

# Homework 8

## Einhüllende & nicht lineare Dynamik Kompression

Stefan Röhl

Technische Universität München, Arcisstraße 21, Munich, Germany

Email: stefan.roehl@tum.de

### I. EINHÜLLENDE

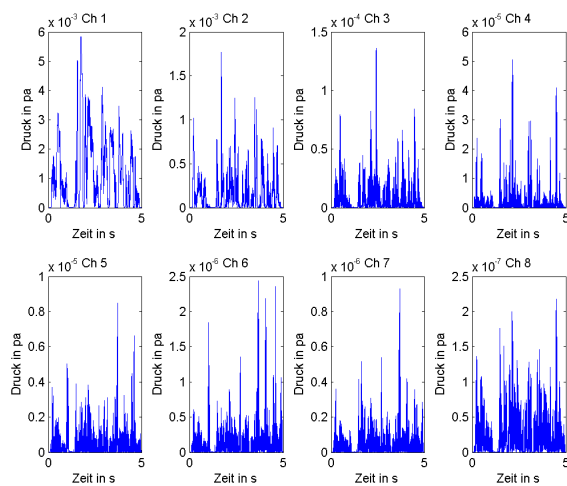


Figure 1. Einhüllende von 8 Kanälen

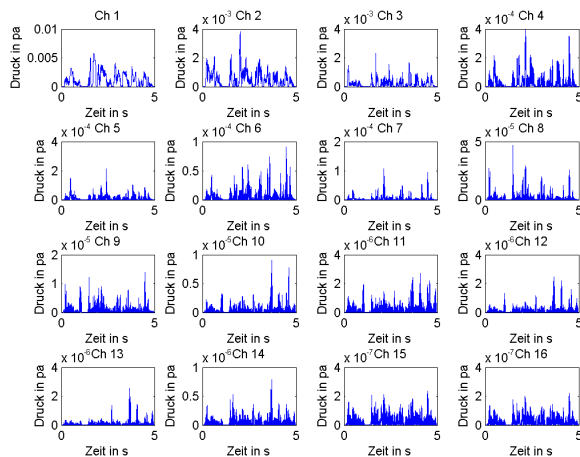


Figure 2. Einhüllende von 16 Kanälen

die Einhüllenden des Signal berechnet. Die jeweiligen Einhüllenden für jeden Kanal der beiden Filterbänke sind in Abbildung 1 und 2 zu sehen. Auffällig ist das Kanäle für tiefe Frequenzen einen weitaus höheren Pegel aufweisen, als Kanäle für hohe Frequenzen.

- a) Das zu verarbeitende Sprachsignal enthält folgenden englischen Satz: *In language infinitely many words can be written with a small set of letters.* Dieses Signal wurde zuerst mit einer 8- bzw. 16-Kanal-Filterbank in unterschiedliche Frequenzbänder aufgeteilt. Von den Ausgängen jedes Kanals wurden mit der Hilbert Transformation

- b) In Abbildung 3 ist das originale Zeitsignal zu erkennen, welches eine Ausdehnung von 5 Sekunden hat. Darunter in Abbildung 4 ist das komplette Frequenzspektrum des unveränderten Signals zu sehen. Da es sich um ein menschliches Sprachsignal handelt, liegen die Frequenzanteile zwischen 100 und ca. 8kHz. Das Kurzzeit-Spektrogramm

