MAXIMIZING A HIGH DIMENSIONAL POSTERIOR USING A GENETIC ALGORITHM

IMAGE SEGMENTATION POST-PROCESSING

KATY MCKEOUGH


7873
z = 3.689
Counts
8
6
4
2
0
10307
z = 2.686
Counts
20
15
10
5
0
10309
z = 2.186
Counts
80
60
40
20
0

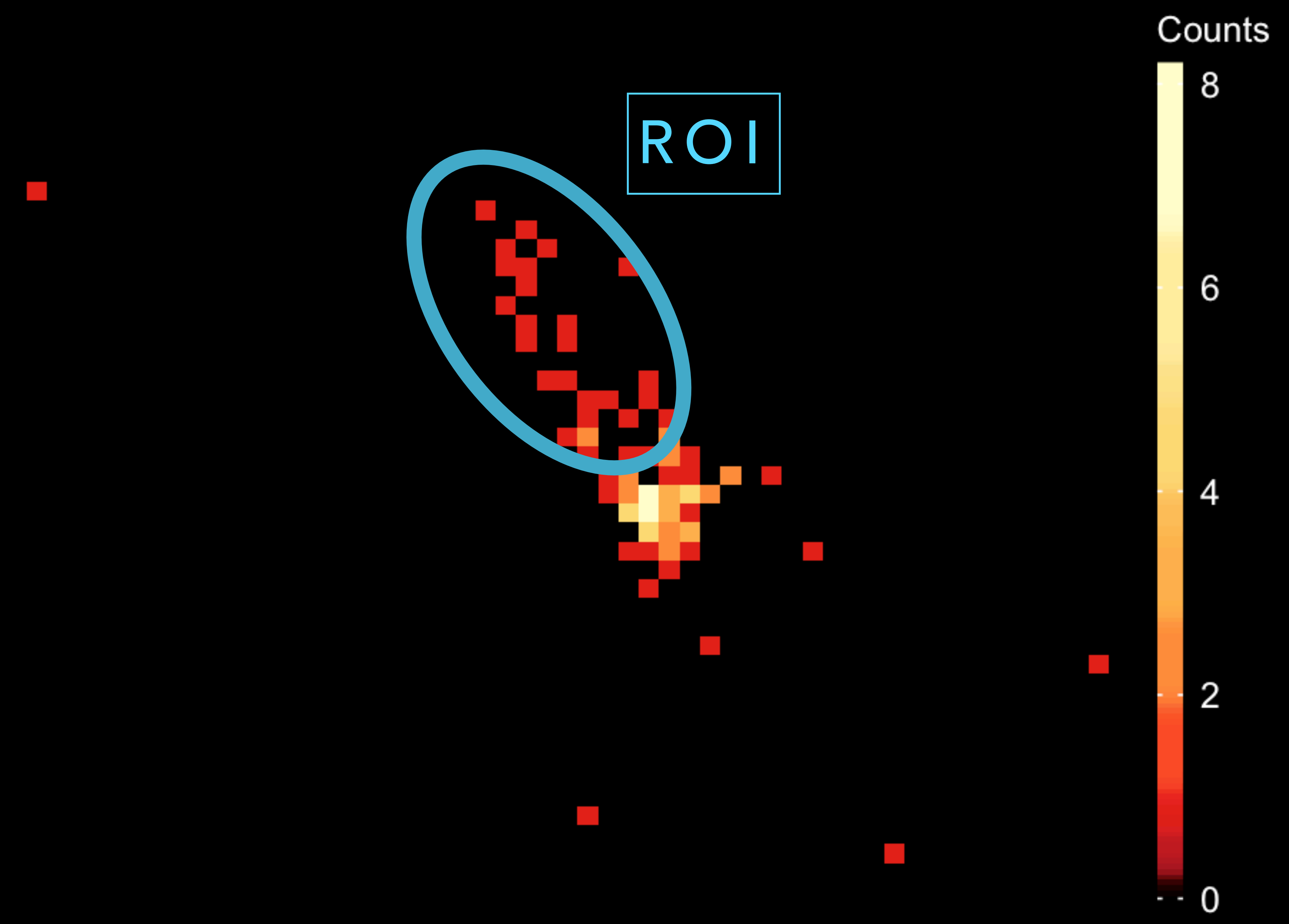

7873
Counts
8
6
4
2
0
10307
Counts
20
15
10
5
0
10309
Counts
80
60
40
20
0

MOTIVATION

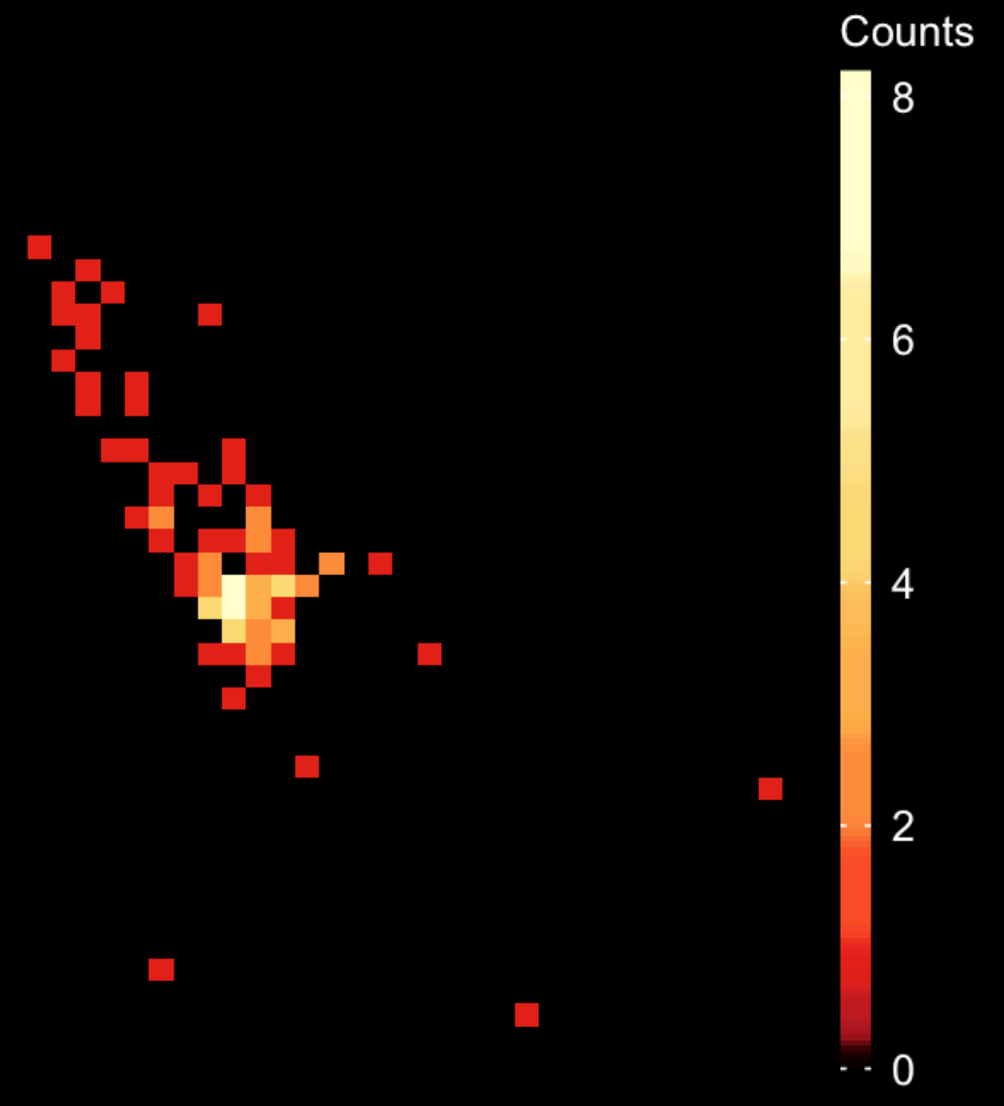
- **This is very difficult**
- Interested in morphology of complex astronomical objects
  - Irregular shapes
  - Low photon counts
  - Small structures, low resolution
  - Diffuse sources (no edges)
- Can not always rely on other wavelengths to help out

