Automatic classification of galaxy morphology with ZEST+.

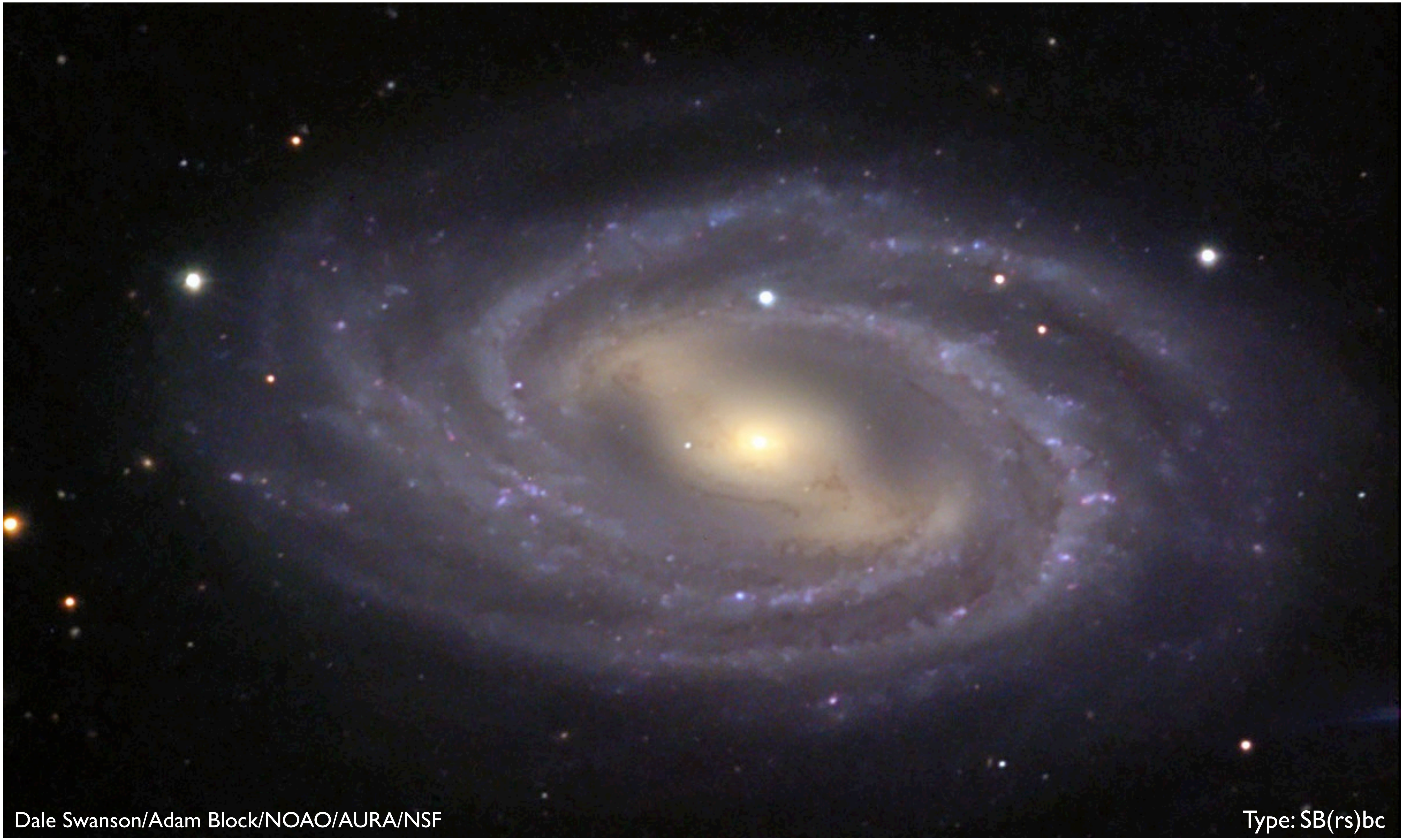
Mariano Ciccolini
Institut für Astronomie, ETHZ

Automatic classification of galaxy morphology with ZEST+.

Mariano Ciccolini
Institut für Astronomie, ETHZ

- Galaxies and their morphologies
- Support Vector Machines
- ZEST+

M109: A barred spiral galaxy



Early, purely descriptive systems

W. Herschel & J. Herschel (1780 -1860)

- very faint
- Faint
- Bright
- small (with definite borders)
- exceedingly large

J. Dreyer (1888)

- brightness
- form
- concentration

M. Wolf (1908)

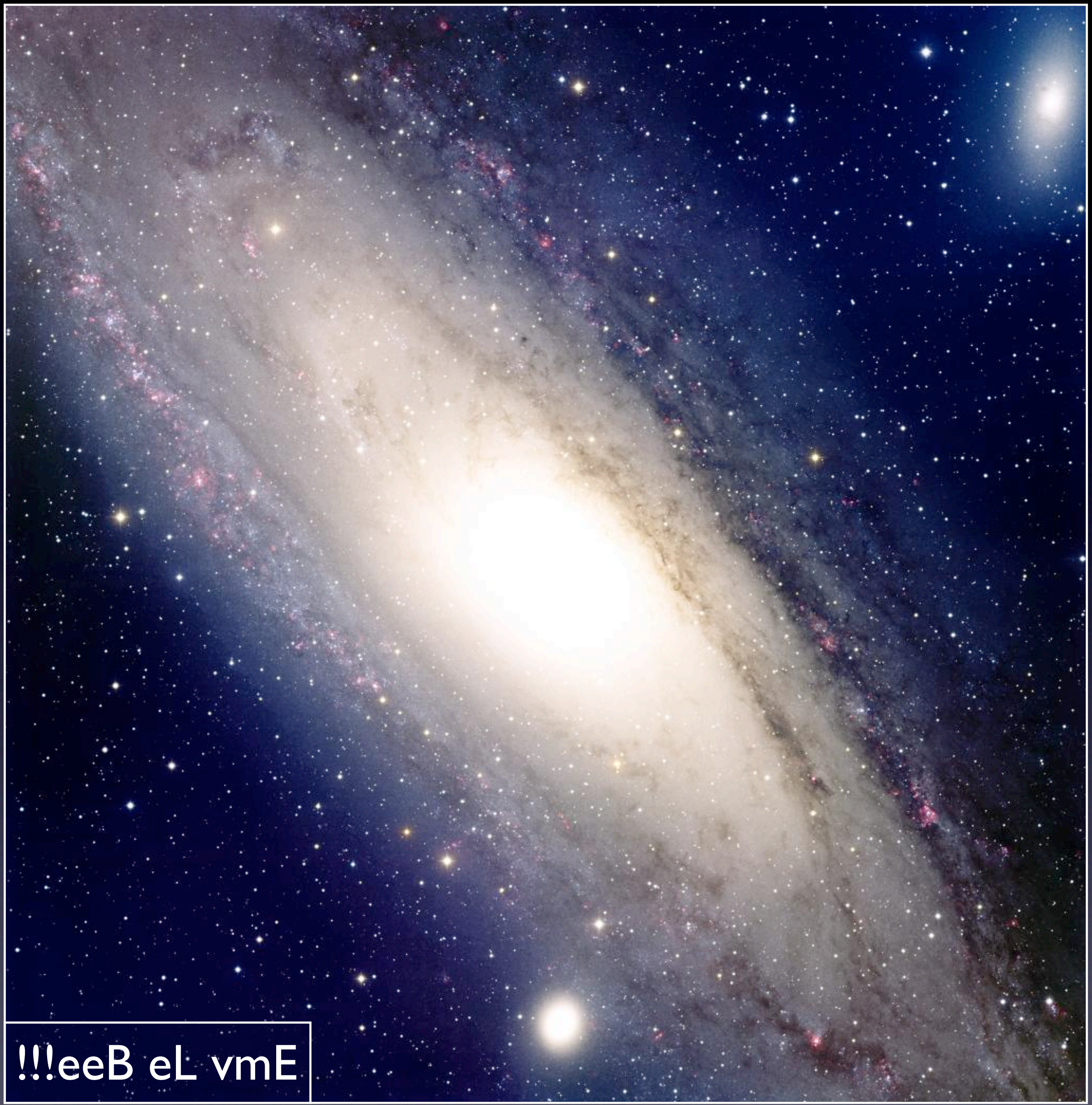
- 17 classes (g) - (w)

Early, purely descriptive systems

W. Herschel & J. Herschel (1780 -1860)
- very faint
- Faint
- Bright
- small (with definite borders)
- exceedingly large

J. Dreyer (1888)
- brightness
- form
- concentration

M. Wolf (1908)
- 17 classes (g) - (w)



T.A.Rector and B.A.Wolpa/NOAO/AURA/NSF

Early, purely descriptive systems

W. Herschel & J. Herschel (1780 -1860)

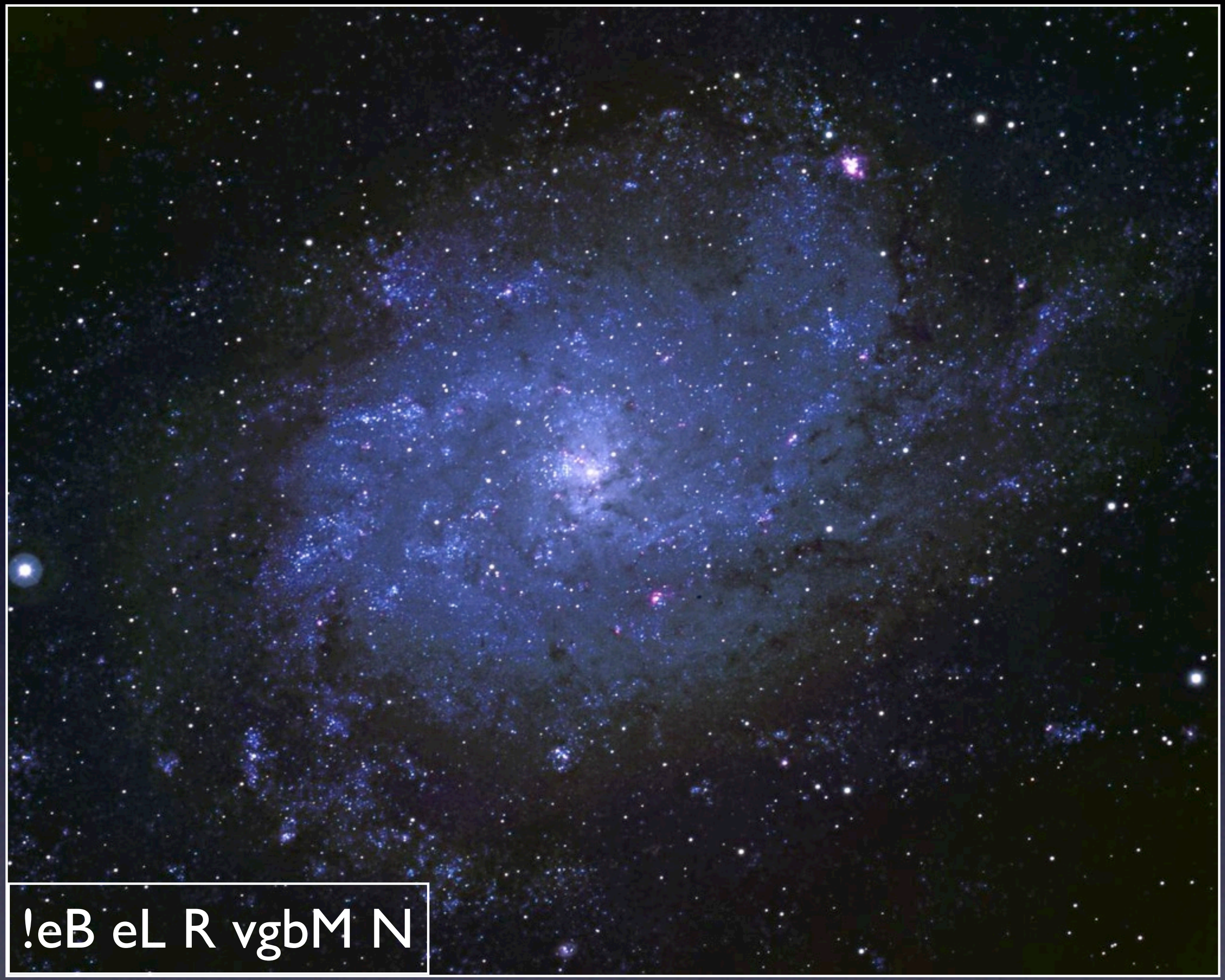
- very faint
- Faint
- Bright
- small (with definite borders)
- exceedingly large

J. Dreyer (1888)

- brightness
- form
- concentration

M. Wolf (1908)

- 17 classes (g) - (w)



Bill Schoening/NOAO/AURA/NSF

Early, purely descriptive systems

W. Herschel & J. Herschel (1780 -1860)

- very faint
- Faint
- Bright
- small (with definite borders)
- exceedingly large

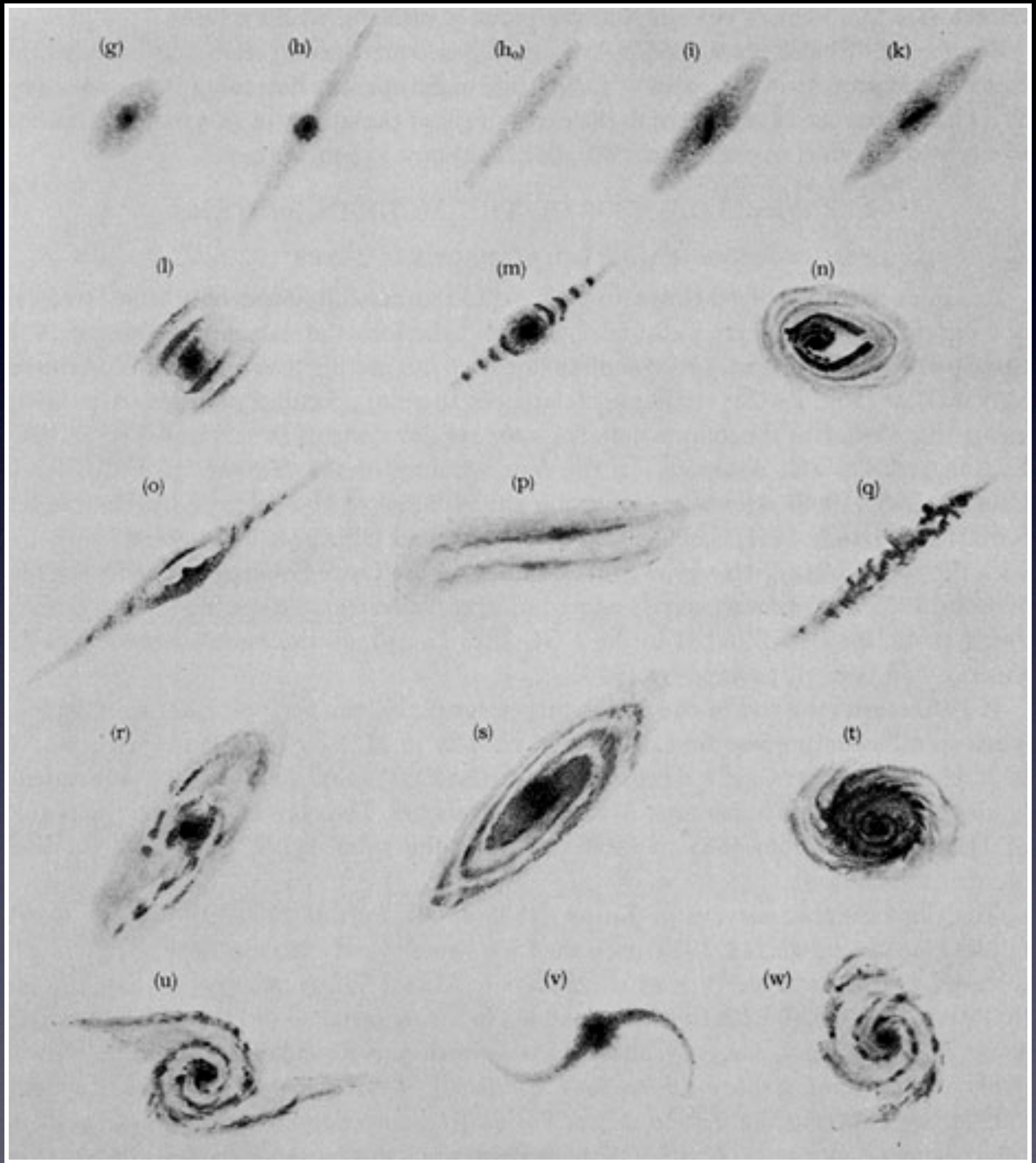
J. Dreyer (1888)

- brightness
- form
- concentration

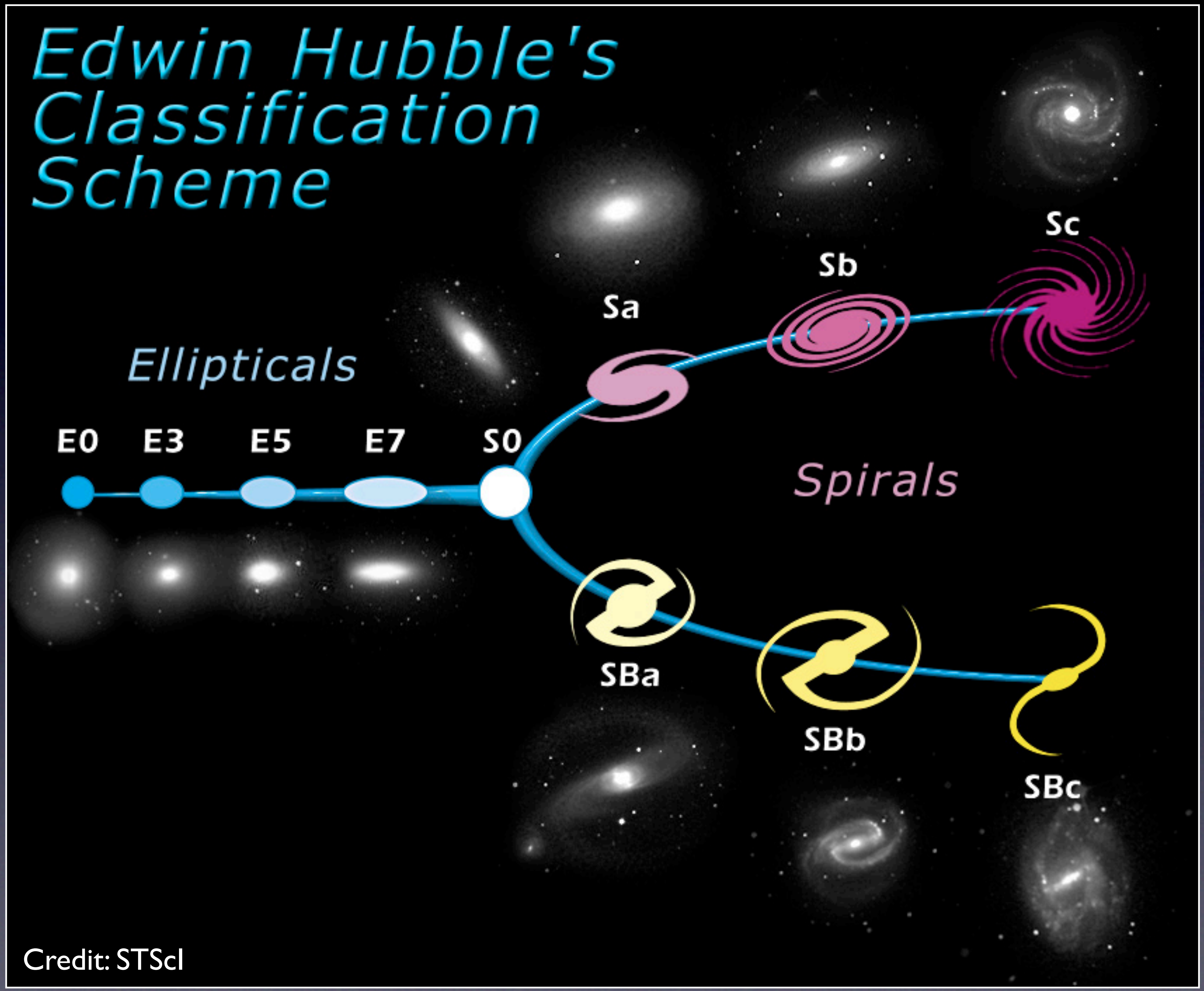
M. Wolf (1908)

- 17 classes (g) - (w)

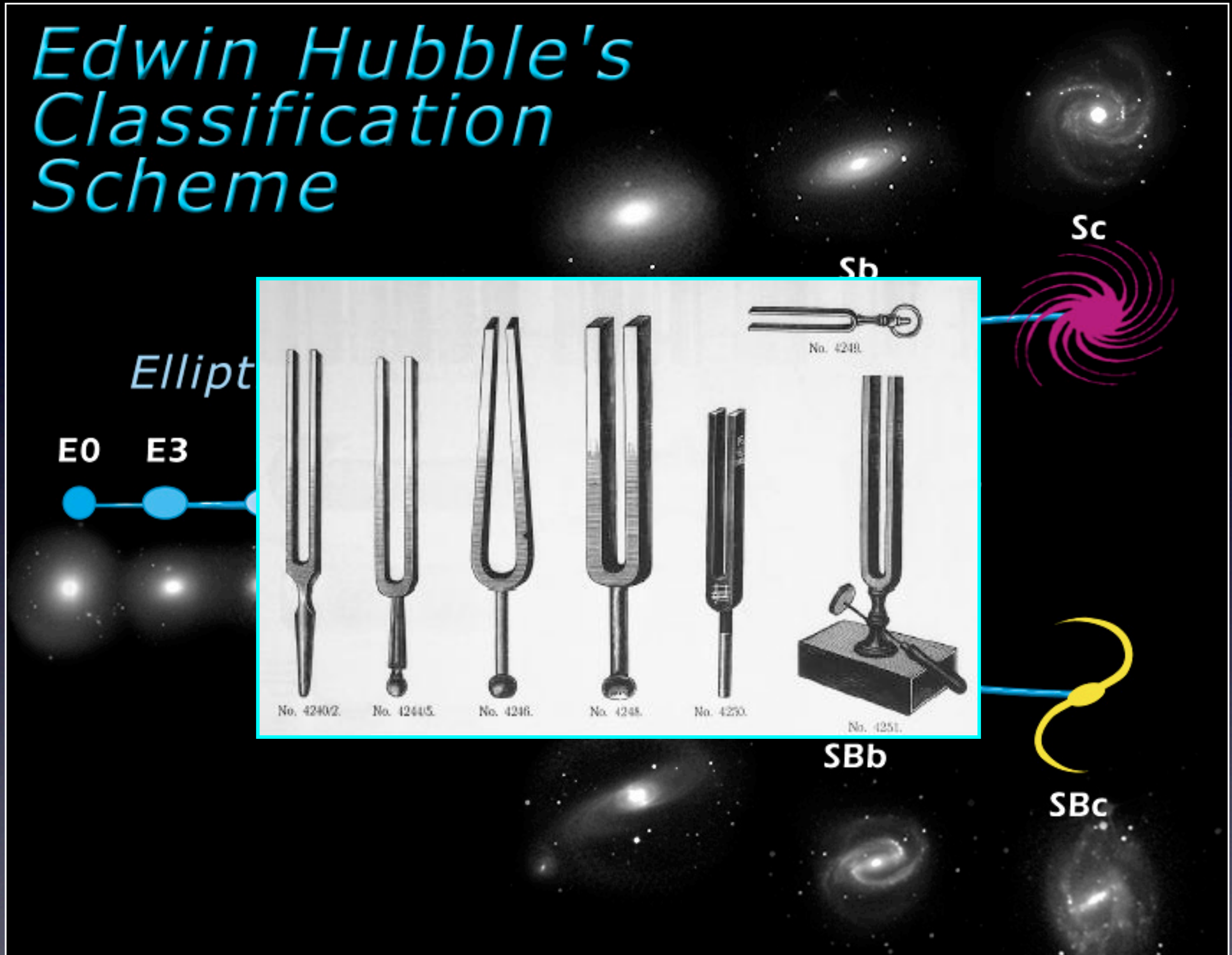
"Es gibt keine zwei Nebelflecken am Himmel, die sich gleichen. Trotzdem geht das Bestreben der Beobachter seit Herschel dahin, die kleinen Nebelflecken zu klassifizieren."



E. Hubble (1936): Tuning fork diagram

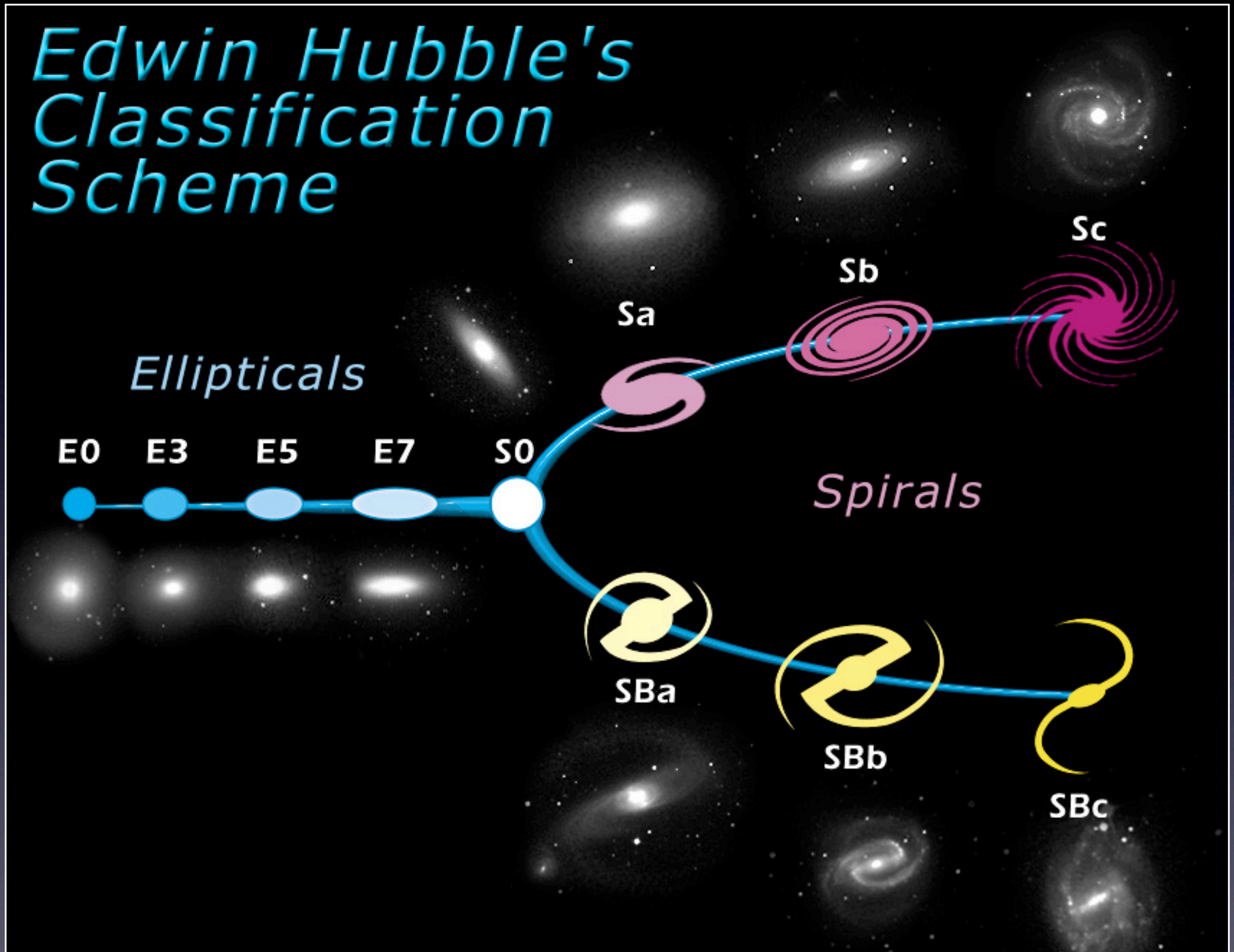


E. Hubble (1936): Tuning fork diagram



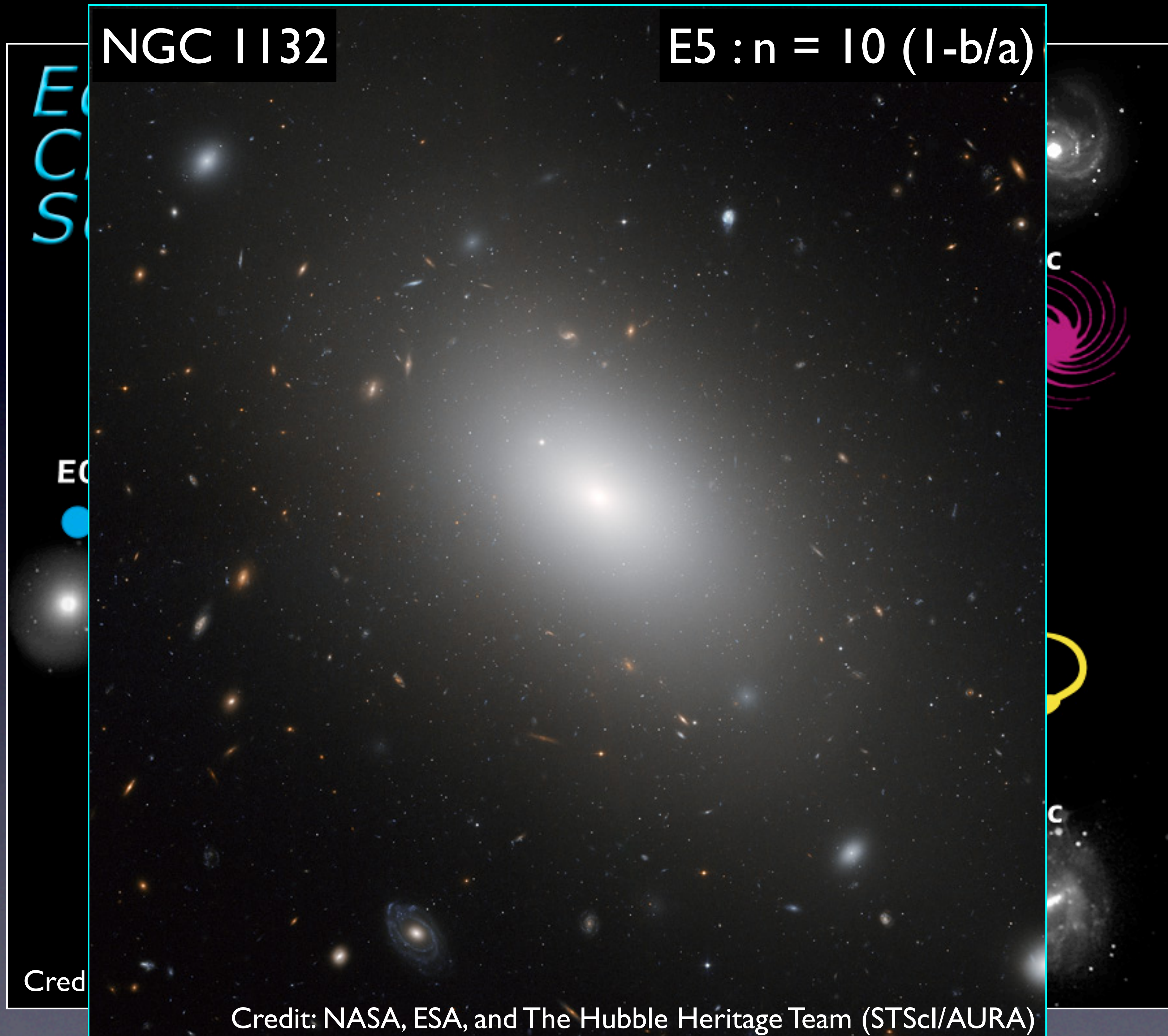
Credit: STScI

E. Hubble (1936): Tuning fork diagram



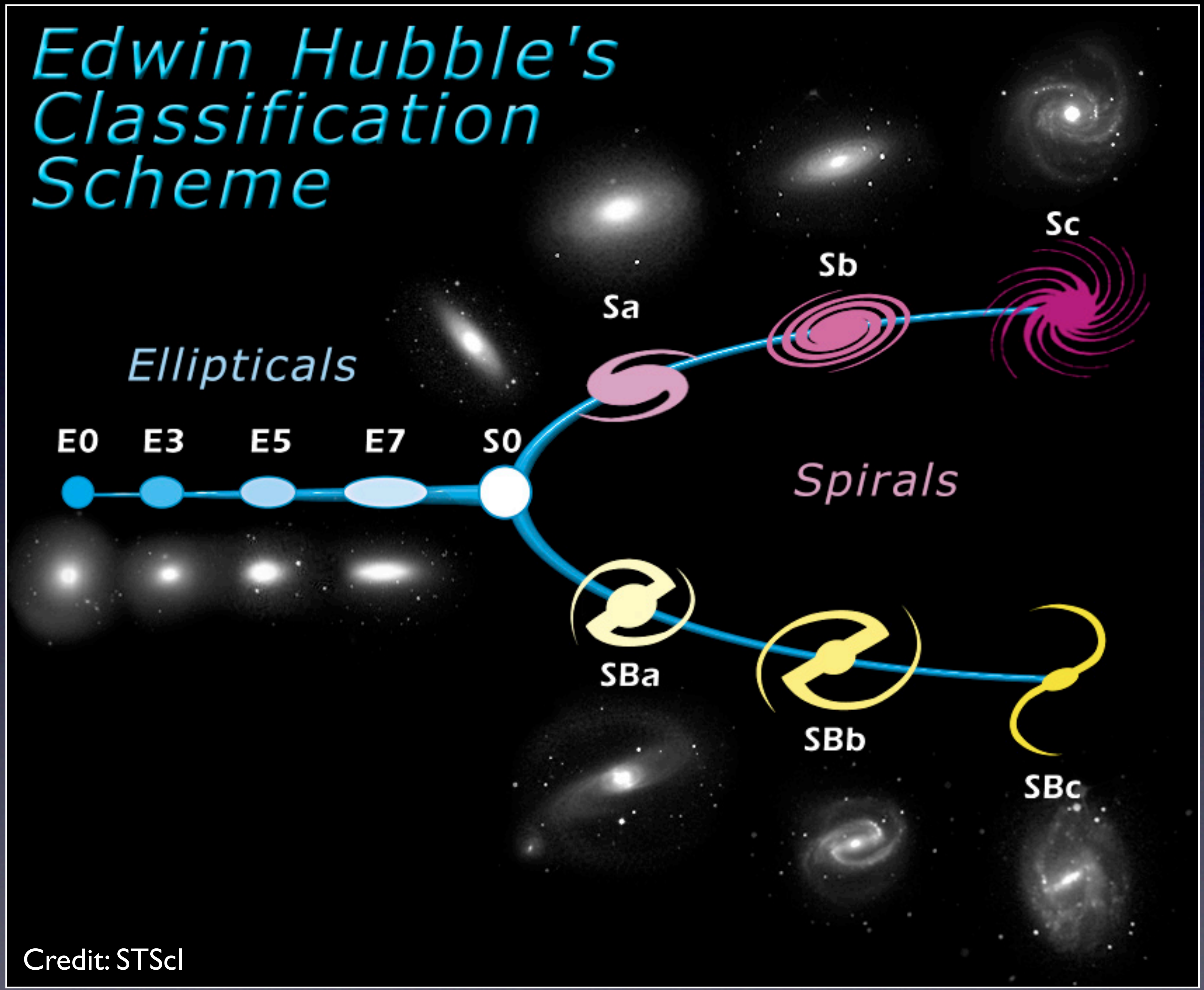
Credit: STScI

E. Hubble (1936): Tuning fork diagram



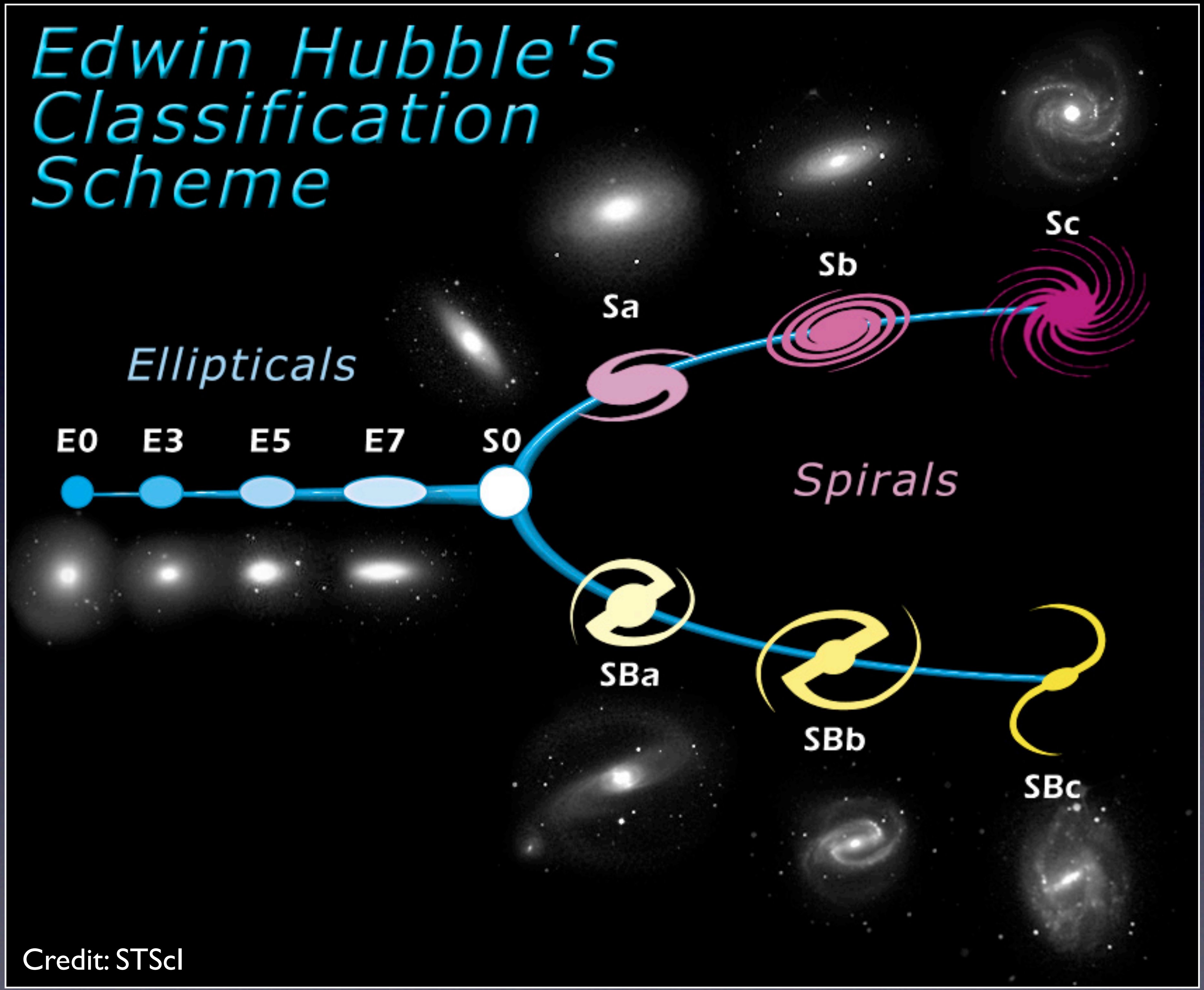
Credit: NASA, ESA, and The Hubble Heritage Team (STScI/AURA)

E. Hubble (1936): Tuning fork diagram



E. Hubble (1936): Tuning fork diagram

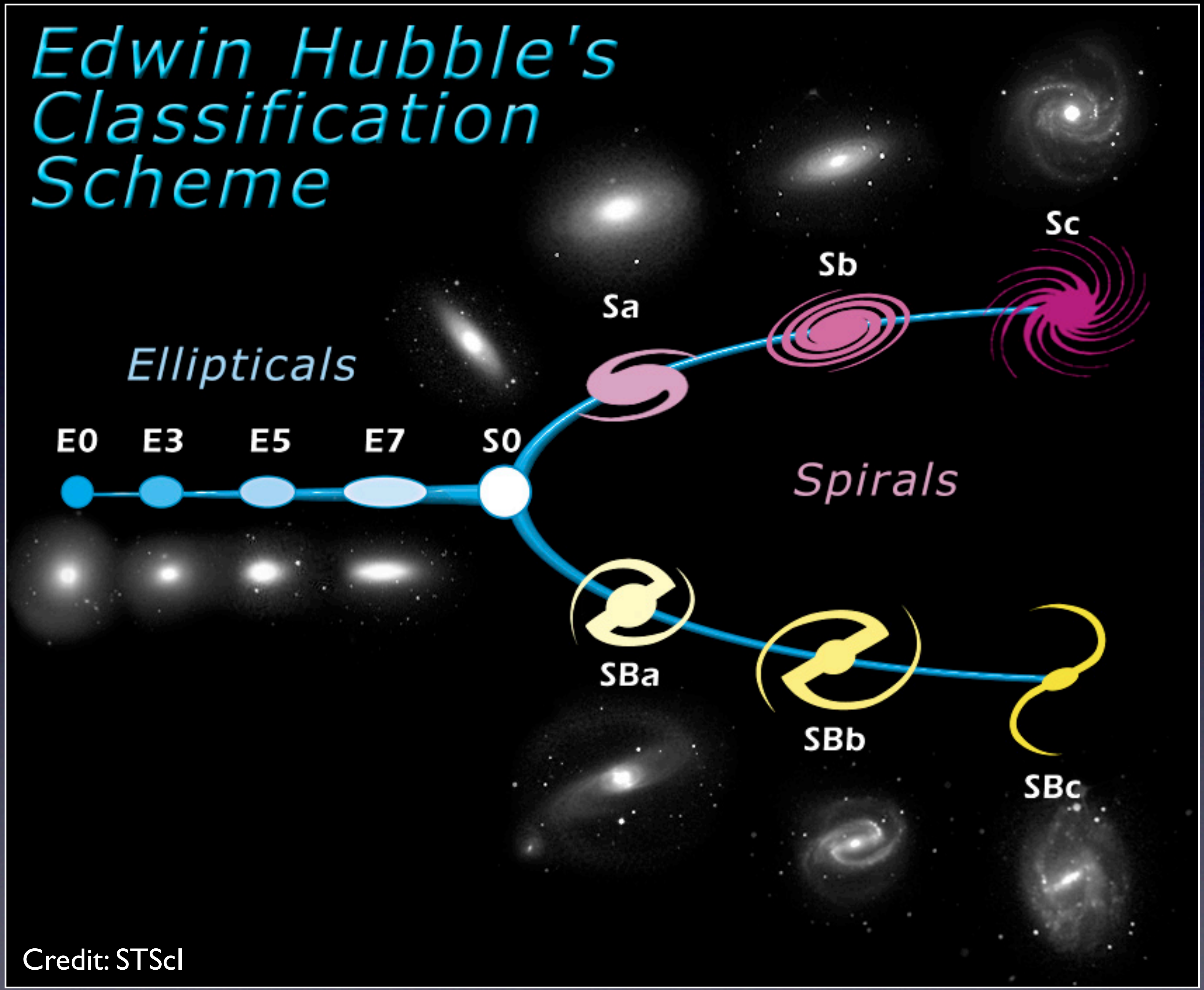
E. Hubble (1936): Tuning fork diagram



E. Hubble (1936): Tuning fork diagram

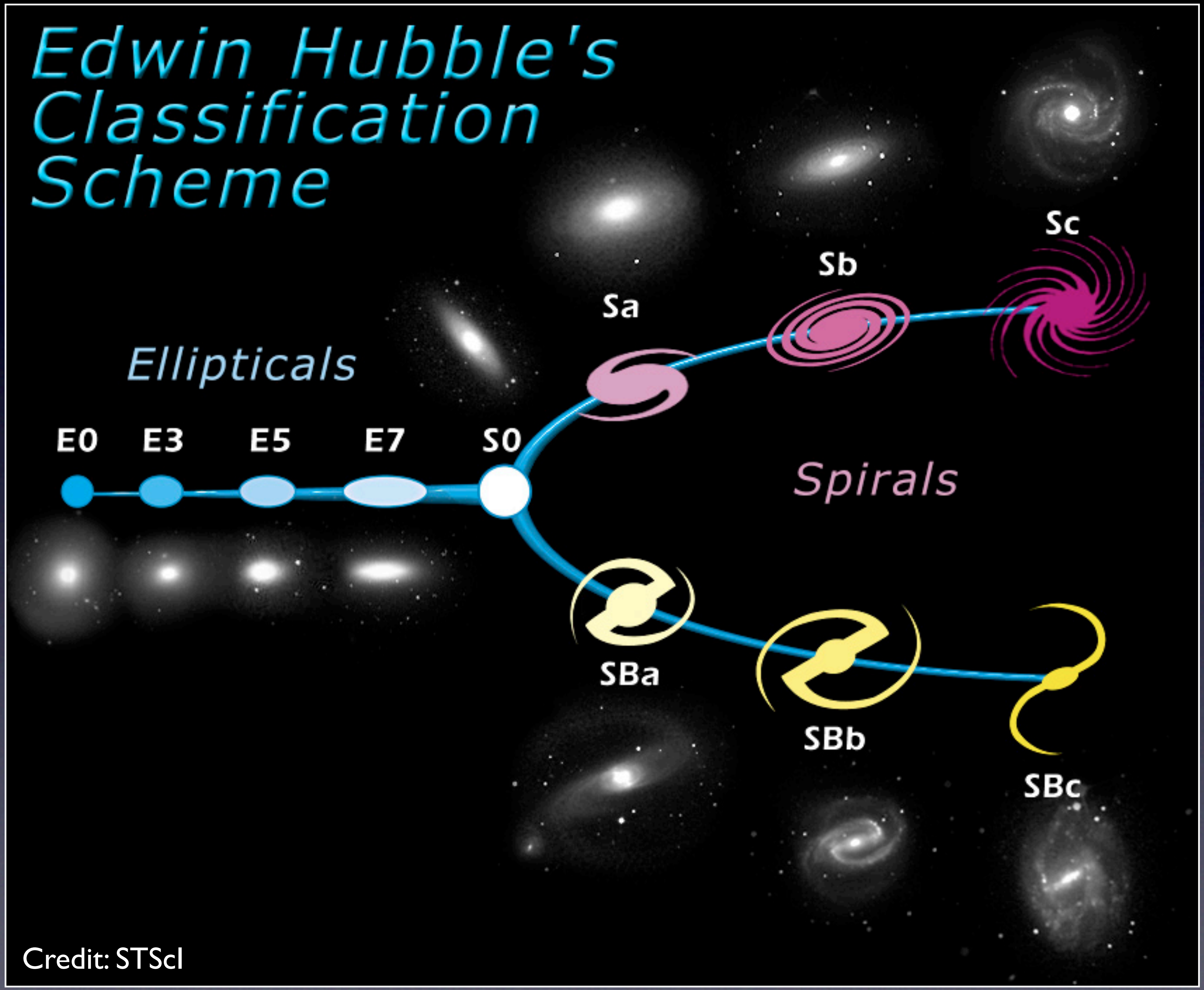
Credit: NASA, ESA, and The Hubble Heritage Team (STScI/AURA)
Credit: STScI