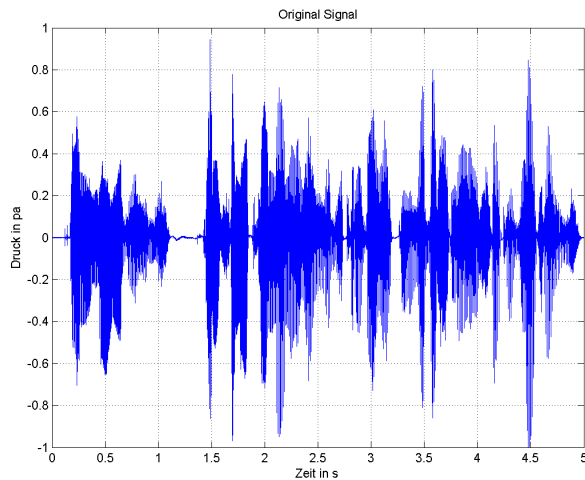


Figure 3. Original Signal

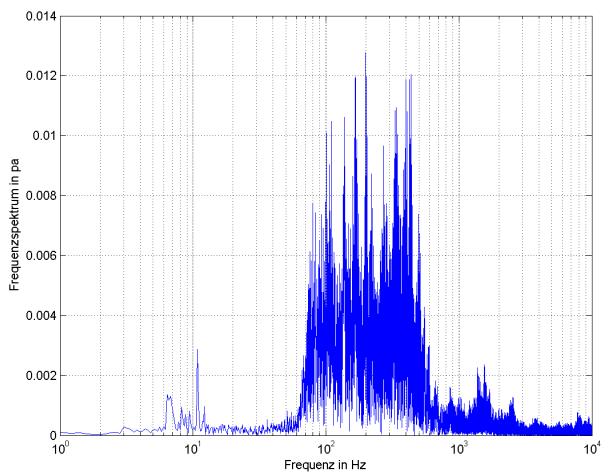


Figure 4. Frequenzspektrum Original Signal

zeigt die Frequenzanteile nicht komplett für den ganzen Signalverlauf sondern immer für den jeweiligen Zeitabschnitt im Signal. Auch hier sieht man wieder die Unterschiede zwischen dem original Signal und den gefilterten Signalen. Man muss jedoch beachten, dass es sich bei den Signalen aus der Filterbank nur noch um die Einhüllenden des Signals handelt und diese nicht direkt mit dem original Signal vergleichbar sind. Man sieht jedoch auch wieder deutlich die höhere Frequenzauflösung der 16-Kanal-Filterbank im Gegensatz zur 8-Kanal-Filterbank. Des weiteren fällt auf, dass die Kanäle nicht richtig ausgelastet sind, nur wenige Stellen sind heller gefärbt als dunkelblau. Dies soll mit einer Dynamikkompensation verbessert werden.

