current



A. Edelen, C. Emma, B. O'Shea, A. Scheinker, A. Hanuka, S. Gessner, et al.

Deep RL learns how to optimally control a system by interacting with it over time (builds understanding of system)

Pyramid of ML

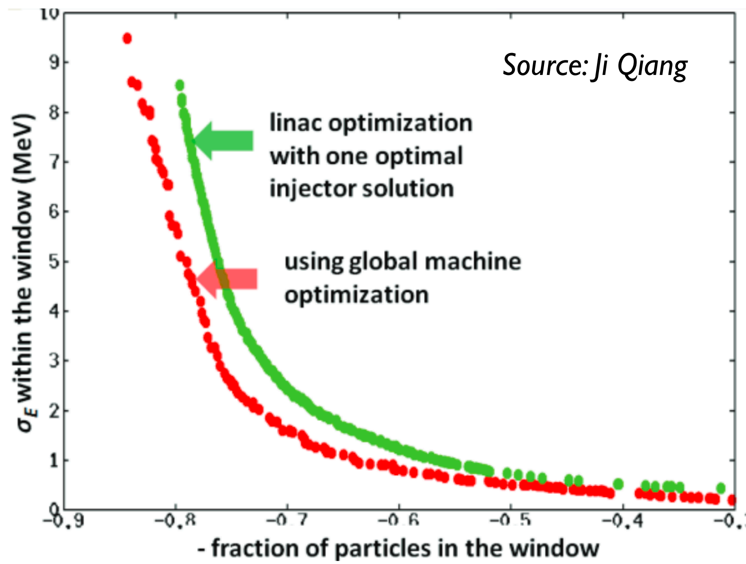
Machine Control and Understanding			
Diagnostics (Need information to make decisions)		Control (How to make decisions)	
Edge Radiation Emittance Diagnostics (E326)	Virtual TCAV Predictive Diagnostics (E327)	Adaptive Feedback (E325)	Learned Control (Reinforcement Learning, New proposal)
Non-destructive, single shot continuous monitoring of emittance of high-current beams	Longitudinal phase space diagnostics, always on, and for extremely short bunches	Stable, high-quality beams through control of unmodeled accelerator behavior	Fast, high-quality control of extreme beams by exploiting learned FACET-II responses

Synergistic experiments, individual success enhances all research

Science Motivation

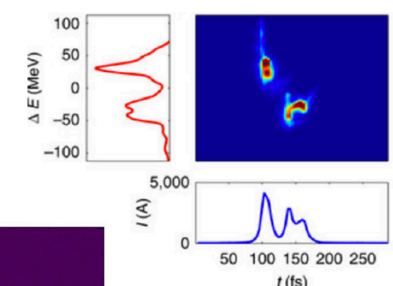
Major limitations in the way accelerator tuning is done:

- Piecemeal tuning of subsystems *(known to be sub-optimal)*
- Indirect use of high-dimensional diagnostics *(e.g. images)*
- Often a lack of accurate online models
  - → *Potentially limiting factors in control of extreme beams*

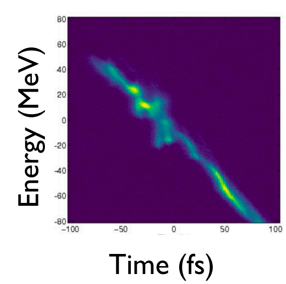


More global view can enable better control:

- Fully exploit unknown system-wide sensitivities + nonlinearities
- Faster switching between setups *(if using global representation of machine)*
- Better handling of parameter tradeoffs *(e.g. jitter, matching, longitudinal phase space)*



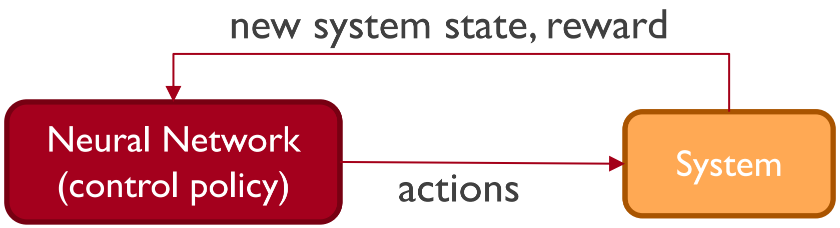
A. Marinelli, et al., Nat. Commun. **6**, 6369 (2015)



