Towards the next continuum limit of ε'

Non-perturbative renormalization and software optimization

Jonas Hildebrand

Lattice 2026

July 26 – August 1, 2026

University of Connecticut

The RBC & UKQCD collaborations

BNL & RBRC
Peter Boyle
Taku Izubuchi
Chulwoo Jung
Christopher Kelly
Patrick Oare
Amarjit Soni (emeritus)
Xin-Yu Tuo
Tianle Wang
Shuhei Yamamoto

CERN
Matteo Di Carlo
Felix Erben
Andreas Jüttner (Southampton)

Columbia University
Ryan Abbott
Norman Christ
Jonah Eick
Rajiv Eranki
Sarah Fields
Ceran Hu
Yikai Huo
Erik Lundstrum
Robert Mawhinney

University of Connecticut
Thomas Blum
Jonas Hildebrand
Luchang Jin
Vaishakhi Moningi
Anton Shcherbakov
Douglas Stewart
Joshua Swain

DESY Zeuthen
Raoul Hodgson

University of Edinburgh
Matthew Black
Luigi Del Debbio
Vera Gülpers
Maxwell T. Hansen
Nils Hermansson Truedsson
Ryan Hill
Rajnandini Mukherjee
Antonin Portelli
Teseo San Jose

Johannes Gutenberg University of Mainz
Alessandro Barone

KEK
Shigemi Ohta (emeritus)

Liverpool Hope
Nicolas Garron (Liverpool)

University of Liverpool
Tobias Tsang
Aurelia Zuchanke

Massachusetts Institute of Technology
En-Hung Chao

University of Milano Bicocca
Francesca A. Bresciani
Mattia Bruno
Gabriele Morandi

Peking University
Xu Feng

University of Regensburg
Daniel Knüttel
Christoph Lehner
Julian Parrino
Andreas Völklein
Haoyu Wang

RIKEN R-CCS
Masaaki Tomii

University of Siegen
Oliver Witzel

University of Southampton
Bipasha Chakraborty
Ahmed Elgaziari
Jonathan Flynn
Ollie Heald
Joe McKeon
Chris Sachrajda

Stony Brook University
Sergey Syritsyn (RBRC)

Kaons and CP violation

CP eigenstates: $\boxed{K_1} = \dfrac{K^0 + \bar{K}^0}{\sqrt{2}}$, $\quad \boxed{K_2} = \dfrac{K^0 - \bar{K}^0}{\sqrt{2}}$

CP even $\qquad$ CP odd

Mass eigenstates: $|K_S\rangle = |K_1\rangle + \varepsilon|K_2\rangle \qquad |K_L\rangle = |K_2\rangle - \varepsilon|K_1\rangle$

$$|K_L\rangle \xrightarrow{0.2\%} |\pi\pi\rangle$$

$$|K_L\rangle \xrightarrow{31.1\%} |\pi\pi\pi\rangle$$

Indirect CP violation: $|K_1\rangle \to |\pi\pi\rangle \ \sim \varepsilon$

Direct CP violation: $|K_2\rangle \to |\pi\pi\rangle \ \sim \varepsilon'$

$$\frac{\Gamma(K_L \to \pi^0\pi^0)}{\Gamma(K_S \to \pi^0\pi^0)} \Bigg/ \frac{\Gamma(K_L \to \pi^+\pi^-)}{\Gamma(K_S \to \pi^+\pi^-)} = 1 - 6\,\mathrm{Re}\left(\frac{\varepsilon'}{\varepsilon}\right)$$

$$\mathrm{Re}\left(\frac{\varepsilon'}{\varepsilon}\right)_{\mathrm{exp}} = 16.6(2.3) \times 10^{-4}$$

[NA48 & KTeV]

$K \to \pi\pi$ amplitudes and ε'

Pion scattering phase shifts at m_K

$$\varepsilon' = \frac{i\omega e^{i\boxed{\delta_2-\delta_1}}}{\sqrt{2}}\left[\frac{\mathrm{Im}(A_2)}{\mathrm{Re}(A_2)} - \frac{\mathrm{Im}(A_0)}{\mathrm{Re}(A_0)}\right] \qquad \omega = \frac{\mathrm{Re}(A_2)}{\mathrm{Re}(A_0)}$$