E. Hubble (1936): Tuning fork diagram



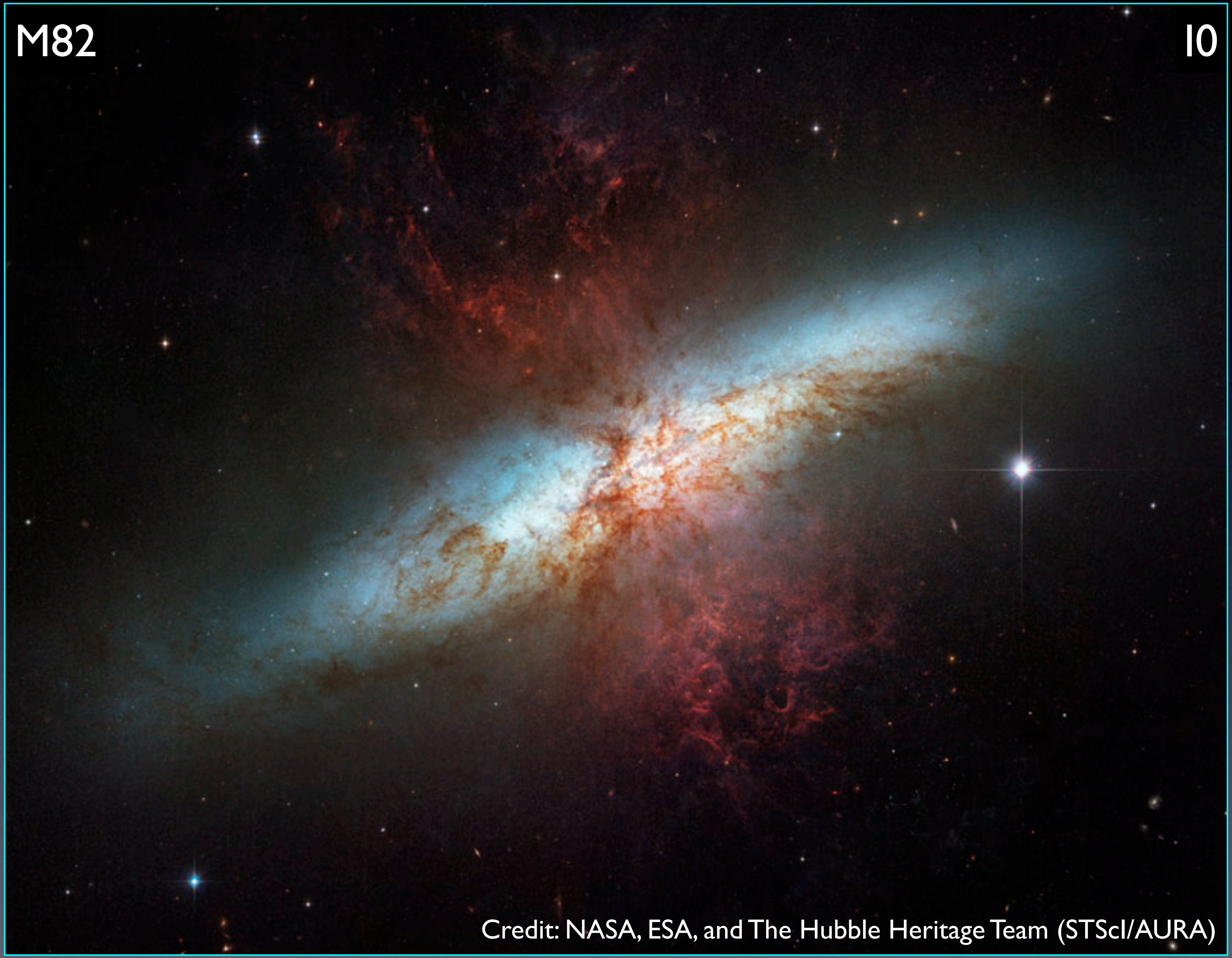
E. Hubble (1936): Tuning fork diagram

E. Hubble (1936): Tuning fork diagram

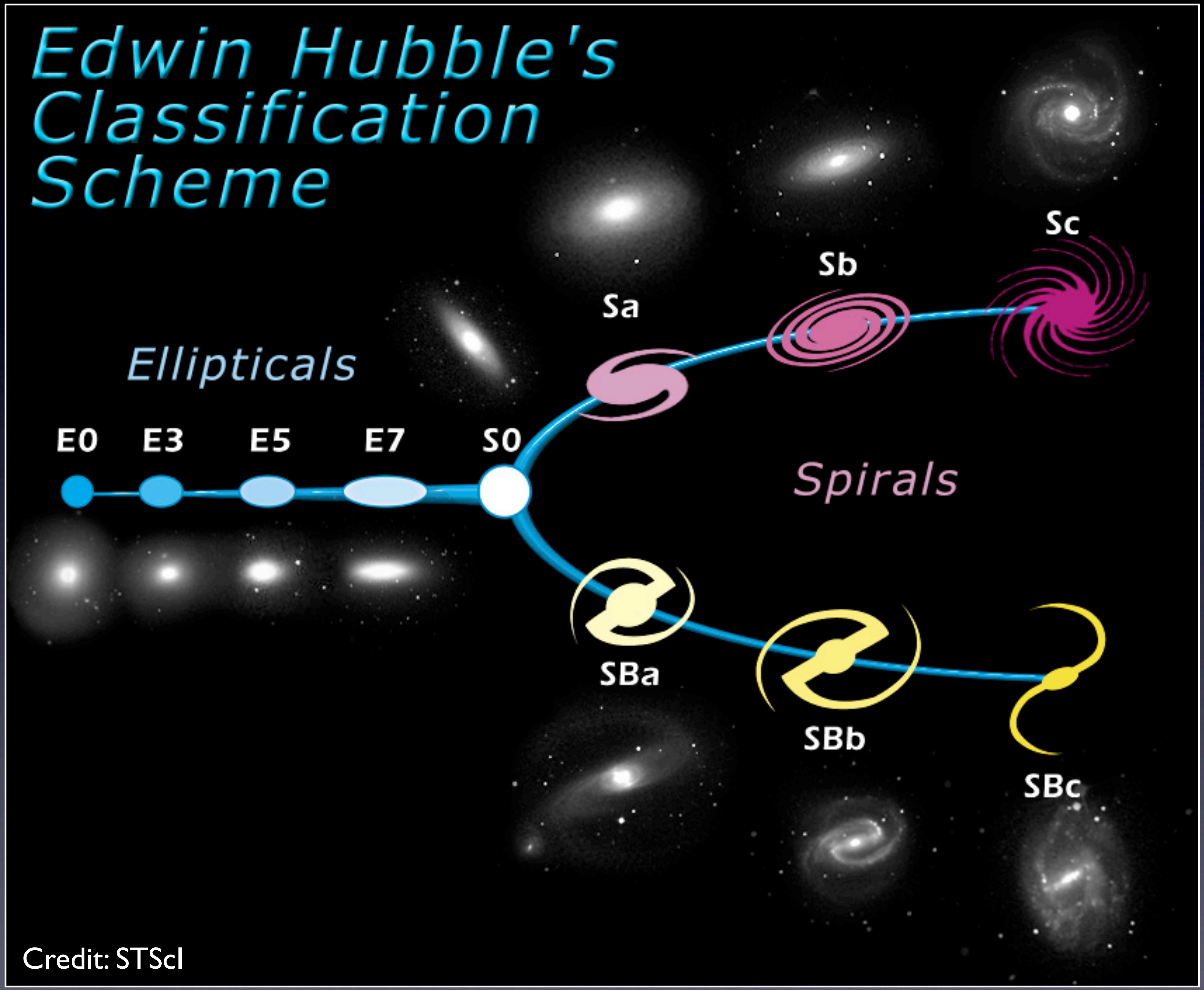


Credit: STScI

E. Hubble (1936): Tuning fork diagram

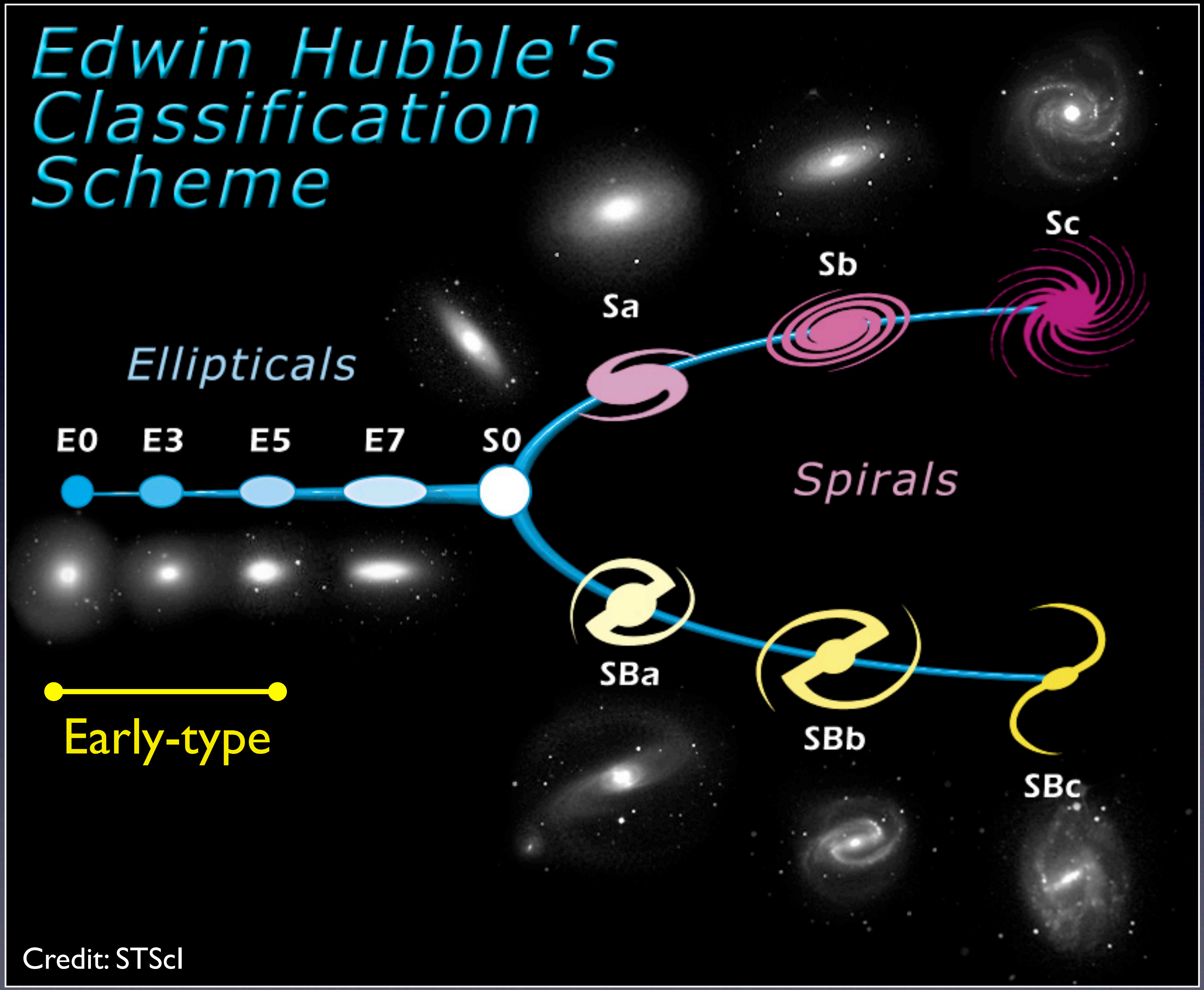


E. Hubble (1936): Tuning fork diagram



E. Hubble (1936): Tuning fork diagram

Late-type

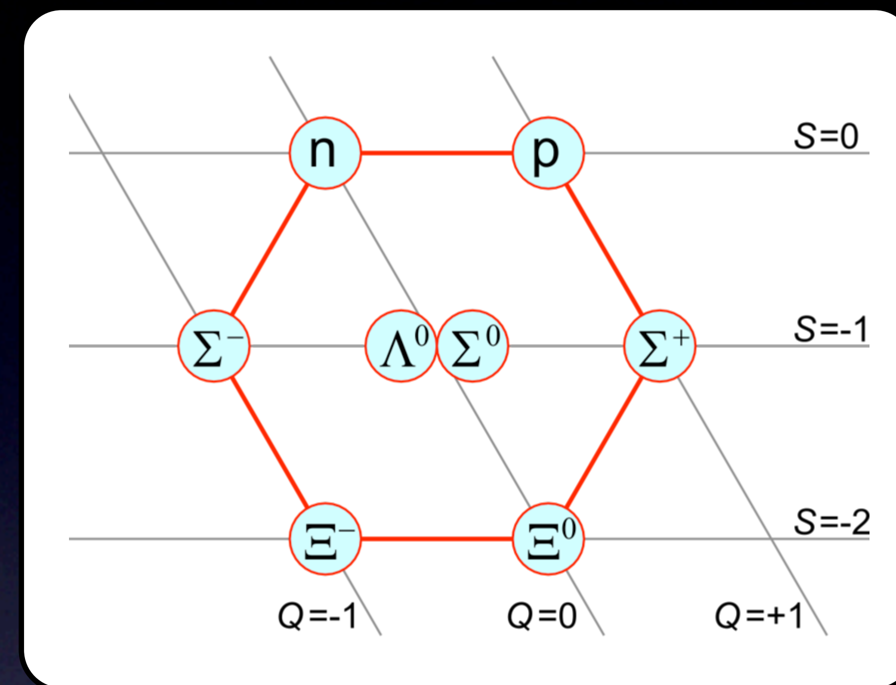


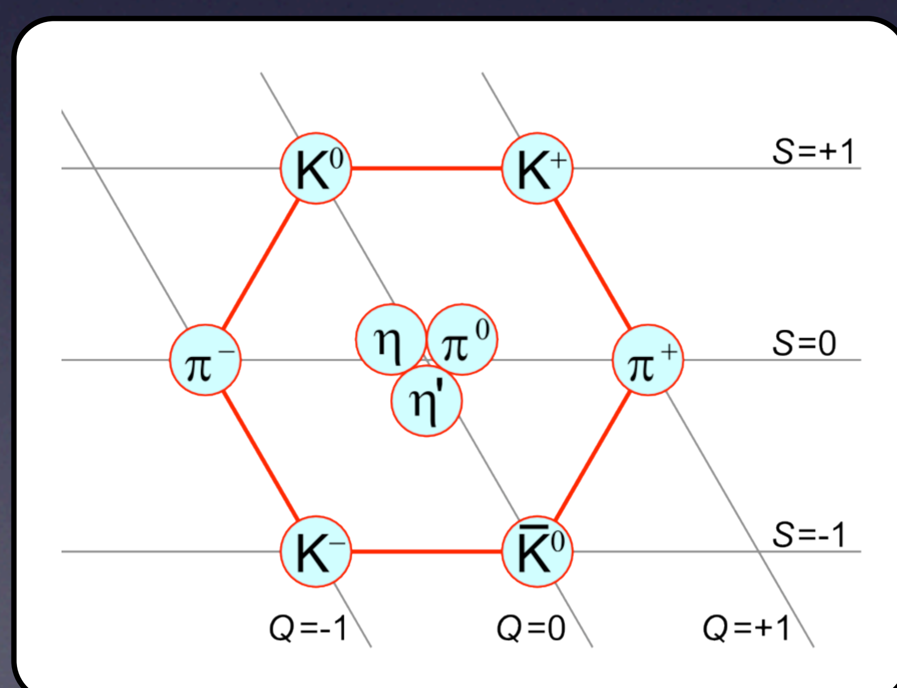
Early-type

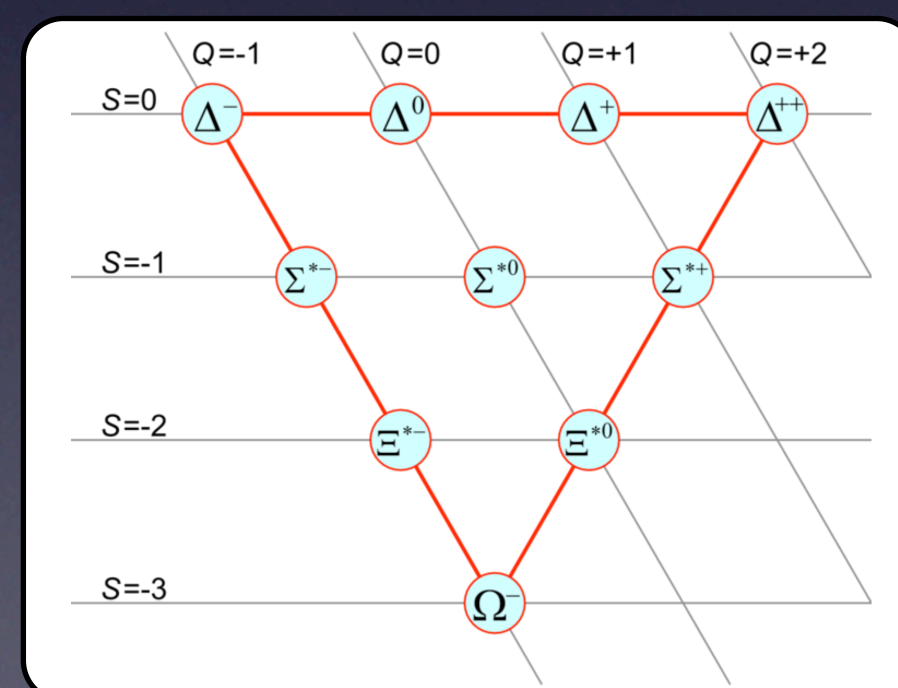
Why classify?

"The ultimate purpose of the classification is to understand galaxy formation and evolution." A. Sandage

HEP example: **Hadrons classification scheme** → **Eight-fold way** → **Quark model & QCD**







Parametric approach

Light decomposition: Fit a 1D/2D model to surface brightness profile

Sérsic profile:

$$I(R) = I_e \exp\left\{-b_n\left[\left(\frac{R}{R_e}\right)^{1/n} - 1\right]\right\}$$

$$\int_0^{R_e} I(R) = \frac{1}{2}\int_0^{\infty} I(R) \quad \Rightarrow \quad \Gamma(2n) = 2\gamma(2n, b_n)$$

n=1 : Exponential profile
n=4 : de Vaucouleurs profile

$$I(R) = I_4(R) + I_1(R) \quad \Rightarrow \quad \frac{B}{D} = \frac{\int_0^{\infty} I_4(R)}{\int_0^{\infty} I_1(R)}$$

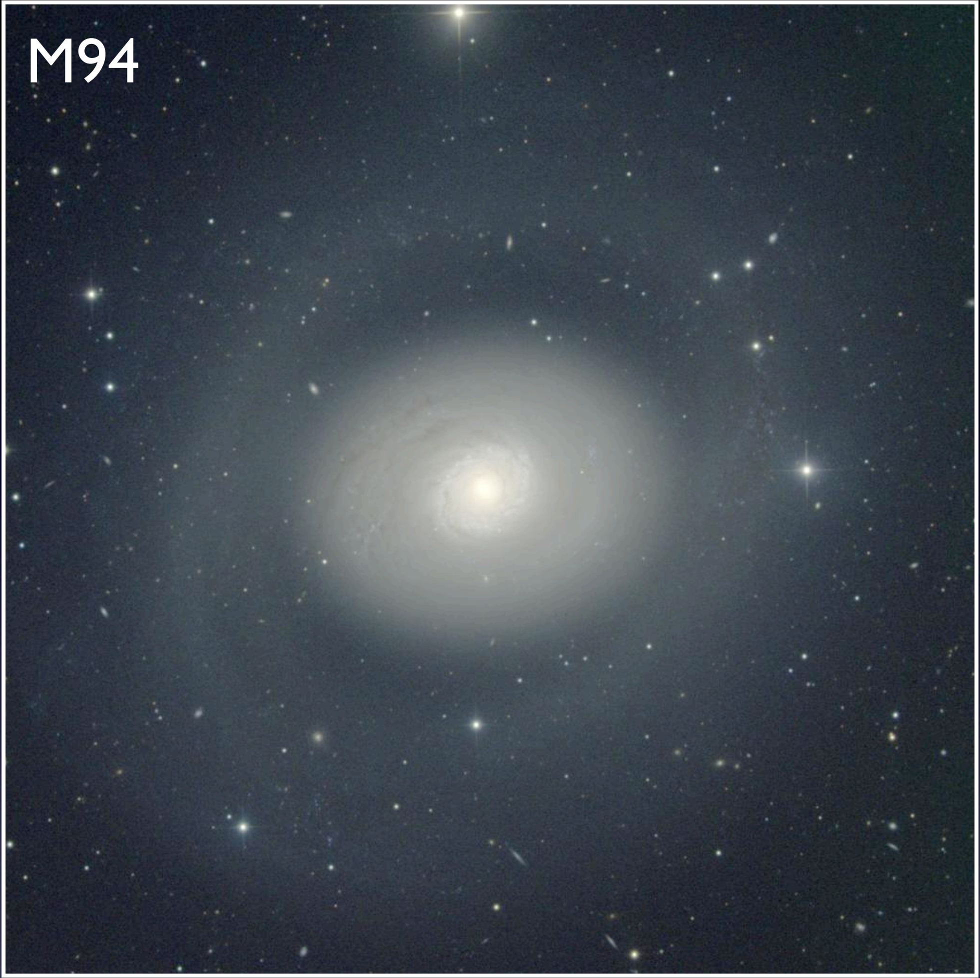
Other approaches:

- Isophote fitting
- Fourier decomposition
- Shapelets

Parametric approach limitations

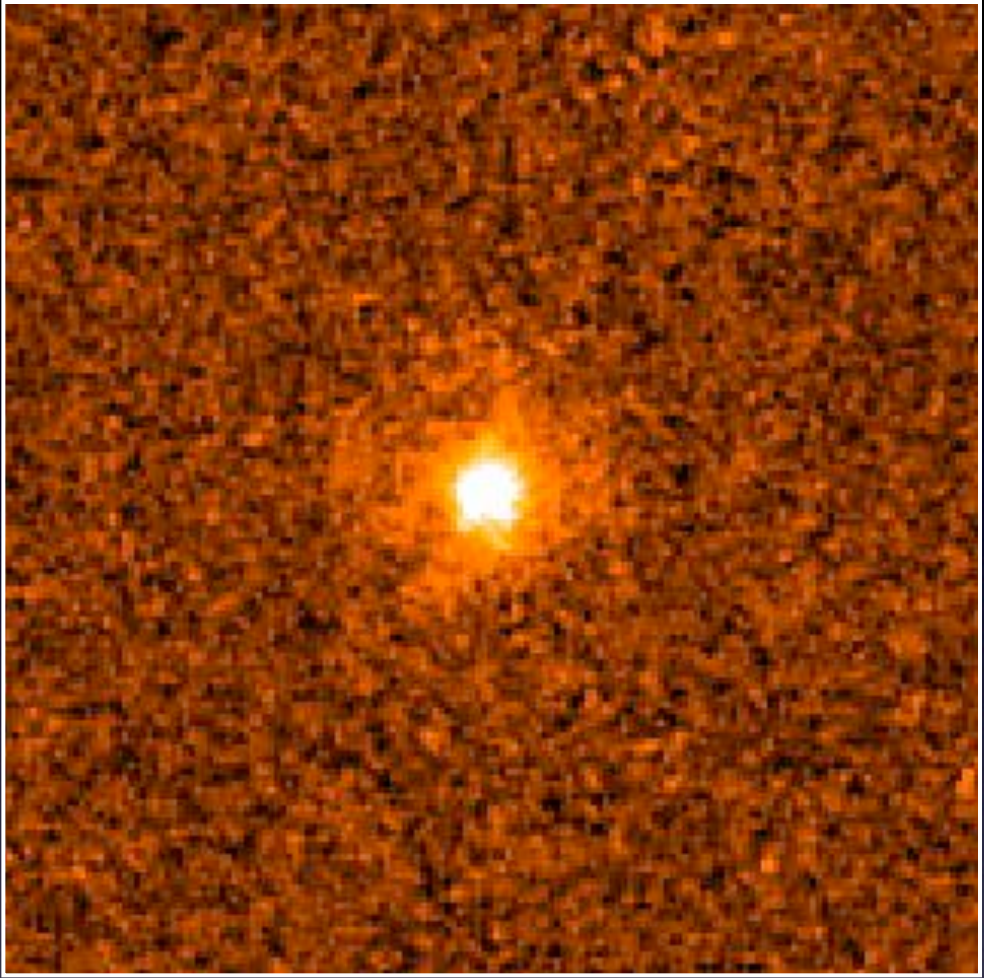
Redshift	Lookback time (Gy)	Morphology
z < 0.3	< ~3.5	Hubble scheme applies in full detail.
z ~ 0.5	~5	Barred spirals rare Underdeveloped arms
z > 0.6	~6	Mergers and Irr increase z~1 : 30% off HS

Parametric approach limitations



Hillary Mathis, N.A.Sharp/NOAO/AURA/NSF

Nearby galaxy (z ~ 0.001)



COSMOS HST-ACS survey

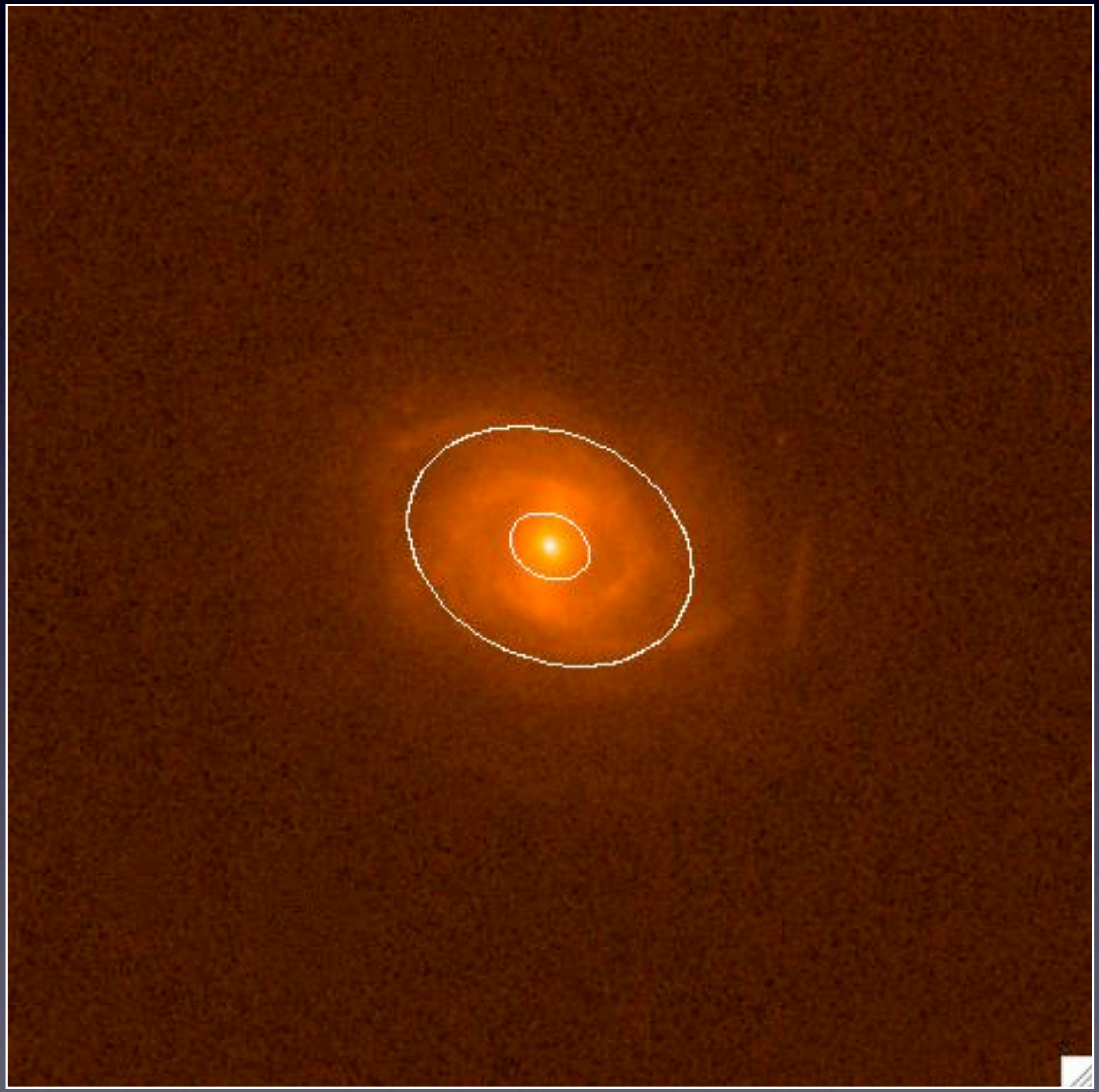
Distant galaxy (z ~ 1.2)

Non-parametric approach: Concentration

$$C = \log\left(\frac{r_{20}}{r_{80}}\right)$$

$$r_{20} \quad : \sum_{i:r(i)\leq r_{20}} I_i < 0.20\, I_{\text{tot}}$$

$$r_{80} \quad : \sum_{i:r(i)\leq r_{80}} I_i < 0.80\, I_{\text{tot}}$$

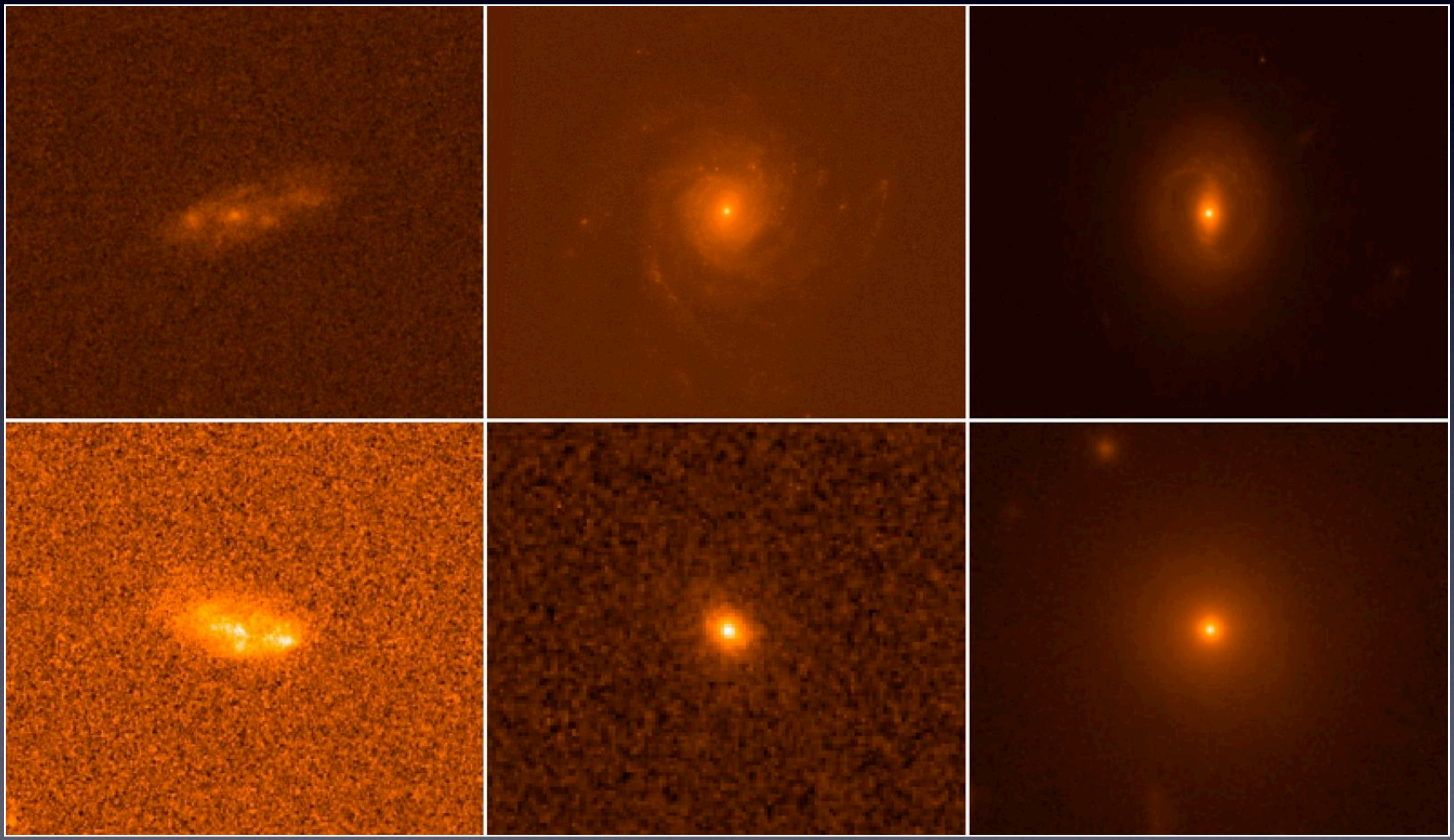


Non-parametric approach: Concentration

$$C = \log\left(\frac{r_{20}}{r_{80}}\right)$$

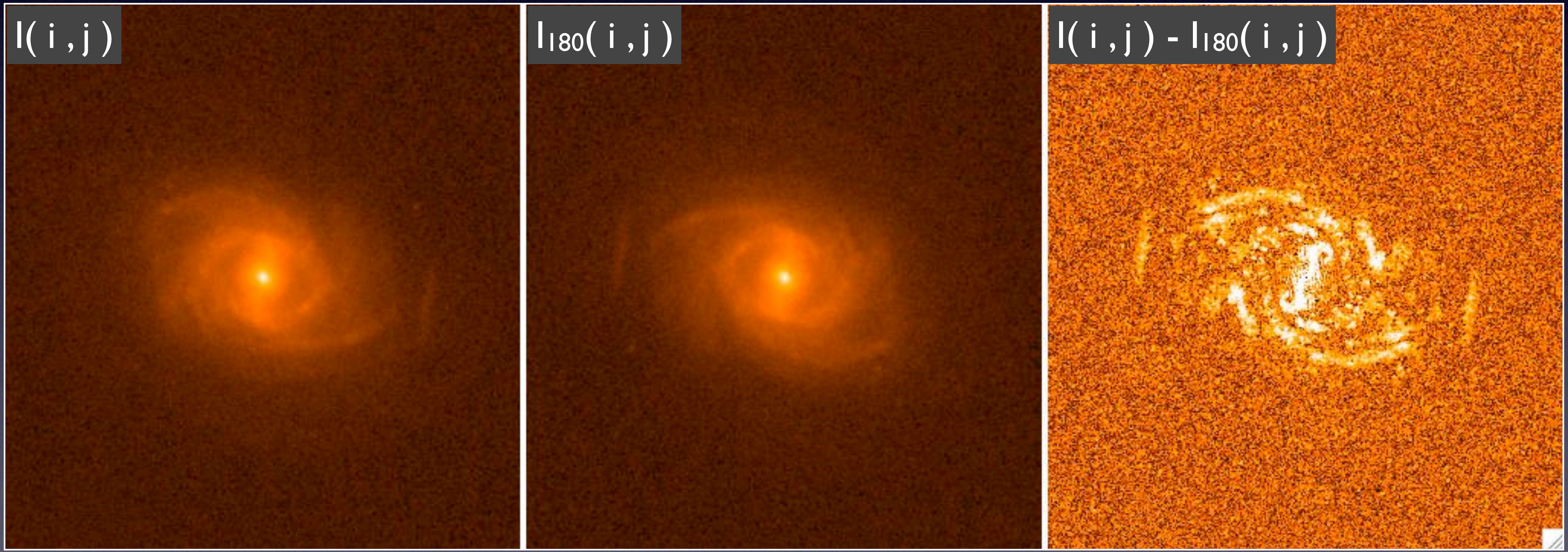
$$r_{20} \quad : \quad \sum_{i:r(i)\leq r_{20}} I_i < 0.20\, I_{\mathrm{tot}}$$

$$r_{80} \quad : \quad \sum_{i:r(i)\leq r_{80}} I_i < 0.80\, I_{\mathrm{tot}}$$



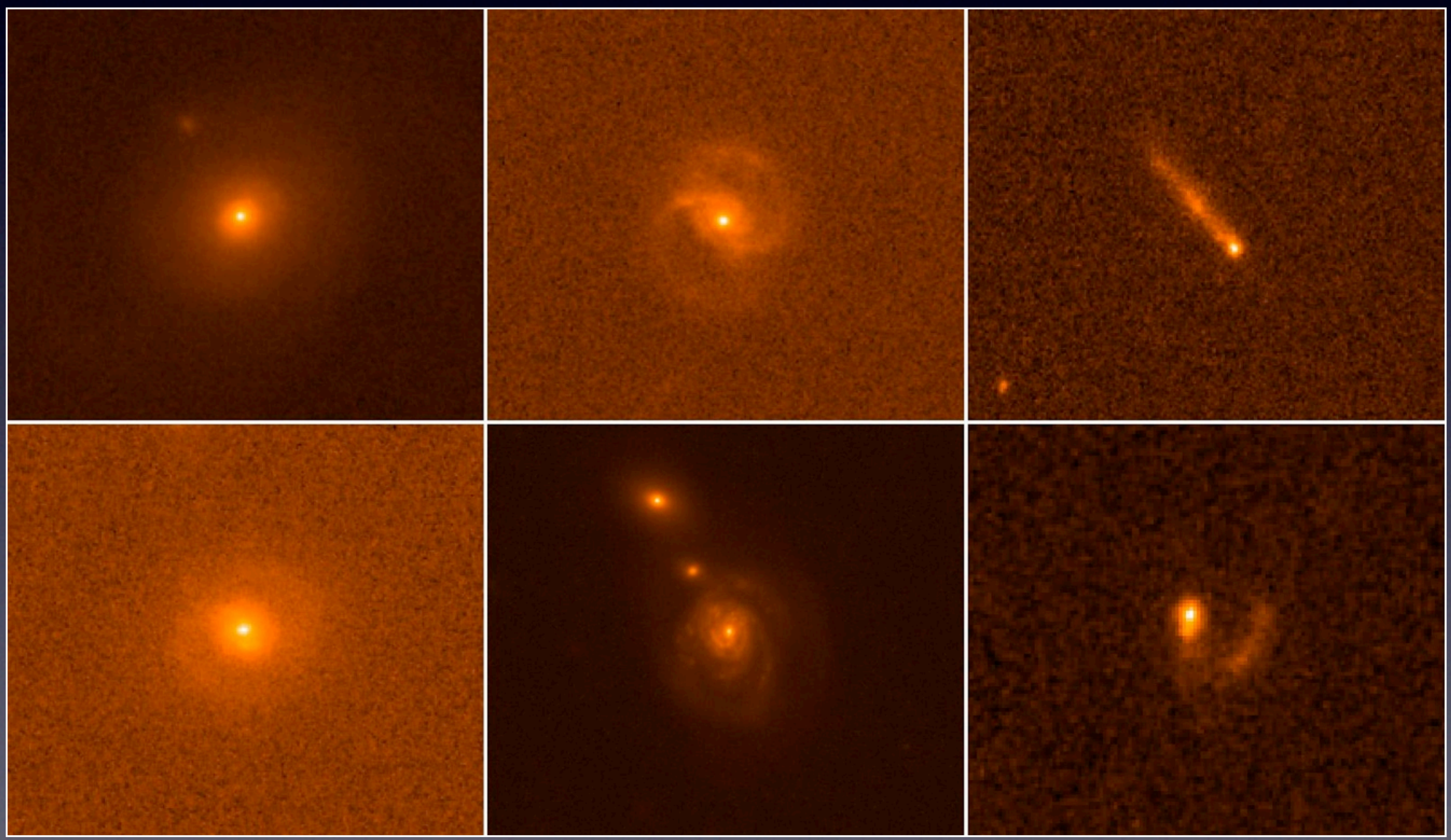
Non-parametric approach: Asymmetry

$$A_0 = \frac{1}{2} \frac{\sum_{ij} |I(i,j) - I_{180}(i,j)|}{\sum_{ij} |I(i,j)|}, \qquad A = A_0 - A_{\rm bkg}$$



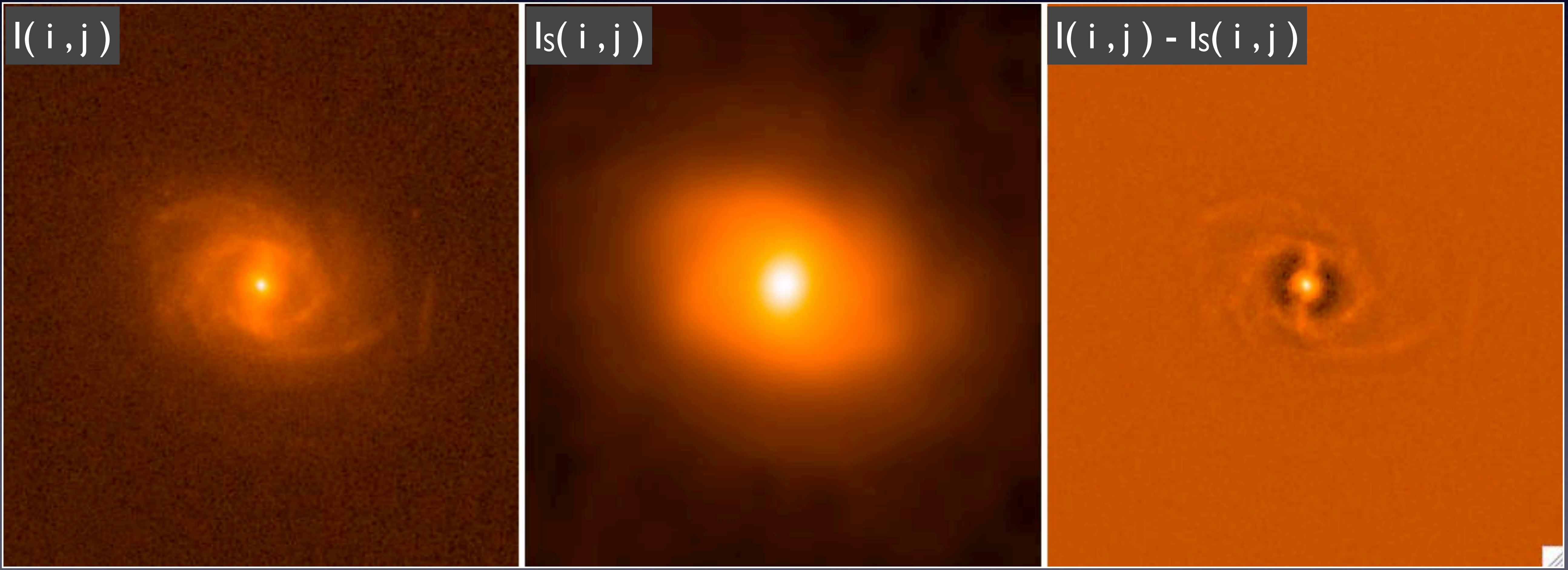
Non-parametric approach: Asymmetry

$$A_0 = \frac{1}{2}\frac{\sum_{ij}|I(i,j) - I_{180}(i,j)|}{\sum_{ij}|I(i,j)|}, \qquad A = A_0 - A_{\rm bkg}$$



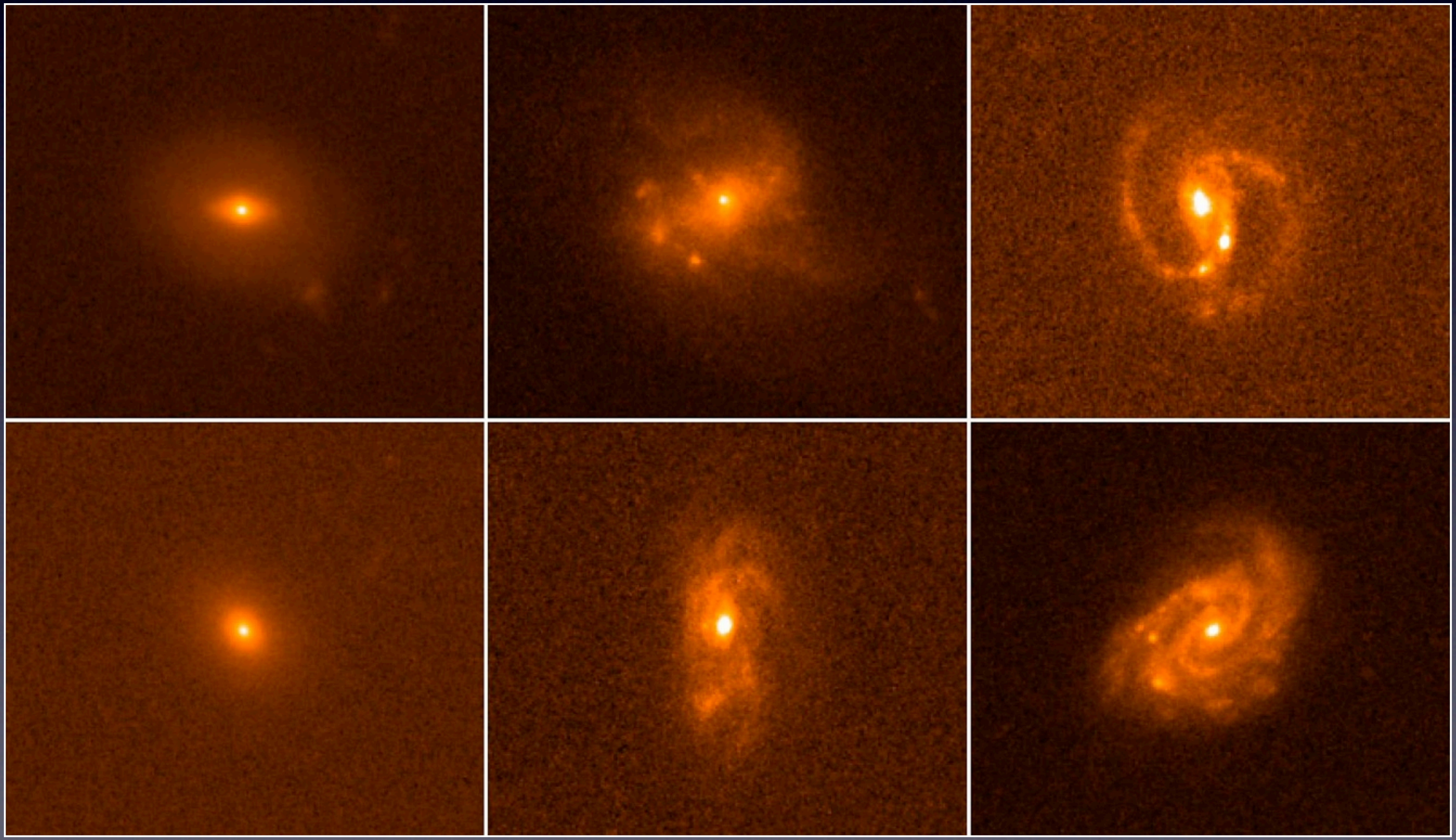
Non-parametric approach: Clumpiness

$$S_0 = \frac{1}{2}\frac{\sum'_{ij} I(i,j) - I_S(i,j)}{\sum_{ij} |I(i,j)|}, \qquad S = S_0 - S_{\rm bkg}$$



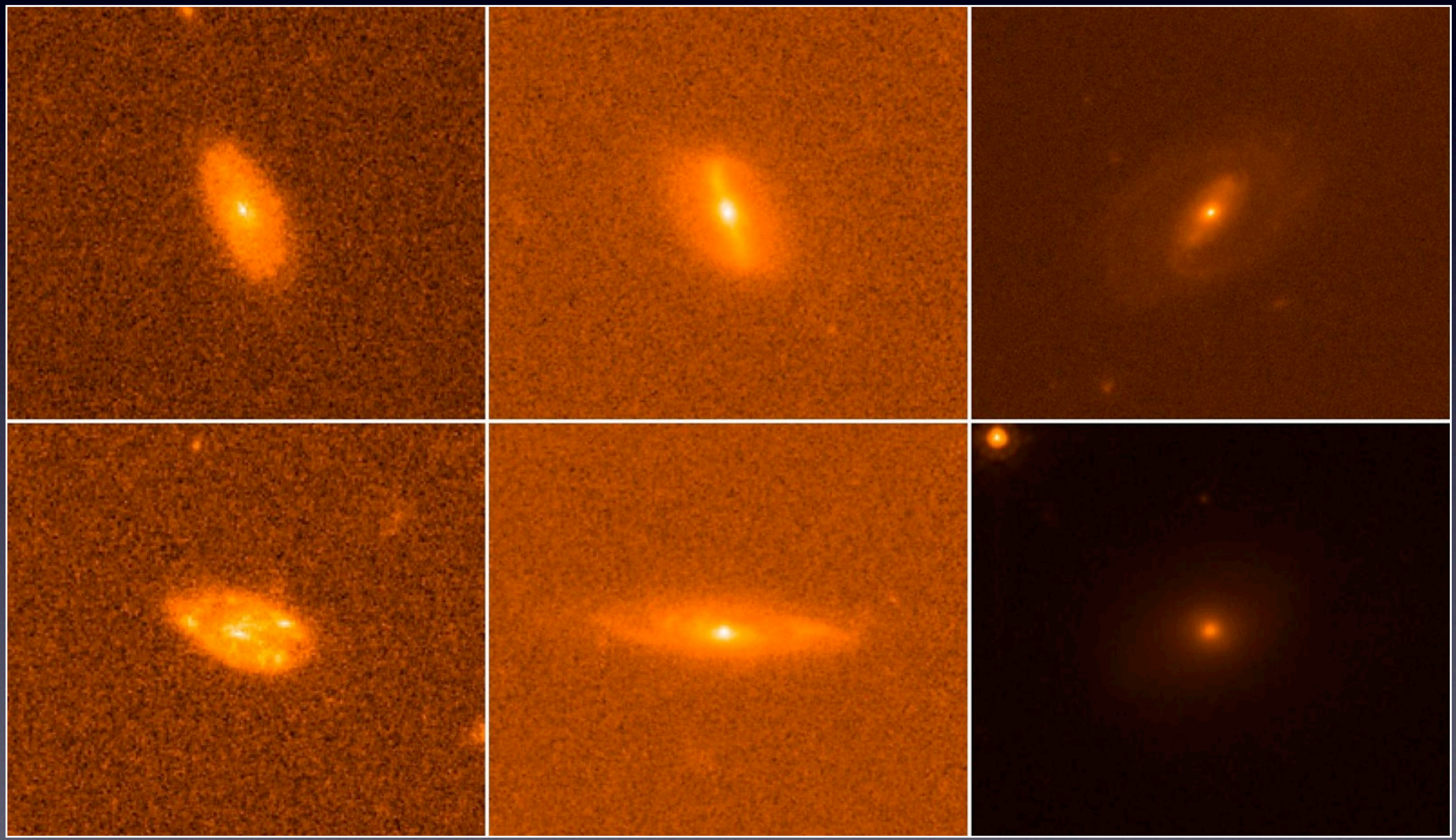
Non-parametric approach: Clumpiness

$$S_0 = \frac{1}{2} \frac{\sum'_{ij} I(i,j) - I_S(i,j)}{\sum_{ij} |I(i,j)|}, \qquad S = S_0 - S_{\rm bkg}$$



Non-parametric approach: Gini coefficient

$$G = \frac{1}{2\bar{I}\,N\,(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} |I_i - I_j|$$



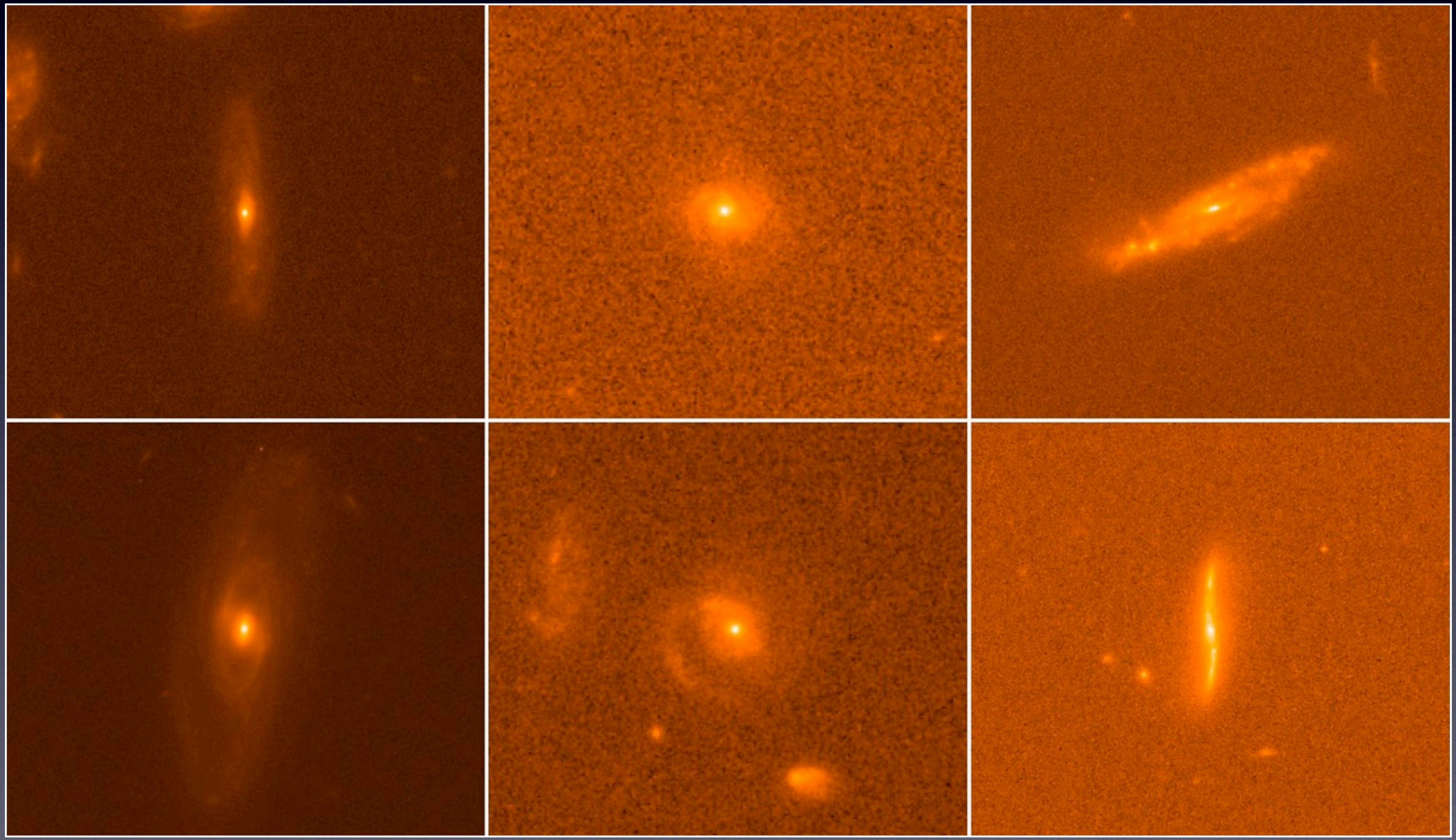
Non-parametric approach: M_{20}

$$x_c = \frac{\sum_{i\in\mathcal{E}} x_i I_i}{\sum_{i\in\mathcal{E}} I_i}$$

$$y_c = \frac{\sum_{i\in\mathcal{E}} y_i I_i}{\sum_{i\in\mathcal{E}} I_i}$$

$$M_i = I_i[(x_i - x_c)^2 + (y_i - y_c)^2]$$

$$M_{\rm tot} = \sum_{i\in\mathcal{E}} M_i$$

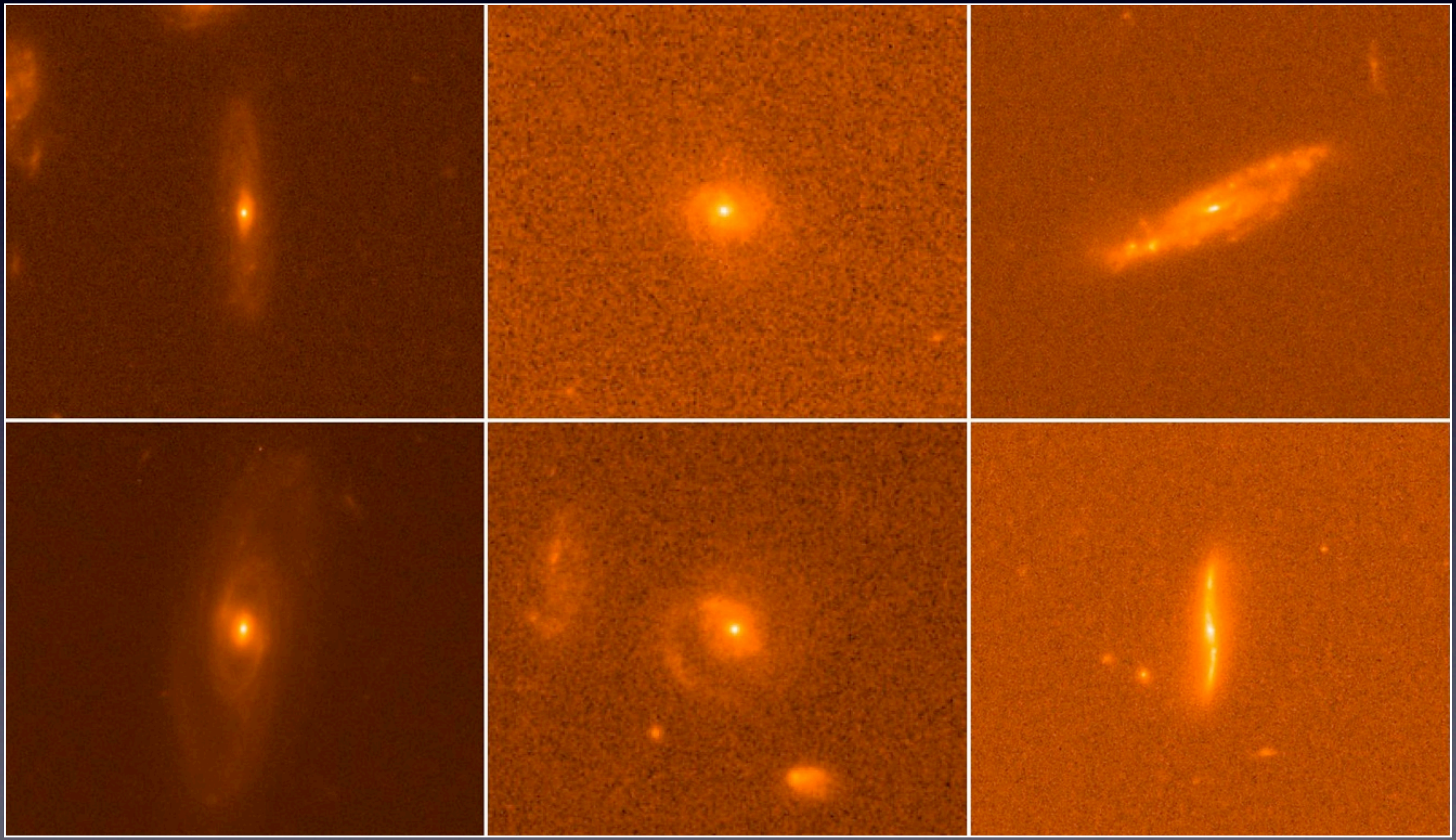


Non-parametric approach: M_{20}

$$x_c = \frac{\sum_{i\in\mathcal{E}} x_i I_i}{\sum_{i\in\mathcal{E}} I_i}$$

$$y_c = \frac{\sum_{i\in\mathcal{E}} y_i I_i}{\sum_{i\in\mathcal{E}} I_i}$$

$$M_{20} = \log\left(\frac{\sum_{i\in\mathcal{E}_{20}} M_i}{M_{\rm tot}}\right)$$



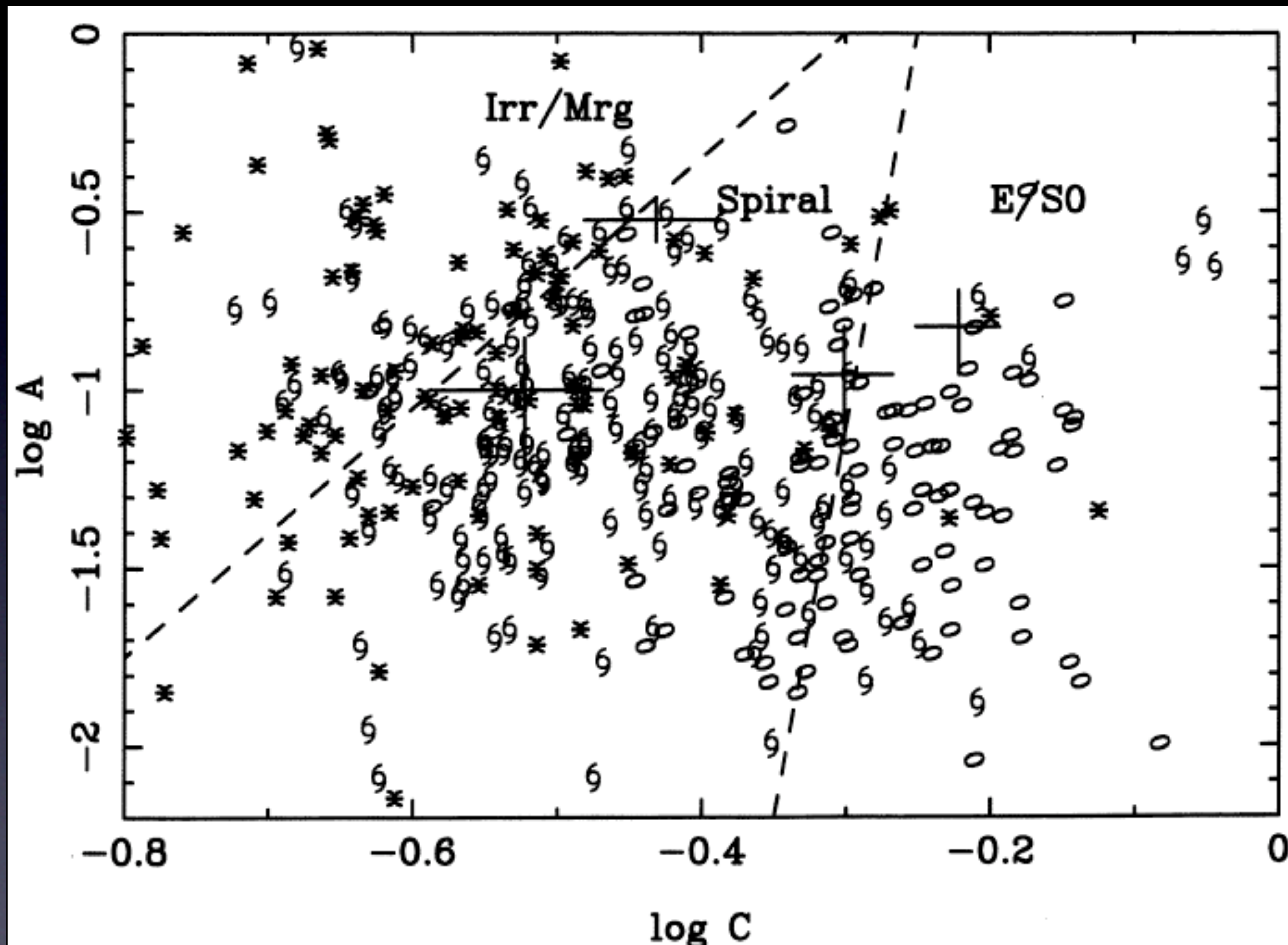
Non-parametric approach: Towards automatic classification

R. Abraham et al. (1996):	C-A plane separation
C. J. Concelice (2003):	C-A-S volume separation
J. Lotz et al. (2004):	C-A and G-M_{20} separation

Can these results be generalized?

