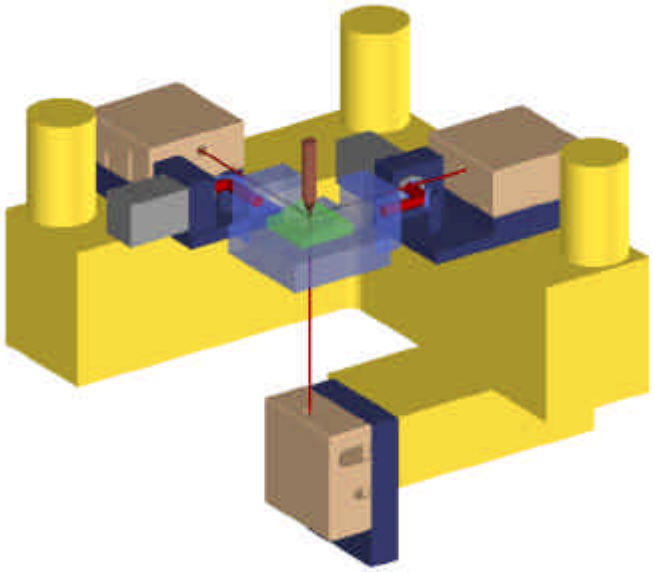
Untersuchung von Verfahren zur Messdatenreduktion und –kompression für den Einsatz in einer Nanomessmaschine

Dipl.-Ing. T. Machleidt
PD Dr.-Ing. habil. K.-H. Franke
Fachgebiet Graphische Datenverarbeitung / TU-Ilmenau

Gliederung:

- Einleitung / Motivation
- Zielstellung der Reduktion
- Kompressionsverfahren
- Ergebnisse und Bewertung
- Zusammenfassung



TU Ilmenau
Leitung: Prof. Jäger

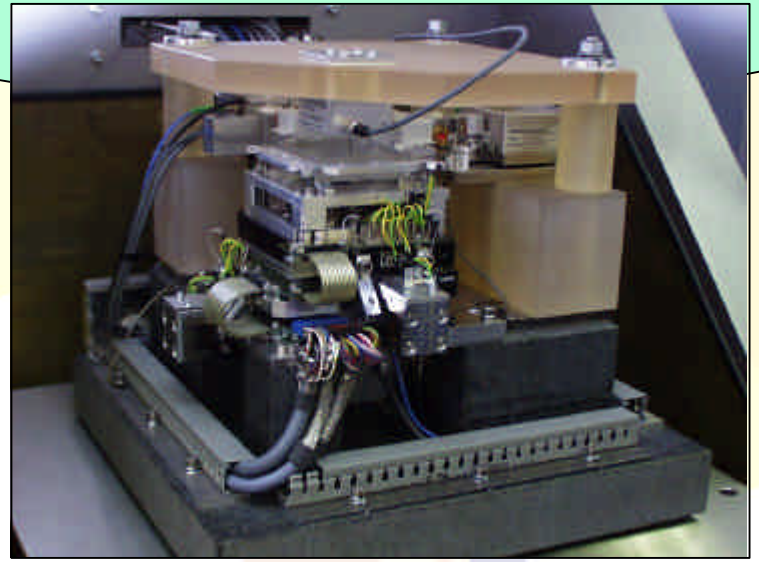
SIOS Meßtechnik GmbH

TMWFK[1]

zbs

Nanopositionier- und Nanomessmaschine

- Seit 1996 durch TMWFK[1] gefördert
- 2000 konnte erste Maschine fertig gestellt werden
- Seit Mitte 2002 SFB an der TUI
- Anfang 2003 Thüringer Forschungspreis 2002 verliehen



