Near-End Listening Enhancement for Mobile Phones

Bastian Sauert

6th Workshop on Speech in Noise: Intelligibility and Quality

Marseille, France – January 10, 2014

Institute of Communication Systems and Data Processing
Prof. Dr.-Ing. Peter Vary

RWTH AACHEN UNIVERSITY

Introduction: Near-End Listening Enhancement

Near-end listener in background noise

- Higher listening effort
- Reduced speech intelligibility

Approach

- *Preprocess* clean far-end speech
- *Enhance* intelligibility in near-end noise

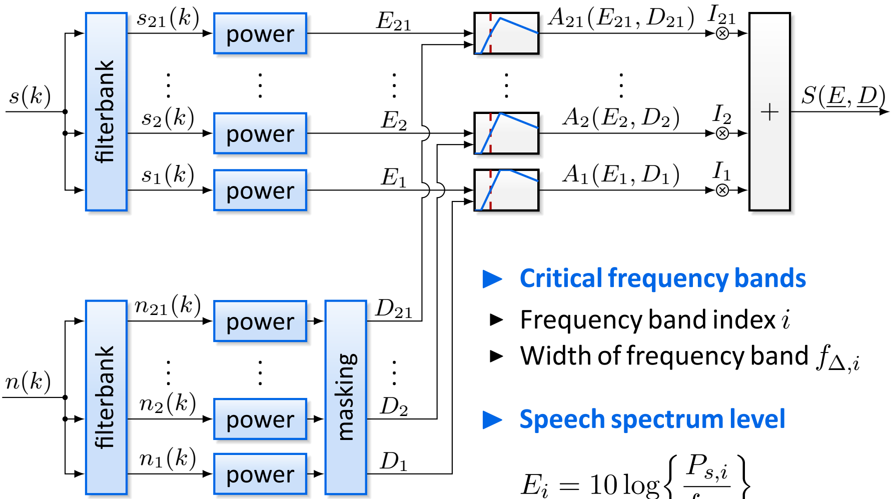
Near-End Listening Enhancement

- **Most proposed algorithms in literature**
  - Noise-independent processing
  - Heuristic approaches

- **Objective criterion**
  - *Speech Intelligibility Index* (SII) **[ANSI S3.5-1997]**

- **Consider requirements and restrictions of realistic scenarios**
  - Dynamic adaptation to ambient noise characteristics **[Sauert 2006a,b]**
  - Blind noise estimation from microphone signal **[Sauert 2011]**
  - Loudspeaker restrictions **[Sauert 2010a]**
  - Prevent hearing damage **[Sauert 2006a,b]**
  - Low delay **[Sauert 2008]**

Speech Intelligibility Index (SII) [ANSI S3.5-1997]

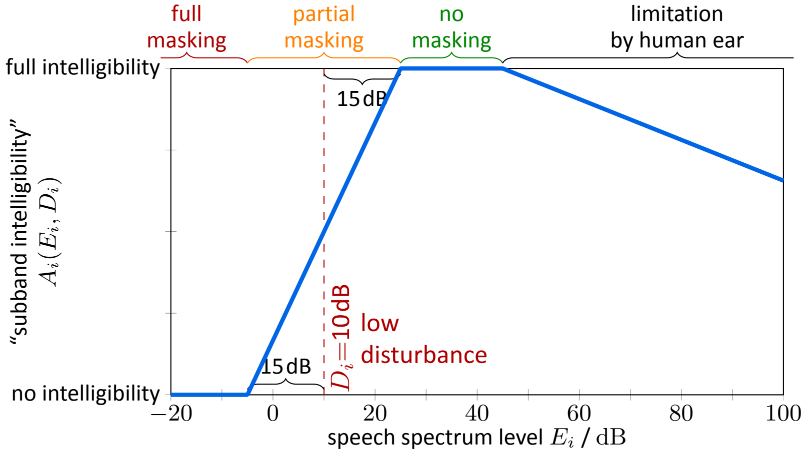


- **Critical frequency bands**
  - Frequency band index i
  - Width of frequency band $f_{\Delta,i}$
- **Speech spectrum level**