- Parameter space dimensionality.
- Automatic, non-linear, region detection.
- Number of classes.
- Prediction.

Non-parametric approach: Towards automatic classification



Non-parametric approach: Towards automatic classification

R. Abraham et al. (1996):	C-A plane separation
C. J. Concelice (2003):	C-A-S volume separation
J. Lotz et al. (2004):	C-A and G-M_{20} separation

Can these results be generalized?

- Parameter space dimensionality.
- Automatic, non-linear, region detection.
- Number of classes.
- Prediction.

Pattern Recognition: Supervised Learning

Algorithms:

- Principal component analysis
- Artificial neural networks
- κ-nearest neighbours
- Support Vector Machines

Toy example:

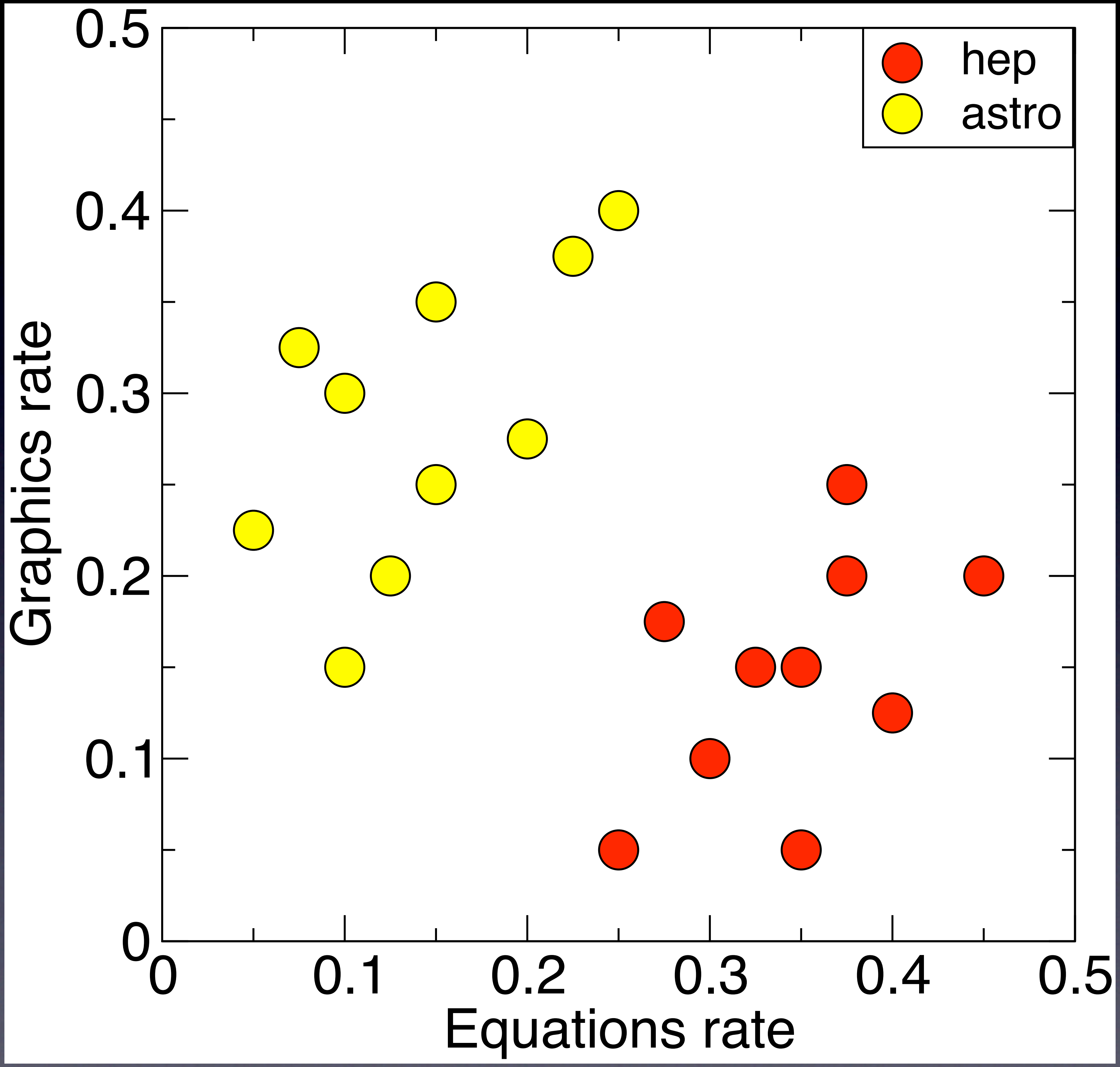
Label / Target		Patterns
Astronomer:	+1	Equations & Graphics rate
Particle Physicist:	-1	

Data = Pattern + Label

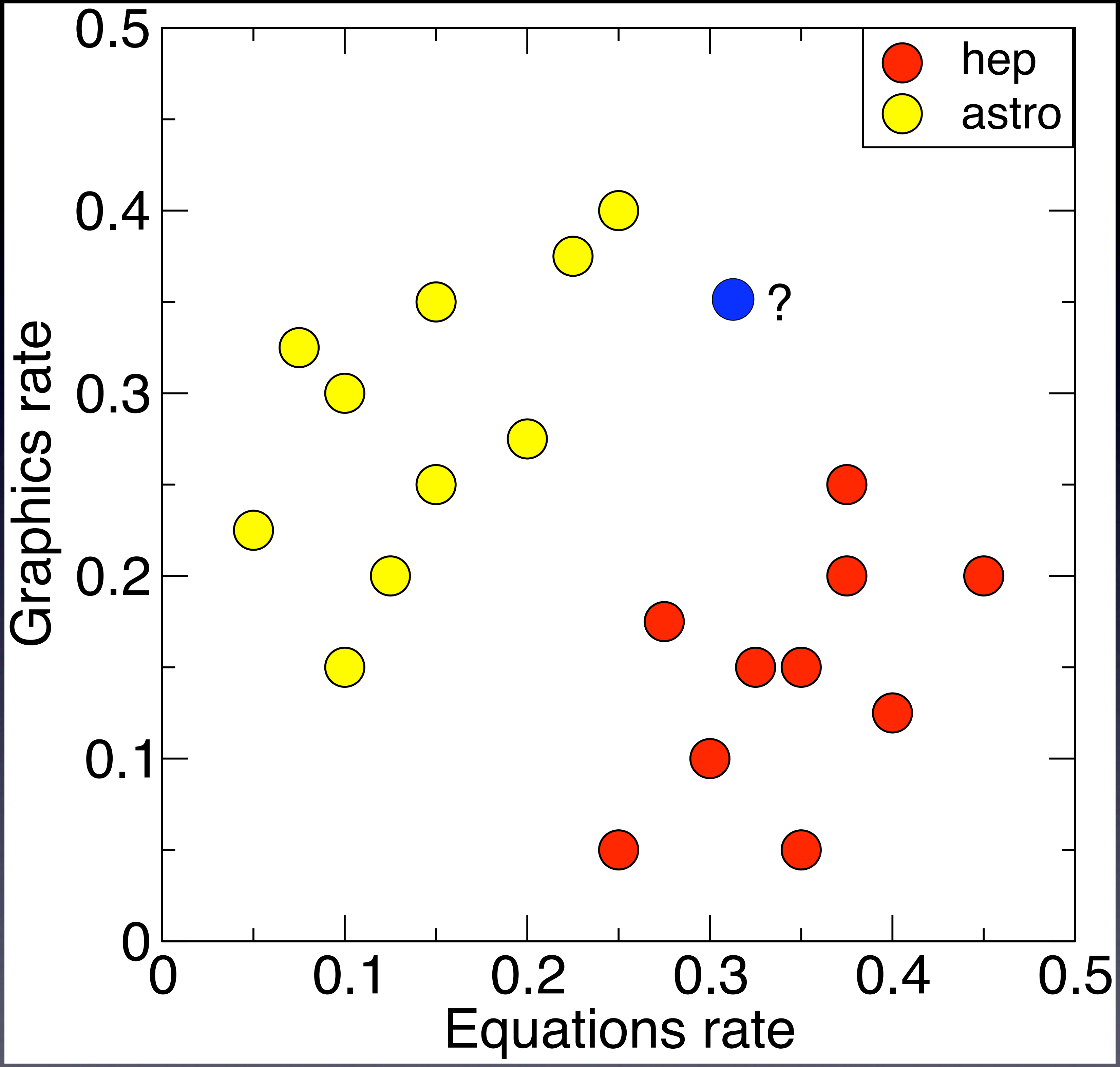
Pattern Recognition: Supervised Learning

Training sample		
Patterns (x_1 , x_2)		Label (y)
Equations	Images	
0.05	0.23	+1
0.08	0.33	+1
0.10	0.30	+1
0.13	0.20	+1
0.25	0.05	-1
0.23	0.18	-1
0.33	0.15	-1
0.33	0.05	-1
...	...	...

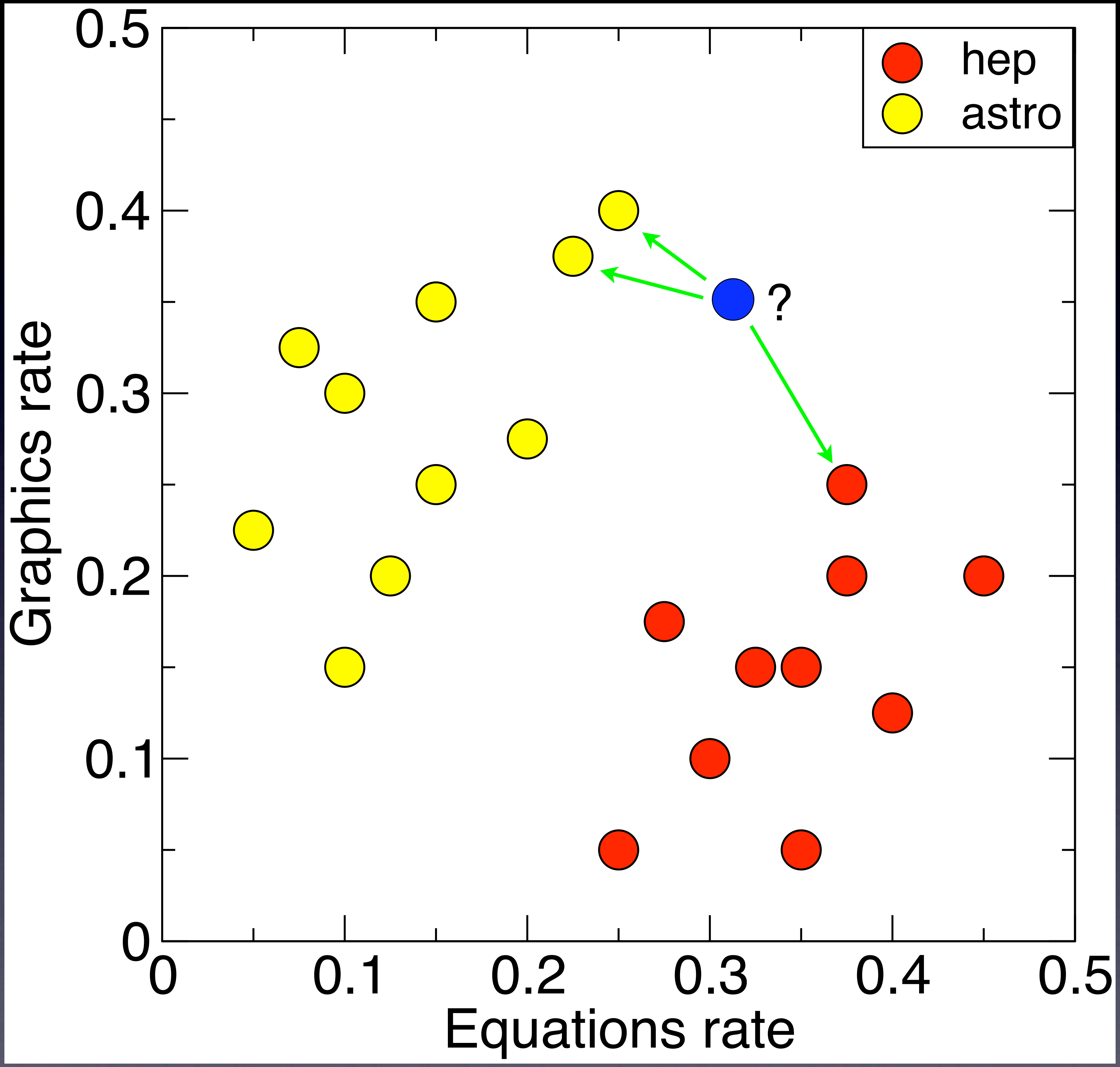
Pattern Recognition: Supervised Learning



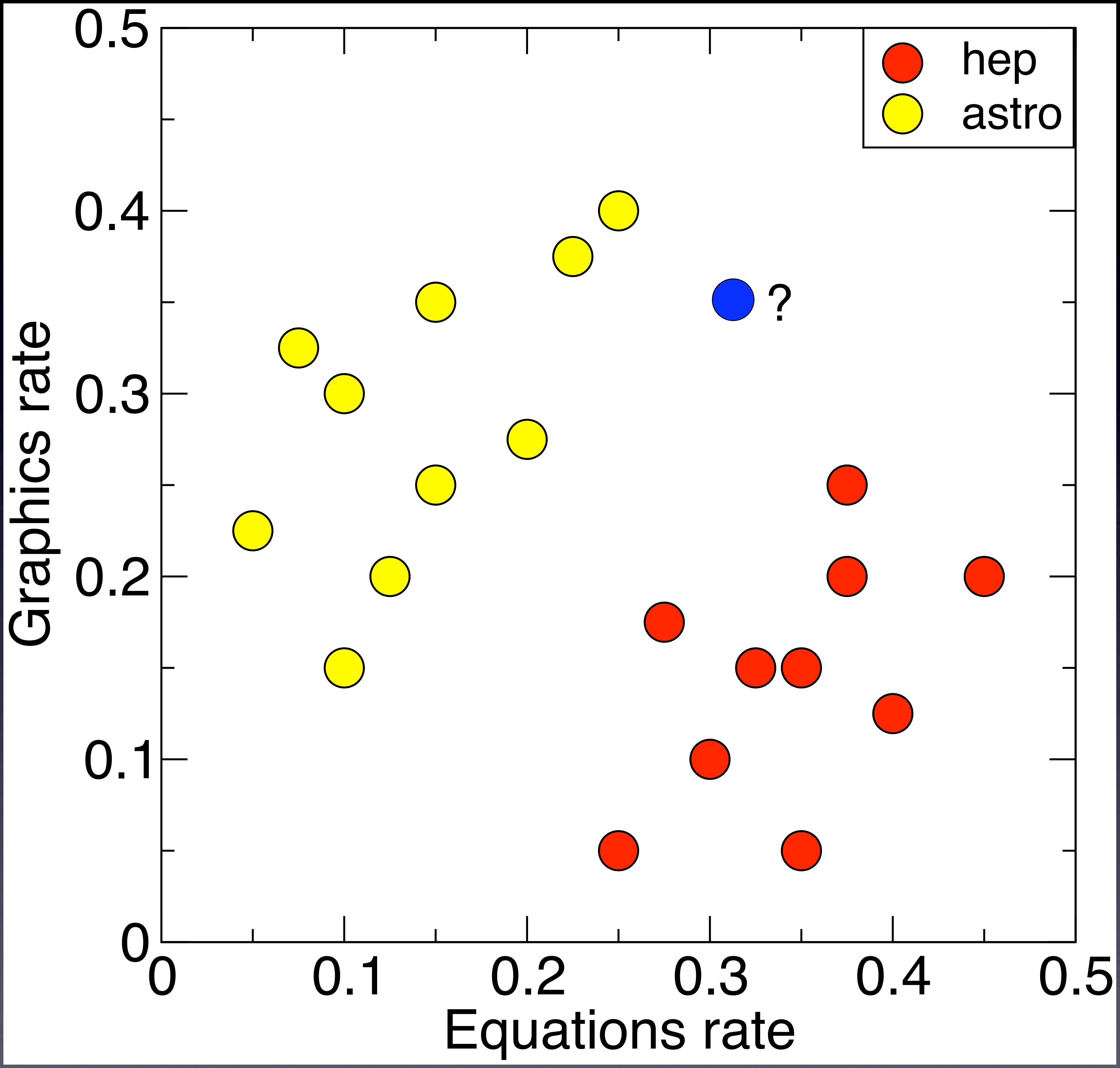
Pattern Recognition: Supervised Learning



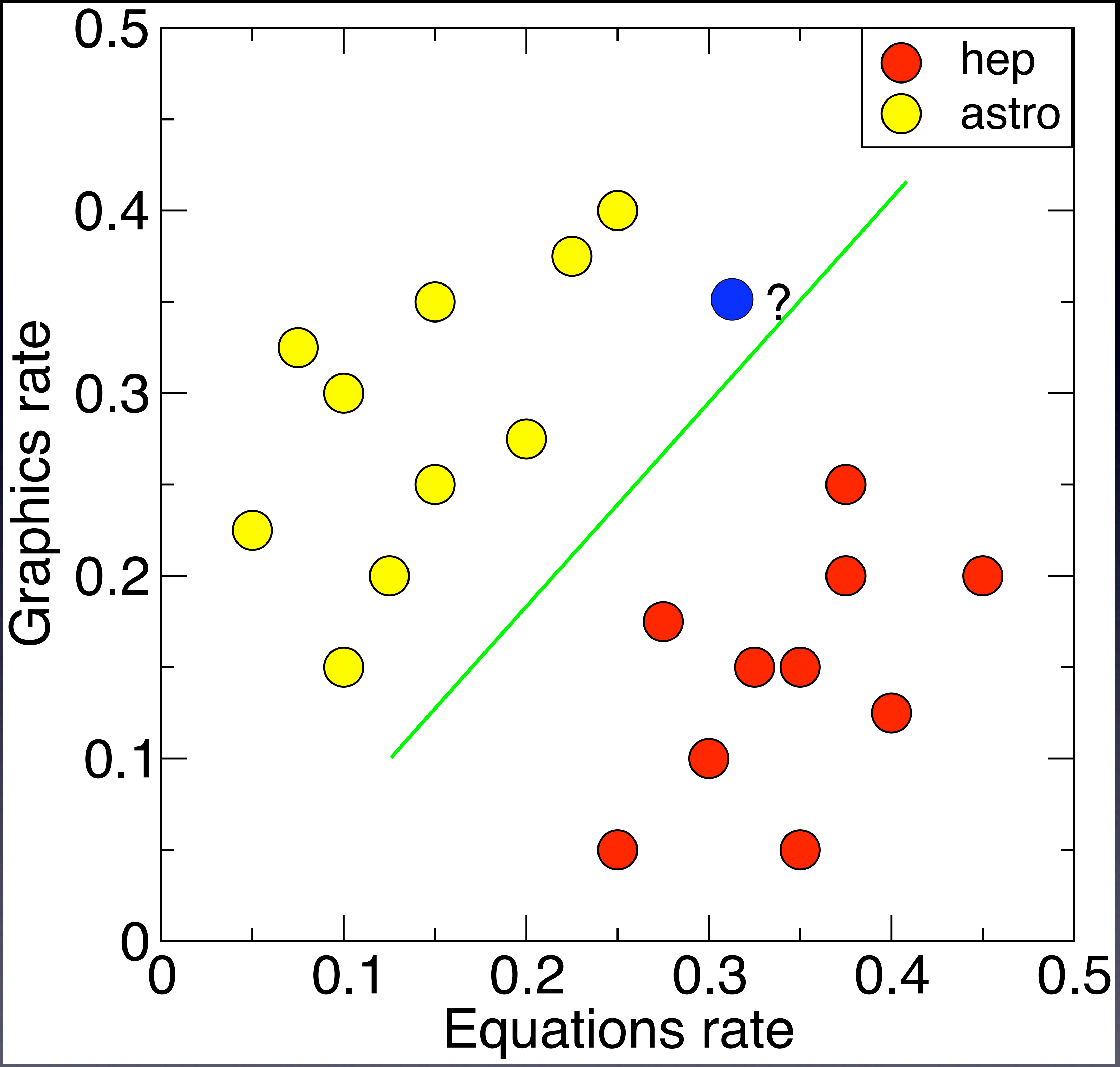
Pattern Recognition: Supervised Learning



Pattern Recognition: Supervised Learning



Pattern Recognition: Supervised Learning

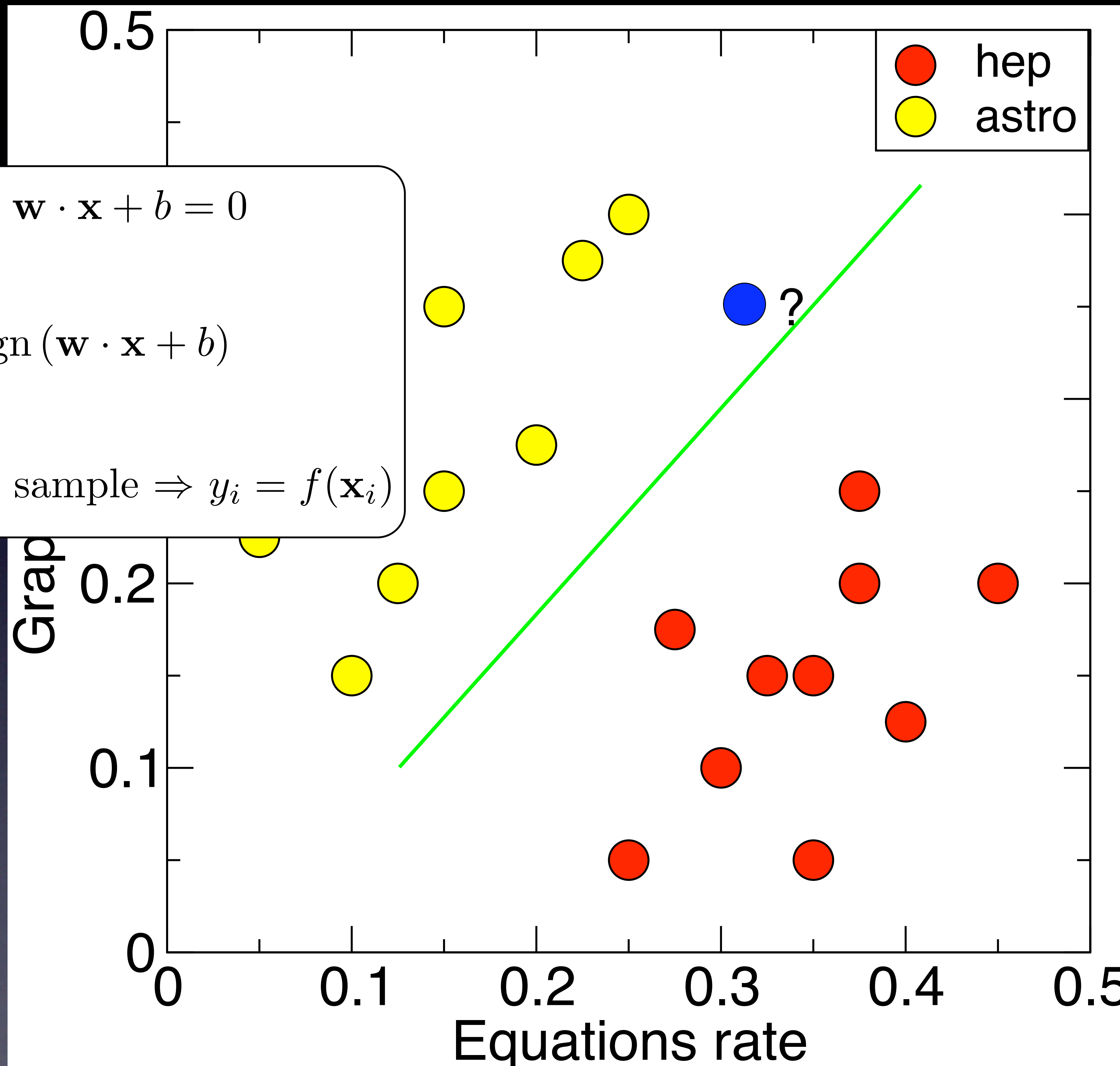


Pattern Recognition: Supervised Learning

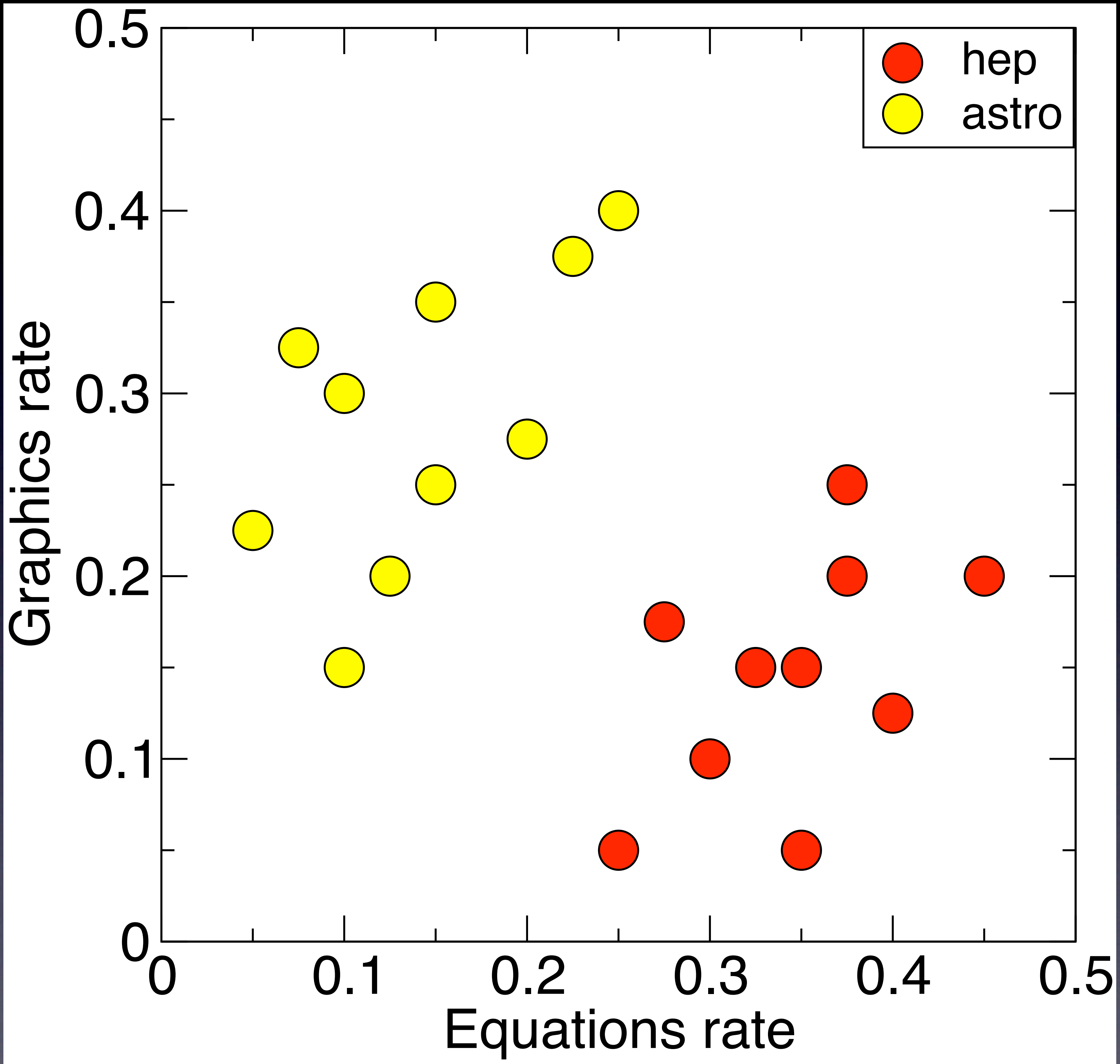
$$\mathbf{x} \in \Pi \;\Leftrightarrow\; \mathbf{w} \cdot \mathbf{x} + b = 0$$

$$f(\mathbf{x}) = \mathrm{sgn}\,(\mathbf{w} \cdot \mathbf{x} + b)$$

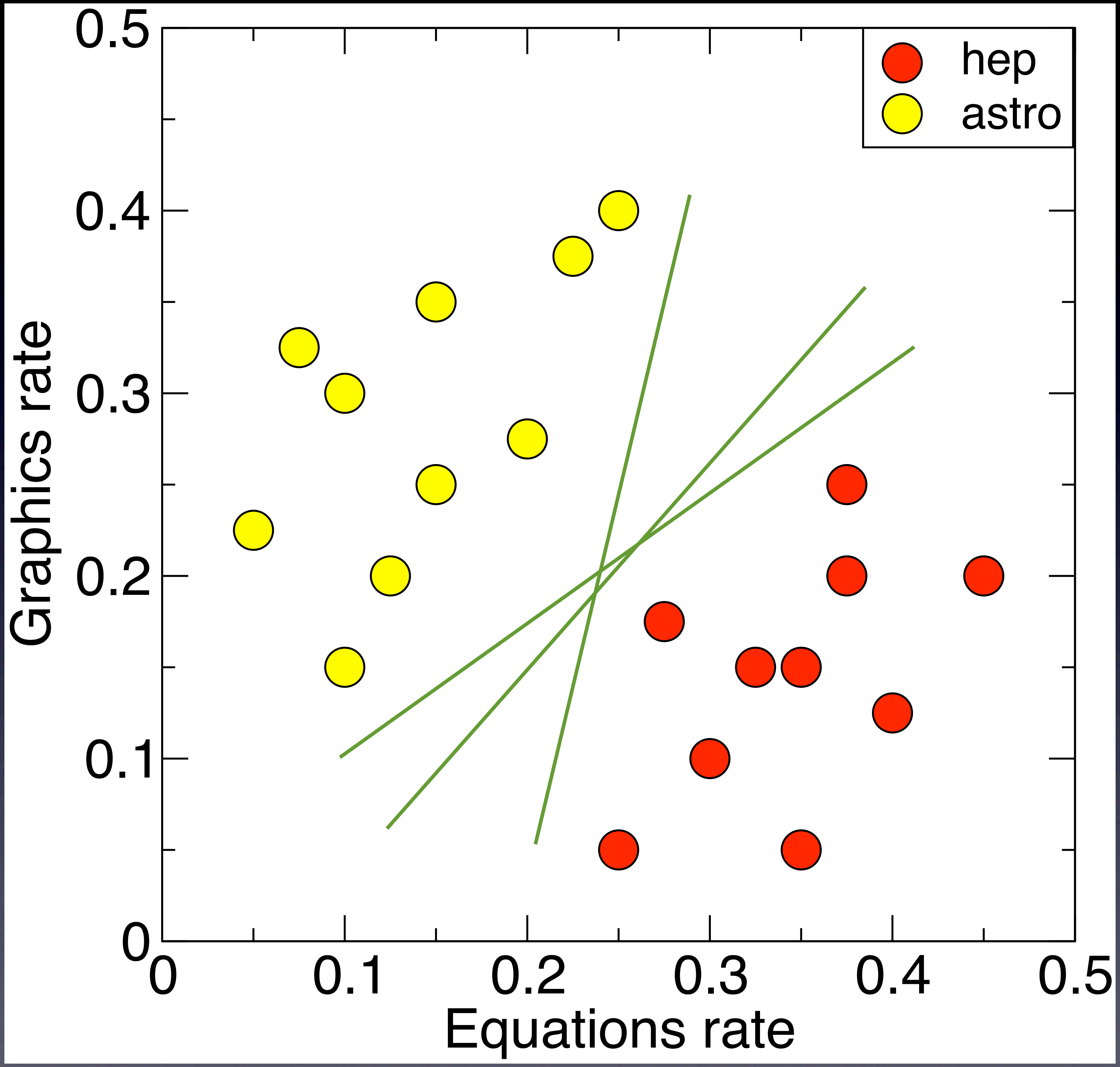
$$(\mathbf{x}_i, y_i) \text{ in training sample} \Rightarrow y_i = f(\mathbf{x}_i)$$



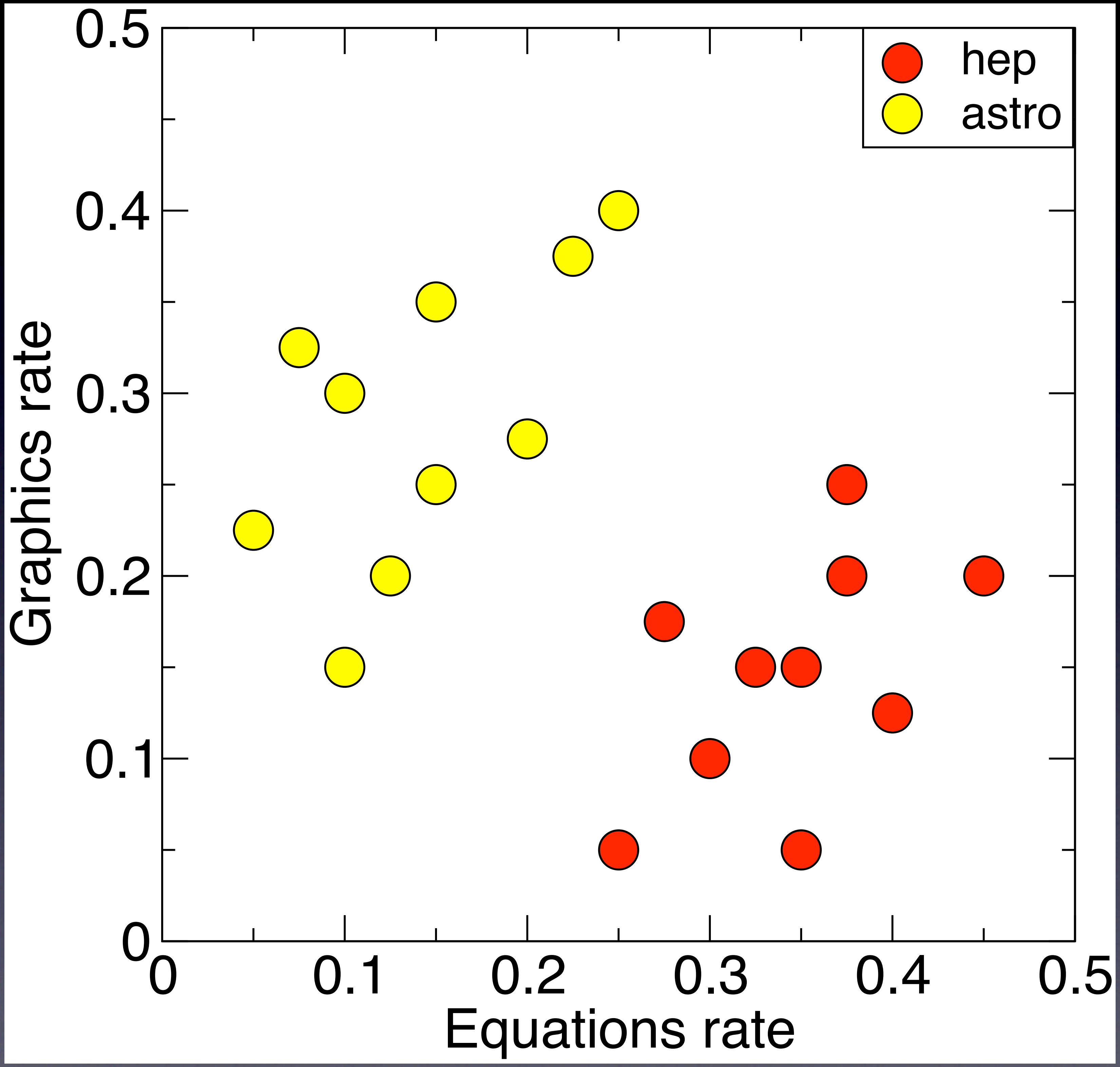
Pattern Recognition: Separating hyperplanes



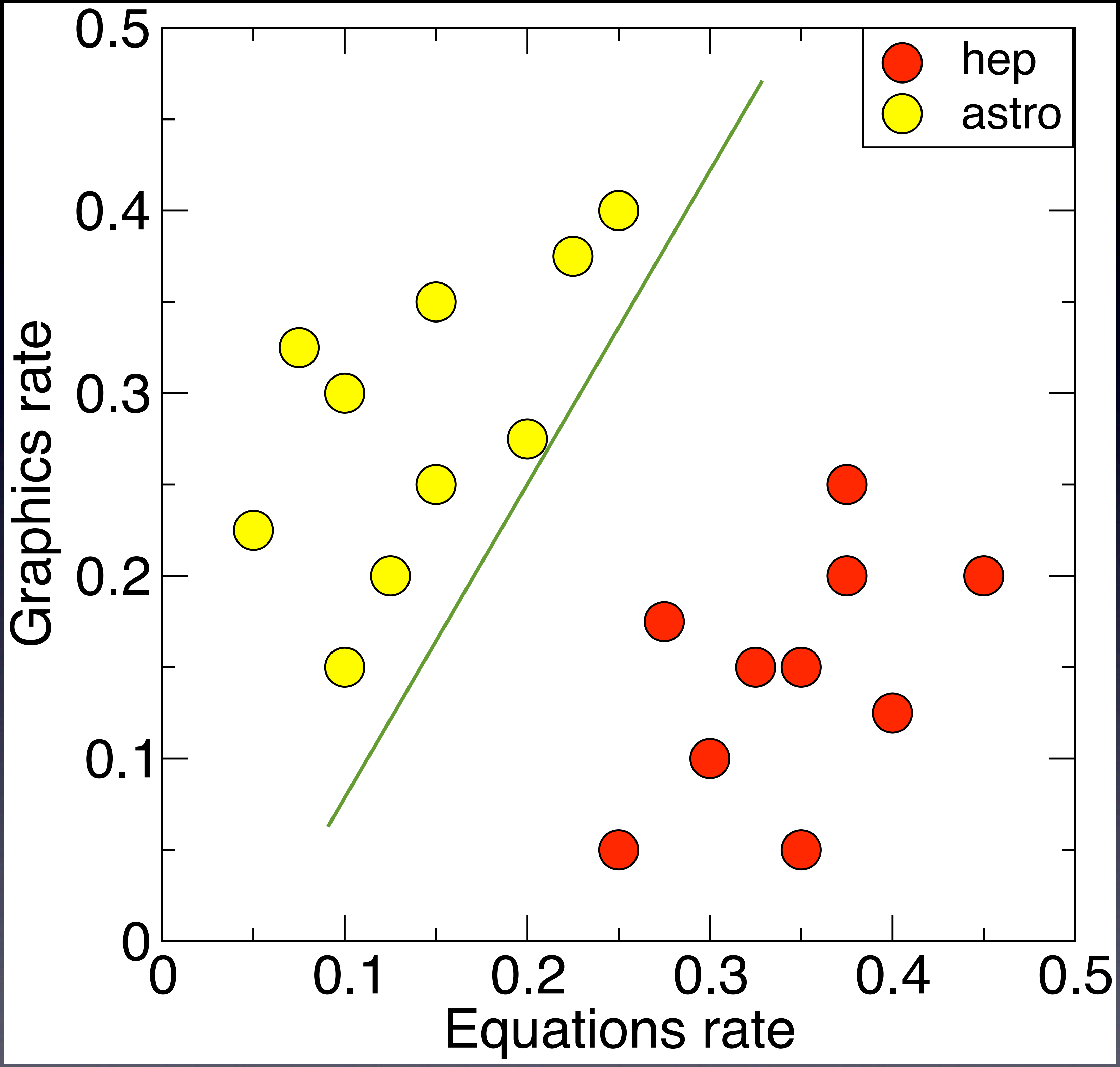
Pattern Recognition: Separating hyperplanes



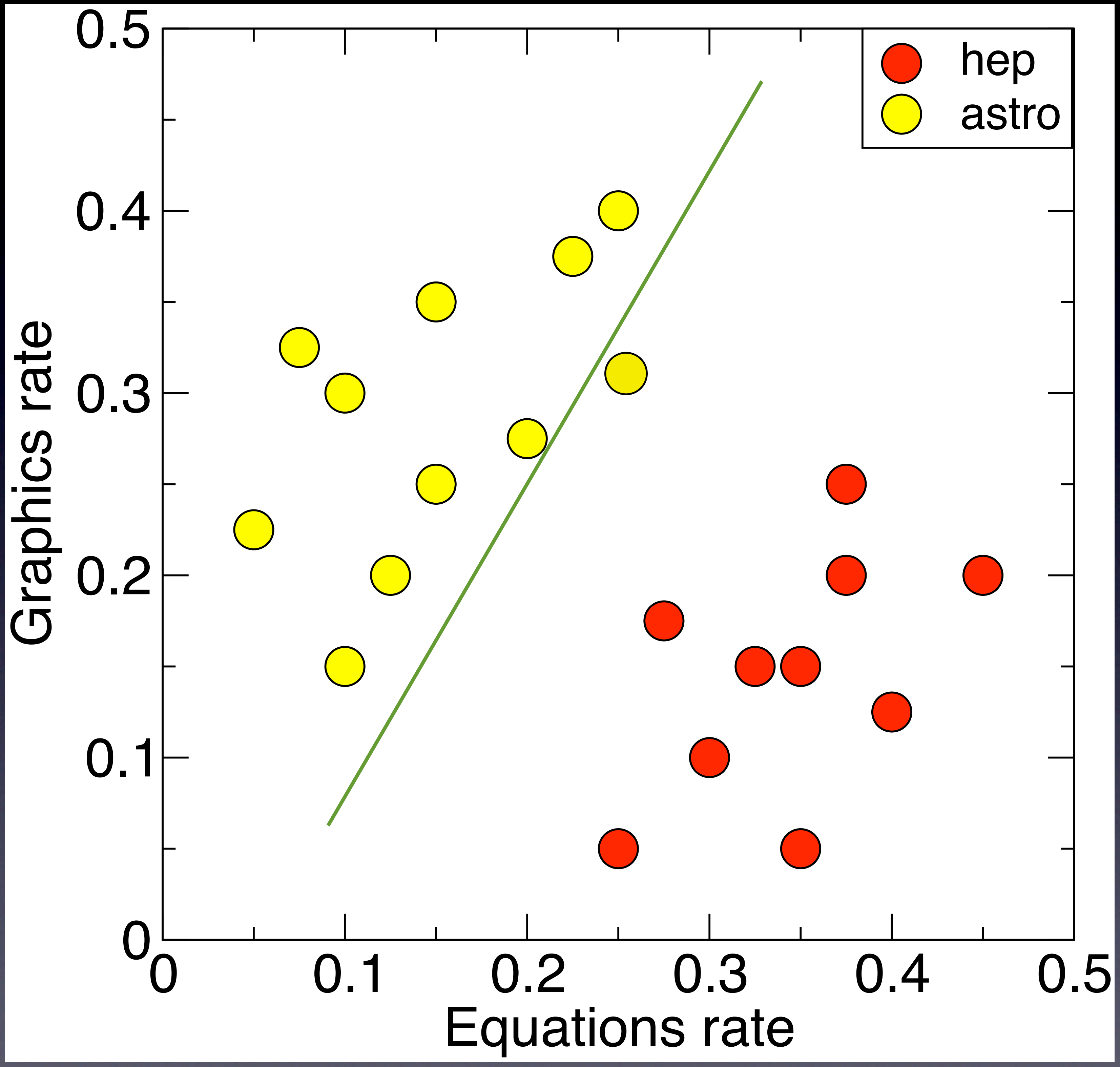
Pattern Recognition: Separating hyperplanes



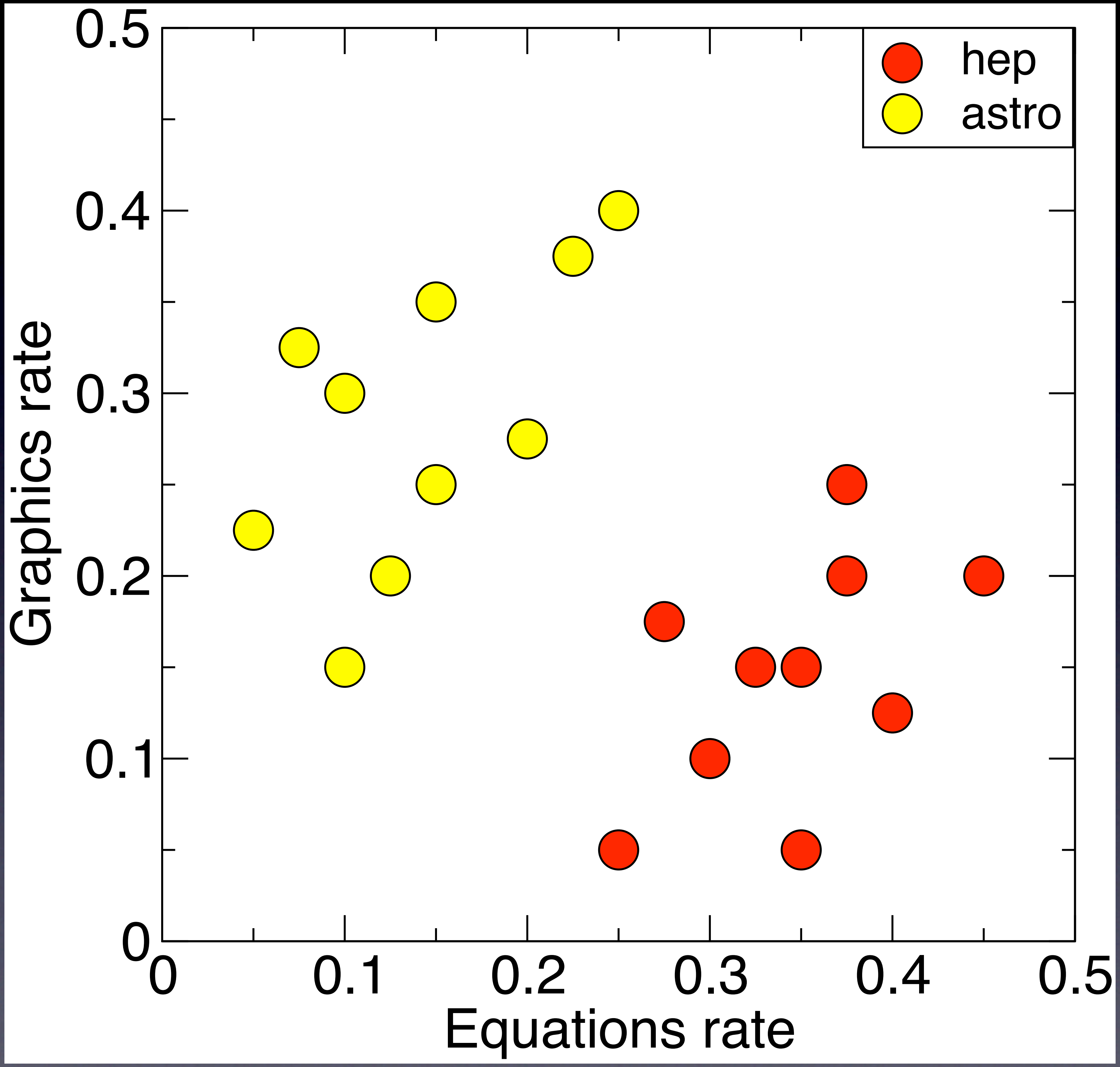
Pattern Recognition: Separating hyperplanes



