

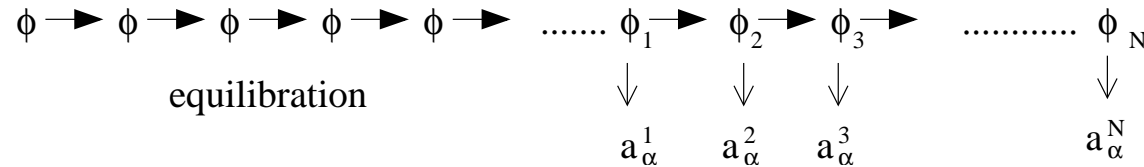
Part II: Optimized error analysis for expensive data

Ulli Wolff



- ⇒ Definition of the problem, basic formulas
- ⇒ Binning strategy, Γ -method
- ⇒ Description of software for download
- ⇒ Sample applications
 - all details found in paper [hep-lat/0306017v3](https://arxiv.org/abs/hep-lat/0306017v3)
 - transparencies: www-com.physik.hu-berlin.de → talk Orsay 2003
 - software download: www-com.physik.hu-berlin.de/ALPHAsoft

The problem



- $a_\alpha^i, i = 1, \dots, N$ successive MC estimates of observables with **exact** means A_α
- recording started ($i = 1$) **after equilibration**
- we want to estimate A_α and functions $F = f(A_1, A_2, \dots)$

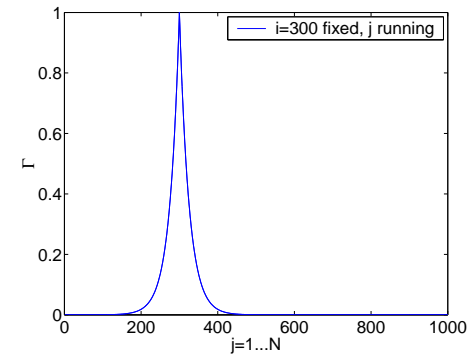
obvious estimators: $\bar{a}_\alpha = \frac{1}{N} \sum_{i=1}^N a_\alpha^i, \quad \bar{F} = f(\bar{a}_1, \bar{a}_2, \dots)$

statistical error: $\sigma_{\alpha\alpha}^2 = \langle (\bar{a}_\alpha - A_\alpha)^2 \rangle \quad \sigma_F^2 = \langle (\bar{F} - F)^2 \rangle \approx \sum_{\alpha\beta} \frac{\partial f}{\partial A_\alpha} \frac{\partial f}{\partial A_\beta} \sigma_{\alpha\beta}^2$

- $\langle \dots \rangle \leftrightarrow$ infinite ensemble of Markov chains (same algorithm, independent random #)
- simplification for this talk: **only one α** \rightarrow index dropped (easy to generalize)

$$\sigma^2 = \frac{1}{N^2} \sum_{i,j=1}^N \langle (a^i - A)(a^j - A) \rangle = \frac{1}{N^2} \sum_{i,j=1}^N \Gamma(i - j) \approx \frac{1}{N} C$$

with $C = \sum_{t=-\infty}^{\infty} \Gamma(t)$, $\Gamma(t) \sim \exp(-|t|/\tau)$, $N \gg \tau$



- $\langle (a^i - A)(a^{i+t} - A) \rangle = \Gamma(t) = \Gamma(-t) =$ autocorrelation function at equilibrium
- $\Gamma(t \neq 0)$ depends on the **algorithm**, $\Gamma(0)$ is the (**static**) variance
- C is like a susceptibility, $2\tau_{\text{int},A} = C/\Gamma(0)$

\Rightarrow we need to estimate C (\rightarrow matrix $C_{\alpha\beta}$ in general)

