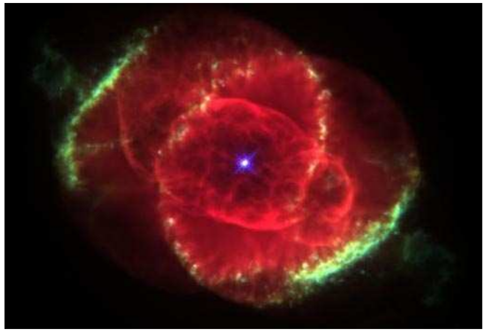
# Equivalence of Fluctuation Scale Dependence and Autocorrelations

Duncan Prindle

MIT Workshop

April, 2005

STAR



CENPA

Center for Experimental Nuclear Physics and Astrophysics

UNIVERSITY · OF · WASHINGTON · 1861 ·

# Agenda

- Fluctuations on binned spaces
- Covariances and Pearson’s coefficient
- Autocorrelation density ratio for $n$ and $p_\mathrm{t}$
- Autocorrelations directly from pair ratios
- Fluctuation/autocorrelation integral equation
- Inverting the integral equation
- Autocorrelations *vs* conditional distributions

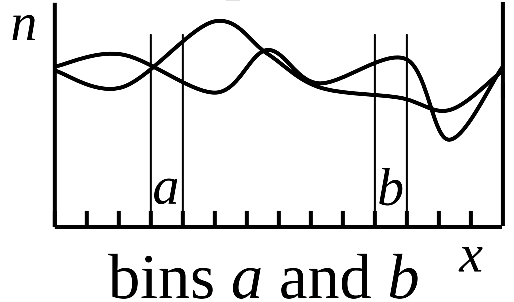
# Fluctuations on Binned Spaces

mixed-pair reference

bin contents: particle count *n*

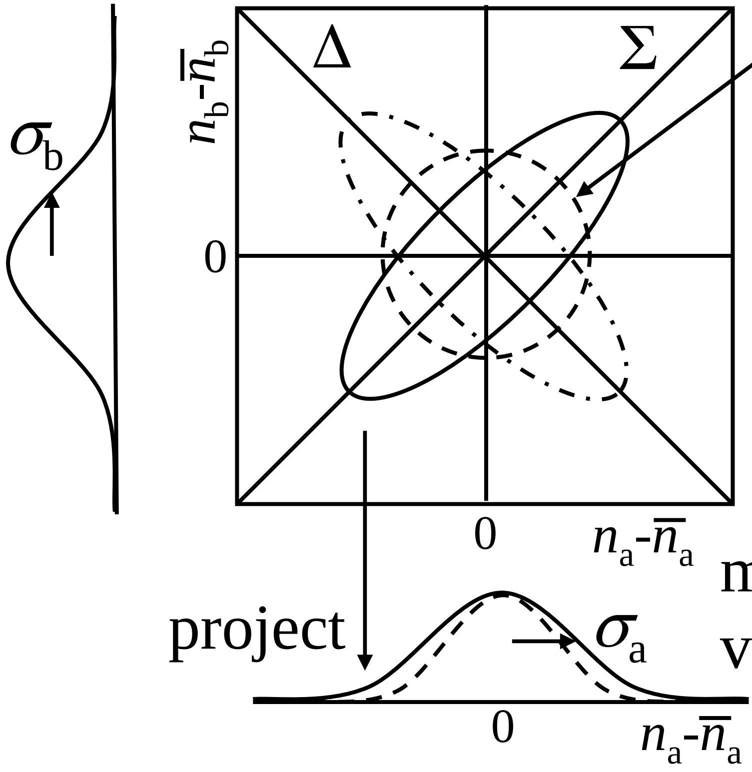
could also be $p_t$ sum

multiple events



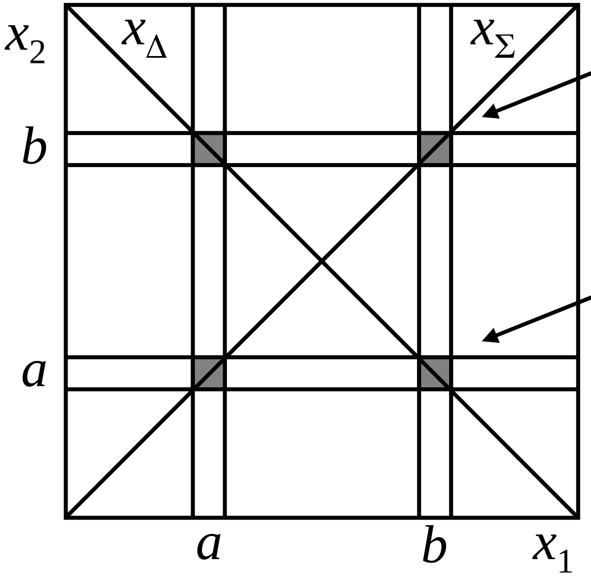
bins *a* and *b*

single-particle space



project

marginal variances



variance

covariance

product space

two-particle space

$$\sigma^2_{a,b} = \overline{(n-\bar{n})^2_{a,b}}$$

**variances for bins *a, b***

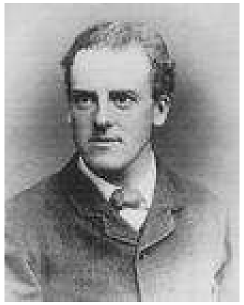
**covariance *between a* and *b***

$$\sigma^2_{ab} = \overline{(n-\bar{n})_a(n-\bar{n})_b}$$

$$\equiv \sigma^2_\Sigma - \sigma^2_\Delta$$

***variances depend on bin size or scale***

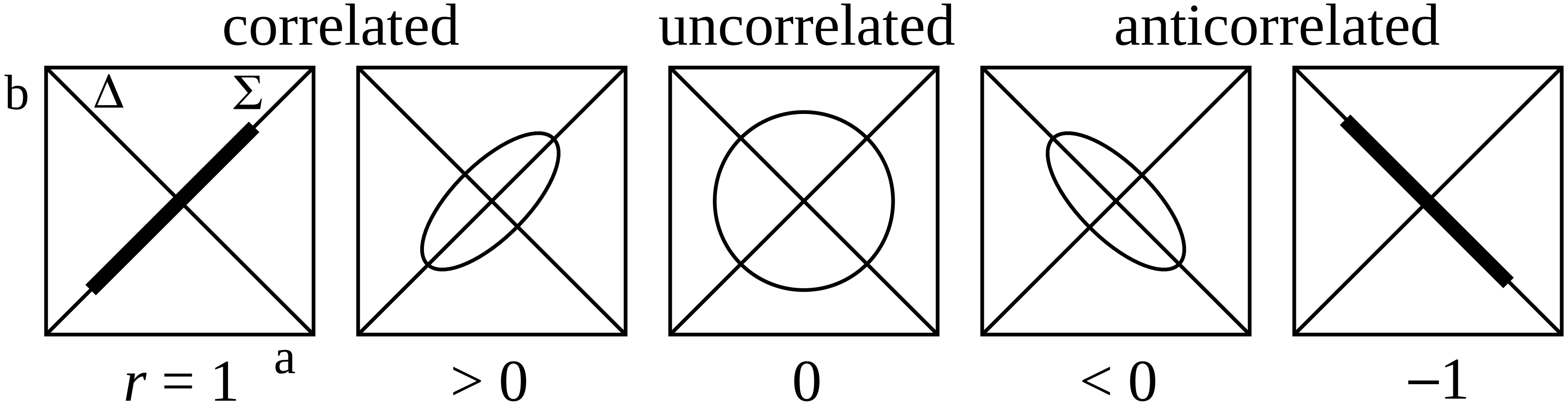
# Pearson's Correlation Coefficient



Karl Pearson, 1857-1936

$$r = \frac{\sigma_{ab}^2}{\sqrt{\sigma_a^2 \sigma_b^2}} = \frac{\sigma_\Sigma^2 - \sigma_\Delta^2}{\sigma_\Sigma^2 + \sigma_\Delta^2} \in [-1,1]$$ covariance *relative to* marginal variances

geometric mean of marginal variances



- Normalized fluctuation measure – the most basic correlation measure
- Relates fluctuations in two bins – are they correlated or not?

