Reinforcement Learning for FLASH Dose Delivery Optimization

Jonathan Edelen, Joshua Einstein-Curtis, Christopher Hall, Morgan Henderson (RadiaSoft)

Auralee Edelen, Jorge Diaz Cruz (SLAC)

February 7th 2024

RL4AA 2024: Salzburg Austria

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Accelerator R&D and Production award number DE-SC0023641

RadiaSoft is an industry leader in high-level research & design and scientific consulting for beamline physics and machine learning

- Diverse staff covering a wide range of scientific and software expertise
  - Seven full time software developers with expertise in Java, Python, C++, and Perl.
  - Eleven full time computational scientists and engineers with expertise in a variety of specialty disciplines, including: optics, controls, embedded systems, and computational modeling.
  - Three administrative staff and one marketing professional.
- Regular contributions to more than 20 workshops and conferences annually, including accelerator education

Computational Support	User Support	Operational Systems
Sirepo: browser based computational gateway	Image processing: Sample identification and noise reduction	Embedded systems for edge AI and LLRF
Jupyterhub: customized computational environment	Data analysis and visualization workflows	Digital twins and online modeling
Shielding design and radiation transport simulations	Sirepo https://sirepo.com	

Contract R&D / Engineering Services

Industrial Applications of Accelerators

Security and Defense
- Directed energy
- Single effects testing

Medical
- Proton Therapy
- E-beam therapy
- X-ray therapy

Sterilization
- Medical device sterilization
- Food irradiation
- Wastewater treatment

Imaging
- X-ray sources
- Gamma-ray sources
- electron microscopy

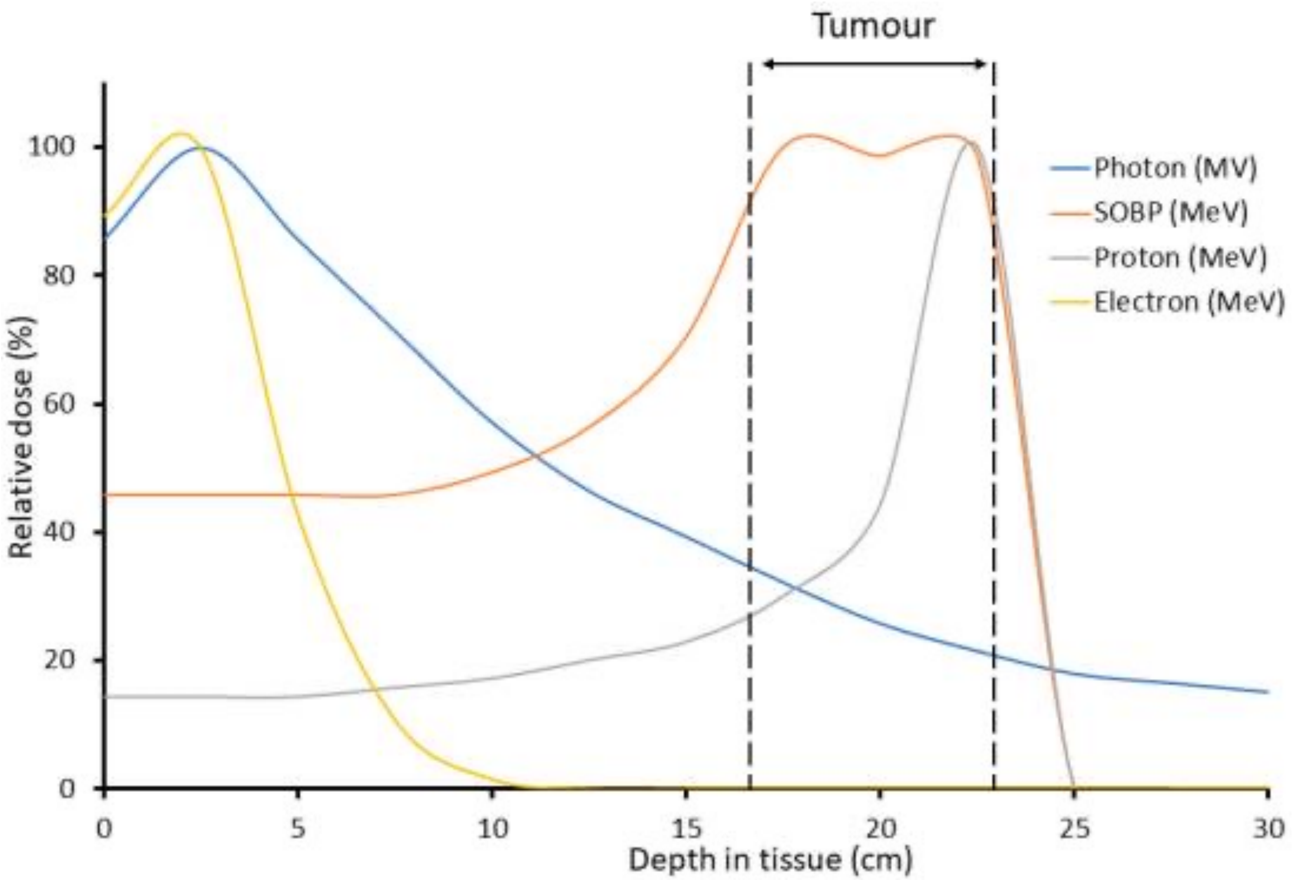
Manufacturing
- Polymer treatment
- Industrial Curing
- Welding
- Ion Implantation

- Next generation of industrial systems are increasing in complexity
  - Distributed RF generation
  - Multi-cavity designs for higher energy

What is FLASH Radiotherapy?