A. Marinelli, IPAC'18

Comprehensive, system-wide control is likely to be a key factor in improving custom control of extreme beams, but this is a difficult task

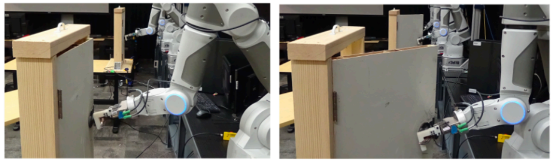
Deep Reinforcement Learning



- Control policy maps states to actions
- Policy is learned over time based on performance *(quantified by the "reward")*
- Neural network enables use of diverse signal types *(e.g. scalars, images, time series)*
- Often learns a system model simultaneously *(map states + actions to expected reward)*

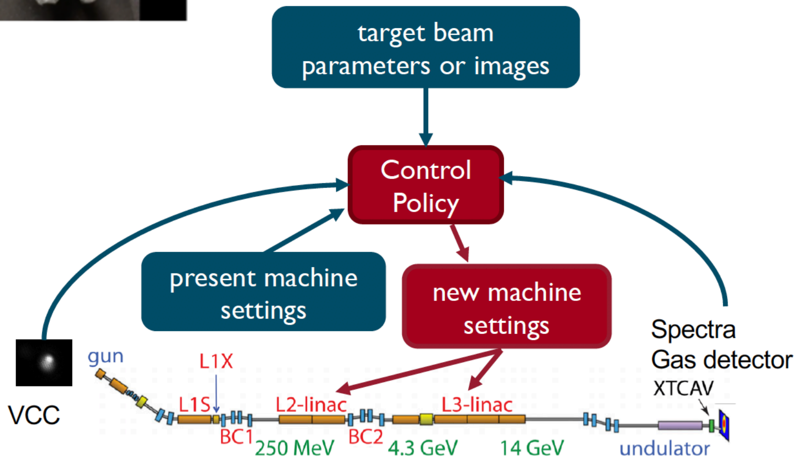
Nagabandi, et al., 2019

Gu, et al., 2016



Appeal for accelerator control:

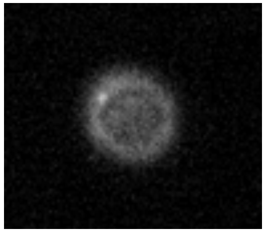
- Suitable for large, nonlinear systems
- Exploit machine-wide sensitivities + directly use complicated diagnostic information
- Leverage information from past observations
- Transfer between similar designs
- Well-established in other fields (e.g. robotic control) → but accelerators have unique challenges

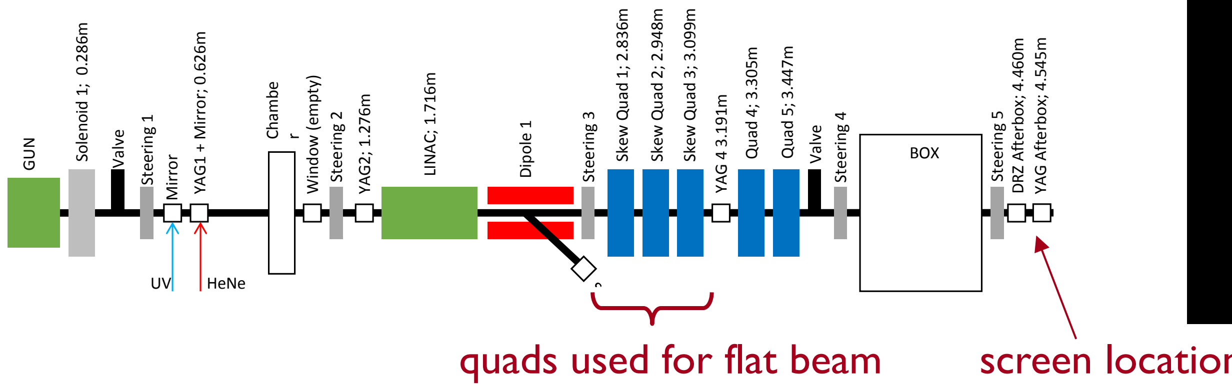


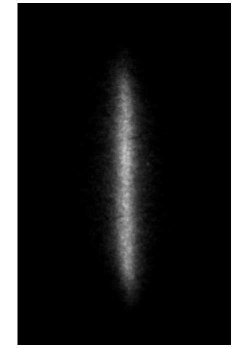
Deep RL is well-suited to accelerator control, but dedicated R&D is needed to bring it to full fruition