Lellouch–Lüscher finite volume factor

Renormalization matrix Lattice QCD and pQCD

$$A_I = \boxed{F}\frac{G_F}{\sqrt{2}}V_{us}^* V_{ud}\sum_{ij}\boxed{[z_i(\mu)+\tau y_i(\mu)]}\boxed{Z_{ij}(\mu)}\boxed{\langle(\pi\pi)_I|Q_j^{\mathrm{lat}}|K\rangle}$$

Wilson coefficients pQCD

Hadronic matrix elements Lattice QCD

Physical point ensembles

Iwasaki gauge action

$2 + 1$f Möbius domain wall fermions

48I

- $48^3 \times 96 \times 24$
- $a^{-1} = 1.73$ GeV
- $L = 5.5$ fm
- $m_\pi = 139$ MeV

64I

- $64^3 \times 128 \times 12$
- $a^{-1} = 2.359$ GeV
- $L = 5.4$ fm
- $m_\pi = 139$ MeV

96I

- $96^3 \times 192 \times 12$
- $a^{-1} = 2.71$ GeV
- $L = 6.9$ fm
- $m_\pi = 135$ MeV

Three-point continuum limit directly attacks finite lattice-spacing systematic error.

Part 1: Contraction software optimization

A2A propagators

We write the quark propagator in two parts: the low mode part, a spectral decomposition truncated after $N_{\rm ev}$ low modes, and a high mode part using a deflated noise-source solve, where we subtract off the deflation guess:

$$D^{-1}(y,x) = D^{-1}_{\rm low}(y,x) + D^{-1}_{\rm high}(y,x) = \sum_{i=1}^{N_{\rm ev}} \frac{u_i u_i^\dagger}{\lambda_i} + \sum_{h=1}^{N_{\rm hit}} D^{-1}_{\rm defl}\, \eta_h\, \eta_h^\dagger$$

$$D^{-1}_{\rm defl} = D^{-1} - \sum_{j=1}^{N_{\rm ev}} \frac{u_j u_j^\dagger}{\lambda_j}$$

The η_h come from $N_{\rm hit}$ random, independent noise vectors diluted in spin, colour and time indices,

$$\left\{ \left(\eta^{(1)}_{[1]}, \ldots, \eta^{(1)}_{[N_{\rm hit}]}\right), \ldots, \left(\eta^{(N_d)}_{[1]}, \ldots, \eta^{(N_d)}_{[N_{\rm hit}]}\right) \right\} \qquad \text{with} \quad \langle\langle \eta_{[r]}(x)\, \eta^\dagger_{[r]}(y) \rangle\rangle = \delta_{x,y} \quad ; \quad \eta^{(i)}_{[r]}(x)\eta^{(j)\dagger}_{[s]}(y) = 0$$

Traditionally we use a single hit ($\mathcal{O}(1000)$ sources), see Masaaki Tomii's talk for a multi-hit study for statistical error reduction.

A2A propagators

The A2A propagator has a similar structure between low and high parts:

$$D_{\mathrm{A2A}}^{-1} = \sum_{i=1}^{N_{\mathrm{ev}}} \frac{u_i u_i^\dagger}{\lambda_i} + \sum_{h=1}^{N_{\mathrm{hit}}} \underbrace{D_{\mathrm{defl}}^{-1} \eta_h}_{\psi} \eta_h^\dagger$$

Thus we can construct hybrid A2A vectors

$$w_i = \left\{ \frac{u^{(1)}}{\lambda_1}, \dots, \frac{u^{(N_{\mathrm{ev}})}}{\lambda_{N_{\mathrm{ev}}}}, \eta^{(1)}, \dots, \eta^{(N_d)} \right\}, \qquad v_i = \left\{ u^{(1)}, \dots, u^{(N_{\mathrm{ev}})}, \psi^{(1)}, \dots, \psi^{(N_d)} \right\}$$

such that the A2A propagator can be written

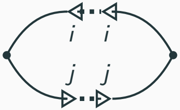
$$D^{-1}(x, y) = \sum_{i}^{N_{\mathrm{ev}}+N_d} |v_i(x)\rangle \langle w_i(y)|$$

[Foley et. al. 2005]

