Tensor Trains: defeating the curse of dimensionality

F. Alexander Wolf

Arnold Sommerfeld Center for Theoretical Physics, LMU Munich

Institute for Computational Biology
Helmholtz Center Munich
16 Feb 2015

Motivation

Kazeev, Khammash, Nip & Schwab, PLoS Comput. Biol. **10** e1003359 (2014)

OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Direct Solution of the Chemical Master Equation Using Quantized Tensor Trains

Vladimir Kazeev[1], Mustafa Khammash[2]*, Michael Nip[3], Christoph Schwab[4]

1 Seminar für Angewandte Mathematik, ETH Zürich, Zürich, Switzerland, **2** Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland, **3** Department of Mechanical Engineering, UC Santa Barbara, Santa Barbara, California, United States of America, **4** Seminar für Angewandte Mathematik, ETH Zürich, Zürich, Switzerland

simulation package. This allows us, on the one hand, to validate the accuracy of the QTT-based solutions obtained here and, on the other hand, to provide evidence of the dramatic increase in

17 March 2014 | Volume 10 | Issue 3 | e1003359

efficiency afforded by the new deterministic approach: Monte Carlo simulations on 1500 cores of a high-performance cluster were matched in accuracy and outperformed in the wall-clock time by a MATLAB implementation running on a notebook.

Outline

- Tensor Trains in statistical physics
- Solving the Chemical Master Equation using Tensor Trains

Statistics / Statistical physics

Statistics

data

- ▷ properties of the probability distribution
  - ▷ underlying laws / mechanisms / causes

Statistics / Statistical physics

Statistics

data

▷ properties of the probability distribution

▷ underlying laws / mechanisms / causes

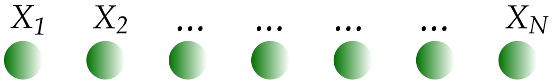
Statistical physics

natural laws / microscopic interactions

▷ beast of a probability distribution

▷ emergent macroscopic behavior / emergent correlations

Generic example (i): noninteracting 1d Ising model

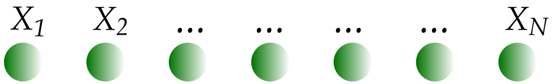


System described by vector of random variables $\boldsymbol{X} \in \{0,1\}^N$ with joint probability mass function

$$p(\boldsymbol{x}) = \tfrac{1}{Z} e^{-H(\boldsymbol{x})/T}, \quad H(\boldsymbol{x}) = \sum_{n=1}^{N} x_n$$

normalized to $Z = \sum_{\boldsymbol{x}} e^{-H(\boldsymbol{x})/T}$.

Generic example (i): noninteracting 1d Ising model



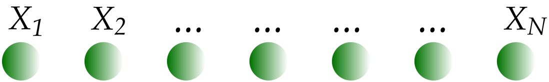
System described by vector of random variables $\boldsymbol{X} \in \{0,1\}^N$ with joint probability mass function

$$\boldsymbol{p}_{\boldsymbol{x}} = \tfrac{1}{Z} e^{-H(\boldsymbol{x})/T}, \quad H(\boldsymbol{x}) = \sum_{n=1}^{N} x_n$$

normalized to $Z = \sum_{\boldsymbol{x}} e^{-H(\boldsymbol{x})/T}$.

▷ $\boldsymbol{p}$ has 2^N components $\boldsymbol{x} \in \{(0,0,...,0),(0,0,...,1),\dots\}$.

Generic example (i): noninteracting 1d Ising model



System described by vector of random variables $\boldsymbol{X} \in \{0,1\}^N$ with joint probability mass function

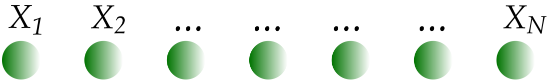
$$\boldsymbol{p}_{\boldsymbol{x}} = \tfrac{1}{Z} e^{-H(\boldsymbol{x})/T}, \quad H(\boldsymbol{x}) = \sum_{n=1}^{N} x_n$$

normalized to $Z = \sum_{\boldsymbol{x}} e^{-H(\boldsymbol{x})/T}$.

▷ $\boldsymbol{p}$ has 2^N components $\boldsymbol{x} \in \{(0,0,...,0),(0,0,...,1),\dots\}$.