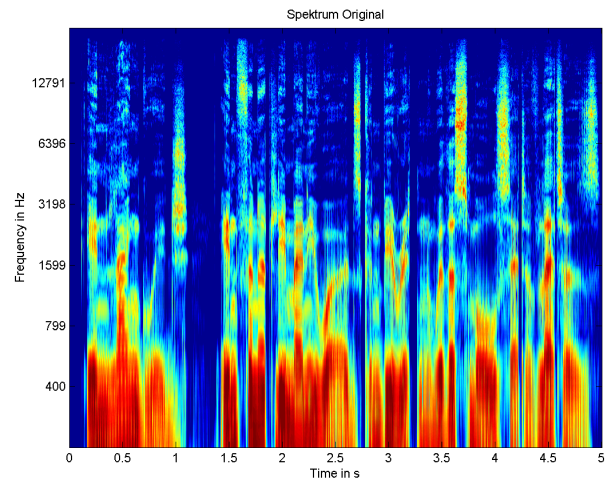


Figure 5. Kurzzeit-Spektrogramm Original Signal

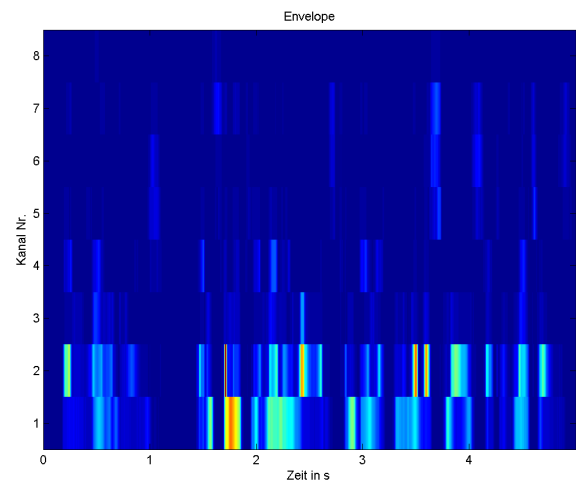


Figure 6. Kurzzeit-Spektrogramm der Einhüllenden mit 8 Kanälen

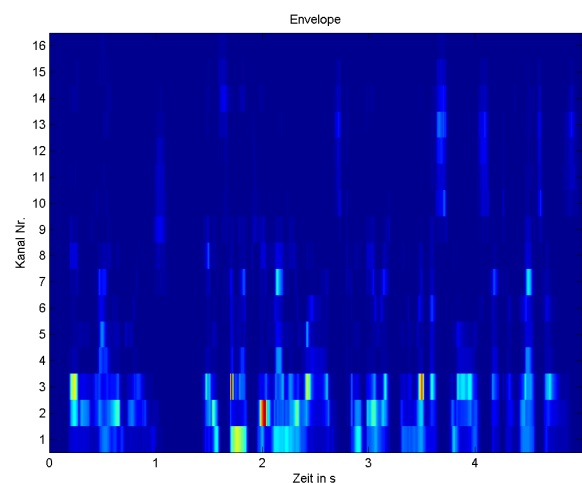


Figure 7. Kurzzeit-Spektrogramm der Einhüllenden mit 16 Kanälen

Das Summenspektrum zeigt den mittleren Pegel eines

Kanals - also ein bestimmtes Frequenzband - über der Zeit. In diesem Spektrum kann man erkennen, dass besonders die hohen Frequenzen nur einen kleinen Anteil am Gesamtpegel haben. Dies ist sowohl bei der 8 und der 16-Kanal-Filterbank der Fall.

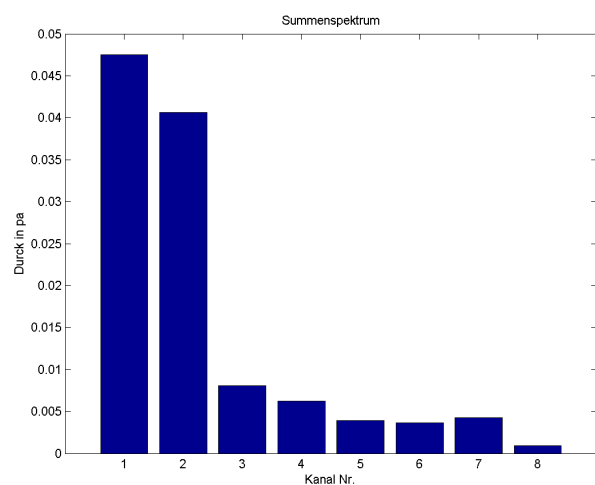


Figure 8. Summenspektrum der Einhüllenden mit 8 Kanälen

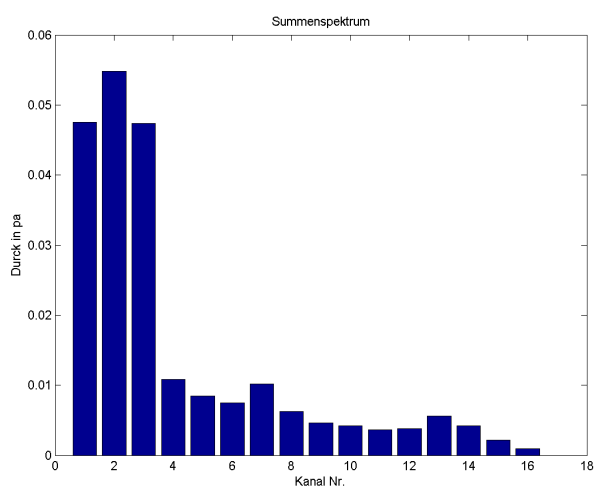


Figure 9. Summenspektrum der Einhüllenden mit 16 Kanälen

## II. DYNAMIKKOMPRESSION

- a) Da das gesprochene Signal einen ziemlich großen Dynamikbereich hat, welcher so nur schlecht zu kodieren ist, wird eine nicht lineare Dynamikkompensation vorgenommen. Diese verstärkt leise Signale mehr als bereits laute Signale. Somit ist der unterschied zwischen lauten und leisen Abschnitten nicht mehr so groß und es muss nur noch ein kleinerer Lautstärkebereich abgebildet werden. Nach der Dynamikkompensation befinden sich die Pegel der Einhüllenden alle in der Größenordnung  $10^{-3}$  Pa (vgl. Abb. 10 und 11). Nicht so wie vor der Dynamikkompensation, wo Kanäle mit niedrigen Frequenzen Pegel in der Größenordnung von  $10^{-1}$  Pa aufweisen und Kanäle mit hohen Frequenzen im Bereich von  $10^{-7}$  Pa arbeiten (vgl. Abb. 2).