A. Edelen, E .Cropp, A. Hanuka, C. Emma, P. Musumeci, P. Denham, A. Scheinker

Recent Example



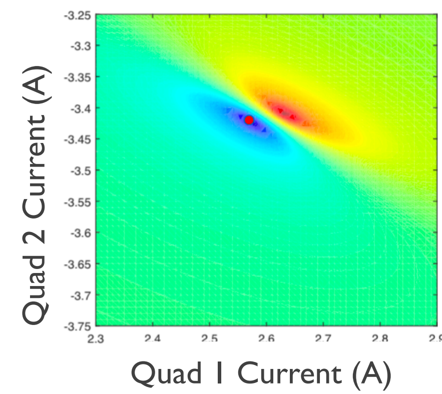




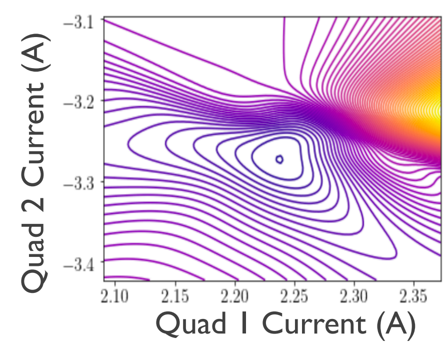
RL on the round-to-flat beam transform at UCLA Pegasus:

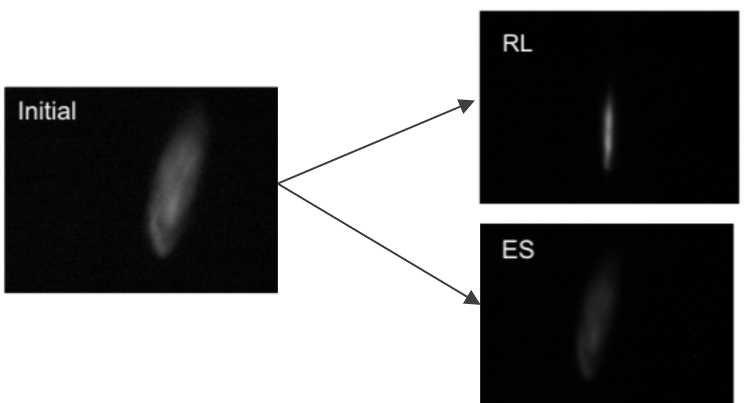
- Trained offline using learned model
- Transferred to machine for retraining
- *One trained, RL had fastest convergence compared with other methods*