- Ultrafast delivery of radiation
  - dose rates that are several orders of magnitude greater than those used in conventional radiotherapy
  - 40 Gy/s (FLASH) *vs* 0.5–5 Gy/min (conventional)

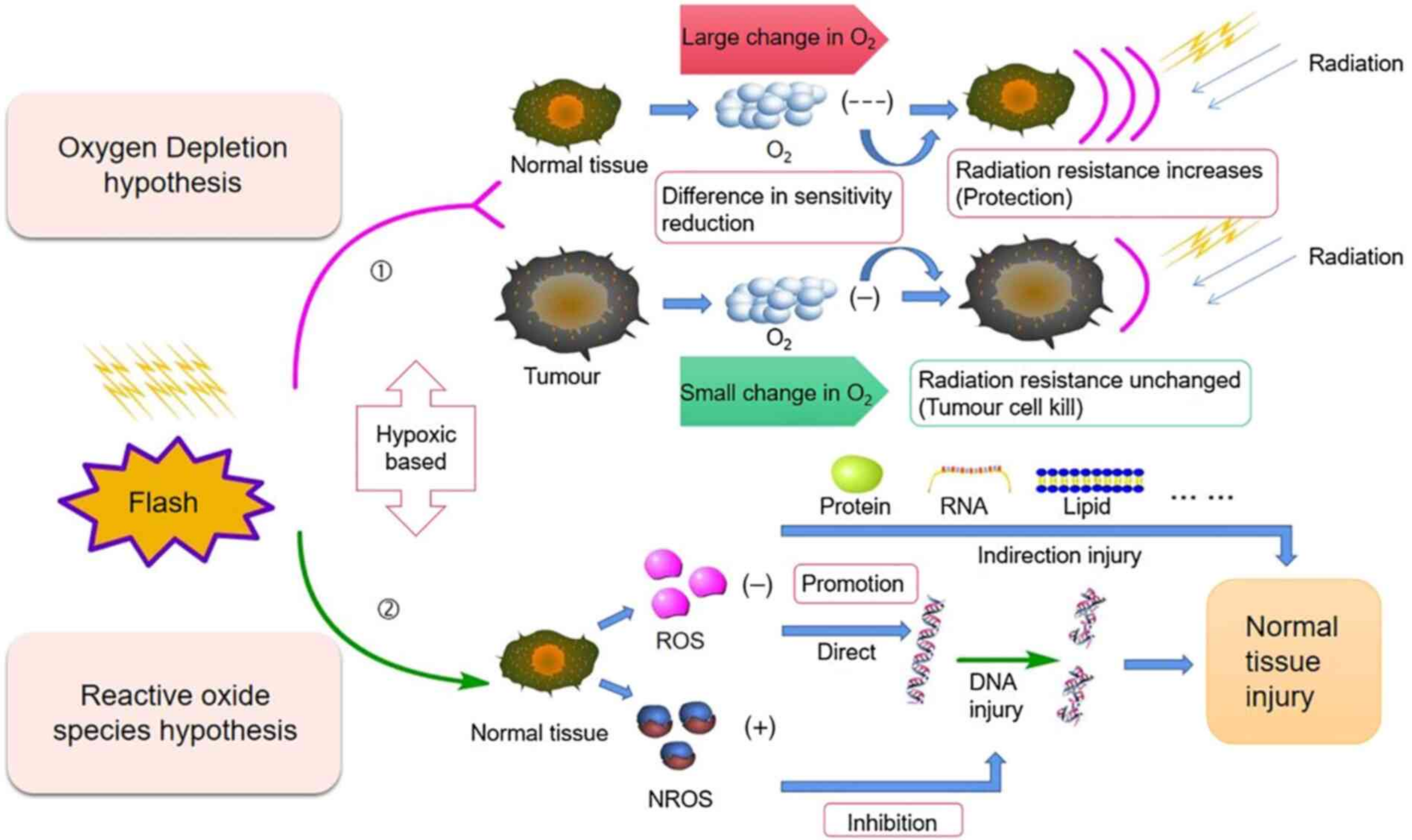
- Preclinical data suggesting that FLASH could achieve better disease control with fewer side effects
  - Improved safety and efficacy (confirmed by clinical trials)



Hughes JR, Parsons JL. FLASH Radiotherapy: Current Knowledge and Future Insights Using Proton-Beam Therapy. Int J Mol Sci. 2020 Sep 5;21(18):6492. doi: 10.3390/ijms21186492. PMID: 32899466; PMCID: PMC7556020.