[1]Thüringer Ministerium für Wissenschaft, Forschung und Kunst

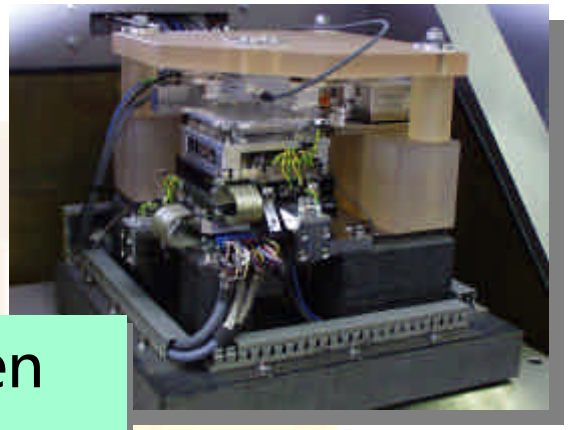
Messvolumen: $25x25x5mm^3$

Auflösung: 0.1nm

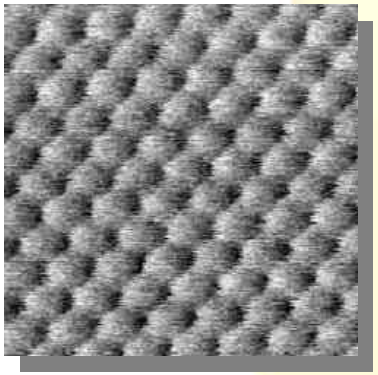
Weitere Informationen über die NPM-Maschine und ihre Entwickler finden Sie unter:
http://www.tu-ilmenau.de / http://www.sios.de / http://www.zbs-ilmenau.de

NMM
25x25x5mm³
Auflösung 0.1nm

Nanomessmaschine



Ziel? Aufgabe?



Vermessung von Oberflächen im Nanometerbereich
(z.Z. 25x25x5mm³, Auflösung von ca. 0.1nm)

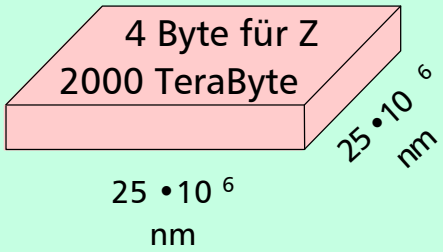
Anforderungen

an die eingesetzte Hardware

an die Messdaten-erfassung und -verarbeitung

Motivation: Messdatenreduktion

Datenvolumen bei
1 nm Auflösung



34000 Festplatten je 73 GByte

Messzeit bei 160 μs pro Messung
ca. 3170 Jahre

Wie viele Messdaten entstehen?

Wie groß ist der Datenstrom?

Wozu Reduktion der Daten?

Zusammenhang zwischen Messziel und Datenreduktion?

d.h. komplette Oberfläche kann nicht abgetastet werden

Motivation: Messdatenreduktion

Sensor-Arrays
(z.B. 64x64 Sensoren)

50 μm x 50 μm $\xrightarrow{2\ \text{nm}}$ 625 Millionen Messpunkte (ca. 2 GByte)

Reduzierung der Datenmenge erforderlich
(anwendungsspezifisch ➲ Datenübertragung lokal u. global,
Archivierung, Visualisierung, Verarbeitung)