$$E_i = 10 \log \left\{ \frac{P_{s,i}}{f_{\Delta,i}} \right\}$$

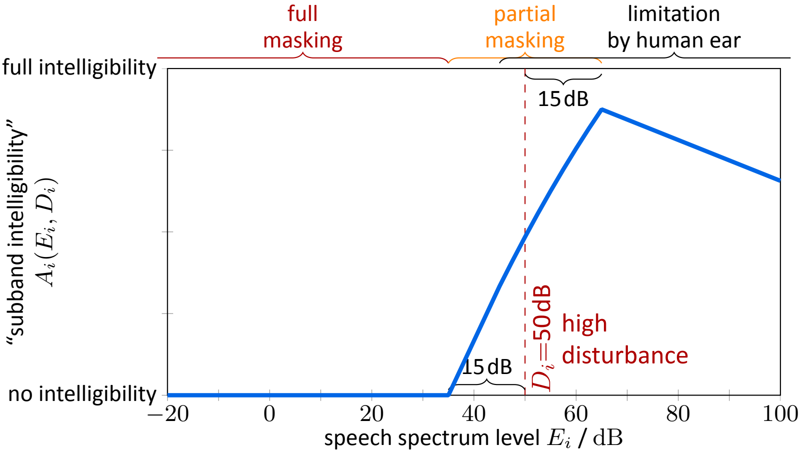
Band Audibility Function $A_i(E_i, D_i)$ in Subband i

- **Low disturbance case**



Band Audibility Function $A_i(E_i, D_i)$ in Subband i

- **High disturbance case**



Main Steps of SII-Based Optimization Algorithm [Sauert 2009]

1. **Find optimum output speech spectrum level $\underline{E}_{\text{opt}}(\kappa)$**
   which maximizes the speech intelligibility index $S(\underline{E}, \underline{D})$
   given the current disturbance spectrum level $\underline{D}(\kappa)$

$$\underline{E}_{\text{opt}}(\kappa) = \arg\max_{\underline{E}^{\text{out}}} \left[\sum_i I_i \cdot A_i(E_i^{\text{out}}, D_i(\kappa)) \right]$$

2. **Calculate spectral weights**
   to reach optimum output speech spectrum level
   with received far-end (input) speech signal

$$W_i(\kappa) = 10^{[E_{\text{opt},i}(\kappa) - E_i^{\text{in}}(\kappa)]/20}$$

3. **Apply weights to received far-end speech signal**

Optimization with Constraint of Total Output Power

- **Three cases of total audio power constraint**
  - *Without* limitation of total audio power
  - *No increase* of total audio power: $\mathfrak{P}^{\text{max}}(\kappa) = \mathfrak{P}_s^{\text{in}}(\kappa)$
  - *Increase* of total power *up to limit* of loudspeaker

SII-Based Optimization *without* Constraint of Total Power

- **Find optimum output speech spectrum level $\underline{E}_{\text{opt}}(\kappa)$**

$$\underline{E}_{\text{opt}}(\kappa) = \arg\max_{\underline{E}^{\text{out}}} \left[\sum_i I_i \cdot A_i(E_i^{\text{out}}, D_i(\kappa)) \right] \quad \text{given } D_i(\kappa)$$

- **Constraint:** Protection of listener's ear

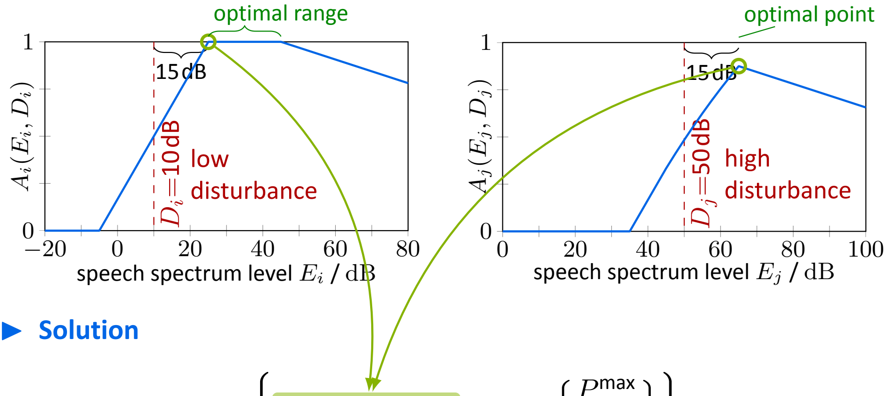
$$E_i^{\text{out}} \leq 10 \log\left\{ \frac{P_s^{\text{max}}}{f_{\Delta,i}} \right\} \quad \text{with} \quad 10 \log\left\{ \frac{P_s^{\text{max}}}{P_0} \right\} = 95\,\text{dB}_{\text{SPL}}$$