# Object and Reference Distributions

object distribution: *sibling* pairs

reference distribution: *mixed* pairs

***fluctuations***

variances, covariance

$$\sigma_a^2 = \overline{(n-\overline{n})_a^2} \qquad \sigma_{ab}^2 = \overline{(n-\overline{n})_a(n-\overline{n})_b}$$

reference is bin mean values

variance differences

$$\Delta\sigma_{ab}^2 = \overline{(n-\overline{n})_a^2} - \overline{(n-\overline{n})_b^2}$$

two bins, two scales,
two distributions

***correlations***

pair densities

$$\rho_{obj}(x_1, x_2) \qquad \rho_{ref}(x_1, x_2)$$

correlated pairs

$$\Delta\rho \equiv \rho_{obj}(x_1, x_2) - \rho_{ref}(x_1, x_2)$$

correlated pairs *per pair*

$$\Delta\rho / \rho_{ref} = \rho_{obj}(x_1, x_2) / \rho_{ref}(x_1, x_2) - 1$$

correlated pairs *per particle*

$$\Delta\rho / \sqrt{\rho_{ref}} \equiv \left\{\rho_{obj}(x_1, x_2) - \rho_{ref}(x_1, x_2)\right\} / \sqrt{\rho_{ref}}$$

# Autocorrelations

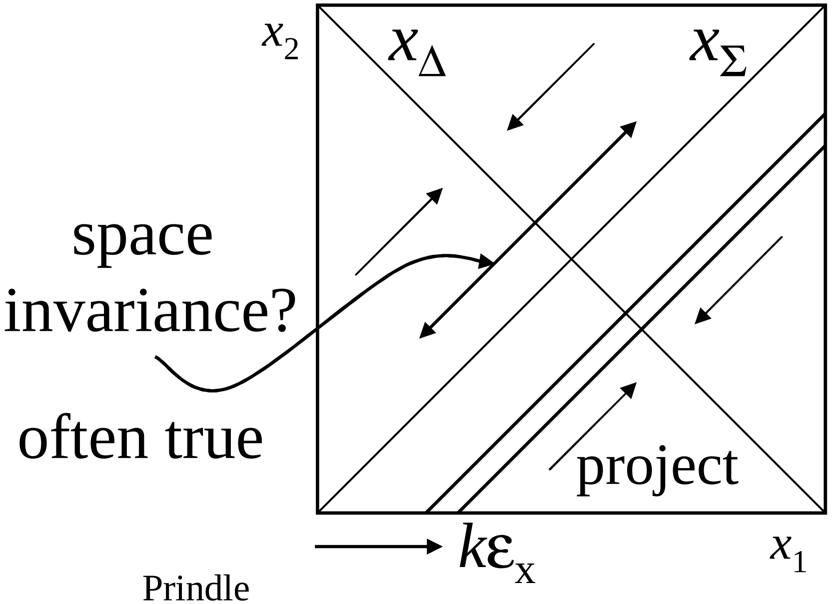
the autocorrelation is a projection *by averaging* of a two-point distribution onto difference variable(s) $x_\Delta$

$$\rho(x_1, x_2)$$

$$(\varphi_1, \varphi_2) \rightarrow \ \varphi_\Delta \equiv \varphi_1 - \varphi_2 \quad \varphi_\Sigma \equiv \varphi_1 + \varphi_2$$

$$(\eta_1, \eta_2) \rightarrow \ \eta_\Delta \equiv \eta_1 - \eta_2 \quad \eta_\Sigma \equiv \eta_1 + \eta_2$$

for space-invariant (on $x_\Sigma$) correlations this is a lossless projection



autocorrelation:
projection onto
difference variable

autocorrelations can be in the form of densities $\rho\,(x_\Delta)$
or histograms $A_k(\varepsilon_x) \cong \varepsilon_x^2 \rho(2k\varepsilon_x)$

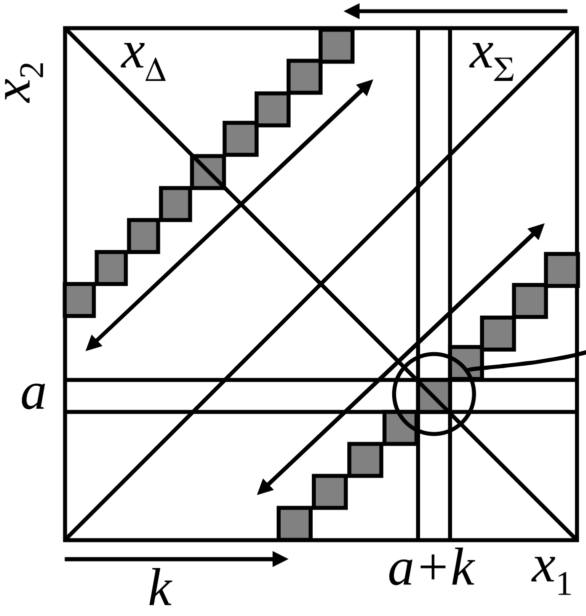
# Autocorrelation Density Ratio

Pearson

$$r = \frac{\sigma_{ab}^2}{\sqrt{\sigma_a^2\sigma_b^2}} = \frac{\overline{(n-\bar{n})_a(n-\bar{n})_b}}{\sqrt{\overline{(n-\bar{n})_a^2}\;\overline{(n-\bar{n})_b^2}}} \rightarrow \frac{\overline{(n-\bar{n})_a(n-\bar{n})_b}}{\sqrt{\bar{n}_a\bar{n}_b}}$$

Poisson values

k



$$\frac{\Delta A_k(n)}{\sqrt{A_{k,ref}(n)}} \equiv \left\{\frac{\overline{(n-\bar{n})_a(n-\bar{n})_{a+k}}}{\sqrt{\bar{n}_a\bar{n}_{a+k}}}\right\}_{\bar{a}}$$

average over *a* on $k^{\text{th}}$ diagonal

***density ratio – what we use***

$$\frac{\Delta A_k(n)}{\varepsilon_{x_\Delta}\sqrt{A_{k,ref}(n)}} \equiv \frac{\Delta\rho(n;k\varepsilon_{x_\Delta})}{\sqrt{\rho_{ref}(n;k\varepsilon_{x_\Delta})}}$$

autocorrelation density ratio is *diagonal average* of modified Pearson’s coefficients