▷ Remark $2^{100} \simeq 10^{30} \simeq 10^{15}$ TB.

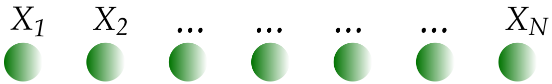
Generic example (i): noninteracting 1d Ising model



Compute correlations via $\mathrm{cov}(X_n, X_m) = \langle X_n X_m \rangle - \langle X_n \rangle \langle X_n \rangle$.

$$\langle X_n X_m \rangle = \sum_{\boldsymbol{x}} x_n x_m \boldsymbol{p}_{\boldsymbol{x}}$$

Generic example (i): noninteracting 1d Ising model



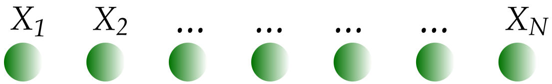
Compute correlations via cov$(X_n, X_m) = \langle X_n X_m \rangle - \langle X_n \rangle \langle X_n \rangle$.

$$\langle X_n X_m \rangle = \sum_{\boldsymbol{x}} x_n x_m \boldsymbol{p}_{\boldsymbol{x}}$$

▷ Naive brute force: 2^N operations necessary.

▷ Monte Carlo: sampling in space of 2^N states.

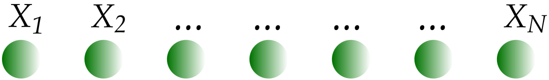
Generic example (i): noninteracting 1d Ising model



But: non-interacting degrees of freedom X_n imply full *separability*

$$
\begin{aligned}
\boldsymbol{p_x} = \boldsymbol{p}_{x_1,x_2,\ldots,x_N} &= \tfrac{1}{Z} e^{-\sum_{n=1}^{N} x_n/T} \\
&= \tfrac{1}{Z} A_{x_1} A_{x_2} \ldots A_{x_N}, \quad A_{x_n} = e^{-x_n/T}
\end{aligned}
$$

Generic example (i): noninteracting 1d Ising model



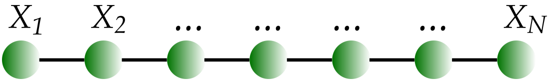
But: non-interacting degrees of freedom X_n imply full *separability*

$$
\begin{aligned}
\boldsymbol{p_x} = \boldsymbol{p}_{x_1,x_2,\ldots,x_N} &= \tfrac{1}{Z} e^{-\sum_{n=1}^{N} x_n/T} \\
&= \tfrac{1}{Z} A_{x_1} A_{x_2} \ldots A_{x_N}, \quad A_{x_n} = e^{-x_n/T}
\end{aligned}
$$

Compute correlations in $2N$ operations . . .

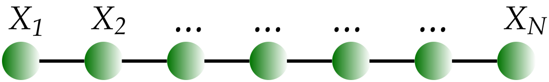
$$
\begin{aligned}
\langle X_n X_m \rangle &= \frac{1}{Z} \Big(\sum_{x_n} x_n A_{x_n} \Big) \Big(\sum_{x_m} x_m A_{x_m} \Big) \prod_{k \neq n,m}^{N} \Big(\sum_{x_k} A_{x_k} \Big) \\
&= \langle X_n \rangle \langle X_m \rangle \quad \ldots \quad \text{there are none.}
\end{aligned}
$$

Generic example (ii): interacting 1d Ising model



$$\widetilde{\boldsymbol{p}}_{\boldsymbol{x}} = \tfrac{1}{Z} e^{-H(\boldsymbol{x})/T}, \quad H(\boldsymbol{x}) = -\sum_{n=1}^{N-1} x_n x_{n+1}$$

Generic example (ii): interacting 1d Ising model



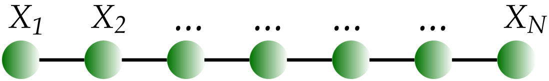
$$\widetilde{\boldsymbol{p}}_{\boldsymbol{x}} = \tfrac{1}{Z} e^{-H(\boldsymbol{x})/T}, \quad H(\boldsymbol{x}) = -\sum_{n=1}^{N-1} x_n x_{n+1}$$

▷ Is just a "discrete Gaussian" (continuous if $X_n \in \mathbb{R}$) with

$$\mathsf{cov}(\boldsymbol{x}, \boldsymbol{y})^{-1} = \begin{pmatrix} 0 & \frac{2}{T} & 0 & \dots & 0 \\ \frac{2}{T} & 0 & \frac{2}{T} & \dots & 0 \\ 0 & \frac{2}{T} & 0 & \frac{2}{T} & \dots \\ 0 & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

▷ Correlations by inverting or diagonalizing the covariance matrix.