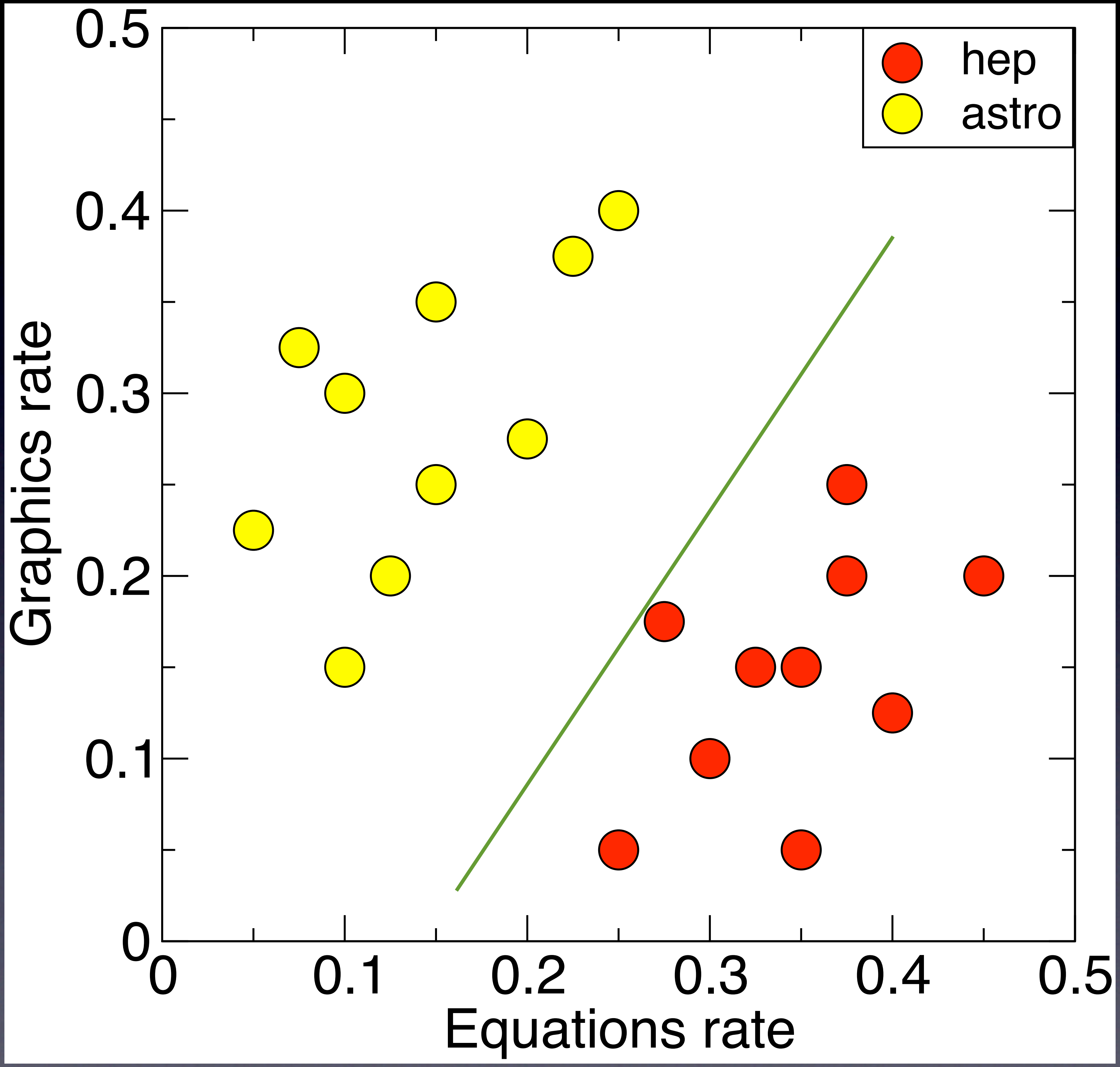
Pattern Recognition: Separating hyperplanes



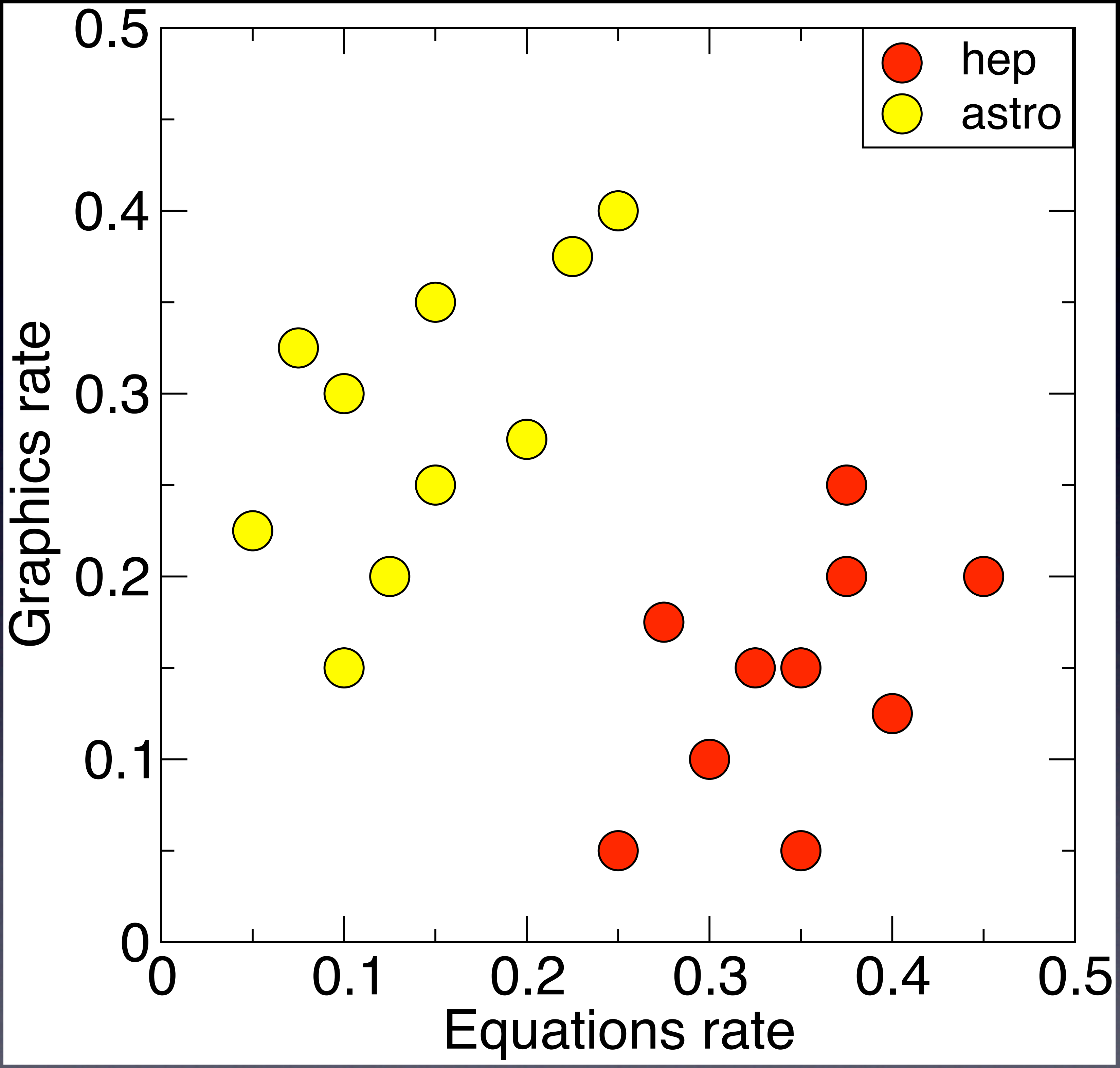
Pattern Recognition: Separating hyperplanes



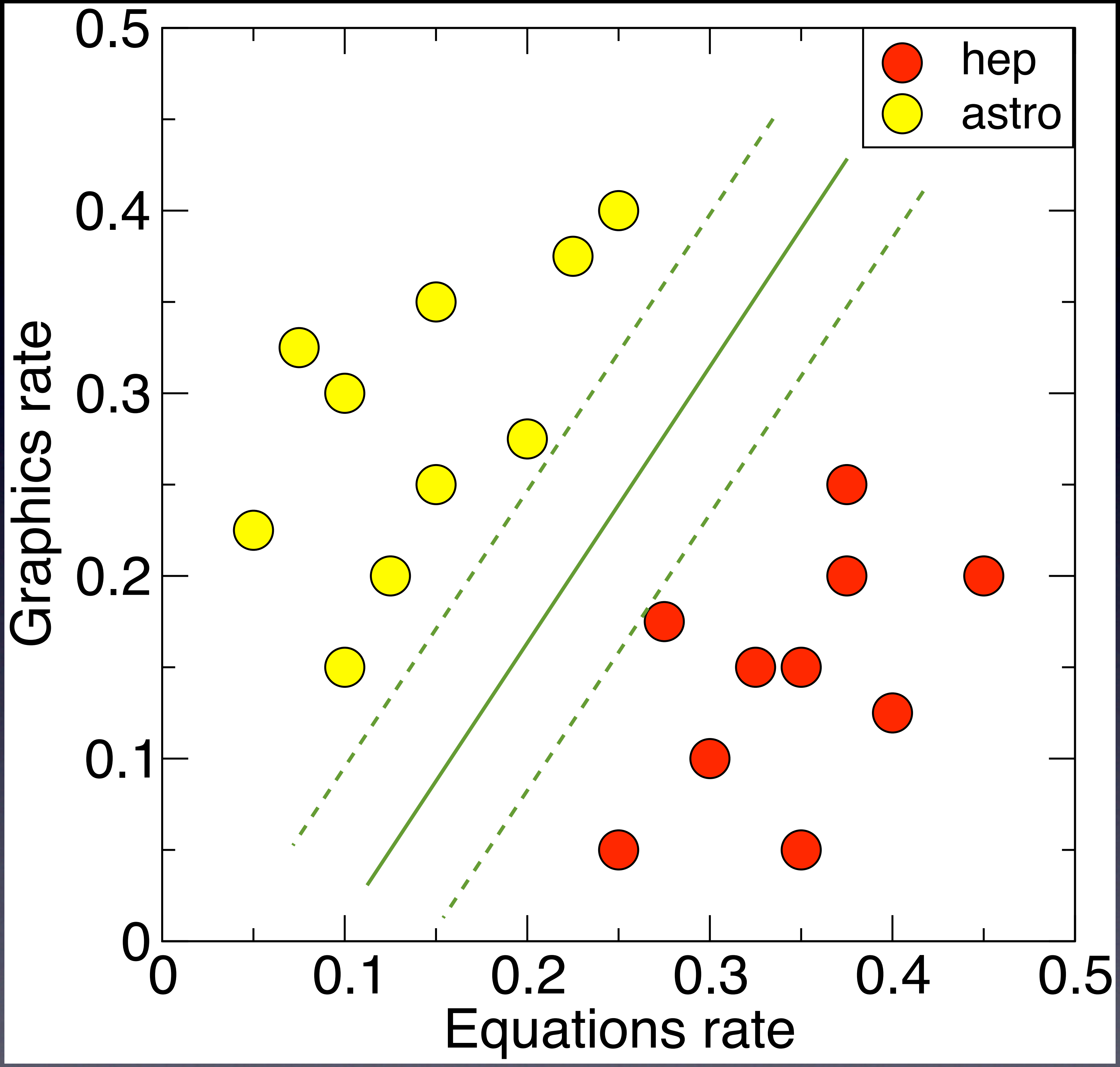
Pattern Recognition: Separating hyperplanes



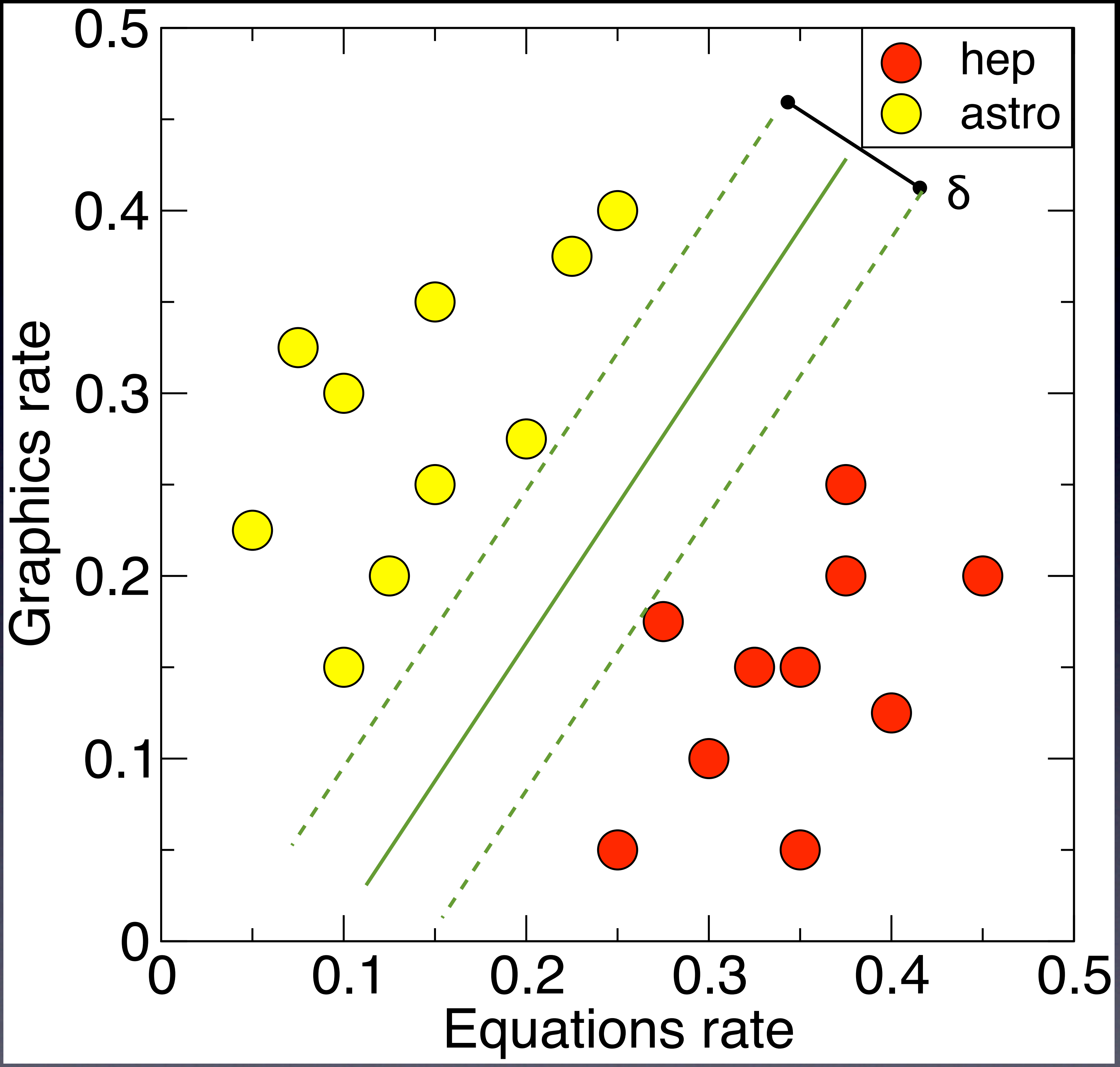
Pattern Recognition: Separating hyperplanes



Pattern Recognition: Separating hyperplanes



Pattern Recognition: Separating hyperplanes



Pattern Recognition: Separating hyperplanes

$$\mathbf{x} \in \Pi \;\Leftrightarrow\; \mathbf{w}\cdot\mathbf{x} + b = 0$$

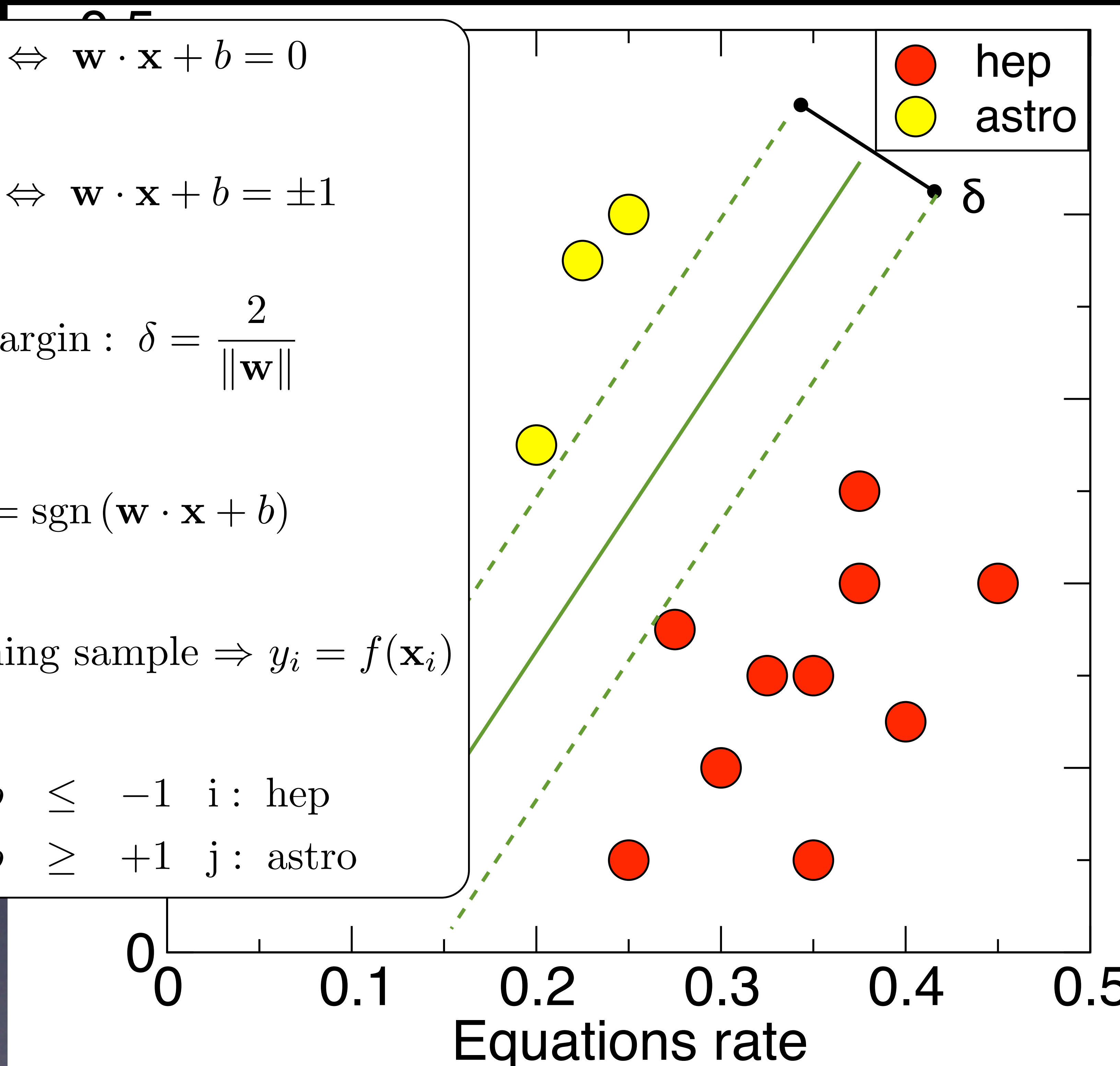
$$\mathbf{x} \in \Pi^{\pm} \;\Leftrightarrow\; \mathbf{w}\cdot\mathbf{x} + b = \pm 1$$

$$\Rightarrow \quad \text{Margin}: \; \delta = \frac{2}{\|\mathbf{w}\|}$$

$$f(\mathbf{x}) = \text{sgn}\left(\mathbf{w}\cdot\mathbf{x} + b\right)$$

$$(\mathbf{x}_i, y_i) \text{ in training sample} \Rightarrow y_i = f(\mathbf{x}_i)$$

$$\begin{array}{lclr} \mathbf{w}\cdot\mathbf{x}_i + b & \leq & -1 & \text{i}: \text{ hep} \\ \mathbf{w}\cdot\mathbf{x}_j + b & \geq & +1 & \text{j}: \text{ astro} \end{array}$$



Optimal hyperplane: Formal approach

$$(\mathbf{x_1}, y_1), \ldots, (\mathbf{x}_m, y_m) \in \mathcal{X} \times \{\pm 1\} \quad \text{with} \quad \mathcal{X} \subset \Re^N$$

$$(x_i, y_i) \longleftrightarrow P(x, y)$$

$$f : \mathcal{X} \to \{\pm 1\} \quad : \quad f(\mathbf{x}_i) = y_i$$

$$P(x, y) \longrightarrow (\mathbf{x}, y) \in \mathcal{X} \quad \Rightarrow \quad f(\mathbf{x}) = y$$

$$\mathbf{x} \in \Pi \;\Leftrightarrow\; \mathbf{w} \cdot \mathbf{x} + b = 0$$

$$\mathbf{x} \in \Pi^{\pm} \;\Leftrightarrow\; \mathbf{w} \cdot \mathbf{x} + b = \pm 1$$

$$\Rightarrow \quad \text{Margin}: \; \delta = \frac{2}{\|\mathbf{w}\|}$$

$$f(\mathbf{x}) = \text{sgn}\,(\mathbf{w} \cdot \mathbf{x} + b)$$

$$(\mathbf{x}_i, y_i) \text{ in training sample} \Rightarrow y_i = f(\mathbf{x}_i)$$

$$\begin{aligned} \mathbf{w} \cdot \mathbf{x}_i + b &\leq -1 \quad \text{i}: \text{ hep} \\ \mathbf{w} \cdot \mathbf{x}_j + b &\geq +1 \quad \text{j}: \text{ astro} \end{aligned}$$

Optimal hyperplane: Formal approach

$$\text{minimize} \qquad \tau(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2$$

$$\text{s.t.} \qquad y_i\left((\mathbf{w}\cdot\mathbf{x}_i) + b\right) \geq 1, \quad i = 1, \ldots, m$$

$$\alpha_i \geq 0 \quad i = 1, \ldots, m$$

$$\mathcal{L}(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{m} \alpha_i (y_i\left((\mathbf{w}\cdot\mathbf{x}_i) + b\right) - 1)$$

Karush-Kuhn-Tucker

$$\sum_{i=1}^{m} \alpha_i\, y_i = 0,$$

$$\mathbf{w} = \sum_{i=1}^{m} \alpha_i\, y_i\, \mathbf{x}_i$$

$$0 \leq \alpha_i$$

$$0 \leq (y_i\left(\mathbf{w}\cdot\mathbf{x}_i + b\right) - 1)$$

$$0 = \alpha_i\left(y_i\left(\mathbf{w}\cdot\mathbf{x}_i + b\right) - 1\right)$$

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^{m} y_i\, \alpha_i (\mathbf{x}\cdot\mathbf{x}_i) + b\right)$$

Optimal hyperplane: Formal approach

Wolfe dual problem

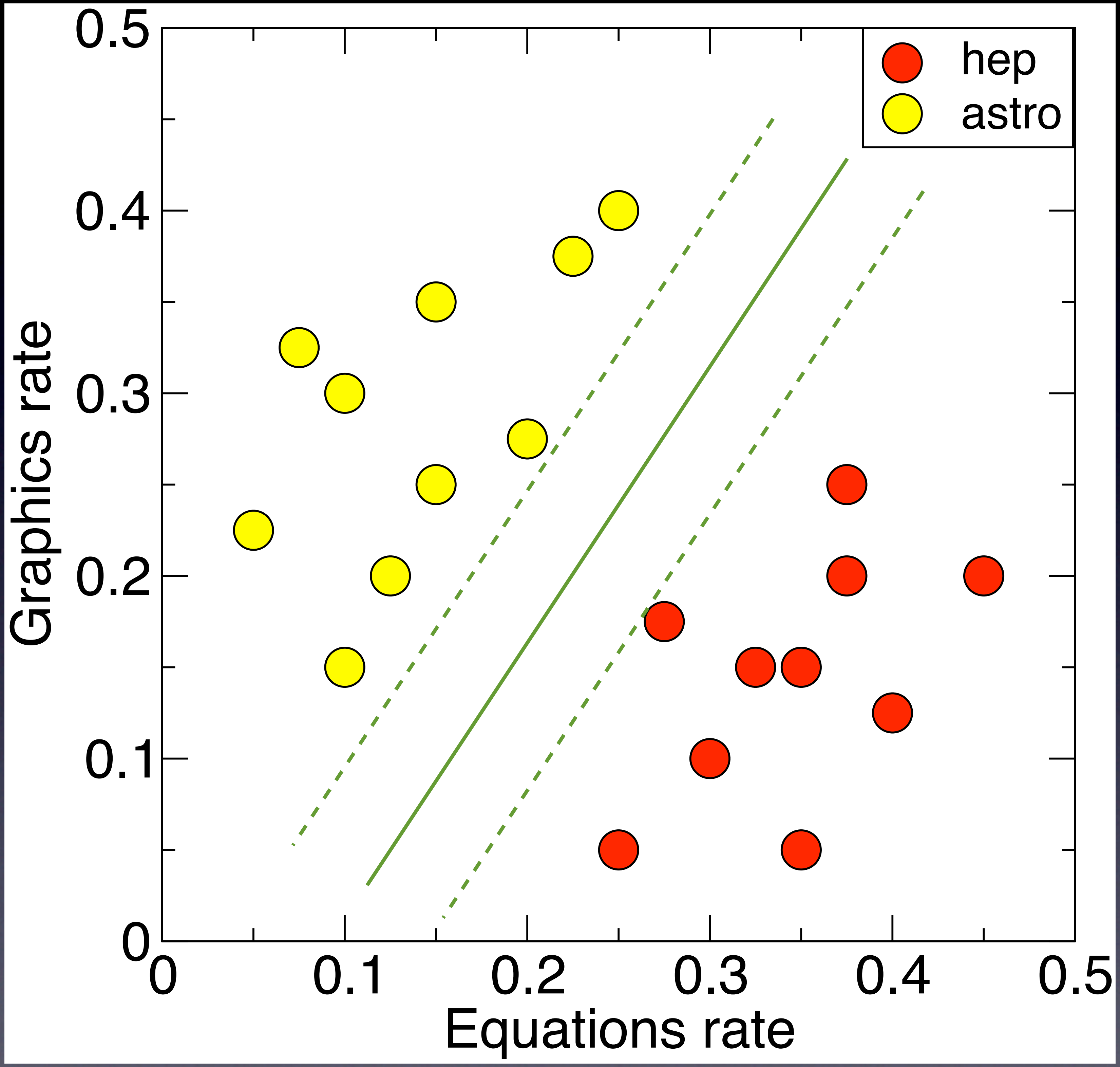
$$\text{maximize} \qquad W(\boldsymbol{\alpha}) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \, \alpha_j \, y_i \, y_j \, (\mathbf{x}_i \cdot \mathbf{x}_j)$$

$$\text{s.t.} \qquad \alpha_i \geq 0 \quad i = 1, \ldots, m \quad \text{and} \quad \sum_{i=1}^{m} \alpha_i y_i = 0$$

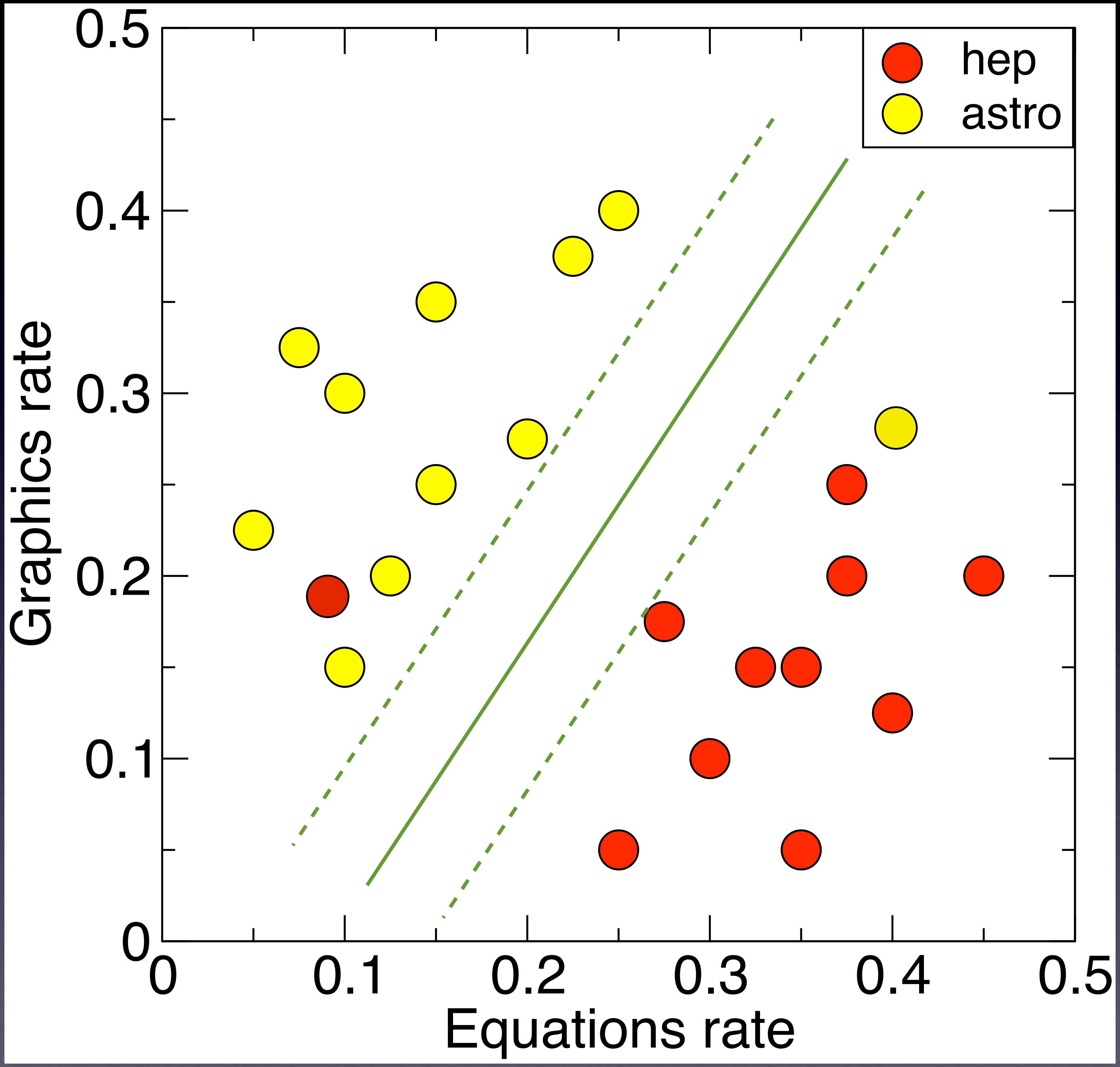
$$f(\mathbf{x}) \quad = \quad \text{sgn}\left(\sum_{i=1}^{m} y_i \, \alpha_i (\mathbf{x} \cdot \mathbf{x}_i) + b \right)$$

$$b \quad = \quad 1 - \sum_{i=1}^{m} \alpha_i \, y_i \, (\mathbf{x}_i \cdot \mathbf{x}_j) \quad \text{with} \quad \alpha_j \neq 0$$

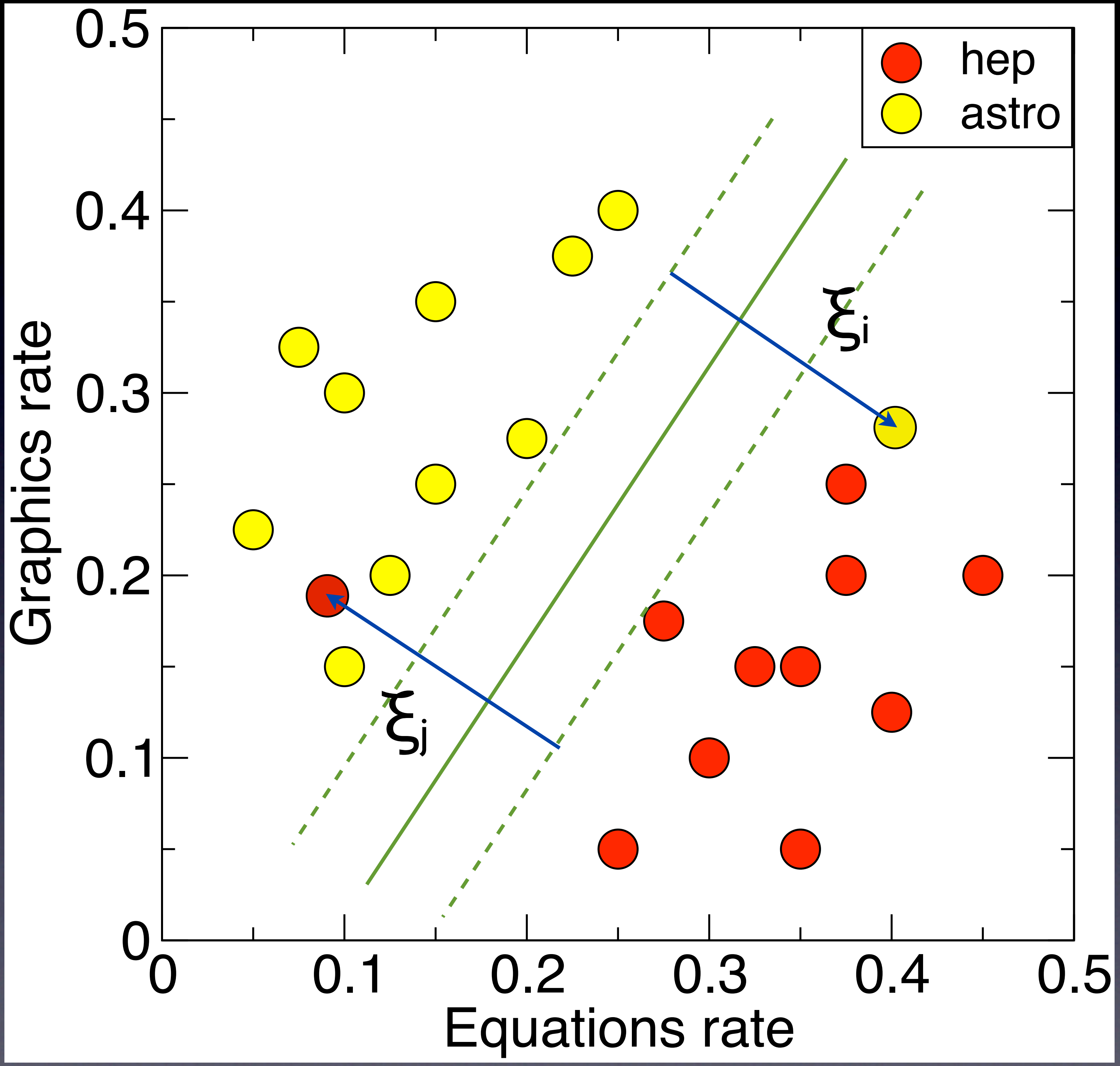
Optimal hyperplane: Non-linear separable data



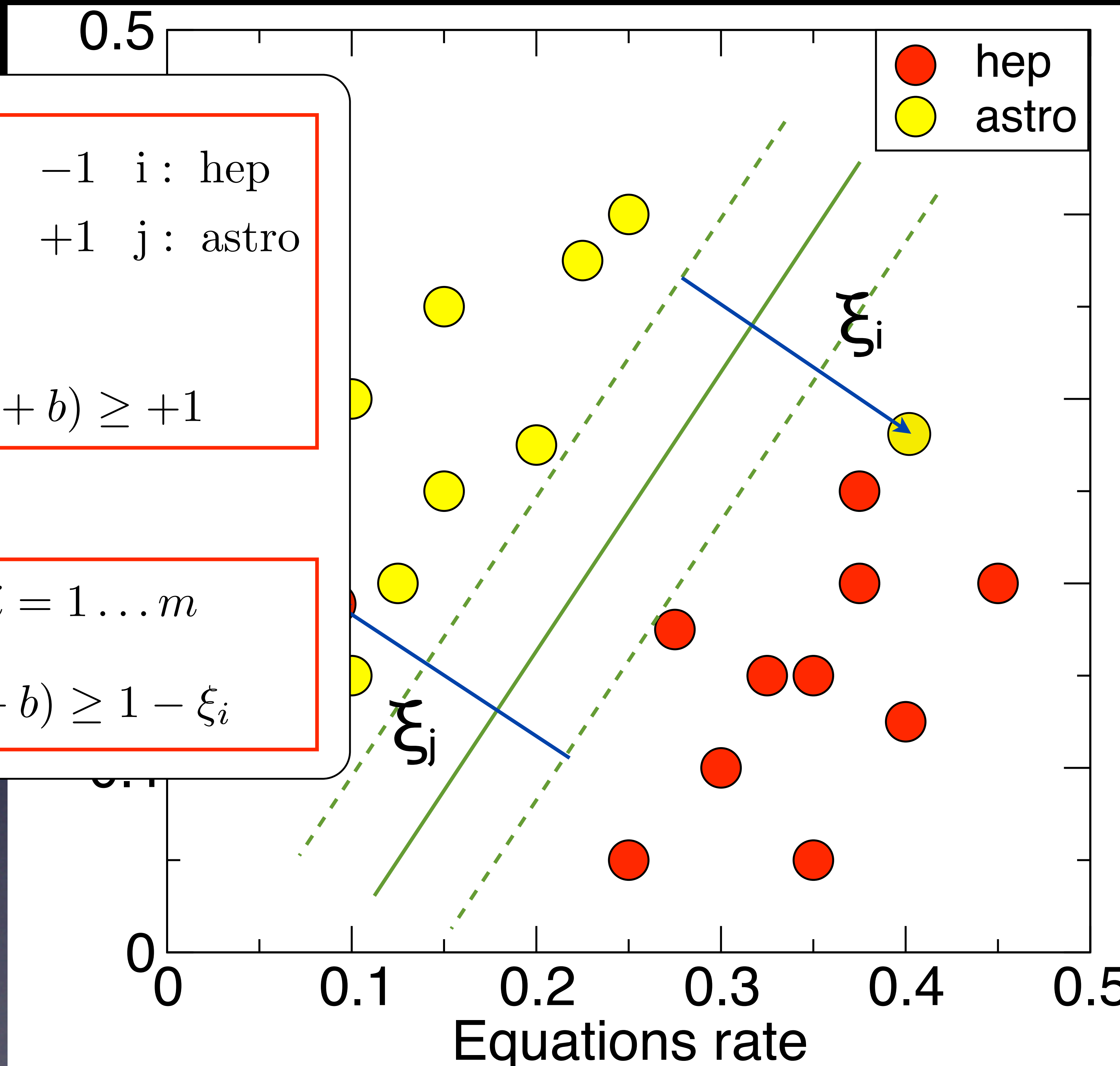
Optimal hyperplane: Non-linear separable data



Optimal hyperplane: Non-linear separable data



Optimal hyperplane: Non-linear separable data



$$
\begin{aligned}
\mathbf{w}\cdot\mathbf{x}_i + b &\leq -1 \quad \text{i : hep} \\
\mathbf{w}\cdot\mathbf{x}_j + b &\geq +1 \quad \text{j : astro}
\end{aligned}
$$

$$y_i\left(\mathbf{w}\cdot\mathbf{x}_i + b\right) \geq +1$$

$$\xi_i \geq 0 \quad i = 1\ldots m$$

$$y_i\left(\mathbf{w}\cdot\mathbf{x}_i + b\right) \geq 1 - \xi_i$$

Optimal hyperplane: Non-linear separable data

$$\text{minimize} \qquad \tau(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{m}\xi_i$$

$$\text{s.t.} \qquad \xi \geq 0 \quad \text{and} \quad y_i\left(\mathbf{w}\cdot\mathbf{x}_i + b\right) \geq 1 - \xi_i, \quad i = 1,\ldots,m$$

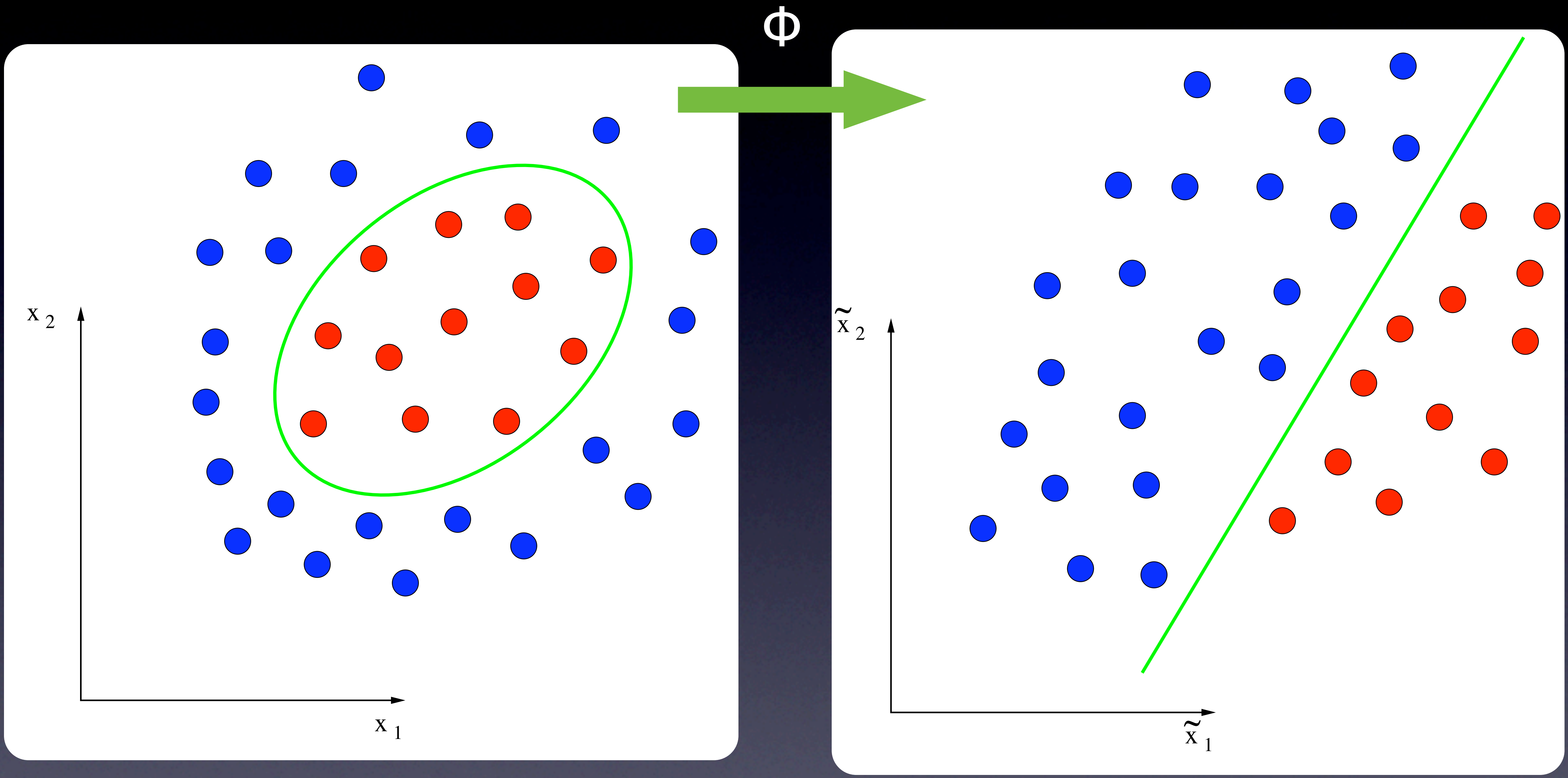
$$\text{maximize} \qquad W(\boldsymbol{\alpha}) = \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{m}\alpha_i\,\alpha_j\,y_i\,y_j\,(\mathbf{x}_i\cdot\mathbf{x}_j)$$

$$\text{s.t.} \qquad 0 \leq \alpha_i \leq C \quad i = 1,\ldots,m \quad \text{and} \quad \sum_{i=1}^{m}\alpha_i y_i = 0$$

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^{m} y_i\,\alpha_i(\mathbf{x}\cdot\mathbf{x}_i) + b\right)$$

$$b = 1 - \sum_{i=1}^{m}\alpha_i\,y_i\,(\mathbf{x}_i\cdot\mathbf{x}_j) \quad \text{with} \quad 0 < \alpha_j < C \Rightarrow \xi_j = 0$$

Support vector machines: Mapping data



Support Vector Machines: Kernel methods

$$(\mathbf{x_1}, y_1), \ldots, (\mathbf{x}_m, y_m) \in \mathcal{X} \times \{\pm 1\} \quad \text{with} \quad \mathcal{X} \subset \Re^N$$

$$(x_i, y_i) \longleftrightarrow P(x, y)$$

$$f : \mathcal{X} \to \{\pm 1\} \quad : \quad f(\mathbf{x}_i) = y_i$$

$$P(x, y) \longrightarrow (\mathbf{x}, y) \in \mathcal{X} \quad \Rightarrow \quad f(\mathbf{x}) = y$$

$$\begin{aligned} k : \mathcal{X} \times \mathcal{X} &\longrightarrow \Re \\ (x, x') &\longrightarrow k(x, x') \end{aligned}$$

$$\text{If } \mathbf{x},\ \mathbf{x}' \in \Re^N \Rightarrow$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbf{x} \cdot \mathbf{x}'$$

$$\begin{aligned} \Phi : \mathcal{X} &\longrightarrow F \subset \Re^N \\ x &\longrightarrow \mathbf{x} \end{aligned}$$

$$k(x, x') \quad \equiv \quad \mathbf{x} \cdot \mathbf{x}' = \Phi(x) \cdot \Phi(x')$$

Support Vector Machines: Kernel methods

$$(\mathbf{x_1}, y_1), \ldots, (\mathbf{x}_m, y_m) \in \mathcal{X} \times \{\pm 1\} \quad \text{with} \quad \mathcal{X} \subset \Re^N$$

$$(x_i, y_i) \longleftrightarrow P(x, y)$$

$$f : \mathcal{X} \to \{\pm 1\} \quad : \quad f(\mathbf{x}_i) = y_i$$

$$P(x, y) \longrightarrow (\mathbf{x}, y) \in \mathcal{X} \quad \Rightarrow \quad f(\mathbf{x}) = y$$

$$\begin{aligned} k : \mathcal{X} \times \mathcal{X} &\longrightarrow \Re \\ (x, x') &\longrightarrow k(x, x') \end{aligned}$$

$$\text{If } \mathbf{x},\ \mathbf{x}' \in \Re^N \Rightarrow$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbf{x} \cdot \mathbf{x}'$$

$$\begin{aligned} \Phi : \mathcal{X} &\longrightarrow F \subset \Re^N \\ x &\longrightarrow \mathbf{x} \end{aligned}$$

$$k(x, x') \equiv \mathbf{x} \cdot \mathbf{x}' = \Phi(x) \cdot \Phi(x')$$

Support Vector Machines: Kernel methods

$$(\mathbf{x_1}, y_1), \ldots, (\mathbf{x}_m, y_m) \in \mathcal{X} \times \{\pm 1\} \quad \text{~~with } \mathcal{X} \subset \Re^N\text{~~}$$

$$\text{maximize} \quad W(\boldsymbol{\alpha}) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \, \alpha_j \, y_i \, y_j \, (\mathbf{x}_i \cdot \mathbf{x}_j)$$

$$\text{s.t.} \quad 0 \le \alpha_i \le C \quad i = 1, \ldots, m \quad \text{and} \quad \sum_{i=1}^{m} \alpha_i y_i = 0$$

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^{m} y_i \, \alpha_i (\mathbf{x} \cdot \mathbf{x}_i) + b \right)$$

$$b = 1 - \sum_{i=1}^{m} \alpha_i \, y_i \, (\mathbf{x}_i \cdot \mathbf{x}_j) \quad \text{with} \quad 0 < \alpha_j < C \Rightarrow \xi_j = 0$$

$k : \mathcal{X}$ ...

$(x$ ...

If $\mathbf{x},\ \mathbf{x}' \in \Re^N \Rightarrow$

$$k(\mathbf{x}, \mathbf{x}') = \mathbf{x} \cdot \mathbf{x}'$$

$$k(x, x') \equiv \mathbf{x} \cdot \mathbf{x}' = \Phi(x) \cdot \Phi(x')$$

Support Vector Machines: Kernel methods

$$(\mathbf{x_1}, y_1), \ldots, (\mathbf{x}_m, y_m) \in \mathcal{X} \times \{\pm 1\} \quad \text{~~with}\ \ \mathcal{X} \subset \Re^N\text{~~}$$

$$\text{maximize} \quad W(\boldsymbol{\alpha}) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i\, \alpha_j\, y_i\, y_j\, (\mathbf{x}_i \cdot \mathbf{x}_j)$$

$$\text{s.t.} \quad 0 \le \alpha_i \le C \quad i = 1, \ldots, m \quad \text{and} \quad \sum_{i=1}^{m} \alpha_i y_i = 0$$

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^{m} y_i\, \alpha_i (\mathbf{x} \cdot \mathbf{x}_i) + b \right)$$

$$b = 1 - \sum_{i=1}^{m} \alpha_i\, y_i (\mathbf{x}_i \cdot \mathbf{x}_j) \quad \text{with} \quad 0 < \alpha_j < C \Rightarrow \xi_j = 0$$

$k : \mathcal{X}$ ...

$(x$ ...