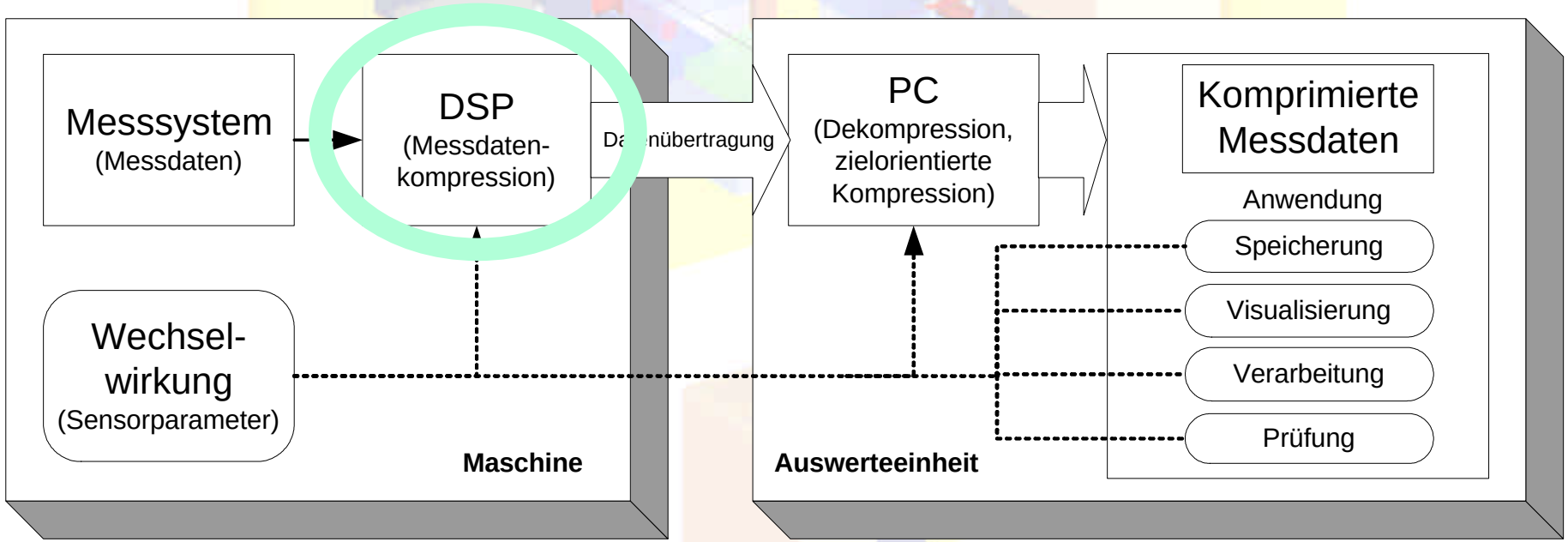
verlustfrei
(Lineare Prädiktive
Kodierung)

verlustbehaftet
(ADPCM, krümmungs-
abhängige Datenreduktion)

abstrakte 3D-Daten-
primitive (Triangulation,
Regelgeometrien)

Analyse: Messdatenreduktion

Zielstellung: spezielle Verfahren für die NPM entwickeln



Schematischer Aufbau der Nanomessmaschine

Analyse: Messdatenkompression im DSP

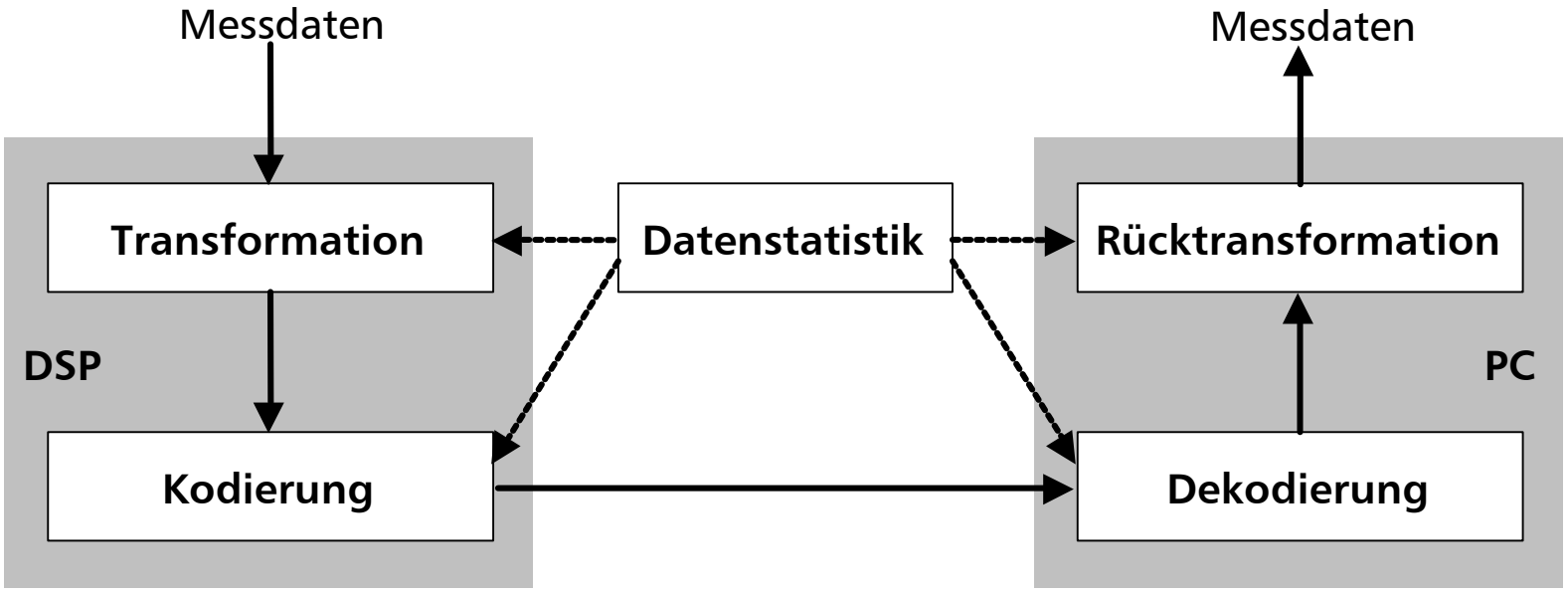
Randbedingungen:

- Kompression des Datenstroms **während** der Messung
- Störungen in den Messdaten können nicht ausgeschlossen werden
- Begrenzte Ressourcen im DSP vorhanden (Rechenleistung, Speicher)
- Kompressionsverfahren dürfen Mess- und Regelsystem nicht beeinträchtigen

Vorgehensweise:

- **verlustfreie** Kompression der Messdaten
- Ursprungsdaten direkt zu Kodieren ist nicht effizient
- Ursprungsdaten mittels vorgeschalteter Transformation in eine Folge gut komprimierbarer Daten umwandeln
- Einsatz von Verfahren der Optimalkodierung (z.B. Huffman-Kodierung)
- Anpassung der Verfahren an die Messdatenstatistik (Musterdatensätze oder adaptiv)

Verfahren: Messdatenkompression im DSP



Einfaches Schema des optimalen Datenflusses vom DSP zum PC

Transformation der Messdaten:

- Transformation der Messdaten in den ganzzahligen Bereich unter Verwendung der maximalen Auflösung
- Differenzbildung benachbarter Messdaten
- Schätzung (Prädiktion) des aktuellen Messwertes aus den vorangegangenen Messwerten und Bildung der Differenz

Verfahren: Messdatentransformation

Umwandlung vom Datentyp Double in den Datentyp Integer (64 Bit auf 32 Bit)

- Wandlung der Messdaten von der Einheit Meter in Ångström
- Daten in Integer konvertieren (Abschneiden der Nachkommastellen)

Differenzbildung benachbarter Messwerte

- Ziel ist Folge gut kodierbarer Messdaten (Verringerung der Entropie)

$$d[n] = z[n] - z[n-1]$$

z[n] – aktueller Messwert

Lineare Prädiktion

- Ziel ist Folge gut kodierbarer Messdaten (Verringerung der Entropie)

$$\hat{z}[n] = \sum_{k=1}^{p} a_k \cdot z[n-k] \qquad e[n] = z[n] - \hat{z}[n]$$

P – Ordnung

a - Prädiktorkoeffizienten

$\hat{z}[n]$ – Schätzwert

e(n) - Prädiktionsfehler

Verfahren: Kodierung

Was ist Kodierung?