Simulation

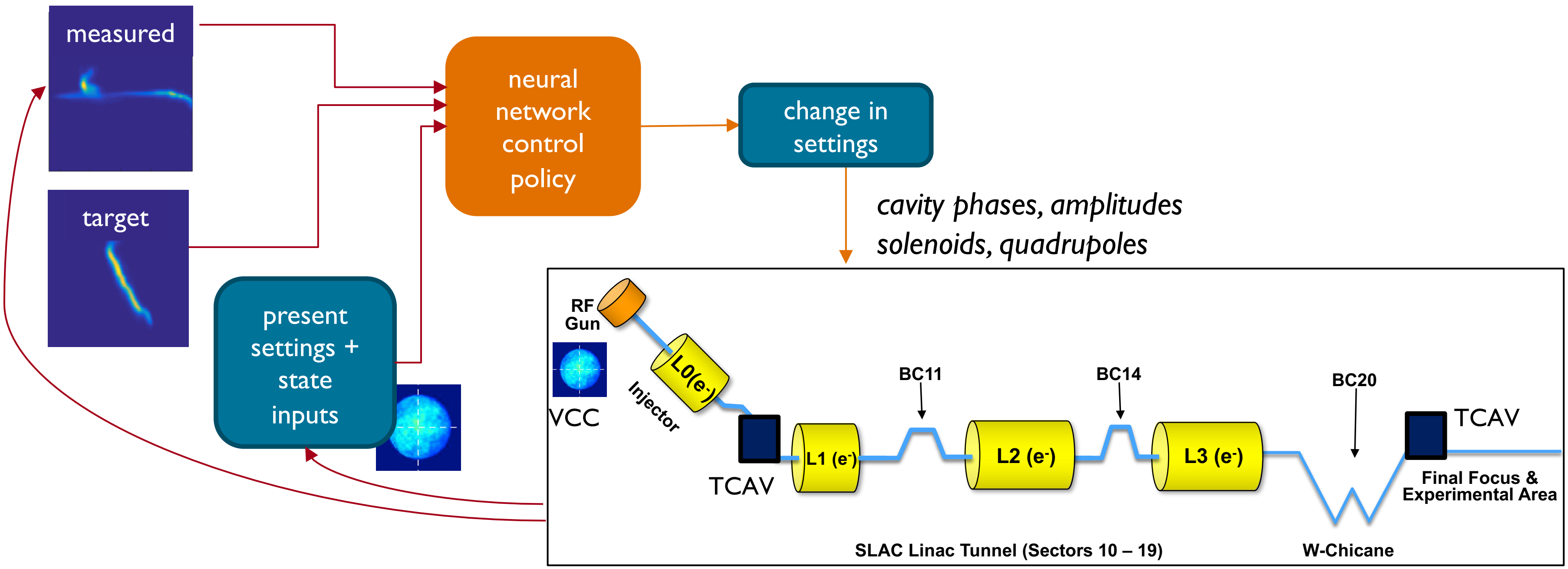


Neural Network





Conceptual Experimental Layout



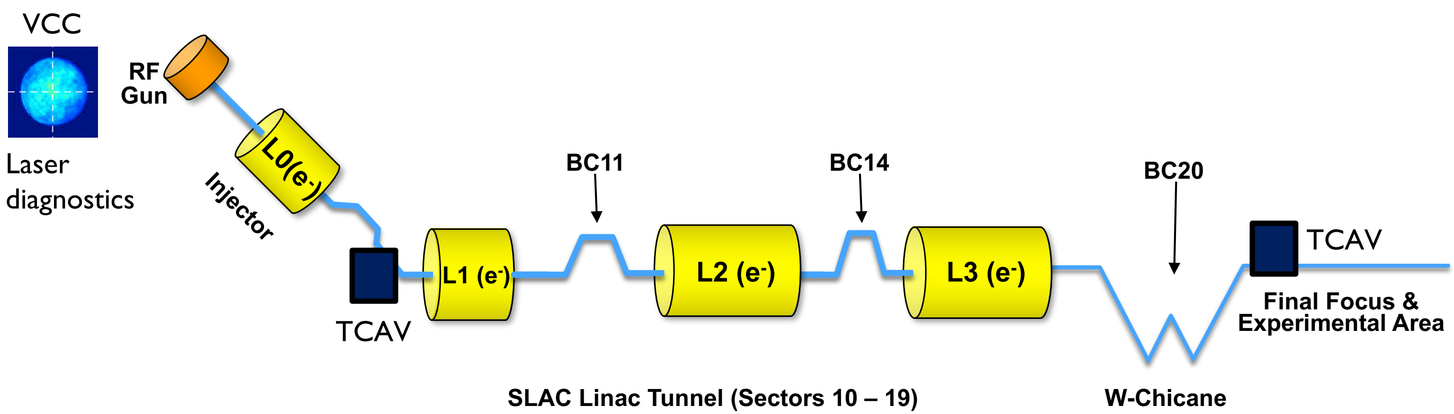
Three stages: injector, post-BC20, then combined control

Train controller in simulation, leverage passive data from other experiments in training, then test + retrain on machine