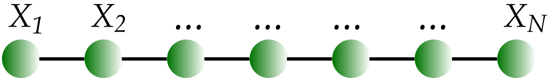
Generic example (ii): interacting 1d Ising model



But: two-body interactions imply "almost – separability"

$$Z \sum_{\boldsymbol{x}} \widetilde{\boldsymbol{p}}_{\boldsymbol{x}} = \sum_{\boldsymbol{x}} e^{x_1 x_2 / T} e^{x_2 x_3 / T} \ldots$$

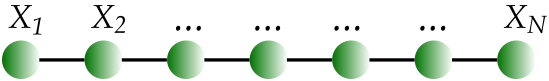
Generic example (ii): interacting 1d Ising model



But: two-body interactions imply "almost – separability"

$$Z \sum_{\boldsymbol{x}} \widetilde{\boldsymbol{p}}_{\boldsymbol{x}} = \sum_{\boldsymbol{x}} A_{x_1,x_2} A_{x_2,x_3} \ldots$$

Generic example (ii): interacting 1d Ising model

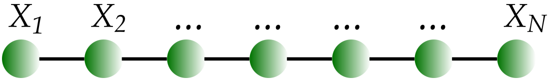


But: two-body interactions imply "almost – separability"

$$Z \sum_{\boldsymbol{x}} \widetilde{\boldsymbol{p}}_{\boldsymbol{x}} = \sum_{\boldsymbol{x}} A_{x_1,x_2} A_{x_2,x_3} \ldots$$

$$= \mathsf{tr}_{\mathsf{all}} AA \ldots, \qquad A_{x_n,x_{n+1}} = e^{x_n x_{n+1}/T}, \quad A \in \mathbb{R}^{2\times 2}$$

Generic example (ii): interacting 1d Ising model



But: two-body interactions imply “almost – separability”

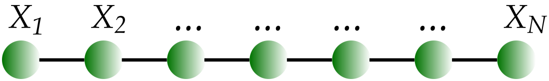
$$Z \sum_{\boldsymbol{x}} \widetilde{\boldsymbol{p}}_{\boldsymbol{x}} = \sum_{\boldsymbol{x}} A_{x_1,x_2} A_{x_2,x_3} \ldots$$

$$= \mathsf{tr}_{\mathsf{all}} AA \ldots, \qquad A_{x_n,x_{n+1}} = e^{x_n x_{n+1}/T}, \quad A \in \mathbb{R}^{2\times 2}$$

▷ Compare to non-interacting case

$$Z \sum_{\boldsymbol{x}} \boldsymbol{p}_{\boldsymbol{x}} = \sum_{\boldsymbol{x}} A_{x_1} A_{x_2} \ldots, \qquad A_{x_n} = e^{-x_n/T}$$

Generic example (ii): interacting 1d Ising model

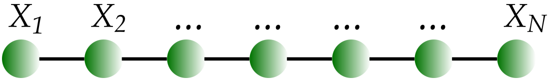


Compute correlations in $2^3 N$ operations (N matrix products)

$$\langle X_n X_m \rangle_{\widetilde{\boldsymbol{p}}} = \frac{1}{Z} \mathsf{tr}_{\mathsf{all}} \prod_{k=1}^{n-1} \left(A^{[k]} \right) M \prod_{k=n}^{m-1} \left(A^{[k]} \right) M \prod_{k=m}^{N-1} \left(A^{[k]} \right)$$

$$\text{where} \quad M = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Generic example (ii): interacting 1d Ising model



Compute correlations in $2^3 N$ operations (N matrix products)