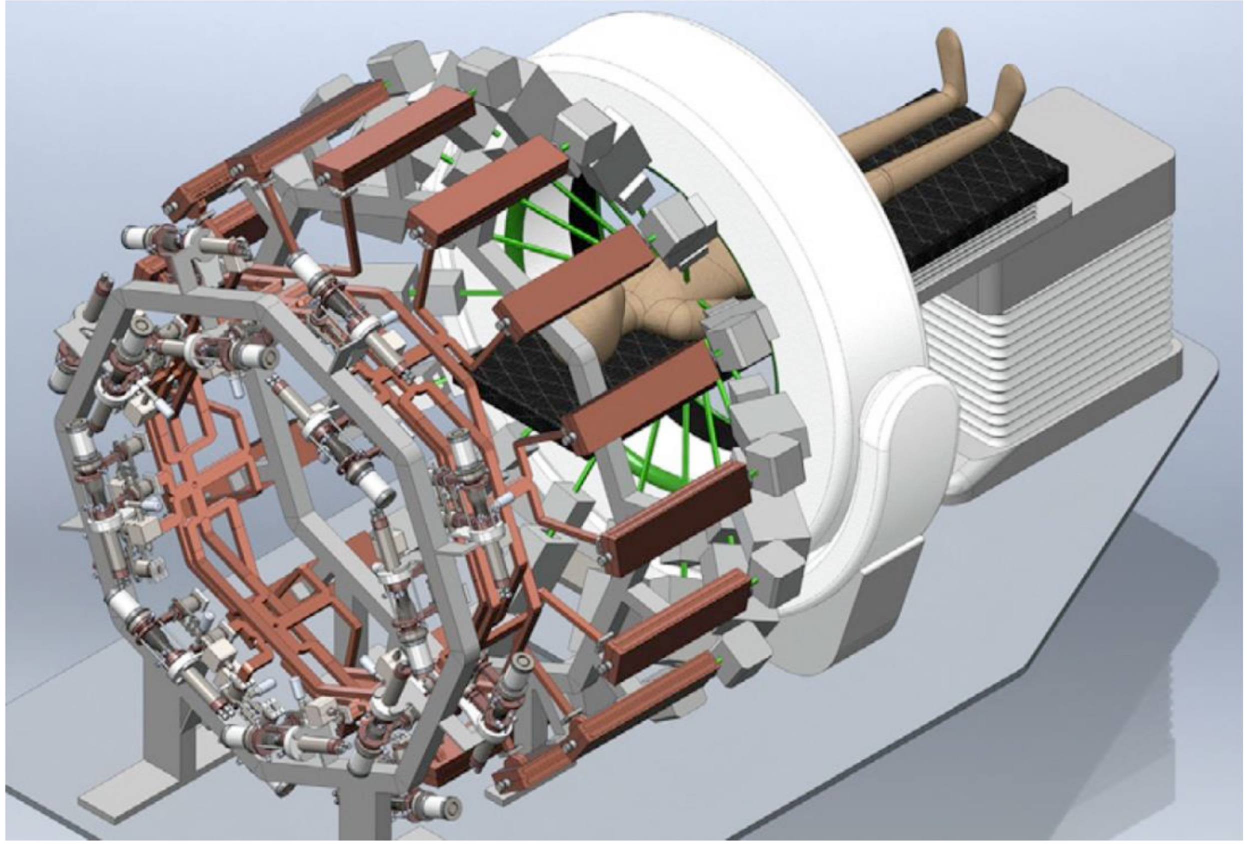
How does FLASH Radiotherapy work?

- Right: Mechanistic diagram of the oxygen consumption hypothesis and ROS hypothesis:
  - Top: Oxygen consumption hypothesis: High-dose transient irradiation reduces the presence of oxygen, and this effect is greater on normal cells, resulting in stronger radiation resistance.
  - Bottom: Reactive Oxygen Species (ROS) levels that causes DNA, RNA, protein and lipid injury, and an increase in the protective non-reactive oxygen species (NROS) levels that inhibits DNA injury.



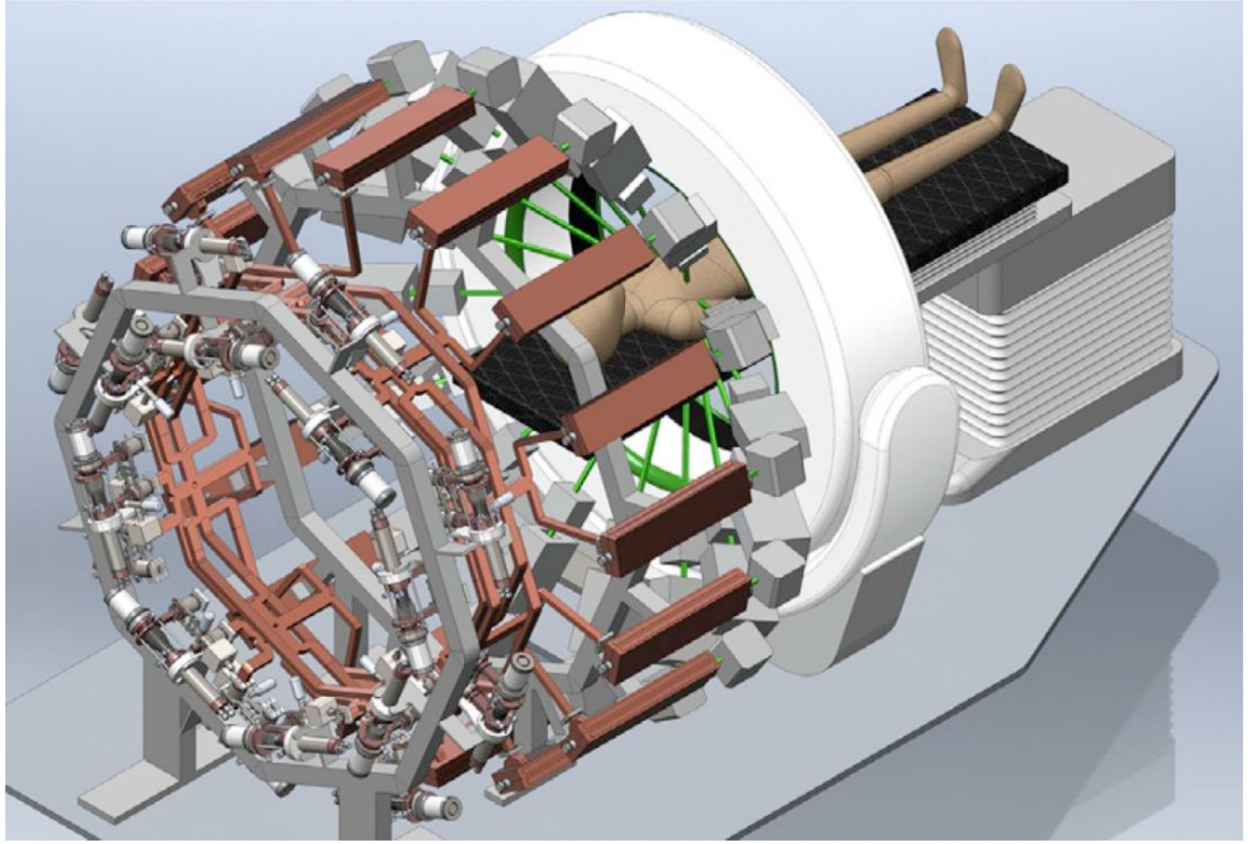
PHASER: A solution for FLASH-RT

- PHASER (pluridirectional high-energy agile scanning electronic radiotherapy)
  - 16 klystrinos power combined to drive a given linac with 5.3 MW of peak power
  - Switching between LINACs occurs at 300ns