Binning strategy

- divide $N = N_B \times B$ measurements into bins of B **consecutive** measurements each
- form bin-averages $b^k = \frac{1}{B} \sum_{i \in \text{bin}\#k} a^i$
- mean $\bar{b} = \frac{1}{N_B} \sum_{k=1}^{N_B} b^k \equiv \bar{a}$ identical
- **one neglects autocorrelations between b^k** and uses the standard estimator for the error:

$$\bar{\sigma}_{\text{bin}}^2 = \frac{1}{N_B(N_B - 1)} \sum_{k=1}^{N_B} (b^k - \bar{b})^2$$

this implies an estimate \bar{C}_{bin} for C with bias and statistical error

$$\langle \bar{C}_{\text{bin}} - C \rangle \sim -\frac{\tau}{B} C \quad \langle (\bar{C}_{\text{bin}} - C)^2 \rangle \approx \frac{2}{N_B} C^2 = \frac{2B}{N} C^2$$

- **B has to be chosen to compromise between these**

Γ -strategy

estimate autocorrelation function $\Gamma(t)$ and C directly [Sokal], estimator:

$$\bar{\Gamma}(t) = \frac{1}{N-t} \sum_{i=1}^{N-t} (a^i - \bar{a})(a^{i+t} - \bar{a}), \quad \langle \bar{\Gamma}(t) - \Gamma(t) \rangle \simeq -\frac{C}{N}$$

