Observations on how to improve electron efficiency measurements

Jeffrey Berryhill, Kalanand Mishra
Fermilab

Egamma meeting
(November 18, 2010)

Motivation

Uncertainty in the electroweak measurements (see Georgios' talk in Physics plenary on last Tuesday): [all numbers are in percent]

Source	$W \to e\nu$	$Z \to e^+e^-$	W^+ (e)	W^- (e)	W^+/W^- (e)	W/Z (e)
Lepton reconstruction & identification	3.9	5.9	5.1	5.1	5.2	3.0
Momentum scale & resolution	2.0	0.6	2.2	1.8	0.4	2.0
$\not\!\!E_T$ scale & resolution	1.8	n/a	1.6	1.9	0.4	1.8
Background subtraction / modeling	1.3	0.1	1.1	1.5	0.7	1.3
PDF uncertainty for acceptance	0.8	1.1	0.9	1.5	1.7	0.9
Other theoretical uncertainties	1.3	1.3	1.3	0.9	1.3	1.0
Total	5.1	6.2	6.1	6.2	5.7	4.4

✦For precision EWK analyses $e^\pm$ efficiency is the most significant uncertainty.

-If we want to improve upon our existing W,Z measurements then this is the place to start. Reducing syst uncertainty in efficiency is important in this case.

-For differential measurements bin-by-bin statistical uncertainty is large.

-W^+/W^- cross section ratio (3 pb^{-1}) = 1.433 ± 0.020 (stat) ± 0.052 (syst), where syst is predominantly the stat error from Z tag&probe e^+/e^- efficiency ratio and μ^+/μ^- efficiency ratio.

✦For other CMS measurements/searches (e.g., top, Exotica, SUSU, Higgs) electron efficiency still constitutes a significant source of systematic uncertainty.

Tag&probe machinery: PhysicsTools/TagAndProbe

An important tool for precision efficiency measurements

✦*PhysicsTools/TagAndProbe* is a lightweight ROOT-based framework that allows one to compute any kind of electron/muon efficiency

✦Highly flexible design: user can easily configure
- *tag, probe, passing probe* selections
- which variables to store in the Tree

✦User can compute efficiency as a function of any of the probe variables
- in any number of dimensions, e.g., E_T, η, ϕ, nJets, MET, Ht,...

✦Backend fitter allows user to specify how to estimate signal yields for numerator and denominator.

✦User has full control over efficiency computation. No black boxes.
- comes with official CMSSW releases, with periodic updates.
- wide user base; adopted for the first EWK, Quarkonia, top papers.

Details at: https://twiki.cern.ch/twiki/bin/view/CMS/TagAndProbe
https://twiki.cern.ch/twiki/bin/view/CMS/ElectronTagAndProbe

Hypernews: *Physics Analysis Tools:* hn-cms-physTools@cern.ch

Savannah: https://savannah.cern.ch/projects/physicstool/

An instantiation of tag&probe efficiency sequence

Used in inclusive W,Z cross section and ratio measurements, EWK-10/002

Tag Selection

- GsfElectrons.
- Super cluster within $|\eta|$ acceptance
- $E_T > 20$ GeV
- Isolation and Id cuts as in WP80
- Matched to lowest unprescaled electron trigger

Probe Selection

- $E_T > 20$ GeV, $|\eta|$ in acceptance
- Fit the tag-probe invariant mass to get the number of signal events.

Obtain factorized efficiencies for passing probes:

SuperCluster → GsfElectron → WP80/WP95/EleSel → HLT

offline electron reconstruction efficiency with respect to acceptance

trigger efficiency w.r.t. offline selection

Estimated efficiencies:

ε_{REC} : SuperCluster → GSF electron

ε_{WP80} : GSF electron → electron WP80

ε_{TRG} : electron WP80 → Trigger

Methodology to apply data-driven efficiency

✦**Efficiencies are determined from MC and corrected with data** [details on next slide]

$$\epsilon_{X} = \epsilon_{MC\text{-}X} \times \rho_{eff\text{-}X}, \quad \rho_{eff\text{-}X} = \frac{\epsilon_{TNP\text{-}X}(\mathrm{data})}{\epsilon_{TNP\text{-}X}(\mathrm{MC})}$$