BACKGROUND

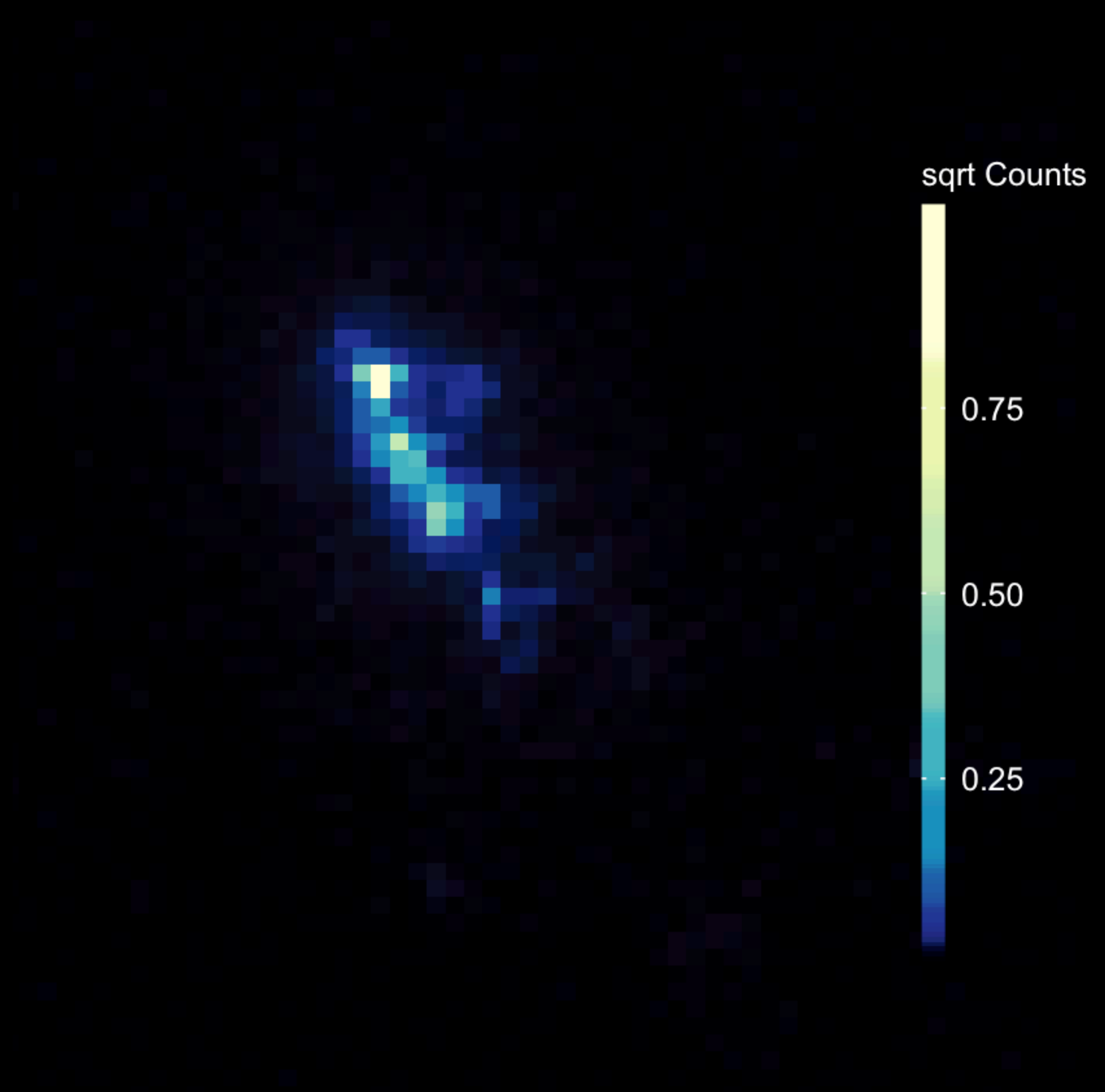
- **Region of Interest (ROI)** - region containing source, separate from the background (e.g. the jet or a partition of the jet)
- Previous work tests whether or not a jet exists in a predefined ROI (McKeough et al. 2016, Stein et al. 2015)
- Multi-phase image segmentation finds minimal boundary around ROI (McKeough et al. TBD)



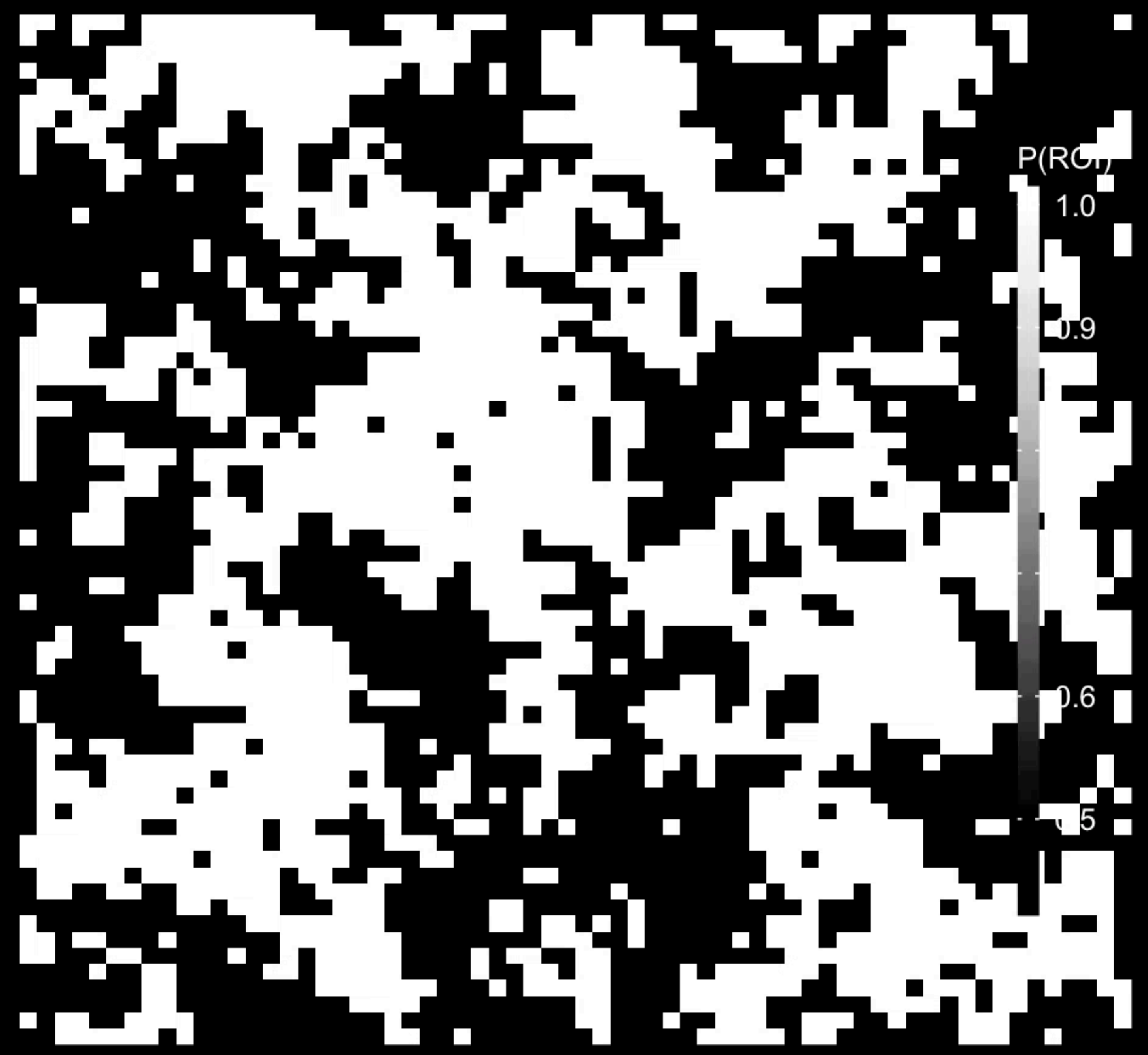
MULTI-PHASE IMAGE SEGMENTATION



X-Ray Counts



Expected
Multi-scale Counts
(LIRA)