- **21-dimensional non-convex quadratic optimization problem with bound constraint**

$\Rightarrow$ **21 *one*-dimensional non-convex quadratic optimization problems with bound constraint**

Bounded SII-Based Optimization

[Sauert 2009]



▶ Solution

$$E_{\text{opt},i} = \min\left\{ D_i(\kappa) + 15\,\text{dB},\ 10\log\left\{\frac{P_s^{\max}}{f_{\Delta,i}}\right\}\right\}$$

- A more sophisticated solution tries to retain color of tone

SII-Based Optimization *with* Constraint of Total Power

- **Find optimum output speech spectrum level $\underline{E}_{\text{opt}}(\kappa)$**

$$\underline{E}_{\text{opt}}(\kappa) = \arg\max_{\underline{E}^{\text{out}}} \left[\sum_i I_i \cdot A_i(E_i^{\text{out}}, D_i(\kappa)) \right] \quad \text{given } D_i(\kappa)$$

- **Constraint 1:** Protection of listener's ear

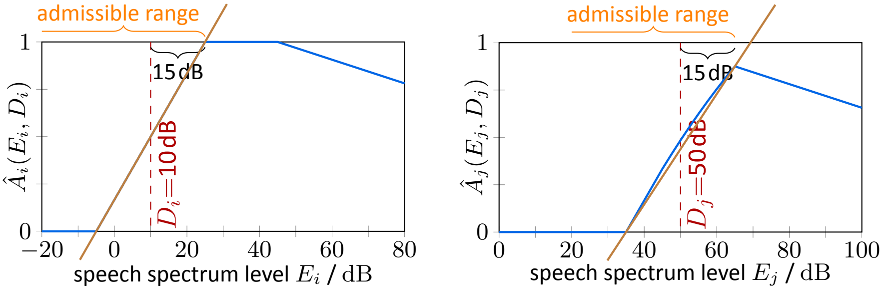
$$E_i^{\text{out}} \leq 10 \log \left\{ \frac{P_s^{\text{max}}}{f_{\Delta,i}} \right\} \quad \text{with} \quad 10 \log \left\{ \frac{P_s^{\text{max}}}{P_0} \right\} = 95\,\text{dB}_{\text{SPL}}$$

- **Constraint 2:** Total output power

$$\sum_i P_{s,i}^{\text{out}} = \sum_i f_{\Delta,i} \cdot 10^{E_i^{\text{out}}/10} \overset{!}{\leq} \mathfrak{P}^{\text{max}}(\kappa)$$

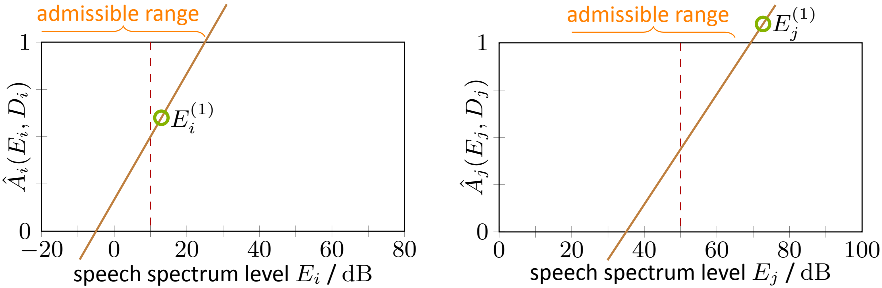
- **21-dimensional non-convex quadratic optimization problem with exponential constraint**

Recursive Closed-Form Optim. (OptSIIrecur) [Sauert 2010b]



- **Admissible range up to $D_i + 15\,\mathrm{dB}$**
- **Linear approximation $\hat{A}_i(E_i, D_i)$**
  - Reasonably good below $D_i + 15\,\mathrm{dB}$
  - No match above $D_i + 15\,\mathrm{dB}$

Recursive Closed-Form Optim. (OptSIIrecur) [Sauert 2010b]