giving

$$\bar{C}_\Gamma(W) = \bar{\Gamma}(0) + 2 \sum_{t=1}^W \bar{\Gamma}(t)$$

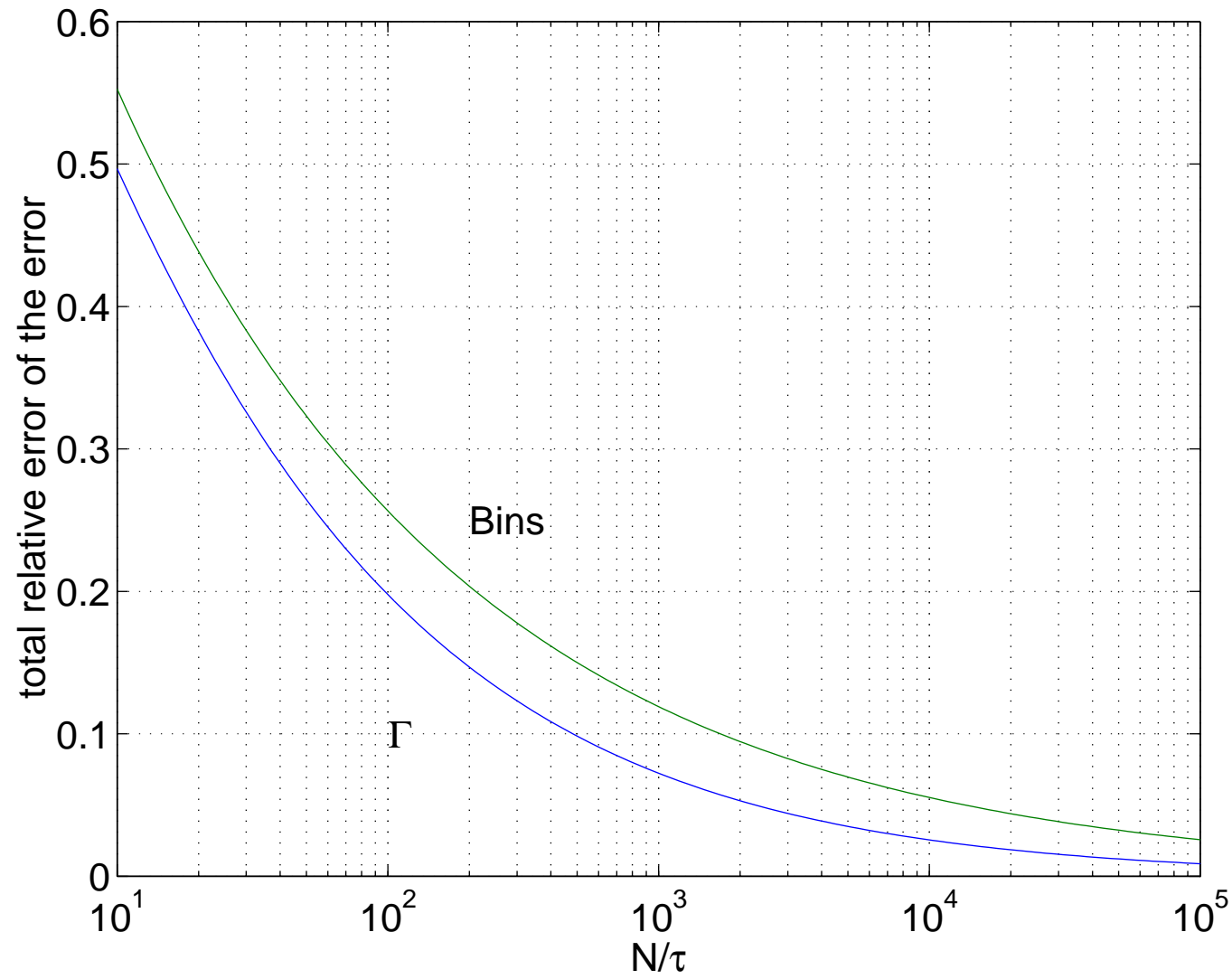
with bias and statistical error

$$\langle \bar{C}_\Gamma - C \rangle \sim -\exp(-W/\tau) C \quad \langle (\bar{C}_\Gamma - C)^2 \rangle \approx \frac{2(2W+1)}{N} C^2$$

- $2W$ is similar to B and has to be chosen to compromise....but exp

Relative errors of estimates for σ^2 , σ_F^2 (error of the error!)

Quantity	Binning	Γ
sys + stat =	$\tau/B + \sqrt{2B/N}$	$\exp(-W/\tau) + \sqrt{2(2W+1)/N}$
sum minimal for	$B = \tau(2N/\tau)^{1/3}$	$W \simeq \ln(N/\tau)\tau/2$
value	$3/2(2N/\tau)^{-1/3} \propto N^{-1/3}$	$\sqrt{2 \ln(N/\tau)\tau/N} \propto N^{-1/2}$
ratio sys/stat =	$1/2$	$1/\ln(N/\tau)$



Matlab implementation

download MATLAB routine `UWerr.m` from www-com.physik.hu-berlin.de/ALPHAsoft

additional features compared to what is discussed here

- estimate $f(A_1, A_2, A_3, \dots)$ for user-defined routine for f or for primary A_α themselves
- handle replica-simulations of arbitrary lengths, combined error analysis, check of statistical compatibility (goodness of “fit” to a constant Q computed)
- optimal window W assessed semi-automatically; it needs a rough estimate of $S = \tau / \tau_{\text{int},F}$
- plots generated for Γ , τ_{int} , replica distribution (see sample applications below)