PHASER: A solution for FLASH-RT

- PHASER (pluridirectional high-energy agile scanning electronic radiotherapy)
  - 16 klystrinos power combined to drive a given linac with 5.3 MW of peak power
  - Switching between LINACs occurs at 300ns

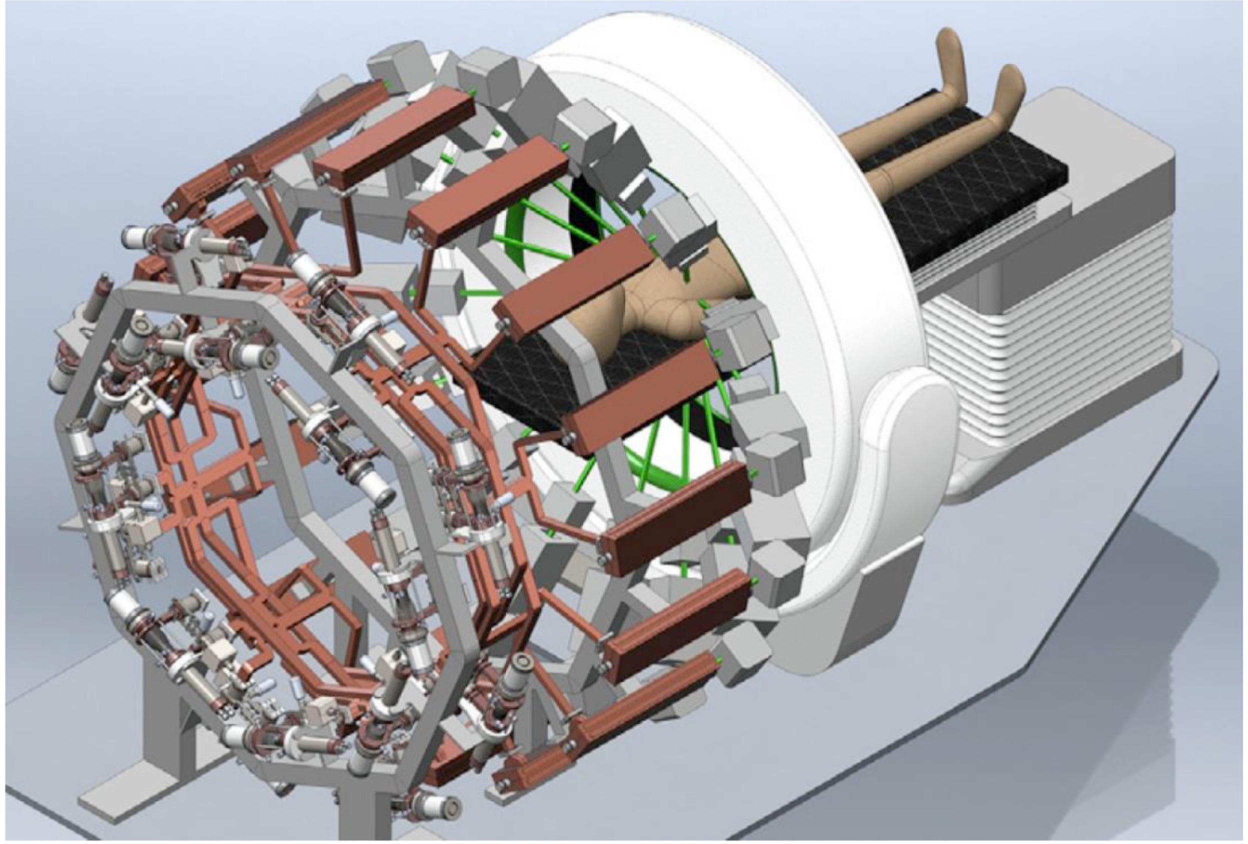


- Understanding the accelerator
  - Modeling the power combining is challenging (compensating for phase and amplitude jitter in the klystrinos)
  - Different LINACs need to operate at different energies
  - Beam steering using magnets

PHASER: A solution for FLASH-RT

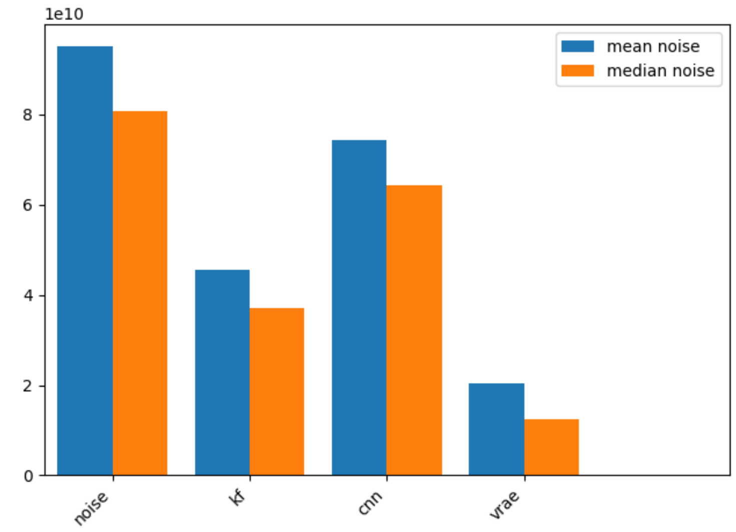
- PHASER (pluridirectional high-energy agile scanning electronic radiotherapy)
  - 16 klystrinos power combined to drive a given linac with 5.3 MW of peak power
  - Switching between LINACs occurs at 300ns