If $\mathbf{x},\ \mathbf{x}' \in \Re^N \Rightarrow$

$$k(\mathbf{x}, \mathbf{x}') = \mathbf{x} \cdot \mathbf{x}'$$

$$k(x, x') \equiv \mathbf{x} \cdot \mathbf{x}' = \Phi(x) \cdot \Phi(x')$$

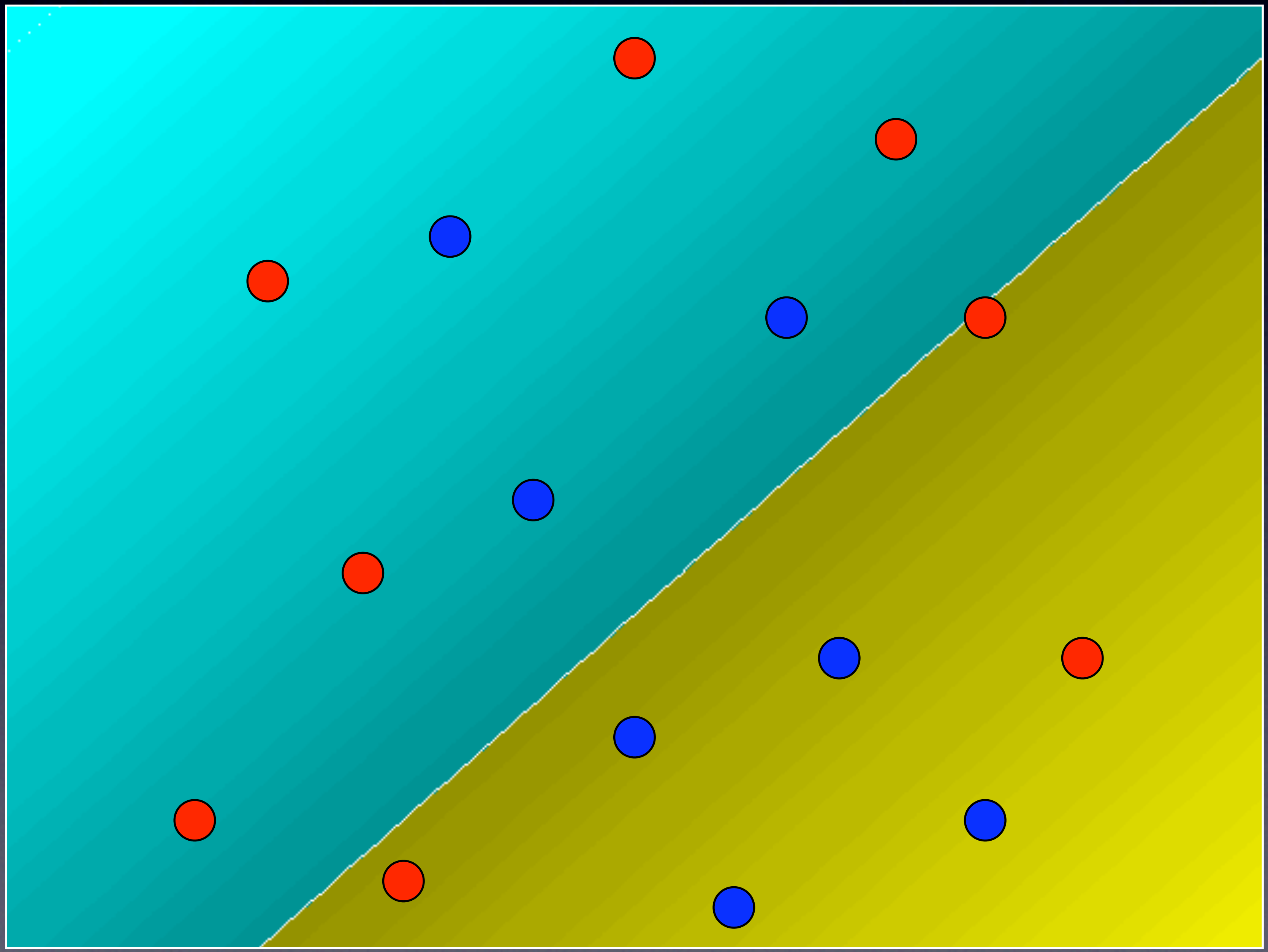
Support Vector Machines: Kernel methods

linear : $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$

polynomial : $K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma\, \mathbf{x}_i \cdot \mathbf{x}_j + r_0)^d$

RBF : $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right)$

sigmoid : $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh\left(\gamma\, \mathbf{x}_i \cdot \mathbf{x}_j + r_0\right)$



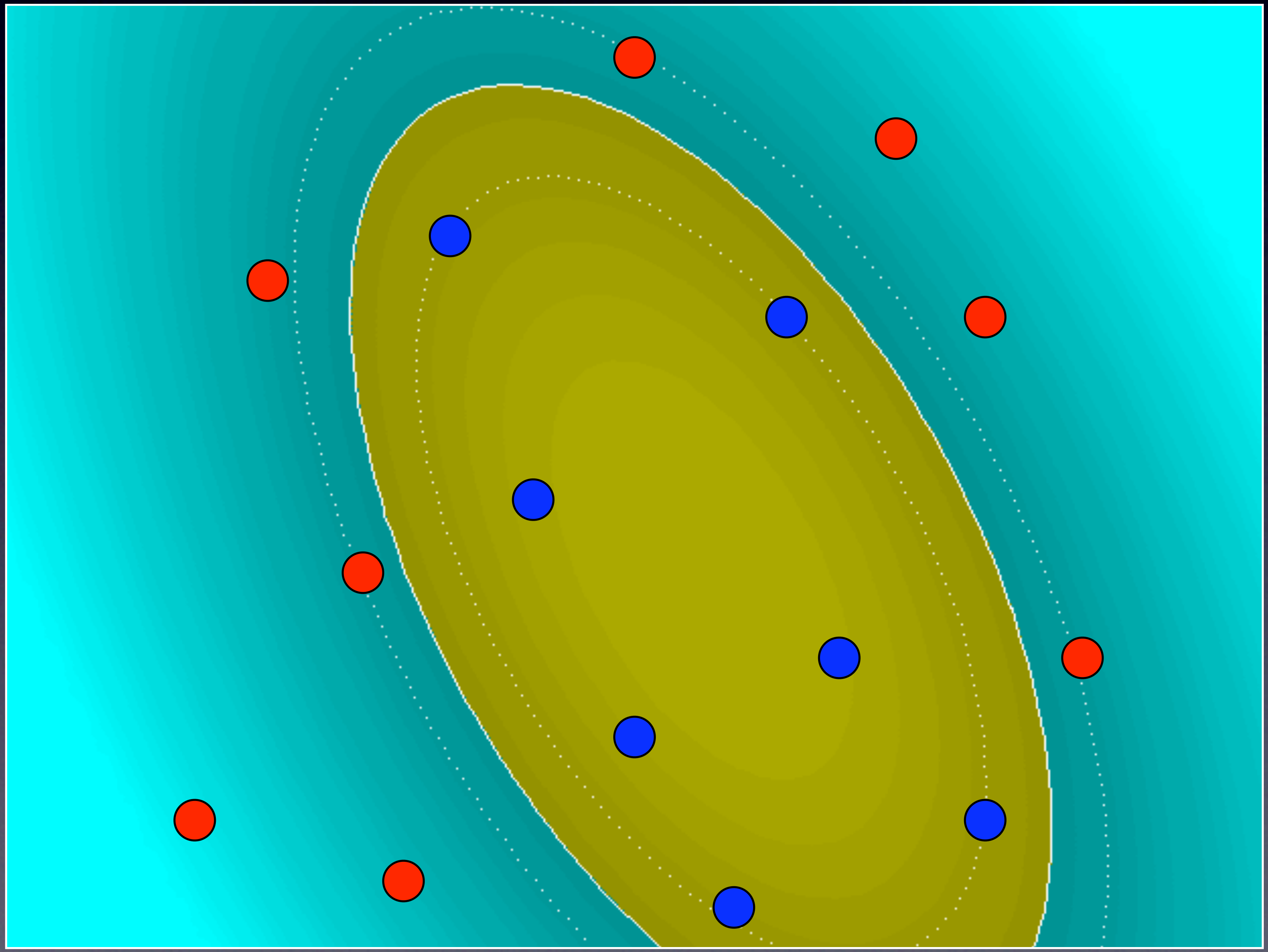
Support Vector Machines: Kernel methods

$$\text{linear}: \quad K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$$

$$\text{polynomial}: \quad K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma\, \mathbf{x}_i \cdot \mathbf{x}_j + r_0)^d$$

$$\text{RBF}: \quad K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right)$$

$$\text{sigmoid}: \quad K(\mathbf{x}_i, \mathbf{x}_j) = \tanh\left(\gamma\, \mathbf{x}_i \cdot \mathbf{x}_j + r_0\right)$$



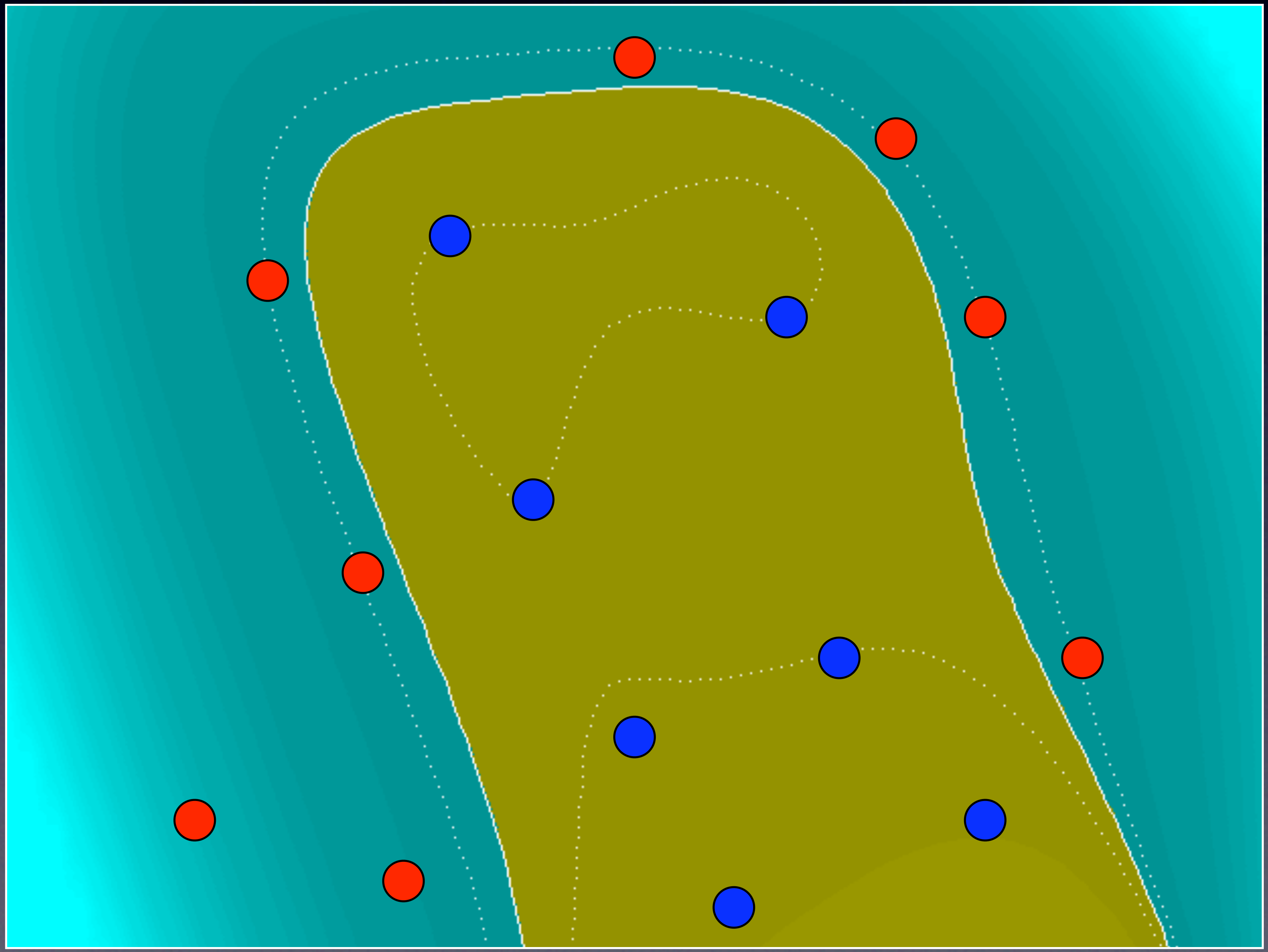
Support Vector Machines: Kernel methods

$$\text{linear}: \quad K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$$

$$\text{polynomial}: \quad K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma\, \mathbf{x}_i \cdot \mathbf{x}_j + r_0)^d$$

$$\text{RBF}: \quad K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2\right)$$

$$\text{sigmoid}: \quad K(\mathbf{x}_i, \mathbf{x}_j) = \tanh\left(\gamma\, \mathbf{x}_i \cdot \mathbf{x}_j + r_0\right)$$



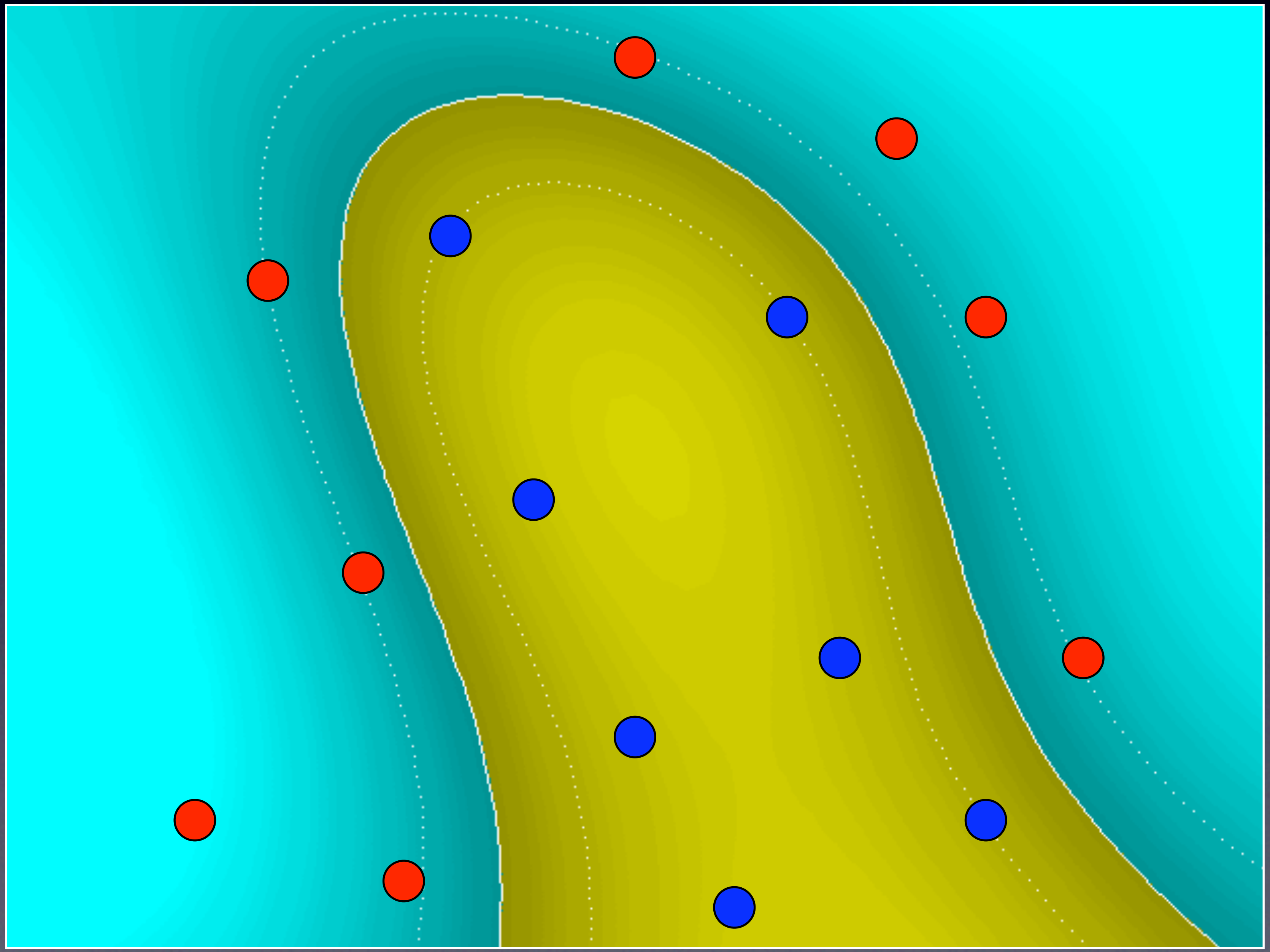
Support Vector Machines: Kernel methods

linear : $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j$

polynomial : $K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma\, \mathbf{x}_i \cdot \mathbf{x}_j + r_0)^d$

RBF : $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma\, \|\mathbf{x}_i - \mathbf{x}_j\|^2\right)$

sigmoid : $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh\left(\gamma\, \mathbf{x}_i \cdot \mathbf{x}_j + r_0\right)$



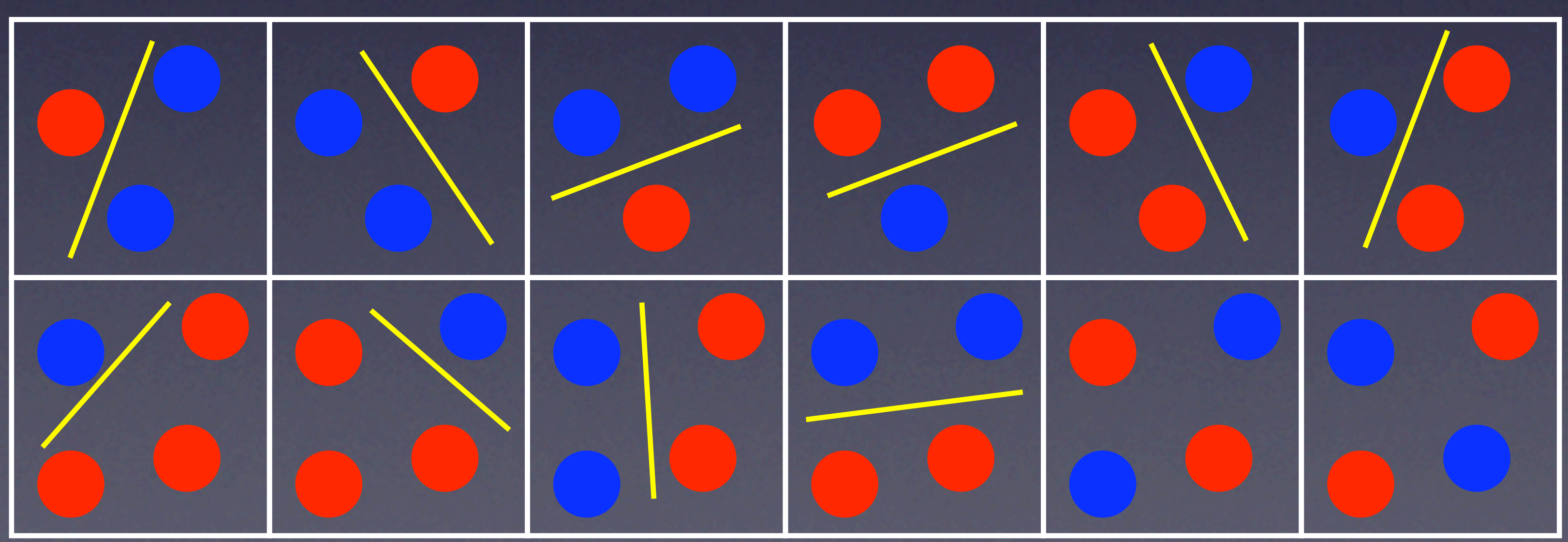
Support Vector Machines: Error constraints

$$R_{\text{emp}}[f] = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{2}|f(x_i) - y_i| \quad , \quad R[f] = \int \frac{1}{2}|f(x) - y|dP(x,y)$$

$$R[f] \leq R_{\text{emp}}[f] + \phi\left(\frac{h}{m}, \frac{\log(\eta)}{m}\right)$$

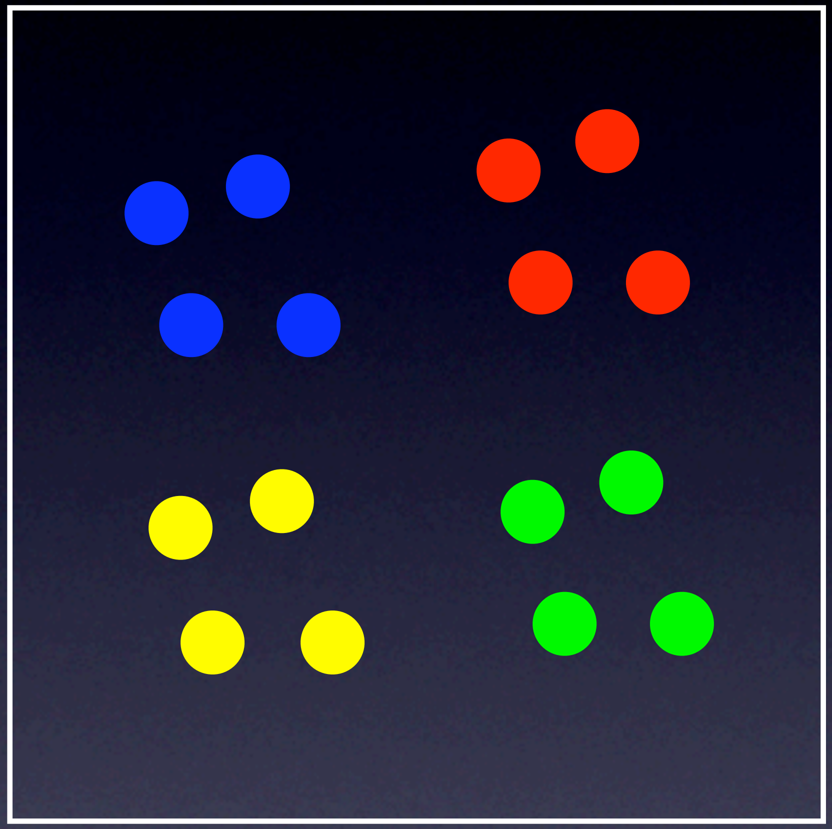
$$\phi\left(\frac{h}{m}, \frac{\log(\eta)}{m}\right) = \sqrt{\frac{h\left(\log\frac{2m}{h} + 1\right) - \log(\eta/4)}{m}}$$

Capacity:



Multiclass Support Vector Machines

2-class SVM generalization to *k* classes.

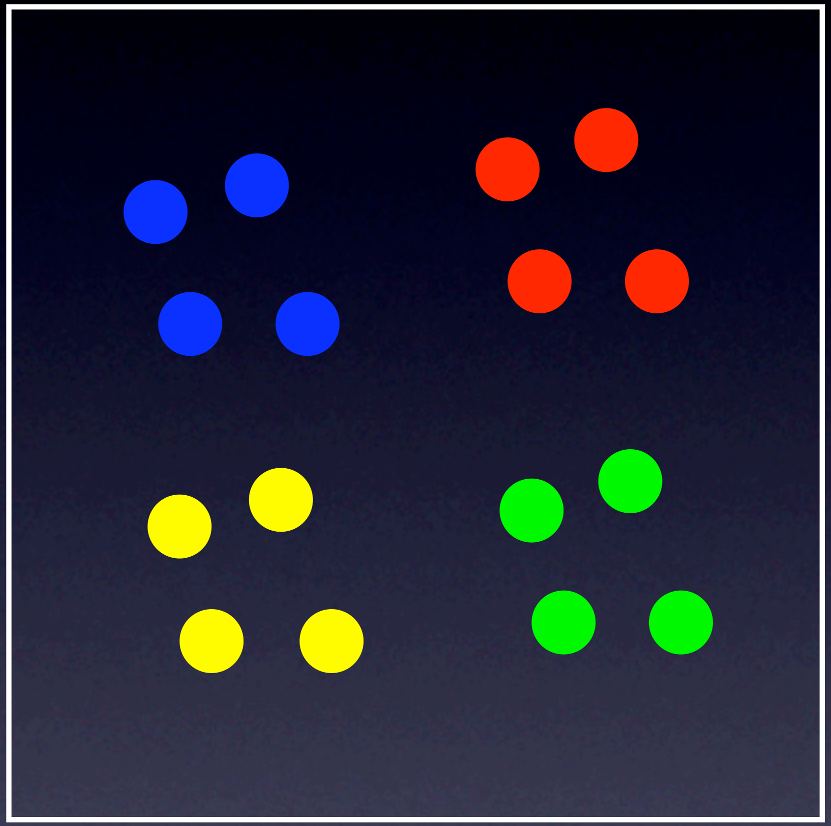


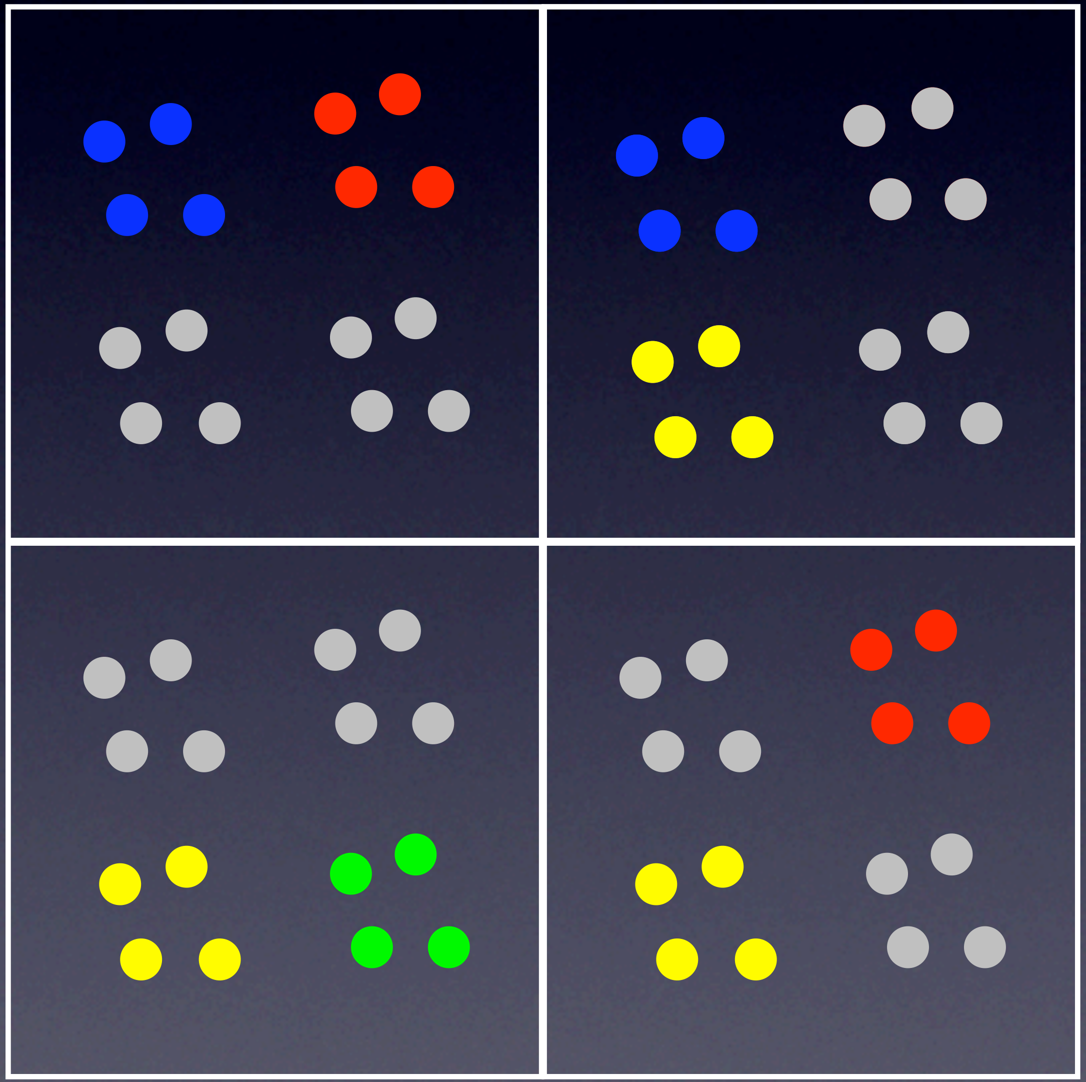
Multiclass Support Vector Machines

2-class SVM generalization to *k* classes.

one-versus-one:

- Train $k\,(k+1)/2$ SVM
- Classification: Poll



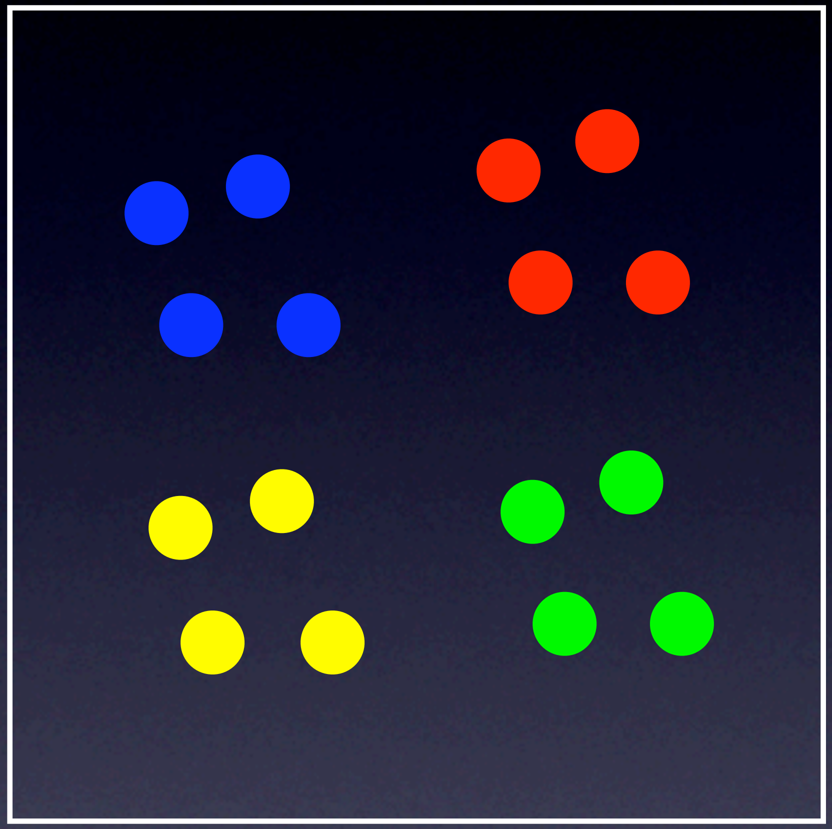


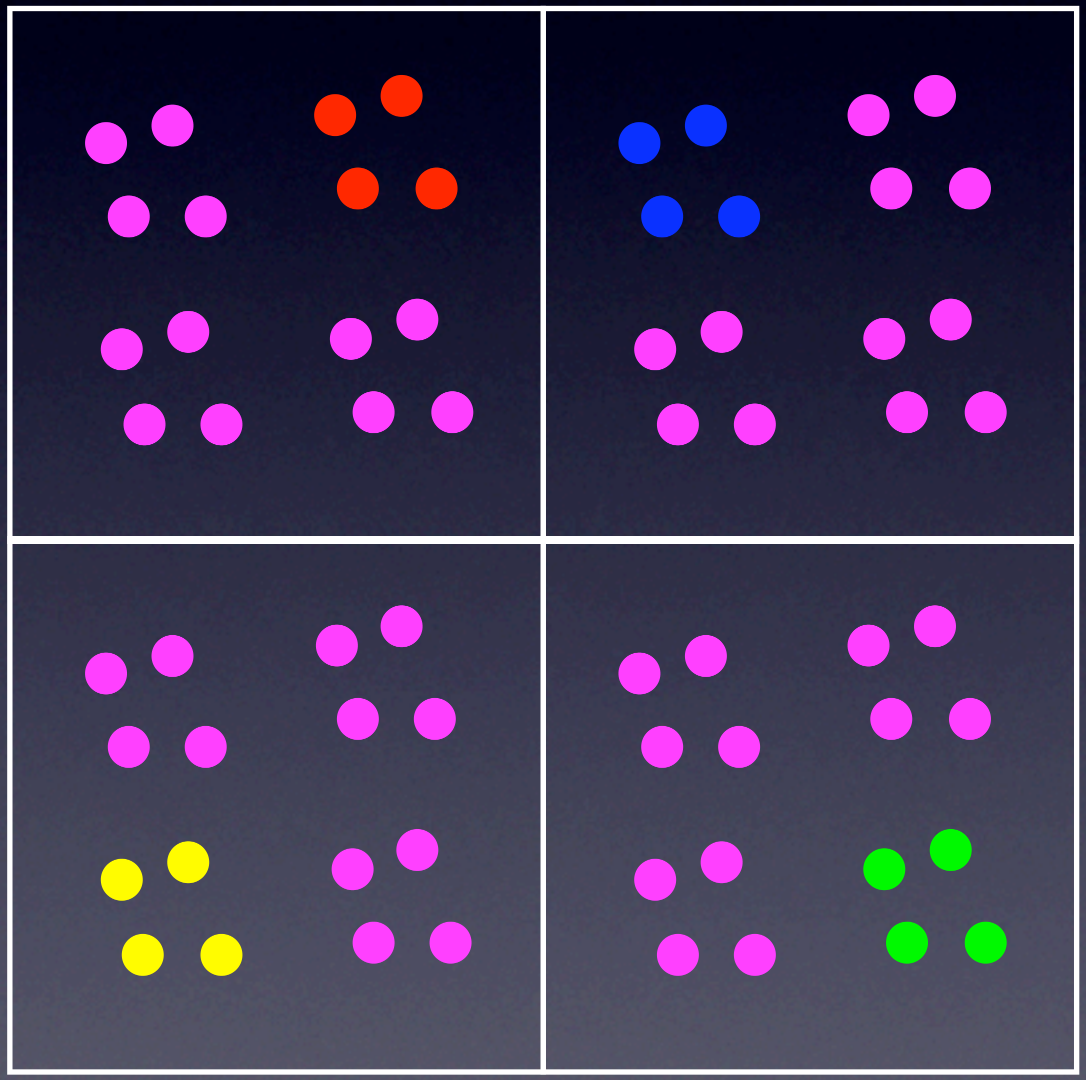
Multiclass Support Vector Machines

2-class SVM generalization to *k* classes.

one-versus-all:

- Train *k* SVM
- Classification: max{d.f.}





Support Vector Machines in HEP

P.Vannerem et al., Classifying LEP data with support vector algorithms (hep-ex/9905027)

In e+ e- → q qbar:
- Charm-tagging.
- Muon identification.

ANN and SVM give consistent results

A. Vaiciulis, SVM in analysis of Top quark production (Nucl. Instrum. Meth. A502 (2003) 492)

Signal and background efficiency consistent with best set of cuts

TMVA

TMVA: Toolkit for Multivariate Analysis

Integrated machine learning environment for CERN's ROOT
http://tmva.sourceforge.net

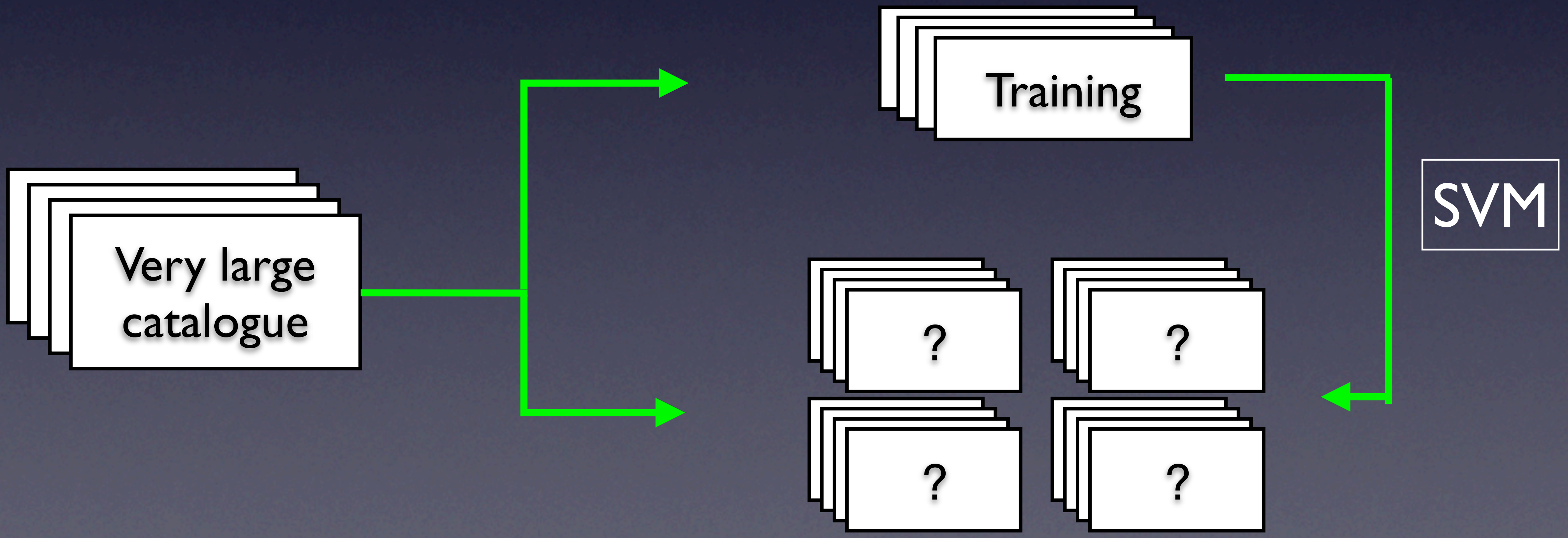
Summary

Class	M
Early	1
Late	2
Irregular	3

Non-parametric coefficients: C, A, S...

$$\mathcal{M} = \mathcal{M}(C,\, A,\, S,\, G,\, M_{20},\, \ldots)$$

Support Vector Machines:



ZEST+: The Zurich estimator of structural types

- The evolution of ZEST+
- General Approach
- Details
- Applications: COSMOS
- Challenges ahead

The evolution of ZEST+

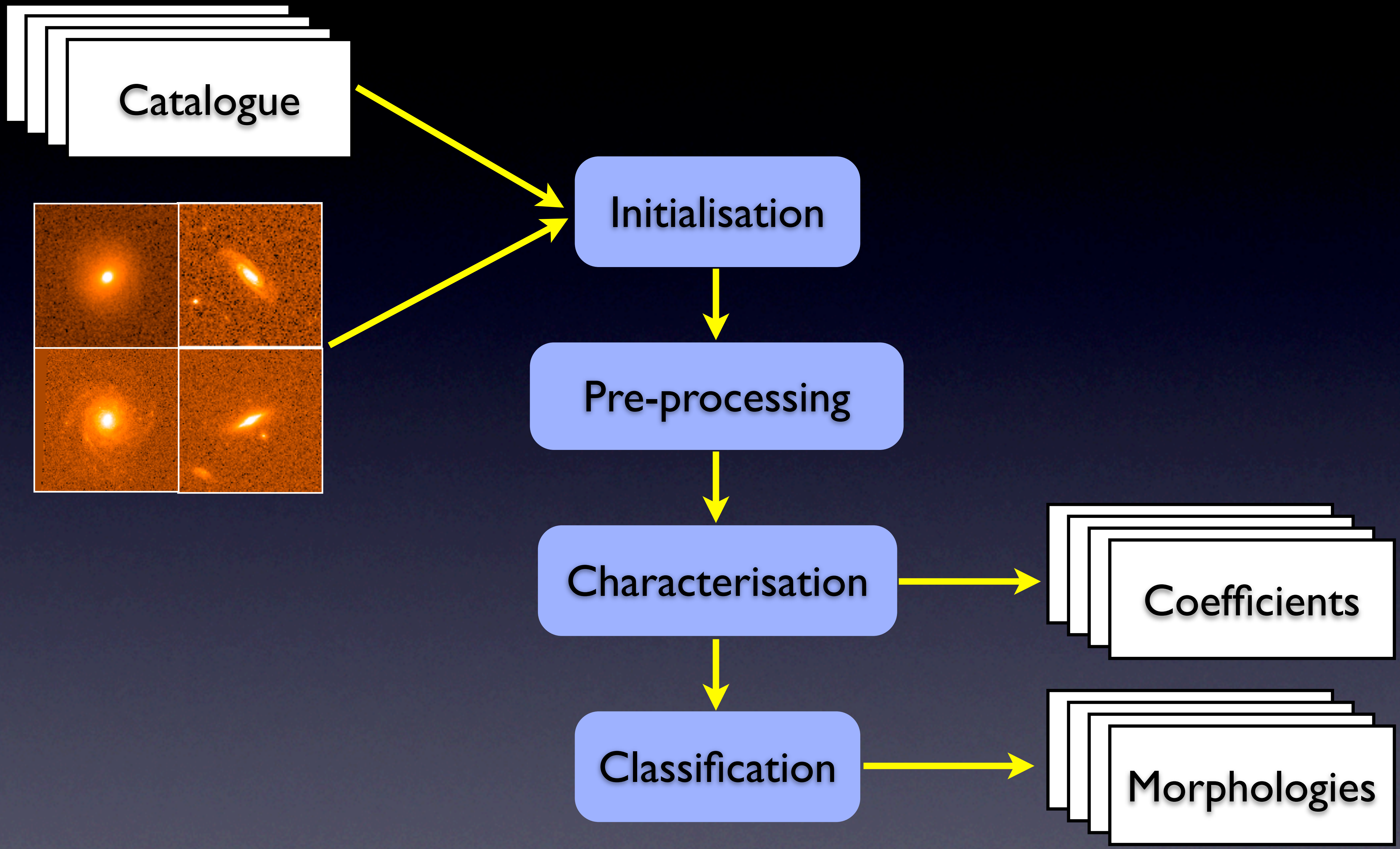
ZEST: C. Scarlata & M. Carollo (2007)

- C, A, G, M_{20}, ϵ, and Sérsic n.
- PCA analysis: 3D classification grid.
- IDL application, no public release.

ZEST+:

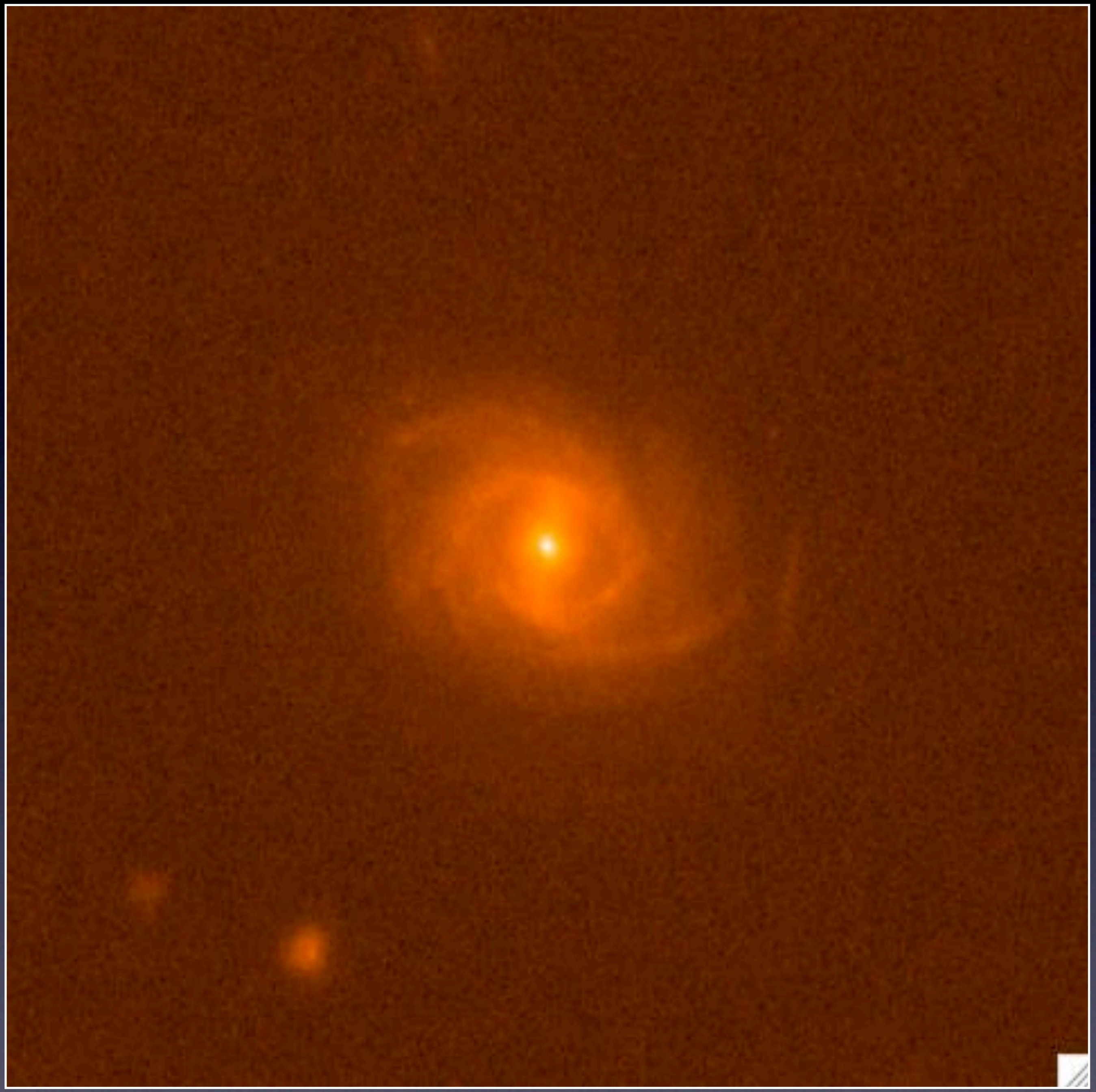
- First C++ version, without classification (E. Weihs).
- Further modifications (T. Bschorr).
- Complete rewrite, new features, SVM classification (M.C.)

ZEST+ Architecture



ZEST+: Pre-processing

- Pre-processing
- Basic segmentation
- Image cleaning
- Segmentation refinement



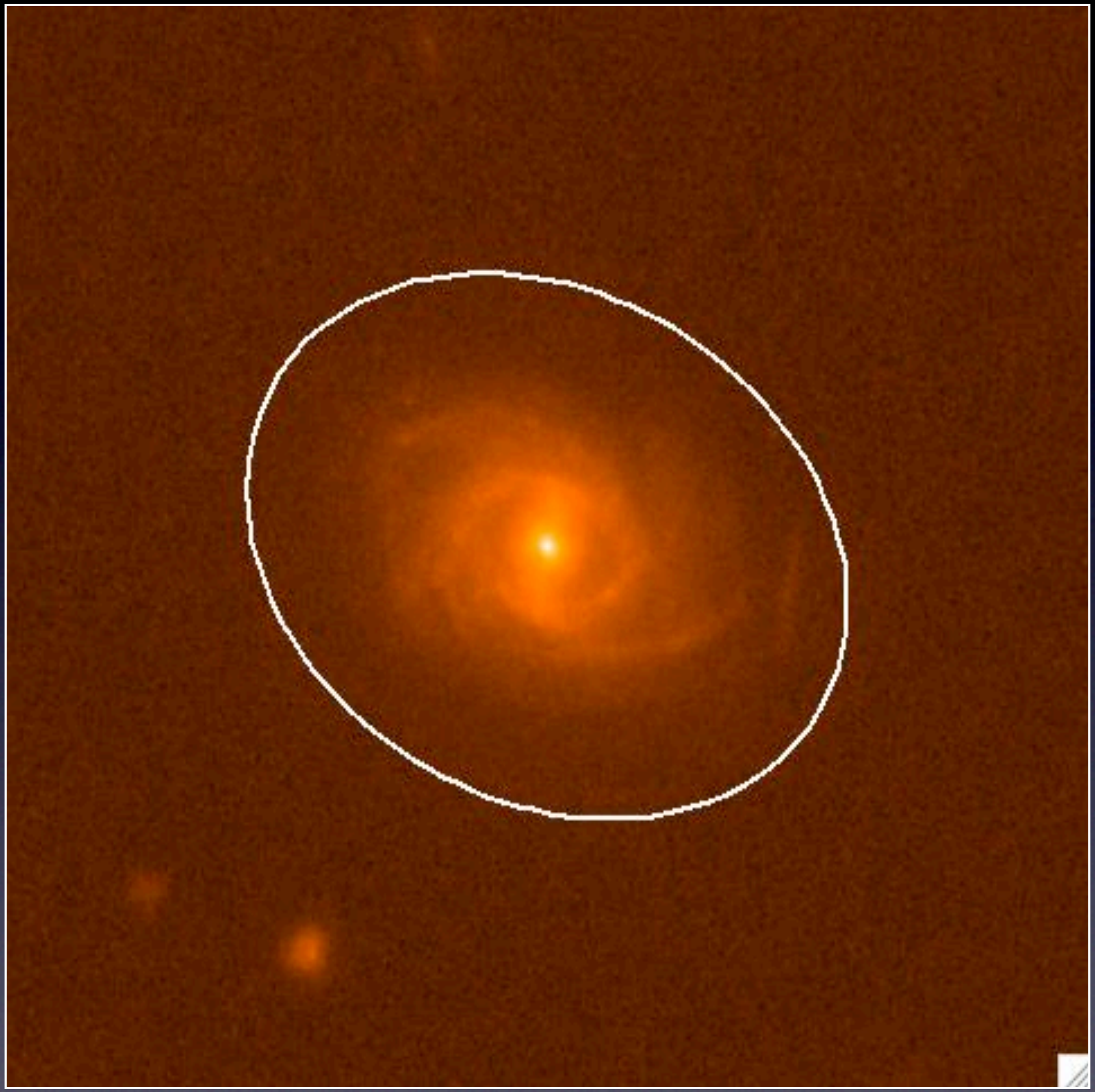
ZEST+: Pre-processing

Pre-processing

Basic segmentation

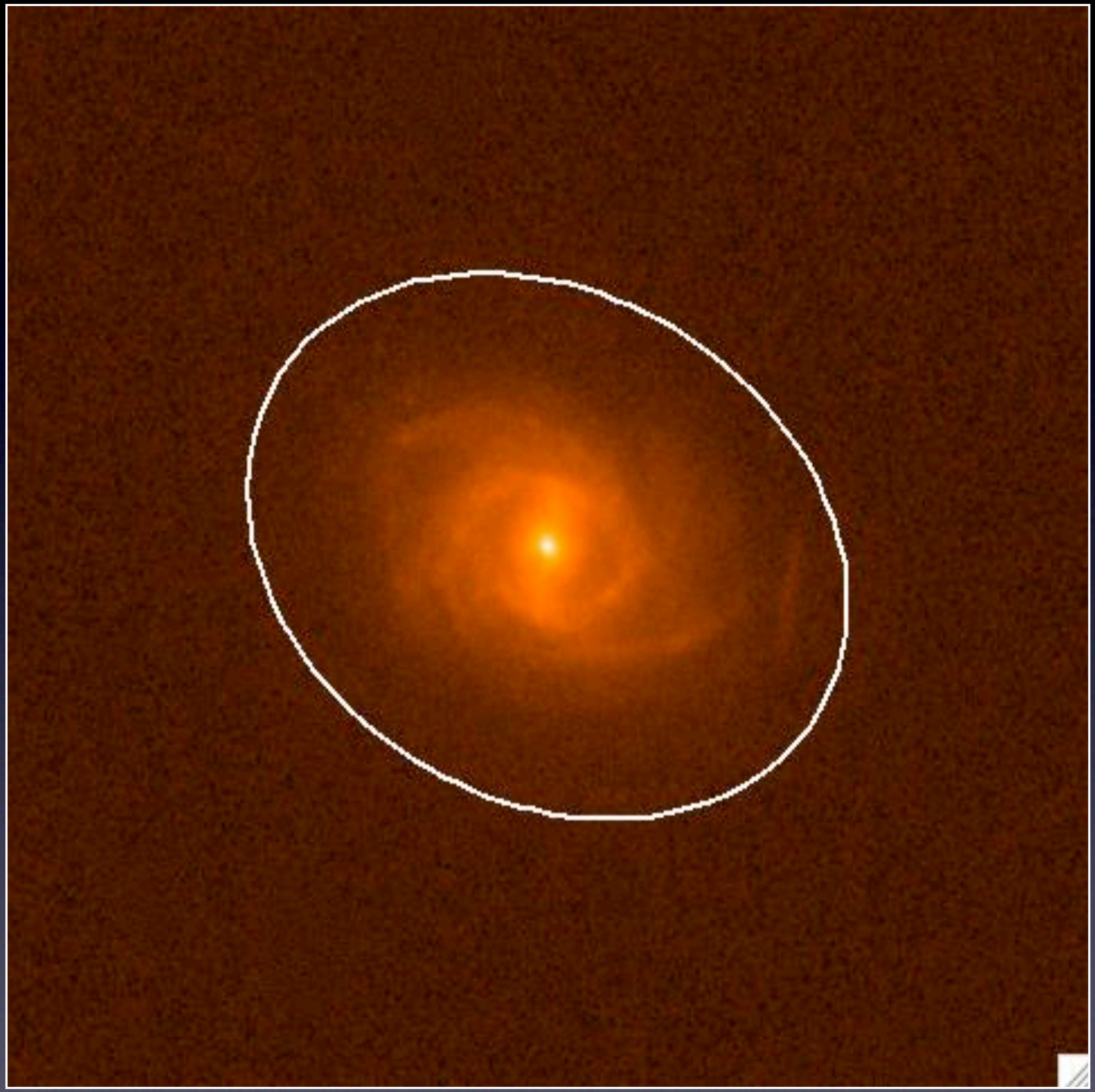
Image cleaning

Segmentation refinement



ZEST+: Pre-processing

- Pre-processing
- Basic segmentation
- **Image cleaning**
- Segmentation refinement