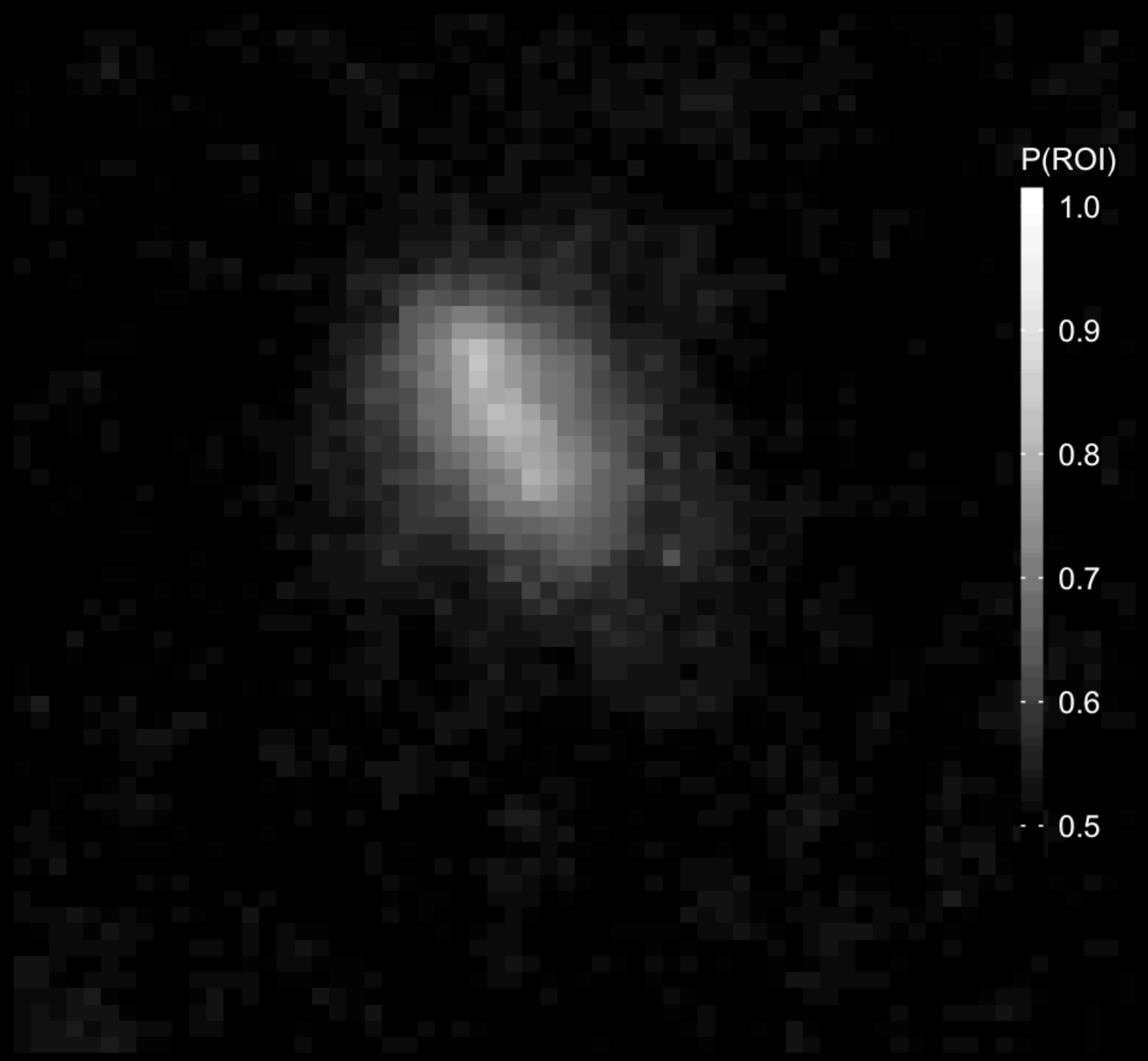


Pixel Assignments

MULTI-PHASE IMAGE SEGMENTATION

- Reconstruct image using LIRA (Λ ; Low-count Image Reconstruction and Analysis)
  - Esch et al. (2004) , Connors & van Dyk (2007)
- Assign each pixel i in the image to either the ROI $z_i = +1$ or the background $z_i = -1$
- Build posterior describing pixel assignments Z

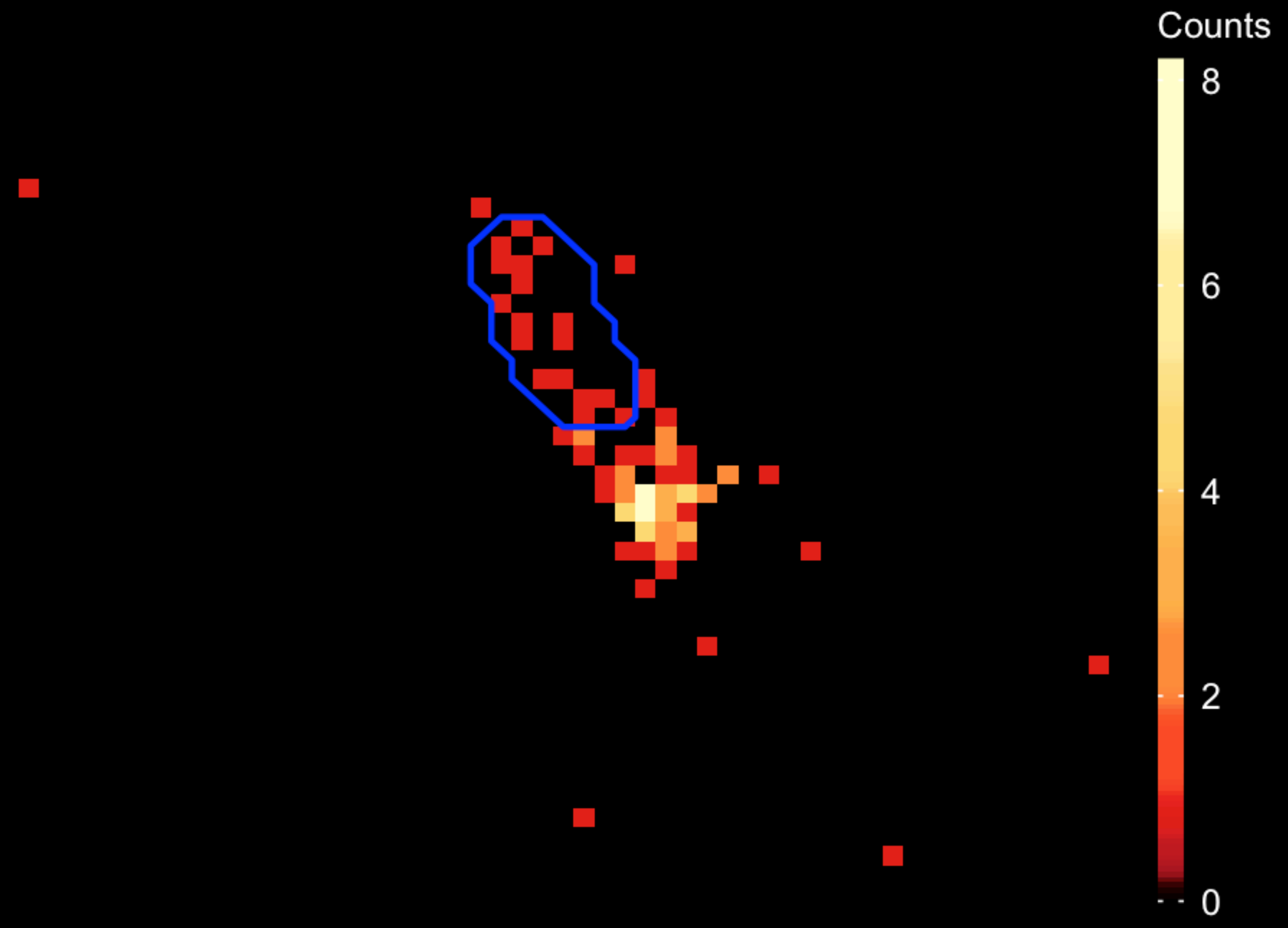
$$p(Z | \Lambda, \theta)$$



THE MINIMAL BOUNDARY

- The **minimal boundary** is defined as the point in which the source can no longer be distinguished from the boundary

- We estimate the minimal boundary by maximizing the posterior distribution on pixel assignments



THE PROBLEM

- Posterior space is discrete, but very large ($2^{64 \times 64}$)
  - Probabilities evaluated at a single observations are too small
  - Not feasible to methodically evaluate posterior at every possible Z

ONE SOLUTION

Compare ratio

$$R = \frac{p(Z_i \mid \Lambda, \theta)}{p(Z_j \mid \Lambda, \theta)} > 1 \quad \Rightarrow \quad p_{\text{max}} = p(Z_i \mid \Lambda, \theta)$$

- Pairwise comparisons easier to calculate
- Able to find global maximum in set of Z through series of pairwise comparisons

AD HOC SET SELECTION

- Creates a smaller set of Z to explore ($64 \times 64 = 4096$)

- If ζ_i is a one-to-one mapping of the z_i where if $z_i = -1 \rightarrow \zeta_i = 0$ and $z_i = +1 \rightarrow \zeta_i = 1$ then the neighborhood statistic is evaluated at each pixel to be

$$\phi_i = \frac{\sum_{j \in d(i,j)=1} \zeta_i \zeta_j}{\sum_{j \in d(i,j)=1} 1}$$

- $\bar{\phi}_i$ is the average neighborhood value across draws from the posterior,

- A collection of images is created by sequentially assigning pixels with the highest $\bar{\phi}_i$ to +1 and the remainder to -1

A BETTER SOLUTION: GENETIC ALGORITHMS

GENETIC ALGORITHMS

- Stochastic search method inspired by the laws of genetics and natural selection — fittest individuals are selected for reproduction in order to produce offspring
- Efficiently optimize over a large space while avoiding getting caught in local extrema
- Use cases in :
  - Medical imaging (Pereira et al. 2014)
  - Astronomy (Rajpaul 2012)
  - Image segmentation (Yu 1998, Sheta et al. 2012)

GENETIC ALGORITHMS

PROS

- Relatively simple than other standard optimization techniques
- Robust to problems with high noise and/or high dimensionality
- High speeds, easily to compute in parallel

NO FREE LUNCH

- “Jack of all problems, but master of none”
- Limited theoretical understanding

OUTLINE:

1. Start with *N* individuals in an **initial population**
2. Repeat until convergence:
   I. **Selection** — select the fittest individuals to become parents for the next generation
   II. **Crossover** — new individuals are created as a combination of two of the selected parents
   III. **Mutation** — each "gene" in an offspring has a probability of mutating
3. Final boundary