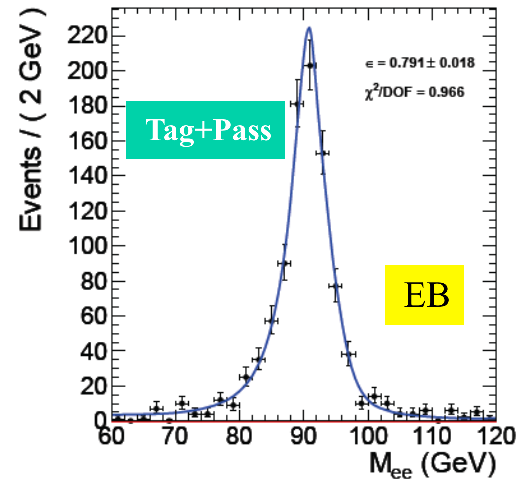
Efficiencies are determined in ECAL Barrel & Endcaps, for e^+ and e^-.

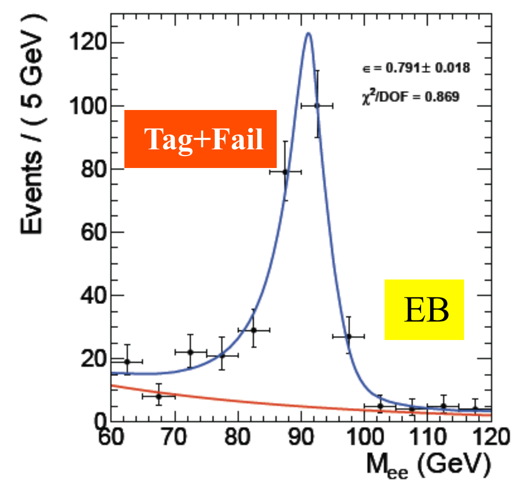
✦**Data/MC correction factors (ρ) are estimated with Tag & Probe** [details on next-to-next slide]

- Tag : WP80 electron matched to trigger object
  Probe: passing candidate from previous stage
- M_{ee} required in window around Z mass, 60–120 GeV

✦**Simultaneous unbinned likelihood fit of M_{ee} in Tag+Pass and Tag+Fail categories:** [details on 3 slides later]

-signal shape templates: (NLO lineshape ⊗ modified Crystal Ball) or (NLO lineshape ⊗ CMS simulation ⊗ Gaussian smearing)
- exponential or polynomial model for background





Why compute $\rho = \varepsilon_{data}/\varepsilon_{MC}$ scale factors ?

✦**Primarily because the electrons in your analysis has different kinematic distribution than in the calibration sample (i.e., Z→ee)**

Example 1: Electrons from W→eν decay have somewhat softer E_T spectrum than electrons from Z→ee decay. Since electron efficiency strongly depends on the E_T and η of the electron, the average efficiency for W electrons is different than Z electrons for the exact same WP80 selection ! These kinematic models are well described by Monte Carlo, therefore data/MC ratio is unbiased.

Example 2: Even in case of binned efficiency, or for Z→ee analysis itself, the electron efficiency may have weaker dependence on variables or conditions that are not accounted for explicitly, e.g., nJets, SumET, underlying event activity, ... If these effects are simulated well then they cancel out in data/MC ratio.

✦**If there is any bias in the tag&probe method itself (either inherent, or arising from the choice of tag or probe) then this will mostly cancel out in the data/MC ratio.**

In our opinion, this is the way analyses should be using data-driven efficiency. Absolute normalization of efficiency to data exposes you to kinematic biases, and should be avoided as much as possible.

Observations on tag, probe, and M_{ee} selections

✦The choice of tag electron collection is dictated by the following factors
- -it should be a well reconstructed and identified electron of high purity
- -should be at least as tight or tighter than the probe
- -should not be super tight to bias probe selection

✦The choice of probe and passing probe, and for that matter efficiency steps, is dictated by the analysis needs.

✦The M_{ee} range used to fit for the Z signal yield depends on the kinematics of the electrons.
- -For inclusive W, Z analysis the 60−120 GeV range was good enough (given the E_T>20 GeV cut for electrons).

✦We do not apply opposite charge (OS) requirement in either the W,Z analysis selection or in the efficiency computation
- -because this doesn't buy us anything (charge misId rate for WP80 is small)
- -and it is hard to compute efficiency for OS requirement with good precision.