Diagnostics and Observables

- LPS diagnostics *(e.g. injector, downstream TCAVs)*
- Emittance measurements
- Upstream inputs: virtual cathode camera, QE map when available, etc.
- Readbacks from settings *(gun solenoid, gun and linac phases / amplitudes, etc.)*

→ *Can be flexible and adapt to installation / commissioning schedule*



Goals, Timeline, and Definition of Success

SLAC

Technical Goals:

Stage 0: Finish simulation campaign in Impact and Lucretia (by mid-November)

Stage 1: Demonstrate RL control for the injector emittance and longitudinal phase space (LPS) in single-bunch mode (6 months)

Stage 2: Demonstrate RL control for linac LPS in single-bunch mode (year 1)

Stage 3: Combined injector + linac control of LPS in single-bunch mode (year 1)

Stage 4: Extend to two-bunch mode (year 1.5 - 2) and other setups

→ *Definition of success: faster or higher-quality tuning than standard methods, new capabilities in control*

Infrastructure Goals:

Delivery of algorithms and interface for regular FACET-II operations (year 1 - 1.5)
Demonstrate extension to other setups (year 1.5 - 2)

→ *Definition of success: tool available and used in routine operation*

Staged approach gradually increases complexity
Success defined by performance (new capabilities) and transition to operations

Potential Future Evolution

Initial studies are to establish the method + infrastructure and demonstrate the approach → philosophy is to start small

Future evolution:

- Scale up to more comprehensive control *(more controlled variables and observed inputs)*
- Adapt to more challenging tasks where comprehensive control is likely to have a large impact on capabilities *(e.g. optimize through plasma stage)*
- Transfer learning to apply to other machines for similar control tasks (e.g. injector optimization, longitudinal phase space manipulation)

Match to Facility + Other FACET-II Experiments

Match to FACET-II R&D Goals and Facility Infrastructure

- R&D with challenging-to-control "extreme" beams *(e.g. high current, high intensity)*
  → RL is a promising approach for fine control of such beams (e.g. strong nonlinear collective effects)
- Variety of experiments requiring custom beams
  → Ideal application for RL (can use learned information from previously-observed setups)
- Suitable diagnostics and data/control infrastructure
  → Essential for making effective use of deep RL (best use is with high-dimensional data)
- Benefit to future FACET-II experiments *→ RL control will likely be an enabling technology for new experimental capabilities (esp. where finer control of extreme beams is required)*

Synergy with Current ML Efforts

- ML-based diagnostics (LPS and edge radiation) can be inputs to RL control
- Joint use of gathered data and machine time
  Data from LPS VD and adaptive feedback can be used in pre-training RL controller, data from RL testing can help train LPS VD
- Adaptive feedback and RL fill complementary niches and can be used in tandem:

Model-independent → no data needed	*Model-dependent → learns to control system over time*
Good for new setups, hardware changes, etc	*Very fast for observed setups, needs more time to learn very different setups*
Excellent for stabilization, drift compensation, fine tuning	*Good for bulk and fine tuning; less robust for stabilization*

Good synergy w/ current ML efforts *(e.g. tandem use of algorithms, can leverage joint data)*
Good match to long-term FACET-II science goals *(e.g. control of extreme beams)*

Desired Facility Upgrades

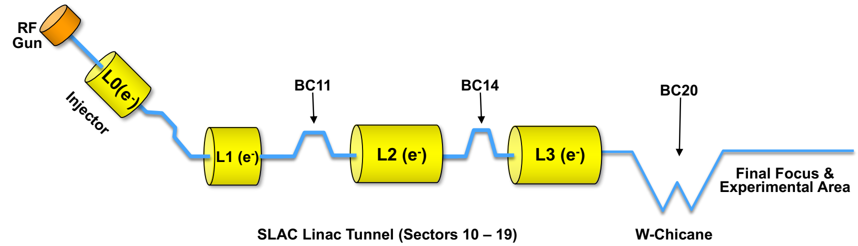
- GPUs connected to control system will be critical for dedicated convolutional neural network deployment and online retraining

  → *for initial testing can use shared desktop GPU (but not suitable for dedicated use)*

