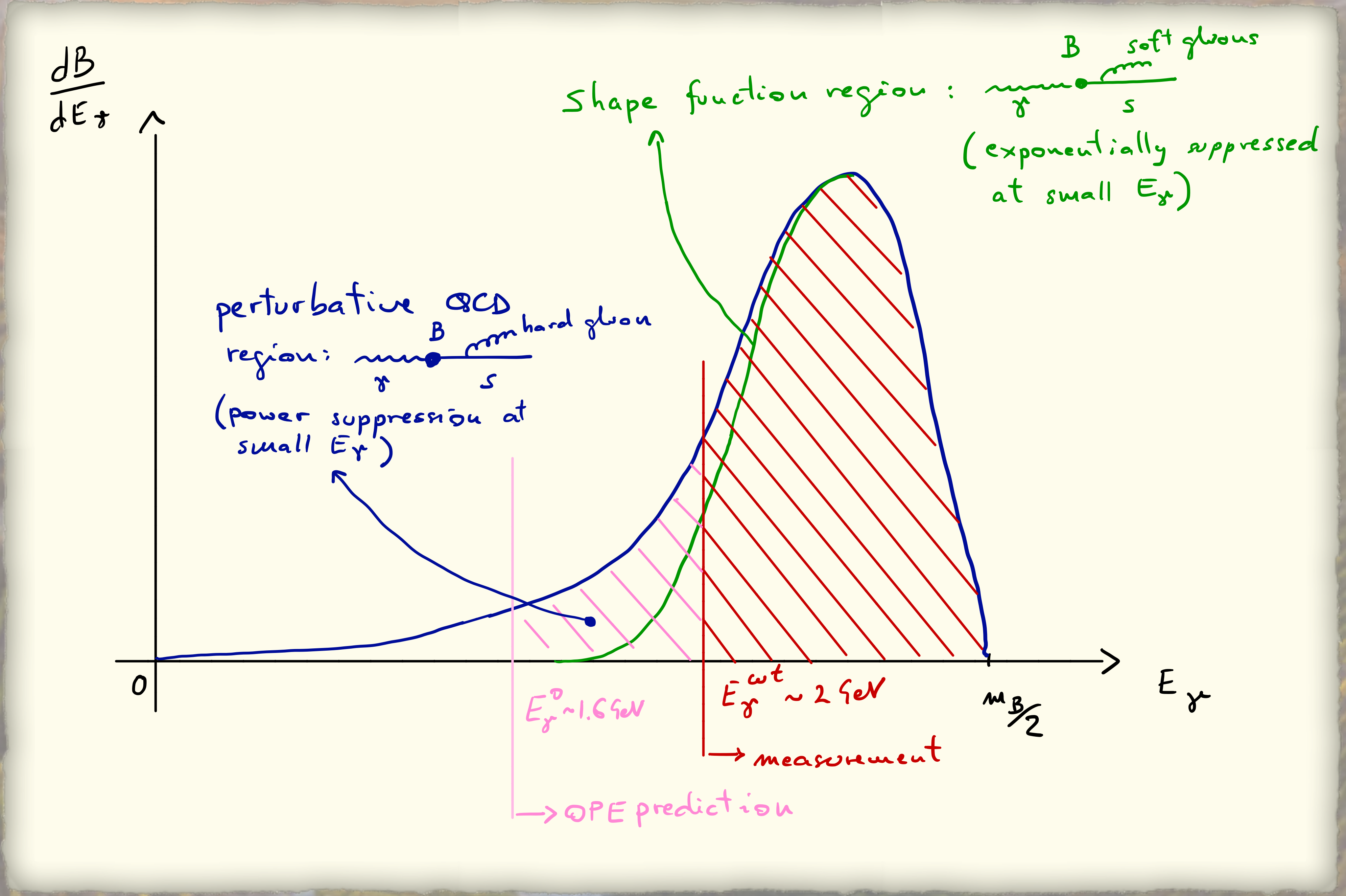
$B \to X_s\gamma$ with Neural Networks†

Enrico Lunghi
Indiana University

Challenges in Semileptonic B Decays
Apr 19-23 2022, Barolo, Italy



†Progress report of work done in collaboration with P. Gambino, M. Misiak, S. Schacht

Motivations

- $B \to X_s \gamma$
  - OPE at NNLO for $E_\gamma^{\rm th} > 1.6$ GeV
  - Spectrum at NNLO: depends on a the (mostly) unknown B-meson Shape Function
  - Experimental data requires $E_\gamma^{\rm exp} > [1.6\text{-}1.9]$ GeV (most precise at large E_γ)

- Fits to the spectrum allow to
  - improve the extrapolation down to 1.6 GeV
  - extract information on m_b and various hadronic matrix elements (on which we already have information from $B \to X_c \ell \nu$)
  - extract the Shape Function itself

- Knowledge of the Shape function is essential to $B \to X_u \ell \nu$ and to the m_{X_s} extrapolation in $B \to X_s \ell \ell$