# What About $p_t$ ?

$$r = \frac{\sigma^2_{ab}}{\sqrt{\sigma^2_a \sigma^2_b}} \sim \frac{\overline{(n-\bar{n})_a (n-\bar{n})_b}}{\sqrt{\bar{n}_a \bar{n}_b}\ \text{Poisson}} \quad \rightarrow \quad \frac{\overline{(p_t - n\,\hat{p}_t)_a (p_t - n\,\hat{p}_t)_b}}{\enclose{circle}{\sigma^2_{\hat{p}_t}}\sqrt{\bar{n}_a \bar{n}_b}\ \text{Poisson}}$$

relative number covariance

relative $p_t$ covariance

$$\left(p_t - \bar{p}_t\right) \equiv \left(p_t - n\hat{p}_t\right) + \hat{p}_t\left(n - \bar{n}\right)$$

implies

compare $\left(p_t - n\hat{p}_t\right)$ with $\hat{p}_t\left(n - \bar{n}\right)$

***we omit the $\sigma^2$ factor in the denominator to facilitate n-$p_t$ comparisons***

**a = b**

$$\Delta\sigma^2_{n/}\left(\delta x\right) \equiv \overline{\left(n\left(\delta x\right) - \bar{n}\left(\delta x\right)\right)^2} / \bar{n}\left(\delta x\right) - 1$$

**scale dependent**

$$\Delta\sigma^2_{p_t:n}\left(\delta x\right) \equiv \overline{\left(p_t\left(\delta x\right) - n\left(\delta x\right)\hat{p}_t\right)^2} / \bar{n}\left(\delta x\right) - \sigma^2_{\hat{p}_t}$$

number and $p_t$ relative covariances have the same structure: Pearson

# Autocorrelations from Pair Ratios

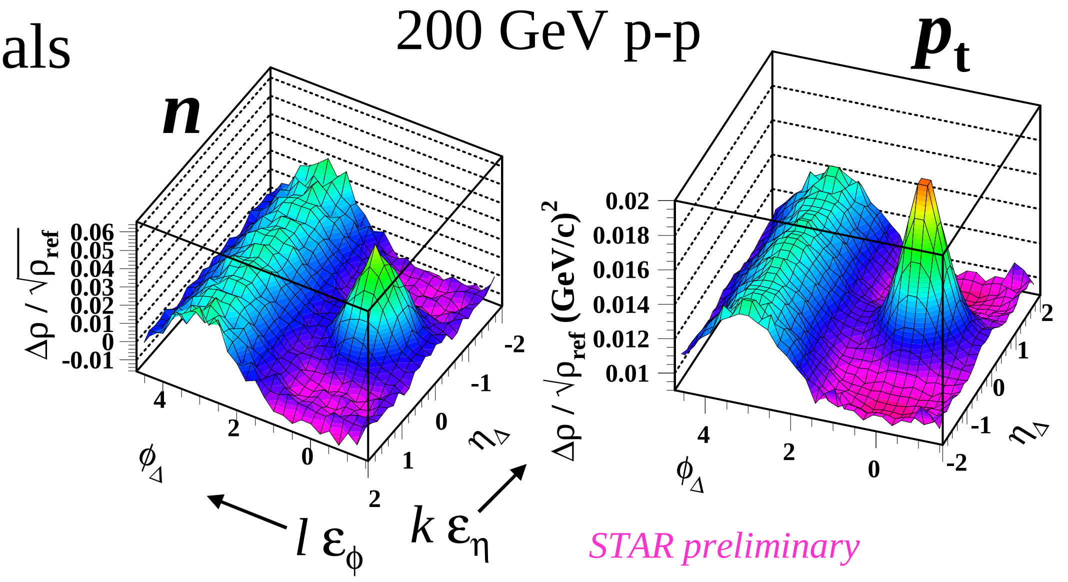
calculate autocorrelations *directly* as pair ratios

$$\frac{\Delta A_{kl}(n,p_t)}{\sqrt{A_{kl,ref}(n,p_t)}}$$

average over *a*,*b* on *k*th, *l*th diagonals

$$\equiv \left\{ \frac{\overline{(n-\bar{n})_{ab}(n-\bar{n})_{a+k,b+l}}}{\sqrt{\bar{n}_{ab}\bar{n}_{a+k,b+l}}} \right\}_{\overline{ab}}$$

$$\equiv \left\{ \frac{\overline{(p_t-n\,\hat{p}_t)_{ab}(p_t-n\,\hat{p}_t)_{a+k,b+l}}}{\sqrt{\bar{n}_{ab}\bar{n}_{a+k,b+l}}} \right\}_{\overline{ab}}$$



‘direct’ look at autocorrelation,

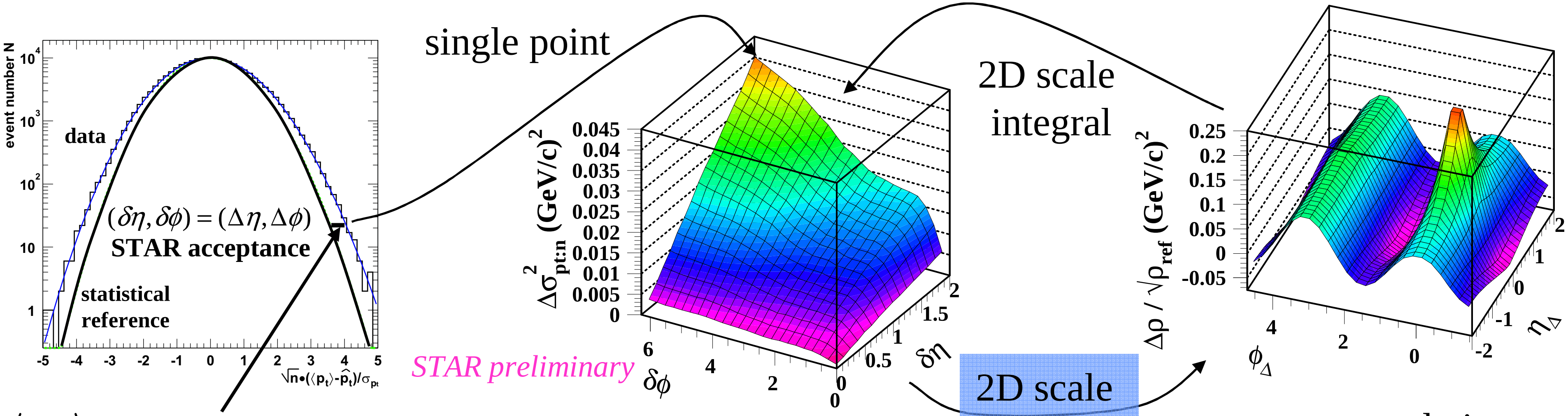
computationally expensive: $O(n^2)$

*joint* (η,φ) autocorrelations

equivalent ratios for *n and* $p_t$

autocorrelations can also be obtained by inverting corresponding fluctuation scale dependence →

