

Von den Gleichungen bis zur Implementierung

Digitale Vorverzerrung in der HF-Kommunikation

Dieser Artikel befasst sich mit den mathematischen Grundlagen der digitalen Vorverzerrung (Digital Predistortion, DPD) und ihrer Implementierung.

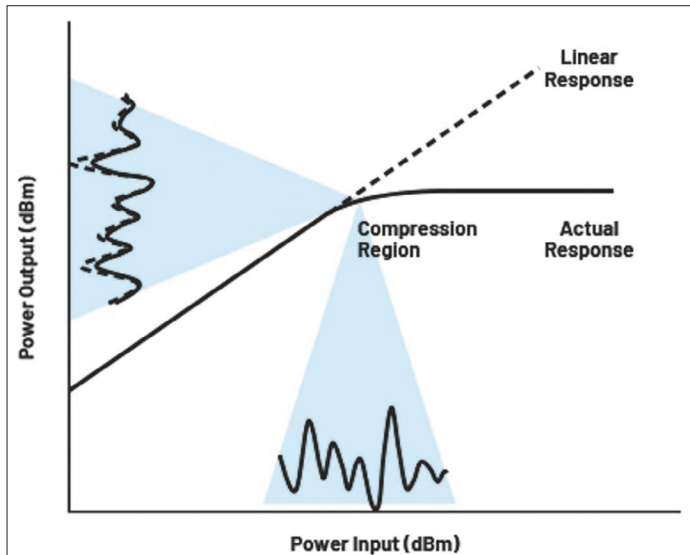


Bild 1: Plot PA-Eingangsleistung gegenüber PA-Ausgangsleistung

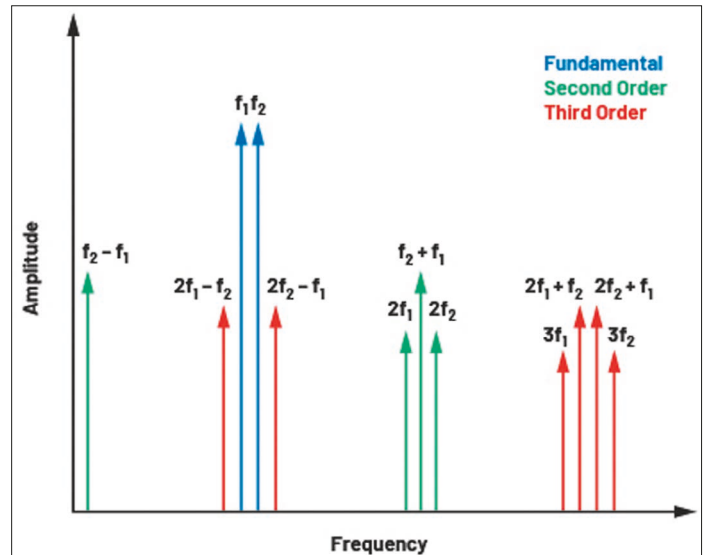


Bild 2: Lage der Intermodulationsprodukte und der harmonischen Störungen bei 2-Ton-Signal am Eingang

Es wird anhand eines Transceivers gezeigt, warum die DPD in modernen Kommunikationssystemen benötigt wird, und untersucht, wie das mathematische Modell die reale Signalverzerrung erfasst.

Einführung

DPD ist in aktuellen zellularen Kommunikationssystemen allgegenwärtig und stellt sicher, dass HF-Leistungsverstärker (PAs) effizient ihre maximale Leistung an eine Antenne liefern. Da 5G mehr Antennen in den Basisstationen bedingt und das Übertragungsspektrum breiter sein muss als bisher, hat sich DPD als eine Schlüsseltechnologie herauskristallisiert, die die Entwicklung effizienter, kostengünstiger und spezifikationskonformer Mobilfunksysteme ermöglicht.

Viele haben ihr eigenes Verständnis von DPD, sei es aus rein mathematischer Sicht oder wegen der eher eingeschränkten Implementierung in einem Mikroprozessor. Dieser Artikel soll ihr Wissen erweitern und sie

befähigen, das Thema besser zu erfassen.

Was ist DPD und warum wird sie verwendet? PAs sollen möglichst energieeffizient arbeiten, damit der größte Teil der dem Verstärker zugeführten Versorgungsleistung in HF-Ausgangsleistung umgewandelt wird. Der Aufmacher skizziert eine Senderstruktur mit und ohne DPD. Dabei ist Folgendes zu beachten. Transistoren sind von Natur aus nichtlinear. Wenn wir nun PAs in ihrem „linearen“ Bereich betreiben (linear ist hier ein relativer Begriff, daher die Anführungszeichen) wie in Bild 1 dargestellt, dann ist die Ausgangsleistung relativ proportional zur Eingangsleistung. Der Nachteil dieser Betriebsart ist, dass der PA im Allgemeinen in einem sehr ineffizienten Bereich arbeitet. Wir setzen PAs kurz vor dem Kompressionspunkt ein, was heißt: Wenn das Eingangssignal um einen bestimmten Betrag (z.B. 3 dB) erhöht wird, steigt die PA-Ausgangsleistung nicht um denselben Betrag (sondern vielleicht nur um 1 dB). Dabei

wird das Signal an diesem Punkt erheblich verzerrt.