Motivations

- $B \to X_s\gamma$
  - OPE at NNLO for $E_\gamma^{\rm th} > 1.6$ GeV
  - Spectrum at NNLO: depends on a the (mostly) unknown B-meson Shape Function
  - Experimental data requires $E_\gamma^{\rm exp} >$ [1.6-1.9] GeV (most precise at large E_γ)
- Fits to the spectrum allow to
  - improve the extrapolation down to 1.6 GeV
  - extract information on m_b and various hadronic matrix elements (on which we already have information from $B \to X_c\ell\nu$)
  - extract the Shape Function itself
- Knowledge of the Shape function is essential to $B \to X_u\ell\nu$ and to the m_{X_s} extrapolation in $B \to X_s\ell\ell$
- In this talk the CLN parameterization will not be mentioned

$B \to X_s\gamma$ with a Wilsonian cutoff

- We work in the **kinetic Scheme** for the b mass and all hadronic parameters: $m_b,\ \mu_\pi^2,\ \mu_G^2,\ \rho_D^3,\ \rho_{LS}^3$ and the Shape Function

- Introduce a **Wilsonian cutoff** $\mu \sim 1\ \mathrm{GeV}$
  - Virtual and real gluons with energy larger than μ are calculated in perturbation theory
  - Gluons with $E_g < \mu$ are intrinsically non-perturbative and are included via the above mentioned non-perturbative parameters and in the Shape Functions

- This approach is rigorous and has been used to the
  - construction of the widely used kinetic scheme for m_b and other hadronic parameters from various small-velocity $B \to X_{u,c}\ell\nu$ sum-rules
  - calculation of moments of the $B \to X_s\gamma$ spectrum
  - extraction of $V_{cb},\ m_b$ and other hadronic parameters from $B \to X_c\ell\nu$
  - extraction of V_{ub} from $B \to X_u\ell\nu$

References

- **Wilsonian cutoff in B physics and kinetic scheme:**
  Bigi, Shifman, Uraltsev, Vainshtein, hep-h/9312359
  Bigi, Shifman, Uraltsev, Vainshtein, hep-ph/9402360
  Bigi, Shifman, Uraltsev, Vainshtein, hep-ph/9405410
  Bigi, Shifman, Uraltsev, Vainshtein, hep-ph/9312359
  Uraltsev, hep-ph/9610425
  Bigi, Shifman, Uraltsev, hep-ph/9703290
  Uraltsev, hep-ph/0010328

- $B \to X_s \gamma$ **moments:**
  Bigi, Uraltsev, hep-ph/0308165
  Benson, Bigi, Uraltsev, hep-ph/041080

- $B \to X_c \ell \nu$ **fits:**
  Benson, Bigi, Mannel, Uraltsev, hep-ph/0302262

- $B \to X_u \ell \nu$ **fits:**
  Gambino, Giordano, Ossola, Uraltsev, 0707.2493

Master Formula

- We write the spectrum as ($\mu \simeq 1\ \mathrm{GeV}$ is the Wilsonian cutoff):

$$\frac{d\Gamma}{dE_\gamma} = \int dk_+ \, f(k_+, \mu) \frac{d\Gamma^{pert}}{dE_\gamma}\left(E_\gamma - \frac{k_+}{2}, \mu\right)$$

$$= \Gamma_0 \sum_{i \le j = 1}^{8} C_i^{\mathrm{eff}\,*}(\mu_b) C_j^{\mathrm{eff}}(\mu_b) \int_{-\infty}^{\lambda} d\kappa \, F(\kappa, \mu) W_{ij}^{pert}(\xi - \kappa, \mu, \mu_b)$$

Shape Function in the kinetic scheme

where $F(\kappa, \mu) = m_b \, f(m_b \kappa, \mu)$

$$m_b = m_b^{\mathrm{kin}}(\mu)$$

$$\xi = 2E_\gamma / m_b$$

$$\lambda = (m_B - m_b)/m_b$$

$$\Gamma_0 = \frac{G_F^2 \alpha m_b^2 m_b^{\overline{\mathrm{MS}}}(\mu_b)^2}{16\pi^4} |V_{tb} V_{ts}^*|^2$$

Master Formula

- For instance, the $|C_7|^2$ hard scattering kernel has the following structure:

virtual corrections ($E_g > \mu$): 0th moment (total rate)

$$W_{77}^{pert}(\xi,\mu,\mu_b) = \left[1 + \frac{\alpha_s(\mu_b)}{\pi} C_F \left(V^{(1)} + \beta_0 \frac{\alpha_s}{\pi} V^{(2,\mathrm{BLM})} + \frac{\alpha_s}{\pi} V^{(2)} \right)\right] \delta(1-\xi)$$

1st moment

$$+ \frac{\alpha_s(\mu_b)}{\pi} C_F \left[B^{(1)} + \beta_0 \frac{\alpha_s}{\pi} B^{(2,\mathrm{BLM})} + \frac{\alpha_s}{\pi} B^{(2)} \right] \delta'(1-\xi)$$