# Fluctuations and Correlations



single point

2D scale integral

STAR preliminary

2D scale inversion

$\langle p_t \rangle$ fluctuation excess

scale dependence

autocorrelation

$$\Delta\sigma^2_{p_t:n}\left(m\varepsilon_\eta, n\varepsilon_\phi\right) = 4\sum_{k=1}^{m}\sum_{l=1}^{n}\overbrace{\left(1-\frac{k-1/2}{m}\right)\left(1-\frac{l-1/2}{n}\right)}^{\text{kernel } K_{mn;kl}}\overbrace{\frac{\Delta A}{\sqrt{A_{ref}}}\left(k\varepsilon_\eta, l\varepsilon_\phi\right)}$$

$$\Delta\sigma^2_{p_t:n}\left(\delta\eta,\delta\varphi\right) = 4\int_0^{\delta\eta} d\eta_\Delta \int_0^{\delta\varphi} d\varphi_\Delta K\left(\delta\eta,\delta\varphi;\eta_\Delta,\varphi_\Delta\right)\frac{\Delta\rho}{\sqrt{\rho_{ref}}}\left(\eta_\Delta,\varphi_\Delta\right)$$

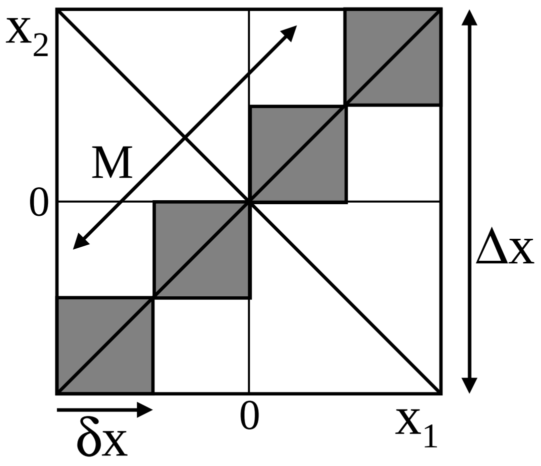
fluctuations ⇔ *integral equation* ⇔ correlations

# Derivation – Part I

To perform the numerical integration the acceptance is divided into microbins of fixed size $\varepsilon_x$

The average over macrobins is rearranged to an average over microbins

macrobin average



$M$ = number of macro-bins in acceptance

what we measure

***average over all macrobins in all events***

$$\Delta\sigma^2_{p_t:n}(\delta x) \equiv \overline{\{p(\delta x) - n(\delta x)\hat{p}\}^2 / \bar{n}(\delta x)} - \overline{\sum_i (p_i - \hat{p})^2}$$

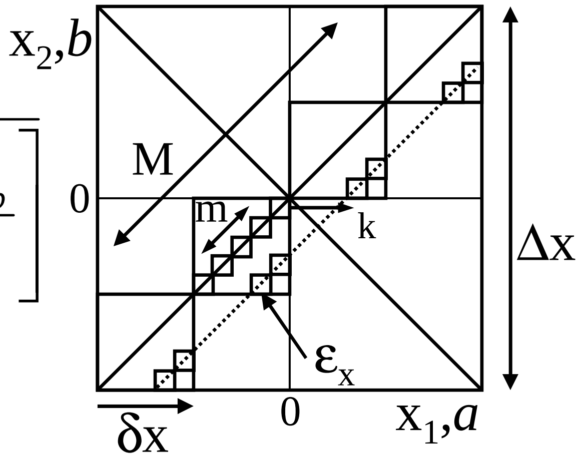
$$= \overline{\sum_{a,b=1}^{m} \frac{\{p_t(\varepsilon_x) - n(\varepsilon_x)\hat{p}_t\}_a \{p_t(\varepsilon_x) - n(\varepsilon_x)\hat{p}_t\}_b}{m\bar{n}(\varepsilon_x)}}$$

$$= \overline{\sum_{k=1-m}^{m-1} K_{m;k} \frac{\bar{n}_k}{\bar{n}} \left[ \frac{1}{m-|k|} \sum_{1\le a,b\le m}^{a-b=k} \frac{\sqrt{\bar{n}_a \bar{n}_b}}{\bar{n}_k} \cdot \frac{\{\cdots\}_a \{\cdots\}_b}{\sqrt{\bar{n}_a \bar{n}_b}} \right]}$$

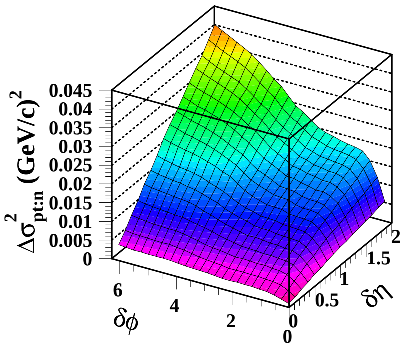
$$\equiv \sum_{k=1-m}^{m-1} K_{m;k} \frac{\Delta A(p_t : n; \varepsilon_x)}{\sqrt{A_{k,ref}(n;\varepsilon_x)}}$$

what we infer

microbin average



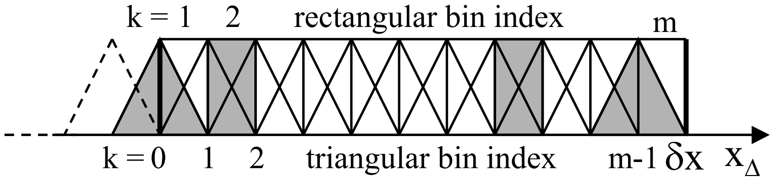
# Derivation – Part II



what we measure

$$\Delta\sigma^2_{p_t:n}(\delta x) \equiv \sum_{k=1-m}^{m-1} K_{m;k} \frac{\bar{n}_k}{\bar{n}} \frac{\Delta A_k(p_t:n;\varepsilon_x)}{\sqrt{A_{k,ref}(n;\varepsilon_x)}}$$

slightly modified binning

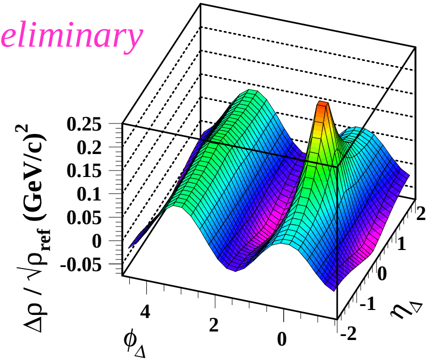


note factor 2 → $$2\sum_{k=1}^{m} K_{m;k} \frac{\bar{n}_k}{\bar{n}} \frac{\Delta A_k(p_t:n;\varepsilon_x)}{\sqrt{A_{k,ref}(n;\varepsilon_x)}}$$

invoke autocorrelation symmetry on $x_\Delta$
and use rectangular bin index $[1,m]$

1D $$= 2\sum_{k=1}^{m} \varepsilon_x K_{m;k} \frac{\bar{n}_k}{\bar{n}} \frac{\Delta\rho(p_t:n;k\varepsilon_x)}{\sqrt{\rho_{ref}(n;k\varepsilon_x)}}$$