We can then write a pion two-point function

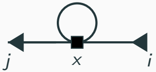
$$
\begin{aligned}
C(t) &= \left\langle O_\pi(x,t)\, O_\pi^\dagger(y,0) \right\rangle = \sum_{\vec{x},\vec{y}} \left\langle \mathrm{Tr}\left[S^{\alpha\beta}(x,y)\,\gamma_5\, S^{\beta\alpha}(y,x)\,\gamma_5 \right] \right\rangle \\
&= \sum_{\vec{x},\vec{y}} \left\langle \mathrm{Tr}\left[\sum_i |v_i(x)\rangle \langle w_i(y)|\,\gamma_5 \sum_j |v_j(y)\rangle \langle w_j(x)|\,\gamma_5 \right] \right\rangle \\
&= \sum_{ij} \left\langle \sum_{\vec{x}} \langle w_j(x)|\gamma_5|v_i(x)\rangle \sum_{\vec{y}} \langle w_i(y)|\gamma_5|v_j(y)\rangle \right\rangle \\
&= \left\langle \sum_{ij} \Pi_{ij}^{\gamma_5(\vec{0},0)}(x_0)\, \Pi_{ji}^{\gamma_5(\vec{0},0)}(y_0) \right\rangle
\end{aligned}
$$

We avoid a V^2 volume sum through constructing meson fields which we contract over mode indices



Extended meson fields

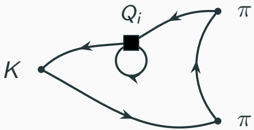
We can express the self-loop with two external legs as a meson-field-like object – the extended meson field. Generally



$$\sim \; \Pi^{4q}_{ij}(t) = \sum_{\vec{x}} v^{a,\alpha}_j(x) L^{a,\alpha;b,\beta}(x) v^{b,\beta}_i(x)$$

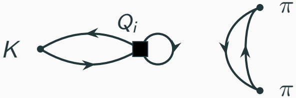
where L is a propagator field like object representing the loop propagator and both gamma structures. These extended meson fields contribute to two topologies in $K \to \pi\pi$ three point functions:

Type 3



$$\sim \left\langle \sum_{ijkl} \Pi^K_{ij}(t_K)\, \Pi^{4q}_{jk}(t_Q)\, \Pi^{\pi}_{kl}(t_1)\, \Pi^{\pi}_{li}(t_1) \right\rangle$$

Type 4



$$\sim \left\langle \sum_{ijk} \Pi^K_{ij}(t_K) \Pi^{4q}_{jk}(t_Q) \times \sum_{mn} \Pi^{\pi}_{nm}(t_1) \Pi^{\pi}_{mn}(t_1) \right\rangle$$

Optimization tools

Spin–color contractions on field objects $\Rightarrow$ threaded loops over site:

$$L^{\alpha a;\beta b}(x)\, v_i^{\beta b}(x) = z_i^{\alpha a}(x)$$

`accelerator_for` + coalesced reads/writes on GPU utilizes SIMD vector width as a second threading dimension for scalar instructions: Larger thread count and wider memory transactions on GPUs.

Final reduction: full spin–colour and spatial reduction, A2A outer product

$$\Pi_{ij}^{4q}(t) = \sum_{\vec{x}} \left\langle v_i^s(x) \middle| z_j^l(x) \right\rangle = \sum_{\vec{x},\alpha,a} v_i^{\alpha,a}(x)\, z_j^{\alpha,a}(x)$$

replace a threaded, serial, block-strided sum with highly optimized BLAS batched GEMM routines

$$C_{MN} = A_{MK}\, B_{KN}, \qquad M = N_i,\ N = N_j,\ K = N_{\mathrm{sc}} \times N_{xyz}$$

Very well saturated GPU – multiple batches of full tiled operations.

[Qian et. al. 2013]

[Boyle et. al. 2015]

Meson field optimization

With EMF offloaded, the standard meson field code is the next dominant piece – but the full momentum set is a constraint on the GEMM.

$$\Pi_{ij}^{\gamma_5}(\vec{p}, t) = \sum_{\vec{x}} e^{-i\vec{p}\cdot\vec{x}} \langle v_j(x)|\gamma_5|w_i(x)\rangle$$

- Cost grows as $\mathcal{O}(N_{\mathrm{mom}} \times N_\gamma)$; both are far too small to serve as GEMM outer indices.
- Settled design: explicit loop over $(\gamma, \vec{p})$, gamma and phase applied to the right vectors in their own kernels $\Rightarrow$ GEMM structure identical to EMF.
- HDF5 output per $(\gamma, \vec{p})$ pair $\Rightarrow$ parallel writes, so IO no longer dominates.