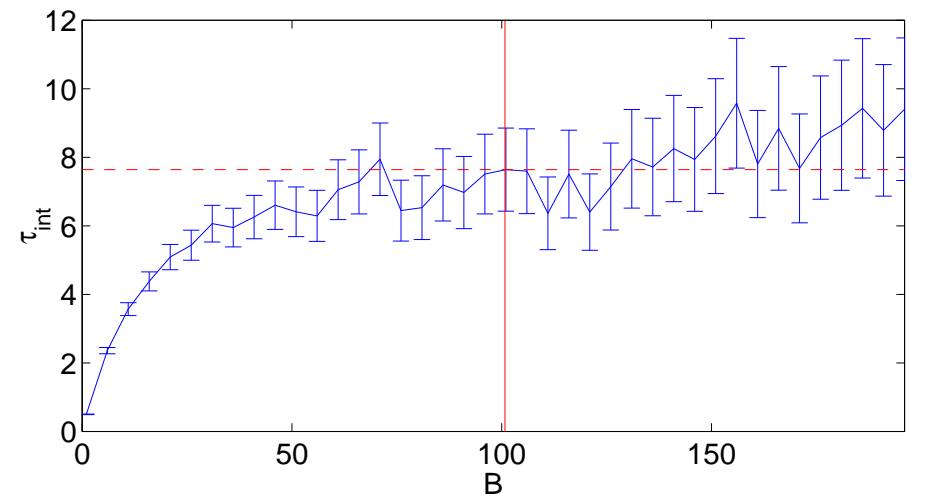
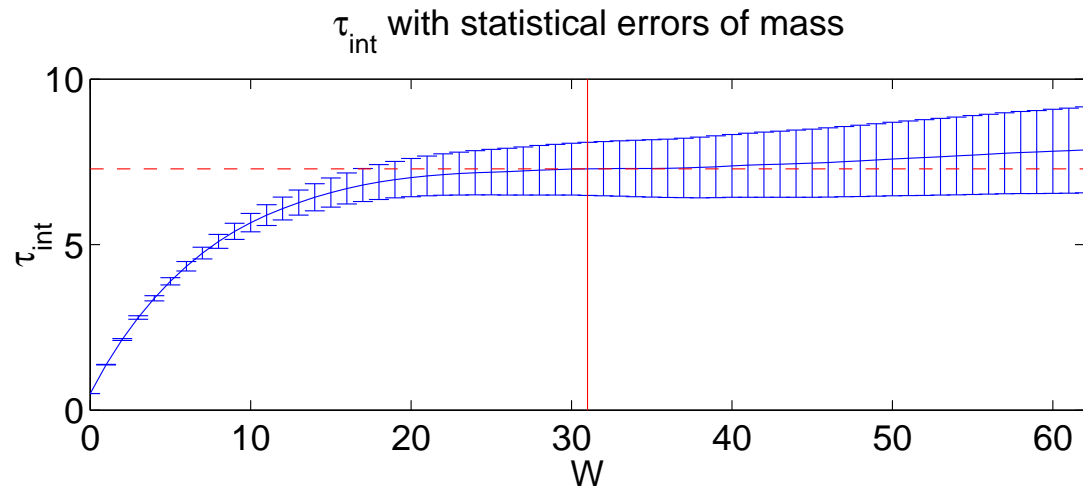