Diese Verzerrungen treten abhängig vom Eingangssignal

Die Autorin:

Claire Masterson ist Systemingenieurin in der Wireless Systems Group bei Analog Devices Limerick und arbeitet an der Systemimplementierung, Softwareentwicklung sowie der Entwicklung und Verifizierung von Algorithmen. Sie erhielt einen B.A.I. und einen Dokortitel vom Trinity College Dublin und kam nach ihrem Abschluss 2011 zu ADI. Ihr besonderes Interesse gilt der Anwendung digitaler Signalverarbeitung in realen Systemen, insbesondere bei der Entwicklung von 5G- und 6G-Systemen und DPD-Implementierungen der nächsten Generation. Sie ist zu erreichen unter [claire.masterson@analog.com](mailto:masterson@analog.com).

Autorin:
Claire Masterson
Analog Devices
www.analog.com

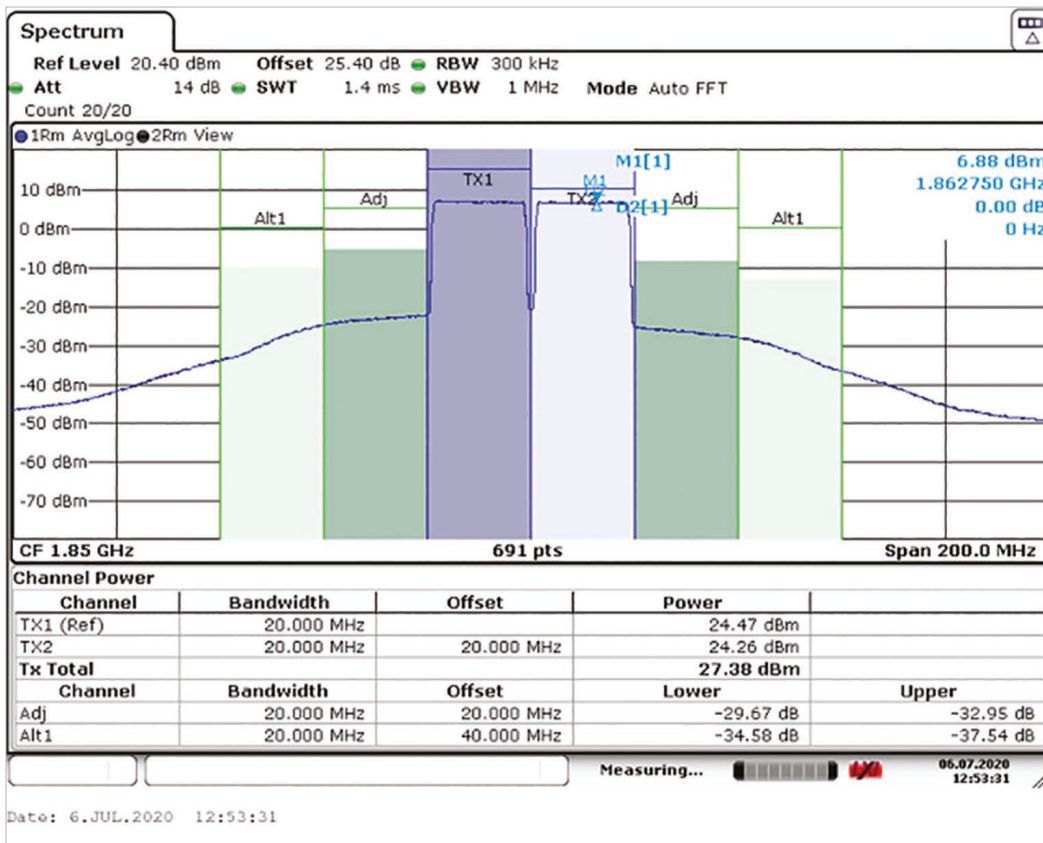


Bild 3: 2x20 MHz Träger an einem SKY66391-12 RF PA mit Mittenfrequenz 1850 MHz

bei bestimmten Frequenzen auf. Bild 2 zeigt diese und die Beziehung zwischen den Grundfrequenzen und deren Mischprodukten. In HF-Systemen müssen nur die Verzerrungen kompensiert werden, die in der Nähe der Grundschwingung auftreten, d.h., die Intermodulationsprodukte ungerader Ordnung. Die Filterung im System kümmert sich um die Produkte außerhalb des Bandes (Oberwellen und Intermodulationsprodukte gerader Ordnung). Bild 3 zeigt den Ausgang eines PAs, der in der Nähe seines Kompressionspunktes betrieben wird. Die Intermodulationsprodukte (insbesondere die dritter Ordnung) sind deutlich zu erkennen. Sie sehen aus wie „Röcke“ um das gewünschte Signal.