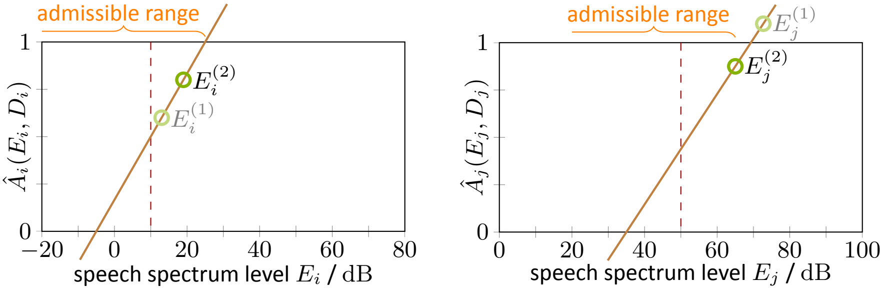


- **Approach:** Method of Lagrange multipliers
- **Step 1:** Closed-form solution $\forall\ i$

$$E_i^{(1)} = 10 \log \left\{ \frac{\Gamma_i}{\sum_\zeta \Gamma_\zeta} \cdot \frac{\mathfrak{P}^{\text{max}}(\kappa)}{f_{\Delta,i}} \right\} \text{ with } \Gamma_i = I_i \cdot \frac{\partial \hat{A}_i(E_i, D_i(\kappa))}{\partial E_i}$$

  - Solution might be out of admissible range: $E_i^{(1)} > D_i(\kappa) + 15\,\text{dB}$

Recursive Closed-Form Optim. (OptSIIrecur) [Sauert 2010b]



- **Step $\upsilon = 2, 3, \ldots$**
  - Limit $E_i^{(\upsilon-1)}$ to admissible range
  - Repeat closed-form solution for remaining frequency bands

- **Final solution**

$$\underline{E}_{\text{opt}} = \underline{E}^{(\Upsilon)}$$ with usually $\Upsilon \leq 2$ recursion steps

Analysis of Optimum Solution

[Sauert 2012]

- **Low SNR:** Bandpass character approximately independent of noise

$$W_{\text{opt},i}(\kappa) \approx \sqrt{\frac{I_i}{\sum\limits_{\zeta} I_\zeta} \cdot \frac{\mathfrak{P}^{\text{max}}(\kappa)}{P_{s,i}^{\text{in}}(\kappa)}} \quad \text{for all subbands } i$$

- **Medium SNR:** Spectral shape of output speech roughly follows noise

$$E_{\text{opt},i}(\kappa) \rightarrow D_i(\kappa) + 15\,\text{dB}$$

- **High SNR:** Transparent spectral weights

$$W_{\text{opt},i}(\kappa) \rightarrow 1$$

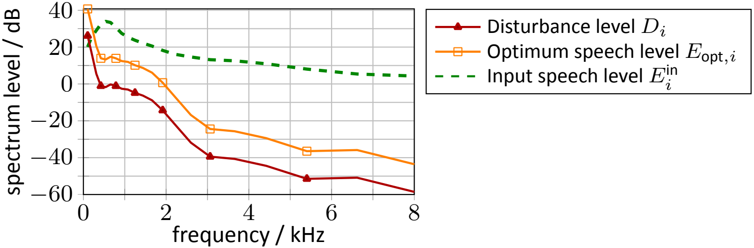
Analysis for Narrow Bandpass Noise [Sauert 2012]

- **Medium SNR:** $E_{\text{opt},i}(\kappa) \to D_i(\kappa) + 15\,\text{dB}$
- **Special noise scenarios**, e.g., car interior noise
  - Noise energy is concentrated in few subbands

  $\Rightarrow$ Speech energy will be concentrated in few subbands

  - Strong lowpass effect, not covered by SII

- **Example: Car interior noise** *without increase* of total audio power ⊜



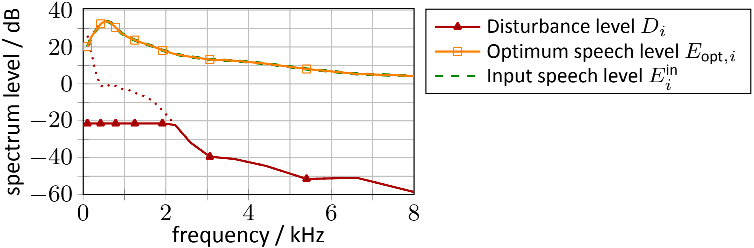
A Priori Limitation of Disturbance (OptSIIrecurDist)