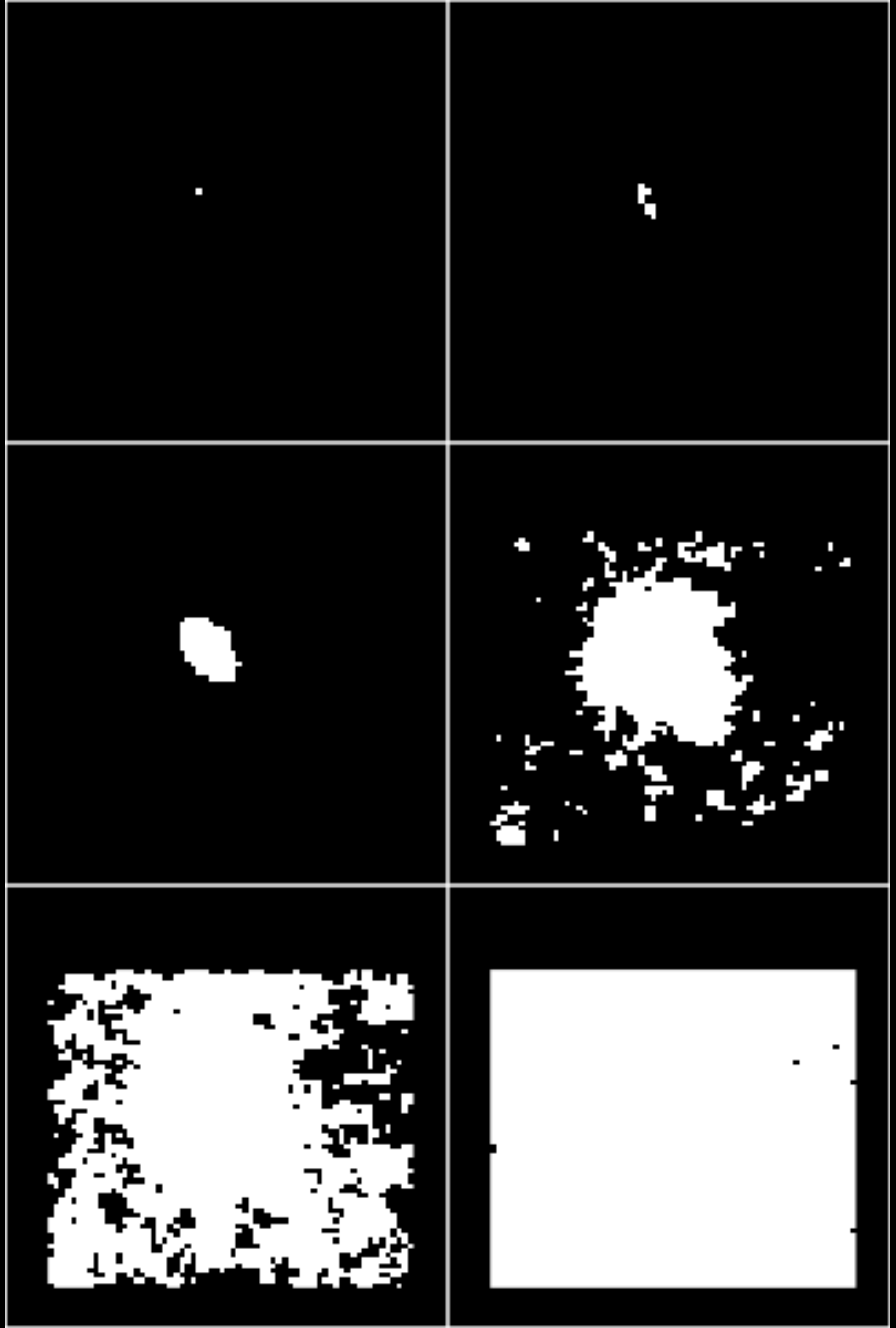
OUTLINE:

1. Start with N individuals in an **initial population**
2. Repeat until convergence:
   I. **Selection** — select the fittest individuals to become parents for the next generation
   II. **Crossover** — new individuals are created as a combination of two of the selected parents
   III. **Mutation** — each "gene" in an offspring has a probability of mutating
3. Final boundary

INITIAL POPULATION

- Initial population can be entirely random, or generated from a “best guess”

- We will use the ad hoc selection method to generate the initial population of ($64 \times 64 = 4096$) pixel assignments (Z)



OUTLINE:

1. Start with *N* individuals in an **initial population**
2. Repeat until convergence:
   I. **Selection** — select the fittest individuals to become parents for the next generation
   II. **Crossover** — new individuals are created as a combination of two of the selected parents
   III. **Mutation** — each "gene" in an offspring has a probability of mutating
3. Final boundary

SELECTION

- Create a **fitness function** to evaluate how "fit" each individual in the population is relative to one another.
  - Optimize over the posterior: $p(Z \mid \Lambda, \theta)$
- Select N_{select} individuals to become parents and reproduce, based on the fitness function

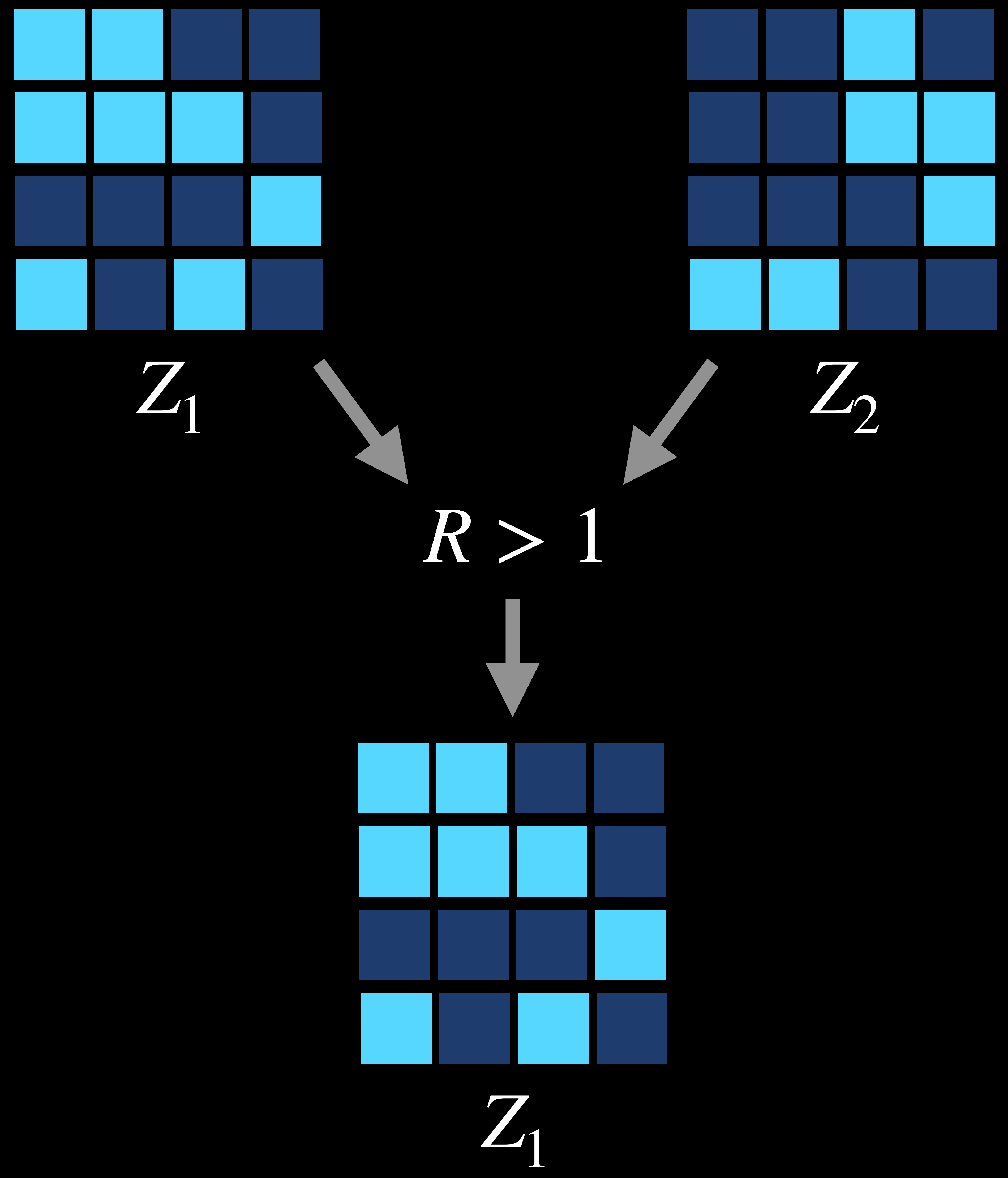
SELECTION

There are numerous types of selection, three main procedures are

- *Rank selection* — rank all Z by the fitness function. Select the N_{select} -th fittest individual observations
- *Roulette selection* — assign a probability to each Z based on a fitness function and randomly draw N_{select} individuals based on distribution (fittest are most likely to be selected)
- *Tournament selection* — create a bracket tournament where two Z face off in each round, the fittest wins and moves on; repeat until N_{select} are selected

TOURNAMENT SELECTION

- Repeat N_{select} times:
  - Select 2 pixel assignments (Z_1, Z_2) completely at random to be in a tournament
  - Evaluate $R = \dfrac{P(Z_1 \mid \lambda, \theta)}{P(Z_2 \mid \lambda, \theta)}$
  - If $R > 1$ then Z_1 is selected, if $R < 1$ then Z_2 is selected



TOURNAMENT SELECTION

Reasons for using tournament selection:

- Allows for pairwise comparisons
- Easy to implement in parallel (Muhlenbein 1989)
- Relatively small time complexity $O(n)$ compared to standard roulette $O(n^2)$ and ranking selection methods $O(n \ln n)$
- Smaller tournament brackets encourage diversity (Goldberg & Deb 1991)

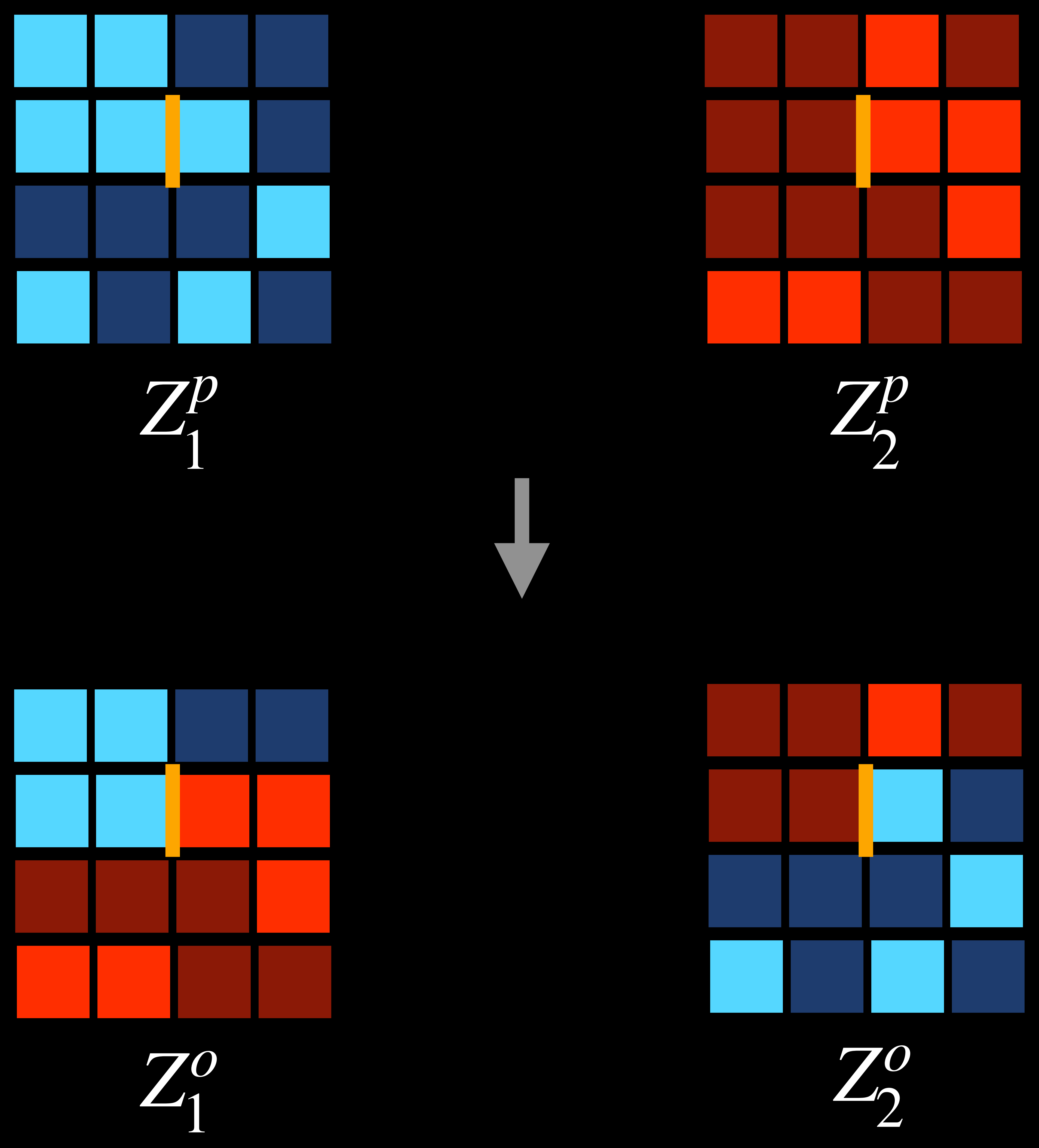
OUTLINE:

1. Start with N individuals in an **initial population**
2. Repeat until convergence:
   I. **Selection** — select the fittest individuals to become parents for the next generation
   II. **Crossover** — new individuals are created as a combination of two of the selected parents
   III. **Mutation** — each "gene" in an offspring has a probability of mutating
3. Final boundary

CROSSOVER (REPRODUCTION)

Once selected, the "parents" pair up to produce "offspring" based on their pixel assignments

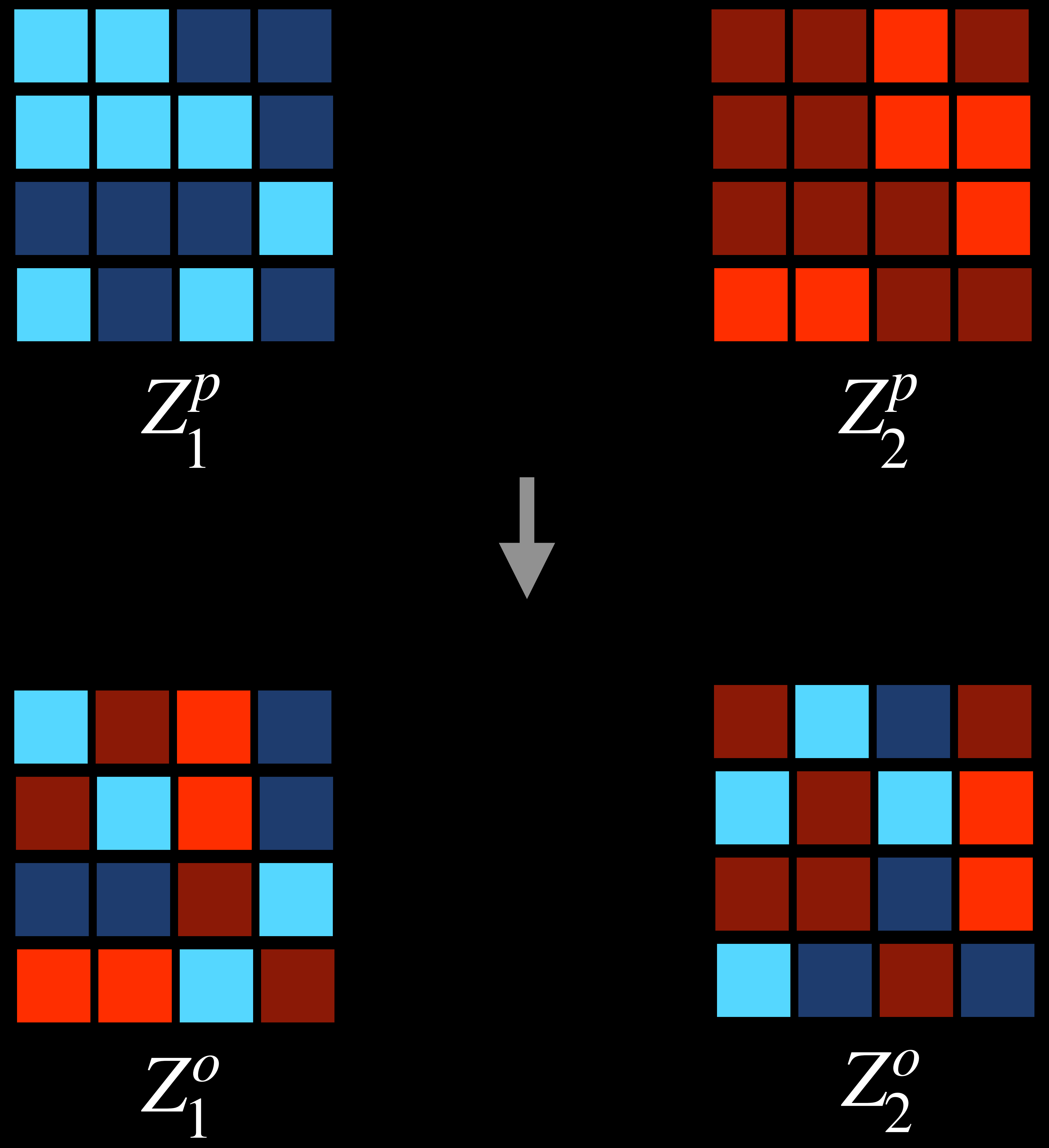
- *One point crossover* — select a random pixel, the offspring get all assignments from on parent before that pixel and all assignments after from that pixel onward
- *Uniform crossover* — each pixel has an equal chance from being the same as one parent or the other



CROSSOVER (REPRODUCTION)

Once selected, the "parents" pair up to produce "offspring" based on their pixel assignments