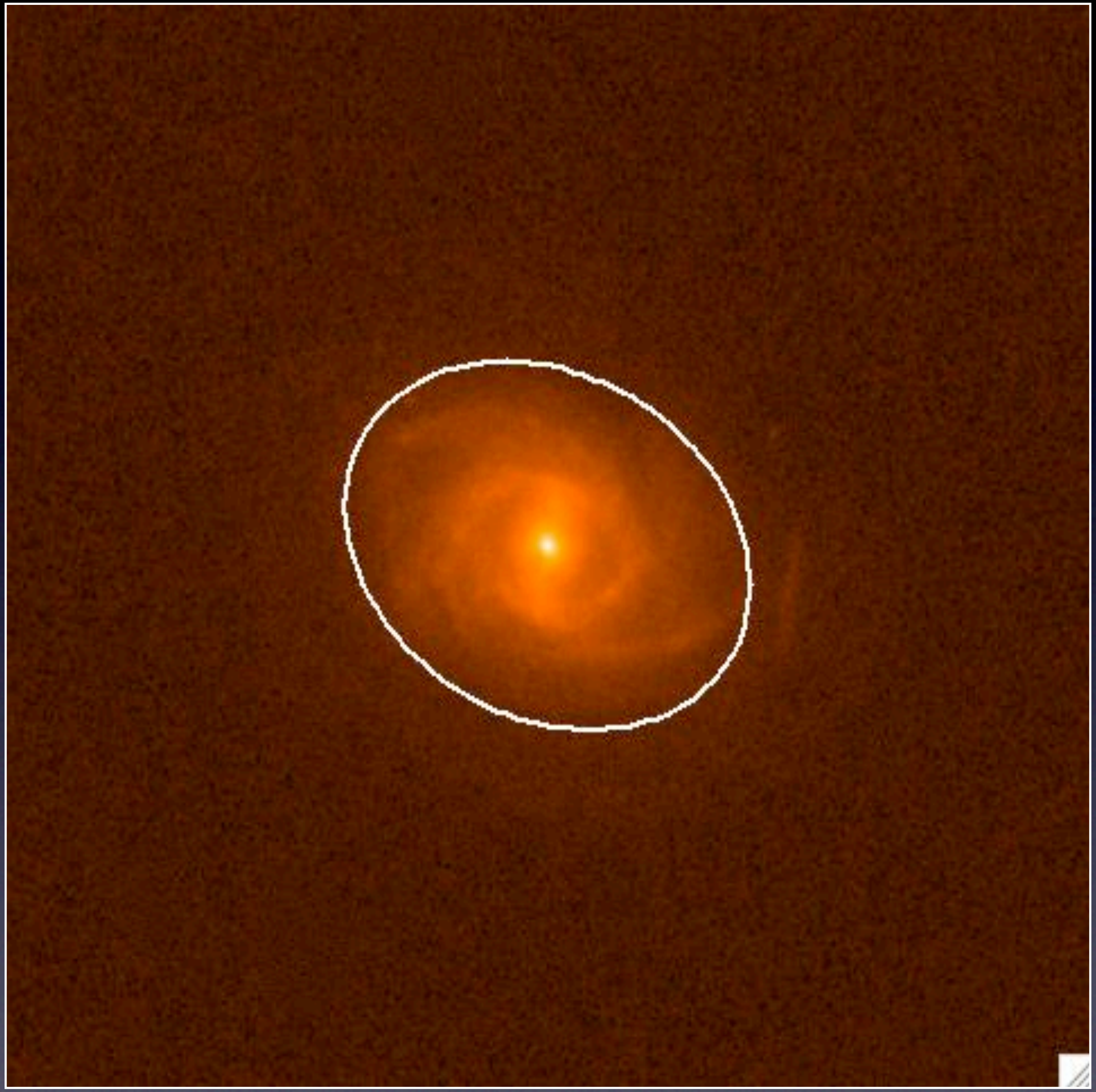
ZEST+: Pre-processing

Pre-processing

Basic segmentation

Image cleaning

Segmentation refinement



ZEST+: Segmentation refinement

Galaxy’s center: Center of asymmetry

Galaxy’s size: Petrosian radius

$$\eta(R) = \frac{2\pi \int_0^R I(R')dR'}{\pi R^2 I(R)}$$

$$\frac{1}{\eta({R^\alpha}_p)} = \alpha$$

$$R_p \equiv R_p^{0.2}$$

ZEST+: Characterisation

Characterisation

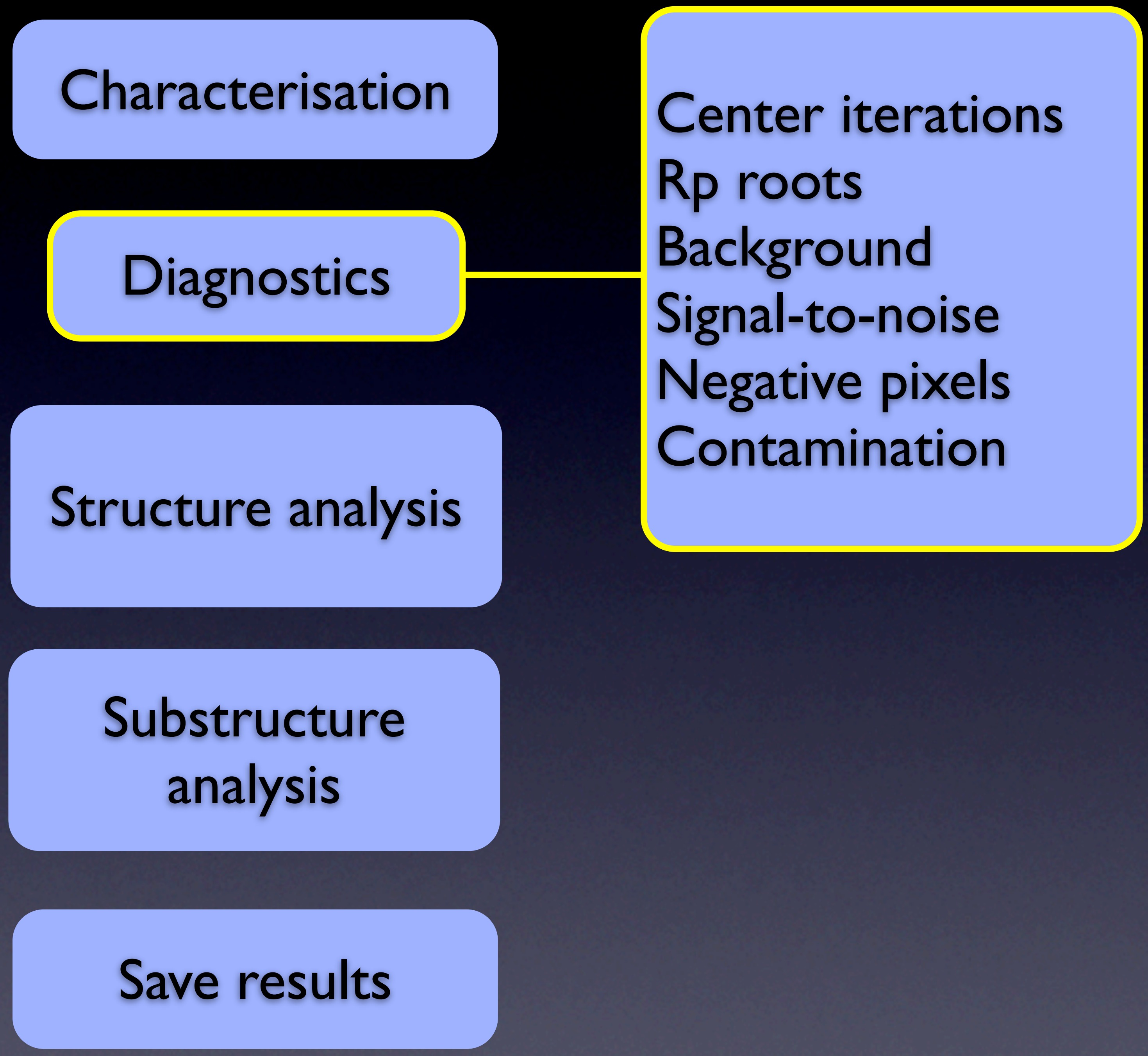
Diagnostics

Structure analysis

Substructure analysis

Save results

ZEST+: Characterisation



ZEST+: Characterisation

- Characterisation
- Diagnostics
- Structure analysis — C, A, S, G, M_{20}
- Substructure analysis
- Save results

ZEST+: Characterisation

Characterisation

Diagnostics

Structure analysis

Substructure analysis

Selection
ϵ, C, A, G, M_{20}

Save results

ZEST+: Characterisation

Characterisation

Diagnostics

Structure analysis

Substructure analysis

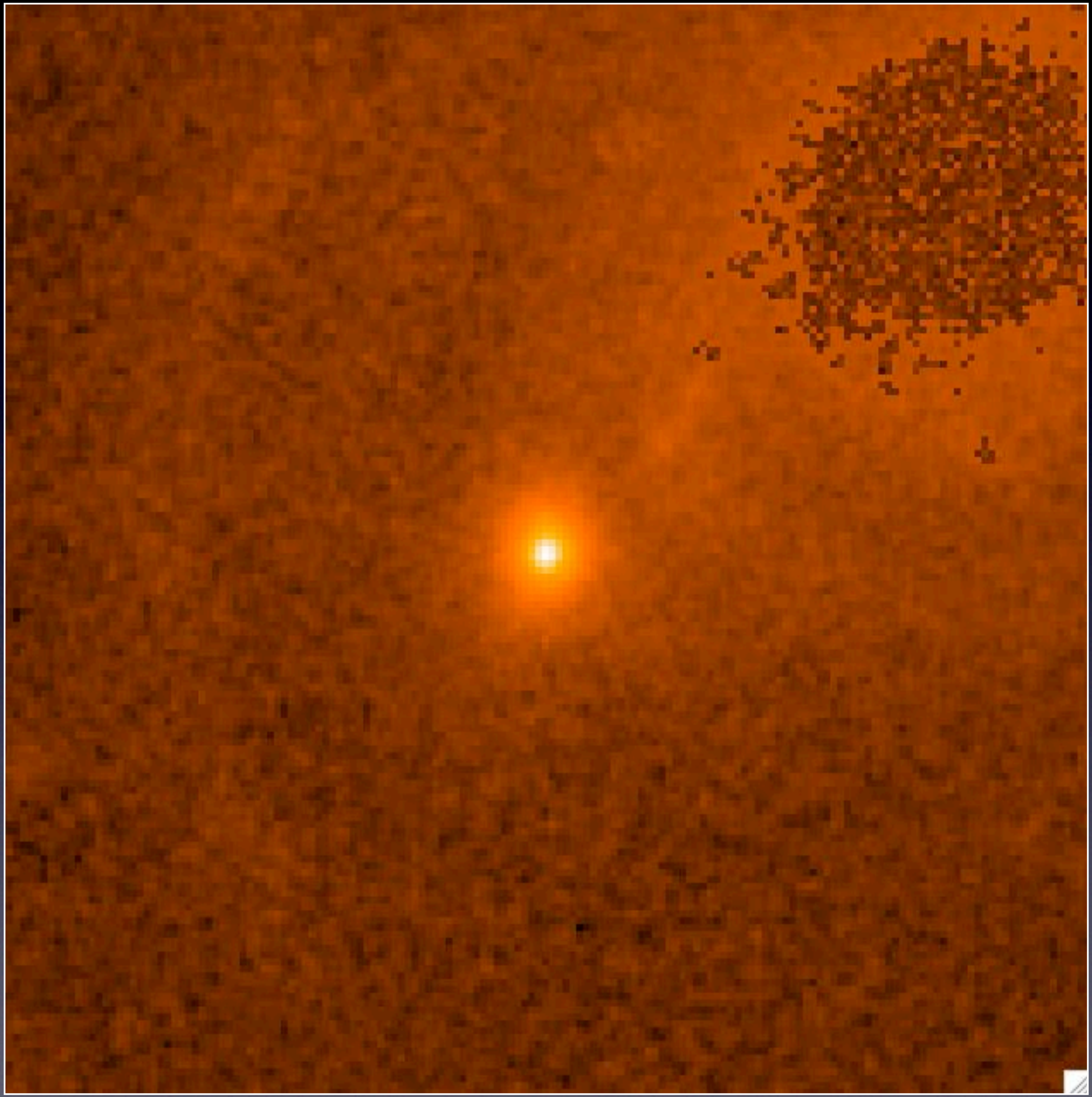
Save results

Id	…	C	A	S	M_{20}	G	ϵ	C	A	G	M_{20}	Err.

ZEST+: Coefficients revisited

Differences from ideal case:
- Negative pixels
- Low signal-to-noise ratio
- Background artefacts

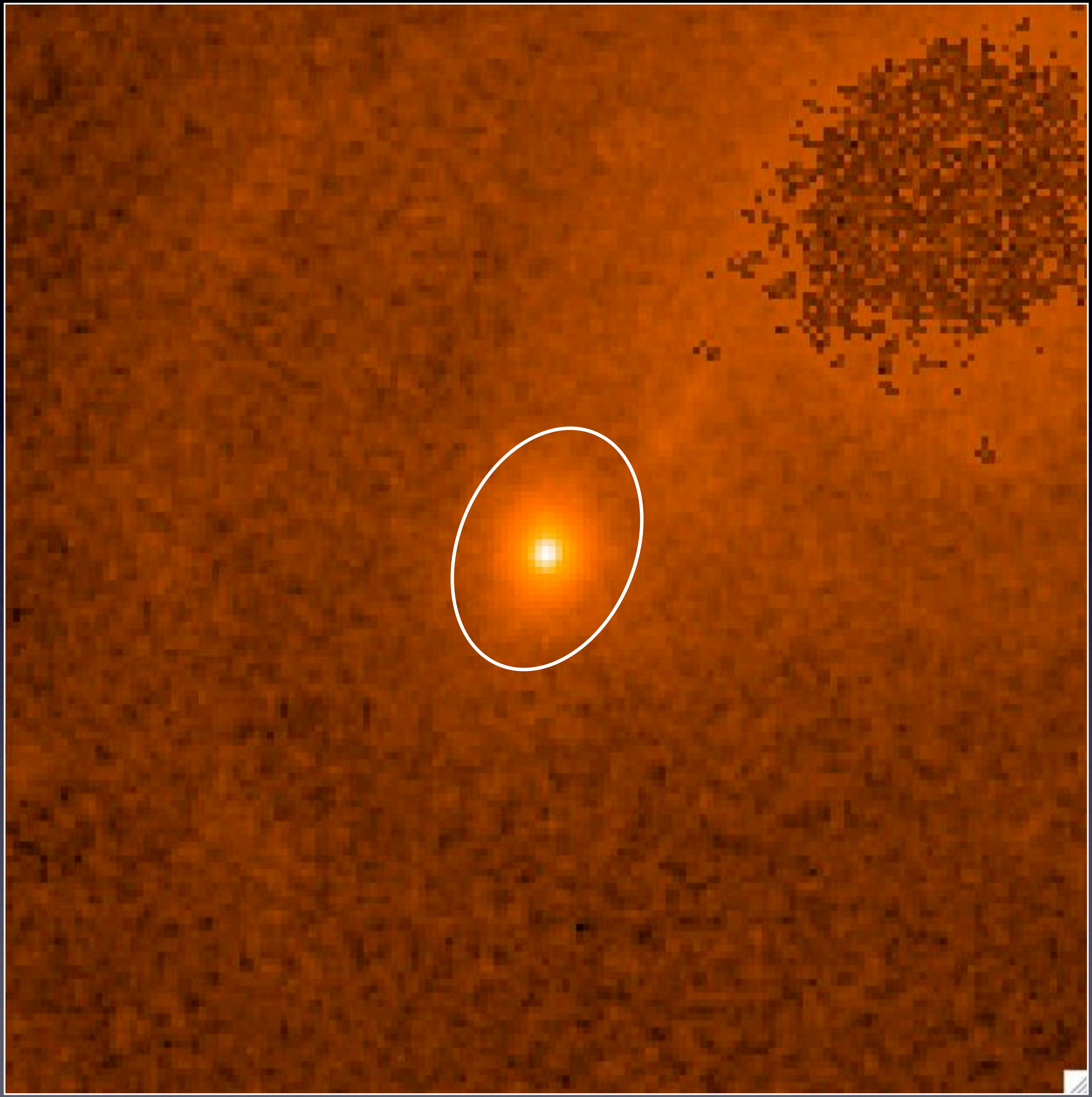
$$\begin{aligned} A &= A_0 - A_{\rm bkg} \\ S &= S_0 - S_{\rm bkg} \\ G &= G(I_j) \rightarrow G(|I_j|) \end{aligned}$$



ZEST+: Coefficients revisited

Differences from ideal case:
- Negative pixels
- Low signal-to-noise ratio
- Background artefacts

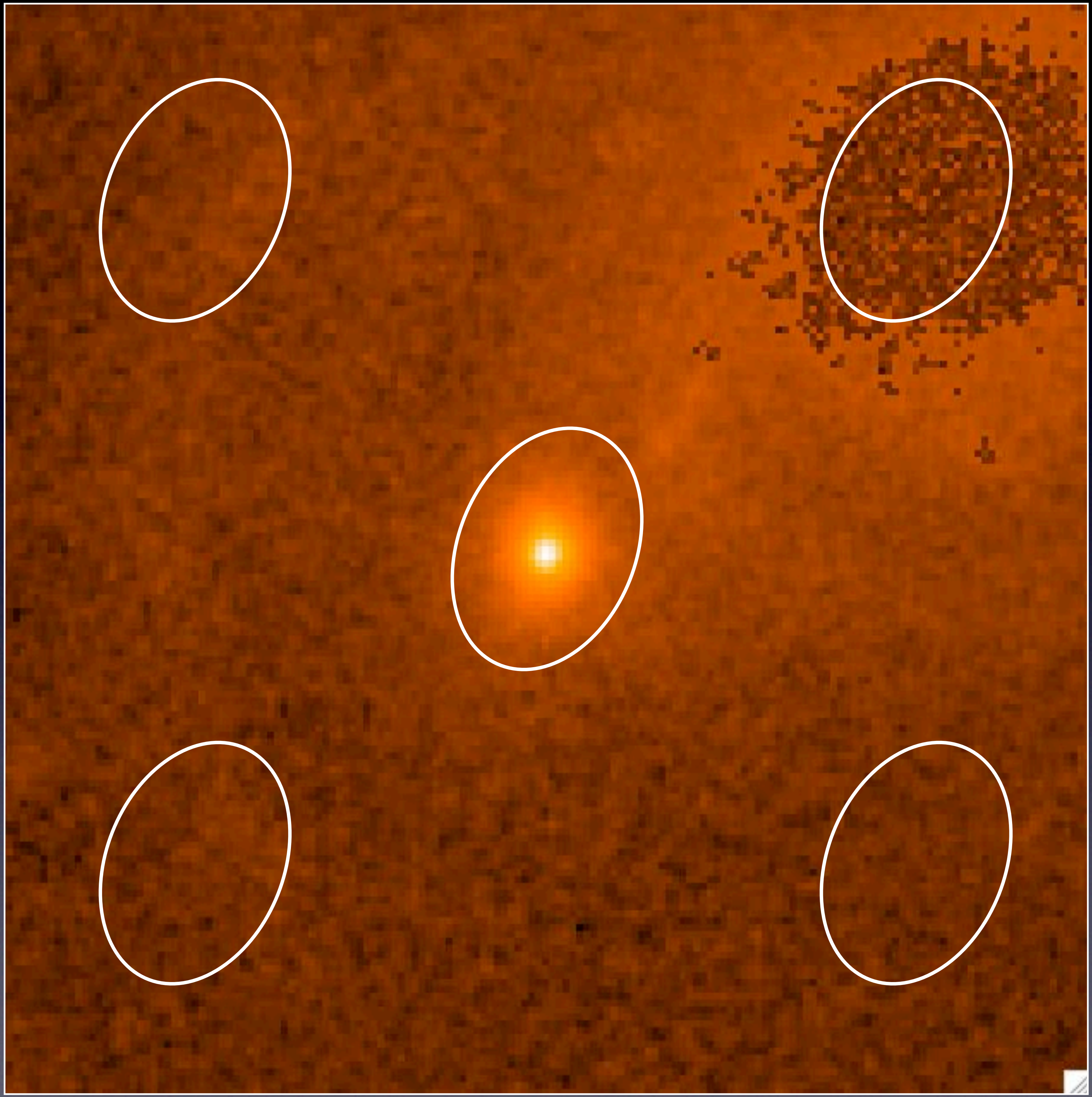
$$\begin{aligned} A &= A_0 - A_{\rm bkg} \\ S &= S_0 - S_{\rm bkg} \\ G &= G(I_j) \rightarrow G(|I_j|) \end{aligned}$$



ZEST+: Coefficients revisited

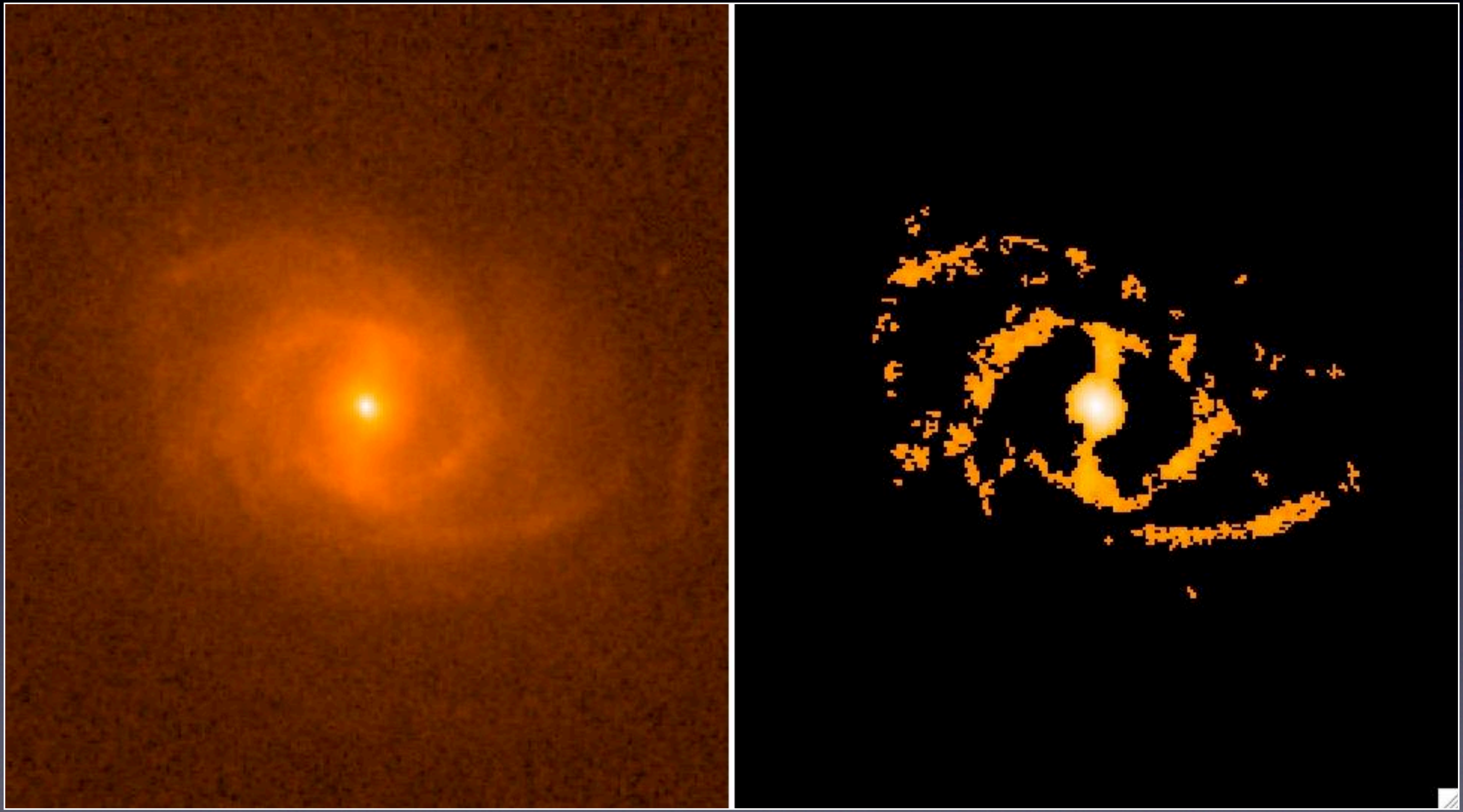
Differences from ideal case:
- Negative pixels
- Low signal-to-noise ratio
- Background artefacts

$$
\begin{aligned}
A &= A_0 - A_{\rm bkg} \\
S &= S_0 - S_{\rm bkg} \\
G &= G(I_j) \rightarrow G(|I_j|)
\end{aligned}
$$



ZEST+: Substructure analysis

1. Self-subtract smoothed image.
2. Threshold and eliminate isolated pixels.
3. Measure all morphological coefficients.



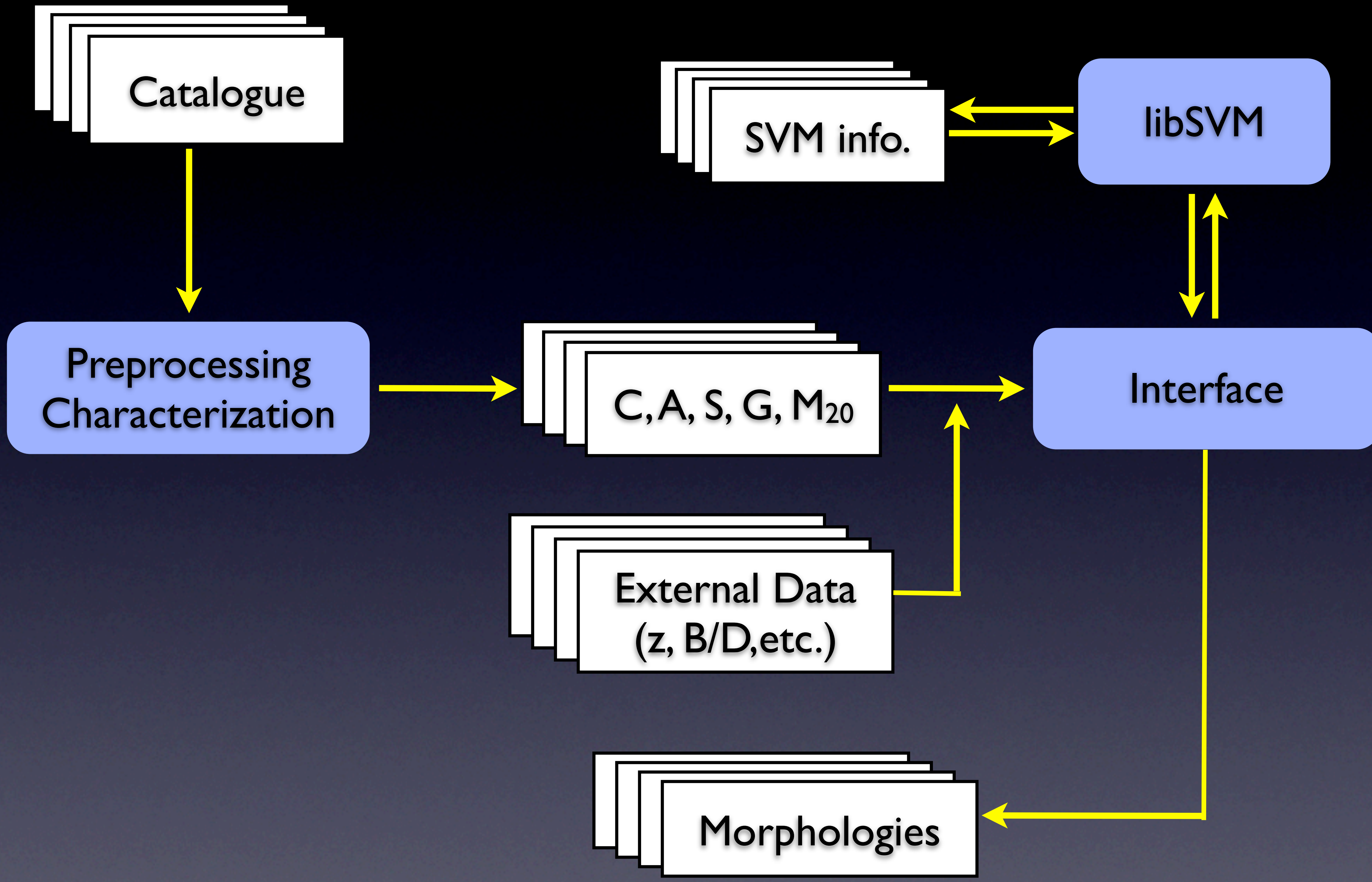
ZEST+: Classification

Algorithm: Support Vector Machines

Implementation: libSVM-3.88 (C. Chang & C. Lin 2001)

- SVM stand-alone C applications.
- SVM library
- Kernels: Linear, polynomial, RBF, sigmoid and user provided.
- Multiclass algorithm: one-versus-one.
- Supports SVM probabilities.

ZEST+: Classification



Recent ZEST+ applications

- Automatic preprocessing of large datasets.
- Mask preparation to use with external fitting programs.
- Petrosian radius calculation in simulated galaxies.
- Morphological data calculation for a recent VLT proposal.
- Star-forming E/S0 galaxies study.
- Substructure identification for tidal features study.
- Morphological analysis in the COSMOS survey.

Cosmic Evolution Survey

cosmos

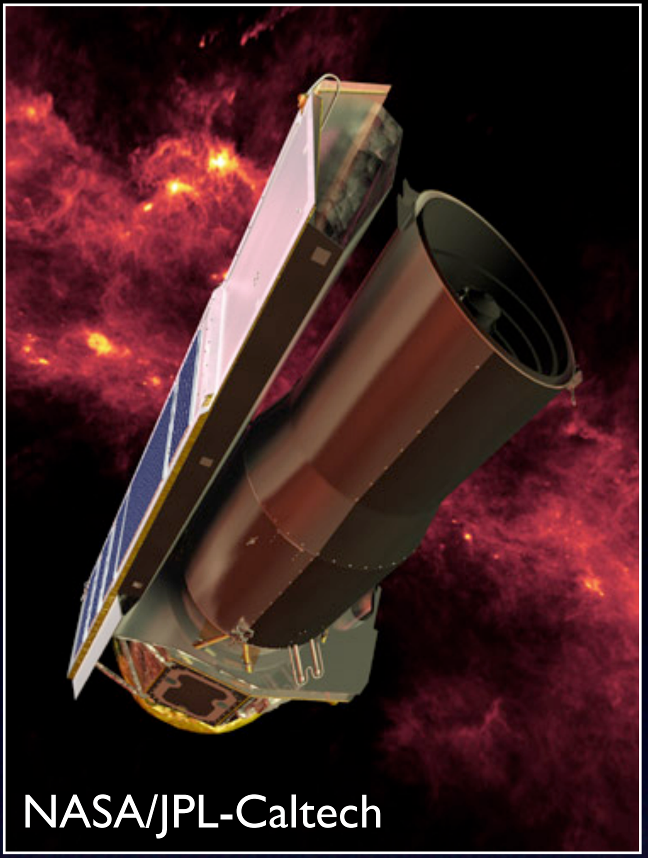
Objective

Probe galaxy formation and evolution as a function of z and LSS environment

- HST Treasury Project with ACS
- Largest HST survey
- 2 square degrees equatorial field
- 2 million objects $I_{AB} > 27$ mag
- Up to z ~ 5


ESO/H.Zodet


Image courtesy of ESA


NASA/JPL-Caltech


Subaru Telescope, NAOJ

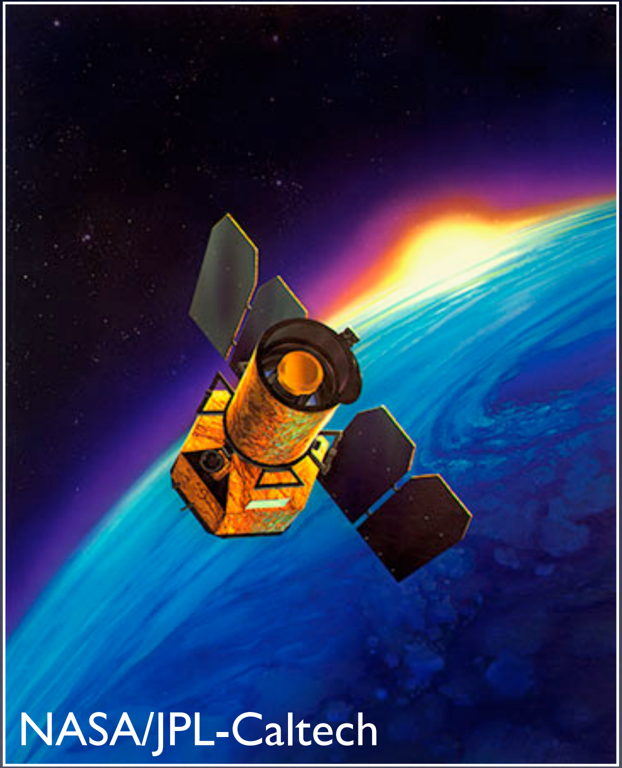
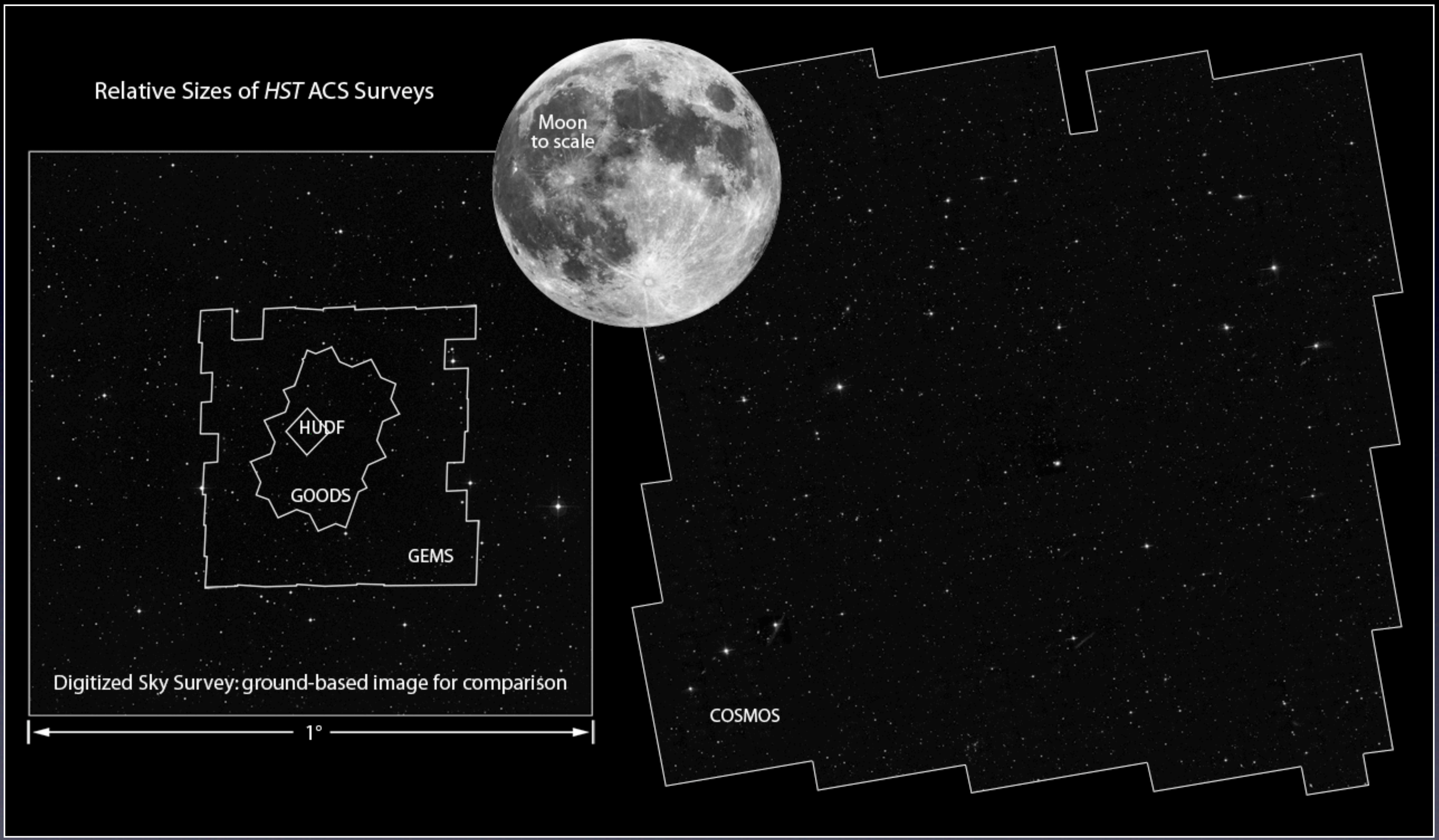