48I Timing comparison

Meson field	no. fields	GPU walltime (s)	CPU walltime (s)	improvement factor
$\bar{s}\gamma_5 d$	1	37.56	492.0	13.1
$\bar{s}\sigma^{\mu\nu}G_{\mu\nu}d$	1	117.9	960.9	8.15
K (3pt)	7	165.3	779.4	4.72
K (2pt)	14	334.9	1852.4	5.53
σ	7	447.5	1558	3.49
π	27	723.3	6009	8.31
EMF	2	2108.4	17452	8.28
Total walltime	–	3897.3	27651	7.09

64I timings and projections

Module	walltime (s)	node count	node hours
Eigenvector read	504.6	64	8.971
Light low modes	1471	64	26.15
Light high modes (deflated, 1 hit)	15814	64	281.1
Strange high modes (1 hit)	17476	32	155.3
CG job total	35265	–	471.5
A2A vector read	523.8	256	37.25
$\bar{s}\gamma_5 d$ field	37.63	256	2.676
$\bar{s}\sigma^{\mu\nu}G_{\mu\nu}d$ field	243.1	256	17.29
σ field (3 smearing widths)	1347	256	95.81
K fields (3 smearing widths)	1754	256	124.7
π field	1077	256	76.59
EMF (2 loop flavors)	2841	256	202.1
Smearing/Sparsening	630.1	256	44.81
MF job total	8454	–	601.2
Total	43719	–	1072.7

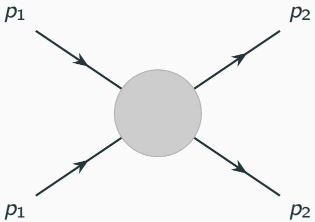
Part 2: Non-perturbative renormalization

Four-quark operator renormalization

We renormalize the four-quark operators multiplicatively

$$\hat{Q}_i^{\mathrm{RI}}(\mu) \;=\; \sum_{j\in C} Z_{ij}^{\mathrm{RI}\leftarrow\mathrm{lat}}(\mu)\,\hat{Q}_j^{\mathrm{lat}}$$

$$\frac{Z_{ij}(\mu)}{Z_q^2(\mu)}\,P_{4,m}^{\alpha\beta\gamma\delta}\left\langle E_m \sum_x e^{2ipx} Q_i^{\mathrm{lat}}(x)\right\rangle_{\mathrm{amp}}^{\alpha\beta\gamma\delta} = F_{im}$$



$$p_1^2 = p_2^2 = (p_1 - p_2)^2 = \mu^2$$

RI-SMOM: amputated, Landau-gauge-fixed Green's functions projected onto a given spin–colour structure are set equal to their tree-level value

External states E_m carry the kinematics that set the scale; twisted BCs in the valence sector relax the Fourier-mode constraint.

Window constraint and step scaling

$$A_I \sim \sum_{ij} [z_i(\mu) + \tau\, y_i(\mu)]\, Z_{ij}(\mu)\, \left\langle (\pi\pi)_I \middle| Q_j^{\mathrm{lat}} \middle| K \right\rangle$$

- All scales and schemes must match $\Rightarrow$ RI is an intermediate scheme; truncating the perturbative series at low scales dominates the error $\Rightarrow$ evolve μ non-perturbatively.
- Rome–Southampton window $\Lambda^2_{\mathrm{QCD}} \ll \mu^2 \ll (\pi/a)^2$: μ must be small relative to discretization effects. This upper bound is what motivates step scaling.
- Renormalize at a low μ inside the window, then move to smaller physical volume / finer lattice spacing to allow a higher renormalization scale μ' within the window:

$$\Sigma^s(\mu', \mu, a) = \frac{Z_{ij}(\mu')}{Z_{ij}(\mu)}, \quad \sigma^s(\mu', \mu) = \lim_{a\to 0} \Sigma^s(\mu', \mu, a), \quad \left\langle O^{\mathrm{RI}}(\mu') \right\rangle = \sum_{j\in C} \sigma^s_{ij}(\mu', \mu) \left\langle O_j^{\mathrm{RI}}(\mu) \right\rangle$$