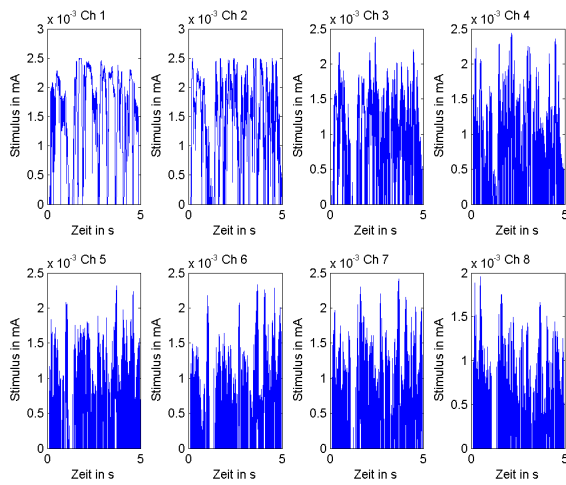


Figure 10. Einhüllenden mit Dynamikkompensation von 8 Kanälen

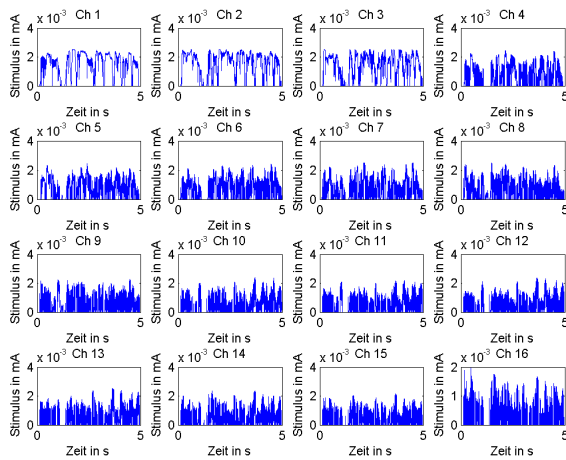


Figure 11. Einhüllenden mit Dynamikkompensation von 16 Kanälen

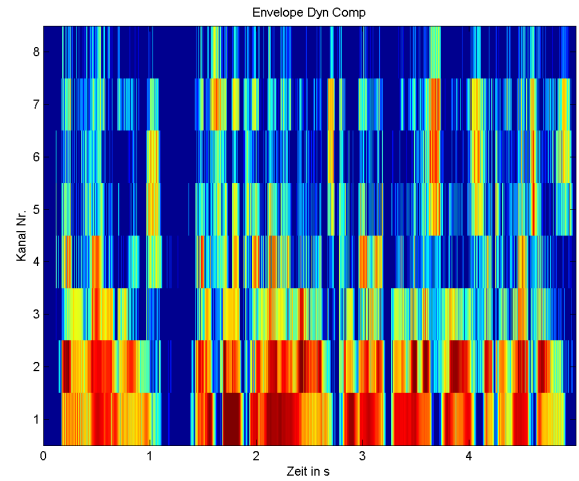


Figure 12. Kurzzeit-Spektrogramm der Einhüllenden mit 8 Kanälen nach Dynamikkompensation