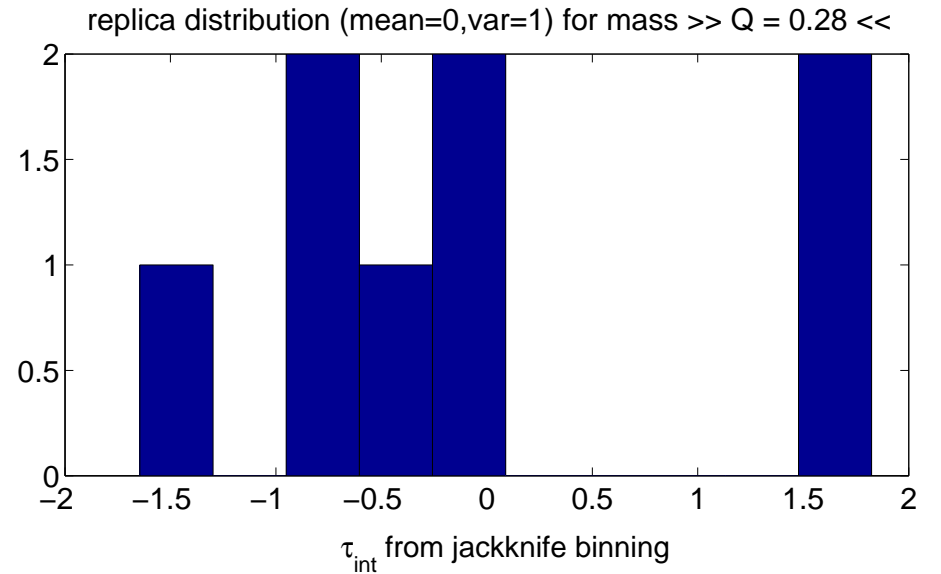
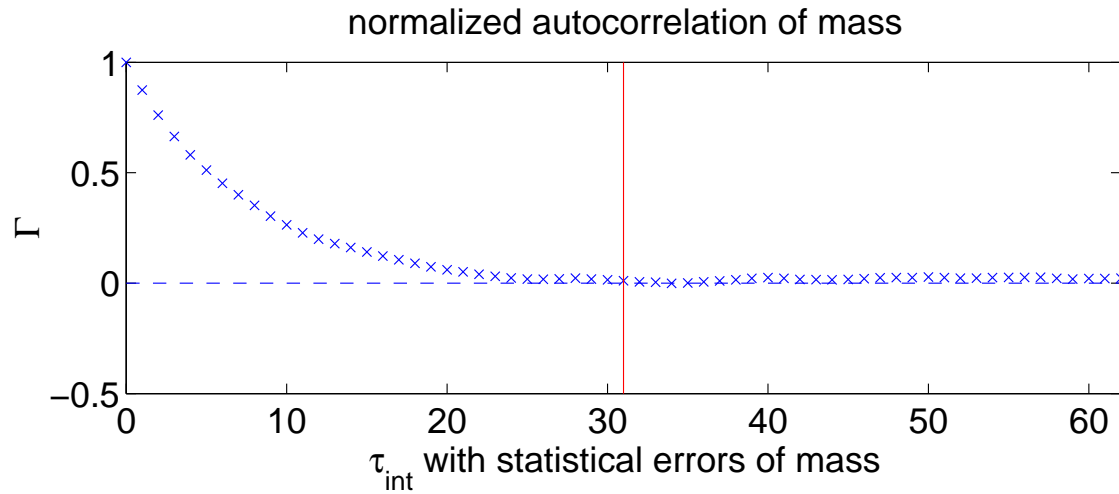
Application 1: Synthetic data