Observations on backend simultaneous fitter

Why choose NLO lineshape (with additional smearing) ?

✦For signal, no simple parametric shape is able to describe the data
-Because of physics reasons

✦In case of dilepton mass peak from Z decay, we are not dominated by resolution but by the following main effects:
- -interference between γ* and Z (this changes the Breit-Wigner behavior)
- -final state radiation from lepton (this causes a long low mass tail)
- -in case of electrons, there is further effect due to escale and resolution
- -poorly reconstructed leptons in case of Tag+Fail events.

✦Overall resolution in data is slightly worse than in MC

Currently the stat. error is overestimated by 10−15% [in fact by $\sqrt{2}/\sqrt{(2-\varepsilon)}$]

See backup slides

These choices were made in order to compute the efficiency with high precision. People computing tag&probe efficiency for other analyses need to make their own choices. The example scripts in tag&probe package are exactly that − meant to serve as *example scripts*. They are highly configurable and easy to modify.

How uncertainties add up ...

AN-2010/349

I will focus on systematic uncertainties in the next slides

	Topology	Monte-Carlo		Data				Data/Monte-Carlo
		ϵ	$\Delta\epsilon$	ϵ	$\Delta\epsilon$ (stat)	$\Delta\epsilon$ (syst)	$\Delta\epsilon$ (total)	
SC→ Reco	EB	0.9851	0.0002	0.986	0.005	0.012	0.013	1.001 ± 0.013
	EE	0.9629	0.0004	0.962	0.008	0.012	0.015	0.999 ± 0.016
	EB e^-	0.9854	0.0002	0.985	0.005	0.012	0.013	0.999 ± 0.013
	EE e^-	0.9624	0.0004	0.967	0.002	0.012	0.012	1.005 ± 0.012
	EB e^+	0.9849	0.0002	0.987 [a]	0.004	0.012	0.013	1.002 ± 0.012
	EE e^+	0.9634	0.0004	0.951	0.003	0.012	0.012	0.987 ± 0.012
Reco→ WP80	EB	0.8547	0.0004	0.791	0.018	0.020	0.027	0.925 ± 0.032
	EE	0.7488	0.0006	0.692	0.020	0.020	0.028	0.924 ± 0.037
	EB e^-	0.8545	0.0005	0.793	0.027	0.020	0.034	0.929 ± 0.040
	EE e^-	0.7489	0.0009	0.701	0.038	0.020	0.043	0.936 ± 0.057
	EB e^+	0.8549	0.0005	0.780	0.031	0.020	0.037	0.913 ± 0.044
	EE e^+	0.7487	0.0009	0.701	0.029	0.020	0.035	0.936 ± 0.047
WP80→ HLT	EB	0.997	0.0001	0.989	0.003	0.001	0.0032	0.992 ± 0.003
	EE	0.988	0.0003	0.992	0.005	0.001	0.0051	1.004 ± 0.005
	EB e^-	0.997	0.0001	0.986	0.005	0.001	0.0051	0.989 ± 0.005
	EE e^-	0.988	0.0004	0.994	0.006	0.001	0.0061	1.006 ± 0.006
	EB e^+	0.996	0.0001	0.994	0.004	0.001	0.0042	0.998 ± 0.004
	EE e^+	0.989	0.0004	0.990	0.007	0.001	0.0071	1.001 ± 0.007

Sources of systematic uncertainty

- Signal Shape
  - Extend Mee window to include more of the low mass tail, 50-120 GeV
  - Construct data-driven signal shapes by tightening selection on Tag+Fail
    - Fit with these templates, difference w.r.t nominal fit is the systematic
- Background Model
  - Consider power-law ($1/M^{\alpha}$) as alternative model to exponential
  - Fix α to value found from fit to dijet data and generate pseudo-experiments
  - Fit each trial with exponential, measure bias
- Energy Scale /Resolution
  - Apply corrections ± uncertainties to the MC, measure difference in yield