- **Problem:** D_i has high values in only few subbands
- **Countermeasure:** Restrict D_i relative to average (in dB)

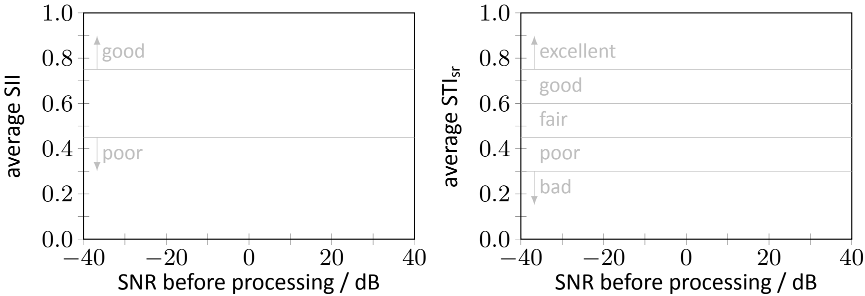
$$D_i'(\kappa) = \min\left\{ D_i(\kappa),\ \frac{1}{I}\sum_{\zeta} D_\zeta'(\kappa) + D_\Delta \right\} \quad \text{with} \quad D_\Delta = 7\,\mathrm{dB}$$

- **Independent of overall disturbance level**

- **Example: Car interior noise** *without increase* of total audio power ⊜

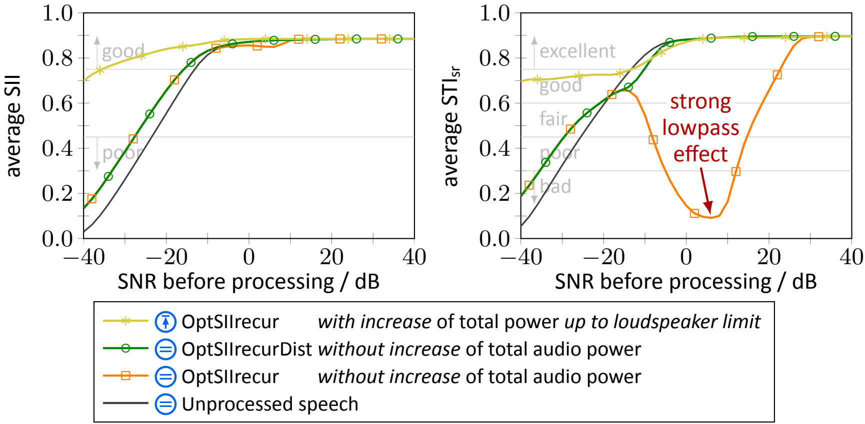


Simulation Results



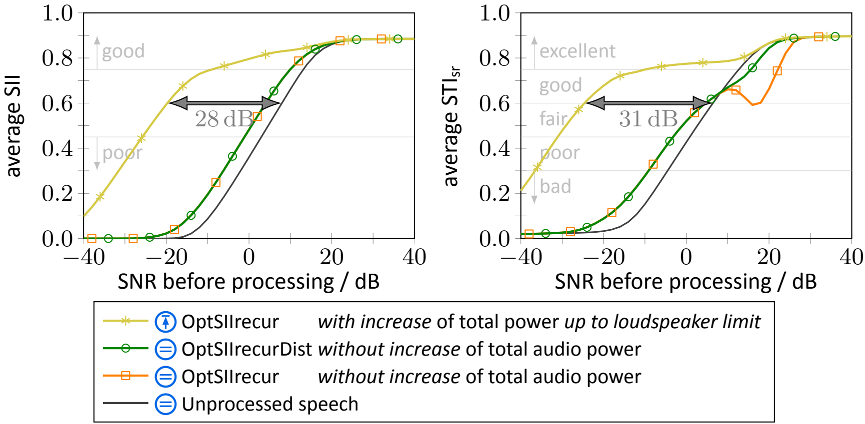
- **Audio data**
  - Speech: TIMIT database at sample rate 8 kHz with level 62.35 dB_{SPL}
  - Noise: NOISEX-92 database with level according to SNR
- **Measure:** *Speech Intelligibility Index* (SII)
- **Measure:** *Speech-based revised Speech Transmission Index* (STI_{sr})
  - Revised Speech Transmission Index (STI_r) **[Steeneken & Houtgast 1999]**
  - Extended to speech signals as stimuli **[Goldsworthy & Greenberg 2004]**