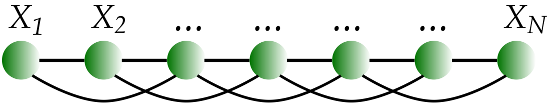
$$\langle X_n X_m \rangle_{\widetilde{\boldsymbol{p}}} = \frac{1}{Z} \mathsf{tr}_{\mathsf{all}} \prod_{k=1}^{n-1} \left(A^{[k]} \right) M \prod_{k=n}^{m-1} \left(A^{[k]} \right) M \prod_{k=m}^{N-1} \left(A^{[k]} \right)$$

where $\quad M = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

▷ Compare to non-interacting case ($2N$ operations)

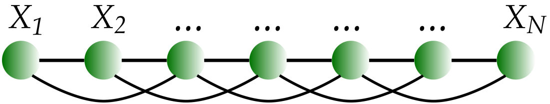
$$\langle X_n X_m \rangle_{\boldsymbol{p}} = \frac{1}{Z} \Big(\sum_{x_n} x_n A_{x_n} \Big) \Big(\sum_{x_m} x_m A_{x_m} \Big) \prod_{k \neq n,m} \Big(\sum_{x_k} A_{x_k} \Big)$$

Generic example (iii): three-body interacting Ising model



$$\hat{\boldsymbol{p}}_{\boldsymbol{x}} = \frac{1}{Z} e^{-H(\boldsymbol{x})/T}, \quad H(\boldsymbol{x}) = -\sum_{n=1}^{N-2} x_n x_{n+1} x_{n+2}$$

Generic example (iii): three-body interacting Ising model



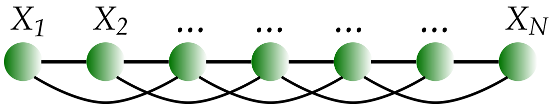
$$\hat{\boldsymbol{p}}_{\boldsymbol{x}} = \frac{1}{Z} e^{-H(\boldsymbol{x})/T}, \quad H(\boldsymbol{x}) = -\sum_{n=1}^{N-2} x_n x_{n+1} x_{n+2}$$

$$Z \sum_{\boldsymbol{x}} \hat{\boldsymbol{p}}_{\boldsymbol{x}} = \sum_{\boldsymbol{x}} \prod_{n=1}^{N-2} A_{x_n, x_{n+1}, x_{n+2}}$$

$$A_{x_n, x_{n+1}, x_{n+2}} = e^{x_n x_{n+1} x_{n+2}/T}$$

$$A \in \mathbb{R}^{2\times 2\times 2}$$

Generic example (iii): three-body interacting Ising model



$$\hat{\boldsymbol{p}}_{\boldsymbol{x}} = \tfrac{1}{Z} e^{-H(\boldsymbol{x})/T}, \quad H(\boldsymbol{x}) = -\sum_{n=1}^{N-2} x_n x_{n+1} x_{n+2}$$

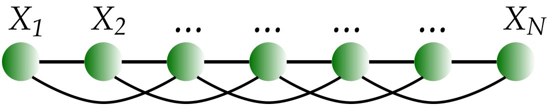
$$Z \sum_{\boldsymbol{x}} \hat{\boldsymbol{p}}_{\boldsymbol{x}} = \sum_{\boldsymbol{x}} \prod_{n=1}^{N-2} A_{x_n, x_{n+1}, x_{n+2}}$$

$$A_{x_n, x_{n+1}, x_{n+2}} = e^{x_n x_{n+1} x_{n+2}/T}$$
$$A \in \mathbb{R}^{2\times 2\times 2}$$

$$= \sum_{\boldsymbol{x}'} \prod_{n=1}^{N-2} B_{x'_n, x'_{n+1}} B^t_{x'_{n+1}, x'_{n+2}}$$

$$B_{x'_n, 2x_{n+1}+x_{n+2}} = A_{x_n, x_{n+1}, x_{n+2}}$$
$$B \in \mathbb{R}^{2\times 4}$$