- Extended data archiving / acquisition capabilities

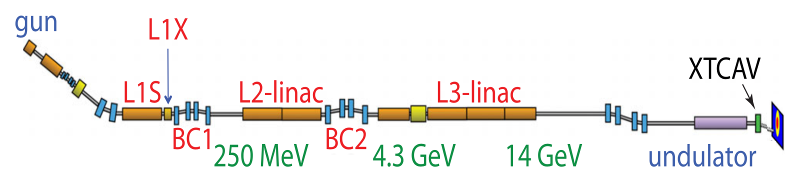
  → *e.g. more flexibility with beam synchronous acquisition for cameras*

Collaboration

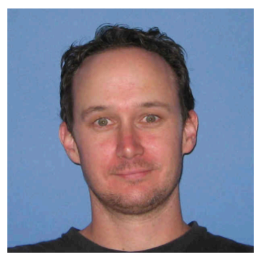


LCLS injector and linac layout is similar to that of FACET-II, and both require delivery of customized LPS

→ *Opportunity to demonstrate transfer of RL algorithms between accelerator systems*



Long-term aim is to apply to experiments (e.g. tuning for CSR studies, plasma experiments, etc.)





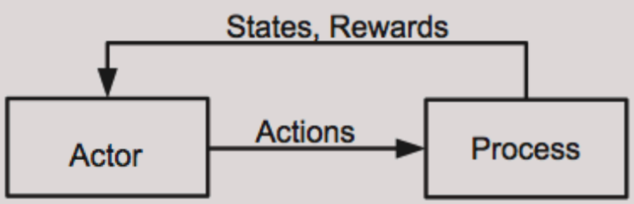
Backups

Timeline

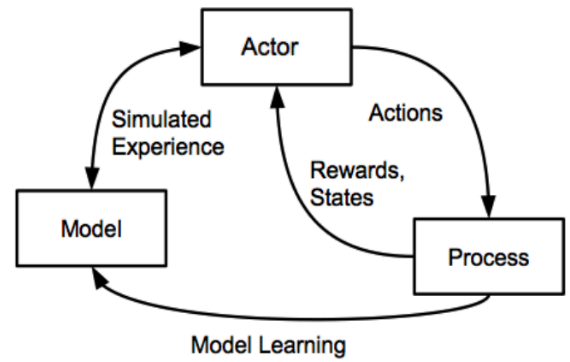
- Simulation campaign: *underway with Impact (injector) and Lucretia (linac); estimated finish by end of November 2020*
- Conceptual design: *done, will adapt to available instrumentation as needed*
- Experimental design (90%): *mid December 2020*
- Ready for installation: *end of December 2020*
- First science: *nominal beam parameters or commissioning beam is ok*
- Stages: first target injector, then linac + combined control

Neural Network Policies and Reinforcement Learning

Actor-only Methods



- Actor is a control policy
- Maps states to actions
- Reward provides training signal

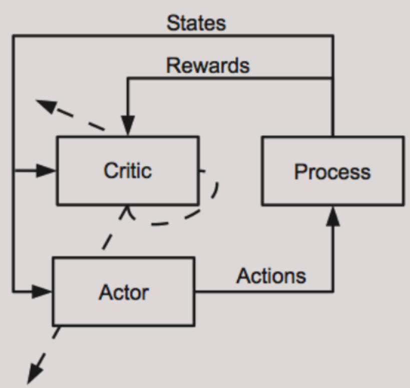


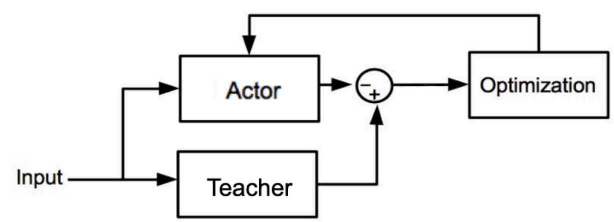
Can train on models first to get a good initial solution before deployment

Actor-Critic Methods

- Critic maps states or state/action pairs to an estimate of long-term reward
- Could be a NN, tabular, etc.
- Critic provides training signal to actor

Without actor: use an optimization algorithm with the critic





Can use supervised learning to first approximate the behavior of a different control policy