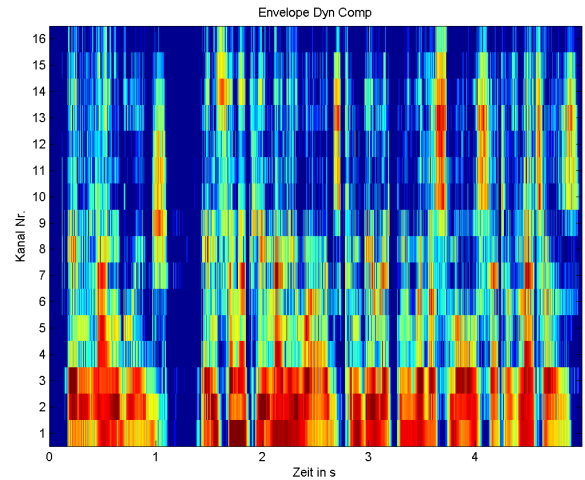


Figure 13. Kurzzeit-Spektrogramm der Einhüllenden mit 16 Kanälen nach Dynamikkompensation

- b) Auch in den Kurzzeit-Spektrogrammen ist nun eine wesentlich höhere Auslastung zu erkennen. Da Pegel der verschiedenen Kanäle nun wesentlich näher beieinander liegen, kann man diese besser kodieren was sich eben auch in der Farbkodierung zeigt (vgl. Abb. 12 und 13). Selbst die Kanäle mit hohen Frequenzen zeigen nun deutliche Pegel und das original Signal wird deutlich besser repräsentiert. Besonders die Darstellung mit 16 Kanälen ähnelt viel mehr dem Originalspektrum (vgl. Abb. 5).

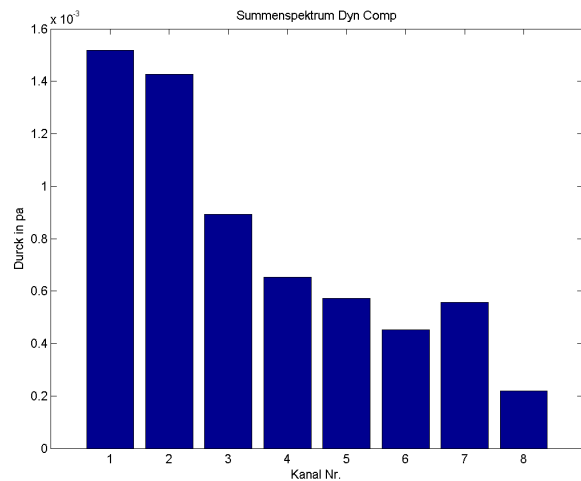


Figure 14. Summenspektrum der Einhüllenden mit 8 Kanälen nach Dynamikkompression

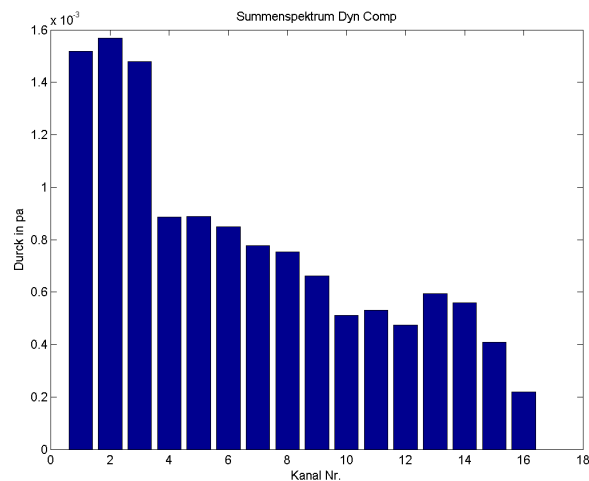


Figure 15. Summenspektrum der Einhüllenden mit 16 Kanälen nach Dynamikkompression

Auch in den Summenspektren sieht man eine deutliche Anhebung der hohen Frequenzen. Jedoch muss man auch feststellen, dass die Pegel der tiefen Frequenzbänder gestaut wurden. Somit spielt sich der Dynamikbereich nur noch zwischen  $10^{-3}\text{Pa}$  und  $10^{-4}\text{Pa}$  ab, das heißt die Frequenzbänder liegen vom Pegel her auf ungefähr dem selben Level. Dies ermöglicht es die selbe Kodierung für alle Kanäle zu verwenden, ohne zu viele Informationen durch Clipping oder schlecht Auflösung in den niedrigen Bereichen zu bekommen.