Generic example (iii): three-body interacting Ising model



$$\hat{\boldsymbol{p}}_{\boldsymbol{x}} = \tfrac{1}{Z} e^{-H(\boldsymbol{x})/T}, \quad H(\boldsymbol{x}) = -\sum_{n=1}^{N-2} x_n x_{n+1} x_{n+2}$$

$$Z \sum_{\boldsymbol{x}} \hat{\boldsymbol{p}}_{\boldsymbol{x}} = \sum_{\boldsymbol{x}} \prod_{n=1}^{N-2} A_{x_n, x_{n+1}, x_{n+2}}$$

$$A_{x_n, x_{n+1}, x_{n+2}} = e^{x_n x_{n+1} x_{n+2}/T}$$

$$A \in \mathbb{R}^{2\times 2\times 2}$$

$$= \sum_{\boldsymbol{x}'} \prod_{n=1}^{N-2} B_{x'_n, x'_{n+1}} B^t_{x'_{n+1}, x'_{n+2}}$$

$$B_{x'_n, 2x_{n+1}+x_{n+2}} = A_{x_n, x_{n+1}, x_{n+2}}$$

$$B \in \mathbb{R}^{2\times 4}$$

Tensor Train format ▷ $\frac{1}{2}(2^3 + 4^3)N$ operations

Summary Part I

- Write function $v : \{0, 1, ..., d\}^N \to \mathbb{F}$, $d, N \in \mathbb{N}$ as vector $\boldsymbol{v}_{\boldsymbol{x}} = v(\boldsymbol{x})$, $\boldsymbol{v} \in \mathbb{F}^{d^N}$, that is indexed *and parametrized* by $\boldsymbol{x} \in \{0, 1, ..., d\}^N$.

Summary Part I

- Write function $v : \{0, 1, ..., d\}^N \to \mathbb{F}$, $d, N \in \mathbb{N}$ as vector $\boldsymbol{v_x} = v(\boldsymbol{x})$, $\boldsymbol{v} \in \mathbb{F}^{d^N}$, that is indexed *and parametrized* by $\boldsymbol{x} \in \{0, 1, ..., d\}^N$.

  If $\boldsymbol{v_x} = v(\boldsymbol{x})$ does *not* couple *all* index components x_n among each other, there is a low rank TT representation.

Summary Part I

- Write function $v : \{0, 1, ..., d\}^N \to \mathbb{F}$, $d, N \in \mathbb{N}$ as vector $\boldsymbol{v_x} = v(\boldsymbol{x})$, $\boldsymbol{v} \in \mathbb{F}^{d^N}$, that is indexed *and parametrized* by $\boldsymbol{x} \in \{0, 1, ..., d\}^N$.

  If $\boldsymbol{v_x} = v(\boldsymbol{x})$ does *not* couple *all* index components x_n among each other, there is a low rank TT representation.

  This reduces computational cost in manipulations of such a function (vector) from exponential to linear in system size.

Summary Part I

- Write function $v : \{0, 1, ..., d\}^N \to \mathbb{F}$, $d, N \in \mathbb{N}$ as vector $\boldsymbol{v}_{\boldsymbol{x}} = v(\boldsymbol{x})$, $\boldsymbol{v} \in \mathbb{F}^{d^N}$, that is indexed *and parametrized* by $\boldsymbol{x} \in \{0, 1, ..., d\}^N$.

  If $\boldsymbol{v}_{\boldsymbol{x}} = v(\boldsymbol{x})$ does *not* couple *all* index components x_n among each other, there is a low rank TT representation.

  This reduces computational cost in manipulations of such a function (vector) from exponential to linear in system size.

- Approach stems from quantum many-body physics.

  White, Phys. Rev. Lett. **69** 2863 (1992)

Summary Part I

- Write function $v : \{0, 1, ..., d\}^N \to \mathbb{F}$, $d, N \in \mathbb{N}$ as vector $\boldsymbol{v_x} = v(\boldsymbol{x})$, $\boldsymbol{v} \in \mathbb{F}^{d^N}$, that is indexed *and parametrized* by $\boldsymbol{x} \in \{0, 1, ..., d\}^N$.

  If $\boldsymbol{v_x} = v(\boldsymbol{x})$ does *not* couple *all* index components x_n among each other, there is a low rank TT representation.

  This reduces computational cost in manipulations of such a function (vector) from exponential to linear in system size.

- Approach stems from quantum many-body physics.