Step scaling in previous PBC $K \to \pi\pi$ calculation

2023 PBC $K \to \pi\pi$ calculation [Blum et. al. 2023]

Measurement ensemble

- 24ID ensemble: $24^3 \times 64$, $a^{-1} = 1.023$ GeV, physical pion mass
- renormalized at two scales: $\mu_1 = 1.3$ GeV, $\mu_2 = 1.48$ GeV

Step scaling ensemble

- 32IDFine ensemble: $32^3 \times 64$, $a^{-1} = 3.148$ GeV, $m_\pi = 371$ MeV
- renormalized at low scales $\mu_1 = 1.236$ GeV, $\mu_2 = 1.514$ GeV, and PT matching scale $\mu_h = 4.006$ GeV

Systematics

- finite quark mass
- scale difference between step scaling and measurement ensembles
- step scaling matrix at a single lattice spacing
- PT matching scale – $\mu_h = 4.006$ GeV

Proposed step scaling ensembles

Step scaling function at multiple lattice spacings $\Rightarrow$ continuum limit. This requires matching integer Fourier modes between the physical-point ensembles and between the two fine ensembles.

$\Rightarrow$ **F** ensembles with half 48I lattice spacing, $a^{-1} = 3.53$ GeV, similar dimensions 48^4

$\Rightarrow$ **VF** ensembles with half 64I lattice spacing, $a^{-1} = 4.71$ GeV, similar dimensions $64^3 \times 96$

Two practical lever arms:

- second volume for both F, VF lattice spacing ($32^3 \times 48$ for F, 48^4 for VF)
- second β value for each volume / lattice spacing (3 on the 32F set)

Adds many more possible momenta modes for interpolation, to match scales as exactly as possible. Target ~ 8 GeV PT matching scale.

Extending the set of ensembles

F ensembles (target $a^{-1} = 3.53$)

- base set at $SU(3)$ symmetric point at physical strange, $m_{u/d} = m_s = 0.016715$ (32F3f, 48F3f)
- two mass variations on small volume:

  32F3f-ll: $m_{u/d} = 0.0088$

  32F3f-hs: $m_s = 0.0352$
- one variant with fixed global topology: 32F3f-Q15
- dynamical charm on both volumes:

  32F4f: $m_c = 0.187$

  48F4f: $m_c = 0.187$

VF ensembles (target $a^{-1} = 4.71$)

- base set at $SU(3)$ symmetric point, $m_{u/d} = m_s = 0.0125$ (48VF3f, 64VF3f)
- dynamical charm on large volume:

  64VF4f: $m_c = 0.142$ (physical)
- 3 charm mass variations on small volume:

  48VF4fc3: $m_c = 0.106$

  48VF4fc1: $m_c = 0.142$ (physical)

  48VF4fc2: $m_c = 0.178$

[Lehner (previous talk)]

Conclusion

Meson field optimization

- GPU meson field code runs efficiently on leadership-class machines
- code available through Grid, applicable to any meson field
- production running on 64I beginning early August; 100+ configs by end of 2026
- multiple hits, valence charm, BSM operators more accessible with optimized contraction software

Step scaling ensembles

- much higher PT matching scale possible, target ~ 8 GeV
- continuum limit of step scaling function
- systematic dynamical quark mass dependence studies possible
- 3f–4f matching non-perturbatively with ensembles of similar lattice spacing
- step scaling in full 4-flavour calculation
- beginning NPR measurements in early August

Backup

EMF code structure

Each EMF contraction type follows the same general procedure, each step of which is associated with a new GPU kernel. Take type 1 as an example (colour mixed, no separate loop trace):

$$\Pi_{ij}^{4q,\,\mathrm{type}\,1}(t) = \sum_{\vec{x}} \left\langle v_j^{a,\alpha}(x)\right| \left(\Gamma_\mu^1\right)^{\alpha\beta} \sum_k \left|v_k^{a,\beta}(x)\right\rangle \left\langle w_k^{b,\gamma}(x)\right| \left(\Gamma_\mu^2\right)^{\gamma\delta} \left|v_i^{b,\delta}(x)\right\rangle$$