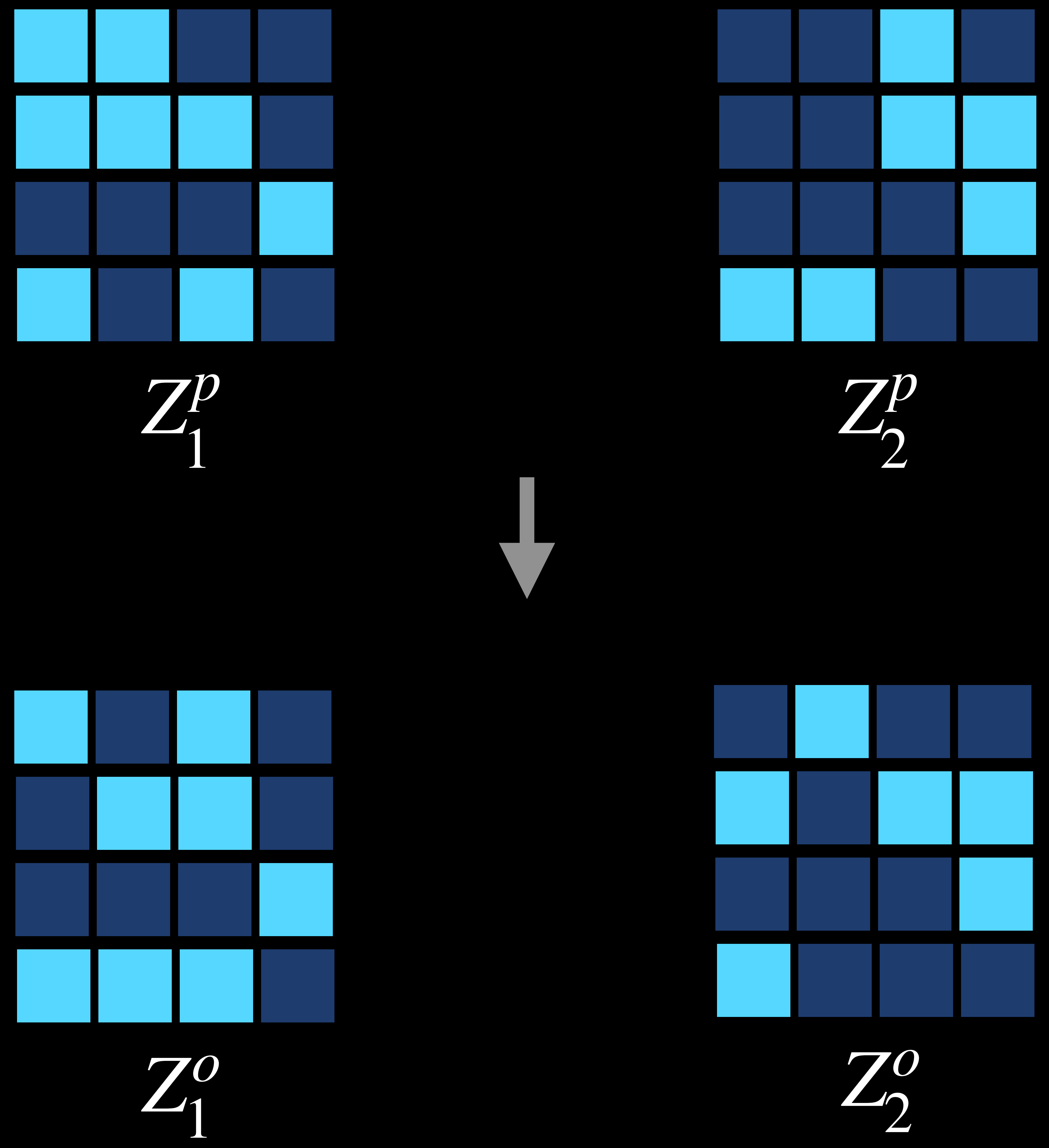
- *One point crossover* — select a random pixel, the offspring get all assignments from on parent before that pixel and all assignments after from that pixel onward
- *Uniform crossover* — each pixel has an equal chance from being the same as one parent or the other



REPRODUCTION

Once selected, the "parents" pair up to produce "offspring" based on their pixel assignments

- *One point crossover* — select a random pixel, the offspring get all assignments from on parent before that pixel and all assignments after from that pixel onward
- *Uniform crossover* — each pixel has an equal chance from being the same as one parent or the other

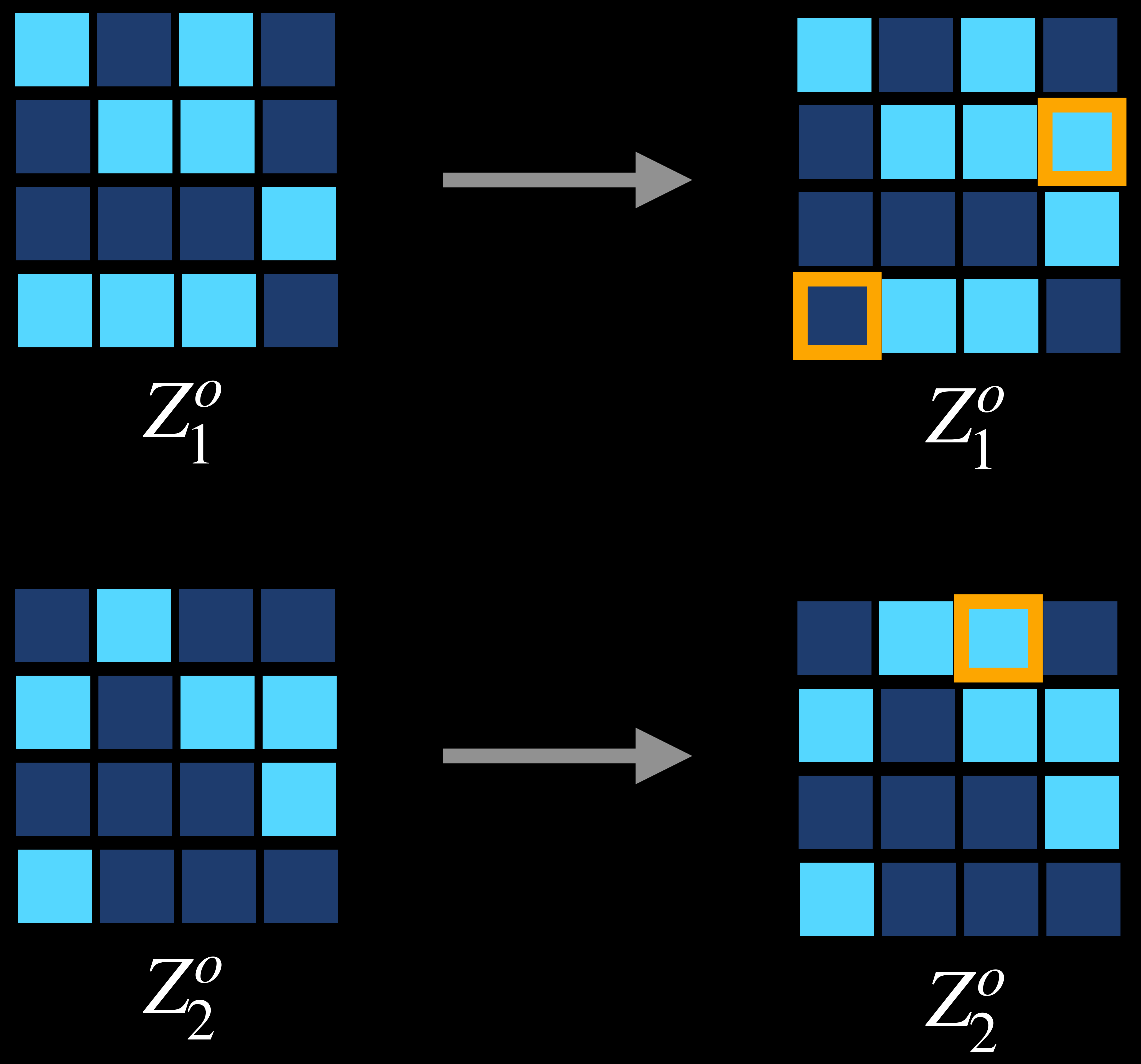


OUTLINE:

1. Start with *N* individuals in an **initial population**
2. Repeat until convergence:
    I. **Selection** — select the fittest individuals to become parents for the next generation
    II. **Crossover** — new individuals are created as a combination of two of the selected parents
    III. **Mutation** — each pixel in an offspring has a probability of mutating
3. Final boundary

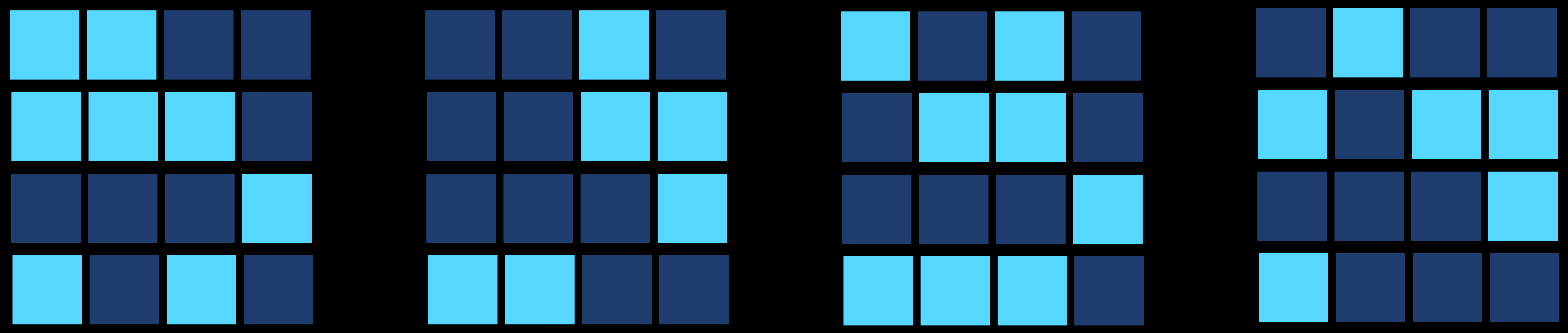
MUTATION

- *Bit Flip Mutation* — select one or more random bits and flip them
  - A bit will flip with probability $1/\ell$ where ℓ is the length of the gene sequence