- artificial noisy autocorrelated data for 'time-slice correlation function' $A_1 = G(z)$, $A_2 = G(z + 1)$
- measure exactly known $m = f(A_1, A_2) = \ln(A_1/A_2) = 0.2$
- noise (auto)correlated between a_1^i, a_2^i (with $\tau = 4, 8$ components) such that $\tau_{\text{int},m} = 7.92$
- 8 replica, 1000 measurements each, $\sigma_m = 0.0142$

```
>> Nr=[1000 1000 1000 1000 1000 1000 1000 1000]; S=1;
>> [value,dvalue,ddvalue,tauint,dtauint,Q]= ...
    UWerr(Data,S,Nr,'mass',@effmass,1,2);
>> [value,dvalue,ddvalue,tauint,dtauint,Q]
```

```
ans = 0.2128    0.0134    0.0008    7.2909    0.8021    0.2827
```

```
exact: 0.2      0.0142      7.92
```

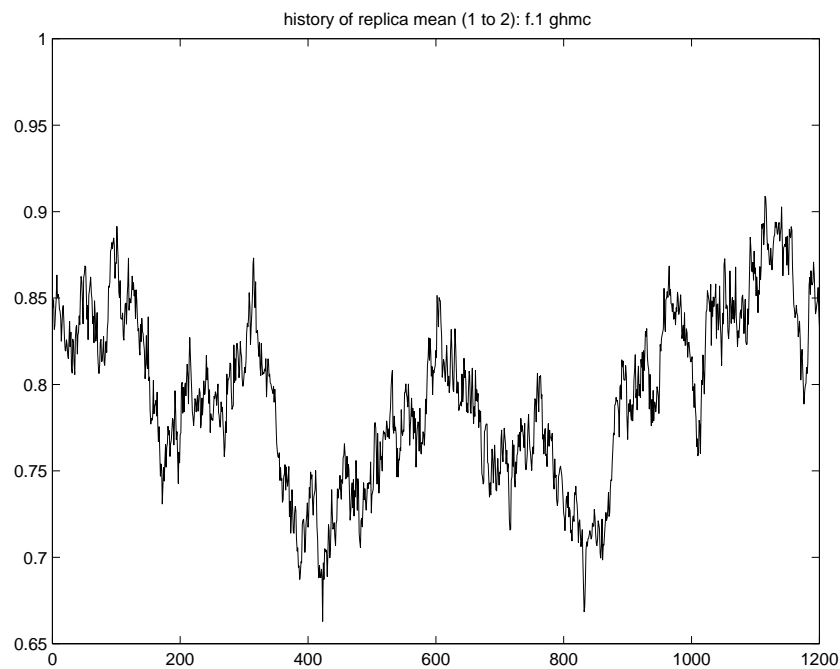


Z_P with Dynamical fermions

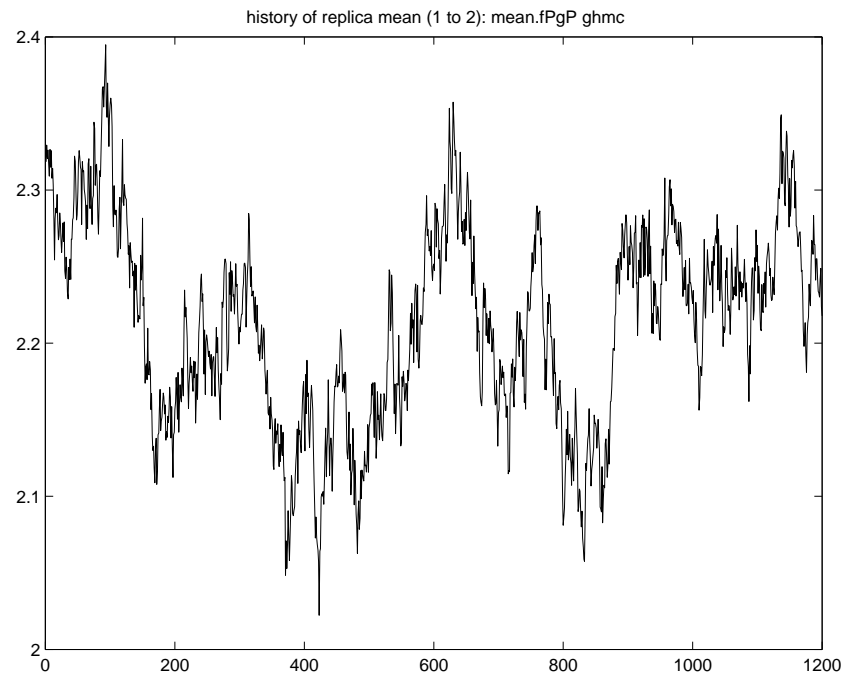
From the ALPHA running mass project with two lite flavors, a typical difficult run:
HMC with 2 pseudofermions, $L = 24$, $\beta = 8.02599$, $\kappa = 0.133063$, $\theta = 0.5$