Simulation Results for Car Interior Noise



- ⊜ **2–5 dB SNR change without power increase**
- ⤒ **26–31 dB SNR change with power increase**

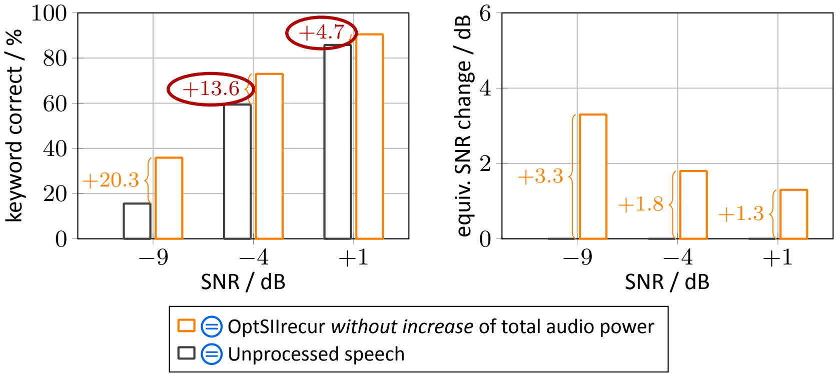
Simulation Results for Speech Babble Noise



- ⊜ **2–5 dB SNR change without power increase**
- ⍑ **26–31 dB SNR change with power increase**

Large Scale Listening Test

[Cooke, Sauert et al. 2013]

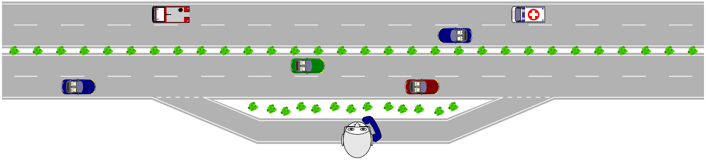


- **'The Listening Talker' (LISTA) project funded by EU (2010-2013)**
- **Invited to large scale listening test with 139 listeners**

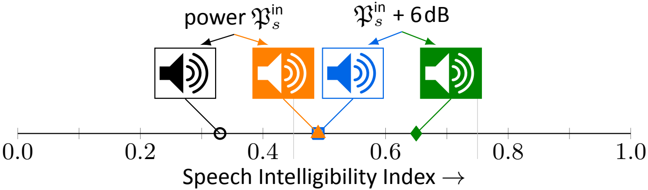
Significant improvement of word recognition rate by OptSIIrecur

Performance in Real Environment

▶ Near-end listener at a motorway station



SNR: -3 dB



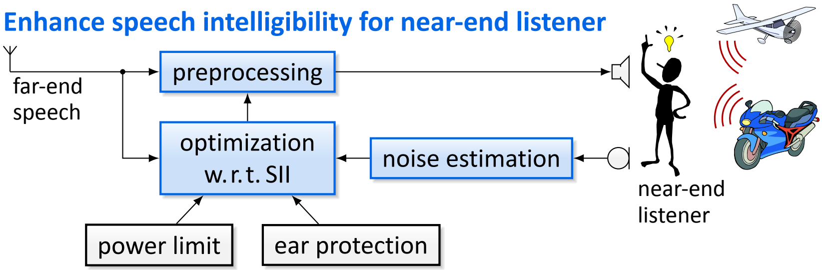
- ○ ⊜ Unprocessed speech
- ▲ ⊜ OptSIIrecurDist *without increase* of total audio power
- □ ⊕ Flat amplification *by 6 dB*
- ◆ ⊕ OptSIIrecurDist with increase of total audio power *by 6 dB*

Applications for Near-End Listening Enhancement

- Mobile phone
- Hands-free kit
- Binaural headset & headphones
- Car multimedia
- Public address system
- Hearing aids

Summary

- **Enhance speech intelligibility for near-end listener**



- **Up to 22 percentage points increase of word recognition rate** ***without increase*** **of total audio power ⊜**
- **Developed concepts can be applied in many different devices**
- **For details please refer to my PhD thesis:**

  `http://www.ind.rwth-aachen.de/~bib/sauert14/`