Step 1: loop propagator calculation

$$S_{ab}^{\beta\gamma}(x,x) \;=\; \sum_i \left|v_i^{a,\beta}(x)\right\rangle \left\langle w_i^{b,\gamma}(x)\right|$$

Mode inner product, spin–color outer product for all sites. Same kernel for all four types.

$$\Pi_{ij}^{4q,\,\mathrm{type}\,1}(t) = \sum_{\vec{x}} \left\langle v_j^{\delta}(x)\right| \Gamma_\mu^{1,\,\alpha\beta}\, S_{ab}^{\beta\gamma}(x,x)\, \Gamma_\mu^{2,\,\gamma\delta}\, |v_i^{\alpha}(x)\rangle$$

EMF code structure

Step 2: "loop contraction" – evaluates the quantity between the external A2A vectors

$$L_{ab}^{\alpha\delta}(x) \;=\; \Gamma_{\mu}^{\alpha\beta}\, S_{ab}^{\beta\gamma}(x,x)\, \Gamma_{\mu}^{\gamma\delta} \qquad \longrightarrow \qquad \text{full propagator field for type 1}$$

$$\Longrightarrow \quad \Pi_{ij}^{4q,\,\mathrm{type}\,1}(t) = \sum_{\vec{x}} \left\langle v_j^{\alpha,a}(x) \middle| L_{ab}^{\alpha\delta}(x) \middle| v_i^{\delta,b}(x) \right\rangle$$

Step 3: LoopRight – contracts the result of step 2 with the right external A2A vector

$$L_{ab}^{\alpha\delta}(x)\, v_i^{\delta,b}(x) \;=\; z_i^{\alpha,a}(x)$$

Step 4: final reduction

$$\Pi_{ij}^{4q,\,\mathrm{type}\,1}(t) = \sum_{\vec{x}} \langle v_j(x) | z_i(x) \rangle$$

spin–colour inner product
A2A index outer product

Full set of new ensembles: F ensembles

Ensemble	Dimensions	L[fm]	β	a^{-1}	$a_{t_0}^{-1}$	$a_{w_0}^{-1}$	$m_{u,d}$	m_s	m_c
32F3f-ll	$32^3 \times 48$	1.769	2.44	3.53	3.57	3.43	0.0088	0.016715	∞
32F3f-hs	$32^3 \times 48$	1.769	2.44	3.53	3.57	3.43	0.016715	0.0352	∞
32F3f-Q15	$32^3 \times 48$	1.769	2.44	3.53	3.57	3.43	0.016715	0.016715	∞
32F3f-0	$32^3 \times 48$	1.769	2.44	3.53	3.57	3.43	0.016715	0.016715	∞
32F3f-1	$32^3 \times 48$	1.769	2.41	3.53	3.57	3.43	0.0176	0.0176	∞
32F3f-2	$32^3 \times 48$	1.769	2.47	3.53	3.57	3.43	0.0176	0.0176	∞
32F4f-1	$32^3 \times 48$	1.769	2.39	3.53	3.57	3.43	0.0176	0.0176	0.187
48F3f-1	48^4	2.653	2.41	3.53	3.57	3.43	0.0176	0.0176	∞
48F3f-2	48^4	2.653	2.47	3.53	3.57	3.43	0.0176	0.0176	∞
48F4f	48^4	2.653	2.39	3.53	3.57	3.43	0.0176	0.0176	0.187

Full set of new ensembles: VF ensembles

Ensemble	Dimensions	L[fm]	β	a^{-1}	$m_{u,d}$	m_s	m_c
48VF3f-1	48^4	2.011	2.56	4.71	0.0125	0.0125	∞
48VF3f-2	48^4	2.011	2.62	4.71	0.0125	0.0125	∞
48VF4fc1	48^4	2.011	2.54	4.71	0.0125	0.0125	0.142
48VF4fc2	48^4	2.011	2.54	4.71	0.0125	0.0125	0.178
48VF4fc3	48^4	2.011	2.54	4.71	0.0125	0.0125	0.213
64VF3f-1	$64^3 \times 96$	2.681	2.56	4.71	0.0125	0.0125	∞
64VF3f-2	$64^3 \times 96$	2.681	2.62	4.71	0.0125	0.0125	∞
64VF4f	$64^3 \times 96$	2.681	2.54	4.71	0.0125	0.0125	0.142