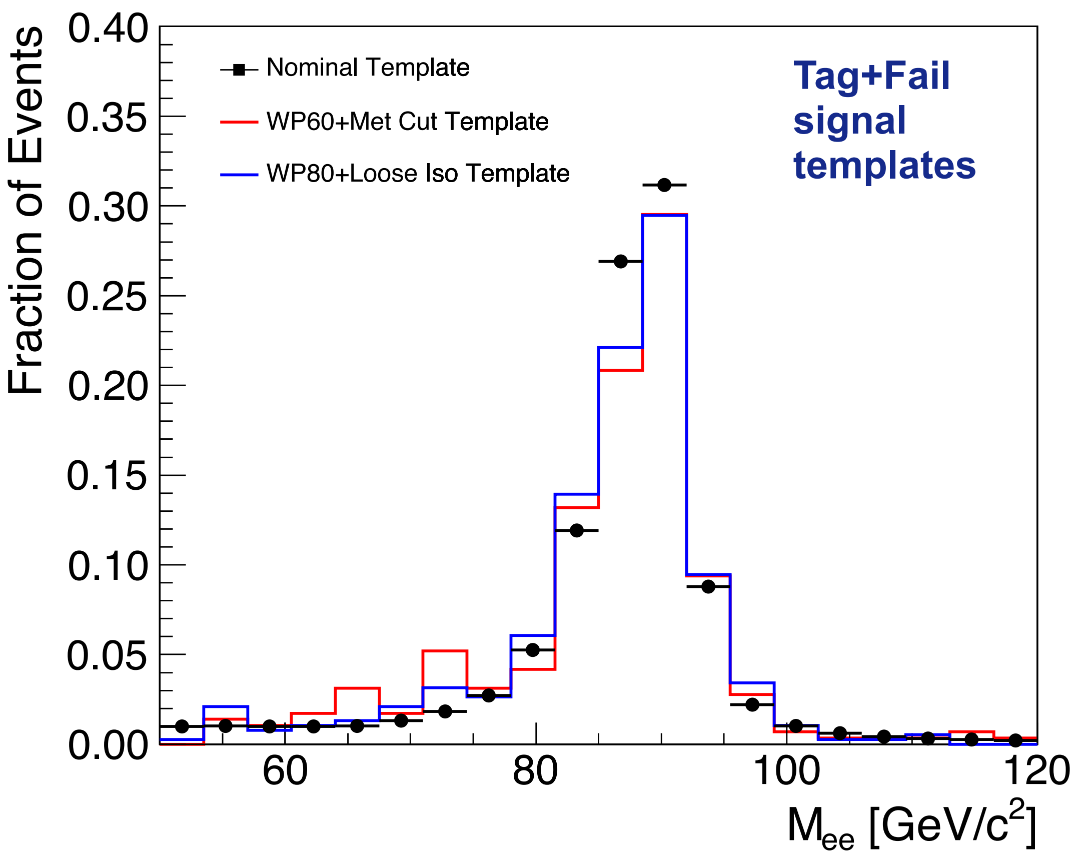
Source	% ε_{reco}	% $\varepsilon_{reco\text{-}WP95}$	% $\varepsilon_{reco\text{-}WP80}$	% $\varepsilon_{WP80\text{-}HLT}$	% $\varepsilon_{WP80\text{-}HLT}$
Background Model	0.06	0.25	0.24	0.01	< 0.00
Energy Scale	0.1	0.1	0.2	< 0.00	0.1
Signal Shape	1.2	1.0	2.0	-	-

I will focus on this in the next slides

Main culprit: Tag+Failed signal template

What was done in inclusive W,Z analysis ?

Take signal template for Tag+FailingProbe events from MC. Use data-driven template to validate the technique, and compute systematic uncertainty.



- Data driven shapes have large uncertainty,
  - •due to extrapolations involved,
  - •unknown escale & resolution for failing electrons,
  - •imprecise knowledge of the material model
  - •and of farction of showering electrons etc.

-Overall good agreement between data and MC signal templates.

-Currently try all these templates

⇒ syst.= largest deviation in eff

What is the difficulty with TF signal template ?

✦No one quite knows the energy scale or resolution for poorly reconstructed electrons

- many/most of the signal events in Tag+Fail sample have such electrons
- in presence of very high background: S/B <1 or << 1

✦Therefore, it is hard to know if the background has large slope or signal has large tail :(

✦Currently it is not possible to extract EXACT data-driven template with good purity.

- one can only extract template by using tighter tag selection
- or applying additional cleaning cuts to probe, which may or may not be unbiased
- or applying topological cuts, e. g., MET cut.