2nd moment

$$+ \frac{1}{2} \frac{\alpha_s(\mu_b)}{\pi} C_F \left[C^{(1)} + \beta_0 \frac{\alpha_s}{\pi} C^{(2,\mathrm{BLM})} + \frac{\alpha_s}{\pi} C^{(2)} \right] \delta''(1-\xi)$$

gluon Bremsstrahlung ($E_g > \mu$)

$$+ \frac{\alpha_s(\mu_b)}{\pi} C_F \left[H^{(1)}(\xi,\mu) + \beta_0 \frac{\alpha_s}{\pi} H^{(2,\mathrm{BLM})}(\xi,\mu) + \frac{\alpha_s}{\pi} H^{(2)}(\xi) \right] \theta(1-\xi)\theta(\xi)$$

- Virtual contributions can be either calculated directly or extracted by requiring that each moment matches the OPE result

Master Formula

- **Soft-gluon contributions (with $E_g < 1\ \mathrm{GeV}$) have been absorbed into the SF**
- The moments of the photon spectrum depend only on corresponding moments of the SF. For instance:

$$\int d\xi \int d\kappa F(\kappa) W_{77}^{pert}(\xi - \kappa) = \int d\kappa\, F(\kappa) \int d\xi W_{77}^{pert}(\xi - \kappa) = \int d\kappa\, F(\kappa) \int d\xi' W_{77}^{pert}(\xi')$$

total rate at NNLO

$$= 1 + O(1/m_b^2)$$

- Power corrections in the OPE for the moments are reproduced by appropriate matching conditions for the moments of the Shape Function. For instance, we find:

$$\int d\kappa\, F(\kappa, \mu) = 1 - \frac{\mu_\pi^2 + 3\mu_G^2}{2m_b^2} - \frac{11\rho_D^3 - 9\rho_{LS}^3}{6m_b^3} + \frac{\alpha_s}{\pi}\left[C_F\left(\frac{4\eta}{3} + \frac{3\eta^2}{4} + \frac{11\eta^3}{18}\right)\frac{\mu_\pi^2}{m_b^2} + \left(\frac{67}{36} - \frac{26\pi^2}{27} - \frac{9L_b}{4} + \frac{16\eta}{3} + 3\eta^2 + \frac{22\eta^3}{9}\right)\frac{\mu_G^2}{m_b^2}\right]$$

Master Formula

- We include effects of all other operators in a similar way.
- For instance the 78 kernel is:

$$W_{78}^{pert}(\xi,\mu,\mu_b) = C_F \frac{\alpha_s(\mu_b)}{\pi}\Big[\Big(V_{78}^{(1)} + \frac{\alpha_s}{\pi}V_{78}^{(2)}\Big)\delta(1-\xi) + \frac{\alpha_s}{\pi}B_{78}^{(2)}\delta'(1-\xi) + \frac{1}{2}\frac{\alpha_s}{\pi}C_{78}^{(2)}\delta''(1-\xi)$$
$$+\Big(H_{78}^{(1)}(\xi,\mu) + \frac{\alpha_s}{\pi}H_{78}^{(2)}(\xi,\mu)\Big)\theta(1-\xi)\theta(\xi)\Big]$$

The spectrum is obtained by convoluting this kernel with *same* Shape Function introduced for 77 because soft and collinear end-point singularities factorize