NASA/JPL-Caltech


Image courtesy of NRAO/AUI

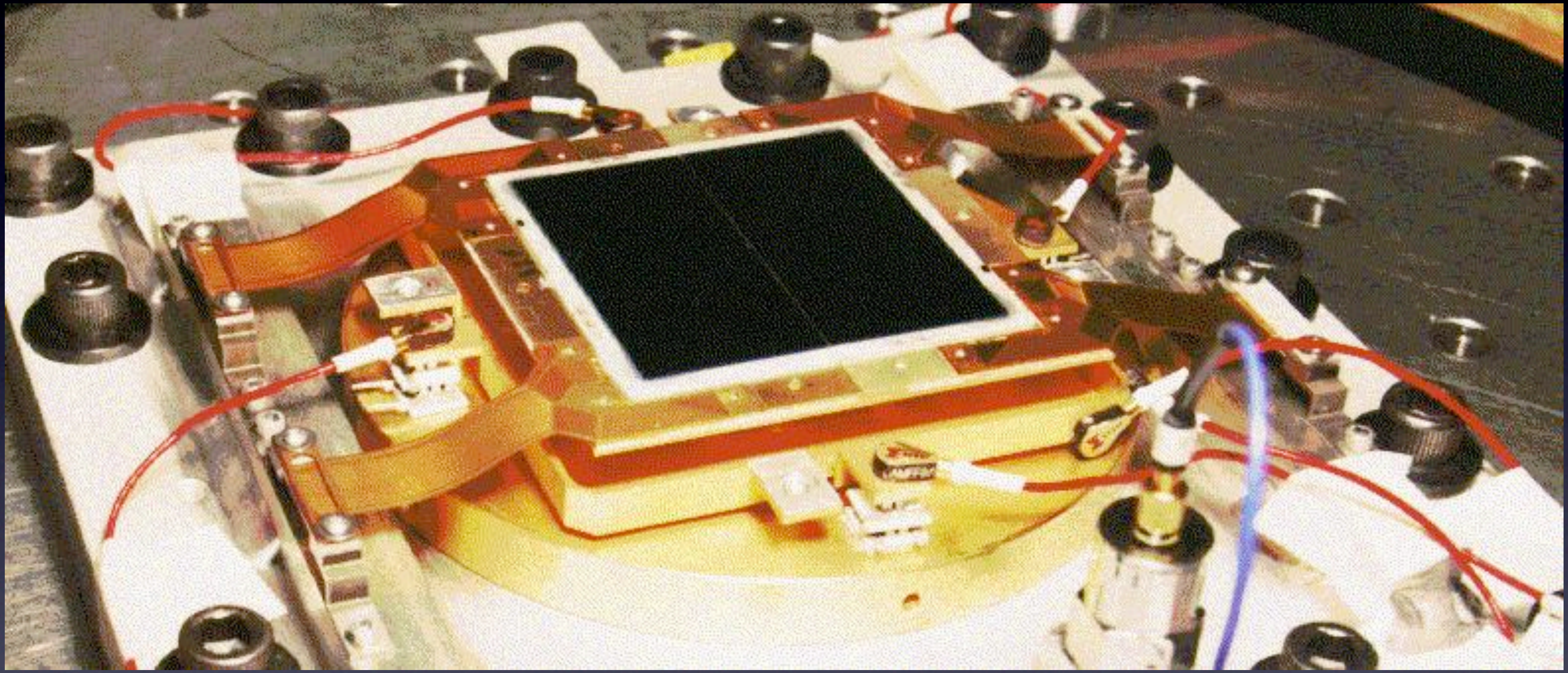
Chandra, UKIRT, NOAO,CFHT, and others

The COSMOS field

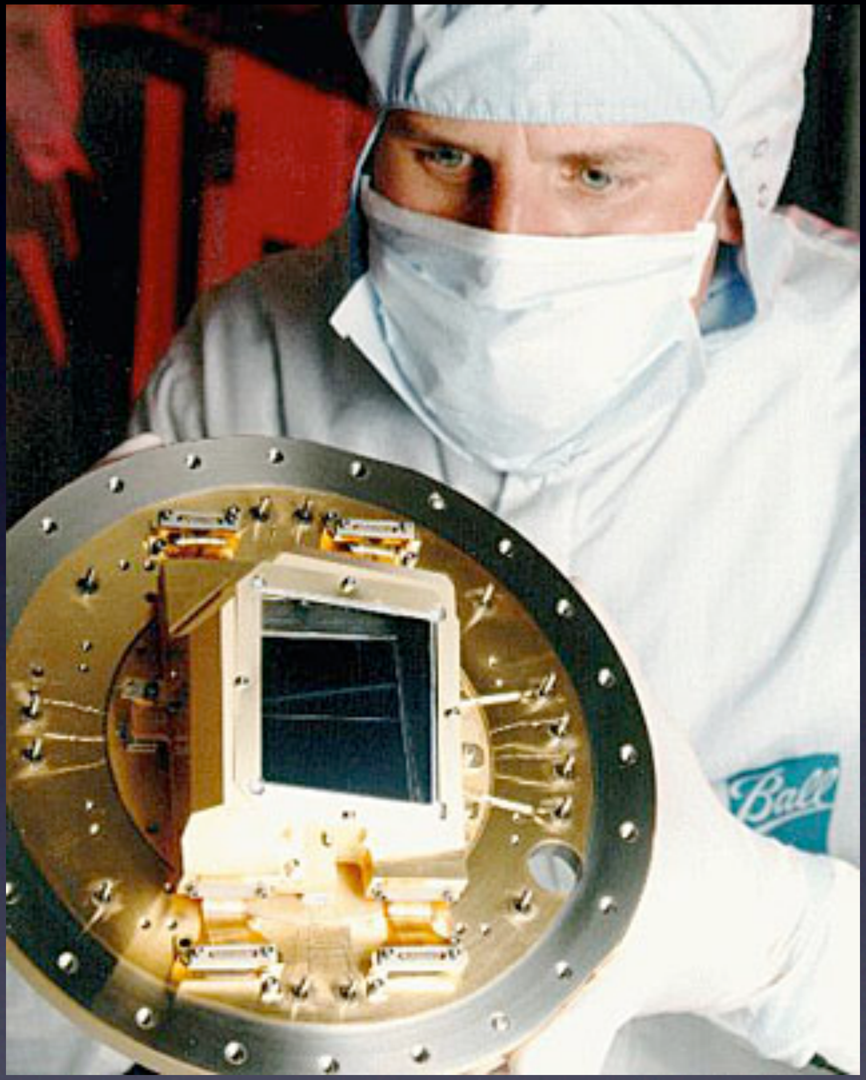


HST - ACS camea

HST - ACS camea

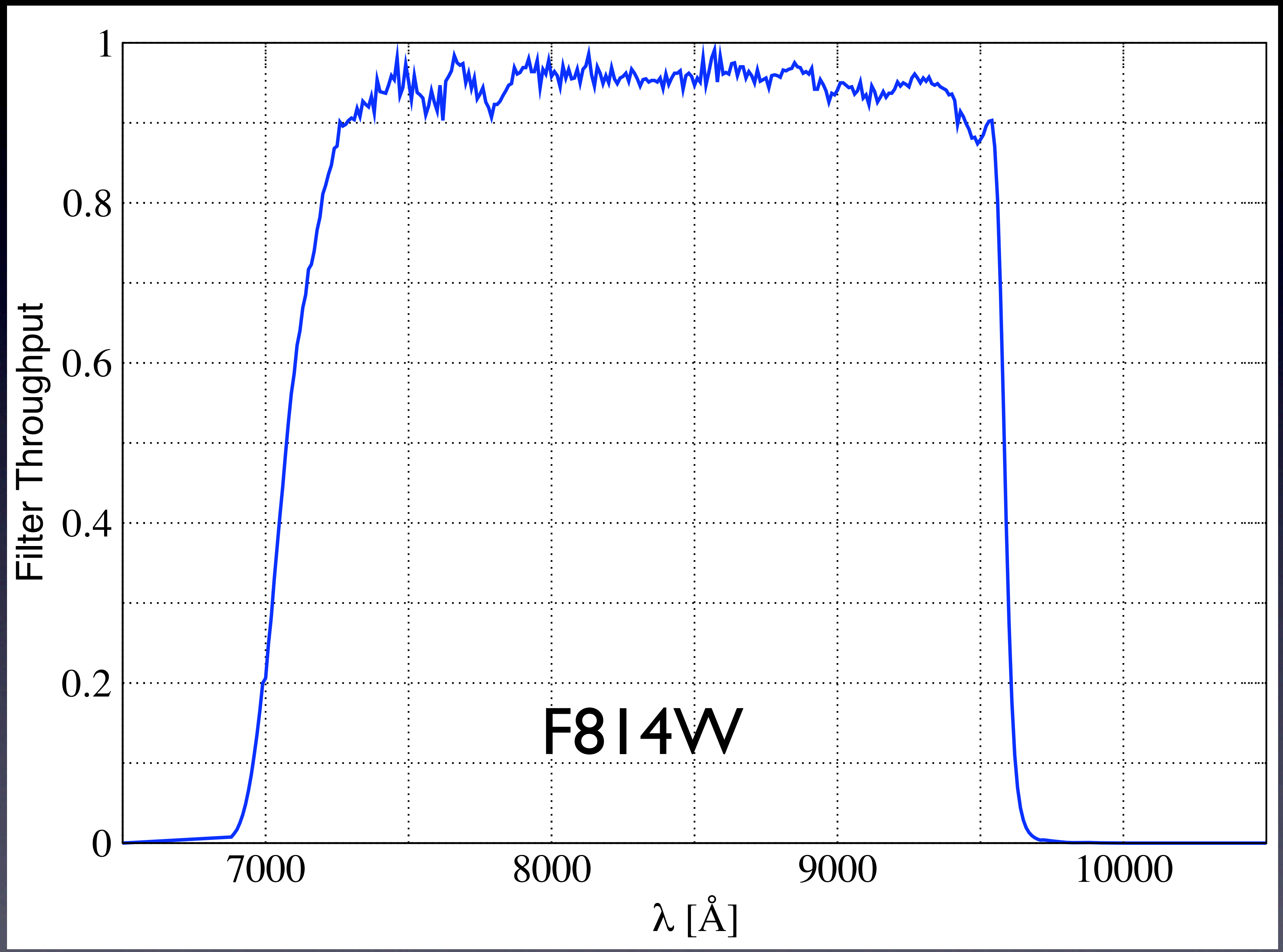


HST - ACS camea



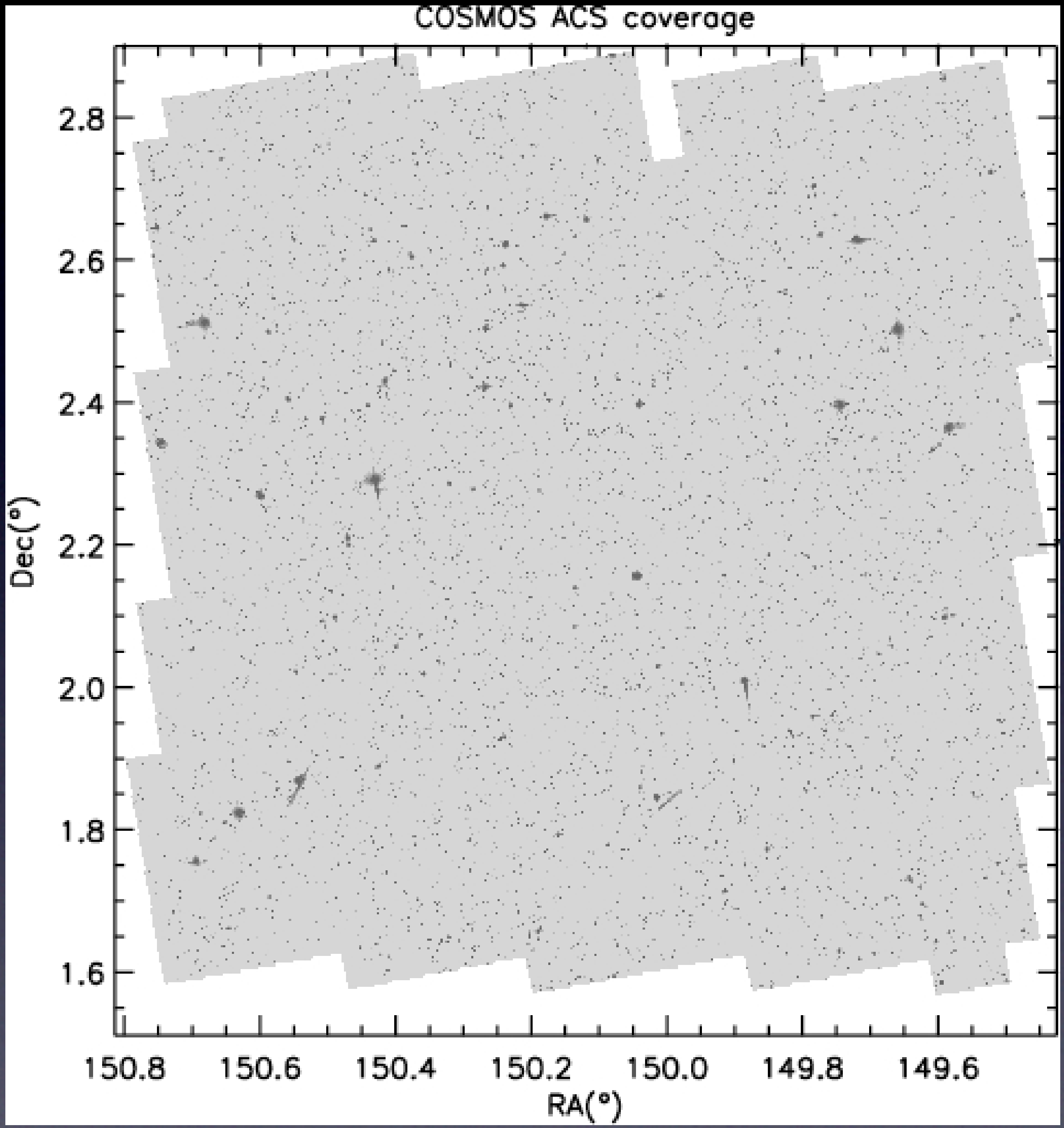
HST - ACS camea

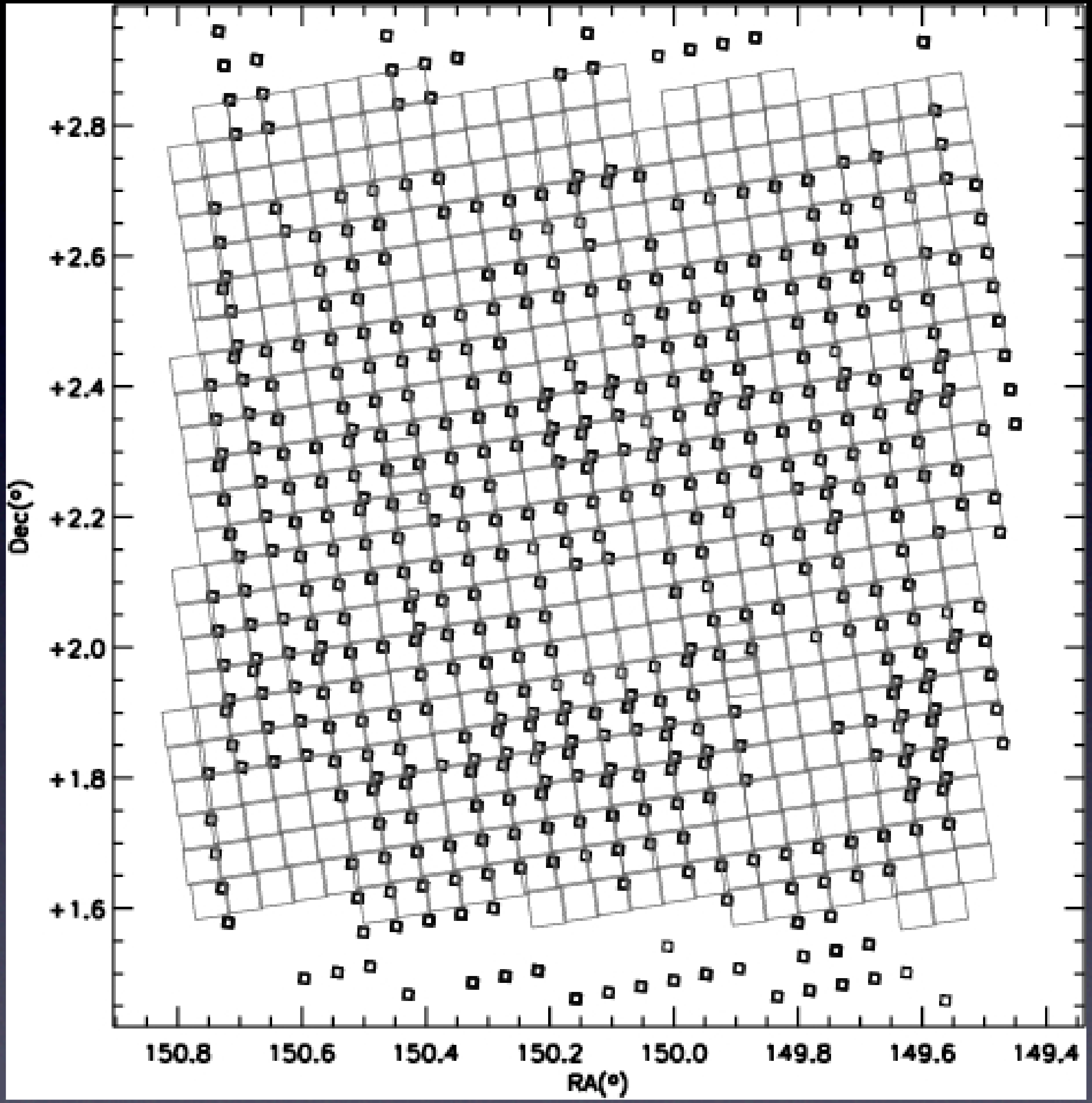
Visual ~ 4000 Å - 7000 Å



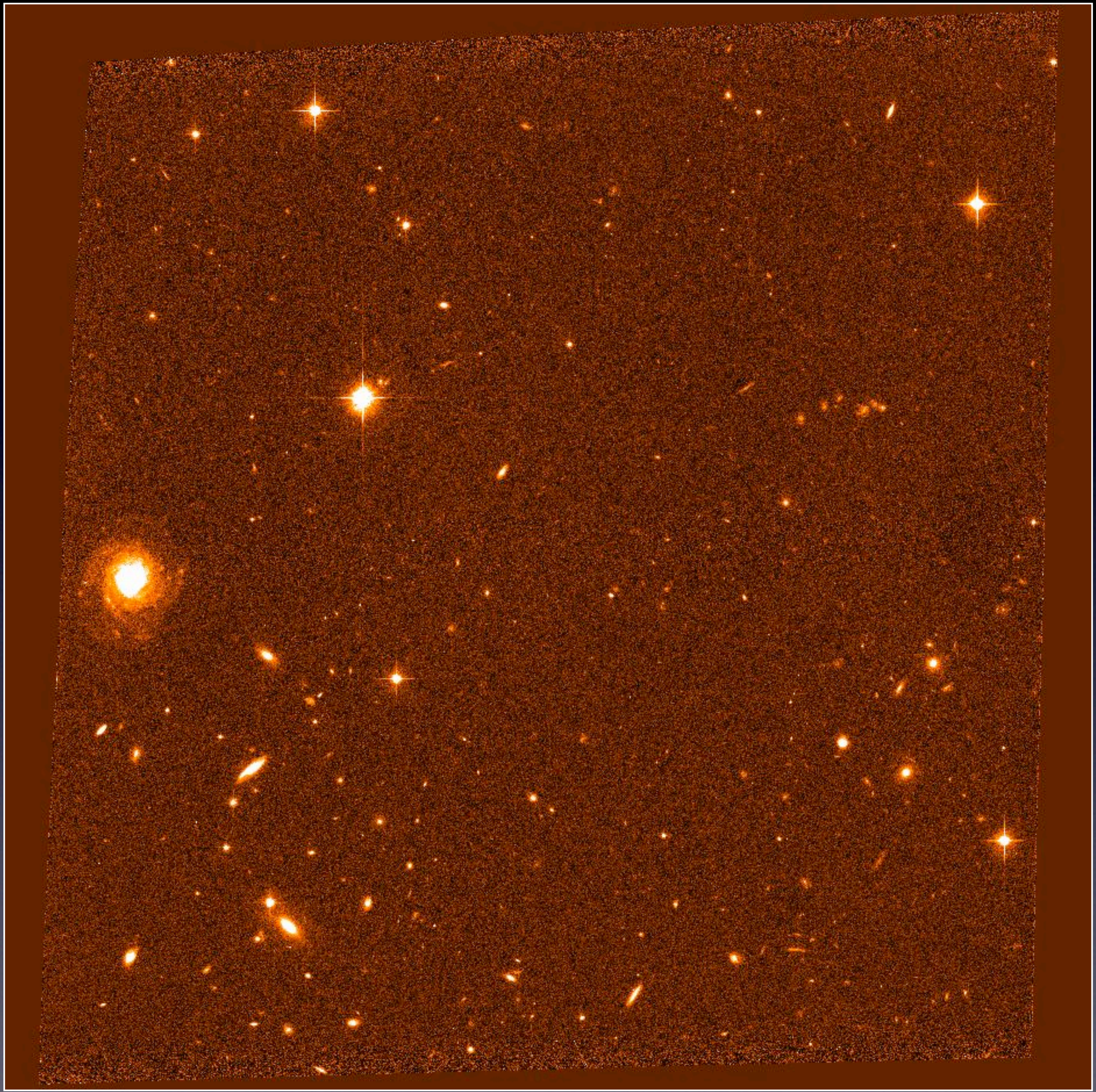
HST - ACS Data

HST - ACS Data

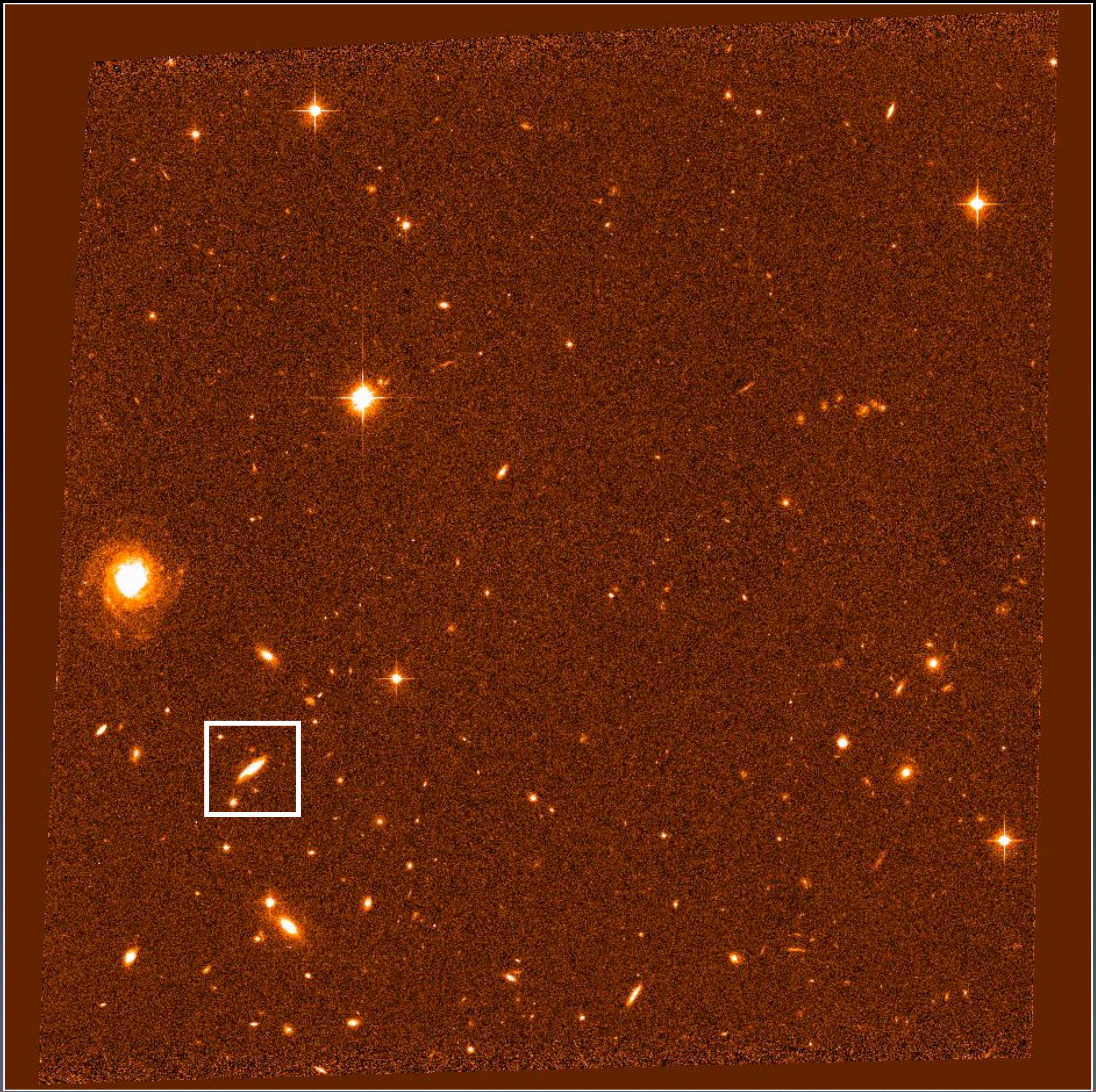

COSMOS ACS coverage
Dec(°)
2.8
2.6
2.4
2.2
2.0
1.8
1.6
150.8
150.6
150.4
150.2
150.0
149.8
149.6
RA(°)


Dec(°)
+2.8
+2.6
+2.4
+2.2
+2.0
+1.8
+1.6
150.8
150.6
150.4
150.2
150.0
149.8
149.6
149.4
RA(°)

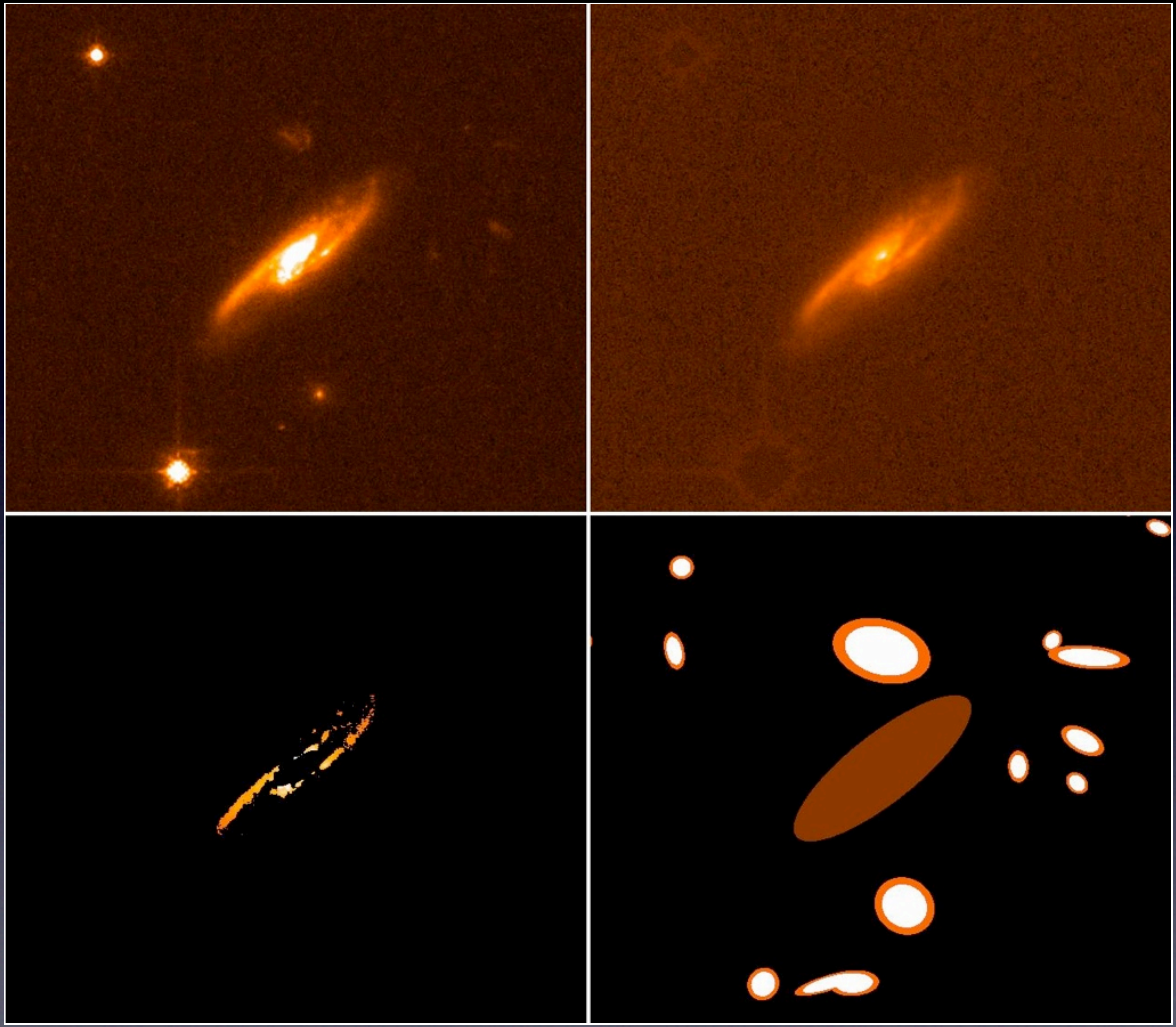
HST-ACS data: From the instrument to ZEST+



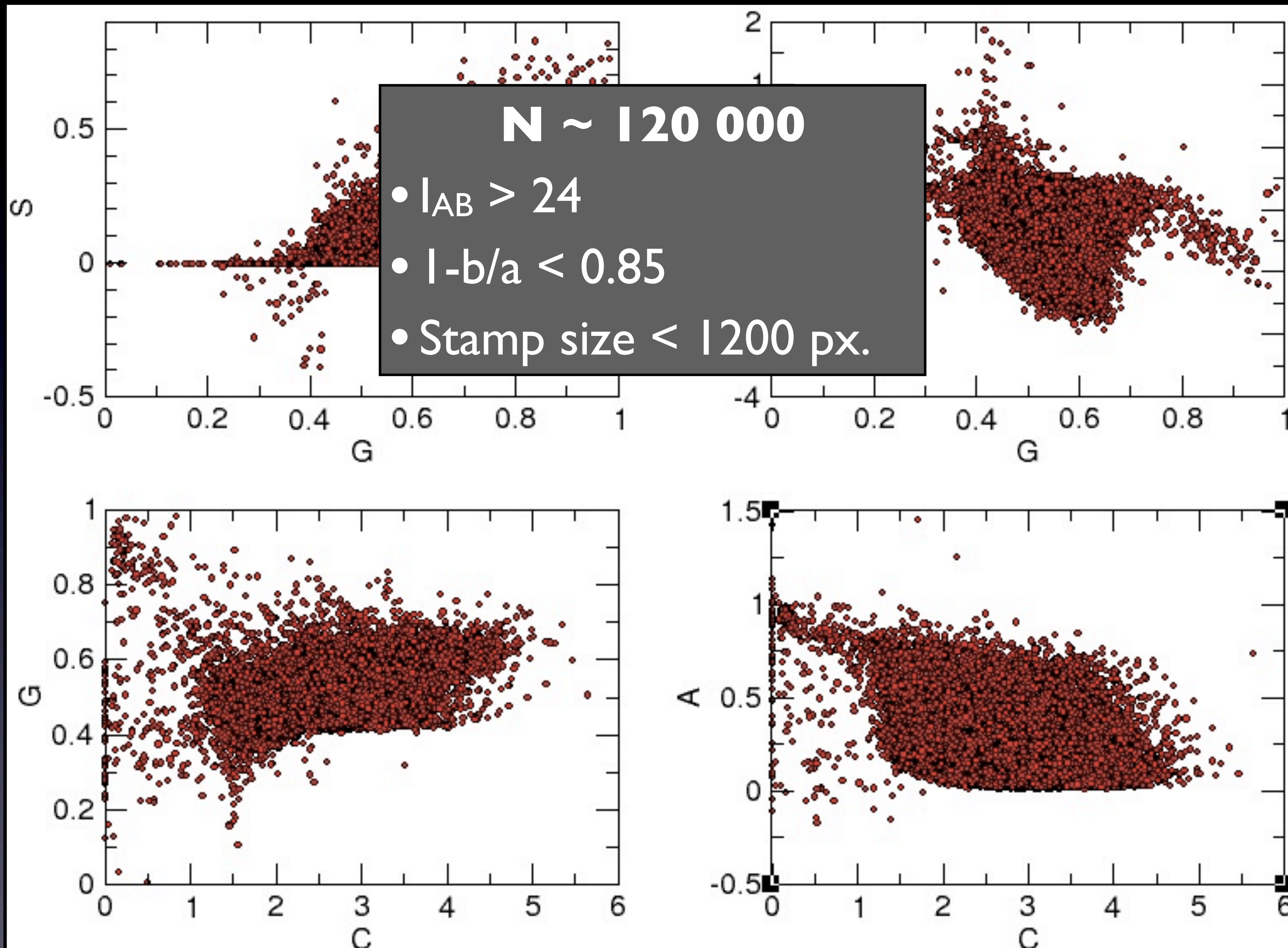
HST-ACS data: From the instrument to ZEST+



HST-ACS data: From the instrument to ZEST+



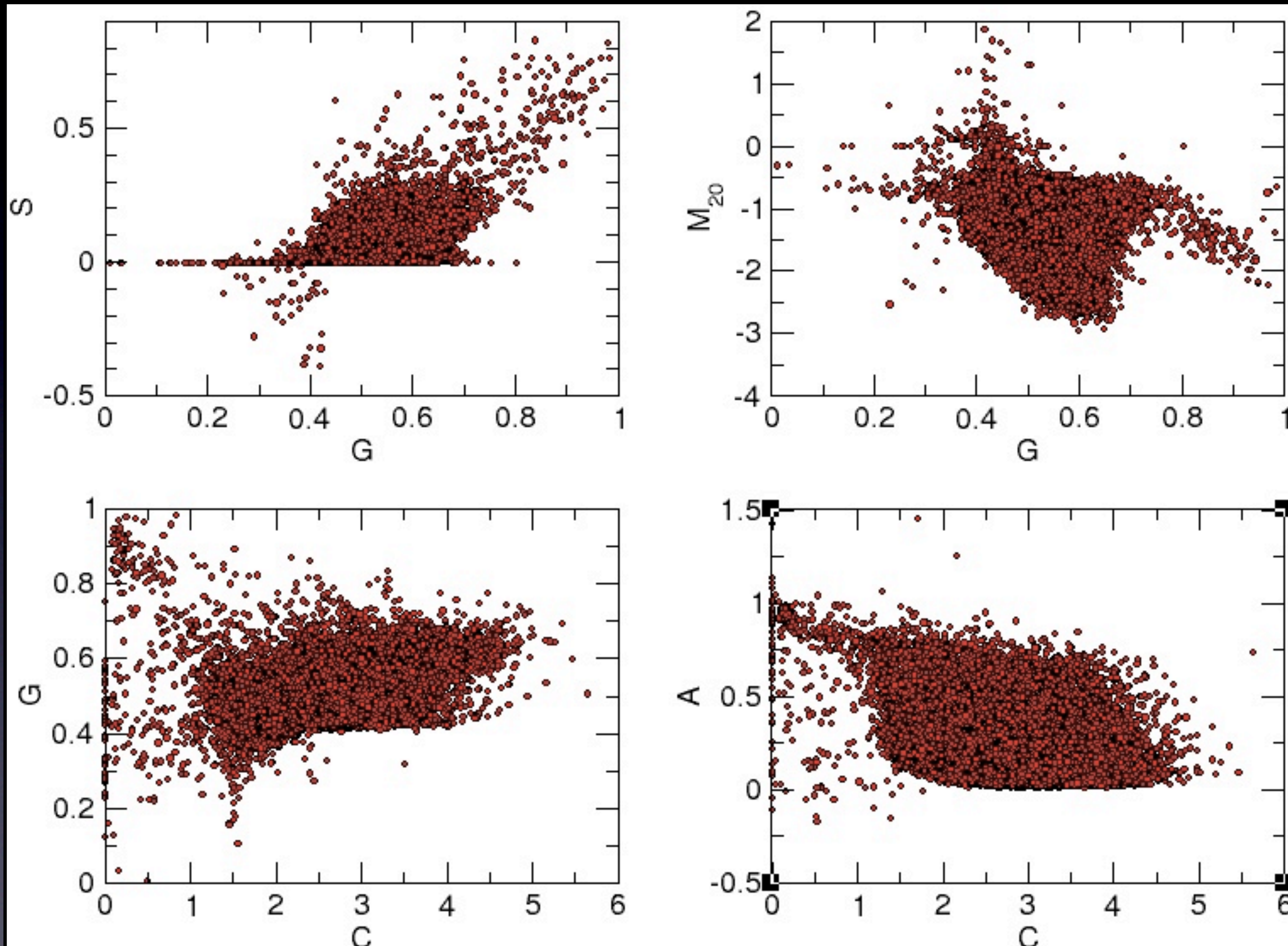
ZEST+ results



N ~ 120 000

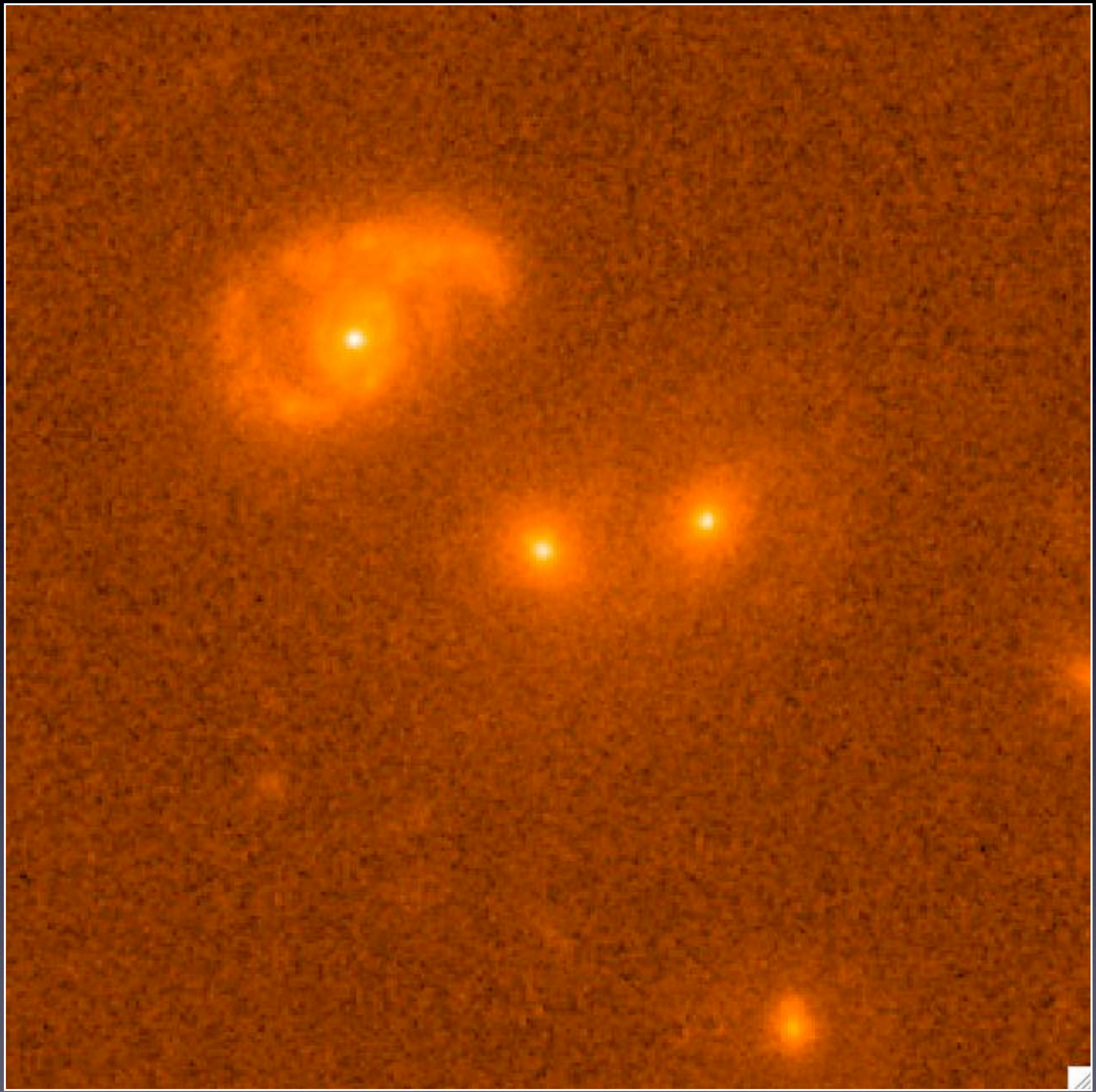
- $I_{AB} > 24$
- 1-b/a < 0.85
- Stamp size < 1200 px.

ZEST+ results



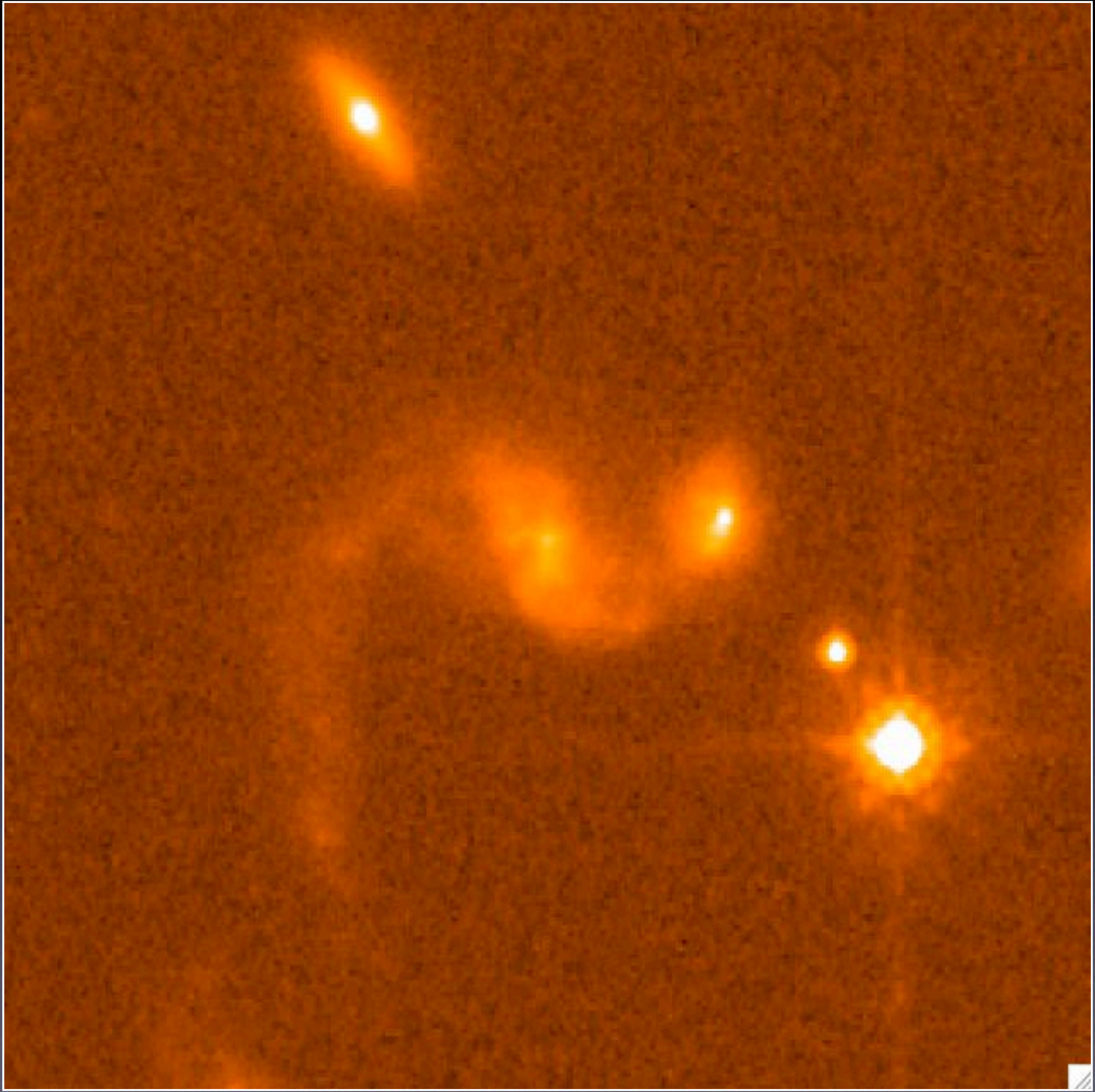
Challenges: Cleaning errors tolerance

Bright nearby objects



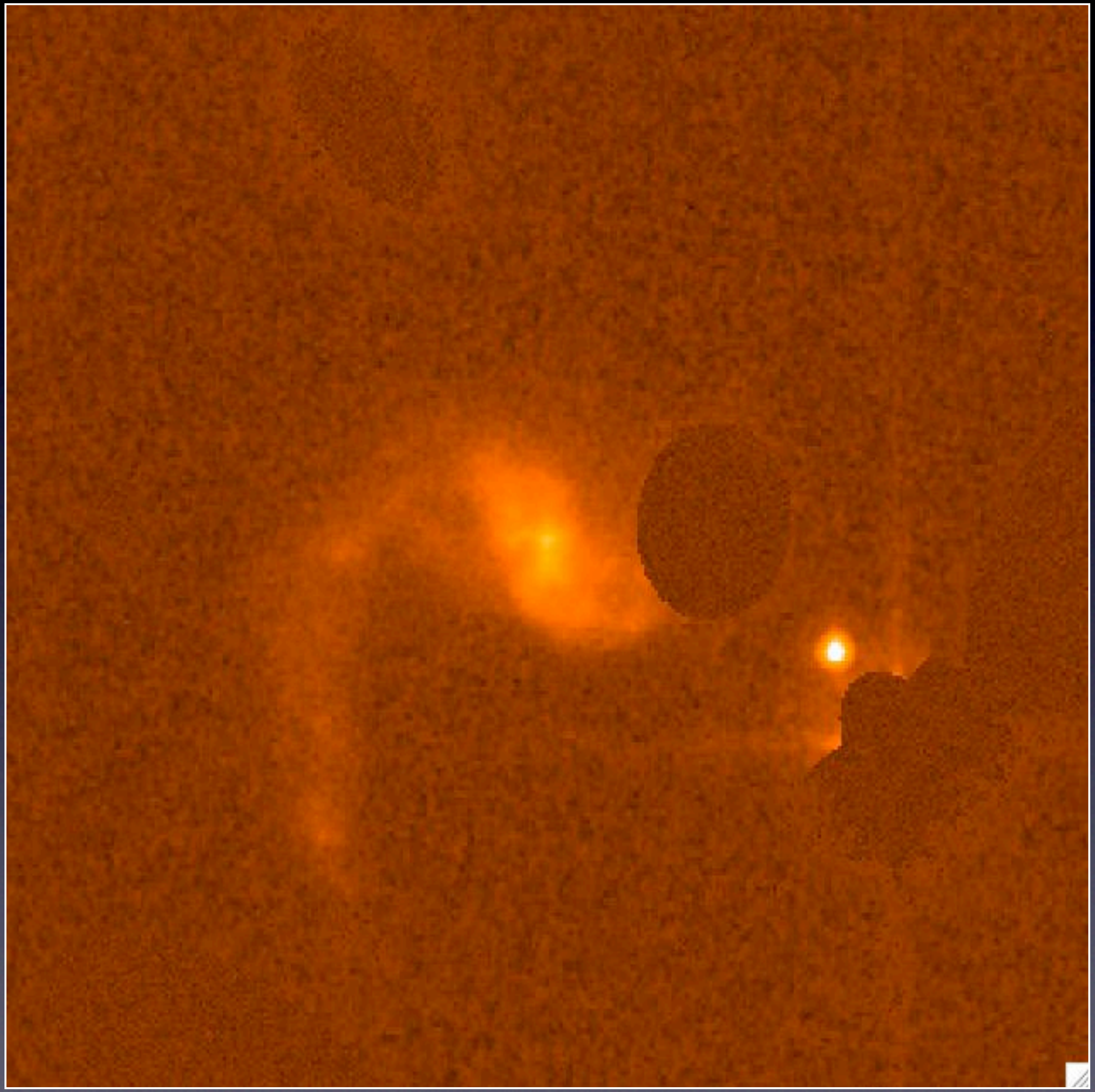
Challenges: Cleaning errors tolerance

Object truncation



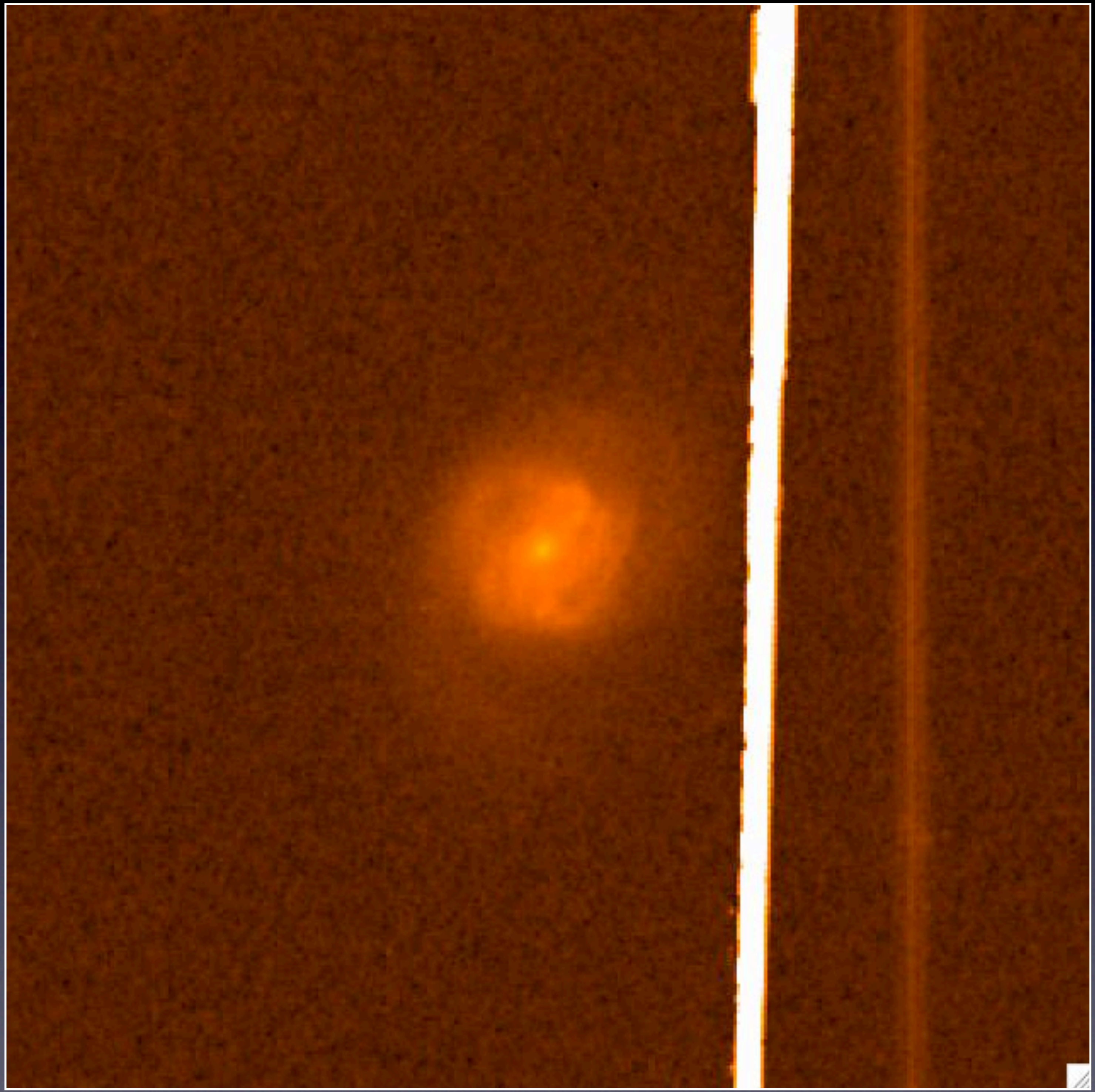
Challenges: Cleaning errors tolerance

Object truncation



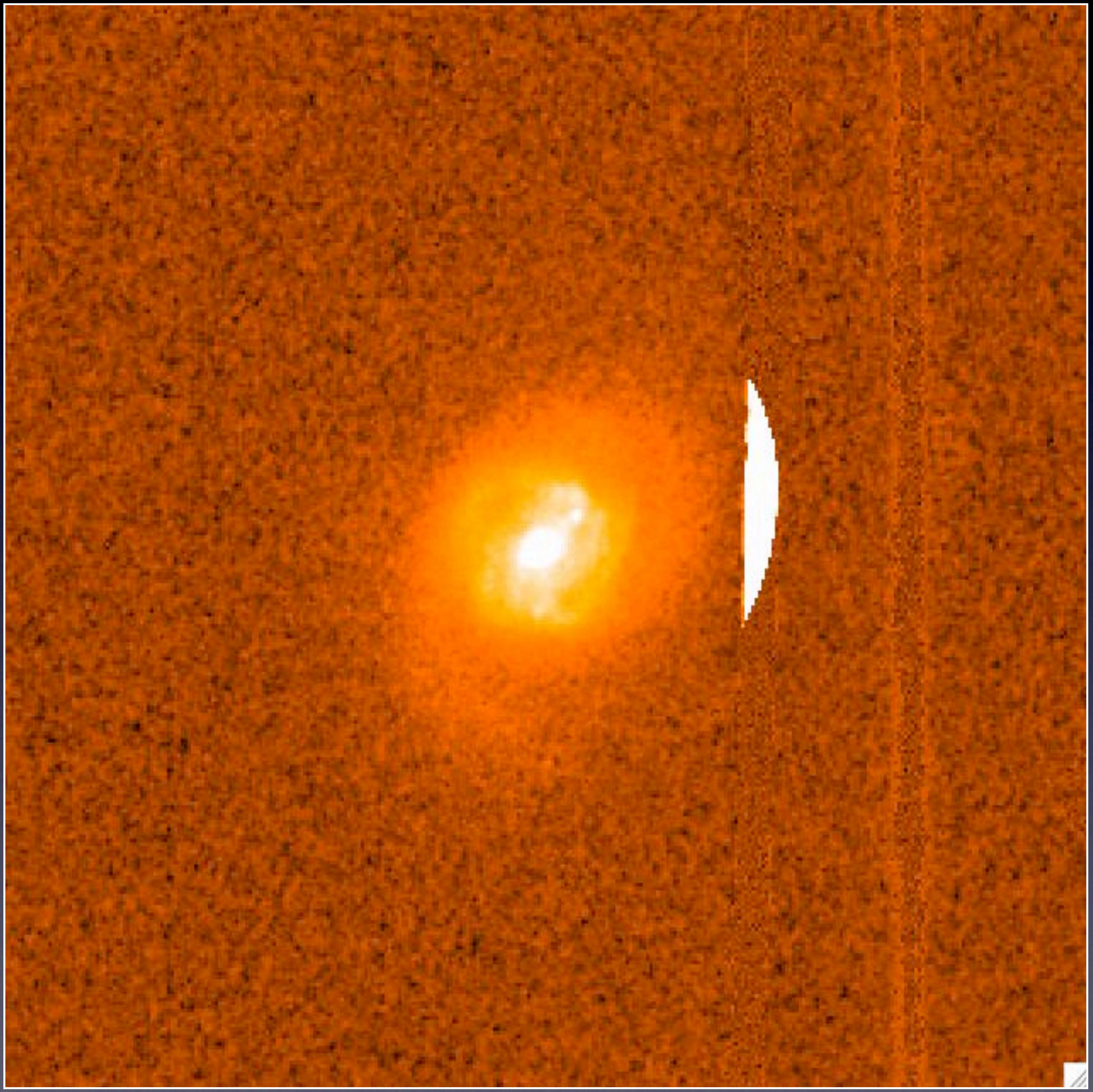
Challenges: Cleaning errors tolerance

Background features



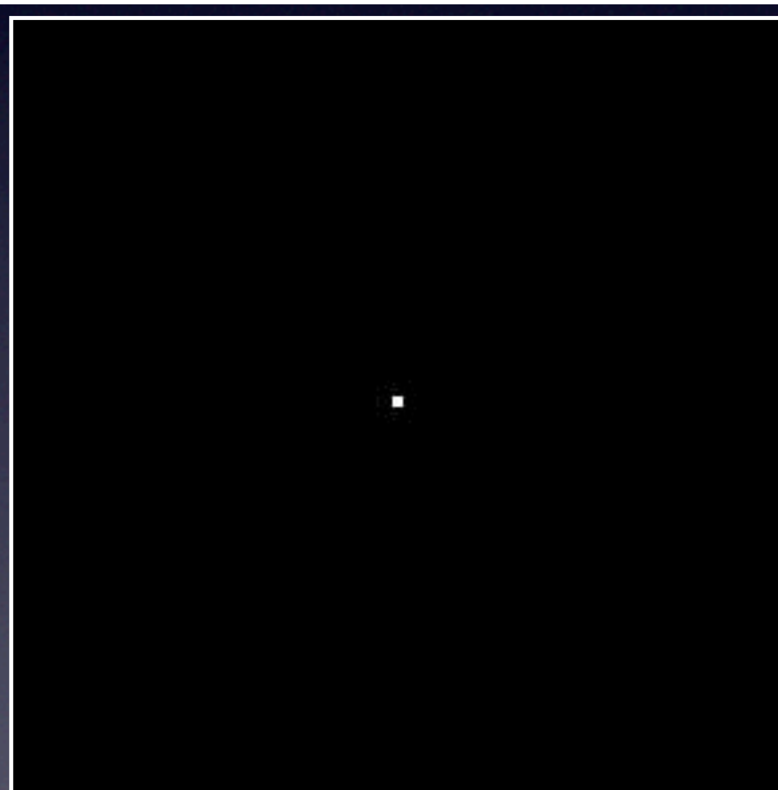
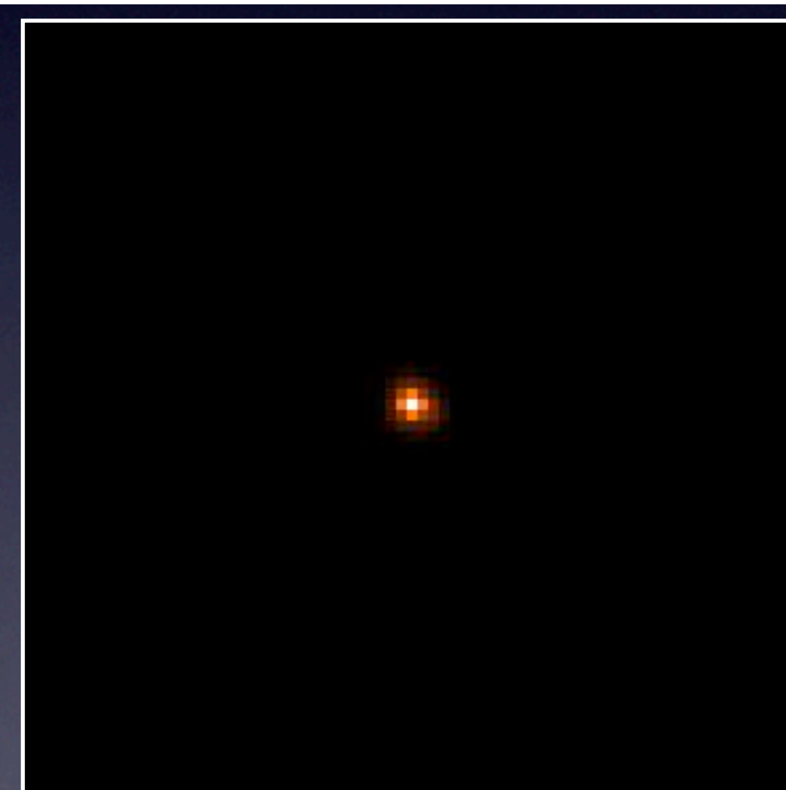
Challenges: Cleaning errors tolerance

Background features



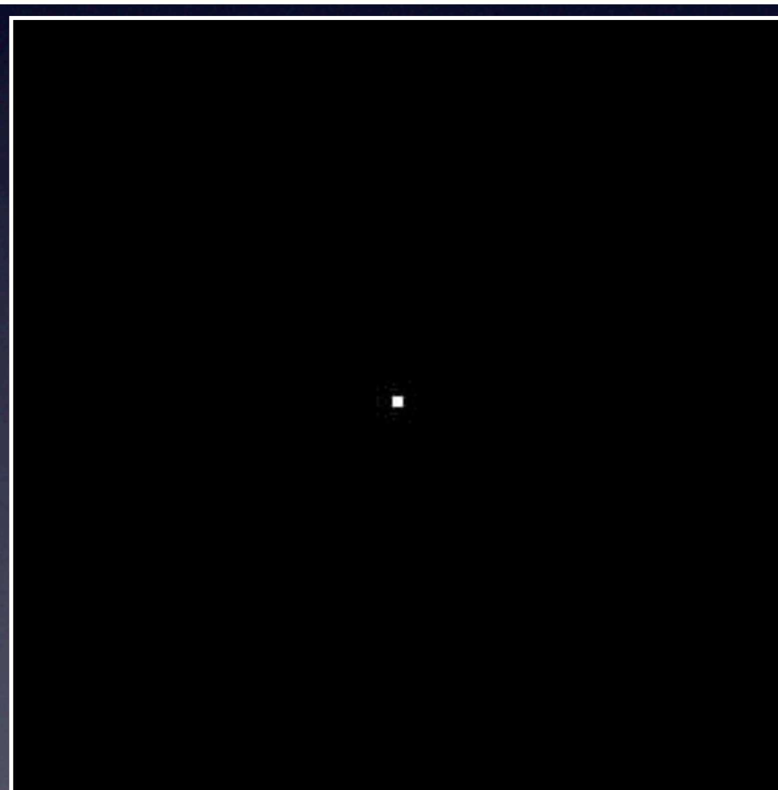
Challenges: Point Spread Function

$$\tilde{I}(x,y) = I(x,y) \star \mathcal{F}(x,y) + n(x,y)$$

No PSF	With PSF
	

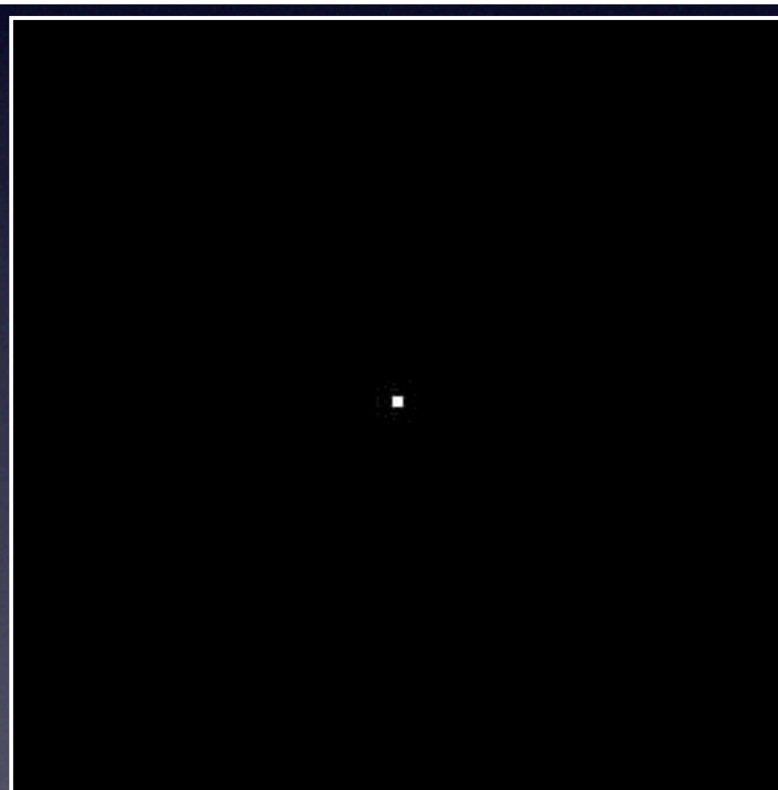
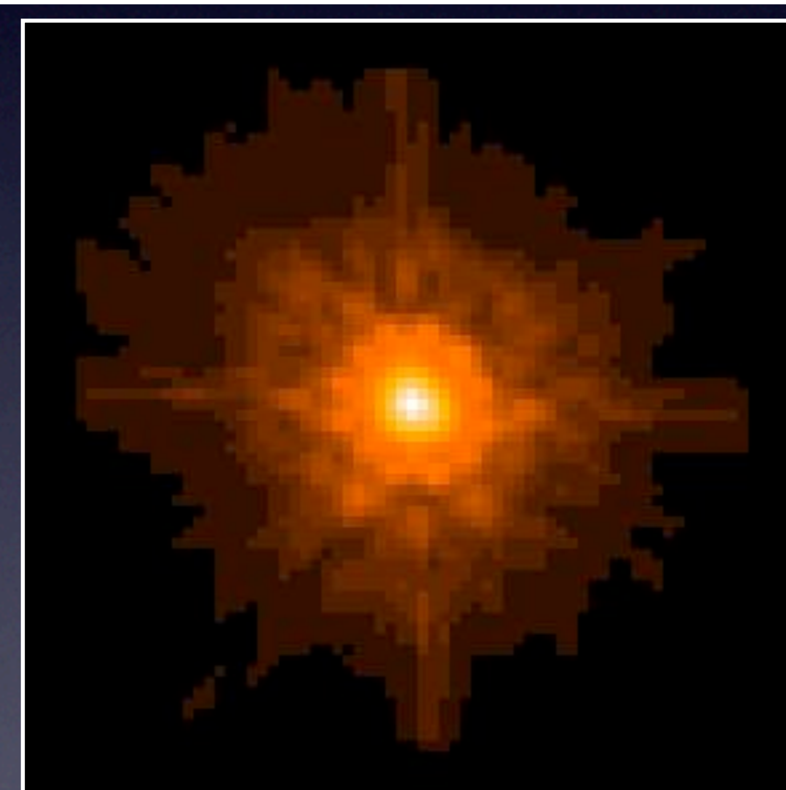
Challenges: Point Spread Function

$$\tilde{I}(x,y) = I(x,y) \star \mathcal{F}(x,y) + n(x,y)$$

No PSF	With PSF
	

Challenges: Point Spread Function

$$\tilde{I}(x,y) = I(x,y) \star \mathcal{F}(x,y) + n(x,y)$$

No PSF	With PSF
	

Challenges: Point Spread Function

Problem I: PSF effects

- Parametric Approach
- Richardson-Lucy Deconvolution
- Simulations

$$\tilde{I}(x,y) = I(x,y) \star \mathcal{F}(x,y) + n(x,y)$$

Problem II: PSF determination

- Analytic PSF
- PSF fitting
- HST: TinyTim
- Blind deconvolution

Conclusions

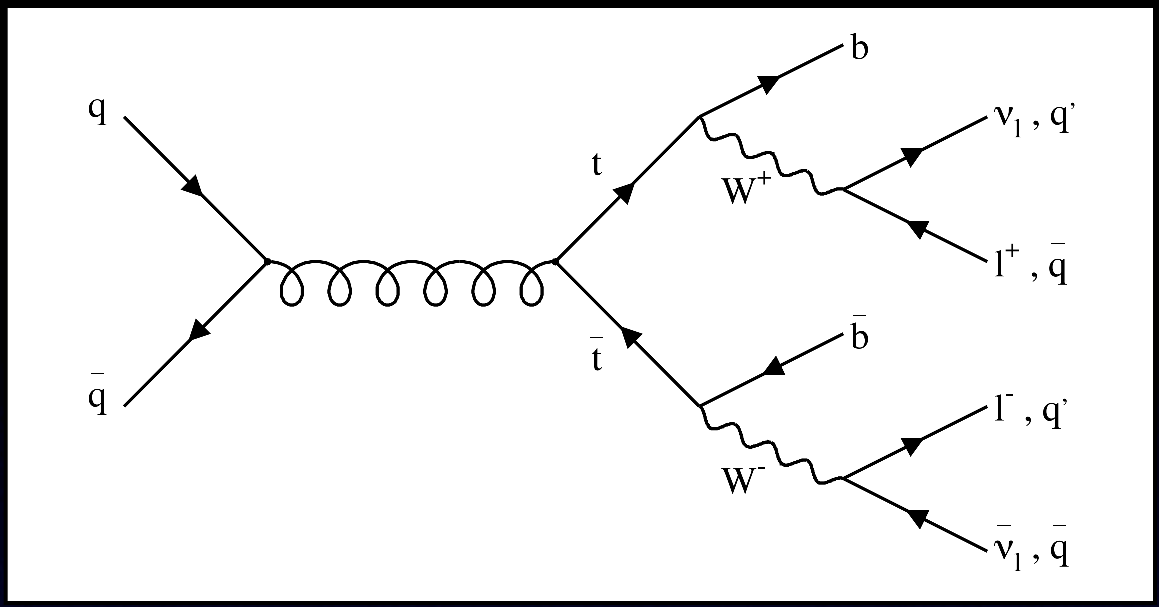
ZEST+:

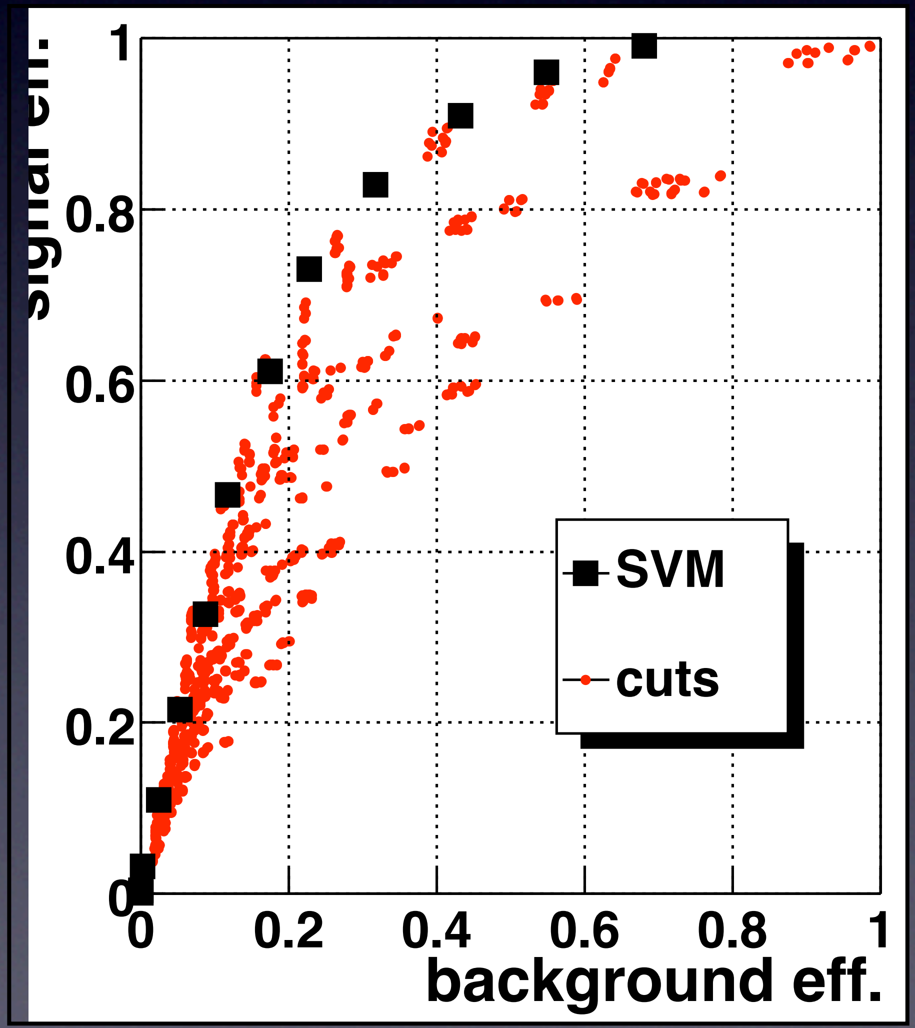
- uses a comprehensive set of non-parametric morphology descriptors.
- analyses substructure.
- uses a state-of-the-art classification algorithm.
- supports the use of ad-hoc external data.
- is flexible and fast.
- is fully documented and its source is freely available.
- is a portable, modular, standalone application written in C++.