*STAR preliminary*



2D $x \to (\eta, \phi)$

$$\Delta\sigma^2_{p_t:n}\left(m\varepsilon_\eta, n\varepsilon_\phi\right) = 4\sum_{k=1}^{m}\varepsilon_\eta\sum_{l=1}^{n}\varepsilon_\phi\left(1-\frac{k-1/2}{m}\right)\left(1-\frac{l-1/2}{n}\right)\frac{\Delta\rho(p_t:n)}{\sqrt{\rho_{ref}(n)}}\left(k\varepsilon_\eta, l\varepsilon_\phi\right)$$

what we infer

computationally cheap: $O(n)$

# Inversion and Regularization

Fredholm integral equation:

$$\underset{\mathbf{D}}{\Delta\sigma^2_{p_t:n}\left(m\varepsilon_\eta, n\varepsilon_\phi\right)} = 4\sum_{k=1}^{m}\varepsilon_\eta\sum_{l=1}^{n}\varepsilon_\phi \overset{K_{\text{mn,kl}}}{\underbrace{\left(1-\frac{k-1/2}{m}\right)\left(1-\frac{l-1/2}{n}\right)}_{\mathbf{T}}} \underset{\mathbf{I}}{\frac{\Delta\rho(p_t:n)}{\sqrt{\rho_{ref}(n)}}\left(k\varepsilon_\eta, l\varepsilon_\phi\right)}$$

or $\mathbf{D} = \mathbf{T}\mathbf{I} + \mathbf{N}$ matrix equation

We know **T** - given **D** we solve for **I:**

$$\mathbf{I}_\alpha = \mathbf{T}_\alpha^{-1}(\mathbf{D}_\alpha - \mathbf{N}_\alpha)$$

$\mathbf{T}^{-1}$ is a differentiation, acts as a high-pass filter

$\mathbf{I}_\alpha$ is treated as a matrix of free values in a $\chi^2$ fit subject to:

the $\mathbf{T}^{-1}\mathbf{N}$ noise term must be reduced by smoothing or ‘regularization’

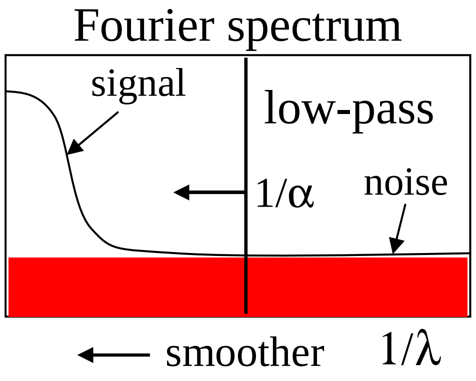
Tikhonov *regularization* - minimize:

$$\chi^2_\alpha \equiv \| \underset{\substack{\downarrow \\ \text{data-image mismatch}}}{\mathbf{D} - \mathbf{T}\mathbf{I}_\alpha} \|^2 + \alpha \| \underset{\substack{\downarrow \\ \text{small-}\lambda\text{ noise on image}}}{\mathbf{L}\mathbf{I}_\alpha} \|^2$$

*e.g.*, local gradient

# Choosing α

$\mathbf{T}_\alpha^{-1}$ is effectively a ‘high-pass’ filter: small-wavelength noise is increased

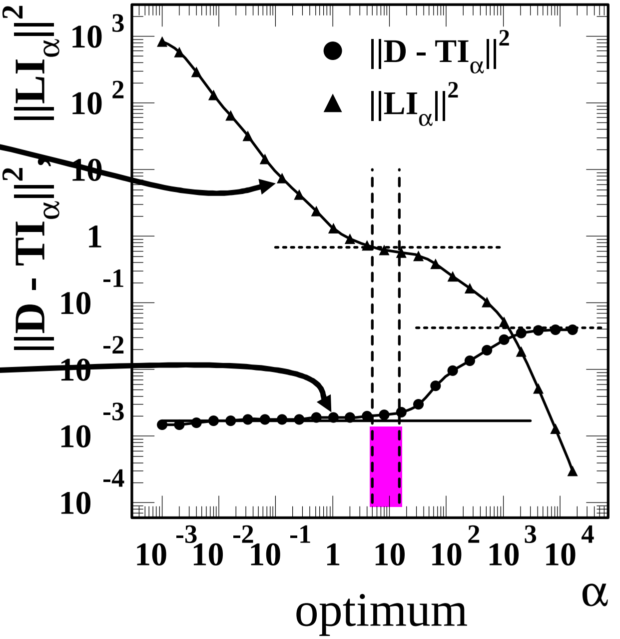


$\alpha$ controls a compensating low-pass filter: small values retain all data information, but also all noise larger values reduce noise, and finally signal

**noise**

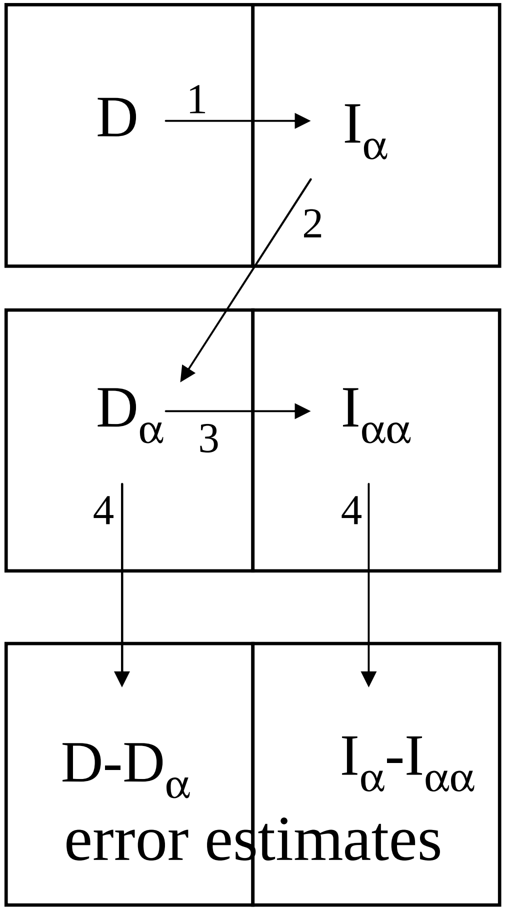
$$\chi^2_\alpha \equiv \| \mathbf{D} - \mathbf{T}\mathbf{I}_\alpha \|^2 + \alpha \| \mathbf{L}\mathbf{I}_\alpha \|^2$$

**signal loss**



# Error Estimation

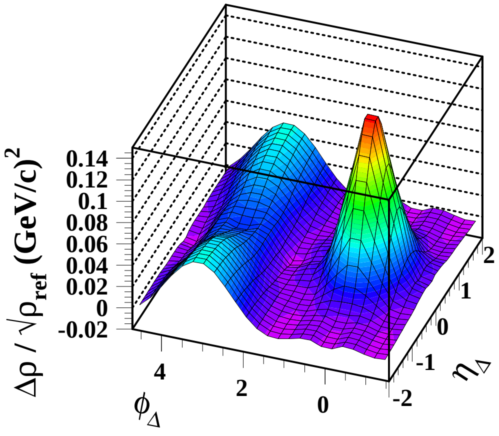
1. Invert data to get smoothed autocorrelation
2. Integrate forward to get smoothed ‘data’
3. invert smoothed ‘data’ a second time
4. differences in either case estimate combination of residual statistical error and smoothing distortion