Korchemski and Sterman, hep-ph/9902341
Akhoury and Rothstein, hep-ph/9512303

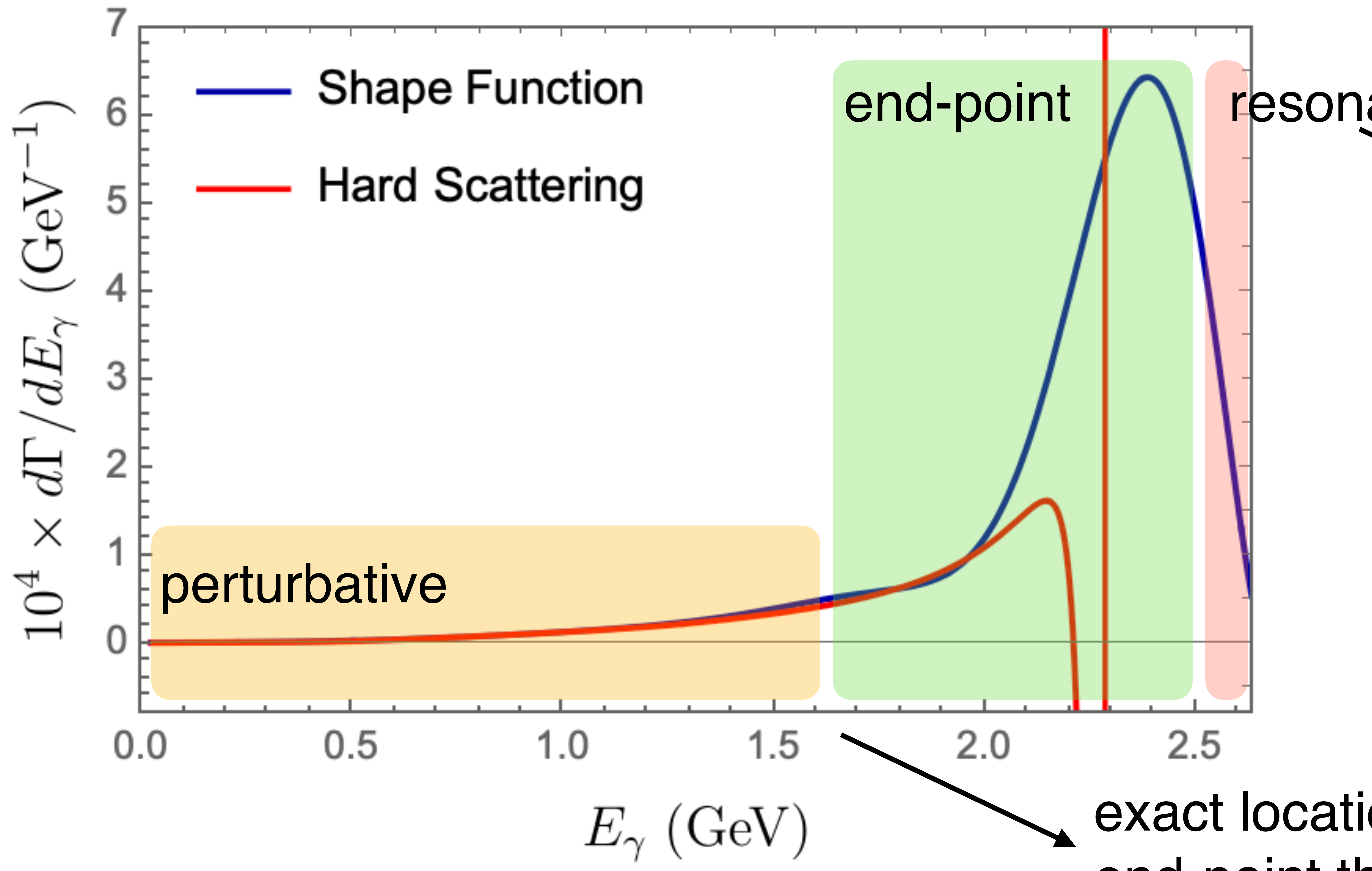
- For all other kernels we have only BLM contributions at NNLO:

$$W_{ij}^{pert}(\xi,\mu,\mu_b) = C_F \frac{\alpha_s(\mu_b)}{\pi}\Big[\Big(V_{ij}^{(1)} + \beta_0\frac{\alpha_s}{\pi}V_{ij}^{(2\text{BLM})}\Big)\delta(1-\xi) + \Big(H_{ij}^{(1)}(\xi,\mu) + \beta_0\frac{\alpha_s}{\pi}H_{ij}^{(2\text{BLM})}(\xi,\mu)\Big)\theta(1-\xi)\theta(\xi)\Big]$$

- Resolved contributions are not present in this formalism and need to be taken into account separately [Lee, Neubert, Pax, hep-ph/0609224]

Master Formula

- Shape Function vs Hard Scattering spectra:



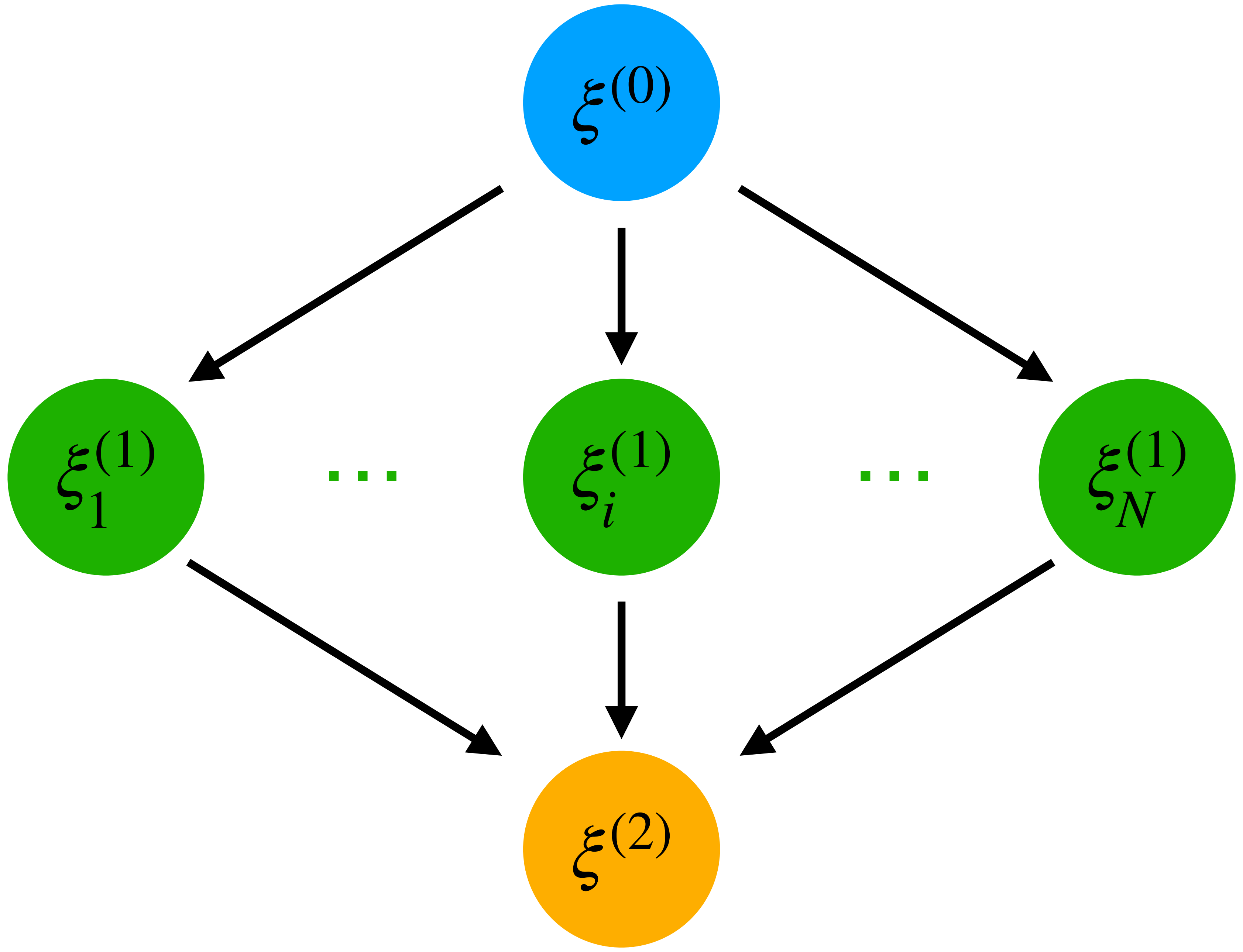
no Shape Function description: need to integrate over large enough E_γ range

