Implementation of the Neuberger operator on GPUs

Lattice 2010 — Sardinia

Friday, 18th of June, 2010

Bjoern Walk
bwalk@kph.uni-mainz.de
Institut für Kernphysik, Mainz

Outline

- Physical background
- Neuberger operator
- Low-mode projection
- Performance results
- Zero-mode expansion
- Conclusions/Outlook

Motivation

Problem statement

Calculate $K \to \pi\pi$ to investigate origins of the $\Delta I = 1/2$-rule

Do this:

- keeping the charm active, *[Giusti et. al., 2004]*
- outside the GIM-limit,
- on large volumes ($L \leq 2\,\mathrm{fm}$)

A lot **computational power** and **good algorithms** are required.

Transition amplitudes of $K \to \pi\pi$

- Parametrization of the transition amplitudes

$$T\left(K \to \pi\pi|_{I=\alpha}\right) = \mathrm{i}A_\alpha \mathrm{e}^{\mathrm{i}\delta_\alpha}, \qquad \alpha = 0, 2$$

- **Experimental data:**

$$|A_0|/|A_2| \sim 22.1 \qquad \Delta I = 1/2\text{-rule}$$

Mechanism of the enhancement not yet understood **quantitatively**!

Chiral symmetry

Chiral symmetry is an important aspect

- because of **mixing patterns** in the operator basis
- and to make contact to ChPT and determine **LEC**s

Ginsparg-Wilson relation

$$\gamma_5 D + D\gamma_5 = \bar{a} D \gamma_5 D$$

Solutions: **Neuberger overlap**, domain wall, ...

Neuberger overlap fermions

- Definition in terms of the Wilson-Dirac operator D_w

$$D = \frac{1 + \gamma_5 \,\mathrm{sign}(Q)}{\bar{a}}, \qquad Q = \gamma_5 (a D_\mathrm{w} - 1 - s)$$

- **Sign function** of an operator is defined by series expansion

$$\mathrm{sign}(Q) \simeq X P_n(X^2), \qquad X \equiv Q / \|Q\|$$

- Depending on n this is an expensive operation!

Minmax approximation to sign(Q) *[Giusti, Hoelbling, Lüscher, Wittig]*

$$\text{sign}(Q) \simeq X P_n(X^2), \qquad P_n(y) = \sum_k^n c_k T_k(y)$$

- $T_k(y)$ to be **Chebyshev polynomials** for numerical reasons
- Find c_k via minmax approximation and minimization of δ

$$\delta = \max_{\varepsilon \le y \le 1} |h(y)|, \qquad h(y) \equiv 1 - \sqrt{y}\, P_n(y)$$

Low-mode projection

[Giusti, Hoelbling, Lüscher, Wittig]

In the the ε-**regime** low-lying eigenmodes cannot be ignored

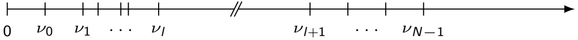


Figure: Positive spectrum of $Q^\dagger Q$

- Dirac-Operator Q gets **ill-conditioned**
- Separate the few lowest modes and treat them exactly
- Do this in such a way that error remains controllable

Low-mode projection

[Giusti, Hoelbling, Lüscher, Wittig]

- Determine low-lying eigenmodes via **Ritz functional**
- Let V be the subspace spanned by those low-modes
- Introduce projectors

$$\mathbb{P}_{\pm} = \sum_{\nu_k \lesseqgtr 0} u_k \otimes (u_k)^{\dagger}, \qquad u_k \in V$$

- With that

$$\mathrm{sign}(Q) \simeq \mathbb{P}_{+} - \mathbb{P}_{-} + (1 - \mathbb{P}_{+} - \mathbb{P}_{-})\, X P_n(X^2)$$

Low-mode projection algorithm

1. Find lowest eigenvalue ν_0 of Q via Ritz with CG
2. Project to the orthogonal subspace spanned by ν_0

$$\mathbb{P}_0 z = z - u_0 \langle u_0 | z \rangle$$

3. Introduce the projected operator $Q_0 = \mathbb{P}_0 Q \mathbb{P}_0$
4. Repeat with $Q \to Q_0$, k^{th} projector is given

$$\mathbb{P}_k z = z - \sum_j^k u_j \langle u_j | z \rangle$$

Utilization of the GPU

- Procedure is **very intensive** in terms of computing time
- Lots of applications of Q are neccessary

- Many vector operations and scalar products involved

Utilization of the GPU

- Procedure is **very intensive** in terms of computing time
- Lots of applications of Q are neccessary

- Many vector operations and scalar products involved

GPU is generally very good at those

Utilization of the GPU

- Procedure is **very intensive** in terms of computing time
- Lots of applications of Q are neccessary

has been done successfully on GPU *[Barros, Babich, Brower, Clark, Rebbi]*

- Many vector operations and scalar products involved

GPU is generally very good at those

Optimisations

For the Wilson-Dirac operator

- Ensure coalescing by **dedicated data layout**
- Minimize memory usage with **SU(3) reconstruction**

  [Bunk, Sommer, 1985]
- Use an **index theme** which supports texture cache

For the Neuberger operator

- Make sure the **whole algorithm** stays on the GPU
- Optimally **overlap** calculation with communication