DPD charakterisiert diese Verzerrungen, beobachtet das PA-Ausgangssignal und verändert das Eingangssignal so, bis das PA-Ausgangssignal sich dem Idealwert nähert und nicht vom linearen Betrieb in den nichtlinearen Betrieb übergeht. Dies kann nur unter ganz bestimmten Bedingungen effizient erfolgen: Wir müssen den Verstärker und das Eingangssignal so konfigurieren, dass der Verstärker etwas komprimiert (Ausgangs-1dB-Kompressionspunkt, OP1dB), aber nicht vollständig in die Sättigung geht.

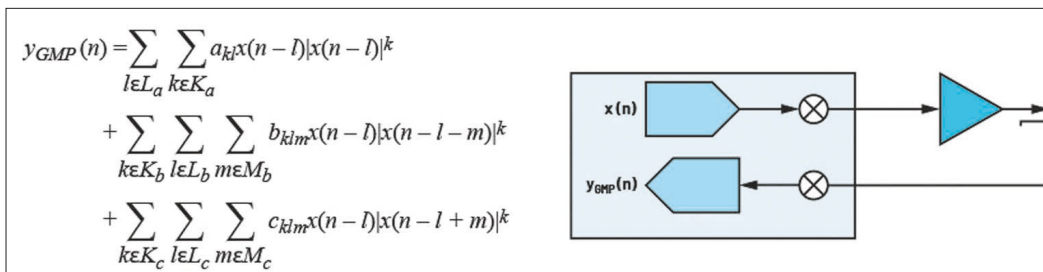


Bild 4: GMP zur Modellierung der PA-Verzerrungen [1]

Mathematik zur PA-Verzerrungen

Die Arbeit [1] ist zwar ein bahnbrechendes Werk, für eine Einführung in das Thema aber vielleicht etwas zu kompliziert. Versuchen wir also zunächst, den dortigen Generalized-Memory-Polynomial-Ansatz (GMP) aufzuschlüsseln, um zu einem intuitiveren Verständnis der mathematischen Zusammenhänge zu gelangen!

Die Volterra-Reihen sind das mathematische Rückgrat der DPD und werden verwendet, um nichtlineare Systeme mit Speicher zu modellieren. „Speicher“ bedeutet einfach, dass die aktuelle Ausgabe des Systems von den aktuellen und vergangenen

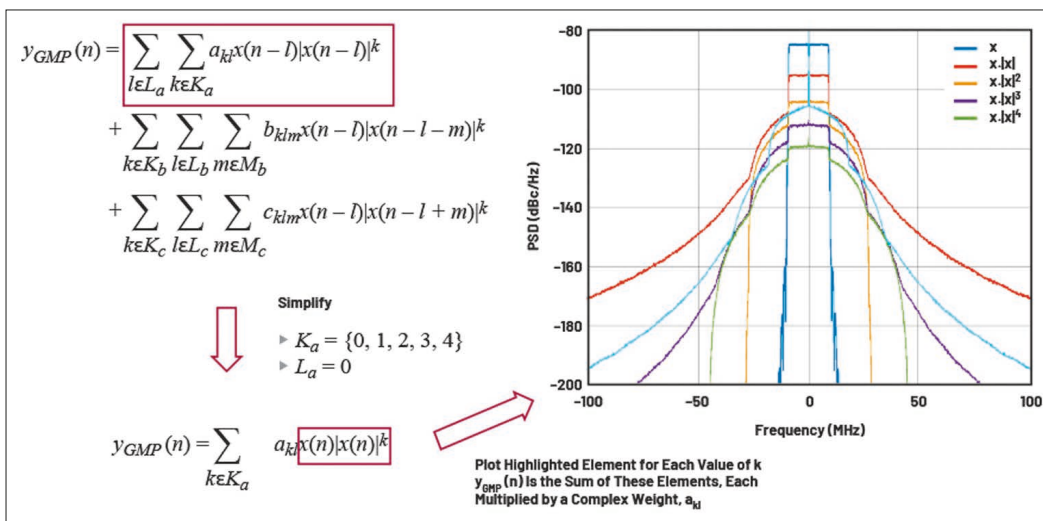


Bild 5: Plot des Effektes der Ordnung (k) auf Signale in der Frequenzdomäne des Signals x