✦Thus one needs to extrapolate the data-driven template, which has statistical uncertainty associated with it

Improving signal shape systematics

Here are some possible solutions:

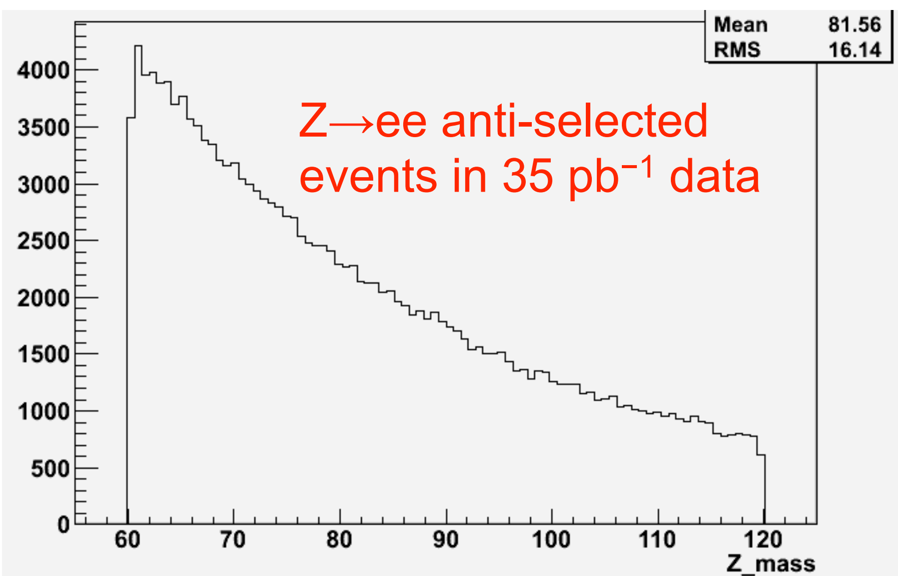
1. Improve electron selection
   - The Tag+Fail sample is dominated by background from QCD dijet events.
   - It will help a lot if we can somehow improve electron selection so that the dijet rejection is significantly higher for the same fixed signal efficiency.
2. Unbiased probe cleaning   We already exhausted this avenue.
3. More thoughtful modeling of the electron scale and resolution
   - For poorly reconstructed electrons which dominate the Tag+Fail sample.
   - There are some ideas for which we need input from POG/DPG.
4. Apply some loose isolation and Id cuts to failed sample
   - The three efficiencies (i.e., reco, selection, trigger) will no longer be uncorrelated, because the denominator of a given step will not be the numerator of the previous step, as is the case now.
   - One would need to compute correlations using pseudo experiments or full sim. This again is not scalable if one has to compute 3 efficiencies in 100 bins each.

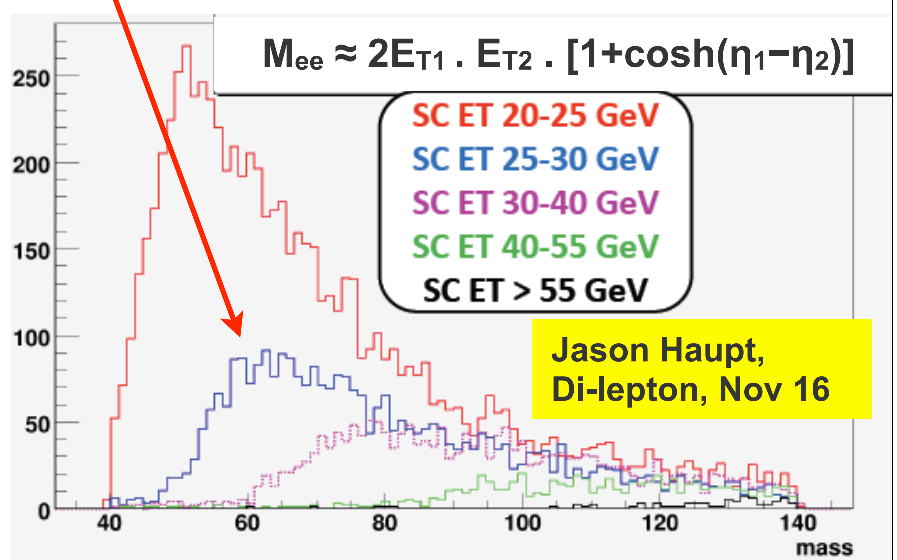
Solution #1 is preferable, #3 can be tried, #4 is messy and error-prone.