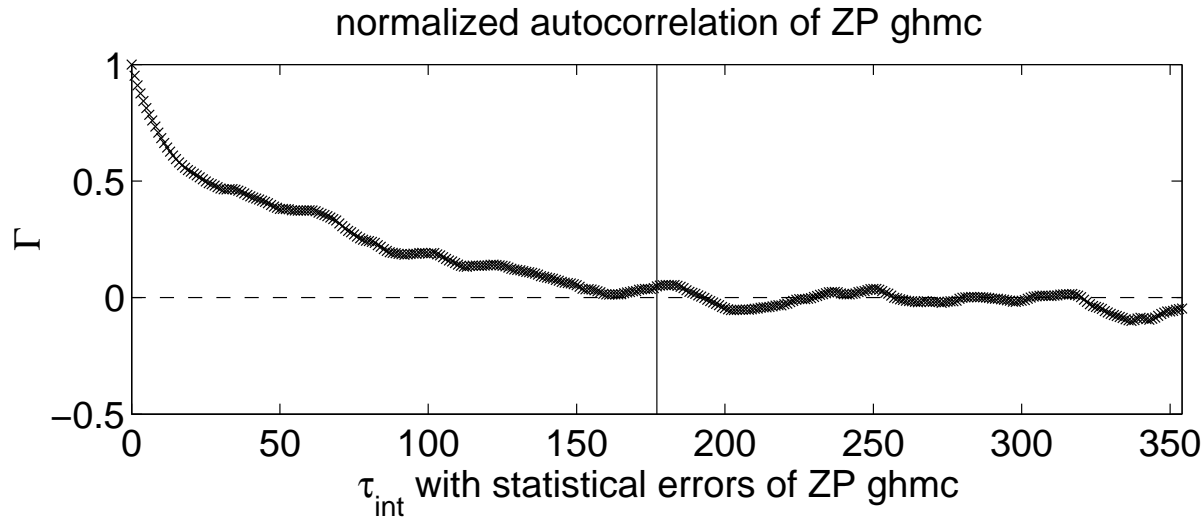
$$Z_P(L) = \frac{\sqrt{3f_1}}{f_P(L)}, \quad f_1, f_P \text{ are SF correlations of } \bar{\psi}\gamma_5\psi, \quad L = \mu^{-1} = \text{running-scale}$$