c) Das aufgenommene Signal muss für die Stimulation in Innenohr noch in eine Stromstärke umgerechnet werden. Die folgenden Abbildungen zeigen wie welcher Schallpegel mit der vorher vorgenommenen Dynamikkompression in welche Stromstärken umgewandelt werden. Dabei sind die Abbildungen 16 17 und 18 linear gezeichnet, wohingegen die Abbildungen 19 20 und 21 logarithmisch aufgetragen sind. Die Markierung  $I_{MCL}$  steht für die höchste zulässige Stromstärke, die noch keine Schmerzen beim Patienten auslöst und somit nicht überschritten werden darf. Daher knickt die Kennlinie an dieser Stelle ab und es kommt zu Clipping. Die Markierung  $I_{THR}$  steht für den minimalisierten möglichen Strom, der kodiert werden kann. Signale mit kleinerem Pegel werden nicht ausgegeben. Alle Pegel dazwischen werden nicht-linear auf Stromstärken umgerechnet, wobei kleine Pegel mehr verstärkt werden als große. Es wurden drei verschiedene Kompressionsraten verwendet. Bei einer Kompressionsrate von 250 sieht man am logarithmischen Plot, dass sich die Kompression fast logarithmisch verhält, da im Plot eine Gerade zu sehen ist. Bei höheren Kompressionsraten wie 500 oder 1000 verhält sich die Kompression schon über-logarithmisch, da auch der logarithmische Plot keine Gerade mehr ist. Dies heißt hier werden kleine Pegel noch differenziert betrachtet und diesen ein nochmal weitaus größerer Dynamikbereich zur Verfügung gestellt, als den lauten Pegeln.