Improving background systematics

✦Essentially all the background in Tag+Fail sample come from QCD dijet.
✦From fitting the dijet data, or Z anti-selection events (where both legs fail WP95 criteria), we know that the spectrum falls as a power low ~$1/M^{\alpha}$.
-which can be well reproduced by an exponential or polynomial function in a narrow mass range (say, 60−120 GeV). We currently use all these.

✦A possible improvement would be to take the background template directly from anti-selected events. This will allow to model the kinematic sculpting introduced when computing efficiency as a function of E_T.





Backup

Will stat. error reduce by √N? Yes, but ...

✦For the inclusive W, Z analysis the stat error will go down by √N
- efficiency error in any particular E_T, η bin will be large
- however, in inclusive analysis one only needs single-bin average efficiency, so stat uncertainty will be small.

✦Differential analyses (Z p_T, Z rapidity, $W^{\pm}$ asymmetry, Z forward-backward asymmetry, ...) need efficiency binned in E_T, η, cosθ*, nJets, ..
- large statistical uncertainty in any given bin
- for ~10x bins and ~10x more data, similar error budget per measurement as is currently the case with 3 pb^{-1} in inclusive analysis
- it pays well, therefore, for people doing differential analysis to treat this uncertainty estimation carefully.

✦CMS must use full statistics to reduce uncertainty to remain competitive
- make use of the correlations to our advantage
- compute the correlations explicitly, or devise means to avoid them.

Why stat. error estimation is typically incorrect?

When tags and probes are indistinguishable, there are three classes of events:

TT: both leptons satisfy tag selection and both lie in the phase space bin (PSB) to be tested

TP: one lepton satisfies tag selection, is not in the test PSB, and the other satisfies tag selection and is in the test PSB

TF: one lepton satisfies tag selection, and the other is in the test PSB and fails tag selection.

$$\varepsilon = (2N_{TT} + N_{TP})/(2N_{TT} + N_{TP} + N_{TF})$$

$$d\varepsilon^2 = [\varepsilon^*(1-\varepsilon) / (2N_{TT} + N_{TP} + N_{TF})] * (1 + [2N_{TT}/(2N_{TT}+N_{TP})] (1-\varepsilon))$$

(counting "probes" instead of events underestimates error by a factor of up to √2, typically people multiply the uncertainty by √2 which causes the overestimation.)

-For binned efficiencies, double-tag events T_iT_j introduce pairwise statistical correlations for each pair (ε_i, ε_j)

-In order to estimate statistical CORRECTLY one must

- either track the covariance matrix of the binned efficiencies,
- or devise efficiency estimators which rely on pairwise disjoint sets of events.

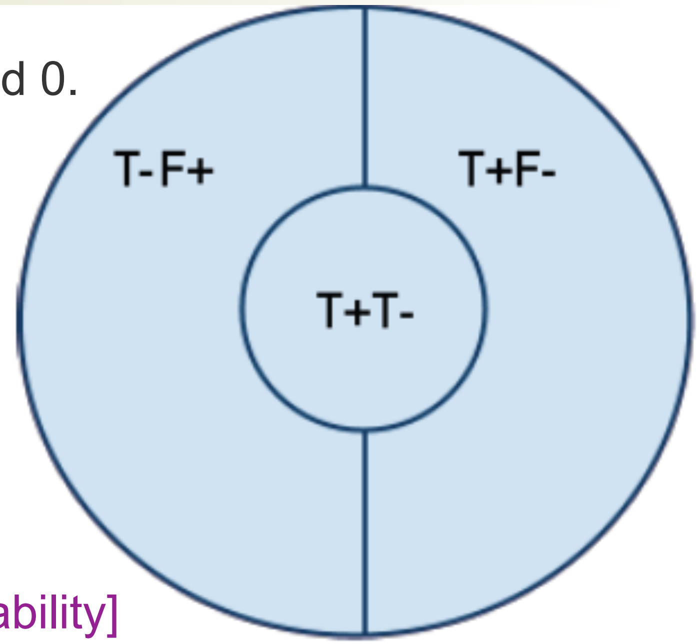
Correlation in charged efficiency estimation

In what follows, charge misid is ignored and assumed 0.
Let's examine "one-bin" case first "Greedy" design:
Count all double-tag events and all events with appropriately charged failing probes.