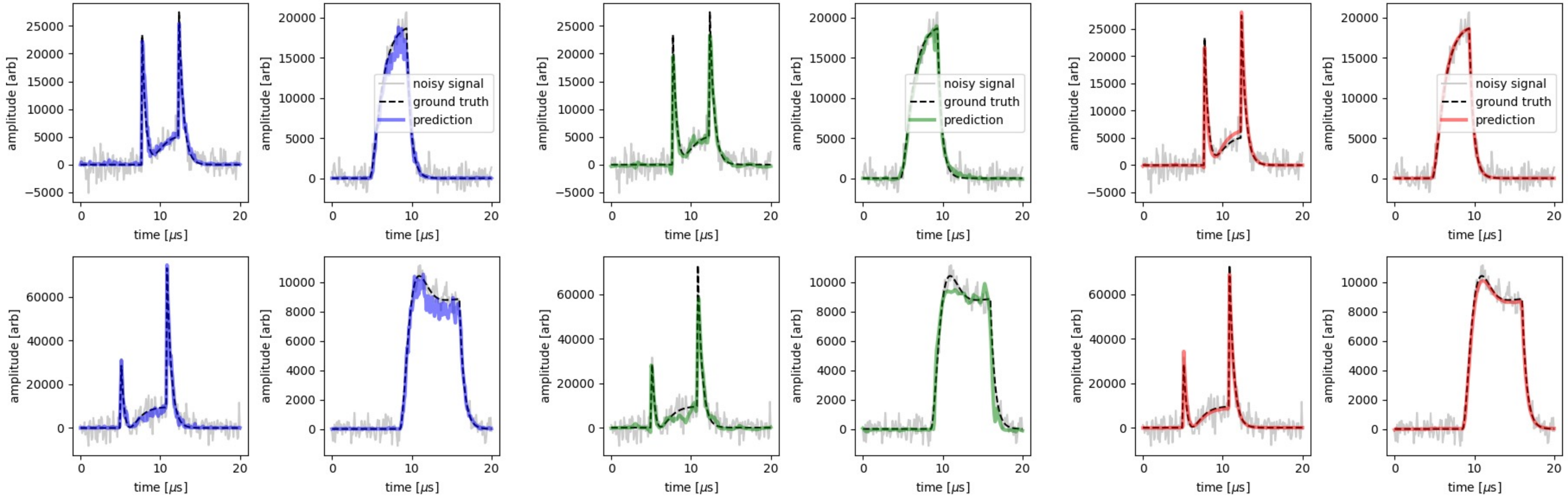
- Understanding the accelerator
  - Modeling the power combining is challenging (compensating for phase and amplitude jitter in the klystrinos)
  - Different LINACs need to operate at different energies
  - Beam steering using magnets
- Patient treatment
  - Rapid computation of optimal dose
  - Compensation for breathing



Reducing Noise in RF Signals

- Comparison of noise reduction methods
  - Kalman filter includes noise term in model
  - 1D CNN autoencoder
- Noise statistics computed from FFT of the model error with the clean data.





Kalman Filter

Convolutional Neural Network

VRAE

A toy model for power combining

- The RF power at a given station is the sum of the inputs from the klystrinos

Gain error (beam current, cathode voltage, output cavity)

Mechanical phase errors (fabrication)

$$V_{rf} = \sum_{i=1}^{16} (A_i \pm \delta A_i) \cos(\phi_i^{set} + \delta\phi_i^e + \delta\phi_i^m)$$

Phase set point (control parameter / action)

Electrical phase error (LLRF)

- Beam energy is actively controlled by adjusting set point phase of the different RF sources