  White, Phys. Rev. Lett. **69** 2863 (1992)

  In the quantum TT community, the most used algorithm is an optimization that operates on an arbitrarily parameterized TT:

  Solve linear system $H\boldsymbol{v} = \lambda \boldsymbol{v}$ for the lowest eigenvalue

$$\min_{\boldsymbol{v}} \frac{(\boldsymbol{v}, H\boldsymbol{v})}{(\boldsymbol{v}, \boldsymbol{v})}, \qquad \boldsymbol{v} \in \mathbb{C}^{d^N}, \ H \in \mathbb{C}^{d^N \times d^N}$$

Outline

- ▷ Tensor Trains in statistical physics
- ▷ Solving the Chemical Master Equation using Tensor Trains

Solving the chemical master equation using Tensor Trains

Kazeev, Khammash, Nip & Schwab, PLoS Comput. Biol. **10** e1003359 (2014) Dolgov & Khoromskij, arXiv:1311.3143 (2013)

N reacting molecules in thermal equilibrium are described by a jump Markov process: the number of molecules of one species corresponds to one component of a random vector $\boldsymbol{X}(t) \in \{0, 1, \ldots, n_{\mathsf{max}}\}^N$.

Solving the chemical master equation using Tensor Trains

Kazeev, Khammash, Nip & Schwab, PLoS Comput. Biol. **10** e1003359 (2014)   Dolgov & Khoromskij, arXiv:1311.3143 (2013)

N reacting molecules in thermal equilibrium are described by a jump Markov process: the number of molecules of one species corresponds to one component of a random vector $\boldsymbol{X}(t) \in \{0, 1, \ldots, n_{\text{max}}\}^N$.

The corresponding probability density function $\boldsymbol{p}(t)$, where $\boldsymbol{p}_{\boldsymbol{x}}(t)$ is the probability for that a certain population number configuration $\boldsymbol{x} = (n_1, n_2, ..., n_N)$ occurs at time t, evolves according to a linear ODE:

$$\frac{d}{dt}\boldsymbol{p}(t) = H\boldsymbol{p}(t)$$

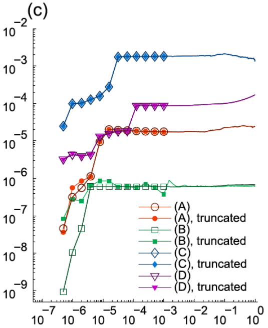
H describes chemical reactions parametrized by propensities $\omega(\boldsymbol{x})$ and coupling terms.

Example: enzymatic futile cycle

Kazeev, Khammash, Nip & Schwab, PLoS Comput. Biol. **10** e1003359 (2014)

$$X + E^{f}_{+} \underset{k_{+2}}{\overset{k_{+1}}{\rightleftharpoons}} E^{b}_{+}, \quad X^{*} + E^{f}_{-} \underset{k_{-2}}{\overset{k_{-1}}{\rightleftharpoons}} E^{b}_{-},$$

$$E^{b}_{+} \xrightarrow{k_{+3}} E^{f}_{+} + X^{*}, \quad E^{b}_{-} \xrightarrow{k_{-3}} E^{f}_{-} + X,$$



▷ State space truncated to $2^{22} \simeq 4 \cdot 10^{6}$.

▷ "… 10^{10} Monte Carlo simulations (every 10000 draws taking at least 110 seconds, amounting to the overall CPU time over 10^{8} seconds) …"

Tensor Trains in the literature

- ▷ Physics: Schollwöck, Rev. Mod. Phys. **77**, 259 (2005) / Schollwöck, Annals of Physics **326**, 96 (2011)
- ▷ Applied mathematics: no review yet, but vivid research activities.
- ▷ "Data science": few very recent treatments.
- ▷ Biology: One journal article.

Tensor Trains in the literature

- ▷ Physics: Schollwöck, Rev. Mod. Phys. **77**, 259 (2005) / Schollwöck, Annals of Physics **326**, 96 (2011)
- ▷ Applied mathematics: no review yet, but vivid research activities.
- ▷ "Data science": few very recent treatments.
- ▷ Biology: One journal article.

Thanks for your attention!

Dolgov, S. & B. Khoromskij, 2013, 1311.3143.

Kazeev, V., M. Khammash, M. Nip & C. Schwab, 2014, PLoS Comput. Biol. **10**, e1003359.

Schollwöck, U., 2005, Rev. Mod. Phys. **77**, 259.

Schollwöck, U., 2011, Annals of Physics **326**, 96.

White, S. R., 1992, Phys. Rev. Lett. **69**, 2863.