exact location of the perturbative/ end-point threshold depends on Shape Function

Shape Function with a Neural Network

[Gambino, Healey, Mondino, 1604.07598]

- Neural Networks can be used to provide **unbiased** parameterizations of continuous functions
- We consider a NN with one input ($\xi^{(0)}$), one layer with N nodes ($\xi_i^{(1)}$) and a single output ($\xi^{(2)}$)



$$\kappa = \xi^{(0)}$$

$$\downarrow$$

weights thresholds

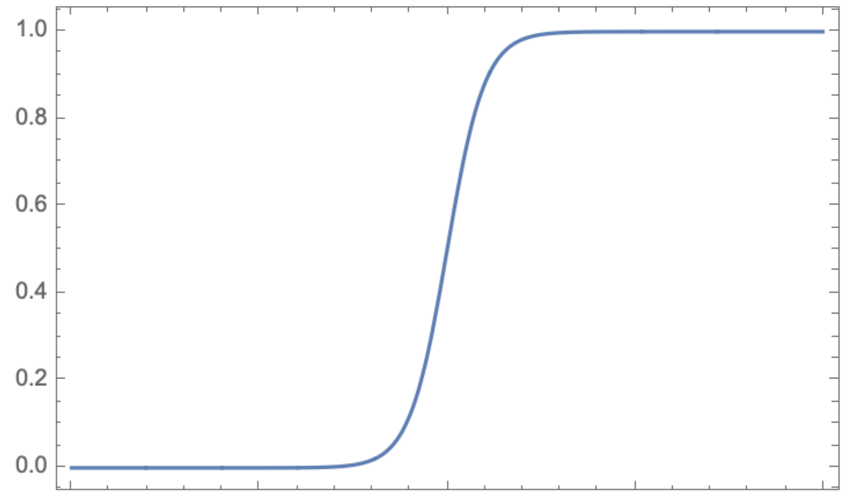
$$\xi_i^{(1)} = g(\xi^{(0)} w_i^{(1)} - \theta_i^{(1)})$$

$$\downarrow$$

3N+1 parameters

$$\xi^{(2)} = g(\sum_{i=1}^{N} \xi_i^{(1)} w_i^{(2)} - \theta^{(2)}) = N(\{w,\theta\},\kappa)$$

- g is a non-linear activation function. We adopt a sigmoid: $g(x) = \dfrac{1}{1+e^{-x}}$



Shape Function with a Neural Network

- The actual Shape Function we consider contains an *underlying function* whose purpose is to help speeding the training without (possibly) introducing a strong bias

- Finally we impose few further conditions: positivity of the overall Shape Function and an explicit cut-off ($\kappa < \lambda$)

- For instance we consider:

$$F(\kappa) = \bar{w}_1 e^{\bar{w}_2 \kappa} (\lambda - \kappa)^{\bar{w}_3} \; \Big| N(\{w, \theta\}, \kappa) \Big| \; \theta(\lambda - \kappa)$$

- The extra parameters $\bar{w}_i$ (with $\bar{w}_1 > 0$) are treated on equal footing with the other weights