$$E_{beam} \propto V_{rf} \qquad\qquad P_{rf} \propto V_{rf}^2$$

Tuning the Beam Energy

Optimization

Lookup Table

Inverse Models

Too slow to converge

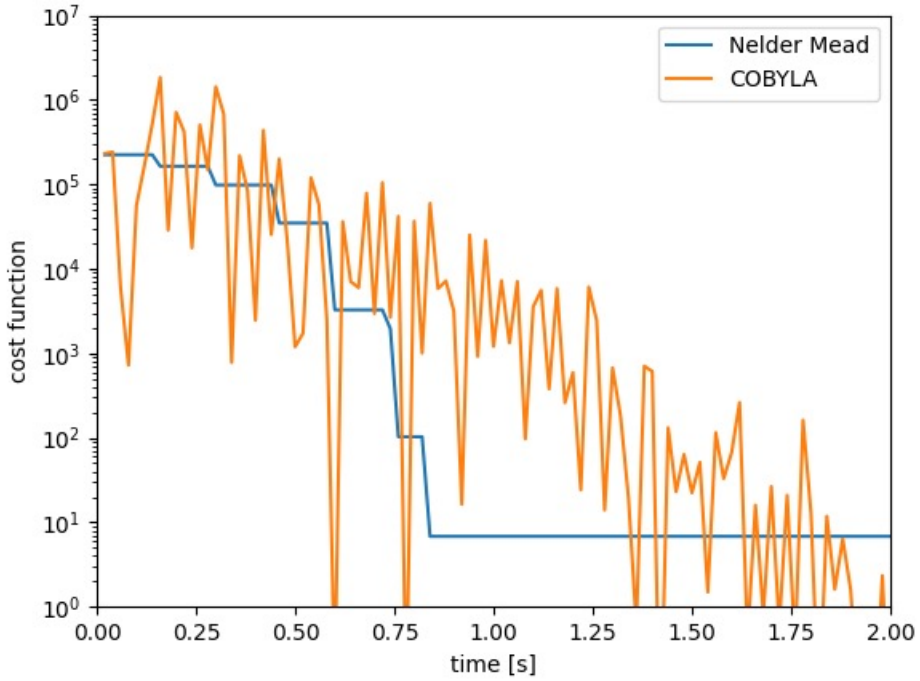
Sensitive to noise and perturbations

Nonstationary System

Problem is not invertible / highly degenerate

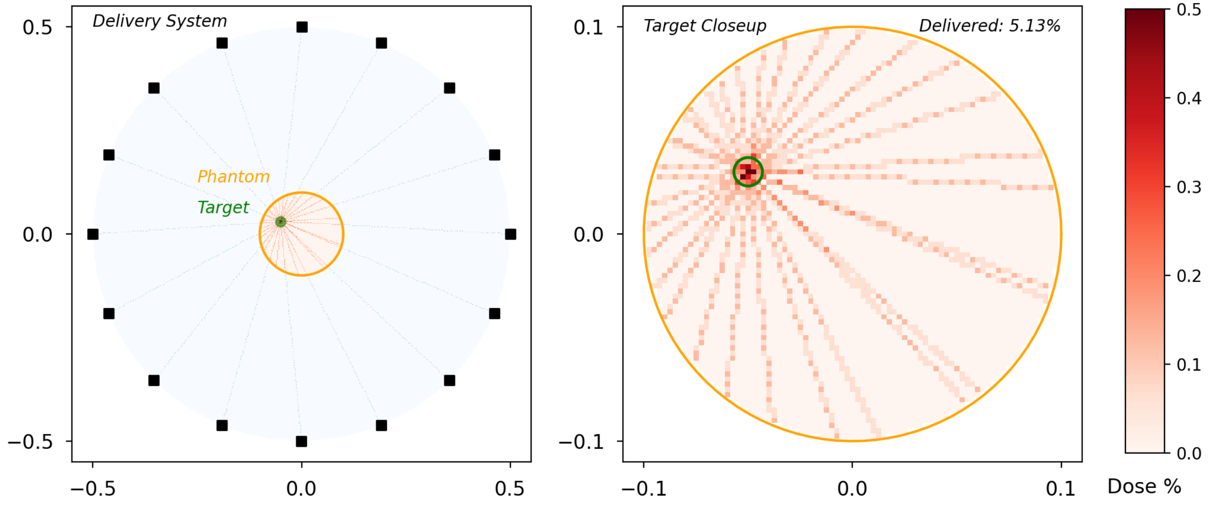
Optimization Studies for RF

- Comparing optimization routines
  - Nelder-Mead and COBYLA
  - Completion time assumes 50Hz operation of the linac
  - Higher repetition rate possible
    - software could limit the optimizer update rate