NEXT GENERATION

- Pool parents and offspring and begin selection process again:



Z_1 Z_2 Z_3 Z_4

OUTLINE:

1. Start with N individuals in an **initial population**

2. Repeat until convergence:

   I. **Selection** — select the fittest individuals to become parents for the next generation

   II. **Crossover** — new individuals are created as a combination of two of the selected parents

   III. **Mutation** — each "gene" in an offspring has a probability of mutating

3. Final boundary

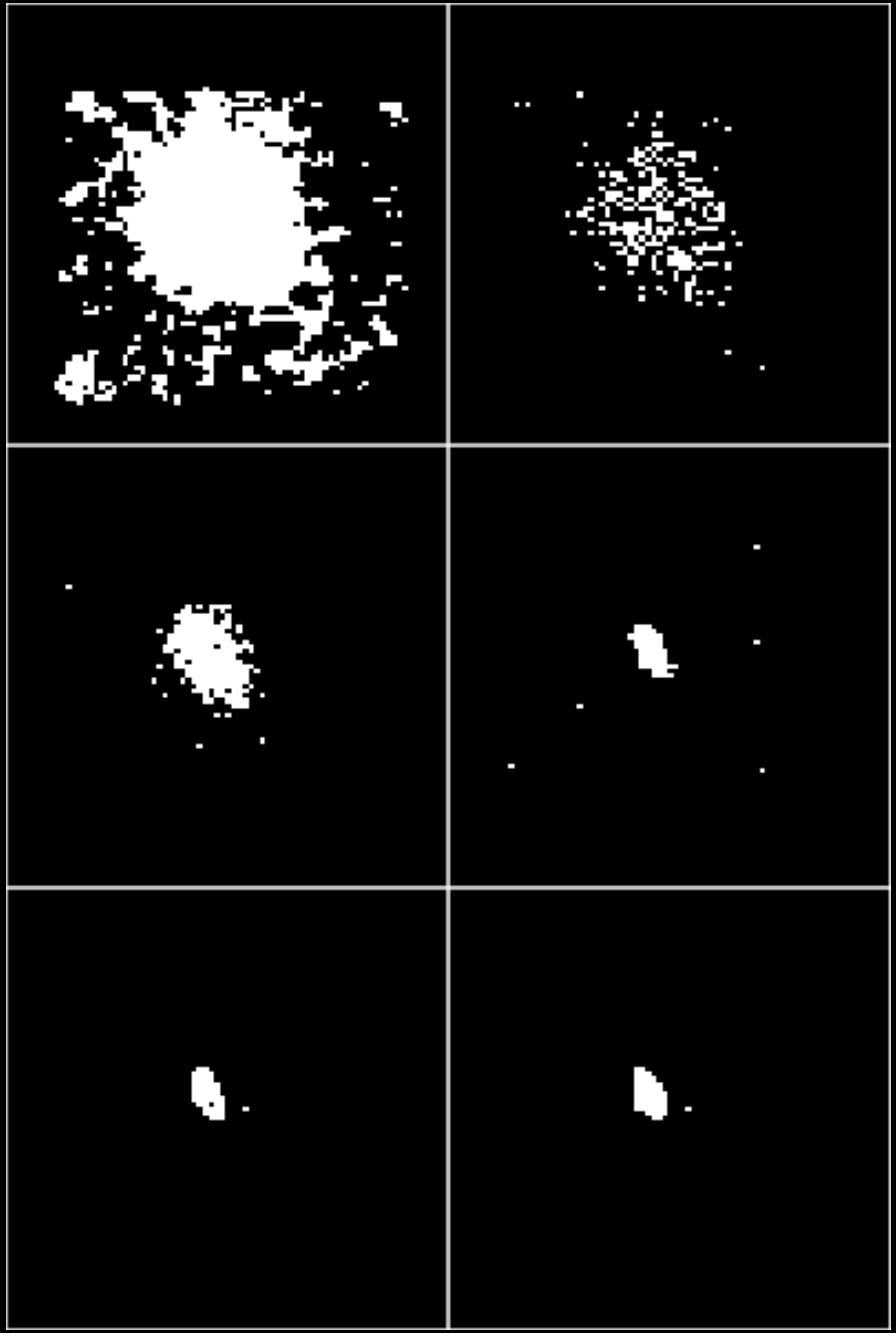
CONVERGENCE

- Stop when all but 10% of the pixels are identical classified

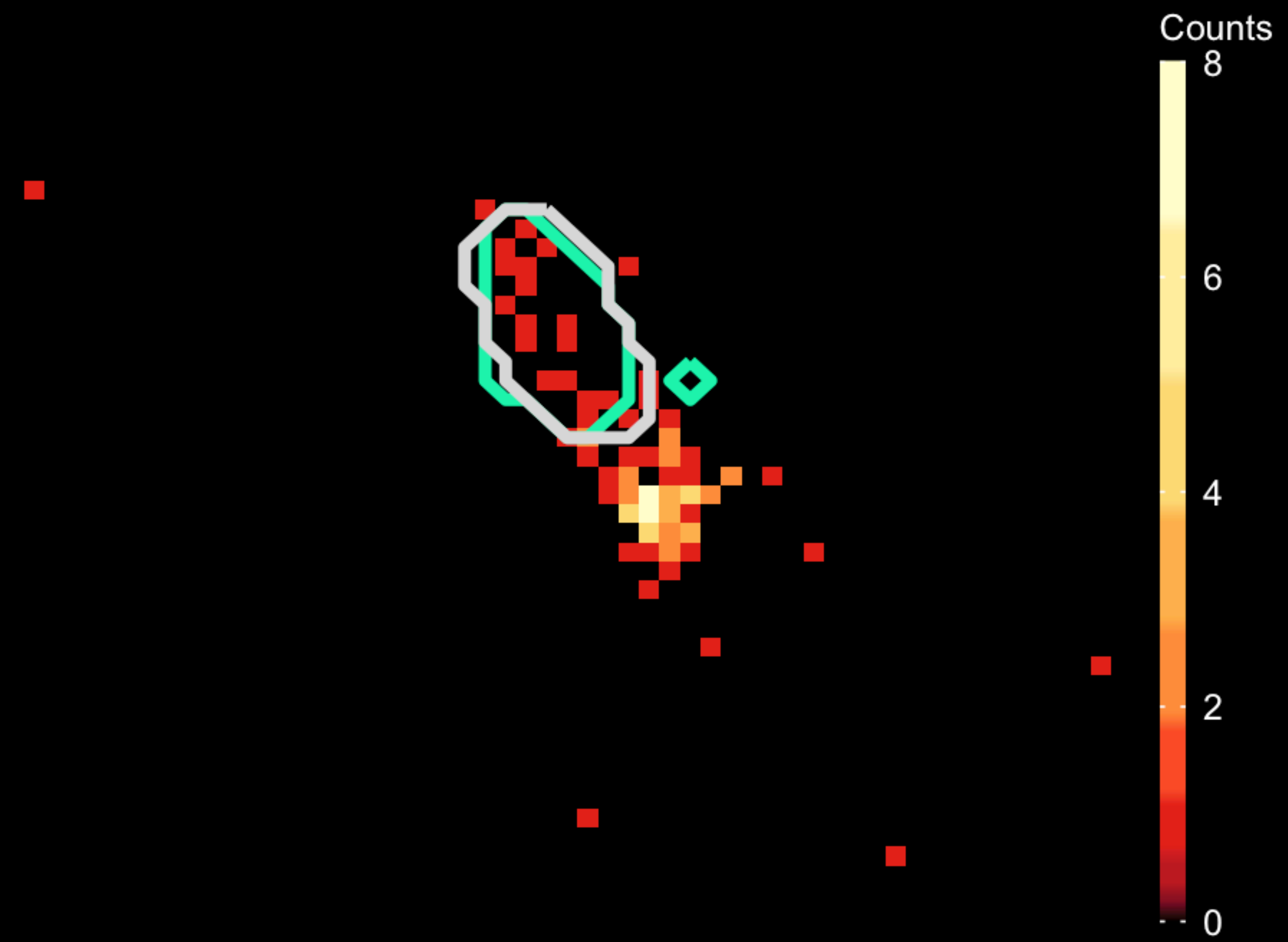
- Stop after a maximum number of iterations (1000)