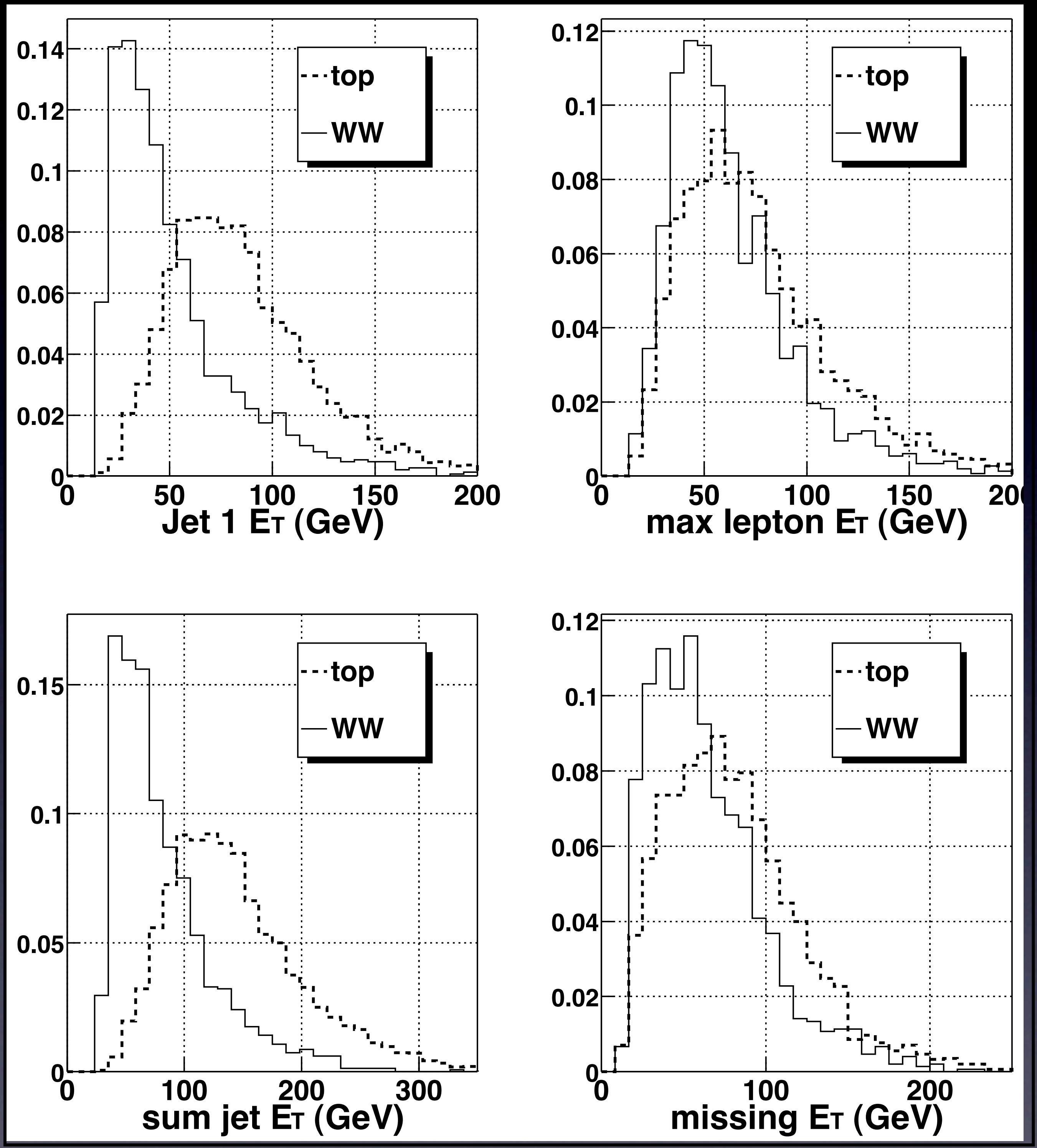
More information:

- *http://www.astro.phys.ethz.ch*
- *http://cosmos.astro.caltech.edu*
- *http://www.kernel-machines.org*
- *http://www.csie.ntu.edu/~cjlin/libsvm*

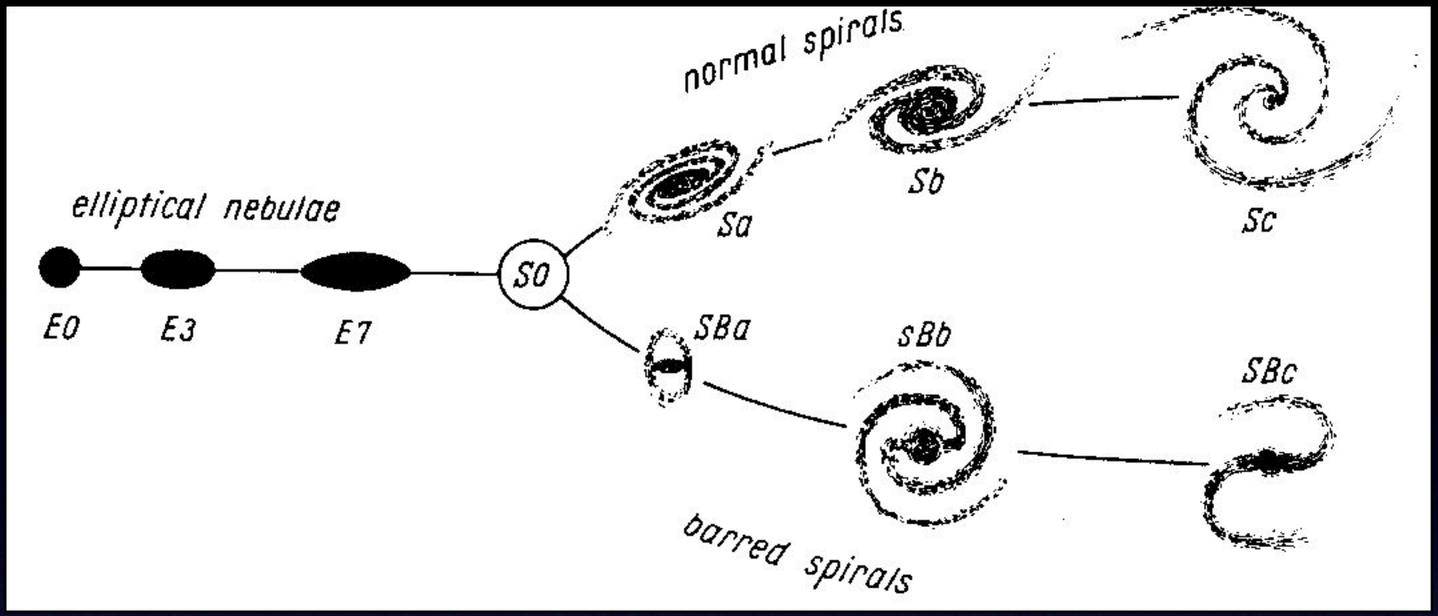
Top quark SVM analysis *(A. Vaiciulis, 2003)*





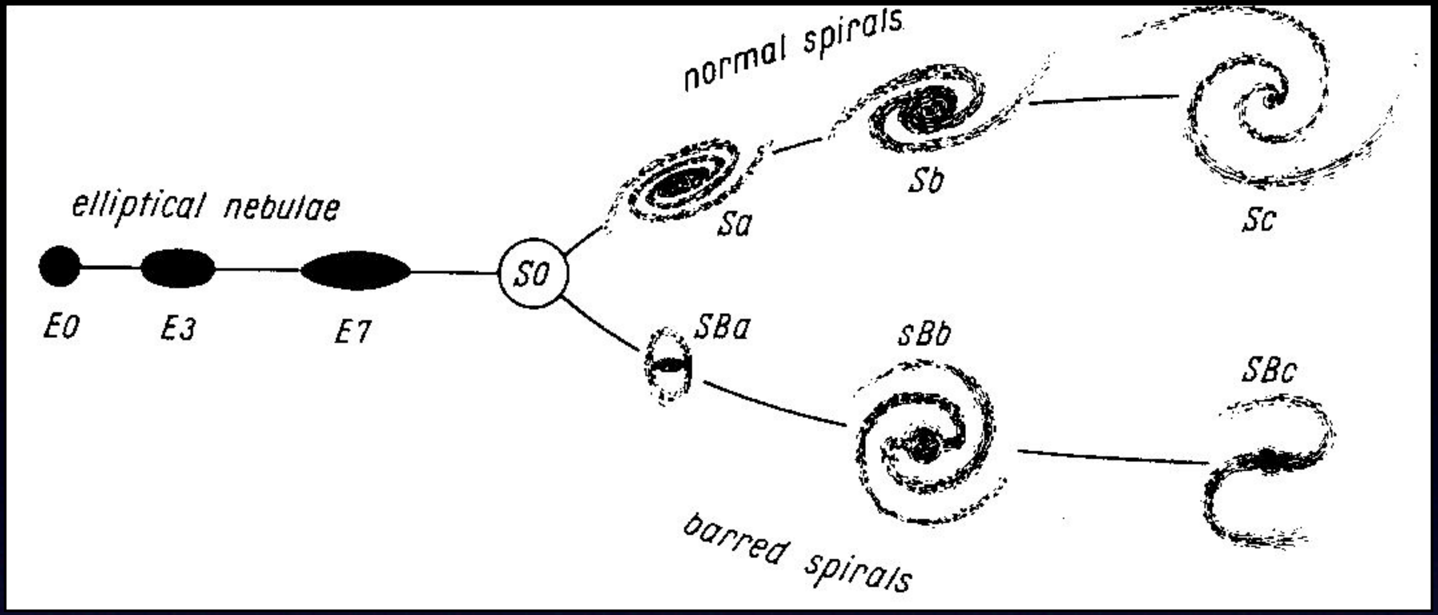


Revised Hubble system



- Narrowing of the bins
- Extension: Sd, Im, Am
- Arm attachment: (r) and (s)
- Rings

Revised Hubble system

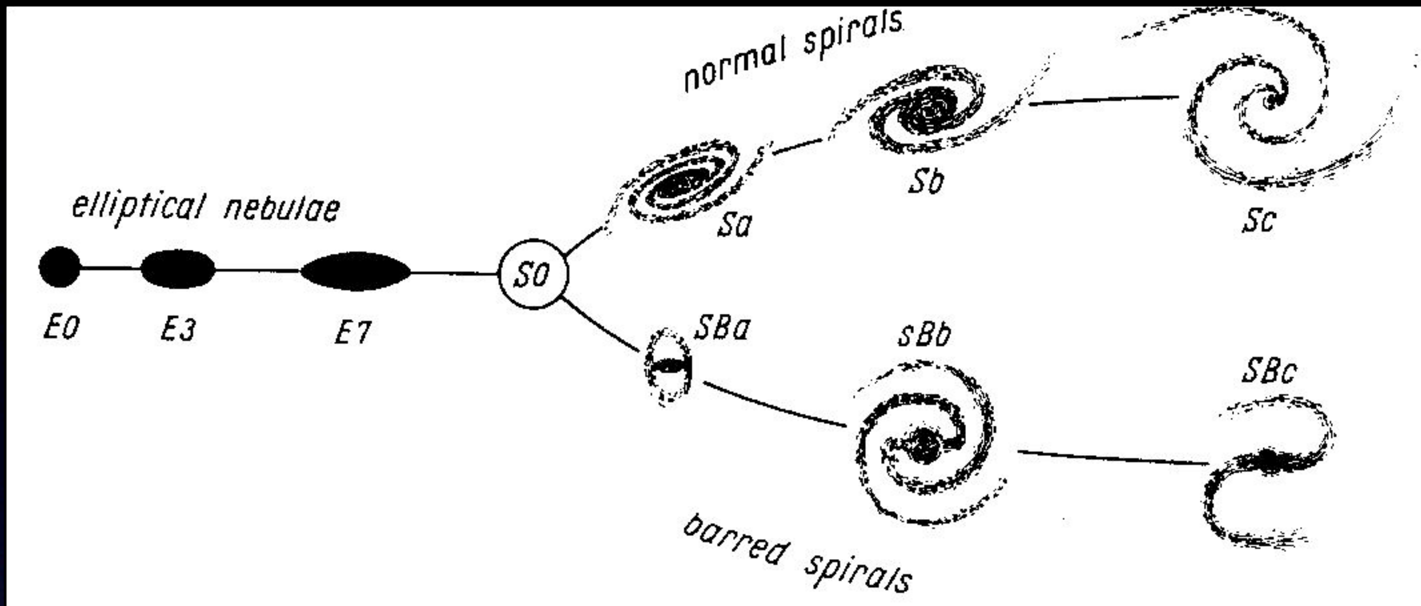


- Narrowing of the bins
- Extension: Sd, Im, Am
- Arm attachment: (r) and (s)
- Rings

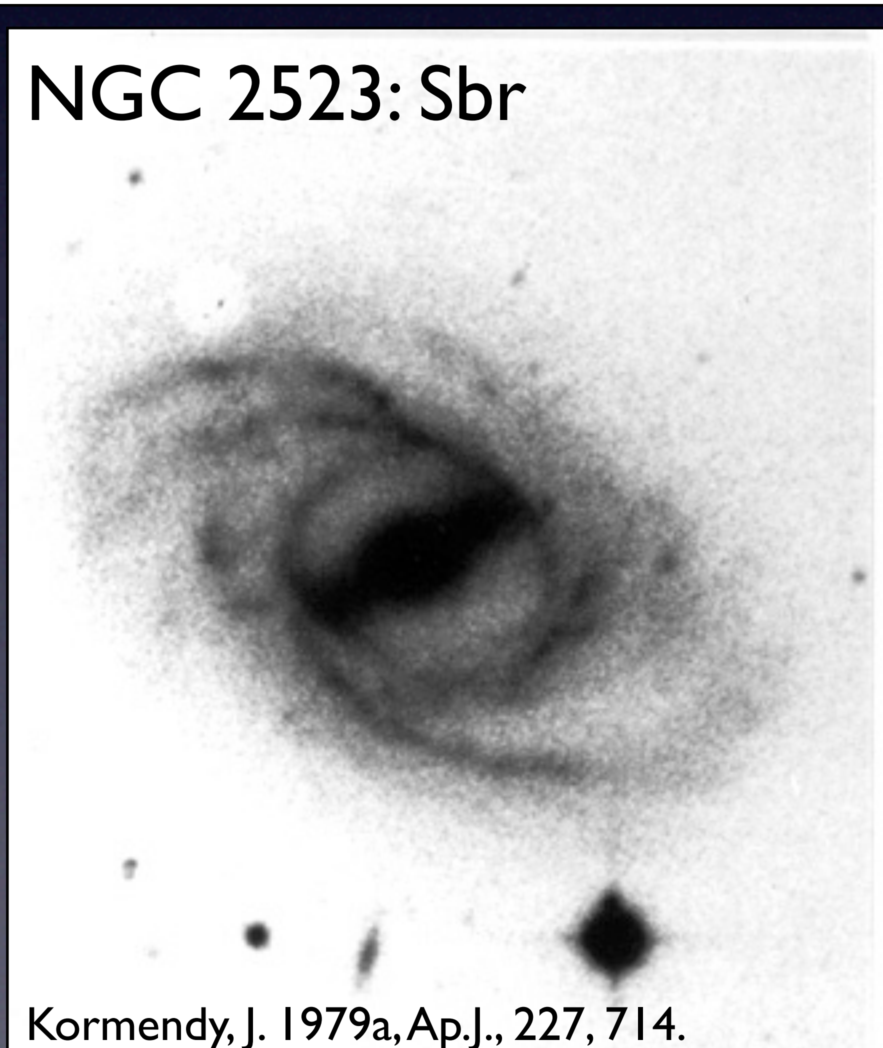
<u>Modern Notation</u>

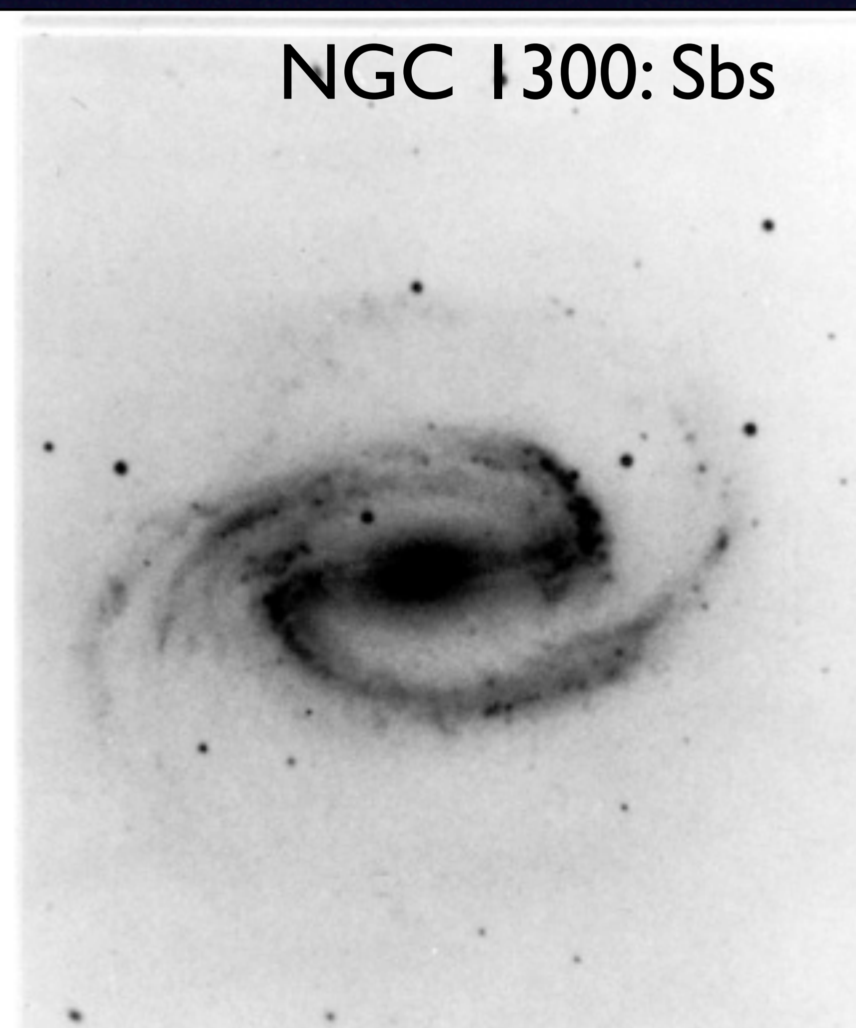
E - S0 - S0a - Sab - Sb - Sbc - Sc - Scd - Sd - Sd/m - Sm - Im

Revised Hubble system



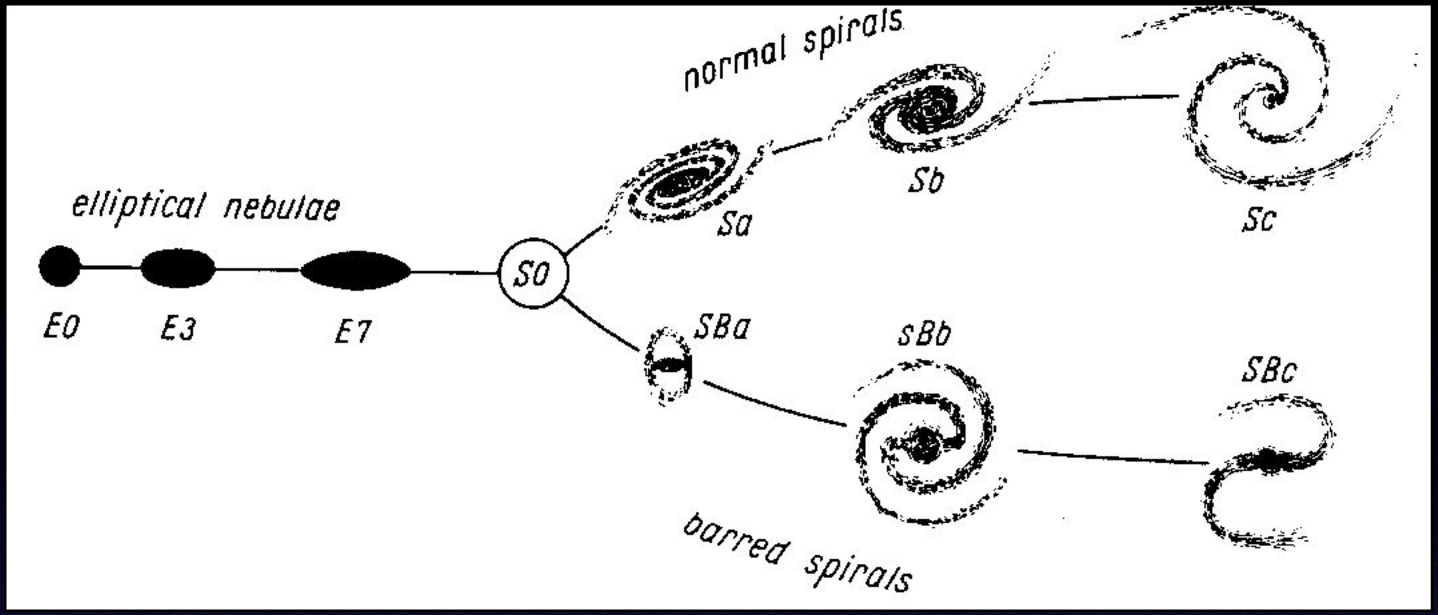
- •Narrowing of the bins
- •Extension: Sd, Im, Am
- •Arm attachment: (r) and (s)
- •Rings



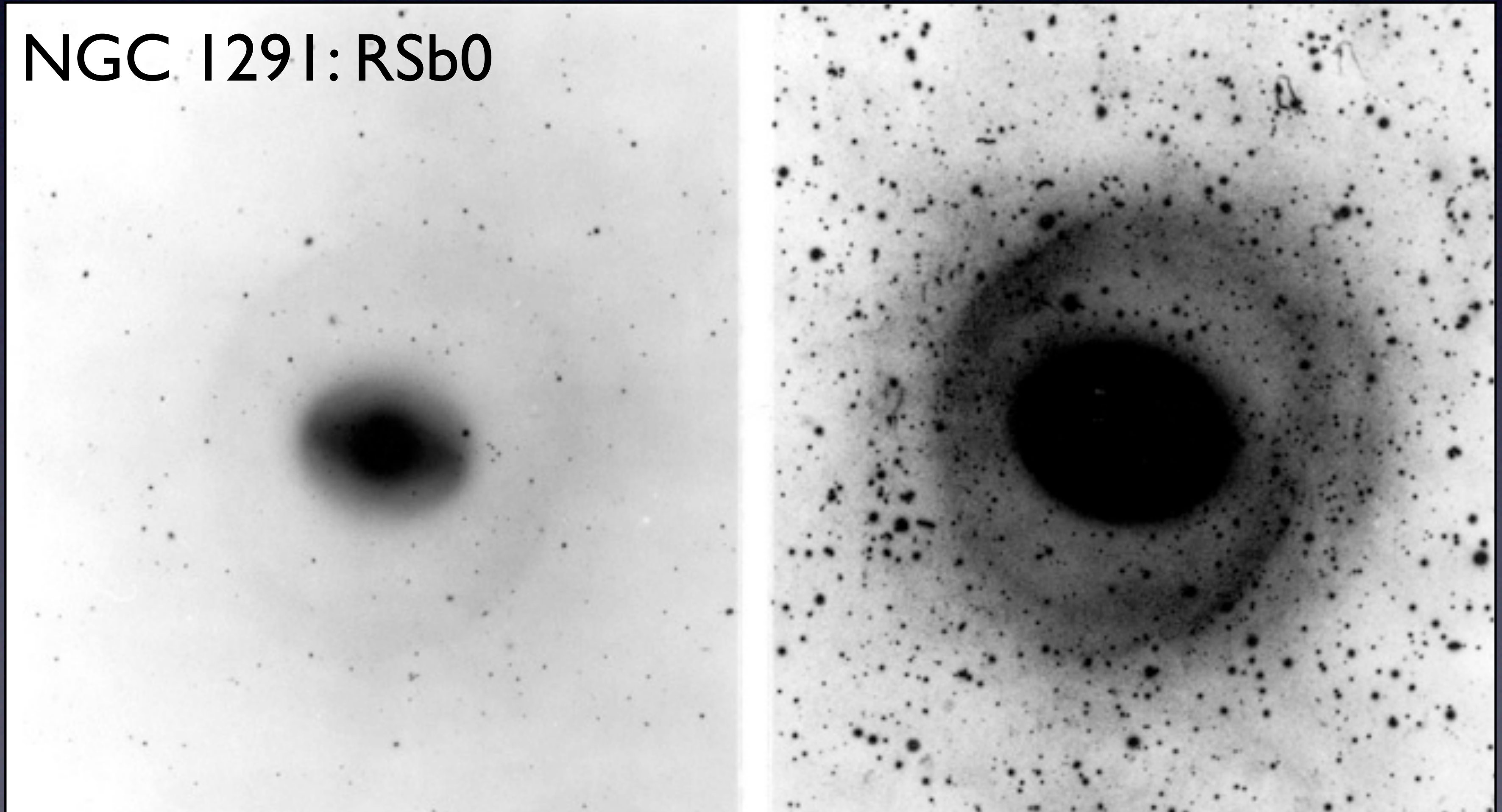


Kormendy, J. 1979a, Ap.J., 227, 714.

Revised Hubble system



- Narrowing of the bins
- Extension: Sd, Im, Am
- Arm attachment: (r) and (s)
- Rings

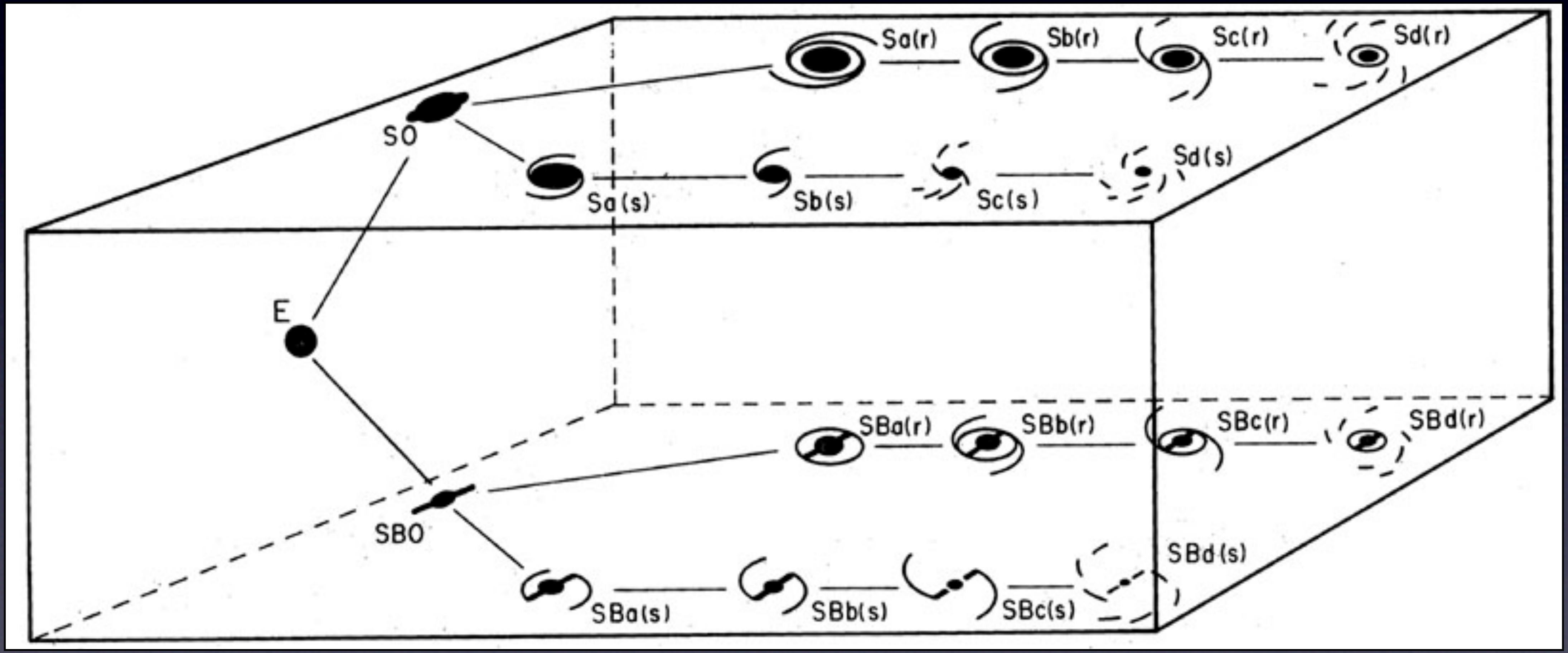


Kormendy, J. 1979a, Ap.J., 227, 714.

Visualization of the revised system

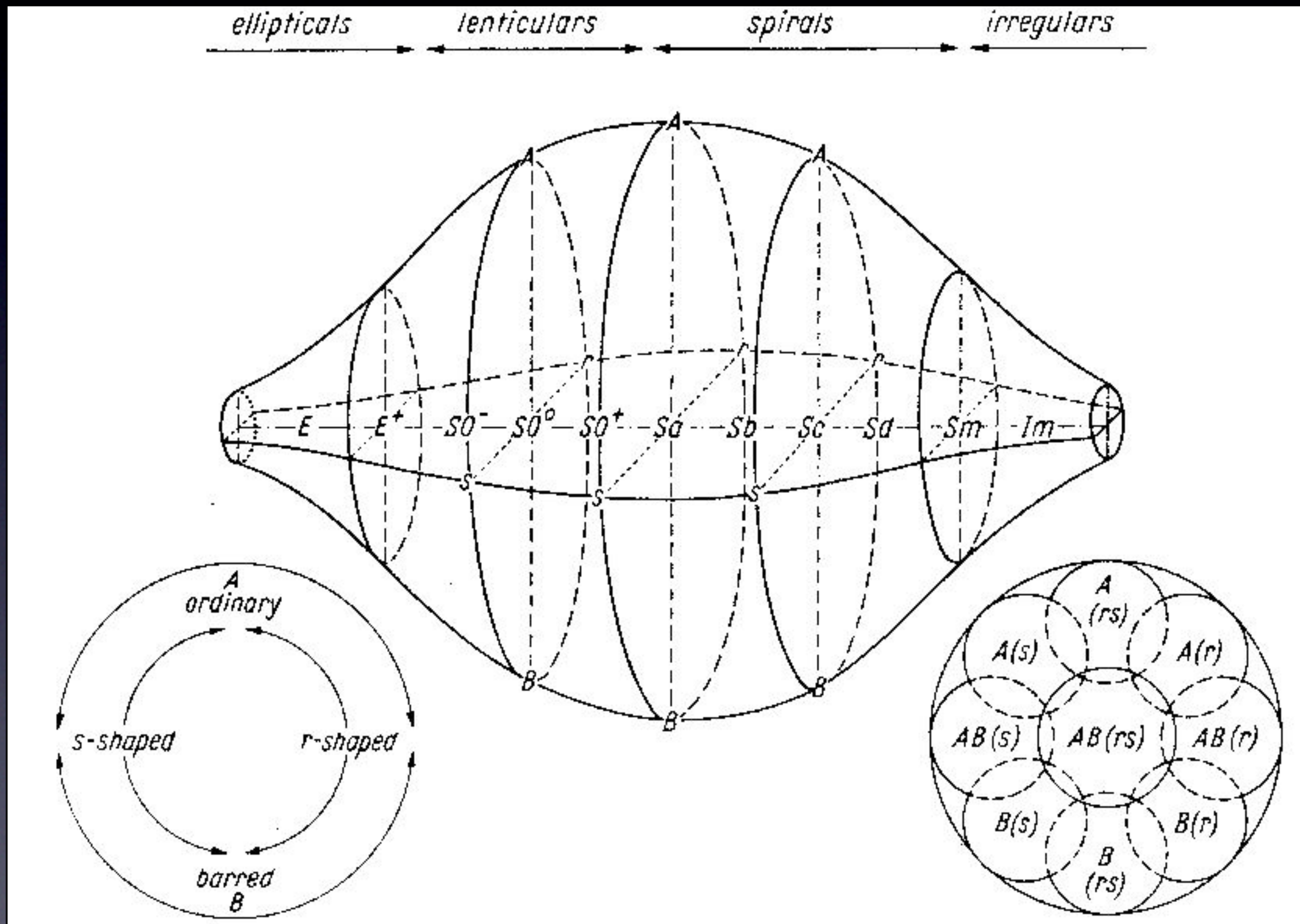
Visualization of the revised system

Hodge (1966)



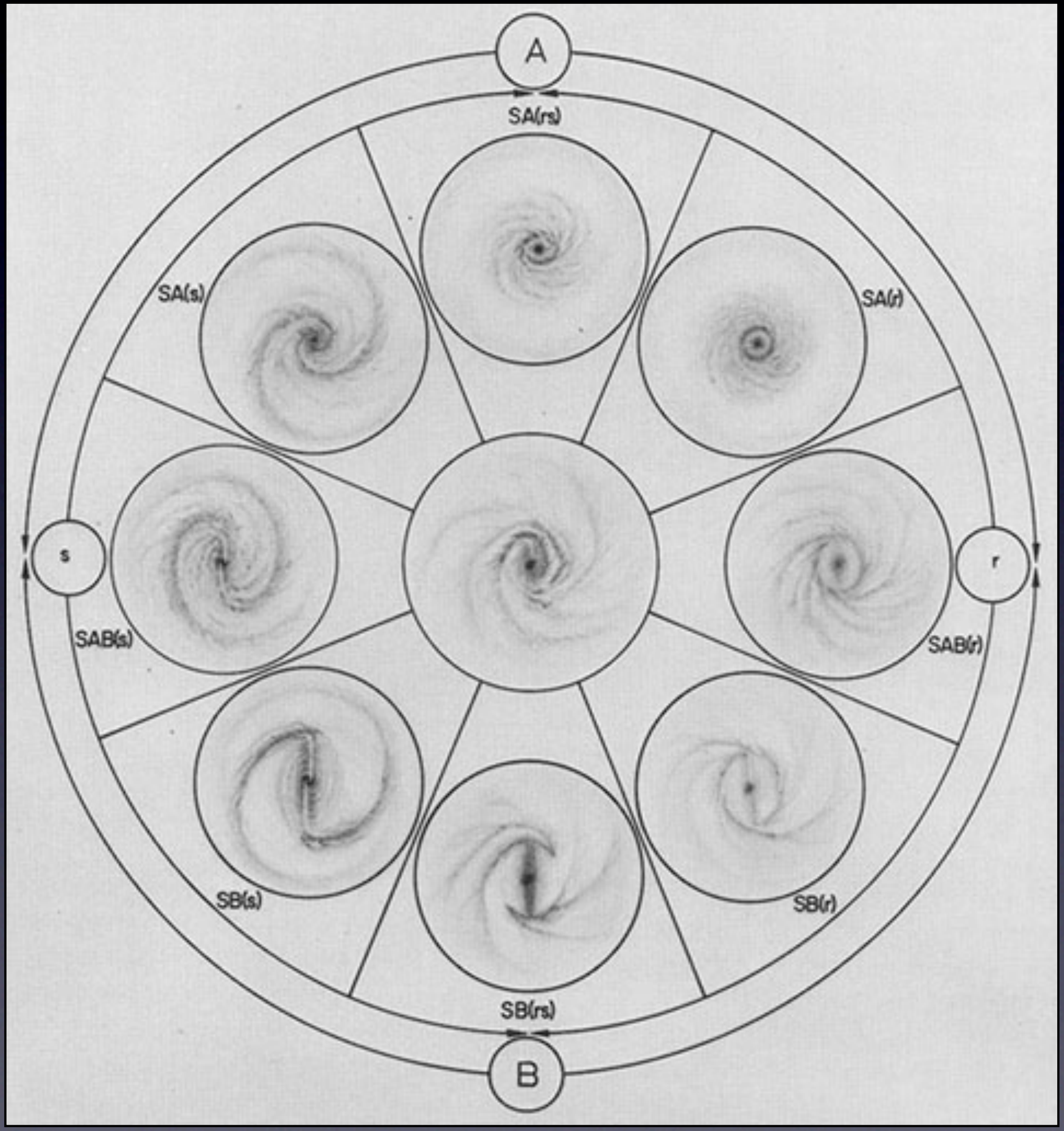
Visualization of the revised system

de Vaucouleurs (1959)

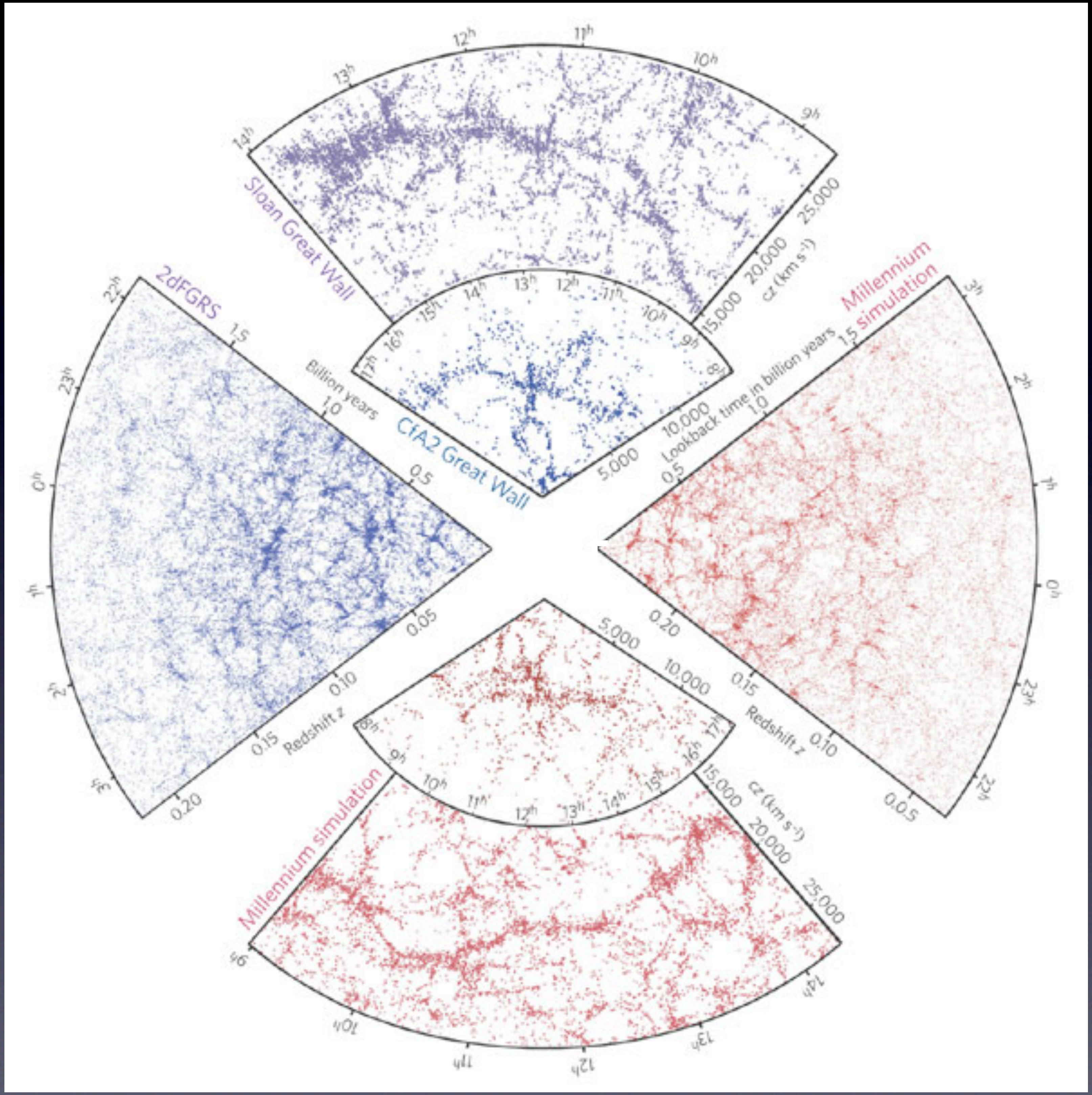


Visualization of the revised system

de Vaucouleurs (1959)

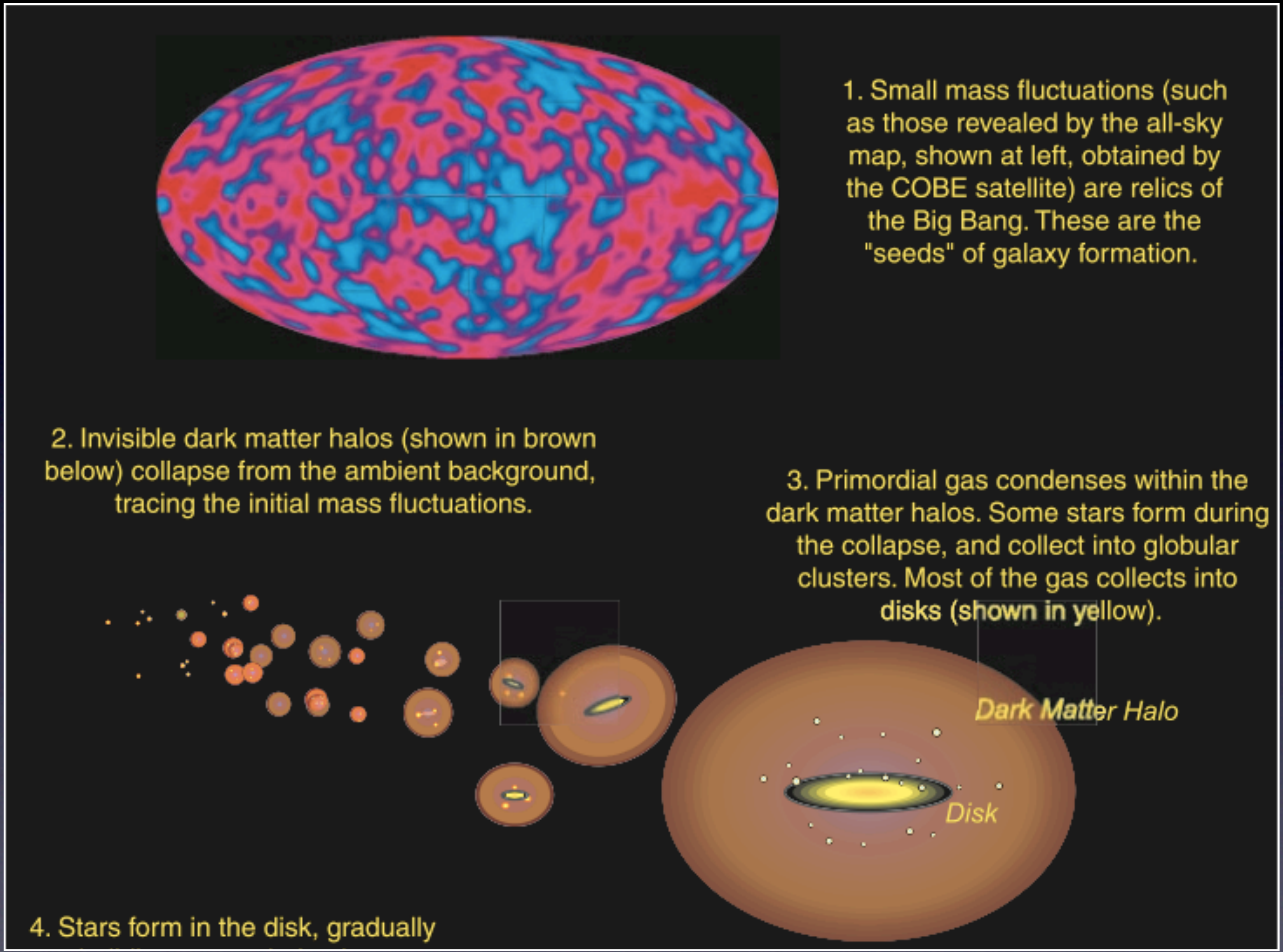


The Universe large scale structure



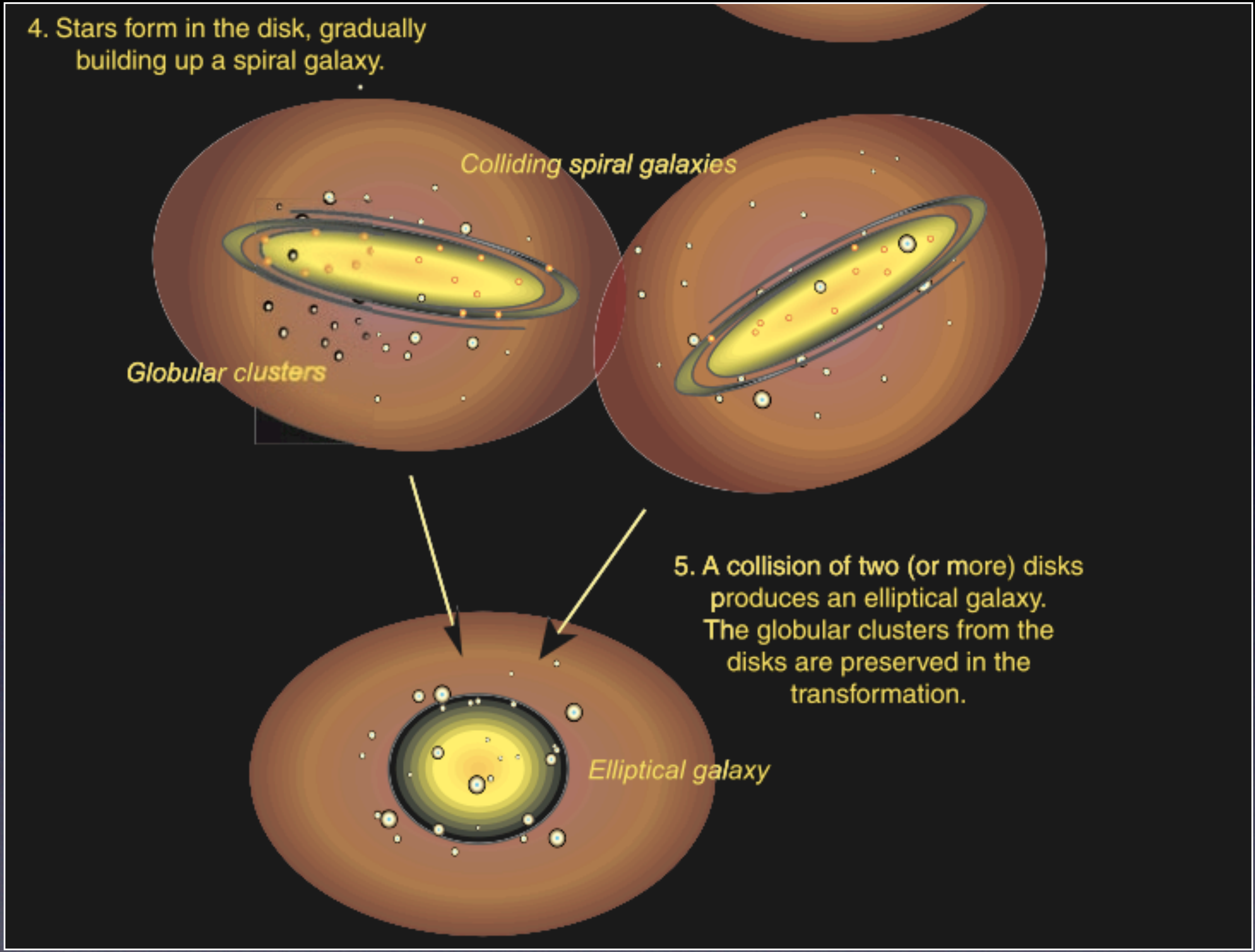
Volker Springel, Carlos S. Frenk and Simon D. M. White
Nature 440, 1137-1144 (27 April 2006)

Galaxies: The hierarchical formation model



R. G. Abraham, Science **293**, 1273 (2001)

Galaxies: The hierarchical formation model



R. G. Abraham, Science **293**, 1273 (2001)