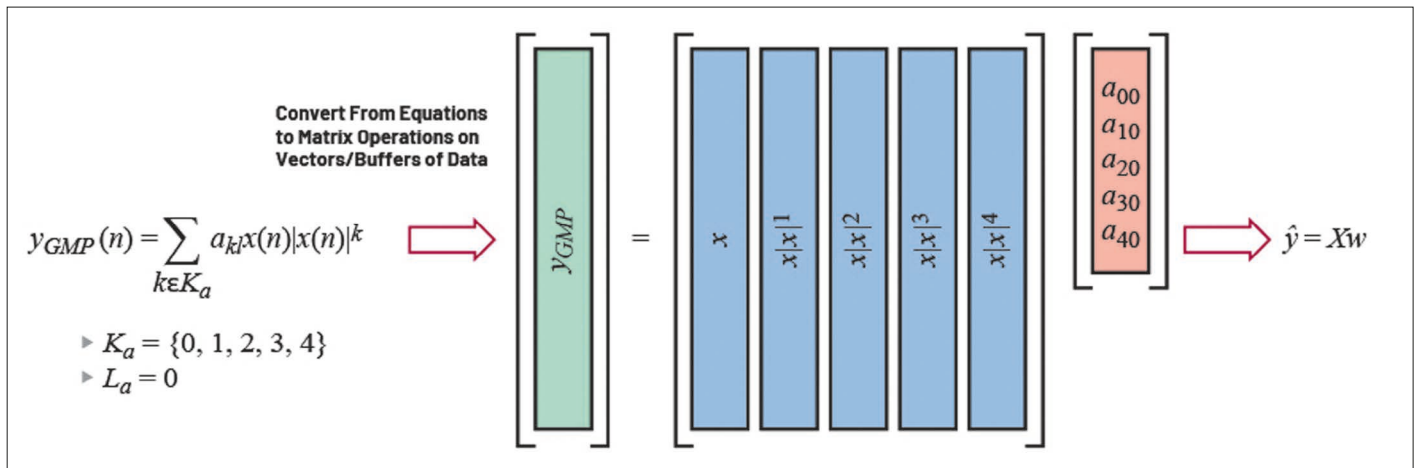


Bild 6: Umwandlung der vereinfachten Gleichung in eine Matrixform mittels Datenpuffer (so wie sie digital umgesetzt wird)

Eingaben abhängen kann. Die Volterra-Reihen sind sehr allgemein (und daher leistungsfähig) und werden auch in vielen Bereichen außerhalb der Elektrotechnik verwendet. Für die DPD in PAs können die Volterra-Reihen verschlankt und so gestaltet werden, dass sie in digitalen Echtzeitsystemen besser implementierbar und stabiler sind. GMP ist ein solcher verschlankter Ansatz.

Bild 4 beschreibt, wie GMP verwendet wird, um die Beziehung zwischen dem Eingang x des PA und seinem Ausgang y zu modellieren. Es ist zu sehen, dass die drei separaten Summenblöcke der Gleichung einander sehr ähnlich sind. Konzentrieren wir uns zunächst auf den ersten, der unten rot hervorgehoben ist. Der Term $|x(\dots)|^k$ ist die Hüllkurve des Eingangssignals, wobei k die Polynomordnung ist. l bezieht Memory in das System ein. Wenn $L_a = \{0, 1, 2\}$ ist, dann ermöglicht das Modell, dass die Ausgabe $y_{GMP}(n)$ von der aktuellen Eingabe $x(n)$ und den vergangenen Eingaben $x(n-1)$ und $x(n-2)$ abhängt.

In Bild 5 wird der Effekt der Polynomordnung k auf einen Beispielvektor untersucht. Der Vektor x ist ein einzelner 20-MHz-Träger und wird im komplexen Basisband gezeigt. Die GMP-Modellierungsgleichung wird durch Entfernen der Speicherkomponente vereinfacht. Die Darstellungen von

$x|x|^k$ zeigen eine deutliche Ähnlichkeit mit der in Bild 3 sichtbaren realen Verzerrung.

Jede Polynomordnung (k) und Speicherverzögerung (l) hat eine zugehörige komplexe Gewichtung (a_{kl}). Nachdem die Komplexität des Modells bestimmt wurde (welche Werte von k und l einbezogen werden), muss nach diesen Gewichtungen auf der Grundlage realer Beobachtung des PA-Ausgangs für ein bekanntes Eingangssignal gelöst werden. In Bild 6 wird die vereinfachte Gleichung in eine Matrixform umgewandelt. Die verwendete mathematische Notation ermöglicht eine übersichtliche Darstellung des Modells. Für die tatsächliche Implementierung von DPD auf Puffern mit digitalen Daten ist es jedoch am einfachsten und repräsentativer, alles in Matrixform zu betrachten.