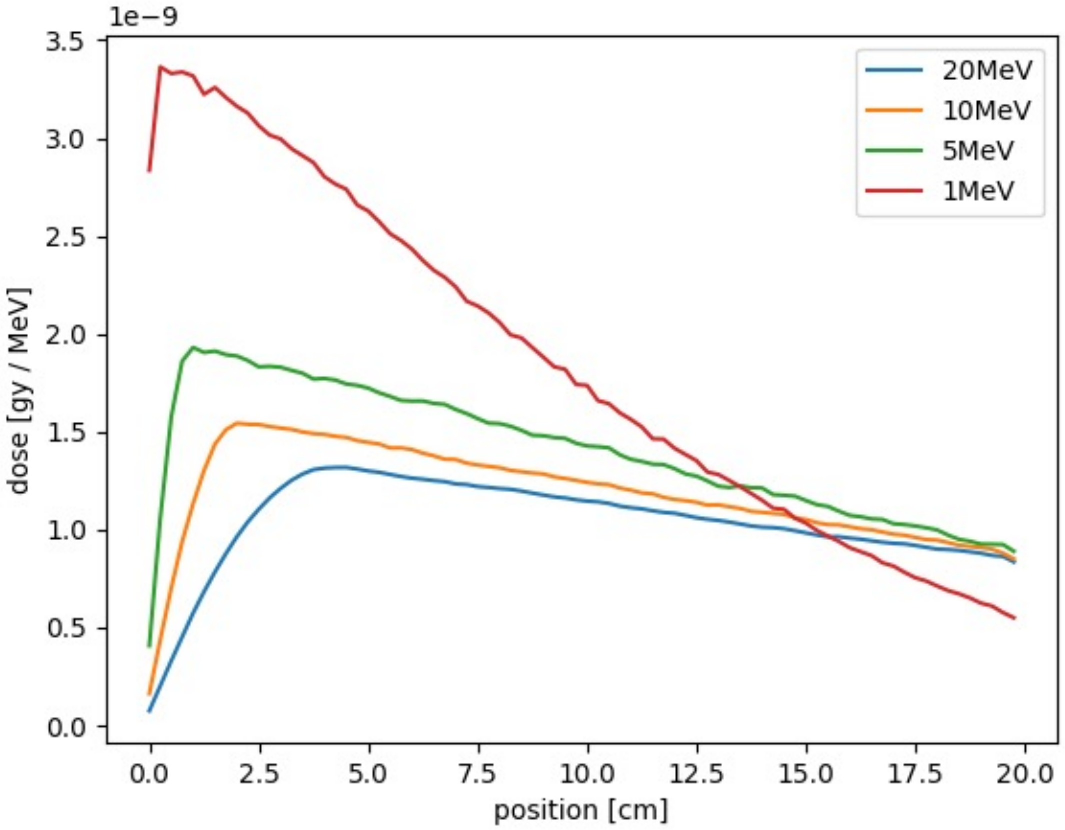
A Toy Model for A PHASER-like system

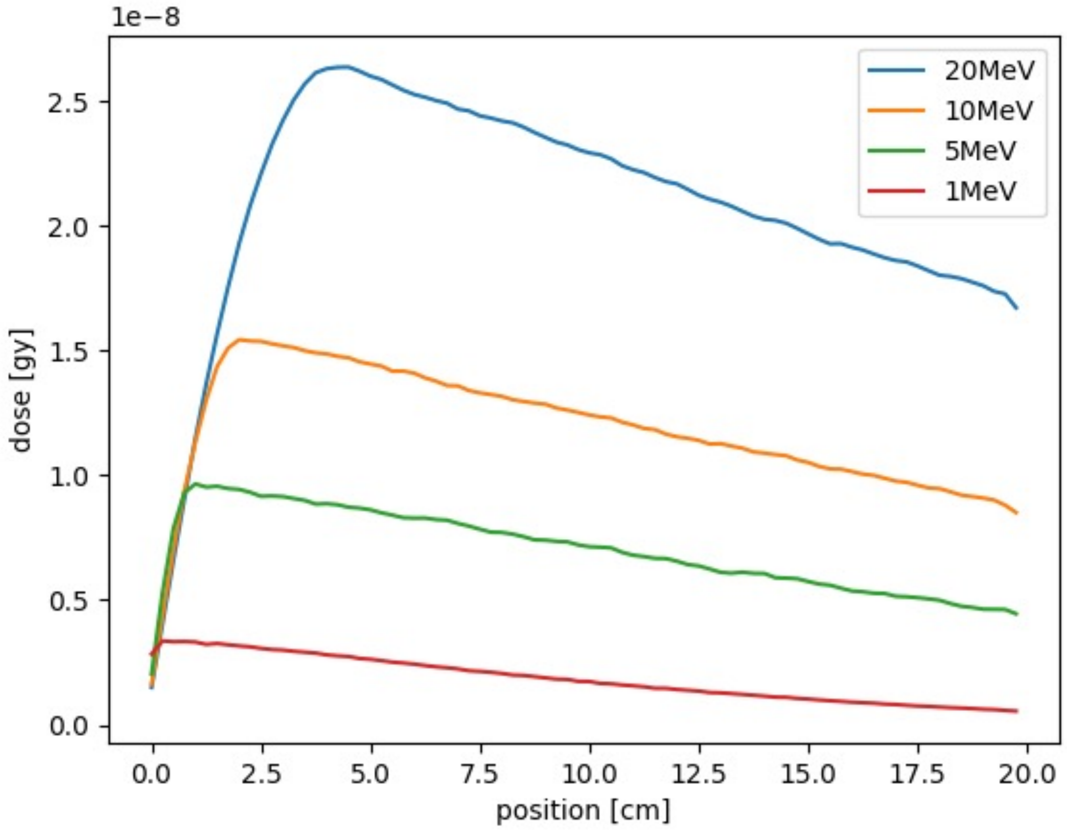
- Sixteen different x-ray sources with the ability to tune energy and steering to optimize the dose delivery profile
    - Toy model assumes a single target plane (2D)
    - Energy deposited in a water phantom
    - Energy range of the x-rays is 1-20 MeV
    - Can adjust commensurate with PHASER parameters
- Water phantom simulated in GEANT-4
    - Modeling 1-D energy loss / deposition



Dose delivery simulations in GEANT-4

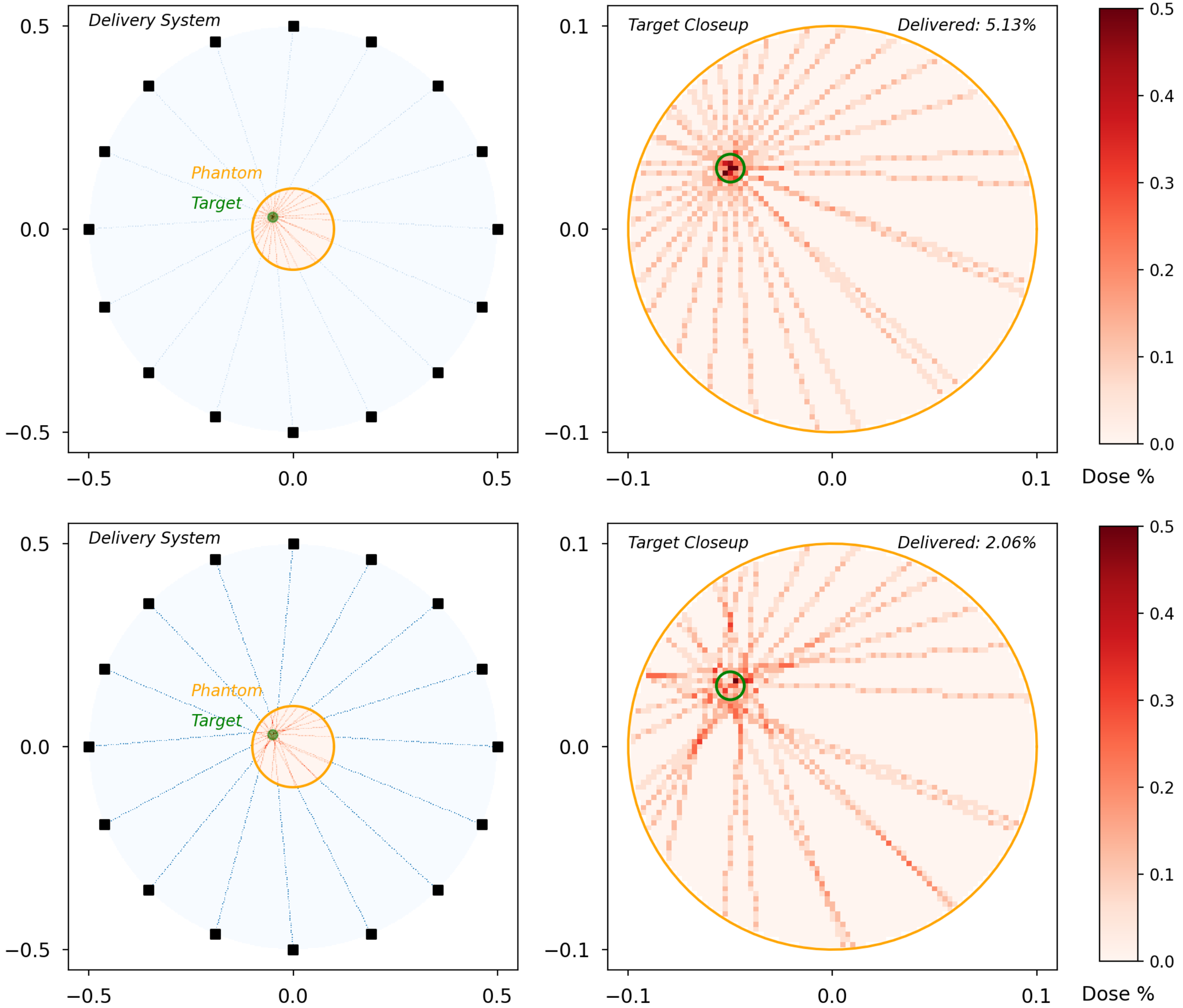
- Compute the energy loss as a function of position inside a water phantom
  - Scan energy then use interpolating function to generate a continuous control knob for the RL model
  - Dose computed for 10^6 x-rays – realistic beams would deliver ~10k times this dose