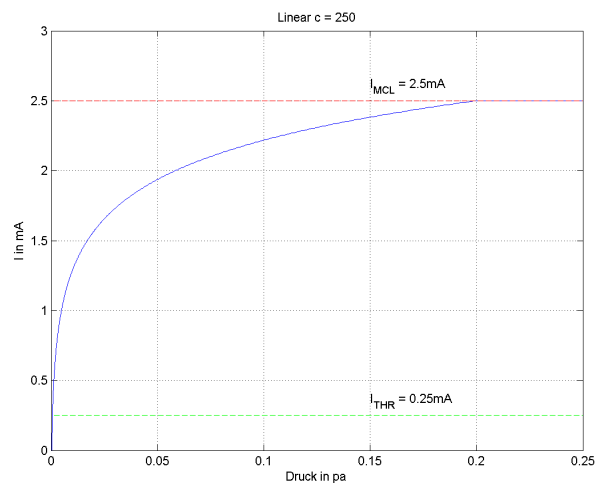


Figure 16. Kompressionskennlinie Linear mit Kompressionsrate 250

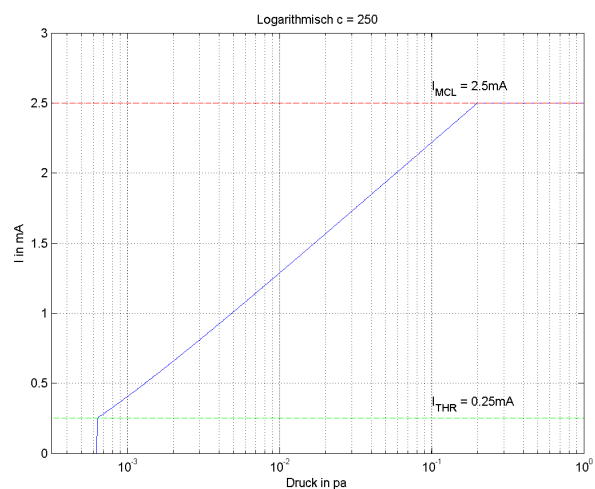


Figure 19. Kompressionskennlinie Logarithmisch Linear mit Kompressionsrate 250

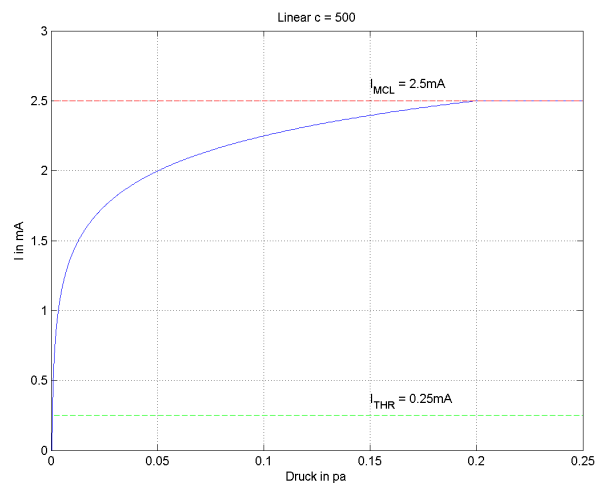


Figure 17. Kompressionskennlinie Linear mit Kompressionsrate 500

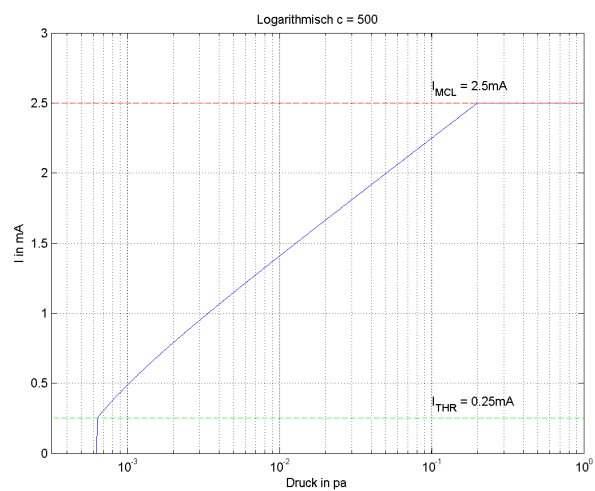


Figure 20. Kompressionskennlinie Logarithmisch Linear mit Kompressionsrate 500

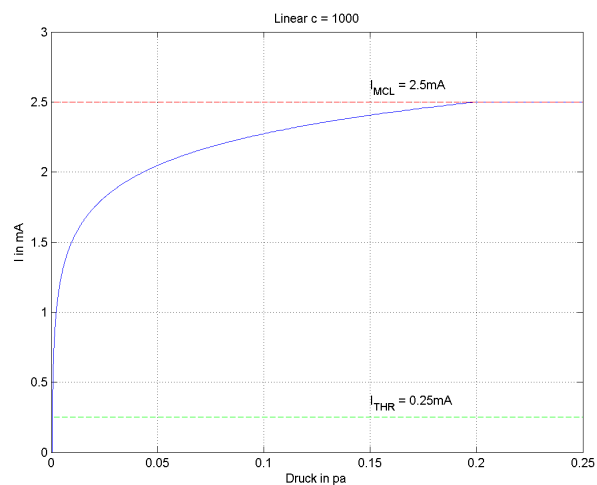


Figure 18. Kompressionskennlinie Linear mit Kompressionsrate 1000

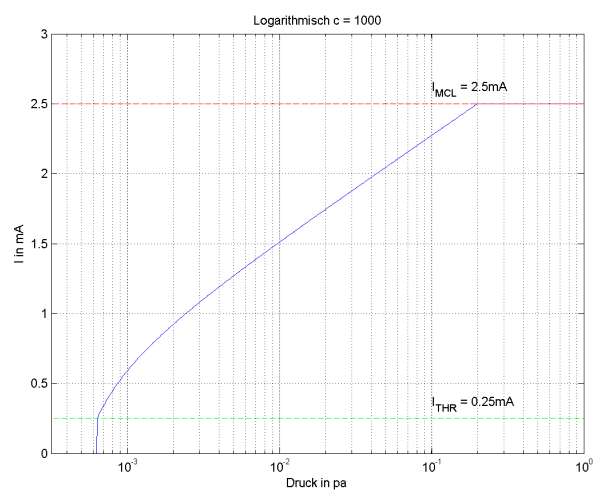


Figure 21. Kompressionskennlinie Logarithmisch Linear mit Kompressionsrate 1000