$\varepsilon+ = P / (P + T^{-}F^{+})$

$\varepsilon- = P / (P + T^{+}F^{-})$

$\varepsilon+$ and $\varepsilon-$ are now correlated via the common sample $P \equiv T^{+}T^{-}$. What is the effect on $\varepsilon^{+}/\varepsilon^{-}$?



$r = \varepsilon^{+}/\varepsilon^{-} = (P + F^{-}) / (P + F^{+})$ [omitting T for readability]

$dr^2 = [(F^{+} - F^{-})^2\, dP^2 + (P+F^{-})^2\, (dF^{+})^2 + (P+F^{+})^2\, (dF^{-})^2] / (P+F^{+})^4$

In the limit of charge symmetry and no background ($dN \sim \sqrt{N}$)

$dr^2 = 2 * dF^2 / (P+F)^2$, for $F = F^{+} = F^{-}$

If you naively ignore the correlation and compute $dr^2/r^2 = (d\varepsilon^{+}/\varepsilon^{+})^2 + (d\varepsilon^{-}/\varepsilon^{-})^2$, then you get

$dr^2 = 2*dF^2/(P+F)^2 * (1/\varepsilon)$

i.e., variance is OVER-estimated by ε ($= \varepsilon+ = \varepsilon-$)

So, in order not to inflate stat error one needs to take into account the correlations.

Bin-to-bin correlations

✦Unlike the charge-blind efficiencies, there are no pair-wise correlations between $\varepsilon^{\pm}_{i}$ and $\varepsilon^{\pm}_{j}$ due to double-tag events.

✦There are still correlations between ε^{+}_{i} and ε^{-}_{j} because they recycle the same passing events.

✦Those correlations are identical to the one-bin case just considered, i.e. propagating errors naively for + and − separately overcovers by 5-10%.

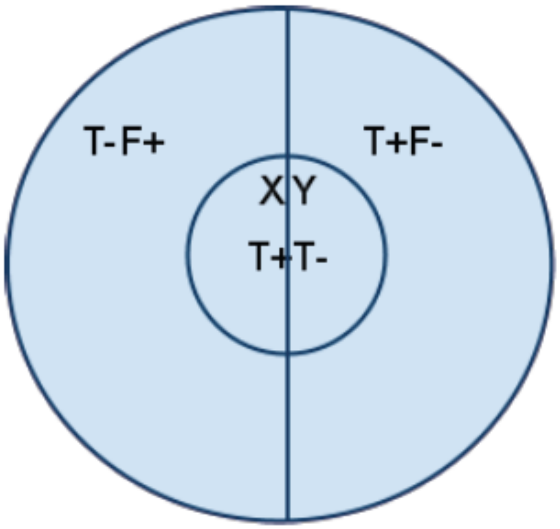
What if you split samples to avoid correlation ?

"Split" design: Randomly assign half of double tags to + eff's and half to − eff's

$\varepsilon^+ = P_X / (P_X + F^+/2)$, $\varepsilon^- = P_Y / (P_Y + F^-/2)$

No correlations, but half the stats for passing probes.

NB: Failing probes have not been split in half, and thus the factor 2 in the eff formulae



$(d\varepsilon^+)^2 = [2F^{+2}\, dP^2 + P^2\, dF^{+2}]/(P + F^+)4$ split case

$(d\varepsilon^+)^2 = [F^{+2}\, dP^2 + P^2\, dF^{+2}]/(P + F^+)4$ greedy case

In the limit of no background the variance ratio between the two cases favors the greedy design by a factor 2−ε. What is the impact on the efficiency ratio though?

$r = \varepsilon+/\varepsilon- = [2P_X *(2P_Y + F^-)] / [2P_Y*(2P_X + F^+)]$

$dr^2_{\ split} = dr^2_{\ greedy} + 8(1-\varepsilon)2 * dP^2/P^2$

Moral of the story

- Using greedy case to obtain tag and probe efficiency ratios provides superior errors.
- Naively propagating +/- errors overcovers somewhat.
- A simple modification to tag and probe fit (three categories instead of two) can get the extra 5-10% savings in charged efficiency measurement.