Betrachten wir kurz die zweite und dritte Zeile der Gleichung in Bild 5, die der Einfachheit halber ignoriert werden. Wenn m auf Null gesetzt wird, sind diese Zeilen identisch mit der ersten. Diese Zeilen ermöglichen es, dass Verzögerungen (sowohl positive als auch negative) zwischen dem Hüllkurventerm und dem komplexen Basisbandsignal hinzugefügt werden. Diese werden als nacheilende und voreilende Crossterme bezeichnet und können die Modellierungsgenauigkeit von DPD erheblich verbessern. Sie bieten einen zusätzlichen Freiheitsgrad bei unseren Versuchen, das Verhalten des Verstärkers zu modellieren. Zu beachten ist, dass M_b , M_c , K_b und K_c keine Nullen enthalten; andernfalls würden sich Terme aus der ersten Zeile wiederholen.

Wie legen wir also die Ordnung des Modells fest, die Anzahl der

Speicherterme und welche Crossterme wir hinzufügen sollten? Hier kommt etwas „schwarze Magie“ ins Spiel. Wir können uns bis zu einem gewissen Grad von unserem Wissen über die Physik der Verzerrung leiten lassen. Die Art des Verstärkers, die Komponenten mit denen er aufgebaut ist und die Bandbreite des Signals, die gefordert ist, wirken sich alle auf die Modellierungsbedingungen aus und ermöglichen es einem auf diesem Gebiet erfahrenen Ingenieur, das zu verwendende Modell einzugrenzen. Darüber hinaus ist jedoch auch ein gewisses Maß an Trial and Error erforderlich.

Der letzte Aspekt ist die Frage, wie nach den Gewichtungskoeffizienten aufzulösen ist, nachdem nun eine Modellierungsstruktur zur Verfügung steht. Aus praktischer Sicht besteht die Tendenz, die Umkehrung des

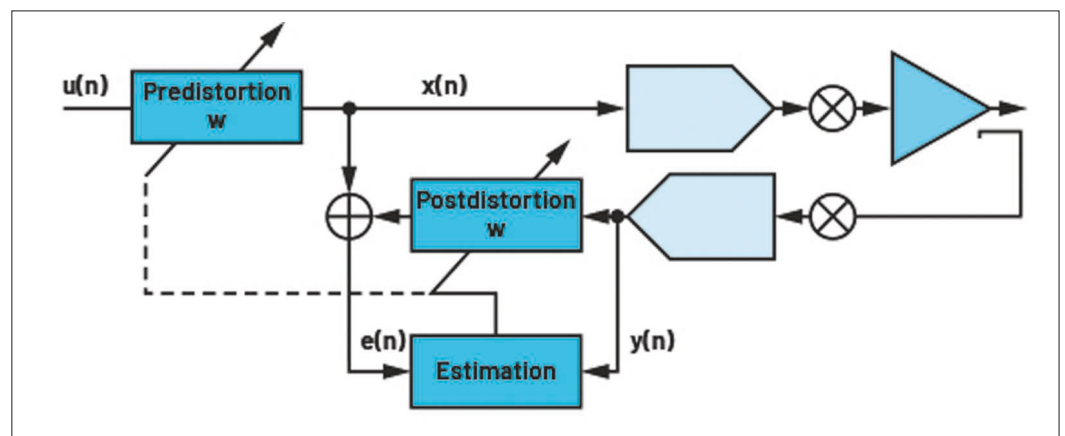


Bild 7: Das Blockdiagramm zeigt die indirekte Implementierung der Modellierung und Predistortion