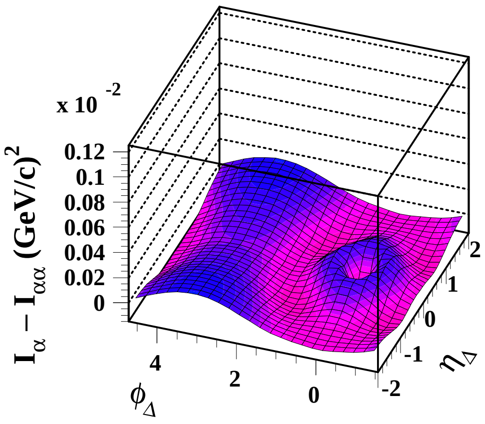
$$\begin{tikzcd}
D \arrow[r, "1"] & I_\alpha \arrow[dl, "2"] \\
D_\alpha \arrow[r, "3"] \arrow[d, "4"] & I_{\alpha\alpha} \arrow[d, "4"] \\
D-D_\alpha & I_\alpha-I_{\alpha\alpha}
\end{tikzcd}$$

error estimates

***Hijing quench-off***



**inverted image**

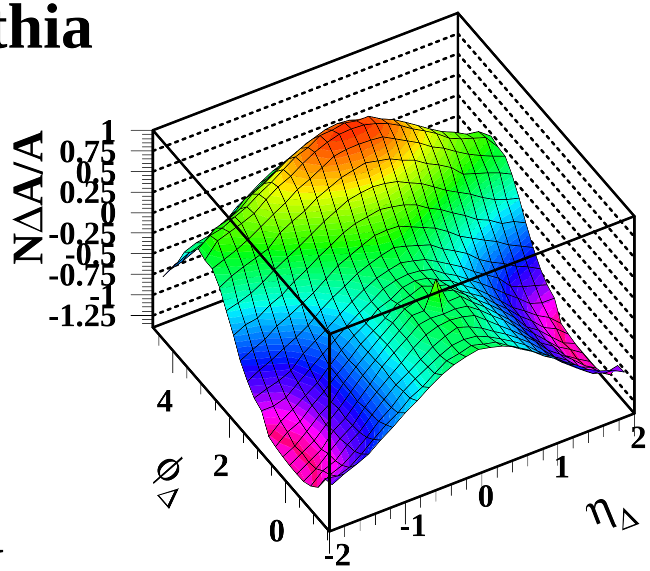


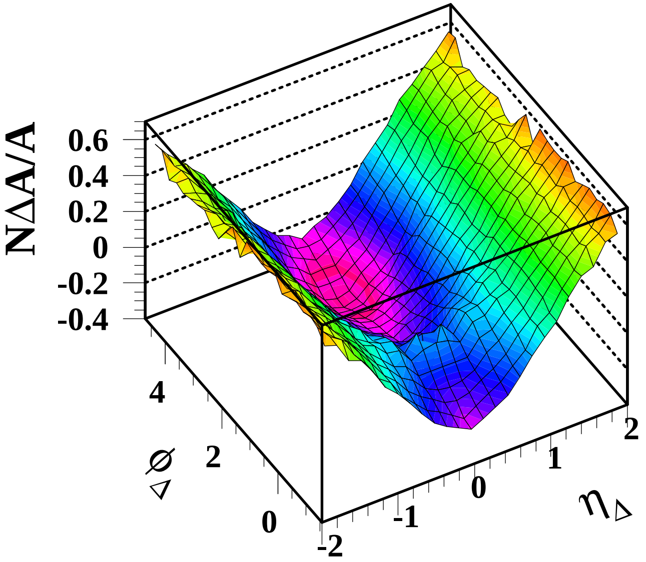
**smoothing error**

# Inversion Precision: Comparisons

**Pythia**

**direct**



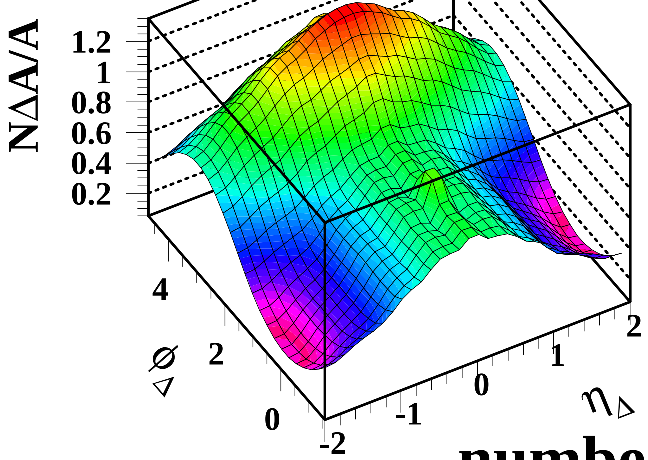


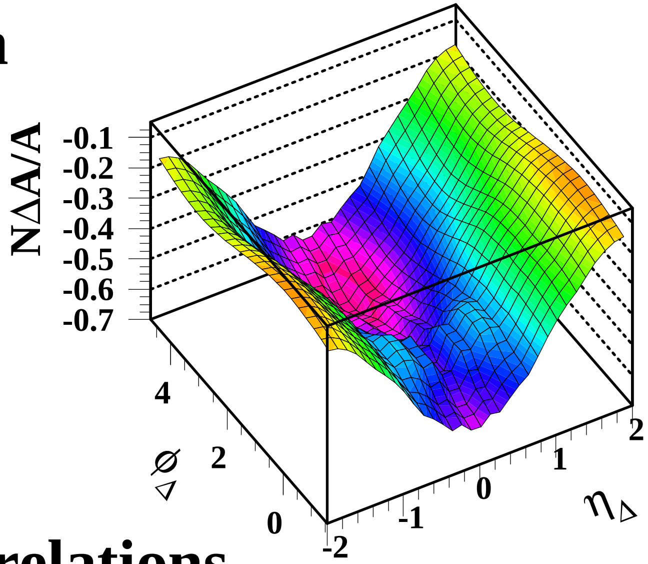
**CI**

**CD**

$\approx 25\Delta\rho / \sqrt{\rho_{ref}}$

**inversion**

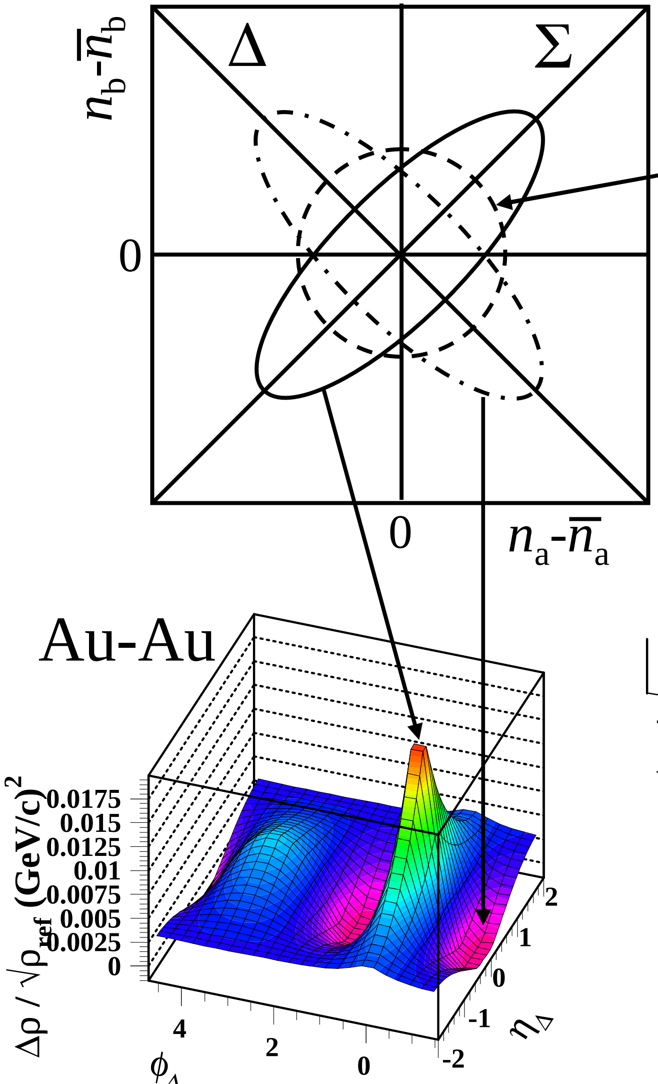




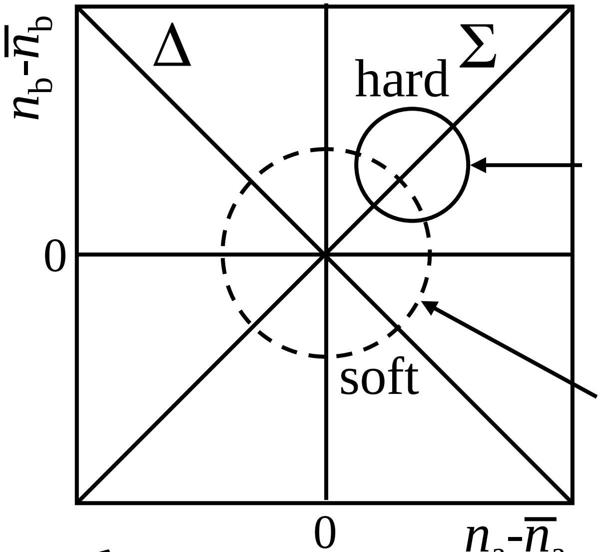
**number autocorrelations**

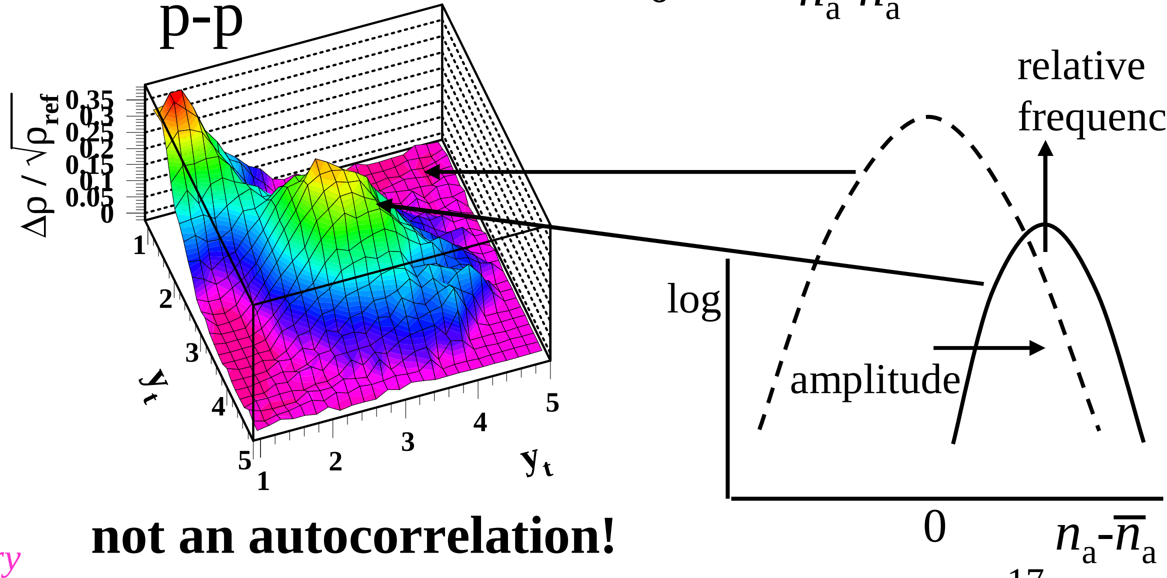