- In the preliminary tests we have run we considered N=6, implying 22 adjustable weights

Training

- We adopt a **genetic algorithm**
- Each full iteration begins with a “parent” SF with random weights and ends with a “child” SF for which a χ^2 comprised of experimental (BaBar,Belle) and theoretical (SF moments) information is below a given threshold.
- Each successful iteration yields a replica: a possible Shape Function which is fully compatible with both experimental and theoretical constraints.
- There are many (important) details related to
  - **overtraining** (e.g. divide χ^2 into training and validation)
  - **avoiding local minima**
- Distribution of replicas is not expected to be strongly affected/biased by minor changes in the chosen training scheme

χ^2 calculation

- Training a Neural Network requires an enormous number of χ^2 evaluations
- Binned experimental results are presented in the B and $\Upsilon(4S)$ rest frames and require multiple numerical integrations
- An excellent solution is **approximating the NN Shape Function with a cubic spline**.
- We found that dividing the $\kappa \in [-1,\lambda]$ into 50 intervals (with $4 \times 50 =$ 200 spline coefficients) yields $0.1\ \%$ approximations to realistic Shape Functions and to their corresponding spectra and integrated observables
- All observables can be easily calculated for each of the 200 basic spline building blocks. The χ^2 can then be easily obtained by contracting the latter with the 200 spline coefficients

χ^2 calculation

- An alternative approach we considered is projecting the NN Shape Function onto a (hopefully) small set of basis functions
- As an example we tried a basis inspired by the one adopted by the SIMBA collaboration:

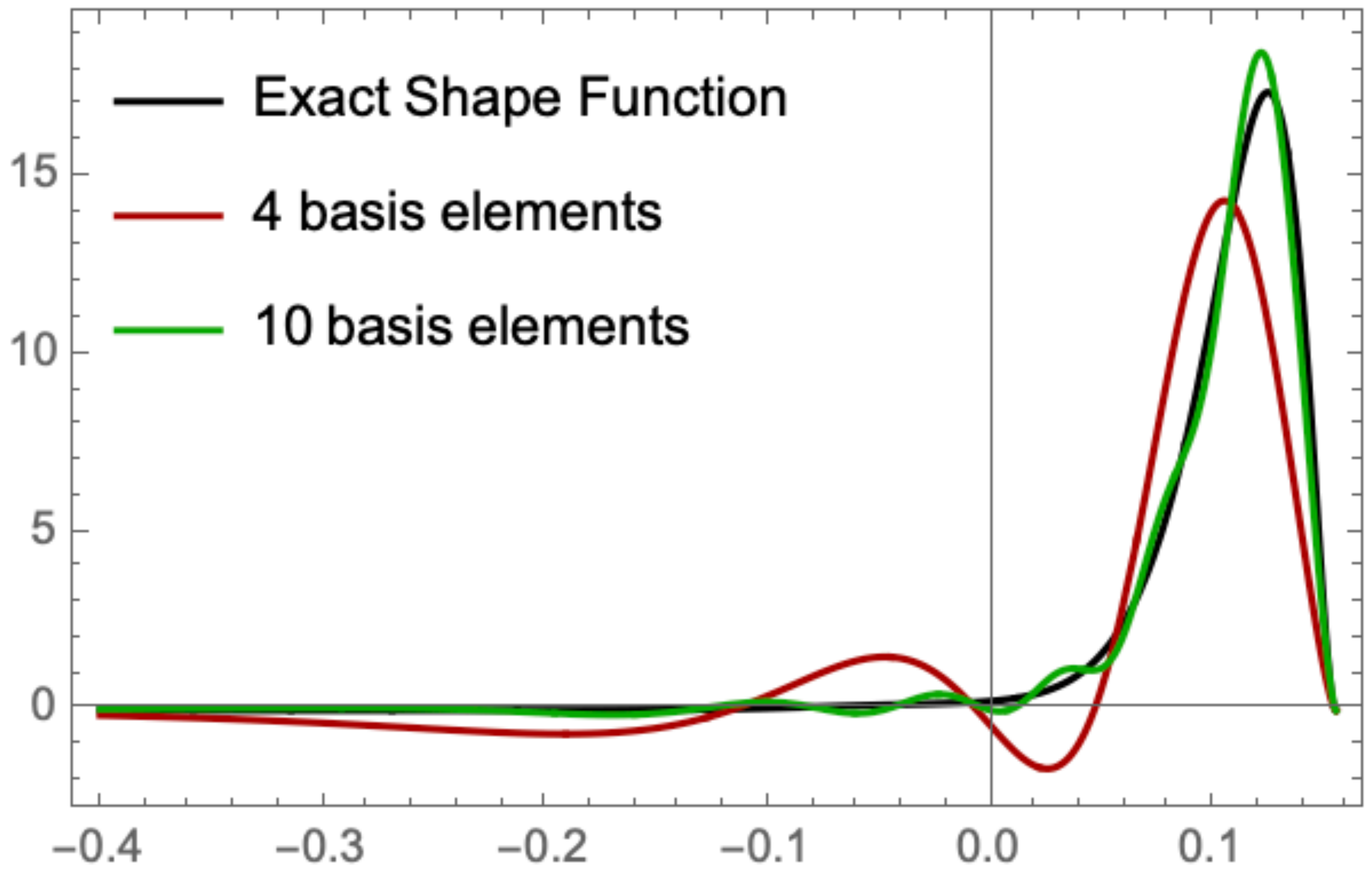
$$F(\kappa) = \sum_{n=0}^{\infty} c_n f_n(1 - \kappa/\lambda)$$