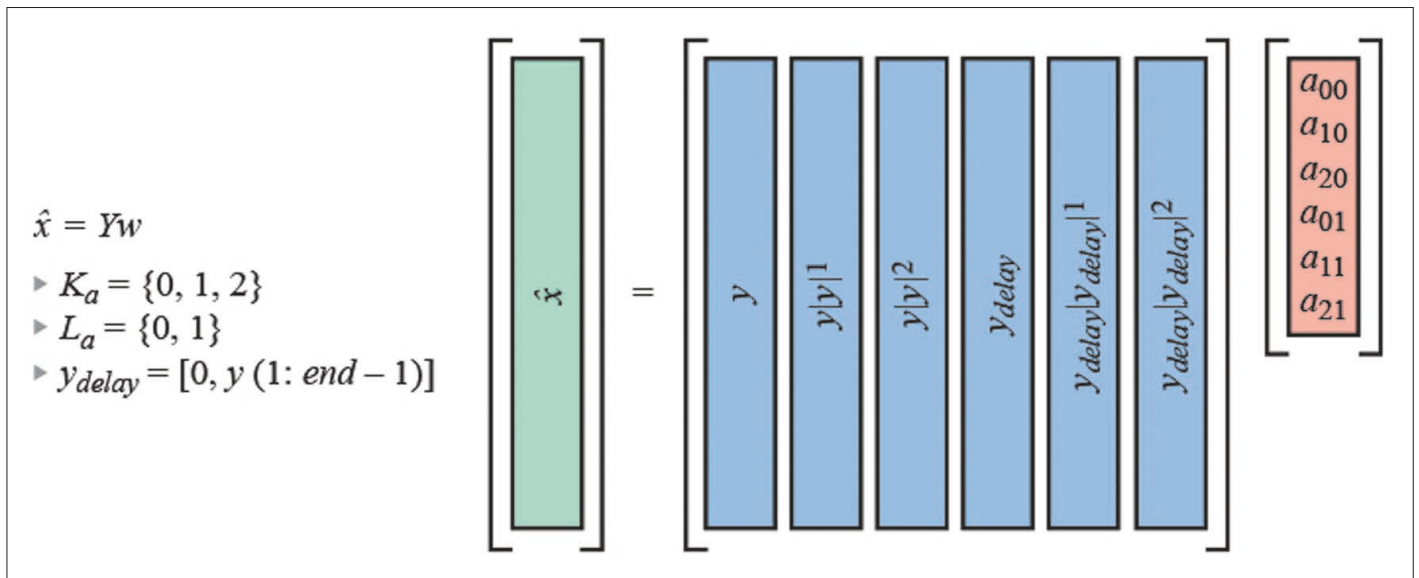


Bild 8: Umgekehrte Näherungsgleichung in Matrixform. Hier wurde ein Speicher eingefügt

oben beschriebenen Modells zu berechnen. Es stellt sich heraus, dass diese Modellkoeffizienten eine schöne Reziprozität aufweisen, da dieselben Gewichtungen zur Nachverzerrung des erfassten PA-Ausgangsvektors zur Beseitigung von Nichtlinearitäten und zur Vorverzerrung des durch die PA gesendeten Sendesignals verwendet werden können, damit der PA-Ausgang so linear wie möglich erscheint. Bild 7 zeigt ein Blockdiagramm zur Bestimmung der Gewichtungskoeffizienten und der Vorverzerrung.

Für das inverse Modell wird die in Bild 6 dargestellte Matrixgleichung vertauscht, so dass $X^{\wedge} = Yw$ entsteht. Hier wird die Matrix Y auf die gleiche Weise gebildet wie X im anderen Fall, wie in Bild 8 dargestellt. In diesem Beispiel wurde ein Speicherterm eingefügt und die Anzahl der einbezogenen Polynomordnungen wurde reduziert. Y ist nicht quadratisch (es handelt sich um eine hohe, schlanke Matrix), so dass dies mithilfe der „Pseudo-Inversen“ der Matrix erreicht wird:

$$w = (Y^H Y)^{-1} Y^H x$$

Dies löst w im Sinne der kleinsten Quadrate, d.h., es minimiert das Quadrat der Differenz zwischen X^{\wedge} und Yw , was ja gewollt ist!

Dies kann noch etwas verfeinert werden, um zu berücksichtigen, dass es in einer Applikation mit sich ändernden Signalen angewendet wird. Hier werden die Koeffizienten eingeschränkt, indem sie ausgehend von ihrem vorherigen Wert aktualisiert werden. μ ist ein konstanter Wert zwischen 0 und 1, der bestimmt, wie stark sich die Gewichtungen pro Iteration ändern können. Ist $\mu = 1$ und $w_0 = 0$, so kehrt diese Gleichung sofort zur grundlegenden Lösung der kleinsten Quadrate zurück. Wenn μ auf einen Wert kleiner als 1 gesetzt wird, dauert es eine Anzahl von Iterationen, bis die Koeffizienten konvergieren.

$$w_{i+1} = w_i + \mu(Y^H Y)^{-1} Y^H e, e = x - \hat{x}$$

Beachten Sie, dass die hier beschriebenen Modellierungs- und Abschätztechniken nicht die einzigen Möglichkeiten sind, DPD durchzuführen. Techniken wie die Modellierung auf der Grundlage der

dynamischen Abweichungsreduzierung (Dynamic Deviation Reduction, DDR) können anstelle von oder zusätzlich zu dieser verwendet werden. Die beschriebenen Abschätzungsverfahren zur Lösung der Koeffizienten können ebenfalls auf zahlreiche andere Arten durchgeführt werden.

Implementierung

DPD wird im digitalen Basisband implementiert, im Allgemeinen in einem Mikroprozessor oder einem FPGA. Die Radio-Verse-Transceiver-Produkte von ADI, wie z. B. die ADRV902x-Familie, verfügen über integrierte Mikroprozessorkerne, deren Struktur speziell für eine einfache DPD-Implementierung ausgelegt ist.