- Vorgang der Umwandlung einer Information (aus einer bestimmten Darstellung in eine andere Darstellung)
- Rückwandlung wird als Dekodierung bezeichnet

Huffman Kodierung

- zu jedem Wert wird die Auftrittswahrscheinlichkeit bestimmt
- anhand dieser Wahrscheinlichkeiten wird ein Binärbaum erzeugt (Werte in den Blättern, Wahrscheinlichkeiten in den inneren Knoten)
- Kode wird über Pfad im Baum von der Wurzel bis zum Blatt bestimmt (im linken Teilbaum steht Kode „0“, im rechten Teilbaum steht Kode „1“)

Arithmetische Kodierung

- zu jedem Wert wird die Auftrittswahrscheinlichkeit ermittelt
- Wahrscheinlichkeiten teilen den Raum zwischen 0 und 1 in Teilbereiche
- auftretender Wert -> der jeweilige Teilbereich wird wieder nach den Wahrscheinlichkeiten unterteilt
- Übertragen wird eine Zahl aus dem kleinsten Teilbereich

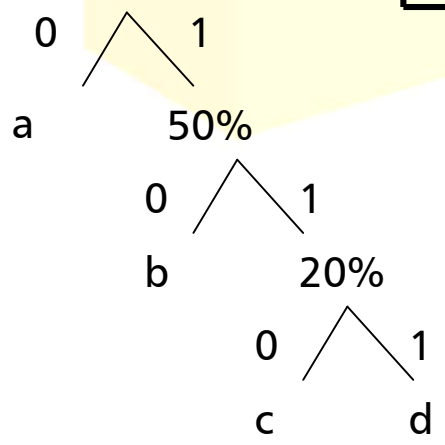
Verfahren: Kodierung

Beispiel Huffman-Kodierung

Zeichensatz:

Zeichen	Wahrscheinlichkeit (%)
a	50
b	30
c	10
d	10

Baum:

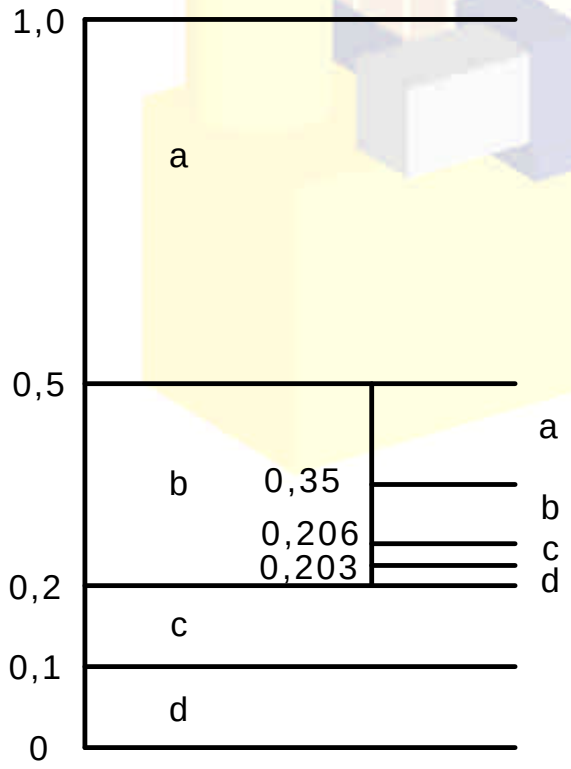


Kodetabelle:

Zeichen	Kode
a	0
b	10
c	110
d	111

Verfahren: Kodierung

Beispiel arithmetische Kodierung

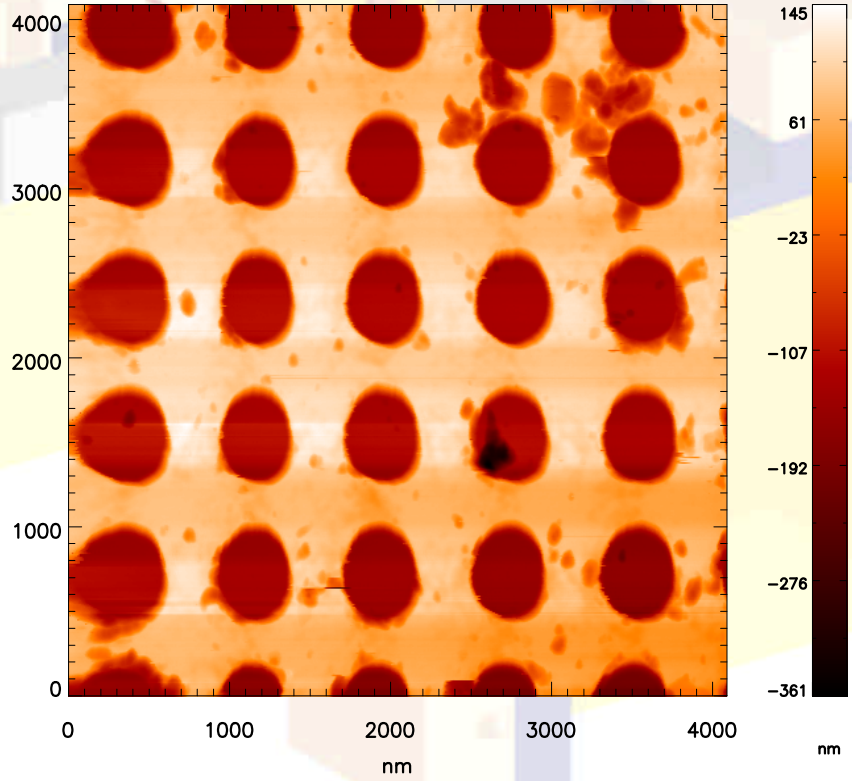


Zeichen	Wahrscheinlichkeit (%)
a	50
b	30
c	10
d	10

für die Zeichenkette bc wird eine Zahl zwischen 0,203 und 0,206 kodiert

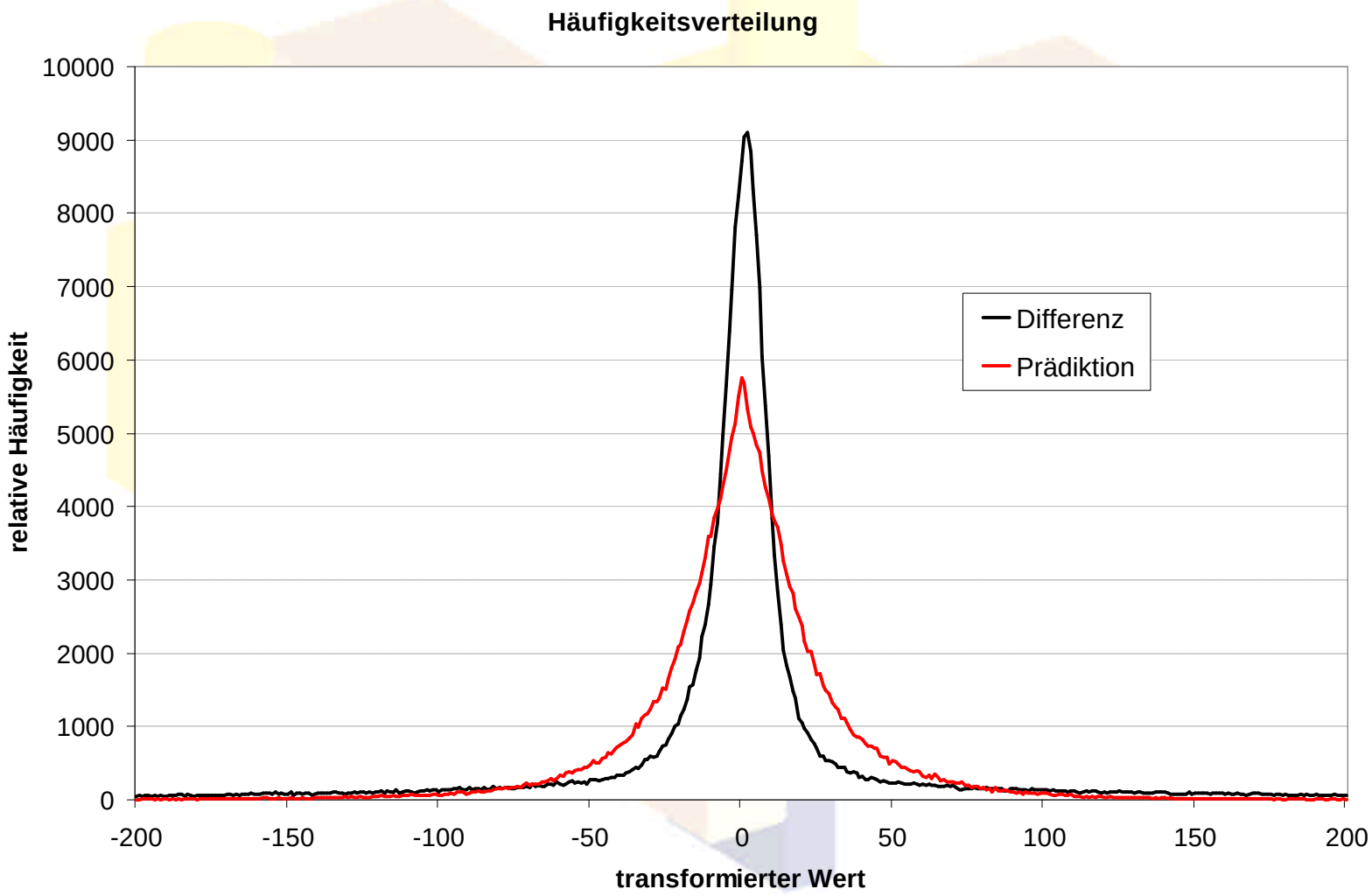
Anwendung: Beispieldaten

Messdaten NPM
(499 x 500 Messpunkte)



Datenvolumen ohne Kompression:
8Byte * 499 = 3992 Byte/Messzeile

Anwendung: Häufigkeitsverteilung für transformierten Messwert



Anwendung: Ergebnisse der Messdatenkompression im DSP

Transformation	**Differenz**		**Lineare Prädiktion**	
Kodierungsverfahren	Huffman	Arithmetisch	Huffman	Arithmetisch
Anzahl kodierten Linien	500	500	500	500
Entropie	7,2501 Bit	7,2501 Bit	7,0019 Bit	7,0019 Bit
Mittlere kodewortlänge **Mittlere kod. Linienlänge**	7,297 Bit 455,13 Byte	7,293 Bit 454,92 Byte	7,0563 Bit 440,14 Byte	7,0484 Bit 439,64 Byte
Mittlere Rechenzeit je Linie:				
· **Kodierung**	5,9 ms	5,3 ms	9,2 ms	6,2 ms
· **Dekodierung**	56,5 ms	30,6 ms	43,1 ms	40,7 ms
Datenmenge, die im Arbeitsspeicher des DSP zu halten ist	2375 KB[1] Kodebuch	32 KB Häufigkeiten	2872 KB[1] Kodebuch	32 KB Häufigkeiten

[1]Dieser Wert ist variabel.

Ergebnisse der Simulation auf P4 2,2 GHz

Zusammenfassung und Ausblick

- Nanomesstechnik stellt hohe Anforderungen an die Datenkompression
- Messdatenkompression im DSP als erster Schritt
- Anwendung von Transformationen zur Verbesserung der Redundanz
- Vorstellung von optimalen Kodierverfahren
- Simulationen mit den vorgestellten Verfahren ergaben einen Kompressionsfaktor von 8

- Ausbau der Linearen Prädiktion (Adaptive Ordnungs- und Koeffizientenberechnung, Einbeziehung der Charakteristik des Sensors)
- Untersuchung von adaptiven Kodierverfahren
- Integration der Verfahren in das DSP-System der Nanomessmaschine
- Bewertung der Verfahrenskombination für unterschiedliche Messstrategien und Oberflächen
- Untersuchung von zielorientierten Datenkompressionsverfahren im PC