$$f_n(x) = \sqrt{Y(x)}\sqrt{2n+1}\ P_n\left[y(x)\right]$$

$$Y(x) = \frac{128}{3} x^3 e^{-4x} \qquad y(x) = -1 + 2\int_0^x Y(w)\,dw$$

Shape Functions with this "typical" spread are approximated by few basis elements

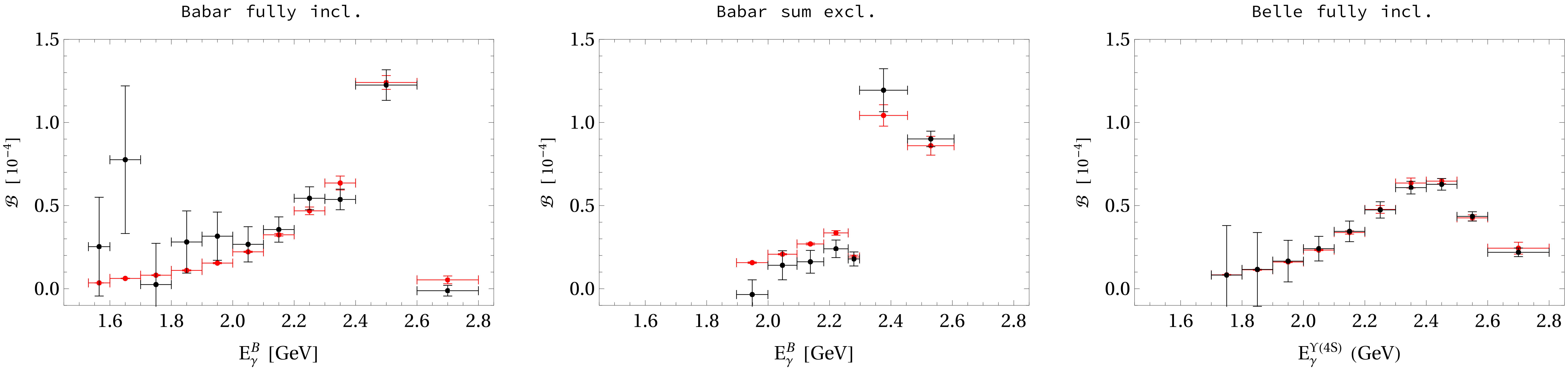
- Need many basis functions to reasonably approximate SFs with widths different from the "built-in" one
- Basis elements with large n yield highly oscillating spectra and usually partake into delicate cancellations