Die DPD-Implementierung in eingebettete Software umfasst zwei Aspekte. Der erste ist der DPD-Aktuator, in dem die Vorverzerrung der live übertragenen Daten in Echtzeit durchgeführt

wird, und der zweite ist die DPD-Adaption Engine, in der die DPD-Koeffizienten auf der Grundlage von Beobachtungen des PA-Ausgangs aktualisiert werden.

Der Schlüssel dazu, wie DPD in Echtzeit implementiert wird, ist die Verwendung von Lookup-Tabellen (LUTs). Diese ermöglichen es, teure Laufzeitberechnungen durch ein einfacheres Array-Indexing zu ersetzen. Betrachten wir also nun, wie der DPD-Aktuator eine Vorverzerrung auf ein übertragenes Datensample anwendet. Die Notation ist in Bild 7 dargestellt, wobei $u(n)$ das zu übertragende Rohdatensample ist und $x(n)$ die vorverzerrte Version. Bild 9 zeigt die Berechnungen, die erforderlich sind, um ein vorverzerrtes Sample für ein bestimmtes Szenario zu erhalten. Es handelt sich hierbei um ein relativ begrenztes Beispiel mit der höchsten Polynomordnung dritter Ordnung und nur einem Speicherabgriff

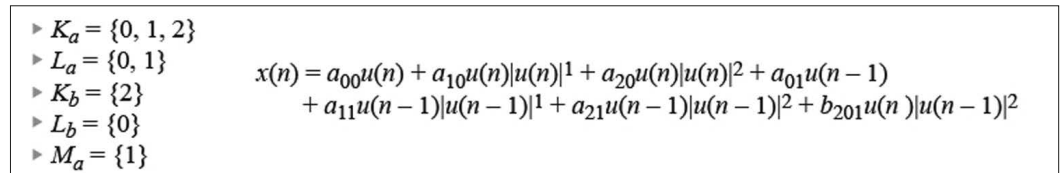


Bild 9: Vorverzerrungsberechnung für den Fall dritter Ordnung mit einem Speicherabgriff und einem Crossterm-Element dritter Ordnung

$$x(n) = u(n) [a_{00} + a_{10}|u(n)|^1 + a_{20}|u(n)|^2] + u(n-1) [a_{01} + a_{11}|u(n-1)|^1 + a_{21}|u(n-1)|^2] + u(n) [b_{201}|u(n-1)|^2]$$

LUT1

LUT2

LUT3

Bild 10: Neugruppierung von Gleichungselementen zur Darstellung der LUT-Struktur

sowie einem einzigen Crossterm. Selbst in diesem Fall sind eindeutig viele Multiplikations-, Potenz- und Additionsberechnungen erforderlich, um dieses eine Daten-Sample zu erhalten.

Hier kommen die LUTs ins Spiel, um die Echtzeitberechnungen zu erleichtern. Die Gleichung in Bild 9 kann in die Gleichung in Bild 10 umgeschrieben werden, wobei die Daten, die in die LUTs eingegeben werden, deutlicher werden. Jede LUT enthält das Ergebnis des hervorgehobenen Elements der Gleichung für eine große Anzahl möglicher Werte für $|u(n)|$. Die Auflösung hängt von der Größe der LUTs ab, die in der verfügbaren Hardware implementiert werden kann. Der Betrag des aktuellen Eingangssamples wird in Abhängigkeit von der Auflösung der LUT quantisiert und als Index für den Zugriff auf das richtige LUT-Element für den jeweiligen Eingang verwendet.

Bild 11 zeigt, wie die LUTs in die vollständige Implementierung des Vorverzerrungs-Aktuators für unser Beispiel eingebunden sind. Man beachte, dass dies nur eine von vielen möglichen Implementierungen ist. Ein Beispiel für eine Änderung, die unter Beibehaltung des gleichen Ausgangs vorgenommen werden könnte, ist die Verschiebung des Verzögerungselements z^{-1} auf die rechte Seite von LUT2.

Die Adaption-Engine hat die Aufgabe, die Koeffizienten zu ermitteln, die zur Berechnung der LUT-Werte im Aktuator verwendet werden. Dazu muss der in den Gleichungen 1 und 2 beschriebene w -Vektor gelöst werden. Die Pseudoinversen-Matrix-Operation $(Y^H Y)^{-1} Y^H$ ist sehr rechenintensiv. Gleichung 1 kann wie folgt umgeschrieben werden:

$$Y^H Y w = Y^H x$$

Ist $C_{YY} = Y^H Y$ und $C_{Yx} = Y^H x$, ergibt sich für Gleichung 3:

$$C_{YY} w = C_{Yx}$$

C_{YY} ist eine quadratische Matrix und kann mit Hilfe der Cholesky-Zerlegung in das Produkt aus einer oberen Dreiecksmatrix L und ihrer konjugierten Transponierung ($C_{YY} = L^H L$) zerlegt werden. Dies ermöglicht es uns, nach w aufzulösen, indem wir eine Dummy-Variable z einführen und sie wie gezeigt auflösen:

$$L^H z = C_{Yx}$$

Dann setzen wir diese Dummy-Variable wieder ein und lösen sie auf für:

$$L w = C_{Yx}$$

Da es sich bei L und L^H um obere bzw. untere Dreiecksmatrizen handelt, lassen sich Gleichung 5 und Gleichung 6 leicht und mit minimalem Rechenaufwand lösen, um w zu erhalten. Jedesmal, wenn die Adaption-

Engine läuft und neue Werte für w gefunden werden, müssen die Aktuator-LUTs entsprechend aktualisiert werden. Die Adaption kann in regelmäßigen oder unregelmäßigen Abständen erfolgen nach Beobachtung des PA-Ausgangs oder Änderungen des zu übertragenden Signals.

Die Implementierung von DPD in ein eingebettetes System erfordert eine Reihe von Überprüfungen und Abgleichen, um dessen Stabilität zu gewährleisten. Es ist äußerst wichtig, dass die übertragenen Datenpuffer und die Daten des Erfassungspuffers zeitlich aufeinander abgestimmt sind, um sicherzustellen, dass die zwischen ihnen hergestellte mathematische Beziehung korrekt ist und bei ihrer Anwendung über die Zeit Bestand hat. Wenn diese Anpassung verlorengeht, werden die von der Adaption-Engine zurückgegebenen Koeffizienten das System nicht korrekt vorverzerrten, was zu Instabilität im System führen kann. Der vorverzerrte Aktuatorausgang sollte auch überprüft werden, um sicherzustellen, dass das Signal den DAC nicht in die Sättigung bringt.

Die Studie [2] ist eine gute Quelle für die Anwendung von DPD in einem drahtgebundenen Kommunikationssystem mit ultraweiter Bandbreite.

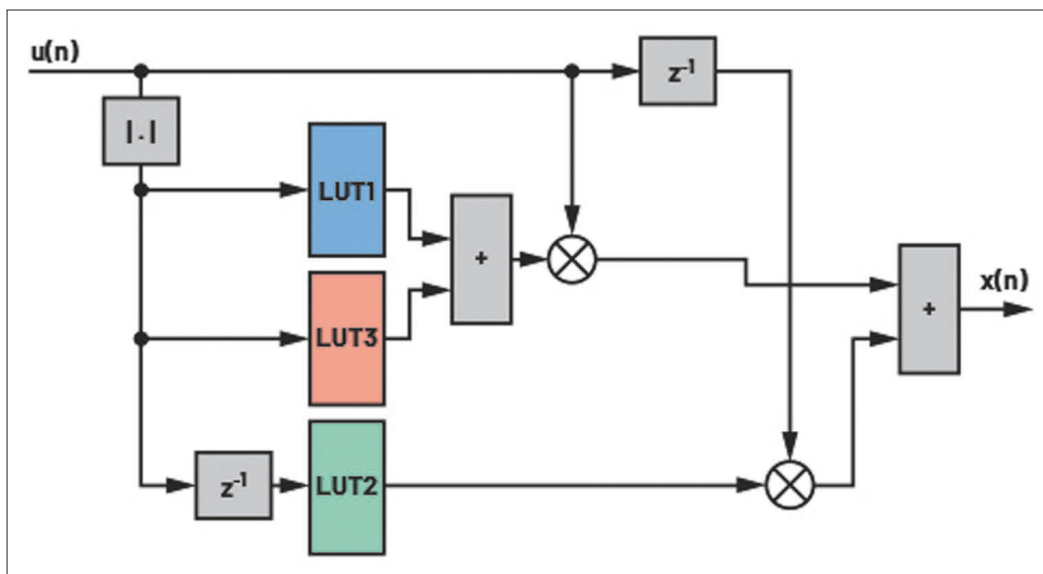


Bild 11: Blockdiagramm der möglichen Implementierung von DPD mittels LUT

Referenzen

[1] Dennis R. Morgan, Zhenxiang Ma, Jaehyeong Kim, Michael G. Zierdt and John Pastalan: "A Generalized Memory Polynomial Model for Digital Predistortion of RF Power Amplifiers" IEEE Transactions on Signal Processing, Vol. 54, No. 10, October 2006

[2] Patrick Pratt and Frank Kearney: "Ultrawideband Digital Predistortion (DPD): The Rewards (Power and Performance) and Challenges of Implementation in Cable Distribution Systems" Analog Dialogue, Vol. 51, No. 3, July 2017 ◀