history f_1

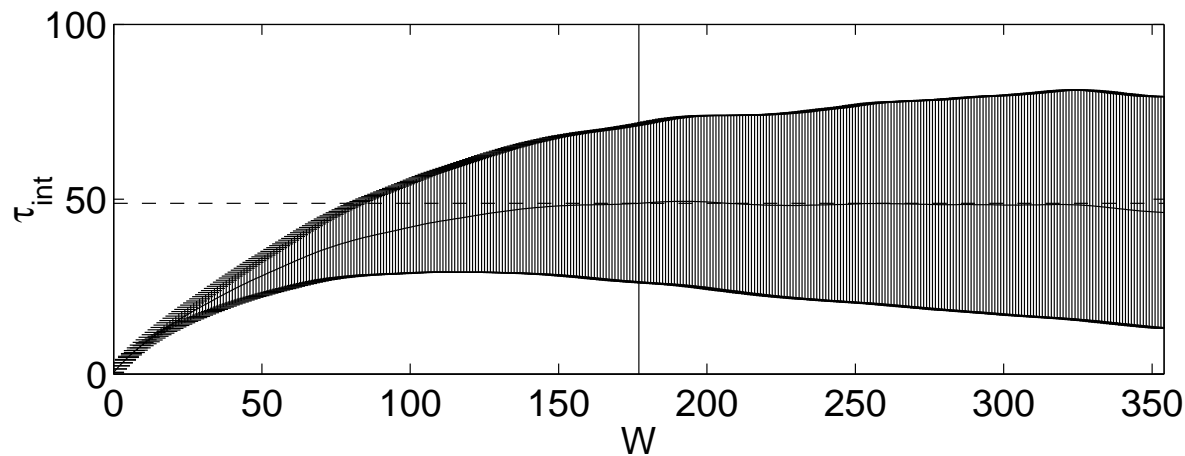


history f_P



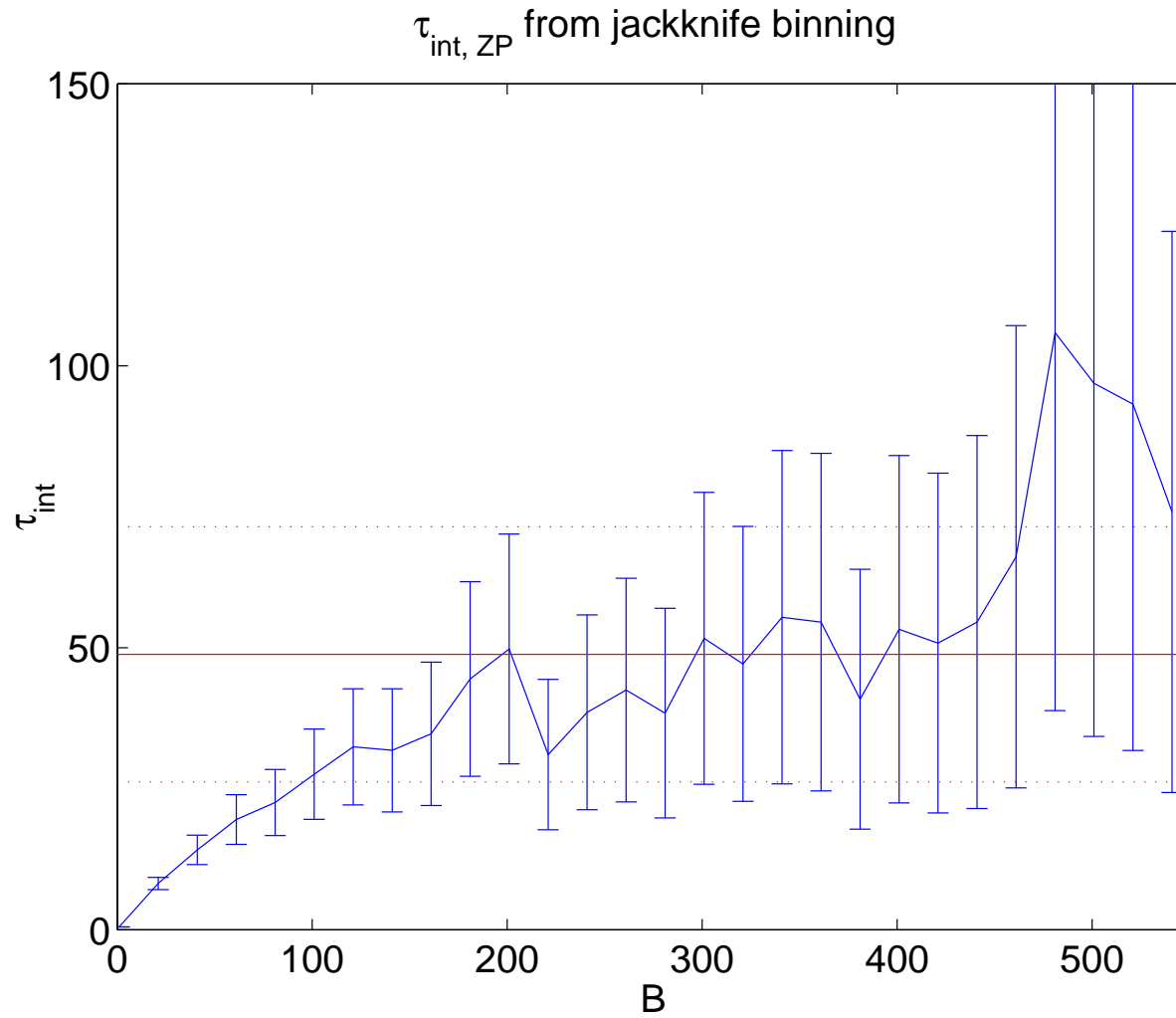


OBSERVABLE: f1
 2 x 1200 measurements
 mean-value = 0.79325
 error = 0.01838
 tau_int = 88.92
 error(Sokal) = 46.29
 Q-value = 0.80



OBSERVABLE: ZP
 2 x 1200 measurements
 mean-value = 0.69676
 error = 0.00322
 tau_int = 48.84
 error(Sokal) = 22.62
 Q-value = 0.37

Jackknife for Z_P



Correlation f_1