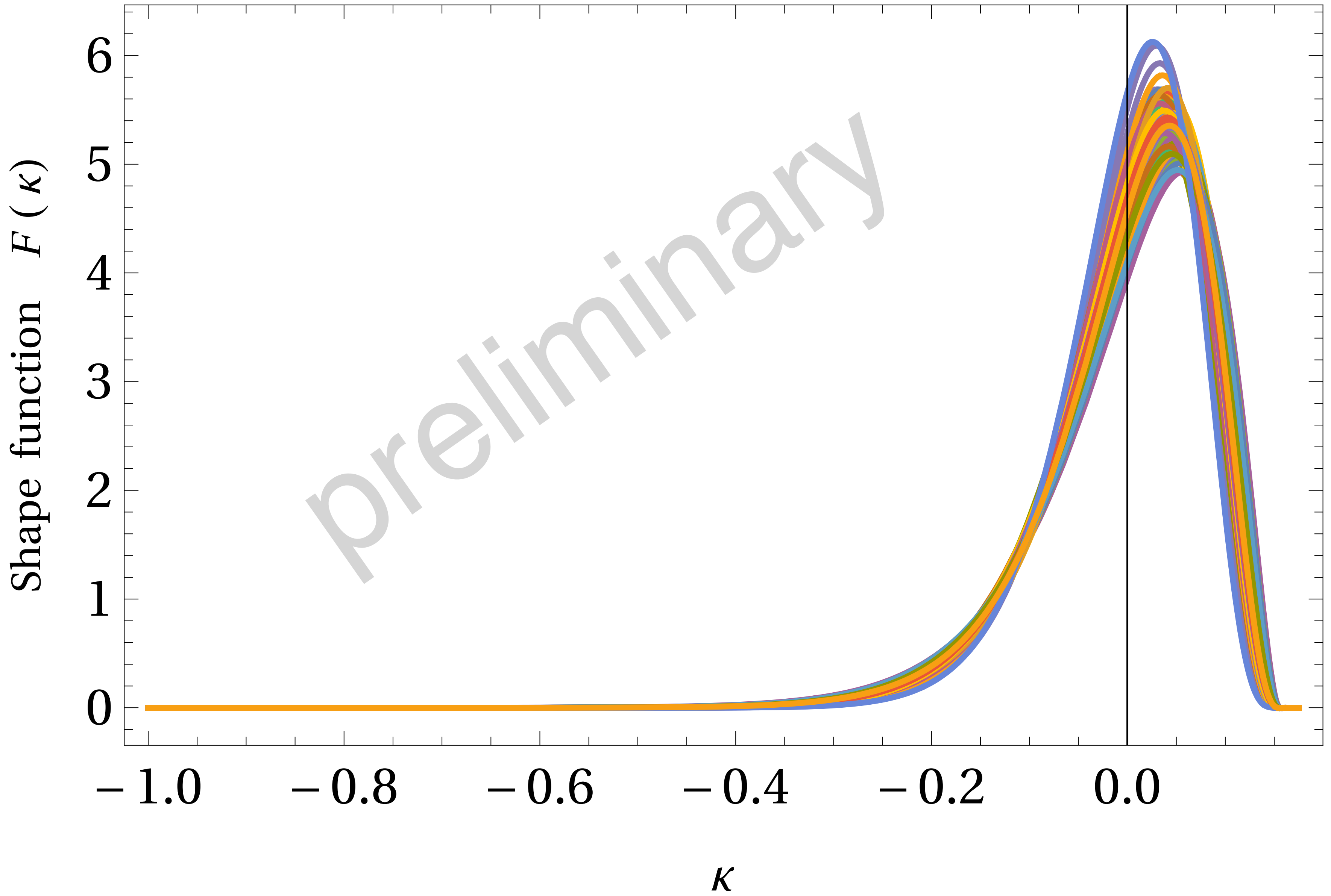


Experimental data

- We considered data from 2012 BaBar fully inclusive and sum over exclusive analyses (in the B rest frame) and 2016 Belle results (in the $\Upsilon(4S)$ rest frame):

Some very preliminary results





- Some NNLO corrections missing
- Still working on training
- …

Comparison with SCET approach

[SIMBA collaboration: Bernlochner, Lacker, Ligeti, Stewart, F. Tackmann, K. Tackmann]

- Within SCET the separation of soft and hard-collinear scales is achieved within an effective theory framework (based on the method of regions and dim-reg): this corresponds to a different scheme for the (leading) SF

- At leading power the SCET SF is universal for all inclusive B decays. At subleading power several additional universal SCET SFs appear. Unfortunately $B \to X_s\gamma$, $B \to X_u\ell\nu$ and $B \to X_s\ell\ell$ depend on different combinations of leading and subleading SFs.

- **Effectively both the kinetic and SCET approach extract from $B \to X_s\gamma$ one effective SF which can only be used judiciously (i.e. with added uncertainties) in other inclusive B decays**

- Within SCET it is possible to resum hard-collinear logs. We have $Q_h \sim m_b = 4.57$ GeV, $Q_{hc} \sim \sqrt{m_b(m_B - m_b)} = 1.8$ GeV and $Q_s \sim (m_B - m_b) = 0.71$ GeV. The relevant logs are therefore quite small: $\log[Q_h/Q_{hc}] = \log[Q_{hc}/Q_s] = 0.93$

- I conclude that these two approaches are effectively equivalent and offer a complementary approach to a simultaneous analysis of inclusive radiative and (rare-)semileptonic B decays

Implications for $B \to X_s \ell\ell$

- SF needed for extrapolation in m_{X_s} and to improve the EvtGen Monte Carlo event generator which is the heart of Belle, BaBar and Belle II analyses.

[EvtGen: Ryd, Lange, Kuznetsova, Versille, Rotondo, Kirkby, Wuerthwein, Ishikawa; Maintained by J. Back, M. Kreps and T. Latham at University of Warwick]

- Hadronic spectrum is based on the Fermi motion implementation presented in Ali, Hiller, Handoko, Morozumi hep-ph/9609449:

$$\frac{d\Gamma_B}{ds\,du\,dp} = \int du' \frac{m_b(p)^2}{m_B} p \left[\frac{4}{\sqrt{\pi p_F^3}} \exp(-p^2/p_F^2)\right] \left(u'^2 + 4 m_b(p)^2 s\right)^{-1/2} \left[\frac{d\Gamma_b}{ds\,du}\right]_{m_b \to m_b(p)}$$

parton level with momentum dependent b mass

- We need to urgently update the code!
- Work in progress on the complete triple differential rate at $O(\alpha_s)$

[T. Huber, T. Hurth, J. Jenkins, EL, in preparation]

```
{
  pb = _calcprob->FermiMomentum(_pf);

  // effective b-quark mass
  mb = mB*mB + _mq*_mq - 2.0*mB*sqrt(pb*pb + _mq*_mq);
  if ( mb>0. && sqrt(mb)-_ms  < 2.0*ml ) mb= -10.;
}
mb = sqrt(mb);
```