Hardware comparison

	GTX285	**GTX480**	**C20*0**
Number of Cores	240	480	448
Amount of memory	2	1.5	3–6
Shader clock rate	1476	1401	1150
Memory clock rate	1242	1848	1500
Memory bandwidth	159.0	177.4	144.0

Benchmark results for BLAS functions

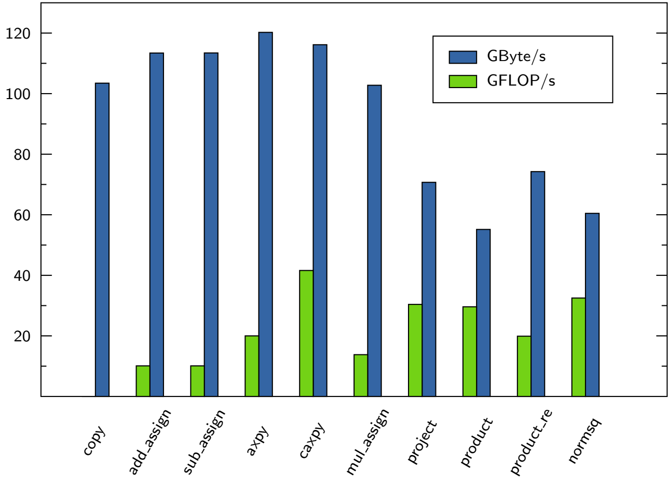


Figure: GeForce GTX285, Lattice size is 16^4, single-precision

Benchmark results for Wilson-Dirac operator

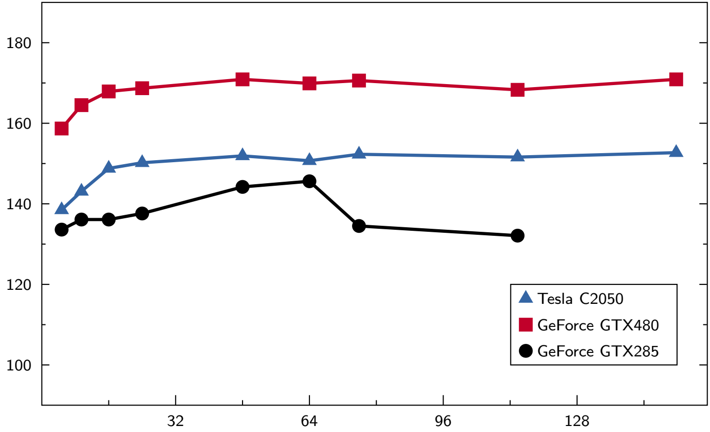
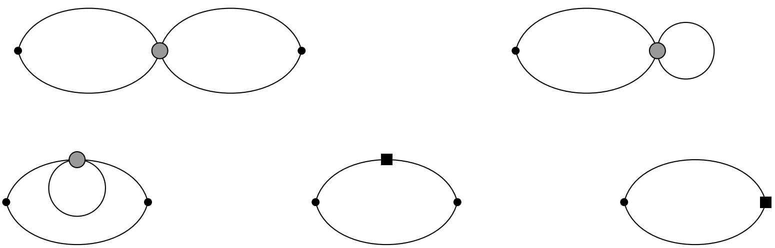


Figure: Lattice spatial volume is fixed at 24^3, single-precision

Diagrams for $K \to \pi\pi$



- ● $Q_1^{\pm} = ([O_1]_{suud} \pm [O_1]_{sudu}) - (u \to c)$
- ■ $Q_2^{\pm} = (m_u^2 - m_c^2) m_d (\bar{s} P_+ d) + m_s (\bar{s} P_d)$
- • $P^a(x) = \mathrm{i}(\bar{\psi}\gamma_5 T^a \psi)(x)$

Correlation functions

$K \longrightarrow \pi$

$$C^{\pm}_{1,2;\nu}(x_0, y_0) = \sum_{\mathbf{xy}} \langle [P(x)]_{du} Q_{1,2}(0) [P(y)]_{us} \rangle_\nu$$

$K \longrightarrow$ vacuum

$$K^{\pm}_{1,2;\nu}(x_0) = \sum_{\mathbf{x}} \langle Q_{1,2}(0) [P(y)]_{ds} \rangle_\nu$$

Inversion of the Neuberger operator

- Pseudo-scalar density allows for **zero-mode saturation**
- Spectral decomposition

$$S(x,y) = \sum_{i}^{|\nu|} \frac{v_i(x) \otimes v_i(y)^\dagger}{mV} + \text{ higher modes}$$

With that

- introduce a **zero-mode source** for inversion,
- correlators can be obtained by **scalar products** of zero-modes

Inversion of the Neuberger operator

- Often the projected propagator $P_{\mp}S(x,y)P_{\pm}$ occurs
- Zero-modes have **definite chirality**

$$\gamma_5 v_i(x) = \chi v_i(x), \qquad \chi = \nu/|\nu|$$

- Hence, they **do not contribute** to projected propagator
- Still one explicit inversion for the propagator neccessary

Conclusion

- Neuberger overlap fermions are **very compute intensive**
- In the ε-regime **low-modes** cannot be ignored
- **Zero-mode expansion** as alternative approach for the propagator (work in progess)
- GPUs are a **good utility** for our investigations

Thank you for your attention!