A Toy Model for A PHASER-like system

- Compute target direction and energy deposition in the phantom
  - Compute beam entry and exit point and associated path length inside the phantom
  - Use path length data to compute the energy deposition curve
  - Compute a histogram of the energy deposition weighted by the dose data from GEANT-4
- Target defined as a circular region with a fixed center point
- Beam steering and energy are the model input parameters
- Right shows the dose distribution for two cases
  - Correct beam steering (top)
  - Random errors in the steering (bottom)



RL Methods

- Problem setup
  - PHASER-like system with unknown targeting offsets
  - Phantom targets with randomized size & location
  - Actions space consists of source targets & energies
- Actor-critic approach with deterministic policy gradient (DPG)
  - Actor & critic both represented as deep neural networks (DNNs)
  - Policy & action-value networks trained concurrently
  - Avoids need for importance sampling
- Rewards & penalties tailored to dose delivery problem
  - Reward for percentage of successfully delivered dose
  - Penalty for extraneous patient radiation above threshold
  - Penalty for targeting offset errors

Critic Loss

$$J_Q = \frac{1}{T}\sum_{t=1}^{T}\Big(R_t + \gamma Q_{target} - Q^w\big(s_t, \mu_\theta(s_t)\big)\Big)^2$$

Actor Loss

$$J_a = \frac{1}{T}\sum_{t=1}^{T} Q^w\big(s_t, \mu_\theta(s_t)\big)$$

Targeting Error Penalty

$$p_{err} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\varepsilon} - \varepsilon)^2$$

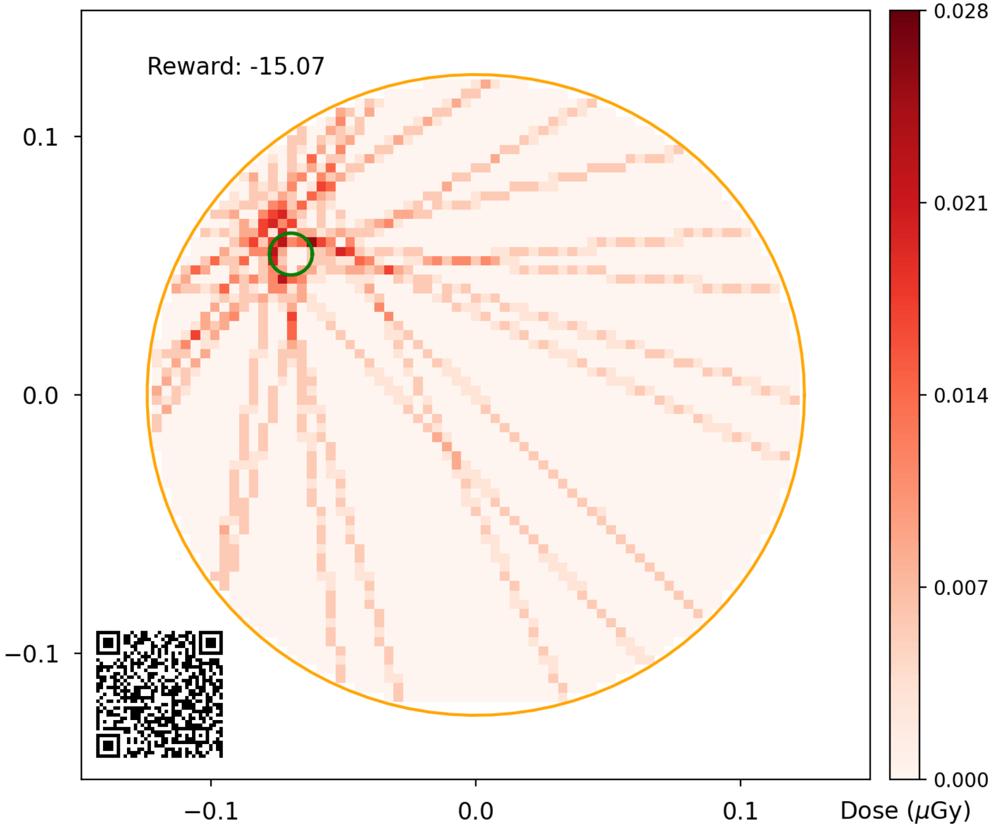
Radiation Penalty

$$p_{rad} = \frac{1}{1 + e^{-\rho\left(\frac{\gamma}{\gamma_c} - 0.5\right)}}$$

RL Results & Ongoing Efforts

- Achieved slightly better-than-nominal targeting
  - Nominal targets assume no targeting offsets
  - Still highly sensitive to variations between episodes
    - Known drawback to NN actor-critic approach
- Optimizing training scheme
  - Learning/discount rates
  - Relative weighting of terms within rewards & losses
- Exploring additional network architectures
  - Began with basic 3-layer NNs, ReLU activation, etc.
- Seeking to reach optimal or near-optimal controls
  - Perfectly correcting for targeting offsets
  - Energy controls accounting for target depth, size, etc.
- Eventually, add complexity
  - e.g. irregular target shapes, real-time target deformation, etc.

PHASER toy model transitioning from nominal control settings ($R = -15.07$) to RL agent settings ($R = -14.93$)



Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.