# Two Correlation Types

correlated and anticorrelated



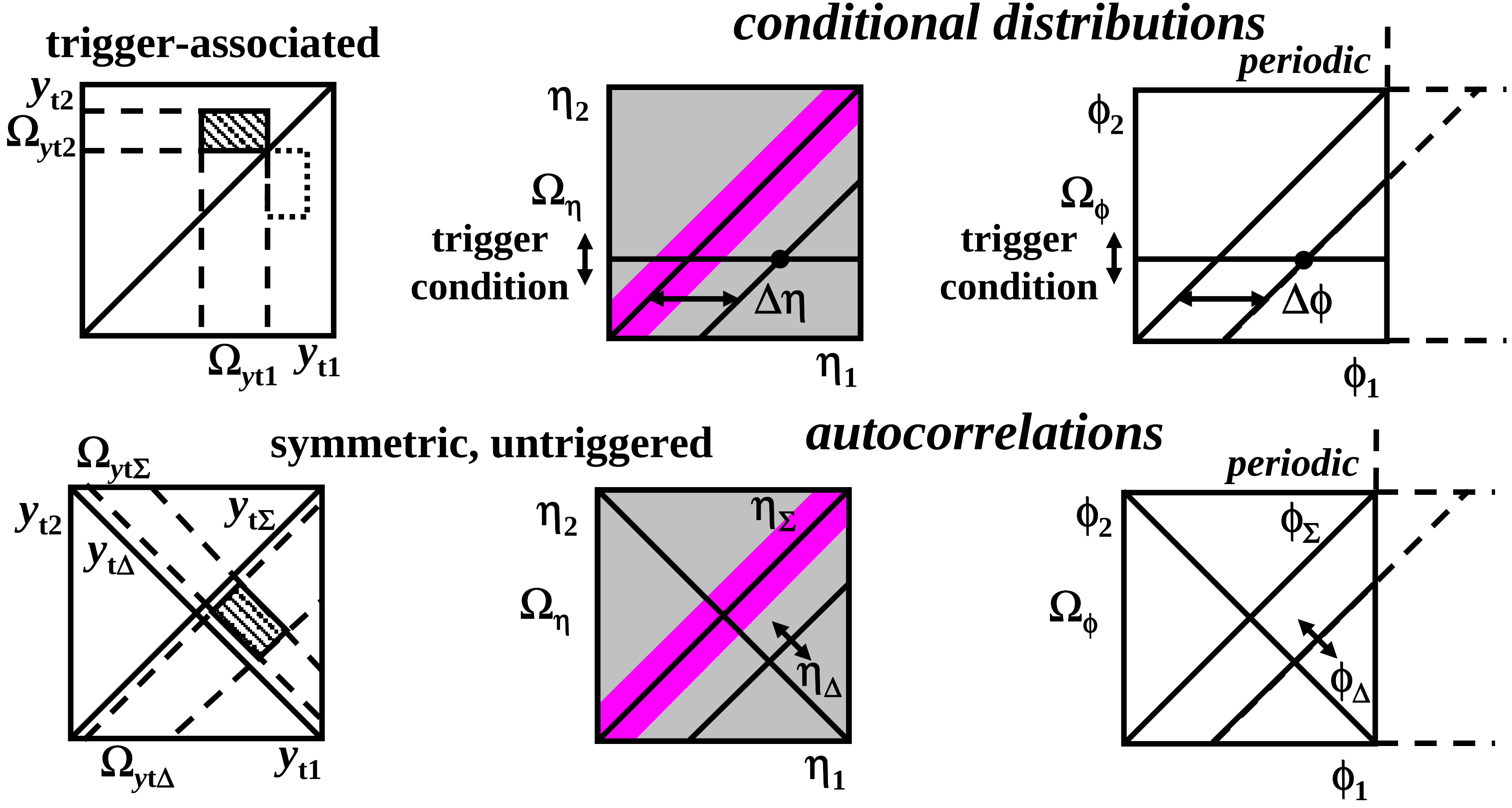
rare events





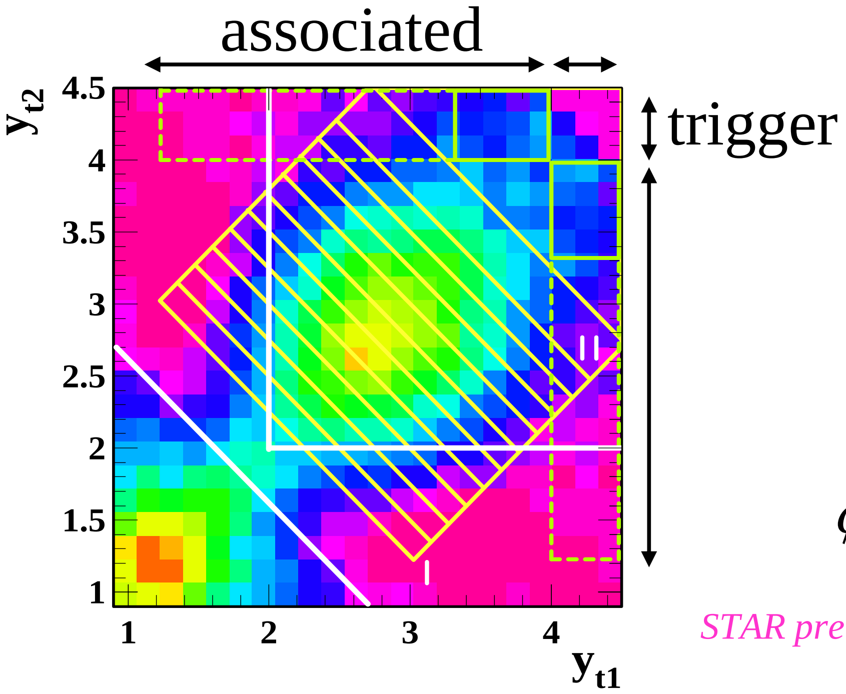
**not an autocorrelation!**

# Autocorrelations and Conditional Distributions



$$\frac{\Delta\rho_2(n;\eta_\Delta,\phi_\Delta)}{\sqrt{\rho_{ref}(n;\eta_\Delta,\phi_\Delta)}} \quad vs \quad \frac{1}{N_{trigger}}\frac{dN_{pair}}{d\Delta\phi} \equiv \frac{\int_{\Omega_\eta} d\Delta\eta \; d^2N_{pair}(\Delta\eta,\Delta\phi)/d\Delta\eta\, d\Delta\phi}{N_{trigger}}$$

# Leading-particle *vs* Symmetric

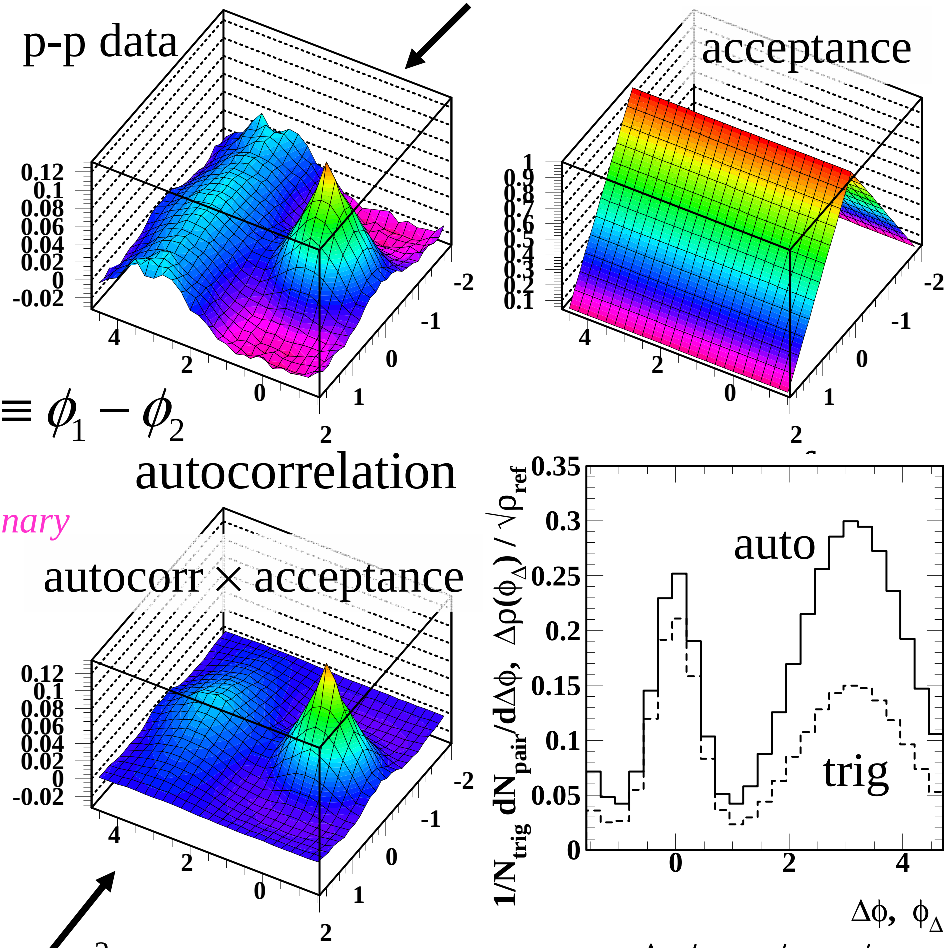


***mixed-pair reference subtraction***

$$\frac{\Delta\rho(n;\eta_\Delta,\phi_\Delta:\Omega_{y_{t\Sigma}},\Omega_{y_{t\Delta}})}{\sqrt{\rho_{ref}(n;\eta_\Delta,\phi_\Delta:\Omega_{y_{t\Sigma}},\Omega_{y_{t\Delta}})}}$$

$$\phi_\Delta \equiv \phi_1 - \phi_2$$

*STAR preliminary*



$$\frac{1}{N_{trig}}\frac{d^2N_{pair}}{d\Delta\eta d\Delta\phi} - \text{'background'}$$

$$\Delta\phi \equiv \phi - \phi_{trig}$$

- The distribution above is sibling – mixed, with all background removed
- Triggered backgrounds require model subtraction

# Summary

- An integral equation connects fluctuation scale dependence with corresponding autocorrelations
- Inversion of the integral equation produces autocorrelations equivalent to direct pair counting
- Fluctuations are thereby directly interpretable in terms of the underlying two-particle correlations
- Autocorrelations are complementary to leading-particle techniques for studying nuclear collisions