FINAL BOUNDARY

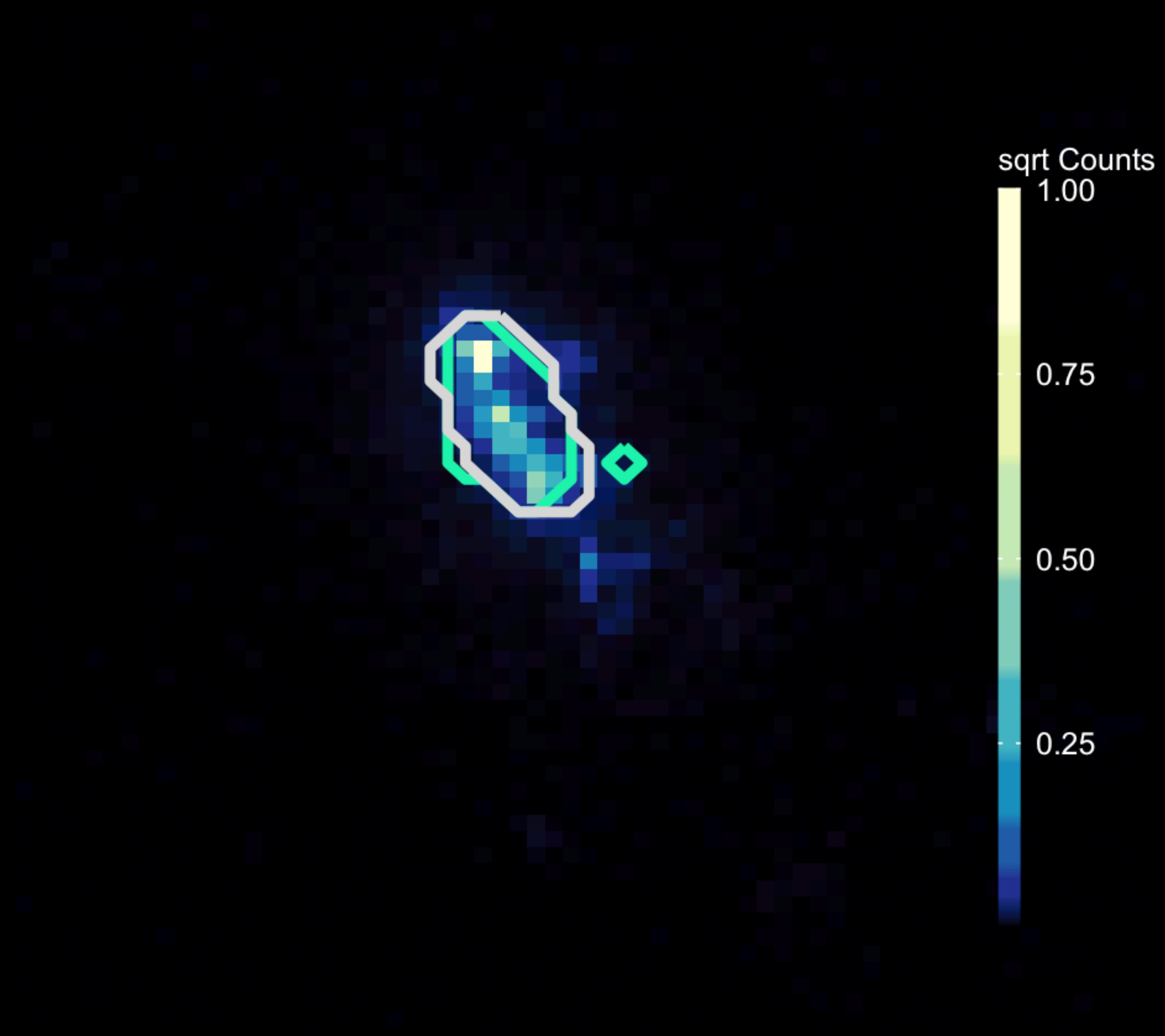
- Find global maximum within unique Z of final generation of



APPLICATION TO OBS ID 7873



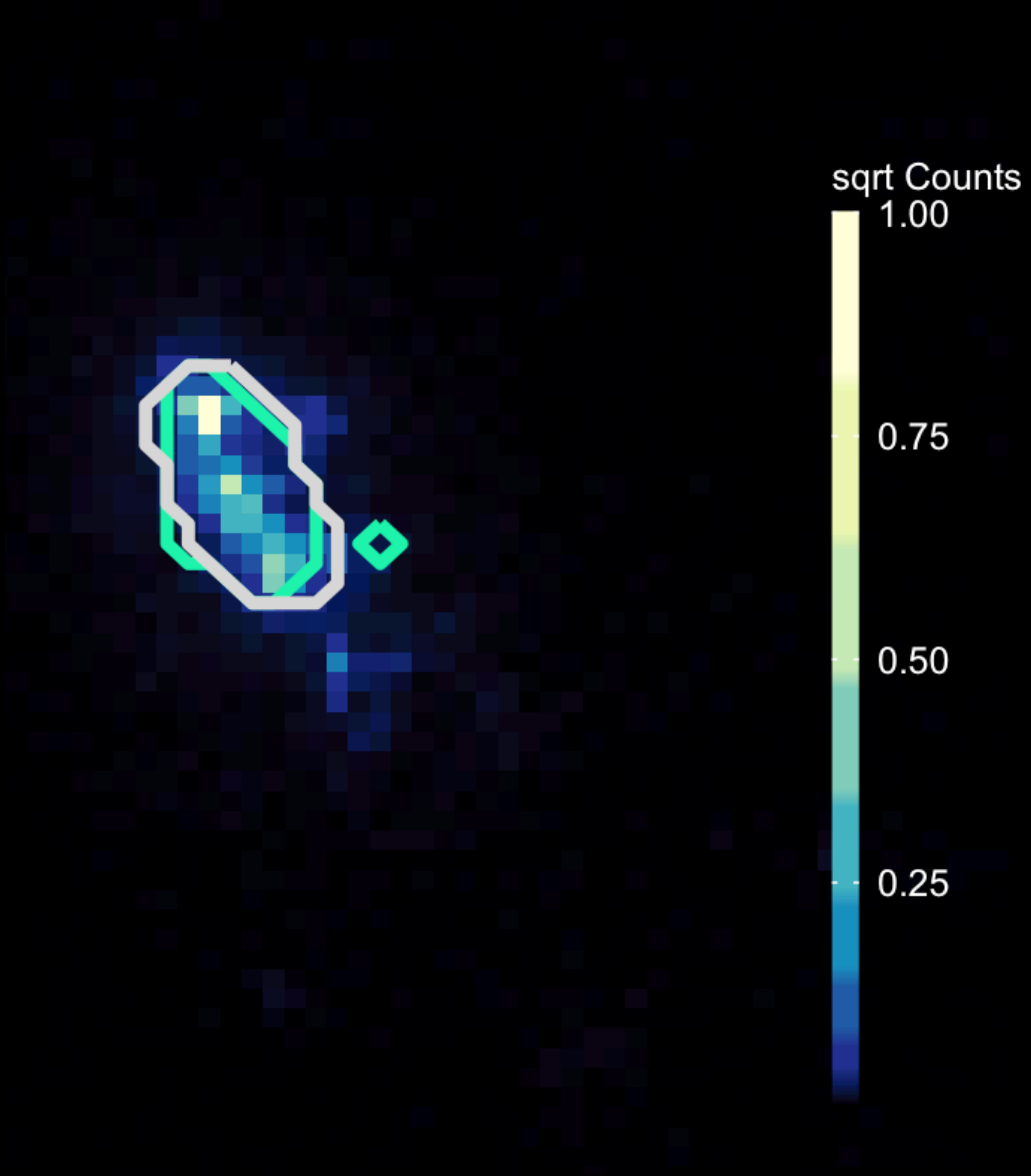
X-Ray Counts



Expected Multi-scale Counts (LIRA)

AN IMPROVEMENT?

- The Z maximized using the genetic algorithm maximizes better than just using the ad hoc selection : $R = 1.5 \times 10^{12}$
- Ad hoc version looks at fixed number of possible boundaries
  - ($64 \times 64 = 4096$)
- Genetic algorithm explores many more possibilities in the relevant region of the posterior
  - > 100,000 pixel assignments considered



POSSIBLE EXTENSIONS

- Uncertainty — is there a way get an error bound on our final estimate
- Smarter mutations
  - Probability of mutation correlated with whether or not the pixel matches it's neighbors (Yu, 1998)
  - Pixels are swapped with local pixels rather than flipped randomly (scramble or swap mutations)

REFERENCES

- Connors, A. and D. A. vanDyk (2007).
- Goldberg, D. E. and K. Deb (1991).
- McKeough, K., A. Siemiginowska, C. C. Cheung, L. Stawarz, V. L. Kashyap,N. Stein, V. Stampoulis, D. A. van Dyk, J. F. C. Wardle, N. P. Lee, D. E.Harris, D. A. Schwartz, D. Donato, L. Maraschi, and F. Tavecchio (2016).
- Pereira, D. C., R. P.Ramos, and M. Z. do Nascimento (2014).
- Rajpaul, V. (2012).
- Sheta, A., M. S. Braik, and S. Al-jahdali (2012).
- Stein, N. M.,D. A. Van Dyk, V. L. Kashyap, and A. Siemiginowska (2015).
- Yu, M. (1998).